# Introduction to R

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

Biomedical Data Sciences

# Course Overview

- Two day course
- One lecture and three half-day practicals
- Basic functionality - data classes and manipulation
- Statistics - data exploration, visualisation, and regression
- Genomic ranges - basic bioinformatics analysis

# What is R?

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

# 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

- Available for various hardware and software
- Multiple operating systems
- Ported to different hardware

# Community and extendability

- 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 almost any data type - 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

# 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/