

PSTAT 100 Project report

Racial Disparities in Fatal Police Shootings

Jordan Tran, Alyssa Keehan, and Jasmine Kwok.

Author contributions

Jordan contributed to making an abstract, data description, results, and making multiple visualizations.

Alyssa contributed by creating the heat map and writing out the Background, aims and discussion parts.

Jasmine contributed by writing the data description, methods, results and some visualizations.

Abstract

Unjust police shootings have sparked national outrage in the United States and have started international social movements, such as Black Lives Matter that protest against incidents of racially motivated police violence. The media spotlight has been shined on numerous infamous cases of potentially prejudiced police shooting victims, like the shooting of Breonna Taylor. Our project aims to explore whether police shootings are racially biased and if certain racial groups are targeted more. We also aim to look specifically at California which is the State that has the most police shootings and to see if there are any racial disparities in its victims. We find that there is a difference between the racial distribution of police shooting victims and the racial distribution of the population and that African Americans are overly represented in the population of police shooting victims. We determine that our county demographic data on education and income have no correlated variables with black shooting victim population and that California reflects our previous findings and also that Hispanics are overly represented in the State's population of police shooting victims as well.

Introduction

Background

The frequent and lawful police killings in just the last year have contributed to some of the hardest months of the pandemic. With the media coverage of the Black Lives Matter Movements, it seems that so many Police Shooting Victims are African American. However, we noticed that in total there are more White victims than Black in police killings. The connection between police killings and race, especially, has become an even larger issue and has led to mass protesting and looting all over the country. In light of these events circling the wrongful police killings, we intend to create an educational, meaningful, and informative project that is relevant to the current society and helpful for further research in updating future police funding and practices.

Our goal is to perform an in-depth analysis on factors, like race, that were accounted for in thousands of fatal police shootings in recent years. With this research, we hope to learn the specific variables that contribute to racially biased fatalities and explore potential solutions to prevent them from occurring.



Gordon Parks, Untitled, New York, New York, 1963. Courtesy and © The Gordon Parks Foundation.

Aims

The first direction we would like to go in for this project is to detect any notable differences between the distributions of races in the regional population and those who were victims of police shootings. After seeing that the proportion of Black Victims is higher than the proportion of the Blacks in our population, we decided to investigate the demographics of a particular region where a police killed a Black victim. We want to see if there is any indication in geography or demographics that may contribute to high Black fatalities in police shootings. Finally, we decided to look specifically at California, the state with the highest number of police killings, to see if its racial distribution can be related to geography. Since California has one of the highest relative Hispanic populations in the country, it would be interesting to see how the distribution of the minority population differs with the number of victims by race.

Materials and methods

Datasets

Data description

Our data is an up to date log of police killings in the United States from the Washington Post for the past five years. It contains general notes about the event and even information about the victim, police and station. The information comes from several different news sources, social media posts and police reports. The data started being logged after the incident in 2014 where Michael Brown, an unarmed black man was killed by the police. A post investigation of this incident showed that the FBI severely undercounted the number of police-caused fatalities and the reason being many departments failed to require reports of this kind. The Washington Post's database is kept up-to-date as fatal shootings and facts emerge from the Post's collective methods.

The data is publicly available:

Tate, Julie, et al. "Fatal Force: Police Shootings Database." The Washington Post, WP Company, 22 Jan. 2020, www.washingtonpost.com/graphics/investigations/police-shootings-database/.

Sample and measurement information

The initial dataset, Washington Post's data of police fatal shooting incidents from Jan. 1, 2015, was recorded by culling local news reports, law enforcement websites, social media, as well as monitoring independent databases such Killed by Police and Fatal Encounters. The population of this data only includes all fatal shootings in the United States by a police officer in the line of duty and excludes deaths of people in custody, fatal shootings by off-duty officers, or non-shooting deaths. The sampling frame consists of all entities in the population that were publicly reported in any form of way since Jan. 1, 2015. The sampling mechanism is a census which indicates that our sample consists of all entities in our sampling frame. The sample consists of 6280 observations of fatal police shootings from Jan 2, 2015 to May 9, 2021. Our data is administrative and has no scope of inference.

Data structure

For this study, the observational units are incidents of a fatal shooting in the United States by a police officer in the line of duty. An observation was made at each shooting in which a police officer on duty shoots and kills a civilian since January 1st, 2015 to May 9th, 2021 which is the date we downloaded our dataset. The particular details about each killing are indicated using the variable descriptions in Table 1.

