Hannah Garrett
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: hands-on data exploration and evaluation

    Data analytics: broader term that includes data analysis as a necessary part

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

    Murder – independent, continuous, ratio

    Assault – independent, continuous, ratio

    Urban pop – independent, continuous, ratio

    States – categorical, nominal

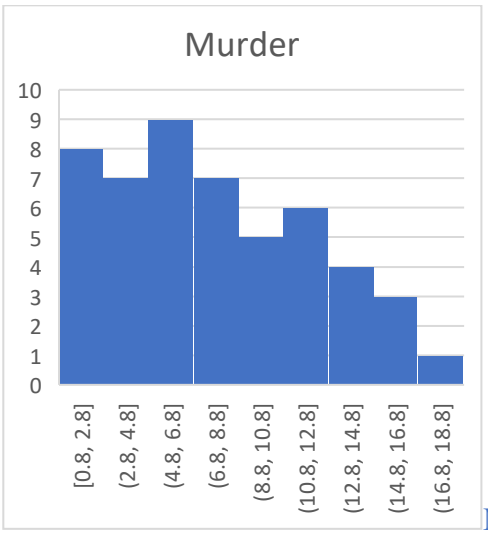
    c. What is the difference between interval and ratio data?

    Interval: data where numbers have equal spacing between adjacent values without an absolute zero

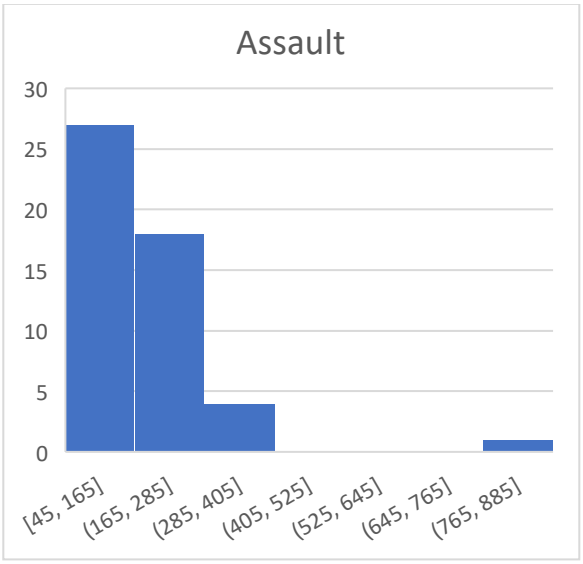
    Ratio: When numbers have units of equal magnitude and rank order on a scale with an absolute zero.

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

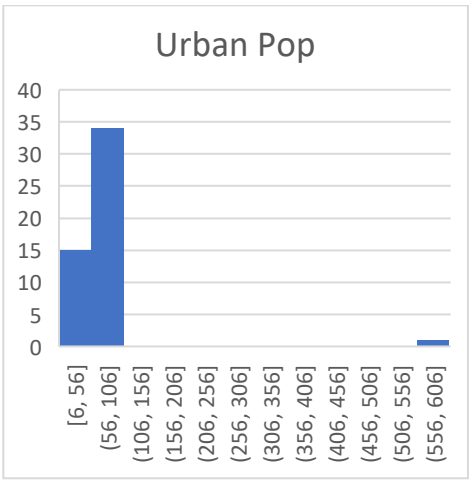
    It is the first kind of data analysis performed on a data set and is applied to large volumes of data, such as census data.



Most data lie between 4.8 and 6.8 and the data is right skewed.



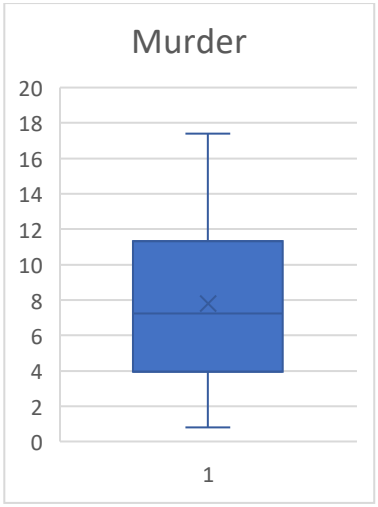
Most data is between 45 and 165. The data is right skewed and has an outlier between 765 and 885.



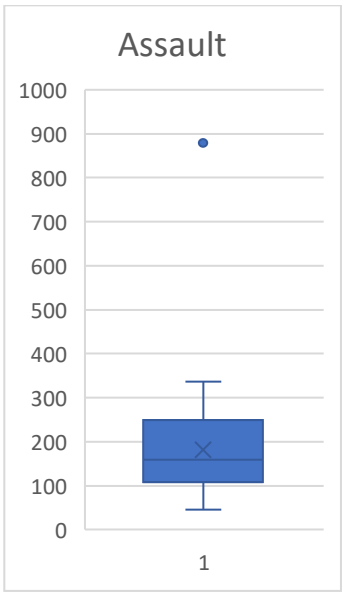
Most of the data is between 56 and 106. The outliers fall between 556 and 606.

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

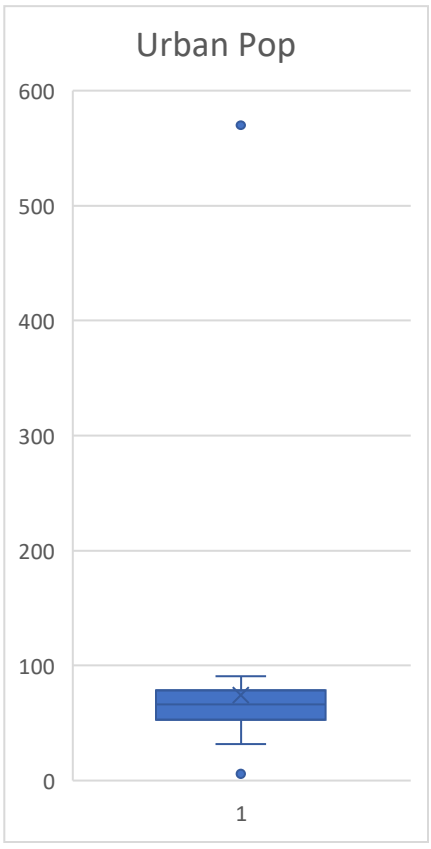
Values that represent the spread of the data



IQR = 7.175



IQR = 140



IQR = 24.5

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

Numbers that describe the center of a distribution.

	Murder	Assault	Urban Pop
Mean	7.788	182.18	74.2
Median	7.25	159	66
Mode	13.2	120	80
Standard Deviation	4.311735	128.2332	72.67049
Variance	18.59106	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.

Used to determine why something happened.
Correlation between murder and assault – 0.649313
Correlation between murder and urban pop – (-0.18617)
Correlation between assault and urban pop – (-0.14066)

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.