Workflows for Reproducible Research with R & Git

(Jupyter) Notebooks & Binder

Mitglied der Leibniz Gemeinschaft

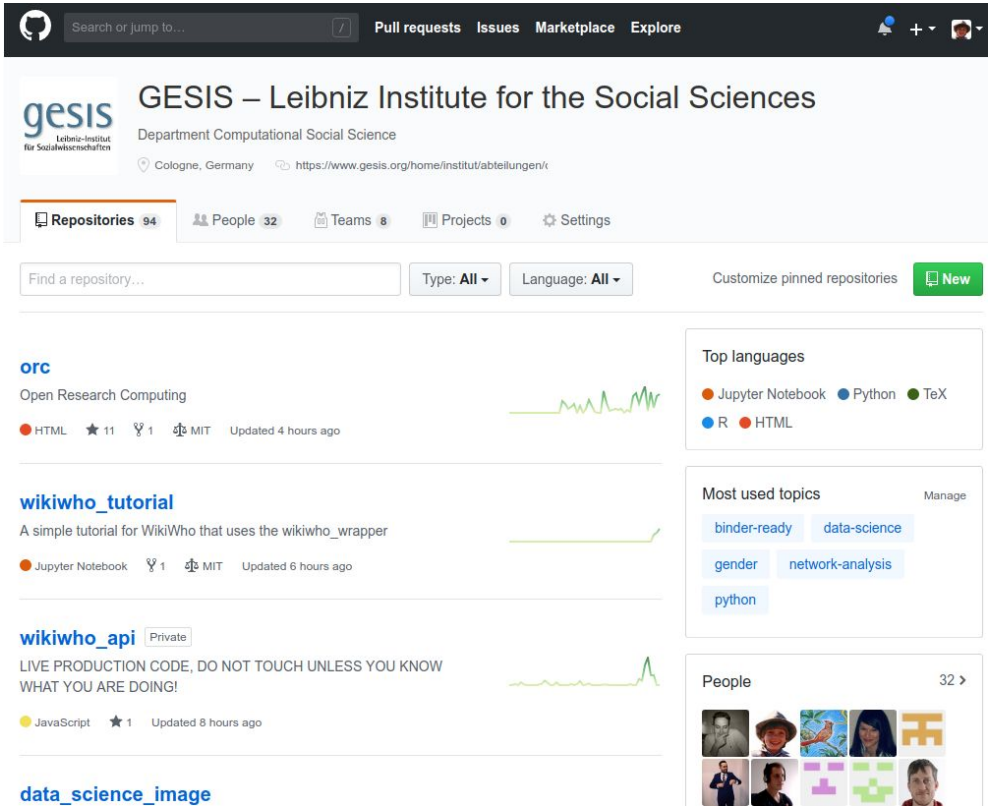
Organizers: Bernd Weiß
Johannes Breuer
Arnim Bleier

