Effect of Race on US Police Killings

Prepared by

Alex Creighton, Brenna Harris, Noah Stracqualursi

University of Vermont

December 8, 2020

Table of Contents

Abstract……2

Introduction……2

Data Description……4

Descriptive Statistics……6

Methods and Results……11

Body Cameras and Race……11

Location and Race……13

Conclusion……17

Limitations and Recommendations……20

Abstract

Race is one of the most divisive and highly-discussed topics in the United States, and the sub-topic of police shootings is largely at the forefront of the race topic. Our research seeks to gain a deeper analysis in the issue of police killings, looking to gain insight on the question: "Is race a factor affecting the killings of US citizens by police?" Three of the facets of the dataset are analyzed - the rate of shootings based on the race of victims, the body camera usage by police officers during police killings, and the location of the killings. The data being studied offers insights into police killings of suspected criminals/felons. We utilized a dataset that was compiled in order to gain "insights and analyze the story around racism in America". The dataset contains variables which describe the situation and background of each police killing of the individuals listed. We evaluated these variables individually as well as in correlation with other variables, in order to determine whether race potentially affects police killings. Overall, this study found that there is the potential for race to be a determining factor behind police killings, however, we cannot confidently conclude that this is true.

Introduction

Throughout the history of the United States, the treatment of minorities - particularly black people - can be considered questionable, to put it lightly. In recent years, this treatment has been one of the most scrutinized and divisive issues facing the country. Specifically, the seemingly overwhelming number of police killings of black people over the past decade (and more) has been considered, discussed, and protested more than ever in the past few years. Our

research seeks to take a deeper analysis into this issue, utilizing the statistical tests and methods learned in this course to evaluate data gathered from several sources regarding police killings. The dataset used to conduct this analysis is described below in the Data Description section.

Overall, this research seeks to gain insight on the question: “Is race a factor affecting the killings of US citizens by police?” In order to evaluate this question, three of the facets of the dataset are analyzed - the rate of shootings based on the race of victims, the body camera usage by police officers during police killings, and the location of the killings. While there are many other elements that affect the analysis of police killings and their relation to race, based on the dataset that was utilized and the prominent aspects of this issue that are discussed, we believe that the three facets that are analyzed have the ability to offer a deeper insight into the effect of race on police shootings.

Through the use of logistic regression - including chi-squared and Goodman-Kruskal tests - and a variety of tables, charts, and plots showing the distribution and proportions of the aspects of police killings being studied, we conducted research and analysis on the dataset and overarching issue. Plots displaying the proportions of certain factors, such as the manner of death, race of the victim, and whether they were armed, allow the ability for initial analysis of whether race affects police killings. Further logistic regression testing offers the ability to model the probability of a certain class or event existing; in this case, the probability that race has an effect on police killings, using factors such as body camera use and location to assist in the analysis of the issue.

The following section describes the dataset that is used; it details how the data was obtained, the way in which data was collected by the owner of the dataset, and descriptions of the variables within the dataset. Then, the next section provides descriptive statistics. This section includes counts and frequency tables to offer initial analysis on individual variables from the dataset. Following the descriptive statistics are the methods and results, displayed in one section. Methods and Results describes the methods utilized to analyze the data, including plots and tables, as well as the results achieved and the analysis of such results. Following that, we conclude, restating the overall question driving our research and discussing how the research was able to provide insight into the issue. Lastly, the Limits and Recommendations section highlights any limits that may affect the study and its results, and offers advice and general thoughts regarding future studies analyzing the effect of race on police shootings.

Data Description

The data being studied (obtained from https://www.kaggle.com/ahsen1330/us-police-shootings) offers insights into police killings of suspected criminals/felons; it contains variables which describe the situation and background of each police killing of the individuals listed. The data was obtained via Kaggle - the owner of the dataset, Ahsen Nazir, was inspired by the news of police killings, particularly in the earlier period of 2020, to gather data on the subject to gather "insights and analyze the story around racism in America". Nazir collected data from various sources on Kaggle regarding police killings from January 2015 to July 2020.

The raw data was downloaded and any incorrect values or missing information was filled in using other datasets. Additionally, the data was prepared for visualization, normalization, and categorization, as the goal of compiling this dataset was to provide an easily understandable overall insight into police killings. The data offers the ability to evaluate cases of police killings on multiple levels, including race and mental illness. Each variable in the dataset is included below:

Variable	Description	Response
id	Identification number of victim	Number
name	Name of victim	Name
data	Date of shooting	Date (Month/Day/Year)
manner_of_death	How victim was killed	Descriptive word (shot, shot and Tasered)
armed	Whether victim was armed	Descriptive word (gun, unarmed, knife, etc.)
age	Age of victim	Age (Number)
gender	Gender of victim	M/F
race	Race of victim	Race (Black, White, Hispanic, etc.)
city	City of occurrence	City name
state	State of occurrence	State abbreviation (51 states, Washington D.C. included)
signs_of_mental_illness	Signs of mental illness in victim	TRUE/FALSE
threat_level	Threat level of victim on officer	Descriptive word (attack, undetermined, other, etc.)
flee	Whether victim was fleeing	Descriptive word (Not fleeing, Car, Foot, etc.)
body_camera	Whether officer obtained body camera footage	TRUE/FALSE
arms_category	Category of weapon held by victim	Brief description of weapon

Table 1: Variables from dataset including description and response (way in which variable is answered in dataset).

