PPUA 5263 – Final Project Report Surya Menon

Crime Analysis in Boston

Introduction

For my final project, I was interested in examining the distribution of crime incidents in the city of Boston between 2015 and 2019, and learning more about the context into how crime actually manifests in the city. Specifically, I wanted to examine general spatial crime trends and patterns in the city, and furthermore whether there is a relationship (geographically and statistically, such as correlation) between crime in Boston and demographic factors such as race, education level, and income.

Examining crime patterns alongside demographics may give us a better understanding of how public safety in Boston has evolved. Analyzing crime incidents can inform us about which areas of the city are experiencing the most crime incidents, what types of crimes are most prevalent, how these patterns may have changed across years, and which populations are most vulnerable to crime incidents in the city. With more clarity on how crime is occurring in the city and which areas and populations are at highest risk, the city can develop policies that better target these issues and more effectively allocate resources to law enforcement agencies. Furthermore, by identifying potentially relevant factors (i.e. demographics, location) that contribute to crime occurrences, we could additionally create some sort of prediction model to identify areas in which crimes may occur. In general, better understanding crime distribution in the city can help in designing initiatives that improve the overall security and safety in the city.[1]

Methodology

Since our analysis is focused on the city of Boston, we first imported relevant shapefiles specific to the city. In particular, we imported a shapefile for the 2010 Census tracts for Boston, which are some of the smallest geographic units with which to examine Census data; we chose this standard geographic unit to get a more fine-grained level of analysis.[2] We additionally imported a Boston neighborhoods shapefile from the Boston Planning and Development Agency (BPDA) to help provide more context and interpretability of the Census tract results, since neighborhood divisions are more commonly understood and digestible geographic units than tracts.[3] We conducted all of our subsequent analyses on top of these base shapefiles (Figure 1).

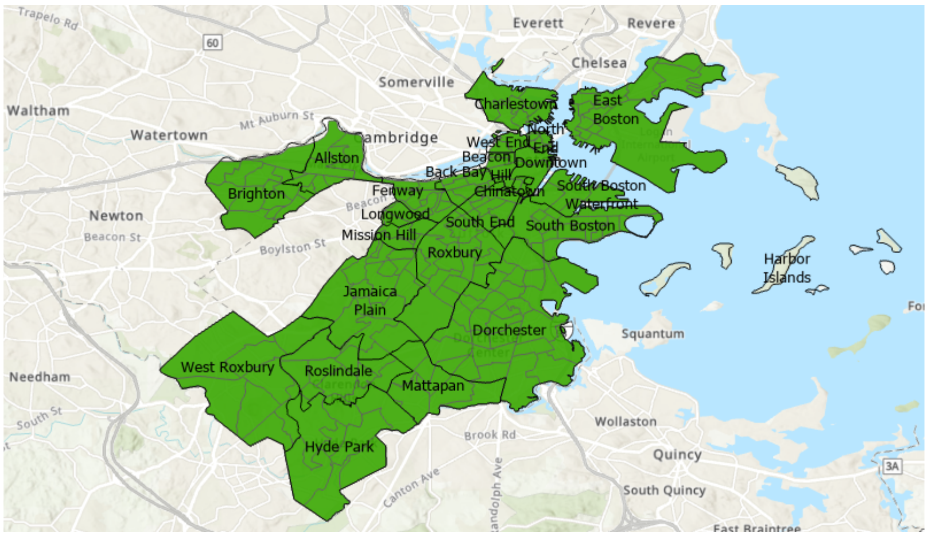

Figure 1: Boston Neighborhood and Census Tract Boundaries

To initially look at general crime trends in the city of Boston, we utilized the Crime Incidents Reports dataset, which is available on Boston's open data hub, Analyze Boston.[4] This dataset was created by the Boston Police Department (BPD) and describes preliminary details regarding incidents in which the BPD was called to respond between August 2015 and present date (as of this report, we had incidents reported until October 13, 2019). The dataset includes fields such as: incident classification, date of incident, district, offense type, and the coordinates for the location of the incident (latitude and longitude values). Because of the spatial information included in this dataset, we were able to later spatially link this dataset with other relevant data files (i.e. Census demographics).

