Crime in Boston

COURSERA CAPSTONE PROJECT

Zoe Zhang

Introduction

This final project explores if a specific neighborhood in Boston is safe and if more Police Station is associated with lower crime incidents, by looking at a city's Crime Incident open data, and the Police Station venue data from Foursquare.

The historical data set has a time and space dimension for different types of crimes in the city. Analysis and visualization of the crime types and their trends can help us to understand the crime dynamics in the city.

Problem Statement

Predict the probability of some rare event (violent crimes in this case) happening on a specific date with a spatial feature, to inform the police force can use this to concentrate on specific parts of the city each day to prevent violent crimes or at least increase the chance of patrolling there.

1. Is there seasonality when crimes are more likely to occur in a year? Has the crime incidents been growing or trending down overall?
2. What is / are the safest neighborhoods in Boston with the lowest Crime incidence rate?
3. Which neighborhoods are most safe in terms of how dense Police Stations are placed in a given Neighborhood?

Notes: Normalization by unit area method was adopted to address the square miles variance across all Neighborhoods. Limitation may exist by adopting a geographic area unit normalization. For instance, it does not take into account the population in a

neighborhood or covered by a specific Police Station, or number of police officers in a Police Station.

Methodology

Data Collection

The crime incident open data in Boston neighborhoods is studied to answer the above questions. Crime incident reports are provided by Boston Police Department (BPD) to document the initial details surrounding an incident to which BPD officers respond. This is a dataset containing records from the new crime incident report system, which includes a reduced set of fields focused on capturing the type of incident as well as when and where it occurred, specifically, these elements are included:

- Incident number: Unique identifier for the incident
- Offense Code Group: Incidents are grouped with crime types
- Reporting Area: The code for the reporting areas or zones.
- Occurred On Date: Date of the incident
- Lat: Latitude of the crime location
- Long: Longitude of the crime location

This csv dataset can be obtained here: Crime Incident Reports (August 2015 - To Date) [Source - New System]

Boston City data containing neighborhood boundaries will be obtained from the data source: Boston Neighborhoods

Additional data related to locations and police department will be obtained via the FourSquare API utilized via the Request library in Python.

Data Preparation

- Data will be sorted based on crime incidents.

- Check for NA, and drop rows where there is no value.
- Add Neighborhood field to crime incidents dataframe, based on if the set of geocodes of a crime incident location is in the Boston Neighborhood geojson data. The former is Point, while the Boston Neighborhood geojson file uses Multipolygon type.
- There will be three merges of dataframes during the analysis, which may appear at different stages.
- Aggregate historical crime data for each neighborhood.
- Add calculated fields (normalization) to dataframes.

Data Exploration

- Explore the trends of crime occurence and if there is any pattern / seasonality.
- Explore the geographic distribution, i.e. which neighborhood has the highest crime incidents.
- Explore the Police Station number, density and geolocation for each neighborhood.
- The data will be visually assessed using graphing from various Python libraries.

Correlation

- Explore the association between numbers of Police Station and crime incidents. Does more Police Stations necessarily make a neighborhood safer?
- Explore the association between Police Station density (per square mile) and the crime incidents per square mile

Discussion and Result

Crime Incidents Trend and Seasonality

The year-trend chart shows the actual count of crime incidents by year. Year 2015 and 2020 crime incident numbers were annualized because the open data does not provide a full year worth of data.

Annual Trend: As we could see from the Year Trend table 1 and Figure 1, 2017 saw the largest number of annual crime incidents during the past 5 years and 2020, and since then crime incidents have been consistently trending down.

Seasonality: From Figure 2 & 3, it seems that summer months May through August tend to observe a larger number of crime incidents, while winter months November through February tend to see a relatively smaller number of crimes. Summer season in 2017 in particular, the crime incident rates were the highest amongst all months. *Note May 2017 does not seem to be a full month, though no standardization was performed.*

Table 1. Crime Incidents Trend by Year

	2015	2016	2017	2018	2019	2020
Count	53,597	99,430	103,380	98,888	98,073	42,694
Annualized	98,948	99,430	103,380	98,888	98,073	85,388
% Change		*0.5%*	*4.0%*	*-4.3%*	*-0.8%*	*-12.9%*

