

21  
2C

The International  
Interactive Computing  
Collaboration

# Reproducible Computational Environments with Binder

Sarah Gibson  
she/her

[doi.org/10.5281/zenodo.10075621](https://doi.org/10.5281/zenodo.10075621)



Hello, 🙌  
My name is  
Sarah Gibson

I'm an **open source infrastructure engineer** at 2i2c.

My research background is Astrophysics 🌌

I am a JupyterHub team member and support the infrastructure that powers mybinder.org

I am also a core team member of *The Turing Way* and hold a fellowship from the Software Sustainability Institute, advocating for best software practices in research

# Why do I care?

Irreproducible research  
wastes time!



# What does Reproducibility mean?

		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

# What does Reproducibility mean?

		Data	
		Same	Different
Analysis	Same	Repeatable Reproducible	Replicable
	Different	Robust	Generalisable

# Academic errors have real world effects

	B	C	I	J	K	L	M
2			Real GDP growth				
3			Debt/GDP				
4	Country	Coverage	30 or less	30 to 60	60 to 90	90 or above	30 or less
26			3.7	3.0	3.5	1.7	5.5
27	Minimum		1.6	0.3	1.3	-1.8	0.8
28	Maximum		5.4	4.9	10.2	3.6	13.3
29							
30	US	1946-2009	n.a.	3.4	3.3	-2.0	n.a.
31	UK	1946-2009	n.a.	2.4	2.5	2.4	n.a.
32	Sweden	1946-2009	3.6	2.9	2.7	n.a.	6.3
33	Spain	1946-2009	1.5	3.4	4.2	n.a.	9.9
34	Portugal	1952-2009	4.8	2.5	0.3	n.a.	7.9
35	New Zealand	1948-2009	2.5	2.9	3.9	-7.9	2.6
36	Netherlands	1956-2009	4.1	2.7	1.1	n.a.	6.4
37	Norway	1947-2009	3.4	5.1	n.a.	n.a.	5.4
38	Japan	1946-2009	7.0	4.0	1.0	0.7	7.0
39	Italy	1951-2009	5.4	2.1	1.8	1.0	5.6
40	Ireland	1948-2009	4.4	4.5	4.0	2.4	2.9
41	Greece	1970-2009	4.0	0.3	2.7	2.9	13.3
42	Germany	1946-2009	3.9	0.9	n.a.	n.a.	3.2
43	France	1949-2009	4.9	2.7	3.0	n.a.	5.2
44	Finland	1946-2009	3.8	2.4	5.5	n.a.	7.0
45	Denmark	1950-2009	3.5	1.7	2.4	n.a.	5.6
46	Canada	1951-2009	1.9	3.6	4.1	n.a.	2.2
47	Belgium	1947-2009	n.a.	4.2	3.1	2.6	n.a.
48	Austria	1948-2009	5.2	3.3	-3.8	n.a.	5.7
49	Australia	1951-2009	3.2	4.9	4.0	n.a.	5.9
50							
51			4.1	2.8	2.8	=AVERAGE(L30:L44)	

# Academic errors have real world effects

	B	C	I	J	K	L	M
2			Real GDP growth				
3			Debt/GDP				
4	Country	Coverage	30 or less	30 to 60	60 to 90	90 or above	30 or less
26			3.7	3.0	3.5	1.7	5.5
27	Minimum		1.6	0.3	1.3	-1.8	0.8
28	Maximum		5.4	4.9	10.2	3.6	13.3
29							
30	US	1946-2009	n.a.	3.4	3.3	-2.0	n.a.
31	UK	1946-2009	n.a.	2.4	2.5	2.4	n.a.
32	Sweden	1946-2009	3.6	2.9	2.7	n.a.	6.3
33	Spain	1946-2009	1.5	3.4	4.2	n.a.	9.9
34	Portugal	1952-2009	4.8	2.5	0.3	n.a.	7.9
35	New Zealand	1948-2009	2.5	2.9	3.9	-7.9	2.6
36	Netherlands	1956-2009	4.1	2.7	1.1	n.a.	6.4
37	Norway	1947-2009	3.4	5.1	n.a.	n.a.	5.4
38	Japan	1946-2009	7.0	4.0	1.0	0.7	7.0
39	Italy	1951-2009	5.4	2.1	1.8	1.0	5.6
40	Ireland	1948-2009	4.4	4.5	4.0	2.4	2.9
41	Greece	1970-2009	4.0	0.3	2.7	2.9	13.3
42	Germany	1946-2009	3.9	0.9	n.a.	n.a.	3.2
43	France	1949-2009	4.9	2.7	3.0	n.a.	5.2
44	Finland	1946-2009	3.8	2.4	5.5	n.a.	7.0
45	Denmark	1950-2009	3.5	1.7	2.4	n.a.	5.6
46	Canada	1951-2009	1.9	3.6	4.1	n.a.	2.2
47	Belgium	1947-2009	n.a.	4.2	3.1	2.6	n.a.
48	Austria	1948-2009	5.2	3.3	-3.8	n.a.	5.7
49	Australia	1951-2009	3.2	4.9	4.0	n.a.	5.9
50							
51			4.1	2.8	2.8	=AVERAGE(L30:L44)	

