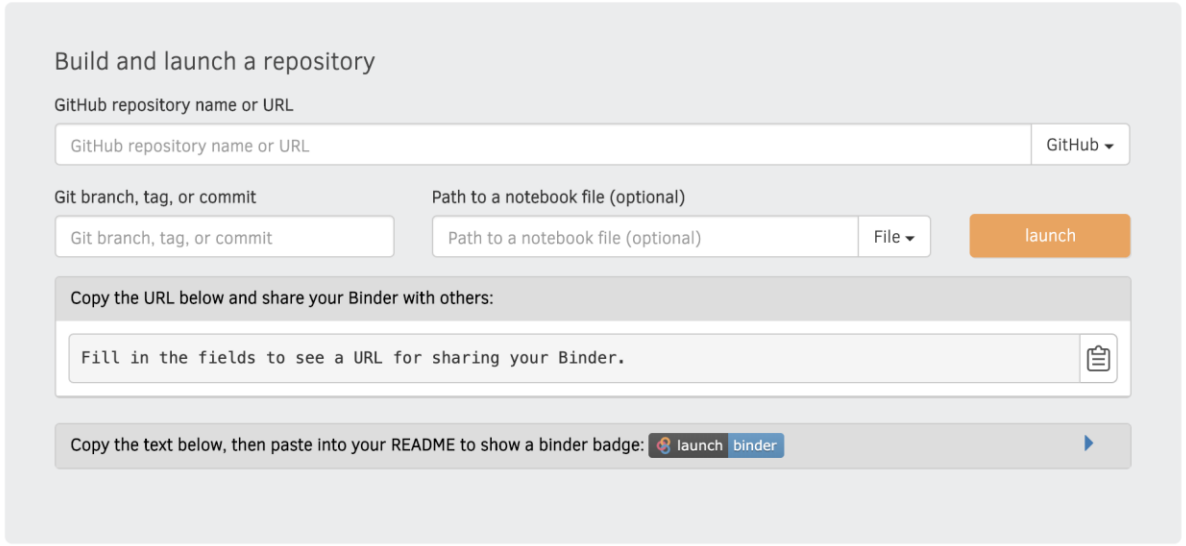
Sharing Reproducible Computational Environments with

