

Case study: immigration and accidents in Barcelona

Celine Heng, Donata Buozyte, Joseph Kong, Max Schäpers

03/01/2022

Motivation

In 2017 the number of new car registrations in Barcelona was about 190 thousand¹. Considering this, it is not surprising that the number of victims in accidents was at about 12 thousand with a total of 241 serious injuries. Especially in the central district Eixample the number of victims is high compared to the other districts.

On the other hand, Eixample seems to be a popular district among immigrants from all over the world, as more than 50 thousand people decided to settle there.

Considering these two observations, one might conclude that the immigrants could be a reason why the amount of accidents and injuries is significantly higher than in other districts.

Data Preparation

In the data analysis the two data tables regarding the accidents in 2017 and regarding the immigrants by nationality from the Barcelona data table are used. An additional data table with the population density per district is defined.²

The accidents data table is melted to create columns with the type of injury and the number of injuries of that type per accident. Further new columns include the sum of injuries (per type and in total) per district, the sum of injuries in Eixample and non-Eixample districts and their ratio in percent.

The immigrants per nationality data table is prepared in a similar way and includes the same additional columns but with the sums of the number of immigrants.

Lastly, all previously prepared data tables are merged to enable the analysis of the relationships between the number of injuries and the number of immigrants.

Data Analysis

The following visualizations in Figure 1 show the mentioned spike of vehicle accidents and also the injuries due these accidents in the district Eixample. Compared to all the other districts, about 30% of all injuries happened in Eixample and especially the amount of mild injuries is significantly higher than in other districts.

¹<https://www.catalannews.com/business/item/number-of-new-cars-registered-up-7-in-2017>

