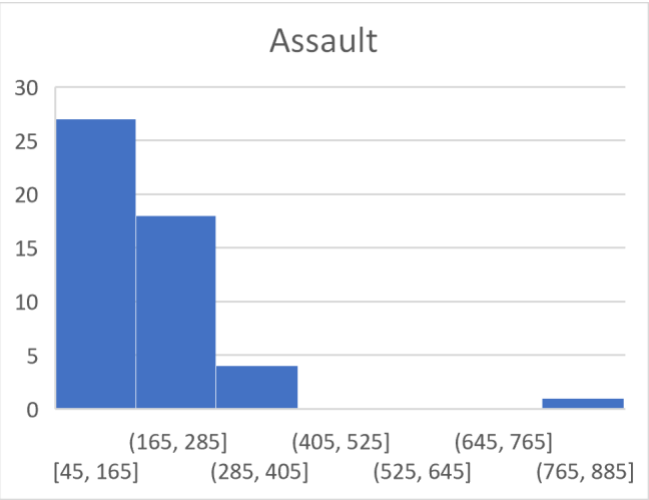
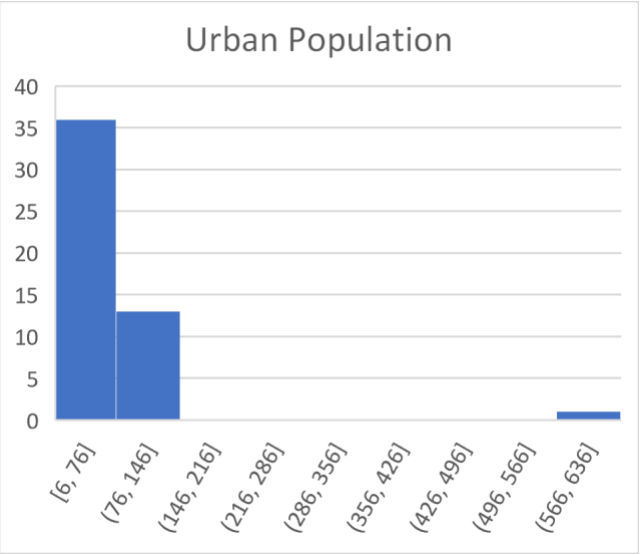
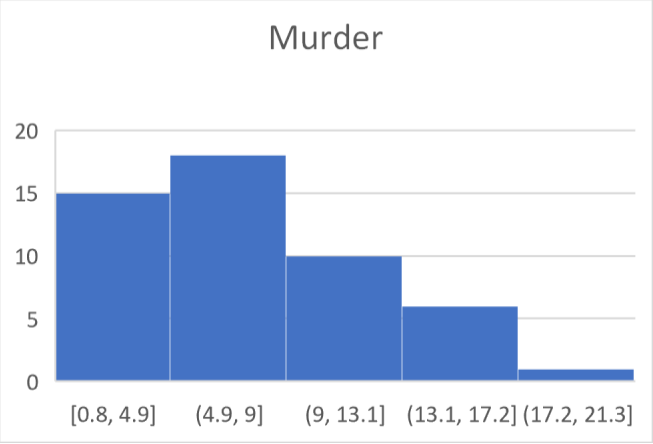


**InClassAssignment1(Group of two)**  
**CS160-02**  
**Introduction to Data Science**  
**Spring 2023**

**Working on Techniques for Analyzing Data**

**Instructions:** Complete the following activities for this project.

1. Create a new GitHub repository named Assignment1\_XXX, where XXX are your initials.
2. Using excel (to generate the result) and word documents (type answers and paste the results) work on the following questions and submit your work using **pdf** format.
  - a. What are the differences between data analysis and data analytics?  
Data analysis- Hands on data exploration  
Data analytics- broader term that includes data analysis as a broader term
  - b. Comment on variable types of Murder, Assault, and urban pop.  
All variables are independent. State column is categorical. Ratio data. States are categorical nominal.
  - c. What is the difference between interval and ratio data?  
Interval- numbers have equal spacing with an absolute zero.  
Ratio- numbers rank on a scale with an absolute zero
  - d. What is descriptive analysis? Represent the data of Murder, Assault, and urban pop. Comment on the distribution. It is the first type of analysis used on large scales of data. Each category of data is right skewed.



- e. What is a measure of dispersion? They are values that represent the measure of the spread of data.

Calculate the interquartile range of those three variables

Murder iqr- 7.715

Assualt iqr- 140

Urban pop iqr- 24.5

There are outliers on the assault and urban population box plots.

For murder it is right skewed because the median is 7.25 and the mean is 7.788. The mode is 13.2.

For assault it is right skewed because the mean is 109, while the median is 249. The mode is 120.

For urban pop it is right skewed because the mean is 74.2, and the median is 159. The mode is 80.

Although the mode does not necessarily fit, it is still right skewed because of the mean and median.

- f. What is the measure of centrality? Find the measurement of centrality: mean, median, mode

These are numbers that describe the center of distribution.

For murder it is right skewed because the median is 7.25 and the mean is 7.788. The mode is 13.2.

For assault it is right skewed because the mean is 109, while the median is 249. The mode is 120.

For urban pop it is right skewed because the mean is 74.2, and the median is 159. The mode is 80.

- g. What are diagnostic analytics? Find diagnostic analysis for pair of variables.

Diagnostic analytics are used to tell why something happened.

The correlation for murder and assault is 0.649313.

The correlation for murder and urban pop is -0.18617.  
The correlation for urban pop and assault is -0.14066.

3. Using the instructions provided by GitHub, create a git repository named **DS160InClassAssignment**, and push your pdf file to it. Each of you needs to submit your work.

**Submission:**

Paste a link to your GitHub repository in the area provided for this assignment and submit it by class time.