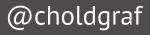




Open Infrastructure for open science How Binder Powers an Open Stack in the Cloud Chris Holdgraf, UC Berkeley and Project Jupyter





andrewosh



betatim



freeman-lab



mbmilligan



sgibson91



~





captainsafia

henchc

mpacer

tgeorgeux



JamiesHQ



parente



willingc



choldgraf





rgbkrk

jhamrick





Ruv7

Zsailer

jzf2101







ellisonbg













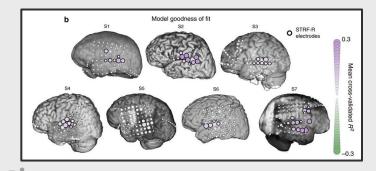
Alfred P. Sloan Foundation

GORDON AND BETTY MOORE FOUNDATION

A bit about me then...

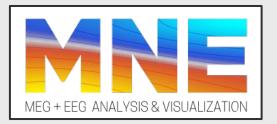
Berkeley Neuroscience

Cognitive Neuroscience





Open Source



@choldgraf

A bit about me now...

Research and Open Source



Education and Open Source





@choldgraf



a *community* of people and an *ecosystem* of open tools and standards for interactive computing



create things that are **language-agnostic** and **modular**. Empower people to use **other open tools**.

Aside: Jupyter and the last mile problem



 $\dot{}$

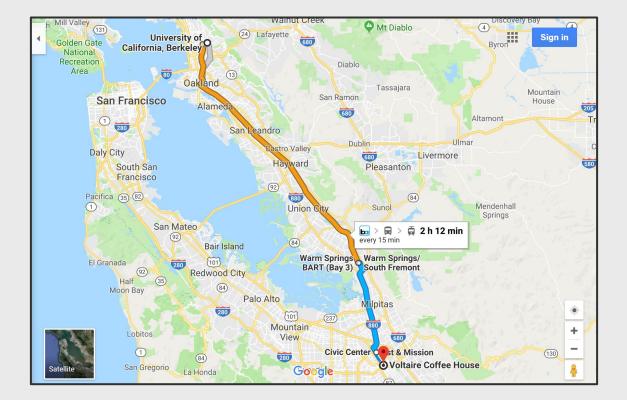


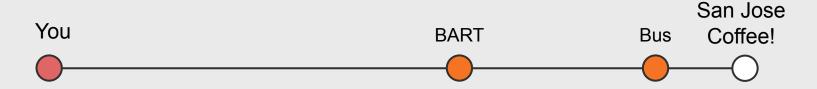
One option: walk there by myself

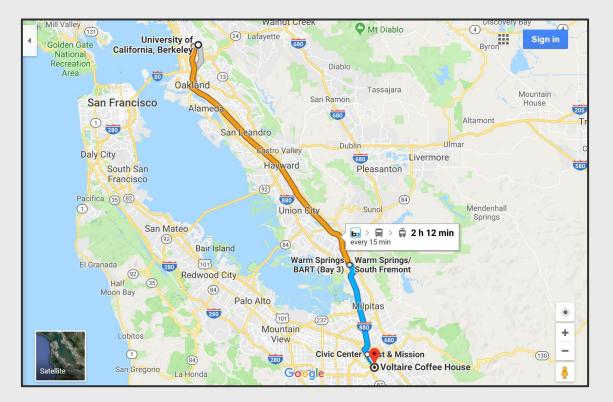
Another option: pay somebody to drive me

My favorite option: use public infrastructure

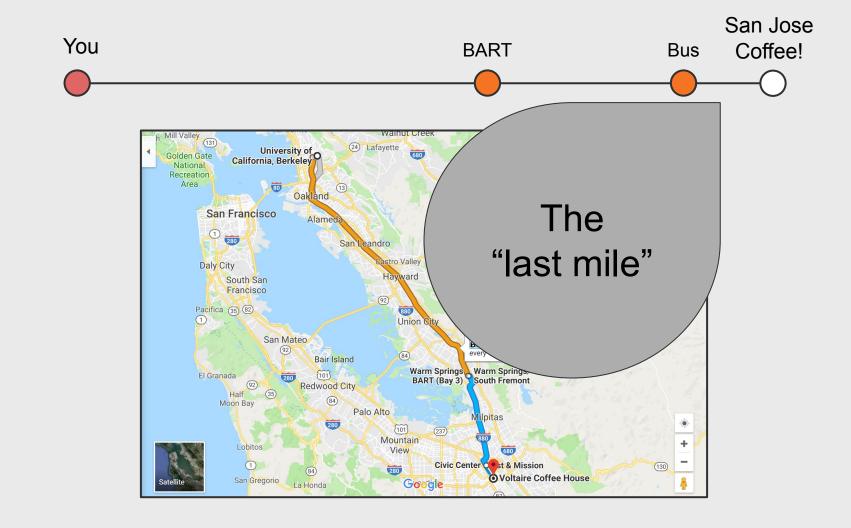
You San Jose Coffee!







~



~

Public infrastructure gets us closer to our goal.

It makes the last mile shorter.

How does **Jupyter** fit in to this?



