Exploratory Data Analysis of Boston Crime Data

EAS 503, Fall Semester
Final Project Report

Submitted by:

Group Number 7

Chaitanya Pawa
José Barajas
Syed Aqhib Ahmed

Motivation

In efforts to improve efficiency and overall performance of Boston first responders, Boston Crime data has been analyzed via exploratory data analysis methods to help answer some of the following questions:

- Are there any areas that exhibit high crime rates, but lack the resources to respond to those crimes (police stations, fire stations, etc.)?
- Do certain hours of the day, days of the week, or months exhibit disproportional amounts of crime?
- How does crime differ by its frequency?
- How does crime differ by its type?

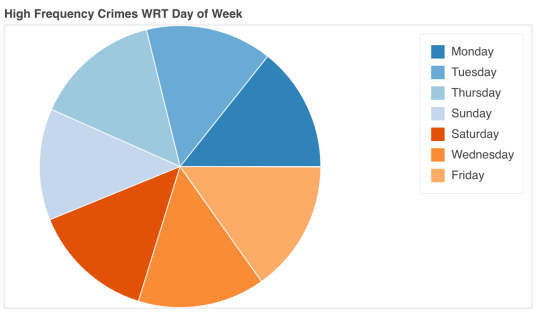
Data

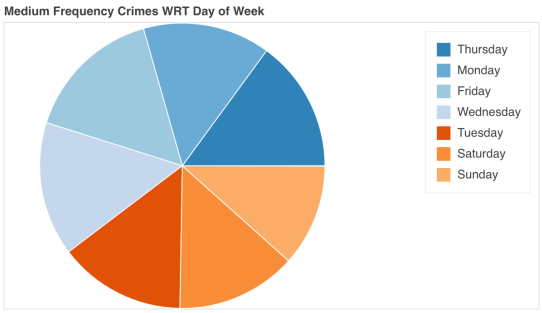
- Data on Boston crime incidents, location of police stations, locations of fire stations were acquired via Analyze Boston.
- Additionally, the data was subsetted by types of crimes with high frequency crimes (>4000), medium frequency (>4000 and >500), and low frequency crimes.

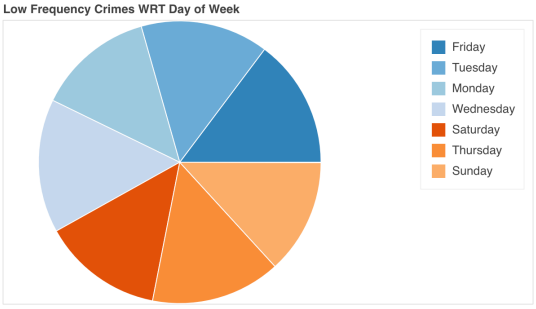
Analyzing Crimes by Day of the Week

Crimes were analyzed over different days of the week. Below the pie charts indicate the portion of crime committed for each day of the week for high, medium, and low frequency crimes.

- Crimes do not seem to vary by day of the week.