Table 1: variable descriptions and units for each variable in the final dataset.

Name	Variable description	Type	Units of measurement
id	unique identifier for each victim	Nominal	None
name	name of the victim	Nominal	None
date	date of the fatal shooting	Nominal	YYYY-MM-DD
manner_of_death	manner of victim's death	Nominal	None
armed	what the victim was armed with	Nominal	None
age	age of the victim	Numeric	Years
gender	gender of the victim	Nominal	None
race	race of the victim	Nominal	None
county	county where the fatal shooting took place	Nominal	None
state	state where the fatal shooting took place	Nominal	None
signs_of_mental_illness	indicator if the victim had a history of mental health issues, expressed suicidal intentions or was experiencing mental distress at the time of the shooting	Nominal	None
threat_level	threat level of victim to the police	Nominal	None
flee	indicator if the victim was moving away from officers	Nominal	None
body_camera	indicator if an officer had a body camera on at the incident	Nominal	None
longitude	longitude location of the shooting expressed as WGS84 coordinates, geocoded from addresses	Numeric	Degrees

Name	Variable description	Type	Units of measurement
latitude	latitude location of the shooting expressed as WGS84 coordinates, geocoded from addresses	Numeric	Degrees
is_geocoding_exact	indicator of the accuracy of the coordinates	Nominal	None
Budget_Per_Capita	police budget per capita of the state where incident occurred	Numeric	Dollars
Age.Percent 65 and Older	the percentage of population that is 65 years old or older	Numeric	Percentage
Age.Percent Under 18 Years	the percentage of population that is under 18 years old	Numeric	Percentage
Age.Percent Under 5 Years	the percentage of population that is under 5 years old	Numeric	Percentage
Education.Bachelor's Degree or Higher	the percentage of population having a bachelor's degree or higher	Numeric	Percentage
Education.High School or Higher	the percentage of population having a high school education or higher	Numeric	Percentage
Ethnicities.American Indian and Alaska Native Alone	the percentage of population who are Native Americans	Numeric	Percentage
Ethnicities.Asian Alone	the percentage of population who are Asians	Numeric	Percentage
Ethnicities.Hispanic or Latino	the percentage of population who are Hispanic or Latino	Numeric	Percentage
Ethnicities.Native Hawaiian and Other Pacific Islander Alone	the percentage of population who are Native Hawaiian or Other Pacific Islander	Numeric	Percentage
Ethnicities.Two or More Races	the percentage of population who belongs to multiple races	Numeric	Percentage
Ethnicities.White Alone, not Hispanic or Latino	the percentage of population who are White	Numeric	Percentage
Income.Median Household Income	median household income value	Numeric	Dollars
Income.Per Capita Income	income per capita	Numeric	Dollars
Income.Persons Below Poverty Level	the percentage of individuals who are below poverty level	Numeric	Percentage
Population.2014 Population	the 2014 population by county	Numeric	Persons
Population.Population per Square Mile	the population per square mile for each county	Numeric	Persons

In preprocessing, the initial dataset did not include details on budget per capita, age, education, racial, and income demographics of the population for each county. Since, these additional information would help with our subsequent analysis, we added the additional information by merging the 2014 county census data. The variables included in the final dataset are displayed in Table 1 while the first few rows of the final dataset are shown in Table 2.

Table 2: Example rows of the final dataset.

id	name	date	manner_of_death	armed	age	gender	race	County	...	Income.Persons	Population.2014	Population.Population
										Below Poverty Level		
3	Tim Elliot	2015-01-02	shot	gun	53.0	M	A	Mason	...	17.3	60711.0	63.3
4	Lewis Lee Lembke	2015-01-02	shot	gun	47.0	M	W	Washington	...	11.4	562998.0	731.4
5	John Paul Quintero	2015-01-03	shot and Tasered	unarmed	23.0	M	H	Sedgwick	...	15.3	508803.0	499.6

