Jupyter: An insider story

Matthias Bussonnier

Berkeley
UNIVERSITY OF CALIFORNIA

BIDS
BERKELEY INSTITUTE
FOR DATA SCIENCE

Slides (likely) on GitHub:
https://github.com/Carreau/talks/

Me

- 2011 - Master degree in Physics
- 2011 - 2014 PhD in Biophysics, Using / Contributing to IPython
- 2012 - … Jupyter/IPython Core dev
- 2014 - … PostDoc at UC Berkeley - Full time on **Jupyter/IPython**

Feel free to interrupt me if you have any questions

carreau

mbussonn

bussonniermatthias@gmail

This is a story about community

How Jupyter came to be

Why is it an attractive community ?

How it is currently used

what are the challenges ?

What we hope it will become

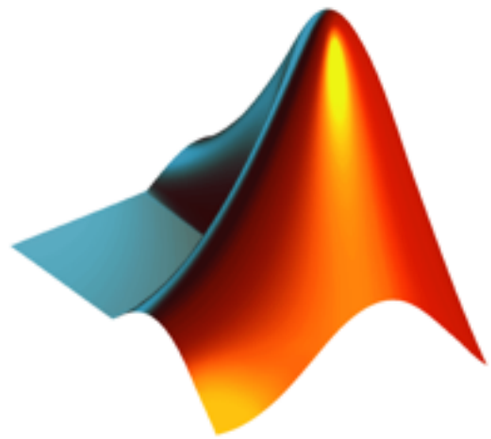
How can you be part of it?

The Dark Ages (< 2001)

FORTRAN

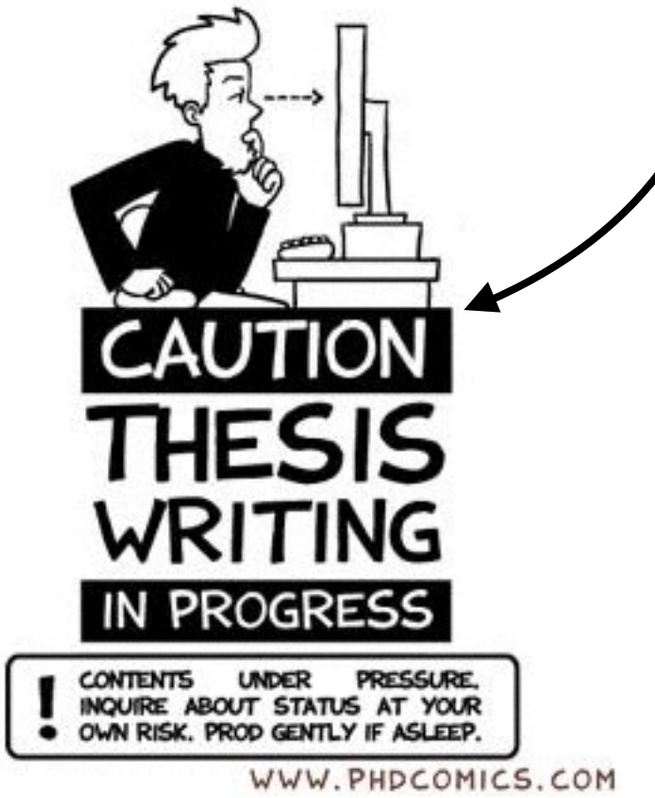
• • •

Perl



Start of a long tradition:

kamsblog.org



IP[y]:
IPython

Start of a long tradition:

IPython 0.1

(PhD can wait)

xkcd.com

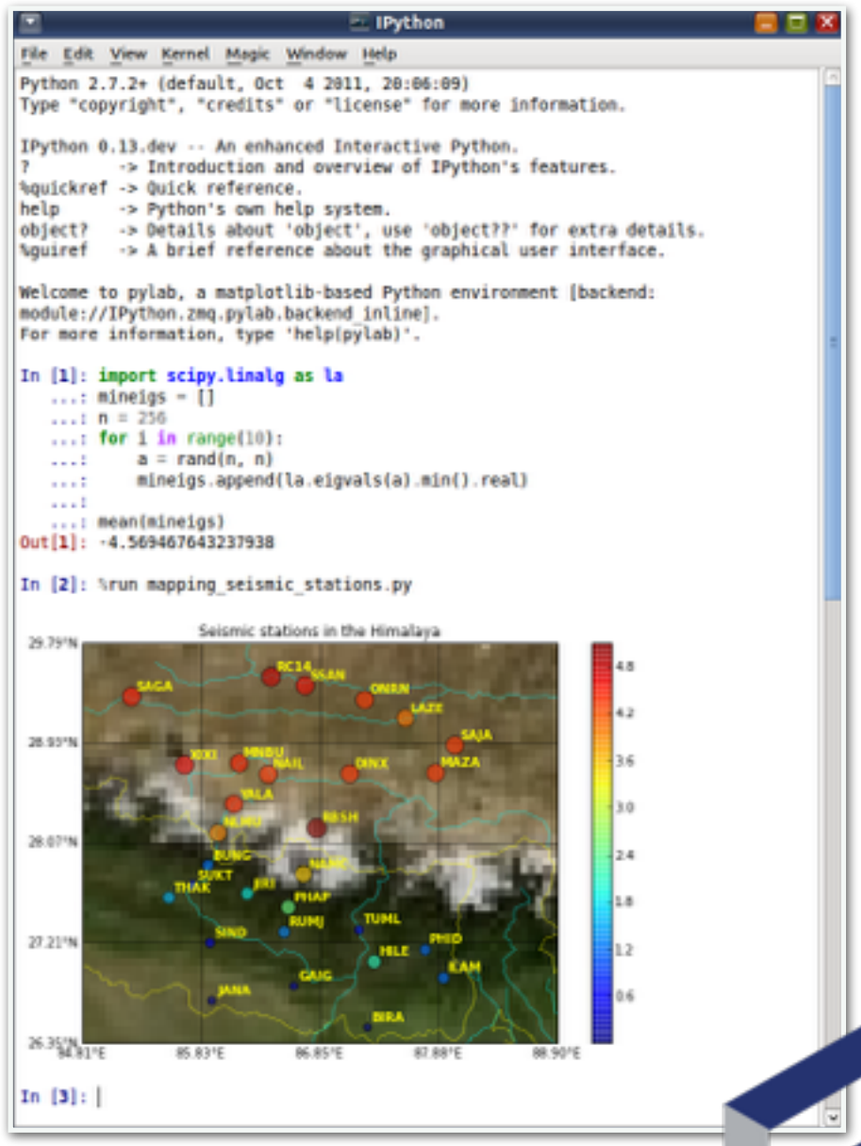
2001-2010ish

NOVEMBER 2001: "JUST AN AFTERNOON HACK"

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

2011-2012

QtConsole



- Nice & hard to use
  - No menu
  - Few shortcuts

  But really useful… can I be of help?