²https://en.wikipedia.org/wiki/Districts_of_Barcelona

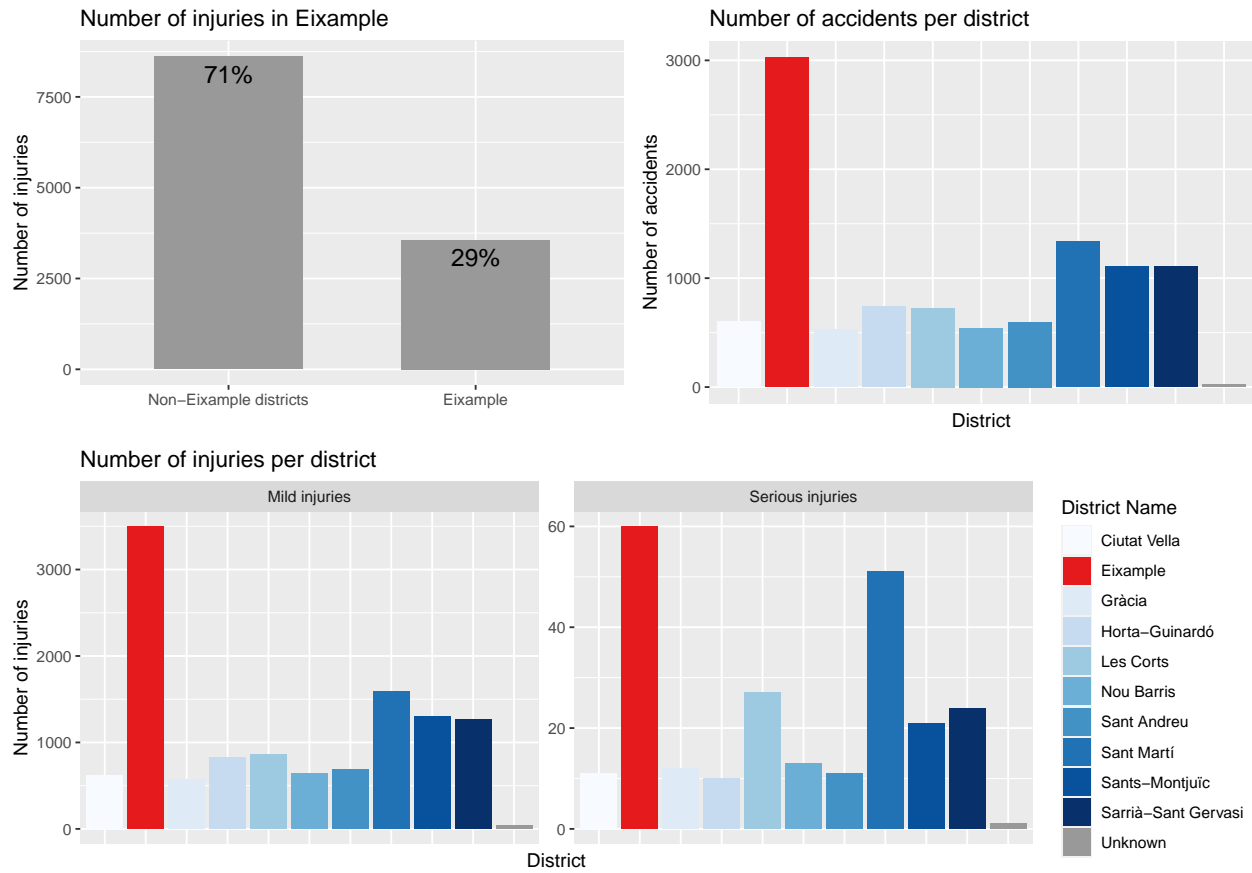


Figure 1: Histograms displaying the ratio between the number of accidents and injuries due to accidents in Eixample and in all other districts either together or separated.

The histogram in Figure 2 displays the observation, that comparatively more immigrants decide to settle in Eixample than in any other district. The distribution of immigrants to each district is also similar to the distribution which can be also observed for the number of accidents and injuries per district in Figure 1.

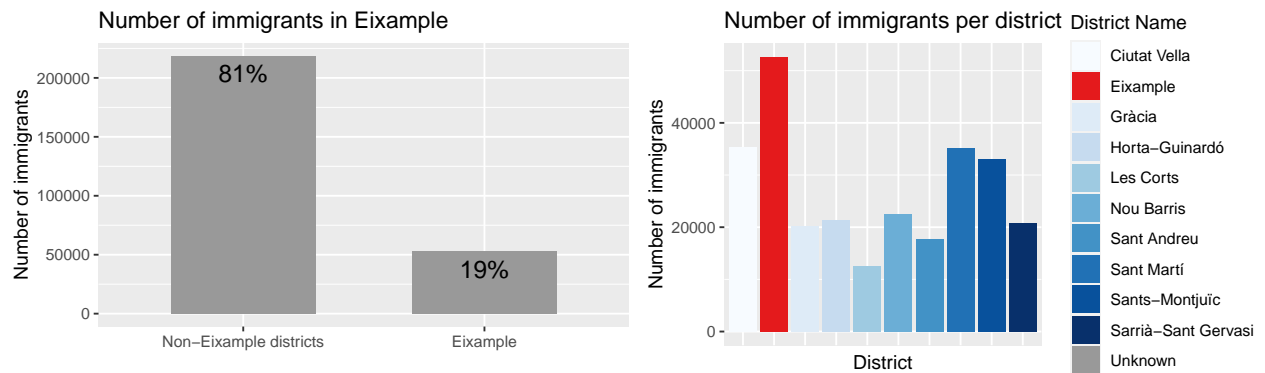


Figure 2: Histograms displaying the ratio between of immigrants to Eixample and to all other districts either together or separated.

Comparing the immigration and number of injuries per district, such as in Figure 3, one could conclude that

they are somehow related. This assumption can be further inspected using the Spearman rank correlation test, as the actual distribution of the data is likely not Gaussian and unknown and as the test is robust to outliers, like for example Eixample. In this case, the correlation between the immigrants and injuries with the null hypothesis that they are not correlated is tested. As the p-value is sufficiently small, this hypothesis is rejected, which implies that they are correlated and supports therefore the aforementioned assumption.³

