pennsylvania
SYSTEM OF CARE PARTNERSHIP

Data 101 – The Basics

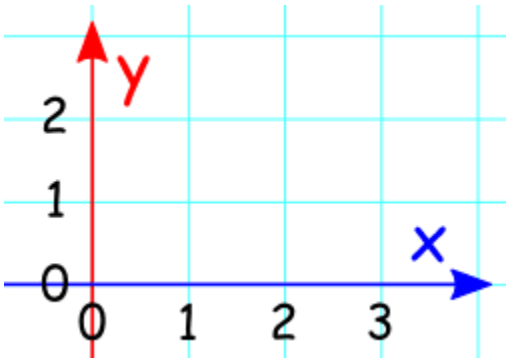
Goal: ***Tips for how to learn to speak the language of data and evaluation. Terms defined with simple explanations.***

Data is any information you are collecting: numbers, statistics, or measurements. It can also be words, observations, or any other inputs.

- **Quantitative** – numbers or things that can be measured or counted. (Ex. I paid $3.00 for 12 ounces of coffee that was brewed at 200 degrees.)
- **Qualitative** – things you can observe but not easily measured or counted – social interactions, feelings, colors, etc. (Ex. My coffee tastes bitter, looks frothy, has a nice aroma, and is in a red cup.)

N = (number) and refers to the total number of people/subjects/items represented in the graph.

Y axis is the vertical (up-down) line on the side of a graph.

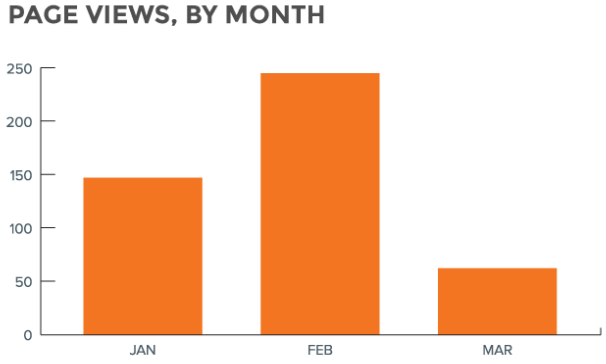


X axis is the horizontal (left-right) line at the bottom of a graph.

A **variable** is something that can be measured or counted. It is something that can increase or decrease depending on the situation that you are measuring. You can find variables as the titles on the X and Y axes.

Cross-sectional Data is something that is only measured once. It is a snapshot of variables at one point in time.

Longitudinal Data is something that is measured repeatedly over time. You can look at the same variable at different times.



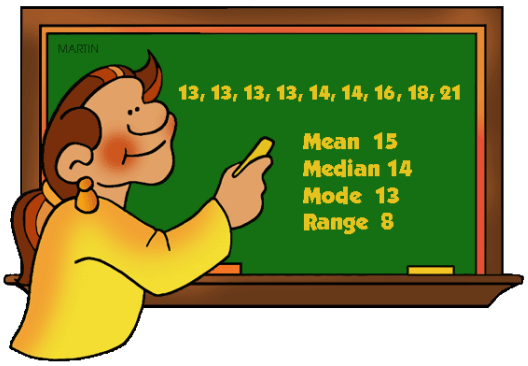
Data Set is a group of variables that can be compared to each other or looked at individually in order to understand how they relate.

Mean – The average, or the sum of all values in a series divided by the number of values.

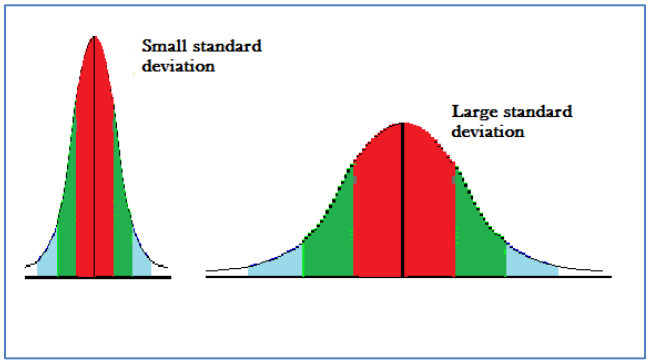
Median – The middle value in a series.

Mode – The most frequent value in a series.

Range – The difference between the highest and lowest values in the series.



Standard Deviation – The measure of how far a variable is from the mean. A low standard deviation shows that the scores are clustered around the mean. A high standard deviation shows that the scores are spread out over a larger range.



Correlation – is where two or more variables demonstrate a positive or negative relationship to each other.

Positive Correlation - when one variable goes up the other goes up (Ex. The more time you spend running on a treadmill, the more calories you will burn).

Negative Correlation – when one variable goes up the other one goes down (Ex. The more you exercise, the less you will weigh).

Correlation does not equal causation! *It can be just a coincidence that things are related. (Ex. There is a strong statistical negative correlation between the increase in lemons imported from Mexico and the decrease in U.S. highway fatalities!)*

p value – how likely something is to NOT be true. (Ex. a p value of 0.05 means that there is a 5% chance that the result was just a coincidence.) The smaller the p value, the more likely the data results are accurate. ***Look for p values at the bottom of data tables to see how significant the results are.*

Statistically significant – means that the data presented is probably true (not due to chance). A p value of 0.05 is considered the standard for data to be "significant" but some people prefer a number even lower.

Reliability – the degree that the results are stable and consistent no matter how many times you repeat it.

Validity – the degree that the data is measuring what it claims to measure.