US Police Fatal Shooting Data Analysis

Group Jauguar: Oumnia Baqacem, Wen Lin, Serena Yin

2022-12-14

I. Introduction

The goal of the project is to analyze data on fatal shootings by police officers in the United States. The analysis aims to understand the factors that play a role in such incidents. Factors we analyzed include Race, Age, Gender, Presence of Officer Body Camera, Casualties' Possession of Weapons, and Casualties' State of Mental illness.

The phenomenon under study is fatal shootings by police in the United States. Fatal shootings by police have been a point of huge controversy in the United States for many years, with concerns about excessive use of force and potential bias playing a significant role in national debates. In recent years, especially as social media has brought greater attention to these incidents, many call for greater transparency and accountability amongst law enforcement

Using a dataset from the Washington Post, which tracks details about each police-involved killing in the United States since 2015 we will explore any relationships between specific demographics of victims and police shootings. We would also look at some variables about the circumstances of the shootings and discuss if they are related to police fatal shootings.

II. Background

The dataset we are analyzing is the: *'data-police-shootings/fatal-police-shootings-data.csv'*

Published by the Washington Post, this data set tracks details about each police-involved killing in the United States. The dataset is regularly updated as new killings occur. For the purposes of this project, we limited the date range to be from the years 2015 to 2020.

The raw data set has 7891 observations across 17 different variables. When limited to our chosen time frame the observations decrease to 59444. The unit of observation is the individual, and each individual is assessed at various levels including both details about the deceased's demographics and circumstances of the shooting.

```
## Rows: 5,944
## Columns: 17
## $ id                      <int> 3, 4, 5, 8, 9, 11, 13, 15, 16, 17, 19, 21, 22,~
## $ name                    <chr> "Tim Elliot", "Lewis Lee Lembke", "John Paul Q~
## $ date                    <chr> "2015-01-02", "2015-01-02", "2015-01-03", "201~
## $ manner_of_death         <chr> "shot", "shot", "shot and Tasered", "shot", "s~
## $ armed                   <chr> "gun", "gun", "unarmed", "toy weapon", "nail g~
## $ age                     <int> 53, 47, 23, 32, 39, 18, 22, 35, 34, 47, 25, 31~
## $ gender                  <chr> "M", "M", "M", "M", "M", "M", "M", "M", "F", "~
## $ race                    <chr> "A", "W", "H", "W", "H", "W", "H", "W", "W", "~
## $ city                    <chr> "Shelton", "Aloha", "Wichita", "San Francisco"~
```

```
## $ state                   <chr> "WA", "OR", "KS", "CA", "CO", "OK", "AZ", "KS"~
## $ signs_of_mental_illness <chr> "True", "False", "False", "True", "False", "Fa~
## $ threat_level            <chr> "attack", "attack", "other", "attack", "attack~
## $ flee                    <chr> "Not fleeing", "Not fleeing", "Not fleeing", "~
## $ body_camera             <chr> "False", "False", "False", "False", "False", "~
## $ longitude               <dbl> -123.122, -122.892, -97.281, -122.422, -104.69~
## $ latitude                <dbl> 47.247, 45.487, 37.695, 37.763, 40.384, 35.877~
## $ is_geocoding_exact      <chr> "True", "True", "True", "True", "True", "True"~
```

The Variables at which each individual is assessed include: Id, Name, Date, Manner of Death, If the victim was armed, Gender, Race, City, State, Signs of Mental Illness, Threat Level, Whether the victim fled, If the shooting officer wore a body camera., longitude, latitude, and if the geocoding is exact.

Fatal police shootings have always been an ongoing event in the US, that results in nearly 1000 deaths annually since 2015, as documented by the Washington Post. Our topic aims to analyze the statistical data and discover trends, if any, in fatal police shootings. The dataset only documents shooting events that lead to civilian death. So in the event that someone got shot and did not die, it will not be part of this dataset. Another important thing to note is that only shootings involving a police officer in the line of duty will be recorded here. Since all records involve shooting, our EDA revolves more around the different factors (if a civilian is armed, if the officer is wearing a body camera, etc) that corresponded to fatal shootings because we lack data on cases where the police offer does not shoot (which we expect there to be an enormous amount of observations). Therefore, we are unable to analyze what factors lead to a fatal shooting and instead can only analyze the occurrence of factors in the event of a shooting.

III. Data Wrangling

Below we showed our code for data wrangling for each plot and commented the code for explanation.

Gender

```
plotdata_gender <- data2 %>%
  filter(gender != "") %>%
  group_by(gender) %>% # this groups the data by gender
  select(gender) %>%
  summarize(count =n()) # counts how many occurrences of each gender is recorded
```

Age

```
plotdata_age <- data2 %>%
# create a variable called age_group
  mutate(age_group = case_when(
    age %in% seq(10,14) ~ "10-14",
    age %in% seq(15,19) ~ "15-19",
    age %in% seq(20,24) ~ "20-24",
    age %in% seq(25,29) ~ "25-29",
    age %in% seq(30,34) ~ "30-34",
    age %in% seq(35,39) ~ "35-39",
    age %in% seq(40,44) ~ "40-44",
```

```
    age %in% seq(45,49) ~ "45-49",
    age %in% seq(50,54) ~ "50-54",
    age %in% seq(55,59) ~ "55-59",
    age %in% seq(60,64) ~ "60-64",
    age %in% seq(65,69) ~ "65-69",
    age %in% seq(70,74) ~ "70-74",
    age %in% seq(75,79) ~ "75-79",
    age %in% seq(80,84) ~ "80-84",
    age %in% seq(85,89) ~ "85-89",
    age %in% seq(90,94) ~ "90-94"
  )) %>%
  # divide age group based on age

  filter(!is.na(age_group))
# filter out observations that have NA vales for age_group
```

Race

```
plot_race <- data2 %>%
  filter(race != "" & race !="N") %>% # remove the samples with blank or unknown race
  group_by(race) %>%
  summarize(count = n()) %>% # count the occurrence of each race
  mutate(race = (case_when(
    race == "A" ~ "Asian",
    race == "B" ~ "Black",
    race == "W" ~ "White",
    race == "H" ~ "Hispanic",
    race == "O" ~ "Other"
  ))) %>% # rename the race so it's easier to read
  arrange(desc(count))
```

Race Density

note for below, we created a table for the US race distribution percentage in 2020 using data from the US Census Bureau

```
plot_race_den <- race_den %>% # plot data for per capita race
  left_join(plot_race, by = "race") %>%
  # combine the percentage with the dataset grouped by race
  mutate(population_mill = 329.5 * percent) %>%
  # multiply the percentage of each race by the total US population in 2020 to obtain the population in
  mutate(density = count/population_mill)
```

State of Being Armed

```
plotdata_armed <- data2 %>%
  group_by(armed) %>%
  filter(armed == "gun" | armed == "knife" | armed == "unarmed")%>%
```

```
  #filtering by the variables we are interested in  looking at
  summarize(count=n()) # counting the number of occurrences
```

State of Armed by Race per million capita

```
plot_armed_race <- data2 %>%
  filter(race != "" & race !="N" & race != "O") %>% # filter out the N/A, unknown, and Other races
  filter(armed == "gun" | armed == "knife" | armed == "unarmed") %>%
  # filter status to only armed with gun or knife or unarmed
  group_by(race, armed) %>%
  summarize(count = n()) %>%
  mutate(race = (case_when( # rename for better presentation
    race == "A" ~ "Asian",
    race == "B" ~ "Black",
    race == "W" ~ "White",
    race == "H" ~ "Hispanic",
  ))) %>%
  right_join(race_den, by = "race") %>%
  mutate(population_mill = 329.5 * percent) %>% # calculate total population for each race
  mutate(per_mill_capita = count/population_mill) # calculate respective per capita
```

```
## `summarise()` has grouped output by 'race'. You can override using the
## `.groups` argument.
```

Signs of Mental IIness

```
plotdata_mental <- data2 %>% # data for signs of mental illness
  mutate (mental_illness = if_else(signs_of_mental_illness=="True",
                                   "Present",
                                   "Not Present"))
# create a variable called mental_illness
# if value for signs_of_mental_illness = "True", mental_illness = "Present"
# if value for signs_of_mental_illness != "True", mental_illness = "Not Present"
```

