

**The  
Alan Turing  
Institute**

---

# **Reproducible Computational Environments with Binder**

**Shared notes:** [hackmd.io/@malvikasharan/BinderJuly2020](https://hackmd.io/@malvikasharan/BinderJuly2020)

**Malvika Sharan & Sarah Gibson**

**Pronouns:** she/her

[@TuringWay](#) [@malvikasharan](#) [@mybinderteam](#), [doi.org/10.5281/zenodo.3974919](https://doi.org/10.5281/zenodo.3974919)



# The Alan Turing Institute

---

## *The Turing Way:*

A guide to reproducible,  
collaborative, ethical and  
inclusive data science



“

*The Turing Way* is an **open source book project** that involves and supports a **diverse research community** in ensuring that reproducible and ethical **data science** is accessible and comprehensible **for everyone**.

”

# The Turing Way Team and Community



Kirstie Whitaker  
Turing Research Fellow

*The Turing Way* is a lightly opinionated guide to reproducible data science.

Our goal is to provide all the information that researchers need at the start of their projects to ensure that they are easy to reproduce at the end.

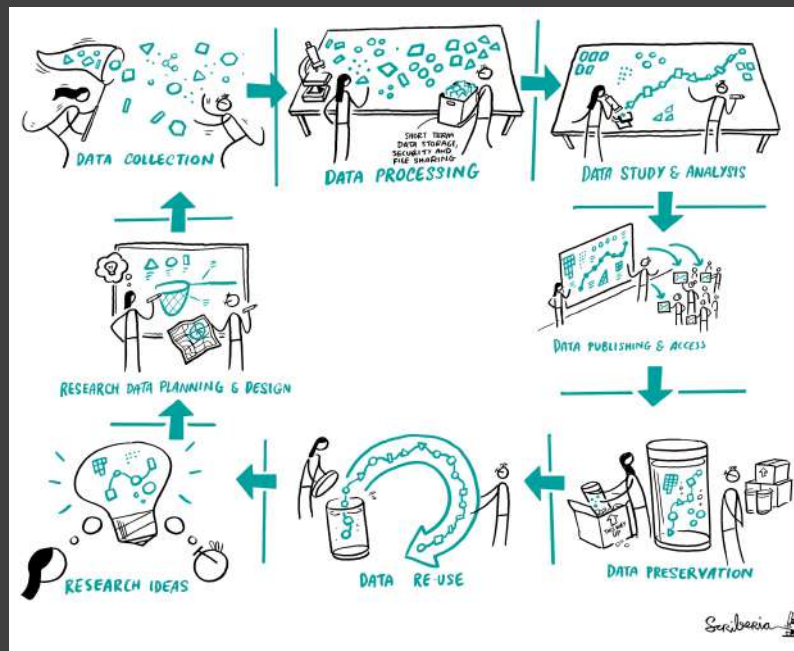
This also means making sure PhD students, postdocs, PIs, and funding teams know which parts of the "responsibility of reproducibility" they can affect, and what they should do to nudge data science to being more efficient, effective, and understandable.



# Started as a book on reproducibility

Same analysis steps on the same dataset produces same answer

This process is overwhelming.



# Started as a book on reproducibility

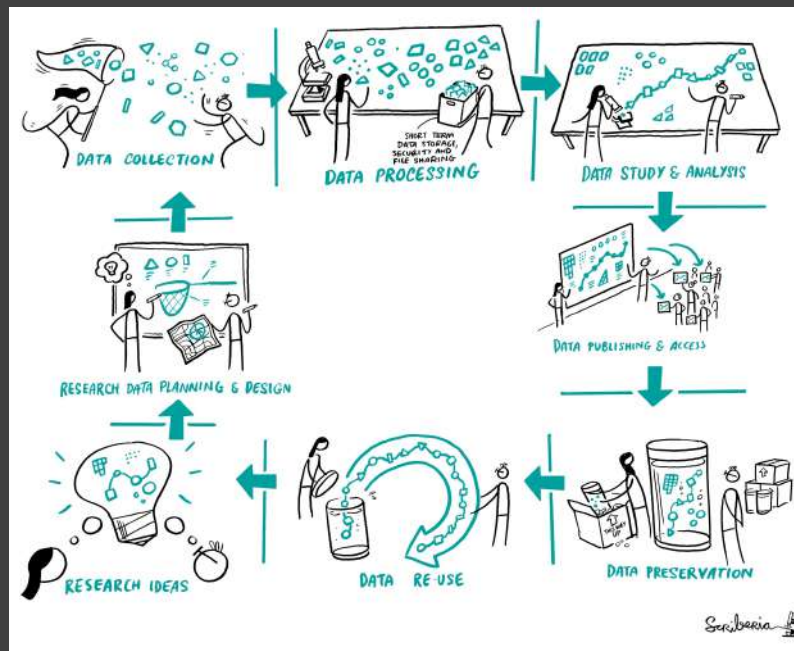
Same analysis steps on the same dataset produces same answer

This process is overwhelming.

