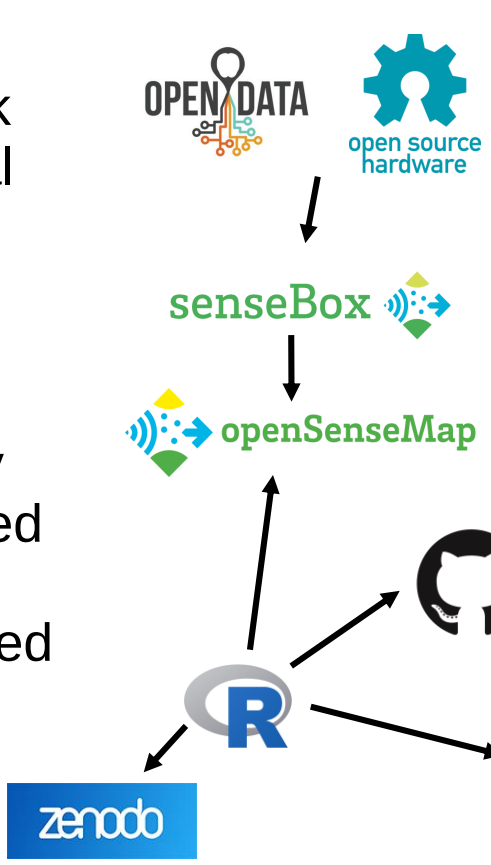
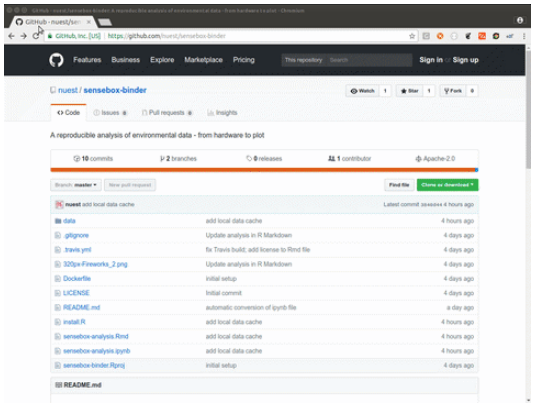
Open Environmental Data Analysis

Daniel Nüst and Thomas Bartoschek, University of Münster, Institute for Geoinformatics, Münster, Germany, {daniel.nuest,bartoschek}@uni-muenster.de

- Overview
- Open
- senseBox
- OSeM
- R Markdown
- Jupyter
- {D,R}ocker
- Binder
- Video
- Abstract + References

Documented, portable, archivable, and one-click interactive computational environments for reproducible research based on Open Source software analysing Open Data measured by Open Hardware published online as interactive website using cloud-based workflows and Linux containers. For free. Awesome.