November 17th, 2022

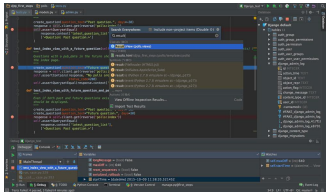
GESIS Library, Cologne

Mitglied der Leibniz Gemeinschaft

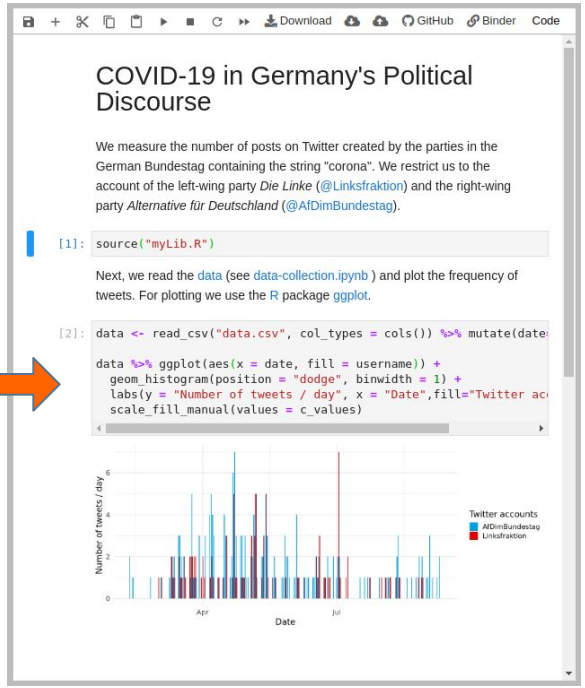
Open Access **Code**



What are Notebooks: Literate Programming



Source code



Natural language

Examples:

- ❑ Jupyter
- ❑ R Markdown
- ❑ Pluto.jl
- ❑ …

Try Jupyter (exercise)

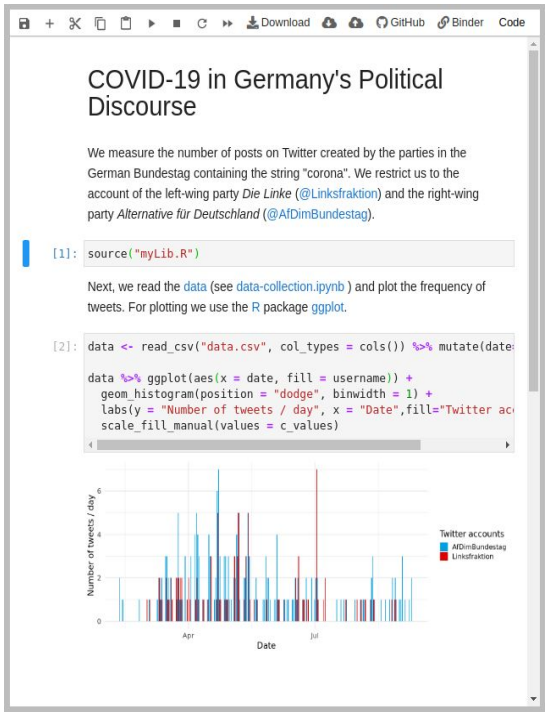
https://mybinder.org/v2/gh/jupyterlab/jupyterlab-demo/master

https://notebooks.gesis.org/binder/v2/gh/arnim/RStan-Binder/master

Computation

Cloud:

- ❏ potentially large Data
- ❏ standardized environment
- ❏ 1-Click reproducibility

