# Police Shootings and Racism in America

Connie Wu

11/09/2020

## I. Introduction

On May 25, 2020, George Floyd was publicly suffocated to death by police officers in Minneapolis, Minnesota. This, along with the murder of Breonna Taylor by Louisville police, incited waves of protests against police brutality across the United States and increased the spread of Black Lives Matter content on social media. The recent boost in attention to the Black Lives Matter movement has once again brought to light the issue of racism in America and its link to police brutality, and more specifically, police use of deadly force. Several studies have already found that race does play a part in who is targeted and killed in police shootings. For example, in a recent VICE News investigation about police shootings, it was found that "Black people were shot more often and at higher rates than people of any other race." [1] Additionally, Edwards et. al performed a study regarding the effects of age, race-ethnicity, and sex on the risk of being killed by lethal force by law enforcement and similarly found that "Black men are about 2.5 times more likely to be killed by police over the life course than are white men" while "Black women are about 1.4 times more likely to be killed by police than are white women." [2] However, even with these studies and their disturbing conclusions, a poll done by AP-NORC in June 2020 found that still 39% of Americans think that police are not more likely to use lethal force against a Black person than a White person. [3] Although this has decreased from an overwhelming 51% in June 2015, there is still a great deal of research that needs to be done in this area to provide more statistical evidence backing the relationship between racism and police use of deadly force.

As a result, I have decided to build off of VICE News' study and investigate data on police shootings further to understand the roles that the race of both victim and officer, as well as other factors such as whether they are carrying a weapon and the total number of victims in the crime, play in fatal versus non-fatal police shootings. Additionally, to account for the varying locations of the homicides, I will be using a dataset found on Kaggle detailing the gun provisions that are upheld by each state see how gun legislation affects lethal vs. non-lethal shootings. I will also add a predictor indicating whether the state in which the homicide occurred requires de-escalation training for police officers. Finally, I will explore how the race demographics of each location relates to police use of deadly force. This will allow me to better understand how racism has manifested itself in America's police system and determine whether current attempts to prevent police use of lethal force are effective or not.

## II. Data

As a basis for this study, I will be using the same dataset that VICE News used. This dataset contains data on officer-involved shootings from 47 of the largest local police departments in America, and more specifically, "information on 4,117 incidents and 4,400 subjects [(victims)] over seven years." [1] VICE News provides 34 variables in the dataset, including dates ranging from January 2010 to September 2016, 47 unique cities, subject race, subject gender, office race, officer gender, the type of weapon the subject was carrying, and whether the shooting was fatal or not. Race was broken up into 6 categories: White (non-Hispanic) (represented as W), Black (non-Hispanic) (B), Asian (A), Latino (L), Other (O), and Unknown (U). Gender was broken up into 3 categories: Male (M), Female (F), and Unknown (U). Weapon type was broken up into

5 values: "gun", "knife", "replica", "other", and "unknown". Additionally, there were cases in which multiple victims and/or multiple officers were present in the shooting. Each of these scenarios was still represented within one row of the dataset, but semi-colons were present in the respective victim and officer columns, separating each individual's information from each other.

Additionally, to assess the effect of different types of legislation on fatal vs. non-fatal police shootings in various states, I will be using a dataset from Kaggle containing 135 variables that detail whether a certain gun provision is absent or present in a certain year and U.S. state for 133 different gun provisions. [4] The years range from 1991 to 2017, and the gun provisions address 14 categories, some of which are dealer regulations, buyer regulations, prohibitions for high-risk gun possession, background checks, ammunition regulations, possession regulations, concealed carry permitting, assault weapons and large-capacity magazines, child access prevention, gun trafficking, and domestic violence. A 1 in the gun provision column represents a presence of the law, and a 0 indicates an absence. I will also add a column to this gun provision dataset indicating whether a state requires de-escalation training based off of information reported by APM Report in 2017, with a 1 meaning the state does require training and a 0 meaning the state does not. There were 16 states that required de-escalation training as of November 2017. [5]

Finally, to examine city demographics as a predictor, I will be using population data from the 2013 American Community Survey that VICE News had already cleaned and standardized to the shootings dataset, meaning the city names can be matched up between the two datasets. VICE News most likely provided only 2013 census data because it is the average year of all the years represented in the shootings dataset. Furthermore, using census data only from 2013 requires that we analyze this data under the assumption that there is no drastic difference in population demographics between 2010 and 2016. The census data includes 7 variables: the police department and the city it's located in, as well as the city's Black, Asian, Hispanic, White, and overall total population in 2013.

