Boston Shooting Incidents Analysis

Text

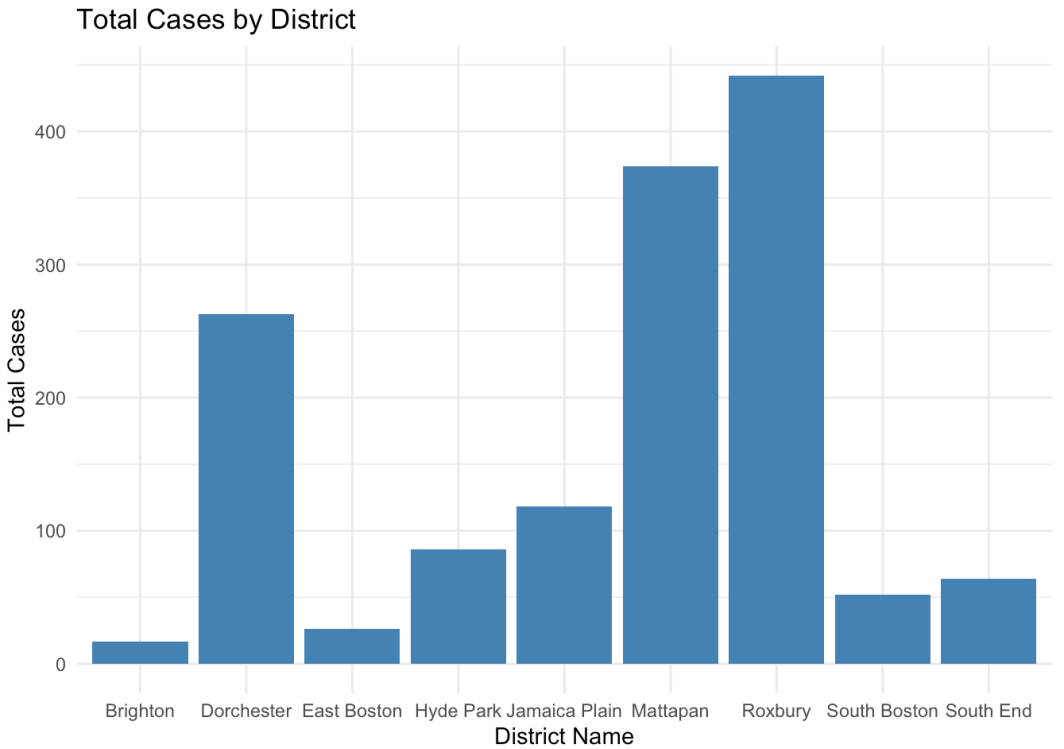
Breaking Boston's Silence: Does Socioeconomic Status Hold the Key to Gun Violence?

In Boston, an unsettling pattern emerges—gun violence casts a long shadow across its diverse neighborhoods. But what factors contribute to this grave issue? I explore shooting incidents over the last decade aims to determine if socioeconomic factors are at the heart of this issue. Could understanding these elements lead to a solution for the violence?

At the center of my analysis is a hypothesis: *improving socioeconomic conditions—specifically income, employment, education, and racial equity—can significantly reduce the prevalence of shootings across Boston's neighborhoods*. By examining data from various neighborhoods, I explore how income, employment, education and race relate to the frequency of shootings.

