

# Final Report on Barcelona

*Karan Dassi*

*5/28/2019*

## Introduction

This study plans to give a 360-degree and in depth exploration of the city of Barcelona, by dissecting information along the following interrelated 4 dimensions: Population, transportation, accidentality, migration and employment.

To explore and gain the most insight from our data, we have based our exploration in the results of a few studies that help us have better understanding about the city of Barcelona. The City Report: Barcelona from Wilco explains the transaction of the city's labor market, demographic change and immigration trend. The articles from "Determinants of urban sprawl in European cities" and "Measuring Negative Synergies of Urban Sprawl and Economic Crisis over Public Transport Efficiency: The Case of Spain" includes the city's urban planning strategy and reviews.

The first section will look at the raw accident data, the second will study immigration and the final part will cover unemployment. Each section will contain the conclusions pertaining to it's subject matter. As an appendix, we are submitting a geographic representation of the most relevant findings, which we believe helps put a very important perspective in the data.

## Few Dataset Description

```
## Classes 'spec_tbl_df', 'tbl_df', 'tbl' and 'data.frame': 10312 obs. of 13 variables:
## $ DISTRICT.NAME : chr "Ciutat Vella" "Ciutat Vella" "Ciutat Vella" "Ciutat Vella" ...
## $ NEIGHBORHOOD.NAME: chr "el Raval" "el Raval" "el Raval" "el Raval" ...
## $ WEEKDAY : chr "Thursday" "Thursday" "Friday" "Tuesday" ...
## $ MONTH : chr "May" "November" "June" "July" ...
## $ DAY : num 18 16 23 4 20 10 27 29 3 28 ...
## $ HOUR : num 0 20 1 19 16 0 15 21 13 7 ...
## $ PART.OF.THE.DAY : chr "Night" "Afternoon" "Night" "Afternoon" ...
## $ MILD.INJURIES : num 3 1 1 1 0 1 1 0 0 3 ...
## $ SERIOUS.INJURIES : num 0 0 0 0 0 0 0 0 0 0 ...
## $ VICTIMS : num 3 1 1 1 0 1 1 0 0 3 ...
## $ VEHICLES.INVOLVED: num 2 1 1 1 2 2 1 1 1 3 ...
## $ LONGITUDE : num 2.18 2.17 2.17 2.18 2.17 ...
## $ LATITUDE : num 41.4 41.4 41.4 41.4 41.4 ...
## - attr(*, "spec")=
## .. cols(
## .. DISTRICT.NAME = col_character(),
## .. NEIGHBORHOOD.NAME = col_character(),
## .. WEEKDAY = col_character(),
## .. MONTH = col_character(),
## .. DAY = col_double(),
## .. HOUR = col_double(),
## .. PART.OF.THE.DAY = col_character(),
## .. MILD.INJURIES = col_double(),
## .. SERIOUS.INJURIES = col_double(),
## .. VICTIMS = col_double(),
## .. VEHICLES.INVOLVED = col_double(),
## .. LONGITUDE = col_double(),
```

```

## .. LATITUDE = col_double()
## .. )

## Classes 'spec_tbl_df', 'tbl_df', 'tbl' and 'data.frame': 14016 obs. of 6 variables:
## $ YEAR : num 2017 2017 2017 2017 ...
## $ DISTRICT.NAME : chr "Ciutat Vella" "Ciutat Vella" "Ciutat Vella" "Ciutat Vella" ...
## $ NEIGHBORHOOD.NAME: chr "el Raval" "el Barri Gòtic" "la Barceloneta" "Sant Pere, Santa Caterina i
## $ GENDER : chr "Male" "Male" "Male" "Male" ...
## $ AGE : chr "0-4" "0-4" "0-4" "0-4" ...
## $ NUMBER_P : num 224 50 43 95 124 191 176 158 220 143 ...
## - attr(*, "spec")=
## .. cols(
## .. YEAR = col_double(),
## .. DISTRICT.NAME = col_character(),
## .. NEIGHBORHOOD.NAME = col_character(),
## .. GENDER = col_character(),
## .. AGE = col_character(),
## .. NUMBER_P = col_double()
## .. )

## Classes 'spec_tbl_df', 'tbl_df', 'tbl' and 'data.frame': 14208 obs. of 9 variables:
## $ Year : num 2017 2017 2017 2017 2017 ...
## $ Month : chr "January" "January" "January" "January" ...
## $ District Code : num 1 1 1 1 2 2 2 2 2 2 ...
## $ District Name : chr "Ciutat Vella" "Ciutat Vella" "Ciutat Vella" "Ciutat Vella" ...
## $ Neighborhood Code: num 1 2 3 4 5 6 7 8 9 10 ...
## $ Neighborhood Name: chr "el Raval" "el Barri Gòtic" "la Barceloneta" "Sant Pere, Santa Caterina i
## $ Gender : chr "Male" "Male" "Male" "Male" ...
## $ Demand_occupation: chr "Registered unemployed" "Registered unemployed" "Registered unemployed" "
## $ Number : num 2107 538 537 741 630 ...
## - attr(*, "spec")=
## .. cols(
## .. Year = col_double(),
## .. Month = col_character(),
## .. `District Code` = col_double(),
## .. `District Name` = col_character(),
## .. `Neighborhood Code` = col_double(),
## .. `Neighborhood Name` = col_character(),
## .. Gender = col_character(),
## .. Demand_occupation = col_character(),
## .. Number = col_double()
## .. )