First, real pull request on a project I enjoyed:

- 2 changed files with 14 additions.
- 12 (remainings) comments[1]
- 7 Days in the making
  - (That's 2 lines / day)
- I was **not** a Python expert:
  - "Semicolon at end of line are unnecessary"

I probably worked 50h on that. If 1000 users gain 30s each week for a year that's 833h. SO it was worth it.

[1] (GitHub did not kept intermediate commit at the time)

~500 Pull requests later,

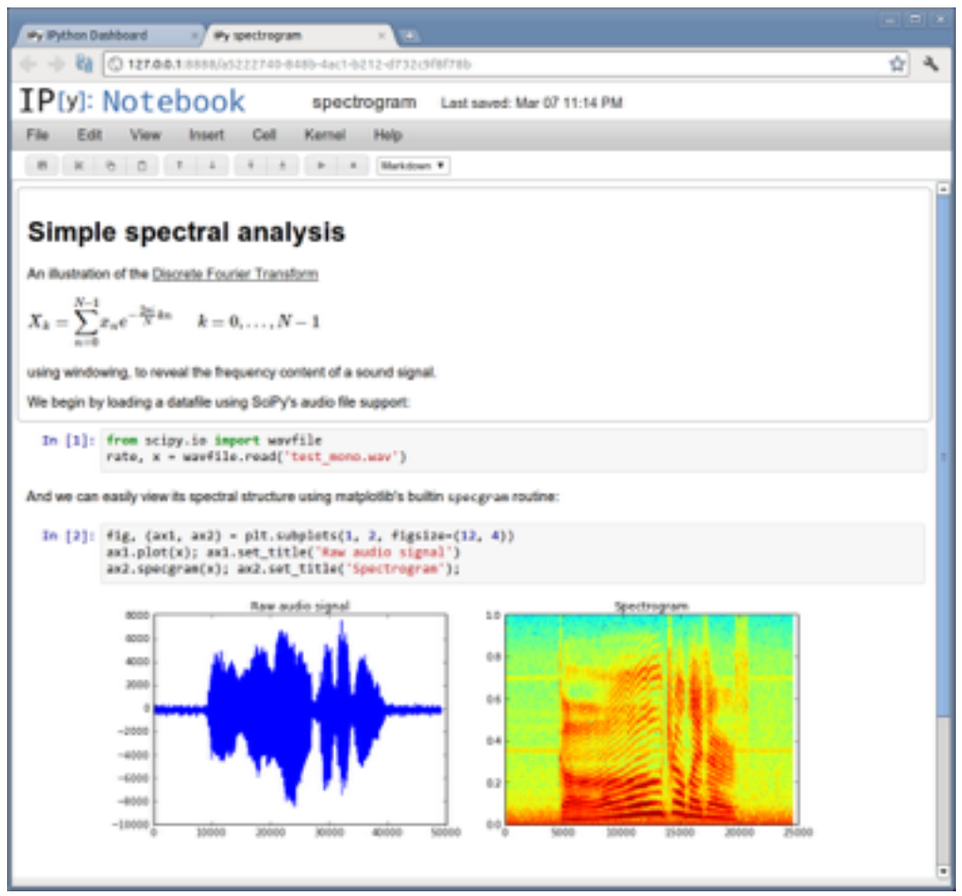
"I'll likely spend two days teaching someone how to do it, than doing it myself"

That's how I was welcomed to the project, that's how we (try to) welcome new contributors.

(I'm still not a Python Expert)

Not the first notebook

6th Notebooks Prototype



WHY THIS ONE WORKED

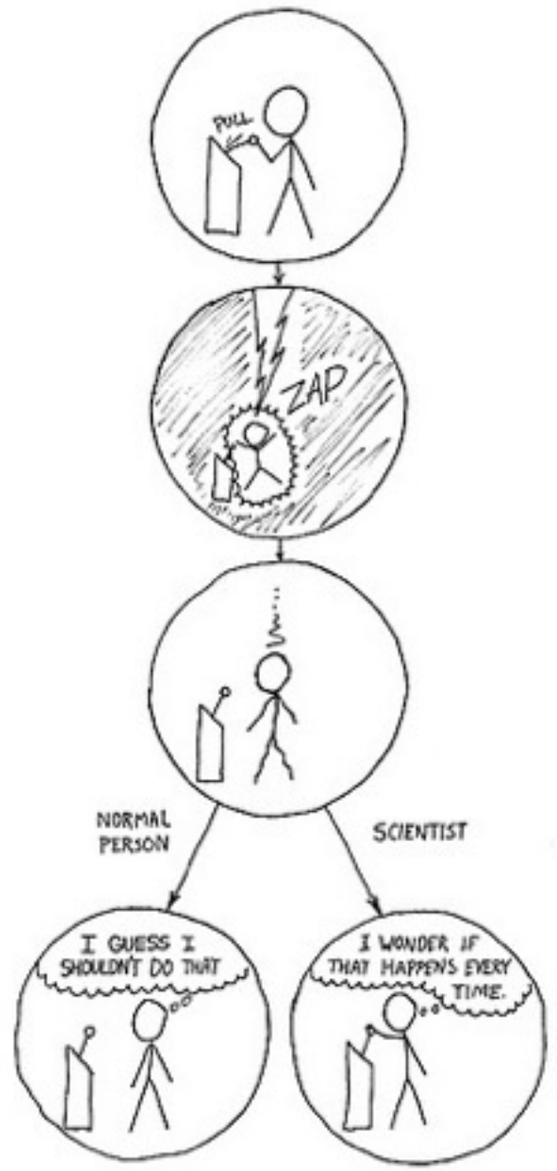
- Mature Web technologies
  - Websocket vs Ajax
  - Javascript (v8)
  - MathJax…
  - JSON vs XML
- More mature protocol
- Separation of Concern
- Feedback from Sage Math

- The Community

(SciPy 2012 good example)

