1- Introduction to the R language

Alex Sanchez, Miriam Mota, Ricardo Gonzalo and Mireia Ferrer

Statistics and Bioinformatics Unit. Vall d'Hebron Institut de Recerca

Readme

- License: Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License http://creativecommons.org/licenses/by-nc-sa/4.0/
- You are free to:
  - **Share** : copy and redistribute the material
  - **Adapt** : rebuild and transform the material
- Under the following conditions:
  - **Attribution** : You must give appropriate credit, provide a link to the license, and indicate if changes were made.
  - **NonCommercial** : You may not use this work for commercial purposes.
  - **Share Alike** : If you remix, transform, or build upon this work, you must distribute your contributions under the same license to this one

Introduction to R

Outline

- A first contact with R & Rstudio.
  - How does one work with R
- A primer of data import
  - Reading data into R
- A primer of communication
  - R Notebooks and RMarkdown

A first contact with R, Rstudio and the tidyverse

What is R?

- R is a *language and environment* for statistical computing and graphics.
- R provides a wide variety of statistical and graphical techniques, and is highly extensible.
- It compiles and runs on a wide variety of UNIX platforms and similar systems Windows and MacOS.

R PRO's (why you are here!)

- The system is
  - free (as in *free beer*)
  - It's platform independent
  - It is constantly improving (2 new versions/year)
- It is a statistical tool
  - Implements almost every statistical method that exists
  - Great graphics (Examples)
  - Simple reporting tools
  - Also state-of-the-art in Bioinformatics through the Bioconductor Project.
- Programming language
  - Easy to automate repetitive tasks (Example_1.1)
  - Possibility to create user friendly web interfaces with a moderate effort. (Examples)

