Copenhagen Summer University

| 📅 | 19 – 23 AUGUST 2019, 9:00 – 16:30 | 👛 | EUR 2,680 / DKK 19,900 EXCL. DANISH VAT |

Copenhagen Summer University

DATA SCIENCE WITH RSTUDIO

Data is everywhere but does not generate value by itself. So how do you get more value out of your data? And how do you learn to become an efficient and well-structured data analyst? By learning about RStudio.

RStudio is an extremely versatile modern data science environment, which utilizes a range of statistical methods to carry out most tasks via the unprecedented back catalogue of packages for the R programming language. R is, moreover, suitable for automatizing repetitive tasks and for ensuring accuracy and reproducibility of your analyses. Finally, R and RStudio offer seamless integration with other programming languages such as Python and C++, with databases and with large-scale analytics tools such as Apache Spark.

WHAT YOU WILL LEARN

You will learn to build a complete data analysis pipeline in RStudio. This includes learning R programming techniques for:

– Data import from multiple sources

– Data manipulation and visualization

– Automatic and interactive report generation

– Database integration

– Large-scale computations

In addition to the technical programming skills, you will also learn to apply a conceptual framework to data analysis, where all the steps of a data analysis are automatized via a programmatic pipeline.

COURSE CONTENT

The course is based on RStudio and a collection of modern R packages. The focus will be on learning to exploit the full potential of these tools, which can serve as an infrastructure for almost any conceivable data analysis in R.

Core elements:

– RStudio: An integrated development environment for R, which supports interactive data analysis, building of data analysis pipelines, and R software development

– Tidyverse: A framework and collection of R packages centered on the concept of tidy data

– Visualization: High-quality figures are created from structured specifications using the R package ggplot2

– Reproducible analysis: Automatic and reproducible reports are written and generated using R Markdown

– Interactive communication: Reactive web applications for interactive presentations of data and analyses are written using Shiny

– Cloud computing: How to leverage cloud resources for big data and large-scale analyses

PARTICIPANTS

R is a programming language, and the course takes a programmatic approach to data analysis. Therefore, in order to benefit fully from the course, participants should be interested in and willing to program. The course requires only modest statistical and mathematical knowledge, but participants should at least know about mean, variance and simple linear regression.

The course is for:

– People with some experience in SAS, Matlab or Python programming for data analysis, but with no or limited experience with R

– People with general programming and/or database experience, but limited or no experience with data analysis and data modeling

– People with some R experience and an interest in learning RStudio, R Markdown, and Tidyverse

R (www.r-project.org) and RStudio (www.rstudio.com) are open source and available free of change. Participants are expected to bring a laptop with these programs installed.

COURSE DIRECTORS

Niels Richard Hansen, Professor, Department of Mathematical Sciences, University of Copenhagen

Anders Tolver, Associate Professor, Department of Mathematical Sciences, University of Copenhagen

TEACHING MATERIAL

Participants will receive a copy of the book R for Data Science (2016) by Garrett Grolemund and Hadley Wickham.

COURSE FEE

EUR 2,680 / DKK 19,900 excl. Danish VAT. Fee includes teaching, course materials and all meals during the course.