**BBC** Sign in News Sport Weather iPlayer Search

## NEWS

Home UK World Business Politics Tech Science Health Family & Education

### Magazine

#### Reinhart, Rogoff... and Herndon: The student who caught out the profs

By Ruth Alexander  
BBC News

© 20 April 2013

f t Share

This week, economists have been astonished to find that a famous academic paper often used to make the case for austerity cuts contains major errors. Another surprise is that the mistakes, by two eminent Harvard professors, were spotted by a student doing his homework.

On 4 January 2010, the Munich Iliad in Atlanta. At the annual meeting of the American Economic Association, Professor Carmen Reinhart and the former chief economist of the International Monetary Fund, Ken Rogoff, are presenting a research paper called 'Growth in a Time of Debt'.



<https://statmodeling.stat.columbia.edu/2013/04/16/memo-to-reinhart-and-rogooff-i-think-its-best-to-admit-your-errors-and-go-on-from-there>  
<https://www.bbc.co.uk/news/magazine-22223190>

[doi.org/10.5281/zenodo.10075621](https://doi.org/10.5281/zenodo.10075621)

# Humans are the hardest part of reproducibility



SPRINGER NATURE



Is not considered  
for promotion

Held to higher  
standards than  
others

Publication  
bias towards  
novel findings

# Barriers to reproducible research

Requires  
additional skills

Plead the 5th

Support  
additional users

Takes time

# What actually goes into reproducible research?



Scriberia 

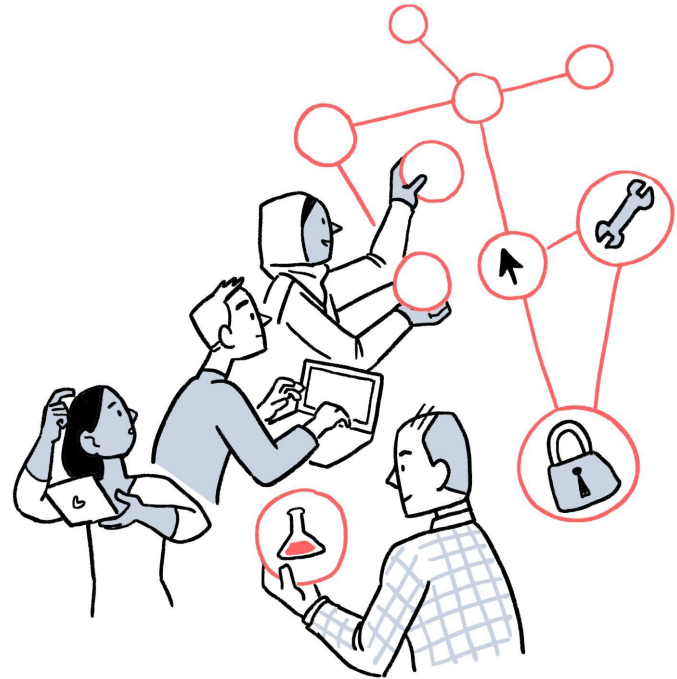
# What goes into reproducible research?

