Jupyter: An insider story

Matthias Bussonnier







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

(OMG I'm not wearing Green)













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?



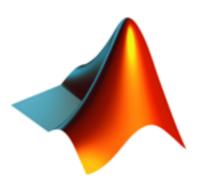
The Dark Ages (< 2001)



FORTRAN

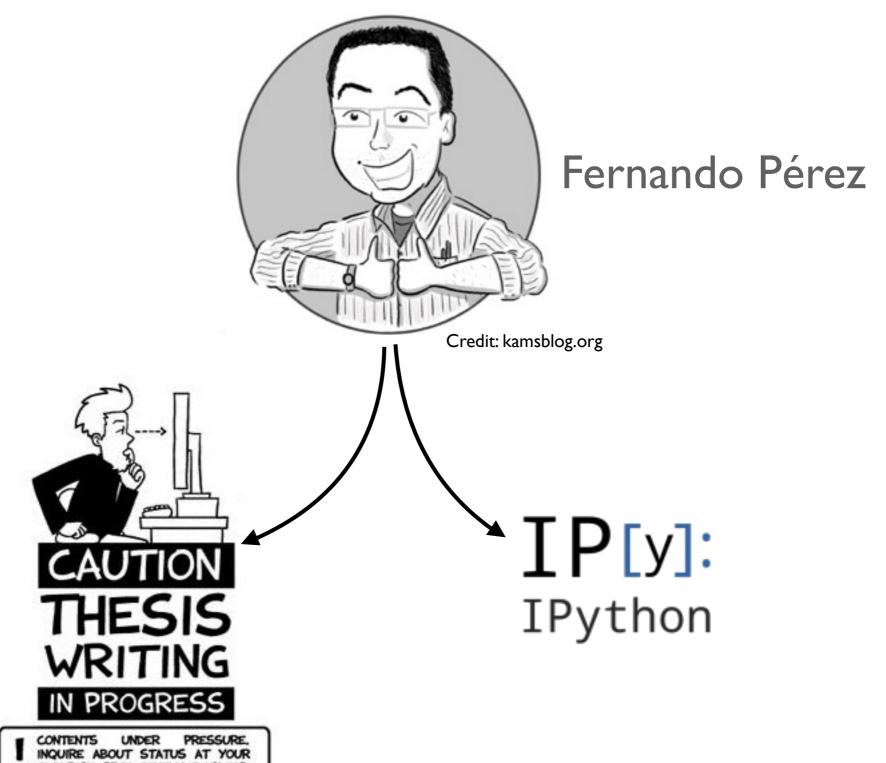
• • •







Start of a long tradition:



OWN RISK. PROD GENTLY IF ASLEEP.

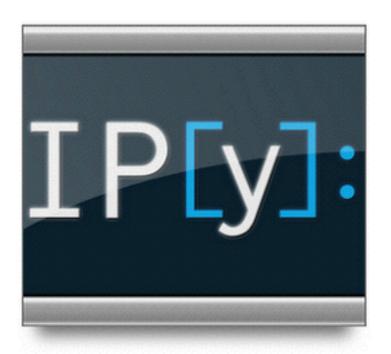
WWW. PHDCOMICS. COM



Start of a long tradition:



THE BEST THESIS DEFENSE IS A GOOD THESIS OFFENSE. XKCd.com

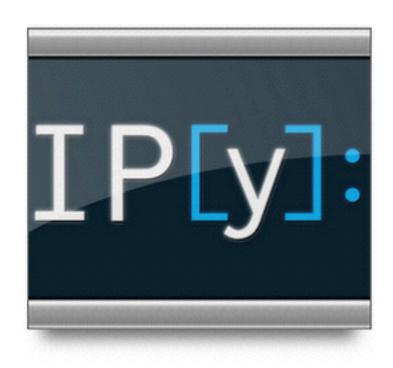


IPython 0.1

(PhD can wait)