Presence of a Police Body Camera

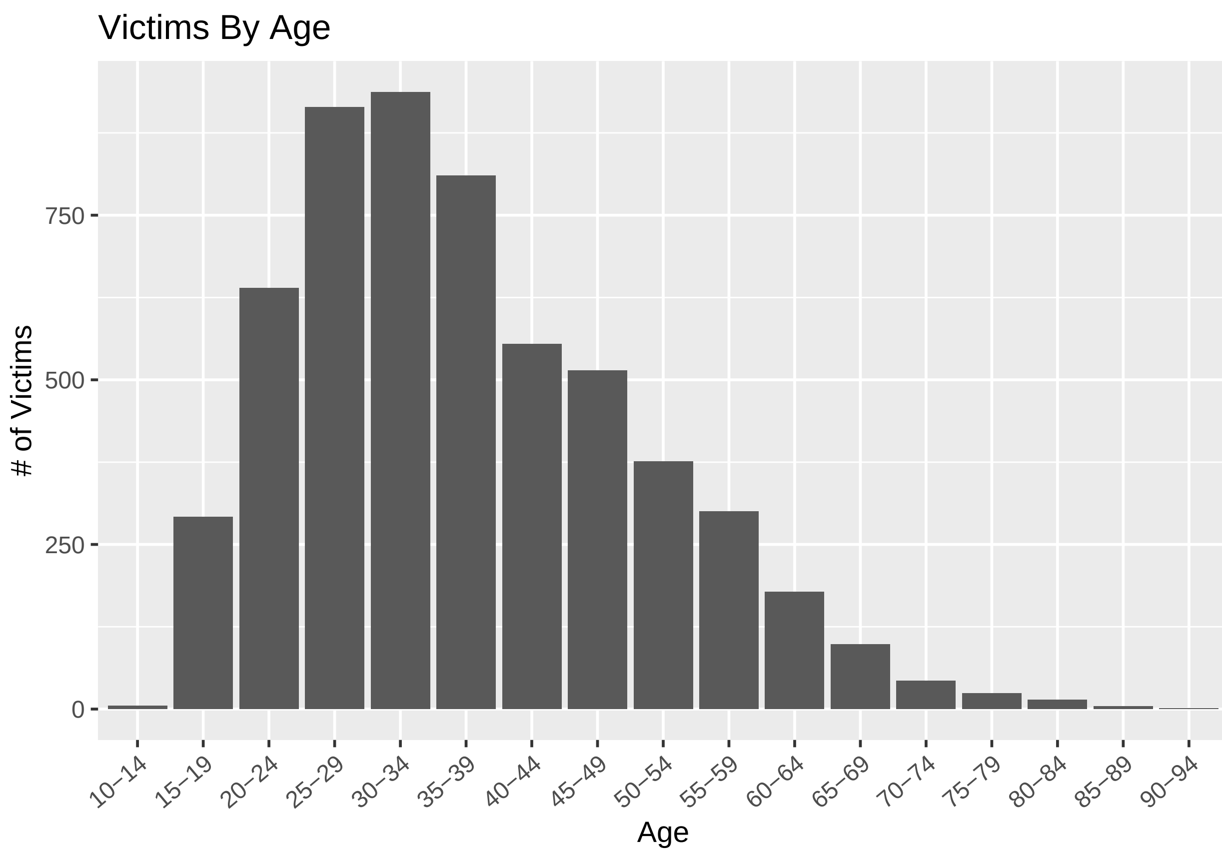
```
plotdata_camera <- data2 %>% #data for body_camera
  mutate (body_camera = if_else(body_camera=="True","Camera Present", "Camera NOT Present"))
 # if the value is "True", make it "Camera Present"
 # if the value is not "True", make it "Camera NOT Present"
```

State Body Camera Funding

We did not include the code for this plot data (called `camera_funding`) because of its length. In summary, we created a boolean variable `camera_funding` that indicates if a state funds body camera or not. We then manually inserted the value for each of the 50 states using the `case_when` function, which is why the code is long(check appendix if needed). Lastly we renamed the values in the column to “Funded” or “Not Funded” depending on its respective boolean value.

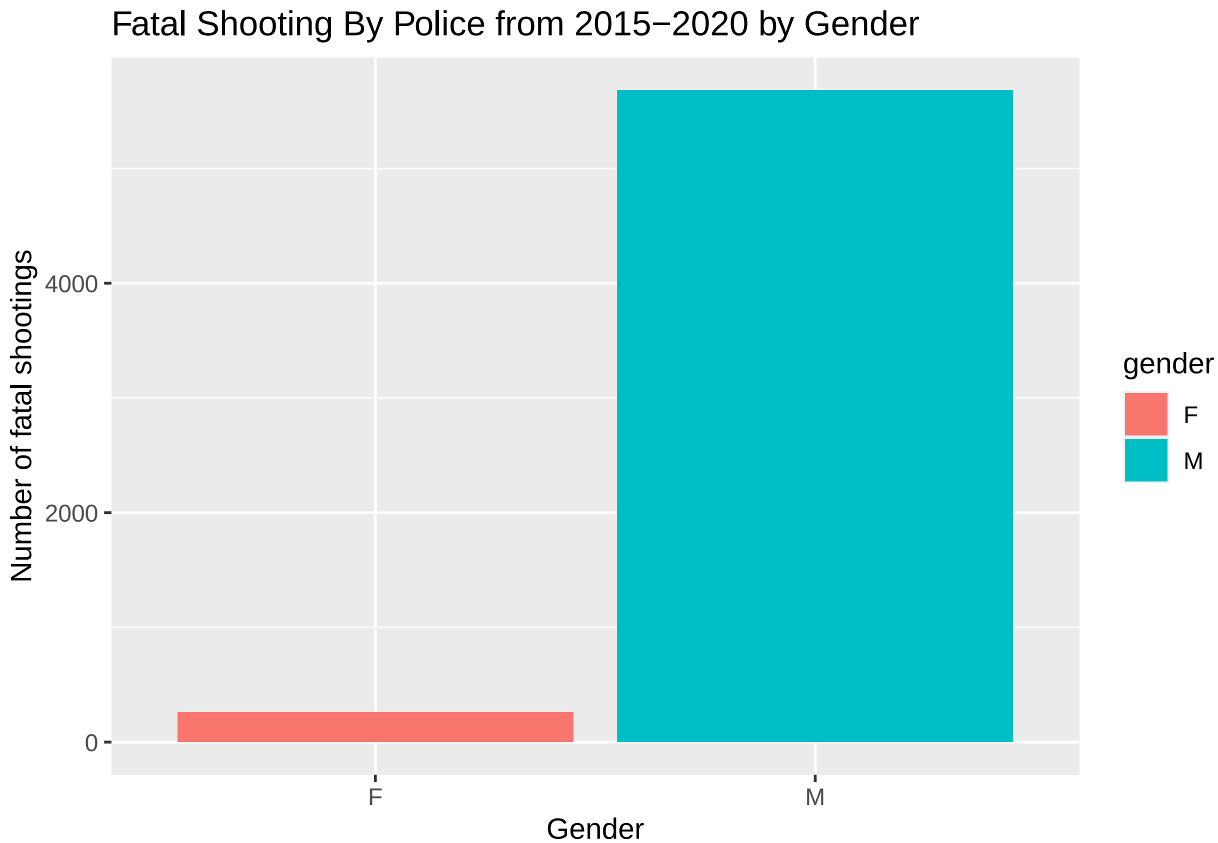
IV. Exploratory Analysis

Age Plot



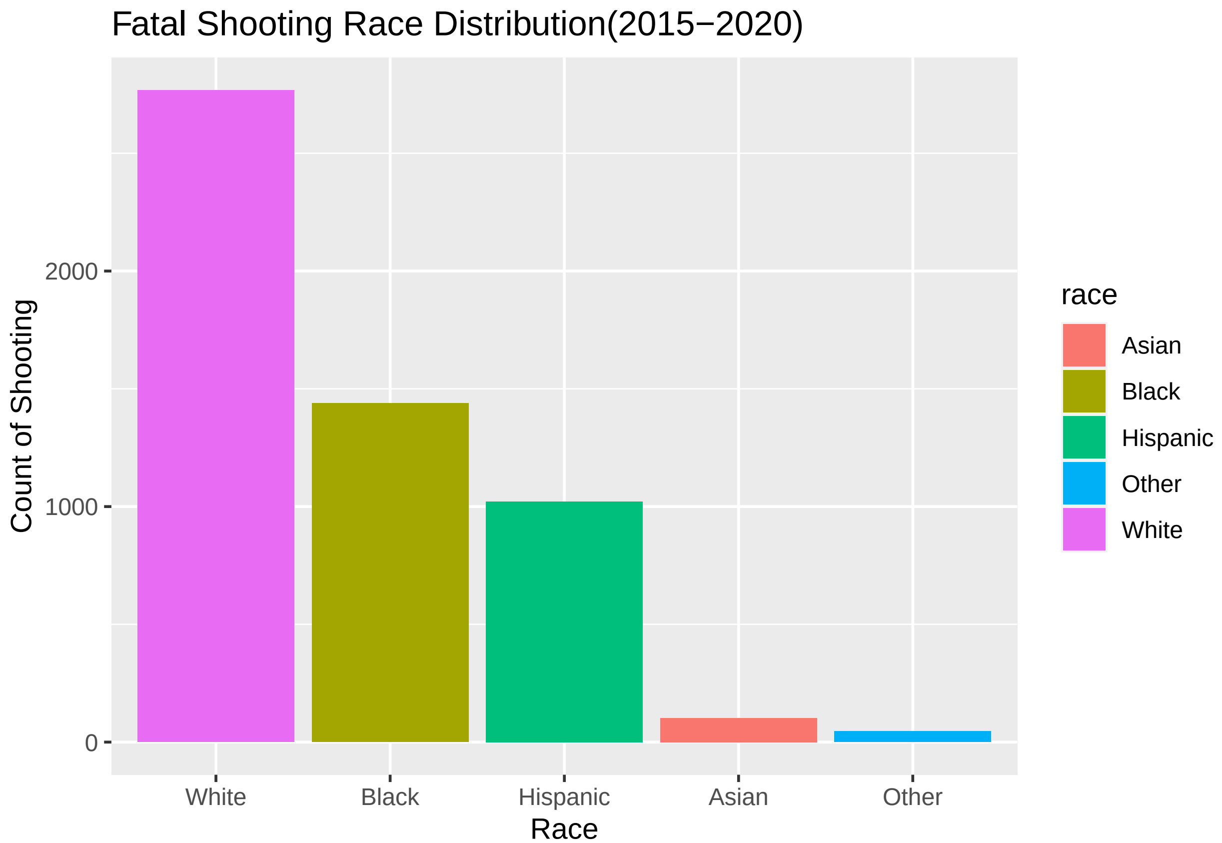
To learn the age patterns of victims, we divided the victims into age groups and made a histogram to show the distribution of victims among different age groups in fatal shooting incidents. As we can see, this histogram has an unsymmetrical unimodal distribution and has a peak at the age group of 30-34. From the plot, we can see that the majority of victims shot and killed by police were between the age of 20-49. We also noticed that this distribution is pretty close to the age distribution of crimes in the United States.

