Crimes in the City of Boston during 2022

Joaquin Alvarez Galnares

February 2023

Abstract

This report corresponds to the first exam of Causal Inference, taught by Prof. Mauricio Romero. The associated code for replication and data can be found **here**.

Figure 1: This figure was constructed using R and corresponds to the map of Boston.

Figure 1 was built with the help of Esteban Moro's blog.

1 Context

Analyze Boston manages open data for the City of Boston. In particular, the website publishes yearly data of crime incident reports from the Boston Police Department. In this report we will analyze the data from 2022 using R as a statistical programming language that will be helpful to get an insight about the criminal situation in the City of Boston during 2022.

2 The general landscape

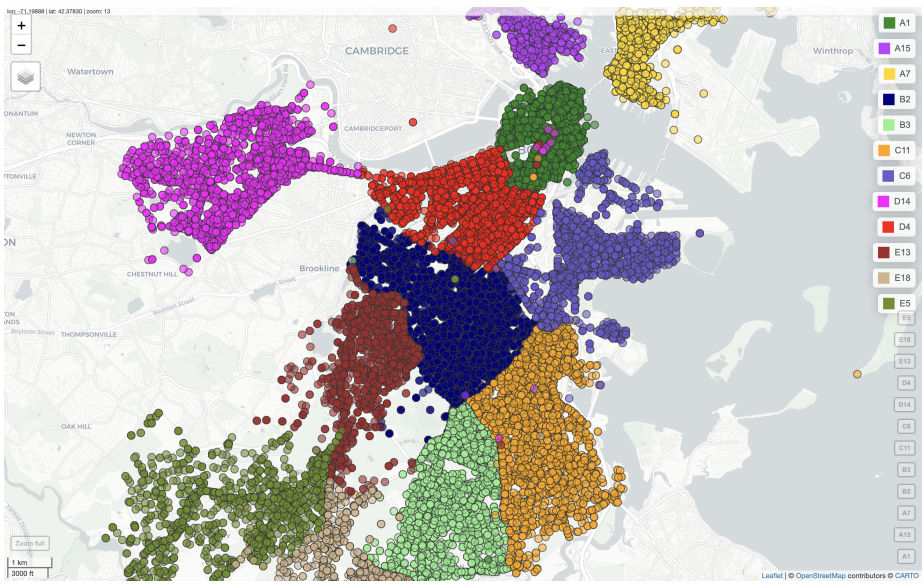

Figure 2: Map of the crimes stratified by district.

We recover practically all the City of Boston by plotting the locations of the reports. It should be emphasized that the data set contains not only crime events but also other type of events such as suicide reports. We should keep in mind the structure and the distribution of the districts throughout the city since we will exploit it during our analysis later on.

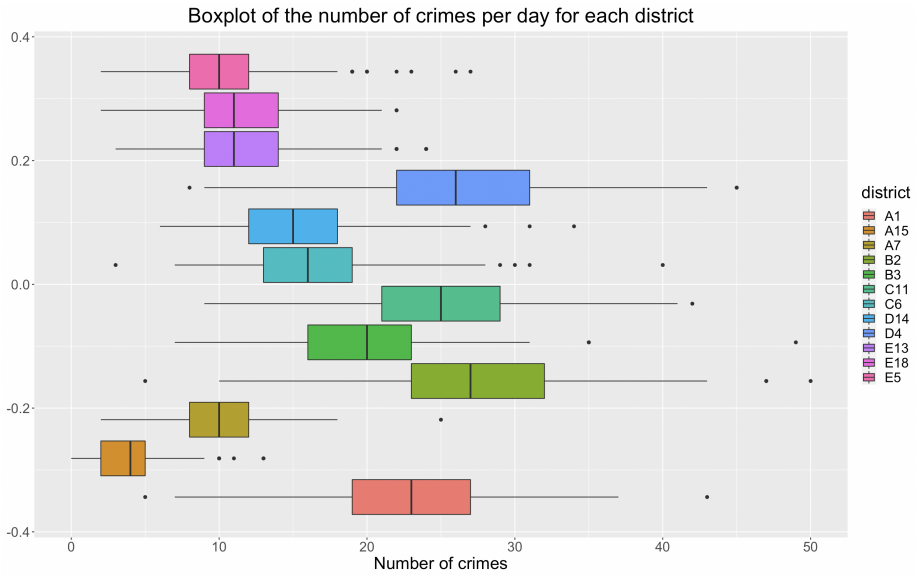

Figure 3: Box plot of daily crimes per district. Bigger arms represent bigger variance. Points outside the arms correspond to outlier observations. The middle line in each box corresponds to the median of crimes in each district.

