Introduction to Python

John Poe
jdpoe223@gmail.com

Course Description

Python has become one of the most popular languages for statistics and machine learning on the planet. Python's flexibility to handle virtually any kind of data combined with the truly astounding range of analytical techniques that it offers means that it provides enormous value to academics. Despite this fact, it is rarely if ever taught to people in the social and behavioral sciences or public policy. Most courses and textbooks that do teach Python at an introductory level are designed by people with computer science backgrounds who tend to hide the value Python provides for the social sciences.

In this 4-day introductory workshop, we will highlight some of the more astoundingly impressive and useful tools that Python offers that are of particular use to social scientists. Note that this is not the standard list of libraries that one would learn in an introductory workshop on Python. It is designed explicitly for the typical ICPSR audience and so places less emphasis on most of the computer science-oriented libraries. This workshop cannot make you an expert in Python. It can and will give you a sense of some of the tools Python makes available and how to learn more about them.

Given the size of this workshop it is not possible to help each participant with software installation individually. Therefore, it is not required that you follow along with the material on your own machine. You will be given copies of all scripts and output as well as data through a Dropbox link and can follow along with an html version of the script and output.

Below is a preliminary list of topics to be covered each day. This list is subject to change depending on how quickly we go through material.

You can find course materials here
Monday: Basics of Python and Anaconda

• Overview of the workshop
• Installing Python 3 and the Anaconda Build
• The basics of Jupyter notebooks for Python
• A brief introduction to pandas

Tuesday: Basics Statistical Modeling in Python

• Data management and graphics
• Basic linear modeling
• Advanced & generalized linear modeling

Wednesday: Scikit-learn

• A basic introduction
• Linear modeling
• Classification
• Text Analysis

Thursday: Integration with Other Software

• Python and R
• Python and Stata