A limitation of this dataset is that there are no recorded instances of sexual assault or rape, which are quite severe crimes that would provide more insight into an important aspect of public safety within the city, and could potentially result in interesting analyses when examining the types of crimes that occur near colleges and universities in the area. Another limit of this dataset is the lack of information about the reported incident and subsequent outcome, such as the weather, the number of people involved, or whether an arrest occurred; some of these details could contribute to some other interesting analyses, including: are certain crimes correlated with specific weather conditions, do certain parts of the city have higher arrests, etc.

The crime incidents dataset is organized in tabular format (one record per incident) with longitude and latitude coordinates for each incident, so we created a feature class of point geographic entities (using the coordinates to define the locations) with the XY Table To Point tool. We also put our data into a projected coordinate system—in this case, WGS 1984 UTM Zone 19N, Meters—in order to allow for linear units of measurement, which were necessary when running some of the spatial analysis tools.

We mapped the crime incident points across all years and since there are over 400,000 occurrences in the dataset, it is hard to visualize any sort of patterns or clustering in the dataset when just examining the points (Figure 2). We additionally added a time slider to visualize the crime incidents by year, but we did not observe any real spatial or clustering differences over time. To start trying to parse the crime data for patterns, we used the Summary Statistics tool to look at the frequency of each *Offense_Group* (i.e. larceny, motor vehicle accident, homicide, etc.). There are 67 distinct crime types for this attribute, and instead of visualizing the patterns of all of these different categories, we decided to examine the distributions of the some of the top *serious* crimes, which are specified as Part I offenses by the FBI Uniform Crime Reporting

(URC) Program, throughout the city.[5] The summary file identified the top five serious crimes as: Larceny, Larceny from Motor Vehicle, Aggravated Assault, Auto Theft, and Residential Burglary. When using unique colors symbology to look at the distribution of these five crimes, we again did not identify clusters of crimes in specific parts of the city or any significant trends over time.

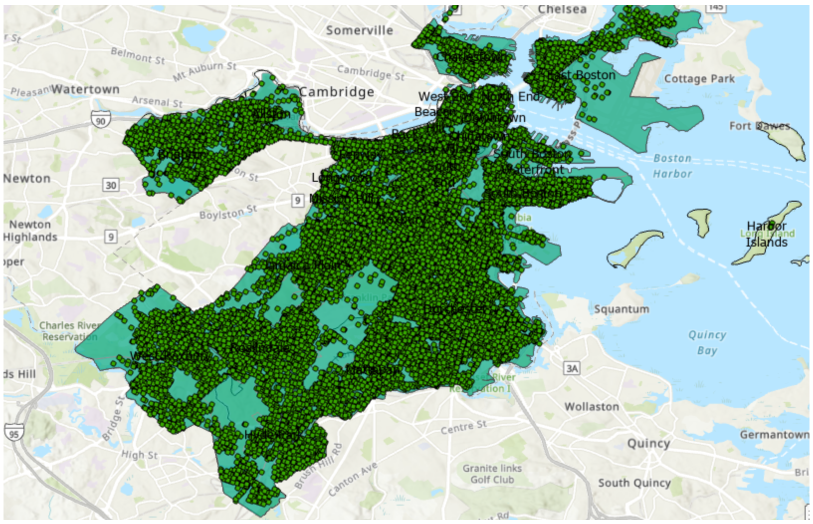

Figure 2: Boston Crime Incident Points, 2015-2019

Since looking at crime incidents by points is not very interpretable, to get a better understanding of crime distribution in the city, we implemented heat map symbology, which utilizes the kernel density method to identify relatively denser areas of crime incidents. We examined this distribution between years and over time, as well as in relation to the locations of police stations throughout the city.[6] Additionally, to further identify statistically significant hot and cold crime spots and trends in the city—including new, intensifying, and persistent clusters—we ran the Emerging Hot Spot Analysis (EHSA) tool on the crime incidents data.

Next we incorporated demographics to analyze crime distribution in Boston alongside race, income, and education level at a Census tract level. We specifically imported US Census Bureau American Community Survey (ACS) five-year summary data for Massachusetts tracts, which included social, economic, and housing indicator estimates between 2013-2017; since we are looking at smaller geographic units, we are limited to summary estimates for these measures.[7] We joined this dataset with our Boston Census tract layer to limit our demographic data to the city rather than the entire state. To effectively interpret crime incidents by Census tract, we spatially joined this demographic data with our crime data points in order to acquire crime counts per tract (i.e. join counts); we also selected incidents by the top two serious crimes (aggravated assault, larceny) and spatially joined on these subsets.

