Group Members: Daniel Kim, Judy Fu, Kaid'dub Pavel, Vivian Ouzounis

Data and Audience

Depending on the specifics of your project, you should outline the answers to these (and perhaps other) questions:

What is the dataset you'll be working with? Please include background on who collected the data, where you accessed it, and any additional information we should know about how this data came to be.

Link to Dataset: https://data.seattle.gov/Public-Safety/Use-Of-Force/ppi5-g2bj
The data set we will be working with is the "Use of Force" in the Seattle Police Department. This data is a compilation of date and time, incident number, incident type, precinct, sector, and beat (type of beating), officer ID, subject ID, subject race, and subject gender. The data was provided by the City of Seattle by a dataset owner named 6506 Tony Baden. The data was collected in 2017, but has been updated as of late on February 14, 2023.

Who is your target audience? Depending on the domain of your data, there may be a variety of audiences interested in using the dataset. You should home in on one of these audiences.

The target audience is people who are interested in the role of police in communities, police reform/abolition, the effectiveness of the use of force policy, the presence of policing of certain marginalized communities, and more. This data can be used by anyone who is interested in understanding the role of the police in communities and the effects of their actions. If we were to home in on a very specific audience, it would be police reform/abolition groups seeking to prove whether or not the police have too much power and whether or not they're abusing that power.

What does your audience want to learn from your data? Please list out at least 3 specific questions that your project will answer for your audience.

1. Is there a racialized pattern for police's use of force?
2. What precinct contains the highest proportion of BIPOC communities and how is it policed in comparison to non-BIPOC precincts?
3. Is there a correlation between gender and the use of force?

Technical Description

This section of your proposal asks you to think through the specific analytical steps you'll need to complete throughout the project.

How will you be reading in your data (i.e., are you using an API, or is it a static .csv/.json file)?

We'll use a static .csv file.

What types of data-wrangling (reshaping, reformatting, etc.) will you need to do to your data?

The .csv format will be basically ready to use in rstudio, but we will have to:
- Order the data by incident level (level 1, 2, 3) to create a ranking.
- Edit date/time into new columns (create new columns for date and time separately).

What (major/new) libraries will be using in this project (no need to list common libraries that are used in many projects such as dplyr)

We will use the tidyverse package to utilize readr, stringr, magrittr, and ggplot2.

What questions, if any, will you be answering with statistical analysis/machine learning?

We can use statistical analysis to answer the audience questions that we listed above. In particular, we want to see if there is a pattern associated between the use of force in police towards a certain race or gender. We could potentially convert this into meaningful information, such as whether or not police are abusing their powers by attacking minority groups specifically.

What major challenges do you anticipate?

A challenge will be deciding what to plot, and how to best visualize the data for the audience. Additionally, figuring out how to create a well-detailed map of the data in RStudio will be tricky as well.