01_introduction-IPython-notebook

June 24, 2015

1 Efficient Data Analysis with the IPython Notebook

1.1 Objectives

- Become familiar with the **IPython Notebook**.

- Introduce the IPython **landscape**.

- Getting started with **exploratory data analysis** in Python

- Conducting reproducible data analyis and computing experiments

- How do you currently:

    - wrangle data?
    - visualize results?
    - Analysis: machine learning, stats
    - Parallel computing
    - Big data

1.2 What is Python?

Python is a general-purpose programming language that blends procedural, functional, and object-oriented paradigms

Mark Lutz, Learning Python

- Simple, clean syntax
- Easy to learn
- Interpreted
- Strong, dynamically typed
- Runs everywhere: Linux, Mac, and Windows
- Free and open
- Expressive: do more with fewer lines of code
- Lean: **modules**
- Options: **Procedural**, **object-oriented**, and **functional**.

1.2.1 Abstractions

- Python provides high-level abstraction
- Performance can be on par with compiled code if right approach is used

1.3 IPython and the IPython Notebook

1.4 IPython

- Platform for interactive computing
- Shell or browser-based notebook
- Project Jupyter
- Language independent notebook
- Can be used with R, Julia, bash ...

1.4.1 IPython Notebook

http://blog.fperez.org/2012/01/ipython-notebook-historical.html
Interactive web-based computing, data analysis, and documentation.

- One document for code and output
- Run locally and remote
- Document process
- Share results

1.4.2 Integrate Code and Documentation

- Data structure ouput

- Inline plots

- Conversation sytle programming (Literate programming)

- Telling a data story

- Great for *iterative* programming.

  - Data analysis
  - Quick scripts
  - Prototyping

- 2 type of cells:

  - `Markdown` for documentation
  - `Code` for execution programs

```
In [1]: 2+4
```

```
Out[1]: 6
```

```
In [2]: print("Hello world!")
```

```
Hello world!
```

1.4.3 Locally and Remote

- Run locally
- Connect to the cloud (e.g AWS)
- Connect to supercomputer (e.g. XSEDE Resource)
- Add compute power:
  - mpi4py
  - IPython Parallel
  - spark big distributed data
  - Numbapro GPU
  - ...

1.4.4 Documentation and Sharing

1.4.5 Keyboard Shortcuts

1.4.6 Markdown and LaTeX

- Markdown
- Latex $y = \sqrt{a+b}$

1.4.7 Images

<img src='https://s3.amazonaws.com/research_computing_tutorials/monty-python.png' width="300">

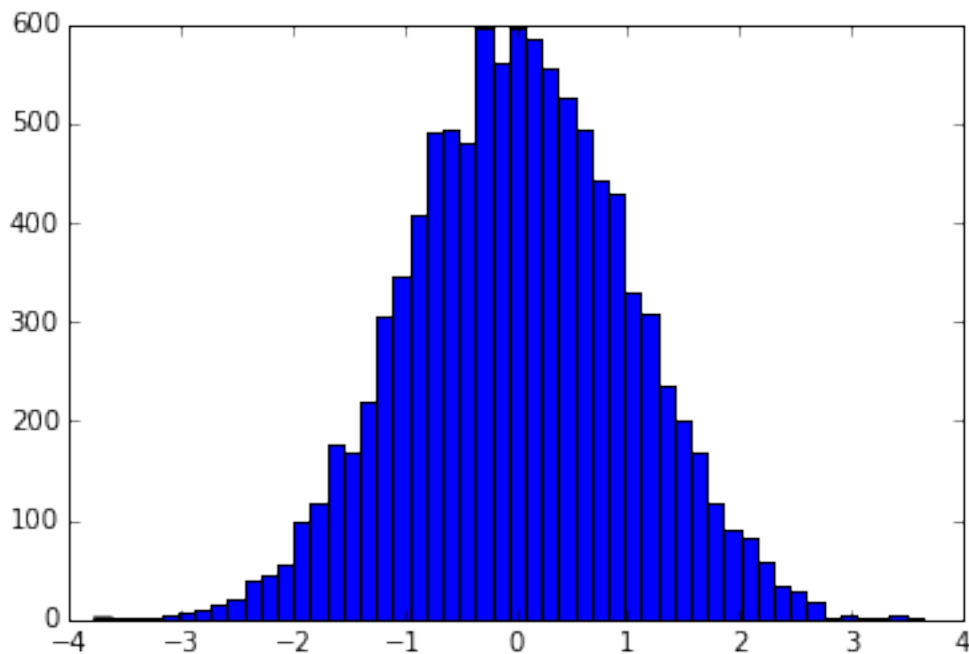This is an image:

1.4.8 Embeded Plots

```
In [39]: %matplotlib inline
         import matplotlib.pyplot as plt
         import numpy as np
         x = np.random.randn(10000)
         print(x)
```

```
[ 0.59719566  0.76673608 -0.69192902 ...,  1.25618813  0.67444306
  1.07282385]
```

Plot a Histogram of x

```
In [41]: plt.hist(x, bins=50)
         plt.show()
```

1.4.9 Customizable

- Custom CSS
- Custom javascript libraries
- Create your own output format.
- Tools and workflow