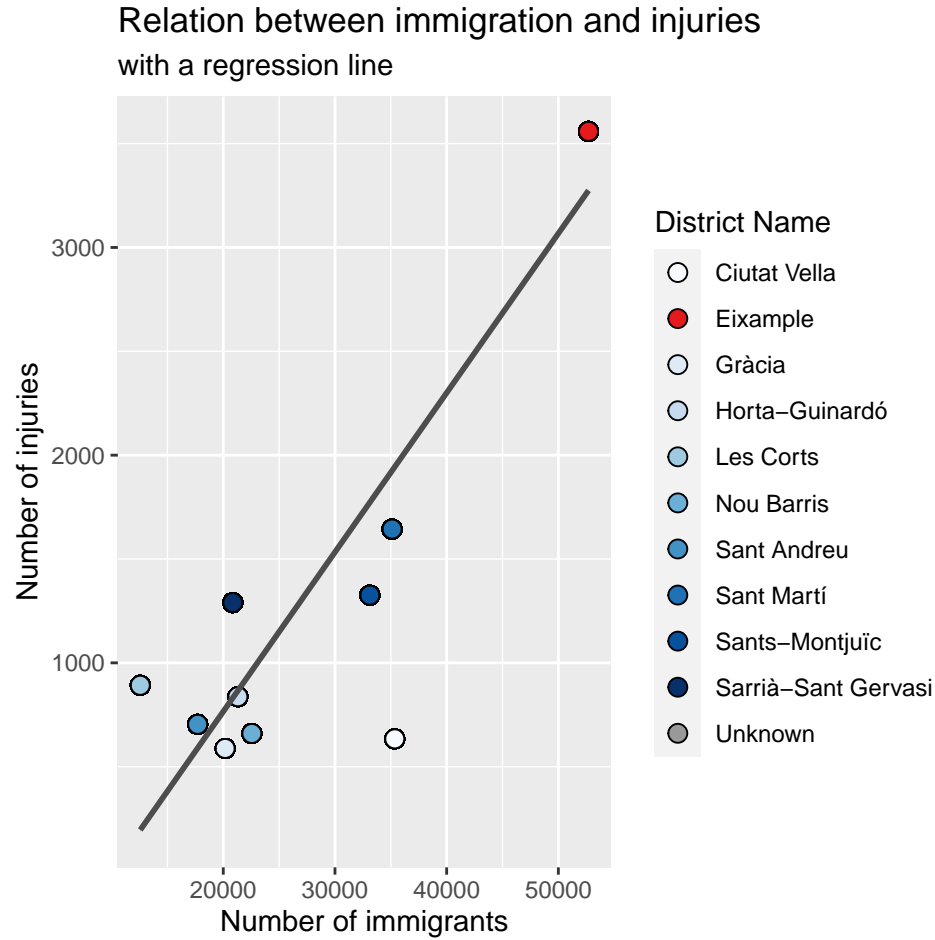


Figure 3: Scatter plot displaying the relation between the number of immigrants and the number of injuries with an regression line showing the positive correlation between the axes.

```
##
## Spearman's rank correlation rho
##
## data: immigrants and total_injuries_per_dist
## S = 3.262e+11, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.776888
```

Although the Spearman test is robust to outliers, it is useful to perform it without the district Eixample, as it could be interpreted as a significant outlier in the original plot in Figure 3.

³Remark: it should be noted that the number of available data points is quite low, which heavily influences the plot, the regression line, the Spearman test and therefore the conclusions.

In this second test the value ρ is about half of the original value, which indicates that the possible outlier has a huge influence on the Spearman test. Nevertheless, the p-value is significantly low and hence the null hypothesis is still rejected.

```
##
## Spearman's rank correlation rho
##
## data: immigrants and total_injuries_per_dist
## S = 3.2147e+11, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.3758764
```

But as Eixample is the central district of Barcelona, it is obvious that it also has the highest population among all districts with about 266 thousand residents. The population density in thousand residents per km² in this district is 35.7, and therefore the highest among all districts in Barcelona, as it's size is only 7.46 km². As a comparison, the population of Sants-Montjuïc is about 182 thousand, while it's size is 21.35 km², hence its density is only about 8.5.⁴

The following visualization in Figure 4 shows that for similar population density the number of injuries stays nearly the same with the only slight exception being Sant Martí. This could be due to its three beaches, which are popular among locals and tourists, or the high amount of factories, which both usually cause more traffic in the according areas.⁵

⁴https://en.wikipedia.org/wiki/Districts_of_Barcelona

⁵<https://www.barcelona-journal.de/reisetipps/die-stadt/sant-marti.html>, [https://de.wikipedia.org/wiki/Sant_Mart%C3%AD_\(Barcelona\)](https://de.wikipedia.org/wiki/Sant_Mart%C3%AD_(Barcelona))

