

Exploratory Data Analysis on 2015 Police Killings

Aman Raj

September 28, 2024

Abstract

This report conducts an exploratory data analysis (EDA) of the 2015 police killings dataset, focusing on the racial and demographic distribution of victims, their cause of death, whether they were armed, and the population statistics of their respective locations. Specifically, we analyze the following variables: raceethnicity, cause, armed, share_black and share_white.

Introduction

Police violence, especially in the United States, has been a subject of intense public debate. In 2015, numerous individuals were killed by police, prompting questions about racial disparities and the use of force. This analysis aims to explore these questions by examining race/ethnicity, cause of death, and population characteristics where the incidents occurred.

Data Description

The dataset used is titled “Police Killings in 2015” from Kaggle, which provides detailed records of individuals killed by law enforcement in 2015. Below are the variables of interest:

- raceethnicity: The race or ethnicity of the deceased.
- cause: The cause of death (e.g., gunshot, taser, etc.).
- armed: Indicates whether the deceased was armed.
- share_black: The share of the population that is Black (alone, not in combination with other races).
- share_white: Share of population that is non-Hispanic white.

Note that the dataset contains 34 columns and 467 entries. Out of the 34, variables of interest are the aforementioned 5.

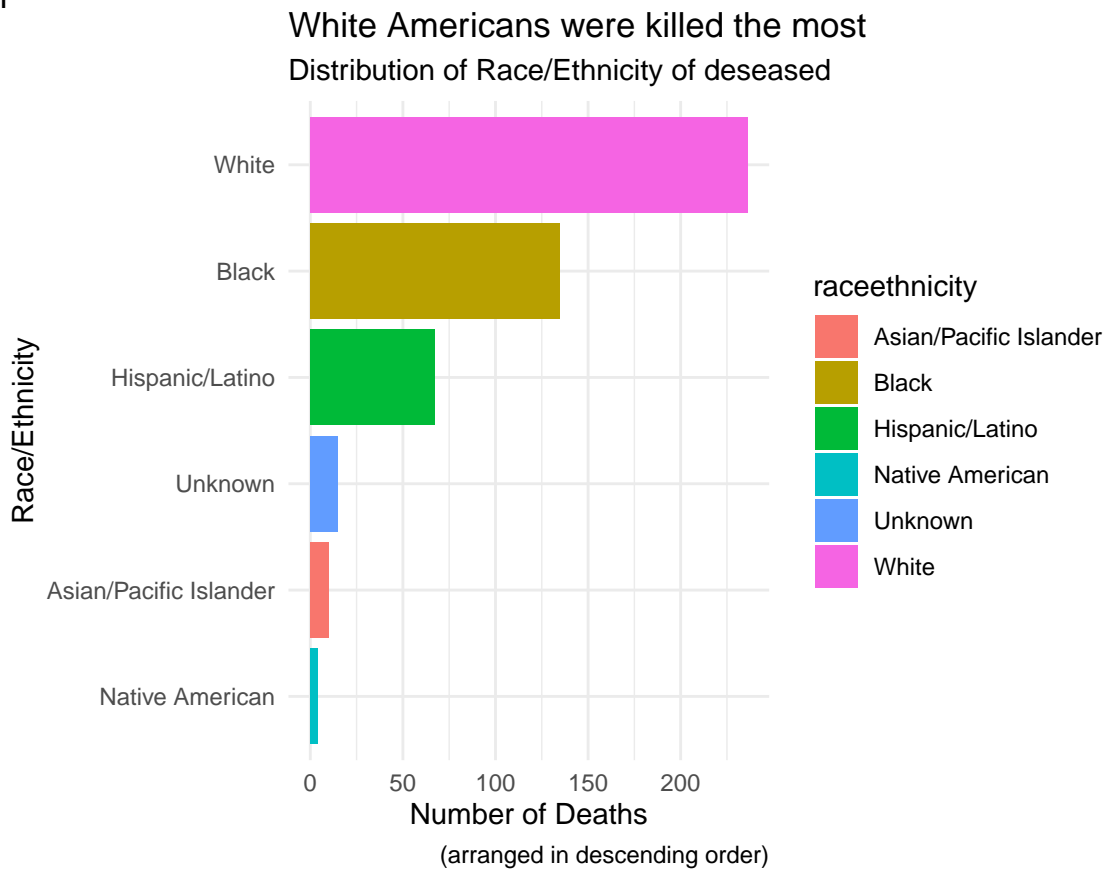
Exploratory Data Analysis

In this section, we will analyze the data, focusing on the key variables mentioned.

