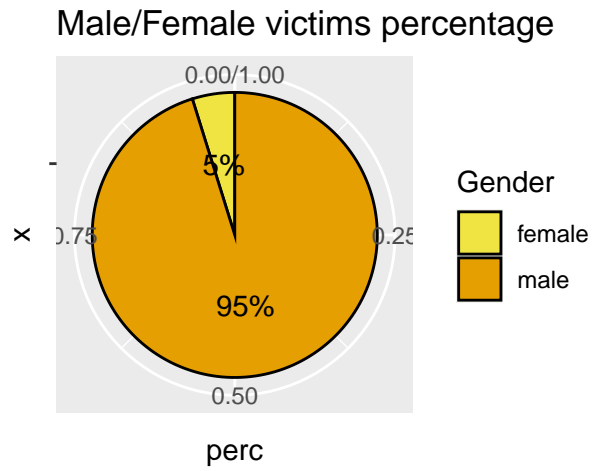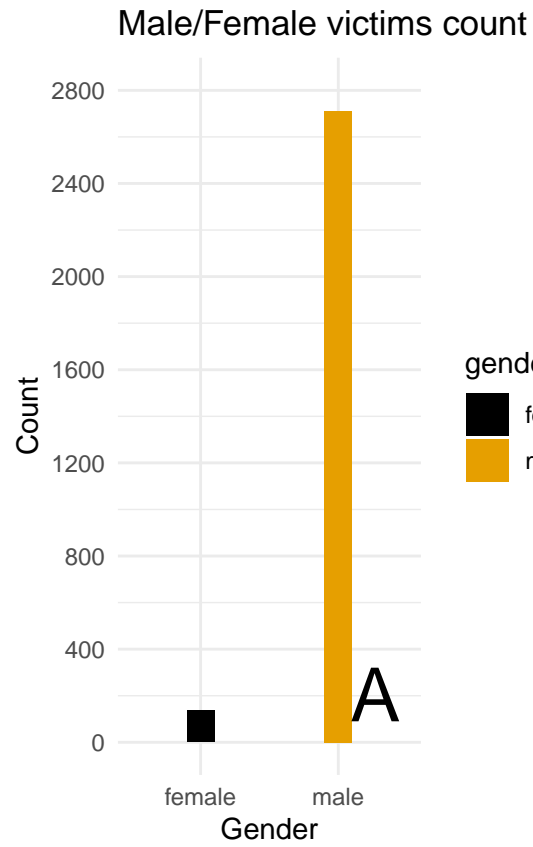# Term Project Part 1 - Police shooting data

```
## # A tibble: 2 x 2
##   gender 'Male/Female victims Count'
##   <chr>                       <int>
## 1 male                         2711
## 2 female                        137
```

```
## # A tibble: 15 x 2
##    armed_with                  n
##    <chr>                   <int>
##  1 gun                      1680
##  2 knife                     520
##  3 unarmed                   248
##  4 