

# Exploratory Data Analysis Project

Jazmin Borrayo

09/23/2024

```
"C:/Users/jazmi/OneDrive/Documents/math130/fatal-police-shootings-data.xlsx"
```

```
## [1] "C:/Users/jazmi/OneDrive/Documents/math130/fatal-police-shootings-data.xlsx"
```

```
library(ggplot2)  
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':  
##  
##   filter, lag
```

```
## The following objects are masked from 'package:base':  
##  
##   intersect, setdiff, setequal, union
```

```
library(knitr)  
knitr::opts_chunk$set(warning = FALSE, message = FALSE)
```

## Introduction

The data set this project analyzes is “Police Shootings”. This data consists of an excel file on police shootings in 2015 and was compiled by the Washington Post. This data set contains 14 variables and 3960 observations. Race, age, and gender are the variables that will be examined. The research questions are as follows: Is there a relationship between race, age, and gender regarding the likelihood of being shot by the police? Are certain groups more likely to be shot by the police than others?

```
library(readxl)  
police <- read_excel("C:/Users/jazmi/OneDrive/Documents/math130/fatal-police-shootings-data.xlsx",  
                    sheet=1, col_names=TRUE)  
police_filtered <- na.omit(police)
```

# Univariate Exploration

## Variable: Race

### Summary Statistics

The table below displays the summary statistics for race as a nominal variable, essentially what is considered a categorical variable.

```
table(police_filtered$race)
```

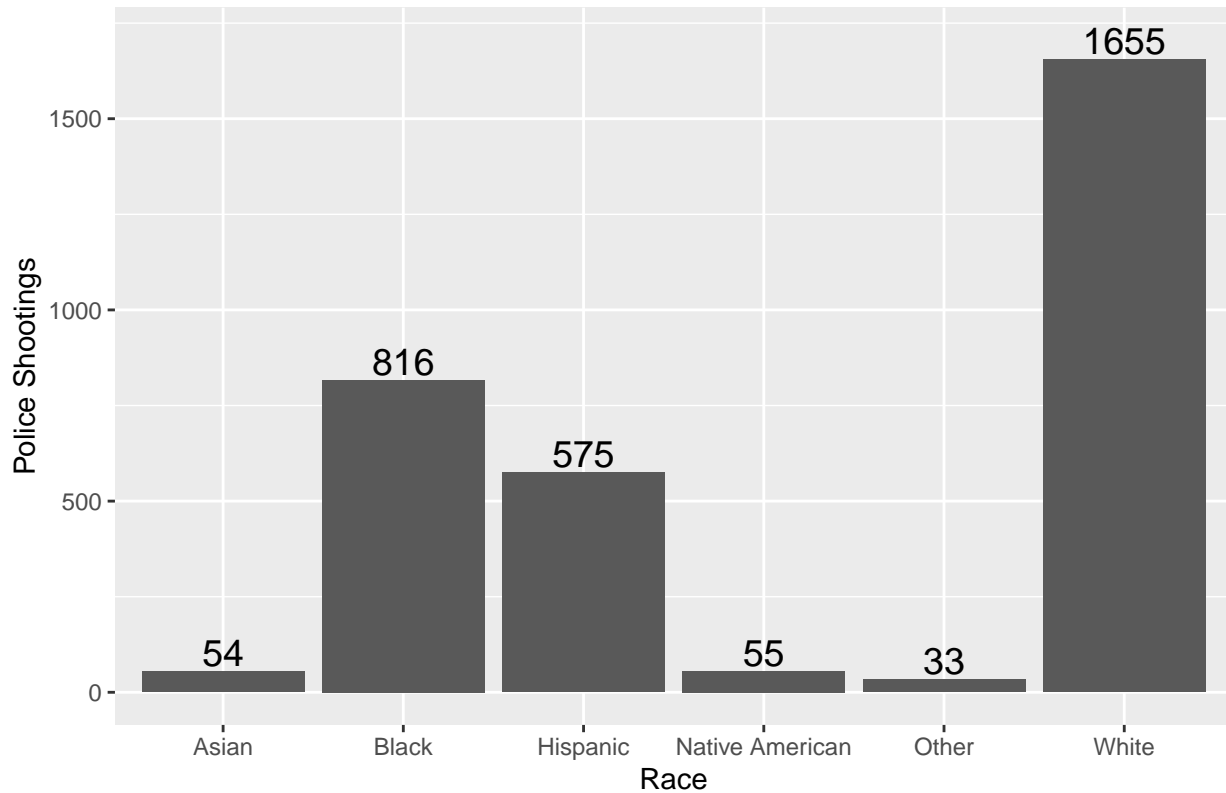
```
##  
##   A   B   H   N   O   W  
##  54 816 575  55  33 1655
```

```
police_filtered$race <- factor(police_filtered$race,  
                               labels=c("Asian", "Black", "Hispanic", "Native American", "Other", "White"))
```

