Using R

Göran Broström

2016-11-30

What is R?

R is

- a language and environment for *statistical computing* and *graphics*.
- a *GNU project* originating from the *S* language and environment which was developed at *Bell Laboratories* (formerly AT&T, now Lucent Technologies) by *John Chambers* and colleagues.
- a different implementation of S.
- 25 years old.

From The R project

What is R?

- R provides a wide variety of statistical and graphical techniques, and is *highly extensible*.
- One of R's strengths is the ease with which *well-designed publication-quality plots* can be produced, including mathematical symbols and formulae where needed.

The R environment

R includes

- an effective *data handling* and *storage* facility,
- operators for *calculations on arrays*, in particular *matrices*,
- intermediate *tools for data analysis*,
- superb *graphical facilities*, and
- a simple and effective *programming language* which includes
  - *conditionals*,
  - *loops*,
  - user-defined recursive functions and
  - *input* and *output* facilities.

Why R?

- **R** is *free*:
  - as in *free beer*,
  - as in *free speech*.
- **R** can be installed on as many computers you like
- **R** is *open-source*.
- **R** has thousands (7000+ today) of *add-on packages*.

What is RStudio?

RStudio allows you to *run R* in a more *user-friendly* environment

- It is *free* and *open-source*,
- Tools for *interacting* statistical *analysis* with report *writing*.
  - *R markdown*
  - *knitr*
  - LaTeX
- Tools for *version control* (git, gitHub)
- Support.

Why RStudio?

Also,

- It is *available* on all (Linux, Mac, Windows) platforms.
- It looks *exactly the same* on all platforms.
  - It *behaves exactly the same* on all platforms.

What is reproducible research?

- It can be *reproduced by others*,
- It can be reproduced *by yourself* (6 months later)!

Focus is on reproducing results *from a single set of observations*.

What is reproducible research *not*?

- We do *not* mean *replicable*.
  - [Replicable:] If the *results* can be replicated on *another* study population.
- It is rather about rules of *disciplin* and *documentation*.
- *RStudio* gives you the *tools*.

Version control

- Must be *easy* to go back to *earlier versions* of your research:
  - You detect *errors*,
  - You get *new data*,
  - The journal asks you to perform *separate analyses* for men and women,
  - etc.
- There are *tools* to help:
  - git, by Linus Torvalds,
    - gitHub