During our analysis of this dataset, we focus on a few of these variables to evaluate. Specifically, we focus on the victims' race, whether they were armed, police body cameras, and location. We surmised that these variables have the most potential to analyze whether race is a factor in police killings; our reasoning and analysis will be discussed in the coming sections.

Descriptive Statistics

The dataset that was obtained for the purpose of this study contains exclusively categorical variables, therefore we were not able to calculate means or variances. In order to provide an initial insight, we determined the counts of several variables and plotted their frequency, with the purpose of displaying the proportions of each of them. This allows us to analyze the variables individually and determine conclusions that may inform the methods we use and results we achieve in the following section. We did not include the chi-squared correlations of variables in this section, as some such calculations are done in the Methods and Results section. Below shows the frequency plots for manner of death, armed, race, state, threat level, and body camera.

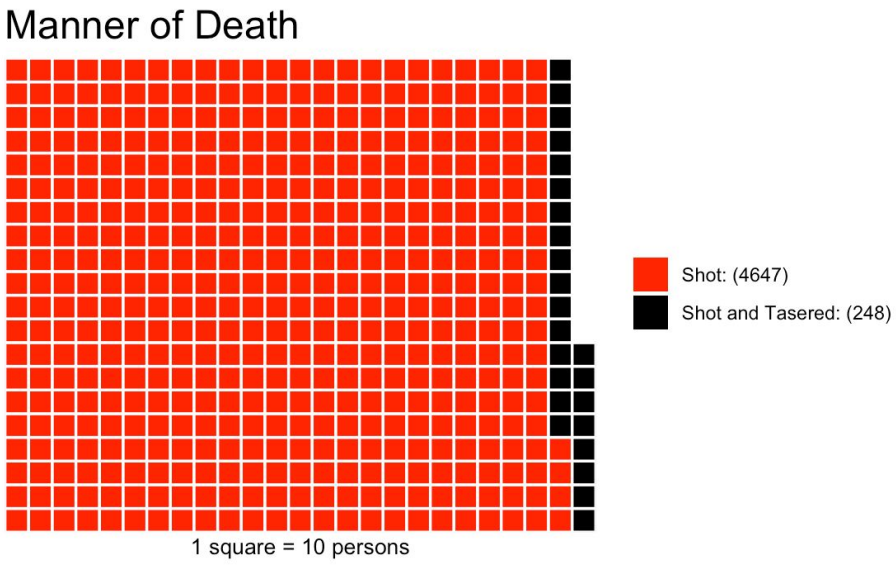

Figure 1: From the dataset, 4,647 of the victims were shot and 248 were shot and tasered by the police.

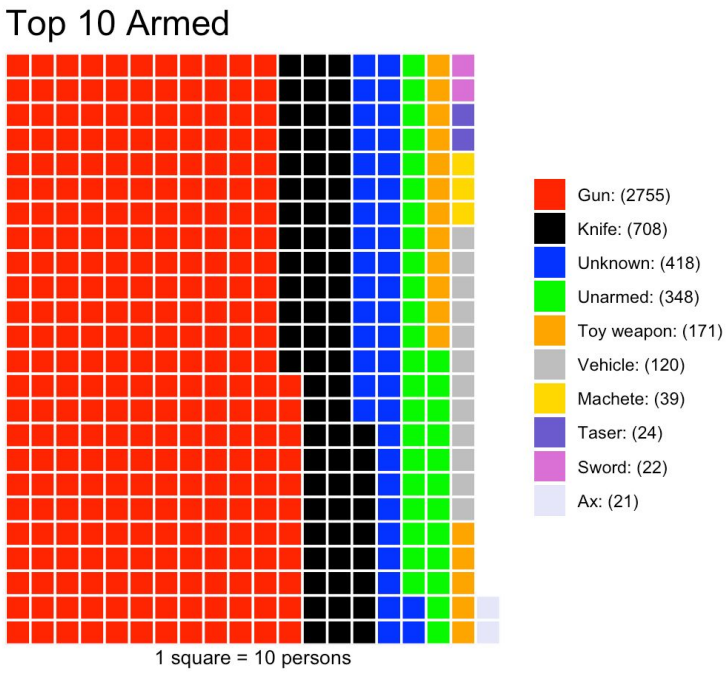

Figure 2: This table displays the top 10 frequencies of weapons/items that the victims were "armed" with when they were killed by the police.