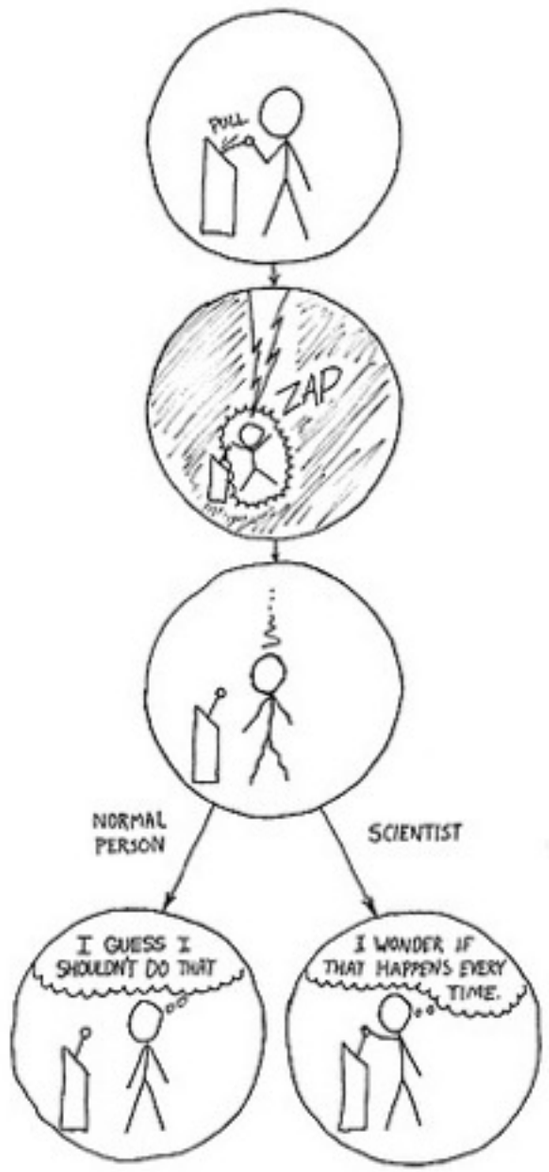
The Community

- Started mostly with scientists
  - dog fooding
  - we build tools we need
- Most of us are not software engineers
  - You can likely help us to make it better
  - Developing Jupyter is not hard, it's long
- The Core team does publish a lot of notebooks.
  - Gallery of interesting notebook
  - Nbviewer
  - Ligo
    - All by Community



The Community

- We love to write software for our users
  - Writing software should be fun
- We love to teach
  - We prefer to let people do PR instead of taking over.
  - We don't like to just give the solution away.
- We are Learning
  - We know that we don't know (especially JS)
  - We can't be expert in 50 languages.



"Data Scientists love Python so much that they write Python all Day.

IPython developper love Python so much, that they write JavaScript all day."

–Jake van der Plas

The Community

jupyter.org/community.html

Jupyter General Mailing List

A Google Group for general Jupyter discussions of Jupyter's use.

Jupyter in Education Mailing List

A Google Group for general Jupyter discussions of Jupyter's use in education.

Jupyter on GitHub x 65 535 repositories

GitHub is our primary place to report issues with Jupyter.

Jupyter Developer Chatrooms

A place for contributors to ask and receive brief Jupyter development questions.

Jupyter on Stack Overflow

A third party question and answer site for programmers to share information about Jupyter.

- We do not distinguish developers and users.
  - No -dev vs -user mailing list.

- Contributions is not only “writing code” or “fixing bugs”

- Dev meetings recorded and on Youtube (weekly recap of who did what)

How is Jupyter used,

& Challenges

Desktop Application

Kernel-gateway

Data Science

Blogging

Detection of Gravitational waves

"What's programming ?"

<your use case>

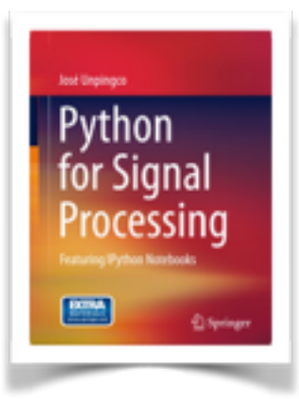
Multi-user deploy

rackspace

Deploying JupyterHub for Education

On Demande notebook Server

binder



Hard cover Books

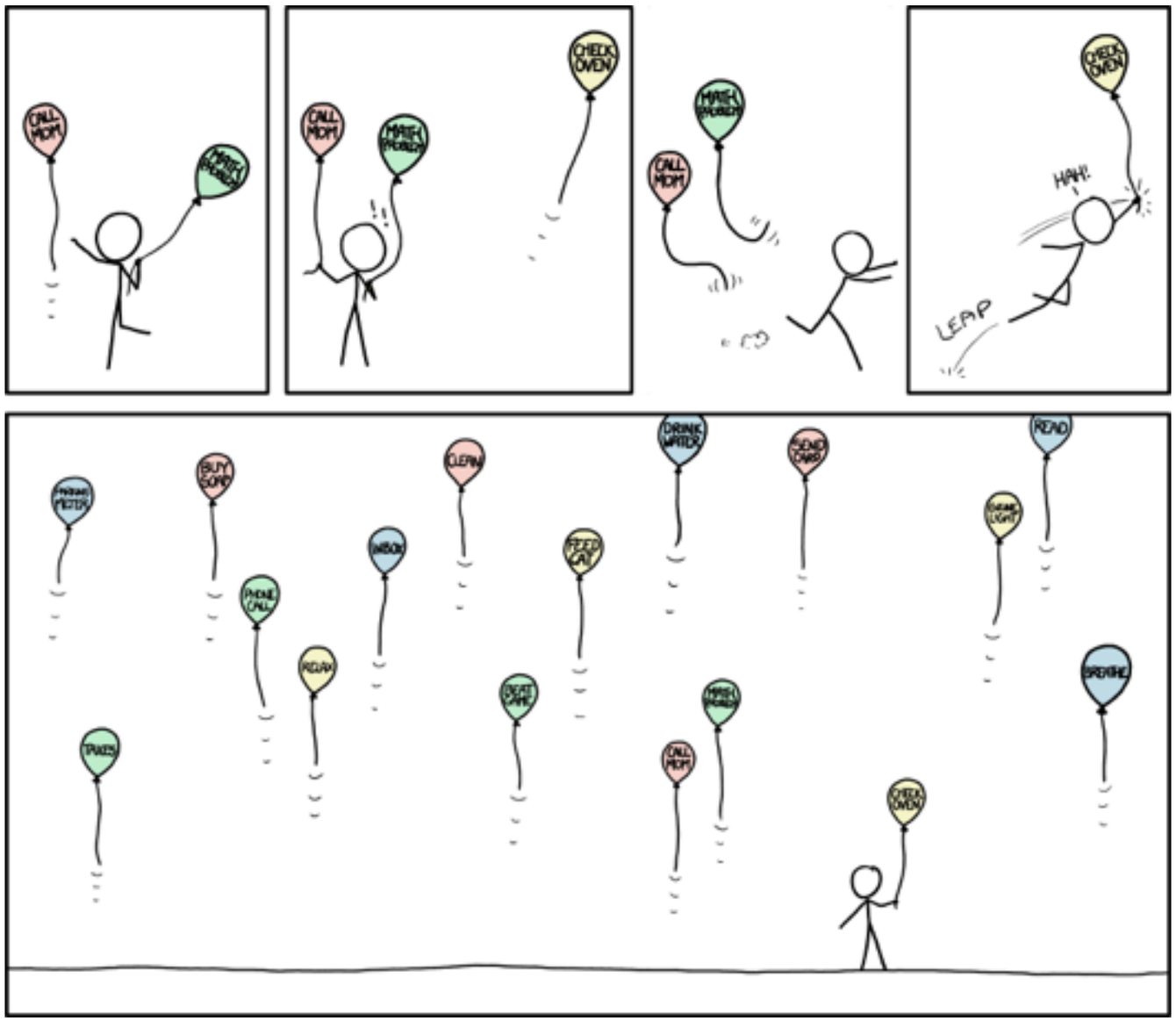
O'Reilly Data

O'REILLY®

Biggest Deployment:

Your Laptops !

