Methodological report of the JEP-CEV-HRDAG Joint Project on data integration and statistical estimation*

March 6, 2025†

Contents

*The following researchers from participated in this project. CEV: Paula Amado, William Acero, Camilo Argoty, Giovanny Babativa, Luz Karime Bernal, Alejandro Castro, María Juliana Durán, Esteban Espinosa, Valentina Gómez, Andrea González, and María Alejandra Ortíz. HRDAG: Patrick Ball, Maria Gargiulo, Valentina Rozo, and Tarak Shah. JEP: Juan Guillermo Dueñas, Luis Ernesto Jaimes, Camila Lozano, Mary Julieth Murillo, and Harbey Rodríguez.

†This report is a translation of the Spanish report dated March 6, 2025 and available here.

1 Introduction

Since mid-2020, the Colombian Truth Commission (*Comisión para el Esclarecimiento de la Verdad, la Convivencia y la No repetición* - CEV), the Special Jurisdiction for Peace (*Jurisdicción Especial para la Paz* - JEP), and the Human Rights Data Analysis Group (HRDAG) joined the project *Integration of data and statistical estimates of victims in the context of the armed conflict.* The objective of this project was to provide the entities of the Comprehensive System of Truth, Justice, Reparation and Non-Repetition (*Sistema Integral de Verdad, Justicia, Reparación y No Repetición* - SIVJRNR) with solid scientific arguments to examine the magnitude of violence by estimating levels of underreporting and identifying victimization patterns of homicides,[1] enforced disappearances, kidnappings, forced displacement, and child recruitment.

This project consisted of four components:

1. Linking records from databases about the armed conflict in Colombia;
2. Creating estimates and analysis of underreporting related to violations of Human Rights and International Humanitarian Law;
3. Advising the implementation of data analysis related to these estimates;
4. Strengthening capacity within the entities of the SIVJRNR for contrasting data sources and implementing statistical models that strengthen the analysis of data about human rights.

The project was developed in four phases. Phase 1 concluded on December 18, 2020 and Phase 2 on April 14, 2021. In September 2021, the Phase 3 report, which detailed the methodologies implemented to date, was presented. Finally, this report presents updates to the Phase 3 report, including the inclusion of new databases, methodological decisions, and analyses.

1.1 What is our goal?

The foundation for understanding human rights violations is documentation that, in recent decades, has been done using databases. It is possible to choose which cases to register and each project has a universe of interest defined by the criteria it chooses. These may include a focus on an ethnic group or a geographic location, recording only dead bodies or cases reported to the police, cases that have been investigated, cases reported to the media, cases of people seeking reparation for their losses in the conflict, or cases investigated in detail to study their relationship with the armed conflict. These are just some of the ways in which the creators of the databases define their universe of interest. In the last 40 years many people, organizations, and State institutions have created hundreds of databases with information on violence in Colombia.

We usually want to know about patterns: was there more violence in 2001 or in 2002? Which armed group committed the highest proportion of homicides? However, any database, even the best ones, contains only one image of the world according to the criteria defined for that project. A database can tell us whether people *reported* more homicides to this database in 2001 or 2002, but the patterns in the database are not necessarily related to the patterns in the world. Perhaps, the researchers in 2001 were very active and in 2002 they were replaced by others with less interest in doing fieldwork, which is why we saw the peak in 2001. This peak would tell us more about the project, not about the violence. Each database is finite and limited by the project's budget, its ability to reach affected communities and gain their trust, and by the ability of its researchers to record all the information they receive. The databases are necessarily partial, including only some events that are part of the universe of interest. This is not a criticism of any database. All of the databases underreport victims and answer a key question: what have we documented? A good database allows us to know what is in it.

[1] "Homicide" is understood as the intentional death of persons by one of the actors in the armed conflict. It is also understood as the deaths of civilian persons in the course of hostilities between armed actors, either during combat or indiscriminate attacks, or by the activation of explosive devices, anti-personnel mines, or other devices. In this project it was not possible to discriminate between deaths of civilians and those of combatants that occurred during the armed conflict. This analysis is planned for future projects.

Statistical arguments require something more. We use statistics to learn about the world, not about patterns in the data. Statistical arguments require that we go beyond the data to estimate what is true in the world. With limited data we cannot know the truth of the world, so we have to approximate using the data we have, and probability models created and tested in the last hundreds of years. We have to consider the missing information from all available records and the entire universe of documented and undocumented cases. Again, our data is limited, so our result is not a single number, it is a range in which the true answer is likely to be. An estimate gives up the precise definition of what a database contains in order to approximate the truth in the world. Both methods have important roles in understanding the world: good databases tell us about the cases they contain, while good estimates tell us about the world. We need both.

Our goal with this project is to create a statistical analysis of the patterns of armed conflict in Colombia. By statistical analysis, we mean that we want to take into account two fundamental components of rigorous mathematical measures: bias and variance. Bias, in a technical sense, is the difference between our estimate and reality. We cannot measure it because we do not know exactly what the truth is. However, we have over a century of mathematics, statistical theory, and practice showing us how to use data in statistical models to minimize bias. Variance is the measure of how accurate our estimates are. Variance is a formal, mathematical measure of uncertainty, expressed in an interval. For example, in the joint project we found that there were more than 450,000 homicides in the armed conflict. More formally, we find that we have records of more than 450,000 victims with first and last names in one or more databases, who we estimate were victims of the armed conflict. This last part, "we estimate they were victims of the armed conflict", is not accurate because there are thousands of victims for whom we do not know whether their homicides occurred or not within the context of the conflict. So, the formal expression of our estimate is based on the information that we have and its use to consider other information that we do not have about exactly which cases are or are not related to the armed conflict. We estimate that there are between 436,747 and 464,585 homicide victims related to the armed conflict.

Including information about whether or not a record is part of the armed conflict is only the first step. There are many more victims who were not documented or who were documented with only partial information. For example, we have cases without the name of the victim (such as "peasant from farm X") or records of bodies that were not identified. We know that there are many more bodies hidden in clandestine graves, burned, or in rivers. Some of those homicides are misclassified as accidents and many are not documented because the families of the victims or witnesses decided, probably with reason, that reporting the death could create a problem for them. Considering all the information we have and using statistical models that have been used to estimate human populations since the mid-20th century, we estimate that there were more than 800,000 homicides in the context of the armed conflict. Since it is an estimate, it has uncertainty interval between 777,852–852,756.

What do we mean by an "estimate"? We have described two types of estimates that we use in our project. The first is called "statistical imputation" (note that this has nothing to do with imputation in legal terms). Many of our records have missing information about important fields. Continuing with the previous example, we have 374,567 records for which we are sure that the homicide is related to the armed conflict and 75,752 records for which we are sure that the homicide is not related to the armed conflict. However, we have another 104,050 records for whom we do not know if their homicide occurred in the context of the armed conflict or not.

Why should we make estimates? We might decide to calculate the statistics only with the 374,567 victims for which we are certain. But, ignoring the 104,050 victims for which we are uncertain makes a very strong assumption. In statistics this assumption is known as "missing **completely** at random" and it means that if we use only the 374,567 homicides we know occurred in the context of the armed conflict, we are assuming that the proportion of the 104,050 unknown cases pertaining to the armed conflict is identical to the proportion we observe. This shows that in every analysis, without exception, we make of a statistic is based on a model, even if the analysts do not explicitly mention it.

The problem of assuming that the missing data is missing completely at random is even more complicated when we analyze a second variable. For example, if we want to understand the relationship of violence over time by comparing homicides in 2002 with homicides in 2015. If we ignore the 104,050 cases for which we do not know whether they occurred in the context of the armed conflict or not, we are assuming that if we included these 104,050 cases as having occurred in the context of the armed conflict, our understanding of the relationship between these two years would not change. This assumption is not correct, as we show in Section 6.1. Whereas in 2002 almost all cases that were part of the armed conflict were documented as such, in 2015 we found that a substantial fraction of the homicides for which we do not have information on their relationship to the armed conflict are probably related to it. Estimating the level of underreporting is critical because the cases least likely to be documented are the cases of those most likely to be excluded from the data: people from rural areas, from minority ethnic groups, with low incomes, and from families who have safety concerns. Statistical models try to recognize these missing cases, as we show in the findings. If we cannot document each victim, we should at least try to estimate how many there are and that their suffering is not totally silenced.

Different projects have different criteria for including data. These criteria reflect the ideas of the project about what they are trying to learn and express and, therefore, the statistics that come from each project are usually different (see for example the comparison of the Unique Registry for Victims and the National Center for Historic Memory in Suárez 2018). In our project, we try to create statistical analyzes that include all victims of homicide, kidnapping, enforced disappearance and child recruitment during the armed conflict. We want to present our estimates as scientifically rigorous, which means that we have to use a model to reduce bias and describe the variance. In doing so, there are two types of missing information that we have to address. The first type of missing information are the missing fields, which are not recorded in each case, such as whether or not the event occurred within the context of the armed conflict or the presumed perpetrator. We work on this gap using statistical imputation which is explained in Section 4. The second type missing information is missing records—victims' whose human rights violations have not been documented. For this type of missing information we estimate the levels of underreporting as described in Section 6.

We have many datasets and many victims appear in more than one of them. The deduplication is perhaps the biggest part of the project. The records that refer to the same victim are called "coreferent records". Some databases also contain more than one record for the same victim, which may be a result of how the information was extracted from the original database, or because the project was unable to identify the same victims. We handle both forms of deduplication at the same time, as explained in Section 3. One observation is that sources tend to group together. That is, groups of sources often document the same cases, with relatively few cases in common with other sources. We call this analysis "documentation patterns" and we show it in Section 5.

1.2 Data overview

The CEV and the JEP received hundreds of data sets from multiple entities or social and victim organizations. Since this project requires information at the victim level, after reviewing and validating the data, deterministic and probabilistic techniques were implemented which allowed the consolidation of the records into a database that integrates the records of a total of 112 data sets. They provided information on homicide, enforced disappearance, kidnapping, child recruitment, forced displacement, and exile[2] of civil society organizations, victims' collectives, and official institutions. We limit ourselves to these acts of violence because we focus on those that people usually experience only once. Displacement is the exception, which we include because of the large amount of data.

This is the first project that integrates such a large number of databases with a human rights approach. With a quick review, you can see that the databases span different time periods, regions, victim populations, and thus tell different stories. Each database has partial information, and therefore also biases. This is not a criticism of the databases, on the contrary, it is a basic characteristic of any information-gathering process. The entities or organizations that contributed information to the SIVJRNR and therefore to the project are:

1. Agency for Reintegration and Normalization (*Agencia para la Reincorporación y la Normalización* - ARN);

[2] Only the CEV database provided information on victims of exile, so it was not possible to carry out data integration or estimates of underreporting.

2. Codebac Association (*Asociación Codebac*);
3. Colombian Association of Victims of Kidnapping and Enforced Disappearance (*Asociación Colombiana de Víctimas de Secuestro y Desaparición Forzada* - ACOMIDES);
4. National Association of Peasant Users (*Asociación Nacional de Usuarios Campesinos* - ANUC);
5. Nuevo Amanecer Victims' Association (*Asociación de Víctimas Nuevo Amanecer*); 6.Colombian Chancellery (*Cancillería de Colombia*);
6. Center for Indigenous Cooperation (*Centro de Cooperación al Indígena* - CECOIN);
7. National Center for Historical Memory (*Centro Nacional de Memoria Histórica* - CNMH);
8. Figures and Concepts (*Cifras y Conceptos*);
9. José Alvear Restrepo Lawyers' Collective (*Colectivo de Abogados José Alvear Restrepo* - CAJAR);
10. Colombian Commission of Jurists (*Comisión Colombiana de Juristas* - CCJ);
11. Commission for the Clarification of Truth, Coexistence and Non-Repetition (*Comisión para el Esclarecimiento de la Verdad, la Convivencia y la No repetición* - CEV);
12. Committee of Solidarity with Political Prisoners (*Comité de Solidaridad con los Presos Políticos* - CSPP);
13. Black communities of the Domingodó River (*Comunidades Negras de la cuenca del río Domingodó)*;
14. Presidential Council for Human Rights and International Affairs (*Consejería Presidencial para los Derechos Humanos y Asuntos Internacionales* - SIDDHH);
15. National Council for the Fight against Kidnapping and other Attacks against Personal Liberty (*Consejo Nacional de Lucha contra el Secuestro y demás Atentados contra la Libertad Personal* - CONASE);
16. Coordination Colombia Europe United States (*Coordinación Colombia Europa Estados Unidos* - CCEEU)
17. Affirmative Caribbean Corporation (*Corporación Caribe Afirmativo*);
18. Corporation for the Defense and Promotion of Human Rights Reiniciar (*Corporación para la Defensa y Promoción de los Derechos Humanos Reiniciar*);
19. Regional Corporation for the Defense of Human Rights (*Corporación Regional para la Defensa de los Derechos Humanos* - CREDHOS);
20. Administrative Department of Security (*Departamento Administrativo de Seguridad* - DAS).
21. Ombudsman's Office (*Defensoría del Pueblo*);
22. National Union School (*Escuela Nacional Sindical*);
23. Colombian Federation of Education Workers (*Federación Colombiana de Trabajadores de la Educación* - FECODE);
24. Office of the Attorney General of the Nation (*Fiscalía General de la Nación* - FGN);
25. Foging Futures Foundation (*Fundación Forjando Futuros*);
26. Bonds of Dignity Foundation (*Fundación Lazos de Dignidad*);
27. Free Country Foundation (*Fundación País Libre*);
28. Universities Teachers Group (*Grupo Docentes Universidades*);
29. Colombian Institute of Family Welfare (*Instituto Colombiano de Bienestar Familiar* - ICBF);
30. Institute of Studies for Development and Peace (*Instituto de Estudios para el Desarrollo y la Paz* - INDEPAZ);
31. National Institute of Legal Medicine and Forensic Sciences (*Instituto Nacional de Medicina Legal y Ciencias Forenses* - INML);
32. Special Jurisdiction for Peace (*Jurisdicción Especial para la Paz* - JEP);
33. Military Criminal Justice (*Justicia Penal Militar* - JPM);
34. Ministry of National Defense (*Ministerio de Defensa Nacional*);
35. Southern Colombian Observatory for Human Rights, Peace, and Territory (*Observatorio Surcolombiano de Derechos Humanos, Paz y Territorio* - OBSURDH);
36. Land Observatory (*Observatorio de Tierras*);
37. National Indigenous Organization of Colombia (*Organización Nacional Indígena de Colombia* - ONIC);
38. Medellin Ombudsman (*Personería de Medellín*);
39. Office of the General Inspector of the Nation (*Procuraduría General de la Nación* - PGN);
40. National Police (*Policía Nacional* - PONAL);
41. International Society for Human Rights (*Sociedad Internacional para los Derechos Humanos* - SIDDHH);

42. We are Defenders (*Somos Defensores*);
43. Special Administrative Unit for Comprehensive Care and Reparation for Victims (*Unidad Administrativa Especial para la Atención y Reparación Integral a las Víctimas*)

The importance of this project relies on the fact that the records of human rights violations suffer from two types of missing information: missing fields and underreporting.[3] Therefore, it is not correct to analyze patterns of violence from databases that register victims. The observed data does not always reflect reality, but rather the documentation work done by a certain project.

The first type of missing information is missing fields, which are limited to the documented or recorded data. All the databases have different fields: some are related to the event (such as the year, municipality, department, and perpetrator), while others are related to the person (first name, last name, age, gender, and ethnicity, among others). However, when organizations or institutions document acts of violence, they do not always have complete information and it is normal for some fields to remain empty for some records. There are many situations in which this can occur. For example, there may be a record of a victim, but the exact day on which the event occurred is not known. Or, perhaps, we have the name and last name, but not the ID (*cédula* in Colombia). In some other cases, the ethnicity may not be known, among other reasons. In addition to empty fields, it is also possible that a field may contain incorrect information, such as an error in the *cédula* or a spelling error in the first name or last name.

The second type of missing information is underreporting. Underreporting goes beyond the recorded data and occurs because there are victims who are not documented. This can occur for a number of reasons. For example, it may be because the victim or their relatives are afraid to file a complaint and opt for silence or that the act of violence took place in a very distant place, where the organizations or institutions cannot reach. It can also occur because although they were victims within the context of the armed conflict, a different kind of violence was registered instead of the one they suffered or because their bodies were found but have not been identified. It might be that their bodies were thrown into rivers or common graves and there is no record. Unlike the first type of missing information, in this case it is not known how many records are missing.

Despite joining the 112 sources of information, these types of missing information are present and both types of missingness—missing fields or underreporting— can be a source of bias. As a result, it is necessary to find a way to address the missing information, which becomes a challenge for research.

To correct the two types of missing information it is necessary to follow three major steps. First, integrate and deduplicate existing information. Second, statistically impute the missing fields. That is, use statistical methods that allow inference of plausible values for the characteristics of the victims that have missing fields. Third, estimate the universe of victims, including underreporting, using multiple systems estimation.

1.3 Workflow

The "quantum" workflow defined by HRDAG was used in the project. According to it, all tasks must be self-contained and self-documented. The project has six large tasks, as shown in Figure 1. It begins with the "individual" task in which the data is processed for each of the data sets from each organization. There data is imported, cleaned, standardized, filtered, and exported. Once the variables and the content of the different databases are standardized, in the sense of having the same names for the same variables, the same use of dates, etc., there are two parallel tasks. On one hand, there is the record linkage task. As the name suggests, this task links the records from the various sources. This process is also known as deduplication. On the other hand, there is the "support variables" task, in which a neural network is trained to generate variables that will help with the next task. When these two tasks are finalized, the next step is statistical imputation. With the results of the imputation, it is possible to make estimates. Finally, from all the previous tasks, we arrive at the results.

[3] More details on this can be seen in this video.

The original and intermediate data were stored on a CEV server. We used a GitHub repository for version control, but it is worth noting that the repository contains only the code and not the data. The repository has two levels.

- On the first level there are the tasks. For example, record linkage, imputation, and multiple systems estimation (MSE)
- On the second level, within each of these tasks, there are at least three directories: input, src, and output. The input includes the files to be read. src includes the scripts that are to be executed on the input files. The results of the src scripts are written to the output directory. The input and output information were only stored on the CEV server, while the contents of src were available on GitHub. In addition, in each task there is a Makefile, which guarantees replicability.

Having the contents of the tasks in these three directories guarantees that each task is self-contained. Also, the tasks are self-documenting because everything someone would need to understand them is in the files. More information about the workflow is available in Ball (2016) and Shah (2019).

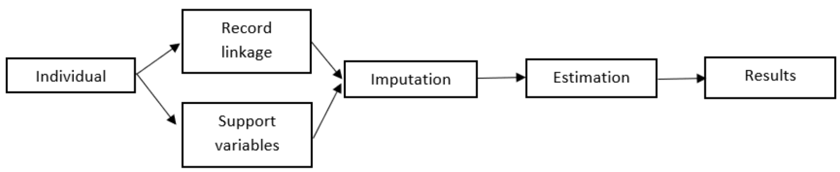

Figure 1: Workflow.

1.4 Summary of this report

This report is divided into 9 sections, including this introduction. In the second section, we present the statistical findings of the project. In the third section, we detail the record linkage or deduplication process. This task is the one that integrates the information from the 112 data sets and is divided into 6 steps, which are presented in their respective order. In the fourth section, we present the statistical imputation of missing fields, which corrects the first type of missing information. In addition, we present how we work with specialized and non-specialized databases about the armed conflict, as well as with enforced and non-enforced disappearances. In the fifth section, we show the documentation patterns after statistical imputation. In other words, the armed conflict victims documented by different organizations. In the sixth section, we present the method implemented to estimate the underreporting of populations by multiple systems estimation. This is the method we use to fill the second type of missing information (underreporting). In the seventh section we present two characteristics of the estimates: bias and variance. In the eighth section we show the limitations and outline directions for future work. Finally, in the ninth section, we conclude with the report annexes.

2 Statistical findings

In this section we present the findings of the project. Each of the types of violence corresponds to a different period of time depending on the availability of data, as follows: enforced disappearance (1985–2016), forced displacement (1985–2019), homicide (1985–2018), child recruitment (1990–2017), and kidnapping (1990–2018).

We will present three different ways of understanding the data. First, the observed data, which is limited by the two types of missing information (missing fields and missing records) previously discussed. Second, statistically imputed data, which solves the problem of missing fields with a measure of the uncertainty. Third, population estimates, which address the two types of missing data (missing fields and missing records) and allow us to calculate the uncertainty that comes from them. Note that for displacement, statistical imputation or estimates are not included. This is because of the dominance of Single Registry of Victims (*Registro Único de Víctimas* - RUV) in the documenting of victims of this kind of violence. Throughout the chapters of the Final Report, the most probable statistic for statistical imputation has been presented. In Section 4 we describe the process in which we make 10 replicates, which are then combined. The most probable data is the mode of the 10 replicates.

To begin, let's focus on the observed data. That is, in the second column of each table. The first thing that becomes evident is that homicides are the main documented kind of violence that occurred in the armed conflict (excluding forced displacement) with 374,567 victims in Table 2, followed by enforced disappearance with 110,351 victims in Table 1, kidnapping with 46,319 victims in Table 4, and child recruitment with 14,669 in Table 3.

The only types of violence that occurred within the context of the armed conflict by definition are child recruitment and forced displacement. So, for the other kinds of violence it was necessary to impute the variable "is conflict" to determine whether a victim was a victim of a human rights violation in the context of the armed conflict. In addition, for enforced disappearance it was necessary to statistically impute the variable "is enforced disappearance" and for the child recruitment it was necessary to statistically impute the age variable.

When we do the statistical imputation to address the problem of missing fields, the total number of documented victims changes for some types of violence. For example, in the case of enforced disappearance, we observed 110,351 cases of disappearances identified as enforced disappearance in the context of the armed conflict, while 14,539 disappearances were documented without information on whether they were enforced or if they belong to the armed conflict.[4] In statistical imputation, the imputation model assigns values to each of the missing fields. After filling in the missing values in all the records, we can use them to recalculate quantities of interest. We estimate that there is an average of 121,768 victims of forced enforced disappearance related to the armed conflict that corresponds to the most probable number of documented victims. Statistical imputation also measures the uncertainty in the number of enforced disappearances, so the tables also show the range (with lower and upper bounds). In the case of enforced disappearance, Table 1 shows the range: 119,777–123,759.

When statistically imputing the missing fields, in the third column of the tables, it is interesting that the lower limit that is furthest from the observed value is for homicide. This would indicate that this is the act of violence to which most victims are attributed, of whom it was not known whether or not they were related to the armed conflict. This is followed by enforced disappearance and kidnapping. Based on the statistical imputation, in Colombia there would be 450,666 victims of homicide, 121,768 victims of enforced disappearance, 50,770 victims of kidnapping, and 16,238 victims of child recruitment.

In the group of columns on the right side of each table we show the estimates of the total number of victims, including those who were not documented in the sources used in this project. The best way to understand these estimates is to consider their range. For example, for homicides, the interval shown in Table 2 is between 777,852 and 852,756 victims. Regarding enforced disappearances, Table 1 shows that in Colombia there would be between 204,395 and 225,410 victims. Tables 3 and 4 show the estimates of victims of kidnapping child recruitment victims, respectively.

Finally, regarding the estimates, we can see that homicides are the main kind of violence with between 777,852 and 852,756 victims. For enforced disappearances, in Colombia there would be between 204,395 and 225,410 victims. For child recruitment, the victims would be between 27,101 and 40,828. For kidnapping, the victims would be between 74,768 and 92,849. Finally, by deduplicating the registries, in Colombia there would be 7,752,964 victims of displacement with first name, last name, place, and date.

[4] See Section 4.3 for a discussion of our process of identifying these conditions.

Table 1: Distribution of the number of victims of enforced disappearance according to their relationship to the armed conflict (1985–2016)[a]

	Observed	Imputed			Estimated		
Relationship to the armed conflict	N.	N(inf)	N(imp)	N(sup)	N(inf)	N(est)	N(sup)
Victims related to the armed conflict	110,351	119,777	121,768	123,759	204,395	214,418	225,410
Victims not related to the armed conflict	47,207	48,337	50,328	52,319	-	-	-
No Information	14,539	-	-	-	-	-	-

[a] This table includes only the victims related to the armed conflict according to the defined rules. It is also important to note that the words: inf, sup indicate the credible interval: inferior limit (from the imputation or the estimate) and superior limit (from the imputation or the estimate).

Table 2: Distribution of the number of victims of homicide according to their relationship to the armed conflict (1985–2018)[a]

	Observed	Imputed			Estimated		
Relationship to the armed conflict	N.	N(inf)	N(imp)	N(sup)	N(inf)	N(est)	N(sup)
Victims related to the armed conflict	374,567	436,747	450,666	464,585	777,852	813,707	852,756
Victims not related to the armed conflict	75,752	89,784	103,703	117,622	-	-	-
No Information	104,050	-	-	-	-	-	-

[a] It is important to note that the words: inf, sup indicate the credible interval: inferior limit (from the imputation or the estimate) and superior limit (from the imputation or the estimate).

Table 3: Distribution of the number of victims of child recruitment according to their relationship to the armed conflict (1990–2017)[a]

	Observed	Imputed			Estimated		
Relationship to the armed conflict	N.	N(inf)	N(imp)	N(sup)	N(inf)	N(est)	N(sup)
Victims related to the armed conflict	14,669	16,027	16,238	16,449	27,101	32,806	40,828

[a] By definition all the children recruitment are related to the armed conflict. Yet, this table includes the imputation of the age variable, since it includes only the victims under the age of 18 at the moment of their recruitment.
[b] It is important to note that the words: inf, sup indicate the credible interval: inferior limit (from the imputation or the estimate) and superior limit (from the imputation or the estimate).

Table 4: Distribution of the number of victims of kidnapping according to their relationship to the armed conflict (1990–2018)[a]

	Observed	Imputed			Estimated		
Relationship to the armed conflict	N.	N(inf)	N(imp)	N(sup)	N(inf)	N(est)	N(sup)
Victims related to the armed conflict	46,319	50,234	50,770	51,306	74,768	82,765	92,849
Victims not related to the armed conflict	4,647	4,647	5,102	5,638	-	-	-
No Information	4,906	-	-	-	-	-	-

[a] It is important to note that the words: inf, sup indicate the credible interval: inferior limit (from the imputation or the estimate) and superior limit (from the imputation or the estimate).

Table 5: Distribution of the number of victims of forced displacement according to their relationship to the armed conflict (1985–2019)

	Observed
Relationship to the armed conflict	N.
Victims related to the armed conflict	7,752,964

In the graphs of Figure 2 we find the observed (black line), statistically imputed (blue line) and estimated (green area) victims for the four types of violence for which we used multiple systems estimation. The y-axis shows the number of victims, while the x-axis shows the year. The green dots show instances where the the variance is so large that it does not fit on the graph. The first thing we observe is that the estimates follow the trends of documented victims. When analyzing each of the acts of violence, the scales of the vertical axis should be noted. As we mentioned before, the magnitudes of homicide are much higher than those of other acts of violence (except forced displacement).

For all acts of violence, there is an evident peak of violence at the beginning of 2000. This peak could have been particularly high for enforced disappearance, child recruitment, and kidnapping, in which the variance is so wide that it is not presented in the graph. Also, it is worth noting that the cases of child recruitment and kidnapping in recent years have reached almost zero and underreporting has been reduced. However, cases persist and, especially for homicides, underreporting is still present.

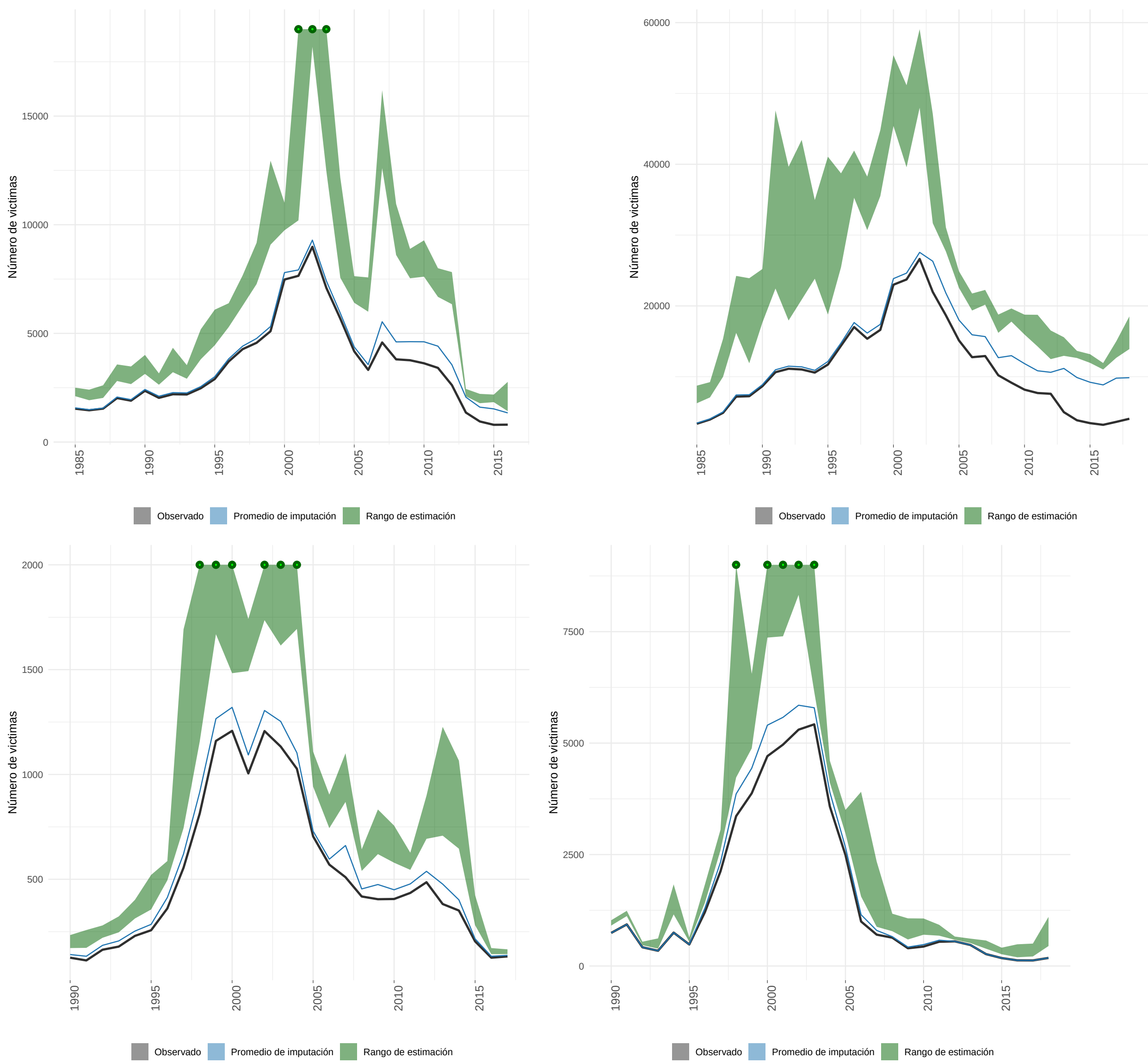

Figure 2: Observed, imputed, and estimated victims of enforced disappearance (uper-left), homicide (upper-right), child recruitment (lower-left) and kidnapping (lower-right) by presumed perpetrator

Figure 3 shows the 10 main departments for each type of violence. It is clear that Antioquia ranked first in three types of violence (enforced disappearance, homicide, and kidnapping). It is worth noting, in addition to Antioquia, that Cesar, Meta, Nariño, Norte de Santander, and Valle del Cauca are among the 10 departments with the most victims for the four types of violence. Also, for each kind of violence the order of the departments is approximately the same no matter what data we use (observed, statistically imputed, or estimated). For example, if we consider only the observed data (black bar), Antioquia has the highest number of homicides. This is also the case with the statistical imputation data (blue bar) and the estimates of undocumented victims (green bar).

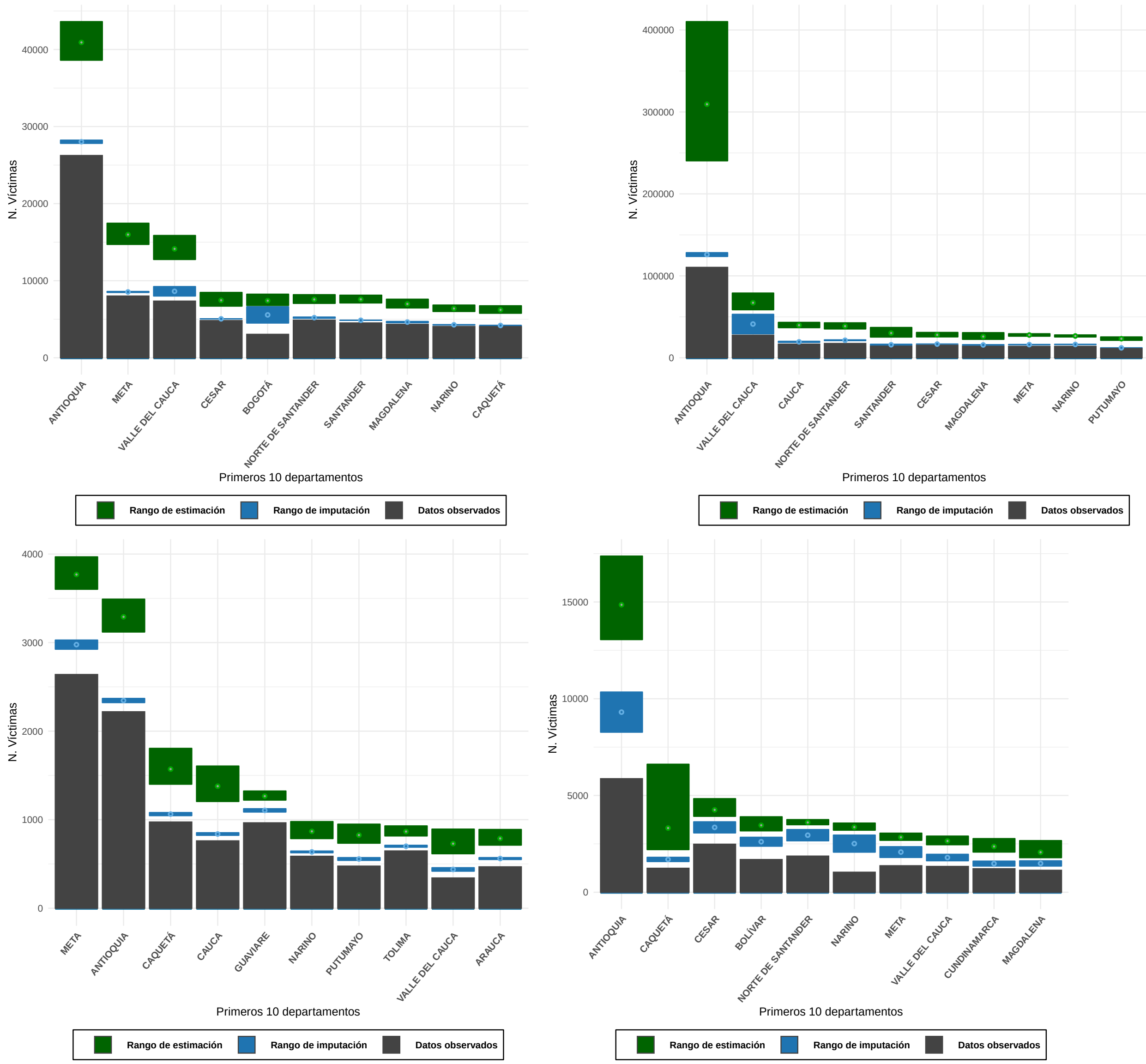

Figure 3: 10 main departments of enforced disappearance (upper-left), homicide (upper-right), child recruitment (lower-left) and kidnapping (lower-right)

In the following tables we observe the sex of the victims for the different types of violence. It is evident that men were the main direct victims of all the acts of violence. For the observed data we have the following results. The proportion of male victims of forced displacement is 0.48. For enforced disappearance, its proportion would be 0.85 of the victims, while in homicide it would be 0.91, in recruitment 0.69 and in kidnapping 0.78.

Regarding the statistical imputation, we find the following. In enforced disappearance, the proportion of men would be between 0.82 and 0.84 of the victims, while in homicide it would be between 0.88 and 0.93, for child recruitment 0.68 and 0.72, and kidnapping between 0.77 and 0.79. In the case of displacement—using the integrated data—the proportion of male victims would be between 0.48 and 0.48.

Now, for the estimates. We find that in the case of enforced disappearance, the proportion of men would be between 0.79 and 0.87 of the victims, while for homicide they would be between 0.88 and 0.87, for child recruitment 0.55 and 1, and for kidnapping between 0.69 and 0.87. In the case of displacement and based on the integrated and statistically imputed data, the proportion of male is 0.48.

Table 6: Displacement victims by sex (1985–2019)[a]

	Observed		
Sex	N.	P(NA)	P
FEMALE	3,994,536	0.52	0.52
MALE	3,750,968	0.48	0.48
Obs. with NAs	7,460	0.00	-
Total	7,752,964	1.00	1.00

[a] The total row shows the columns that can be added and whose result should be 1

Table 7: Enforced disappearance victims by sex (1985–2016)

	Observed			Imputed						Estimated					
Sex	N.	P(NA)	P	N(inf)	N(imp)	N(sup)	P(inf)	P(imp)	P(sup)	N(inf)	N(est)	N(sup)	P(inf)	P(est)	P(sup)
MALE	92,934	0.84	0.85	100,088	101,292	102,496	0.82	0.83	0.84	161,406	169,535	178,545	0.79	0.83	0.87
FEMALE	16,662	0.15	0.15	19,591	20,477	21,363	0.16	0.17	0.18	32,310	34,922	37,994	0.16	0.17	0.19
Obs. with NAs	755	0.01	-	-	-	-	-	-	-	-	-	-	-	-	-
Total	110,351	1.00	1.00	-	121,769	-	-	1.00	-	-	-	-	-	1.00	-

[a] This table includes only the observations that were imputed as part of the armed conflict and as enforced disappearances (IS CONFLICT and IS FORCED DIS).
[b] It is important to note that the columns: inf , med and sup indicate the credible interval. inf: inferior bound (from the imputation or the estimate), P(med) mean of the imputation and P(est): mean of the estimation and sup: superior bound (from the imputation or the estimate)
[c] The imputation includes missing values from the variables SEX and IS CONFLICT
[d] The total row shows the columns that can be added and whose result should be 1.

Table 8: Homicide victims by sex (1985–2018)[a]

	Observed			Imputaded						Estimaded					
Sex	N.	P(NA)	P	N(inf)	N(imp)	N(sup)	P(inf)	P(imp)	P(sup)	N(inf)	N(est)	N(sup)	P(inf)	P(est)	P(sup)
MALE	333,633	0.89	0.91	396,135	408,680	421,225	0.88	0.91	0.93	684,144	716,851	752,722	0.88	0.92	0.97
FEMALE	33,509	0.09	0.09	40,178	41,986	43,794	0.09	0.09	0.10	54,390	60,692	69,424	0.07	0.08	0.09
Obs. with NAs	7,425	0.02	-	-	-	-	-	-	-	-	-	-	-	-	-
Total	374,567	1.00	1.00	-	450,666	-	-	1.00	-	-	-	-	-	1.00	-

[a] This table includes only the observations that were imputed as part of the armed conflict.
[b] It is important to note that the columns: inf , med and sup indicate the credible interval. inf: inferior bound (from the imputation or the estimate), P(med) mean of the imputation and P(est): mean of the estimation and sup: superior bound (from the imputation or the estimate)
[c] The imputation includes missing values from the variables SEX and IS CONFLICT
[d] The total row shows the columns that can be added and whose result should be 1.

Table 9: Child recruitment victims by sex (1990–2017)

	Observed			Imputed						Estimated					
Sex	N.	P(NA)	P	N(inf)	N(imp)	N(sup)	P(inf)	P(imp)	P(sup)	N(inf)	N(est)	N(sup)	P(inf)	P(est)	P(sup)
MALE	8,489	0.58	0.69	10,977	11,314	11,651	0.68	0.7	0.72	16,270	21,262	29,915	0.55	0.72	1.00
FEMALE	3,825	0.26	0.31	4,558	4,924	5,290	0.28	0.3	0.33	7,326	8,290	9,547	0.25	0.28	0.32
Obs. with NAs	2,355	0.16	-	-	-	-	-	-	-	-	-	-	-	-	-
Total	14,669	1.00	1.00	-	16,238	-	-	1.0	-	-	-	-	-	1.00	-

[a] This table includes only the observations that were imputed as part of the armed conflict.
[b] It is important to note that the columns: inf , med and sup indicate the credible interval. inf: inferior bound (from the imputation or the estimate), P(med) mean of the imputation and P(est): mean of the estimation and sup: superior bound (from the imputation or the estimate)
[c] The imputation includes missing values from the variables SEX and IS CONFLICT
[d] The total row shows the columns that can be added and whose result should be 1.

Table 10: Kidnapping victims by sex (1990–2018)

	Observed			Imputed						Estimated					
Sex	N.	P(NA)	P	N(inf)	N(imp)	N(sup)	P(inf)	P(imp)	P(sup)	N(inf)	N(est)	N(sup)	P(inf)	P(est)	P(sup)
MALE	35,526	0.77	0.78	38,989	39,479	39,969	0.77	0.78	0.79	53,484	59,415	67,296	0.69	0.76	0.87
FEMALE	10,033	0.22	0.22	11,086	11,291	11,496	0.22	0.22	0.23	16,514	18,309	20,570	0.21	0.24	0.26
Obs. with NAs	760	0.02	-	-	-	-	-	-	-	-	-	-	-	-	-
Total	46,319	1.01	1.00	-	50,770	-	-	1.00	-	-	-	-	-	1.00	-

[a] This table includes only the observations that were imputed as part of the armed conflict.
[b] It is important to note that the columns: inf , med and sup indicate the credible interval. inf: inferior bound (from the imputation or the estimate), P(med) mean of the imputation and P(est): mean of the estimation and sup: superior bound (from the imputation or the estimate)
[c] The imputation includes missing values from the variables SEX and IS CONFLICT
[d] The total row shows the columns that can be added and whose result should be 1.

Finally, we present the findings by the presumed armed actor or perpetrator. The first thing that becomes evident is that the paramilitaries are the main actors presumed responsible for both enforced disappearances and homicides, while the FARC are the main actor presumed responsible for child recruitment. When we study the aggregate, it is evident that these two groups occupy the first and second place in each of the types of violence and the ranges of the estimates do not overlap. In other words, for enforced disappearances and homicides, the paramilitaries were the main armed actor, followed by the FARC. While for child recruitment, the FARC was the main armed actor, followed by the paramilitaries. It should be noted that the relative order of responsibility is consistent between the observed data, the statistical imputation, and the estimate of the total population of victims, including those that were not documented by any data source.

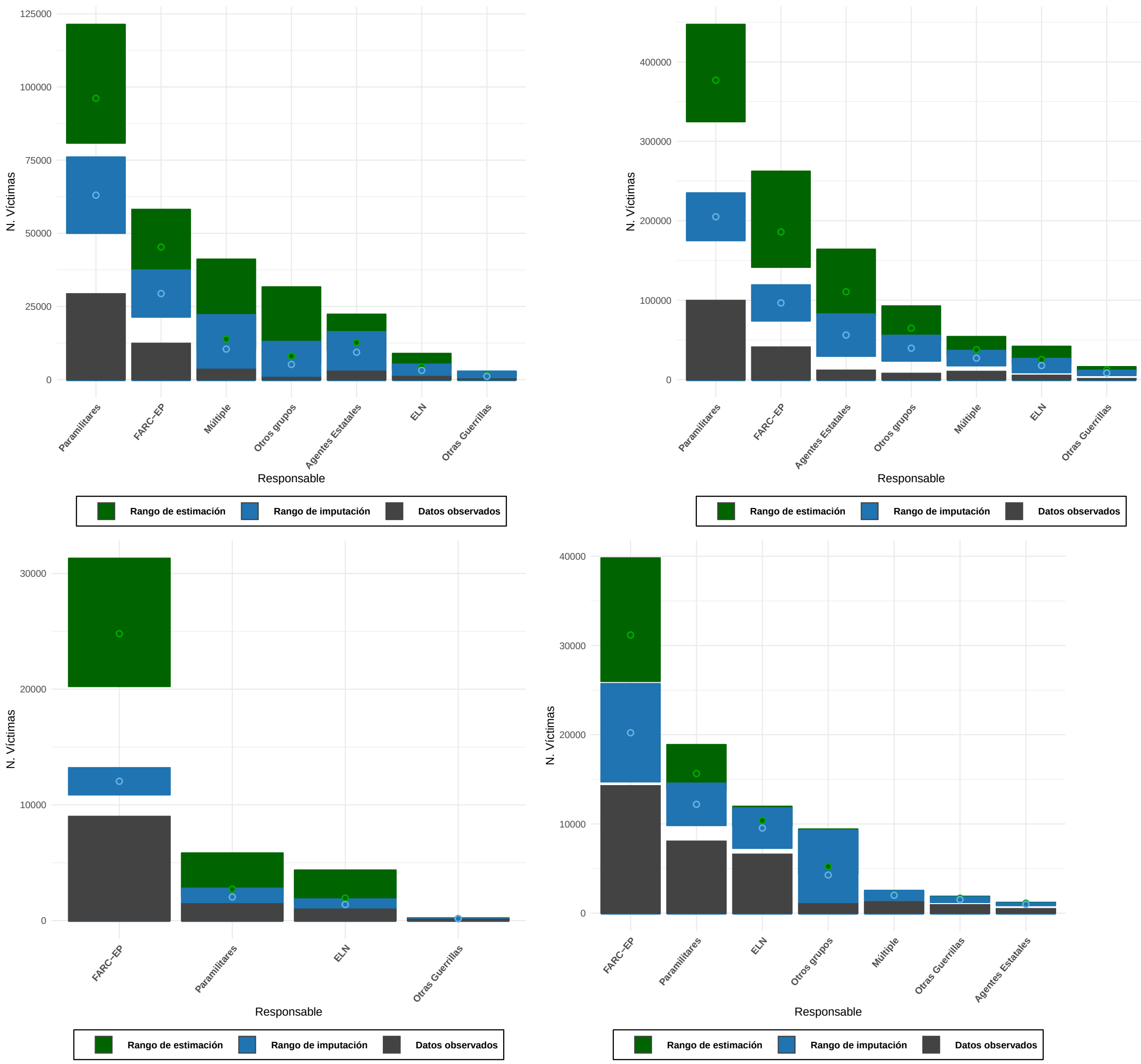

Figure 4: Presumed perpetrator of enforced disappearance (upper-left), homicide (upper-right), child recruitment (lower-left) and kidnapping (lower-right) for observed, imputated and estimated victims

In the following graphs, we analyze the timeline of each of the types of violence by perpetrator. It is striking that there seems to be a direct and positive relationship between the different armed actors. That is, they follow similar trends over time. For homicide, there is evidence of a peak in violence allegedly committed by state agents in 2007, while violence by paramilitaries and the FARC was on the decline. Regarding child recruitment, the FARC's peak of violence occurred prior to that of the paramilitaries, ELN and other guerrillas. Although the variance is wide, it would seem that there was a peak by the paramilitaries and the FARC between 2010 and 2013. Finally, for kidnapping, some peaks are observed in the 1990s by the FARC and the ELN. Since 1996, there has been an increase in cases, which reached their peak at the beginning of 2000 for different actors. This is the type of violence in which the role of the other actors is seen the most, with a high variance. Yet, they are present in different years. As was the case with other analyses, the pattern of the observed data follows the pattern of statistical imputation and estimation of the entire victim population.

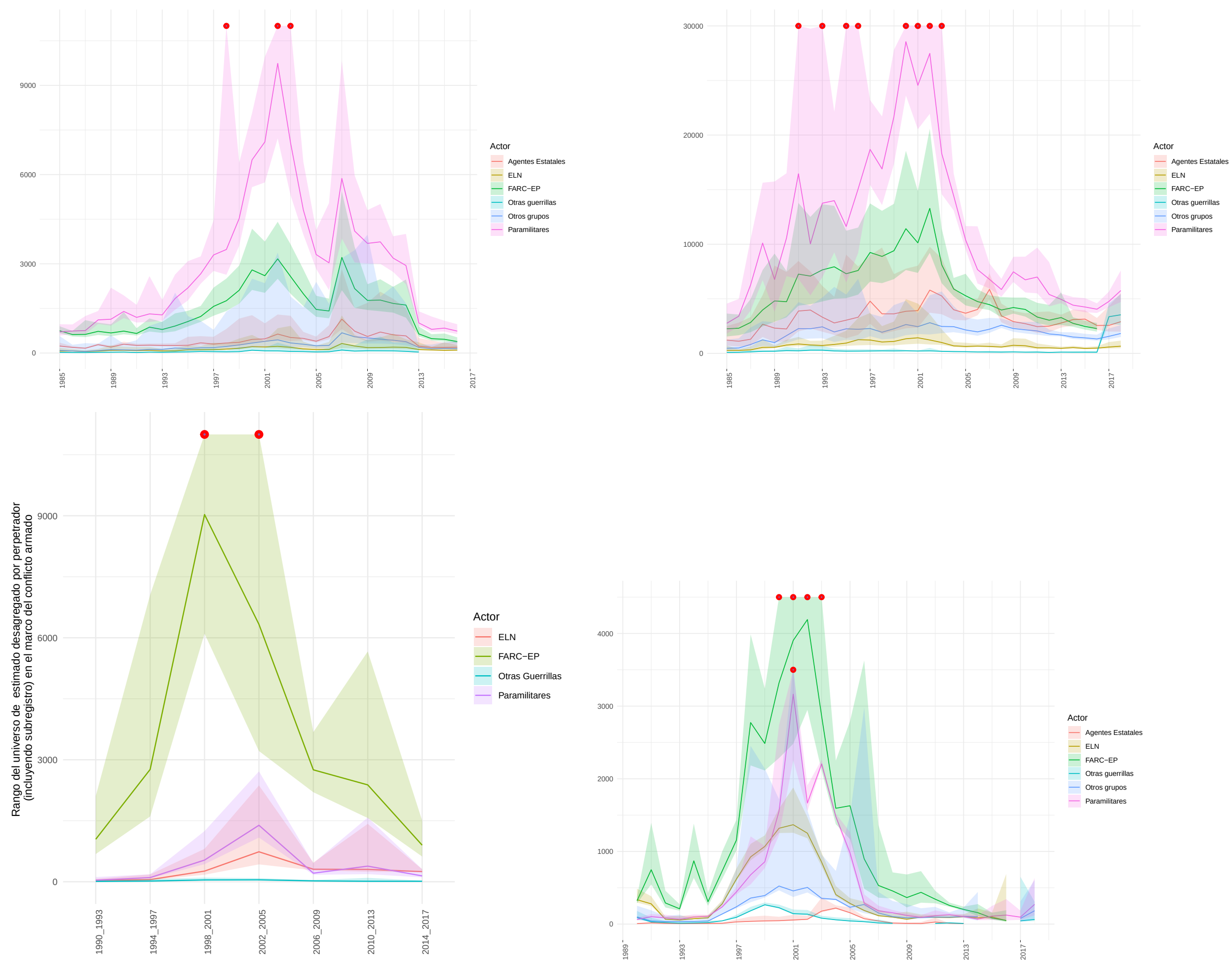

Figure 5: Estimates of enforced disappearance (upper-left), homicide (upper-right), child recruitment (lower-left) and kidnapping (lower-right) by presumed prepatrator

The following tables show the presumed responsibilities by armed group. When analyzing the data integration for displacement and records with information of the presumed perpetrator, the unspecified guerrillas would be the main actor with a proportion of 0.51, followed by paramilitaries with 0.37.

It is interesting that for enforced disappearances and homicide the proportion of responsibilities observed is, respectively, 0.58 and 0.56 for paramilitaries, 0.25 and 0.23 for FARC, and 0.06 and 0.07 for state agents. That is, the proportion of these two events in terms of observed data is similar. On the other hand, for child recruitment, the FARC's responsibility is 0.76, for paramilitaries is 0.12, and the ELN's is 0.08. While for kidnapping the responsibility of the FARC is 0.44, the paramilitaries' is 0.25, and the ELN's is 0.2. For displacement, the other guerrillas are presumably responsible for 0.51, followed by paramilitaries with 0.37.

The observed data shows the importance of using statistics. The tables clearly show the high proportion of data without information on the perpetrator. This is a critically missing field in displacement (0.67), enforced disappearance (0.54) and homicide (0.52), while in kidnapping it reaches 0.29.

Regarding the statistical imputation for enforced disappearance, the presumed perpetrator would continue to be the paramilitaries (0.41–0.62) with more than double that of the FARC (0.18–0.31). When making the estimate for enforced disappearance, the main actor would continue to be the paramilitaries (0.44–0.67) with more than double that of the FARC (0.21–0.32).

It is worth noting how the variance in responsibilities increases when we use estimates. For example, the responsibility of the multiple actors and State increases between the observed and estimated data, going from 0.07 to between 0.06 and 0.23 and between 0.05 and 0.12, respectively. For its part, that of "other groups" increases from 0.02 to between 0.03 and 0.17.

Regarding the presumed responsibility of homicide with imputed data, first there would be the paramilitaries with (0.39–0.52), the FARC with (0.16–0.27) and the State with (0.07–0.18). While when making the estimates we have that the paramilitaries would be the main presumed perpetrator with (0.4–0.55), followed by the FARC (0.17–0.32) and the State (0.1–0.2).

There are two points to note. First, the order of the groups presumed to have perpetrated the violence from highest to lowest is the same for the statistical imputation and the estimate of the population of all the victims: first the paramilitaries, followed by the FARC and state agents. Neither statistical imputation nor estimation change this main finding. Second, by comparing the intervals of the statistical imputation and the estimate, we see that the intervals do not intersect. For example, the interval of the statistical imputation for the paramilitaries does not intersect with the interval of the FARC. This means that the two estimates are statistically significantly different, that is, there is statistical evidence to affirm that they are different.

Regarding recruitment of children and adolescents, the FARC is the main actor in both observed and statistically imputed and estimated data. In the observed data their responsibility is 0.76. When making the statistical imputation, the proportion would be between 0.67 and 0.82. When estimating, the proportion would oscillate between 0.67 and 1. They would be followed by the paramilitaries with between 0.07 and 0.19.

Regarding kidnapping, the FARC would again be the main presumed perpetrator. In the observed data they would be responsible for the 0.44. While in the statistically imputed data its responsibility would be between 0.29 and 0.51. Regarding the estimate, the responsibility of the FARC would have a range between 0.39 and 0.59, followed by the paramilitaries with between 0.21 and 0.28, and the ELN with between 0.15 and 0.18.

Table 11: Presumed perpetrator of displacement (1985–2019)[a]

	Observed		
Presumed perpetrator	N.	P(NA)	P
Obs. with NAs	5,208,856	0.67	-
Unspecified guerrillas	1,309,028	0.17	0.51
Paramilitary	947,073	0.12	0.37
Multiple	126,922	0.02	0.05
Other groups	114,883	0.01	0.05
FARC-EP	37,542	0.00	0.01
State	7,326	0.00	0.00
ELN	1,334	0.00	0.00
-	7,752,964	0.99	0.99

[a] The row with total shows the columns that can be added and whose result should be 1
[b] As was shown in the documentation patterns, most of the records of displacement come from the RUV. In many cases there are unspecified guerrillas.

Table 12: Presumed perpetrator of enforced disappearance (1985–2016)[a]

	Observed			Imputed						Estimated					
Presumed perpetrator	N.	P(NA)	P	N(inf)	N(imp)	N(sup)	P(inf)	P(imp)	P(sup)	N(inf)	N(est)	N(sup)	P(inf)	P(est)	P(sup)
Paramilitary	29,368	0.27	0.58	49,960	63,029	76,098	0.41	0.52	0.62	80,789	96,131	121,418	0.44	0.53	0.67
FARC-EP	12,451	0.11	0.25	21,330	29,410	37,490	0.18	0.24	0.31	37,643	45,301	58,213	0.21	0.25	0.32
Multiple	3,566	0.03	0.07	3,566	10,448	22,239	0.00	0.09	0.18	10,523	13,844	41,199	0.06	0.08	0.23
State	2,907	0.03	0.06	2,907	9,359	16,442	0.02	0.08	0.14	9,888	12,662	22,384	0.05	0.07	0.12
Other groups	770	0.01	0.02	770	5,200	13,089	0.00	0.04	0.11	5,284	8,009	31,723	0.03	0.04	0.17
ELN	1,120	0.01	0.02	1,120	3,194	5,320	0.01	0.03	0.04	3,288	4,304	8,993	0.02	0.02	0.05
Other guerrillas	275	0.00	0.01	275	1,128	2,949	0.00	0.01	0.02	1,143	1,437	2,788	0.01	0.01	0.02
Obs. with NAs	59,894	0.54	-	-	-	-	-	-	-	-	-	-	-	-	-
-	110,351	1.00	1.01	-	121,768	-	-	1.01	-	-	-	-	-	1.00	-

[a] This table includes only the observations that have been imputed as part of the armed conflict and enforces disappearences (IS CONFLICT and IS FORCED DIS).
[b] It is important to note that the columns: inf , med and sup, indicate the credible interval. inf: inferior bound (from the imputation or the estimate), P(med) mean of the imputation and P(est): mean of the estimation and sup: superior bound (from the imputation or the estimate)
[c] The imputation includes the missing values for PRESUMED PERPETRATOR, IS CONFLICT, and IS FORCED DIS.
[d] The row with total shows the columns that can be added and whose result should be 1
[e] The estimates by presumed perpetrator have a small downward bias.

Table 13: Presumed perpetrator of homicide (1985–2018)

	Observed			Imputed						Estimated					
Presumed perpetrator	N.	P(NA)	P	N(inf)	N(imp)	N(sup)	P(inf)	P(imp)	P(sup)	N(inf)	N(est)	N(sup)	P(inf)	P(est)	P(sup)
Paramilitary	99,837	0.27	0.56	174,867	205,028	235,189	0.39	0.45	0.52	324,602	376,928	447,276	0.40	0.46	0.55
FARC-EP	41,161	0.11	0.23	73,746	96,592	119,438	0.16	0.21	0.27	141,374	185,934	262,475	0.17	0.23	0.32
State	12,008	0.03	0.07	29,364	56,094	82,824	0.07	0.12	0.18	80,926	110,652	164,263	0.10	0.14	0.20
Other groups	8,048	0.02	0.05	23,252	39,606	55,960	0.05	0.09	0.12	50,265	64,799	92,879	0.06	0.08	0.11
Multiple	10,508	0.03	0.06	17,347	27,124	36,901	0.04	0.06	0.08	30,503	37,695	54,384	0.04	0.05	0.07
ELN	5,635	0.02	0.03	8,668	17,725	26,782	0.02	0.04	0.06	19,482	25,306	42,107	0.02	0.03	0.05
Other guerrillas	1,473	0.00	0.01	5,097	8,496	11,895	0.01	0.02	0.03	10,233	12,391	16,439	0.01	0.02	0.02
Obs. with NAs	195,897	0.52	-	-	-	-	-	-	-	-	-	-	-	-	-
-	374,567	1.00	1.01	-	450,665	-	-	0.99	-	-	-	-	-	1.01	-

[a] This table includes only the observations that have been imputed as part of the armed conflict (IS CONFLICT).
[b] It is important to note that the columns: inf , med and sup, indicate the credible interval. inf: inferior bound (from the imputation or the estimate), P(med) mean of the imputation and P(est): mean of the estimation and sup: superior bound (from the imputation or the estimate)
[c] The imputation includes the missing values for PRESUMED PERPETRATOR and IS CONFLICT.
[d] The row with total shows the columns that can be added and whose result should be 1.

Table 14: Presumed perpetrator of child recruitment (1990–2017)

	Observed			Imputed						Estimated					
Presumed perpetrator	N.	P(NA)	P	N(inf)	N(imp)	N(sup)	P(inf)	P(imp)	P(sup)	N(inf)	N(est)	N(sup)	P(inf)	P(est)	P(sup)
FARC-EP	9,013	0.61	0.76	10,857	12,038	13,219	0.67	0.75	0.82	20,227	24,798	31,322	0.67	0.82	1.00
Paramilitary	1,457	0.10	0.12	1,457	2,038	2,804	0.09	0.13	0.17	2,101	2,714	5,851	0.07	0.09	0.19
ELN	997	0.07	0.08	997	1,391	1,876	0.06	0.09	0.12	1,441	1,938	4,373	0.05	0.06	0.14
Other guerrillas	116	0.01	0.01	131	174	217	0.01	0.01	0.01	142	172	242	0.00	0.01	0.01
Obs. with NAs	2,777	0.19	-	-	-	-	-	-	-	-	-	-	-	-	-
-	14,636	0.99	1.00	-	16,129	-	-	1.00	-	-	-	-	-	1.00	-

[a] This table includes only the observations that have been imputed as part of the armed conflict (IS CONFLICT).
[b] It is important to note that the columns: inf , med and sup, indicate the credible interval. inf: inferior bound (from the imputation or the estimate), P(med) mean of the imputation and P(est): mean of the estimation and sup: superior bound (from the imputation or the estimate)
[c] The imputation includes the missing values for PRESUMED PERPETRATOR, IS CONFLICT, and AGE.

Table 15: Presumed perpetrator of kidnapping (1990–2018)[a]

	Observed			Imputed						Estimated					
Presumed perpetrator	N.	P(NA)	P	N(inf)	N(imp)	N(sup)	P(inf)	P(imp)	P(sup)	N(inf)	N(est)	N(sup)	P(inf)	P(est)	P(sup)
FARC-EP	14,291	0.31	0.44	14,709	20,223	25,737	0.29	0.40	0.51	25,968	31,180	39,826	0.39	0.46	0.59
Paramilitary	8,077	0.17	0.25	9,818	12,200	14,582	0.19	0.24	0.29	13,825	15,644	18,905	0.21	0.23	0.28
ELN	6,618	0.14	0.20	7,264	9,538	11,812	0.14	0.19	0.23	9,774	10,378	11,988	0.15	0.15	0.18
Other groups	1,059	0.02	0.03	1,059	4,276	9,323	0.00	0.08	0.18	4,380	5,235	9,440	0.07	0.08	0.14
Multiple	1,317	0.03	0.04	1,450	2,009	2,568	0.03	0.04	0.05	2,029	2,076	2,178	0.03	0.03	0.03
Other guerrillas	958	0.02	0.03	1,195	1,519	1,843	0.02	0.03	0.04	1,569	1,673	1,903	0.02	0.02	0.03
State	514	0.01	0.02	840	1,005	1,170	0.02	0.02	0.02	1,053	1,114	1,222	0.02	0.02	0.02
Obs. with NAs	13,485	0.29	-	-	-	-	-	-	-	-	-	-	-	-	-
-	46,319	0.99	1.01	-	50,770	-	-	1.00	-	-	-	-	-	0.99	-

[a] This table includes only the observations that have been imputed as part of the armed conflict (IS CONFLICT).
[b] It is important to note that the columns: inf , med and sup, indicate the credible interval. inf: inferior bound (from the imputation or the estimate), P(med) mean of the imputation and P(est): mean of the estimation and sup: superior bound (from the imputation or the estimate)
[c] The imputation includes the missing values for PRESUMED PERPETRATOR and IS CONFLICT.
[d] The row with total shows the columns that can be added and whose result should be 1.

3 Record linkage

Given that the project's inputs are different data sets and that the same victim can be registered in more than one of them, it is necessary to link the records. This process is also known as "deduplication" and consists of creating a single record for each victim to avoid double counting. A summary of the extensive research on this topic is available in Herzog, Scheuren, and Winkler (2007).

Our deduplication process follows the steps described in Christen (2012). Specifically, we start with the imported of data. Second, we generate blocks of records, and third, we generate features. Fourth, we create a pairwise model. Fifth, we create groups records. Sixth, we join and export the result. Finally, we audit the results to demonstrate that the integration is correct. Below we explain each of these steps in detail.

3.1 Data preparation

Although many of the fields of the databases described in the appendix are shared across multiple files, it is still necessary to standardize the records we received. For example, it is possible that some databases have the departments in upper case, others in lower case, some with accents, etc. Also, the names of the variables can be different. For example, some databases may refer to the municipality code as `divipola` while others use `coddane`. With data standardization we keep the original information, but we make the necessary changes in the fields so that they are the same between databases. That is to say, that the variables that refer to the same thing have the same name and that the information is in the same format.

Once the data is standardized, the next step to be able to deduplicate is to have the minimum information necessary to identify whether or not two records refer to the same person. It is important to note that not all records have complete fields. For example, although it might be thought that a solution would be to join by *cédula* (the Colombian ID), this information is not always available. Therefore, we must define the minimum information necessary to identify whether or not two records refer to the same victim. In this case, we define the victim's name, victim's last name, and the department and year of the violence act as the minimum fields. In addition, the recorded instance of violence must be homicide, enforced disappearance, kidnapping, child recruitment, forced displacement, and/or exile to be included in this analysis. After filtering, we are only left with the records that have the minimum information necessary for deduplication and that are direct victims. At the end of this step the consolidated file from all data sets contains 12,863,977 records.

3.2 Generation of blocks of records

Intuitively, the goal of record linkage is to analyze each of the possible pairs of records and decide whether or not they are about the same person. This is called being "coreferent." However, there are too many records to compare all possible pairs. We have $\binom{n}{2}$ possible combinations. Since in this case our n is 12,863,977 records, we need to generate trillions of record pairs. This figure is almost the number of cells in the human body and is many trillions of times greater than what a supercomputer can process. To reduce the problem to a tractable size we must start by identifying only the pairs that have some chance of being coreferent, for which we will describe the generation of training data.

3.2.1 Generation of training data

Since it is not possible for a human or a computer to analyze all the possible pairs of records, we generate blocks based on different criteria and from these we identify the records that are similar and could correspond to the same person to use as training data.

In this case, the training data wre generated based on human expertise. Specifically, this person is called the "oracle" (Christen, Vatsalan, and Wang 2015). For this project, the oracle was HRDAG record-linkage expert Michelle Dukich, who reviewed hundreds of blocks from which the records were generated.[5]

[5]For more details see this presentation.

To reduce the magnitude of the task we selected groups of 50 to 200 records that share some characteristics. For example, records of people with the first name “MARÍA”, who have suffered any of the acts of violence in Antioquia in March 2001; or records of people who suffered homicide, whose first surname begins with “LOP”. The oracle separated these records into groups of records that, according to her expert judgment, refer to the same person. She did this based on human intuition and looking at the complete record information. In other words, it was not limited to the fact that the names were the same, but included an analysis of the names, place of events, date, etc.

All pairs of records referring to the same person within each group, called “positive pairs”, were extracted. In total, 1,082,479 positive pairs and 1,717,201 negative pairs were identified. This is the training data for the block generation model.

One possible concern would be that the training data was labeled incorrectly. To prove this, in December 2021 we asked 5 members of the Information Analysis Group (*Grupo de Análisis de la Información*) at the JEP to do the oracle exercise.[6] The decisions made by each of them were compared with the oracle's decisions. The results showed that the intersubjective agreement, that is, how much they agreed with the oracle, was greater than 0.9 in 4 of the cases.[7] When analyzing the records in which there were differences between the oracle and the analysts, the oracle had identified pairs of people that the JEP analysts did not. The analysts recognized that in these cases the records would refer to the same person.

3.2.2 Blocking

The blocking model uses the training data (labeled by the oracle) to find the “rules” that define whether or not two records correspond to the same person. In other words, it learns to imitate the decisions of the expert. Finding these blocks of potential pairs is a challenging problem in statistics and computer science.[8]

The “rules” are generated from the training data to create blocks of records that are very similar to each other. For example, because they share the three initial letters of the last name, the act of violence is the same, and it occurred on the same date. This rule combines three fields (last name, type of violence, and date), which must match for a pair of records to potentially correspond to the same person. This combination of fields is known as a *logical conjunction*.

Of course, each rule by itself covers only a few of all possible pairs. Therefore, more than one rule is added. The fact of having more than one rule is known as *logical disjunction*. The rules are added until all (or almost all) of the positive pairs of the training data are found within the logical disjunction. That is, until most of the positive pairs that were identified by the oracle in the training data are identified by the rules.

The rule generation task must be done carefully, otherwise it could include too many records. For example, if we consider pairs defined by having the same first and last names, there would be more than 1.5 trillion possible pairs for a computational capacity of approximately 30 million pairs. Then, a balance must be found between rules that include the largest number of possible pairs in an efficient way to be analyzed within computational capacity. The final disjunction of conjunctions is shown below.

[6] They analyzed 230,582, 185,187, 409,301, 197,983, and 354,279 pairs of records, respectively.

[7] The intersubjective agreement is calculated with Cohen's Kappa, which takes 1 as the maximum value. The Kappas were 0.968, 0.975, 0.972, 0.971, and 0.763. The fifth analyst only reached a value of 0.763, however, she recognized that she had not carried out the exercise with sufficient dedication.