2001-2010ish



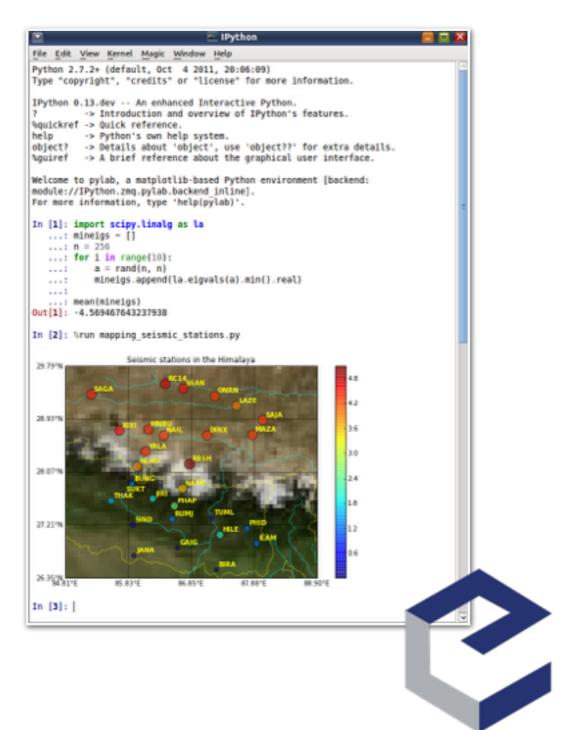
NOVEMBER 2001: "JUST AN AFTERNOON HACK"

- 259 Line Python script.
- sys.ps | -> ln[N].
- sys.displayhook -> Out[N], caches results.
- Plotting, Numeric, etc.



2011-2012

QtConsole



Nice but can be 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.
- I2 (remaining) comments[I]
- 7 Days in the making
 - (That's 2 lines / day)
- I was <u>not</u> 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.



~500 Pull requests later,

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

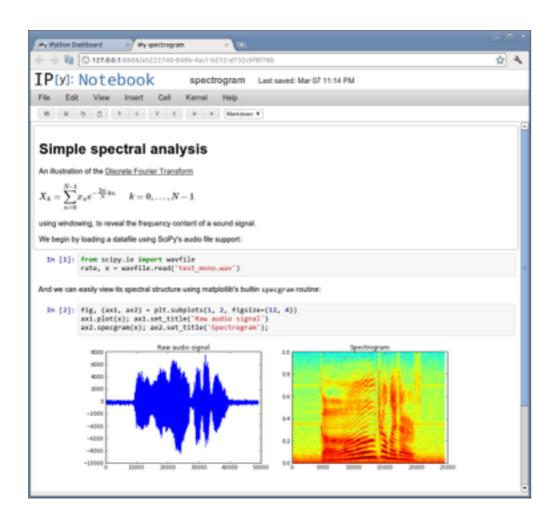
That's how I was welcomed to the project, that's how we (try to) welcome new contributors.

(I'm still not a Python Expert)



Not the first notebook

6th Notebooks Prototype



WHYTHIS ONE WORKED

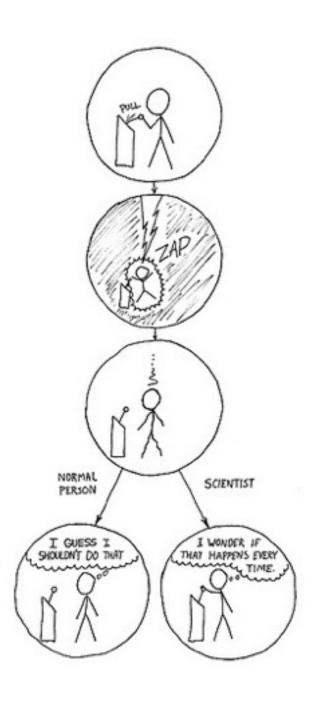
- Mature Web technologies
 - Websocket vs Ajax
 - Javascript (v8)
 - MathJax...
 - JSON vs XML
- More mature protocol
- Separation of Concern
- Feedback from Sage Math
- The Community

(SciPy 2012 good example)



