

Practical Steps Towards Reproducible Research

Heidi Seibold / @HeidiBaya
Freelance Open Science trainer, IGDORE

WORKSHOP

ENABLING REPRODUCIBILITY IN DATA SCIENCE

📅 Thursday, 09.06.2022 · 9.30 a.m. - 1 p.m. 📍 Online

Slides and drawings are licensed
under CC-BY 4.0



?

?

Please ask questions in the
HedgeDoc

?

?

?

•

Why are you here?

→ Breakout room (4 minutes)

**The
Economist**

OCTOBER 19TH-20TH 2013

economist.com

Washington's lawyer surplus

How to do a nuclear deal with Iran

Investment tips from Nobel economists

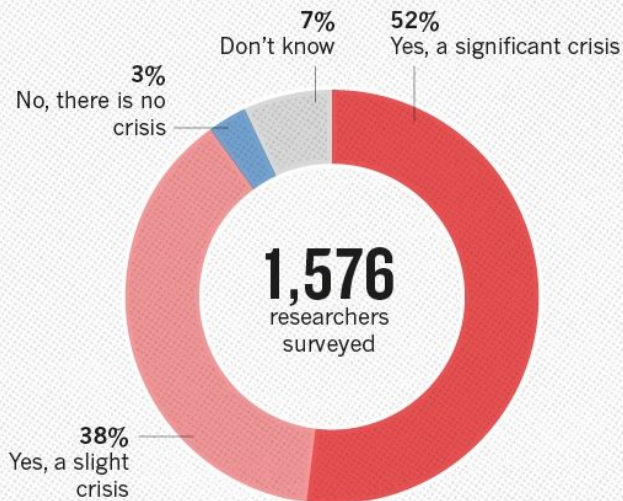
Junk bonds are back

The meaning of Sachin Tendulkar

**HOW
SCIENCE
GOES
WRONG.**

99
Einsteinium

IS THERE A REPRODUCIBILITY CRISIS?



Why Most Published Research Findings Are False

John P. A. Ioannidis

A computational reproducibility study
featuring longitudinal data analyses

Heidi Seibold^{*,1,2,3,4}, Severin Czerny¹, Siona Decke¹, R
Steffen Fohr¹, Nico Hahn¹, Rabea Hartmann¹, Christo
Dario Lepke¹, Verena Loidl¹, Maximilian Mandl¹, Sara
Alexander Piehler¹, Elio Rojas¹, Stefanie Schmid¹, Han
Lennart Schneider¹, Xiao-Yin To¹, Viet Tran¹, Antje V
Wagner¹, Maria Waize¹, Hannah Wecker¹, Rui Yang¹,

1 LMU Munich, Germany

2 University of Bielefeld, Germany

3 Helmholtz Zentrum München, Germany

4 LMU Open Science Center

[DOI: 10.1371/journal.pmed.0020124](https://doi.org/10.1371/journal.pmed.0020124)

Is this what we want?

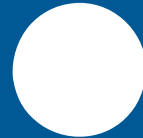
**How to
become a
science
champion?**





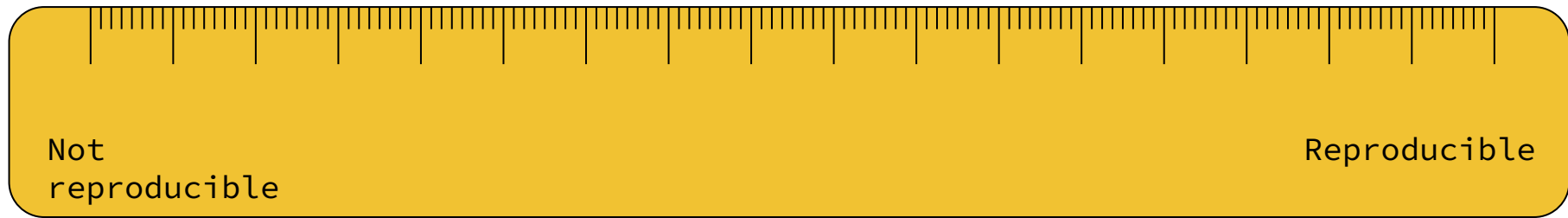
**What makes a science
champion?**

→ Answer in Hedge Doc



		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

From [The Turing Way](#)



