Welcome!

This slide and more info at: http://bit.ly/JPyBloom

Before we get started please fill out this poll:

http://etc.ch/gw2Z

If you're feeling ambitious, install Anaconda with Jupyter at:
https://www.continuum.io/downloads

Learning Jupyter

A hands-on night with PyLadies
at Bloomberg

Jamie Whitacre
whitacre@berkeley.edu
Berkeley Institute for Data Science

Tonight

A Jupyter overview + demo

Lab time to install Jupyter and try it out!

Jupyter is great for learning Python

basic knowledge of Python is assumed for this talk

Survey and install info at: http://bit.ly/JPyBloom

Live Resources

Live humans here tonight:

Jupyter Team Members

Friends of Jupyter

Extraordinary PyLadies

Survey and install info at: http://bit.ly/JPyBloom

Other Resources

- **Our website: Jupyter.org**

- **Online documentation:**

  - "Official" documentation written by the project: https://jupyter.readthedocs.io/en/latest/

  - Documentation written by the community: http://jupyter-notebook-beginner-guide.readthedocs.io/en/latest/index.html

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

  Now you can also search for notebooks on GitHub as well

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

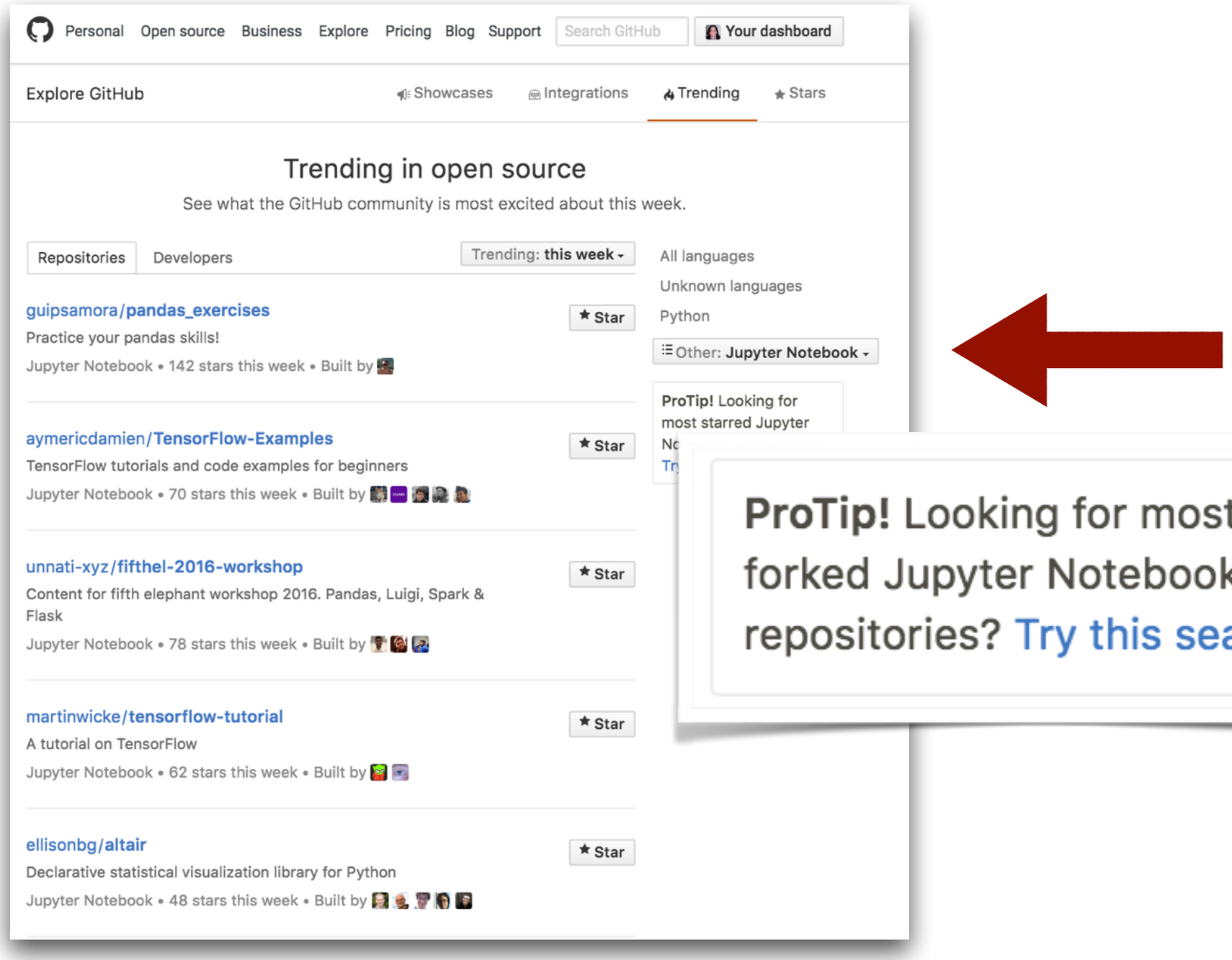
This section contains academic papers that have been published in the peer-reviewed literature or pre-print sites such as the ArXiv that include one or more notebooks that enable (even if only partially) readers to reproduce the results of the publication. If you include a publication here, please link 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

Over 500k Notebooks on GitHub

Check out “Trending Repositories”



https://github.com/trending/jupyter-notebook?since=weekly

About Me

- Jupyter Technical Project Manager based out of the Berkeley Institute for Data Science at UC Berkeley

- Background in Bioinformatics & Scientific Data Systems Engineering

- How data can help us tell stories about the earth and everything on it.


jupyter

What is it?

Nathan Cummins

jupyter Welcome to P

File Edit View Insert Cell

jupyter

Welcome to the

This Notebook Server wa

WARNING

Don't rely on this serv

Your server is hosted thar

Run some Python

To run the code below:

1. Click on the cell to se
2. Press SHIFT+ENTER

A full tutorial for using the

```
In [ ]: %matplotlib inline

        import pandas as pd
        import numpy as np
        import matplotlib
```

jupyter Lorenz Differential Equations (autosaved)

File Edit View Insert Cell Kernel Help Python 3

Code Cell Toolbar: None

Exploring the Lorenz System

In this Notebook we explore the Lorenz system of differential equations:

$$\dot{x} = \sigma(y - x)$$
$$\dot{y} = \rho x - y - xz$$
$$\dot{z} = -\beta z + xy$$

This is one of the classic systems in non-linear differential equations. It exhibits a range of complex behaviors as the parameters (σ, β, ρ) are varied, including what are known as *chaotic solutions*. The system was originally developed as a simplified mathematical model for atmospheric convection in 1963.

```
In [7]: interact(Lorenz, N=fixed(10), angle=(0.,360.),
                σ=(0.0,50.0),β=(0.,5), ρ=(0.0,50.0));
```

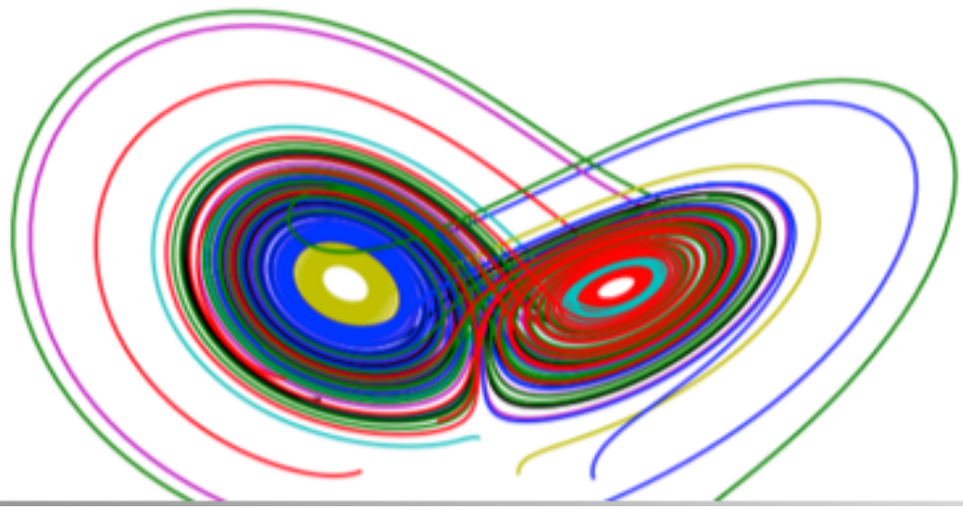
angle 308.2

max_time 12

σ 10

β 2.6

ρ 28



Jupyter in the wild

- Open source - which means anyone can download it and use it for free.

- Digital Journalism, Scientific research, Commercial products, Classrooms, Coding Bootcamps, Online Learning, Conferences, Data Engineering Pipelines, High Performance Computing (HPC) . . .

- Physics, Astronomy, Biology, Economics & Finance, Social Sciences, Geo Sciences, Digital Humanities . . .

Jupyter Notebook

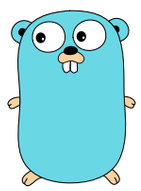
- A digital “document” that gives you a way to record your thoughts and CODE in a single place

- Provides a window into your data, serves as a tool to explore, understand, communicate, and tell stories about your data in a systematic and reproducible way

- Is the evolution of the IPython Notebook into a language agnostic computing environment supporting 60+ programming languages including Python, R, Julia, and many others.

- Allows you to use the many Python libraries available to developers

Jupyter Protocol
is language agnostic



Jupyter Notebook

- A digital “document” that gives you a way to record your thoughts and CODE in a single place

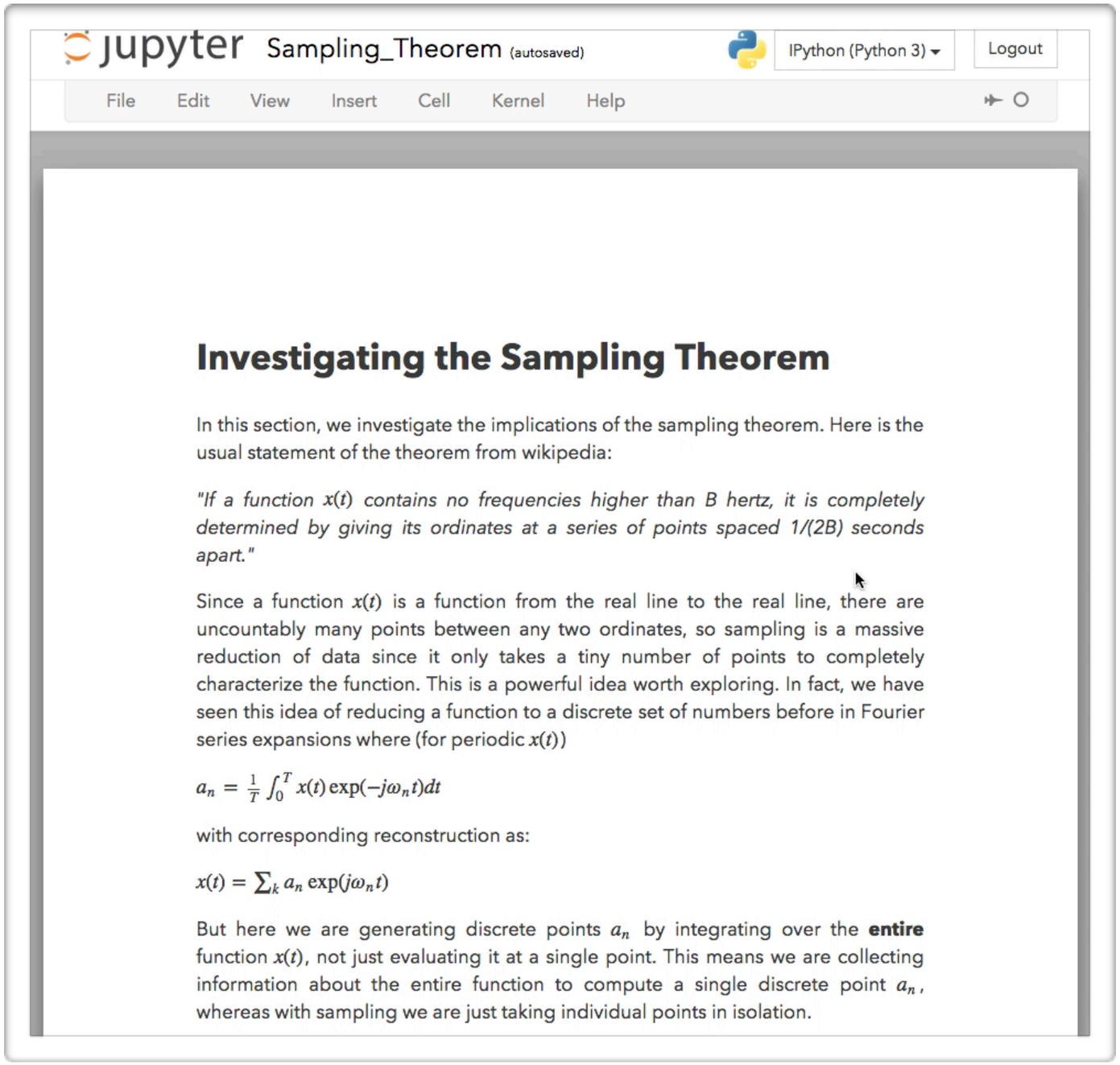
- Provides a window into your data, serves as a tool to explore, understand, communicate, and tell stories about your data in a systematic and reproducible way