- Version control
  - Snapshots in time



# What goes into reproducible research?

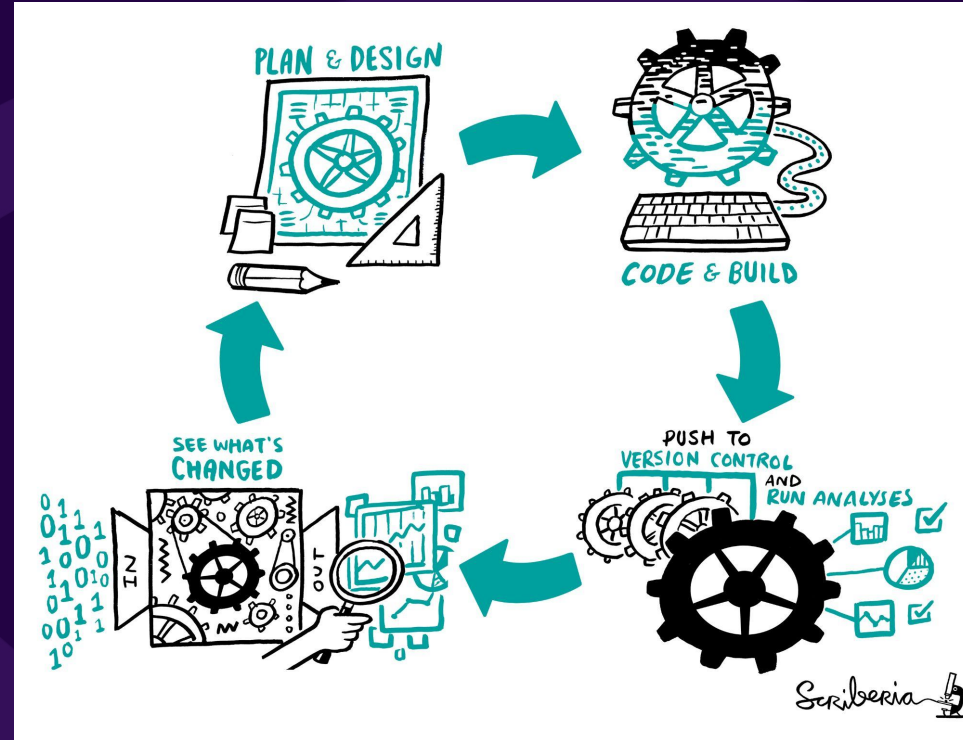
- Version control
  - Snapshots in time
- Testing
  - What's changed?



Scriberia 

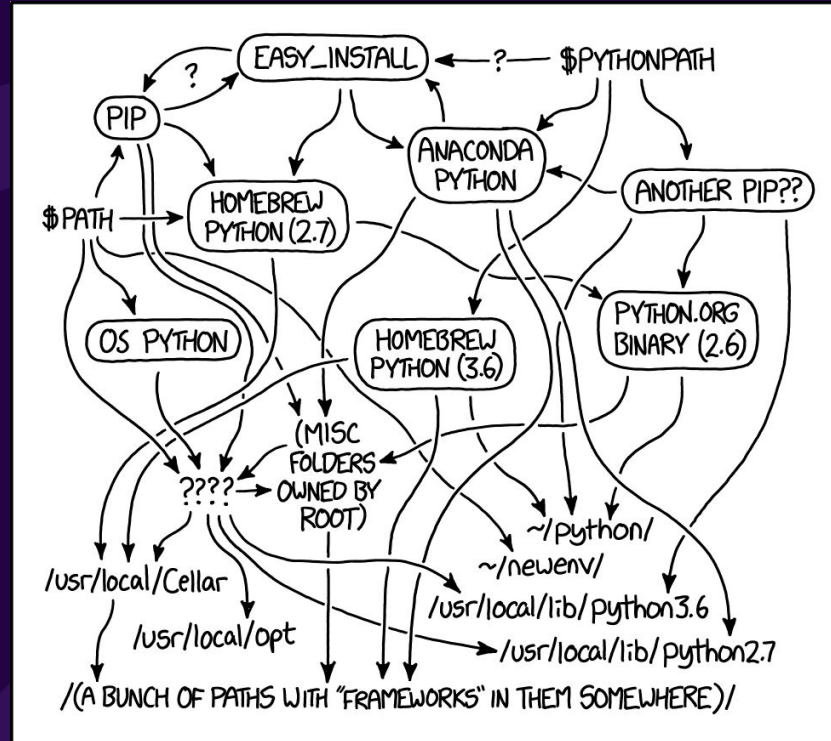
# What goes into reproducible research?

- Version control
  - Snapshots in time
- Testing
  - What's changed?
- Continuous Integration
  - Automatically test each change as you make it
  - Be explicit about what's changed!



# What goes into reproducible research?

- Version control
  - Snapshots in time
- Testing
  - What's changed?
- Continuous Integration
  - Automatically test each change as you make it
  - Be explicit about what's changed!
- Software environment management



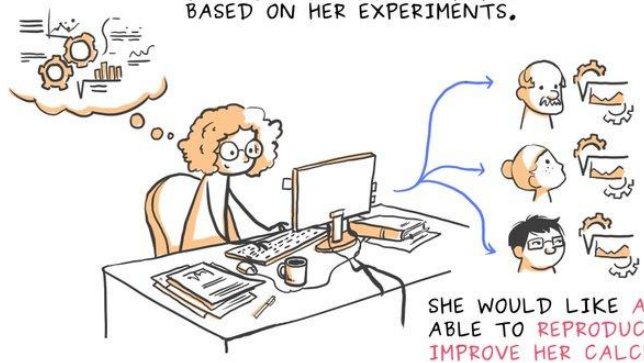
# Project Binder

provides a computational environment without the toil of installing packages





JANE HAS WRITTEN A PAPER BASED ON HER EXPERIMENTS.