Gender Plot



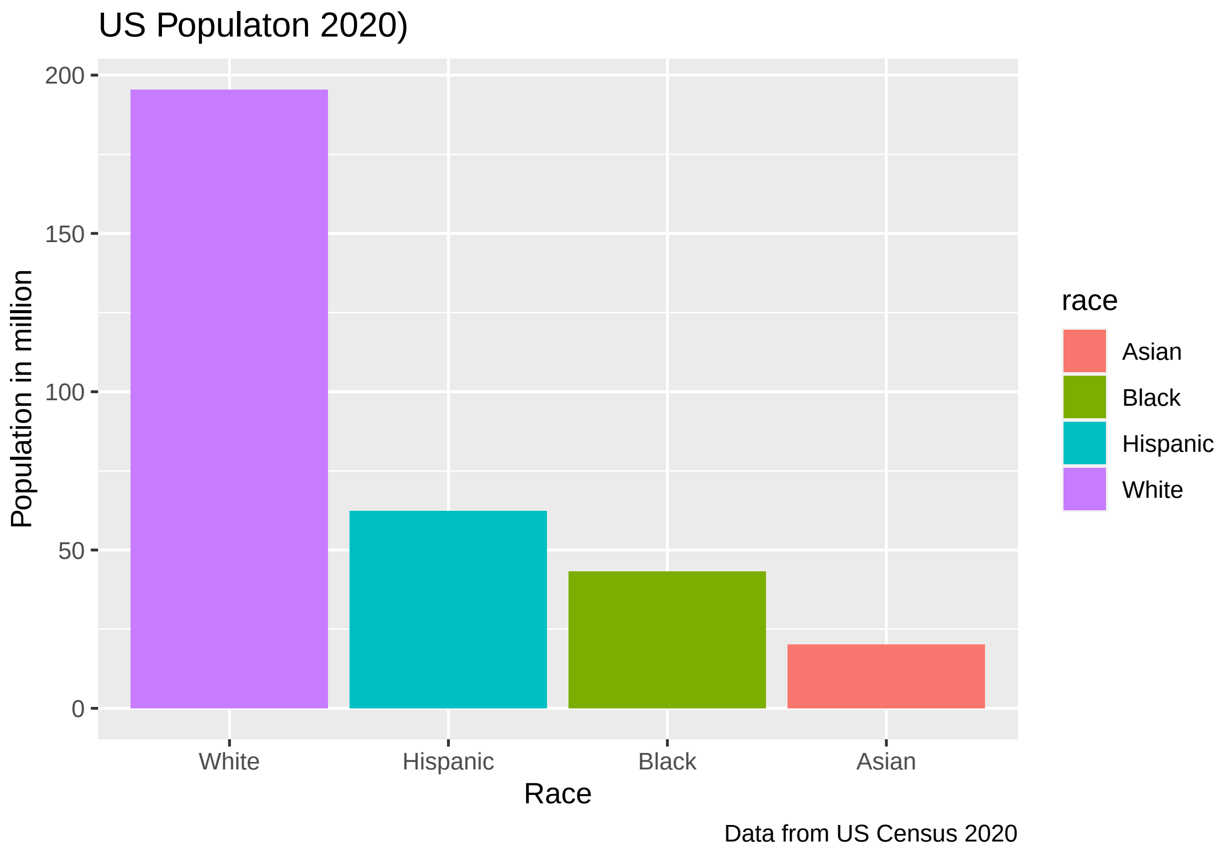
This plot visualizes deaths by police shootings according to gender. It is clear from the chart that males make up the vast majority of deaths by police shooting despite only making up half of the population.

Race Plot



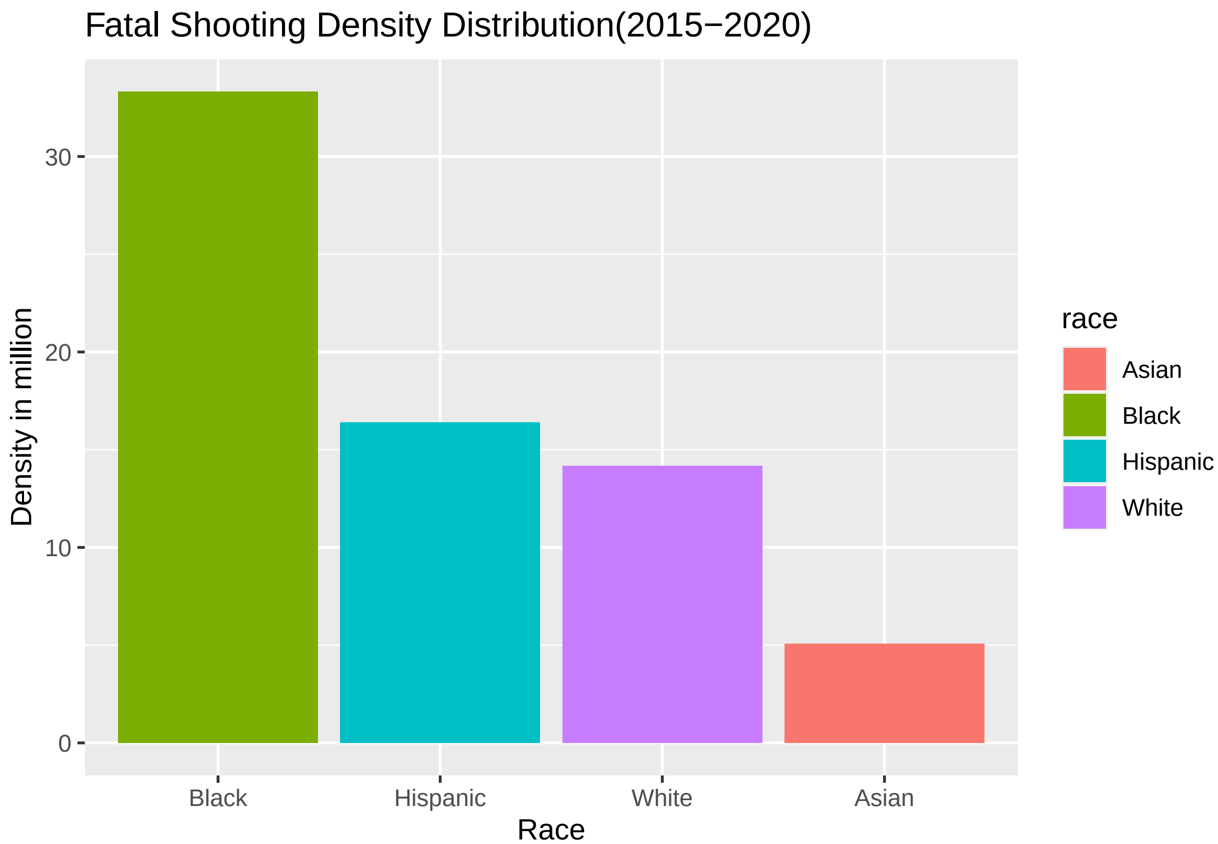
This plot aims to visualize the count of police fatal shootings for each race in the dataset. We also remove the samples where the race is unknown or blank, so we can have a better understanding of the known races. The result below shows that most of the counts of fatal shootings are against White, with nearly double the count of the second highest race, Black. However, we also noticed that there's more outcry about police violence among the black population so we wanted to investigate why.