### Graphs

```
ggplot(police_filtered, aes(x=race)) +  
  geom_bar() +  
  ylab("Police Shootings") + xlab("Race") +  
  ggtitle("Police Shootings Based on Race") +  
  geom_text(aes(y=after_stat(count) + 50, label=after_stat(count)), stat='count', size =5)
```

## Police Shootings Based on Race



This bar chart displays and compares the total amount of individuals by race involved in police shootings. The groups most impacted are “White”, followed by “Black” and “Hispanic”.

## Variable: Age

### Summary Statistics

The information below displays the summary statistics for age as a continuous numerical variable.

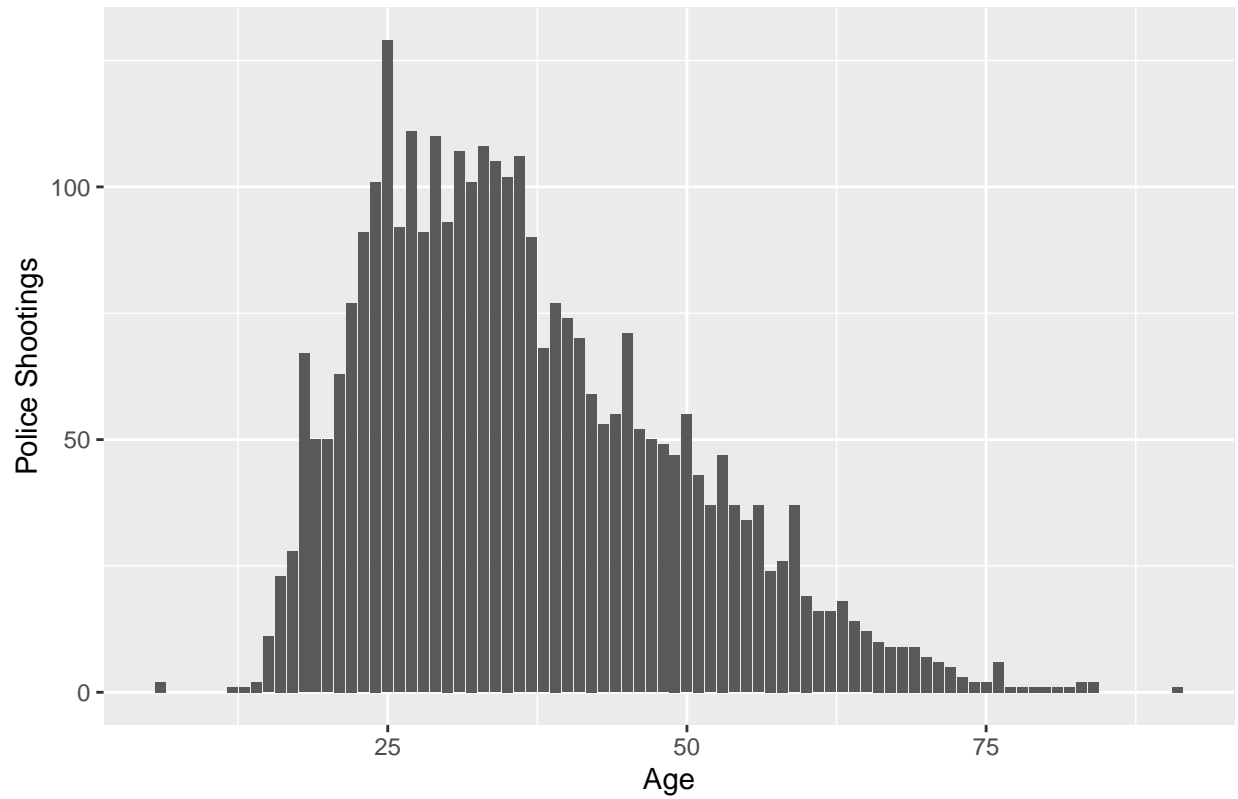
```
summary(police_filtered$age)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      6.00  27.00   34.00   36.52  45.00   91.00
```