- Is the evolution of the IPython Notebook into a language agnostic computing environment supporting 60+ programming languages including Python, R, Julia, and many others.

- Allows you to use the many Python libraries available to the Python community

Jupyter Notebook

- Notebook = Sequence of text cells & code cells
- Publicly documented
- text cell = markdown + latex
- Machine readable, easy to understand (input + output)
- Transformable (metadata)



https://github.com/unpingco/Python-for-Signal-Processing

A Tool for Interactive Computing

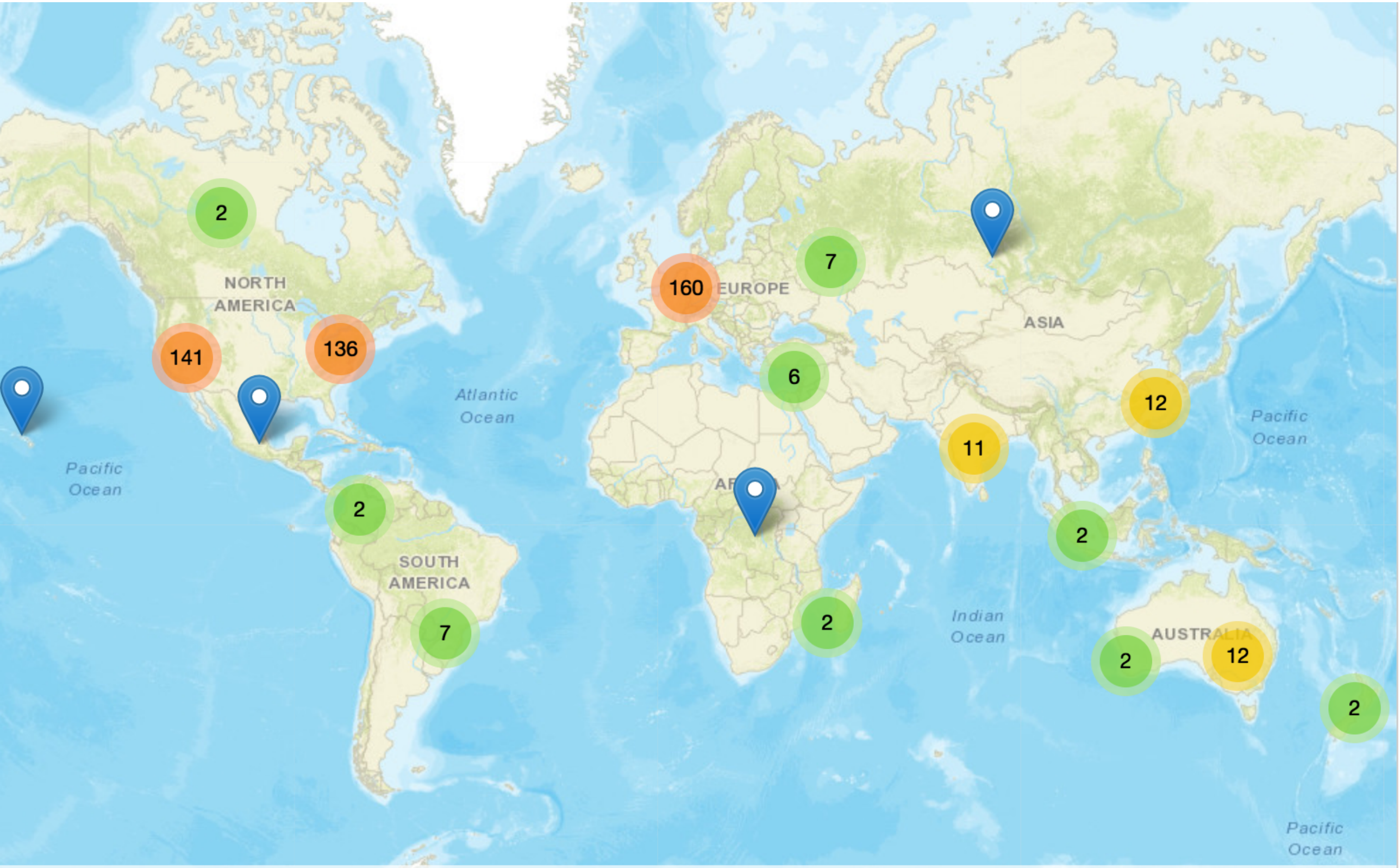
- A dialogue between the human and the computer.
- Assemble ideas using the computer as playground, as "data microscope"
- Building blocks for scientific computing and data science

Project Jupyter

Jupyter team: where *all the credit* goes

Plus ~ 500 more Open source contributors!

Committers to IPython/Jupyter repositories



By Stuart Geiger, Jamie Whitacre, Matthias Bussonnier, and Nami Saghaei

Funding and partnerships

ALFRED P. SLOAN FOUNDATION

THE LEONA M. AND HARRY B. HELMSLEY CHARITABLE TRUST

GORDON AND BETTY MOORE FOUNDATION

U.S. DEPARTMENT OF ENERGY

NEW! We're hiring!