DOI 10.5281/zenodo.1135139

https://github.com/nuest/sensebox-binder/

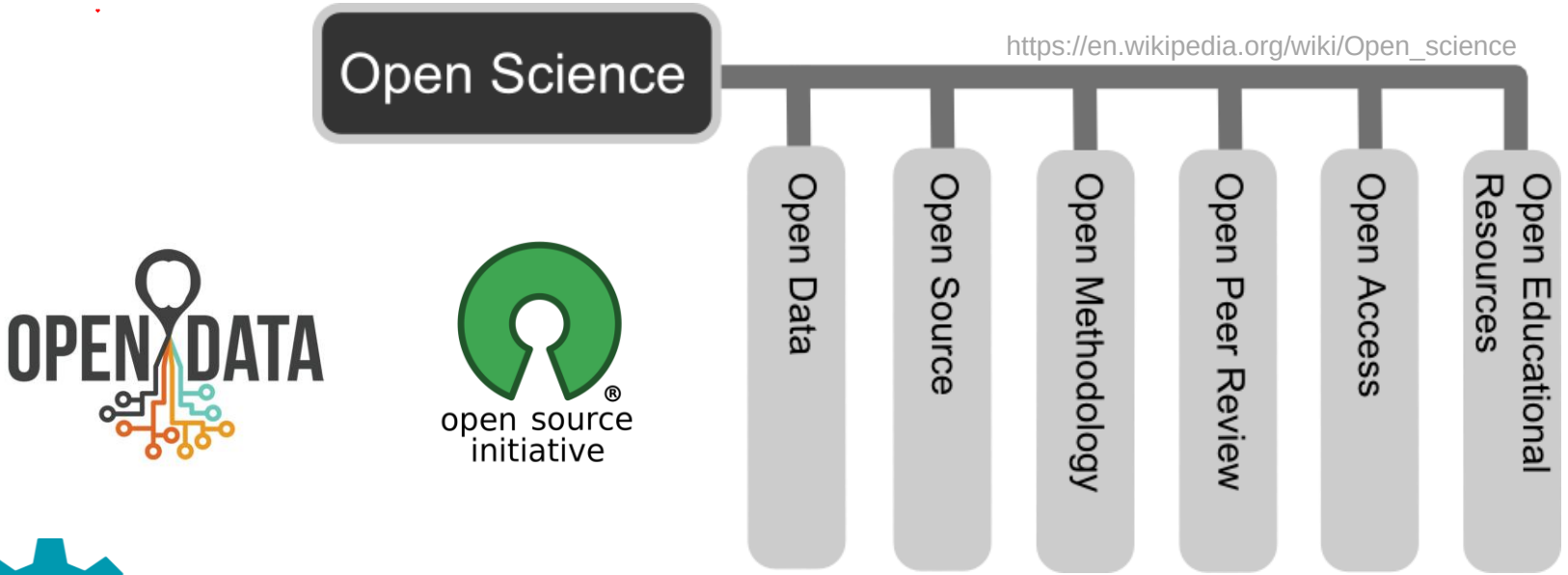
Open Environmental Data Analysis

Daniel Nüst and Thomas Bartoschek
Institute for Geoinformatics
{daniel.nuest,bartoschek}@uni-muenster.de

- Overview
- Open
- senseBox
- OSeM
- R Markdown
- Jupyter
- {D,R}ocker
- Binder
- Video
- Abstract + References

DOI 10.5281/zenodo.1135139

Code Blog



Open Environmental Data Analysis

Daniel Nüst and Thomas Bartoschek
Institute for Geoinformatics
{daniel.nuest,bartoschek}@uni-muenster.de

- Overview
- Open
- senseBox
- OSeM
- R Markdown
- Jupyter
- {D,R}ocker
- Binder
- Video
- Abstract + References

DOI 10.5281/zenodo.1135139

Code Blog

senseBox

https://sensebox.de/en/

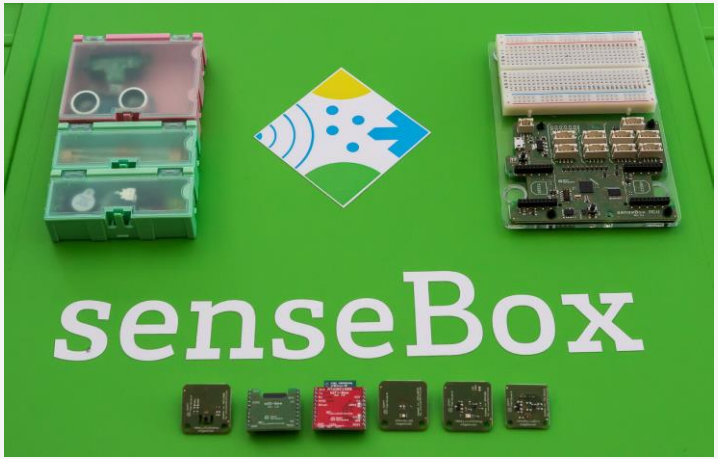
senseBox:home

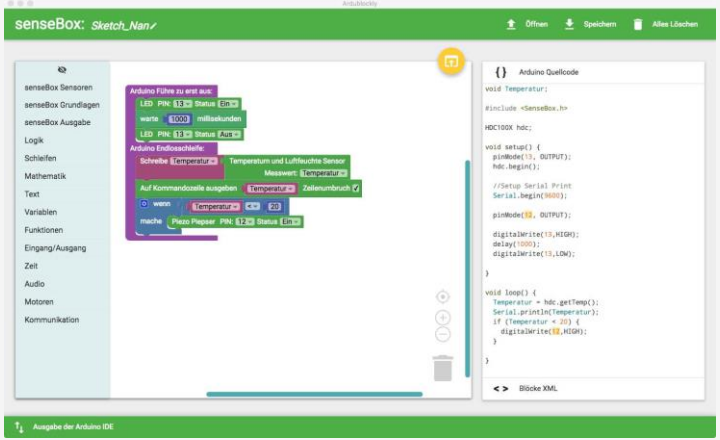
Citizens with senseBox:home can use the technology for their own local research, to collect data by themselves or to contribute to the openSenseMap sensor network.

senseBox:edu

Additionally, we created senseBox:edu for schools and junior scientists, which is an experimental box with didactical concepts, instructions and project ideas included.