```

The accidents data set contains the accidents recorded by police office in the city of Barcelona in 2017. There are total 10,312 observations in the accidents data set with 13 variables - District Name, Neighborhood Name, Weekday, Month, Day, Hour, Part of the day, Mild Injuries, Serious Injuries, Victims, Vehicles Involved, Longitude and Latitude.

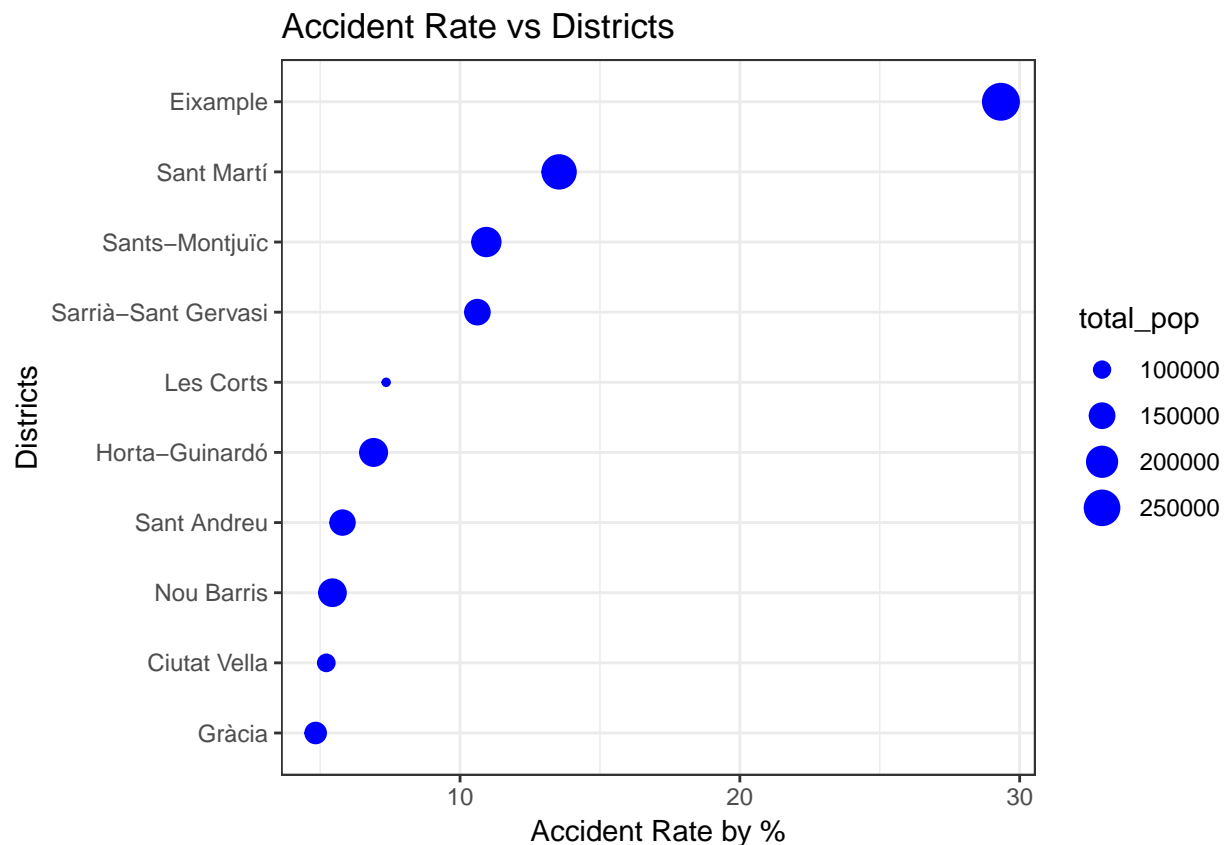
The population data set contains the total 2017 population of Barcelona. There are total 14,016 observations in the population data set with 6 variables - Year, District.Name, Neighborhood.Name, Gender, Age, and Number. The unemployment dataset contains the employment records for the people in Barcelona in the year 2017, there is a total of 14,208 observations for 9 variables.

