Term Project Part 1 - Police shooting data

```
##
##    F    M
##  166 3497
```

```
## # A tibble: 2 x 2
##   gender     n
##   <chr>  <int>
## 1 M       3497
## 2 F        166
```

```
## # A tibble: 82 x 2
##    armed              n
##    <chr>          <int>
##  1 gun             2159
##  2 knife            578
##  3 unarmed          277
##  4 toy weapon       148
##  5 undetermined     121
##  6 vehicle           63
##  7 unknown weapon    44
##  8 machete           34
##  9 Taser             16
## 10 sword             16
## # i 72 more rows
```

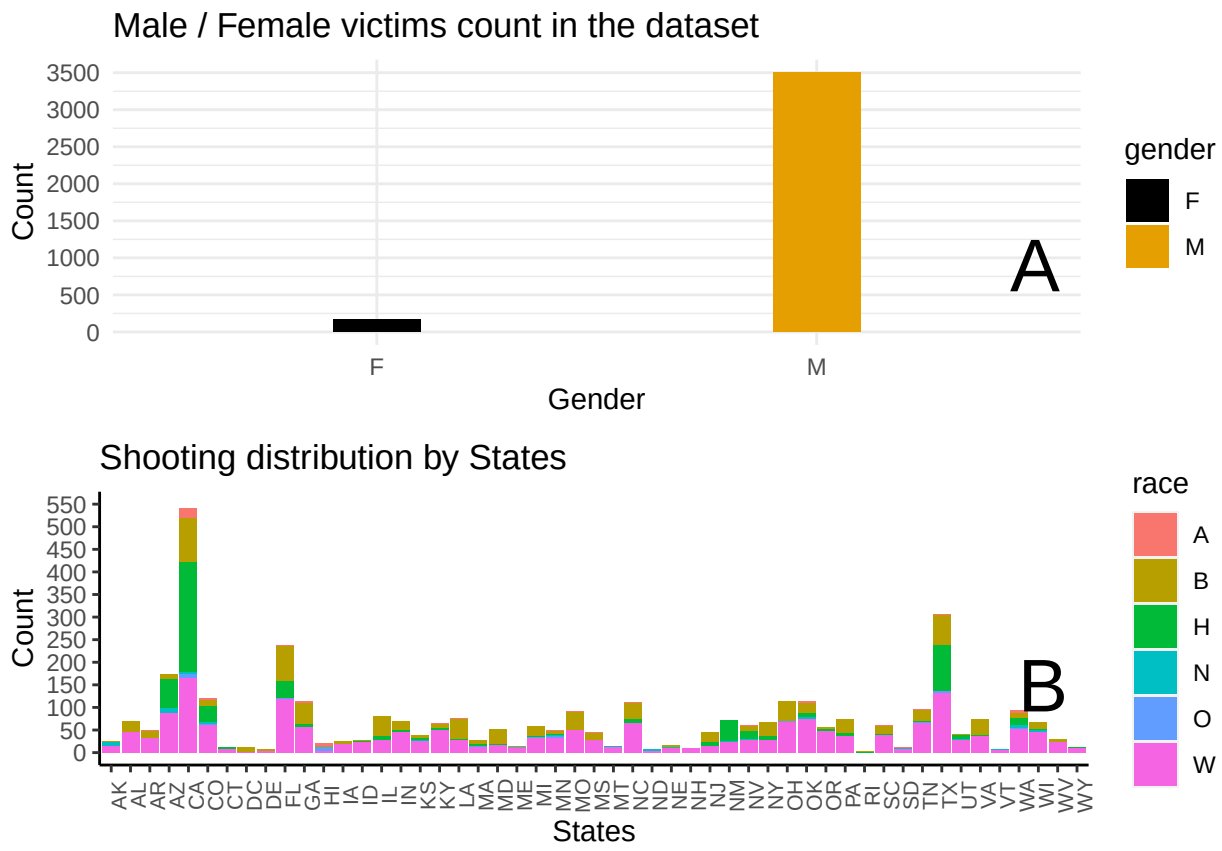
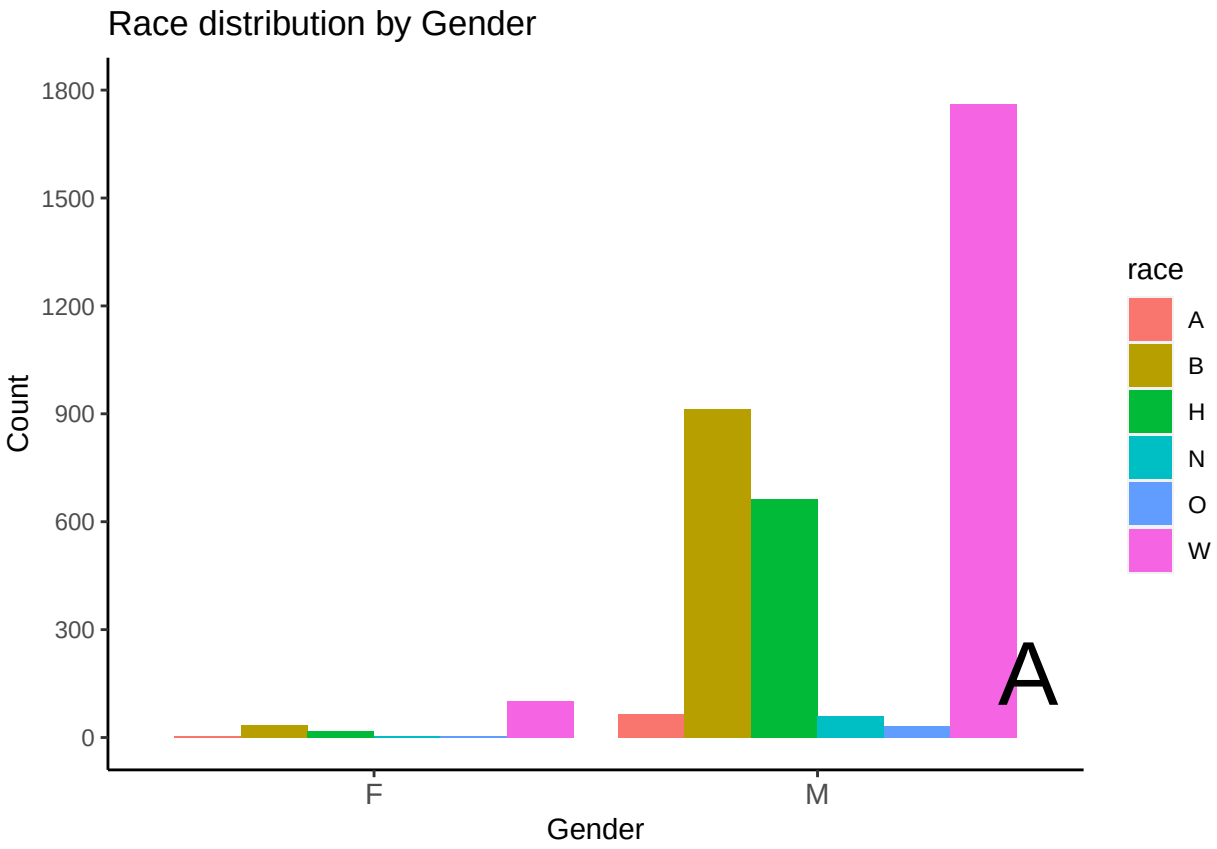


