

Dr. Gerhard Schöfl

Introductory Data Analysis with R

24–25 April 2014, 9.00–17.00 h

MMZ, Carl-Zeiss-Str. 8, SR 205

R is a software environment for statistical computing, graphics, and data analysis. It has become hugely popular among data scientists due to its open source nature and its extensibility via packages. In contrast to other popular statistical software, R is a fully fledged high-level programming language. While somewhat daunting for the beginner, the long-term benefits of having a programming language at your disposal when dealing with data are immense. Especially once your data-sets outgrow the average Excel spreadsheet.

In this course we will cover the basics of computational data analysis with R. You will learn to perform simple programming tasks, how to read data into R, and we will talk about how data is represented in R. We will cover the basics of data munging (the often neglected skill of cleaning, pruning, and generally bending data sets into a shape amenable for further analysis) and data visualisation. A guided tour through a real world analysis, starting from obtaining and cleaning the data, exploratory analysis, statistical inference, and graphical report, will complete the course.

A prerequisite for this course is an affinity for working with a computer. Prior exposure to R is a benefit but not necessary.