School of Social Sciences
Chair for Social Data Science and Methodology

Introduction to R

The University of Mannheim
Chair for Social Data Science and Methodology
University of Mannheim, School of Social Sciences

Launch: Spring 2023

Course format: Online course (self-study)
Access link: `https://ilias.uni-mannheim.de/goto.php?target=crs_1334688&client_id=ILIAS`
Open to: All students of the University of Mannheim
Instructor: Lion Behrens, M.Sc.

Responsible chair holder: Prof. Florian Keusch
Contact: `f.keusch@uni-mannheim.de`

Course Description

This self-paced open online course provides an introduction to the statistical programming language R and is designed for both B.A. and M.A. students of the University of Mannheim. The course material is centered around the core tasks that are necessary to perform statistical analyses underlying quantitative research in the social sciences. Upon successful completion of the course, students are capable to perform the complete workflow of a quantitative research project in R, including data import, data cleaning and wrangling, exploratory data analysis, statistical modeling and visualization, as well as professionally reporting the results. Additionally, students will have had first exposure to writing efficient code (that includes the use of loops and functions).

The material is constructed around six core modules. Each core module is divided in a number of individual chapters. In each chapter, (i) a set of PDF slides, (ii) a pre-recorded video lecture, as well as (iii) illustrative and ready-to-use R code are provided. At suitable points in the material, students are forwarded to short self-contained online exercises for self-study to practice the material.

Prerequisites

There are formally no prerequisites for this course except of an open mind, the willingness to learn, and (hopefully) the ability to have fun while coding. The course material assumes that students are familiar with basic statistical concepts such as descriptive statistics, statistical inference and regression modeling (linear regression, logistic regression) at the B.A. level and does not cover any background on statistical modeling itself. Rather, we focus on how to implement statistical analyses in R.

Additional Resources

While we believe that upon completing this course, students will be well-equipped to perform their statistical analyses in R on their own, there is a large number of resources available online to further improve your R coding skills. Some of these are better than others. The following is a short list of recommended material:

- Wickham, Hadley and Garrett Grolemund, 2017. R for Data Science - Import, Tidy, Transform, Visualize, and Model Data. O'Reilly. Access: `https://r4ds.had.co.nz/`.
- Wickham, Hadley, 2019. Advanced R. CRC Press. Access: `https://adv-r.hadley.nz/index.html`.
- Imai, Kosuke, 2018. Quantitative Social Science: An Introduction. Princeton University Press.

Installing R and RStudio

R is the statistical programming language that this course is based on and can be downloaded for free at `http://www.r-project.org/`. RStudio is a dynamic user interface, which makes coding in R handy, efficient and visually pleasing. RStudio is a product provided by Posit and can be downloaded for free at `https://posit.co/download/rstudio-desktop/`. In Module 1 *The R Environment* of this course, an installation manual for R and RStudio can be found that provides a step-by-step guide how to set these up.

Course Outline

Getting Started

---

1 The R Environment

- 1.1 Overview of the Course
- 1.2 Why R?
- 1.3 R, RStudio, Basic Functionality
  (*Self-paced exercise*: `https://lionbehrens.shinyapps.io/1-3exercise`)
- 1.4 Object Classes
- 1.5 Accessing, Subsetting and Naming Objects
  (*Self-paced exercise*: `https://lionbehrens.shinyapps.io/1-5exercise`)
- 1.6 Tips to Get Started

R Basics

2 Data Manipulation

- 2.1 Data Wrangling - Tidyverse Philosophy
- 2.2 Loading and Storing Data
- 2.3 Re-Naming and Re-Ordering Rows and Variables - rename(), relocate(), arrange()
- 2.4 Subsetting Rows and Variables - select(), filter()
- 2.5 Transforming Variables - Class conversions, mutate(), recode()
- 2.6 Transforming Variables - summarize(), group_by()
- 2.7 Merging Data Frames
- 2.8 Missing Values
- 2.9 The Data Wrangling Pipeline

3 Exploratory Data Analysis

- 3.1 Summary Statistics, Frequency Tables, Cross-Tabulations, Correlation Matrices

4 Data Visualization With ggplot2

- 4.1 Constructing One Plot Step-by-Step
- 4.2 Plotting Anything

5 Regression Modeling

- 5.1 Overview
- 5.2 Visualizing and Exporting Results
- 5.3 Generalized Linear Models

Efficient Workflow

6 For-Loops and Functions

- 6.1 For-loops
- 6.2 Functions