

Computer programming E140

Christian Dudel

September 4, 2024

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 experience 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/EDSD2024>
- ▶ <https://osf.io/p6rh8/>

Contact

- ▶ Email: dudel@demogr.mpg.de
- ▶ Office: 359 (3rd floor, east wing)
- ▶ Twitter: @c_dudel
- ▶ 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, identification, survey methodology

Course schedule

September 4 (Wed), 09:30-11:30: Introduction and basics

September 4 (Wed), 13:30-15:30: Descriptive statistics

September 5 (Thu), 14:00-15:00: Tutorial

September 6 (Fri), 09:30-11:30: Data visualization

September 6 (Fri), 13:30-15:30: Data handling

September 9 (Mon), 09:30-11:30: Programming (1)

September 11 (Wed), 13:30-15:30: Programming (2)

September 12 (Thu), 14:00-15:00: Tutorial

September 13 (Fri), 09:30-11:30: Programming (3)

October 18 (Fri), 09:30-11:30: Example

Types of session

- ▶ Regular session (7)
- ▶ Tutorial session (2)
- ▶ Final example session (1)

Regular sessions

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

Tutorial sessions

- ▶ You solve exercises
- ▶ I am there to help
- ▶ Exercises and solutions are available online (GitHub/OSF)
- ▶ Voluntary
- ▶ Has to be distinguished from the (mandatory!) assignment

Final example session

- ▶ Real application
- ▶ Several options
 - ▶ Birth register data (fertility rates etc.)
 - ▶ Agent-based simulation (Schelling's segregation model)
 - ▶ COVID-19 fatality
- ▶ Open for suggestions!

Assignment: Overview

- ▶ One mandatory assignment
- ▶ Assignment handed out on September 13
- ▶ Deadline: October 20
- ▶ 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 20, 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 20
- ▶ You can work in groups
- ▶ Pass/fail

Other dates

September 26 (Fri), 13:00-17:00: Social Demography Recruitment Day (prelim. title)

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 21,000 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/>