Data Analysis with Pandas

Jeff Reback

@jreback

StrataNYC 2015

September 29, 2015

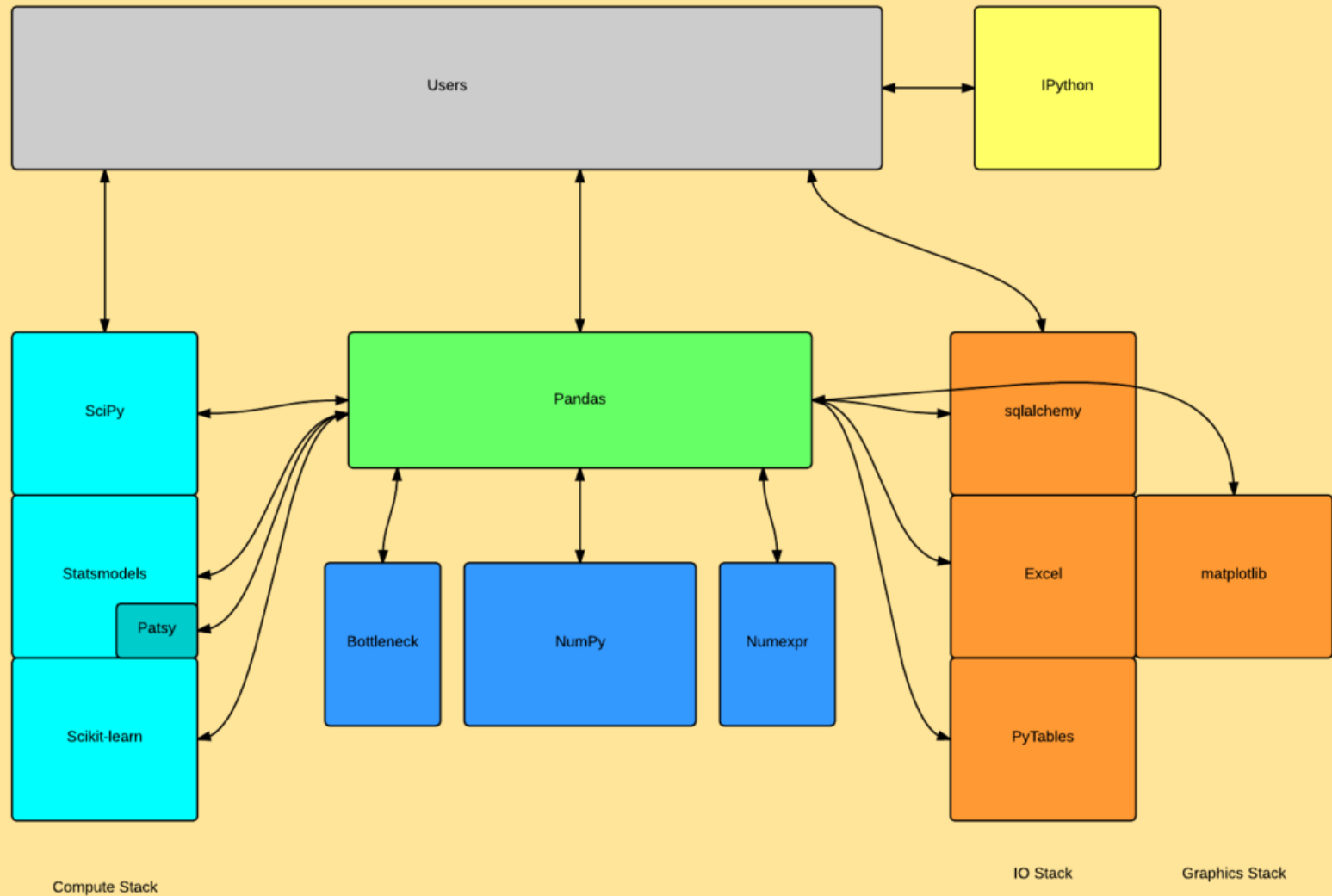https://github.com/jreback/StrataNYC2015/tree/master/data_analysis

Jeff Reback

@jreback

- former quant
- currently working on projects at Continuum
- core commiter to pandas for last 3 years
- manage pandas since 2013

Why Pandas?

- Vectorization for the masses
- ETL
- Fast and Efficient DataFrame
- Interoperability with ecosystem
- Database-like
- User friendly API
- Munging & data prep is a big part of the pipeline

http://pandas.pydata.org/

Users

IPython

SciPy

Pandas

sqlalchemy

Statsmodels

Patsy

Bottleneck

NumPy

Numexpr

Excel

matplotlib

Scikit-learn

PyTables

Compute Stack

IO Stack

Graphics Stack

1. Data IO

2. Examine the Data

3. Grouping

4. Tidy Data

5. ML & Stats

How to contribute

`https://github.com/pydata/pandas/issues`

This Talk

https://github.com/jreback/StrataNYC2015/tree/master/data_analysis

@jreback