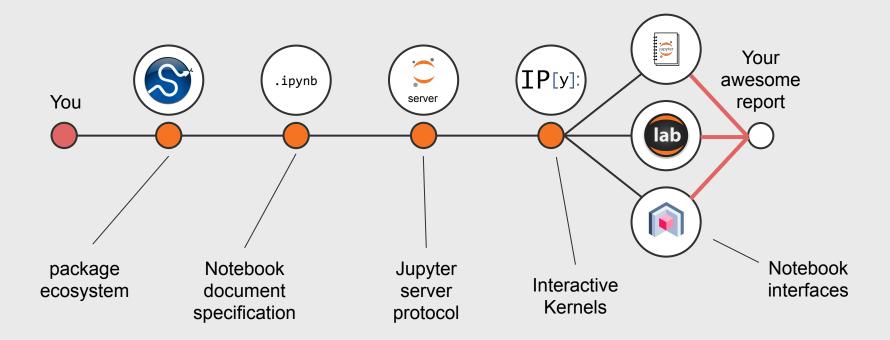


One option: build it from scratch

Another option: pay somebody for a product

My favorite option: build on open, modular tools

Jupyter shortens the last mile by creating and leveraging public infrastructure



Back to our talk...

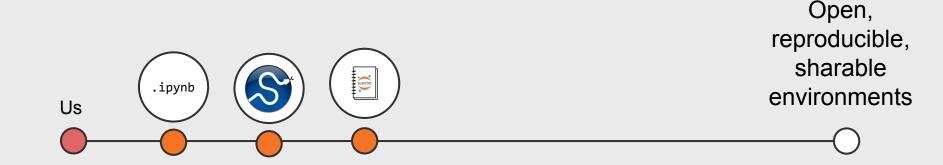
~

The science is the code

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 John Claerbout) WaveLab and Reproducible Research, 1995

Our mission for this talk



 $\dot{\mathbf{C}}$

Part 1: from your laptop to the cloud with JupyterHub

(some) data science should be taught to everyone (no, really)

Here's what this means at Berkeley...



How can we connect people with computation?

What is JupyterHub? Host pre-configured data science environments on shared infrastructure