replica                   131
##  5 blunt_object               89
##  6 undetermined               62
##  7 vehicle                    37
##  8 other                      31
##  9 unknown                    24
## 10 gun;knife                  14
## 11 vehicle;gun                 4
## 12 other;gun                   3
## 13 blunt_object;blunt_object   2
## 14 knife;blunt_object          2
## 15 gun;vehicle                 1
```
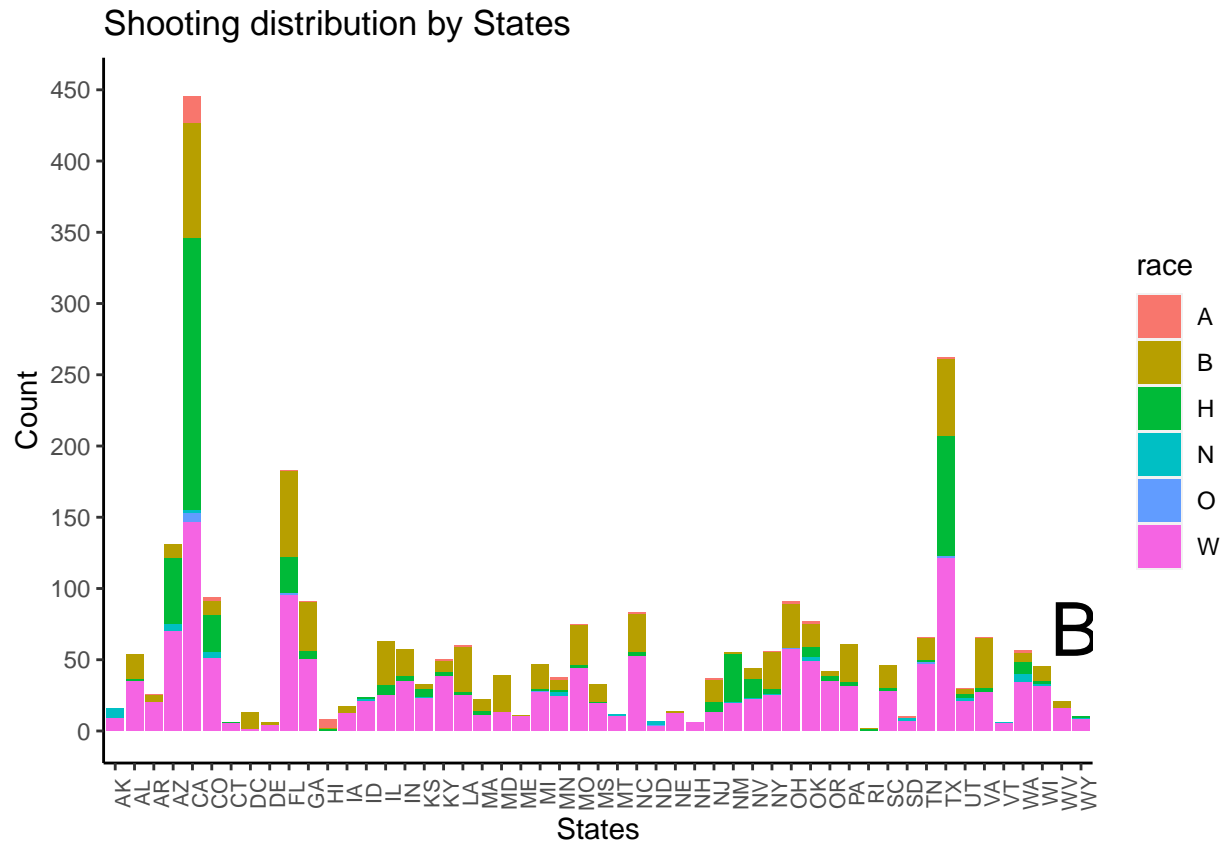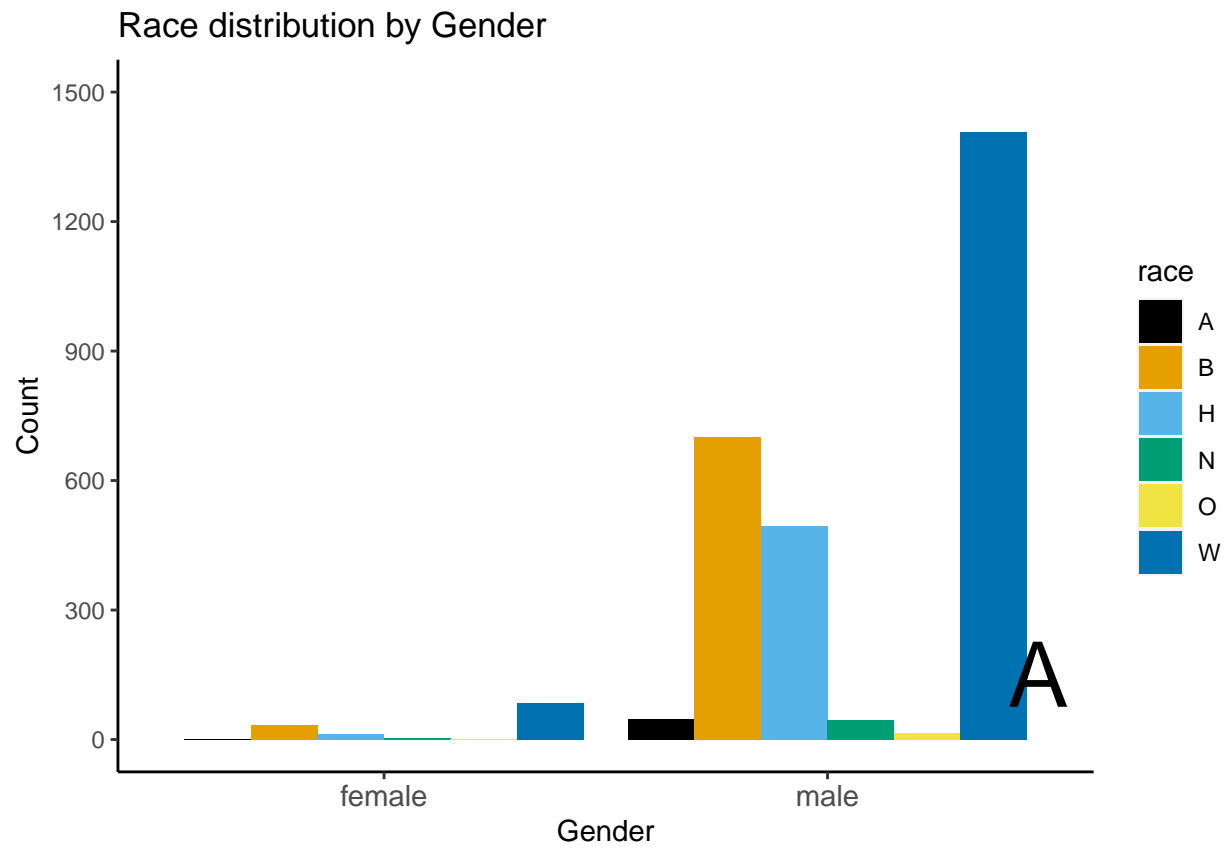
Male/Female victims count

