

Introduction to Jupyter

Presented by Ben Winjum

IDRE

July 7, 2020

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



Project Jupyter exists to develop open-source software, open-standards, and services for interactive computing across dozens of programming languages.



Julia, Python, R
and now over 100 other languages

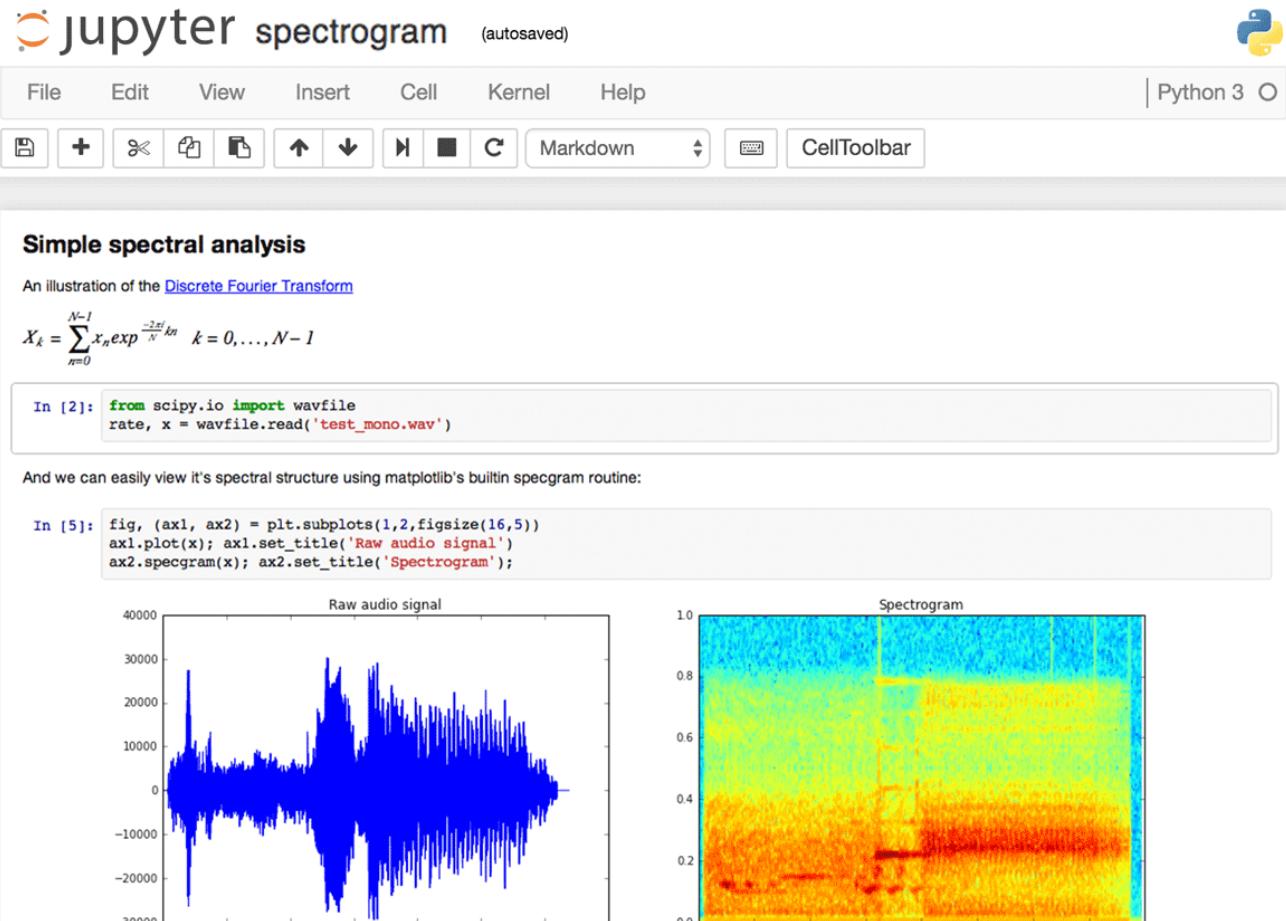


A central component of the Jupyter project
is the Jupyter Notebook

Observations February 16th			
2. S. pris. mar 6 H. 12	O **		
3. manc'	** O	*	
2. x6n:	O ***	*	
3. manc'	O *	*	
3. Ho. s.	* O	*	
4. manc'	* O	**	
6. manc'	** O	*	
8. manc H. 13.	* * * O		
10. manc'	* * * O	*	
11.	* * O	*	
12. H. 4 way:	* O	*	
13. manc'	* * * O	*	
14. clare.	* * * O	*	

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.