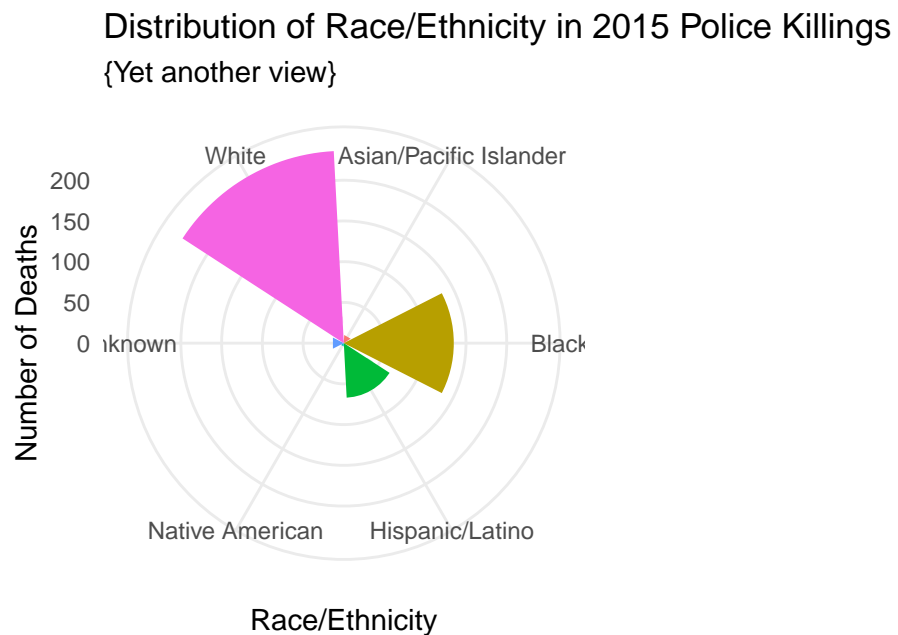
1. Distribution of Race/Ethnicity of Deceased

We’ll begin by analyzing the distribution of race and ethnicity in the dataset.

G 1.1



G 1.2

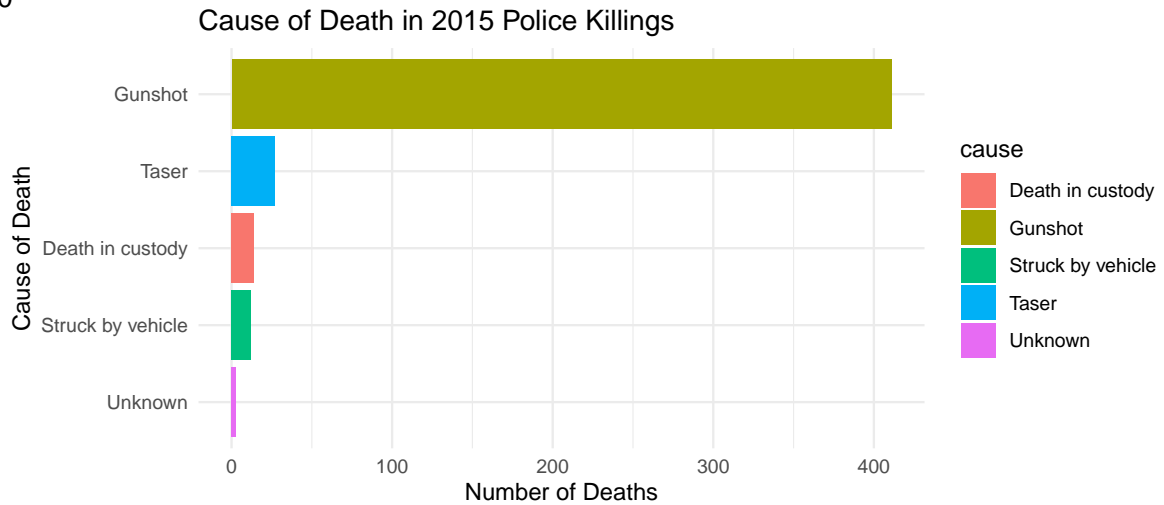


According to the graphs G1.1 and G1.2, it appears prima facie that most Americans who were killed in 2015 by the Police are Whites. Let us analyse more and see if this is the case.

