Introduction to R and Bioconductor

2022-8-25, Franziska Greulich

- a statistical environment
- A programming language for data analysis and visualization

Why use R?

- Complete statistical enviroment and programming language
- Efficient functions and data structures for data analysis
- Powerful graphics
- Free access to growing number of packages
- Excelent documentation and build-in help system
- Free, open-source and avaliable for all OSs

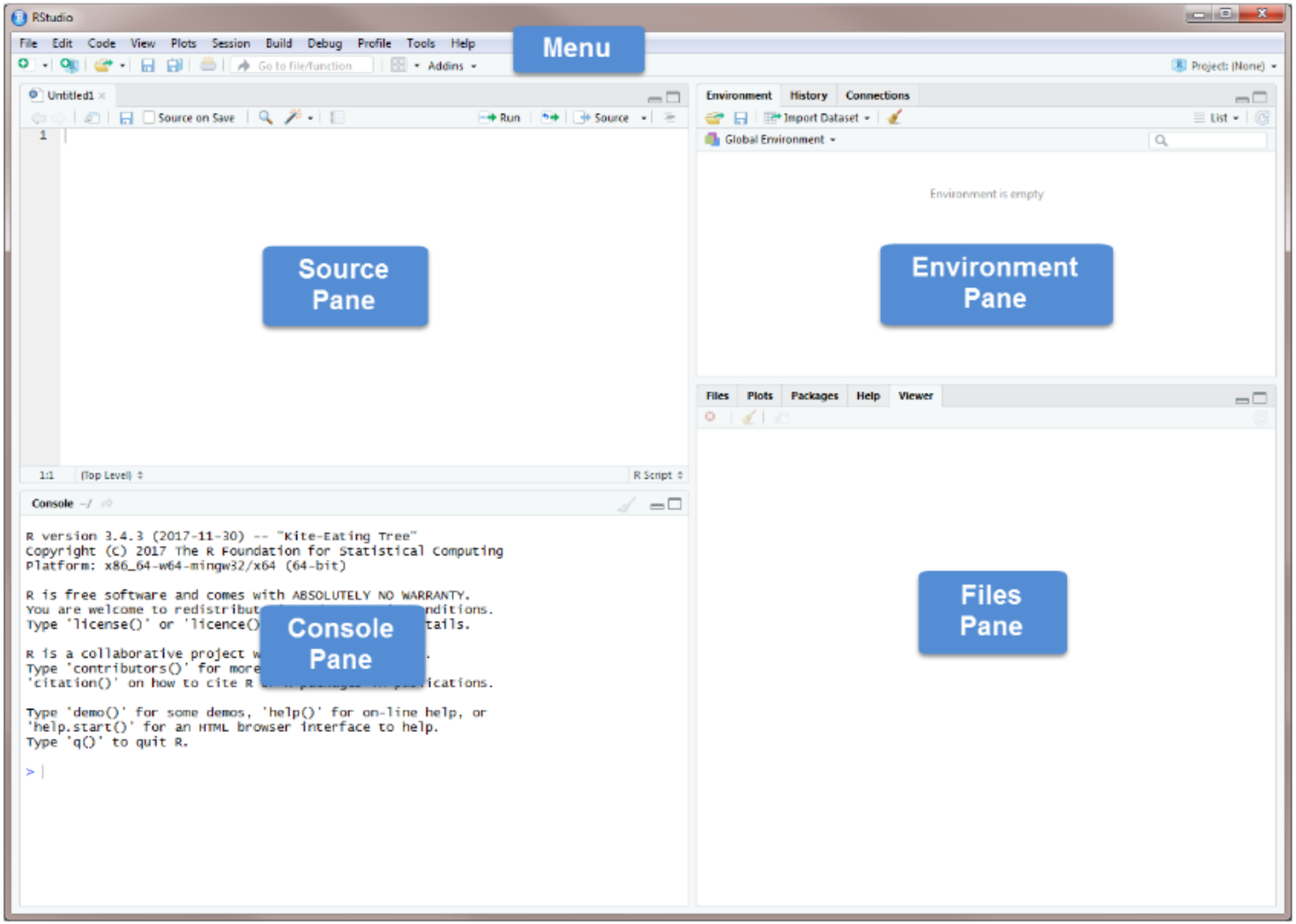
Packages and repositories

1. CRAN – The Comprehensive R Archive Network:
   - > 14K packages
   - Purpose: general data analysis
   - https://cran.r-project.org/
2. Bioconductor:
   - > 2K packages
   - bioscience data analysis
   - Includes extensive vignettes, courses and training
   - https://www.bioconductor.org/
3. Rstudio packages
   - https://www.rstudio.com/products/rpackages/