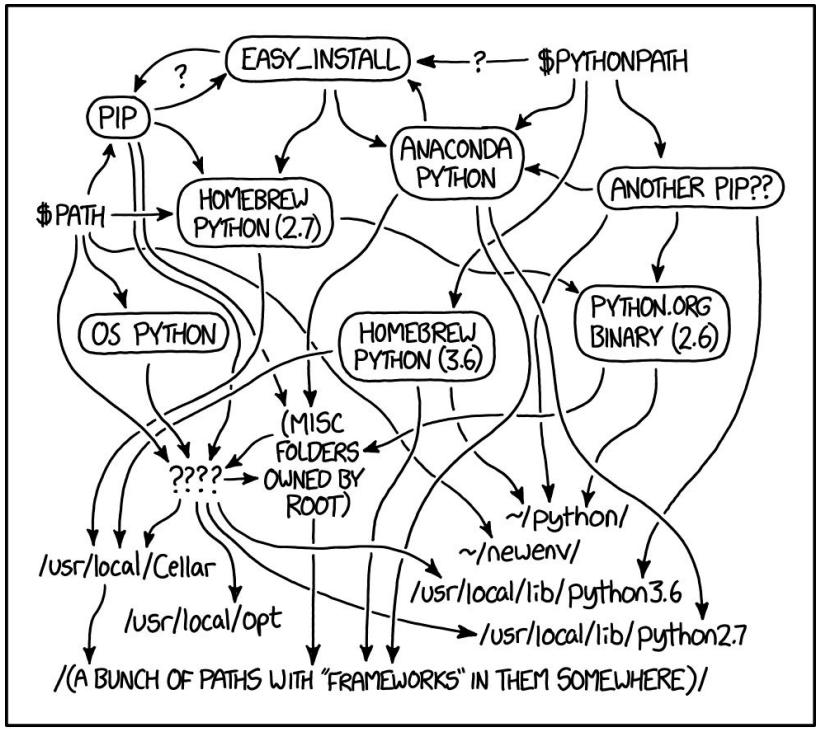


Personal Computer:

- ❏ only small data
- ❏ every environment different
- ❏ time consuming to set up

Mitglied der Leibniz Gemeinschaft

The environment matters



MY PYTHON ENVIRONMENT HAS BECOME SO DEGRADED
THAT MY LAPTOP HAS BEEN DECLARED A SUPERFUND SITE.

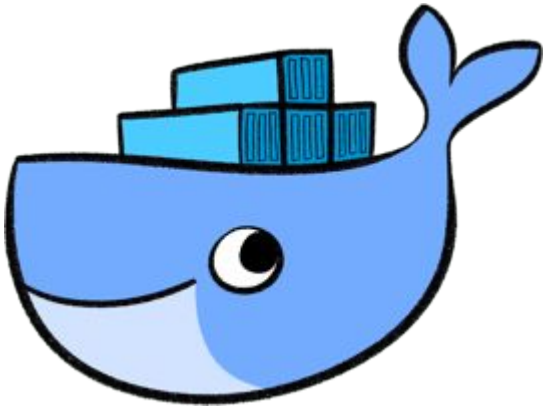
https://xkcd.com/1987/

Is “Lockdown” the Solution?

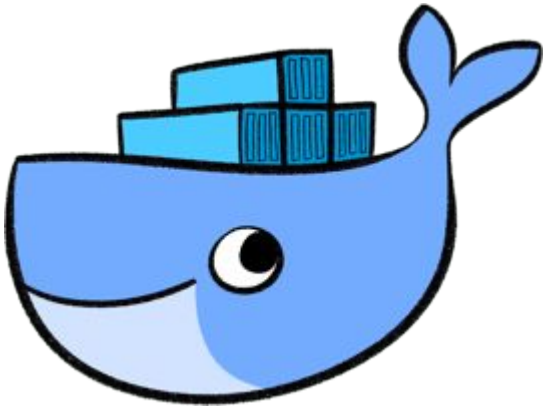
Only the administrators control the environment.

But … “my Work is Special”

Is Docker the Solution?



Is Docker the Solution?