2. Cause of Death Analysis

Next, we analyze the causes of death for victims.

G 2.0

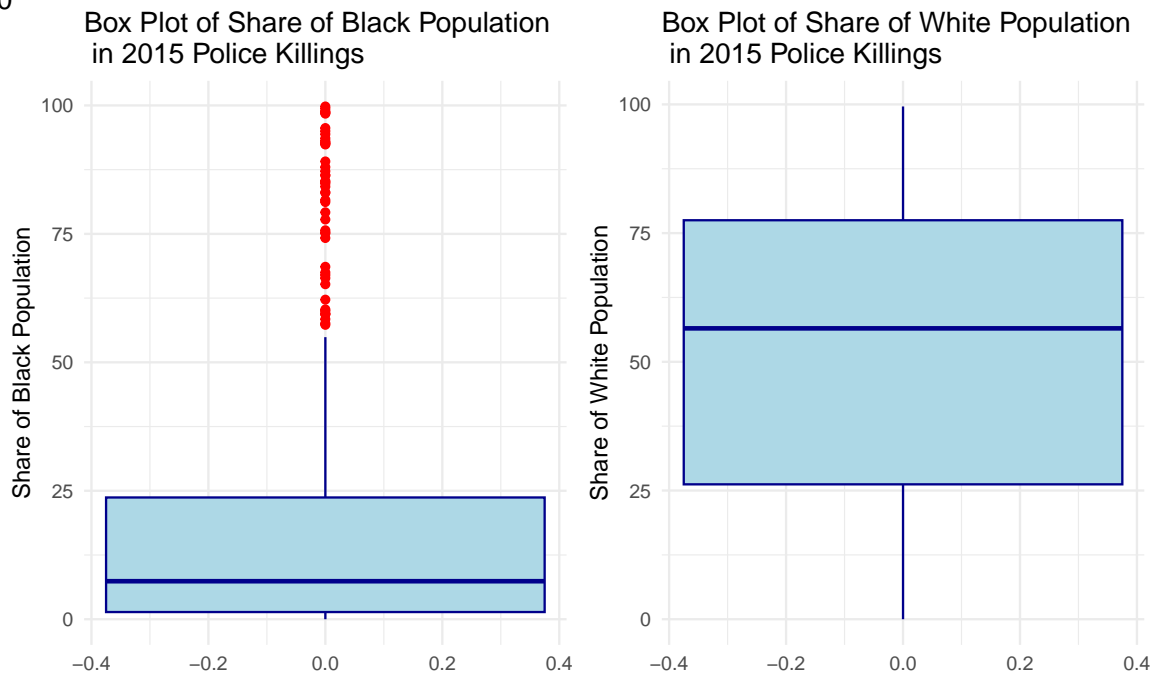


G 2.0 clearly suggests that most of the killings happened due to Gunshot. The other causes of death are negligible with respect to Gunshot.

3. Share of Black Population

We'll explore the distribution of the share of Black residents in the areas where the incidents occurred.