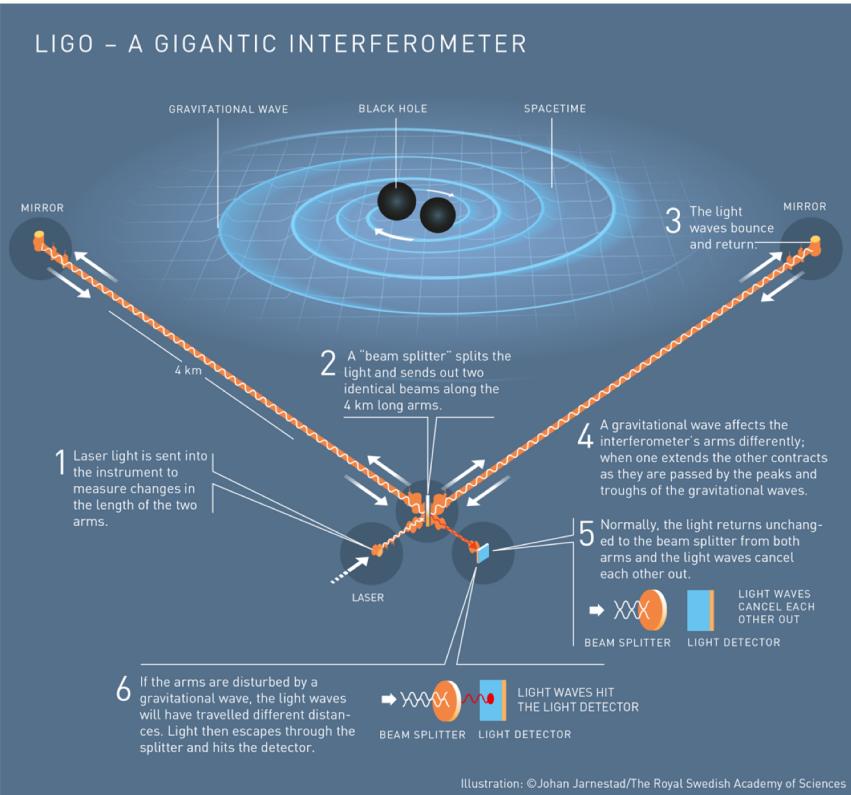
Let's look at an example: learning how to plot in Python with Matplotlib

The screenshot shows a Jupyter Notebook Viewer window titled "Jupyter Notebook Viewer". The URL in the address bar is nbviewer.jupyter.org/github/jrjohansson/scientific-python-lectures/blob/master/Lecture-4-Matplotlib.ipynb. The page title is "matplotlib - 2D and 3D plotting in Python" by J.R. Johansson (jrjohansson@gmail.com). It includes links to the latest IPython notebook version and the author's GitHub page. A code cell in the notebook is shown:

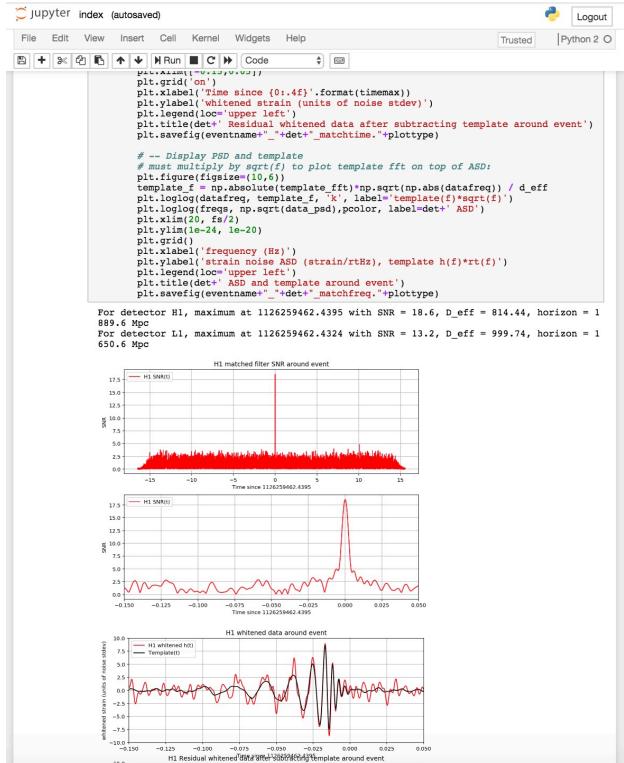
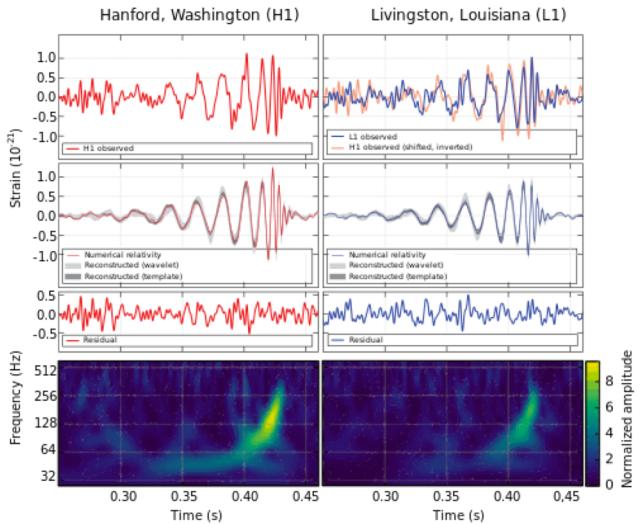
```
In [1]: # This line configures matplotlib to show figures embedded in the notebook,
# instead of opening a new window for each figure. More about that later.
# If you are using an old version of IPython, try using '%pylab inline' instead.
%matplotlib inline
```