SHE WOULD LIKE ANYONE TO BE ABLE TO REPRODUCE, CHECK, AND IMPROVE HER CALCULATIONS

STEP 1

SHE DESCRIBES THE EXPERIMENTS AS A Jupyter NOTEBOOK, MIXING:

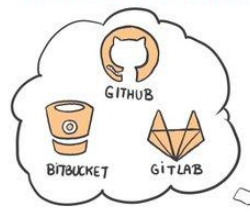


PROSE CODE & VISUALIZATION AND RESOURCES: SOURCE CODE, DATA, MEDIA...

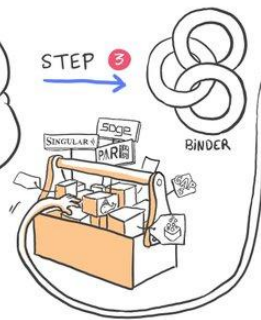
STEP 2

SHE PUBLISHES THEM ON A PUBLICLY HOSTED REPOSITORY

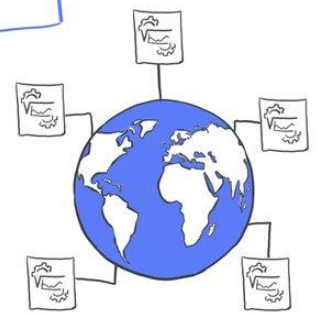
SHE MAKES THAT REPOSITORY BINDER-READY BY DESCRIBING THE SOFTWARE REQUIRED TO RUN THE NOTEBOOK



STEP 3



- CONFIGURATION ✓
- NOTEBOOK ✓
- RESOURCES ✓

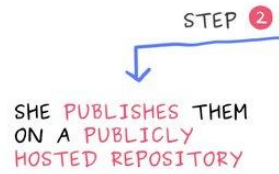
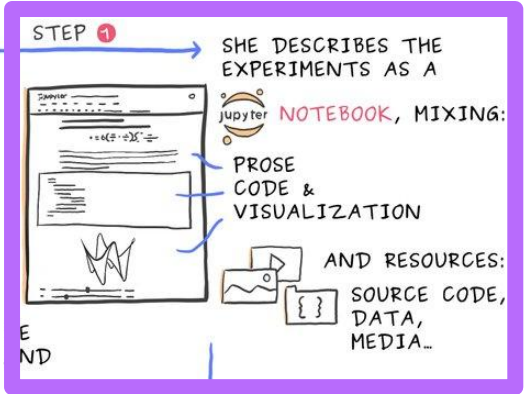


EVERYONE CAN NOW RUN AND REPRODUCE HER COMPUTATIONS

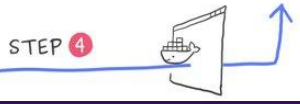
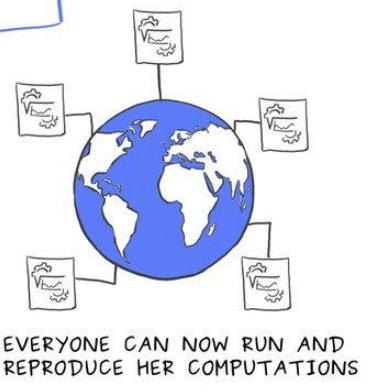
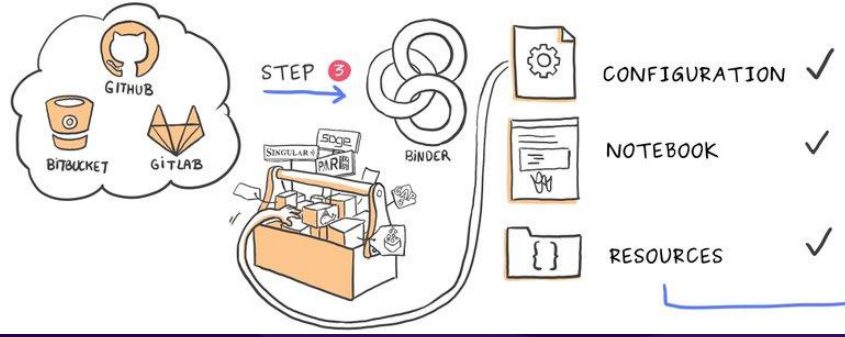
STEP 4

