Data Analysis and Visualization in R (IN2339)

Case Study

[Esra Yucal, Navin Kumar Trivedi, Utku Atay, Vinit Kumar]

2022-01-23

Motivation

The motivation of this case study is to find an underlying reason for the stagnant population of the City of Barcelona even with an increasing number of deaths as compared to births in the corresponding years. The goal of this case study is to deep dive into the provided data, finding a proper explanation for the same.

Data Preparation

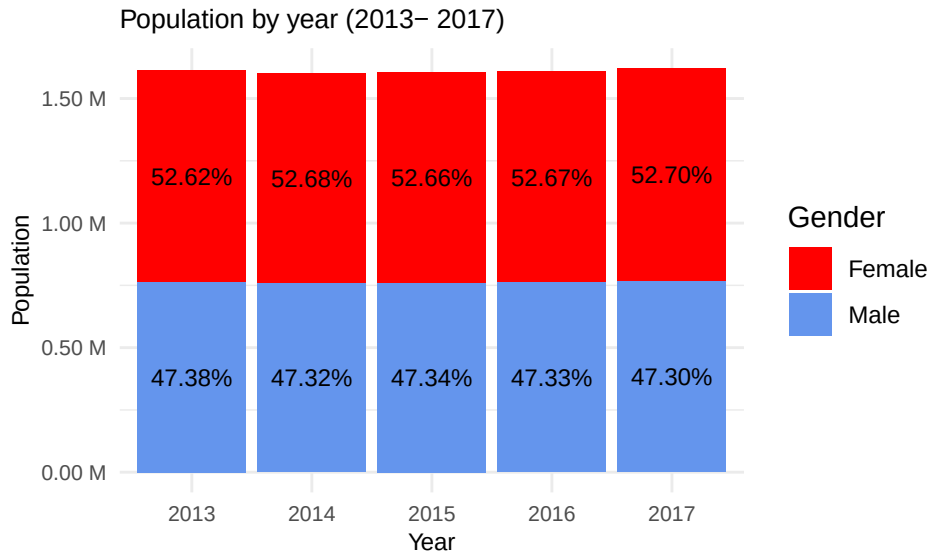
In this section, we perform the needed data preparation steps needed for the analysis. Unnecessary chunk codes are omitted in the compiled pdf-file. Firstly, we load some libraries, then read the csv files and examine each relevant file one by one.

After that, we try to tidy data in order to work with them more clearly. "Births", "Immigration by Sex" and "Immigration by Nationality" datasets have improper colnames like "District Name", because they have spaces in the column names. We changed the columns by using "make.names" function and convert them appropiate names like "District.Name" We change "Neighbor.Code" and "District.Code" variables from integer to factor for all files. Additionally "Year" data is also given in integer, therefore we change it also into factor because data includes a certain number of years. Finally, we tidy and manipulate data with functions like melt and cast. We continue to use these manipulation techniques when necessary in the Data Analysis section.

