

INTRODUCTION TO R

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

COURSE OVERVIEW

- 2-day Introduction to R – aim to get everyone at the same level
- Two lectures today: Introduction to R and RStudio Cloud – set up this and Discord
- Three practicals

1

- reproducible work
- importing data
- data manipulation

2

- data visualization
- analysis
- writing functions

3

- bioconductor
- genomic ranges
- basic analysis

WHAT IS R?

- Different from most statistical software – you can't sidestep the blinking cursor
- Developed by Ross Ihaka and Robert Gentleman
- Language for statistical analysis and graphics
- Inspired by the programming language, S
- Maintained by R Development Core Team
- Large group of primarily statisticians



Source - R. Ihaka

THE BENEFITS OF R

- **Free** – especially attractive compared to other software
- No hidden costs
- **Open source** - source code can be viewed and modified by anyone
- Actively maintained
- Stable and reliable
- Instead of relying on a small team of developers
- Bugs and issues can be detected and resolved far quicker

THE BENEFITS OF R

- Compiles and runs on **multiple operating systems**
- Can even port it to different hardware
- Large and active **community**
- Happy to provide support and information
- Scripting language - write your own **functions**
- Novel methods quickly developed into packages by users
- Versatile and stays up-to-date – other programs may never get niche methods

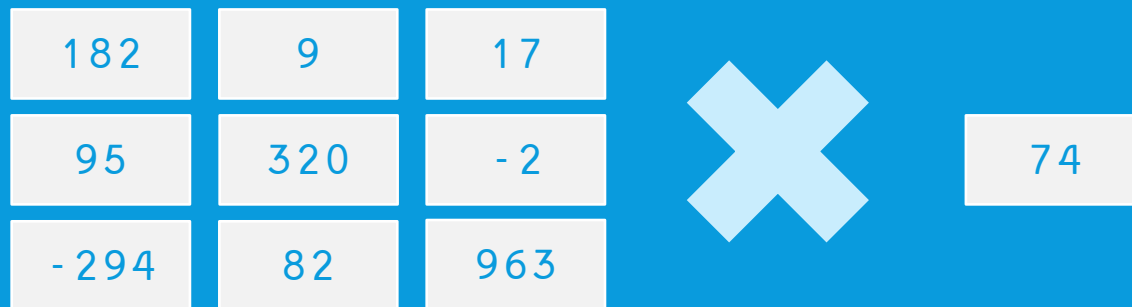


THE BENEFITS OF R

- **Integrative** - plays well with others
- Individuals extended R to nest into their current workflows
- Can read in multiple data types - html, sav, databases, IDAT
- Implement other languages - C, Python, Java – no need to import / export
- Output many file formats - markdown, sav, pdfs

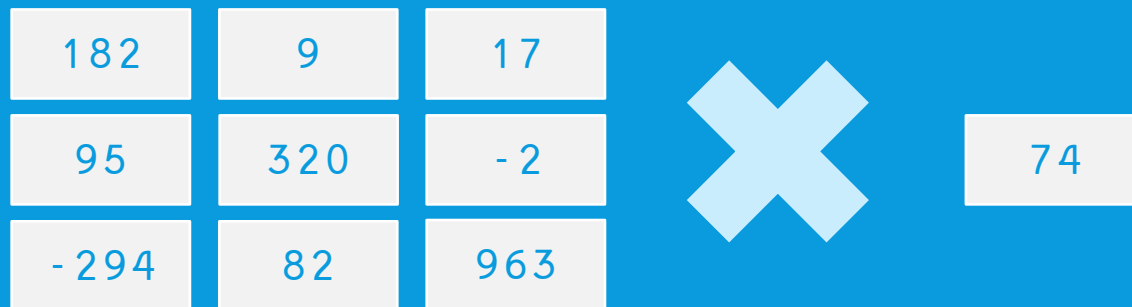
THE BENEFITS OF R

- Easy **parallelization** - meaning faster calculations
- Optimized for **vector** operations - better than scalar languages or loops