### Graphs

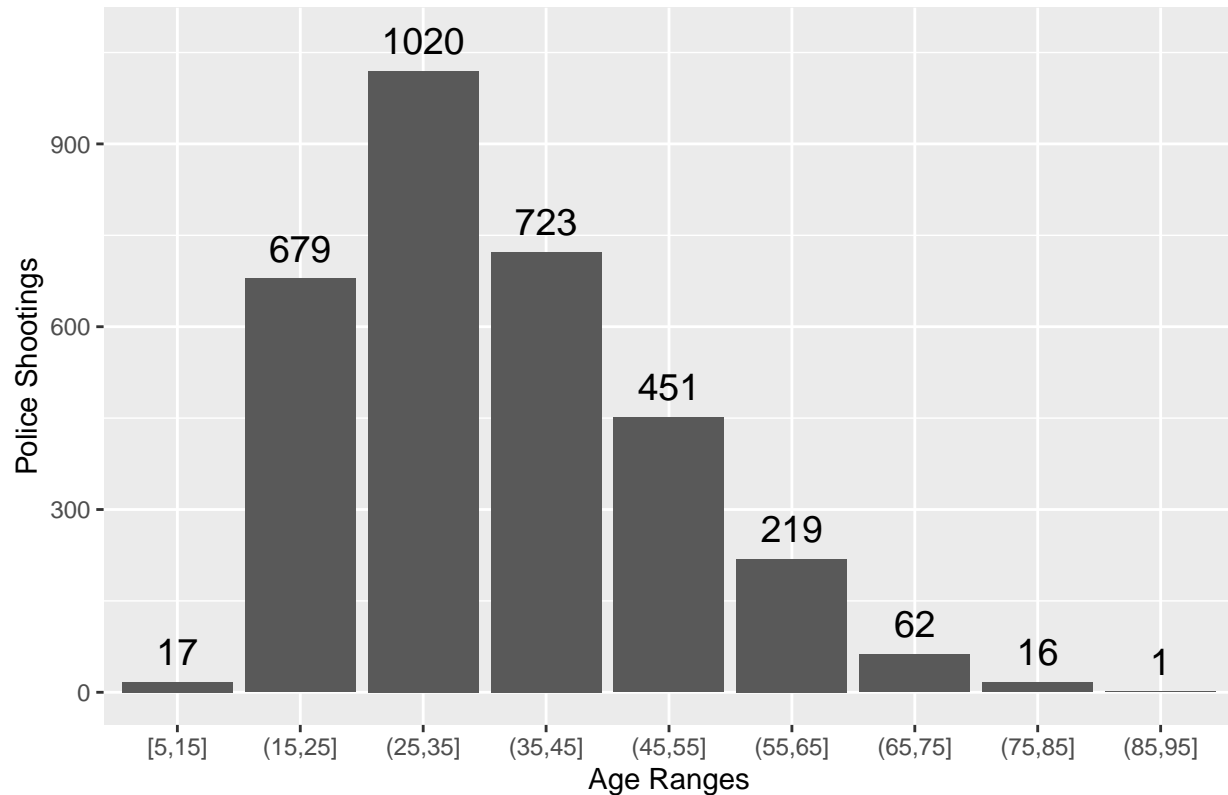
```
ggplot(police_filtered, aes(x = age)) +
  geom_bar() +
  ylab("Police Shootings") + xlab("Age") +
  ggtitle("Police Shootings Based on Age")
```

## Police Shootings Based on Age



```
police_filtered$age_range <- cut_width(police_filtered$age, width = 10)
ggplot(police_filtered, aes(x = age_range)) +
  geom_bar() +
  ylab("Police Shootings") + xlab("Age Ranges") +
  ggtitle("Police Shootings by Age Range")+
  geom_text(aes(y=after_stat(count) + 50, label=after_stat(count)), stat='count', size =5)
```

## Police Shootings by Age Range



These bar charts show the distribution of police shootings based on the age of victims. The first bar chart displays only the distribution of police shootings based on the age of the victims. The second bar chart is grouped into age ranges to display better trends of distribution of those involved in police shootings. The second chart shows higher trends among those between the ages of 15-45.

