Python Workshop

When: May 21, 2019 (5 - 7 PM)

Where: 33 Oxford St. (Maxwell-Dworkin), G125

Who is this for: This is an introductory tutorial on working with data. This is aimed primarily at Harvard researchers and students who are already familiar with python but want to learn about handling data in python.

What are the topics:

This tutorial will be an introduction to working with data in python via the numpy/pandas modules. Some of the topics we might cover include

- Different ways of working with data in Python
- Data handling via dataframes in Pandas

(Material for this tutorial at the end of this page)

Python installation and Jupyter Notebook on your laptop:

If you want to work through the tutorial material during the workshop, bring your laptop with Anaconda version of python installed.

Please visit the link below for Anaconda installation:
https://www.anaconda.com/distribution/

Suggestions:

- Choose Python 3 (3.6 as of this writing)
- On Mac, if the graphical installer does not work, choose the command line installer.

Running Jupyter Notebook:

Helpful links on Anaconda, ipython notebooks etc:

https://docs.anaconda.com/anaconda/user-guide/faq/
http://nbviewer.ipython.org/github/catherinedevlin/mpfwf_exercises/blob/master/setup.windows.ipynb

Tutorial Materials

The rest of this page assumes you have installed Anaconda and the various python binaries are available in a terminal (Mac and Linux) or command prompt (Windows). Download the Ipython Notebooks below and put them in the same directory.

On Windows, the simplest option is to put these files in the “Ipython Notebooks” directory in the “My Documents” directory under “Documents” (i.e Documents --> My Documents --> IPython Notebooks).

You can open Ipython Notebook on various OSs as follows:

On windows, start the ipython notebook using the launcher under Anaconda in the Start Menu.

On Mac:

Double click on the launcher (should be available on the desktop) and choose ipython notebook. In the notebook, navigate to the folder which contains the tutorial notebooks.

On Linux, open a terminal (and on Windows, open a command prompt) and change to the directory where your notebooks are. Then type:
jupyter notebook

from the directory where you have all the following files (you need to have Anaconda bin directory in the path). Once the jupyter notebook server and the browser are up, you will see the files with 'ipynb' extension in the dashboard. Clicking on it will open it.

Ipython (jupyter) Notebooks and data for this tutorial:

introduction-to-pandas.ipynb

python_tutorial_multiple_events.ipynb

total_petroleum_consumption_thousand_barrels_per_day.csv

multiple_events.csv

The following notebook with sqlite/sql to pandas interface is optional:

nyc_complaints_sql.ipynb

Data for the above notebook (a small chunk of original 2009 data)

311_Service_Requests_for_2009_part_01.csv

Tutorial assumes you are familiar with basic python. You can find material on the web (just google) on introductory python. You may also look at the basic python tutorial in the notebook:

python-basics-python3-02012018.ipynb

custom.css — for changing the appearance of the notebook (optional)