Visual Analysis: The Story Through Numbers

Neighborhood Impact: Boston's Safety Divide

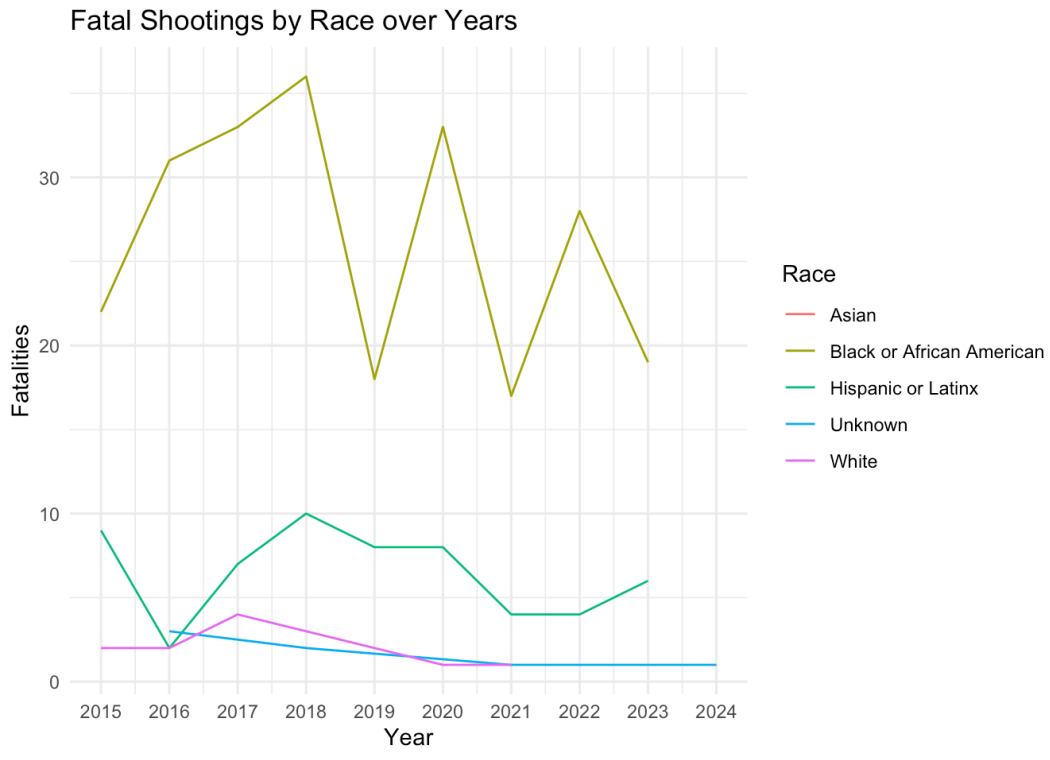


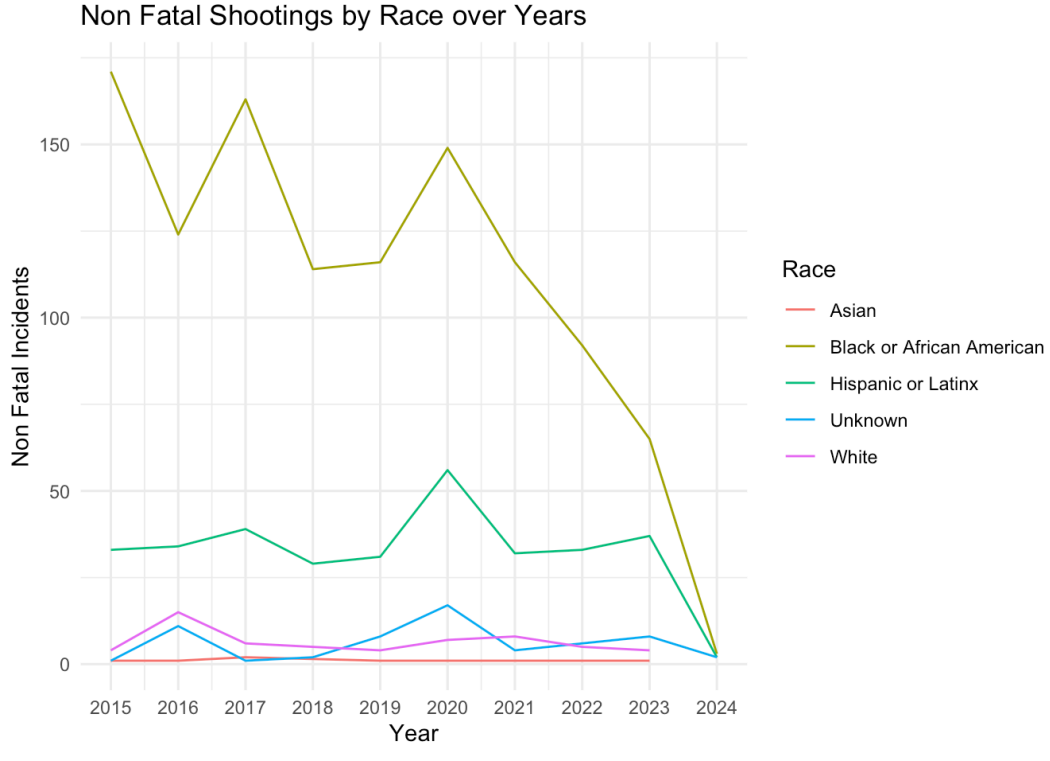
The distribution of gun violence across Boston reveals significant disparities that echo through its neighborhoods. The bar chart above showcases the total number of shooting incidents by district, highlighting a varied landscape of safety across the city. Notably, districts like Mattapan, Roxbury and Dorchester experience far higher frequencies of violence compared to quieter areas like the East Boston and Brighton. This visualization not only illustrates these differences but sets the stage for a deeper exploration of underlying causes. What can these variations tell us about the resources and challenges faced by residents of different districts?

Racial Disparities and Fatal Outcomes

The data indicates significant racial disparities in shooting incidents. Across all years and districts, the races with the highest number of shooting incidents are Black and African American, as well as Hispanic or Latinx. This is the case for both Fatal and Non-Fatal shootings.

The following graphs display the number of Fatal and Non-Fatal shootings by race every year. Although the number of incidents fluctuates a bit for each race every year, the relative number of incidents for each race compared to the others stays around the same. White, Asian, and “Unknown” consistently have an extremely low number of victims.