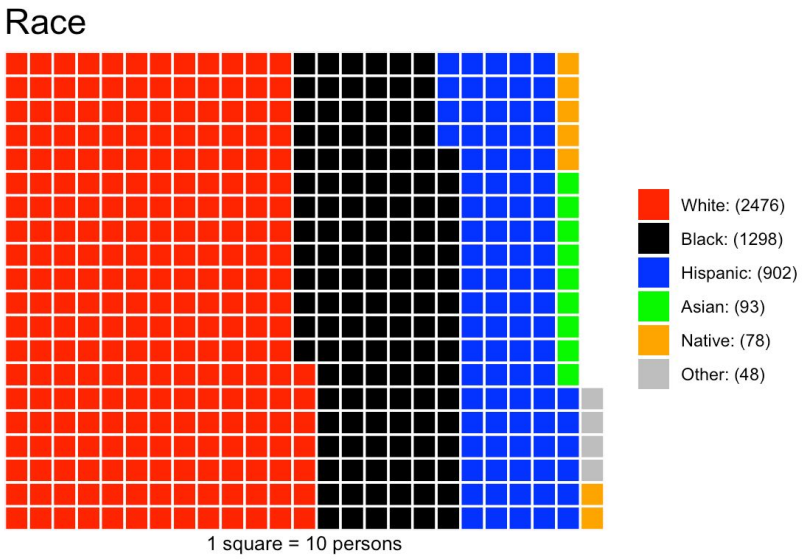

Figure 3: The visual breakdown of the race of each of the victims killed by the police.

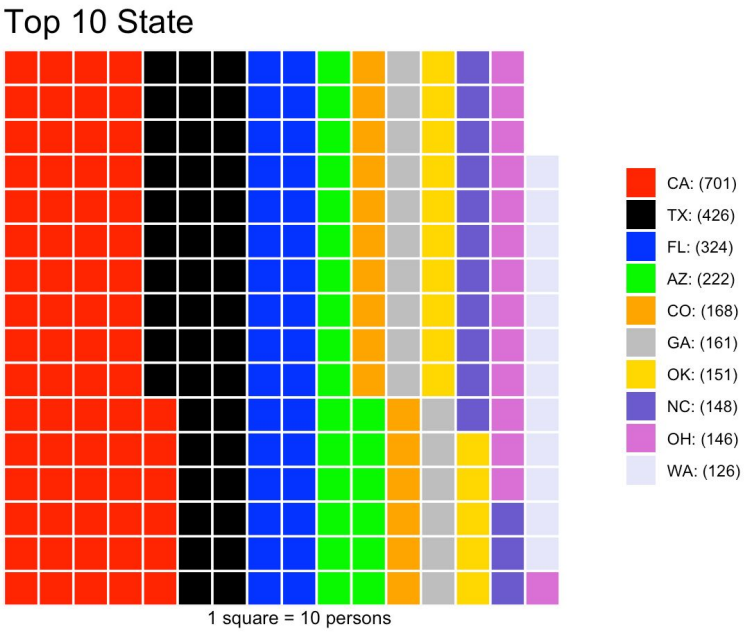

Figure 4: The representation of the frequency of killings by police in the top 10 states with the most occurrences.

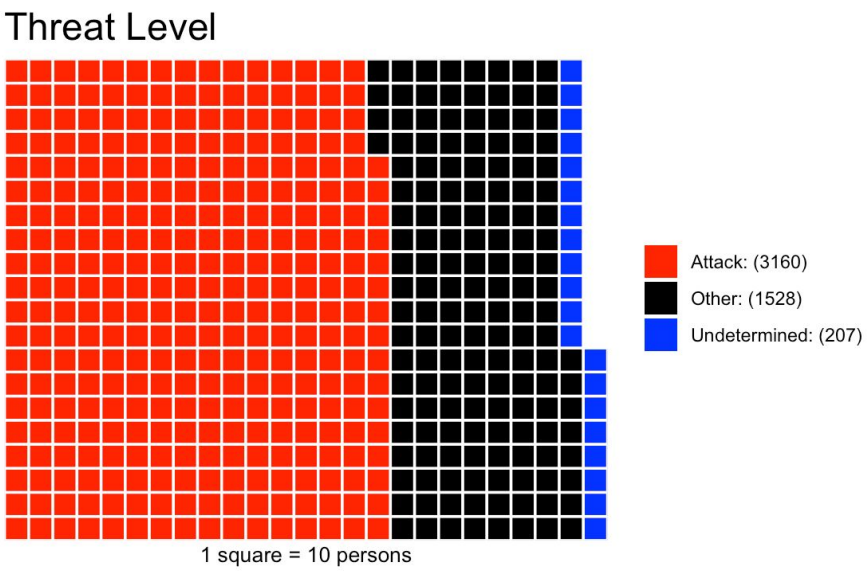

Figure 5: This displays the frequencies of the reported threat levels of the victims that were killed by the police.

