Name: Steve G. Mwangi
Date: October 13, 2019
Assignment: TMATH 390 R Lab 2 Document.

C1. (4) Submit your R scripts to Canvas. You can upload them directly to your assignment as *.R documents.
Submitted

C2.
Choose 1 quantitative variable: Net Worth
Choose 1 qualitative variable: Gender
Labels Gender can take: Male and Female

```
> #C2.
> #Determine working directory with
> getwd()
[1] "C:/Users/steve/Desktop/UWT/Fall Classes/TMATH 390/R Documents/R Assignments/R_Lab_2"
> #Change working directory to: C:\Users\steve\Desktop\UWT\Fall Classes\TMATH 390\R Documents\R Assignments\
  #R_Lab_2
> setwd("C:/Users/steve/Desktop/UWT/Fall Classes/TMATH 390/R Documents/R Assignments/R_Lab_2")
> #Read csv file of my data
> data.df = read.csv("data.csv")
> head(data.df)
  Rank          Name Net.worth YTD.Change    Industry Gender             Continent
1    1     Jeff Bezos     108.6     -16.20  Technology   Male North and South America
2    2     Bill Gates     104.9      15.60  Technology   Male North and South America
3    3 Bernard Arnault      92.2      21.60    Consumer   Male                Europe
4    4  Warren Buffett      82.2      -1.64 Diversified   Male North and South America
5    5  Amancio Ortega      68.3       7.81    Consumer   Male                Europe
6    6   Larry Ellison      67.1      10.10  Technology   Male North and South America

>
```

C3.

```
> #Quantitative variable
> #C3.
> #Quantitative variable chosen: Net Worth
> #Producing a summary of Quantitative Net worth column
> summary(data.df$Net.worth)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  19.20   23.10   34.10   40.72   52.58  108.60
> #Producing a summary of Qualitative Gender column
> summary(data.df$Gender)
Female   Male
     8     42
```

C4. Describe what the summary command returns for a quantitative variable.

```
> #Producing a summary of Quantitative Net worth column
> summary(data.df$Net.worth)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  19.20   23.10   34.10   40.72   52.58  108.60
```

The command finds in the net-worth column:

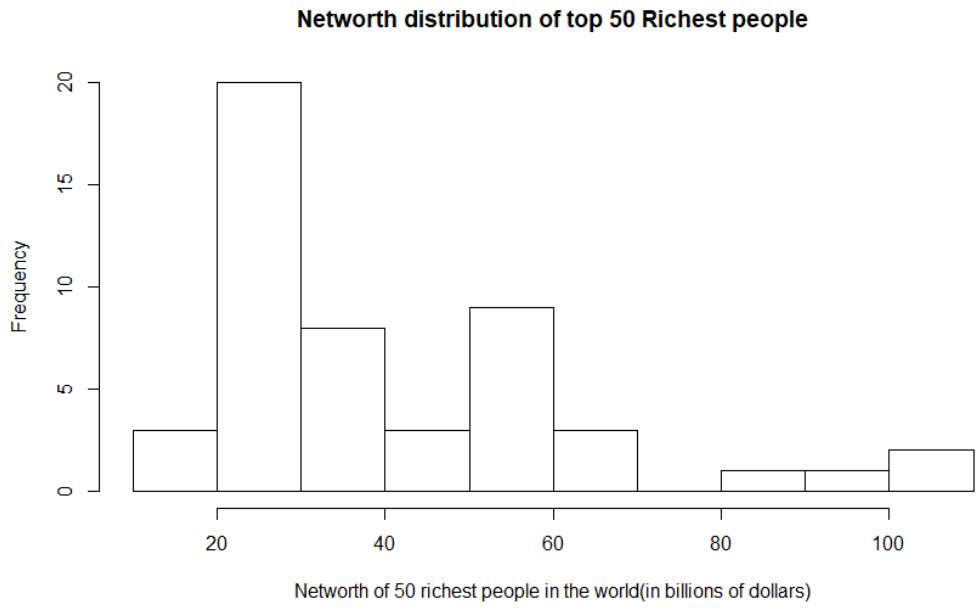
1. minimum value
2. 1st quartile value
3. Median value
4. Mean of all the values
5. 3rd quartile value
6. Maximum value

Summary for the qualitative variable Gender, brings up the frequency of the two values in the column:

```
> #Producing a summary of Qualitative Gender column
> summary(data.df$Gender)
Female   Male
     8     42
```

C5. Histograms to visualize net worth column

```
#C5.
> #Histogram of Net Worth Column
> #Command creates a histogram of chosen column, in the dataframe data.df
> #the xlab argument writes test to label the x-axis
> # main argument gives it title
> hist(data.df$Net.worth, xlab="Networth of 50 richest people in the world(in billions of dollars)",
    main="Networth distribution of top 50 Richest people")
```

