

[◀ Back to Week 3](#)[✕ Lessons](#)[Prev](#)[Next](#)

A comma-separated values (CSV) file is a plain text file that is used to store tabular data. Each line of the file is one row of the data. A row consists of fields that are typically separated by commas.

For example, consider the following table of numbers:

32	5	17
8	62	3
14	9	7

This could be stored in a CSV file with the following contents:

```
1 32,5,17
2 8,62,3
3 14,9,7
```

Different programs can then read and write this data while storing it internally in some more convenient format. CSV files are often used to share data among database and spreadsheet programs that do not share a common internal format, for example.

Format

Unfortunately, the format of a CSV file is not well specified. Problems arise because you might want to have commas in the data. Imagine, for example, that one of the fields is a person's name in **Last, First** format. The comma between the last and first names should remain in the field, not be interpreted as a separator between fields. So, even though they are called **comma-separated values**, delimiters other than commas can be used to separate fields (such as a space, semicolon, tab, etc.). There is no universally adhered to standard.

Another way to solve some of the problems is to use quotes around strings that should be in a single field. Then the delimiter is not treated as a separator when it is within a quoted string. But, then you have the same problem when you need to have the quote character within a field.

Generally, CSV files, therefore, can use just about any character as the field separator (or delimiter) and quote character. The file itself does not specify which characters are being used. So, the user must either know ahead of time or look at the first few lines of the file and try to figure it out. So, you might see a "CSV" file that looks like this:

```
1 "Apples;Oranges";3;7
2 "Pears";8;13
```

In this case, you can guess that " is the quote character and ; is the field separator. Spaces around the field separator are generally ignored.

As CSV files are used to store tabular data, you should expect that each line will have the same number of fields. Sometimes, though, some fields may be empty. You would usually see this as follows:

```
1 32,5,17
2 8,,
3 14,,7
```

In this case, you can see that the the second and third fields of the second row are empty and the second field of the third row is empty. But, you may also see rows with different numbers of fields. Typically, this would be interpreted as the final fields that are missing being empty.

Expectations

Most well-formed CSV files adhere to the following "rules":

1. Fields are separated by a single delimiter character, which is often a comma (,).
2. Rows are separated by a newline character.
3. Fields are interpreted as plain text.
4. Fields can be quoted by a quote character, which is often a double quote (").
5. Quoted fields can contain the delimiter character and/or newlines within them.
6. Each row contains the same number of fields in the same order.

If a CSV file follows these "rules", it is easier to use and interpret (as long as you know which characters are being used as the delimiter and quote characters). Note that CSV files are not required to follow these conventions. While most files will, there are some programs that do not adhere to these rules and follow their own conventions.

If a CSV file does not follow these rules, you often need to fix the files manually by editing them yourself to conform to some set of rules that the program you are using can handle.

Field Labels

Sometimes the first row of a CSV file will be a header row that contains the names of the fields instead of data:

```
1 "Last name","First name","Title"
2 "Rixner","Scott","Professor"
3 "Warren","Joe","Professor"
```

Mark as completed