[8] For a further explanation of the blocking model (sometimes called indexing) see Christen (2012), Chapter 4, Herzog, Scheuren, and Winkler (2007), Bilenko, Kamath, and Mooney (2006), Michelson and Knoblock (2006) and this blog post by HRDAG.

```
(apellido_1_meta1 ∧ nombre_1_meta2 ∧ synth_hecho_ymd) ∨
(cens_apellido_1 ∧ cens_nombre_1 ∧ nombre_2 ∧ sort_nm_1) ∨
(dept_hecho_desaparicion ∧ desaparicion_ymd ∧ sort_nm_1 ∧ sort_nm_1_last_3) ∨
(cens_sort_nm_1 ∧ dept_hecho_homicidio ∧ homicidio_ymd) ∨
(muni_hecho_reclutamiento ∧ nombre_2_last_3 ∧ sexo ∧ sort_nm_n ∧ sort_nm_n_meta2) ∨
(dept_hecho_homicidio ∧ homicidio_yrmo ∧ sort_nm_n_meta1 ∧ sort_nm_n_meta2) ∨
(secuestro_ymd) ∨
(apellido_1 ∧ homicidio_ymd) ∨
(cens_nombre_1 ∧ desaparicion_yrmo) ∨
(apellido_2_first_3 ∧ nombre_2_last_3 ∧ reclutamiento_yr ∧ sort_nm_n_meta2) ∨
(homicidio_ymd ∧ nombre_2_first_3) ∨
(apellido_2_first_3 ∧ desaparicion_ymd ∧ etnia ∧ nombre_1_first_3 ∧ nombre_1_last_3 ∧ nombre_2_meta2) ∨
(cens_sort_nm_1 ∧ sort_nm_n_meta2 ∧ synth_hecho_ymd) ∨
(desaparicion_ymd ∧ dob_ym ∧ nombre_2_meta2) ∨
(apellido_1 ∧ cens_apellido_1 ∧ nombre_2 ∧ secuestro_yrmo ∧ sort_nm_1_meta1) ∨
(apellido_2_first_3 ∧ desaparicion_ymd ∧ nombre_2_meta1) ∨
(apellido_2_last_3 ∧ cens_dob_ymd ∧ etnia ∧ muni_hecho_desaparicion ∧ nombre_2_first_3 ∧ nombre_2_last_3) ∨
(apellido_1_meta1 ∧ muni_hecho_desaparicion ∧ nombre_2_meta2) ∨
(cens_nombre_1 ∧ homicidio_ymd ∧ nombre_1_last_3) ∨
(apellido_2 ∧ cens_dob_ym ∧ dob_ymd ∧ nombre_2 ∧ sort_nm_n) ∨
(cens_apellido_1 ∧ cens_nombre_1 ∧ cens_sort_nm_1 ∧ homicidio_yrmo) ∨
(muni_hecho_secuestro ∧ nombre_2_first_3 ∧ sort_nm_n) ∨
(apellido_2 ∧ cens_apellido_1 ∧ cens_sort_nm_n ∧ nombre_2 ∧ sort_nm_1) ∨
(cens_sort_nm_n ∧ desaparicion_yrmo ∧ sort_nm_1) ∨
(apellido_1 ∧ cens_apellido_1 ∧ desaparicion_ymd) ∨
(apellido_1_last_3 ∧ desaparicion_ymd ∧ nombre_2) ∨
(cens_apellido_1 ∧ homicidio_yrmo) ∨
(apellido_1_first_3 ∧ sort_nm_1 ∧ synth_hecho_ymd ∧ synth_muni_code) ∨
(apellido_2 ∧ cens_nombre_1 ∧ desplazamiento_yr ∧ nombre_2 ∧ sort_nm_1) ∨
(apellido_2_meta1 ∧ homicidio_ymd) ∨
(muni_hecho_desaparicion ∧ sort_nm_1_first_3 ∧ sort_nm_n) ∨
(apellido_1_first_3 ∧ apellido_2_last_3 ∧ cens_nombre_1 ∧ homicidio_yrmo) ∨
(apellido_1 ∧ cens_dept_hecho_homicidio ∧ cens_nombre_1 ∧ muni_hecho_homicidio) ∨
(cens_sort_nm_n ∧ desaparicion_ymd) ∨
(apellido_1_last_3 ∧ desaparicion_ymd ∧ synth_muni_code) ∨
(apellido_1_last_3 ∧ apellido_2_meta2 ∧ desaparicion_yrmo ∧ sexo) ∨
(synth_id) ∨
(apellido_1 ∧ apellido_2 ∧ nombre_1) ∨
(apellido_1 ∧ dob_ymd ∧ nombre_1) ∨
(nac_sorted)
```

To understand the rules, it is worth taking a few examples. The last rule (`nac_sorted`) considers all pairs of records that share first and middle names, and first and second last names, with *tokens* sorted. For example, if there was a victim Juan Gabriel Pérez Ramírez, we order him as Pérez Ramírez Juan Gabriel. While Ana Rodríguez would be Rodríguez NA Ana NA. Now, the penultimate rule considers those with the same first and last name, as well as with the same date of birth. For its part, the `synth_id` rule considers all pairs of records that share the identity card number or the number of some other identification document.[9]

Although there are many rules in this list, it is necessary to remember that a record pair has to satisfy only one of these rules to be included among the candidate pairs. Each time a new positive pair is found, the rule list is generated anew. The search algorithm covers 99,7% of the labeled positive pairs. It is worth emphasizing that from this we have blocks of potential records that may belong to the same person, but whether or not these records actually refer to the same person has not yet been addressed. So far there are then a total of 105,734,852 pairs of records to be analyzed.

[9] Many *cédula* numbers were reported with errors. For example, because the numbers in a pair of records match exactly except for one character. Similarly, *cédula* numbers were frequently reported in the "other document" field and vice versa. There are also many registries that lack information about *cédula* number.

3.3 Feature generation

Once we have smaller blocks, it is necessary to analyze the possible pairs of records within each block. Again, it would be easy for a person to identify whether or not two records are coreferent. For example, because the names, violent acts, and date match. It may also be because the dates of birth are very close or because the names sound similar, even though they are spelled differently.[10] [11]

116 comparison measures (for example, comparing the first 3 letters of the name) were generated from a total of 45 fields (or variables). None of these measures are sufficient on its own, but when combined, they provide a sufficient starting point for estimating the probability that two records refer to the same person.

3.4 Pairwise model

We create a model that predicts a weight for a pair of records to refer to the same person. The features generated in the previous task are the independent variables of the model, and the dependent variable is taken from the training data labeled by the oracle.

3.4.1 Model evaluation

In this case, we address the question: do record X and record Y refer to the same person? To answer it we can use a binary classification model, such as a *random forest* or logistic regression (James et al. 2013). The `XGBoost` (Chen and Guestrin 2016) algorithm has been found to consistently generate the best results for this application. `XGBoost` is a regression classifier based on boosted trees, and since the variable to be predicted is binary, we decided to use a logistic function as our objective function.

The model is trained using the labeled training data. In other words, the algorithm learns the relative weights (or importance) of the features that best predict the positive and negative labels in the subset of data selected to train. These weights are then used by the model to predict the likelihood that each pair refers to the same person. The relative importance for the ten variables identified by the algorithm as most informative are shown in Figure 6. Keep in mind that the graph is illustrative to understand the relative importance of the variables, but the scale is not proportional to it. That is, `lv_full_lastname_name` is not 7 times more important than `nac_sorted_jaccard3`, but it is more important.

[10] See this wikipedia explanation about metaphones.

[11] See the wikipedia explanations of the Levenshtein distance and the Jaro-Winkler distance.

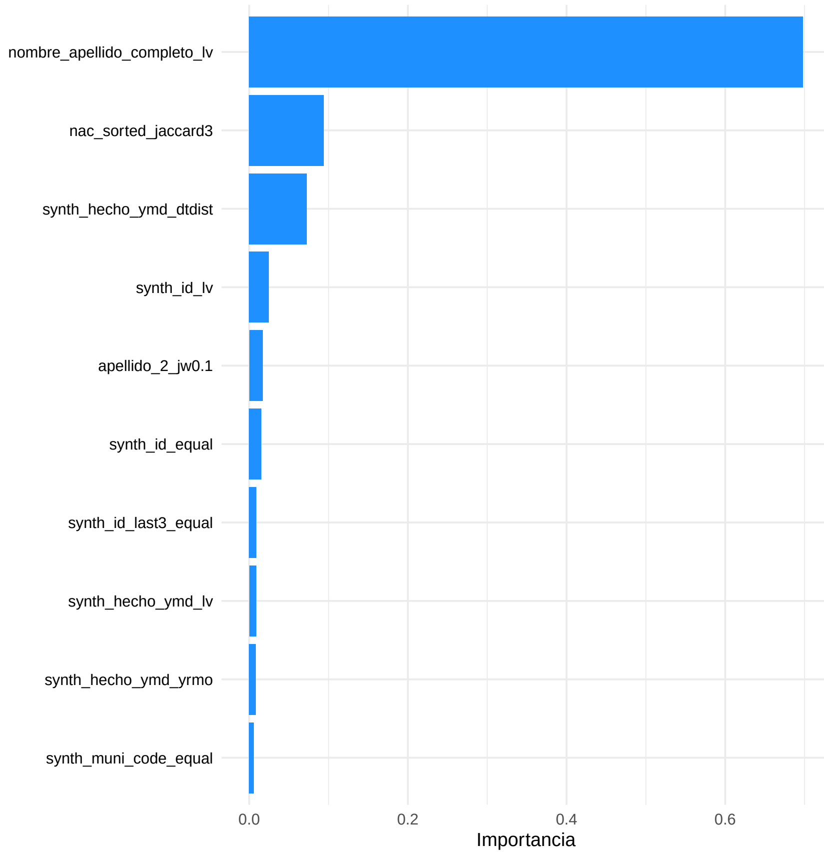

Figure 6: Top 10 of the most important variables in the model fitting.

In this model, most of the explanatory power of the model comes from the variable `nombre_apellido_completo_lv` (`name_lastName_full_lv`). This variable measures the difference between the text strings of the full names of the first record and the second record in the pair under consideration. The difference in this measure is the numbr of characters needed to transform one text string to the other. It is followed by the Jaccard index, which measures the similarity between the complete name and surname in both records. Third, the coincidence of the dates. The similarity of the *cédula*, last name, and municipality are also important for the classifier.

How good is the classification? The model is tested using the labeled test data set. These data were not used in the training process, and therefore allow us to measure the performance of the model on data that the model has never seen. Across different machine learning metrics, this model performs very well. The F_1 score, recall, precision, accuracy, and Matthews correlation coefficient are greater than 0.98.

These measurements are calculated to show a complete assessment of the model, but we prefer two graphical methods to assess the model: separation and calibration, shown in Figure 7.

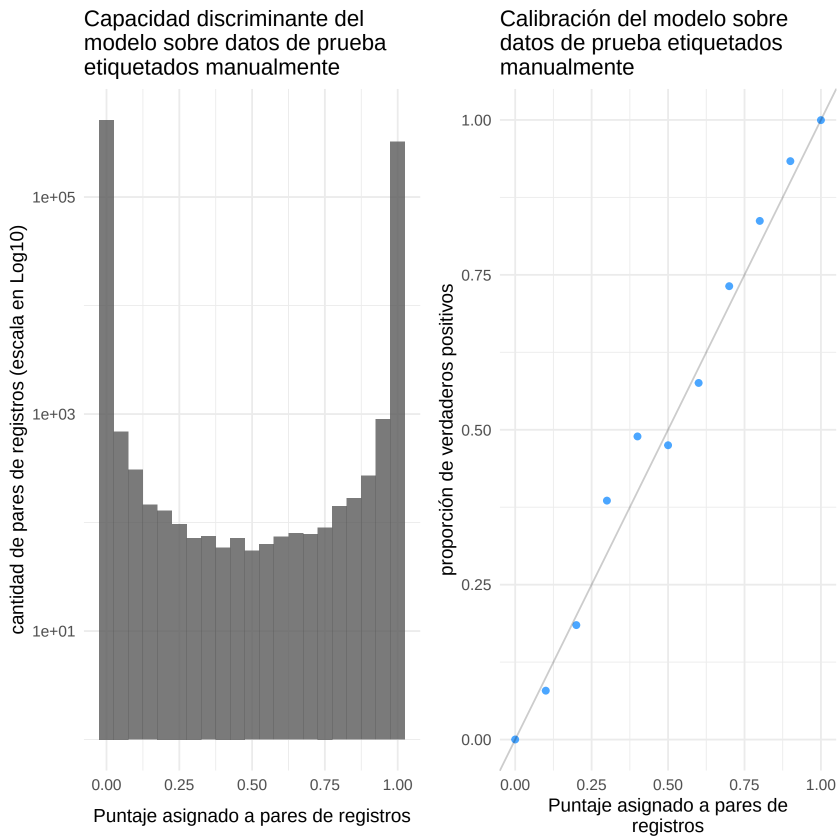

Figure 7: Perfomance evaluation of the model through a separation analysis (0 for no match and 1 for match) and calibration.

The left panel of Figure 7 shows the number of pairs in the training set for each of the scores predicted by the model. It is worth noting that the y-axis is on a logarithmic scale. The two highest bars are at 0 and 1, which indicates that the model is able to discriminate most pairs well, since it assigns 0 to pairs of records that do not refer to the same person or 1 if they do. Therefore, the model is reliable. On the other hand, for most pairs with a probability between 0 and 1 there are around only 100 pairs. So the model is well separated.

The right panel of Figure 7, the calibration plot, shows the extent to which the model-predicted score are an accurate representation of model certainty or uncertainty. The figure shows the proportion of pairs with each value that the oracle labeled as positive. A value of 0.5 indicates that the model has no way of knowing if a pair with this value is positive or negative. In general, it is expected that in a well-calibrated model, the score assigned to pairs of records and the proportion of true positives are equal, so the line $y = x$ is presented in the graph. There it is observed that the points are very close to the line indicating a good calibration.

The combination of good separation and good calibration means that the model is reproducing the decisions of the oracle. That is, the model learned to mimic human decisions about whether or not a pair of records refer to the same person. Once the model is trained, it is applied to the entire data set of 105,734,852 candidate record pairs, and the scores are saved for the clustering task.

3.5 Clustering

The training data for clustering comes from records reviewed by the oracle. Similar to the process for the blocking model, in this step the oracle reviewed groups of records,[12] such as those shown in the dendrogram. The groups are connected components, and therefore the oracle separates them into matching groups that refer to the same person. This way of generating training data allows to generate many positive pairs (every pair in a match group is a positive pair) and many negative pairs (every pair of records in a connected component that is not in the same match group is a negative pair). This is an efficient use of the limited amount of labeled training data.

At this time we have identified whether or not each pair of records within each block corresponds to the same person. However, the same person could be in two pairs of records or more. Then, it is necessary to group the pairs of records that correspond to the same person.

Grouping consists of two steps. First, the records that have some connection are grouped, considering a very low probability of matching (we have used $cc_t = 0.2$). If a record A has a link to a record B ≥ 0.2, and B has a link to a record C, then A, B, and C form a "connected component" (see Hagberg, Schult, and Swart 2008), using the language of graph theory. This step separates our records into relatively small partitions of possible groupings.

Record linkage is achieved thanks to a series of criteria in the algorithm. For example, comparing the first three letters of the records. However, the model is not always certain whether or not a pair of records belong to the same person and, in many cases, assigns them with a probability close to 0.5. According to percolation theory (see Berchenko et al. (2009) and Ming Li and Zhang (2021)), the possible pairs eventually merge and connect into one huge structure, known as a spanning cluster. So at some point by having so many pairs that might refer to the same person, the number of blocks grows exponentially. However, given the computational limitations, only a limited number of blocks can be considered.

Each cluster is then separated using hierarchical agglomerative clustering, using the `fastcluster` algorithm and software (Müllner 2013). This method organizes the records within a cluster into smaller subclusters that are close to each other.

Then it is possible to partition the subgroups into groups of matching records that refer to a person. We have used an "average" link with a "flat" link threshold of $f_t = 0.65$.[13] The f_t hyperparameter is a distance, not a match probability, so this is the complement $(1 - p)$ of the scores returned by the model. Thus, if the model indicated a score of 0.8 between two pairs, which when approaching 1 indicates that it is highly likely that this pair refers to the same person, the distance between them for clustering will be 0.2.

The clustering method is best understood with a visualization. Therefore, one of these groupings is presented below in a graph called "dendrogram". Figure 8 shows the distance between pairs of records and how the subclusters are similar to each other. So, the horizontal axis shows the distance between subclusters, while the lines show the distance required to join the subclusters.

In this case, on the left is the ID or *cédula*. The colored boxes show whether or not the cédula. is the same for two or more records. In this case, only the third to last and the second to last records have the same *cédula*. It is followed by name 1, covered by the black boxes, which show that in this case name 1 matches for all records. The dendrogram shows us that records 5 and 6 are very similar, they are connected by the red line, as well as records 1, 2, and 3. Since we allow a f_t of 0.65, in this case these records would be the only ones linked together.

[12] For a discussion of generating cluster-based training data, and a comparison with pairwise training data, see this presentation (see near the minute 17:30) given in 2016 at a conference on record linkiage at the Isaac Newton Institute for Mathematical Sciences at the University of Cambridge.

[13] We use the `scipy fcluster` function to partition the clusters

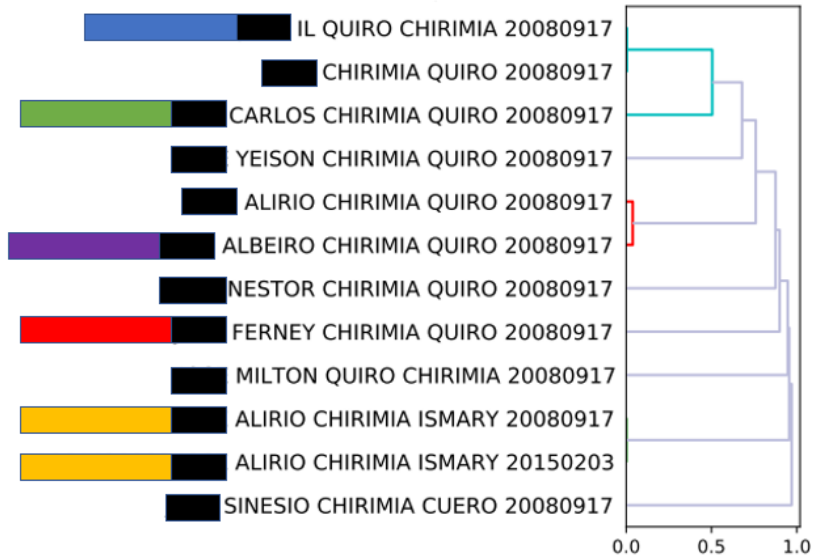

Figure 8: Dendrogram

3.6 Merge and export

The grouping method groups all records that refer to a person into match groups. It is possible that the same victim has different information depending on the source. For this reason, it is necessary to establish a series of reclassification decisions.

3.6.1 Reclassification

3.6.1.1 Types of violence Once the work of identifying the pairs of records that correspond to the same person has been done, it is necessary to establish what to do when the same victim has two or more registered types of violence. Based on the expertise of researchers on the conflict, a series of reclassification decisions were made.

According to the knowledge of the researchers, the following decisions were made. First, if once we deduplicate the same victim has the same type of violence, that type is used. Second, if in the deduplication the same victim is registered as a victim of homicide and another type of violence, both events are maintained. In other words, in the cases of homicide and kidnapping, homicide and enforced disappearance, homicide and recruitment, and homicide and displacement, both types of violence remain.

The main ambiguity is found in the combinations of kidnapping, enforced disappearance, and recruitment, which are observed in Table 16. There we see the proportions, taking the first type of violence as the denominator. For example, we see that the 0.08 of enforced disappearances is also kidnapping. While the 0.19 of kidnappings are also enforced disappearances. There are many more disappearances than kidnappings, so the proportions are different. We count a victim only once in each of these three events.

The reclassification rules were made in this case according to the amount of information required to report each type of violence. Specifically, to make a recruitment report, it is feasible to assume that there is information that the minor was taken by an illegal armed group. While to make the kidnapping report, it is feasible to think that compensation was requested in exchange. On the other hand, for enforced disappearance there is no information of any kind. Based on these criteria, it is possible to assume that if a record is documented as recruitment or kidnapping, it is because there is more information than in the case of enforced disappearance. So, in cases of ambiguity, the reclassification of recruitment of children and adolescents was made, followed by kidnapping, and finally enforced disappearance.

Table 16: Proportion of two types of violence documented for the same victim

Types of violence	Proportion
Disappearance with kidnapping	0.08
Disappearance with recruitment	0.01
Kidnapping with disappearance	0.19
Kidnapping with recruitment	0.01
Recruitment with disappearance	0.11
Recruitment with kidnapping	0.04

Finally, it is important to emphasize that there are also victims who have homicide and other types of violence. In these cases, we maintain all instances of violence. The ratio of disappearance to homicide is 0.16, that of kidnapping to homicide is 0.08, and that of recruitment to homicide is 0.08.

3.6.1.2 Ethnicity It was decided to prioritize groups historically discriminated against. Specifically, if a record has an associated ethnic group historically discriminated against and, in addition, "mestizo", the historically discriminated group is maintained. Based on these rules, we reclassified 490 records from mestizo to indigenous, 357 records from mestizo to Afro-Colombian[14] and 4 records from mestizo to Rom.

3.6.1.3 Other variables If there are many records in the same group and the values for a variable other than the type of violence or ethnicity are different, we take the mode. That is, it is the most common value. In case there is a tie, we pick randomly.

3.7 Results

Table 17 shows the confusion matrix of the training data. A "true positive" is a pair of records that both the oracle and the model identified as a match. Whereas a "true negative" is a pair that both the oracle and the model identified as a non-match. A "false positive" is a pair that the oracle identified as not being coreferent, but the model grouped together. Finally, a "false negative" is a pair that the oracle identified as coreferent, but the model did not.

The clustering method hyperparameters (cc_t, f_t) can be adjusted to change the balance between false positives and false negatives. We choose to have a balance number of false positives and false negatives. However, since our priority is identifying the different victims of the conflict, we are willing to accept a slightly higher rate of false positives, instead of failing to identify different people. This is shown in Table 17, where we observe a little more than 200 more false positives than false negatives. However, we see that false negative or false positive records are few compared to true positives and true negatives. These results indicate that good classification performance.

Table 17: Final confusion matrix for hand-labeled data

Measure	Count
False Negative	714
False Positive	962
True Negative	1,716,239
True Positive	1,081,765

[14]Throughout the text we will refer to black, Afro-Colombian, Raizal, and/or Palenquera people as Afro-Colombian.

3.8 Audit

The record linkage model works under one premise: all coreferent pairs, that is, those that refer to the same person, were covered by a block. However, it is possible that this assumption is not met. We perform four verification tests regarding: possible human errors, the data preparation scheme, overlap expectations, and the correct identification of coreferent pairs. We detail each test in the following sections.

3.8.1 Human errors

The record linkage task takes place after cleaning the data from each of the databases. However, there could be human errors in the database processing scripts. One possible source of error could be related to selecting the wrong variables. On multiple occasions, databases include more than one variable with information related to first or last name, which are the fundamental fields for record linkage. For example, this may be because the databases include information from other people, such as witnesses. In this case, it would be possible that the wrong variable was taken. Therefore, we identified all columns with first name/last name information and created a file with the information for each database and the correct variables. Using this file, we audited all scripts and verified that that correct variables were being selected.

3.8.2 Data preparation scheme

Variable names do not always correspond to the data contents that a user may assume. For example, a database may include the variable “name” but it contains a code instead of a name. For this, we generate histograms with different measures to evaluate the content of the variables. Specifically, we analyze the number of characters and the distribution of letters in first name 1, first name 2, last name 1, last name 2, and full name fields of the records.

In the following figures, we show histograms for the length of the first name 1 and last name 1 fields for 20 data sources, which cover more than 90% of unique victims. We observe consistency across all sources, suggesting that none of the name variables we selected are behaving strangely and do indeed contain first or last name information.

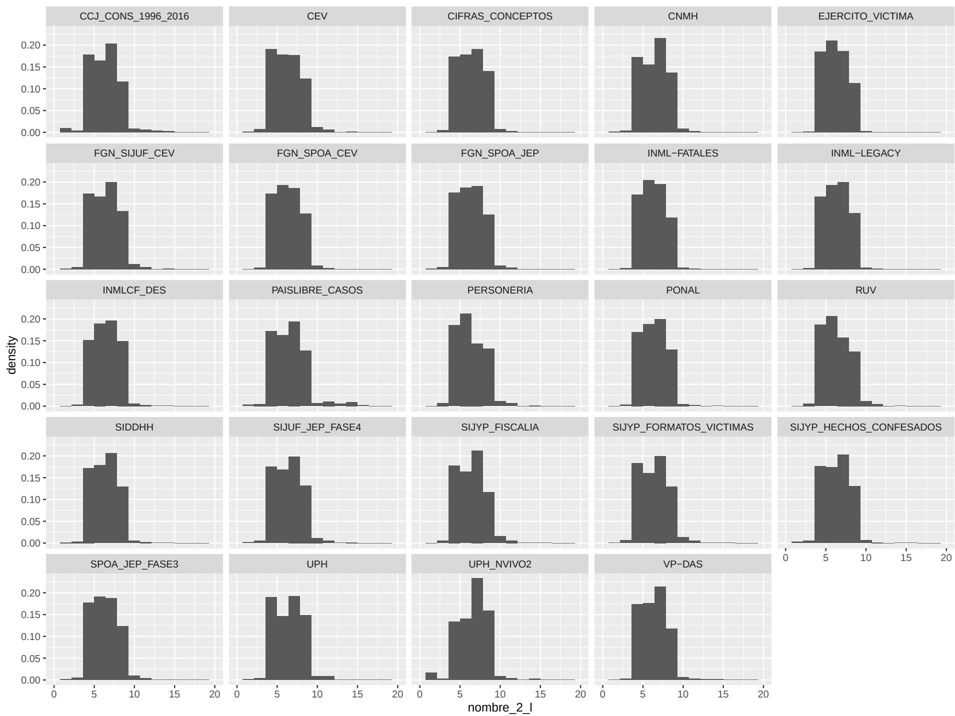

Figure 9: Distribution of the quantity of characters in name 1

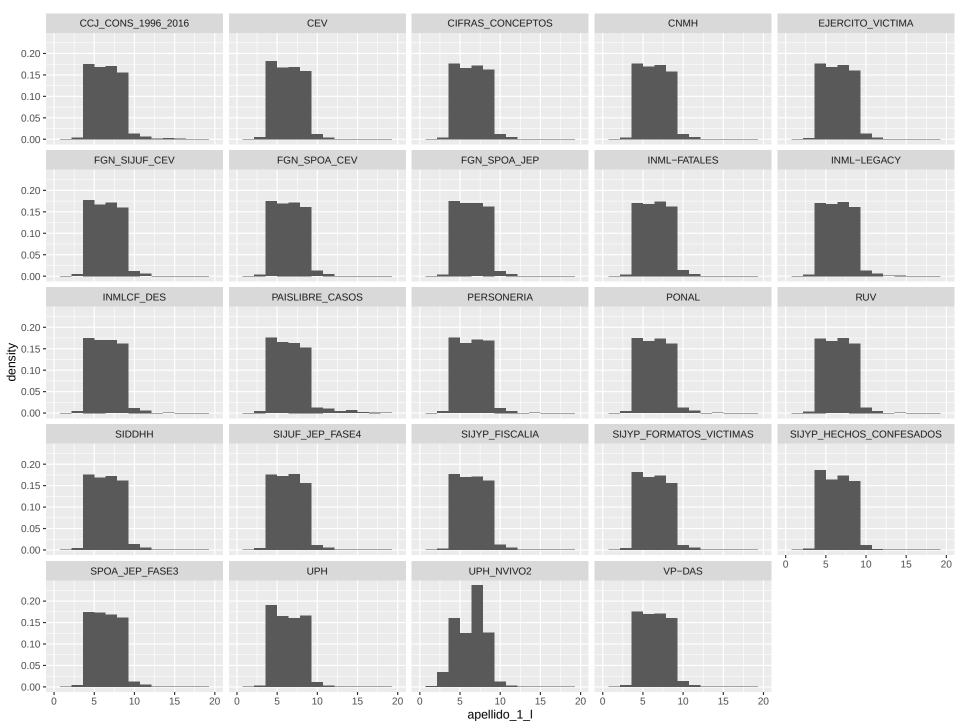

Figure 10: Distribution of the quantity of characters in last name 1

Similarly, the following histograms show the distribution of letters in the first name 1 and last name 1, again for 20 data sources covering over 90% of unique victims. Again, we did not find atypical behaviors.

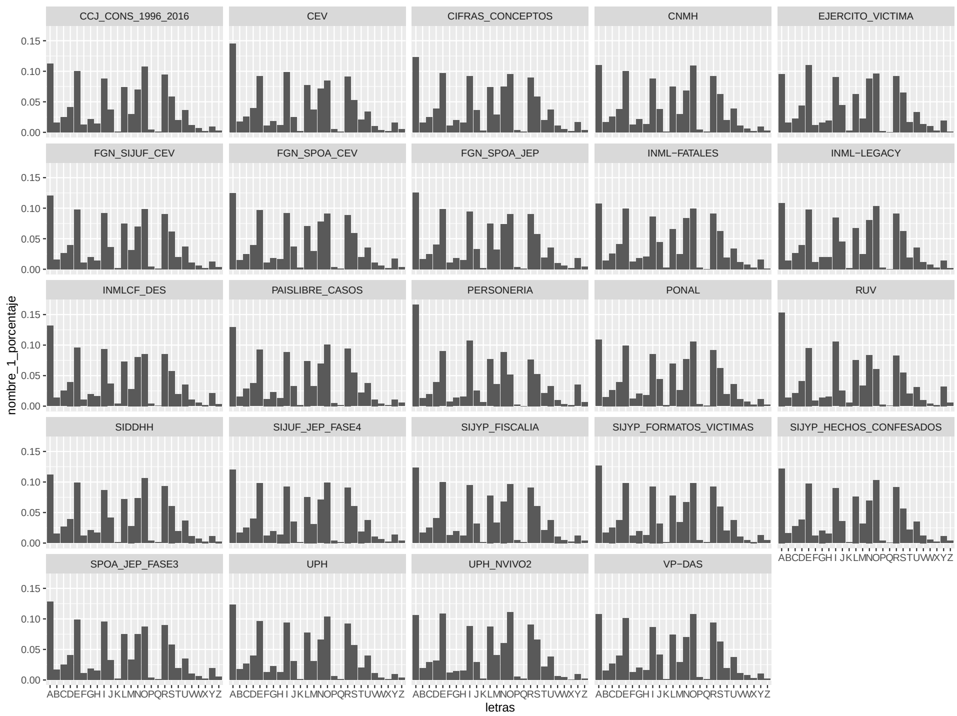

Figure 11: Distribution of the quantity of characters in name 1 by letter

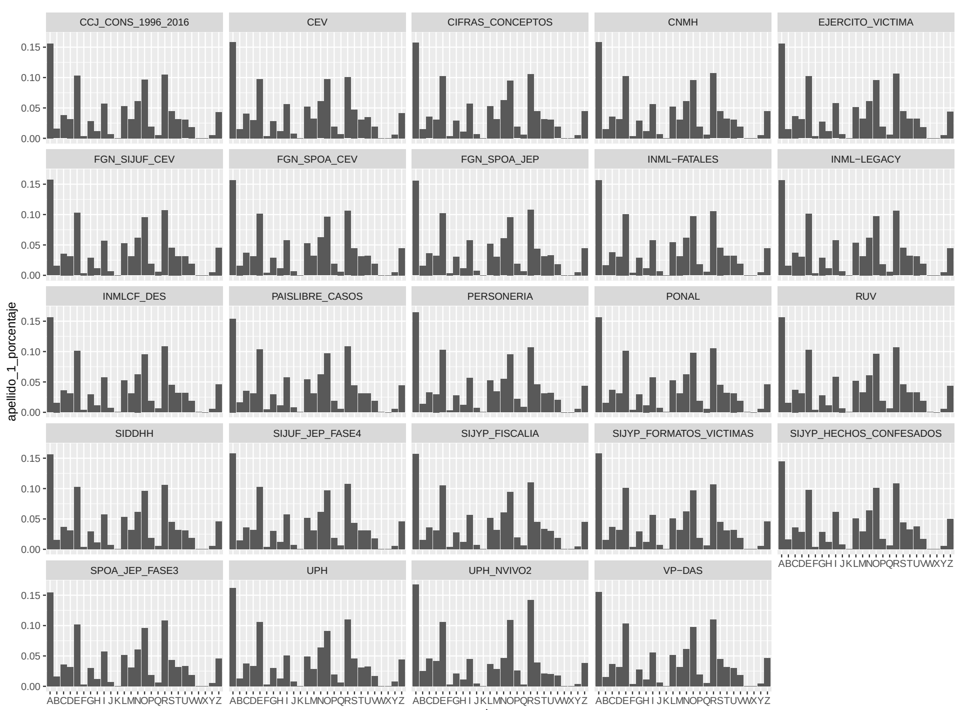

Figure 12: Distribution of the quantity of characters in last name 1 by letter

3.8.3 Overlap expectations

Some of the databases are built with other databases as input. So we expect some minimal overlaps that we know about ex-ante. Specifically, we have that:

$overlap(A, B) = \frac{A \cap B}{|A|}$

This overlap is not symmetrical, since it is possible that organization A has taken organization B as a source of information, but the opposite is not necessarily true. Thanks to the documentation or knowledge of the databases, we can find out about some cases in which the projects took other organizations' data as an input. From this, we define an expected overlap and test whether it is fulfilled.

The results are shown in Table 18, where we show that the minimum overlap expectations are met.

Table 18: Overlap expectations

Violence	Base 1	Base 2	Min overlap expected	Overlap
Enforces dis	INML	CNMH	0.3	0.64
Displacement	FGN	RUV	0.5	0.75
Homicide	INML	FGN	0.5	0.92
Recruitment	ARN	ICBF	0.5	0.66
Kidnapping	País Libre	CNMH	0.3	0.64

3.8.4 Identification of coreferent pairs

If the coreferent records looked different enough they would not be covered by blocks. If this occurred, the records may not be linked to the same victim. One possible cause of differences among truly coreferent records could be information corruption during data processing For example, if we had a coding error that swapped the first name with the last name of some records. This issue would affect not only the block, but also the sampling process we use to generate training data. Therefore, it would be difficult to assess whether such errors are occurring in block generation.

Thanks to the fact that we store intermediate data at each stage of data processing, we can retrieve original versions of the input data. So, we sent a sample of potential links to the oracle, and she identified 3,665 coreferent pairs, of which 83 were undiscovered by the blocks.

We used these results to generate additional training data for the blocking task. First, we generated pairs based on blocks of non-standardized fields, and then filtered out previously discovered pairs that our classification model labeled as likely matches. An innovation here was our ability to consider multiple representations of the same record during pair generation. Some registries have multiple versions of the victim's name, sometimes spelled differently, and during data standardization, we select one of them as canonical.

Finally, we ran the block generation rule lookup again with the enriched training data. The resulting rules continued to work well on our existing training data, and also covered all 83 previously undiscovered pairs that the oracle identified. In summary, given all the tests and model results we have high confidence in the ability of our model to identify unique victims.

3.8.5 Final results of record linkage

We started the match with 12,863,977 records and ended with 8,775,884 individual victims. Some victims in the deduplicated database suffered more than one type of violence (see Section 3.6.1.1). Other victims in this database are filtered (see Section 8.1.1) and therefore are not considered for any acts of violence. Calculations for the number of victims in the deduplicated database for each type of violence are presented in Section 2.

4 Statistical imputation of missing data

The challenge of statistical imputation of missing fields is knowing the probability that each missing field should be assigned to a certain category or value (e.g., age). The fields that were left for each of the integrated databases of the 4 types of violence[15] are: sex, ethnicity, age, perpetrator, and municipality. However, since we also have specialized and non-specialized databases, it was necessary to impute two additional variables: "is conflict-related" and "is enforced disappearance." We present these statistical imputations in detail below.

4.1 Missing fields

In Table 19 the proportion of records missing information for each variable. We see that "perpetrator" is the most critical variable, as it has up to 70% of records missing information. Ethnicity also has notable levels of missingness, ranging from 39% of records of recruitment and 59% of records of disappearances. We see that the variables of sex and age tend to be more complete. Note, that these statistic refer to all records, including records of violence that occurred outside the context of the armed conflict.

Table 19: Missing fields according to the kind of violence.

Kind of violence	Age	Ethnicity	Perpetrator	Sex
Disappearance	0.06	0.59	0.7	0
Homicide	0.1	0.43	0.66	0.04
Recruitment	0.1	0.39	0.23	0.19
Kidnapping	0.23	0.44	0.33	0.01

[a] This table includes the registries for all the years.
[a] In all variables there are missing fields. Given the two decimals, they can be rounded to zero.

4.2 Is conflict

In the project we received both specialized and non-specialized databases. The specialized databases are all those that are limited to documenting what happened in the context of the armed conflict. Among them is the Observatory of Memory and Conflict (*Observatorio de Memoria y Conflicto*) of the National Center for Historical Memory (*Centro Nacional de Memoria Histórica*), the Single Registry of Victims (*Registro Único de Víctimas*) of the Unit for Comprehensive Care and Reparation for Victims (*Unidad para la Atención y Reparación Integral a las Víctimas*), and the database of interviews conducted by the Truth Commission, among others. Then, we build an binary variable that classifies the records as being related to the armed conflict or not. We refer to this variable as the "is conflict" variable. It takes a value of 1 in the case of the specialized databases because we know with certainty that these databases only document violence that occurred within the context of the armed conflict.

There are also non-specialized databases. Among these are those of the National Institute of Legal Medicine and Forensic Sciences (*Instituto Nacional de Medicina Legal y Ciencias Forenses - INML*) and those of the Attorney General's Office (*Fiscalía General de la Nación - FGN*), among others. In this case, we know that some records correspond to the armed conflict, for example, because the perpetrator was an armed group, and the "is conflict" variable would take a value of 1. Additionally, we also know that some records do not correspond to the armed conflict, for example, because the violence was documented as being common crime. In these instances, the "is conflict" variable would take a value of 0. However, there are thousands of records for which it is not clear whether the violence took place in the context of the armed conflict.

It makes sense to assume that some of the searches by Police, FGN, and INML occurred within the context of the armed conflict, but how many? We can use machine learning to help answer this question by identifying

[15] Homicide, disappearance, kidnapping and recruitment. Exile is not included due to the lack of enough data. Displacement is also not included given the high influence of the RUV, as shown in Section 5.5.

characteristics in common between records that relate to violence that occurred in the context of the armed conflict and common characteristics of violence that occurred outside the context of the conflict. We can then use this information to estimate whether a record missing this information relates to the conflict. In the next subsection we present the missing data for the "is conflict" variable and our decisions to generate training data to use both specialized and non-specialized databases.

4.2.1 Missing data for the "is conflict" variable

Table 20 shows the proportion of records missing information about whether they occurred in the context of the armed conflict. We find that enforced disappearance and homicide have the highest proportion of missingness. In the case of disappearance, we do not have information on whether 30% of the records occurred within the context of the conflict. While in the case of homicide this percentage is 20%.

We use machine learning to classify the records for which we are missing this information. To do this, we need training data. That is, example data that meet and do not meet a condition. In this case, the condition is whether a record relates to the armed conflict. We used the ACOMIDES, CCJ, FGN, JPM, País Libre, and INML databases along with a set of expert criteria to define whether a record occurred within the context of the conflict. To create the training data, we assigned records a value of 0 if they were not related to the armed conflict and 1 if they were. This training data was derived from human expertise. In this case, we worked with CEV experts on the armed conflict who developed a series of rules to label the training data.

The experts developed two types of rules. First, there are rules for which there is no doubt that the event occurred or did not occur within the context of the conflict. In these cases, a value of 1 or 0 is directly assigned. This type of rule is used to assigned values of the "is conflict" variable when the perpetrator is an armed group or in instances of common crime. Although these rules cover most records, they are not enough. We need more training data, especially for cases that occurred outside of the context of the armed conflict; we need more records with known values to train the model.

As a result, the experts identified a series of probabilistic rules for consummated homicides documented by the Prosecutor's Office's Oral Accusatory Penal System (*Sistema Penal Oral Acusatorio - SPOA*). These rules allow us to include uncertainty in our analysis and acknowledge that it exists. We list them below:[16]

- It is unlikely that it occurred in the context of the armed conflict. For example, in the case in which the crime is "abuse of authority due to failure to report Article 417 of the Penal Code". A probability of 0.75 is assigned that the "is conflict" variable is false, that is, `is_conflict == 0`.
- It is not clear whether it occurred within the context of the armed conflict. For example, "crimes against life and personal integrity." A probability of 0.5 is assigned that the "is conflict" variable is true, that is, `is_conflict == 1`.
- It is very likely that it occurred in the context of the armed conflict. For example, "acts of barbarism Art. 145 of the Penal Code". A probability of 0.75 is assigned that the "is conflict" variable is true, that is, `is_conflict == 1`.

We make the probabilistic assignment only once and thanks to it, it is possible to convert some values from `NA` to 0 and others to 1. If the record does not comply with any of the rules, it remains as `NA`. It is worth remembering that it is valuable to preserve as many records as possible for those for which there is no clarity as to whether or not they are linked to the armed conflict so that statistical imputation can explore the impact of this uncertainty on the estimates.

[16] They are available in detail in the annex at 9.4.1.

Table 20: NA proportion for is conflict.

Kind of violence	NA proportion
Disappearance	0.3
Homicide	0.2
Recruitment	0
Kidnapping	0.08

[a] By definition child recruitment occurred in the context of armed conflict.

4.3 Is enforced disappearance

Similar to needing to know whether an event occurred in the context of the armed conflict, we also need know whether disappearances are enforced or not.[17] Again, we need training data, and again, we turned to the expertise of CEV researchers. They identified cases that could immediately be classified as enforced disappearances or as missing persons cases. For example, because the alleged perpetrator is an armed actor (the "is enforced disappearance" variable takes a value of 1) or because it was connected to common crime (the "is enforced disappearance" variable takes a value of 0). However, we needed more training data to be able to classify records that were not initially assigned a value of 1 or 0.

Again, we addressed this using a series of probabilistic rules. This exercise was conducted using the National Institute of Legal Medicine and Forensic Sciences databases based on a series of milestones in specific departments and years, as well as the victim's group membership. According to these rules, probabilities of 0, 0.5, 0.6, 0.7, 0.8, 0.9, and 1 that the disappearances were enforced were assigned.[18]

As was the case with the "is conflict" variable, we seek to recognize that uncertainty exists. The probabilistic assignment was made only once, converting some NA values to 0 and others to 1. If a record is in a department-year defined by the experts as "1 - 0.7", we randomly draw a value between 0–1 and if the random value is less than 0.7, we assign the "is enforced disappearance" variable to be 1. If we randomly draw a value greater than 0.7 the value of the "is enforced disappearance" variable remains NA. In contrast, consider a record in a department-year defined by researchers as "0 - 0.5". In this case, if the randomly selected value is less than 0.5, we assign the "is enforced disappearance" variable to be 0. Otherwise, the variable remains NA. It is important to note that in each rule we try to leave as many records as NA as possible so that the model can explore the variation created through imputation. Our goal is to generate enough training data labeled with 1s or 0s to train the support variable models and the statistical imputation.

We are interested in enforced disappearances that occurred in the context of the armed conflict. If we take these two conditions into account, the records to be imputed are those with missing values in the "is conflict" and/or "is enforced disappearance" variables. In this case, the proportion of data to be statistically imputed is 0.31.

4.4 Support variables for statistical imputation of missing data

Imputing missing data based only on the variables in the standardized databases would be very difficult. We would need to complete the missing fields based only on information about ethnicity, age, sex, department, municipality, year is conflict, and is enforced disappearance.

The original data sets, however, contain more information. Some databases document, for example, the profession of the victim. Others, the weapon used in a killing or the exact location. Many also have text fields that describe the violent events. This information could help the imputation model. If the model could

[17] In Spanish there is no distinction between a "missing" or a "disappeared" person. The categories are "disappearance" (*desaparición*) or "enforced disappearance" (*desaparición forzada*).

[18] Details are available in Appendix 9.4.2.

take all of the record information into account, it would surely make better imputation decisions. We create support variables for this purpose.

Support variables capture the heterogeneous and non-standardized information from the original databases from which the records come. The narrative and descriptive fields are especially rich in information. To create the support variables, all record fields are transformed to a sequence of words.

In general, the variables that are useful for imputation are those that meet two characteristics: (i) they are strongly correlated, but not collinear, with the variables to be imputed; and (ii) they are not missing for any records in the complete data set (Van Buuren, Boshuizen, and Knook 1999). Our support variables comply with these two characteristics.

Since the original data comes in a variety of formats, it is necessary to generate a standard structure that allows for all information to be used in a single model. To do this, we combine all record information into a single sequence. We cleaned this sequence of text, eliminating numbers, punctuation, and stop words and lemmatized each word.[19] Then, we took a list of the 2000 most common words and converted each word in the sequences into an index position of this list.

By converting the words, it was possible to train a *long short-term memory* recurrent neural network [LSTM; Hochreiter and Schmidhuber (1997)]. We used Tensorflow and Keras to train the model. The LSTM layer of the model extracts 32 neurons, which are functions that incorporate the four LSTM processes at each iteration (or "epoch") of model training (the neuron consists of four processes, which includes a cell, an entry door, an exit door, and a forget door). The neurons are fed by a sigmoid function, which is essentially logistic regression, fitting a binary dependent variable.

To train the neural network for a particular variable, we begin with the records that have no missing fields in that variable. Since the sigmoid model works best with balanced data, we randomly select an equal number of records that meet and do not meet the condition. For example, for the sex variable we take the same number of records that meet the condition of being a man and not being a man. From this selection, we randomly select 75% of the records to train the model and reserve 25% for testing model fit and performance.

This information will aid the statistical imputation model in searching for the records that are most "similar" to those that have missing fields and, in this way, make better decisions about the group to which they belong. In simple words, the support variables identify additional "clues" for statistical imputation, such as the description of the alleged perpetrator, the exact location, etc. to identify the category to which the records would belong. These "clues" are not created by humans, but learned from the neural network. In the next subsection we show the fit of the model of the support variables and then the statistical imputation of the missing fields.

4.4.1 Model fit

Table 21 shows the fit of the support variable models. The AUC is greater than 0.93 for all support variables and the Pearson correlation coefficient is greater than 0.79. These measures indicate that the support variable models are reliable in representing the observed probability that each variable is (or is not) relevant for each record.[20]

[19] See this Github repository for a lemmatization dictionary for Spanish.

[20] The AUC is an indicator of the performance of a classification model, which takes values between 0 (the model has the worst possible separation ability) and 1 (it has the best possible separation ability). The Pearson coefficient takes values between –1 (the variables are perfectly inversely related) and 1 (the variables are perfectly related).

Table 21: Model fit statistics, correlation coefficiente between support variables and original variables, proportion of missing data imputed.

Variable	AUC	Pearson coefficient	Missing fields proportion
ADULTO	0.942	0.799	0.534
CONFLICT	0.999	0.994	0.091
ELN	0.998	0.983	0.819
EST	0.999	0.995	0.819
EST_PAR	0.981	0.895	0.819
FARC	0.999	0.992	0.819
FEM	0.999	0.991	0.424
FORCED_DIS	0.982	0.893	0.978
GUE	0.999	0.996	0.819
GUE_OTRA	0.937	0.848	0.819
INDIGENA	0.999	0.999	0.552
MASC	0.999	0.991	0.424
MENOR	0.942	0.802	0.534
MESTIZO	0.997	0.987	0.552
AFROCOLOMBIANO	0.999	0.994	0.552
NINO	0.994	0.975	0.534
OTRO	0.993	0.941	0.819
PARA	0.999	0.993	0.819
POSDES	0.999	0.996	0.819

In summary, support variables are variables that are not missing for any record and can aid the statistical imputation of missing data, but that do not determine the result of the imputation process.

4.5 Statistical imputation of missing data for sex, ethnicity, age, perpetrator, municipality, is conflict, and is enforced disappearance

The variables for which missing fields must be complete, that is, the variables to be imputed are: sex, ethnicity, age, perpetrator, municipality, is conflict, and is enforced disappearance. One possible solution for addressing the missing fields would be to delete the records that do not have complete information. However, this approach has at least two problems. First, it means assuming that cases without information have the same distribution as those with information. We can think about the consequences using the ethnicity field in the case of homicides. This assumption implies that the 43% of homicide victims who do not have a registered ethnic group have the same distribution as those. This is not necessarily true, and it is possible that *mestizo* people are more documented than others for a variety of reasons. Second, deleting incomplete records would mean losing a large percentage of the available data, especially for homicide and enforced disappearance. Another solution could be to keep the records that are missing information, but exclude those that do not have complete information for a particular analysis. However, this would ignore that records with missing fields might have different information. For example, if the alleged perpetrator of the records of violence missing a particular fields are different from those where the field values are known, the story we would know about the conflict would be different. Rather than deleting incomplete records, it is necessary to statistically impute missing fields.

One intuition regarding the missing values of a particular variable is that they follow a pattern similar to the observed values of the variable conditional on all observed values. This is known as "missing at random" (MAR). This is a fairly plausible assumption, indicating that records that are similar on other variables should have similar values on the missing field.

In order to impute missing fields conditional on all the values in a record, the combination of all other variables, including the support variables, are used. This is done for all missing fields in a record and is called *multiple imputation* (Rubin 1987, 1996).[21] The approach described here is a *fully conditional specification*.[22]

When used correctly, the multiple imputation method produces unbiased estimates for the imputed variables and incorporates the uncertainty introduced by the statistical imputation of missing data into the final estimates of the statistics of interest (van Buuren 2018, chap. 2).[23]

This method starts by filling in the missing values with some value. Then the value for each variable is predicted using all other columns. There are many possible models to do this prediction. In this case, we use a method known as *predictive mean matching*. This method calculates the predicted value of the missing field based a randomly selected record close to it. This is a robust method, which performs better than others such as decision trees.

The statistical imputation of missing data is sequential, since the missing values of all variables are completed. The process of predicting and completing each variable is known as a replicate. Each of these replicates has a random component to help the model compute. This randomness corresponds to the selection of the *donor* record from which the value of the missing field to be filled in is taken. The results from each replicate are slightly different, reflecting the uncertainty in the value, so the process is performed multiple times. This method is known as "multiple imputation" and was proposed by Rubin (1978). Years later, Van Buuren, Boshuizen, and Knook (1999) proposed multiple imputation using chained equations.[24] [25] This is the standard method used to address missing fields and is used in disciplines such as political science and epidemiology, among others.

In Table 22 we show the impact of statistical imputation in the context of homicides and the "is conflict" variable. The "total records" column shows the total number of homicides (without being restricted to conflict-related homicides) documented by the National Institute of Legal Medicine (INML), the National Police (PONAL), the Colombian Commission of Jurists (CCJ), and the General Prosecutor's Office of the Nation (FGN). Since we do not know whether some homicides occurred in the context of the armed conflict, we use the model described in this section to impute the "is conflict" variable. After applying the imputation model, we can estimate the number of homicides that likely occurred in the context of the conflict and measure the associated uncertainty. According to our estimates, out of the 155997 homicides registered by the INML 92757 likely occurred as part of the armed conflict. In other words, about 59%. For PONAL, CCJ, and FGN these percentages would be 55%, 90% and 64%, respectively. It is worth remembering that both in the "total records" and in the "records related to the conflict" columns there are victims that are documented by more than one of the databases. In other words, there are coincidences between the victims registered by INML, PONAL, CCJ, and FGN.[26]

Table 22: Total registries and registries related to the armed conflict

Dataset	Total records	Records related to the armed conflict
INML	155.997	92.757
PONAL	151.407	83.431
CCJ	38.131	34.266
FGN	145.068	92.439

[a] These registries are not limited by time. They might be different from the ones presented in the findings section.

[21] The R package `mice` is used for the calculations (van Buuren and Groothuis-Oudshoorn 2011).

[22] See van Buuren (2018), specifically Section 4.5.1. Also see Kropko et al. (2014), who shows that for categorical variables like the ones used here, conditional specifications (like the ones used here) tend to produce better imputations than normal multivariate models. See also Cro (2017) and Heymans and Eekhout (2019).

[23] This book is available online here.

[24] Specifically, we use 10 replicates in this analysis (White, Royston, and Wood 2011).

[25] Once the replicates are in place, they are combined according to the rules outlined by Rubin (1987).

[26] We do statistical imputation 10 times. Each of these times is known as a "replicate". For example, in the first replicate the INML and PONAL have 39,725 matching records, while the INML and FGN have 27,976 and the FGN and PONAL have 17,809 matches.

5 Documentation patterns

As previously mentioned, it is likely that the same victim is documented in more than one database. This fact is key to carry out multiple systems estimation (MSE). Before estimation, it is necessary to examine the overlaps between the different databases for each type of violence.

We use UpSet plots for this purpose. Since we are interested in studying the documentation patterns of the victims related to the armed conflict, in this section we present UpSet plots created after imputing the "is conflict" variable. These graphs make it possible to identify how many victims are documented by different combinations of databases using the first replicate of the statistical imputation process. Also, it is important to highlight that we present the databases and combinations of databases that document the largest number of victims. However, these are not the only sources or combinations of sources that document records of a particular type of violence.

Additionally, it is worth remembering that there is a difference between the dynamics of *violence* and the patterns of *documentation*. The samples held in the databases are convenience samples, which show what is observable. However, there may be violence that is not documented and that is part of the dynamics of violence (Price and Ball 2015). The following UpSet plots present the *documentation* patterns.

5.1 Homicide

Each row represents an organization and the black circles show when an organization is "active". In this case, the first column shows that RUV is the source that reported the most unique victims of homicide and that these victims were not documented by any other source. In the second column, we observe that RUV, CNMH, and SIJyP documented the second largest number of unique victims, that is victims who were documented by all three sources and no other sources. The third place is occupied by FGN alone (these are records of *homicidios consumados* only).

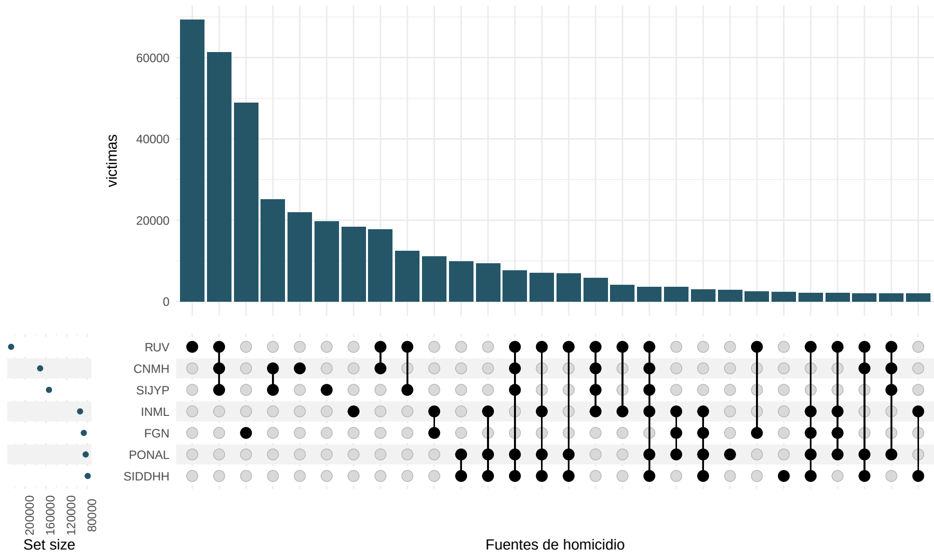

Figure 13: Documentation pattern between sources for homicide

5.2 Disappearance

INML is the source that documented the largest number of unique victims, followed by victims documented by INML, CNMH, RUV, and SIJyP, but not documented by any other sources. Third place is occupied by victims who were documented by both CNMH and RUV, but not any of the other sources.

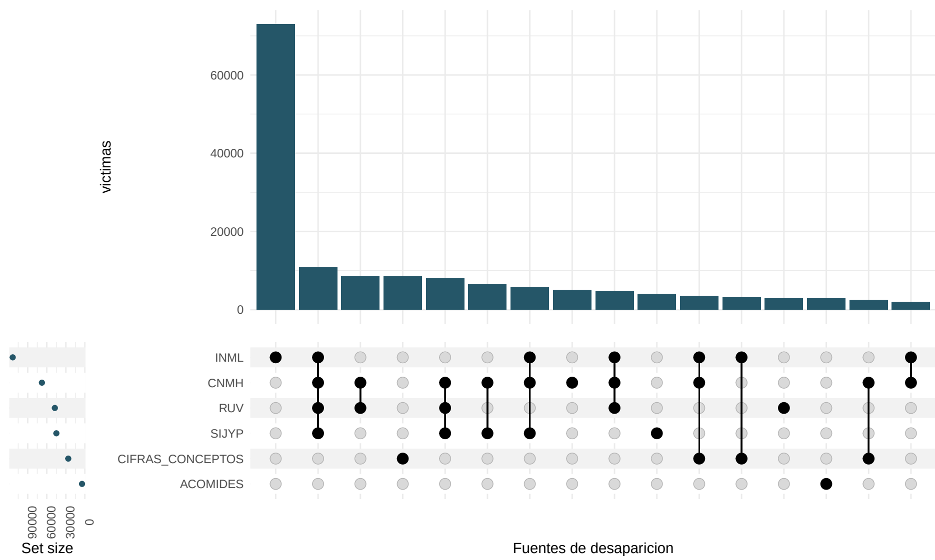

Figure 14: Documentation pattern between sources for enforced disappearance

5.3 Child recruitment

In the case of child recruitment, FGN documented the most unique victims. It is followed by victims documented only by SIJyP and then victims documented by FGN and JEP and no other sources. It it worth reiterating that the FGN data are of victims documented by the institution, which were reviewed by JEP analysts.

5.4 Kidnapping

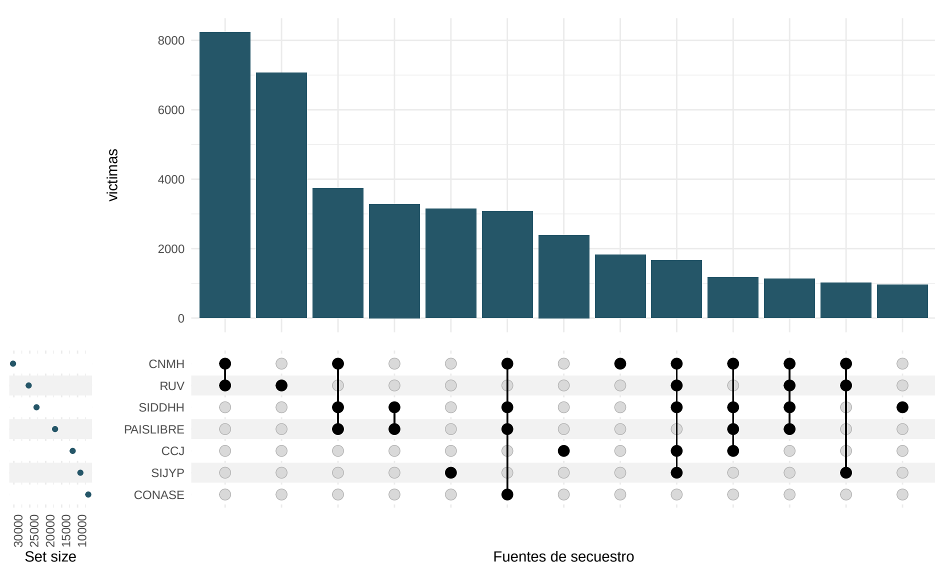

Figure 15: Documentation pattern between sources for kidnapping

In the case of kidnapping, the largest number of victims were documented by CNMH and RUV, but not any other source. This is followed by victims documented only by RUV and then victims documented by CNMH, País Libre, and the database of the Human Rights Information system of the Presidential Counsel for Human Rights (*Consejería Presidencial para los Derechos Humanos*).

5.5 Forced displacement

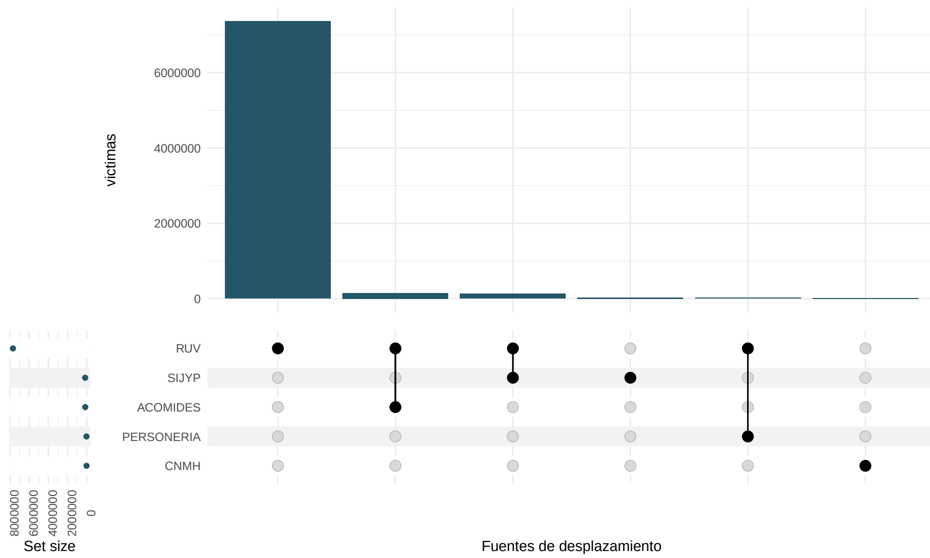

Figure 16: Documentation pattern between sources for displacement

Upon deduplicating the data on forced displacement, the enormous weight of the RUV database. Here, only direct victims of enforced displacement who suffered at least one instance of displacement were considered. In the case of multiple displacements, only information about one instance of displacement is taken, using the mode of all records associated with the same victim. RUV alone documents the most unique victims followed by victims who are documented by RUV and ACOMIDES, but not any other sources and victims documented by RUV and SIJyP, but no other sources.

6 Multiple systems estimation

To estimate the total population of victims of various human rights violations and, therefore, the underreporting in the documentation of violence, we used a technique called *multiple systems estimation* (MSE), also known as capture-recapture. MSE is a class of statistical models that has been used to study human and animal populations since the late 1890s. Chao (2001) provides a review of the technical development of this method, while Bird and King (2018) provide a historical look at its applications to human populations. Lum, Price, and Banks (2013), Ball and Price (2019), Ball and Price (2018) and Rozo Ángel and Ball (2019) show applications to human rights issues, including in the context of armed conflict.

The intuition behind this method is as follows:[27] imagine two dark rooms. We want to know their sizes, but we cannot see inside them. The only tool we have for exploring sizes is a handful of rubber balls. The rubber balls do not make any sound when they collide with the walls, ceiling, or floor, but they do make a small noise—*click*—when they collide with each other. We throw the rubber balls in the first room and hear many *clicks* in a row. We take the balls again and throw them into the second room with the same force.

[27]HRDAG uses this analogy in many reports that use MSE.

Now we hear *clicks* but less frequently. Our intuition is that the second room is larger because the rubber balls spread out more and therefore collided less frequently.

In statistical terms, the size of the "room" is the size of the population of victims of a type of violence that we want to estimate and are "throwing" the sources to document the population of victims. When two or more of the sources document the same victim it is as if they "collided" by making a *click*. We use these documentation patterns to estimate the size of the total population of victims of a specific type of violence, including those that were never documented in our sources (underreporting).

6.1 Technical details

MSE was originally implemented in 1783 by the mathematician Frances Pierre-Simon Laplace with two data sources (Amorós 2014) and four fundamental assumptions (Scheuren 2004). Although modern MSE methods do not necessarily require these assumptions, it is useful to know about them to understand how the methodology works. Here are the assumptions:

- 1. The estimated population is "closed", that is, members of the population are neither created nor deleted during the documentation period.

In this analysis, the objective is to estimate the total population of victims of enforced disappearance, forced displacement, homicide, child recruitment, and kidnapping. For events that can occur more than once to a victim (with the exception of homicide), only one instance of the event is included. So, it is possible to affirm that the populations of victims of the five kinds of violence do not change retroactively. With the exception of rare cases where a victim is documented as having been killed, but is later found alive, there is no way to leave the victim population after entering.

- 2. Record linkage is accurate.

MSE requires a consolidated database that includes each of the victims and indicates the sources that each victim was documented on. This database is the result of the record linkage process. Although it is not possible to ensure that record linkage is perfect as this assumption suggests, the semi-supervised process allows evaluation of record linkage performance. In Section 3, we show that the precision in the calibration of the implemented model is sufficient to have a low rate of false positives and negatives, which allows confidence in the record linking process. Furthermore, it has been shown that even with poor record linkage, MSE results are reliable (J. Johndrow, Lum, and Dunson 2018).

- 3. Being documented in one system or source does not affect the probability of being documented in any other ("list or source independence").

It is helpful to use an example to understand what source independence is. Imagine that we have a population in a municipality where a series of acts of violence occurred in the context of the armed conflict. Each of the victims in this population has many characteristics: their race, sex, ethnicity, among others. One of their characteristics is political ideology: it is likely that some victims from the municipality have an ideology closer to the right and others to the left. However, once the violence occurs, different organizations come to document what happened. Let's start with an organization that has a right-wing political ideology or is more associated with this political leaning: some victims with an ideology closer to the left may not feel comfortable sharing what happened with the organization. Thus, it is probable that the list of this organization reflects more the victimization suffered by people on the right than those on the left, not because of a prejudice of the organization but because of trust processes. In this case, the observation probability correlation between the two sources would be negative.

This assumption can be relaxed by using three or more data sources and using log-linear models to estimate the interaction between subgroups of sources (Bishop, Fienberg, and Holland 1975; Cormack 1989; Chao

1987). For the Bayesian approach see Madigan and York (1997). However, the log-linear approach is limited to a few data sources. It also requires the analyst to select a specific model, or else calculate a weighted average over all possible models, which helps control for Assumption 4. This control is only to the extent that heterogeneity in capture probabilities is correlated with being recorded on a specific data source.

- 4. The probability of being documented by a particular system or source is equal for all members of the population ("capture homogeneity").

The probability of capture (probability of being recorded by a source) is constant across sources. Deviations from this assumption are called "capture heterogeneity" in the statistical literature (Chao 1987). In reality, this is not usually the case with all victims, as some are more likely to be recorded than others. For example, the murder of a human rights lawyer in Bogotá is more likely to be recorded than the murder of a peasant woman in a village. Also, projects may not be equally likely to collect data in different years. For example, a specific source may have had a budget cut in one year and stopped documenting human rights violations. So, the probability of documenting a victim on a specific year from a specific source would be less than in a different year.

This can be controlled in different ways. The classical approach from Sekar and Deming (1949) reduces capture heterogeneity by grouping similar records together. This process is known as "stratification." In the project we always stratify by year and in some cases by year and department or other variables. Stratification creates small lists, e.g., year1-dept1, year1-dept2, . . . , yearX-dept32, which give the model more information.

When the model cannot handle capture heterogeneity, we observer multimodal posterior distributions. That is, the model has difficulty in determining an answer without more information. Many studies of MSE probability theory have found that this may be due to a pattern in the data, leading to "unidentifiability" of a consistent response.[28] For a discussion and a proposal to solve this aspect, which we do not use in the project, see J. E. Johndrow, Lum, and Manrique-Vallier (2019). For a discussion of how to select the strata to deal with multimodal estimation, see Hoover Green and Ball (2019).

With three or more independent data sources there are several ways to reduce the effect of capture heterogeneity. The classical approach points out that capture heterogeneity normally generates dependency between sources. One approach to control for capture heterogeneity during the 1990s was to use Rasch models (Darroch et al. 1993; Fienberg, Johnson, and Junker 1999). In this method, a latent variable captures the individual catchability variation for each record i (expressed as θ_i) and the individual effect of listing j expressed as logit $(p_{ij}) = \beta_j + \theta_i$. Although this approach was a breakthrough, it is still limited because it assumes that effects between sources or lists and individual effects do not interact.

In this analysis we use a method that covers Assumptions (3) and (4) simultaneously, overcoming the limitations of the log-linear approach. We use a type of non-parametric Bayesian MSE called *"Bayesian Non-Parametric Latent-Class Capture-Recapture"* [LCMCR; Manrique-Vallier (2016)].[29] The LCMCR model estimates a posterior probability distribution of the likely values for the total number of victims (those observed plus our estimate of the size of the underreporting) in the stratum of interest.

6.2 Implementation with imputed data

The variance of chained models is propagated. In our case, error from multiple imputation is propagated into the MSE estimates. When the statistical imputation of missing data has high uncertainty, the variance will be larger, and as a result the variance of the estimates will also be larger.[30]

In general, the variance of the imputation is greater than that of MSE. This is because the variance of the missing data imputation is based on the assumption that the replicates are normally distributed (left-hand side), while the estimates are log-normally distributed (right-hand side). This happens because in the estimates we know the minimum (the documented victims).

[28] See Link (2004).
[29] We use the package `LCMCR` developed by Professor Daniel Manrique-Vallier in `R` for all estimates in our analysis.
[30] See Zhou and Reiter (2010).

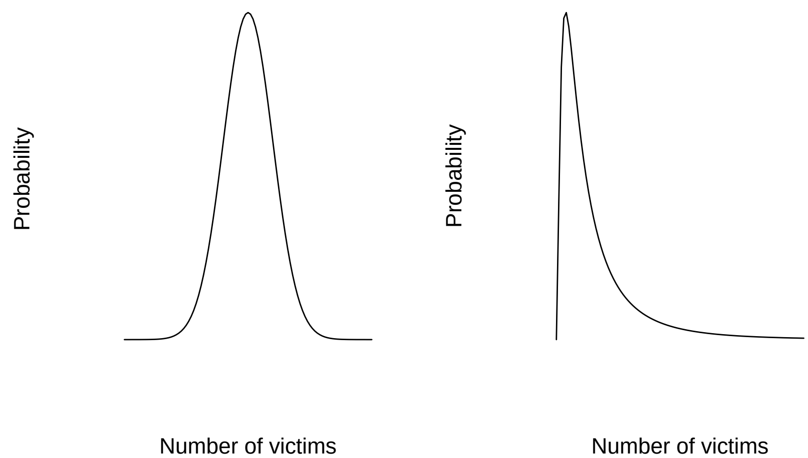

Figure 17: Normal y log-normal distributions.

A table with as many rows as the number of victims and as many columns as the number of sources is used as the input to the LCMCR model. If a victim was documented by a source, it takes the value of 1, otherwise it takes the value of 0. The estimate is then calculated using this table of zeros and ones. However, the data is stratified before calculating the estimates to address potential issues with capture heterogeneity. It is necessary that in each stratum there is more than one victim observed in at least three sources. We only estimate strata for which there were at least 3 lists. Our sampler uses a burn in of 10,000 data points and we draw 1,000 samples from the posterior distribution, one ever 500 iterations.

