

# Tools and Workflows for Reproducible Research in the Quantitative Social Sciences

## Jupyter Notebooks & Binder

Organizers:

Bernd Weiß  
Johannes Breuer  
**Arnim Bleier**

November 18th, 2022

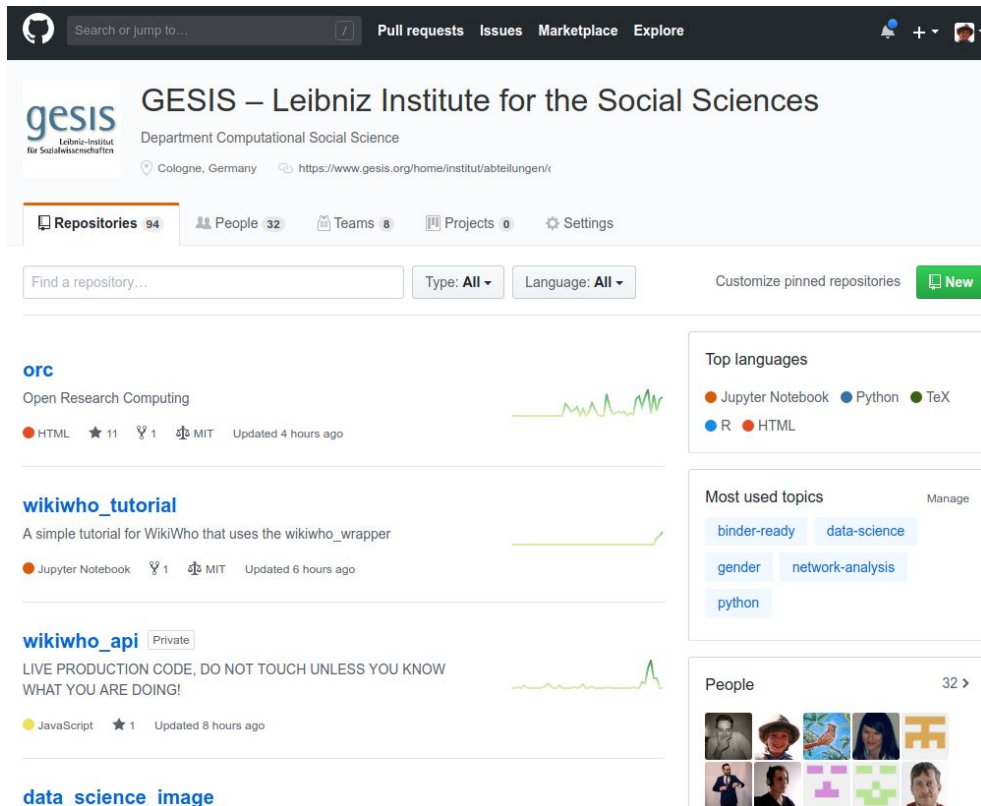
# GESIS Library, Cologne



Mitglied der

Leibniz  
Gemeinschaft

# Open Access Code



The screenshot shows the GitHub profile page for GESIS – Leibniz Institute for the Social Sciences. The header includes the GitHub logo, a search bar, and navigation links for Pull requests, Issues, Marketplace, and Explore. The profile section displays the GESIS logo, the full name of the institute, and its location in Cologne, Germany. Below this, repository statistics are shown: 94 Repositories, 32 People, 8 Teams, and 0 Projects. A search bar for repositories is present, along with filters for Type (All) and Language (All). The main content area lists three repositories: 'orc' (Open Research Computing), 'wikiwho\_tutorial', and 'wikiwho\_api'. Each repository entry includes its description, a language icon (HTML, JavaScript, or Jupyter Notebook), star count, license (MIT), and update time. A green line graph indicates repository activity. On the right side, there are two panels: 'Top languages' showing Jupyter Notebook, Python, TeX, R, and HTML; and 'Most used topics' including binder-ready, data-science, gender, network-analysis, and python. At the bottom right, a 'People' section shows 32 contributors with their profile pictures.

Search or jump to...