**Data Processing**

Although the VICE News dataset was already relatively clean with regards to the victim's data, there was still quite a bit of cleaning that needed to be done for the race and gender data of the officers. To clean these columns, I replaced all values that were not "W", "B", "A", "O", or "U" in the race column with the most informed guess that I could make about what the values represented. For example, "A/PI" values were taken to represent Asian/Pacific Islander and thus were replaced with "A", and "A/W" or other values with a "/" in them were typically taken to represent multi-racial individuals and as a result were replaced with "O." Similar procedures were carried out in the gender column for officers for values that were not "M", "F", or "U".

After this initial cleaning, I created new columns in the same dataset that represented whether any victims of race Black, White, Asian, or Other were present in the crime, respectively, and if any male or female victims were present in the crime, respectively. I also added new columns to represent the same information for officers (for each race and gender, whether each was present), and a new column to represent whether any victim involved in the shooting was shot fatally. Additionally, I made sure that each weapon type had its own column. All of these added columns had values of either 0 or 1, with 0 representing an absence of the variable and 1 a presence. Finally, I filtered out 2,063 rows that had only unknown and NA values for the fatality of the shooting or the races or genders of the victims or officers, as these rows did not provide sufficient information for my analysis, and I selected only the columns I needed, such as whether any victim of the shooting was shot fatally or not, the genders and races of the victims and officers, and the weapons that the victims were carrying, if any.

For the gun provisions dataset, I created a new column that sums up the total number of laws listed in the dataset that each state had in 2013 (for the same reason that VICE News used 2013). Additionally, as I mentioned earlier, I appended a column to this dataset representing whether a state requires de-escalation training for police officers as of November 2017. 2017 was used under the assumption that no major event occurred between September 2016 and November 2017 that caused a sudden increase in the number of states that require de-escalation training for police officers. Finally, for the census data, the only cleaning that

needed to be done was extracting the state that each city was in from the department column so that the census could be easily joined with the cleaned shootings dataset by city name, as well as the gun provisions dataset by abbreviated state name.

