

# Exploratory Data Analysis

Thomas Delgado

2/23/2022

## Code Setup

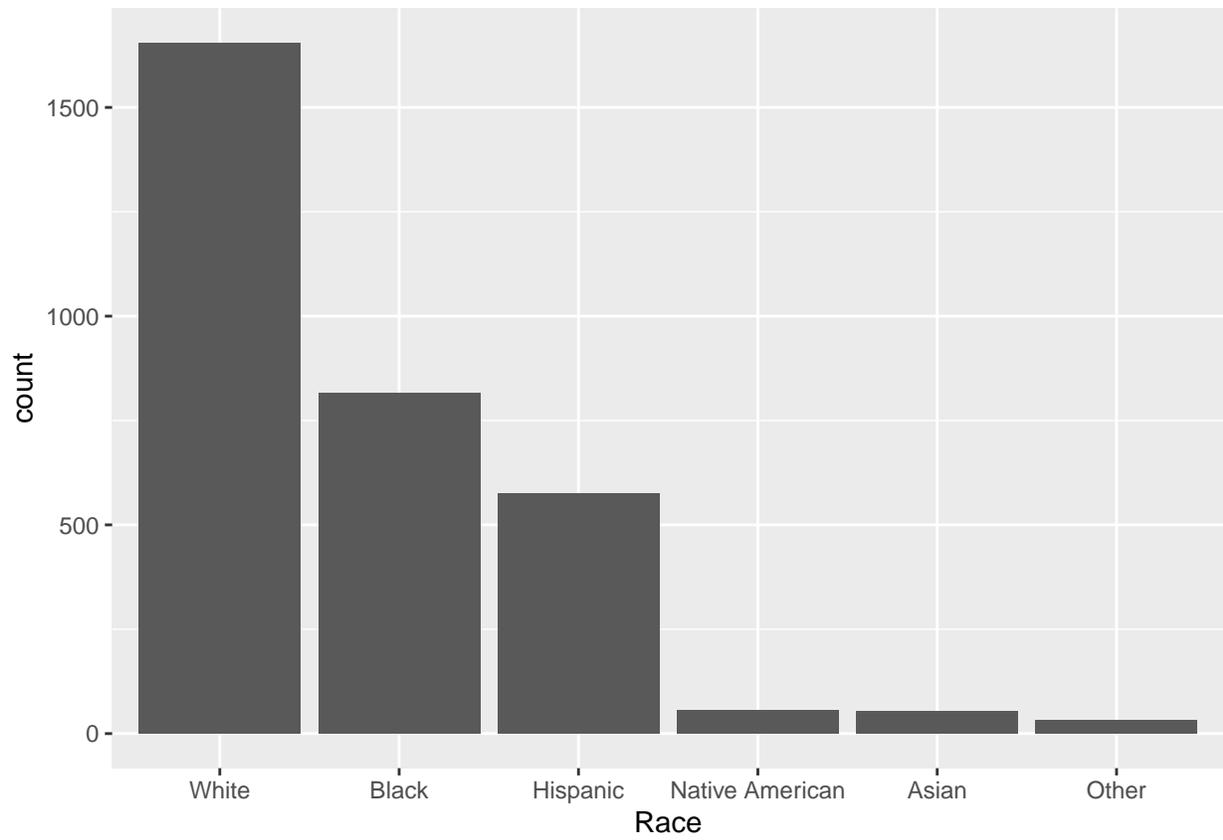
## Introduction

With the aftermath of the murder of George Floyd, much attention has turned to the use of fatal force by law enforcement officers. As a result, a focus on the use of body cameras by officers has arisen to provide an evidentiary basis for investigations pertaining to the use of fatal force by officers, especially in cases involving a person of color. For this data exploration, I will be using a record of fatal police shootings between 2015 and January of 2018 that has been compiled by *The Washington Post* to see if there is a relationship between the assigned race of a victim and whether or not a body camera was being used.

