

# Open Science, Reproducible Research, and R

## An Overview

Ernest Guevarra

2024-01-22

**Go to the people. Live with them. Learn from them. Love them.**

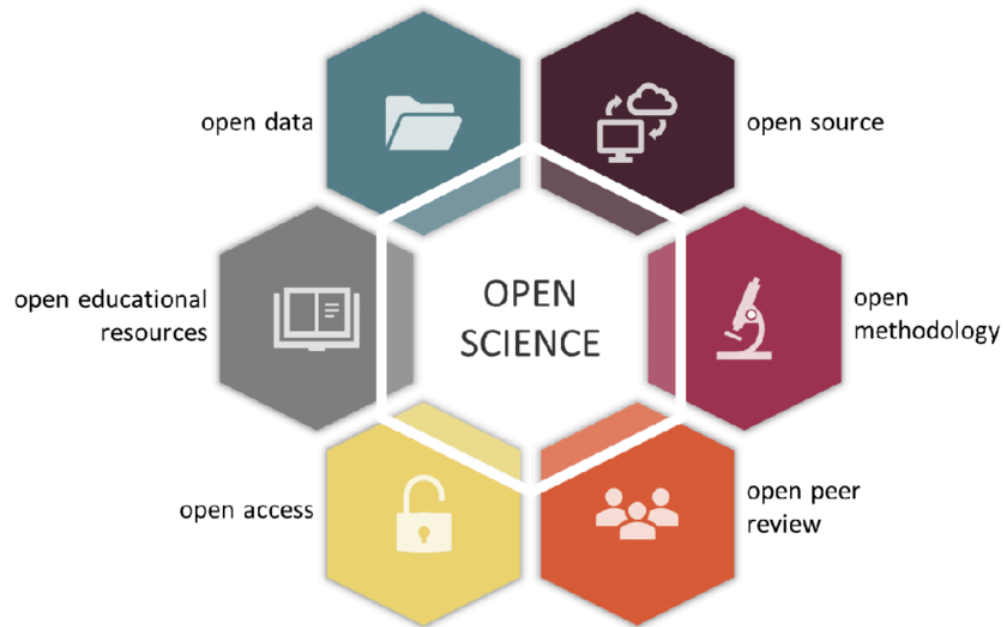
**Start with what they know. Build with what they have.**

**But with the best leaders, when the work is done, the task accomplished, the people will say "We have done it ourselves".**

# Outline

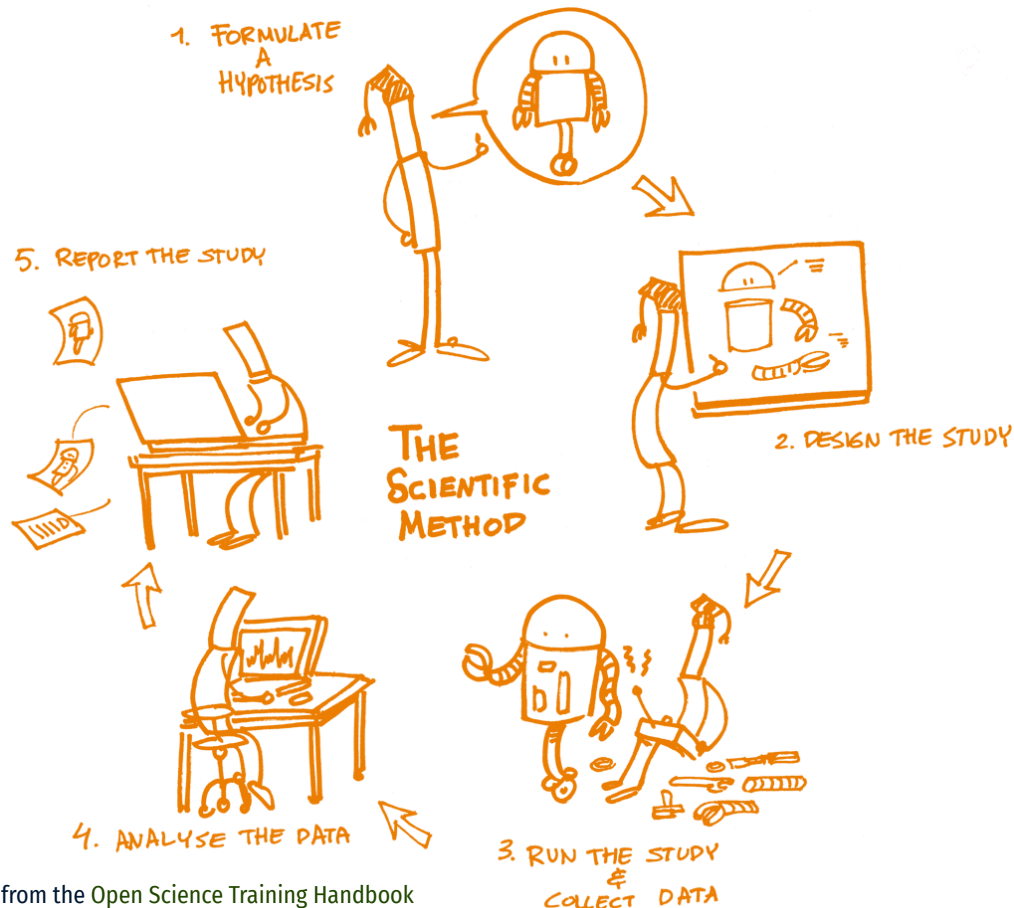
1. What is Open Science?
2. What is Reproducible Research?
3. Why Reproducible Research?
4. How to do Reproducible Research?
5. Why R?