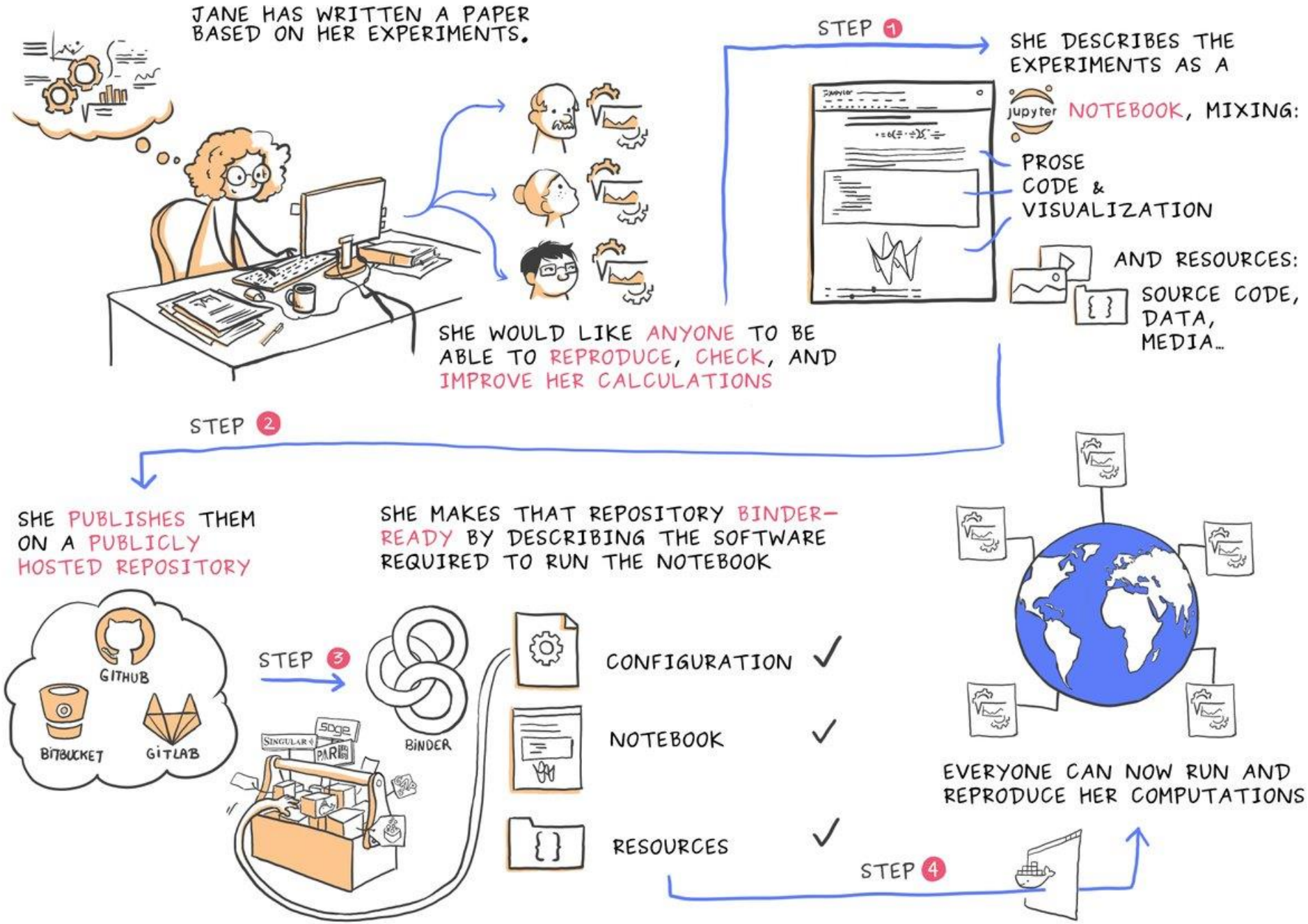
Packaging software is hard…

- Hardware
- Software
- Programming language
- Packages and libraries
- The code itself
- Resources

What is binder?

- Public, free to use service
- Provides hardware, software and code
- Creates a link to a browser window where code, running in the Cloud, can be explored interactively
- https://mybinder.org



