Jessica Bonin

Analysis of Environmental Data Lecture

Week 3 Reading Questions

September 19, 2021

Worked with Juliana Berube and Julia Vineyard

1. Which of the plot types show every data point?

* Scatterplots
* Cleveland dotplots
* Pairplots

2. Which of the plot types show aggregated or summarized data?

* Histograms
* Boxplots

3. Explain what a conditional variable means in the context of graphical data exploration.

A conditional variable is a third variable that is used to categorize the data into subsections. For example, if you wanted to compare height and weight, you may want to add the conditional variable of sex (male and female). This is because the relationship between height and weight will be dependent upon each sex.

4. List *at least three* of the common measures of spread or dispersion that were mentioned in

the readings.

* Range
* Variance
* Interquartile Range

5. Choose *two of the measures* in your list and explain how they capture different aspects of the concept of spread.

* Range: The range is the difference in value from your maximum value and your minimum value. This simply tells you how spread out the data are.
* Variance: The variance expresses the spread of data around the mean of the data. It tells you how close the data is from the mean.

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.

* One important reason to perform data exploration is to find outliers or errors in the dataset. An easy way to visualize this is by creating a scatterplot or boxplot and exploring the extreme outliers.
* Another important reason to perform data exploration is to explore the relationships between variables in your dataset. By plotting points in a scatterplot with different variables on the x and y axes, you can visually observe associations between the two. You can then run statistical analyses like linear regressions or other models to fit the data into functional relationships.