IP[y]:
IPython

jupyter

Project Jupyter at BIDS

building the tools and the institutions for data science

Fernando Pérez
(@fperez_org & fperez@lbl.gov)

LBL & UC Berkeley

BERKELEY LAB

Berkeley
UNIVERSITY OF CALIFORNIA

BIDS
BERKELEY INSTITUTE
FOR DATA SCIENCE

A bit about me

- Particle physics, applied mathematics, neuroscience
  - Constant element: *computing in science*
- Building tools to use computers for thinking and communicating (in science).
- Building projects to change the role of computers in science
  - Open tools for scientific computing: IPython & friends…
  - The Numfocus foundation
  - BIDS: the Berkeley Institute for Data Science

“The purpose of computing is insight, not numbers”

–Hamming’62

IPython: CU Boulder, 2001

or how to best procrastinate on a Physics dissertation

```
/bin/bash

In [13]: run ~/scratch/error
reps: 5
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
/home/fperez/scratch/error.py in <module>()
     70 if __name__ == '__main__':
     71     #explode()

---> 72     main()
     73     g2='another global'

/home/fperez/scratch/error.py in main()
     60     array_num = zeros(size,'d')
     61     for i in xrange(reps):
---> 62         RampNum(array_num, size, 0.0, 1.0)
     63     RNtime = time.clock()-t0
     64     print 'RampNum time:', RNtime

/home/fperez/scratch/error.py in RampNum(result, size, start, end)
     43     tmp = zeros(size+1)
     44     step = (end-start)/(size-1-tmp)
---> 45     result[:] = arange(size)*step + start
     46
     47 def main():

ValueError: shape mismatch: objects cannot be broadcast to a single shape

In [14]:
```

November 2001: "Just an afternoon hack"

- 259 Line Python script.
- `sys.ps1` -> `In [N]`.
- `sys.displayhook` -> `Out[N]`, caches results.
- Plotting, Numeric, etc.

~2014 (Openhub stats)

- 19,279 commits
- 442 contributors
- Total Lines: 187,326
- Number of Languages : 7 (JS, CSS, HTML, ...)

A rapidly growing community

Plus ~ 500 more Open source contributors!

Current and recent 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

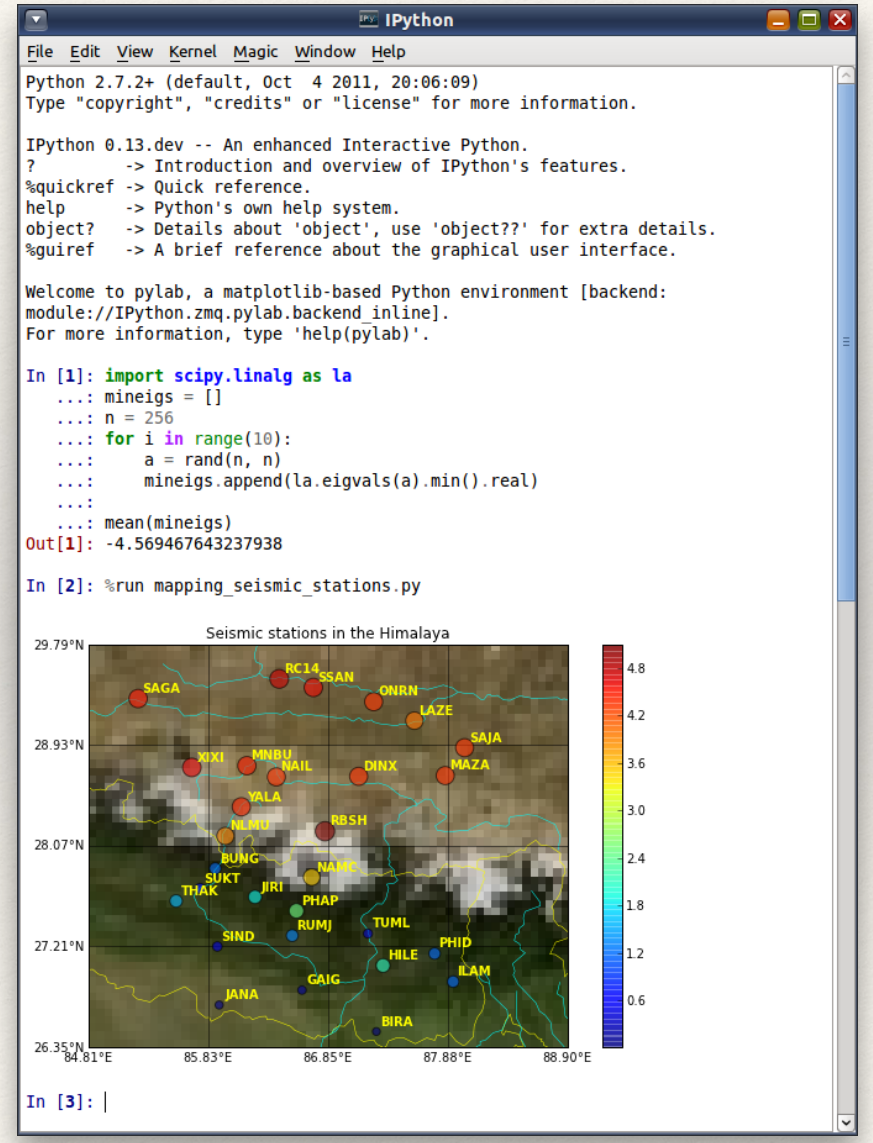
Google

Bloomberg

Beyond the Terminal…

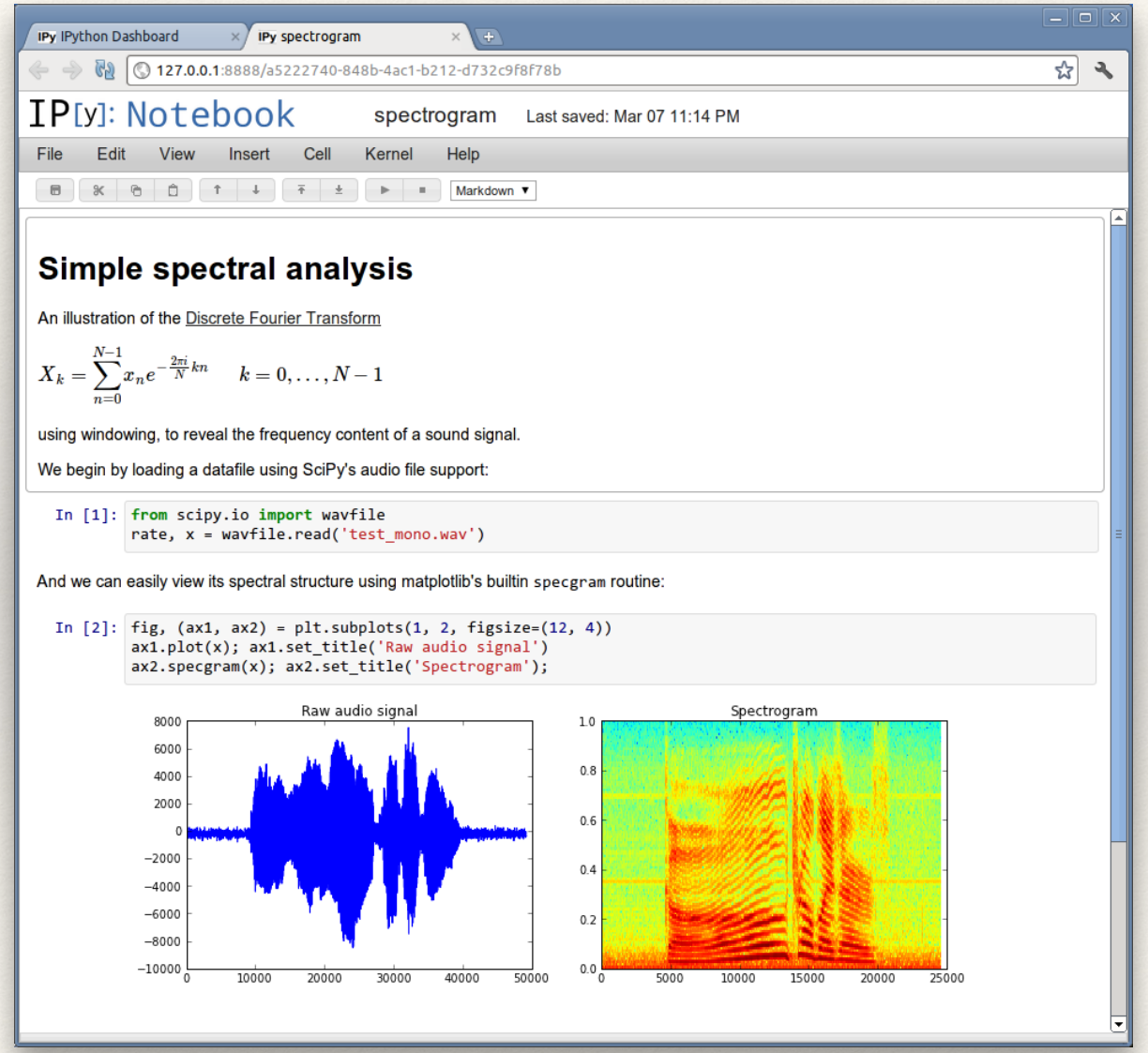
- The REPL as a network protocol
- Kernels
  - execute code
- Clients
  - Read input
  - Present output

Simple abstractions enable rich, sophisticated clients



2011: The IPython Notebook

- Rich web client
- Text & math
- Code
- Results
- Share, reproduce.



The Notebook: “Literate Computing”

Computational Narratives

- Computers deal with *code and data*.
- Humans deal with narratives that *communicate*.

Literate Computing (*not* Literate Programming)

narratives anchored in a live computation, that communicate a story based on data and results.

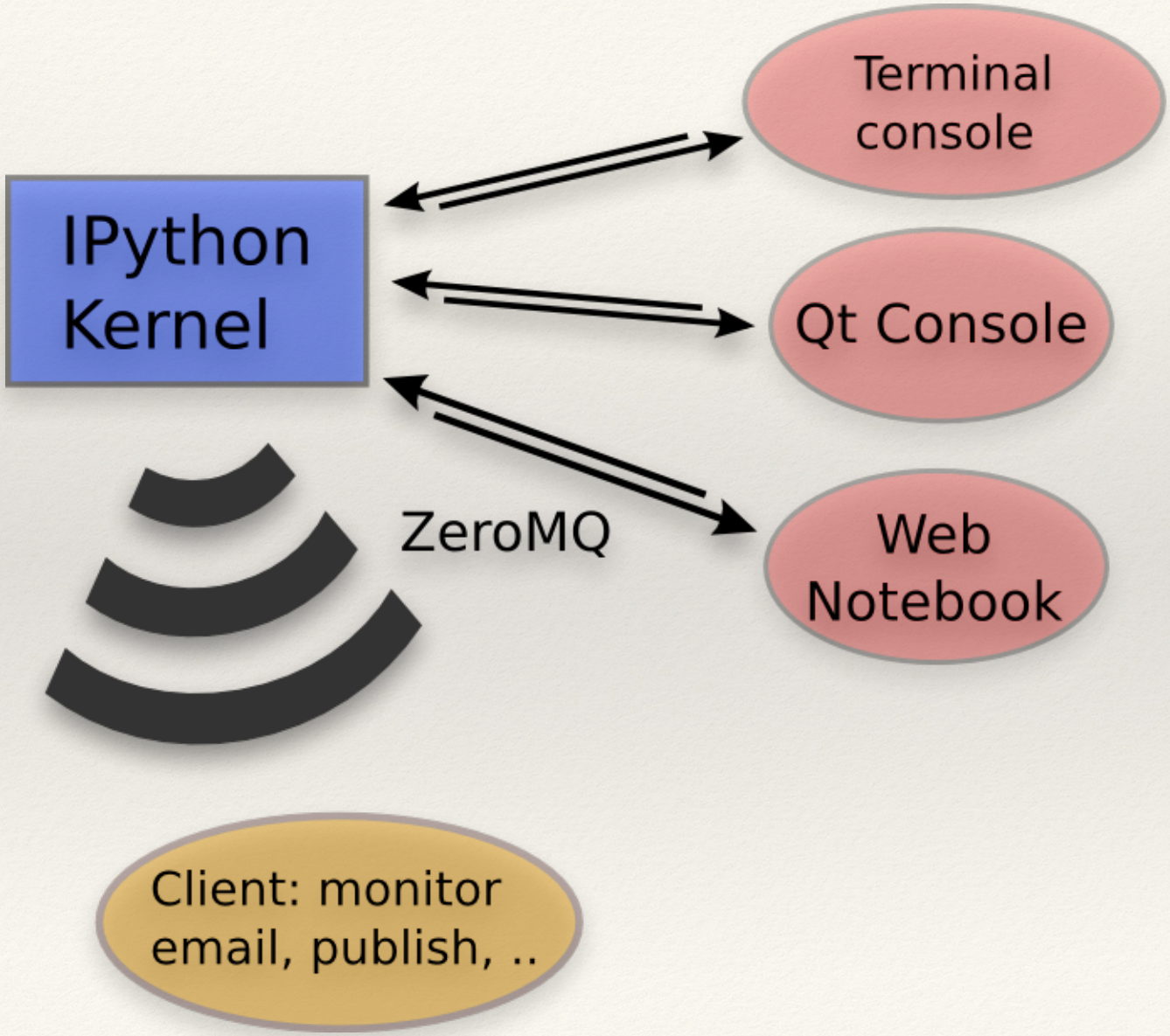
Cf: Mathematica, Maple, MuPad, Sage…

From IPython to Project Jupyter

IP[y]:
IPython

jupyter

A simple and generic architecture



Not just about Python: Kernels in any language

- IPython "Official", we ship it.
- IJulia
- IRKernel
- IHaskell
- IFSharp
- Ruby
- IScala
- IErlang
- **Lots more! ~50 and counting**

“Why is it called IPython,

if it can do Julia, R, Haskell, Ruby, … ?”

