

Kaitlyn Valentinetti

September 21, 2022

ECo 602: Michael Nelson

Week 2 Reading Questions

1. The following plot types show every data point: scatterplot, QQ plot, Cleveland dotplot, and coplot.
2. The following plot types show aggregated or summarized data: histogram and boxplot.
3. A conditioning variable is a grouping factor that allows the data we are exploring to be aggregated into bins for a graphical model. They can be nominal or continuous—in the case of the latter, the bins may overlap with one another. Conditioning variables provide more information about the data because it can be used as a third or fourth dimension of your data that, and they make the graph simpler to see patterns because they are separated into slices for each bin. Examples of conditioning variables are sex and body mass.
4. Three common measures of spread or dispersion that are mentioned in the text are standard deviation, median absolute deviation, and range.
5. The standard deviation is the square root of the variance which is the mean squared deviation from the expected value. Since the standard deviation is square rooted, it's in the same units as the measurement variable and is interpretable. It helps us visualize the likelihood of an event to fall near the mean. The range is the minimum value of data subtracted from the maximum value. This is useful to see the absolute difference from the low end to the high end of the data points and get an idea of the total spread.
6. It is important to perform data exploration prior to your analysis because it saves a lot of time during data analysis and makes the analysis much easier. Data exploration is graphically beneficial as it allows you to discover which graph best represents your data and the concept you are discussing. For example, the normal distribution model is frequently used as a reference framework for experiments, and it shows the single central tendency of a frequency distribution. It is universally used as a baseline to compare the frequency distribution of other experiments. The comparison tells you about the probability of the data falling around the mean. Exploration also allows you to screen your numerical data for problems before you fully commit to a sophisticated model. An

example of a plot you would use to screen the data for problems could be a box-and-whiskers plot. A box and whisker plot measures the lowest and highest values, the 25% and 75% quartiles, and the median. In some instances, it could be normal to have outliers (lowest and highest values) that are far away from the median, but it could indicate there is an error in the data entry which is exactly why boxplots are useful in data exploration. Even if there wasn't an error, seeing an outlier on the boxplot could be a reason to transform the data to reduce the impact of the outlier on the analysis.