The Community

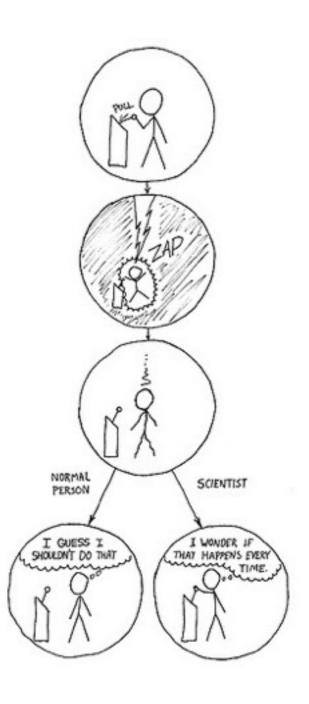
- Started mostly with scientists
 - dog fooding
 - we build tools we need
- Most of us are not software engineers
 - You can likely help us to make it better
 - Developing Jupyter is not hard, it's long
- The "Core" team does publish a lot of notebooks
 - Gallery of interesting notebook
 - Nbviewer
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





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 × 65 535 repositories

GitHub is our primary place to report issues with Jupyter.



Jupyter Developer Chatroom S

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

Google





Bloomberg

O'REILLY"

Desktop Application

CONTINUUM









Kernel-gateway





















Data Science





Detection of Gravitational waves

Blogging

"What's programming?"

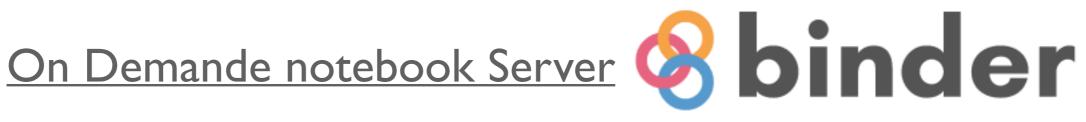
<your use case>

Multi-user deploy

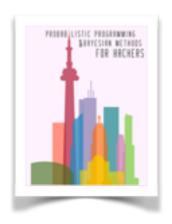


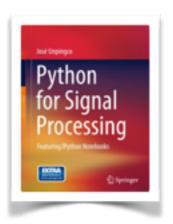


Deploying JupyterHub for Education









Hard cover Books

O'Reilly Data





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













Can't respond to everything

- Really hard <u>not</u> 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 Documentation
 - Write features
 - Use Jupyter
 - Review other's Pull Requests



Can't respond to everything

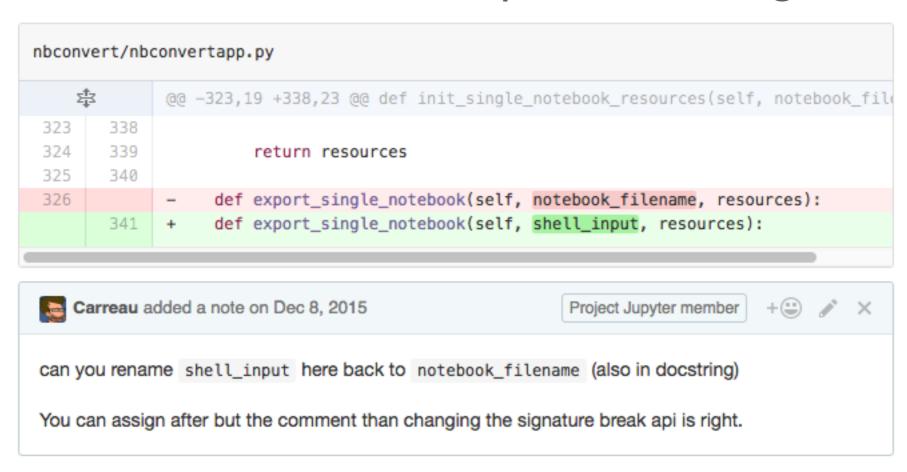
- Bug reports for beginner user
 - Pandas, Matplotlib, Anaconda, Python, Windows, OS X, IBM workbench.
 - Many "Beginner" user don't make the difference Jupyter/the language/library they use.
- Bug report from advance user.
 - Race condition when deploying JupyterHub on 50 nodes with a couple hundreds of users.



Lead to subtle issues

Time, makes us communicate in short, confusing sentences.

Often with lots of implicit knowledge.