Pull requests Issues Marketplace Explore

**gesis** Leibniz-Institut für Sozialwissenschaften

**GESIS – Leibniz Institute for the Social Sciences**

Department Computational Social Science

Cologne, Germany <https://www.gesis.org/home/institut/abteilungen/>

**Repositories** 94 **People** 32 **Teams** 8 **Projects** 0 **Settings**

Find a repository... Type: All Language: All Customize pinned repositories **New**

**orc**  
Open Research Computing  
HTML ★ 11 1 MIT Updated 4 hours ago

**wikiwho\_tutorial**  
A simple tutorial for WikiWho that uses the wikiwho\_wrapper  
Jupyter Notebook 1 MIT Updated 6 hours ago

**wikiwho\_api** Private  
LIVE PRODUCTION CODE, DO NOT TOUCH UNLESS YOU KNOW WHAT YOU ARE DOING!  
JavaScript ★ 1 Updated 8 hours ago

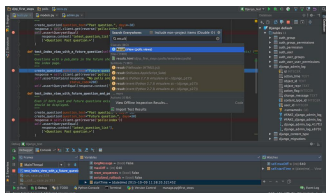
**data\_science\_image**

**Top languages**  
Jupyter Notebook Python TeX  
R HTML

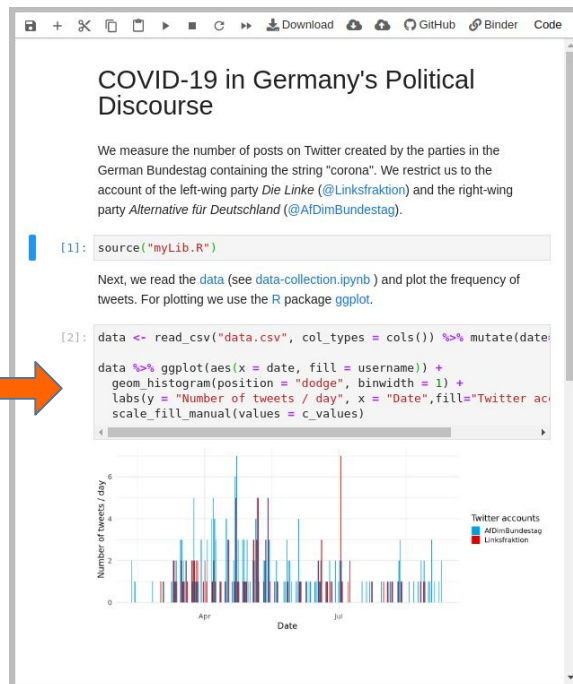
**Most used topics** Manage  
binder-ready data-science  
gender network-analysis  
python

**People** 32 >

# 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



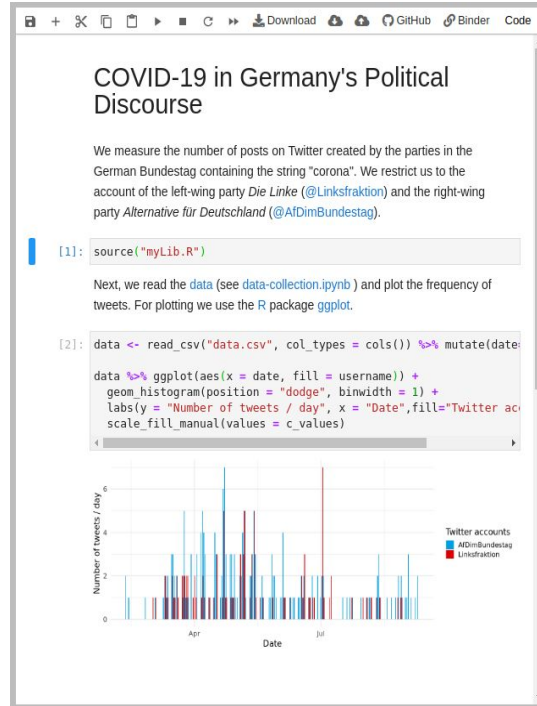
## Cloud:

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

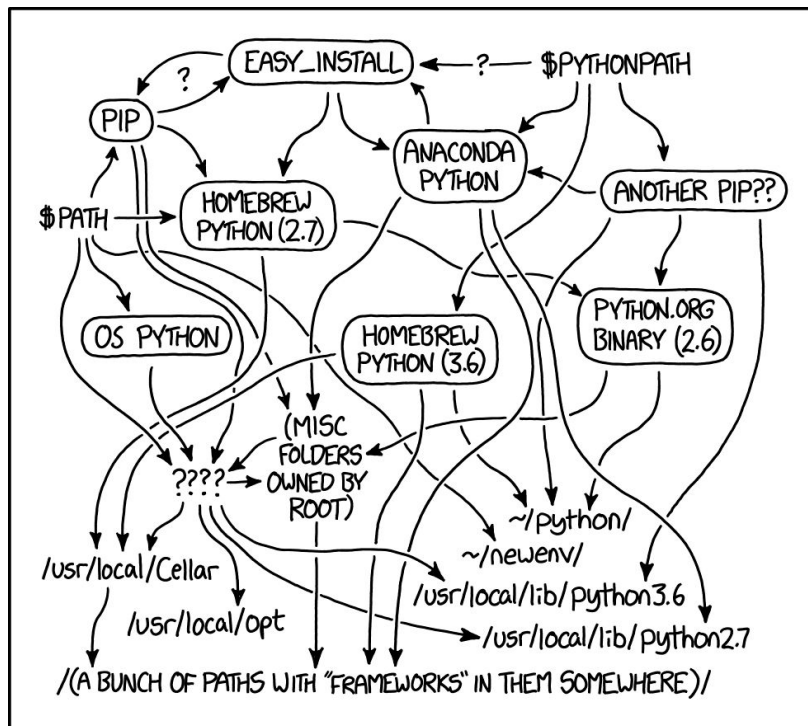


## Personal Computer:

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



# The environment matters



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

# Is “Lockdown” the Solution?



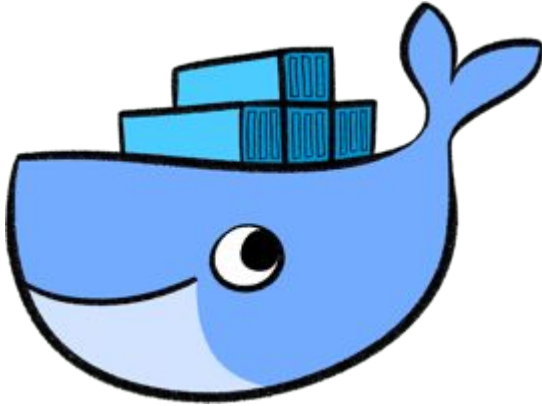
Only the administrators  
control the environment.



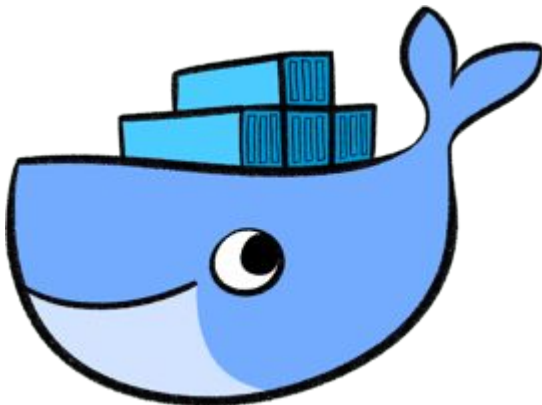
But ... “my Work is Special”



# 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 xvfz apache-couchdb-1.3.1.tar.gz

RUN apt-get install -y make

RUN cd /tmp/apache-couchdb-\* ; ./configure && make install

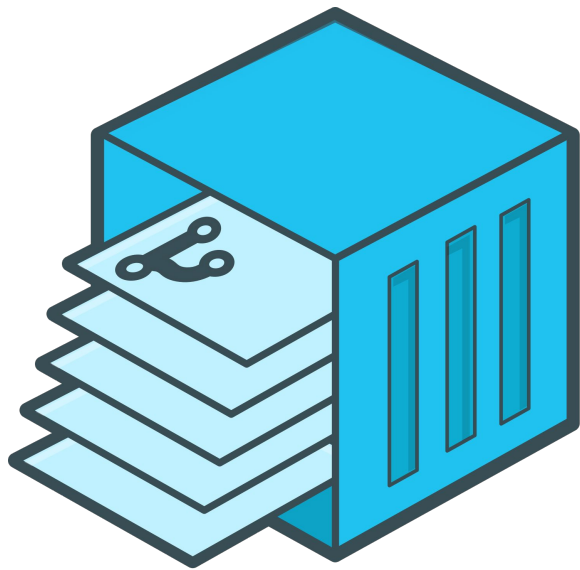
RUN printf "[httpd]\nport = 8101\nbind\_address = 0.0.0.0" >  
/usr/local/etc/couchdb/local.d/docker.ini

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.