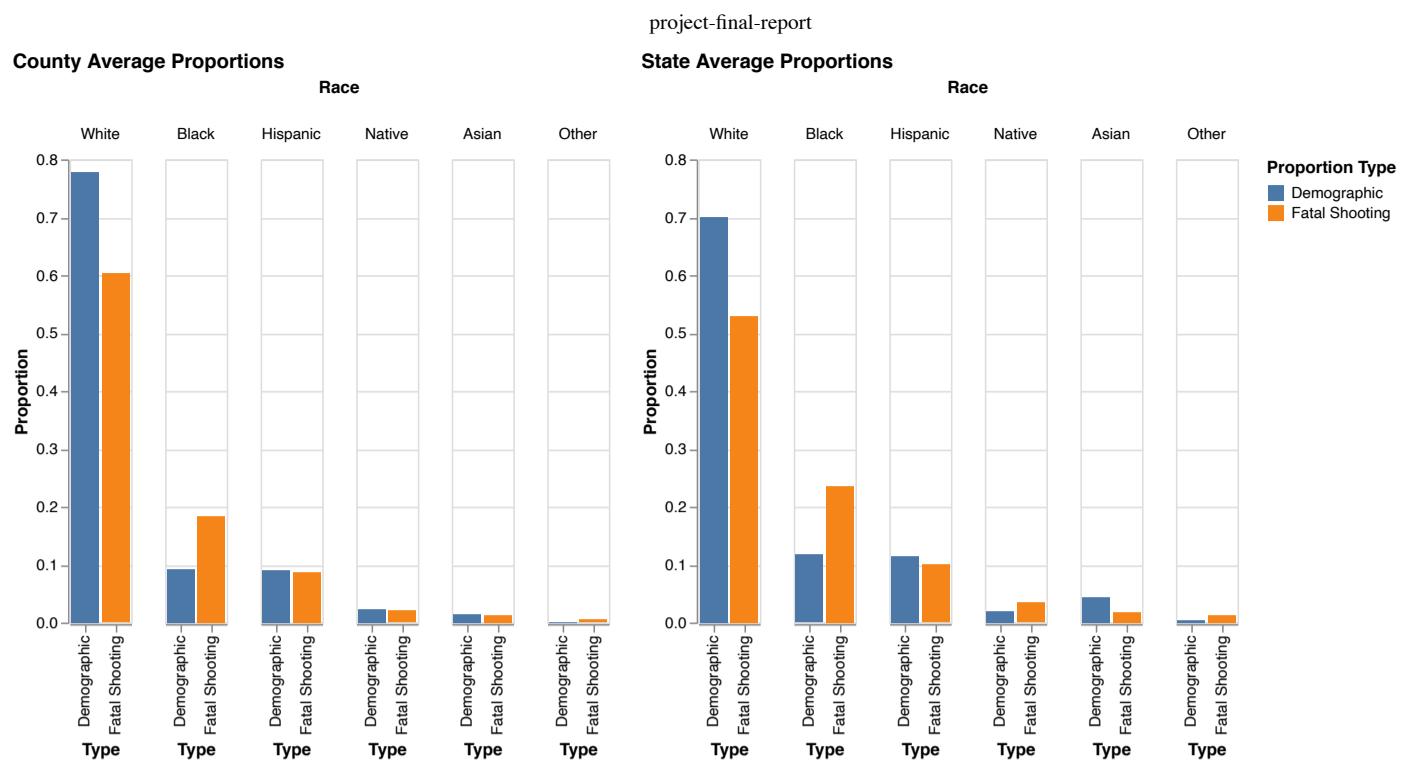
Methods

Our final direction for this project was to look specifically at race and other regional demographics to see if it had a notable relationship with the number of killings by race in that area. To start, we created bar charts to represent the average proportions of total racial populations and total killings per county and state. With this, we would be able to see if there are any notable distinctions between the values of a certain race. After deciding to focus on the proportion of black victims in our dataset, we merged and melted the demographic data of the victims with the region data of the shooting locations. Then we created a heat map to visualize potential patterns of covariation between the variables of the combined dataset. We then performed a hypothesis test comparing the absolute differences between the racial distribution of the victims and the total population of each county. Since our focus was on the difference in the black population of victims and total, we decided to look at computing the absolute difference and the total variation distance test statistics. The absolute difference test statistics focused primarily on comparing the difference between the black proportion of those killed in police killings and the total populations in counties. The total variation distance measurement was used to calculate that same test statistic for all racial categories included in our data. To better understand racial killings in California, we grouped the killings by each race and calculated their proportions as well as aggregated the census data to obtain the racial demographics proportion for California. A grouped bar chart was created to compare the proportion of racial demographics represented in fatal shootings to the actual proportion of each race in California. We also created a horizontal bar chart displaying the 10 counties in California with highest total police killings with the racial distribution for each of the counties.