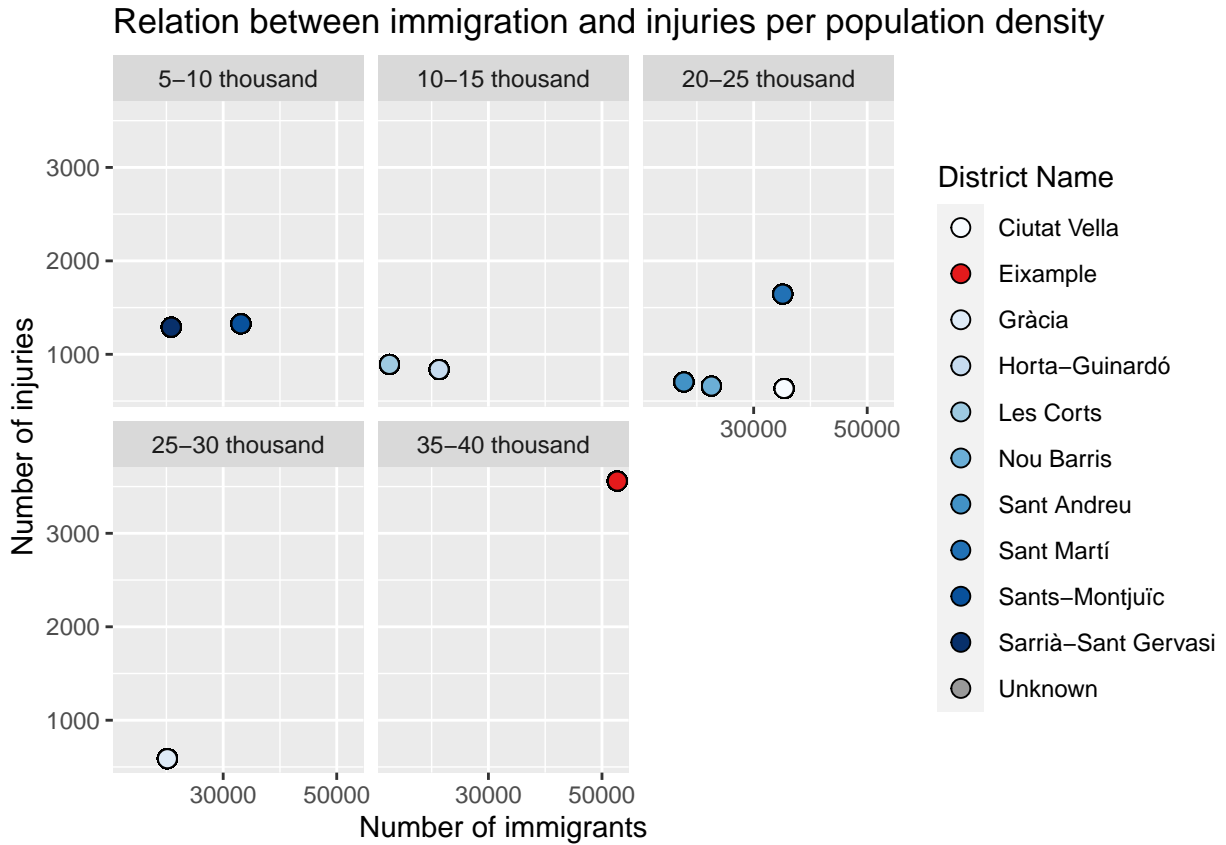


Figure 4: Scatter plot displaying the relation between the number of immigrants and the number of injuries faceted by the population density of the districts.

Conclusion

Based on the analysis we can conclude that there is a certain relationship between the amount of injuries due to accidents and the immigration in a district, but this relationship is due to the population density per district. Hence although they are related, more immigration does not cause a higher amount of injuries in accidents. But Figure 4 also shows, that the population density does not directly cause the high number of injuries, as districts with a really low population density can have significantly more injuries than others. The actual relationship between those two factors will not be further analysed here, but could for example be related to the position of the district or the housing options, attractions, parks, etc. in the district.