Statistical Analysis on Police Killings to Generate Inferences and Find Trends Regarding Fatalities

Drake Wagner, Sofia Mohammed, Rachel Wells, Kayleigh Zigler

Table of Contents

ABSTRACT

The topic of police killings has been an issue recently noted in the media more than ever. In 2019, 1,004 people were killed by the police. Of these 1,004 people, 370 were white, 235 were black, 158 were Hispanic, and 241 listed as unknown or other (Statista Research Department, 2020). These statistics led us to choose to research data of police killings. The purpose of our study was to see if we could find trends in different variables linked to being shot by the police. The data that we selected is a representative sample of police killings in the United States in 2015. Important variables as noted in this research paper include Age, Ethnicity, Unemployment rate, College, and if they were armed or not. To analyze the data, we came up with four research questions. These include "Is there a relationship between the share of the over 25 population with a bachelor's degree or higher and the tract level unemployment rate," "Is the average age of people killed by police officers less than 30 years," "Is the proportion of blacks killed while unarmed greater than whites," and "Is there a significant disproportion in which races are killed by police officers." Using regression, statistical inference, and other statistical methods we found that the average age of people killed by police officers is less than 30 years, there is a moderately negative correlation between the unemployment rate and the share of people with a bachelor's degree or higher, there is not enough evidence to claim that the proportions of black and white unarmed people are different, and that there is a significant difference in the proportion of races affected by police killings. Through our data analysis, we were able to see different factors that correlate with police killings in 2015 and were able to make the data relevant to today's society.

INTRODUCTION

The data used throughout this project is titled "Police Killings" and was created by the organization FiveThirtyEight. The dataset contains 465 values and incorporates data from the guardian

and U.S. census that was calculated in the 2015 5-year American Community Survey. Initially, the intention of this dataset was to analyze the locations of police killings, but this dataset includes much more, such as age, race, gender, and armed. As we examined this dataset we were intrigued by these key variables, especially due to the present-day relevance of police killings. Many movements have risen due to trends in police killings and thereby sparked our interest in four research questions. These questions include, "Is there a relationship between the share of the over 25 population with a bachelor's degree or higher and the tract level unemployment rate," "Is the average age of people killed by police officers less than 30 years," "Is the proportion of blacks killed while unarmed greater than whites," and "Is there a significant disproportion in which races are killed by police officers." We used many techniques to evaluate these questions such as hypothesis testing, confidence intervals, regression, data wrangling, and data visualization.

RESEARCH QUESTIONS

Is the average age of people killed by police officers less than 30 years?

Based on the data of Police Killings in 2015, the mean age of people who were killed by police officers was 22.25 years old. As seen in Figure 1, we created a box plot that entails that the mean average is less than 30 years of age as the median age was 20 years old through data visualization. We wanted to further analyze this data to figure out the true mean interval and to hypothesize that on average, people who were killed are less than 30 years of age. To conduct this, we need to do a one-sample confidence interval for means and a one-sample hypothesis test for means. Our null hypothesis was that the mean age of people who were killed were 30 years of age. Our alternative hypothesis was that the mean age of people who were killed is less than 30 years old. Within the data set, we assume that the variables are independent from each other and hold all other variables fixed. We also assume that the sample is representative of the entire population and is randomly selected. As some ages were unknown resulting in

a few outliers, we were able to omit them through data wrangling by creating a new vector. After conducting a confidence interval for means, we can conclude that we are 95% confident that the true age mean of people who were killed by police officers is between 21.09449 and 23.40659. Based on the results of our hypothesis test for means, our test yielded a p-value of 2.2e-16. As our p-value is less than 5% and is small, we had enough statistical evidence to reject the null that the average age of people killed by police officers is less than 30 years. These results are prevalent in our society as we see many young individuals being killed by police officers. Based off of the Los Angeles Times, "The early 20's are a particularly dangerous time for young men..."(Khan, "Getting killed by police is a leading cause of death for young black men in America").

Is there a relationship between the share of over 25 population with a bachelor's degree or higher and unemployment rate?

