Introduction to Jupyter

Presented by Ben Winjum
IDRE
April 17, 2019

Outline

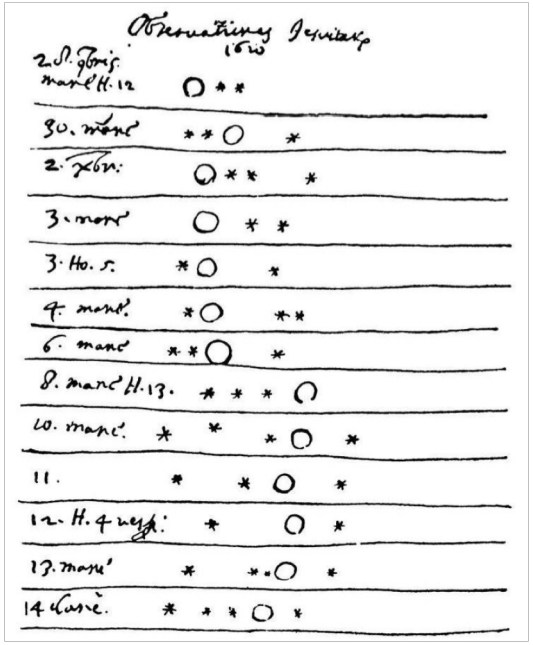
- What is Jupyter?
- Interactive – notebook basics
- A slightly deeper view of notebooks and the growing use of Jupyter
- Interactive – making a notebook from scratch
  - Including interactive elements
- The research and education ecosystem the Jupyter enables
- Interactive – JupyterLab

http://jupyter.org/

Julia, **Pyt**hon, **R**

and now over 100 other languages

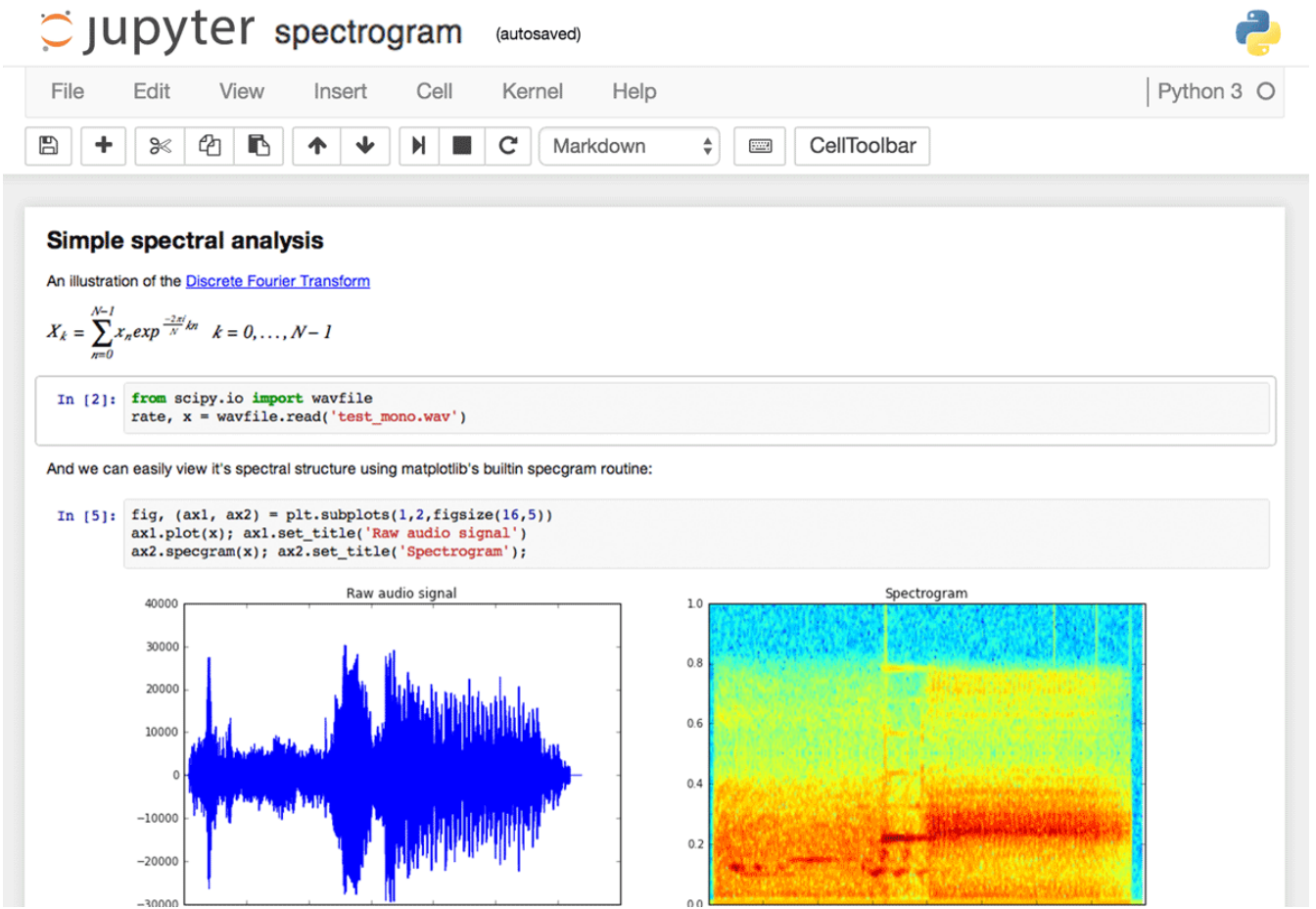
Supported kernels: https://github.com/jupyter/jupyter/wiki/Jupyter-kernels



