Variable Types

5 min

In our tree census, we are collecting data about two types of variables: one that we measure (height) and one that we categorize (species).

The difference between measuring and categorizing is so important that the data itself is termed differently:

- [Variables](#) that are measured are **Numerical** variables

- Variables that are categorized are **Categorical** variables.

Numerical variables

Numerical variables are a combination of the measurement and the unit. Without the unit, a numerical variable is just a number.

Imagine I go into a cafe and ask the barista for 3. Three what? ☕? 🍩? 💵? Or my friend asks how far Toledo is and I say 300. 300 miles? Kilometers? Minutes? Without units, numbers don't mean anything.

There are two ways to get a number: by counting and measuring. Counting gives us whole numbers and **discrete** variables. Measuring gives us potentially partial values and **continuous** variables.

In our tree census, we are measuring the height of our trees in feet (indicated in the variable name, 'Height (ft)'), a continuous variable.

Categorical variables

Categorical variables describe characteristics with words or relative values.

In the tree census, trees species are described with words like London Plane, Honeylocust, or Pin Oak. This is the best description and encodes all the information we need about the species. This kind of categorical variable is a **nominal variable** which literally means a named value.

We also captured whether or not our trees grew alone. In our 'Single' variable, there were only two options: Yes and No. This is called a **dichotomous variable**. Dichotomous variables have only 2 logical possibilities, "on/off", "yes/no", "true/false", "0/1", there's no middle ground and no 3rd option. If there is a logical third option, it's not a dichotomous variable.

Finally, let's say that we wanted to capture how "pretty" we thought each tree was. This isn't really a thing we can measure, but we can subjectively say on a scale of 1 to 5, how pretty we think each tree is. The prettiest trees are a 5, the least pretty trees are a 1.

That ranking is inherently ordered and therefore called an **ordinal variable**.

Ordinal variables are really popular in survey design "on a scale of 1-5 how much do you agree with this statement?" This is called a **likert** scale. They also show up in the Olympics and other competitions where someone wins 1st, 2nd, or 3rd place.

Ordinal variables can get a little confusing because they are often represented as numbers. But they don't represent measurements or counts, they represent categories. For example, let's say an Olympian wins Gold and Bronze medals, it doesn't make sense to say that they averaged Silver. The same is true of likert scales: there's no average between "Very pretty" and "Pretty."

Instructions

Consider the ranking of the trees to the right. There are 3 categorizations:

- o   'Place', represented by the awards, 1st, 2nd, and 3rd place,

- o   'Ranking', which represents the categorization of our trees in our dataset, and

- o   'Prettiness', which is the verbal description of how pretty each tree is.

Place is an ordinal variable just like our Ranking. Take a moment to consider the relationship between Place, Ranking, and the nominal description.

Note that the "least pretty" tree is still beautiful 🌳 ❤️

Prettiest
5

Pretty
3

Least Pretty
1