When a stratum satisfies these conditions, the stratum is estimated using LCMCR. We generate 1,000 samples from the posterior distribution of the number of probable victims for the 10 replicates created using statistical imputation. We then combine the results from the 10 replicates to calculate the point estimate and the associated uncertainty interval.While this requires subsequent estimates to be close to normally distributed, MSE estimates generally have a long right tail.

Following Chao (1987), we transform our data according to $\theta(N) = log(N - n)$, where n is the number of observed victims in the stratum. We do this transformation for two reasons. First, the lower limit of the number of victims in a stratum is delimited by the number of victims observed in that stratum. Uncertainty calculations for estimates should not have lower bounds lower than the number of documented victims in the stratum. Using the quantity `N - n` ensures that the credible interval for the estimate is always greater than the number of documented victims in the stratum.[31] Second, to calculate the final estimates and credibility intervals, we use the combination rules described in Rubin (1987). These rules assume that the distribution of N is normal, but as we explained, the posterior distributions have a long right tail. The distribution of N approximates a log-normal distribution, so $log(N)$ approximates a normal distribution. Combining the two: the distribution of $\theta(N)$ is guarantees that the credible interval is always equal to or greater than the number of documented victims in the stratum and is generally approximately normal so we can apply the rules described in Rubin (1987) without violating the assumptions.

We apply this transformation to the samples from the posterior distributions of N for each of the ten replicates. Finally, we combine the ten estimates of θ, $\hat{\theta}_{IM}$, and their standard errors, s.e.$(\hat{\theta}_{IM})$ by following the combination rules described in Rubin (1987) and undoing the transformation to get results at the original scale.[32] The multiple imputation framework was created based on the Bayesian paradigm and the

[31]See Chao (2015) p. 12 note 2 in the code.
[32]See Gelman et al. (2013) p. 453 for a brief summary of the rules we used to combine the results of the 10 replicates.

combination rules were developed to approximate intervals of credibility with a specified probability (Rubin 1996). In our case we use credibility intervals of equal probability of 95%. As a result of applying the combination rules we have $\hat{N}_{IM} = n + \exp(\hat{\theta}_{IM})$ and $IC_{95\%} = n + \exp(\hat{\theta}_{IM} \pm t \times \widehat{\text{s.e.}}(\hat{\theta}_{IM}))$, where $IC_{95\%}$ is the 95% credibility interval and t is the 97.5th percentile of a t-distribution with degrees of freedom calculated according to the rules described in Rubin (1987).

7 About the estimates

There are two main characteristics of the estimates: bias and variance. In Section 6.2 we described the estimate of variance. Bias represents the distance between the true value and the expected value with the estimate procedure. This occurs for two main reasons: i) variation in the probability of observation or capture heterogeneity and ii) over-stratification. For its part, the variance (or uncertainty) means that, in not having all the records, the estimation process involves uncertainty. That is, the results could change depending on the sources of variation (or uncertainty). Then, the applied methods make it possible to estimate this variance, generating a range where the true value can likely be found. As a result, the estimates do not provide *one* certain figure. There are 4 sources of uncertainty, which we explain in detail in this section: i) the relationship between documented victims and the overall universe of victims, ii) the density of information on the lists, iii) the variation in the probability of observation, and iv) the statistical imputation of missing fields. We explain each source in greater detail in the following sections.

7.1 Bias

Bias occurs when there is a systematic difference between reality and the estimate's results. But why would there be bias? This could happen if the information we have were missing some populations. For example, if none of the databases recorded violence against the Afro-Colombian population, our estimates would be biased: they would not show the violence suffered by this population either. Systematically, our estimates would exclude the violence that Afro-Colombians have experienced, so they would not show us the reality of the violence. Another example would be an exterminated population for which there is no record: our estimates would be biased because they would be omitting these victims. There are two sources of bias: variation in the probability of observation and over-stratification. Below we explain each of them.

- 1. Variation in probability of observation/capture heterogeneity: Variation in probability of observation occurs when different individuals have a different probability of being documented in a list. For example, there are two individuals: a recognized lawyer in a capital city and a peasant woman in a village. If both are killed, the projects are more likely to document the murder of the lawyer (maybe with a probability of 0.9), while the homicide of the peasant would have a lower probability of being recorded (for example, 0.05). So, it is possible that due to randomness in the databases there is no record of murdered peasant women. When making the estimates, we would fail to estimate the underreporting for peasant women, but we would do it well for lawyers. Our estimates would be biased by not including peasant victims.

  The capture heterogeneity is not given only by the profession of a victim. For example, it is likely that in more urban municipalities there is a higher record of victimization. Likewise, it is likely that in the most recent periods there is less underreporting thanks to connectivity or awareness of human rights or ease of reporting. Also, it is possible that in places with more security there is higher documentation, among others. These differences can lead to the estimates being wrong and biased by underestimating the populations with the lowest capture probability.

  To understand where the variation in observation probability comes from, it is helpful to understand the process of collecting the data. The lists are built based on at least three unique characteristics of each institution or organization i) its mandate, ii) its budget, and iii) its logistical capacity. This leads to projects documenting what is within their capabilities and mission, so there will be victims that are not included. For example, some may specialize in acts of violence committed against a specific sector of society (trade unionists, women, Afro-Colombians, etc.) and others may focus on a region of the country (the Pacific, a specific department, a city, etc.).

The capture heterogeneity creates bias, since given the characteristics of the organizations and institutions that document the victims, they will have more or less records of some type of victims depending on the sociodemographic, temporal, or geographical characteristics. Thus, if we made estimates based on lists with similar characteristics, our results would be biased: we would be showing the reality of the victims who are more likely to be registered by the organizations. However, in this project we have more than 100 databases from more than 40 different organizations and institutions, which is why we consider that they manage to cover a large part of the types of victims and allow bias to be mitigated. However, this does not eliminate the bias that could exist due to temporal or regional variables.

- 2. "Over-stratification": One of the ways of dealing with capture heterogeneity is through stratification. A stratum, according to Särndal (2003), is a partition of the population. That is, the union of all the sets reproduces the population and the intersection between any pair of sets is empty. This is done to achieve homogeneity of the characteristic of interest in each stratum and thus reduce uncertainty. In other words, stratification is a strategy by which an initial list is taken and based on it smaller lists are created with records that share characteristics, for example gender and ethnicity. There are then lists of "Afro-Colombian women", "Afro-Colombian men", "indigenous women", and so on for all possible combinations of the categories of the two characteristics. By separating the big list into smaller lists, the model has more information (more lists) to make the estimate.

  In this project we always stratify by year at minimum to recognize the differences that could lead to a smaller or larger probability of observing victims in some years. Intuitively, we might consider to better to use all available variables to stratify and thus "guarantee" that the model can handle potential differences in the observation probability across different groups, it is possible to over-stratify, which can result in biased estimates. Let's consider what over-stratification means: it means that we use so many variables to make small lists of all the combinations, for example "woman, Afro-Colombian, in department X, of age, age Y, in year Z, and victimized by actor J", that these lists would end up having very few records. Not only because few victims would meet these characteristics, but also because not all records have information in all fields, so many would be left out of these mini-lists. By leaving records out, that is, leaving part of the truth out, we would have a biased result: it would be similar to the example of not including Afro-Colombians or the massacred population. In the next subsection we detail why this occurs.

Although there is the possibility that the estimate is not completely free of bias, as is the case with any quantitative research, the advantages over a descriptive analysis are evident for at least two reasons. First, there are ways to identify when an estimate may be biased. We can compare the number of records prior to stratification vs. after stratification. Thus, it is possible to identify if a high percentage of records is being left out and if that could result in statistical bias. As previously mentioned, stratification is a task that requires finding balance: while creating small lists based on characteristics that might help the model to recognize and integrate variation in capture probability to reduce bias, creating too many of these lists would remove information from the model and increase the bias. It is necessary to try different stratification schema to find the optimal balance.

The second advantage is related to the key question of bias: if we were in this situation, would we be underestimating or overestimating the number of victims? In the specific case of the model used in the project, previous research has shown that the estimates tend to have a downward bias. That is, they tend to underestimate the number of victims.[33]

[33] See Rivest (2011) and the limitations section.

7.1.1 Details on bias due to overstratification

Throughout the project we observed that in some cases the estimate of underreporting tended to be very small. In other words, the model suggested that almost all cases were being documented in at least one list. But, by adding several strata, the underreporting increased. For example, if we estimated separately the years between 1986 and 1989, we would find low underreporting. However, when making the estimate for the period between 1986 and 1989 we found a greater underreporting.

As we mentioned in Section 6.1 the MSE algorithm we use is "Bayesian Non-Parametric Latent-Class Capture-Recapture" (Manrique-Vallier 2016). We contacted Professor Manrique-Vallier to discuss the findings we obtained when estimating years individually and jointly. Manrique-Vallier explained to us that, in effect, the LCMCR model, as it is currently implemented, tends to produce conservative estimates as long as they are adjusted with few records and several lists. The reason for this behavior is the a priori specification of capture probabilities at the list level. These probabilities are modeled so that the priors are independent and uniform over an interval $(0, 1)$ to represent the ignorance of their true value. However, when many lists are taken together, this very specification implies that some of these lists may have a high probability of capturing individuals. So, when there are a large number of lists, the prior probability of most individuals being captured by at least one list tends to be considerable. This means that a complete enumeration is likely to be considered a priori.

This effect is not problematic as long as there is enough data to exceed the specification of the prior probability. However, when there are a large number of lists (which would mean that the prior would suggest a complete enumeration) and few records, the prior may be too strong. So the a priori specification would be difficult to overcome, resulting in a conservative estimate close to the observed value.

Professor Manrique-Vallier explained that this is why we sometimes find significant differences between finely stratified estimates and less stratified estimates. If we consider the same number of lists, the a priori specification will suggest almost complete enumeration with the same strength, regardless of the number of records in the substrata. Therefore, if we have a large number of lists (which suggests full enumeration), and some substrata with very few records, estimates in these substrata could be more conservative than in the aggregate, resulting in disparities. We have been aware of this challenge and have created strata that appear to have enough data for us to work around this limitation. Although not always possible, for most strata, we have been able to make adequate estimates.

In future work, Professor Manrique-Vallier suggests that he will create mechanisms to set more specific priors in λ_j instead of assuming "flat" priors.

7.2 Variance

An estimate is based on partial information from reality and has a degree of uncertainty (or "variance" in statistical terms). What does "uncertainty" mean in the context of statistics? It means that part of the results is a consequence of the randomness due to the documented records, that is, if some of the documented records were different from those that are available, then the expected result would be different. But how much could the results change depending on the documented records that are analyzed? This can be measured from the variance and with it we can determine the uncertainty of the estimate, in this way the result is not a specific value but a range where the truth is expected to be found.[34] Uncertainty then allows us to know all possible scenarios with a degree of certainty.

Many times people don't trust a range because it doesn't answer the question of what *the* number is. However, this ignores the fact that if we did descriptive statistics only based on the observed data, we would not know the truth, but rather what has been documented. In this case, there would be no variance because there is absolute certainty about what was documented. Yet, there would be an immeasurable bias: what was documented does not reflect reality, but there would be no way to calculate how distant what happened in

[34] This statement is based on the interpretation of the Bayesian "credibility interval". Frequentists use "confidence intervals", which use a different interpretation of uncertainty.

the conflict is from what was documented. On the other hand, thanks to the estimates it is possible to reduce the uncertainty of how many victims were left out to a measurable and interpretable range. This range allows us to examine patterns and trends. In the specific case of this project, one of the most interesting estimates is the universe of victims. That is, both the victims documented by the organizations and entities and those that were not registered. Like any estimate, we have uncertainty here (our estimates are represented in ranges). But why is there uncertainty? There are four reasons for this:

- 1. The relationship between documented victims and the universe of victims: The more documentation there is of the victims compared to the universe (the total number of victims, both documented and undocumented), the less uncertainty there will be. Intuitively, the fact that organizations and institutions have the capacity, in aggregate, to document a situation that is closer to reality, the less uncertainty we will have about what reality is like. The same thing happens with the model: as we have less missing information about reality, since the organizations managed to capture a high percentage of the total cases (i.e., they left a low percentage unrecorded), we will have more certainty about how many cases could have gone unrecorded. Therefore, stronger the relationship between registered victims and total victims, the lower the variance will be.

- 2. The density of information on the lists: The more coincidences there are in the victims registered by different organizations, the lower the variance. To understand the intuition behind this, we can imagine that the estimation seeks to answer how big a dark room is.[35] The form of knowing the size is through rubber balls (equivalent to the records of each list), which are thrown in the room. As we hear more hits between the balls, this will mean that the room is smaller: there is less room for them to move without bumping into each other. This intuition is transferable to the case in question. The more coincidence between the lists, in other words, the higher the density, the less uncertainty the estimate will have.[36]

- 3. The variation in the observation probability: Although our model is currently the best to include the variation in capture probability, it is possible that with the documented information, such as sex, ethnicity, year, and municipality, it is not enough to show the differences in it.

To estimate the universe of victims, we take the intersection of all the lists. In other words, we know exactly on which lists the victim “Juan Pérez, murdered on May 2, 1990 in Medellín, Antioquia” is documented. Depending on the number of intersections between lists, that is, of matches between records, there will be more or less uncertainty. If we have more coincidences, the variance will be lower: intuitively the lists are capturing the same population, so it is likely that the universe of victims is similar to that documented. In this case there will be less uncertainty. Therefore, the inability to fully capture the variation of the observation probability creates uncertainty.

- 4. Statistical imputation of missing fields: In order to know the universe of victims we need to carry out record linkage on our large database of input records, eliminating duplicate records. After completing record linkage, we need to complete the fields that are missing information. That is, assign race, ethnicity, sex, municipality and whether or not the event occurred within the context of the armed conflict to the records that are missing information in these fields. This task of completing the missing fields is known as “imputation” and was explained in detail in Section 4. The higher the proportion of records with complete information relative to the total number of records, the lower the uncertainty.

In short, there are four sources of uncertainty that cause estimates to move within a range. While it is tempting to use recorded data and stick to descriptive statistics, doing so would leave out all undocumented

[35]HRDAG uses this analogy in most studies where MSE is used.

[36]We know which databases documented each victim thanks to the results of record linkage. From this process, we have a single database where each row corresponds to a unique, documented victim and we create a series of columns, each corresponding to one of the lists, indicating whether a victim was documented by a particular list. These cells take the value of 0 if the victim was not registered in the project’s list and 1 if the victim was registered.

victims, thereby incurring bias. The four sources of uncertainty include the relationship between documented victims and the total universe of victims, the density of lists we have, the variation in the probability of observation, and the statistical imputation of missing fields. Although uncertainty does not allow us to answer what *the number* is, it does allow us to identify trends and patterns. The statistic reported as the "estimate" corresponds to the statistic that is most probable given the posterior distribution.

8 Limitations and future work

"In all fictional works, each time a man is confronted with several alternatives he chooses one and eliminates the others (...) He creates, in this way, diverse futures, diverse times which themselves also proliferate and fork" (Jorge Luis Borges, 1944)

In the short story "The Garden of Forking Paths" by Jorge Luis Borges, the characters discuss a possible labyrinth created by the protagonist's ancestors. The labyrinth represents the idea that whenever the characters reach a point where they must choose two paths, the story splits into two copies and the characters follow both paths.

This tale has been used by great statisticians, such as Gelman and Loken (2014), to describe the process of scientific research. Whenever researchers make a project design decision, one path is followed, even though there is another one. We are aware that there are many paths that we do not take in our work. Each time we did, we left unexplored trails that could have led to other results.

Throughout this document we have shown the decisions we made and we have justified why we considered these decisions to be the best. We believe that we chose the paths that tend to influence the results least, and therefore our results should remain as close to the original data as the models allow. Our results reflect the combination of assumptions and decisions we made. We believe that the assumptions are possible and plausible and that the decisions are the best. However, there are other possible paths.

In this section we compile those moments and present the limitations of the project based on the data, record linkage, statistical imputation, and multiple systems estimation. For each of these elements, we return to the key decisions, the why, the other possible paths, and the consequences of having opted for a particular decision.

8.1 Data

Any quantitative research is limited by the data it has. We start from a fundamental assumption and identify five limitations related to the data.

The assumption from which we start in the project is that the vast majority of the records we receive are victims of violence that actually occurred. We are not concerned that they are not because they did not happen, but because they do not represent what the database tells us they represent. For example, because there could be records within homicide databases that have been attempted homicides or because the year that was left in the database is the year of the complaint and not of the occurrence, among other possible systematic errors in the data.

Regarding the limitations, to begin with it was not possible to include the data from SIJUF (from the Prosecutor's Office) in the analyses. We were also unable to include SPOA data for acts of violence other than "homicide". Therefore, due to the design of the Prosecutor's system, two of the largest non-specialized sources of information in the country were excluded. We will explain this in detail below in the 8.1.1 subsection.

Second, as shown in the section on documented violence, the data about displacement is concentrated in the RUV. Since MSE is based on overlaps (and non-overlaps) between lists, the estimates are affected by the high importance of the RUV. So, we recommend working with the replicate files created through statistical imputation to study this violence. That is, it was not possible to calculate the underreporting or use MSE.

Third, although we also intended to include exile in this project, there was not enough data to do so. So, it was excluded from the research.

Fourth, although the CEV has been studying the conflict since 1958, data prior to 1985 is scarce. Even if it were possible to impute and estimate, the variance would be very high, so it is not recommended to use the results for this period of time. This is the first decision we made and it implies that each type of violence has adequate data for different periods. As we show in the tables, homicide has the most data available between 1985 and 2018, enforced disappearance between 1985 and 2016, kidnapping between 1990 and 2018, child recruitment between 1985 and 2017, and forced displacement between 1985 and 2019 (but without statistical imputation or estimates).

Finally, one of the hypotheses often held regarding victimization is that the activities the victim engaged in—such as being a human rights defender, journalist, or trade unionist—are a fundamental differentiating element. However, this information was scarce in the databases we received for this project, so it was not possible to statistically impute missing fields of this variable. However, some databases specialized in violence faced by a certain population. This is the case, for example, for trade unionists. In situations where these specialized databases were available for certain sectors, we were able to perform specific analyses. However, this was not the case for all populations of interest and there were relevant and important groups that could not be analyzed.

8.1.1 About the data of the Attorney General's Office

Two of the large official databases in the country that include acts of violence are the Accusatory Oral Criminal System (*Sistema Penal Oral Acusatorio* - SPOA) and the Justice Information System of the Attorney General's Office (*Sistema de Información de Justicia de la Fiscalía* - SIJUF) of the Attorney General's Office. However, these systems use **cases** as their unit of observation, not **people** or **violations**. That is, in the same case there is more than one person. For example, the victim and her parents, the victim and the complainants, etc. So, the question of "who did what to whom?" cannot be answered. (Ball 2008).

However, the Attorney General's Office does have a variable for "homicide" (*homicidio consumado* in Spanish) within the SPOA. This allowed us to use include data about homicides. That is, we do not have SPOA data for kinds of violence other than homicide. On the other hand, due to SIJUF design issues, we could not include it in the analysis.

The Prosecutor's Office has other databases that were delivered to the project apart from the SPOA and SIJUF. Specifically, these are databases related to child recruitment, which were seized by the institution. These databases were used within the project.

In addition, the JEP verified cases of the Patriotic Union (*Unión Patriótica* - UP) and social movement leaders registered by the Attorney General's Office, which we were able to include for this reason.

In short, the SPOA and the SIJUF are designed to study "cases" and not "victims." Therefore, we use the variable "homicide" to identify the victims of homicides from the SPOA. In summary, from the Attorney General's Office we only include victims of consummated homicide and databases on recruitment, social movement leaders, and members of the UP.

8.2 Record linkage

The results of the record linkage, as described in Section 3, are the basis from which the statistical imputation of missing data and the multiple systems estimates were made. We identified two decisions we made that could have been different: the minimum fields considered and the relationship between false positives and false negatives. Next, we explain each one of them.

8.2.1 Minimum fields

To be able to link records it is necessary to have some minimum field information. Specifically, we opted for one first name, one last name, year, department, and type of violence. However, not all records had these fields and, therefore, there are victims who are not explicitly included in the project. However, these are included in the estimates of underreporting we construct using MSE.

We could have taken other minimum fields, such as name, last name, and year, or first name, last name, department, and year. However, we believe that given the large number of records, it was necessary to have more information so that the oracle could generate accurate training data. We believe that with the five variables we chose there is a good balance between the amount of information that the oracle has available and the number victims that are excluded because they do not have all the minimum fields.

8.2.2 False positives vs. false negatives

This trade-off is specific to the record linkage model, which is a classification problem. As its name implies, the algorithm seeks to classify an object of interest into a group. In the case of record linkage, we want to learn whether a record is part of group of records.

In data science there are different useful measures to study the performance of a model. As we explained in Section 3, among them are recall and precision. Recall shows us the performance of a classifier against false negatives. That is, the records that are identified as different victims, even if they are the same victim. For its part, precision shows us the performance against false positives. That is, the records that are identified as the same victim even though they are not.

It is possible to prioritize one of the two measures. On the one hand, we could decide to focus on minimizing false negatives (maximizing recall) or we could minimize false positives (maximizing precision). In the case of record linkage, we opted for a balance between the two, accepting a slightly higher rate of false negatives, rather than failing to identify different people. We chose this to balance the risk of over-identifying and under-identifying. That is, between not recognizing that two or more records can be the same victim, and that two or more records are the same victim. In Table 17 we present the result of the model, where the balance between false negatives and false positives is evident, with a slightly higher number of false positives than false negatives

This is another decision which could have been different. Here the the quantity of possible paths is enormous, according to different ratios between false negatives and false positives. There are many possible scenarios for the final deduplicated records based on this balance. From the project, we believe that the best decision was to allow a slightly higher false positive match rate without allowing the difference to become too extreme. Otherwise, we would have ended up double-counting some victims or with an undercount. However, using a different ratio of false negatives and false positives, the results of record linkage would have been different. Finally, it is important to mention that scientific literature has shown that even if record linkage performs poorly, the results of MSE are still reliable (J. Johndrow, Lum, and Dunson 2018).

8.2.3 Cases of ambiguity

One of the key issues in statistical imputation of missing fields is the decisions made in cases of ambiguity. As we show in Section 3, the last steps of record linkage are union and export. At that point it is necessary to decide what to do with the records that have ambiguity in some variables. That is, they have different information in the same field.

The decision to prioritize certain types of violence was done based on the analysis of experts on the armed conflict. Specifically, we decided that if, after deduplication, the same record had child recruitment and disappearance and/or kidnapping, we would code the record as child recruitment. If a record had kidnapping and disappearance, we would code the record as kidnapping.

The decision to prioritize types of violence is made based on the analysis of armed conflict experts. Specifically, it was decided that if when doing the deduplication we have homicide and displacement for the same record, homicide is taken. If the same registry has child recruitment, enforce disappearance and kidnapping, child recruitment is taken. While if there is enforced disappearance and kidnapping, enforced disappearance is considered.

We prioritized ethnicity with the aim of prioritizing groups that have historically been discriminated against. In the event that a record had two different ethnic groups, indigenous was prioritized, followed by Afro-Colombian, and then Rrom. Based on the expertise of researchers working on ethnicity, we learned that in some cases indigenous, Afro-Colombian, or Rrom people discover that they have been registered in databases as "mestizo". So, we decided to prioritize the references to indigenous, Afro-Colombian, and Rrom when they occurred in coreferent pairs.

Ambiguity, however, is not limited to the type of violence or the ethnicity of the victim, it also applies to other variables. In these cases, we take the non-missing value that most frequently appears. In the case of ties, we chose the value of the field randomly among the most frequent values. However, these rules could have been different and the results of the statistical imputation would have also been different. Yet, we believe that we made the best possible decisions in line with knowledge about the conflict, with respect to recognition of ethnic populations, and probability.

8.3 Statistical imputation

Once record linakge is complete, some records will still have missing fields. Then, it is necessary to make the statistical imputation. The decisions to be made in this step can be divided into three large groups: those related to the use of specialized and non-specialized databases, those related to the variables used to provide information in the statistical imputation, and those related to the algorithm used.

8.3.1 Specialized datasets vs. non-specialized datasets

One of the big decisions to be made in the project was whether to limit ourselves to the use of specialized data sets that only documented violence that occurred in the context of the armed conflict or to also use non-specialized ones. We decided to include both types of databases in order to include a significant amount of official data. Specifically, those from INML-CF, FGN, and PONAL. The direct consequence of this is that we had to statistically impute the variable "is conflict", as we explain in the next subsection.

8.3.2 Extensions of the statistical imputation

As we show in Section 6.2, different variables have different proportions of missing information. The most critical ones are: is conflict; is enforced disappearance, and the perpetrator.

Our first decision was regarding training data. We chose probabilistic rules based on expert criteria for the "is conflict" and "is enforced disappearance" variables. However, we could have done this process deterministically or with different rules. We think we made the best decisions for two reasons. First, because the rules were created by expert researchers from the CEV based on knowledge of the conflict dynamics. Second, our probabilistic approach recognizes that not all instances of violence that comply with a rule meet (or do not meet) a certain condition. Although we trust the criteria of the experts who defined the rules, we must recognize that, if we took other rules, the statistical imputation of missing fields would also change.

The other major challenge relates to how we addressed perpetrator. In our case, we decided to divide the guerrillas into groups: ARC-EP, ELN, and "others". However, we could have grouped them all together as "guerrillas" and the results would have been different. Given the magnitude of the documented violence of the FARC-EP and the ELN in the years of the study relative to that of other guerrilla groups, we considered it important to analyze their specific cases as an exercise contributing to the truth. In the next subsection we give more details about this decision.

We decided to use all possible variables for imputation. We only eliminated variables that were constant or collinear. We could have done the statistical imputation of missing data only based on certain variables. However, we believed that by providing the model more complete information, but without falling into the curse of dimensionality by adding a large number of variables, which was possible thanks to the creation of the support variables, we could achieve better results.

As a result, one of the fundamental decisions we made was to create support variables to help the imputation (see Section 4.4). We opted for this approach to be able to use all of the information from the original databases and learn to identify clues that could help us answer whether or not an act of violence occurred within the context of the armed conflict, whether or not a disappearance was enforced, the sex of the victim, whether or not they were a minor, the alleged perpetrator, and the municipality in which the event occurred. In the case of armed conflict or enforced disappearance, a clue, for example, could be that the narratives includes that the person had strong political beliefs. It could also be the mention of specific confrontations or references to the groups as "subversive" or "terrorist". We want to emphasize that these words or phrases are not chosen by us. These are not rules, nor are they identified by humans. They are identified by the model because the words and phrases are correlated with the variable to be predicted. We believe that this is the best decision because it allows us not to lose valuable information from the original records, especially the text, but the statistical imputation could be done without the help of support variables.

8.3.3 Imputation of the unknown guerrilla records

As we previously mentioned, some records have "guerrilla" as the perpetrator without specifying which group. Distinguishing between different guerrilla groups is essential for the study of patterns of violence, which could be different for each armed group, including among the various guerrilla groups (Gutiérrez-Sanín and Wood 2017). We used statistical imputation model with the help of the support variables detailed in Section 4 to assign these cases to a specific armed group. Given the period of time of analysis of the project, we limit ourselves to three categories: FARC-EP, ELN, and other guerrillas.

Prior to imputing the alleged perpetrator for cases that have unspecified "guerrillas," we studied the records before exporting the results of the record linkage process. At that point we know which records from the various sources belong to the same victim. We found that the behavior of victims with at least one unspecified "guerrilla" record differs from the behavior of victims with at least one record belonging to a specific guerrilla group. When studying all the records that belong to the same victim, in the 44% of the cases in which at least one of the records has unspecified "guerrilla" there is at least one other record that belongs to the same victim and is recorded as victims of paramilitaries, the State, "multiple", or another group (such as criminal gangs). In other words, the information on the perpetrator contained in the various records that refer to the same person is contradictory. Instead, this contradiction appears in 15% of records of victims with at least one record with "guerrilla-FARC", and 20% of records of victims with at least one registration with "guerrilla-ELN". That is, unspecified "guerrilla" records are linked to groups other than guerrillas (mainly paramilitaries) at a rate more than double that of records that do have a specific guerrilla group. This high rate of contradiction makes us skeptical of the data quality of the unspecified "guerrilla" value. We then included this information in the support variables so that the records could be statistically imputed.

There are several mechanisms that could explain the inaccuracy of these records. For example, it may be that when reporting the violence, the witness did not know with certainty which group was responsible or confuses one group with another. It could also be a conscious decision to reduce the risk of retaliation for reporting violence from a specific group. These are speculations, we do not know why there is this deficiency in these records, we only know that the unspecified "guerrilla" information in these cases is not consistent with other records linked to the same victim.

Although we chose to impute the perpetrator with the statistical model detailed in Section 4, we could have taken another approach, such as assigning the cases with the unspecified "guerrilla" value by randomly selecting FARC, ELN, or another guerrilla group, following the proportion from the observed data. That is, 0.8 to FARC-EP, 0.15 to ELN, and 0.05 to other guerrillas. From there would be possible to do the imputation. We find that the ratio of those presumed responsible changes as follows: the estimate of the proportion attributed to paramilitaries is reduced from 0.45 to 0.4[37] and the estimate for FARC increases from 0.21 to 0.27.[38] The high level of records with unspecified "guerrilla" and perpetrators other than guerrilla especially affects the "multiple" category, increasing it from 0.06 to 0.12[39] while the other perpetrators are decreased a bit. We note that using this proportional imputation, the estimated presumed responsibility ratios for paramilitaries and FARC-EP are within the estimated credibility intervals without using unspecified "guerrilla." So, the main impact of using unspecified "guerrilla" is that the category of "multiple" perpatrators is increased.

To be clear, this experiment is not a reliable finding. It is a sensitivity analysis to show the maximum possible impact of the allocation of unspecified "guerrilla" information. Using this imputation would not change the substantive finding that paramilitaries are the main suspects in homicides, with the FARC-EP in second place. What this experiment emphasizes is that using "guerrilla" without directly specifying it amplifies the contradiction in the presumed perpetrator field, leading to a higher estimate of multiple perpetrators.

Two key problems with this approach are worth studying. First, the assignment of the proportions does not include the high level of contradiction that we observe. This experiment assumes that the probability of a specific group within unspecified "guerrilla" is the same as in the observed data for groups that do have a specific guerrilla assigned, even though the detail is missing. We do not know the true distribution, but it is very likely that a substantial fraction of unspecified "guerrillas" should be assigned to multiple or paramilitaries.

Second, as we show, the data for the unspecified "guerrilla" variable has uncertainty. We are sure that some records should be attributed to FARC-EP and a smaller fraction to ELN. However, other records should be attributed to paramilitaries, other groups, and State agents. We do not know the correct proportion for the different actors. Determining these proportions and determining their certainty is exactly the task of statistical imputation. In particular, if there is uncertainty in the data, the best outcome is for the estimated proportions to have credibility intervals wide enough to represent the range of possible values. We feed the information to the imputation model through the support variables and stratification schemes, and the model shows us the range of variation. Sometimes we need to assign some inaccurate cases with probability rules from experts, as we did for the "is conflict" and "is enforced disappearance" variables. In these cases we try to assign a size to the intervals. If we do not know the true value of a variable, it is better to leave it unknown so that it can contribute to the uncertainty, as measured by the credible interval.

We did other experiments with unspecified "guerrilla", such as assigning the proportion to the three guerrilla categories and some to multiple, leaving others for the statistical imputation model. All these experiments led to results similar to those we show in Table 13. That is, in each of our experiments, the estimates for the paramilitary groups and the FARC-EP are within the credibility interval shown in Table 13. This means that for the paramilitary and FARC-EP results, in a statistical sense, the experiments were not significantly different than the results in Table 13. There are two consistent differences between our experiments and the Table 13: the multiple category is always larger than in the Table 13 and the credibility intervals are smaller. Neither of these results seems likely to us, so we prefer the support variables approximation.

The best approximation would be a more advanced statistical imputation model in which unspecified guerrilla information would guide the statistical imputation without over-determining the results. We use support variables for this and we think it is the best way to do it. With the support variables, we are not removing the "guerrilla" information from the records, but rather using it with the rest of the information contained in the record to make a probabilistic prediction of the alleged perpetrator. Also the assumption about the behavior of missing data in our imputation model—"missing at random"—is much more justifiable than the improbable assumption required to impute these cases in a proportional way and is not ignorant of the fact that some of these records probably correspond to paramilitaries, the state, or other armed groups. Without having more information on these records that only report "guerrilla" as the alleged perpetrator, our approach to imputation is the most technically rigorous that is possible with the information and technology at our disposal.

[37]With a 95% credible interval of 0.4 (0.35–0.45).
[38]With a 95% credible interval of 0.27 (0.23–0.32).
[39]With a 95% credible interval of 0.12 (0.09–0.14).

8.3.4 Classification model

There are different classification models that can be used for the full conditional specification of the statistical imputation. Until Phase 3 of the project, we used CART, a decision tree that has several advantages. Among them is that it is robust to collinearity between the predictor variables and the ability to fit data of different types (binary, categorical, or continuous). CART also fits non-linear relationships and complex interactions without making assumptions about the parameters (Burgette and Reiter 2010; Akande 2015). However, its computational cost is high.

In Phase 4 of the project we opted to use PMM, as mentioned in Section 6. We chose this approach because it has different advantages over CART. First, it preserves variance better than CART; we can trust the error estimates of this algorithm more. Second, it's about 15 times faster, which is computationally important given the very high number of records we used in the project. Third, PMM can also fit to different data types and is robust to poorly spefieid models. As a result, managed to have more efficient results. However, the results will vary according to the algorithm that is selected.

8.4 MSE

The third and final step is MSE. Again, there were a number of possibilities and we made the decisions that we considered to be the most appropriate. These decisions are specifically related to the estimation method, stratification, and the variance estimation method. Additionally, we identify a limitation related to the variance.

8.4.1 Method

There are different estimation methods for MSE. We use the LCMCR method. We could have used another method, such as Capture-Recapture Estimation using Bayesian Model Averaging (DGA). DGA allows us to have a more specific and interpretable prior probability (see Hoover Green and Ball 2019), but LCMCR allows us to use more sources, since DGA is limited to 5 sources.[40]

The cost of using LCMCR is that, as we show in Section 7, this method is downward biased. So our results underestimate violence. However, we believe that it was better to assume this bias in order to take advantage of the rich information that dozens of organizations have documented.

8.4.2 Stratification

The more variables we use to stratify the data, the less capture heterogeneity there will be. However, the cost of further stratification is greater bias. Returning to the example used in Section 7 in which there was a lawyer in Bogotá with a capture probability of 0.9 and a peasant woman in a village with a capture probability of 0.05. Here we could stratify by different variables: year, department, sex, and even the municipality. We would then be creating small lists of records that comply with the possible combinations of these variables. However, if we limit ourselves, for example, to the municipalities of only one department, we could exclude victims like the peasant woman with a very low probability of being captured. As a result, the model will not have a way to estimate the level of underreporting. Thus, there is a relationship between the sample size (which depends on the stratum) and how well the model estimates all victims, including those unlikely to be documented.

Finally, it is important to emphasize that stratification should serve the analytical needs of the project. For example, if we need to compare the number of estimated victims by armed group, we must stratify by armed

[40] We could also have used a frequentist method, like Rcapture, but frequentist methods are also limited in the number of lists they can use, like DGA, or are limited in the number of interactions. We think it is hard to justify this in practice because of the number of lists we have. There would be no way to use all of them and we would need to justify the selection of some over others.

group. Choosing a stratification schema is always a balance between the amount of data in each stratum, control of capture heterogeneity, and the analytical needs of the project.

8.4.3 Method to estimate the variance

As we show in Section 6 we follow Chao (1987) and transform the estimates. This is the transformation we apply to the samples of the posterior distributions for each of the ten replicates. However, LCMCR leads to a Bayesian estimate, and therefore it is theoretically possible to make a fully Bayesian estimate of the variance. Following Zhou and Reiter (2010) it would be possible to combine all samples from the posterior distributions. However, Zhou and Reiter (2010) propose a minimum of 30 replicates and at best 100 replicates.

However, on average, the 10 replicates we have in each stratum require between 3 and 5 CPU hours to estimate with LCMCR. If we were to estimate between 3 and 10 times as many replicates, that would result in between 9 and 50 CPU hours in each stratum. In the case of sex, if we stratified by year and sex between 1990 and 2015, we would have 50 strata. This would suggest between 450 and 2,500 CPU hours for this analysis alone. With our current technology and budget, this was not possible.

With computational improvements, it would be recommended to use the method proposed by Zhou and Reiter (2010). In Figure 18 we show the comparison of the credible intervals using the Chao (1987) method against the credible interval using the Zhou and Reiter (2010) method. On the left side we observe the 10 posterior distributions of the 10 replicates for the same stratum. The green lines show the credible interval following Chao (1987), while the blue lines show the credible interval following Zhou and Reiter (2010). So, we see that with Chao (1987) the estimated variance is greater than the variance following Zhou and Reiter (2010). For its part, on the right side we see all the subsequent imputations together.

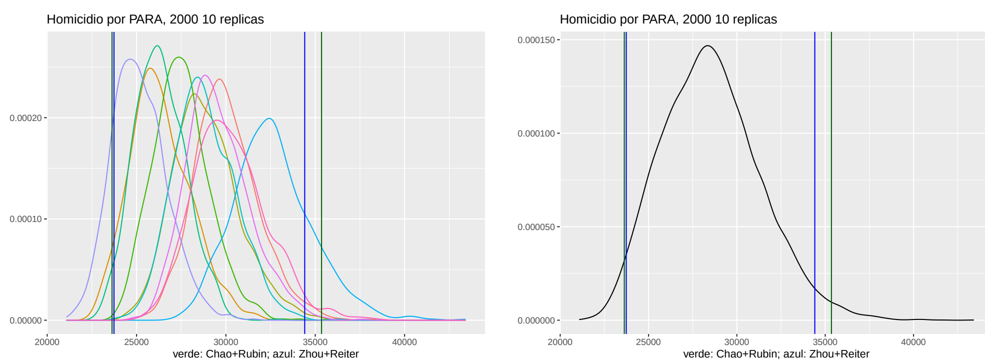

Figure 18: Credibility examples for paramilitary in 2002.

The difference we see between these two approaches could be related to an idea from Rubin (1996) called "super-efficiency". Rubin observed that when an imputation model uses more information than the estimate model, the variance estimated by his rules will tend to be conservative. "Conservative" in this context means that the estimated variance could be larger than the true variance. In our case, the imputation model uses many variables that are not in the estimation model. Specifically, support variables are used only in imputation. So the MSE estimates will be "super efficient". As a result, variance estimated by Rubin's method will tend to overestimate the variance, which might explain why the credible interval of Zhou and Reiter (2010) is a bit smaller than the credible interval estimated by using the transformation proposed by Chao (1987) and the rules proposed by Rubin (1987). See Cro (2017) for further explanation of this effect.

8.5 Future work

In this section we have emphasized the current limitations and other paths we could have taken. As we show, some of the limitations are due to computational capacity, others because science has not yet made the respective developments. We have identified three possible improvements for future work.

First, the prior distribution could be modified in the LCMCR method. This would allow the capture heterogeneity to be incorporated. For example, information about the likelihood of records being documented or not documented could be included.

Second, the correlation of the observations in the strata could be incorporated. With the current method, an estimate is made for each stratum separately, assuming their independence. In real life, the strata are likely to be correlated. This would also allow us to “borrow” information between strata to make estimates in smaller strata. For example, if a particular source had a low observed probability of homicides in Casanare in 2003, it is likely that this source would also have a low observed probability of homicides in Arauca in 2003 and Casanare in 2002. If there were not enough information for Casanare in 2003, incorporating the correlation of the observations would allow us to take information between the strata to make the estimates. Such a model would also allow us to make estimates for smaller strata, such as municipalities or specific months.

Third, statistical imputation of missing fields and estimation could be integrated into a single model. As previously shown, linking the models together leads to propagation of the variance from statistical imputation into the estimation. This could be changed when integrating the models.

Manrique-Vallier, Ball, and Sadinle (2022) present advances and future work on multiple systems estimation/

9 Appendix

9.1 Acknowledgments

We are grateful to the donors whose support made this work possible, including core support from the John D. and Catherine T. MacArthur Foundation, The Oak Foundation, and the Foundation to Promote Open Society, and project-specific support from the Swiss Embassy Human Security Division, the British Embassy Conflict, Stability and Security Fund, and the Justus Liebig University Giessen, with funds awarded by the German Foreign Federal Office.

9.2 About the organizations that were part of the project

9.2.1 Commission for the Clarification of Truth, Coexistence and Non-Repetition (*Comisión para el Esclarecimiento de la Verdad, la Convivencia y la No Repetición* - CEV)

Within the framework of the Final Agreement for the termination of the conflict and the construction of a stable and lasting peace, signed between the Government of Colombia and the FARC-EP, through Legislative Act 01 of 2017 and Decree 588 of 2017, the Commission for the Clarification of Truth, Coexistence and No Repetition was created. This is a temporary and extrajudicial mechanism of the Comprehensive Truth System, Justice, Reparation and Non-Repetition (SIVJRNR), to find out the truth of what happened within the context of the armed conflict and contribute to clarifying the violations and offenses committed during it and offer a broad explanation of its complexity to the whole of society.

9.2.2 Special Jurisdiction for Peace (*Jurisdicción Especial para la Paz* - JEP)

The Special Jurisdiction for Peace (JEP) is the justice component of the SIVJRNR, created by the Peace Agreement between the National Government and the FARC-EP. The JEP has the function of administering transitional justice and knowing the crimes committed within the framework of the armed conflict before December 1, 2016. The existence of the JEP may not exceed 20 years.

The JEP was created to satisfy the rights of victims to justice, offer them truth and contribute to their reparation, with the purpose of building a stable and lasting peace. The work of the JEP focuses on the most serious and representative crimes of the armed conflict, according to selection and prioritization criteria that have been defined by the law and the magistrates. In particular, the JEP may know about the crimes committed by ex-combatants of the FARC-EP, members of the public, other state agents and civil third parties. Regarding the latter two, the Constitutional Court clarified that their participation in the JEP would be voluntary.

9.2.3 Human Rights Data Analysis Group

The Human Rights Data Analysis Group is a non-profit, non-partisan organization15 that applies scientific methods to the analysis of human rights violations around the world. This work began in 1991 when Patrick Ball began developing databases for human rights groups in El Salvador. HRDAG grew at the American Association for the Advancement of Science from 1994–2003, and at the Benetech Initiative from 2003– 2013. In February 2013, HRDAG became an independent organization based in San Francisco, California; contact details and more information is available on HRDAG's website (https://hrdag.org).

HRDAG is composed of applied and mathematical statisticians, computer scientists, demographers, and social scientists. HRDAG supports the protections established in the Universal Declaration of Human Rights, the International Covenant on Civil and Political Rights, and other international human rights treaties and instruments. HRDAG scientists provide unbiased, scientific results to human rights advocates to clarify human rights violence.

9.3 Data

To see a full description of the datasets that were used in the project, see the Spanish report available here.

9.4 Scripts

9.4.1 Scripts used for the "is conflict" variable

```
cat(readLines(files$lideres), sep = "\n")
```

9.4.1.1 FGN - Líderes

```
## #
## # Authors:     Valentina Gómez
## # Maintainers  Valentina Gómez, PB
## # Copyright   2022, HRDAG,
## # =========================
## # CO-SIVJRNR-data/individual/FGN/is-ca/src/is-ca-fgn_lideres.R
##
## # ------ setup
## Sys.setlocale("LC_CTYPE", "en_US.UTF-8")
## set.seed(19481210)
##
## require(pacman)
## p_load(argparse, dplyr, here, arrow, assertr, stringr, lubridate, readr)
##
## stopifnot(endsWith(getwd(),"is-ca"))
##
## parser <- ArgumentParser()
## parser$add_argument("--input_fgn_lideres",
##                     default=here::here("individual/FGN/clean/output/fgn_lideres.parquet"))
## parser$add_argument("--examples",
##                     default = "individual/FGN/is-ca/hand/example_fgn_lideres.csv")
## parser$add_argument("--output",
##                     default = "output/fgn_lideres.parquet")
## args <- parser$parse_args()
##
## # ------- function
##
## # ------- main
##
## # NB: we are only retaining homicidios for consumado=SI; all is_conflict
## data_fgn_lideres <- read_parquet(args$input_fgn_lideres) %>%
##  mutate(is_conflict = 1)
##
## # names(data_fgn_lideres)
## # nrow(data_fgn_lideres)
## #
## # table(data_fgn_lideres$conflicto_armado_caso, useNA = "always")
## # table(data_fgn_lideres$vinculacion_persona, useNA = "always")
## # table(data_fgn_lideres$grupo_delito, useNA = "always")
```

```
## #
## # caso_is_ca = c("FARC-EP", "PARAMILITARES", "OTRAS GUERRILLAS", "SI")
## #
## # viol_is_ca = c("DESAPARICION FORZADA", "HOMICIDIO DOLOSO", "DESPLAZAMIENTO")
## #
## # viol_na_ca = c("AMENAZAS", "FEMINICIDIO", "LESIONES PERSONALES")
## #
## #
## # conflict_fgn_lideres <- data_fgn_lideres %>%
## #   mutate(is_conflict = case_when(grupo_delito %in% viol_is_ca &
## #                                    conflicto_armado_caso %in% caso_is_ca ~ as.integer(1),
## #                                  grupo_delito %in% viol_na_ca ~ NA_integer_,
## #                                  grupo_delito %in% viol_is_ca &
## #                                    conflicto_armado_caso == "NO" ~ as.integer(0),
## #                                  is.na(conflicto_armado_caso) &
## #                                    grupo_delito == "DESPLAZAMIENTO" ~ as.integer(1),
## #                                  conflicto_armado_caso == "SIN GRUPO IDENTIFICADO" &
## #                                    grupo_delito %in% viol_is_ca ~ as.integer(1),
## #                                  TRUE ~ NA_integer_))
## #
## # table(conflict_fgn_lideres$is_conflict, useNA = "always")
## #
## # # 0    1   <NA>
## # #566  348   5186
## #
## # nrows <- nrow(conflict_fgn_lideres)
## #
## # sample <- conflict_fgn_lideres %>%
## #   select(recordid, is_conflict)
## #
## # sample <- sample_frac(sample, 0.4)
## #
## # write.table(sample, file = args$examples, sep = "|", quote = FALSE, row.names = FALSE)
##
## write_parquet(data_fgn_lideres, args$output)
##
## #done!
```

```
cat(readLines(files$jtr), sep = "\n")
```

9.4.1.2 FGN - UP

```
## #
## # Authors:     SAC
## # Maintainers: SAC, PB
## # Copyright:   2022, HRDAG, GPL v2 or later
## # ========================================
## # CO-SIVJRNR-data/individual/FGN/is-ca/src/is-ca-fgn-jtr-up.R
##
## # ----- setup
## Sys.setlocale("LC_CTYPE", "en_US.UTF-8")
```

```
## set.seed(19481210)
##
## require(pacman)
## p_load(argparse, dplyr, here, arrow, assertr, stringr, lubridate)
##
## stopifnot(str_detect(Sys.getlocale(), "LC_CTYPE=en_US.UTF-8"))
## stopifnot(str_ends(getwd(), "FGN/is-ca"))
##
##
## parser <- ArgumentParser()
## parser$add_argument("--input",
##                     default=here::here("individual/FGN/clean/output/up-fgn-transicional.parquet"))
## parser$add_argument("--output",
##                     default = "output/up-fgn-transicional.parquet")
##
## args <- parser$parse_args()
##
## # ------- function
##
## # ------- main
##
## up_fgn_jtr <-  read_parquet(args$input)
##
## up_fgn_jtr <- up_fgn_jtr %>%
##   mutate(is_conflict = 1)
##
## glimpse(up_fgn_jtr)
##
## up_fgn_jtr %>% write_parquet(args$output)
##
## #done
```

```
cat(readLines(files$sijuf_cv), sep = "\n")
```

9.4.1.3 FGN - SIJUF - CEV

```
## #
## # Authors:     VG
## # Maintainers  VG, PA, PB, JGD
## # Copyright   2022, HRDAG,
## # =========================
## # CO-SIVJRNR-data/individual/FGN/is-ca/src/is-ca-FISCALIA_datos_sijuf_cv.R
##
## # following the email from Folco Zaffalon (8 abr 2022, 17:16 PDT),
## # and the accompanying file
## # ‘SIJUF_casos_perpetrador_ocupacion victima_delitos.docx‘
## #
## # Verde = SI CONFLICTO (1)
## # Azul = MUY PROBABLE 0.75~(is_conflict == 1)
## # Amarillo = NO SE SABE 0.5~(is_conflict == 1)
## # Naranja = POCO PROBABLE 0.75~(is_conflict == 0)
```

```
## # Rojo = NO CONFLICTO (0)
##
## # ------ setup
## Sys.setlocale("LC_CTYPE", "en_US.UTF-8")
## set.seed(19481210)   # <-- the seed is really important in this run!
##
## require(pacman)
## p_load(argparse, dplyr, here, arrow, assertr, stringr, lubridate, readr)
##
## stopifnot(endsWith(getwd(),"is-ca"))
##
## parser <- ArgumentParser()
## parser$add_argument("--input_SIJUF_CEV",
##                     default=here::here("individual/FGN/clean/output/FISCALIA_datos_sijuf_cv.parquet")
## parser$add_argument("--examples",
##                     default = "individual/FGN/is-ca/hand/example_SIJUF.csv")
## parser$add_argument("--output",
##                     default = "output/FISCALIA_datos_sijuf_cv.parquet")
## args <- parser$parse_args()
##
## # ------- function
##
## # ------- main
##
## data_SIJUF_CEV <- read_parquet(args$input_SIJUF_CEV)
##
## perp_is_ca = c("AUC - AUTODEFENSAS UNIDAS DE COLOMBIA",
##                 "AGENTES_ESTATALES","ELN - EJERCITO DE LIBERACION NACIONAL",
##                "EPL - EJERCITO POPULAR DE LIBERACION",
##                "ERP - EJERCITO REVOLUCIONARIO DEL PUEBLO",
##                "FARC - FUERZAS ARMADAS REVOLUCIONARIAS COLOMB",
##                "GUERRILLA", "OTROS GRUPOS PARAMILITARES", "PARAMILITARES")
##
## perp_no_ca = c("COMUNIDADES INDIGENAS", "DELINCUENCIA COMUN")
##
## data_SIJUF_CEV$fecha_hechos <- as.Date(data_SIJUF_CEV$fecha_hechos, "%d/%m/%Y")
##
##
## #IS CA
##
## casos_verde <- c("CASOS AUTO 92 CORTE CONSTITUCIONAL", "CASOS DEMANDADOS CIDH",
##                  "CASOS EN EL SIDH", "CASOS OIT",
##                  "CASOS SAN JOSE DE APARTADO", "CASOS UP",
##                  "COMPULSAS - JUSTICIA Y PAZ", "CONFLICTO ARMADO",
##                  "ELN", "FARC-EP",
##                  "GENERO, NINEZ Y ADOLESCENCIA HOMICIDIO PRESENTANDO COMO BAJA POR FUERZA PUBLICA",
##                  "MARCHA PATRIOTICA", "MASACRES",
##                  "PARAPOLITICA", "PRESUNTOS HOMICIDIOS AGENTES DEL ESTADO",
##                  "RELACION DE CIVILES CON ACTORES DE CONFLICTO", "RESTITUCION DE TIERRAS",
##                  "VICTIMAS DE MINAS ANTIPERSONALES",
##                  "VICTIMAS MIEMBROS DE LAS ONG'S", "VICTIMAS",
##                  "VICTIMAS SINDICALISTAS",
##                  "VIOLENCIA SEXUAL CONFLICTO ARMADO",
##                  "VIXCTIMAS DEFENSORES DE DDHH")
```

```
##
## casos_amarillo <- c("AFRODESCENDIENTES", "BANDAS EMERGENTES", "VIOLENCIA DE GENERO")
##
## casos_azul <- c("VICTIMAS INDIGENAS", "VICTIMAS LGBT", "PERIODISTAS", "VINCULO FUNC. PUBLICOS CON GRU
##
## casos_rojo <- c("TRATA TRANSNACIONAL")
##
## perp_verde <- c("AGENTE DEL ESTADO", "AGUILAS NEGRAS", "AUC", "CORDILLERA ELN", "GAO", "GUERRILLA",
##                 "INDEPENDIENTES", "LOS RASTROJOS", "M19", "MACHETEROS", "PARAMILITARES",
##                 "ELN", "EPL", "ERP", "ERG", "FARC", "GAO")
## perp_azul <- c("FUERZA PUBLICA",  "AGENTE DEL ESTADO")
## perp_naranja <- c("INDEPENDIENTES")
##
## prof_verde <- c("ACTIVIDADES RELACIONADAS CON EL SINDICALISMO",
##                "ACTIVIDADES RELACIONADAS ORG. CIVICAS Y CAMPESINAS", "ACTIVISTA DERECHOS HUMANOS",
##                "ACTIVISTA SINDICAL", "ALCALDE", "ALMIRANTE (FUERZA PUBLICA)",
##                "DESMOVILIZADO", "DESPLAZADO",
##                "DIRIGENTE ORGANIZACION DERECHOS HUMANOS", "DIRIGENTE SINDICAL", "INFANTE DE MARINA",
##                "INTEGRANTE AUTODEFENSAS", "INTEGRANTE GUERRILLA",
##                "JUEZ PENAL CIRCUITO", "JUEZ PENAL MUNICIPAL",
##                "JUEZ PROMISCUO",  "LIDER ORGANIZACION CAMPESINA", "LIDER ORGANIZACION COMUNITARIA",
##                "LIDER ORGANIZACION INDIGENA", "LIDER ORGANIZACION POLITICA",  "MARINERO ARMADA NACIO
##                "MIEMBRO, AFILIADO, ACTIVISTA DE LA UNION PATRIOTICA", "SECRETARIO DE SINDICATO" )
##
## prof_azul <- c("AEROTECNICO SUBJEFE FUERZA AEREA", "AGENTE POLICIA NACIONAL",
##               "BRIGADIER GENERAL DEL EJERCITO NACIONAL", "BRIGADIER GENERAL INFANTERIA DE MARINA",
##               "BRIGADIER GENERAL POLICIA NACIONAL", "CABO PRIMERO DEL EJERCITO NACIONAL",
##               "CABO PRIMERO POLICIA NACIONAL", "CABO SEGUNDO DEL EJERCITO NACIONAL",
##               "CABO SEGUNDO POLICIA NACIONAL", "CABO TERCERO DEL EJERCITO NACIONAL",
##               "CAPITAN DE CORBETA ARMADA NACIONAL", "CAPITAN DE FRAGATA ARMADA NACIONAL",
##               "CAPITAN DE LA FUERZA AEREA", "CAPITAN DE LA MARINA",
##               "CAPITAN DE NAVIO ARMADA NACIONAL", "CAPITAN DEL EJERCITO NACIONAL",
##               "CAPITAN POLICIA NACIONAL", "CONCEJAL", "CONTRA ALMIRANTE ARMADA NACIONAL",
##               "CONTRALOR MUNICIPAL", "CORONEL DEL EJERCITO NACIONAL",
##               "CORONEL INFANTERIA DE MARINA", "CORONEL POLICIA NACIONAL",
##               "DEFENSOR DEL PUEBLO", "DEFENSOR DEL PUEBLO DELEGADO", "DIPUTADO",
##               "EMPLEADOS AREA DE FISCALIA", "EMPLEADOS DEL AREA DEL CTI DE LA FISCALIA",
##               "ENFERMERAS", "ENFERMERAS JEFES",  "ESTUDIANTES UNIVERSITARIOS", "EXALCALDE",
##               "FISCAL DELEG. TRIBUNAL SUP.", "FISCAL ESPECIALIZADO", "FISCAL LOCAL",
##               "FISCAL SECCIONAL", "GANADERO",
##               "GENERAL DE LA FUERZA AEREA", "GENERAL DE LA POLICIA NACIONAL",
##               "GENERAL DE LA REPUBLICA","GENERAL DE LA REPUBLICA (FUERZA PUBLICA)",
##               "GENERAL DEL EJERCITO NACIONAL", "GENERAL INFANTERIA DE MARINA",
##               "GOBERNADOR",  "GOBERNADOR DEPARTAMENTO",  "INDIGENA",  "INDIGENTE",
##               "INGENIEROS DE PETROLEOS",  "INTENDENTE POLICIA NACIONAL", "INVESTIGADOR JUDICIAL C.T.I
##               "INVESTIGADOR JUDICIAL D.A.S.", "INVESTIGADOR JUDICIAL POLICIA NACIONAL",
##               "MAGISTRADO CONSEJO DE ESTADO",  "MAGISTRADO CONSEJO SECC. JUDICATURA",
##               "MAGISTRADO CONSEJO SUP. JUDICATURA",  "MAGISTRADO TRIBUNAL ADMINISTRATIVO",
##               "MAGISTRADO TRIBUNAL SUPERIOR", "MAGISTRADOS DE TRIBUNALES", "MAYOR DE LA FUERZA AEREA
##               "MAYOR DEL EJERCITO NACIONAL", "MAYOR GENERAL DEL EJERCITO NACIONAL",
##               "MAYOR GENERAL POLICIA NACIONAL", "MAYOR INFANTERIA DE MARINA",
##               "MAYOR POLICIA NACIONAL",  "MIEMBROS CIVILES ARMADA NACIONAL",
##               "MIEMBROS CIVILES DE LA MARINA",  "MIEMBROS CIVILES EJERCITO NACIONAL",
##               "MIEMBROS CIVILES POLICIA NACIONAL",  "MIEMBROS FUERZAS MILITARES Y POLICIAS",
```

```
##                 "MIEMBROS RETIRADOS ARMADA NACIONAL", "MIEMBROS RETIRADOS EJERCITO NACIONAL",
##                 "MIEMBROS RETIRADOS FUERZA AEREA  MIEMBROS RETIRADOS FUERZAS MILITARES Y DE POLICIA",
##                 "MIEMBROS RETIRADOS POLICIA NACIONAL", "OCUPACIONES EXCLUSIVAS DE LAS FUERZAS MILITARES
##                 "OFICIALES DE CUBIERTA", "OFICIALES DE LAS FUERZAS MILITARES", "OFICIALES DE POLICIA",
##                 "OTROS MIEMBROS ARMADA NACIONAL", "OTROS MIEMBROS DE LA MARINA", "OTROS MIEMBROS EJERCI
##                 "OTROS MIEMBROS FUERZA AEREA", "OTROS MIEMBROS POLICIA NACIONAL",
##                 "PRESIDENTE DE SINDICATO", "PROCURADOR DELEGADO", "PROFESORES DE EDUCACION BASICA",
##                 "PROFESORES UNIVERSITARIOS", "PROMOTORES DE SALUD", "REINSERTADO",
##                 "REPRESENTANTE A LA CAMARA", "SACERDOTE", "SARGENTO MAYOR DEL EJERCITO NACIONAL",
##                 "SARGENTO MAYOR POLICIA NACIONAL", "SARGENTO PRIMERO DEL EJERCITO NACIONAL",
##                 "SARGENTO PRIMERO POLICIA NACIONAL", "SARGENTO SEGUNDO DEL EJERCITO NACIONAL",
##                 "SARGENTO SEGUNDO POLICIA NACIONAL", "SARGENTO VICE PRIMERO POLICIA NACIONAL",
##                 "SARGENTO VICEPRIMERO DEL EJERCITO NACIONAL", "SARGENTO VICEPRIMERO INFANTERIA DE AVIA
##                 "SENADOR DE LA REPUBLICA", "SOLDADO EJERCITO NACIONAL",  "SOLDADO FUERZA AEREA",
##                 "SUB COMISARIO POLICIA NACIONAL", "SUB INTENDENTE POLICIA NACIONAL",
##                 "SUB OFICIAL DE LA MARINA", "SUB OFICIAL JEFE ARMADA NACIONAL",
##                 "SUB OFICIAL PRIMERO ARMADA NACIONAL", "SUB OFICIAL SEGUNDO ARMADA NACIONAL",
##                 "SUB TENIENTE POLICIA NACIONAL", "SUBOFICIALES DE LA POLICIA",
##                 "SUBOFICIALES DE LAS FUERZAS MILITARES", "SUBTENIENTE DEL EJERCITO NACIONAL", "TECNICOS
##                 "TENIENTE CORONEL DE LA FUERZA AEREA", "TENIENTE CORONEL DEL EJERCITO NACIONAL",
##                 "TENIENTE CORONEL INFANTERIA DE MARINA", "TENIENTE CORONEL POLICIA NACIONAL",
##                 "TENIENTE DE CORBETA ARMADA NACIONAL", "TENIENTE DE FRAGATA ARMADA NACIONAL",
##                 "TENIENTE DE LA FUERZA AEREA", "TENIENTE DE LA MARINA",
##                 "TENIENTE DE NAVIO ARMADA NACIONAL", "TENIENTE DEL EJERCITO NACIONAL",
##                 "TENIENTE POLICIA NACIONAL", "TRABAJADORES EN SERVICIO SECIAL Y COMUNITARIO", "TRABAJA
##
## prof_amarillo <- c("AEROTECNICO JEFE FUERZA AEREA", "AEROTECNICO PRIMERO FUERZA AEREA",
##                     "AGRICULTOR", "AGRONOMOS Y ESPECIALISTAS AGRICOLAS", "ALBANIL", "ARTESANOS",
##                     "ASISTENTES ADMINISTRATIVOS", "ASISTENTES DE CONTABILIDAD",
##                     "ASISTENTES DE JUZGADOS, TRIBUNALES",
##                     "AUXILIARES DE ENFERMERIA", "AUXILIARES DE ODONTOLOGIA",
##                     "AUXILIARES DE TRIBUNALES, JUZGADOS",
##                     "BIOLOGOS, BOTANICOS, ZOOLOGOS Y RELACIONADOS",
##                     "BRIGADIER GENERAL DE LA FUERZA AEREA", "CAJEROS",
##                     "CAJEROS DE SERVICIOS FINANCIEROS", "CARNICEROS", "COMERCIANTE",
##                     "COMERCIANTE DE GANADO", "COMISARIOS, INSPECTORES DE POLICIA",
##                     "CONDUCTOR O AUXILIAR DE TRANSPORTE", "CONTRALOR DEPARTAMENTA",
##                     "COTERO", "DESEMPLEADO",
##                     "DIRECTORES Y ADMINISTRADORES DE EDUCACION BASICA Y DIRECTORES Y GERENTES GENERALES
##                     "DIRECTORES Y GERENTES GENERALES PRODUCCION DE BIENES", "DIRECTORES Y GERENTES GEN
##                     "EBANISTA", "INVESTIGADORES Y ANALISTAS DE POLITICA", "EDILES",
##                     "ELECTRICISTA", "EMBAJADOR",  "EMPLEADOS DE BANCA, SEGUROS Y OTROS SERVICIOS FINAN
##                     "ESTUDIANTE PRIMARIA", "ESTUDIANTES SECUNDARIA", "FARMACEUTICOS",
##                     "FOTOGRAFOS", "FUNCIONARIOS DE PROGRAMAS EXCLUSIVOS DE LA ADMINIS ",
##                     "GEOLOGOS, GEOQUIMICOS Y GEOFISICOS ", "GERENTES DE ARQUITECTURA Y CIENCIAS",
##                     "GUARDIANES DE PRISION", "INGENIERO DE VIAS Y TRANSPORTE", "INGENIERO SANITARIO Y 
##                     "INGENIEROS CIVILES", "INGENIEROS DE MINAS",  "INGENIEROS DE SISTEMAS",
##                     "INGENIEROS ELECTRICOS Y ELECTRONICOS", "INGENIEROS INDUSTRIALES",
##                     "INGENIEROS MECANICOS", "INGENIEROS METALURGICOS",  "NGENIEROS QUIMICOS",
##                     "INSPECTOR DE CONSTRUCCION",  "INSPECTORES DE SANIDAD, SEGURIDAD Y SALUD OCUPACION
##                     "INSTRUMENTADORES QUIRURGICOS",  "INVESTIGADORES Y CONSULTORES, DESARROLLO ECONOMI
##                     "JEFE DE MISION DIPLOMATICA",  "JEFES DE BIBLIOTECA, PUBLICACIONES Y EMPLEADOS DE"
##                     "JEFES DE OFICINA EN GENERAL",  "JEFES DE REGISTRO, DISTRIBUCION Y PROGRAMACION",
##                     "JUEZ CIVIL CIRCUITO", "JUEZ CIVIL MUNICIPAL",  "JUEZ DE FAMILIA", "JUEZ DE MENORES
```

```
##                     "JUEZ DE PAZ", "JUEZ LABORAL",  "MATEMATICOS, ESTADISTICOS Y ACTUARIOS",
##                     "MECANICO AUTOMOTRIZ",  "MEDICOS GENERALES", "MERCADERISTAS", "MEDICOS ESPECIALISTA
##                     "MIEMBROS DEL PODER EJECUTIVO LEGISLATIVO Y JUDICIAL", "MINERO",
##                     "MINISTRO DEL DESPACHO", "MOTOTAXISTA",  "NOTARIO PUBLICO", "OBRERO",
##                     "OCUPACIONES PROFESIONALES EN ORGANIZACION Y ADMINI", "ODONTOLOGOS",
##                     "ORIENTADORES EDUCATIVOS", "OTRAS ACTIVIDADES DE LA CONSTRUCCION", "OTRAS ACTIVIDAD
##                     "OTRAS OCUPACIONES RELIGIOSAS", "OTRO", "OTROS ARTISTAS", "OTROS ESTUDIANTES",
##                     "OTROS INVESTIGADORES JUDICIALES", "OTROS JUECES", "OTROS SUPERVISORES", "PANADEROS
##                     "PELUQUEROS, ESTILISTAS Y AFINES",  "PENSIONADO",  "PERSONAL DIRECTIVO DE LA ADMINI
##                     "PATRULLERO POLICIA NACIONAL",  "PERIODISTAS",  "PERSONERO",  "POLICIAS",  "PRESID
##                     "PROCURADOR DELEGADO", "PROFESORES DE EDUCACION BASICA", "POLITOLOGO",
##                     "PROFESORES DE EDUCACION MEDIA", "PROFESORES DE PREESCOLAR", "PSICOLOGOS",
##                     "RECICLADOR", "REGISTRADOR NACIONAL ESTADO CIVIL", "SECRETARIAS", "SOLDADOR",
##                     "SUPERVISORES DE VIGILANCIA", "TECNOLOGOS Y TECNICOS AGROPECUARIOS", "TERAPEUTAS OC
##                     "TRABAJADORA SEXUAL", "VICE ALMIRANTE ARMADA NACIONAL", "VIGILANTES Y GUARDIANES DE
##
## prof_naranja <- c("ABOGADO", "ADMINISTRADOR DE EMPRESAS", "ADMINISTRADORES DE COMERCIO AL POR MENOR"
##                     "ADMINISTRADORES DE EDUCACION SUPERIOR Y FORMACION",
##                     "ADMINISTRADORES DE INMUEBLES", "ADMINISTRADORES Y OPERADORES DE SISTEMAS",
##                     "ANALISTAS DE SISTEMAS", "ANUNCIADORES Y LOCUTORES",
##                     "ARBITROS", "ARQUITECTOS", "ASEADORES Y SERVICIO DOMESTICO",
##                     "ASISTENTES DE PERSONAL Y SELECCION", "ADMINISTRADOR DE EMPRESAS",
##                     "ADMINISTRADORES DE COMERCIO AL POR MENOR  ADMINISTRADORES DE EDUCACION SUPERIOR Y
##                     "ADMINISTRADORES DE INMUEBLES", "ADMINISTRADORES Y OPERADORES DE SISTEMAS",
##                     "ANALISTAS DE SISTEMAS", "ANUNCIADORES Y LOCUTORES", "ASEADORES Y SERVICIO DOMESTIC
##                     "ASISTENTES DE PERSONAL Y SELECCION", "AUXILIARES ADMINISTRATIVOS",
##                     "AUXILIARES DE ALMACEN Y BODEGA", "AUXILIARES DE CONTABILIDAD", "AUXILIARES DE OFIC
##                     "AUXILIARES DE PERSONAL Y NOMINA", "AUXILIARES DE SERVICIO A PASAJEROS",
##                     "AYUDANTES DE COCINA Y CAFETERIA", "CARTEROS Y MENSAJEROS", "CHEFS", "COCINERO",
##                     "COMERCIANTE DE ESMERALDA", "COMPRADORES", "DEPORTISTA",
##                     "DIRECTOR DEPARTAMENTO ADMINISTRATIVO",
##                     "DIRECTORES DE FISCALIAS DIRECTORES DE PROGRAMAS DE ESPARCIMIENTO Y DEPORT",
##                     "DIRECTORES Y GERENTES SERVICIOS FINANCIEROS",
##                     "DISENADORES DE TEATRO, MODA Y EXHIBICION Y OTROS C",
##                     "DISENADORES GRAFICOS Y DIBUJANTES ARTISTICOS", "DISENADORES INDUSTRIALES",
##                     "DOMESTICADORES Y TRABAJADORES DEL CUIDADO DE ANIMA",
##                     "ECONOMISTAS", "EMPLEADOS DE INFORMACION Y SERVICIO AL CLIENTE",
##                     "EMPLEADOS DE PUBLICACION Y AFINES", "EMPLEADOS DE VENTAS Y SERVICIOS DE AREOLINEAS
##                     "FISICOS Y ASTRONOMOS", "FISIOTERAPEUTAS",
##                     "GERENTES DE COMERCIO AL POR MENOR", "GERENTES DE EMPRESAS DE TELECOMUNICACIONES",
##                     "GERENTES DE MEDIOS DE COMUNICACION Y ARTES ESCENIC", "GERENTES DE OTROS SERVICIOS
##                     "GERENTES DE OTROS SERVICIOS A LAS EMPRESAS", "GERENTES DE OTROS SERVICIOS ADMINIST
##                     "GERENTES DE PROGRAMAS DE POLITICA DE DESARROLLO EC GERENTES DE SEGUROS, BIENES RA
##                     "GERENTES DE SERVICIOS DE COMERCIO EXTERIOR", "GERENTES DE VENTAS, MERCADEO Y PUBL
##                     "GERENTES FINANCIEROS", "GUIAS DE VIAJES Y TURISMO", "HIGIENISTAS DENTALES",
##                     "MPULSADORES Y DEMOSTRADORES", "LAVACARROS", "MECANOGRAFOS", "MESEROS", "MUSICOS Y
##                     "OFICIALES DE BOMBEROS", "ORNAMENTADOR", "OTROS INGENIEROS", "OTROS INSPECTORES",
##                     "OTROS INSTRUCTORES", "PARTICULAR", "PATRONISTAS PRODUCTOS TEXTILES, CUERO Y PIEL"
##                     "PILOTOS, INGENIEROS E INSTRUCTORES DE VUELO", "PINTORES, ESCULTORES Y OTROS ARTIST
##                     "PLOMERO", "PRACTICANTES DE LA MEDICINA ALTERNATIVA", "PROGRAMADORES DE SISTEMAS",
##                     "QUIMICOS", "RECEPCIONISTAS Y OPERADORES DE CONMUTADOR", "RECREACIONISTAS",
##                     "REPRESENTANTES DE VENTAS NO TECNICAS", "SUPERINTENDENTE DE SALUD", "SUPERINTENDENT
##                     "SUPERINTENDENTE Y GERENTE DE ESTABLECIMIENTOS PUBL", "SUPERVISORES DE VENTAS",
##                     "TESORERO", "TOPOGRAFOS", "VENDEDORES DE MOSTRADOR", "VENDEDORES, VENTAS TECNICAS"
```

```
##
## prof_rojo <- c("ACTIVIADES RELACIONADAS CON LA ZAPATERIA", "ACTIVIDADES DE LATONERIA Y PINTURA AUTOM
##                "ACTIVIDADES DE PESCA", "ACTIVIDADES RELACIONADAS CON EL HOGAR",
##                "ACTIVIDADES RELACIONADAS CON LA CARPINTERIA", "ACTIVIDADES RELACIONADAS CON LA TAUROM
##                "AGENTES DE ADUANA Y OTROS AGENTES", "BARMAN", "BOMBEROS",
##                "CONSEJEROS DE EMPLEO", "CONTADORES Y AUDITORES FINANCIEROS",
##                "CONTRALOR GENERAL DE LA REPUBLICA", "CONTROLADORES DE TRAFICO AEREO", "DIGITADORES",
##                "OPERADORES DE JUEGOS MECANICOS Y DE SALON", "ORGANIZADORES DE EVENTOS",
##                "OTRAS OCUPACIONES DE SERVICIOS PERSONALES", "OTRAS OCUPACIONES ELEMENTALES DE LAS VEN
##                "OTRAS OCUPACIONES ELEMENTALES DE LOS SERVICIOS OTRAS OCUPACIONES PROFESIONALES EN TER
##                "PRESIDENTE DE LA REPUBLICA",
##                "TECNICOS DENTALES",
##                "TECNICOS EN GRABACION DE AUDIO Y VIDEO", "TECNICOS EN TRANSMISION",
##                "TECNICOS OPTICOS",
##                "TECNICOS Y MECANICOS DE INSTRUMENTOS DE AERONAVEGA  TECNICOS Y MECANICOS DE INSTRUMEN
##                "TECNOLOGOS Y TECNICOS DE LABORATORIO MEDICO Y PATO",
##                "TECNOLOGOS Y TECNICOS EN ARQUITECTURA", "TECNOLOGOS Y TECNICOS EN CIENCIAS BIOLGICAS"
##                "TECNOLOGOS Y TECNICOS EN ELECTROCARDIOGRAFIA Y ELE", "TECNOLOGOS Y TECNICOS EN GEOLOG
##                "TECNOLOGOS Y TECNICOS EN INGENIERA CIVIL TECNOLOGOS Y TECNICOS EN INGENIERA ELECTRICA
##                "TECNOLOGOS Y TECNICOS EN QUIMICA APLICADA TECNOLOGOS Y TECNICOS FORESTALES Y DE RECUR
##                "TRABAJADORES DE ESTACION DE SERVICIO")
##
## one <- as.integer(1)
## zero <- as.integer(0)
##
## conflict_SIJUF_CEV <- data_SIJUF_CEV %>%
##   mutate(prob = runif(nrow(.))) %>%
##   mutate(is_conflict = case_when(
##     vinculacion == "PERPETRADOR" ~ zero,
##     perpetrador %in% perp_verde ~ one,
##     perpetrador %in% perp_azul & prob < 0.75 ~ one,
##     perpetrador %in% perp_naranja & prob < 0.75 ~ zero,
##     perpetrador %in% perp_is_ca & vinculacion == "VICTIMA" ~ one,
##     perpetrador1 %in% perp_is_ca &  vinculacion == "VICTIMA" ~ one,
##     perpetrador2 %in% perp_is_ca & vinculacion == "VICTIMA" ~ one,
##     perpetrador3 %in% perp_is_ca & vinculacion == "VICTIMA" ~ one,
##     perpetrador %in% perp_no_ca ~ zero,
##     perpetrador1 %in% perp_no_ca ~ zero,
##     perpetrador2 %in% perp_no_ca ~ zero,
##     perpetrador3 %in% perp_no_ca ~ zero,
##     profesion %in% prof_verde ~ one,
##     profesion %in% prof_azul & prob < 0.75 ~ one,
##     profesion %in% prof_amarillo & prob < 0.5 ~ one,
##     profesion %in% prof_naranja & prob < 0.75 ~ zero,
##     profesion %in% prof_rojo ~ zero,
##     calidad_profesion %in% prof_naranja & prob < 0.75 ~ zero,
##     categoria %in% casos_verde ~ one,
##     categoria %in% casos_azul & prob < 0.75 ~ one,
##     categoria %in% casos_amarillo & prob < 0.5 ~ one,
##     categoria %in% casos_rojo ~ zero,
##     vinculacion == "OTRA" ~ NA_integer_,
##     TRUE ~ NA_integer_)) %>%
##   select(-prob)
##
```

```
##
## table(conflict_SIJUF_CEV$is_conflict, useNA = "always")
##
## # before Folco's changes
## # 0      1       <NA>
## #70429   1444 112509
##
## # after Folco's changes
## # 0      1      <NA>
## # 72386  2669  109327
##
## nrows <- nrow(conflict_SIJUF_CEV)
##
## sample  <- conflict_SIJUF_CEV %>%
##   filter(is_conflict ==0) %>%
##   select(recordid, is_conflict)
##
## sample <- sample_frac(sample, 0.4)
##
## # write.table(sample, file = args$examples, sep = "|", quote = FALSE, row.names = FALSE)
## write_parquet(conflict_SIJUF_CEV, args$output)
##
##
## # ---------------
##
##
## # names(data_SIJUF_CEV)
## # nrow(data_SIJUF_CEV)
## #
## # table(data_SIJUF_CEV$grupo_delito, useNA = "always")
## # table(data_SIJUF_CEV$clasificacion_conflicto_armado, useNA = "always")
## #
## # #Filtrando base: contiene y no contiene el hecho de desaparición
## # desa <- data_SIJUF_CEV %>%
## #   filter(grupo_delito == "DESAPARICION FORZADA")
## #
## # no_desa <- data_SIJUF_CEV%>%
## #   filter(!grupo_delito == "DESAPARICION FORZADA")
## #
## # #recategorizando variable
## #
## # desaparicion <- desa %>%
## #   mutate(car_per_cat=case_when(
## #     str_detect(calidad_profesion, "AUTODEFENSAS") ~ "AUTODEFENSAS",
## #     str_detect(calidad_profesion, "COMERCIANTE") ~ "COMERCIANTE",
## #     str_detect(calidad_profesion,"HOGAR|AMA DE CASA|DOMESTICO")~ "HOGAR",
## #     str_detect(calidad_profesion, "AGRICULTOR|AGROPECUARIO|AGRO") ~ "AGRICULTOR",
## #     str_detect(calidad_profesion,"VIGILANCIA|SEGURIDAD") ~ "SEGURIDAD",
## #     str_detect(calidad_profesion, "DESMOVILIZADO|REINSERTADO|EXCOMBATIENTE") ~ "DESMOVILIZADO",
## #     str_detect(calidad_profesion, "AEROTECNICO|TECNICO ARMADA NACIONAL$|ARMADA|INFANTERIA|FUERZA A
## #     str_detect(calidad_profesion,"SARGENTO|CAPITAN|BRIGADIER|POLICIA|CABO|SOLDADO BACHILLER|TENIEN
## #     str_detect(calidad_profesion, "VENDEDOR|VENTAS") ~ "TRABAJO INDEPENDIENTE",
## #     calidad_profesion == "ALBANIL" ~ "TRABAJO INDEPENDIENTE",
## #     calidad_profesion == "CONSTRUCCION" ~ "TRABAJO INDEPENDIENTE",
```

```
## #     str_detect(calidad_profesion, "AFRODESCENDIENTE|INDIGENA") ~ "GRUPOS ETNICOS",
## #     calidad_profesion == "MINERO" ~ "MINERO",
## #     str_detect(calidad_profesion, "EBANISTA|CARNICERO|GANADERO|OBRERO|CONSTRUCCION|VENDEDOR|VENTAS
## #     str_detect(calidad_profesion, "CAMPESINAS|JORNALERO|PESCADOR|CAMPO|CAMPESINO|PESCA") ~ "CAMPESI
## #     str_detect(calidad_profesion,"DESPLAZADO")~ "VICTIMA CONFLICTO",
## #     str_detect(calidad_profesion, "JUDICIAL|JUEZ|FISCAL|MAGISTRADO|PODER EJECUTIVO") ~ "FUNCIONARI
## #     str_detect(calidad_profesion, "FUNCIONARIO|CONTRALOR|PROCURADOR|PERSONERO|DIRECTOR DE CARCEL|NO
## #     str_detect(calidad_profesion, "EDIL|UP|POLITICO|MARCHA PATRIOTICA|CANDIDATO|DIPUTADO|CONCEJAL|A
## #     str_detect(calidad_profesion, "ESTUDIANTE") ~ "ESTUDIANTE",
## #     str_detect(calidad_profesion,"DERECHO|TRIBUNAL|JUZGADO")~ "SECTOR JUDICIAL",
## #     str_detect(calidad_profesion,"INDIGENTE") ~ "HABITANTE DE CALLE",
## #     str_detect(calidad_profesion,"TRABAJADORA SEXUAL") ~ "TRABAJO SEXUAL",
## #     str_detect(calidad_profesion, "EDUCATIVO|PROFESOR|EDUCADOR|EDUCACION|DOCENTES") ~ "EDUCACION",
## #     str_detect(calidad_profesion, "SINDICAL|SINDICATO")~ "SINDICALISTA",
## #     str_detect(calidad_profesion, "COMUNIDAD LGBT") ~ "COMUNIDAD LGBT",
## #     str_detect(calidad_profesion, "CONDUCTOR|TRANSPORTE|TRANSITO") ~ "TRANSPORTE",
## #     str_detect(calidad_profesion, "MIEMBRO GUERRILLA|GUERRILLERO|FARC|ELN|GUERRILLA") ~ "GUERRILLE
## #     str_detect(calidad_profesion, "BIBLIOTECA|ADMINISTRADOR DE EMPRESAS") ~ "PROFESIONALES",
## #     str_detect(calidad_profesion,"MEDICO") ~ "SALUD",
## #     calidad_profesion == "OTRO" ~ NA_character_,
## #     calidad_profesion ==  "OTRAS OCUPACIONES ELEMENTALES DE LOS SERVICIOS" ~ NA_character_,
## #     calidad_profesion == "OTRAS ACTIVIDADES U OFICIOS"~ NA_character_,
## #     calidad_profesion == "PARTICULAR" ~ NA_character_,
## #     calidad_profesion =="EMPLEADOS DE PUBLICACION Y AFINES" ~ NA_character_,
## #     TRUE  ~ calidad_profesion))
##
## #done!
```

```
cat(readLines(files$spoa_cv), sep = "\n")
```

