

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 while data analytics is a broader term that includes data analysis as a part of it.

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

Murder= independent/continuous/ratio

Assault=independent/continuous/ratio

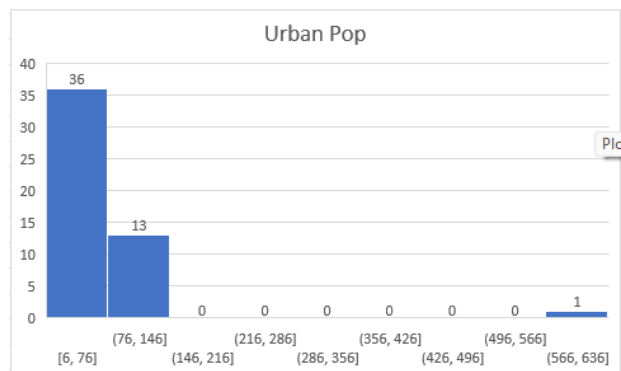
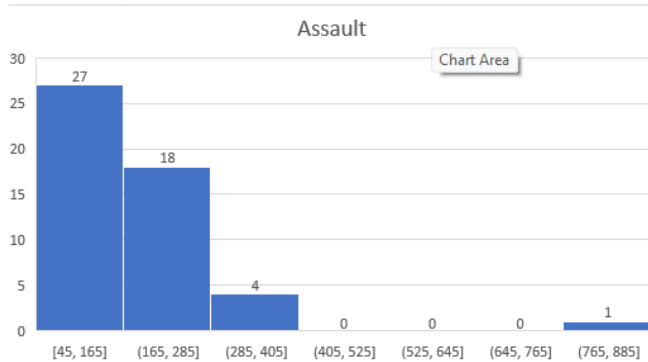
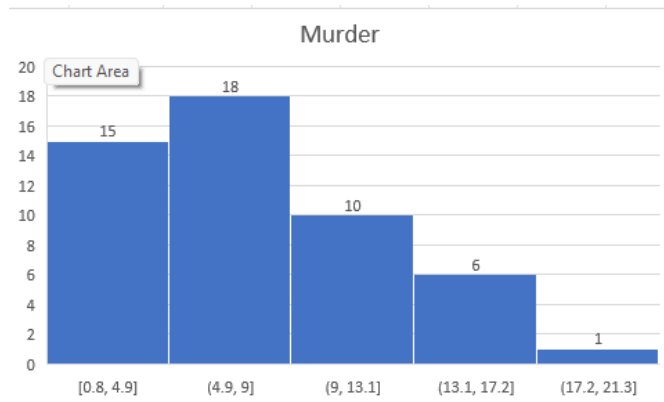
Urban Pop=independent/continuous/ratio

- 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, while ratio data is when numbers have units of equal significance and magnitude and rank order 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.

All three of the histograms for Murder, assault, and urban pop are all skewed to the right. While the murder data set is closest to a normal distribution, it is still right skewed, and both the assault and urban pop data sets have outliers that skew the data sets to the right.



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

A measure of dispersion are the values that represent the spread of the data. The IQR for the variables are as follows:

Murder= 7.175

Assault=140

Urban Pop=24.5

	Q1	Q3	IQR	Mean	Median	Mode
Murder	4.075	11.25	7.175	7.788	7.25	13.2
Assault	109	249	140	182.1836	159	120
Urban Pop	53.25	77.75	24.5	74.2	66	80

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

The measure of centrality are a group of numbers that describes the center of a data set.

Murder-> Mean=7.788, Median=7.25, Mode=13.2

Assault->Mean=182.1836, Median=159, Mode=120

Urban Pop-> Mean=74.2, Median=66, Mode=80

	Q1	Q3	IQR	Mean	Median	Mode
Murder	4.075	11.25	7.175	7.788	7.25	13.2
Assault	109	249	140	182.1836	159	120
Urban Pop	53.25	77.75	24.5	74.2	66	80

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

Diagnostic analytics, such as the correlation technique, are used to determine why something happened.

The correlation between Murder and Assault has value of 0.6494. This positive value means that there is a positive correlation between the two, although the relationship is not incredibly strong as it isn't very close to 1.0.

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.