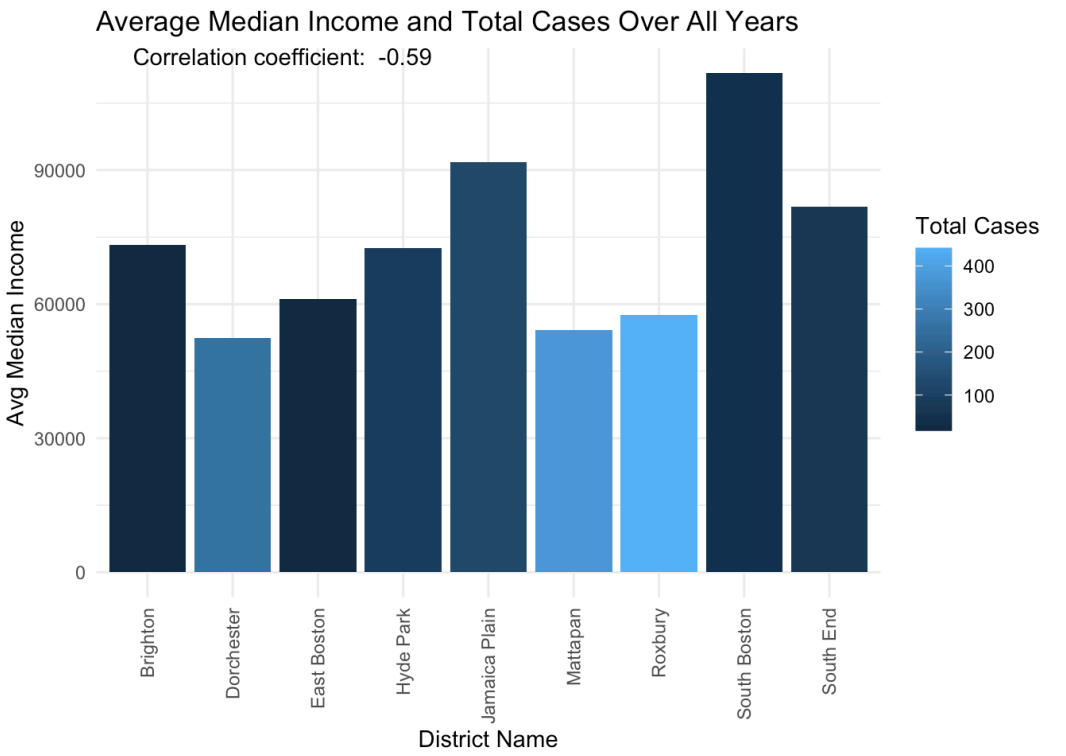




The Relationship Between Income and Safety

As I dive deeper into the socioeconomic layers of Boston, my analysis shifts to one of the most telling indicators of community safety: economic stability. In this section, I examine how the average median income across Boston's

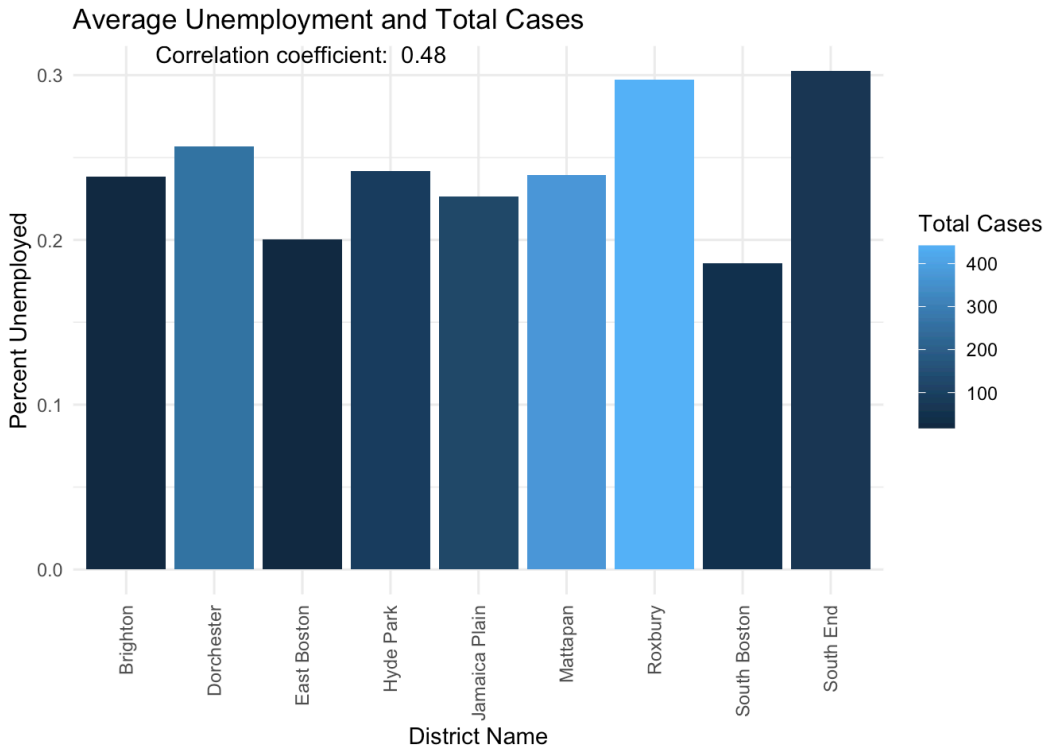
neighborhoods correlates with the number of shooting incidents they experience. The upcoming graph will visually showcase this relationship, revealing how economic conditions might directly influence the safety and well-being of residents.