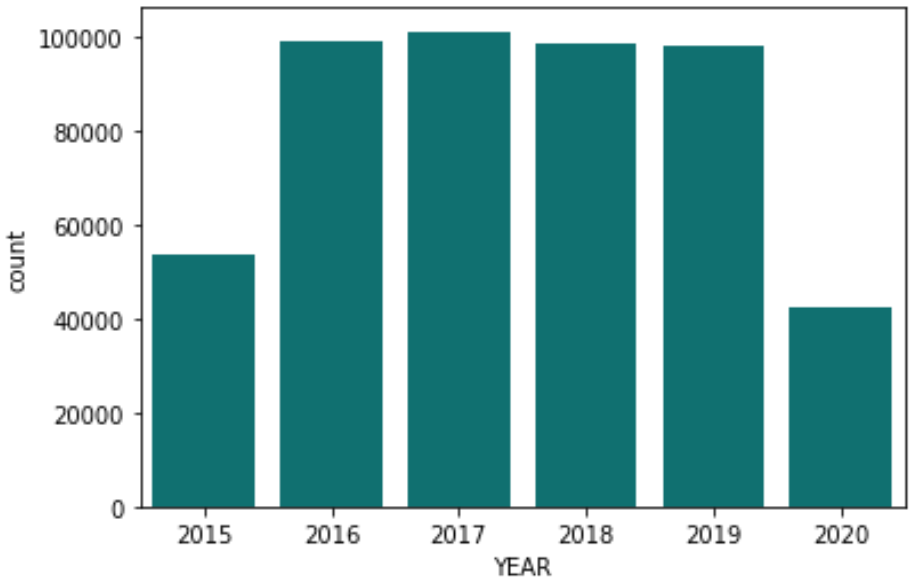

Figure 1. Crime Incidents Trend by Year

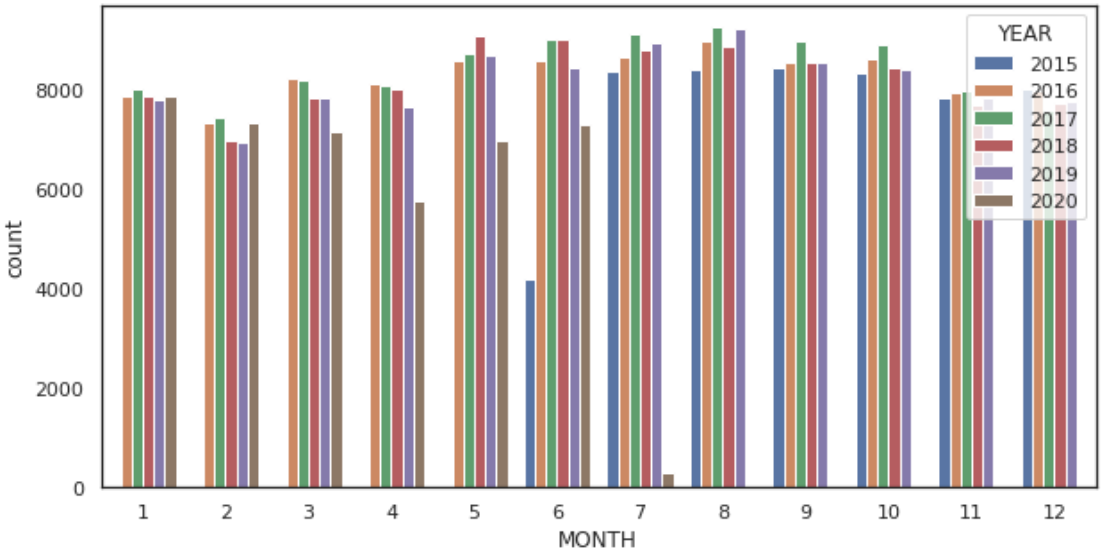

Figure 2. Crime Incidents Trend by Month/Year

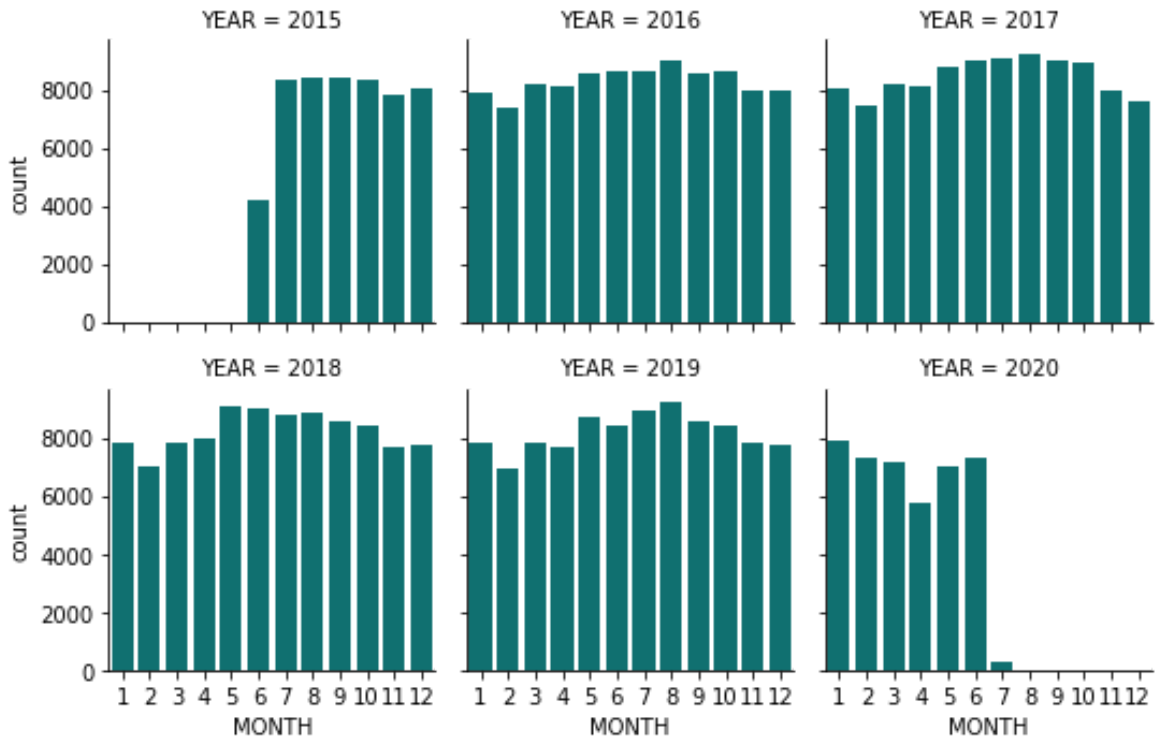

Figure 3. Crime Incidents Trend by Month in Separate Subplots

Most Common Crime Group in Boston

According to the RMS Crime Incident Field Explanation by Boston Police Department / Analyze Boston, the Crime Code Group is an internal category of Crime Description, while Crime Description is the primary descriptor of incident. There are 282 Crime Incident Descriptions, and categorized into 67 Crime Code Groups. Crime Code Group makes better sense of giving the crime dynamic (Figure 4 & 5).

During the data exploration, I noticed that one single incident may be tagged with more than one Crime Description, and/or Crime Code Grouping. The sum of the counts grouped by either fields may exceed the distinct count of incidents (i.e. incident number, or another unique identifier). Also there is no Crime Code Group in 2020.

With that, Figure 4 below shows the top 10 most common Crime Offense Groups in Boston. Motor vehicle accidents, larceny, medical assistance (needed or dispatched), investigate person, other, drug violation, simple assault, vandalism, verbal dispute, and

investigate property are the most common types of incidents that occurred in Boston during the past five years, 2015 - 2019.

There was no significant variance observed during the past five years (table or charts not shown), though the order of top 10 Crime Groups may have shifted slightly. In that regard, these fields are simply explored as a component of data description, but not followed along further in this project.