Results

Racial representations in the population and victims of police shootings

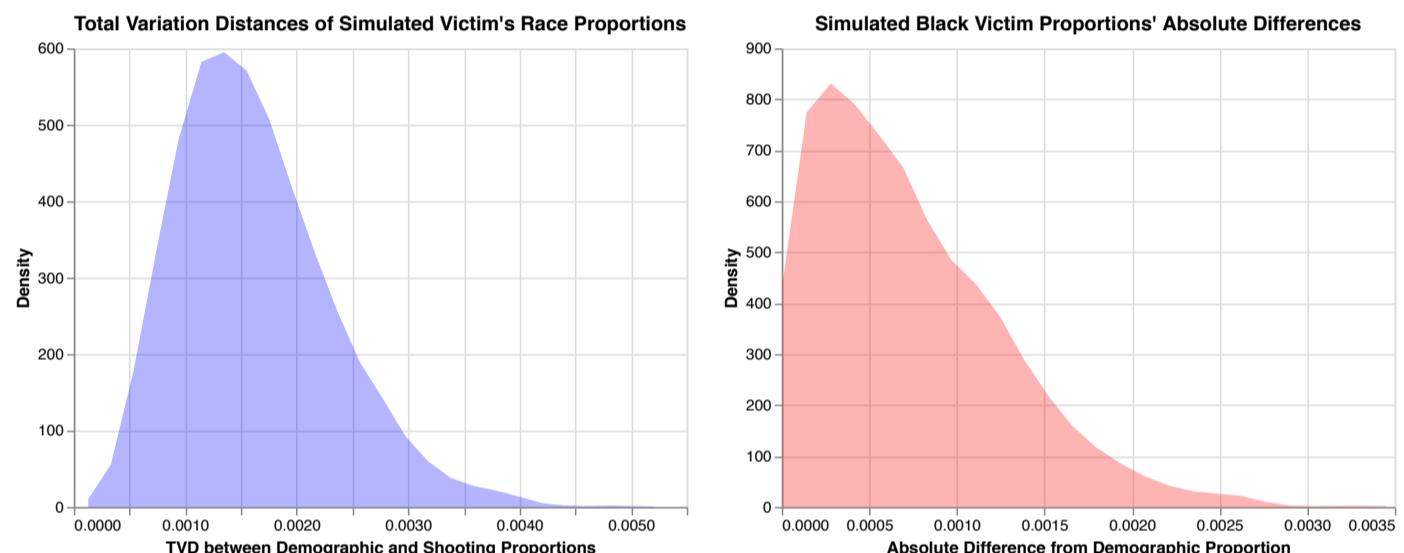
Figure 2: County and State Average proportions of racial demographics of police shooting victims and the racial demographics of the population



Exploratory analysis focused on the difference between the racial distribution of police shooting victims and the county average proportions of police shooting victims by race. Figure 2 shows the county and state proportion distributions of racial representations in the population and victims of police shootings.

Hypothesis tests on racial proportions of police shooting victims

Figure 3: Density graphs of simulated test statistics for our two above hypothesis tests.



H₀: There is no difference between the racial distribution of police shooting victims and the racial distribution of the population.

H₁: There is a difference between the racial distribution of police shooting victims and the racial distribution of the population.

A hypothesis test was performed using the test statistic of total variation distance between our observed police shooting victim racial proportions and our population racial distributions. Our observed total variation distance was .092 and a p-value of 0 was found. Our p-value of 0 makes us reject our null hypothesis and provides strong evidence towards our alternative hypothesis that there is a difference between the racial distribution of police shooting victims and the racial distribution of the population.