9.4.1.4 FGN - SPOA - CEV

```
## #
## # Authors:     Maria Ortiz, Valentina Gómez, Paula Amado, PB
## # Maintainers  Maria Ortiz
## # Copyright   2022, HRDAG,
## # =========================
## # CO-SIVJRNR-data/individual/FGN/is-ca/src/is-ca-FISCALIA_datos_spoa_cv.R
##
## # ------ setup
##
## Sys.setlocale("LC_CTYPE", "en_US.UTF-8")
## set.seed(19481210)
##
## require(pacman)
## p_load(argparse, dplyr, here, arrow, assertr, stringr, lubridate)
##
## stopifnot(endsWith(getwd(),"is-ca"))
##
## parser <- ArgumentParser()
## parser$add_argument("--input_SPOA_CEV",
```

```
##                     default=here::here("individual/FGN/clean/output/FISCALIA_datos_spoa_cv.parquet"))
## parser$add_argument("--examples",
##                     default = "individual/FGN/is-ca/hand/example_SPOA.csv")
## parser$add_argument("--output",
##                     default = "individual/FGN/is-ca/output/FISCALIA_datos_spoa_cv.parquet")
## args <- parser$parse_args()
##
## # ------- function
##
## # ------- main
##
## data_SPOA_CEV <- read_parquet(args$input_SPOA_CEV)
##
##
## # #IS -CA
##
## viol_is_ca = c("DESAPARICION FORZADA", "DESPLAZAMIENTO", "HOMICIDIO DOLOSO",
##                "RECLUTAMIENTO ILICITO")
##
## perp_is_ca = c("AUC - AUTODEFENSAS UNIDAS DE COLOMBIA",
##                "AGENTES_ESTATALES","ELN - EJERCITO DE LIBERACION NACIONAL",
##                "FARC - FUERZAS ARMADAS REVOLUCIONARIAS DE COLOMBIA",
##                "FUERZA PUBLICA - DELITOS", "OTRAS GUERRILLAS")
##
## cat_caso_is_ca = c("AUTO 092 (LIGA DE MUJERES DESPLAZADAS", "CASOS UP",
##                    "CONFLICTO ARMADO", "CRIMEN DE GUERRA",
##                    "ELN", "FARC-EP", "HOMICIDIO PRESENTADO COMO BAJA POR FUERZA PUBLICA",
##                    "HOMICIDIO PRESENTANDO COMO BAJA POR FUERZA PUBLICA",
##                    "LESA HUMANIDAD", "PARAPOLITICA", "POLICIA NACIONAL VINCULOS GRUPOS ILEGALES",
##                    "PRESUNTOS HOMICIDIOS AGENTES DEL ESTADO", "UNION PATRIOTICA",
##                    "MILITARES VINCULOS GRUPOS ILEGALES","RECLUTAMIENTO DE MENORES",
##                    "VINCULO FUNC. PUBLICOS CON GRUPOS ILEGALES",
##                    "RELACION DE CIVILES CON ACTORES DE CONFLICTO",
##                    "VINCULOS F.P CON GRUPOS AL MARGEN DE LA LEY", "VIOLACION DD.HH",
##                    "VIOLENCIA SEXUAL CONFLICTO ARMADO", "VIXCTIMAS DEFENSORES DE DDHH")
##
## # is -ca colores
##
##
## perp_verde <- c("AUC - AUTODEFENSAS UNIDAS DE COLOMBIA",
##                        "AUTODEFENSAS",
##                        "AUTODEFENSAS UNIDAS DE COLOMBIA",
##                        "BANDAS EMERGENTES",
##                        "EJERCITO DE LIBERACION NACIONAL",
##                        "ELN - EJERCITO DE LIBERACION NACIONAL",
##                        "EPL - EJERCITO POPULAR DE LIBERACION",
##                        "ERG - EJERCITO REVOLUCIONARIO GUEVARISTA",
##                        "ERP - EJERCITO REVOLUCIONARIO DEL PUEBLO",
##                        "FARC",
##                        "FARC - FUERZAS ARMADAS REVOLUCIONARIAS DE COLOMBIA",
##                        "FARC-EP",
##                        "FARC - FUERZAS ARMADAS REVOLUCIONARIAS COLOMB",
##                        "FARC.",
##                        "FUERZA PUBLICA",
```

```
##                          "MJBC - MOVIMIENTO JAIME BATEMAN CAYON",
##                          "OTROS GRUPOS PARAMILITARES",
##                          "AGENTES_ESTATALES",
##                          "AUC - AUTODEFENSAS UNIDAS DE COLOMBIA",
##                          "AUTODEFENSAS",
##                          "AUTODEFENSAS UNIDAS DE COLOMBIA",
##                          "BANDAS EMERGENTES",
##                          "CAP - COMANDO ARMADO POPULAR",
##                          "EJERCITO DE LIBERACION NACIONAL",
##                          "ELN - EJERCITO DE LIBERACION NACIONAL",
##                          "EPL - EJERCITO POPULAR DE LIBERACION",
##                          "ERP - EJERCITO REVOLUCIONARIO DEL PUEBLO",
##                          "FARC",
##                          "FARC - FUERZAS ARMADAS REVOLUCIONARIAS COLOMB",
##                          "FARC-EP",
##                          "FARC.",
##                          "FP - DELITOS",
##                          "FUERZA PUBLICA",
##                          "GUERRILLA",
##                          "MPLICADO FUNCIONARIO PUBLICO",
##                          "MOVIMIENTO 26 DE ABRIL CARLOS PIZARRO LEON",
##                          "OTROS GRUPOS PARAMILITARES",
##                          "PARAMILITARES",
##                          "AGENTES_ESTATALES",
##                          "AUC - AUTODEFENSAS UNIDAS DE COLOMBIA",
##                          "AUTODEFENSAS",
##                          "BANDAS EMERGENTES",
##                          "EJERCITO DE LIBERACION NACIONAL",
##                          "ELN - EJERCITO DE LIBERACION NACIONAL",
##                          "EPL - EJERCITO POPULAR DE LIBERACION",
##                          "ERG - EJERCITO REVOLUCIONARIO GUEVARISTA",
##                          "ERP - EJERCITO REVOLUCIONARIO DEL PUEBLO",
##                          "FARC - FUERZAS ARMADAS REVOLUCIONARIAS COLOMB",
##                          "MOVIMIENTOS POLITICOS",
##                          "PARAMILITARES")
##
## perp_rojo <- c("DELINCUENCIA COMUN","COMUNIDADES INDIGENAS")
##
## perp_amarillo <- c("CARTEL DEL NORTE DEL VALLE","NARCOTRAFICO",
##                      "CMR - COMANDO MILICIANO REVOLUCIONARIO",
##                      "FP - DELITOS",
##                      "IMPLICADO FUNCIONARIO PUBLICO")
##
## perp_naranja <- c("MOVIMIENTOS POLITICOS")
##
## #a35_titulo
## titulo_rojo <- c("DE LA PROTECCION DE LA INFORMACION Y DE LOS DATOS",
##                    "DE LOS DELITOS CONTRA LA SALUD PUBLICA",
##                    "DE LOS DELITOS CONTRA LOS ANIMALES",
##                    "DE LOS DELITOS CONTRA LOS DERECHOS DE AUTOR",
##                    "DECRETOS",
##                    "DELITOS CONTRA EL ORDEN ECONOMICO SOCIAL",
##                    "DELITOS CONTRA EL ORDEN ECONOMICO Y SOCIAL",
##                    "DELITOS CONTRA LA ADMINISTRACION DE PUBLICA",
```

```
##                  "DELITOS CONTRA LA EFICAZ Y RECTA IMPARTICION DE JUSTICIA",
##                  "DELITOS CONTRA LA FAMILIA",
##                  "DELITOS CONTRA LA FE PUBLICA",
##                  "DELITOS CONTRA LA INTEGRIDAD MORAL",
##                  "DELITOS CONTRA LA SALUD PUBLICA")
##
## titulo_verde <- c("DE LOS DELITOS CONTRA EL REGIMEN CONSTITUCIONAL Y LEGAL",
##                   "DELITOS CONTRA LAS PERSONAS Y BIENES PROTEGIDOS POR EL DEREC",
##                   "DELITOS CONTRA PERSONAS Y BIENES PROTEGIDOS POR EL D.I.H.")
##
## titulo_amarillo <- c("DE LOS DELITOS CONTRA LOS RECURSOS NATURALES Y EL MEDIO AMBIENTE",
##                      "DELITOS CONTRA LA LIBERTAD INDIVIDUAL Y OTRAS GARANTIAS",
##                      "DELITOS CONTRA LA LIBERTAD, INTEGRIDAD Y FORMACION SEXUALES",
##                      "DELITOS CONTRA LA VIDA Y LA INTEGRIDAD PERSONAL",
##                      "LEYES",
##                      "SIN DELITO")
##
## titulo_naranja <- c("DELITOS CONTRA EL PATRIMONIO ECONOMICO",
##                     "DELITOS CONTRA LA EXISTENCIA Y SEGURIDAD DEL ESTADO",
##                     "DELITOS CONTRA LA SEGURIDAD PUBLICA",
##                     "DELITOS CONTRA LOS RECURSOS NATURALES Y EL MEDIO AMBIENTE",
##                     "DELITOS CONTRA MECANISMOS DE PARTICIPACION DEMOCRATICA")
##
## #tipo_hecho
## delito_rojo <- c("ABANDONO ART 128 C.P.",
##                  "ABANDONO ART. 127 C.P.",
##                  "ABANDONO DE HIJO FRUTO DE ACCESO CARNAL VIOLENTO ABUSIVO O DE INSEMINACION ARTIFIC
##                  "ABUSO DE CONFIANZA CALIFICADO. ART. 250 C.P.",
##                  "ABUSO DE CONFIANZA. ART. 249 C.P.",
##                  "ACAPARAMIENTO. ART. 297 C.P.",
##                  "ADOPCION IRREGULAR. ART. 232",
##                  "AGIOTAJE. ART. 301",
##                  "APLICACION FRAUDULENTA DE CREDITO OFICIALMENTE REGULADO ART. 311 C.P.",
##                  "APROVECHAMIENTO DE ERROR AJENO O CASO FORTUITO. ART. 252 C.P.",
##                  "ASESORAMIENTO Y OTRAS ACTUACIONES ILEGALES ART. 421 C.P.",
##                  "ASESORAMIENTO Y OTRAS ACTUACIONES ILEGALES ART. 421 C.P. INCISO 1",
##                  "ASOCIACION PARA LA COMISION DE UN DELITO CONTRA LA ADMINISTRACION PUBLICA ART. 434
##                  "CALUMNIA. ART. 221 C.P.",
##                  "CAPTACION MASIVA Y HABITUAL DE DINEROS ART. 316 C.P.",
##                  "CAZA ILEGAL ART. 336",
##                  "CELEBRACION INDEBIDA DE CONTRATO DE SEGURO ART. 172 C.P.",
##                  "CIRCULACION Y USO DE EFECTO OFICIAL O SELLO FALSIFICADO. ART. 281 C.P.",
##                  "CIRCUNSTANCIA DE AGRAVACION PUNIBLE ART 377B CP LEY 1453 DE 2011",
##                  "COHECHO IMPROPIO ART. 406 C.P.",
##                  "COHECHO POR DAR U OFRECER ART. 407 C.P.",
##                  "COHECHO PROPIO ART. 405 C.P.",
##                  "CONTRABANDO ART. 319 C.P.",
##                  "CONTRATO SIN CUMPLIMIENTO DE REQUISITOS LEGALES ART. 410 C.P.",
##                  "CORRUPCION DE ALIMENTOS, PRODUCTOS MEDICOS O MATERIAL PROFILACTICO ART. 372 C.P.",
##                  "CORRUPCION PRIVADA ART. 250A C.P.",
##                  "DAÑO EN MATERIA PRIMA, PRODUCTO AGROPECUARIO O INDUSTRIAL. ART. 304 C.P.",
##                  "DAÑO INFORMATICO ART 269D LEY 1273 DE 2009",
##                  "DE LA PRESTACION, ACCESO O USO ILEGALES DE LOS SERVICIOS DE TELECOMUNICACIONES ART
##                  "DEFRAUDACION A LOS DERECHOS PATRIMONIALES DE AUTOR. ART. 271 C.P.",
```

```
##                    "DEFRAUDACION DE FLUIDOS. ART. 256 C.P.",
##                    "DENEGACION DE INSCRIPCION ART. 396 C.P.",
##                    "DESTRUCCION, SUPRESION U OCULTAMIENTO DE DOCUMENTO PUBLICO. ART. 292 C.P.",
##                    "DESTRUCCION, SUPRESION Y OCULTAMIENTO DE DOCUMENTO PRIVADO. ART. 293 C.P.",
##                    "EJERCICIO ARBITRARIO DE LA CUSTODIA DE HIJO MENOR DE EDAD ART. 230A C.P. AD. LEY 8
##                    "EJERCICIO ILICITO DE ACTIVIDAD MONOPOLISTICA DE ARBITRIO RENTISTICO ART. 312 C.P."
##                    "EMISION Y TRANSFERENCIA ILEGAL DE CHEQUE. ART. 248 C.P.",
##                    "EMISIONES ILEGALES. ART. 276 C.P.",
##                    "ENRIQUECIMIENTO ILICITO ART. 412 C.P.",
##                    "ENRIQUECIMIENTO ILICITO DE PARTICULARES ART. 327 C.P.",
##                    "ENRIQUECIMIENTO ILICITO DE PARTICULARES ART. 327 C.P. INFERIOR 100 SMLM",
##                    "ESTAFA. ART. 246 C.P.",
##                    "ESTIMULO AL USO ILICITO ART. 378 C.P.",
##                    "EXPORTACION O IMPORTACION FICTICIA ART. 310 C.P.",
##                    "FABRICACION Y COMERCIALIZACION DE SUSTANCIAS NOCIVAS PARA LA SALUD ART. 374 C.P.",
##                    "FALSA AUTOACUSACION ART. 437 C.P.",
##                    "FALSA DENUNCIA ART. 435 C.P.",
##                    "FALSA DENUNCIA CONTRA PERSONA DETERMINADA ART. 436 C.P.",
##                    "FALSEDAD EN DOCUMENTO PRIVADO. ART. 289 C.P.",
##                    "FALSEDAD IDEOLOGICA EN DOCUMENTO PUBLICO. ART. 286 C.P.",
##                    "FALSEDAD MARCARIA. ART. 285 C.P.",
##                    "FALSEDAD MARCARIA. ART. 285 C.P. INCISO 1",
##                    "FALSEDAD MATERIAL EN DOCUMENTO PUBLICO. ART. 287 C.P.",
##                    "FALSEDAD PARA OBTENER PRUEBA DE HECHO VERDADERO. ART. 295 C.P.",
##                    "FALSEDAD PERSONAL. ART. 296 C.P.",
##                    "FALSIFICACION DE EFECTO OFICIAL TIMBRADO. ART. 280 C.P.",
##                    "FALSIFICACION DE MONEDA NACIONAL O EXTRANJERA. ART. 273 C.P.",
##                    "FALSIFICACION O USO FRAUDULENTO DE SELLO OFICIAL. ART. 279 C.P.",
##                    "FALSO TESTIMONIO ART. 442 C.P.",
##                    "FAVORECIMIENTO ART. 446 C.P.",
##                    "FAVORECIMIENTO DE CONTRABANDO ART. 320 C.P.",
##                    "FECUNDACION Y TRAFICO DE EMBRIONES ART. 134 C.P.",
##                    "FEMINICIDIO ART. 104A C.P.",
##                    "FRAUDE EN LA INSCRIPCION DE CEDULAS ART. 389 C.P.",
##                    "FRAUDE PROCESAL ART. 453 C.P.",
##                    "FUGA DE PRESOS ART. 448 C.P.",
##                    "GESTION INDEBIDA DE LOS RECURSOS SOCIALES. ART. 260 C.P.",
##                    "HOMICIDIO CULPOSO ART. 109 C.P.",
##                    "HOMICIDIO POR PIEDAD ART. 106 C.P.",
##                    "HURTO CALIFICADO. ART. 240 C.P.",
##                    "HURTO POR MEDIOS INFORMATICOS Y SEMEJANTES ART. 269I LEY 1273 DE 2009",
##                    "HURTO. ART. 239 C.P.",
##                    "ILICITA EXPLOTACION COMERCIAL ART. 303",
##                    "ILICITO APROVECHAMIENTO DE LOS RECURSOS NATURALES RENOVABLES ART. 328 C.P.",
##                    "IMITACION O SIMULACION DE ALIMENTOS, PRODUCTOS O SUSTANCIAS ART. 373 C.P.",
##                    "IMPEDIMENTO O PERTURBACION DE LA CELBRACIÓN DE AUDIENCIAS PUBLICAS ART. 454C C.P. A
##                    "INASISTENCIA ALIMENTARIA ART. 233 C.P.",
##                    "INCENDIO ART. 350 C.P.",
##                    "INCESTO. ART. 237 C.P.",
##                    "INDUCCION A AYUDA AL SUICIDIO ART. 107 C.P.",
##                    "INFIDELIDAD A LOS DEBERES PROFESIONALES ART. 445 C.P.",
##                    "INJURIA. ART. 220 C.P.",
##                    "INTERES INDEBIDO EN LA CELEBRACION DE CONTRATOS ART. 409 C.P.",
##                    "INTERVENCION EN POLITICA ART. 422 C.P.",
```

```
##                    "INVASION DE AREAS DE ESPECIAL IMPORTANCIA ECOLOGICA ART. 337 C.P.",
##                    "INVASION DE TIERRAS O EDIFICACIONES. ART. 263 C.P.",
##                    "IRRESPETO A CADAVERES. ART. 204 C.P.",
##                    "LESIONES AL FETO ART. 125 C.P.",
##                    "LESIONES CULPOSAS AL FETO ART. 126 C.P.",
##                    "LESIONES CULPOSAS ART. 120 C.P.",
##                    "LESIONES CULPOSAS ART. 120 C.P. CON PERDIDA ANATOMICA O FUNCIONAL DE UN ORGANO O MI
##                    "LESIONES CULPOSAS ART. 120 C.P. CON PERTURBACION FUNCIONAL PERMANENTE ART.114 INCIS
##                    "LESIONES CULPOSAS ART. 120 C.P. CON PERTURBACION FUNCIONAL TRANSITORIA ART.114 INCI
##                    "LESIONES CULPOSAS ART. 120 C.P. CON PERTURBACION PSIQUICA PERMANENTE ART.115",
##                    "LESIONES CULPOSAS ART. 120 C.P. CON PERTURBACION PSIQUICA TRANSITORIA ART.115",
##                    "LESIONES CULPOSAS ART. 120 C.P. INCISO 1",
##                    "LESIONES CULPOSAS ART. 120 C.P.AGRAVADO POR NO LICENCIA O LICENCIA SUSPENDIDA ART.
##                    "LESIONES PERSONALES CULPOSAS ART 120 C.P. CON INCAPACIDAD MAYOR 30 DIAS MENOR 90 DI
##                    "LESIONES PERSONALES CULPOSAS ART 120 C.P. CON INCAPACIDAD MAYOR 90 DIAS ART. 112 C
##                    "LESIONES PERSONALES CULPOSAS ART 120 C.P. CON INCAPACIDAD MENOR 30 DIAS ART.112 C.
##                    "LESIONES PERSONALES CULPOSAS ART. 120 C.P CON DEFORMIDAD FISICA AFECTA ROSTRO ART.
##                    "LESIONES PERSONALES CULPOSAS ART. 120 C.P CON DEFORMIDAD FISICA PERMANENTE ART. 113
##                    "LESIONES PERSONALES CULPOSAS ART. 120 C.P. CON DEFORMIDAD FISICA TRANSITORIA ART.
##                    "MALTRATO ANIMAL ART. 339A C.P.",
##                    "MALTRATO MEDIANTE RESTRICCION A AL LIBERTAD FISICA. ART. 230 C.P.",
##                    "MALTRATO POR DESCUIDO O ABANDONO EN PERSONA MAYOR DE 60 AÑOS",
##                    "MALVERSACION Y DILAPIDACION DE BIENES FAMILIARES. ART. 236 C.P.",
##                    "MANIPULACION FRAUDULENTA DE ESPECIES INSCRITAS EN EL REGISTRO NACIONAL DE VALORES
##                    "OBSTRUCCION DE OBRAS DE DEFENSA Y ASISTENCIA ART. 364 C.P.",
##                    "OBTENCION DE DOCUMENTO PUBLICO FALSO ART. 288 C.P.",
##                    "OCULTAMIENTO, ALTERACION O DESTRUCCION DE ELEMENTO MATERIAL PROBATORIO ART.454B C.
##                    "OCULTAMIENTO, RETENCION Y POSESION ILICITA DE CEDULA ART. 395 C.P.",
##                    "OFRECIMIENTO ENGAÑOSO DE PRODUCTOS Y SERVICIOS. ART. 300 C.P.",
##                    "OMISION DE DENUNCIA DE PARTICULAR ART. 441 C.P.",
##                    "OMISION DE DENUNCIA DE PARTICULAR ART. 441 C.P. MODIFICADO ART. 18 LEY 1121 DE 200
##                    "OMISION DEL AGENTE RETENEDOR O RECAUDADOR ART. 402 C.P.",
##                    "PANICO ECONOMICO ART. 302 C.P.",
##                    "PARTO O ABORTO PRETERINTENCIONAL ART. 118 C.P.",
##                    "PECULADO CULPOSO ART. 400 C.P.",
##                    "PECULADO POR APLICACION OFICIAL DIFERENTE ART. 399 C.P.",
##                    "PECULADO POR APROPIACION ART. 397 C.P.",
##                    "PECULADO POR USO ART. 398 C.P.",
##                    "PESCA ILEGAL ART. 335",
##                    "PORNOGRAFIA CON MENORES ART. 218 C.P.",
##                    "PORTE DE SUSTANCIAS ART. 383 C.P.",
##                    "PREVARICATO POR ACCION ART. 413 C.P.",
##                    "PREVARICATO POR OMISION ART. 414 C.P.",
##                    "PROPAGACION DE EPIDEMIA ART. 369 C.P.",
##                    "PROPAGACION DEL VIRUS DE INMUNODEFICIENCIA HUMANA O DE LA HEPATITIS B ART. 370 C.P
##                    "PROXENETISMO CON MENOR DE EDAD ART. 213A C.P.",
##                    "PROXENETISMO CON MENOR DE EDAD ART. 213A C.P. ADICIONADO LEY 1329 DE 2009",
##                    "RECEPTACION ART. 327 C C.P.",
##                    "RECEPTACION ART. 447 C.P.",
##                    "REPETIBILIDAD DEL SER HUMANO ART. 133 C.P.",
##                    "SABOTAJE. ART. 199 C.P.",
##                    "SIN DELITO",
##                    "SINIESTRO O DAÑO DE NAVE ART. 354 C.P.",
##                    "SOBORNO ART. 444 C.P.",
```

```
##                  "SOBORNO EN LA ACTUACION PENAL ART. 444A C.P. AD. LEY 890/2004 ART.10",
##                  "SUMINISTRO A MENOR ART. 381 C.P.",
##                  "SUPRESION, ALTERACION O SUPOSICION DEL ESTADO CIVIL. ART. 238 C.P.",
##                  " SUSTRACCION DE BIEN PROPIO. ART. 254 C.P.",
##                  "SUSTRACCION DE COSA PROPIA AL CUMPLIMIENTO DE DEBERES CONSTITUCIONALES O LEGALES A
##                  "TRAFICO DE MIGRANTES ART. 188 C.P. MOD. LEY 747/2002 ART.1",
##                  "TRAFICO DE MONEDA FALSIFICADA. ART. 274 C.P.",
##                  "TRAFICO DE NIÑAS, NIÑOS Y ADOLESCENTES ART 188C LEY 1453 DE 2011",
##                  "TRAFICO, ELABORACION Y TENENCIA DE ELEMENTOS DESTINADOS A LA FALSIFICACION DE MONE
##                  "TRANSFERENCIA NO CONSENTIDA DE ACTIVOS VALIENDOSE DE ALGUNA MANIPULACION INFORMATI
##                  "TRATA DE PERSONAS ART. 188A C.P.",
##                  "TRATA DE PERSONAS TRANSNACIONAL ART. 188A C.P.",
##                  "TURISMO SEXUAL. ART.219 C.P. MOD. POR ART 23 L, 1336 DE 2009",
##                  "URBANIZACION ILEGAL ART. 318 C.P.",
##                  "USO DE DOCUMENTO FALSO. ART. 291 C.P.",
##                  "USO DE SOFTWARE MALICIOSO ART 269E LEY 1273 DE 2009",
##                  "USO ILEGITIMO DE PATENTES ART. 307 C.P.",
##                  "USO ILEGITIMO DE PATENTES ART. 307 C.P. INCISO 1. EL QUE FABRIQUE PRODUCTO SIN AUT
##                  "USURA ART. 305 C.P.",
##                  "USURA ART. 305 C.P. INCISO 1 A CAMBIO DE PRESTAMO DE DINERO O POR CONCEPTO DE VENT
##                  "USURPACION DE AGUAS. ART. 262 C.P.",
##                  "USURPACION DE DERECHOS DE PROPIEDAD INDUSTRIAL Y DERECHOS DE OBTENTORES DE VARIEDA
##                  "UTILIZAC.O FACILITAC.MEDIOS DE COMUNICAC.PARA OFRECER ACTIV. SEXUALES CON MENORES 
##                  "VIOLACION A LA LIBERTAD RELIGIOSA. ART. 201 C.P.}VIOLACION A LOS DERECHOS MORALES 
##                  "VIOLACION A LOS DERECHOS MORALES DE AUTOR. ART. 270 C.P. N.1",
##                  "VIOLACION A LOS DERECHOS PATRIMONIALES DE AUTOR Y DERECHOS CONEXOS ART. 271 C.P. M
##                  "VIOLACION AL REGIMEN LEGAL O CONSTITUCIONAL DE INHABILIDADES E INCOMPATIBILIDADES 
##                  "VIOLACION DE HABITACION AJENA. ART. 189 C.P.",
##                  "VIOLACION DE LA RESERVA INDUSTRIAL O COMERCIAL ART. 308 C.P.",
##                  "VIOLENCIA INTRAFAMILIAR ART. 229 C.P.")
##
## delito_naranja <- c("ABUSO DE AUTORIDAD POR OMISION DE DENUNCIA ART. 417 C.P. INCISO 1",
##                     "ACCESO ABUSIVO A UN SISTEMA INFORMATICO ART 269A LEY 1273 DE 2009",
##                     "ACCESO ABUSIVO A UN SISTEMA INFORMATICO. ART. 195 C.P.",
##                     "ACOSO SEXUAL ART. 210A C.P. ADICIONADO LEY 1257 DE 2008",
##                     "ACTO SEXUAL VIOLENTO CON MENOR DE CATORCE AÑOS ART. 209 C.P.",
##                     "ACTO SEXUAL VIOLENTO. ART. 206 C.P.",
##                     "DEMANDA DE EXPLOT.SEX. COMERC. MENOR DE 18 AÑOS ADIC. ART 217A C.P. LEY 1329 DE
##                     "DISPARO DE ARMA DE FUEGO ART. 356A CP LEY 1453 DE 2011",
##                     "DISPARO DE ARMA DE FUEGO CONTRA VEHICULO ART. 356 C.P.",
##                     "EXPLOTACION DE MENORES DE EDAD",
##                     "FRAUDE A RESOLUCION JUDICIAL ART. 454 C.P.",
##                     "FRAUDE AL SUFRAGANTE ART. 388 C.P.",
##                     "HOMICIDIO ART. 103 C.P.",
##                     "INSTIGACION A DELINQUIR ART. 348 C.P.",
##                     "INSTIGACION A DELINQUIR ART. 348 C.P. INCISO 1",
##                     "INTERCEPTACION DE DATOS INFORMATICOS, ART 269C LEY 1273 DE 2009",
##                     "LESIONES ART. 111 C.P.",
##                     "LESIONES ART. 113 C.P.",
##                     "LESIONES CON PERDIDA ANATOMICA O FUNCIONAL DE UN ORGANO O MIEMBRO ART.116",
##                     "LESIONES CON PERTURBACION FUNCIONAL PERMANENTE ART.114 INCISO 2",
##                     "LESIONES CON PERTURBACION PSIQUICA PERMANENTE ART.115",
##                     "LESIONES CON PERTURBACION PSIQUICA TRANSITORIA ART.115",
##                     "LESIONES PERSONALES AGRAVADAS POR CIRCUNSTANCIAS ART 104 C.P. ART. 119 C.P.",
```

```
##                        "LESIONES PERSONALES ART 120 C.P. CON INCAPACIDAD MENOR 30 DIAS ART.112 C.P. INC
##                        "LESIONES PERSONALES CON DEFORMIDAD FISICA AFECTA ROSTRO ART. 113 C.P.",
##                        "LESIONES PERSONALES CON DEFORMIDAD FISICA PERMANENTE ART. 113 C.P.INCISO 2",
##                        "LESIONES PERSONALES CON INCAPACIDAD MAYOR 30 DIAS MENOR 90 DIAS ART. 112 C.P. I
##                        "OBSTACULIZACION ILEGITIMA DEL SISTEMA INFORMATICO O RED DE TELECOMUNICACION ART
##                        "OMISION DE APOYO ART. 424 C.P.",
##                        "OMISION DE SOCORRO ART. 131 C.P.",
##                        "PERTURBACION DE ACTOS OFICIALES ART. 430 C.P. INCISO 1",
##                        "PERTURBACION DE CERTAMEN DEMOCRATICO ART. 386 C.P.",
##                        "PERTURBACION DE LA POSESION SOBRE INMUEBLE. ART. 264 C.P.",
##                        "PERTURBACION EN SERVICIO DE TRANSPORTE COLECTIVO U OFICIAL ART. 353 C.P.",
##                        "SECUESTRO SIMPLE ART. 168 C.P.",
##                        "SECUESTRO SIMPLE ART. 168 C.P. AGRAVADO CUANDO SE TRAFIQUE CON LA PERSONA SECUE
##                        "SECUESTRO SIMPLE ART. 168 C.P. AGRAVADO SOBREVENIR MUERTE O LESIONES PERSONALES
##                        "SIMULACION DE INVESTIDURA O CARGO ART. 426 C.P.",
##                        "SUPLANTACION DE SITIOS WEB PARA CAPTURAR DATOS PERSONALES ART 269G LEY 1273 DE 
##                        "TENENCIA, FABRICACION Y TRAFICO DE SUSTANCIA U OBJETOS PELIGROSOS ART. 358 C.P.
##                        "TRAFICO DE INFLUENCIAS DE SERVIDOR PUBLICO ART. 411 C.P.",
##                        "TRAFICO DE SUSTANCIAS PARA PROCESAMIENTO DE NARCOTICOS ART. 382 C.P.",
##                        "TRAFICO FABRICACION O PORTE DE ESTUPEFACIENTES ART. 376 C.",
##                        "ULTRAJE A EMBLEMAS Y SIMBOLOS PATRIOS ART. 461 C.P.",
##                        "USURPACION DE FUNCIONES PUBLICAS ART. 425 C.P.",
##                        "UTILIZACION ILICITA DE EQUIPOS TRANSMISORES O RECEPTORES. ART. 197 C.P.",
##                        "UTILIZACION INDEBIDA DE INFLUENCIAS DERIVADAS DEL EJERCICIO DE FUNCION PUBLICA 
##                        "VIOLACION DE INMUNIDAD DIPLOMATICA ART. 465 C.P.",
##                        "VOTO FRAUDULENTO ART. 391 C.P")
##
##
## delito_verde <- c("ABORTO FORZADO EN PERSONA PROTEGIDA ART. 139E",
##                      "ACCESO CARNAL ABUSIVO EN PERSONA PROTEGIDA MENOR DE CATORCE AÑOS ART. 138",
##                      "ACCESO CARNAL VIOLENTO EN PERSONA PROTEGIDA ART. 138 C.P.",
##                      "ACTOS DE TERRORISMO ART. 144 C.P.",
##                      "ACTOS SEXUALES CON PERSONA PROTEGIDA MENOR DE CATORCE AÑOS ART. 139A",
##                      "ACTOS SEXUALES VIOLENTOS EN PERSONA PROTEGIDA ART. 139 C.P.",
##                      "ADIC.L.579/2002 ART.2.EMPLEO,PRODUCCION,COMERCIALIZACION Y ALMACENAMIENTO DE MINA
##                      "ADIC.L.579/2002 ART.3.AYUDA E INDUCCION AL EMPLEO, PRODUCCION Y TRANSFERENCIA DE 
##                      "ADMINISTRACION DE RECURSOS RELACIONADOS CON ACTIVIDADES TERRORISTAS ART. 345 C.P.
##                      "AMENAZAS CONTRA DEFENSORES DE DERECHOS HUMANOS Y SERVIDORES PUBLICOS ART. 188E",
##                      "APOLOGIA DEL GENOCIDIO ART. 102 C.P.",
##                      "ASONADA ART. 469 C.P.",
##                      "ATAQUE CONTRA OBRAS E INSTALACIONES QUE CONTIENEN FUERZAS PELIGROSAS ART. 157 C.P
##                      "ATENTADOS A LA SUBSISTENCIA Y DEVASTACION ART. 160 C.P.",
##                      "CONSPIRACION ART. 471 C.P.",
##                      "CONSTREÑIMIENTO A APOYO BELICO ART. 150 C.P.",
##                      "DEPORTACION, EXPULSION, TRASLADO O DESPLAZAMIENTO FORZADO DE POBLACION CIVIL ART.
##                      "DESCONOCIMIENTO DE HABEAS CORPUS ART. 177 C.P.",
##                      "DESNUDEZ FORZADA EN PERSONA PROTEGIDA ART. 139D",
##                      "DESPLAZAMIENTO FORZADO ART. 180 C.P.",
##                      "DESPOJO EN EL CAMPO DE BATALLA ART. 151 C.P.",
##                      "DESTRUCCION DE BIENES E INSTALACIONES DE CARACTER SANITARIO ART. 155 C.P.",
##                      "DESTRUCCION DEL MEDIO AMBIENTE ART. 164 C.P.",
##                      "DESTRUCCION O UTILIZACION ILICITA DE BIENES CULTURALES Y DE LUGARES DE CULTO ART.
##                      "DESTRUCCION Y APROPIACION DE BIENES PROTEGIDOS ART. 154 C.P.",
##                      "DETENCION ARBITRARIA ESPECIAL ART. 176 C.P.",
```

```
##                    "DETENCION ILEGAL Y PRIVACION DEL DEBIDO PROCESO ART. 149 C.P.",
##                    "EMBARAZO FORZADO EN PERSONA PROTEGIDA ART. 139C",
##                    "ESCLAVITUD SEXUAL EN PERSONA PROTEGIDA ART. 141A",
##                    "EXACCION O CONTRIBUCIONES ARBITRARIAS ART. 163 C.P.",
##                    "EXCLAVITUD SEXUAL EN PERSONA PROTEGIDA ART. 141A",
##                    "GENOCIDIO ART. 101 C.P.",
##                    "HOMICIDIO EN PERSONA PROTEGIDA ART. 135 C.P.",
##                    "HOSTILIDAD MILITAR ART. 456 C.P.",
##                    "LESIONES EN PERSONA PROTEGIDA ART. 136 C.P.",
##                    "OBSTACULIZACION DE TAREAS SANITARIAS Y HUMANITARIAS ART. 153 C.P.",
##                    "OMISION DE MEDIDAS DE PROTECCION A LA POBLACION CIVIL ART. 161 C.P.",
##                    "OMISION DE MEDIDAS DE SOCORRO Y ASISTENCIA HUMANITARIAS ART. 152 C.P.",
##                    "PERFIDIA ART. 143 C.P.",
##                    "PRIVACION ILEGAL DE LA LIBERTAD ART. 174 C.P.",
##                    "PROLONGACION ILICITA DE LA PRIVACION DE LA LIBERTAD ART. 175 C.P.",
##                    "PROSTITUCION FORZADA EN PERSONA PROTEGIDA ART. 141",
##                    "PROSTITUCION FORZADA O ESCLAVITUD SEXUAL ART. 141 C.P.",
##                    "REBELION ART. 467 C.P.",
##                    "RECLUTAMIENTO ILICITO ART. 162 C.P.",
##                    "REPRESALIAS ART. 158 C.P.",
##                    "SEDICION ART. 468 C.P.",
##                    "SEDUCCION, USURPACION Y RETENCION ILEGAL DEL MANDO ART. 472 C.P.",
##                    "TOMA DE REHENES ART. 148 C.P.",
##                    "TORTURA CONTRA PERSONA PROTEGIDA ART. 137 C.P.",
##                    "TRATOS INHUMANOS Y DEGRADANTES Y EXPERIMENTOS BIOLOGICOS EN PERSONA PROTEGIDA ART
##                    "UTILIZACION DE MEDIOS Y METODOS DE GUERRA ILICITOS ART. 142 C.P.",
##                    "UTILIZACION ILEGAL DE UNIFORMES E INSIGNIAS ART. 346 C.P.")
##
## delito_amarillo <- c("ABORTO ART. 122 C.P.",
##                       "DAÑO EN BIEN AJENO. ART. 265 C.P.",
##                       "ABORTO SIN CONSENTIMIENTO ART. 123 C.P.",
##                       "ABUSO DE AUTORIDAD POR OMISION DE DENUNCIA ART. 417 C.P.",
##                       "ACCESO CARNAL ABUSIVO CON MENOR DE CATORCE AÑOS. ART. 208 C.P.",
##                       "ACCESO CARNAL O ACTO SEXUAL ABUSIVOS CON INCAPAZ DE RESISTIR. ART. 210 C.P.",
##                       "ACCESO CARNAL ABUSIVO CON MENOR DE CATORCE ANOS. ART. 208 C.P.",
##                       "ABUSO DE AUTORIDAD POR ACTO ARBITRARIO O INJUSTO ART. 416",
##                       "ABUSO DE CONDICIONES DE INFERIORIDAD. ART. 251 C.P.",
##                       "ACCESO CARNAL O ACTO SEXUAL EN PERSONA PUESTA EN INCAPACIDAD DE RESISTIR ART. 
##                       "ACCESO CARNAL O ACTO SEXUAL EN PERSONA PUESTA EN INCAPACIDAD DE RESISTIR ART. 
##                       "ACCESO CARNAL VIOLENTO. ART. 205 C.P.",
##                       "ACTOS SEXUALES CON MENOR DE CATORCE AÑOS. ART. 209 C.P.",
##                       "ACTOS SEXUALES CON MENOR DE CATORCE ANOS. ART. 209 C.P.",
##                       "ALTERACION DE RESULTADOS ELECTORALES ART. 394 C.P.",
##                       "ALTERACION DESFIGURACION Y SUPLANTACION DE MARCAS DE GANADO. ART. 243 C.P.",
##                       "ALZAMIENTO DE BIENES. ART. 253 C.P.",
##                       "CONCUSION ART. 404 C.P.",
##                       "CONSERVACION O FINANCIACION DE PLANTACIONES ART. 375 C.P.",
##                       "CONTRABANDO DE HIDROCARBUROS Y SUS DERIVADOS ART. 319-1 C.P.",
##                       "ABUSO DE FUNCION PUBLICA ART. 428 C.P.",
##                       "ACTOS DE DISCRIMINACION RACIAL ART. 147 C.P.",
##                       "ACTOS DE RACISMO O DISCRIMINACIÓNART. 134 A",
##                       "CONSTREÑIMIENTO A LA PROSTITUCION ART. 214 C.P.",
##                       "CONSTREÑIMIENTO AL SUFRAGANTE ART. 387 C.P.",
##                       "CONSTREÑIMIENTO ILEGAL ART. 182 C.P.",
```

```
##                          "CONTAMINACION AMBIENTAL ART. 332 C.P.",
##                          "CONTAMINACION AMBIENTAL POR EXPLOTACION DE YACIMIENTO MINERO O HIDROCARBURO AR
##                          "CORRUPCION DE SUFRAGANTE ART. 390 C.P.",
##                          "DESTINACION ILEGAL DE COMBUSTIBLE ART. 327D C.P.",
##                          "EXPLOTACION ILICITA DE YACIMIENTO MINERO Y OTROS MATERIALES ART. 338 C.P.",
##                          "EXTORSION. ART. 244 C.P.",
##                          "FABRICACION, IMPORTACION, TRAFICO, POSESION Y USO DE ARMAS QUIMICAS, BIOLOGICA
##                          "FABRICACION, TRAFICO Y PORTE DE ARMAS DE FUEGO O MUNICIONES ART. 365 C.P.",
##                          "FAVORECIMIENTO DE CONTRABANDO DE HIDROCARBUROS Y SUS DERIVADOS ART. 320-1 C.P.
##                          "FAVORECIMIENTO DE LA FUGA ART. 449 C.P.",
##                          "FAVORECIMIENTO DEL DELITO DE CONTRABANDO DE HIDROCARBUROS O SUS DERIVADOS ART.
##                          "FAVORECIMIENTO POR SERVIDOR PUBLICO ART. 322 C.P.",
##                          "FAVORECIMIENTO POR SERVIDOR PUBLICO ART. 322 C.P. INCISO 1",
##                          "OMISION DE CONTROL ART. 325 C.P.",
##                          "REVELACION DE SECRETO ART. 418 C.P. INCISO 1",
##                          "SECUESTRO EXTORSIVO ART. 169 C.P.",
##                          "TERRORISMO ART. 343 C.P.",
##                          "TESTAFERRATO ART. 326 C.P.",
##                          "TESTAFERRATO ART. 326 C.P. MENOR 100 SALARIOS",
##                          "TRAFICO, TRANSPORTE Y POSESION DE MATERIALES RADIOACTIVOS O SUSTANCIAS NUCLEAR
##                          "USO DE MENORES DE EDAD LA COMISION DE DELITO ART 188D LEY 1453 DE 2011",
##                          "USURPACION DE TIERRAS. ART. 261 C.P.",
##                          "UTILIZACION INDEBIDA DE INFORMACION OFICIAL PRIVILEGIADA. ART. 420 C.P.",
##                          "UTILIZACION INDEBIDA DE INFORMACION PRIVILEGIADA. ART. 258 C.P.",
##                          "VIOLACION DE LA LIBERTAD DE TRABAJO. ART. 198 C.P.",
##                          "FRAUDE DE SUBVENCIONES",
##                          "HOMICIDIO PRETERINTENCIONAL ART. 105 C.P.",
##                          "INDUCCION A LA PROSTITUCION ART. 213 C.P.",
##                          "LAVADO DE ACTIVOS ART. 323 C.P.",
##                          "USURPACION DE MARCAS Y PATENTES ART. 306 C.P.",
##                          "DAÑO")
##
## delito_azul <- c("CONSTREÑIMIENTO PARA DELINQUIR ART. 184 C.P.",
##                  "ACTOS DE BARBARIE ART. 145 C.P.",
##                  "AMENAZAS A TESTIGOS ART. 454A C.P. AD. LEY 890 DE 2004 ART.13",
##                  "AMENAZAS ART. 347 C.P.",
##                  "APODERAMIENTO DE AERONAVES, NAVES O MEDIOS DE TRANSPORTE COLECTIVO. ART. 173 C.P."
##                  "APODERAMIENTO DE LOS HIDROCARBUROS, SUS DERIVADOS, BIOCOMBUSTIBLES O MEZCLAS QUE L
##                  "CONCIERTO PARA DELINQUIR ART. 340 C.P.",
##                  "DAÑO EN LOS RECURSOS NATURALES ART. 331 C.P.",
##                  "DAÑO EN OBRAS DE UTILIDAD SOCIAL ART. 351 C.P.",
##                  "DAÑO EN OBRAS O ELEMENTOS DE LOS SERVICIOS DE COMUNICACIONES, ENERGIA Y COMBUSTIBL
##                  "DESAPARICION FORZADA ART. 165 C.P.",
##                  "DESTINACION ILICITA DE MUEBLES O INMUEBLES ART. 377 C.P.",
##                  "DIVULGACION Y EMPLEO DE DOCUMENTOS RESERVADOS. ART. 194 C.P.",
##                  "EMPLEO ILEGAL DE LA FUERZA PUBLICA ART. 423 C.P.",
##                  "EMPLEO O LANZAMIENTO DE SUSTANCIAS U OBJETOS PELIGROSOS ART. 359 C.P.",
##                  "ENTRENAMIENTO PARA ACTIVIDADES ILICITAS ART. 341 C.P.",
##                  "ESPIONAJE ART. 463 C.P.",
##                  "FABRICACION, TRAFICO Y PORTE DE ARMAS DE USO PRIVATIVO DE LAS FUERZAS ARMADAS ART.
##                  "HOSTIGAMIENTO POR MOTIVOS DE RAZA, RELIGIÓN, IDEOLOGÍA, POLÍTICA, U ORIGEN NACIONA
##                  "OBSTRUCCION A VIAS PUBLICAS QUE AFECTAN EL ORDEN PUBLICO. ART.353A C.P. LEY 1453 D
##                  "TORTURA ART. 178 C.P.",
##                  "UTILIZACION DE ASUNTO SOMETIDO A SECRETO O RESERVA ART. 419 C.P.",
```

```
##                   "VIOLACION DE DATOS PERSONALES ART 269F LEY 1273 DE 2009",
##                   "VIOLACION DE HABITACION AJENA POR SERVIDOR PUBLICO ART. 190 C.P.",
##                   "VIOLACION DE LOS DERECHOS DE REUNION Y ASOCIACION. ART. 200 C.P.",
##                   "VIOLACION ILICITA DE COMUNICACIONES O CORRESPONDENCIA DE CARACTER OFICIAL. ART. 19
##                   "VIOLACION ILICITA DE COMUNICACIONES. ART. 192 C.P.",
##                   "VIOLENCIA CONTRA SERVIDOR PUBLICO ART. 429 C.P.")
##
## delito_na <- c("ABORTO ART. 122 C.P.",
##               "ABORTO SIN CONSENTIMIENTO ART. 123 C.P.",
##               "ABUSO DE AUTORIDAD POR OMISION DE DENUNCIA ART. 417 C.P.",
##               "ACCESO CARNAL ABUSIVO CON MENOR DE CATORCE ANOS. ART. 208 C.P.",
##               "ACCESO CARNAL O ACTO SEXUAL ABUSIVOS CON INCAPAZ DE RESISTIR. ART. 210 C.P.",
##               "ACCESO CARNAL O ACTO SEXUAL EN PERSONA PUESTA EN INCAPACIDAD DE RESISTIR. ART. 207 C.P
##               "ACCESO CARNAL VIOLENTO. ART. 205 C.P.",
##               "ACTOS SEXUALES CON MENOR DE CATORCE ANOS. ART. 209 C.P.",
##               "ALTERACION DE RESULTADOS ELECTORALES ART. 394 C.P.",
##               "ALTERACION DESFIGURACION Y SUPLANTACION DE MARCAS DE GANADO. ART. 243 C.P.",
##               "ALZAMIENTO DE BIENES. ART. 253 C.P.",
##               "CONCUSION ART. 404 C.P.",
##               "CONSERVACION O FINANCIACION DE PLANTACIONES ART. 375 C.P.",
##               "CONTRABANDO DE HIDROCARBUROS Y SUS DERIVADOS ART. 319-1 C.P.",
##               "DAÑO EN BIEN AJENO. ART. 265 C.P.",
##               "DAÑO EN MATERIA PRIMA, PRODUCTO AGROPECUARIO O INDUSTRIAL. ART. 304 C.P.",
##               "DESTINACION ILICITA DE MUEBLES O INMUEBLES ART. 377 C.P.",
##               "DISPARO DE ARMA DE FUEGO CONTRA VEHICULO ART. 356 C.P.",
##               "DIVULGACION Y EMPLEO DE DOCUMENTOS RESERVADOS. ART. 194 C.P.",
##               "EXPLOTACION ILICITA DE YACIMIENTO MINERO Y OTROS MATERIALES ART. 338 C.P.",
##               "FABRICACION Y COMERCIALIZACION DE SUSTANCIAS NOCIVAS PARA LA SALUD ART. 374 C.P.",
##               "FAVORECIMIENTO DE CONTRABANDO DE HIDROCARBUROS Y SUS DERIVADOS ART. 320-1 C.P.",
##               "FAVORECIMIENTO DEL DELITO DE CONTRABANDO DE HIDROCARBUROS O SUS DERIVADOS ART. 6 DECRE
##               "FRAUDE AL SUFRAGANTE ART. 388 C.P.",
##               "FRAUDE EN LA INSCRIPCION DE CEDULAS ART. 389 C.P.",
##               "FUGA DE PRESOS ART. 448 C.P.",
##               "HOMICIDIO ART. 103 C.P.",
##               "HURTO CALIFICADO. ART. 240 C.P.",
##               "HURTO. ART. 239 C.P.",
##               "ILICITO APROVECHAMIENTO DE LOS RECURSOS NATURALES RENOVABLES ART. 328 C.P.",
##               "INCENDIO ART. 350 C.P.",
##               "INSTIGACION A DELINQUIR ART. 348 C.P.",
##               "INSTIGACION A DELINQUIR ART. 348 C.P. INCISO 1",
##               "INTERVENCION EN POLITICA ART. 422 C.P.",
##               "INVASION DE AREAS DE ESPECIAL IMPORTANCIA ECOLOGICA ART. 337 C.P.",
##               "INVASION DE TIERRAS O EDIFICACIONES. ART. 263 C.P.",
##               "IRRESPETO A CADAVERES. ART. 204 C.P.",
##               "LAVADO DE ACTIVOS ART. 323 C.P.",
##               "MALTRATO MEDIANTE RESTRICCION A AL LIBERTAD FISICA. ART. 230 C.P.",
##               "OBSTRUCCION DE OBRAS DE DEFENSA Y ASISTENCIA ART. 364 C.P.",
##               "OMISION DE APOYO ART. 424 C.P.",
##               "OMISION DE SOCORRO ART. 131 C.P.",
##               "PERFIDIA ART. 143 C.P.",
##               "PERTURBACION DE ACTOS OFICIALES ART. 430 C.P. INCISO 1",
##               "PERTURBACION DE CERTAMEN DEMOCRATICO ART. 386 C.P.",
##               "PERTURBACION DE LA POSESION SOBRE INMUEBLE. ART. 264 C.P.",
##               "PERTURBACION EN SERVICIO DE TRANSPORTE COLECTIVO U OFICIAL ART. 353 C.P.",
```

```
##              "RECEPTACION ART. 447 C.P.",
##              "RECEPTACION ART. 327 C C.P.",
##              "REVELACION DE SECRETO ART. 418 C.P. INCISO 1",
##              "SIMULACION DE INVESTIDURA O CARGO ART. 426 C.P.",
##              "SINIESTRO O DAÑO DE NAVE ART. 354 C.P.",
##              "TENENCIA, FABRICACION Y TRAFICO DE SUSTANCIA U OBJETOS PELIGROSOS ART. 358 C.P.",
##              "TRAFICO DE SUSTANCIAS PARA PROCESAMIENTO DE NARCOTICOS ART. 382 C.P.",
##              "TRAFICO FABRICACION O PORTE DE ESTUPEFACIENTES ART. 376 C..",
##              "ULTRAJE A EMBLEMAS Y SIMBOLOS PATRIOS ART. 461 C.P.",
##              "USO DE DOCUMENTO FALSO. ART. 291 C.P.",
##              "USURPACION DE FUNCIONES PUBLICAS ART. 425 C.P.",
##              "UTILIZACION ILICITA DE EQUIPOS TRANSMISORES O RECEPTORES. ART. 197 C.P.",
##              "UTILIZACION INDEBIDA DE INFORMACION PRIVILEGIADA. ART. 258 C.P.")
##
## # Rules is conflict.
##
## one <- as.integer(1)
## zero <- as.integer(0)
##
## conflict_spoa <- data_SPOA_CEV %>%
##   mutate(prob = runif(nrow(.))) %>%
##   mutate(is_conflict = case_when(
##     tipo_vinculacion == "PERPETRADOR" ~ zero,
##     perpetrador %in% perp_verde ~ one,
##     perpetrador1 %in% perp_verde ~ one,
##     perpetrador2 %in% perp_verde ~ one,
##     perpetrador3 %in% perp_verde ~ one,
##     perpetrador4 %in% perp_verde ~ one,
##     perpetrador %in% perp_amarillo & prob < 0.5 ~ one,
##     perpetrador1 %in% perp_amarillo & prob < 0.5 ~ one,
##     perpetrador2 %in% perp_amarillo & prob < 0.5 ~ one,
##     perpetrador3 %in% perp_amarillo & prob < 0.5 ~ one,
##     perpetrador4 %in% perp_amarillo & prob < 0.5 ~ one,
##     perpetrador %in% perp_naranja & prob < 0.75 ~ zero,
##     perpetrador1 %in% perp_naranja & prob < 0.75 ~ zero,
##     perpetrador2 %in% perp_naranja & prob < 0.75 ~ zero,
##     perpetrador3 %in% perp_naranja & prob < 0.75 ~ zero,
##     perpetrador4 %in% perp_naranja & prob < 0.75 ~ zero,
##     perpetrador %in% perp_rojo ~ zero,
##     perpetrador1 %in% perp_rojo ~ zero,
##     perpetrador2 %in% perp_rojo ~ zero,
##     perpetrador3 %in% perp_rojo ~ zero,
##     perpetrador4 %in% perp_rojo ~ zero,
##     perpetrador %in% perp_is_ca &  tipo_vinculacion == "VICTIMA" ~ one,
##     perpetrador1 %in% perp_is_ca &  tipo_vinculacion == "VICTIMA" ~ one,
##     perpetrador2 %in% perp_is_ca & tipo_vinculacion == "VICTIMA" ~ one,
##     perpetrador3 %in% perp_is_ca & tipo_vinculacion == "VICTIMA" ~ one,
##     perpetrador4 %in% perp_is_ca & tipo_vinculacion == "VICTIMA" ~ one,
##     delito %in% delito_na ~ NA_integer_,
##     delito %in% delito_verde ~ one,
##     delito %in% delito_azul & prob < 0.75 ~ one,
##     delito %in% delito_amarillo & prob < 0.5 ~ one,
##     delito %in% delito_naranja & prob < 0.75 ~ zero,
##     delito %in% delito_rojo ~ zero,
```

```
##     titulo %in% titulo_verde ~ one,
##     titulo %in% titulo_amarillo & prob < 0.5 ~ one,
##     titulo %in% titulo_naranja & prob < 0.75 ~ zero,
##     titulo %in% titulo_rojo ~ zero,
##     tipo_vinculacion == "OTRA" ~ NA_integer_,
##     TRUE ~ NA_integer_)) %>%
##   select(-prob)
##
##
## sample <- conflict_spoa  %>%
##   filter(is_conflict ==0) %>%
##   select(recordid, is_conflict)
##
## # ------- training data
##
## sample <- sample_frac(sample, 0.2)
##
## write.table(sample, file = args$examples, sep = "|", quote = FALSE, row.names = FALSE)
##
## write_parquet(conflict_spoa, args$output)
## #done
```

```
cat(readLines(files$is_ca_sijuf_jep_fase4), sep = "\n")
```

