Delia Mahoney
Sept 15, 2021
Week 2 Reading Q's
*worked solo

1. *Which of the plot types show every data point?*
    a. Scatterplots, Cleveland dot plots, QQ plots, and co plots show every data point
2. *Which of the plot types show aggregated or summarized data?*
    a. Histograms and box plots show aggregated and summarized data.
3. Explain what a conditional variable means in the context of graphical data exploration.
    a. A conditional variable in the context of graphical data exploration is a third variable upon which the first two are dependent. In the set of the data from two primary variables, the conditional variable is something that differs among the collection of the two data sets, and possibly represents a discrepancy in the sampling not otherwise identified.
4. List *at least three* of the common measures of spread or dispersion that were mentioned in the readings.
    a. Variance
    b. Standard Deviation
    c. Coefficient of Variation
    d. Median Absolute Deviation
    e. Range
    f. Interquartile Range
5. Choose *two of the measures* in your list and explain how they capture different aspects of the concept of spread.
    a. Standard Deviation- SD measures the amount of variance from the mean. It is a calculated value (square root of variance) that describes ranges along the distribution that classify the dispersal of the data points. The value indicates how far removed from the mean the point is, and where it fall in relation to other points. In a normal distribution, the first SD contains 68% of the data, the second 95%, and the third 99.7%.
    b. Interquartile Range- The IQR gives the middle 50% of a set of data, the range from the 25th to 75th percentiles. From the median, the data set is broken down again, dividing the data on each side in half again. The distance between these two new points is the Interquartile range. This is useful for identifying where the bulk of the data lies, and how outliers can affect these values.
6. List two of the important reasons to perform data exploration (numerical and/or graphical). For each of the two reasons you identify, describe the quantities or plots you would use and the insight you would gain.
    a. One important reason to perform data exploration is to make sure your data makes sense. If your metrics or graphic that you produce seem unusual after the exploration, then it's an indication that something may be wrong and you should go back and reevaluate or double check your calculations. Box plots or Cleveland plots could indicate outliers, and it may be worth checking to see if those points are from observation or if an error occurred.
    b. Another reason for data exploration is to take the opportunity to look for new patterns in the data. The exploration may lead you to ask new questions about your data which could lead to new findings or analyses. For instance, the exploration could yield values that indicate an outside influence at play that might be worth researching.

Delia Mahoney
Sept 15, 2021
Week 2 Reading Q's
*worked solo