The Figure 3 allows us to visualize multiple properties and make multiple comparisons simultaneously about the crimes in each district in Boston. For example, we can appreciate a lot of variance for D4, B2 districts, because the arms of the box plots are very long. The same for E5, whereas A15 has very low variance, and the smallest mean. Actually the box plots allow us to see a very interesting insight: those neighborhoods with a big mean in the number of crimes per day also tend to have a bigger variance! This suggests a strong (and positive) correlation between mean and variance! We verify that in the Figure 4.

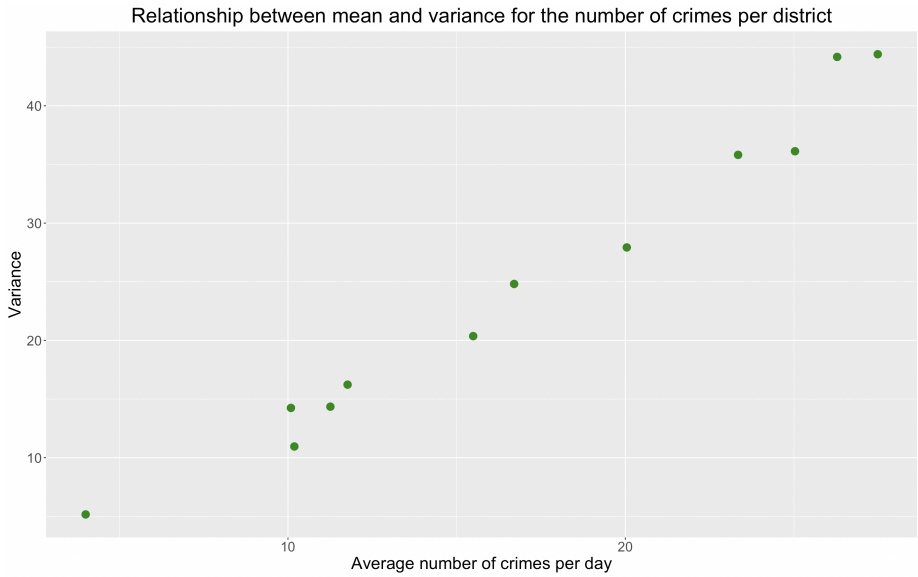

Figure 4: Linear relationship between mean and variance of the crimes for each district.

We obtained a correlation equal to 0.988 between the mean and the variance for the number of crimes in each district, which confirms the intuition we pointed out with the Figure 3. In simple words: for those districts that have a big number of crimes per day (on average), there is more uncertainty (variance) about the number of crimes we will observe on an arbitrarily chosen day.

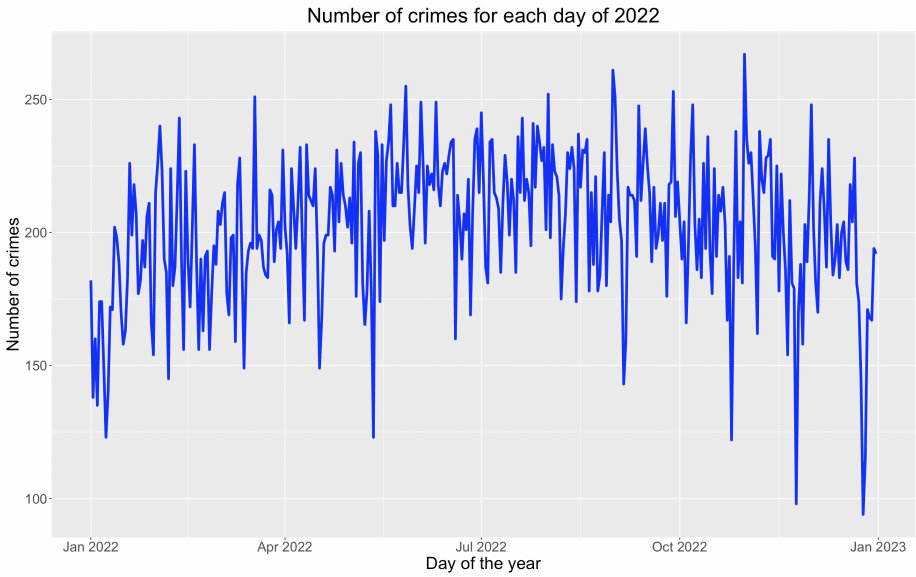

Figure 5: Time-series of the number of crimes in Boston during 2022.