Notably, neighborhoods like South Boston and the Jamaica Plain, which hold the highest income levels, also exhibit some of the lowest rates of gun violence, contrasting sharply with areas like Mattapan and Roxbury. These findings suggest that improving economic conditions in lower-income districts could be a key strategy in reducing violence.

Employment Status—A Safety Indicator?

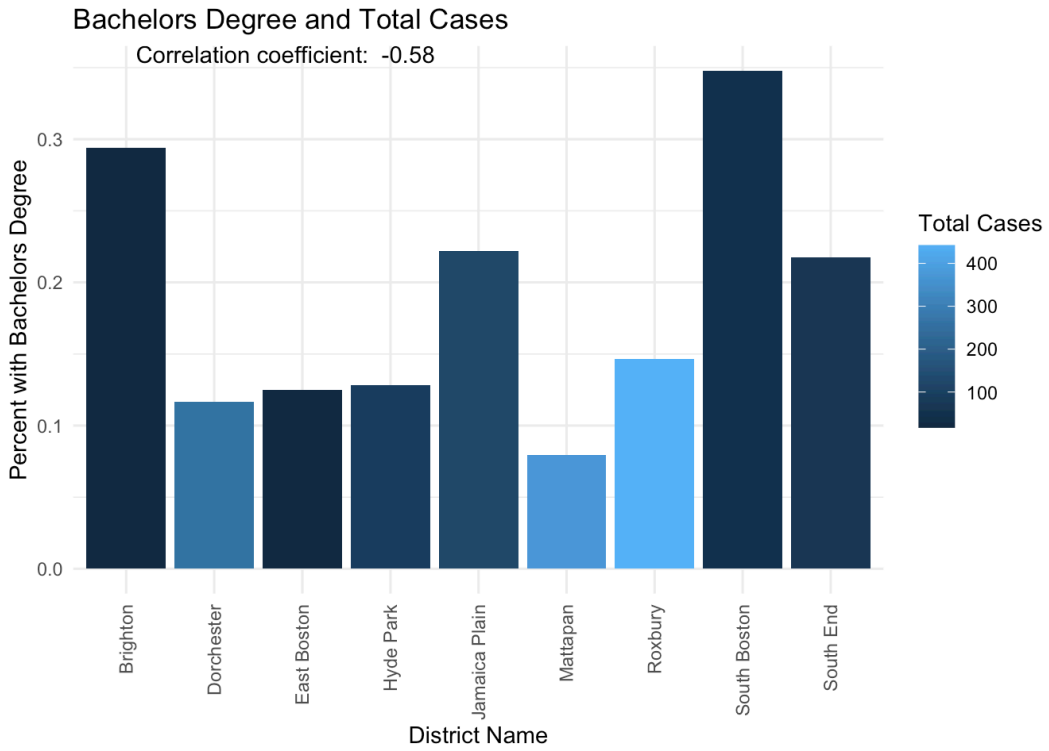
I examine another vital socioeconomic indicator that may significantly influence urban safety: unemployment rates across Boston's districts. The following graph presents the correlation between average unemployment rates and the total number of shooting incidents, offering insights into how joblessness might increase the risks of violence. The data suggests a substantial link between higher unemployment and increased shooting incidents, hinting that employment stability could be a pivotal factor in enhancing community safety.



The graph depicts a clear trend where districts with higher unemployment rates, such as Mattapan and Roxbury, also report a greater number of shooting incidents. This visualization highlights the potential role of economic distress, manifesting as unemployment, in contributing to social instability and violence. Particularly striking is the high unemployment and corresponding number of cases in Roxbury, underscoring the urgency for targeted economic and employment support in such areas. On the other hand, districts like South and East Boston, which feature relatively lower unemployment rates, witness fewer shootings, supporting the theory that job creation and economic vitality might serve as effective deterrents to violence.

Educational Attainment and Community Safety

The graph below illustrates the percentage of the population with bachelor's degrees by district alongside the total number of shooting cases. It is evident that districts like Brighton and South Boston, which boast higher percentages of degree holders, report significantly fewer shootings compared to areas like Mattapan and Dorchester, where fewer residents have completed higher education. This trend is visually represented and suggests that enhancing educational opportunities could serve as a preventive measure against urban violence.



Data Source

The primary data set comes from data.boston.gov and contains data regarding shootings in the Boston area. Data was collected by the Boston Regional Intelligence Center under the Boston Police Department Bureau of Intelligence and Analysis. The data is from 2015 forward.

Dataset Overview

The dataset consists of 1925 rows and 8 columns, detailing both fatal and non-fatal shooting incidents. Here's a breakdown of the unchanged variables from the original data set:

Variable	Type	Description
incident_num	integer	Unique identifier for each shooting incident.
shooting_date	date	Date when the shooting occurred.
district	string	Police district where the incident was reported
shooting_type_v2	string	Specifies if the shooting was fatal or non-fatal.
victim_gender	string	Gender of the victim involved.
victim_race	string	Race of the victim.
victim_ethnicity_NIBRS	string	Ethnicity of the victim according to NIBRS standards.

multi_victim	boolean	Indicates if multiple victims were involved.

