Lina Clifford

Week 3: Reading Questions

Worked with: Liz Clark, Olivia Dinkelacker, Laura Haynes

9/22/2022

1) **Which of the plot types show every data point?**
    a. Scatterplot, coplot, and QQ-plots show every data point.

2) **Which of the plot types show aggregated or summarized data?**
    a. Histogram, Cleveland Dot plot, boxplot, and coplots show aggregated or summarized data.

3) **Explain what a conditional variable means in the context of graphical data exploration.**
    a. The conditional variable is the variable by which data are binned or grouped in the context of coplots, conditional histograms, and other conditional plots. A conditional variable could be sex when looking at data of bird wing span.

4) **List at least three of the common measures of spread or dispersion that were mentioned in the readings.**
    a. Three common measures of spread or dispersion mentioned in the reading include skewness, range, and standard deviation.

5) **Choose two of the measures in your list and explain how they capture different aspects of the concept of spread.**
    a. Skew reflects non-normality and captures how asymmetrical the mean of a data set is. Range captures the total range of values in the data from the minimum value to the maximum value. Range helps us to understand how widely spread the dataset is in its entirety.

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 reason to perform data exploration is to scan your data for possible bias or error in collection. You could use a scatter plot to do this and compare the skew/distribution of the data to what you expected from previous similar experiments. If there is extreme

skew in the distribution of points within the scatterplot, it may be warranted to check for errors in data collection.

b. Another reason to perform data exploration is to get a sense of how your variables are related to one another. One way to do this would be measuring correlation, either with Pearson Correlation or Spearman Correlation.