# Managing your R code towards reproducibility

DZG Grad Meeting in Evolutionary Biology,
Bielefeld 2022

David García-Callejas

# Summary

- Why this workshop?
- General remarks on reproducibility
- What will you learn today?

# why this workshop?

*“Every analysis you do on a dataset will have to be redone 10-15 times before publication. Plan accordingly.”*

Trevor Branch

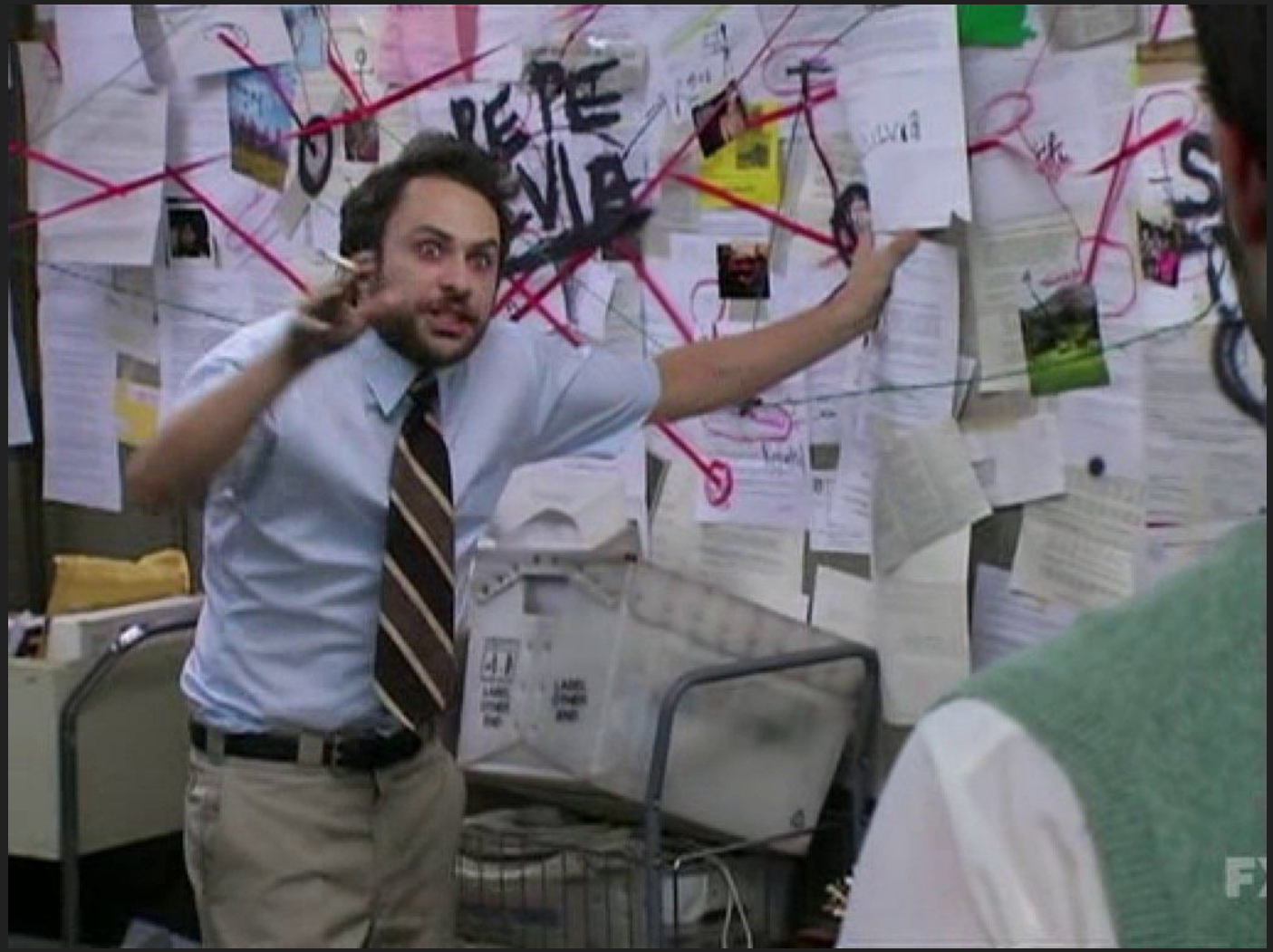
FX

**Table 1 | A manifesto for reproducible science.**

| Theme | Proposal | Examples of initiatives/potential solutions (extent of current adoption) | Stakeholder(s) |
|---|---|---|---|
| Methods | Protecting against cognitive biases | All of the initiatives listed below (* to ****)<br>Blinding (**) | J, F |
| | Improving methodological training | Rigorous training in statistics and research methods for future researchers (*)<br>Rigorous continuing education in statistics and methods for researchers (*) | I, F |
| | Independent methodological support | Involvement of methodologists in research (**)<br>Independent oversight (*) | F |
| | Collaboration and team science | Multi-site studies/distributed data collection (*)<br>Team-science consortia (*) | I, F |
| Reporting and dissemination | Promoting study pre-registration | Registered Reports (*)<br>Open Science Framework (*) | J, F |
| | Improving the quality of reporting | Use of reporting checklists (**)<br>Protocol checklists (*) | J |
| | Protecting against conflicts of interest | Disclosure of conflicts of interest (***)<br>Exclusion/containment of financial and non-financial conflicts of interest (*) | J |
| Reproducibility | Encouraging transparency and open science | Open data, materials, software and so on (* to **)<br>Pre-registration (**** for clinical trials, * for other studies) | J, F, R |
| Evaluation | Diversifying peer review | Preprints (* in biomedical/behavioural sciences, **** in physical sciences)<br>Pre- and post-publication peer review, for example, Publons, PubMed Commons (*) | J |
| Incentives | Rewarding open and reproducible practices | Badges (*)<br>Registered Reports (*)<br>Transparency and Openness Promotion guidelines (*)<br>Funding replication studies (*)<br>Open science practices in hiring and promotion (*) | J, I, F |

Estimated extent of current adoption: *, <5%; **, 5–30%; ***, 30–60%; ****, >60%. Abbreviations for key stakeholders: J, journals/publishers; F, funders; I, institutions; R, regulators.

Munafò et al. (2017)
*Nature Human Behaviour*

**Table 1 | A manifesto for reproducible science.**

| Theme | Proposal | Examples of initiatives/potential solutions (extent of current adoption) | Stakeholder(s) |
|---|---|---|---|
| Methods | Protecting against cognitive biases | All of the initiatives listed below (* to ****)<br>Blinding (**) | J, F |
| | Improving methodological training | Rigorous training in statistics and research methods for future researchers (*)<br>Rigorous continuing education in statistics and methods for researchers (*) | I, F |
| | Independent methodological support | Involvement of methodologists in research (**)<br>Independent oversight (*) | F |
| | Collaboration and team science | Multi-site studies/distributed data collection (*)<br>Team-science consortia (*) | I, F |
