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

<div align="center">

Working on Techniques for Analyzing Data

</div>

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?

   The difference between data analysis and data analytics is analysis is a subset of data analytics that includes specific processes, and data analytics is the broad field of using data and tools to make business decisions.

   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?

   The difference between interval and ratio data is interval data has equal spacing between adjacent values and doesn't have an absolute zero, this data can also represent values below zero. Ratio data are numbers that have equal 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.

Descriptive analysis is the process of using current and historical data to identify trends and relationships. It's sometimes called the simplest form of data analysis because it describes trends and relationships but doesn't dig deeper.

Murder is a right skewed data set, the mean is greater then the median.

Assault is a right skewed data set as well with an outlier in the distribution, the mean is greater then the mode as well as it's greater then the median. Median is also greater then the mode which all make it right skewed.
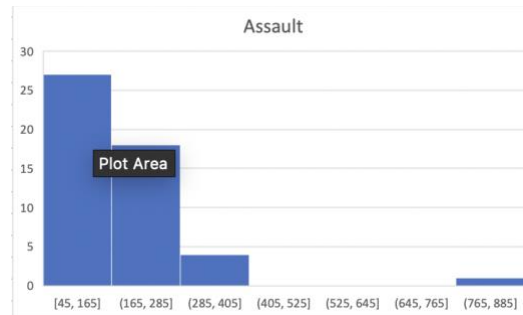
Urban Pop is a right skewed data distribution too with an outlier. The mean is greater than the median in this data set.
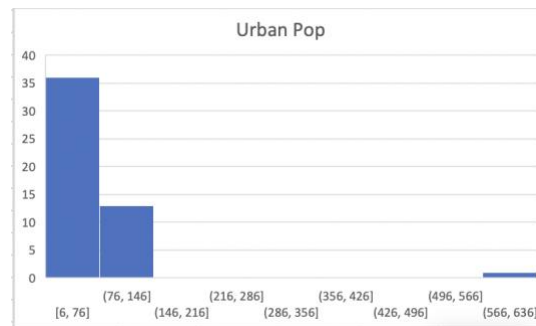
e. What is a measure of dispersion? Calculate the interquartile range of those three variables

Measure of dispersion is a measure of spread of data about the mean.

Murder

IQR = 7.175

Variance = 18.97

Standard Deviation = 4.35

Assault

IQR = 140

Variance = 16779.3

Standard Deviation = 129.5

Urban Pop

IQR = 24.5

Variance = 5388.7

Standard Deviation = 73.4

	Murder	Assault	UrbanPop
Quartile 1	4.075	109	53.25
Quratile 3	11.25	249	77.75
IQR	7.175	140	24.5
Variance	18.97047	16779.33	5388.776
Standard Deviation	4.35551	129.5351	73.40828

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

Measure of centrality is the number that describes the center of distribution.

Murder mean: 7.78

Murder median: 7.25

Murder mode: 13.2

Assault mean: 182.18

Assault median: 159

Assault mode: 120

UrbanPop mean: 74.2

UrbanPop median: 66

UrbanPop mode: 80

	Murder	Assault	UrbanPop
Mean	7.788	182.1836	74.2
Mode	13.2	120	80
Median	7.25	159	66

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

Diagnostic analytics is the process of using data to determine the causes of trends and correlations between variables.

The correlation between murder and assault is 0.649. Because this number is positive and closer to 1 than to 0 and -1, I'd say that the correlation between the two variable is positive.

The correlation between assault and urban pop is -0.1406. Because this number is close to 0 it means there isn't a meaningful correlation between the data.

The correlation between Murder and urban pop is -0.186. Because this number is close to 0 it means that there isn't a meaningful correlation between the data.

Correlation	
0.649375912	Murder & Assult
-0.14066314	Assult & Urban Pop
-0.18616956	Murder & Urban Pop

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.