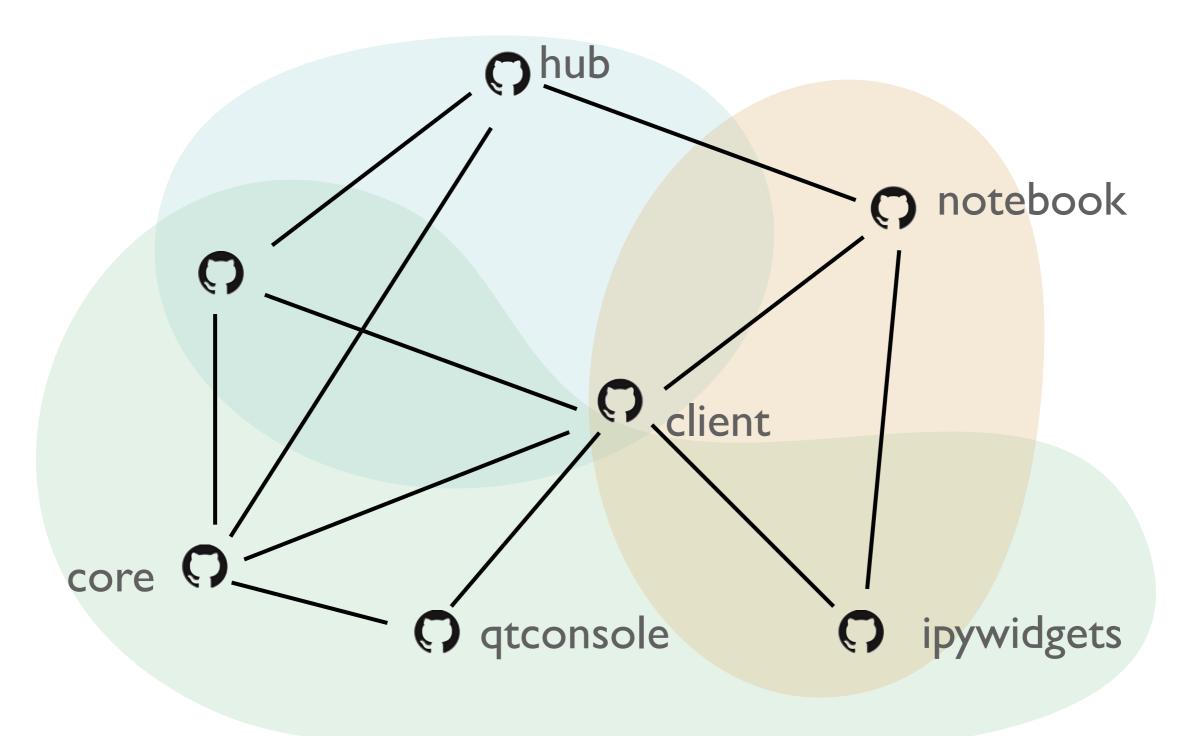


Time, makes us communicate in short, confusing sentences.

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



Newsletter 1 I March 15 I Hello World

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

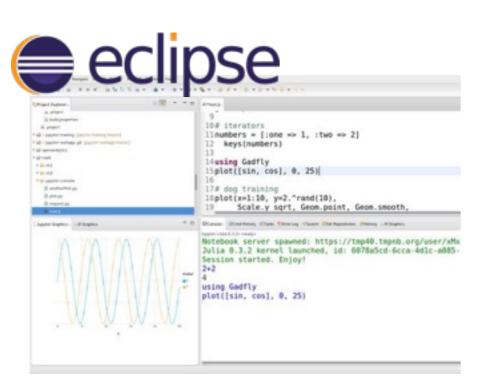
This Thursday and Friday, O'Reilly Media and the Harvard-Smithsonian Center for Astrophysics Library and the Harvard Law School Library will be hosting the third official JupyterDay event, complete with talks, discussions, hands-on training sessions, and

