

# EDA Project

Asha Venegas

2024-09-23

## Introduction

The data set "Fatal Police Shootings" is a comprehensive look at police shootings resulting in death across the United States. The dataset includes children below the age of 18 as well as adults of varying races, armament and genders. In addition, the dataset provides a look at whether or not the police officer on the scene had an on and recording body camera. There are 3960 observations with 14 variables in the dataset. Focusing on only a few of these, there is expected to be a higher rate of those shot fatally for those who are young adult, black males and the police officer with the body camera off or unavailable.

```
summary(police_shootings)
```

```
##      id      name      date
## Min.   : 3   Length:3960   Min.   :2015-01-02 00:00:00
## 1st Qu.:1146  Class :character   1st Qu.:2015-12-28 18:00:00
## Median :2208  Mode  :character   Median :2017-01-07 12:00:00
## Mean   :2219                      Mean   :2016-12-31 04:52:00
## 3rd Qu.:3294                      3rd Qu.:2018-01-07 00:00:00
## Max.   :4361                      Max.   :2019-01-07 00:00:00
##
## manner_of_death  armed      age      gender
## Length:3960     Length:3960   Min.   : 6.00   Length:3960
## Class :character Class :character 1st Qu.:27.00   Class :character
## Mode  :character Mode  :character Median :35.00   Mode  :character
##
##                      Mean   :36.85
##                      3rd Qu.:45.00
##                      Max.   :91.00
##                      NA's   :152
##      race      city      state
## Length:3960     Length:3960   Length:3960
## Class :character Class :character Class :character
## Mode  :character Mode  :character Mode  :character
##
##
##
##
## signs_of_mental_illness threat_level      flee      body_camera
## Mode :logical          Length:3960     Length:3960     Mode :logical
## FALSE:3028             Class :character Class :character FALSE:3527
## TRUE :932              Mode  :character Mode  :character TRUE :433
##
##
##
##
```

```
police_shootings$Races <- fct_recode(police_shootings$race, "Asian" = "A", "Black" = "B", "Hispanic" = "H", "White" = "W",
"Native" = "N", "Other" = "O")
table(police_shootings$Races) %>% prop.table()
```

```
##
##      Asian      Black      Hispanic      Native      Other      White
## 0.01708205 0.25959115 0.18454215 0.01736208 0.01036124 0.51106133
```

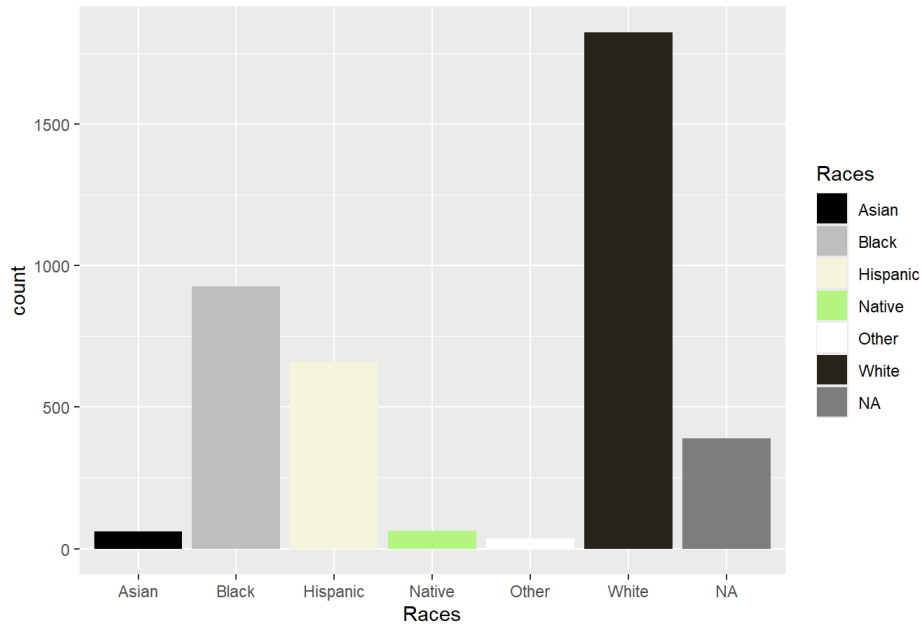
```
summary(police_shootings$Races)
```

```
##      Asian      Black      Hispanic      Native      Other      White      NA's
##      61       927       659       62       37       1825     389
```

```
library(RColorBrewer)
```

```
ggplot(police_shootings, aes(x=Races, fill=Races)) + geom_bar() +
  scale_fill_manual(values=c("black", "grey", "beige", "#B7F684", "white", "#28231D", "#FFFFCC")) + ggtitle("Summary of Race C
ount")
```

Summary of Race Count



Out of 3,960 recorded deaths, there are 927 recordings of those who are black and 3,033 non-black individuals.