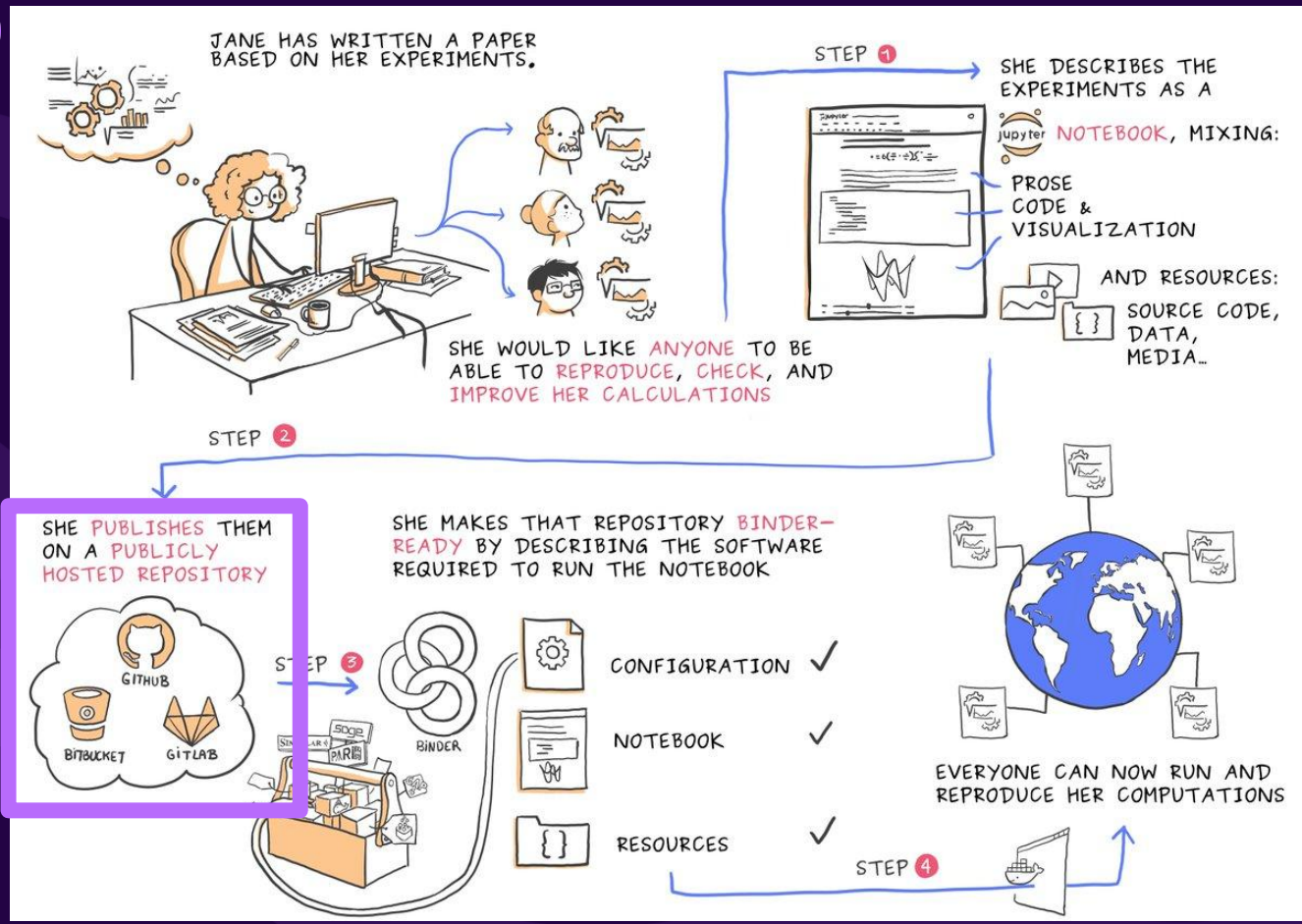


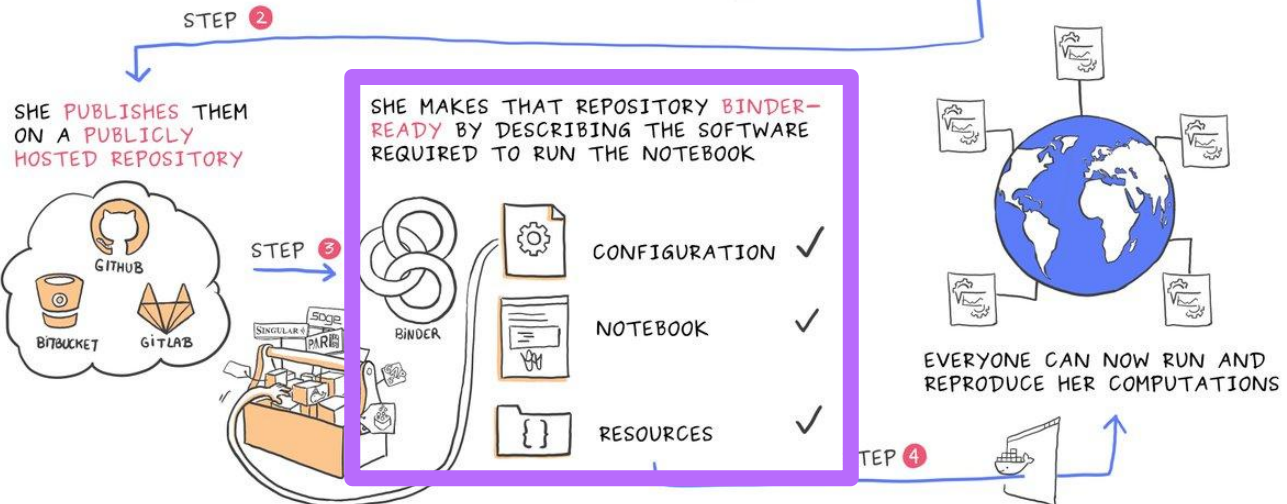
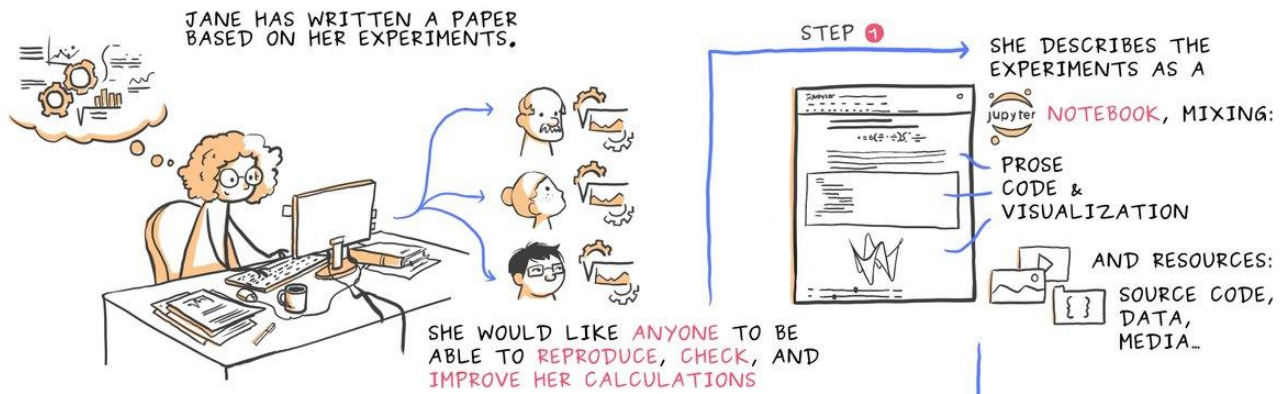




SHE MAKES THAT REPOSITORY BINDER-READY BY DESCRIBING THE SOFTWARE REQUIRED TO RUN THE NOTEBOOK