IPython

- Interactive Python shell at the terminal
- Kernel for this protocol in Python
- Tools for Interactive Parallel computing
- Network protocol for interactive computing
- Clients for protocol
  - Console
  - Qt Console
  - Notebook
- Notebook file format & tools (nbconvert...)
- Nbviewer

IPython … Jupyter

- Interactive Python shell at the terminal
- Kernel for this protocol in Python
- Tools for Interactive Parallel computing

- Network protocol for interactive computing
- Clients for protocol
  - Console
  - Qt Console
  - Notebook
- Notebook file format & tools (nbconvert...)
- Nbviewer

Language Agnostic

What's in a name?

- *Inspired* by the open languages of science:
  - Julia, Python & R
  - *not* an acronym: *all languages* equal class citizens.
- Astronomy and Scientific Python:
  - A long and fruitful collaboration
- Galileo's notebooks:
  - the original, open science, data-and-narrative papers
  - Authorea: "Science was Always meant to be Open"

The Jupyter Notebook Ecosystem

nbviewer: seamless notebook sharing

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

Reproducible Research (2012): Paper, Notebooks and Virtual Machine

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

The ISME Journal
Multidisciplinary Journal of Microbial Ecology

Login Cart

Search go Advanced search

Journal home > Archive > Commentaries > Full text

Journal home
Advance online publication
About AOP
Current issue
Archive
Focuses

Commentary

The ISME Journal (2013) **7,** 461–464; doi:10.1038/ismej.2012.123; published online 25 October 2012

Collaborative cloud-enabled tools allow rapid, reproducible biological insights

Open

FULL TEXT
Previous | Next
Table of contents
Download PDF
Send to a friend
View interactive PDF in

agan-Kelley[1,12], William Anton Walters[2,12],
onald[3,6,12], Justin Riley[4], Brian E Granger[5],
nzalez[6], Rob Knight[7,8], Fernando Perez[9] and J
poraso[10,11]

Group in Applied Science and Technology, University of
t Berkeley, Berkeley, CA, USA
nt of Molecular, Cellular and Developmental Biology,
f Colorado at Boulder, Boulder, CO, USA
's Institute, University of Colorado at Boulder, Boulder, CO,

ducational Innovation and Technology, Massachusetts
Technology, Cambridge, MA, USA
epartment, California Polytechnic State University, San Luis
USA
nt of Computer Science, University of Colorado at Boulder,
), USA

This notebook is intended to calculate the positions of primers in an alignment, using functions from PrimerProspector.

Import the needed functions, and define the primer sequences

In [8]:
```
# Code modified from PrimerProspector library slice_aligned_region.py (development version
)

# Imports and definitions
from string import lower, upper
from operator import itemgetter

from cogent import LoadSeqs, DNA
from cogent.core.alphabet import AlphabetError
from cogent.align.align import make_dna_scoring_dict, local_pairwise
from cogent.parse.fasta import MinimalFastaParser
from cogent.core.moltype import IUPAC_DNA_ambiguities

DNA_CODES = ['A', 'C', 'T', 'G', 'R', 'Y', 'M', 'K',
 'W', 'S', 'B', 'D', 'H', 'V', 'N']


# Note that these are all written 5'->3', the reverse primers are reverse complemented for
 the local alignment

# If one wanted to test different primers, they would be defined here.

# 27f/338r = V2 (also includes V1, but generally just referred to as V2)
# 349f/534r = V3
# 515f/806r = V4
# 967f/1046r = V6
# 1391f/1492r = V9

primer_seqs = {
            '27f':'AGAGTTTGATCMTGGCTCAG',
            '338r':DNA.rc('GCTGCCTCCCGTAGGAGT'),
            '349f':'GYGCASCAGKCGMGAAW',
            '534r':DNA.rc('ATTACCGCGGCTGCTGG'),
            '515f':'GTGCCAGCMGCCGCGGTAA',
            '806r':DNA.rc('GGACTACVSGGGTATCTAAT'),
            '967f':'CAACGCGAAGAACCTTACC',
            '1048r':DNA.rc('CGRCRGCCATGYACCWC'),
            '1391f':'TGYACACACCGCCCGTC',
            '1492r':DNA.rc('GGCTACCTTGTTACGACTT'),
            '1391r':'TGYACACACCGCCCGTC' # Need this rather than forward primer to get prope
r 3' position of reverse version
            }

reference_aligned_file = '/home/ubuntu/qiime_software/gg_otus-4feb2011-release/rep_set/gg_
76_otus_4feb2011_aligned.fasta'
```

Instructions and supporting data for the QIIME/IPython/StarCluster demo at the *2012 NIH Cloud Computing the Microbiome* workshop and our corresponding paper in the ISME Journal.

The analysis made use of the IPython Notebook, QIIME, StarCluster, PyCogent, and PrimerProspector. All of these tools are pre-installed in the `ami-9f69c1f6` public Amazon EC2 instance, which was used in this study.

Supporting Files

The IPython notebooks supporting this study can be viewed here and are available here in PDF format:

- NIH Cloud Demo (Complete)
- NIH Cloud Demo (Fast)
- Timing*
- Variable Region Position Boundaries
- Pearson v Robinson-Foulds Distances
- V3 and V4 Regions Only

* Note that the Timing notebook is for reference as related to the paper only - it will not be directly reproducible on re-runs of the above notebooks as it relies on the semi-manual creation of the `tasks.log` file. The `tasks.log` file used to generate the original timing data is available for download here.

The Greengenes reference OTU collection used in this study is available for download here.

The IPython notebook files (`.ipynb`) are available for download here.

The tree metadata mapping file used in generating the coloring categories in the 3D PCoA plot is available here.

The paper for this analysis, "Collaborative cloud-enabled tools allow rapid, reproducible biological insights", is available here.

Reproducing the analysis

Four `m2.4xlarge` instances were booted using StarCluster to create a 32 core cluster with approximately 280GB of RAM (70GB per 8 core instance). This was used for the full analysis (a more complete analysis then was done during the workshop, where the workshop analysis was optimized to run quickly). To support the large quantity of data that is generated during the analysis, you should create an EBS volume which will be attached to the running instance. A 20 GB volume will be sufficient. The volume used for running these notebooks is available as `snap-75eb8005`.

To reproduce the analyses presented in this paper you should install StarCluster locally, and configure it according to the instructions on the StarCluster website. You can then add the following to your `~/.starcluster/config` file:

```
[plugin ipcluster]
setup_class = starcluster.plugins.ipcluster.IPCluster
enable_notebook = true
# If you leave notebook_passwd out, a random password
# will be generated instead.
notebook_passwd = YOUR-PASSWORD

[cluster qiime-ipython]
node_image_id = ami-9f69c1f6
cluster_user = ubuntu
keyname = YOUR-KEY
cluster_size = 4
node_instance_type = m2.4xlarge
plugins = ipcluster
volumes = qiime-ipython-data

[volume qiime-ipython-data]
VOLUME_ID = YOUR-VOLUME-ID
MOUNT_PATH = /home/ubuntu/data
```