| Reporting and dissemination | Promoting study pre-registration | Registered Reports (*)<br>Open Science Framework (*) | J, F |
| | Improving the quality of reporting | Use of reporting checklists (**)<br>Protocol checklists (*) | J |
| | Protecting against conflicts of interest | Disclosure of conflicts of interest (***)<br>Exclusion/containment of financial and non-financial conflicts of interest (*) | J |
| Reproducibility | Encouraging transparency and open science | Open data, materials, software and so on (* to **)<br>Pre-registration (**** for clinical trials, * for other studies) | J, F, R |
| Evaluation | Diversifying peer review | Preprints (* in biomedical/behavioural sciences, **** in physical sciences)<br>Pre- and post-publication peer review, for example, Publons, PubMed Commons (*) | J |
| Incentives | Rewarding open and reproducible practices | Badges (*)<br>Registered Reports (*)<br>Transparency and Openness Promotion guidelines (*)<br>Funding replication studies (*)<br>Open science practices in hiring and promotion (*) | J, I, F |

Estimated extent of current adoption: *, <5%; **, 5–30%; ***, 30–60%; ****, >60%. Abbreviations for key stakeholders: J, journals/publishers; F, funders; I, institutions; R, regulators.

Munafò et al. (2017)
*Nature Human Behaviour*

# Reproducibility

*“A scientific article is advertising, not scholarship. The actual scholarship is the full software environment, code and data, that produced the result.”*

Claerbout and Karrenbach (1992)

# Reproducibility

- When is a study reproducible?

# Reproducibility

- When is a study reproducible?
- Reproducibility != Repeatability

# Reproducibility

- When is a study reproducible?
- Reproducibility != Repeatability
- Reproducibility is a gradient

# Reproducibility

Benefits of a reproducible workflow

- Automatic re-generation of results/repetitive tasks
- Easy to correct or update results
- Simplifies collaboration
- Publishing code helps spotting mistakes
- Publishing code helps the review process
- Reproducibility increases the impact of the study
- Saves time and effort in future projects

# Reproducibility

Criteria for a reproducible workflow

- Raw data is available
- Raw data has been validated
- Raw data is properly documented
- Raw data is stored in open formats
- Raw data is openly accessible in an online repository
- All data management and analysis is performed through computer code
- Computer code is properly documented
- Computer code generates the final tables and figures
- Data and code use version control systems
- All files are contained in a parent folder
- Parent folder organized in subfolders
- Raw data always separated from derived results
- There exists a README file describing objectives and organization of the study
- It is possible to install the necessary software in different machines
- The final manuscript, data, and computer code are publicly accessible, and their license is specified

# what will we learn today?

- up to 10:30: how to organize your R code
- 11:00 - 12:30: setting up your Github account and sharing Rstudio projects. Bonus: generating Rmarkdown files combining text, code, and figures.