Practical steps towards computational reproducibility



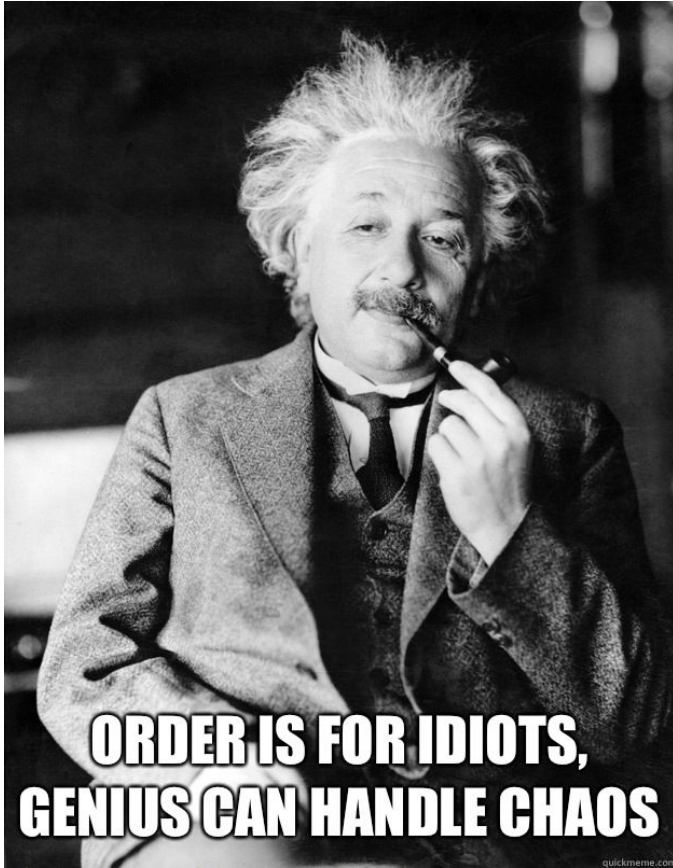
Practical steps towards reproducible research

- Get organized!
- Use Open Source Software
- Use Version Control
- Make your work available online

Get organized!



Good organisation



Let's not pretend: we're not geniuses ;P

Good organisation ... starts simple

- No sending emails with files
- Nice file organisation
- Good naming

Dear colleagues,

attached you find the first public version of the [REDACTED] protocol. Please have a look and do comment. We can also meet to aggregate our reviews.

► 1 attachment: StudyProposal [REDACTED]_Validation_V1_250918docx.docx

```
— Makefile
— references_topic.bib
— manuscript_topic.Rnw
— abstract.tex
— introduction.tex
— methods.tex
— application.Rnw
— simulation.Rnw
— discussion.tex
— journal_correspondence
  — journal2
    — response1.txt
    — p2p_1.tex
  — journal1
    — response.txt
— analysis
  — Makefile
  — try_stuff.Rmd
  — simulation.Rnw
  — simulation.R
  — analysis_study1.R
  — analysis_study2.R
  — data
    — clean_data.R
    — data_study2.csv
    — data_study1.csv
    — simulation_results.RData
  — data_original
    — data_study1_original.csv
    — data_study2_original.csv
  — basis
    — functions_simulation.R
    — functions_analyses.R
```

