Declaration on Plagiarism

Name:	Kirthy Francis
Student Number:	19210588
Programme:	MCM
Module Code:	CA682
Assignment Title:	Data Visualisation
Submission Date:	15 Dec 2019
Module Coordinator:	Dr Suzanne Little

I declare that this material, which I now submit for assessment, is entirely my own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my work. I understand that plagiarism, collusion, and copying are grave and serious offences in the university and accept the penalties that would be imposed should I engage in plagiarism, collusion or copying. I have read and understood the Assignment Regulations. I have identified and included the source of all facts, ideas, opinions, and viewpoints of others in the assignment references. Direct quotations from books, journal articles, internet sources, module text, or any other source whatsoever are acknowledged and the source cited are identified in the assignment references. This assignment, or any part of it, has not been previously submitted by me or any other person for assessment on this or any other course of study.

I have read and understood the referencing guidelines found at http://www.dcu.ie/info/regulations/plagiarism.shtml, https://www4.dcu.ie/students/az/plagiarism and/or recommended in the assignment guidelines

Name: Kirthy Francis Date: 15 December 2019

Uniform Crime Reports (Boston)

Abstract

In ordinary language, a **crime** is an unlawful act punishable by a state or other authority. Uniform Crime Reporting (UCR) Program compiles official data on crime in the United States which is then published by the Federal Bureau of Investigation (FBI) (from Wikipedia). According to the UCR, there are three main categories in crime – part 1, part 2 and part 3. Part 1 being the most severe crimes. Using data on crime reports from the Boston Police Department, the following are the questions I aim to answer using the visualizations.

What is the most common category of crime and what are the crimes that come under each category? When are crimes from each category likely to occur? Which areas of Boston are affected by which category of crime?

From the visualisations, we find that Part 3 crimes are the most common crimes and part 1 crimes less common. It could also be concluded that most crimes take place late at night.

1. Dataset

The data obtained for this study was downloaded in CSV format from Analyze Boston. Analyze Boston is the City of Boston's open data hub which provides many datasets for analysis. The dataset consists of crime incident reports that are reported by the Boston Police Department (BPD) to document the initial details surrounding an incident to which BPD officers respond. The data has been collected from August 2015 till November 2019. The data consists of the different offense codes and their description i.e. the different crimes, offense groups, the location where the crime occurred including latitude and longitude, date and time the crime was committed among many others in 17 attributes. Data types include latitude and longitude, date, integer and characters. The raw dataset consisted of 445328 rows in 74.3 MB. After cleaning, the dataset consists of 395773 rows in 67.8 MB.

2. Data Exploration, Processing, Cleaning and/or Integration

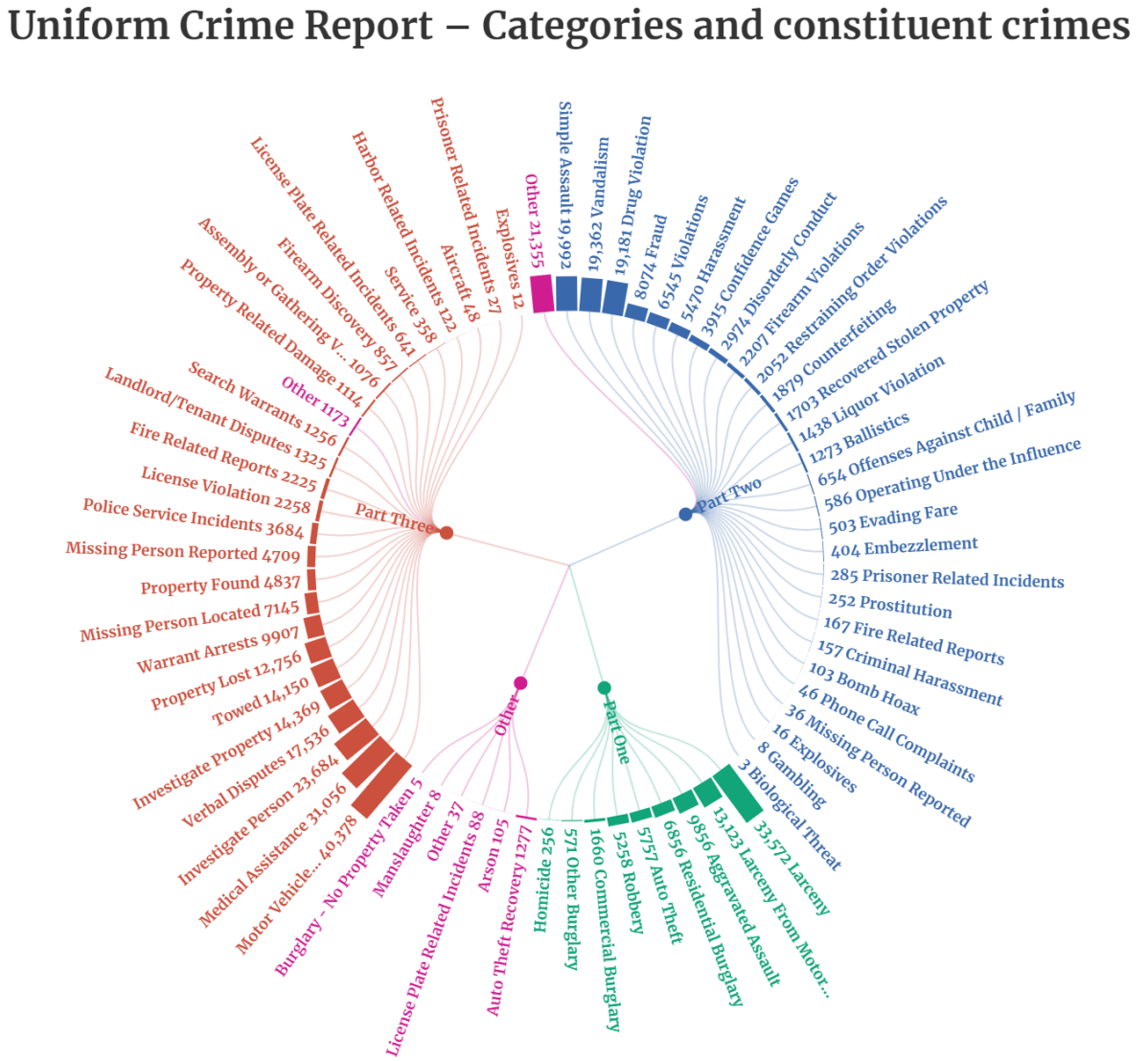
The dataset contained many null values in the DISTRICT, REPORTING_AREA, UCR_PART, STREET and Lat columns. I removed all the rows which contained null values in these columns using python. The SHOOTING column contains values only when the value is true and since I will not be using this column, I dropped it. I also dropped the Location column which was a repetition of the Lat and Long column.