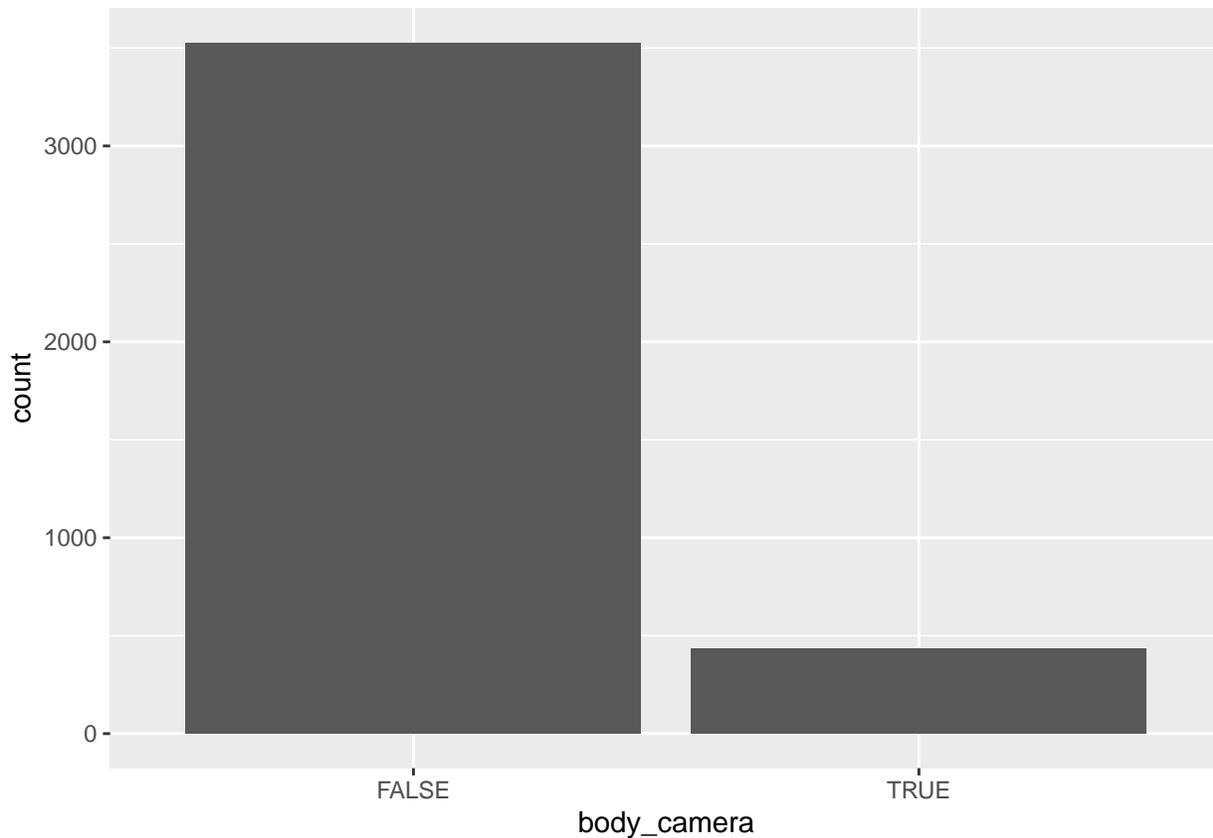
## Univariate exploration

First, we will look at the overall rates and distribution of body camera use and the demographics of those being exposed to fatal force.

```
shootings$race_exp <- fct_recode(shootings$race, "Asian" = "A", "Black" = "B", "Hispanic" = "H", "Native  
shootings %>%  
  na.omit(race_exp) %>%  
  ggplot(aes(x=forcats::fct_infreq(race_exp))) +  
  geom_bar() +  
  xlab("Race")
```



```
ggplot(shootings, aes(x=body_camera)) + geom_bar()
```