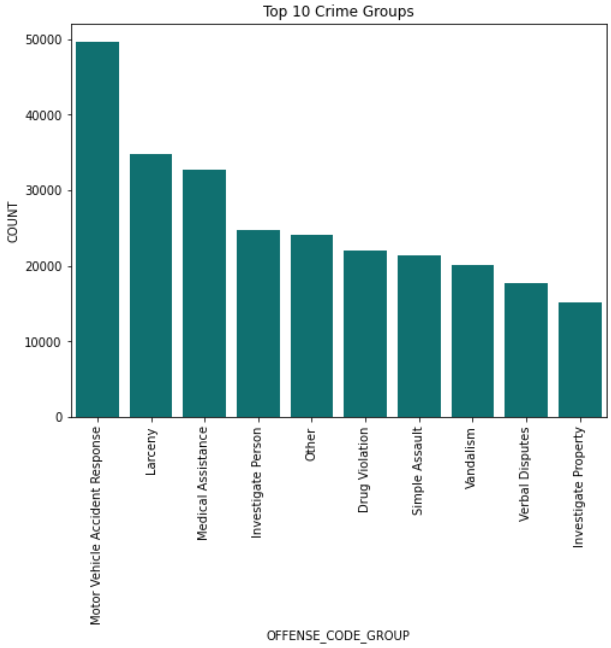

Figure 4. Top 10 Most Common Crime Groups in Boston during 2015-2019

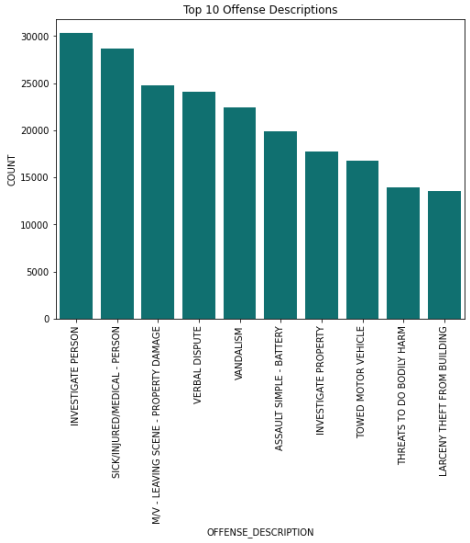

Figure 5. Top 10 Most Common Incidents in Boston during 2015-2020

Crime Incidents Distribution by Neighborhoods

Neighborhood field was added to the crime incidents dataframe, based on if the set of geocodes of a crime incident location is in the Boston Neighborhood geojson data.

Figure 6 illustrates the crime incidents in Boston. Intuitively, the darker the color, the higher the incident rate, and the less “safe” the neighborhood.

