

Week 3 Reading Questions

Desireé Smith

1. Which of the plot types show every data point? Scatterplot, coplot, QQ-plots
2. Which of the plot types show aggregated or summarized data? Histogram, box and whisker plot
3. Conditional plot, conditioning variable, and related terms occurred throughout the Zuur and McGarigal readings.

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

A conditional variable means that the variable that is set in place may or may not happen. You have to see if a condition is met or not to see what the outcome will be. You can look at previous finding to see they likelihood of it occurring, but it is not set in stone.

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

The three common measures of spread/ dispersion are variance, standard deviation and coefficient of variation.

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

The standard deviation is the square root of the variance. It will show us how clustered or how spread out the data is in relation to the mean. The coefficient of variation is the ratio of the SD to the mean so it will show us the level of dispersion around the mean. If the CV is small, then it shows us that there is less dispersion, and we can use it to see a precise estimate.

6. Consider a dataset that you have collected or worked with.

If you haven't worked much with existing datasets hypothesize a dataset that you might collect for your research.

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.

While I was in undergrad, I got the opportunity to intern with a Marine Science lab. I was helping a PhD candidate collect data so she could create a mesocosm for her specimen. It was important for us to perform data exploration so we can create something that would be an accurate representation of the specimen's real habitat. This would help ensure that the results are like what would be found in nature. We went out in the field and measured

the length of the blades of grass and pneumatophores roots so we could replicate it in the mesocosm. We also measure the amount of sunlight in the salt marshes to see how much shade we would need. To graphically display the length of the vegetation I would use a box plot. This would allow us to see the mean height of the vegetation and replicate that in the mesocosm. Another reason to perform data exploration is for us to make predictions. It can help show us a pattern in the data that we can then use to make prediction with future data. We can use inferential statistics to help make prediction about how the species in the wild would behave based on the research.