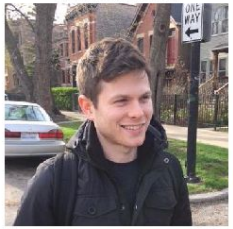
Sharing reproducible computations on Binder

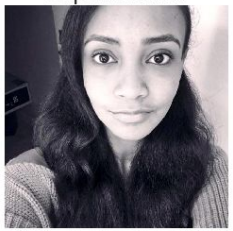
Lindsey Heagy, UC Berkeley & Project Jupyter

andrewosh



betatim

captainsafia



Carreau

choldgraf

consideRatio

ellisonbg

fperez

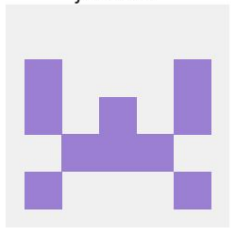
freeman-lab

henchc

JamiesHQ

jhamrick

jzf2101



lheagy

mbmilligan

minrk

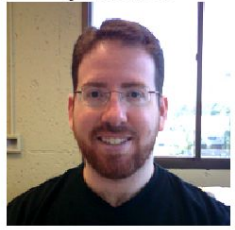
mpacer

parente

rgbkrk

Ruv7

ryanlovett



sgibson91

takluyver

tgeorgeux

willingc

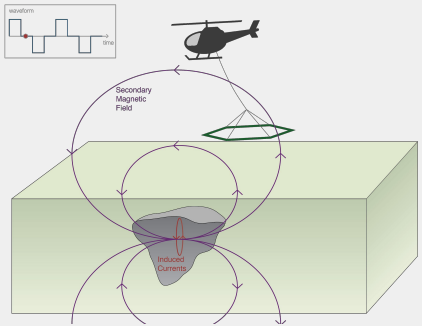
yuvipanda

Zsailer

you?

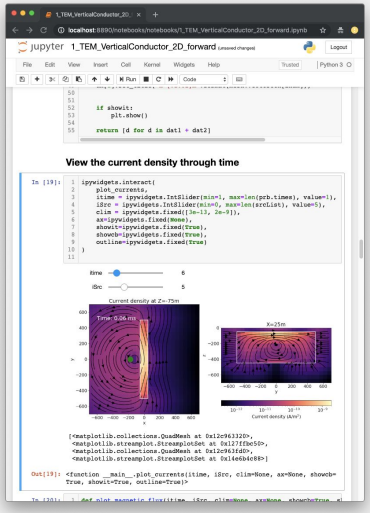
👋 hello (a bit about me)

geophysical inversions



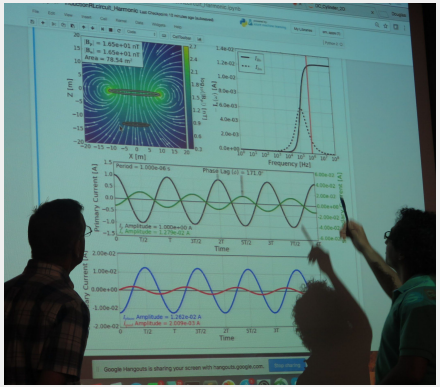
THE UNIVERSITY OF BRITISH COLUMBIA

open-source software



simpeg

open research & education



GeoSci.xyz

Jupyter, geoscience + data science

jupyter + 

Berkeley UNIVERSITY OF CALIFORNIA

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

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
> WaveLab and Reproducible Research, 1995

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
> WaveLab and Reproducible Research, 1995

the science is the code

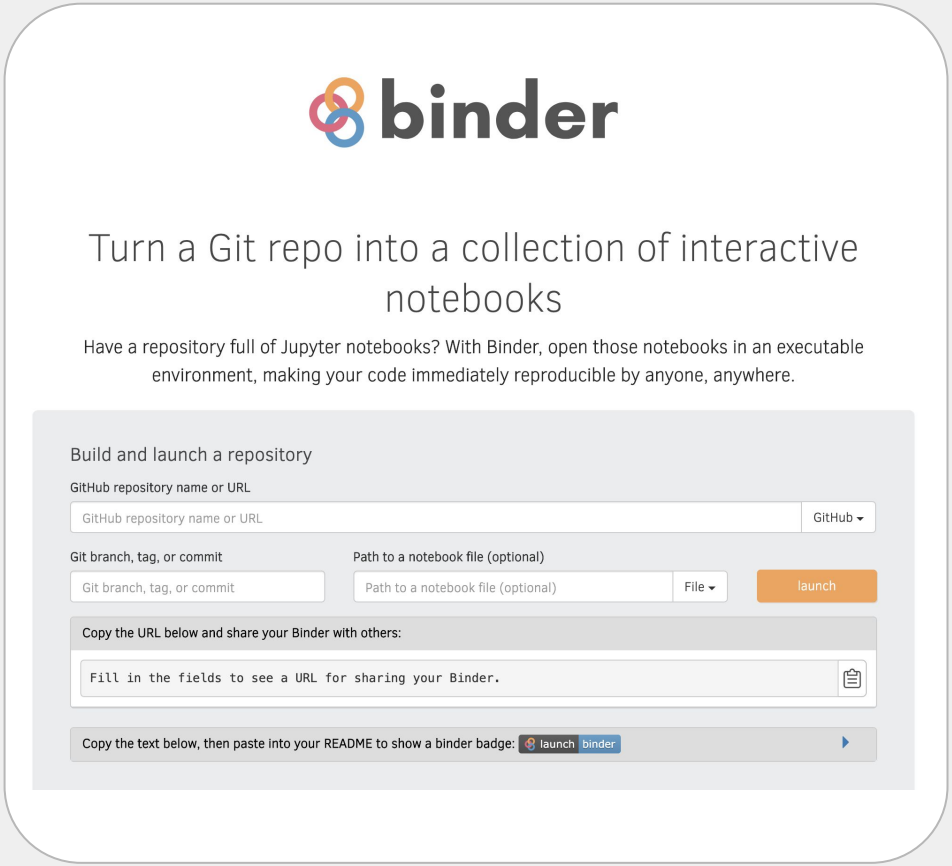
An article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures.