After merging the three datasets, the final dataset consists of 2,054 observations.
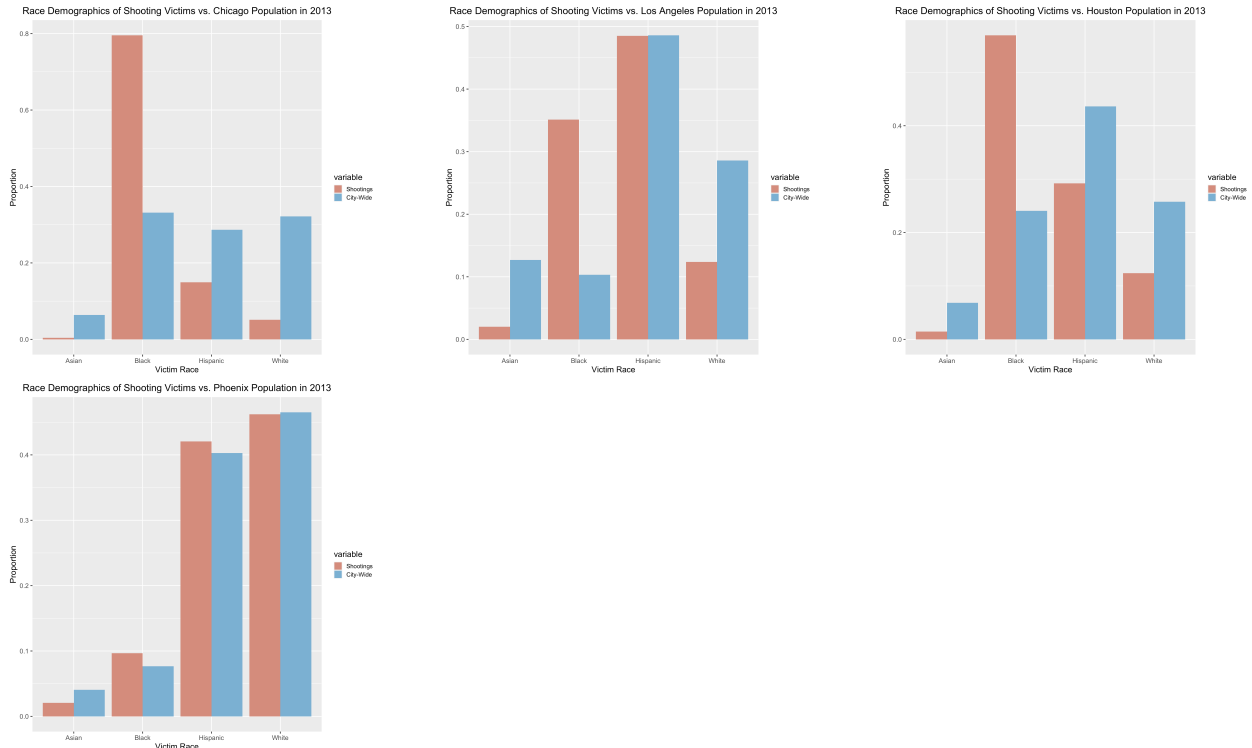
**Exploratory Data Analysis**



Figure 1: Race Proportions of Shooting Victims vs. City Population Demographics of the Most Popular 4 Cities in the Data

Plotted above in *Figure 1* are the race proportions of shooting victims in the top 4 most popular cities in the dataset (Chicago, Los Angeles, Houston, and Phoenix) vs. the race proportions of those cities' total populations in 2013. Because the bars representing shootings in the Black race category are taller than the bars representing city-wide population in all 4 plots, it is evident Black people are disproportionately the victims of police shootings compared to city race proportions. This can especially be seen in Houston, where Black people make up less than 25% of the city population yet more than 50% of the victim population of police shootings.

Additionally, there appears to be 1,906 incidents in the final dataset in which there was only one victim, and 148 with multiple victims. Among all shootings with only one victim, it appears in *Figure 2* that the victim is more often Black than some other race. However, it seems that Black victims and most victims of the other race types are more likely to be shot non-fatally than fatally, whereas White victims seem more likely to be shot fatally. Finally, it is important to note that the numbers of victims that are of race types Asian and Other are relatively small. Because of this, we will combine Asian and Other race types into one single race category called "Other" for both officers and victims for our models.

