

## MINI ARTICLE

# Studying the role of age, gender and location in disappearances reported between 2020 and 2021

Robert Benassai Dalmau,<sup>\*†</sup> Martina Moreno Fina,<sup>‡</sup> Nil Masó Castro,<sup>‡</sup> and Adrián Chacón Candia<sup>‡</sup>

<sup>†</sup>Departament de Física de la Matèria Condensada, Barcelona, Spain.

<sup>‡</sup>Universitat de Barcelona Institute of Complex Systems, Barcelona, Spain.

\*Corresponding author. Email: [rbenasda7@alumnes.ub.edu](mailto:rbenasda7@alumnes.ub.edu)

### Abstract

Disappearance cases are a common topic on news coverages around the world. Only in Catalonia, thousands of people are reported missing each year. In this article, a data set of disappearance reports from 2020 to 2021 has been studied through the generation of geographical heat maps and disappearances vs. age analysis. Results show that missing reports are filed most often for youngsters around 14 to 16 years old, and that, in the adult years, women disappear more than twice as much as men do. This last result could be explained through the rise in awareness and cases of gender violence in the last couple of decades, although a more in depth analysis with more data should be carried out to fully support this idea. The geographical analysis concluded that most cases were reported in the metropolitan areas of Barcelona, while the regions of Girona and Ponent had the most amount of disappearances per number of inhabitants. This statistical analysis could become important to highlight the most susceptible and vulnerable collectives and develop new preventive mechanisms and strategies to further protect these individuals.

**Keywords:** disappearances; police; gender violence

## 1. Background

Each year, thousands of disappearance reports are filed just in Catalonia alone. The magnitude of this number is worrying, and by analysing some aspects of these reports, an answer can be found to questions such as how do age, gender or even location affect the frequency of these events.

The data used in this article has been retrieved from Generalitat's online open data portal, *Dades Obertes Catalunya* [6]. The data set has been provided by Direcció General de la Policia (Departament d'Interior), who have available all the details about each disappearance case. The data set contains information related to the disappeared

individuals, specifically the dates (all between 2020 and 2021), ages, genders and police regions [5] of each report.

## 2. Methods

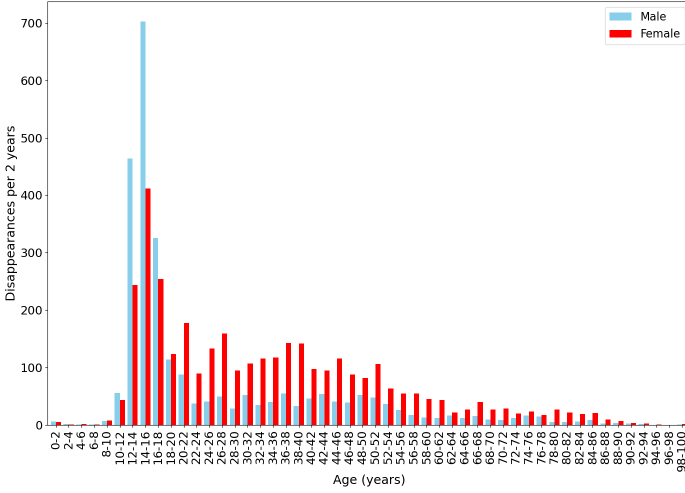
To treat the data, the *Pandas* [4] library from Python has been used. Data has been stored in *Data Frames*, table-like objects with the same row-column structure as the original data set.

In order to make a bar plot of the disappearances in age intervals of two years (figure 1), a complementary *Data Frame* has been created. This one contained three columns: *age*, *female disappearances* and *male disappearances*. This way, gender and age could be compared simultaneously in the same plot.