-- Buckheit and Donoho
WaveLab and Reproducible Research, 1995

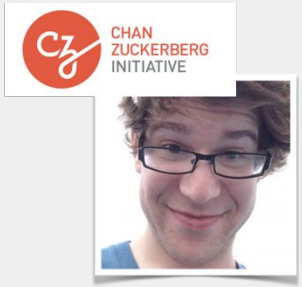
(and a place to run the code?)

live, computational environments, running on the cloud, built from your research repositories

mybinder.org

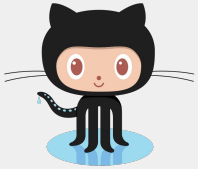


origins



github.com/freeman-lab

+



GitHub

+

explicit dependencies

+

docker

kubernetes

- creates reproducible containers from repositories (repo2docker)
- generates user sessions that serve these containers (JupyterHub)
- provides an interface to create, share, and use these sessions (BinderHub)
- demonstrates the above as a free public service/tech demo (mybinder.org)

open science

(why?)

- in order to collaborate
- to build on the work of others
- for others to build upon your work
- to make revisions to your paper when you hear back from reviewers in 8 months

open science

(how?)

- complete set of instructions
- complete development environment
- a place to run the code

open science

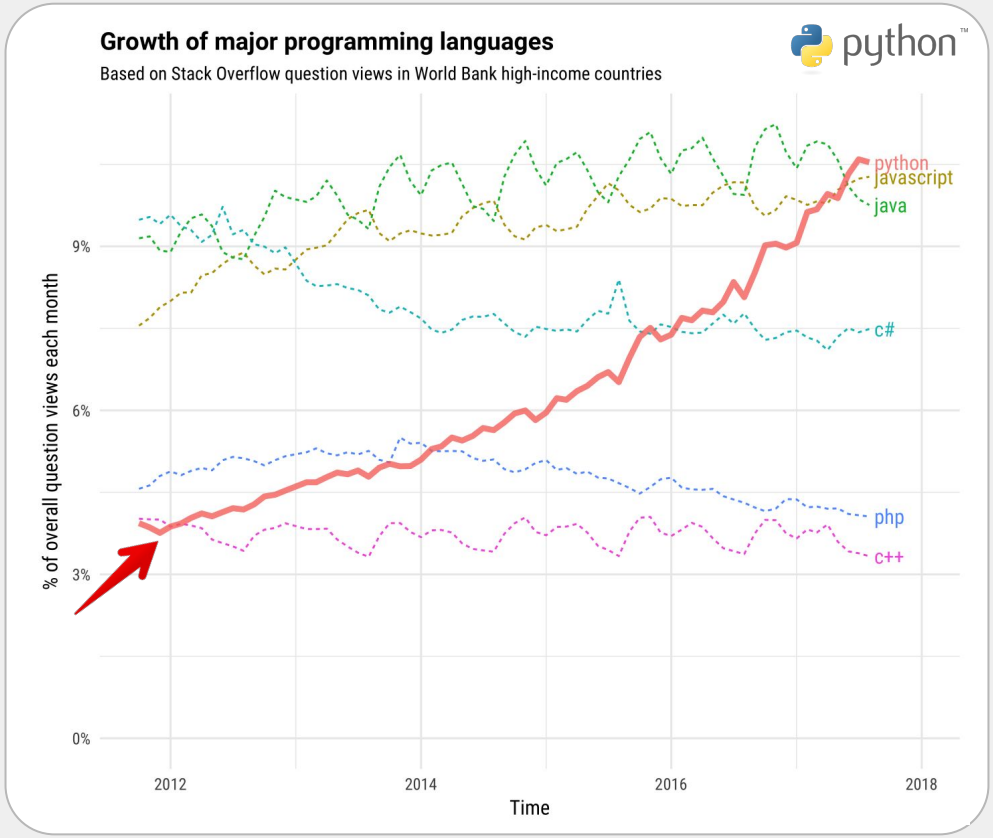
(how?)

- complete set of instructions
- complete development environment
- a place to run the code