##Bivariate Analysis

```
police_shootings$Races <- fct_recode(police_shootings$race, "Asian" = "A", "Black" = "B", "Hispanic" = "H", "White" = "W",
"Native" = "N", "Other" = "O")
table(police_shootings$Races)
```

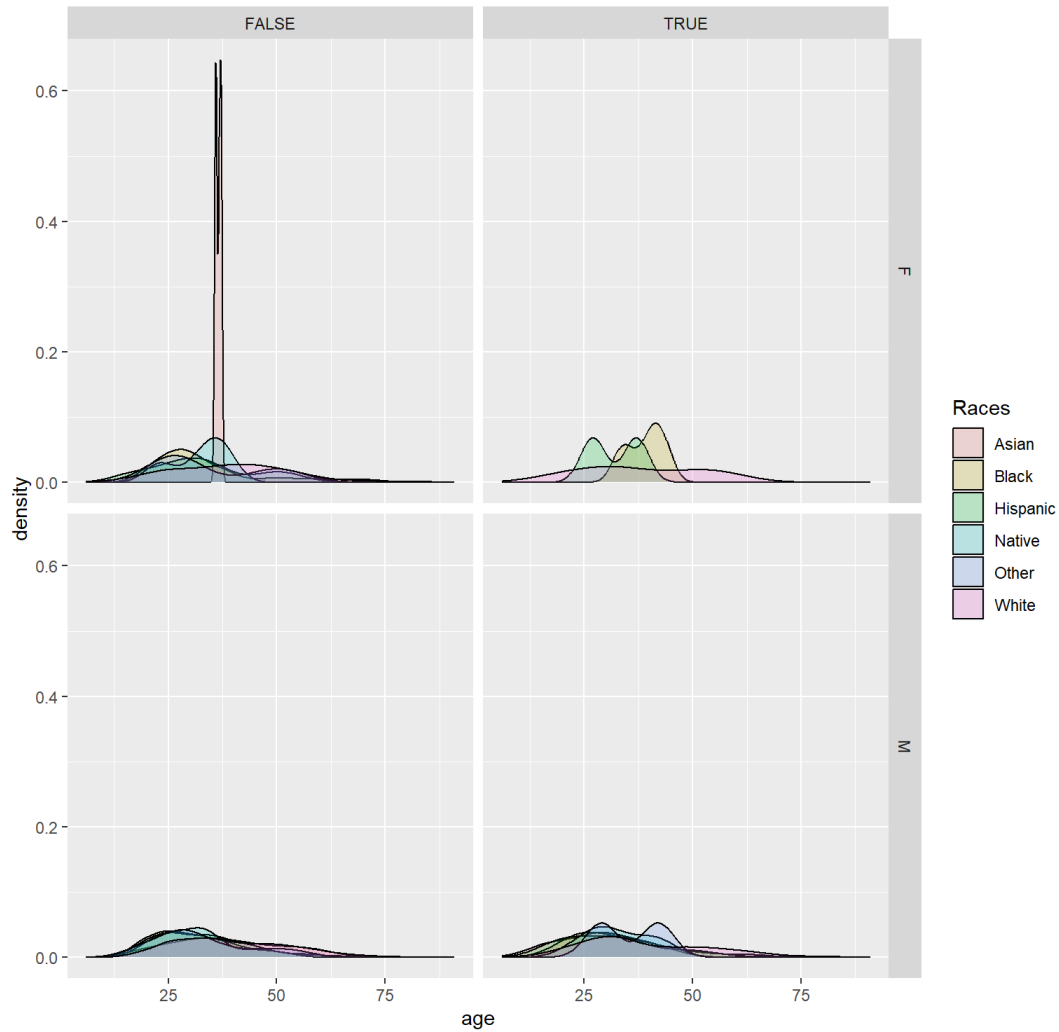
```
##
##   Asian   Black Hispanic   Native   Other   White
##     61    927    659      62     37    1825
```

```
police_shootings %>% select(age, Races , gender, body_camera) %>% na.omit() %>%
  ggplot(aes(x=age, fill=Races)) +
  geom_density(alpha=.2) + facet_grid(gender~body_camera) + ggtitle("Comparing Gender and Race to Police Body Cameras Being
On or Off")
```

```
## Warning: Groups with fewer than two data points have been dropped.
```