In today's society, we are taught that a college degree will open doors that cannot be comparable to those who do not have a college education. Based on the U.S. Bureau of Labor Statistics, people with less than a high school diploma had the highest unemployment rate of 7.4% (Chen, 2017). As education increased, the unemployment rate lowered. We thought the police killings data would be a perfect way to test how much of an effect the share of the over twenty-five population with a bachelor's degree or higher can have on a tract-level unemployment rate. We also wanted to see if we could use the share of the twenty-five and older population to predict the unemployment rate. We decided to first conduct a scatter plot, seen in figure 4 with a line of best fit to highlight the relationship. This graph displays that as the share of college degree holders increased, the unemployment rate decreased. To test the correlation we produced a correlation coefficient that gave us the value -0.474. Based on this correlation coefficient value, there is only a moderately negative correlation between the two variables. We

wanted to further explore this relationship by creating a simple linear regression. Creating the linear regression summary, we were able to see that the y-intercept is 0.164 and the slope for the college variable is -0.205. We then looked at the overall p-value, which we found to be far less than 5% as it is 2.2e-16 and the p-value for the college variable is 2e-16. We could then use these findings to make a linear model to make a prediction for the unemployment rate when the college rate is 0.3. We found based on our linear model if the share of the 25 and overpopulation with a bachelor's degree or higher is 0.3, then the unemployment rate is predicted to be 0.103. Unfortunately, our regression model showed that the multiple r-squared value is only 22.28%. This coefficient of determination means that our prediction is only 22.28% accurate, which is lower than what is preferred for observational studies and accounts for 22.28% variance. Overall, the relationship between obtaining a college degree and unemployment is weaker than anticipated based on our data.

Is the proportion of blacks killed while unarmed greater than whites?

In the past decade Black Lives Matters, a large activist movement, has organized many gatherings to protest police killings of unarmed individuals. According to the United States Code of Federal Regulations Title 10, police officers should only use deadly force if the force is in self-defense in which the officer believes they are in imminent danger, to prevent death or serious bodily harm when it's to protect the security of nuclear weapons, and for apprehension when escaping with the uses of a weapon or explosive (United States, Dept. of Energy 999-1000). With the law in mind, there should be very few reasons to shoot at an unarmed individual, however, it does occur, and with the increase in protests, the issue has received a lot of media coverage especially when the killed individual is not white. Many cite police racism and bias leading to unarmed black being killed. For this reason, we decided to examine if there is a statistical difference in the proportion of white individuals killed when unarmed compared to unarmed black individuals. We conducted a two-sample hypothesis test to test our alternative hypothesis being the proportion of blacks killed while unarmed is greater than the proportion of whites killed while

unarmed using the “Police Killings” data set. The null hypothesis was that there would be no difference between the proportion of black unarmed individuals and white unarmed individuals killed. We did so by wrangling the data to have two datasets. One dataset included all whites that were unarmed and included 51 elements. The other dataset included all the blacks that were unarmed and contained 32 elements. The police killings dataset included 236 white individuals and 135 black individuals. This data regarding white and black individuals and their arment can be viewed in figure 3. After conducting the hypothesis test we obtained a p-value of 0.3684 and therefore did not have enough statistical evidence to reject the null. The proportion of black unarmed individuals killed was 0.2370370 and the proportion of white unarmed individuals was 0.2161017. Looking at the data at 5% significance, there is not enough evidence to claim that the proportions are statistically different.

Is there a significant disproportion in which races are killed by police officers?

Police killings correlating with the ethnicity of the victim has been a common study in recent years. Oftentimes these analyses are focused on the rates at which a single ethnicity (usually black) is shot with a firearm by an officer, compared to the rates at which white individuals are shot with a firearm. (Lawson, K. (2015). Police Shooting of Black Men and Implicit Racial Bias: Can't We All Just Get Along. U. Haw. L. Rev., 37, 339.) Past studies have been useful in determining methods of prevention, such as adding implicit bias courses in police officer training (Smith, R. J. (2015). Reducing Racially Disparate Policing Outcomes: Is Implicit Bias Training the Answer. U. Haw. L. Rev., 37, 295.). In our particular study, rather than focusing on one or two particular ethnicities, we performed a multiple inference model to consider if there is a significant difference when accounting for the 5 recognized races in the United States of America (see figure 4). While the data frame consists of a 6th variable under race listed as “other,” this has been removed to prevent skewed data or potential confounding variables.