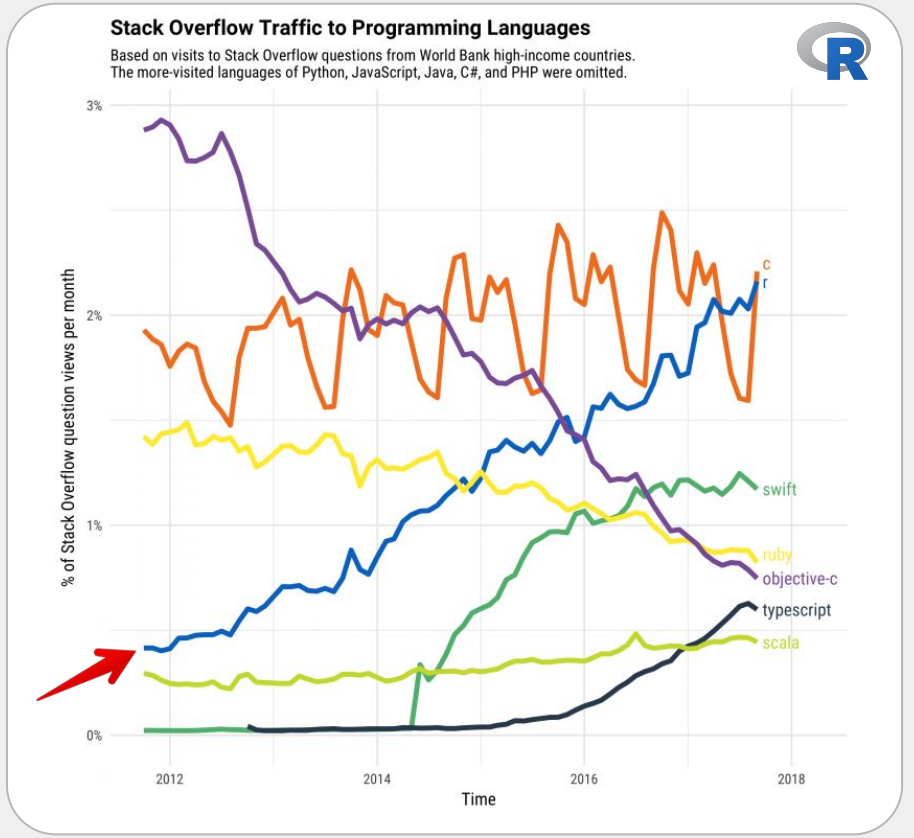
complete set of instructions

open-source languages are the raw material

complete set of instructions



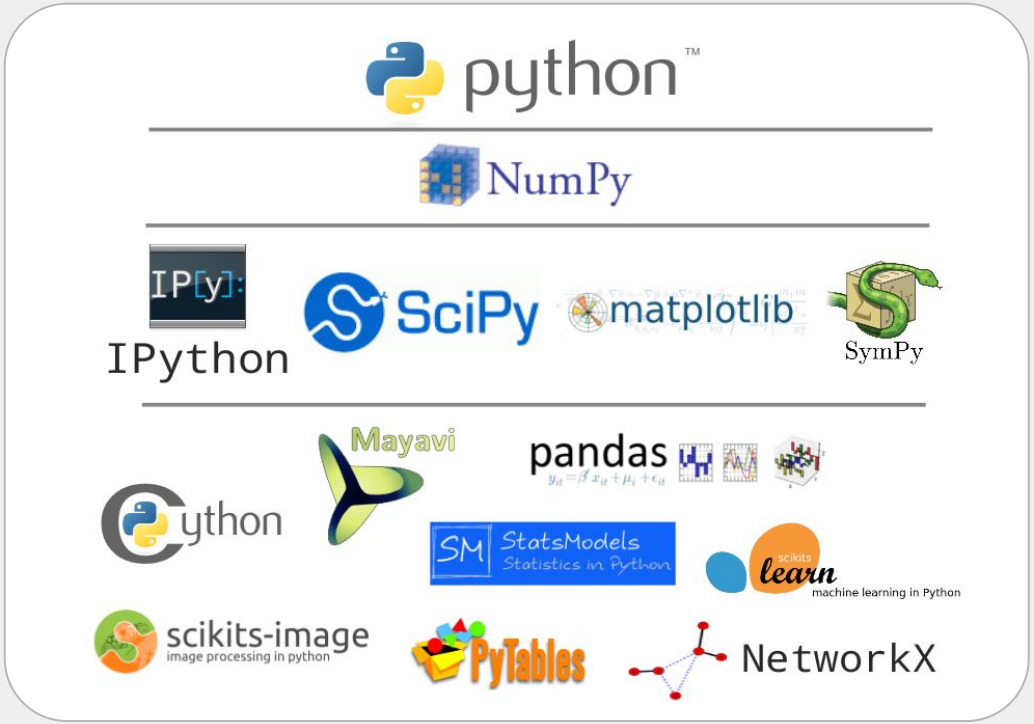
https://stackoverflow.blog/2017/09/06/incredible-growth-python/



