Julia Nguyen
Katherine Qian
Youbeen Shim
Catherine Sun

Analysis Process

Alongside the *Baltimore Calls for Service* dataset, three other external datasets were utilized: a *Gun Offender Registry* dataset of gun offenders registered with the Baltimore Police Department, a *Victims* dataset of information on crimes that involved victims, and an *Arrests* dataset of top arrest charges processed at Baltimore's Central Booking & Intake Facility. These paired well with the *Baltimore Calls for Service* dataset since they provided various information on victims, arrests, offenders, weapons, etc, that would have happened at the same date, time, and location of calls.

First, different kinds of questions were asked for data exploration and to delve into the main prompts of the challenge. When and where were these calls and crimes happening? What kinds of weapons were involved in the calls and where? Who were the people arrested and what were their ages, their race? Where were victim crimes happening? What and who were involved?

After coming up with several curiosities, it was time for an abundance of data cleaning and manipulation. All of the data cleaning, manipulation, and visualizations were performed through RStudio with the help of the following packages: *tidyverse, lubridate, stringr, ggmap*. Pipelining was the main technique utilized.

Data cleaning was primarily for merging the datasets as well as plotting the longitudes and latitudes onto the map. The longitudes and latitudes were standardized across all datasets along with the date and time. Once all datasets were cleaned and the external datasets were merged with call (individually rather than all at once), then commenced the exploration and analysis of the aforementioned curiosities with an Exploratory Data Analysis (EDA). This EDA answered some interesting questions and resulted in new ones.

Location was a very important aspect to look at, so heat maps were heavily applied to this challenge. They provided an understanding of which possible approaches by the Baltimore police department and possible implementation of policies could have impacts on specific locations and areas.

The use of more simple visualizations was also helpful where we looked at the number of calls throughout each day to the police, and a bar graph depicting age and race groups that were arrested at the time of calls. For visualizations involving demographics, the data was first standardized into percentages to account for the status quo distribution of the race in the Baltimore population.

When it came to the visualizations, how the audience would perceive the data and interpret the message being delivered was taken into consideration. The type of graph was chosen based on clarity and effectiveness; the types of heat maps were chosen based on the same criteria and on the color and patterns that were easy on the eye.

With the aid of the visualizations, potential ways in which Baltimore police could effectively and efficiently prevent crime and innovative approaches to increase public safety for citizens and police were suggested.

Word Count: 472

Alongside the Baltimore Calls for Service dataset, other external datasets were used: Gun Offender Registry dataset, Victims dataset (crimes that involved victims), and Arrests dataset (top arrest charges processed at Baltimore's Central Booking & Intake Facility). All of the data cleaning, manipulation, and visualizations were performed through RStudio, w/packages tidyverse, lubridate, stringr, ggmap. Pipelining was the main technique utilized; external data was merged on long/lat and date.