O'REILLY

SIMONS FOUNDATION

POWERED BY rackspace® the open cloud company

CONTINUUM ANALYTICS

IBM

Microsoft

ENTHOUGHT SCIENTIFIC COMPUTING SOLUTIONS

Google

Bloomberg

Notebook Workflows: The Big Picture

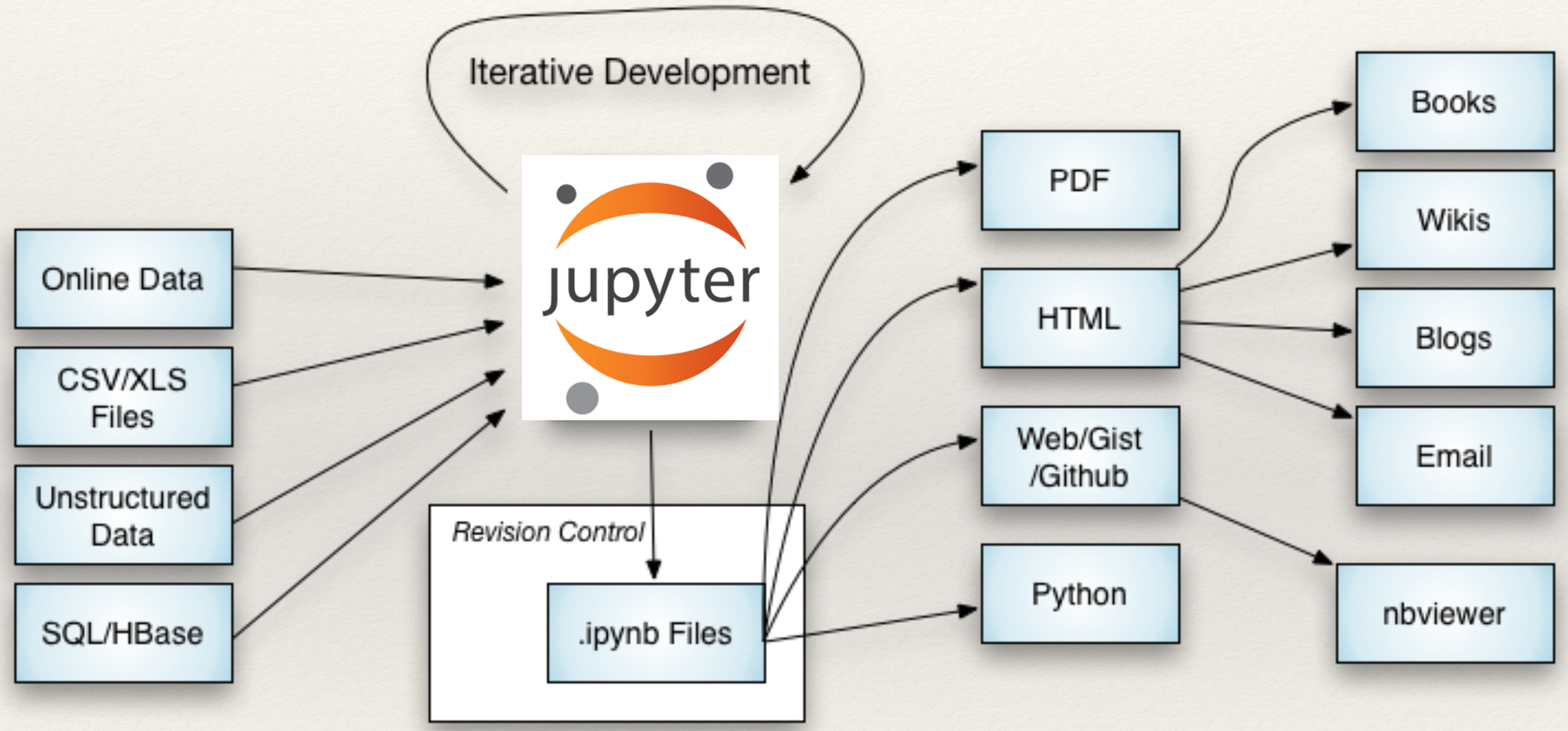


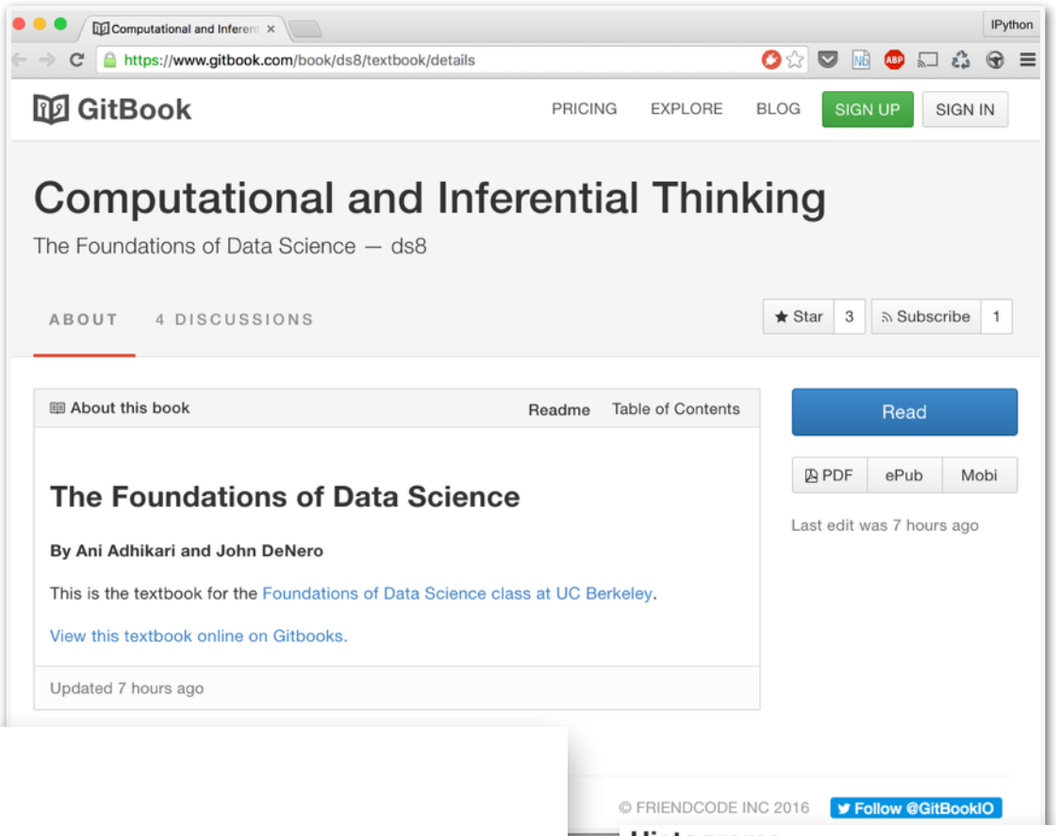
Image credit: Joshua Barratt

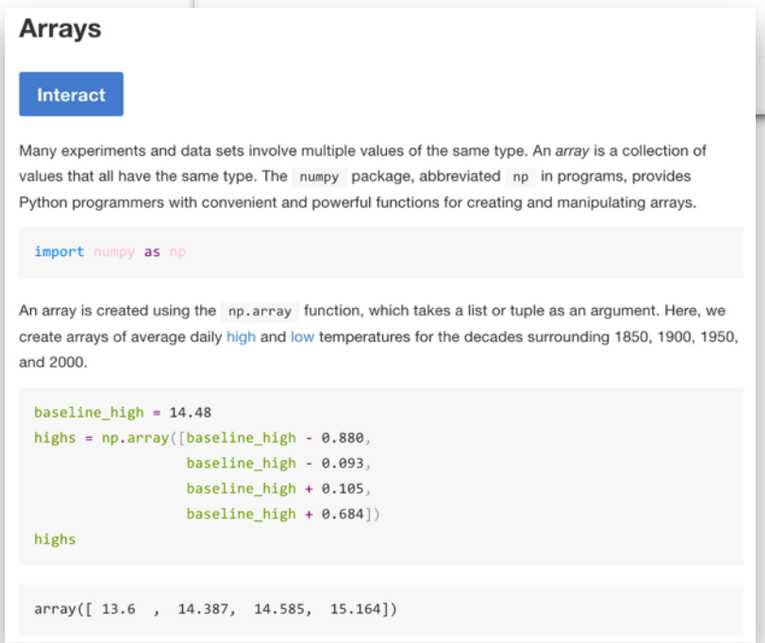
Data Analysis Workflow

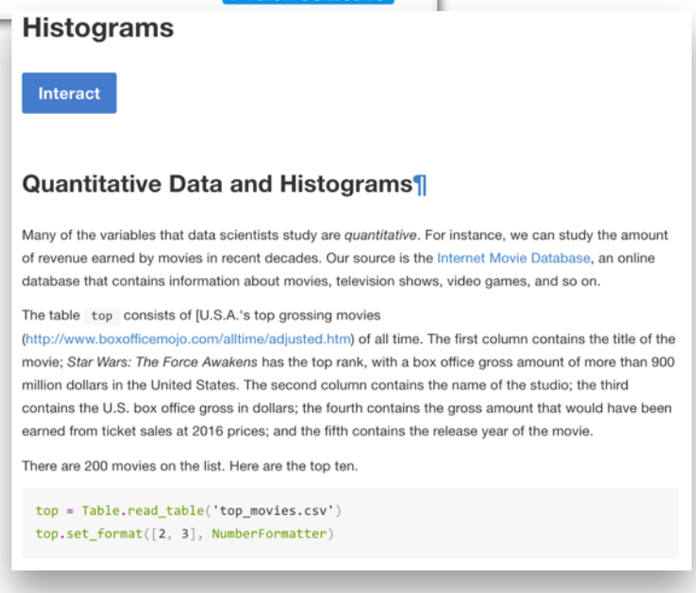
1. Pose a question or problem
2. Acquire data
3. Explore the data by writing & running code
4. Prepare and clean the data
5. Complete final analysis & visualize results
6. Write up analysis for publication (blogs, journals, etc.)
7. Share what you've done
8. Reproduce what other people have done

Berkeley's *Foundations of Data Science*

- New curriculum aimed at all freshmen at UC Berkeley
- Interactive textbook is Jupyter Notebooks
- Course deployment is JupyterHub
  - Off Jess Hamrick's work







data.berkeley.edu, data8.org

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.

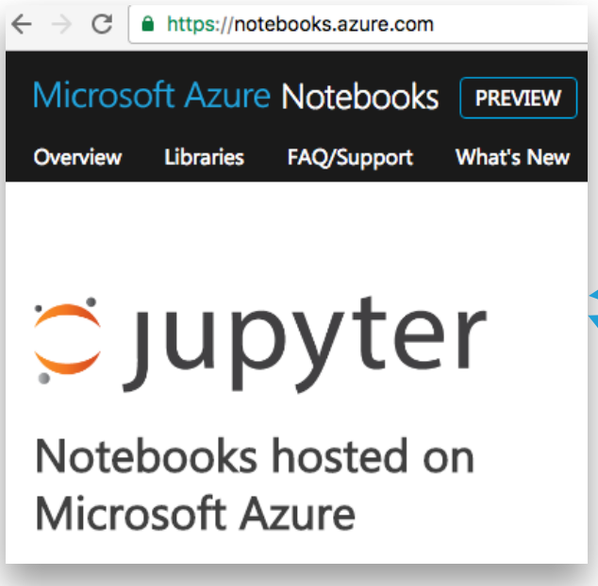
Container friendly

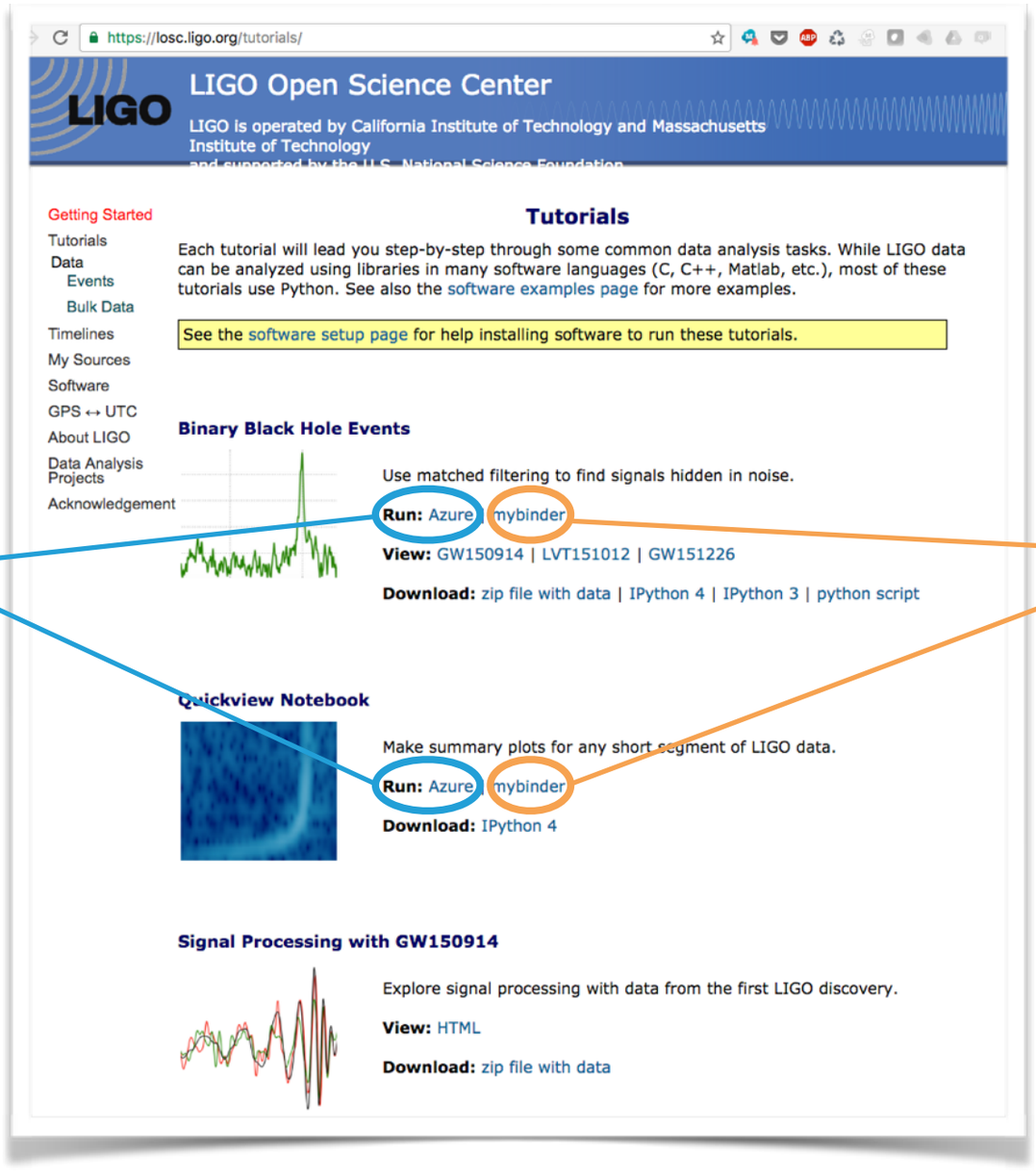
Use Docker containers to scale your deployment and isolate user processes using a growing ecosystem of prebuilt Docker containers.