Data Analysis

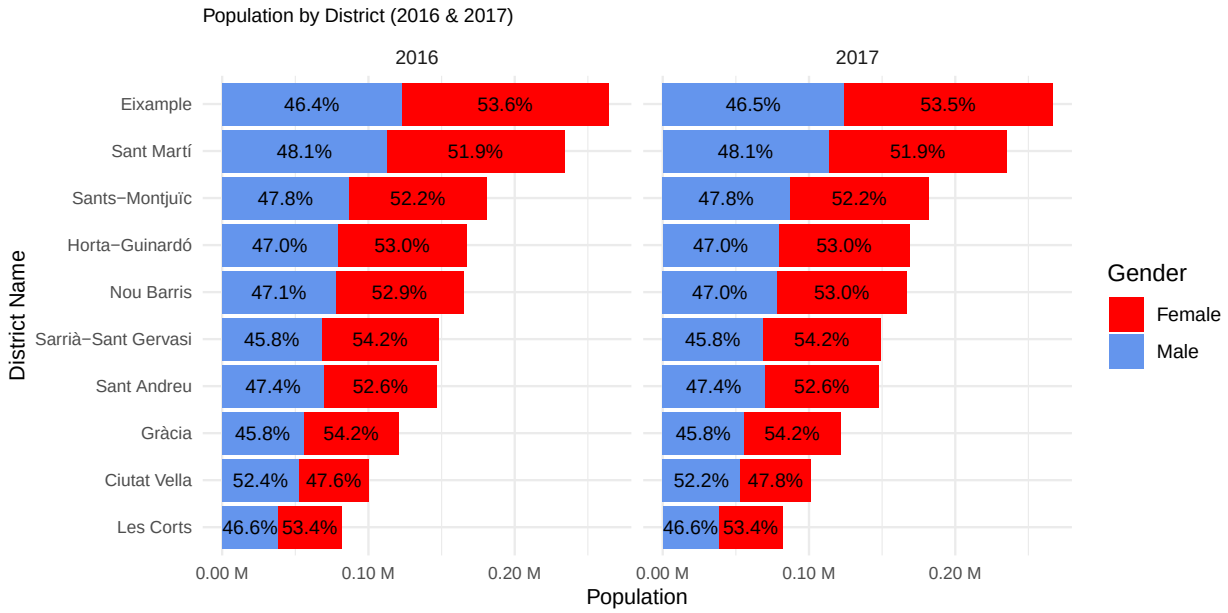
Population by Year & Gender

Initially, we start with the population dataset and show the population of Barcelona by Gender and Year from 2013-2017. The population increases slightly in recent years. The portion of female and male in the population remained almost the same. Female's percentage is somewhat more then male's one.



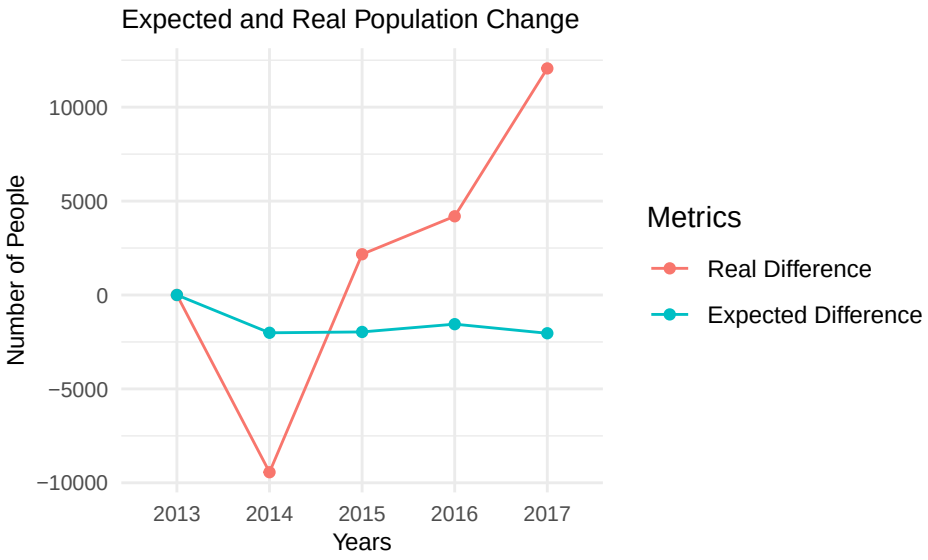
Population by District & Gender

Afterwards we analyze the district population. Now, we show the population of Barcelona by District and Gender for the Year 2016 and 2017. Eixample and Sant Marti are the most populated districts in Barcelona. The district's population and gender distribution in district are almost same in both years.



