Lab 2 (100 points + 10 points BONUS) - Describing Distributions with Graphs and Numbers
Objectives: Numerical Summaries, Histograms, and Boxplots

Remember to use the cleaned data set that you generated in Lab 1.

A. (10 points) Online Prelab

B (45 pts) Average Test Scores (Data Set: USData cleaned) We are interested in the graphical and numeric summaries for the adjusted average test score for a college admission exam (TestScore).

1. (10 points) Code. The code is the script (commands) either in R or in the Editor in SAS. Remember that you need to include the code or procedure for inputting the data into your software package.
2. (2 points) Find the five-number summary for these data.
3. (5 points) Calculate the 1.5 IQR upper and lower limits for the outliers. Are there any outliers according to the 1.5 IQR rule (just answer yes or no, and explain why you know)? This part may be done by hand. If done by hand, all work needs to be provided. If done via computer code, then the code must be listed.
4. (5 points) Make a modified boxplot. Describe the distribution by stating whether it is symmetrical, left, or right skewed, and if there are outliers. Please indicate which features of the plot you used to classify the type of skewness and determine whether outliers are present.
5. (5 points) Make a histogram of the data. Describe the distribution by stating whether it is symmetrical, left, or right skewed, and if there are outliers. Please indicate which features of the plot you used to classify the type of skewness and determine whether outliers are present.
6. (5 points) Are the data points that you considered outliers in the histogram and the boxplot the same or different? If they are different, please provide a possible explanation for the difference.
7. (5 points) Obtain the sample mean, $\bar{x}$, and the sample standard deviation, s. Indicate the location of the mean and median on the histogram in Question 5. This may be done by hand or computer software. $\bar{x}$ is the average of the average values and the standard deviation of the average values for the counties in the sample.
8. (3 points) Do you think the median is close to the mean? Please explain your rationale.
9. (5 points) If you only had one measure to describe the central location of the distribution of TestScore, which would you choose? Please explain your answer.

C (45 pts). Larcenies (Data Set: USData Cleaned). We are interested in the graphical and numeric summaries of number of larcenies (thefts of personal property) out of 100,000 people (LarceniesPerPopulation).

1. (10 points) Code
2. (2 points) Find the five-number summary.
3. (5 points) Calculate the 1.5 IQR upper and lower limits for the outliers. Are there any outliers according to the 1.5 IQR rule (just answer yes or no, and explain why you know)? This part may be done by hand. If done by hand, all work needs to be provided. If done via computer code, then the code must be listed.

4. (5 points) Make a modified boxplot. Describe the distribution by stating whether it is symmetrical, left, or right skewed, and if there are outliers. Please indicate which features of the plot you used to classify the type of skewness and determine whether outliers are present.
5. (5 points) Make a histogram of the data. Describe the distribution by stating whether it is symmetrical, left, or right skewed, and if there are outliers. Please indicate which features of the plot you used to classify the type of skewness and determine whether outliers are present.
6. (5 points) Are the data points that you considered outliers in the histogram and the boxplot the same or different? If they are different, please provide a possible explanation for the difference.
7. (5 points) Obtain the sample mean, $\bar{x}$, and the sample standard deviation, s. Indicate the location of the mean and median on the histogram generated in Question 5. This may be done by hand or computer software.
8. (3 points) Do you think the median is close to the mean? Please explain your rationale.
9. (5 points) If you only had one measure to describe the central location of the distribution of LarceniesPerPopulation, which would you choose? Please explain your answer.

D. (10 points) BONUS. We do not discuss how to make graphs of categorical variables in this class; however, this is very important. Make a pie chart for Region. Note: You will not get coding help for bonus questions.

1. (5 points) Code
2. (5 points) If the labels are abbreviations which are not clear, state what each label represents. What information did you learn from the pie chart? Please explain your answer.