Programming with R for Reproducible Research

Martin Mächler, Seminar für Statistik, ETH Zurich

First version: Feb. 2014; update: Feb. 23, 2016

1 Course *Programming with R for Reproducible Research*

1. Prerequisites:

   - Both parts of "Using R for Data Analysis"
   - Laptop with R ($\geq$ 3.2.1) and RStudio / StatET / ESS, or similar "R IDE" installed
   - One semester of statistics

2. Duration: 2 hours $\times$ 7 weeks ($= \frac{1}{2}$ semester), corresponds to "1 G" of a full semester

3. Credits: 1 ECTS

4. Exam: "Written", respectively at computer, at the end of the teaching block, April 19 ??

5. Lecture Notes: Written in "Reproducible Research" (Sweave) Style;

6. Textbook: Used very loosely: *The Art of R Programming* by Norman Matloff

   (a) Polybuchhandlung, CHF 45.-

7. **Many** online resources.

   (a) A very sophisticated (and hence not 100% correct) one: Advanced R by Hadley Wickham

2 Outline - Topics

1. Programming with Rd

   (a) R Data Types, notably list()s, lapply, etc
      - quick review (of prerequisites)
      - Slides from "Using R part 2"

   (b) John Chambers: To understand computations in R, note that
      - Everything that exists is an object.
      - Everything that happens is a function call.

   (c) *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

   (d) First steps with parallel execution: package parallel, even on your notebook

   (e) Object Orientation in R: S3, S4, Reference classes

   (f) Better understanding of packages and their namespaces (see below)

   (g) R Functions as "Closures": Example splinefun

   (h) Environments

   (i) Expressions ($_{\text{substitute}}$(), quote(), eval()_ etc)

2. Reproducible Research and Data Analysis

   (a) 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*.

   (b) Why reproducibility is very important
- CRAN task view "Reproducible reasearch"

   (c) Reproducible Data Analysis: R code and Report

   (d) Reproducible Research: Theory, Illustrations, Simulation

   (e) Sweave and knitr – implementation and examples

3. Towards Writing your own R Package

   (a) Understanding Namespaces

   (b) Design, Testing, Documentation