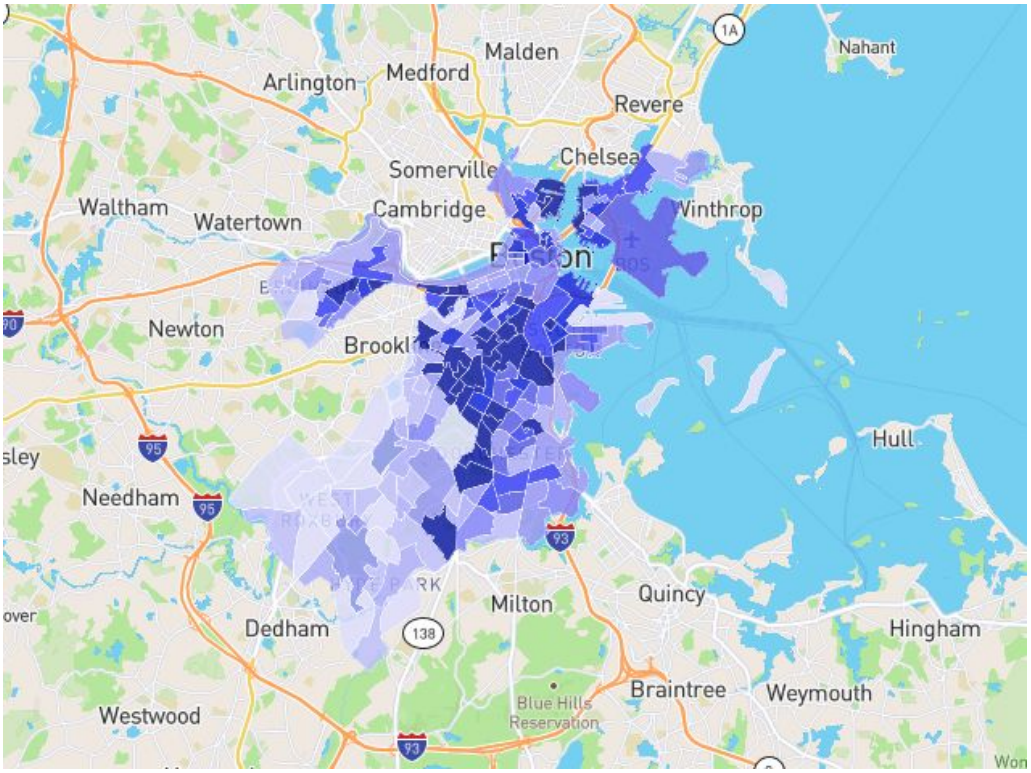

Figure 6. Crime Incidents in Boston Across Neighborhoods

Figure 7 is a map of Boston Neighborhoods. Figure 8 is a map of Boston Police Stations. Both maps have popup labels showing Neighborhood name and Police Station name.