# What is Open Science?



- movement to make scientific research and dissemination accessible to anyone
- aims for greater transparency in research and removes barriers for sharing outputs, resources, methods or tools at any stage of the research process

# What is Reproducible Research?

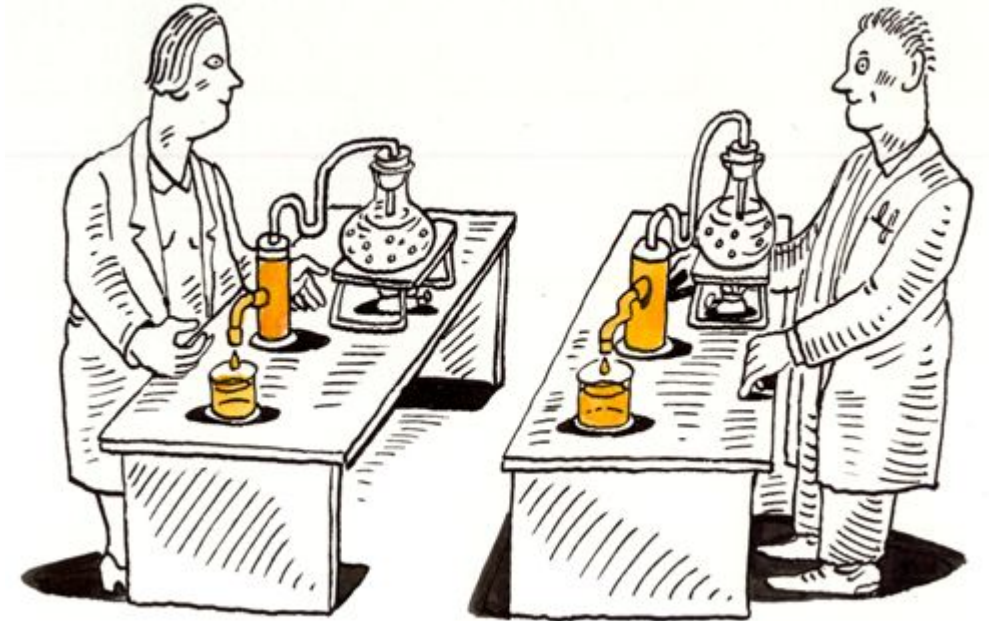


from the Open Science Training Handbook

- The concept of reproducibility is directly linked to the scientific method
  1. Formulating a hypothesis
  2. Designing the study
  3. Running the study and collecting the data
  4. Analyzing the data
  5. Reporting the study
- Each of these steps should be clearly reported by providing clear and open documentation, and thus making the study transparent and reproducible.