G 3.0



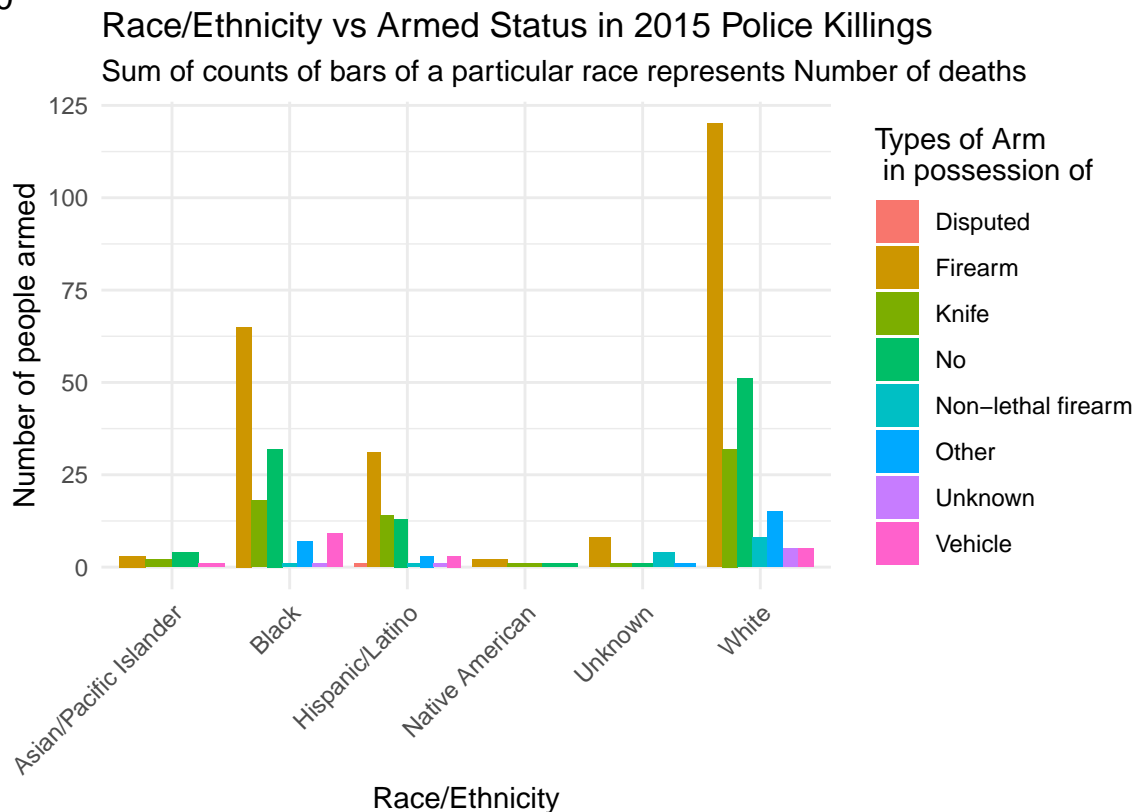
- G 3.0 suggests that if we look at a particular entry which represents a deceased, there is data for the area where he lives and the percentage of blacks and whites in that particular area. The box plot perfectly captures that information.

- Also, plotting the two side by side tells us that percentage of white people at any given geographic location where the incidents took place is always more than the blacks which defies in some sense our conclusions drawn from G 1.0

4. Race/Ethnicity vs Armed Status

Finally, we analyze the relationship between race/ethnicity and whether the deceased was armed.

G 4.0



The graph suggests that:

- if we look at a subset of people who were killed, namely the people who were deceased but also were armed, the maximum of those carried Firearm.
- So although we saw in G 1.1 and 1.2 that white Americans were killed the most, note that one of the major reasons they were was the fact that a major population of white Americans had arms as is suggested by G 4.0

Results and Discussion

The analysis reveals the following key insights:

1. **Race/Ethnicity Distribution:** The majority of victims were White, followed by Black and Hispanic/Latino. However, when taking into account the population sizes of the areas, Black people are disproportionately affected.

2. **Cause of Death:** Most deaths were caused by gunfire, which is consistent with police use of firearms during confrontations.
3. **Armed Status:** A significant portion of the victims were unarmed, raising important questions about the necessity of lethal force in these cases.
4. **Population and Share of Black Population:** The incidents often occurred in areas with a moderate to high share of Black residents, suggesting a racial component to the geographic distribution of police violence.

Conclusion

The analysis suggests that Black individuals were disproportionately affected by police killings in 2015. Further, a significant number of the deceased were unarmed, calling into question the justification for lethal force in many instances. These findings are consistent with broader concerns about racial disparities in policing practices in the United States.

References

1. *R for Data Science (2nd Edition)*, by Hadley Wickham and Garrett Grolemund. Available at: [R_Hadley](#)
2. Dataset: "Police Killings in 2015". Available at: [Kaggle](#)

Additional Notes:

1. *Visualizations:* This report utilizes various visualization techniques such as bar charts, boxplot and stacked bar chart.
2. *Population Analysis:* The share of the Black population (`share_black`) and that of White population (`share_white`) are analyzed to provide a more demographic context to the incidents.