Analysis of Environmental Data
Reading Questions – Week 3
Olivia Dinkelacker, with: Liz Clark, Lina Clifford, Laura Haynes

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

Scatterplot, QQplot, Coplot

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

Histogram, Boxplot, Coplot, Cleveland dotplot

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

The conditioning variable is an additional variable by which the data set is divided (two variables conditioned by a third variable). The conditional variable can me nominal or continuous, but have to be categorial. If the conditional variable is e.g. week, then the data is divided by week and shows the relation of two variables by every week.

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

Standard deviation
Variance
Median absolute deviation

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

Variance measure of how data points differ from the mean
captures the mean squared deviation from the mean or the expected value.

The median absolute deviation (MAD) captures the median absolute deviation from the median.

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.

Data exploration is important to see how our data is distributed. We can use a histogram to do that. If the histogram of our data is bell-shaped, it is normally distributed. If the histogram shows a long tail, it is not normally distributed. Another important thing is to check for unexpected patterns, correlations or skewness. We could use a scatterplot to get more insight into that.