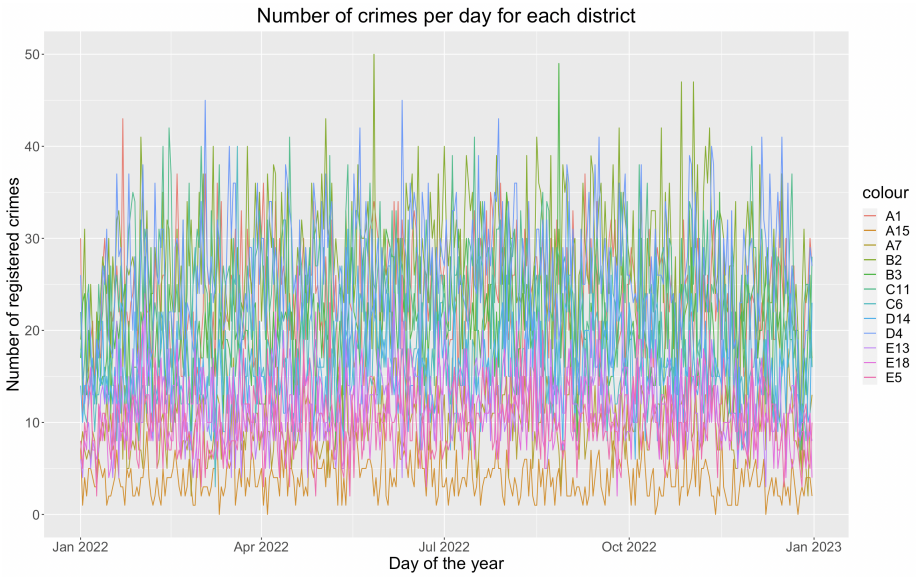

Figure 6: Stratified time-series by district.

Decomposing Figure 5, Figure 6 allows us to visualize idiosyncratic tenden-

cies for each district. For example, coherent with the remarks we made about Figure 3, we can observe high volatility for the number of crimes in district B2 (the green peaks at the top correspond to that district), whereas A15 presents a stable trajectory (orange at the bottom) with low variance, and a low average number of crimes per day.

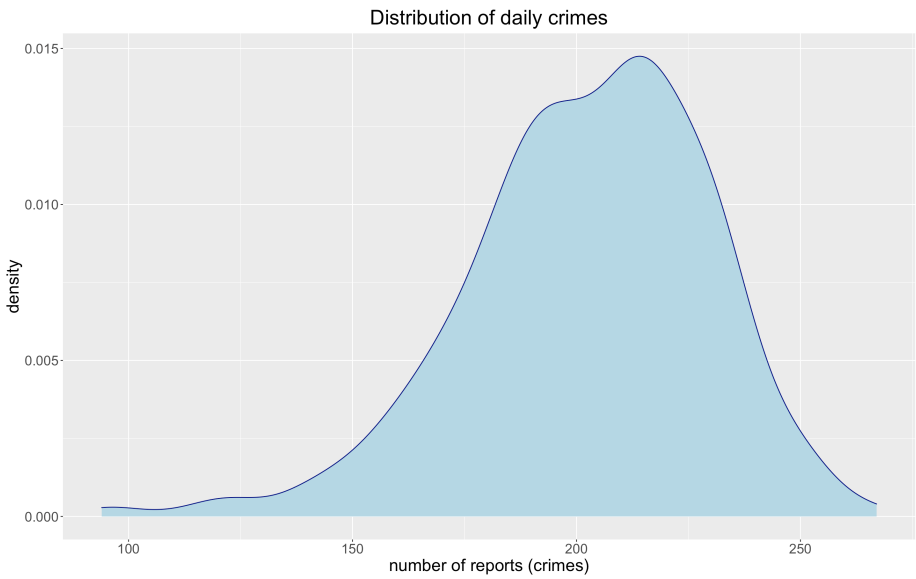

Figure 7: Distribution of the total number of crimes per day.

We appreciate a bell-shaped distribution that's not symmetric and has a heavy left tail. The mean number of crimes per day was 202.3342 and it has a standard deviation of 27.15742.