Data Cleaning and Processing

The data cleaning and initial processing are documented in the attached R script, load_and_clean_data.R. Key steps included:

- Removing entries with unknown or missing critical information to ensure data quality.
- Reformatting the `victim_ethnicity_NIBRS` into a boolean `v_hispanic_or_latinx` column for clearer demographic analysis.
- Simplifying the `shooting_type_v2` into a boolean `fatal` column to distinctly categorize incidents.
- Standardizing district names through a mapping dictionary to align with geographical descriptors used in public records.

Integration of Additional Data Sources

I used the tidycensus package to add census data from the American Community Survey to the original data set. I consider key indicators such as employment status, unemployment rate, median income, percent of single-parent households, and percentage of homes that are owner occupied. I also pulled demographic data and geolocation data to help analyze the distribution of incidents across races between Boston districts.

```
 [1] "district_name"
 [2] "year"
 [3] "incident_num"
 [4] "Date"
 [5] "Time"
 [6] "district"
 [7] "victim_gender"
 [8] "victim_race"
 [9] "multi_victim"
[10] "v_hispanic_or_latinx"
[11] "fatal"
[12] "Total."
[13] "GEOID_ZCTA5_20"
[14] "NAME.x"
[15] "total_house_units"
```

```
[16] "non_married_household"
[17] "bachelors_25"
[18] "one_parent"
[19] "married_households"
[20] "income_deficit"
[21] "household_medincome"
[22] "not_enrolled_school"
[23] "not_in_labor_force"
[24] "pct_below_poverty_level"
[25] "vacancy_status"
[26] "total_pop"
[27] "White_alone"
[28] "Black_or_African_American_alone"
[29] "American_Indian_and_Alaska_Native_alone"
[30] "Asian_alone"
[31] "Native_Hawaiian_and_Other_Pacific_Islander_alone"
[32] "aggregate_time_to_work"
[33] "GEOID_TRACT_20"
[34] "NAME.y"
[35] "geometry"
```

Analysis Introduction

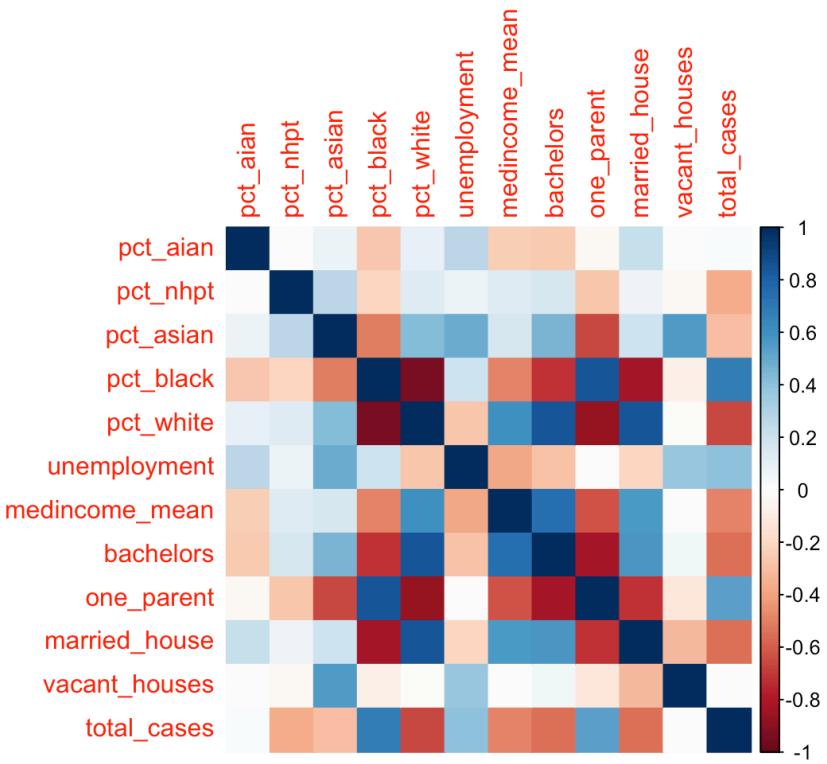
I hypothesize that higher socioeconomic status, indicated by increased income levels, stable employment, and higher educational attainment, is associated with reduced instances of gun violence. To test this hypothesis, I employ statistical models, including several OLS models and a geospatial weighted regression to examine the relationships between spatial, socioeconomic, and demographic data. This approach will assess the strength and significance of these factors in relation to urban violence, providing a data-driven basis for recommendations and policy considerations.