https://stackoverflow.blog/2017/10/10/impressive-growth-r/

complete set of instructions

mature ecosystems of tools



http://www.focusedsupport.com/blog/getting-setup-with-scientific-python/

complete set of instructions

web-native interfaces for interacting with code

complete set of instructions

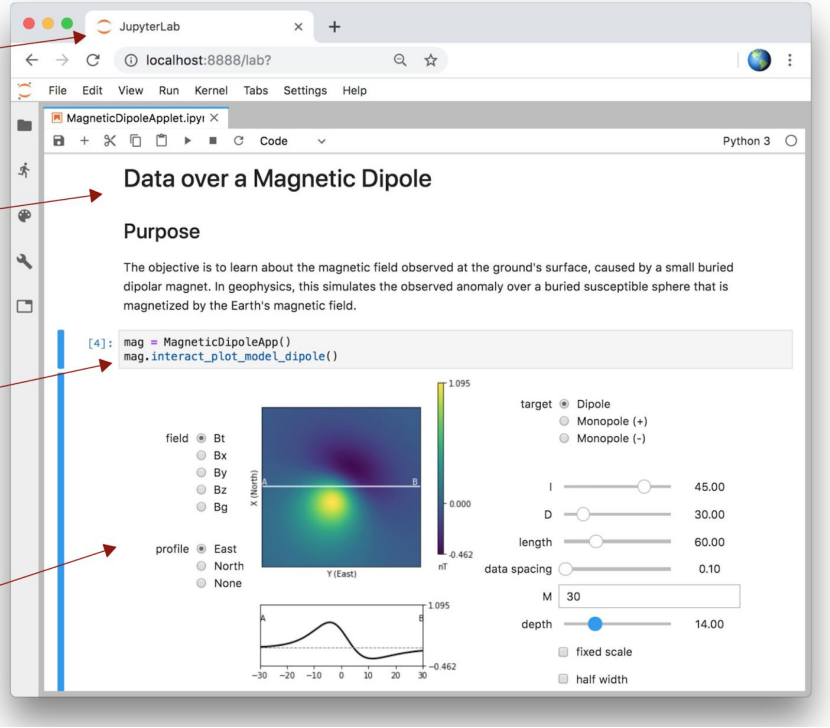
capture the steps

Runs in web browser

Text, comments, equations

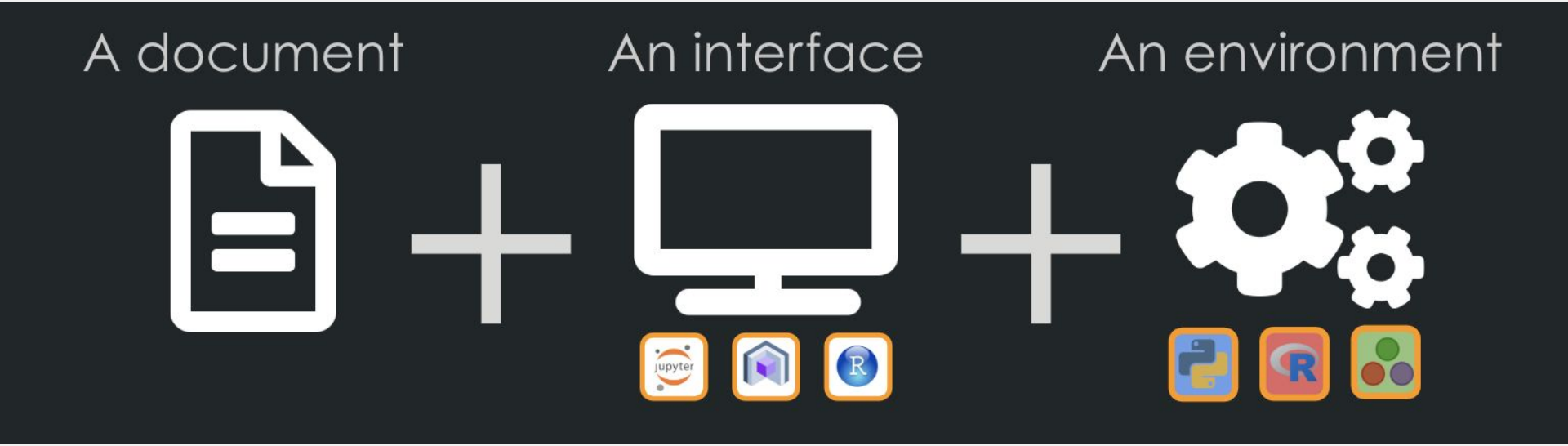
Lines of code
(Python, Julia, R, Matlab, …)

Output: values, images, plots, tables, interactive widgets



complete set of instructions

capture the steps: what is a notebook?



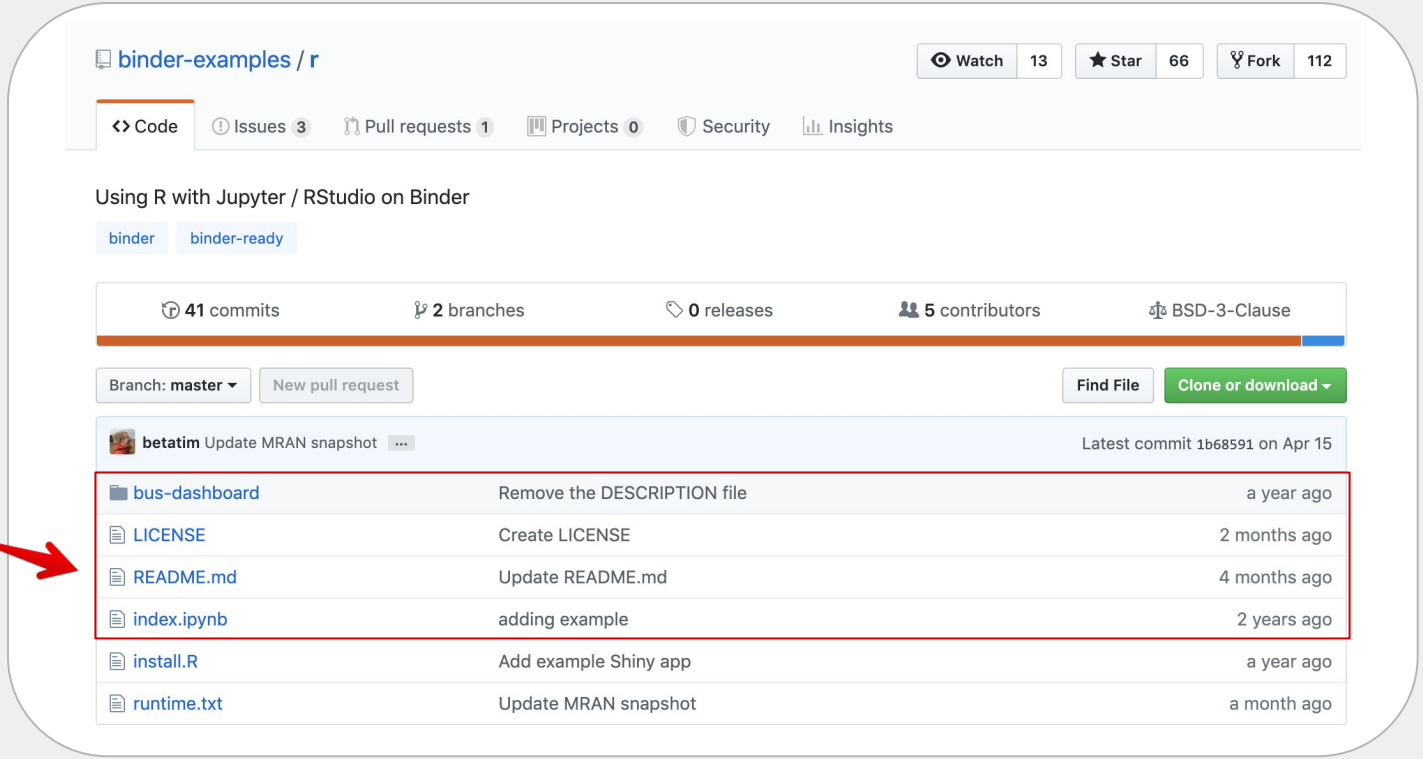
complete set of instructions

maintenance and sharing

- version control
- issue tracking
- licensing
- integrations with
  - testing services
  - documentation hosting
  - ...

complete set of instructions

maintenance and sharing



https://github.com/binder-examples/r

open science

(how?)