Challenges



(As usual there is ~~an~~ two xkcd for that)

"Clients may maintain connections to the server for no more than 86,400 seconds per day. If you need additional time, you may contact IERS to file a request for up to one additional second."

Can’t follow everything

plotly

NOVARTIS

EMC {code}

CfA Harvard-Smithsonian Center for Astrophysics

GenePattern

Can't respond to everything

- Really hard **not** to respond to someone asking a question
  - Don't know
  - Some else will respond the wrong thing
    - You should give others the change to get involved.
  - Respond too late

- What should I prioritize
  - Fix Bugs / Leave them open for new comers to fix ?
  - Write Docs
  - Write features
  - Use Jupyter to figure out what's wrong
  - Review other's Pull Requests

Can’t respond to everything

- Bug reports for beginner user
  - Pandas, Matplotlib, Anaconda, Python, Windows, OS X, IBM workbench.
  - Many “Beginner” user don’t make the difference Jupyter/the language/library they use.

- Bug report from advance user.
  - Race condition when deploying JupyterHub on 50 nodes with a couple hundreds of users.

Lead to subtle issues

Time, makes us communicate in short, confusing sentences.
Often with lots of implicit knowledge.

nbconvert/nbconvertapp.py

```
@@ -323,19 +338,23 @@ def init_single_notebook_resources(self, notebook_fil
323 338
324 339         return resources
325 340
326     -    def export_single_notebook(self, notebook_filename, resources):
    341 +    def export_single_notebook(self, shell_input, resources):
```

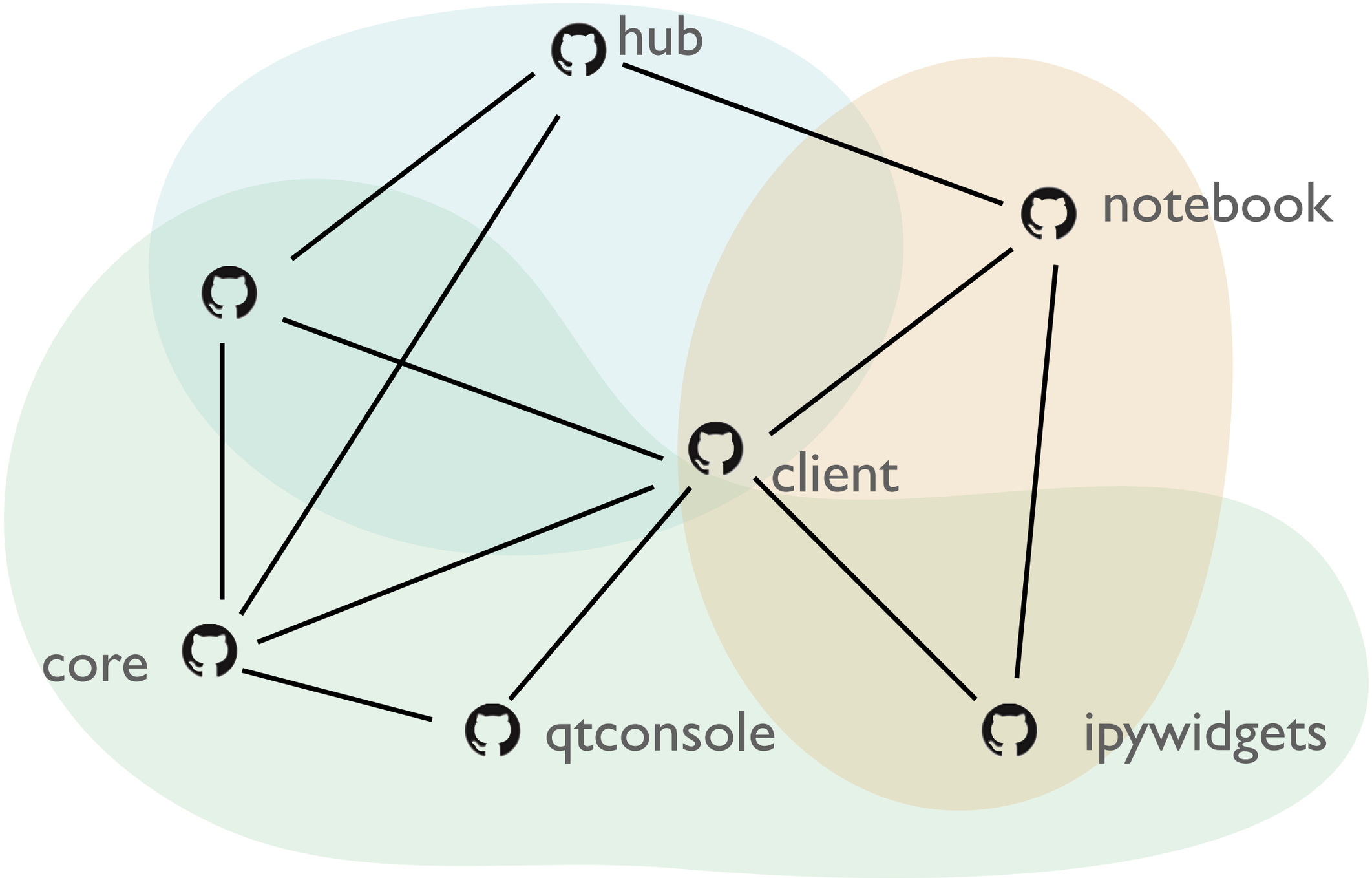
Carreau added a note on Dec 8, 2015 — Project Jupyter member

can you rename `shell_input` here back to `notebook_filename` (also in docstring)

You can assign after but the comment than changing the signature break api is right.

Time, makes us communicate in short, confusing sentences.
Often with lots of implicit knowledge.

The Big Split



"Core" developers can focus on a subset of component

hard time figuring out where to open a bug

Automation

Communication

Community

An open source project is more than just the license.

Documentation

Trust

Bus Factor

DevOps

Jupyter

- Is, and starting to appear more as an ecosystem than single tool.

- The "core"
  - Provide a default set of component
  - set a basic & sane configuration.
    - (I'm looking at you Vim and Emacs)
  - Is a "reference" implementation

- People from allover the world
  - Most of the best extensions, usage, demo, … are not in the core.

Not everyone can read/respond to 200 email/day
Try a more curated channel

github.com/jupyter/newsletter

Greetings! This update is the first installment of a newsletter we've started to help keep our global community up to speed on all things Juptyer. Keeping in line with Jupyter's nature, it is our goal that the newsletter be a collaborative effort. If you have ideas or content for the newsletter, feel free to add it to our Newsletter Git repository - or just open an issue with interesting links if you are unfamiliar with pull requests. If you know of anyone who may have an interest in learning about or getting involved with the project, share our subscription link. We welcome feedback from the community.