9.4.1.5 FGN - SIJUF - JEP FASE 4

```
## #
## # Authors:     PA
## # Maintainers  VG, PA, PB, JGD
## # Copyright   2022, HRDAG,
## # =========================
## # CO-SIVJRNR-data/individual/FGN/is-ca/src/is-ca-sijuf_jep_fase4.R
##
## # following the email from Folco Zaffalon (8 abr 2022, 17:16 PDT),
## # and the accompanying file
## # ‘SIJUF_casos_perpetrador_ocupacion victima_delitos.docx‘
## #
## # Verde = SI CONFLICTO (1)
## # Azul = MUY PROBABLE 0.75~(is_conflict == 1)
## # Amarillo = NO SE SABE 0.5~(is_conflict == 1)
## # Naranja = POCO PROBABLE 0.75~(is_conflict == 0)
## # Rojo = NO CONFLICTO (0)
##
## # ------ setup
## Sys.setlocale("LC_CTYPE", "en_US.UTF-8")
## set.seed(19481210)   # <-- the seed is really important in this run!
##
## require(pacman)
## p_load(argparse, dplyr, here, arrow, assertr, stringr, lubridate, readr)
##
## stopifnot(endsWith(getwd(),"is-ca"))
```

```
##
## parser <- ArgumentParser()
## parser$add_argument("--sijuf_jep_fase4",
##                     default=here("individual/FGN/clean/output/sijuf_jep_fase4.parquet"))
## parser$add_argument("--examples",
##                     default = "individual/FGN/is-ca/hand/example_sijuf_jep.csv")
## parser$add_argument("--output",
##                     default = "output/sijuf_jep_fase4.parquet")
## args <- parser$parse_args()
##
## # ------- function
##
## # ------- main
##
## data_sijuf_jep_fase4 <- read_parquet(args$sijuf_jep_fase4)
##
## perp_is_ca = c("AUC - AUTODEFENSAS UNIDAS DE COLOMBIA",
##                "AGENTES_ESTATALES","ELN - EJERCITO DE LIBERACION NACIONAL",
##                "EPL - EJERCITO POPULAR DE LIBERACION",
##                "ERP - EJERCITO REVOLUCIONARIO DEL PUEBLO",
##                "FARC - FUERZAS ARMADAS REVOLUCIONARIAS COLOMB",
##                "GUERRILLA", "OTROS GRUPOS PARAMILITARES", "PARAMILITARES")
##
## perp_no_ca = c("COMUNIDADES INDIGENAS", "DELINCUENCIA COMUN")
##
## perp_verde <- c("AGENTE DEL ESTADO", "AGUILAS NEGRAS", "AUC", "CORDILLERA ELN", "GAO", "GUERRILLA",
##                 "INDEPENDIENTES", "LOS RASTROJOS", "M19", "MACHETEROS", "PARAMILITARES",
##                 "ELN", "EPL", "ERP", "ERG", "FARC", "GAO")
##
## perp_azul <- c("FUERZA PUBLICA",  "AGENTE DEL ESTADO")
##
## perp_naranja <- c("INDEPENDIENTES")
##
## delito_verde <- c("ACCESO CARNAL VIOLENTO EN PERSONA PROTEGIDA ART. 138 C.P.",
##                   "ACTOS DE TERRORISMO ART. 144 C.P.",
##                   "ACTOS SEXUALES VIOLENTOS EN PERSONA PROTEGIDA ART. 139 C.P.",
##                   "ADMINISTRACION DE RECURSOS CON FINES TERRORISTAS ART. 17 DEC180 DE 1988",
##                   "ADMINISTRACION DE RECURSOS RELACIONADOS CON ACTIVIDADES TERRORISTAS ART. 345 C.P.
##                   "APOLOGIA DEL GENOCIDIO ART. 102 C.P.",
##                   "ASONADA ART. 128 C.P.",
##                   "ASONADA ART. 469 C.P.",
##                   "ATAQUE CONTRA OBRAS E INSTALACIONES QUE CONTIENEN FUERZAS PELIGROSAS ART. 157 C.P
##                   "ATENTADOS A LA SUBSISTENCIA Y DEVASTACION ART. 160 C.P.",
##                   "CONSPIRACION ART. 130 C.P.",
##                   "CONSPIRACION ART. 471 C.P.",
##                   "CONSTRENIMIENTO A APOYO BELICO ART. 150 C.P.",
##                   "DEPORTACION, EXPULSION , TRASLADO O DESPLAZAMIENTO FORZADO DE POBLACION CIVIL ART
##                   "DEPORTACION, EXPULSION, TRASLADO O DESPLAZAMIENTO FORZADO DE POBLACION CIVIL ART.
##                   "DESPLAZAMIENTO FORZADO ART. 180 C.P.",
##                   "DESPLAZAMIENTO FORZADO LEY 589/2000",
##                   "DESPOJO EN EL CAMPO DE BATALLA ART. 151 C.P.",
##                   "DESTRUCCION Y APROPIACION DE BIENES PROTEGIDAS ART. 154 C.P.",
##                   "DESTRUCCION Y APROPIACION DE BIENES PROTEGIDOS ART. 154 C.P.",
##                   "ENTRENAMIENTO PARA ACTIVIDADES SICARIALES ART. 3 DEC.1194 DE 1989",
```

```
##                    "GENOCIDIO ART. 101 C.P.",
##                    "HOMICIDIO EN PERSONA PROTEGIDA ART. 135 C.P.",
##                    "HOMICIDIO EN PERSONAS PROTEGIDAS ART. 135 C.P.",
##                    "INSTIGACION A LA GUERRA ART. 458 C.P.",
##                    "INSTIGACION AL TERRORISMO ART. 8 DEC. 180 DE 1988",
##                    "INSTIGACION O CONSTRENIMIENTO PARA INGRESO A GRUPOS TERRORISTAS ART. 6 DEC. 180 DE
##                    "INSTRUCCION Y ENTRENAMIENTO AL TERRORISMO ART. 15 DEC.180 DE 1988",
##                    "LESIONES EN PERSONA PROTEGIDA ART. 136 C.P.",
##                    "LESIONES PERSONALES FINES TERRORISTAS ART. 31 DEC. 180 DE 1988",
##                    "OMISION DE INFORMES SOBRE ACTIVIDADES TERRORISTAS ART. 4 DEC. 180 DE 1988",
##                    "OMISION DE MEDIDAS DE PROTECCION A LA POBLACION CIVIL ART. 161 C.P.",
##                    "OMISION DE MEDIDAS DE SOCORRO Y ASISTENCIA HUMANITARIA ART. 152 C.P.",
##                    "OMISION DE MEDIDAS DE SOCORRO Y ASISTENCIA HUMANITARIAS ART. 152 C.P.",
##                    "REBELION ART. 125 C.P. MOD. ART. 1 DEC. 1857 DE 1989",
##                    "REBELION ART. 467 C.P.",
##                    "RECLUTAMIENTO ILICITO ART. 162 C.P.",
##                    "REPRESALIAS ART. 158 C.P.",
##                    "SEDICION ART. 126 C.P.",
##                    "SEDICION ART. 468 C.P.",
##                    "SEDUCCION, USURPACION Y RETENCION ILEGAL DEL MANDO ART. 472 C.P.",
##                    "TERRORISMO ART. 343 C.P.",
##                    "TERRORISMO ART.187 C.P.",
##                    "TOMA DE REHENES ART. 148 C.P.",
##                    "TORTURA CONTRA PERSONA PROTEGIDA ART. 137 C.P.",
##                    "TORTURA EN PERSONA PROTEGIDA ART. 137 C.P.",
##                    "TRATOS INHUMANOS Y DEGRADANTES Y EXPERIMENTOS BIOLOGICOS EN PERSONA PROTEGIDA ART
##                    "TRATOS INHUMANOS Y DEGRADANTES Y EXPERIMENTOS BIOLOGICOS EN PERSONA PROTEGIDA ART
##                    "UTILIZACION DE MEDIOS Y METODOS DE GUERRA ILICITOS ART. 142 C.P.")
##
## delito_azul <- c("ACTOS DE BARBARIE ART. 145 C.P.",
##                   "ACTOS DE DISCRIMINACION RACIAL ART. 147 C.P.",
##                   "AMENAZAS ART. 347 C.P.",
##                   "AMENAZAS PERSONALES O FAMILIARES ART. 26 DEC. 180 DE 1988",
##                   "DESAPARICION FORZADA ART. 165 C.P.",
##                   "DESAPARICION FORZADA LEY 589/2000",
##                   "DESTRUCCION O UTILIZACION ILICITA DE BIENES CULTURALES Y DE LUGARES DE CULTO ART. 1
##                   "DETENCION ARBITRARIA ESPECIAL ART. 176 C.P.",
##                   "DETENCION ARBITRARIA ESPECIAL ART. 274 C.P.",
##                   "DETENCION ILEGAL Y PRIVACION DEL DEBIDO PROCESO ART. 149 C.P.",
##                   "DISPARO DE ARMA DE FUEGO CONTRA VEHICULO ART. 356 C.P.",
##                   "DISPARO DE ARMA DE FUEGO CONTRA VEHICULO ART.195 C.P.",
##                   "EMPLEO ILEGAL DE LA FUERZA PUBLICA ART. 159 C.P.",
##                   "EMPLEO ILEGAL DE LA FUERZA PUBLICA ART. 423 C.P.",
##                   "EMPLEO O LANZAMIENTO DE SUSTANCIAS U OBJETOS PELIGROSOS ART. 198 C.P.",
##                   "EMPLEO O LANZAMIENTO DE SUSTANCIAS U OBJETOS PELIGROSOS ART. 359 C.P.",
##                   "EMPLEO O LANZAMIENTO DE SUSTANCIAS U OBJETOS PELIGROSOS CON FINES TERRORISTAS ART.
##                   "EMPLEO, PRODUCCION, COMERCIALIZACION Y ALMACENAMIENTO DE MINAS ANTIPERSONAL ARCT 3
##                   "ENRIQUECIMIENTO ILICITO DERIVADO DE SECUESTRO ART. 6 LEY 40 DE 1993",
##                   "ENTRENAMIENTO PARA ACTIVIDADES ILICITAS ART. 341 C.P.",
##                   "EXACCION O CONTRIBUCIONES ARBITRARIAS ART. 163 C.P.",
##                   "EXTORSION ART. 355 C.P.",
##                   "EXTORSION. ART. 244 C.P.",
##                   "FABRICACION Y TRAFICO DE ARMAS DE FUEGO O MUNICIONES ART. 201 C.P.",
##                   "FABRICACION Y TRAFICO DE ARMAS Y MUNICIONES DE USO PRIVATIVO DE LAS FUERZAS ARMADAS
```

```
##                  "FABRICACION, IMPORTACION, TRAFICO POSESION Y USO DE ARMAS QUIMICAS, BIOLOGICAS Y NU
##                  "FABRICACION, IMPORTACION, TRAFICO, POSESION Y USO DE ARMAS QUIMICAS, BIOLOGICAS Y N
##                  "FABRICACION, TRAFICO Y PORTE DE ARMAS DE FUEGO O MUNICIONES ART. 365 C.P.",
##                  "FABRICACION, TRAFICO Y PORTE DE ARMAS DE USO PRIVATIVO DE LAS FUERZAS ARMADAS ART.
##                  "INCITACION A LA INSUBORDINACION POR MEDIOS PUBLICOS ART. 28 LEY 29 DE 1944",
##                  "OMISION DE INFORMES EN SECUESTRO ART. 9 LEY 40 DE 1993",
##                  "OMISION DE SOCORRO ART. 131 C.P.",
##                  "PERTENENCIA A BANDAS DE SICARIOS ART. 2 DEC-LEY 1194789 ACLP ART. 6 DEC 2266791",
##                  "PERTURBACION DE ACTOS OFICIALES ART. 430 C.P.",
##                  "PERTURBACION DE CERTAMEN DEMOCRATICO ART. 386 C.P.",
##                  "PROSTITUCION FORZADA O ESCLAVITUD SEXUAL ART. 141 C.P.",
##                  "SECUESTRO EXTORSIVO AGRAVADO ART. 170 C.P.",
##                  "SECUESTRO EXTORSIVO ART. 169 C.P.",
##                  "SECUESTRO EXTORSIVO ART. 268 C.P.",
##                  "SIMULACION DE INVESTIDURA O CARGO ART. 426 C.P.",
##                  "SINIESTRO O DANO DE NAVE ART. 193 C.P.",
##                  "SINIESTRO O DANO DE NAVE ART. 354 C.P.",
##                  "SUPLANTACION DE AUTORIDAD ART. 20 DEC. 180 DE 1986",
##                  "TESTAFERRATO ART. 326 C.P.",
##                  "TESTAFERRATO ART.6 DEC. LEY 1856 DE 1989",
##                  "TORTURA ART. 178 C.P.",
##                  "TORTURA ART. 279 C.P.",
##                  "UTILIZACION ILEGAL DE UNIFORMES E INSIGNIAS ART. 346 C.P.",
##                  "UTILIZACION ILEGAL DE UNIFROMES O INSIGNIAS ART. 19 DEC. 180 DE 1988",
##                  "UTILIZACION ILICITA DE EQUIPOS TRANSMISORES O RECEPTOR ART. 16 DEC. 180 DE 1988",
##                  "UTILIZACION ILICITA DE EQUIPOS TRANSMISORES O RECEPTORES ART. 197 C.P.",
##                  "UTILIZACION ILICITA DE EQUIPOS TRANSMISORES O RECEPTORES. ART. 197 C.P.",
##                  "VIOLACION DE LOS DERECHOS DE REUNION Y ASOCIACION 292 C.P.",
##                  "VIOLACION DE LOS DERECHOS DE REUNION Y ASOCIACION. ART. 200 C.P.")
##
## delito_naranja <- c("ACCESO ABUSIVO A UN SISTEMA INFORMATICO ART. 195 C.P.",
##                     "ACCESO ABUSIVO A UN SISTEMA INFORMATICO. ART. 195 C.P.",
##                     "ACCESO CARNAL MEDIANTE ENGANO ART. 301 C.P.",
##                     "ASOCIACION PARA LA COMISION DE UN DELITO CONTRA LA ADMINISTRACION PUBLICA ART. 4
##                     "CONTRABANDO ART. 15 LEY 383/97",
##                     "CONTRABANDO ART. 319 C.P.",
##                     "CONTRABANDO DE HIDROCARBUROS O SUS DERIVADOS DTO 1900/02",
##                     "CONTRATO SIN CUMPLIMIENTO DE REQUISITOS LEGALES ART. 146 C.P.",
##                     "CONTRATO SIN CUMPLIMIENTO DE REQUISITOS LEGALES ART. 410 C.P.",
##                     "DE LA EXTINCION DE DOMINIO",
##                     "DESTINACION DE MUEBLES E INMUEBLES ART. 34 LEY 30 DE 1986",
##                     "DESTINACION ILICITA DE MUEBLES O INMUEBLES ART. 377 C.P.",
##                      "HOMICIDIO ART. 103 C.P.",
##                     "HURTO ART. 349 C.P.",
##                     "HURTO ART. 349 C.P.MOD. ART. 1 N.11 LEY 23 DE 1991",
##                     "HURTO. ART. 239 C.P.",
##                     "PISTAS DE ATERRIZAJE ILEGALES (RESPONSABILIDAD DEL DUENO POSEEDOR O ARRENDATARIO
##                     "REINGRESO ILEGAL AL PAIS ART. 185 C.P.",
##                     "REVELACION DE SECRETO ART. 154 C.P.",
##                     "REVELACION DE SECRETO ART. 418 C.P.",
##                     "SECUESTRO SIMPLE ART. 168 C.P.",
##                     "SECUESTRO SIMPLE ART. 269 C.P.",
##                     "SEDUCCION USURPACION Y RETENCION ILEGAL DEL MANDO ART. 131 C.P.",
##                     "SOBORNO ART. 174 C.P.",
```

```
##                      "SOBORNO ART. 174 C.P.",
##                      "SOBORNO ART. 444 C.P.",
##                      "SOBORNO EN LA ACTUACION PENAL ART 444A C.P.",
##                      "TRAFICO DE INFLUENCIAS DE SERVIDOR PUBLICO ART. 411 C.P.",
##                      "TRAFICO DE MONEDA FALSIFICADA ART. 208 C.P.",
##                      "TRAFICO DE MONEDA FALSIFICADA. ART. 274 C.P.",
##                      "TRATA DE PERSONAS ART. 311 C.P",
##                      "TRATA DE PERSONAS. ART. 215 C.P.",
##                      "USURPACION DE AGUAS. ART. 262 C.P.",
##                      "USURPACION DE FUNCIONES PUBLICAS ART. 161 C.P.",
##                      "VIOLACION DE HABITACION AJENA ART. 284 C.P.",
##                      "VIOLACION DE HABITACION AJENA. ART. 189 C.P.",
##                      "VIOLACION DE INMUNIDAD DIPLOMATICA ART. 465 C.P.")
##
## delito_rojo <- c("ABANDONO ART. 127 C.P.",
##                   "ABANDONO ART. 346 C.P.",
##                   "ABANDONO DEL CARGO ART. 156 C.P.",
##                   "ABUSO DE CONFIANZA ART. 358 C.P.",
##                   "CONFIANZA ART. 358 C.P.MOD. ART. 1 N.16 LEY 23 DE 1991",
##                   "ABUSO DE CONFIANZA CALIFICADO ART. 250 C.P.",
##                   "ABUSO DE CONFIANZA. ART. 249 C.P.",
##                   "ACCESO O PRESTACION ILEGAL DE LOS SERVICIOS DE TELECOMUNICACIONES ART.6 LEY 422/98
##                   "ACTO SEXUAL MEDIANTE ENGANO ART. 302 C.P.",
##                   "ADOPCION ILEGAL DE MENORES ART. 267 DEC. 2737 DE 1989",
##                   "AGIOTAJE. ART. 301",
##                   "ALTERACION Y MODIFICACION DE CALIDAD CANTIDAD PESO O MEDIDA ART. 231 C.P.",
##                   "ALTERACION Y MODIFICACION DE CALIDAD, CANTIDAD, PESO O MEDIDA. ART. 299 C.P.",
##                   "APLICACION FRAUDULENTA DE CREDITO OFICIALMENTE REGULADO ART. 241 C.P.",
##                   "APLICACION FRAUDULENTA DE CREDITO OFICIALMENTE REGULADO ART. 311 C.P.",
##                   "APROVECHAMIENTO DE ERROR AJENO O CASO FORTUITO ART. 361 C.P.",
##                   "APROVECHAMIENTO DE ERROR AJENO O CASO FORTUITO. ART. 252 C.P.",
##                   "ASESORAMIENTO Y OTRAS ACTUACIONES ILEGALES ART. 421 C.P.",
##                   "BIGAMIA ART. 260 C.P.",
##                   "CALUMNIA ART. 314 C.P.",
##                   "CALUMNIA INDIRECTA ART. 315 C.P.",
##                   "CALUMNIA. ART. 221 C.P.",
##                   "CAPTACION MASIVA Y HABITUAL ART. 208 N.3 DEC. 663 DE 1993",
##                   "CAZA ILEGAL ART. 336",
##                   "CELEBRACION INDEBIDA DE CONTRATO DE SEGURO ART. 12 LEY 40 DE 1993",
##                   "CELEBRACION INDEBIDA DE CONTRATO DE SEGURO ART. 172 C.P.",
##                   "CELEBRACION INDEBIDA DE CONTRATOS DE SEGUROS ART. 172 C.P.",
##                   "CIRCULACION Y USO DE EFECTO OFICIAL O SELLO FALSIFICADO. ART. 281 C.P.",
##                   "COHECHO IMPROPIO ART. 142 C.P.",
##                   "COHECHO IMPROPIO ART. 406 C.P.",
##                   "COHECHO POR DAR U OFRECER ART. 143 C.",
##                   "COHECHO POR DAR U OFRECER ART. 407 C.P.",
##                   "COHECHO PROPIO ART. 141 C.P.",
##                   "COHECHO POR DAR U OFRECER ART. 407 C.P.",
##                   "COHECHO PROPIO ART. 141 C.P.",
##                   "COHECHO PROPIO ART. 405 C.P.",
##                   "CONSTANCIAS Y ESTADOS FINANCIEROS FALSOS ART. 43 LEY 222/95",
##                   "CORRUPCION DE ALIMENTOS Y MEDICINAS ART. 206 C.P.",
##                   "CORRUPCION DE ALIMENTOS, PRODUCTOS MEDICOS O MATERIAL PROFILACTICO ART. 372 C.P.",
##                   "DEFRAUDACION A LAS RENTAS DE ADUANA ART.17 LEY 383/97",
```

```
##                     "DEFRAUDACION A LOS DERECHOS PATRIMONIALES DE AUTOR. ART. 271 C.P.",
##                     "DEFRAUDACION DE FLUIDOS ART. 256 C.P.",
##                     "DEFRAUDACION DE FLUIDOS. ART. 256 C.P.",
##                     "DEL ACCESO ILEGAL O PRESTACION ILEGAL DE LOS SERVICIOS DE TELECOMUNICACIONES. ART.
##                     "DEL URBANIZADOR ILEGAL ART. 367A CP.",
##                     "DESTRUCCION SUPRESION OCULTA DOCUMENTO PRIVADO ART. 224 C.P.",
##                     "DESTRUCCION SUPRESION OCULTA DOCUMENTO PUBLICO ART. 223 C.P",
##                     "DESTRUCCION, SUPRESION U OCULTAMIENTO DE DOCUMENTO PUBLICO. ART. 292 C.P.",
##                     "DESTRUCCION, SUPRESION Y OCULTAMIENTO DE DOCUMENTO PRIVADO. ART. 293 C.P.",
##                     "DISPOSICION DE BIEN PROPIO GRAVADO CON PRENDA ART. 364 C.P.",
##                     "DISPOSICION DE BIEN PROPIO GRAVADO CON PRENDA. ART. 255 C.P.",
##                     "DIVULGACION Y EMPLEO DE DOCUMENTOS RESERVADOS ART. 289 C.P.",
##                     "DIVULGACION Y EMPLEO DE DOCUMENTOS RESERVADOS. ART. 194 C.P.",
##                     "EJECUCION O REPRODUCCION DE OBRAS ARTISTICAS SIN AUTORIZACION DEL TITULAR DEL DERE
##                     "EJERCICIO ARBITRARIO DE LA CUSTODIA DE HIJO MENOR DE EDAD ART 230A C.P.",
##                     "EJERCICIO ILICITO DE ACTIVIDAD MONOPOLISTICA DE ARBITRIO RENTISTICO ART. 312 C.P."
##                     "EJERCICIO ILICITO DE ACTIVIDADES MONOPOLISTICAS DE ARBITRIO RENTISITICO ART. 241A
##                     "EMISION Y TRANSFERENCIA ILEGAL DE CHEQUE. ART. 248 C.P.",
##                     "EMISION Y TRANSFERENCIA ILEGAL DE CHEQUES ART. 357 C.P.",
##                     "EMISION Y TRANSFERENCIA ILEGAL DE CHEQUES ART. 357 C.P. MOD. ART. 1 N.15 LEY 23 DE
##                     "EMISIONES ILEGALES. ART. 276 C.P.",
##                     "ENRIQUECIMIENTO ILICITO ART. 412 C.P.",
##                     "ENRIQUECIMIENTO ILICITO DE PARTICULARES ART. 1 DEC. 1895 DE 1989",
##                     "ENRIQUECIMIENTO ILICITO DE PARTICULARES ART. 327 C.P.",
##                     "ENRIQUECIMIENTO ILICITO DE SERVIDOR PUBLICO ART. 148 C.P.",
##                     "ESTAFA ART. 356 C.P.MOD. ART. 4 N.1 LEY 23 DE 1991",
##                     "ESTAFA. ART. 246 C.P.",
##                     "ESTIMULO A LA PROSTITUCION DE MENORES ART. 312 C.P",
##                     "ESTIMULO A LA PROSTITUCION DE MENORES. ART. 217 C.P.",
##                     "ESTIMULO AL USO ILICITO ART. 378 C.P.",
##                     "EXISTENCIA, CONSTRUCCION Y UTILIZACION ILEGAL DE PISTAS DE ATERRIZAJE ART. 385 C.P
##                     "EXPERIMENTACION ILEGAL EN ESPECIES ANIMALES O VEGETALES ART. 334 C.P.",
##                     "EXPLORATACION O EXPLORACION ILICITA MINERA O PETROLERA ART. 244 C.P. MOD. ART. 21
##                     "EXPLOTACION ILICITA DE YACIMIENTO MINERO Y OTROS MATERIALES ART. 338 C.P.",
##                     "FALSA AUTOACUSACION ART. 168 C.P.",
##                     "FALSA AUTOACUSACION ART. 437 C.P.",
##                     "FALSA DENUNCIA ART. 166 C.P.",
##                     "FALSA DENUNCIA ART. 435 C.P.",
##                     "FALSA DENUNCIA CONTRA PERSONA DETERMINADA ART. 167 C.P.",
##                     "FALSA DENUNCIA CONTRA PERSONA DETERMINADA ART. 436 C.P.",
##                     "FALSEDAD DERECHOS DE AUTOR ART. 51 LEY 44 DE 1993",
##                     "FALSEDAD EN DOCUMENTO PRIVADO ART. 221 C.P.",
##                     "FALSEDAD EN DOCUMENTO PRIVADO. ART. 289 C.P.",
##                     "FALSEDAD IDEOLOGICA EN DOCUMENTO PUBLICO ART. 219 C.P",
##                     "FALSEDAD IDEOLOGICA EN DOCUMENTO PUBLICO. ART. 286 C.P.",
##                     "FALSEDAD MARCARIA ART. 217 C.P.",
##                     "FALSEDAD MARCARIA. ART. 285 C.P.",
##                     "FALSEDAD MATERIAL EMPLEADO OFICIAL DOCUMENTO PUBLICO ART.218 CP",
##                     "FALSEDAD MATERIAL EN DOCUMENTO PUBLICO. ART. 287 C.P.",
##                     "FALSEDAD MATERIAL PARTICULAR DOCUMENTO PUBLICO ART. 220 C.P.",
##                     "FALSEDAD PERSONAL ART. 227 C.P.",
##                     "FALSEDAD PERSONAL PARA OBTENCION DOCUMENTO PUBLICO ART. 226 CP",
##                     "FALSEDAD PERSONAL. ART. 296 C.P.",
##                     "FALSIFICACION DE EFECTO OFICIAL TIMBRADO. ART. 280 C.P.",
```

```
##                    "FALSIFICACION DE MONEDA NACIONAL O EXTRANJERA ART. 207 C.P.",
##                    "FALSIFICACION DE MONEDA NACIONAL O EXTRANJERA. ART. 273 C.P.",
##                    "FALSO TESTIMONIO ART. 172 C.P.",
##                    "FALSO TESTIMONIO ART. 442 C.P.",
##                    "FAVORECIMIENTO ART. 176 C.P. MOD. ART. 6 LEY 356/97",
##                    "FAVORECIMIENTO ART. 446 C.P.",
##                    "FAVORECIMIENTO CONTRABANDO DE HIDROCARBUROS - DERIVADOS DTO 1900/02",
##                    "FAVORECIMIENTO DE CONTRABANDO ART. 320 C.P.",
##                    "FAVORECIMIENTO DE CONTRABANDO ART.16 LEY 383/97",
##                    "FECUNDACION Y TRAFICO DE EMBRIONES ART. 134 C.P.",
##                    "FRAUDE A RESOLUCION JUDICIAL ART. 454 C.P.",
##                    "FRAUDE A RESOLUCION JUDICIAL ART.184 C.P.",
##                    "FRAUDE PROCESAL ART. 182 C.P.",
##                    "FRAUDE PROCESAL ART. 453 C.P.",
##                    "INDEBIDA DE LOS RECURSOS SOCIALES. ART. 260 C.P.",
##                    "GESTION INDEBIDA DE RECURSOS SOCIALES ART. 260 C.P.",
##                    "HOMICIDIO CULPOSO ART. 109 C.P.",
##                    "HOMICIDIO CULPOSO ART. 329 C.P.",
##                    "HOMICIDIO POR PIEDAD ART. 106 C.P.",
##                    "HOMICIDIO POR PIEDAD ART.326 C.P.",
##                    "HOMICIDIO PRETERINTENCIONAL ART. 105 C.P.",
##                    "IMITACION O SIMULACION DE ALIMENTOS, PRODUCTOS O SUSTANCIAS ART. 373 C.P.",
##                    "INASISTENCIA ALIMENTARIA ART. 233 C.P.",
##                    "INASISTENCIA ALIMENTARIA ART. 263 C.P.",
##                    "INCESTO ART. 259 C.P",
##                    "INCESTO. ART. 237 C.P.",
##                    "INDUCCION A AYUDA AL SUICIDIO ART. 107 C.P.",
##                    "INDUCCION A LA PROSTITUCION ART. 213 C.P.",
##                    "INDUCCION A LA PROSTITUCION ART. 308 C.P.",
##                    "INDUCCION O AYUDA AL SUICIDIO ART. 327 C.P.",
##                    "INFIDELIDAD A LOS DEBERES PROFESIONALES ART. 175 C.P.",
##                    "INFIDELIDAD A LOS DEBERES PROFESIONALES ART. 445 C.P.",
##                    "INJURIA ART. 313 C.P.",
##                    "INJURIA POR VIAS DE HECHO ART. 319 C.P.",
##                    "INJURIA. ART. 220 C.P.",
##                    "INSEMINACION ARTIFICIAL O TRANSFERENCIA DE OVULO FECUNDADO NO CONSENTIDAS. ART. 18
##                    "INTERES ILICITO EN LA CELEBRACION DE CONTRATOS ART. 145 C.P.",
##                    "INTERES INDEBIDO EN LA CELEBRACION DE CONTRATOS ART. 409 C.P.",
##                    "LESIONES AL FETO ART. 125 C.P.",
##                    "LESIONES ART. 111 C.P.",
##                    "LESIONES CULPOSAS AL FETO ART. 126 C.P.",
##                    "LESIONES CULPOSAS ART. 120 C.P.",
##                    "LESIONES PERSONALES CULPOSAS ART. 340 C.P.",
##                    "LESIONES PERSONALES CULPOSAS ART. 340 C.P.MOD. ART. 12 LEY 228 DE 1995",
##                    "MALVERSACION Y DILAPIDACION DE BIENES FAMILIARES. ART. 236 C.P.",
##                    "MATRIMONIO ILEGAL ART. 261 C.P.",
##                    "MENDICIDAD Y TRAFICO DE MENORES. ART. 231 C.P.",
##                    "MODALIDAD CULPOSA DEL FAVORECIMIENTO DE LA FUGA ART. 180 C.P.",
##                    "MODALIDAD CULPOSA DEL FAVORECIMIENTO DE LA FUGA ART. 450 C.P.",
##                    "MUERTE DE HIJO FRUTO DE ACCESO CARNAL VIOLENTO, ABUSIVO O DE INSEMINACION ARTIFICI
##                    "OBTENCION DE DOCUMENTO PUBLICO FALSO ART. 288 C.P.",
##                    "OCULTAMIENTO RETENCION Y POSESION ILICITA DE CEDULA ART. 257 C.P.",
##                    "OCULTAMIENTO, ALTERACION O DESTRUCCION DE ELEMENTO MATERIAL PROBATORIO ART. 454B",
##                    "OCULTAMIENTO, RETENCION Y POSESION ILICITA DE CEDULA ART. 395 C.P.",
```

```
##                   "OMISION DE DENUNCIA DE PARTICULAR",
##                   "OMISION DE DENUNCIA DE PARTICULAR ART. 441 C.P.",
##                   "OMISION DEL AGENTE RETENEDOR O RECAUDADOR ART. 402 C.P.",
##                   "PANICO ECONOMICO ART. 302 C.P.",
##                   "PARTO O ABORTO PRETERINTENCIONAL ART. 118 C.P.",
##                   "PECULADO CULPOSO ART. 137 C.P.",
##                   "PECULADO CULPOSO ART. 400 C.P.",
##                   "PECULADO POR APLICACION OFICIAL DIFERENTE ART. 136 C.P.",
##                   "PECULADO POR APLICACION OFICIAL DIFERENTE ART. 399 C.P.",
##                   "PECULADO POR APROPIACION ART. 133 C.P.",
##                   "PECULADO POR APROPIACION ART. 397 C.P.",
##                   "PECULADO POR EXTENSION ART. 138 C.P",
##                   "PECULADO POR USO ART. 134 C.P.",
##                   "PECULADO POR USO ART. 398 C.P.",
##                   "PERFIDIA ART. 143 C.P.",
##                   "PORNOGRAFIA CON MENORES ART. 218 C.P.",
##                   "PREVARICATO POR ACCION ART. 149 C.P.",
##                   "PREVARICATO POR ACCION ART. 413 C.P",
##                   "PREVARICATO POR ASESORAMIENTO ILEGAL ART. 151 C.P.",
##                   "PREVARICATO POR OMISION ART. 150 C.P.",
##                   "PREVARICATO POR OMISION ART. 414 C.P.",
##                   "PROPAGACION DEL VIRUS DE INMUNODEFICIENCIA HUMANA O DE LA HEPATITIS B ART. 370 C.P
##                   "REPETIBILIDAD DEL SER HUMANO ART. 13 C.P.",
##                   "REPETIBILIDAD DEL SER HUMANO ART. 133",
##                   "ALTERACION O SUPOSICION DEL ESTADO CIVIL. ART. 238 C.P.",
##                   "SUSTRACCION DE BIEN PROPIO. ART. 254 C.P.",
##                   "TRAFICO INFLUENCIAS OBTENER FAVOR SERVIDOR PUBLICO ART. 147 C.P.",
##                   "TURISMO SEXUAL. ART. 219 C.P.",
##                   "URBANIZACION ILEGAL ART. 318 C.P.",
##                   "USURA ART. 235 C.P.",
##                   "USURA ART. 305 C.P.",
##                   "USURPACION DE MARCAS Y PATENTES ART. 236 C.P.",
##                   "USURPACION DE MARCAS Y PATENTES ART. 306 C.P.",
##                   "UTILIZACION INDEBIDA DE INFORMACION PRIVILEGIADA. ART. 258 C.P.",
##                   "VIOLACION A LOS DERECHOS MORALES DE AUTOR ART. 270 C.P.",
##                   "VIOLACION A LOS DERECHOS MORALES DE AUTOR. ART. 270 C.P.",
##                   "VIOLACION A LOS MECANISMOS DE PROTECCION DE LOS DERECHOS PATRIMONIALES DE AUTOR Y (
##                   "VIOLACION A LOS MECANISMOS DE PROTECCION DE LOS DERECHOS PATRIMONIALES DE AUTOR Y (
##                   "VIOLACION AL REGIMEN LEGAL O CONSTITUCIONAL DE INHABILIDADES E INCOMPATIBILIDADES A
##                   "VIOLACION DE FRONTERAS PARA EXPLOTACION DE RECURSOS NATURALES ART. 123 C.P.",
##                   "VIOLACION DE MEDIDAS SANITARIAS ART. 203",
##                   "VIOLACION REGIMEN LEGAL INHABILIDADES INCOMPATIBILIDADES ART.144 CP.",
##                   "VIOLENCIA INTRAFAMILIAR ART. 22 LEY 294 DE 1996",
##                   "VIOLENCIA INTRAFAMILIAR ART. 229 C.P.",
##                   "VIOLENCIA SEXUAL ENTRE CONYUGES ART. 25 LEY 294 DE 1996")
##
## delito_amarillo <- c("ABORTO ART. 122 C.P.",
##                       "ABORTO ART. 343 C.P.",
##                       "ABORTO SIN CONSENTIMIENTO ART. 123 C.P.",
##                       "ABORTO SIN CONSENTIMIENTO ART. 344 C.P.",
##                       "ABUSO DE AUTORIDAD POR ACTO ARBITRARIO O INJUSTO ART. 152 C.P.",
##                       "ABUSO DE AUTORIDAD POR ACTO ARBITRARIO O INJUSTO ART. 416",
##                       "ABUSO DE AUTORIDAD POR OMISION DE DENUNCIA ART. 153 C.P.",
##                       "ABUSO DE AUTORIDAD POR OMISION DE DENUNCIA ART. 417 C.P.",
```

```
##                            "ABUSO DE CIRCUSTANCIAS DE INFERIORIDAD ART. 360 C.P.",
##                            "ABUSO DE CONDICIONES DE INFERIORIDAD. ART. 251 C.P.",
##                            "ABUSO DE FUNCION PUBLICA ART. 162 C.P.",
##                            "ABUSO DE FUNCION PUBLICA ART. 428 C.P.",
##                            "CARNAL ABUSIVO CON INCAPAZ DE RESISTIR ART. 304 C.P/97",
##                            "ACCESO CARNAL ABUSIVO CON MENOR ART. 303 C.P",
##                            "CARNAL ABUSIVO CON MENOR DE CATORCE ANOS. ART. 208 C.P.",
##                            "ACCESO CARNAL O ACTO SEXUAL ABUSIVOS CON INCAPAZ DE RESISTIR. ART. 210 C.P.",
##                            "ACCESO CARNAL O ACTO SEXUAL EN PERSONA PUESTA EN INCAPACIDAD DE RESISTIR. ART.
##                            "ACCESO CARNAL VIOLENTO ART. 298 C.P.",
##                            "ACCESO CARNAL VIOLENTO. ART. 205 C.P.",
##                            "ACOSO SEXUAL ART. 210 A",
##                            "ACTO SEXUAL EN PERSONA PUESTA EN INCAPACIDAD DE RESISTIR ART. 300 C.P. 97",
##                            "ACTO SEXUAL VIOLENTO ART. 299 C.P.",
##                            "ACTO SEXUAL VIOLENTO. ART. 206 C.P.",
##                            "ACTOS CONTRARIOS A LA DEFENSA DE LA NACION ART. 460 C.P.",
##                            "ACTOS SEXUALES CON MENOR DE 14 ANOS ART. 305 C.P",
##                            "ACTOS SEXUALES CON MENOR DE CATORCE ANOS. ART. 209 C.P.",
##                            "ALTERACION DE RESULTADOS ELECTORALES ART. 394 C.P.",
##                            "ALTERACION DE RESULTADOS ELECTORALES ART.256 C.P.",
##                            "ALTERACION DESFIGURACION Y SUPLANTACION DE MARCAS DE GANADO ART.354 C.P.",
##                            "ALTERACION, DESFIGURACION Y SUPLANTACION DE MARCAS DE GANADO. ART. 243 C.P.",
##                            "ALZAMIENTO DE BIENES ART. 362 C.P.",
##                            "ALZAMIENTO DE BIENES. ART. 253 C.P.",
##                            "AMENAZAS A TESTIGO ART. 454A C.P.",
##                            "APODERAMIENTO DE AERONAVES, NAVES O MEDIOS DE TRANSPORTE COLECTIVO. ART. 173 C
##                            "APODERAMIENTO DE HIDROCARBUROS, SUS DERIVADOS, BIOCOMBUSTIBLES O MEZCLAS QUE L
##                            "APODERAMIENTO Y DESVIO DE AERONAVE ART. 28 DEC. 180 DE 1988",
##                            "CONCIERTO PARA DELINQUIR ART. 186 C.P",
##                            "CONCIERTO PARA DELINQUIR ART. 340 C.P.",
##                            "CONCIERTO PARA DELINQUIR LEY 589/2000",
##                            "CONCUSION ART. 404 C.P.",
##                            "CONCUSION ART.140 C.P.",
##                            "CONSERVACION O FINANCIACION DE PLANTACIONES ART. 375 C.P.",
##                            "CONSTRENIMIENTO A LA PROSTITUCION ART. 214 C.P.",
##                            "CONSTRENIMIENTO A LA PROSTITUCION ART. 309 C.P.",
##                            "CONSTRENIMIENTO AL ELECTOR Y VIOLENCIA Y FRAUDE ELECTORALES ART. 249 Y ART. 25
##                            "CONSTRENIMIENTO AL SUFRAGANTE ART. 387 C.P.",
##                            "CONSTRENIMIENTO ILEGAL ART. 182 C.P.",
##                            "CONSTRENIMIENTO ILEGAL ART. 276 C.P.",
##                            "CONSTRENIMIENTO PARA DELINQUIR ART. 184 C.P.",
##                            "CONSTRENIMIENTO PARA DELINQUIR ART. 277 C.P.",
##                            "CONTAMINACION AMBIENTAL ART. 247 C.P. MOD. ART. 24 LEY 491/99",
##                            "CONTAMINACION AMBIENTAL ART. 332 C.P.",
##                            "CONTAMINACION AMBIENTAL CULPOSA POR EXPLOTACION DE YACIMIENTO MINERO O HIDROCA
##                            "CONTAMINACION AMBIENTAL CULPOSA POR EXPORTACION DE YACIMIENTO MINERO O HIDROCA
##                            "CONTAMINACION DE AGUAS ART. 205 C.P.",
##                            "CONTAMINACION DE AGUAS ART. 371 C.P.",
##                            "CORRUPCION DE ELECTOR ART. 251 C.P.",
##                            "CORRUPCION DE SUFRAGANTE ART. 390 C.P.",
##                            "CULTIVO CONSERVACION FINANCIACION PLANTACIONES ART. 32 LEY 30/86",
##                            "DANO EN BIEN AJENO ART. 370 C.P.",
##                            "DANO EN BIEN AJENO ART. 370 C.P.MOD. ART. 1 N.19 LEY 23 DE 1991",
##                            "DANO EN BIEN AJENO. ART. 265 C.P.",
```

```
##                          "DANO EN LOS RECURSOS NATURALES ART. 246 C.P.",
##                          "DANO EN LOS RECURSOS NATURALES ART. 331 C.P.",
##                          "DANO EN MATERIA PRIMA, PRODUCTO AGROPECUARIO O INDUSTRIAL. ART. 304 C.P.",
##                          "DANO EN OBRAS O ELEMENTOS DE LOS SERVICIOS DE COMUNICACIONES, ENERGIA Y COMBUS
##                          "DANO EN OBRAS O ELEMENTOS DE LOS SERVICIOS DE COMUNICACIONES, ENERGIA Y COMBUS
##                          "DANO O AGRAVIOS A PERSONAS O COSAS DESTINADAS AL CULTO ART. 296 C.P. MOD. ART.
##                          "DESTRUCCION DE BIENES E INSTALACIONES DE CARACTER SANITARIO ART. 155 C.P.",
##                          "DESTRUCCION DEL MEDIO AMBIENTE ART. 164 C.P",
##                          "EXTRADICION",
##                          "FABRICACION Y COMERCIALIZACION DE SUSTANCIAS NOCIVAS PARA LA SALUD ART. 374 C.
##                          "FAVORECIMIENTO DE LA FUGA ART. 179 C.P.",
##                          "FAVORECIMIENTO DE LA FUGA ART. 449 C.P.",
##                          "FAVORECIMIENTO DEL VOTO FRAUDULENTO ART. 253 C.P.",
##                          "FRAUDE AL SUFRAGANTE ART. 388 C.P.",
##                          "FRAUDE EN LA INSCRIPCION DE CEDULAS ART. 389 C.P.",
##                          "FUGA DE PRESOS ART. 178 C.P.",
##                          "FUGA DE PRESOS ART. 448 C.P.",
##                          "HOMICIDIO ART. 323 C.P.",
##                          "HURTO AGRAVADO ART. 11 LEY 228 DE 1995",
##                          "HURTO CALIFICADO ART. 350 C.P.",
##                          "HURTO CALIFICADO. ART. 240 C.P.",
##                          "HURTO DE HIDROCARBUROS O SUS DERIVADOS DTO 1900/02",
##                          "ILICITO APROVECHAMIENTO DE LOS RECURSOS NATURALES RENOVABLES ART. 328 C.P.",
##                          "ILICITO APROVECHAMIENTO DE RECURSOS BIOLOGICOS ART. 242 C.P. MOD. ART. 19 LEY
##                          "IMPEDIMENTO Y PERTURBACION DE CEREMONIA RELIGIOSA. ART. 202 C.P.",
##                          "INCENDIO ART. 189 C.P.",
##                          "INCENDIO ART. 350 C.P.",
##                          "INSTIGACION A DELINQUIR ART. 188 C.P.",
##                          "INSTIGACION A DELINQUIR ART. 348 C.P.",
##                          "INSTIGACION A DELINQUIR LEY 589/2000",
##                          "INTERVENCION EN POLITICA ART. 158 C.P.",
##                          "INTERVENCION EN POLITICA ART. 422 C.P.",
##                          "INVASION DE AEREAS DE ESPECIAL IMPORTANCIA ECOLOGICA ART. 243 C.P. MOD. ART. 2
##                          "INVASION DE AREAS DE ESPECIAL IMPORTANCIA ECOLOGICA ART. 337 C.P.",
##                          "INVASION DE TIERRAS O EDIFICACIONES. ART. 263 C.P.",
##                          "INVASION DE TIERRAS O EDIFICIOS ART. 367 C.P.",
##                          "IRRESPETO A CADAVERES ART. 297",
##                          "IRRESPETO A CADAVERES. ART. 204 C.P.",
##                          "LAVADO DE ACTIVOS ART. 247A CP.",
##                          "LAVADO DE ACTIVOS ART. 323 C.P.",
##                          "LESIONES PERSONALES ART. 331 C.P.",
##                          "MALTRATO CONSTITUTIVO DE LESIONES PERSONALES ART. 23 LEY 294 DE 1996",
##                          "MALTRATO MEDIANTE RESTRICCION A AL LIBERTAD FISICA. ART. 230 C.P.",
##                          "MALTRATO MEDIANTE RESTRICCION A LA LIBERTAD FISICA ART. 230 C.P.",
##                          "MALTRATO MEDIANTE RESTRICCION DE LA LIBERTAD FISICA ART. 24 LEY 294 DE 1996",
##                          "OBSTACULIZACION DE TAREAS SANITARIAS Y HUMANITARIAS ART. 153 C.P.",
##                          "OBSTRUCCION DE OBRAS DE DEFENSA Y ASISTENCIA ART. 364 C.P.",
##                          "OMISION DE ACTO PROPIO EN DELITOS DE EXTORSION O SECUESTRO ART. 33 LEY 40 DE 1
##                          "OMISION DE APOYO ART. 160 C.P.",
##                          "OMISION DE APOYO ART. 424 C.P.",
##                          "OMISION DE CONTROL ART. 247B CP.",
##                          "PERTURBACION DE ACTOS OFICIALES ART. 165 C.P.",
##                          "PERTURBACION DE LA POSESION SOBRE INMUEBLE. ART. 264 C.P.",
##                          "PERTURBACION DE LA POSESION SOBRE INMUEBLES ART. 368 C.P.",
```

```
##                        "PERTURBACION DE LOS SERVICIOS DE COMUNICACION, ENERGIA Y DE COMBUSTIBLES ART. 
##                        "PERTURBACION ELECTORAL 248 C.P.",
##                        "PERTURBACION EN EL SERVICIO DE TRANSPORTE COLECTIVO U OFICIAL ART. 192 C.P.",
##                        "PERTURBACION EN SERVICIO DE TRANSPORTE COLECTIVO U OFICIAL ART. 353 C.P.",
##                        "POSESION DE SUSTANCIAS PARA EL PROCESAMIENTO DE NARCOTICOS ART. 43 LEY 30 DE 1
##                        "PRIVACION ILEGAL DE LA LIBERTAD ART. 174 C.P.",
##                        "PRIVACION ILEGAL DE LA LIBERTAD ART. 272 C.P.",
##                        "PROLONGACION ILICITA DE LA PRIVACION DE LA LIBERTAD ART. 175 C.P.",
##                        "PROLONGACION ILICITA DE LA PRIVACION DE LA LIBERTAD ART. 273 C.P.",
##                        "RECEPTACION ART. 177.- C.P.",
##                        "RECEPTACION ART. 447 C.P.",
##                        "SIMULACION DE INVESTIDURA O CARGO ART. 163 C.P.",
##                        "TENENCIA FABRICACION Y TRAFICO DE SUSTANCIAS U OBJETOS PELIGROSOS ART. 197 C.P
##                        "TENENCIA, FABRICACION Y TRAFICO DE SUSTANCIA U OBJETOS PELIGROSOS ART. 358 C.P
##                        "TRAFICO DE SUSTANCIAS PARA PROCESAMIENTO DE NARCOTICOS ART. 382 C.P.",
##                        "TRAFICO FABRICACION O PORTE DE ESTUPEFACIENTES ART. 33 LEY 30/86",
##                        "TRAFICO FABRICACION O PORTE DE ESTUPEFACIENTES ART. 376 C..",
##                        "TRAFICO, TRANSPORTE Y POSESION DE MATERIALES RADIOACTIVOS O SUSTANCIAS NUCLEAR
##                        "ULTRAJE A EMBLEMAS Y SIMBOLOS PATRIOS ART. 461 C.P.",
##                        "USO DE DOCUMENTO FALSO. ART. 291 C.P.",
##                        "USO DE DOCUMENTO PUBLICO FALSO ART. 222 C.P.",
##                        "USURPACION DE FUNCIONES PUBLICAS ART. 425 C.P.",
##                        "USURPACION DE TIERRAS ART. 365 C.P.",
##                        "USURPACION DE TIERRAS. ART. 261 C.P.",
##                        "VIOLACION DE HABITACION AJENA POR SERVIDOR PUBLICO ART. 190 C.P.",
##                        "VIOLACION DE LA LIBERTAD DE TRABAJO ART. 290 C.P.",
##                        "VIOLACION DE LA LIBERTAD DE TRABAJO. ART. 198 C.P.",
##                        "VIOLACION ILICITA DE COMUNICACIONES ART. 288 C.P.",
##                        "VIOLACION ILICITA DE COMUNICACIONES O CORRESPONDENCIA DE CARACTER OFICIAL ART.
##                        "VIOLACION ILICITA DE COMUNICACIONES O CORRESPONDENCIA DE CARACTER OFICIAL. ART
##                        "VIOLACION ILICITA DE COMUNICACIONES. ART. 192 C.P.",
##                        "VIOLENCIA CONTRA EMPLEADO OFICIAL ART. 164 C.P.",
##                        "VIOLENCIA CONTRA SERVIDOR PUBLICO ART. 429 C.P.")
##
##
## delito_na <- c("ABORTO ART. 122 C.P.",
##                "ABORTO SIN CONSENTIMIENTO ART. 123 C.P.",
##                "ABUSO DE AUTORIDAD POR OMISION DE DENUNCIA ART. 417 C.P.",
##                "ACCESO CARNAL ABUSIVO CON MENOR DE CATORCE ANOS. ART. 208 C.P.",
##                "ACCESO CARNAL O ACTO SEXUAL ABUSIVOS CON INCAPAZ DE RESISTIR. ART. 210 C.P.",
##                "ACCESO CARNAL O ACTO SEXUAL EN PERSONA PUESTA EN INCAPACIDAD DE RESISTIR. ART. 207 C
##                "ACCESO CARNAL VIOLENTO. ART. 205 C.P.",
##                "ACTOS SEXUALES CON MENOR DE CATORCE ANOS. ART. 209 C.P.",
##                "ALTERACION DE RESULTADOS ELECTORALES ART. 394 C.P.",
##                "ALTERACION, DESFIGURACION Y SUPLANTACION DE MARCAS DE GANADO. ART. 243 C.P.",
##                "ALZAMIENTO DE BIENES. ART. 253 C.P.",
##                "CONCUSION ART. 404 C.P.",
##                "CONSERVACION O FINANCIACION DE PLANTACIONES ART. 375 C.P.",
##                "CONTRABANDO DE HIDROCARBUROS O SUS DERIVADOS DTO 1900/02",
##                "DANO EN BIEN AJENO. ART. 265 C.P.",
##                "DANO EN MATERIA PRIMA, PRODUCTO AGROPECUARIO O INDUSTRIAL. ART. 304 C.P.",
##                "DESTINACION ILICITA DE MUEBLES O INMUEBLES ART. 377 C.P.",
##                "DISPARO DE ARMA DE FUEGO CONTRA VEHICULO ART. 356 C.P.",
##                "DISPARO DE ARMA DE FUEGO CONTRA VEHICULO ART.195 C.P.",
```

```
##                 "DIVULGACION Y EMPLEO DE DOCUMENTOS RESERVADOS. ART. 194 C.P.",
##                 "DIVULGACION Y EMPLEO DE DOCUMENTOS RESERVADOS ART. 289 C.P.",
##                 "EXPLOTACION ILICITA DE YACIMIENTO MINERO Y OTROS MATERIALES ART. 338 C.P.",
##                 "FABRICACION Y COMERCIALIZACION DE SUSTANCIAS NOCIVAS PARA LA SALUD ART. 374 C.P.",
##                 "FAVORECIMIENTO CONTRABANDO DE HIDROCARBUROS - DERIVADOS DTO 1900/02",
##                 "FRAUDE AL SUFRAGANTE ART. 388 C.P.",
##                 "FRAUDE EN LA INSCRIPCION DE CEDULAS ART. 389 C.P.",
##                 "FUGA DE PRESOS ART. 178 C.P.",
##                 "FUGA DE PRESOS ART. 448 C.P.",
##                 "HOMICIDIO ART. 103 C.P.",
##                 "HURTO CALIFICADO. ART. 240 C.P.",
##                 "HURTO CALIFICADO ART. 350 C.P.",
##                 "HURTO. ART. 239 C.P.",
##                 "ILICITO APROVECHAMIENTO DE LOS RECURSOS NATURALES RENOVABLES ART. 328 C.P.",
##                 "INCENDIO ART. 350 C.P.",
##                 "INCENDIO ART. 189 C.P.",
##                 "INSTIGACION A DELINQUIR ART. 348 C.P.",
##                 "INSTIGACION A DELINQUIR ART. 188 C.P.",
##                 "INTERVENCION EN POLITICA ART. 158 C.P.",
##                 "INTERVENCION EN POLITICA ART. 422 C.P.",
##                 "INVASION DE AREAS DE ESPECIAL IMPORTANCIA ECOLOGICA ART. 337 C.P.",
##                 "INVASION DE TIERRAS O EDIFICACIONES. ART. 263 C.P.",
##                 "INVASION DE TIERRAS O EDIFICIOS ART. 367 C.P.",
##                 "IRRESPETO A CADAVERES ART. 297",
##                 "IRRESPETO A CADAVERES. ART. 204 C.P.",
##                 "LAVADO DE ACTIVOS ART. 323 C.P.",
##                 "LAVADO DE ACTIVOS ART. 247A CP",
##                 "MALTRATO MEDIANTE RESTRICCION A AL LIBERTAD FISICA. ART. 230 C.P.",
##                 "MALTRATO MEDIANTE RESTRICCION A LA LIBERTAD FISICA ART. 230 C.P.",
##                 "MALTRATO MEDIANTE RESTRICCION DE LA LIBERTAD FISICA ART. 24 LEY 294 DE 1996",
##                 "OBSTRUCCION DE OBRAS DE DEFENSA Y ASISTENCIA ART. 364 C.P.",
##                 "OMISION DE APOYO ART. 424 C.P.",
##                 "OMISION DE APOYO ART. 160 C.P.",
##                 "OMISION DE SOCORRO ART. 131 C.P.",
##                 "PERFIDIA ART. 143 C.P.",
##                 "PERTURBACION DE ACTOS OFICIALES ART. 430 C.P.",
##                 "PERTURBACION DE ACTOS OFICIALES ART. 165 C.P.",
##                 "PERTURBACION DE CERTAMEN DEMOCRATICO ART. 386 C.P.",
##                 "PERTURBACION DE LA POSESION SOBRE INMUEBLE. ART. 264 C.P.",
##                 "PERTURBACION DE LA POSESION SOBRE INMUEBLES ART. 368 C.P.",
##                 "PERTURBACION EN SERVICIO DE TRANSPORTE COLECTIVO U OFICIAL ART. 353 C.P.",
##                 "PERTURBACION EN EL SERVICIO DE TRANSPORTE COLECTIVO U OFICIAL ART. 192 C.P.",
##                 "RECEPTACION ART. 447 C.P.",
##                 "RECEPTACION ART. 177.- C.P.",
##                 "REVELACION DE SECRETO ART. 418 C.P",
##                 "REVELACION DE SECRETO ART. 154 C.P.",
##                 "SIMULACION DE INVESTIDURA O CARGO ART. 426 C.P.",
##                 "SIMULACION DE INVESTIDURA O CARGO ART. 163 C.P.",
##                 "SINIESTRO O DANO DE NAVE ART. 354 C.P.",
##                 "SINIESTRO O DANO DE NAVE ART. 193 C.P.",
##                 "TENENCIA FABRICACION Y TRAFICO DE SUSTANCIAS U OBJETOS PELIGROSOS ART. 197 C.P.",
##                 "TENENCIA, FABRICACION Y TRAFICO DE SUSTANCIA U OBJETOS PELIGROSOS ART. 358 C.P.",
##                 "TRAFICO DE SUSTANCIAS PARA PROCESAMIENTO DE NARCOTICOS ART. 382 C.P.",
##                 "TRAFICO FABRICACION O PORTE DE ESTUPEFACIENTES ART. 376 C..",
```

```
##                  "TRAFICO FABRICACION O PORTE DE ESTUPEFACIENTES ART. 33 LEY 30/86",
##                  "ULTRAJE A EMBLEMAS Y SIMBOLOS PATRIOS ART. 461 C.P.",
##                  "USO DE DOCUMENTO FALSO. ART. 291 C.P.",
##                  "USO DE DOCUMENTO PUBLICO FALSO ART. 222 C.P.",
##                  "USURPACION DE FUNCIONES PUBLICAS ART. 425 C.P.",
##                  "UTILIZACION ILICITA DE EQUIPOS TRANSMISORES O RECEPTORES ART. 197 C.P.",
##                  "UTILIZACION ILICITA DE EQUIPOS TRANSMISORES O RECEPTOR ART. 16 DEC. 180 DE 1988",
##                  "UTILIZACION ILICITA DE EQUIPOS TRANSMISORES O RECEPTORES. ART. 197 C.P.",
##                  "UTILIZACION INDEBIDA DE INFORMACION PRIVILEGIADA. ART. 258 C.P.")
##
## prof_verde <- c("ACTIVIDADES RELACIONADAS CON EL SINDICALISMO",
##                  "ACTIVIDADES RELACIONADAS ORG. CIVICAS Y CAMPESINAS", "ACTIVISTA DERECHOS HUMANOS",
##                  "ACTIVISTA SINDICAL", "ALCALDE", "ALMIRANTE (FUERZA PUBLICA)",
##                  "DESMOVILIZADO", "DESPLAZADO",
##                  "DIRIGENTE ORGANIZACION DERECHOS HUMANOS", "DIRIGENTE SINDICAL", "INFANTE DE MARINA"
##                  "INTEGRANTE AUTODEFENSAS", "INTEGRANTE GUERRILLA",
##                  "JUEZ PENAL CIRCUITO", "JUEZ PENAL MUNICIPAL",
##                  "JUEZ PROMISCUO",  "LIDER ORGANIZACION CAMPESINA", "LIDER ORGANIZACION COMUNITARIA",
##                  "LIDER ORGANIZACION INDIGENA", "LIDER ORGANIZACION POLITICA",  "MARINERO ARMADA NACIO
##                  "MIEMBRO, AFILIADO, ACTIVISTA DE LA UNION PATRIOTICA", "SECRETARIO DE SINDICATO" )
##
## prof_azul <- c("AEROTECNICO SUBJEFE FUERZA AEREA", "AGENTE POLICIA NACIONAL",
##                 "BRIGADIER GENERAL DEL EJERCITO NACIONAL", "BRIGADIER GENERAL INFANTERIA DE MARINA",
##                 "BRIGADIER GENERAL POLICIA NACIONAL", "CABO PRIMERO DEL EJERCITO NACIONAL",
##                 "CABO PRIMERO POLICIA NACIONAL", "CABO SEGUNDO DEL EJERCITO NACIONAL",
##                 "CABO SEGUNDO POLICIA NACIONAL", "CABO TERCERO DEL EJERCITO NACIONAL",
##                 "CAPITAN DE CORBETA ARMADA NACIONAL", "CAPITAN DE FRAGATA ARMADA NACIONAL",
##                 "CAPITAN DE LA FUERZA AEREA", "CAPITAN DE LA MARINA",
##                 "CAPITAN DE NAVIO ARMADA NACIONAL", "CAPITAN DEL EJERCITO NACIONAL",
##                 "CAPITAN POLICIA NACIONAL", "CONCEJAL", "CONTRA ALMIRANTE ARMADA NACIONAL",
##                 "CONTRALOR MUNICIPAL", "CORONEL DEL EJERCITO NACIONAL",
##                 "CORONEL INFANTERIA DE MARINA", "CORONEL POLICIA NACIONAL",
##                 "DEFENSOR DEL PUEBLO", "DEFENSOR DEL PUEBLO DELEGADO", "DIPUTADO",
##                 "EMPLEADOS AREA DE FISCALIA", "EMPLEADOS DEL AREA DEL CTI DE LA FISCALIA",
##                 "ENFERMERAS", "ENFERMERAS JEFES",  "ESTUDIANTES UNIVERSITARIOS", "EXALCALDE",
##                 "FISCAL DELEG. TRIBUNAL SUP.", "FISCAL ESPECIALIZADO", "FISCAL LOCAL",
##                 "FISCAL SECCIONAL", "GANADERO",
##                 "GENERAL DE LA FUERZA AEREA", "GENERAL DE LA POLICIA NACIONAL",
##                 "GENERAL DE LA REPUBLICA","GENERAL DE LA REPUBLICA (FUERZA PUBLICA)",
##                 "GENERAL DEL EJERCITO NACIONAL", "GENERAL INFANTERIA DE MARINA",
##                 "GOBERNADOR",  "GOBERNADOR DEPARTAMENTO",  "INDIGENA",  "INDIGENTE",
##                 "INGENIEROS DE PETROLEOS",  "INTENDENTE POLICIA NACIONAL", "INVESTIGADOR JUDICIAL C.T
##                 "INVESTIGADOR JUDICIAL D.A.S.", "INVESTIGADOR JUDICIAL POLICIA NACIONAL",
##                 "MAGISTRADO CONSEJO DE ESTADO",  "MAGISTRADO CONSEJO SECC. JUDICATURA",
##                 "MAGISTRADO CONSEJO SUP. JUDICATURA",  "MAGISTRADO TRIBUNAL ADMINISTRATIVO",
##                 "MAGISTRADO TRIBUNAL SUPERIOR", "MAGISTRADOS DE TRIBUNALES", "MAYOR DE LA FUERZA AERE
##                 "MAYOR DEL EJERCITO NACIONAL", "MAYOR GENERAL DEL EJERCITO NACIONAL",
##                 "MAYOR GENERAL POLICIA NACIONAL", "MAYOR INFANTERIA DE MARINA",
##                 "MAYOR POLICIA NACIONAL",  "MIEMBROS CIVILES ARMADA NACIONAL",
##                 "MIEMBROS CIVILES DE LA MARINA",  "MIEMBROS CIVILES EJERCITO NACIONAL",
##                 "MIEMBROS CIVILES POLICIA NACIONAL",  "MIEMBROS FUERZAS MILITARES Y POLICIAS",
##                 "MIEMBROS RETIRADOS ARMADA NACIONAL", "MIEMBROS RETIRADOS EJERCITO NACIONAL",
##                 "MIEMBROS RETIRADOS FUERZA AEREA  MIEMBROS RETIRADOS FUERZAS MILITARES Y DE POLICIA",
##                 "MIEMBROS RETIRADOS POLICIA NACIONAL", "OCUPACIONES EXCLUSIVAS DE LAS FUERZAS MILITAR
```

```
##                 "OFICIALES DE CUBIERTA", "OFICIALES DE LAS FUERZAS MILITARES", "OFICIALES DE POLICIA"
##                 "OTROS MIEMBROS ARMADA NACIONAL", "OTROS MIEMBROS DE LA MARINA", "OTROS MIEMBROS EJERC
##                 "OTROS MIEMBROS FUERZA AEREA", "OTROS MIEMBROS POLICIA NACIONAL",
##                 "PRESIDENTE DE SINDICATO", "PROCURADOR DELEGADO", "PROFESORES DE EDUCACION BASICA",
##                 "PROFESORES UNIVERSITARIOS", "PROMOTORES DE SALUD", "REINSERTADO",
##                 "REPRESENTANTE A LA CAMARA", "SACERDOTE", "SARGENTO MAYOR DEL EJERCITO NACIONAL",
##                 "SARGENTO MAYOR POLICIA NACIONAL", "SARGENTO PRIMERO DEL EJERCITO NACIONAL",
##                 "SARGENTO PRIMERO POLICIA NACIONAL", "SARGENTO SEGUNDO DEL EJERCITO NACIONAL",
##                 "SARGENTO SEGUNDO POLICIA NACIONAL", "SARGENTO VICE PRIMERO POLICIA NACIONAL",
##                 "SARGENTO VICEPRIMERO DEL EJERCITO NACIONAL", "SARGENTO VICEPRIMERO INFANTERIA DE AVIA
##                 "SENADOR DE LA REPUBLICA", "SOLDADO EJERCITO NACIONAL",  "SOLDADO FUERZA AEREA",
##                 "SUB COMISARIO POLICIA NACIONAL", "SUB INTENDENTE POLICIA NACIONAL",
##                 "SUB OFICIAL DE LA MARINA", "SUB OFICIAL JEFE ARMADA NACIONAL",
##                 "SUB OFICIAL PRIMERO ARMADA NACIONAL", "SUB OFICIAL SEGUNDO ARMADA NACIONAL",
##                 "SUB TENIENTE POLICIA NACIONAL", "SUBOFICIALES DE LA POLICIA",
##                 "SUBOFICIALES DE LAS FUERZAS MILITARES", "SUBTENIENTE DEL EJERCITO NACIONAL", "TECNICO
##                 "TENIENTE CORONEL DE LA FUERZA AEREA", "TENIENTE CORONEL DEL EJERCITO NACIONAL",
##                 "TENIENTE CORONEL INFANTERIA DE MARINA", "TENIENTE CORONEL POLICIA NACIONAL",
##                 "TENIENTE DE CORBETA ARMADA NACIONAL", "TENIENTE DE FRAGATA ARMADA NACIONAL",
##                 "TENIENTE DE LA FUERZA AEREA", "TENIENTE DE LA MARINA",
##                 "TENIENTE DE NAVIO ARMADA NACIONAL", "TENIENTE DEL EJERCITO NACIONAL",
##                 "TENIENTE POLICIA NACIONAL", "TRABAJADORES EN SERVICIO SECIAL Y COMUNITARIO", "TRABAJA
##
## prof_amarillo <- c("AEROTECNICO JEFE FUERZA AEREA", "AEROTECNICO PRIMERO FUERZA AEREA",
##                     "AGRICULTOR", "AGRONOMOS Y ESPECIALISTAS AGRICOLAS", "ALBANIL", "ARTESANOS",
##                     "ASISTENTES ADMINISTRATIVOS", "ASISTENTES DE CONTABILIDAD",
##                     "ASISTENTES DE JUZGADOS, TRIBUNALES",
##                     "AUXILIARES DE ENFERMERIA", "AUXILIARES DE ODONTOLOGIA",
##                     "AUXILIARES DE TRIBUNALES, JUZGADOS",
##                     "BIOLOGOS, BOTANICOS, ZOOLOGOS Y RELACIONADOS",
##                     "BRIGADIER GENERAL DE LA FUERZA AEREA", "CAJEROS",
##                     "CAJEROS DE SERVICIOS FINANCIEROS", "CARNICEROS", "COMERCIANTE",
##                     "COMERCIANTE DE GANADO", "COMISARIOS, INSPECTORES DE POLICIA",
##                     "CONDUCTOR O AUXILIAR DE TRANSPORTE", "CONTRALOR DEPARTAMENTA",
##                     "COTERO", "DESEMPLEADO",
##                     "DIRECTORES Y ADMINISTRADORES DE EDUCACION BASICA Y DIRECTORES Y GERENTES GENERALE
##                     "DIRECTORES Y GERENTES GENERALES PRODUCCION DE BIENES", "DIRECTORES Y GERENTES GEN
##                     "EBANISTA", "INVESTIGADORES Y ANALISTAS DE POLITICA", "EDILES",
##                     "ELECTRICISTA", "EMBAJADOR",  "EMPLEADOS DE BANCA, SEGUROS Y OTROS SERVICIOS FINAN
##                     "ESTUDIANTE PRIMARIA", "ESTUDIANTES SECUNDARIA", "FARMACEUTICOS",
##                     "FOTOGRAFOS", "FUNCIONARIOS DE PROGRAMAS EXCLUSIVOS DE LA ADMINIS ",
##                     "GEOLOGOS, GEOQUIMICOS Y GEOFISICOS ", "GERENTES DE ARQUITECTURA Y CIENCIAS",
##                     "GUARDIANES DE PRISION", "INGENIERO DE VIAS Y TRANSPORTE", "INGENIERO SANITARIO Y
##                     "INGENIEROS CIVILES", "INGENIEROS DE MINAS",  "INGENIEROS DE SISTEMAS",
##                     "INGENIEROS ELECTRICOS Y ELECTRONICOS", "INGENIEROS INDUSTRIALES",
##                     "INGENIEROS MECANICOS", "INGENIEROS METALURGICOS",  "NGENIEROS QUIMICOS",
##                     "INSPECTOR DE CONSTRUCCION",  "INSPECTORES DE SANIDAD, SEGURIDAD Y SALUD OCUPACION
##                     "INSTRUMENTADORES QUIRURGICOS",  "INVESTIGADORES Y CONSULTORES, DESARROLLO ECONOMI
##                     "JEFE DE MISION DIPLOMATICA",  "JEFES DE BIBLIOTECA, PUBLICACIONES Y EMPLEADOS DE
##                     "JEFES DE OFICINA EN GENERAL",  "JEFES DE REGISTRO, DISTRIBUCION Y PROGRAMACION",
##                     "JUEZ CIVIL CIRCUITO", "JUEZ CIVIL MUNICIPAL",  "JUEZ DE FAMILIA", "JUEZ DE MENORE
##                     "JUEZ DE PAZ", "JUEZ LABORAL",  "MATEMATICOS, ESTADISTICOS Y ACTUARIOS",
##                     "MECANICO AUTOMOTRIZ",  "MEDICOS GENERALES", "MERCADERISTAS", "MEDICOS ESPECIALIST
##                     "MIEMBROS DEL PODER EJECUTIVO LEGISLATIVO Y JUDICIAL", "MINERO",
```

```
##                      "MINISTRO DEL DESPACHO", "MOTOTAXISTA",  "NOTARIO PUBLICO", "OBRERO",
##                      "OCUPACIONES PROFESIONALES EN ORGANIZACION Y ADMINI", "ODONTOLOGOS",
##                      "ORIENTADORES EDUCATIVOS", "OTRAS ACTIVIDADES DE LA CONSTRUCCION", "OTRAS ACTIVID
##                      "OTRAS OCUPACIONES RELIGIOSAS", "OTRO", "OTROS ARTISTAS", "OTROS ESTUDIANTES",
##                      "OTROS INVESTIGADORES JUDICIALES", "OTROS JUECES", "OTROS SUPERVISORES", "PANADER
##                      "PELUQUEROS, ESTILISTAS Y AFINES",  "PENSIONADO",  "PERSONAL DIRECTIVO DE LA ADMI
##                      "PATRULLERO POLICIA NACIONAL",  "PERIODISTAS",  "PERSONERO",  "POLICIAS",  "PRESI
##                      "PROCURADOR DELEGADO", "PROFESORES DE EDUCACION BASICA", "POLITOLOGO",
##                      "PROFESORES DE EDUCACION MEDIA", "PROFESORES DE PREESCOLAR", "PSICOLOGOS",
##                      "RECICLADOR", "REGISTRADOR NACIONAL ESTADO CIVIL", "SECRETARIAS", "SOLDADOR",
##                      "SUPERVISORES DE VIGILANCIA", "TECNOLOGOS Y TECNICOS AGROPECUARIOS", "TERAPEUTAS
##                      "TRABAJADORA SEXUAL", "VICE ALMIRANTE ARMADA NACIONAL", "VIGILANTES Y GUARDIANES
##
## prof_naranja <- c("ABOGADO", "ADMINISTRADOR DE EMPRESAS", "ADMINISTRADORES DE COMERCIO AL POR MENOR"
##                     "ADMINISTRADORES DE EDUCACION SUPERIOR Y FORMACION",
##                     "ADMINISTRADORES DE INMUEBLES", "ADMINISTRADORES Y OPERADORES DE SISTEMAS",
##                     "ANALISTAS DE SISTEMAS", "ANUNCIADORES Y LOCUTORES",
##                     "ARBITROS", "ARQUITECTOS", "ASEADORES Y SERVICIO DOMESTICO",
##                     "ASISTENTES DE PERSONAL Y SELECCION", "ADMINISTRADOR DE EMPRESAS",
##                     "ADMINISTRADORES DE COMERCIO AL POR MENOR  ADMINISTRADORES DE EDUCACION SUPERIOR Y
##                     "ADMINISTRADORES DE INMUEBLES", "ADMINISTRADORES Y OPERADORES DE SISTEMAS",
##                     "ANALISTAS DE SISTEMAS", "ANUNCIADORES Y LOCUTORES", "ASEADORES Y SERVICIO DOMESTI
##                     "ASISTENTES DE PERSONAL Y SELECCION", "AUXILIARES ADMINISTRATIVOS",
##                     "AUXILIARES DE ALMACEN Y BODEGA", "AUXILIARES DE CONTABILIDAD", "AUXILIARES DE OFI
##                     "AUXILIARES DE PERSONAL Y NOMINA", "AUXILIARES DE SERVICIO A PASAJEROS",
##                     "AYUDANTES DE COCINA Y CAFETERIA", "CARTEROS Y MENSAJEROS", "CHEFS", "COCINERO",
##                     "COMERCIANTE DE ESMERALDA", "COMPRADORES", "DEPORTISTA",
##                     "DIRECTOR DEPARTAMENTO ADMINISTRATIVO",
##                     "DIRECTORES DE FISCALIAS DIRECTORES DE PROGRAMAS DE ESPARCIMIENTO Y DEPORT",
##                     "DIRECTORES Y GERENTES SERVICIOS FINANCIEROS",
##                     "DISENADORES DE TEATRO, MODA Y EXHIBICION Y OTROS C",
##                     "DISENADORES GRAFICOS Y DIBUJANTES ARTISTICOS", "DISENADORES INDUSTRIALES",
##                     "DOMESTICADORES Y TRABAJADORES DEL CUIDADO DE ANIMA",
##                     "ECONOMISTAS", "EMPLEADOS DE INFORMACION Y SERVICIO AL CLIENTE",
##                     "EMPLEADOS DE PUBLICACION Y AFINES", "EMPLEADOS DE VENTAS Y SERVICIOS DE AREOLINEA
##                     "FISICOS Y ASTRONOMOS", "FISIOTERAPEUTAS",
##                     "GERENTES DE COMERCIO AL POR MENOR", "GERENTES DE EMPRESAS DE TELECOMUNICACIONES",
##                     "GERENTES DE MEDIOS DE COMUNICACION Y ARTES ESCENIC", "GERENTES DE OTROS SERVICIOS
##                     "GERENTES DE OTROS SERVICIOS A LAS EMPRESAS", "GERENTES DE OTROS SERVICIOS ADMINIS
##                     "GERENTES DE PROGRAMAS DE POLITICA DE DESARROLLO EC GERENTES DE SEGUROS, BIENES RA
##                     "GERENTES DE SERVICIOS DE COMERCIO EXTERIOR", "GERENTES DE VENTAS, MERCADEO Y PUBL
##                     "GERENTES FINANCIEROS", "GUIAS DE VIAJES Y TURISMO", "HIGIENISTAS DENTALES",
##                     "MPULSADORES Y DEMOSTRADORES", "LAVACARROS", "MECANOGRAFOS", "MESEROS", "MUSICOS Y
##                     "OFICIALES DE BOMBEROS", "ORNAMENTADOR", "OTROS INGENIEROS", "OTROS INSPECTORES",
##                     "OTROS INSTRUCTORES", "PARTICULAR", "PATRONISTAS PRODUCTOS TEXTILES, CUERO Y PIEL"
##                     "PILOTOS, INGENIEROS E INSTRUCTORES DE VUELO", "PINTORES, ESCULTORES Y OTROS ARTIS
##                     "PLOMERO", "PRACTICANTES DE LA MEDICINA ALTERNATIVA", "PROGRAMADORES DE SISTEMAS",
##                     "QUIMICOS", "RECEPCIONISTAS Y OPERADORES DE CONMUTADOR", "RECREACIONISTAS",
##                     "REPRESENTANTES DE VENTAS NO TECNICAS", "SUPERINTENDENTE DE SALUD", "SUPERINTENDEN
##                     "SUPERINTENDENTE Y GERENTE DE ESTABLECIMIENTOS PUBL", "SUPERVISORES DE VENTAS",
##                     "TESORERO", "TOPOGRAFOS", "VENDEDORES DE MOSTRADOR", "VENDEDORES, VENTAS TECNICAS"
##
## prof_rojo <- c("ACTIVIADES RELACIONADAS CON LA ZAPATERIA", "ACTIVIDADES DE LATONERIA Y PINTURA AUTOM
##                 "ACTIVIDADES DE PESCA", "ACTIVIDADES RELACIONADAS CON EL HOGAR",
```

```
##                 "ACTIVIDADES RELACIONADAS CON LA CARPINTERIA", "ACTIVIDADES RELACIONADAS CON LA TAURO
##                 "AGENTES DE ADUANA Y OTROS AGENTES", "BARMAN", "BOMBEROS",
##                 "CONSEJEROS DE EMPLEO", "CONTADORES Y AUDITORES FINANCIEROS",
##                 "CONTRALOR GENERAL DE LA REPUBLICA", "CONTROLADORES DE TRAFICO AEREO", "DIGITADORES",
##                 "OPERADORES DE JUEGOS MECANICOS Y DE SALON", "ORGANIZADORES DE EVENTOS",
##                 "OTRAS OCUPACIONES DE SERVICIOS PERSONALES", "OTRAS OCUPACIONES ELEMENTALES DE LAS VE
##                 "OTRAS OCUPACIONES ELEMENTALES DE LOS SERVICIOS OTRAS OCUPACIONES PROFESIONALES EN TE
##                 "PRESIDENTE DE LA REPUBLICA",
##                 "TECNICOS DENTALES",
##                 "TECNICOS EN GRABACION DE AUDIO Y VIDEO", "TECNICOS EN TRANSMISION",
##                 "TECNICOS OPTICOS",
##                 "TECNICOS Y MECANICOS DE INSTRUMENTOS DE AERONAVEGA  TECNICOS Y MECANICOS DE INSTRUME
##                 "TECNOLOGOS Y TECNICOS DE LABORATORIO MEDICO Y PATO",
##                 "TECNOLOGOS Y TECNICOS EN ARQUITECTURA", "TECNOLOGOS Y TECNICOS EN CIENCIAS BIOLGICAS
##                 "TECNOLOGOS Y TECNICOS EN ELECTROCARDIOGRAFIA Y ELE", "TECNOLOGOS Y TECNICOS EN GEOLO
##                 "TECNOLOGOS Y TECNICOS EN INGENIERA CIVIL TECNOLOGOS Y TECNICOS EN INGENIERA ELECTRIC
##                 "TECNOLOGOS Y TECNICOS EN QUIMICA APLICADA TECNOLOGOS Y TECNICOS FORESTALES Y DE RECU
##                 "TRABAJADORES DE ESTACION DE SERVICIO")
##
## categoria_ca <- c("PARAPOLITICA" ,"FARC-EP","VICTIMAS PERIODISTAS",
##                    "PRESUNTOS HOMICIDIOS AGENTES DEL ESTADO","CONFLICTO ARMADO","CASOS UP" ,
##                    "HOMICIDIO PRESENTANDO COMO BAJA POR FUERZA PUBLICA",
##                    "CASOS EN EL SIDH","VIXCTIMAS DEFENSORES DE DDHH",
##                    "VINCULO FUNC. PUBLICOS CON GRUPOS ILEGALES","MASACRES",
##                    "CASOS DEMANDADOS CIDH","CASOS SAN JOSE DE APARTADO",
##                    "RESTITUCION DE TIERRAS",
##                    "VICTIMAS INDIGENAS","CASOS AUTO 92 CORTE CONSTITUCIONAL",
##                    "BANDAS EMERGENTES" ,
##                    "RELACION DE CIVILES CON ACTORES DE CONFLICTO",
##                    "COMPULSAS - JUSTICIA Y PAZ","VICTIMAS SINDICALISTAS" ,
##                    "ELN","CASOS OIT",
##                    "VICTIMAS DE MINAS ANTIPERSONALES","VIOLENCIA SEXUAL CONFLICTO ARMADO",
##                    "VICTIMAS MIEMBROS DE LAS ONG'S","MARCHA PATRIOTICA")
##
## one <- as.integer(1)
## zero <- as.integer(0)
##
## conflict_sijuf_jep <- data_sijuf_jep_fase4 %>%
##   mutate(prob = runif(nrow(.))) %>%
##   mutate(is_conflict = case_when(
##     a29_tipo_persona == "PERPETRADOR" ~ zero,  # irrelevant, will be filtered
##     perpetrador1 %in% perp_no_ca ~ zero,
##     perpetrador2 %in% perp_no_ca ~ zero,
##     perpetrador3 %in% perp_no_ca ~ zero,
##     a67_delito_unico %in% delito_rojo ~ zero,
##     a63_delito %in% delito_rojo ~ zero,
##     a23_delito_ppal %in% delito_rojo ~ zero,
##     perpetrador1 %in% perp_verde ~ one,
##     perpetrador2 %in% perp_verde ~ one,
##     perpetrador3 %in% perp_verde ~ one,
##     a63_delito %in% delito_verde ~ one,
##     a67_delito_unico %in% delito_verde ~ one,
##     a24_categoria_caso %in% categoria_ca ~ one,
##     a44_categoria %in% categoria_ca ~ one,
```

```
##      a23_delito_ppal %in% delito_verde ~ one,
##      perpetrador1 %in% perp_is_ca &  a29_tipo_persona == "VICTIMA" ~ one,
##      perpetrador2 %in% perp_is_ca & a29_tipo_persona == "VICTIMA" ~ one,
##      perpetrador3 %in% perp_is_ca & a29_tipo_persona == "VICTIMA" ~ one,
##      perpetrador1 %in% perp_naranja & prob < 0.75 ~ zero,
##      perpetrador2 %in% perp_naranja & prob < 0.75 ~ zero,
##      perpetrador3 %in% perp_naranja & prob < 0.75 ~ zero,
##      a23_delito_ppal %in% delito_naranja & prob < 0.75 ~ zero,
##      a67_delito_unico %in% delito_naranja & prob < 0.75 ~ zero,
##      a63_delito %in% delito_naranja & prob < 0.75 ~ zero,
##      perpetrador1 %in% perp_azul & prob < 0.75 ~ one,
##      perpetrador2 %in% perp_azul & prob < 0.75 ~ one,
##      perpetrador3 %in% perp_azul & prob < 0.75 ~ one,
##      a23_delito_ppal %in% delito_azul & prob < 0.75 ~ one,
##      a23_delito_ppal %in% delito_amarillo & prob < 0.5 ~ one,
##      a63_delito %in% delito_azul & prob < 0.75 ~ one,
##      a63_delito %in% delito_amarillo & prob < 0.5 ~ one,
##      a67_delito_unico %in% delito_azul & prob < 0.75 ~ one,
##      a67_delito_unico %in% delito_amarillo & prob < 0.5 ~ one,
##      a63_delito %in% delito_na ~ NA_integer_,
##      a29_tipo_persona == "OTRA" ~ NA_integer_,
##      a67_delito_unico %in% delito_na ~ NA_integer_,
##      a23_delito_ppal %in% delito_na ~ NA_integer_,
##      TRUE ~ NA_integer_)) %>%
##    select(-prob)
##
## sample <- conflict_sijuf_jep  %>%
##    filter(is_conflict ==0) %>%
##    select(recordid, is_conflict)
##
## # ------- training data
##
## sample <- sample_frac(sample, 0.2)
##
## write.table(sample, file = args$examples, sep = "|", quote = FALSE, row.names = FALSE)
##
## write_parquet(conflict_sijuf_jep, args$output)
```

```
cat(readLines(files$is_ca_spoa_jep_fase3), sep = '\n')
```

