Objectives:

* To begin to appreciate that ***data***are numbers with a context that are studied for a purpose
* To learn to recognize different classifications of ***variables***
* To become familiar with the fundamental concept of ***variability***
* To discover the notion of the ***distribution*** of a variable
* To gain some exposure to the types of questions that statistical reasoning addresses
* To begin to gather experience describing distributions of data verbally

The numbers that you use to describe a situation, for example the number of something or the category of something etc are considered variables. Numbers in and unto themselves are not data. ***context*** is what is needed to make a number a piece of data.

Since the context of the number is important in the determination of if it is data then we should always describe the context when conveying the number. *i.e.* It is better to say that “the smallest number of states visited is 12 “as opposed to “the smallest item in the data set is 12” (No context - difficult to understand meaning.)

***Variable:*** any characteristic of a person or thing that can be assigned a number or category

***Measurement or Quantitative variable:*** one that can assume a range of numerical values (Number of states)

***Categorical variable:*** one that records a category designation (Republican, Democrat, Conservative, etc.)

***Binary categorical variables:*** a variable that has exactly two possible categories (gender, coin flip)

***Observational unit or case:*** The person or thing that is assigned the number or category (each person in the class when asking the topic questions would be an observational unit)

***Distribution of variables:*** The pattern of ***variability*** of a set of data, much of statistics is finding way to express this distribution and analyzing the meaning or effects of the distribution.

An example of a visual display of a distribution is the ***bar graph*** *– specifically for expressing categorical variable distributions.*

Another visual display of distribution of data is the ***dotplot***. dotplots are used for relatively small data sets of *quantitative variables*

Numerical display of the spread of a distribution: *Standard deviation*.

***Frequency:*** The number of occurrences of each possible response in a dataset

***Tally:*** To count the number of occurrences of each numerical value when a fairly small number of values are possible.