BU Department of Computer Science

bu.edu/cs

Analyzing and Improving Boston Policing

Eli Saracino

Inspiration

The Boston Police Department has been criticized in recent years for its handling of stop and frisks, with citizen's arguing that the BPD disproportionately targets minorities. Bringing an in-depth statistical analysis to this problem might shed light on the relevance of the issues the BPD is charged with, as well as what they might be able to do to improve in some of these areas.

The data used in these analyses, referenced below, includes a set of **239,525 crime incidents** reported in Boston from June 2015 through November 2017, along with their respective locations. Data about FIOs in Boston was supplied from a set of **152,230 reported incidents**, which included location as well as race of the suspect stopped.

Figure 1

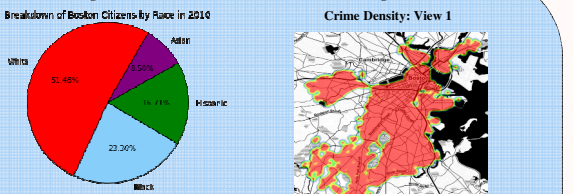


Figure 3

Crime Density: View 1

Figure 2

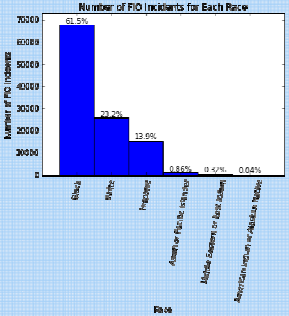
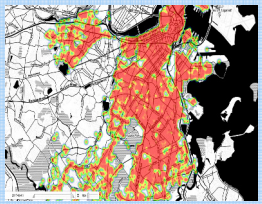


Figure 4



An initial look at the data shows discrepancies between the racial consistency of Boston and the frequency at which the BPD stops individuals of specific races. Looking at the density of crimes over Boston also shows specific hotspots for crime: specifically centered around the corner of Washington and Dudley Street, and tapering off toward the southwest.

References

"Population Estimates, July 1, 2016, (V2016)." UNITED STATES QuickFacts from the US Census Bureau. N.p., n.d. Web. 04 December 2017.
"Boston Police Department FIO | City of Boston | Open Data." *City of Boston's LEGACY Open Data Portal*. N.p., n.d. Web. 03 December 2017.
"Crime Incident Reports (August 2015 - To Date)." *Analyze Boston*, Boston Police Department, data.boston.gov/dataset/crime-incident-reports-august-2015-to-date-source-new-system.http://www.cityofboston.gov/images_documents/Redistricting_tcm3-10195.pdf

Stop and Frisk Disparity

Looking to see if the patterns surrounding crime in the area justify the policing behavior determined by the FIO reports, there were some jarring discrepancies. Relying on the 2010 census data for comparison, a disparity between the number of FIO incidents in an area and the number of crimes *in* that area is apparent. Note how the racial composition of Roxbury and the South End appear to contribute to the number of FIOs in those districts.

Figure 5

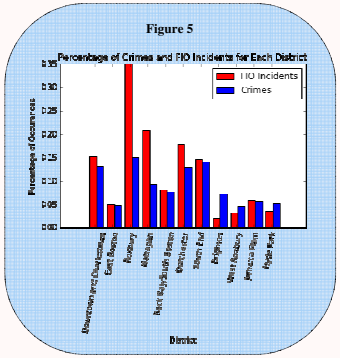


Figure 6

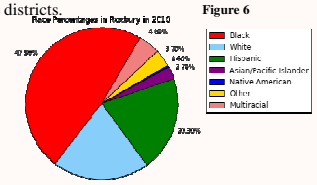
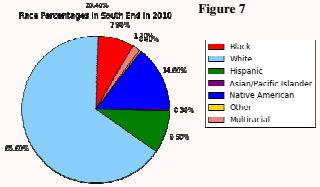


Figure 7



Comparing the percentage of crimes in Boston policing districts to the percentage of FIO incidents from 2012 to June 2015 was essential in determining whether or not the FIO incidents could be explained by the crime in the area or if other factors contributed. Despite having almost the same percentage of crimes, the South End had significantly fewer FIO incidents, and is predominately white whereas Roxbury is predominantly black.

Regression Analysis

Figure 8

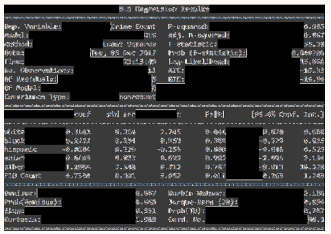
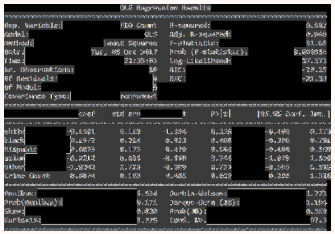


Figure 9



Hypothesis: Using linear regression, I investigated a potential correlation between the **racial composition of a neighborhood** and the **number of reported Crimes/FIOs in that neighborhood.**

Results: Figure 8 shows a definitive correlation between the number of white individuals in a district and the amount of crime in that district. While there is a similar correlation for black individuals, the confidence interval is centered around 0 for that estimate, suggesting a weak correlation. **Figure 9** shows a distinct correlation between the number of black people in a district and the number of FIOs in that district, with a negative correlation for white individuals.

The K-means Improvement

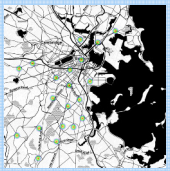
Figure 10

Figure 11

Figure 12

Figure 13

Figure 14



Clearly, the BPD is targeting individuals for stop and frisk with some bias. My priority now became improving the effectiveness of the BPD to react to crime in an efficient, unbiased way.

Running k-means++ on the Crime Incident dataset for several values of k allows for a tunable attribute that specifies exactly where the BPD should be utilizing it's resources. This could offer a strictly statistical way for the BPD to improve it's policing in an unbiased way. Let k be the number of areas where the BPD can expend resources, and this map will show you exactly where those resources should be deployed. Shown above in **Figures 10-14** are the results of k-means++ clustering for each of $k = 1$, 5, and 20.

Conclusions

- The racial disparity is most profound in districts such as Roxbury and Mattapan. While crime is generally higher in those districts, the FIOs in those districts are disproportionate to the amount of crime.
- The regression results indicated that the "whiter" the neighborhood, the less likely for there to be FIOs in that neighborhood, even though there is exists a positive correlation the number of white individuals in an area, and the amount of crime in that area.
- The BPD could improve their system by deploying officers to the locations specified by the statistical analysis generated by k-means++. In this way, policing would go where the crime goes-not necessary where minorities are.

@BUCompSci