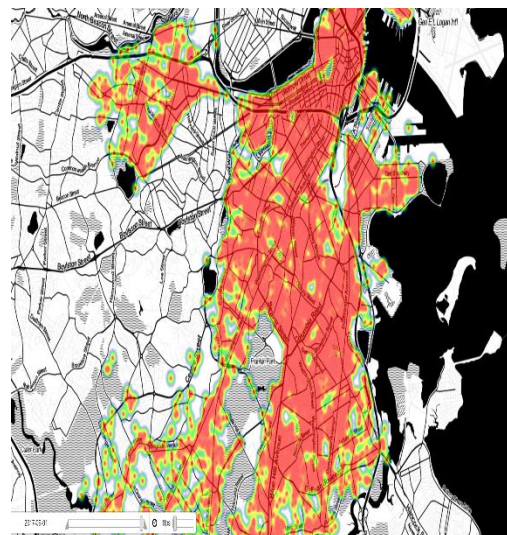


While the project I undertook for this course began as a simple overview of crime in Boston and how the Boston Police Department is combating that crime, it developed into a much more thorough investigation of the deep-seated issues facing the BPD, and the evaluation of statistical constructs that might be used to help solve some of those issues. Described below are the both my research findings and general approach to the problems raised, as well as a suggested solution to begin to combat the faults identified.

Initially, I looked to simply gather as many statistics about crime in Boston as possible. To understand the scope of what the Boston Police Department has to deal with, it was necessary to flesh out the problem (**Figure 1** gives a heatmap of the crimes investigated). To this end, online data resources were endlessly useful. While most of the datasets used are fully fleshed out in the README.md associated with the contents of the project, I can say without hesitation that the *Boston Open Data Portal* (now *Analyze Boston*) has played an important part in the collection and processing of most of the main datasets investigated. In becoming familiar with the data, it became clear that there were distinct crime hotspots located throughout Boston.

In particular, I became interested in comparing these how the BPD handled policing these hotspots with regard to their **Field Interrogation and Observation** (hereby FIO), or stop and frisk, program. Recent years have found the police department coming under fire for this program, accusing it of

Figure 1



disproportionately targeting minority citizens. In order to better arm myself with the knowledge I would need to try to improve Boston policing, I needed to determine the veracity of these claims.

This idea sparked much of the statistical analysis used in the project. At a high level, I wanted to answer whether or not the BPD did have some racial bias in who they choose to stop and frisk, and, if they did, what might be done to combat the issue. A quick, face-value, look at the breakdown of statistics showed this to be true. **Figure 2** conveys that, while Boston is a majority white city, there exists a noticeable skew toward black individuals when the BPD chooses targets for their FIOs. While this is not necessary meaningful on its own, it was certainly cause for concern, at deepened my curiosity into the subject.

As such, I decided it would be particularly helpful to examine each of Boston's policing districts on a case-by-case basis. **Figure 3**, on the next page, shows the result of this examination, and there are some jarring irregularities. Take, for instance, Roxbury and the South End. The locations have roughly same amount of crime as a percentage of the entire dataset, but Roxbury sees an excessive increase in the percentage of FIOs. When looking more intently at the racial composition of each of these neighborhoods, the striking difference between them is that Roxbury is a majority black neighborhood, while the South End is majority white. This, too, is not explicit proof of wrongdoing by the BPD, but it certainly

