Interactive Computing with Jupyter

Jason Grout, Bloomberg

Me

- 2007 PhD in Mathematics, BYU
- 2007 - 2009 Postdoc (Iowa State), SageMath developer
- 2009 - 2014 Asst Prof of Mathematics, Drake U., SageMath
- 2014 - … Bloomberg, IPython/Jupyter developer

Feel free to interrupt me if you have any questions

jasongrout

jgrout6@bloomberg.net

Poll

Raise your hand if you have used Python or Jupyter

- with a language other than Python?
- with the web-based Jupyter notebook?
- with the interactive widgets in the notebook?

NOVEMBER 2001: "JUST AN AFTERNOON HACK"

- **259 Line Python script.** (https://gist.github.com/fperez/1579699)
- **sys.ps1, displayhook**
- **Plotting, Numeric, etc.**

2014 (OPENHUB STATS)

- **19,279 commits**
- **442 contributors**
- **Total Lines: 187,326**
- **Number of Languages: 7**

2015 "THE BIG SPLIT"

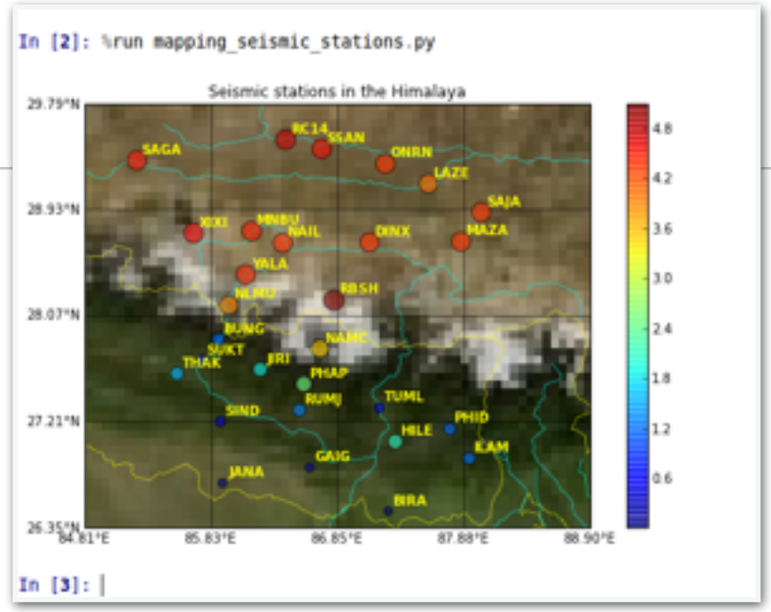
Dozen of projects… and counting

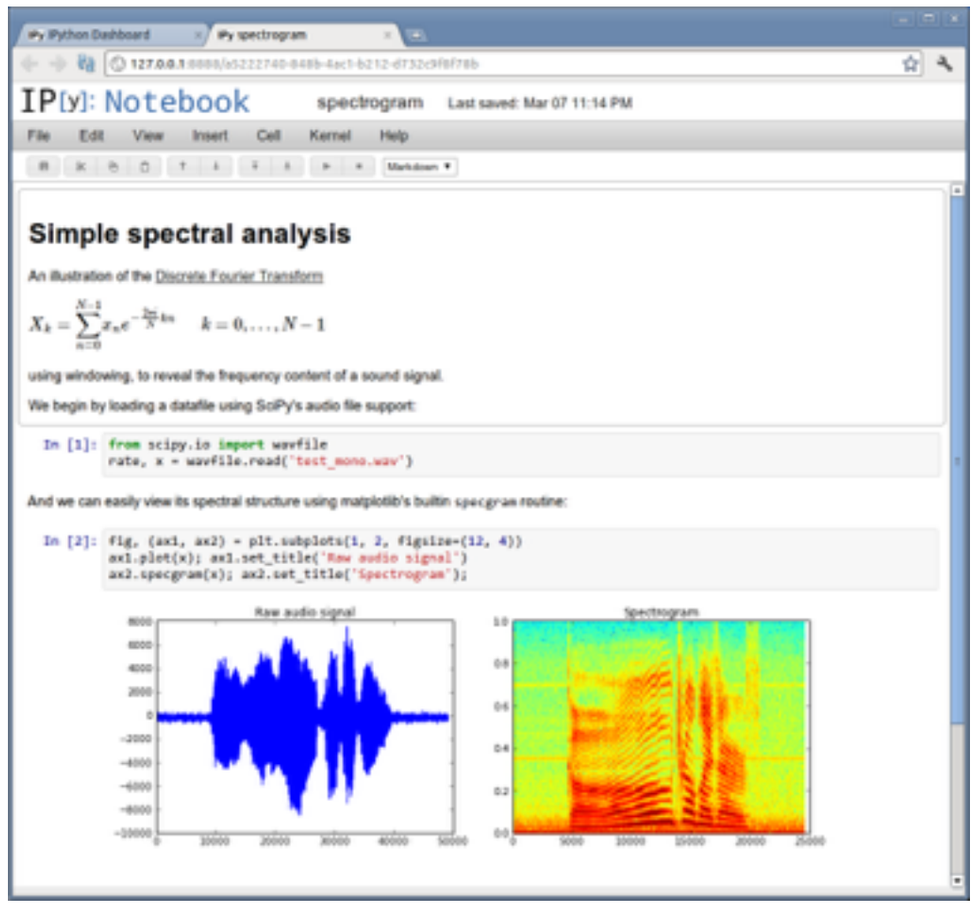
Improve over the terminal

- The REPL as a network protocol
- Kernels
  - execute code
- Clients
  - Read input
  - Present output

Simple abstractions enable rich, sophisticated clients

- The Notebook: http://try.jupyter.org
  - Rich web client
  - Text & math
  - Code
  - Results
  - Share, reproduce.





Jupyter vs IPython

- Network protocol for interactive computing
- Clients for protocol
  - Console
  - Qt Console
  - Notebook
- Notebook file format & tools (nbconvert…)
- JupyterHub
- Nbviewer
- NbGrader
- Tmpnb
- …

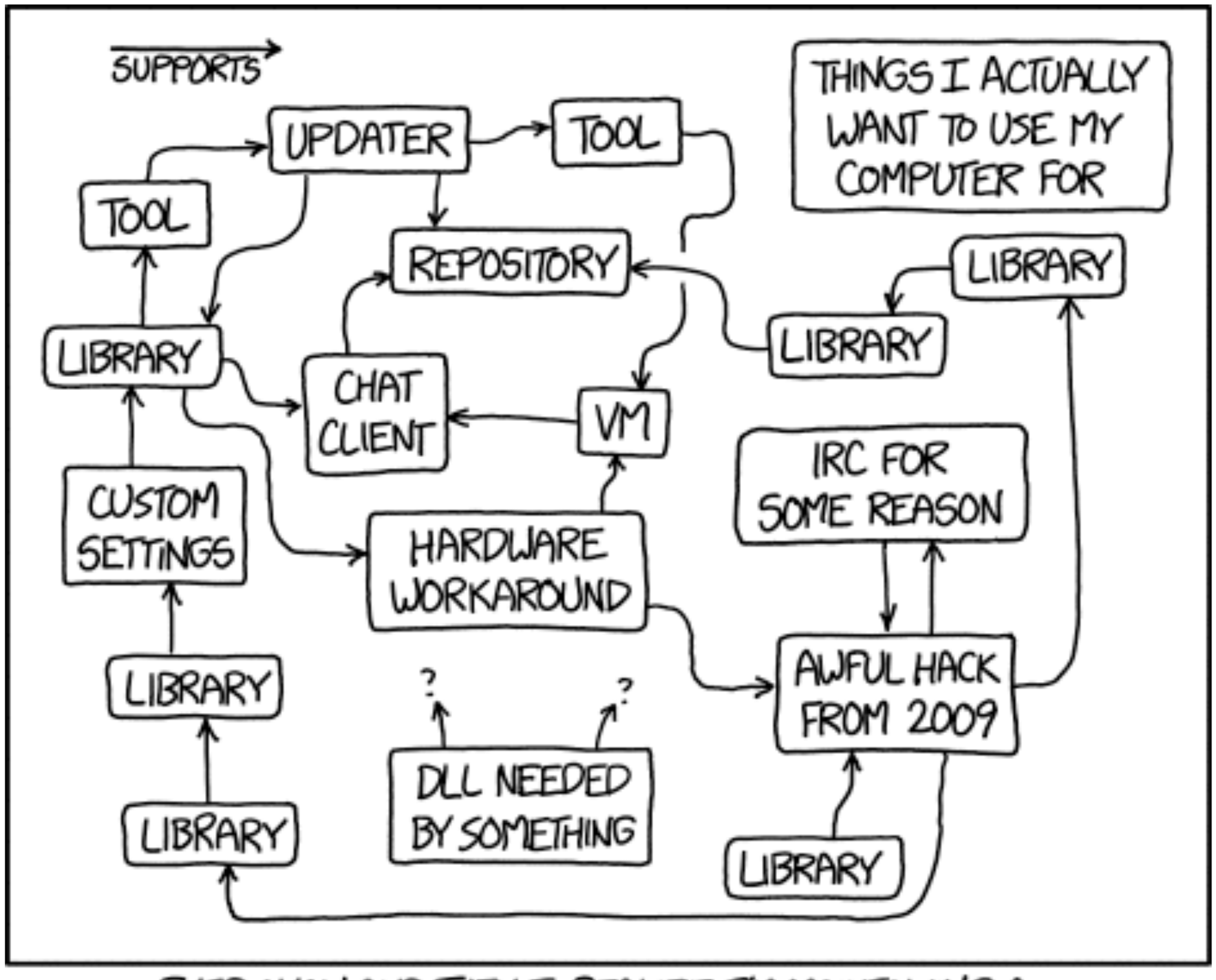
- Interactive Python shell at the terminal
- Kernel for this protocol in Python
- Tools for Cross-Language integration
- Tools for Interactive Parallel computing
- The “reference” kernel for Jupyter

Why Jupyter?

