

**Introduction of Jupyter
Notebook
&
Data Analyses using plotting
packages**

By Linda Li



Jupyter Notebook is a Web based application, that allows you to write and execute code, analyze data, embed content, and share reproducible work.

Uses include: data cleaning and transformation, numerical simulation, statistical modeling, machine learning and much more.

Publishing is flexible: PDF, HTML, ipynb, dashboards, slides, and more. Code cells are based on an input and output format.

Jupyter Notebook Documentation:

<https://jupyter-notebook.readthedocs.io/en/latest/>

Installation & Accessing Jupyter Notebook

1. Pip install, windows user can install with setuptools

2. Anaconda

<http://jupyter-notebook-beginner-guide.readthedocs.io/en/latest/install.html>,

<http://jupyter.readthedocs.io/en/latest/install.html>

3. Microsoft Azure account <https://studio.azureml.net>

4. Google Cloud Platform Datalab

5. <https://try.jupyter.org>

Jupyter Notebook vs Python Shell

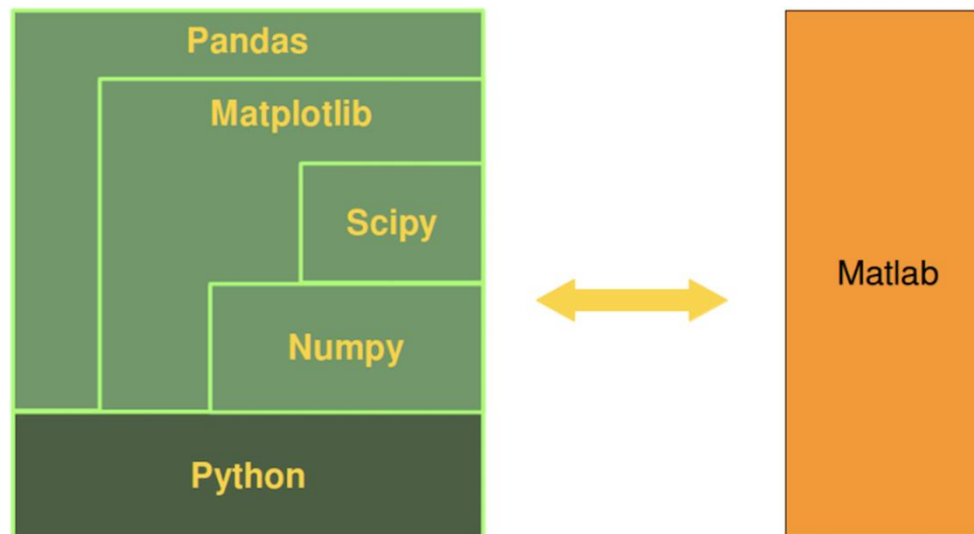
- Jupyter Notebook goes back and forth
- Jupyter persists the code and data output
- Jupyter Notebook saves to pdf file (HTML, ipynb..) for sharing

Data Analysis Packages

- Numpy
- Pandas

Numpy is a science calculation package from Python

Numpy is largely related to matrix manipulation and linear algebra. Its position in the python stack is in the following chart:



Numpy is used to simplify mathematical operation, for example – multiply matrices

- Traditional matrix multiplication hand-coded with Python involves loop inside the loop
- With Numpy :
`new_matrix=np.multiply(matrix1, matrix2)`

Pandas

Pandas DataFrame - Operation on data sets to help automation and analysis on the data sets. Pandas offers DataFrame to facilitate data set manipulation and analysis, it also provides simple charting capability on top of matplotlib.pyplot.

An example of Pandas DataFrame demonstrates the followings:

Data wrangling

Charting

Seaborn

Cartopy for Geospatial charting

priceline.com

Source from the public travel data set:

Download data: Air transport, passengers carried (1970-2016):

<https://data.worldbank.org/indicator/IS.AIR.PSGR?view=chart>

priceline.com

Many Thanks to Arup Danda – Priceline VP of Data, Technology for review