

Computer programming E140

Christian Dudel

September 23, 2020

COVID-19

- ▶ Course will be completely online
- ▶ First online course I teach, different from previous courses
- ▶ Challenging situation for everyone, let me know if too much
- ▶ Feel free to skip classes (everyone needs to do assignment, though)

Zoom

- ▶ Please mute your microphone when not speaking
- ▶ Feel free to turn off video
- ▶ Feel free to interrupt me any time (unmute first)...
- ▶ ...for instance, if you cannot hear me well...
- ▶ ...but be aware that there might be some delay
- ▶ I might miss things you write in the Zoom chat

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

Contact

- ▶ Slack (see email)
- ▶ Email: dudel@demogr.mpg.de
- ▶ Office: 359 (3rd floor, east wing)
- ▶ Twitter: @c_dudel
- ▶ Website: <https://sites.google.com/view/cdudel>

Things I work on/I am interested in

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

Course schedule

September 23 (Wed), 16:30-18:30 (CEST)

September 25 (Fri), 16:30-18:30 (CEST)

September 30 (Wed), 16:30-18:30 (CEST)

October 2 (Fri), 16:30-18:30 (CEST)

October 7 (Wed), 16:30-18:30 (CEST)

October 9 (Fri), 6:30-18:30 (CEST)

October 14 (Wed), 16:30-18:30 (CEST)

Materials

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

- ▶ <https://github.com/christiandudel/EDSD2020>
- ▶ <https://osf.io/c6jru/>

Slack and (voluntary) exercises

- ▶ For each session, there will be some voluntary exercises to solve
- ▶ If you have any questions regarding the exercises you can post them on Slack
- ▶ Solutions will be available online (GitHub/OSF)
- ▶ These voluntary exercises have to be distinguished from the (mandatory!) assignment

Assignment: Overview

- ▶ One mandatory assignment
- ▶ Assignment handed out on October 7
- ▶ Deadline: October 13
- ▶ Assignment will consist of several exercises: “Do this and that with R”, “Solve this problem with R”, etc.

Assignment: Your solutions

- ▶ You submit R code as solutions (via email)
- ▶ R code should be commented, explaining what is happening
- ▶ Code should work “out-of-the-box” without errors

Assignment: Deadline

- ▶ Deadline assignment: October 13, 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 “out-of-the-box”* ... 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
- ▶ You can work in groups
- ▶ Pass/fail

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 16,000 packages available
- ▶ Commonly used in both science and industry
- ▶ Tons of R-related materials: Books, journals, conferences, forums, tutorials. . .
- ▶ Many methods are already implemented in R

Why use RStudio?

- ▶ R is the programming language
- ▶ RStudio is a tool to use R more efficiently
- ▶ 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://cran.r-project.org/>
- ▶ R-Studio: <https://www.rstudio.com/>