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

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

Today: mybinder.org

binder

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

github.com/freeman-lab

1 **Tell us your GitHub repo**

user/project OR github url

2 **Configure dependencies**

> none for basic Python projects

requirements.txt for pip Python projects

environment.yml for conda Python projects

Dockerfile for custom builds

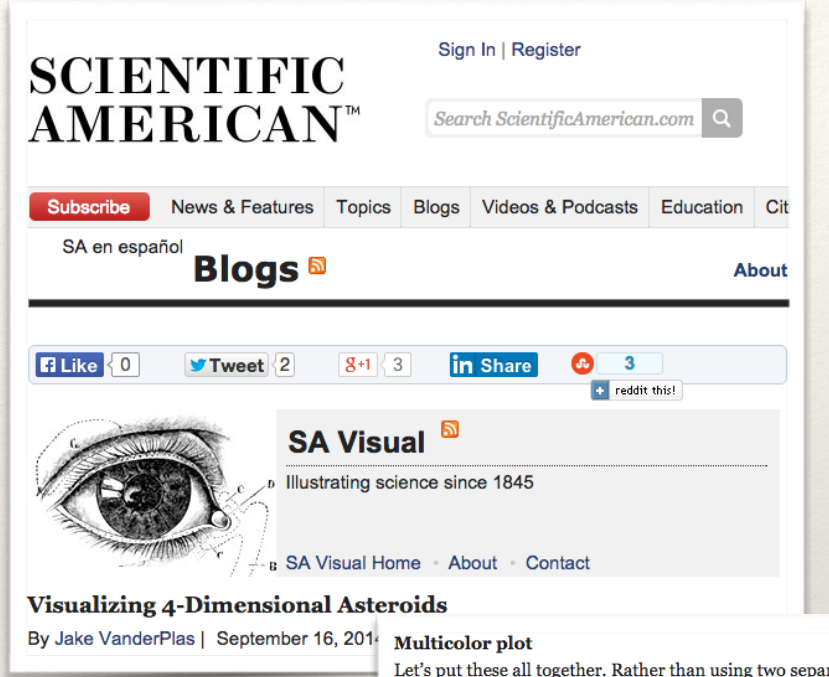
3 **Attach services**

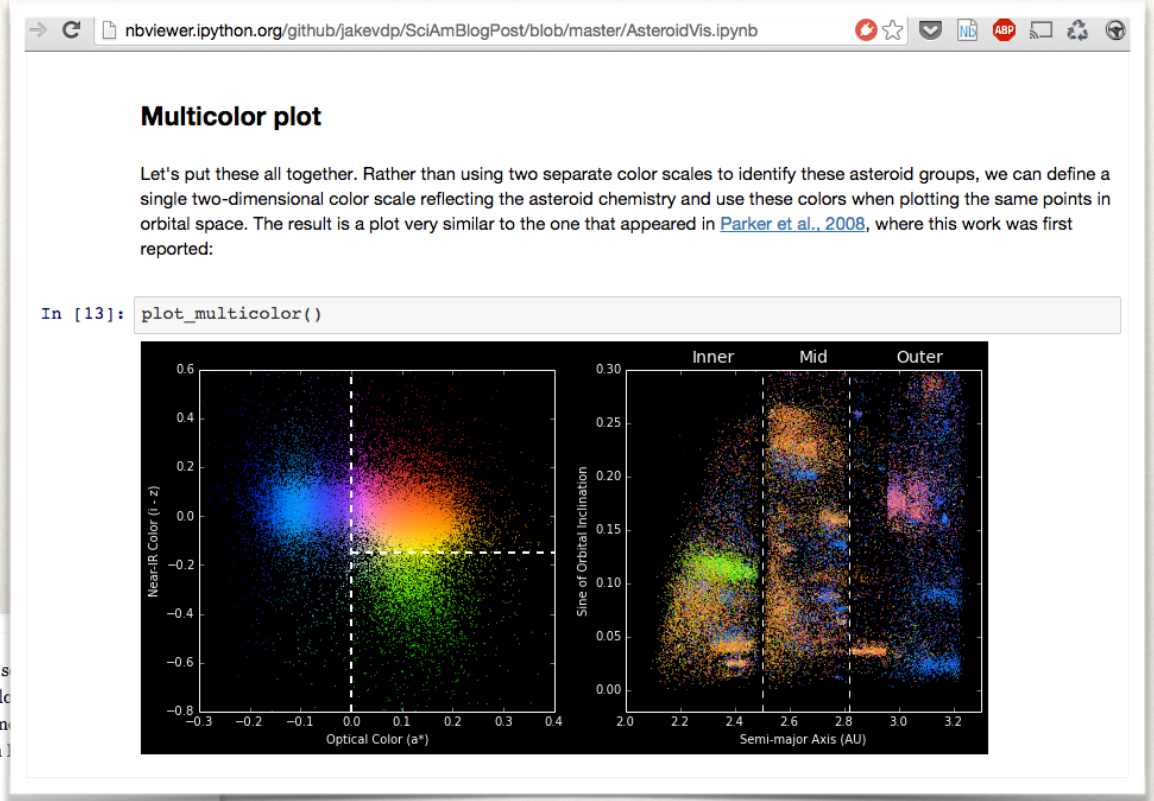
Postgres

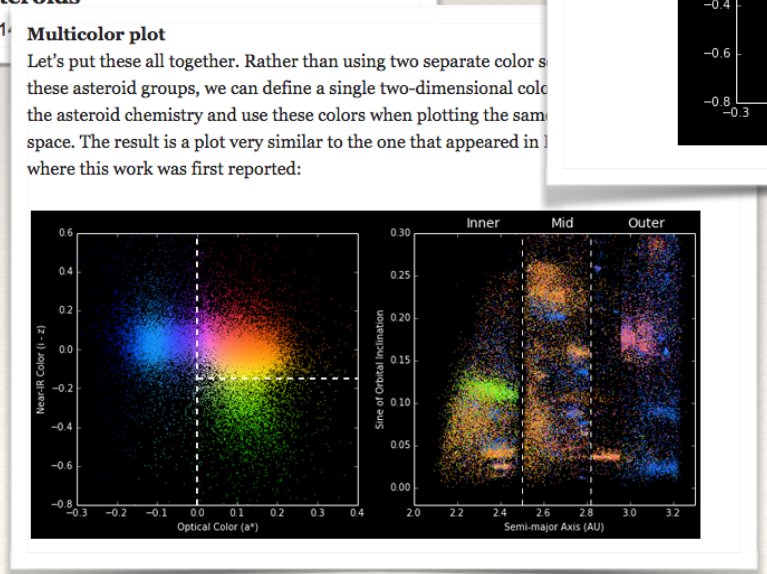
Spark

make my binder

Scientific Blogging







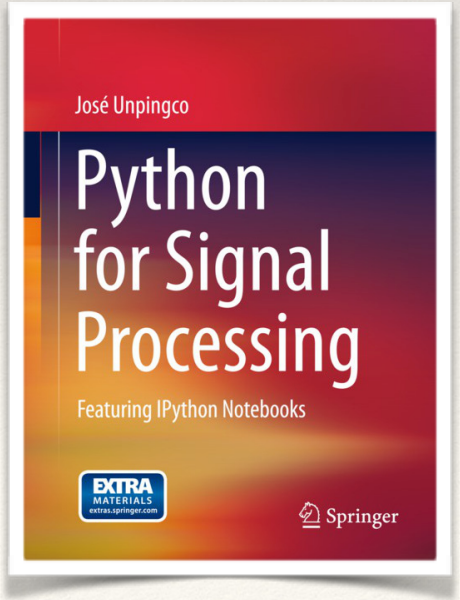
Jake van der Plas @ UW

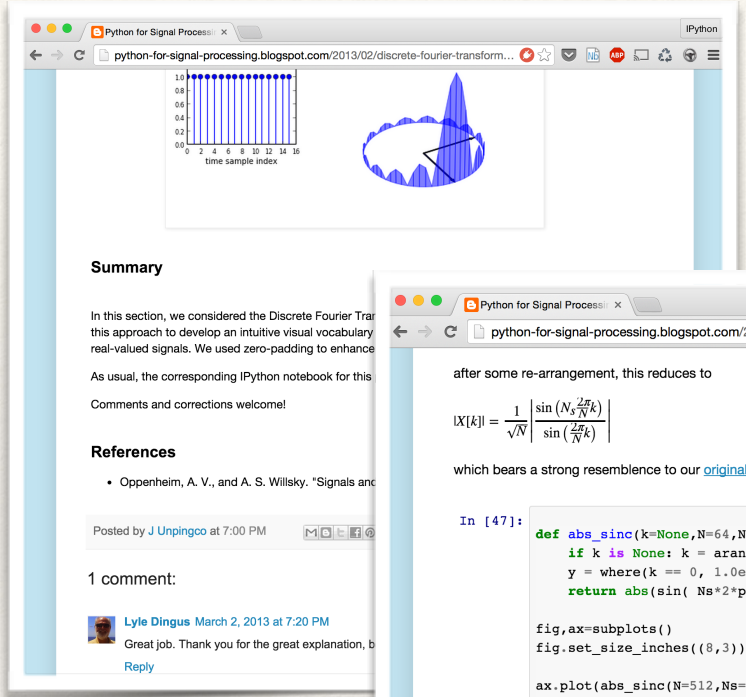
http://blogs.scientificamerican.com/sa-visual/2014/09/16/visualizing-4-dimensional-asteroids

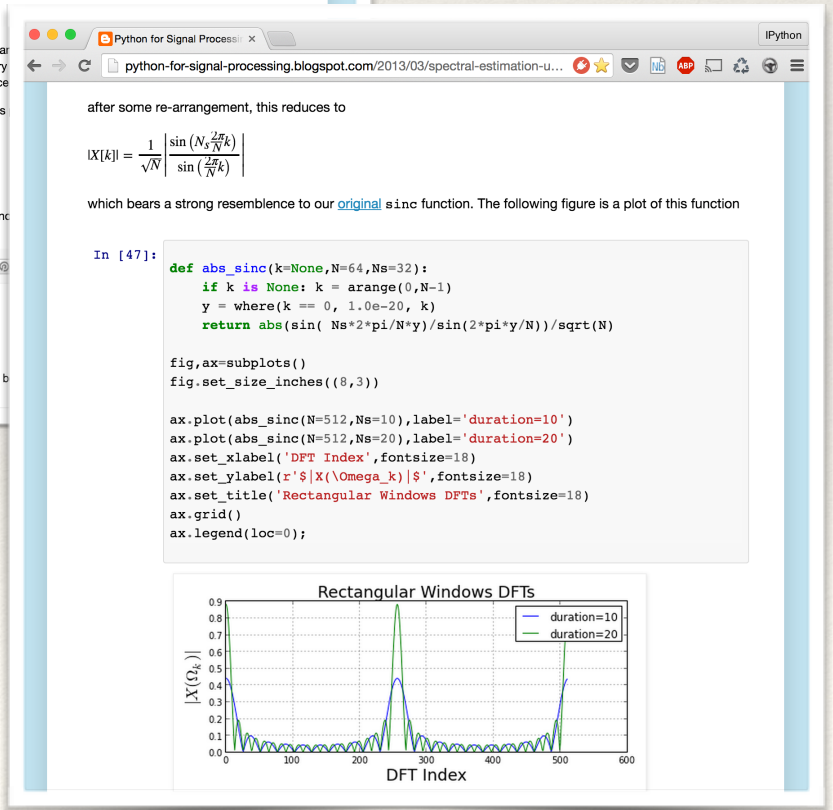
Executable books

Python for Signal Processing, by José Unpingco

- Springer hardcover book
- Chapters: IPython Notebooks
- Posted as a blog entry
- All available as a Github repo







University Courses

	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

These are just some we are aware of!

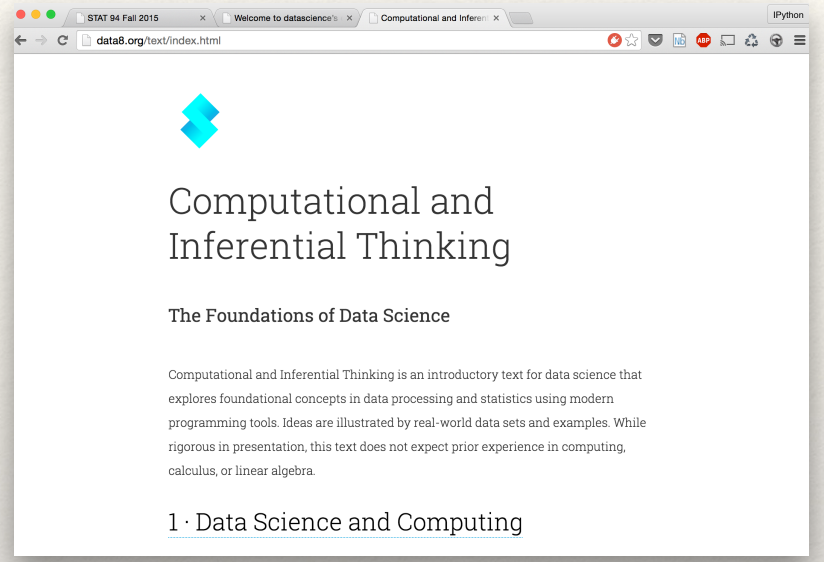
New Jupyter in Education Mailing List:

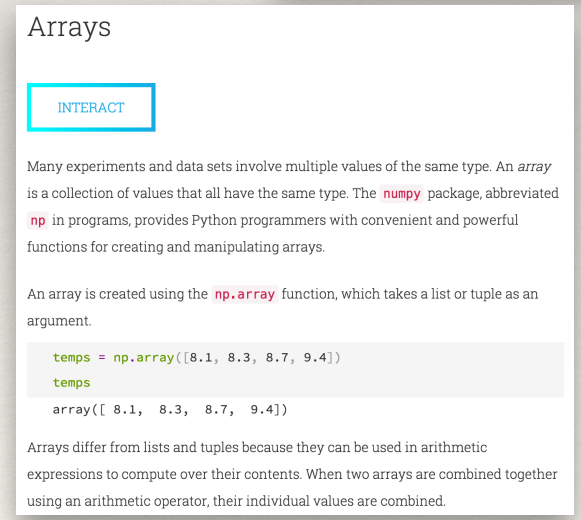
https://groups.google.com/forum/#!forum/jupyter-education

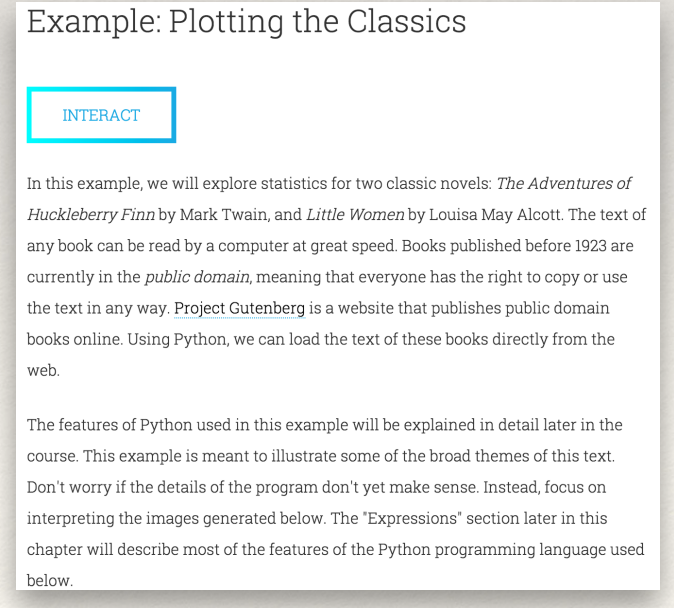
Berkeley's *Foundations of Data Science*

- New curriculum aimed at all freshmen at UC Berkeley
- Interactive textbook is Jupyter Notebooks
- Course deployment is JupyterHub