Code meets data

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

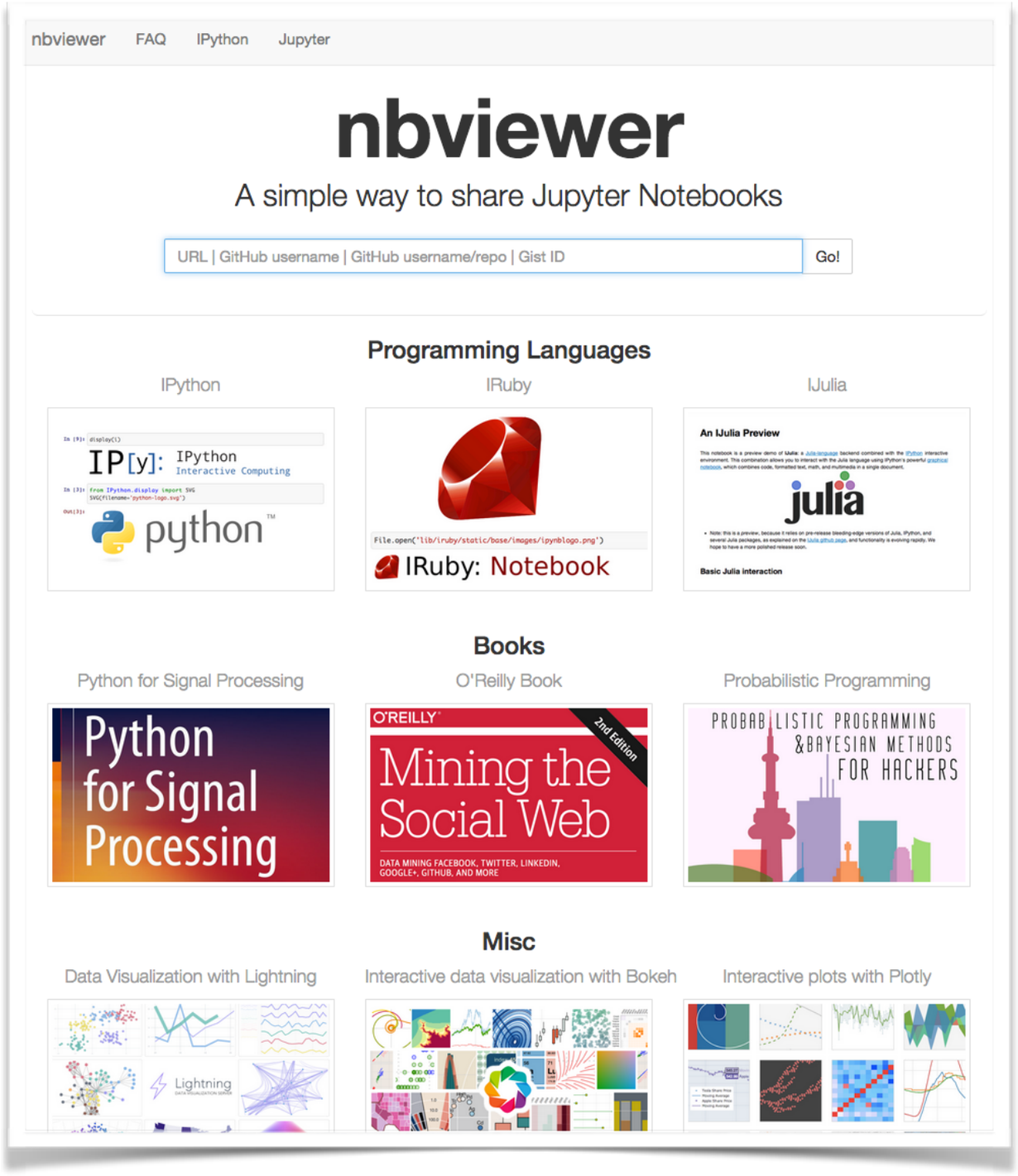
LIGO: Open Science with Jupyter



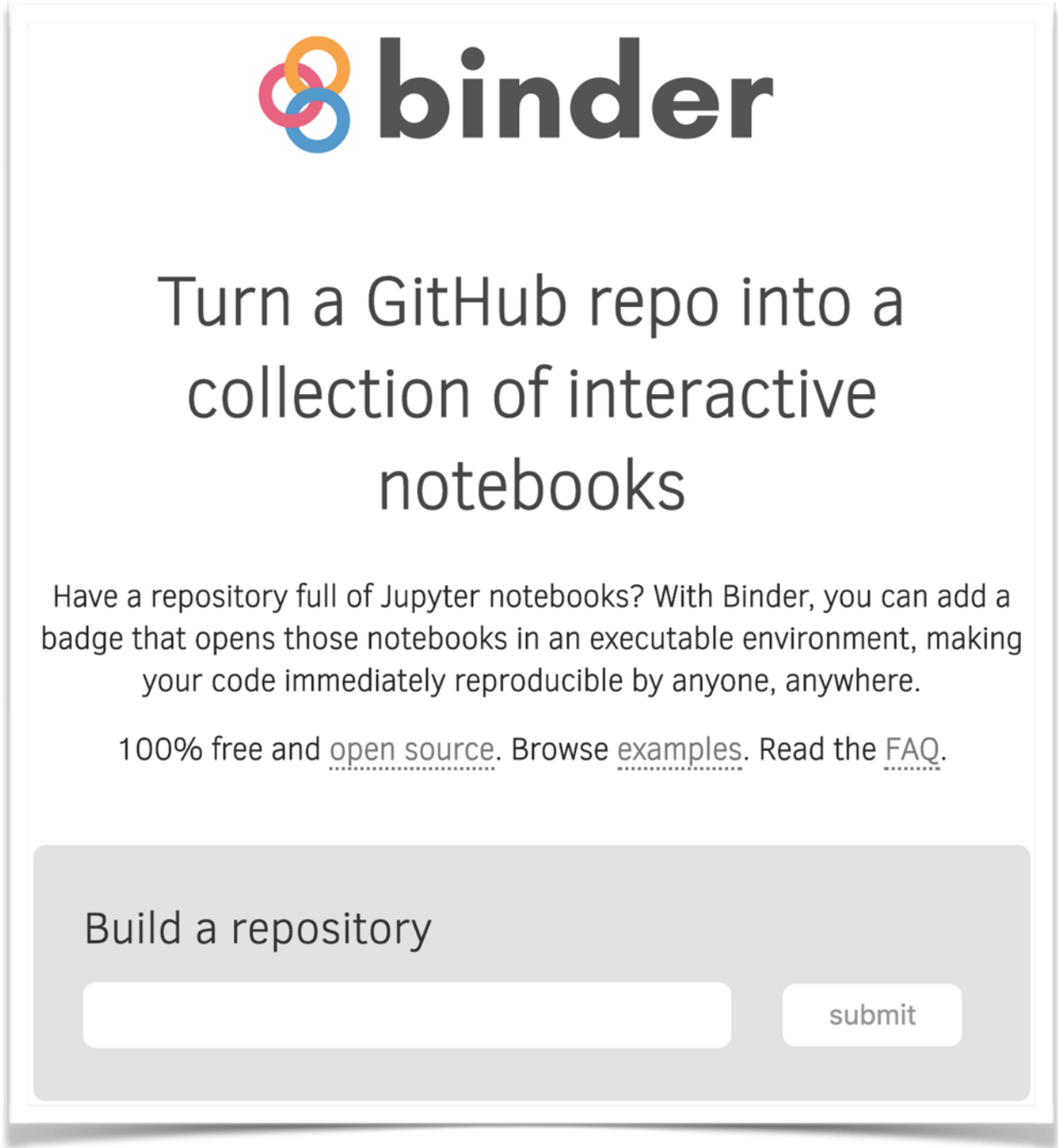


nbviewer: seamless notebook sharing

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



mybinder.org



github.com/freeman-lab

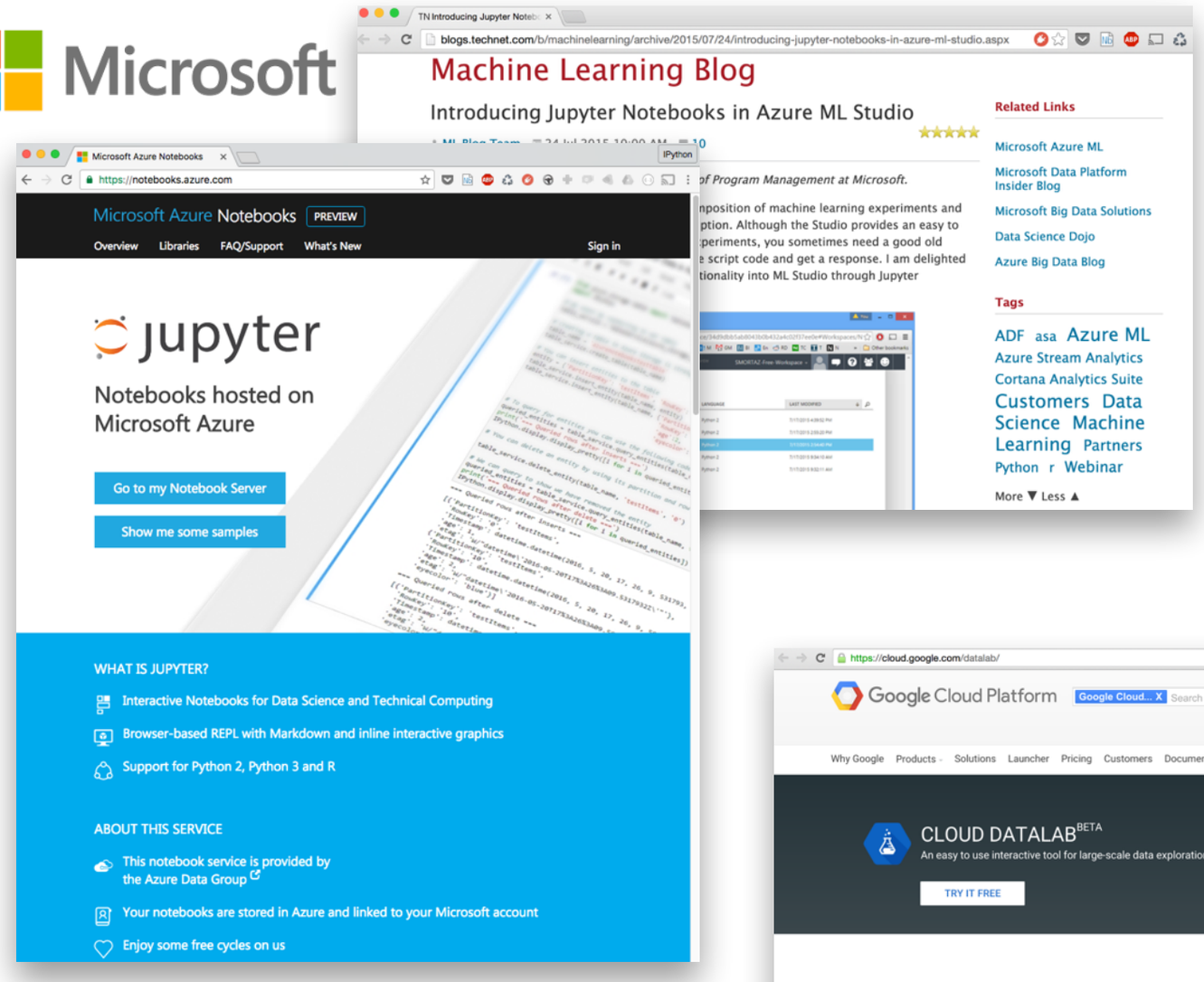
github.com/andrewosh

Andrew Osheroff's SciPy'16 talk:
https://www.youtube.com/watch?v=OK6M4w7LYIc

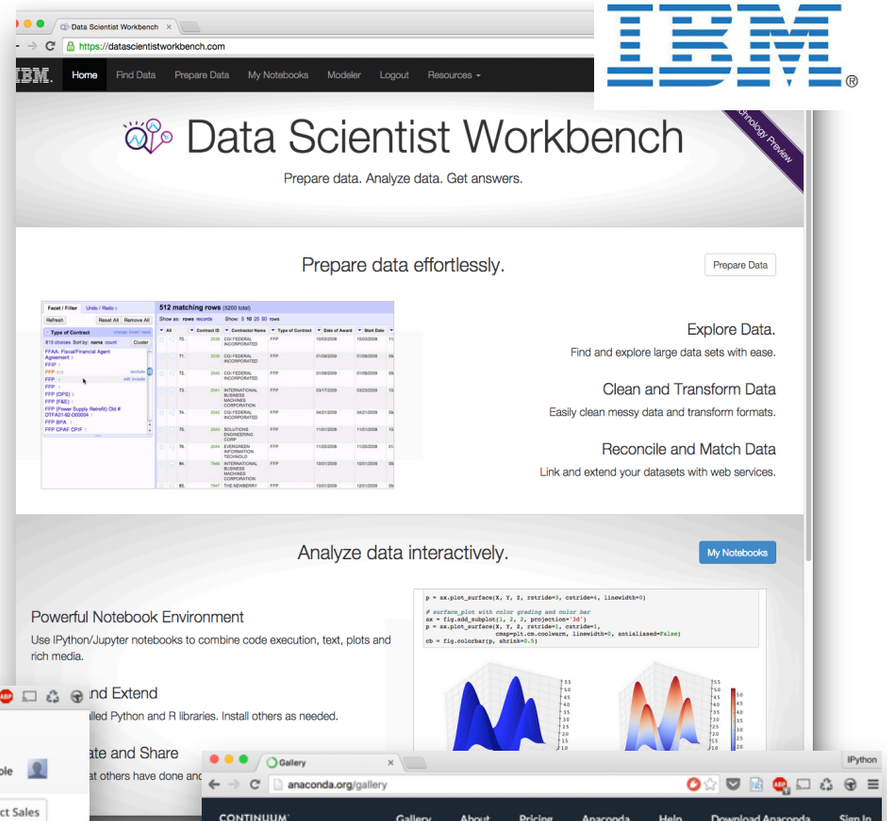
These foundations have become infrastructure