Naming

NO

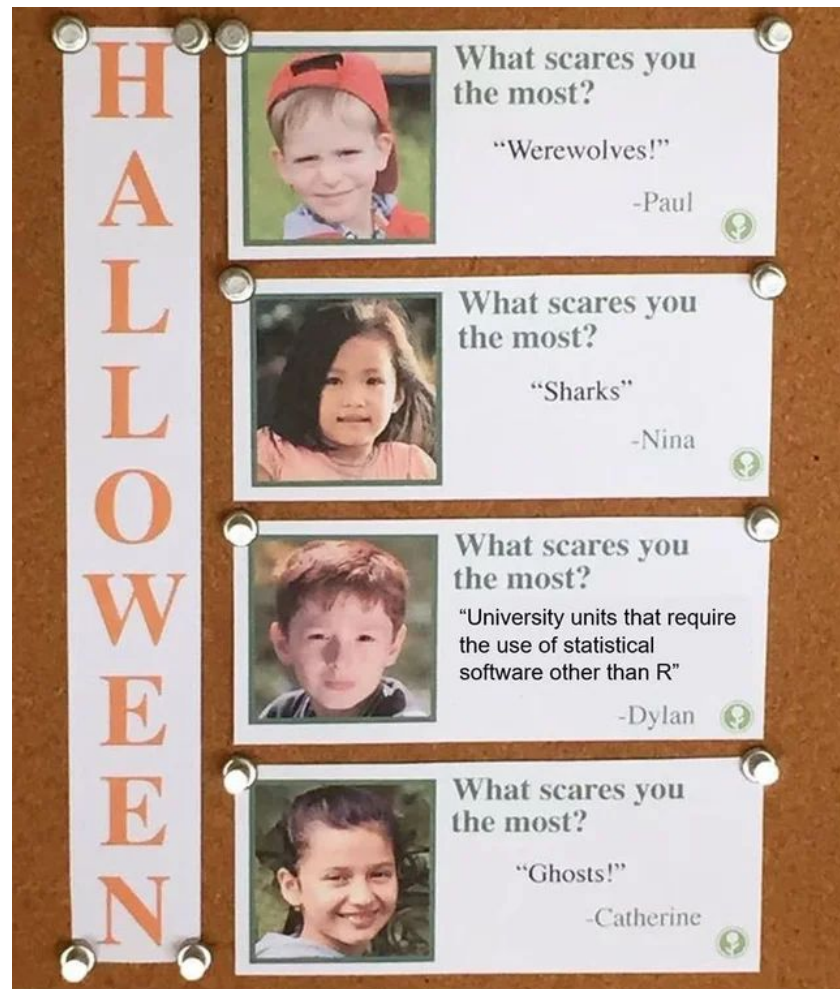
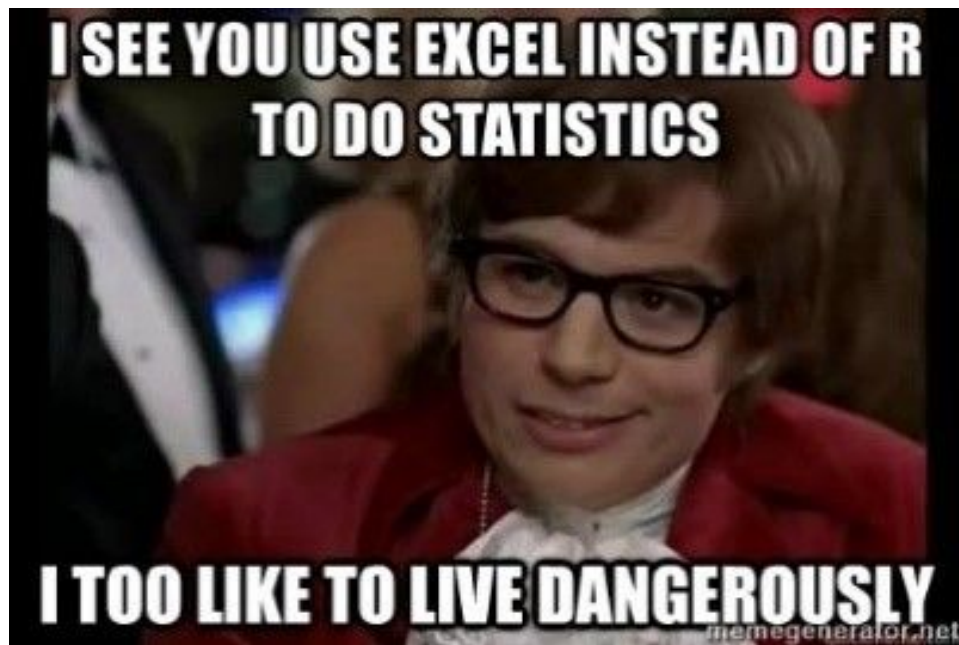
- Myabstract.docx
- Joe's Filenames Use Spaces and Punctuation.xlsx
- figure 1.png
- fig 2.png
- JW7d^(2sl@deletethisandyourcareerisoverWx2*.txt