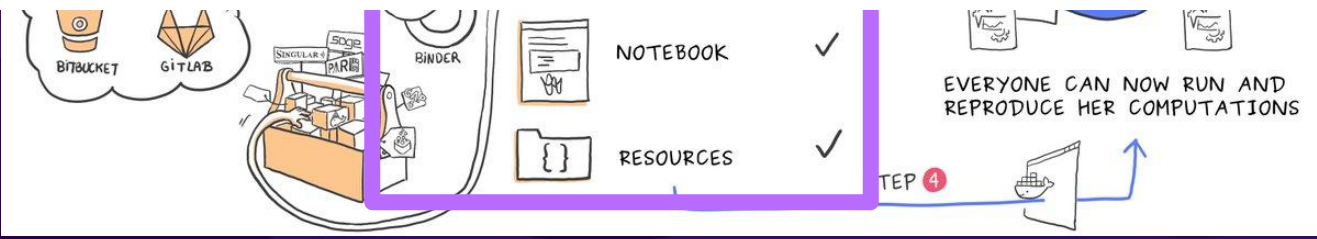




choldgraf Update requirements.txt 21a328d on 21 Jun 2 contributors

5 lines (3 sloc) | 46 Bytes Raw Blame History

```
1 numpy==1.16.*
2 matplotlib==3.*
3 seaborn==0.8.1
4
```