Microsoft, IBM, Google, Continuum…

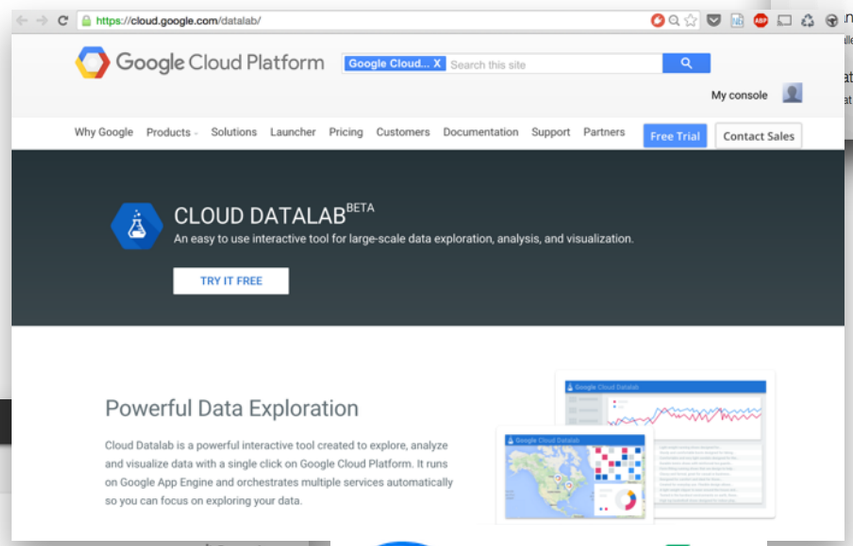
Microsoft

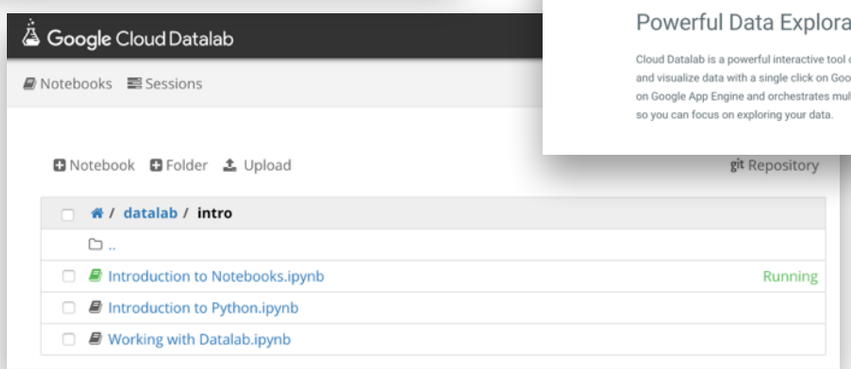


IBM

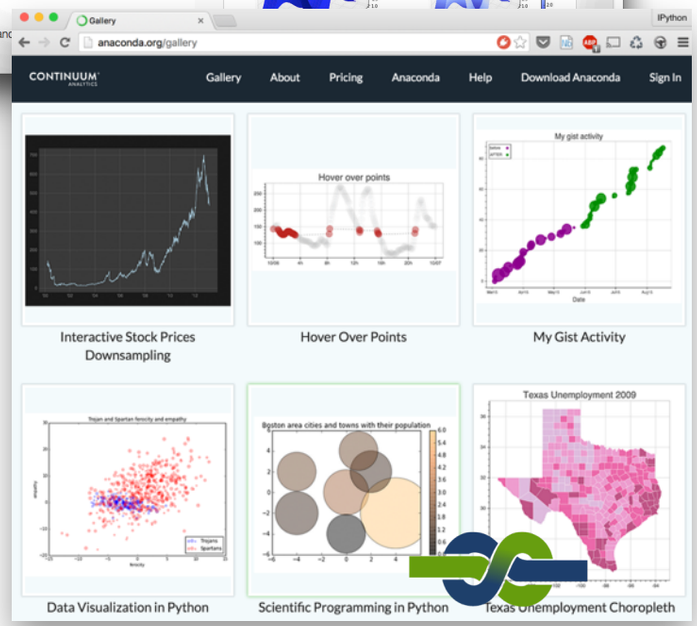


Google





Google

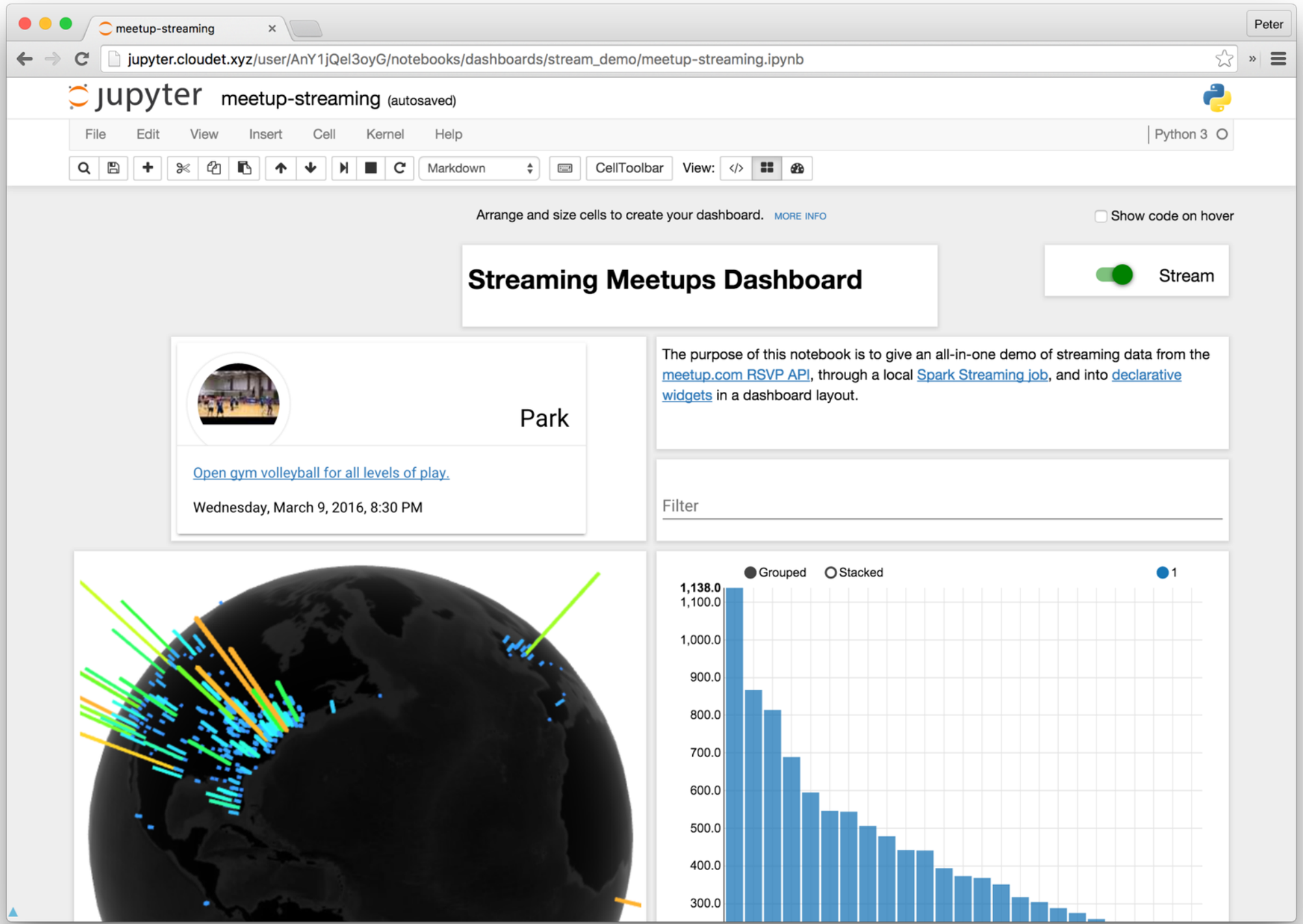


CONTINUUM ANALYTICS

In Development

Dashboards

https://github.com/jupyter-incubator/dashboards



Jupyter Dashboards & Declarative Widgets

Will help users to:

- Pull interactive web component widgets into a notebook
- Arrange widgets in a grid- or report-like layout
- Bundle notebooks and widgets for deployment as dashboards
- Serve notebook-defined dashboards as standalone web apps

nbdime

note**b**ook **di**ff & **me**rge

- tools for diffing and merging notebooks
- command-line rendering of diffs (outputs elided)

  nbdiff
- html rendering of rich diffs (JupyterLab)