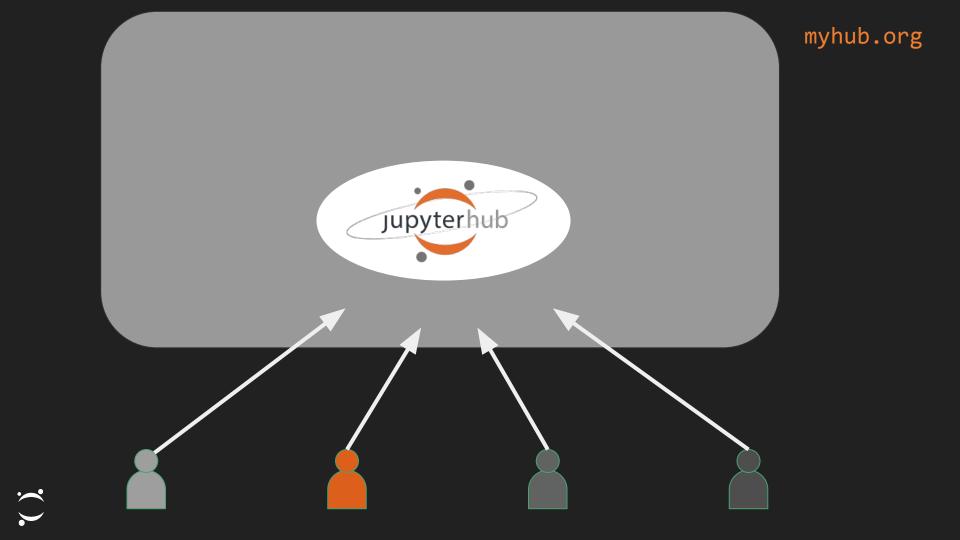


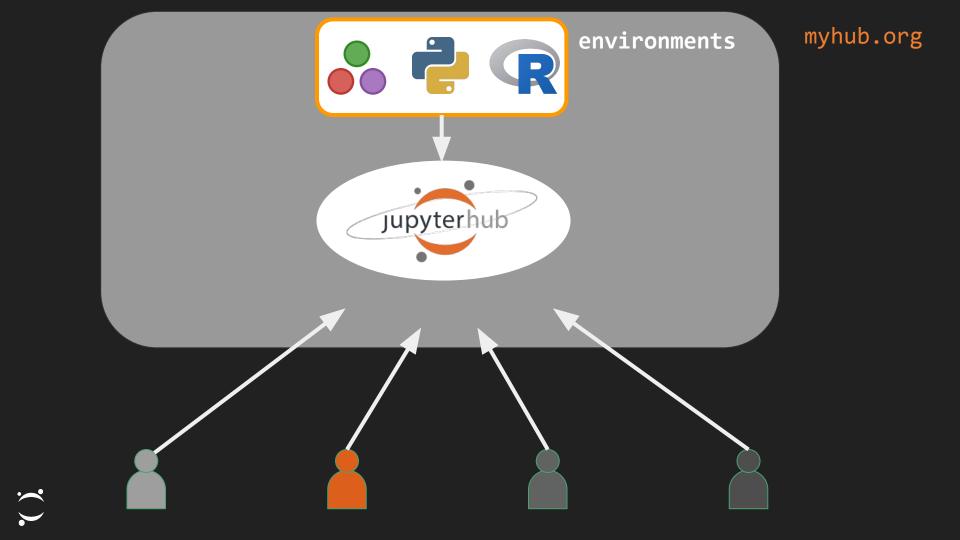
jupyter.org/hub

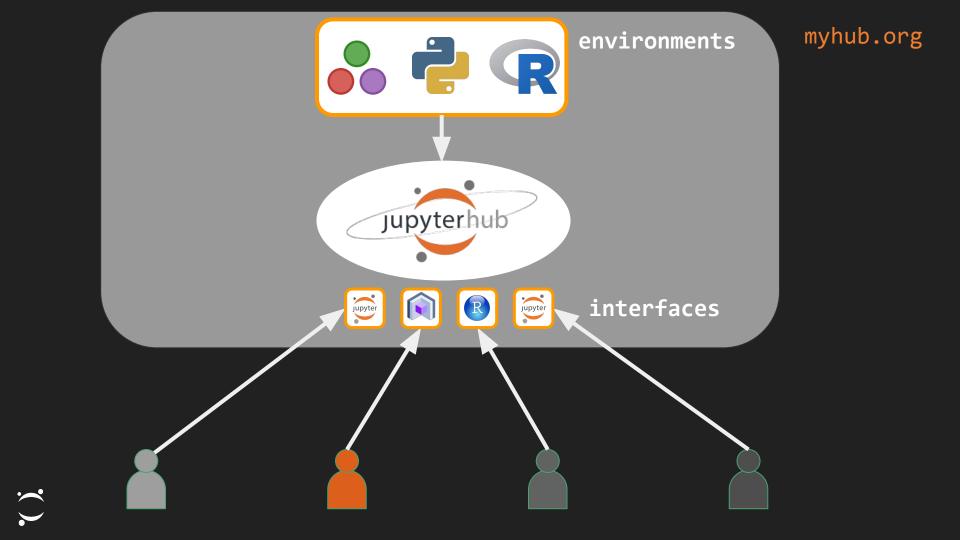
myhub.org

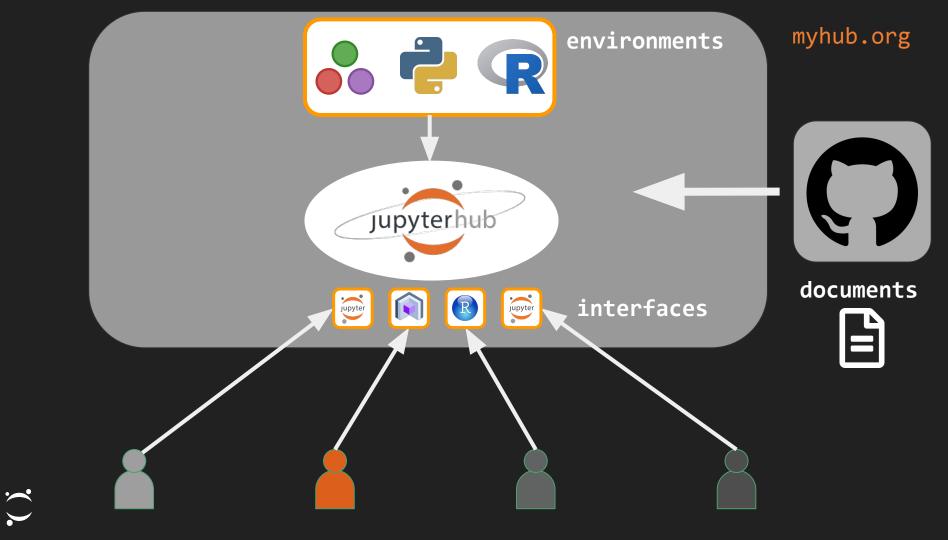
My fancy machine in the cloud

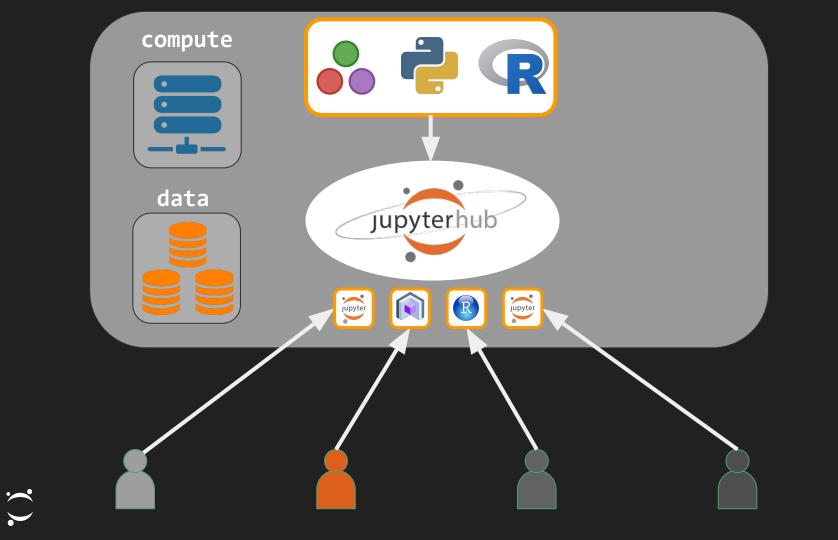


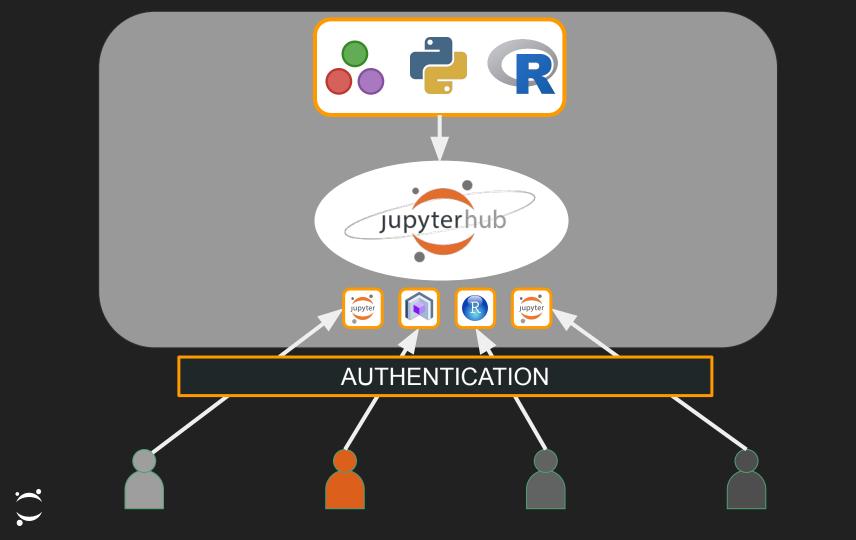












Chris Is Trying A Live Demo

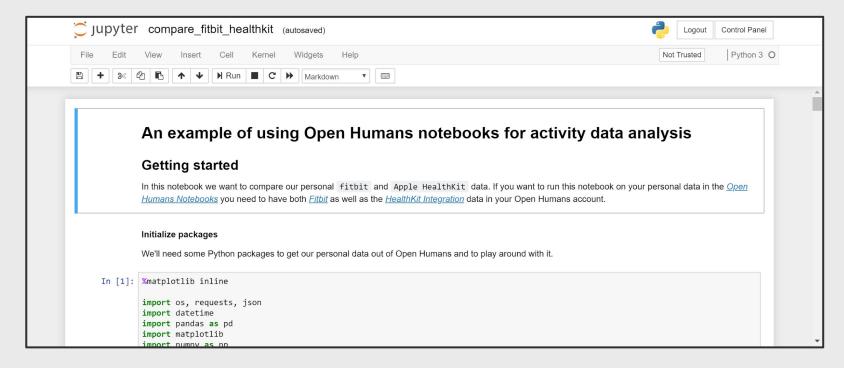
Hopefully he doesn't embarrass himself too badly.

* Press ESC to cancel and return to Windows.

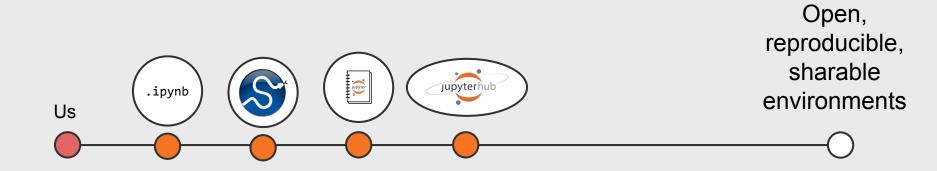
- * Press ENTER to close this application that is not responding. You will lose any unsaved information in this application.
- * Press CTRL+ALT+DEL again to restart your computer. You will lose any unsaved information in all applications.

Textbook link Interact link





notebooks.openhumans.org