We were now able to visualize crimes and demographics on the same map (representing crimes with graduated symbols, and demographics with graduated colors) and interpret whether there are relationships between these factors. We also created some new measures in order to develop analyses that are more directly comparable when conducting analyses; instead of looking at direct counts (i.e. of crime incidents, individuals of color, etc.), we calculated the rates of individuals within the square mileage of a tract. The Census data included a field for the area of the tract in acres, which we converted to square mileage (0.0015625 square miles per 1 acre). We then created a crime per square mileage metric by dividing the join count by square mileage for each tract. We also looked at this measure for the aggravated assault and larceny subsets.

For demographic measures, we were specifically interested in whether income, education, and race in a tract was related to the crimes occurring in the area. The dataset already had a median household income measure, but we wanted to look at the rate (per square mileage) of individuals that were white (and of color) and individuals with some higher education. We created a count of minority populations in a tract (by adding populations of black, Hispanic, Asian, and multiracial individuals) as well as white populations, and a count of individuals with some education (by combining the measures of some college, bachelor's, master's, and doctorate degrees included in the ACS data), and then divided each of these counts by the square mileage of the tract. These new demographics variables were then implemented in our analyses.

Since our demographic measures are five-year estimates, we wanted to look at general measures of these indicators over the time period of the crime incidents we were exploring (2015-2019). As a result, we employed Natural Breaks (Jenks) classification when visualizing our different demographic metrics. When investigating these demographics with crime incidents to make relative comparisons, we considered examining crimes in general and across individual years and evaluated both Natural Breaks and Standard Deviation (i.e. looking at distance from the mean) classifications when interpreting our results.

We finally attempted to conduct an initial regression analysis of the dataset by running ordinary least squares (OLS) to predict crimes per square mileage with our calculated measures of race, education, and income as potential explanatory variables. This analysis was very rudimentary, and we will briefly discuss these results in the *Analysis* section.

Analysis

In looking at the density of crimes with a heat map, we see that certain parts of the city—in particular areas in downtown Boston, Roxbury, and Dorchester—generally have more dense distributions of crimes, especially compared to tracts in West Roxbury and Roslindale (Figure 3). We found that this trend was present when looking holistically across all years of data, as well as by individual year. Additionally, we observed that most police stations in the city are generally located close to or within dense crime clusters, which could indicate that law enforcement is appropriately positioned to respond to criminal activity in the area, but that the city could possibly benefit from providing additional resources and locations to the police department in order to handle these areas with higher concentrations of crime.