Count

2800

2400

2000

1600

1200

800

400

0

female   male

Gender

gender

■ female

■ male

A

Male/Female victims percentage

0.00/1.00

5%

x   0.75   0.25

95%

0.50

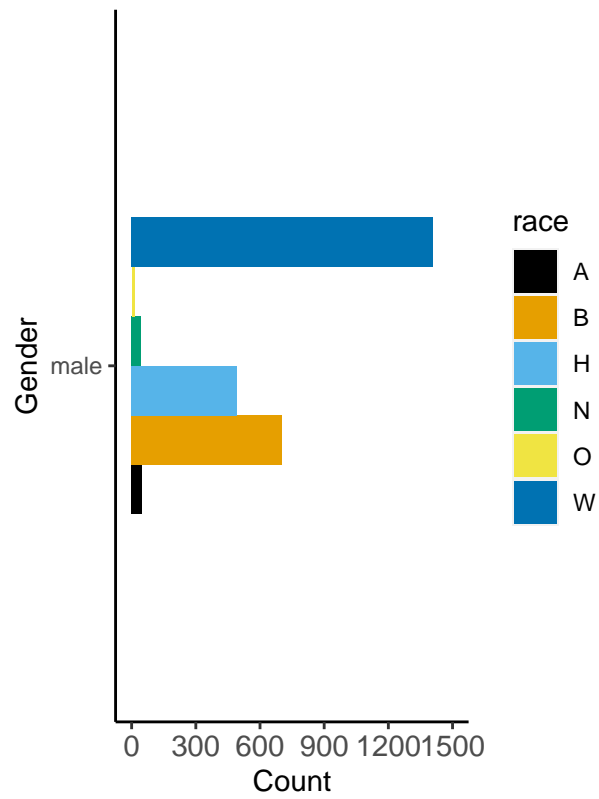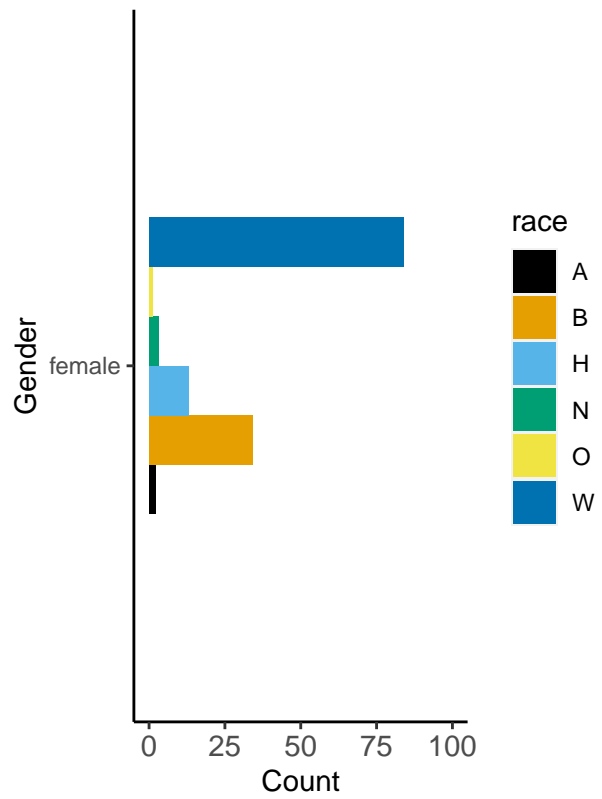perc

Gender

□ female

■ male

Figure 1. Panel A shows Male/Female victims count in the datatset and Panel B shows shooting distribution of race by states.