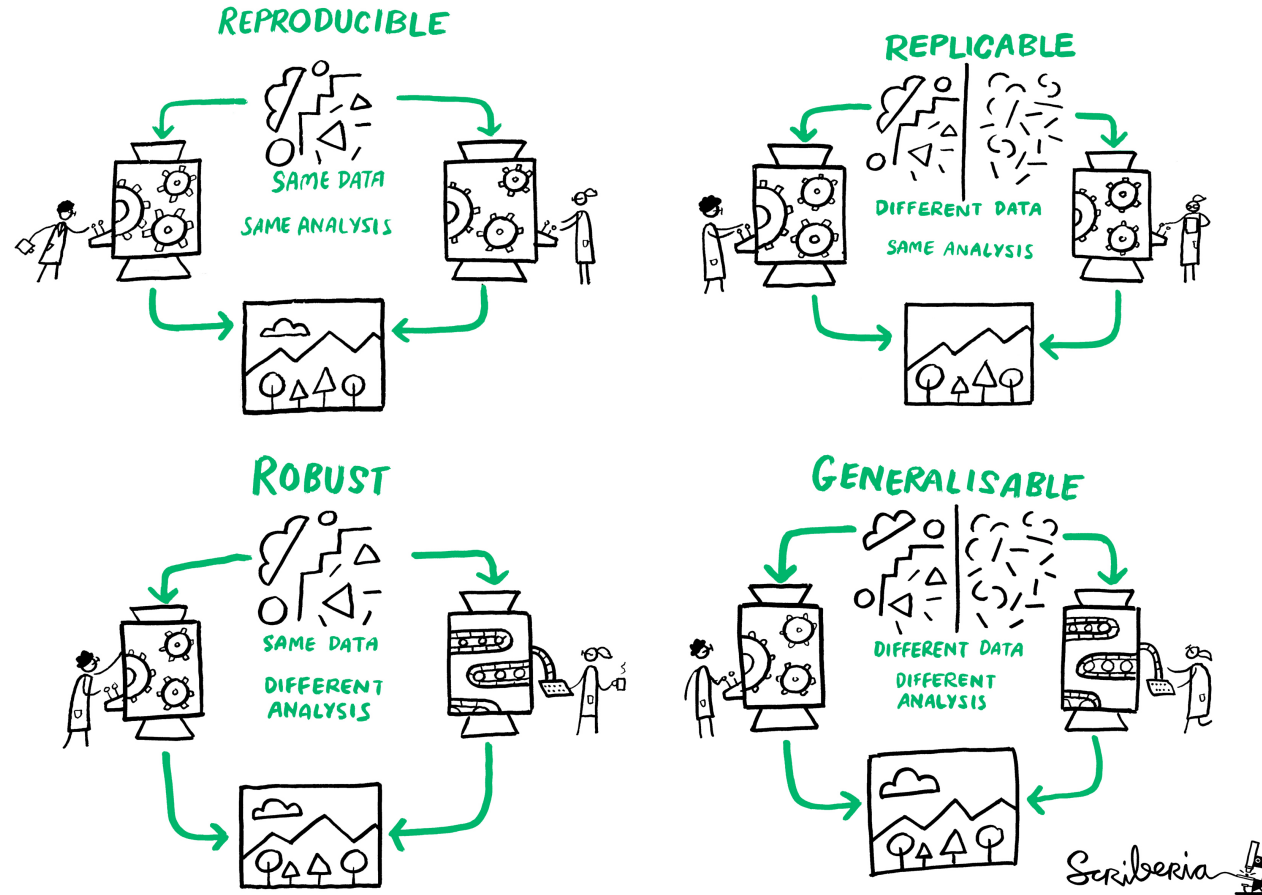
# What is Reproducible Research?

- Research papers with accompanying software tools that allow the reader to directly reproduce the results and employ the methods that are presented in the research paper (Gentleman and Lang, 2004)
- Research data and code are made available so that others are able to reach the same results as are claimed in scientific outputs (Open Science Training Handbook)
- The standard of reproducibility calls for the data and the computer code used to analyze the data be made available to others (Peng, 2012)