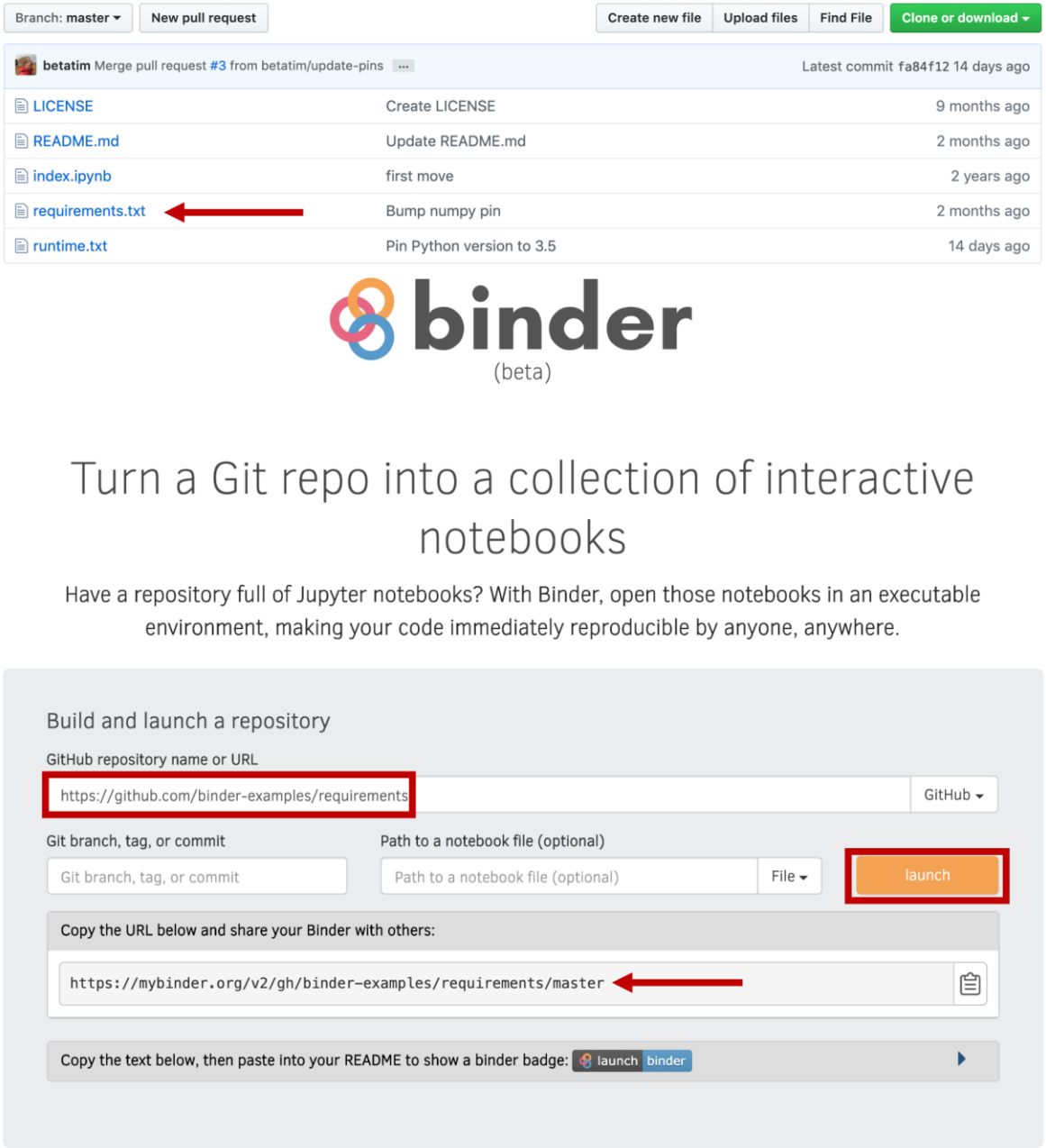


Courtesy of Juliette Taka: https://twitter.com/mybinderteam/status/1082556317842264064

Easy as 1, 2, 3…

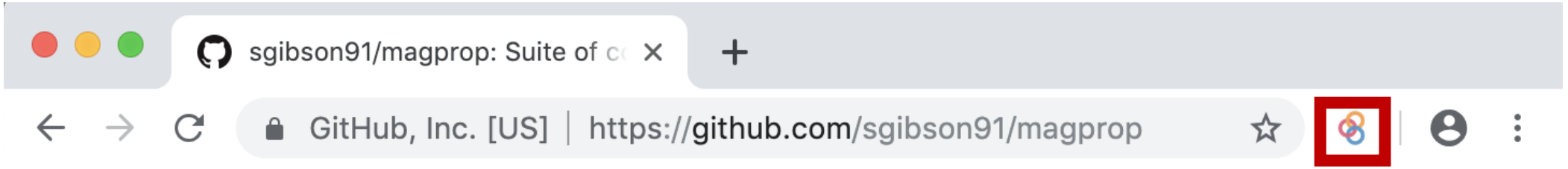
- Step 1: Create a config file for your project
- Step 2: Enter repo URL into mybinder.org
- Step 3: Hit launch!

bit.ly/zero-to-binder-solo



Open With Binder browser extension

- Available for Chrome and Firefox
- Launch a repo in Binder directly from the GitHub page!