A central component of the Jupyter project is the Jupyter Notebook

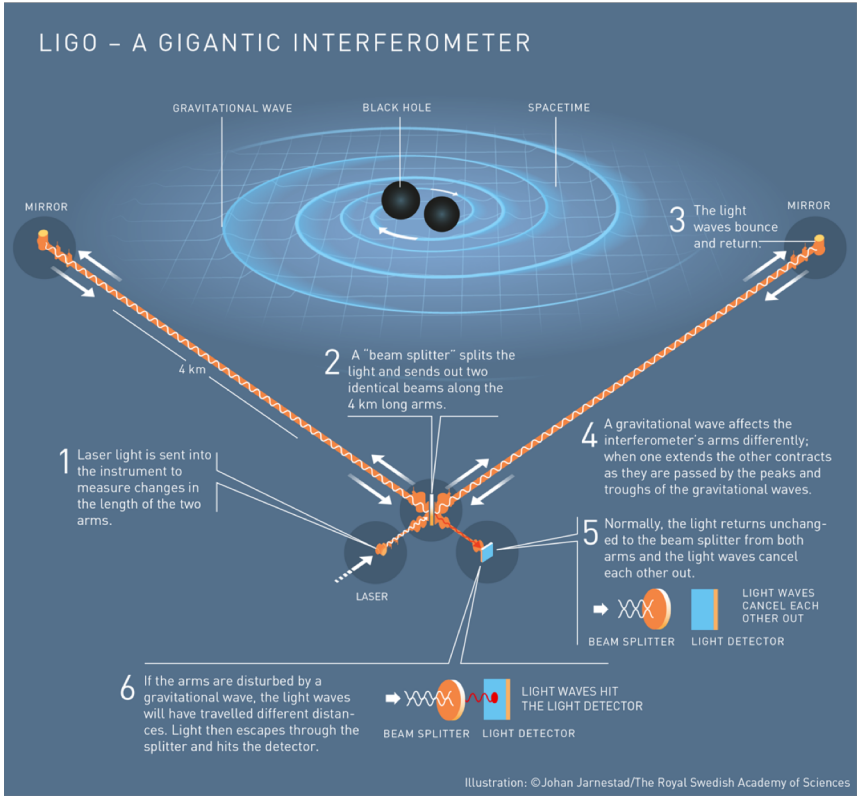
The Jupyter Notebook
an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text.

Uses include:
data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.



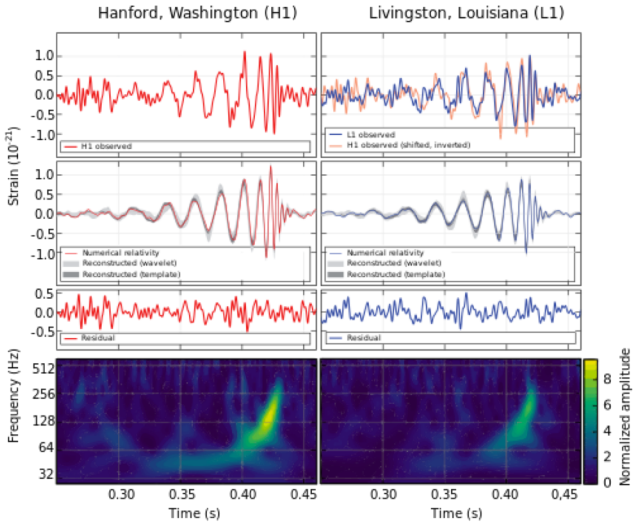
http://jupyter.org/

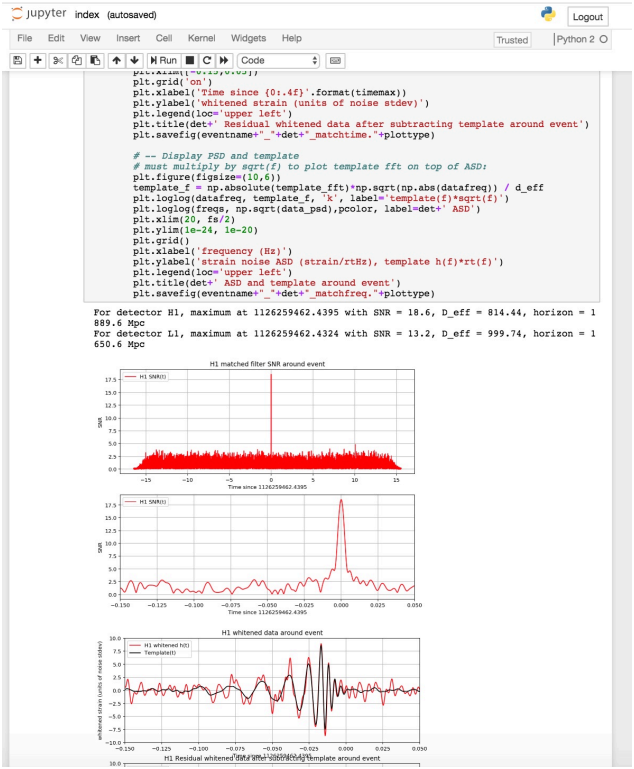
Jupyter Exploration of 2017 Physics Nobel Work



http://physics.aps.org/featured-article-pdf/10.1103/PhysRevLett.116.061102
https://www.gw-openscience.org/tutorials/

