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

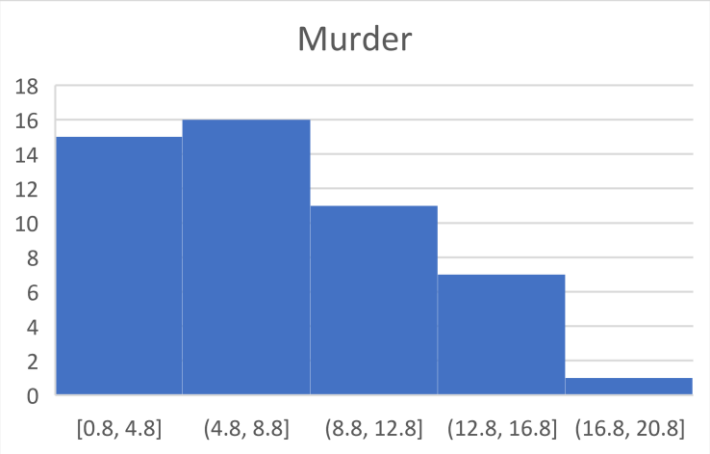
Working on Techniques for Analyzing Data

Instructions: Complete the following activities for this project.

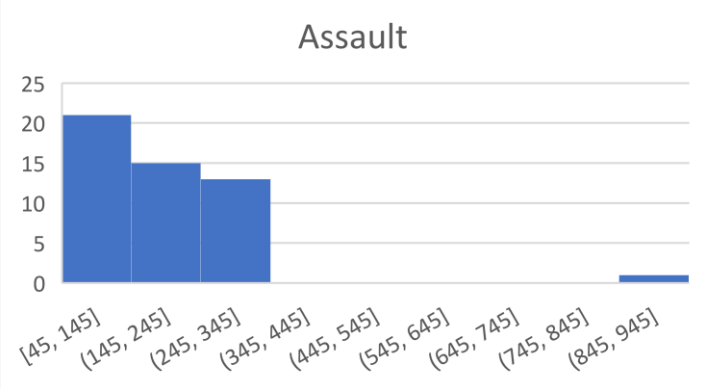
1. Create a new GitHub repository named Assignment1_XXX, where XXX are your initials.
2. Using excel (to generate the result) and word documents (type answers and paste the results) work on the following questions and submit your work using **pdf** format.

    a. What are the differences between data analysis and data analytics?
        a. Data analysis looks at data that happened to explain while data analytics uses data to model the future or predict an outcome.
    b. Comment on variable types of Murder, Assault, and urban pop.
        a. All are independent.
        b. The states are categorical and nominal.
        c. All are ratio.
    c. What is the difference between interval and ratio data?
        a. Ratio data is on a scale with an absolute 0, Interval data does not have an absolute 0 but all numbers have equal spacing between adjacent values.
    d. What is descriptive analysis? Represent the data of Murder, Assault, and urban pop. Comment on the distribution.
        a. It is the first kind of data analysis performed on a data set. Techniques used are Frequency distribution, measures of centrality, and dispersion of distribution.

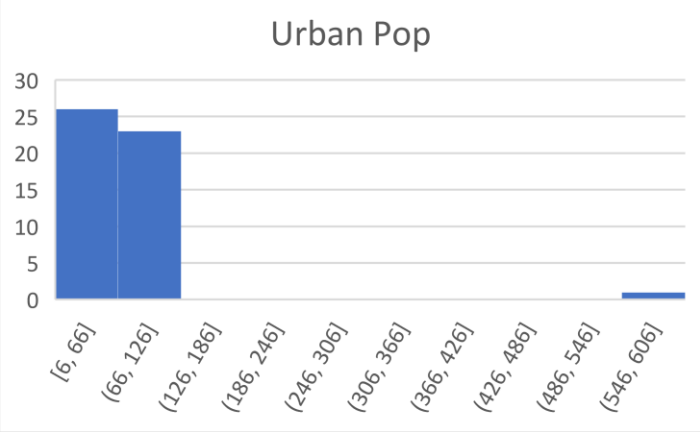
b. Murder distribution is skewed right. Mode > Median > Mean.



c. Assault distribution is skewed right with high outlier values. Mode > Median > Mean.



d. Urban pop is skewed right with a high value outlier. Mode > Median > Mean.



e. What is a measure of dispersion? Calculate the interquartile range of those three variables
    a. Murder = Q3(11.25) – Q1(4.075)= **IQR (7.175)**
    b. Assault = Q3(249) – Q1(109)= **IQR (140)**
    c. Urban pop = Q3(77.75) – Q(53.25)= **IQR (24.5)**
f. What is the measure of centrality? Find the measurement of centrality: mean, median, mode
    a. Murder: Mean- 7.788, Median- 7.25, Mode- 13.2
    b. Assault: Mean- 182.18, Median- 159, Mode-120
    c. Urban pop: Mean- 74.2, Median- 66, Mode- 80
g. What are diagnostic analytics? Find diagnostic analysis for pair of variables.
    a. These determine why something happened. The most used technique is correlation.
        i. Murder vs Assault correlation: 0.649313315
        ii. Murder vs Urban pop correlation: -0.186169557
        iii. Assault vs Urban pop correlation: -0.140657654

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.