<https://nbviewer.jupyter.org/github/jrjohansson/scientific-python-lectures/blob/master/Lecture-4-Matplotlib.ipynb>

Jupyter Exploration of 2017 Physics Nobel Work



LIGO & Open Science



Jupyter notebooks engage learners

Convert from kern format to MusicXML

```
In [12]: c = converter.parse('/Users/carol/Downloads/duet.edokomuzi.krn')
c.show()
```

Out[12]:

5

10

music21: a toolkit for computer-aided musicology

```
def closedPosition(self):
    return self._newChordObject with:
        >>> chord = Chord("C4", "G5")
        >>> chord2 = chord.closedPosition()
        >>> print(chord2.lyric)
        'C4' + g4

newChord = copy.deepcopy(self)
tempChordObject = newChord._pitch
chordLyrics = self._lyrics
for thisPitch in tempChordObject:
    while thisPitch in chordLyrics:
        thisPitch = thisPitch + 1
    tempChordObject.append(thisPitch)
newChord._lyrics = tempChordObject
```

What is music21?

music21 is a toolkit for helping musicians and other creative humans answer questions about music easily and simply. It's a general purpose library. "I wonder how often Bach does oct." or "I wonder which hand was the best to use there in this order," or "I bet we know more about Renaissance counterpoint (or Indian raga) or post-tonal pitch structures than anyone else in the world." It has a program to automatically write more of them. Then you can help you with your work.

How simple is music21 to use?

Extremely: After starting Python and typing "from music21 import *" you can do all of these things with only a single line of music21 code:

```
converter.parse('kernFormat').show() # or display('kernFormat')
```

Display a short melody in musical notation:

```
converter.parse('kernFormat').show()
```

Print the twelve-tone matrix for a tone row (In this case the opening of Schoenberg's Fourth String Quartet):

```
print(music21.sets.RhythmSet([(1,1),(1,1),(1,1),(1,1),(1,1),(1,1),(1,1),(1,1),(1,1),(1,1),(1,1),(1,1)])
```

or since all of the *Music21* school rooms are really *Mutable* objects, you can type

```
print(music21.sets.RhythmSet([1,2,3,4,5,6,7,8,9,10,11,12])._lyrics)
```

Convert a file from Humdrum's "kern" data format to MusicXML, for editing in Muse or Sibelius:

```
converter.parse('humdrumFormat').write('musicxmlFormat')
```



Michael Scott Cuthbert ([cuthbert \[at\] mit.edu](mailto:cuthbert@mit.edu)) is Associate Professor of Music and Homer A. Burnell Career Development Professor at M.I.T.

1

From C. Willing, JupyterCon 2017

So How can you build amazing notebooks?



Before we try to build amazing things, let's figure out how the tools work

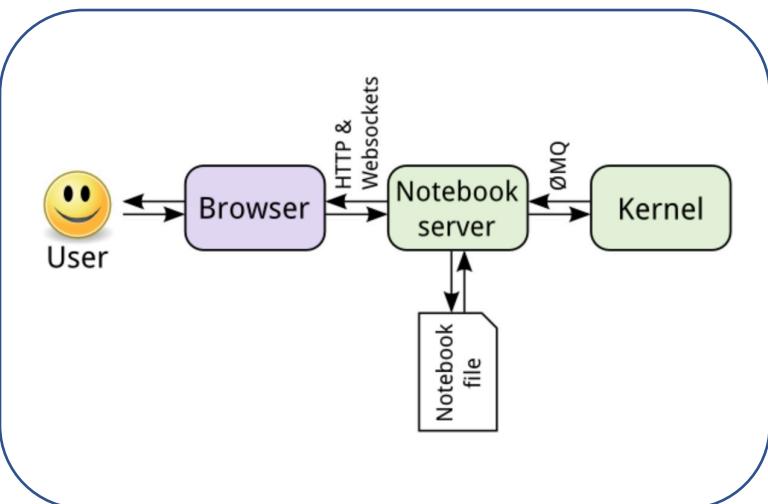
Go to

<https://github.com/benjum/idre-intro-to-jupyter>

The Jupyter Notebook

Jupyter is comprised of several components

Front-end:



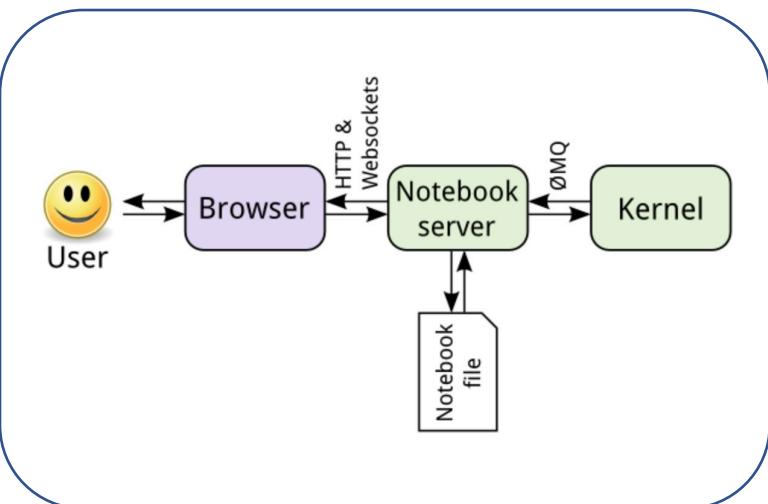
1. **Web Application:** Browser-based tool for interactive development of notebook documents
2. **Notebook Document:** A representation of all content visible in the web application, including inputs and outputs of the computations, explanatory text, mathematics, images, and rich media representations 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.

The Jupyter Notebook

Jupyter is comprised of several components

Back-end:

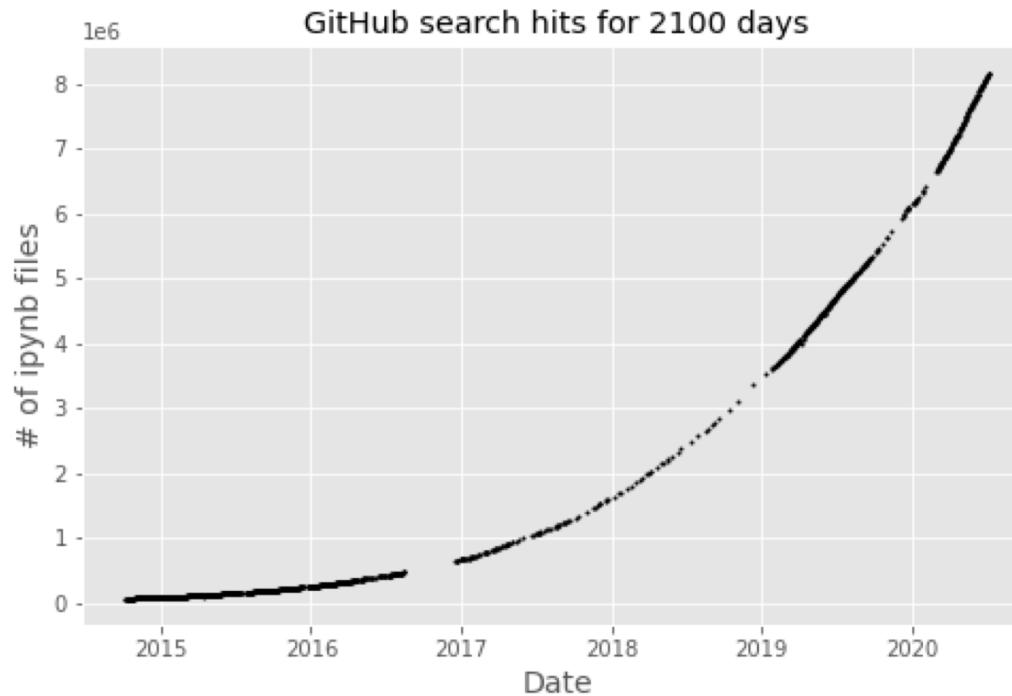
1. **Kernel:** A separate process responsible for running user code. Jupyter is capable of interfacing with many programming languages.
2. **Notebook Server:** Communicates with kernel and routes the programming language to the web browser.



