0A: Introduction to Programming in R for the Social Sciences

Essex Summer School 2023

Johannes Karreth

Associate Professor, Department of Politics & International Relations
Ursinus College, USA
Email: jkarreth@ursinus.edu
Course website: http://www.jkarreth.net/intror-essex.html (with links to course materials)
Course meetings: M–F, July 3–7, 2022 / 10:00-13:30 BST BST/UTC+1 via Zoom (core course time)
Teaching assistant: Maya Jeyabraba, University of Essex (maiyoraa.jeyabraba@essex.ac.uk)

This course will run in a virtual format in 2023.

Course description and goals

The course will introduce participants to the open-source statistical software R. R is a highly versatile software environment suitable for introductory and advanced quantitative social science and data analysis. The course offers participants a near-complete foundation to use R in other courses at ESS.
Specifically, the course will explore the following topics:

- Introduction to the R language and software architecture
- Use of the tidyverse suite of R packages
- Incorporating R code and document production (R Markdown)
- Workflow, reproducibility, and version control in R
- Data import and data management, including working with "messy" datasets
- Data visualization
- Basic functions

Upon successful completion of this course, participants will have acquired intermediate R skills aligned with most ESS courses that rely on R. The course is suitable for researchers at the beginning of their quantitative training as well as researchers with advanced background in quantitative social science wishing to acquire a new, free, open-source, and highly versatile set of tools. A workflow for reproducible data analysis is also a core element of the course. The course content will be reinforced through regular hands-on exercises and frequent feedback from the instructor.

Remote learning setup

Lectures will run on Zoom. Participants are invited to interrupt and ask questions any time. Much of the course will consist of 1-on-1 or small-group exercises with the teaching staff. These will take place during the core course time. I will also be available for virtual office hours during every day of the workshop, both during fixed times and by appointment.

Prerequisites

This is an introductory, highly accessible course. No prior knowledge of statistical or quantitative methods, or R, or computer programming is required. Participants with experience in other tools (e.g. SPSS, Stata, or SAS) will find the course structure helpful to transfer their skillsets into R.

Literature

We will use the following text, which is available at no cost as an e-book at `https://r4ds.hadley.nz`:

- Wickham, Hadley, Mine Çetinkaya-Rundel, and Garrett Grolemund. 2023. *R for Data Science*. Sebastopol, CA: O'Reilly.

Further readings and materials will be made available to participants during the course.

Software and Preparation

Participants will be asked to install R and RStudio on their personal laptops during the first course meeting. We will go over how to use these programs on the first day of the course, using a detailed tutorial with step-by-step instructions. We will also have time to catch up on installation problems on the first day.

Course schedule

For each day, the core reading usually provides substantial details for the units discussed on that day. A typical course period will consist of the following:

- **Lectures** are self-contained mini-units mixing lecture and discussion.
- **Exercises** are problem sets that participants complete during the core course time to reinforce the material learned in the course on that respective day.

The following time slots and topics will likely be modified as the course proceeds. The most current version of this document can be found at `http://www.jkarreth.net/intror-essex.html`. Chapter numbers refer to the chapters in the 2nd edition of R4DS (see above).

Day	Unit	Topic	Chapter
Mo 7/3	1	Introduction to the R language and software architecture	1-2
Mo 7/3	2	Data visualization & workflow	3-5
Tu 7/4	3	Data transformation	5
Tu 7/4	4	Exploratory data analysis	10-11
We 7/5	5	Data import & data management	8, 13-21
Th 7/6	6	Working with R objects efficiently and effectively	26-27
Th 7/6	7	R for statistical modeling	
Fr 7/7	8	Presentation: RMarkdown/Quarto	29-30
Fr 7/7	9	Presentation: Tables and graphs	12
Fr 7/7	10	Best practices and Q&A	