JupyterDays Boston

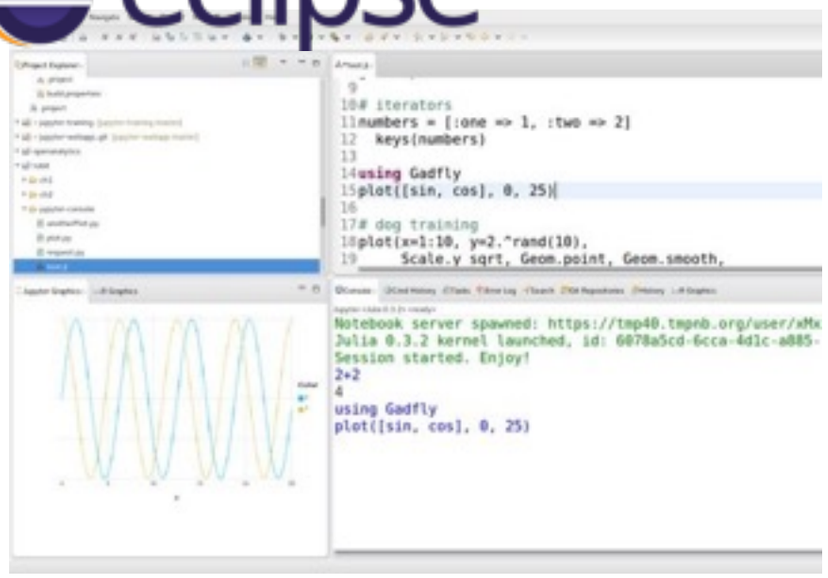
This Thursday and Friday, O'Reilly Media and the Harvard-Smithsonian Center for Astrophysics Library and the Harvard Law School Library will be hosting the third official JupyterDay event, complete with talks, discussions, hands-on training sessions, and

Please come help with the newsletter.