Twitter
Facebook
Github





Open Environmental Data Analysis

Daniel Nüst and Thomas Bartoschek
Institute for Geoinformatics
{daniel.nuest,bartoschek}@uni-muenster.de

openSenseMap

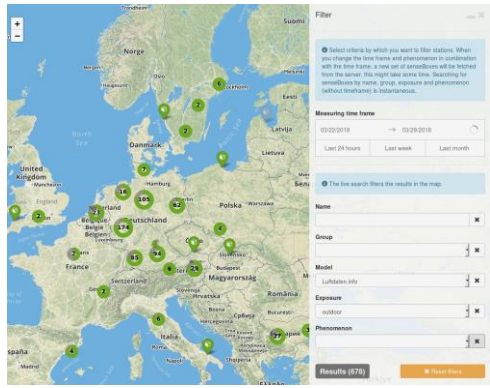
- Overview
- Open
- senseBox
- OSeM
- R Markdown
- Jupyter
- {D,R}ocker
- Binder
- Video
- Abstract + References

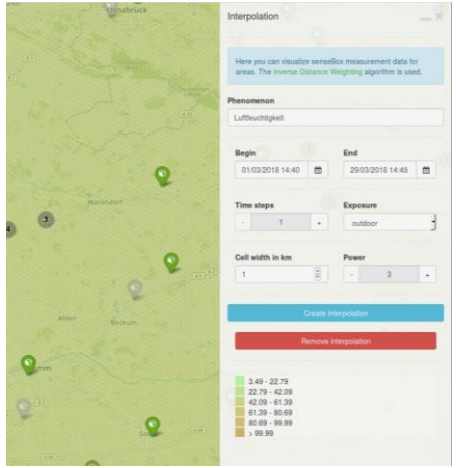
DOI 10.5281/zenodo.1135139

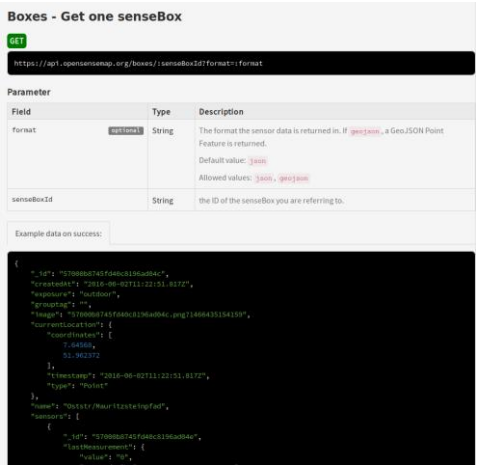
Code Blog

Browse
Analyse
Download
API

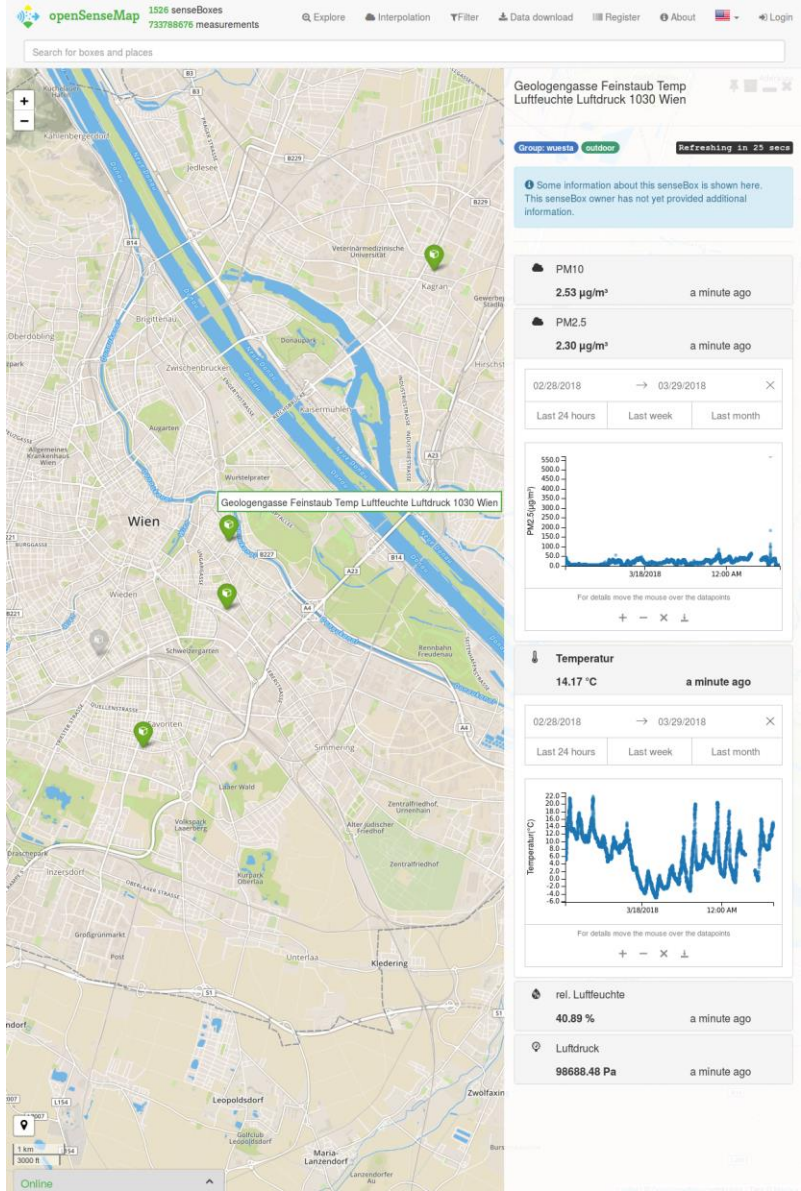
1561 senseBoxes
748029324 measurements







https://opensensemap.org



Open Environmental Data Analysis

Daniel Nüst and Thomas Bartoschek
Institute for Geoinformatics
