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 hands-on data exploration and evaluation, it typically looks at what happened in order to explain.

    Data analytics is a broader term that includes data analysis as a necessary part, it models the future or predicts a result.

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

    State: Categorical, nominal, ratio

    Murder: Independent, continuous, ratio

    Assault: Independent, continuous, ratio

    Urban Pop.: Independent, continuous, ratio

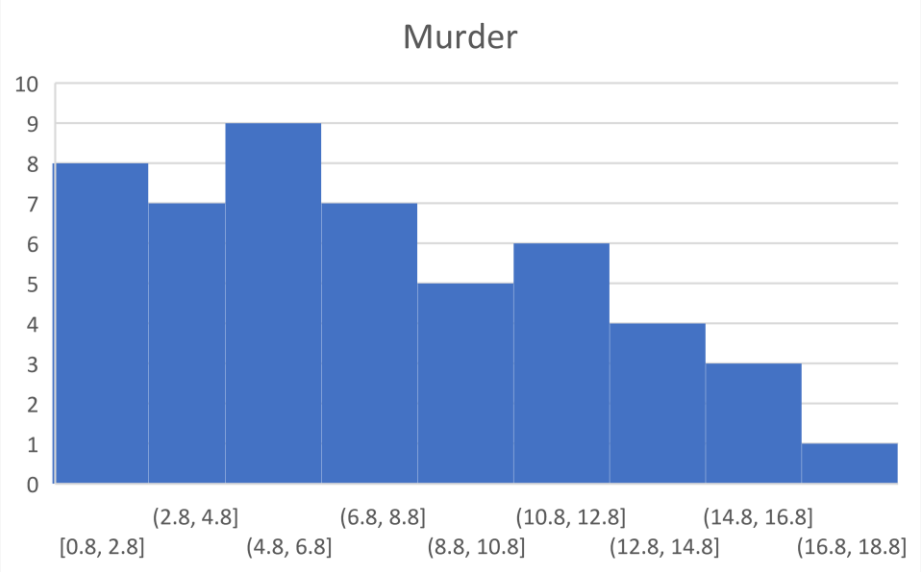
    c. What is the difference between interval and ratio data?

    Interval data is when numbers have equal spacing between adjacent values without an absolute zero. They hold no true zero and can represent values below zero.

    Ratio data has an absolute zero and the numbers have units of equal magnitude and rank order.

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

Descriptive analysis is the first thing you do when looking at data. You have to understand the data you are looking at and it is applied to large volumes of data.



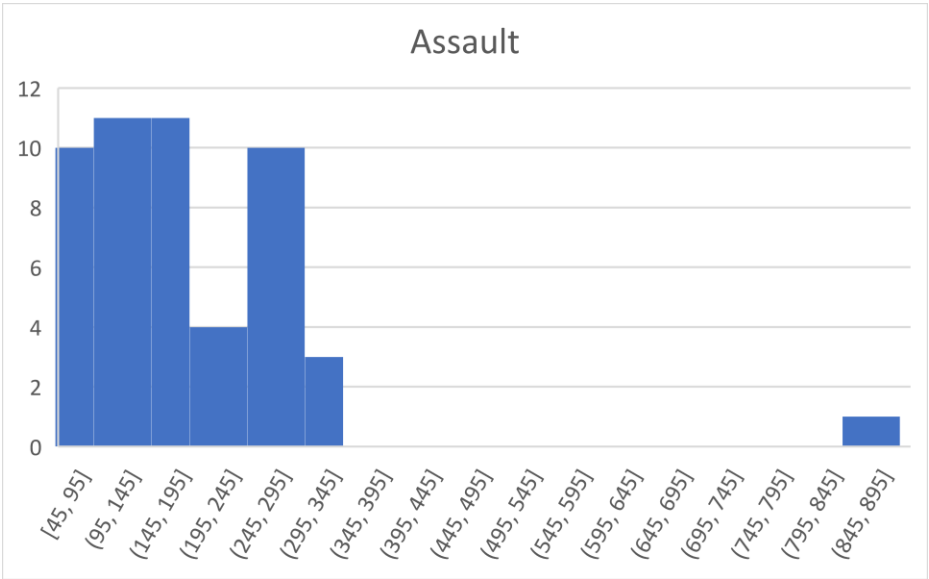
Murder:

Mean: 7.788

Median: 7.25

Mode: 13.2

This data is right skewed, because the mean is greater than the median, with no outliers



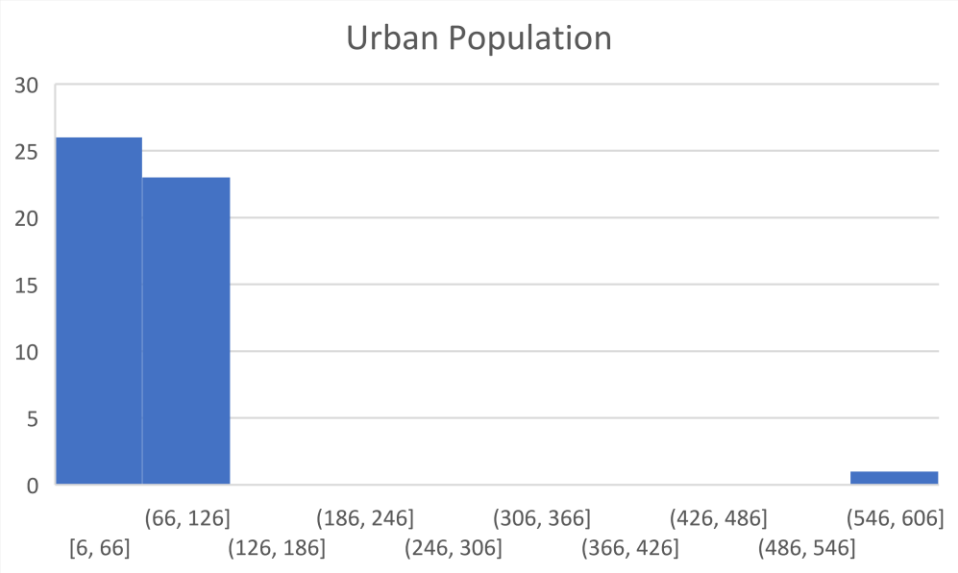
Assault:

Mean: 182.1836

Median: 159

Mode: 120

This data is right skewed, because the mean is greater than the median, which is greater than the mode, with an outlier between 845 and 895



Urban Population:

Mean: 74.2

Median: 66

Mode: 80

This data is uniform with an outlier between 546 and 606

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

A measure of dispersion is a value that represents the spread of the data. For example, the range, IQR, variance, and standard deviation.

Murder:

Range: 16.6

IQR: 7.175

Variance: 18.59

Standard Deviation: 4.31

Assault:

Range: 834

IQR: 140

Variance: 16443.75

Standard Deviation: 128.23

Urban Population:

Range: 564

IQR: 24.5

Variance: 5281

Standard Deviation: 72.67

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

A measure of centrality is a number that describes the center of a distribution. For example, the mean, median, and mode.

Murder:

Mean: 7.788

Median: 7.25

Mode: 13.2

Assault:

Mean: 182.1836

Median: 159

Mode: 120

Urban Population:

Mean: 74.2

Median: 66

Mode: 80

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

Diagnostic analytics are used to determine why something happened; the most used technique is correlation.

The correlation between Murder and Assault is 0.6494

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.