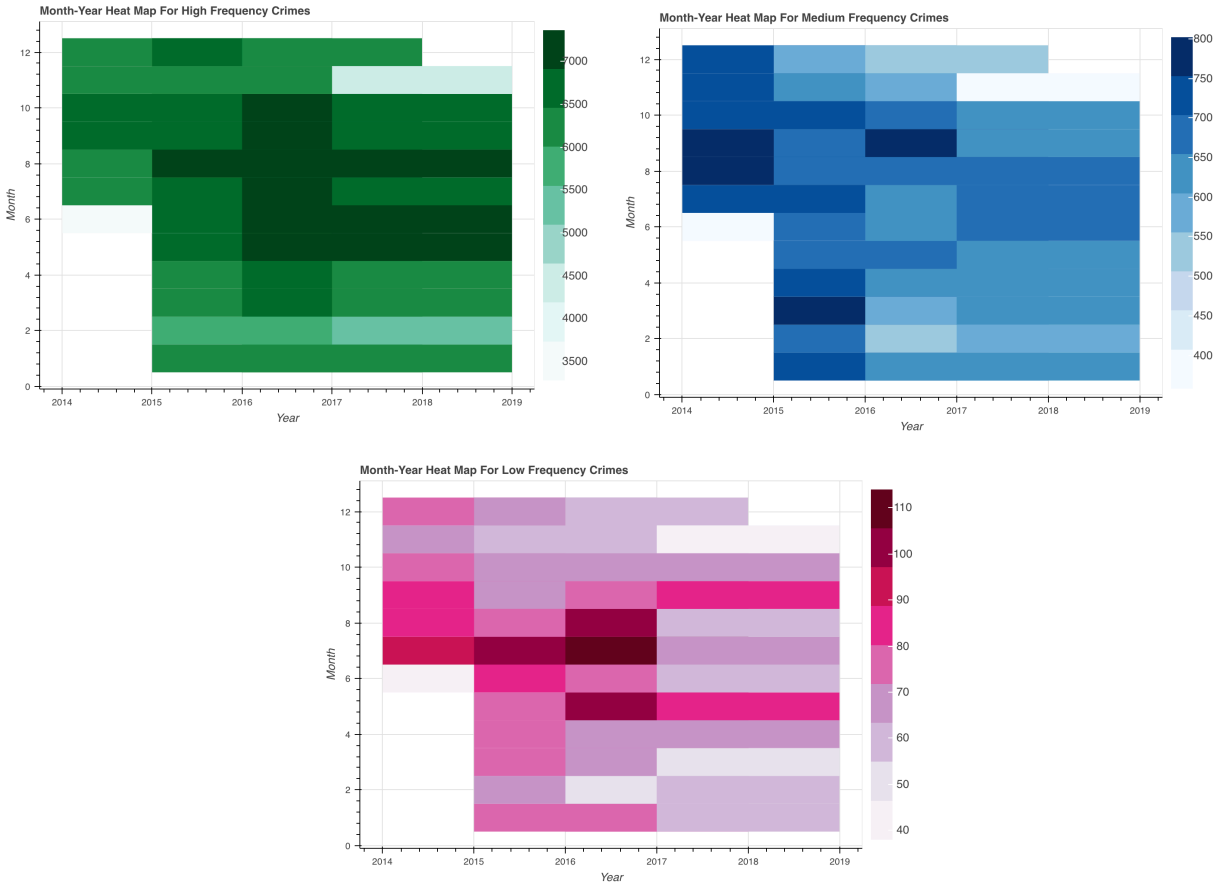




Analyzing Crimes Month by Month

Crimes were analyzed on a month by month basis over the past 6 years. Three heat maps for high, medium, and low frequency crimes were made in order to see how crimes vary by month. Color intensity indicates the average amount of crimes for that month

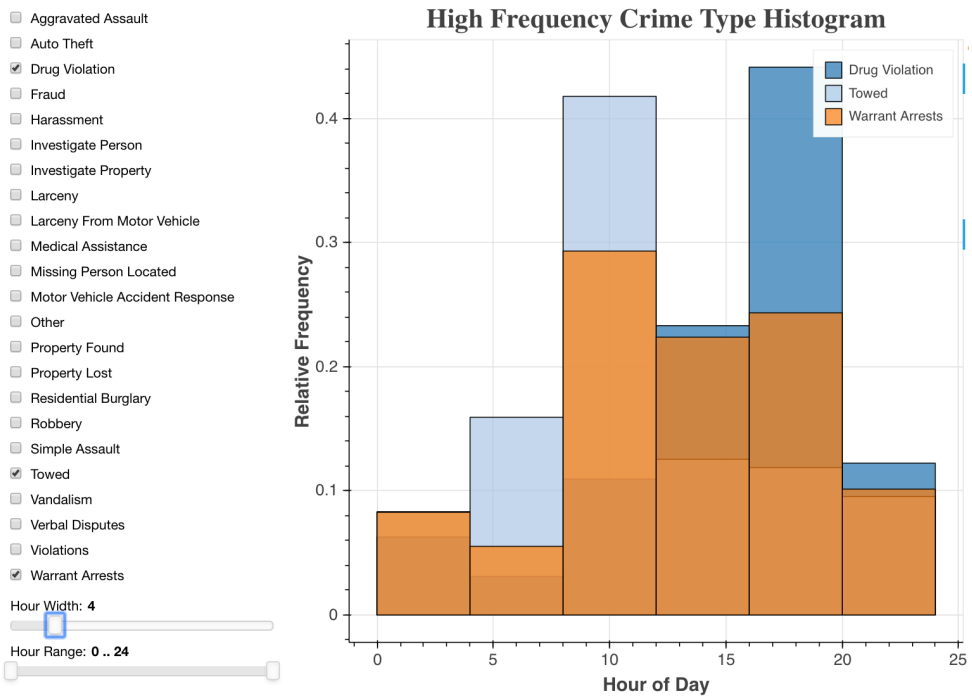
- Summer months (especially August) seem to consistently have a greater crime rate in all categories.
- Post 2017, medium and low frequency crimes seem to have dropped off (relatively), while high frequency crimes only intensified in rate.
- Winter months see a consistent decrease in crime for all categories.