  nbdiff-web
- git integration

  git-nbdiffdriver

https://github.com/jupyter/nbdime

nbflow

Jess Hamrick

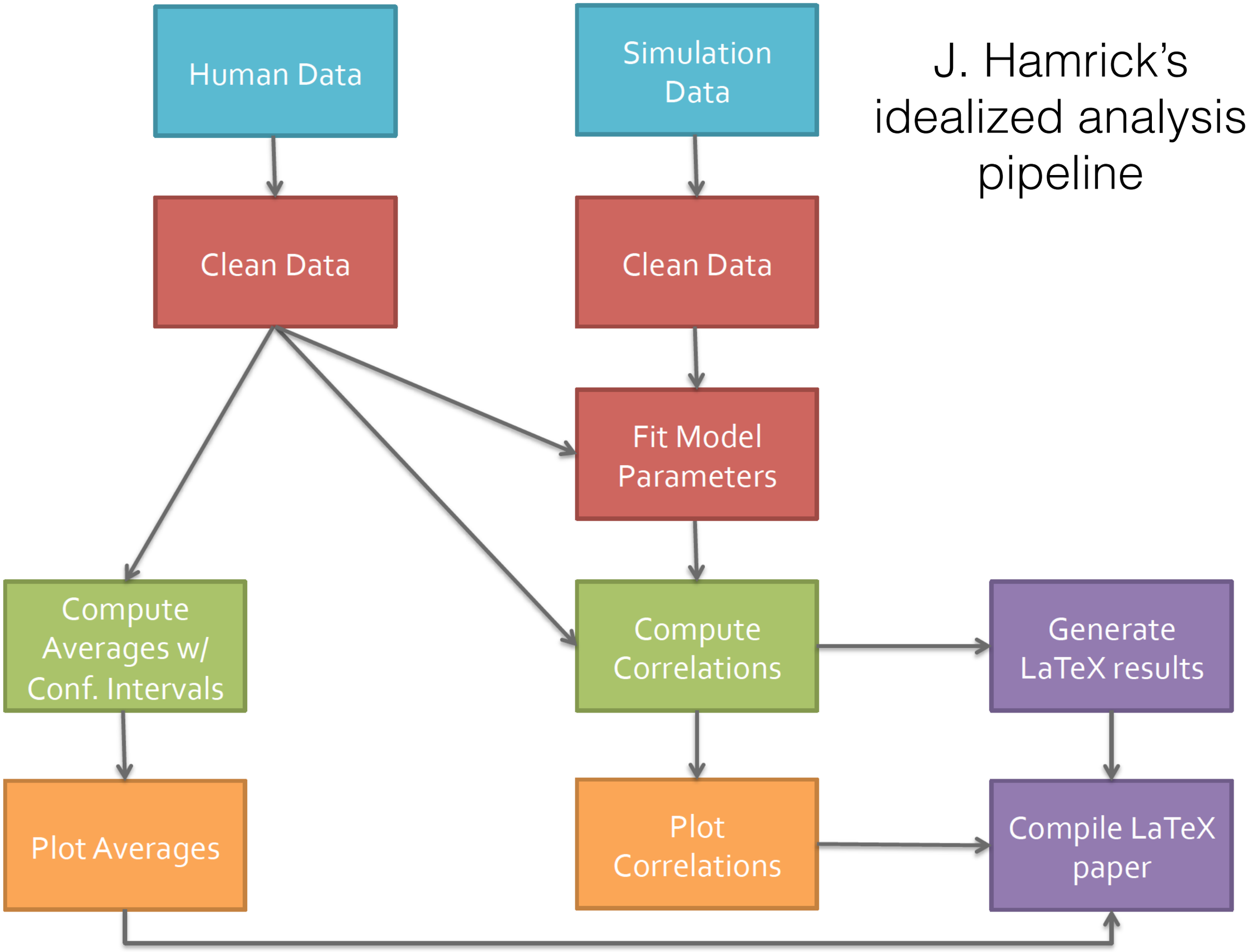
- Reproducible, One-Button Workflows with the Jupyter Notebook and SCons

- Ideal analysis pipelines

- Notebooks can be SCons commands (Python 2 only for now)

- Jess' nbflow repo: http://tinyurl.com/nbflow-example

- For Jess' SciPy 2016 talk: youtube.com/watch?v=Fc2W930NJs8



J. Hamrick's idealized analysis pipeline

JupyterLab:
Building Blocks for Interactive Computing

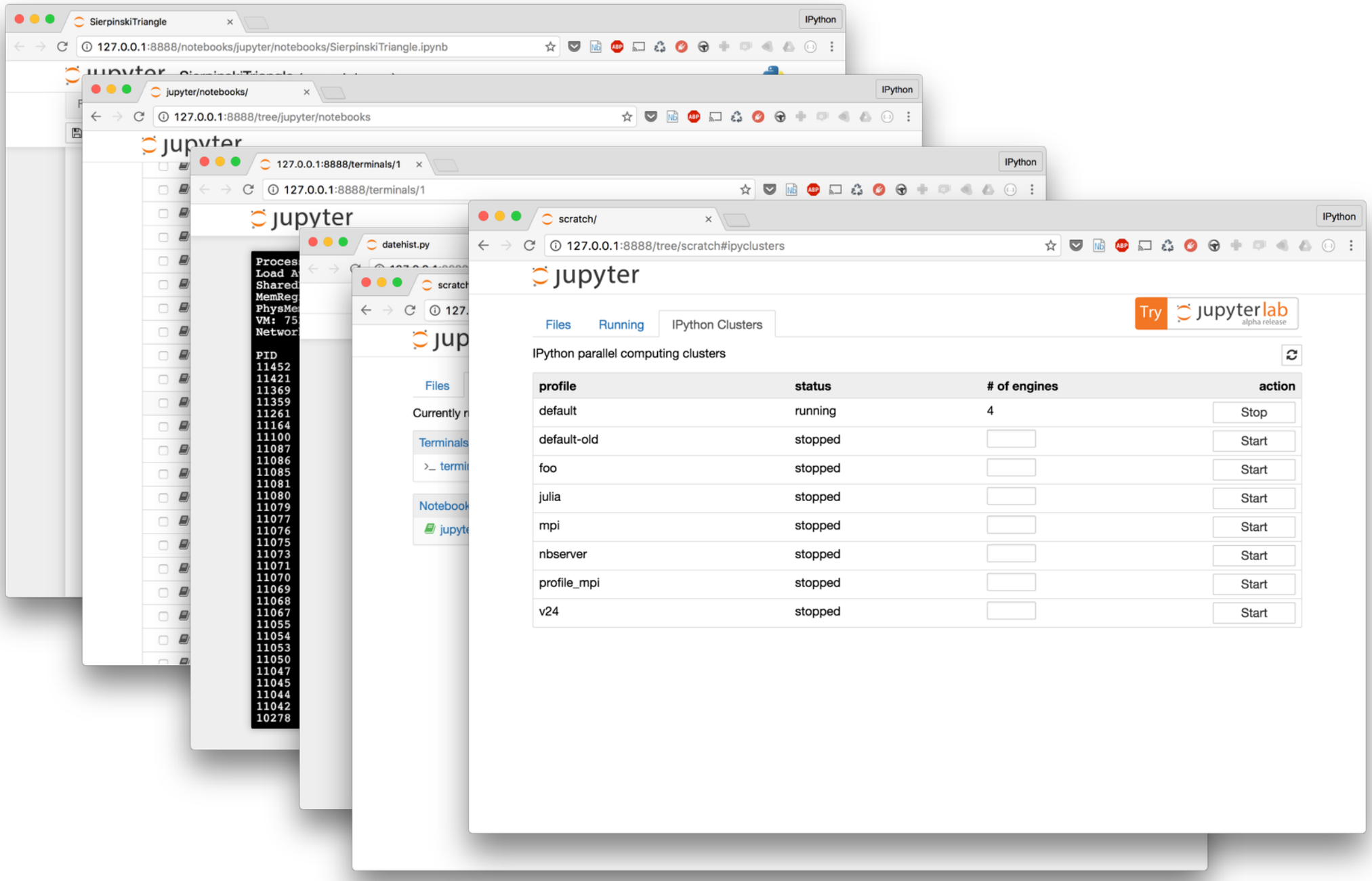
Brian E. Granger, Cal Poly
Jason Grout, Bloomberg LP
Chris Colbert, Continuum
Sylvain Corlay, Bloomberg
Afshin Darian, Continuum
Cameron Oelsen, Cal Poly
Fernando Perez, LBNL/ Berkeley
Steven Silvester, Continuum
David Willmer

JupyterLab

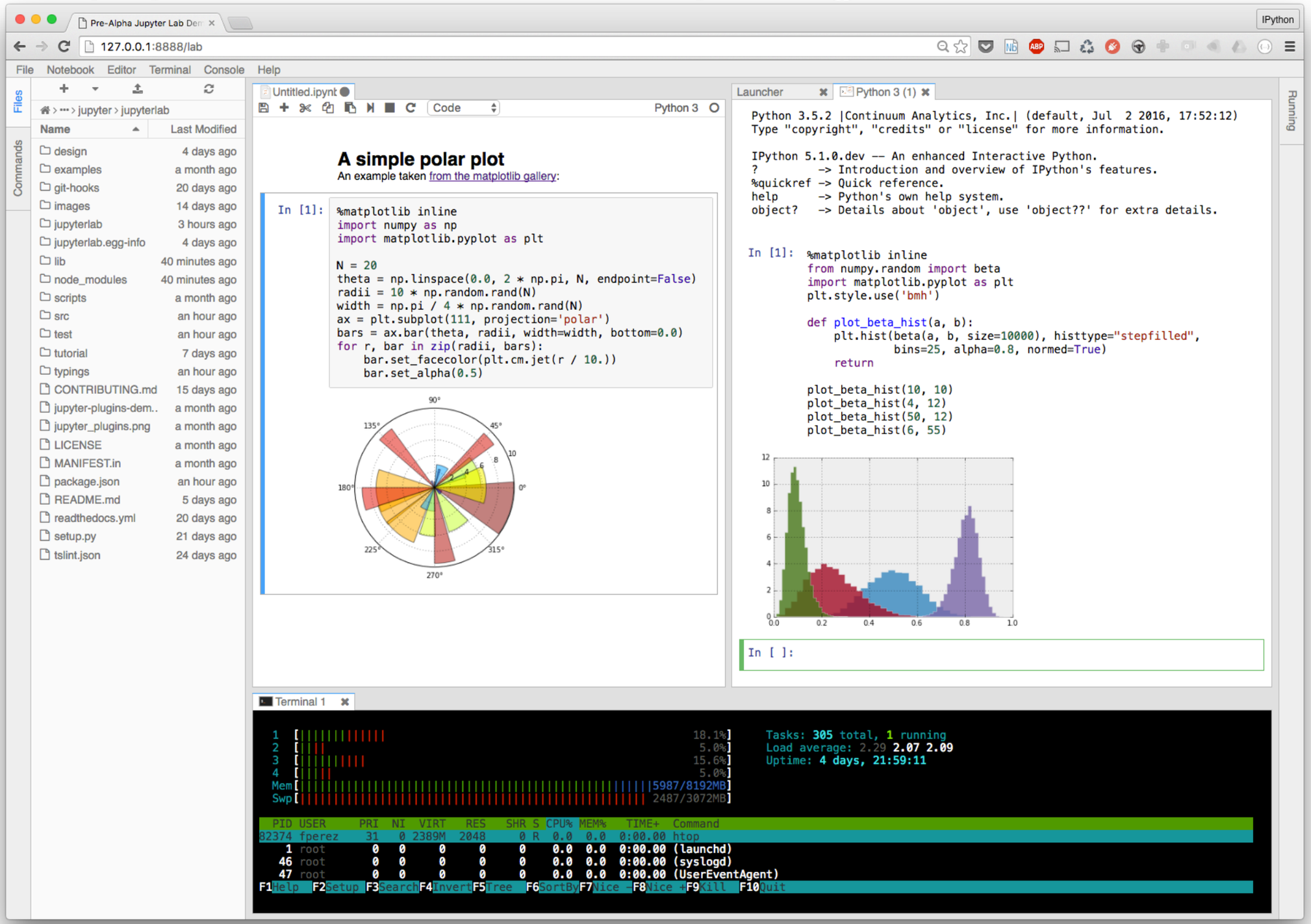
https://github.com/jupyter/jupyterlab

- JupyterLab is the natural evolution of the Jupyter Notebook user interface
- JupyterLab is an IDE: *Interactive* Development Environment
- Flexible user interface for assembling the fundamental building blocks of interactive computing
- Modernized JavaScript architecture based on npm/webpack, plugin system, model/view separation
- Built using PhosphorJS (http://phosphorjs.github.io/)
- Design-driven development process

The “Notebook”?



JupyterLab: unifying these ideas



https://raw.githubusercontent.com/jupyter/jupyterlab/master/jupyter-plugins-demo.gif

jupyter/jupyterlab*

*JupyterLab is a early developer preview, and is not suitable for general usage yet. Features and implementation are subject to change.

