IP[y]:
IPython

Berkeley
Division of
Data Sciences

jupyter

Jupyter meets the Earth:
an open, collaborative approach for Earth data science

Fernando Pérez

Lindsey Heagy

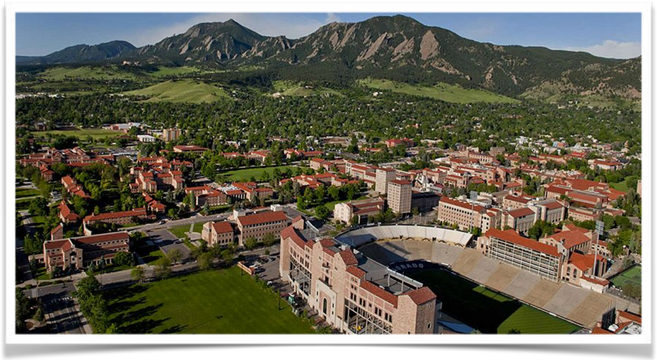
BIDS
BERKELEY INSTITUTE
FOR DATA SCIENCE

University of California, Berkeley
DEPARTMENT OF STATISTICS

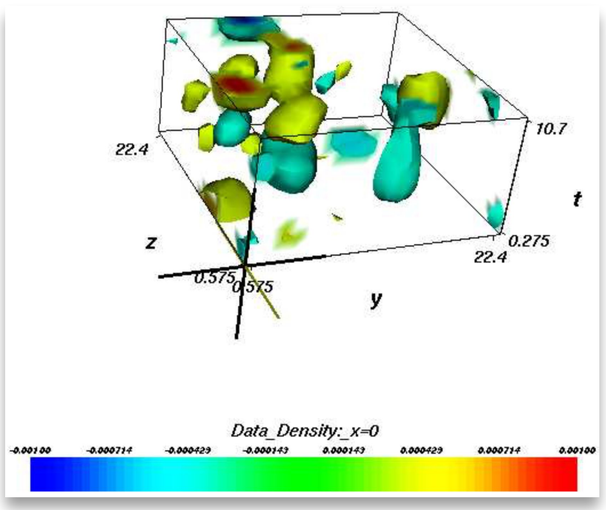
BERKELEY LAB

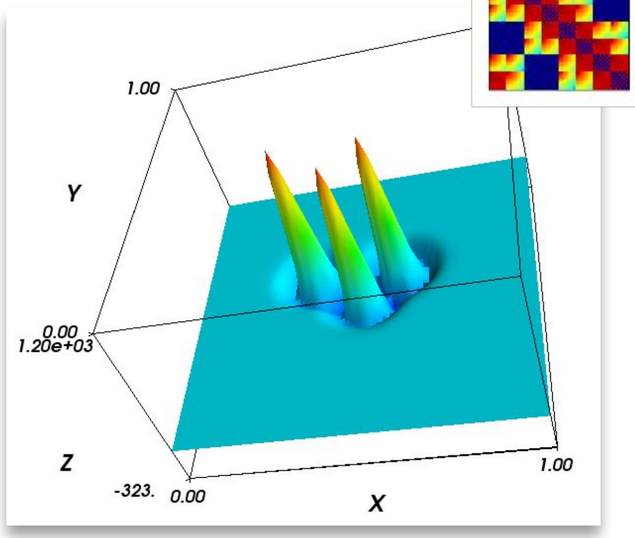
👋! a bit about me

Boulder

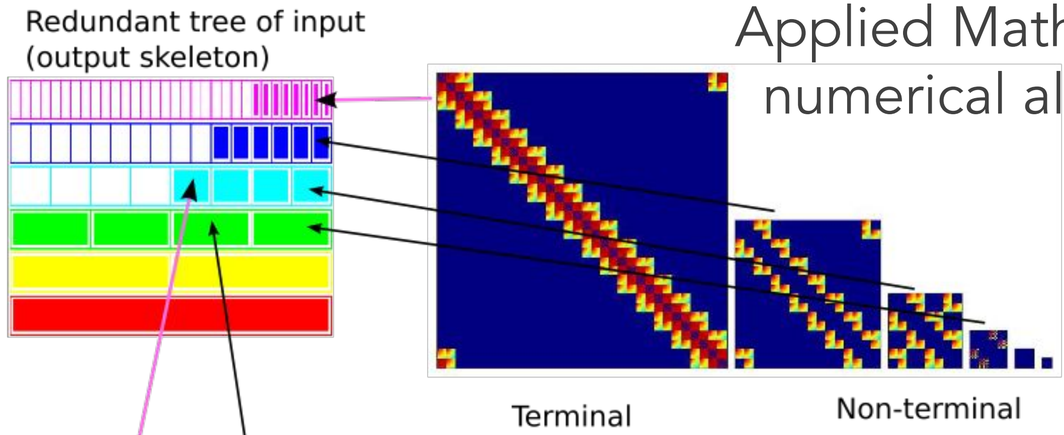


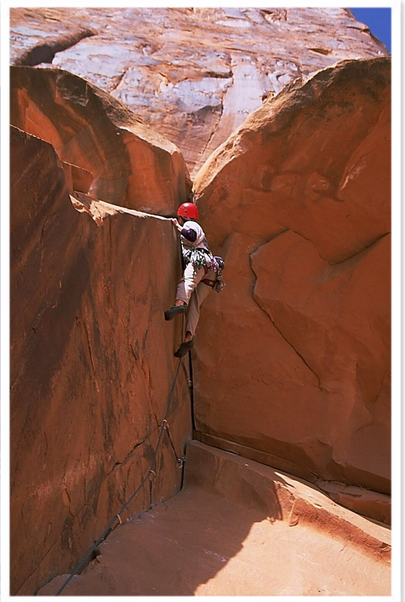
Physics PhD: Lattice QCD Simulations





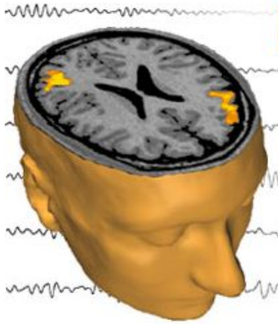
Applied Math Postdoc: numerical algorithms





BRAIN IMAGING CENTER

UNIVERSITY OF CALIFORNIA, BERKELEY



The Henry H. Wheeler, Jr. Brain Imaging Center (BIC) is one of four Technology Centers established under the auspices of the Helen Wills Neuroscience Institute. It is a campus-wide resource that supports advanced brain imaging technologies dedicated solely to basic brain research.

BERKELEY LAB

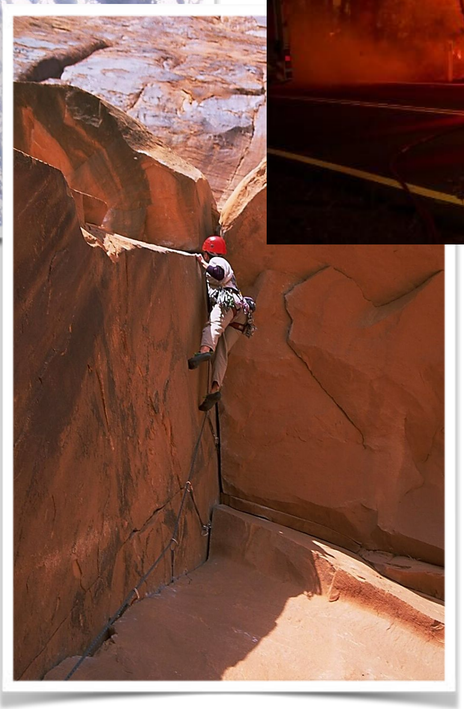
BIDS

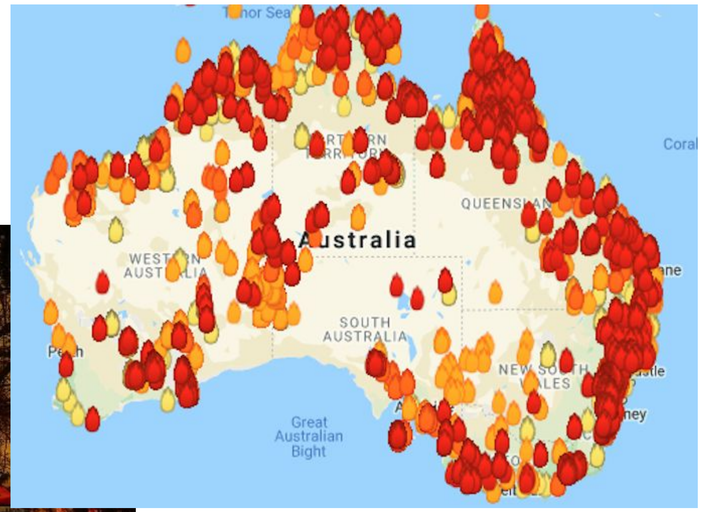
BERKELEY INSTITUTE FOR DATA SCIENCE

University of California, Berkeley

DEPARTMENT OF STATISTICS

Wait, why geoscience??





what drives progress in (geo)science?

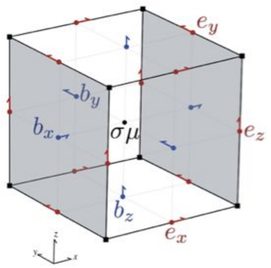
$$\nabla \times \mathbf{e} = -\frac{\partial \mathbf{b}}{\partial t}$$

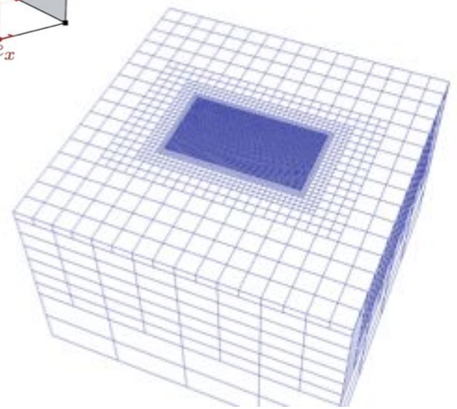
$$\nabla \times \mathbf{h} = \mathbf{j} + \frac{\partial \mathbf{d}}{\partial t}$$

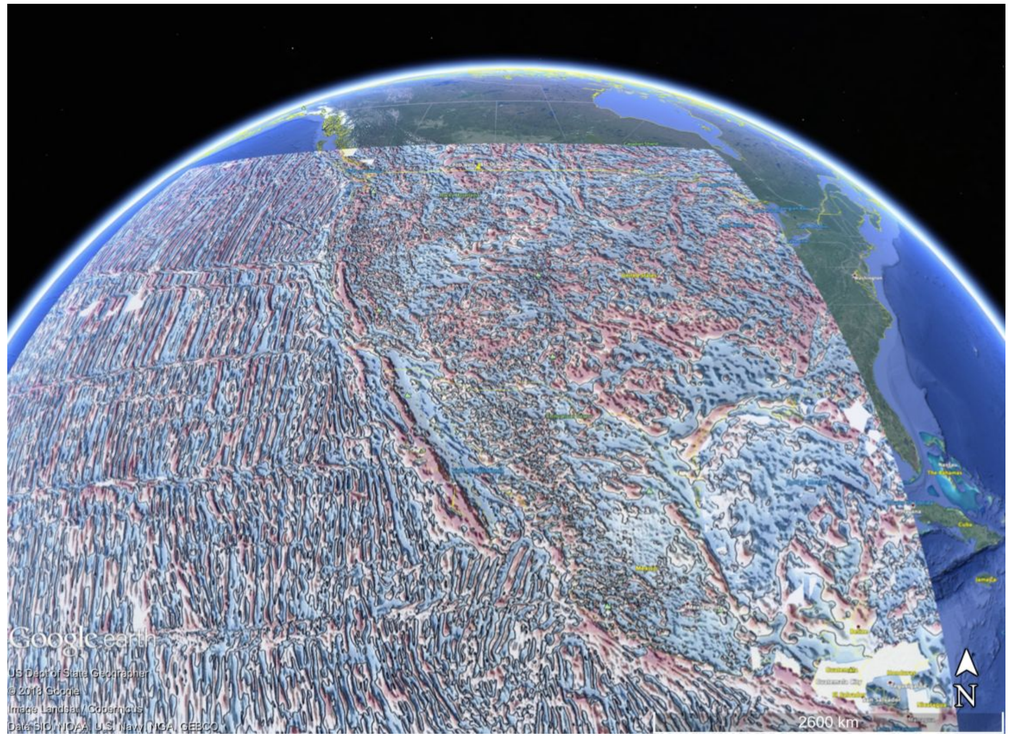
Theory & Ideas

Observations & Data

Simulations & Computation







EMAG2: Earth Magnetic Anomaly Grid (2-arc-minute resolution). Image credit: Dom Fournier (toolkit.geosci.xyz)

After Hamman, 2018

"The gears of the engine are starting to grind"