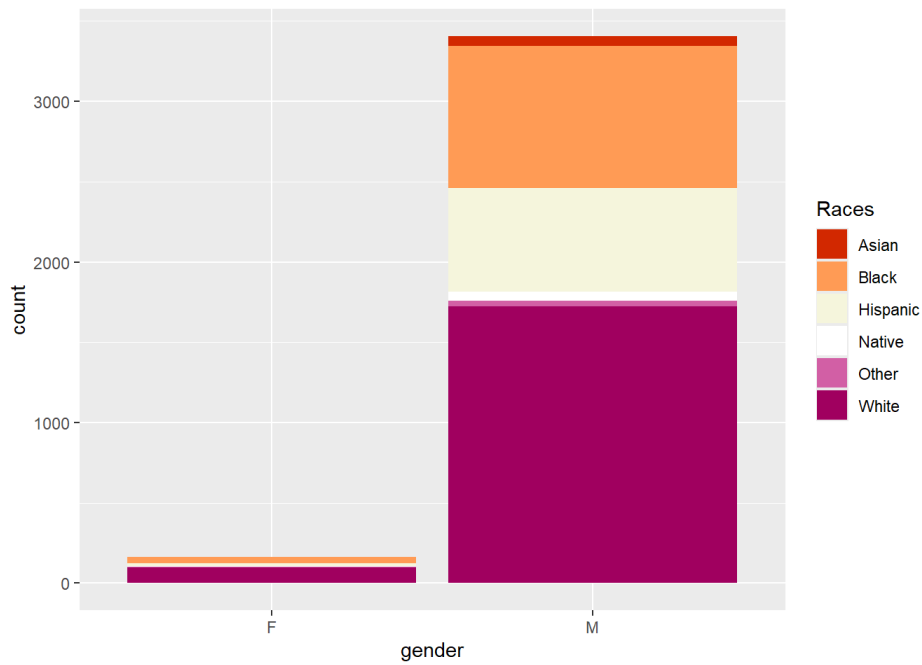
```
## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf
```

## Comparing Gender and Race to Police Body Cameras Being On or Off



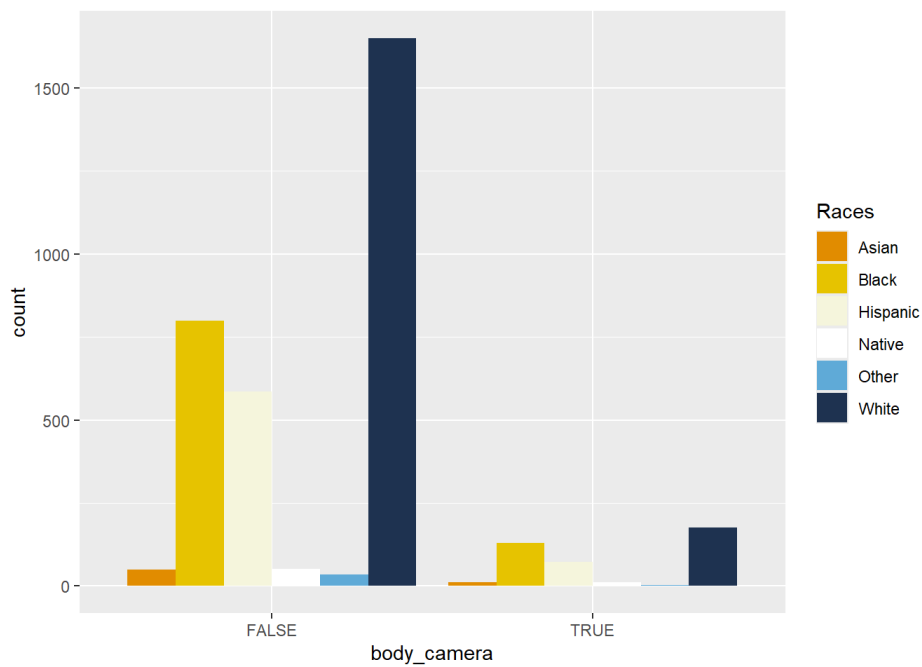
Comparing between genders (m=male, f=female), race, age, and whether police body cameras were on (TRUE) or off (FALSE), native men over the age of 25 had the highest deaths when police body cameras were off. Men of racial ethnicity labeled as "other" around the ages of 25 and mid 40's had the highest death count when body cameras were on. When body cameras were on, black women in their 40's were the most killed and asian women around the age of 30 were fatally shot the most when body cameras were off.

```
police_shootings %>% select(gender, Races) %>% na.omit %>% ggplot(aes(x=gender, fill=Races)) + geom_bar() + scale_fill_manual(values=c("#D62800", "#FF9B56", "beige", "white", "#D462A6", "#A40062"))
```