A Chi-Squared Test was conducted on the five variables to perform this analysis. Our null hypothesis stated that there was no difference in the proportion of races killed by police officers. Our

alternate hypothesis stated that there was a difference in the proportion of races killed by police officers. Using the dataset for our analyses, a p-value of less than 2.2e-16 was output, giving us enough evidence to reject the null hypothesis.

CONCLUSION

In conclusion, we were able to analyze trends and inferences regarding police killings using the sample dataset “Police Killings,” and statistical techniques. These findings are valuable due to the prevalence of police killings and potential use for predicting trends in police caused fatalities. Each of our questions was chosen in an attempt to either find a trend or generate a statistical inference. Through a hypothesis and confidence interval, we were able to discover to state with 95% confidence that the average age of a police killing victim is between 21.09449 and 23.40659. In our second research question, we found that the share of 25 and older population with a bachelor's degree or higher had a weak moderate negative correlation of -0.474 to the tract-level unemployment rate. For the third question, we found that the proportion of black unarmed individuals killed was not statistically different compared to the proportion of white unarmed men through a hypothesis test. Lastly, our analysis of our fourth research question indicated that there was a significant difference in the proportions of races affected by police killings. These findings give us insight on how the variables in our dataset relate to police killings and could be further used in predicting trends in fatalities.

Though we were able to pinpoint some interesting findings with our chosen dataset, there were some limitations. For instance, when we were first brainstorming research questions to investigate, we were interested in building a model using some of the factors in the data set to predict the likelihood of someone being a victim of a police killing. However, we were not able to do so since there was no response variable for being killed in the dataset and everyone's name in the dataset was killed. This thereby restricted us on being able to predict the data. We were also limited in finding trends with the

chosen dataset since all the data provided was from the year 2015. As a result, we couldn't analyze trends over time to see if there has been an increase or decrease in killings throughout the years. In addition, there are some unknowns within some variables and we, therefore, had to omit the incomplete data.