Analyzing Crimes by Hour of the Day

Crimes were analyzed over a 24 hour period. Three interactive histograms for high, medium, and low frequency crimes were made in order to see how crimes vary by hour of the day.

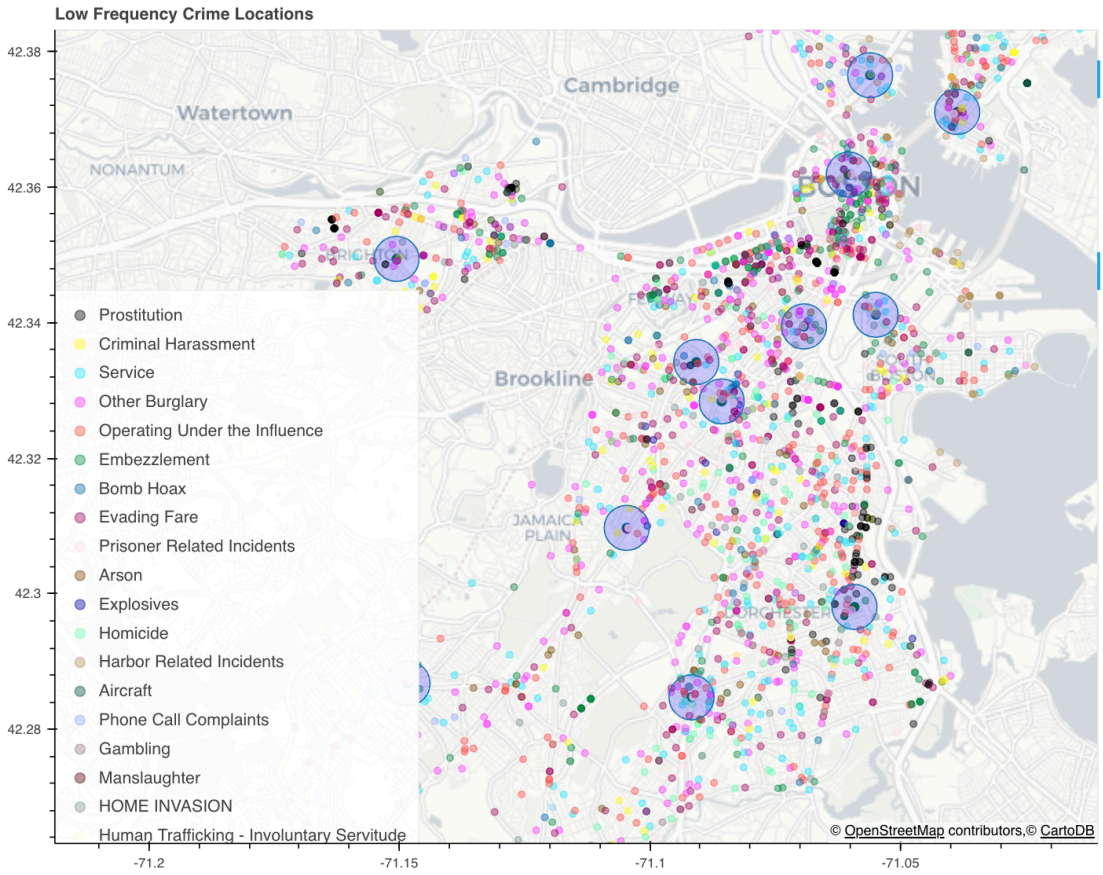
- High frequency crimes:
  - Most tended to skew during the afternoon and evening.
  - Specifically, crimes involving drug violations have a significant spike between 4PM - 8PM
  - Towed violations has an overwhelming skew between 8AM and Lunch time
- Medium frequency crimes:
  - Tended to skew towards mid day
  - Not as predictable as high frequency crimes
- Low frequency crimes:
  - Had more outliers (Biological threats, human trafficking, etc.)
  - Intuition holds for the rest (i.e. we expect offenses such as driving under the influence happen during fringe hour of the day).