&lt;&gt; Code

! Issues 2

Pull requests 0

Projects 0

Wiki

Security

Insights

Branch: master ▾

conda / environment.yml

Find file

Copy path

 betatim Update environment.yml

89dd429 on 11 Dec 2018

4 contributors



14 lines (13 sloc) | 161 Bytes

Raw

Blame

History



```
1 name: example-environment
2 channels:
3   - conda-forge
4 dependencies:
5   - numpy
6   - psutil
7   - toolz
8   - matplotlib
9   - dill
10  - pandas
11  - partd
12  - bokeh
13  - dask
```



<> Code Issues 0 Pull requests 0 Projects 0 Wiki Security Insights

Branch: master binder-r-description / DESCRIPTION

Find file Copy path

gedankenstuecke first commit 70f8b8e on 18 Sep 2018

1 contributor

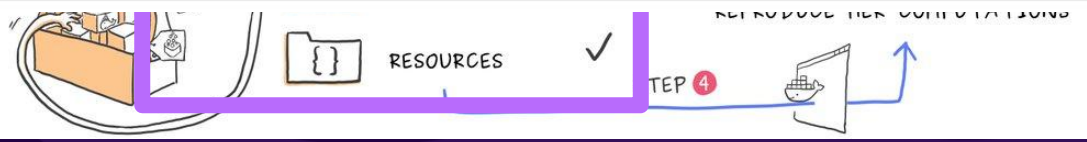
8 lines (7 sloc) 282 Bytes

Raw Blame History

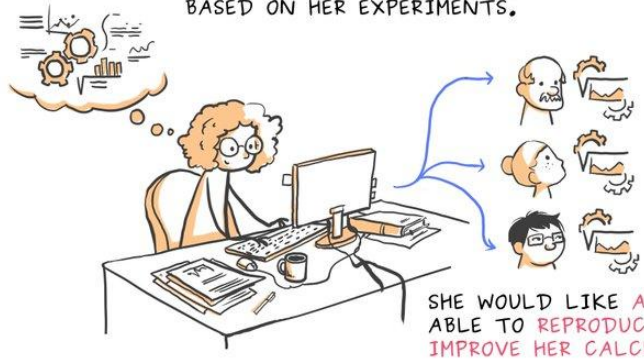
```

1 Package: binderdescription
2 Version: 0.1
3 Date: 2018-09-18
4 Title: Binder R DESCRIPTION support
5 Description: Test that automatically building R packages works
6 Author: Bastian Greshake Tzovaras <bgresshake@googlemail.com>
7 Maintainer: Bastian Greshake Tzovaras <bgresshake@googlemail.com>

```



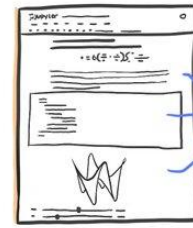
JANE HAS WRITTEN A PAPER BASED ON HER EXPERIMENTS.



SHE WOULD LIKE ANYONE TO BE ABLE TO REPRODUCE, CHECK, AND IMPROVE HER CALCULATIONS

STEP 1

SHE DESCRIBES THE EXPERIMENTS AS A Jupyter NOTEBOOK, MIXING:



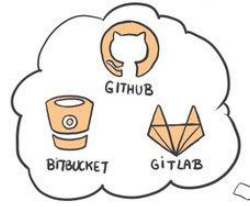
PROSE CODE & VISUALIZATION

AND RESOURCES: SOURCE CODE, DATA, MEDIA...

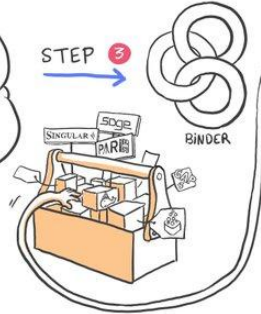
STEP 2

SHE PUBLISHES THEM ON A PUBLICLY HOSTED REPOSITORY

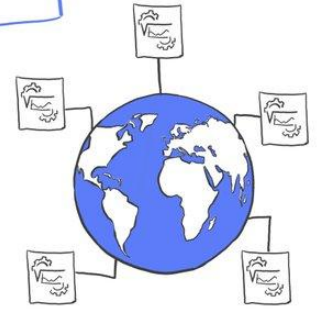
SHE MAKES THAT REPOSITORY BINDER-READY BY DESCRIBING THE SOFTWARE REQUIRED TO RUN THE NOTEBOOK



STEP 3



- CONFIGURATION ✓
- NOTEBOOK ✓
- RESOURCES ✓



EVERYONE CAN NOW RUN AND REPRODUCE HER COMPUTATIONS

STEP 4



# Accessing an environment via a URL

- Entire information of the project is encoded in the URL
- Easily accessible, easily shareable

## Build and launch a repository

GitHub repository name or URL

GitHub ▾

Git branch, tag, or commit



Path to a notebook file (optional)

File ▾

launch

Copy the URL below and share your Binder with others:





# Kirstie Whitaker

“I like to use Binder when working with my students and collaborators because I can very easily check the analysis on my phone! While feeling fun, Binder also requires version control, the computational environment and a new build for each change.