The purpose of computing is insight, not numbers.

- Richard Hamming, 1962

Wikipedia



EVERY NOW AND THEN I REALIZE I'M MAINTAINING A HUGE CHAIN OF TECHNOLOGY SOLELY TO SUPPORT ITSELF.

xkcd.com/1579

Gaining Insight

- Friction distracts focus and causes blisters
- Context switches cause friction
  - Switching/managing tool chains
  - Exploration vs typing

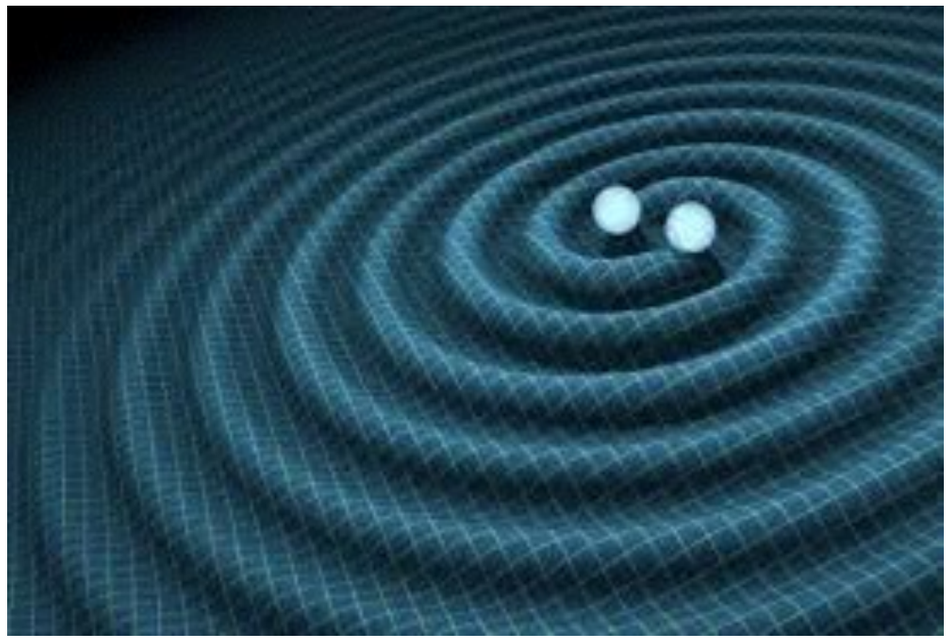
Reduce Friction

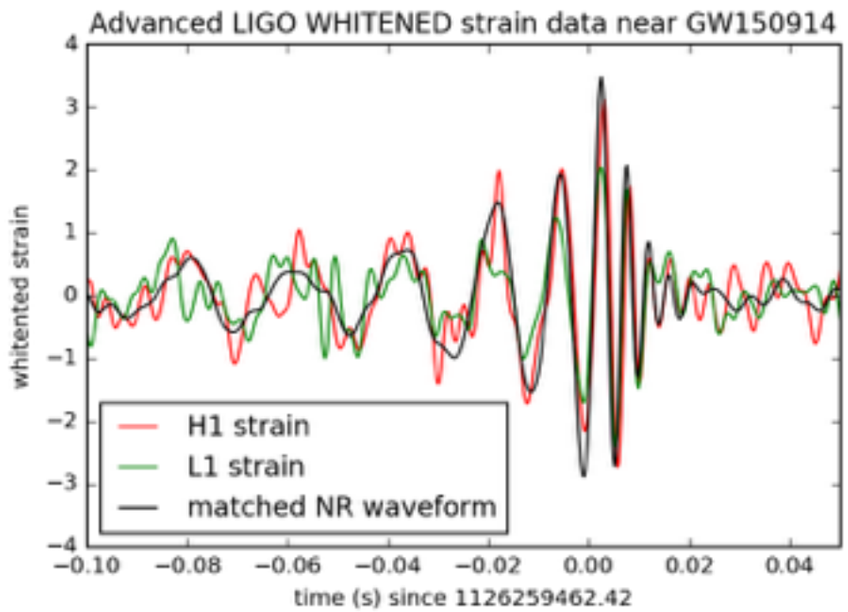
Demo

Giving Insight

Explanation and Interactive Computation Together

LIGO

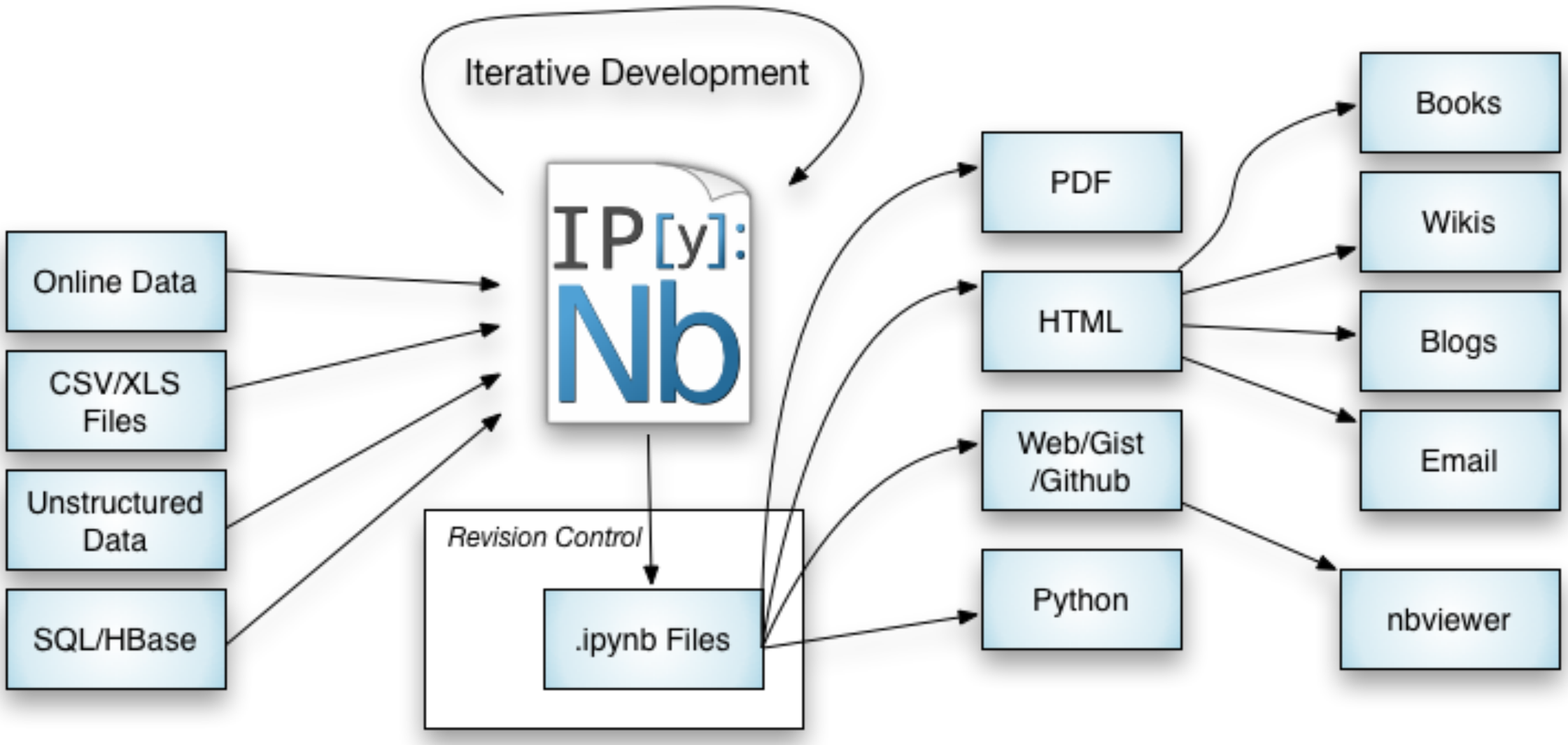




https://www.youtube.com/watch?v=TWqhUANNFXw

