Project - descriptive statistics

PCS - UConn, 2022

PUT YOUR NAME HERE

General Instructions

- Answer the questions by inserting R code and necessary comments. Your output must contain the R code (do not use the echo=FALSE option).

The Data

In a Washington Post article from October 6, 2015, titled "Zero correlation between state homicide rate and state gun laws", Eugene Volokh examined the relationship between the total number of gun deaths (per 100,000 people) and gun laws.

Source: The Washinton Post, Oct. 6 2015

To do that, Volokh used the '**Brady score**', which represents how difficult it is to obtain a gun in a state. A low Brady scores means a low level of gun restrictions (so it is easier to obtain a gun), and a high score means that it is harder to get a gun (i.e. there are stricter gun laws).

The data set, available in the file guns.csv, contains the following variables:

- Jurisdiction: State/jurisdiction name
- Homicide rate: Number of homicide death per 100,000 people
- Gun accident rate: Number of gun accident death per 100,000 people
- Sum: The summation of homicide rate and gun accident rate
- Brady score: The Brady score represents how difficult it is to obtain a gun in a state.
- Brady grade: Another interpretation of Brady Score

Question 1

a) Load the data `guns.csv` into R and store it as **gunData** (use the `read.csv` function).
b) Attach the data set (and remember to detach it at the end of the file)
c) Check the dimensions and structre of the dataset

```
# put your code here
```

Question 2

We now create a new categorical variable **bScoreCategory** which divides the Brady Scores into two categories: high (H) if the score is greater than 9, and low (L) if it is less than or equal to 9. Next, we convert the variable to be a factor. (uncomment the following three lines of R code).

```
#bScoreCategory <- rep("L", dim(gunData)[1])
#bScoreCategory[which(Brady.score > 9)] <- "H"
#bScoreCategory <- as.factor(bScoreCategory)
```

a) Display the content of the new variable.
b) Use the `table` function to show the number of H's and the number of L's.

```
# put your code here
```

Question 3

a) Create a new logical variable **Homicide_index** which divides the Homicide.rate into two categories: TRUE if the rate is greater than 4.3, and FALSE is it is lower or equal to 4.3.
b) Display the content of the new variable.
c) Use the `table` function to show the number of TRUE's and the number of FALSE's.

```
# put your code here
```

Question 4

a) Use the logical variable your created in question 3 to create two numerical variables: HRate which contains all the Homicide.rate values that are greater than 4.3, and LRate which contains all remaining Homicide.rate values.
b) Produce a side-by-side boxplot comparing HRate and LRate Use at least two more options in the `boxplot` function to improve the plot and make it more appealing.

```
# put your code here
```

Question 5

a) Find the 2 states which have the highest numbers of gun deaths (per 100,000 people) and provide their Brady score and their number of gun deaths.
b) Produce a scatterplot of variables **Sum** versus **Brady.score**. Make the title as **Number of gun death per 100,000 people by Brady Score**, x axis name as **Brady Score**, y axis name as **Gun Deaths** and change the point symbol to solid rhombus. Click the link to find more plot options.
c) The text variable **Jurisdiction** contains the name of each state. Use the function `text` to label each point with corresponding names. (hint: link)
d) Color the two states which have the highest number of gun death per 100,000 people with red (or choose your preferred colors). The other states remain black.
e) Find all the states/jurisdictions which have weak restrictions for buying guns (use Brady score less than 0), but also has a low number of gun deaths per 100,000 people (use less than 4). Use the functions `which` and `intersect`. Display it as a table with the state/jurisdiction names, the Brady score, and the number of gun deaths per 100,000 people.
f) Summarize your findings from the scatterplot.

```
# put your code here
```