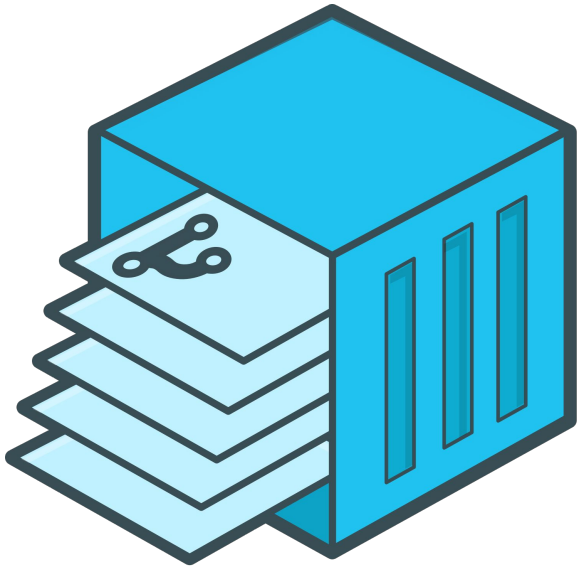


Dockerfile

```
FROM ubuntu

RUN echo "deb http://us.archive.ubuntu.com/ubuntu/ precise universe" >> /etc/apt/sources.list
RUN apt-get -y update
RUN apt-get install -y g++
RUN apt-get install -y erlang-dev erlang-manpages erlang-base-hipe erlang-eunit erlang-nox
erlang-xmerl erlang-inets

RUN apt-get install -y libmozjs185-dev libicu-dev libcurl4-gnutls-dev libtool wget

RUN cd /tmp ; wget
http://www.bizdirusa.com/mirrors/apache/couchdb/source/1.3.1/apache-couchdb-1.3.1.tar.gz

RUN cd /tmp && tar xvzf apache-couchdb-1.3.1.tar.gz
RUN apt-get install -y make
RUN cd /tmp/apache-couchdb-* ; ./configure && make install

RUN printf "[httpd]\nport = 8101\nbind_address = 0.0.0.0" >
/usr/local/etc/couchdb/local.d/docker.ini

EXPOSE 8101

CMD ["/usr/local/bin/couchdb"]
```

https://github.com/kstaken/dockerfile-examples/blob/master/couchdb/Dockerfile

Build Docker Images from a Git Repository



jupyter-repo2docker is a tool for building and running Docker images from source code repositories.

https://repo2docker.readthedocs.io

What does jupyter-repo2docker ?

Consider you want to build and run a simple binder repository

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

How would you proceed?

1) git clone https://github.com/binder-examples/requirements
2) pip install -r requirements.txt
3) jupyter notebook

What does jupyter-repo2docker ?

Consider you want to build and run a simple binder repository

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

How would you proceed using repo2docker?

```
jupyter-repo2docker https://github.com/binder-examples/requirements
```

(Some) supported Environment Configuration Files

requirements.txt

```
numpy==1.13.1
matplotlib==2.0.2
seaborn==0.8.1
```

or

environment.yaml

```
name: example-environment
Channels:
 - conda-forge
dependencies:
 - python
 - numpy
```

install.R

```
install.packages("tidyverse", repos =
"https://cloud.r-project.org/",
dependencies=TRUE)
```