The Future

Code Edition

eclipse

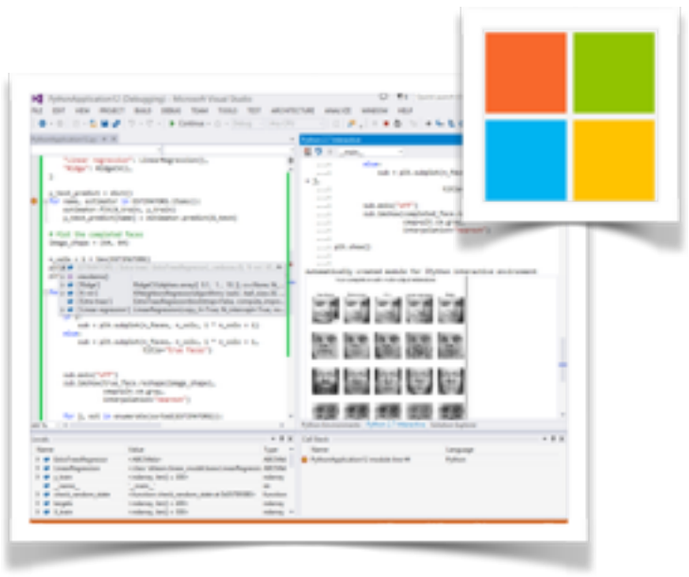


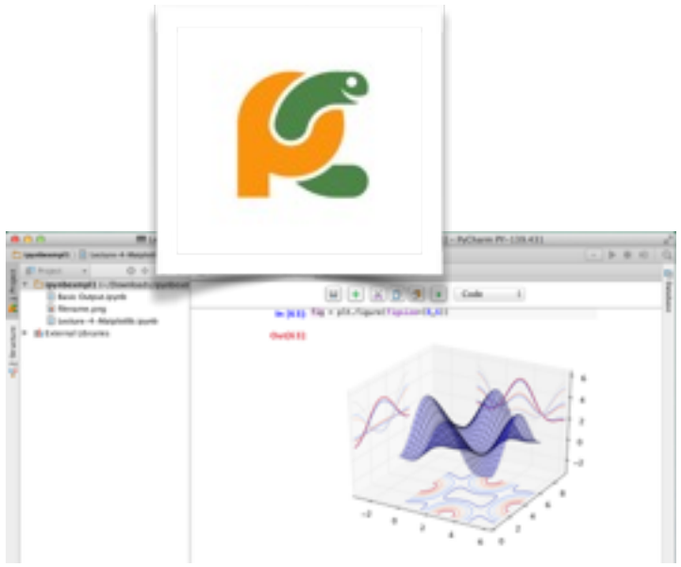
nteract

hydrogen

ick

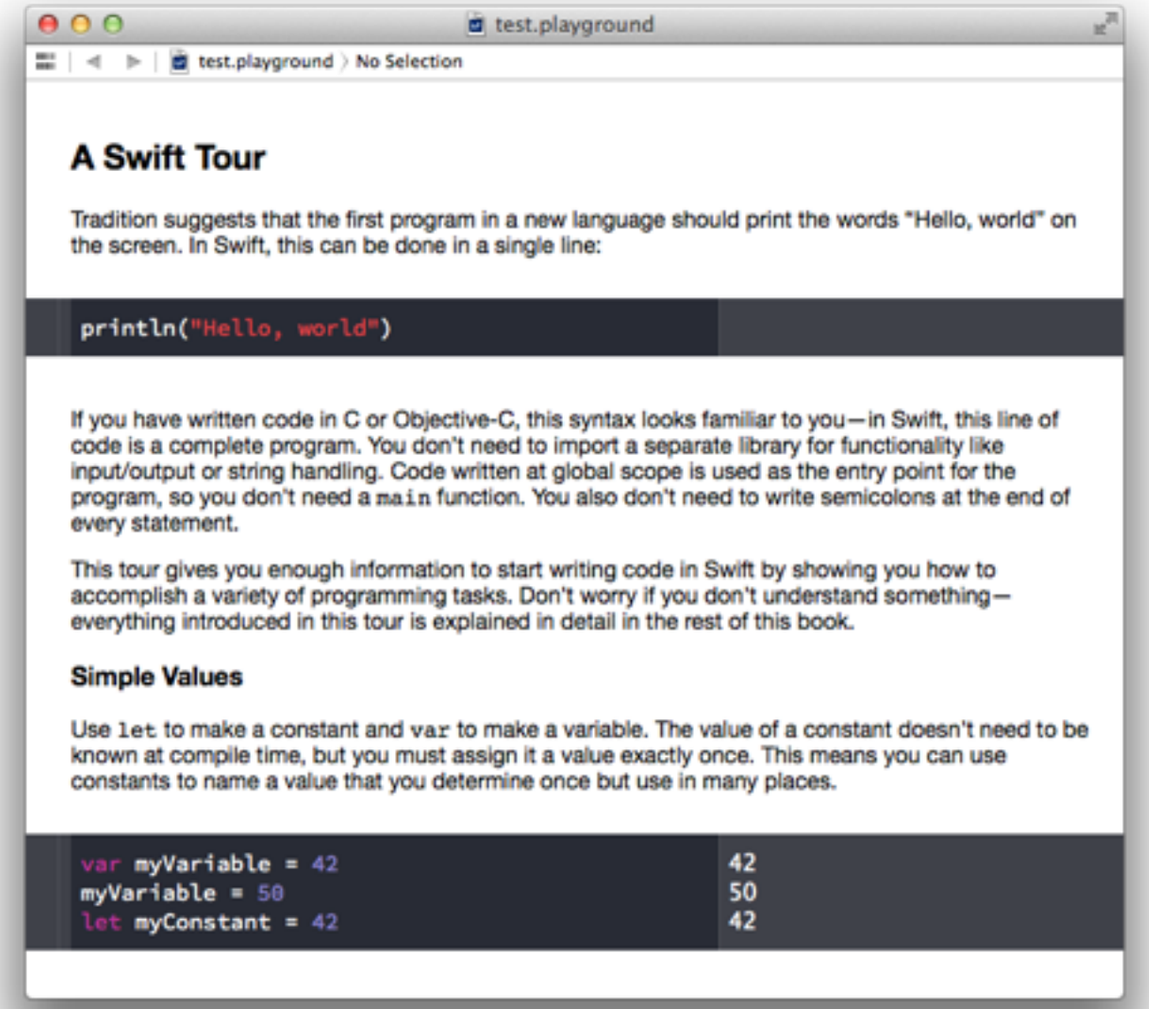
PTVS



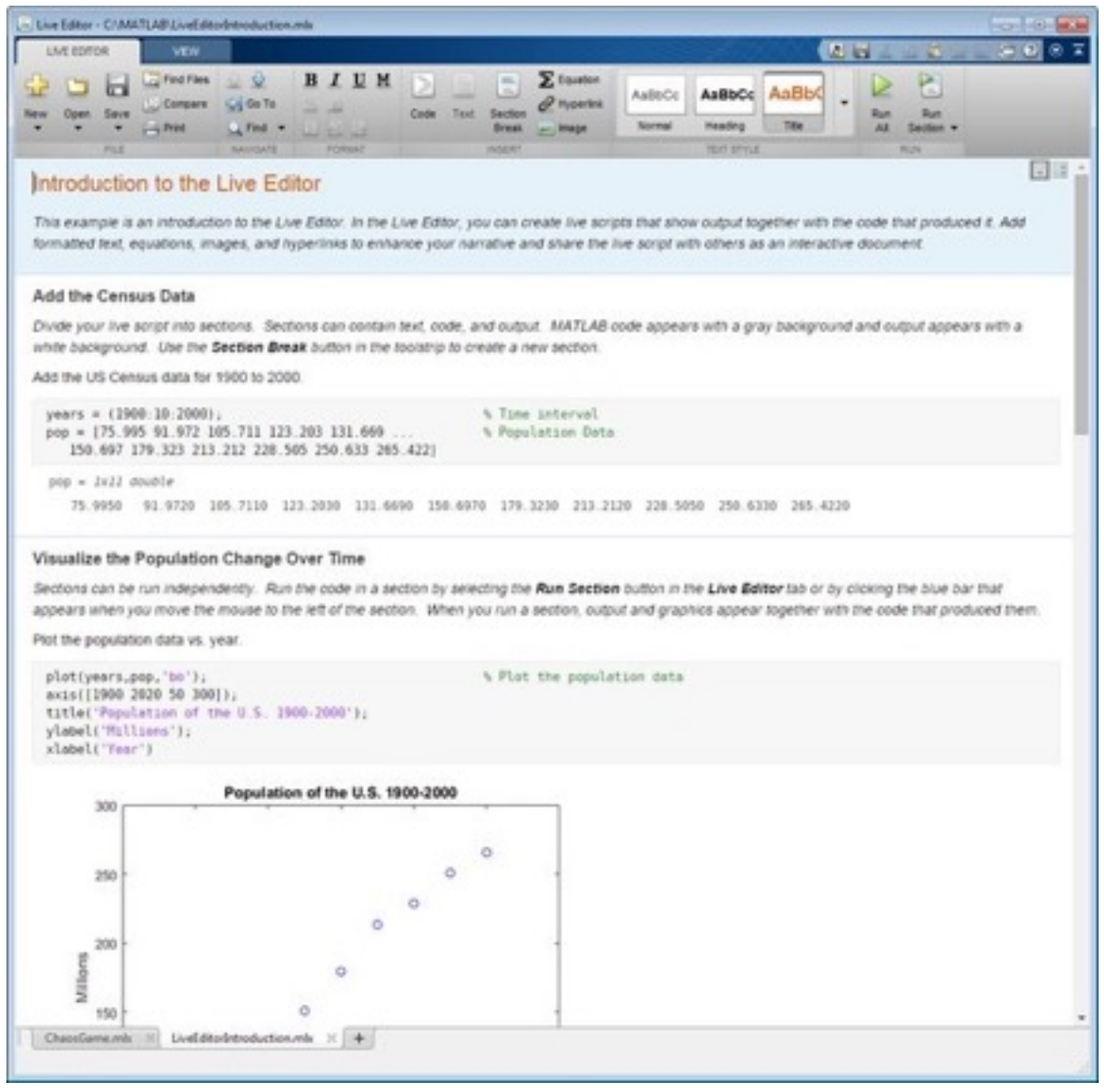


Pycharm

Think of Jupyter-* as a reference implementation



Fruit-Company



2DArray-Laboratory

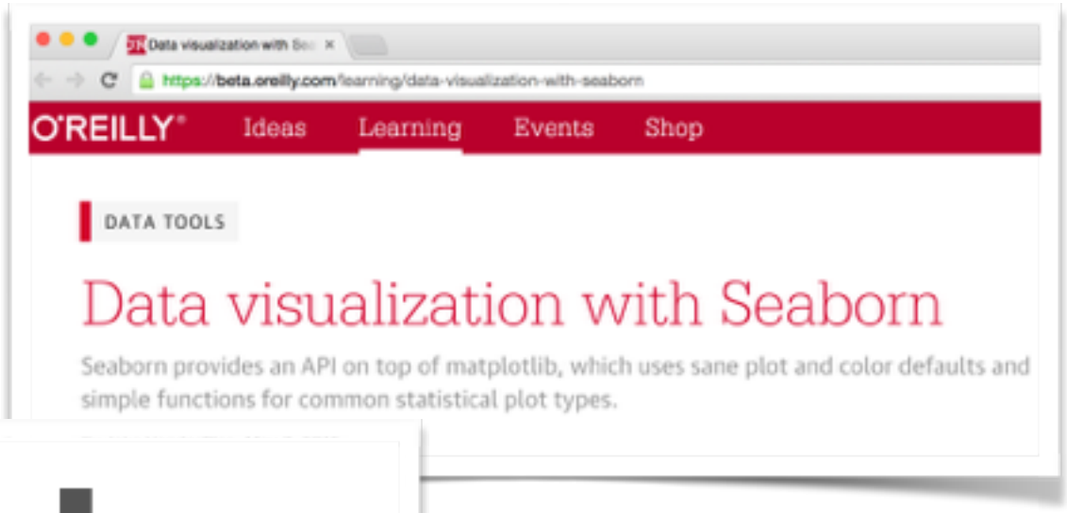
jupyterdays 30
March 17-18 | Boston

The Future

Publication

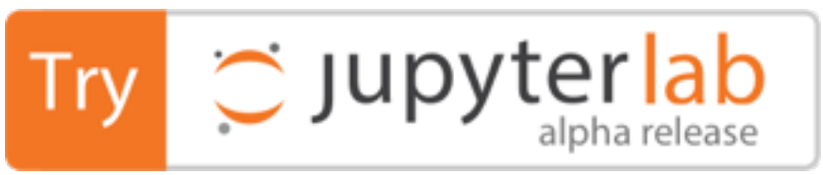
Simplifying author workflow, and make user visible document more interactive.

