Assignment 5.2 - Week 9&10 in R

Aarti Ramani

2023-08-09

Load required libraries

```
library(readxl)
library(ggplot2)
library(plotly)
library(dplyr)
```

Read xls into a dataframe

```
crime_df <- read.csv("C:/Masters/GitHub/Summer2023/DSC640-Data Presentation & Visualization/Week9&10/ex6-2/crimer
atesbystate-formatted.csv")
nrow(crime_df)
```

```
## [1] 52
```

```
head(crime_df,5)
```

```
##            state murder forcible_rape robbery aggravated_assault burglary
## 1 United States     5.6          31.7   140.7              291.1    726.7
## 2       Alabama     8.2          34.3   141.4              247.8    953.8
## 3        Alaska     4.8          81.1    80.9              465.1    622.5
## 4       Arizona     7.5          33.8   144.4              327.4    948.4
## 5       Arkansas    6.7          42.9    91.1              386.8   1084.6
##   larceny_theft motor_vehicle_theft
## 1        2286.3               416.7
## 2        2650.0               288.3
## 3        2599.1               391.0
## 4        2965.2               924.4
## 5        2711.2               262.1
```

```
education_df <- read.csv("C:/Masters/GitHub/Summer2023/DSC640-Data Presentation & Visualization/Week9&10/ex6-2/ed
ucation.csv")
nrow(education_df)
```

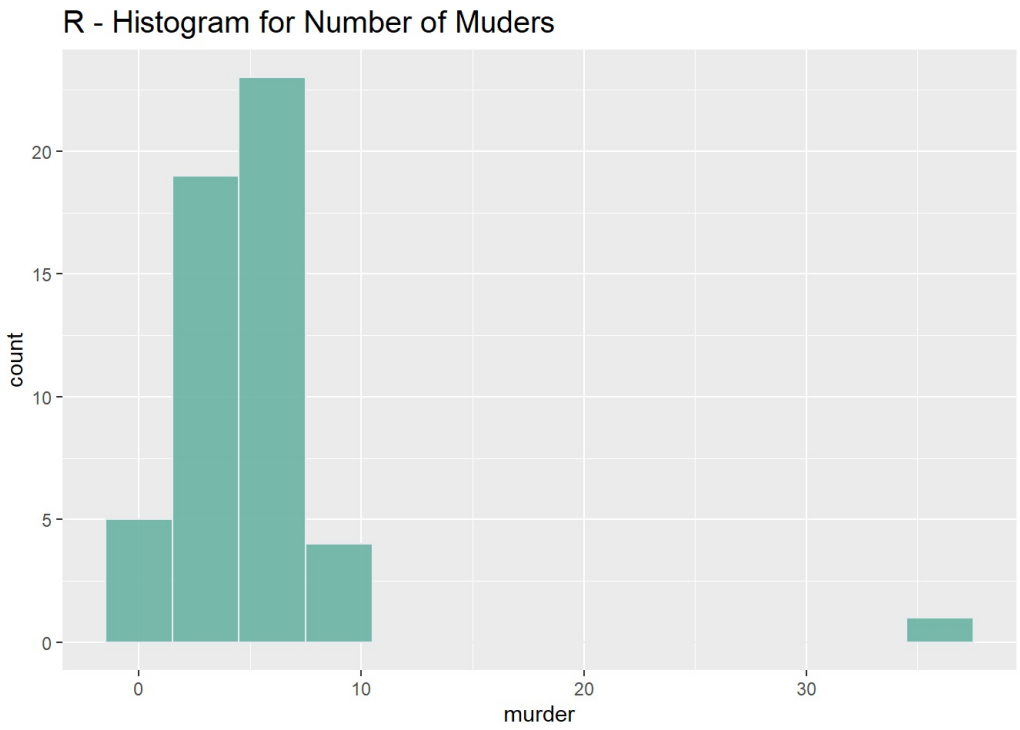
```
## [1] 52
```

```
head(education_df,5)
```

```
##           state reading math writing percent_graduates_sat pupil_staff_ratio
## 1 United States     501  515     493                    46              7.9
## 2       Alabama     557  552     549                     7              6.7
## 3        Alaska     520  516     492                    46              7.9
## 4       Arizona     516  521     497                    26             10.4
## 5      Arkansas     572  572     556                     5              6.8
##   dropout_rate
## 1          4.4
## 2          2.3
## 3          7.3
## 4          7.6
## 5          4.6
```