Figure 1. Panel A shows Male/Female victims count in the datatset and Panel B shows shooting distribution of race by states.


Race distribution by Gender
Count
1800
1500
1200
900
600
300
0
F
M
Gender
race
A
B
H
N
O
W
A

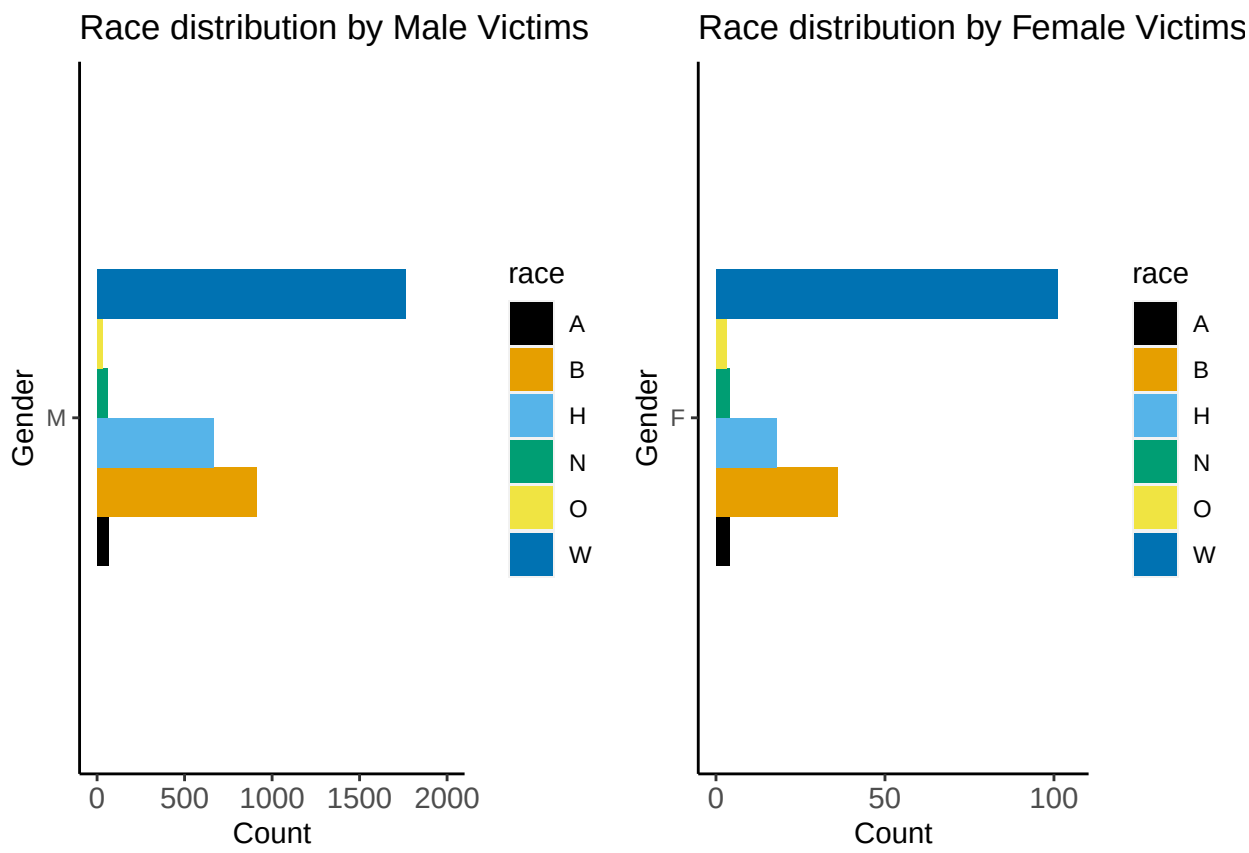


Figure 2. Panel A shows race distribution by gender

Male/Female victims of race distribuition are shown separately.