Types of config files… Python with pip

pip freeze > requirements.txt

4 lines (3 sloc) | 45 Bytes — Raw | Blame | History

```
1  numpy==1.16.*
2  matplotlib==3.*
3  seaborn==0.8.1
```

https://github.com/binder-examples/requirements/blob/master/requirements.txt

Types of config files… Python with conda

conda env export > environment.yml

14 lines (13 sloc) | 161 Bytes

```yaml
name: example-environment
channels:
  - conda-forge
dependencies:
  - numpy
  - psutil
  - toolz
  - matplotlib
  - dill
  - pandas
  - partd
  - bokeh
  - dask
```

https://github.com/binder-examples/python-conda_pip/blob/master/environment.yml

Types of config files... **runtime.txt**

- Specify a Python 2.7 environment
- (Can also be achieved using environment.yml and conda)

1 lines (1 sloc) | 10 Bytes | Raw | Blame | History

```
1  python-2.7
```

https://github.com/binder-examples/python2_runtime/blob/master/runtime.txt

Types of config files... R environments

install.R <- This is a made-up file!

6 lines (5 sloc) | 148 Bytes — Raw | Blame | History

```
1  install.packages("tidyverse")
2  install.packages("rmarkdown")
3  install.packages("httr")
4  install.packages("shinydashboard")
5  install.packages('leaflet')
```

If you have an R package, the DESCRIPTION and NAMESPACE files are (almost) enough to describe the dependencies

https://github.com/binder-examples/r/blob/master/install.R
https://github.com/binder-examples/binder-r-description

Some caveats for R users…

- Need a runtime.txt file
  - Binder uses MRAN to pull packages
  - MRAN takes daily snapshots of CRAN
- Some packages take a long time to install…
  - E.g. tidyverse
  - Either only list the required packages or create bespoke Dockerfile
  - https://github.com/binder-examples/rocker
  - https://github.com/karthik/holepunch

Types of config files… All the cool things!

Binder supports:

- Julia, Stencila, Bokeh, Octave (free MatLab)…
- apt and nix package managers
  - E.g. LaTeX, vim…

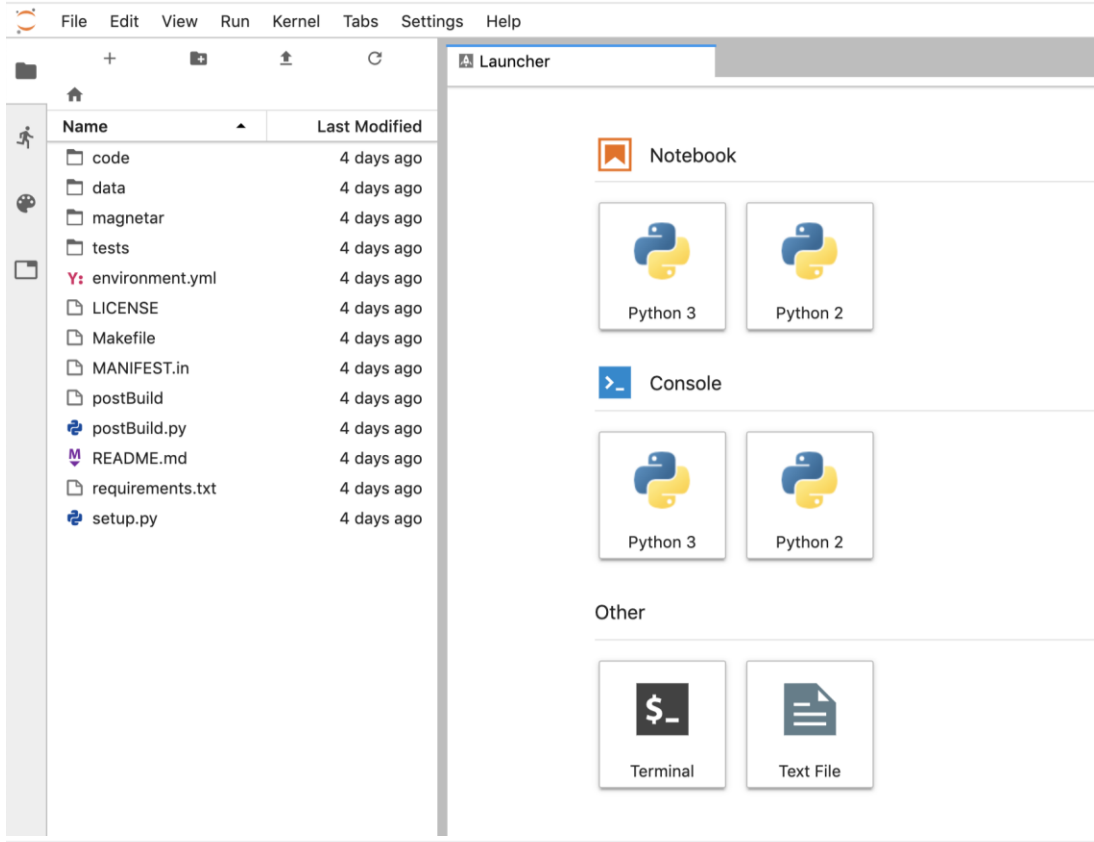
https://mybinder.readthedocs.io/en/latest/config_files.html