Binder makes it much easier to share responsibility with busy PIs.”



<https://www.turing.ac.uk/people/researchers/kirstie-whitaker>

Build and launch a repository

GitHub repository name or URL

GitHub ▾

Git branch, tag, or commit

Path to a notebook file (optional)

File ▾

launch



Redirect User to  
mybinder.org/some\_url **6**

**1** Clone GitHub Repo



**2** Build image  
according to  
instructions  
contained within the  
repo



**5** Make image accessible at  
mybinder.org/some\_url



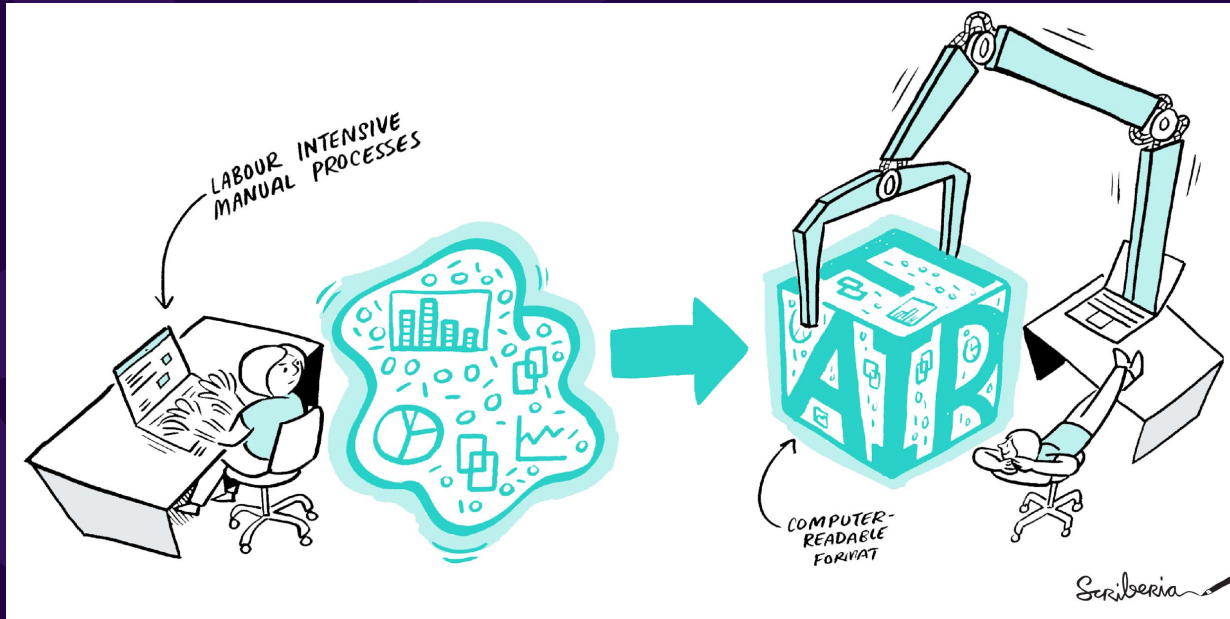
**4** Allocate  
computational  
resources

**3** Execute image



# Containerisation tends towards Portability

- Modern datasets are big! Need to take the software to the data
- Need to develop the analysis in the current environment
- Move software stacks between systems without costly setup time
- Reduce toil of one-off setups: events, trainings, workshops, etc



# What next?



# Call to Action

- Whether it's HPC centres or cloud datacentres - they're all just someone else's computers
- Need to port software stacks across systems
- How can concepts from Binder's API or `repo2<container>` be integrated into HPC operations to assist this?
  - Reducing toil for one-time setups
  - Facilitating ease-of-access to the development environment with all the expected tooling

# Thank You

Illustrations created by Scriberia  
with *The Turing Way* community.  
Used under a CC-BY 4.0 licence.  
DOI: [10.5281/zenodo.3332807](https://doi.org/10.5281/zenodo.3332807)