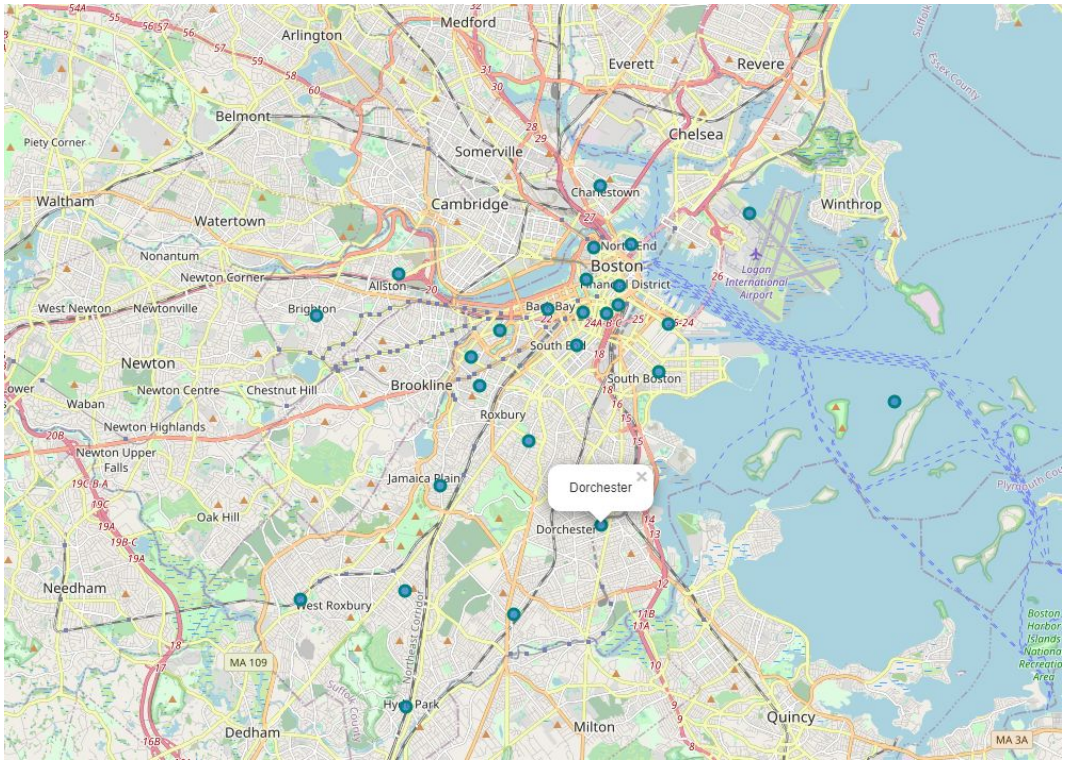

Figure 7. Map of Boston Neighborhoods

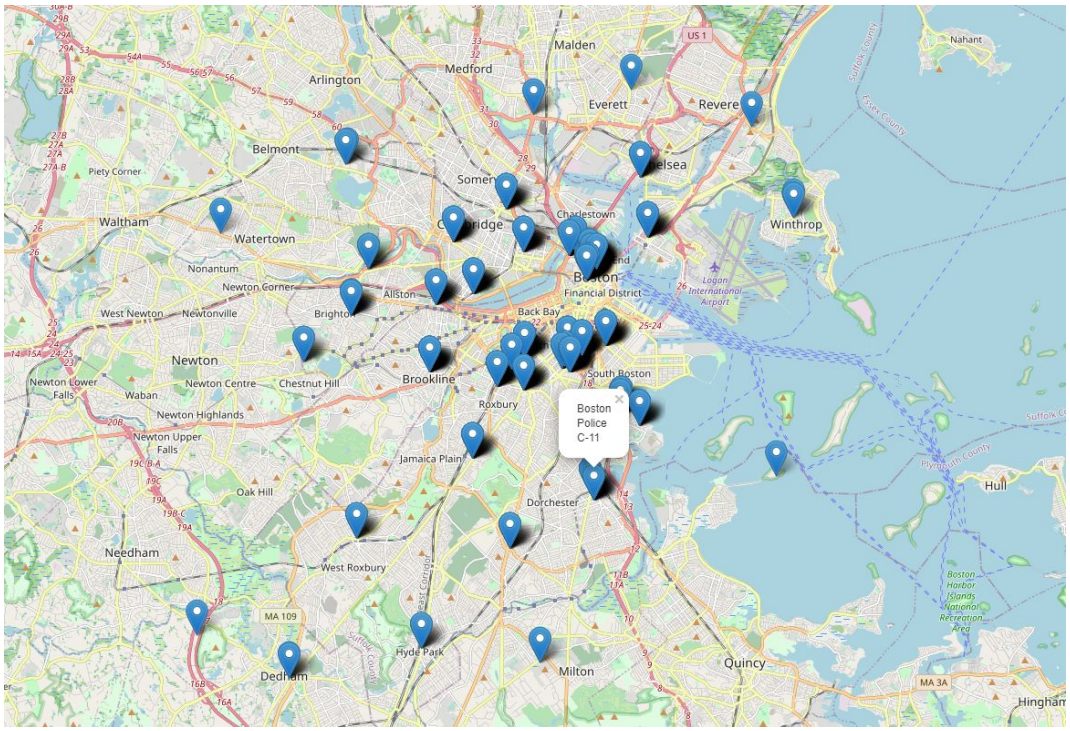

Figure 8. Map of Boston Police Stations

Table 2. Boston Police Stations Details

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Allston	42.358405	-71.128563	Boston University Police Station	42.352361	-71.118470	Police Station
1	Allston	42.358405	-71.128563	Mass State Police H5 Brighton	42.362389	-71.143929	Police Station
2	Allston	42.358405	-71.128563	Harvard University Police Department	42.369880	-71.112101	Police Station
3	Allston	42.358405	-71.128563	MIT Police (Building W89)	42.355546	-71.104676	Police Station
4	Allston	42.358405	-71.128563	Boston Police D-14	42.349644	-71.150468	Police Station
...	...	...	...	...	...	...	...
462	West Roxbury	42.280642	-71.160027	Boston Police E-5	42.287091	-71.148411	Police Station
463	West Roxbury	42.280642	-71.160027	Dedham Police Headquarters	42.248231	-71.173848	Police Station
464	West Roxbury	42.280642	-71.160027	Dedham Police Central Dispatch	42.248048	-71.174021	Police Station
465	West Roxbury	42.280642	-71.160027	Boston Police E-18	42.256470	-71.124222	Police Station
466	West Roxbury	42.280642	-71.160027	Norfolk County House Of Corrections	42.260140	-71.208710	Police Station

Correlation between Police Station and Crime Incidents

So far, we have looked at the crime incident trend, and if there is any seasonality when incidents are more likely to occur in a year, neighborhood safety based on single criteria crime incident rates. The other question is, would a neighborhood with more police stations necessarily have fewer crime incidents? Or would a neighborhood with more densely located police stations have fewer crime incidents per square mile? As noted in the Methodology section, I am using the number of police stations, and the number of police stations per square mile as a proxy to compare neighborhoods police force staffing, and unit area normalization method. The caveat is it may not take into account the actual numbers of police officers, which might vary across police stations, and the actual population. Also, the Neighborhood may not necessarily be the same as police station's district or reporting area.