http://mybinder.org/repo/minrk/ligo-binder/GW150914_tutorial.ipynb

Sharing a notebook



NbViewer

- Zero-install reading of notebooks
- Just share a URL
- nbviewer.org

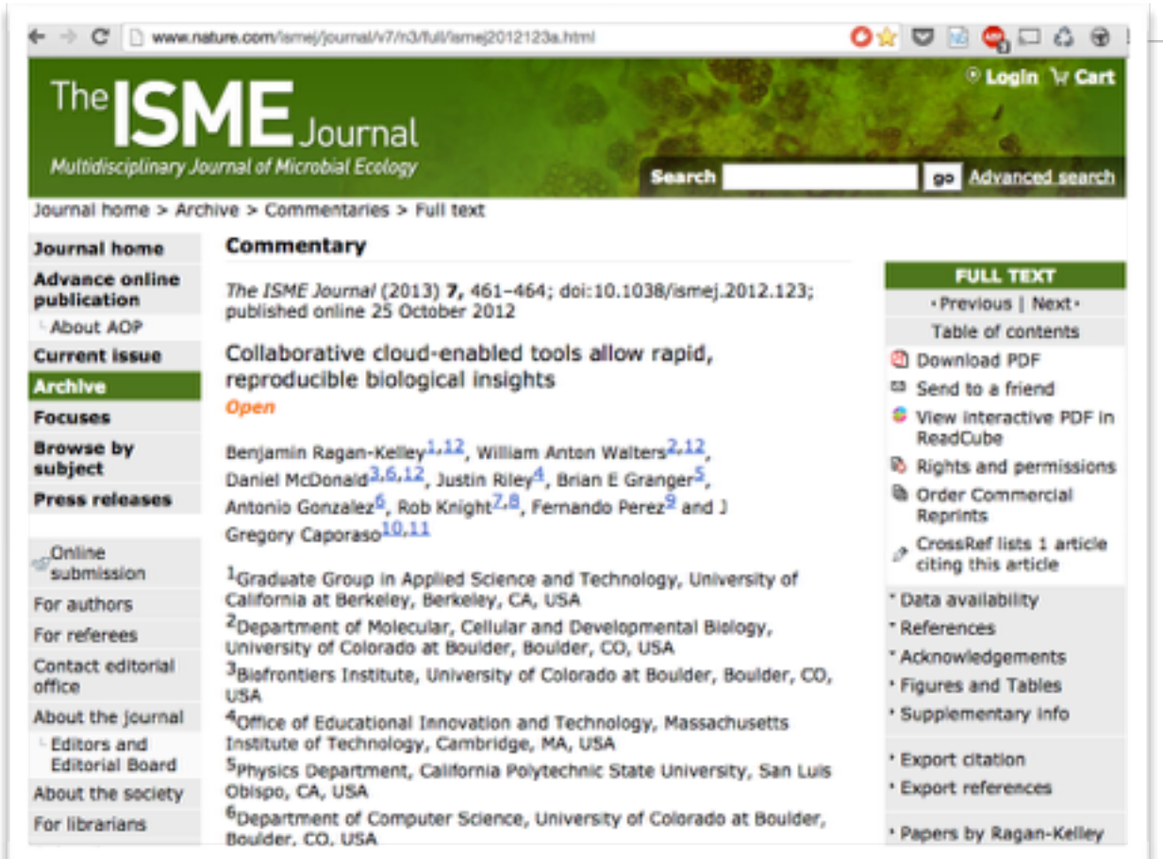
  - convert to HTML on the fly.

Sharing:

git push, or dropbox sync.

- Since May GitHub renders Notebooks
- Powered by `nbconvert`, the library that deals with `.ipynb` -> *
- Over 200,000 notebooks on GitHub

Papers

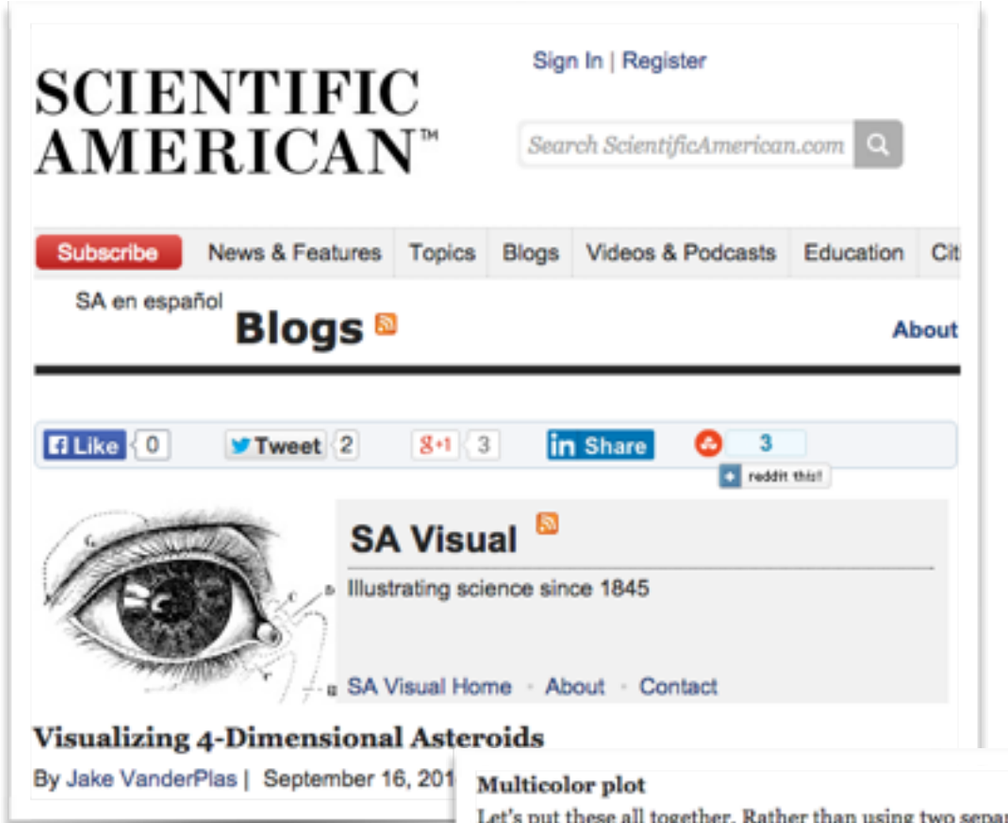


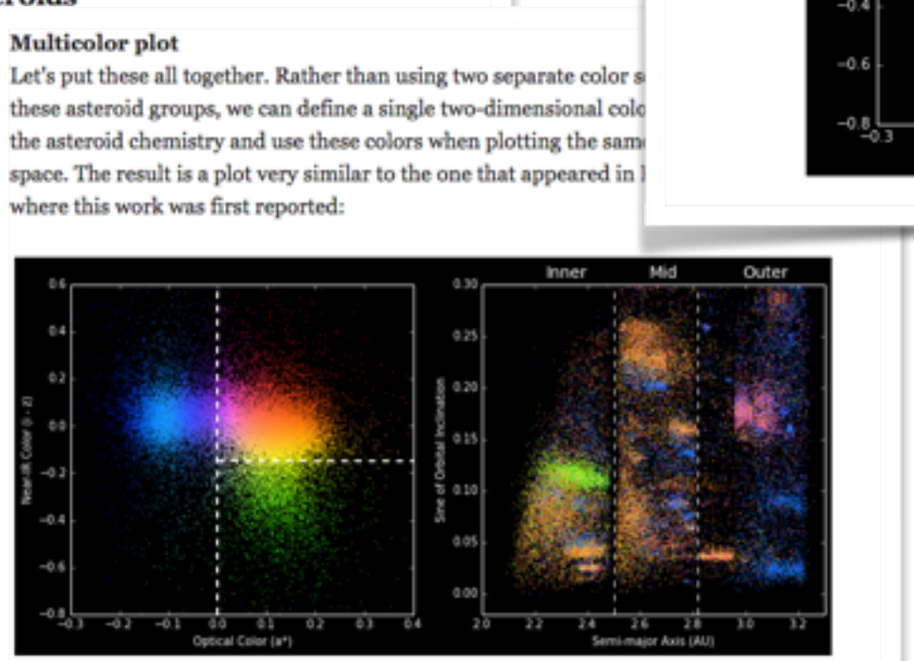
http://www.nature.com/ismej/journal/v7/n3/full/ismej2012123a.html

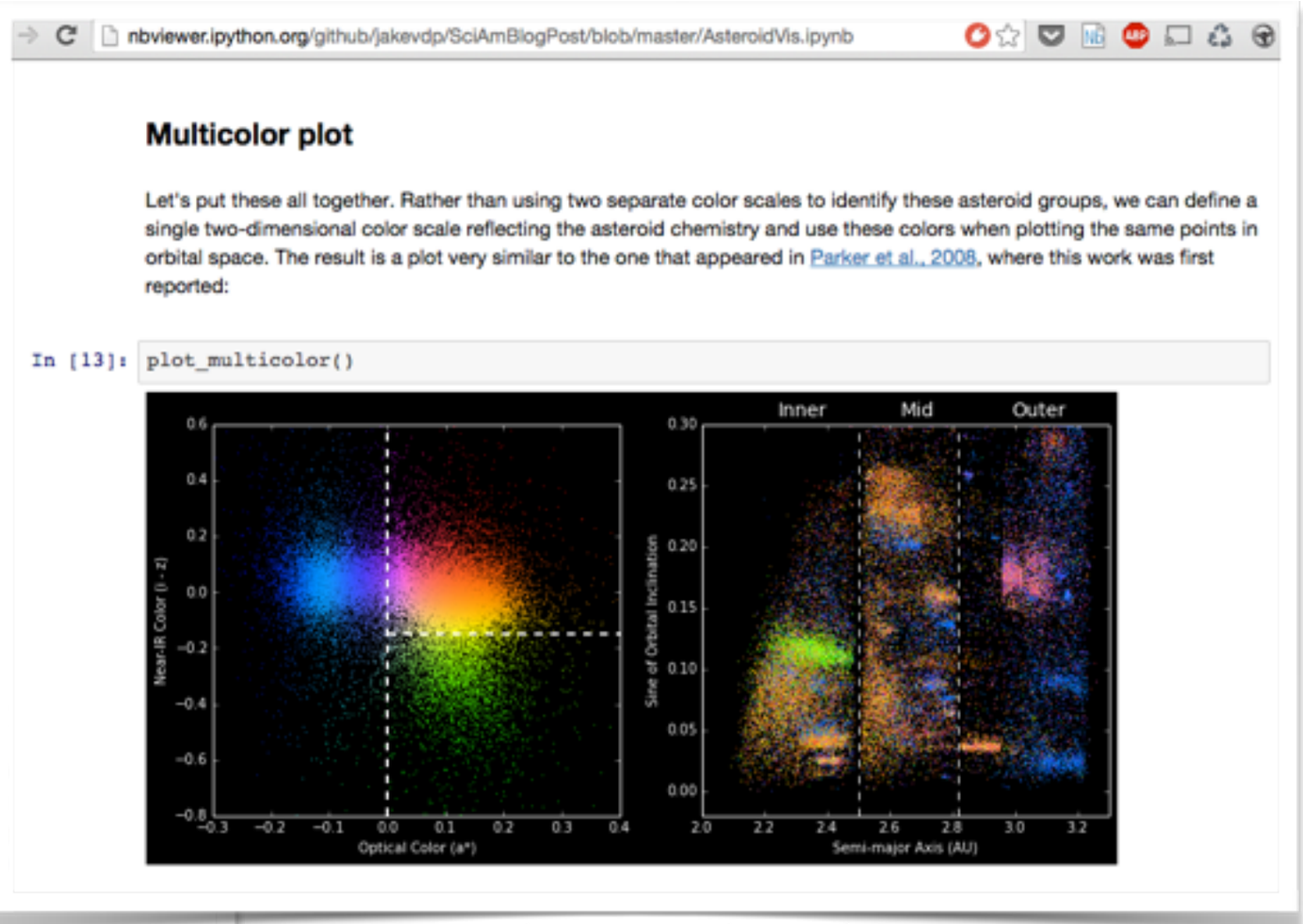
http://qiime.org/home_static/nih-cloud-apr2012