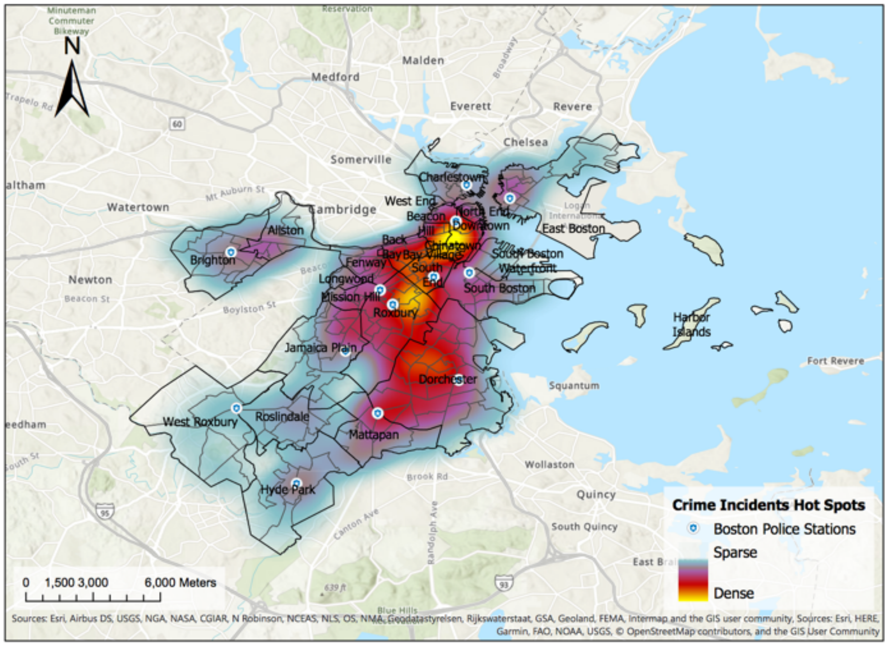

Figure 3: Heat Map of Crime Incidents in Boston, 2015-2019

When examining our EHSA results to further understand the clustering of crime incidents in the city over time, we see that there are persistent hot spots for crime in the southern parts of Dorchester and Roxbury, and a few of the smaller tracts of downtown Boston, indicating that these areas have continuously experienced significantly higher incidents of crime relative to the rest of the city (Figure 4). We also noticed some emerging hot spots beginning to border these areas with persistent hot spots, which may signal that areas with crimes are expanding to surrounding areas (and perhaps law enforcement has failed to fight and deter criminal activity in these areas). However, we also see that there are certain small clusters—particularly near northern Roxbury and Dorchester—of diminishing hot spots, which could indicate changes and development in these parts of the city (i.e. gentrification, possible influx of high-income individuals in the area). We would need demographic data across individual years to further examine this phenomenon and explore the potential causes of crime reduction in these specific tracts that border existing and growing hot spots for crime.

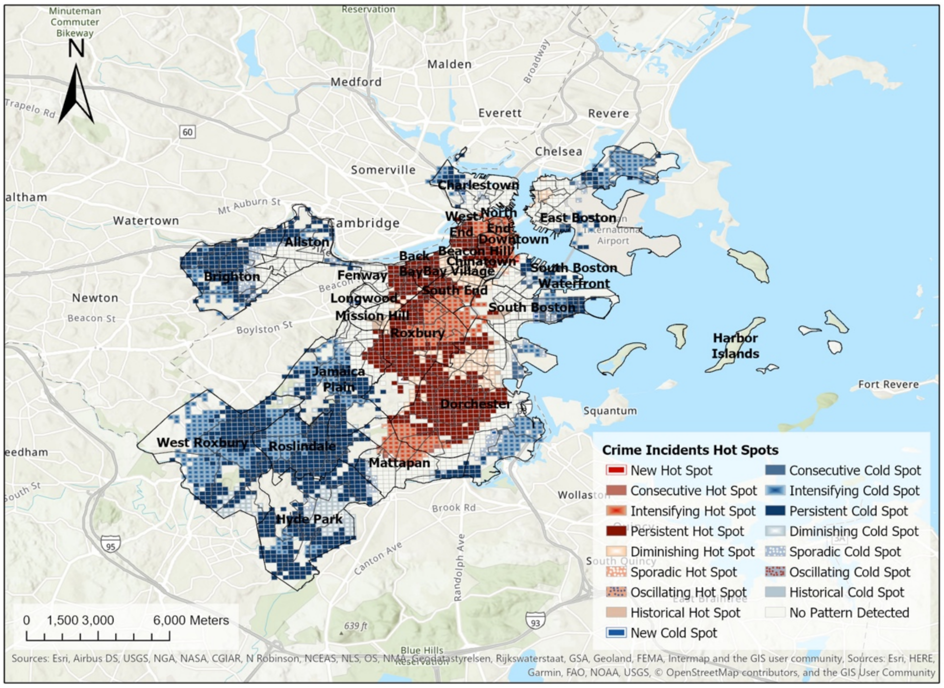

Figure 4: Emerging Hot Spot Analysis – Boston Crime Incidents, 2015-2019

We did not really find any significant differences or deviations in patterns when examining our measure of crime per square mile by individual year versus across all years; when exploring differences by year (using Standard Deviation classification), we found that the crime per square mile distribution reflected the patterns we found in our hot spot analysis (larger measures in tracts in Dorchester and Roxbury). Because there was little variation by year, we decided to look at a general measure of crime across all years and utilized the Jenks classification for the rest of our analyses.