Race Plot by Population



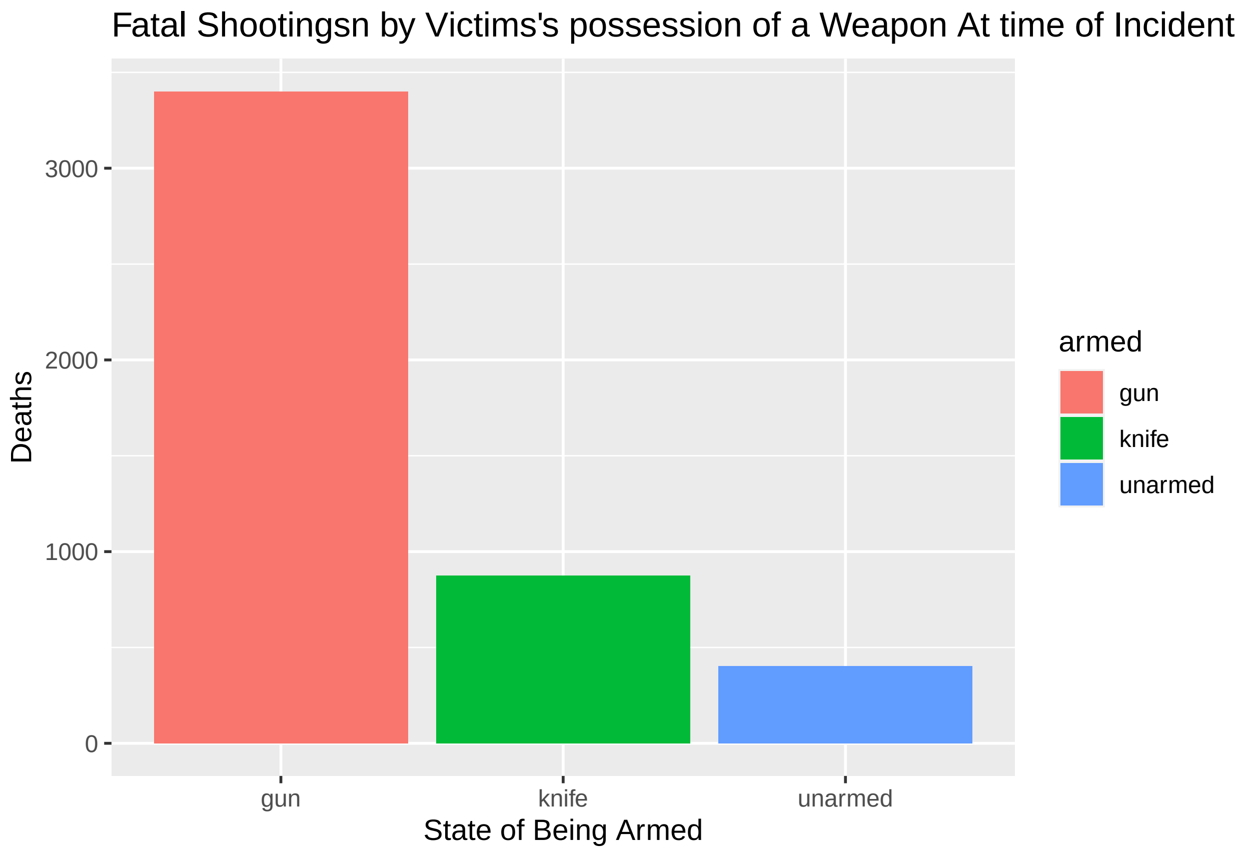
This plot shows the population for the four races that we have the most data on. Since the population in the US is different for each race, so the difference in shooting count could be explained. However, it is important to note that these differences in distribution are also not consistent between the races. For example, the black pop is nearly 1/4 the white population but the shooting count for the black population is roughly 1/2 of the white population.

Race Plot by Density



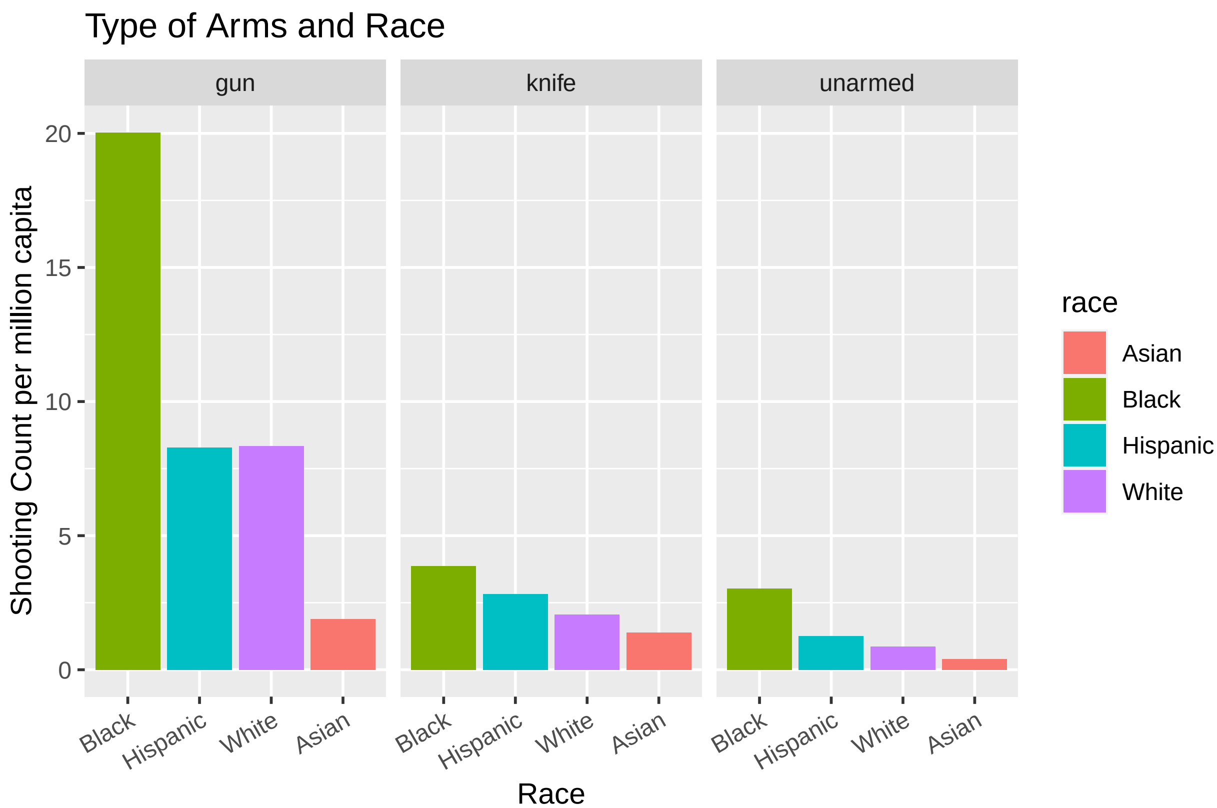
This plot aims to explore more into the race category because we feel that count alone is not representative because the population of each race is different. The plot below should count per million people in the respective race. We can see that although White population has the highest count, Black population actually has the highest per capita in million rate, represented in green where roughly 33 black people are shot among a million of them.

Posession of a Weapon Plot



This plot aims to visualize whether there are any identifiable trends between possession of a weapon and a fatal shooting occurring. The generated plot indicates that victims in possession of a gun were the most common among those in police shootings, following victims with knives, and lastly victims who were unarmed. This intuitively makes sense as firearms pose the most danger/risk with knives following.

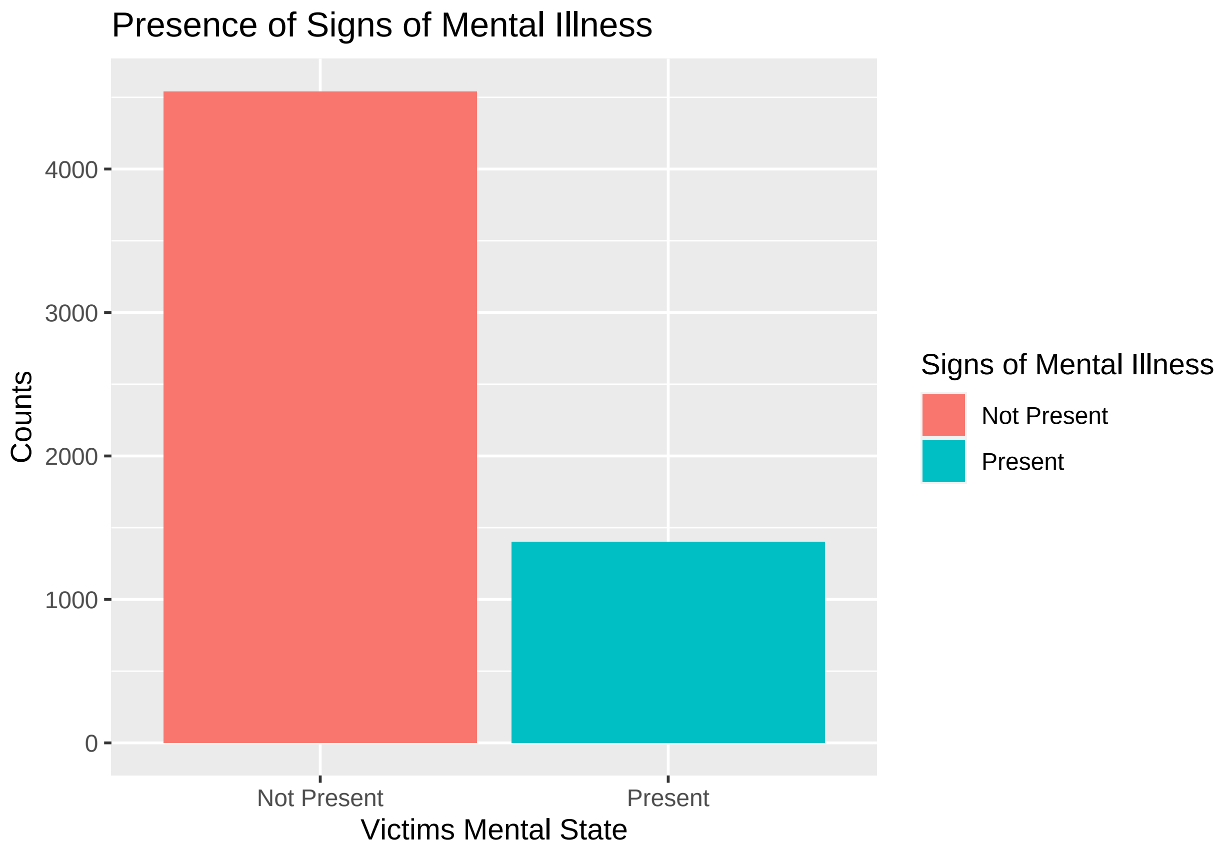
Race and State of Armed Plot



a from U.S. Census 2020. Other races not included for the simplicity of the fatal shooting data set