Analysis

Descriptive Statistics: Correlation of Relevant Factors

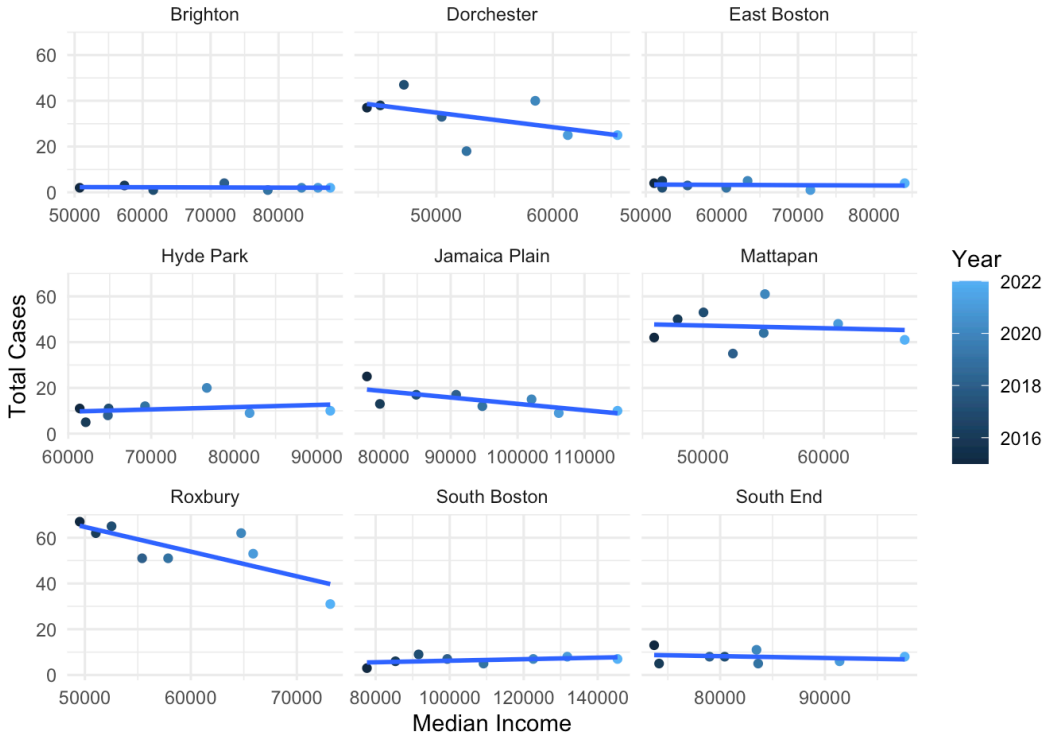
To understand the underlying relationship between the independent and dependent variables, I calculated the correlation matrix to display the direction

and magnitude of the relationships present in the data. As shown by the bottom row of the table, the dependent variable total cases is positively correlated with unemployment, the proportion of single parent households in a given district, and percentage of the population that is Black or African American. For interpretation, it is important to note that this does not imply a causal effect, and the following analysis will explain why certain demographics are more correlated with total incidents by further exploring the indivdiual impact of each variable across each of Boston's districts.

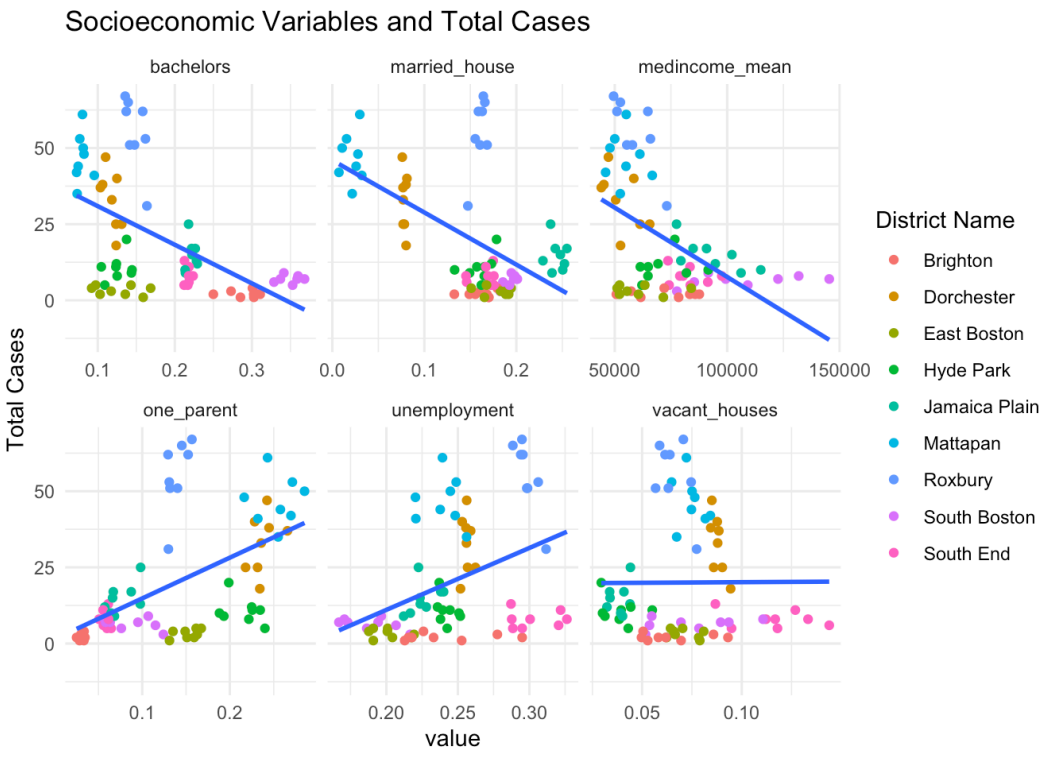


Exploratory Analysis: Relationships Between Variables

I plot median income against the total number of cases to assess whether the direction and magnitude of correlation over all districts is still relevant. Considering Dorchester and Roxbury districts are where a significant proportion of incidents are reported, I see that as median income increases, total number of cases decreases indicating employment stability (or potentially higher pay) may play a significant role in hindering gun violence. Similarly in Mattapan there appears to be a slightly negative effect indicating the importance of median income in the top three districts for the total number of cases.