1.5 years, >20 chapters,

>170 contributors,


a successful community



# Expansion of Scope

- Reproducibility
- Project Design
- Collaboration
- Communication
- Ethics

*It's about 'the journey'*



## The Turing Way

Welcome


- Guide for Reproducible Research
- Guide for Project Design
- Guide for Communication
- Guide for Collaboration
- Guide for Ethical Research
- Community Handbook

Powered by [Jupyter Book](#)

←

## Welcome

The Turing Way is an Open Access community-driven guide to reproducible, ethical, inclusive and collaborative data science.



Our goal is to provide all the information that researchers need at the start of their projects to ensure that they are easy to reproduce at the end.





## The Turing Way

Search this book...

Welcome

Guide for Reproducible Research

Overview

Open Research

Version Control

Licensing

Research Data Management

Reproducible Environments

BinderHub

Code quality

Code Testing

Code Reviewing Process

Continuous Integration

Reproducible Research with Make

Research Compendia

Credit for Reproducible Research

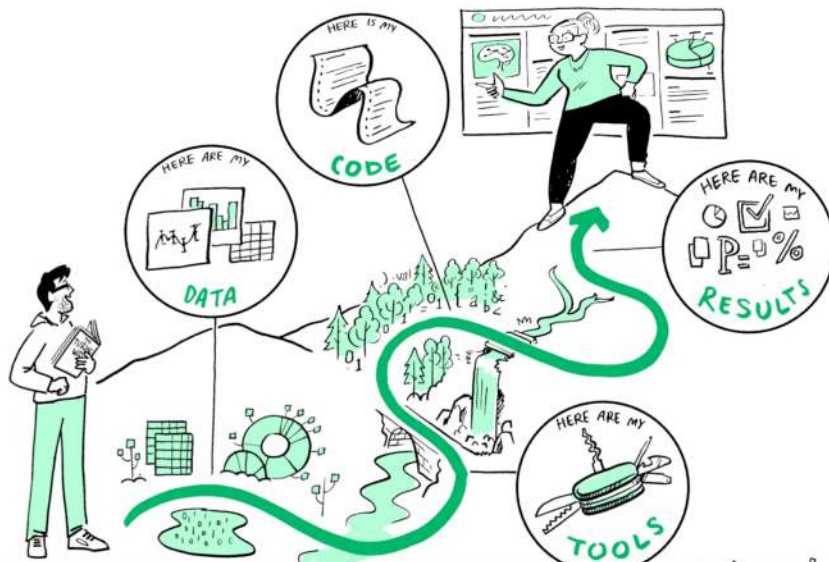
Risk Assessment

# Guide for Reproducible Research

*This guide covers topics related to skills, tools and best practices for research reproducibility.*

*The Turing Way defines reproducibility in data research as data and code being available to fully rerun the analysis.*