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

Estimate of public Jupyter notebooks on GitHub



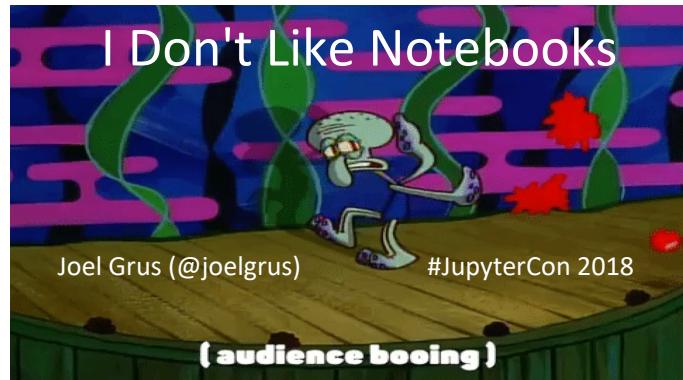
Not all uses of Jupyter are for clearly documented research and education

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

... and not all uses of Jupyter are completely clear either

- Methods for interacting with code in notebooks are different than for conventional software development
 - Hidden state
 - Out-of-order execution
 - Not quite as convenient to copy/paste large chunks of code
 - Added complexity and abstraction for integration with narrative layers



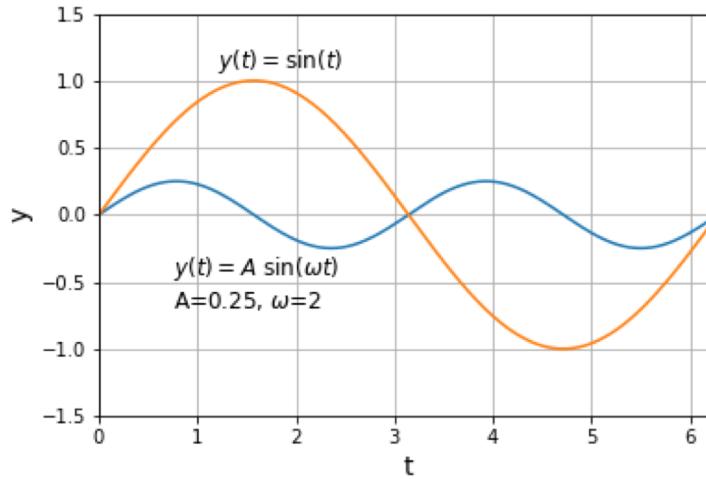
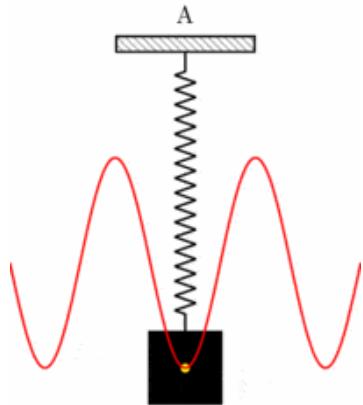
Taken in part from Joel Grus humorous talk "I Don't Like Notebooks"
<http://blog.revolutionanalytics.com/2018/09/notebooks-literate-programming.html>