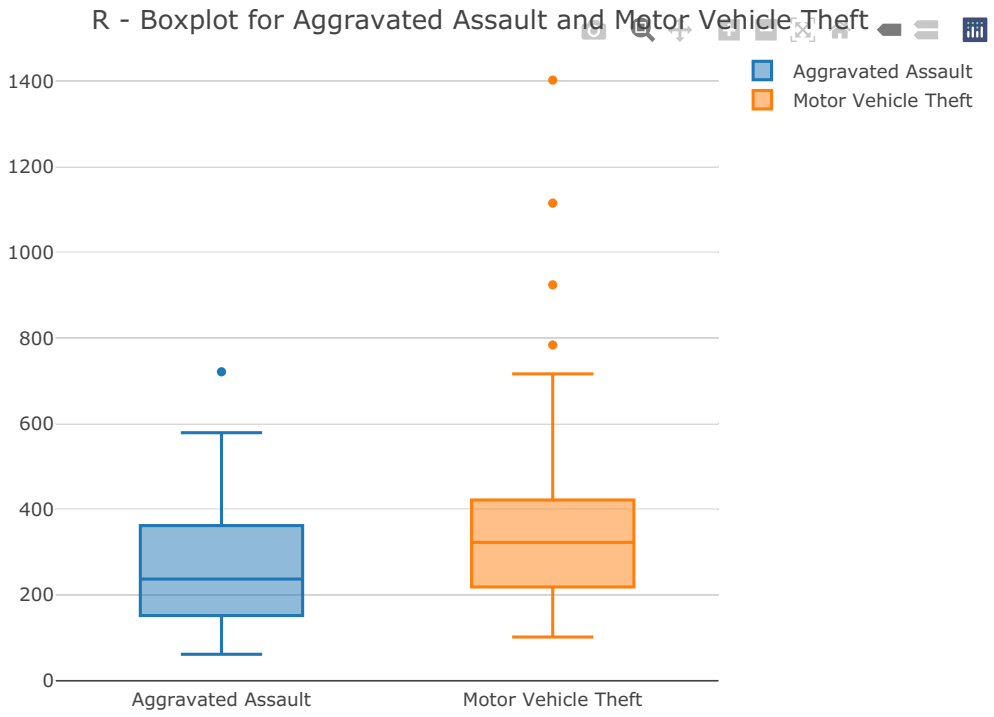
R - Histogram

```
p <- crime_df %>%
  ggplot( aes(x=murder)) +
    geom_histogram( binwidth=3, fill="#69b3a2", color="#e9ecef", alpha=0.9) +
    ggtitle("R - Histogram for Number of Muders") +
    theme(plot.title = element_text(size=15))
p
```



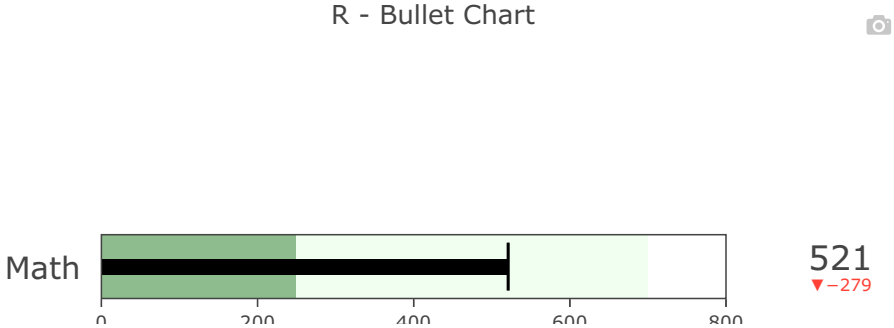
R - Box Plot

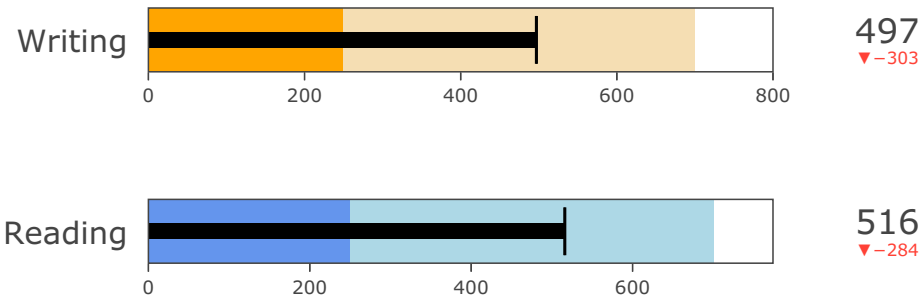
```
fig <- plot_ly(type = "box", y = crime_df$aggravated_assault, name="Aggravated Assault")
fig <- fig %>% add_trace(y = crime_df$motor_vehicle_theft, name="Motor Vehicle Theft")

fig <- fig %>% layout(title = "R - Boxplot for Aggravated Assault and Motor Vehicle Theft")

fig
```