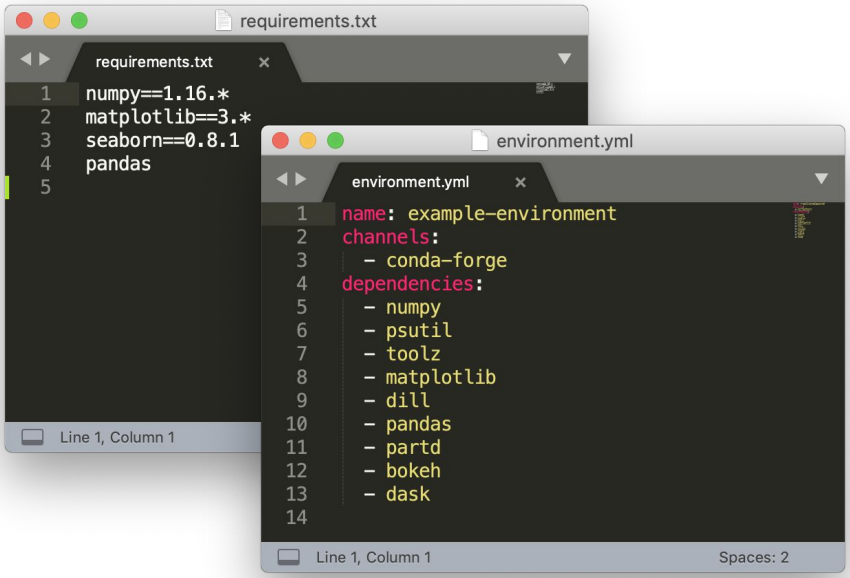
- complete set of instructions
- complete development environment
- a place to run the code

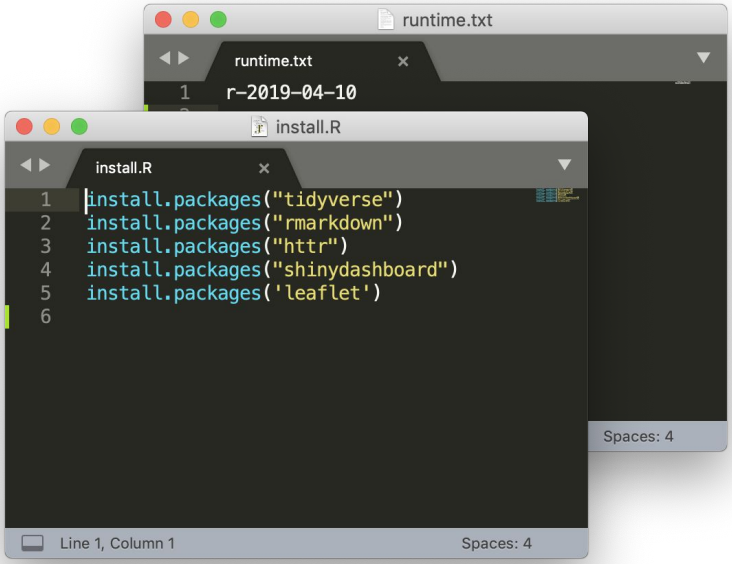
repo2docker

repo2docker deterministically build a docker image
from a repository with documented dependencies

complete development environment

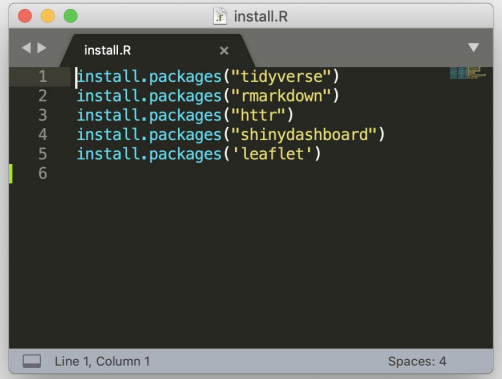
define dependencies following community standards of practice