Improved multiscale, nonlinear, noisy models

CMIP6 ~15-30 PB

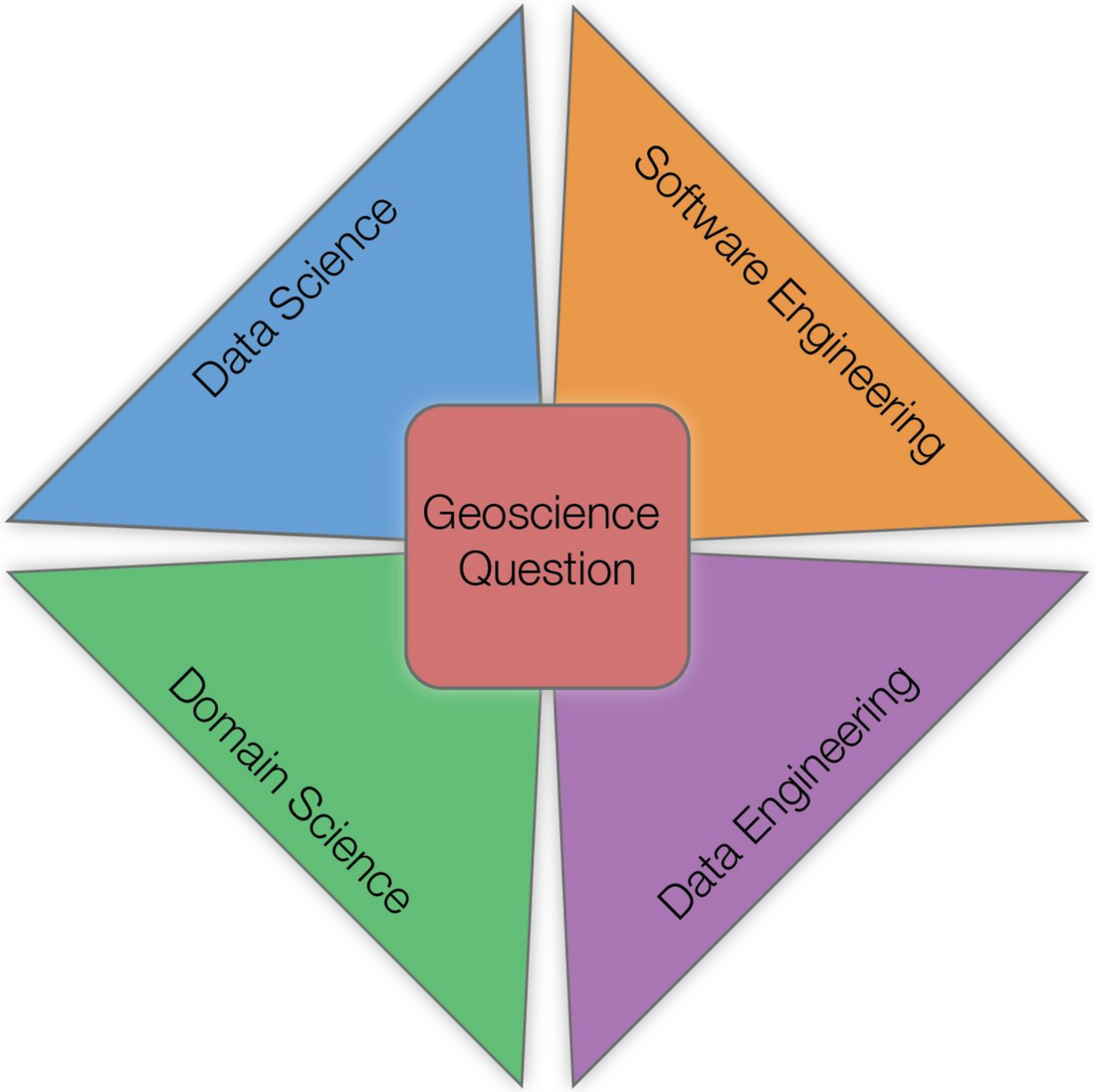
Experimental Data rates of TB/day+

Heterogeneous, multimodal

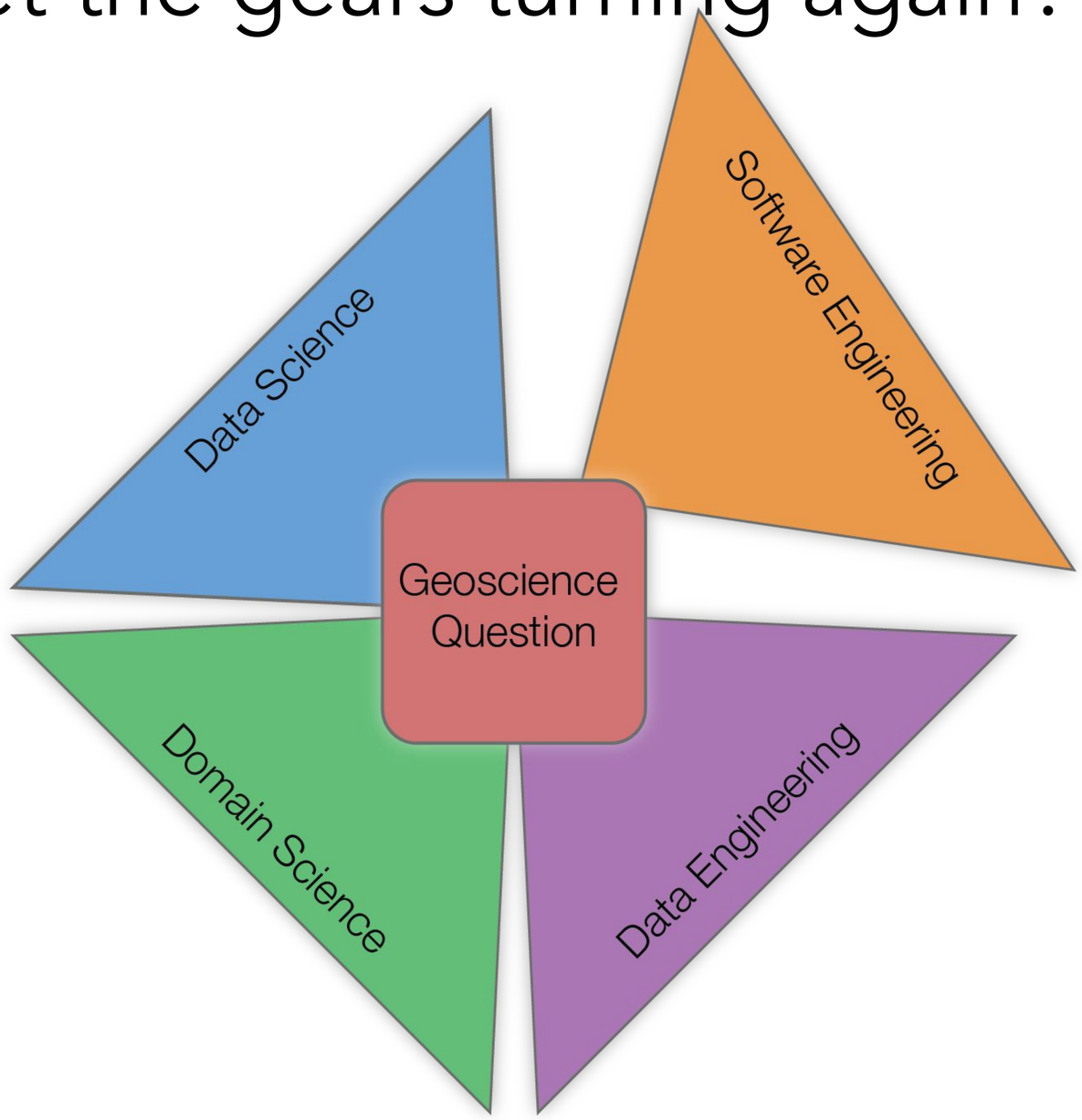
Exascale computing
Cloud engineering skills
Machine Learning pipelines

After Hamman, 2018

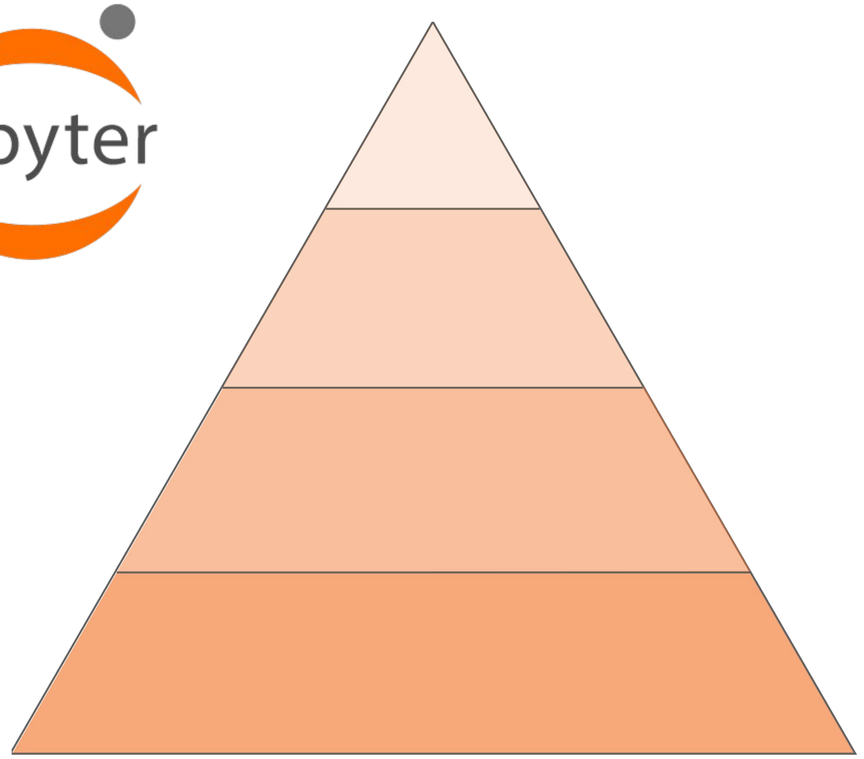
how do we get the gears turning again?



how do we get the gears turning again?



OSS: more than software

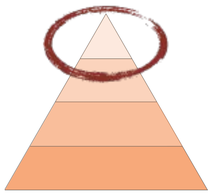


Services and content

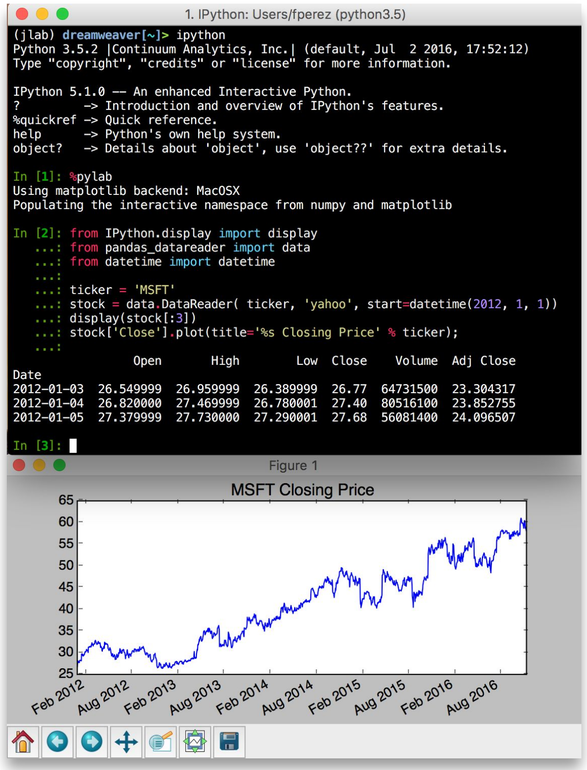
Software

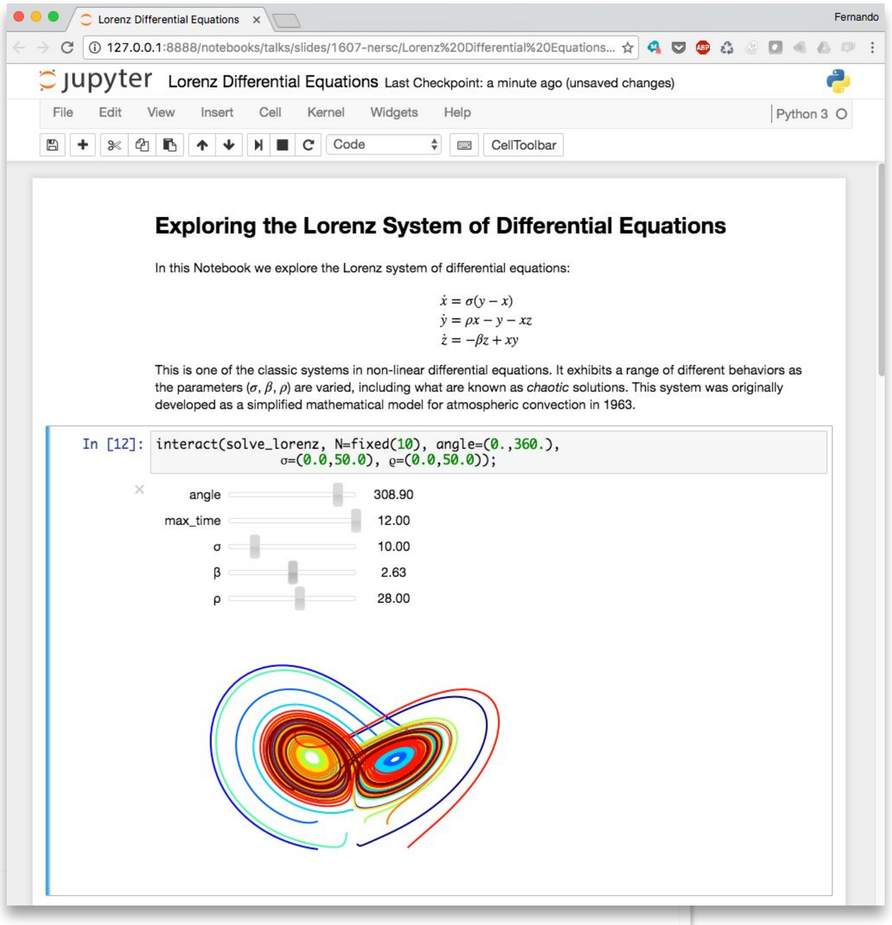
Standards and Protocols

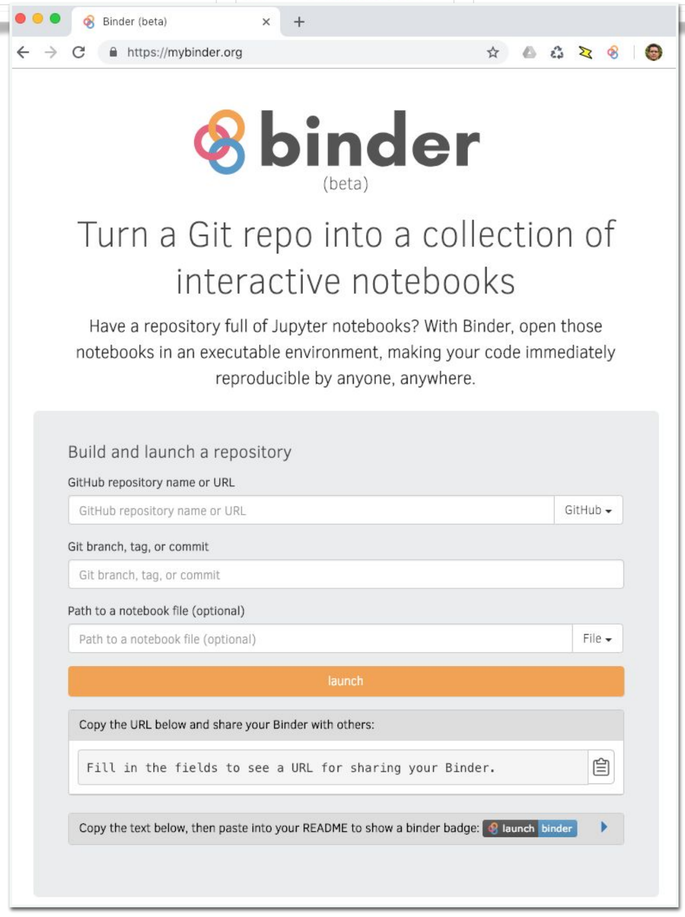
Community



Content/Services

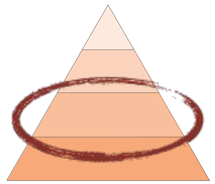






Jupyter for Organizations

JupyterHub is a multiuser version of the notebook designed for centralized deployments in companies, university classrooms and research labs.



A language agnostic protocol

Scala

perl

jupyter

julia

ERLANG

Spark

python™

IP[y]:
IPython

~100 different kernels: https://github.com/jupyter/jupyter/wiki/Jupyter-kernels

Community: formalized governance

Me :)

Brian Granger

Cal Poly, Amazon

Steering Council

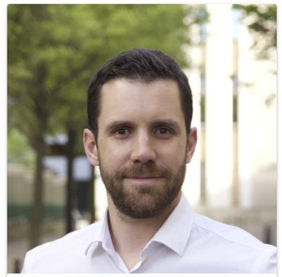
The role of the Jupyter Steering Council is to ensure, through working with and serving the broader Jupyter community, the long-term well-being of the project, both technically and as a community. The Jupyter Steering Council currently consists of the following members (in alphabetical order).

Damian Avila

Anaconda, Inc.
@damianavila on GitHub

Matthias Bussonnier

UC Merced
@Carreau on GitHub



Sylvain Corlay

QuantStack
@sylvaincorlay on GitHub

Institutional Partners

Institutional Partners are organizations that support the project by employing Jupyter Steering Council members. Current Institutional Partners include:

ANACONDA. · Bloomberg · NETFLIX · CAL POLY SAN LUIS OBISPO

Berkeley UNIVERSITY OF CALIFORNIA · QuantStack Scientific Computing · TWO SIGMA · JPMORGAN CHASE & CO.

UNIVERSITY OF CALIFORNIA MERCED · aws

Sponsors

Project Jupyter receives direct funding from the following sources:

THE LEONA M. AND HARRY B. HELMSLEY CHARITABLE TRUST · ALFRED P. SLOAN FOUNDATION · GORDON AND BETTY MOORE FOUNDATION · Google

rackspace the #1 managed cloud company

fastly · Horizon 2020 European Union funding for Research & Innovation · Microsoft · QUANSIGHT

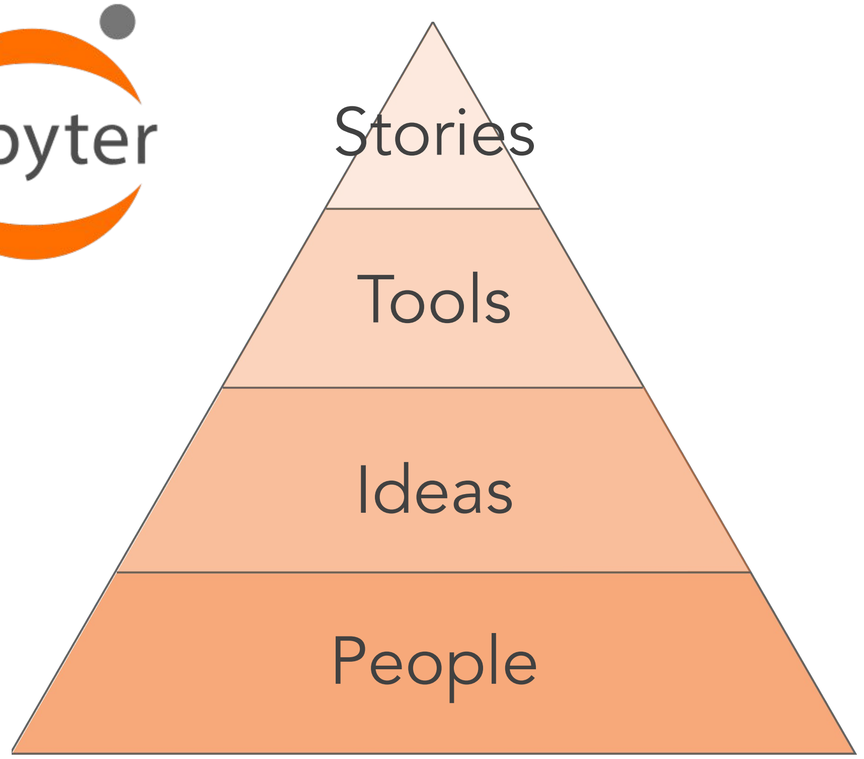
SCHMIDT FUTURES

Formal fiscal sponsorship

NUMFOCUS

OPEN CODE = BETTER SCIENCE

More than software, woven into science



Services and content: impact

Software

Standards and Protocols: ecosystem

Community: innovation & resiliency

OSS supports CORE Science*

Collaborative

Open

Reproducible

Extensible

* With a nod to the FAIR principles of open data

Collaborative?