With the objective of making a geographical heat map of the disappearances in each police region (RP), the *Geopandas* [2] library from Python has been used. This library uses *Data Frames* which contain an extra column with the contours of the geographical areas. The data set considered contained information about of the RP where the disappearance occurred. As a map containing the polygons (spatial contours) of each RP could not be found online, other maps containing "comarques" and municipalities [3] have been used to generate a new map which included the RPs. This has been achieved by merging the appropriate "comarques" and municipalities, according to official sources [5], into the bigger RPs using the *dissolve* method in *Geopandas*. This method essentially joins the Polygons corresponding to a common column value. In this case the program joined all the municipalities and "comarques" that were in the same RP. In order to do this, a column with the corresponding RP of each municipality had to be added to the original *Geopandas Data Frame*. The detailed procedure can be found in the project's GitHub repository ([https://github.com/robedamo/data\\_analysis\\_project/tree/main/Robert](https://github.com/robedamo/data_analysis_project/tree/main/Robert)).

## 3. Results

In figure 1 a bar plot with the disappearance reports in age intervals of two years is represented. For each age interval men and women are shown separately such that a comparison between male and female disappearances can be made. Notably, the maximum amount of reports occurs in the age range from 14 to 16 years old for both genders, although more missing males are reported at that age. From the legal Spanish adult age (18 years old) onwards, male disappearances reduce drastically, while female ones, although also decrease from the maximum, are more recurrent during the adult lifespan than men's. To put this fact into numbers, between 18 up to 60 years old, a total of 2273 women were reported to be missing between 2020 and 2021 against 1026 disappeared men.



**Figure 1.** Bar plot of male and female disappearances in age intervals of two years.

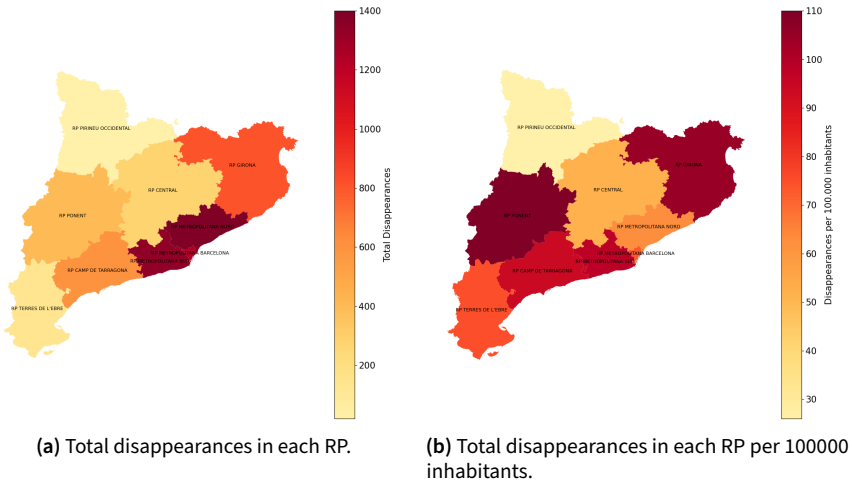
Overall, 57% of the reported disappearances corresponded to men and 43% to women. It can be tested statistically if men missing reports are equally probable as women ones. Making the statistical hypothesis that the probability of a man disappearing is the same as for a woman  $p_m = p_w = 0.5$ , and considering that the distribution follows a binomial process, the mean ( $\mu = p_m$ ) is normally distributed. Here, the value of a man disappearance report is 1 and the value for a woman disappearance report is 0. The standard deviation of the sample is  $\sigma = \sqrt{\frac{p_m p_w}{n}}$  with  $n = 6219$  reports. Then, the  $z$  value obtained is:

$$z = \frac{0.57 - \mu}{\sigma} = 11.04 \quad (1)$$

Which gives a p-value of nearly 1. With this result, the hypothesis that  $p_m = p_w = 0.5$  can be rejected with nearly 100% confidence. This result may seem categorical, but many factors have been ignored in this analysis, such as age or RP location, which may strongly influence these probabilities.

If the focus is now put in where the cases were reported, a heat map can be built showing the total amount of disappearance reports in each RP (figure 2 (a)). Arguably, a more representative way of visualising the data is to show the total disappearances relative to the population of each RP (measured in disappearances per 100000 inhabitants) (figure 2 (b)).