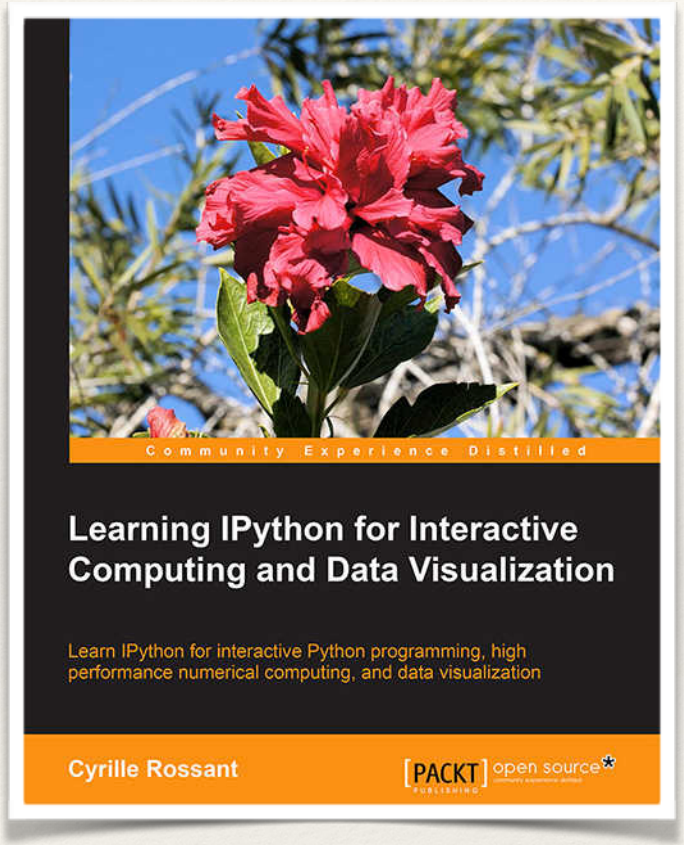
http://data8.org



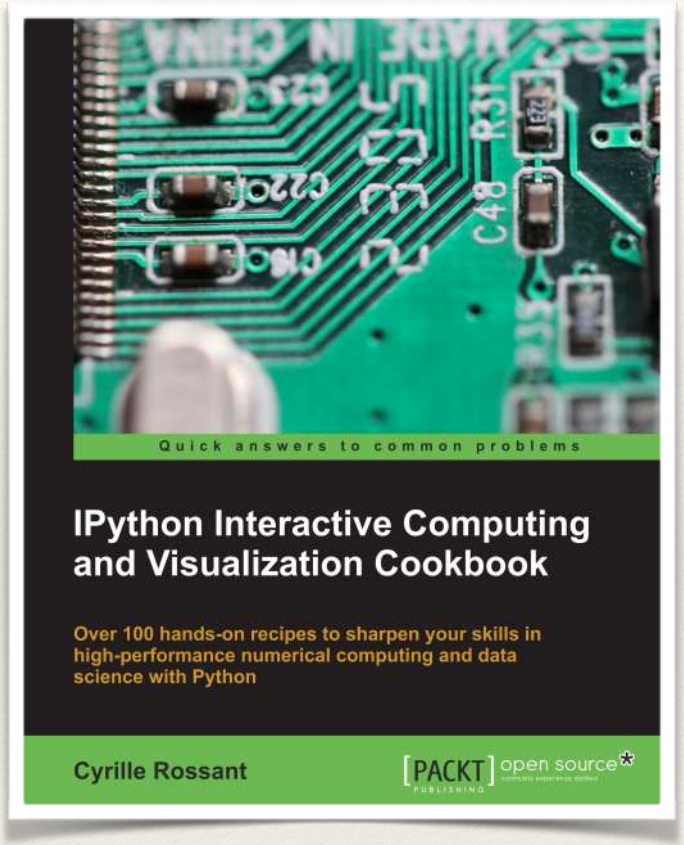




Books about IPython



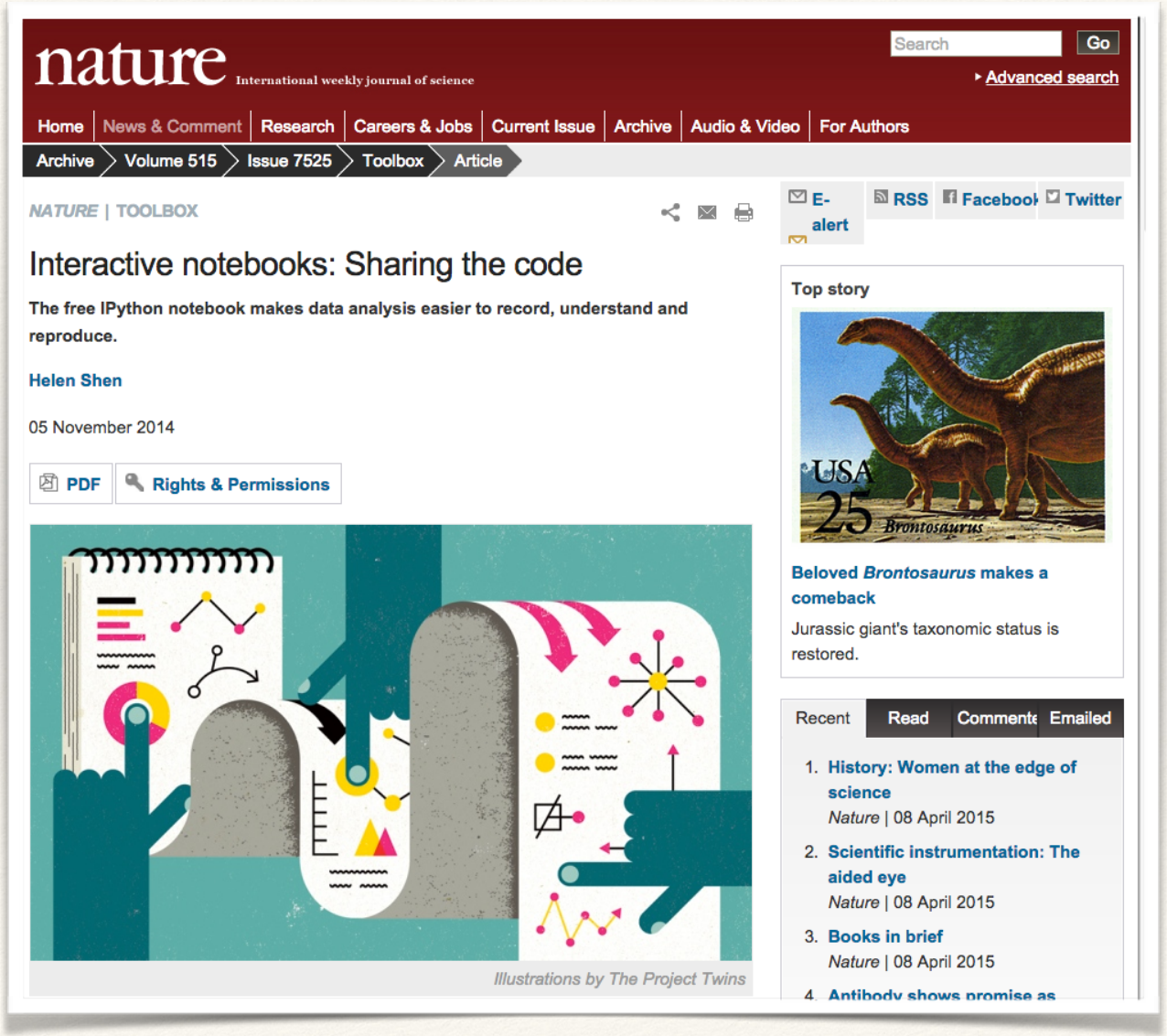
Learning IPython for Interactive Computing and Data Visualization



IPython Interactive Computing and Visualization Cookbook

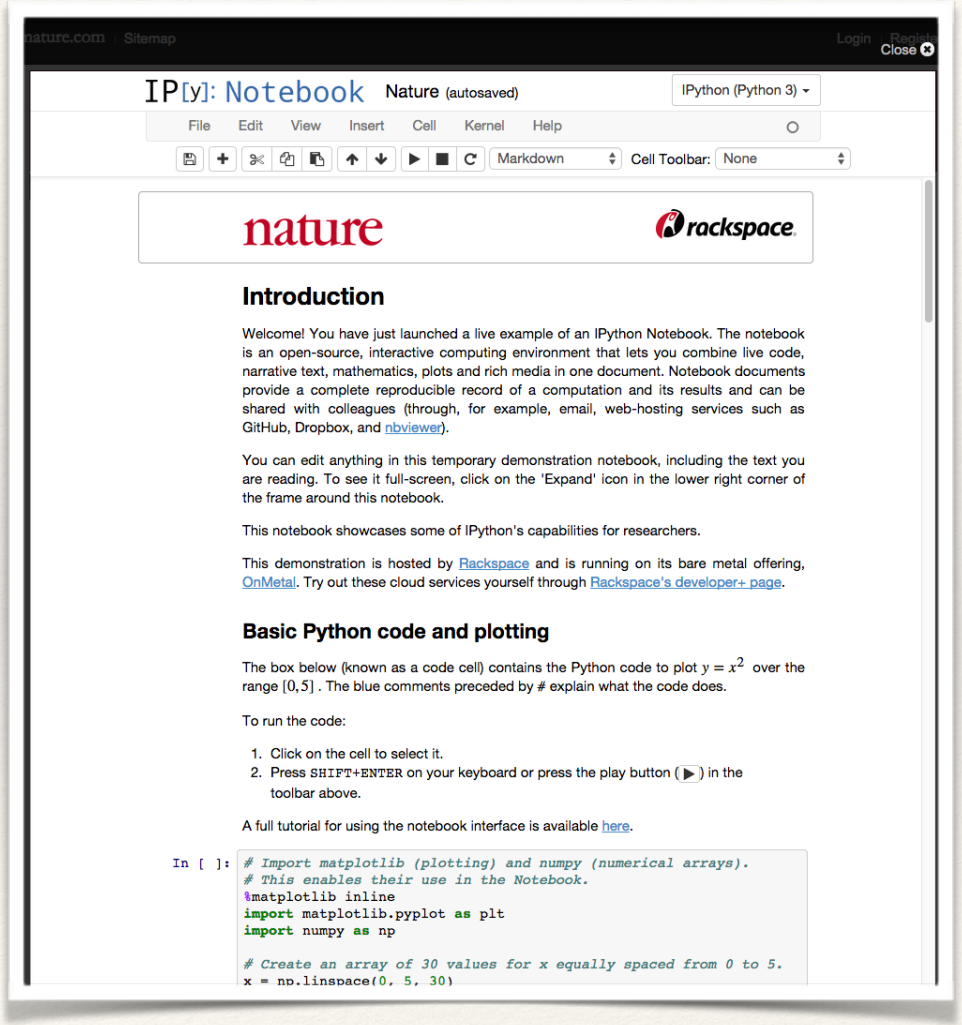
Cyrille Rossant
cyrille.rossant.net

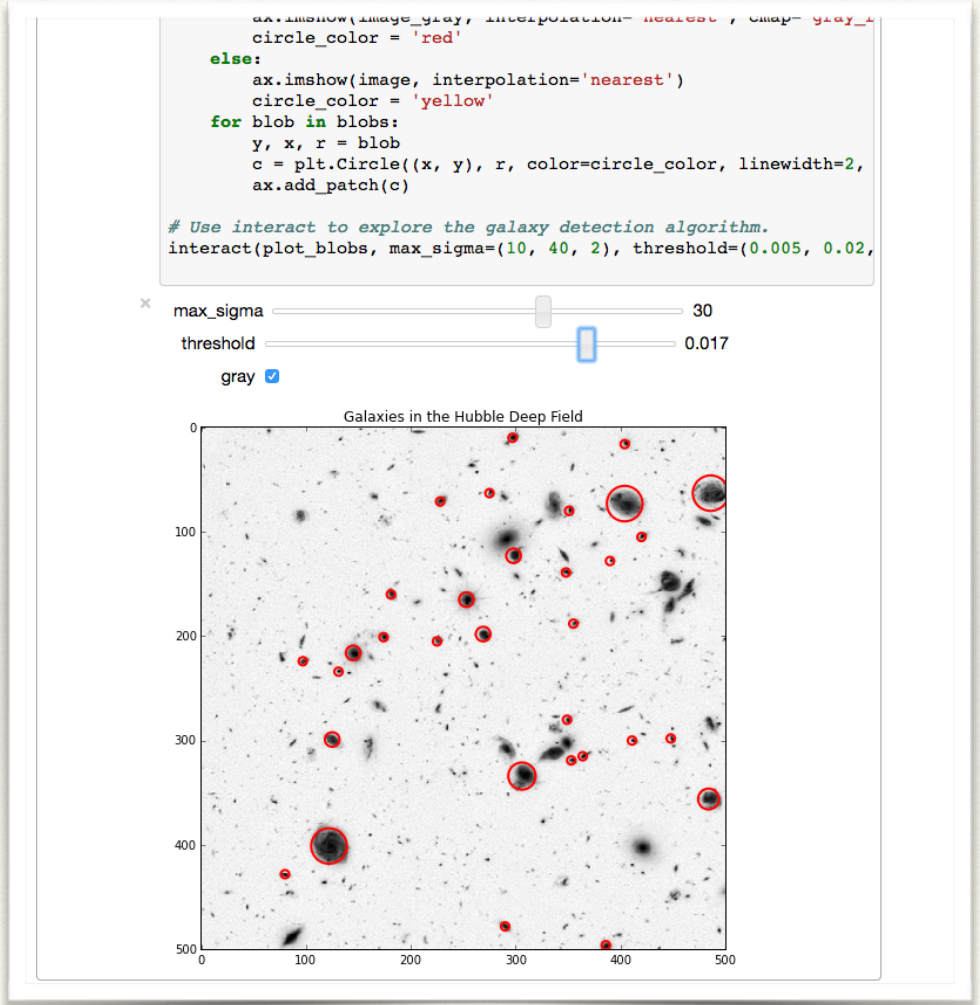
Changing the scientific culture



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

Executable papers: the future?





