Problem Set #1

Name(s): _____

Q1

Use the "cars" dataset (openintro) to answer the following questions:

a) Create a histogram and boxplot (box and whiskers) for the 'mpgCity' variable (attach)

b) Determine the skewness of the 'mpgCity' distribution (actual skewness value)
   - You will need to do a library statement before using the "skewness" function - library(moments)

c) Calculate the mean and the median for the 'mpgCity' variable.

d) Does each of the previous tasks remain consistent in terms of skewness? Does the boxplot and histogram demonstrate the direction of skewness that the skewness measurement provides. Does the relationship between the mean and median support such skewness? Elaborate...

Q2

Use the "cars" dataset (openintro) to answer the following questions:

a) Create a scatterplot for the 'weight' (x) and 'mpgCity' (y) variable. Use the "plot" function (attach)

b) Determine the covariance and correlation measures between variables of interest in part (a)

c) Explain the connection between the scatterplot and metrics calculated in part (b)

d) Does such findings in the previous tasks make "intuitive" sense? Elaborate...

Q3

Construct a box and whiskers plot with the following data (x):

```
x <- c(1, 3, 4, 5, 7, 4, 5, 4, 2, 9, 10, 4, 3, 5, 11, 1, 3, 3, 5, 4)
```

You must provide the R code or mathematical approach for each of the following questions (a-f) in your response. Just do not provide the answer!

a) What is the median?

b) What is Q3 (75% Percentile)?

c) What is Q1 (25% Percentile)?

d) What is the IQR of the data set?
e) What is the lower fence?

f) What is the upper fence?

g) How many "extreme" values are present in the data set?

h) Generate the box and whiskers plot using the "boxplot" R function (attach). Ensure the boxplot created is consistentt with the results derived in the previous tasks.

Q4

Use the "cars" dataset (openintro) to answer the following questions:

You must provide the R code for each of the following questions (a-c) in your response.

a) How many 'midsize' (count) cars are in the data set "cars"? Use the "table" function with the variable of interest, 'type'

b) Create a barplot to confirm your answer in part (a) and attach

c) Build a contingency table with the variables 'type' and 'driveTrain'. How many front wheel drive small cars exist in the data set?

Q5

Use the "beaver1" dataset (datasets) to answer the following questions:

Run the following code to generate two samples of 'temp' from the "beaver1" dataset. Each will be assigned a new name, 'temp1' and 'temp2'. Each sample will consist of 15 observations.

```
set.seed(1986)
temp1 <- sample(beaver1$temp, 15)
temp2 <- sample(beaver1$temp, 15)
```

You must provide the R code for each of the following questions (a-b) in your response.

a) Calculate the sample variance, sample stadard deviation, sample IQR for each sample

b) Since both samples were taken from the same population, compare your findings for each measure of spread between samples. Elaborate...