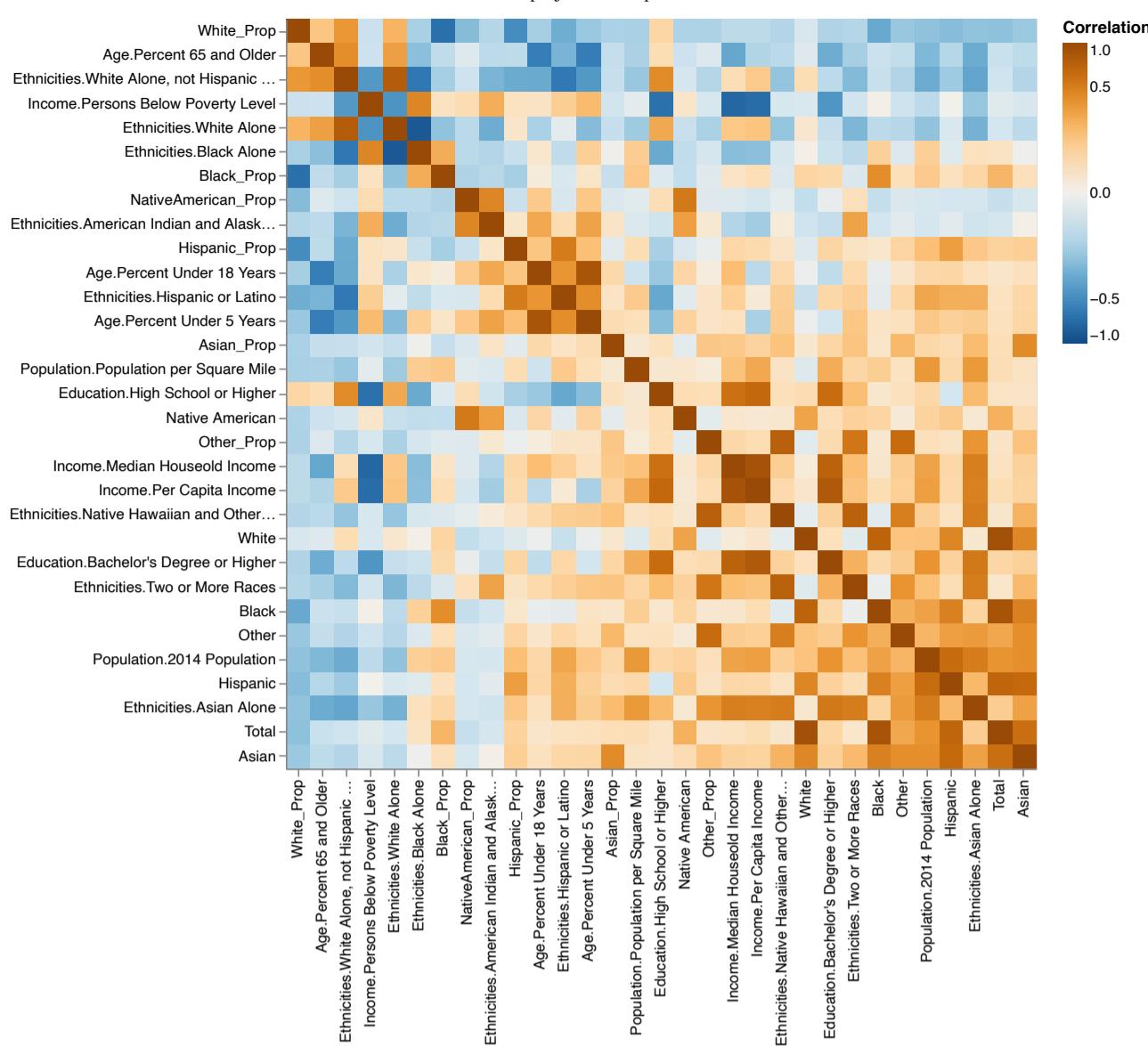
H₀: African Americans are equally represented in the population of a county and the county's population of police shooting victims.

H₁: African Americans are not equally represented in the population of a county and the county's population of police shooting victims.

An additional hypothesis test was conducted with a focus on the African American population. The test statistic used was the absolute difference between the proportion of police shooting African American victims and the population proportion of African Americans. Our observed absolute difference was .139 and it has the result of a p-value of 0. Our p-value of 0 makes us reject the null and provides strong evidence towards our alternative hypothesis that African Americans are not equally represented in the population of a county and the county's population of police shooting victims.

