ECON 370 Quantitative Economics with Python

Lecture: Working with Data in Python

Zhen Huo [zh335@nyu.edu]
Matthew McKay [mm8390@nyu.edu]
Thomas Sargent [thomas.sargent@nyu.edu]

Spring 2016

Agenda

1. Class Schedule (Remaining 6 Lectures)
2. Working with Data in Python (Pandas)

Class Schedule

19th April	Working with Data in Python (Pandas)
21st April	Asset Pricing (Thomas Sargent)
26th April	Asset Pricing (Thomas Sargent)
28th April	McCall Search Model (Zhen Huo)
03rd May	McCall Search Model (Zhen Huo)
05th May	Review

Pandas

Pandas is the workhorse for data work in Python that is built on top of **NumPy**

Some things that Pandas is very good at:

1. Easy handling of missing data (represented as NaN) in floating point as well as non-floating point data
2. Automatic and explicit data alignment: objects can be explicitly aligned to a set of labels, or the user can simply ignore the labels and let Series, DataFrame, etc. automatically align the data for you in computations
3. Hierarchical labeling of axes (possible to have multiple labels per tick)

Reference: http://pandas-docs.github.io/pandas-docs-travis/

Pandas - Continued

Operations:

1. Powerful, flexible group by functionality to perform split-apply-combine operations on data sets, for both aggregating and transforming data
2. Intelligent label-based slicing, fancy indexing, and subsetting of large data sets
3. Intuitive merging and joining data sets
4. Flexible reshaping and pivoting of data sets

Reference: http://pandas-docs.github.io/pandas-docs-travis/

Pandas - Continued

IO

1. Robust IO tools for loading data from
    - flat files (CSV and delimited),
    - Excel files,
    - databases,
    - and saving / loading data from the ultrafast HDF5 format

Reference: http://pandas-docs.github.io/pandas-docs-travis/

Pandas - Continued

Specialized Data Types: TimeSeries

1. Time series-specific functionality:
    - date range generation and frequency conversion,
    - moving window statistics,
    - moving window linear regressions,
    - date shifting and lagging, etc.

Reference: http://pandas-docs.github.io/pandas-docs-travis/

Working with Data in Python

1. intro-python-and-data-analysis.ipynb
2. pandas-explore-fred-data.ipynb

Other Useful Packages

There are a lot of additional packages that are useful for working with data in Python

1. NetworkX (https://networkx.github.io/)
2. odo (https://github.com/blaze/odo)
3. dask (http://dask.pydata.org/en/latest/)
4. statsmodels (http://statsmodels.sourceforge.net/)
5. Scikit-learn (http://scikit-learn.org/stable/)
6. Beautiful Soup, HTML5lib, ...
7. Matplotlib, Plotly, Bokeh ...