Growth Rate

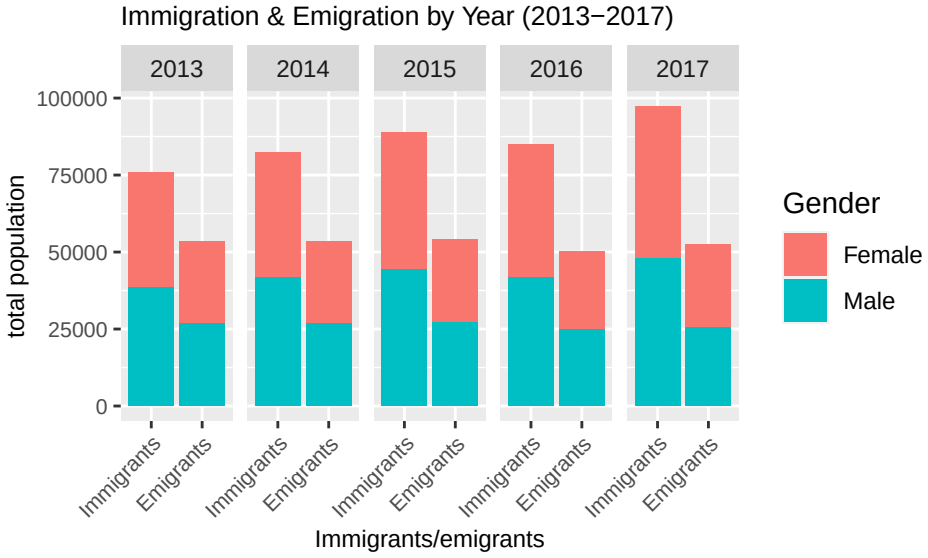
As already mentioned above, the population of Barcelona is increasing slightly over years. However when we examine the births and deaths statistics, it could be expected that population is decreasing because the number of deaths is higher than births. The difference between births and deaths is called expected population change while the current population change is named real population change. Now we plot those changes over years.



The graph indicates that real change is greater than expected one, which means there should be another factor, which cause the population increase in Barcelona. Our hypthese is that there is relationship between population increase and immigrants.

Immigrants by year and gender

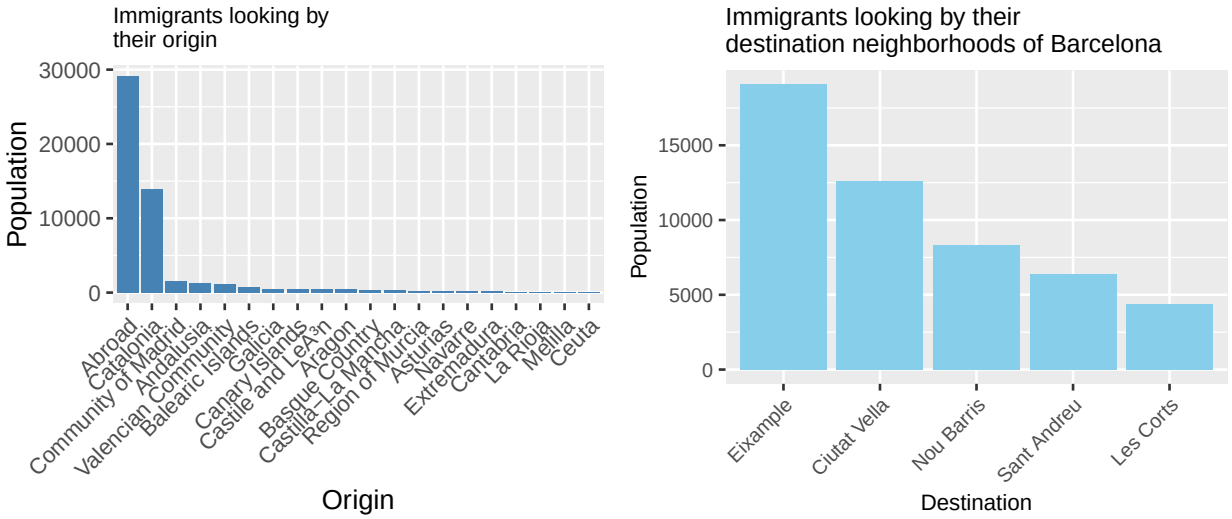
The file "immigrants_emigrants_by_sex.csv" has been used, which contains information about immigration and emigration numbers based on year, gender, district name, neighborhood name, and so on.



Between the years 2013 and 2017, immigration and emigration were examined, as well as by gender using the geom bar. According to the bar chart, immigration remained relatively stable between 2013 and 2017.However, as the graph shows, immigration has increased significantly since 2013. Furthermore, the majority of the immigrants were female.