9.4.1.6 FGN - SPOA - JEP FASE 3

```
## #
## # Authors:     PA, VG
## # Maintainers  VG, PA, PB, JGD
## # Copyright   2022, HRDAG,
## # ========================
## # CO-SIVJRNR-data/individual/FGN/is-ca/src/is-ca-spoa_jep_fase3.R
##
## # following the email from Folco Zaffalon (8 abr 2022, 17:16 PDT),
## # and the accompanying file
```

```
## # ‘SIJUF_casos_perpetrador_ocupacion victima_delitos.docx‘
## #
## # Verde = SI CONFLICTO (1)
## # Azul = MUY PROBABLE 0.75~(is_conflict == 1)
## # Amarillo = NO SE SABE 0.5~(is_conflict == 1)
## # Naranja = POCO PROBABLE 0.75~(is_conflict == 0)
## # Rojo = NO CONFLICTO (0)
##
## # ------ setup
## Sys.setlocale("LC_CTYPE", "en_US.UTF-8")
## set.seed(19481210)   # <-- the seed is really important in this run!
##
## require(pacman)
## p_load(argparse, dplyr, here, arrow, assertr, stringr, lubridate, readr)
##
## stopifnot(endsWith(getwd(),"is-ca"))
##
## parser <- ArgumentParser()
## parser$add_argument("--spoa_jep_fase3",
##                     default=here("individual/FGN/clean/output/spoa_jep_fase3.parquet"))
## parser$add_argument("--examples",
##                     default = "individual/FGN/is-ca/hand/example_spoa_jep.csv")
## parser$add_argument("--output",
##                     default = "output/spoa_jep_fase3.parquet")
## args <- parser$parse_args()
##
## # ------- function
##
## # ------- main
##
## data_spoa_jep_fase3 <- read_parquet(args$spoa_jep_fase3)
##
##
## # #IS -CA
##
## viol_is_ca = c("DESAPARICION FORZADA", "DESPLAZAMIENTO", "HOMICIDIO DOLOSO",
##                "RECLUTAMIENTO ILICITO")
##
## perp_is_ca = c("AUC - AUTODEFENSAS UNIDAS DE COLOMBIA",
##                "AGENTES_ESTATALES","ELN - EJERCITO DE LIBERACION NACIONAL",
##                "FARC - FUERZAS ARMADAS REVOLUCIONARIAS DE COLOMBIA",
##                "FUERZA PUBLICA - DELITOS", "OTRAS GUERRILLAS")
##
## cat_caso_is_ca = c("AUTO 092 (LIGA DE MUJERES DESPLAZADAS", "CASOS UP",
##                    "CONFLICTO ARMADO", "CRIMEN DE GUERRA",
##                    "ELN", "FARC-EP", "HOMICIDIO PRESENTADO COMO BAJA POR FUERZA PUBLICA",
##                    "HOMICIDIO PRESENTANDO COMO BAJA POR FUERZA PUBLICA",
##                    "LESA HUMANIDAD", "PARAPOLITICA", "POLICIA NACIONAL VINCULOS GRUPOS ILEGALES",
##                    "PRESUNTOS HOMICIDIOS AGENTES DEL ESTADO", "UNION PATRIOTICA",
##                    "MILITARES VINCULOS GRUPOS ILEGALES","RECLUTAMIENTO DE MENORES",
##                    "VINCULO FUNC. PUBLICOS CON GRUPOS ILEGALES",
##                    "RELACION DE CIVILES CON ACTORES DE CONFLICTO",
##                    "VINCULOS F.P CON GRUPOS AL MARGEN DE LA LEY", "VIOLACION DD.HH",
##                    "VIOLENCIA SEXUAL CONFLICTO ARMADO", "VIXCTIMAS DEFENSORES DE DDHH")
```

```
##
## # is -ca colores
##
##
## perp_verde <- c("AUC - AUTODEFENSAS UNIDAS DE COLOMBIA",
##                 "AUTODEFENSAS",
##                 "AUTODEFENSAS UNIDAS DE COLOMBIA",
##                 "BANDAS EMERGENTES",
##                 "EJERCITO DE LIBERACION NACIONAL",
##                 "ELN - EJERCITO DE LIBERACION NACIONAL",
##                 "EPL - EJERCITO POPULAR DE LIBERACION",
##                 "ERG - EJERCITO REVOLUCIONARIO GUEVARISTA",
##                 "ERP - EJERCITO REVOLUCIONARIO DEL PUEBLO",
##                 "FARC",
##                 "FARC - FUERZAS ARMADAS REVOLUCIONARIAS DE COLOMBIA",
##                 "FARC-EP",
##                 "FARC - FUERZAS ARMADAS REVOLUCIONARIAS COLOMB",
##                 "FARC.",
##                 "FUERZA PUBLICA",
##                 "MJBC - MOVIMIENTO JAIME BATEMAN CAYON",
##                 "OTROS GRUPOS PARAMILITARES",
##                 "AGENTES_ESTATALES",
##                 "AUC - AUTODEFENSAS UNIDAS DE COLOMBIA",
##                 "AUTODEFENSAS",
##                 "AUTODEFENSAS UNIDAS DE COLOMBIA",
##                 "BANDAS EMERGENTES",
##                 "CAP - COMANDO ARMADO POPULAR",
##                 "EJERCITO DE LIBERACION NACIONAL",
##                 "ELN - EJERCITO DE LIBERACION NACIONAL",
##                 "EPL - EJERCITO POPULAR DE LIBERACION",
##                 "ERP - EJERCITO REVOLUCIONARIO DEL PUEBLO",
##                 "FARC",
##                 "FARC - FUERZAS ARMADAS REVOLUCIONARIAS COLOMB",
##                 "FARC-EP",
##                 "FARC.",
##                 "FP - DELITOS",
##                 "FUERZA PUBLICA",
##                 "GUERRILLA",
##                 "MPLICADO FUNCIONARIO PUBLICO",
##                 "MOVIMIENTO 26 DE ABRIL CARLOS PIZARRO LEON",
##                 "OTROS GRUPOS PARAMILITARES",
##                 "PARAMILITARES",
##                 "AGENTES_ESTATALES",
##                 "AUC - AUTODEFENSAS UNIDAS DE COLOMBIA",
##                 "AUTODEFENSAS",
##                 "BANDAS EMERGENTES",
##                 "EJERCITO DE LIBERACION NACIONAL",
##                 "ELN - EJERCITO DE LIBERACION NACIONAL",
##                 "EPL - EJERCITO POPULAR DE LIBERACION",
##                 "ERG - EJERCITO REVOLUCIONARIO GUEVARISTA",
##                 "ERP - EJERCITO REVOLUCIONARIO DEL PUEBLO",
##                 "FARC - FUERZAS ARMADAS REVOLUCIONARIAS COLOMB",
##                 "MOVIMIENTOS POLITICOS",
##                 "PARAMILITARES")
```

```
##
## perp_rojo <- c("DELINCUENCIA COMUN","COMUNIDADES INDIGENAS")
##
## perp_amarillo <- c("CARTEL DEL NORTE DEL VALLE","NARCOTRAFICO",
##                    "CMR - COMANDO MILICIANO REVOLUCIONARIO",
##                    "FP - DELITOS",
##                    "IMPLICADO FUNCIONARIO PUBLICO")
##
## perp_naranja <- c("MOVIMIENTOS POLITICOS")
##
## #a35_titulo
## titulo_rojo <- c("DE LA PROTECCION DE LA INFORMACION Y DE LOS DATOS",
##                  "DE LOS DELITOS CONTRA LA SALUD PUBLICA",
##                  "DE LOS DELITOS CONTRA LOS ANIMALES",
##                  "DE LOS DELITOS CONTRA LOS DERECHOS DE AUTOR",
##                  "DECRETOS",
##                  "DELITOS CONTRA EL ORDEN ECONOMICO SOCIAL",
##                  "DELITOS CONTRA EL ORDEN ECONOMICO Y SOCIAL",
##                  "DELITOS CONTRA LA ADMINISTRACION DE PUBLICA",
##                  "DELITOS CONTRA LA EFICAZ Y RECTA IMPARTICION DE JUSTICIA",
##                  "DELITOS CONTRA LA FAMILIA",
##                  "DELITOS CONTRA LA FE PUBLICA",
##                  "DELITOS CONTRA LA INTEGRIDAD MORAL",
##                  "DELITOS CONTRA LA SALUD PUBLICA")
##
## titulo_verde <- c("DE LOS DELITOS CONTRA EL REGIMEN CONSTITUCIONAL Y LEGAL",
##                   "DELITOS CONTRA LAS PERSONAS Y BIENES PROTEGIDOS POR EL DEREC",
##                   "DELITOS CONTRA PERSONAS Y BIENES PROTEGIDOS POR EL D.I.H.")
##
## titulo_amarillo <- c("DE LOS DELITOS CONTRA LOS RECURSOS NATURALES Y EL MEDIO AMBIENTE",
##                      "DELITOS CONTRA LA LIBERTAD INDIVIDUAL Y OTRAS GARANTIAS",
##                      "DELITOS CONTRA LA LIBERTAD, INTEGRIDAD Y FORMACION SEXUALES",
##                      "DELITOS CONTRA LA VIDA Y LA INTEGRIDAD PERSONAL",
##                      "LEYES",
##                      "SIN DELITO")
##
## titulo_naranja <- c("DELITOS CONTRA EL PATRIMONIO ECONOMICO",
##                     "DELITOS CONTRA LA EXISTENCIA Y SEGURIDAD DEL ESTADO",
##                     "DELITOS CONTRA LA SEGURIDAD PUBLICA",
##                     "DELITOS CONTRA LOS RECURSOS NATURALES Y EL MEDIO AMBIENTE",
##                     "DELITOS CONTRA MECANISMOS DE PARTICIPACION DEMOCRATICA")
##
## #tipo_hecho
## delito_rojo <- c("ABANDONO ART 128 C.P.",
##                  "ABANDONO ART. 127 C.P.",
##                  "ABANDONO DE HIJO FRUTO DE ACCESO CARNAL VIOLENTO ABUSIVO O DE INSEMINACION ARTIFICI
##                  "ABUSO DE CONFIANZA CALIFICADO. ART. 250 C.P.",
##                  "ABUSO DE CONFIANZA. ART. 249 C.P.",
##                  "ACAPARAMIENTO. ART. 297 C.P.",
##                  "ADOPCION IRREGULAR. ART. 232",
##                  "AGIOTAJE. ART. 301",
##                  "APLICACION FRAUDULENTA DE CREDITO OFICIALMENTE REGULADO ART. 311 C.P.",
##                  "APROVECHAMIENTO DE ERROR AJENO O CASO FORTUITO. ART. 252 C.P.",
##                  "ASESORAMIENTO Y OTRAS ACTUACIONES ILEGALES ART. 421 C.P.",
```

```
##                    "ASESORAMIENTO Y OTRAS ACTUACIONES ILEGALES ART. 421 C.P. INCISO 1",
##                    "ASOCIACION PARA LA COMISION DE UN DELITO CONTRA LA ADMINISTRACION PUBLICA ART. 434
##                    "CALUMNIA. ART. 221 C.P.",
##                    "CAPTACION MASIVA Y HABITUAL DE DINEROS ART. 316 C.P.",
##                    "CAZA ILEGAL ART. 336",
##                    "CELEBRACION INDEBIDA DE CONTRATO DE SEGURO ART. 172 C.P.",
##                    "CIRCULACION Y USO DE EFECTO OFICIAL O SELLO FALSIFICADO. ART. 281 C.P.",
##                    "CIRCUNSTANCIA DE AGRAVACION PUNIBLE ART 377B CP LEY 1453 DE 2011",
##                    "COHECHO IMPROPIO ART. 406 C.P.",
##                    "COHECHO POR DAR U OFRECER ART. 407 C.P.",
##                    "COHECHO PROPIO ART. 405 C.P.",
##                    "CONTRABANDO ART. 319 C.P.",
##                    "CONTRATO SIN CUMPLIMIENTO DE REQUISITOS LEGALES ART. 410 C.P.",
##                    "CORRUPCION DE ALIMENTOS, PRODUCTOS MEDICOS O MATERIAL PROFILACTICO ART. 372 C.P.",
##                    "CORRUPCION PRIVADA ART. 250A C.P.",
##                    "DAÑO EN MATERIA PRIMA, PRODUCTO AGROPECUARIO O INDUSTRIAL. ART. 304 C.P.",
##                    "DAÑO INFORMATICO ART 269D LEY 1273 DE 2009",
##                    "DE LA PRESTACION, ACCESO O USO ILEGALES DE LOS SERVICIOS DE TELECOMUNICACIONES ART
##                    "DEFRAUDACION A LOS DERECHOS PATRIMONIALES DE AUTOR. ART. 271 C.P.",
##                    "DEFRAUDACION DE FLUIDOS. ART. 256 C.P.",
##                    "DENEGACION DE INSCRIPCION ART. 396 C.P.",
##                    "DESTRUCCION, SUPRESION U OCULTAMIENTO DE DOCUMENTO PUBLICO. ART. 292 C.P.",
##                    "DESTRUCCION, SUPRESION Y OCULTAMIENTO DE DOCUMENTO PRIVADO. ART. 293 C.P.",
##                    "EJERCICIO ARBITRARIO DE LA CUSTODIA DE HIJO MENOR DE EDAD ART. 230A C.P. AD. LEY 8
##                    "EJERCICIO ILICITO DE ACTIVIDAD MONOPOLISTICA DE ARBITRIO RENTISTICO ART. 312 C.P."
##                    "EMISION Y TRANSFERENCIA ILEGAL DE CHEQUE. ART. 248 C.P.",
##                    "EMISIONES ILEGALES. ART. 276 C.P.",
##                    "ENRIQUECIMIENTO ILICITO ART. 412 C.P.",
##                    "ENRIQUECIMIENTO ILICITO DE PARTICULARES ART. 327 C.P.",
##                    "ENRIQUECIMIENTO ILICITO DE PARTICULARES ART. 327 C.P. INFERIOR 100 SMLM",
##                    "ESTAFA. ART. 246 C.P.",
##                    "ESTIMULO AL USO ILICITO ART. 378 C.P.",
##                    "EXPORTACION O IMPORTACION FICTICIA ART. 310 C.P.",
##                    "FABRICACION Y COMERCIALIZACION DE SUSTANCIAS NOCIVAS PARA LA SALUD ART. 374 C.P.",
##                    "FALSA AUTOACUSACION ART. 437 C.P.",
##                    "FALSA DENUNCIA ART. 435 C.P.",
##                    "FALSA DENUNCIA CONTRA PERSONA DETERMINADA ART. 436 C.P.",
##                    "FALSEDAD EN DOCUMENTO PRIVADO. ART. 289 C.P.",
##                    "FALSEDAD IDEOLOGICA EN DOCUMENTO PUBLICO. ART. 286 C.P.",
##                    "FALSEDAD MARCARIA. ART. 285 C.P.",
##                    "FALSEDAD MARCARIA. ART. 285 C.P. INCISO 1",
##                    "FALSEDAD MATERIAL EN DOCUMENTO PUBLICO. ART. 287 C.P.",
##                    "FALSEDAD PARA OBTENER PRUEBA DE HECHO VERDADERO. ART. 295 C.P.",
##                    "FALSEDAD PERSONAL. ART. 296 C.P.",
##                    "FALSIFICACION DE EFECTO OFICIAL TIMBRADO. ART. 280 C.P.",
##                    "FALSIFICACION DE MONEDA NACIONAL O EXTRANJERA. ART. 273 C.P.",
##                    "FALSIFICACION O USO FRAUDULENTO DE SELLO OFICIAL. ART. 279 C.P.",
##                    "FALSO TESTIMONIO ART. 442 C.P.",
##                    "FAVORECIMIENTO ART. 446 C.P.",
##                    "FAVORECIMIENTO DE CONTRABANDO ART. 320 C.P.",
##                    "FECUNDACION Y TRAFICO DE EMBRIONES ART. 134 C.P.",
##                    "FEMINICIDIO ART. 104A C.P.",
##                    "FRAUDE EN LA INSCRIPCION DE CEDULAS ART. 389 C.P.",
##                    "FRAUDE PROCESAL ART. 453 C.P.",
```

```
##                     "FUGA DE PRESOS ART. 448 C.P.",
##                     "GESTION INDEBIDA DE LOS RECURSOS SOCIALES. ART. 260 C.P.",
##                     "HOMICIDIO CULPOSO ART. 109 C.P.",
##                     "HOMICIDIO POR PIEDAD ART. 106 C.P.",
##                     "HURTO CALIFICADO. ART. 240 C.P.",
##                     "HURTO POR MEDIOS INFORMATICOS Y SEMEJANTES ART. 269I LEY 1273 DE 2009",
##                     "HURTO. ART. 239 C.P.",
##                     "ILICITA EXPLOTACION COMERCIAL ART. 303",
##                     "ILICITO APROVECHAMIENTO DE LOS RECURSOS NATURALES RENOVABLES ART. 328 C.P.",
##                     "IMITACION O SIMULACION DE ALIMENTOS, PRODUCTOS O SUSTANCIAS ART. 373 C.P.",
##                     "IMPEDIMENTO O PERTURBACION DE LA CELBRACIÓN DE AUDIENCIAS PUBLICAS ART. 454C C.P. A
##                     "INASISTENCIA ALIMENTARIA ART. 233 C.P.",
##                     "INCENDIO ART. 350 C.P.",
##                     "INCESTO. ART. 237 C.P.",
##                     "INDUCCION A AYUDA AL SUICIDIO ART. 107 C.P.",
##                     "INFIDELIDAD A LOS DEBERES PROFESIONALES ART. 445 C.P.",
##                     "INJURIA. ART. 220 C.P.",
##                     "INTERES INDEBIDO EN LA CELEBRACION DE CONTRATOS ART. 409 C.P.",
##                     "INTERVENCION EN POLITICA ART. 422 C.P.",
##                     "INVASION DE AREAS DE ESPECIAL IMPORTANCIA ECOLOGICA ART. 337 C.P.",
##                     "INVASION DE TIERRAS O EDIFICACIONES. ART. 263 C.P.",
##                     "IRRESPETO A CADAVERES. ART. 204 C.P.",
##                     "LESIONES AL FETO ART. 125 C.P.",
##                     "LESIONES CULPOSAS AL FETO ART. 126 C.P.",
##                     "LESIONES CULPOSAS ART. 120 C.P.",
##                     "LESIONES CULPOSAS ART. 120 C.P. CON PERDIDA ANATOMICA O FUNCIONAL DE UN ORGANO O MI
##                     "LESIONES CULPOSAS ART. 120 C.P. CON PERTURBACION FUNCIONAL PERMANENTE ART.114 INCIS
##                     "LESIONES CULPOSAS ART. 120 C.P. CON PERTURBACION FUNCIONAL TRANSITORIA ART.114 INCI
##                     "LESIONES CULPOSAS ART. 120 C.P. CON PERTURBACION PSIQUICA PERMANENTE ART.115",
##                     "LESIONES CULPOSAS ART. 120 C.P. CON PERTURBACION PSIQUICA TRANSITORIA ART.115",
##                     "LESIONES CULPOSAS ART. 120 C.P. INCISO 1",
##                     "LESIONES CULPOSAS ART. 120 C.P.AGRAVADO POR NO LICENCIA O LICENCIA SUSPENDIDA ART.
##                     "LESIONES PERSONALES CULPOSAS ART 120 C.P. CON INCAPACIDAD MAYOR 30 DIAS MENOR 90 DI
##                     "LESIONES PERSONALES CULPOSAS ART 120 C.P. CON INCAPACIDAD MAYOR 90 DIAS ART. 112 C
##                     "LESIONES PERSONALES CULPOSAS ART 120 C.P. CON INCAPACIDAD MENOR 30 DIAS ART.112 C.
##                     "LESIONES PERSONALES CULPOSAS ART. 120 C.P CON DEFORMIDAD FISICA AFECTA ROSTRO ART.
##                     "LESIONES PERSONALES CULPOSAS ART. 120 C.P CON DEFORMIDAD FISICA PERMANENTE ART. 11
##                     "LESIONES PERSONALES CULPOSAS ART. 120 C.P. CON DEFORMIDAD FISICA TRANSITORIA ART. 
##                     "MALTRATO ANIMAL ART. 339A C.P.",
##                     "MALTRATO MEDIANTE RESTRICCION A AL LIBERTAD FISICA. ART. 230 C.P.",
##                     "MALTRATO POR DESCUIDO O ABANDONO EN PERSONA MAYOR DE 60 AÑOS",
##                     "MALVERSACION Y DILAPIDACION DE BIENES FAMILIARES. ART. 236 C.P.",
##                     "MANIPULACION FRAUDULENTA DE ESPECIES INSCRITAS EN EL REGISTRO NACIONAL DE VALORES 
##                     "OBSTRUCCION DE OBRAS DE DEFENSA Y ASISTENCIA ART. 364 C.P.",
##                     "OBTENCION DE DOCUMENTO PUBLICO FALSO ART. 288 C.P.",
##                     "OCULTAMIENTO, ALTERACION O DESTRUCCION DE ELEMENTO MATERIAL PROBATORIO ART.454B C.
##                     "OCULTAMIENTO, RETENCION Y POSESION ILICITA DE CEDULA ART. 395 C.P.",
##                     "OFRECIMIENTO ENGAÑOSO DE PRODUCTOS Y SERVICIOS. ART. 300 C.P.",
##                     "OMISION DE DENUNCIA DE PARTICULAR ART. 441 C.P.",
##                     "OMISION DE DENUNCIA DE PARTICULAR ART. 441 C.P. MODIFICADO ART. 18 LEY 1121 DE 200
##                     "OMISION DEL AGENTE RETENEDOR O RECAUDADOR ART. 402 C.P.",
##                     "PANICO ECONOMICO ART. 302 C.P.",
##                     "PARTO O ABORTO PRETERINTENCIONAL ART. 118 C.P.",
##                     "PECULADO CULPOSO ART. 400 C.P.",
```

```
##                   "PECULADO POR APLICACION OFICIAL DIFERENTE ART. 399 C.P.",
##                   "PECULADO POR APROPIACION ART. 397 C.P.",
##                   "PECULADO POR USO ART. 398 C.P.",
##                   "PESCA ILEGAL ART. 335",
##                   "PORNOGRAFIA CON MENORES ART. 218 C.P.",
##                   "PORTE DE SUSTANCIAS ART. 383 C.P.",
##                   "PREVARICATO POR ACCION ART. 413 C.P.",
##                   "PREVARICATO POR OMISION ART. 414 C.P.",
##                   "PROPAGACION DE EPIDEMIA ART. 369 C.P.",
##                   "PROPAGACION DEL VIRUS DE INMUNODEFICIENCIA HUMANA O DE LA HEPATITIS B ART. 370 C.P
##                   "PROXENETISMO CON MENOR DE EDAD ART. 213A C.P.",
##                   "PROXENETISMO CON MENOR DE EDAD ART. 213A C.P. ADICIONADO LEY 1329 DE 2009",
##                   "RECEPTACION ART. 327 C C.P.",
##                   "RECEPTACION ART. 447 C.P.",
##                   "REPETIBILIDAD DEL SER HUMANO ART. 133 C.P.",
##                   "SABOTAJE. ART. 199 C.P.",
##                   "SIN DELITO",
##                   "SINIESTRO O DAÑO DE NAVE ART. 354 C.P.",
##                   "SOBORNO ART. 444 C.P.",
##                   "SOBORNO EN LA ACTUACION PENAL ART. 444A C.P. AD. LEY 890/2004 ART.10",
##                   "SUMINISTRO A MENOR ART. 381 C.P.",
##                   "SUPRESION, ALTERACION O SUPOSICION DEL ESTADO CIVIL. ART. 238 C.P.",
##                   " SUSTRACCION DE BIEN PROPIO. ART. 254 C.P.",
##                   "SUSTRACCION DE COSA PROPIA AL CUMPLIMIENTO DE DEBERES CONSTITUCIONALES O LEGALES AR
##                   "TRAFICO DE MIGRANTES ART. 188 C.P. MOD. LEY 747/2002 ART.1",
##                   "TRAFICO DE MONEDA FALSIFICADA. ART. 274 C.P.",
##                   "TRAFICO DE NIÑAS, NIÑOS Y ADOLESCENTES ART 188C LEY 1453 DE 2011",
##                   "TRAFICO, ELABORACION Y TENENCIA DE ELEMENTOS DESTINADOS A LA FALSIFICACION DE MONED
##                   "TRANSFERENCIA NO CONSENTIDA DE ACTIVOS VALIENDOSE DE ALGUNA MANIPULACION INFORMATIC
##                   "TRATA DE PERSONAS ART. 188A C.P.",
##                   "TRATA DE PERSONAS TRANSNACIONAL ART. 188A C.P.",
##                   "TURISMO SEXUAL. ART.219 C.P. MOD. POR ART 23 L, 1336 DE 2009",
##                   "URBANIZACION ILEGAL ART. 318 C.P.",
##                   "USO DE DOCUMENTO FALSO. ART. 291 C.P.",
##                   "USO DE SOFTWARE MALICIOSO ART 269E LEY 1273 DE 2009",
##                   "USO ILEGITIMO DE PATENTES ART. 307 C.P.",
##                   "USO ILEGITIMO DE PATENTES ART. 307 C.P. INCISO 1. EL QUE FABRIQUE PRODUCTO SIN AUTO
##                   "USURA ART. 305 C.P.",
##                   "USURA ART. 305 C.P. INCISO 1 A CAMBIO DE PRESTAMO DE DINERO O POR CONCEPTO DE VENTA
##                   "USURPACION DE AGUAS. ART. 262 C.P.",
##                   "USURPACION DE DERECHOS DE PROPIEDAD INDUSTRIAL Y DERECHOS DE OBTENTORES DE VARIEDAD
##                   "UTILIZAC.O FACILITAC.MEDIOS DE COMUNICAC.PARA OFRECER ACTIV. SEXUALES CON MENORES D
##                   "VIOLACION A LA LIBERTAD RELIGIOSA. ART. 201 C.P.}VIOLACION A LOS DERECHOS MORALES D
##                   "VIOLACION A LOS DERECHOS MORALES DE AUTOR. ART. 270 C.P. N.1",
##                   "VIOLACION A LOS DERECHOS PATRIMONIALES DE AUTOR Y DERECHOS CONEXOS ART. 271 C.P. MO
##                   "VIOLACION AL REGIMEN LEGAL O CONSTITUCIONAL DE INHABILIDADES E INCOMPATIBILIDADES A
##                   "VIOLACION DE HABITACION AJENA. ART. 189 C.P.",
##                   "VIOLACION DE LA RESERVA INDUSTRIAL O COMERCIAL ART. 308 C.P.",
##                   "VIOLENCIA INTRAFAMILIAR ART. 229 C.P.")
##
## delito_naranja <- c("ABUSO DE AUTORIDAD POR OMISION DE DENUNCIA ART. 417 C.P. INCISO 1",
##                     "ACCESO ABUSIVO A UN SISTEMA INFORMATICO ART 269A LEY 1273 DE 2009",
##                     "ACCESO ABUSIVO A UN SISTEMA INFORMATICO. ART. 195 C.P.",
##                     "ACOSO SEXUAL ART. 210A C.P. ADICIONADO LEY 1257 DE 2008",
```

```
##                        "ACTO SEXUAL VIOLENTO CON MENOR DE CATORCE AÑOS ART. 209 C.P.",
##                        "ACTO SEXUAL VIOLENTO. ART. 206 C.P.",
##                        "DEMANDA DE EXPLOT.SEX. COMERC. MENOR DE 18 AÑOS ADIC. ART 217A C.P. LEY 1329 DE
##                        "DISPARO DE ARMA DE FUEGO ART. 356A CP LEY 1453 DE 2011",
##                        "DISPARO DE ARMA DE FUEGO CONTRA VEHICULO ART. 356 C.P.",
##                        "EXPLOTACION DE MENORES DE EDAD",
##                        "FRAUDE A RESOLUCION JUDICIAL ART. 454 C.P.",
##                        "FRAUDE AL SUFRAGANTE ART. 388 C.P.",
##                        "HOMICIDIO ART. 103 C.P.",
##                        "INSTIGACION A DELINQUIR ART. 348 C.P.",
##                        "INSTIGACION A DELINQUIR ART. 348 C.P. INCISO 1",
##                        "INTERCEPTACION DE DATOS INFORMATICOS, ART 269C LEY 1273 DE 2009",
##                        "LESIONES ART. 111 C.P.",
##                        "LESIONES ART. 113 C.P.",
##                        "LESIONES CON PERDIDA ANATOMICA O FUNCIONAL DE UN ORGANO O MIEMBRO ART.116",
##                        "LESIONES CON PERTURBACION FUNCIONAL PERMANENTE ART.114 INCISO 2",
##                        "LESIONES CON PERTURBACION PSIQUICA PERMANENTE ART.115",
##                        "LESIONES CON PERTURBACION PSIQUICA TRANSITORIA ART.115",
##                        "LESIONES PERSONALES AGRAVADAS POR CIRCUNSTANCIAS ART 104 C.P. ART. 119 C.P.",
##                        "LESIONES PERSONALES ART 120 C.P. CON INCAPACIDAD MENOR 30 DIAS ART.112 C.P. INCI
##                        "LESIONES PERSONALES CON DEFORMIDAD FISICA AFECTA ROSTRO ART. 113 C.P.",
##                        "LESIONES PERSONALES CON DEFORMIDAD FISICA PERMANENTE ART. 113 C.P.INCISO 2",
##                        "LESIONES PERSONALES CON INCAPACIDAD MAYOR 30 DIAS MENOR 90 DIAS ART. 112 C.P. I
##                        "OBSTACULIZACION ILEGITIMA DEL SISTEMA INFORMATICO O RED DE TELECOMUNICACION ART
##                        "OMISION DE APOYO ART. 424 C.P.",
##                        "OMISION DE SOCORRO ART. 131 C.P.",
##                        "PERTURBACION DE ACTOS OFICIALES ART. 430 C.P. INCISO 1",
##                        "PERTURBACION DE CERTAMEN DEMOCRATICO ART. 386 C.P.",
##                        "PERTURBACION DE LA POSESION SOBRE INMUEBLE. ART. 264 C.P.",
##                        "PERTURBACION EN SERVICIO DE TRANSPORTE COLECTIVO U OFICIAL ART. 353 C.P.",
##                        "SECUESTRO SIMPLE ART. 168 C.P.",
##                        "SECUESTRO SIMPLE ART. 168 C.P. AGRAVADO CUANDO SE TRAFIQUE CON LA PERSONA SECUE
##                        "SECUESTRO SIMPLE ART. 168 C.P. AGRAVADO SOBREVENIR MUERTE O LESIONES PERSONALES
##                        "SIMULACION DE INVESTIDURA O CARGO ART. 426 C.P.",
##                        "SUPLANTACION DE SITIOS WEB PARA CAPTURAR DATOS PERSONALES ART 269G LEY 1273 DE
##                        "TENENCIA, FABRICACION Y TRAFICO DE SUSTANCIA U OBJETOS PELIGROSOS ART. 358 C.P.
##                        "TRAFICO DE INFLUENCIAS DE SERVIDOR PUBLICO ART. 411 C.P.",
##                        "TRAFICO DE SUSTANCIAS PARA PROCESAMIENTO DE NARCOTICOS ART. 382 C.P.",
##                        "TRAFICO FABRICACION O PORTE DE ESTUPEFACIENTES ART. 376 C.",
##                        "ULTRAJE A EMBLEMAS Y SIMBOLOS PATRIOS ART. 461 C.P.",
##                        "USURPACION DE FUNCIONES PUBLICAS ART. 425 C.P.",
##                        "UTILIZACION ILICITA DE EQUIPOS TRANSMISORES O RECEPTORES. ART. 197 C.P.",
##                        "UTILIZACION INDEBIDA DE INFLUENCIAS DERIVADAS DEL EJERCICIO DE FUNCION PUBLICA
##                        "VIOLACION DE INMUNIDAD DIPLOMATICA ART. 465 C.P.",
##                        "VOTO FRAUDULENTO ART. 391 C.P")
##
##
## delito_verde <- c("ABORTO FORZADO EN PERSONA PROTEGIDA ART. 139E",
##                   "ACCESO CARNAL ABUSIVO EN PERSONA PROTEGIDA MENOR DE CATORCE AÑOS ART. 138",
##                   "ACCESO CARNAL VIOLENTO EN PERSONA PROTEGIDA ART. 138 C.P.",
##                   "ACTOS DE TERRORISMO ART. 144 C.P.",
##                   "ACTOS SEXUALES CON PERSONA PROTEGIDA MENOR DE CATORCE AÑOS ART. 139A",
##                   "ACTOS SEXUALES VIOLENTOS EN PERSONA PROTEGIDA ART. 139 C.P.",
##                   "ADIC.L.579/2002 ART.2.EMPLEO,PRODUCCION,COMERCIALIZACION Y ALMACENAMIENTO DE MINA
```

```
##                     "ADIC.L.579/2002 ART.3.AYUDA E INDUCCION AL EMPLEO, PRODUCCION Y TRANSFERENCIA DE 
##                     "ADMINISTRACION DE RECURSOS RELACIONADOS CON ACTIVIDADES TERRORISTAS ART. 345 C.P.
##                     "AMENAZAS CONTRA DEFENSORES DE DERECHOS HUMANOS Y SERVIDORES PUBLICOS ART. 188E",
##                     "APOLOGIA DEL GENOCIDIO ART. 102 C.P.",
##                     "ASONADA ART. 469 C.P.",
##                     "ATAQUE CONTRA OBRAS E INSTALACIONES QUE CONTIENEN FUERZAS PELIGROSAS ART. 157 C.P
##                     "ATENTADOS A LA SUBSISTENCIA Y DEVASTACION ART. 160 C.P.",
##                     "CONSPIRACION ART. 471 C.P.",
##                     "CONSTREÑIMIENTO A APOYO BELICO ART. 150 C.P.",
##                     "DEPORTACION, EXPULSION, TRASLADO O DESPLAZAMIENTO FORZADO DE POBLACION CIVIL ART.
##                     "DESCONOCIMIENTO DE HABEAS CORPUS ART. 177 C.P.",
##                     "DESNUDEZ FORZADA EN PERSONA PROTEGIDA ART. 139D",
##                     "DESPLAZAMIENTO FORZADO ART. 180 C.P.",
##                     "DESPOJO EN EL CAMPO DE BATALLA ART. 151 C.P.",
##                     "DESTRUCCION DE BIENES E INSTALACIONES DE CARACTER SANITARIO ART. 155 C.P.",
##                     "DESTRUCCION DEL MEDIO AMBIENTE ART. 164 C.P.",
##                     "DESTRUCCION O UTILIZACION ILICITA DE BIENES CULTURALES Y DE LUGARES DE CULTO ART.
##                     "DESTRUCCION Y APROPIACION DE BIENES PROTEGIDOS ART. 154 C.P.",
##                     "DETENCION ARBITRARIA ESPECIAL ART. 176 C.P.",
##                     "DETENCION ILEGAL Y PRIVACION DEL DEBIDO PROCESO ART. 149 C.P.",
##                     "EMBARAZO FORZADO EN PERSONA PROTEGIDA ART. 139C",
##                     "ESCLAVITUD SEXUAL EN PERSONA PROTEGIDA ART. 141A",
##                     "EXACCION O CONTRIBUCIONES ARBITRARIAS ART. 163 C.P.",
##                     "EXCLAVITUD SEXUAL EN PERSONA PROTEGIDA ART. 141A",
##                     "GENOCIDIO ART. 101 C.P.",
##                     "HOMICIDIO EN PERSONA PROTEGIDA ART. 135 C.P.",
##                     "HOSTILIDAD MILITAR ART. 456 C.P.",
##                     "LESIONES EN PERSONA PROTEGIDA ART. 136 C.P.",
##                     "OBSTACULIZACION DE TAREAS SANITARIAS Y HUMANITARIAS ART. 153 C.P.",
##                     "OMISION DE MEDIDAS DE PROTECCION A LA POBLACION CIVIL ART. 161 C.P.",
##                     "OMISION DE MEDIDAS DE SOCORRO Y ASISTENCIA HUMANITARIAS ART. 152 C.P.",
##                     "PERFIDIA ART. 143 C.P.",
##                     "PRIVACION ILEGAL DE LA LIBERTAD ART. 174 C.P.",
##                     "PROLONGACION ILICITA DE LA PRIVACION DE LA LIBERTAD ART. 175 C.P.",
##                     "PROSTITUCION FORZADA EN PERSONA PROTEGIDA ART. 141",
##                     "PROSTITUCION FORZADA O ESCLAVITUD SEXUAL ART. 141 C.P.",
##                     "REBELION ART. 467 C.P.",
##                     "RECLUTAMIENTO ILICITO ART. 162 C.P.",
##                     "REPRESALIAS ART. 158 C.P.",
##                     "SEDICION ART. 468 C.P.",
##                     "SEDUCCION, USURPACION Y RETENCION ILEGAL DEL MANDO ART. 472 C.P.",
##                     "TOMA DE REHENES ART. 148 C.P.",
##                     "TORTURA CONTRA PERSONA PROTEGIDA ART. 137 C.P.",
##                     "TRATOS INHUMANOS Y DEGRADANTES Y EXPERIMENTOS BIOLOGICOS EN PERSONA PROTEGIDA ART
##                     "UTILIZACION DE MEDIOS Y METODOS DE GUERRA ILICITOS ART. 142 C.P.",
##                     "UTILIZACION ILEGAL DE UNIFORMES E INSIGNIAS ART. 346 C.P.")
##
## delito_amarillo <- c("ABORTO ART. 122 C.P.",
##                        "DAÑO EN BIEN AJENO. ART. 265 C.P.",
##                        "ABORTO SIN CONSENTIMIENTO ART. 123 C.P.",
##                        "ABUSO DE AUTORIDAD POR OMISION DE DENUNCIA ART. 417 C.P.",
##                        "ACCESO CARNAL ABUSIVO CON MENOR DE CATORCE AÑOS. ART. 208 C.P.",
##                        "ACCESO CARNAL O ACTO SEXUAL ABUSIVOS CON INCAPAZ DE RESISTIR. ART. 210 C.P.",
##                        "ACCESO CARNAL ABUSIVO CON MENOR DE CATORCE ANOS. ART. 208 C.P.",
```

```
##                        "ABUSO DE AUTORIDAD POR ACTO ARBITRARIO O INJUSTO ART. 416",
##                        "ABUSO DE CONDICIONES DE INFERIORIDAD. ART. 251 C.P.",
##                        "ACCESO CARNAL O ACTO SEXUAL EN PERSONA PUESTA EN INCAPACIDAD DE RESISTIR ART. 
##                        "ACCESO CARNAL O ACTO SEXUAL EN PERSONA PUESTA EN INCAPACIDAD DE RESISTIR ART. 
##                        "ACCESO CARNAL VIOLENTO. ART. 205 C.P.",
##                        "ACTOS SEXUALES CON MENOR DE CATORCE AÑOS. ART. 209 C.P.",
##                        "ACTOS SEXUALES CON MENOR DE CATORCE ANOS. ART. 209 C.P.",
##                        "ALTERACION DE RESULTADOS ELECTORALES ART. 394 C.P.",
##                        "ALTERACION DESFIGURACION Y SUPLANTACION DE MARCAS DE GANADO. ART. 243 C.P.",
##                        "ALZAMIENTO DE BIENES. ART. 253 C.P.",
##                        "CONCUSION ART. 404 C.P.",
##                        "CONSERVACION O FINANCIACION DE PLANTACIONES ART. 375 C.P.",
##                        "CONTRABANDO DE HIDROCARBUROS Y SUS DERIVADOS ART. 319-1 C.P.",
##                        "ABUSO DE FUNCION PUBLICA ART. 428 C.P.",
##                        "ACTOS DE DISCRIMINACION RACIAL ART. 147 C.P.",
##                        "ACTOS DE RACISMO O DISCRIMINACIÓNART. 134 A",
##                        "CONSTREÑIMIENTO A LA PROSTITUCION ART. 214 C.P.",
##                        "CONSTREÑIMIENTO AL SUFRAGANTE ART. 387 C.P.",
##                        "CONSTREÑIMIENTO ILEGAL ART. 182 C.P.",
##                        "CONTAMINACION AMBIENTAL ART. 332 C.P.",
##                        "CONTAMINACION AMBIENTAL POR EXPLOTACION DE YACIMIENTO MINERO O HIDROCARBURO AR
##                        "CORRUPCION DE SUFRAGANTE ART. 390 C.P.",
##                        "DESTINACION ILEGAL DE COMBUSTIBLE ART. 327D C.P.",
##                        "EXPLOTACION ILICITA DE YACIMIENTO MINERO Y OTROS MATERIALES ART. 338 C.P.",
##                        "EXTORSION. ART. 244 C.P.",
##                        "FABRICACION, IMPORTACION, TRAFICO, POSESION Y USO DE ARMAS QUIMICAS, BIOLOGICA
##                        "FABRICACION, TRAFICO Y PORTE DE ARMAS DE FUEGO O MUNICIONES ART. 365 C.P.",
##                        "FAVORECIMIENTO DE CONTRABANDO DE HIDROCARBUROS Y SUS DERIVADOS ART. 320-1 C.P.
##                        "FAVORECIMIENTO DE LA FUGA ART. 449 C.P.",
##                        "FAVORECIMIENTO DEL DELITO DE CONTRABANDO DE HIDROCARBUROS O SUS DERIVADOS ART.
##                        "FAVORECIMIENTO POR SERVIDOR PUBLICO ART. 322 C.P.",
##                        "FAVORECIMIENTO POR SERVIDOR PUBLICO ART. 322 C.P. INCISO 1",
##                        "OMISION DE CONTROL ART. 325 C.P.",
##                        "REVELACION DE SECRETO ART. 418 C.P. INCISO 1",
##                        "SECUESTRO EXTORSIVO ART. 169 C.P.",
##                        "TERRORISMO ART. 343 C.P.",
##                        "TESTAFERRATO ART. 326 C.P.",
##                        "TESTAFERRATO ART. 326 C.P. MENOR 100 SALARIOS",
##                        "TRAFICO, TRANSPORTE Y POSESION DE MATERIALES RADIOACTIVOS O SUSTANCIAS NUCLEAR
##                        "USO DE MENORES DE EDAD LA COMISION DE DELITO ART 188D LEY 1453 DE 2011",
##                        "USURPACION DE TIERRAS. ART. 261 C.P.",
##                        "UTILIZACION INDEBIDA DE INFORMACION OFICIAL PRIVILEGIADA. ART. 420 C.P.",
##                        "UTILIZACION INDEBIDA DE INFORMACION PRIVILEGIADA. ART. 258 C.P.",
##                        "VIOLACION DE LA LIBERTAD DE TRABAJO. ART. 198 C.P.",
##                        "FRAUDE DE SUBVENCIONES",
##                        "HOMICIDIO PRETERINTENCIONAL ART. 105 C.P.",
##                        "INDUCCION A LA PROSTITUCION ART. 213 C.P.",
##                        "LAVADO DE ACTIVOS ART. 323 C.P.",
##                        "USURPACION DE MARCAS Y PATENTES ART. 306 C.P.",
##                        "DAÑO")
##
## delito_azul <- c("CONSTREÑIMIENTO PARA DELINQUIR ART. 184 C.P.",
##                  "ACTOS DE BARBARIE ART. 145 C.P.",
##                  "AMENAZAS A TESTIGOS ART. 454A C.P. AD. LEY 890 DE 2004 ART.13",
```

```
##                   "AMENAZAS ART. 347 C.P.",
##                   "APODERAMIENTO DE AERONAVES, NAVES O MEDIOS DE TRANSPORTE COLECTIVO. ART. 173 C.P."
##                   "APODERAMIENTO DE LOS HIDROCARBUROS, SUS DERIVADOS, BIOCOMBUSTIBLES O MEZCLAS QUE L
##                   "CONCIERTO PARA DELINQUIR ART. 340 C.P.",
##                   "DAÑO EN LOS RECURSOS NATURALES ART. 331 C.P.",
##                   "DAÑO EN OBRAS DE UTILIDAD SOCIAL ART. 351 C.P.",
##                   "DAÑO EN OBRAS O ELEMENTOS DE LOS SERVICIOS DE COMUNICACIONES, ENERGIA Y COMBUSTIBL
##                   "DESAPARICION FORZADA ART. 165 C.P.",
##                   "DESTINACION ILICITA DE MUEBLES O INMUEBLES ART. 377 C.P.",
##                   "DIVULGACION Y EMPLEO DE DOCUMENTOS RESERVADOS. ART. 194 C.P.",
##                   "EMPLEO ILEGAL DE LA FUERZA PUBLICA ART. 423 C.P.",
##                   "EMPLEO O LANZAMIENTO DE SUSTANCIAS U OBJETOS PELIGROSOS ART. 359 C.P.",
##                   "ENTRENAMIENTO PARA ACTIVIDADES ILICITAS ART. 341 C.P.",
##                   "ESPIONAJE ART. 463 C.P.",
##                   "FABRICACION, TRAFICO Y PORTE DE ARMAS DE USO PRIVATIVO DE LAS FUERZAS ARMADAS ART.
##                   "HOSTIGAMIENTO POR MOTIVOS DE RAZA, RELIGIÓN, IDEOLOGÍA, POLÍTICA, U ORIGEN NACIONA
##                   "OBSTRUCCION A VIAS PUBLICAS QUE AFECTAN EL ORDEN PUBLICO. ART.353A C.P. LEY 1453 D
##                   "TORTURA ART. 178 C.P.",
##                   "UTILIZACION DE ASUNTO SOMETIDO A SECRETO O RESERVA ART. 419 C.P.",
##                   "VIOLACION DE DATOS PERSONALES ART 269F LEY 1273 DE 2009",
##                   "VIOLACION DE HABITACION AJENA POR SERVIDOR PUBLICO ART. 190 C.P.",
##                   "VIOLACION DE LOS DERECHOS DE REUNION Y ASOCIACION. ART. 200 C.P.",
##                   "VIOLACION ILICITA DE COMUNICACIONES O CORRESPONDENCIA DE CARACTER OFICIAL. ART. 19
##                   "VIOLACION ILICITA DE COMUNICACIONES. ART. 192 C.P.",
##                   "VIOLENCIA CONTRA SERVIDOR PUBLICO ART. 429 C.P.")
##
## delito_na <- c("ABORTO ART. 122 C.P.",
##                 "ABORTO SIN CONSENTIMIENTO ART. 123 C.P.",
##                 "ABUSO DE AUTORIDAD POR OMISION DE DENUNCIA ART. 417 C.P.",
##                 "ACCESO CARNAL ABUSIVO CON MENOR DE CATORCE ANOS. ART. 208 C.P.",
##                 "ACCESO CARNAL O ACTO SEXUAL ABUSIVOS CON INCAPAZ DE RESISTIR. ART. 210 C.P.",
##                 "ACCESO CARNAL O ACTO SEXUAL EN PERSONA PUESTA EN INCAPACIDAD DE RESISTIR. ART. 207 C
##                 "ACCESO CARNAL VIOLENTO. ART. 205 C.P.",
##                 "ACTOS SEXUALES CON MENOR DE CATORCE ANOS. ART. 209 C.P.",
##                 "ALTERACION DE RESULTADOS ELECTORALES ART. 394 C.P.",
##                 "ALTERACION DESFIGURACION Y SUPLANTACION DE MARCAS DE GANADO. ART. 243 C.P.",
##                 "ALZAMIENTO DE BIENES. ART. 253 C.P.",
##                 "CONCUSION ART. 404 C.P.",
##                 "CONSERVACION O FINANCIACION DE PLANTACIONES ART. 375 C.P.",
##                 "CONTRABANDO DE HIDROCARBUROS Y SUS DERIVADOS ART. 319-1 C.P.",
##                 "DAÑO EN BIEN AJENO. ART. 265 C.P.",
##                 "DAÑO EN MATERIA PRIMA, PRODUCTO AGROPECUARIO O INDUSTRIAL. ART. 304 C.P.",
##                 "DESTINACION ILICITA DE MUEBLES O INMUEBLES ART. 377 C.P.",
##                 "DISPARO DE ARMA DE FUEGO CONTRA VEHICULO ART. 356 C.P.",
##                 "DIVULGACION Y EMPLEO DE DOCUMENTOS RESERVADOS. ART. 194 C.P.",
##                 "EXPLOTACION ILICITA DE YACIMIENTO MINERO Y OTROS MATERIALES ART. 338 C.P.",
##                 "FABRICACION Y COMERCIALIZACION DE SUSTANCIAS NOCIVAS PARA LA SALUD ART. 374 C.P.",
##                 "FAVORECIMIENTO DE CONTRABANDO DE HIDROCARBUROS Y SUS DERIVADOS ART. 320-1 C.P.",
##                 "FAVORECIMIENTO DEL DELITO DE CONTRABANDO DE HIDROCARBUROS O SUS DERIVADOS ART. 6 DEC
##                 "FRAUDE AL SUFRAGANTE ART. 388 C.P.",
##                 "FRAUDE EN LA INSCRIPCION DE CEDULAS ART. 389 C.P.",
##                 "FUGA DE PRESOS ART. 448 C.P.",
##                 "HOMICIDIO ART. 103 C.P.",
##                 "HURTO CALIFICADO. ART. 240 C.P.",
```

```
##                  "HURTO. ART. 239 C.P.",
##                  "ILICITO APROVECHAMIENTO DE LOS RECURSOS NATURALES RENOVABLES ART. 328 C.P.",
##                  "INCENDIO ART. 350 C.P.",
##                  "INSTIGACION A DELINQUIR ART. 348 C.P.",
##                  "INSTIGACION A DELINQUIR ART. 348 C.P. INCISO 1",
##                  "INTERVENCION EN POLITICA ART. 422 C.P.",
##                  "INVASION DE AREAS DE ESPECIAL IMPORTANCIA ECOLOGICA ART. 337 C.P.",
##                  "INVASION DE TIERRAS O EDIFICACIONES. ART. 263 C.P.",
##                  "IRRESPETO A CADAVERES. ART. 204 C.P.",
##                  "LAVADO DE ACTIVOS ART. 323 C.P.",
##                  "MALTRATO MEDIANTE RESTRICCION A AL LIBERTAD FISICA. ART. 230 C.P.",
##                  "OBSTRUCCION DE OBRAS DE DEFENSA Y ASISTENCIA ART. 364 C.P.",
##                  "OMISION DE APOYO ART. 424 C.P.",
##                  "OMISION DE SOCORRO ART. 131 C.P.",
##                  "PERFIDIA ART. 143 C.P.",
##                  "PERTURBACION DE ACTOS OFICIALES ART. 430 C.P. INCISO 1",
##                  "PERTURBACION DE CERTAMEN DEMOCRATICO ART. 386 C.P.",
##                  "PERTURBACION DE LA POSESION SOBRE INMUEBLE. ART. 264 C.P.",
##                  "PERTURBACION EN SERVICIO DE TRANSPORTE COLECTIVO U OFICIAL ART. 353 C.P.",
##                  "RECEPTACION ART. 447 C.P.",
##                  "RECEPTACION ART. 327 C C.P.",
##                  "REVELACION DE SECRETO ART. 418 C.P. INCISO 1",
##                  "SIMULACION DE INVESTIDURA O CARGO ART. 426 C.P.",
##                  "SINIESTRO O DAÑO DE NAVE ART. 354 C.P.",
##                  "TENENCIA, FABRICACION Y TRAFICO DE SUSTANCIA U OBJETOS PELIGROSOS ART. 358 C.P.",
##                  "TRAFICO DE SUSTANCIAS PARA PROCESAMIENTO DE NARCOTICOS ART. 382 C.P.",
##                  "TRAFICO FABRICACION O PORTE DE ESTUPEFACIENTES ART. 376 C..",
##                  "ULTRAJE A EMBLEMAS Y SIMBOLOS PATRIOS ART. 461 C.P.",
##                  "USO DE DOCUMENTO FALSO. ART. 291 C.P.",
##                  "USURPACION DE FUNCIONES PUBLICAS ART. 425 C.P.",
##                  "UTILIZACION ILICITA DE EQUIPOS TRANSMISORES O RECEPTORES. ART. 197 C.P.",
##                  "UTILIZACION INDEBIDA DE INFORMACION PRIVILEGIADA. ART. 258 C.P.")
##
## # Rules is conflict.
##
## one <- as.integer(1)
## zero <- as.integer(0)
##
## conflict_spoa <- data_spoa_jep_fase3 %>%
##   mutate(prob = runif(nrow(.))) %>%
##   mutate(is_conflict = case_when(
##     a02_tipo_persona == "PERPETRADOR" ~ zero,
##     perpetrador1 %in% perp_verde ~ one,
##     perpetrador2 %in% perp_verde ~ one,
##     perpetrador3 %in% perp_verde ~ one,
##     perpetrador1 %in% perp_amarillo & prob < 0.5 ~ one,
##     perpetrador2 %in% perp_amarillo & prob < 0.5 ~ one,
##     perpetrador3 %in% perp_amarillo & prob < 0.5 ~ one,
##     perpetrador1 %in% perp_naranja & prob < 0.75 ~ zero,
##     perpetrador2 %in% perp_naranja & prob < 0.75 ~ zero,
##     perpetrador3 %in% perp_naranja & prob < 0.75 ~ zero,
##     perpetrador1 %in% perp_rojo ~ zero,
##     perpetrador2 %in% perp_rojo ~ zero,
##     perpetrador3 %in% perp_rojo ~ zero,
```

```
##     perpetrador1 %in% perp_is_ca &  a02_tipo_persona == "VICTIMA" ~ one,
##     perpetrador2 %in% perp_is_ca & a02_tipo_persona == "VICTIMA" ~ one,
##     perpetrador3 %in% perp_is_ca & a02_tipo_persona == "VICTIMA" ~ one,
##     a70_delito_a %in% delito_na ~ NA_integer_,
##     a70_delito_a %in% delito_verde ~ one,
##     a70_delito_a %in% delito_azul & prob < 0.75 ~ one,
##     a70_delito_a %in% delito_amarillo & prob < 0.5 ~ one,
##     a70_delito_a %in% delito_naranja & prob < 0.75 ~ zero,
##     a70_delito_a %in% delito_rojo ~ zero,
##     a35_titulo %in% titulo_verde ~ one,
##     a35_titulo %in% titulo_amarillo & prob < 0.5 ~ one,
##     a35_titulo %in% titulo_naranja & prob < 0.75 ~ zero,
##     a35_titulo %in% titulo_rojo ~ zero,
##     a02_tipo_persona == "OTRA" ~ NA_integer_,
##     TRUE ~ NA_integer_)) %>%
##   select(-prob)
##
## table(conflict_spoa$is_conflict, useNA = "always")
##
## # 0          1      <NA>
## # 1137614  923465   37592
##
## sample <- conflict_spoa  %>%
##   filter(is_conflict ==0) %>%
##   select(recordid, is_conflict)
##
## # ------- training data
##
## sample <- sample_frac(sample, 0.2)
##
## write.table(sample, file = args$examples, sep = "|", quote = FALSE, row.names = FALSE)
##
##
## write_parquet(conflict_spoa, args$output)
```

```
cat(readLines(files$is_ca), sep = "\n")
```

9.4.1.7 FGN - IS_CA

```
## #
## # Authors:     Maria Ortiz
## # Maintainers  Maria Ortiz
## # Copyright   2020, HRDAG,
## # ========================
## # CO-SIVJRNR-data/individual/FGN/is-ca/src/is-ca.R
##
## # ------ setup
##
## Sys.setlocale("LC_CTYPE", "en_US.UTF-8")
## set.seed(19481210)
##
```

```
## require(pacman)
## p_load(argparse, dplyr, here, arrow, assertr, stringr, lubridate)
##
## stopifnot(endsWith(getwd(),"is-ca"))
##
## parser <- ArgumentParser()
## parser$add_argument("--input",
##                     default=here::here("individual/FGN/clean/output/fgn_lideres.parquet"))
## parser$add_argument("--output",
##                     default = "output/fgn_lideres.parquet")
##
## args <- parser$parse_args()
##
## # ------- function
##
## # ------- main
##
## data <-  read_parquet(args$input)
##
## data <- data %>%
##         mutate(is_conflict = 1)
##
## glimpse(data)
##
## write_parquet(data, args$output)
##
## #done
```

```
cat(readLines(files$is_ca_desaparecido), sep = "\n")
```

9.4.1.8 INML - Desaparecidos

```
## #
## # Authors:     JG
## # Maintainers  JG
## # Copyright   2020, HRDAG,
## # ========================
## # CO-SIVJRNR-data/individual/INML/is-ca/src/is-ca-legacy.R
##
## # ------ setup
##
## Sys.setlocale("LC_CTYPE", "en_US.UTF-8")
## set.seed(19481210)
##
## require(pacman)
## p_load(argparse, here, arrow, readr, assertr, dplyr, stringr, lubridate)
##
## stopifnot(endsWith(getwd(),"is-ca"))
##
## parser <- ArgumentParser()
## parser$add_argument("--input",
```

```
##                     default=here::here("individual/INML/clean/output/inml-desaparecidos.parquet"))
## parser$add_argument("--examples",
##                     default = here::here("individual/INML/is-ca/hand/examples-desaparecidos.csv"))
## parser$add_argument("--output",
##                     default = "individual/INML/is-ca/output/inml-desaparecidos.parquet")
## args <- parser$parse_args()
##
## # ------- main
##
## inml_str <-  read_parquet(args$input) %>%
##   mutate(perpetrator = organizacion_responsable,
##          tipo_hecho = clasificacion_de_la_desaparicion,
##          #contexto = circunstancia_de_los_hechos,
##          #mecanismo = causa_de_muerte_definitiva,
##          social_group = grupo_vulnerable)
##
## inml_str <- inml_str %>%
##   mutate_at(vars(tipo_hecho),
##             ~case_when(. == "SIN INFORMACION" ~ NA_character_,
##                        . == "PARA VERIFICACION DE IDENTIDAD" ~NA_character_,
##                        TRUE ~ .)) %>%
##   verify(nrow(.) == nrow(inml_str)) %>%
##   mutate_at(vars(social_group),
##             ~case_when(. == "NINGUNA" ~ NA_character_,
##                        . == "OTROS" ~ NA_character_,
##                        . == "SIN INFORMACION" ~ NA_character_,
##                        . == "NO DILIGENCIADO" ~ NA_character_,
##                        TRUE ~ .)) %>%
##   verify(nrow(.) == nrow(inml_str))
##
## perp_vars <- inml_str %>%
##   group_by(perpetrator, tipo_hecho, social_group) %>%
##   count() %>%
##   arrange(desc(n))
##
## conflict_perps <- c("AUTODEFENSAS INDEPENDIENTES", "AUTODEFENSAS UNIDAS DE COLOMBIA",
##                     "DISIDENCIAS DE LOS PRINCIPALES GRUPOS SUBVERSIVOS", "EJERCITO DE LIBERACION NACI
##                     "FUERZAS ARMADAS REVOLUCIONARIAS DE COLOMBIA", "EJERCITO NACIONAL",
##                     "POLICIA NACIONAL")
##
## inml_str <- inml_str %>%
##   mutate(is_conflict = case_when(
##     perpetrator %in% c("EJERCITO NACIONAL", "POLICIA NACIONAL") & is.na(tipo_hecho) ~ NA_integer_,
##     perpetrator %in% conflict_perps ~ as.integer(1),
##     is.na(perpetrator) & tipo_hecho == "DESAPARICION PRESUNTAMENTE FORZADA" ~ as.integer(1),
##     TRUE ~ NA_integer_))
##
## is_conflict_dist <- inml_str %>%
##   group_by(is_conflict, perpetrator, tipo_hecho, social_group) %>%
##   count() %>%
##   arrange(desc(is_conflict, n))
##
## # there is no records with is_conflict = 0
##
```

```
## write_parquet(inml_str, args$output)
```

```
cat(readLines(files$is_ca_fatales), sep = "\n")
```