When mapping our crime measure against various demographics, we generally saw that the distribution of crimes per square mile was positively correlated with the minority population measure per square mile; we saw that areas in the city that had higher counts of crime incidents (i.e. most of the tracts in Dorchester and Roxbury and downtown Boston, relative to the rest of the city) also tended to have larger representations of minority populations (Figure 5). This result possibly confirms some research about minority populations being disproportionally exposed to and affected by crimes compared to white individuals.[8] We also interestingly found a similar trend when spatially relating crime per square mile with the concentration of individuals with some higher education; we found increased rates of crimes in areas with larger rates of college educated individuals, which appears to be most prevalent in downtown Boston tracts (Figure 6). This result is somewhat surprising, and seems counterintuitive to what may be expected, specifically that areas with more educated individuals would generally have less crime activity.[9]

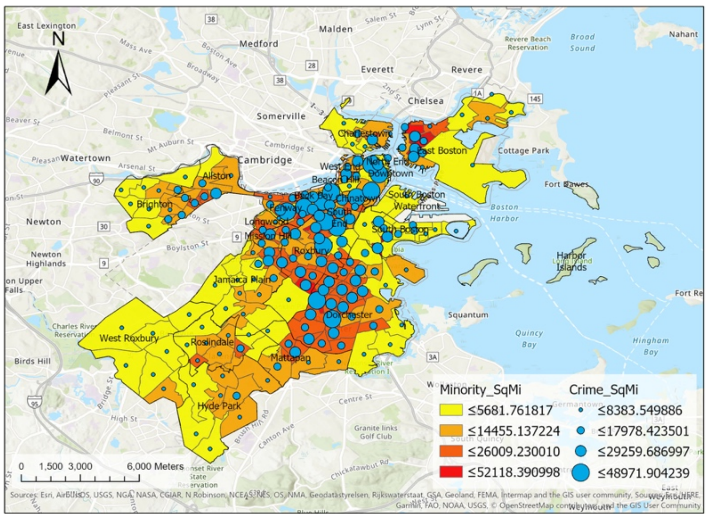

Figure 5: Crimes Incidents and Minority Population Boston Crime Incidents, 2015-2019

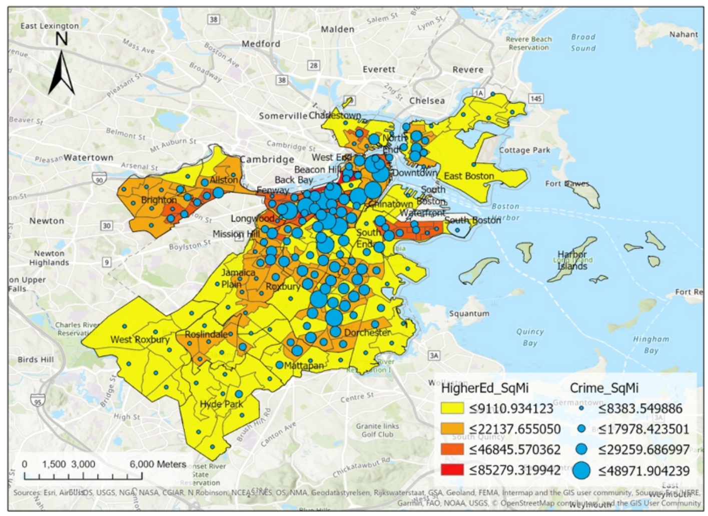

Figure 6: Crimes Incidents and Higher Education Boston Crime Incidents, 2015-2019

When looking at median household income for Boston tracts, we did not find a simple correlative relationship. We observed that overall, tracts with the highest measures of crime per square mile are either in located in areas with lower median incomes, or within tracts that border lower income areas of Boston (Figure 7). In particular, we noticed that in Dorchester, higher crime counts are clustered both in low income areas (i.e. southern Dorchester) and high-income tracts that are bordering lower income regions, such as eastern Dorchester tracts. We observed similar behavior in parts of downtown Boston, including in Chinatown (which is relatively lower in income compared to the rest of the city), and the South End tracts, which have both higher median income and crime incidents and border lower income areas, such as tracts in the northern part of Roxbury.