How can users package+share their work?

Part 2: packaging and sharing your environment with repo2docker

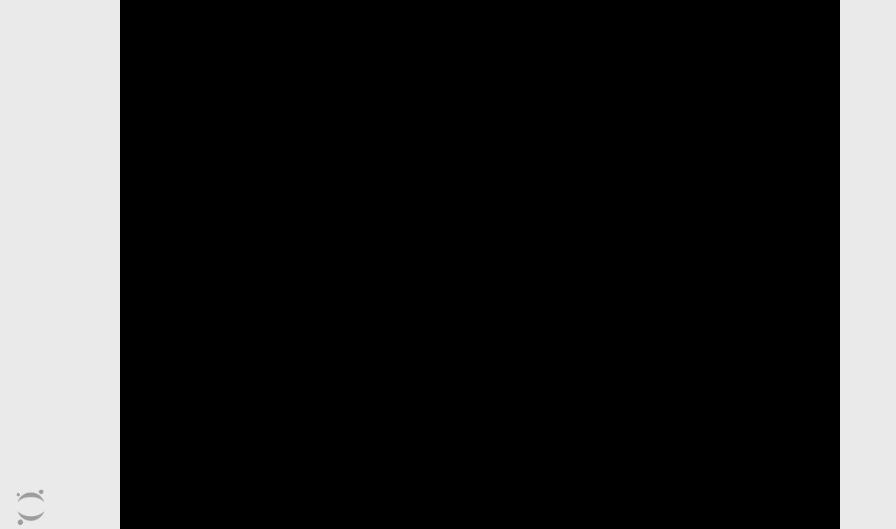
What is repo2docker? Convert a repository into a Docker image that runs the code inside.



repo2docker.readthedocs.io

github.com/minrk/ligo-binder

~



$jupyter repo2docker \$

>

https://github.com/minrk/ligo-binder

$\scriptstyle inter \ in$

> https://github.com/minrk/ligo-binder

```
Cloning into
'/var/folders/.../T/repo2dockermu6z66sd'...
Using CondaBuildPack builder
Step 1/31 : FROM buildpack-deps:bionic
 ---> 29f4eef41002
Step 2/31 : ENV DEBIAN_FRONTEND=noninteractive
 ---> Using cache
 ---> ee1ba7c4f5f4
Step 3/31 : RUN apt-get update && apt-get
install --yes --no-install-recommends locales &&
apt-get purge && apt-get clean && rm -rf
/var/lib/apt/lists/*
```