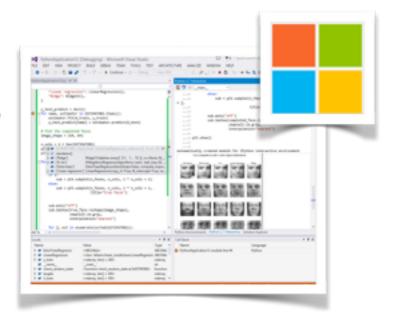


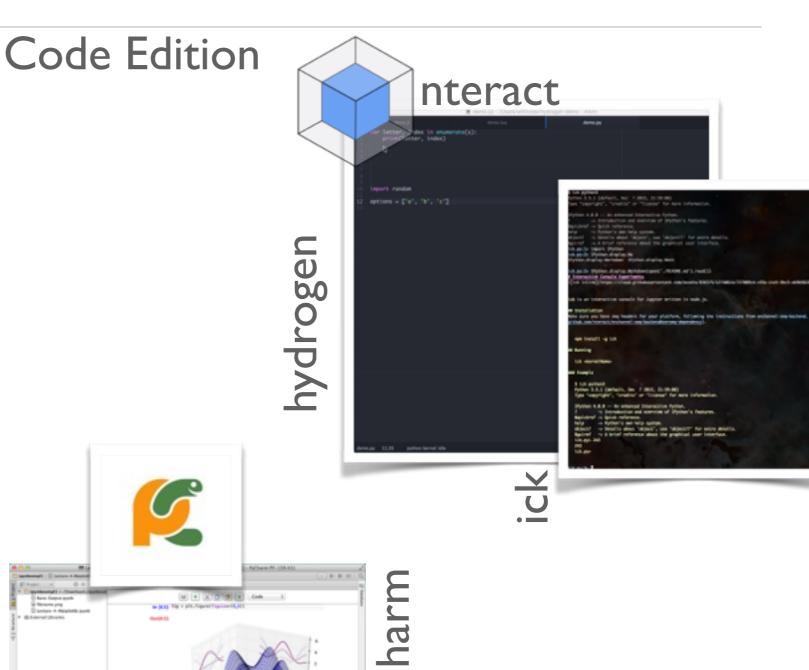
Please come help with the newsletter.



