

Initial Analysis

William Scott Breckwoldt

College of Professional Studies: Analytics, Northeastern University

ALY 6070: Communication and Visualization for Data Analytics

Professor Alli Maurer, PhD

January 27, 2022

Initial Analysis

Introduction

The Initial Analysis of the Crime Incident Reports in Boston (August 2015 to September 2020) includes data visualizations using dashboards, scorecards, spatial data representations, and use of annotations, complemented by evaluations, interpretations, and recommendations.

The goal of this assignment is to implement the skills I have learned in Module 1, 2, and 3 involved in designing dashboards, utilizing graphic designs, creating and presenting effective data visualizations, and using ethical strategies. My dashboards will “tell a story” with a narrative flow. The graphic design concepts will enhance accessibility and aesthetics. I will choose effective data visualizations that are relevant to the context and are concise. These visualizations will be presented such that the focus is on key data insights. Finally, I will use ethical strategies and create visualizations that are not biased and will not mislead the audience.

In the following analysis, I will focus on the achieving the goals written above while answering the questions provided by Dr. Maurer:

- What are the variables in the dataset?
- What correlations, patterns, and trends do you see?
- Who will be the audience of the data? What question may they ask about the data?
- What other questions may the audience present?
- How could a dashboard be used to show data clearly?
- What types of graphs and charts can be used to clearly explain the data and answer the business question?

After the analysis, I will provide a conclusion, in which I will provide my overall interpretations of the data visualizations, observations, and significant takeaways from the analysis.

Analysis

Dataset and Variables

The three files used in this assignment include the dataset, which is the Crime Incident Reports (August 2015 – September 2020).csv and the RMS_Crime_Incident_Field_Explanations.xlsx and the RMS_Offense_Codes.xlsx. The RMS_Crime_Incident_Field_Explanations.xlsx provides explanations for some variables. The RMS_Offense_Codes.xlsx shares the offense description associated with the offense numerical code.

The Crime Incident Reports (August 2015 – September 2020).csv dataset has 17 variables and 512,364 observations. I will list the variables with their descriptions from the RMS_Crime_Incident_Field_Explanations.xlsx file and my own observations:

1. INCIDENT_NUMBER: 8-to-10-digit internal Boston Police Department report number
2. OFFENSE_CODE: Numerical code of offense description
3. OFFENSE_CODE_GROUP: Internal categorization of the offense
4. OFFENSE_DESCRIPTION: Primary descriptor of incident
5. DISTRICT: What district the crime was reported in
6. REPORTING_AREA: Reporting area number associated with where the crime was reported from

7. SHOOTING: Indicated a shooting took place; N or 0: shooting did not take place, Y or 1: shooting did take place
8. OCCURRED_ON_DATE: Earliest date and time the incident could have taken place
9. YEAR: Year of the incident
10. MONTH: Month of the incident
11. DAY_OF_WEEK: Day of the week of the incident
12. HOUR: Hour of the day of the incident (0-23)
13. UCR_PART: Universal crime reporting number (1,2,3)
14. STREET: Street name the incident took place on
15. Lat: Latitude where the incident took place
16. Long: Longitude where the incident took place
17. Location: combined latitude and longitude where the incident took place

My initial thoughts were that trends in time and location and type and amount of crime may be identified using data visualizations in Tableau. To make this dataset usable in Tableau, I saved the .csv as a .xlsx file.