from <https://www.displayr.com/what-is-reproducible-research>

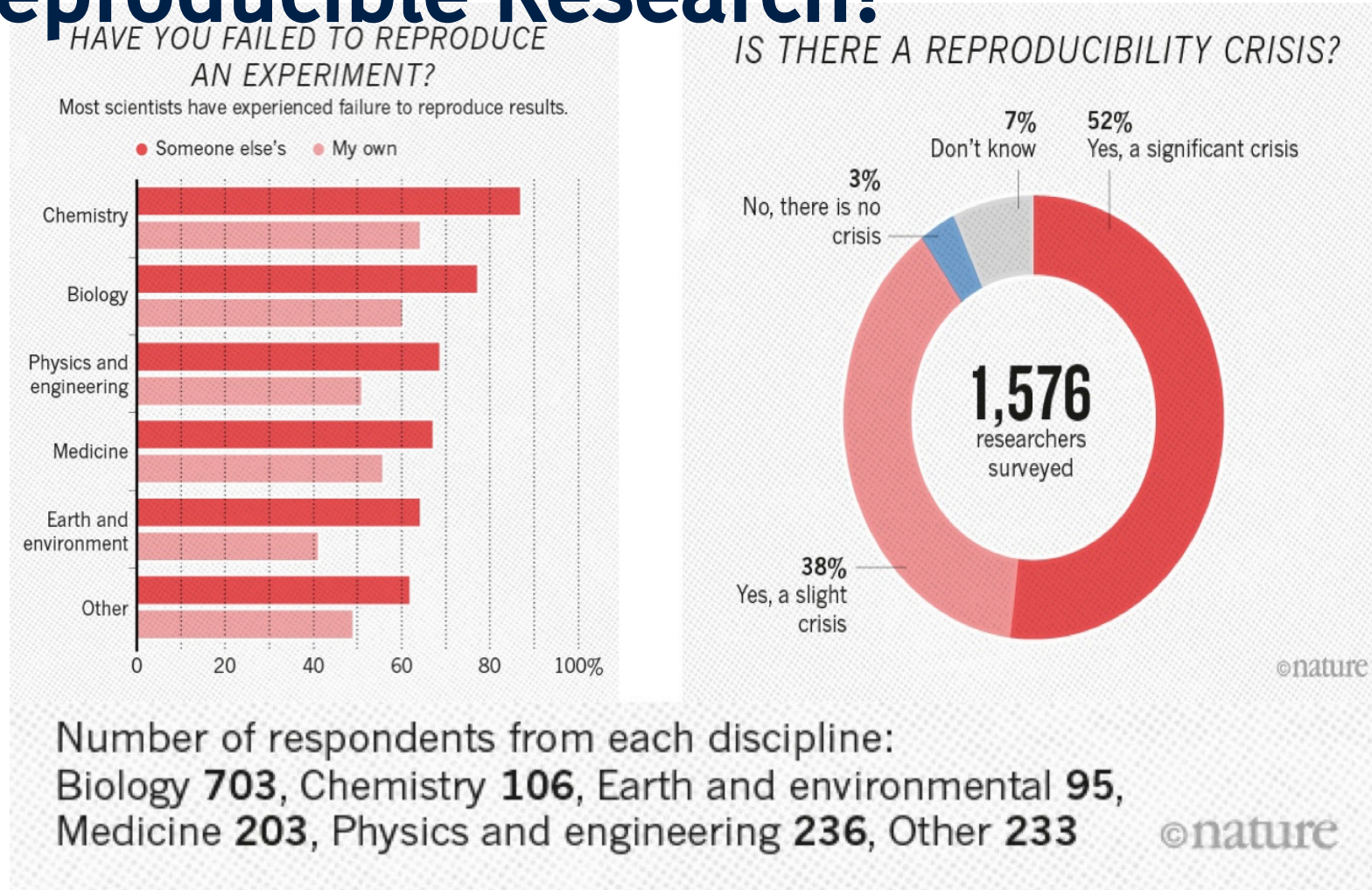
# Differentiating Reproducible Research



This image was created by Scriberia for The Turing Way community and is used under a CC-BY licence.