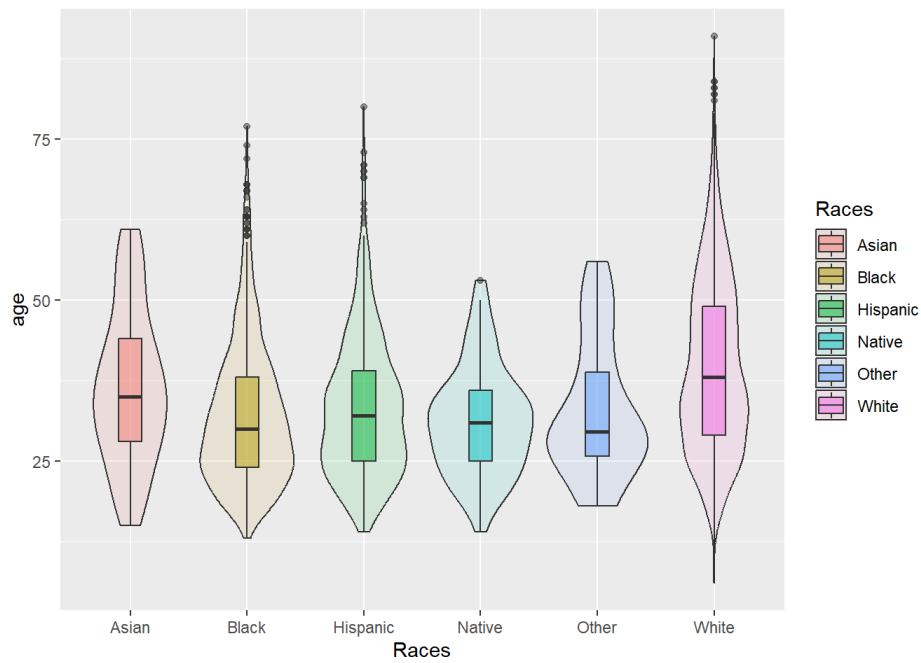
Between females and males of all races, men make up the majority of those shot dead in this dataset with the highest of those killed both men and women being black and white individuals.

```
police_shootings %>% select(body_camera, Races) %>% na.omit %>% ggplot(aes(x=body_camera, fill=Races)) + geom_bar(position = "dodge") + scale_fill_manual(values=c("#E38D00", "#E7C601", "beige", "white", "#5FAAD7", "#1F3554"))
```



White people are the majority of those shot with both body camera on and with police body camera off. Black people are second to white in both conditions.

```
police_shootings %>% select(Races, age) %>% na.omit %>% ggplot(aes(x=Races, y=age, fill=Races)) + geom_violin(alpha=.1) + geom_boxplot(alpha=.5, width=.2)
```



Black people are the second youngest group of individuals whom are fatally shot by police (around the age of 30). Those whose race did not match the available categories (“other”) are the youngest group around the ages of 27-28.

## Conclusion

Upon exploring gender, race, age, and police body camera functionality, multiple different groups came out as the greatest populations shot dead by police. Unlike my prediction, white men over the age of 40 made up the majority of those fatally shot by police. Male individuals whom are Native were the largest population shot by police whose body cameras were off. However, young black women and men often came out as second highest/largest in each univariate and bivariate analyses.