



University of Torino, Earth Science Department

**COST Action ES0601**

**HOME: Advances in HOMogenisation METHODS for climate series: an integrated approach**

# R Summer School

September 7<sup>th</sup>-10<sup>th</sup>, 2009

Physics Department, University of Patras, Greece

---

---

# Overview

- R: introduction and basics
  - Reading and writing files
  - Programming
  - Function definition
  - Introduction to graphics
  - Statistical functions
  - Graphics in R
  - Introductory statistics in R
-

---

# **Introduction and Basics**

---

---

# Introduction

Statistics:

A collection of methods to transform *data* in *information*.

Steps for a good statistical analysis:

1. Accurate collection and “preparation” of data (long time is required!)
2. Development and/or application of statistical methods.

To help in the second step many statistical software are available.

---

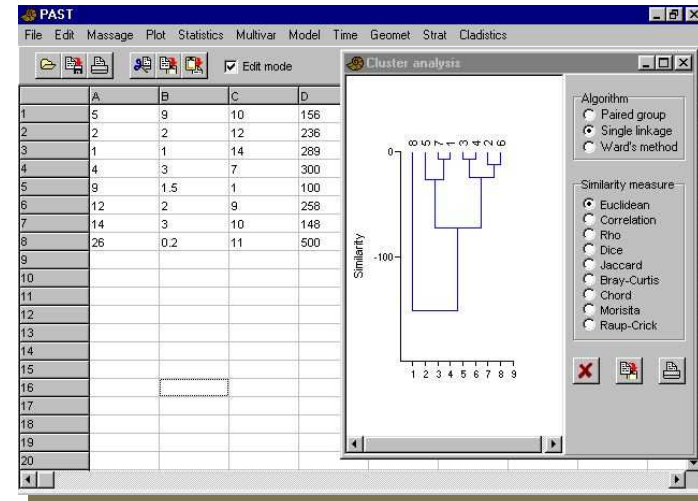
---

# Statistical Software

- “Spreadsheet”  
Excell, OpenCalc
  - Commercial statistical software  
SPSS, Statistica, Matlab
  - Free statistical software  
R, PAST
  - Software developed *ad hoc* by the user
-

# Data management

- “Spreadsheet”  
(Excell, OpenOffice, PAST,...)



- Consolle with  
command line  
(R, Matlab,...)

```

DATAPLOT - [Graphic1]
File Edit View State Window Help

#####
Dataplot
Interactive Graphics & Data Analysis Language
James J. Filliben & Alan Heckert
Computing & Applied Mathematics Laboratory
National Institute of Standards & Technology
301-975-2855 & 301-975-2899
#####

Number of Observations per Variable = 20000
Number of Variables = 10
Total Internal Data Space Size = 200000
Substitution/Replacement Character = ^

1. Welcome to DATAPLOT (version 8/1999). For assistance at
any time during a run, enter HELP .

2. For a list of new commands/capabilities,
enter NEWS (updated 8/1999)

3. The alternate Graphics output files (creatable
via the DEVICE 2 and 3 commands) are
dnp1f.tex and dnp2f.tex respectively.

Paused | Input pending in Graphic1
```

---

# Spreadsheet

## Advantages:

- World-wide diffusion
- Usually very easy to use
- Free software (as OpenOffice Calc) are available

## Disadvantages:

- Limited statistical functions
-

---

# Commercial software

## Advantages:

- Many statistical functions
- Big-size program
- Excellent manuals, help

## Disadvantages:

- Very expensive!
  - Quite complicated to use
-

---

# Open Source Statistical Software

## Advantages:

- FREE
- Oriented for specific fields of research, so very complete.
- Frequently updated
- Community of users's support

## Disadvantages:

- Not so easy to use
-

---

**R**

**...a very good compromise  
among all this options...**

---

## What is R?



R is a “**language** and **environment** for statistical computing”

implementation of the **S** *programming language* which was developed at Bell Laboratories by Rick Becker, John Chambers and Allan Wilks

R as an “environment” within which statistical techniques are implemented.

