

3.3.1

Import and Inspect the Data

Before stopping for the day, you take a look at the dataset again with Tom. You want to be familiar with the data before you write any code. Seth tells you to download the dataset and save it in your Election_Analysis folder, and then go see Tom to inspect the data.

Import the Data

The first step is to import the data. In your Election_Analysis folder, create a subfolder named "Resources." This is where you'll save the dataset.

Click the following link to download the dataset into the Resources folder.

[Download the dataset](https://2u-data-curriculum-team.s3.amazonaws.com/dataviz-online/v2/module_3/election_results.csv) [_\(https://2u-data-curriculum-team.s3.amazonaws.com/dataviz-online/v2/module_3/election_results.csv\)](https://2u-data-curriculum-team.s3.amazonaws.com/dataviz-online/v2/module_3/election_results.csv)

When the download is complete, you should see a file named `election_results.csv` in the Resources folder.

Inspect the Data

A **CSV file** is one of the most common data formats that is used with large datasets. When data is exported from websites or databases, they are primarily exported in the form of a CSV file.

CSV stands for comma-separated value. This means that each value is separated by a comma. If you open `election_results.csv` with a text editor like VS Code, you'll see that each row, or value, is separated by a comma.

```
Ballot ID, County, Candidate
1009260, Jefferson, Charles Casper Stockham
1030085, Jefferson, Charles Casper Stockham
1031522, Jefferson, Charles Casper Stockham
1037115, Jefferson, Charles Casper Stockham
1035173, Jefferson, Charles Casper Stockham
1032693, Jefferson, Charles Casper Stockham
1030087, Jefferson, Charles Casper Stockham
1014984, Jefferson, Charles Casper Stockham
1034482, Jefferson, Charles Casper Stockham
1021108, Jefferson, Charles Casper Stockham
1015868, Jefferson, Charles Casper Stockham
1006093, Jefferson, Charles Casper Stockham
1016857, Jefferson, Charles Casper Stockham
```

Before we do any data analysis, we need to inspect the data.



REWIND

Remember, when inspecting the data focus on the following questions.

- How many columns and rows are there?
- What types of data are present?
- Is the data readable, or does it need to be converted in some way?

Open `election_results.csv` in Excel so you can see the number of columns and rows. A text editor will not give this level of detail. You may notice that this file looks very similar to an Excel file with an `.xlsx` extension. You should also notice that there are three headers: Ballot ID, County, and Candidate. Here are the first 10 rows:

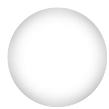
	A	B	C
1	Ballot ID	County	Candidate
2	1323913	Jefferson	Charles Casper Stockham
3	1005842	Jefferson	Charles Casper Stockham
4	1880345	Jefferson	Charles Casper Stockham
5	1600337	Jefferson	Charles Casper Stockham
6	1835994	Jefferson	Charles Casper Stockham
7	1772756	Jefferson	Charles Casper Stockham
8	1920023	Jefferson	Charles Casper Stockham
9	1040408	Jefferson	Charles Casper Stockham
10	1018412	Jefferson	Charles Casper Stockham

NOTE

Sometimes CSV files can have **headers**. Headers are essentially a way to label each column of data in the file.

If we scroll down to the last row, you'll see there are 369,712 rows.

369703	4762851	Arapahoe	Raymon Anthony Doane
369704	4768093	Arapahoe	Raymon Anthony Doane
369705	4196905	Arapahoe	Raymon Anthony Doane
369706	4299985	Arapahoe	Raymon Anthony Doane
369707	4620283	Arapahoe	Raymon Anthony Doane
369708	4714953	Arapahoe	Raymon Anthony Doane
369709	4497542	Arapahoe	Raymon Anthony Doane
369710	4085849	Arapahoe	Raymon Anthony Doane
369711	4592018	Arapahoe	Raymon Anthony Doane
369712	4660518	Arapahoe	Raymon Anthony Doane



REWIND

There's a keyboard shortcut to get to the last row of an Excel file so that you don't have to keep scrolling. Place the cursor in a column that doesn't contain any empty cells and press Command and the down arrow key (for macOS) or CTRL and the down arrow key (for Windows).

The data consists of a number for the ballot ID and a name for the county and candidate, respectively. At this point, we don't know how many counties and candidates are in the file. Overall, the data appears readable and there are no unusual row values that we can see—but let's use Python to check anyway.

© 2020 - 2022 Trilogy Education Services, a 2U, Inc. brand. All Rights Reserved.