

An Introduction to R

Gerardo Ferrara

Master in Economics and Complexity, Collegio Carlo Alberto

Fall 2013

What is R?

R is an integrated suite of software facilities for data manipulation, calculation and graphical display. Among other things it has:

- Effective data handling and storage facilities.
- Large, checked, integrated collection of tools for data analysis.
- Extensive and powerful graphics abilities, that are tightly linked with its analytic abilities.
- A well developed, simple and effective programming language which includes conditionals, loops, user defined recursive functions and input and output facilities.

Features of R

- Runs on almost any standard computing platform/OS (even on the PlayStation 3).
- Frequent releases (annual + bugfix releases); active development.
- Graphics capabilities very sophisticated and better than most stat packages.
- Useful for interactive work, but contains a powerful programming language for developing new tools.
- Very active and vibrant user community; R-help and R-devel mailing lists and Stack Overflow.

It is free!

With free software, you are granted:

- The freedom to run the program, for any purpose.
- The freedom to study how the program works, and adapt it to your needs.
- The freedom to improve the program, and release your improvements to the public, so that the whole community benefits.

Design of the R System

R functionality is divided into a number of packages.

- The *base* R system contains, among other things, the base package which is required to run R and contains the most fundamental functions.
- The other packages contained in the *base* system include: **utils**, **stats**, **datasets**, **graphics**, **grDevices**, **grid**, **methods**, **tools**, **parallel**, **compiler**, **splines**, **tcltk**, **stats4**.
- There are also Recommend packages: **boot**, **class**, **cluster**, **codetools**, **foreign**, **KernSmooth**, **lattice**, **mgcv**, **nlme**, **rpart**, **survival**, **MASS**, **spatial**, **nnet**, **Matrix**.
- There are 4,728 packages on CRAN (as of Sept 6, 2013) that have been developed by users and programmers around the world.

Why not just use a spreadsheet in Excel?

- Excel is not efficient when we need to analyze large datasets.
- Excel has limited built-in statistical tools.
- Excel has limited data manipulation capabilities.

Social Network Analysis with R

R is a useful tool for Social Network Analysis with many advantages over traditional software packages. With a little coding and patience, one can produce analyses and visualizations that better suit the problem at hand with this singular platform, rather than learning GUI application after application. Among other things it has:

- Packages: *igraph*, *sna*, *network*, *statnet*.
- Centrality measures: *degree*, *betweenness*, *closeness*, *transitivity*.
- Clusters: *clusters*, *no.clusters*.
- Cliques: *cliques*, *largest.cliques*, *maximal.cliques*, *clique.number*.
- Community detection: *fastgreedy.community*, *springlass.community*.

Some Useful Books on R

- Knell, J. R., and Braun, J., Introductory R: A Beginners Guide to Data Visualisation and Analysis using R, Springer, 2013.
- Maindonald, J., and Braun, J., Data Analysis And Graphics Using R, Cambridge University Press, 2007.
- Verzani, J., Using R for introductory statistics, Chapman & Hall/CRC, Boca Raton, FL, 2005.
- A longer list of books is at <http://www.r-project.org/doc/bib/R-books.html>