Studio

- Integrated development environment (IDE) for R (but also C, C++, python, bash…)
- Provides simple graphical user interface
- Open-source and commercial for Windows, Mac and Linux
- Supports version control, package authoring and documents

Studio


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Menu
Untitled1
Source on Save
Run
Source
Source Pane
Environment History Connections
Import Dataset
Global Environment
Environment is empty
Environment Pane
Files Plots Packages Help Viewer
Files Pane
Console
R version 3.4.3 (2017-11-30) -- "Kite-Eating Tree"
Copyright (C) 2017 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
Type 'license()' or 'licence()
R is a collaborative project w
Type 'contributors()' for more
'citation()' on how to cite R
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
Console Pane

Studio exercise (10 min)

1. Open Rstudio
2. Type 2+4 into the console and hit ENTER
3. Create a new R Script: /File/New File/R Script
4. Write 2+4 into your script and hit CTRL+ENTER. What does the console show?
5. Assign 2+4 to a variable x and hit CTRL+ENTER. What does the console show? What happened to your environment pane?
   - Assignment operators in R: =, <-
6. Type x into the console and hit ENTER.

Working „reproducible “

1. Check your R and Rstudio version:

   $ R.version$version.string

   $ RStudio.Version()$version

Those should be R version 4.2.x and Rstudio version 2022.7.1.xxx

Working „reproducible “

1. Work in **R projects**
   - version controlled via git locally and globally (Github)
   - comes with own package enviroment
   - combine data, code and results

Working „reproducible “

Exercise: (5 min)

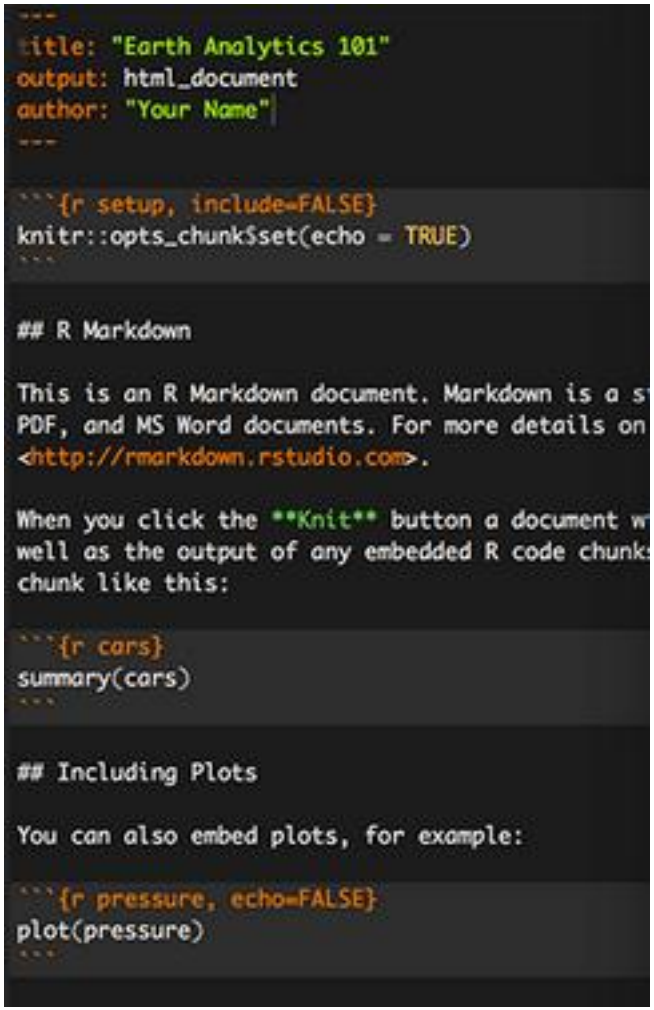
- Create an R project: /File/New Project/
- Select a New Directory/New Project and call it test (choose a local location)
- Check the boxes „Create a git repository“ and „Use renv with this project“
- Create Project

markdown

2. Use **Rmarkdown** documents

- rmarkdown and knitr R packages for literate programming **linking statistical analysis and reporting of results** in one document
- Uses the markup language
- Interleave code chunks with descriptive text and results
- .Rmd documents can be rendered into HTML, DOC or PDF

markdown



YAML header (---)