# 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

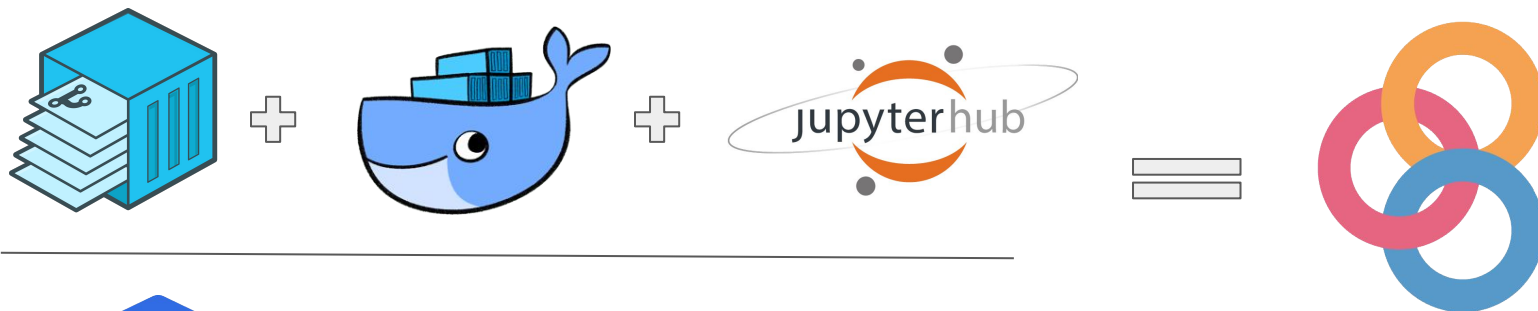
r-2018-07-27



```
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?



# kubernetes

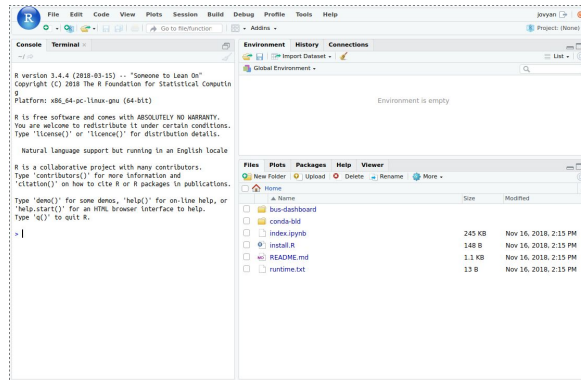
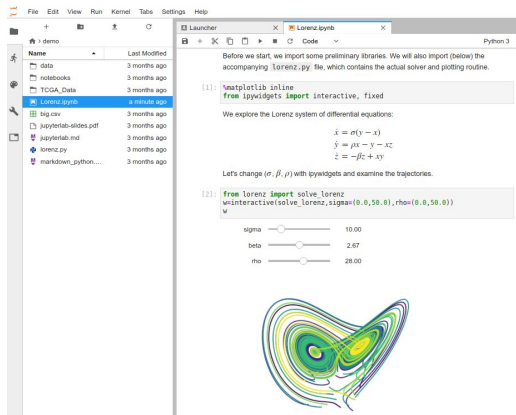


Have a look at the Open Source Project:

<https://github.com/jupyterhub/binderhub/>

Join the Binder chat for questions:

<https://gitter.im/jupyterhub/binder>



# Deployments



[mybinder.org](https://mybinder.org)

[notebooks.gesis.org](https://notebooks.gesis.org)

Turn

Have a repository full of jupyter notebooks? With Pangeo-Binder cluster, access data

Build and launch a repository

GitHub repository name or URL

Git branch, tag, or commit

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:

**PANGEO**

Turn a

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

Git branch, tag, or commit

Path to a notebook file (optional)

Git branch, tag, or commit

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:

 binder (beta)

gesis

Leibniz Institute for the Social Sciences

GESIS Notebooks (beta)

Home Your Server Binder Gallery

Leibniz

Leibniz Gemeinschaft



<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](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 = Jupytertext**

<https://jupytertext.readthedocs.io/en/latest/>

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