Firstly, Figure 8 illustrates the Number of Police Stations on the horizontal axis, and number of crime incidents on the vertical axis, and the square mile of a neighborhood on the size of the circle marker.

- As the chart shows, Dorchester neighborhood has the largest number of crime incidents, followed by Roxbury, and then Downtown. These three neighborhoods

have 19, 25, and 28 Police Stations within the boundary of the neighborhoods, respectively.

- Longwood, South Boston Waterfront, North End, and Beacon Hill neighborhoods have the lowest crime incident rates, and ones of the highest numbers of police stations.
- There is no significant correlation between the number of police stations and crime incidents.
- When excluding Dorchester, no significant correlation was observed either.

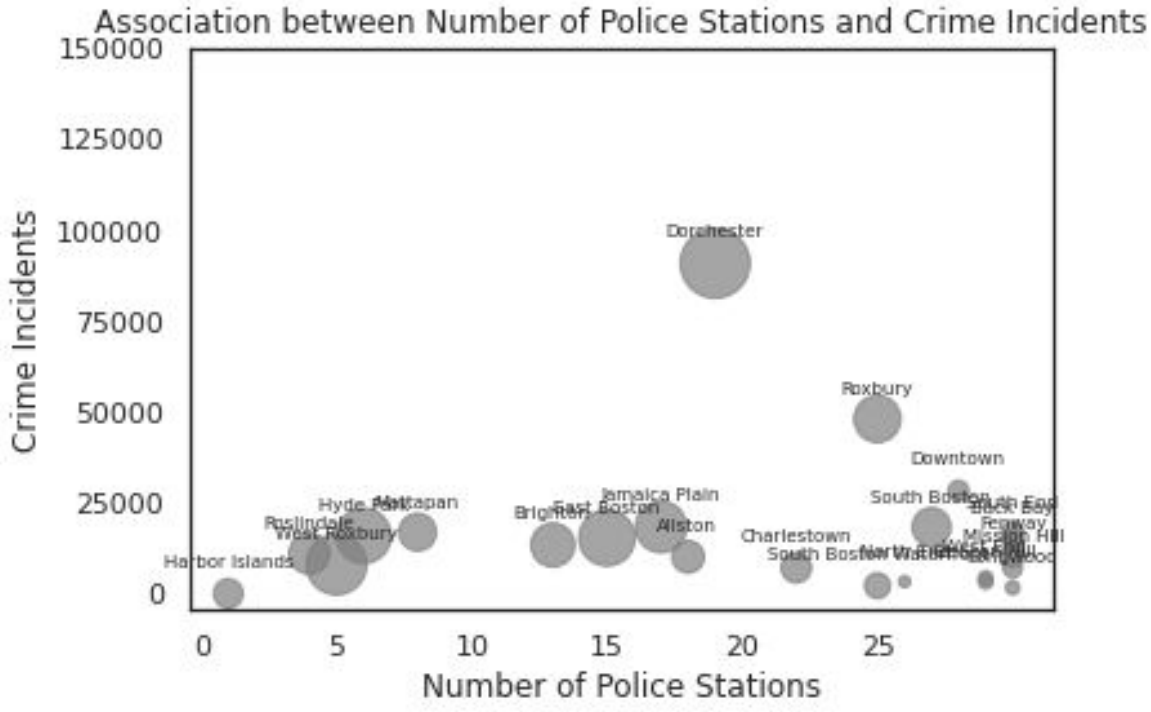

Figure 9. Association between Number of Police Stations and Crime Incidents

Then, Figure 9 illustrates the Police Stations density on the horizontal axis, and crime incident density on the vertical axis, and again the square mile of a neighborhood as the size of the circle marker.

- As the chart shows, Downtown has the largest crime incident rates per square mile, and a higher percentile of police station density.

- Comparatively, Downtown has similar police station density, but is almost twice as high as Back Bay's crime incidents density.
- Comparing North End and Mission Hill, North End has almost 2.5 folds higher police station density than Mission Hill, however the crime incidents rate is at the approximately same level as Mission Hill.
- Neighborhoods in the suburbs have a lower number of police stations, larger square mile areas, and lower police station density. However, the crime numbers are among the highest (top 50th percentile).