THE BENEFITS OF R

- CPU: 1 instruction per cycle, 3 operands per instruction



THE BENEFITS OF R

- CPU: 1 instruction per cycle, 3 operands per instruction
- Scalar language or loop:

9 instructions = 9 cycles

182	×	74
95	×	74
-294	×	74
9	×	74
320	×	74

17	×	74
-2	×	74
963	×	74
82	×	74

THE BENEFITS OF R

- CPU: 1 instruction per cycle, 3 operands per instruction
- Vectorized language:

1 instruction

[182 , 95 , -294 , 9 , 320 , 17 , -2 , 963 , 82]



74



9 operands = 3 cycles

[182 , 95 , -294]



74

[9 , 320 , 17]



74

[182 , 95 , -294]

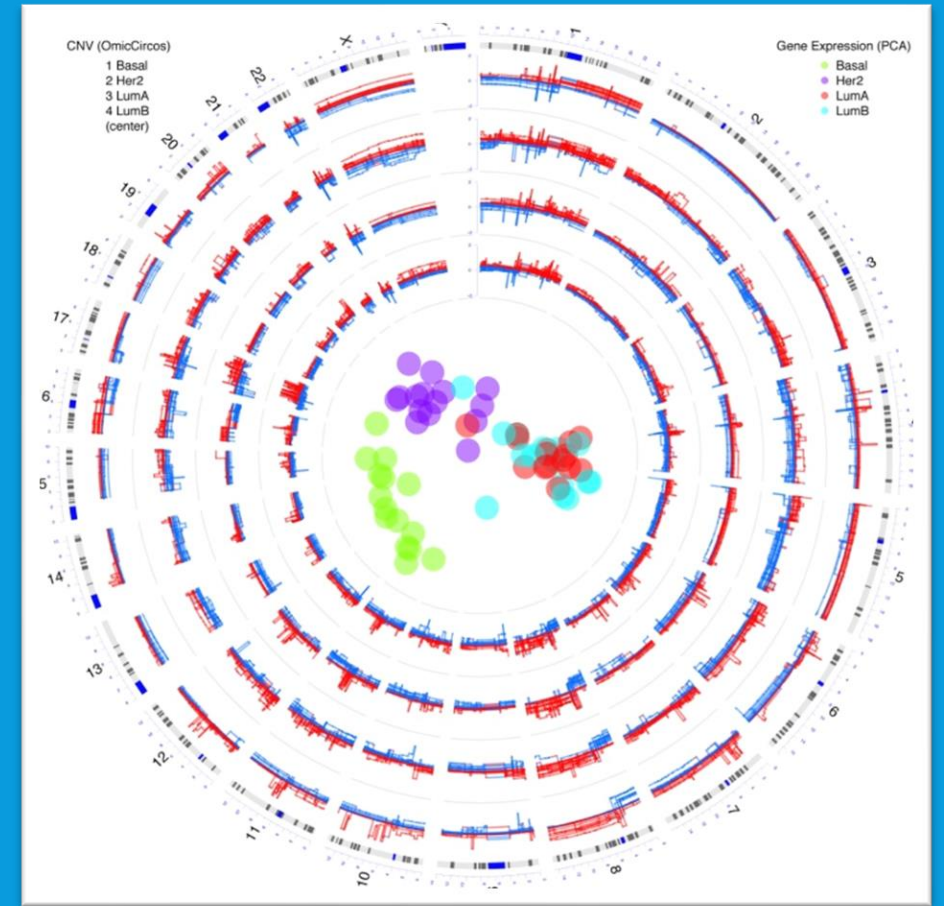


74

3 times as fast!