{daniel.nuest,bartoschek}@uni-muenster.de

R Markdown

- Overview
- Open
- senseBox
- OSeM
- **R Markdown**
- Jupyter
- {D,R}ocker
- Binder
- Video
- Abstract + References

DOI 10.5281/zenodo.1135139

Code Blog

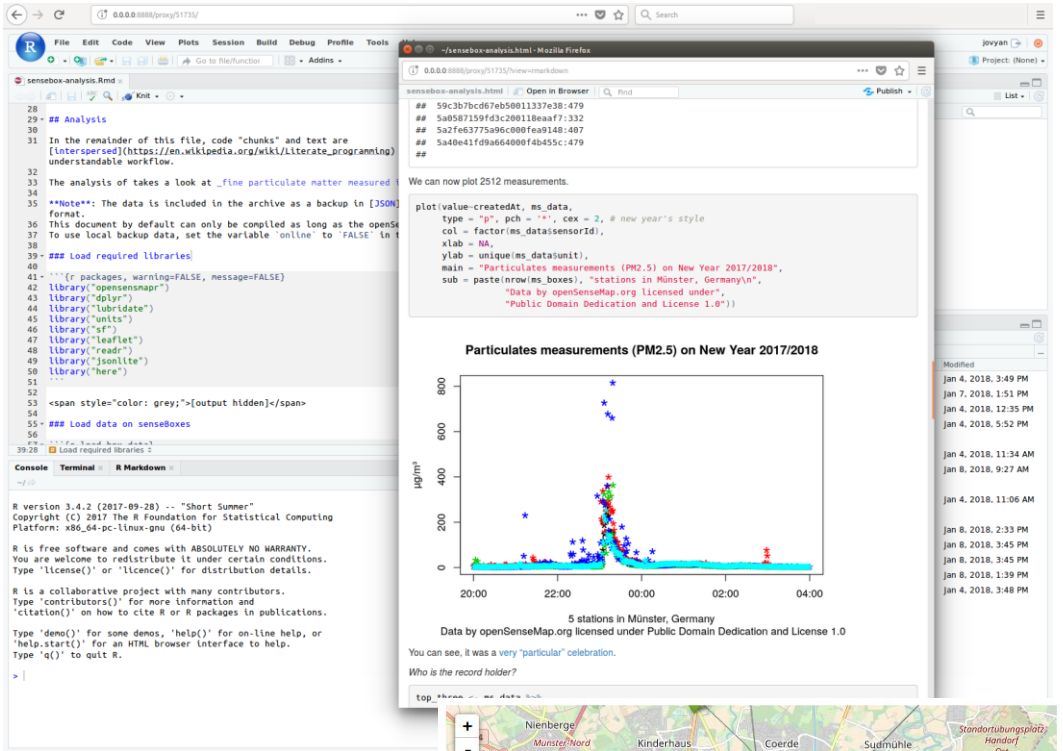

```
### Load data on senseBoxes
```

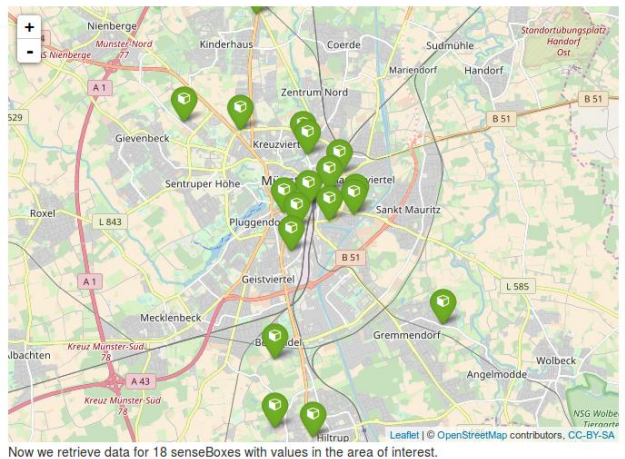
```
```{r load_box_data}
library("opensensmapr")
online <- TRUE # access online API or use local data backup?
analysis_date <- lubridate::as_datetime("2018-01-01 00:00:00")

if (online) {
  # retrieve data from openSenseMap API
  all_boxes <- osem_boxes()
  pm25_boxes <- osem_boxes(
    exposure = 'outdoor',
    date = analysis_date, # ±4 hours
    phenomenon = 'PM2.5'
  )