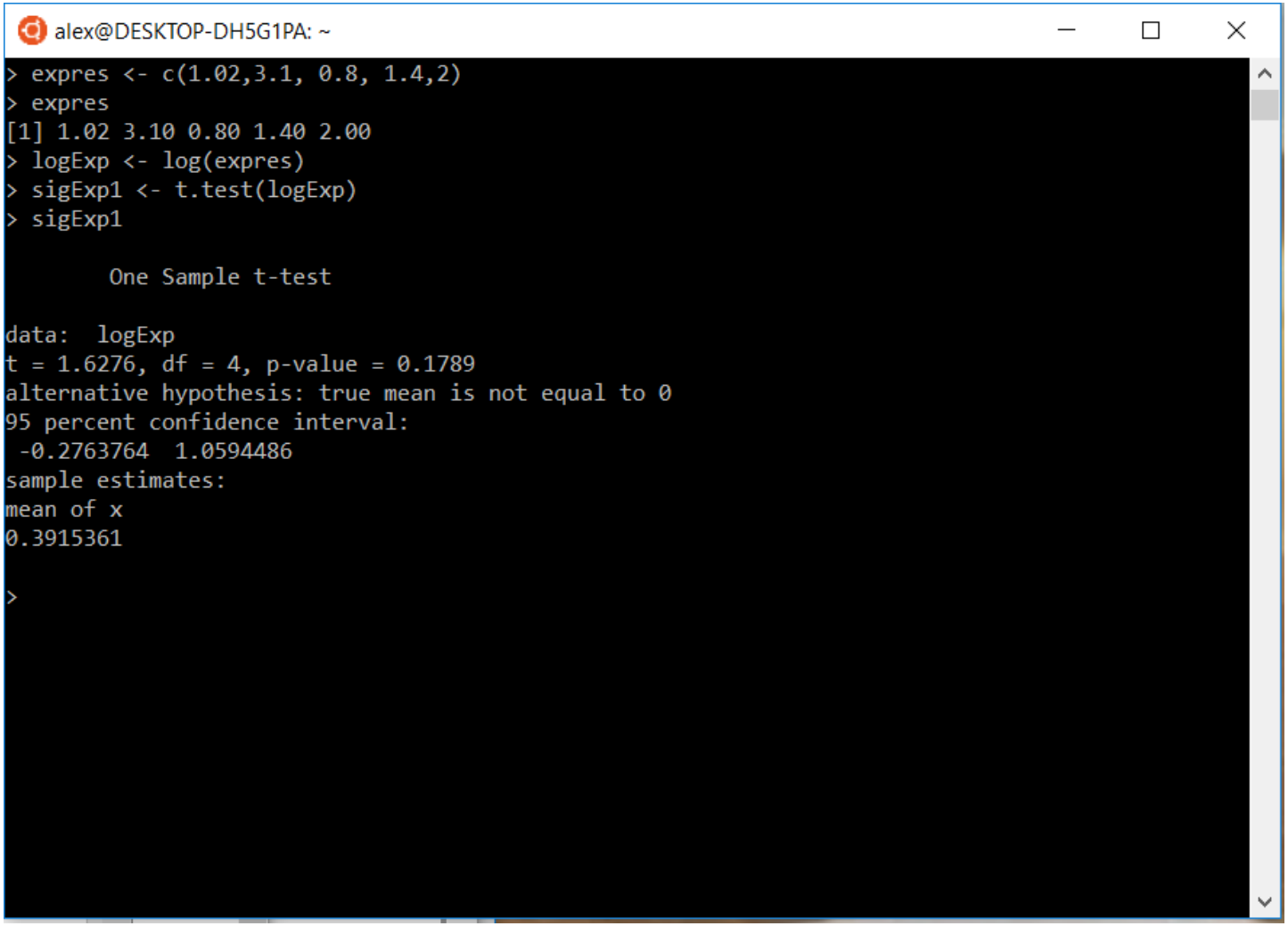
R CON's

- R is mainly used issuing commands from a console
  - less user friendly that almost any other statistical tool you may know.
- Constantly having new versions may affect our projects
- Not necessarily the best language nor suitable for every existing task