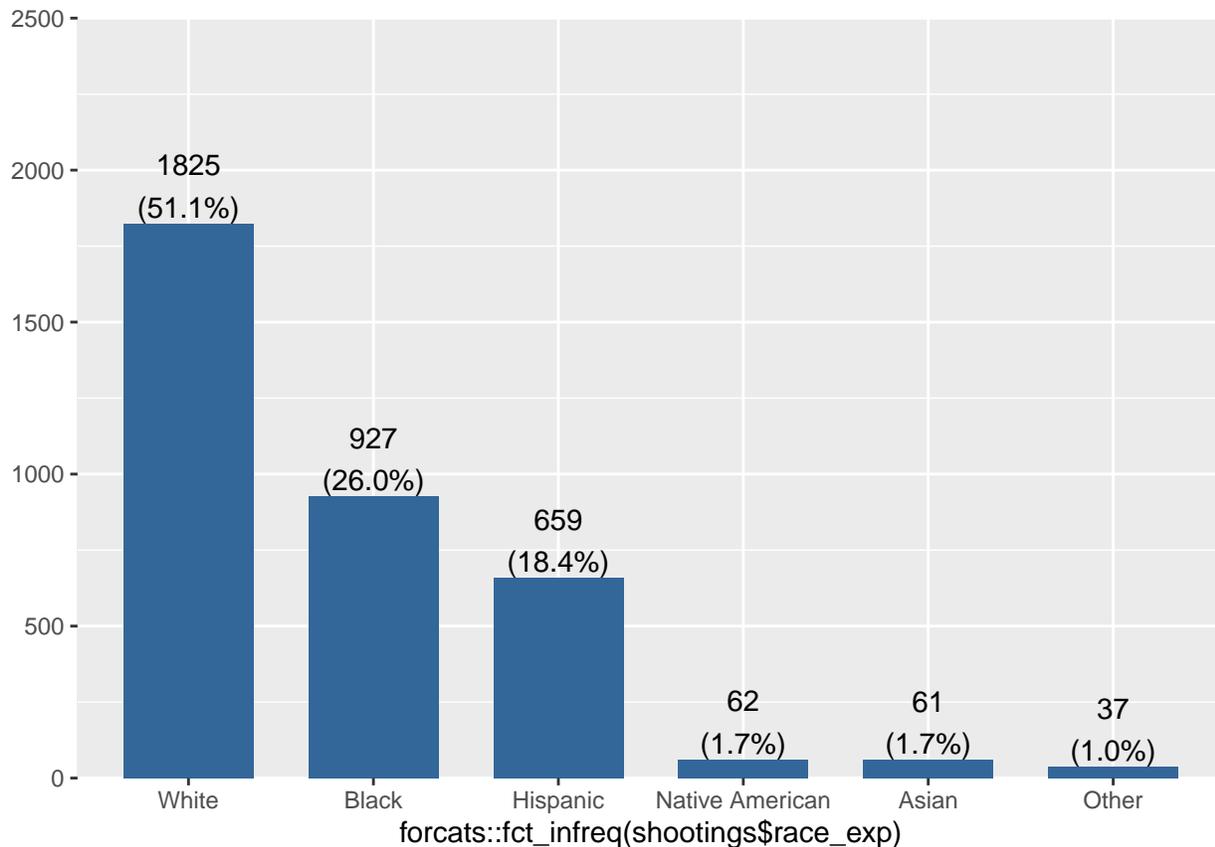


Based on this initial exploration, it appears that most of the fatal shootings involved white individuals and the majority of these incidents did not have a body camera in use when the incident took place. This is consistent with the population demographics of the United States and the lack of federal laws requiring the use of body cameras for officers. When considering population demographics however, the raw number of the population does not provide the context in which individuals not belonging to the majority experience certain effects. For this, a comparison of population proportion provides greater insight.

```
table(shootings$race_exp) %>%
  prop.table() %>%
  round(3)
```

```
##
##      Asian      Black  Hispanic Native American      Other
##      0.017      0.260      0.185      0.017      0.010
##      White
##      0.511
```

```
plot_frq(forcats::fct_infreq(shootings$race_exp))
```



Here, we see the demographic proportion of victims of fatal police shootings. While Blacks and Hispanics characterize 26% and 18.5% of the fatal shootings respectively, in 2016 those same populations only characterize 13.3% and 17.8% of the overall United States population. This highlights that there is an over-representation of people of color in fatal police shootings.