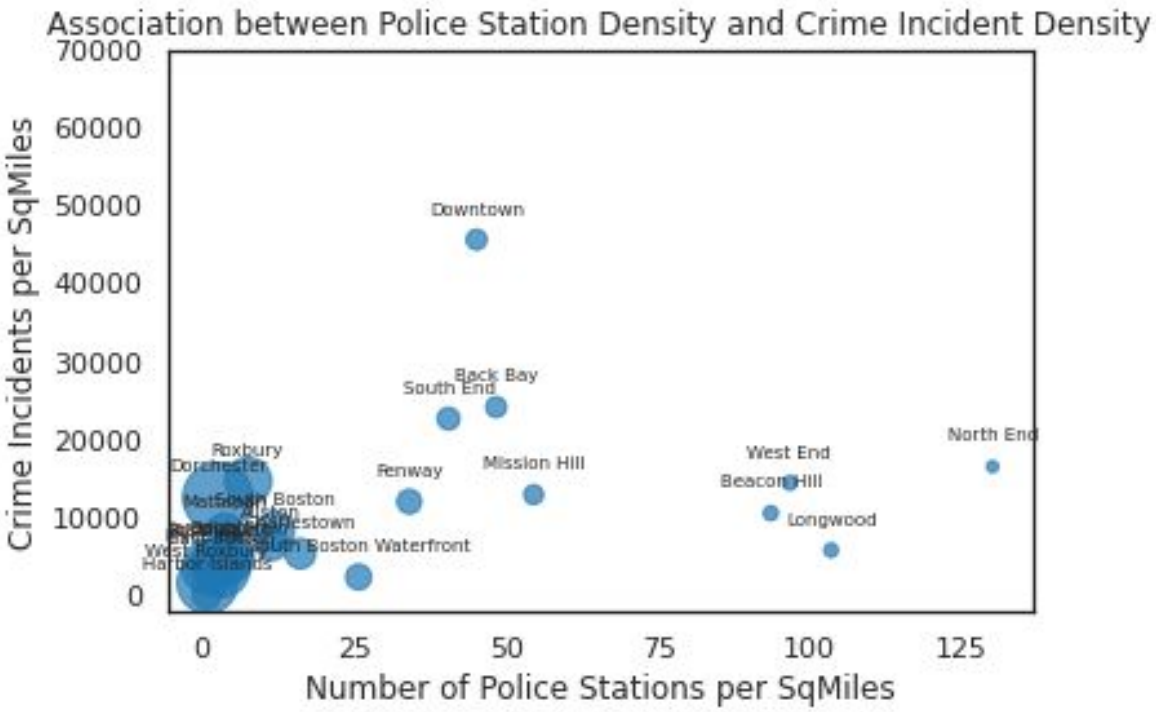

Figure 10. Association between Police Stations and Crime Incident Density

Conclusion

Crime incidents have been trending down since 2017.

- During 2015 and 2020 (through May, annualized), Boston has the highest number of crime incidents in 2017.

- Since then crime incidents have been consistently trending down.

Crime incidents are more likely to occur during summer time, and least likely during winter.

- Specifically, summer months May through August had the highest numbers of crime incidents, while winter months November through February had the lowest incident occurrences.

Crime Categories/Types do not change significantly over the years.

- Motor vehicle accidents, larceny, *medical assistance (needed or dispatched),* investigate person, other, drug violation, simple assault, vandalism, verbal dispute, and investigate property are among the most common types of incidents occurred in Boston during the past five years, 2015 - 2019.
- No significant variance was observed during the past five years regarding the crime incidents groups, though the order may have shifted slightly.

More number of police stations is not necessarily associated with lower number of crime incidents, or higher police station density (number of police stations per square mile) not necessarily correlated to a lower crime incidents density.

- Downtown has similar police station density, but is almost twice as high as Back Bay's crime incidents density.
- North End has almost 2.5 folds higher police station density than Mission Hill, however the crime incidents rate is at the approximately same level as Mission Hill.

As future opportunities, additional data points may need exploring and plugging in as predictors for a better fit model, e.g. demographics including population, median household income, crime category stratification, etc.