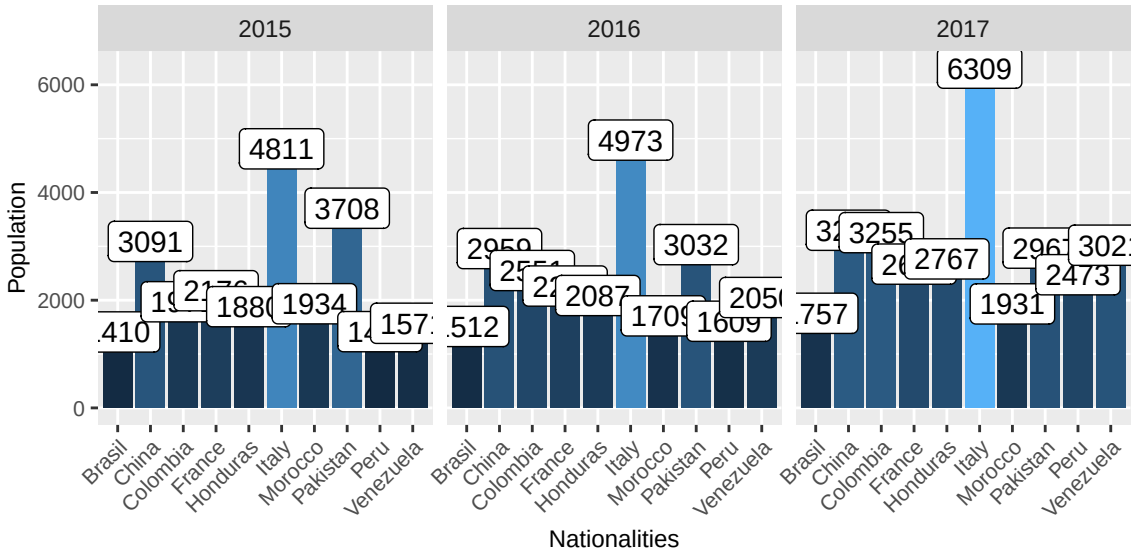
Immigrants by district and population change in the district

The file "immigrants_emigrants_by_destination2.csv" was used to investigate the number of people who immigrate to and emigrate from Barcelona, as well as the neighborhoods where they live or move in 2017. After a quick check of data and a search of resources, districts of Barcelona were discovered and used for data separation (filtering) based on immigration and emigration moves.



The first plot shows the distribution of immigrants based on their origin location, while the second plot shows the distribution of immigrants based on their destination to Barcelona neighborhoods.Vast majority of immigrants are from countries other than Spain.(coming from abroad) The majority of the immigration has been dissolved to the Eixample and Sant Marti neighborhoods of Barcelona.