LIGO & Open Science

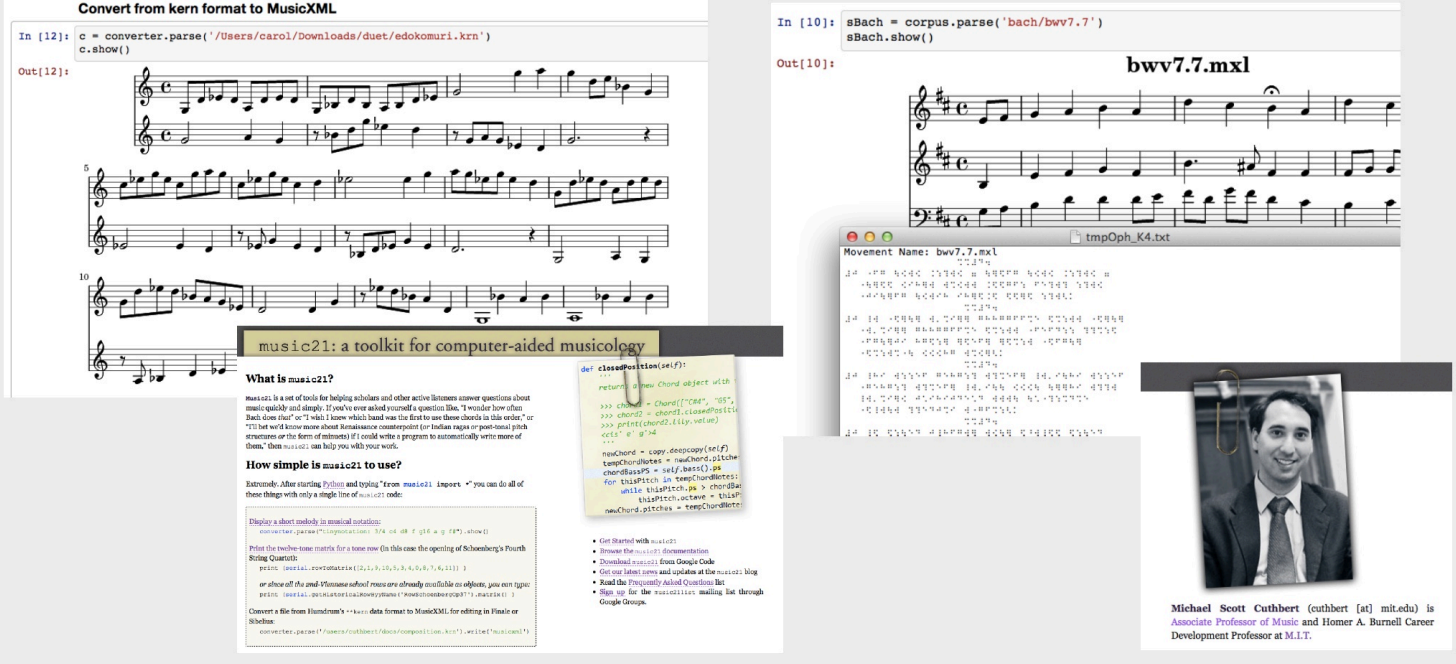




http://physics.aps.org/featured-article-pdf/10.1103/PhysRevLett.116.061102

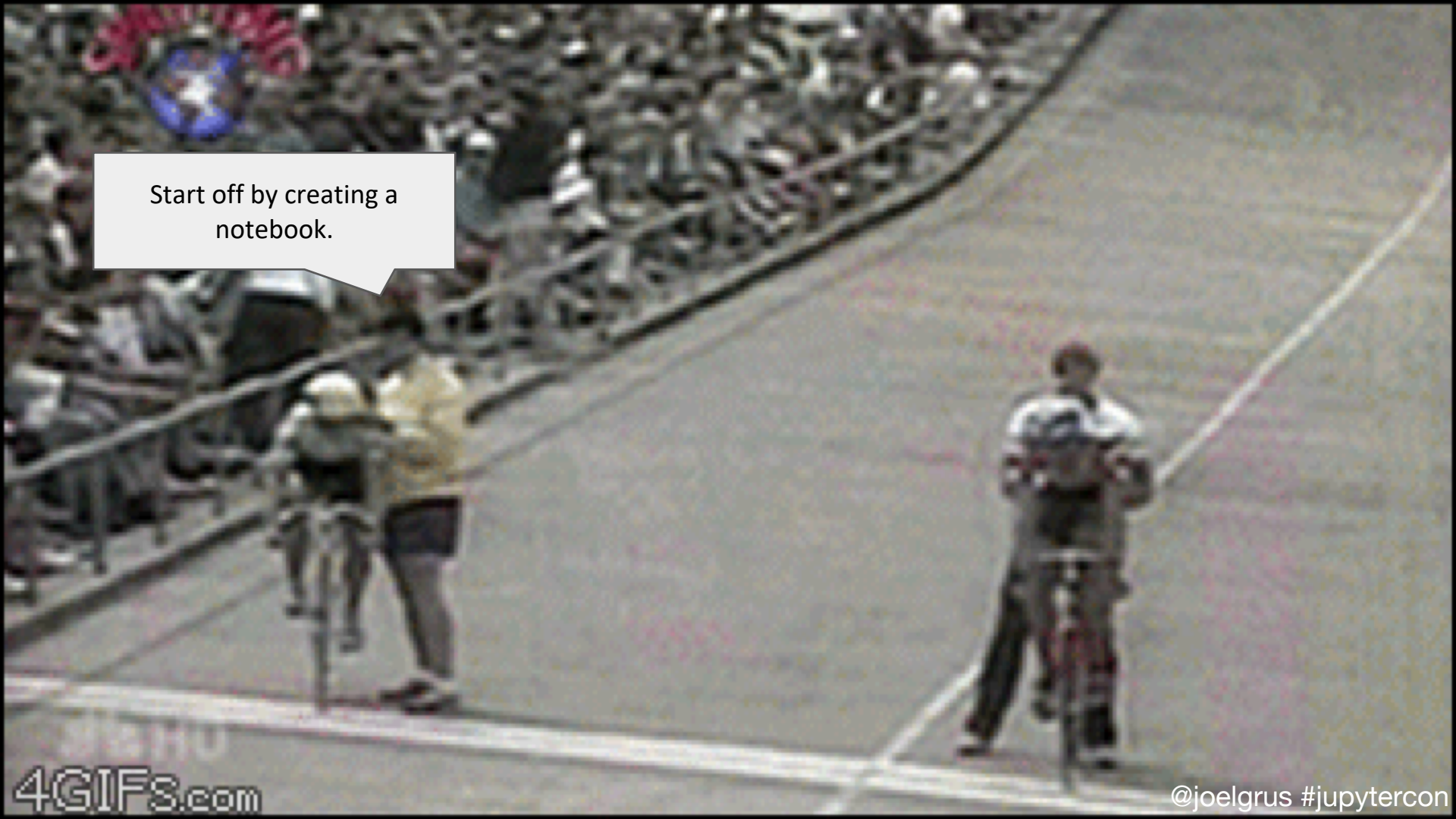
https://www.gw-openscience.org/tutorials/