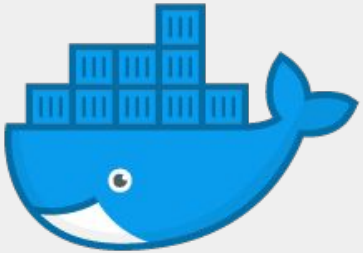
complete development environment

repo2docker



dependencies

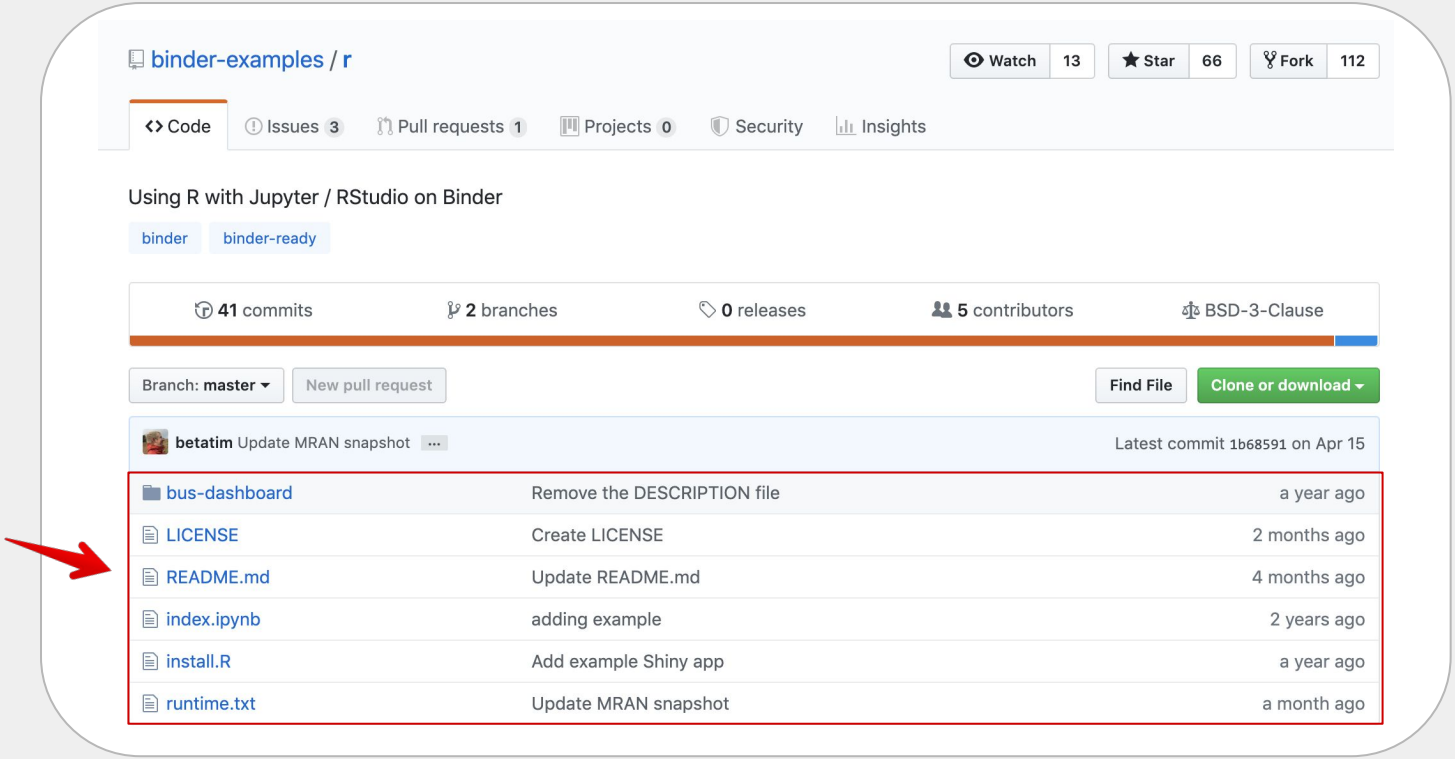
repo2docker



container file

complete development environment

define dependencies



https://github.com/binder-examples/r

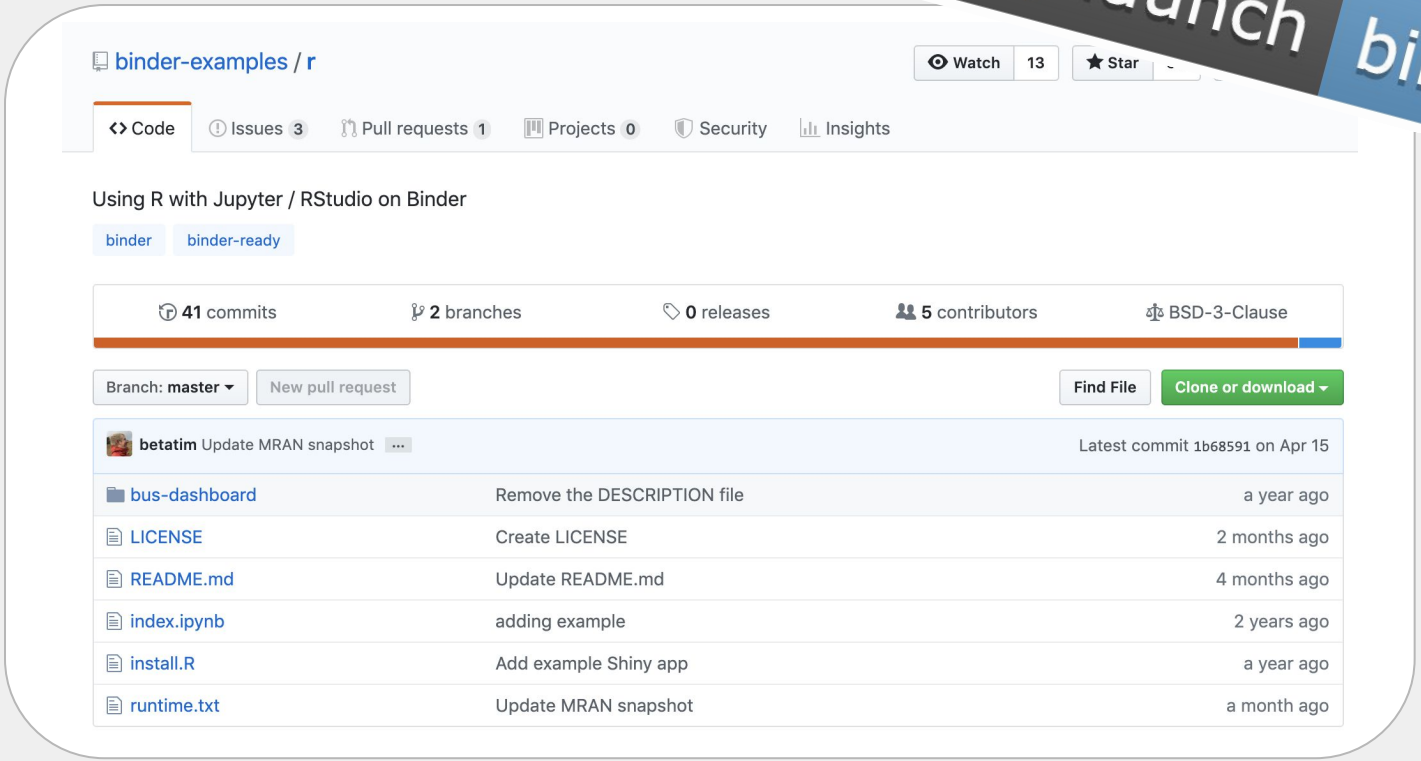
open science

(how?)