Race distribution by Gender

Race distribution by Male Victims

Race distribution by Female Vict

Shooting distribution by Age & Gender

Shooting Distribution by Race and Gender
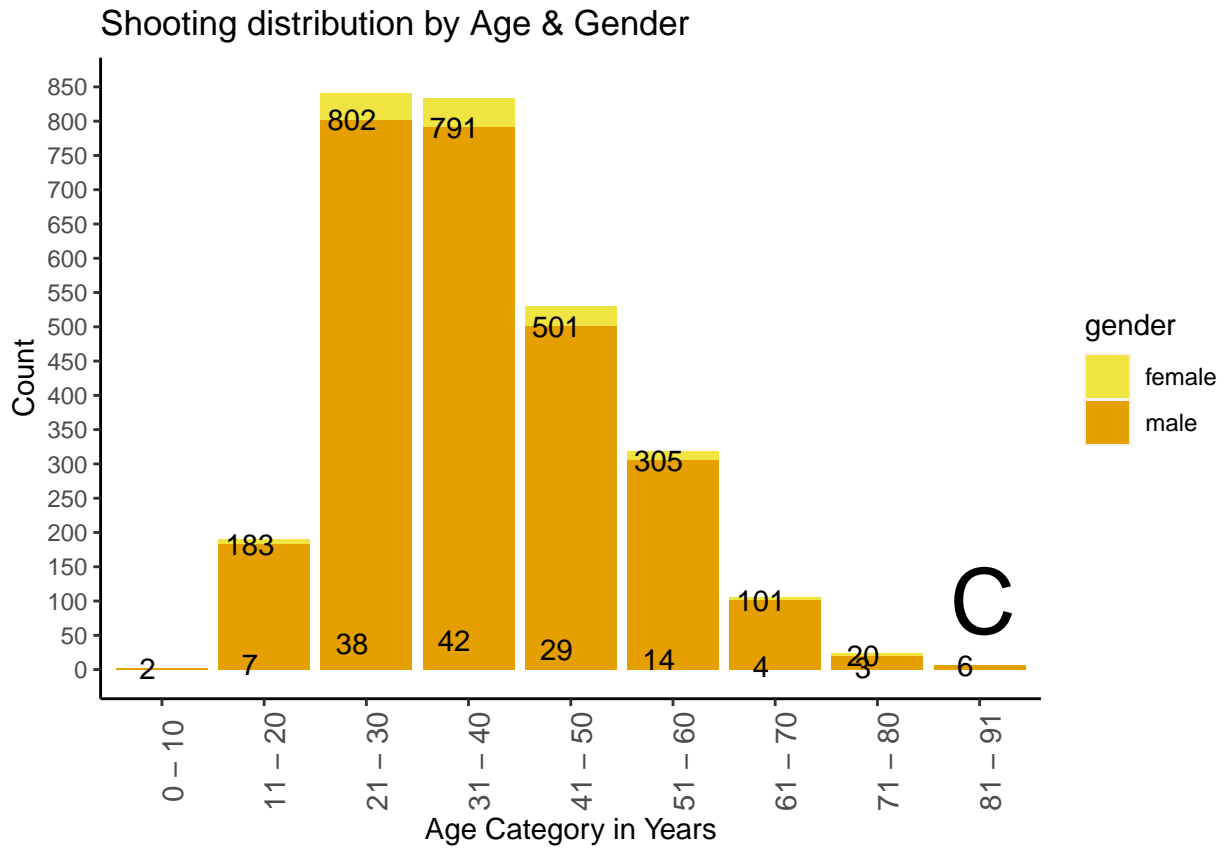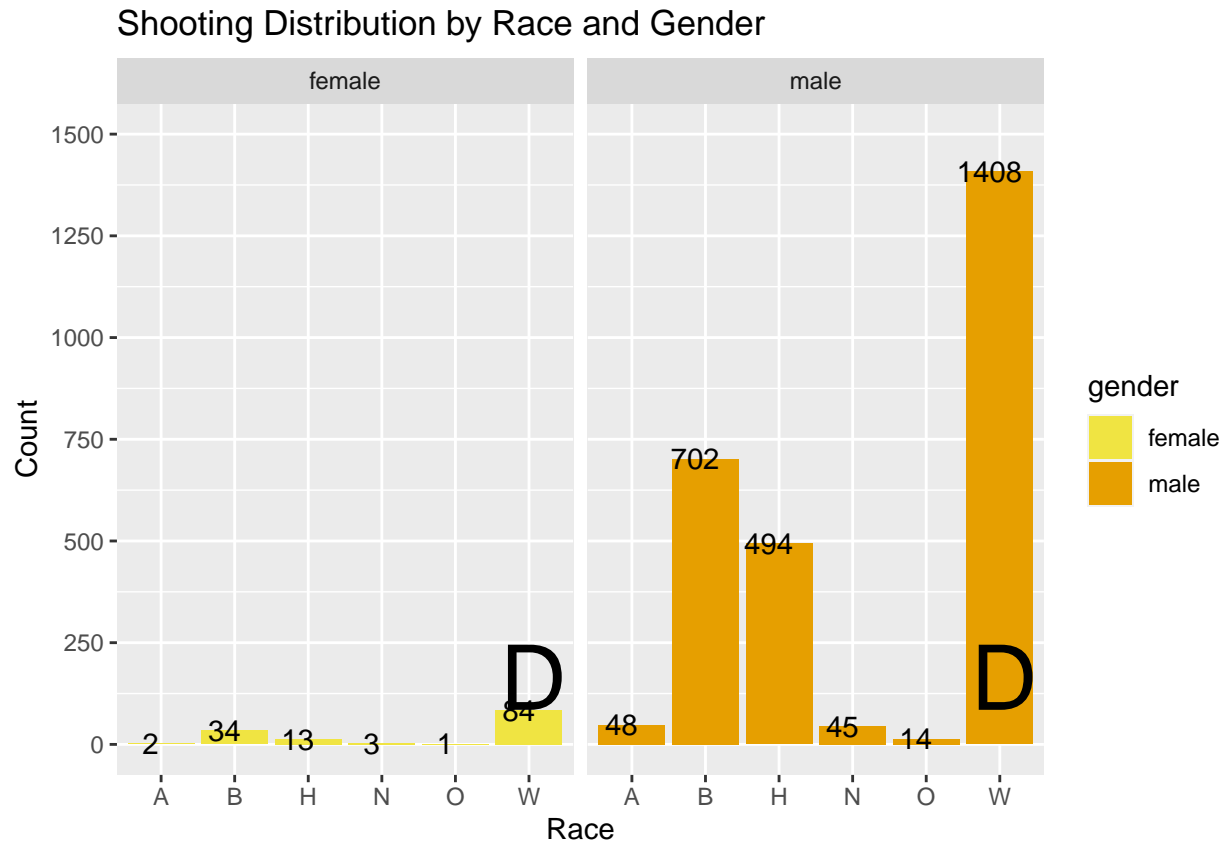
Figure 2. Panel A shows race distribution by gender. panel B shows Male/Female victims of race distribuition are shown separately.

panel C shows Shooting distribuition by age & gender and Panel D shows Shooting distribuition by Race and Gender.

