

Types of variables

StatPREP Class Activity

Orientation

There are two major types of variables:

- **Quantitative**, where the *value* of the variable is a number.
- **Categorical**, where the value of the variable is one of a set of labels. That is, the value tells which *category* or *group* the row falls into.

Quantitative variables cover an *interval* of the number line. That interval is described by two numbers: the minimum value and the maximum value of the variable.

Quantitative variables also have a *spread*. There are different ways to measure the spread. In this lesson, we'll use the *difference* between the maximum and minimum value. Note that the *interval* is two numbers while the *spread* is a single number.

The *spread* is said to measure the *variability* in the variable. Other ways to quantify the spread, which we won't use in this lesson, are the *standard deviation* and the *variance*.

For categorical variables, we don't use the concepts *interval* or *spread*. The possible or allowed values of a categorical variable are called the *levels* of that variable. For example, the levels of a categorical variable describing "commute type" might be: walk, bike, bus, drive, etc. The levels of a categorical variable like "language spoken" might be English, Spanish, Chinese, and so on.

The levels of many categorical variables are *unordered*. This means that the concept of *between* doesn't naturally apply. For instance, there is no natural way that Spanish is between English and Chinese.

Some categorical variables are *ordinal*, which is just to say that there is a natural order to the levels. An example is a variable recording "opinion," which might have levels Disagree, Neutral, Agree.

Activity

Open the [Point Plot Little App](https://dtkaplan.shinyapps.io/LA_point_plot/). (See footnote¹). Set the data frame to NHANES and the sample size to be several hundred.

¹https://dtkaplan.shinyapps.io/LA_point_plot/

1. Pick a *quantitative response* variable and a *categorical explanatory* variable. You might have to use trial and error to find such variables but once you do, it will be evident in the graph. Please don't use ID or SurveyYr – stick to the variables that are about the individual being surveyed.

Write down the names of the variables you selected. For the categorical variable, write down each of the levels.

- Measure the *interval* spanned by the quantitative variable for each group defined by the categorical variable. You can use the measuring stick built into the app. (That is, click at a point in the graphics frame and drag the mouse to select a vertical interval.)

Write down the interval spanned by the values of the response variable for each categorical level of the explanatory variable.

Level of explanatory variable	Interval spanned by response variable
-------------------------------	---------------------------------------

- Also, measure and record the *spread* of the quantitative variable for each categorical level.
2. Using the selector for the explanatory variable, find five more categorical variables in the NHANES data. At least two of your five should be *ordinal* variables.

For each, write down the levels, and whether the variable is unordered or ordinal. . . .