  # update local data
  all_json <- toJSON(all_boxes, digits = NA, pretty = TRUE)
  write(all_json, file = here("data/all_boxes.json"))
  pm25_json <- toJSON(pm25_boxes, digits = NA, pretty = TRUE)
  write(pm25_json, file = here("data/pm25_boxes.json"))
} else {
  # load data from file and fix column types
  all_boxes_file <- fromJSON(here("data/all_boxes.json"))
```

```
library("leaflet")
leaflet() %>%
  addTiles() %>%
  addAwesomeMarkers(data = ms_boxes,
             popup = ~paste0("<b>Name:</b> ", name, "<br><b>Id:</b> ",
                             "<a href='https://opensensemap.org/explore/", X_id, "' ",
                             "target='_blank'>", X_id, "</a>"),
             label = ~name,
             icon = sense_icon)
```





Now we retrieve data for 18 senseBoxes with values in the area of interest.

Rendered by Travis CI & published on GitHub Pages:
https://nuest.github.io/sensebox-binder/sensebox-analysis.html
```

Open Environmental Data Analysis

Daniel Nüst and Thomas Bartoschek
Institute for Geoinformatics
{daniel.nuest,bartoschek}@uni-muenster.de

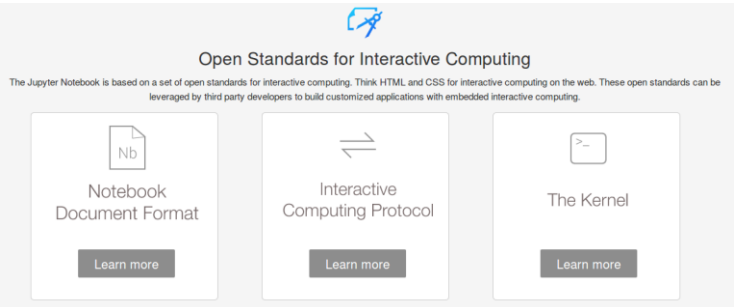
DOI 10.5281/zenodo.1135139

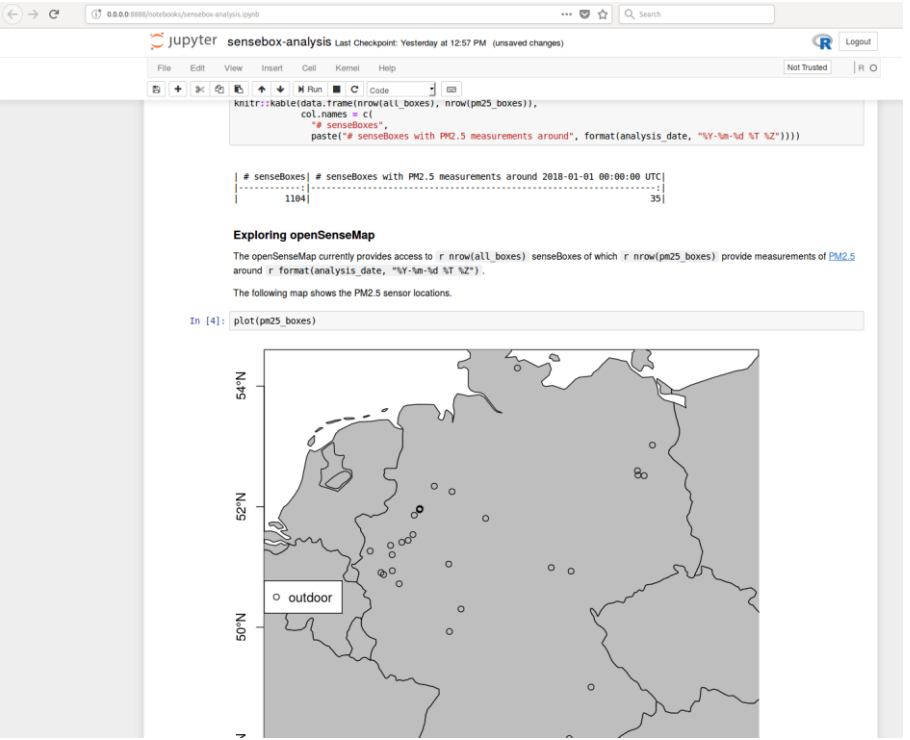
Code **Blog**

The Jupyter Notebook

The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.