Step 1: get the repo

git clone https://github.com/me/myproject

Step 2: Identify requirements

scholdgraf Update README.md	
README.md	Update README.md
index.ipynb	first move
E requirements.txt	Pin requirements.txt to values that actually exist

Step 2: Identify requirements

💡 yuvipanda s/Anaconda/conda/ 🚥	
E README.md	s/Anaconda/conda/
environment.yml	Require a stable install of dask
index.ipynb	updating to new syntax

Step 2: Identify requirements

betatim Update README.md	
bus-dashboard	Remove the DESCRIPTION file
README.md	Update README.md
index.ipynb	adding example
install.R	Add example Shiny app
🖹 runtime.txt	Remove dockerfile, add runtime.txt

Step 3: generate **Dockerfile**

setup the runtime environment

COPY conda/install-miniconda.bash /tmp/install-miniconda.bash COPY conda/environment.py-3.6.frozen.yml /tmp/environment.yml RUN bash /tmp/install-miniconda.bash && \ rm /tmp/install-miniconda.bash /tmp/environment.yml

• • •

. . .

Step 3: generate **Dockerfile**

assemble the environment for the repo

```
# Copy and chown stuff.
COPY src/ ${HOME}
RUN chown -R ${NB_USER}:${NB_USER} ${HOME}
```

Run assemble scripts! These will actually build the spec # in the repository into the image. USER \${NB_USER} RUN \${KERNEL_PYTHON_PREFIX}/bin/pip install --no-cache-dir \ -r "requirements.txt"

 $\mathbf{\dot{}}$

Step 4: build (& push) image

docker build -t myimage
docker push myimage

https://github.com/minrk/ligo-binder Cloning into '/var/folders/.../T/repo2dockermu6z66sd'... Using CondaBuildPack builder Step 1/31 : FROM buildpack-deps:bionic ---> 29f4eef41002 Step 2/31 : ENV DEBIAN_FRONTEND=noninteractive ---> Using cache ---> ee1ba7c4f5f4 Step 3/31 : RUN apt-get update && apt-get install --yes --no-install-recommends locales && apt-get purge && apt-get clean && rm -rf /var/lib/apt/lists/*

. . . .

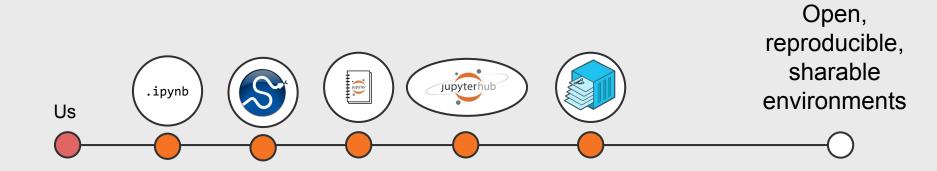
\$ jupyter repo2docker \

Some supported configuration files

- environment.yml conda
- requirements.txt pip
- **REQUIRE** julia
- install.R R
- apt.txt apt
- setup.py pip
- postBuild shell
- runtime.txt
- Dockerfile docker
- yournewbuildpack.txt

Guiding principles of repo2docker

- Repos should be **human** and **machine readable**
- Use existing specifications and standards
- Support many languages and interfaces
- Be lightweight and tightly-scoped, but extendable

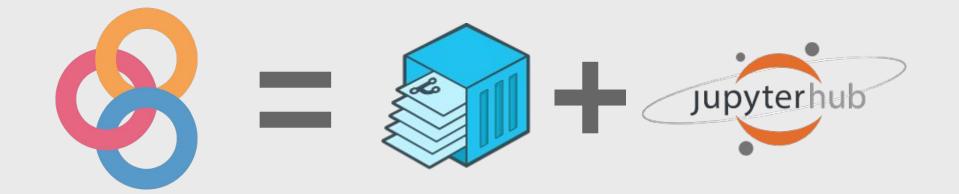


Part 3: tying this together with BinderHub

What is BinderHub? One-click sharable, interactive, reproducible environments from your public git repository



mybinder.org
binderhub.readthedocs.io



 \sim

BinderHub is open tech...

- Built on Kubernetes
- Cloud-agnostic
- Scalable
- Community driven and deployable by anyone

One example: mybinder.org



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.

GitHub repo or URL			
GitHub repository name or link			1
Git branch, tag, or commit	Path to a notebook file (optional)		
master	Path to a notebook file (optional)	File 🕶	
		File 🕶	launch
Copy the URL below and share your	r Binder with others:		
Fill in the fields to see a URL for sharing your Binder.			Ê

Chris Is Trying A Live Demo

Hopefully he doesn't embarrass himself too badly.

- * Press ESC to cancel and return to Windows.
- * Press ENTER to close this application that is not responding. You will lose any unsaved information in this application.
- * Press CTRL+ALT+DEL again to restart your computer. You will lose any unsaved information in all applications.

mybinder.org

binder-examples/requirements





HOME ARCHIVE EVENTS JUPYTER WEBSITE

Project Jupyter Follow Project Jupyter exists to develop open-source software, open standards, and services for interactive and reproducible computing. Feb 20 · 5 min read

JupyterLab is Ready for Users

We are proud to announce the beta release series of JupyterLab, the nextgeneration web-based interface for <u>Project Jupyter</u>.

tl;dr: JupyterLab is ready for daily use (<u>installation</u>, <u>documentation</u>, <u>try it</u> * <u>with Binder</u>)



Mybinder.org is an open service...

