

1.2.1: What Are Data?

The first important point about data is that data *are* - meaning that the word “data” is plural (though some people disagree with me on this). You might also wonder how to pronounce “data” – I say “day-tah” but I know many people who say “dah-tah” and I have been able to remain friends with them in spite of this. Now if I heard them say “the data is” then that would be bigger issue...

2.1.1 Qualitative data

Data are composed of *variables*, where a variable reflects a unique measurement or quantity. Some variables are *qualitative*, meaning that they describe a quality rather than a numeric quantity. For example, in my stats course I generally give an introductory survey, both to obtain data to use in class and to learn more about the students. One of the questions that I ask is “What is your favorite food?”, to which some of the answers have been: blueberries, chocolate, tamales, pasta, pizza, and mango. Those data are not intrinsically numerical; we could assign numbers to each one (1=blueberries, 2=chocolate, etc), but we would just be using the numbers as labels rather than as real numbers; for example, it wouldn’t make sense to add the numbers together in this case. However, we will often code qualitative data using numbers in order to make them easier to work with, as you will see later.

2.1.2 Quantitative data

More commonly in statistics we will work with *quantitative* data, meaning data that are numerical. For example, here Table 2.1 shows the results from another question that I ask in my introductory class, which is “Why are you taking this class?”

Table 2.1: Counts of the prevalence of different responses to the question “Why are you taking this class?”

Why are you taking this class?	Number of students
It fulfills a degree plan requirement	105
It fulfills a General Education Breadth Requirement	32
It is not required but I am interested in the topic	11
Other	4

Note that the students’ answers were qualitative, but we generated a quantitative summary of them by counting how many students gave each response.

2.1.2.1 Types of numbers

There are several different types of numbers that we work with in statistics. It’s important to understand these differences, in part because programming languages like R often distinguish between them.

Binary numbers. The simplest are binary numbers – that is, zero or one. We will often use binary numbers to represent whether something is true or false, or present or absent. For example, I might ask 10 people if they have ever experienced a migraine headache, recording their answers as “Yes” or “No”. It’s often useful to instead use *logical* values, which take the value of either `TRUE` or `FALSE`. We can create these by testing whether each value is equal to “Yes”, which we can do using the `==` symbol. This will return the value `TRUE` for any matching “Yes” values, and `FALSE` otherwise. These are useful to R knows how to interpret them natively, whereas it doesn’t know what “Yes” and “No” mean.

In general, most programming languages treat truth values and binary numbers equivalently. The number 1 is equal to the logical value `TRUE`, and the number zero is equal to the logical value `FALSE`.

Integers. Integers are whole numbers with no fractional or decimal part. We most commonly encounter integers when we count things, but they also often occur in psychological measurement. For example, in my introductory survey I administer a set of questions about attitudes towards statistics (such as “Statistics seems very mysterious to me.”), on which the students respond with a number between 1 (“Disagree strongly”) and 7 (“Agree strongly”).

Real numbers. Most commonly in statistics we work with real numbers, which have a fractional/decimal part. For example, we might measure someone’s weight, which can be measured to an arbitrary level of precision, from whole pounds down to micrograms.

This page titled [1.2.1: What Are Data?](#) is shared under a [CC BY-NC 2.0](#) license and was authored, remixed, and/or curated by [Russell A. Poldrack](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

- [2.1: What Are Data?](#) by [Russell A. Poldrack](#) is licensed [CC BY-NC 4.0](#). Original source: <https://statstinking21.github.io/statstinking21-core-site>.