Papers with code as AMI/VMs

Blogs







Jake van der Plas @ UW

http://blogs.scientificamerican.com/sa-visual/2014/09/16/visualizing-4-dimensional-asteroids

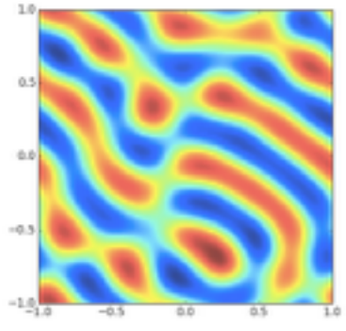
Courses, MOOCS

	Course	University	Instructor
0	Data Science and Visualization with Python	Santa Clara	Brian Granger
1	Python for Data Science	UC Berkeley	Josh Bloom
2	Introduction to Data Science	UC Berkeley	Michael Franklin
3	Working with Open Data	UC Berkeley	Raymond Yee
4	Introduction to Signal Processing	UC Berkeley	Miki Lustig
5	Data Science (CS 109)	Harvard University	Pfister and Blitzstein
6	Practical Data Science	NYU	Josh Attenberg
7	Scientific Computing (ASTR 599)	University of Washington	Jake Vanderplas
8	Computational Physics	Cal Poly	Jennifer Klay
9	Introduction to Programming	Alaskan High School	Eric Matthes
10	Aerodynamics-Hydrodynamics (MAE 6226)	George Washington University	Lorena Barba
11	HyperPython: hyperbolic conservation laws	KAUST	David Ketcheson
12	Quantitative Economics	NYU	Sargent and Stachurski
13	Practical Numerical Methods with Python	4 separate universities + MOOC	Barba, et al.
14	Data Science: Algorithms	Columbia - Lede Program	Chris Wiggins
15	Data Science: Databases	Columbia - Lede Program	Chris Wiggins
16	Data Science: Foundations	Columbia - Lede Program	Chris Wiggins
17	Data Science: Platforms	Columbia - Lede Program	Chris Wiggins

Lorena A. **Barba group**

Announcing "Practical Numerical Methods with Python" MOOC

Posted on 07.26.2014



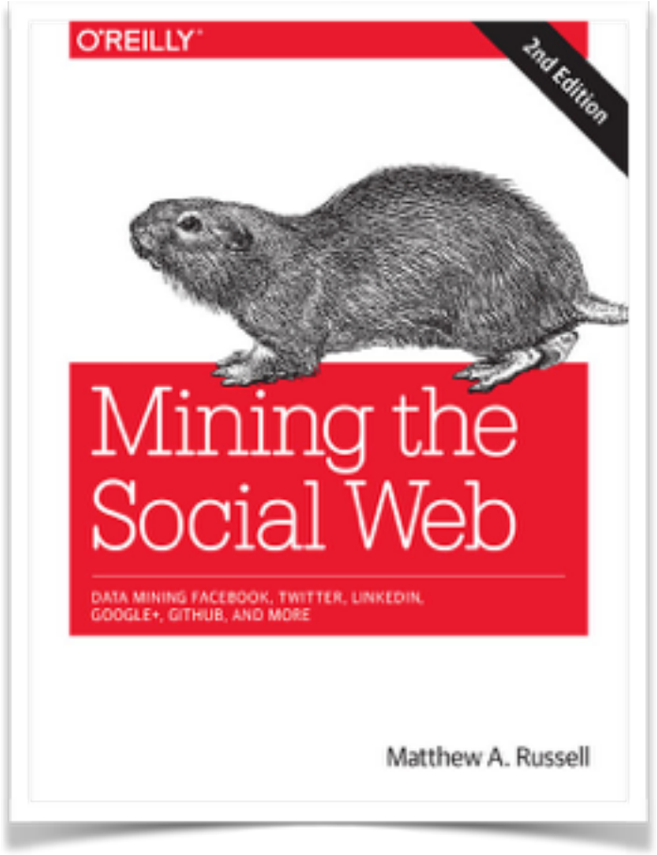
Pattern formation:

- solution for a reaction-diffusion system like:

$$u_t = \delta D_1 \nabla^2 u + f(u,v)$$
$$v_t = \delta D_2 \nabla^2 v + g(u,v)$$

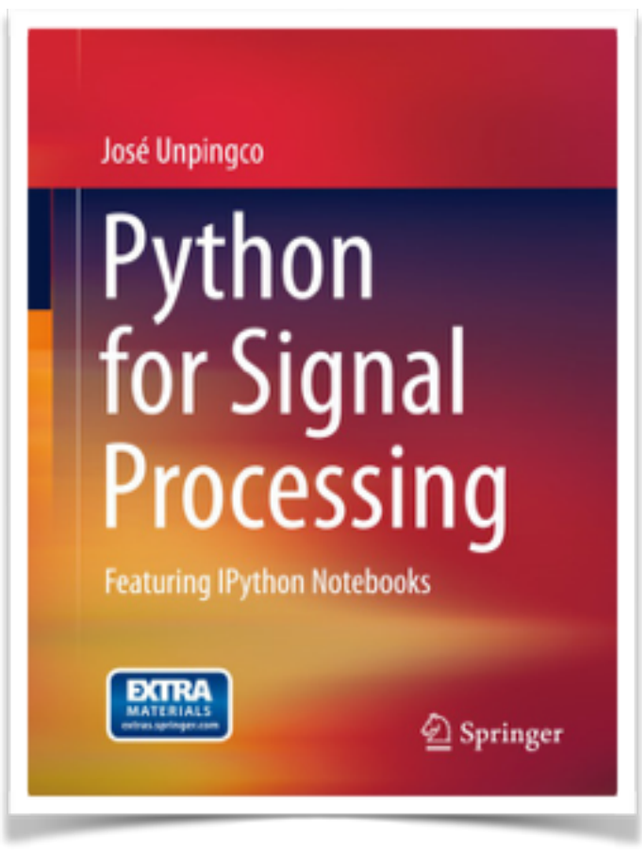
An example of the types of problems we will learn to solve in this course, among others governed by differential equations.

Books



By Matthew Russell

By Cameron Davidson-Pilon



By José Unpingco

You can download and execute the books locally.

Check The Gallery

A gallery of interesting IPython Notebooks

Fernando Perez edited this page 8 days ago · 229 revisions

This page is a curated collection of IPython notebooks that are notable for some reason. Feel free to add new content here, but please try to only include links to notebooks that include interesting visual or technical content; this should *not* simply be a dump of a Google search on every ipynb file out there.

Important contribution instructions: If you add new content, please ensure that for any notebook you link to, the link is to the rendered version using nbviewer, rather than the raw file. Simply paste the notebook URL in the nbviewer box and copy the resulting URL of the rendered version. This will make it much easier for visitors to be able to immediately access the new content.

Note that Matt Davis has conveniently written a set of bookmarklets and extensions to make it a one-click affair to load a Notebook URL into your browser of choice, directly opening into nbviewer.

