

Data analysis with R

Lecturer: Maria Chiara Debernardi

Language

English

Course description and objectives

R is an open-source programming language specifically designed for data analysis. Thanks to a rich ecosystem of free packages, it supports a wide range of statistical tasks - from descriptive and inferential analyses to data mining, machine learning, and text analysis - and it is increasingly used in both academia and industry.

This course introduces the R language and the RStudio (Posit) development environment, focusing on the core workflow for data analysis, with particular attention to the *tidyverse*.

During the course, students will learn the fundamentals of R's data structures and syntax, and will practice a modern, reproducible approach to data analysis through hands-on exercises.

Upon successful completion of this course, students will be able to:

- work with the main data objects in R and manage data types
- perform descriptive analyses using tables, summary statistics, and effective visualizations
- compute and interpret confidence intervals, hypothesis tests, linear regression models, and ANOVA
- write simple ad hoc functions and use basic programming constructs
- use key *tidyverse* packages to build clear and readable analysis pipelines

Important notice: this course focuses on R language and RStudio. It is not intended to replace a formal course in Statistics; therefore, statistical theory will be addressed only at an introductory level.

Audience

The course is open to all Bocconi students. In particular, it is intended for students from the second year of the Bachelor's program onwards who are interested in carrying out statistical analyses, both for their thesis project and for their future professional work.

Prerequisites

No prior coding experience or knowledge of R is assumed.

Students should be comfortable with basic computer operations and common productivity software (e.g., spreadsheets).

Basic knowledge and understanding of descriptive and inferential statistics are strongly recommended.

Guidelines

Registration:

You can sign up for the course only through the yoU@B student Diary, in the "**Sign-up for various activities**" box (please note that the box appears only when registrations open. Before then it will not be visible).

You may only cancel your registration by Diary **no later** than the registration deadline for the course itself. No other methods of cancellation are allowed.

Registration will be confirmed a few days before the start of the course through a message posted in the yoU@B student Diary.

Attendance:

- Attendance of **75% or more** of the lessons: earning of the **Open Badge**
- Attendance of **less than 25%** of the lessons: placement on the **Exclusion List**

Duration

16 academic hours

Teaching mode

This course is taught **in person** only. Online mode will not be provided.

Calendar

Lecture	Date	Time	Room
1	Wed 25/03/2026	18.15 - 19.45	N33 (L. Del Vecchio)
2	Thu 26/03/2026	18.15 - 19.45	N33 (L. Del Vecchio)
4	Wed 01/04/2026	18.15 - 19.45	N33 (L. Del Vecchio)
3	Thu 02/04/2026	18.15 - 19.45	N33 (L. Del Vecchio)
5	Wed 15/04/2026	18.15 - 19.45	N33 (L. Del Vecchio)
6	Thu 16/04/2026	18.15 - 19.45	N33 (L. Del Vecchio)

7	Wed 22/04/2026	18.15 - 19.45	N33 (L. Del Vecchio)
8	Wed 29/04/2026	18.15 - 19.45	N33 (L. Del Vecchio)

Note: lessons will be held in a standard classroom. **Students are required to bring their own laptop.**

Syllabus of the course

Lecture	Topics
1	Introduction to R and RStudio <ul style="list-style-type: none"> - R and the RStudio (Posit) IDE - Packages and CRAN repositories - Getting help in R - First steps with R - Scripts and projects <i>Exercises</i>
2	Data in R <ul style="list-style-type: none"> - Atomic data types: numbers and strings - Vectors: the basic building block - Matrices/arrays, lists, data frames (<i>tibbles</i>) - Type conversion and coercion - Factors <i>Exercises</i>
3	A first look at data <ul style="list-style-type: none"> - The <i>tidyverse</i> - Importing and exporting data - Frequency distributions - Univariate and bivariate summary statistics <i>Exercises</i>
4	Data visualization and manipulation <ul style="list-style-type: none"> - Data manipulation with <i>dplyr</i> - Chaining operations with pipes - Visualizing data with <i>ggplot2</i> and other graphics packages <i>Exercises</i>

Lecture	Topics
5	Statistical inference in R <ul style="list-style-type: none"> - A practical workflow for statistical analysis - Confidence intervals and hypothesis tests - Linear regression and ANOVA - Variable transformations and feature engineering basics - Missing data handling - Detecting and treating outliers <i>Exercises</i>
6	Programming with R <ul style="list-style-type: none"> - Control structures (if/else, for/while) - Writing your own functions - Reproducible reporting <i>Exercises</i>
7	Time series analysis <ul style="list-style-type: none"> - Working with dates and times in R - The <i>lubridate</i> package - Exploratory analysis and autoregressive models <i>Exercises</i>
8	Data mining in R <ul style="list-style-type: none"> - Typical problems and how to address them in R - Model training and evaluation (e.g., <i>tidymodels</i>, <i>caret</i>) - Model explainability (e.g., <i>DALEX</i>, <i>iml</i>, <i>lime</i>) <i>Exercises</i>

Software used

- 1) **R language** (r-project.org): latest stable release (R 4.5.2 or later)
Download from a [CRAN mirror](#) the version suitable for your operating system
- 2) **RStudio Desktop** ([Posit](#)): latest stable Desktop Open Source / Free version (2026.01.1+403) downloadable [here](#)

Suggested bibliography

H. Wickham, M. Çetinkaya-Rundel, G. Grolemund, *R for Data Science* (2nd Edition), O'Reilly Media, 2023

The book is freely available [online](#).

Available seats

This activity is limited to **110** participants. Registration cannot be conducted once capacity is reached or after the registration period ends.

Please note that you may unsubscribe from ITEC courses only through the yoU@B Diary and **only before the registration deadline.**