Python for Social Science

Ling 596: Special Topics
Spring, 2014: No prerequisites
gawron@mail.sdsu.edu
TuTh 1400-1515 PSFA 413

Course description

Increasingly, social scientists find themselves facing exponentially larger data sets available on the internet and elsewhere without suitable tools to deal with them. Many social scientists end up using spreadsheet programs for their data-processing tasks and spend hours clicking around or copying and pasting, and then repeating the process for other data files.

This course will show you how to use computing tools freely available via the scripting language Python to use your data more powerfully and effectively. The course touches on many topics of theoretical interest in the emerging field of data science, such as social networks and data visualization, but the focus is on manipulating data so that you can tailor it to the needs of your particular project. The course targets social science students and will assume no prior programming knowledge. Although many of the techniques are relevant to linguistics, economics, and geography, the course focuses on techniques that are applicable to a wide range of data sources, including images, social network data, web pages, blogs. [Course website: http://www-rohan.sdsu.edu/~gawron/python_for_ss]

Sample Topics

1. Python Basics
2. Text data (regular expressions) and web data (Blogs/Twitter)
3. Extracting data sets from big data sources (e.g., census data)
4. Social networks (visualizing social groups)
5. Connecting to your stat package (data frames)
6. Using geocoders, Google maps, and Google earth (online data sources)

Example visualizations

An online book discussion
Visualizing House votes
The Enron Email network
Zachary’s karate club network