AMERICAN UNIVERSITY STAT 614 HOMEWORK 4

INSTRUCTIONS: Submit your R code, output, comments, and narratives using two files, (Rmarkdown file and a Word file) by way of the knitting process.

**library(tidyverse) library(ggplot2)**

Explore and analyze the **midwest** data frame

Highlight and run **midwest** to call the data frame

1) How many rows and columns does the **midwest** data frame have?

2) Name three variables of the **midwest** data frame that are categorical.

3) What is the Native American population of Bond county Illinois ?

4) What does the variable **percollege** stand for or represent?

5) Use and show R code to produce a histogram for the variable **percollege**

6) How would you characterize the distribution of the data modeled by your histogram; skewed left, skewed right, or normal?

7) What is the relationship between the mean and the median?

8) Use and show R code to produce a boxplot for the variable **percollege**.

9) Does the boxplot indicate that the data has outliers?

10) Use the boxplot or the summary function to find the five number summary (Minimnm, 1st Quartile, Median, 3rd Quartile, Maximum)

11) Use and show R code to find the standard deviation of the data.

12) Use and show R code to find the Mean of the data,

13) Use and show R code to find the IQR.

14) Use and show R code to produce a stem and leaf plot.

15) In three of four sentences explain the relationship that you see between the stem and leaf plot and the histogram.

16) Use and show Tidyverse R code to produce a scatter plot for the variables **percollege** and **poppovertyknown**. Use **percollege** as the independent variable and **poppovertyknown** as the dependent variable.

17) In four of five sentences, explain what relationship exists between the variables as indicated by the scatter plot. Is there an extreme outlier in the scatter plot?