Analysis and Visualization of Big Data (2022), **1**, 1–8

OPEN DATA JOURNAL

MINI ARTICLE

Hate Crimes in Catalonia

Adrián Chacón Candia,[*†] Martina Moreno Fina,[‡] Robert Benassai Dalmau,[‡] and Nil Masó Castro[‡]

†Departament de Física de la Matèria Condensada, Barcelona, Spain.
‡Universitat de Barcelona Institute of Complex Systems, Barcelona, Spain.
[*]Corresponding author. Email: achacoca9@alumnes.ub.edu

Abstract

Hate crimes are legal infractions that target a specific victim, place or object due to it's belonging (or it's supposed one) to a certain race, nationality, language, religion or physical and mental limitations and other similar factors. This article studies the distribution of hate crimes all over Catalonia and how different populations (regarding gender or age) suffer from them. The study has focused on the years 2020 and 2021.

Keywords: hate crimes; police regions; inhabitants; gender and age studies

1. Background

1.1 Data

The first we need to know about the data that has been studied is it's origin. This study has mainly been done with the following data sets:

- *Portal de dades obertes de Catalunya*[5]: this data-set provides the information to conduct the study since it categorizes in 16 different ways each of the almost 2.000 crimes that were documented. Regarding that information our study will only focus deeply on the categories that provided a better and more conclusive study.

- *Institut Cartogràfic i Geològic de Catalunya*[3]: This data-set provides the necessary information to do the corresponding mapping of Catalonia to later do heat maps of the territory. It is also necessary other subdivisions found in *AMB*[1] and also *IDESCAT*[4].

- Population of each *Regió Policial*, this data was extracted from a newspaper letter listed in the bibliography [2].

All this data has been collected by the police body of Catalonia (*Mossos d'Esquadra*) and has undergone the FAIR principles to make the data accessible to everybody and not disclose personal information regarding the people in the study. One example of that could be that the information of hate crimes in populations under 20.000 inhabitants has been grouped in larger regions to difficult the location of the people in the study.

1.2 Background of analysis

This study strives to provide as much information regarding hate crimes as it can cover. First, we want to categorize the crimes and see if patters emerge from those categorizations. Once we have done that a more thorough study on specific categories is conducted to try to figure out what patterns or correlations emerge from comparing them. All of this is done with the goal to reduce hate crimes and provide information to find more effective ways to battle it.

2. Methods

The transformation of the data set into the results can be divided in different sections as documented in the "Chaky_Police.ipynb" code and later in this section. The code is done with Python (Jupyter Notebook) and the necessary libraries are Pandas, Sodapy, Matplotlip.pyplot, Numpy and Geopandas. It can be found on GitHub: Chaky81 - learn_git_a_bit.

The evolution of the project was the following:

2.1 First contact with the data-frame

First of all we need to import the data about hate crimes and turn it into a data-frame.

```
    #!/usr/bin/env python

# make sure to install these packages before running:
# pip install pandas
# pip install sodapy

import pandas as pd
from sodapy import Socrata

# Unauthenticated client only works with public data sets. Note 'None'
# in place of application token, and no username or password:
client = Socrata("analisi.transparenciacatalunya.cat", None)

# Example authenticated client (needed for non-public datasets):
# client = Socrata(analisi.transparenciacatalunya.cat,
#                 MyAppToken,
#                 username="user@example.com",
#                 password="AFakePassword")

# First 2000 results, returned as JSON from API / converted to Python
    list of
# dictionaries by sodapy.
results = client.get("gci6-2ubm", limit=2000)

# Convert to pandas DataFrame
results_df = pd.DataFrame.from_records(results)
```

2.2 Knowledge

After completing the first step we then try to familiarize ourselves with the data-frame and it's categories correcting all the typos that appeared. Moreover, we also study the properties and actions that panda enables us to do with "*results_df*".

2.3 First data management

In this section we started to subdivide the data-frame "*results_df*" to better understand the data and also make it more manageable. In this section we conclude that our study will be mainly focused on the different police regions that Catalonia has and also the age and gender of the people in the collected data. This conclusion was derived from the plots made from all the different categories since it was made visible which categories where more or less workable.

An example of unworkable data was the location of the crime since the response was very diverse and each category had a low frequency in most of the cases as it can be seen in the plot "tipus_de_lloc_dels_fets_total.png" from the previously mentioned.

2.4 Bar Plots and other calculations

Now that we knew we wanted to talk about age, gender and police regions we worked on each topic individually.

2.4.1 Age

Regarding age, we wanted to see how it related with being the aggressor or the victim of a hate crime and also compare the number of hate crimes for a given age with the total population of that age group.

2.4.2 Gender

For the gender studies we wanted to know how the gender related to being an aggressor or a victim of a hate crime.

2.4.3 Police Regions

Regarding police regions we looked into the number of hate crimes in each region and also it's density normalizing it with it's total population.

2.5 Mapping

Having learnt Geopandas during the course, and being able to plot a map a main objective of the course, a more detailed study was made for the 2.4.3 section. Instead of histograms we thought it would be much more visual and understandable to plot a Catalonia heat map regarding the number of hate crimes per police region. To do so, we needed to create a Catalonia map subdivided in police regions since there wasn't one available.

