Ollie Murphy
ECo 602
Week 3 Reading Questions

1) Which of the plot types show every data point?
    a) QQ Plot
    b) Scatterplot
    c) Cleaveland dotplot
    d) coplot
2) Which of the plot types show aggregated or summarized data?
    a) Histogram
    b) boxplot
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) In the context of graphical data exploration, a conditional variable is an additional variable (often a third variable) that breaks down the relationship between two (or more) main variables that are being plotted. In the case of a coplot, for example, multiple scatterplots of two variables are divided by their relationship to a third variable—the conditional variable. For a more specific example, consider a hypothetical study of small mammal abundance as it relates to the number of outdoor cats in a particular area. In a coplot, the scatterplots might plot number of observed rodents versus number of observed cats, divided by the conditional variable of the time of year that the counts were taken. Other plots, like the boxplot, can also be split by conditional variables.
4) List *at least three* of the common measures of spread or dispersion that were mentioned in the readings.
    a) Median absolute deviation
    b) Range
    c) Interquartile Range
5) Choose *two of the measures* in your list and explain how they capture different aspects of the concept of spread.
    a) Range vs. Interquartile range. Range looks at the distance between the minimum and maximum data points, while interquartile range shows the distance between the 25^{th} and 75^{th} quartiles of the data points. Range, then, will give you an idea of the absolute range of the data, including any outliers. Interquartile range, however, will give you an idea of the range of the middle 50% of the data, leaving out the ends and any outliers therein.
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.
    a) First, to identify any outliers in the data sets. You could use a boxplot to identify possible outliers, and a comparison of the boxplot with a Cleaveland dotplot (especially if the data is ordered by a relevant conditional variable) to decide if the points are indeed outliers.
    b) Second, to identify relationships and possible correlation between pairs of variables, you could use either a scatterplot or, for data matrices with more than two variables, a scatterplot matrix. Examining the resulting plots would help you gain insight to how certain variables are related, especially in the case of variables that you may have assumed to be independent.