Questions, Data Visualizations, and Observations

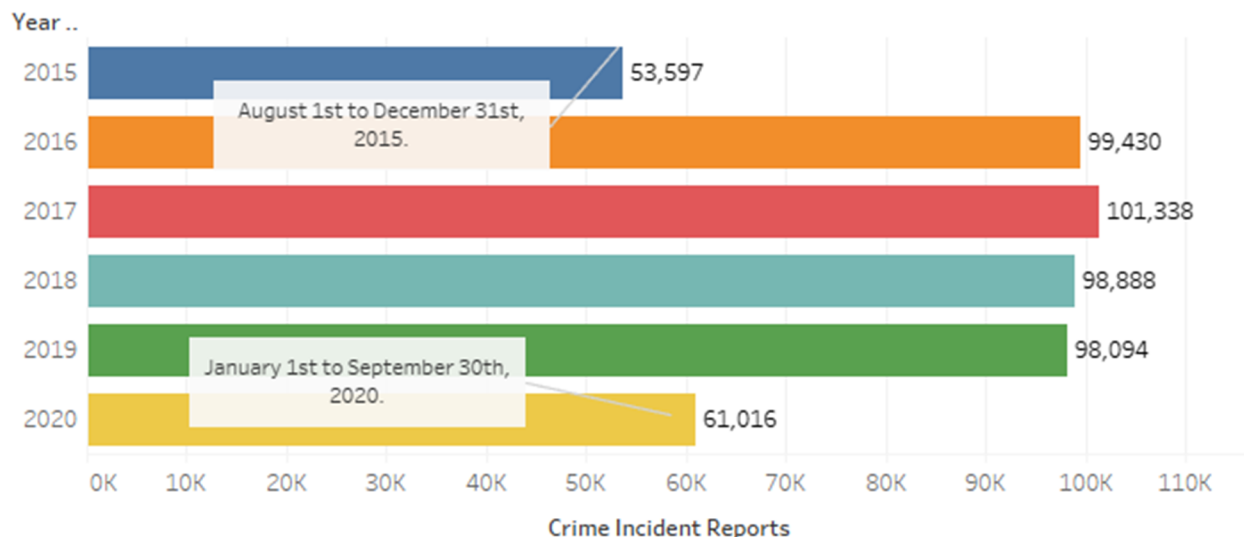
Those interested in this dataset and the data visualizations one can present would be the Boston Police Department and citizens of Boston. Through data visualizations the Boston Police department could get a good idea of where to focus their resources and citizens could get an idea of the safer and more dangerous areas and times. However, it should be noted that I am only an analytics student and a subject matter expert, a criminal justice or law professional, may provide better insight on my data visualizations and observations.

How many crime incidents have been in recorded annually in Boston?

First, I decided to present data visualizations that represent crime and time. The following data visualization is a simple horizontal bar chart that demonstrates crime incident reports per year.

