# An R reproducibility toolkit for the practical researcher

Elio Campitelli ⬤ [1,2,3] **and Paola Corrales** ⬤ [1,2,3]

1 Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales, Departamento de Ciencias de la Atmósfera y los Océanos. Buenos Aires, Argentina. **2** CONICET – Universidad de Buenos Aires. Centro de Investigaciones del Mar y la Atmósfera (CIMA). Buenos Aires, Argentina. **3** CNRS – IRD – CONICET – UBA. Instituto Franco-Argentino para el Estudio del Clima y sus Impactos (IRL 3351 IFAECI). Buenos Aires, Argentina.

Summary

An R reproducibility tool-kit for the practical researcher is a workshop designed to teach researchers, data scientists, and anyone who uses R for research open source tools to work in a reproducible way. The materials are organised in four modules and can be taught or followed in order or independently of each other. Each section includes materials, exercises and detailed exploration of each topic.

The workshop is available at reproducibility.rocks and licensed under the Creative Commons Attribution-ShareAlike 4.0 International License.

Statement of Need

Reproducibility should be part of any data analysis work, but it's particularly important in scientific research. Achieving full reproducibility is challenging and requires, time, resources and specialised tools and skills. This workshop empowers researchers so they can maximize the reproducibility of their work by leveraging existing tools in the R ecosystem –such as RMarkdown, renv, and others–, version control and containers.

The course material is suitable for learners to follow themselves, including exercises and challenges they can complete. The notes for instructors have instructions intended for anyone who would like to teach and/or re-use the material along with slides, assessment checkpoints, and templates for shared document and feedback form.

Description of Learning Modules

The workshop is intended for researchers, data scientists, and anyone who uses R to generate documents and who wants to collaborate with other people (or themselves in the future). Basic prior experience with R is highly recommended.

The materials are organized in four 3-hour modules, ideally covered over a few weeks to allow time for spaced practice (Benjamin & Tullis, 2010) and homework assessment. Each module has specific learning objectives (Table 1) and associated homework with the common goal of taking an existing non-reproducible research project and pushing it further in the reproducibly spectrum with each step.

Table 1: Overview of the 4 modules and learning objectives.

Day #	Learning Objectives	Topics
Day 1: The workbench	1. Create an R project with a specific and useful structure; 2. Extend the use of RMarkdown files for reports and papers; 3. Use LaTeX templates to customize RMarkdown reports.	1. What is reproducibility anyway?; 2. A tidy project is a happy project 3. Reporting with R Markdown
Day 2: Sharing and Contributing	1. Use git to track your work; Use git to collaborate with others; How to set up project documentation; Organizing a project.	1. here are the portable paths, 2. Getting the most of git; 3. Research compendia with rrtools
Day 3: Managing Dependencies and data	1. Manage package dependencies inside a project; 2. How to pack and share data associated wth a project.	1. Managing R dependencies with renv; 2. Sharing data
Day 4: Building Reproducible Computational Environments with Docker	1. Using docker to share a project.	1. Reproducible computational environment with Docker

Instructional design

The workshop can be taught on-line or in person and while we encourage self-taught students to study with a group, it is possible to follow the material alone. The material is taught using a combination of discussions, live coding, demonstration sessions, and exercises (Table 2). For each topic there is a multiple choice question to assess previous familiarity with the topic. Depending on the answers the instructor can decide if the topic needs to be covered in full or if it can be partially or entirely skipped.

Table 2: Pedagogical methods used throughout the workshop

Method	Description
Discussions	Instructors present a topic and ask questions to the students to motivate a conversation.
Live coding or demonstrations	Instructors write code or demonstrate how a tool works and invite students to follow along.
Exercises	Students are invited to solve short exercises, either simply repeating what the instructor just showed, doing something similar, or working in groups to solve a particular problem.
Continuous anonymous feedback	Students are asked at the end of each day to give feedback on 1. Something you like and we should keep and 2. Something you don't like that we should improve.

The Notes for instructors section includes secondary materials and templates that can help to prepare for the workshop.

Experience of use in teaching and learning situations

We delivered this workshop three times on-line. Participants were graduate and undergraduate students, experienced researchers, and private and government sector researchers. Each time two instructors where always present and shared the tasks of teaching, answering questions and troubleshooting technical issues. This also helps to keep the agenda on track and the general organization of the workshop.

The anonymous feedback was positive and many recommendations or requests from the students were included in the following days. The feedback also helped to shape the materials and organisation of the second edition. In particular, the first edition was divided into only three modules, but this proved to be not enough time so the content was reorganised into the current four modules.

Other people have used the materials to learn specific tools and apply them to their work independently.

Story of the project

Elio and Paola started using R in 2014 and by the time they needed to write their undergrad dissertations they rapidly started using rmarkdown and git. They continued using R during their PhDs and started adding new tools to their repertory to make their work easier and as reproducible as possible. In 2021 Carlo Pecoraro from Physalia courses contacted Elio to invite them to teach a workshop. Elio and Paola took the opportunity to compile what they been learning and using for years in a hands-on workshop.

References

Benjamin, A. S., & Tullis, J. (2010). What makes distributed practice effective? *Cognitive Psychology*, *61*(3), 228–247. https://doi.org/10.1016/j.cogpsych.2010.05.004