2.5.1 Map Making

Since there is not a map generated that is limited by RPs, it is necessary to create one through a map divided in "comarques". First of all, we create a dictionary with all RPs and its "comarques". Also, we create a new inverted dictionary, which has all the RPs for each "comarca" since it will be useful in the next steps. Due to the fact that the "comarca" Barcelonès has different "municipis" that correspond to different police regions, it is necessary to add them to the dictionary created. Then, we create a new map delimited by the RPs through joining all the areas of the "comarques" of the same RP. The result is a DataFrame that has one column for RPs and one for the geometry, which consists on the features of the multipolygons and polygons. Finally, we add to this DataFrame the number of hate crimes as a new column. Furthermore, the heat map will be provided with the name of each police region and a title on the heat scale to make it more comprehensible.

2.5.2 Map Plots

For the map plots our goal is to show the number and the density of hate crimes committed in each police region. For the first map we do as the previous section mentions but for the latter we need to divide the number of hate crimes per the population of the given region.

3. Results

3.1 Age

As stated before, regarding age, we wanted to see the ratio of victims per total aggressions (victims + aggressors) for a given age and also the ratio of hate crimes for each age group compared to the total population.

The results were the following:

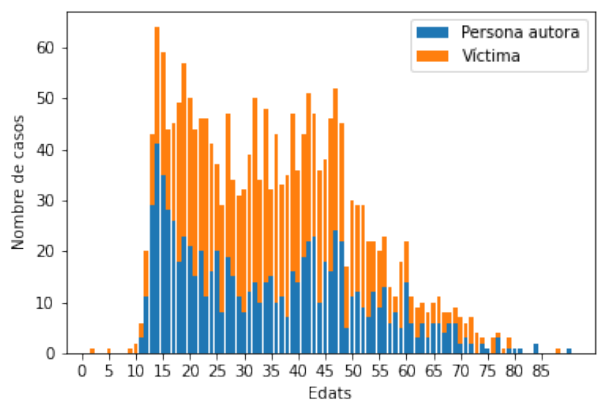

Figure 1. Histogram that represents the number of agressors (*Persona autora*) and victims (*víctima*) for a given age.

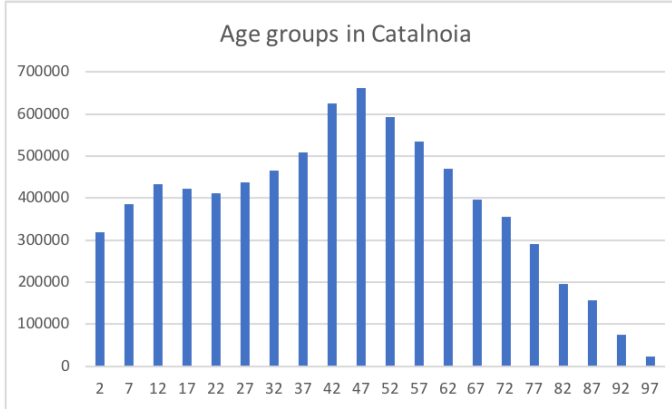

Figure 2. Population of different age groups in Catalonia, the values of the histogram represent the mean age of a group that goes from [x-2,x+2] where x is the value that appears in the plot.

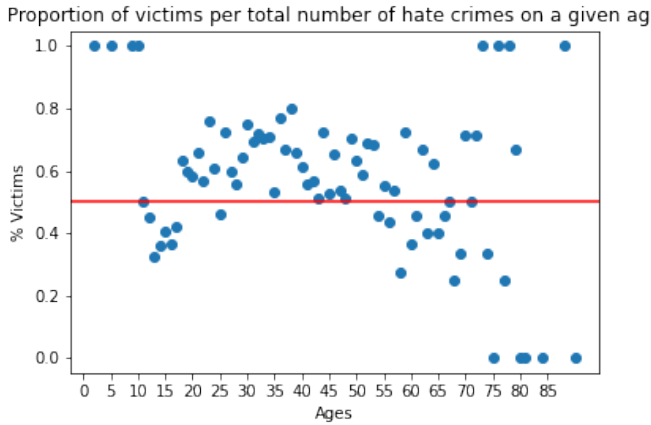

Figure 3. Proportion of victims per total agressions on a given age with the 0.5 value ploted as a read line.

The conclusion extracted from the figures in this section were that, if you compare Fig.1 with Fig.2 you can see a clear spike in the teenager ages in 1 that is not corresponded in 2. Thus, this leads to conclude that the young population (teenagers) commits a much higher ratio of hate crimes than the other age groups.
Furthermore, regarding the victim ratio for the different ages (Fig.3) we can only say that the results were none conclusive since there's too much dispersion with the data and it is located near the expected value, 0.5, which means that the number of victims is the same as the number of aggressors.

3.2 Gender

For gender we computed the same ratio as the section before.

Table 1. Gender studies regarding hate crimes

Gender	Victim	Aggressor	Ratio
Male	730	636	0.53
Female	408	195	0.68

The results demonstrate that most of the hate crimes are committed and received by males. However, when women are involved, they are much likely to be enrolled as a victim and not an aggressor.

