Course Description

This course consists of two parts. The first part teaches fundamental programming concepts, assuming students have no prior knowledge of the field. The second part builds upon the first and teaches programming skills specific to data analysis.

part I - class 1

- what is a computer - hardware and the operating system
- what is programming (and some background about programming languages, compilers/interpreters, and R)
- installing R, RStudio, and libs

part I - class 2
• setting a working directory, writing code (initial experience)
• input and evaluation - variables, printing, auto-printing, syntax

part I - class 3
• objects and attributes
• data types - vectors (and scalars), matrices, lists
• data types - factors
• missing values

part I - class 4
• data frames
• names in R
• reading and writing data
• reading large tables
• reading various textual formats
• connections

part I - class 5
• subsetting - basics
• subsetting lists
• subsetting matrices

part I - class 6
• partial matching
• subsetting - removing missing values
• vectorized operations and logical operators

part I - class 7
• control structures (intro)
• control structures - if else
• control structures - for-loop
• control structures - while-loops

part I - class 8
• sets and logical operations
• control structures - repeat loop

part I - class 9
• functions - intro
• functions in R
• scoping rules - symbol binding

part I - class 10
• scoping rules - scoping revisited
• scoping rules - optimization example
• coding standards
• dates and times

part I - class 11
• special loop functions in R: lapply, sapply
• special loop functions: apply
• special loop functions: mapply

part I - class 12
• special loop functions: tapply
• the split function
• debugging - diagnosing a problem

part I - class 13
• debugging - tools
• the str and summary functions
• code profiling
• classes - a brief info
• simulation - generating random numbers (if time permits)

part II - class 1
• the data science process

part II - class 2
• data visualization

part II - class 3
• data transformation with dplyr

part II - class 4
• exploratory data analysis (EDA)

part II - class 5
• communicating with others - knitr
• R projects

part II - class 6
• tibbles with tibble
• importing data with readr

part II - class 7
• tidy data with tidyr
• relational data with dplyr

part II - class 8
Course Goals

The knowledge and experience acquired in the course build basic general programming skills and, in addition, train students in accessing data from various sources, cleaning data, performing various transformations, and representing data graphically. Students also learn how to perform an initial evaluation of data and preliminary data analysis thereby developing initial insights regarding the data at hand.

Grading

Final exam: 55%; home assignments 45%.

- One must pass the exam (i.e. obtain a minimum grade of 60) to pass the course.
- One must submit a minimum number of home assignments to be eligible to take the exam (further details regarding assignment requirements and grading are available on the course's moodle page.

NOTE: Class attendance is mandatory. Exemptions will be given only in accordance with Reichman University's bylaws. Students who do not comply will not be allowed to take the exam.

Learning Outcomes

Upon successfully completing the course, students will be able to:

- write code for reading data from various sources,
- clean data,
- perform various data transformations,
- represent data graphically and perform basic-level EDA.

Students will also be able to perform basic evaluation and preliminary analysis of data, thus developing initial insights regarding the data at hand.

---

📖 **Lecturer Office Hours**

Officially: Thursdays 14:30 to 16:00 (by appointment).

Unofficially: If we're on campus, I'm at my office nearly daily and am happy to meet any time, as long as I don't have prior commitments. If we're off-campus, I can be reached by mail. Zoom meetings will be scheduled for issues that can be better handled that way.

---

⏰ **Teaching Assistant**

By appointment

---

📢 **Reading List**

No additional reading material is required for this course.

The second part of the course follows *R for Data Science* by Hadley Wickam, which is a recommended resource for the course and beyond.