Reading Questions Week 3, 25Sept2022, Rosalyn Bathrick

- **Q1 (1 pt.):** Which of the plot types show every data point?
  - Scatterplot
  - Cleveland dotplot
  - Coplot
- **Q2 (1 pt.):** Which of the plot types show aggregated or summarized data
  - Histogram
  - Boxplot
  - QQ plot
- **Q3 (3 pts.):** Explain what a conditional variable means in the context of graphical data exploration.
  - A conditional variable is a variable that is used to split data into different plots or analyses to understand how the variable impacts the data. ie. Two histograms that separate measurements taken from female/male in a species to show the difference between them.
- **Q4 (1 pt.):** List *at least three* of the common measures of spread or dispersion that were mentioned in the readings.
  - Interquartile range, between the 25^{th} 75^{th} quartiles
  - Standard deviation
  - Coefficient of variance
- **Q5 (2 pts.):** Choose *two of the measures* in your list and explain how they capture different aspects of the concept of spread.
  - The interquartile range captures the center of the data – how much of the data falls between 25%-75%, with 50% being the median.
  - Standard deviation describes how far, on average, each data point falls from the mean
- **Q6 (5 pts.):** 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.

One important reason to perform data exploration is to summarize associations between variables and to get a sense of the relationships. I would use the scatterplot matrix for this visualization, as it provides a number of plots depicting relationships between two variables, and a summary of the correlation coefficients.

A second important reason to perform data exploration is to understand spread of a univariate variable. I would use a boxplot, which would help me visualize the IQR, median, range, and outliers. Comparing variables in boxplots can be helpful in understanding how much they overlap.