Introduction to R

Maarten van Iterson & Koen Dekkers

Biomedical Data Sciences

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, Social Sciences, Bioinformatics, etc
- 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 that software works) and obligatory documentation
- Mailing list, example worksflows, 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 experience)
- Many functions for problems (use Google to find them)

Some nice examples

- R graph gallery: `http://www.r-graph-gallery.com/`
- R web application: `https://shiny.rstudio.com/gallery/`
- Bioconductor workflows:
  `https://www.bioconductor.org/help/workflows`

Learning R

- Online course material, for example:
  `https://www.datacamp.com/`
- Books
- R course LUMC
- Practical (brief introduction)