Python Installation Instructions

December 10, 2016

Lab on Non-Democratic Politics
University of Southern California

Before the training session, please go through the following steps. They should take about an hour. If you get stuck, stop by our office; it's far easier to work these things out in person.

1) Download Python: https://www.python.org/downloads/

It is *very* important that you download Python 2 rather than Python 3. Although 3 is newer, 2 has better modules for text analysis.

2) Download Eclipse. This is my favorite editor for writing Python code because it plays nice with non-English languages.

https://www.eclipse.org/downloads/download.php?file=/oomph/epp/neon/R1/eclipse-inst-mac64.tar.gz

3) Set up Eclipse for Python. You can follow this guide:

https://www.ics.uci.edu/~pattis/common/handouts/pythoneclipsejava/eclipsepython.html

4) Install easy_install. Go to Terminal, and type the following:

```
curl https://bootstrap.pypa.io/ez_setup.py -o - | sudo python
```

If you are prompted for your password, type it in and press enter.

5) Install pip. Go to Terminal, and type the following:

```
sudo easy_install pip
```

If you are prompted for your password, type it in and press enter.

6) Pip is a downloader, so let's use it to download some Python modules that will allow us to do nifty things. In Terminal, type the following:

```
pip install codecs
pip install datetime
pip install selenium
pip install nltk
```

Great! Now we can write out files with non-English script, use dates, and animate browsers. NLTK provides support for a bunch of different text analysis methods.

7) Next, we'll be using Firefox to scrape websites. Download yourself a nice vintage copy of it here (macs only; windows and linux navigate from /45.0/):

https://ftp.mozilla.org/pub/firefox/releases/45.0/mac/en-US/

8) Now the hard stuff is done! If you want to familiarize yourself with Python, you can consult this guide, which we'll draw from in the training session.

http://web.stanford.edu/~zlotnick/teaching.html

Also, check out Stack Overflow. If you google "how do I X stackoverflow python", you will always find the answer.

9) Familiarize yourself with the NYT archive, which we will scrape.

http://query.nytimes.com/search/sitesearch/

10) Choose a term you're interested in scraping from the NYT archive, like "Russia" or "election" or "Caroline." In future sessions, we'll learn how to create time series data about your term. Then, we'll cover how to run regressions in R and to produce pretty pictures.

11) Start thinking about long-term projects. We're scraping the NYT archive because it's perfect for learning about lots of things you typically encounter when scraping websites. If there is a website you'd like to scrape for a project or independent study (e.g. Russian news portals, Lisa =)), keep us in the loop about this.