Jessica Martinez

Analysis of Environmental Data

Week 3 Reading Questions

(\*I worked independently on this assignment\*)

**Plots**

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

**The plot types that show every data point are: scatterplots, coplots, Cleveland dotplots, qqplot**

**QQ plot, scatterplot, Cleveland dotplots, coplots**

* ~~Histogram~~
* Scatterplot
* Cleveland dotplot
* ~~Boxplot~~
* QQ plot
* Coplot

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

**Histograms, boxplot,**

The Plot types that show aggregated or summarized data are Boxplot, histograms,

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

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

**Conditioning Variables**

Conditional plot, conditioning variable, and related terms occurred throughout the Zuur and McGarigal readings.

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

**For boxplots, the conditioning varible always has to be categorical**

**Coplots: can condition on a continuous or categorical**

**Conditioning variables help to display three dimensional data**

In the context of graphical data exploration, a conditional varible is

* A conditional plot (a coplot) is a scatter plot of two variables when conditioned on a third varible. The third varible is the conditioning varible. This variable can have both values either continuous or categorical. In the continuous variable, we created subsets by dividing them into a smaller range of values. In categorical variables, the subsets are created based on different categories
* Coplots: shows slices through a 3 dimensional space
* When we plot a coplot with our conditioning varible as a categorical varible we can create multiple scatterplots at the levels of our conditioning varible
* When we plot a coplot with our confitioning carible as a continuous varible, bins may be different widths, causing some overlap (depends how the binning algorthim works)
* From McGargial:
  + Coplot: scatterplot of two variables conditioned on a third (Ie the relationship between x and y given z)
  + In cases involving the relationship between a dependent and independent varible, the relationship may be obscured by the effects of other variables.
* From Zurr: conditional values can be nominal or conditional. From nomial variables, there is no overlap in ranges of the conditional variable. For continuous variables we can allow for some overlap in the ranges of the conditional variables, and the number of graphs can be modified.

**Dispersion**

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

* From McGarigal🡪 measures of spread (measure the dipersion of the data set or how spread the data is)
  + Variance: mean squared deviation from the mean or expected value
  + Standard deviation: root mean squared deviation from the mean or expected value
  + Coefficient of variation: normalized measure of spread: relative standard deviation

Text, letter

Description automatically generated

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

**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.