US Mass Shooting - EDA

Satish

10/3/2017

```
install.packages("tidyr")
install.packages("tidyverse")
install.packages("plotly")
```

```
require(tidyr)
require(tidyverse)
require(plotly)
```

```
# Loading US Mass Shooting Data to a dataframe
data_ms <- read_csv('/Users/satishreddychirra/Document/Data Management & Processing/HW MV/Mass Shootings
```

```
## Parsed with column specification:
## cols(
##   `S#` = col_integer(),
##   Title = col_character(),
##   Location = col_character(),
##   Date = col_character(),
##   Summary = col_character(),
##   Fatalities = col_integer(),
##   Injured = col_integer(),
##   `Total victims` = col_integer(),
##   `Mental Health Issues` = col_character(),
##   Race = col_character(),
##   Gender = col_character(),
##   Latitude = col_double(),
##   Longitude = col_double()
## )
```

```
# Displaying first 10 records from the dataset
head(data_ms,10)
```

```
## # A tibble: 10 x 13
##     `S#`                                Title                Location
##    <int>                                <chr>                   <chr>
##  1     1        Las Vegas Strip mass shooting            Las Vegas, NV
##  2     2           San Francisco UPS shooting        San Francisco, CA
##  3     3    Pennsylvania supermarket shooting          Tunkhannock, PA
##  4     4 Florida awning manufacturer shooting         Orlando, Florida
##  5     5     Rural Ohio nursing home shooting       Kirkersville, Ohio
##  6     6             Fresno downtown shooting       Fresno, California
##  7     7     Fort Lauderdale airport shooting Fort Lauderdale, Florida
##  8     8               Cascade Mall shooting           Burlington, WA
##  9     9          Baton Rouge police shooting          Baton Rouge, LA
## 10    10               Dallas police shooting            Dallas, Texas
## # ... with 10 more variables: Date <chr>, Summary <chr>, Fatalities <int>,
## #   Injured <int>, `Total victims` <int>, `Mental Health Issues` <chr>,
## #   Race <chr>, Gender <chr>, Latitude <dbl>, Longitude <dbl>
```

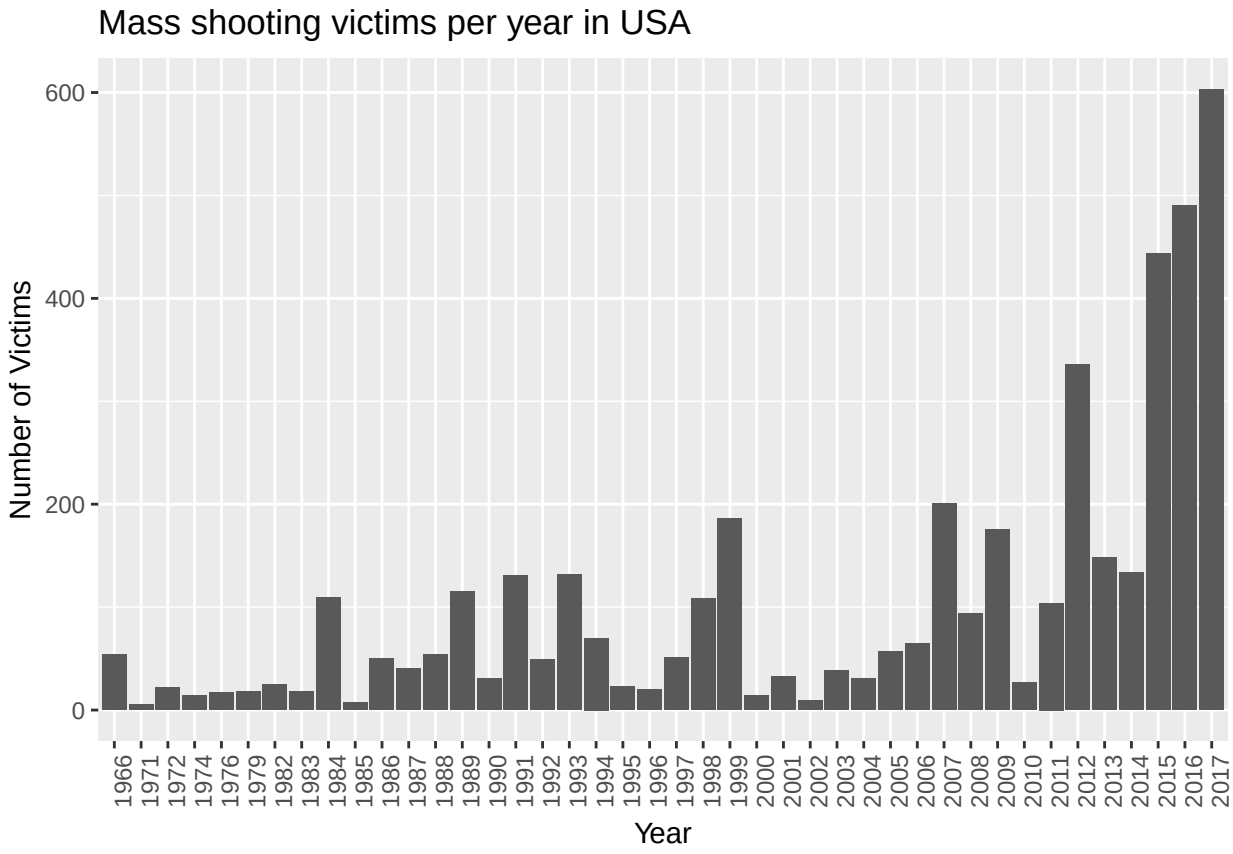
Step 1: Perform exploratory data analysis on the dataset, using the techniques learned in class. Calculate