Table of Contents

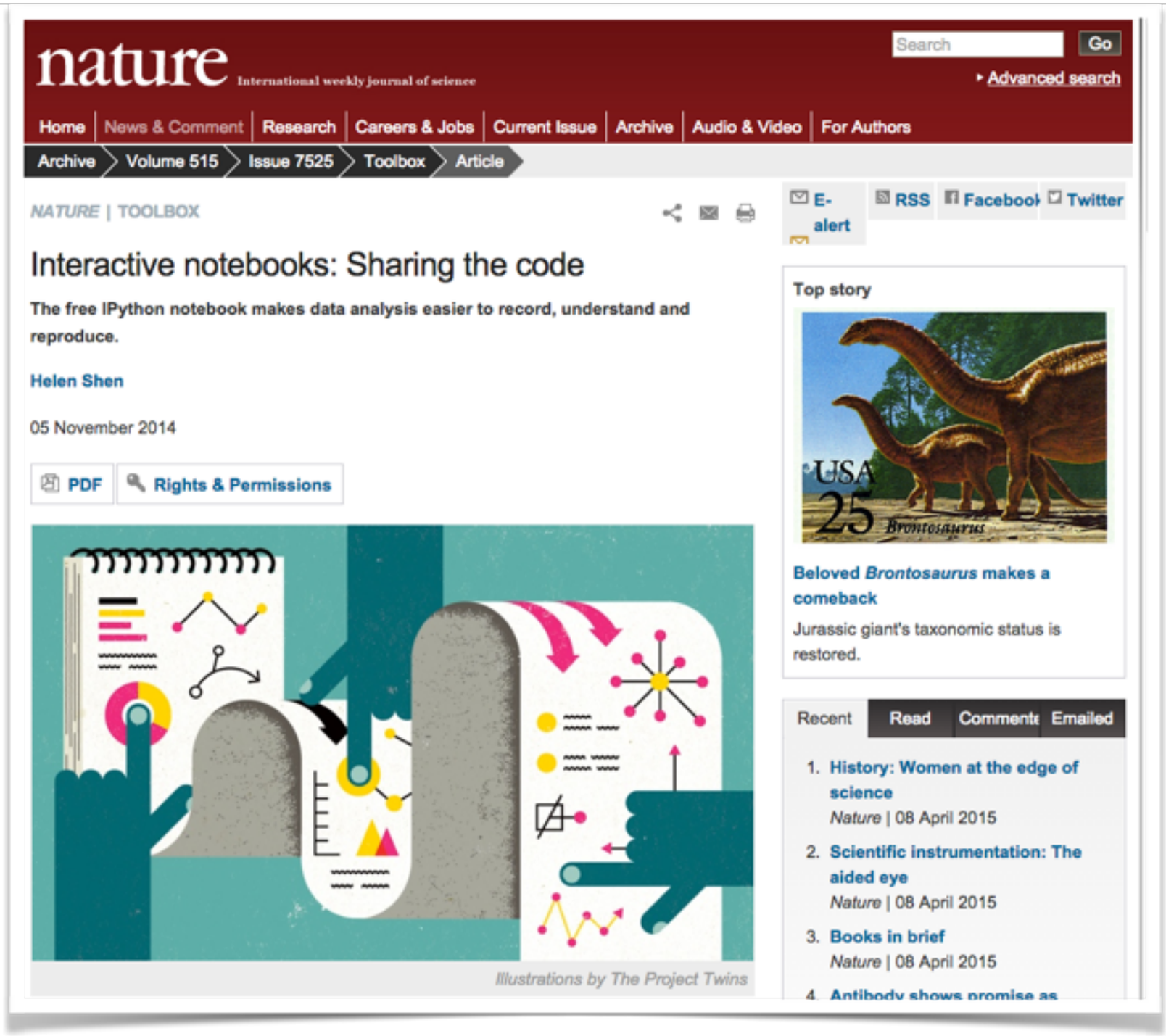
Reproducible academic publications

This section contains academic papers that have been published in the peer-reviewed literature or pre-print sites such as the ArXiv that include one or more notebooks that enable (even if only partially) readers to reproduce the results of the publication. If you include a publication here, please link to the journal article as well as providing the nbviewer notebook link (and any other relevant resources associated with the paper).

1. Reply to 'Influence of cosmic ray variability on the monsoon rainfall and temperature': a false-positive in the field of solar-terrestrial research by Benjamin Laken, 2015. Reviewed article will appear in JASTP. The IPython notebook reproduces the full analysis and figures exactly as they appear in the article, and is available on Github: link via figshare.
2. The probability of improvement in Fisher's geometric model: a probabilistic approach, by Yoav Ram and Lilach Hadany. (Theoretical Population Biology, 2014). An IPython notebook, allowing figure reproduction, was deposited as a supplementry file.
3. Stress-induced mutagenesis and complex adaptation, by Yoav Ram and Lilach Hadany (Proceedings B, 2014). An IPython notebook, allowing figures reproduction, was deposited as a supplementry file.
4. Automatic segmentation of odor maps in the mouse olfactory bulb using regularized non-negative matrix factorization, by J. Soelter et al. (Neuroimage 2014, Open Access). The notebook allows to reproduce most figures from the paper and provides a deeper look at the data. The full code repository is also available.
5. Multi-tiered genomic analysis of head and neck cancer ties TP53 mutation to 3p loss, by A. Gross et al. (Nature Genetics 2014). The full collection of notebooks to replicate the results.
6. powerlaw: a Python package for analysis of heavy-tailed distributions, by J. Alstott et al.. Notebook of examples in manuscript, ArXiv link and project repository.
7. Collaborative cloud-enabled tools allow rapid, reproducible biological insights, by B. Ragan-Kelley et al.. The main notebook, the full collection of related notebooks and the companion site with the Amazon AMI information for reproducing the full paper.
8. A Reference-Free Algorithm for Computational Normalization of Shotgun Sequencing Data, by C.T. Brown et al.. Full notebook, ArXiv link and project repository.
9. The kinematics of the Local Group in a cosmological context by J.E. Forero-Romero et al.. The Full notebook and also all the data in a github repo.

https://github.com/ipython/ipython/wiki/A-gallery-of-interesting-IPython-Notebooks

Replicating, simpler for readers



http://www.nature.com/news/interactive-notebooks-sharing-the-code-1.16261

binder

Turn a GitHub repo into a collection of interactive notebooks powered by Jupyter and Kubernetes.

Have a repo full of Jupyter notebooks? With Binder, you can add a badge that opens those notebooks in an executable environment, making your code immediately reproducible by anyone, anywhere.

100% free and open source. Check out a bunch of examples.
(Currently in testing, let us know if you run into trouble!)

1

Tell us your GitHub repo

user/project OR github url

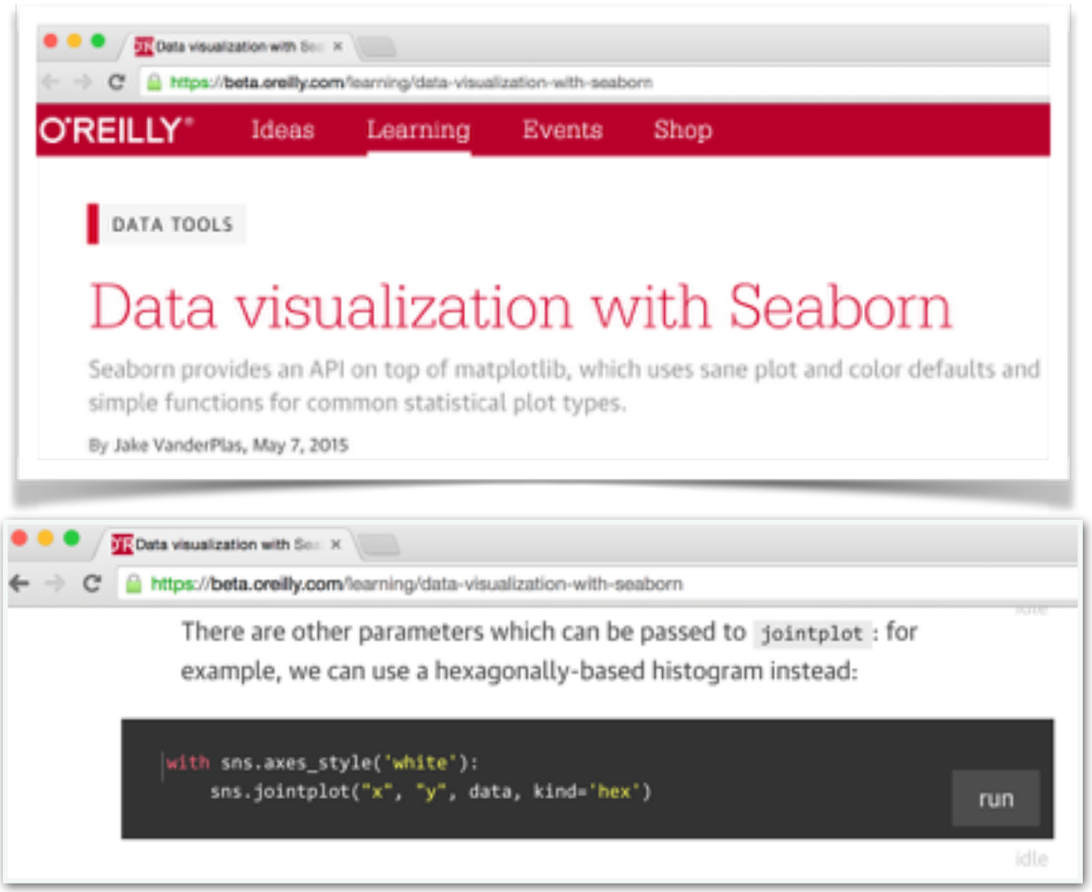
This should contain Jupyter notebooks. If one of them is called index.ipynb it will be where your Binder starts. Any extra folders or files (e.g. data) will be included. See an example repo that uses Binder.

