# Term Project Part II - Police shooting data (v2)

## Code Used To Make These Plots

Below are the code chunks used to make these plots.

```
# Use the police shooting data and complete the following:
#
# 1.Describe the police shooting data set by filling out
# the table below (follow the 'age' example)
#
# attribute     definition          data type    % of missing data
# age   age of the victim at the
# time of the incident                number  0%
#
# 2. In the demo code, the correlation between state
# incident count and state population was examined.
# What other correlation analyses could be done?
# For example, does age correlate with weapon used?
# Perform a correlation analysis (Pearson or X-squared)
# between a pair of variables of your choice.

rm(list = ls())

library(ggplot2)
library(dplyr)
library(gridExtra)
library(tidyverse)
library(hrbrthemes)
library(ggthemes)
library(corrplot)
library(patchwork)
library(scales)
library(ggcorrplot)


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

#view the data
str(shooting_orig)
sapply(shooting_orig, typeof)
```

```r
# #examine missing data
filter(shooting_orig, !complete.cases(shooting_orig))
# apply(shooting_orig, 2, function(x) sum(is.na(x))) %>%
#   sort(decreasing=TRUE)

# Get % of Missing data
apply(shooting_orig, 2, function(x) percent(mean(is.na(x)),accuracy=1)) %>%
  sort(decreasing=TRUE)

# Split date & time
shooting_orig <- shooting_orig %>%
  mutate(date =  format(as.POSIXct(date, format='%Y-%m-%d'))) %>%
  mutate(year = year(date)) %>%
  mutate(month = month(date))

group <- function (string){
  if(is.na(string)) return ("NA")
  if(string == "unarmed") return ("Unarmed")
  else if (string %in% c("undetermined", "unknown")) return ("Undetermined")
  else if (string == "vehicle") return ("Vehicle")
  else if (string %in% c("gun", "gun;knife",
                         "other;gun",  "replica",
                         "gun;vehicle", "vehicle;gun"))  return ("Gun")
  else if (string %in% c("knife", "blunt_object",
"blunt_object;blunt_object",
                         "sharp object", "knife;blunt_object"))
                       return ("Sharp Object")
  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
shooting_orig <- 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(state, age, gender, race, ageCategory, armedType, year)
```

```r
# state region mapping
state_region_mapping <- data.frame(
 state = c(
 "CA", "CO", "WA", "MT", "ID", "OR", "WY", "UT", "NV", "HI",
 "ND", "SD", "MN", "WI","MI", "NE","KS", "MO", "IA", "IL", "IN", "OH",
 "NY", "NJ", "VT", "PA", "MD", "NH", "CT", "RI", "DE", "DC", "MA",
 "AZ","NM", "OK", "TX",
 "SC", "NC", "GA", "FL", "LA", "MS", "AL", "AR", "TN", "KY", "WV", "ME",
"VA", "AK"),
 region = c(
 "West", "West", "West", "West", "West", "West", "West", "West","West",
"West",
 "Midwest", "Midwest", "Midwest", "Midwest","Midwest", "Midwest", "Midwest",
"Midwest",
 "Midwest", "Midwest", "Midwest", "Midwest",
 "Northeast", "Northeast", "Northeast", "Northeast", "Northeast",
"Northeast",
 "Northeast", "Northeast","Northeast", "Northeast", "Northeast",
 "Southwest", "Southwest", "Southwest", "Southwest",
 "Southeast", "Southeast","Southeast", "Southeast", "Southeast", "Southeast",
 "Southeast", "Southeast","Southeast", "Southeast", "Southeast",
"Southeast","Southeast",
 "Southeast"))
# Join the Region column on ufo dataset
shooting_orig_region <- shooting_orig %>%
  filter(!is.na(state))%>%
  filter(!is.na(age))%>%
  filter(!is.na(gender))%>%
  filter(gender!="non-binary")%>%
  filter(!is.na(race))%>%
  filter(!is.na(armedType))%>%
