

COURSE SYLLABUS

Programming Tools

Instructors:

János Károly DIVÉNYI
Economics Department
Central European University
Fall 2015
Course level: MA, first year
1 Credits (2 ECTS Credits)

Course Description

The course is aimed at students having minimal or no programming experience to teach them basic programming principles and make them familiar with R which they can use for data analysis.

Course Prerequisites

No fear of computers.

Learning Outcomes

By the end of this course, students will be able to:

- use R for analyzing data: solve basic tasks of data exploration, data manipulation, and data visualization
- organize their work in a transparent and replicable way
- realize the importance of reading documentations and other sources of getting help so that they can learn new tools by themselves

Course Requirements

The course is Pass/Fail. There is going to be two problem sets consisting of several tasks requiring the use of the tools and principles learned during the class. Each of the problem sets has to be passed to get a final Pass at the end of the course. The problem sets are going to be posted on 25th September and 1st October, and are due at 24:00 on 29th September and 5th October, respectively. Late submissions are equivalent to Fail.

For self-assessment and recalling important concepts from previous classes, mini-quizzes will be offered at the beginning of classes.

Course Schedule

23 September Motivation: why code? Introduction to R

- be productive, comprehensible to others, able to keep track
- Rstudio
- data structures: atomic vector, (matrix, array), list, data frame
- R packages
- getting help: documentation, Stack Overflow, Google

24 September Programming in R. Exploratory data analysis I.

- statements: loops, conditionals
- functions
- R package: *dplyr*
- Filter, select, arrange, mutate, summarise

25 September Exploratory data analysis II. Data manipulation

- Chaining
- Mutate_each, summarise_each, sample
- R package: *tidyr*
- Gather, spread

30 September Data visualization

- Grammar of graphics
- R package: *ggplot2*
- Scatterplot, density, grouping, fitting regression line
- Detour to interactive visualization: R package *ggvis*

1 October Replication guidelines. How a data analysis project should be performed?

- directory organization, usage of relative paths
- documentation
- automation
- helping others to replicate your project