Before modeling, I would like to assess the linearity of each of the dependent variables with total cases. Given the individual scatter plots, it appears that median income and bachelors degree have increasing marginal effects, so I incorporate a polynomial term for each of these variables in the non linear regression model below. As part of the summary statistics, it is important to consider the following coefficients for each trend line; bachelor's degree, number of married households, and median income displayed negative effects while single parent households, unemployment, and vacant houses displayed positive effects.



Model Predictions

The regression summaries output the results for two distinct models: a log regression that includes all of the variables for district, race and socioeconomic factors, and a non-linear regression with socioeconomic and race regressors. I performed variable selection utilizing AIC stopping criterion for both models to limit the potential for collinearity. When controlling for all factors, the percentage of white and black population were chosen in the variable selection process further demonstrating the importance of considering the segregation of Boston neighborhoods shown below. For the non linear regression, all variables selected are significance at the 5% level.

Log Regression Results

	Dependent variable:
	log(total_cases)
pct_white	-8.629** (3.569)
district_nameDorchester	0.320 (1.237)
district_nameEast Boston	-0.988* (0.569)
district_nameHyde Park	-1.438 (1.290)

district_nameJamaica Plain	0.237 (0.708)
district_nameMattapan	0.178 (1.619)
district_nameRoxbury	1.043 (0.957)
district_nameSouth Boston	1.881*** (0.435)
district_nameSouth End	0.517 (0.395)
medincome_mean	-0.00001** (0.00001)
married_house	18.466** (7.089)
Constant	4.496** (2.007)
Observations	72
R^2	0.921
Adjusted R^2	0.906
Residual Std. Error	0.358 (df = 60)
F Statistic	63.349*** (df = 11; 60)
Note:	*p<0.1;* ***p<0.05;*** p<0.01

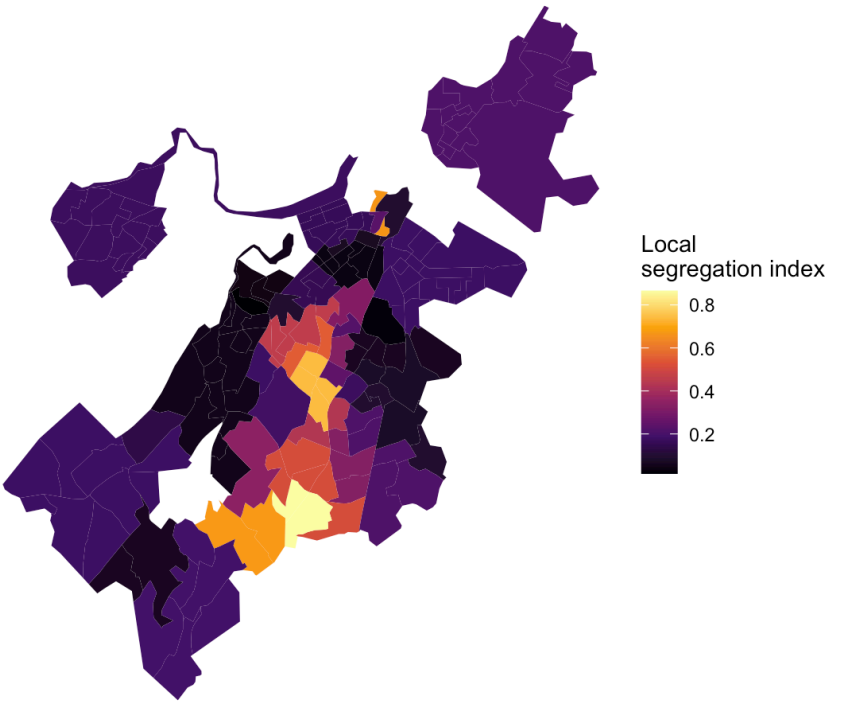
Non-Linear Regression Results

	Dependent variable:
	total_cases
unemployment	343.088*** (63.809)
medincome_mean	-0.001** (0.001)
I(medincome_mean2)	0.000** (0.000)
bachelors	-549.914*** (159.894)
I(bachelors2)	1,129.142*** (366.320)
vacant_houses	-234.031*** (76.868)
Constant	55.624** (21.757)
Observations	72
R^2	0.568
Adjusted R^2	0.528
Residual Std. Error	13.585 (df = 65)
F Statistic	14.246*** (df = 6; 65)
Note:	*p<0.1;* ***p<0.05;*** p<0.01