The Future

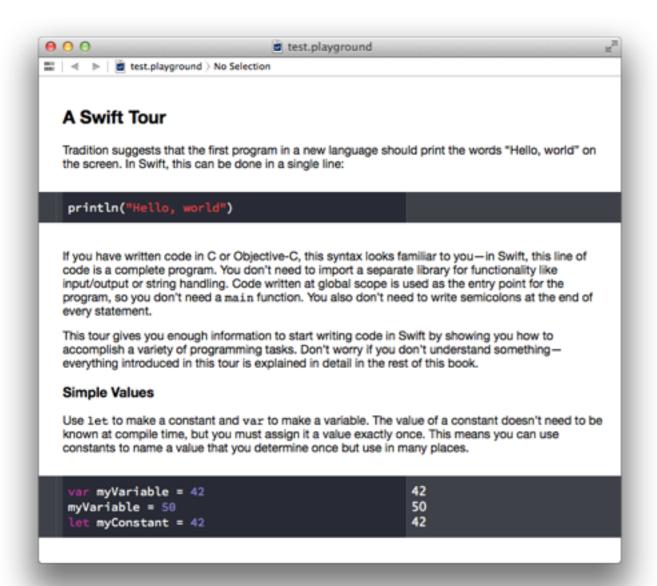




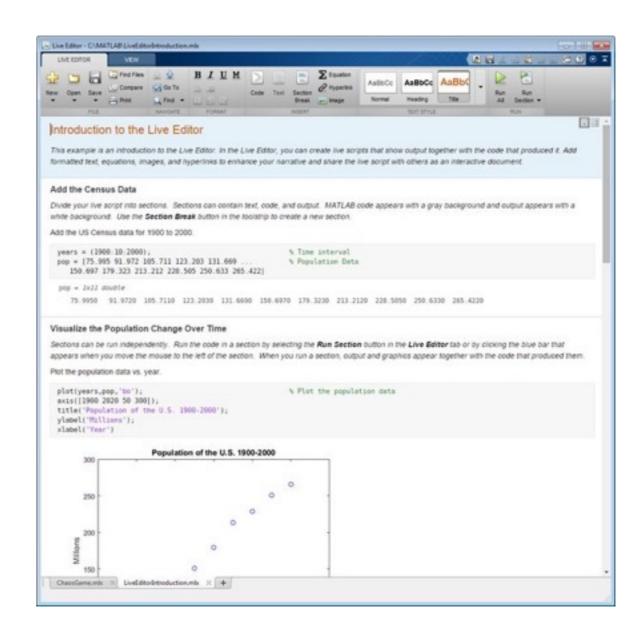


Think of Jupyter-* as a reference implementation Jupyterdays 29





Fruit-Company



2DArray-Laboratory

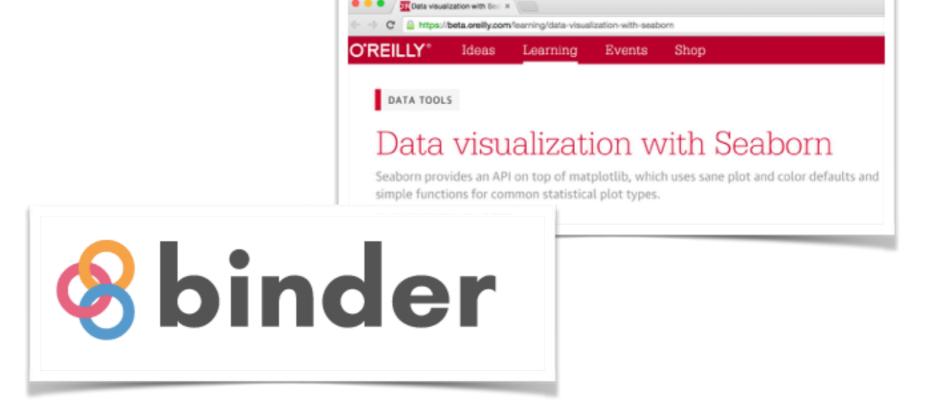


The Future

Publication

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

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





The Future

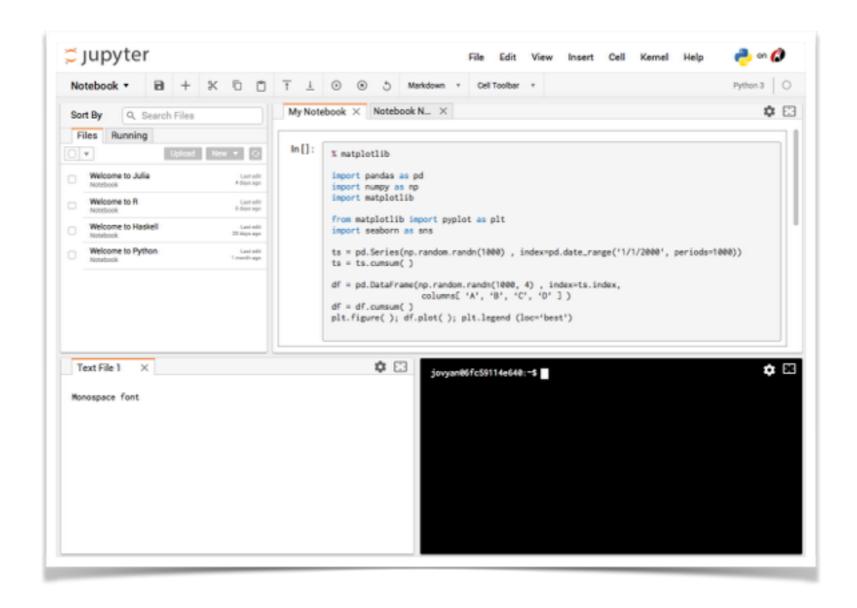
Where is the current core going? (currently on master)