https://jupyter.org





Synchonise R Markdown and Jupyter Notebook with **ipyrmd** on every Rmd save:

ipyrmd --to ipynb --from Rmd -y -o sensebox-analysis.ipynb sensebox-analysis.Rmd

https://github.com/chronitis/ipyrmd

Daniel Nüst and Thomas Bartoschek
Institute for Geoinformatics
{daniel.nuest,bartoschek}@uni-muenster.de

DOI 10.5281/zenodo.1135139

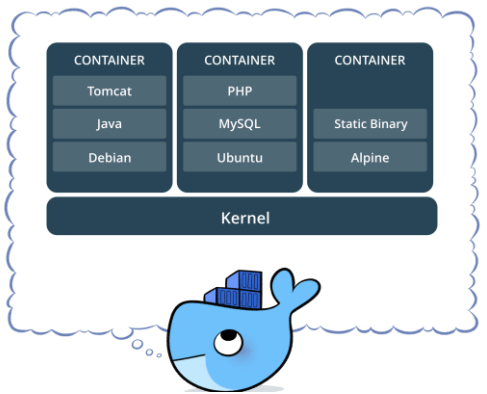
Code Blog

Docker & Rocker

https://github.com/rocker-org/binder/blob/master/Dockerfile

```
1  FROM rocker/geospatial:3.4.2
16  ENV HOME /home/${NB_USER}
17  WORKDIR ${HOME}
18
19  RUN apt-get update && \
20      apt-get -y install python3-venv python3-dev && \
45  CMD jupyter notebook --ip 0.0.0.0

1  FROM rocker/binder:3.4.2
2
3  # Copy repo into ${HOME}, make user own $HOME
4  USER root
5  COPY . ${HOME}
6  RUN chown -R ${NB_USER} ${HOME}
7  USER ${NB_USER}
8
9  ## run any install.R script we find
10 RUN if [ -f install.R ]; then R --quiet -f install.R; fi
```



https://docker.com

Documented, portable, archivable, and one-click interactive **computational environments** for reproducibility based on `Dockerfiles`

The versioned stack

image	description
r-ver	Specify R version in docker tag. Builds on `debian:stable`
rstudio	Adds rstudio
tidyverse	Adds tidyverse & devtools
verse	Adds tex & publishing-related packages
geospatial	Adds geospatial libraries

https://www.rocker-project.org/images/

```
daniel@gin-nuest:~$ docker pull rocker/r-ver:3.1.0
3.1.0: Pulling from rocker/r-ver
4176fe04cefe: Already exists
50c17ea00d6a: Pull complete
Digest: sha256:ca237b5b5f23cbb7465ff60777acce0de8d3293079541ff36814e5c7e730f03c
Status: Downloaded newer image for rocker/r-ver:3.1.0
daniel@gin-nuest:~$ docker run -it rocker/r-ver:3.1.0

R version 3.1.0 (2014-04-10) -- "Spring Dance"
Copyright (C) 2014 The R Foundation for Statistical Computing
Platform: x86_64-unknown-linux-gnu (64-bit)

> getOption("repos")
                                                    CRAN
"https://mran.microsoft.com/snapshot/2014-09-17"
> 1 + 1
[1] 2
>
```

https://www.rocker-project.org

Open Environmental Data Analysis

Daniel Nüst and Thomas Bartoschek
Institute for Geoinformatics
{daniel.nuest,bartoschek}@uni-muenster.de

- Overview
- Open
- senseBox
- OSeM
- R Markdown
- Jupyter
- {D,R}ocker
- Binder
- Video
- Abstract + References

DOI 10.5281/zenodo.1135139

Code Blog

https://mybinder.org

binder (beta)

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.

Build and launch a repository

GitHub repo or URL

GitHub repository name or link

Git branch, tag, or commit: master

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

Binder Team @mybinderteam · 27. März

A brief #humblebrag: over the last 30 days mybinder.org has served >100000 sessions to ~72000 distinct users from 169 countries! 😎

https://twitter.com/mybinderteam/status/978603170250330112

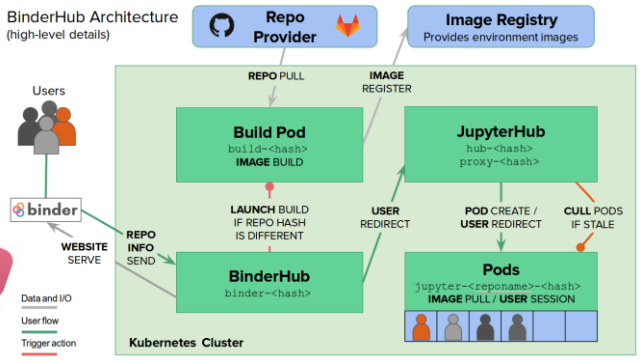
launch binder

How it works

1. **Enter your repository information**
   Provide in the above form a URL or a GitHub repository that contains Jupyter notebooks, as well as a branch, tag, or commit hash. Launch will build your Binder repository. If you specify a path to a notebook file, the notebook will be opened in your browser after building.