Female victims characteristics by race, age and armed type
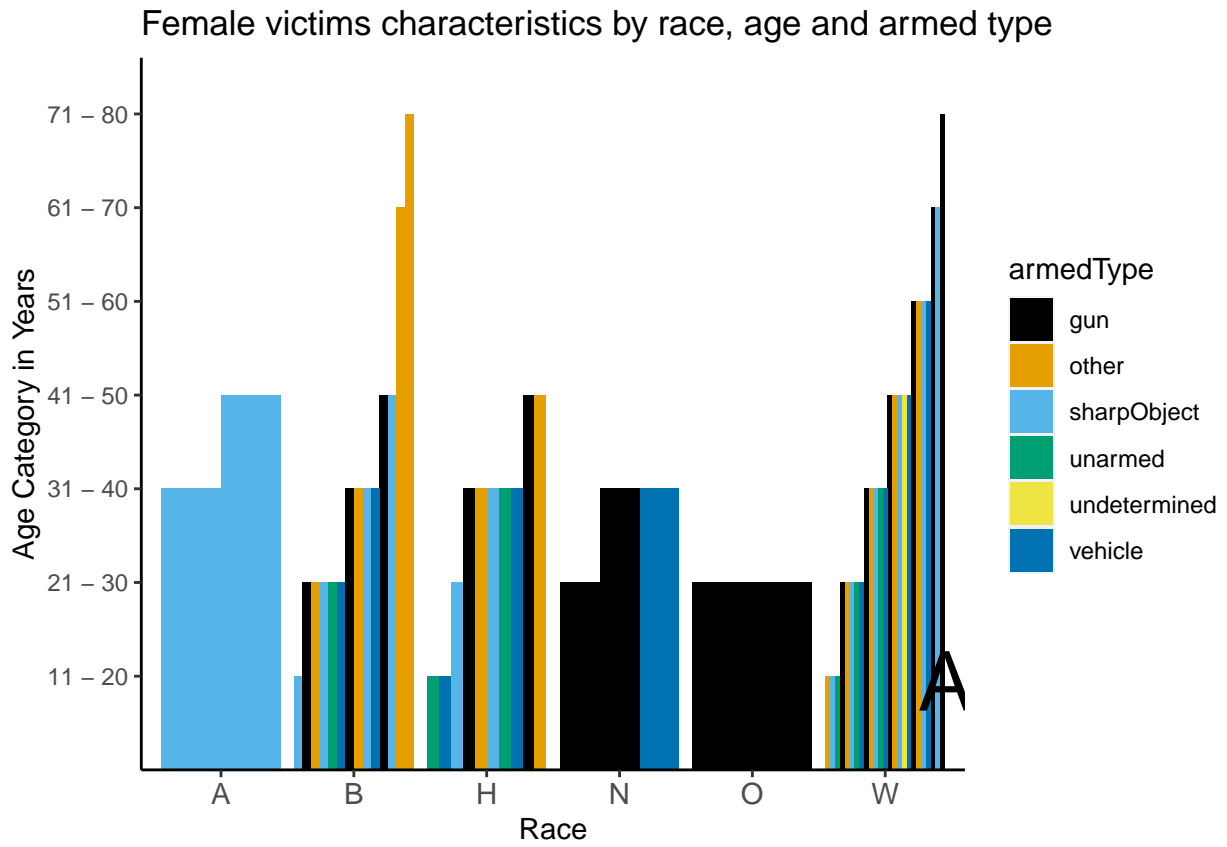
Figure 3. Panel A shows Male victims characteristics based on age, race, weapons.

Panel B shows Female victims characteristics based on age, race, weapons.

Summary:

Police shooting data set csv file was loaded to the data frame and analyzed. Total number of male and female victims count from the given data set are shown. Overall there more number male victims than female victims. As we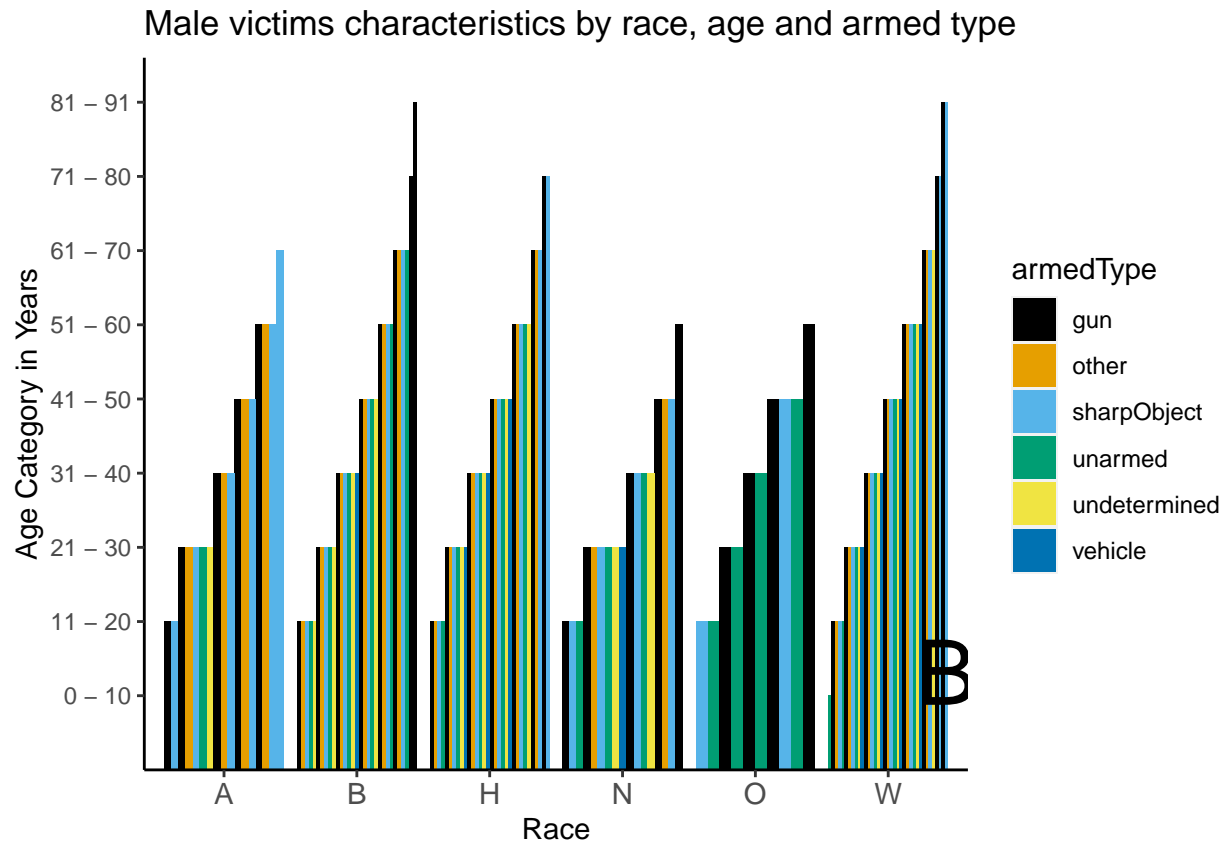 noticed age distribution, Younger people between 20-40 years of age died more than children and aged people. Male and female victims who are between 20-40 years of age are died more than other generations. When looked at the race distribution of male/female victims, White race tops both the male/female lists. Next on the list is black race and the least one is Others. When different characteristics like age, race, gender and weapons are analyzed, both male and female victims age ranges from 6 years and up to 91 years. Male victims use and carry variety of armed types whereas in female victims not all the race uses variety of weapons. Most races of female victims limit their weapons to gun, other and sharp objects. Only black & white races of females have older people, whereas in male victims older peoples are in almost all the races