## Variable: Gender

### Summary Statistics

The table below shows the summary statistics for gender as a categorical variable, specifically a nominal variable.

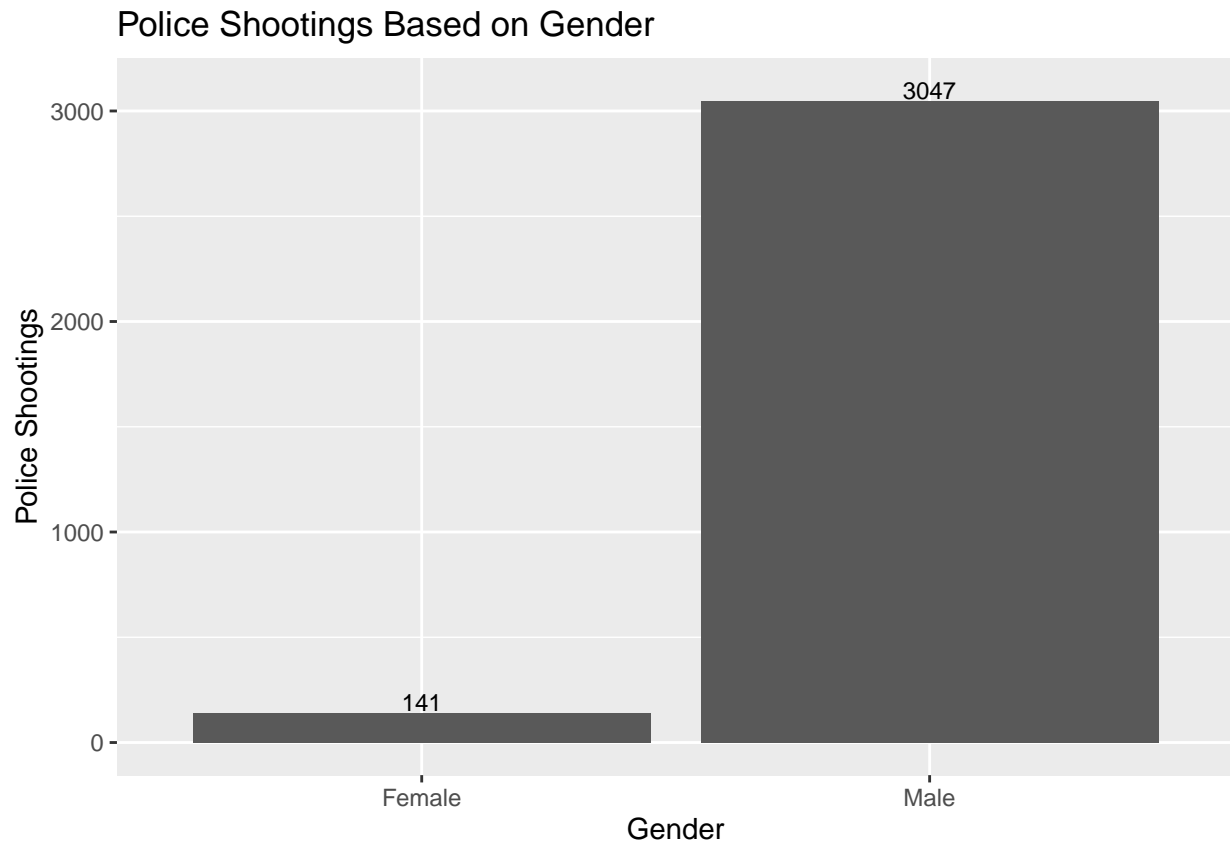
```
table(police_filtered$gender)
```

```
##  
##   F   M  
## 141 3047
```

```
police_filtered$gender <- factor(police_filtered$gender, labels=c("Female", "Male"))
```

