

Master in International and Development Studies

Academic year 2022-2023

Fundamentals of R

MINT - Autumn - 3 ECTS

Friday 23rd September, 14:15 to 16:00
Friday 30th September, 14:15 to 17:45
Friday 7th October, 14:15 to 17:45
Friday 14th October, 14:15 to 17:45
Friday 21st October, 14:15 to 16:00

Course Description

R is a programming language and open software that allows users to import, transform, and analyse diverse types of data. Academics, governments, and industry use R for statistical analysis, data visualisation, and machine learning. This course is a hands-on introduction to programming with R, starting from scratch. While we prioritise application, we are conscious that no technique is a substitute for clear conceptual thinking. Classes cover a core task in R by illustrating them in a topical case study with their conceptual underpinnings.

Syllabus

Course Prerequisites

This course assumes no prior knowledge of R software or programming. While this is not a statistics course and no prior statistics knowledge is required, prior knowledge of the introductory statistics can be useful.

Learning outcomes

By the end of this course, students should be able to (1) perform simple data analysis, (2) communicate findings with visualisations, and (3) produce integrated reports using R.

The course provides students with the basic tools so that they can continue independently with more complex tasks in R and leverage the community for help whenever necessary.

LECTURERS

Henrique Sposito,
PhD Candidate IRPS,
henrique.sposito@graduateinstitute.ch

Livio Silva-Muller,
PhD Candidate ANSO,
livio.silva@graduateinstitute.ch

Course policies

Learning R can be **fRustRating**. This course is designed to allow proper time for each task and provide the necessary support to students.

This course is not reading extensive, but students are expected to spend time with the assignments and getting acquainted with R. Classes have mandatory and suggested readings, indicated in each class in the course outline below and in the course Moodle. Readings are available in the embedded links or in the course Moodle page.

The number of places is limited. Priority will be given to 2nd year MADIS OR MINT students. We recommend this course for students that are taking or plan to take other courses using R.

The lecturers reserve the right to make alterations to this syllabus.

Assessment

- *Attendance and participation (25%)*

Students should attend courses and engage with peers.

- *Weekly assignments (25%)*

Students should complete three small assignments in R to get acquainted with the software and internalise learning outcomes. Weekly assignments will be made available after the respective lecture they cover and be due before the following lecture.

- *Final assignment (50%):*

Students should complete a final report in R that reviews all the weekly assignments.

Course Materials

Students should try to have [R](#) and [RStudio](#) installed before the first lecture.

Class materials (readings, scripts, case studies, and presentations) will be available on the moodle page. Case studies will be designed and selected according to the students' interests.

While there is no required textbook for this course, we recommend the following books:

Douglas, A., Ross, D., Mancini, F., Couto, A. & Lusseau, D. (2022). [An Introduction to R](#).

Wickham, H., & Grolemund, G. (2017). [R for data science: import, tidy, transform, visualize, and model data](#).

Wickham, H. (2016). [ggplot2: Elegant Graphics for Data Analysis](#).

Xie, Y., Allaire, J. J., & Grolemund, G. (2018). [R markdown: The definitive guide](#).

Mesquita, E. B., & Fowler, A. (2021). Thinking clearly with data: A guide to quantitative reasoning and analysis. Princeton University Press.

Healy, K. (2018). Data Visualisation: a practical introduction. Princeton University Press.

COURSE OUTLINE

Introduction to R (Friday 23rd September, 14:15 to 16:00)

We start our course by understanding what R is, how to open the software, and a few basic concepts. To illustrate the skills we hope to work through during the semester with R, we will delve into a case study on “who makes international regimes possible?”.

- What is R language and software?
- What is R used for?
- How to interact with R studio?

Mandatory readings:

- Thinking Clearly in a Data-Driven Age (Chapter 1) in Mesquita, E. B., & Fowler, A. (2021). Thinking clearly with data: A guide to quantitative reasoning and analysis. Princeton University Press.
- [Getting started with R and RStudio](#) (Chapter 1) in Douglas, A., Ross, D., Mancini, F., Couto, A. & Lusseau, D. (2022). [An Introduction to R](#).

Suggested readings:

- Introduction and preliminaries (Chapter 1) in Venables, W. N., Smith, D. M., & R Development Core Team. (2022). [An introduction to R](#).

Block 1: Objects, functions, and data types (Friday 30rd September, 14:15 to 17:45)

R holds different objects with diverse classes. This lecture helps students understand what R classes are and how R functions operate on these. Moreover, we will learn how to properly get external data into R, one of the most important tasks new users should be able to perform. RStudio and several other R packages have made this, and many other, task easier and flexible.

- What are object classes and functions in R?
- How to perform arithmetic operations in R?
- How to import text (.csv) and excel (.xlsx) files to R?
- How explore newly imported data in R?