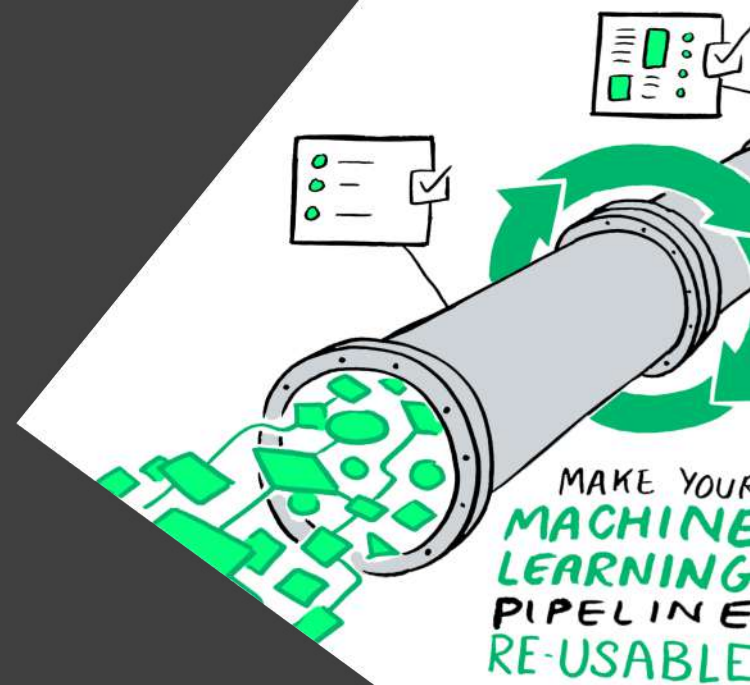
There are several definitions of reproducibility in use, and we discuss these in more detail in the [Definitions of Reproducibility](#) section of this chapter. While it is absolutely fine for us each to use different words, it will be useful for you to know how *The Turing Way* defines *reproducibility* to avoid misunderstandings when reading the rest of the handbook.





---

# What is Reproducibility?



		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

Kirstie Whitaker's talk at PyData LDN: <https://youtu.be/IG3PcZ6EhiU>

<https://the-turing-way.netlify.com/reproducibility/03/definitions.html>

@TuringWay, @drsarahlgibson, @mybinderteam, [doi.org/10.5281/zenodo.3974919](https://doi.org/10.5281/zenodo.3974919)

# The science is the code

*An article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures.*

Buckheit and Donoho

(paraphrasing John Claerbout)

WaveLab and Reproducible Research, 1995

@TuringWay, @drsarahlgibson,

Slide courtesy of Chris Holdraf and the Jupyter Team @mybinderteam, [doi.org/10.5281/zenodo.3974919](https://doi.org/10.5281/zenodo.3974919)

# The science is the code

*An article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures.*

Buckheit and Donoho

(paraphrasing John Claerbout)

WaveLab and Reproducible Research, 1995

@TuringWay, @drsarahlgibson,

Slide courtesy of Chris Holdraf and the Jupyter Team @mybinderteam, [doi.org/10.5281/zenodo.3974919](https://doi.org/10.5281/zenodo.3974919)

---

# Upsetting take home message

Sharing your code and  
data isn't enough



The computational environment includes:

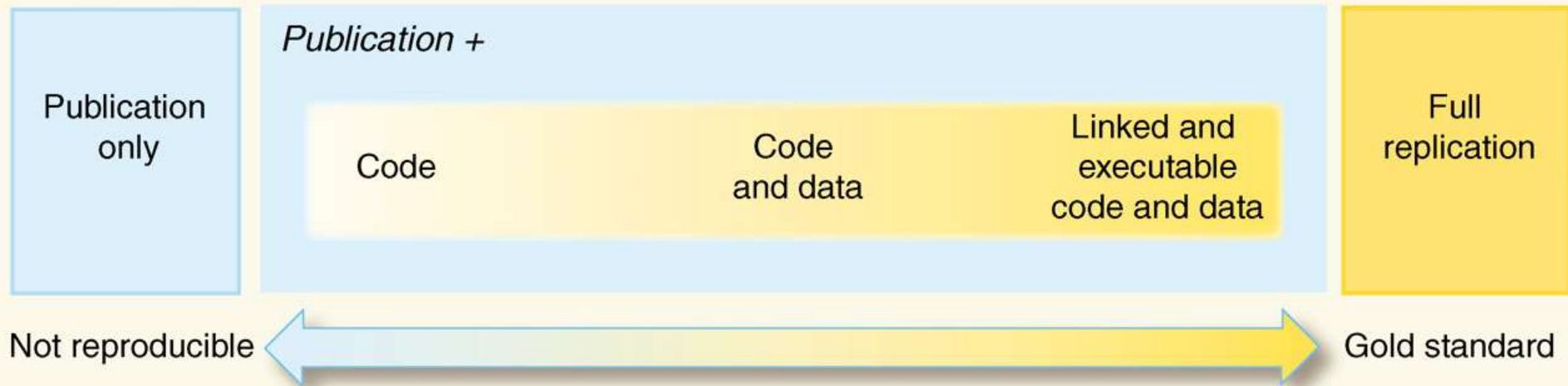
- Hardware (GPU, CPU)
- Operating system (mac, windows, linux)
- Software
  - Language version
  - Package version(s)