Similarly, as seen in *Figure 3*, Black victims make up the highest proportion of victims in shootings with multiple victims. However, in these shootings, all race types are more likely to be shot non-fatally than fatally, with Latinos having the highest proportion of being shot fatally out of all other race types.
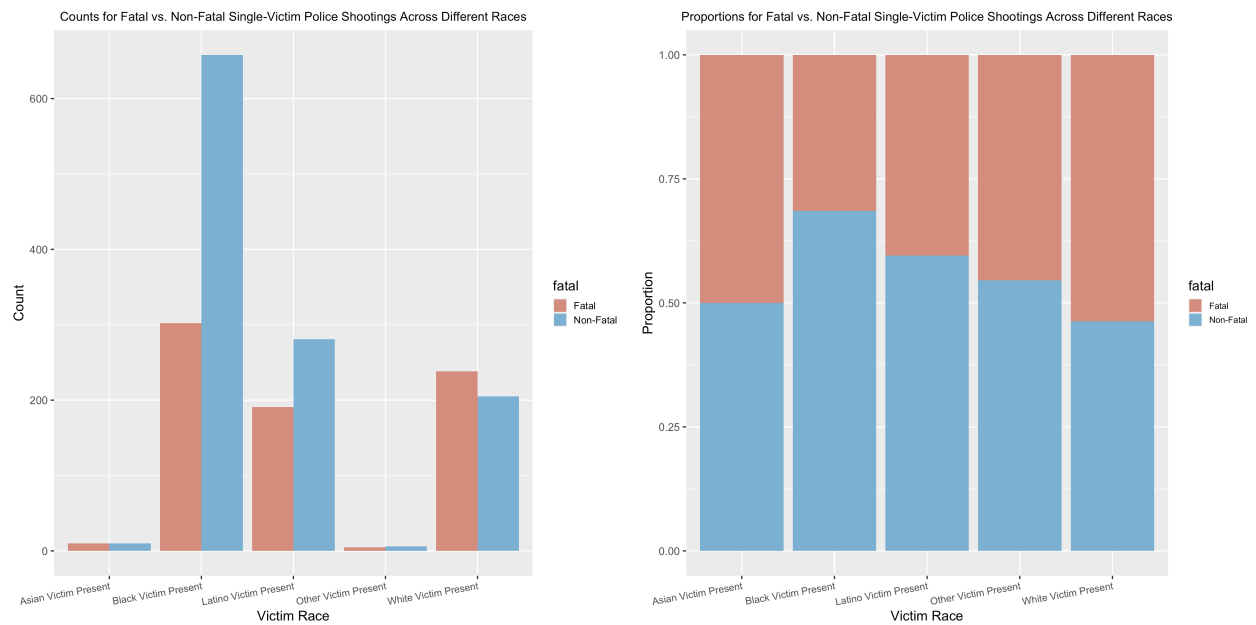
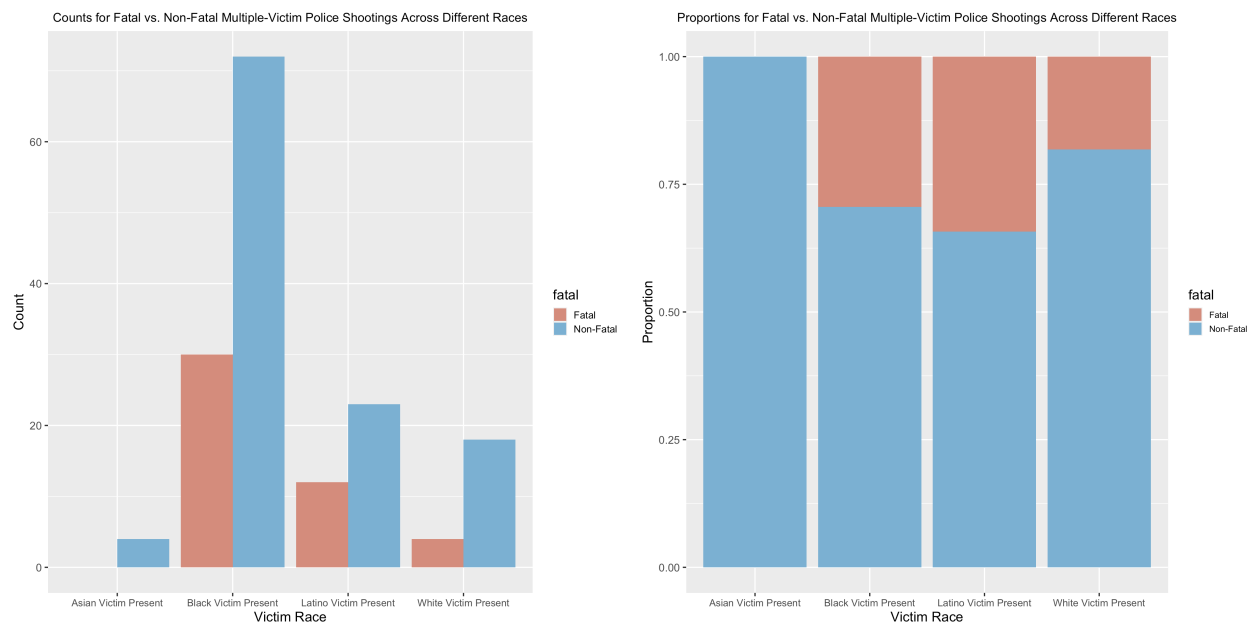Figure 2: Counts (left) and Proportions (right) of Different Races Across Single Victim Shootings



