

# Introduction to R

Maarten van Iterson

Department of Medical Statistics and Bioinformatics, Section Molecular  
Epidemiology, Leiden University Medical Center

## Statistical programming language

- A language for data analysis and graphics
- Ross Ihaka and Robert Gentleman (1993) based on the statistical programming language S (Chambers, 1976)
- currently maintained by a large groups of primarily statisticians all over the world

## Open source

- supports various operating systems
- command-line and many graphical user interface, i.e. RStudio
- easily extendable using packages available at CRAN
- computationally-intensive tasks can be written in C, C++ and Fortran code
- large and active community; R journal, Annual conference, specialist mailing list, etc.

## Many application domains

- Econometrics, Genetics, Pharmacokinetics, SocialSciences, Bioinformatics, etc (CRAN Task View)
- can handle almost any data format: .xlsx, .sav, html, xml, images, binary data formats, connect to databases, etc.
- can generate figures in any format: png, jpg, pdf, etc.
- can generate reports in any format: pdf, markdown, html, docx, etc.
- easy parallelization and efficient reading of large data

## BioConductor

- repository specific for bioinformatic analysis
- expression, DNA methylation, copy number, proteomics, metabolomics, genetics, etc.
- not only methods but annotation and data as well
- > 1000 interrelated packages
- high quality software (reviewed), daily build system (guarantees software works) and obligatory documentation
- mailing list, example workflows, course material, etc.

## Some disadvantages

- syntax not always intuitive
- dynamic language; continuously changing (often more efficient and additional features)
- biannually new releases (incl. bug fixes) so keep uptodate
- many solutions for the same problem often one robust and most efficient (requires experiences)
- too many functions don't know which to use (just google e.g. `cran r t test`)

## Some nice examples

- r graph gallery
- r web application
- bioconductor workflows

## Learning R

- large amount of online course material
- many books
- R course LUMC
- practical brief introduction