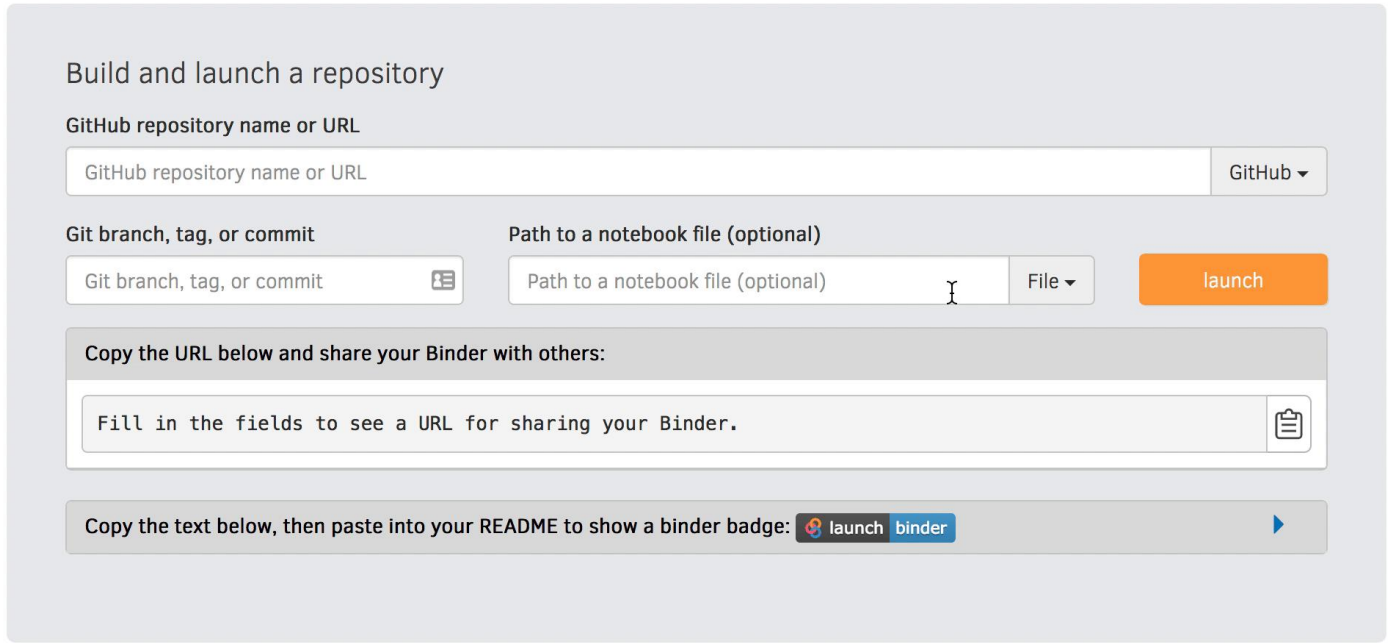
- Java for plotting: https://github.com/twosigma/beakerx
- Octave JupyterBook: https://joergbrech.github.io/Modellbildung-und-Simulation/intro
- Jupyter kernel for C++: https://github.com/QuantStack/xeus-cling
- Multi-language: https://github.com/binder-examples/multi-language-demo
- Continuously build notebook containers: https://github.com/binder-examples/continuous-build