Similar visualizations were used for analyzing the medium and low frequency crimes.

Analyzing Crimes by Location

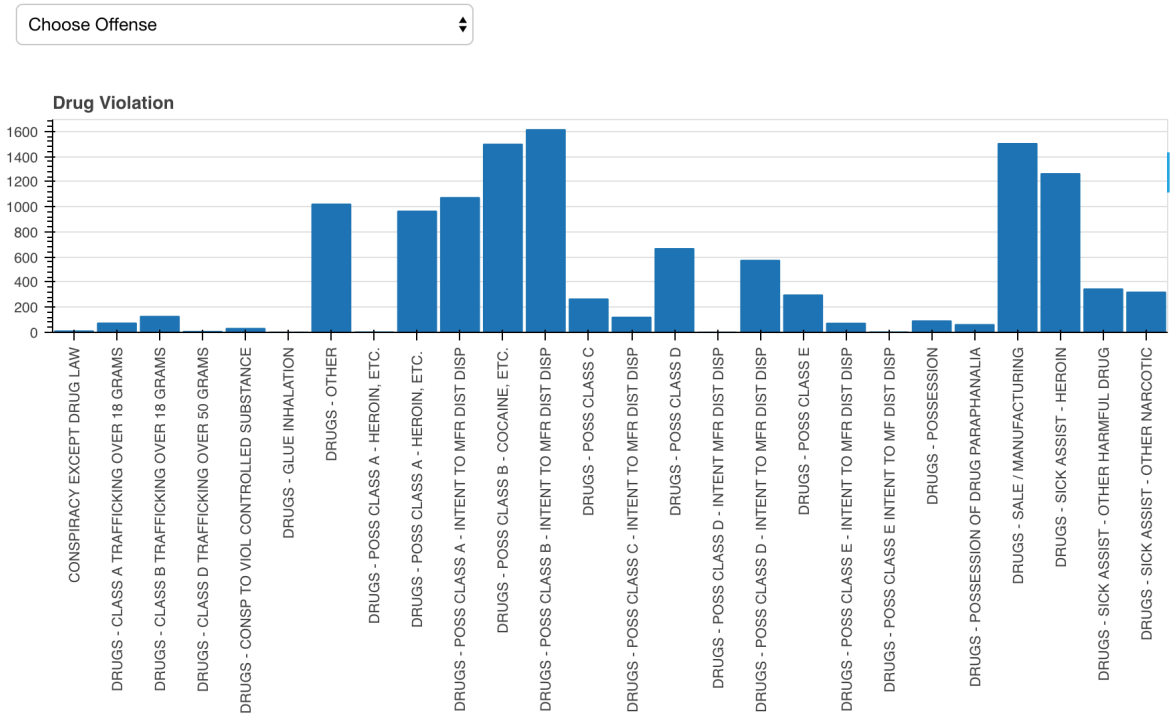
Crimes segregated by type are marked on the map, areas with a proportion of that crime type are more intense in that crime type's associated color. Additionally, response stations are boldly marked.



The areas with type and frequency of the crimes can be analyzed according to their locations along with locations of the response stations to recommend for any changes in the response services.

Analyzing Crimes by Offense Description

The offense description is a more detailed set of information on the nature of the crime type, for instance operating under the influence can be further categorized by operating under the influence of alcohol or drugs.



Future Work

In the future, these results can be employed when assessing:

- Where new police stations should be built by analyzing areas with higher crime rates, but lack police stations.
- How frequently should officers go on patrols by looking at temporal data.

Extend results to 911 operators:

- We can assess the need for operators by analyzing how crime varies temporally

Extending project analysis:

- Creating a predictive model depending on the geographical and temporal data, the caller can be immediately transferred to an operator with the experience and resources tailored to the callers need.
- Extend analysis to fire stations and assessing incidents that require the fire department.