Josep Carreras LEUKAEMIA Research Institute

Introduction to 

Angelika Merkel (Head of Bioinformatics Unit IJC)
14/11/2023

http://www.carrerasresearch.org/en/units/bioinformatics

Workshop overview

1. Why R and what is R?

2. Introducing RStudio -> POSIT (July 2022)

3. Practical session: Get Started with R (based on R Programming for Data Science (D. Peng, 2022))

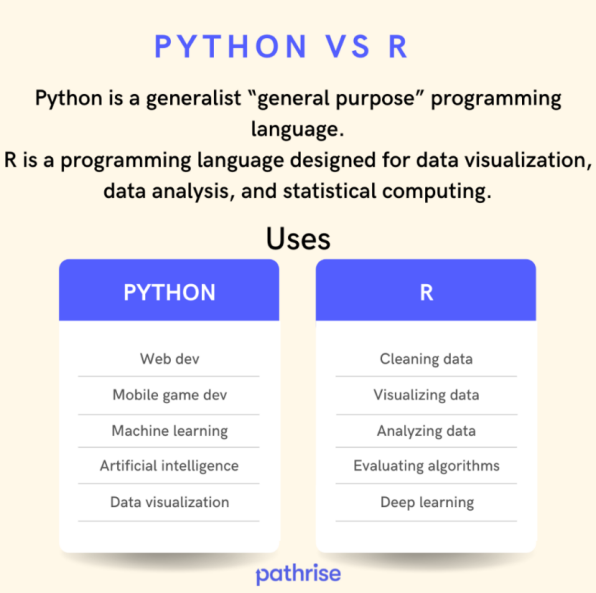
4. Practical session: Data analysis in R - Example analysis

All course material can be found here:
https://ijcbit.github.io/Workshops/

RStudio course server

https://rstudio1.services.carrerasresearch.org/

Why learn R?

1. Statistical computing and graphics
2. Biological data analysis and data science
3. Free + open source, backed by a large interdisciplinary community
4. 

https://www.pathrise.com/guides/python-vs-r-data-science-languages-to-master/

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

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

2005: `ggplot2` 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

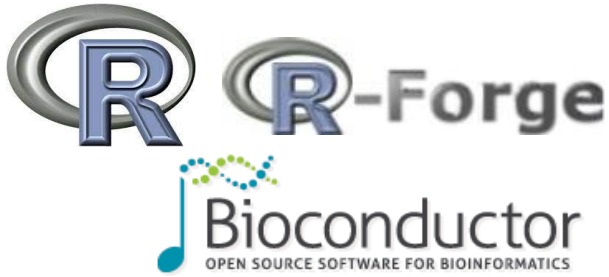
2018: `Tidyverse` package collection for tidy data & data science (by Hadley Wickham)

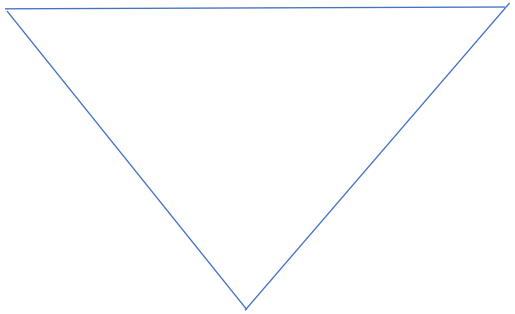
R - More than just data analysis

	Extension	Output formats	Utilities
R script	.R	.csv, png, jpeg, .rds, .RData	Textfiles, images (plots), compressed R objects
R markdown	.rmd	HTML, docx, LaTeX (PDF)	Webpages, documents, notebooks, presentations
Quarto	.qmd	HTML, docx, ppt, LaTeX (PDF)	Webpages, documents, presentations
R sweave	.rnw	LaTeX (PDF)	documents, presentations
R Shiny	App.R, server.R		Interactive web applications

R - More than just a programming language

Code repositories and collaborative development environments





Integrated development environment (IDE)