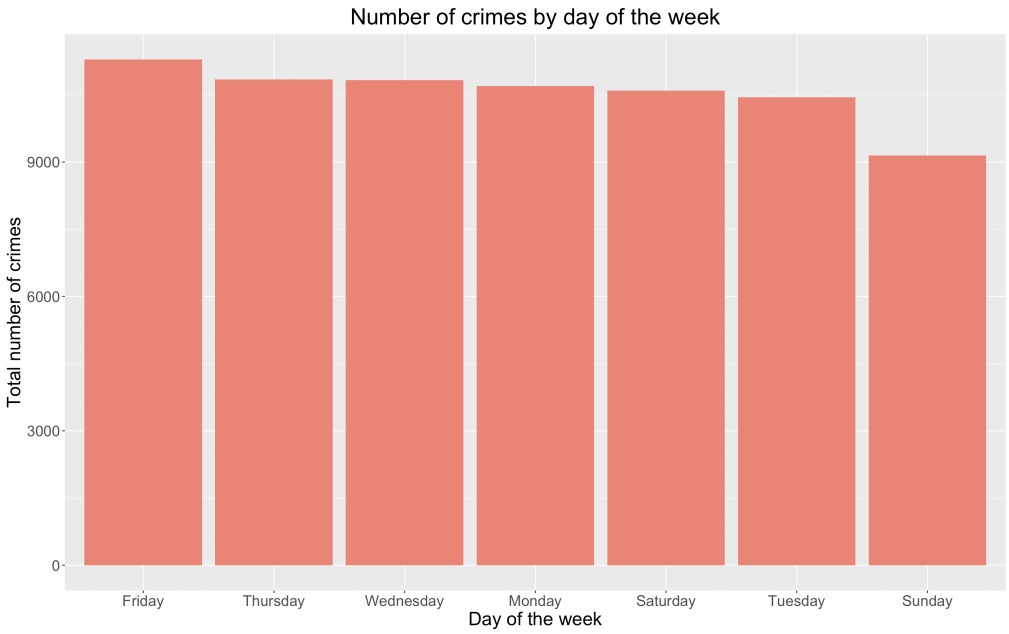

Figure 8: Bar chart of the number of crimes per day in descending order.

Friday is the day with most police reports, while Sunday is the day with less police reports. We could speculate about why this is the case, for example, people make plans during Fridays that make them more vulnerable to assaults and robberies.

3 The Boston Marathon

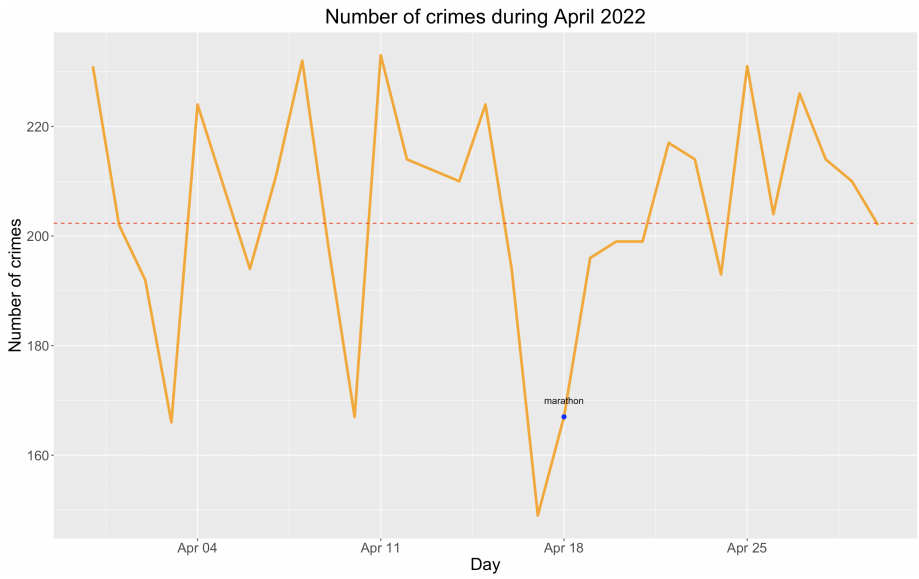

Figure 9: Dashed line corresponds to the average number of crimes per day.

The Boston Marathon of 2022 took place in April 18, 2022. It is interesting to observe that the number of crimes is very low during the weekend of the marathon, actually below average. It is important to keep in mind that the Boston Marathon is a big event that involves a lot of tourism and activities throughout the whole city. Moreover, in the 2013 Boston Marathon there was a terrorist attack that consisted of a bombing close to the finish line of the race. Thus, it should be no surprise that the authorities take security *very seriously* during this event. And indeed, the authorities manage to take down crimes during the weekend as we can see in Figure 9.