IPython: an afternoon hack, 2001

Boulder

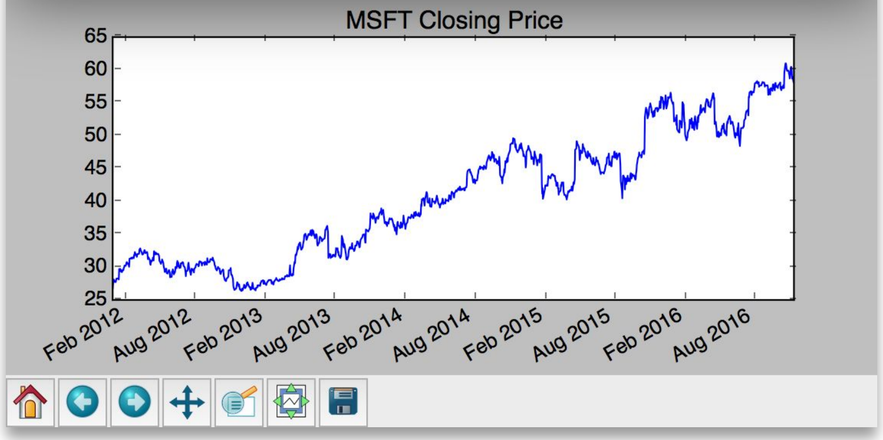
ipython-0.0.1.py

```
Globals for SI units (including g=9.8)     : _load_units   = %(_load_units)s
Starting number for prompt counter         : _prompt_ini   = %(_prompt_ini)s
Number of history items to store in cache  : _cache_size   = %(_cache_size)s
"""

#*****************************************************************************
# Configure here
_load_Numeric   = 1
_load_Gnuplot   = 1
_load_gracePlot = 1
_load_units     = 1
_cache_size     = 1000
_prompt_ini     = 1

# *** Don't modify below unless you know what you're doing. ***

# Crude first version, with minimal object structure. This could be done much
# better, by defining a Cache class (probably using weak references or
# generators). But it seems to work ok. Haven't checked for memory circularity
# problems, though.

#*****************************************************************************
#       Copyright (C) 2001 Fernando P@rez. <fperez@pizero.colorado.edu>
#
#   Distributed under the terms of the GNU General Public License.
#
#   The full text of the GPL is available at:
#
#                  http://www.gnu.org/copyleft/gpl.html
#*****************************************************************************
__author__ = 'Fernando P@rez. <fperez@pizero.colorado.edu>'
__version__ = '0.1'

#****************************************************************************
nitions

rompt1:
e interactive prompt like Mathematica's."""
r__(self):
rn '\nIn['+`_prompt_count`+']:= '

rompt2:
e interactive continuation prompt."""
r__(self):
rn ' ... '+' '*(len('In['+`_prompt_count`+']:= ')-3)

#****************************************************************************
efinitions

_print(arg):
ing with history cache management.

invoked everytime the interpreter needs to print, and is activated
by setting the variable sys.displayhook to it."""

global _p,_pp,_ppp,_cache,_prompt_count
```

1. IPython: Users/fperez (python3.5)

```
(jlab) dreamweaver[~]> ipython
Python 3.5.2 |Continuum Analytics, Inc.| (default, Jul  2 2016, 17:52:12)
Type "copyright", "credits" or "license" for more information.

IPython 5.1.0 -- An enhanced Interactive Python.
?         -> Introduction and overview of IPython's features.
%quickref -> Quick reference.
help      -> Python's own help system.
object?   -> Details about 'object', use 'object??' for extra details.

In [1]: %pylab
Using matplotlib backend: MacOSX
Populating the interactive namespace from numpy and matplotlib

In [2]: from IPython.display import display
   ...: from pandas_datareader import data
   ...: from datetime import datetime
   ...:
   ...: ticker = 'MSFT'
   ...: stock = data.DataReader( ticker, 'yahoo', start=datetime(2012, 1, 1))
   ...: display(stock[:3])
   ...: stock['Close'].plot(title='%s Closing Price' % ticker);
   ...:
                Open       High        Low  Close    Volume  Adj Close
Date
2012-01-03  26.549999  26.959999  26.389999  26.77  64731500  23.304317
2012-01-04  26.820000  27.469999  26.780001  27.40  80516100  23.852755
2012-01-05  27.379999  27.730000  27.290001  27.68  56081400  24.096507

In [3]:
```

Figure 1



Multiple stakeholders, team effort

- Academic scientists
- Educators
- Industry
- Government
- Media/journalism
- 1500+ community volunteers!

Open?

Dimensions of Openness

- Open source code
- Open (FAIR) data
- Open access publications & artifacts
- Open standards: interoperability (even with proprietary tools)
- Open community: all welcome (and mean it!)
- …

Reproducible?
The foundation of collaboration!

the science more than the paper

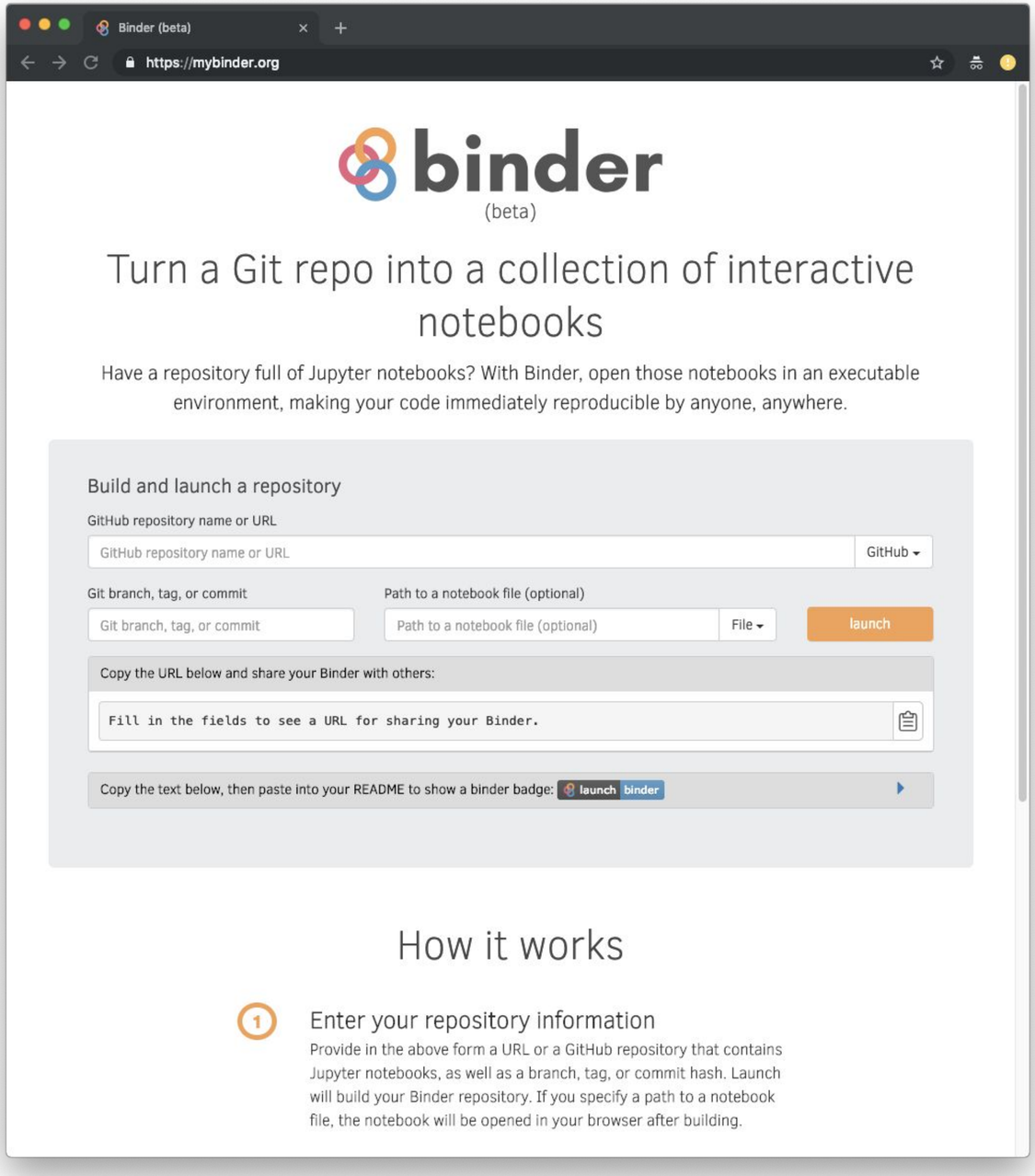
An article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures.

-- Buckheit and Donoho (paraphrasing Claerbout)
WaveLab and Reproducible Research, 1995

(and a place to run the code?)

binder

shareable, interactive, reproducible environments from your public git repository



mybinder.org

Black holes! LIGO, Sept 14, 2015

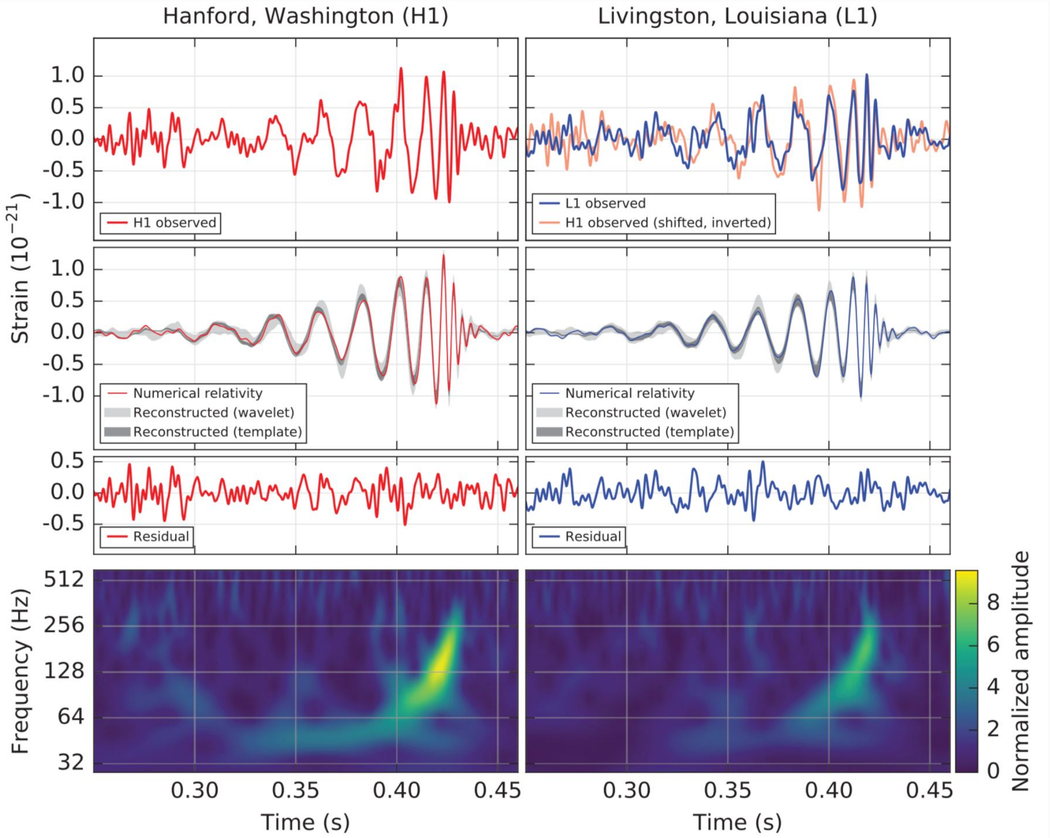


FIG. 1. The gravitational-wave event GW150914 observed by the LIGO Hanford (H1, left column panels) and Livingston (L1, right column panels) detectors. Times are shown relative to September 14, 2015 at 09:50:45 UTC. For visualization, all time series are filtered with a 35–350 Hz bandpass filter to suppress large fluctuations outside the detectors' most sensitive frequency band, and band-reject

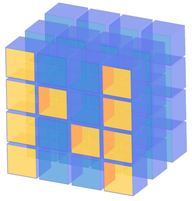
Make sound files

Make wav (sound) files from the filtered, downsampled data, +-2s around the event.

```
# make wav (sound) files from the whitened data, +-2s around the event.
from glob import glob
from IPython.display import display, Audio

from scipy.io import wavfile

# function to keep the data within integer limits, and write to wavfile:
def write_wavfile(filename,fs,data):
    d = np.int16(data/np.max(np.abs(data)) * 32767 * 0.9)
    wavfile.write(filename,int(fs), d)

tevent = 1126259462.422          # Mon Sep 14 09:50:45 GMT 2015
deltat = 2.                      # seconds around the event

# index into the strain time series for this time interval:
indxt = np.where((time >= tevent-deltat) & (time < tevent+deltat))

# write the files:
write_wavfile("GW150914_H1_whitenbp.wav",int(fs), strain_H1_whitenbp[indxt])
write_wavfile("GW150914_L1_whitenbp.wav",int(fs), strain_L1_whitenbp[indxt])
write_wavfile("GW150914_NR_whitenbp.wav",int(fs), NR_H1_whitenbp)

for wav in glob('*whitenbp.wav'):
    display(wav)
    display(Audio(filename=wav))
```

```
'GW150914_H1_whitenbp.wav'
```

0:00

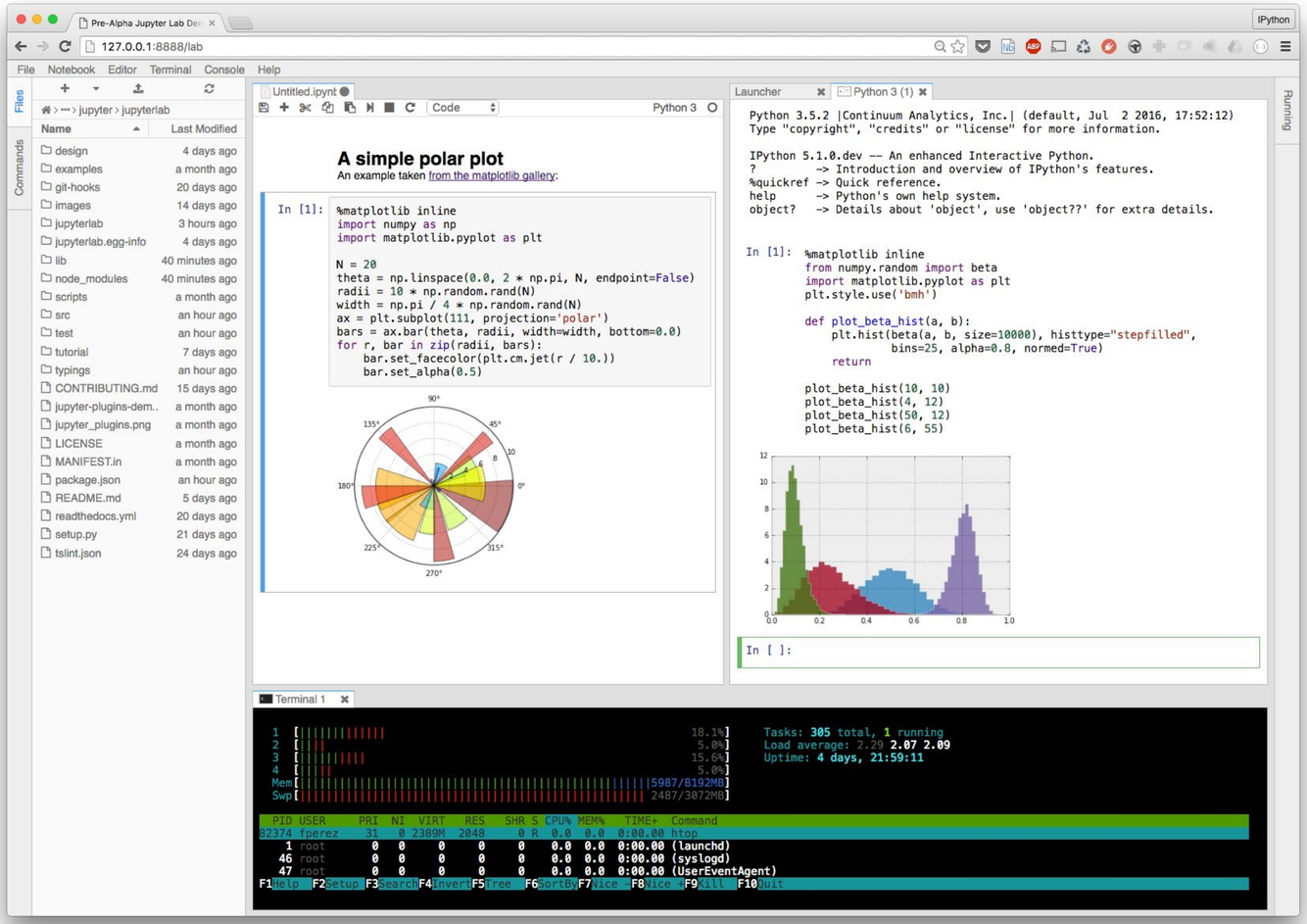


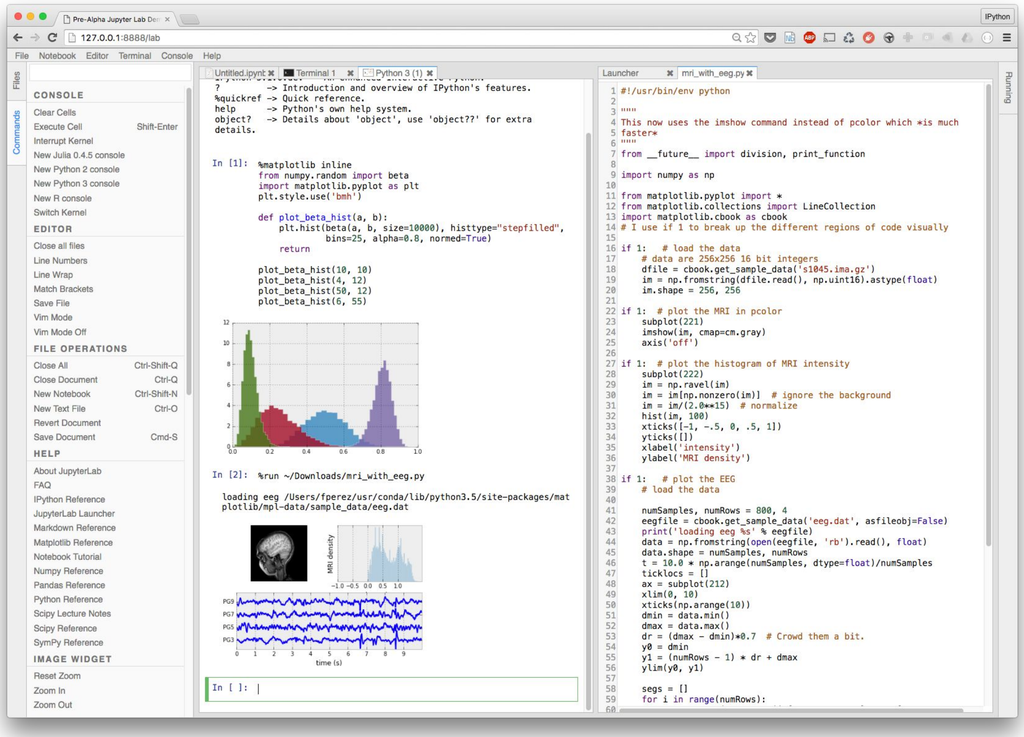
http://bit.ly/black-holes-woop

Audio

Extensible?

JupyterLab: a grand unified theory of Jupyter





Huge Team Effort!

C. Colbert, S. Corlay, A. Darian, B. Granger, J. Grout, P. Ivanov, I. Rose, S. Silvester, C. Willing, J. Zosa-Forde ...

ANACONDA.

