

Shifting Science and Research Culture Towards Openness and Reproducibility Through Education and Training


Luke W. Johnston¹, Helene Baek Juel ², Bettina Lengger ³, Daniel R. Witte ^{1,4}, Hannah Chatwin ⁵, Malene Revsbech Christiansen ², Anders Aasted Isaksen ⁴

1. *Steno Diabetes Center Aarhus*; 2. *Novo Nordisk Foundation Center for Basic Metabolic Research*; 3. *Novo Nordisk Foundation Center for Biosustainability*; 4. *Department of Public Health, Aarhus University*; 5 *University of Southern Denmark*

Other instructors and contributors include Stine Scheuer, Mario Garcia Urena, Anders Askeland, Omar Silverman, Andreas Eiset, and Signe Storgaard.

Contact:
Luke W. Johnston, MSc, PhD
Email: luke.johnston@rm.dk
 0000-0003-4169-2616
posters.lwjohnst.com/2021/dda-ws

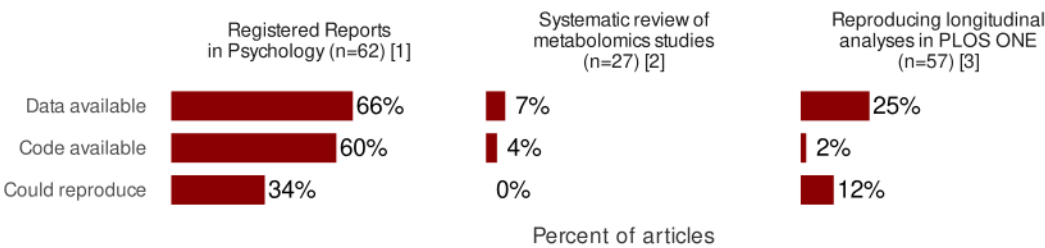


Reproducibility, a core principle of science, is rarely done

- Reproducibility is when the same result is *independently* produced by others using the same data and same code/analysis plan.
- Estimating the reproducibility of scientific studies is currently very difficult because of:
 - Nearly non-existent publishing of code/data
 - General lack of awareness of and training in it
- *Non-replication* is a known major problem, but extent of non-reproducible results is unknown. Barriers to addressing the problem include:
 - Lack of incentives to be reproducible
 - Emphasis on novelty and original work

We don’t share as much as we should



There are few studies on the extent of code and data availability, and whether study results can be reproduced. Figure shows results of some of them: 1) [10.1177/2515245920918872](https://doi.org/10.1177/2515245920918872), 2) [10.1007/s11306-017-1299-3](https://doi.org/10.1007/s11306-017-1299-3), 3) [10.1371/journal.pone.0251194](https://doi.org/10.1371/journal.pone.0251194).

Fundamental changes are needed, including education and training

Our **aim** was then to create an open, re-usable, and beginner-friendly learning module on how biomedical researchers can do **Reproducible Research** using the **R** statistical program (course is abbreviated to “r-cubed” or R3).

Using key principles: Evidence-based learning and teaching practices; Mixture of activities (practicing, listening, reading, typing, discussing); Openly licensed (CC-BY); Publicly accessible; Create safe and supportive environment for learning; Use modern and beginner-friendly software and workflows (e.g. R); Documentation to also re-use for other instructors.



Introductory course details and website

Reproducible Research in R:
An introductory workshop on modern data analyses and workflows

Table of contents

- 1 **Welcome!**
- 2 Syllabus
- 3 Schedule
- 4 Management of R projects
- 5 Version control with Git
- 6 Data management and wrangling
- 7 Analytically reproducible documents
- 8 Data visualization
- Appendix
- A Pre-course tasks
- B Code of Conduct
- C Group project assignment

Want to help out or contribute?
If you find any typos, errors, or places where the text may be improved, please let us know by providing feedback either in the feedback survey (given during class) or by using GitLab.

On  GitLab open an **issue** or submit a **merge request** by clicking the “Edit this page ” link at the side of this page.

