Working with .csv files

In an earlier lesson, you explored comma separated values (.csv) files and how to import them into a Google Sheet or download them from a website. Data analysts use .csv files often, so throughout this course you will use .csv files to transfer data into data analysis programs for further analysis and visualization. .csv files are plain text files with an organized table structure that includes rows and columns. The values in each row are separated by commas. The table structure makes .csv files easy to understand, edit, manipulate, and use for data analysis.
A major advantage of .csv files is their widespread compatibility. They can be imported and exported by a vast range of data analysis tools and software programs.

Download .csv files

To use .csv files and upload them to data analysis programs you will first need to download them to your local device. Downloading a .csv file from a website can vary depending on your operating system or internet browser. Here are some ways you can download a .csv file:

- **Click the download link or .csv attachment:** Locate the link for the .csv file or attachment on the website. Click on it, and the download process will start.
- **Right-click and Save: Right-click** on the data table or element containing the .csv data. Choose **Save as…** or a similar option. Name the file and make sure the extension on the file is ".csv".
- **Force download:** You can use the **Alt** key on your keyboard while clicking the link. This will trigger the download, and you will be able to find the .csv file in your Downloads folder.

Note: When using the Chrome browser or ChromeOS, .csv files may open in a new tab instead of downloading to your machine. If this happens, follow these instructions:

- Select **File** from the menu bar, then select **Save as Google Sheets**. This will open the .csv file as a Google Sheet.
- Select **File** from the menu bar, then select **Download** from the dropdown menu, then select **Comma Separated Values (.csv)**.

Upload .csv files

You will often need to upload .csv files during the data analysis process. Here is how you do this:

- **Locate the upload option:** Each data analysis platform will have a designated button, menu option, or drag and drop area labeled **Upload** or **Import**. This is where you will upload your .csv file.
- **Choose your .csv file:** Click **Upload** or **Import** on the platform you are using to open your file explorer. **Select** your .csv file. If you just downloaded a .csv file from the web, it will be located in your computer's **Downloads** folder.
- **Initiate the upload:** Once you've selected your .csv file, click **Upload** or **Import.**The platform may display a progress bar or message indicating that the upload is complete.

Note: Some platforms have restrictions on the file size or format of .csv files. Make sure your .csv files adhere to these requirements before uploading.

Key takeaways

Data analysis programs help us extract insights and knowledge from data. Using .csv files is essential in data analysis. Understanding how to easily download data from the web or add your data to these programs will allow you to complete data cleaning, visualizations, analysis, and so much more!

Step-by-Step: Import data from spreadsheets and databases

This reading outlines the steps the instructor performs in the following video, Importing data from spreadsheets and databases. The video teaches you how to import a .csv file into a Google Sheet so you can analyze the data.
Keep this guide open as you watch the video. It can serve as a reference if you need additional context or clarification while following the video steps. This is not a graded activity, but you can complete these steps to practice the skills demonstrated in the video.

What you'll need

To follow along with the examples in this video, open a blank spreadsheet. Additionally, download the following .csv file:
Example Spreadsheet - Entertainment Expenses - Sheet1
CSV File

Example 1: Use the menu to import a .csv file

Sometimes, you'll need to import the data from a .csv file into a Google Sheet. A .csv file saves data in a table format. Follow the steps below to bring the data from a .csv file into a new spreadsheet.
Note: The .csv file you import will differ from the .csv file the instructor imports in the video.