2. **We build a Docker image of your repository**
   Binder will search for a dependency file, such as requirements.txt or environment.yml, in the repository's root directory (more details on more complex dependencies in documentation). The dependency files will be used to build a Docker image. If an image has already been built for the given repository, it will not be rebuilt. If a new commit has been made, the image will automatically be rebuilt.

3. **Interact with your notebooks in a live environment!**
   A JupyterHub server will host your repository's contents. We offer you a reusable link and badge to your live repository that you can easily share with others.



Open Environmental Data Analysis

Daniel Nüst and Thomas Bartoschek
Institute for Geoinformatics
{daniel.nuest,bartoschek}@uni-muenster.de

DOI 10.5281/zenodo.1135139

Code Blog

What happens?

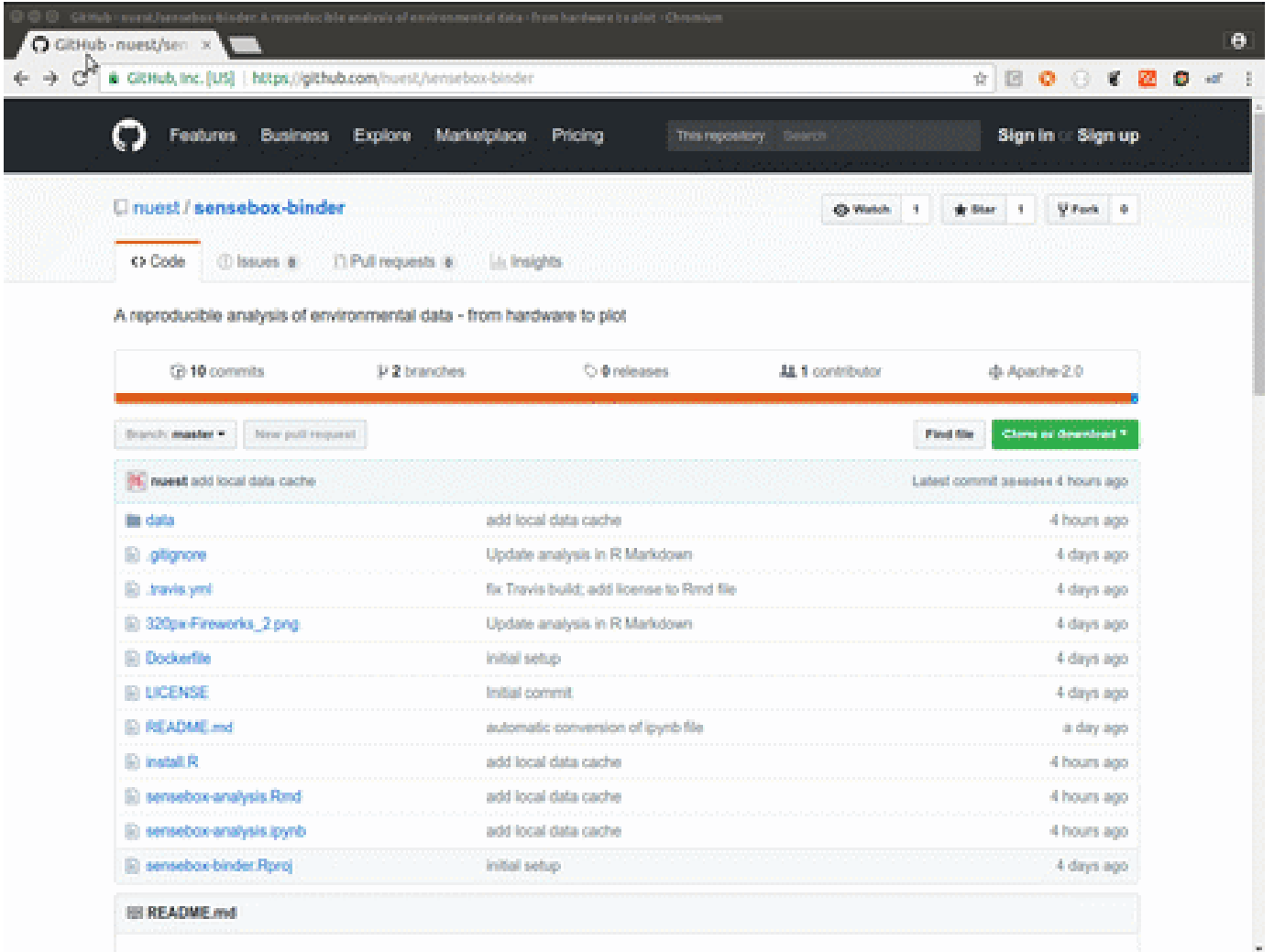
Click on “Launch binder”