Figure 3: Counts (left) and Proportions (right) of Different Races Across Multiple Victim Shootings

Additionally, it should be noted that there are no victims with race type "Other" in shootings where there are multiple victims.

## III. Methods

Because I want to explore the relationship between the number of victims in a police shooting affects and whether the shooting is fatal or not, I decided to create two separate models: one for shootings with only one victim, and one for shootings with multiple victims. Both models are random effect logistic regression models because the range that the response variable can take in logistic regression is the most reasonable given that I am predicting whether a victim is shot fatally or not (labeled 1 and 0, respectively). Additionally, I will be using random effects to control for correlation within cities.

To create these two models, I first performed backwards selection using simple logistic regression models on all relevant predictors in order to get a sense of which predictors are most statistically significant in prediction shooting fatality. Then, I added potential interactions that I wanted to explore one at a time and performed likelihood ratio tests with each new model iteration to check if the added interaction improved the previous model. Whether an interaction was statistically significant or not was determined by whether the likelihood ratio test's p-value was less than 0.05. Finally, I put these variables obtained from backwards selection into a random effects linear regression model with city as the random intercept and evaluated the fit of this model using cross validation, which I will describe in more detail in the Results section. From there, I tweaked the model by experimenting with additional predictors, interactions, and random effects between city and other predictors in order to obtain the best fitting model.

The final model for single-victim shootings is written out below:

$$log(\frac{p_i}{(1-p_i)}) = \beta_0 + \beta_1 \ SubjectRace_i + \beta_2 \ OfficerBlack_i + \beta_3 \ OfficerWhite_i + \beta_4 \ MultipleOfficers_i$$

$$+\beta_{5j} \ VictimWithGun_{ij} + \beta_6 \ VictimWithOther_i + \beta_7 \ VictimWithReplica_i + \beta_8 \ VictimWithKnife_i + \beta_9 \ VictimUnarmed_i + b_i$$

$$b_i \sim N(0, \sigma_b^2)$$

$$SubjectRace = \text{White*, Black, Latino, Other}$$

In the model above, it is important to note that the predictor for whether the victim was carrying a gun is varied by city. This means that there is both a random intercept for city as well as a random slope for whether the victim was carrying a gun.

The final model for multiple-victim shootings is written out below:

$$log(\frac{p_i}{(1-p_i)}) = \beta_0 + \beta_1 \ black\_victim\_present_i + \beta_2 \ white\_victim\_present_i + \beta_3 \ latino\_victim\_present_i +$$

$$\beta_4 \ (\frac{black_i}{city\_pop_i}) + \beta_5 \ black\_officer\_present_i + \beta_6 \ white\_officer\_present_i +$$

$$\beta_7 \ latino\_officer\_present_i + \beta_8 \ victim\_armed\_gun_i + \beta_9 \ white\_officer\_present_i * black\_victim\_present_i$$

## IV. Results

**Single Victims Model**

**Multiple Victims Model**

TO DO: - fix my models – backward selection - random effects by city, maybe race | city - more interactions

**Analysis Plan**

Preliminarily, it appears that none of the legislation variables were significant in predicting whether one was shot fatally or not, and thus these were not included in either of the models. Additionally, victim and officer race both are significant individually and as part of interaction terms, which will be interesting to quantify. As expected, the coefficient for subject race type Black is positive in both models, meaning Black victims are more likely to be shot fatally. In the next revision, when quantifying the effects of each of these predictors on police shooting fatality, I will be analyzing the coefficients in terms of odds ratios, since I used logistic regression models. I also will most likely change the baseline to be White for both victim and officer race, since I want to investigate racism against minorities specifically in the law enforcement system.