Learn More About Jupyter

- jupyter.org
- jupyter.readthedocs.org
- GitHub repos
- O'Reilly Tutorials
    - Advanced Jupyter Notebook Deployment by Jonathan Frederic
    - Jupyter Notebook for Data Science Teams by Jonathan Whitmore
- Mailing lists
- Conferences: PyData, SciPy, JupyterDays, PyCon, Strata
- YouTube
- Google it

Create tools for Data Science!
Contribute to Jupyter!

jupyter.org

github.com/jupyter

Thank You

whitacre@berkeley.edu
@jupyterJW
@ProjectJupyter @IPythonDev

Try it out at

try.jupyter.org

Jupyter Walkthrough

- Layout of screens
  - Notebook
  - File directory
  - other
- Basic Features
  - Create notebook
  - Add Cell
  - Execute Cell
  - Delete Cell
  - Keyboard shortcuts
  - Cell Magic
- Working with Data
  - Plots
  - Tables
  - Widgets
- Advanced
  - Multi-language
    - R

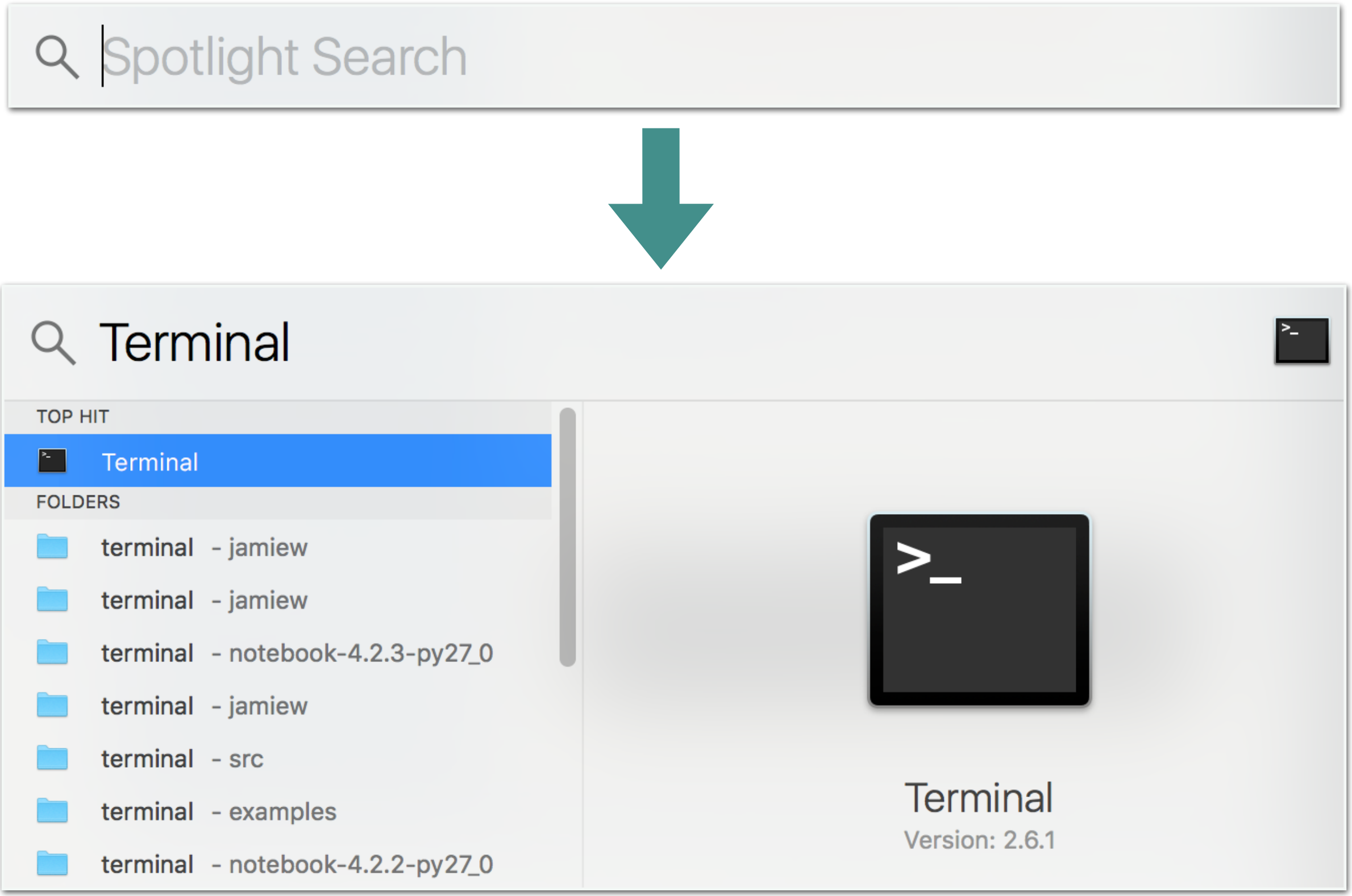
Install Jupyter

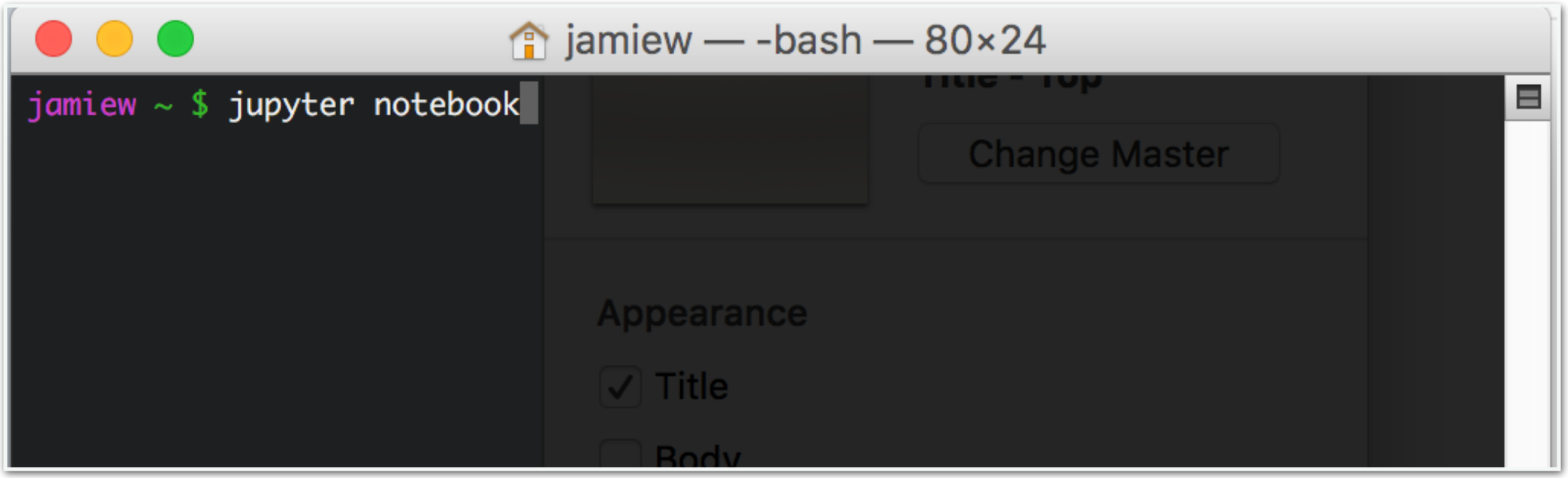
- Please take this poll before we get started:

http://etc.ch/gw2Z

- Easiest way to install Jupyter is by going to continuum.io/downloads and downloading Anaconda

- Anaconda is a free, open source scientific computing package that will put all the good stuff on your computer without much trouble


Spotlight Search
Terminal
TOP HIT
Terminal
FOLDERS
terminal - jamiew
terminal - jamiew
terminal - notebook-4.2.3-py27_0
terminal - jamiew
terminal - src
terminal - examples
terminal - notebook-4.2.2-py27_0
Terminal
Version: 2.6.1


jamiew — -bash — 80×24
jamiew ~ $ jupyter notebook
Change Master
Appearance
Title
Body

localhost:8888/tree

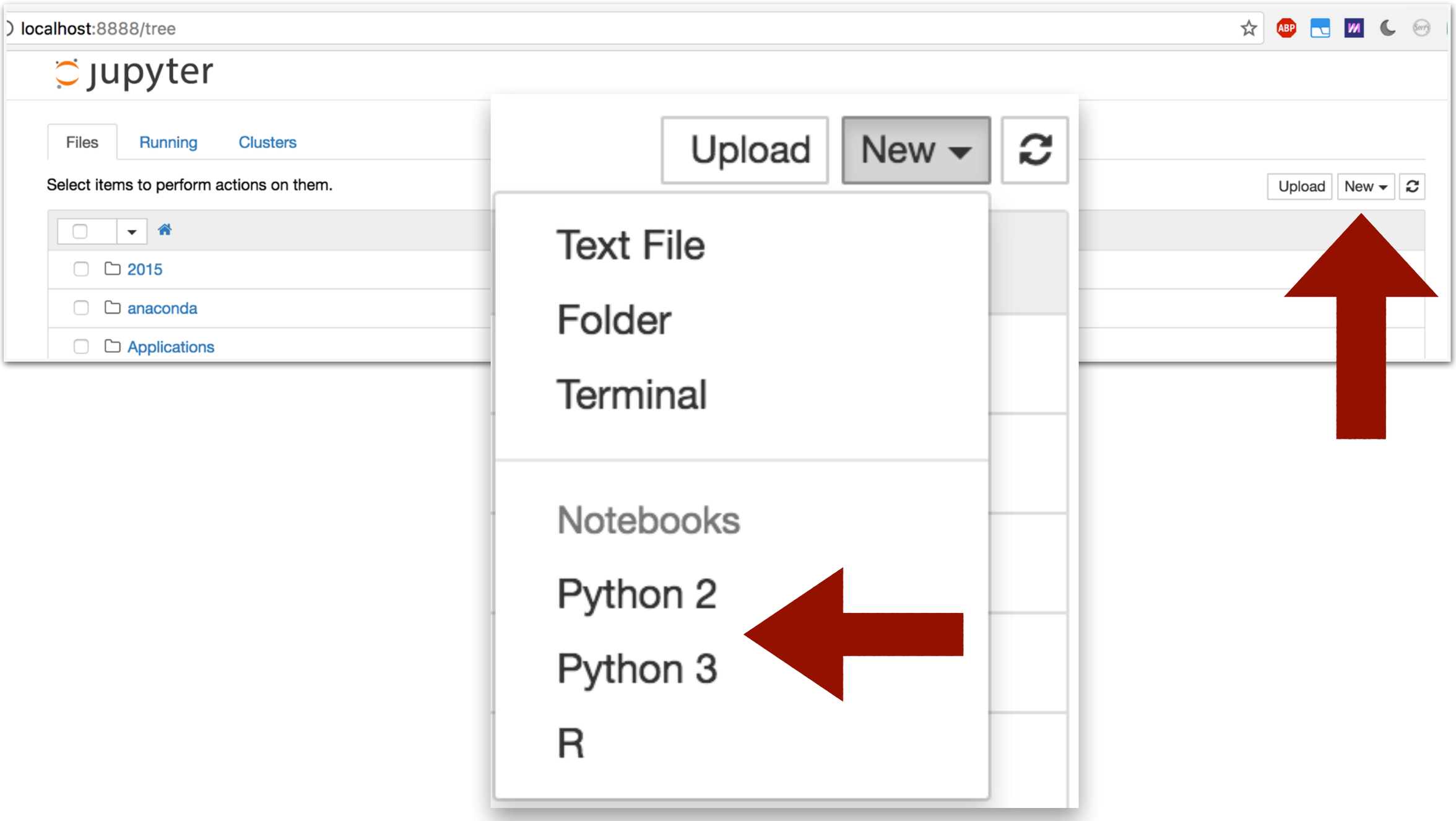
jupyter

Files | Running | Clusters

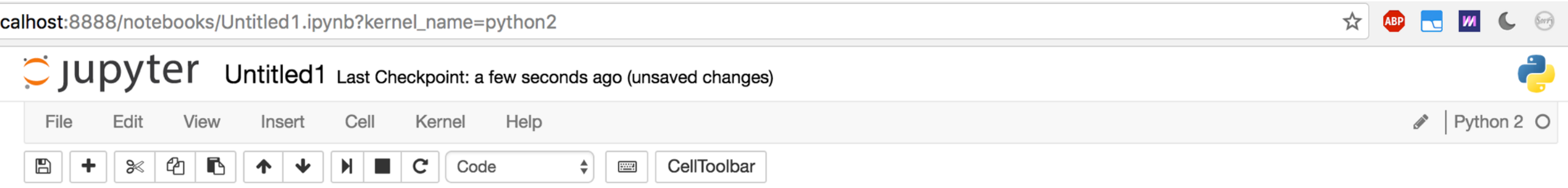
Select items to perform actions on them.