- O'reilly Thebee
- Kernel Gateway
- Binder
- nbconvert
- Static widgets
- Dashboards



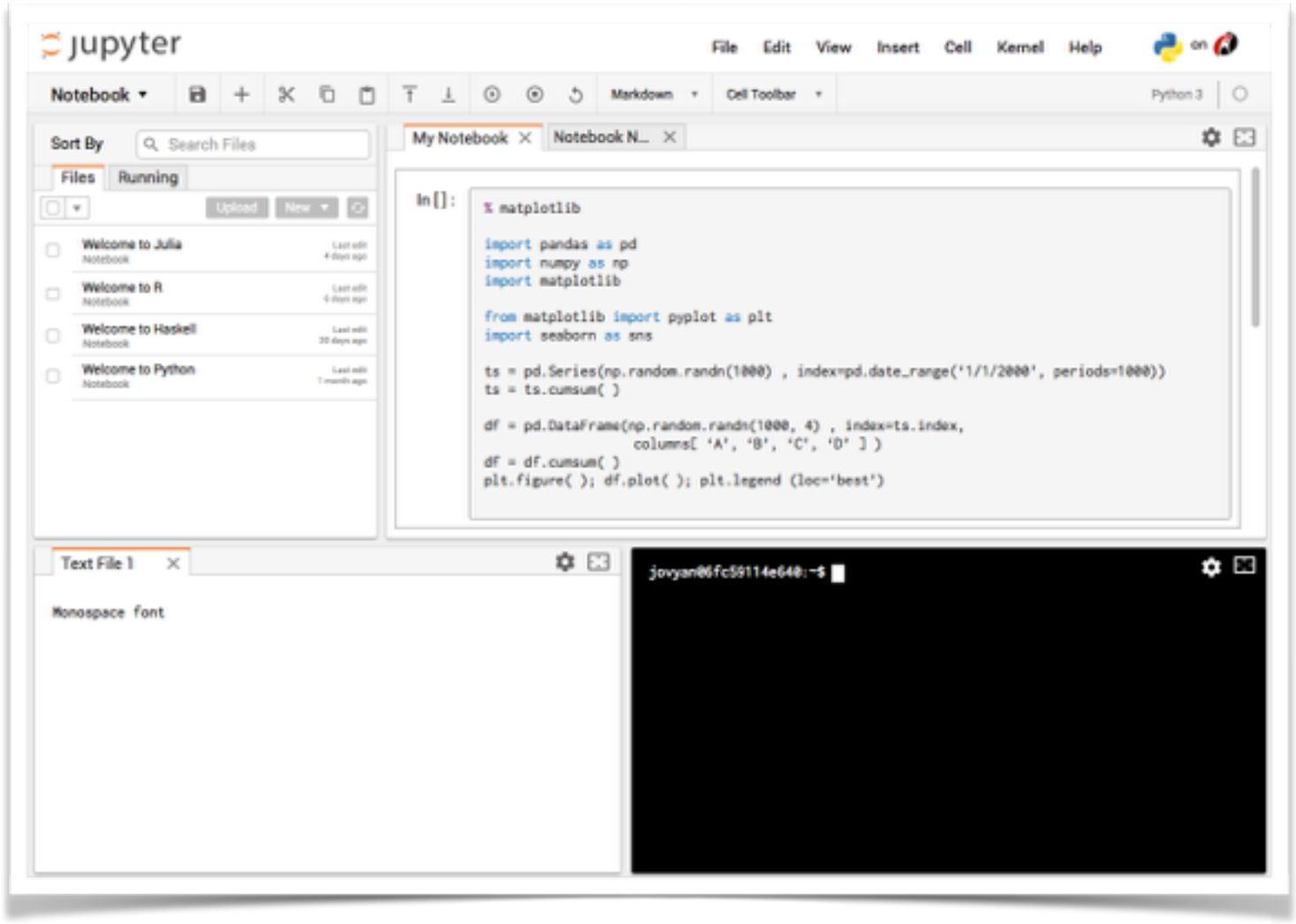
The Future

Where is the current core going ?
(currently on master)



Demo

Each pannel is an independant component



Dasboard

jupyterdays 34
March 17-18 | Boston

jupyter scotch_dashboard (unsaved changes)

File Edit View Insert Cell Kernel Help | Python 3

Markdown CellToolbar View:

```
    js = Javascript("IPython.notebook.events.trigger('select.factors_keys', {factors_keys: ['%s']});" % name)

    return display(html, js)
```

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.

In [15]:
```
prompt_w = widgets.HTML(value=tmpl.render(name='Aberfeldy'))
prompt_w
```

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.

In [16]:
```
picker_w = widgets.interact(on_pick_scotch, Scotch=list(sim_df.index))
```

Scotch Dalmore

```
/home/main/anaconda2/envs/python3/lib/python3.4/site-packages/ipykernel/__main__.py:2: FutureWarning: order is deprecated, use sort_values(...)
  from ipykernel import kernelapp as app
```

	Similarity
RoyalLochnagar	0.970362
Dalluaine	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.

In [17]:
```
radar_w = RadarWidget(df=features_df)
radar_w
```

And finally, we'll include a footer in our dashboard with some attribution information.

jupyter scotch_dashboard (autosaved)

File Edit View Insert Cell Kernel Help | Python 3

Markdown CellToolbar View:

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

	Similarity
RoyalLochnagar	0.970362
Dalluaine	0.959166
GlenOrd	0.945333
Glendullan	0.945112
BlairAthol	0.944089

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.

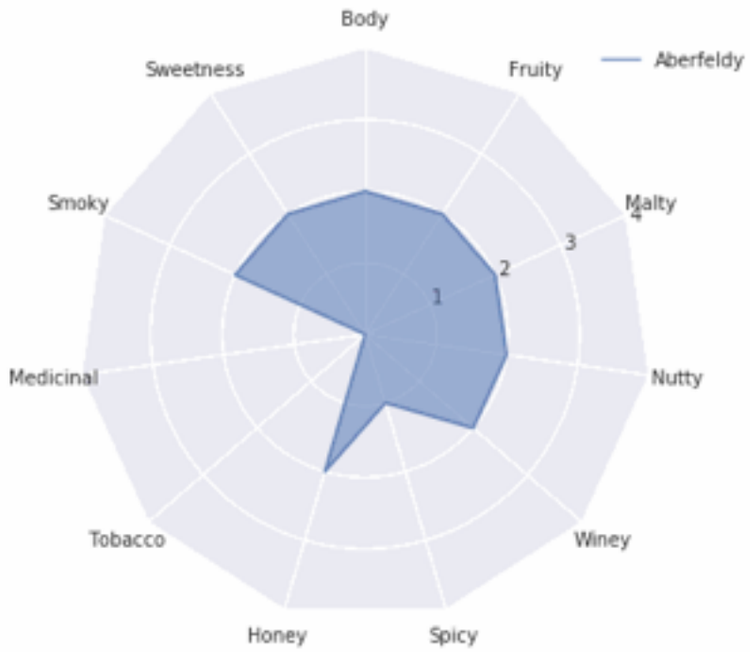
We will have a short intro line indicating which scotch you're examining, followed by a dropdown of all the scotches to pick from. Then on the selection of a scotch, a recommendation table will display the top 5 most similar scotches to your original selection. We will also have a radar chart to show 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.

