Computer programming E140

Christian Dudel

September 2, 2025

What will be covered in this course?

- Software: R (and RStudio)
- Mostly basic things in these programs
- Course will not cover many things
- Course will not be deep

Goals

At the end of this course...

- ...you understand basic concepts of R
- ...you can do a basic analysis in R

Prerequisites

- Basic demographic knowledge (e.g., you know what a ‘rate’ is)
- Basic statistical knowledge (e.g., you know what a ‘mean’ is)
- First experince using statistical software (Stata, Excel, SAS, SPSS, R, . . .)

Already an expert?

- Write a function that takes any integer as input and returns TRUE if the integer is a prime number, otherwise it returns FALSE
- Write a function that takes any integer as input and returns its amicable number if it exists, otherwise it returns FALSE
- Check one of the many tasks on rosettacode.org solved with R
- Reproduce or replicate one of the many reproducible/replicable articles available at Demographic Research
- Work on your own analysis

Materials

Materials will be available from GitHub, also mirrored on OSF:

- https://github.com/christiandudel/EDSD2025
- https://osf.io/zspkq/

Contact

- Email: dudel@demogr.mpg.de
- Office: 359 (3rd floor, east wing)
- Website: http://www.christiandudel.com

Things I work on/I am interested in

- *Topics*: Labor markets, fertility, health
- *Concepts*: Stratification, inequality, life courses, aging
- *Methods*: Longitudinal data analysis, causal inference, (partial) identification, survey methodology

Course schedule

- September 2 (Tue), 09:30 - 11:30: Basics
- September 4 (Thu), 09:30 - 11:30: Descriptive statistics
- September 9 (Tue), 09:30 - 11:30: Data visualization
- September 11 (Thu), 13:30 - 15:30: Data handling
- September 22 (Mon), 09:30 - 11:30: Example
- September 24 (Wed), 13:30 - 15:30: Programming (1)
- October 14 (Tue), 13:30 - 15:30: Programming (2)
- October 15 (Wed), 13:30 - 15:30: Programming example

Types of session

- Regular session (6)
- Example session (2)

Regular sessions

- I show things
- You bring your laptop and follow
- Always possible to ask questions!

Example sessions

- Real applications
- Combine material from previous sessions

(Voluntary) exercises

- For some sessions, there will be voluntary exercises
- Exercises and solutions are available online (GitHub/OSF)
- These voluntary exercises have to be distinguished from the (mandatory!) assignment

Assignment: Overview

- One mandatory assignment
- Assignment handed out on September 22
- Deadline: October 17
- Assignment will consist of several tasks

Assignment: Your solutions

- You submit R code as solutions (via email)
- R code should be commented, explaining what is happening
- Code should work without errors

Assignment: Deadline

- Deadline assignment: October 17, 12am/midnight/24:00 (CEST/Berlin time)
- Send your solutions to me (dudel@demogr.mpg.de)
- You will get a confirmation (might take a bit, sorry)
- I might get back to you if I have problems with your file(s)
- It is your responsibility that your files are working!

Assignment: Groups

- You can work in groups
- Actually, I strongly suggest you work in groups!
- Please not more than five people per group
- Please submit your solutions only once per group
- Make clear who is member of the group when submitting

Assignment: Grading

- You can either “pass” or “fail”
- Your code should. . .
- . . . work without errors
- . . . be well-documented: Comments!
- . . . should be (somewhat) efficient. If one step can do the work then don’t use two or more!

Assignment: Summary

- One assignment consisting of coding tasks
- You submit code as solutions, via email
- Deadline: October 17
- You can work in groups
- Pass/fail

Other dates

- September 23 (Tue), 16:00: MPIDR Summer Party
- September 26 (Fri): EDSD Social Demography Day
- October 9 (Thu): Staff outing
- Drinks?

What is R?

- R is an open source statistical programming language
- First release in 1995
- Used for data analysis and statistical programming

Why use R?

- Free, open source
- Can easily be extended
- More than 22,600 packages available on CRAN
- Many methods are already implemented in R
- Commonly used in both science and industry
- Many R-related materials: Books, journals, conferences, forums, tutorials...

Why use RStudio?

- R is the programming language
- RStudio is a tool to use R more efficiently (IDE)
- Features:
- Syntax highlighting, code folding
- Project management (e.g., GitHub)
- Markdown support
- ...

Disclaimer

- R is not the only statistical software and it is fine if you prefer something else
- RStudio is not the only IDE/editor for R (ESS, RKWward, Tinn-R, ...)
- R can be used in many different ways
- Example: base R vs tidyverse vs data.table vs specialized packages
- I do things in certain ways, and this course will follow that
- This does not mean that the solutions from this course are the only or the best way to do things

What do you need to get started?

- R: https://cloud.r-project.org/
- R-Studio: https://www.rstudio.com/