

# Understanding Data Types

## Understanding Data Types

Like reading a recipe before beginning to cook, understanding the data types in your inputs helps ensure that you have all the necessary ingredients to build a successful workflow.

Data types are not important just for a computer's processing, but for your own processing, too. Understanding the differences between types of data helps derive significance from data values and develop an efficient approach to producing meaningful results.

## The Five Data Types

In Designer, a column of data is categorized as one of five types: String, Numeric, DateTime, Boolean, or Spatial. In this lesson we'll explore these types in the context of data from an input we will use in our workflow: a survey of trees in New York City. Click a data type to learn more about it.

### Strings

A string is a sequence of characters, such as letters, numbers, symbols, and spaces. The name of a tree, for example, is a value that should be categorized as a string. Even something like a tree's ID, which contains only numbers, is a value that should be assigned a string data type since it's a qualitative attribute. Qualitative values are those that describe qualities of an object that can be categorized, not measured. These values should not be used for mathematical computations. The length, or the count of characters, of string values can vary. In Designer, there are two string subtypes to consider: variable & fixed. For data that varies in length, like the names of trees, variable string types should be used. If all values are expected to be the same length, like a six (6) digit Tree ID code, then fixed is a more efficient subtype.

### Numbers

Numeric data is quantitative, meaning that it can be measured. Values such as the height of a tree or count of trees in a neighborhood are examples of this data type. Designer recognizes seven (7) numeric data types that include integers, or whole numbers, and decimals. Bytes and integers accommodate whole numbers of different sizes, spanning from numbers containing values from zero (0) to 255, all the way to nine quintillion. Decimal types (float, double and fixed) allow for additional precision in the form of decimal values. Float and double are more exacting than fixed, with double being the most precise option, whereas fixed decimals are configured to fit your needs.

### DateTime

DateTime data types can be thought of as three types: Dates, Times and the combination of the two. Values of this type are familiar and easy to identify but can be formatted many ways. Datetime datatypes are important for correctly performing calculations such as "How many trees were planted in the past six months?" or even "What day was it 15,000 days ago?"

## **Boolean**

Boolean data types are the smallest data types in Designer. This data type only has two logical values: 0, or False, and not 0, or True. It is popular to use this datatype when just wanting to categorize a value as one of two values, such as a tree that is "Dead" or "Alive".

## **Spatial**

A spatial data type is made up of either a point represented by an X and Y coordinate, a line, or a polygon.

## **Viewing Data Types**

While you may not know the data types in your inputs before building your workflow, Designer provides a window into this information, so you can easily view the data types with which you are working at any point in your workflow. After running a workflow, simply select a tool to view its metadata in the Results window.

## **Viewing Metadata**

Click the Input Data tool to view the data types associated with the data at the beginning of the workflow.

Then, click the "Metadata" button in the Results Window.

Rather than the actual data values, you now see the metadata associated with the values in each column of our input. Metadata is simply data about data, such as its data type, size, source and description (if available). It appears that all our columns in this input are categorized as variable strings, or V\_Strings.

View the metadata from another tool in the workflow: the Sort tool. Click the Sort tool.

Then, click the "Metadata" button in the Results Window.

Along the workflow, changes have been applied to data types to allow for different analytical processes, resulting in a data stream whose data types are different from when it was first read in. Checking data types often while developing your workflow can help avoid some common pitfalls such as unnecessarily consuming memory, causing a workflow to take a long time to run, encountering error messages, or outputting incorrect results.

## **Outputs**

Depending on the file type used to read in or output data in Designer, the data types you need for your analysis may not be available. Some file types, such as Alteryx databases, will recognize many different data types while others, such as CSVs, are limited to recognizing strings. If your data is not categorized with a data type that is appropriate for your analysis, don't worry! You can configure and change the underlying data type as part of a workflow.