9.4.1.9 INML - Fatales

```
## #
## # Authors:     JG
## # Maintainers  JG
## # Copyright   2020, HRDAG,
## # =========================
## # CO-SIVJRNR-data/individual/INML/is-ca/src/is-ca-legacy.R
##
## # ------ setup
##
## Sys.setlocale("LC_CTYPE", "en_US.UTF-8")
## set.seed(19481210)
##
## require(pacman)
## p_load(argparse, here, arrow, readr, assertr, dplyr, stringr, lubridate)
##
## stopifnot(endsWith(getwd(),"is-ca"))
##
## parser <- ArgumentParser()
## parser$add_argument("--input",
##                     default=here::here("individual/INML/clean/output/inml-fatales.parquet"))
## parser$add_argument("--examples",
##                     default = here::here("individual/INML/is-ca/hand/examples-fatales.csv"))
## parser$add_argument("--output",
##                     default = "individual/INML/is-ca/output/inml-fatales.parquet")
## args <- parser$parse_args()
##
## # ------- function
##
## # ------- main
##
## inml_str <-  read_parquet(args$input) %>%
##   mutate(perpetrator = presunto_agresor,
##          tipo_hecho = manera_de_muerte_definitiva,
##          contexto = circunstancia_de_los_hechos,
##          mecanismo = causa_de_muerte_definitiva,
##          social_group = condicion_de_vulnerabilidad)
##
## inml_str <- inml_str %>%
##   mutate_at(vars(tipo_hecho),
##             ~case_when(. == "INDETERMINADA" ~ NA_character_,
##                        . == "EN ESTUDIO" ~ NA_character_,
##                        . == "SIN INFORMACION" ~ NA_character_,
##                        . == "POR DETERMINAR" ~ NA_character_,
##                        . == "NS/NR-SININFORMACION" ~ NA_character_,
##                        . == "POR DETERMINAR" ~ NA_character_,
```

```
##                         . == "VIOLENTA - SIN DETERMINAR" ~ NA_character_,
##                         TRUE ~.)) %>%
##   verify(nrow(.) == nrow(inml_str)) %>%
##   mutate_at(vars(contexto),
##             ~case_when(. == "NS/NR - SIN INFORMACION" ~ NA_character_,
##                         . == "NO APLICA" ~ NA_character_,
##                         TRUE ~ .)) %>%
##   verify(nrow(.) == nrow(inml_str)) %>%
##   mutate_at(vars(mecanismo),
##             ~case_when(. == "NS/NR - SININFORMACION" ~ NA_character_,
##                         . == "INDETERMINADA" ~ NA_character_,
##                         . == "EN ESTUDIO" ~ NA_character_,
##                         . == "SIN INFORMACION" ~ NA_character_,
##                         . == "POR DETERMINAR" ~ NA_character_,
##                         TRUE ~ .)) %>%
##   verify(nrow(.) == nrow(inml_str)) %>%
##   mutate_at(vars(social_group),
##             ~case_when(. == "NINGUNA" ~ NA_character_,
##                         . == "OTROS" ~ NA_character_,
##                         . == "SIN INFORMACION" ~ NA_character_,
##                         TRUE ~ .)) %>%
##   verify(nrow(.) == nrow(inml_str))
##
## perp_vars <- inml_str %>%
##   group_by(perpetrator, contexto, tipo_hecho, mecanismo) %>%
##   count() %>%
##   arrange(desc(n))
##
## # validando categorias por hecho y grupo social para perpetradores relacionados con fuerza pública
## ffmm_perps <- c("SERVICIOS DE INTELIGENCIA", "FUERZAS MILITARES", "POLICIA",
##                 "FUERZAS IRREGULARES")
##
## ffmm_social <- inml_str %>%
##   filter(perpetrator %in% c(ffmm_perps, "GRUPOS DE SEGURIDAD PRIVADA")) %>%
##   group_by(perpetrator, tipo_hecho, social_group) %>%
##   count() %>%
##   arrange(desc(n))
##
## perps_is_ca <- c("SERVICIOS DE INTELIGENCIA", "FARC",
##                  "OTRAS GUERRILLAS", "PARAMILITARES - AUTODEFENSAS", "ELN",
##                  "FUERZAS MILITARES", "POLICIA", "FUERZAS IRREGULARES")
##
## perps_not_ca <- c('MADRASTA', 'ABUELO (A)', 'HERMANO (A)', 'VECINO', 'AMIGO',
##                   'COMPANERO (A) PERMANENTE','EX ESPOSO (A)', 'COMPANERO (A) DE TRABAJO',
##                   'EX-NOVIO (A)','MADRE', 'DELINCUENCIA COMUN',
##                   'CLIENTE', 'PADRE', 'PANDILLAS', 'OTROS FAMILIARES CIVILES O CONSANGUINEOS',
##                   'AMANTE', 'ARRENDADOR', 'ARRENDATARIO', 'CONOCIDO SIN NINGUN TRATO',
##                   'CUNADO (A)', 'EMPLEADO (A)', 'EMPLEADOR','ENCARGADO DEL MENOR', 'ESPOSO (A)',
##                   'EX-AMANTE','HIJO (A)', 'MADRASTRA', 'NOVIO (A)',
##                   'PADRASTRO', 'PRIMO (A)', 'PROFESOR (A)', 'PROVEEDOR', 'SUEGRO (A)', 'TIO (A)')
##
## contexto_is_ca <- c("ACCION GUERRILLERA", "ACCION PARAMILITAR", "CONFLICTO ARMADO",
##                     "ENFRENTAMIENTO ARMADO", "ASESINATO POLITICO", "CONFLICTO POLITICO",
##                     "ACCION MILITAR", "SECUESTRO", "TERRORISMO")
```

```
##
## contexto_not_ca <- c("RINA", "ATRACO CALLEJERO", "MALTRATO DE PAREJA",
##                      "MALTRATO ENTRE OTROS FAMILIARES",
##                      "ROBO ENTIDAD BANCARIA O COMERCIAL", "ROBO RESIDENCIA - MORADA", "ROBO VEHICULO
##                      "VIOLENCIA CONTRA GRUPOS DESCALIFICADOS O MARGINALES",
##                      "RESPONSABILIDAD MEDICA / ENFERMERIA", "RESPONSABILIDAD ODONTOLOGICA")
##
## inml_str <- inml_str %>%
##   mutate(is_conflict = case_when(
##     # validado con el campo social_group: varios registros son posteriormente asignados como
##     # NA cuando no hay grupo social identificado
##     contexto == "ACCIONMILITAR" & is.na(social_group) ~ NA_integer_,
##     contexto == "ENFRENTAMIENTOARMADO" & is.na(social_group) ~ NA_integer_,
##     contexto %in% contexto_is_ca ~ as.integer(1),
##     contexto %in% contexto_not_ca ~ as.integer(0),
##     contexto == "VENGANZA-AJUSTEDECUENTAS" &
##       str_detect(social_group, "(GREMIAL|CAMPESINO|DESMOVILIZADOS|FUNCIONARIOS JUDICIALES|GRUPOS ARMA
##     contexto == "VENGANZA-AJUSTEDECUENTAS" ~ as.integer(0),
##     contexto == "VENGANZA-AJUSTEDECUENTAS" & is.na(social_group) ~ NA_integer_,
##     tipo_hecho == "NATURAL" ~ as.integer(0),
##     tipo_hecho == "VIOLENTA-ACCIDENTAL" | contexto == "AUTOLESIONINVOLUNTARIA(ACCIDENTE)" ~ as.integ
##     tipo_hecho == "VIOLENTA-SUICIDIO" | contexto == "AUTOLESIONVOLUNTARIA(SUICIDIO)" ~ as.integer(0)
##     tipo_hecho == "VIOLENTA-TRANSITO" ~ as.integer(0),
##     perpetrator %in% c(ffmm_perps, "GRUPOS DE SEGURIDAD PRIVADA") & is.na(social_group) ~ NA_integer
##     perpetrator %in% perps_is_ca ~ as.integer(1),
##     perpetrator %in% perps_not_ca ~ as.integer(0),
##     perpetrator == "GRUPOS DE SEGURIDAD PRIVADA" ~ as.integer(0),
##     perpetrator == "GRUPOS DE SEGURIDAD PRIVADA" & social_group == "PRESUNTO COLABORADOR GRUPO ILEGA
##     perpetrator %in% c('NARCOTRAFICANTES', 'GRUPOS DE SEGURIDAD PRIVADA') ~ NA_integer_,
##     TRUE ~ NA_integer_)) %>%
##   verify(nrow(.) == nrow(inml_str))
##
## conflict_perps <- c("ELN", "FARC", "OTRAS GUERRILLAS", "PARAMILITARES - AUTODEFENSAS")
##
## inml_str <- inml_str %>%
##   mutate(is_conflict = case_when(is_conflict == 0 & perpetrator %in% conflict_perps ~ as.integer(1),
##                                  TRUE ~ is_conflict))
##
## is_conflict_dist <- inml_str %>%
##   group_by(is_conflict, perpetrator, contexto, social_group) %>%
##   count() %>%
##   arrange(desc(is_conflict, n))
##
## inconsistent_recs <- inml_str %>%
##   filter(is_conflict == 0  & perpetrator %in% conflict_perps) %>%
##   verify(nrow(.) == 0)
##
## sample_no_conflict <- inml_str %>%
##   filter(is_conflict == 0) %>%
##   group_by(contexto) %>%
##   slice_sample(n = 100) %>%
##   ungroup() %>%
##   select(recordid, is_conflict) %>%
##   distinct()
```

```
##
## write_delim(sample_no_conflict, args$examples, delim = "|")
## write_parquet(inml_str, args$output)
## #done
```

```
cat(readLines(files$is_ca_legacy), sep = "\n")
```

9.4.1.10 INML - Legacy

```
## #
## # Authors:     JG
## # Maintainers  JG
##
## # Copyright   2020, HRDAG,
## # ========================
## # CO-SIVJRNR-data/individual/INML/is-ca/src/is-ca-legacy.R
##
## # ------ setup
## Sys.setlocale("LC_CTYPE", "en_US.UTF-8")
## set.seed(19481210)
##
## require(pacman)
## p_load(argparse,  here, arrow, readr, assertr, dplyr, tidyr, stringr, lubridate)
##
## stopifnot(endsWith(getwd(),"is-ca"))
##
## parser <- ArgumentParser()
## parser$add_argument("--input",
##                     default = here::here("individual/INML/clean/output/inml-legacy.parquet"))
## parser$add_argument("--examples",
##                     default = here::here("individual/INML/is-ca/hand/examples-legacy.csv"))
## parser$add_argument("--output",
##                     default = here::here("individual/INML/is-ca/output/inml-legacy.parquet"))
## args <- parser$parse_args()
##
## inml <- read_parquet(args$input) %>%
##   mutate(N_str = str_count(violation, ",") +1)
##
## data_viol_0 <- filter(inml, is.na(N_str))
## data_viol_1 <- filter(inml, N_str ==1) # tipo_hecho
## data_viol_2 <- filter(inml, N_str ==2) # tipo_hecho, mecanismo
## data_viol_3 <- filter(inml, N_str ==3) # contexto, tipo_hecho, mecanismo
##
## data_viol_0 <- data_viol_0 %>%
##   mutate(contexto = NA_character_,
##          tipo_hecho = NA_character_,
##          mecanismo = NA_character_)
##
## data_viol_1 <- data_viol_1 %>%
##   mutate(contexto = NA_character_,
##          tipo_hecho = violation,
```

```
##          mecanismo = NA_character_)
##
## data_viol_2 <- data_viol_2 %>%
##   mutate(contexto = NA_character_) %>%
##   separate(violation, c("tipo_hecho", "mecanismo"), sep =",", remove = FALSE)
##
## data_viol_3 <- data_viol_3 %>%
##   separate(violation, c("contexto", "tipo_hecho", "mecanismo"), sep = ",", remove = FALSE)
##
## inml_str <- bind_rows(list(data_viol_0, data_viol_1, data_viol_2, data_viol_3)) %>%
##   verify(nrow(.) == nrow(inml))
##
## inml_str <- inml_str %>%
##   mutate_at(vars(tipo_hecho),
##             ~case_when(. == "INDETERMINADA" ~ NA_character_,
##                        . == "ENESTUDIO" ~ NA_character_,
##                        . == "SININFORMACION" ~ NA_character_,
##                        . == "PORDETERMINAR" ~ NA_character_,
##                        . == "NS/NR-SININFORMACION" ~ NA_character_,
##                        . == "PORESTABLECER" ~ NA_character_,
##                        . == "VIOLENTA-SINDETERMINAR" ~ NA_character_,
##                        TRUE ~.)) %>%
##   verify(nrow(.) == nrow(inml_str)) %>%
##   mutate_at(vars(contexto),
##             ~case_when(. == "NS/NR-SININFORMACION" ~ NA_character_,
##                        TRUE ~ .)) %>%
##   verify(nrow(.) == nrow(inml_str)) %>%
##   mutate_at(vars(mecanismo),
##             ~case_when(. == "NS/NR-SININFORMACION" ~ NA_character_,
##                        . == "INDETERMINADA" ~ NA_character_,
##                        . == "ENESTUDIO" ~ NA_character_,
##                        . == "SININFORMACION" ~ NA_character_,
##                        . == "12" ~ NA_character_,
##                        . == "PORESTABLECER" ~ NA_character_,
##                        TRUE ~ .)) %>%
##   verify(nrow(.) == nrow(inml_str)) %>%
##   mutate_at(vars(social_group),
##             ~case_when(. == "NINGUNA" ~ NA_character_,
##                        . == "OTROS" ~ NA_character_,
##                        . == "SIN INFORMACION" ~ NA_character_,
##                        TRUE ~ .)) %>%
##   verify(nrow(.) == nrow(inml_str))
##
##
## N_suicidio <- inml_str %>%
##   filter(violation == "VIOLENTA-SUICIDIO" | contexto == "AUTOLESIONVOLUNTARIA(SUICIDIO)") %>%
##   nrow(.)
##
## N_accidental <- inml_str %>%
##   filter(violation == "VIOLENTA-ACCIDENTAL" | contexto == "AUTOLESIONINVOLUNTARIA(ACCIDENTE)") %>%
##   nrow(.)
##
## N_transito <- filter(inml, violation == "VIOLENTA-TRANSITO") %>% nrow(.)
##
```

```
## perp_vars <- inml_str %>%
##   group_by(perpetrator, contexto, tipo_hecho, mecanismo) %>%
##   count() %>%
##   arrange(desc(n))
##
## perp_contexto <- inml_str %>%
##   group_by(perpetrator, contexto) %>%
##   count() %>%
##   arrange(desc(n))
##
## perp_tipo_hecho <-inml_str %>%
##   group_by(perpetrator, tipo_hecho) %>%
##   count() %>%
##   arrange(desc(n))
##
## perp_mecanismo <- inml_str %>%
##   group_by(perpetrator, mecanismo) %>%
##   count() %>%
##   arrange(desc(n))
##
## # validando categorias por hecho y grupo social para perpetradores relacionados con fuerza pública
## ffmm_perps <- c("SERVICIOS DE INTELIGENCIA", "FUERZAS MILITARES", "POLICIA",
##                 "FUERZAS IRREGULARES")
##
## ffmm_social <- inml_str %>%
##   filter(perpetrator %in% c(ffmm_perps, "GRUPOS DE SEGURIDAD PRIVADA")) %>%
##   group_by(perpetrator, tipo_hecho, social_group) %>%
##   count() %>%
##   arrange(desc(n))
##
## perps_is_ca <- c("SERVICIOS DE INTELIGENCIA", "FARC",
##                  "OTRAS GUERRILLAS", "PARAMILITARES - AUTODEFENSAS", "ELN",
##                  "FUERZAS MILITARES", "POLICIA", "FUERZAS IRREGULARES")
##
## perps_not_ca <- c('MADRASTA', 'ABUELO (A)', 'HERMANO (A)', 'VECINO', 'AMIGO',
##                   'COMPANERO (A) PERMANENTE','EX ESPOSO (A)', 'COMPANERO (A) DE TRABAJO',
##                   'EX-NOVIO (A)','MADRE', 'DELINCUENCIA COMUN',
##                   'CLIENTE', 'PADRE', 'PANDILLAS', 'OTROS FAMILIARES CIVILES O CONSANGUINEOS',
##                   'AMANTE', 'ARRENDADOR', 'ARRENDATARIO', 'CONOCIDO SIN NINGUN TRATO',
##                   'CUNADO (A)', 'EMPLEADO (A)', 'EMPLEADOR','ENCARGADO DEL MENOR', 'ESPOSO (A)',
##                   'EX-AMANTE','HIJO (A)', 'MADRASTRA', 'NOVIO (A)',
##                   'PADRASTRO', 'PRIMO (A)', 'PROFESOR (A)', 'PROVEEDOR', 'SUEGRO (A)', 'TIO (A)')
##
## contexto_is_ca <- c("ACCIONGUERRILLERA", "ACCIONPARAMILITAR", "CONFLICTOARMADO",
##                     "ENFRENTAMIENTOARMADO", "ASESINATOPOLITICO", "CONFLICTOPOLITICO",
##                     "ACCIONMILITAR", "SECUESTRO", "TERRORISMO")
##
## contexto_not_ca <- c("RINA","ACCIDENTEDETRABAJO", "ACCIDENTEDETRANSPORTE", "AGRESIONPORANIMALES",
##                      "ATRACOCALLEJERO", "AUTOLESIONINVOLUNTARIA(ACCIDENTE)", "AUTOLESIONVOLUNTARIA(S
##                      "DESASTRENATURAL", "DESASTRENONATURALOACC.MASIVO", "ECONOMICO",
##                      "EMBRIAGUEZ(ALCOHOLICAYNOALCOHOLICA)", "EXCESOVELOCIDAD", "MALTRATODEPAREJA",
##                      "MALTRATOENTREOTROSFAMILIARES", "POSIBLESFALLASMECANICAS",
##                      "ROBOENTIDADBANCARIAOCOMERCIAL", "ROBORESIDENCIA-MORADA", "ROBOVEHICULO",
##                      "VIOLACIONNORMASDETRANSITOPEATONES", "VIOLACIONOTRASNORMASDETRANSITO",
```

```
##                         "VIOLENCIACONTRAGRUPOSDESCALIFICADOSOMARGINALES",
##                         "RESPONSABILIDADMEDICA/ENFERMERIA", "RESPONSABILIDADODONTOLOGICA")
##
## inml_str <- inml_str %>%
##   mutate(is_conflict = case_when(
##     # validado con el campo social_group: varios registros son posteriormente asignados como
##     # NA cuando no hay grupo social identificado
##      contexto == "ACCIONMILITAR" & is.na(social_group) ~ NA_integer_,
##      contexto == "ENFRENTAMIENTOARMADO" & is.na(social_group) ~ NA_integer_,
##      contexto %in% contexto_is_ca ~ as.integer(1),
##      contexto %in% contexto_not_ca ~ as.integer(0),
##      contexto == "VENGANZA-AJUSTEDECUENTAS" &
##        str_detect(social_group, "(GREMIAL|CAMPESINO|DESMOVILIZADOS|FUNCIONARIOS JUDICIALES|GRUPOS AR
##      contexto == "VENGANZA-AJUSTEDECUENTAS" ~ as.integer(0),
##      contexto == "VENGANZA-AJUSTEDECUENTAS" & is.na(social_group) ~ NA_integer_,
##      tipo_hecho == "NATURAL" ~ as.integer(0),
##      tipo_hecho == "VIOLENTA-ACCIDENTAL" | contexto == "AUTOLESIONINVOLUNTARIA(ACCIDENTE)" ~ as.inte
##      tipo_hecho == "VIOLENTA-SUICIDIO" | contexto == "AUTOLESIONVOLUNTARIA(SUICIDIO)" ~ as.integer(0
##      tipo_hecho == "VIOLENTA-TRANSITO" ~ as.integer(0),
##      perpetrator %in% c(ffmm_perps, "GRUPOS DE SEGURIDAD PRIVADA") & is.na(social_group) ~ NA_intege
##      perpetrator %in% perps_is_ca ~ as.integer(1),
##      perpetrator %in% perps_not_ca ~ as.integer(0),
##      perpetrator == "GRUPOS DE SEGURIDAD PRIVADA" ~ as.integer(0),
##      perpetrator == "GRUPOS DE SEGURIDAD PRIVADA" & social_group == "PRESUNTO COLABORADOR GRUPO ILEG
##      perpetrator %in% c('NARCOTRAFICANTES', 'GRUPOS DE SEGURIDAD PRIVADA') ~ NA_integer_,
##      TRUE ~ NA_integer_)) %>%
##   verify(nrow(.) == nrow(inml_str))
##
##
## is_conflict_dist <- inml_str %>%
##   group_by(is_conflict, perpetrator, contexto, social_group, tipo_hecho) %>%
##   count() %>%
##   arrange(desc(is_conflict, n))
##
## sample_no_conflict <- inml_str %>%
##   filter(is_conflict == 0) %>%
##   group_by(contexto) %>%
##   slice_sample(n = 100) %>%
##   ungroup() %>%
##   select(recordid, is_conflict) %>%
##   distinct()
##
## write_delim(sample_no_conflict, args$examples, delim = "|")
## write_parquet(inml_str, args$output)
##
## #done
```

```
cat(readLines(files$is_ca_inmlcs), sep = "\n")
```

9.4.1.11 INML - Es conflicto

```
## #
```

```
## # Authors:     JG
## # Maintainers  JG
## # Copyright   2020, HRDAG,
## # ========================
## # CO-SIVJRNR-data/individual/INMLCF/is-ca/src/is-ca.R
##
## # ------ setup
##
## Sys.setlocale("LC_CTYPE", "en_US.UTF-8")
## set.seed(19481210)
##
## require(pacman)
## p_load(argparse, here, arrow, readr, assertr, dplyr, stringr, lubridate)
##
## stopifnot(endsWith(getwd(),"is-ca"))
##
## parser <- ArgumentParser()
## parser$add_argument("--input",
##                     default=here::here("individual/INMLCF/is-forced-dis/output/inmlcf-des.parquet"))
## parser$add_argument("--examples",
##                     default = here::here("individual/INML/is-ca/hand/examples-desaparicion.csv"))
## parser$add_argument("--output",
##                     default = "output/inmlcf.parquet")
## args <- parser$parse_args()
##
## #--- funciones
##
## # ------- main
##
## inmlcf <- read_parquet(args$input)
##
## inmlcf <- inmlcf %>%
##   mutate(is_conflict = case_when(
##     # Related to conflict
##     ancestro_racial %in% c("INDIGENA", "NEGRO") ~ as.integer(1),
##     pertenencia_grupal == "GRUPOS ETNICOS" & is_forced_dis == 1 ~ as.integer(1),
##     pertenencia_grupal == "DESPLAZADOS" ~ as.integer(1),
##     pertenencia_grupal == "LIDERES" ~ as.integer(1),
##     pertenencia_grupal == "CAMPESINOS" ~ as.integer(1),
##     pertenencia_grupal == "SERVIDOR PUBLICO" ~ as.integer(1),
##     pertenencia_grupal == "SINDICALISTA" ~ as.integer(1),
##     pertenencia_grupal == "PERIODISTA" ~ as.integer(1),
##     pertenencia_grupal == "POLITICO - DIRIGENTE POLITICO" ~ as.integer(1),
##     pertenencia_grupal == "MISION MEDICA - TRABAJADORES SALUD" ~ as.integer(1),
##     pertenencia_grupal == "HERIDO Y/O ENFERMO BAJO PROTECCION SANITARIA O MEDICA" ~ as.integer(1),
##     pertenencia_grupal == "RELIGIOSOS" & is_forced_dis == 1 ~as.integer(1),
##     pertenencia_grupal == "DESMOVILIZADOS" & is_forced_dis == 1 ~as.integer(1),
##     pertenencia_grupal == "ADICTOS" & is_forced_dis == 1 ~as.integer(1),
##     pertenencia_grupal == "EDUCADORES" & is_forced_dis == 1 ~as.integer(1),
##     clasificacion_desaparicion == "SECUESTRO" ~ as.integer(1),
##     clasificacion_desaparicion == "RECLUTAMIENTO" ~ as.integer(1),
##     (clasificacion_desaparicion == "DESAPARICION PRESUNTAMENTE FORZADA" | is_forced_dis ==1) & is.na
##     #pertenencia_grupal == "TRABAJADORA SEXUAL" & clasificacion_desaparicion == "DESAPARICION PRESUN
##     perpetrator %in% c("GUERRILLAS", "ESTADO", "PARAMILITARES", "BANDAS EMERGENTES") ~ as.integer(1)
```

```
##     # not related to conflict
##     clasificacion_desaparicion == "DESASTRE NATURAL" ~ as.integer(0),
##     pertenencia_grupal == "TRABAJADORA SEXUAL" & is.na(clasificacion_desaparicion) ~ as.integer(0),
##     clasificacion_desaparicion == "TRATA DE PERSONAS" ~ as.integer(0),
##     perpetrator %in% c('PANDILLAS') ~ as.integer(0),
##     perpetrator %in% c('FAMILIARES') ~ as.integer(0),
##     pertenencia_grupal == "RELCUSO" ~ NA_integer_,
##     TRUE ~ NA_integer_))
##
## is_conflict_dist <- inmlcf %>%
##   group_by(is_conflict, perpetrator) %>%
##   count() %>%
##   arrange(desc(is_conflict, n))
##
##
## is_conflict_desap <- inmlcf %>%
##   group_by(is_conflict, perpetrator, clasificacion_desaparicion) %>%
##   count() %>%
##   arrange(desc(is_conflict, n))
##
## sample_no_conflict <- inmlcf %>%
##   filter(is_conflict == 0) %>%
##   group_by(perpetrator) %>%
##   slice_sample(n = 50) %>%
##   ungroup() %>%
##   select(recordid, is_conflict) %>%
##   distinct()
##
## write_delim(sample_no_conflict, args$examples, delim = "|")
## write_parquet(inmlcf, args$output)
##
## #done
```

9.4.2 Scripts and Excel used for the "is enforced disappearance" variable

```
cat(readLines(files$is_forced_dis_fgn), sep = "\n")
```

9.4.2.1 FGN - Es desaparición forzada

```
## #
## # Authors:     Maria Ortiz
## # Maintainers  Maria Ortiz
## # Copyright   2021, HRDAG,
## # ========================
## # CO-SIVJRNR-data/individual/FGN/is-forced-dis/src/is-forced-dis.R
##
## # ------ setup
##
## Sys.setlocale("LC_CTYPE", "en_US.UTF-8")
## set.seed(19481210)
##
```

```
## require(pacman)
## p_load(argparse, dplyr, here, arrow, assertr, stringr, lubridate)
##
## stopifnot(endsWith(getwd(),"is-forced-dis"))
##
## parser <- ArgumentParser()
## parser$add_argument("--input",
##                     default=here::here("individual/FGN/clean/output/vinculacion.parquet"))
## parser$add_argument("--output",
##                     default = "output/fgn.parquet")
##
## args <- parser$parse_args()
##
## # ------- function
##
## # ------- main
##
## data <-  read_parquet(args$input)
##
## if (str_detect(args$input, "hv_bloque_oriental")){
##   data <- data %>%
##     mutate(is_forced_dis = case_when(tipo_hecho == "DESAPARICION FORZADA" ~ as.integer(1),
##                                      TRUE ~ as.integer(0)))
##
## } else if (str_detect(args$input, "lideres")){
##   data <- data %>%
##     mutate(is_forced_dis = case_when(grupo_delito == "DESAPARICION FORZADA" ~ as.integer(1),
##                                      TRUE ~ as.integer(0)))
##
## } else if (str_detect(args$input, "jep_consolidado")){
##   data <- data %>%
##     mutate(is_forced_dis = case_when(hecho == "DESAPARICION FORZADA" ~ as.integer(1),
##                                      TRUE ~ as.integer(0)))
##
## } else if (str_detect(args$input, "up-fgn-transicional")){
##   data <- data %>%
##     mutate(is_forced_dis = case_when(afectacion_1_adj == "DESAPARICION FORZADA" ~ as.integer(1),
##                                      TRUE ~ as.integer(0)))
##
## } else if (str_detect(args$input, "vinculacion")){
##   data <- data %>%
##     mutate(is_forced_dis = case_when(hecho == "DESAPARICION FORZADA" ~ as.integer(1),
##                                      TRUE ~ as.integer(0)))
## }
##
## glimpse(data)
## table(data$is_forced_dis, useNA = "always")
##
## write_parquet(data, args$output)
##
## #done
```

```r
cat(readLines(files$is_forced_dis_fgn_sijuf_cv), sep = "\n")
```

9.4.2.2 FGN - SIJUF - CEV

```
## # ---------------------------------
## # Authors:     PA
## # Maintainers  PA, VG, PB
## # Copyright   2022, HRDAG, GPL-2 or newer
## # ---------------------------------
## # CO-SIVJRNR-data/individual/FGN/is-forced-dis/src/is-forced-dis-FISCALIA_datos_sijuf_cv.R
##
## # note that this script outputs only the recs that are disappearances.
## # the output should not be used for anything except to feed the is-forced-dis
## # model.
## # ------ setup
##
## Sys.setlocale("LC_CTYPE", "en_US.UTF-8")
## set.seed(19481210)
##
## pacman::p_load(argparse, readxl, dplyr, here, arrow, assertr,
##                stringr, stringi, lubridate, logger)
##
##
## parser <- ArgumentParser()
## parser$add_argument("--input", default="../clean/output/FISCALIA_datos_sijuf_cv.parquet")
## parser$add_argument("--dane", default=here("share/DANE/output/dipo.parquet"))
## parser$add_argument("--dept_anio_probs", default=here("individual/INMLCF/is-forced-dis/hand/inmlcf-i
## parser$add_argument("--output", default = "output/FISCALIA_datos_sijuf_cv.parquet")
## args <- parser$parse_args()
##
## one <- as.integer(1)
## zero <- as.integer(0)
##
## # ------- main
## # Federico & Liliana's work is in this file
## probs <- read_excel(args$dept_anio_probs) %>%
##   mutate(departamento_desaparicion = stri_trans_general(departamento_desaparicion, "Latin-ASCII")) %
##   rename(depto_hechos = departamento_desaparicion) %>%
##   mutate(direction = ifelse(is.na(direction), 1, direction)) %>%
##   mutate(prob = as.numeric(prob)) %>%
##   mutate(prob = ifelse(is.na(prob), 0.5, prob)) %>%
##   rename(prob_forced_dis = prob) %>%
##   mutate(yy_hecho = year)
## log_info("probs read")
##
## data <- read_parquet(args$input) %>%
##   filter(grupo_delito == "DESAPARICION FORZADA") %>%
##   left_join(probs, by = c("depto_hechos", "yy_hecho")) %>%
##   mutate(prob = runif(nrow(.)))
## rm(probs)
## log_info("data initially read")
##
```

```
## etapa_caso_des <- c("ETAPA DE INSTRUCCION","INVESTIGACION","JUICIO","EJECUCION DE PENAS","ETAPA JUIC
## etapa_na_des <- c("QUERELLABLE","INDAGACION","TERMINACION ANTICIPADA","ETAPA DE INVESTIGACION PRELIM
## hecho_no_desp <- c("HOMICIDIO DOLOSO", "DESPLAZAMIENTO","RECLUTAMIENTO ILICITO","FEMINICIDIO")
##
## log_info("starting big case_when")
## data <- data %>%
##   mutate(is_forced_dis = case_when(
##     etapa %in% etapa_caso_des ~ one,
##     (direction == 1) & (prob_forced_dis > prob) ~ one,
##     (direction == 0) & (prob_forced_dis > prob) ~ zero,
##     TRUE ~ NA_integer_)) %>%
##   select(-prob_forced_dis, -prob, -direction) %>%
##   write_parquet(args$output)
##
## print(table(data$is_forced_dis, useNA="always"))
## log_info("done.")
##
## #done
```

```
cat(readLines(files$is_forced_dis_fgn_spoa_cv), sep = "\n")
```

9.4.2.3 FGN - SPOA - CEV

```
## # ---------------------------------
## # Authors:     PA
## # Maintainers  PA, VG, PB
## # Copyright   2022, HRDAG, GPL-2 or newer
## # ---------------------------------
## # CO-SIVJRNR-data/individual/FGN/is-forced-dis/src/is-forced-dis-FISCALIA_datos_spoa_cv.R
##
## # note that this script outputs only the recs that are disappearances.
## # the output should not be used for anything except to feed the is-forced-dis
## # model.
## # ------ setup
##
## Sys.setlocale("LC_CTYPE", "en_US.UTF-8")
## set.seed(19481210)
##
## pacman::p_load(argparse, readxl, dplyr, here, arrow, assertr,
##                stringi, stringr, lubridate, logger)
##
## parser <- ArgumentParser()
## parser$add_argument("--input", default=here("individual/FGN/clean/output/FISCALIA_datos_spoa_cv.parq
## parser$add_argument("--dept_anio_probs", default=here("individual/INMLCF/is-forced-dis/hand/inmlcf-i
## parser$add_argument("--output", default = "output/FISCALIA_datos_spoa_cv.parquet")
## args <- parser$parse_args()
##
## one <- as.integer(1)
## zero <- as.integer(0)
##
## # ------- main
```

```
## # Federico & Liliana's work is in this file
## probs <- read_excel(args$dept_anio_probs) %>%
##   mutate(departamento_desaparicion = stri_trans_general(departamento_desaparicion, "Latin-ASCII")) %
##   rename(dpto_hechos = departamento_desaparicion) %>%
##   mutate(direction = ifelse(is.na(direction), 1, direction)) %>%
##   mutate(prob = as.numeric(prob)) %>%
##   mutate(prob = ifelse(is.na(prob), 0.5, prob)) %>%
##   rename(prob_forced_dis = prob) %>%
##   mutate(yy_hecho = year)
## log_info("probs read")
##
## data <- read_parquet(args$input) %>%
##   filter(grupo_delito == "DESAPARICION FORZADA") %>%
##   left_join(probs, by=c("dpto_hechos", "yy_hecho")) %>%
##   mutate(prob = runif(nrow(.)))
## rm(probs)
## log_info("data read")
##
## etapa_caso_des <- c("ETAPA DE INSTRUCCION","INVESTIGACION","JUICIO","EJECUCION DE PENAS","ETAPA JUIC
## etapa_na_des <- c("QUERELLABLE","INDAGACION","TERMINACION ANTICIPADA","ETAPA DE INVESTIGACION PRELIMI
## hecho_no_desp <- c("HOMICIDIO DOLOSO", "DESPLAZAMIENTO","RECLUTAMIENTO ILICITO","FEMINICIDIO")
##
## log_info("beginning case_when")
## data <- data %>%
##   mutate(is_forced_dis = case_when(
##     etapa_caso %in% etapa_caso_des ~ one,
##     (direction == 1) & (prob_forced_dis > prob) ~ one,
##     (direction == 0) & (prob_forced_dis > prob) ~ zero,
##     TRUE ~ NA_integer_)) %>%
##   select(-prob_forced_dis, -prob, -direction) %>%
##   write_parquet(args$output)
##
## print(table(data$is_forced_dis, useNA="always"))
## log_info("done.")
##
## #done
```

```r
cat(readLines(files$is_forced_dis_sijuf_jep), sep = "\n")
```

9.4.2.4 FGN - SIJUF - JEP

```
## # ---------------------------------
## # Authors:     PA
## # Maintainers  PA, VG, PB
## # Copyright   2022, HRDAG, GPL-2 or newer
## # ---------------------------------
## # CO-SIVJRNR-data/individual/FGN/is-forced-dis/src/is-forced-dis-sijuf-jep-fase4.R
##
## # note that this script outputs only the recs that are disappearances.
## # the output should not be used for anything except to feed the is-forced-dis
## # model.
```

```
## # ------ setup
## Sys.setlocale("LC_CTYPE", "en_US.UTF-8")
## set.seed(19481210)
##
## pacman::p_load(argparse, readxl, dplyr, here, arrow, assertr,
##                stringi, stringr, lubridate, logger)
##
## parser <- ArgumentParser()
## parser$add_argument("--input", default=here("individual/FGN/clean/output/sijuf_jep_fase4.parquet"))
## parser$add_argument("--dept_anio_probs", default=here("individual/INMLCF/is-forced-dis/hand/inmlcf-is
## parser$add_argument("--output", default = "output/sijuf_jep_fase4.parquet")
## args <- parser$parse_args()
##
## one <- as.integer(1)
## zero <- as.integer(0)
##
## # ------- main
## # Federico & Liliana's work is in this file
## probs <- read_excel(args$dept_anio_probs) %>%
##   mutate(departamento_desaparicion = stri_trans_general(departamento_desaparicion, "Latin-ASCII")) %>
##   rename(a18_depto_hechos = departamento_desaparicion) %>%
##   mutate(direction = ifelse(is.na(direction), 1, direction)) %>%
##   mutate(prob = as.numeric(prob)) %>%
##   mutate(prob = ifelse(is.na(prob), 0.5, prob)) %>%
##   rename(prob_forced_dis = prob) %>%
##   mutate(yy_hecho = year)
## log_info("probs read")
##
## data <- read_parquet(args$input) %>%
##   filter(str_detect(a23_delito_ppal, "DESAPARICION FORZADA")) %>%
##   left_join(probs, by = c("a18_depto_hechos", "yy_hecho")) %>%
##   mutate(prob = runif(nrow(.)))
## rm(probs)
## log_info("data read initially")
##
## etapa_caso_des <- c("ETAPA DE INSTRUCCION","INVESTIGACION","JUICIO","EJECUCION DE PENAS","ETAPA JUIC
## etapa_na_des <- c("QUERELLABLE","INDAGACION","TERMINACION ANTICIPADA","ETAPA DE INVESTIGACION PRELIMI
##
## log_info("beginning case_when")
## data <- data %>%
##   mutate(is_forced_dis = case_when(
##     a10_etapa %in% etapa_caso_des ~ one,
##     (direction == 1) & (prob_forced_dis > prob) ~ one,
##     (direction == 0) & (prob_forced_dis > prob) ~ zero)) %>%
##   select(recordid, is_forced_dis) %>%
##   write_parquet(args$output)
## log_info("made is_forced_dis, write complete")
## print(table(data$is_forced_dis, useNA="always"))
##
## #done
```

```
cat(readLines(files$is_forced_dis_spoa_jep3), sep = "\n")
```

9.4.2.5 FGN - SPOA - JEP FASE 3

```
## # --------------------------------
## # Authors:     PA
## # Maintainers  PA, VG, PB
## # Copyright   2022, HRDAG,
## # --------------------------------
## # CO-SIVJRNR-data/individual/FGN/is-forced-dis/src/is-forced-dis-spoa-jep-fase3.R
##
## # note that this script outputs only the recs that are disappearances.
## # the output should not be used for anything except to feed the is-forced-dis
## # model.
## # ------ setup
##
## Sys.setlocale("LC_CTYPE", "en_US.UTF-8")
## set.seed(19481210)
##
## require(pacman)
## p_load(argparse, readxl, dplyr, here, arrow, assertr,
##        stringr, stringi, lubridate, logger)
##
##
## parser <- ArgumentParser()
## parser$add_argument("--input", default="../clean/output/spoa_jep_fase3.parquet")
## parser$add_argument("--dept_anio_probs", default=here("individual/INMLCF/is-forced-dis/hand/inmlcf-is
## parser$add_argument("--output", default = "output/spoa_jep_fase3.parquet")
## args <- parser$parse_args()
##
## one <- as.integer(1)
## zero <- as.integer(0)
##
## # ------- main
## # Federico & Liliana's work is in this file
## probs <- read_excel(args$dept_anio_probs) %>%
##   mutate(departamento_desaparicion = stri_trans_general(departamento_desaparicion, "Latin-ASCII")) %
##   rename(a43_departamento = departamento_desaparicion) %>%
##   mutate(direction = ifelse(is.na(direction), 1, direction)) %>%
##   mutate(prob = as.numeric(prob)) %>%
##   mutate(prob = ifelse(is.na(prob), 0.5, prob)) %>%
##   rename(prob_forced_dis = prob) %>%
##   mutate(yy_hecho = year)
## log_info("probs read")
##
## data <- read_parquet(args$input) %>%
##   filter(str_detect(a39_descripcion, "DESAPARICION FORZADA")) %>%
##   left_join(probs, by = c("a43_departamento", "yy_hecho")) %>%
##   mutate(prob = runif(nrow(.)))
## rm(probs)
## log_info("data read initially")
##
## etapa_caso_des <- c("ETAPA DE INSTRUCCION","INVESTIGACION","JUICIO","EJECUCION DE PENAS","ETAPA JUIC
## etapa_na_des <- c("QUERELLABLE","INDAGACION","TERMINACION ANTICIPADA","ETAPA DE INVESTIGACION PRELIM
```

```
##
## log_info("case_when begins")
## data <- data %>%
##   mutate(is_forced_dis = case_when(
##     a24_etapa %in% etapa_caso_des ~ one,
##     (direction == 1) & (prob_forced_dis > prob) ~ one,
##     (direction == 0) & (prob_forced_dis > prob) ~ zero,
##     TRUE ~ NA_integer_)) %>%
##   select(-prob_forced_dis, -prob, -direction) %>%
##   write_parquet(args$output)
##
## print(table(data$is_forced_dis, useNA="always"))
## log_info("done.")
##
## #done
```

```
cat(readLines(files$is_forced_dis_spoa_jep), sep = "\n")
```

9.4.2.6 FGN - SPOA - JEP

```
## # ---------------------------------
## # Authors:     PA
## # Maintainers  PA, VG, PB
## # Copyright   2022, HRDAG,
## # ---------------------------------
## # CO-SIVJRNR-data/individual/FGN/is-forced-dis/src/is-forced-dis-spoa-jep.R
##
## # note that this script outputs only the recs that are disappearances.
## # the output should not be used for anything except to feed the is-forced-dis
## # model.
## # ------ setup
##
## Sys.setlocale("LC_CTYPE", "en_US.UTF-8")
## set.seed(19481210)
##
## require(pacman)
## p_load(argparse, readxl, dplyr, here, arrow, assertr,
##        stringr, stringi, logger, lubridate)
##
##
## parser <- ArgumentParser()
## parser$add_argument("--input", default=here("individual/FGN/clean/output/spoa_jep.parquet"))
## parser$add_argument("--dept_anio_probs", default=here("individual/INMLCF/is-forced-dis/hand/inmlcf-i
## parser$add_argument("--output", default = "output/spoa_jep.parquet")
## args <- parser$parse_args()
##
## one <- as.integer(1)
## zero <- as.integer(0)
##
## # ------- main
## # Federico & Liliana's work is in this file
```

```
## probs <- read_excel(args$dept_anio_probs) %>%
##   mutate(departamento_desaparicion = stri_trans_general(departamento_desaparicion, "Latin-ASCII")) %
##   rename(departamento = departamento_desaparicion) %>%
##   mutate(direction = ifelse(is.na(direction), 1, direction)) %>%
##   mutate(prob = as.numeric(prob)) %>%
##   mutate(prob = ifelse(is.na(prob), 0.5, prob)) %>%
##   rename(prob_forced_dis = prob) %>%
##   mutate(yy_hecho = year)
## log_info("probs read")
##
## # ------- main
##
## data <- read_parquet(args$input) %>%
##   filter(str_detect(descripcion, "DESAPARICION FORZADA")) %>%
##   left_join(probs, by = c("departamento", "yy_hecho")) %>%
##   mutate(prob = runif(nrow(.)))
## rm(probs)
## log_info("data read initially")
##
## etapa_caso_des <- c("ETAPA DE INSTRUCCION","INVESTIGACION","JUICIO","EJECUCION DE PENAS","ETAPA JUIC
## etapa_na_des <- c("QUERELLABLE","INDAGACION","TERMINACION ANTICIPADA","ETAPA DE INVESTIGACION PRELIM
##
## data <- data %>%
##   mutate(is_forced_dis = case_when(
##     etapa %in% etapa_caso_des ~ as.integer(1),
##     (direction == 1) & (prob_forced_dis > prob) ~ one,
##     (direction == 0) & (prob_forced_dis > prob) ~ zero,
##     TRUE ~ NA_integer_)) %>%
##   select(-prob_forced_dis, -prob, -direction) %>%
##   write_parquet(args$output)
##
## print(table(data$is_forced_dis, useNA="always"))
## log_info("done.")
##
## # done
```

```
cat(readLines(files$is_forced_dis_inml), sep = "\n")
```

9.4.2.7 INML - Es desaparición forzada

```
## #
## # Authors:     Valentina Gómez
## # Maintainers  Valentina Gómez, Paula Amado, PB
## # Copyright   2022, HRDAG, GPL-2 or better
## # ========================
## # CO-SIVJRNR-data/individual/INMLCF/is-forced-dis/src/is-forced-dis.R
##
## # ------ setup
## pacman::p_load(argparse, dplyr, here, tidyr,
##                arrow, assertr, stringr, stringi,
##                lubridate, readr, readxl, yaml)
```

```
##
## stopifnot(str_detect(Sys.getlocale(), "LC_CTYPE=en.US.UTF-8"))
## set.seed(19481210)
## one <- as.integer(1)
## zero <- as.integer(0)
##
## parser <- ArgumentParser()
## parser$add_argument("--input", default=here::here("individual/INMLCF/clean/output/inmlcf-des.parquet
## parser$add_argument("--dept_anio_probs", default="hand/inmlcf-is-forced-dis-prob.xlsx")
## parser$add_argument("--dane", default=here::here("share/DANE/output/dipo.parquet"))
## parser$add_argument("--output", default = "output/inmlcf.parquet")
## args <- parser$parse_args()
##
##
## # ------- main
## # Federico & Liliana's work is in this file
## probs <- read_excel(args$dept_anio_probs) %>%
##   mutate(departamento_desaparicion = stri_trans_general(departamento_desaparicion, "Latin-ASCII")) %
##   mutate(direction = ifelse(is.na(direction), 1, direction)) %>%
##   mutate(prob = as.numeric(prob)) %>%
##   mutate(prob = ifelse(is.na(prob), 0.5, prob)) %>%
##   rename(prob_forced_dis = prob) %>%
##   mutate(yy_hecho = year)
##
## inmlcf <- read_parquet(args$input) %>%
##   mutate(departamento_desaparicion = stri_trans_general(departamento_desaparicion, "Latin-ASCII")) %
##   mutate(yy_hecho = yy_fecha_desaparicion) %>%
##   mutate(perpetrator =
##            case_when(str_detect(presuncion_responsabilidad, "(AUTODEFENSAS|BLOQUE)") ~ "PARAMILITARES
##                      str_detect(presuncion_responsabilidad, "(DISIDENCIAS|FARC|SECRETARIADO|FRENTE|C
##                      str_detect(presuncion_responsabilidad, "(NACIONAL|ESTADO|DIVISION|POLICIA|GAULA
##                      str_detect(presuncion_responsabilidad, "CARTEL") ~ "NARCOTRAFICANTES",
##                      str_detect(presuncion_responsabilidad, "BANDA") ~ "BANDAS EMERGENTES",
##                      str_detect(presuncion_responsabilidad, "(PADRE|MADRE|OTROS FAMILIARES)") ~ "FAM
##                      str_detect(presuncion_responsabilidad, "NINGUNA") ~ NA_character_,
##                      TRUE ~ presuncion_responsabilidad)) %>%
##   left_join(probs, by = c("departamento_desaparicion", "yy_hecho")) %>%
##   mutate(prob = runif(nrow(.)))
##
## inmlcf <- inmlcf %>%
##   mutate(is_forced_dis = case_when(
##     clasificacion_desaparicion == "DESAPARICION PRESUNTAMENTE FORZADA" ~ one,
##     pertenencia_grupal == "DESMOVILIZADOS" & is.na(clasificacion_desaparicion) ~ one,
##     pertenencia_grupal == "FUNCIONARIOS JUDICIALES" ~ one,
##     pertenencia_grupal == "SERVIDOR PUBLICO" ~ one,
##     pertenencia_grupal == "POLITICO - DIRIGENTE POLITICO" ~ one,
##     pertenencia_grupal == "SINDICALISTA" ~ one,
##     pertenencia_grupal == "RECLUSO" ~ one,
##     pertenencia_grupal == "MISION MEDICA - TRABAJADORES SALUD" ~ one,
##     pertenencia_grupal == "PERIODISTA" ~ one,
##     pertenencia_grupal == "GRUPOS ETNICOS" &
##       is.na(clasificacion_desaparicion) &
##       between(yy_fecha_desaparicion, 1998, 2004) ~ one,
##     clasificacion_desaparicion == "SECUESTRO" ~ one,
```

```
##     clasificacion_desaparicion == "RECLUTAMIENTO" ~ one,
##     clasificacion_desaparicion == "DESASTRE NATURAL" ~ zero,
##     str_detect(estado_desaparicion, "ANULADO") ~ zero,
##     str_detect(estado_desaparicion, "MUERTO") & (0.8 > prob) ~ one,
##     str_detect(estado_desaparicion, "VIVO") & (0.8 > prob) ~ zero,
##     is.na(prob_forced_dis) ~ NA_integer_,
##     (direction == 1) & (prob_forced_dis > prob) ~ one,
##     (direction == 0) & (prob_forced_dis > prob) ~ zero,
##     TRUE ~ NA_integer_)) %>%
##   select(-prob_forced_dis, -prob, -direction, -yy_hecho) %>%
##   write_parquet(args$output)
##
## print(table(inmlcf$is_forced_dis, useNA="always"))
## #done
```

```
library(readxl)
df <- read_excel(files$excel)
knitr::kable(df, "latex", longtable = T, booktabs = T)
```

9.4.2.8 INML - EXCEL

dept_code_hecho	year	departamento_desaparicion	prob	direction
5	1980	ANTIOQUIA	1	-
8	1980	ATLÁNTICO	1	-
11	1980	BOGOTÁ D.C	1	-
13	1980	BOLÍVAR	1	-
15	1980	BOYACÁ	1	-
17	1980	CALDAS	1	-
18	1980	CAQUETÁ	1	-
85	1980	CASANARE	1	-
20	1980	CESAR	1	-
27	1980	CHOCÓ	1	-
25	1980	CUNDINAMARCA	1	-
41	1980	HUILA	1	-
47	1980	MAGDALENA	1	-
50	1980	META	1	-
52	1980	NARIÑO	1	-
54	1980	NORTE DE SANTANDER	1	-
63	1980	QUINDIO	1	-
68	1980	SANTANDER	1	-
-	1980	SIN INFORMACIÓN	1	0
70	1980	SUCRE	1	-
73	1980	TOLIMA	1	-
76	1980	VALLE DEL CAUCA	1	-
5	1981	ANTIOQUIA	1	-
11	1981	BOGOTÁ D.C	1	-
15	1981	BOYACÁ	1	-

85	1981	CASANARE	1	-
19	1981	CAUCA	1	-
41	1981	HUILA	1	-
50	1981	META	1	-
68	1981	SANTANDER	1	-
76	1981	VALLE DEL CAUCA	1	-
5	1982	ANTIOQUIA	1	-
81	1982	ARAUCA	1	-
8	1982	ATLÁNTICO	1	-
11	1982	BOGOTÁ D.C	1	-
13	1982	BOLÍVAR	1	-
15	1982	BOYACÁ	1	-
17	1982	CALDAS	1	-
18	1982	CAQUETÁ	1	-
19	1982	CAUCA	1	-
20	1982	CESAR	1	-
27	1982	CHOCÓ	1	-
23	1982	CÓRDOBA	1	-
25	1982	CUNDINAMARCA	1	-
95	1982	GUAVIARE	1	-
41	1982	HUILA	1	-
44	1982	LA GUAJIRA	0.5	-
47	1982	MAGDALENA	1	-
50	1982	META	1	-
52	1982	NARIÑO	1	-
54	1982	NORTE DE SANTANDER	1	-
86	1982	PUTUMAYO	1	-
63	1982	QUINDIO	1	-
66	1982	RISARALDA	1	-
68	1982	SANTANDER	1	-
-	1982	SIN INFORMACIÓN	1	0
-	1982	SIN INFORMACIÓN (VENEZUELA)	1	0
70	1982	SUCRE	1	-
73	1982	TOLIMA	1	-
76	1982	VALLE DEL CAUCA	1	-
81	1983	ARAUCA	1	-
11	1983	BOGOTÁ D.C	1	-
18	1983	CAQUETÁ	1	-
19	1983	CAUCA	1	-
25	1983	CUNDINAMARCA	1	-
47	1983	MAGDALENA	1	-
50	1983	META	1	-
86	1983	PUTUMAYO	1	-
76	1983	VALLE DEL CAUCA	1	-
5	1984	ANTIOQUIA	1	-
81	1984	ARAUCA	1	-
11	1984	BOGOTÁ D.C	1	-
15	1984	BOYACÁ	1	-
18	1984	CAQUETÁ	1	-

19	1984	CAUCA	1	-
23	1984	CÓRDOBA	1	-
25	1984	CUNDINAMARCA	1	-
95	1984	GUAVIARE	1	-
41	1984	HUILA	1	-
47	1984	MAGDALENA	1	-
50	1984	META	1	-
54	1984	NORTE DE SANTANDER	1	-
63	1984	QUINDIO	1	-
68	1984	SANTANDER	1	-
76	1984	VALLE DEL CAUCA	1	-
99	1984	VICHADA	1	-
91	1985	AMAZONAS	1	-
5	1985	ANTIOQUIA	0.8	-
81	1985	ARAUCA	1	-
8	1985	ATLÁNTICO	1	-
11	1985	BOGOTÁ D.C	1	-
13	1985	BOLÍVAR	1	-
15	1985	BOYACÁ	1	-
17	1985	CALDAS	1	-
18	1985	CAQUETÁ	1	-
85	1985	CASANARE	1	-
19	1985	CAUCA	1	-
20	1985	CESAR	1	-
27	1985	CHOCÓ	1	-
23	1985	CÓRDOBA	1	-
25	1985	CUNDINAMARCA	1	-
94	1985	GUAINÍA	1	-
95	1985	GUAVIARE	1	-
41	1985	HUILA	1	-
44	1985	LA GUAJIRA	1	-
47	1985	MAGDALENA	1	-
50	1985	META	1	-
52	1985	NARIÑO	1	-
54	1985	NORTE DE SANTANDER	1	-
86	1985	PUTUMAYO	1	-
63	1985	QUINDIO	1	-
66	1985	RISARALDA	1	-
68	1985	SANTANDER	1	-
-	1985	SIN INFORMACIÓN	1	0
-	1985	SIN INFORMACIÓN (VENEZUELA)	1	0
70	1985	SUCRE	1	-
73	1985	TOLIMA	0.8	-
76	1985	VALLE DEL CAUCA	1	-
99	1985	VICHADA	1	-
5	1986	ANTIOQUIA	0.8	-
81	1986	ARAUCA	0.8	-
11	1986	BOGOTÁ D.C	1	-
13	1986	BOLÍVAR	1	-

15	1986	BOYACÁ	0.8	-
17	1986	CALDAS	1	-
18	1986	CAQUETÁ	1	-
85	1986	CASANARE	1	-
19	1986	CAUCA	1	-
20	1986	CESAR	1	-
23	1986	CÓRDOBA	1	-
25	1986	CUNDINAMARCA	1	-
94	1986	GUAINÍA	1	-
95	1986	GUAVIARE	1	-
41	1986	HUILA	1	-
44	1986	LA GUAJIRA	1	-
-	1986	LARA	1	0
47	1986	MAGDALENA	1	-
50	1986	META	1	-
52	1986	NARIÑO	1	-
54	1986	NORTE DE SANTANDER	1	-
86	1986	PUTUMAYO	1	-
63	1986	QUINDIO	1	-
66	1986	RISARALDA	1	-
68	1986	SANTANDER	1	-
-	1986	SIN INFORMACIÓN	1	0
-	1986	SIN INFORMACIÓN (VENEZUELA)	1	0
70	1986	SUCRE	1	-
73	1986	TOLIMA	1	-
76	1986	VALLE DEL CAUCA	1	-
99	1986	VICHADA	1	-
5	1987	ANTIOQUIA	0.8	-
81	1987	ARAUCA	1	-
8	1987	ATLÁNTICO	1	-
11	1987	BOGOTÁ D.C	1	-
13	1987	BOLÍVAR	1	-
15	1987	BOYACÁ	0.8	-
17	1987	CALDAS	1	-
18	1987	CAQUETÁ	1	-
85	1987	CASANARE	1	-
19	1987	CAUCA	1	-
20	1987	CESAR	1	-
27	1987	CHOCÓ	1	-
23	1987	CÓRDOBA	1	-
25	1987	CUNDINAMARCA	1	-
95	1987	GUAVIARE	1	-
41	1987	HUILA	1	-
44	1987	LA GUAJIRA	1	-
47	1987	MAGDALENA	1	-
50	1987	META	1	-
52	1987	NARIÑO	1	-
54	1987	NORTE DE SANTANDER	1	-
86	1987	PUTUMAYO	1	-

63	1987	QUINDIO	1	-
66	1987	RISARALDA	1	-
68	1987	SANTANDER	1	-
-	1987	SIN INFORMACIÓN	1	-
70	1987	SUCRE	1	-
73	1987	TOLIMA	1	-
76	1987	VALLE DEL CAUCA	1	-
97	1987	VAUPÉS	1	-
99	1987	VICHADA	1	-
5	1988	ANTIOQUIA	0.7	-
81	1988	ARAUCA	1	-
8	1988	ATLÁNTICO	1	-
11	1988	BOGOTÁ D.C	1	-
13	1988	BOLÍVAR	1	-
15	1988	BOYACÁ	1	-
17	1988	CALDAS	0.8	-
18	1988	CAQUETÁ	1	-
85	1988	CASANARE	1	-
19	1988	CAUCA	1	-
20	1988	CESAR	1	-
27	1988	CHOCÓ	1	-
23	1988	CÓRDOBA	1	-
25	1988	CUNDINAMARCA	0.8	-
95	1988	GUAVIARE	1	-
41	1988	HUILA	1	-
44	1988	LA GUAJIRA	1	-
47	1988	MAGDALENA	1	-
50	1988	META	1	-
52	1988	NARIÑO	1	-
54	1988	NORTE DE SANTANDER	1	-
86	1988	PUTUMAYO	1	-
63	1988	QUINDIO	1	-
66	1988	RISARALDA	1	-
68	1988	SANTANDER	0.8	-
-	1988	SIN INFORMACIÓN	1	0
-	1988	SIN INFORMACIÓN (VENEZUELA)	1	0
70	1988	SUCRE	1	-
73	1988	TOLIMA	0.8	-
76	1988	VALLE DEL CAUCA	0.8	-
99	1988	VICHADA	1	-
91	1989	AMAZONAS	1	-
5	1989	ANTIOQUIA	0.8	-
81	1989	ARAUCA	1	-
8	1989	ATLÁNTICO	0.8	-
11	1989	BOGOTÁ D.C	1	-
13	1989	BOLÍVAR	1	-
15	1989	BOYACÁ	0.8	-
17	1989	CALDAS	0.8	-
18	1989	CAQUETÁ	1	-

85	1989	CASANARE	1	-
19	1989	CAUCA	1	-
20	1989	CESAR	1	-
27	1989	CHOCÓ	1	-
23	1989	CÓRDOBA	1	-
25	1989	CUNDINAMARCA	1	-
94	1989	GUAINÍA	1	-
95	1989	GUAVIARE	1	-
41	1989	HUILA	1	-
44	1989	LA GUAJIRA	1	-
47	1989	MAGDALENA	1	-
-	1989	MÉRIDA	1	0
50	1989	META	1	-
52	1989	NARIÑO	1	-
54	1989	NORTE DE SANTANDER	1	-
86	1989	PUTUMAYO	1	-
63	1989	QUINDIO	1	-
66	1989	RISARALDA	1	-
68	1989	SANTANDER	1	-
-	1989	SIN INFORMACIÓN	1	0
-	1989	SIN INFORMACIÓN (VENEZUELA)	1	0
70	1989	SUCRE	1	-
73	1989	TOLIMA	1	-
76	1989	VALLE DEL CAUCA	0.8	-
97	1989	VAUPÉS	1	-
99	1989	VICHADA	1	-
5	1990	ANTIOQUIA	0.8	-
81	1990	ARAUCA	1	-
8	1990	ATLÁNTICO	0.8	-
11	1990	BOGOTÁ D.C	0.8	-
13	1990	BOLÍVAR	0.8	-
15	1990	BOYACÁ	0.8	-
17	1990	CALDAS	0.8	-
18	1990	CAQUETÁ	1	-
85	1990	CASANARE	1	-
19	1990	CAUCA	1	-
20	1990	CESAR	1	-
27	1990	CHOCÓ	1	-
23	1990	CÓRDOBA	1	-
25	1990	CUNDINAMARCA	0.8	-
-	1990	FLORIDA	1	0
95	1990	GUAVIARE	-	-
41	1990	HUILA	1	-
44	1990	LA GUAJIRA	0.7	-
47	1990	MAGDALENA	1	-
50	1990	META	1	-
52	1990	NARIÑO	1	-
54	1990	NORTE DE SANTANDER	1	-
-	1990	PORTUGUESA	1	0

86	1990	PUTUMAYO	1	-
63	1990	QUINDIO	1	-
66	1990	RISARALDA	0.8	-
68	1990	SANTANDER	-	-
-	1990	SIN INFORMACIÓN	1	0
70	1990	SUCRE	1	-
73	1990	TOLIMA	0.8	-
76	1990	VALLE DEL CAUCA	0.7	-
97	1990	VAUPÉS	1	-
99	1990	VICHADA	1	-
5	1991	ANTIOQUIA	0.7	-
81	1991	ARAUCA	1	-
8	1991	ATLÁNTICO	0.7	-
11	1991	BOGOTÁ D.C	0.8	-
13	1991	BOLÍVAR	0.8	-
15	1991	BOYACÁ	0.7	-
17	1991	CALDAS	0.7	-
18	1991	CAQUETÁ	1	-
85	1991	CASANARE	1	-
19	1991	CAUCA	1	-
20	1991	CESAR	1	-
27	1991	CHOCÓ	1	-
23	1991	CÓRDOBA	1	-
25	1991	CUNDINAMARCA	0.8	-
94	1991	GUAINÍA	1	-
95	1991	GUAVIARE	1	-
41	1991	HUILA	1	-
44	1991	LA GUAJIRA	0.7	-
47	1991	MAGDALENA	1	-
50	1991	META	0.8	-
52	1991	NARIÑO	1	-
54	1991	NORTE DE SANTANDER	1	-
86	1991	PUTUMAYO	1	-
63	1991	QUINDIO	1	-
66	1991	RISARALDA	0.8	-
68	1991	SANTANDER	1	-
-	1991	SIN INFORMACIÓN	1	0
70	1991	SUCRE	1	-
73	1991	TOLIMA	0.8	-
76	1991	VALLE DEL CAUCA	0.7	-
0	1992	-	1	0
91	1992	AMAZONAS	1	-
5	1992	ANTIOQUIA	0.8	-
81	1992	ARAUCA	1	-
8	1992	ATLÁNTICO	0.7	-
11	1992	BOGOTÁ D.C	0.7	-
13	1992	BOLÍVAR	1	-
15	1992	BOYACÁ	0.8	-
17	1992	CALDAS	0.7	-

18	1992	CAQUETÁ	1	-
85	1992	CASANARE	1	-
19	1992	CAUCA	1	-
20	1992	CESAR	1	-
27	1992	CHOCÓ	1	-
23	1992	CÓRDOBA	1	-
25	1992	CUNDINAMARCA	0.8	-
-	1992	ESMERALDAS	1	0
95	1992	GUAVIARE	1	-
41	1992	HUILA	0.8	-
44	1992	LA GUAJIRA	0.8	-
47	1992	MAGDALENA	1	-
50	1992	META	0.8	-
52	1992	NARIÑO	1	-
54	1992	NORTE DE SANTANDER	1	-
86	1992	PUTUMAYO	1	-
63	1992	QUINDIO	0.8	-
66	1992	RISARALDA	0.8	-
88	1992	SAN ANDRES Y PROVIDENCIA	1	0
68	1992	SANTANDER	0.8	-
-	1992	SIN INFORMACIÓN	1	0
70	1992	SUCRE	1	-
73	1992	TOLIMA	0.7	-
76	1992	VALLE DEL CAUCA	0.7	-
97	1992	VAUPÉS	1	-
99	1992	VICHADA	1	-
5	1993	ANTIOQUIA	0.8	-
81	1993	ARAUCA	1	-
8	1993	ATLÁNTICO	0.8	-
-	1993	BARINAS	1	0
11	1993	BOGOTÁ D.C	0.7	-
13	1993	BOLÍVAR	1	-
15	1993	BOYACÁ	0.7	-
17	1993	CALDAS	0.7	-
18	1993	CAQUETÁ	1	-
85	1993	CASANARE	1	-
19	1993	CAUCA	1	-
20	1993	CESAR	1	-
27	1993	CHOCÓ	1	-
23	1993	CÓRDOBA	1	-
25	1993	CUNDINAMARCA	0.8	-
-	1993	FLORIDA	1	0
95	1993	GUAVIARE	1	-
41	1993	HUILA	0.8	-
44	1993	LA GUAJIRA	0.8	-
47	1993	MAGDALENA	1	-
50	1993	META	1	-
52	1993	NARIÑO	1	-
54	1993	NORTE DE SANTANDER	1	-

-	1993	PANAMÁ	1	0
86	1993	PUTUMAYO	1	-
63	1993	QUINDIO	0.8	-
66	1993	RISARALDA	0.7	-
68	1993	SANTANDER	1	-
-	1993	SIN INFORMACIÓN	1	0
70	1993	SUCRE	1	-
-	1993	TÁCHIRA	1	0
73	1993	TOLIMA	0.7	-
76	1993	VALLE DEL CAUCA	0.8	-
99	1993	VICHADA	1	-
-	1993	ZULIA	1	0
5	1994	ANTIOQUIA	0.7	-
81	1994	ARAUCA	1	-
8	1994	ATLÁNTICO	0.7	-
11	1994	BOGOTÁ D.C	0.6	-
13	1994	BOLÍVAR	1	-
15	1994	BOYACÁ	0.8	-
17	1994	CALDAS	0.7	-
18	1994	CAQUETÁ	1	-
85	1994	CASANARE	1	-
19	1994	CAUCA	1	-
20	1994	CESAR	1	-
27	1994	CHOCÓ	1	-
23	1994	CÓRDOBA	1	-
25	1994	CUNDINAMARCA	0.7	-
-	1994	DISTRITO CAPITAL	-	-
94	1994	GUAINÍA	1	-
95	1994	GUAVIARE	1	-
41	1994	HUILA	0.8	-
44	1994	LA GUAJIRA	0.8	-
47	1994	MAGDALENA	1	-
50	1994	META	0.8	-
52	1994	NARIÑO	1	-
54	1994	NORTE DE SANTANDER	1	-
86	1994	PUTUMAYO	1	-
63	1994	QUINDIO	0.8	-
66	1994	RISARALDA	0.8	-
68	1994	SANTANDER	0.8	-
-	1994	SIN INFORMACIÓN	1	0
70	1994	SUCRE	1	-
73	1994	TOLIMA	0.7	-
76	1994	VALLE DEL CAUCA	0.8	-
97	1994	VAUPÉS	1	-
99	1994	VICHADA	1	-
5	1995	ANTIOQUIA	0.6	-
-	1995	APURE	1	0
81	1995	ARAUCA	1	-
8	1995	ATLÁNTICO	0.7	-

11	1995	BOGOTÁ D.C	0.6	-
13	1995	BOLÍVAR	1	-
15	1995	BOYACÁ	0.7	-
17	1995	CALDAS	0.7	-
18	1995	CAQUETÁ	1	-
85	1995	CASANARE	1	-
19	1995	CAUCA	1	-
20	1995	CESAR	1	-
27	1995	CHOCÓ	1	-
23	1995	CÓRDOBA	1	-
25	1995	CUNDINAMARCA	0.8	-
94	1995	GUAINÍA	1	-
95	1995	GUAVIARE	1	-
41	1995	HUILA	0.8	-
44	1995	LA GUAJIRA	0.8	-
47	1995	MAGDALENA	1	-
50	1995	META	0.8	-
52	1995	NARIÑO	1	-
54	1995	NORTE DE SANTANDER	1	-
86	1995	PUTUMAYO	1	-
63	1995	QUINDIO	1	-
66	1995	RISARALDA	0.7	-
68	1995	SANTANDER	0.8	-
-	1995	SIN INFORMACIÓN	1	0
-	1995	SIN INFORMACIÓN (VENEZUELA)	1	0
70	1995	SUCRE	1	-
73	1995	TOLIMA	0.7	-
76	1995	VALLE DEL CAUCA	0.8	-
99	1995	VICHADA	1	-
0	1996	(NA)	1	0
91	1996	AMAZONAS	1	-
5	1996	ANTIOQUIA	0.6	-
-	1996	APURE	1	0
81	1996	ARAUCA	1	-
8	1996	ATLÁNTICO	0.7	-
11	1996	BOGOTÁ D.C	0.6	-
13	1996	BOLÍVAR	1	-
15	1996	BOYACÁ	0.7	-
17	1996	CALDAS	0.7	-
18	1996	CAQUETÁ	1	-
85	1996	CASANARE	1	-
19	1996	CAUCA	1	-
20	1996	CESAR	1	-
27	1996	CHOCÓ	1	-
23	1996	CÓRDOBA	1	-
25	1996	CUNDINAMARCA	0.8	-
-	1996	DISTRITO CAPITAL	1	-
94	1996	GUAINÍA	1	-
95	1996	GUAVIARE	1	-