Geospatial Analysis

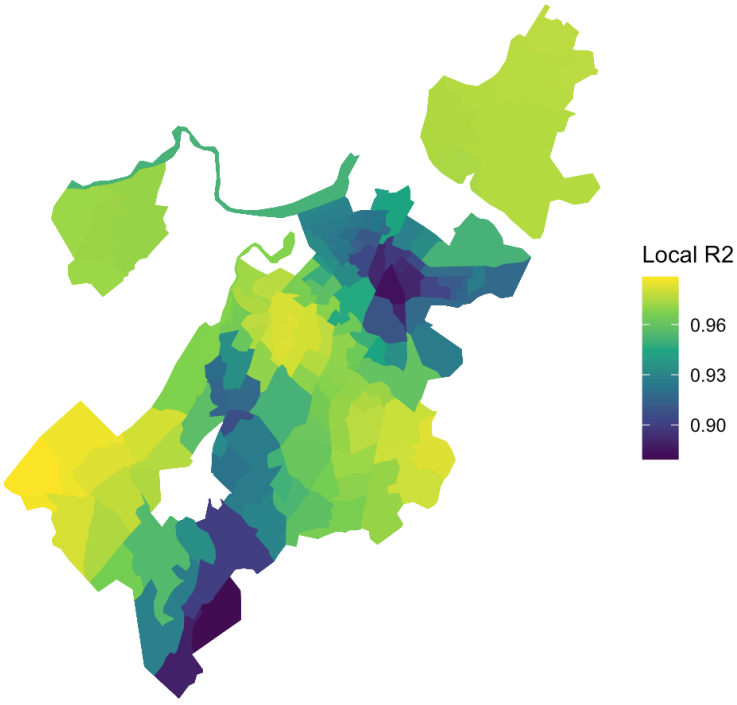
The segregation map shows the local segregation index of different areas in

Boston. The higher an area's index is (from 0 to 1), the less racially diverse it is. Notably, the most segregated areas are Roxbury and Mattapan, the top two districts for total number of cases.

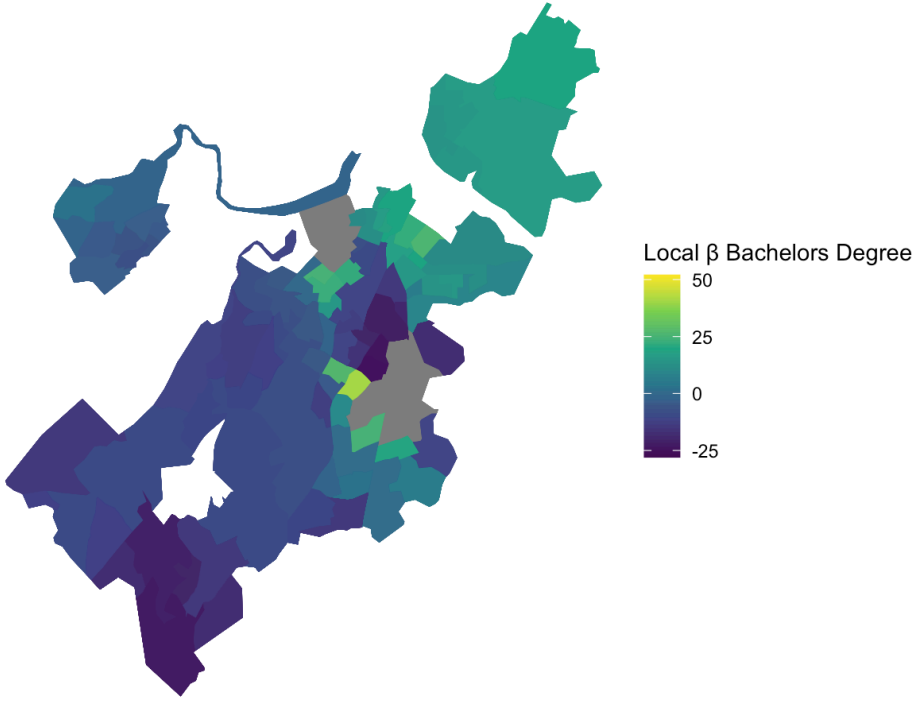


As shown by the descriptive statistics and regression results, the data is highly correlated with location, therefore the regressors cannot be treated as randomly assigned and they violate the independence assumption for OLS. To overcome this limitation, I perform a geospatial robust weighted regression from the GWmodel R package which computes a local regression for each location, while including a distance decay function, or a weight added to the sum of linear combinations, which specifies how observations outside the current location will be incorporated. A kernel bandwidth is implemented to determine the cutoff distance for observations included. The local parameters returned by the model are plotted below.

This map shows how well the model performs in predicting cases among the Boston districts. Across the map, the score was close to 1, indicating around 90% of the variance in our dependent variable was captured by the model.



This map is another exemplification that as more bachelor's degrees are attained, the number of cases decreases. This is especially true in areas that have a higher segregation index, so I infer that higher level educational attainment has a significant effect on the number of cases. That is, as educational attainment increases, total cases decrease as displayed by the heatmap of the models parameter for education.


Local β Bachelors Degree
50
25
0
-25