Introduction

Utrecht University Winter School: Introduction to R

Utrecht University

Kyle M. Lang

Department of Methodology & Statistics
Utrecht University

Plan for the Day

After this lecture, we'll spend the rest of the day working through interactive R scripts.

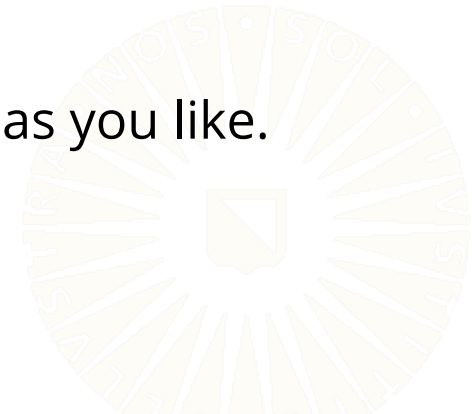
- The scripts live in the "code" directory.
- You should follow along.
- You will solve embedded practice problems.

We'll stop for a one-hour lunch break between 13:00 and 14:00.

- We'll take a few other breaks in between.

The course will finish at 18:00.

- You're free to access the course materials for as long as you like.



What is R?

R is a holistic (open-source) software system for data analysis and statistical programming.

- R is an implementation of the S language.
  - Developed by John Chambers and colleagues
    - Becker and Chambers (1984)
    - Becker, Chambers, and Wilks (1988)
    - Chambers and Hastie (1992)
    - Chambers (1998)
- Introduced by Ihaka and Gentleman (1996).
  - Currently maintained by the *R Core Team*.
- Support by thousands of world-wide contributors.
  - Anyone can contribute an R package to the *Comprehensive R Archive Network* (CRAN)
  - Must conform to the licensing and packaging requirements.

What is R?

I prefer to think about R as a *statistical programming language*, rather than as a data analysis program.

- R **IS NOT** its GUI (no matter which GUI you use).
- You can write R code in whatever program you like (e.g., RStudio, EMACS, VIM, Notepad, directly in the console/shell/command line).
- R can be used for basic (or advanced) data analysis, but its real strength is its flexible programming framework.
  - Tedious tasks can be automated.
  - Computationally demanding jobs can be run in parallel.
  - R-based research *wants* to be reproducible.
  - Analyses are automatically documented via their scripts.

Getting R

You can download R, for free, from the following web page:

- `https://www.r-project.org/`

You will also need a proper text editor/IDE. For those who are just learning R, I recommend **RStudio**:

- `https://www.rstudio.com/`

How R Works

R is an interpreted programming language.

- The commands you enter into the R *Console* are executed immediately.
- You don't need to compile your code before running it.
- In this sense, interacting with R is similar to interacting with other syntax-based statistical packages (e.g., SAS, STATA, Mplus).

How R Works

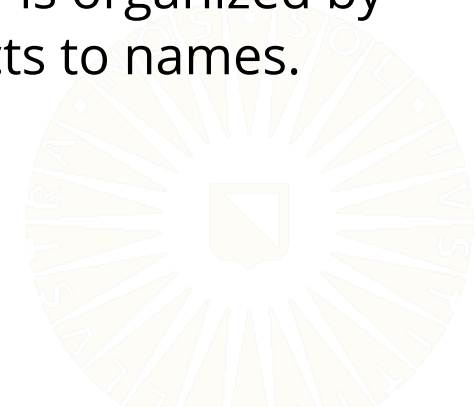
R mixes the *functional* and *object-oriented* programming paradigms.

Functional

- R is designed to break down problems into functions.
- Every R function is a first-class object.
- R uses pass-by-value semantics.

Object-Oriented

- Everything in R is an object.
- R functions work by creating and modifying R objects.
- The R workflow is organized by assigning objects to names.

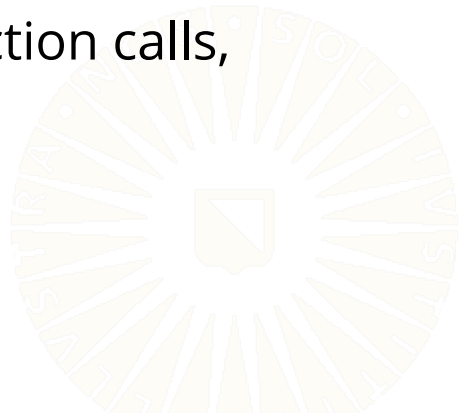


Interacting with R

When working with R, you will write *scripts* that contain all of the commands you want to execute.

- There is no “clicky-box” Tom-foolery in R.
- Your script can be run interactively or in “batch-mode”, as a self-contained program.

The primary purpose of the commands in your script will be to create and modify various objects (e.g., datasets, variables, function calls, graphical devices).



What's Next?

Now, we'll start in on the meat-and-potatoes of the course

- The interactive R scripts

We have seven potential topics:

1. Basic commands
2. Data objects
3. Data I/O
4. Data manipulation
5. Data analysis
6. Data visualization
7. Simple programming

References

Becker, R. A., & Chambers, J. M. (1984). *S: an interactive environment for data analysis and graphics*. Monterey, CA: Wadsworth and Brooks/Cole.

Becker, R. A., Chambers, J. M., & Wilks, A. R. (1988). *The new S language*. London: Chapman & Hall.

Chambers, J. M. (1998). *Programming with data: A guide to the S language*. New York: Springer Science & Business Media.

Chambers, J. M., & Hastie, T. J. (1992). *Statistical models in s*. London: Chapman & Hall.

Ihaka, R., & Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, *5*(3), 299–314.