Code Used To Make These Plots

Below are the code chunks used to make these plots.

```r
# Use the police shooting data and what you've learned so far in this course
# to answer the following questions:
# 1. How many male and female victims were presented in the data set?
# 2. What are the characteristics of male vs. female victims in terms
# of age, race, weapons, etc?
# Using race as an example, how many male victims were White, Black,
# Hispanic, Native American, Asian, and Other,
# in comparison, how many female victims were White, Black,
# Hispanic, Native American, Asian, and Other?
# 3. How would you summarize your findings? Post your code and
# write a paragraph. These, and all future exercises,
# will also be part of your project report.


rm(list = ls())

library(ggplot2)
library(reshape2)
library(lubridate)
library(dplyr)
library(gridExtra)
library(tidyverse)
library(hrbrthemes)
library(ggthemes)


# Load the data
shooting_orig <- read_csv("fatal-police-shootings-data1.csv",
                          col_names = TRUE, na="",  show_col_types = FALSE)
shooting_orig <- na.omit(shooting_orig)

cbPalette <- c("#F0E442", "#E69F00", "#56B4E9", "#009E73","#999999" ,
               "#0072B2", "#D55E00", "#CC79A7","#BAAAA3")
custom_color = c("darkgreen", "red","blue")

# boxplot(shooting_orig$age)
# #
# #How many male and female victims were presented in the data set?
# sort(table(shooting_orig$gender), decreasing =FALSE)
# count(shooting_orig, gender, sort = TRUE)

#How many male and female victims were presented in the data set?
# sort(table(shooting_orig$gender), decreasing =FALSE)
count(shooting_orig, gender, sort = TRUE, name="Male/Female victims Count")


# How many male & female victims were presented in the data set.
shdat_gender <- shooting_orig %>%
```

```r
  filter(!is.na(gender))%>%
  group_by(gender) %>%
  arrange(gender)%>%
  summarise(count = n())

shdat_gender1 <- shooting_orig %>%
  group_by(gender) %>% # Variable to be transformed
  count() %>%
  ungroup() %>%
  mutate(perc = `n` / sum(`n`)) %>%
  arrange(perc) %>%
  mutate(labels = scales::percent(perc))

# race vs gender
shdat_race <- shooting_orig %>%
  filter(!is.na(gender))%>%
  filter(!is.na(race))%>%
  group_by(gender, race) %>%
  summarise(count = n(), .groups = "drop")

shdat_race1 <- shooting_orig %>%
  filter(!is.na(gender))%>%
  filter(!is.na(race))%>%
  group_by(gender, race, age) %>%
  summarise(count = n(), .groups = "drop")


# Male vs Race
shdat_male <- shooting_orig %>%
  filter(!is.na(gender))%>%
  filter(!is.na(race))%>%
  filter(gender=="male")%>%
  group_by(gender, race) %>%
  tally()

# Female vs Race
shdat_female <- shooting_orig %>%
  filter(!is.na(gender))%>%
  filter(!is.na(race))%>%
  filter(gender=='female')%>%
  group_by(gender, race) %>%
  tally()

# Create group armedTyPe for weapons
count(shooting_orig, armed_with, sort=TRUE)

group <- function (string){
  if(is.na(string)) return ("NA")
  if(string == "unarmed") return ("unarmed")
  else if (string == "undetermined") return ("undetermined")
  else if (string == "vehicle") return ("vehicle")
  else if (string %in% c("gun", "toy weapon", "gun and knife",
                         "gun and car", "BB gun",
```

```r
                           "guns and explosives", "gun and vehicle",
                           "hatchet and gun", "gun and sword",
                           "machete and gun", "vehicle and gun",
                           "pellet gun"))  return ("gun")
  else if (string %in% c("knife", "ax", "sword", "box cutter",
                           "hatchet", "sharp object",
                           "scissors", "meat cleaver", "pick-axe",
                           "straight edge razor",
                           "pitchfork", "chainsaw", "samurai sword",
                           "spear")) return ("sharpObject")
  else return ("other")
}
#apply group on each element of a vector
shooting_orig$armedType <- sapply(shooting_orig$armed_with, group)

# create age category to consolidate different age groups
# Identify different characteristics based on age, weapons, race and gender
shdata_char <- shooting_orig %>%
  mutate(ageCategory = case_when(between(age, 0, 10) ~ "0 - 10",
                                 between(age, 11, 20) ~ "11 - 20",
                                 between(age, 21, 30) ~ "21 - 30",
                                 between(age, 31, 40) ~ "31 - 40",
                                 between(age, 41, 50) ~ "41 - 50",
                                 between(age, 51, 60) ~ "51 - 60",
                                 between(age, 61, 70) ~ "61 - 70",
                                 between(age, 71, 80) ~ "71 - 80",
                                 between(age, 81, 91) ~ "81 - 91")) %>%
  select(gender, age, race, ageCategory, armedType, )

# Male victims age & armed type chars
shdata_male_chars <- shdata_char %>%
  filter(gender=='male')%>%
  group_by(ageCategory) %>%
  select(ageCategory, gender, race, armedType)

# female victims age & armed type chars
shdata_female_chars <- shdata_char %>%
  filter(gender=='female')%>%
  group_by(ageCategory) %>%
  select(ageCategory, gender, race, armedType)


# 1. How many male and female victims were presented in the data set?

# bar plots to show gender counts in the dataset

# g1 <- ggplot(shdat_gender, aes(x=gender, y=count, fill=gender)) +
#   geom_bar(stat = "identity", width = 0.5) +
#   ggtitle("Male / Female victims count in the dataset")+
#   scale_y_continuous(breaks = seq(0,3500,100), limits = c(0, 3500) )
#   # coord_flip()

g1 <- ggplot(shdat_gender1, aes(x = "", y = perc, fill = gender)) +
  geom_col(color = "black") +
```

```r
  geom_text(aes(label = labels),
            position = position_stack(vjust = 0.5)) +
  ggtitle("Male/Female victims percentage")+
  guides(fill = guide_legend(title = "Gender")) +
  # scale_fill_viridis_d() +
  scale_fill_manual(values=cbPalette)+
  coord_polar(theta = "y")

g2 <- ggplot(shdat_gender, aes(x=gender, y=count, fill=gender)) +
  geom_bar(stat = "identity", width=0.2) +
  # scale_fill_manual(values=cbPalette)+
  scale_fill_colorblind()+
  ggtitle("Male/Female victims count")+
  labs(y = "Count", x = "Gender" )+
  scale_y_continuous(breaks = seq(0,2800,400), limits = c(0, 2800) )+
  theme_minimal()+
  annotate(geom = 'text', label = 'A ', x = Inf, y = -Inf,
           hjust = 1, vjust = -1, size =10)+
  # coord_flip()+
  theme(text = element_text(size = 11))

g3 <- ggplot(shooting_orig, aes(x=state, fill=race)) +