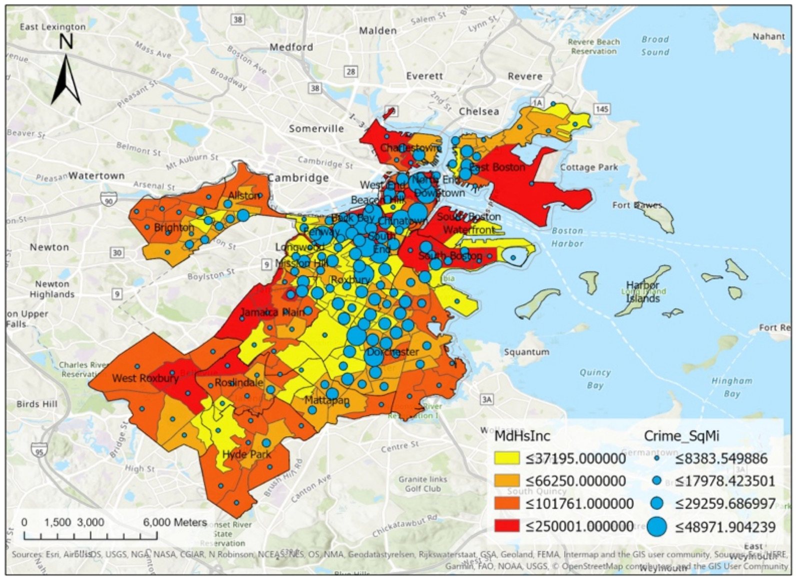

Figure 7: Crimes Incidents and Median Income – Boston Crime Incidents, 2015-2019

We also decided to look at the two most frequent serious crimes reported in the dataset, aggravated assault and larceny, since it would prudent from a public safety perspective to design policies that are focused on reducing and preventing high occurrences of very severe crimes in the city. When looking at both of these incident subsets by year, we again did not really observe any variations or trends over time, so we decided to look at the clustering across all five years in our analysis.

We found that larceny incidents per square mile seemed somewhat positively correlated with larger rates of highly educated populations across all years (Figure 8). Generally, we observed that tracts in the city with relatively more individuals with some higher education per square mile tended to also have relatively higher rates of larceny incidents. Perhaps we observe this trend because students may be targeted for this specific crime, as they are individuals pursuing higher education that also may be more vulnerable to theft and other property crimes.[10]

Further analysis into the breakdown of individuals with higher education would be needed to better understand this relationship.

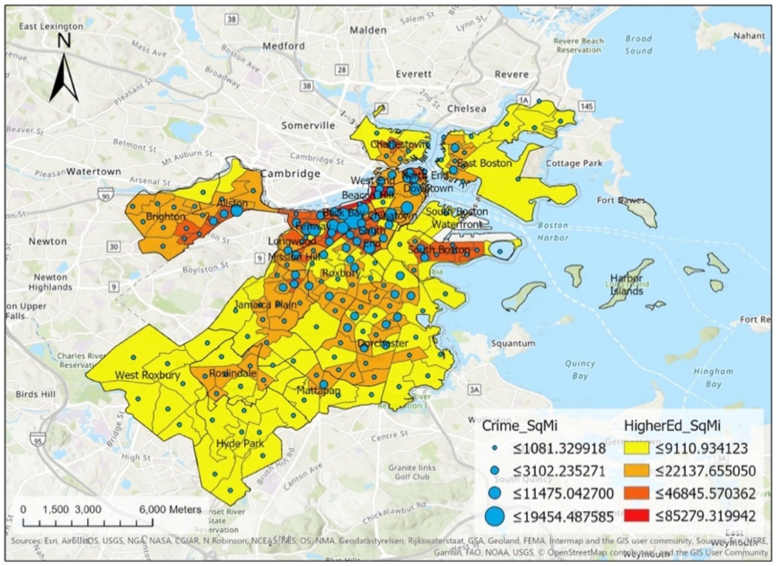

Figure 8: Larceny Incidents and Higher Education – Boston Crime Incidents, 2015-2019

We also found that incidents of aggravated assault appeared even more positively related to minority populations per square mile than what we observed when looking at our general crime count (Figure 9). Through this further analysis, we were able to better understand some of the specific crimes to which certain populations of the city are more vulnerable.