**And all the interactions between the layers**



# You need the computational environment too

## Reproducibility Spectrum





---

# What is Binder?

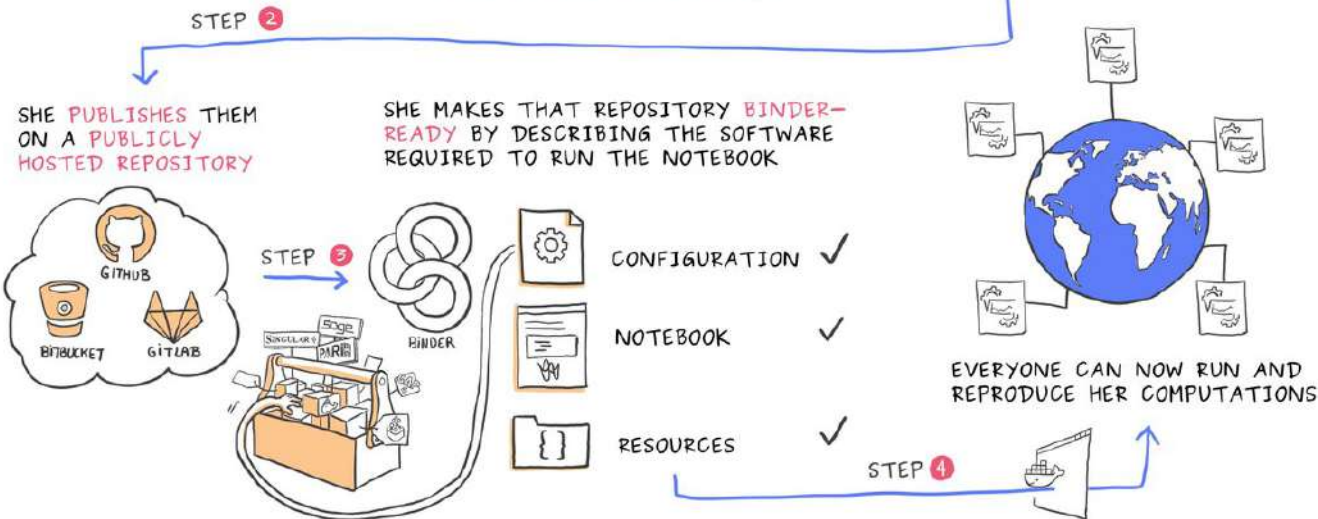
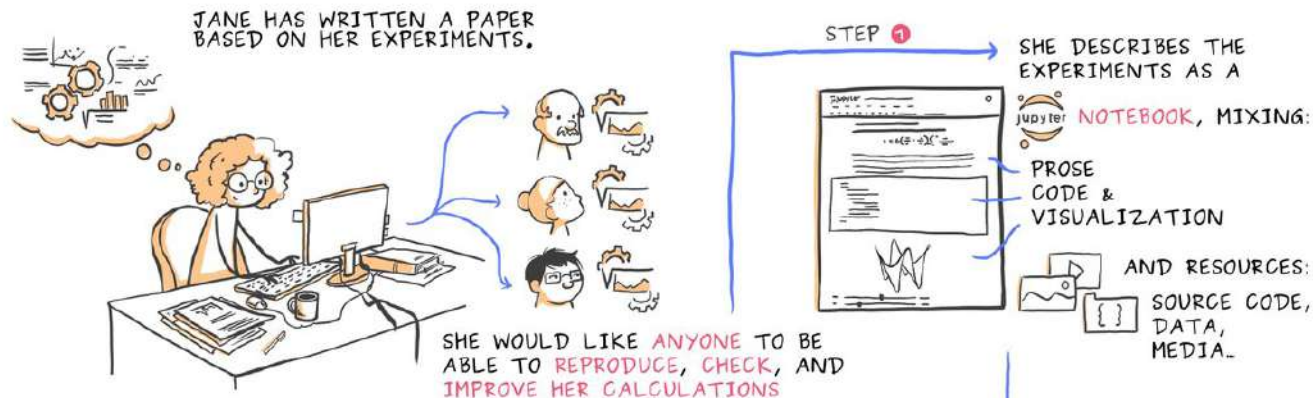


- Project Binder is a global community
- Binder can host interactive code in a browser



