

Do I Need to Learn Programming to Download Big Data?



You want to download and analyze “Big Data” – such as messages or network data from Twitter or Facebook or Instagram. But you’ve never done it before, and you’re wondering, “Do I need to learn computer programming?” Here are some decision rules, laid out in the form of brief case studies.

One-Shot Download with Limited Analysis

Let’s say you have one organization you’re interested in studying on Twitter and want to download all of its tweets.

You are doing only basic analyses in a spreadsheet like Excel. In this case, if you have a PC, you can likely get away with something like NodeXL – an add-on to Excel. **VERDICT: COMPUTER PROGRAMMING LIKELY NOT NECESSARY**

One-Shot Download with Analysis in Other Software

Let's start with the same data needs as above: a one-shot download from one (or several) organizations on Twitter. You wish to undertake extensive analyses of the data but can rely on some other software to handle the heavy lifting – maybe a qualitative analysis tool such as ATLAS or statistical software such as SAS, R, or Stata. Each of those tools has its own programming capabilities, so if you're proficient in one of those tools – and your data-gathering needs are relatively straightforward – you *might* be able to get away with not learning programming. **VERDICT: COMPUTER PROGRAMMING MAY BE UNNECESSARY**

Anything Else

In almost any other situation, I would recommend learning a programming language. Why is this necessary? For one case, let's say you wish to download tweets for a given hashtag over the course of an event. In this case you'll want to use a database – even a simple database like SQLite – to avert duplicates from being downloaded. The programming language, meanwhile, helps you download the tweets and “talk” to the database. In short, if you are downloading tweets *more than once* for the same sample of organizations, you should probably jump to learning a programming language. Similarly, if you have any need at all for *manipulating the data* you download – merging, annotating, reformulating, adding new variables, collapsing by time or organization, etc. – then a programming language becomes highly desirable. Finally, if you have any interest in or need of medium- to advanced-level *analysis of*

the data, then a programming language is similarly highly desirable. **VERDICT: PICK A PROGRAMMING LANGUAGE AND LEARN IT**

Conclusion

Not *everyone* needs to learn a programming language to accomplish their social media data downloading objectives. If your needs fall into one of the simple cases noted above then you may wish to skip it and focus on other things. On the other hand, if you are going to be doing data downloads again in the future, or if you have anything beyond basic downloading needs, or if you want to tap into sophisticated data manipulation and data analysis capabilities, then you should seriously consider learning to program.

Learning a programming language is a challenge. Of that there is little doubt. Yet the payoff in improved productivity alone can be substantial. Add to that the powerful analytical and data visualization capabilities that open up to the researcher who is skilled in a programming language. Lastly, leaving aside the buzzword “Big Data,” programming opens up a world of new data found on websites, social media platforms, and online data repositories. I would thus go so far as to say that any researcher interested in social media is doing themselves a great disservice by not learning some programming. For this very reason, one of my goals on this site is to provide guidance to those who are interested in getting up and running on Python for conducting academic and social media research. If you are a beginner, I’d recommend you work through [the tutorials listed here](#) in order.