Figure 2

Breakdown of Boston Citizens by Race in 2010

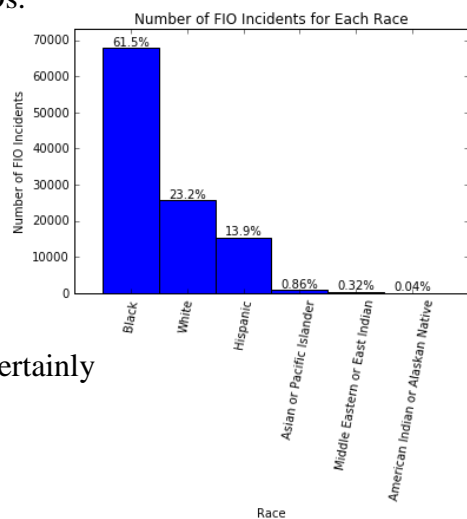
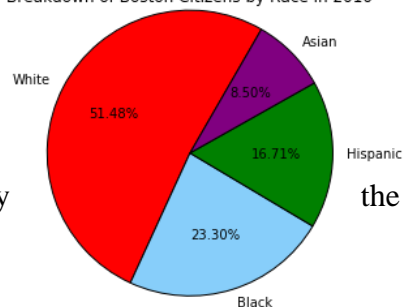
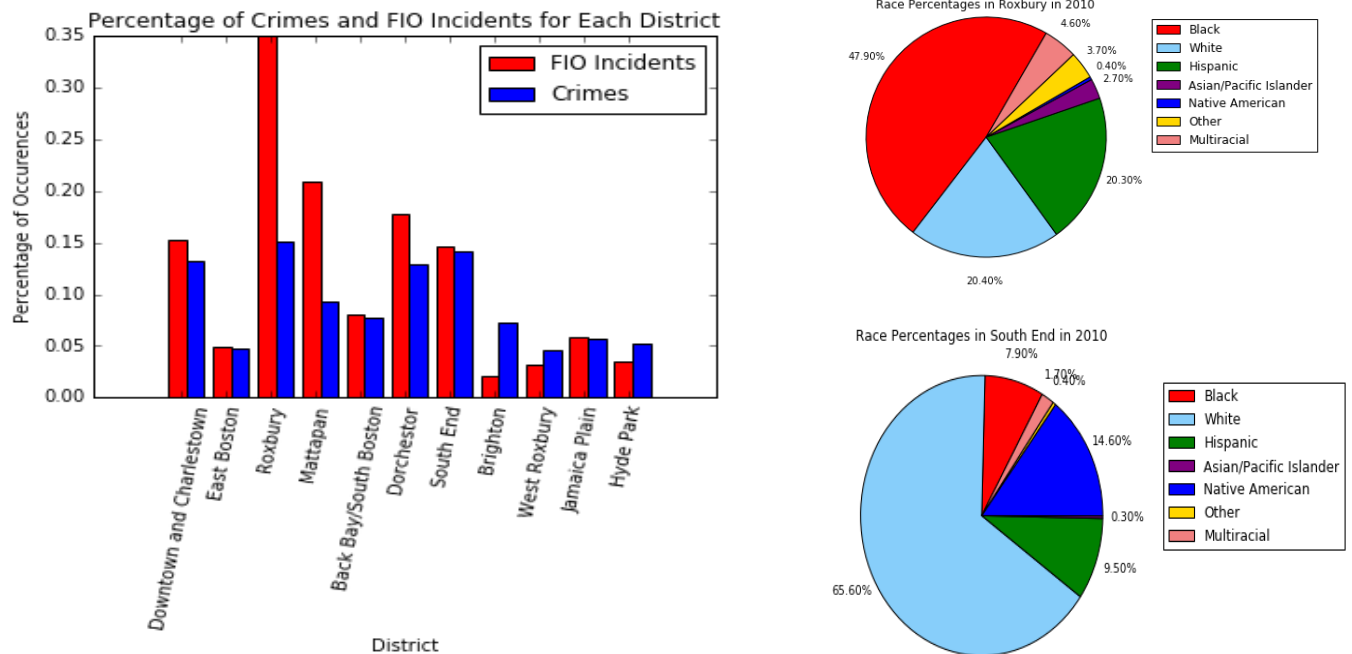


Figure 3



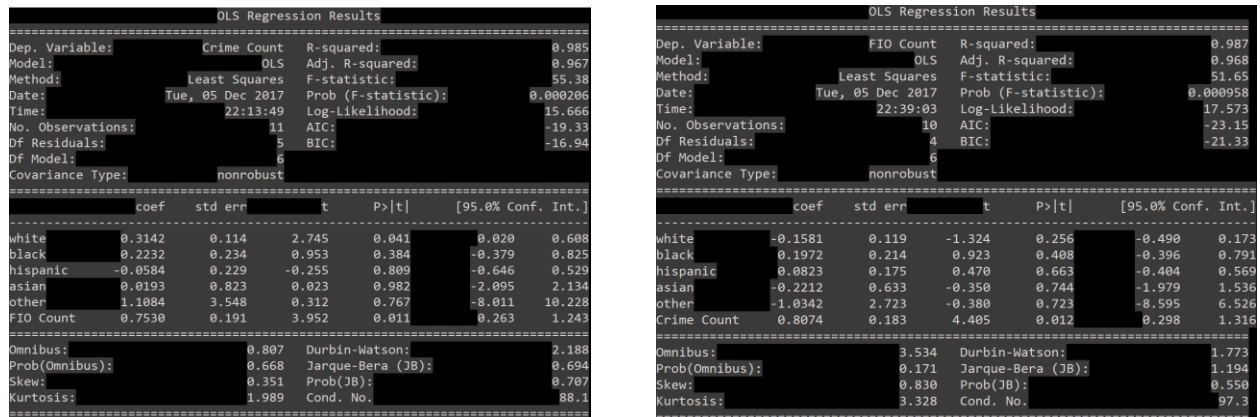
raised some new questions: namely, how much does the racial composition of a neighborhood play in the number of FIOs that take place in that neighborhood?

This is where regression analysis began to play an important role. I tried to model a regression such that the input variables would be the racial consistency of each of Boston's policing districts as a percentage of the population of that district. The dependent variables, then, could be the amount of crime in a given district, as well as the number of FIOs in that district. In this way, I could effectively examine how much each race contributes to the statistics investigated above.