- Project Binder is a global community
- [mybinder.org](https://mybinder.org) allows launching interactive environments in a browser





mybinder.org

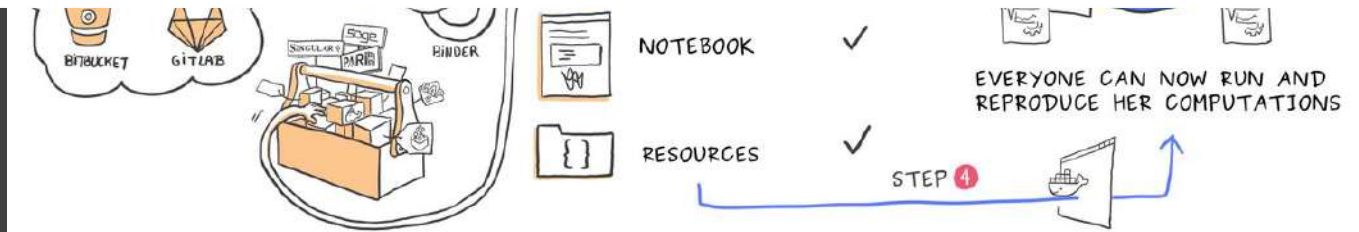
choldgraf Update requirements.txt 21a328d on 21 Jun

2 contributors

5 lines (3 sloc) 46 Bytes

Raw Blame History

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



		Data	
		Same	Different
Analysis	Same	<div>Repeatable</div> Reproducible	Replicable
	Different	Robust	Generalisable

Kirstie Whitaker's talk at PyData LDN: <https://youtu.be/IG3PcZ6EhiU>

<https://the-turing-way.netlify.com/reproducibility/03/definitions.html>

@TuringWay, @drsarahlgibson,  
@mybinderteam, [doi.org/10.5281/zenodo.3974919](https://doi.org/10.5281/zenodo.3974919)



Branch: master ▾

[the-turing-way](#) / [workshops](#) / [boost-research-reproducibility-binder](#) / [workshop-presentations](#) / [zero-to-binder-python.md](#)[Go to file](#)

...



sgibson91 Tidy up the Python Binder tutorial ✓

Latest commit 808b9db on 3 Jun [History](#)

1 contributor

294 lines (197 sloc) | 13.9 KB

[Edit on HackMD](#)[Raw](#)[Blame](#)

# From Zero to Binder in Python!

Sarah Gibson, *The Alan Turing Institute*

[The Turing Way](#) - making reproducible Data Science "too easy not to do"

Based on Tim Head's *Zero-to-Binder* workshops which can be found here: <http://bit.ly/zero-to-binder> and <http://bit.ly/zero-to-binder-rise>

To follow these instructions on your own machine, follow this link: <http://bit.ly/zero-to-binder-python>

Binder can take a long time to load, but this doesn't necessarily mean that you Binder will fail to launch. You can always refresh the window if you see the "... is taking longer to load, hang tight!" message.

See HackMD for the link: <https://hackmd.io/@malvikasharan/BinderJuly2020>



Thanks to [Google Cloud](#) and [OVH](#) for sponsoring our computers 🙌



Starting repository: `jupyterhub/binder-billing/master`

New to Binder? Check out the [Binder Documentation](#) for more information.

Build logs

show

Here's a non-interactive preview on [nbviewer](#) while we start a server for you. Your binder will open automatically when it is ready.



JUPYTER

FAQ



binder-billing

Thanks to [Google Cloud](#) and [OVH](#) for sponsoring our computers 🙌



Starting repository: `jupyterhub/binder-billing/master`

New to Binder? Check out the [Binder Documentation](#) for more information.

Build logs

show

Here's a non-interactive preview on [nbviewer](#) while we start a server for you. Your binder will open automatically when it is ready.