Exploring the linear correlation values of the racial counts and proportions of police shooting victims

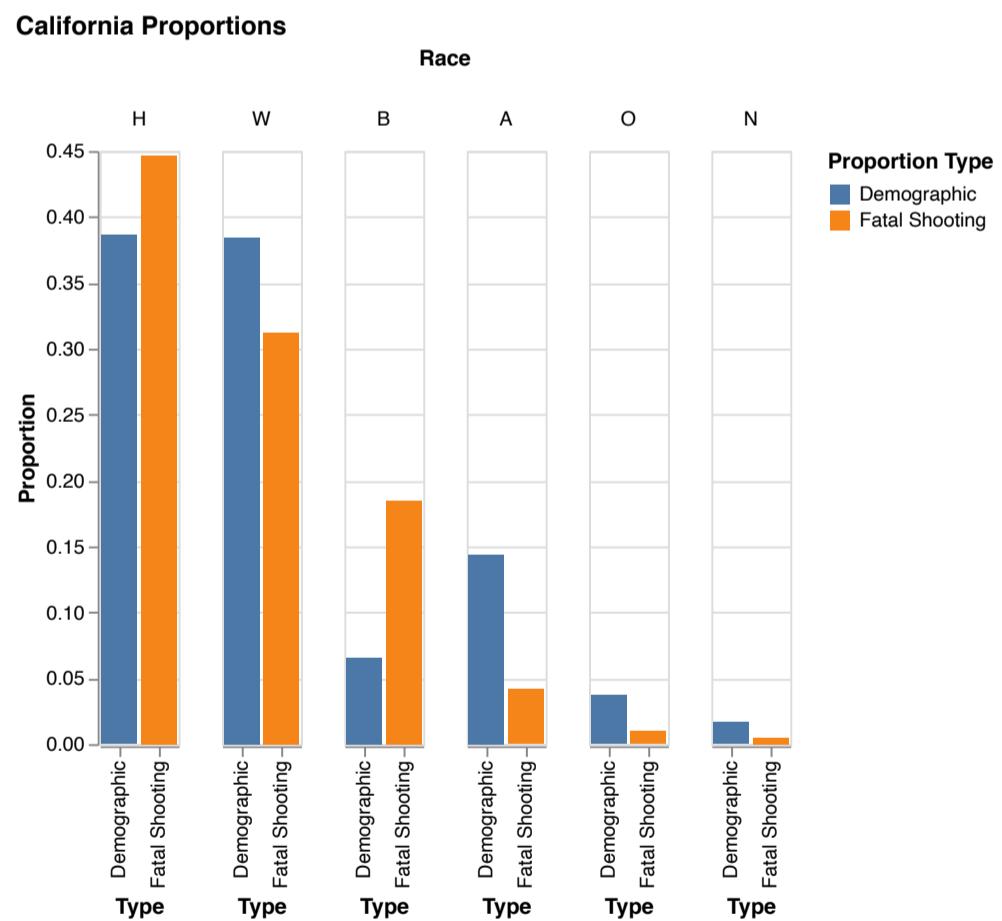
Figure 4: Heat Map showing correlation values between economic, education, racial, and police shooting racial demographics.



Linear correlation values of all racial counts and proportions against various economic and demographic variables was calculated. We found no significant linear correlation between the black population and the other demographic data by county.

California Police Shootings

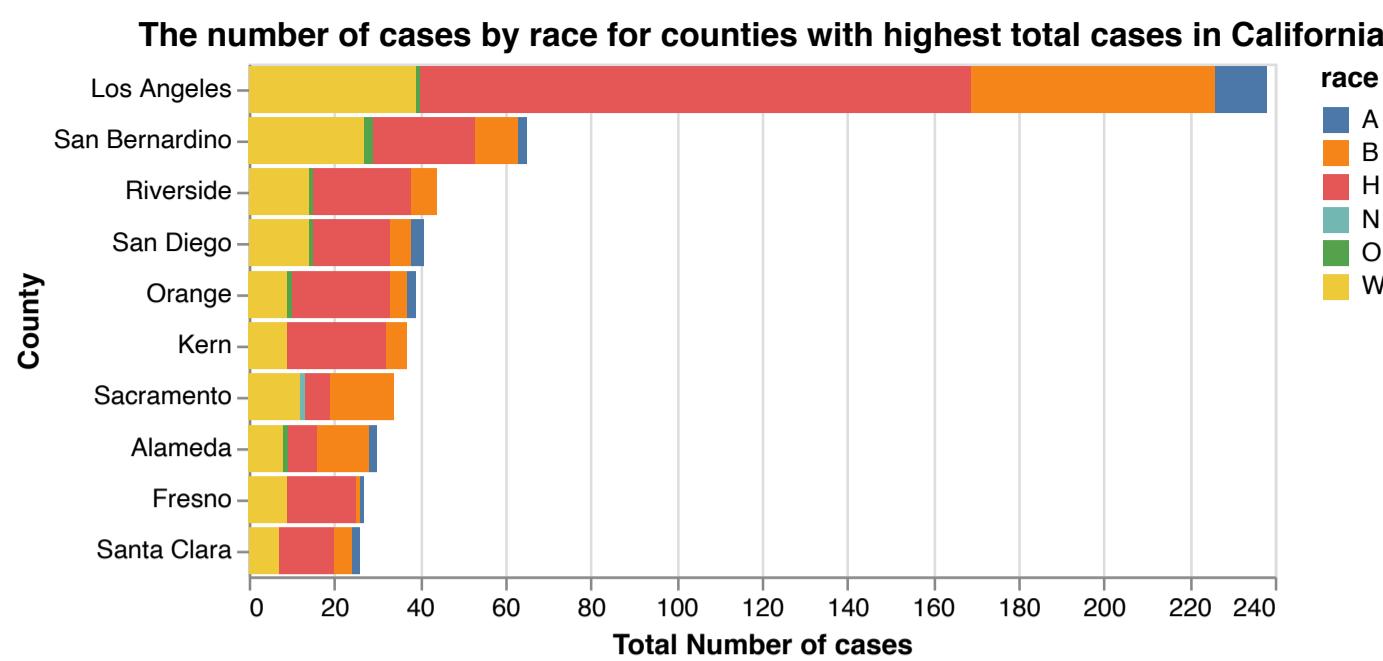
Figure 5: California Proportions of racial demographics on police shooting victims