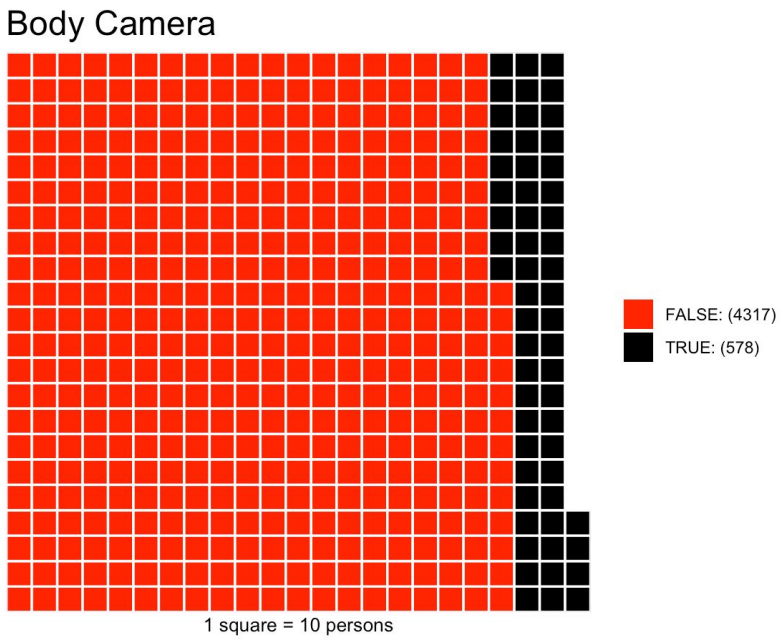

Figure 6: This plot displays the frequency that a police officer had body camera footage of the incident when they killed the victims. (FALSE: no body camera footage, TRUE: body camera footage)

Each of these plots displays the frequencies of certain factors during the occurrences of police killings from the dataset. These plots are a visual representation of the proportions of each of the potential aspects of the variables that are listed. These frequency plots allow for us to analyze each of the variables individually, however, further testing will provide more in-depth insight into the issue. Such testing includes analyzing the effect of variables on each other, correlations, and proportions. Tests for further analysis are included in the Methods and Results section. The main findings from initial evaluation of the variables individually are included below.

The large proportion of victims reportedly had weapons on their person when they were killed by police, and only 7.11% of the victims were reportedly unarmed (Figure 2). We found that more than half (50.58%) of the victims killed by police were white, whereas only just over a quarter (26.52%) were black and less than half (48.44%) were people of color overall; these percentages exclude the “Other” category (Figure 3). Regarding the breakdown of states (Figure 4), California had the most occurrences of police killings and had a greater frequency over the second-most state (Texas) by a factor of 1.65; six of the top 10 states with occurrences of police killings were southern states (California not included). In 65.56% of the occurrences police officers reported that they were attacked by the victims (Figure 5), however, the large majority of these reports cannot be substantiated, as figure 6 shows that 88.19% of the occurrences did not have body camera footage of the incident.

Methods and Results

Body Cameras and Race

The dataset includes six race types who were killed as a result of police shootings: Asian, Black, Hispanic, Native, White, and Other. Regardless of whether it was "necessary" for the police officer in each instance to shoot and kill the victim, we sought to determine whether the victim's race had any relation to whether or not the police officer had their body camera on. To start, we created a bar chart using the counts from each race and whether or not a body camera was turned on.

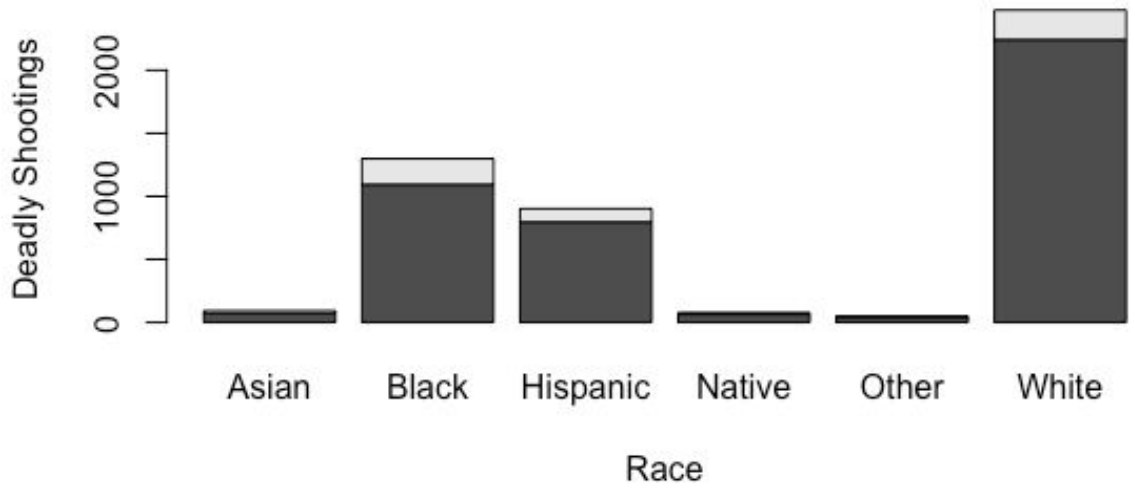

Figure 7: Number of deadly shootings by race; black shading represents body cameras off, white shading represents body cameras on.

Based on the plot in Figure 7, the proportion of body cameras on appears to be similar among Black and White people, but there is almost twice the number of deadly

shootings of White people rather than Black people. It is difficult to determine anything meaningful from this figure alone, so we added this data into a table and performed tests to analyze further.