Figure 1

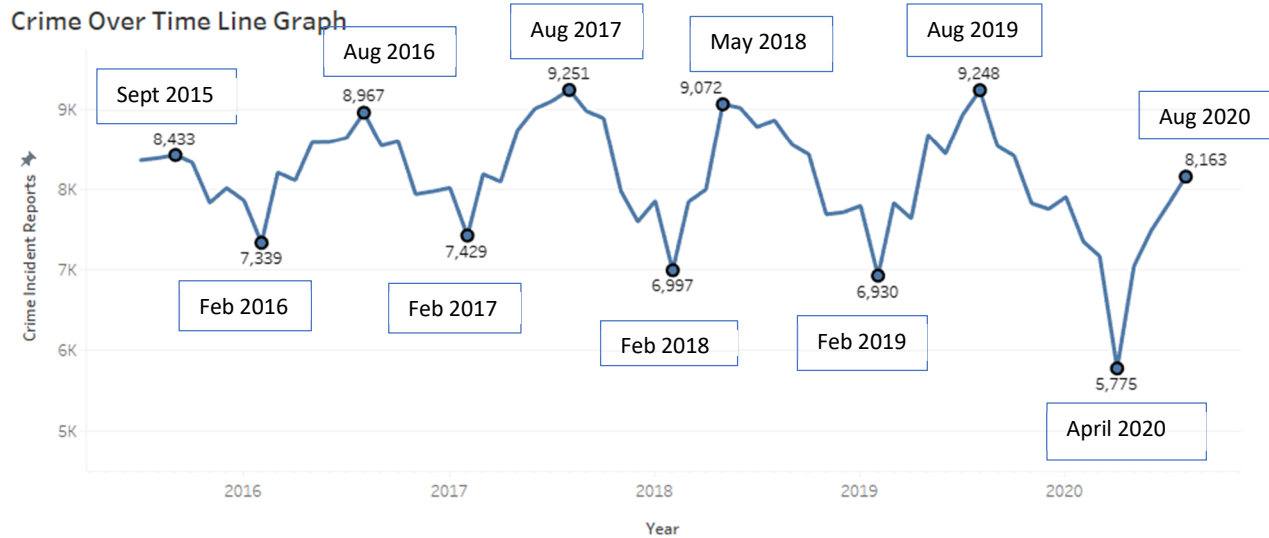
Crime by Year Bar Plot



Unfortunately, the dataset does not include data for the entire years of 2015 and 2020. However, this visualization is still insightful. We can see that 2017 had the highest number of crime incident reports followed by 2016, 2018, and 2019. It appears that crime in Boston has been declining since 2017.

What months have the most and least amount of crime incident reports?

Figure 2

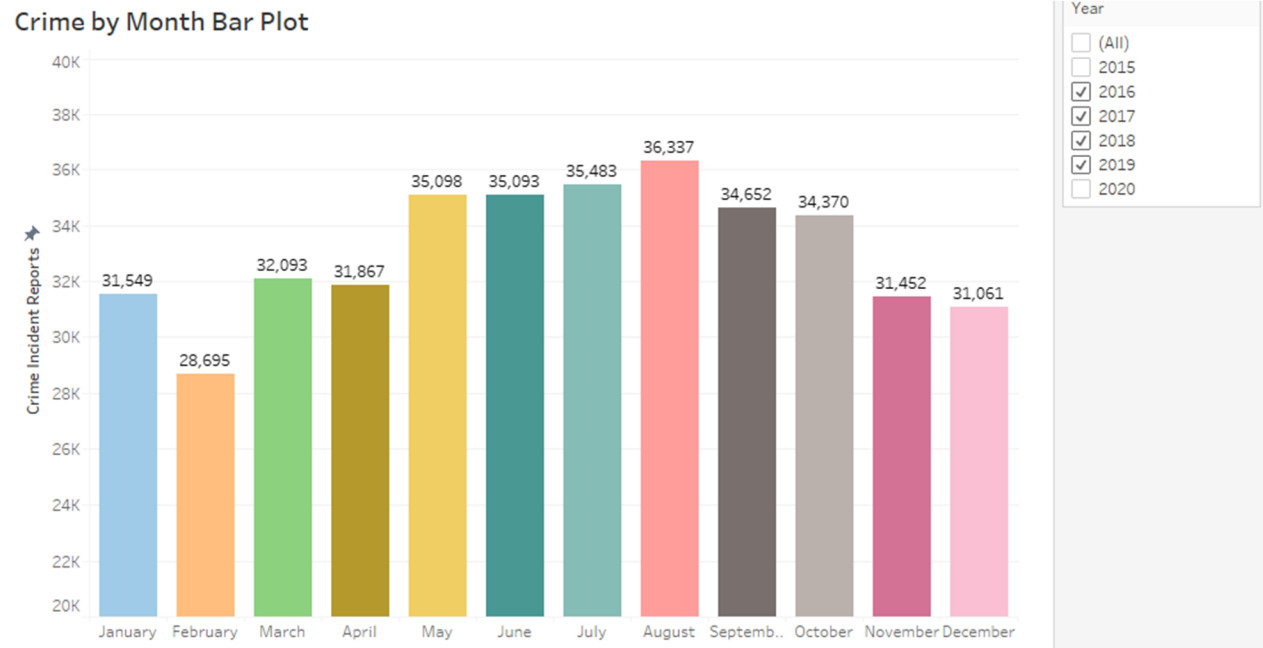


This line graph demonstrates crime incident reports by month overtime. The y-axis limits are set to 4,500 and 9,750. These limits show a more drastic difference in the peaks and troughs, as opposed to a y-axis that began at 0.

