Johanna Ravenhurst
Submitted: September 25, 2022
I did not work with other students.

Week 3 Reading Questions – ECO 602

this week's readings are:

- McGarigal Chapter 3: Exploration, sections 1 - 5
- Zuur Chapter 4: Exploration, section 4.1.
    - You don't need to read about the last two plot types: lattice graphs and design/interaction plots
- Bolker Chapter 2: Data Analysis and Graphics
    - Read Sections 2.1 through 2.3. There are a lot of helpful R examples and tips.

PLOTS

Consider the following types of plots described in the McGarigal and Zuur readings:

- Histogram
- Scatterplot
- Cleveland dotplot
- Boxplot
- QQ plot
- coplot

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

- Cleveland dotplot
- Scatterplot
- QQ plot
- coplot

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

- Histogram – aggregated to show the distribution of the data
- Boxplot – mean/median, spread

CONDITIONING VARIABLES

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

- It can be useful to create graphs that are conditional on a certain variable, such as species or sex. This means that the data are separated based on that conditional variable so that we can see if there are any potential differences in the data based on something like the sex of the animal.

DISPERSION

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

- Variance
- Standard deviation
- Range

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

- The range tells us just the minimum and maximum values of the data, so it gives information about the extreme values for our data, or the boundaries of the spread.
- The variance tells us how much the data varies around the mean, so it tells us more about how the data is distributed within the range of values.

DATA EXPLORATION

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.

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.

- It's important to perform data exploration in order to see if there are outliers in your data that might impact that analysis. A boxplot could show if the data for a certain variable includes extreme outliers. It's possible that outliers could be due to measurement errors, so it is worthwhile to examine them closely.
- Conducting data exploration to investigate missing values in your data can be very useful when deciding which variables to keep in your analysis. If a given variable contains an excessive amount of missing observations, it might not be useful if you need to include all or most of the records in your final analysis. I would use summary functions and cross-tabulations to investigate missing data for my variables of interest.