Race/Camera	Asian	Black	Hispanic	Native	Other	White
FALSE	77	1094	796	66	42	2242
TRUE	16	204	106	12	6	204

Table 2: Table of police shootings with the race of the dead as column headers and whether or not body cameras were on during the shooting as the row headers.

Looking at Table 2, we evaluate the counts of officers who had body camera footage and those who did not, divided by each race from the data. To see if the proportions are equal for each race, we first used a chi-squared test of independence. From the chi-squared test, we found a test statistic of 35.836 with 5 degrees of freedom. This corresponded to a p-value of 1.024 * 10^-6, or approximately 0, meaning that we reject the null hypothesis and conclude that at least one of the proportions is different. Following the chi-squared test, we performed the Goodman-Kruskal's Gamma test to see if there is a difference in the proportions. From this, we obtained a gamma value of -0.204 which translates to a z-score of -3.833. Using that z-score, we found a p-value of 0.000126, therefore we would reject the null hypothesis and conclude that there is a difference in proportions of body camera footage versus no body camera footage by race.

Finding low p-values in both the chi-squared and Goodman-Kruskal's Gamma tests, we are able to conclude that the proportions of body camera footage are different by race.

This suggests that there is some factor that affects the use of body cameras by race during deadly shootings. This conclusion leads us to consider the possibility that the factor in question may be race, however, we are unable to determine whether that is the case.

Location and Race

The data that was utilized provided us with the location in which each shooting occurred. We sought to determine whether the usage of body cameras in police shootings differed by state. To do this, we created a data frame of shootings by state and created a frequency variable for the number of shootings that occurred in each state. To understand the data better, we created the visuals below.