Jupyter notebooks engage learners



From C. Willing, JupyterCon 2017

So …. How can you build amazing notebooks?


Start off by creating a notebook.
4GIFs.com
@joelgrus #jupytercon

Before we can move mountains, it helps to learn how our tools work

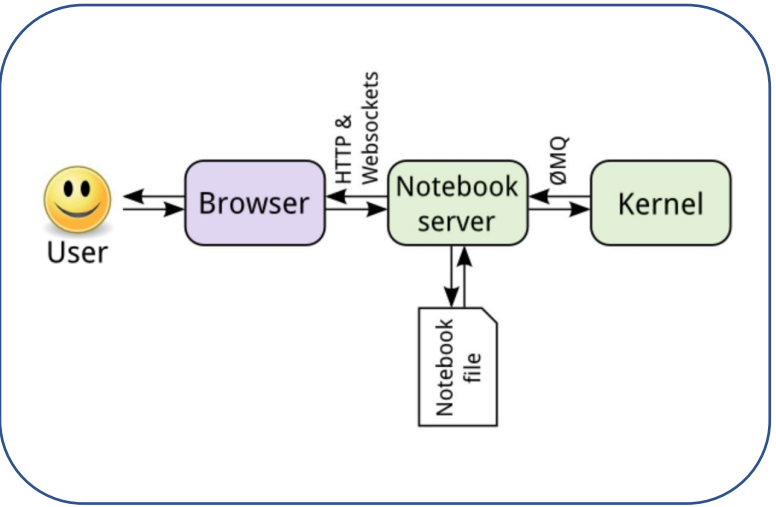
Notebook basics

go to https://jupyter.idre.ucla.edu

```
$ cd work

$ git clone git://github.com/benjum/idre-intro-to-jupyter.git
```

The Jupyter Notebook



Jupyter is comprised of several components

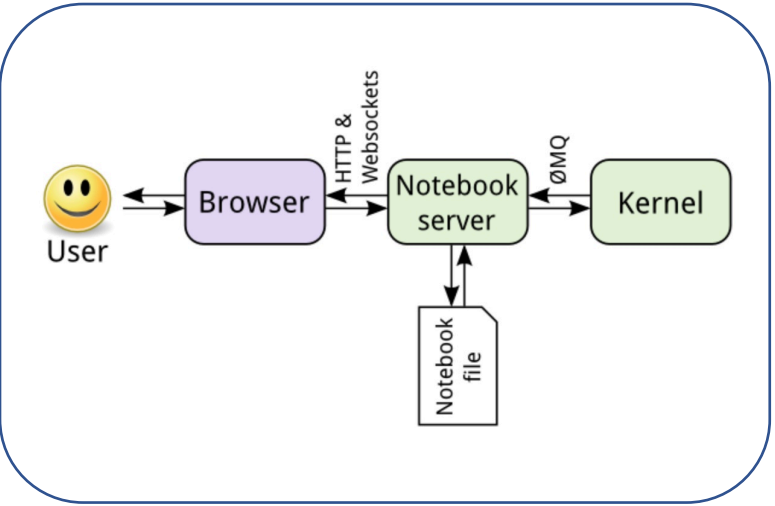
Front-end:

1. **Web Application**: Browser-based tool for interactive development of notebook documents

2. **Notebook Document**: A representation of all content visible in the web application, including inputs and outputs of the computations, explanatory text, mathematics, images, and rich media representations of objects. These documents are internally JSON files and are saved with the .ipynb extension. Since JSON is a plain text format, they can be version-controlled and shared with colleagues.