Bloomberg

Notebooks++

File Edit View Run Kernel Tabs Settings Help

Files Running Commands Cell Tools Tabs

Lorenz.ipynb

Code Python 3

The Lorenz Differential Equations

Before we start, we import some preliminary libraries. We will also import (below) the accompanying `lorenz.py` file, which contains the actual solver and plotting routine.

In [1]:
```
%matplotlib inline
from ipywidgets import interactive, fixed
```

We explore the Lorenz system of differential equations:

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

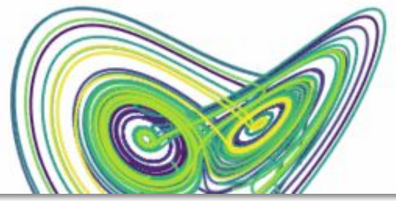
Let's change (σ, β, ρ) with ipywidgets and examine the trajectories.

In [2]:
```
from lorenz import solve_lorenz
w=interactive(solve_lorenz,sigma=(0.0,50.0),rho=(0.0,50.0))
w
```

sigma 10.10

beta 2.63

rho 23.30

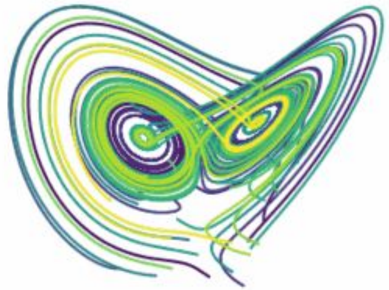


Output View

sigma 10.10

beta 2.63

rho 23.30



Lorenz.pdf

1 The Lorenz Differential Equations

Before we start, we import some preliminary libraries. We will also import (below) the accompanying `lorenz.py` file, which contains the actual solver and plotting routine.

```
In [1]: %matplotlib inline
        from ipywidgets import interactive, fixed
```

We explore the Lorenz system of differential equations:

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

Let's change (σ, β, ρ) with ipywidgets and examine the trajectories.

```
In [2]: from lorenz import solve_lorenz
        w=interactive(solve_lorenz,sigma=(0.0,50.0),rho=(0.0,50.0))
        w

interactive(children=(FloatSlider(value=10.0, description='sigma', max=50.0), FloatSlider(valu
```

Beyond notebooks

File Edit View Run Kernel Tabs Settings Help

Launcher | markdown_python.m

```
1  # Markdown
2
3  This is a regular markdown file. In JupyterLab you can open and
   edit
4  the file in the file editor, or in the markdown viewer. As you
   edit the
5  file the rendered markdown will automatically update.
6
7  $$ x+\alpha+ \sum_{i=0}y_i $$
8
9  # Including Python code
```

markdown_python.m

```
import numpy as np
import pandas as pd
data = {
    'x': np.random.rand(100),
    'y': np.random.rand(100),
    'color': np.random.rand(100),
    'size': 100.0*np.random.rand(100)
}
df = pd.DataFrame(data)
df.head()
```

[1]:

	x	y	color	size
0	0.617536	0.508976	0.945603	8.052391
1	0.051126	0.322041	0.237476	49.005570
2	0.247221	0.476783	0.875143	81.492632
3	0.713989	0.480705	0.462035	7.631572
4	0.312106	0.888199	0.640439	53.201591

[ ]:

markdown_python.m

Markdown

This is a regular markdown file. In JupyterLab you can open and edit the file in the file editor, or in the markdown viewer. As you edit the file the rendered markdown will automatically update.

$$x+\alpha+\sum_{i=0}y_i$$

Including Python code

Here is a block of Python code in the markdown file:

```
a = 100
```

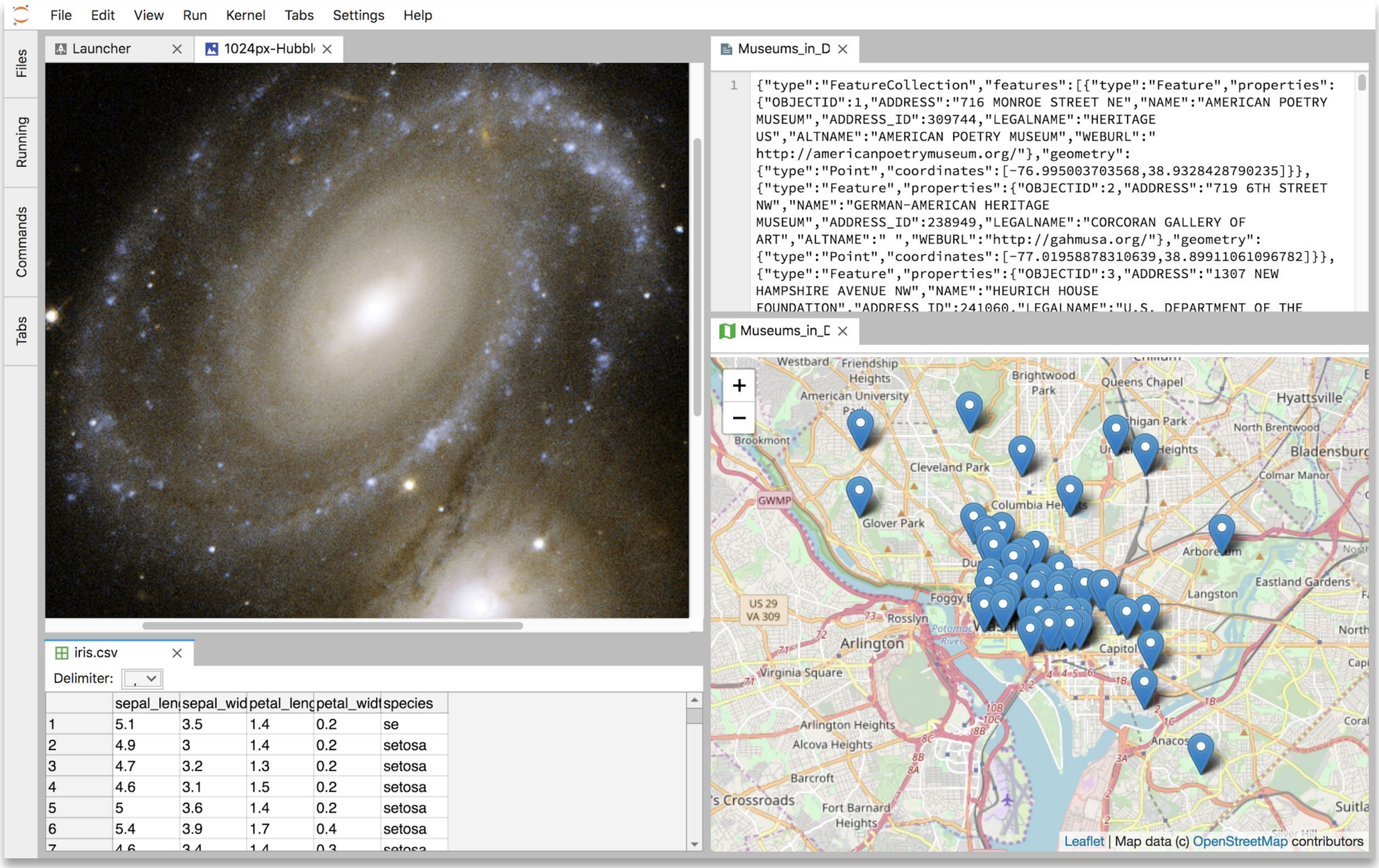
Let's attach a Python 3 Kernel and Console to this markdown file. Then we can select lines of code in the markdown file and run them in the console by pressing `Shift+Enter`. Let's do something more complicated:

First import `matplotlib`, `numpy` and `pandas`, and create a data frame:

```
%matplotlib inline
from matplotlib import pyplot as plt
from matplotlib import style
import numpy as np
import pandas as pd
data = {
    'x': np.random.rand(100),
    'y': np.random.rand(100),
    'color': np.random.rand(100),
    'size': 100.0*np.random.rand(100)
}
df = pd.DataFrame(data)
df.head()
```

0 1 ipythongfm Ln 7, Col 4 Spaces: 4 markdown_python.md

Data



	sepal_len	sepal_wid	petal_leng	petal_widt	species
1	5.1	3.5	1.4	0.2	se
2	4.9	3	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa
7	4.6	3.4	1.4	0.3	setosa

JupyterLab is extensible: FlyBrainLab

An Interactive Computing Platform for the Fly Brain

BIONET Group, Columbia University

http://www.bionet.ee.columbia.edu

Aurel A. Lazar (PI)

Tingkai Liu

Mehmet K. Turkcan

Chung-Heng Yeh

Yiyin Zhou

http://fruitflybrain.org

National infrastructure, from K-12 to HPC

cybera

CYBERA › NETWORK › SERVICES › PROJECTS › NEWS & EVENTS › MEMBERSHIP › CONTACT US ›

Jupyter 'All-in-One' Science Platform

Learning and sharing in a flexible, collaborative and interactive way.

Jupyter is an integrative application that incorporates math, science and engineering tools, along with communication and visualization resources, in one web-based platform. Simply put: It enables a broad suite of computing capabilities on any device that has an internet connection. For free.

Cybera and the Pacific Institute for the Mathematical Sciences (PIMS) have teamed up to increase access to, and awareness of, Jupyter. Cybera is hosting the platform on its Rapid Access Cloud, and is offering free access (and advice on how to get started) to Canada's public and innovation sectors.

Who Is Jupyter Useful For?

Use Cases

Pacific Institute for the Mathematical Sciences

J. Colliander,
I. Allison,
B. Carra

compute canada | calcul canada

Home

About

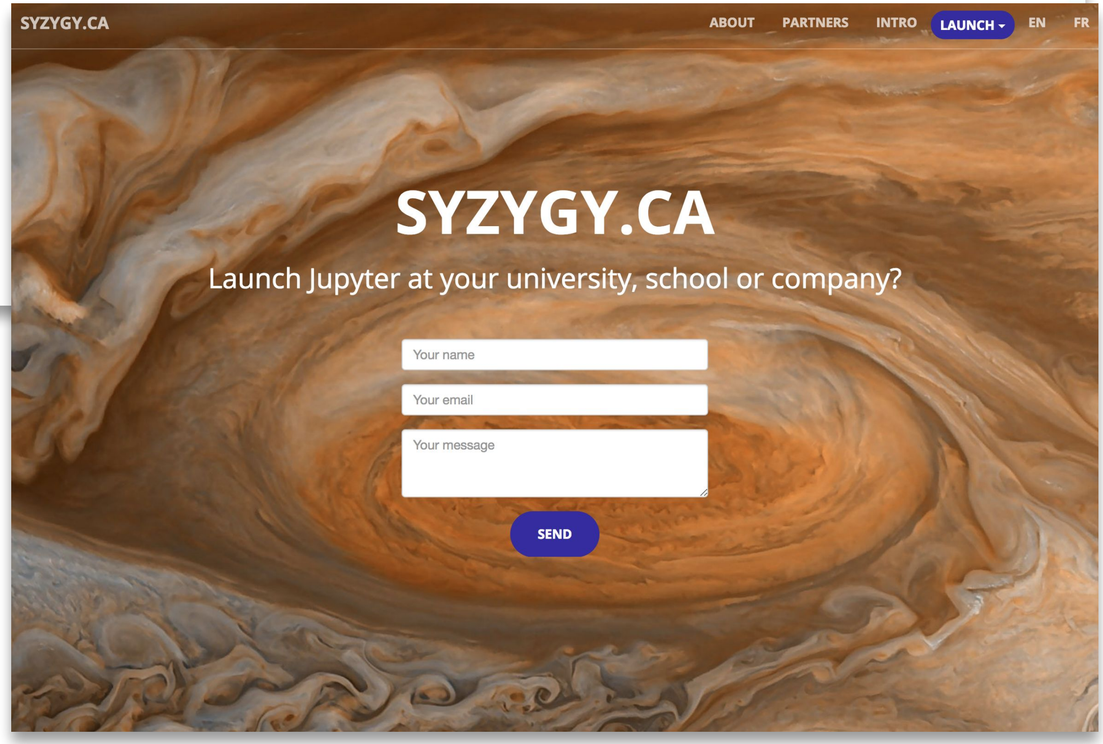
Renewing Canada's Advanced Research Computing Platform