## Multivariate exploration

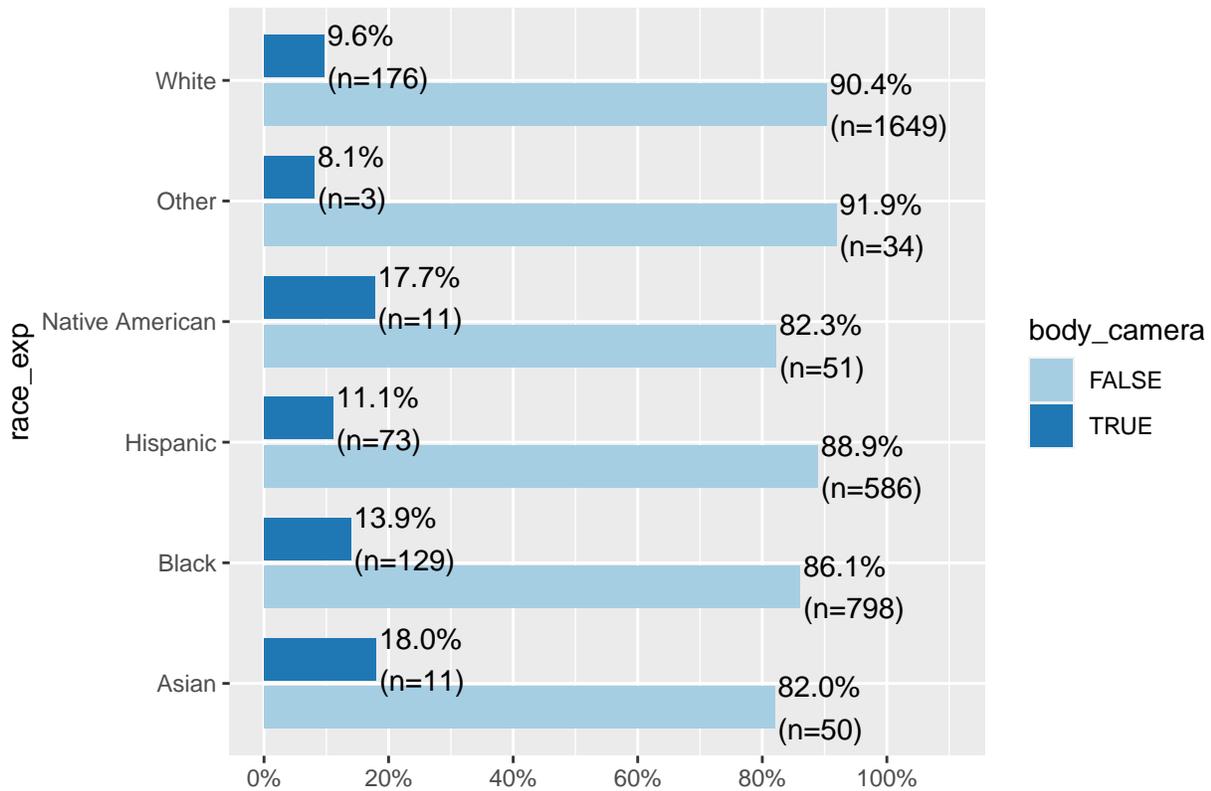
Since we can see that people of color are more likely to fall victim to a fatal shooting, let us explore whether or not there is a correlation in the lack of body camera footage when race is considered.

```
table(shootings$race_exp, shootings$body_camera)
```

```
##
##           FALSE TRUE
## Asian           50  11
## Black           798 129
## Hispanic        586  73
## Native American  51  11
## Other           34   3
## White          1649 176
```

```
plot_xtab(shootings$race_exp, shootings$body_camera,
  margin="row",
  coord.flip = "TRUE",
  expand.grid = "TRUE",
  title = "Body Camera Use by Race")
```