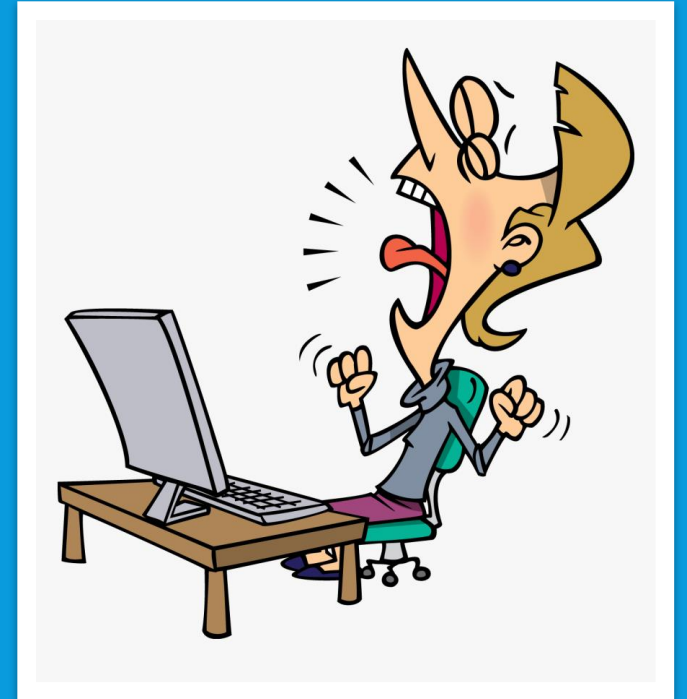
THE BENEFITS OF R

- Stunning graphics – e.g. geo-spatial data
- Bioinformatics packages – e.g. OmicCircos
- Used by Google, Twitter, and Facebook
- Build interactive web applications with Shiny



THE DISADVANTAGES OF R

- Working in R can be quite frustrating
- Syntax – unintuitive and inconsistent
- Steep learning curve - command-line type user interface
- Requires experience - multiple options for any problem
- Continuously changing - need to stay up-to-date
- Nothing to stop packages using the same named functions
- No official support - sometimes poor documentation



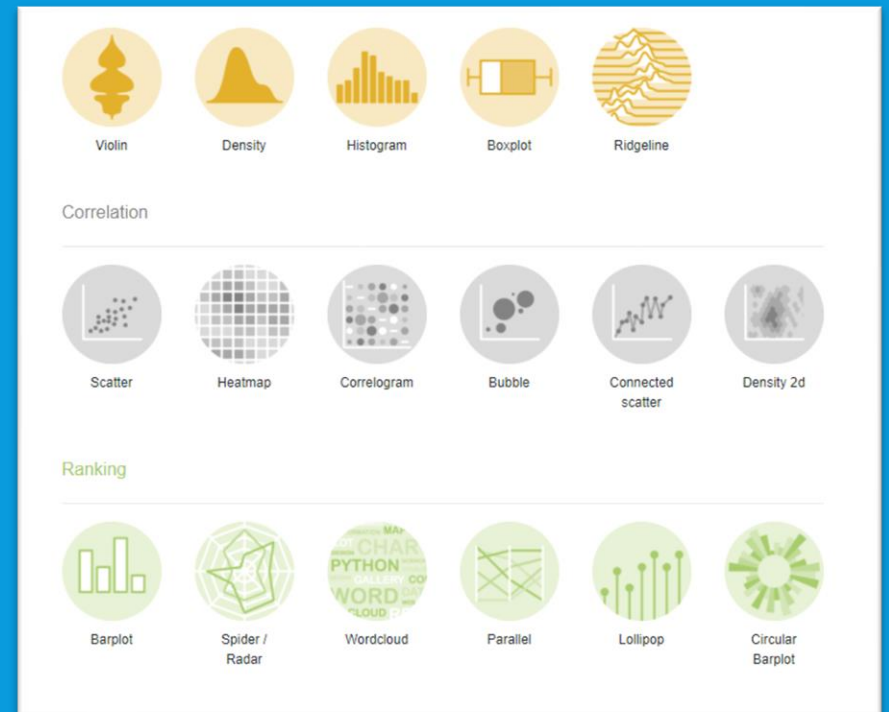
BIOCONDUCTOR



- Software repository - over 1,500 bioinformatics packages
- Analyse, annotate, and visualize high-throughput -omic data
- Have some quality standards
- Obligatory vignettes - high quality documentation
- Daily build system - guarantees software works
- Basics introduced in the 3rd practical

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 – often online using R Bookdown
- R course at LUMC



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 – often online using R Bookdown
- R course at LUMC