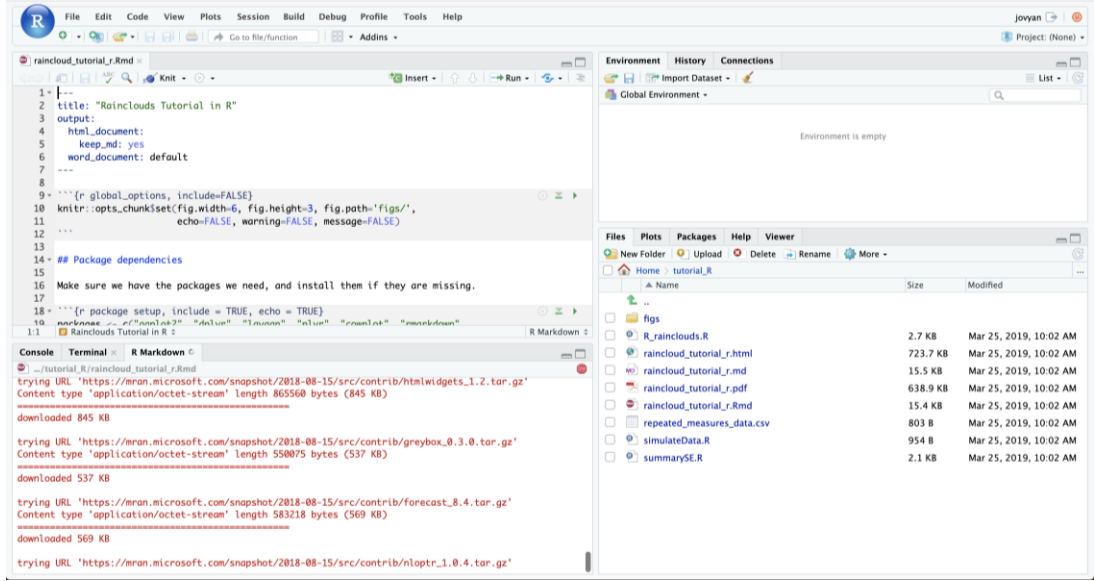
More cool things… IDEs

JupyterLab and RStudio are installed by default

Accessed via lab or rstudio values to ?urlpath= URL argument







postBuild scripts… Getting data and more

- Executed in the container after the environment has been built
- Bake datasets into the environment
- Execute scripts before the user arrives in the environment
- bash commands

Executable File | 12 lines (10 sloc) | 200 Bytes

```
1  set -ex
2
3  invoke build --env-name=root --no-kernel
4  invoke demofiles
5  invoke talk -t demo
6  rm -rf demofiles
7  rm -rf notebooks
8  rm -rf narrative
9  rm -rf slides
10 rm demo/notebooks/Julia.ipynb
11 jupyter lab clean
```

