Elyse McCormick
ECO 602
9/23/2022

Weekly Reading Questions 3

**All questions answered alone*

Q1 (1 pt.): Which of the plot types show every data point?

Scatterplots, QQ plots, and coplots show every data point, others either aggregate or summarize the data.

Q2 (1 pt.): Which of the plot types show aggregated or summarized data?

Boxplots, Cleveland dot plots, and histograms show aggregated or summarized data.

Q3 (3 pts.): Explain what a conditional variable means in the context of graphical data exploration.

A conditional variable is just the idea that you can use some variable to see if certain conditions have been met and use those conditions to explore and organize your data. For example, a conditional variable in a coplot is your "organizing" variable across the top box, where you can organize your smaller scatter plots within that conditional variable. In the case of a coplot, you may have a conditional variable of sex, (i.e. did it meet the condition of male or female?) and then you can organize your scatter plots for other variables (like mass, size, etc) depending on whether the organism was male or female.

Q4 (1 pt.): List at least three of the common measures of spread or dispersion that were mentioned in the readings.

Variance, standard deviation, and range are three common measures of spread in a data set.

Q5 (2 pts.): Choose two of the measures in your list and explain how they capture different aspects of the concept of spread.

Variance indicates the mean squared deviation from an expected value, the standard deviation shows the root mean squared deviation from the expected value, and the range is the absolute range of values present in the dataset.

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.

Numerical exploration of the data, something like summary statistics, is a great way to explore data because it gives you a general sense of your average values by looking at the mean, it gives you a sense of how variable your data is through the standard deviation, and it gives a look at the median and mode of the data as well, which can tell you how the data is centered and if it's repetitive.

Graphical exploration can be a great way to understand what your data actually look like when plotted. You can see the average dispersion in a histogram, you can see every single point plotted in a scatterplot and compare variables within the dataset to see if there are any emerging trends. You can also organize your data with something like a coplot to see if those trends change when compared to a conditional variable (like sex).