http://www.nature.com/news/ipython-interactive-demo-7.21492?article=1.16261

Notebook Workflows: The Big Picture

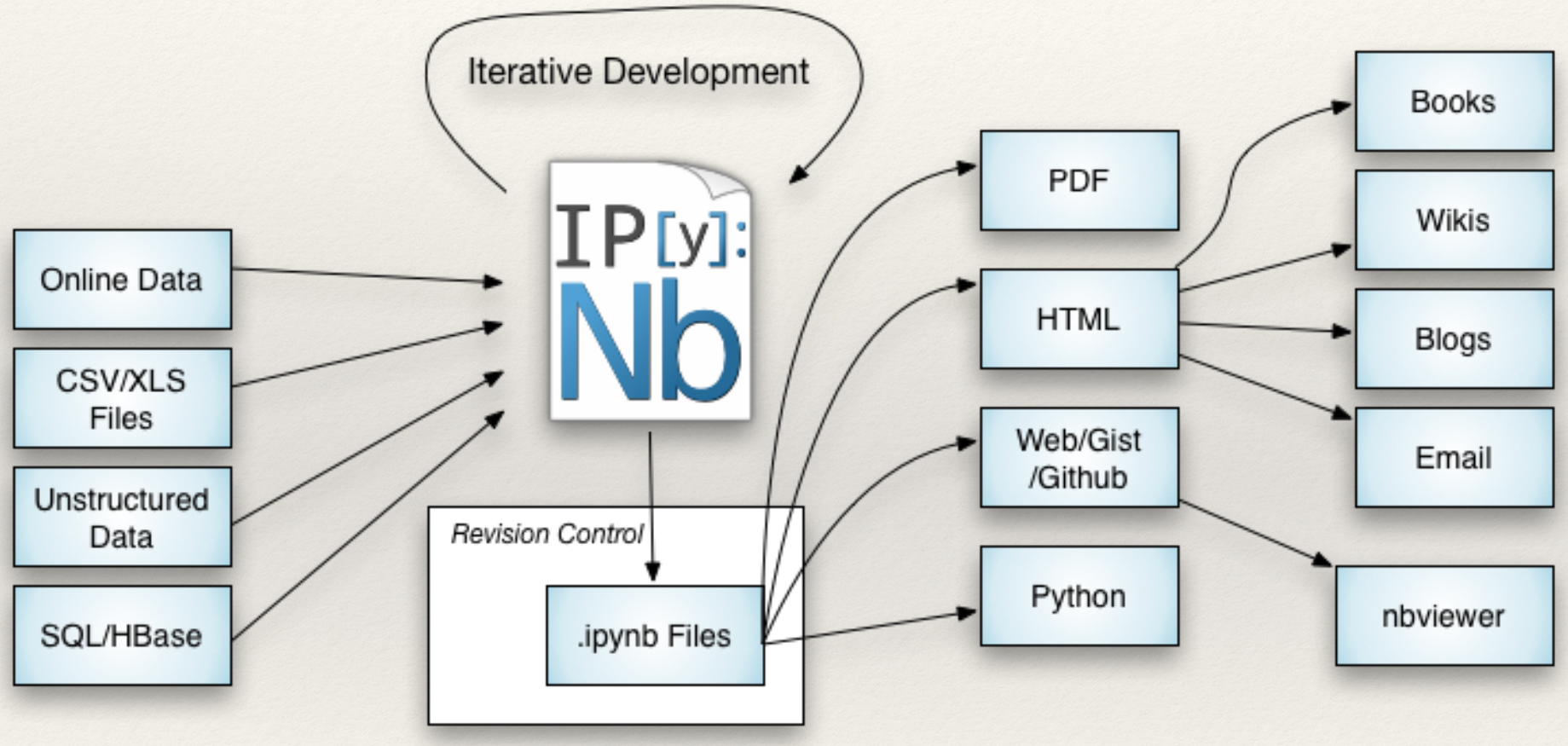


Image credit: Joshua Barratt

Lots more! The IPython 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

Growth: opportunity and challenge

Project Jupyter gets $6M to expand collaborative data science software

JULY 7, 2015

Tags: CRD

PALO ALTO, Calif. July 7, 2015 — Three foundations pledged $6M over the next three years to Project Jupyter, an open-source software project that supports scientific computing and data science across a wide range of programming languages via a large, public, open and inclusive community.

Fernando Perez and Brian Granger discuss the architecture of Project Jupyter, as its scope expands to reach data science applications in over 40 programming languages. Photo credit: Adriana Restrepo

Fernando Perez of University of California, Berkeley and Lawrence Berkeley National Laboratory (Berkeley Lab's) Computational Research Division and Brian Granger of California Polytechnic University, San Luis Obispo will lead the project at their institutions. Perez and Granger's efforts with Project Jupyter are the result of their work developing IPython, a popular user interface for interactive computing across multiple programming languages.

With this award from the Leona M. and Harry B. Helmsley Charitable Trust, Alfred P. Sloan Foundation, and Gordon and Betty Moore Foundation, these researchers will expand and improve the capabilities of the Jupyter Notebook, a web-based platform that allows scientists, researchers and educators to combine live code, equations, narrative text and rich media into a single, interactive document.

JupyterHub: multiuser support

Jupyter for Organizations

JupyterHub is a multiuser version of the notebook designed for centralized deployments in companies, university classrooms and research labs.

Pluggable authentication

Manage users and authentication with PAM, OAuth or integrate with your own directory service system. Collaborate with others through the Linux permission model.

Centralized deployment

Deploy the Jupyter Notebook to all users in your organization on centralized servers on- or off-site.

Container friendly

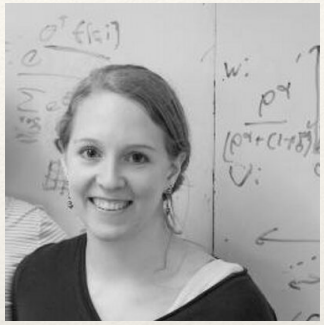
Use Docker containers to scale your deployment and isolate user processes using a growing ecosystem of prebuilt Docker containers.

Code meets data

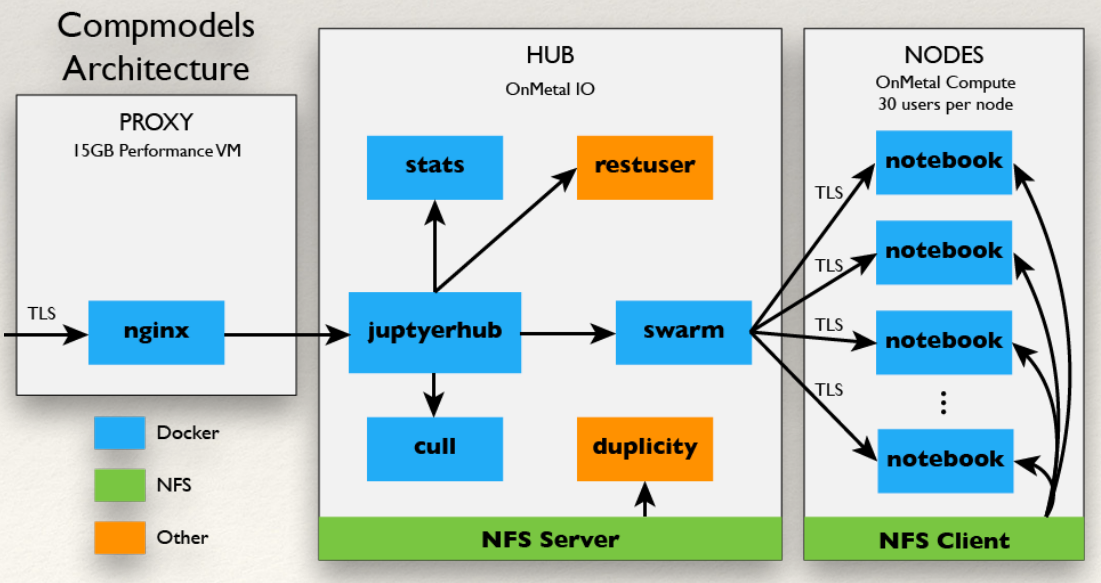
Deploy the Notebook next to your data to provide unified software management and data access within your organization.

JupyterHub in Education @ Berkeley

- Computationally intensive course, ~220 students
- Fully hosted environment, zero-install, spring 2015.
- Homework management and grading (w B. Granger)
- Now powers data8.org - Cal's new *Foundations of Data Science*, (fall 2015).