## Barcelona accident analysis in 2017

### Hypothesis 1 in accidents:

Accidents would be spread evenly in the entire city as most areas in Barcelona are accessible for both car and walkers.

```
## # A tibble: 10 x 4
##   DISTRICT.NAME      total_vic vic_rate total_pop
##   <chr>            <dbl>   <dbl>   <dbl>
## 1 Eixample          3562    29.3   266416
## 2 Sant Martí        1645    13.5   235513
## 3 Sants-Montjuïc    1329    10.9   181910
## 4 Sarrià-Sant Gervasi 1290    10.6   149279
## 5 Les Corts         894     7.36   82033
## 6 Horta-Guinardó    839     6.91  168751
## 7 Sant Andreu       704     5.8   147594
## 8 Nou Barris        661     5.44  166579
## 9 Ciutat Vella      634     5.22  101387
## 10 Gràcia           588     4.84  121347
```



### Analysis and finding:

Our initial research shows that Barcelona is relatively small city with 39.34 mi<sup>2</sup> which is 0.58 times as big as Washington D.C; therefore, our assumption is that the accidents would be spread evenly in the entire city as most areas in Barcelona are accessible for both car and walkers. However, our result indicates that most accidents were actually clustering in the district Eixample and then Sant Marti, Sants Montjuic and

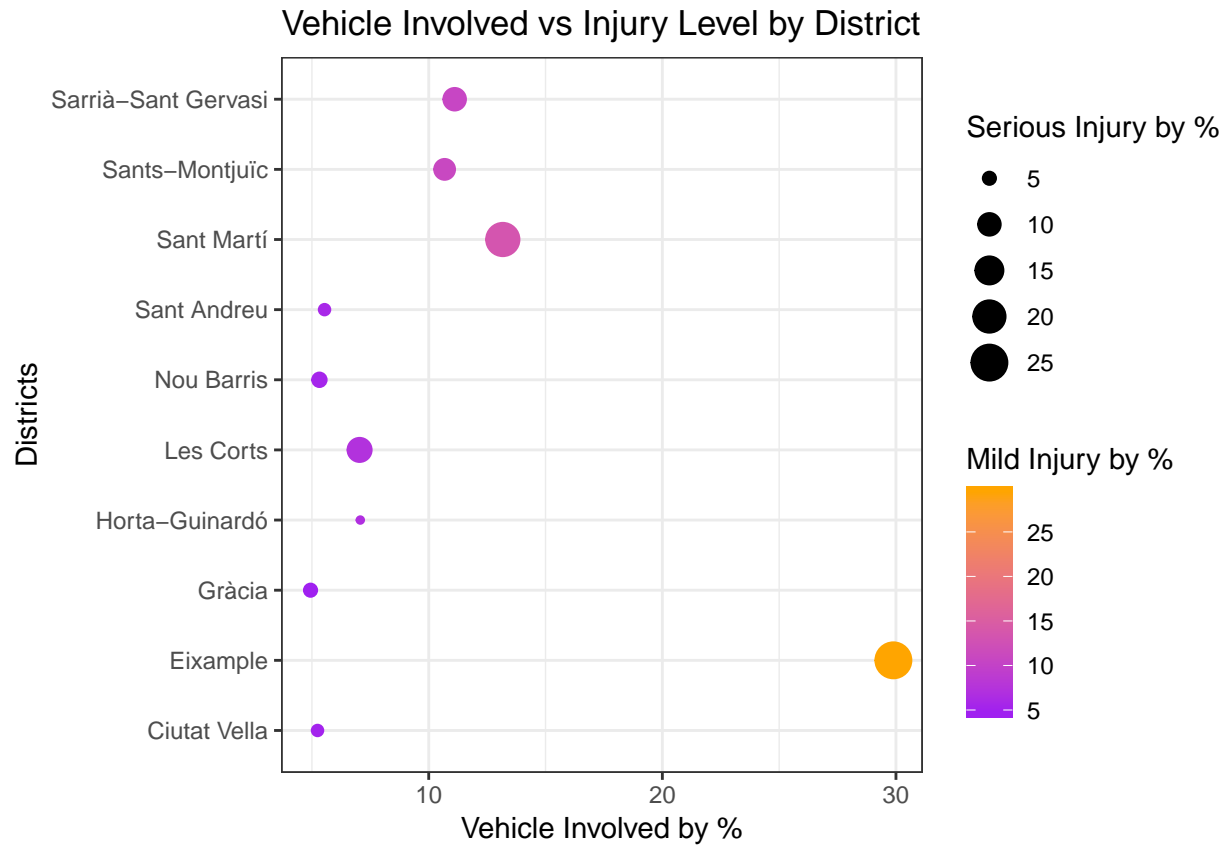
Sarria-Sant Gervasi with 29.33%, 13.54%, 10.94% and 10.62% accident rate. We have also incorporated population data set in our analysis and it shows that the higher the accident rate is, the higher the population is within the district.