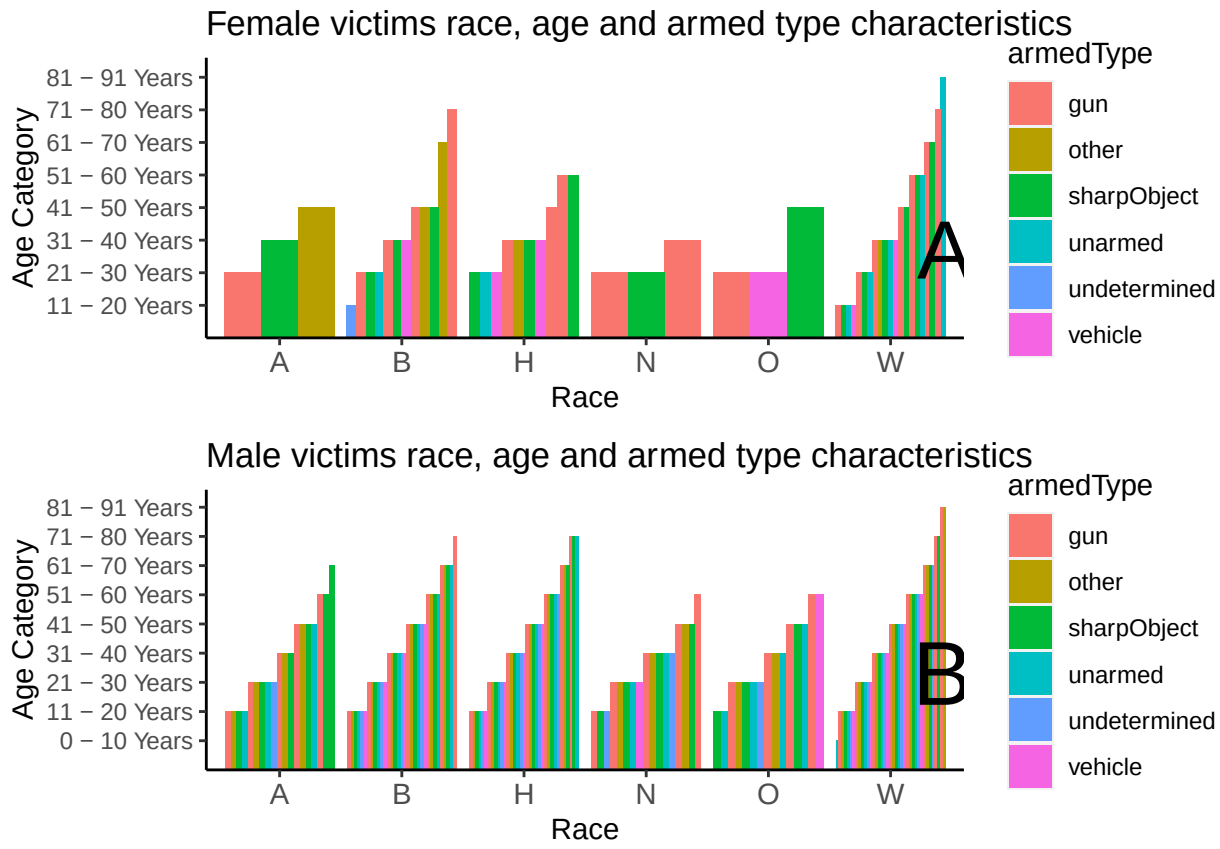


Figure 3. Panel A shows Male victims characteristics based on age, race, weapons.

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.

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.

```
# 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-data.csv",
                          col_names = TRUE, na="")

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 & female victims were presented in the data set.
shdat_gender <- shooting_orig %>%
  filter(!is.na(gender))%>%
  group_by(gender) %>%
  arrange(gender)%>%
```

```
  summarise(count = n())

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

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

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

# Create group armedTyPe for weapons
count(shooting_orig, armed, 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",
                         "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, 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 Years",
                                 between(age, 11, 20) ~ "11 - 20 Years",
                                 between(age, 21, 30) ~ "21 - 30 Years",
                                 between(age, 31, 40) ~ "31 - 40 Years",
                                 between(age, 41, 50) ~ "41 - 50 Years",
                                 between(age, 51, 60) ~ "51 - 60 Years",
                                 between(age, 61, 70) ~ "61 - 70 Years",
                                 between(age, 71, 80) ~ "71 - 80 Years",
                                 between(age, 81, 91) ~ "81 - 91 Years")) %>%
  select(gender, age, race, ageCategory, armedType)

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

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

```
# box 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()

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 in the dataset")+
  labs(y = "Count", x = "Gender" )+
  scale_y_continuous(breaks = seq(0,3500,500), limits = c(0, 3500) )+
  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(0,550,50), limits = c(0, 550))+
  annotate(geom = 'text', label = 'B ', x = Inf, y = -Inf,
           hjust = 1, vjust = -1, 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, g3, ncol=1)
```

```
g4 <- ggplot(shdat_race, aes(x=gender, y= count, fill=race)) +
  geom_bar(stat="identity", position=position_dodge())+
  scale_y_continuous(breaks = seq(0,1800,300), limits = c(0, 1800))+
  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") +
  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(0,2000,500), limits = c(0, 2000))+
  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(0,105,50), limits = c(0, 105))+
  scale_fill_colorblind()+
  coord_flip()+
  labs(title = "Race distribution by Female Victims",
       x = "Gender", y="Count") +
```

```
  # geom_text(aes(label = 'C '), hjust = -1)+
  # annotate(geom = 'text', label = 'C ', x = Inf, y = -Inf,
  #          hjust = 1, vjust = -1, size =10)+
  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))

  # theme(text = element_text(size = 11))+

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

```
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")
  labs(title = "Female victims race, age and armed type characteristics",
       x = "Race", y="Age Category") +
  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")+
  labs(title = "Male victims race, age and armed type characteristics",
       x = "Race", y="Age Category") +
  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,  g8, ncol=1)
```