runtime.txt

```
r-2018-07-27
```

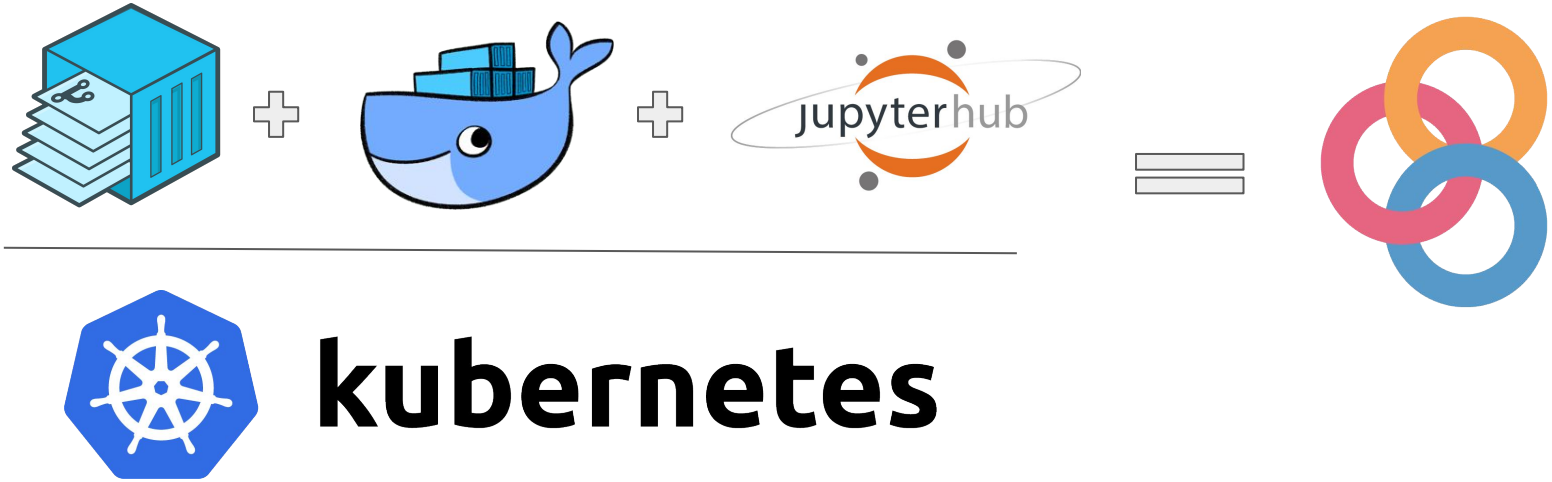

See https://repo2docker.readthedocs.io/en/latest/config_files.html

Terminal - arnim@KOL16001 ~

File Edit View Terminal Tabs Help

```
arnim@KOL16001 ~ $ jupyter-repo2docker https://github.com/binder-examples/requirements
Picked Git content provider.
Cloning into '/tmp/repo2dockerto2bblgt'...
remote: Enumerating objects: 6, done.
remote: Counting objects: 100% (6/6), done.
remote: Compressing objects: 100% (5/5), done.
remote: Total 6 (delta 0), reused 4 (delta 0), pack-reused 0
Unpacking objects: 100% (6/6), done.
Reusing existing image (r2dhttps-3a-2f-2fgithub-2ecom-2fbinder-2dexamples-2frequirementsd0583e9), not building.[I 02:02:06.578
 NotebookApp] Writing notebook server cookie secret to /home/arnim/.local/share/jupyter/runtime/notebook_cookie_secret
[I 02:02:06.931 NotebookApp] JupyterLab extension loaded from /srv/conda/lib/python3.6/site-packages/jupyterlab
[I 02:02:06.931 NotebookApp] JupyterLab application directory is /srv/conda/share/jupyter/lab
[I 02:02:06.941 NotebookApp] nteract extension loaded from /srv/conda/lib/python3.6/site-packages/nteract_on_jupyter
[I 02:02:06.943 NotebookApp] Serving notebooks from local directory: /home/arnim
[I 02:02:06.943 NotebookApp] The Jupyter Notebook is running at:
[I 02:02:06.943 NotebookApp] http://127.0.0.1:44831/?token=a49e0def6bba998835161f511426a0c19163bc55471f7ce2
[I 02:02:06.943 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[W 02:02:06.943 NotebookApp] No web browser found: could not locate runnable browser.
[C 02:02:06.944 NotebookApp]

    Copy/paste this URL into your browser when you connect for the first time,
    to login with a token:
        http://127.0.0.1:44831/?token=a49e0def6bba998835161f511426a0c19163bc55471f7ce2
```

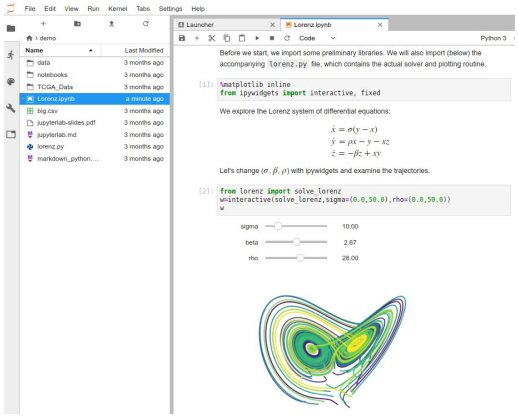
What is BinderHub?



