Programming with R for Reproducible Research

Martin Mächler, Seminar für Statistik, ETH Zurich

Before first lecture, Feb.18, 2014

1 Course *Programming with R for Reproducible Research*

- Prerequisites:
  - Both parts of "Using R for Data Analysis"
  - Laptop with R ($\geq$ 3.0.1) and RStudio / StatET / ESS, or similar "R IDE" installed
  - One semester of statistics
- Duration: 2 hours $\times$ 7 weeks (= $\frac{1}{2}$ semester),
  corresponds to "1 G" of a full semester
- Credits: 1 ECTS
- Exam: "Written", respectively at computer, at the end of the teaching block,
  April 8 or 15.
- Lecture Notes: Written in "Reproducible Research" (Sweave) Style;
- Textbook: Used very loosely: *The Art of R Programming* by Norman Matloff
  - Polybuchhandlung, CHF 45.-
- **Many** online resources. A very sophisticated (and hence not 100% correct) one:
  Advanced R by Hadley Wickham

2 Outline - Topics

- Programming with R
  - R Data Types, notably list()s, lapply, etc
    - quick review (of prerequisites)
    - Slides from "Using R part 2"
  - John Chambers: To understand computations in R, note that
    - Everything that exists is an object.
    - Everything that happens is a function call.
  - *10.0 times 0.1 is hardly ever 1.0*
    ("The elements of Programming Style", Kernighan and Plauger, 1974): – computer numerics and R FAQ 7.31
  - First steps with parallel execution: package parallel,
    even on your notebook
  - Object Orientation in R: S3, S4, Reference classes

  - Better understanding of packages and their namespaces (see below)
  - R Functions as "Closures": Example splinefun
  - Environments
  - Expressions (substitute(), quote(), eval() etc)
- Reproducible Research and Data Analysis
  - This document is written in Emacs "*Org Mode*"
    * show source, options
    * one can do R and C and more with "/Org Babel/", but that is Emacs only.
    * We will use and learn a bit: *Sweave* and *knitr*.
  - Why reproducibility is very important
    * CRAN task view "Reproducible reasearch"
  - Reproducible Data Analysis: R code and Report
  - Reproducible Research: Theory, Illustrations, Simulation
  - Sweave and knitr – implementation and examples
- Towards Writing your own R Package
  - Understanding Namespaces
  - Design, Testing, Documentation