

INTRODUCTION TO R

Lucy Sinke - I.j.sinke@lumc.nl

COURSE OVERVIEW

- 2 day course: 2 lectures + 3 practical sessions
- Lectures: Introduction to R and RStudio Cloud
- Practicals

1

- reproducible work
- importing data
- data manipulation

2

- data visualization
- analysis
- writing functions

3

- bioconductor
- genomic ranges
- basic analysis

WHAT IS R?

- Language for statistical analysis and graphics
- Developed by Ross Ihaka and Robert Gentleman
- Inspired by the programming language, S
- Maintained by R Development Core Team
- Large group of primarily statisticians



Source - R. Ihaka

FREE AND OPEN SOURCE

- Compared to other software
- No hidden costs
- Source code can be viewed and modified by anyone
- Actively maintained
- Stable and reliable

PLATFORM SUPPORT

- Multiple operating systems
- Ported to different hardware

COMMUNITY

- Large and active community
- Happy to provide support and information on developments
- Scripting language - write your own functions
- Import other user written functions as packages
- Incredibly versatile and stays up-to-date

INTEGRATIVE

- Widely used - econometrics to bioinformatics
- Individuals extended R to combine with current workflows
- Can read in multiple data types - html, sav, databases
- Implement other languages - C, Python, Java
- Output many file formats - markdown, sav, pdfs

COMPUTATION

- Intensive tasks can be written in lower level languages
- Easy parallelization - meaning faster calculations
- Optimized for vector operations - better than loops

VISUALIZATIONS

- Stunning graphics - especially using ggplot2
- Used by Google, Twitter, and Facebook
- Build interactive web applications with Shiny

DISADVANTAGES

- Not always intuitive syntax
- Steep learning curve - command-line type user interface
- Requires experience - multiple options for any problem
- Continuously changing - need to stay up-to-date
- No official support - sometimes poor documentation

BIOCONDUCTOR

- Software repository - over 1,500 bioinformatics packages
- Analysis of high-throughput genomic data
- Integration of annotations, such as GO
- Obligatory vignettes - high quality documentation
- Daily build system - guarantees software works

RESOURCES

- R graph gallery: <http://www.r-graph-gallery.com/>
- R web application: <https://shiny.rstudio.com/gallery/>
- Bioconductor: <https://bioconductor.org/>
- Online courses: <https://www.datacamp.com/>
- Books
- R course at LUMC: https://barmsijs.lumc.nl/R_course/