For sensitivity analysis, I plan to perform cross-validation on the dataset by holding out 10% of the data to use as a test set and training my models on the other 90% of the data as a way to assess prediction accuracy. For model diagnostics, I will plot binned residual plots and ROC curves to assess how well my models fit the data.

**Strengths and Limitations**

One strength of my analysis is that I am considering many variables as predictors, from individual level data to contextual predictors, such as state gun legislation. This allows me to thoroughly explore and determine whether the relationships between a wide range of potential predictors and police shooting fatality in the context of police brutality are significant in any types of shootings (single vs. multiple victims, for example). There are several limitations to this study. For example, the VICE News dataset spans over the years 2010 to 2016, and neither of my models will be exploring time series data whatsoever. Thus, the aspect of time may be playing some sort of effect on my response variable that I am not accurately capturing in my models. There are also several other confounding factors that have not been accounted for, such as how well one can shoot at baseline or what training they have received with regards to shooting. Additionally, when cleaning the data, there were several "Unknown" values in the fatality column. Unknown values were assumed to represent "No" in this study, although this may not necessarily be true, to allow for more straightforward cleaning of the data and not lose too many observations in the data. Finally, there were less than 200 cases in which there were multiple victims in my final dataset, which is a relatively small sample size. Thus, it may be necessary to take the results from the multiple-victim model with a grain of salt.

# V. Discussion and Conclusion

# VI. References

[1] McCann, Allison, et al. "Police Shoot Far More People than Anyone Realized, a VICE News Investigation Reveals." VICE News, VICE News, 12 Dec. 2017, news.vice.com/en_us/article/xwvv3a/shot-by-cops.

[2] Edwards, Frank, et al. "Risk of Being Killed by Police Use of Force in the United States by Age, Race–Ethnicity, and Sex." PNAS, PNAS, 20 Aug. 2019, www.pnas.org/content/116/34/16793.

[3] Press, Associated. "Sweeping Change In US Views Of Police Violence, New Poll Finds." KPBS Public Media, KPBS, 18 June 2020, www.kpbs.org/news/2020/jun/18/sweeping-change-us-views-police-violence-new-poll-/.

[4] https://www.kaggle.com/jboysen/state-firearms

[5] Gilbert, Curtis. "Most States Neglect Ordering Police to Learn De-Escalation Tactics to Avoid Shootings." Not Trained to Not Kill | APM Reports, APM Reports, 16 Sept. 2020, www.apmreports.org/story/2017/05/05/police-de-escalation-training.

[6] DeGue, Sarah et al. "Deaths Due to Use of Lethal Force by Law Enforcement: Findings From the National Violent Death Reporting System, 17 U.S. States, 2009-2012." American journal of preventive medicine vol. 51,5 Suppl 3 (2016): S173-S187. doi:10.1016/j.amepre.2016.08.027
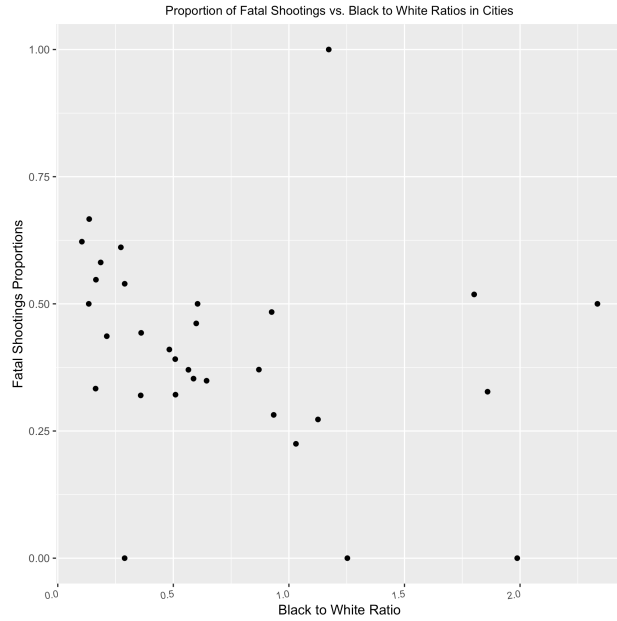
## VII. Appendix