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



Jupyter is a fantastic tool for education and research

Demo time: Let's make a notebook about a sine wave



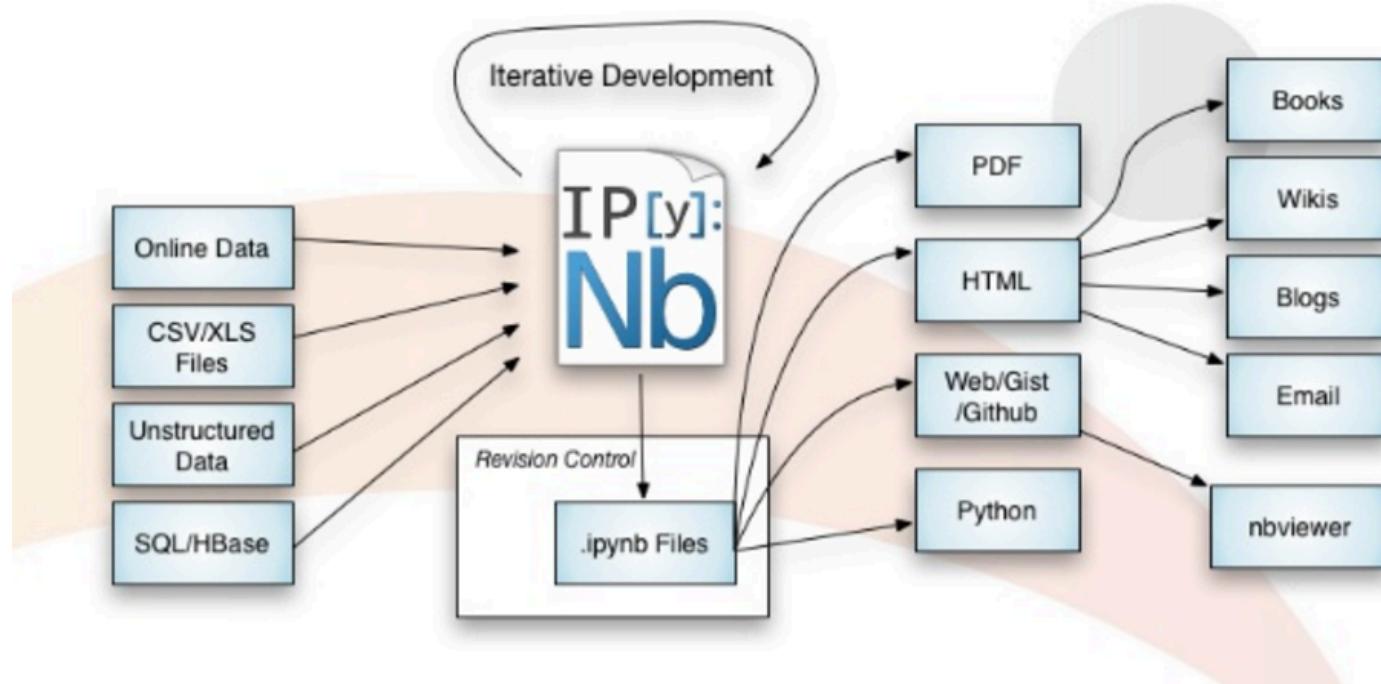
Jupyter and Education



<https://data.berkeley.edu/news/coursefuture>

Educational Examples

- Interactive notes
 - [12 Steps to Navier-Stokes](#), Lorena Barba
- Online textbooks
 - [An Introduction to Applied Bioinformatics](#), Greg Caporaso
- Using research software for education
 - [JupyterPIC](#) for plasma physics classes at UCLA



Jupyter and Journalism

News from California, the natio x +

latimes.com

bookmarks science gateways cesmii-presentation r docker jupyterhub Software Carpentry Computational Inf... PLOS Biology: Bes... Other

tions

Los Angeles Times

SUBSCRIBE NOW \$1 for 4 weeks LOG IN



As coronavirus cases surge, L.A. County death toll tops 3,500

Young adults who think 'they are invincible' increasingly infected by coronavirus, Newsom says

Newsom, coming under fire, ramps up enforcement and adds counties to watchlist

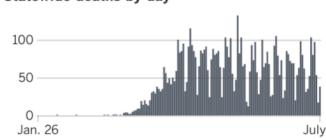
Jelly Belly and Yeezy got PPP loans. Which other California companies benefited?

Top medical officer for California prisons ousted amid worsening coronavirus outbreak

Coronavirus cases statewide »
As of July 6, 8:10 p.m. Pacific

confirmed	deaths
273,576	6,459

Statewide deaths by day



Jan. 26 July 5

California » L.A. County » Orange County »

We are moving into [Stage 3 of reopening the state](#). Lower-risk businesses can now reopen with social distancing guidelines.

FREE FOR ALL READERS >

<https://github.com/datadesk/california-coronavirus-data>

How to share

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



Turn a GitHub repo into a collection of
interactive notebooks

Have a repository full of Jupyter notebooks? With Binder, open those notebooks in an executable environment, making your code immediately reproducible by anyone, anywhere.

Build and launch a repository

GitHub repository name or URL
 GitHub ▾

Git branch, tag, or commit
 Path to a notebook file (optional) File ▾

Copy the URL below and share your Binder with others:

<https://github.com/jupyter/nbconvert>

nbgitpuller link generator: <https://jupyterhub.github.io/nbgitpuller/link>

Atlantic Monthly
April 5, 2018



Genomic analysis of elongated skulls and extensive female-biased immigration in early Medieval Bavaria

Krishna R. Veeramah^a, Andreas Rott^{b,1}, Melanie Groß^{c,1}, Lucy van Dorp^d, Saioa López^e, Karola Kirsanow^f, Christian Sell^c, Jens Blöcher^c, Daniel Wegmann^{f,g}, Vivian Link^{f,g}, Zuzana Hofmanová^{f,g}, Joris Peters^{b,h}, Bernd Trautmann^b, Anja Gairhosⁱ, Jochen Haberstroh^j, Bernd Päffgen^k, Garrett Hellenthal^d, Brigitte Haas-Gebhardⁱ, Michaela Harbeck^{b,2,3}, and Joachim Burger^{c,2,3}

