Using R for reproducible data management and analyses in Forest Sciences

Post-graduate program in Forest Sciences / UnB

Discipline – Transfer of forestry information (Code: EFL360457)

Credits - 4

Workload – 60h (30h presential + 30h of personal work)

Schedule - First semester - Wednesday 14-16h

Focus areas - Nature conservation, Forest management and Technology and use of forest resources

Level– Master and PhD

Professor – Géraldine Derroire, geraldine.derroire@cirad.fr

Learning outcomes

At the end of this module, the students will be able to:

- Define the concept of reproducible research and cite the main tools that it requires
- Understand the basis of R and R studio (interface, packages, basic functions...)
- Import, manipulate and export data in R
- Apply good practices in data management
- Organise their work in a R studio project
- Use the main packages of the Tidyverse for tidy analyses
- Choose the appropriate graphics to explore their data and create them using ggplot
- Implement basic statistical tests in R
- Understand and implement more advanced programming techniques (loops, functions...)
- Use the tool Quarto to produce a report or a presentation with text and code

Targeted audience

This module is open to students from the three focus areas, level Master and PhD.

No previous experience with R is required, as the course will start with the basics.

Teaching methods

The teaching methods will emphasise on interactivity and practical application. For this, each class will contain a mix of formal teaching, practical exercises, and discussion.

The module will be taught in a mix of English and Portuguese: most presentations will be in English, but questions and students interventions can be made in either English or Portuguese. Course material (bibliographic references, resources, video, etc) will be provided in a balanced mix of English and Portuguese. This pedagogical choice aims at encouraging the students to work and communicate in English, in a friendly and inclusive way.

Program

Session	Topic
1	Introduction of the course, the professor and the students Principles of reproducible science
2	Getting started with R 1/3
3	Getting started with R 2/3
4	Getting started with R 3/3
5	Manipulating data
6	Tidyverse 1/2
7	Tidyverse 2/2
8	Graphical exploration 1/2
9	Graphical exploration 2/2
10	Basic statistics with R 1/2
11	Basic statistics with R 2/2
12	Going further with R 1/2
13	Going further with R 2/2
14	Version control with git and GitHub
15	Letterate programming using quarto

Assessment

Student's learning will be evaluated through:

- Short exercises given at the end of each class (1/3 of the final mark)
- A personal project using the tools covered in class (2/3 of the final mark). This project will be submitted as a pdf file and a GitHub repository. It will comport some data exploration and basic analyses. Students will be encouraged to work on their own data if they have already collected some. Alternatively, data can be provided by the professor. In any case, students are strongly encouraged to discuss the topic of their work early on with the professor. They will not be evaluated on the relevance of their

scientific questions and results, but on their capacity to implement a reproducible workflow in R to explore and test their hypotheses.

Practical aspects

Students will be encouraged to practice on their own laptop, in order to have all the programs installed for their personal work during and after the module. Should this be a problem, please inform the professor no later than the 15th of March so that an alternative solution can be looked for.

Students are asked to have installed the required programs before the class. For this, a detailed tutorial will be given.

Recommended bibliography

Additional resources will be proposed during the module.

Wickham, H., Çetinkaya-Rundel, M., & Grolemund, G. (2023). *R for data science* (2nd éd.). O'Reilly Media. http://r4ds.had.co.nz/

Wickham, H. (2017). *ggplot2 : Elegant graphics for data analysis* (2nd éd.). Springer. https://ggplot2-book.org/

Navarro, D. (2019) *Learning statistics with R: A tutorial for psychology students and other beginners.* https://learningstatisticswithr.com/book/

Damiani, A. Milz, B., Lente, C., Falbel, D., Correa, F.? Trecenti, J., Luduvice, N., Lacerda, T., Amorim, W. (2022) *Ciência de Dados em R*. https://livro.curso-r.com/index.html