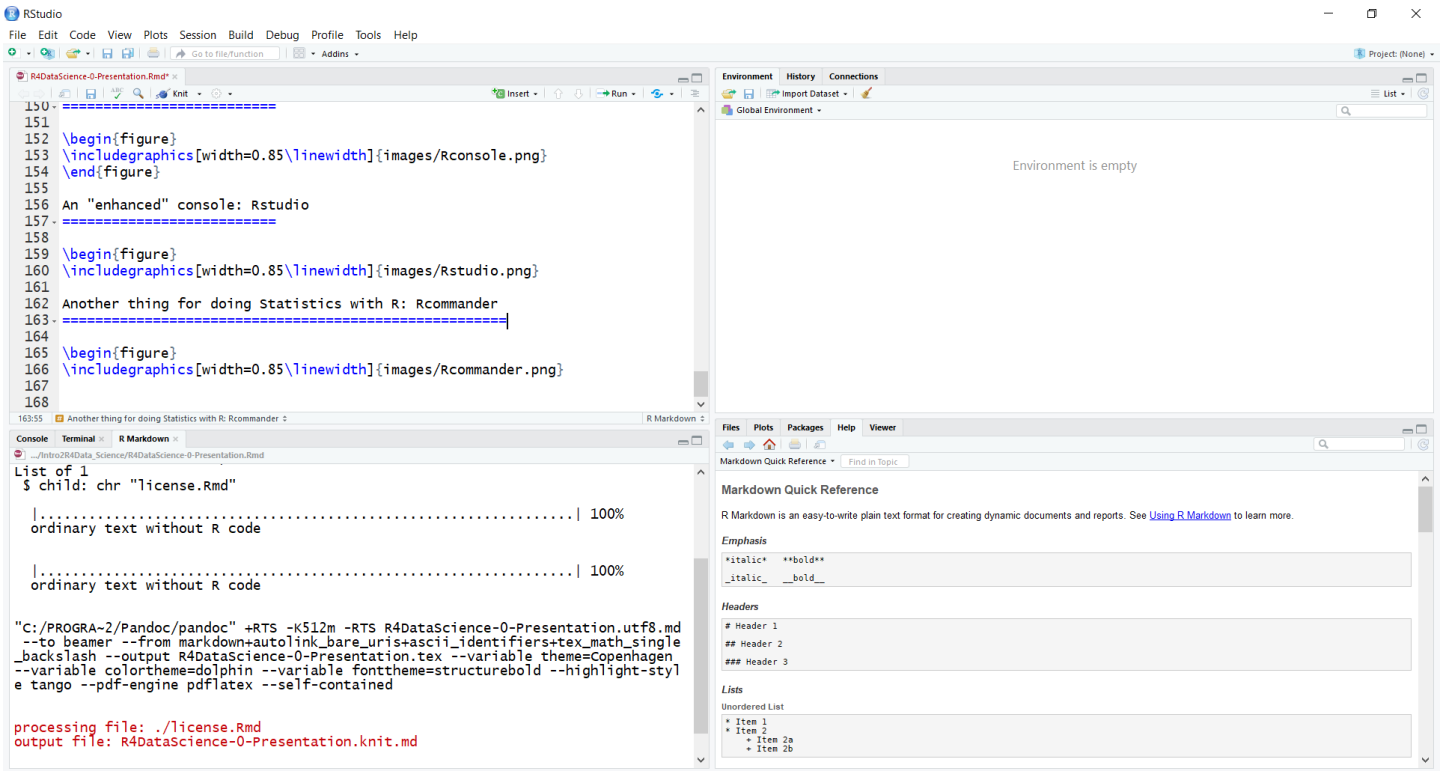
How is R used

- Traditionally R was used from an Operating System console ("Terminal")
- This is an intimidating approach for many users
- A variety of options exist to decrease the learning curve.
  - Use a supportive development environment such as **Rstudio**
  - Use an interface to Statistical tools, such as **Rcommander** or ::DeduceR** allowing to concetrate an Statistics, not in commands.

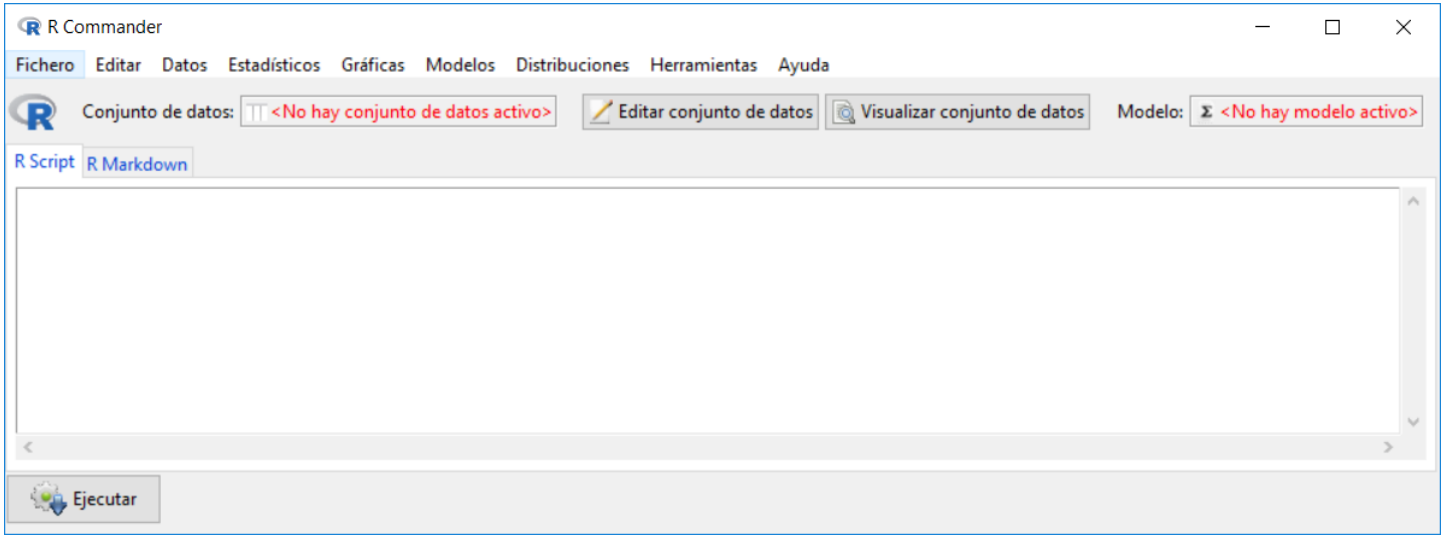
A raw R console in linux



An "enhanced" console: Rstudio



Something that is not a console: Rcommander



Using R

Commands, Objects and Functions

Some explanations here

Packages and datasets

Some explanations here

The tidyverse

Some explanations here

Some basic data types

Some explanations here

Getting data into R

Importing data with Rstudio

Some explanations here

Reading Excel files

Some explanations here

Reading text files

Some explanations here

Interlude: Summarizing data

Some explanations here

Using Rmarkdown to produce dynamic output

Reproducible research with R notebooks

Some explanations here

Dynamic reports with Rmarkdown

Some explanations here