

Stella Dee

Week 3 Questions

I did not work with any other students.

Q1. Scatterplots, coplots, and Cleveland dotplots typically show every data point.

Q2. Histograms, boxplots and QQ plots typically show aggregated or summarized data.

Q3. In the context of graphical data exploration, a conditional variable describes what the data is grouped by. For example, in a coplot of penguin allometric measurements, a reasonable conditional variable could be sex. In a chart or graph, a conditional variable may sometimes (but not always) make more sense to view in the form of a color or symbology decision as opposed to a numerical value.

Q4. Spread or dispersion can be measured as being symmetric, skewed, or bimodal.

Q5. Considering symmetry shows the researcher whether the distribution from the mean is roughly equal when you consider the range from 25-75th percentile. Bimodal spread has multiple peaks. However, spread could be both bimodal and symmetric, if the two peaks are evenly spaced from the overall mean. Spread could also be bimodal but not symmetric, if one peak contained more of the data points than the other. For this reason, they capture different aspects of the idea of spread.

Q6. Two important reasons to perform data exploration include beginning to determine if there could be a relationship between variables and gaining an intuition for what the shape of that relationship may be. For example, in my research, I hope to determine if the degree of pruning has an effect on the damping ratio of trees, and if so, if that relationship might be described as linear, exponential, or some other mathematical descriptor. To do so I might use a scatterplot and attempt to fit a line to gain some insight into this question.