

Matthew Jusino

### Week 3 Reading Questions

(I worked on these alone)

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

Scatterplot, Cleveland dotplot, QQ plot

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

Boxplot, Histogram, Coplot

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

A conditional variable is a variable representing certain “conditions” for separating data to look at differences between groups within the data. For instance, sex is a common conditional variable when doing ecological analysis, as you often want to look at the differences in response to variables in males and females. Another thing you could separate by is status (alive or dead) to see if there might be some correlation between the variables you’re looking at and whether the subject is still living or not. If all members of the species above a certain length are dead, perhaps there is some correlation between length and mortality.

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

Variance, Standard Deviation, Coefficient of Variation, Median Absolute Deviation, Range, Interquartile Range

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

Interquartile Range shows the range of values between the 25<sup>th</sup> and 75<sup>th</sup> quantiles of data, which captures a significant chunk of the data and allows for inferences to be made about the total spread. Coefficient of variation is useful for comparing spread of data for variables on different scales, as it’s expressed as a percentage from the standard deviation divided by the mean, which means that the Coefficient of Variation puts all variables on the same scale, allowing you to compare their dispersion.

- **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.

Using the plot I created yesterday in class for Delomys, a good reason to perform data exploration is to allow one to easily visualize the data. I created a coplot showing the body mass of Delomys subjects,

comparing living and dead. This allowed me to visualize a potential relationship between body mass and mortality in the Delomys, though this was not the case, as the two box plots were comparable. You can also perform data exploration as a way to determine patterns in the behavior or growth rate of a species. If I plot body mass vs age of a species on a scatterplot, then fit a model to it, I can attempt to quantify the growth rate of the species, allowing me to predict the body mass of species based on age, or vice versa, in the future.