summary statistics that are of interest to you and create plots using ggplot2 that show your findings.

```
# Year wise mass shooting in USA
df_msy <- data_ms %>%
            select(Date, `Total victims`) %>%
            group_by(format(as.Date(Date, format="%m/%d/%Y"),"%Y")) %>%
            summarise(sum(`Total victims`))
```

```
## Warning in strptime(x, format, tz = "GMT"): unknown timezone 'zone/tz/
## 2017c.1.0/zoneinfo/America/New_York'
```

```
colnames(df_msy) <- c('Year','Total_Victims')

# Plotting mass shooting victims per year in USA
yw <- ggplot(data=df_msy, mapping = aes(x = Year, y = Total_Victims)) +
        geom_bar(stat = "identity") +
        xlab("Year") + ylab("Number of Victims") + labs(title = "Mass shooting victims per year in USA")
        theme(axis.text.x = element_text(angle = 90))

yw
```



We can see from the above that, there is a abrupt increase in the number of victims effected due to mass shooting in last three years.

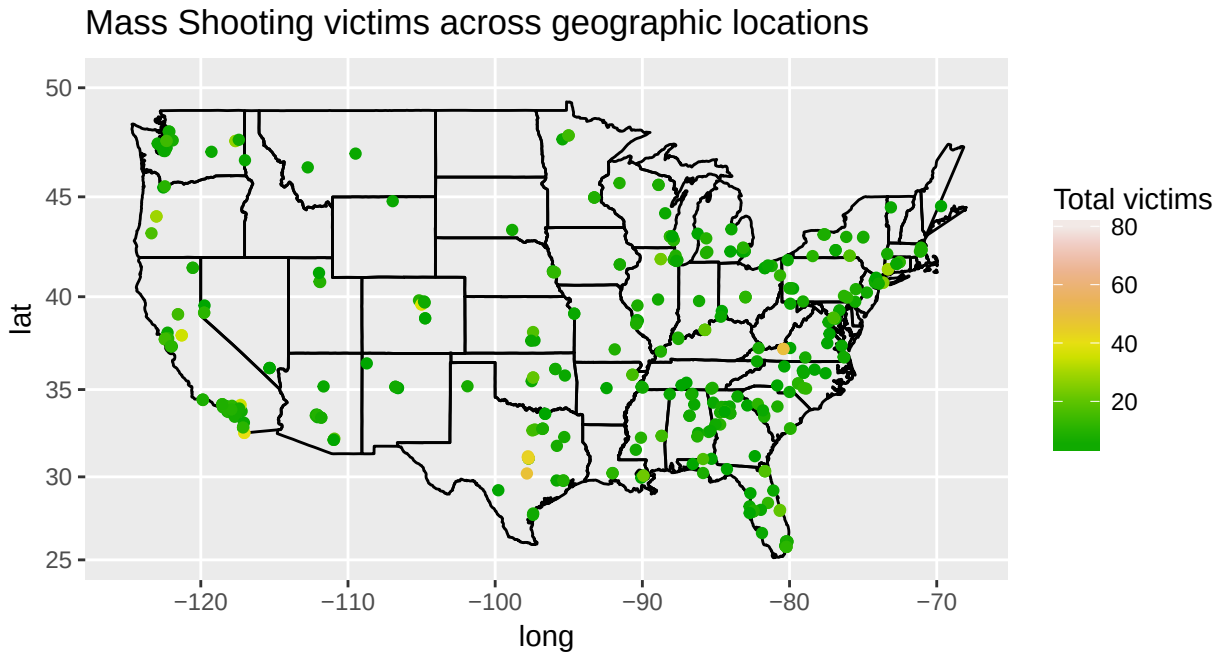
```
# Mass Shooting victims as per the geographic location
statesUSA <- map_data("state")
```

```
##
## Attaching package: 'maps'
```

```
## The following object is masked from 'package:purrr':
##
##     map
```

```
ms <- ggplot(data = statesUSA) +
        geom_polygon(aes(x = long, y = lat, group = group), fill = NA, color = "black") +
        coord_map() +
        geom_point(data = na.omit(data_ms), aes(x =Longitude, y = Latitude, color = `Total victims`)) +
        scale_x_continuous(limits = c(-125,-68)) +
        scale_y_continuous(limits = c(25,50)) +
        scale_colour_gradientn(colours = terrain.colors(10)) +
        labs(title = "Mass Shooting victims across geographic locations")

ms
```

