

Setting up Your Computer to Use My Python Code for Downloading Twitter Data

I frequently get requests for how to download social media data in general, as well as for help on how to run code I have written to download and analyze the data I analyzed for a particular piece of research. Often, these requests are from people who are excited about doing social media research but have yet to gain much experience in using computer programming. For this reason, I have created a [set of tutorials](#) designed precisely for such users.

I am always happy to share the code I've used in my research. That said, there are barriers to actually using someone else's code. One of the key barriers is getting your own computer "set up" to actually run the code. The aim of this post is to walk you through the steps needed to run and modify the code I've written to download and analyze social media data.

Step One: Download and Install Python

As I write about [here](#), for Unix, Windows, and Mac users alike I'd recommend you install [Anaconda Python 2.7](#). This distribution of Python is free and easy to install. Moreover, it includes most of the add-on packages necessary for scientific computing, including Numpy, Pandas, iPython, Statsmodels, SQLAlchemy, and Matplotlib.

Go to [this tutorial](#) for instructions on how to install and run Anaconda Python.

Step Two: Install Python Add-On Packages

Anaconda Python comes pre-installed with *almost* everything you need. There are a couple of modules you will have to install

manually:

Twython – for accessing the Twitter data

and

simplejson – for parsing the [JSON](#) data that is returned by the Twitter [API](#) (Application Programming Interface).

Assuming you are on a Mac and using Anaconda Python, the simplest way is to use [pip](#). On a Mac or Linux machine, you would simply open the Terminal and type *pip install Twython* and *pip install simplejson*. If you're on a PC, please take a look at [Wayne Xu's tutorial](#) (see Slide #8).

Step Three: The Database

I generally download my Twitter data into an SQLite database. SQLite is a common relational database. It is lightweight and easy to use, and comes preinstalled in Anaconda Python.

You may already know other ways of downloading social media data, such as NodeXL in Excel for Windows. Why then would you want to using SQLite? There are two reasons. First, SQLite is better plugged into the Python architecture, which will come in handy when you are ready to actually manipulate and analyze the data. Second, if you are downloading tweets more than once, a database is the much better solution for a simple reason: it allows you to write a check for duplicates in the database and stop them from being inserted. This is an almost essential feature that you cannot easily implement outside of a database.

Also know that once you have downloaded the data into an SQLite database, you can view and edit the data in the same manner as an Excel file, and even export the data into CSV format for viewing in Excel. To do this, simply download and install [Database Browser for SQLite](#). If you use Firefox, you can alternatively use a plug-in called [SQLite Manager](#).

Step Four: Accessing the Twitter API

Almost all of my Twitter code grabs data from the Twitter API, which sets up procedures for reliably accessing the Twitter data. Beginning in 2013 Twitter made it more difficult to access its APIs. Now [OAuth authentication](#) is needed for almost everything. This means you need to go on Twitter and create an 'app.' You won't actually use the app for anything – you just need the password and authentication code. You can [create your app here](#). For more detailed instructions on creating the app take a look at slides 4 through 6 of Wayne Xu's (my excellent former PhD student) tutorial [tutorial](#).

Step Five: Start Using the Code

Once you've completed the above four steps you will have all the necessary building blocks for successfully running [my Python scripts](#) for downloading and parsing Twitter data. Happy coding!