R Programming

Coursera Course by John Hopkins University

INSTRUCTORS: Dr. Jeff Leek, Dr. Roger D. Peng, Dr. Brian Caffo

---

Overview of R, R data types and objects, reading and writing data

Installing R & RStuido

- This was covered in the previous course.

Swirl

- swirl teaches you R programming and data science interactively, at your own pace, and right in the R console.
- Start swirl
  - install the package "swirl" if you haven't yet
  - Everytime you want to run swirl execute:
    - library("swirl")
    - swirl()
  - You'll then be prompted to install a course
  - Help page for swirl

History of S and R programming

- What is S?
  - R is a dialect of S
  - S was developed by John Chambers and others at Bell Labs
  - Initiated in 1976 as an internal statistical analysis environment, implemented as Fortran libraries
    - Early versions did not contain functions for statistical modeling
  - Version 3 was released in 1988, which was rewritten in C and began to resemble the system that we have today.
  - Version 4 was released in 1998 and is the version we use today.

    * This version is documented in *Programming with Data* by John Chambers (the green book)
  - Insightful sells its implementation of the S language under the name *S-PLUS*, which includes a number of fancy features, mostly GUIs.
  - S won the Association for Computing Machinery's Software System Award in '98

  - (More about S)[https://web.archive.org/web/20181014111802/ect.bell-labs.com/sl/S/]
- What is R?
  - R was developed by Ross Ihaka and Robert Gentleman, they documented thier experience in a (1996 JCGS paper)[https://amstat.tandfonline.com/doi/abs/10.1080/10618600.1996.10474713].

  - In 1995, R become free software after Martin Machler convinced Ross & Robert to use the GNU (General Public License)

  - Versions
    * R version 1.0.0 was released in 2000

    * R version 3.0.2 is released in Dec. 2013

  - Syntax is similar to S, making it easy for S-PLUS users to switch over

  - Runs on almost any standard computing platform/OS (even on the PS3)

  - Frequent releases; active development and communities

  - Funtionality is divided into modular packages as to keep it "lean"

  - It's free!

  - What is free about Free Software?
    * Freedom 0: freedom to run the program, for any purpose

    * Freedom 1: freedom to study how the program works, and adapt it to one's needs. Which implies access to the source code
    * Freedom 2: freedom to redistribute copies
    * Freedom 3: freedom to improve the program, and release your improvements to the public, or to sell them.
    * These are outlined by the (Free Software Foundation)[https://www.fsf.org/]
- Drawbacks of R
  - Essentially based on 40 year old technology,the original S language
  - Little build support for dynamic or 3D graphics. Although there are packages for such
  - Functionality is based on consumer demand and use contributions, if a feature is not present you'll have to build it.

  - Objects that are manipulated in R have to be stored in the physical memory of the computer, as such if an object is bigger than the memory you'll be unable to load it into memory

  - Not ideal for all possible situations (but this is a drawback of all software packages)

*Design of the R System
+ "base" R system that can be downloaded from (CRAN)[http://cran.r-project.org] (krey-an).
- contains the packages: **utils, stats, datasets, graphics, grDevices, grid, methods, tools, parallel, compiler, splines, tcltk, stats4**.
- "Recommended" packages: **boot, class, cluster, codetools, foreign, KernSmooth, lattice, mgcv, nlme, rpart, survival, MASS, spatial, nnet, Matrix**.

+ Packages are available all around the web, but packages on CRAN have to meet a certain level of quality.

- Some Useful Books on S/R
  - Chambers (2008). *Software for Data Analysis*, Springer.
  - Chambers (1998). *Programming with Data*, Springer.
  - Venables & Ripley (2002). *Modern Applied Statistics with S*, Springer.
  - Venables & Ripley (2000). *S Programming*, Springer.
  - Pinheiro & Bates (2000). *Mixed-Effects Models in S and S-Plus*, Springer.
  - Murrell (2005). *R Graphics*, Chapman & Hall/CRC Press.
  - (Additional Books)[http://www.r-project.org/doc/bib/R-books.html]

Differences between atomic data types

Basic Arithmetic operations

Subset R objects using the "[", "[[", and "$" operators and logical vectors

The explicit coercion feature of R

Removing missing (NA) values from a vector

Control structures, functions, scoping rules, dates and times

Loop functions, debugging tools

Simulation, code profiling