Image: https://jupyter.readthedocs.io/en/latest/architecture/how_jupyter_ipython_work.html

The Jupyter Notebook



Jupyter is comprised of several components

Back-end:

1. **Kernel**: A separate process responsible for running user code. Jupyter is capable of interfacing with many programming languages.

2. **Notebook Server**: Communicates with kernel and routes the programming language to the web browser.

Image: https://jupyter.readthedocs.io/en/latest/architecture/how_jupyter_ipython_work.html

Different ways to view notebooks

- Lab notebook
  - Single author… and meant primarily for single viewer
  - Can be split into parts before they get too long
  - Can be split into different topics
  - Not meant to be anything other than place for experimentation and development
- Deliverable report
  - Single- or team-authored
  - Meant to share
  - Fully polished
  - Store the final analysis and outputs
- Interactive playground
  - Carefully crafted educational narratives to invite interaction
  - Interleaving regular notebook cells with test cells

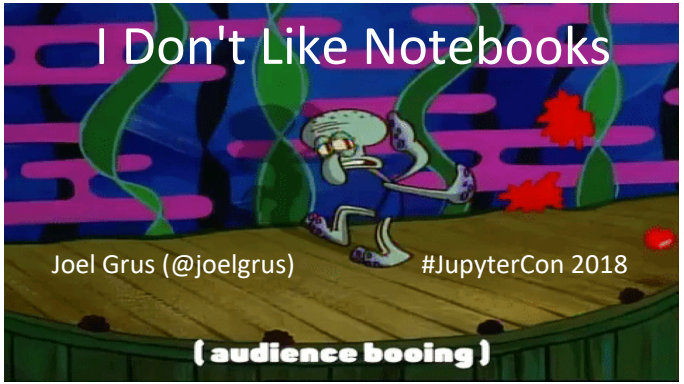
Not all uses of Jupyter are for clearly documented research and education

A. Rule, A. Tabard, J. Hollan. *Exploration and Explanation in Computational Notebooks.* ACM CHI Conference on Human Factors in Computing Systems, Apr 2018, Montréal, Canada.

- In a scan of > 1 million notebooks on GitHub, 1/4 had no explanatory text
- In analyzing 200 academic notebooks, < 40% discussed reasoning or explained results
- In interviews with 15 academic data analysts, most considered computational notebooks personal, exploratory, and messy

https://hal.archives-ouvertes.fr/hal-01676633/document

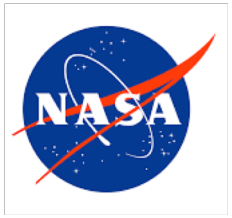
“Why not to use Jupyter”



- Hidden state and out-of-order execution
- Notebooks are difficult for beginners
- Notebooks encourage bad habits
- Notebooks discourage modularity and testing
- Jupyter’s autocomplete, linting, and way of looking up the help are awkward
- Notebooks encourage bad processes
- Notebooks hinder reproducible + extensible science
- Notebooks make it hard to copy and paste into Slack/Github issues
- Errors will always halt execution
- Notebooks make it easy to teach poorly
- Notebooks make it hard to teach well

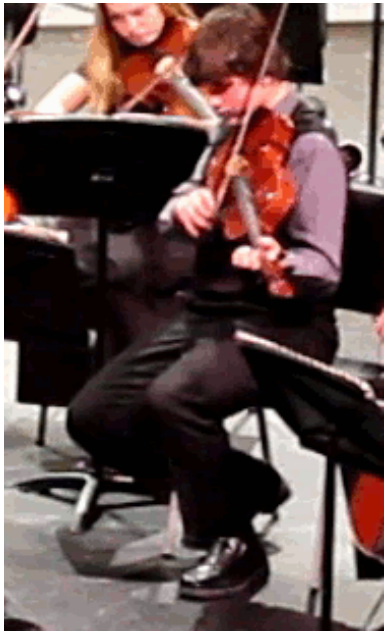
http://blog.revolutionanalytics.com/2018/09/notebooks-literate-programming.html

Nevertheless, Jupyter tools are being used by some major companies, and they are finding it in their best interest to contribute to open source tools related to Jupyter



So… is Jupyter a good environment for developing code?

So… is Jupyter a good environment for developing code?



YES!

But… like any beautiful instrument, it helps to develop technique

… now let’s practice making one

Atlantic Monthly
April 5, 2018

The Scientific Paper Is Obsolete

Here's what's next.

PNAS / Richard Goerg / Getty / The Atlantic

Novel uses of Jupyter: Digital Publication (takes a long time for adoption to actually occur)

The "Paper" of the Future

Alyssa Goodman (Harvard University)
Josh Peek (Space Telescope Science Institute)
Alberto Accomazzi (Harvard-Smithsonian Center for Astrophysics (CFA))
Chris Beaumont (Harvard-Smithsonian Center for Astrophysics (CFA))
Christine L. Borgman (UCLA - University of California, Los Angeles)
Hope How-Huan Chen (Harvard University)
Merce Crosas (Harvard University)
Christopher Erdmann (North Carolina State University)
August Muench
Alberto Pepe (Authorea Team)
Curtis Wong (Microsoft)
show less