### Graphs

```
ggplot(police_filtered, aes(gender)) +
  geom_bar() +
  ylab("Police Shootings") + xlab("Gender") +
  ggtitle("Police Shootings Based on Gender") +
  geom_text(aes(y=after_stat(count) + 50, label=after_stat(count)), stat='count', size =3)
```

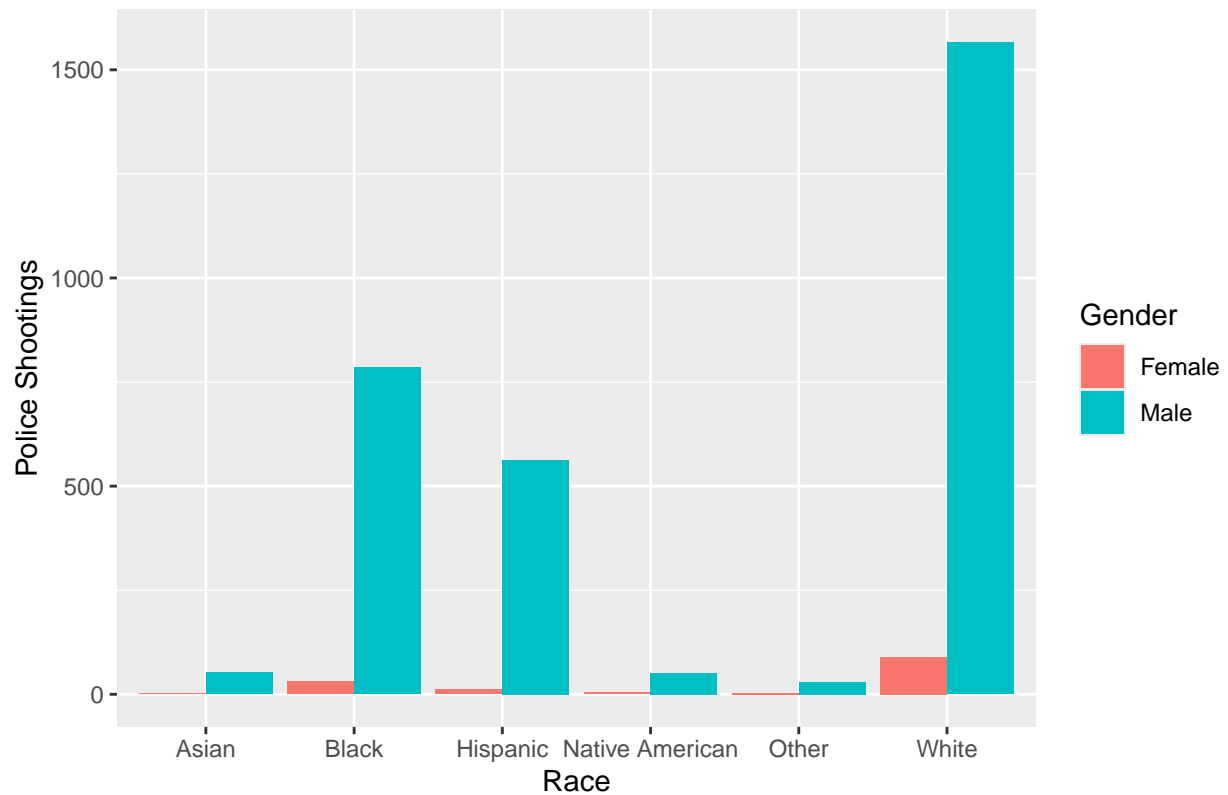


This bar chart compares the total amount of individuals involved in police shootings based on gender, which displays the disproportionate amount of men involved in police shootings.

## Bivariate Exploration

```
ggplot(police_filtered, aes(x=race, fill=gender)) +
  geom_bar(position = "dodge") +
  ylab("Police Shootings") + xlab("Race") +
  ggtitle("Police Shootings Based on Gender and Race") +
  scale_fill_discrete(name="Gender")
```