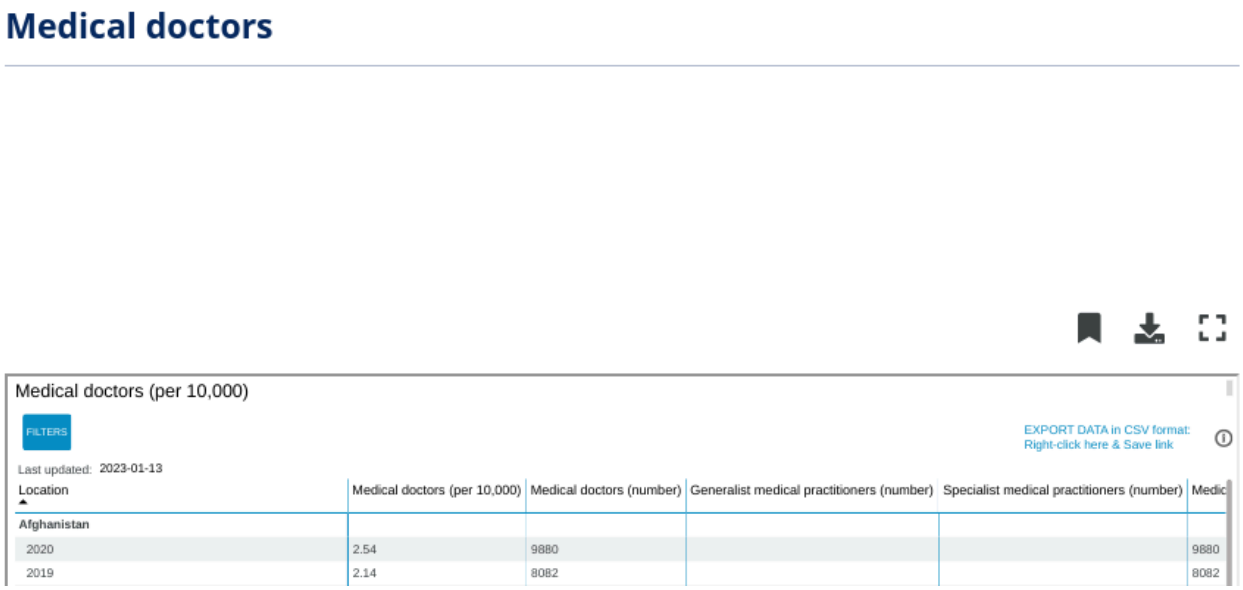
1. In the menu, select **File** then **Import**. The **Import file window** will pop up.
2. Select **Upload** then **Browse** to select the .csv file to import. For this example, select the **.csv** file you downloaded in the previous section
3. Next select the **Import location**. You can:
    a. Create a new spreadsheet
    b. Insert the .csv data as a new sheet
    c. Replace spreadsheet
    d. Replace current sheet
    e. Append (add) the data to the current spreadsheet
    f. Replace the data starting with a specific cell.
4. Next, select the **Separator type**. Google Sheets defaults to automatically detecting separator, or delimiter, type. To manually set the delimiter type, select the dropdown menu under Separator type and choose the separator.
5. Next, determine if you would like the text to be imported with or without formatting. In the box next to **Convert text to numbers, dates, and formulas**, keep the checkmark if you want text data to be formatted.
6. Select **Import data**. The data in the .csv file will be loaded into your sheet, and you can begin using it.

Example 2: Download data from the Global Health Observatory

You can download public data from the internet, such as data from the World Health Organization and the Global Health Observatory.
Note: The Global Health Observatory's website has been updated since this video was filmed. Follow these instructions to download the .csv file the instructor uses.

1. Navigate to the Global Health Observatory workforce statistics database.
2. Scroll to navigate to the **Medical doctors** table.
3. Then, scroll over **EXPORT DATA in .csv format** in the table. Right-click, then select **Save link as...**
4. The .csv file will download to your computer as **data.csv**.

Note: If you already have a .csv file named **data.csv**, your computer will add a number to the file name.



Medical doctors (number), General medical practitioners (number), and Specialist medical practitioners (number). The EXPORT DATA in CSV format is also in the screenshot

