



## DEPARTMENT OF MEDICINE, HUDDINGE

### **H7F6003, Intermediate R –Data Science and Visualization Techniques Beyond Base R, 3 credits (hec)**

R fortsättning - Datavetenskap och visualiseringstekniker bortom grundkunskap i R, 3 högskolepoäng

*Third-cycle level / Forskarnivå*

---

#### **Approval**

This syllabus was approved by the The Committee for Doctoral Education on 2024-02-13, and is valid from autumn semester 2024.

#### *Responsible department*

Department of Medicine, Huddinge, Faculty of Medicine

#### *Contributing department/s*

Department of Laboratory Medicine

#### **Prerequisite courses, or equivalent**

Knowledge and/or skills in basic R, equivalent to course "H1F5300: Get started with R – Programming Basics, Data Analysis and Visualisation", "H5F2953: Statistics with R - from Data to Publication Figure" or corresponding courses.

#### **Purpose & Intended learning outcomes**

The course is practical, where the aim is to:

Teach students how to work with more advanced concepts of the programming environment R and Rstudio, such as functional programming, simple algorithm development, advanced data manipulation and visual presentation, Rmarkdown and the Tidyverse.

After attending the course, the student should be able to:

- Data wrangle, tidy up messy data and structure data for analysis (e.g., convert from wide to long format)
- Identify situations suitable to the use of functions, loops and conditionals

- Construct their own algorithms, incorporating self-created functions, loops and conditionals
- Work with .REnv, for version control and authentication
- Use RMarkdown to create easier markup and navigation, as well as create PDF files and websites
- Use version control for collaboration
- Create their own packages from scratch, as well as “source” other R files from another script
- Advanced visualisation with ggplot and an introduction to interactive presentations of data
- Evaluate the efficiency of code, selecting the best option of, for example functions, to solve a certain problem

## Course content

The intermediate R course covers more complex topics and builds upon the foundation established in basic R courses. The course contains:

- Recommended coding conventions and best coding practices
- Functional programming, including advanced techniques for writing functions in R, such as closures, anonymous functions, and higher-order functions
- Object-Oriented Programming: An introduction to the basics of object-oriented programming in R, including classes, objects, and inheritance
- Advanced data manipulation, including topics such as regular expressions, string manipulation, and the use of the tidyverse packages for data cleaning and manipulation
- Advanced data visualization, including the use of advanced visualization techniques, such as interactive visualizations using shiny, and visualizing complex data using the ggplot2 package.

The course will also emphasize the use of best practices for reproducibility and collaboration. Introducing the concepts of writing modular and reusable code, using version control with Git, and using RMarkdown for reproducible reporting.

## Forms of teaching and learning

Distance learning with online ZOOM lectures and labs. Videos, quizzes and tasks in Canvas. Group and individual exercises and assignments. Reviewing other students’ code and interaction with other students. Individual project work. Each session will be structured around a specific concept, for example writing simple algorithms (like binary search or greedy algorithms). Each session will be comprised of two days. The first day the concept is introduced with a lecture, then quizzes and tasks, to gradually increase the autonomy of the students to use the specific concept. The second day will be a lab where students are supposed to use the concept to perform a specific action, receiving formative feedback from one peer and one teacher. The last day of every second week will be a larger either individual or group exercise, where the student is required to combine introduced concepts into a whole. This exercise will be reviewed by a fellow student who will have the opportunity to comment on ways to improve the work.

### *Language of instruction*

The course is given in English

## **Grading scale**

Pass (G) /Fail (U)

## **Compulsory components & forms of assessment**

Participation during lectures, quizzes, labs, individual and group projects and reviews of other students' projects. Absence from lectures can be compensated by finishing an additional task, quizzes can be finished at any time during the course. The activity of the labs can be finished at any time during the course, but there will only be support from teachers during the specific lab occasion. The remaining elements can not be compensated during the course.

Students are graded based on two exams.

## **Course literature**

Recommended e-book: "Advanced R" by Hadley Wickham (2nd online edition)