- complete set of instructions
- complete development environment
- a place to run the code

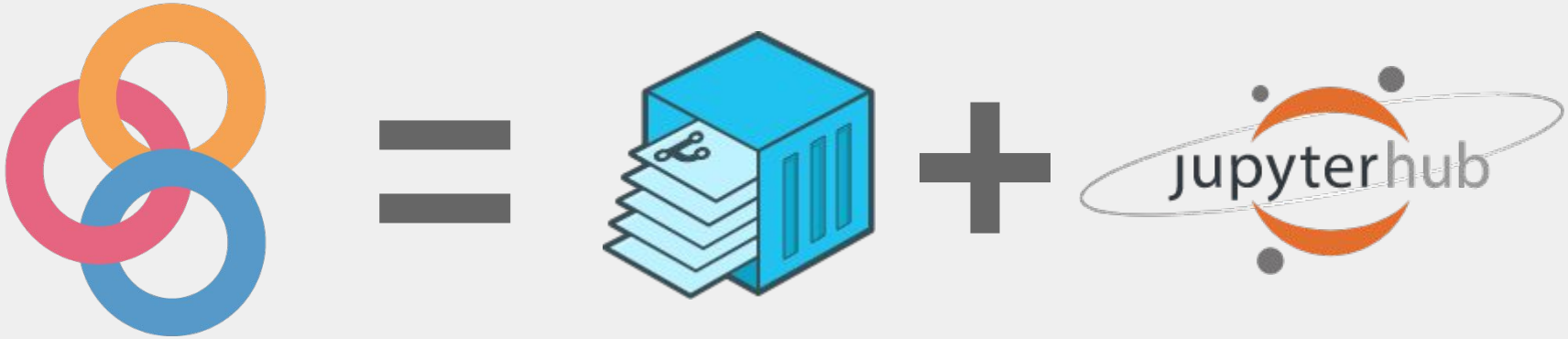
a place to run the code

binder: computational environments on the cloud



a place to run the code

binder!



binder

repo2docker

JupyterHub
(next talk!)

Turn a Git 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 repository name or URL | GitHub

Git branch, tag, or commit

Git branch, tag, or commit

Path to a notebook file (optional)

Path to a notebook file (optional) | File

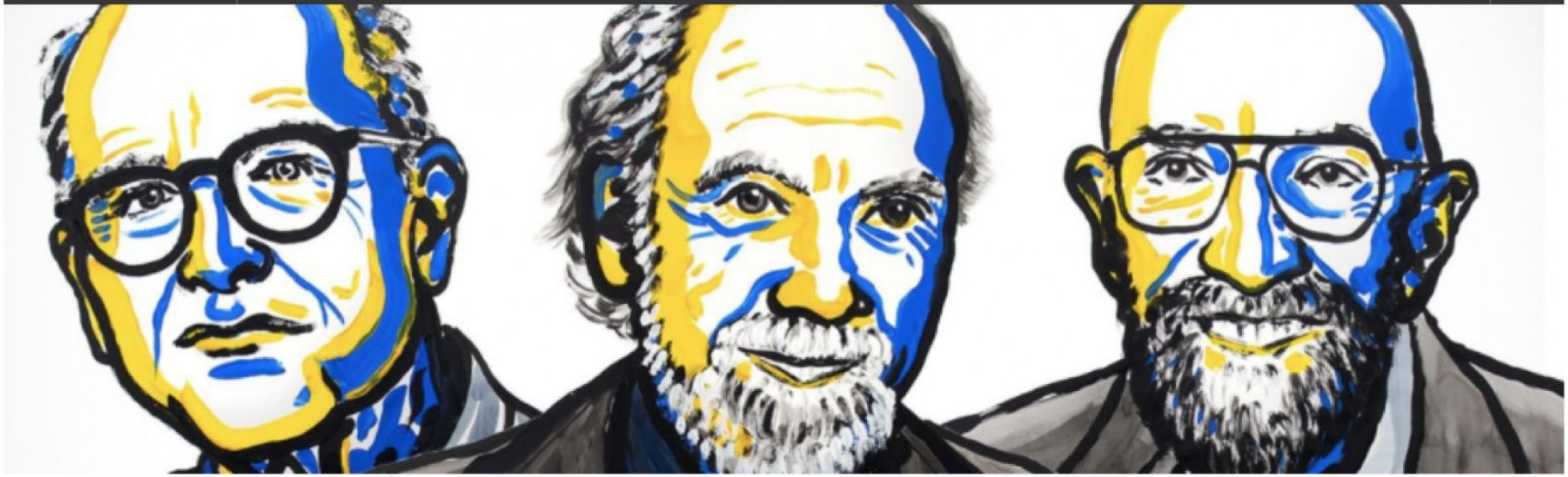
launch

Copy the URL below and share your Binder with others:

```
Fill in the fields to see a URL for sharing your Binder.
```

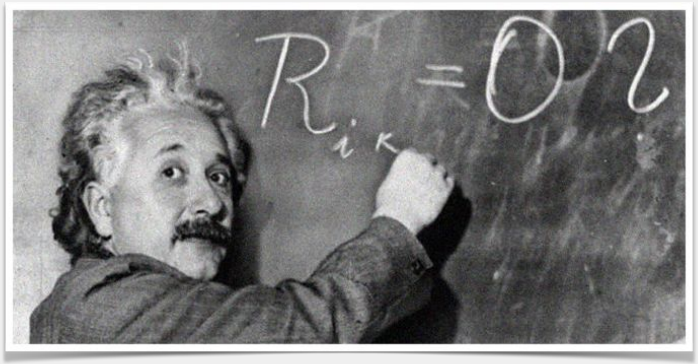
Copy the text below, then paste into your README to show a binder badge: launch binder