My strategy here was to avoid blank space and highlight the differences in crime per month. I hope and think that they will not be misled after addressing that and labeling the peaks and troughs.

In Tableau, by hovering over the points one can see the number of crimes reported in each month of a year for each point. In this .png image, I was only able to label the number of crime incidents in the peaks and troughs of the line graph, then add text boxes to display the month and year.

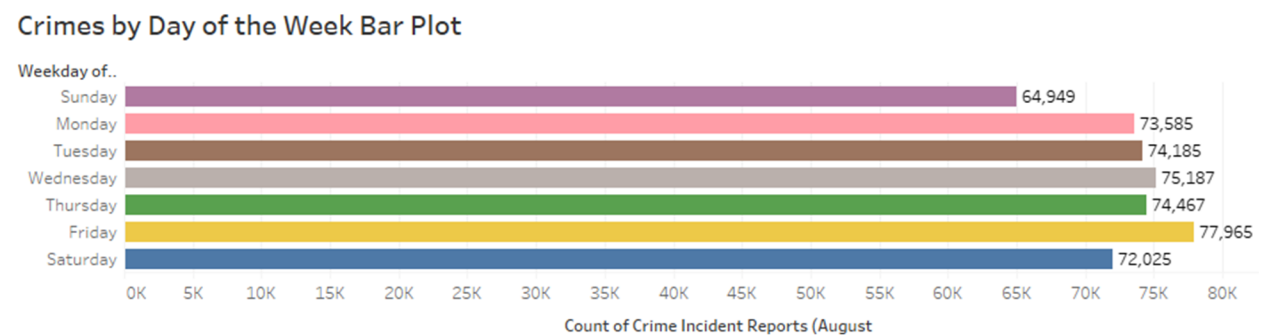
Despite September 2015 and May 2018, it appears that August generally has the greatest number of crime reports. Furthermore, despite April of 2020, which had the least number of crime reports (5,775), February generally has the lowest number of crimes reported in the year. Let's examine this in the next data visualization.

Figure 3

This vertical bar plot contains data between January 1st, 2016, and December 31st, 2019, as shown by the filter in the top right. This is because data was not recorded for the entire years of 2015 and 2020. It also should be noted that the y-axis begins at 20,000 and ends at 40,000. This was to avoid taking up too much space and such that we could clearly see the labels and differences in the plot. Again, the audience should understand this, such that the graph is not misleading but informative.

We can see that as suggested in the line graph, August (36,337 incidents) has the most crime reported, and February (28,695 incidents) has the least crime reported. It seems that in the winter crime incident reports are less frequent than in the summer.

What days of the week have the least and most crime incident reports?