Demo



Each pannel is an independant component

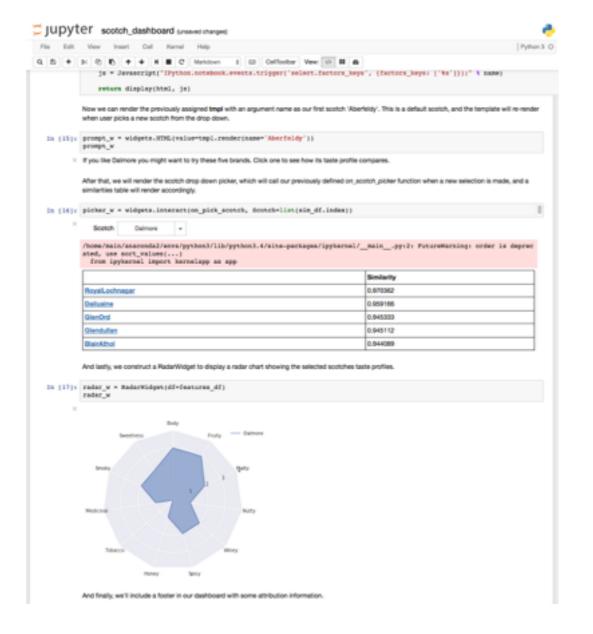


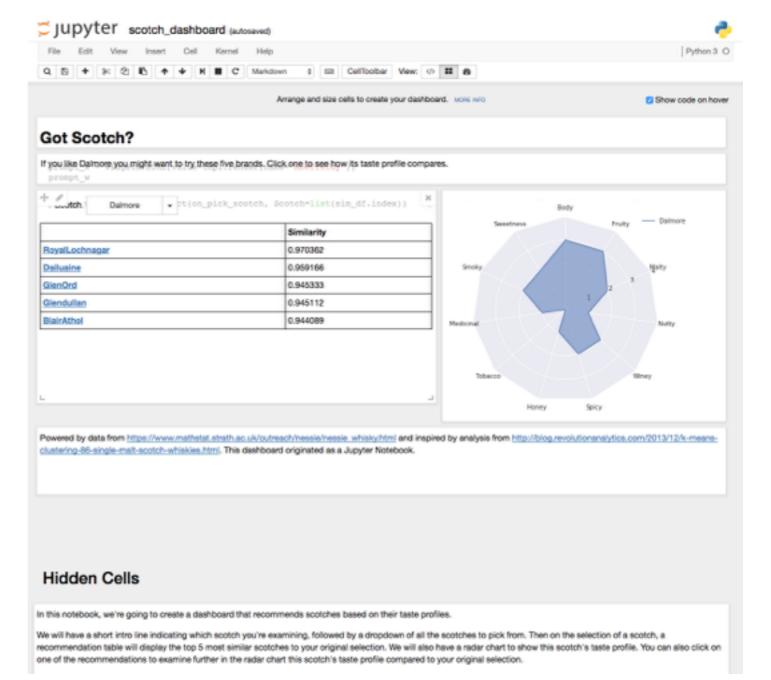


Dasbæard











Got Scotch?

Aberfeldy

Scotch

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

	Similarity
BlairAthol	0.975610
Benromach	0.975610
Benrinnes	0.965345
Scapa	0.963072
Auchroisk	0.963036



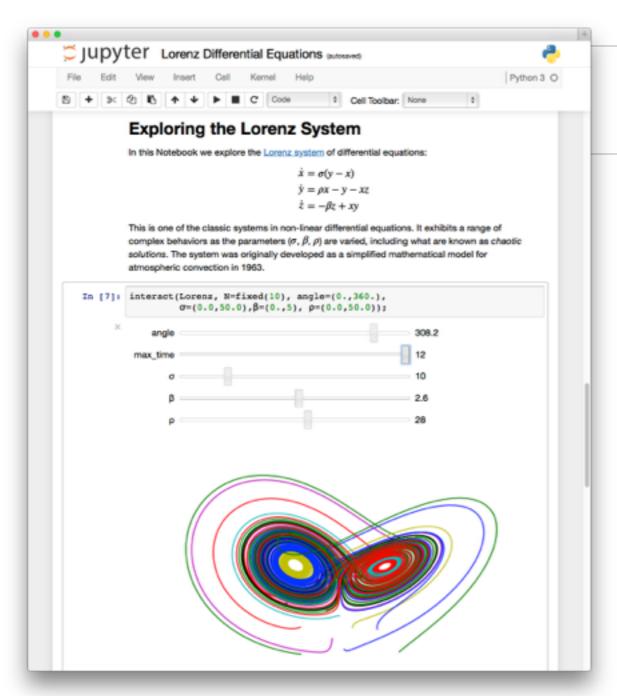
Powered by data from https://www.mathstat.strath.ac.uk/outreach/nessie/nessie whisky.html and inspired by analysis from http://blog.revolutionanalytics.com/2013/12/k-meansclustering-86-single-malt-scotch-whiskies.html. This dashboard originated as a Jupyter Notebook.



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
```







SIMONS FOUNDATION









A few new people

- CalPoly:
 - 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...



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]