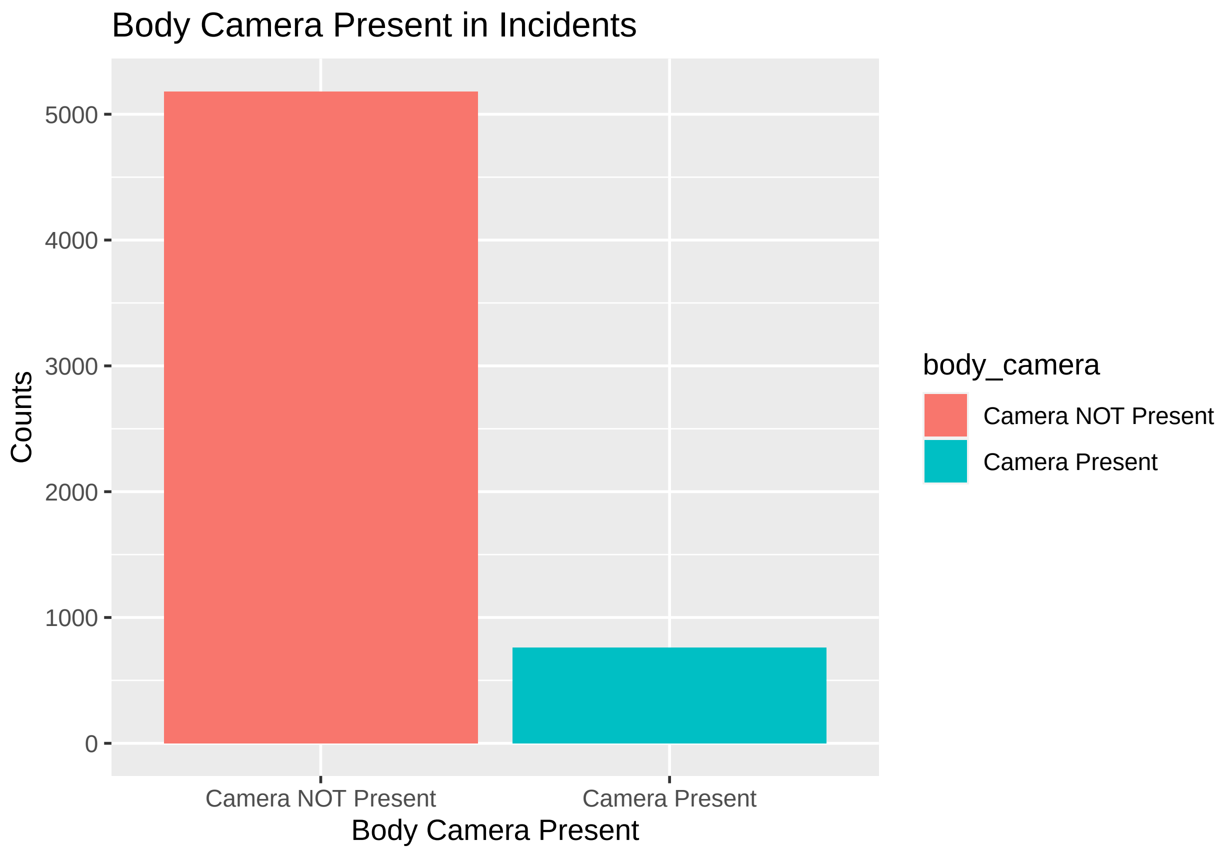
This plot separates the per capita in million race distribution by three types of arms: gun, knife, and unarmed. We wish to see if there is any difference in the fatal shooting distribution trend for the types of weapon.

Signs of Mental Illness Plot



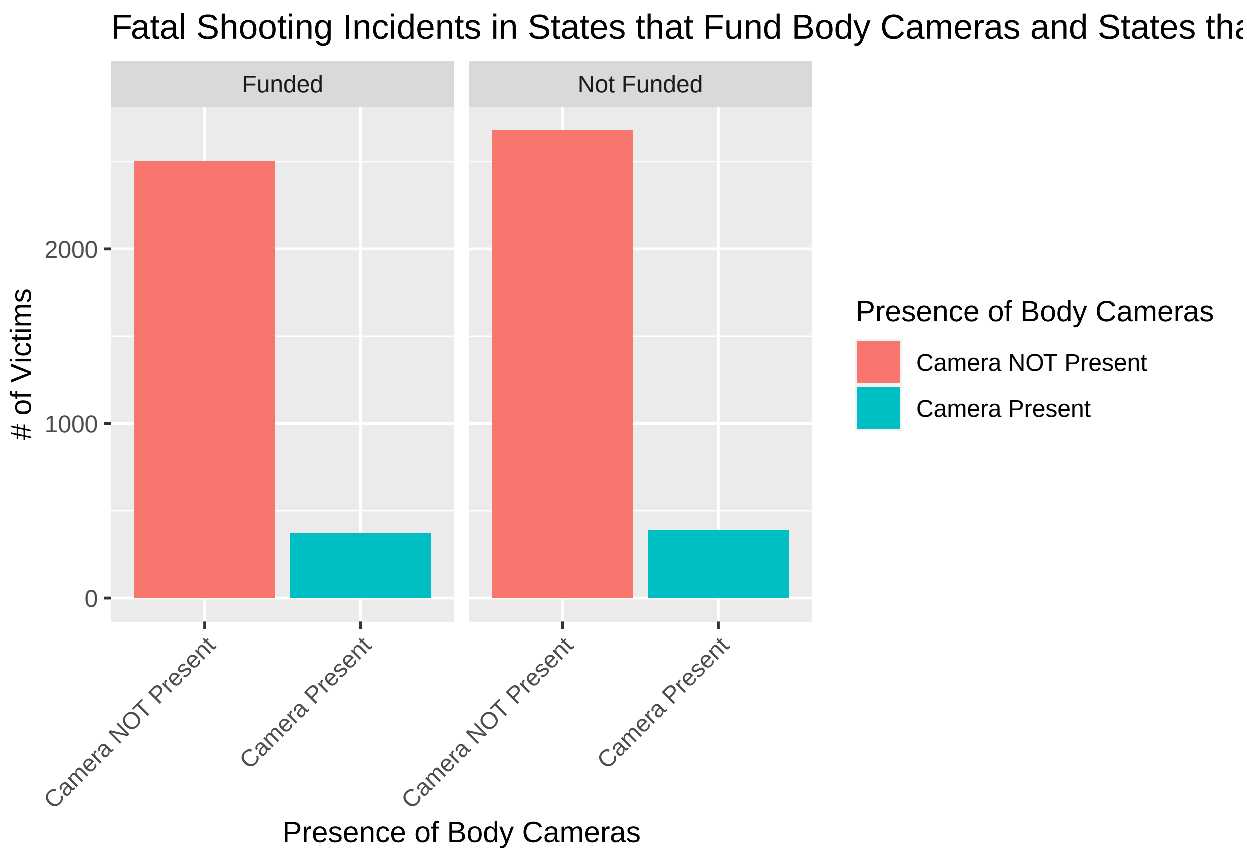
We made a bar graph to visualize the count of victims separated by their mental state. The given dataset defines active signs of mental illness as having a history of mental health issues or experiencing mental distress at the time of the shooting. From the plot below, we can see that more victims with no signs of mental illness were fatally shot. However, it is still noticeable that about a quarter of those killed exhibited signs of mental illness, which is since the reasons that caused those victims to show signs of threat could be lack of medical care.

Presence of a Police Body Camera Plot



This plot aims to answer whether or not the presence of body cameras correlates to fatal shooting incidence. As we can see from the plot, an overwhelming majority of incidents happened when the police did not wear body cameras. There can be many reasons that lead to these unevenly distributed counts. "Prior to May 2020, South Carolina was the only state to require widespread adoption of body-worn cameras" (Body-Worn Camera Laws Database). As of January 2018, there have been only thirteen states and the District of Columbia legislated funding for body cameras. We wonder if this correlation could be explained in that the police officers might make more careful decisions with the presence of the camera because they will take full responsibility for their actions that are being recorded.