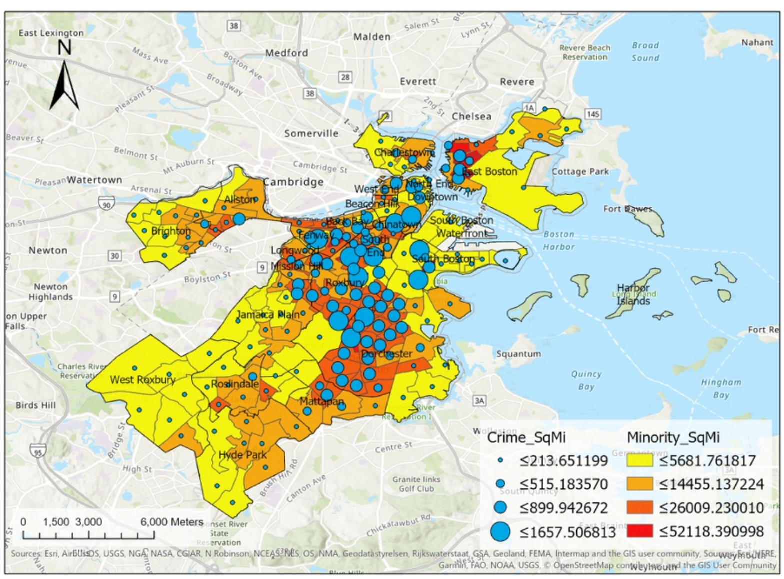

Figure 9: Aggravated Assault Incidents and Minority Population – Boston Crime Incidents, 2015-2019

Finally, when performing a simple OLS linear regression analysis to predict crimes per square mile using some demographic measures, we found that our measures of minority population and higher education per square mile were identified as statistically significant variables for the model, both with positive associations with our crime response measure. However, we also discovered that our simple model is biased due to a significant Jarque-Bera statistic, so our model may be influenced by outliers or contain some nonlinearity; as a result we may have an unbalanced distribution of residuals, as opposed to an expected normal distribution.[11] When we mapped these residuals, we observed instances of extreme underprediction by the model (greater than 2.5 standard deviations) as well as many instances of overprediction (between -1.5 and -0.5 standard deviations), so we may need additional variables in our model to better capture our crime metric (Figure 10).

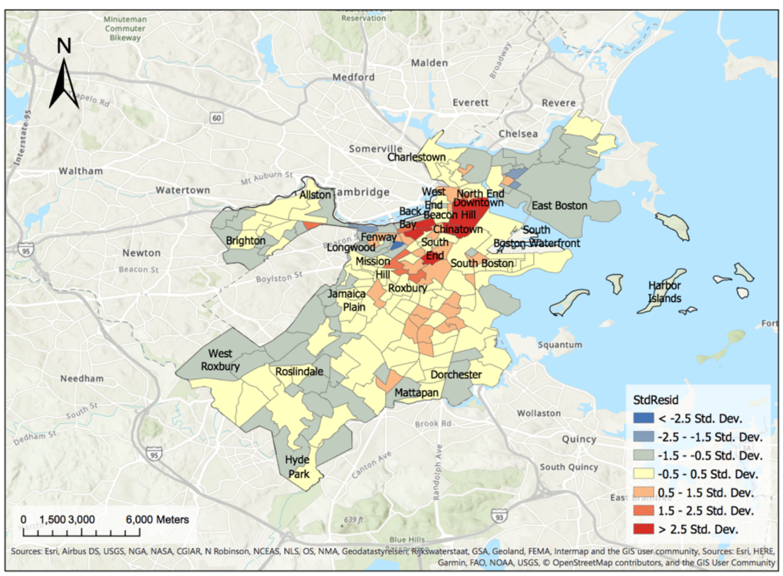

Figure 10: Ordinary Least Squares Regression Residuals – Boston Crime Incidents, 2015-2019

Lastly, we looked at the adjusted R^2 value to measure the model's predictive performance; in this case the value is only about 0.40, which indicates that we may need additional explanatory variables to better capture all of the variation of our crime rate. While our

OLS results are not able to fully model how crime is manifesting, they confirm that some of our demographic measures are relevant in understanding this phenomenon in the city and give us further confidence in our earlier analyses.

Conclusion

Based on our analyses, we found that parts of Dorchester, Roxbury, and downtown Boston had higher concentrations of crime across all five years of our crime incidents data; we additionally identified that these areas had persisting and emerging crime hot spots, indicating that crime has consistently affected communities in these neighborhoods and may be intensifying over time. While there are police stations located around these hotspots, perhaps it would be beneficial to target more resources to law enforcement agencies in these areas.