YES

- 2014-06-08_abstract-for-sla.docx
- Joes-filenames-are-getting-better.xlsx
- Fig01_scatterplot-talk-length-vs-interest.png
- Fig02_histogram-talk-attendance.png
- 1986-01-28_raw-data-from-challenger-o-rings.txt

Use Open Source Software

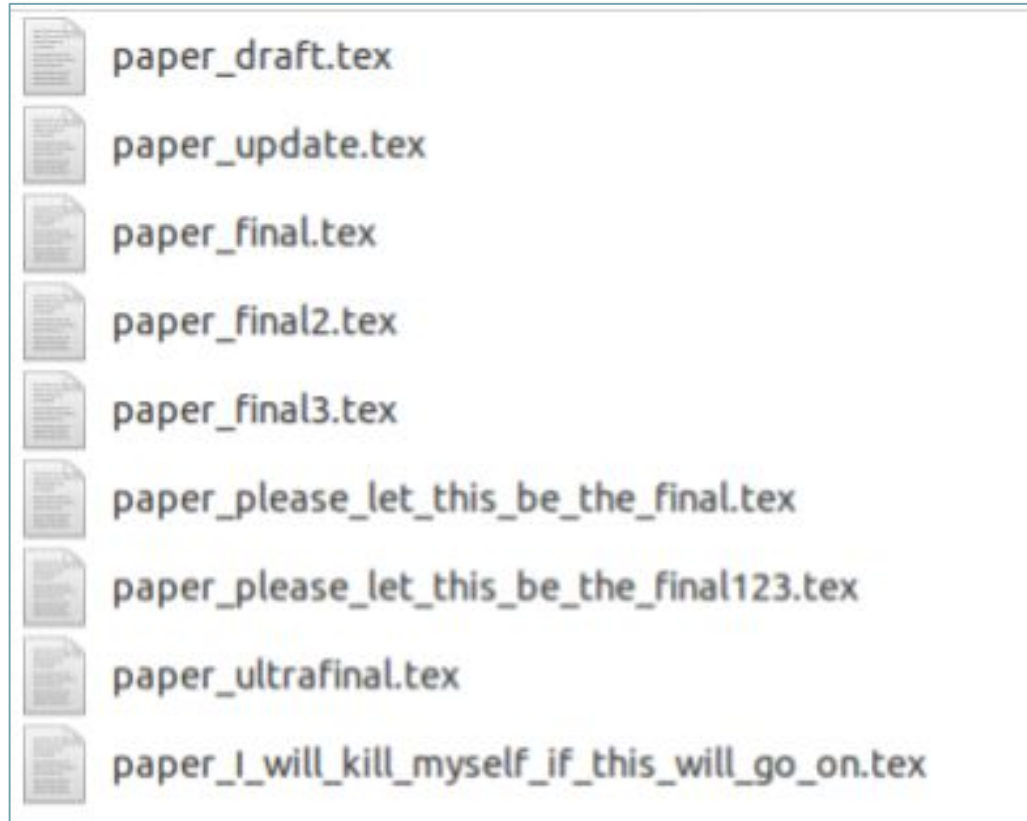




Use Version Control



Version Control in the olden days



Real Version Control

(Including Backup)


