Quiz 1

1. The locations of the observations of a variable correspond to where the observations sit on the number line or in space. The scale of a variable is how dispersed the observations are across the number line. The distribution of a variable defines the frequency of different observations across the variable's range.

2. Boxplots can be used to describe the scale of a variable because they visualize the min, max, range, and quantiles of a variable's observations. Enumerative plots are also good at displaying the dispersion of a variable.

3. Scatter plots show the form, strength, and dependence no external circumstances of the relationship between two variables. The value of the dependent variable determines the relative location of the symbol on the x-axis, and the value of the dependent variable determines the position of the symbol on the y-axis.

4. Ggplot2 builds on the idea of *The Grammar o Graphics* and allows you to build a plot using layers of geometric elements, stats, scales, facets, legends, and themes. This allows plots using ggplot2 to be extendable and flexible, but can come at the cost of the code being longer.

5. The advantage of using a bubble plot for a choropleth map is that the value of the variable can be easily visualized regardless of the geographic area of the variable's region itself. For example, when using a bubble plot for Oregon county population, we can easily see that Multnomah county has a large population, even though its area is relatively small. This allows for the representation of variables to be decoupled from the area or shape of the geography.

6. Using descriptive statistics for describing data allows us to see the exact values of certain statistics (or at least approximations in the case of chi-squared). Looking at a scatter plot might allow us to guess at the strength of the relationship between to variables, but statistics like the chi-square statistic give us a means of measuring the relationship numerically. However, descriptive statistics cannot always show the whole picture'. These statistics summarize the data, and some information is therefore either not shown or lost. Not all observations are enumerated.

7. Trellis/lattice style plots are especially useful for finding the dependence of the relationship between variables on some additional variables. Coplots allow us to see scatter diagrams of one variable as the function of another variable, conditioned by a third variable. This conditioning variable allows us to see differences in the form and strength of the relationship of the first two by some grouping (i.e. 'Yes' as a function of population, conditioned by the country of the voters).