4 Dangerous streets

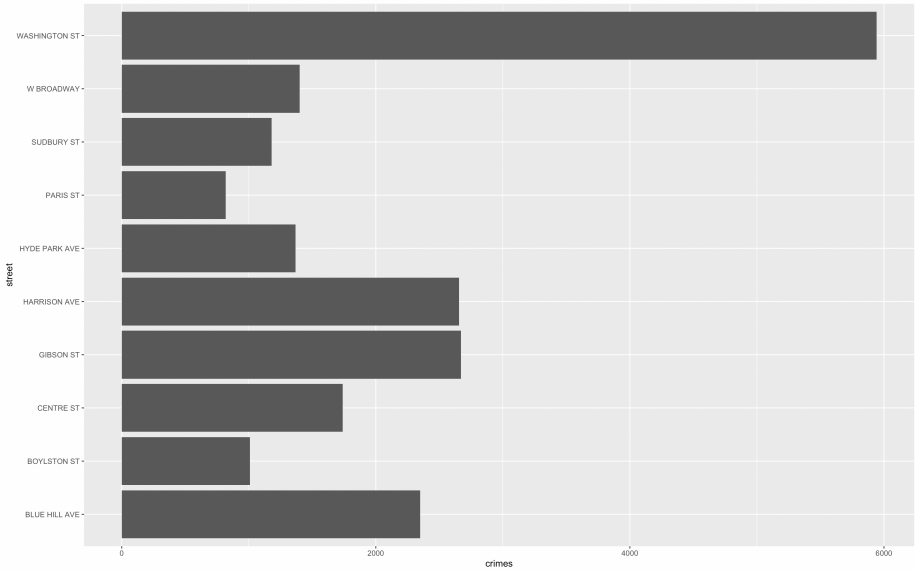

Figure 10: Streets with most registered police reports.

If we try to find an explanation for this situation, something must be considered: Washington St. is the street with most registered crimes in Boston, but it is also the longest street in Boston. So it's reasonable that it has many crimes. It is a large street, thus more representative than shorter streets.

5 Beacon Hill and the Back Bay

Beacon Hill and the Back Bay are the most expensive residential neighborhoods in the city. As the following figures show, they are very safe zones compared to the rest of Boston.