The results were interesting, to say the least. **Figure 4**, on the next page, shows each regression's breakdown, with the correlations found for each race being an indicator of how much each race contributes to the statistics seen in that district. When crime was the dependent variable, there was, in fact, a positive correlation between the number of black individuals in a neighborhood and the number of crimes in that neighborhood, to be sure. However, there was an

even *higher* positive correlation between the number of white individuals in a given neighborhood and the amount of crime in that neighborhood, indicating that, in general, “whiter”

Figure 4



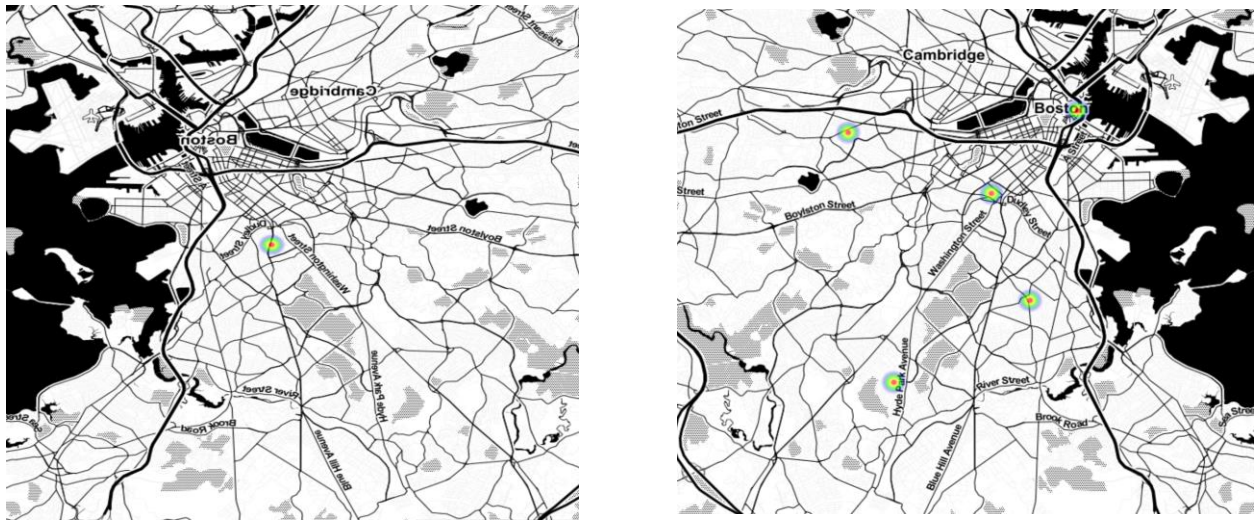
neighborhoods have slightly more crime.

When the dependent variable was the number of FIOs in a neighborhood, the results were as one might have expected, given the preliminary findings of this report: there was a negative correlation between the number of white individuals in a neighborhood and the number of stop and frisks in that neighborhood, and there was a positive correlation between the number of black individuals in a neighborhood and the number of stop and frisk in that neighborhood. Truly, it becomes apparent that the “whiter” the neighborhood, the fewer the number of FIOs occurring in that neighborhood become.

These results were not perfect, however. Because the dataset used to conduct this experiment was small, only consisting of as many points as there are policing districts in Boston, the confidence of each of these figures is relatively weak, as can be seen in the 95% Conf. Int. column of each regression in **Figure 4**. Still, the results found were enough to cement the veracity of the claims brought against the BPD over the racial bias exemplified in their FIO program.

I became determined to employ some strategy to combat this bias. Because the problem seemed to stem from a racial bias, I decided one solution would be to try to remove race from the equation entirely. **Figure 8** conveys the proposed idea; by running k-means on the crime incident reports numerous times, with k equaling everything in the range from one to twenty, I attempted to map out Boston's crime centers: the hubs in Boston around which crime is most frequent.

Figure 8



The result is a tunable map that displays exactly where the BPD needs to deploy its resources in order to maximize the effectiveness of combating crime while minimize the bias that has become to innate to their struggle. Should the BPD only have enough resources to service, say, five of these locations on a specified day, then the right map of **Figure 8** shows exactly where they should dedicate those resources to. In this way, race is taken off the table as a factor in where the Boston Police Department conducts its policing; they strictly adhere to those locations that optimize the way they can fight crime, as determined by the crime incident reports.

While this certainly isn't the definitive answer to the systemic bias that may cause such perturbances as the racial skew in the BPD's FIO program, this system allows a fair way of

utilizing the BPD's resources to the fullest. In serving Boston's districts proportionally to the amount of crime in those areas, Boston policing would go where the *crime* goes, *not* where the minorities are.