Mandatory readings:

- [Some R basics](#) (Chapter 2) in Douglas, A., Ross, D., Mancini, F., Couto, A. & Lusseau, D. (2022). [An Introduction to R](#).
- [Data in R](#) (Chapter 3) in Douglas, A., Ross, D., Mancini, F., Couto, A. & Lusseau, D. (2022). [An Introduction to R](#).
- Data import (Chapter 11) in Wickham, H., & Grolemund, G. (2017). [R for data science: import, tidy, transform, visualize, and model data](#).

Suggested readings:

- Simple manipulations; numbers and vectors (Chapter 2) in Venables, W. N., Smith, D. M., & R Development Core Team. (2022). [An introduction to R](#).

- Grolemund, G., & Wickham, H. . (2011). Dates and Times Made Easy with lubridate. Journal of Statistical Software, 40(3), 1–25. <https://doi.org/10.18637/jss.v040.i03>
- Lists and data frames (Chapter 6) and Reading data from files (Chapter 7) in Venables, W. N., Smith, D. M., & R Development Core Team. (2022). [An introduction to R.](#)

Block 2: Wrangling and cleaning data (Friday 7th October, 14:15 to 17:45)

Data often does not come ready for the analysis you want to perform. Names within variables might not match (e.g. USA, US, United States), observations are missing (e.g. NA), among other human or programmatic coding errors. In this lecture, we will go through the fundamentals of (re)shaping data for our purposes using the [dplyr](#) package.

- How to create and modify variables in data?
- How to group, filter, and summarise information contained in variables?
- How to merge multiple datasets?
- How to detect inconsistencies in data and harmonise strings?
- How to treat missing data?

Mandatory readings:

- [Data transformation](#) (Chapter 5) in Wickham, H., & Grolemund, G. (2017). [R for data science: import, tidy, transform, visualize, and model data.](#)
- [Data Wrangling with dplyr in R](#) (Gonzalez, K. 2021)
- [Regular Expressions](#) (Chapter 17) in Peng, R. D. (2022). [R programming for data science.](#)
- [Handling missing data in R](#)

Suggested readings:

- Wickham, H. . (2014). Tidy Data. Journal of Statistical Software, 59(10), 1–23. <https://doi.org/10.18637/jss.v059.i10>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. A., François, R., & Yutani, H. (2019). [Welcome to the Tidyverse.](#) Journal of open source software, 4(43), 1686.
- [Regular Expressions](#) with the [stringr](#) R package
- Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). [mice: Multivariate imputation by chained equations in R.](#) Journal of statistical software, 45, 1-67.

Block 3: Visualizations (Friday 14th October, 14:15 to 17:45)

Visualising your data is the bread and butter of any analysis you will perform. In this lecture, we will learn the principles of good data visualisation with the [ggplot2](#) package. We introduce how to create different types of plots in R as well as how to customise plots with trends, annotations, and labels to make them “tell a story”.

- How to create bar charts, scatter plots, box plots, and line trends?
- How to quickly visualise variation using facets?
- How to include trend lines in plots?
- How to plot multiple relationships between variables?
- How to annotate information to make plots more informative?

Mandatory readings:

- [Data visualisation](#) (Chapter 3) in Wickham, H., & Grolemund, G. (2017). [R for data science: import, tidy, transform, visualize, and model data.](#)

- Look at the Data (Chapter 1) in Healy, K. (2018). Data visualisation: a practical introduction. Princeton University Press.
- Graph Tables, Add Labels, Make Notes (Chapter 5) in Healy, K. (2018). Data visualisation: a practical introduction. Princeton University Press.

Suggested readings:

- [First steps](#) (Chapter 2) in Wickham, H. (2016). [ggplot2: Elegant Graphics for Data Analysis](#).
- [Annotations](#) (Chapter 8) in Wickham, H. (2016). [ggplot2: Elegant Graphics for Data Analysis](#).
- [Graphics for communication](#) (Chapter 28) in Wickham, H., & Grolemund, G. (2017). [R for data science: import, tidy, transform, visualize, and model data](#).

Exchanging with the community (Friday 21st October, 14:15 to 16:00)

Now that you know the fundamentals, you will want to share results and ask for help. In this lecture, we will learn how to use R [Markdown](#) to export reproducible reports in different file formats (e.g. .pdf, .docx, and html). We will also discuss how to ask for help online using minimal reproducible examples. We will use R to illustrate some more advanced research applications and discuss “how the Amazon has been constructed as a problem in Brazilian presidential discourses?”

- What is R Markdown?
- How to export integrated reports ready for sharing?
- How to create minimal reproducible examples?

Mandatory readings:

- Introduction and Bomb Parts: What is a Model? (Chapter 1) in O'neil, C., 2016. Weapons of math destruction: How big data increases inequality and threatens democracy. Broadway books.
- [Basics](#) (Chapter 2) in Xie, Y., Allaire, J. J., & Grolemund, G. (2018). [R markdown: The definitive guide](#).
- [How to make a reprex?](#)

Suggested readings:

- [R Markdown](#) (Chapter 27) in [R for data science: import, tidy, transform, visualize, and model data](#).
- [Weapons of Math Destruction](#) (O'Neil, C. 2016)

Final assignment due on November 13th, 23:59.