Research Data Management

News

15/03/2017 Featured, News

compute canada and pims launch jupyter service for researchers



Microsoft

Microsoft Azure

COCALC

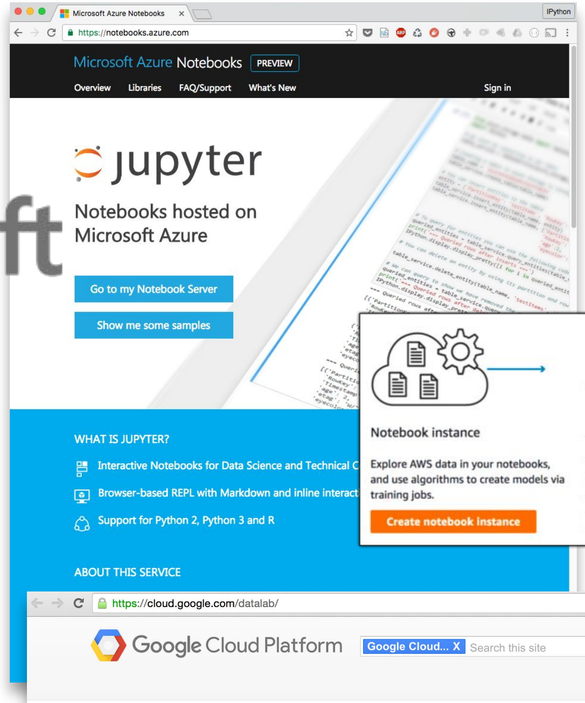
CODE OCEAN

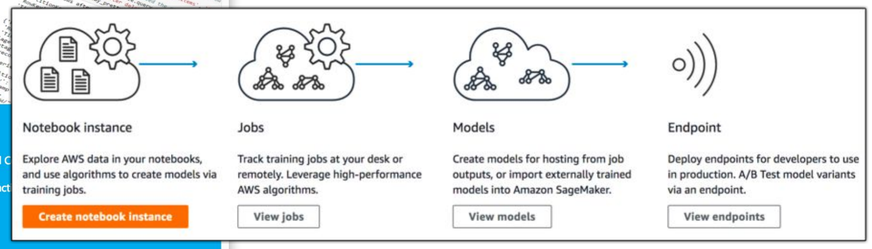
Google

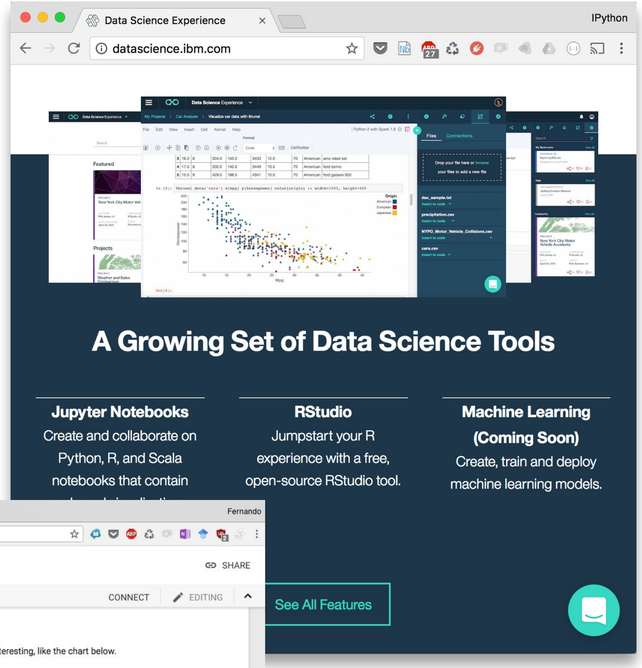
IBM

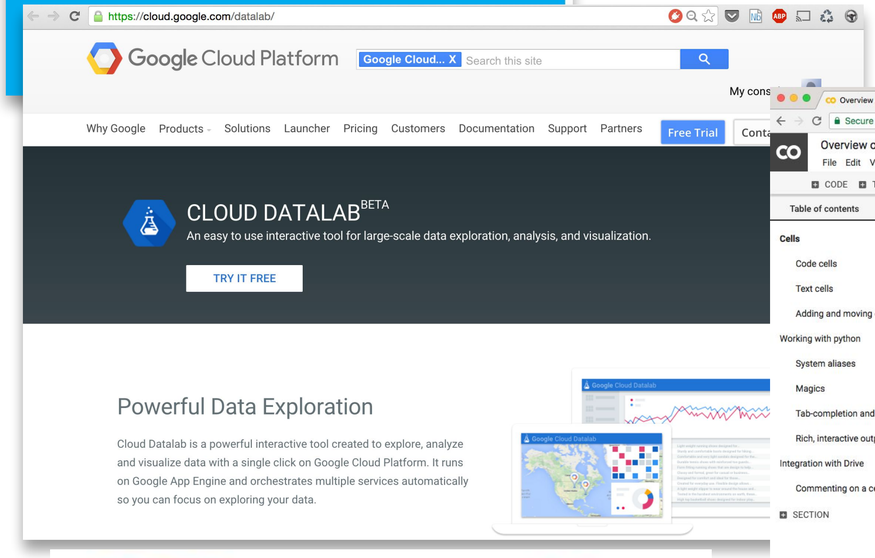
Amazon SageMaker

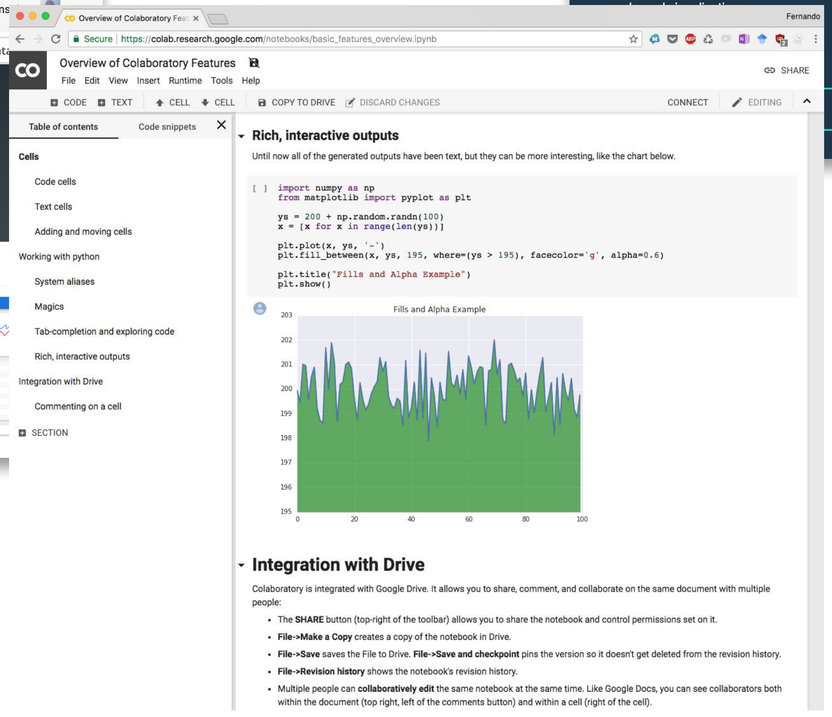
ANACONDA®





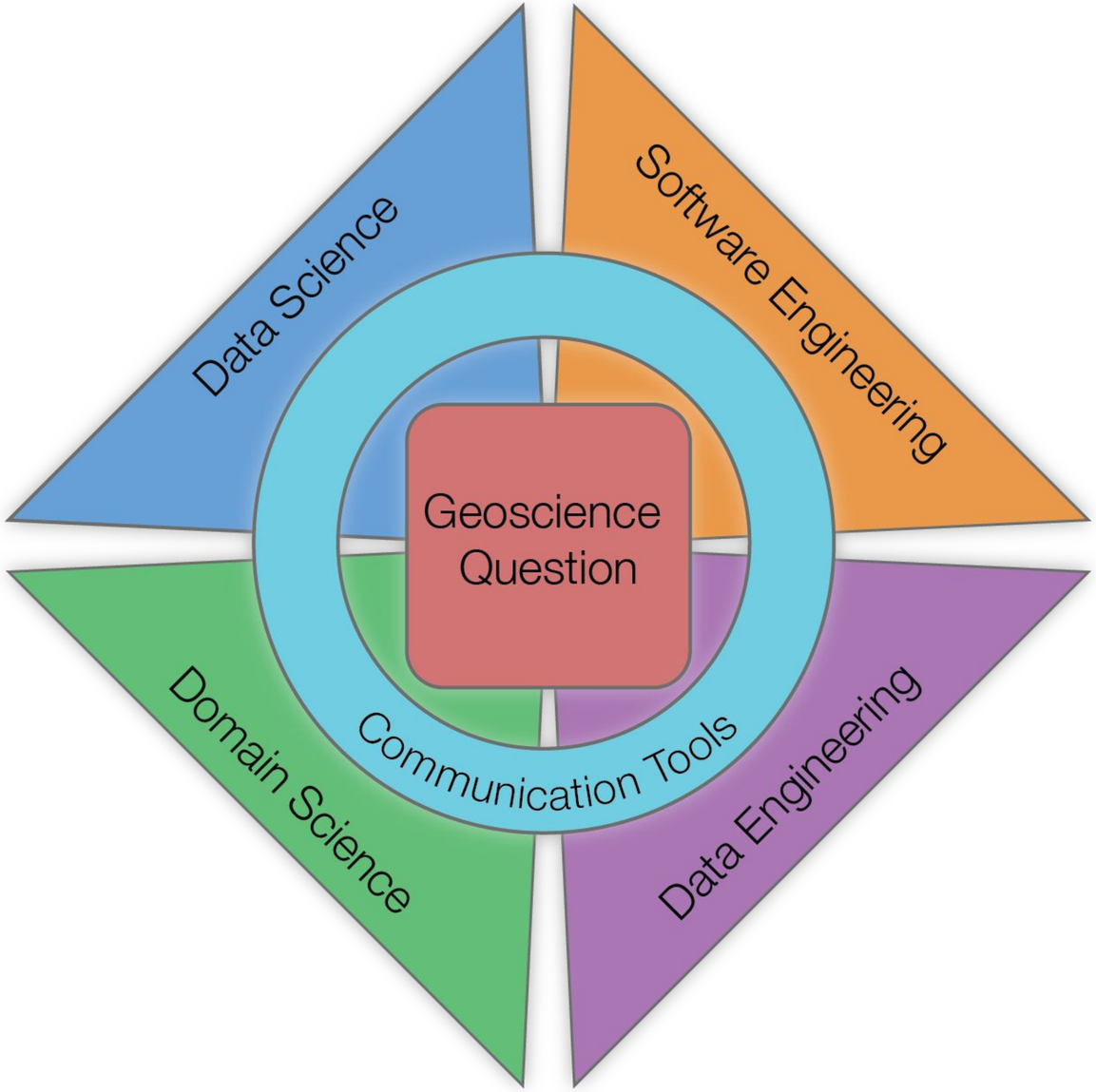




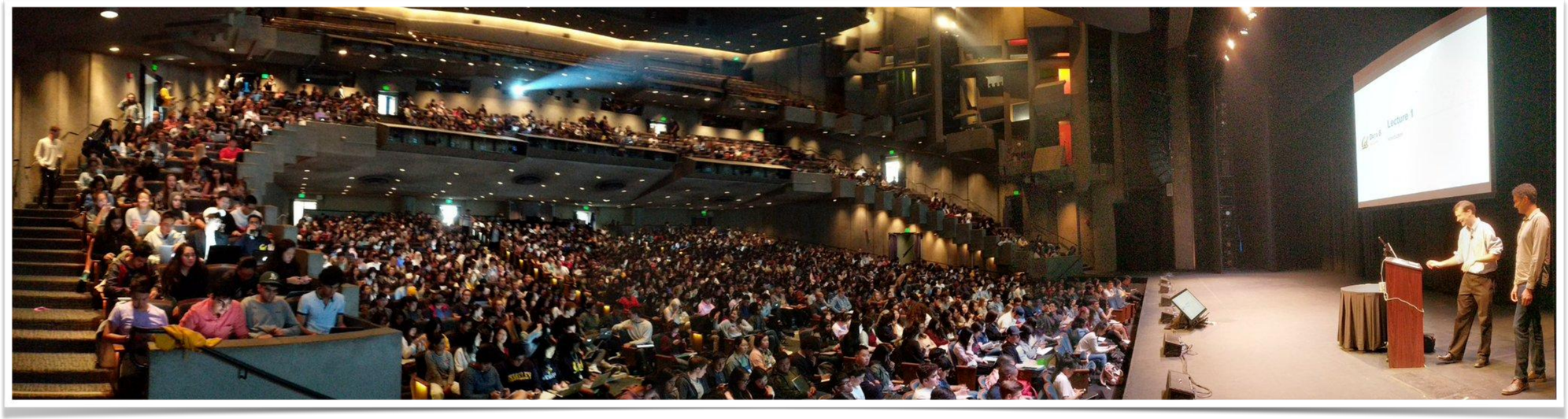


Impact: Research and Education

Impact: Research and Education



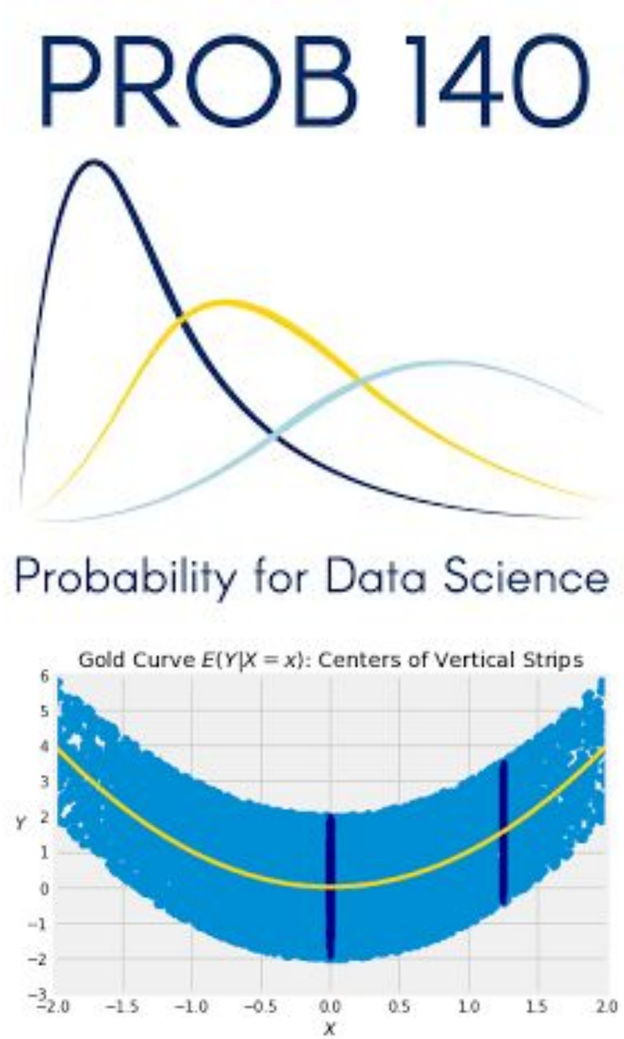
Fall 2018

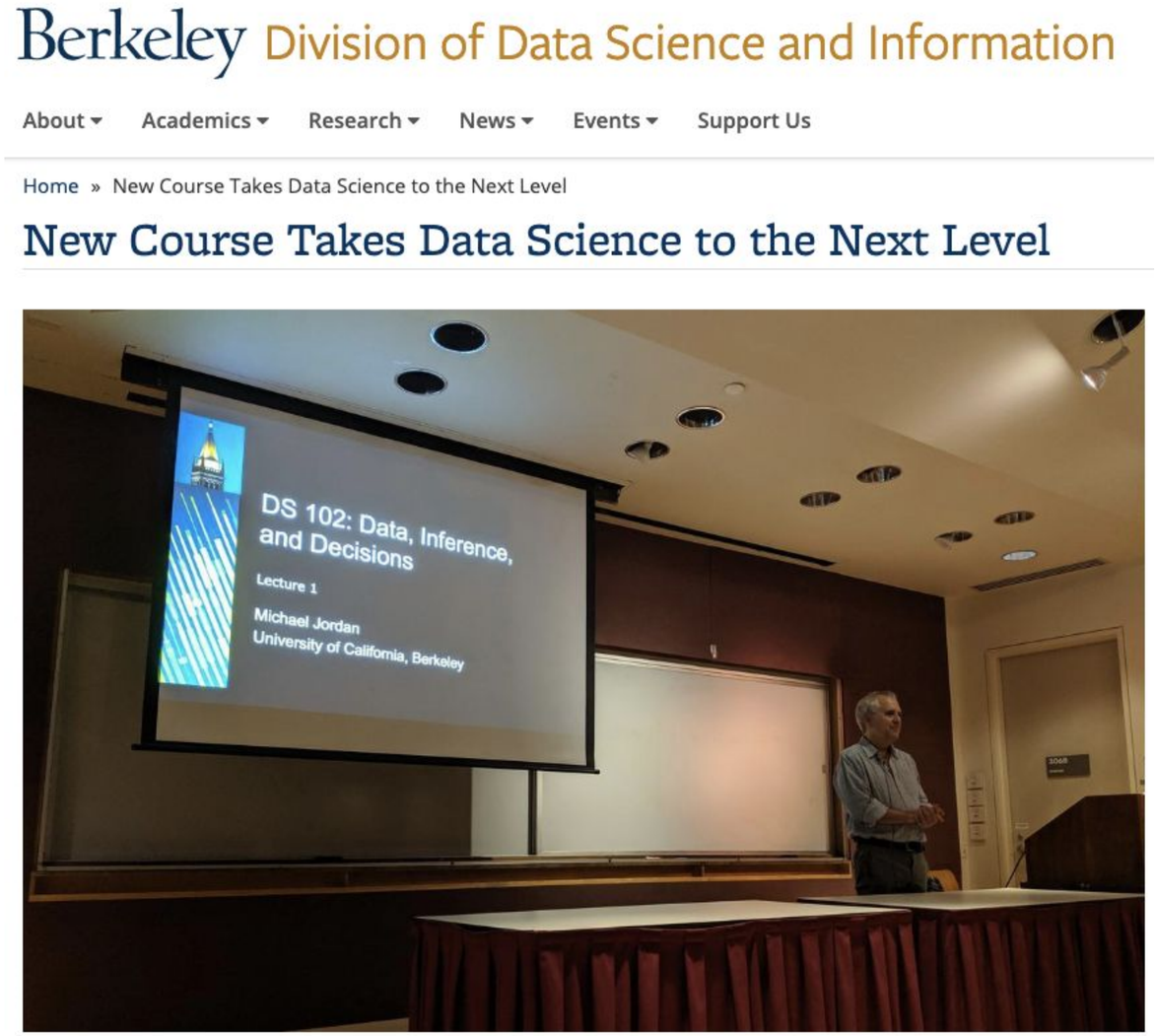


Data 8:
~1,300
students

Data 100:
~800 students

Prob 140 and Data 102





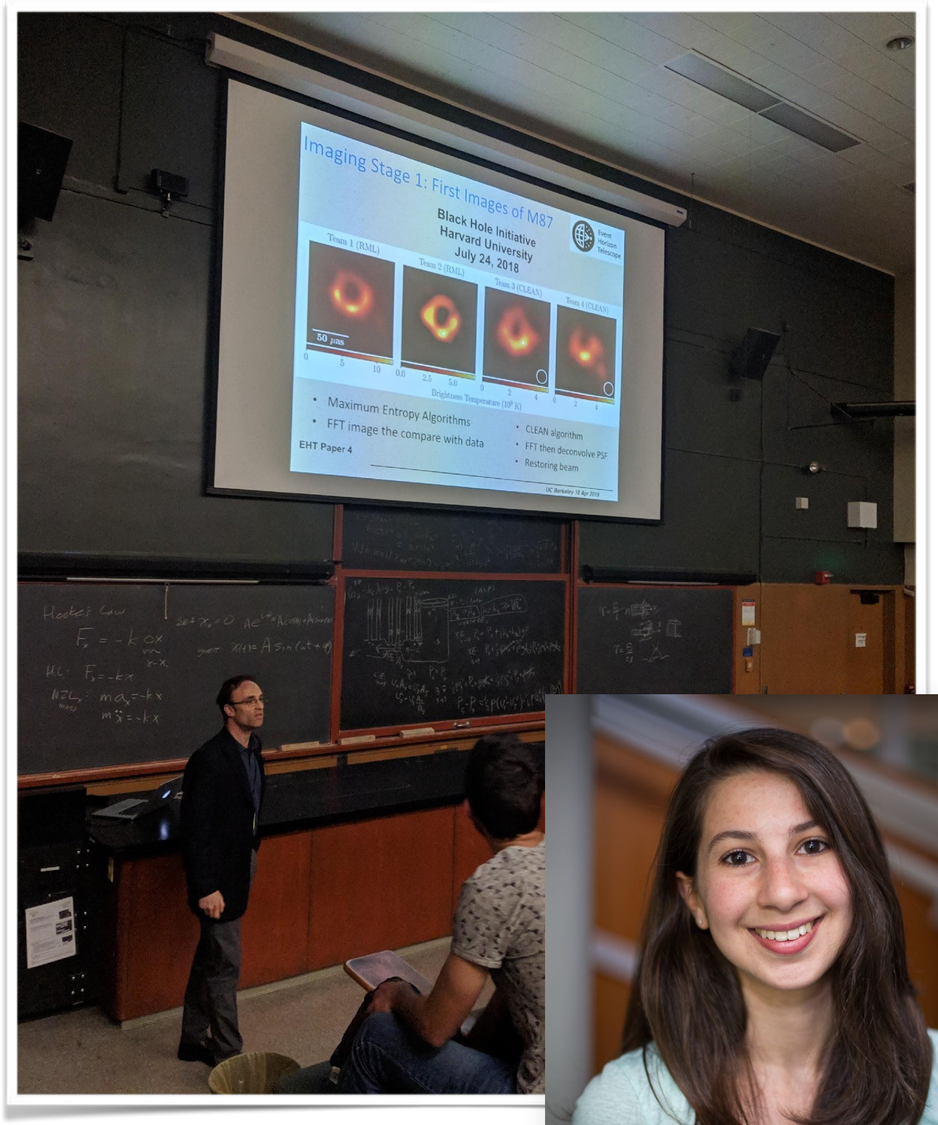
http://prob140.org: Probability for Data Science

http://data102.org: Data, Inference and Decisions

With these tools, we provide

- Broad disciplinary reach and impact of statistical thinking.
- Drastically lowered barriers to student access - intellectual and economic.
- Lowered barriers for faculty* to engage with statistical and computational ideas.
  - (*) often from non computational/statistical domains