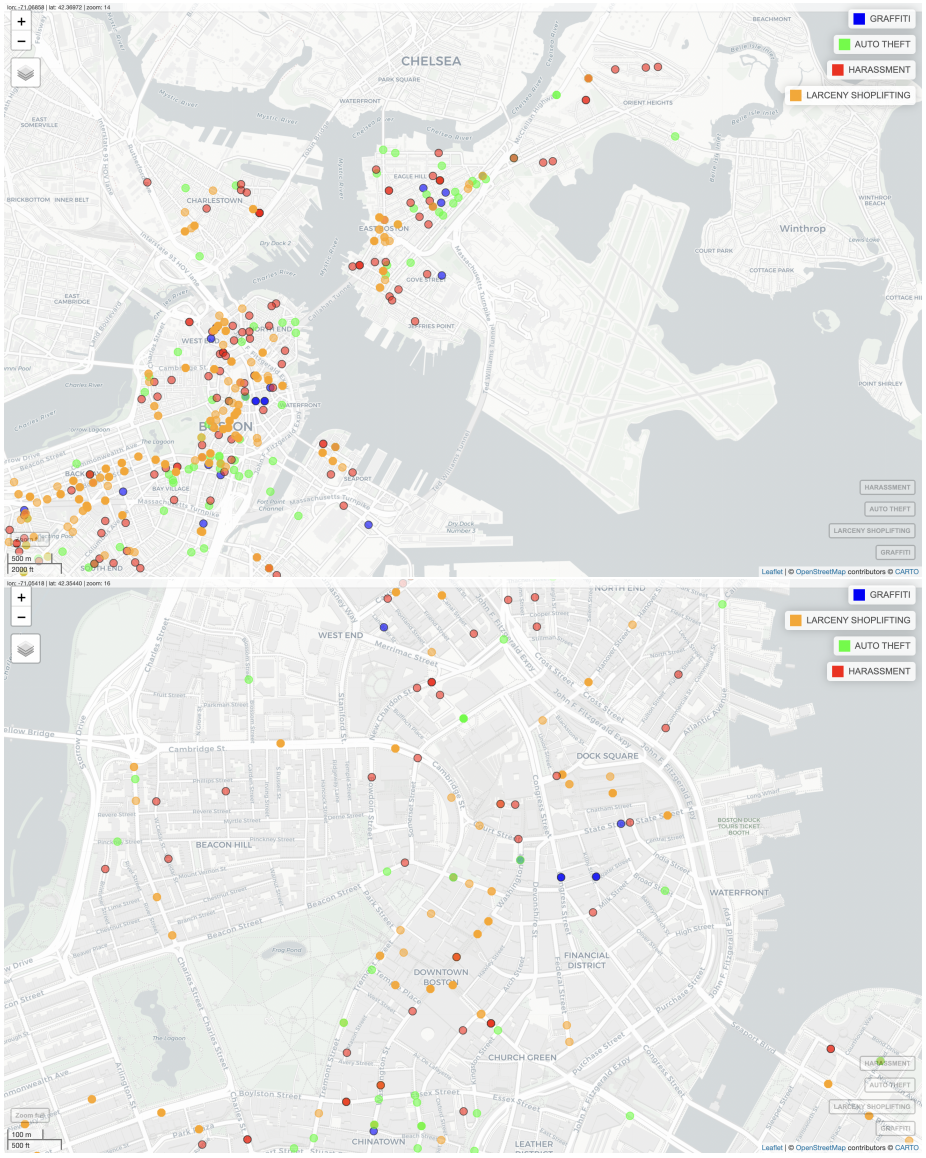

Figure 11: Boston's map with stratified observations of some of the most common types of crime in the city.

Beacon Hill is located at the north of the Boston Common (Boston's main park), whereas the Back Bay is located at the West of the Boston Common. In the upper map of figure 11 we appreciate that there are barely any crimes in Beacon Hill compared to other zones of the city (for some of the most common

types of crime). The lower map in figure 11 makes a zoom into Beacon Hill so that the viewer can appreciate it in more detail.

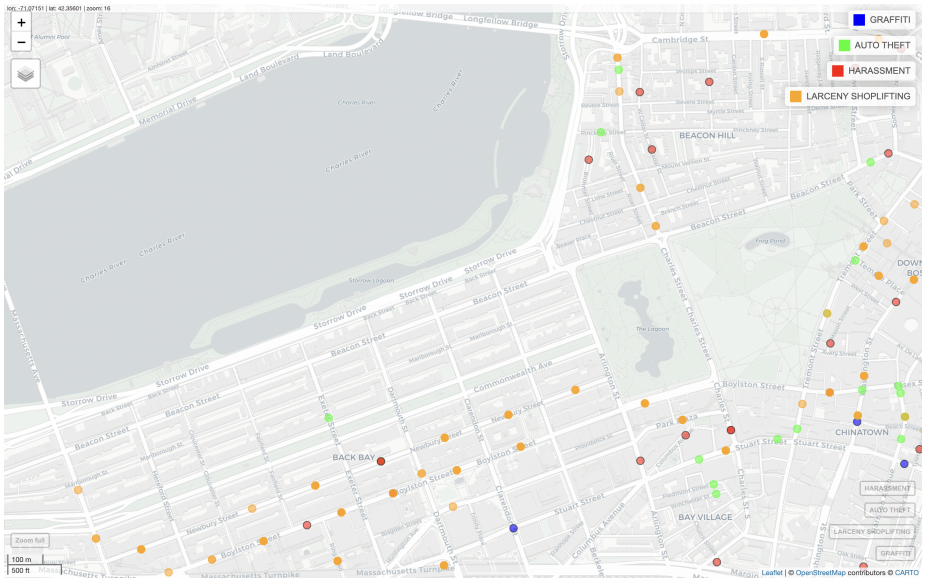

Figure 12: Back Bay (the streets that yield to the Charles River at the west of the Boston Common) area has few crimes compared to the rest of the city.

We observe practically no crimes in the Back Bay (for some of the most common crime types). Notice that Boylston Street and New Bury St. have larceny shoplifting crimes though out all the length of the streets. This shouldn't be something novel if we account for the fact that these streets are packed with stores. Hence, these streets are particularly vulnerable to larceny shoplifting.

For the streets at the North of the Commonwealth avenue we only observe one auto theft, for the categories of crimes being considered.

6 "The first law of geography: Everything is related to everything else, but near things are more related than distant things." Waldo R. Tobler (1970)

We analyze the data having this perspective in mind. However, we should be careful with oversimplifying reality: we are not accounting for population density of each district, nor ethnicity, unemployment rate, average income per resident and many other factors that could impact the relationship between

crimes in each district in order to explain correlated districts which are quite distant between one another.