APPENDIX

Figure 1

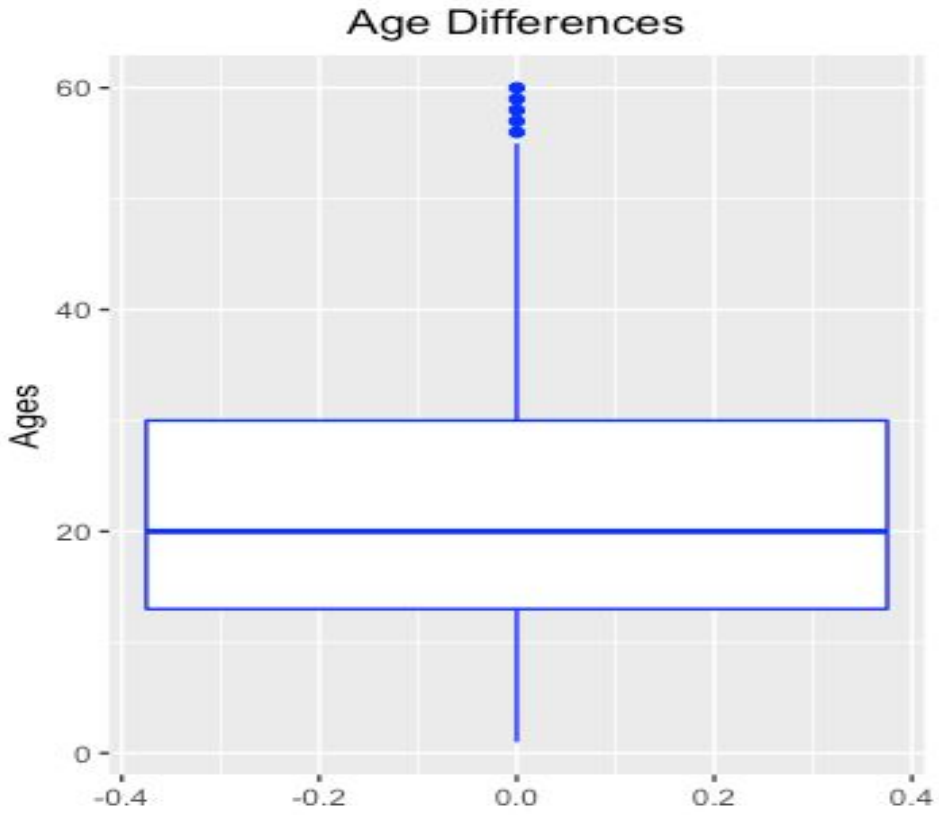


Figure 2

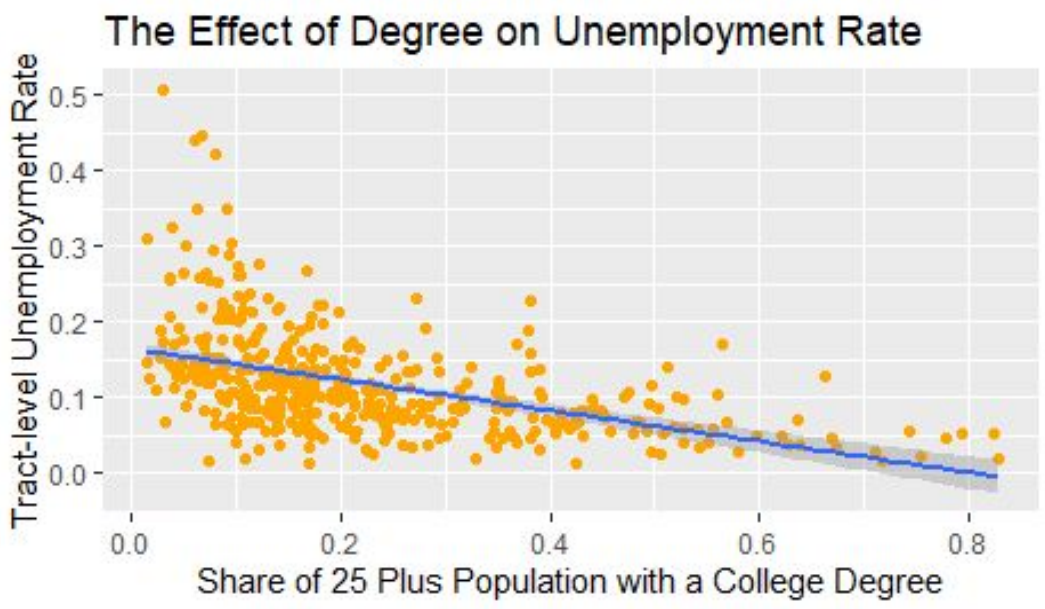


Figure 3

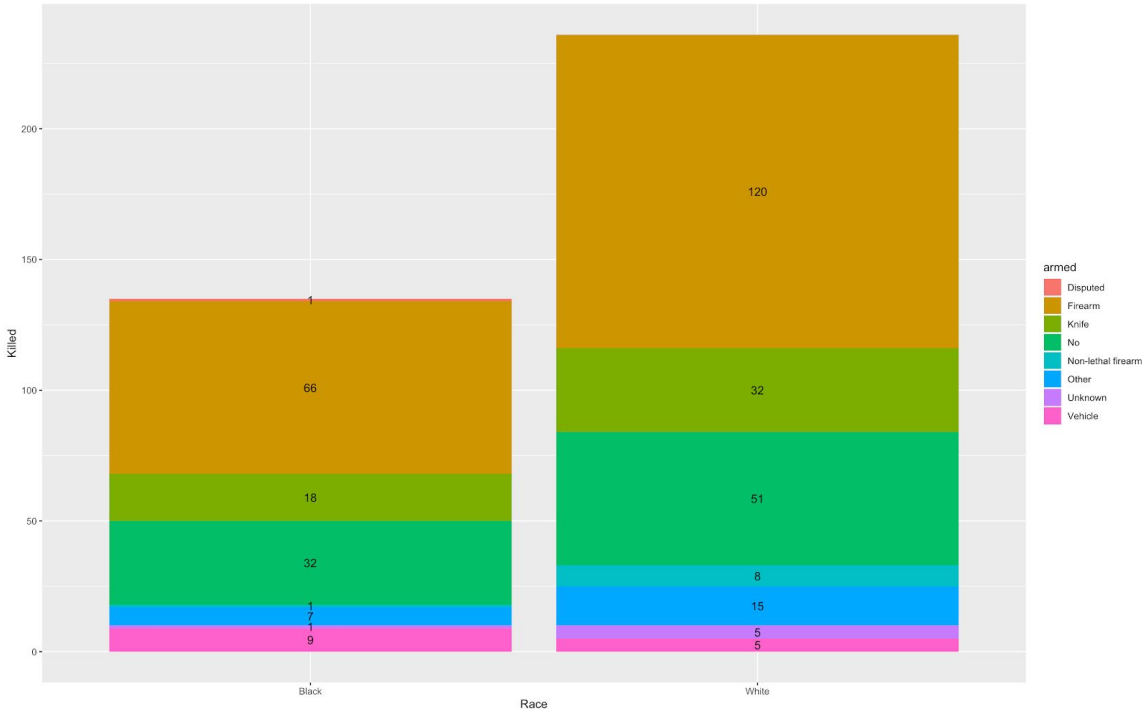
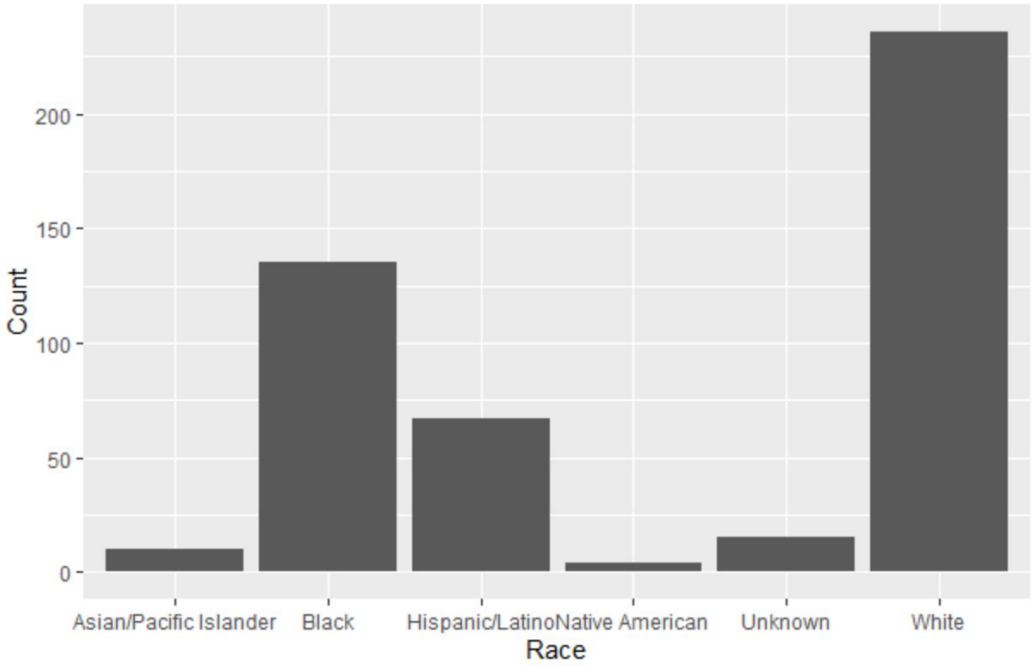


Figure 4


Count
200
150
100
50
0
Asian/Pacific Islander
Black
Hispanic/Latino
Native American
Unknown
White
Race

REFERENCES

Chen, A. (2017, April). More education: Lower unemployment, higher earnings : Career Outlook. Retrieved from https://www.bls.gov/careeroutlook/2017/data-on-display/more-education.htm?view_full

Khan, Amina. “Getting Killed by Police Is a Leading Cause of Death for Young Black Men in America.” *Los Angeles Times*, Los Angeles Times, 16 Aug. 2019, www.latimes.com/science/story/2019-08-15/police-shootings-are-a-leading-cause-of-death-for-black-men

United States, Department of Energy, “Use of Deadly Force”*Code of Federal Regulations*, Government Printing Office, 2011,999-1000

Statista Research Department. (2020, March 31). People shot to death by U.S. police, by race 2020. Retrieved from https://www.statista.com/statistics/585152/people-shot-to-death-by-us-police-by-race/