April 18/19, 2019: Shep Doeleman & Katie Bouman



THE ASTROPHYSICAL JOURNAL LETTERS

First M87 Event Horizon Telescope Results. III. Data Processing and Calibration

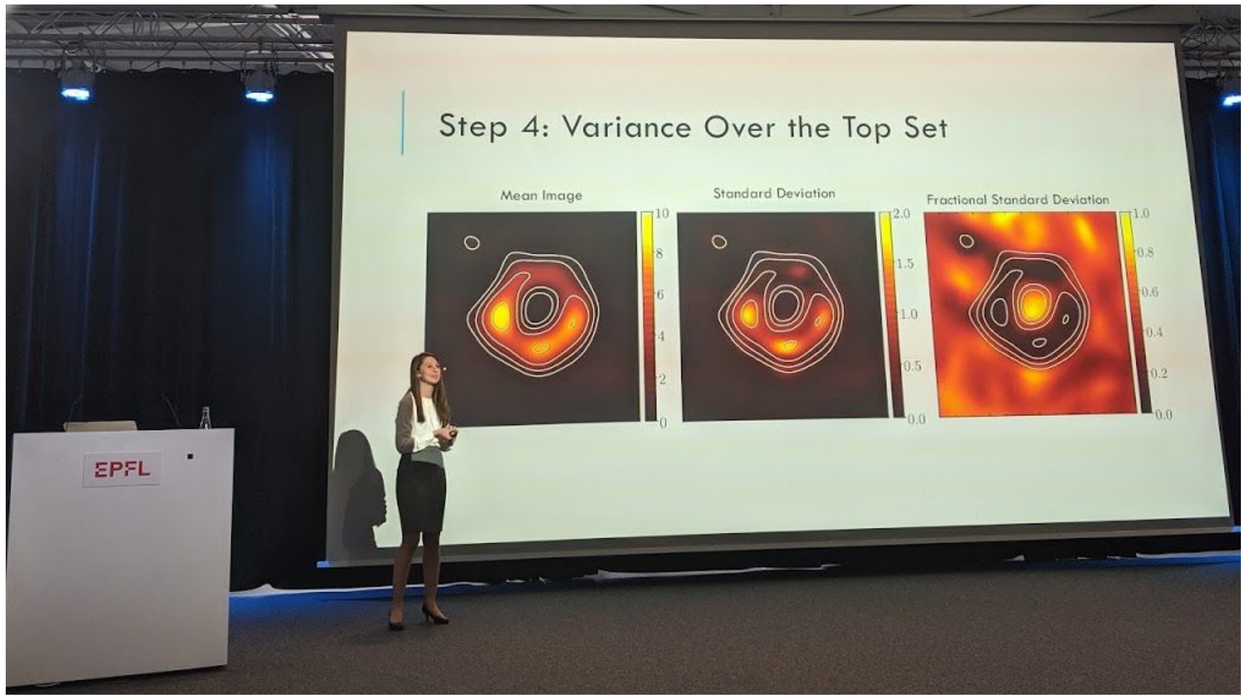
The Event Horizon Telescope Collaboration, Kazunori Akiyama[1,2,3,4], Antxon Alberdi[5], Walter Alef[6], Keiichi Asada[7], Rebecca Azulay[8,9,6], Anne-Kathrin Baczko[6], David Ball[10], Mislav Baloković[4,11], John Barrett[2] +Show full author list

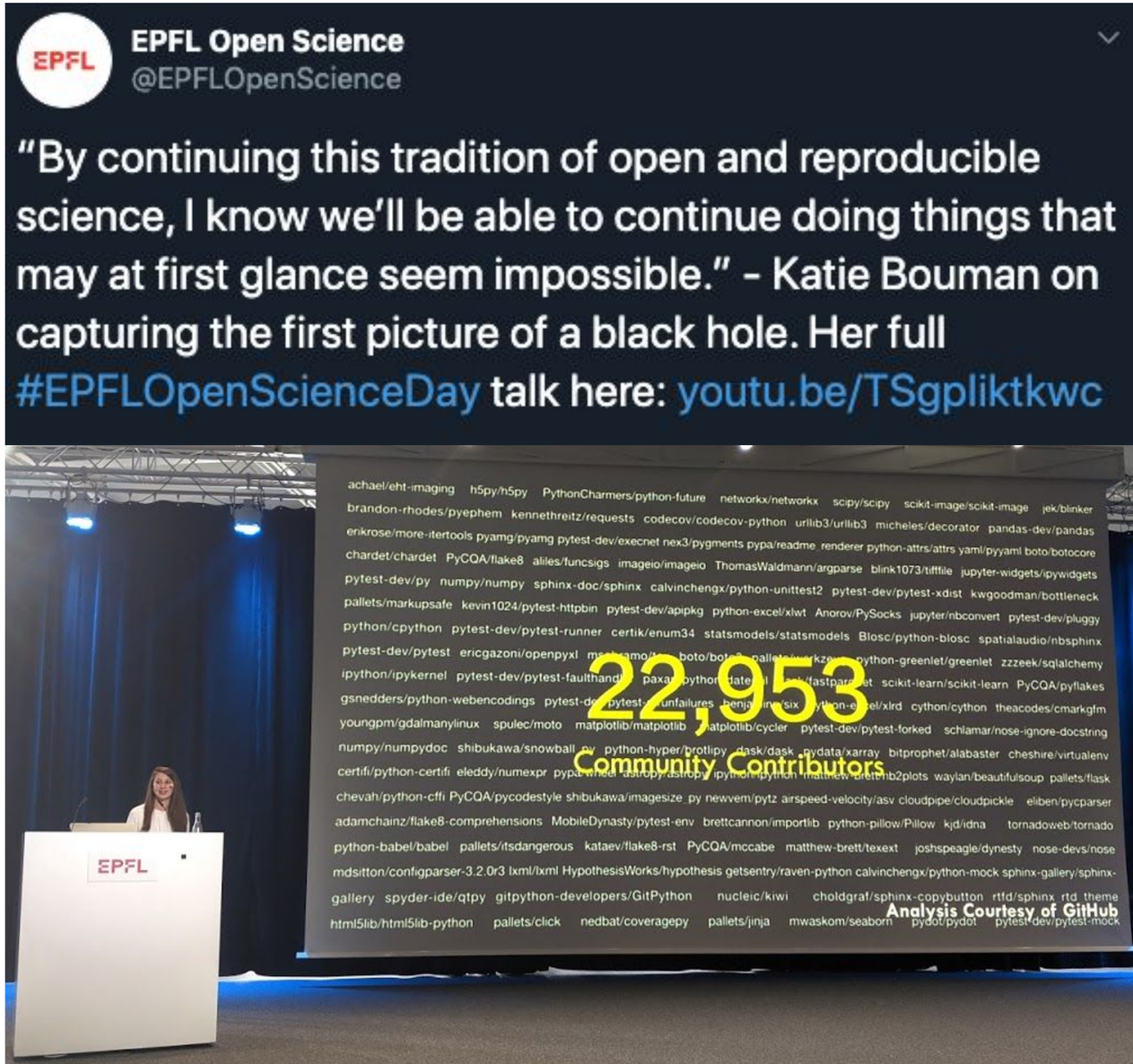
Published 2019 April 10 • © 2019. The American Astronomical Society.
The Astrophysical Journal Letters, Volume 875, Number 1

Software: DiFX (Deller et al. 2011), CALC, PolConvert (Martí-Vidal et al. 2016), HOPS (Whitney et al. 2004), CASA (McMullin et al. 2007), AIPS (Greisen 2003), ParselTongue (Kettenis et al. 2006), GNU Parallel (Tange 2011), GILDAS, eht-imaging (Chael et al. 2016, 2018), Numpy (van der Walt et al. 2011), Scipy (Jones et al. 2001), Pandas (McKinney 2010), Astropy (The Astropy Collaboration et al. 2013, 2018), Jupyter (Kluyver et al. 2016), Matplotlib (Hunter 2007).

Event Horizon Telescope

K. Bouman @ EPFL, Oct'19





Video of talk: https://youtu.be/TSgpliktkwc

CORE Geoscience?

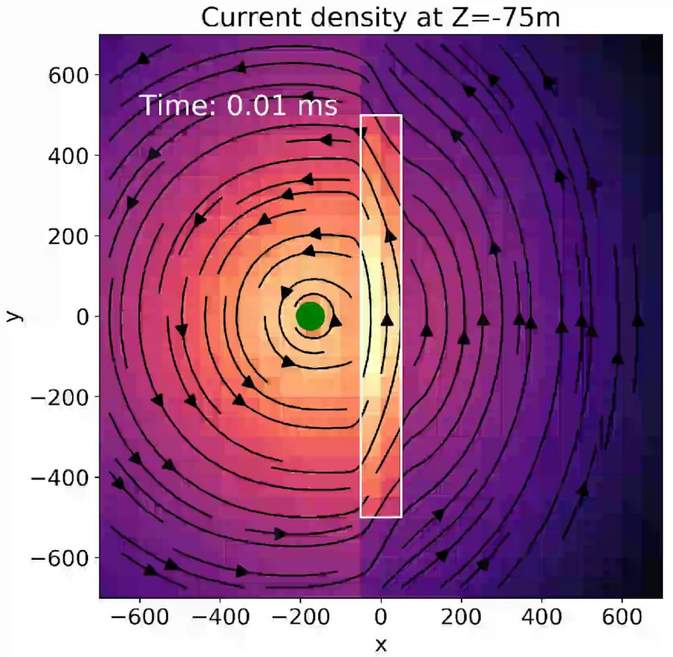
simpeg

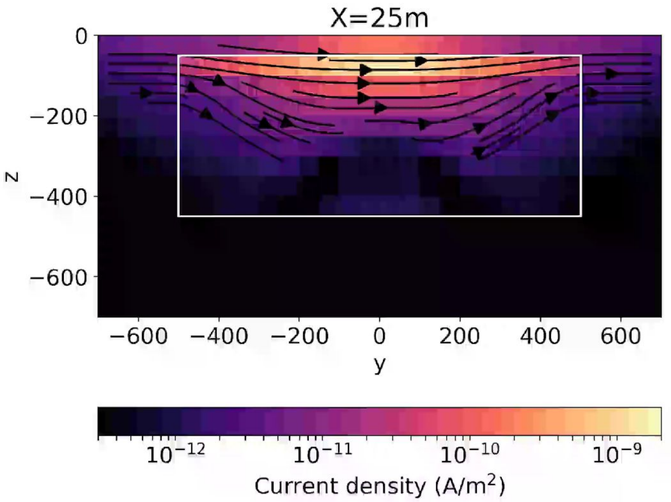
tools by and for researchers

- Modular, multi-physics
  - Gravity
  - Magnetics
  - Direct current resistivity
  - Induced Polarization
  - Electromagnetics
    - Frequency Domain
    - Time Domain
  - Fluid Flow
    - Richards Equation

https://simpeg.xyz

3D Airborne Time Domain EM






Lindsey

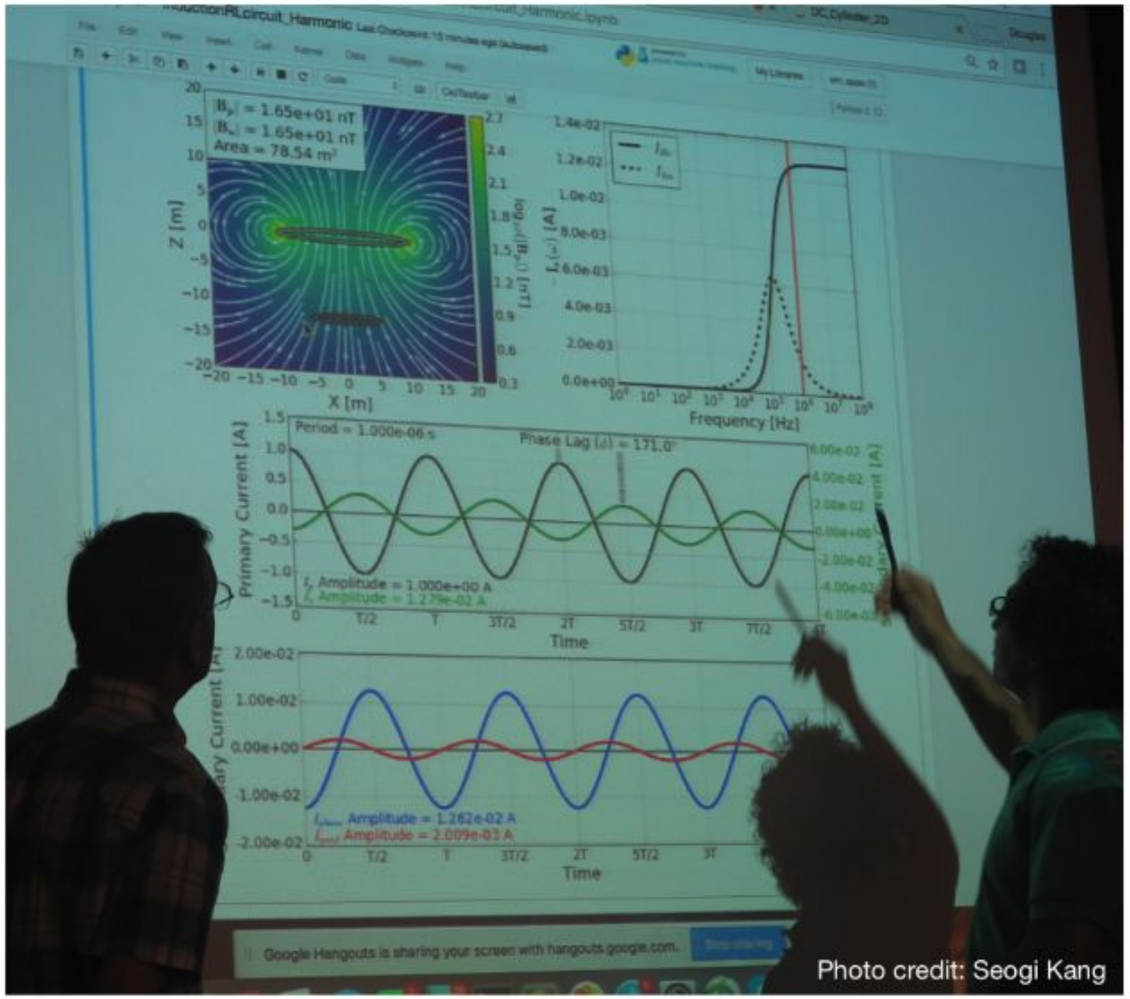

Seogi


Rowan


Doug

$$\nabla \times \vec{E} + i\omega \vec{B} = 0$$

$$\nabla \times \mu^{-1} \vec{B} - \sigma \vec{E} = \vec{J}_s$$



JupyterLab

jupyter.nersc.gov/user/lh...