Presence State Fund vs. Body Camera



The bar graphs visualize the presence of body cameras during the shooting faceted by state's `camera_funding`. From the plots, we can see that the count distribution for both types of states are approximately the same. This indicates that state funding barely has correlation with the presence of body cameras at shootings, and absence of body cameras at the shooting does not have much to do with lack of state funding.

Pairs Plot

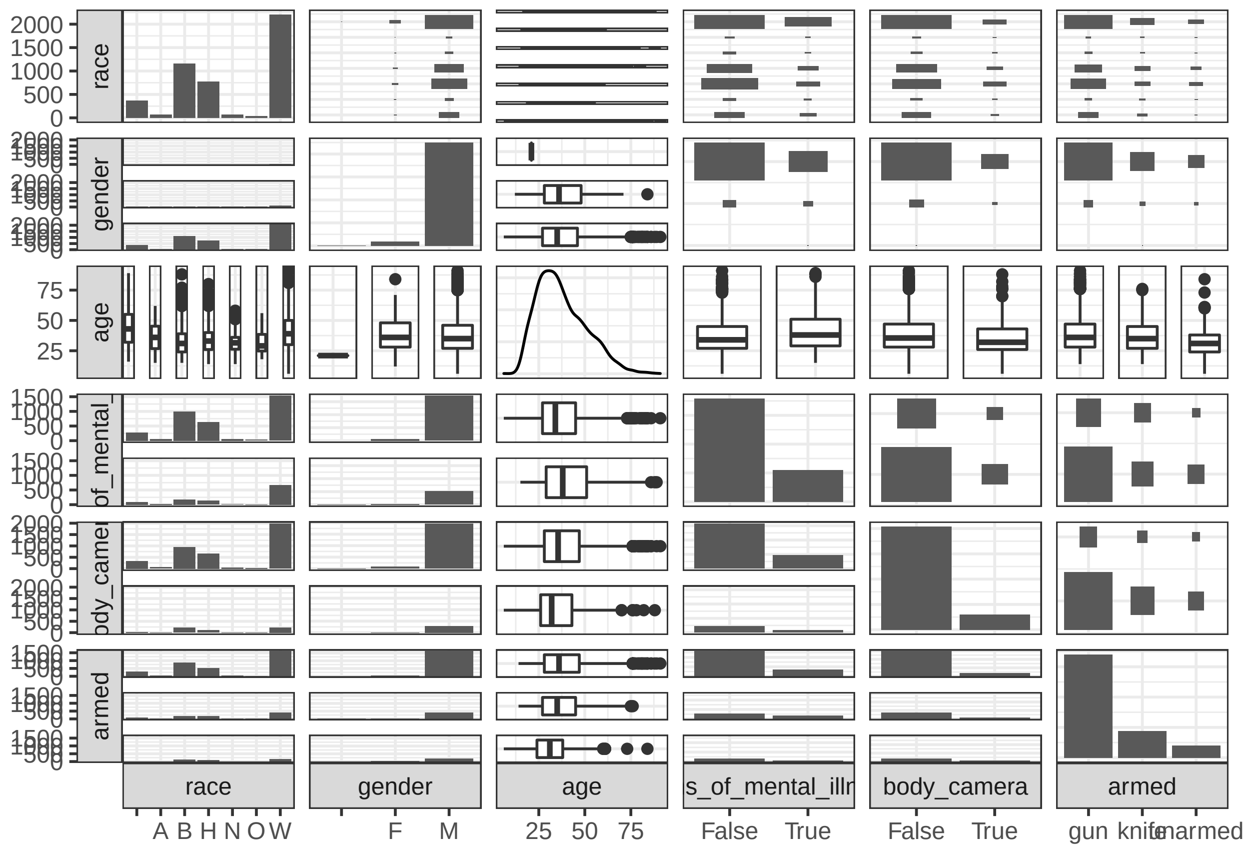
```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```
## Warning: Removed 182 rows containing non-finite values (stat_boxplot).
## Removed 182 rows containing non-finite values (stat_boxplot).
## Removed 182 rows containing non-finite values (stat_boxplot).
## Removed 182 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 182 rows containing non-finite values (stat_density).
```

```
## Warning: Removed 182 rows containing non-finite values (stat_boxplot).
## Removed 182 rows containing non-finite values (stat_boxplot).
## Removed 182 rows containing non-finite values (stat_boxplot).
## Removed 182 rows containing non-finite values (stat_boxplot).
```

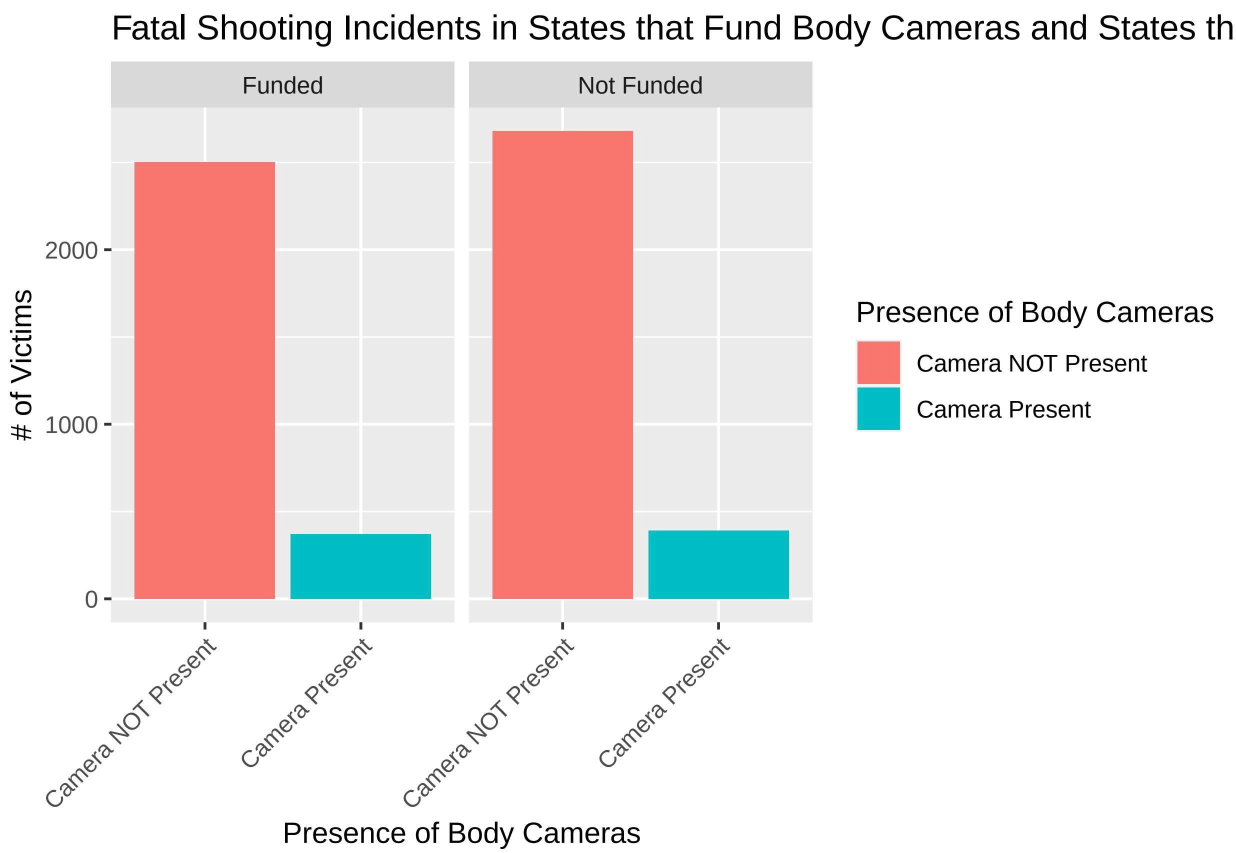
```
## Removed 182 rows containing non-finite values (stat_boxplot).
## Removed 182 rows containing non-finite values (stat_boxplot).
```



The x-axis of the plot shows the variables and the y-axis shows the count of fatal shooting incidence. All graphs aside from the diagonal ones show the relationship between two selected features. Categorical vs. categorical are shown in bar graphs, numerical vs. numerical are shown in line plots, and numerical vs. categorical are shown in steam-and-leaf plots. The diagonal plots compare how the variables within a feature relate to our target, an incidence of fatal shooting. Since none of the diagonal plots are near constant (a straight horizontal line, or two bars of around the same height), we can infer that all of the features are relevant in terms of fatal shooting because their respective variables produced an uneven split of counts. On the other hand, the non-diagonal plots (aka the ones that compare two different features) show the distribution of incidence when we combine the different variables in both features (e.g. a plot of gender distribution when they do not have signs of mental illness and a plot of gender distribution when they do have signs of mental illness). This allows us to consider what to combine when we consider deeper analysis that’s not single-feature, as we can choose the features with low correlation to see what variables in the combination have a higher incidence count.