1 Welcome!

License: CC BY 4.0 | DOI: [10.5281/zenodo.3921893](https://doi.org/10.5281/zenodo.3921893)

The course is designed as a series of participatory live-coding lessons, where the instructor and learner code together, along with hands-on exercises interspersed throughout the course and a final group assignment to do a simple data analysis project. This website contains all of the material for the course, from reading material to exercises to code to images. It is structured as a book, with “chapters” as lessons, given in order of appearance. We make heavy use of the website throughout the course where code-along sessions almost identically follow the material on the website (with slight modifications for time or more detailed explanations).

On this page

- 1 Welcome!
- 1.1 Target audiences
- 1.2 Re-use and licensing
- 1.3 Previous versions
- 1.4 Contributors
- 1.5 Acknowledgements

[View source !\[\]\(749ee395b1ee375a63364ce4c4c25914_img.jpg\)](#)

[Edit this page !\[\]\(bb35e092de4bc1e595cdf863cb994d1f_img.jpg\)](#)

Website that has material for introductory course: r-cubed.rostools.org. See our paper (DOI: [10.21105/jose.00122](https://doi.org/10.21105/jose.00122)) describing the course and how to use it.



Intermediate course details and website

Reproducible Research in R:
An intermediate workshop on modern approaches and workflows to processing data

Table of contents

- 1 **Welcome!**
- 2 Course syllabus
- 3 Schedule
- 4 Pre-course tasks
- 5 Basic setup and workflow
- 6 Importing data, fast!
- 7 Save time, don’t repeat yourself: Making functions
- 8 Save time, don’t repeat yourself: Using functionals
- 9 Processing and joining datasets for cleaning
- 10 Quickly re-arranging data with pivots

Want to help out or contribute?
If you find any typos, errors, or places where the text may be improved, please let us know by providing feedback either in the feedback survey (given during class) or by using GitLab.

On  GitLab open an **issue** or submit a **merge request** by clicking the “Edit this page ” button on the side of this page.

1 Welcome!

License: CC BY 4.0 | DOI: [10.5281/zenodo.4061900](https://doi.org/10.5281/zenodo.4061900)

Reproducibility and open scientific practices are increasingly demanded of, and needed by, scientists and researchers in our modern research environments. As we our tools for generating data become more sophisticated and powerful, we also need to start using more sophisticated and powerful tools for processing it. Training on how to use these tools and build modern data analysis skills is lacking for researchers, even though this work is highly time-consuming and technical. As a consequence of this unawareness of the need for these skills, how exactly data is processed is poorly, if at all, described in scientific studies. This hidden aspect of research could have major impacts on the reproducibility

On this page

- 1 **Welcome!**
- 1.1 Re-use and licensing
- 1.2 Contributors
- 1.3 Acknowledgements

[View source !\[\]\(62acd955ae780b710b1eae66e386f9e5_img.jpg\)](#)

[Edit this page !\[\]\(2802266ba3b75ce94e860a28fc8f46f1_img.jpg\)](#)

Website with the material for the intermediate course: [r-cubed-intermediate.rostools.org](https://r-cubed.intermediate.rostools.org).

Skills from the course are in demand and needed

We’ve done the introductory course 4 times and intermediate 3 times with the DDA. The responses to the pre-course survey question about what they want and expect to learn emphasize the need of these skills and knowledge:

- “... would like to have a more structured workflow ... and to have some tools for reproducibility”
- “[To] manage data in a more organized way...” and “Version control. Sharing of code.”
- “Git and good practice in the context of reproducible and open science”
- “Better structure in my data analyses so that in the future I can clearly see, what I have done and why.”
- “...create pipelines for my analysis that is more reproducible, readable, and that can be easily re-used for my future self and for my collaborators and colleagues”
- “Structured, reproducible approach to using R and [G]it for future use in analyses”
- “...information about open science and reproducibility...”
- “... components about the reproducibility crisis and Open Science problems, so that we could have a context for why we should be incorporating R in our standard practice.”

Next steps and plans

Possible plans include developing an advanced course, short online tutorials, a video series of the material, and to run frequent (online) coding review sessions.

Are you interested in being involved in any of these or current projects? Contact us!

And of course, a HUGE thanks to DDA for hosting these courses and allowing them to grow ♥ ♥