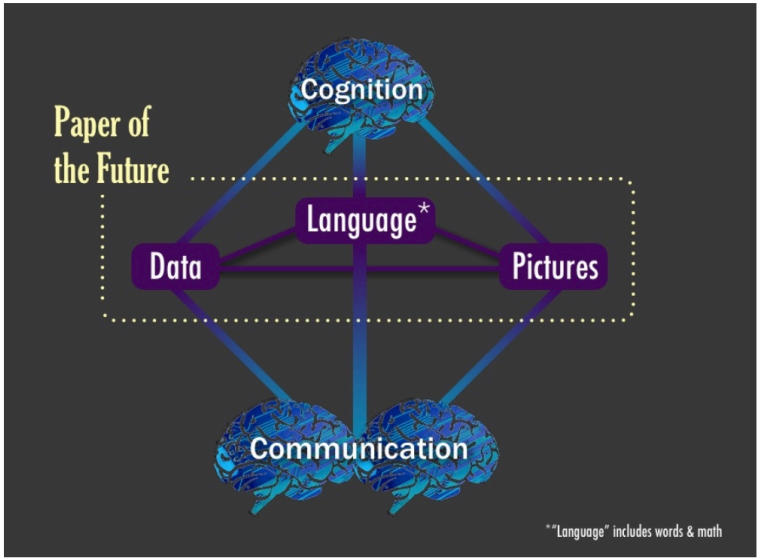


Fig. 1

*The Paper of the Future should include seamless linkages amongst **data**, **pictures**, and **language**, where "language" includes both words and math. When an individual attempts to understand each of these kinds of information, different cognitive functions are utilized: communication is inefficient if the channel is restricted primarily to language, without easy interconnection to data and pictures.*

https://www.authorea.com/users/23/articles/8762-the-paper-of-the-future?commit=d4033594de841d252b3220927b39de4314d26409

R. Connelly and V. Gayle
An investigation of Social Class Inequalities in General Cognitive Ability in Two British Birth Cohorts

"***An innovative aspect of this work is that we go beyond providing the usual supplementary material and make the complete workflow openly available. We take the trailblazing step of rendering the complete workflow accessible and reproducible within a Jupyter notebook*** *…. Jupyter notebooks have been used in Nobel Prize winning high-profile big science applications but are rarely used in Sociology.*"

https://github.com/RoxanneConnelly/Social-Class-Inequalities-in-General-Cognitive-Ability-in-Two-British-Birth-Cohorts

Jupyter and Education

Educational Examples

- Interactive notes
  - 12 Steps to Bioinformatics, Lorena Barba
- Online textbooks
  - An Introduction to Applied Bioinformatics, Greg Caporaso
- Notebook projects (using research software for education)
  - JupyterPIC for UCLA EE/Phys M185

Jupyter and HPC

https://ccit.clemson.edu/research/jupyter/

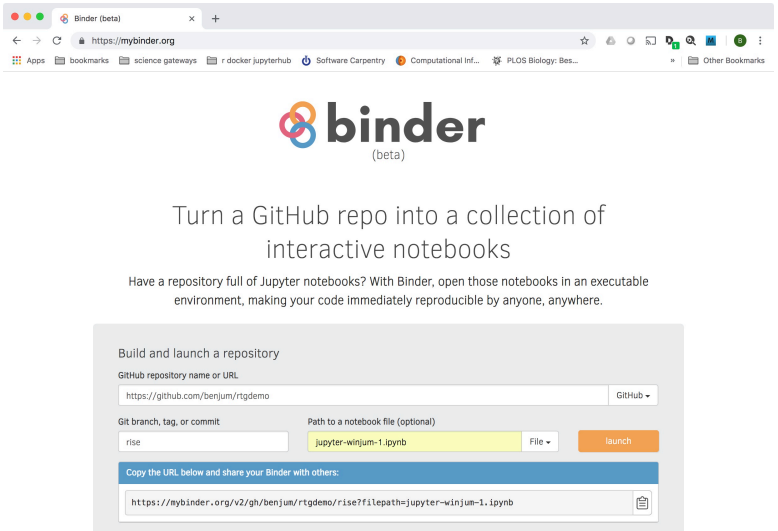
Connecting Notebooks with HPC Clusters

- University HPC Clusters
  - Like UCLA's Hoffman2
  - https://www.hoffman2.idre.ucla.edu/access/jupyter-notebook/

- National Supercomputing Centers
  - Like DOE's NERSC (National Energy Research Scientific Computing Center)
  - http://www.nersc.gov/users/data-analytics/data-analytics-2/jupyter-and-rstudio/