Jess Hamrick @ Cal



K. Kelley
Rackspace

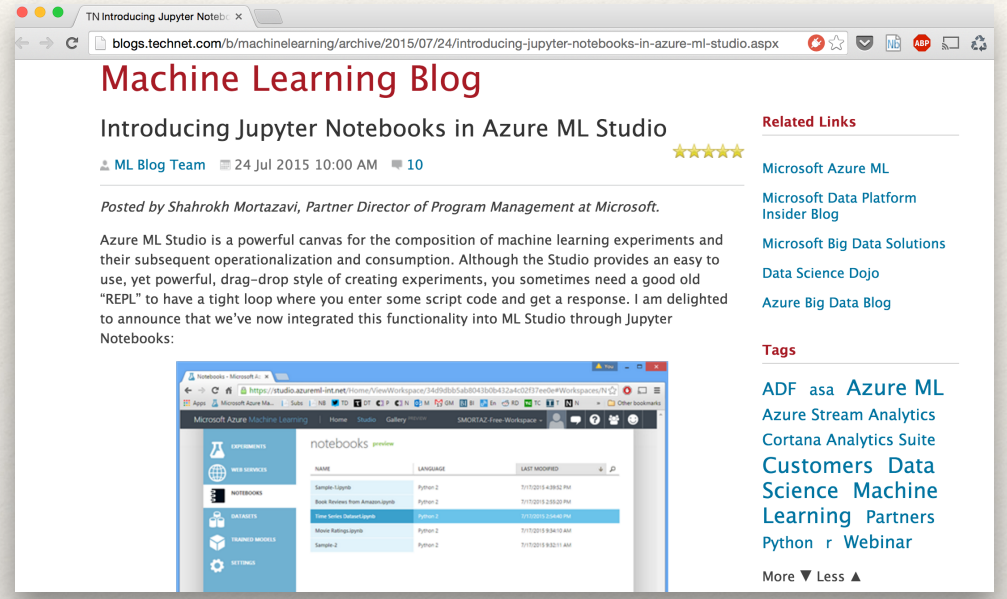
M. Ragan-Kelley
Cal

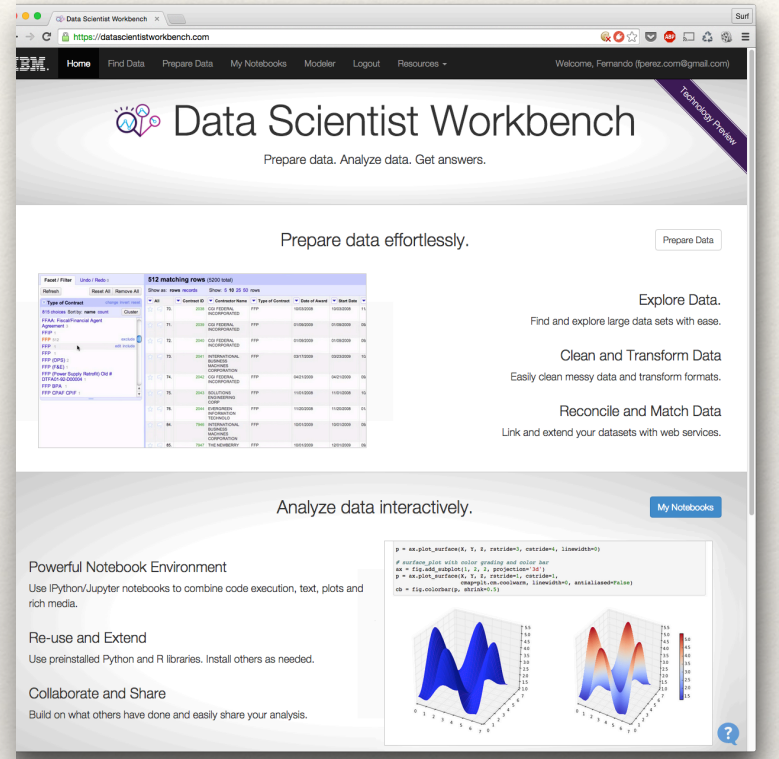
B. Granger
Cal Poly

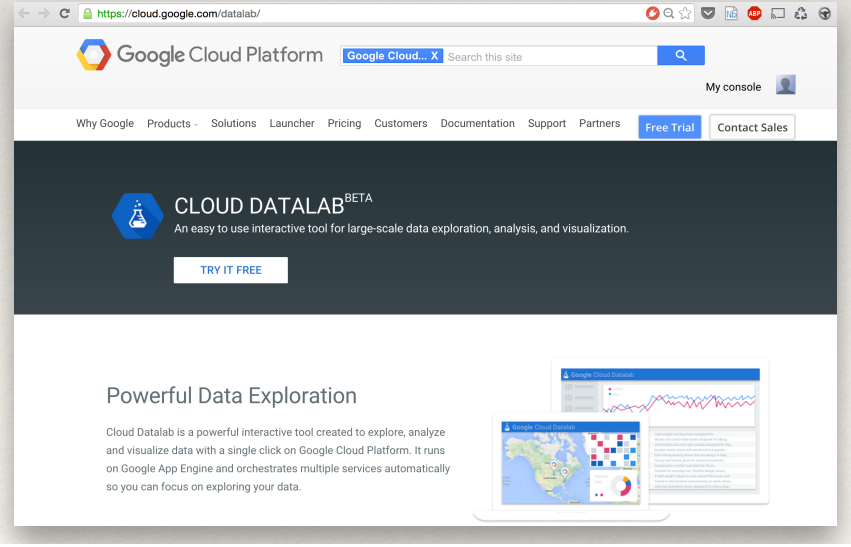
POWERED BY rackspace
the open cloud company

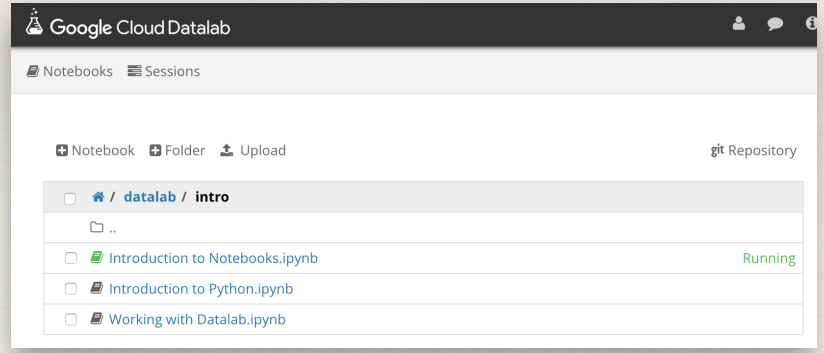
https://developer.rackspace.com/blog/deploying-jupyterhub-for-education

Industry: Microsoft, IBM, Google, …



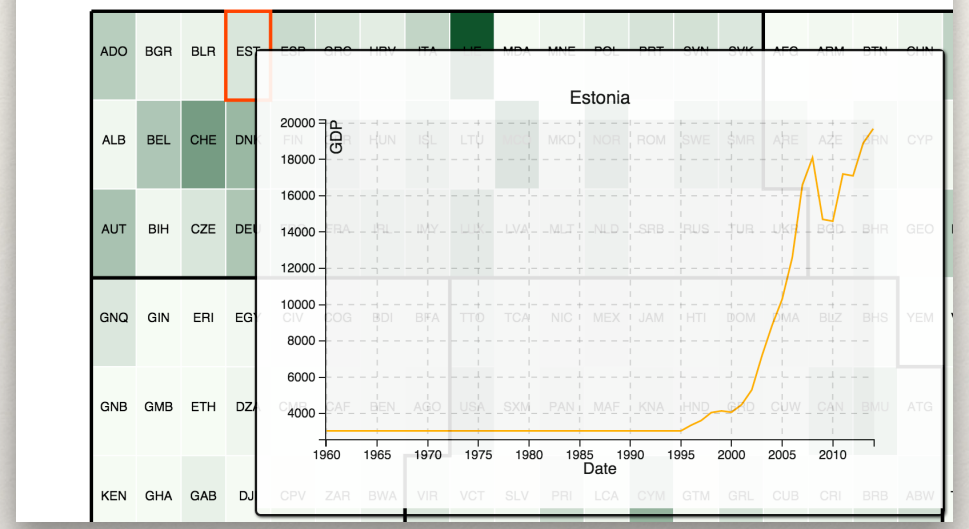




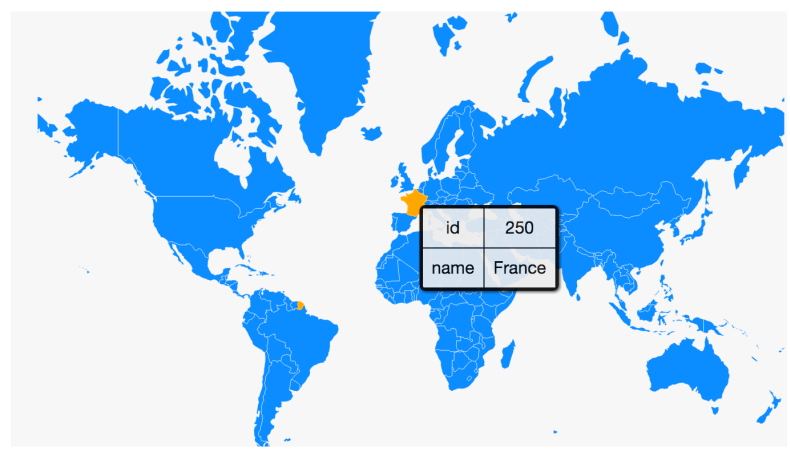


bqplot: interactive plotting from Bloomberg

```
# Custom msg sent when a particular cell is hovered on
market_map.on_hover(hover_handler)
display(market_map)
```




```
def_tt = Tooltip(fields=['id', 'name'])
map_mark = Map(map_data=topo_load('WorldMapData.json'),
               scales={'projection': Mercator()}, tooltip=def_tt)
map_mark.interactions = {'click': 'select', 'hover': 'tooltip'}
Figure(marks=[map_mark])
```



http://github.com/bloomberg/bqplot

O'Reilly Thebe: kernels as a service

Embracing Jupyter Noteboo

https://beta.oreilly.com/ideas/jupyter-at-oreilly

O'REILLY Ideas Learning Events Shop

ON OUR RADAR DATA MONEY ECONOMY

DATA

Embracing Jupyter Notebo O'Reilly

O'Reilly Media is using our Atlas platform to make Jupyter Notebooks class authoring environment for our publishing program.

By Andrew Odewahn, May 7, 2015

Embracing Jupyter Notebooks at O'Reilly

O'Reilly Media is thrilled to announce that we're making IPython Notebooks a first-class authoring environment for our publishing program, on par with Word or our Atlas platform. As part of our move to

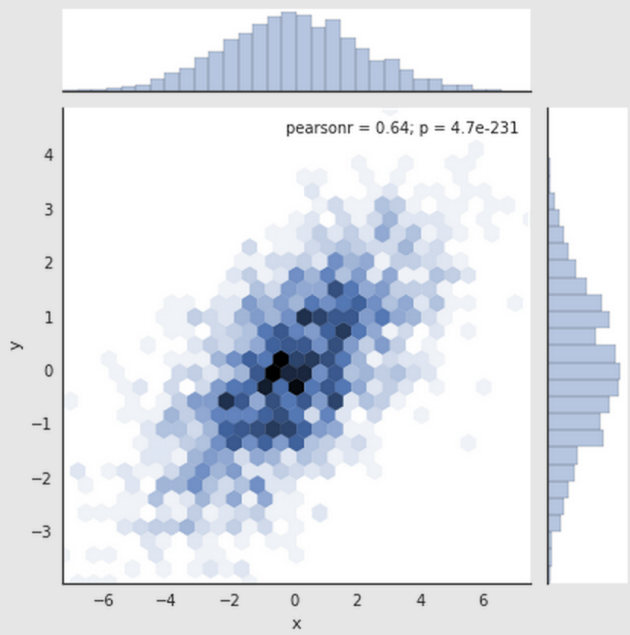
Data visualization with Sea

https://beta.oreilly.com/learning/data-visualization-with-seaborn

idle

There are other parameters which can be passed to `jointplot`: for example, we can use a hexagonally-based histogram instead:

```
with sns.axes_style('white'):
    sns.jointplot("x", "y", data, kind='hex')
```



done

idle

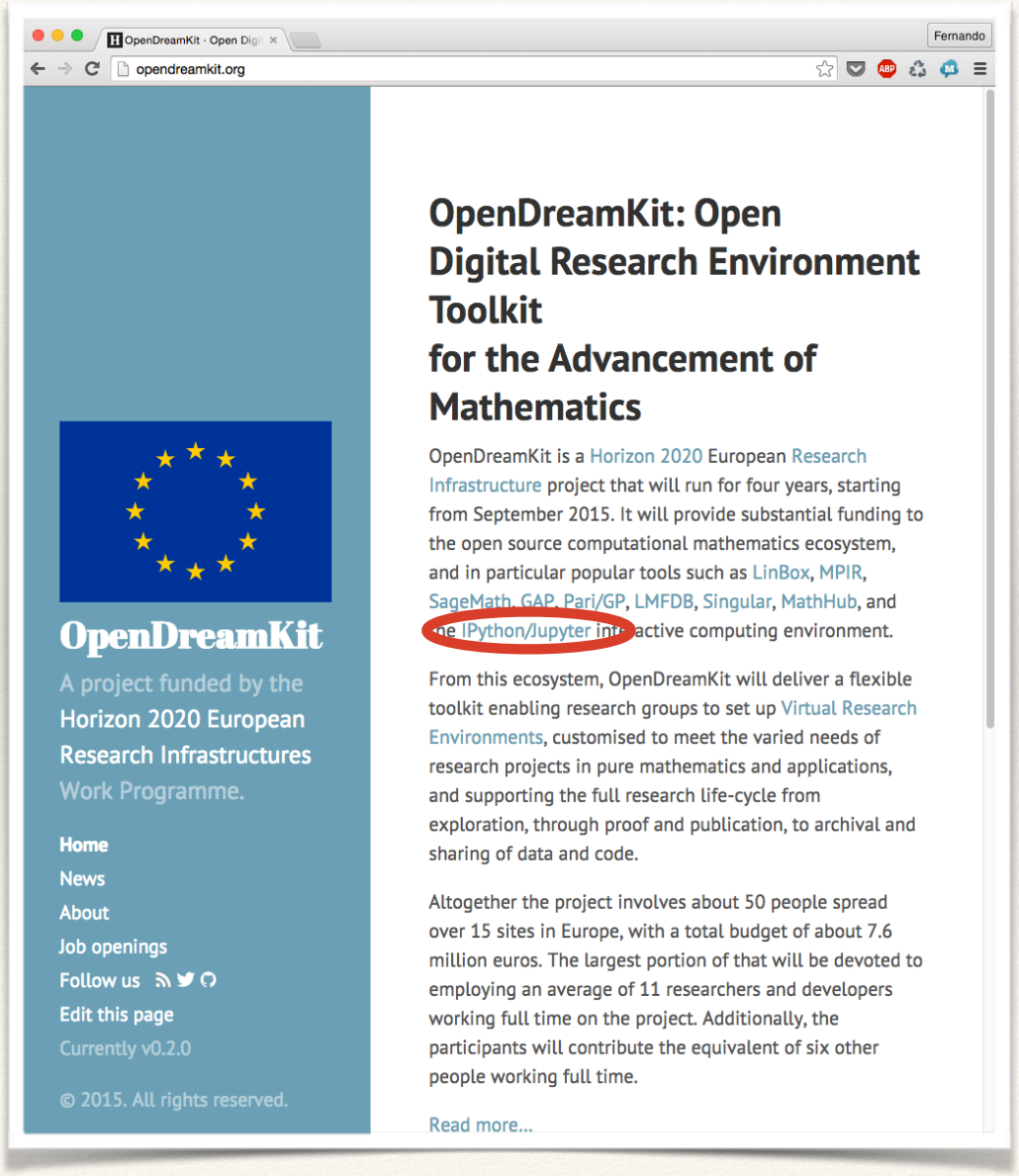
Project Jupyter: Computational Narratives as the Engine of Collaborative Data Science