The proportion of Hispanic and Black victims of the fatal shooting in California were much higher, around 6 percent and 13 percent, than the actual racial proportion in California. In contrast, the proportions of White, Asian, Native, and Other victims were lower in the fatal shootings data than the actual racial proportion in California by around 5 percent, 10 percent, 3 percent, and 1 percent. Ideally, we would hope for the proportion of racial demographics in the fatal shootings data equal the proportion from the census. However, our findings suggest that Hispanic and Black victims are overrepresented in fatal shootings while the other racial groups are underrepresented.

California Shootings by County

Figure 6: Horizontal Bar Chart of the 10 highest cases by county in California



Unsurprisingly, Los Angeles County has the highest total number of fatal shootings around 238 cases. The number of cases then drops significantly to less than 100 for the rest of the counties such as San Bernardino with around 65 cases and Riverside with around 42 cases. Contrary to our expectations, the largest proportion of victims were Hispanic instead of Black or White victims in all the counties above except for Sacramento and Alameda.

Discussion

For our project, we analyzed the police shootings from the past 5 years by specifically looking at racial associations across space. Regionally, our Census data did not show anything remarkable about the various demographics with cases of Black Police Shooting victims. However, we think that there may be confounding variables that might contribute to high police shooting rates if we looked specifically at demographic information separately for each race. Perhaps if we were to explore the crime rates by race, we could see if there may be more potential opportunities for police shooting incidents to occur if a certain race had higher rates of serious crimes. Also, exploring possible implicit bias within the police force that leads to racially motivated police brutality could explain the racial disparities in fatal police shootings. Since our data only includes the instances reported of fatal police shootings, our results do not show the true population of police brutality incidents in the United States. The racial disparities may be more exaggerated when taking into account all police brutality incidents. In addition, since we were unable to effectively find census city data for each police shooting incident, we aggregated the demographics by county which may have underestimated or overestimated extremes within the region.

In the future, we hope to have the resources to explore more about the racial disparities in police shootings in the United States. Instead of just focusing on the Black population we can look deeper into communities with larger minority populations and explore more about their grouped demographics. For example, if we can obtain city racial census data with demographic information such as income, we could explore more population differences that may act as confounding factors on police brutality. There are potentially more disparities in police shooting victims and we would like to further explore them. During our analysis, we noticed that police shooting victims were overwhelmingly male. Further research into the role of gender in being a victim of a police shooting could provide insight into more systematic bias in the police force or clues to other confounding variables to occurrences of police shootings.

7. Bibliography

Peeples, Lynne. "What the Data Say about Police Shootings." *Nature News*, Nature Publishing Group, 4 Sept. 2019, www.nature.com/articles/d41586-019-02601-9.

Tate, Julie, et al. "Fatal Force: Police Shootings Database." The Washington Post, WP Company, 22 Jan. 2020, www.washingtonpost.com/graphics/investigations/police-shootings-database/.