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 data exploration and evaluation while data analytics includes data analysis to make decisions as a necessary part.

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

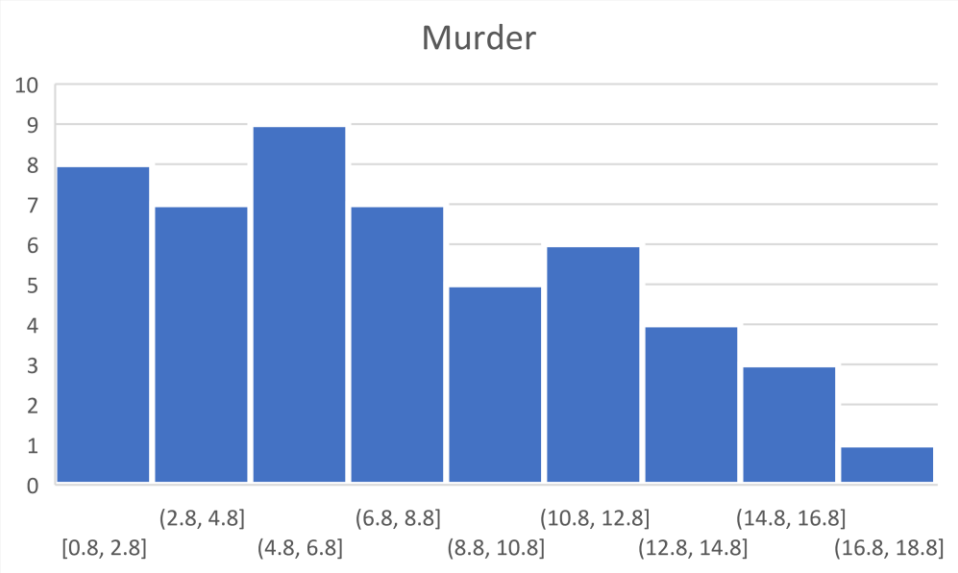
Murder – Independent, continuous, and ratio
Assault – Independent, continuous, and ratio
Urban pop. – Independent, continuous, and ratio
State – Independent, categorical, and ratio

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

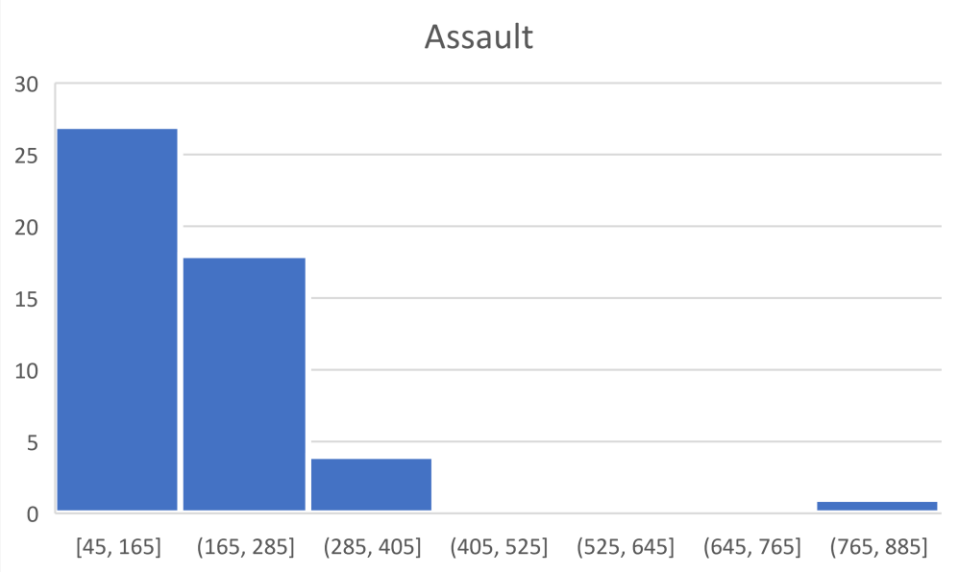
Interval data does not have an absolute zero while ratio data can have an absolute zero.

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

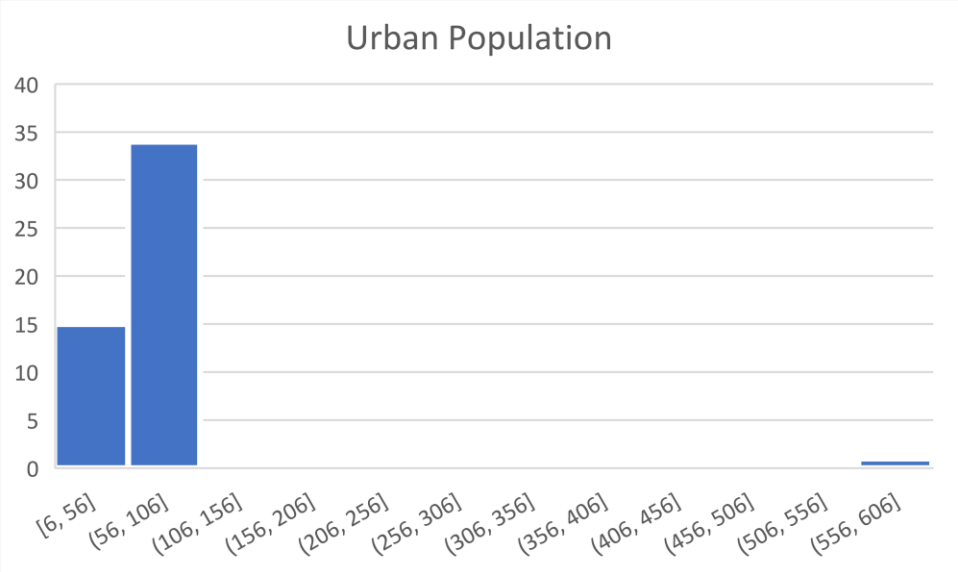
Descriptive analysis is data analysis done on a specific data set.



The Murder distribution is right skewed with the mean is 7.78 murders which is greater than the median at 7.25 murders. While the mode is greatest at 13.2 murders.



The Assault distribution is right skewed with the mean (182.18 assaults) is greater than the median (159 assaults) and the mode (120 assaults). Also, the median is greater than the mode. There is an outlier between 765-885 assaults.



The Urban Population is right skewed with the mean (74.2 people) being greater than the median (66 people), but not greater than the mode (80 people).

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

Measure of dispersion are values that represent the spread of data. This includes range, interquartile range, variance, and standard deviation.

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

Measure of centrality is numbers that represent the center of the distribution. This would be the mean, median, and mode.

	Murder	Assualt	Urban Pop
Mean	7.788	182.1836	74.2
Median	7.25	159	66
Mode	13.2	120	80
Standard Deviation	4.3117347	128.2332	72.67049
Varience	18.591056	16443.75	5281
IQR	7.175	140	24.5
Min	0.8	45	6
Q1	4.075	109	53.25
Q2/Median	7.25	159	66
Q3	11.25	249	77.75
Max	17.4	879	570

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

D

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.