4.2.3

What Are Data Containers?

When we refer to **collecting data**, we mean importing data into a **container**, or structure, and then organizing the data so that we can work with it.

In Pandas, you'll work with two types of containers: the Series and the DataFrame. You've already gotten an overview of both and practiced using them. Now, let's discuss how these containers relate to each other, how they differ, and how we create a DataFrame.

Pandas Containers: Series and DataFrames

A **Series** is a container for a sequence of data. It corresponds to a column in a spreadsheet and can hold any type of data.

A **DataFrame** is the equivalent of two or more Series. It corresponds to a whole spreadsheet. With a DataFrame, we can therefore work with rows and columns (that is, tabular data).

The following image shows an example of two Series and one DataFrame:

Series			Series			DataFrame		
	AAPL			**GOOG**			**AAPL**	**GOOG**
0	0.50		0	0.45		0	0.50	0.45
1	1.37	**+**	1	0.63	**=**	1	1.37	0.63
2	0.62		2	0.55		2	0.62	0.55
3	-0.15		3	-0.04		3	-0.15	-0.04

In the preceding image, notice that one Series contains four Apple stock prices, and the other Series contains four Google stock prices. When combined, the two Series make up a DataFrame that has two columns. One column has the same four Apple prices, and the other column has the same four Google prices.

As mentioned, a Pandas DataFrame is structured like a spreadsheet. But, it offers more-powerful ways for us to express and manipulate data than any available spreadsheet software.

DEEP DIVE ▼

Now, before moving on, check your knowledge with the following assessment:
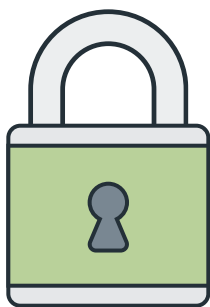
Access Denied

You don't have access to view this resource.

How Do We Create a DataFrame?

One way to create a DataFrame is to collect, or read, data from an external file. This file might be a CSV file or a spreadsheet.

REWIND

A **comma-separated values (CSV) file** is a text file that has the `.csv` file name extension and that follows a specific format for storing a table. For example, if we want to store a table that has two columns named `Apple` and `Google`, we can do so like this:

```
Apple,Google
0.50,0.45
1.34,0.63
0.62,0.55
-0.15,00.04
```

Each line in the CSV file corresponds to a row in the table. And in each row, a comma separates each column of data from the next. Both Pandas and spreadsheet software can read and interpret CSV files.

Now that you've learned what data containers are, you'll next learn about collecting data from a CSV file.