All R functions are stored in “**packages**”

---

## Base packages

They contain the basic functions that allow R to work (e.g. functions for data manipulation, basic statistical analysis and graphics).

They are considered part of the R source code and they are automatically available anytime you run R.

## Contributed packages

These packages allow to implement specialized statistical methods. Some of them (the *recommended* packages) are distributed with any binary distribution of R, the others can be downloaded from the net.

About 2000 packages are available at the moment!

---

---

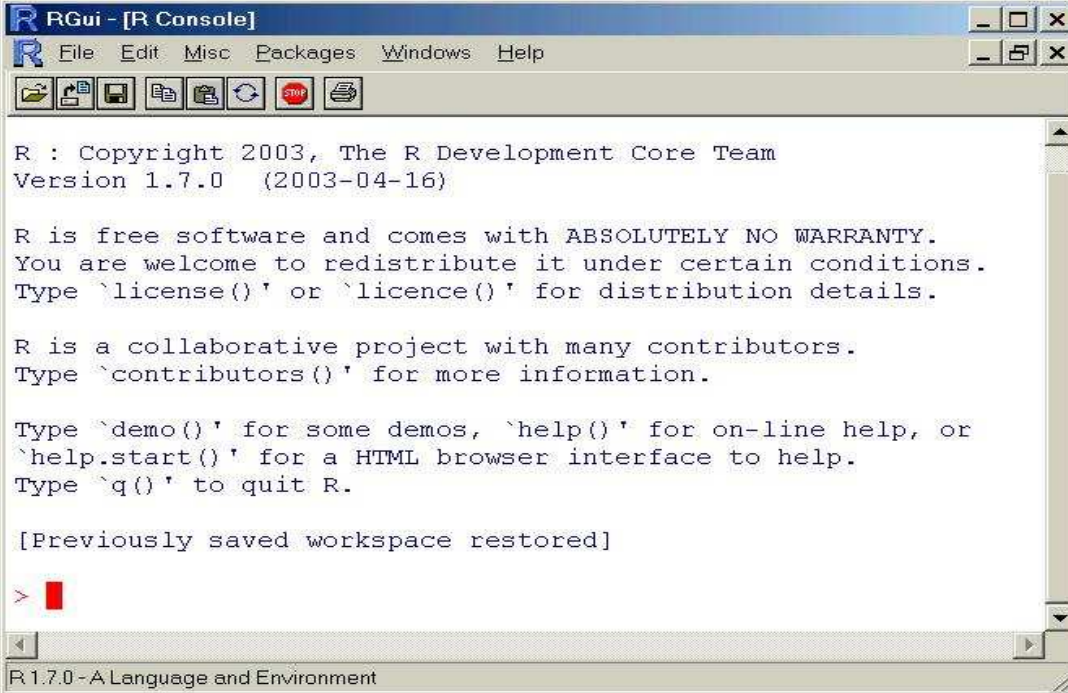
# Why R?

- Free Software
  - Available for Windows, Linux, MacOS
  - Easy data manipulation and storage
  - Graphical facilities for data analysis and display
  - Unlimited possibilities for statistical analysis (flexible)
  - Developement of user-defined functions
  - Excellent help
-

... but  
some time is required to get started to the R language

R interface is a command line: when you start R it will  
appear the prompt:

>\_



```
RGui - [R Console]
File Edit Misc Packages Windows Help
R : Copyright 2003, The R Development Core Team
Version 1.7.0 (2003-04-16)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type `license()` or `licence()` for distribution details.

R is a collaborative project with many contributors.
Type `contributors()` for more information.

Type `demo()` for some demos, `help()` for on-line help, or
`help.start()` for a HTML browser interface to help.
Type `q()` to quit R.

[Previously saved workspace restored]

> █
```

R 1.7.0 - A Language and Environment

---

**After the first day of this school...**

**you will know all the basic commands  
to successfully interact with R**

---