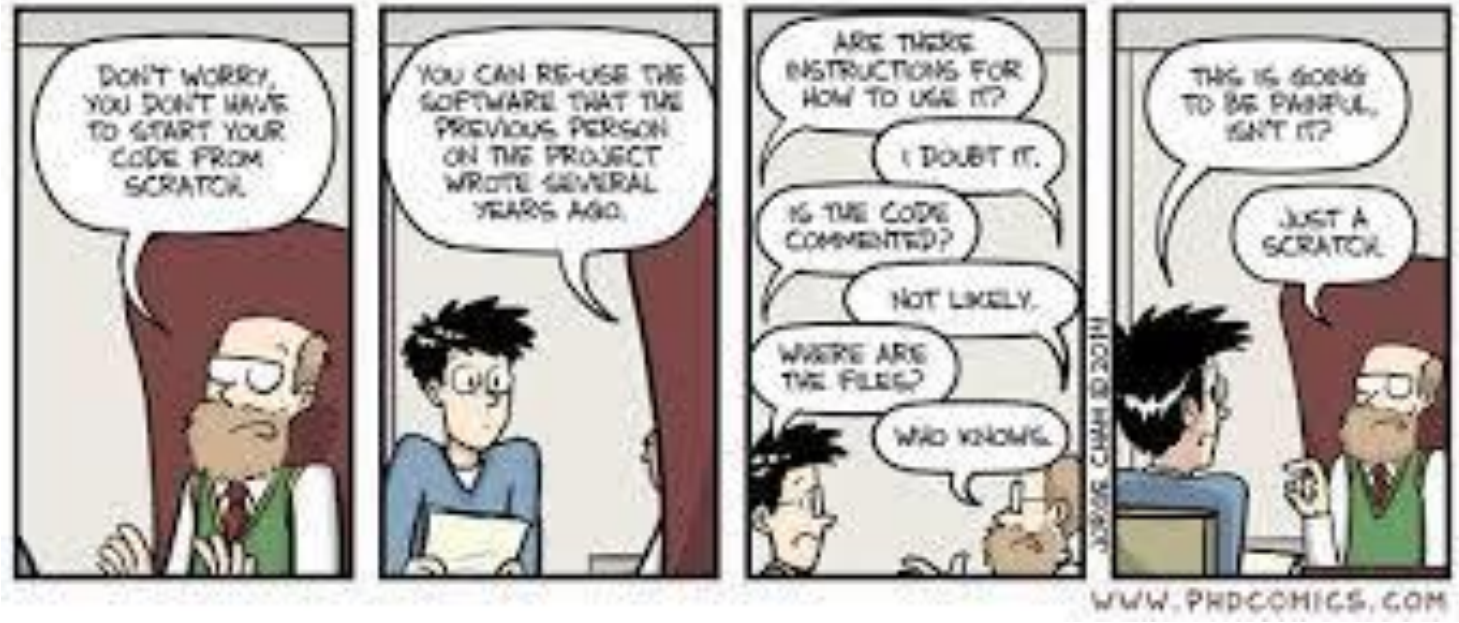
Code chunk (````)

Text simply formatted using markdown

Integrate result tables, plots …

markdown

3. Comment, comment, comment



Further reading

- Rstudio: https://sites.tufts.edu/datalab/files/2018/04/R_RStudio_Basics.pdf; https://psyteachr.github.io/reprores-v2/intro.html
- Version control of R projects: https://swcarpentry.github.io/git-novice/14-supplemental-rstudio/index.html

- Cheatsheets:
  - Rmarkdown: https://www.rstudio.com/wp-content/uploads/2015/02/rmarkdown-cheatsheet.pdf