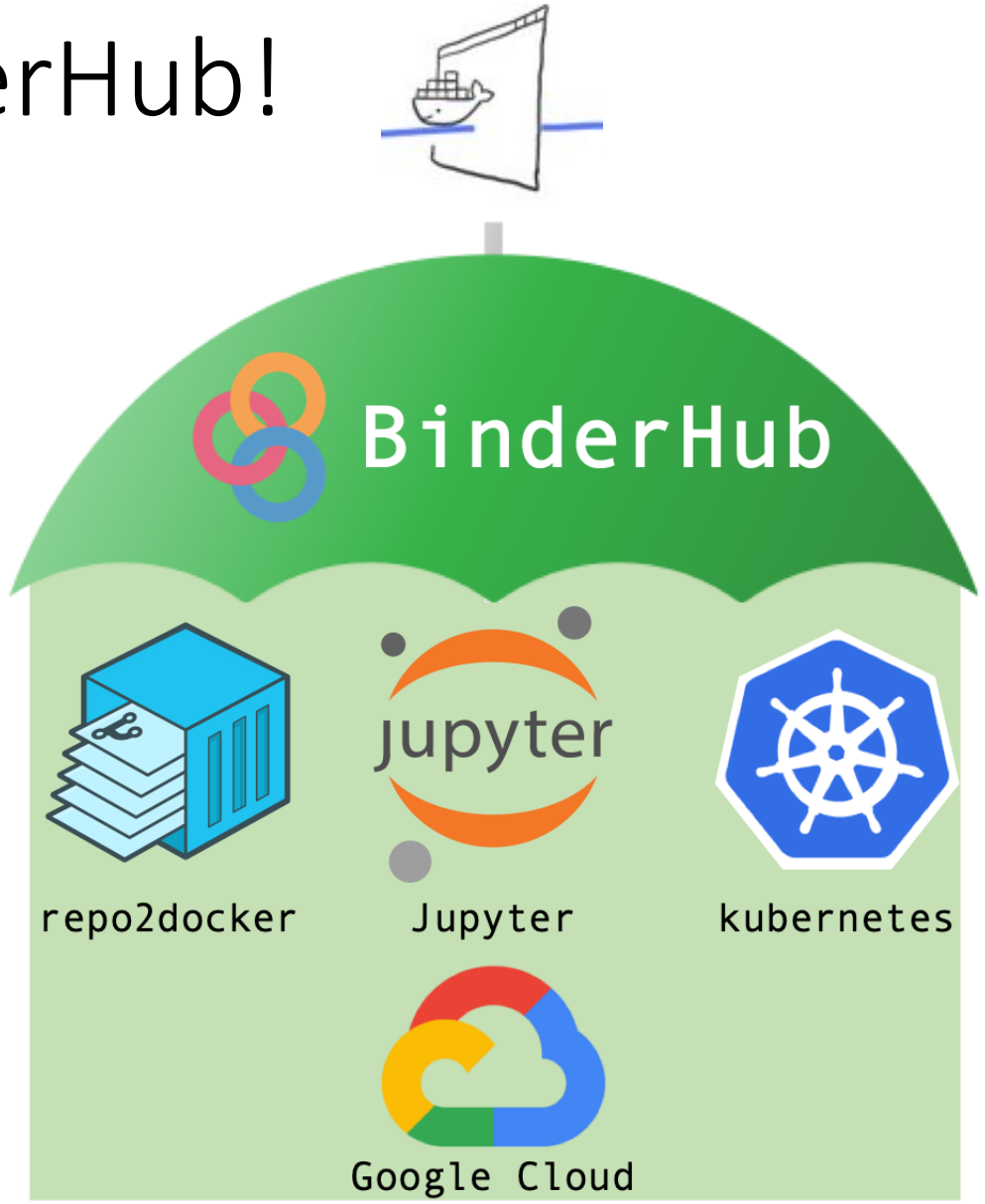
https://github.com/jupyterlab/jupyterlab-demo/blob/master/binder/postBuild

Important things to remember about Binder

- **Changes are not saved!**
  - Anything you've done will be lost when the browser is closed
  - Pushes to the original repo are not currently permitted, though see:
  - https://github.com/jupyterhub/binderhub/issues/623
- **mybinder.org is completely public!**
  - Don't put anything in your binder that you don't want shared with the world
- The image is rebuilt with each new commit, not each launch
  - Relaunches *should* be quicker (but not always)