For answering the question of which category of crime is most common and the crimes that come under each category, I created a radial tree graph that used OFFENSE_CODE_GROUP and UCR_PART attributes. The second question of when most crimes were committed was answered using a histogram using the attributes HOUR and

UCR_PART. The areas in Boston affected by crimes used the attributes Latitude, Longitude, UCR_PART and DISTRICT to show a map which shows the areas affected by each category of crime.

3. Visualisations

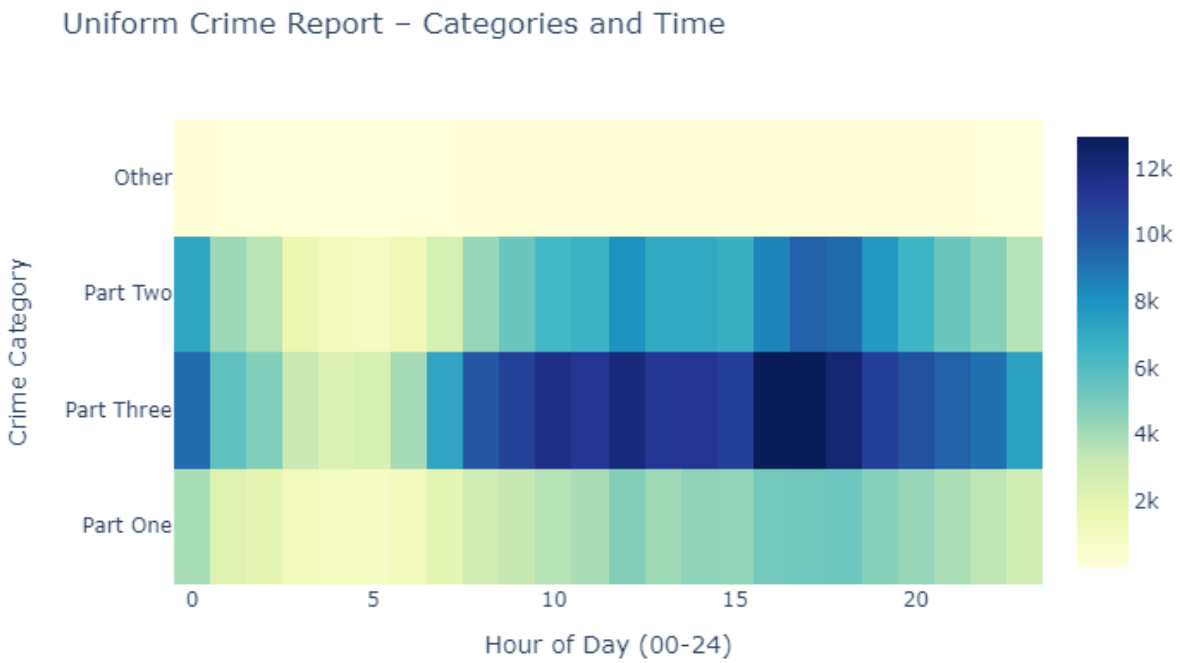
i. Uniform Crime Report – Categories and constituent crimes



This radial tree was created using Flourish Studio. I used a radial tree to show a hierarchical relationship between UCR_PART and OFFENSE_CODE_GROUP attributes. The graph shows the different crime groups and the category under which they fall. I used the Flourish 2 color combination and removed colors like yellow so that all the text labels would be visible. Each category is sorted in decreasing order of the crimes committed. Each crime label also shows the total number of crimes in that crime group. The animation shows the total in each category, here we can see part 3 crimes are most common, followed by the crime groups in the third category and all the crime groups in all categories (figure above).