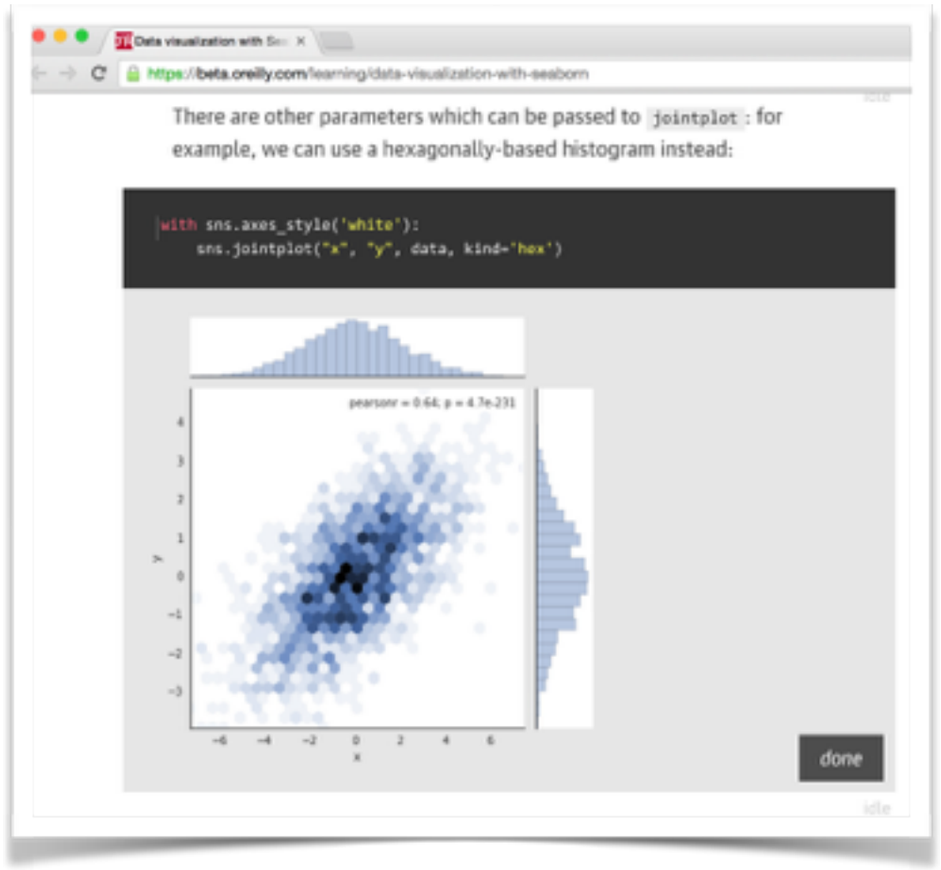
2

Four modes are supported. If

O'Reilly: authoring and delivering executable books

Atlas, ipymd and Thebe

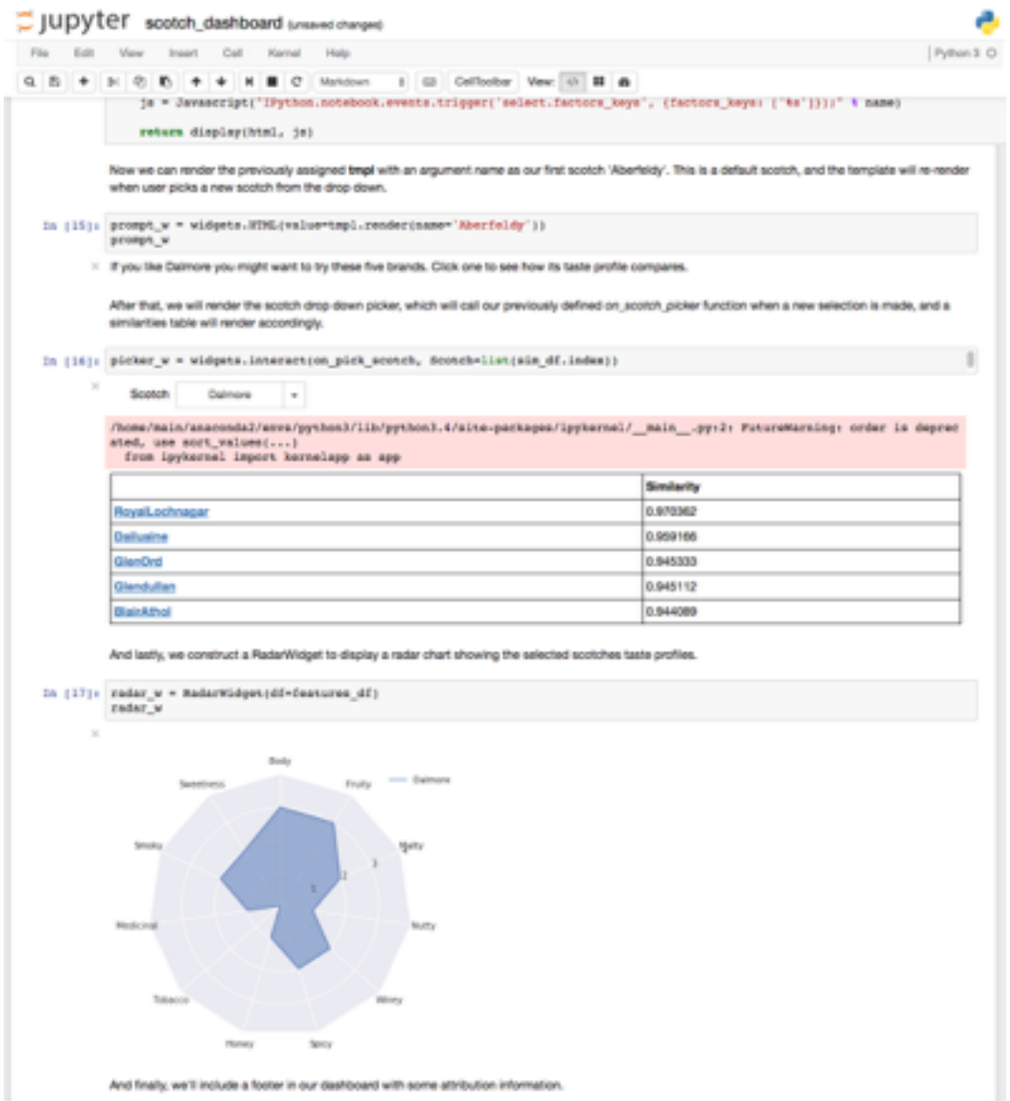


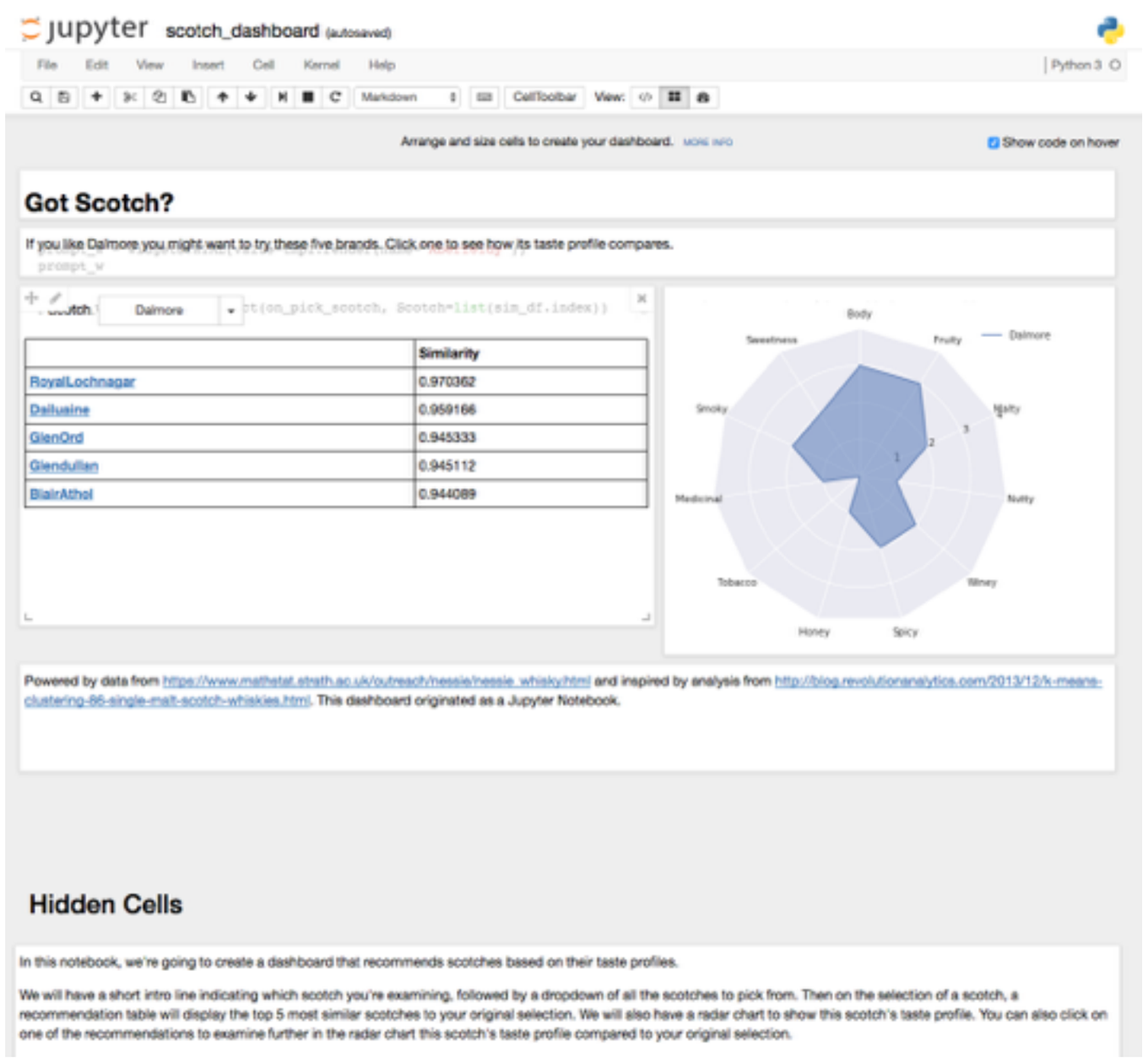


Edited as notebooks* on Atlas

www.oreilly.com/topics/data

Dashboard


jupyter scotch_dashboard (unsaved changes)
Now we can render the previously assigned tmpl with an argument name as our first scotch 'Aberfeldy'. This is a default scotch, and the template will re-render when user picks a new scotch from the drop down.
If you like Dalmore you might want to try these five brands. Click one to see how its taste profile compares.
After that, we will render the scotch drop down picker, which will call our previously defined on_scotch_picker function when a new selection is made, and a similarities table will render accordingly.
Scotch Dalmore
RoyalLochnagar 0.970362
Dailuaine 0.959166
GlenOrd 0.945333
Glendullan 0.945112
BlairAthol 0.944089
And lastly, we construct a RadarWidget to display a radar chart showing the selected scotches taste profiles.
Body
Sweetness
Fruity
Dalmore
Smoky
Nutty
Medicinal
Tobacco
Winey
Honey
Spicy
And finally, we'll include a footer in our dashboard with some attribution information.