41	1996	HUILA	0.8	-
44	1996	LA GUAJIRA	1	-
47	1996	MAGDALENA	1	-
50	1996	META	0.8	-
52	1996	NARIÑO	1	-
54	1996	NORTE DE SANTANDER	1	-
86	1996	PUTUMAYO	1	-
63	1996	QUINDIO	1	-
66	1996	RISARALDA	0.7	-
68	1996	SANTANDER	0.8	-
-	1996	SIN INFORMACIÓN	1	0
70	1996	SUCRE	1	-
-	1996	TÁCHIRA	1	0
73	1996	TOLIMA	0.7	-
76	1996	VALLE DEL CAUCA	0.8	-
-	1996	VARGAS	1	0
97	1996	VAUPÉS	1	-
99	1996	VICHADA	1	-
91	1997	AMAZONAS	1	-
5	1997	ANTIOQUIA	0.7	-
-	1997	APURE	1	0
81	1997	ARAUCA	1	-
8	1997	ATLÁNTICO	0.7	-
11	1997	BOGOTÁ D.C	0.6	-
13	1997	BOLÍVAR	1	-
15	1997	BOYACÁ	0.7	-
17	1997	CALDAS	0.6	-
18	1997	CAQUETÁ	1	-
85	1997	CASANARE	1	-
19	1997	CAUCA	1	-
20	1997	CESAR	0.8	-
27	1997	CHOCÓ	0.6	-
23	1997	CÓRDOBA	0.8	-
25	1997	CUNDINAMARCA	0.7	-
95	1997	GUAVIARE	0.8	-
41	1997	HUILA	0.8	-
44	1997	LA GUAJIRA	0.7	-
47	1997	MAGDALENA	1	-
50	1997	META	0.8	-
52	1997	NARIÑO	1	-
54	1997	NORTE DE SANTANDER	1	-
86	1997	PUTUMAYO	1	-
63	1997	QUINDIO	0.7	-
66	1997	RISARALDA	0.7	-
68	1997	SANTANDER	0.8	-
-	1997	SIN INFORMACIÓN	1	0
70	1997	SUCRE	1	-
-	1997	TÁCHIRA	1	0
73	1997	TOLIMA	0.6	-

76	1997	VALLE DEL CAUCA	0.8	-
97	1997	VAUPÉS	1	-
99	1997	VICHADA	1	-
-	1997	ZULIA	1	0
0	1998	-	1	0
5	1998	ANTIOQUIA	0.7	-
81	1998	ARAUCA	1	-
8	1998	ATLÁNTICO	0.7	-
11	1998	BOGOTÁ D.C	0.6	-
13	1998	BOLÍVAR	1	-
15	1998	BOYACÁ	0.7	-
17	1998	CALDAS	0.6	-
18	1998	CAQUETÁ	0.8	-
85	1998	CASANARE	1	-
19	1998	CAUCA	1	-
20	1998	CESAR	1	-
27	1998	CHOCÓ	1	-
23	1998	CÓRDOBA	1	-
25	1998	CUNDINAMARCA	0.7	-
94	1998	GUAINÍA	1	-
95	1998	GUAVIARE	1	-
41	1998	HUILA	0.7	-
44	1998	LA GUAJIRA	0.7	-
47	1998	MAGDALENA	0.8	-
50	1998	META	0.7	-
52	1998	NARIÑO	1	-
54	1998	NORTE DE SANTANDER	1	-
86	1998	PUTUMAYO	1	-
63	1998	QUINDIO	0.8	-
66	1998	RISARALDA	0.6	-
68	1998	SANTANDER	0.8	-
-	1998	SIN INFORMACIÓN	1	0
70	1998	SUCRE	1	-
73	1998	TOLIMA	0.6	-
76	1998	VALLE DEL CAUCA	0.8	-
97	1998	VAUPÉS	1	-
99	1998	VICHADA	1	-
5	1999	ANTIOQUIA	0.6	-
-	1999	APURE	1	0
81	1999	ARAUCA	1	-
8	1999	ATLÁNTICO	0.7	-
11	1999	BOGOTÁ D.C	0.7	0
13	1999	BOLÍVAR	0.8	-
15	1999	BOYACÁ	0.7	-
17	1999	CALDAS	0.6	-
18	1999	CAQUETÁ	1	-
85	1999	CASANARE	0.8	-
19	1999	CAUCA	1	-
20	1999	CESAR	1	-

27	1999	CHOCÓ	1	-
23	1999	CÓRDOBA	1	-
25	1999	CUNDINAMARCA	0.7	-
-	1999	DISTRITO CAPITAL	1	-
94	1999	GUAINÍA	1	-
95	1999	GUAVIARE	1	-
41	1999	HUILA	0.7	-
44	1999	LA GUAJIRA	0.8	-
47	1999	MAGDALENA	1	-
50	1999	META	0.8	-
-	1999	MIRANDA	1	0
52	1999	NARIÑO	1	-
54	1999	NORTE DE SANTANDER	0.7	-
86	1999	PUTUMAYO	0.8	-
63	1999	QUINDIO	O.6	-
66	1999	RISARALDA	0.6	-
88	1999	SAN ANDRES Y PROVIDENCIA	1	0
68	1999	SANTANDER	0.8	-
-	1999	SIN INFORMACIÓN	1	0
70	1999	SUCRE	1	-
73	1999	TOLIMA	0.7	-
76	1999	VALLE DEL CAUCA	0.7	-
-	1999	VARGAS	1	0
97	1999	VAUPÉS	1	-
99	1999	VICHADA	1	-
0	2000	-	1	0
91	2000	AMAZONAS	1	-
5	2000	ANTIOQUIA	0.5	0
-	2000	APURE	1	0
81	2000	ARAUCA	1	0
8	2000	ATLÁNTICO	0.6	-
11	2000	BOGOTÁ D.C	0.7	0
13	2000	BOLÍVAR	0.7	0
15	2000	BOYACÁ	0.7	0
17	2000	CALDAS	0.7	0
18	2000	CAQUETÁ	0.8	-
85	2000	CASANARE	1	-
19	2000	CAUCA	0.8	-
20	2000	CESAR	0.8	-
27	2000	CHOCÓ	1	-
23	2000	CÓRDOBA	0.7	-
25	2000	CUNDINAMARCA	0.7	0
-	2000	CURAÇAO	1	-
94	2000	GUAINÍA	1	-
95	2000	GUAVIARE	1	-
41	2000	HUILA	0.7	-
44	2000	LA GUAJIRA	0.6	-
47	2000	MAGDALENA	0.7	-
50	2000	META	0.8	-

-	2000	MONTEVIDEO	1	0
52	2000	NARIÑO	1	-
54	2000	NORTE DE SANTANDER	0.8	-
-	2000	PUNTARENAS	1	0
86	2000	PUTUMAYO	0.8	-
63	2000	QUINDIO	0.7	-
66	2000	RISARALDA	0.6	-
68	2000	SANTANDER	0.8	-
-	2000	SIN INFORMACIÓN	1	0
-	2000	SIN INFORMACIÓN (VENEZUELA)	1	0
70	2000	SUCRE	1	-
73	2000	TOLIMA	0.5	0
76	2000	VALLE DEL CAUCA	0.5	0
97	2000	VAUPÉS	1	-
99	2000	VICHADA	1	-
0	2001	-	1	0
5	2001	ANTIOQUIA	0.7	0
-	2001	APURE	1	0
81	2001	ARAUCA	1	-
8	2001	ATLÁNTICO	0.7	-
11	2001	BOGOTÁ D.C	0.8	0
13	2001	BOLÍVAR	0.7	-
15	2001	BOYACÁ	0.7	0
17	2001	CALDAS	0.8	0
18	2001	CAQUETÁ	0.8	-
85	2001	CASANARE	0.8	-
-	2001	CATALUÑA	1	0
19	2001	CAUCA	0.8	-
20	2001	CESAR	0.8	-
27	2001	CHOCÓ	1	-
23	2001	CÓRDOBA	0.8	-
25	2001	CUNDINAMARCA	0.7	0
95	2001	GUAVIARE	1	-
41	2001	HUILA	0.6	0
44	2001	LA GUAJIRA	0.7	-
47	2001	MAGDALENA	1	-
50	2001	META	0.7	-
52	2001	NARIÑO	0.8	-
-	2001	NEW YORK	1	0
54	2001	NORTE DE SANTANDER	0.8	-
86	2001	PUTUMAYO	0.8	-
63	2001	QUINDIO	0.7	-
66	2001	RISARALDA	0.6	-
68	2001	SANTANDER	0.8	-
-	2001	SIN INFORMACIÓN	1	0
70	2001	SUCRE	1	-
-	2001	SUCUMBÍOS	1	0
-	2001	TÁCHIRA	1	0
73	2001	TOLIMA	0.7	0

76	2001	VALLE DEL CAUCA	0.8	-
97	2001	VAUPÉS	1	-
99	2001	VICHADA	1	-
0	2002	-	1	0
91	2002	AMAZONAS	1	-
5	2002	ANTIOQUIA	0.5	-
-	2002	APURE	1	0
81	2002	ARAUCA	0.7	-
8	2002	ATLÁNTICO	0.7	0
11	2002	BOGOTÁ D.C	0.8	0
13	2002	BOLÍVAR	0.8	-
15	2002	BOYACÁ	0.7	0
17	2002	CALDAS	0.8	0
18	2002	CAQUETÁ	0.8	-
85	2002	CASANARE	0.8	-
19	2002	CAUCA	0.7	-
20	2002	CESAR	0.8	-
27	2002	CHOCÓ	1	-
23	2002	CÓRDOBA	1	-
25	2002	CUNDINAMARCA	0.7	0
94	2002	GUAINÍA	1	-
95	2002	GUAVIARE	0.8	-
41	2002	HUILA	0.6	0
44	2002	LA GUAJIRA	0.7	0
47	2002	MAGDALENA	0.7	-
-	2002	MÉRIDA	1	0
50	2002	META	0.7	-
-	2002	MONAGAS	1	0
52	2002	NARIÑO	0.7	-
54	2002	NORTE DE SANTANDER	0.7	0
86	2002	PUTUMAYO	0.7	-
63	2002	QUINDIO	0.7	0
66	2002	RISARALDA	0.8	0
-	2002	SAN JOSÉ	1	0
-	2002	SANTA CRUZ	1	0
68	2002	SANTANDER	1	-
-	2002	SIN INFORMACIÓN	1	0
-	2002	SIN INFORMACIÓN (VENEZUELA)	1	0
70	2002	SUCRE	1	-
-	2002	TÁCHIRA	1	0
73	2002	TOLIMA	0.8	0
76	2002	VALLE DEL CAUCA	0.7	-
97	2002	VAUPÉS	1	-
99	2002	VICHADA	1	-
-	2002	ZULIA	1	0
0	2003	-	1	0
91	2003	AMAZONAS	0.8	-
5	2003	ANTIOQUIA	0.7	-
-	2003	APURE	1	0

81	2003	ARAUCA	1	-
8	2003	ATLÁNTICO	0.8	0
11	2003	BOGOTÁ D.C	0.8	0
13	2003	BOLÍVAR	0.8	-
15	2003	BOYACÁ	0.7	0
17	2003	CALDAS	0.8	0
18	2003	CAQUETÁ	0.8	-
-	2003	CARABOBO	1	0
85	2003	CASANARE	0.7	-
19	2003	CAUCA	1	-
20	2003	CESAR	0.8	-
27	2003	CHOCÓ	1	-
23	2003	CÓRDOBA	1	-
25	2003	CUNDINAMARCA	0.8	0
-	2003	DISTRITO CAPITAL	1	-
94	2003	GUAINÍA	1	-
95	2003	GUAVIARE	0.8	-
41	2003	HUILA	0.6	-
44	2003	LA GUAJIRA	0.7	0
47	2003	MAGDALENA	0.8	-
-	2003	MÉRIDA	1	0
50	2003	META	0.7	-
52	2003	NARIÑO	0.8	-
54	2003	NORTE DE SANTANDER	0.8	-
-	2003	PANAMÁ	1	0
86	2003	PUTUMAYO	0.8	-
63	2003	QUINDIO	0.7	0
66	2003	RISARALDA	0.8	0
68	2003	SANTANDER	0.8	-
-	2003	SIN INFORMACIÓN	1	0
-	2003	SOLOLÁ	1	0
70	2003	SUCRE	1	-
-	2003	TÁCHIRA	1	0
73	2003	TOLIMA	0.8	0
76	2003	VALLE DEL CAUCA	0.5	0
97	2003	VAUPÉS	1	-
99	2003	VICHADA	1	-
0	2004	-	1	0
5	2004	ANTIOQUIA	0.5	-
-	2004	APURE	1	0
81	2004	ARAUCA	1	-
8	2004	ATLÁNTICO	0.6	0
11	2004	BOGOTÁ D.C	0.8	0
13	2004	BOLÍVAR	0.7	-
15	2004	BOYACÁ	0.6	-
17	2004	CALDAS	0.7	0
18	2004	CAQUETÁ	0.8	-
85	2004	CASANARE	0.8	-
19	2004	CAUCA	0.8	-

20	2004	CESAR	0.8	-
27	2004	CHOCÓ	0.8	-
23	2004	CÓRDOBA	0.8	-
25	2004	CUNDINAMARCA	0.6	-
-	2004	DISTRITO CAPITAL	1	-
-	2004	ESMERALDAS	1	0
94	2004	GUAINÍA	1	-
95	2004	GUAVIARE	0.8	-
41	2004	HUILA	0.6	-
44	2004	LA GUAJIRA	0.7	-
-	2004	LARA	1	0
47	2004	MAGDALENA	0.8	-
-	2004	MÉRIDA	1	0
50	2004	META	0.8	-
52	2004	NARIÑO	0.8	-
54	2004	NORTE DE SANTANDER	0.7	-
-	2004	PANAMÁ	1	0
-	2004	PICHINCHA	1	-
86	2004	PUTUMAYO	0.8	-
63	2004	QUINDIO	0.7	-
66	2004	RISARALDA	0.8	-
88	2004	SAN ANDRES Y PROVIDENCIA	1	0
68	2004	SANTANDER	0.8	-
-	2004	SIN INFORMACIÓN	1	0
-	2004	SIN INFORMACIÓN (VENEZUELA)	1	0
70	2004	SUCRE	1	-
-	2004	TÁCHIRA	1	0
73	2004	TOLIMA	0.8	0
76	2004	VALLE DEL CAUCA	0.6	-
97	2004	VAUPÉS	1	-
99	2004	VICHADA	1	-
0	2005	-	1	0
91	2005	AMAZONAS	1	-
5	2005	ANTIOQUIA	0.7	-
-	2005	APURE	1	0
81	2005	ARAUCA	1	-
8	2005	ATLÁNTICO	0.7	0
11	2005	BOGOTÁ D.C	0.8	0
13	2005	BOLÍVAR	1	-
15	2005	BOYACÁ	0.8	0
17	2005	CALDAS	0.7	-
18	2005	CAQUETÁ	1	-
-	2005	CARABOBO	1	0
85	2005	CASANARE	0.8	-
19	2005	CAUCA	1	-
20	2005	CESAR	1	-
27	2005	CHOCÓ	1	-
23	2005	CÓRDOBA	1	-
25	2005	CUNDINAMARCA	0.8	0

94	2005	GUAINÍA	1	-
95	2005	GUAVIARE	0.8	-
41	2005	HUILA	0.7	0
-	2005	IMBABURA	1	0
44	2005	LA GUAJIRA	0.7	0
-	2005	LOJA	1	0
47	2005	MAGDALENA	1	-
50	2005	META	0.8	-
52	2005	NARIÑO	0.8	-
54	2005	NORTE DE SANTANDER	0.8	-
86	2005	PUTUMAYO	0.8	-
63	2005	QUINDIO	0.6	0
66	2005	RISARALDA	0.6	0
88	2005	SAN ANDRES Y PROVIDENCIA	1	0
68	2005	SANTANDER	1	-
-	2005	SIN INFORMACIÓN	1	0
70	2005	SUCRE	1	-
-	2005	TÁCHIRA	1	0
73	2005	TOLIMA	0.7	0
76	2005	VALLE DEL CAUCA	0.7	-
97	2005	VAUPÉS	1	-
99	2005	VICHADA	1	-
-	2005	ZULIA	1	0
0	2006	-	1	0
91	2006	AMAZONAS	1	-
5	2006	ANTIOQUIA	0.7	-
-	2006	APURE	1	0
-	2006	ARAGUA	1	0
81	2006	ARAUCA	1	-
-	2006	ASUNCIÓN	1	0
8	2006	ATLÁNTICO	0.7	-
11	2006	BOGOTÁ D.C	0.8	0
13	2006	BOLÍVAR	0.8	-
15	2006	BOYACÁ	0.7	0
17	2006	CALDAS	0.7	0
18	2006	CAQUETÁ	1	-
85	2006	CASANARE	0.8	-
19	2006	CAUCA	0.8	-
20	2006	CESAR	0.8	-
27	2006	CHOCÓ	0.8	-
23	2006	CÓRDOBA	0.8	-
25	2006	CUNDINAMARCA	0.7	0
-	2006	DISTRITO CAPITAL	1	-
94	2006	GUAINÍA	1	-
95	2006	GUAVIARE	1	-
41	2006	HUILA	0.7	0
-	2006	IMBABURA	1	-
44	2006	LA GUAJIRA	0.7	0
-	2006	LAS PALMAS	1	0

47	2006	MAGDALENA	1	-
50	2006	META	1	-
52	2006	NARIÑO	1	-
54	2006	NORTE DE SANTANDER	1	-
-	2006	PICHINCHA	1	0
86	2006	PUTUMAYO	1	-
63	2006	QUINDIO	0.7	-
66	2006	RISARALDA	0.7	-
68	2006	SANTANDER	1	-
-	2006	SIN INFORMACIÓN	1	0
-	2006	SIN INFORMACIÓN (VENEZUELA)	1	0
70	2006	SUCRE	1	-
-	2006	TÁCHIRA	1	0
73	2006	TOLIMA	0.7	0
76	2006	VALLE DEL CAUCA	0.7	-
97	2006	VAUPÉS	1	-
99	2006	VICHADA	1	-
-	2006	ZULIA	1	0
0	2007	-	1	0
91	2007	AMAZONAS	0.8	-
5	2007	ANTIOQUIA	0.8	-
-	2007	APURE	1	0
81	2007	ARAUCA	1	-
8	2007	ATLÁNTICO	0.7	0
-	2007	BARINAS	1	0
11	2007	BOGOTÁ D.C	0.8	0
13	2007	BOLÍVAR	0.8	-
15	2007	BOYACÁ	0.7	0
17	2007	CALDAS	0.6	-
18	2007	CAQUETÁ	0.6	0
85	2007	CASANARE	0.7	-
19	2007	CAUCA	0.6	0
20	2007	CESAR	0.8	-
27	2007	CHOCÓ	1	-
23	2007	CÓRDOBA	0.8	-
25	2007	CUNDINAMARCA	0.7	-
-	2007	DISTRITO CAPITAL	1	-
-	2007	ESMERALDAS	1	0
94	2007	GUAINÍA	1	-
95	2007	GUAVIARE	0.6	0
41	2007	HUILA	0.7	0
44	2007	LA GUAJIRA	0.7	0
47	2007	MAGDALENA	1	-
50	2007	META	0.6	0
52	2007	NARIÑO	0.6	0
54	2007	NORTE DE SANTANDER	0.8	-
-	2007	PANAMÁ	1	0
-	2007	PICHINCHA	1	0
86	2007	PUTUMAYO	0.6	0

63	2007	QUINDIO	0.7	-
66	2007	RISARALDA	0.7	-
68	2007	SANTANDER	0.8	-
-	2007	SIN INFORMACIÓN	1	0
-	2007	SIN INFORMACIÓN (VENEZUELA)	1	0
70	2007	SUCRE	1	-
-	2007	TÁCHIRA	1	0
73	2007	TOLIMA	0.7	0
76	2007	VALLE DEL CAUCA	0.8	0
97	2007	VAUPÉS	1	-
99	2007	VICHADA	0.6	-
91	2008	AMAZONAS	0.8	-
5	2008	ANTIOQUIA	0.8	-
-	2008	APURE	1	0
81	2008	ARAUCA	1	-
8	2008	ATLÁNTICO	0.7	0
11	2008	BOGOTÁ D.C	0.8	0
13	2008	BOLÍVAR	1	-
15	2008	BOYACÁ	0.7	0
17	2008	CALDAS	0.7	0
18	2008	CAQUETÁ	1	-
85	2008	CASANARE	0.8	-
19	2008	CAUCA	0.8	-
20	2008	CESAR	1	-
27	2008	CHOCÓ	1	-
23	2008	CÓRDOBA	1	-
25	2008	CUNDINAMARCA	0.7	0
-	2008	DISTRITO FEDERAL	1	0
94	2008	GUAINÍA	1	-
95	2008	GUAVIARE	1	-
41	2008	HUILA	0.7	0
44	2008	LA GUAJIRA	0.7	0
47	2008	MAGDALENA	1	-
50	2008	META	0.8	-
52	2008	NARIÑO	0.8	-
54	2008	NORTE DE SANTANDER	1	-
-	2008	PICHINCHA	1	-
86	2008	PUTUMAYO	1	-
63	2008	QUINDIO	0.7	-
66	2008	RISARALDA	0.6	-
88	2008	SAN ANDRES Y PROVIDENCIA	1	0
68	2008	SANTANDER	0.8	-
-	2008	SIN INFORMACIÓN	1	0
-	2008	SIN INFORMACIÓN (VENEZUELA)	1	0
70	2008	SUCRE	1	-
-	2008	TÁCHIRA	1	0
73	2008	TOLIMA	0.7	0
76	2008	VALLE DEL CAUCA	0.7	0
97	2008	VAUPÉS	1	-

99	2008	VICHADA	0.7	-
91	2009	AMAZONAS	1	-
5	2009	ANTIOQUIA	0.7	0
-	2009	APURE	1	0
81	2009	ARAUCA	1	-
8	2009	ATLÁNTICO	0.8	0
11	2009	BOGOTÁ D.C	0.9	0
13	2009	BOLÍVAR	1	-
15	2009	BOYACÁ	0.8	0
17	2009	CALDAS	0.7	0
18	2009	CAQUETÁ	1	-
85	2009	CASANARE	0.8	-
19	2009	CAUCA	1	-
20	2009	CESAR	1	-
27	2009	CHOCÓ	1	-
23	2009	CÓRDOBA	1	-
25	2009	CUNDINAMARCA	0.7	-
-	2009	DISTRITO CAPITAL	1	-
-	2009	DISTRITO FEDERAL	1	0
94	2009	GUAINÍA	1	-
95	2009	GUAVIARE	1	-
41	2009	HUILA	0.7	0
-	2009	ÎLE-DE-FRANCE	1	0
44	2009	LA GUAJIRA	0.7	-
-	2009	LORETO	1	0
47	2009	MAGDALENA	1	-
-	2009	MÉRIDA	1	0
50	2009	META	0.8	-
-	2009	METROPOLITANA	1	0
52	2009	NARIÑO	1	-
54	2009	NORTE DE SANTANDER	0.7	-
-	2009	PANAMÁ	1	0
86	2009	PUTUMAYO	1	-
63	2009	QUINDIO	0.7	-
66	2009	RISARALDA	0.7	-
88	2009	SAN ANDRES Y PROVIDENCIA	1	0
68	2009	SANTANDER	0.8	-
-	2009	SIN INFORMACIÓN	1	0
-	2009	SIN INFORMACIÓN (VENEZUELA)	1	0
70	2009	SUCRE	1	-
-	2009	TÁCHIRA	1	-
73	2009	TOLIMA	0.7	0
-	2009	TRUJILLO	1	0
76	2009	VALLE DEL CAUCA	0.7	-
97	2009	VAUPÉS	1	-
99	2009	VICHADA	0.8	-
-	2009	ZULIA	1	-
91	2010	AMAZONAS	0.8	-
5	2010	ANTIOQUIA	0.7	0

81	2010	ARAUCA	1	-
8	2010	ATLÁNTICO	0.8	0
11	2010	BOGOTÁ D.C	0.9	0
13	2010	BOLÍVAR	0.8	-
15	2010	BOYACÁ	0.7	0
17	2010	CALDAS	0.7	0
18	2010	CAQUETÁ	1	-
-	2010	CARABOBO	1	0
85	2010	CASANARE	0.7	-
19	2010	CAUCA	0.8	-
20	2010	CESAR	0.8	-
27	2010	CHOCÓ	1	-
23	2010	CÓRDOBA	0.8	-
-	2010	CORTÉS	1	-
25	2010	CUNDINAMARCA	0.8	0
-	2010	DISTRITO CAPITAL	1	-
-	2010	FALCÓN	1	0
94	2010	GUAINÍA	1	-
95	2010	GUAVIARE	1	-
41	2010	HUILA	0.7	-
-	2010	IMBABURA	1	0
44	2010	LA GUAJIRA	0.8	-
47	2010	MAGDALENA	0.8	-
-	2010	MÉRIDA	1	0
50	2010	META	0.7	-
-	2010	MONAGAS	1	-
52	2010	NARIÑO	1	-
-	2010	NEW YORK	1	0
54	2010	NORTE DE SANTANDER	0.7	-
-	2010	PANAMÁ	1	0
-	2010	PICHINCHA	1	0
86	2010	PUTUMAYO	1	-
63	2010	QUINDIO	0.7	0
66	2010	RISARALDA	0.8	0
88	2010	SAN ANDRES Y PROVIDENCIA	1	0
68	2010	SANTANDER	0.7	-
-	2010	SIN INFORMACIÓN	1	0
70	2010	SUCRE	1	-
-	2010	SUCUMBÍOS	1	0
-	2010	TÁCHIRA	1	0
73	2010	TOLIMA	0.7	0
-	2010	TRUJILLO	1	0
76	2010	VALLE DEL CAUCA	0.6	0
97	2010	VAUPÉS	1	-
99	2010	VICHADA	1	-
-	2010	ZULIA	1	-
0	2011	(NA)	1	-
91	2011	AMAZONAS	0.8	-
5	2011	ANTIOQUIA	0.8	0

81	2011	ARAUCA	1	-
8	2011	ATLÁNTICO	0.9	0
11	2011	BOGOTÁ D.C	0.9	0
13	2011	BOLÍVAR	1	-
15	2011	BOYACÁ	0.7	0
17	2011	CALDAS	0.7	0
18	2011	CAQUETÁ	1	-
-	2011	CARABOBO	1	-
85	2011	CASANARE	1	-
19	2011	CAUCA	1	-
20	2011	CESAR	0.8	-
27	2011	CHOCÓ	1	-
-	2011	COJEDES	1	0
23	2011	CÓRDOBA	0.8	-
25	2011	CUNDINAMARCA	0.8	0
-	2011	DISTRITO FEDERAL	1	-
94	2011	GUAINÍA	1	-
95	2011	GUAVIARE	1	-
-	2011	GUAYAS	1	0
41	2011	HUILA	0.7	0
-	2011	IMBABURA	1	0
44	2011	LA GUAJIRA	0.7	-
-	2011	LIMA	1	0
47	2011	MAGDALENA	0.8	-
50	2011	META	0.7	0
-	2011	MÉXICO	1	0
52	2011	NARIÑO	0.8	-
54	2011	NORTE DE SANTANDER	0.8	-
-	2011	PANAMÁ	1	0
-	2011	PICHINCHA	1	0
86	2011	PUTUMAYO	1	-
63	2011	QUINDIO	0.7	0
66	2011	RISARALDA	0.7	0
88	2011	SAN ANDRES Y PROVIDENCIA	1	0
-	2011	SANTA CRUZ	1	0
68	2011	SANTANDER	0.8	-
-	2011	SIN INFORMACIÓN	1	0
70	2011	SUCRE	1	-
-	2011	TÁCHIRA	1	0
73	2011	TOLIMA	0.8	0
76	2011	VALLE DEL CAUCA	0.6	0
97	2011	VAUPÉS	1	-
99	2011	VICHADA	0.8	-
-	2011	ZULIA	1	-
91	2012	AMAZONAS	0.8	-
5	2012	ANTIOQUIA	0.8	0
-	2012	APURE	1	0
-	2012	ARAGUA	1	0
81	2012	ARAUCA	1	-

8	2012	ATLÁNTICO	0.8	0
11	2012	BOGOTÁ D.C	0.9	0
13	2012	BOLÍVAR	0.7	-
15	2012	BOYACÁ	0.7	0
-	2012	BRITISH COLUMBIA	1	-
17	2012	CALDAS	0.7	0
18	2012	CAQUETÁ	1	-
-	2012	CARCHI	1	-
85	2012	CASANARE	0.7	-
19	2012	CAUCA	1	-
20	2012	CESAR	0.8	-
27	2012	CHOCÓ	0.8	-
23	2012	CÓRDOBA	1	-
-	2012	CORTÉS	1	0
25	2012	CUNDINAMARCA	0.8	0
-	2012	DISTRITO CAPITAL	1	-
-	2012	DISTRITO FEDERAL	1	0
-	2012	ESMERALDAS	1	0
94	2012	GUAINÍA	1	-
95	2012	GUAVIARE	1	-
41	2012	HUILA	0.8	0
-	2012	ÎLE-DE-FRANCE	1	0
-	2012	JALISCO	1	0
44	2012	LA GUAJIRA	0.7	-
47	2012	MAGDALENA	0.7	-
-	2012	MÉRIDA	1	0
50	2012	META	0.8	-
-	2012	MÉXICO	1	0
-	2012	MICHOACÁN	1	0
52	2012	NARIÑO	0.8	-
54	2012	NORTE DE SANTANDER	0.8	-
-	2012	PANAMÁ	1	0
-	2012	PICHINCHA	1	0
86	2012	PUTUMAYO	1	-
63	2012	QUINDIO	0.8	0
66	2012	RISARALDA	0.8	0
88	2012	SAN ANDRES Y PROVIDENCIA	1	0
68	2012	SANTANDER	1	0
-	2012	SIN INFORMACIÓN	1	0
-	2012	SIN INFORMACIÓN (VENEZUELA)	1	0
70	2012	SUCRE	1	0
-	2012	TÁCHIRA	1	0
73	2012	TOLIMA	0.8	-
-	2012	TRUJILLO	1	0
76	2012	VALLE DEL CAUCA	0.8	0
97	2012	VAUPÉS	1	-
99	2012	VICHADA	1	-
-	2012	VIRGINIA	1	0
-	2012	YARACUY	1	0

-	2012	ZULIA	1	0
91	2013	AMAZONAS	0.8	-
5	2013	ANTIOQUIA	0.7	-
81	2013	ARAUCA	1	0
8	2013	ATLÁNTICO	0.8	0
11	2013	BOGOTÁ D.C	0.9	0
13	2013	BOLÍVAR	0.7	0
-	2013	BONAIRE	1	0
15	2013	BOYACÁ	0.8	0
17	2013	CALDAS	0.8	0
18	2013	CAQUETÁ	1	-
85	2013	CASANARE	0.8	-
19	2013	CAUCA	1	-
20	2013	CESAR	0.8	-
27	2013	CHOCÓ	1	-
23	2013	CÓRDOBA	1	-
25	2013	CUNDINAMARCA	0.8	0
-	2013	ESMERALDAS	1	0
-	2013	FLORIDA	1	0
94	2013	GUAINÍA	1	-
95	2013	GUAVIARE	1	0
-	2013	GUAYAS	1	0
41	2013	HUILA	0.7	0
44	2013	LA GUAJIRA	0.7	-
-	2013	LA PAZ	1	0
-	2013	LARA	1	0
-	2013	LIMA	1	0
47	2013	MAGDALENA	0.7	0
50	2013	META	0.7	-
-	2013	METROPOLITANA	-	0
52	2013	NARIÑO	1	-
54	2013	NORTE DE SANTANDER	0.8	-
-	2013	NORTH CAROLINA	1	0
-	2013	PANAMÁ	1	0
-	2013	PICHINCHA	1	0
86	2013	PUTUMAYO	1	-
63	2013	QUINDIO	0.8	0
66	2013	RISARALDA	0.8	0
88	2013	SAN ANDRES Y PROVIDENCIA	1	0
68	2013	SANTANDER	1	-
-	2013	SÃO PAULO	1	0
-	2013	SIN INFORMACIÓN	1	0
-	2013	SIN INFORMACIÓN (VENEZUELA)	1	0
70	2013	SUCRE	1	-
-	2013	TÁCHIRA	1	0
73	2013	TOLIMA	0.8	0
76	2013	VALLE DEL CAUCA	0.8	0
97	2013	VAUPÉS	1	-
99	2013	VICHADA	0.7	-

0	2014	(NA)	1	0
91	2014	AMAZONAS	1	-
5	2014	ANTIOQUIA	0.8	0
81	2014	ARAUCA	1	-
8	2014	ATLÁNTICO	0.8	0
-	2014	BARINAS	1	0
11	2014	BOGOTÁ D.C	0.9	0
13	2014	BOLÍVAR	0.7	0
15	2014	BOYACÁ	0.8	0
17	2014	CALDAS	0.8	0
18	2014	CAQUETÁ	1	-
85	2014	CASANARE	0.7	0
19	2014	CAUCA	1	-
20	2014	CESAR	0.8	0
27	2014	CHOCÓ	1	-
-	2014	COCLÉ	1	0
23	2014	CÓRDOBA	1	-
25	2014	CUNDINAMARCA	0.8	0
-	2014	CUSCO	1	0
-	2014	DISTRITO FEDERAL	1	0
-	2014	ESMERALDAS	1	0
94	2014	GUAINÍA	1	-
95	2014	GUAVIARE	1	-
-	2014	GUAYAS	1	0
-	2014	HESSEN	1	-
41	2014	HUILA	0.7	0
44	2014	LA GUAJIRA	0.7	0
-	2014	LAMBAYEQUE	1	0
-	2014	LIMA	1	0
47	2014	MAGDALENA	0.7	0
-	2014	MANABÍ	1	0
50	2014	META	0.7	-
-	2014	MÉXICO	1	0
-	2014	MIRANDA	1	0
52	2014	NARIÑO	0.7	-
54	2014	NORTE DE SANTANDER	0.5	0
-	2014	PICHINCHA	1	0
86	2014	PUTUMAYO	1	-
63	2014	QUINDIO	0.8	0
66	2014	RISARALDA	0.8	0
88	2014	SAN ANDRES Y PROVIDENCIA	1	0
-	2014	SANTA CRUZ	1	0
68	2014	SANTANDER	1	-
-	2014	SIN INFORMACIÓN	1	0
-	2014	SIN INFORMACIÓN (VENEZUELA)	1	0
70	2014	SUCRE	1	-
-	2014	TÁCHIRA	1	0
-	2014	TARAPACÁ	1	0
73	2014	TOLIMA	0.8	0

76	2014	VALLE DEL CAUCA	0.7	0
99	2014	VICHADA	0.7	-
-	2014	ZULIA	1	-
91	2015	AMAZONAS	0.8	-
5	2015	ANTIOQUIA	0.7	0
81	2015	ARAUCA	1	-
8	2015	ATLÁNTICO	0.8	0
-	2015	BARINAS	1	-
11	2015	BOGOTÁ D.C	0.9	0
13	2015	BOLÍVAR	0.7	-
15	2015	BOYACÁ	0.8	0
17	2015	CALDAS	0.8	0
18	2015	CAQUETÁ	1	-
85	2015	CASANARE	0.8	-
19	2015	CAUCA	1	-
20	2015	CESAR	1	-
27	2015	CHOCÓ	1	-
23	2015	CÓRDOBA	1	-
25	2015	CUNDINAMARCA	0.8	0
-	2015	DISTRITO CAPITAL	1	-
-	2015	DISTRITO FEDERAL	1	0
-	2015	FALCÓN	1	0
-	2015	FLORIDA	1	0
95	2015	GUAVIARE	1	-
-	2015	GUERRERO	1	-
41	2015	HUILA	0.8	0
-	2015	ÎLE-DE-FRANCE	1	-
-	2015	IMBABURA	1	0
44	2015	LA GUAJIRA	0.7	-
-	2015	LIMA	1	0
47	2015	MAGDALENA	0.6	0
-	2015	MÉRIDA	1	0
50	2015	META	0.8	-
-	2015	METROPOLITANA	1	0
52	2015	NARIÑO	1	0
54	2015	NORTE DE SANTANDER	0.8	-
-	2015	PANAMÁ	1	0
86	2015	PUTUMAYO	1	-
63	2015	QUINDIO	0.7	0
-	2015	RIO DE JANEIRO	1	-
66	2015	RISARALDA	0.8	0
88	2015	SAN ANDRES Y PROVIDENCIA	1	-
68	2015	SANTANDER	1	0
-	2015	SIN INFORMACIÓN	1	-
-	2015	SIN INFORMACIÓN (VENEZUELA)	1	0
70	2015	SUCRE	1	-
-	2015	TÁCHIRA	1	0
73	2015	TOLIMA	0.8	0
76	2015	VALLE DEL CAUCA	0.8	0

99	2015	VICHADA	0.8	-
-	2015	ZULIA	1	0
91	2016	AMAZONAS	0.8	-
5	2016	ANTIOQUIA	0.8	0
-	2016	APURE	1	0
81	2016	ARAUCA	1	-
8	2016	ATLÁNTICO	0.8	0
11	2016	BOGOTÁ D.C	0.9	0
13	2016	BOLÍVAR	1	-
15	2016	BOYACÁ	0.8	0
17	2016	CALDAS	0.8	0
18	2016	CAQUETÁ	1	-
85	2016	CASANARE	0.8	-
19	2016	CAUCA	1	-
20	2016	CESAR	1	-
27	2016	CHOCÓ	1	-
23	2016	CÓRDOBA	1	-
25	2016	CUNDINAMARCA	0.8	0
95	2016	GUAVIARE	1	-
-	2016	GUAYAS	1	-
41	2016	HUILA	0.8	0
44	2016	LA GUAJIRA	0.6	-
-	2016	LIMA	1	0
-	2016	LOJA	1	0
47	2016	MAGDALENA	0.7	-
50	2016	META	1	-
-	2016	METROPOLITANA	1	0
-	2016	MIRANDA	1	0
-	2016	MORELOS	1	0
52	2016	NARIÑO	1	-
54	2016	NORTE DE SANTANDER	0.8	-
-	2016	NUEVO LEÓN	1	0
-	2016	PANAMÁ	1	0
-	2016	PICHINCHA	1	0
86	2016	PUTUMAYO	1	0
63	2016	QUINDIO	0.8	0
-	2016	RIO DE JANEIRO	1	0
66	2016	RISARALDA	0.9	0
88	2016	SAN ANDRES Y PROVIDENCIA	1	0
-	2016	SAN MARCOS	1	0
68	2016	SANTANDER	1	0
-	2016	SÃO PAULO	1	0
-	2016	SIN INFORMACIÓN	1	0
-	2016	SIN INFORMACIÓN (VENEZUELA)	1	0
70	2016	SUCRE	1	-
-	2016	TÁCHIRA	1	0
73	2016	TOLIMA	0.8	0
76	2016	VALLE DEL CAUCA	0.8	0
97	2016	VAUPÉS	1	-

99	2016	VICHADA	0.8	-
5	2017	ANTIOQUIA	1	-
81	2017	ARAUCA	1	-
11	2017	BOGOTÁ D.C	1	-
13	2017	BOLÍVAR	1	-
15	2017	BOYACÁ	1	-
17	2017	CALDAS	1	-
18	2017	CAQUETÁ	1	-
19	2017	CAUCA	1	-
27	2017	CHOCÓ	1	-
23	2017	CÓRDOBA	1	-
95	2017	GUAVIARE	1	-
41	2017	HUILA	1	-
44	2017	LA GUAJIRA	1	-
47	2017	MAGDALENA	1	-
54	2017	NORTE DE SANTANDER	1	-
86	2017	PUTUMAYO	1	-
63	2017	QUINDIO	1	-
66	2017	RISARALDA	1	-
68	2017	SANTANDER	1	-
73	2017	TOLIMA	1	-
76	2017	VALLE DEL CAUCA	1	-
5	2018	ANTIOQUIA	1	-
8	2018	ATLÁNTICO	1	-
15	2018	BOYACÁ	1	-
85	2018	CASANARE	1	-
41	2018	HUILA	1	-
50	2018	META	1	-
52	2018	NARIÑO	1	-
66	2018	RISARALDA	1	-
68	2018	SANTANDER	1	-
-	2018	TÁCHIRA	1	0
76	2018	VALLE DEL CAUCA	1	-
5	2019	ANTIOQUIA	1	-
81	2019	ARAUCA	1	-
11	2019	BOGOTÁ D.C	1	-
95	2019	GUAVIARE	1	-
-	2019	MÉRIDA	1	0
52	2019	NARIÑO	1	-
63	2019	QUINDIO	1	-
68	2019	SANTANDER	1	-
-	2019	SIN INFORMACIÓN	1	0
73	2019	TOLIMA	1	1
76	2019	VALLE DEL CAUCA	1	1
5	2020	ANTIOQUIA	1	1
15	2020	BOYACÁ	1	1
17	2020	CALDAS	1	1
18	2020	CAQUETÁ	-	-
19	2020	CAUCA	1	1

20	2020	CESAR	1	1
23	2020	CÓRDOBA	1	1
50	2020	META	1	1
54	2020	NORTE DE SANTANDER	1	1
68	2020	SANTANDER	1	1
76	2020	VALLE DEL CAUCA	1	1

9.4.3 Additional filters

```
cat(readLines(files$merge), sep = "\n")
```

```
## #!/usr/bin/env Rscript --vanilla
## # vim:set expandtab ts=4 sw=4 ai fileencoding=utf-8:
## #
## # Author: PB
## # Maintainer(s): PB
## # License: (c) HRDAG 2022, GPL v2 or newer
## #
## # -----------------------------------------------------------
## # CO-SIVJRNR-data/match/fase4/merge/src/merge.R
##
## pacman::p_load(argparse, arrow, logger, tidyr, assertr, stringr,
##                dplyr, here, data.table, nnet)
##
## stopifnot(str_detect(Sys.getlocale(), "en_US.UTF-8"))
## stopifnot(str_detect(getwd(), "merge$"))
## set.seed(19481210)
## log_threshold(DEBUG)
##
## vln_types <- c("exilio", "homicidio", "secuestro",
##                "desaparicion", "desplazamiento", "reclutamiento")
##
## perps <- c("AGENTES_ESTATALES", "GRUPOS_POSDESMV_PARAMILITAR",
##            "GUERRILLA", "GUERRILLA_ELN", "GUERRILLA_FARC",
##            "GUERRILLA_OTRA", "OTRO", "PARAMILITARES")
##
## to_rename <- c("yy_hecho", "mm_hecho", "dd_hecho",
##                "dept_code_hecho", "muni_code_hecho",
##                paste0("perp_", perps))
##
## recflds <- c(
##     "recordid",
##     "sexo",
##     "edad",
##     "edad_categoria",
##     to_rename)
##
##
## getmode <- function(x) {
##     # from StackOverflow, w nnet::which.is.max to break ties at random
##     # * prioritizes
```

```
##     #     full munis (e.g., 5012) over dptos w no muni (e.g., 5000)
##     #     full dates (e.g., 20010626) over ymds w no month-day (e.g., 20010000)
##     if (length(x) == 1) return(x[1])
##     x <- x[!is.na(x)]
##     if (length(x) == 1) return(x[1])
##     x2 <- str_subset(x, pattern="000$", negate=TRUE)
##     if (length(x2) == 1) return(x2[1])
##     if (length(x2) > 1) x <- x2
##     if (length(x) == 2) return(sample(x, size = 1))
##     ux <- unique(x)
##     if (length(ux) == 1) return(ux[1])
##     if (length(ux) == 2) return(sample(ux, size = 1))
##     ux[which.is.max(tabulate(match(x, ux)))]
## }
##
##
##
## getargs <- function() {
##     # NB: we only need to pass the output.
##     parser <- argparse::ArgumentParser()
##     parser$add_argument("--input_records",
##                         default = here("match/fase4/import/output/input-records.parquet"))
##     parser$add_argument("--match_groups",
##                         default =  here("match/fase4/cluster-py/output/match-groups.parquet"))
##     parser$add_argument("--fgn_to_filter",
##                         default = here::here("individual/FGN/homcon/output/fgn-to-filter.parquet"))
##     parser$add_argument("--fgn",
##                         default = here::here("individual/FGN/export/output/fgn.parquet"))
##     parser$add_argument("--cnmh_to_filter",
##                         default = here::here("individual/CNMH/filter-archivado/output/cnmh-to-filter
##     parser$add_argument("--rpv",
##                         default = here("match/fase4/merge/output/recordid-p-violation.parquet"))
##     parser$add_argument("--vtype",
##                         default = "promoted")
##     parser$add_argument("--output",
##                         default = here("match/fase4/merge/output/mrd-promoted.parquet"))
##     args <- parser$parse_args()
##
##     args$vtype <- str_sub(str_extract(args$output, "-[a-z]+\\."), 2, -2)
##     args$log <- paste0("output/mrd-", args$vtype, ".log")
##     args
## }
##
##
## get_ir <- function(args) {
##     log_info("data read for {args$vtype} begins")
##     rpv <- read_parquet(args$rpv) %>%
##         filter(vtype == args$vtype) %>%
##         select(recordid)
##     log_debug("got rpv with {nrow(rpv)} recs")
##     ir <- read_parquet(args$input_records) %>%
##         filter(recordid %in% rpv$recordid) %>%
##         left_join(read_parquet(args$match_groups), by='recordid') %>%
##         assert(not_na, match_group_id) %>%
```

```
##         verify(nrow(.) == nrow(rpv))
##     log_info("retained {nrow(ir)} input-recs for vtype={args$vtype}")
##     rm(rpv)
##     ir
## }
##
## rename_and_select <- function(args, ir, vln) {
##     # rename here xforms the cols from their vln_specific names to canonical
##     # colnames
##     vlnt <- paste0("_", vln)
##     cnmh_to_filter <- read_parquet(args$cnmh_to_filter)
##     fgn_to_filter <- read_parquet(args$fgn_to_filter)
##     rns <- ir %>%
##         anti_join(fgn_to_filter) %>%
##         anti_join(cnmh_to_filter) %>%
##         rename_with(~ str_replace(., vlnt, ""), col=all_of(to_rename)) %>%
##         select(all_of(recflds), is_conflict, match_group_id) %>%
##         mutate(ymd_hecho = str_c(
##             str_pad(str_replace_na(yy_hecho, "0000"), width=4, pad="0"),
##             str_pad(str_replace_na(mm_hecho, "00"), width=2, pad="0"),
##             str_pad(str_replace_na(dd_hecho, "00"), width=2, pad="0"))) %>%
##         mutate(ymd_hecho = if_else(startsWith(ymd_hecho, "0"),
##                                    paste0(yy_hecho, "0000"),
##                                    ymd_hecho)) %>%
##         select(-yy_hecho, -mm_hecho, -dd_hecho) %>%
##         mutate(ymd_hecho = na_if(ymd_hecho, "NA0000")) %>%
##         mutate(dept_code_hecho = replace_na(dept_code_hecho, 0)) %>%
##         mutate(muni_code_hecho = if_else(is.na(muni_code_hecho),
##                                          paste0(dept_code_hecho, "000"),
##                                          as.character(muni_code_hecho))) %>%
##         mutate(muni_code_hecho = na_if(muni_code_hecho, "0000")) %>%
##         select(-dept_code_hecho, -recordid) %>%
##         mutate(across(c(edad), as.character))
##     log_debug("rename_and_select returning {nrow(rns)} for {vln}")
##     rns
## }
##
## merge_mg <- function(irs) {
##     log_debug("starting merge")
##     irm <- irs %>%
##         mutate(across(starts_with("perp_"), ~ replace_na(., -1))) %>%
##         group_by(match_group_id) %>%
##         summarize(
##             sexo = getmode(sexo),
##             edad = getmode(edad),
##             edad_categoria = getmode(edad_categoria),
##             muni_code_hecho = getmode(muni_code_hecho),
##             ymd_hecho = min(ymd_hecho, na.rm=TRUE),
##             is_conflict = max(is_conflict, na.rm=TRUE),
##             perp_AGENTES_ESTATALES = max(perp_AGENTES_ESTATALES),
##             perp_GRUPOS_POSDESMV_PARAMILITAR = max(perp_GRUPOS_POSDESMV_PARAMILITAR),
##             perp_GUERRILLA = max(perp_GUERRILLA),
##             perp_GUERRILLA_ELN = max(perp_GUERRILLA_ELN),
##             perp_GUERRILLA_FARC = max(perp_GUERRILLA_FARC),
```

```
##             perp_GUERRILLA_OTRA = max(perp_GUERRILLA_OTRA),
##             perp_OTRO = max(perp_OTRO),
##             perp_PARAMILITARES = max(perp_PARAMILITARES)) %>%
##         ungroup() %>%
##         mutate(across(starts_with("perp_"), ~ na_if(., -1)))
##     log_debug("summarize complete")
##     irm
## }
##
## #---main-----
## args <- getargs()
## log_info("path & locale ok, seed set, log open, run begins.")
##
## ir <- get_ir(args)
## log_info("read {nrow(args)} recs from input-records")
##
## if (args$vtype == "promoted") {
##     # NB: this pulls the same input-recs 4x, once for each promoted vln type
##     # it has to xfrom the vln-specific cols into canonical cols: wide->long
##     # yy_hecho_{vln}, dept_code_hecho_{vln}, perp_AGENTES_ESTATALES_{vln} etc.
##     # once the vln-specific cols are in long form, they can be collapsed on mgi
##     irs <- bind_rows(
##         rename_and_select(args, ir, "homicidio"),
##         rename_and_select(args, ir, "reclutamiento"),
##         rename_and_select(args, ir, "secuestro"),
##         rename_and_select(args, ir, "desaparicion"))
## } else {
##     irs <- rename_and_select(args, ir, args$vtype)
## }
## log_info("irs has {nrow(irs)} recs.")
##
## # irs has canonically named cols so can do a collapse w merge_mg()
## irm <- merge_mg(irs) %>%
##     mutate(muni_code = as.numeric(muni_code_hecho)) %>%
##     mutate(dept_code_hecho = as.integer(muni_code/1000)) %>%
##     mutate(psum = rowSums(across(starts_with("perp_")))) %>%
##     mutate(ymd_int = as.numeric(ymd_hecho)) %>%
##     mutate(yy_hecho = as.integer(ymd_int/10000)) %>%
##     write_parquet(args$output)
## log_info("done with merge, output written")
## log_info("running tests now")
##
##
## # For testing only
## # test that the perps were not conflated
## stopifnot(sum(irm$perp_AGENTES_ESTATALES, na.rm=T) != sum(irm$perp_PARAMILITARES, na.rm=T))
## # test that >80% of irm recs (w known perp) have 1 perp
## stopifnot(between(mean(irm$psum, na.rm=T), 1.01, 2))
## stopifnot(0 <= min(irm$psum, na.rm=T) & max(irm$psum, na.rm=T) <= 8)
## log_info("perp tests ok: are separate, not conflated, no overcounts")
##
## x <- irm %>%
##     assert(within_bounds(1000, 99999), muni_code) %>%
##     select(-muni_code) %>%
```

```
##     assert(within_bounds(19000000, 20200000), ymd_int) %>%
##     select(-ymd_int) %>%
##     assert(not_na, yy_hecho, dept_code_hecho)
## rm(x)
## log_info("test ok: muni_code ok, ymd ok, not_na yy+dept")
##
## log_info("done.")
## # done.
```

```
cat(readLines(files$makei), sep = "\n")
```

```
## #!/usr/bin/env Rscript --vanilla
## # vim: set expandtab ts=4 sw=4 ai fileencoding=utf-8
## #
## # Author: PB
## # Maintainer(s): PB
## # License: (c) HRDAG 2021, GPL v2 or newer
## #
## # -----------------------------------------------------------
## # CO-SIVJRNR-data/match/fase4/merge/src/make-i-vecs.R
## #
##
## pacman::p_load(argparse, arrow, logger, tidyr, assertr, stringr, yaml,
##                purrr, dplyr, glue, here, data.table)
## stopifnot(str_detect(Sys.getlocale(), "en_US.UTF-8"))
## stopifnot(str_detect(getwd(), "merge$"))
## set.seed(19481210)
##
##
## vln_types <- c("exilio", "desplazamiento",
##                "homicidio", "secuestro", "desaparicion", "reclutamiento")
## promoted_vlns <- vln_types[3:length(vln_types)]
##
## getargs <- function() {
##     parser <- argparse::ArgumentParser()
##     parser$add_argument("--recordid_p_violation",
##                         default = "output/recordid-p-violation.parquet")
##     parser$add_argument("--match_groups",
##                         default = here("match/fase4/cluster-py/output/match-groups.parquet"))
##     parser$add_argument("--match_group_violation",
##                         default = here::here("match/fase4/merge/output/match-group-violation.parquet
##     parser$add_argument("--cnmh_to_filter",
##                         default = here::here("individual/CNMH/filter-archivado/output/cnmh-to-filter
##     parser$add_argument("--fgn_to_filter",
##                         default = here::here("individual/FGN/homcon/output/fgn-to-filter.parquet"))
##     parser$add_argument("--fgn",
##                         default = here::here("individual/FGN/export/output/fgn.parquet"))
##     parser$add_argument("--src2in",
##                         default = here::here("match/fase4/merge/hand/src2in.yaml"))
##     parser$add_argument("--output",
##                         default = "output/inclusion-vecs-desaparicion.parquet")
##     args <- parser$parse_args()
##     args$vln <- str_sub(str_extract(args$output, "-[a-z]+\\."), 2, -2)
##     stopifnot(args$vln %in% vln_types)
```

```
##     args$log <- paste0("output/make-i-vecs-", args$vln, ".log")
##     args
## }
##
##
## # --- main -----
## args <- getargs()
##
## log_appender(appender_tee(args$log))
## log_threshold(DEBUG)
## log_info("path & locale ok, seed set, log open, {args$vln} run begins.")
##
## #---setup-----
## in_sources <- yaml.load_file(args$src2in)
## # need to test that all sources in input_records are mapped in in_sources!
##
## if (args$vln %in% promoted_vlns) {
##     target_vln <- "promoted"
## } else {
##     target_vln <- args$vln
## }
##
## # filter for unusable recs here
## cnmh_to_filter <- read_parquet(args$cnmh_to_filter)
## log_info("read {nrow(cnmh_to_filter)} rows to filter from CNMH")
## fgn_to_filter <- read_parquet(args$fgn_to_filter)
## log_info("read {nrow(fgn_to_filter)} rows to filter from FGN")
##
## # recordids, filtered to the appropriate violation level, with source
## input_records <- read_parquet(args$recordid_p_violation) %>%
##     filter(vtype == target_vln) %>%
##     mutate(source = str_replace(source, regex("-"), "_")) %>%
##     select(recordid, source) %>%
##     assert(is_uniq, recordid)
## log_info("read {nrow(input_records)} input-records of type {target_vln}")
##
## input_records <- input_records %>%
##     anti_join(fgn_to_filter)
## log_info("after FGN filter, {nrow(input_records)} input-records retained")
##
## input_records <- input_records %>%
##     anti_join(cnmh_to_filter)
## log_info("after CNMH filter, {nrow(input_records)} input-records retained")
##
## stopifnot(input_records %>% filter(recordid %in% fgn_to_filter$recordid) %>% nrow(.) == 0)
## stopifnot(input_records %>% filter(recordid %in% cnmh_to_filter$recordid) %>% nrow(.) == 0)
## log_info("input_records has no recs that should be filtered")
##
## # match-group-ids, filtered for vln == match-group's _already promoted_ vln
## # vln is in {homicidio, reclutamiento, secuestro, desaparicion}
## # match_group_violation filters match_groups for the appropriate vln type
## match_groups <- read_parquet(args$match_group_violation) %>%
##     filter(violation == args$vln) %>%
##     select(match_group_id) %>%
```

```
##     distinct() %>%
##     left_join(read_parquet(args$match_groups)) %>%
##     distinct()
## n_mgs <- length(unique(match_groups$match_group_id))
## log_info("read {nrow(match_groups)} recordids in {n_mgs} match-groups for {args$vln}")
##
## # filtering. Note that when we're filtering the promoted recs, the number of recs
## # filtered might be greater than the number expected for a specific violation.
## match_groups <- match_groups %>%
##     anti_join(fgn_to_filter)
## n_mgs <- length(unique(match_groups$match_group_id))
## log_info("after FGN filter, {nrow(match_groups)} recordids and {n_mgs} match_groups retained")
## match_groups <- match_groups %>%
##     anti_join(cnmh_to_filter)
## n_mgs <- length(unique(match_groups$match_group_id))
## log_info("after CNMH filter, {nrow(match_groups)} recordids and {n_mgs} match_groups retained")
## n_mgs <- length(unique(match_groups$match_group_id))
## log_info("read {nrow(match_groups)} recordids in {n_mgs} match-groups after FGN- and CNMH-to-filter
## stopifnot(match_groups %>% filter(recordid %in% fgn_to_filter$recordid) %>% nrow(.) == 0)
## stopifnot(match_groups %>% filter(recordid %in% cnmh_to_filter$recordid) %>% nrow(.) == 0)
## log_info("match_groups has no recs that should be filtered")
##
## # in this match, some match-groups include recordids that need to be filtered
## # because they're not of the correct type. This is a consequence of the crude
## # filtering in the match_groups step: recall that (match_group_id, violation)
## # in match-group-violation is *not unique*. Within a match_group_id there
## # may be several violations. the filter in the step below removes the irrelevant
## # recordids.
## in_vecs <- match_groups %>%
##     inner_join(input_records, by="recordid") %>%
##     filter(!is.na(source))
## log_info("in_vecs initialized w {nrow(in_vecs)}, before in_ vars and summary")
##
## stopifnot(in_vecs %>% filter(recordid %in% fgn_to_filter$recordid) %>% nrow(.) == 0)
## stopifnot(in_vecs %>% filter(recordid %in% cnmh_to_filter$recordid) %>% nrow(.) == 0)
## log_info("in_vecs has no recs that should be filtered")
## in_vecs <- in_vecs %>% select(-recordid)
## rm(input_records, match_groups)
##
##
## log_info("starting str_detect:")
## for (i in seq(length(in_sources))) {
##     in_srcs <- in_sources[[i]]
##     in_srcs_char <- str_c(in_srcs, collapse = ', ')
##     in_name <- paste0("in_", names(in_sources)[i])
##     log_info("{in_srcs_char} to {in_name}")
##     in_vecs <- in_vecs %>%
##         select(source) %>%
##         mutate(in_ = as.integer(source %in% in_srcs)) %>%
##         select(-source) %>%
##         rename(!!in_name := in_) %>%
##         bind_cols(in_vecs)
## }
##
```

```
## in_vecs <- in_vecs %>% select(-source) %>% as.data.table
## log_info("as.data.table.")
## # if (args$vln != "homicidio") {
## #     in_vecs <- in_vecs %>% filter(in_FGN == 0) %>% select(-in_FGN)
## # }
##
## log_info("starting sum by match_group_id")
## in_vecs <- in_vecs[, lapply(.SD, sum), by=match_group_id]
##
## print(glue("sum(in_FGN) == {sum(in_vecs$in_FGN)}"))
## # if (args$vln == "homicidio") {
## #     stopifnot("in_FGN" %in% colnames(in_vecs))
## # } else {
## #     print(sum(in_vecs$in_FGN))
## #     stopifnot(! "in_FGN" %in% colnames(in_vecs))
## # }
##
## in_vecs %>% write_parquet(args$output)
## log_info("wrote {nrow(in_vecs)} to {args$output}; done.")
## # done.
```

References

Akande, Olanrewaju Michael. 2015. "A Comparison of Multiple Imputation Methods for Categorical Data." Master's thesis, Duke University.

Amorós, Jaume. 2014. "Recapturing Laplace." *Significance*. https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.1740-9713.2014.00754.x.

Ball, Patrick. 2008. *Quién Le Hizo Qué a Quién*. Benetech.

———. 2016. "The Task Is a Quantum of Workflow." https://hrdag.org/2016/06/14/the-task-is-a-quantum-of-workflow/.

Ball, Patrick, and Megan Price. 2018. "The Statistics of Genocide." *CHANCE* 31 (1): 38–45.

———. 2019. "Using Statistics to Assess Lethal Violence in Civil and Inter-State War." *Annual Review of Statistics and Its Application* 6: 63–84.

Berchenko, Yakir, Yael Artzy-Randrup, Mina Teicher, and Lewi Stone. 2009. "Emergence and Size of the Giant Component in Clustered Random Graphs with a Given Degree Distribution." *Phys. Rev. Lett.* 102 (March): 138701. https://doi.org/10.1103/PhysRevLett.102.138701.

Bilenko, Mikail, Beena Kamath, and Raymond J Mooney. 2006. "Adaptive Blocking: Learning to Scale up Record Linkage." In. Proceedings of the 6th IEEE International Conference on Data Mining. http://bilenko.com/papers/06-icdm.pdf.

Bird, Sheila M, and Ruth King. 2018. "Multiple Systems Estimation (or Capture-Recapture Estimation) to Inform Public Policy." *Annual Review of Statistics and Its Application* 5: 95–118.

Bishop, Yvonne M M, Stephen E Fienberg, and Paul W Holland. 1975. *Discrete Multivariate Analysis: Theory and Practice*. Springer.

Burgette, Lane F, and Jerome P Reiter. 2010. "Multiple Imputation for Missing Data via Sequential Regression Trees." *American Journal of Epidemiology* 172 (9): 1070--6.

Chao, Anne. 1987. "Estimating the Population Size for Capture-Recapture Data with Unequal Capturability." *Biometrics* 43: 783–91.

———. 2001. "An Overview of Closed Capture-Recapture Models." *Journal of Agricultural, Biological, and Environmental Statistics* 6 (2): 158–75.

———. 2015. "Capture-Recapture for Human Populations." *Wiley StatsRef: Statistics Reference Online*, 1–16.

Chen, Tianqi, and Carlos Guestrin. 2016. "XGBoost: A Scalable Tree Boosting System." *CoRR* abs/1603.02754. http://arxiv.org/abs/1603.02754.

Christen, Peter. 2012. *Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection*. New York: Springer.

Christen, Peter, Dinusha Vatsalan, and Qing Wang. 2015. "Efficient Entity Resolution with Adaptive and Interactive Training Data Selection." In *2015 IEEE International Conference on Data Mining*. http://users.cecs.anu.edu.au/~u5170295/papers/icdm-christen-2015.pdf.

Cormack, R M. 1989. "Log-Linear Models for Capture-Recapture." *Biometrics* 45 (2): 395–413. http://www.jstor.org/stable/2531485.

Cro, S. 2017. "Relevant Accessible Sensitivity Analysis for Clinical Trials with Missing Data." PhD thesis, London School of Hygiene & Tropical Medicine.

Darroch, John N, Stephen E Fienberg, Gary F V Glonek, and Brian W Junker. 1993. "A Three-Sample Multiple-Recapture Approach to Census Population Estimation with Heterogeneous Catchability." *Journal of the American Statistical Association* 88 (423): 1137–48. http://www.jstor.org/stable/2290811.

Fienberg, Stephen E, M S Johnson, and B W Junker. 1999. "Classical multilevel and Bayesian approaches to population size estimation using multiple lists." *Journal of the Royal Statistical Society Series A* 162 (3): 383–405. https://doi.org/10.1111/1467-985X.00143.

Gelman, Andrew, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. 2013. *Bayesian Data Analysis, Third Edition*. Chapman; Hall/CRC.

Gelman, Andrew, and E. Loken. 2014. "The Statistical Crisis in Science." *American Scientist*.

Gutiérrez-Sanín, F., and E. Wood. 2017. "What Should We Mean by 'Pattern of Political Violence'? Repertoire, Targeting, Frequency, and Technique." *Cambridge Core* 15.

Hagberg, Aric A., Daniel A. Schult, and Pieter J. Swart. 2008. *Exploring Network Structure, Dynamics, and Function Using NetworkX*. Edited by Ga"el Varoquaux, Travis Vaught, and Jarrod Millman. Pasadena, CA USA.

Herzog, Thomas N, Fritz J Scheuren, and Wiliam E Winkler. 2007. *Data Quality and Record Linkage Techniques*. Springer. http://www.springer.com/computer/database+management+%26+information+retrieval/book/978-0-387-69502-0.

Heymans, M, and I Eekhout. 2019. *Applied Missing Data Analysis with SPSS and (r)studio*. Heymans; Eekhout.

Hochreiter, Sepp, and Jürgen Schmidhuber. 1997. "Long Short-Term Memory." *Neural Computation* 9 (8): 1735–80. https://doi.org/10.1162/neco.1997.9.8.1735.

Hoover Green, Amelia, and Patrick Ball. 2019. "Civilian Killings and Disappearances During Civil War in El Salvador (1980–1992)." *Demographic Research* 41: 781–814. https://www.jstor.org/stable/26850667.

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning*. Vol. 112. Springer.

Johndrow, James E, Kristian Lum, and Daniel Manrique-Vallier. 2019. "Low-risk population size estimates in the presence of capture heterogeneity." *Biometrika* 106 (1): 197–210. https://doi.org/10.1093/biomet/asy065.

Johndrow, James, Kristian Lum, and David Dunson. 2018. "Theoretical Limits of Microclustering for Record Linkage." *Biometrika* 105: 431–46.

Kropko, Jonathan, Ben Goodrich, Andrew Gelman, and Jennifer Hill. 2014. "Multiple Imputation for Continuous and Categorical Data: Comparing Joint Multivariate Normal and Conditional Approaches." *Political Analysis* 22 (4): 497–519. https://doi.org/10.1093/pan/mpu007.

Link, William A. 2004. "Individual Heterogeneity and Identifiability in Capture Recapture Models." *Animal Biodiversity and Conservation* 27 (1): 87–91.

Lum, Kristian, Megan Price, and David Banks. 2013. "Applications of Multiple Systems Estimation in Human Rights Research." *The American Statistician* 67 (4): 191–200.

Madigan, David, and Jeremy C York. 1997. "Bayesian Methods for Estimation of the Size of a Closed Population." *Biometrika* 84 (1): 19–31. http://www.jstor.org/stable/2337552.

Manrique-Vallier, Daniel. 2016. "Bayesian Population Size Estimation Using Dirichlet Process Mixtures." *Biometrics* 72 (4): 1246–54.

Manrique-Vallier, Daniel, Ball P., and M. Sadinle. 2022. "Capture-Recapture for Casualty Estimation and Beyond: Recent Advances and Research Directions."

Michelson, Matthew, and Craig A Knoblock. 2006. "Learning Blocking Schemes for Record Linkage." In *Proceedings of the 21st National Conference on Artificial Intelligence (AAAI)*.

Ming Li, Linyuan Lü, Run-Ran Liu, and Yi-Cheng Zhang. 2021. "Percolation on Complex Networks: Theory and Application." *Physics Reports* 907: 1–68.

Müllner, Daniel. 2013. "Fastcluster: Fast Hierarchical, Agglomerative Clustering Routines for r and Python." *Journal of Statistical Software* 53 (9): 1–18.

Price, Gohdes, M., and P. Ball. 2015. "Documents of War: Understanding the Syrian Conflict." *Royal Statistical Society* 12: 14–19.

Rivest, Louis-Paul. 2011. "A Lower Bound Model for Multiple Record Systems Estimation withHeterogeneous Catchability." *The International Journal of Biostatistics* 7 (1). https://doi.org/10.2202/1557-4679.1283.

Rozo Ángel, Valentina, and Patrick Ball. 2019. "Asesinatos de Lideres Sociales En Colombia: Una Estimación Del Universo-Actualización 2018." Human Rights Data Analysis Group.

Rubin, Donald B. 1978. "Multiple Imputations in Samples Surveys - a Phenomenological Bayesian Approach to Nonresponse." *American Statistical Association* Proceedings of the Survey Research Methods Section: 20–34.

———. 1987. *Multiple Imputation for Nonresponse in Surveys*. Wiley.

———. 1996. "Multiple Imputation After 18 Plus Years." *Journal of the American Statistical Association* 91 (434): 473–89.

Särndal, Swensson, K. E. 2003. *Model Assisted Survey Sampling*. Springr.

Scheuren, Fritz. 2004. "History Corner." *The American Statistician*. https://www.tandfonline.com/doi/abs/10.1198/0003130042926.

Sekar, C Chandra, and W Edwards Deming. 1949. "On a Method of Estimating Birth and Death Rates and the Extent of Registration." *Journal of the American Statistical Association* 44 (245). https://amstat.tandfonline.com/doi/abs/10.1080/01621459.1949.10483294.

Shah, T. 2019".rproj Considered Harmfu." https://hrdag.org/tech-notes/harmful.html.

Suárez, Andrés Felipe. 2018. "Dos Registros, Dos Versiones. Los Usos Políticos de Los Registros Oficiales En Las Luchas Por La Verdad de La Guerra En Colombia." *Nuevo Mundo Mundos Nuevos*.

Van Buuren, S., H. C. Boshuizen, and D. L. Knook. 1999. "Multiple Imputation of Missing Blood Pressure Covariates in Survival Analysis." *Statistics in Medicine* 18 (6): 681–94. https://doi.org/https://doi.org/10.1002/(SICI)1097-0258(19990330)18:6%3C681::AID-SIM71%3E3.0.CO;2-R.

van Buuren, Stef. 2018. *Flexible Imputation of Missing Data*. Boca Raton: Chapman; Hall/CRC. https://stefvanbuuren.name/fimd.

van Buuren, Stef, and Karin Groothuis-Oudshoorn. 2011. "Mice: Multivariate Imputation by Chained Equations in r." *Journal of Statistical Software, Articles* 45 (3): 1–67. https://doi.org/10.18637/jss.v045.i03.

White, Ian R, Patrick Royston, and Angela M Wood. 2011. "Multiple Imputation Using Chained Equations: Issues and Guidance for Practice." *Statistics in Medicine* 30 (4): 377–99. https://doi.org/https://doi.org/10.1002/sim.4067.

Zhou, X, and J. Reiter. 2010. "A Note on Bayesian Inference After Multiple Imputation." *The American Statistician* 64 (2).