^aDepartment of Ecology and Evolution, Stony Brook University, Stony Brook, NY 11794-5245; ^bState Collection for Anthropology and Palaeoanatomy, Bavarian Natural History Collections, 80333 Munich, Germany; ^cPalaeogenetics Group, Institute of Organismic and Molecular Evolution, Johannes Gutenberg University Mainz, 55099 Mainz, Germany; ^dUCL Genetics Institute, Department of Genetics, Evolution and Environment, University College London, WC1E 6BT London, United Kingdom; ^eCancer Institute, University College London, WC1E 6DD London, United Kingdom; ^fDepartment of Biology, University of Fribourg, 1700 Fribourg, Switzerland; ^gSwiss Institute of Bioinformatics, 1700 Fribourg, Switzerland; ^hArchaeoBioCenter and Institute for Palaeoanatomy, Domestication Research and the History of Veterinary Medicine, Ludwig Maximilian University, 80539 Munich, Germany; ⁱBavarian State Archaeological Collection, 80538 Munich, Germany; ^jBavarian State Department of Monuments and Sites, 80539 Munich, Germany; and ^kInstitute of Prehistoric and Prehistoric Archaeology, Ludwig Maximilian University, 80799 Munich, Germany

The Scientific Paper Is Obsolete

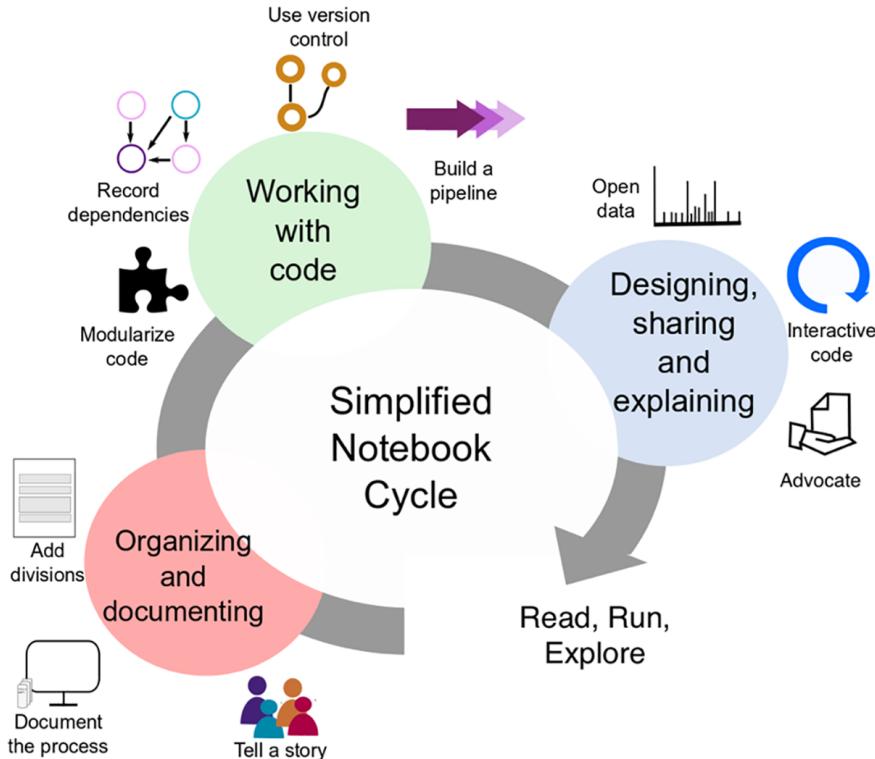
Here's what's next.

Modern European genetic structure demonstrates strong correlations with geography, while genetic analysis of prehistoric humans has indicated at least two major waves of immigration from outside the continent during periods of cultural change. However, population-level genome data that could shed light on the demographic processes occurring during the intervening periods have been absent. Therefore, we generated genomic data from 41 individuals dating mostly to the late 5th/early 6th century

to form in the 5th century AD, and that it emanated from a combination of the romanized local population of the border province of the former Roman Empire and immigrants from north of the Danube (2). While the Baiuvarii are less well known than some other contemporary groups, an interesting archaeological feature in Bavaria from this period is the presence of skeletons with artificially deformed or elongated skulls (Fig. 1*A*).