However,

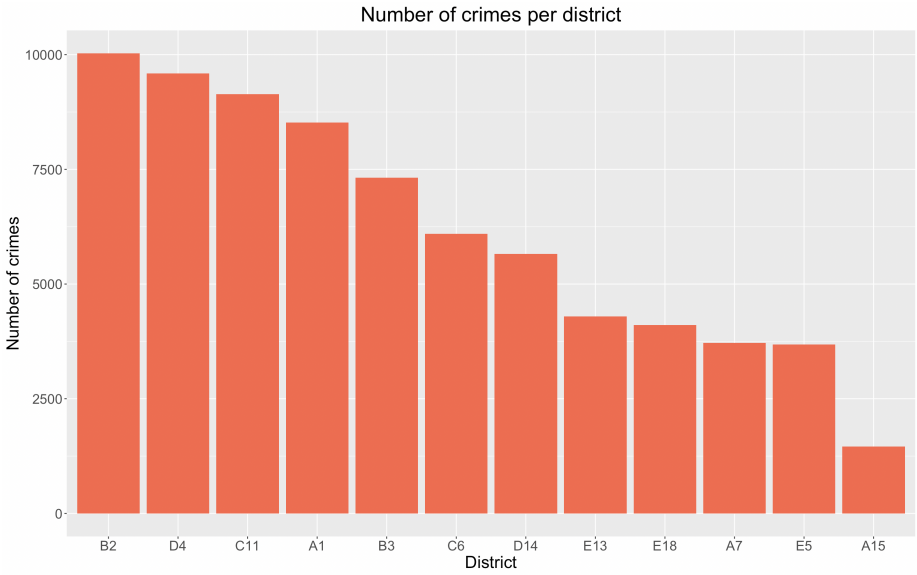

Figure 13: Number of crimes per district.

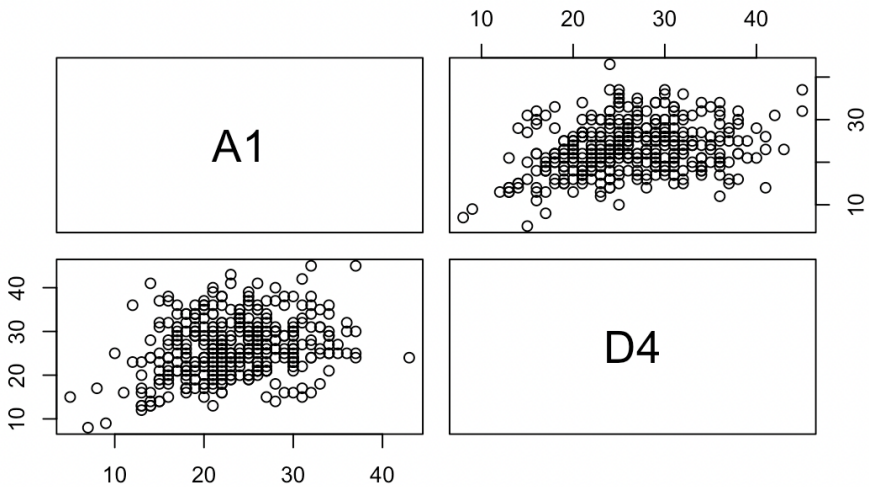

Figure 14: Scatter plot showing joint number of crimes per day in districts A1 and D4, which are neighbor districts with a considerable correlation.

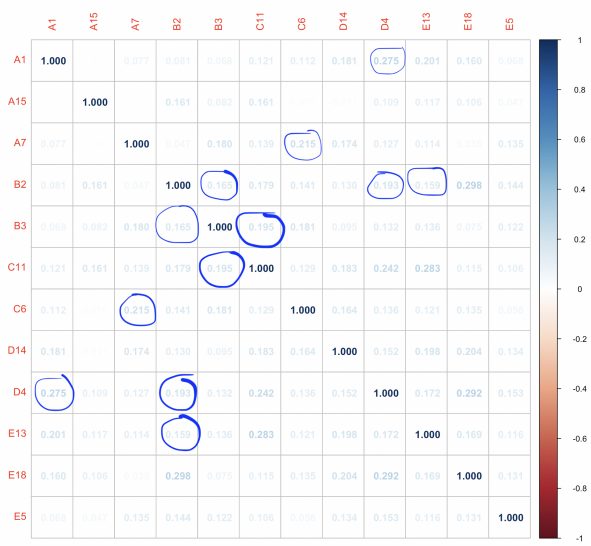

Figure 15: Correlation matrix between districts. Circled numbers correspond to high correlations associated to neighbor districts as can be verified with figure 2.

It should be noted that all correlations are greater than zero. Most of the strongest correlations correspond to neighbor districts (as marked by the circles in figure 15). However there are exceptions, for example, B2 and E18 are highly correlated (the correlation between the number of crimes in both districts is 0.298), however, as seen in figure 2, they are distant to each other.