Example 3: Use the menu to import a .csv file

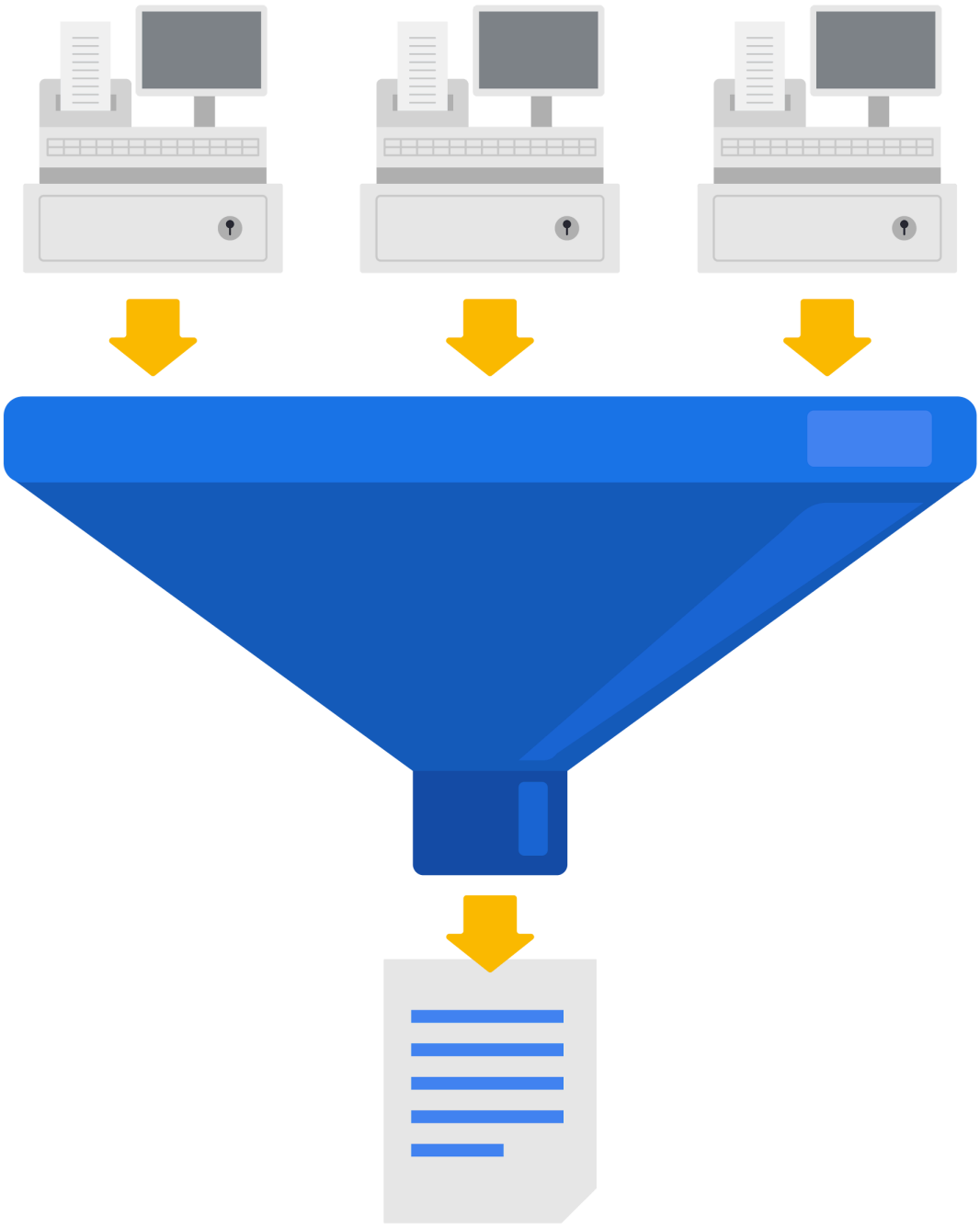
Follow the steps below to bring the .csv file you downloaded, **data.csv**, into a new spreadsheet.

1. Open a blank spreadsheet.
2. In the menu, select **File**, then **Import**. The **Import file window** will pop up.
3. Select **Upload** then **Browse** to select the .csv file to import.
4. Next, select the **Import location**. You can:
    a. Create a new spreadsheet
    b. Insert the .csv data as a new sheet
    c. Replace spreadsheet
    d. Replace current sheet
    e. Append (add) the data to the current spreadsheet
    f. Replace the data starting with a specific cell.
5. Next, select the **Separator type**.
6. Next, determine if you would like the text to be imported with or without formatting. In the box next to **Convert text to numbers, dates, and formulas**, keep the checkmark if you want text data to be formatted.
7. Select **Import data**.
8. The data in the .csv file will be loaded into your sheet, and you can review and title it.

Import data dynamically

As you've learned, you can import data from some data sources, like .csv files into a Google spreadsheet from the **File** menu. Keep in mind that, when you use this method, data that is updated in the .csv will not automatically be updated in the Google Sheet. Instead, it will need to be manually—and continually—updated in the Google Sheet. In some situations, such as when you want to be able to keep track of changes you've made, this method is ideal. In other situations, you might need to keep the data the same in both places, and using data that doesn't update automatically can be time-consuming and tedious. Further, trying to maintain the same dataset in multiple places can cause errors later on.