Artificial cranial deformation (ACD), which is only possible

POPULATION
BIOLOGY



Rule A, Birmingham A, Zuniga C, Altintas I, Huang S-C, Knight R, et al. (2019) *Ten simple rules for writing and sharing computational analyses in Jupyter Notebooks*. PLoS Comput Biol 15(7): e1007007.
<https://doi.org/10.1371/journal.pcbi.1007007>

Jupyter and HPC

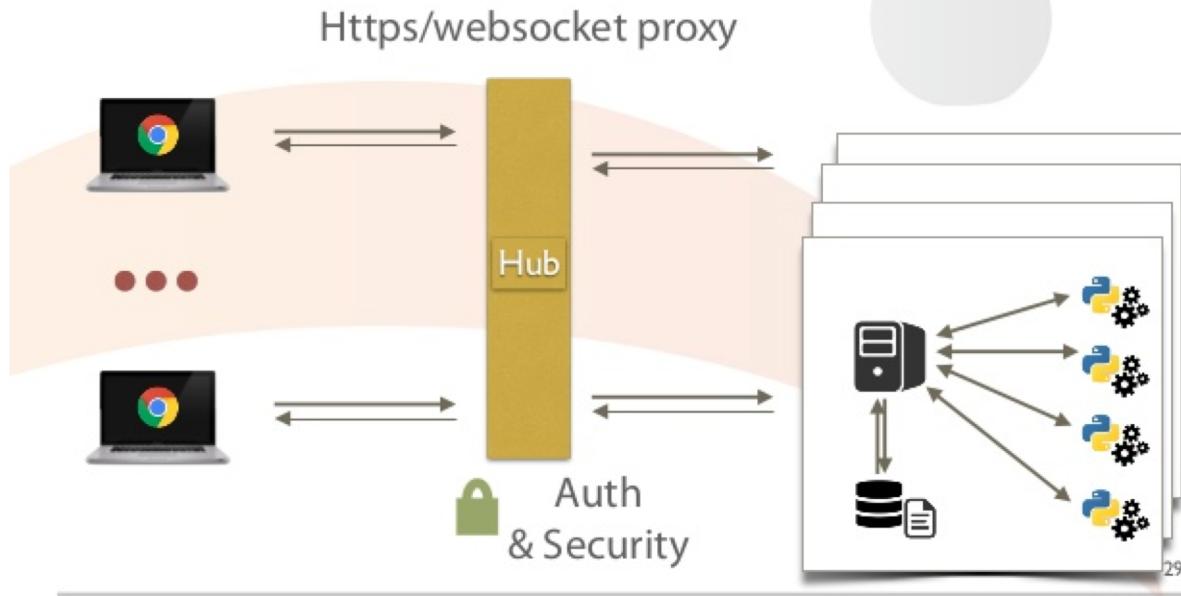


<https://ccit.clemson.edu/research/jupyter/>

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)
 - <https://docs.nersc.gov/services/jupyter/>

jupyterHub



Next Generation: JupyterLab

The screenshot shows the JupyterLab interface. On the left is a sidebar with tabs for Files, Running, Commands, Cell Tools, and Tabs. The Files tab is active, showing a list of notebooks: Data.ipynb (an hour ago), Fasta.ipynb (a day ago), Julia.ipynb (a day ago), Lorenz.ipynb (seconds ago), R.ipynb (a day ago), iris.csv (a day ago), lightning.json (9 days ago), and lorenz.py (3 minutes ago). The Running tab shows no running processes.

The main area has several tabs: Lorenz.ipynb (active), Terminal 1, Console 1, Data.ipynb, README.md. The Lorenz.ipynb tab contains text about the Lorenz system and its differential equations:

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

Below this, it says: "Let's call the function once to view the solutions. For this set of parameters, we see the trajectories swirling around two points, called attractors."

In [4]:

```
from lorenz import solve_lorenz
t, x_t = solve_lorenz(N=10)
```

The Output View tab shows three sliders for parameters: sigma (10.00), beta (2.67), and rho (28.00). Below the sliders is a 2D plot of the Lorenz attractor, which is a complex, fractal-like shape formed by the trajectories of the differential equations.

The code editor tab shows the contents of lorenz.py:

```
def solve_lorenz(N=10, max_time=4.0, sigma=10.0, beta=8./3, rho=28.0):
    """Plot a solution to the Lorenz differential equations."""
    fig = plt.figure()
    ax = fig.add_axes([0, 0, 1, 1], projection='3d')
    ax.axis('off')

    # prepare the axes limits
    ax.set_xlim((-25, 25))
    ax.set_ylim((-35, 35))
    ax.set_zlim(5, 55)

    def lorenz_deriv(x_y_z, t0, sigma=sigma, beta=beta, rho=rho):
        """Compute the time-derivative of a Lorenz system."""
        x, y, z = x_y_z
        return [sigma * (y - x), x * (rho - z) - y, x * y - beta * z]

    # Choose random starting points, uniformly distributed from -15 to 15
    np.random.seed(1)
    x0 = -15 + 30 * np.random(N, 3)
```

Notebooks: in JupyterLab

Final item:
please complete the brief
post-course survey

<https://bit.ly/2BFRPp6>

Any Future Questions:
Email – bwinjum@idre.ucla.edu