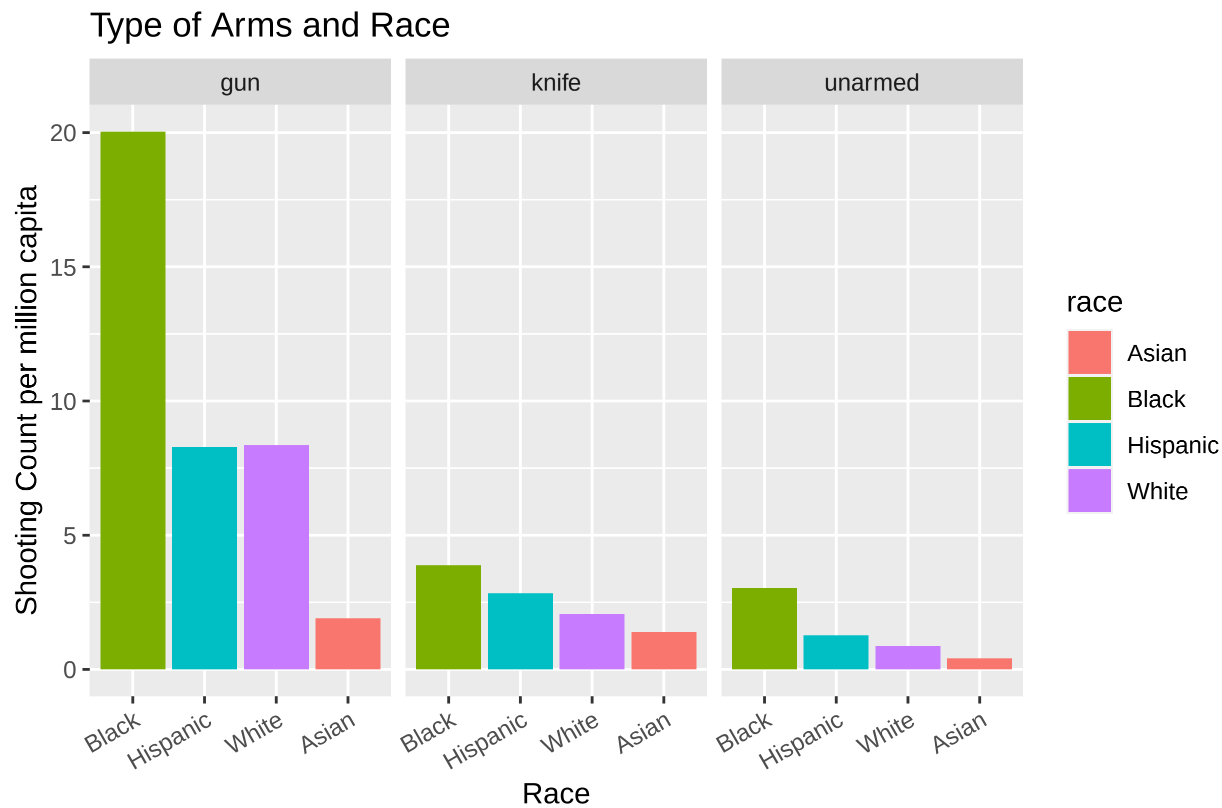
V. Finished Analysis

Body Camera by State Funding



This plot compares states with and without statewide funding for officer body cameras. While the bar graphs do not demonstrate any statistically significant difference between the two groups, a conclusion cannot be drawn regarding the influence of the presence of officer body cameras in a fatal police encounter. This is because state funding does not imply legal obligations for police officers to wear body cameras. Looking at body-camera mandates rather than funding would allow for a more comprehensive analysis. However, only recently have such mandates been passed, and only in a few selective states. In fact, prior to 2020 (where our data analysis ends), only one state required the widespread adoption of body-worn cameras.

Possesion of Weapons by Race per Capita



a from U.S. Census 2020. Other races not included for the simplicity of the fatal shooting data set

This plot reveals the racial demographics of fatal shooting casualties (adjusted for population density) associated with different states of being armed (possession of a gun, possession of a knife, and unarmed). This plot shows a disturbing trend in fatal shootings, Black individuals disproportionately make-up casualties in all status of weapons. In other words, compared to Whites, Hispanic and Asians, Black individuals are more likely subjected to fatal police violence, regardless of whether they are armed or not.

We want to investigate whether race is a determining factor and we would expect the per capita shooting count for each race to stay roughly the same if race does not play any part. This is kind of similar to the relationship between the Hispanic and White populations for all three weapon types. Therefore, the plots show that racial demographics contribute to fatal shooting but we cannot conclude how an individual's race is contributing. Whether the plot indicates possible police bias toward racial demographics cannot be concluded due to the lack of information on the demographics of violent crimes committed. For example, if one race commits violent crime at a higher rate than another it would make sense that that race is disproportionately represented.

However, studies have shown that Black individuals are more likely to be stopped, searched, and arrested by law enforcement, even when controlling for other factors such as the presence of drugs or weapons. This bias extends to the use of lethal force, with Black individuals being 2.5 times more likely to be killed by police than white individuals.

Furthermore, the over-policing of Black and Hispanic communities, combined with a lack of access to education, healthcare, and other essential services, creates a cycle of poverty and violence that disproportionately affects these communities. This, in turn, leads to a higher likelihood of encounters with law enforcement, and ultimately, a higher likelihood of being a victim of fatal police violence.

VI. Conclusion

Through this data analysis we learned that the White population has the highest number of fatal shootings by police in the US, contrary to our expectations that the Black population would be the most affected due to the extensive news coverage of black people being subjected to police violence. However, when adjusting for population distribution we could in fact see that Black people were more likely to be the victims of police fatal shootings per capita.

Further, we became aware that the majority of US states do not require body cameras for on-duty officers, and only less than half of the states provide funding for them while the vast majority of fatal shootings occurred when officers were not wearing body cameras.

Further analysis could be conducted on a dataset that includes both fatal shootings and non-fatal shootings. This analysis would shed light on possible bias in police officers if one demographic was more likely to be given a fatal wound during a police shooting encounter. Additionally, if there was available data that reported on the total violent crime within the US it would be possible to draw conclusions about possible police bias, as the prevalence of violent crime in different communities can be accounted for and controlled for. For example, if the data showed that a certain demographic was overrepresented in fatal police shootings despite being underrepresented in violent crimes it would then be possible to conclude that police are biased toward that community.

Lastly, with data about countries other than the United States, we can see how the U.S. compares to other nations in terms of police violence. This analysis could provide valuable insight into the unique factors contributing to police violence in the US. For example, such a comparison could reveal whether certain policies or practices that are common in the US, but not in other countries, are associated with higher rates of police shootings. Moreover, comparing the data across countries could also shed light on potential solutions to the problem of police violence, by identifying successful strategies that have been implemented in other countries. For example, the data could be used to evaluate the effectiveness of different strategies for reducing police violence, such as de-escalation training for officers or increased accountability measures. By analyzing the data in this way approaches can be identified for reducing police violence and improving relations between law enforcement and the communities they serve. Overall, this comparison would be a valuable addition to our understanding of police violence in the US.

VII. Resources

"Body-Worn Camera Laws Database." National Conference of State Legislatures, 30 Apr. 2021, www.ncsl.org/research/civil-and-criminal-justice/body-worn-cameras-interactive-graphic.aspx.

"U.S. Census Bureau QuickFacts: United States." Census Bureau QuickFacts, www.census.gov/quickfacts/fact/table/US/PST

"Washingtonpost/data-police-shootings: The Washington Post Is Compiling a Database of Every Fatal Shooting in the United States by a Police Officer in the Line of Duty since 2015." GitHub. The Washington Post, 3 Sept. 2020.

