

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

Working on Techniques for Analyzing Data

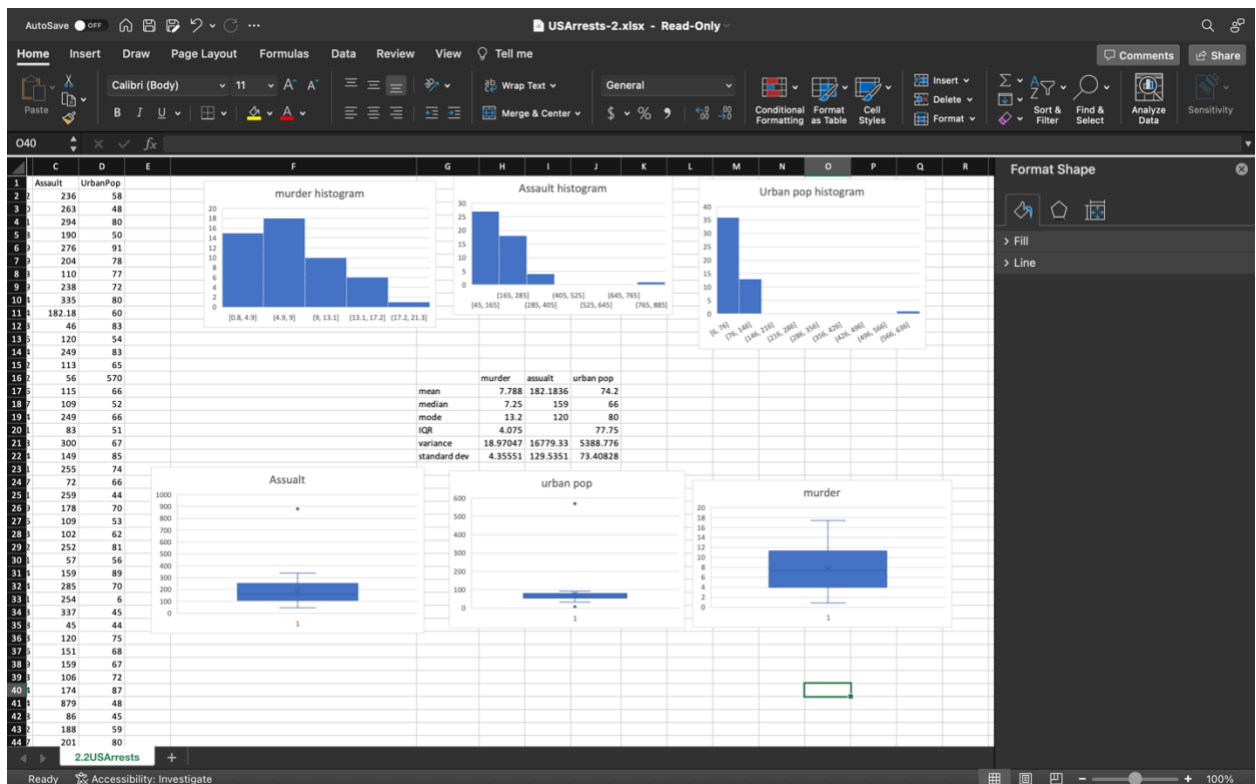
Instructions: Complete the following activities for this project.

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. Data analytics is a 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 = discrete, independent, 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.
 - Ratio data When numbers have units of equal magnitude and rank order on a scale with an absolute zero.
 - Height and weight measure from 0 and above, but never fall below it
 - 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.
 - It is applied to large volumes of data, such as census data.
 - Description and interpretation processes are different steps.
 - Core concepts/techniques:
 - Frequency distribution: histogram
 - Measures of centrality: mean, median, mode
 - Dispersion of distribution: range, interquartile range, variance, standard deviation
- Urban pop and assault both have outliers. Murder and urban pop are left skewed, and assault is right skewed.

- e. What is a measure of dispersion? Calculate the interquartile range of those three variables
 -values that represent the spread of data
- f. What is the measure of centrality? Find the measurement of centrality:
 mean, median, mode
 -numbers that describe the center of a distribution
- g. What are diagnostic analytics? Find diagnostic analysis for pair of variables.

-used to determine why something happened

(murder, assault) = murder is closer to 1 and assault is not. (murder, urban pop)= murder is closer to 1. (assault, urban pop)= assault is closer to 1.



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.