Mybinder.org activity and status:

grafana.mybinder.org/

Our billing: github.com/jupyterhub/binder-billing

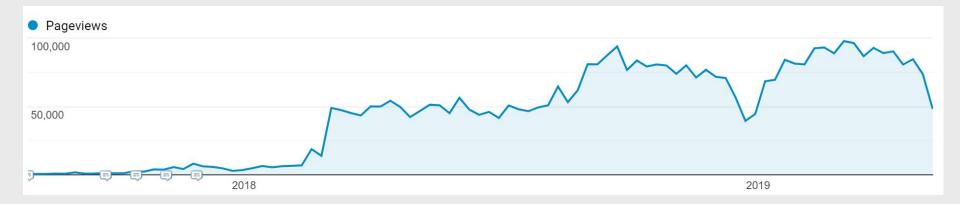
Our operations guide:

mybinder-sre.readthedocs.io/

Our fires:

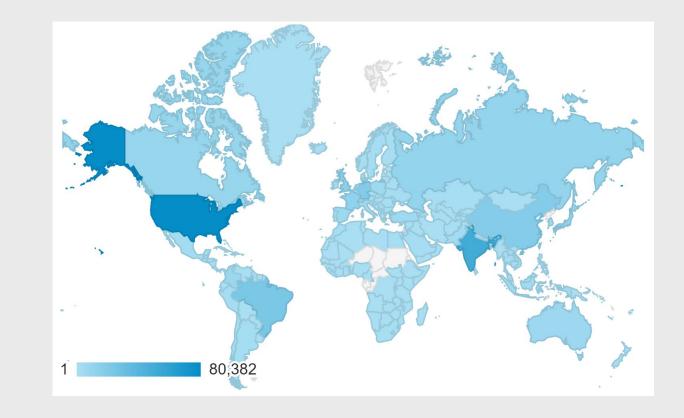
mybinder-sre.readthedocs.io/en/latest/#incident-reports

mybinder.org weekly sessions, last ~year

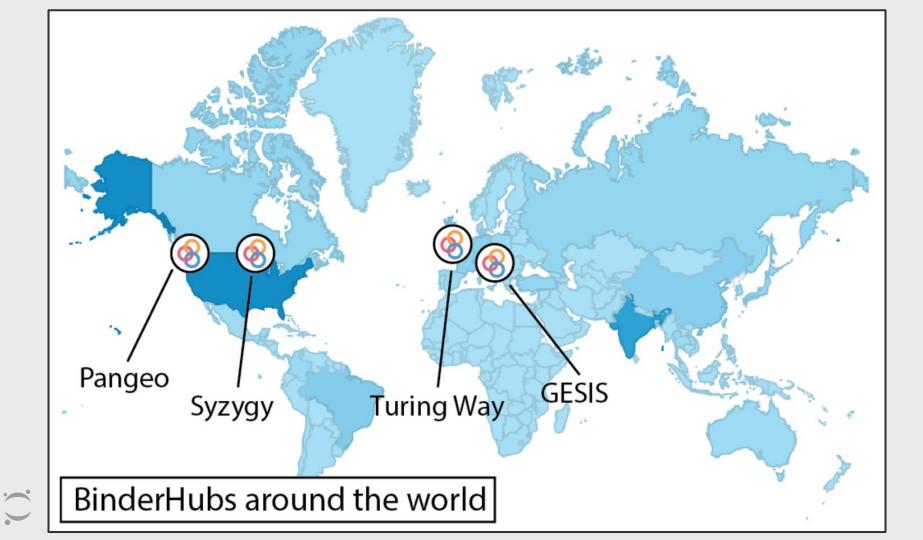


 $\dot{}$

mybinder.org sessions, last month



 $\dot{}$



Some cool new projects (or, stuff you can help out with)

Interactive books with

github.com/jupyter/jupyter-book



Jupyter Book

Home

Search

Getting started

Create your book

Build and publish your book

FAQ

How-to and advanced topics

Features and customization

Markdown files

← TOGGLE SIDEBAR



Interactive code in your book

Sometimes you'd rather let people interact with code *directly on the page* instead of sending them off to a Binder or a JupyterHub. There are currently a few ways to make this happen in Jupyter Book (both of which are experimental).

This page describes how to bring interactivity to your book. Both of these tools use **MyBinder** to provide a remote kernel.

Making your page inputs interactive

🛠 experimental 🛠

If you'd like to provide interactivity for your content without making your readers leave the Jupyter Book site, you can use a project called Thebelab. ■ ON THIS PAGE

MAKING YOUR PAGE INPUTS INTERACTIVE

USING INTERACTIVE WIDGETS ON YOUR PAGE

Publishable documents (with pandoc?)

AEM2018/7th International Workshop on Airborne Electromagnetics

Open source software for simulations and inversions of airborne electromagnetic data

Lindsey J. Heagy Seogi Kang University of British Colu University of Britist Geophysical Inventon Facility Vancouver, BC, Canada Geophysical Inversion Facility Vancouver, BC, Canada /heagy@eos.ubc.ca skang@eos ubc.ca SUMMARY

Inversions of airborne EM data are often an iterative

process, not only requiring that the researcher be able to explore the impact of changing components such as the choice of regularization functional or mode

arameterization, but also often recraiting that forward

simulations be run and fields and fluxes visualized in order to build an understanding of the physical processe

owening what we observe in the data. In the hone of

leveloped the open source so fware package. SimPEG

The software has been designed to be modular and extensible with the goal of allowing researchen to interrogate all of the components and to facilitate the

exploration of new inversion strategies. We present as

werview of the software in its application to airborn EIN and demonstrate its use for visual ring fields and flaxes it forward simulation as well as its flexibility it

regulating and solving the inverse problem. We invest a

parametric inversion, where all of the forward model line

is done on a 3D grid. The results in this paper can b

reproduced using the provided Jupyter notebooks. The Pethon software can also be modified to allow users to

speciment with naturates and earlose the physics of the

Key words: firspancy domain, time domain, inversion

INTRODUCTION & MOTIVATION

Even for a simple model of a conductive vertical plate in a

resistive half-space, the contributions of the diffusing smoke

ring currents which intersect the plate, and vortex currents

induced by the time-surving magnetic flux through the plate to the observed airborne EM signal can be challenging to unavel.

