

## Getting Started with Google Colab

All of the course modules that include hands-on work with Python, SQL, or the R programming language use [Jupyter notebooks](#). To avoid each student needing to install Jupyter on their computer, we are using Jupyter notebooks hosted by Google, as part of [Google Colaboratory](#) (also known as *Google Colab*).

You will need to use the online version of this document so you can follow the links included in it. Jupyter notebooks are used in the course modules *Python for Data Analysis and Visualization*, *Relational Databases and Basic SQL*, *Advanced SQL*, *Machine Learning Using Python*, *Data Mining Using Python*, *Data Mining Using SQL*, *The R Language*, *Network Analysis*, and *Unstructured Data*.

The instructions in this document should be sufficient to get you started, but the end of the document includes links to more information and tutorials for the tools you will be using.

### Step 0. Check your browser

In our experience, Google Colab can be problematic on the *Firefox* browser, but works fine on all others. So please avoid Firefox. Additionally, if you are having mysterious problems with notebooks not opening properly, make sure that you are allowing browser cookies, and that you are not in *Incognito* mode.

### Step 1. Create a Google account

If you already have a Google/Gmail account, proceed to Step 2. Otherwise:

1. Go to the [Google Accounts website](#)
2. Click on the *Create account* link near the bottom of the page
3. Fill in your information on the form on the right side of the page and complete the other required steps. Congratulations, you now have a Google account!

### Step 2. Copy notebooks and data files into your Google Drive file-space

Every Google account owner is entitled to use Google Drive, a file storage service on the cloud, and to use Google Colab to run Jupyter notebooks. You will store and run Jupyter notebooks in your Google Drive file-space. The data files accessed by Jupyter notebooks will need to be stored on your computer; more on that in Step 3.

This step copies the Jupyter notebooks and data files needed for your course module(s) into your file-space.

1. Make sure you're logged into your Google account.
2. Follow one of the following links to go to the public folder containing the notebooks and data files for the course module(s) you are doing:
  - [Basic and Advanced SQL](#)
  - [Python for Data Analysis and Visualization](#)
  - [Data Mining Using Python and SQL](#)
  - [Machine Learning Using Python](#)

- [The R Language](#)
  - [Network Analysis](#)
  - [Unstructured Data](#)
3. Select all of the files by pressing command+a on a Mac or control+a on Windows. Right-click (or control-click) on one of the files to get a drop-down menu and select *Make a copy*.
  4. Go to *My Drive* in the upper-left below the *New* button. You should now be in your own filespace, where you should see the set of files you copied with names starting with *Copy of*. You may want to restore the name of each file to its original, and you will need to do so for the CSV files: right-click (or control-click) on the file to get a drop-down menu that includes *Rename*.

### Step 3. Copy data files onto your computer

The files for each course module include one or more Jupyter notebooks (with extension *.ipynb*) and one or more CSV data files (with extension *.csv*). You will be running the Jupyter notebooks from your Google account, but the CSV files must be stored on your computer.

Once you've copied all of the files for a course module, each file ending in *.csv* will need to be renamed to remove *Copy of*, then downloaded to your computer. (**Important:** If you neglect to remove *Copy of* from the filename, the Jupyter notebooks will not be able to correctly access the file.) Once the file is renamed, right-click (or control-click) on each file and select *Download*. Depending on your system, you may need to drag the downloaded file from the browser window to your computer desktop, or to another convenient location on your computer's file system. *Make sure the downloaded files keep the extension .csv. Depending on the settings, some computers might automatically convert CSV files to extension .xlsx.*

Congratulations, you now have your files set up to run Jupyter notebooks!

**A note on file organization:** If you will be participating in more than one course module, you might decide to create a separate folder on your Google Drive for each one: use the *New* button in the upper-left to create a folder; files can be moved by simply dragging them onto the folder. You will notice that many of the same CSV files are used in more than one module. You don't need to keep separate copies of them on your computer: *A dedicated folder with one copy of each CSV file on your computer can be shared by all of the Jupyter notebooks, and is the most convenient way of organizing your local CSV files.*

### Step 4. Try launching a Jupyter notebook

1. As in Step 2, make sure you're logged into your Google account, and go to your Google Drive page. If needed, navigate to the folder containing the files you will be working with.
2. To launch one of the Jupyter notebooks you copied for the course using Google Colab, simply double-click on the file. Congratulations, you are now running a Jupyter notebook! Professor Widom will guide you in using Jupyter notebooks for your course module.

## Step 4b. Working offline

Unlike Google Sheets, it is not possible to run Jupyter notebooks using Google Colab when you are disconnected from the internet. Jupyter notebooks execute code using Google's cloud computing infrastructure, which is only accessible when online.

### References for Jupyter notebooks

- [Project Jupyter home page](#)
- [Google Colab home page](#)

### References for Python Language

- [History of Python](#)
- [Comparing Python to Other Languages](#)
- [Compiled vs. interpreted languages](#)
- [Interactive Python tutorial](#)
- [Learn Python in 10 minutes](#)
- [Python official documentation](#)

### References for Python Packages

- [Pyplot tutorial](#)
- [Python Data Analysis Library - pandas](#)
- [Pandas Tutorials](#)
- [An Introduction to Scientific Python: Pandas](#)
- [Machine Learning Algorithm Recipes in scikit-learn](#)
- [scikit-learn: KNeighborsClassifier](#)
- [scikit-learn: DecisionTreeClassifier](#)
- [scikit-learn: Naive Bayes](#)
- [scikit-learn: KMeans](#)
- [Networkx Tutorial](#)
- [Python - Text Processing Introduction](#)
- [Python - Regular Expressions Practical Guide](#)
- [Python re - Regular expression operations](#)
- [NLTK Tutorials](#)
- [Python Imaging Library \(PIL/Pillow\) - The Image Module](#)

### References for SQL

- [SQL Tutorial](#) (*Tutorials Point*)
- [SQL Tutorial](#) (*w3schools*)
- [SQLite official documentation](#)
- [SQLite SQL syntax](#)

### References for R Language

- [R Tutorial](#)
- [Quick-R: accessing the power of R](#)