JUPYTER

FAQ



binder-billing

&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
```

&lt;&gt; 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

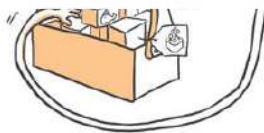
Raw

Blame

History



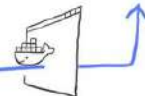
```
1 Package: binderdescription
2 Version: 0.1
3 Date: 2018-09-18
4 Title: Binder R DESCRIPTION support
5 Description: Test that automatically building R packages works
6 Author: Bastian Greshake Tzovaras <bgresshake@googlemail.com>
7 Maintainer: Bastian Greshake Tzovaras <bgresshake@googlemail.com>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



RESOURCES



STEP 4

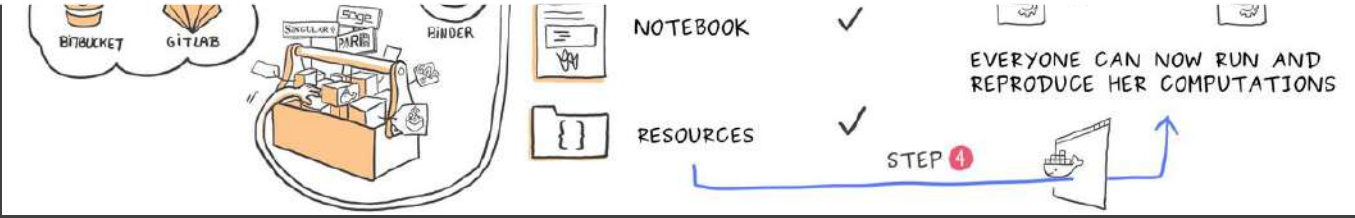


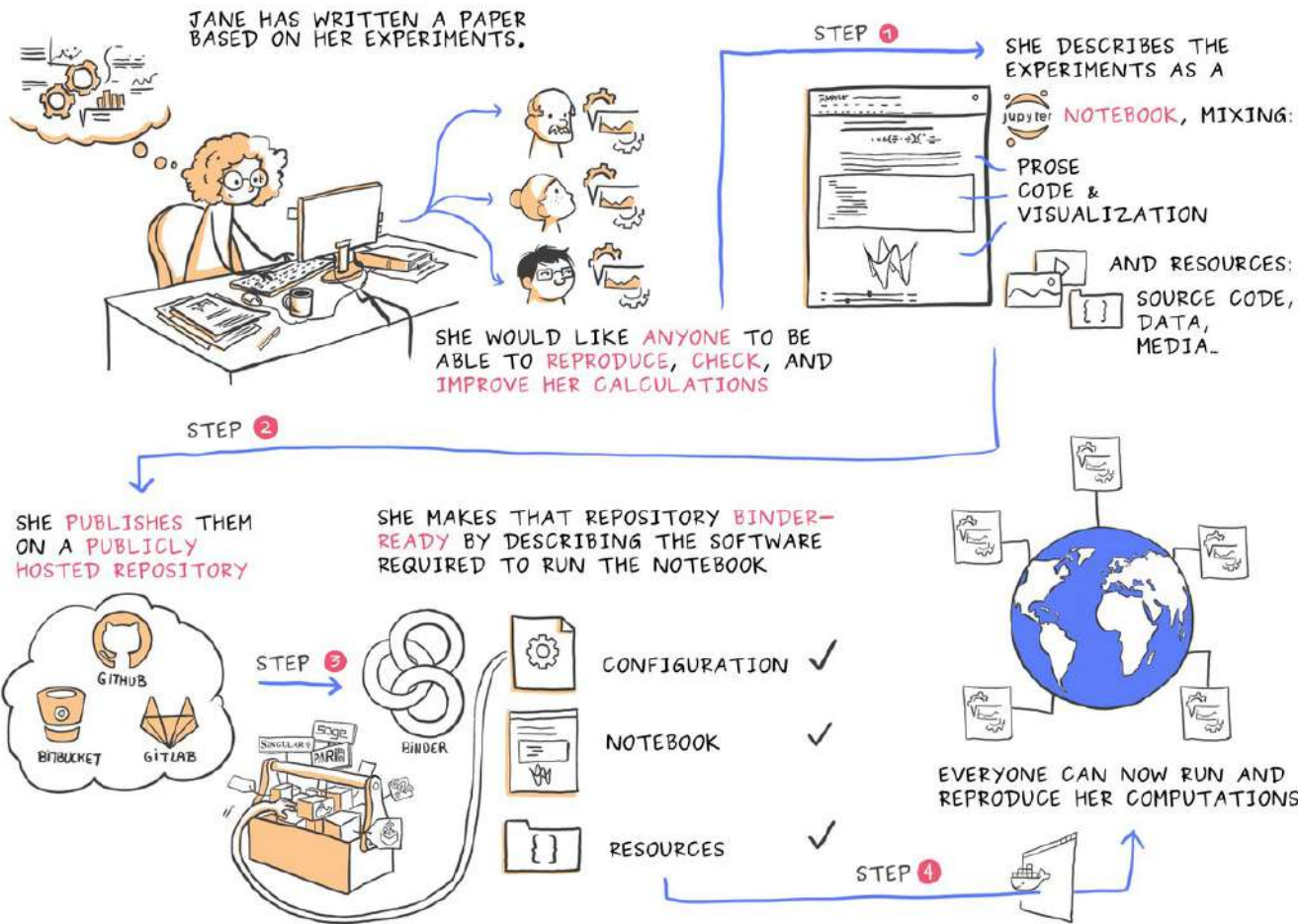
 **betatim** Add example Shiny app

8c01f0d on 31 May 2018

4 contributors 

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





mybinder.org

# Sarah Gibson

---

“It took me a while to feel like I knew enough to contribute to Binder. But the team are always so excited to have my input. Its really motivating to be part of such a welcoming community.”



<https://www.turing.ac.uk/people/researchers/sarah-gibson>

@TuringWay, @drsarahlgibson, @mybinderteam, [doi.org/10.5281/zenodo.3974919](https://doi.org/10.5281/zenodo.3974919)



- Check analysis on my phone
- Share the responsibility with busy PIs
- Requires version control, capturing environment and new build for each change





## Table of Contents

### Getting started with Binder

[Getting started with Binder](#)  
[Common usage patterns in Binder](#)

### How to...

[Choose languages for your environment](#)  
[Configure the user interface](#)  
[Generate custom launch badges for your Binder repository](#)  
[Track repository data on mybinder.org](#)  
[Share your launch](#)

## What is mybinder.org?

**mybinder.org** is a single deployment of a BinderHub instance, managed by the Binder community. It serves as both a public service and a demonstration of the BinderHub technology, though it is by no means the only BinderHub in existence. If you're interested in deploying your own BinderHub for your own uses, please see the [BinderHub documentation](#) and don't hesitate to reach out to the [Binder community](#).

For more information, check out [About mybinder.org](#).