Commit **fe6b5538**  authored 10 months ago by  **HeidiSeibold**

[Browse files](#)

[Options](#) ▾

fix typos

 parent [e8a0da8a](#)  master

 No related merge requests found

Showing **1 changed file** ▾ with **3 additions** and **3 deletions**

[Hide whitespace changes](#)

[Inline](#)

[Side-by-side](#)

▼  **2018_swisscore.tex** 



[View file @ fe6b5538](#)

...	...	@@ -56,8 +56,8 @@ In my research I develop algorithms that detect which patient characteristics
56	56	lead to a positive or negative reaction to a therapy. The algorithms can also
57	57	detect and predict which patients are likely to have side effects from a
58	58	therapy. It is important to acknowledge, that \textbf{personalised medicine does not
59		- necessarily mean, that each patient receives a personalised treatment, but it
60		- can also mean that we find that treatments work similarly across patients and
	59	+ necessarily mean, that each patient receives a personalised treatment, but it
	60	+ can also mean that we find that treatments work similarly across patients and

Prerequisites for Version Control

- **Text**

- Documents/Papers: LaTeX, Markdown
- Analyses: R, Python

- **Willingness to learn something new: Git**

**Make your work available
online**



FAIR Data



Data and supplementary materials have sufficiently rich metadata and a unique and persistent identifier.

FINDABLE



Metadata and data are understandable to humans and machines. Data is deposited in a trusted repository.

ACCESSIBLE



Metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation.

INTEROPERABLE



Data and collections have a clear usage licenses and provide accurate information on provenance.

REUSABLE

Open Code

Steps

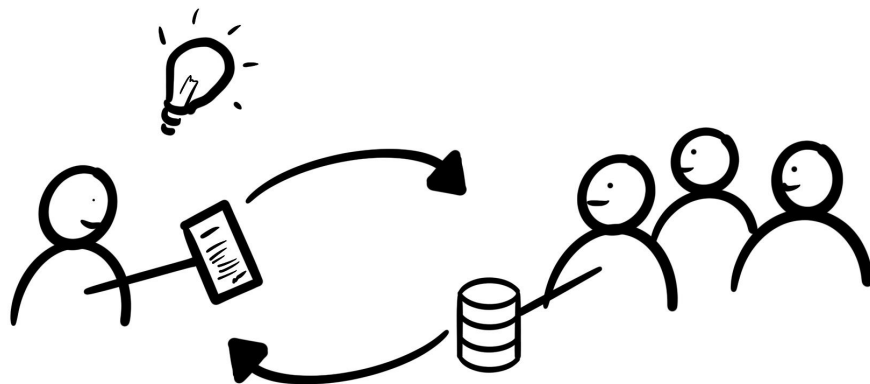
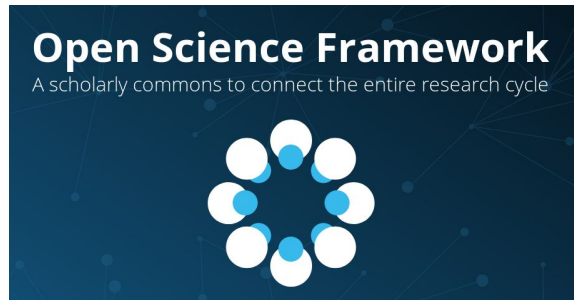
s1 <- "Use scripts (no clicking!)"

s2 <- "Publish code online"

Pro tip

s3 <- "Use version control"