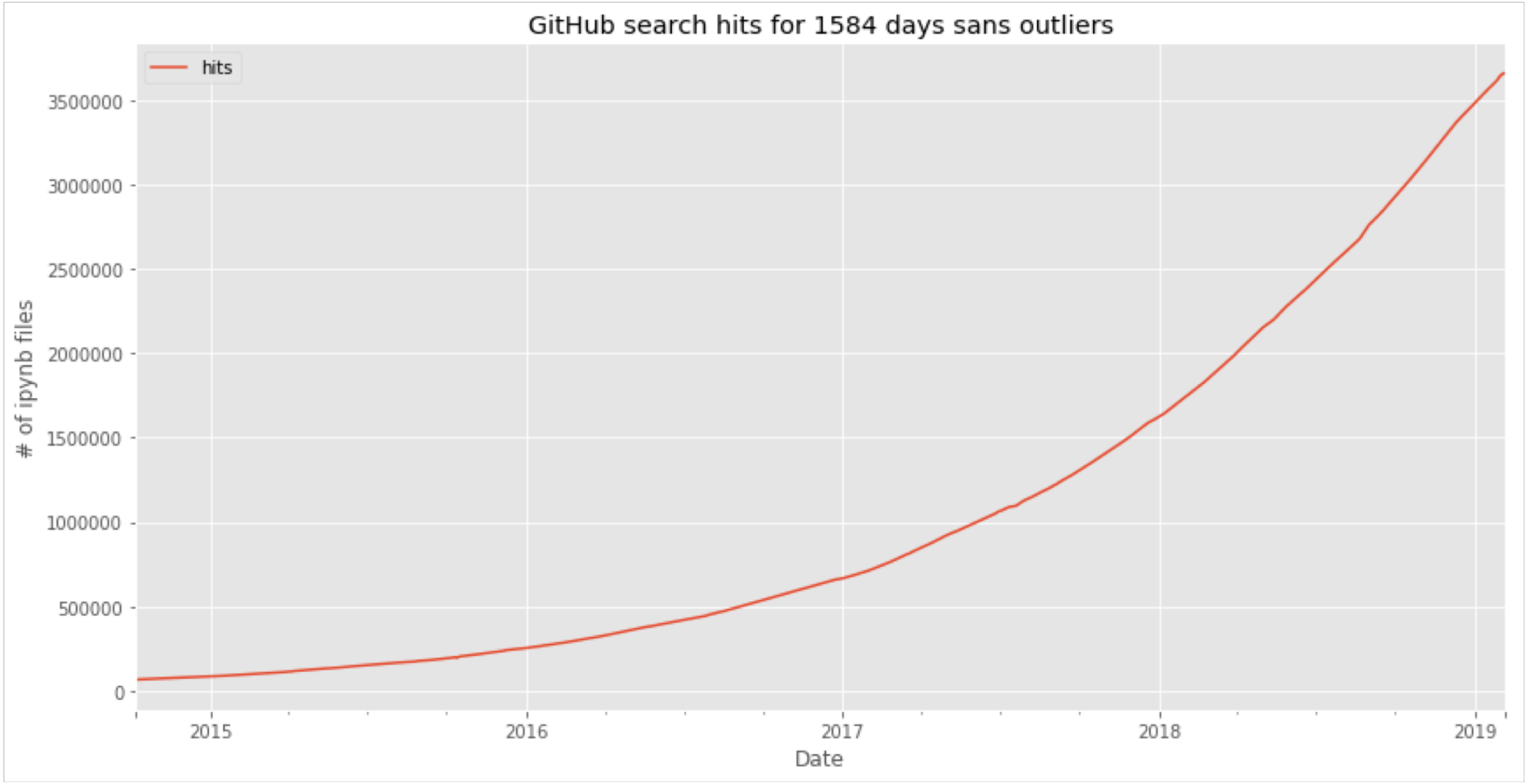
How to share

- Use the notebook menu
- jupyter nbconvert (converts notebooks to html, pdf, etc)
- Binder
- nbviewer.jupyter.org



