Baltimore Police Data Analysis

STAT 440 Case Study 2

**Guangji Zou (Ray)**

Student ID: 301216239

Group Member: Helen Fu and Terry Liu

Email address: [guangjiz@sfu.ca](mailto:guangjiz@sfu.ca)

Introduction:

In this research, the dataset of Baltimore Police includes the 911 call services data from 2015 to 2017. Dataset was formed by more than 290000 records of 911 call services. Each record has the call time date, location, priority, and the description. In this method section, we used the two different methods to describe the dataset of Baltimore of Calls For Service, the three methods are following: data cleaning and Data visualization.

Method:

Our dataset is based on the Open Baltimore of call for service, and source dataset is from 911 calls for service

Given the dataset, first, data cleaning, we try detecting and correcting (or removing) corrupt or inaccurate records from a record set, table, or database and refers to identifying incomplete, incorrect, inaccurate or irrelevant parts of the data and then replacing, modifying, or deleting the dirty or coarse data. Data cleaning may be performed interactively with data wrangling tools, or as batch processing through scripting.

After the cleaning dataset, the data should be similar with the original dataset in R system. As the steps, we random choose five percentage the original data, and we want to remove typographical errors or validating and correcting values against a known list of entities. Then when we view the dataset, we find out there a lot of missing value in this dataset. To solve this problem, we would like to use mice function to impute on the numeric by Predictive Mean Matching (PMM) method. Finally, there are 148510 observations of 6 variable. The all location of the large matrix includes 297020 elements.

Secondly, data visualization (or data visualisation) can help us to describe any effort for understand the significance of data by placing it in a visual context. Our goal of data visualization is to communicate information clearly and label the criminal points and information graphics in each region of Baltimore so that we combined the density map and google map in the real world. Then see clearly about which area had crime activities more often. To get the best result, we used ggmap() function to see the crime location. After that, we can identity the higher and lower risky region which has the high frequency of the emergency case happening and recognize the time periods with high volume of 911 calls within a day.

It appears that in the Downtown area, most of the incidents occur during the afternoon and evening, but near Upton they mostly occur during daytime, and further away from downtown they mostly occur at evening to nighttime.