- **Interactive Computing**
  - Notebooks as interactive applications
  - Modular, reusable UI/UX
  - Software engineering with notebooks
- **Computational Narratives**
  - nbconvert
  - Element filtering
  - Documentation
- **Collaboration**
  - Real time collaboration
  - JupyterHub
- **Sustainability**
  - People
  - Events

UI refactor: Bloomberg/Continuum

- Using phosphor JS framework:
  - https://github.com/phosphorjs/phosphor
- Tiled layout, plugins, much more
- Enable richer layouts, beyond the notebook
  - Text editor, output, variable inspectors, debuggers, …

A notable new European Collaboration



This presents challenges!

- Keeping a healthy community growing
- Engaging newcomers and volunteers
- Managing scope and complexity
- Documenting architecture, process and community

which brings me to…

we also need the right spaces to meet these challenges!

BERKELEY INSTITUTE FOR DATA SCIENCE

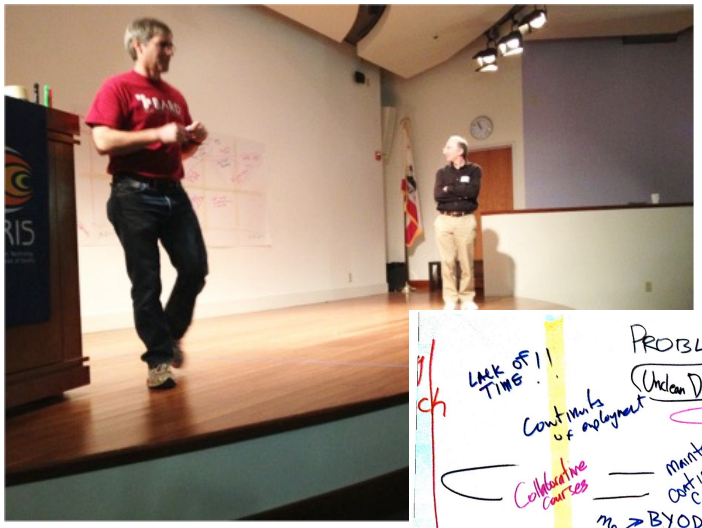
Advancing scientific discovery through collaboration across research domains

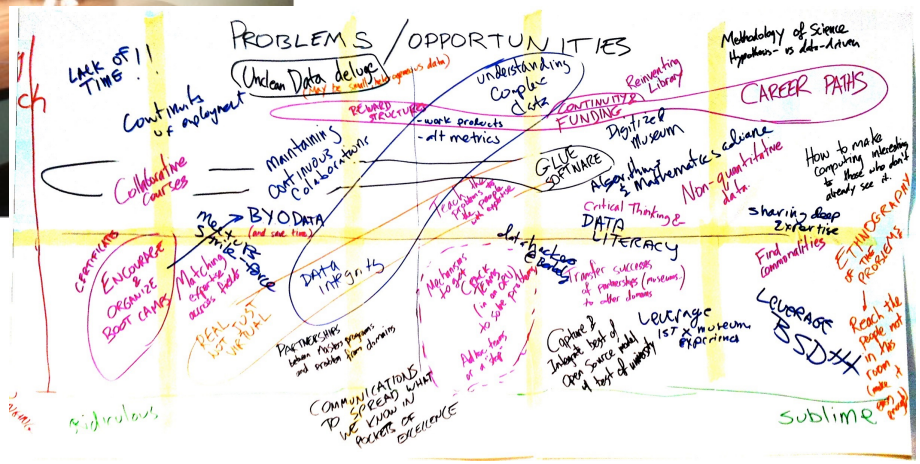
BIDS
BERKELEY INSTITUTE FOR DATA SCIENCE

Berkeley
UNIVERSITY OF CALIFORNIA

Great interest from across the campus

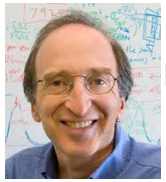
Data Science Workshop held in February 2013 was attended by 80 researchers on three days notice; with follow-up events in May and June (to date 280+ signed up for mailing list)





BERKELEY
Institute for Data Science

Initial Faculty Group



Faculty Lead/PI: **Saul Perlmutter**, Physics, Berkeley Center for Cosmological Physics

Joshua Bloom, Professor, Astronomy; Director, Center for Time Domain Informatics

Henry Brady, Dean, Goldman School of Public Policy

Cathryn Carson, Associate Dean, Social Sciences; Acting Director of Social Sciences Data Laboratory "D-Lab"

David Culler, Professor, EECS

Michael Franklin, Chair, EECS, Co-Director, AMP Lab

Erik Mitchell, Associate University Librarian

Fernando Perez, Researcher, Henry H. Wheeler Jr. Brain Imaging Center

Jasjeet Sekhon, Professor, Political Science and Statistics; Center for Causal Inference and Program Evaluation

Jamie Sethian, Professor, Mathematics

Kimmen Sjölander, Professor, Bioengineering, Plant and Microbial Biology

Philip Stark, Chair, Statistics

Ion Stoica, Professor, EECS, Co-Director, AMP Lab

A 5-year, $37.8 million cross-institutional collaboration

NYU

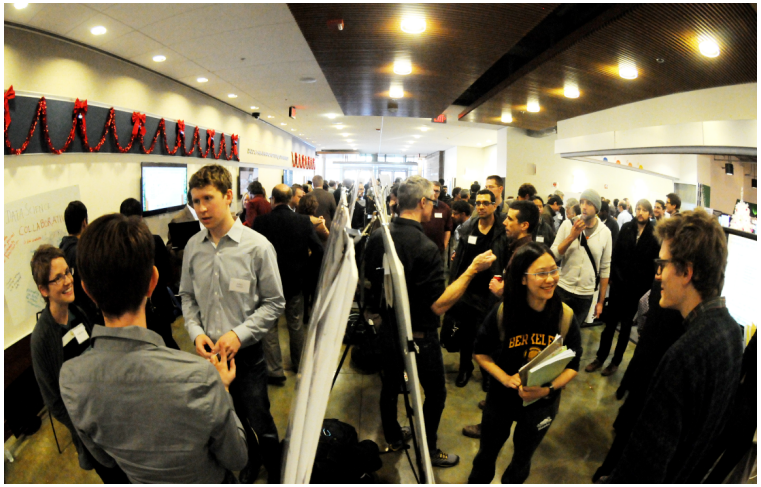
Berkeley
UNIVERSITY OF CALIFORNIA

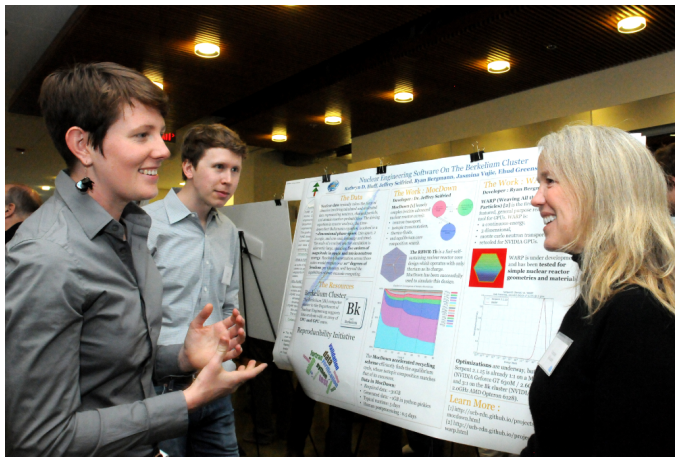
UNIVERSITY of WASHINGTON

GORDON AND BETTY
MOORE
FOUNDATION

ALFRED P. SLOAN FOUNDATION

Launched December 2013





BERKELEY
Institute for
Data Science

Our sponsors

- Foundations
  - Moore and Sloan Foundations $12.5 million
- Industry
  - Siemens
  - State Street
- Institutional
  - UC Berkeley

BIDS Goals

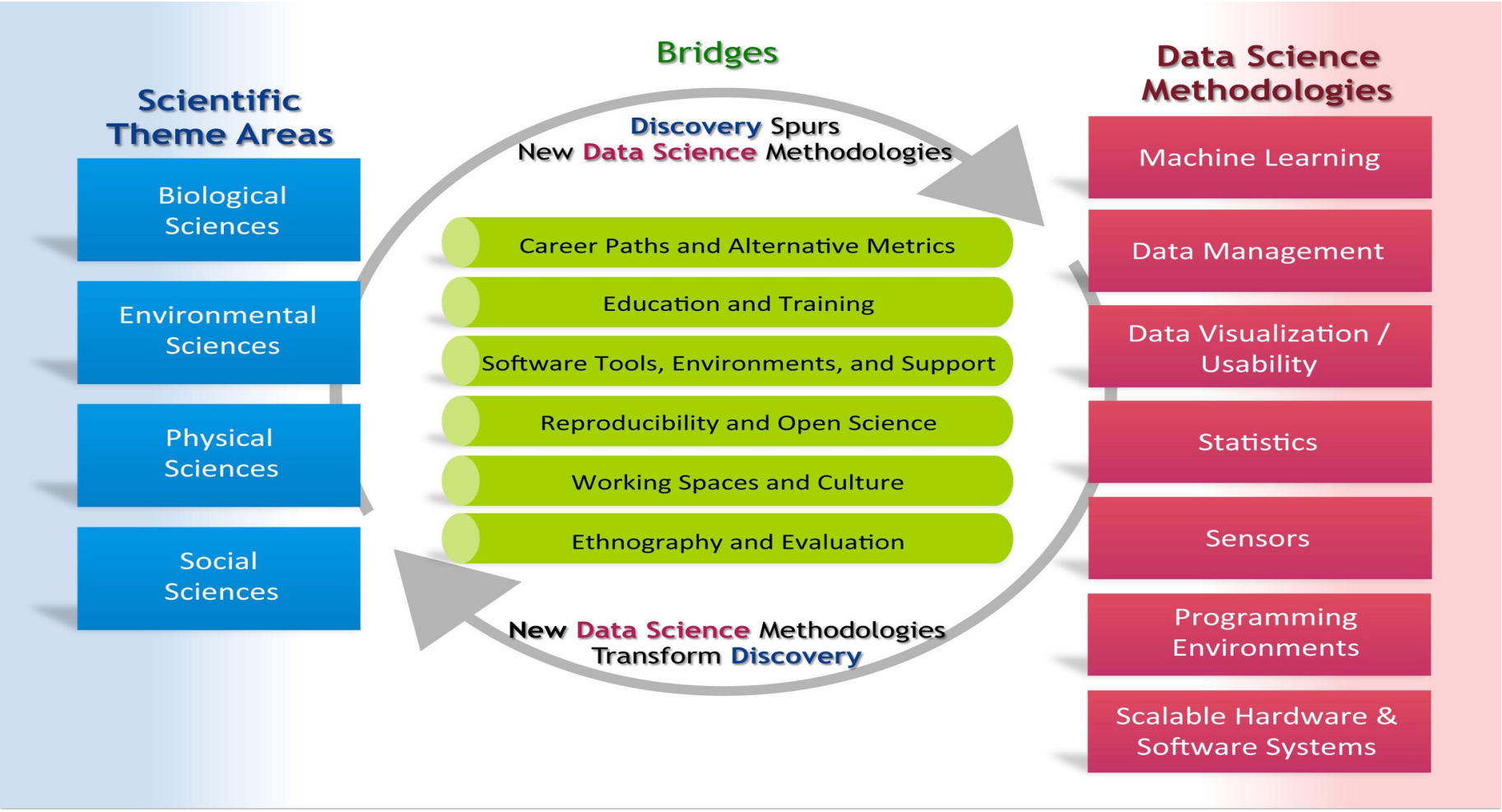
- Support meaningful and sustained interactions and collaborations between
  - Science domains: life science, social science, physical science
  - Methodology fields: computer science, statistics, applied mathematics
- Establish new Data Science career paths that are long-term and sustainable
  - A generation of multi-disciplinary scientists in data-intensive science
  - A generation of data scientists focused on tool development
- Build an ecosystem of analytical tools and research practices
  - Sustainable, reusable, extensible, easy to learn and to translate across research domains
  - Enables scientists to spend more time focusing on their science

We have computing power, we have applied math techniques, we have database approaches, so...