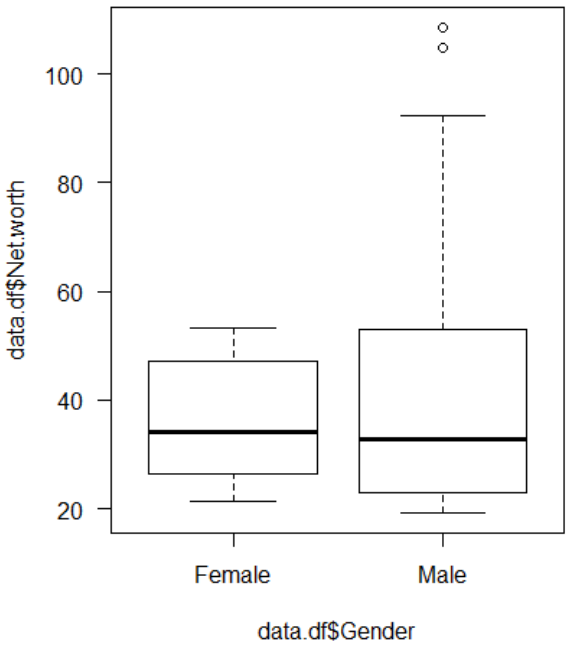


C6.

Right skewed. With many people with net-worth's between 20-60 billion dollars. Few ultra-wealthy with net-worth's above 80 billion dollars. Also, there are no billionaires with net-worth's in the 70-80 billion interval. This might be due to the in accuracies inherent in tracking real time changes in wealth of people, and the fact that most of the net-worth are estimates that do not account for offshore properties, or other properties that have not been publicly disclosed.

C7.

```
> #C7. Boxplots
> # Establishing a graphing window with 1 rows and 2 columns,
> # and las = 1 sets axis labels to be horizontal
> par(mfrow=c(1,2), las = 1)
> #Creating a boxplot.
> # ~ represents a relationship between two variable, with Y on left side, X on right side.
> # Networth(quantitative) across Gender(Qualitative)
> boxplot(data.df$Net.worth~data.df$Gender)
```



C8.

Similarities: Their median net-worth is roughly the same, at about $33 billion dollars

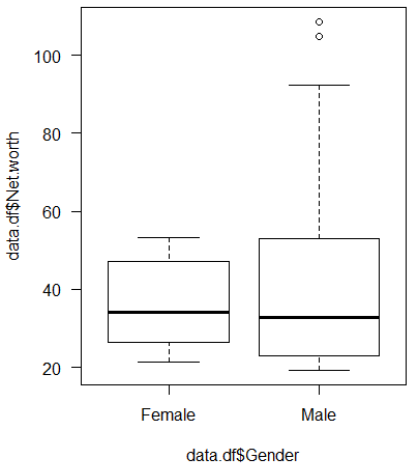
Differences: Larger boxplot for the males, while smaller boxplot for the females. This could be due to the higher number of males (42 in number) in proportion to females (8 in number) as seen from C3.

We see that also the male boxplot has two outliers. This might be accounted for by the fact that there are two ultra-wealthy individuals in the data with net-worth of over 100 billion dollars.

C9.

```
#C8.
> #Individual summary statistics
> #Command to get individual summary statistics
> #FEMALE
> #For mean
> mean(data.df$Net.worth[data.df$Gender=="Female"])
[1] 36.275
> #For median
> median(data.df$Net.worth[data.df$Gender=="Female"])
[1] 34.1
> #Standard Dev.
> sd(data.df$Net.worth[data.df$Gender=="Female"])
[1] 12.15904
> #MALE
> #For mean
> mean(data.df$Net.worth[data.df$Gender=="Male"])
[1] 41.56905
> #For median
> median(data.df$Net.worth[data.df$Gender=="Male"])
[1] 32.7
> #Standard Dev.
> sd(data.df$Net.worth[data.df$Gender=="Male"])
[1] 23.5445
```

C10.



From the numerical summaries, we see that the Male median = 32.7 and Female median = 34.1. This accounts to why their boxplot have a roughly same median line. Furthermore, the standard deviation for Male, 23.5445 is greater than that for Female, 12.15904. This explains why the 1st quartile and 3rd quartile for Male boxplot are widely spaced, compared to that of Female boxplot; the larger deviation from the mean in the Male data compared to that of the female data accounts for the larger boxplot. Also, the histogram for both plots was right skewed, and we see that the boxplot show this since they are larger above the median.