7 Outliers in Cambridge!

It is well known that MIT and Harvard are located at Cambridge. Thus, it is no coincidence that some of the most brilliant minds and scholars in the world actually live there! So an interesting question is, how safe is Cambridge? Out of all police reports in the data set there were only 2 reports in Cambridge! That is not to say that there are no crimes in the zone at all! Sources reveal different information about crimes in Cambridge, for example some blogs report higher crime rates, but one thing remains true: Cambridge is one of the safest zones in Massachusetts. It is also important to consider that Cambridge isn't considered part of Boston *perse*. It is separated by the Charles River as the maps show. So perhaps The Boston Police Department generally isn't responsible for crimes in that zone.

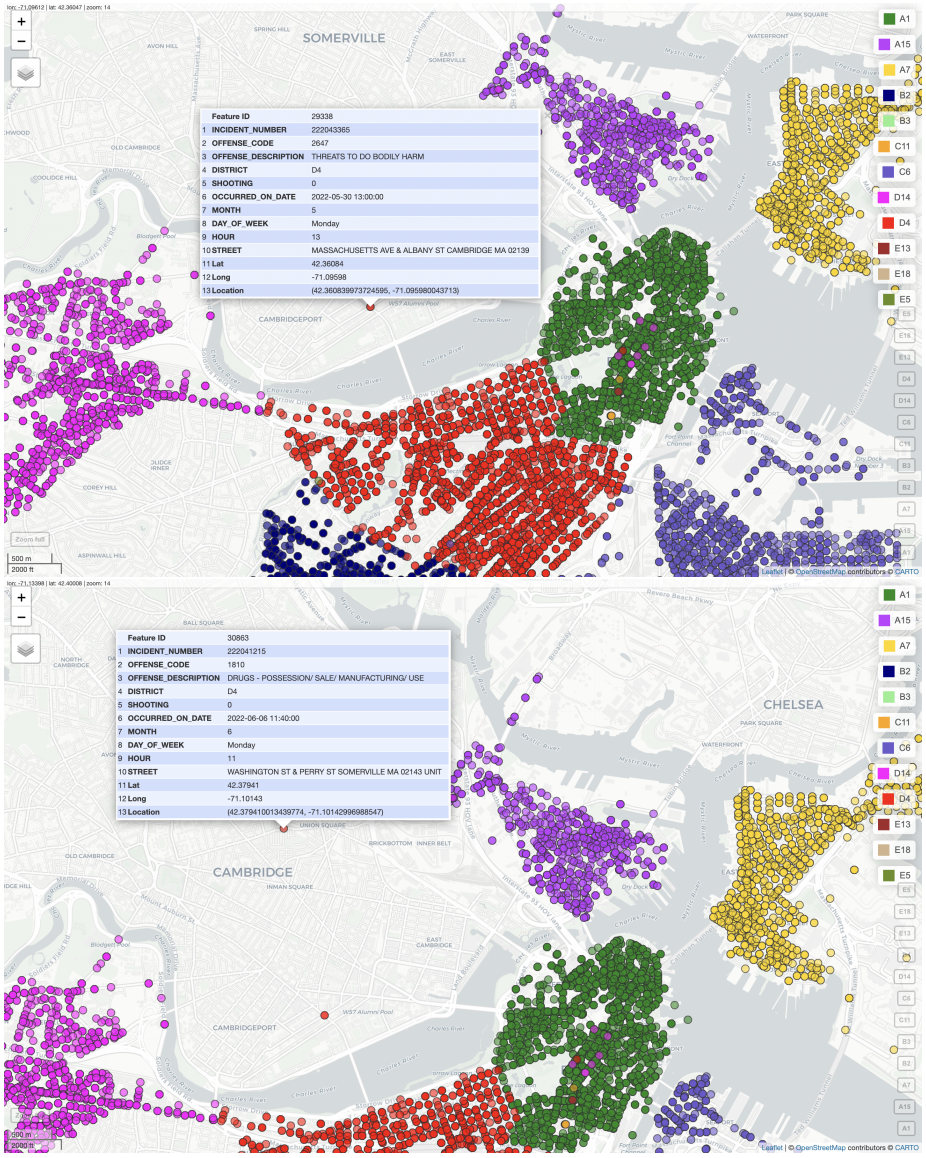

Figure 16: Police Reports in Cambridge with the location and a description. The crime at the bottom happened close to Harvard Square and the one from the map on top occurred near MIT.