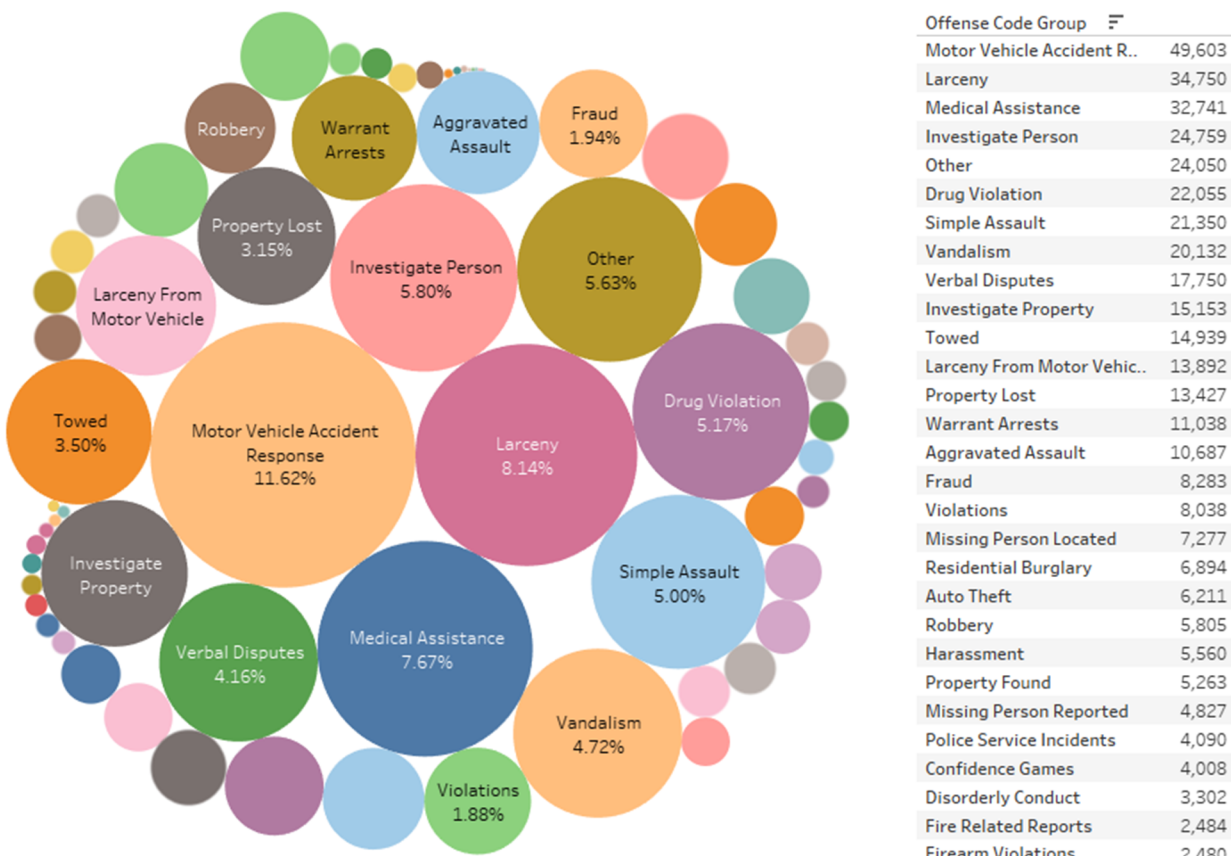
Figure 4

This horizontal bar plot demonstrates crime incident reports by day of the week. Friday has the most (77,965 incidents) followed by Wednesday (75,187 incidents). Sunday has the least (64,949).

What crimes are most frequent?