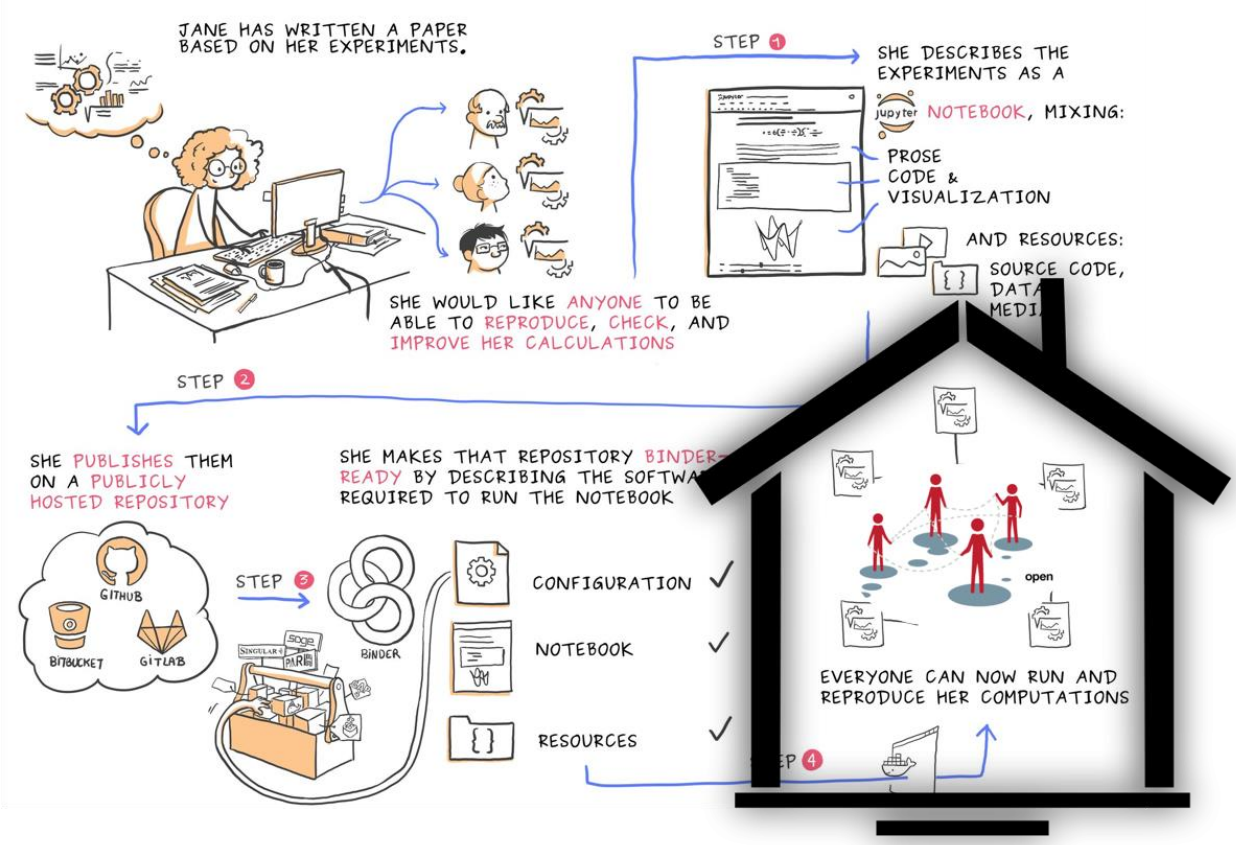
Under the hood… a BinderHub!

- Cloud-based technology
  - currently operating on Google Cloud
  - Kubernetes automatically manages computational resources and deploying containers on to the Cloud
- repo2docker
  - builds a docker container based on a config file and a GitHub repo contents
- JupyterHub
  - connects Cloud server with your browser



Why bother building a BinderHub?

- Authentication
- Private repo access
- Control computational resources
- Choose to share amongst teams or publicly
- They're cool!



Hub23 is coming!

- The BinderHub for Hut23 and Turing research hosted on Azure
- Current implementation:
  - Autoscaling
  - Login via GitHub
  - Can only build public repos
  - Docker images are still pushed to a public DockerHub account
  - Limited to .5 vCPU and 1GB RAM
- Roadmap:
  - Login in via Azure Active Directory as well (minimum plan: tweak the current login schema)
  - Exchange DockerHub for Azure Container Registry (images will now be private)
  - Grant access to build private repos