These challeness are exception if the background has

Dissecting and understanding the physical phenomena that

produce the data which we observe in an airborne EM survey requires that we have the ability to simulate Maxwell's equations and visualize currents and magnetic fields through

time or over a range of freesancies for a variety of models, which might be 1D, 2D, or 3D in space.

Beyond building a fundamental understanding of the physics

governing an airborne EM response, extracting meanineful

have the ability to invert these data. EM inversions are nonlinear and thus the choice of regularization and selection of

AEM2018 - June 17-20, Kolding, Denmart

coloric information from airborne EM data monitors that we

orien source software. finite volume

Rowan Cockett Douglas W. Oldenburg University of British Co. Geophysical Inversion Facility over, BC, Canada rowane Sill gmail.com dougineos ubcca

tuning parameters can have a significant impact on the recovered model. To progress research into extracting more geologic or hydrogeologic information from airborne in an inversion, researchers require the flexibility to experiment with model (e.g. voxel or a parametric model). Looking forward, we also want the ability to interface to other geophysical methods through joint or cooperative inversions.

The classic model of black-box, mornistane software does not the overhead for researchers in going from a new idea to an source Famework and software implementation for Sinvalation and Pammeter Estimation in Geophysics, SimPEG (Cockett et al. 2015). SimJEG includes finite volume simulations and increasion routines for a variety of geophysical applications including Potential Fields, Vadore Zone flow, DC Resistivity, and Electromagnetics. Simulations may be performed or sevent different mesh types, including cylindrically symmetric teshes, 3D tensor meshes, and OcTree meshes. The fields and fluxes computed everywhere in the simulation domain are readily accessible so that they can be easily visualized and explored, particularly when used in the interactive Japor omputing environment (Perez and Granger, 2015). Such simulations and visualizations have proved valuable in the context of geoscience education (Oldenburg et al., 2017) and an be a useful tool for understanding the physical processes that contribute to the data we observe.

Heney et al. (2017) pupyide a complete overview of the structure and intolementation of electromagnetic simulations and inversion within SimPEG. In brief, SimPEG includes staggered-grid, finite volume solutions to the quasi-static Maxwell's exuations in both the time domain and the frequency domain on cylindrically symmetric, 3D tensor meshes, and OcTwe meshes. Variable electric conductivity and magnetic nemeability may be considered. Kang and Oldenburg (2016) have recently extended the implementation to account for

We take a deterministic, gradient-based approach to the inverse problem and consider a Tikhonov style inversion, where the different model parameterizations to be employed. For example, a 2D inversion can be performed for a single line of airborne EM data while the forward simulation is conducted on a 3D mesh, or, similarly, a parametric model (e.g. a plate in a half snace earth) may be considered. SimPEG has been built in a flexible, modular manner, to enable researchers to experiment with components such as which norm is employed in the

Open source software for AEM

SimPEG is implemented in Pethon and is licensed under the permissive MIT license which allows commercial and academic use and adaptation of the software (<u>https://cimnee.vcz</u>); we hope this maximizes the utility of the



the measured data. From there, we will demonstrate a 1D inversion of TDEM data collected over the plate. Our final example will consist of two 2D inversions where the forward simulations are carried out in 3D; the first inversion is for a voxel-based model and the second is for a parametric model. We will conclude with a discussion on extensions of this example, including our vision of how to move forward as we look to address large 3D aithorne IM inversions using the 3imPEG framework.

NUMERICAL EXAMPLES

In the examples that follow, we will use a model of a orductive, vertical plate in a resistive half-space. This is a conductive, vertical parte in a resistive nan-space. This is a problem of relevance in mineral exploration, brings to light a number of fundamental aspects of electromagnetics, and is a challenging inverse reoblem, particularly as 1D inversion annot explain the 3D nature of the model. We will look at the behaviour of currents through time and demonstrate the multi dimensional nature of the EM responses and then conduct 1D and 2D inventions of synfactic airborne TDEM data over the plate. The examples have purposefully been key computationally lightweight so that they can be run in a moderate amount of time on a modern laptor. The source cos for each of these examples is available as Jup vter notebooks a

Airborne TDEM response over a conductive plate The model and simulation mesh we consider is shown in Fig.1. below the surface and the background conductivity is 10⁻¹ S in.

One line of airborne TDEM data is computed along v=0m; the source waveform is a step-off waveform, and vertical db/dr data are collected at 2.1 time-channels between 0.05 and 2.5ms. The simulation is performed in 3D on a tensor mesh.



Firme I. Booth allow doffs and even section triabil through the model of a conductive plate #1.1 S/m) in a resistive half-space (H* The combatic data apported in this simulation are shown in Eiseach peofile line indicates ab , ide at a given time channel. A early times, we see a single peak anomale over the plate and, as

time progresses, a double-park character emerges. To es, we plot the currents in the subsurf through time in Fig 3 and the magnetic flux density in Fig 4.