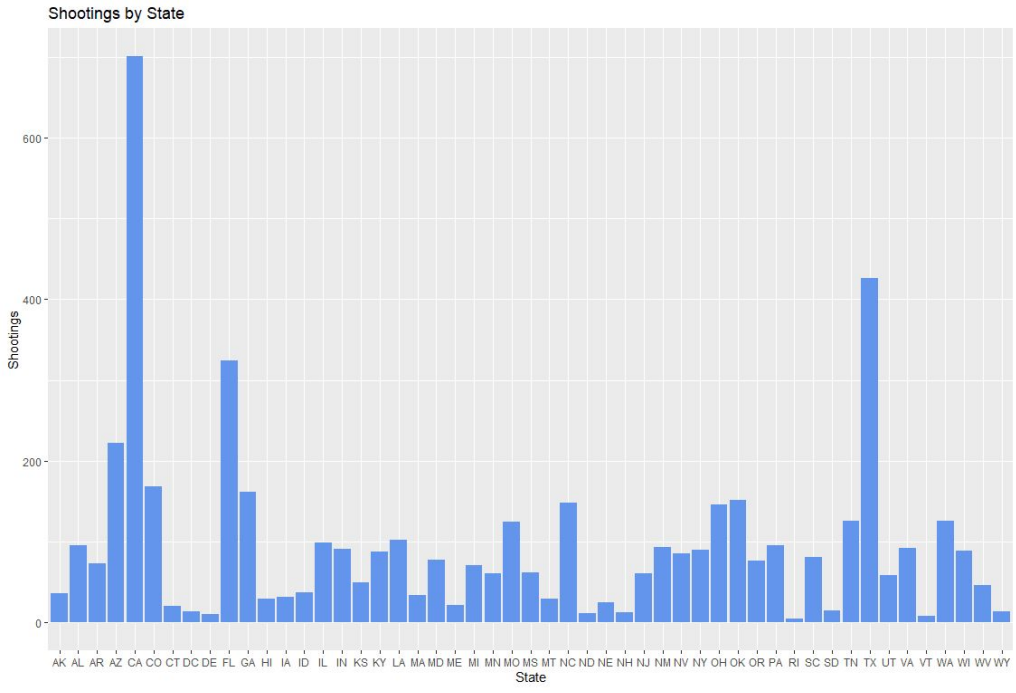

Figure 8: Total number of shootings in each state from January 2015 to June 2020

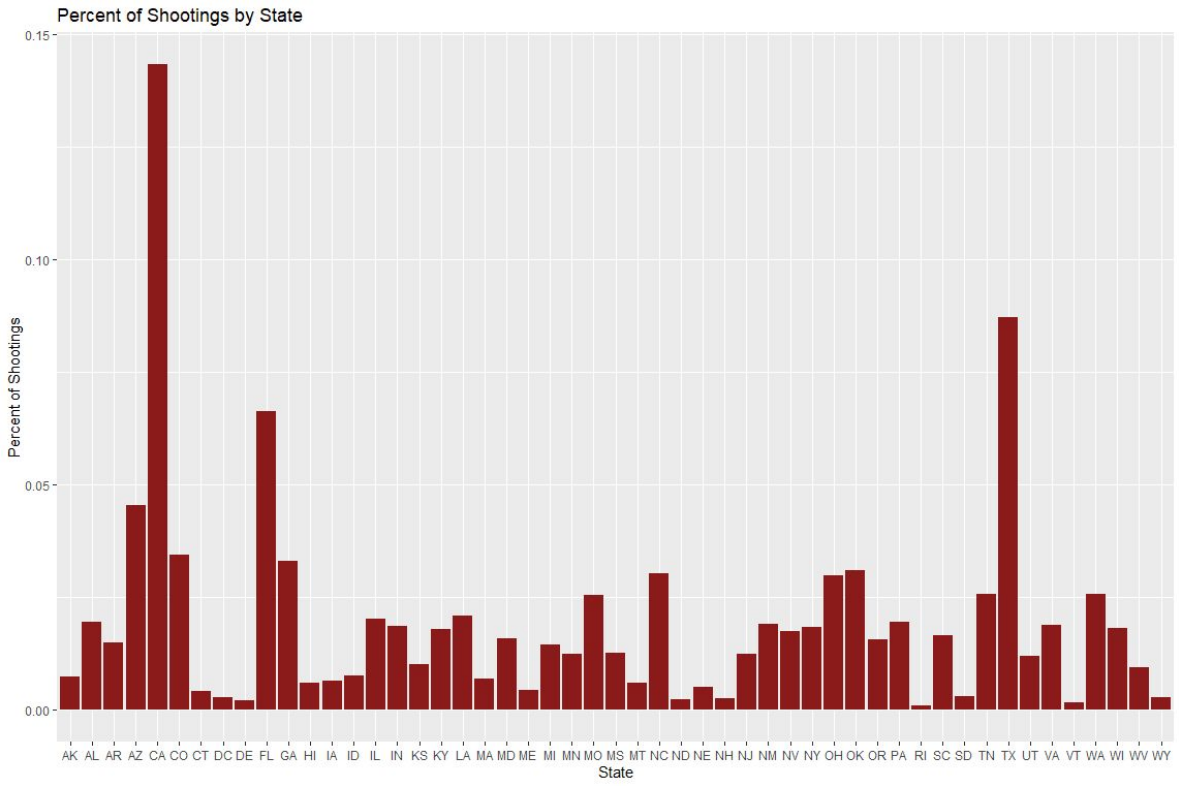

Figure 9: Percent of US police shootings that occurred in each state from January 2015 to June 2020.

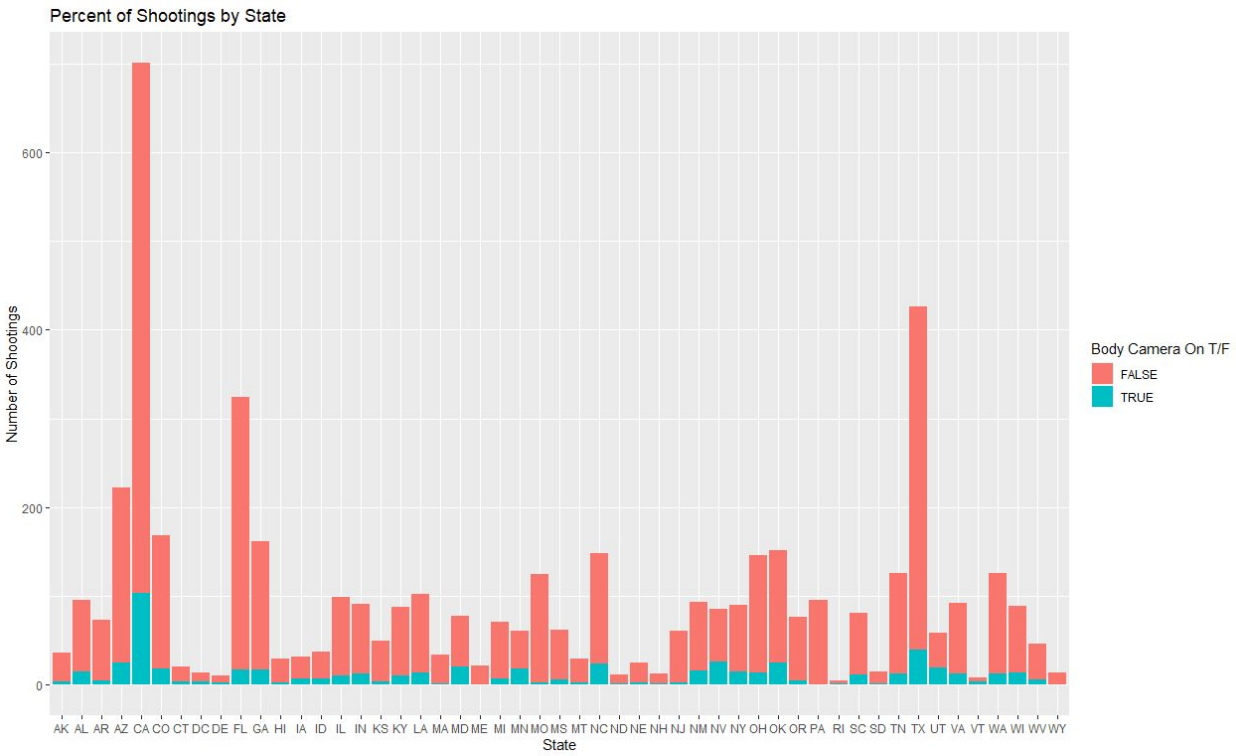

Figure 10: Total number of shootings in each state from January 2015 to June 2020, categorized by whether the officer's body camera was on or off.