Figure 4: Proportion of Fatal Shootings vs. Black to White Ratios in Cities in Dataset (left) and vs. Black Population Percentages in Cities in Dataset (right)

**Single-Victim Shooting**

|  | Estimate | Std. Error | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| (Intercept) | -1.3450848 | 1.1820630 | -1.1379130 | 0.2551568 |
| SubjectRaceB | 0.3195911 | 1.1776248 | 0.2713862 | 0.7860940 |
| SubjectRaceL | 0.4271347 | 1.1793897 | 0.3621659 | 0.7172281 |
| SubjectRaceO | -12.5734499 | 265.8428846 | -0.0472965 | 0.9622769 |
| SubjectRaceW | 1.3157766 | 1.1886011 | 1.1069960 | 0.2682957 |
| I(black/city_pop_2013) | -1.4736305 | 0.4334928 | -3.3994346 | 0.0006753 |
| ob | 0.0720676 | 0.1629737 | 0.4422037 | 0.6583418 |
| ow | 1.6171234 | 1.2851706 | 1.2582947 | 0.2082852 |
| ol | 0.3926620 | 0.1365577 | 2.8754300 | 0.0040348 |
| gun | 0.4290510 | 0.1007981 | 4.2565385 | 0.0000208 |
| SubjectRaceB:ow | -0.9435612 | 1.2874469 | -0.7328933 | 0.4636235 |
| SubjectRaceL:ow | -0.9358752 | 1.2938881 | -0.7233046 | 0.4694928 |
| SubjectRaceO:ow | 13.3356939 | 265.8447552 | 0.0501635 | 0.9599921 |
| SubjectRaceW:ow | -1.4144480 | 1.2997691 | -1.0882302 | 0.2764935 |

**Multiple-Victim Shooting**
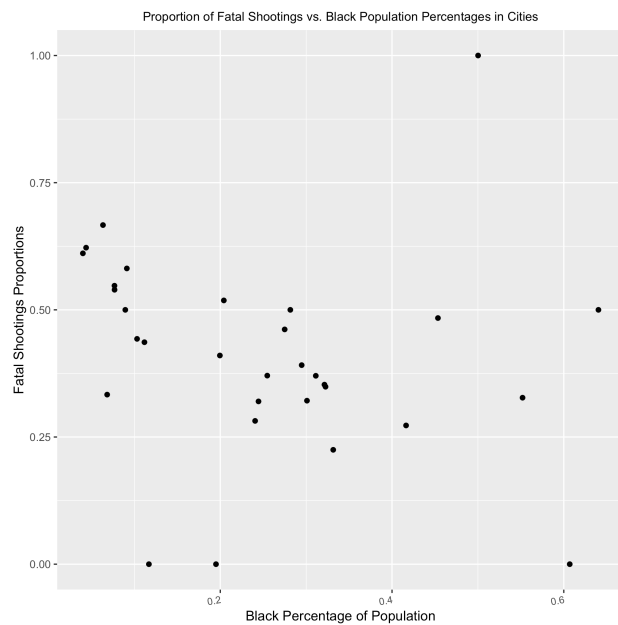
Figure 5: Proportion of Fatal Shootings vs. Black to White Ratios in Cities in Dataset (left) and vs. Black Population Percentages in Cities in Dataset (right)
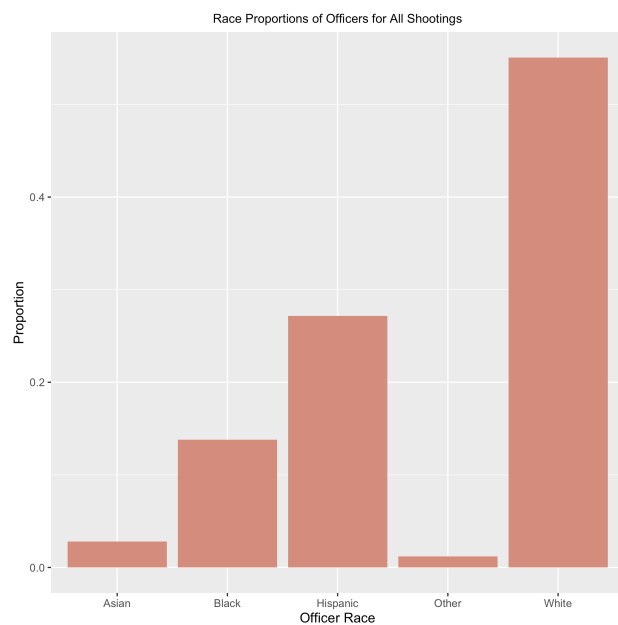