As the result is not supporting our assumption, we dived into more and created a accident count heat map. Based on the heat map (in Appendix 1), we noticed that accidents were spread out from the middle the city which is Eixample to the north, south and west sides of the city where Sant Marti, Sants Montjuic and Sarria-Sant Gervasi locate in respective order. According to Google map attraction locator, most restaurants, bars and hotels in Barcelona are in Eixample and the attractions are spread from Eixample toward Sant Marti, Sants Montjuic and Sarria-Sant Gervasi. Therefore, it supports our first result that most accidents happened in these 4 districts, but our assumption is not correct. The accidents did not spread evenly in the city.

## Hypothesis 2 in accidents:

Serious injury is associated with accidents when vehicle(s) involved.

```
## # A tibble: 10 x 4
##   DISTRICT.NAME      mid_rate ser_rate veh_rate
##   <chr>             <dbl>   <dbl>   <dbl>
## 1 Eixample          29.4     25     29.9
## 2 Sant Martí        13.4     21.2    13.2
## 3 Sarrià-Sant Gervasi 10.6     10     11.1
## 4 Sants-Montjuïc    11.0      8.75    10.7
## 5 Horta-Guinardó     6.95     4.17     7.07
## 6 Les Corts         7.27     11.2     7.04
## 7 Sant Andreu        5.83     4.58     5.54
## 8 Nou Barris         5.44     5.42     5.32
## 9 Ciutat Vella       5.24     4.58     5.24
## 10 Gràcia            4.84      5      4.94
```

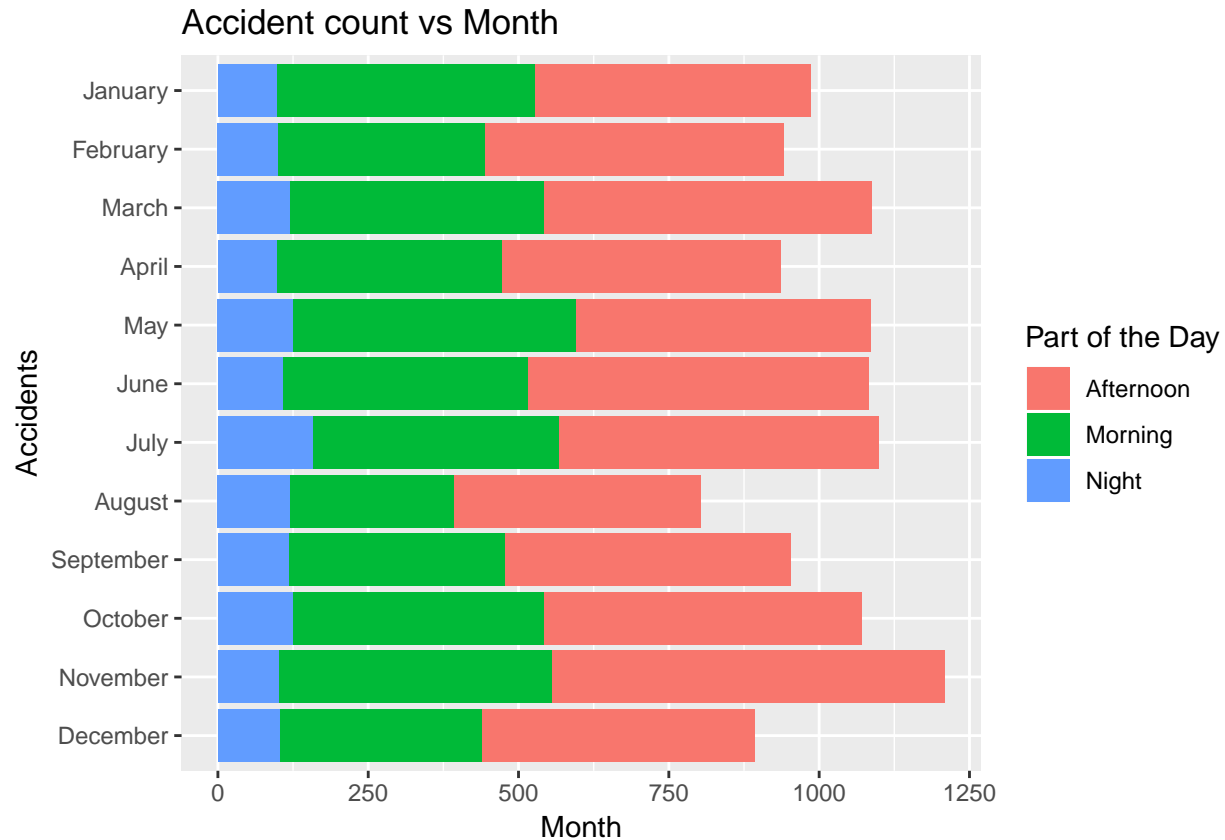


#### Analysis and finding:

In the second analysis, our assumption is that serious injury is associated with accidents when vehicle(s) involved. According to our result, the serious injury rate increases when the vehicle involved rate increases as well in the most districts except Eixample. Since most accidents clustered in Eixample as we mention before, it is reasonable that Eixample had high vehicle involved rate with both high serious and mild injury rates as well. Therefore, our result supports our assumption that when the accident involved with vehicle(s), it tended to lead to serious injury to the victim(s).

