Josep Carreras LEUKAEMIA Research Institute

Introduction to R

Angelika Merkel (Head of Bioinformatics Unit IJC)
21/11/2022

IJC Bioinformatics

Workshop overview

1. Introduction
    1. Background to R
    2. Short introduction to RStudio IDE
2. Get Started with R:
   R Programming for Data Science (D. Peng, 2022)

< Short break >

3. Get Started with data analysis in R:
   Example analysis (html link)

Why learn R?

1. For statistical computing and graphics
2. For biological data analysis and data science
3. Free + open source, backed by a large interdisciplinary community
4. R or Phyton for data analysis?

A little bit of history...

1976: Initiation of **S language** (by John Chambers and others at Bell Labs AT&T, New Jersey) for statistical computing

1991: **Creation of R** (R&R) by Ross Ihaka and Robert Gentleman at Department of Statistics, UC Auckland

1993: R goes public, "R: A language for data analysis and statistics" (Ihaka and Gentleman, 1996)

1995: R under Free Software Foundation GNU license, establishment of R-mailing list (ETH Zurich)
establishment of R Foundation, R Comprehensive Archive Network (TU Vienna)

1997: R "core group" established

2001-2008: R News (newsletter of the R-project)

2002: **Bioconductor v1.0 o**pen-source software for bioinformatics

2005: `ggplot2` open-source data visualization package (by Hadley Wickham)

2009: R-forge collaborative development environment released

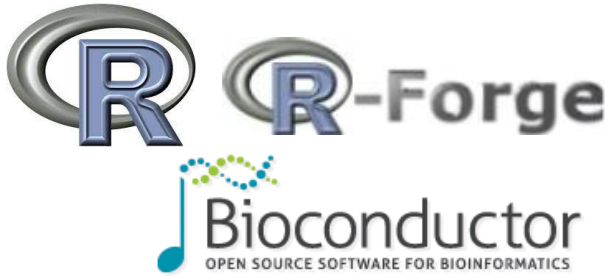
2009: R Journal (super seeds R News)

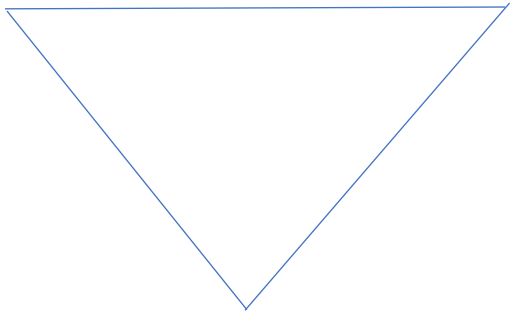
2011: **RStudio IDE v0.92** released; 2016: RStudio IDE v1.1 released

2018: `Tidyverse` package collection for tidy data & data science (by Hadley Wickham)

R - More than just a programming language

Code repositories and collaborative development environments





Integrated development environment (IDE)

Community

R-help -- Main R Mailing List: Primary help

RStudio Integrated Development Environment (IDE)

What is an IDE?

Go to the RStudio browser course

RStudio spaces:

1. Interactive console
2. Source editor
3. Workspace
4. 'Pane' area (Files, plots, package manager, integrated help)

Get started with R

R Programming for Data Science (D. Peng, 2022)

- Chapter 4:
  - Nuts and bolts of R
  - Classes and types of objects
- Chapter 9:
  - Sub-setting (accessing) objects
- Chapter 13:
  - Control structures:
    if-else, for, while, repeat, next, break
- Chapter 14:
  - Functions

How to get help

Inside R and RStudio (integrated help):

```
> ?function()
> function()+F1
```

Community/ web

- Stackoverflow >> R
- R help mailing list
- R-bloggers
- Google is your friend!

Get started with data analysis in R

Example analysis (by A. Merkel with modified parts by D. Peng)

1. Import data
2. Data QC (aka data wrangling or cleaning)
3. Exploratory analysis (incl. base R graphics)
4. Analysis
5. Export results

Further (recommended) topics:

- R graphics
  - ggplot2()
- Data manipulation with R
  - data.table()
  - dplyr (tidyverse)
- Efficient executions in R
  - apply(), sapply(), lapply()
- R for bioinformatics with bioconductor
  - GenomicRanges, Annotation.DB
- R for reproducible research
  - Markdown, github integration, containers (docker/singularity)

Thank you!

IJC Bioinformatics