

# Introduction to Data Analysis and Graphics with R - Session 1

Federico Vegetti

University of Mannheim

*fede.vegetti@gmail.com*

September 4, 2012

# Goals of the course

At the end of this course you should:

- 1 Be able to move comfortably within the R environment.
- 2 Be able to perform basic data analysis, number crunching, plotting and simulations using R.
- 3 Feel free to choose R *when you need it*.
- 4 Know how to find help and information about what you need to do with R.

# Structure of the course

- Seven sessions, roughly 3 hours each
- 06-09, 13-09, 04-10, 11-10, 18-10, 25-10, 08-11
- (15-11 & 22-11)
- The course is not graded, you can either *pass* or *fail*.
- To give you the credits, I want you to:
  - 1 Attend the classes
  - 2 Active participation
  - 3 Do some assignments (from October)
- My office (for the time being):  
A5, 6. Gebäudeteil A. MZES, Room 218.
- Office hours by appointment.  
My email: [fede.vegetti@gmail.com](mailto:fede.vegetti@gmail.com)

# What is R?

- 1 A computer language.
- 2 An interpreter that executes code written in R.
- 3 A graphic system.
- 4 An application that includes the interpreter, the graphic system, packages, user interface.

# Why you want to learn how to use R

- That stats class I need to take for credits is taught in R!
- R is used at the world's largest technology companies (including Google, Microsoft, and Facebook).
- Many other companies and research institutes have migrated to R.
- R is used at statistics classes at universities around the world and by statistics researchers to try new techniques and algorithms.
- Your graphs will look better, no matter what.

# Why you don't feel like learning how to use R

- Steep learning curve.
- You need to master a bit of programming to get (even easy) things done.
- Other softwares can easily get you through basic data analysis via point-and-click.

# R vs. other statistical software packages

- R is free!
  - You may like this the day you won't have a university affiliation.
- You can find implementations for a huge amount of statistical and data analysis algorithms using R.
- Have a problem? Ask the community.
  - There are many R users worldwide. You will be impressed about the amount of discussions boards and forums about R.
- R performs better.
  - To have an idea, try to run a simple multilevel with Stata and with R.

# What are the weak points?

- R requires you to load data sets into memory before processing.
  - This means that you cannot work with very large data sets, unless you have a very large RAM.
- R is a pain for cleaning your data.
- Some other weak points, nothing that concerns us as social science students.

# Some history

- R is based on the computer language S.
- S was developed in 1976 by John Chambers and others at the Bell Labs.
- In the early 1990s, Ross Ihaka and Robert Gentleman (both at the University of Auckland, New Zealand) came out with R after playing around with the S language.
- Allegedly, they called it “R” because of the initials of their first names.
- Since 1997, R is developed by the R Development Core Team.

# Graphical interfaces

- Several interfaces for R—none of them will save you from writing syntax though.
- The most basic one is the graphical user interface (GUI) that you get when you download R.
- Probably the most point-and-click-like is [DeduceR](http://www.deducer.org/pmwiki/pmwiki.php):  
<http://www.deducer.org/pmwiki/pmwiki.php>
- [RExcel](http://rcom.univie.ac.at/) allows you to run R from within Microsoft Excel:  
<http://rcom.univie.ac.at/>
- Many other interfaces, including [Tinn-R](#), [RKward](#), [R Commander](#) and [RStudio](#)

# RStudio

- For this course we will use RStudio
- <http://rstudio.org/>
- It has some advantages:
  - It shows everything you need without opening additional windows.
  - It has some very light point-and-click features.
  - It still requires syntax, but it makes your life easier (e.g. highlights parentheses, adds quotation marks, displays different colors for commands and numbers, etc.).
- However, if you want to use another GUI or text editor, it will make virtually no difference.

# Installing R on your computer

<http://www.r-project.org/>

## The CRAN Comprehensive R Archive Network

- <http://cran.r-project.org/mirrors.html>
- From here you can download the installer, as well as most of the additional packages you need.
- You want to choose a *mirror* i.e. a server located possibly close to where you are (three in Germany).
- R is available for Windows, MacOS X, and Linux.

The site also provides a number of manuals and help facilities:  
<http://ftp.yalwa.org/cran/manuals.html>

# Installing RStudio on your computer

`http://rstudio.org/download/`

- Again, available for Windows, Mac and Linux.
- You will need the “Desktop” version.
- New version 0.96.331 just released.
- You need to download R separately.

# How RStudio looks like