  geom_bar() +
  ggtitle("Shooting distribution by States") +
  # scale_y_continuous(expand = c(0, 0)) +
  labs(y = "Count", x = "States" )+
  scale_y_continuous(breaks = seq(0,450,50), limits = c(0, 450))+
  # scale_fill_manual(values=cbPalette)+
  annotate(geom = 'text', label = 'B ', x = Inf, y = -Inf,
           hjust = 0.6, vjust = -2, size =10)+
  theme(axis.text.x = element_text(size=8, angle=90),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.background = element_blank(),
        axis.line = element_line(colour = "black"))+
  theme(text = element_text(size = 11))

grid.arrange(g2,g1, ncol=2)
grid.arrange(g3, ncol=1)
```

```r
# 2. Using race as an example, how many male victims were White, Black,
# Hispanic, Native American, Asian, and Other,
#  in comparison, how many female victims were White, Black,
# Hispanic, Native American, Asian, and Other?

g4 <- ggplot(shdat_race, aes(x=gender, y= count, fill=race)) +
  geom_bar(stat="identity", position=position_dodge())+
  scale_y_continuous(breaks = seq(0,1500,300), limits = c(0, 1500))+
  scale_fill_colorblind()+
  annotate(geom = 'text', label = 'A ', x = Inf, y = -Inf,
           hjust = 1, vjust = -1, size =12)+
  labs(title = "Race distribution by Gender",
       x = "Gender", y="Count") +
```

```r
    theme(axis.text.x = element_text(size=11, angle=0),
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          panel.background = element_blank(),
          axis.line = element_line(colour = "black"))+
  theme(text = element_text(size = 11))
  # hjust = 1, vjust = -1, size =12)+
  # theme(text = element_text(size = 12, family = 'Garamond'),
  #        legend.position = 'bottom',
  #        legend.direction = 'horizontal')

g5 <- ggplot(shdat_male, aes(x=gender, y= n, fill=race)) +
  geom_bar(stat="identity", width = 0.5, position=position_dodge())+
  # scale_y_continuous(expand = c(0, 0)) +
  scale_y_continuous(breaks = seq(0,1500,300), limits = c(0, 1500))+
  scale_fill_colorblind()+
  labs(title = "Race distribution by Male Victims",
       x = "Gender", y="Count") +
  annotate(geom = 'text', label = 'B ', x = Inf, y = -Inf,
           hjust = 1, vjust = -1, size =12)+
  theme(axis.text.x = element_text(size=11, angle=0),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.background = element_blank(),
        axis.line = element_line(colour = "black"))+
  theme(text = element_text(size = 11))+
    coord_flip()

g6 <- ggplot(shdat_female, aes(x=gender, y= n, fill=race)) +
  geom_bar(stat="identity", width = 0.5, position=position_dodge())+
  # scale_y_continuous(expand = c(0, 0)) +
  scale_y_continuous(breaks = seq(0,100,25), limits = c(0, 100))+
  scale_fill_colorblind()+
  coord_flip()+
  labs(title = "Race distribution by Female Victims",
       x = "Gender", y="Count") +
  theme(axis.text.x = element_text(size=11, angle=0),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.background = element_blank(),
        axis.line = element_line(colour = "black"))+
  theme(text = element_text(size = 11))

#Age and Gender Plot
shooting_orig$gender <- as.factor(shooting_orig$gender)
g9 <- ggplot(shdata_char, aes(x=ageCategory, fill=gender)) +
  geom_bar() +
  labs(title = "Shooting distribution by Age & Gender",
       x = "Age Category in Years", y="Count") +
  scale_y_continuous(breaks = seq(0,850,50), limits = c(0, 850))+
  geom_text( aes(label=after_stat(count)), stat='count',
             nudge_x=-0.14, nudge_y=0.125, width=2 )+
  annotate(geom = 'text', label = 'C ', x = Inf, y = -Inf,
```

```
                    hjust = 1, vjust = -1, size =12)+
      theme(axis.text.x = element_text(size=11, angle=90),
            panel.grid.major = element_blank(),
            panel.grid.minor = element_blank(),
            panel.background = element_blank(),
            axis.line = element_line(colour = "black"))+
      scale_fill_manual(values=cbPalette)

    #Race and Gender Plot
    g10 <- ggplot(shooting_orig[!(shooting_orig$race %in% c(NA)), ],
                  aes(x=race, fill=gender)) + geom_bar() +
      scale_y_continuous(breaks = seq(0,1500,250), limits = c(0, 1500))+
      facet_wrap(~gender) +
      labs(title = "Shooting Distribution by Race and Gender",
           x = "Race", y="Count") +
      geom_text( aes(label=after_stat(count)),
                 stat='count', nudge_x=-0.14, nudge_y=0.125,width=2 )+
      annotate(geom = 'text', label = 'D ', x = Inf, y = -Inf,
               hjust = .8, vjust = -1, size =12)+
      scale_fill_manual(values=cbPalette)


    grid.arrange(g4, ncol=1)
    grid.arrange(g5, g6, ncol=2)
    grid.arrange(g9, ncol=1)
    grid.arrange(g10, ncol=1)


    # # 3. What are the characteristics of male vs. female victims
    # in terms of age, race, weapons, etc?

    g7 <- ggplot(shdata_female_chars, aes(x=race, y=ageCategory , fill=armedType)) +
      geom_bar(stat="identity", position=position_dodge())+
      # ggtitle("Female victims race, age and armed type characteristics")
      # scale_fill_manual(values=cbPalette)+
      scale_fill_colorblind()+
      labs(title = "Female victims characteristics by race, age and armed type",
           x = "Race", y="Age Category in Years") +
      annotate(geom = 'text', label = 'A ', x = Inf, y = -Inf,
               hjust = 0.6, vjust = -1, size =11)+
      theme(axis.text.x = element_text(size=11, angle=0),
            panel.grid.major = element_blank(),
            panel.grid.minor = element_blank(),
            panel.background = element_blank(),
            axis.line = element_line(colour = "black"))+
      theme(text = element_text(size = 11))

    g8 <- ggplot(shdata_male_chars, aes(x=race, y=ageCategory , fill=armedType)) +
      geom_bar(stat="identity", position=position_dodge())+
      # geom_bar(stat = "identity", width = 0.5) +
      ggtitle("Male victims race, age and armed type characteristics")+
      scale_fill_colorblind()+
      labs(title = "Male victims characteristics by race, age and armed type",
           x = "Race", y="Age Category in Years") +
```

```
    annotate(geom = 'text', label = 'B ', x = Inf, y = -Inf,
             hjust = 0.6, vjust = -1, size =12)+
    theme(axis.text.x = element_text(size=11, angle=0),
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          panel.background = element_blank(),
          axis.line = element_line(colour = "black"))+
    theme(text = element_text(size = 11))

grid.arrange(g7, ncol=1)
grid.arrange(g8, ncol=1)
```