  filter(armedType!="NA")%>%
 left_join(state_region_mapping, by = "state") %>%
  select(state, age, gender, race, ageCategory, armedType, year, region)

shooting_orig_reg<- shooting_orig_region %>%
 filter(!is.na(region)) %>%
 group_by(region, age) %>%
 arrange(region)%>%
 summarise(wcount = n())

#create joint data frame to run correlation
(state_count <- arrange(count(shooting_orig_region, state), desc(n)))
(race_count <- arrange(count(shooting_orig_region, race), desc(n)))
(weapon_count <- arrange(count(shooting_orig_region, armedType), desc(n)))

(age_join_arms <- inner_join(weapon_count, shooting_orig_region[,
c("armedType", "age")],
```

```
                              by=c("armedType")))
(year_join_arms <- inner_join(weapon_count, shooting_orig_region[,
c("armedType", "year")],
                              by=c("armedType")))
(race_join_age <- inner_join(race_count, shooting_orig_region[, c("race",
"age")],
                              by=c("race")))

# ** Correlation relationship between two variables**
# Perform the correlation test
cr1 <- cor.test(shooting_orig_reg$age, shooting_orig_reg$wcount)
print("Age vs Region")
print(cr1)
cr2 <- cor.test(race_join_age$age, race_join_age$n)
print("Age vs Race")
print(cr2)
cr3 <- cor.test(age_join_arms$age, age_join_arms$n)
print("Age vs Weapons")
print(cr3)
cr4 <- cor.test(year_join_arms$year, year_join_arms$n)
print("Year vs Weapons")
print(cr4)

# Correlation Matrix
model.matrix(~0+., data=race_join_age) %>%
  cor(use="pairwise.complete.obs") %>%
  ggcorrplot(show.diag=FALSE, type="lower", lab=TRUE, lab_size=2)

model.matrix(~0+., data=shooting_orig_reg) %>%
  cor(use="pairwise.complete.obs") %>%
  ggcorrplot(show.diag=FALSE, type="lower", lab=TRUE, lab_size=2)

model.matrix(~0+., data=year_join_arms) %>%
  cor(use="pairwise.complete.obs") %>%
  ggcorrplot(show.diag=FALSE, type="lower", lab=TRUE, lab_size=2)

model.matrix(~0+., data=age_join_arms) %>%
  cor(use="pairwise.complete.obs") %>%
  ggcorrplot(show.diag=FALSE, type="lower", lab=TRUE, lab_size=2)

# **Correlation relationship Plots**

c1 <- ggplot(shooting_orig_reg) +
  aes(x = age, y = wcount) +
  geom_point(colour = "green") +
  theme_minimal()+
  # scale_fill_colorblind()+
  ggtitle("Age vs Region relationship")+
  labs(y = "count", x = "Age" )+
```

```r
  scale_x_continuous(breaks = seq(0,95,5), limits = c(0, 95) )+
  scale_y_continuous(breaks = seq(0,80,5), limits = c(0, 80) )
c1

c2 <- ggplot(race_join_age) +
  aes(x = age, y = n) +
  # geom_point(colour = "#0c4c8a") +
  geom_point(colour = "red") +
  theme_minimal()+
  # scale_fill_colorblind()+
  ggtitle("Age vs Race relationship")+
  labs(y = "Count", x = "Age" )+
  scale_y_continuous(breaks = seq(0,4000,500), limits = c(0, 4000) )+
  scale_x_continuous(breaks = seq(0,95,5), limits = c(0, 95) )
c2

c3 <- ggplot(age_join_arms) +
  aes(x = age, y = n) +
  # geom_point(colour = "#0c4c8a") +
  geom_point(colour = "blue") +
  theme_minimal()+
  # scale_fill_colorblind()+
  ggtitle("Age vs Weapons relationship")+
  labs(y = "Count", x = "Age" )+
  scale_y_continuous(breaks = seq(0,1850,200), limits = c(0, 1850) )+
  scale_x_continuous(breaks = seq(0,95,5), limits = c(0, 95) )
c3

c4 <- ggplot(year_join_arms) +
  aes(x = year, y = n) +
  geom_point(colour = "#0c4c8a") +
  theme_minimal()+
  # scale_fill_colorblind()+
  ggtitle("Year vs Weapons relationship")+
  labs(y = "Count", x = "Year" )+
  scale_y_continuous(breaks = seq(0,5500,500), limits = c(0, 5500) )+
  scale_x_continuous(breaks = seq(2015,2023,1), limits = c(2015, 2023) )
c4
```