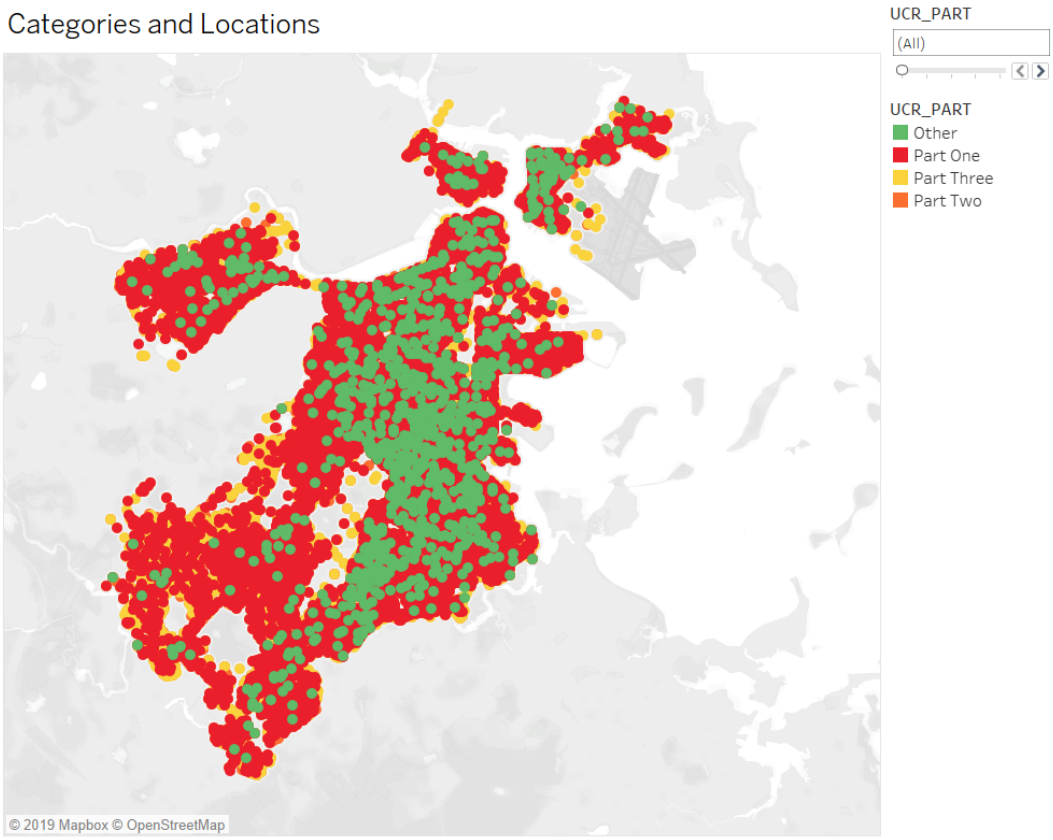
ii. Uniform Crime Report – Categories and Time

The following histogram was created in a jupyter python notebook using the graph objects function from the plotly library. It shows the number of crimes from each category that occurs at which hour of a day. The greater the number, the darker blue the hour of the category will be. A histogram best explains the relationship between the different categories of crime and the time they were committed. Here we can see that most of the crimes occur during the evenings. The others category contains very little data and hence the hours cannot be distinguished as there are only about 1000 records in others combined.



iii. Uniform Crime Report – Categories and Locations

The map shows the different areas in Boston affected by different degrees of crime. The map was created in Tableau. I used red color to represent Part 1 crimes as they are most severe, orange for Part 2 and yellow for Part 3 to represent diminishing severity. Each point on the graph represents a location of crime, the severity of crime and the district. You can select a category to view all the areas where that category of crime has taken place over the time period. You can click on any data point to view all the crimes in the category and district of the selected data point.



4. Conclusion

The visualizations all show that the most common type of crimes in Boston belong to Part 3, which are lesser severe crimes. There are a minority of crimes categorized ‘others’ which could possibly be included in the three categories. Most of the crimes are committed in the evening times primarily between 16:00 and 20:00 hours as is observed from the histogram. The map shows the areas where crimes from the different categories were committed. In the radial graph, the font cannot be increased to improve the clarity of the text. In the histogram, the ‘others’ category contains very small data as compared to Part 1,2 and 3 that the whole category has uniform color and cannot be distinguished between hours.

References

Dataset - https://data.boston.gov/dataset/crime-incident-reports-august-2015-to-date-source-new-system

Histogram - https://www.kaggle.com/ismailsefa/crimes-data-analysis-and-visualization-eda/notebook#Plotly---2D-Histogram-(-Interactive-)

I referenced this histogram to create mine.

Radial graph - https://flourish.studio/