Open Materials



GitLab

Break time



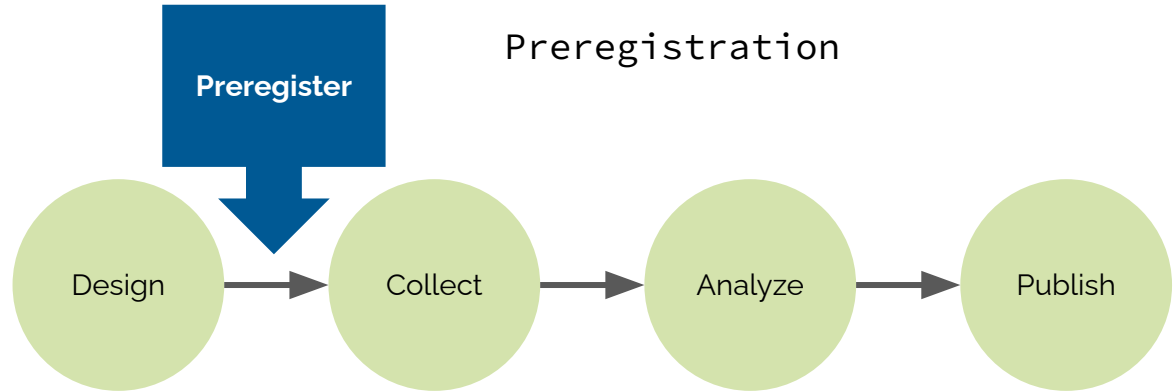
Practical steps towards reproducible research

- Get organized!
- Use Open Source Software
- Use Version Control
- Make your work available online

Other things to consider



Science communication



cos.io/prereg

How hard can it be?

1. Draw an animal on a piece of paper
2. Write instructions on how to draw what you drew (raise your Zoom hand when done)
3. Go to breakout room → draw according to the instructions of the other person
4. Check your result: did you succeed?
5. If you want: upload the original and the reproduced drawing to the HedgeDoc
6. Come back to the main room

How hard can it be?

- What can we learn from that exercise?
- Is it easier or more difficult in your work?

Be a science champion + be selfish



Avoid disaster



Write papers more easily



Convince reviewers



Facilitate continuity of work

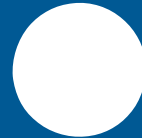


Build your reputation

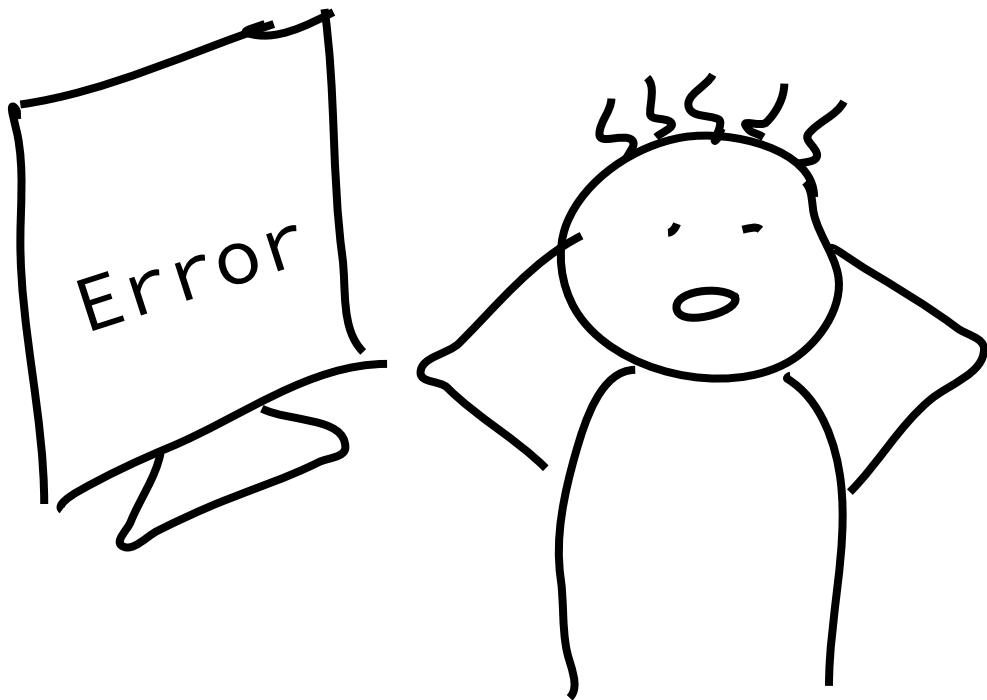


What is stopping you / others?