### Hypothesis 3 for accidents:

If time has an association with accident rates.



### Analysis and finding:

Aside from the analysis with geographical factor, we would also like to know if month and time during the year of 2017 have association with accidents. Our assumptions are most of the accidents would happen at night and there would be higher accident count between May and September than the rest of the month as it is the high season for tourists visiting the city. It appears that our assumptions are not completely supported by our result as most accidents happened during afternoon and the accidents were counted evenly in each month except the month of August.

We conducted more on secondary research to have a better understanding of the accident pattern and we noticed that our assumptions are not accurate mainly because of the cultural factor. In Spain, people tend to have Siesta time which is in the afternoon between 2 - 4pm. During Siesta, shops are closed and workers would either eat in the restaurants or go home and this can lead to higher chances of accidents happen. Additionally, while May - September attracts tourists, August is the vacation month for people in Barcelona. With many locals travel to other cities and countries, it is reasonable to have low accident counts during August. While our assumptions are not accurate, our result is matching to our secondary resources.

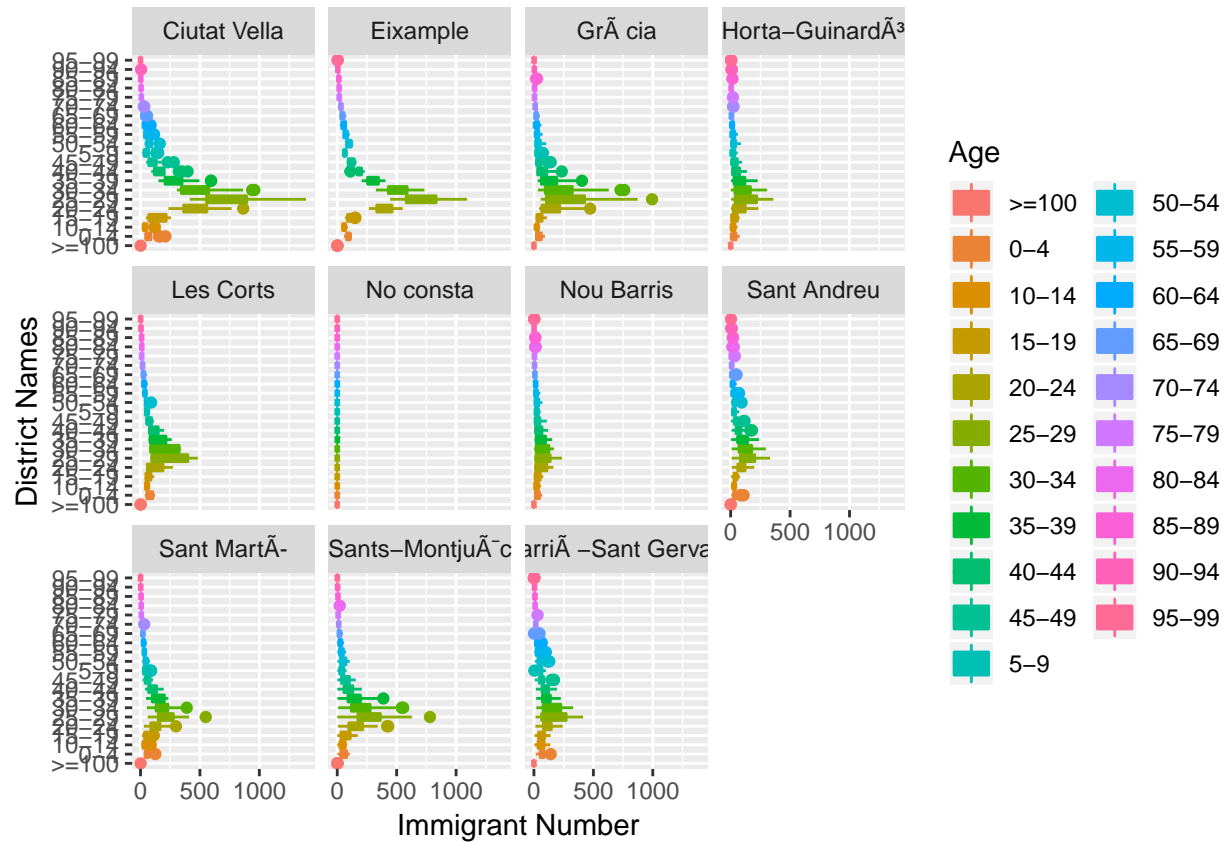
### Immigration Pattern:

In this part, we're looking at the data sets for the immigration to Barcelona. We will be looking at the data sets nationality, age group and sex from the Barcelona data. We will try to analyze graphically to see if there is a pattern between the immigration to Barcelona between these variables.