Fortunately, there are tools to help you automate data imports so you don't need to continually update the data in your current spreadsheet. Take a small general store as an example. The store has three cash registers handled by three clerks. At the end of each day, the owner wants to determine the total sales and the amount of cash in each register. Each clerk is responsible for counting their money and entering their sales total into a spreadsheet. The owner has the spreadsheets set up to import each clerks' data into another spreadsheet, where it automates and calculates the total sales for all three registers. Without this automation, each clerk would have to take turns entering their data into the owner's spreadsheet. This is an example of a dynamic method of importing data, which saves the owner and clerks time and energy. When data is dynamic, it is interactive and automatically changes and updates over time.



In the following sections you'll learn how to import data into Google Sheets dynamically.

`IMPORT` functions in Google Sheets

The `IMPORTRANGE` function

In Google Sheets, the **IMPORTRANGE** function can import all or part of a dataset from another Google Sheet.
To use this function, you need two pieces of information:

1. The URL of the Google Sheet from which you'll import data.
2. The name of the sheet and the range of cells you want to import into your Google Sheet.

Once you have this information, open the Google Sheet into which you want to import data and select the cell into which the first cell of data should be copied. Enter = to indicate you will enter a function, then complete the **IMPORTRANGE** function with the URL and range you identified in the following manner: `=IMPORTRANGE("URL", "sheet_name!cell_range")`. Note that an exclamation point separates the sheet name and the cell range in the second part of this function.
An example of this function is:
`=IMPORTRANGE("https://docs.google.com/thisisatestabc123", "sheet1!A1:F13")`
Note: This URL is for syntax purposes only. It is not meant to be entered into your own spreadsheet.
Once you've completed the function, a box will pop up to prompt you to allow access to the Google Sheet from which you're importing data. You must allow access to the spreadsheet containing the data the first time you import it into Google Sheets. Replace it with a spreadsheet's URL that you have created so you can control access by selecting the Allow access button.
Refer to the Google Help Center's IMPORTRANGE page for more information about the syntax. You'll also learn more about this later in the program.

The IMPORTHTML function

Importing HTML tables is a basic method to extract data from public web pages. This process is often called "scraping." Web scraping made easy introduces how to do this with Google Sheets or Microsoft Excel.
In Google Sheets, you can use the **IMPORTHTML** function to import the data from an HTML table (or list) on a web page. This function is similar to the **IMPORTRANGE** function. Refer to the Google Help Center's IMPORTHTML page for more information about the syntax.

The IMPORTDATA function

Sometimes data displayed on the web is in the form of a comma- or tab-delimited file.
You can use the **IMPORTDATA** function in a Google Sheet to import data into a Google Sheet. This function is similar to the **IMPORTRANGE** function. Refer to Google Help Center's IMPORTDATA page for more information and the syntax.

Explore public datasets

Open data helps create a lot of **public datasets** that you can access to make data-driven decisions. Here are some resources you can use to start searching for public datasets on your own:

- The Google Cloud Public Datasets allow data analysts access to high-demand public datasets, and make it easy to uncover insights in the cloud.
- The Dataset Search can help you find available datasets online with keyword searches.
- Kaggle has an Open Data search function that can help you find datasets to practice with.
- Finally, BigQuery hosts 150+ public datasets you can access and use.

Public health datasets

1. Global Health Observatory data: You can search for datasets from this page or explore featured data collections from the World Health Organization.
2. The Cancer Imaging Archive (TCIA) dataset: Just like the earlier dataset, this data is hosted by the Google Cloud Public Datasets and can be uploaded to BigQuery.
3. 1000 Genomes: This is another dataset from the Google Cloud Public resources that can be uploaded to BigQuery.

Public climate datasets

1. National Climatic Data Center: The NCDC Quick Links page has a selection of datasets you can explore.
2. NOAA Public Dataset Gallery: The NOAA Public Dataset Gallery contains a searchable collection of public datasets.

Public social-political datasets

1. UNICEF State of the World's Children: This dataset from UNICEF includes a collection of tables that can be downloaded.
2. CPS Labor Force Statistics: This page contains links to several available datasets that you can explore.
3. The Stanford Open Policing Project: This dataset can be downloaded as a .csv file for your own use.