

Final Project

Collecting, Storing and Retrieving data course (DA5020), Fall 2019

Milad Tatari

Abstract

Annual income of different jobs varies over time and also locations. A job title might be in a high demand in a time period (like data scientist) or the demand decreases in the market after a few years. In addition, location of the job can significantly affect the income and the number of jobs. For instance, given Boston as an education city with a lot of universities/colleges, probably, it should have many teachers/professors. Here, in this project, I am going to use the data of annual income of all jobs in greater Boston area available in Boston.gov website for different years. This project will discuss what are most frequent job titles and how they have been changed in terms of income over the last 4 years (2015 to 2019) depending on both location and time. Actually, the goal is to address what area is better in terms of income for the selected job titles and how it varied over time.

Data Collecting and analysis

Data is collected from Boston.gov website¹ for our analysis. For each year, there is a dataset as a .CSV file that is downloaded and can be read in Rstudio for the analysis. What we are interested in analyzing here, is the annual income of different job title. For this column, data is not clean since it has to be numeric to do averaging and other math calculation. In order to do this, some regex (Regular Expressions)² is employed to remove “,” and “\$” as a great tool in data analysis. The other tools that are used for data cleaning is “dplyr” package³ to filter some specific job title and group_by different columns. Also, we have used ggplot2 package to visualize the results as a pie chart, bar plot and map plots. To plot the map, as we have the postal codes we need to have the latitude and longitude of different zip code areas. This has been done by zip code package for the

US and then merging the zip code data with our data based on postal codes as a primary key. Given this, we can discuss the results that we have obtained to see what jobs in what area have a higher salary and how it has changed over years.

Results and Discussion

First, we want to analyze the most frequent job titles in Boston area as follows:

Cafeteria Manager, Nurse, Parking Meter Supervisor, Police Officer, School Psychologist, Sr Data Proc Sys Anl, Student Intern and Teacher.

Figure 1 is a pie chart showing the most frequent job titles in Boston area. As it can be seen from Figure 1, the most frequent job titles are “police officers” and “teachers”, so we focus on these two jobs in the current study to see how their salaries have changed over years. It should be noted that I have not used directly the plots of rstudio. I have revised all plots in Adobe Illustrator software for better quality and consistency.

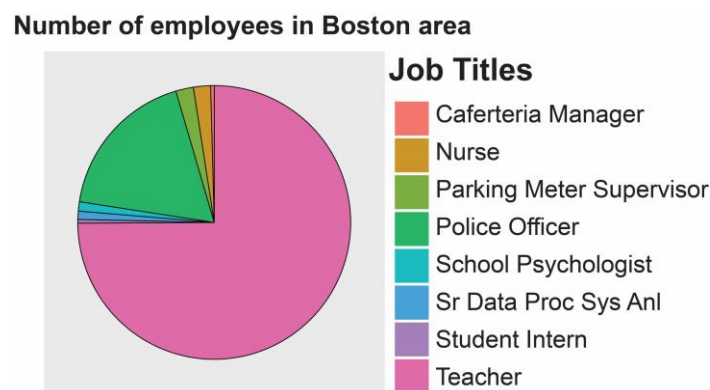


Figure 1. Most frequent job title in Boston area

Figure 2 shows the average annual income of the select jobs in different years. We see that police officers have the highest salary and it is almost increasing from 2015 to 2018. For some job, the trend is not increasing over years for example Nurses and Cafeteria Manager. It gives us the point that maybe, in recent years they are not in high demand. Senior data scientists have also a good

salary and had a huge increase from 2015 to 2016 and then it decreased. Intern students also has the lowest salary as it is not a full time job throughout the year and it is only for 3 to 4 months. Teachers, nurses and data analysts have average annual income of \$80,000

Average annual income for different jobs in Boston

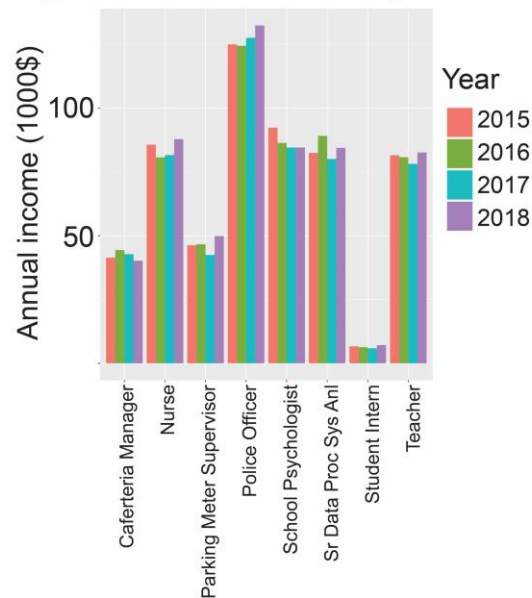


Figure 2. Average annual income of different job titles in Boston area.

Now we want to focus on 5 different areas in Boston as follows (to see the trend of annual income over years and location):

02148 North Boston,

02135 West Boston

02128 East Boston

02136, 02132 South Boston

02110, 02108, 02109, 02113, 02114 Downtown Boston

Figure 3 illustrates the number of police officers and teachers in MA map in the above 5 zones. It can be seen that in all areas, number of teachers is more than that of police officers. The interesting point is that the highest numbers of police officers and teachers are in South Boston which is around 450.



Figure 3. Number of police officers and teachers in different areas of Boston with zoom screen to better show the results.

We have used the zoom-in map plot to compare the trend of annual income change in different areas of Boston as shown in Figure 4 A&B.

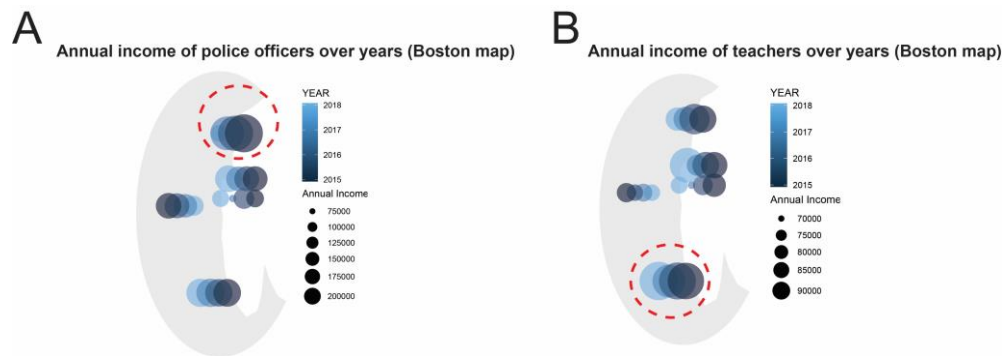


Figure 4. Annual income over years in Boston area for A) Police officers and B) Teachers

Figure 4 shows that the highest income of police officers in north Boston has decreased from 2015 to 2018. In south Boston, the highest income of teachers in south Boston is almost constant in 2015 to 2018. For teachers in east Boston the trend shows that their income also has decreased.

Conclusion

In this project, analyzing the results show that the teachers and police officers have the highest number of employees in Boston. Income of police officers in north Boston is highest among other regions and decreased from 2015 to 2018. Our results also show that the income of teachers in

south Boston is highest among other regions and almost constant from 2015 to 2018. Therefore, the results give us an idea if we want to go with these jobs, which area has the highest salary and how it might change in future.

References

- [1] <https://data.boston.gov/dataset/employee-earnings-report>
- [2] <https://www.regular-expressions.info/>
- [3] <https://cran.r-project.org/web/packages/dplyr/dplyr.pdf>