Open Jupyter Hub

Switch to Rstudio

Render the literate programming document

View workflow, plots, interactive map



https://media.giphy.com/media/l49JRjO65S0WQ1Kyk/giphy.gif

Open Environmental Data Analysis

Daniel Nüst and Thomas Bartoschek
Institute for Geoinformatics
{daniel.nuest,bartoschek}@uni-muenster.de

Overview
Open
senseBox
OSeM
R Markdown
Jupyter
{D,R}ocker
Binder
Video
Abstract + References

DOI 10.5281/zenodo.1135139

Code Blog

Abstract & References

Open Science is a major movement. It is deeply connected with other initiatives on openness, availability, and trans-parency: Open Access, Open Data, Reproducible Research, Free and Open Source Software (FOSS), and Open-source Hardware (OSH, https://en.wikipedia.org/wiki/Open-source_hardware). Free to use platforms supporting Open Science become more recognised and many are FOSS themselves. They range from large, publicly-fundedplayers such as Zenodo (https://zenodo.org), to community projects with first grant funding such as Binder (Pérezet al., 2017).

The senseBox community (https://sensebox.de/en) provides a Do-It-Yourself Toolkit for Citizen Sci-ence as Open Hardware and an Open Data API (https://opensensemap.org) for collecting, sharing, and visualisingobservations, including particulates (Pfeil et. al., 2015). Both approaches were designed following usability engineering methods to enable broad use, i.e. in education or by non-professional users (Wirwahn Bartoschek, 2015).

We present an open environmental data analysis building on all these projects. A central R Markdown (http://rmarkdown.rstudio.com/) document provides a recipe for data retrieval (with backup in plain-text files), analysis, and visualisation with R (https://r-project.org).

The document was created in a containerised environ-ment with Docker (https://docker.com), more specifically the Rocker (Boettiger, 2017) image rocker/binder (https://hub.docker.com/r/rocker/binder/) which provides Jupyter Notebook (https://jupyter.org/) and Rstudio (http://rstudio.com/) user interfaces.

The container's advantages are (a) portability, (b) completeness using con-figuration files and scripts, (c) at least limited archivability, (d) simple integration of browser-based interac-tive user interfaces, and (e) one click online execution with Binder. All resources are published on GitHub (https://github.com/nuest/sensebox-binder) and the Zenodo research data repository (Nüst, 2017).

We demonstrate an open analysis workflow for environmental data utilising numerous Open X projects. It showcases the potential of latest technology and may serve as a template towards development of best practices, but alsoprovides insight into remaining stumbling blocks. Connecting all these parts requires some technical proficiency. In the future reproducible computational workflows should become part of researcher's education, making themaware of simple yet effective practices (e.g. text-based formats, scripted analyses). In the meantime templates and semi-automatic assistants (cf. Nüst et al., 2017) can mitigate challenges.

https://meetingorganizer.copernicus.org/EGU2018/EGU2018-17461.pdf

Geophysical Research Abstracts
Vol. 20, EGU2018-17461
EGU General Assembly 2018

Boettiger, C. and Eddelbuettel, D., 2017. An Introduction to Rocker: Docker Containers for R. The R Journal, RJ-2017-065. url

Nüst, D., 2018. Open environmental data analysis with senseBox, openSenseMap, Jupyter Notebook, RStudio, and BinderHub (Version v2). Zenodo. doi: 10.5281/zenodo.1139929

Nüst, D., Konkol, M., Pebesma, E., Kray, C., Schutzeichel, M., Przibytzin, H., Lorenz, J. Opening the Publication Process with Executable Research Compendia. D-Lib Magazine. 2017. doi: 10.1045/january2017-nuest

Pérez, F., Granger, B., and Ragan-Kelley, B., 2017. Binder: enabling sharing and publication of reproducible computational research. https://figshare.com/s/e9d0ad7bdc4e405cccfa

Pfeil, M., Bartoschek, T., and Wirwahn, J.,2015. OpenSenseMap – A Citizen Science Platform for Publishing and Exploring Sensor Data as Open Data. FOSS4G Conference Proceedings, pp. 122-139, Seoul, 2015. url

Wirwahn, J. and Bartoschek, T., 2015. Usability Engineering for Successful Open Citizen Science. FOSS4G Conference Proceedings, pp. 68-78, Seoul, 2015. url