VIII. Appendix

```
library(tidyr)
library(tidyverse)
library(dplyr)
library(ggplot2)
data <- read.csv("fatal-police-shootings-data.csv")
data2 <- data %>%
```

```
  filter(date <= "2020-12-31")

glimpse(data2)
plotdata_gender <- data2 %>%
  filter(gender != "") %>%
  group_by(gender) %>% # this groups the data by gender
  select(gender) %>%
  summarize(count =n()) # counts how many occurrences of each gender is recorded
plotdata_age <- data2 %>%
# create a variable called age_group
  mutate(age_group = case_when(
    age %in% seq(10,14) ~ "10-14",
    age %in% seq(15,19) ~ "15-19",
    age %in% seq(20,24) ~ "20-24",
    age %in% seq(25,29) ~ "25-29",
    age %in% seq(30,34) ~ "30-34",
    age %in% seq(35,39) ~ "35-39",
    age %in% seq(40,44) ~ "40-44",
    age %in% seq(45,49) ~ "45-49",
    age %in% seq(50,54) ~ "50-54",
    age %in% seq(55,59) ~ "55-59",
    age %in% seq(60,64) ~ "60-64",
    age %in% seq(65,69) ~ "65-69",
    age %in% seq(70,74) ~ "70-74",
    age %in% seq(75,79) ~ "75-79",
    age %in% seq(80,84) ~ "80-84",
    age %in% seq(85,89) ~ "85-89",
    age %in% seq(90,94) ~ "90-94"
  )) %>%
  # divide age group based on age

  filter(!is.na(age_group))
# filter out observations that have NA vales for age_group

plot_race <- data2 %>%
  filter(race != "" & race !="N") %>% # remove the samples with blank or unknown race
  group_by(race) %>%
  summarize(count = n()) %>% # count the occurrence of each race
  mutate(race = (case_when(
    race == "A" ~ "Asian",
    race == "B" ~ "Black",
    race == "W" ~ "White",
    race == "H" ~ "Hispanic",
    race == "O" ~ "Other"
  ))) %>% # rename the race so it's easier to read
  arrange(desc(count))
race_den <- read.csv("us-race-distribution.csv")
plot_race_den <- race_den %>% # plot data for per capita race
  left_join(plot_race, by = "race") %>%
  # combine the percentage with the dataset grouped by race
  mutate(population_mill = 329.5 * percent) %>%
  # multiply the percentage of each race by the total US population in 2020 to obtain the population in
  mutate(density = count/population_mill)
```

```
plotdata_armed <- data2 %>%
  group_by(armed) %>%
  filter(armed == "gun" | armed == "knife" | armed == "unarmed")%>%
  #filtering by the variables we are interested in  looking at
  summarize(count=n()) # counting the number of occurrences

plot_armed_race <- data2 %>%
  filter(race != "" & race !="N" & race != "O") %>% # filter out the N/A, unknown, and Other races
  filter(armed == "gun" | armed == "knife" | armed == "unarmed") %>%
  # filter status to only armed with gun or knife or unarmed
  group_by(race, armed) %>%
  summarize(count = n()) %>%
  mutate(race = (case_when( # rename for better presentation
    race == "A" ~ "Asian",
    race == "B" ~ "Black",
    race == "W" ~ "White",
    race == "H" ~ "Hispanic",
  ))) %>%
  right_join(race_den, by = "race") %>%
  mutate(population_mill = 329.5 * percent) %>% # calculate total population for each race
  mutate(per_mill_capita = count/population_mill) # calculate respective per capita
plotdata_mental <- data2 %>% # data for signs of mental illness
  mutate (mental_illness = if_else(signs_of_mental_illness=="True",
                                   "Present",
                                   "Not Present"))
# create a variable called mental_illness
# if value for signs_of_mental_illness = "True", mental_illness = "Present"
# if value for signs_of_mental_illness != "True", mental_illness = "Not Present"

plotdata_camera <- data2 %>% #data for body_camera
  mutate (body_camera = if_else(body_camera=="True","Camera Present", "Camera NOT Present"))
 # if the value is "True", make it "Camera Present"
 # if the value is not "True", make it "Camera NOT Present"

camera_funding <- plotdata_camera %>%
  # a new variable to indicate if the state has funding for body camera
  mutate (camera_funding = (case_when(
    state == "CA" ~ TRUE,
    state == "NV" ~ TRUE,
    state == "CO" ~ TRUE,
    state == "TX" ~ TRUE,
    state == "IL" ~ TRUE,
    state == "KY" ~ TRUE,
    state == "FL" ~ TRUE,
    state == "SC" ~ TRUE,
    state == "NC" ~ TRUE,
    state == "PA" ~ TRUE,
    state == "NJ" ~ TRUE,
    state == "DC" ~ TRUE,
    state == "CT" ~ TRUE,
    state == "MA" ~ TRUE,
    state == "TX" ~ TRUE,
    state == "AK" ~ FALSE,
```

```
    state == "AK" ~ FALSE,
    state == "WA" ~ FALSE,
    state == "MT" ~ FALSE,
    state == "ND" ~ FALSE,
    state == "MN" ~ FALSE,
    state == "WI" ~ FALSE,
    state == "MI" ~ FALSE,
    state == "NY" ~ FALSE,
    state == "VT" ~ FALSE,
    state == "NH" ~ FALSE,
    state == "ME" ~ FALSE,
    state == "RI" ~ FALSE,
    state == "ID" ~ FALSE,
    state == "WY" ~ FALSE,
    state == "SD" ~ FALSE,
    state == "IA" ~ FALSE,
    state == "IN" ~ FALSE,
    state == "OH" ~ FALSE,
    state == "OR" ~ FALSE,
    state == "NE" ~ FALSE,
    state == "MO" ~ FALSE,
    state == "WV" ~ FALSE,
    state == "MD" ~ FALSE,
    state == "DE" ~ FALSE,
    state == "AZ" ~ FALSE,
    state == "UT" ~ FALSE,
    state == "KS" ~ FALSE,
    state == "AR" ~ FALSE,
    state == "TN" ~ FALSE,
    state == "VA" ~ FALSE,
    state == "NM" ~ FALSE,
    state == "OK" ~ FALSE,
    state == "LA" ~ FALSE,
    state == "MS" ~ FALSE,
    state == "AL" ~ FALSE,
    state == "GA" ~ FALSE,
    state == "HI" ~ FALSE,
    state == "AS" ~ FALSE,
    state == "GU" ~ FALSE,
    state == "MP" ~ FALSE,
    state == "PR" ~ FALSE,
    state == "VI" ~ FALSE,
  ))) %>%
  mutate (camera_funding = if_else(camera_funding==TRUE,"Funded", "Not Funded"))

ggplot(plotdata_age, aes(age_group)) +
  geom_bar() +
  theme(axis.text.x = element_text(angle=40, hjust=1)) +
  labs(
    title = "Victims By Age",
    x = "Age",
    y = "# of Victims"
  )
```

```
ggplot(plotdata_gender,
       aes(x = gender, y = count, fill= gender)) +
  geom_col() +
labs( title = "Fatal Shooting By Police from 2015-2020 by Gender",
  x = "Gender",
  y  = "Number of fatal shootings"
)
ggplot(plot_race, aes(reorder(race, -count), count, fill=race)) +
    geom_col() +
    labs(
      title = "Fatal Shooting Race Distribution(2015-2020)",
      x = "Race",
      y = "Count of Shooting"
    )

ggplot(plot_race_den, aes(reorder(race, -population_mill), population_mill, fill=race)) +
  geom_col() +
    labs(
      title = "US Populaton 2020)",
      x = "Race",
      y = "Population in million",
      caption = "Data from US Census 2020"
    )
ggplot(plot_race_den, aes(reorder(race, -density), density, fill=race)) +
  geom_col() +
    labs(
      title = "Fatal Shooting Density Distribution(2015-2020)",
      x = "Race",
      y = "Density in million"
    )
ggplot(plotdata_armed,
       aes(x= armed, y  = count, fill = armed)) +
  geom_col() +
labs( title = "Fatal Shootingsn by Victims's possession of a Weapon At time of Incident",
      x = "State of Being Armed",
      y = "Deaths"
)
ggplot(plot_armed_race, aes(reorder(race, -per_mill_capita), per_mill_capita, fill=race)) +
  geom_col() +
  facet_wrap(~armed) +
  theme(axis.text.x = element_text(angle=30, hjust=1)) +
    labs(
      title = "Type of Arms and Race",
      x = "Race",
      y = "Shooting Count per million capita",
      caption = "Population Data from U.S. Census 2020. Other races not included for the simplicity of 
    )
ggplot(plotdata_mental, aes(x=mental_illness, fill = mental_illness)) +
  geom_bar() +
labs(title = "Presence of Signs of Mental Illness",
     x = "Victims Mental State",
     y = "Counts",
     fill = "Signs of Mental Illness")
```

```
ggplot(plotdata_camera, aes(x=body_camera, fill=body_camera)) +
  geom_bar() +

labs(title = "Body Camera Present in Incidents",
     x = "Body Camera Present",
     y = "Counts")
ggplot(camera_funding, aes(body_camera, fill = body_camera )) +
  geom_bar() +
  facet_grid(~camera_funding) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  labs(
    title = "Fatal Shooting Incidents in States that Fund Body Cameras and States that Do Not Fund Body
    x = "Presence of Body Cameras",
    y = "# of Victims",
    fill = "Presence of Body Cameras")

pairsdata <- data2 %>%
  filter(armed == "gun" | armed == "knife" | armed == "unarmed")%>%
  select(race, gender, age, signs_of_mental_illness, body_camera, armed)

library(GGally)

ggpairs(pairsdata,
        lower = list(continuous = "cor",combo="box"),
        upper = list(continuous=wrap("smooth", alpha = 0.1),combo="box"),
        switch = "both") +
  theme_bw()
ggplot(camera_funding, aes(body_camera, fill = body_camera )) +
  geom_bar() +
  facet_grid(~camera_funding) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  labs(
    title = "Fatal Shooting Incidents in States that Fund Body Cameras and States that Do Not Fund Body
    x = "Presence of Body Cameras",
    y = "# of Victims",
    fill = "Presence of Body Cameras")
ggplot(plot_armed_race, aes(reorder(race, -per_mill_capita), per_mill_capita, fill=race)) +
  geom_col() +
  facet_wrap(~armed) +
  theme(axis.text.x = element_text(angle=30, hjust=1)) +
    labs(
      title = "Type of Arms and Race",
      x = "Race",
      y = "Shooting Count per million capita",
      caption = "Population Data from U.S. Census 2020. Other races not included for the simplicity of
    )
```