We also observed that the distribution of minority populations and individuals with higher education per square mile are generally positively correlated with the count of crime incidents per square mile of a Census tract. The relationship with the education measure seems a little strange, but perhaps increased crime in these areas could be related to an increased number of college and universities in these areas, where students may be suspectable targets to criminal activity, including theft. Furthermore, when exploring subsets of serious crime incidents, we specifically found that aggravated assaults had an even stronger relationship with non-white individuals in the area. In identifying particularly vulnerable populations to crimes in the city, perhaps we could design policies and initiatives that more effectively prevent these crimes, such as increased public safety programs on college campuses. Additionally, perhaps the city could devote more targeted programs to handle more serious crimes, such as larceny and aggravated assault, that occur more frequently in the city.

When we explored median income in Boston, we found that increased crime counts per square mile occurred in areas of lower income areas that often border higher income tracts, which could be indicative of increased crime in areas of increased income inequality.[12] Developing policies that focus on reducing inequality in the city may additionally help in alleviating increased crime statistics in these areas.

Future research could focus on exploring more temporal aspects of the dataset, including exploring more granularity in crime (i.e. looking at how crime evolves throughout a year by month, or if there are weekly patterns in crime frequency) and expanding the time frame of crime incidents explored. By looking at crimes data across an increased span of years, we could get a better understanding of how crime has evolved over time (i.e. over the decades), especially when evaluated against how demographics have changed over time; for example, it may be illuminating to discover whether gentrification or a reduced minority population trends with a reduced concentration of crime.

Additionally, further exploration into the types of crimes—and developing more general categorizations, such as violent versus property crimes—could provide even more insight into how crime is distributed in the city. Lastly, it would also be interesting to further our spatial analysis by conducting a more comprehensive OLS regression with other metrics (i.e. additional demographics and other relevant measures related to the city, such as economic or environmental factors), or running a geographically weighted regression (GWR), that also accounts for local variations in the relationships between the explanatory and predictor variables.

References –

1. How Can Crime Analysis Help Police Reduce Crime? (2014, August 19). Retrieved November 10, 2019, from https://bjatta.bja.ojp.gov/media/blog/how-can-crime-analysis-help-police-reduce-crime.
2. Census 2010 Tracts. (2019, April 4). Retrieved from https://data.boston.gov/dataset/census-2010-tracts.
3. Boston Neighborhoods. (2017, February 16). Retrieved October 10, 2019, from https://data.boston.gov/dataset/boston-neighborhoods.
4. Crime Incident Reports (August 2015 - To Date) (Source: New System). (2015, June). Retrieved October 10, 2019, from https://data.boston.gov/dataset/crime-incident-reports-august-2015-to-date-source-new-system.
5. UCR Offense Definitions. (n.d.). Retrieved October 10, 2019, from https://www.ucrdatatool.gov/offenses.cfm.
6. Boston Police Stations. (2015, March 2). Retrieved October 10, 2019, from https://data.boston.gov/dataset/boston-police-stations.
7. Basic Characteristics - ACS 2017. (n.d.). Retrieved October 10, 2019, from http://worldmap.harvard.edu/data/geonode:acs_2017_bfs.
8. Neighborhoods and Violent Crime: HUD USER. (2016). Retrieved October 10, 2019, from https://www.huduser.gov/portal/periodicals/em/summer16/highlight2.html.
9. Bell, B., Costa, R., & Machin, S. (2018, October 14). Why education reduces crime. Retrieved October 10, 2019, from https://voxeu.org/article/why-education-reduces-crime.
10. 96 Percent of Campus Crime is Property Crime. These are the Most Unsafe Colleges in America. (2019, November 12). Retrieved November 29, 2019, from https://www.inmyarea.com/research/the-most-unsafe-college-campuses.
11. What they don't tell you about regression analysis. (n.d.). Retrieved November 15, 2019, from https://desktop.arcgis.com/en/arcmap/latest/tools/spatial-statistics-toolbox/what-they-don-t-tell-you-about-regression-analysis.htm.
12. The stark relationship between income inequality and crime. (2018, June 7). Retrieved November 15, 2019, from https://www.economist.com/graphic-detail/2018/06/07/the-stark-relationship-between-income-inequality-and-crime.