3.3 Population ratios

??

In this section we can see how the situation of hate crimes is in each region. In Fig.4 we can see that *RP Metropolitana de Barcelona* is the one that has most of the hate crimes. However, when normalizing to the population, it seems to be that *RP Central* stands out as one of the regions with most hate crimes.

This plots were most useful when comparing the work on this mini article with the other members of the more broad team regarding police knowledge in Catalonia. This study can also be found in a GitHub repository: data_analysis_project.

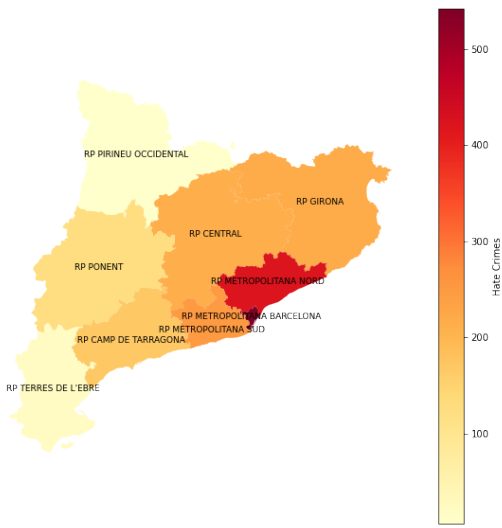

Figure 4. Heat map of Catalonia showing the number of hate crimes committed in each police region.

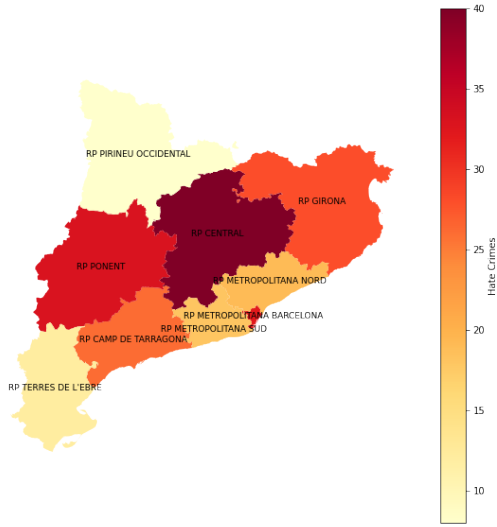

Figure 5. Heat map of Catalonia showing the density of hate crimes committed in each police region.

4. Conclusions / Discussion

First of all, this study has been possible thanks to the open data policy the *Mossos d'esquadra* had with the recollected data. Therefore, this part of the study and also my colleagues' part will also be public to promote data openness since we think it is a useful tool to promote science development.

Regarding the results, the main thing that can be said is that most of our study yields to inconclusive answers about the behaviour or distribution of hate crimes. However, some main bullet-points can be scraped out. We have discovered that age does play a significant role on the number of hate crimes since, in the teenager ages, the number of hate crimes amounted a significant increase compared to the population of the age group. We can also say that, regarding gender, most of the hate crimes have a male figure related to them although, when a women is related, it is usually as the figure of the victim. Finally, regarding the number of hate crimes, one can expect that *RP Metropolitana de Barcelona* is the one with the highest number but it is surprising that *RP Central* stands out when normalizing the population.

A future research would be necessary to better comprehend the behaviour of hate crimes in Catalonia. It is important to add new variables or categories that in this study were dismissed to better complement the information we already have. Moreover, this project would be greatly helped with a much thorough investigation of the other transversal studies about police knowledge[1] since correlations could be met with other police related issues.

All in all, in this study we have obtained some interesting results thanks to data from public research. The most relevant conclusions are that hate crimes are related to age, since we see a spike on those type of crimes in the teenager ages, related to gender, since men are the ones mostly involved but when women appear they usually do as the role of victim and also that *RP Central* has an issue with this type of crime when we consider the ratio for the population in each region.

References

[1] AMB. Àrea metropolitana - municipis metropolitans, 2022. https://www.amb.cat/es/web/area-metropolitana/municipis-metropolitans accessed 25 Nov 2022.

[2] Departament interior de Generalitat de Catalunya, N. D. Quantitat de denúncies a catalunya durant el primer estat d'alarma, per article vulnerat i estat de tramitació, 2021. https://www.naciodigital.cat/manresa/noticia/98203/denuncies-primer-estat-alarma-catalunya-pendents-tc.

[3] ICGC. Divisions administratives, 2022. https://www.icgc.cat/Descarregues/Cartografia-vectorial/Divisions-administratives accessed 25 Nov 2022.

[4] idescat. Superfície. per zones altimètriques. comarques i aran i províncies, 2020. https://www.idescat.cat/indicadors/?id=aec&n=15182 accessed 25 Nov 2022.

[5] interior de Generalitat de Catalunya, D. Víctimes o persones autores de delictes d'odi i discriminació, 2022. https://analisi.transparenciacatalunya.cat/Seguretat/V-ctimes-o-persones-autores-de-delictes-d-odi-i-di/gci6-2ubm.

[1] This studies can be found in the link stated before in the section ??.