## Is mybinder.org free to use?

Yes! Though note that it has relatively [limited computational resources](#).

## How much does running mybinder.org cost?

Great question! If you're interested in the technical costs of running **mybinder.org**, we publish a semi-up-to-date dataset of our costs at the [binder-data](#) repository. In addition, you can explore these costs with the binder link below!



## How can mybinder.org be free to use?

### On this page

What is a Binder?  
What is the Binder community?  
What is BinderHub?  
What is **mybinder.org**?  
Is **mybinder.org** free to use?  
How much does running **mybinder.org** cost?  
How can **mybinder.org** be free to use?  
How much memory am I given when using Binder?  
How long will my Binder session last?  
Can I use mybinder.org for a live demo or workshop?  
How does mybinder.org ensure user privacy?  
How secure is mybinder.org?  
Where can I report a security issue?  
Can I push data from my Binder session back to my repository?  
Can I put my configuration files outside the root of  **v: latest**   
What factors influence how long it takes a Binder session to start?  
Will repos with fewer notebooks launch faster? Should I split my

<https://mybinder.readthedocs.io/en/latest/faq.html#how-much-does-running-mybinder-org-cost>

@TuringWay, @drsarahlgibson, @mybinderteam, doi.org/10.5281/zenodo.3974919

---

# Thank you!

The  
Alan Turing  
Institute

Learn more about Binder:

- [mybinder.readthedocs.io](https://mybinder.readthedocs.io)
- [discourse.jupyter.org](https://discourse.jupyter.org)
- [bit.ly/zero-to-binder-tutorial](https://bit.ly/zero-to-binder-tutorial)



# The Alan Turing Institute

Learn more about reproducibility:

- Book: [the-turing-way.netlify.com](https://the-turing-way.netlify.com)
- Twitter: [twitter.com/turingway](https://twitter.com/turingway)
- Newsletter: [tinyletter.com/TuringWay](https://tinyletter.com/TuringWay)
- Chat: [gitter.im/alan-turing-institute/the-turing-way](https://gitter.im/alan-turing-institute/the-turing-way)
- GitHub: [github.com/alan-turing-institute/the-turing-way](https://github.com/alan-turing-institute/the-turing-way)
- Original artwork by Scriberia: <https://doi.org/10.5281/zenodo.3332808>



Special mentions to these communities:



THE  
CARPENTRIES

moz://a



Software  
Sustainability  
Institute



netherlands  
Science center