https://github.com/jupyter/nbconvert

Estimate of public Jupyter notebooks on GitHub



http://nbviewer.jupyter.org/github/parente/nbestimate/blob/master/estimate.ipynb

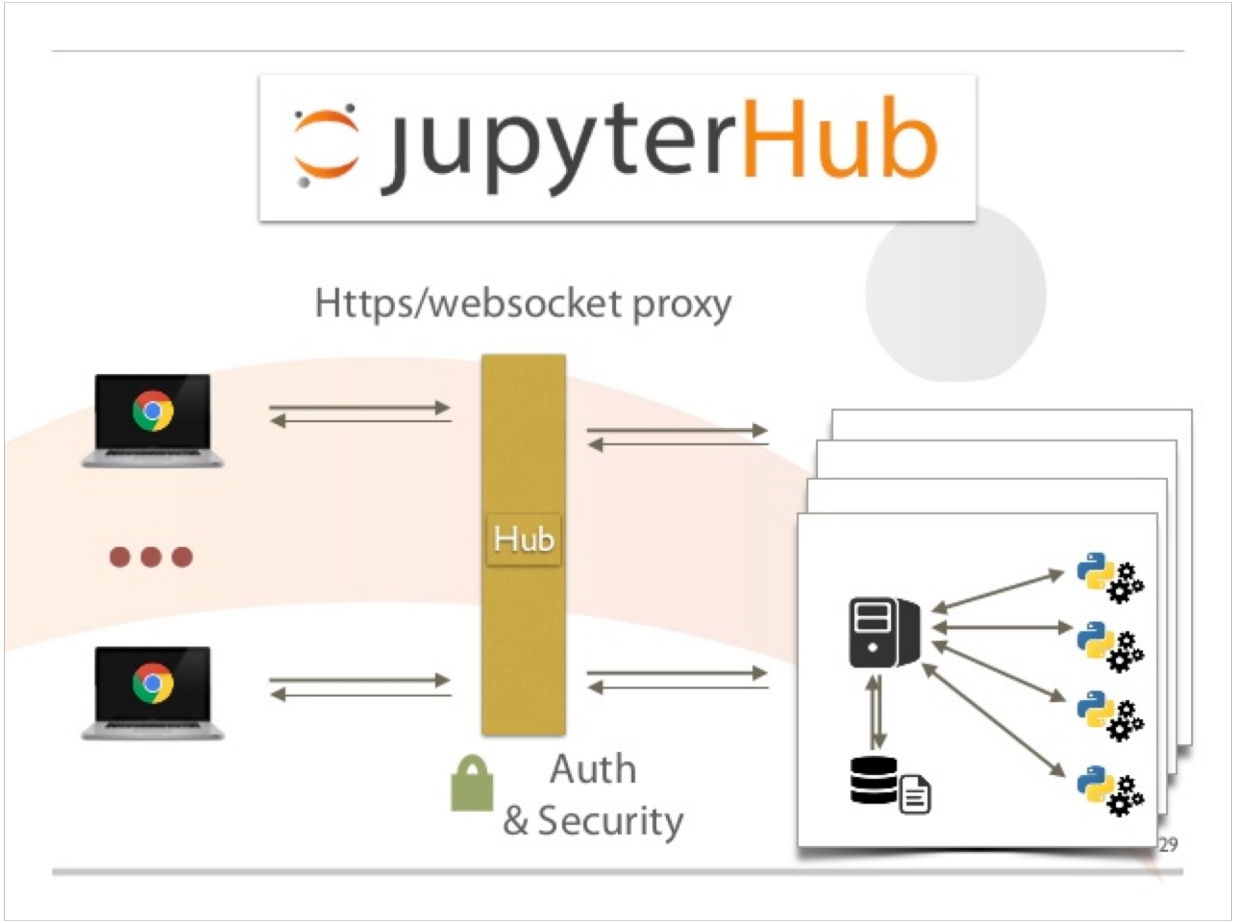


Image: https://jupyterhub.readthedocs.io/en/stable/

Next Generation: JupyterLab

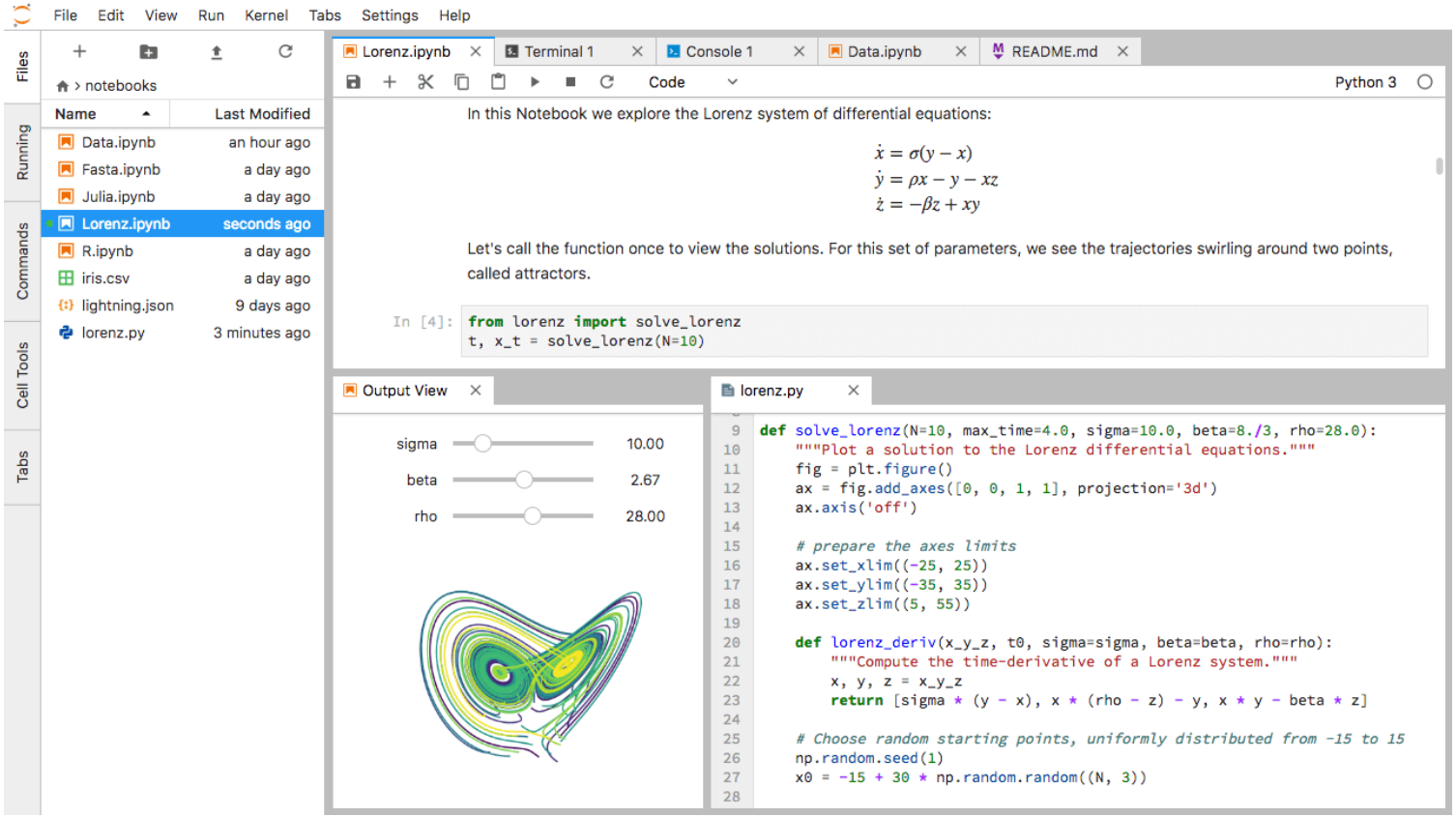


Image: https://jupyterlab.readthedocs.io/en/stable/

Notebooks: in JupyterLab

Final item:
please complete the brief
post-course survey

https://bit.ly/2VaYkVT