Looking at figure 2 (a), the most densely crowded RPs (*RP Metropolitana Nord*, *RP Metropolitana Barcelona* and *RP Metropolitana Sud*) are also the ones with most missing reports overall (see a map of the RPs in [5]). Interestingly, taking into account the population of each RP, *RP Ponent* and *RP Girona* are highlighted as the top two with most disappearances per 100000 inhabitants.



**Figure 2.** Heat maps of disappearances in each RP.

#### 4. Discussion

This data set only contains certain information about the disappearance reports, and the background and causes of each specific case is understandably not provided. This makes the withdrawal of conclusions tough. Nevertheless, some outstanding behaviors shown in the previous results may provide valid arguments in order to discuss certain hypotheses about common patterns.

As seen previously, the age in which most disappearances are reported lies between 14 and 16 years old for both genders. These are difficult years for teenagers, and conflicts at home become more common. Some youngsters decide to run away from home, as their situation and relationship with their parents becomes unsustainable. Other factors such as drugs and alcohol can also influence these decisions, as, at these ages, boys and girls become exposed to these substances and may develop addictions to them [1]. One could infer the contribution of this kind of disappearances if the data set contained the state (active or solved) of the reports and the amount of time these teenagers had been missing, as these "runaway" cases may be resolved in days. It would be interesting to see what percentage of the disappearances are of this kind. Despite this, it is important to remember that these experiences have been traumatic for the family and social surroundings of the victims and, although numbers tend to dehumanize these cases, there were real people involved in them and these matters shall not be taken lightly.

Focusing on the gender perspective of figure 1, it is difficult to find a reason as to why men disappear more often than women at the teenage years, and an answer might lie in the actual reasons of each case.

Regarding the adult age category, data shows that more women are reported to be missing than men. As the possibility of running away from home vanishes, these cases could potentially imply longer term disappearances, related to murders,

kidnappings or other serious crimes. It is known that gender violence is present in our societies, and misogynist crimes are reported each week, be it in the form of rape, domestic violence or even femicides. This misogyny could influence adult women disappearance numbers, and data showed that from 18 to 60 years old, the number of disappeared women doubled men's disappearance numbers between 2020 and 2021. The role of gender violence in these numbers could be known precisely analysing each case individually, but, as mentioned before, this information is understandably not public.

This age bias influences the probability that, given a disappearance report, it corresponds to a man or a woman. The results obtained in the statistical analysis (1) state that these probabilities are not the same, but it ignores the fact that, given an age range, one gender can be more vulnerable than the other.

Lastly, according to the maps in figure 2, it is observed that the most populated areas are the ones with more disappearances as expected. When the number of inhabitants in each RP is taken into account, the RPs from Girona and Ponent are the highlighted ones. It would be useful to have a more detailed geographical description, with smaller divisions such as "comarques" in order to see the incidence of disappearances in each of them.

## References

- [1] Diari de girona, *El 80% dels adolescents que desapareixen ho fan per motius d'addicció*, 2022. <https://www.diaridegirona.cat/successos/2022/03/09/80-dels-adolescents-desapareixen-ho-63626635.html> accessed 24 Nov 2022.
- [2] Geopandas' website, 2022. <https://geopandas.org/en/stable/index.html> accessed 24 Nov 2022.
- [3] Institut cartogràfic i geològic de catalunya, 2022. <https://www.icgc.cat/ca/Descarregues/Cartografia-vectorial/Divisions-administratives> accessed 24 Nov 2022.
- [4] Pandas' website, 2022. <https://pandas.pydata.org/> accessed 24 Nov 2022.
- [5] GENCAT. Regions policials, 2022. [https://interior.gencat.cat/ca/el\\_departament/adreces-i-telefonos/regions\\_policials/](https://interior.gencat.cat/ca/el_departament/adreces-i-telefonos/regions_policials/) accessed 24 Nov 2022.
- [6] GENERALITAT. Dades obertes catalunya, 2022. <https://analisi.transparenciacatalunya.cat/Seguretat/Denuncies-de-persones-desaparegudes/bp4b-qsst> accessed 24 Nov 2022.