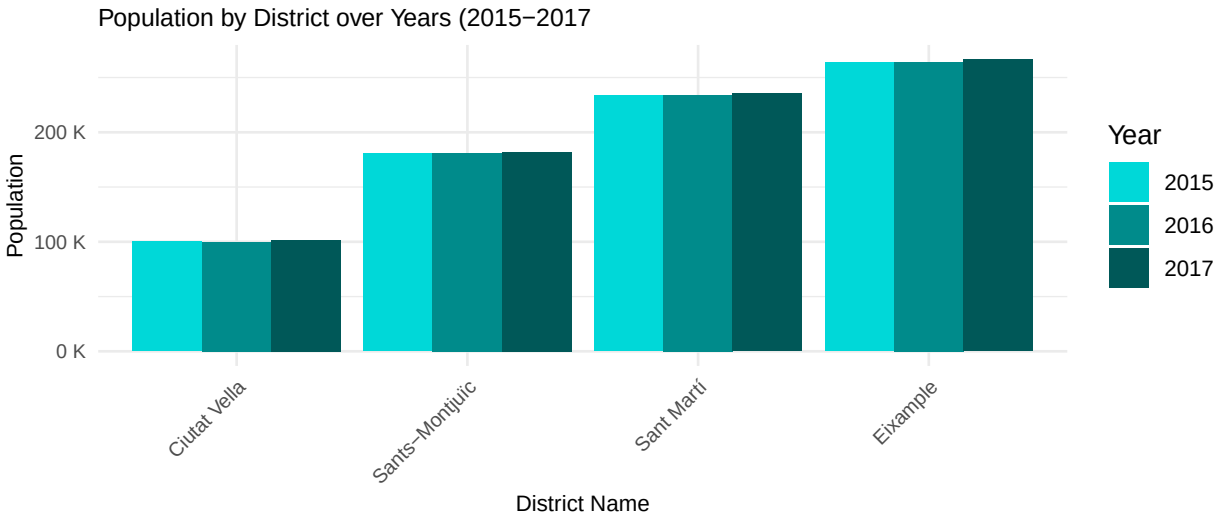
Immigrants by nationality

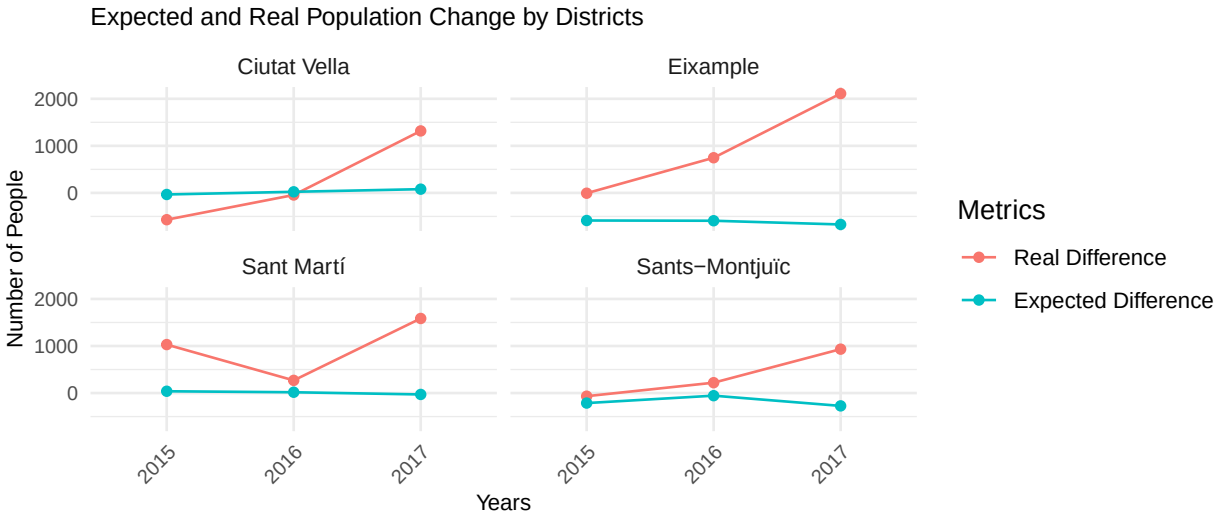


According to the graphs, Italy had the highest rate of immigration. Furthermore, when we look at immigration from countries separately by year, we see that the number of immigrants from other countries does not change much, but immigration from Italy increases every year, and there was a significant increase in 2017.

Population change in most immigrated districts

Lastly, we examine the districts, which are the most immigrated. Those are "Eixample","Sant Martí","Sants-Montjuïc" and "Ciutat Vella". The first graph shows that the population in these districts has increased over years. When we compare real and expected population differences in these districts, it can be clearly said that expected difference is much higher than real one, especially in Eixample and Sants-Montjuïc. These graphs also support our claim, which is population is increased only because of immigrants. Our cofounding factors are population increase and number of immigrations. The former is our dependent variable and the latter is the independent variable.





Conclusion

Hence, we conclude as shown above that there is a difference between the excepted and the real population change. We attribute this to the Immigrants which explains the stagnant population of Barcelona which otherwise should have decreased if we consider the natural change of birth and death over the years.

We also summarize on the population by district with the most immigrants from which we conclude that Eixample has the largest population with Sants-Montjuïc having the highest gap between the expected and the real population change.