mybinder.org



2017 NOBEL PRIZE IN PHYSICS AWARDED FOR DISCOVERY OF GRAVITATIONAL WAVES

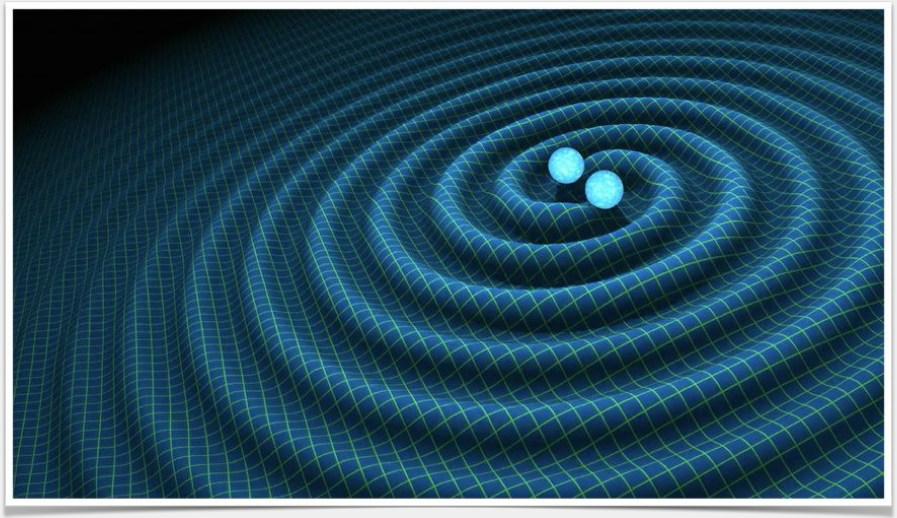
as predicted by Einstein



$$R_{\mu\nu} - \tfrac{1}{2} R\, g_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}$$

Einstein's Field Equations of General Relativity

Annalen der Physik, 1916



LIGO collaboration discovery: Sept 14, 2015

Detection problem:

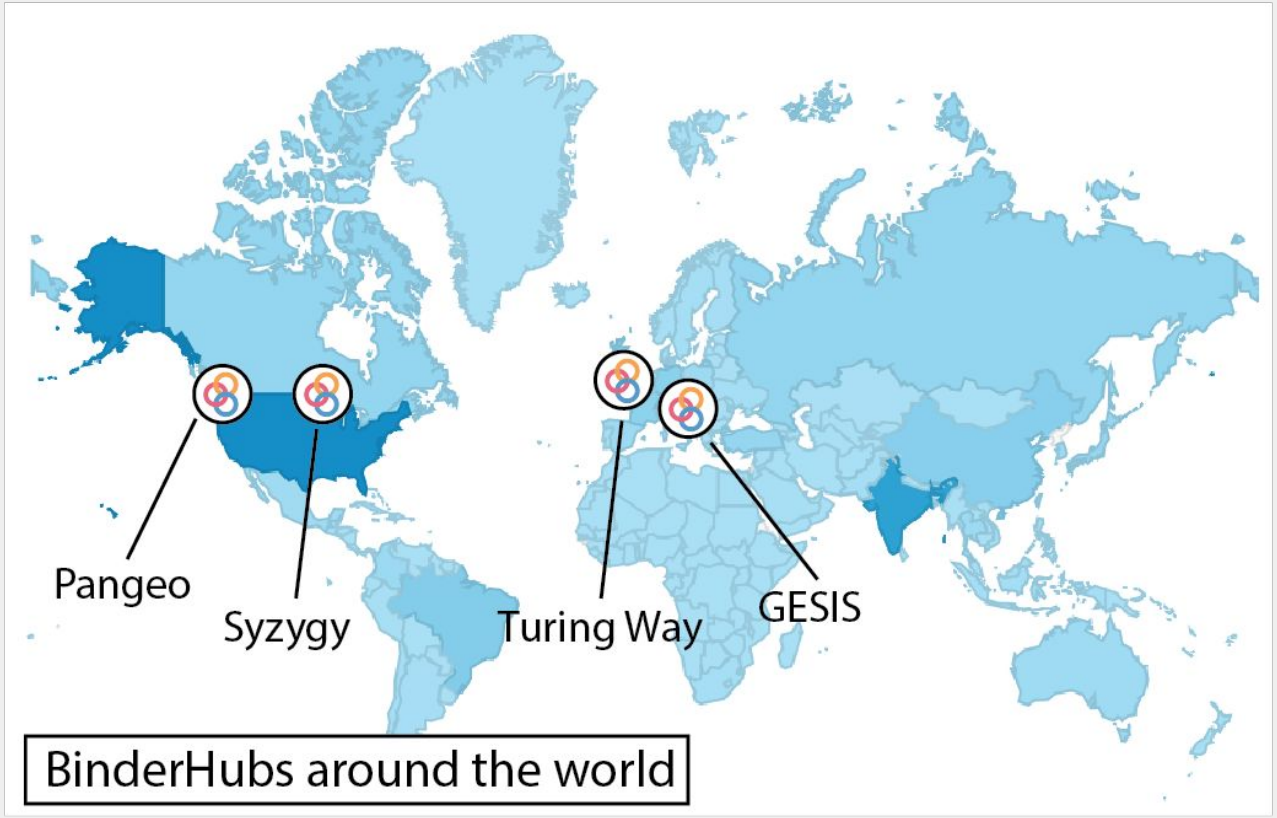
- ~ 1/1000 proton over 4 km.
- Sensitivity ~ 1e-21
- Milky Way: 1e+21m across!

demo time

()

https://www.gw-openscience.org/tutorials/
https://github.com/binder-examples/r

Future


Pangeo
Syzygy
Turing Way
GESIS
BinderHubs around the world

thank you!

try out binder:

`mybinder.org`

connect with the community:

`discourse.jupyter.org`

THE LEONA M. AND HARRY B. HELMSLEY CHARITABLE TRUST

ALFRED P. SLOAN FOUNDATION

GORDON AND BETTY MOORE FOUNDATION