## Hypothesis 2 for immigration:

The younger age groups would make up the majority of the immigrants.



## Analysis and finding:

We can see from the box plot that Ciutat Vella, Eixample, Gràcia and Sants-Montjuïc attract the most immigrants. We can see the green plots, namely the ones representing age group from 25-50 are thicker and have a wider distribution. As the plots turn bluer and redder, the plots are sparser and thinner. From the box plot, we can see that age group 25-50 indeed have the largest immigrant numbers and they have a wider variance of distribution between each district. So our hypotheses were close.



### Hypothesis 3 for immigration:

Males would make up a much larger portion of immigrants than females.



### Analysis and finding:

We can see from the plots that, the immigrant numbers between males are similar. Although to our contrary, the female immigrants are not that much different from male immigrants, even surpassing them in some districts. Also, the female immigrants have a wider distribution, namely a bigger variance in a lot of districts. Again, the districts that attract the most immigrants are Ciutat Vella, Eixample, Gràcia and Les Corts.

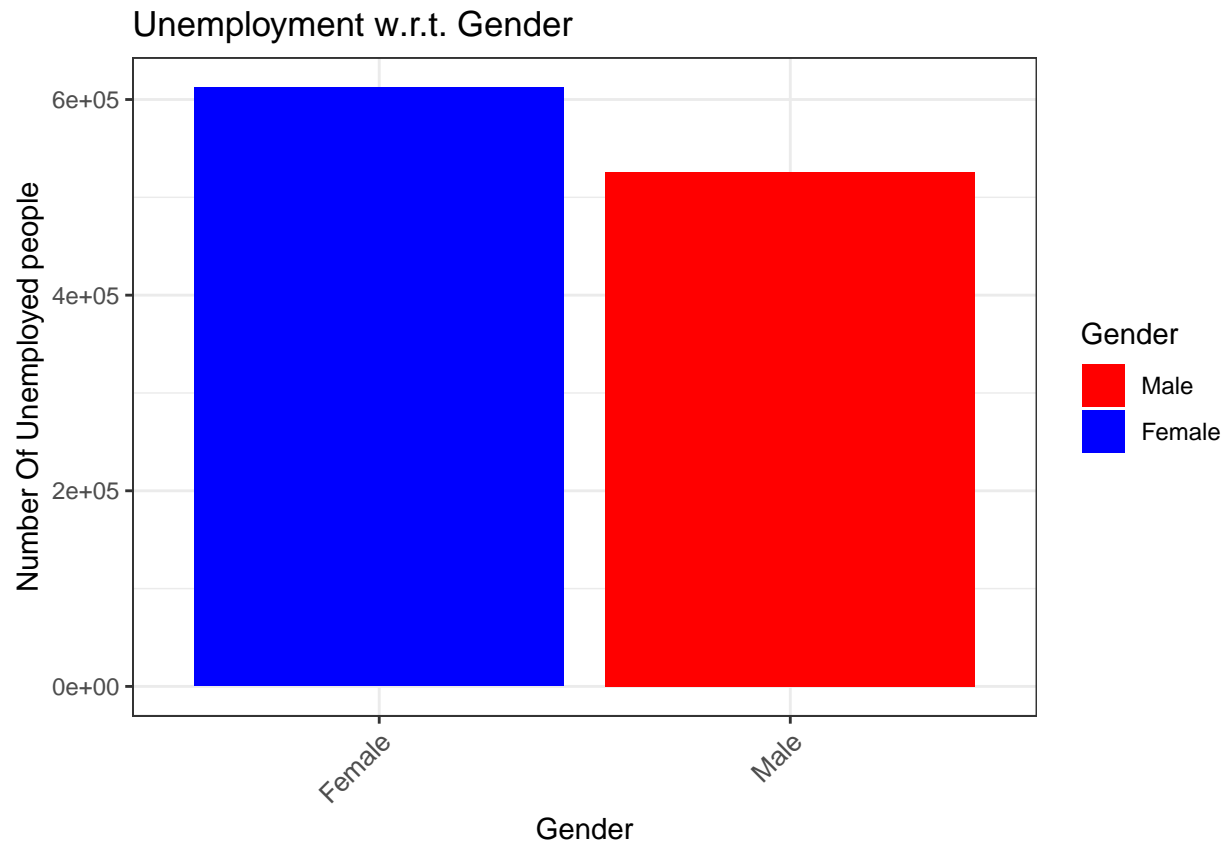
### Unemployment Analysis:

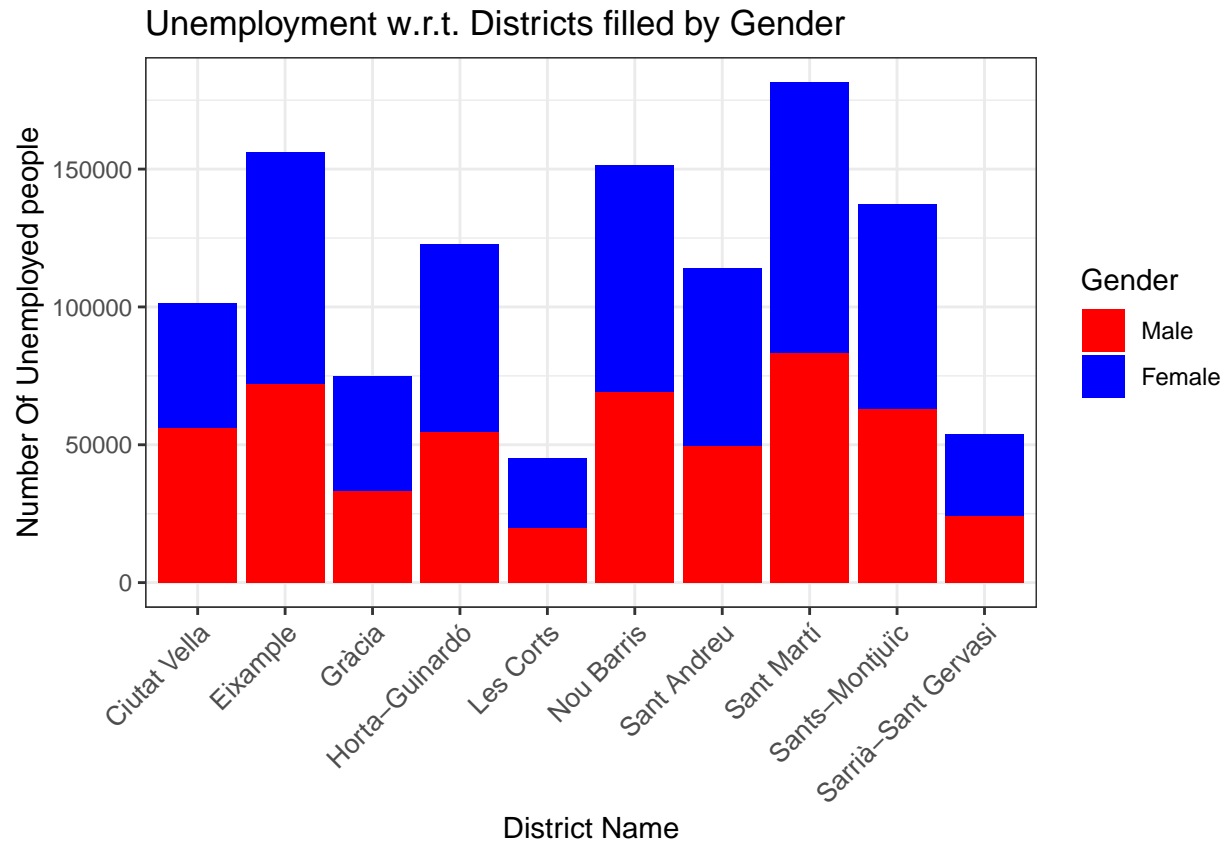
The unemployment data set consists of 14,208 records and a total of 9 variables - Year, Month, District Code, District Name, Neighborhood Code, Neighborhood Name, Gender, Occupation (Referring to type of unemployment), and Number.

Out of all the records we are only concerned about the year 2017 and thus for all the following visualizations, we have used data for unemployment from the year 2017.

Hypothesis 1 for Unemployment:

There will be higher unemployed males in Barcelona than Females.