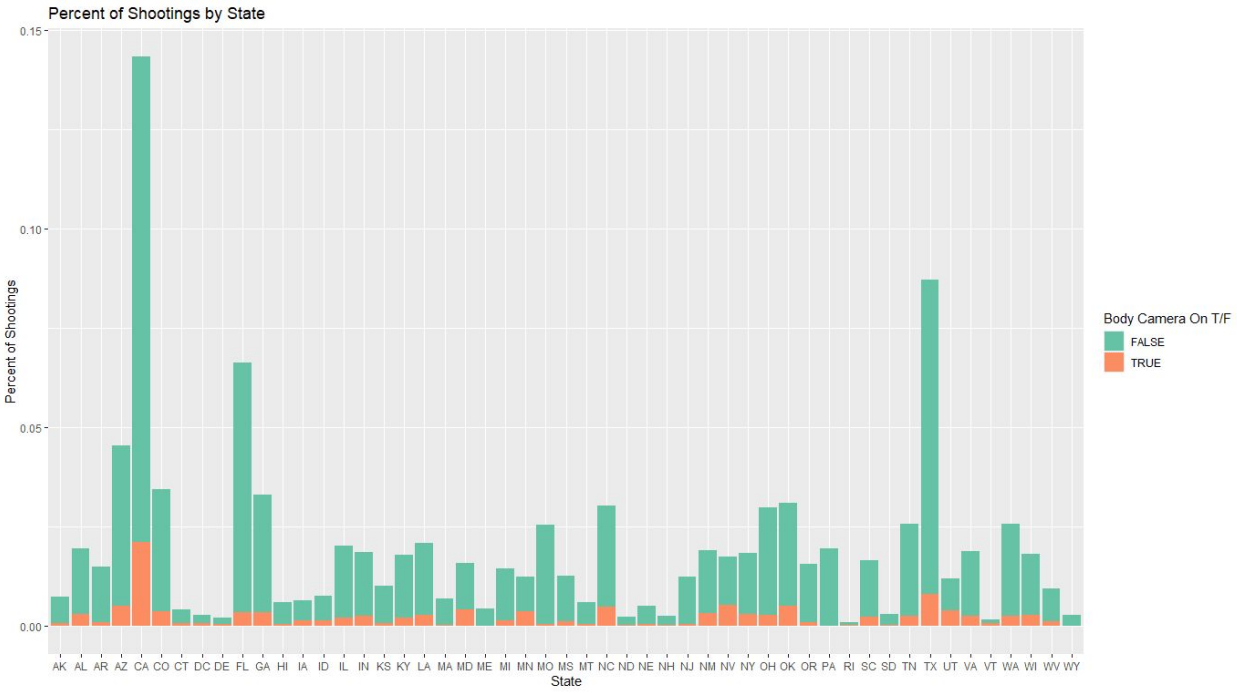

Figure 11: Shootings in each state (as percentage) of total police shootings in the US from January 2015 to June 2020, categorized by whether the officer's body camera was on or off.

Figure 8 displays the total number of shootings that occurred in each state, which shows that the most police killings took place in California. There were 701 police shootings in California, which accounts for 14.32% of the total police shootings in the United States.

After looking at the total number of shootings that occurred in each state, we looked at the proportion of shootings in each state where the officer obtained body camera footage. Figure 10 shows the total number of shootings in each state, categorized by whether or not the officer obtained body camera footage of the incident, and Figure 11 shows us the shootings (categorized by camera on or off) in each state as a percentage of

the total number of shootings in the country. After filtering the data to change the body camera variable into a binary variable, we gained more insight about the shootings that occurred in each state and the United States as a whole. There were 578 shootings that occurred when the officer's body camera was on. When compared to the total number of shootings in the country, only 11.81% were recorded on a body camera. There were three states (Maine, Pennsylvania, and Wyoming) where no shootings occurred on a police camera; of these three states, Pennsylvania had the highest number of shootings (95). The highest percentage of shootings that occurred on camera was in Vermont at 37.5%. Although this is a high proportion, the state had 8 shootings which is the second lowest number of shootings to occur in a state, so the small sample is inflating the percentage.

Our findings when analyzing the location of police killings, along with the proportion of body camera footage obtained in each of the states, allow us to conclude that the proportion of body camera footage does vary by state. The results indicate that body camera footage is, overall, not obtained frequently, but that certain states have higher rates of obtaining body camera footage from police shootings than others. We may be able to note that many of the states with lower percentages of body camera footage are states with historically higher percentages of conservative residents, however, we are unable to confidently determine that this is a factor determining the results.

Conclusion

In recent years, the treatment of minorities, and black people specifically, has been one of the greatest issues facing the United States. The number of police killings of black people over the past decade (and more) has been scrutinized more than ever in the past few years. Our research seeks to take a deeper analysis into the issue of race and police killings, utilizing various statistical tests and methods to evaluate data gathered from several sources regarding police killings. Overall, we sought to gain insight on the question: "Is race a factor affecting the killings of US citizens by police?" To evaluate this question we analyzed the rate of shootings based on the race of victims, the body camera usage by police officers during police killings, and the location of the killings. Analysis of these facets of the data do not allow us to confidently answer the overarching question, however, we believe that the facets that are analyzed have the ability to offer a deeper insight into the effect of race on police shootings.

