Data Analysis in R

General data

Course code:	B19GMC05E
ECTS credits:	3
Type of the course:	General elective
Semester:	Autumn, Semester 3 and 5
Course restrictions:	*Completed Probability and Statistics course is recommended, any programming experience is useful but not a requirement*
Course leader (with availabilities):	**Dániel Kehl, Dr.** kehld@ktk.pte.hu +36 72 501 599/ 63144
Further lecturer(s) (with availabilities):	-

1. Description and aims

The module aims to introduce an open-source script based statistical computing and visualisation environment. Data analysts need tools that operate in many different environments and are capable of handling large datasets. One of these tools is the R project. The course aims to support students to develop basic programming skills. Besides learning basic syntax of the language, the course focuses on classical and even interactive visualisation tools.

2. Intended Learning Outcomes (ILOs)

Upon the successful completion of this course, students should be able to:

1. competently read and write basic data analysis code in R, *(PILO3 and 7)*,
2. write dynamic reports in R *(PILO6)*,
3. confidently handle basic building blocks and syntax of R *(PILO3)*,
4. analyse data at an advanced level *(PILO4)*,
5. select the appropriate technique and method for the solution of such problems *(PILO3)*,
6. articulate the benefits of programming *(PILO2)*.

(The remarks in brackets express each CILO's connection to the Program Intended Learning Outcomes (PILOs).)

3. Content, schedule

1. R and RStudio, getting started (RC2E – Ch 1, R4DS – Ch 1-2)
2. Basic elements of the R language, scalars, vectors, strings, factors, data frames, lists (RC2E – Ch 2, R4DS – Ch 4, 10)
3. Navigating in R (RC2E – Ch 3, R4DS – Ch 6)
4. Reading data into R (RC2E – Ch 4, R4DS – Ch 11)
5. Describing and exploring data (RC2E – Ch 9, R4DS – Ch 7)
6. Visualisation (RC2E – Ch 10, R4DS – Ch 3)
7. Working with factors, strings and dates (RC2E – Ch 7, R4DS – Ch 14-16)
8. Writing custom functions (RC2E – Ch 15, R4DS – Ch 19)

9. Linear models in R (RC2E – Ch 11, R4DS – Ch 23-24)
10. Simple time series analysis (RC2E – Ch 14)
11. Communicate results, R Markdown (RC2E – Ch 16, R4DS – Ch 27)

4. Learning and teaching strategy, methodology

Principal teaching methodologies: pre-reading of book chapter, student presentation, in-class discussion, hands-on in class programming

The essential learning materials are two very widely used handbooks available online for free. Students are expected to read chapters before class and try to answer related questions (CILO 1, 3). In class we discuss problems from the book especially those that were problematic and solve business problems with R (CILO 5, 6). More advanced data (compared to the courses Probability and Statistics and Business Statistics) analytic techniques are discussed (CILO 4). By the end of the semester students set up their own data analysis projects and present their results (CILO 1, 2, 3, 6).

5. Assessment

Formative assessment elements: Oral feedback on in-class activities, discussion of the program code of practical exercises solved at home and during the classes, QA session before midterm test, and drop-in office hours.

Summative assessment elements:

Individual Assessment	100%	**Group Assessment**	0%

Name of the element	**Weight**	**Type**	**Details**	**Retake opportunity**	**Req.***	**Related CILOs**
Midterm	50%	Individual written exam	A written exam in R with various data analysis problems	one retake opportunity	yes	1, 2, 4, 5
Project	50%	Individual oral presentation	Students bring their own data analysis problems and solve them in R	one retake opportunity	yes	3, 6

** Req.: Completion of the element is required to pass the course, irrespective of the performance in other elements.*

6. Learning materials

- Essential

LONG, J. D., TEETOR, P. (2019): R Cookbook, 2nd Edition, https://rc2e.com/
WICKHAM, H., GROLEMUND, G. (2021): R for Data Science, http://r4ds.had.co.nz/

- Recommended

BERENSON, M. L., LEVINE, D., SZABAT, K., AND STEPHAN, D. (2020): Basic Business Statistics: Global Edition 14/e, ISBN: 978-0134684840
DE BROUWER, P. J. S. (2020): The Big R-Book: From Data Science to Learning Machines and Big Data, ISBN: 978-1119632771

7. Further information

International aspects embedded with the course
Not relevant to this course

Ethics, Responsibility & Sustainability (ERS) aspects embedded with the course
Not relevant to this course
Connections to the world of practice of the course
Course leader shares own consulting, data analysis and research experiences.