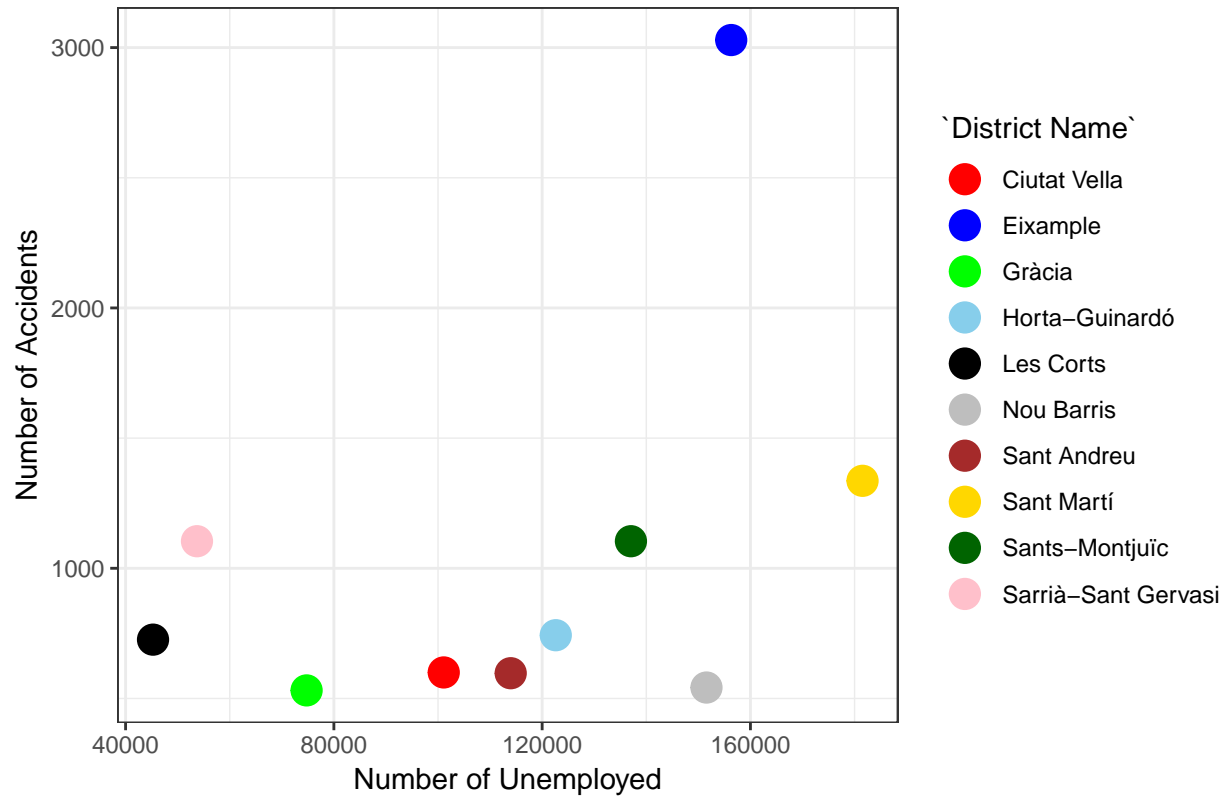
**Analysis and finding:**

To our surprise, there is almost an even distribution of male and female unemployed people in districts of barcelona and a higher number of unemployed females than males as a whole, which breaks our assumption.

**Hypothesis 2 for unemployment:**

Unemployment will be associated with the number of accidents in various districts.

**Number of Accidents for district names filled by unemployment**



**Analysis and finding:**

The top two cities with most number of accidents i.e. Eixample and Sant Marti also have the maximum number of unemployed people but Nou Barris (which means new neighborhoods in Spanish) also has the third highest number of unemployed people but has very few accidents which is due to the fact that Nou Barris transformed from a sub-urban area to an urban area recently and not being a metropolitan city was the most affected by economic crisis a few years back and could not recover fast from that and hence the explanation. It also does not have a lot of vehicles being a newly urban area.

Eixample and Sant Marti are top most populated districts in Barcelona and also have very high unemployment rates.

We can quite confidently say that unemployment has an association with the number of accidents. In other words, higher unemployment rates does affect accident rates in Barcelona.

**References:**

<https://www.barcelona.de/en/barcelona-figueres.html>

<https://mapfight.appspot.com/manhattan-vs-barcelona/manhattan-barcelona-spain-size-comparison>

<https://santorinidave.com/best-time-to-visit-barcelona>

<https://www.tripsavvy.com/why-do-you-sleep-during-siesta-1644327>

<https://www.nytimes.com/2017/07/28/business/spain-europe-economy-recovery-unemployment.html>

[https://www.theepochtimes.com/spain-recovering-from-financial-crisis-but-fears-a-new-recession\\_2725471.html](https://www.theepochtimes.com/spain-recovering-from-financial-crisis-but-fears-a-new-recession_2725471.html)

Ouesteali, Walid and others. "Determinants of urban sprawl in European cities." *Urban Studies Journal* (July, 2015)

Evans, Leonard. "Traffic Fatality Reduction: United States Compared With 25 Other Countries." *American Journal of Public Health* (August 2014)

Alonso, Andres and others. "Measuring Negative Synergies of Urban Sprawl and Economic Crisis over Public Transport Efficiency: The Case Of Spain" *International Regional Science Review* (January, 2017)

Teresa Hontagut. "City Report: Barcelona" Wilco