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

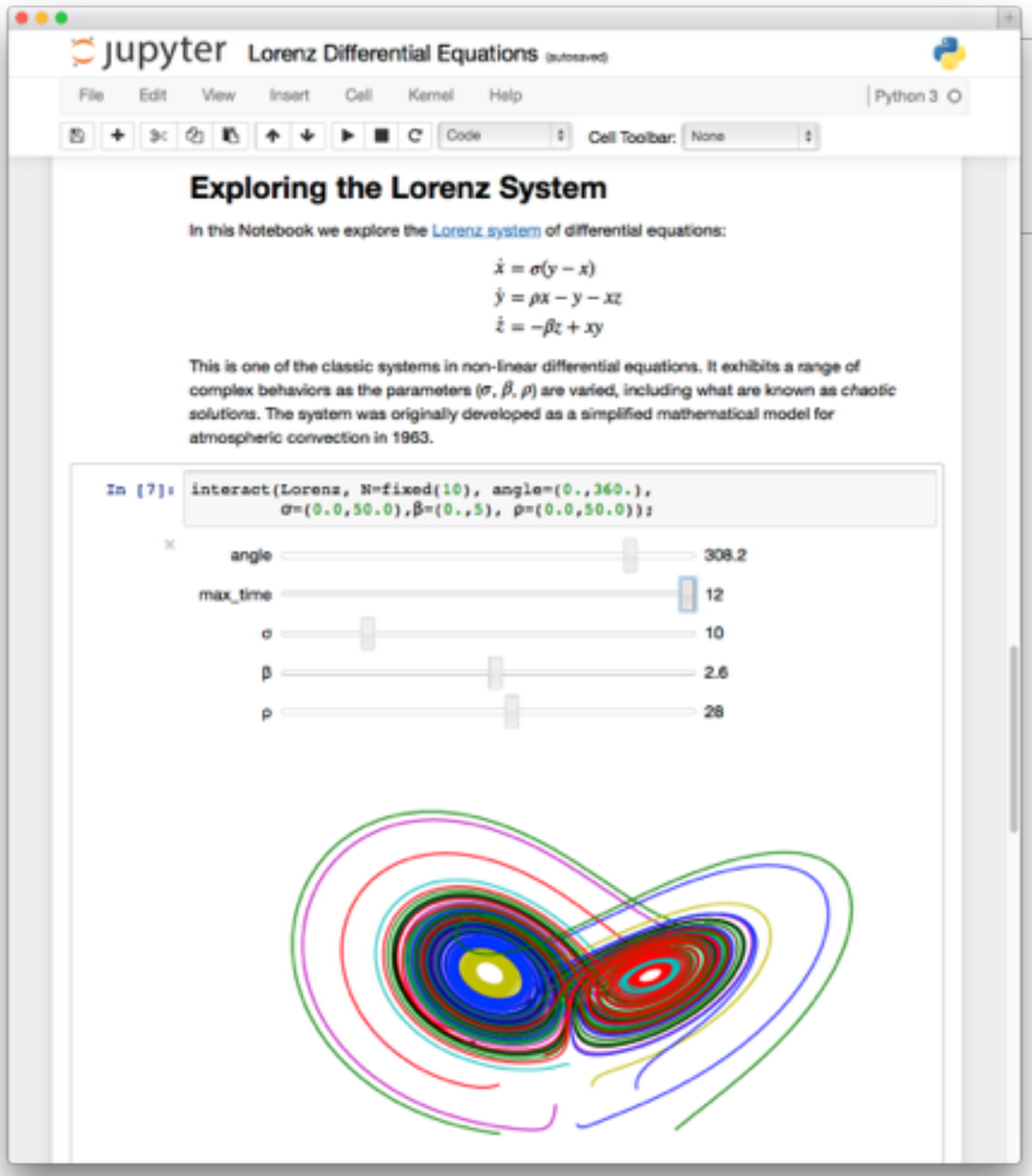


Powered by data from https://www.mathstat.strath.ac.uk/outreach/nessie/nessie_whisky.html and inspired by analysis from http://blog.revolutionanalytics.com/2013/12/k-means-clustering-86-single-malt-scotch-whiskies.html. This dashboard originated as a Jupyter Notebook.

jupyterdays 36
March 17-18 | Boston

incubator & other projects

- Content management
  - Indexing
  - Search
  - Execute notebook as library
  - Execute notebook as templates
  - …
- Separate component
  - Phosphorjs-*
  - nteract-*
  - …
- Kernelless interactivity
  - mpld3
  - bqplot
  - …



Goals

- Make interactive computing more accessible
- Build a set of tools which helps both beginners and advance users.
- Grow a community of user who are knowledgeable on how to use **and develop** the necessary tools
- Find people with interesting ideas.

```
In [4]: for i in (s,s2):
   ...:     print(i)
   ...:
The IPython shell now have syntax highlighting
and real multi-line editing
```

ALFRED P. SLOAN FOUNDATION

GORDON AND BETTY MOORE FOUNDATION

SIMONS FOUNDATION

POWERED BY rackspace
the open cloud company

CONTINUUM ANALYTICS

ENTHOUGHT
SCIENTIFIC COMPUTING SOLUTIONS

Microsoft

Google

Bloomberg

THE LEONA M. AND HARRY B. HELMSLEY CHARITABLE TRUST

A few new people

- Usually at CalPoly, here today:
  - Ana
  - Safia
  - Carole
- Berkeley
  - 2 New post docs
  - Project Manger
- Internships
  - A few students at Berkeley and CalPoly

We want you

- Come on the 29th in Berkeley [1]
- Organize the next JupyterDays !
- Add your project to the newsletter !
- Write a blog post (on blog.jupyter.org ?)
- Participate on the mailing lists/GitHub/gitter…

[1] www.meetup.com/Berkeley-Data-Science-Meetup/events/229439651/

Looking forward to get **you** working with us, and see what **you** made with Jupyter

Time for questions ?

Apologies for the typos, this was written on a mobile phone with autocorrect enabled, in a plane, while arguing on bugs, fixing Twitter (and vice versa). [or not]

jupyterdays 43
March 17-18 | Boston

jupyterdays 44
March 17-18 | Boston