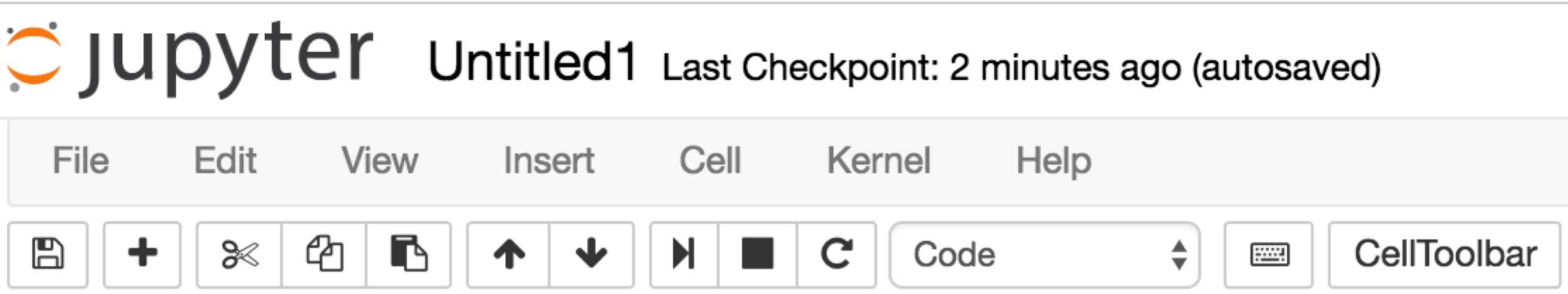
Upload | New | ↻

- 2015
- anaconda
- Applications
- Biorepository Legacy Migration
- bokeh-notebooks-master
- book-exercises
- Coursera-HowToUseGitandGitHub
- datasciencecoursera
- Desktop
- development
- Documents


localhost:8888/tree
jupyter
Files
Running
Clusters
Select items to perform actions on them.
2015
anaconda
Applications
Upload
New
Text File
Folder
Terminal
Notebooks
Python 2
Python 3
R

http://wikiclipart.com/candy-corn-clipart_1783/


calhost:8888/notebooks/Untitled1.ipynb?kernel_name=python2
jupyter
Untitled1
Last Checkpoint: a few seconds ago (unsaved changes)
File
Edit
View
Insert
Cell
Kernel
Help
Python 2
Code
CellToolbar
In [ ]:


jupyter
Untitled1
Last Checkpoint: 2 minutes ago (autosaved)
File
Edit
View
Insert
Cell
Kernel
Help
Code
CellToolbar

In [ ]:

General Commands

- Define a variable
  x = 5
  print(x)

Making Charts

http://matplotlib.org/users/screenshots.html#simple-plot

- Simple plot
- Histograms
- Bar Charts
- Scatter plot for fun

Other great plotting libraries: plot.ly, Seaborn, Altair, Bokeh

Data Frames

http://pandas.pydata.org/pandas-docs/stable/10min.html

github.com/jehuston/pandas_tutorial/blob/master/pandas_tutorial.ipynb

https://github.com/brandon-rhodes/pycon-pandas-tutorial

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

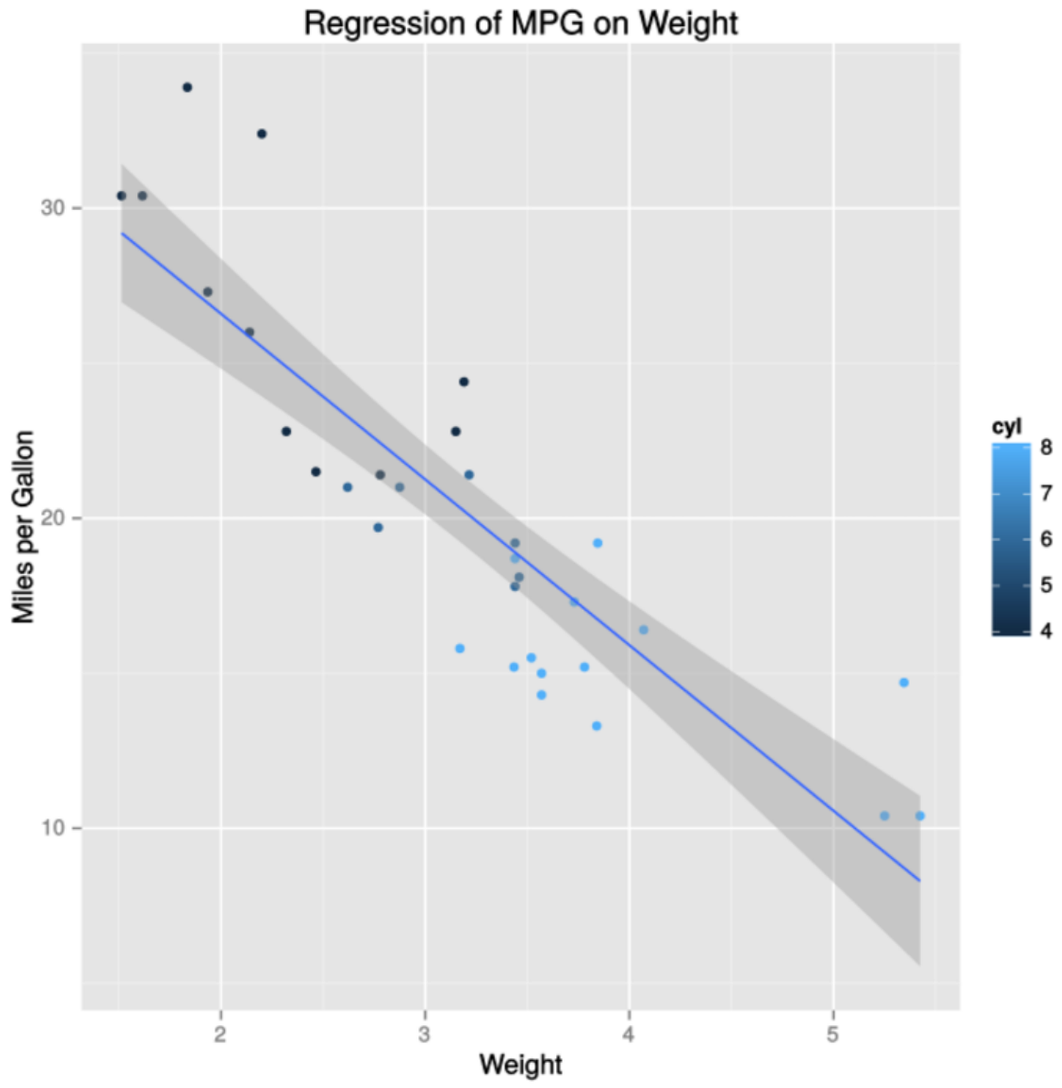
R kernel for Jupyter

https://irkernel.github.io/

https://try.jupyter.org/

In [1]:
```
library(ggplot2)
```

In [2]:
```
qplot(wt, mpg, data=mtcars, geom=c('point', 'smooth'),
   method='lm', formula=y~x, color=cyl,
   main='Regression of MPG on Weight',
   xlab='Weight', ylab='Miles per Gallon')
```



In [ ]:
```
```

Requirements

To run Jupyter with an R kernel, you need at least the following:

- Jupyter. If you already have a Python environment set up, install Jupyter using your preferred tools. If not, installing Anaconda is the quickest way to get everything you need.
- A current R installation.

[Home]

Download

CRAN

R Project

About R
Logo
Contributors
What's New?
Reporting Bugs
Development Site
Conferences
Search

R Foundation

Foundation
Board
Members
Donors
Donate

Help With R

Getting Help

Documentation

Manuals
FAQs
The R Journal
Books
Certification
Other

The R Project for Statistical Computing

https://www.r-project.org/

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To **download R**, please choose your preferred CRAN mirror.

If you have questions about R like how to download and install the software, or what the license terms are, please read our answers to frequently asked questions before you send an email.

News

- The R Foundation welcomes five new ordinary members: Jennifer Bryan, Dianne Cook, Julie Josse, Tomas Kalibera, and Balasubramanian Narasimhan.
- **R version 3.3.2 (Sincere Pumpkin Patch)** has been released on Monday 2016-10-31.
- **The R Journal Volume 8/1** is available.
- The **useR! 2017** conference will take place in Brussels, July 4 - 7, 2017, and details will be appear here in due course.
- **R version 3.3.1 (Bug in Your Hair)** has been released on Tuesday 2016-06-21.
- **R version 3.2.5 (Very, Very Secure Dishes)** has been released on 2016-04-14. This is a rebadging of the quick-fix release 3.2.4-revised.
- **Notice XQuartz users (Mac OS X)** A security issue has been detected with the Sparkle update mechanism used by XQuartz. Avoid updating over insecure channels.
- The **R Logo** is available for download in high-resolution PNG or SVG formats.
- **useR! 2016**, hase taken place at Stanford University, CA, USA, June 27 - June 30, 2016.
- **The R Journal Volume 7/2** is available.
- **R version 3.2.3 (Wooden Christmas-Tree)** has been released on 2015-12-10.
- **R version 3.1.3 (Smooth Sidewalk)** has been released on 2015-03-09.

Installation

BINARY SOURC

1/2) Installing via suppli
(default on Windows + M

You can install all packages using the following lines in ar

```
install.packages(c('repr', 'IRdisplay', '
devtools::install_github('IRkernel/IRkern
# Don't forget step 2/2!
```

To update the IRkernel package, which is not yet on CRAN, you have to rerun the `devtools::` line. For the other packages, a simple `update.packages()` is sufficient.

2/2) Making the kernel available to Jupyter

If you haven't done this already, you will have to make Jupyter see the newly installed R kernel by installing a kernel spec.

The kernel spec can be installed for the current user with the following line from R:

```
IRkernel::installspec()
```

To install system-wide, set `user` to False in the `installspec` command:

```
IRkernel::installspec(user = FALSE)
```

Important!

On OS X, be sure to execute this in R started from the Terminal, not the R App!
(This is because the R app doesn't honor *$PATH* changes in `~/.bash_profile`)

Exercises

```
import matplotlib.pyplot as plt
plt.plot(range(4))
plt.show()  —-> [shift + enter]
```

Find a Notebook you like in the *Gallery of Interesting IPython Notebooks* and try to reproduce the analysis. Make subtle changes or import a completely different data set to see what happens. Share with your neighbors.

Additional resources including Notebooks by PyLadies: https://github.com/willingc/workshop-help

Thanks!

Happy Jupytering!