From the Descriptive Statistics we found information for certain variables from the data individually. We were able to see that the large proportion of victims reportedly had weapons on their person during the time that they were killed by police, and only 7.11% of the victims were reported to be unarmed. Additionally, more than half (50.58%) of the victims killed by police were white, whereas only just over a quarter (26.52%) were black and less than half (48.44%) were people of color overall, excluding the "Other" category. When analyzing the distribution of states, we notice that California had the most occurrences of police killings and had a greater frequency over the second-most state (Texas) by a factor of 1.65; six of the top 10 states with occurrences of police killings were southern states (California not included). We found that in 65.56% of the occurrences

police officers reported that they were attacked by the victims, however, these reports are largely unable to be confirmed as 88.19% of the occurrences did not have body camera footage of the incident.

In the Methods and Recommendations, we first analyzed the impact of race on whether or not the police officers obtained body camera footage of the incident. The bar chart indicated that the proportion of body cameras on appears to be similar among Black and White people, but there is almost twice the number of deadly shootings of White people rather than Black people. We then created a table to analyze each race and their correlation to body camera footage, using the chi-squared test to obtain the test statistic and p-value, and Goodman-Kruskal's Gamma test to evaluate the difference in proportions by race. The low chi-squared p-value of approximately 0 meant that we reject the null hypothesis and conclude that at least one of the proportions is different, and the p-value obtained from the test statistic in Goodman-Kruskal's Gamma test was 0.000126; we again reject the null hypothesis, and conclude that there is a difference in proportions of body camera footage versus no body camera footage by race. The overall results suggest that there is some factor that affects the use of body cameras by race during deadly shootings; this factor may be race, but we are unable to determine whether that is true.

We then analyzed the location of each of the shootings, along with the use of police body cameras in each of those states and occurrences. When evaluating the figures for location, California attains for the largest number and percentage (14.32%) of police killings from the dataset. After looking at the frequency of shootings in each state, we

looked at the proportion of shootings in each state where the officer obtained body camera footage. We filtered the data to change the body camera variable into a binary variable to gain more insight about the shootings that occurred in each state and the United States as a whole. This displays that 578 shootings (11.8%) occurred in the United States when the officer obtained body camera footage. Three states (Maine, Pennsylvania, and Wyoming) had no instances of shootings occurring on a police camera; of these three states, Pennsylvania had the highest number of shootings (95). The highest percentage of shootings that occurred on camera was in Vermont at 37.5%, although the state had 8 shootings - the second lowest number of shootings to occur in a state. While the figures display that many of the states with lower percentages of body camera footage are states with historically higher conservative resident percentages, we cannot confidently conclude that these directly correlate.

Overall, the data indicates that there is the potential for race to be a determining factor regarding police shootings. After evaluating the frequencies of certain variables throughout the occurrences of police killings, analyzing the use of body cameras during shootings, and assessing the locations of both shootings and the use of body cameras, we are able to conclude that there are outside factors that affect police killings. However, the fact that there are outside factors does not allow us to confidently conclude that race is the main factor among them. While there is evidence to suggest that race is the factor driving the differences among variables, further statistical testing has to be done to numerically make such a conclusion.

Limitations and Recommendations

The possible limitations of this study largely do not lie within the statistical analysis of the data, but rather in the data that was utilized. The issue of police killings is greatly controversial, therefore there are many potential factors that affect whether the data and its analysis display an accurate representation of the issue. The lack of quantity of non-deadly shootings provides a lapse in analysis of the issue - while the analysis of police killings does offer the ability to conduct analysis and draw conclusions, police shootings as a whole may lead to different conclusions to be drawn.

In regards to body cameras, there are several factors that might affect this variable. The lack of body cameras in instances of police killings leads us to interpret that factor a certain way, however, lack of use may be due to police budgets and not having cameras at all, broken body cameras, or cameras being purposely turned off. Therefore, a variable such as body camera use allows conclusions to be drawn, but a lack of information provides a limitation to such conclusions.

An overarching possible limitation to the study is the police reporting of these instances. Instances of police killings may involve witnesses, as we have seen with the death of George Floyd, however, often the only people who truly know what happened during these situations are the police officer and the victim. Since the victims of this study were all killed, their account of the story is not able to be heard, so the only reporting of the situation is done by the officer who killed them. If race truly was a factor in that killing, that officer may potentially give a false reporting of the entire situation, which provides a limitation to the data.

Lastly, one limitation to the statistical analysis is the data being exclusively categorical. We were able to conduct appropriate tests and obtain charts and plots to properly analyze the data regarding police killings; however, the lack of quantitative data meant that we were not able to obtain insights such as means and variances, nor were we able to conduct certain tests and display plots. The lack of quantitative data led to the ability to exclusively analyze based on frequencies and proportions, which may limit the overall analysis.