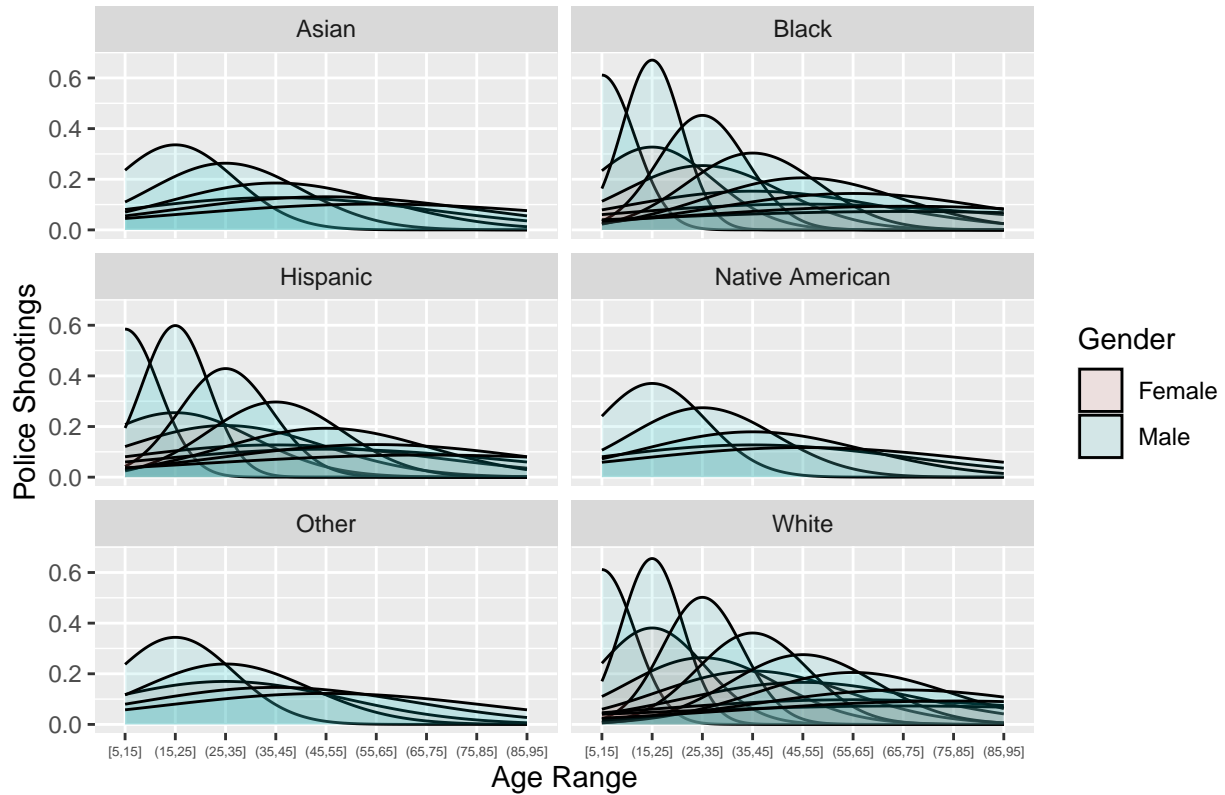
## Police Shootings Based on Gender and Race



This bar chart focuses specifically on gender and race as to see which groups are most commonly involved in police shootings. The results display that “White males” are the most likely group to be involved, then followed by “Black males” and “Hispanic males”.

```
ggplot(police_filtered, aes(x=age_range, fill=gender)) +  
  geom_density(alpha=.1) +  
  facet_wrap(~race, ncol=2) +  
  theme(axis.text.x=element_text(size=5)) +  
  ylab("Police Shootings") + xlab("Age Range") +  
  ggtitle("Police Shootings by Gender, Race, and Age Range") +  
  scale_fill_discrete(name="Gender")
```

## Police Shootings by Gender, Race, and Age Range



This graphic compares the distribution of individuals commonly involved in police shootings. The variables include gender and race compared against age groups. This panel displays two categorical variables against a continuous variable. The graphic has been sorted into individual density plots by race with gender and age range being examined within each plot. The results show that there is higher rates of involvement in police shootings among “White”, “Black”, and “Hispanic” males around the ages of 5-35. There is also a general pattern among younger males around the ages of 15-45 from all race groups to be involved in police shootings.

## Conclusion

This project analyzed the relationship between race, age, and gender regarding the likelihood of being shot by the police. The independent variables were examined and then compared against one another to identify if there is any correlation. The data collected shows that males from “White”, “Black”, and “Hispanic” groups within the ages of 5-35 are more likely to be involved in police shootings.