jupyter scotch_dashboard (autosaved)
Arrange and size cells to create your dashboard. MORE INFO
Show code on hover
Got Scotch?
If you like Dalmore you might want to try these five brands. Click one to see how its taste profile compares.
Scotch Dalmore
RoyalLochnagar 0.970362
Dailuaine 0.959166
GlenOrd 0.945333
Glendullan 0.945112
BlairAthol 0.944089
Body
Sweetness
Fruity
Dalmore
Smoky
Nutty
Medicinal
Tobacco
Winey
Honey
Spicy
Powered by data from https://www.mathstat.strath.ac.uk/outreach/nessie/nessie_whisky.html and inspired by analysis from http://blog.revolutionanalytics.com/2013/12/k-means-clustering-86-single-malt-scotch-whiskies.html. This dashboard originated as a Jupyter Notebook.
Hidden Cells
In this notebook, we're going to create a dashboard that recommends scotches based on their taste profiles.
We will have a short intro line indicating which scotch you're examining, followed by a dropdown of all the scotches to pick from. Then on the selection of a scotch, a recommendation table will display the top 5 most similar scotches to your original selection. We will also have a radar chart to show this scotch's taste profile. You can also click on one of the recommendations to examine further in the radar chart this scotch's taste profile compared to your original selection.

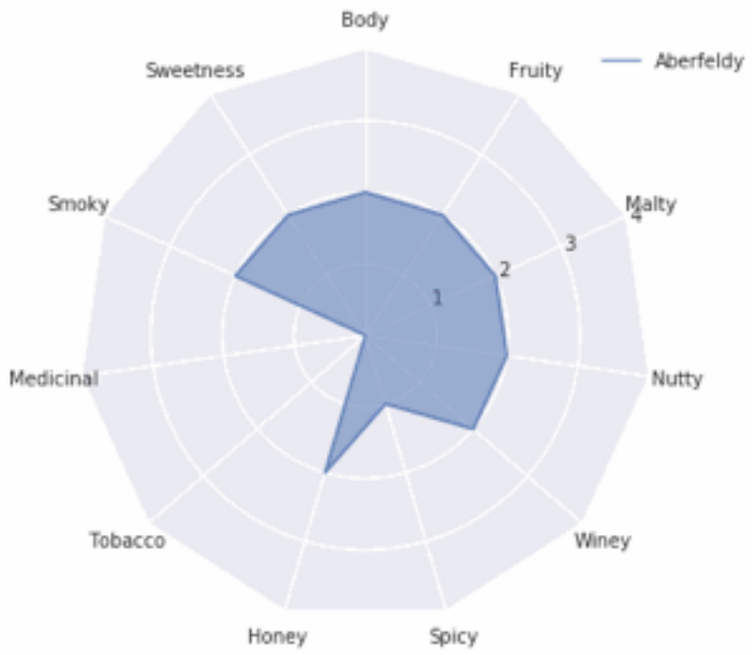
Dashboard

Got Scotch?

If you like Aberfeldy you might want to try these five brands. Click one to see how its taste profile compares.

Scotch Aberfeldy

	Similarity
BlairAthol	0.975610
Benromach	0.975610
Benrinnes	0.965345
Scapa	0.963072
Auchroisk	0.963036



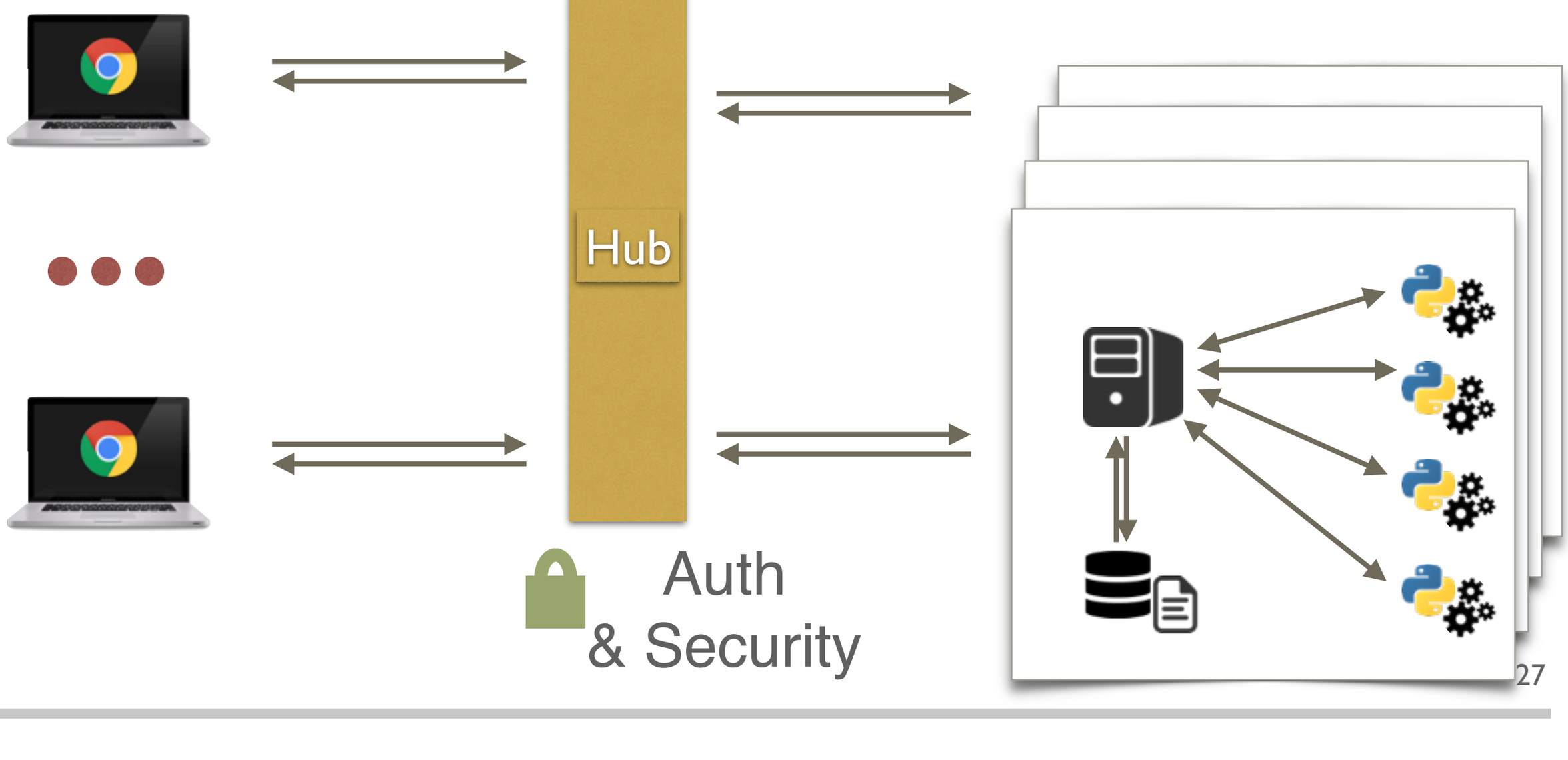
Powered by data from https://www.mathstat.strath.ac.uk/outreach/nessie/nessie_whisky.html and inspired by analysis from http://blog.revolutionanalytics.com/2013/12/k-means-clustering-86-single-malt-scotch-whiskies.html. This dashboard originated as a Jupyter Notebook.

Other incubator & other projects

- Content management
  - Indexing
  - Search
  - Execute notebook as library
  - Execute notebook as templates
- Separate component
  - Kernel Gateway
  - Phosphorjs-*
  - nteract-*
- Interactivity
  - mpld3
  - bqplot

