

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 explores and evaluates hands-on data. Analysis will take in what happened and explain. Data analytics is broader, and it includes data analysis as a particular part of it. Analytics attempt to model and predict the future.

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

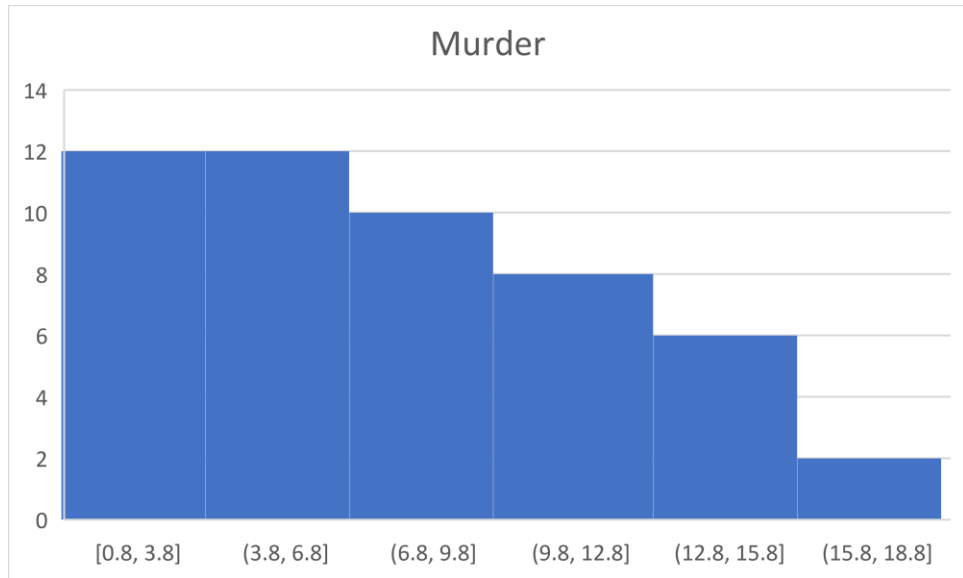
They are all independent, continuous, ratio variables. However, the state column is a categorical, nominal variable type.

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

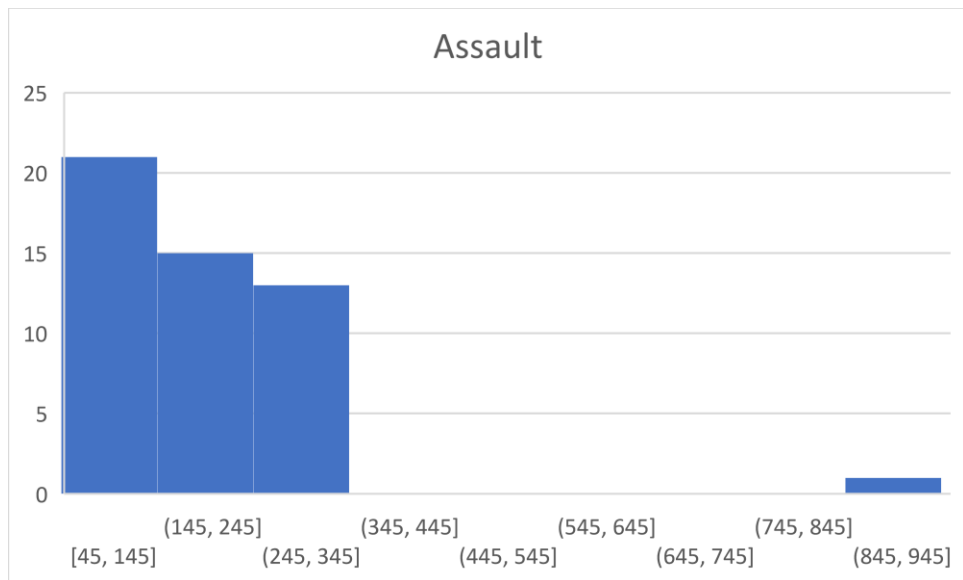
Interval data can be any number, and it does not have a minimum or maximum. Ration data has a minimum and cannot go below that number.

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

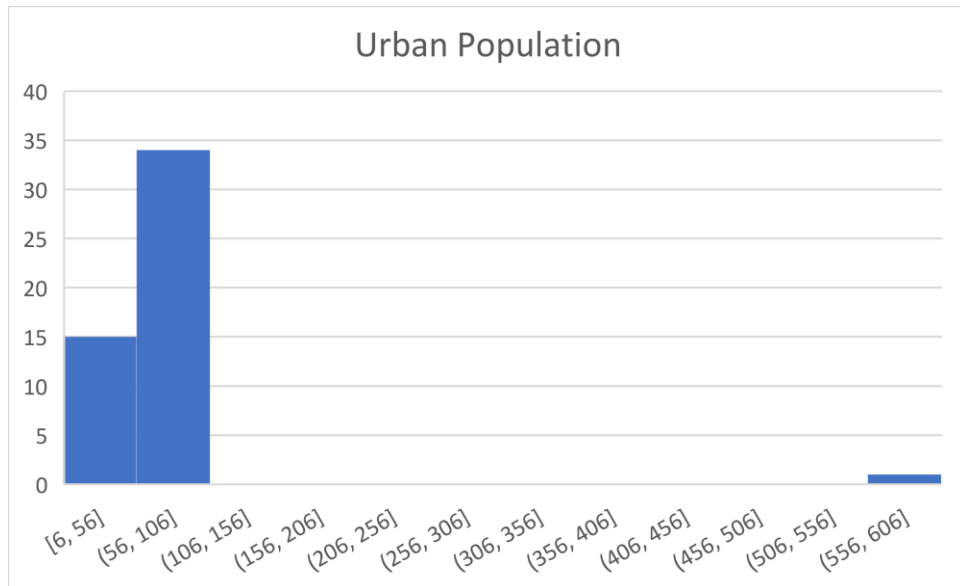
Descriptive analysis refers to the examination of a representation of data.



Right-skewed with the most output between 0.8 and 6.8



Right-skewed with the most output between 45 and 145



Right skewed with most of the output between 56 and 106

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

The measure of dispersion is a measure of the spread of the data.

Murder IQR: 7.2, Assault IQR: 140, Urban pop IQR: 24.5

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

The measure of centrality is an attempt at finding the center of the distribution.

Murder – mean: 7.8, median: 7.3, mode: 13.2

Assault- mean: 182.2, median: 159, mode: 120

Urban pop- mean: 74.2, median: 66, mode: 80

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

Diagnostic analysis is the process of finding out why something happened, usually done with correlation.

The correlation between murder and assault is 65 percent

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.