Dataset: US Police Shootings dataset

(https://www.kaggle.com/ahsen1330/us-police-shootings)

Language/software: Python, Ploty, Dash, Numpy, Pandas

1) **Introduction**: describe your problem: the motivation, the background, the high level questions, etc.

In recent years, police shootings have come to the focus of the public. People looked into the statistics and observed something problematic. “Racism”, “Police Brutality” became some of the most popular topics on social media and political debates during 2020, as people’s attention was drawn to the names “George Floyed”, “Breonna Taylor”, etc.

We would like to take a deep dive into this topic by looking at data from 2015 to present. Is there compelling evidence showing that there is a systemic bias towards a certain demographic? Is the problem more widely present than we realized? And how big of a gap do we have in order to overcome this problem?

We will be exploring these questions through analysis and present our results in an interactive dashboard that can be used by news media, lawmakers, and other stakeholders to understand this subject better.

2) **Detailed analytical questions**: what questions will you ask from the data?

1. Does police shoot a certain demographic more likely than the others?
2. What’s the reason for being shot - were they armed?
3. Which cities have more police shootings than others?
4. Do we see any temporal shooting patterns from January 2015 to June 2020?
5. Does police shoot people of certain races who do not have “threat level” marked as “attack” more often than the others?
6. Does police more likely shoot people whose threat level are not marked as “attack” when his/her body camera is not on?

3) **The data sets**: What are your data sets? Where and how will you collect them?
The main dataset that we will be using is the US Police Shootings dataset (https://www.kaggle.com/ahsen1330/us-police-shootings)
It was found on kaggle and will be downloaded in a .csv file.

4) **Visualization requirement**s: in order to answer the questions in 2), what need to include in your visualization?

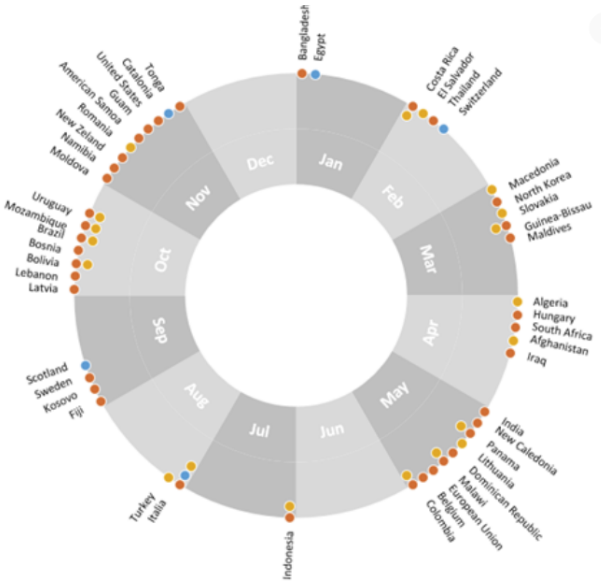
Question 1 & 3:

- Geographical information: city/state +/- other demographic visualization
    - Proportional symbol map. The sizes of the circles represent the total shooting cases in the area. Additionally, animated interaction could be built between

temporal and spatial information so that changes over time across the map could be observed.
    - Heat map. The colors of the region represent the total shooting cases in the area. Additionally, other demographic visualizations could be layer on the top of each city/state to visualize interaction effects. For example, we could use a bar chart to visualize the number of shootings across different races for each state. Then layer each bar chart on the top of the map. Besides, a filter function could support the selection of demographic variables.
- Individual demographic visualization
    - Generate a word cloud to represent the first names of shooted people.
    - Histogram or circles pattern to represent the number of shootings for different age, gender, mental illness state, or race. Interactions could be added between the base map and these individual demographic visualizations.

Question 2:

- Whether they were armed or not:
    - More advanced pie chart like below:



Create a pie chart according to the proportions of “armed”, “unarmed”, and “unknown”. Within the “armed” section, use dots to represent what was being armed. The colors of the dots represent different types of weapons and the size of the dots represent the total number of certain types of weapons.

Question 4:

- Changes of shooting numbers from month to month or quarter to quarter or year to year only:
    - Scatter plot by groups. The x axis is time by month or quarter or year, y axis is the total number of shootings accordingly. The colors of the dots represent different demographic dimensions, like race, gender.

Question 5:

- Population here is narrowed down to the shooted people whose threat levels were marked as “other” and “undetermined”.
    - Bar chart or color pattern. Each bar or bubble represents one race. The height of the bars or size of bubbles represent the total number of shooting cases accordingly in this specified population.

Question 6:

- Population here is narrowed down to the shooted cases in which the police’ body camera is off or not equipped.
    - Bar chart or color pattern. Each bar or bubble represents the categories of “threat level” - either “attack” or “other” and “undetermined”. The height of the bars or size of bubbles represent the total number of shooting cases accordingly in this specified population.

5) **Software**: what software will you be using? You are asked to develop a **dashboard** like product that includes all visualization

Python, Ploty, Dash, Numpy, Pandas

6) **Evaluation:** how will you evaluate your visualization system?

The visualization system we created is complete and well-organized with evaluating all potential factors that can cause the bias of potential shooting by the police. Moreover, all visualizations are clearly presented in the dashboard with the labels that are easy for people to understand.

7) The **team members** and **work distribution**: who are in your team and how the work will be distributed among the team members?

Team members: Fiona Fei, Edison Gu, Shay Sun

Work distribution: The data has been well cleaned and pre-processed. Each of the team members will take two of the questions and make visualizations accordingly. All the individual pieces will be organized into a dashboard at the end. Interaction effects will be added either in the process of making individual visualizations or after the dashboard being built.