R Basics

ECPR Methods Summer School 2019, Budapest

Instructor: Akos Mate, CEU

Email: aakos.mate@gmail.com (or just write me on Twitter (aakos_m) if you are in a hurry)

Office hours: upon appointment (either through email or in persion)

Teaching assistant: Daniel Kovarek, CEU

Email: Kovarek_Daniel@phd.ceu.edu

Office hours: upon appointment (either through email or in persion)

Course description

It is not a stretch to say that R has become one of the main data analysis tools used in and outside of academia. R is an open source programming language, developed for statistical computing, that developed an extremely active user base with an expanding universe of packages.

The guiding logic of the course is to give practical knowledge for the whole data analysis workflow:

1. Importing data
2. Data wrangling/cleaning
3. Data visualization
4. Analysis
5. Reporting the results

It might be strange to switch from SPSS or Stata to R, but the benefits outweigh the efforts of climbing the learning curve. The base R allows us to read different data files into R, manipulate them, create various visualizations and run statistical analysis of any sorts (from basic descriptives to time series analysis, or multilevel regressions). The real value in learning R is that it integrates the research workflow into one environment. It can also be adapted to a broad range of research, from party politics data to ecological modelling.

Course outline

Day 1 (Friday 26 July 13:00–15:00 and 15:30–18:00)

1. Intro to R - notebook

2. Importing, exporting and exploring data - notebook
3. Data wrangling pt.1 - notebook
4. Data wrangling pt.2 - notebook

Day2 (Saturday 27 July 09:00–12:30 and 14:00–17:30)

5. Data visualization with `ggplot2` - notebook
6. Writing functions and iterating in R - notebook
7. Some statistics (linear regression, hypothesis testing) - notebook
8. Intro to Rmarkdown - notebook

Day 1

We start with a general introduction to R and RStudio. We learn how to start coding and set up RStudio to make our workflow as seamless as possible. RStudio is an Integrated Development Environment (IDE) that puts together the R console, a text editor where we write the code, and an object viewer where we can view the data objects we created.

The general introduction will cover how to use R for basic mathematical calculations, and how to create different objects. This part is key because we will cover the base R syntax, how to create / access / remove objects, and how to merge vectors into data frames. These are essential operations for the following sessions.

We also look at how to load data into R from commonly encountered sources, such as .txt, .csv, Excel sheets, Stata, SPSS and SAS save files. After getting data into R, we will perform some basic operations to have a sneak-peek at the data. This includes the usual descriptive statistics and creating histograms and scatterplots.

Day 2

Dedicated to data manipulation and data cleaning, with a hint of data analysis. This is an essential part (which usually takes up the majority of the time) of every analysis. The materials will cover how to set up data in R, what the difference is between the wide and long data format and the more recent push for 'tidy data' in the R community. I will introduce writing loops and functions in R and the 'apply' function family as well as their 'tidy' alternatives.

Similarly to the previous day, all the activities are accompanied with some degree of data visualization, since it is often better to show a figure than a disorienting half-page table. At this point we will have enough results to think about getting them out of R. The course uses RMarkdown to show how to create PDF or html output of our work. There are also several packages developed for getting results out of the R console.

If you have any particular interests during the course, I try to cover those in this final session.

Readings

We will work with practical examples, so no compulsory readings are required. However, for further exploration with R see:

- Wickham, Hadley and Grolemund, Garrett 2017: **R for data science** (available online)
- Healey, Kieran 2019: **Data visualization – A practical introduction**, Princeton University Press (available online)
- Grolemund, Garrett 2014, **Hands-On Programming with R**, O'Reilly
- Wickham, Hadley 2014, **Advanced R**, Chapman and Hall/CRC (online companion)
- Matloff, Norman 2011, **The Art of R Programming: A Tour of Statistical Software Design**, No Starch Press