File Edit View Run Kernel Tabs Settings Help

TDEM_current_plots_src_t

slurm-23917872.out

Code

simpeg37

max_depth 1500

clim_min 1e-8

clim_max 1e-4

model_key: casing | approx_casing | solid_steel | all

view: sigma | mur | e | j | dbdt | dhdt | charge | charge_density

prim_sec: total | primary | secondary | percent

src_ind 0

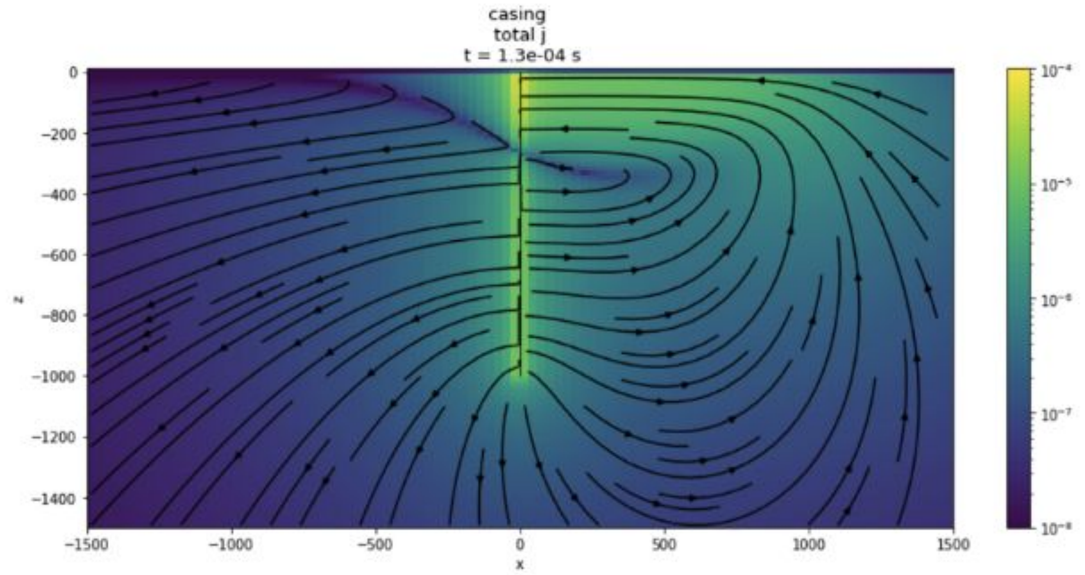
theta_ind 3

time_ind 31

show_mesh

use_aspect

casing_outline



4 2 simpeg37 | Idle Mode: Command Ln 1, Col 1 TDEM_current_plots_src_term.ipynb

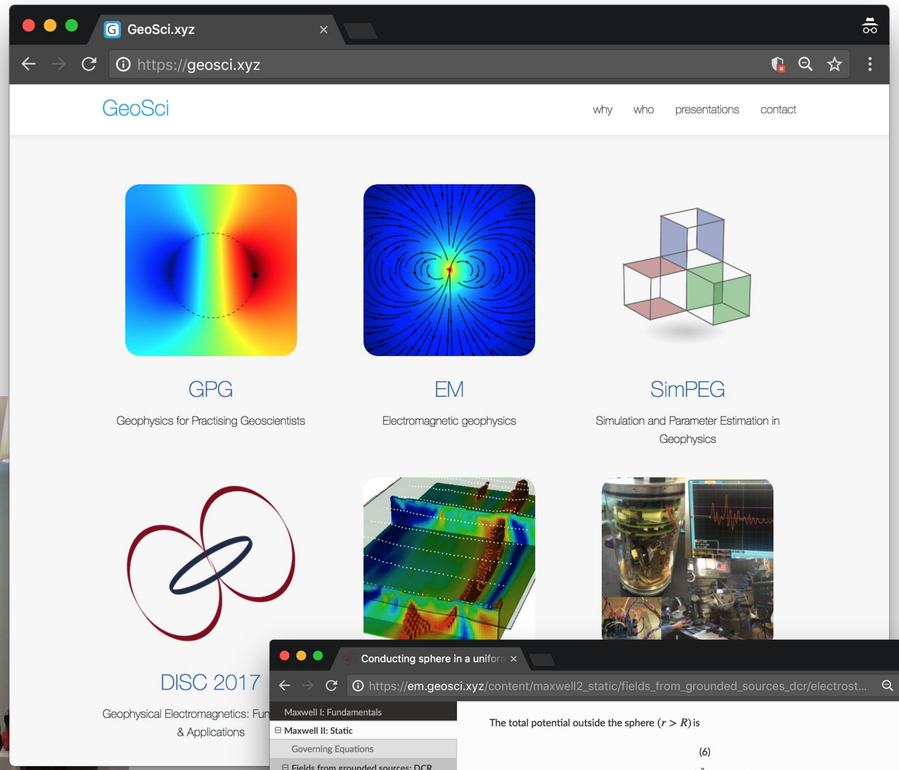
GeoSci.xyz

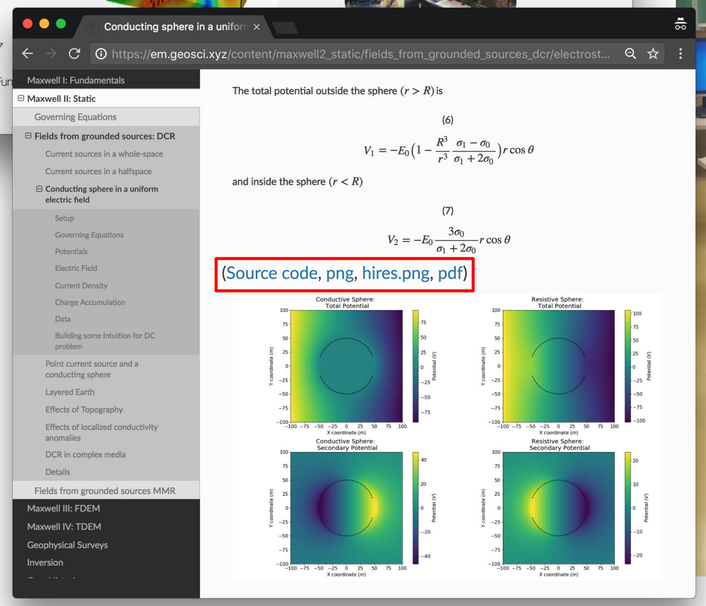
https://geosci.xyz

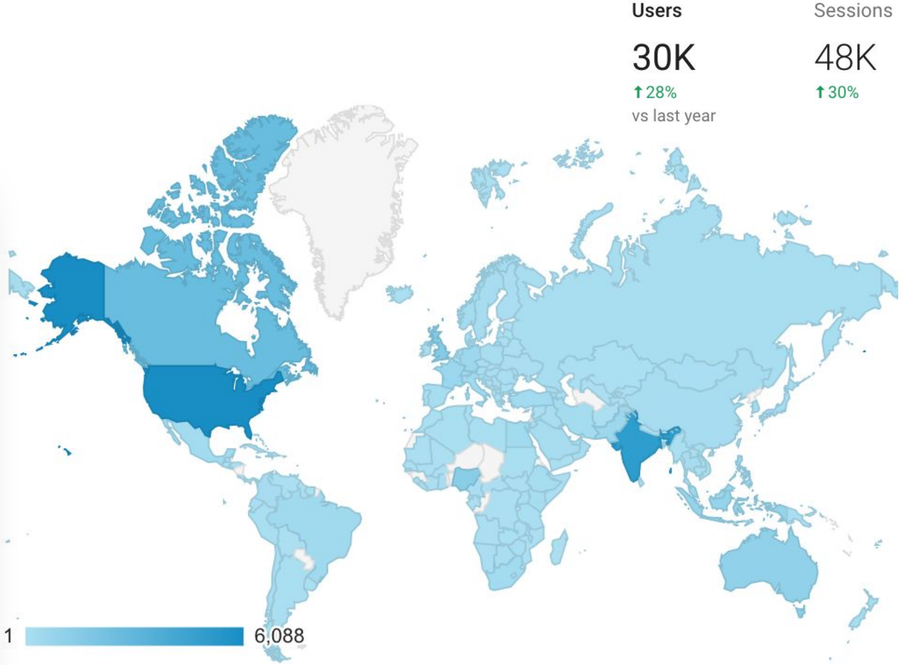
DISC 2017

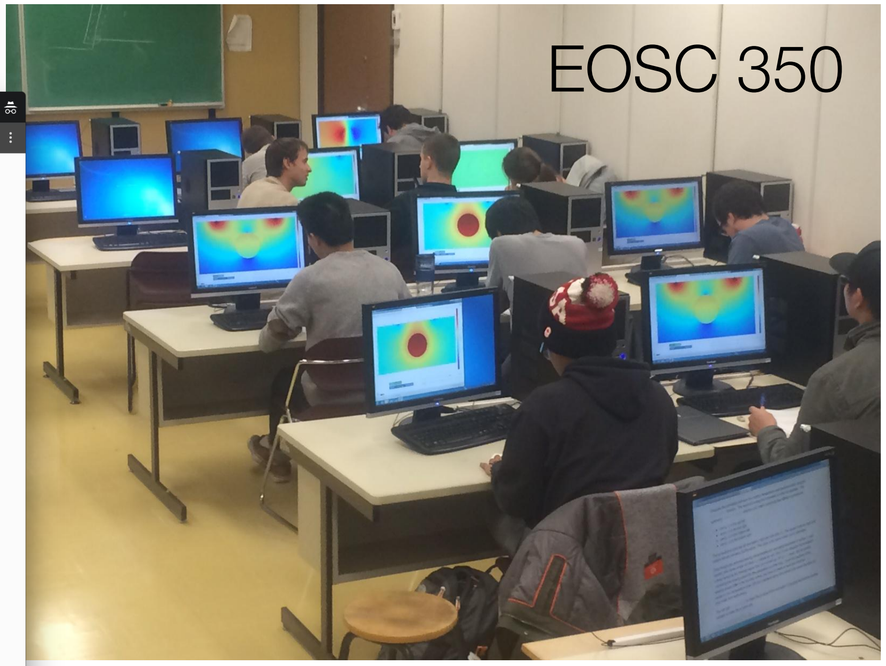
Electromagnetics Fundamentals and Applications

26 locations worldwide









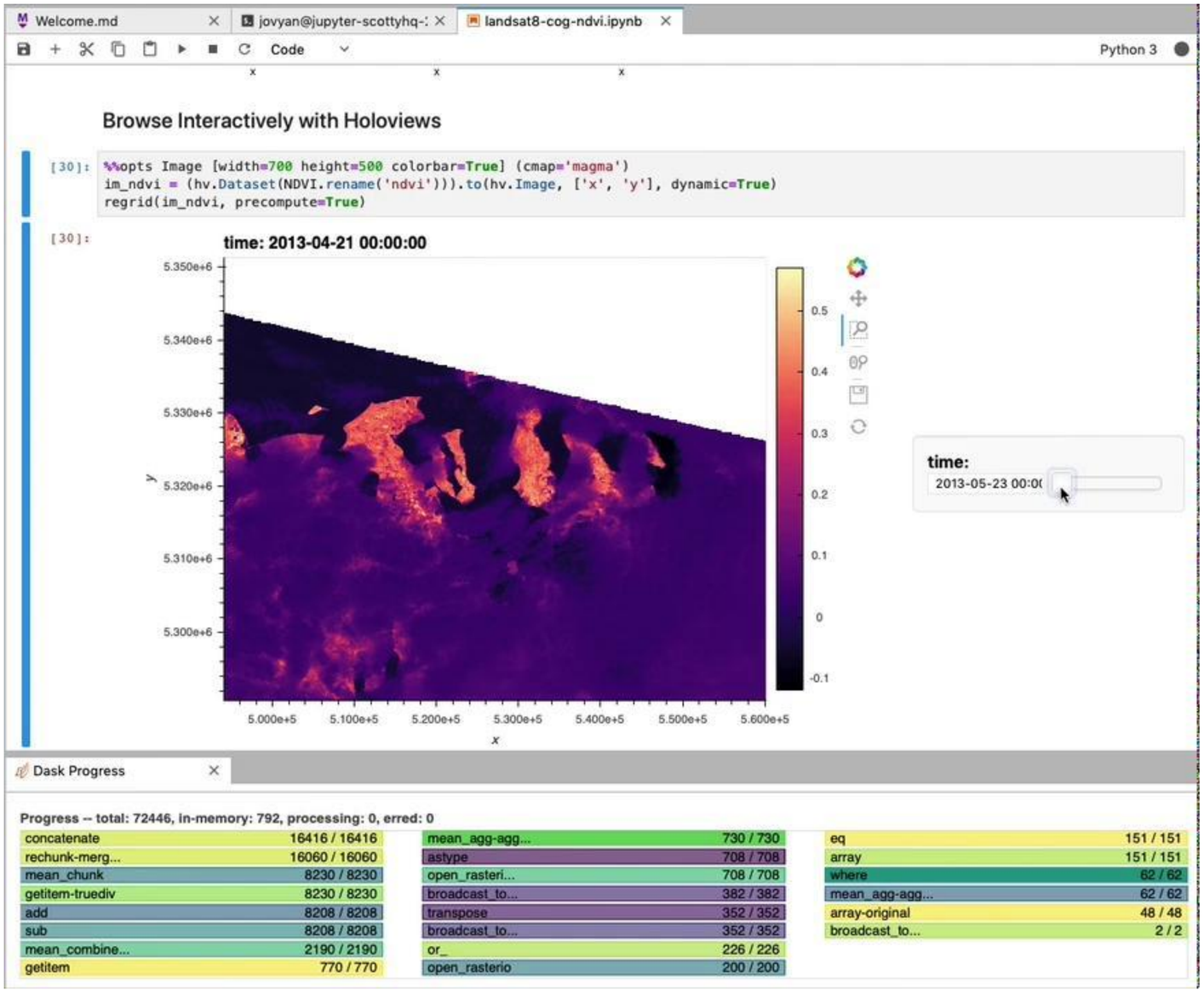
PANGEO

Harnessing the power of cloud computing to study the whole Earth interactively

jupyter — Interactivity

DASK — Distributed computing

xarray — Data models / numerics



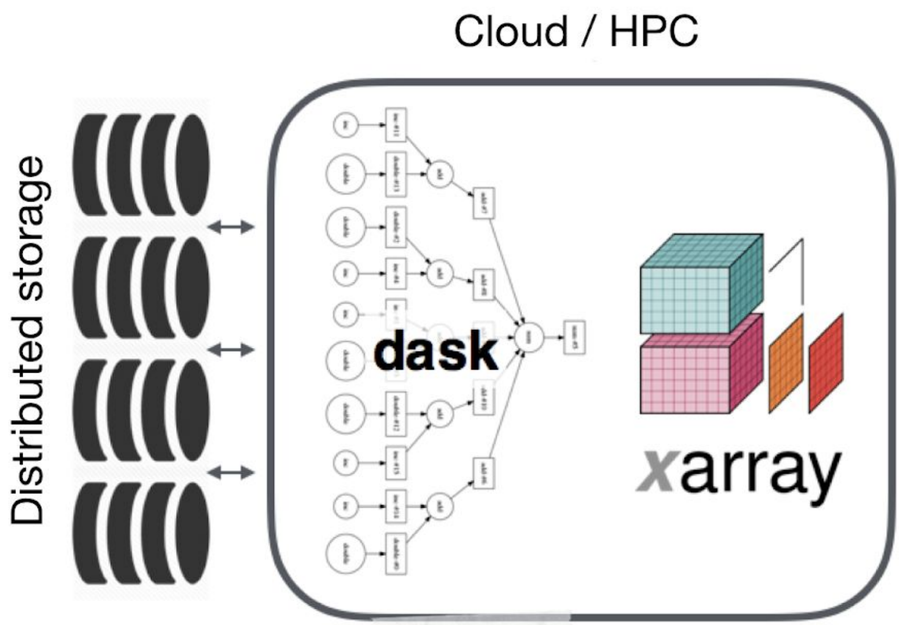
Scott Henderson Follow

Research geophysicist at University of Washington eScience Institute

Oct 1 · 7 min read

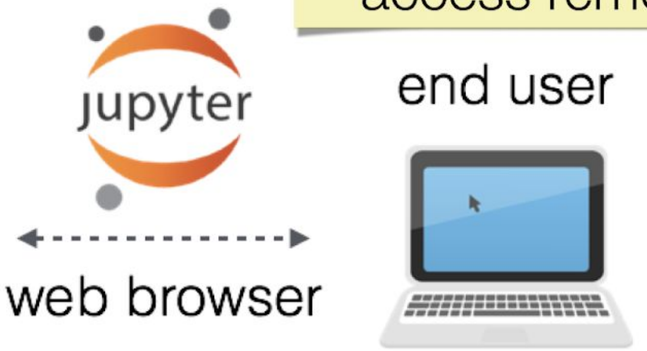
Pangeo Architecture

"Analysis Ready Data" stored on globally-available distributed storage.



Parallel computing system allows users deploy clusters of compute nodes for data processing.

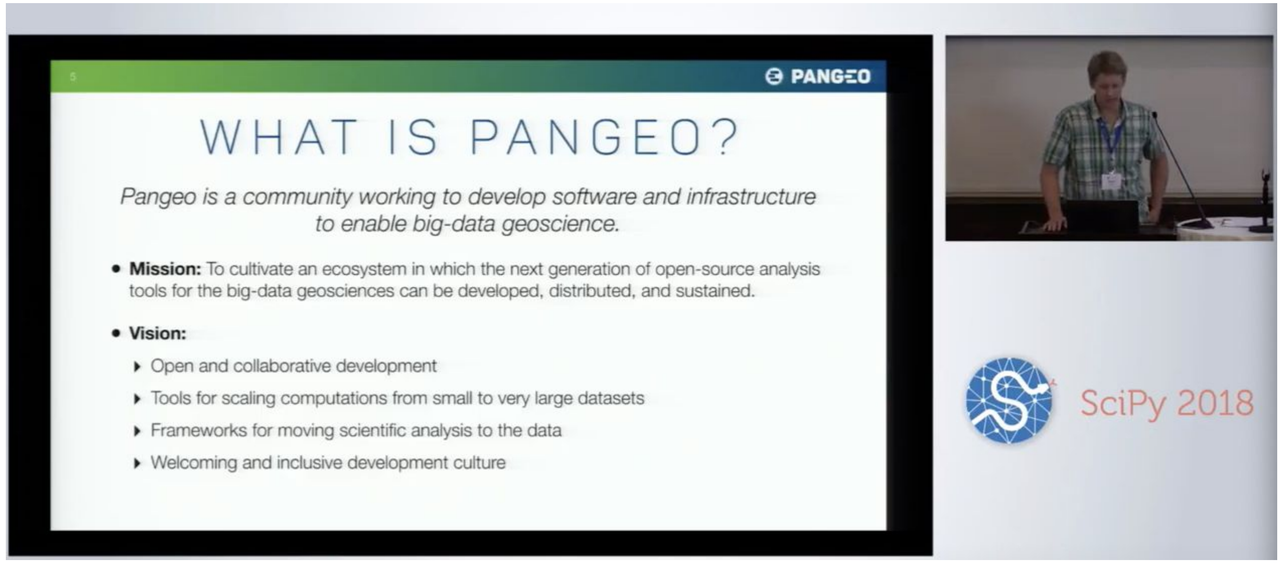
Dask tells the nodes what to do.



Jupyter for interactive access remote systems

Xarray provides data structures and intuitive interface for interacting with datasets

What is Pangeo?



Pangeo: A Big Data Ecosystem for Scalable Earth System Science
Joe Hamman (NCAR), SciPy 2018
https://www.youtube.com/watch?v=2rgD5AJsAbE

The Future of Data-driven Discovery in the Cloud
Ryan Abernathey (Columbia), JupyterCon 2018
https://www.youtube.com/watch?v=7kDYfUe0Zhw

Jupyter meets the Earth: an NSF grant (2M / 3Y)!