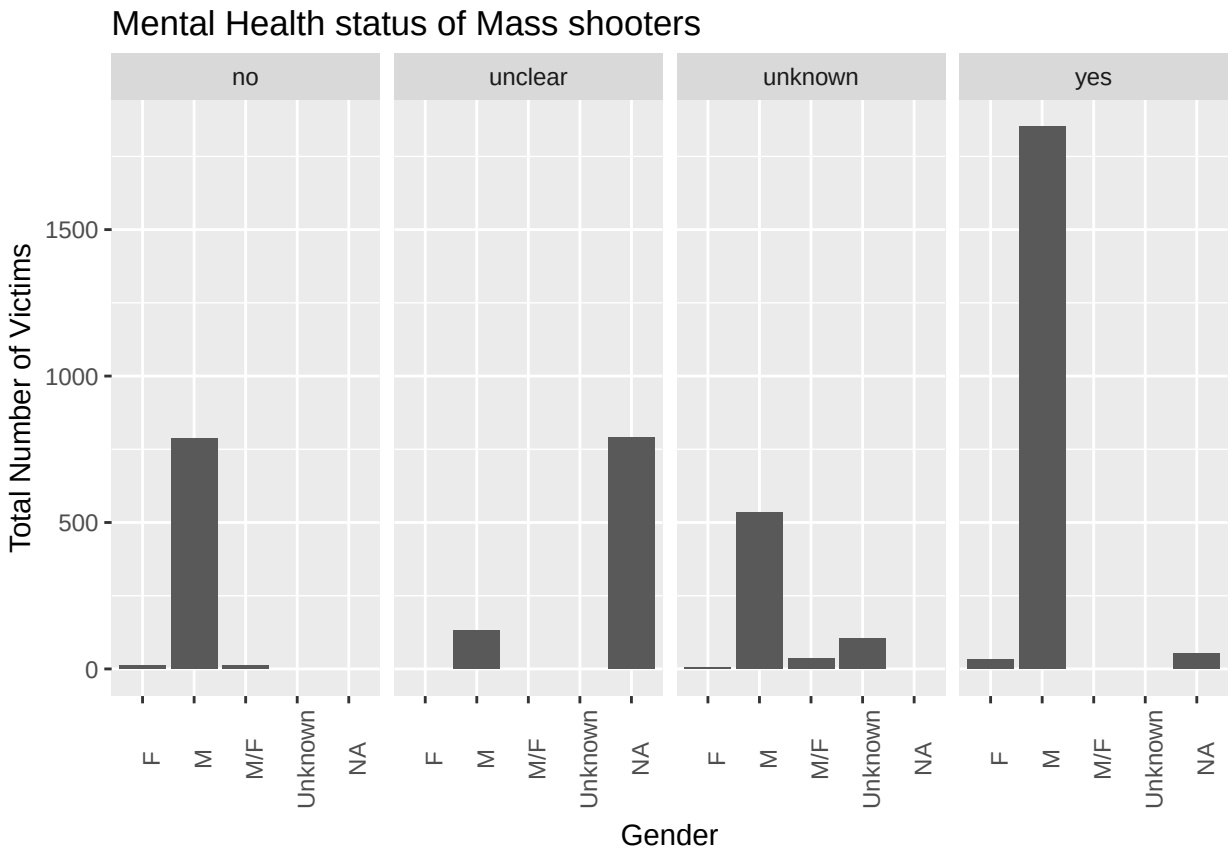
```
## Warning: Removed 3 rows containing missing values (geom_point).
```




```
# Representing the column in a uniform notation
data_ms$Gender <- ifelse(data_ms$Gender == 'Female','F',ifelse(data_ms$Gender == 'Male','M',ifelse(data_

data_hs <- data_ms %>%
              group_by(tolower(`Mental Health Issues`), Gender) %>%
              summarise(sum(`Total victims`))

colnames(data_hs) <- c('Mental_Health_Issues','Gender','Total_Victims')

# Plotting Mental Health status of Mass shooters
mhs <- ggplot(data = data_hs, mapping = aes(x = Gender, y = Total_Victims)) +
        theme(axis.text.x = element_text(angle = 90)) +
        geom_histogram(stat = "identity") + ylab("Total Number of Victims") +
        facet_grid(~Mental_Health_Issues) +
        labs(title = "Mental Health status of Mass shooters")
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```

```
mhs
```



From the above, we can see that majority of mass shooters are men with metal health issues.