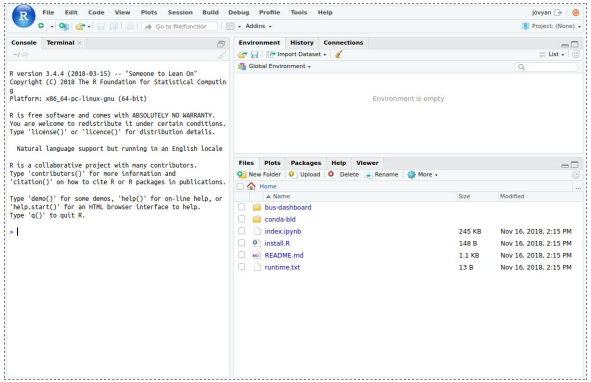
https://github.com/jupyterhub/binderhub

Have a look at the Open Source Project:
https://github.com/jupyterhub/binderhub/

Join the Binder chat for questions:
https://gitter.im/jupyterhub/binder

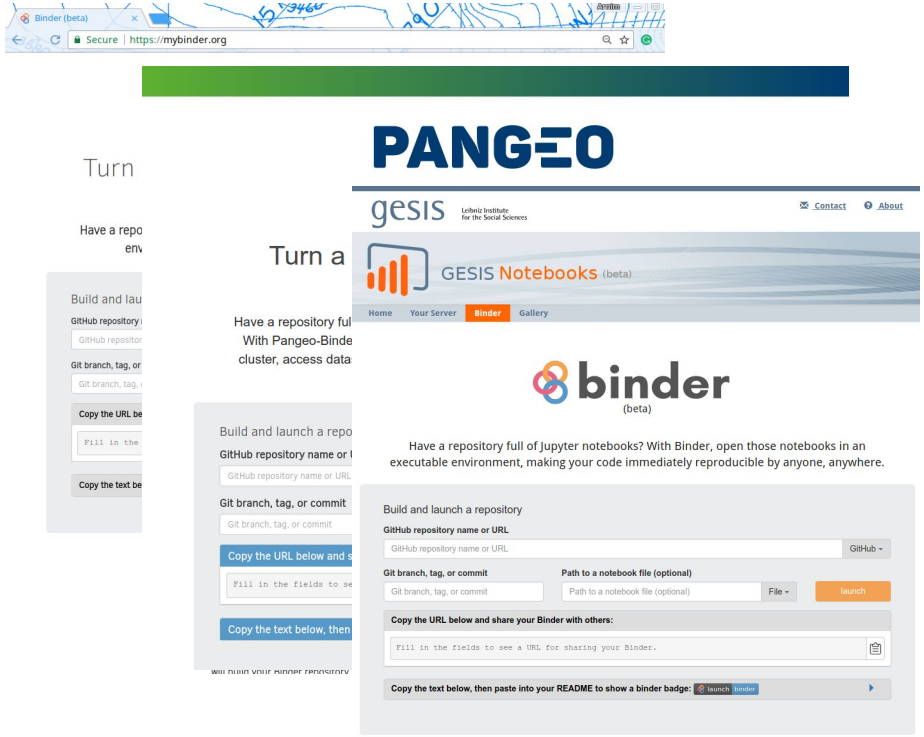


launch binder



launch binder

Deployments



mybinder.org

notebooks.gesis.org

https://mybinder.readthedocs.io/en/latest/about/federation.html

Special thanks to the BinderHub Community

https://github.com/jupyterhub/binderhub/graphs/contributors

and many more who aren't in the GitHub history.

Special thanks to **Tim Head** & **The Turing Way**

for pioneering and sharing training resources

https://build-a-binder.github.io/

https://github.com/alan-turing-institute/the-turing-way/tree/main/workshops

How to **binderize your repository**?

Documentation of the repo2docker Configuration Files
https://repo2docker.readthedocs.io/en/latest/config_files.html

Discourse Jupyter https://discourse.jupyter.org/

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

Working with Jupyter & R Markdown = Jupytext
https://jupytext.readthedocs.io/en/latest/

Our WS demo repository => https://github.com/arnim/ggplot2Demo