R - Bullet Chart

```
bullet_chart_df = education_df[trimws(education_df$state)== "Arizona",]


fig <- plot_ly()
fig <- fig %>%
  add_trace(
    type = "indicator",
    mode = "number+gauge+delta",
    value = bullet_chart_df$reading,
    delta = list(reference = 800),
    domain = list(x = c(0.1, 1), y = c(0, 0.1)),
    title =list(text = "Reading"),
    gauge = list(
      shape = "bullet",
      axis = list(range = c(NULL, 800)),
      threshold = list(
        line= list(color = "black", width = 2),
        thickness = 0.75,
        value = bullet_chart_df$reading),
      steps = list(
        list(range = c(0, 250), color = "cornflowerblue"),
        list(range = c(250, 700), color = "lightblue")),
      bar = list(color = "black")))

fig <- fig %>%
  add_trace(
    type = "indicator",
    mode = "number+gauge+delta",
    value = bullet_chart_df$writing,
    delta = list(reference = 800),
    domain = list(x = c(0.1, 1), y = c(0.3, 0.4)),
    title = list(text = "Writing"),
    gauge = list(
      shape = "bullet",
      axis = list(range = list(NULL, 800)),
      threshold = list(
        line = list(color = "black", width= 2),
        thickness = 0.75,
        value = bullet_chart_df$writing),
      steps = list(
        list(range = c(0, 250), color = "orange"),
        list(range = c(250, 700), color = "wheat")),
      bar = list(color = "black")))

fig <- fig %>%
  add_trace(
    type =  "indicator",
    mode = "number+gauge+delta",
    value = bullet_chart_df$math,
    delta = list(reference = 800 ),
    domain = list(x = c(0.1, 1), y = c(0.6, 0.7)),
    title = list(text = "Math"),
    gauge = list(
      shape = "bullet",
      axis = list(range = list(NULL, 800)),
      threshold = list(
        line = list(color = "black", width = 2),
        thickness = 0.75,
        value = bullet_chart_df$math),
      steps = list(
        list(range = c(0, 250), color = "darkseagreen"),
        list(range = c(250, 700), color = "honeydew")),
      bar = list(color = "black")))

fig <- fig %>% layout(title='R - Bullet Chart')
fig
```




Writing
0
200
400
600
800
497
▼−303

Reading
0
200
400
600
516
▼−284

R - Additional chart : Funnel Chart

```
#create separate dataframes for reading and writing with the state names. Create lists for corresponding state na
mes for plotting
#Sort by reading score

reading_df <- education_df %>% group_by(state,reading) %>% count()  %>% arrange(desc(reading))
read_val <- head(reading_df, 5)[["reading"]]
read_state_val <- head(reading_df, 5)[["state"]] #get the top 5 states

writing_df <- education_df %>% group_by(state,writing) %>% count()  %>% arrange(desc(writing))
write_val <- head(writing_df, 5)[["writing"]]
write_state_val <- head(writing_df, 5)[["state"]] #get the top 5 states


fig <- plot_ly(type = "funnelarea",
    values = read_val, text = read_state_val,
    title = list(position = "top center", text = "Top 5 states with highest reading score",
                 font = list(size = 50)),
    textfont = list(family = "Old Standard TT, serif", size = 13, color = "black"),
    domain = list(x = c(0, 0.4), y =c(0.12, 1)))


fig <- fig %>% add_trace(
    type = "funnelarea",
    scalegroup = "first",
    values = write_val,text = write_state_val,
    title = list(position = "top left", text = "Top 5 states with highest writing score",
                 font = list(size = 50)),
    textfont = list(family = "Old Standard TT, serif", size = 13, color = "black"),
    domain = list(x = c(0.6, 1), y = c(0.12, 1)))


fig
```