jupyterHub

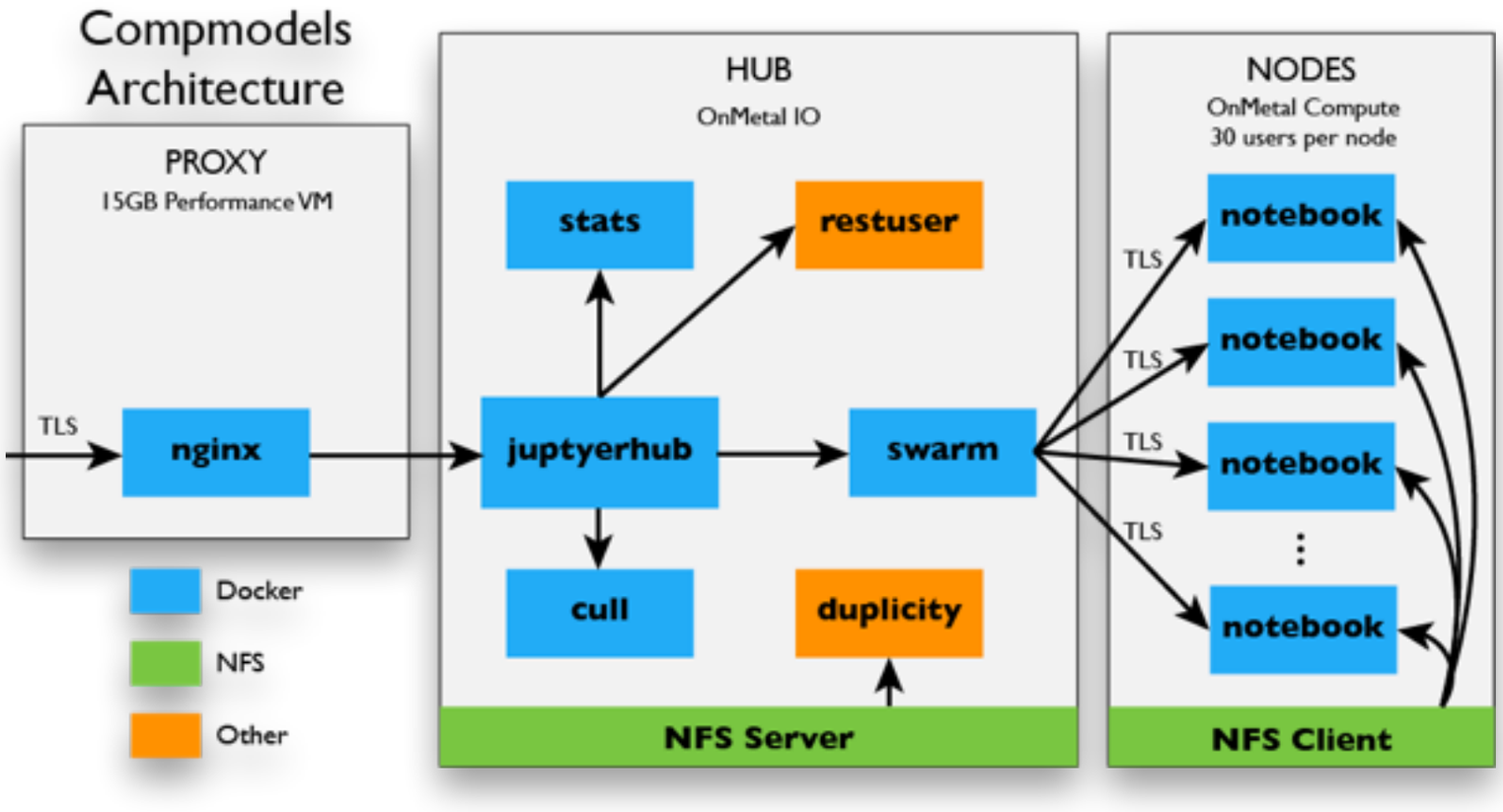
Https/websocket proxy

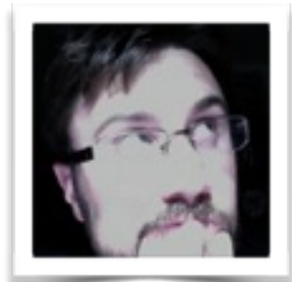


JupyterHub in education

- Computationally intensive course, ~220 students
- Fully hosted environment, zero-install
- Integration with autograding.

Jess Hamrick @ Cal





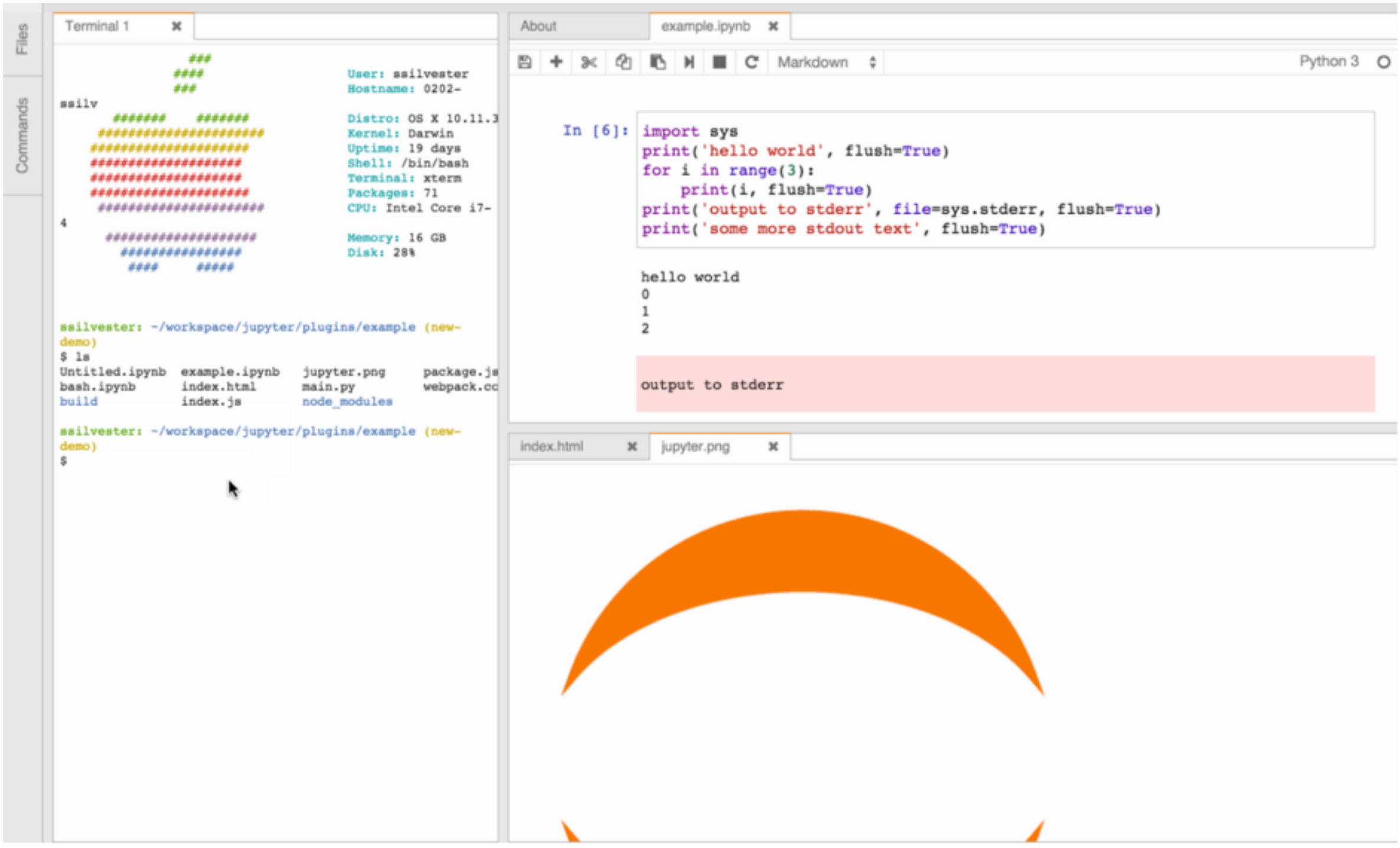
K. Kelley
Rackspace

M. Ragan-Kelley
Simula

B. Granger
Cal Poly

https://developer.rackspace.com/blog/deploying-jupyterhub-for-education

JupyterLab


Files
Commands
Terminal 1
User: ssilvester
Hostname: 0202-ssilv
Distro: OS X 10.11.3
Kernel: Darwin
Uptime: 19 days
Shell: /bin/bash
Terminal: xterm
Packages: 71
CPU: Intel Core i7-4
Memory: 16 GB
Disk: 28%
ssilvester: ~/workspace/jupyter/plugins/example (new-demo)
$ ls
Untitled.ipynb example.ipynb jupyter.png package.js
bash.ipynb index.html main.py webpack.cc
build index.js node_modules
ssilvester: ~/workspace/jupyter/plugins/example (new-demo)
$
About
example.ipynb
Markdown
Python 3
In [6]:
import sys
print('hello world', flush=True)
for i in range(3):
print(i, flush=True)
print('output to stderr', file=sys.stderr, flush=True)
print('some more stdout text', flush=True)
hello world
0
1
2
output to stderr
index.html
jupyter.png

Funding

ALFRED P. SLOAN FOUNDATION

THE LEONA M. AND HARRY B. HELMSLEY CHARITABLE TRUST

GORDON AND BETTY MOORE FOUNDATION

SIMONS FOUNDATION

POWERED BY rackspace® the open cloud company

CONTINUUM ANALYTICS

Microsoft

ENTHOUGHT SCIENTIFIC COMPUTING SOLUTIONS

Google

Bloomberg

The Team

(people that spend a noticeable amount of time on the project, subjective of course)

- **Fernando Pérez** (UC Berkeley)
- **Brian Granger** (CalPoly)
- **Min Ragan-Kelley** (Simula)
- **Thomas Kluyver** (UK)
- **Matthias Bussonnier**
- **Jon Frederic** (Cal Poly)
- **Cameron Oelsen** (Cal Poly)
- **Carole Willing** (Cal Poly)
- **Safia Abdala** (Cal Poly)
- **Ana Ruvalcaba** (Cal Poly)
- Jess Hamrick (UC Berkeley)
- Kyle Kelley (Rackspace)
- Jason Grout (Bloomberg)
- Sylvain Corlay (Bloomberg)
- Kester Tong (Google)
- Nicholas Bollweg (Continuum)
- Damián Avila (Continuum)
- Steven Silvester (Continuum)
- Chris Colbert (Continuum)
- David Willmer (Continuum)
- Peter Parente (IBM)
- Dan Gisolfi (IBM)
- Gino Bustelo (IBM)
- All 400+ GitHub contributors.

Bold: Working full time on IPython/Jupyter, underline: Contribute to Jupyter/IPython with corporate agreement

Credits

- Matthias Bussonnier, Fernando Perez, Carol Willing for slides
- Daniel Lam and Bloomberg Quant Finance Research Team for demos
- Sylvain Corlay for flight simulator demo
- Steve Silvester for JupyterLab screenshot

Thanks!

Questions?