→ Answer in HedgeDoc

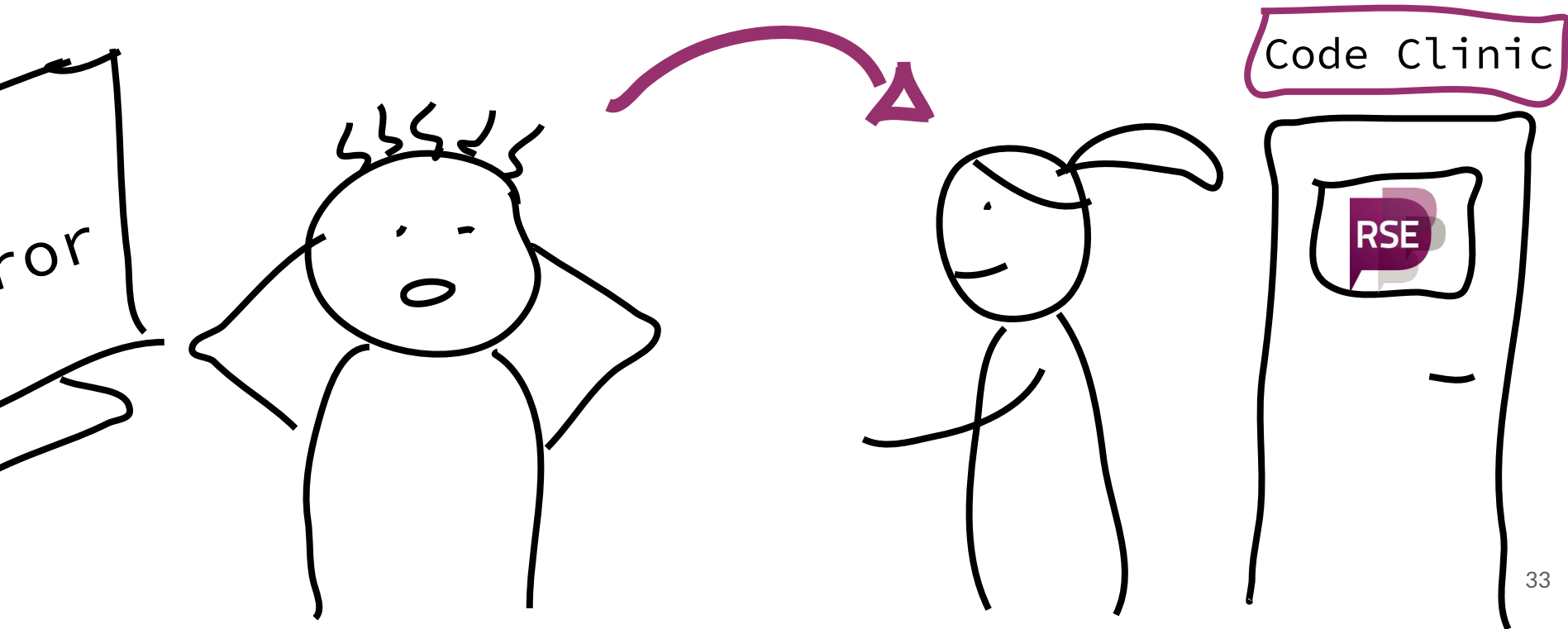


Overwhelmed?



Reproducible
Research
requires
Software Skills

Research Software Engineers can help



Help us help you



**BETTER
SOFTWARE
BETTER
RESEARCH**

Further reading



Ten simple rules for implementing open and reproducible research practices after attending a training course

the-turing-way.netlify.app

<https://anchor.fm/opensciencestories>

Open Science Stories

Let's discuss!

@HeidiBaya