



Week 2 Lecture

Data Visualization Part 2

Last week we kicked off this class. We concluded last week with an introduction to data visualization. This week we will expand our discussion on data visualization by considering quantitative, or numerical data. When graphing Quantitative Data, you may use different techniques when compared to qualitative data. Some examples of graphing quantitative data may include histograms or time series plots.

The following link provides an excellent summary of data visualization for quantitative variables:

<https://mat117.wisconsin.edu/3-visualizing-quantitative-data/>

(Links to an external site.)

The second chapter this week is concerned with correlation and linear regression. As humans, we are often concerned with cause and effect. Data analytics seeks to help us understand the cause and effect of variables. One common way at looking at the association of variables is to develop a scatterplot and determine if there is indeed a relationship.

A scatterplot is a depiction of two variables each plotted on an X-Y diagram. The X axis is the horizontal line, and is the independent variable. The Y axis is the vertical line, and is the dependent variable. The dependent variable is *dependent* on what happens to the X variable. As an example, as the temperature goes up in the summer (temperature is the independent variable) the cost of my air conditioning goes up also (A/C cost is dependent on temperature). This situation can be plotted on a scatter plot showing a strong and positive relationship.

The following link will take you to a history of scatter plots. For nerds like me, this is fascinating:

<https://qz.com/1235712/the-origins-of-the-scatter-plot-data-visualizations-greatest-invention/>

(Links to an external site.)

An important concept needs to be mentioned at this point. That concept is that correlation does not always imply causation. As an example, a recent study was conducted regarding bank robberies and ice cream sales. It was discovered that the peak month of bank robberies is in July. It was also determined that ice cream sales also peaked in July. Conversely, both bank robberies and ice cream sales are both at their lowest points during the month of January.

Instead of providing a dry scientific link to an article, the link below is an interesting way to depict data that is certainly not correlated (at least in a scientific sense):

<https://www.fastcompany.com/3030529/hilarious-graphs-prove-that-correlation-isnt-causation>