**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 is more of a physical form of data exploration and evaluation. Data analysis includes preprocessing as well as transforming and working with the data to determine a hypothesis.**

    **Data analytics includes data analysis. Defines the concept and practice of all activities related to data**

    b. Comment on variable types of Murder, Assault, and urban pop.

    **All variable types are continuous, independent and ratio. State variable is categorical and nominal**

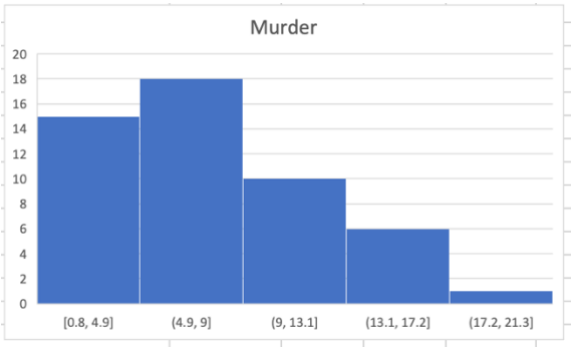
    c. What is the difference between interval and ratio data?

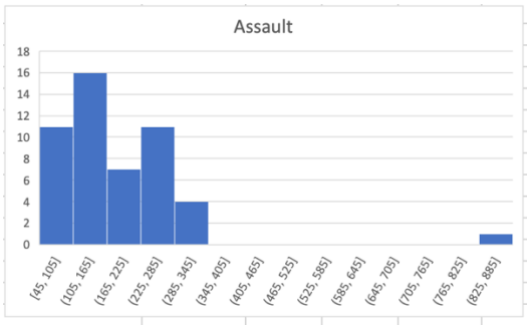
    **Interval data have no absolute zero and have equal spacing between values. (eg. -5, 0, 5, 10, 15)**

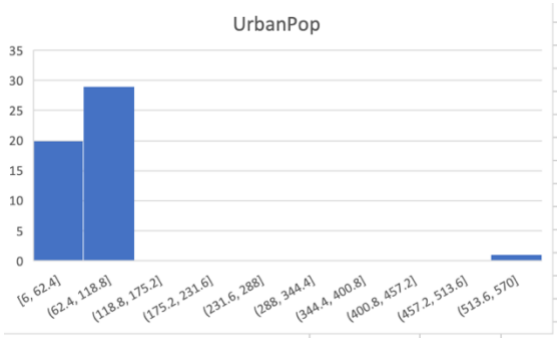
    **Ratio data have an absolute zero and contain units of equal value/magnitude. eg(0, 1, 5, 8)**

    d. What is descriptive analysis? Represent the data of Murder, Assault, and urban pop. Comment on the distribution.

**Descriptive analysis is the kind of data analysis performed on a data set that is applied to larger volumes of data. This can be represented in a histogram.**




**Murder Distribution – right skewed, most of the values are in the first 2 bins between .8 and 9**

**Assault Distribution – right skewed, outlier in the last bin between 825 and 885. Most of the values are between 105 and 165**

**UrbanPop Distribution – right skewed, outlier in the last bin between 514 and 570. Most of the values are between 63 and 119**

e. What is a measure of dispersion? Calculate the interquartile range of those three variables

**The measure of dispersion is the spread of the data. This can be measured by range, interquartile range, standard deviation, or variance.**

**Interquartile range for murder – 7.2**

**Interquartile range for Assault – 140**

**Interquartile range for UrbanPop – 24.5**

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

**Measurement of centrality describes the center of the distribution**

**Murder mean – 7.8**

**Mode – 13.2**

**Median – 7.25**

**Assault mean – 182.2**

**Mode – 120**

**Median – 159**

**UrbanPop mean – 74.2**

**Mode – 80**

**Median – 66**

**Assault is right skewed because mean is greater than mode. Murder is right skewed because mean is greater than median. Urbanpop is right skewed because mean is also greater than median**

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

**Diagnostic analysis is the correlation between variables.**

**Correlation between Murder and Assault - .65 or 65%**

3. Using the instructions provided by GitHub, create a git repository named DS160**InClassAssignment**, 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.