AEM2018 - June 17-20, Kolding, Denmark

Heady, Kang, Cockett, and Oldenburg



Finance 3. Vertical divid data ever the conductive plate. The red data





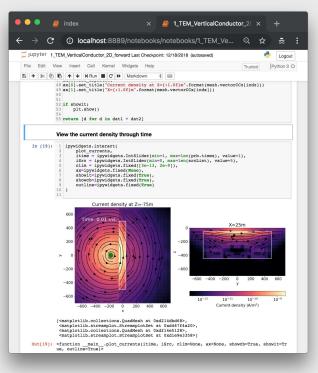


Finner 1 Booth alices doth) and communities (sinks) decoder the current demits in the subsurface at 0.41 ms, 0.1ms and 0.5 ms after shat off. The surve lacetion is shown by the areas dot.

Immediately after the source-current has been shot off, smaller ring currents are induced in the sarth. These horizontal, circular currents reoduce a magnetic field similar to that of a vertical nametic dipole, as seen in Fig 4a. When we are directly over the plate, which has a 100m width, the higher conductivity of the plate results in larger currents and thus a larger magnetifield. This is what gives us the single peak at early time in Fig. 2. The conductivity and thickness of the plate will control the time constant of the decay of these horizontal, circular currents.

At larger offsets, the plate still contributes to a larger signal as the smoke ring currents are channeled into the plate, as can be -0.01ms and t=0.1ms in Fig 3. The excessive currs are more like an electric dipole along the top of the plate, and this gives a magnetic field that is also pointing up at the receiv generated by the channeled currents.

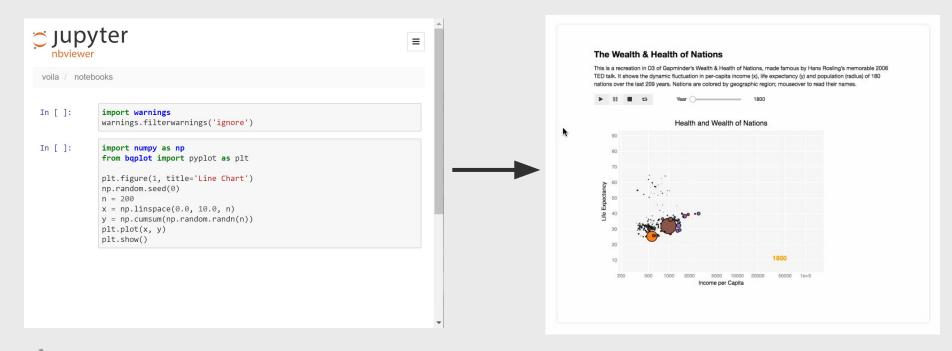
At intermediate times (0.1 ms) the background field has set up strong vortex currents within the plate. By the time that the vortex currents are fully engaged, the galvanic currents are decaying, so at later times (0.5ms) the vortex currents dominate. The votex currents effectively generate a horizontal magneti-



https://github.com/simpeg-research/heagy-2018-aem

HTML dashboards with voila

github.com/QuantStack/voila



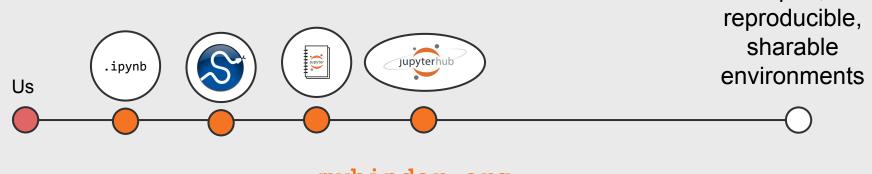
Jupyter makes the "last-mile" problem as small as possible by building modular, open tools.

.ipynb

Open, reproducible, sharable environments

jupyter.org

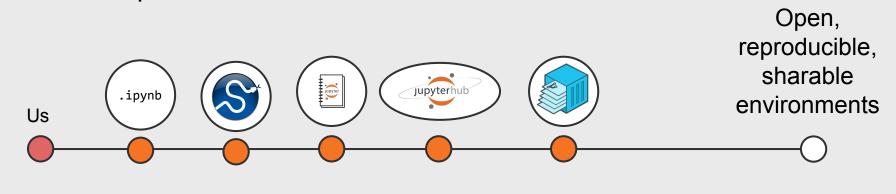
JupyterHub lets you create a shared, interactive analytics environment



Open,

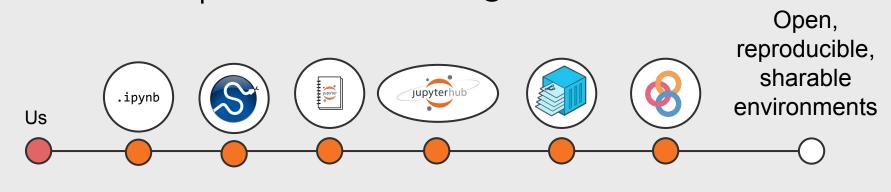
mybinder.org

repo2docker creates reproducible Docker images from a repository



mybinder.org

BinderHub is an open web application to create shareable, reproducible coding environments



Mybinder.org

Get involved with Jupyter

- All of these projects are open source, run by open communities
- Jupyter is a place where *anybody* can participate
- If you'd like to get involved:

```
jupyterhub-team-compass.readthedocs.io
    discourse.jupyter.org
```

@choldgraf