Figure 6: Race Proportions of Officers For All Shootings (left) and For Fatal vs. Non-Fatal Police Shootings Where At Least One Black Victim was Present (right) and
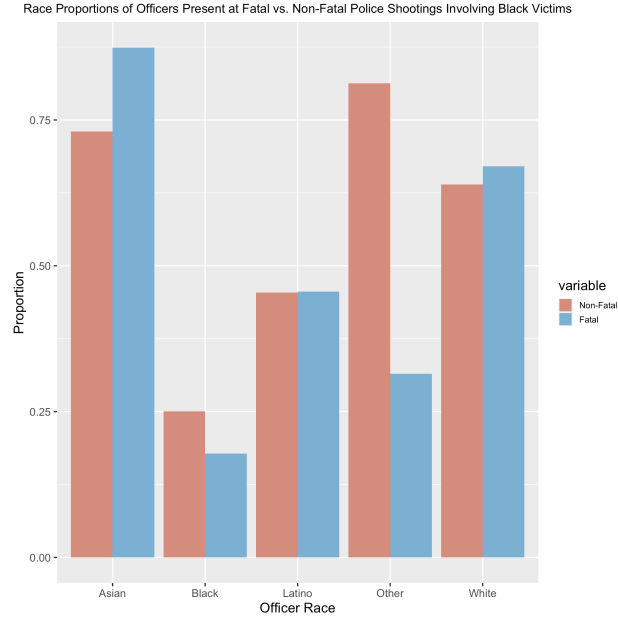
Figure 7: Race Proportions of Officers For All Shootings (left) and For Fatal vs. Non-Fatal Police Shootings Where At Least One Black Victim was Present (right) and

|  | Estimate | Std. Error | z value | Pr(>|z|) |
|---|---|---|---|---|
| (Intercept) | -3.9225831 | 1.5785672 | -2.4849009 | 0.0129588 |
| sb | 2.5412636 | 1.3221039 | 1.9221361 | 0.0545886 |
| sw | -0.4031754 | 0.7352400 | -0.5483589 | 0.5834455 |
| sl | 1.1512831 | 0.8586224 | 1.3408491 | 0.1799694 |
| I(black/city_pop_2013) | -0.4803802 | 2.2628839 | -0.2122867 | 0.8318833 |
| ob | 0.0135013 | 0.5833334 | 0.0231450 | 0.9815346 |
| ow | 2.6209555 | 1.1643627 | 2.2509787 | 0.0243869 |
| ol | 0.3961047 | 0.4422938 | 0.8955691 | 0.3704829 |
| gun | 0.2048378 | 0.4059934 | 0.5045347 | 0.6138857 |
| sb:ow | -1.9515969 | 1.2290236 | -1.5879247 | 0.1123034 |