

CS 3300 Project 1 Final Report

Amanda Ong (ayo3), Annie Zhang (zz229), Billy Zhu (bz83)

- a) Our data describes all the police killings in the United States for the year 2015. We retrieved our data source (in a CSV) file from fivethirtyeight's Github repo: <https://github.com/fivethirtyeight/data/tree/master/police-killings>. The dataset provided by fivethirtyeight contained a lot of different variables. For the purposes of our project, we did not need to address each variable as it would have created too many graphs with too much information. When brainstorming which variables to focus on, we came to consensus that race was the most important one. Too often have we heard about instances of police abuse towards minority; more specifically, towards unarmed minorities. We felt armed was a necessary variable to focus on since a police officer would not have a legitimate reason to kill an unarmed person. We first made graphs for each separate variable. However, seeing how race is such a controversial topic and the whole story needed to be told, we decided to create another graph joining the two variables. The initial data set contained "unknown" values to both the race and armed variable. However, they were in such few frequencies that we felt best to omit these results as it would barely be noticeable next to other categories with a higher count.
- b)
- i) For Graph 1: "Number of people killed by police officers by race", to map the data to the pictogram, I looped through the CSV file to create an array with elements containing the following headings: Race, TotalDeaths, and MaxIndex (to keep track of when to switch colors on the people pictogram). To create the corresponding text labels, I used a linear fontColorScale and a linear sizeScale to display the texts with the corresponding colors and font size. The first text was the largest and the last text was the smallest. Code taken & modified from below links
 - <http://bl.ocks.org/alansmithy/d832fc03f6e6a91e99f4>
 - <https://stackoverflow.com/questions/49040959/how-to-color-specific-parts-of-text-div-in-d3>
 - ii) For Graph 2: "Was there a Threat?", to map the data to visual elements, I first extracted the data from a CSV file using the nest() and rollup() function to count the frequency of each armed category. Having grouped the categories in this variable, I proceeded to use scale functions to help me create my axes and labels. For the y-axis, our frequency counts were not too large so I was able to do a simple linear scale. After looking at some examples for categorical data, I realized I could use the scaleBand() function to help me achieve my results. For the color, I also proceeded to use an ordinal scale to give each bar a distinct color. The data() and enter() functions allowed me to neatly map each value from my array onto the graph. The following link helped us with producing a bar chart: <https://bl.ocks.org/mbostock/3885304>
 - iii) For Graph 3: "Weapons Held At Time of Death, By Race", to map the data to visual elements, I first used the nest(), key(), and rollup() functions to extract the

variables I needed, race and armed status, and the corresponding counts from the data. After these extractions, I included some additional preprocessing to convert the data and counts into a dictionary representing the relative frequencies for each category of armed status for each race. At this point, I could finally proceed with actually creating the graph. Since we wanted to create a horizontal comparative bar graph, I used `scaleBand()` for the y-axis and `scaleLinear()` for the x-axis. I also used an ordinal scale for the color of the different armed categories. Finally, the `data()`, `enter()`, and `stack()` features allowed me to neatly construct the bars for my bar graph. The following link was very helpful for me: <https://bl.ocks.org/mbostock/3886208>

- c) Our visualization investigates what motivated police officers to fatally harm others. Graph 1 shows the number of people killed by police officers based on race. Surprisingly, though we expected police officers to be affected by racial profiling, we found that police officers killed white individuals most frequently.

We then zoom into 2 possible reasons for these findings: Were people killed because they were armed (Graph 2: Was There a Threat?) or were people killed because they were armed *and* because of their race (Graph 3: Weapons Held At Time of Death, By Race). As shown by Graph 2, the majority of people killed possessed a firearm, indicating the police officers killed with a good reason. It is important to note that the number of unarmed individuals that were killed was still surprisingly high. This begs the question: Under what circumstance would an officer kill an unarmed person? However, surprisingly, the next largest group were people who were killed and unarmed.

We investigated that finding in Graph 3. By comparing the length of the red firearm bar and the blue not-armed bar, it is shown that black, Hispanic, and Asian people were more likely to be killed than white people, indicating that there was racial profiling that affected a police officer's decision to shoot. This finding supports the media coverage for police brutality towards minority groups. However, it is important to note that though the group of unarmed Asians stands out in Graph 3, it is not conclusive as the sample size of Asians was limited to 10 people.