1.4.10 Magic Commands

- Built-in useful functions
- % line commands
- %% cell commands

```
In [5]: %lsmagic
```

```
Out[5]: Available line magics:
        %alias  %alias_magic  %autocall  %automagic  %autosave  %bookmark  %cat  %cd  %clear  %colors  %

        Available cell magics:
        %%!  %%HTML  %%SVG  %%bash  %%capture  %%debug  %%file  %%html  %%javascript  %%latex  %%perl  %

        Automagic is ON, % prefix IS NOT needed for line magics.
```

```
In [42]: %timeit y = np.random.randn(100000)
```

```
100 loops, best of 3: 3.46 ms per loop
```

```
In [1]: %ll
```

```
total 2876
-rw-r--r-- 1 tom  63029 Jun 23 11:41 01_exploratory_pandas.ipynb
-rw-r--r-- 1 tom  93630 Jun 24 05:13 01_introduction-IPython-notebook.ipynb
-rw-r--r-- 1 tom 171148 Jun 23 09:38 01_introduction-IPython-notebook.pdf
drwxr-xr-x 5 tom    170 Jun 23 09:35 01_introduction-IPython-notebook_files/
-rw-r--r-- 1 tom 795157 Jun 23 16:35 02_Data_Transfer.pdf
-rw-r--r-- 1 tom 177172 Jun 23 16:35 03_HPC_File_Systems.pdf
-rw-r--r-- 1 tom 227497 Jun 23 12:56 04_python_reading-plotting.ipynb
-rw-r--r-- 1 tom 213926 Jun 24 05:04 05_Data_Conversion_Cleaning.pdf
-rw-r--r-- 1 tom   4709 Jun 24 05:04 06_CSV_to_NetCDF_Exercise.ipynb
-rw-r--r-- 1 tom   9477 Jun 24 05:04 06_CSV_to_NetCDF_Solution.ipynb
-rw-r--r-- 1 tom 298136 Jun 23 16:31 07_python-matplotlib.ipynb
lrwxr-xr-x 1 tom     77 Jun 23 05:51 Walrus_Data -> /Users/tom/Google Drive/Grants/2014_USGS/2015-06-23-
-rw-r--r-- 1 tom  87723 Jun 23 14:10 basic_animation.mp4
-rw-r--r-- 1 tom 191347 Jun 23 05:17 data_overview.png
-rw-r--r-- 1 tom 230492 Jun 23 05:17 ipython-notebook-keyboard.png
-rw-r--r-- 1 tom 172569 Jun 23 05:17 ipython-notebook-sharing.png
-rw-r--r-- 1 tom  77163 Jun 23 05:17 ipython-notebook.png
-rw-r--r-- 1 tom   9674 Jun 23 05:42 rc_logo.png
-rw-r--r-- 1 tom  72981 Jun 23 05:17 traditional_python.png
-rw-r--r-- 1 tom    262 Jun 23 16:33 walrus_animation.mp4
-rw-r--r-- 1 tom   1205 Jun 23 11:41 walrus_behav.png
```

1.4.11 Other Languages: Bash

```
In [43]: %%bash
         ls -l
```

```
total 896
-rw-r--r-- 1 tom staff  27897 Jun 18 13:58 03_overview.ipynb
-rw-r--r-- 1 tom staff  34774 Jun 18 16:35 06_plotting.ipynb
-rw-r--r-- 1 tom staff  91324 Jun 22 10:44 07_introduction-IPython-notebook.ipynb
lrwxr-xr-x 1 tom staff     77 Jun 18 15:59 Walrus_Data -> /Users/tom/Google Drive/Grants/2014_USGS/2015-
-rw-r--r-- 1 tom staff 191347 Jun 21 16:38 data_overview.png
-rw-r--r-- 1 tom staff 230492 Jun 21 17:11 ipython-notebook-keyboard.png
-rw-r--r-- 1 tom staff 172569 Jun 21 17:08 ipython-notebook-sharing.png
-rw-r--r-- 1 tom staff  77163 Jun 21 17:06 ipython-notebook.png
-rw-r--r-- 1 tom staff  72981 Jun 21 17:05 traditional_python.png
```

```python
In [ ]: files = !ls # But glob is a better way
        print files[:5]
```

1.4.12 Keep it all together

```python
In [ ]: %%writefile example.cpp
        #include <iostream>

        int main(){
            std::cout << "hello from c++" << std::endl;
        }
```

```python
In [ ]: %%bash
        g++ example.cpp -o example
        ./example
```

1.5 NBconvert examples

- HTML
- PDF (print) - you have to have LaTex installed
- Slides
- Dynamic Slides
- ReStructured Text (sphinx)

```python
In [2]: !ipython nbconvert --to 'PDF' 01_introduction-IPython-notebook.ipynb
```

```
[NbConvertApp] Converting notebook 01_introduction-IPython-notebook.ipynb to pdf
[NbConvertApp] Support files will be in 01_introduction-IPython-notebook_files/
[NbConvertApp] Making directory 01_introduction-IPython-notebook_files
[NbConvertApp] Making directory 01_introduction-IPython-notebook_files
[NbConvertApp] Making directory 01_introduction-IPython-notebook_files
[NbConvertApp] Writing 33048 bytes to notebook.tex
[NbConvertApp] Building PDF
[NbConvertApp] Running pdflatex 3 times: [u'pdflatex', u'notebook.tex']
[NbConvertApp] PDF successfully created
[NbConvertApp] Support files will be in 01_introduction-IPython-notebook_files/
[NbConvertApp] Making directory 01_introduction-IPython-notebook_files
[NbConvertApp] Making directory 01_introduction-IPython-notebook_files
[NbConvertApp] Making directory 01_introduction-IPython-notebook_files
[NbConvertApp] Writing 171148 bytes to 01_introduction-IPython-notebook.pdf
```