Community

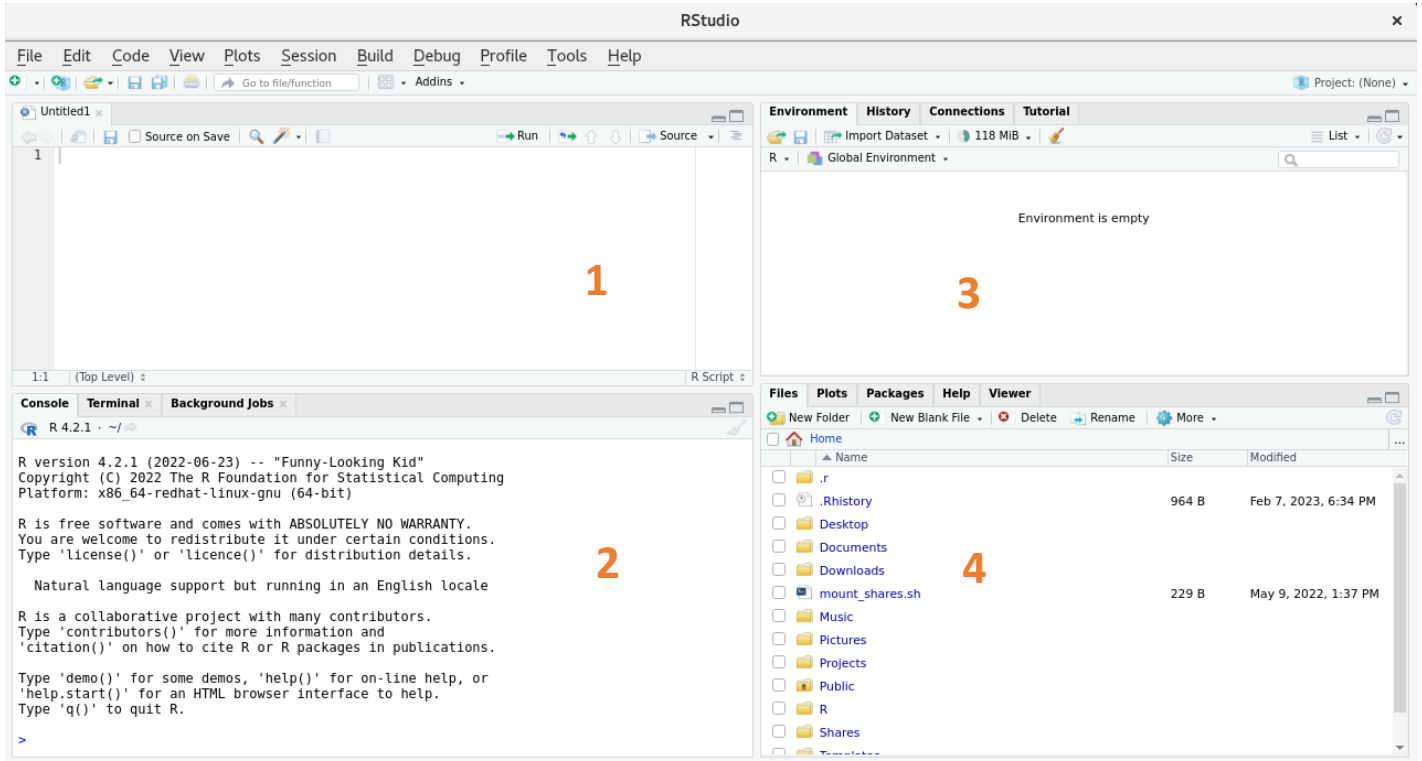
R-help -- Main R Mailing List: Primary help

RStudio: Integrated Development Environment (IDE)

Go to the [RStudio course server]

RStudio spaces:

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

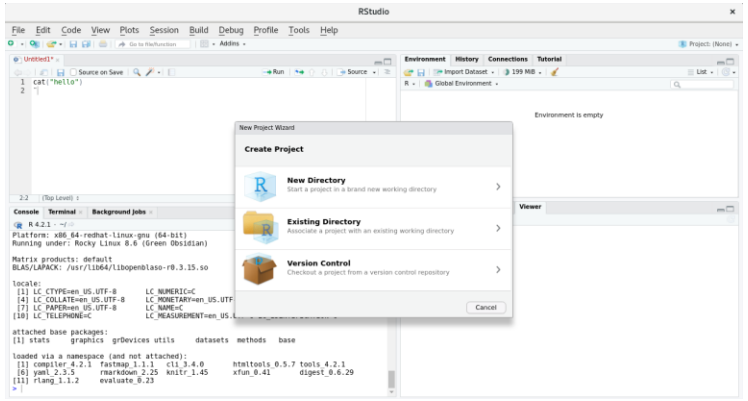


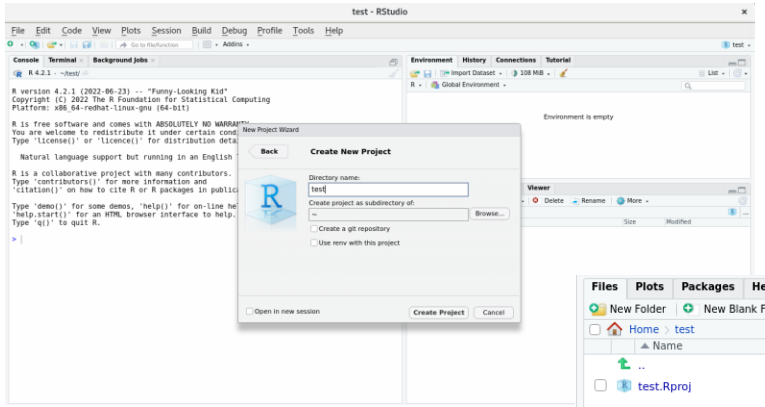
Cheatsheet @ https://rstudio.github.io/cheatsheets/html/rstudio-ide.html

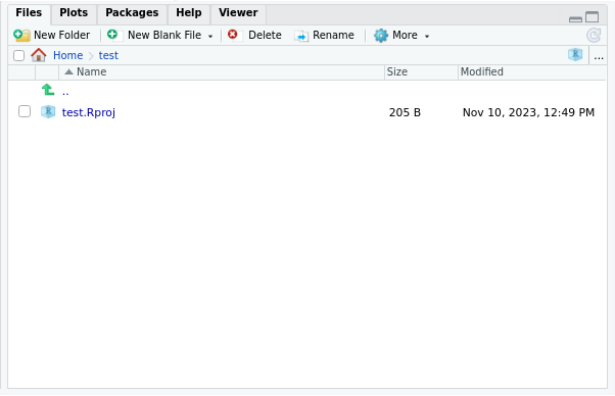
Working with projects

- Everything in one place
- Only relative paths

> File > New Project > New Directory







Practical: 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

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

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!

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)

Further resources

Books:

- R Programming for Data Science (D. Peng, 2022)
- R for data science 2ed (H.Wickham, M. Certinkaya-Rundel & G.Grolemund, 2023)

Tutorials:

- Datanovia

Musings:

- Medium: Towards data science

Thank you!

IJC Bioinformatics