What's missing?

Working Groups

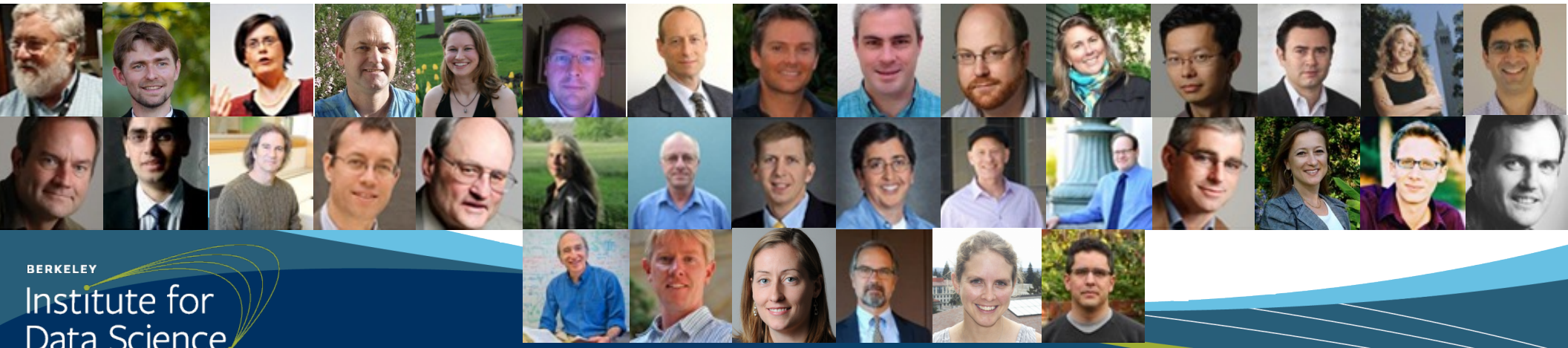
Working to address the major challenges facing major advances in data driven research.



People are at the heart of BIDS

Structured roughly as:

- Competitive fellow positions, typically 50/50 appts, 2-3 y.
- A few full-time positions (tools, ethnography, staff).
- Senior fellows: campus faculty and LBNL scientists.



BERKELEY
Institute for
Data Science

Diverse expertise

- Sociology
- Phylogenomics
- Cosmological Physics
- Nuclear Science
- Neuroscience
- Energy and Resources
- System software
- High-performance computing
- Global Change Biology
- Geospatial
- Statistics
- Environmental science
- Computer Vision
- Distributed computing
- Seismology
- Computer Science
- Astronomy
- Public Policy
- Social Sciences
- Psychology
- Library science
- Molecular & Cell Biology
- Political Science
- Mathematics
- Bioengineering
- City & Regional Planning
- …

Diverse Software Development

http://bids.berkeley.edu/research

BIDS Fellows engage in a range of projects that address the ongoing needs of effectively advancing data-intensive research.

IP[y]: IPython

scikit-image
image processing in python

HOLOS
BERKELEY ECOINFORMATICS ENGINE

The DECIDING FORCE Project

NIMBLE

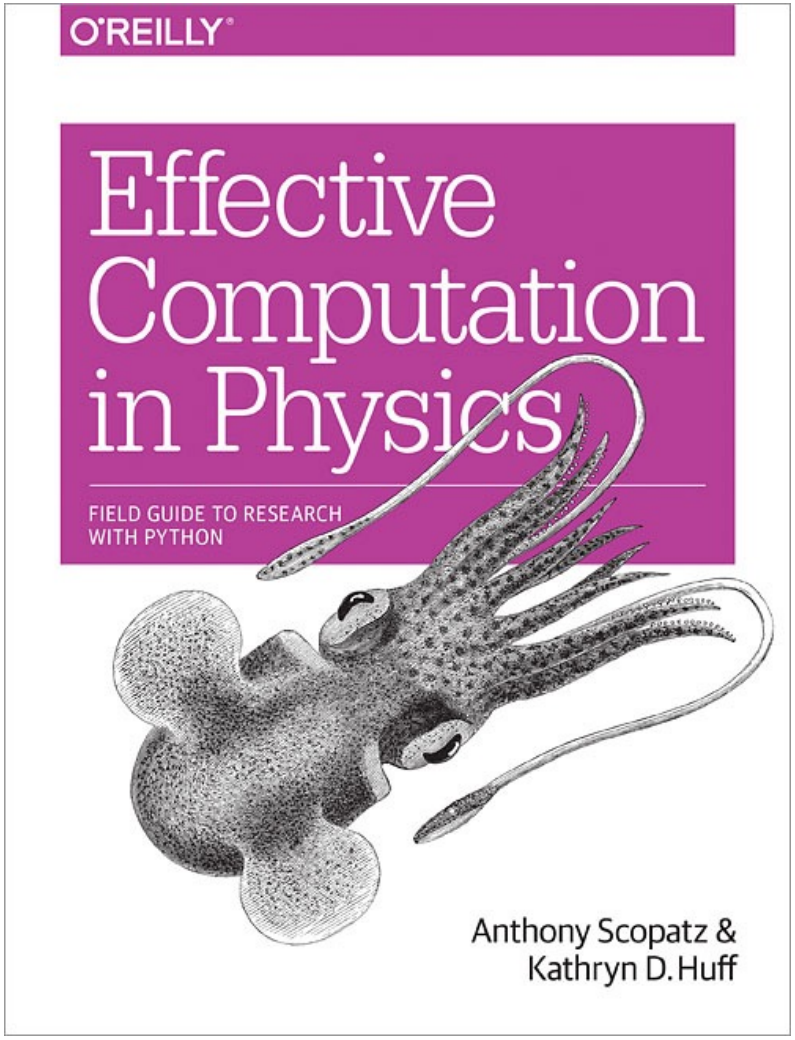
txt_thrshr

COLUS

NumPy

astropy
A Community Python Library for Astronomy

R OpenSci



Katy Huff

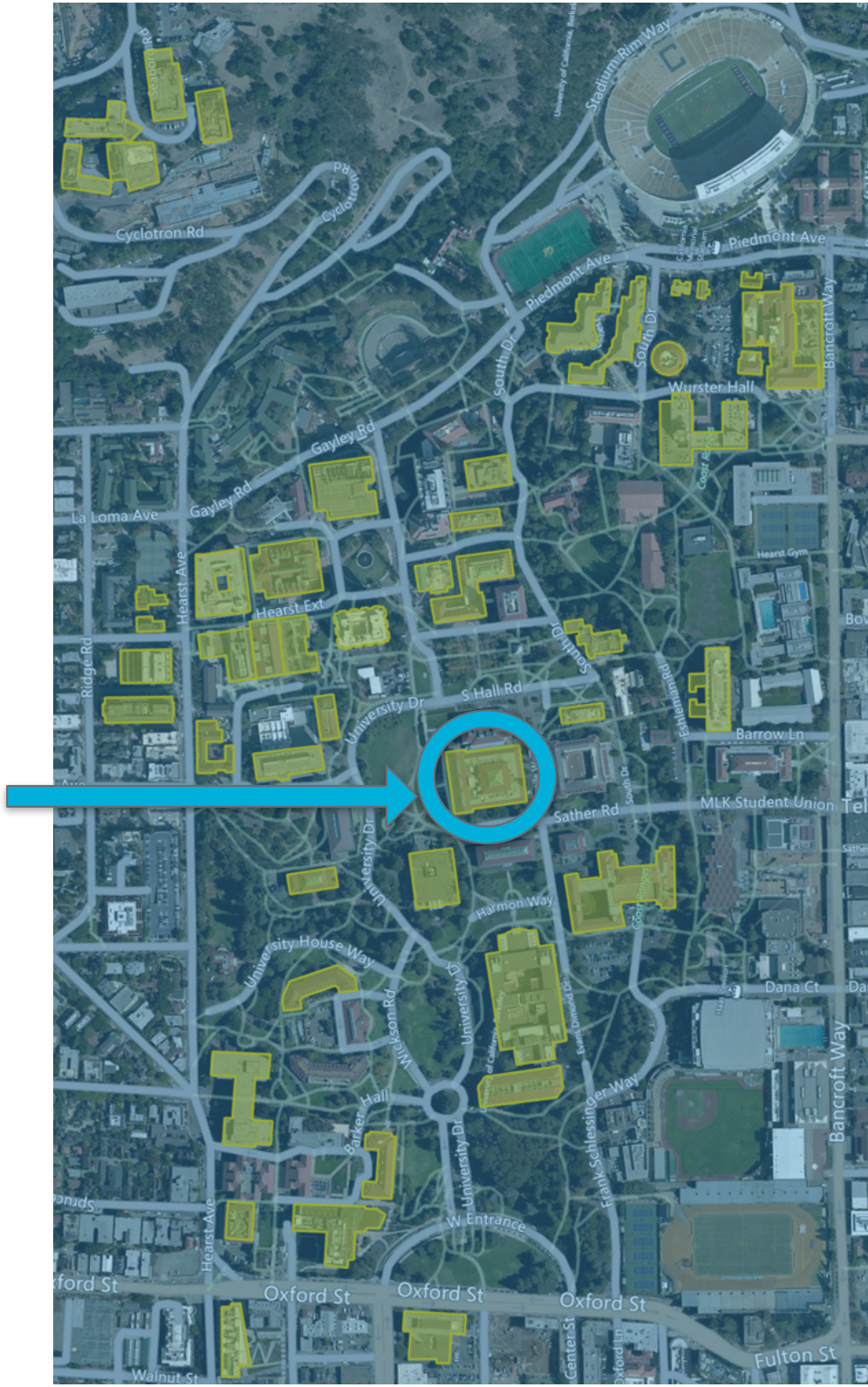
Nuclear Engineering Postdoc
BIDS Data Science Fellow

- physics.codes
- github.com/physics.codes/examples
- shop.oreilly.com/product/0636920033424.do

Our collaborative space

190 Doe Library

Central location that serves as home for data science efforts



Our collaborative space

190 Doe Library

BIDS Events

- Data Science Lecture Series
- BIDS Tea
- The Hacker Within
- Data Science Collaborative
- Data Science Faire
- Technical workshops (Spark, Azure, Data Structures for Data Science, ...)

Joining threads…

- **Project Jupyter**
  - Growing like crazy…
  - We think the next few years will be very interesting
  - But we know the toughest challenges aren't technical!
- **BIDS**
  - is the kind of home I've always wanted for a project like Jupyter
  - it's still an experiment, evolving and changing
  - will change a lot in the next year, as it moves from bootstrapping to forming its own research identity…
  - …and the broader discussion of Data Science @ Berkeley evolves…

Jupyter Project is hiring at Berkeley!

Two postdocs

Project Manager

Thank You!

@fperez_org fperez@lbl.gov

@ProjectJupyter @IPythonDev

Try it out at

try.jupyter.org