Figure 5

Offense Group Bubble Plot and Scorecard Dashboard

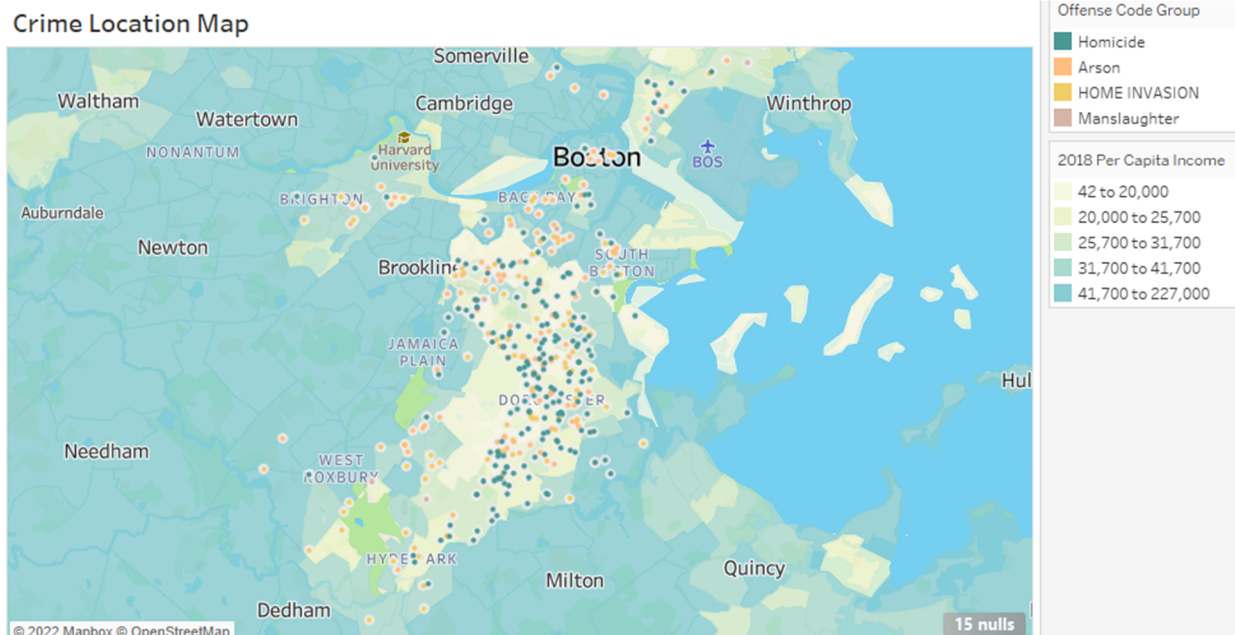


Here is a bubble plot demonstrating offense groups and their frequency as a percentage on the left and an offense group scorecard on the right. The 3 most frequent offense groups are Motor Vehicle Accident Response (11.62%; 49,603), Larceny (8.14%; 34,750), and Medical Assistance (7.67%; 32,741). It must be noted that I excluded the NULL offense group.

What areas are the safest and most dangerous?

The best data visualization technique to use in this scenario would be a map. However, with 512,364 incidents, we cannot include all observations in a screenshot as that would be hard to understand. To display the most safe and dangerous areas, I will only include incidents of homicide, arson, home invasion, and manslaughter.

Figure 6



It is clear by looking at the map, that there is more crime activity in areas with less per capita income. Areas such as Hyde Park, Mattapan, Ashmont, Dorchester, Grove Hall, Roxbury, and East Boston appear to be where most of these crimes are occurring. On the other hand, Back Bay, South End, South Boston, Seaport, Beacon Hill, West End, Financial District, North End and Charlestown all appear to have a low number of these crimes.

Conclusion

In the Analysis, I found trends and patterns in crime and time and area. I identified the year with the most total crime incident reports, as well as the time of year and days of the week with the most and least incidents. Furthermore, the most common type of crimes reported were identified. Finally, I found the neighborhoods that appear to have the most crime activity.

The goal of this assignment was to implement the skills I learned in Module 1, 2, and 3 involved in designing dashboards, utilizing graphic designs, creating and presenting effective data visualizations, and using ethical strategies. My dashboards “tell a story” with a narrative flow. The graphic design concepts enhance accessibility and aesthetics. I choose effective data visualizations that are relevant to the context and are concise. These visualizations will be presented such that the focus is on key data insights. Finally, I used ethical strategies while creating visualizations that are not biased and do not mislead the audience.

I also answered the questions presented by Dr. Maurer. I defined the variables in the dataset. I identified correlations, trends, and patterns displayed in my data visualizations. I assumed my audience

and presented questions they may ask. These questions were answered with graphs, charts, dashboards, spatial data representations, and annotations, complemented by evaluations, interpretations, and recommendations.

It should be noted that this analysis may benefit from a subject matter expert. As an analytics student, not a criminal justice or law professional, I cannot accurately attribute these patterns and trends in my data visualizations. Someone with experience in a field that deals with crime or investigation may provide more insight.