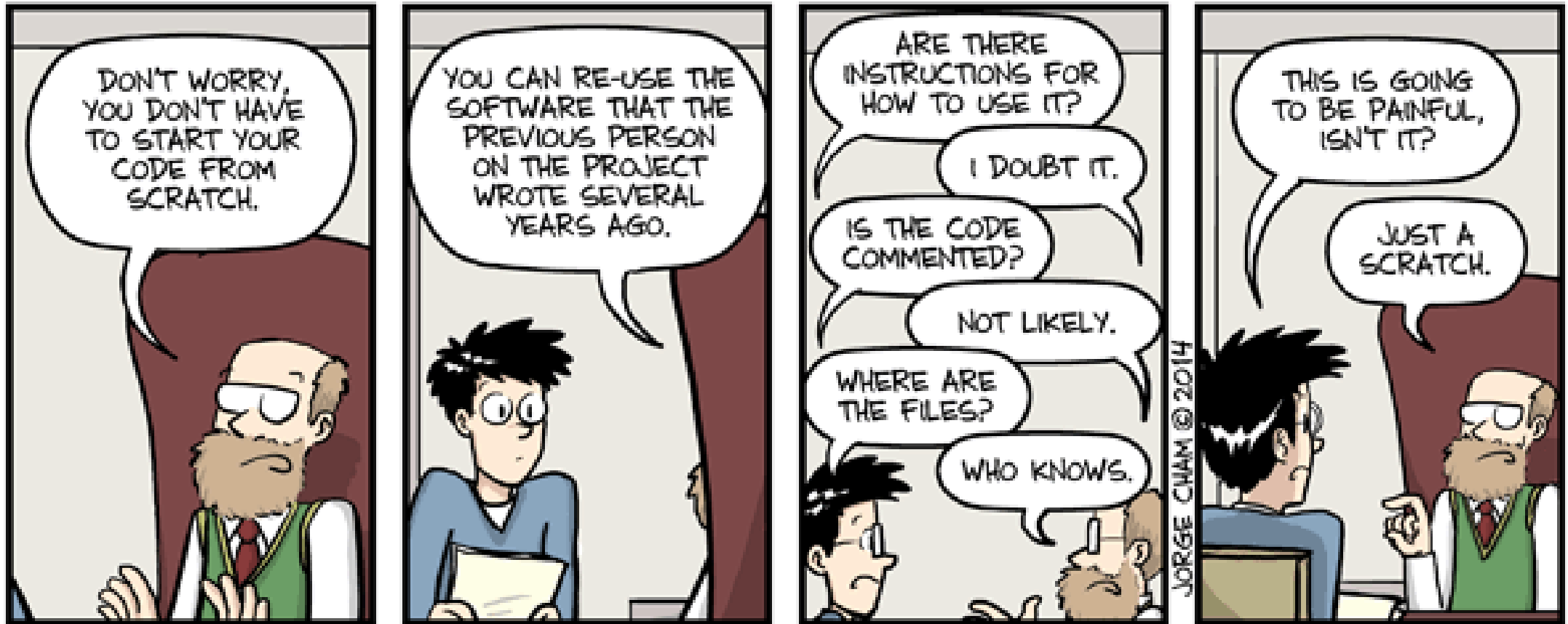
# Why Reproducible Research?



Baker, M. 1,500 scientists lift the lid on reproducibility. Nature 533, 452–454 (2016). <https://doi.org/10.1038/533452a>



# Factors in irreproducible research



[WWW.PHDCOMICS.COM](http://WWW.PHDCOMICS.COM)

# Factors in irreproducible research

- Not enough documentation on how experiment is conducted and data is generated
- Data used to generate original results unavailable
- Software used to generate original results unavailable
- Difficult to recreate software environment (libraries, versions) used to generate original results
- Difficult to rerun the computational steps

# How to do Reproducible Research?

## The reproducibility spectrum

### Data Replication & Reproducibility

PERSPECTIVE

### Reproducible Research in Computational Science

Roger D. Peng

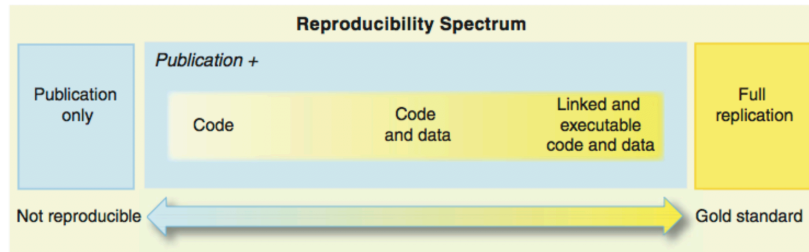


Fig. 1. The spectrum of reproducibility.

## Steps in reproducible research

- Record the project's provenance
- Data and metadata curation
- Establish a testing/analysis workflow
- Test, document, and publish your code
- Share

# Why R for Reproducible Research?

- freely available
- huge user and developer community
- has a robust set of user- and community-developed packages that support reproducible research



Questions?

**Go to the people. Live with them. Learn from them. Love them.**

**Start with what they know. Build with what they have.**

**But with the best leaders, when the work is done, the task accomplished, the people will say "We have done it ourselves".**

**Lao Tzu**

# Thank you!

Slides can be viewed at <https://oxford-ihtm.io/open-reproducible-science/session6.html>

PDF version of slides can be downloaded at <https://oxford-ihtm.io/open-reproducible-science/pdf/session6-open-reproducible-science.pdf>

R scripts for slides available [here](#)