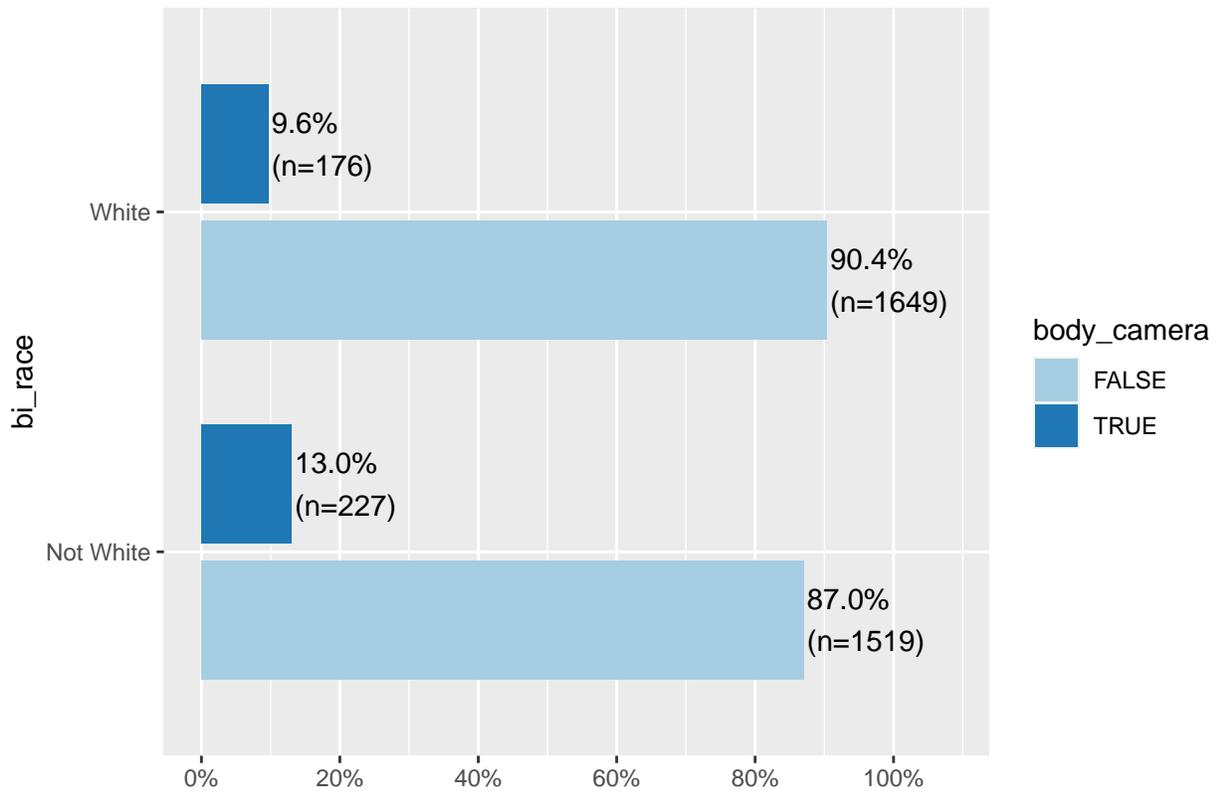
## Body Camera Use by Race



Again, we see that there is an overall lack of body camera use in the event of fatal shootings. Based on the data proportions, white and “other” victims exhibit the lowest rates of body camera use by law enforcement. When considering the raw counts however, white victims have the highest prevalence of body camera usage from the other race categories. To better visualize the potential for racial disparity, let us collapse the provided race categories into “white” and “not white.” With this, we see that the number of fatal shootings becomes more similar than when more discreet racial categories were used and that there appears to be a higher prevalence of body camera use if the victim is not white.

```
shootings$bi_race <- fct_collapse(shootings$race, "Not White" = c("A", "B", "H", "N", "O"), "White" = "W")
plot_xtab(shootings$bi_race, shootings$body_camera,
  margin="row",
  coord.flip = "TRUE",
  expand.grid = "TRUE",
  title = "Body Camera Use by Race")
```

## Body Camera Use by Race



With this, we see that the number of fatal shootings becomes more similar than when more discreet racial categories were used and that there appears to be a higher prevalence of body camera use if the victim is not white.

## Conclusion

From this introductory analysis we can see that people of color are more likely to become victims of an officer involved shooting, but that body camera use does not appear to vary substantially when solely race is considered. This is further highlighted when racial categories are collapsed. However, if the data is viewed solely by raw counts, then it appears that white individuals are over represented in fatal shootings but are the most likely to have body camera footage recorded. Ultimately, this highlights the multiple ways in which data can be viewed and manipulated - especially on an exploratory level such as this - to fit a particular narrative. In an age of ever accessible information, interrogating the way in which data is being presented is fundamental to media literacy. Overall, based on this level of data analysis, there is likely not a correlation between body camera use and race of a victim. Instead, due to the overwhelming lack of body camera usage, there is likely only a circumstantial connection based on the over representation of people of color.