Research use-cases

- Climate data analysis
- Hydrologyd
- Geophysics

PANGEO

Tech developments

- Data discovery
- Interactivity
- Cloud/HPC infrastructure

For more: http://bit.ly/jupytearth

Fernando Pérez

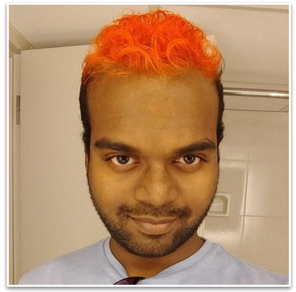
Joe Hamman

Laurel Larsen

Kevin Paul

Lindsey Heagy

Chris Holdgraf

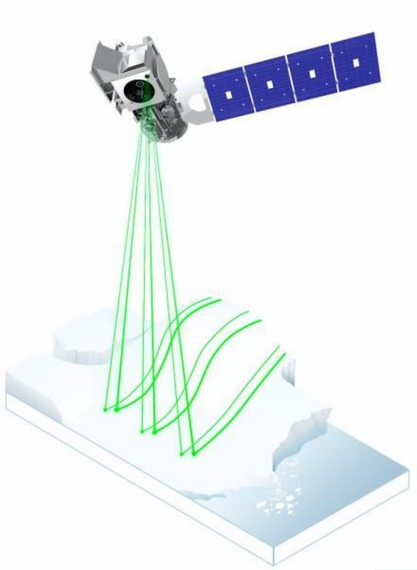


Yuvi Panda

Next-generation LIDAR satellite

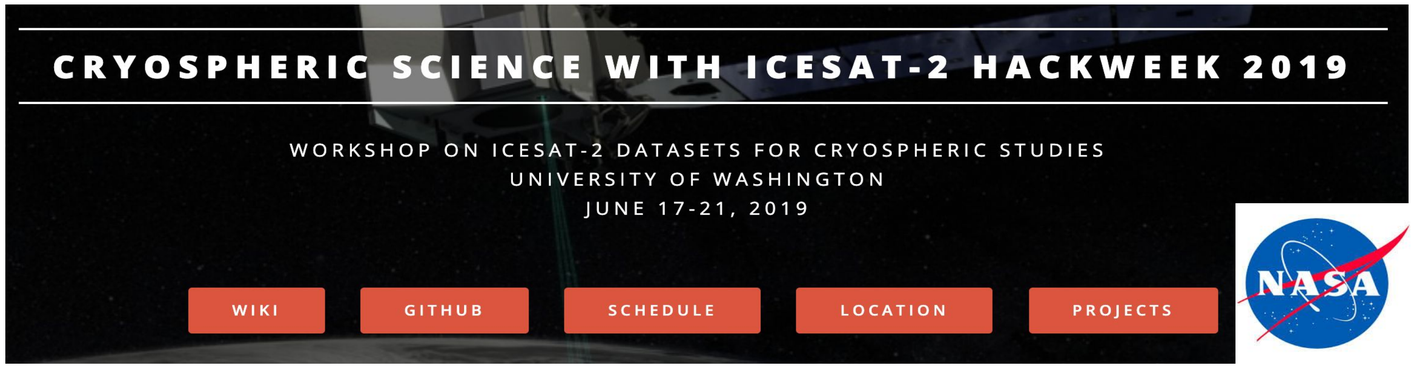
ICESat-2

ICE, CLOUD, AND LAND ELEVATION SATELLITE-2



icepyx

Python tools for
cryospheric sciences



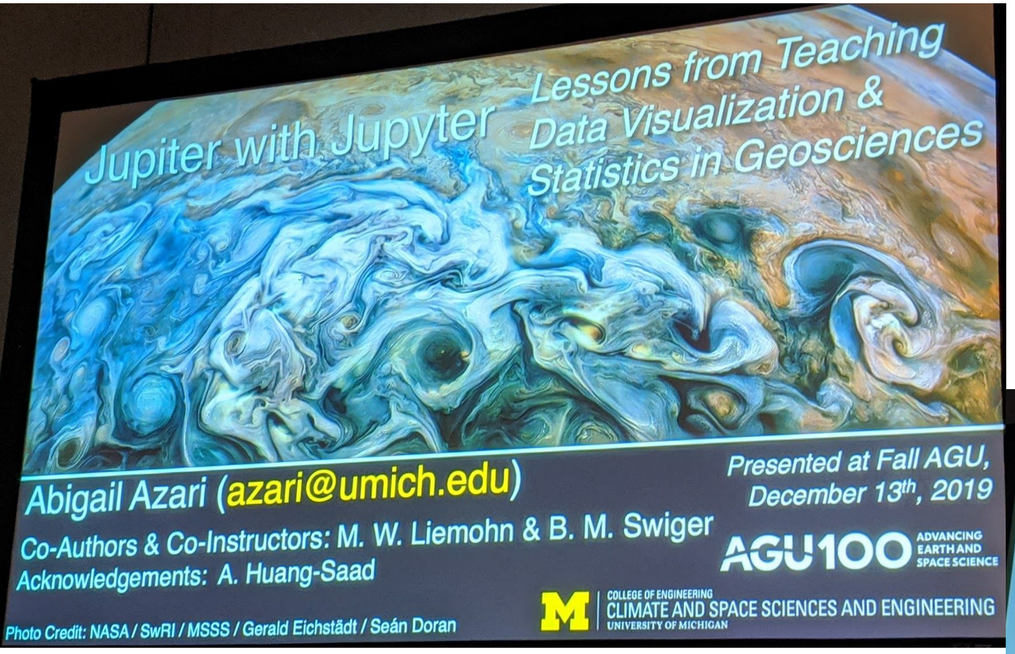
https://icesat-2.gsfc.nasa.gov

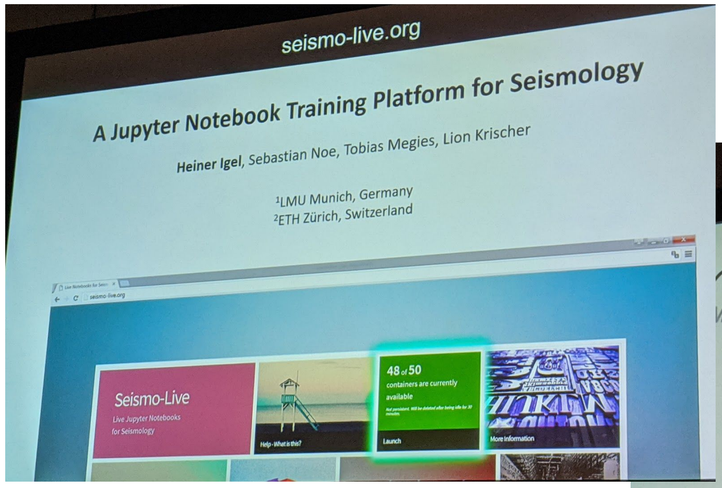
A. Arendt

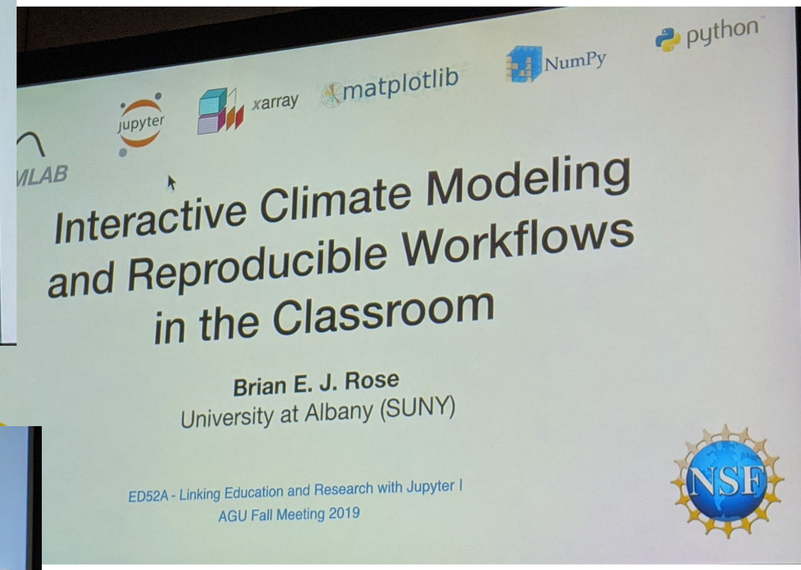
J. Scheik

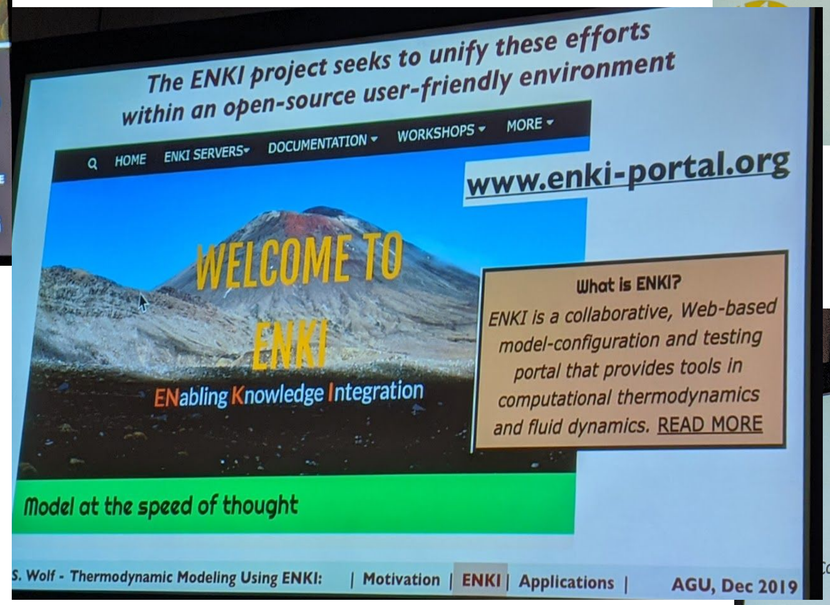
L. Heagy

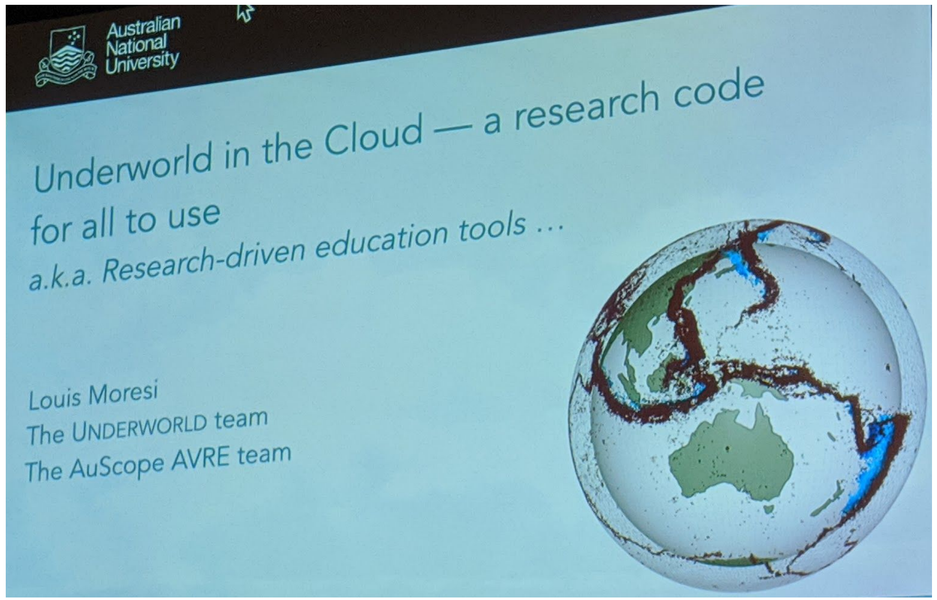
Jupyter at AGU 2019

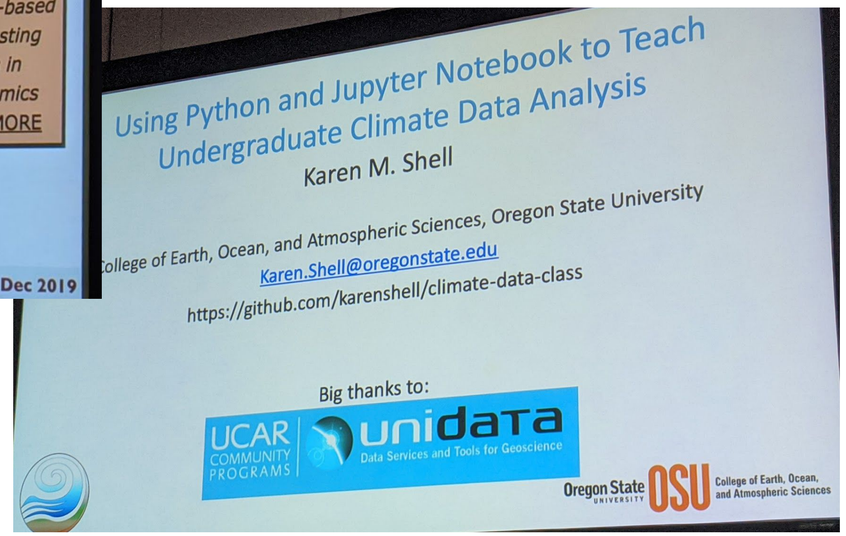












So you want to build Data Science tools in academia…

Jupyter - funding and resources

ALFRED P. SLOAN FOUNDATION

GORDON AND BETTY MOORE FOUNDATION

THE LEONA M. AND HARRY B. HELMSLEY CHARITABLE TRUST

U.S. DEPARTMENT OF ENERGY

CHAN ZUCKERBERG INITIATIVE

SIMONS FOUNDATION

NETFLIX

POWERED BY rackspace® the open cloud company

ANACONDA®

IBM®

Microsoft

Google

Bloomberg

Should I Resign From My Full Professor Job To Work Fulltime On Cocalc?

William Stein • Apr 12, 2019 •

Nearly 3 years ago, I gave a talk at a Harvard mathematics conference announcing that "I am leaving academia to build a company". What I really did is go on *unpaid leave* for three years from my tenured Full Professor position. No further extensions of that leave is possible, so I finally have to decide whether or not to go back to academia or resign.

My unpaid leave is up – what am I going to do?

My third year of unpaid leave from UW is up. I have to decide whether to return to UW or resign. If I return, it turns out that I would have to have at least a 50% appointment. I currently have 50% of one year of teaching in "credits", which means I wouldn't be required to teach for the first year I go back as a 50% appointment. Moreover, the current department chair (John Palmieri) understands and appreciates Sage – he is among the top 10 all time contributors to the source code of Sage!

I have decided to resign. I'm worried about issues of intellectual property; it would be extremely unfair to my employees, investors and customers if I took a 50% UW

Scientific Open Source: *Despite* (direct) federal $$ support

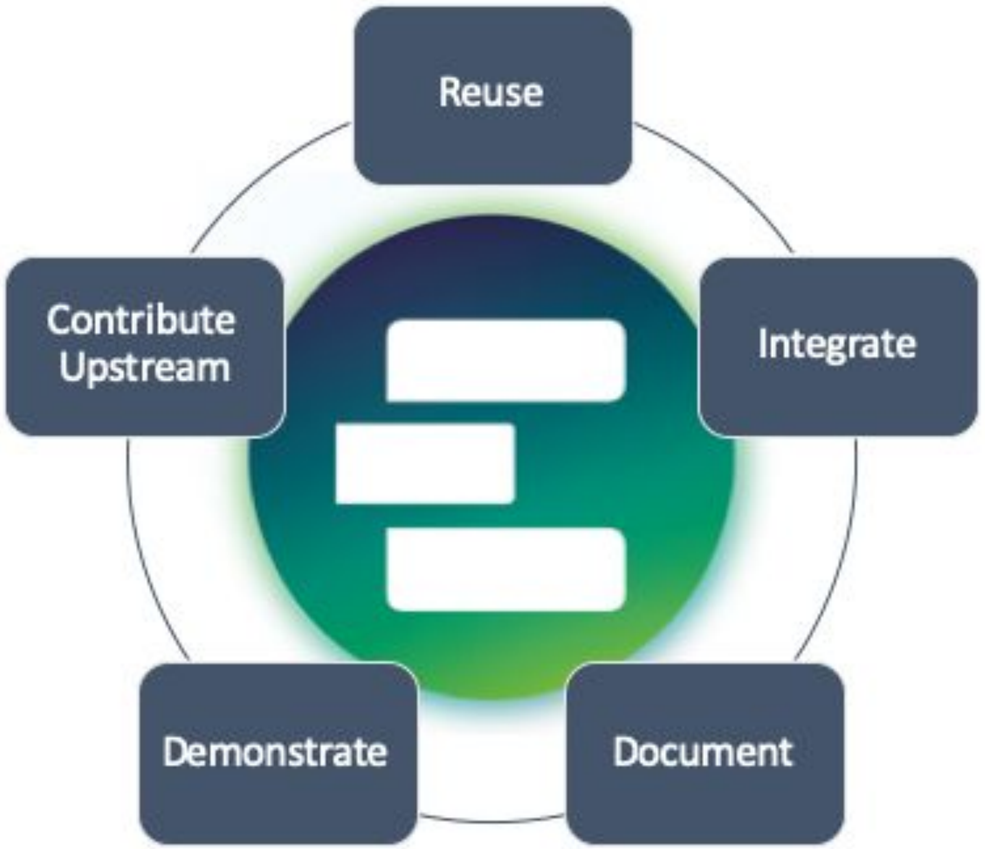
- “Indirectly”, lots of $ have supported Scientific OSS projects/tools.
  - Under the cover of domain-focused work.

Traditional software infrastructure funding

Yes, it's true, the budget is gone again… But you can't deny that now, we get here in an instant!

Quino (Argentinian cartoonist)

The Pangeo Pattern



1. *Reuse:* resist the temptation to build something new
2. *Integrate:* use the modular ecosystem as intended
3. *Document:* explain concrete use cases that impact your community
4. *Demonstrate:* offer public infrastructure (binder.pangeo.io)
5. *Contribute upstream:* help sustain your own ecosystem

https://medium.com/pangeo/the-pangeo-pattern-9a81ca4bad42, Joe Hamman, Nov'19

Challenges: contrasts in cultures and incentives

	Open Source	**Academia**
Credit	Distributed	PI & hierarchy
Output/artifacts	Continuous & Project-specific	Discrete papers
Collaborators	Fluid: professionals, volunteers, …	Structured, funding-dependent
Governance/decision making	Open, community based	Top-down, PI
Authorship	Fluid, roles can evolve, no clear “first/senior” author	Need to say more?
Peer review	Continuous, open, pervasive, friendly	The opposite
Value metric	Utility, need, impact	“Novel and transformative”

Thank you (Bay Area team)

Current (Berkeley, LBNL, Bloomberg)

Stacey Dorton, Lindsey Heagy, Chris Holdgraf, Yuvi Panda, Ryan Lovett, Shreyas Cholia, Shane Canon, Rollin Thomas, Jason Grout

Former Berkeley

Min Ragan-Kelley, Paul Ivanov, Thomas Kluyver, M Pacer, Matthias Bussonnier, Jessica Hamrick, Ian Rose, Jamie Whitacre.

University of California, Berkeley
DEPARTMENT OF STATISTICS

BIDS
BERKELEY INSTITUTE FOR DATA SCIENCE

BERKELEY LAB

NERSC

Bloomberg

Open, collaborative geoscience

- Meet scientific/technical challenges
  - and interdisciplinary ones
- Research/education partnership
- Engage with society/policy makers
- Have a better community experience!

Thank You!

@fperez_org, @lindsey_jh

{fernando.perez, lheagy}@berkeley.edu