Lecture given at the Naresuan University (NU)
Phitsanulok, 17 November 2011

An introduction to R programming language

Hristos Tyralis, Civil Engineer,
Msc in Statistics and Operations Research, PhD candidate
Department of Water Resources and Environmental Engineering
School of Civil Engineering
National Technical University of Athens, Greece
(montchrister@gmail.com, itia.ntua.gr/en/)

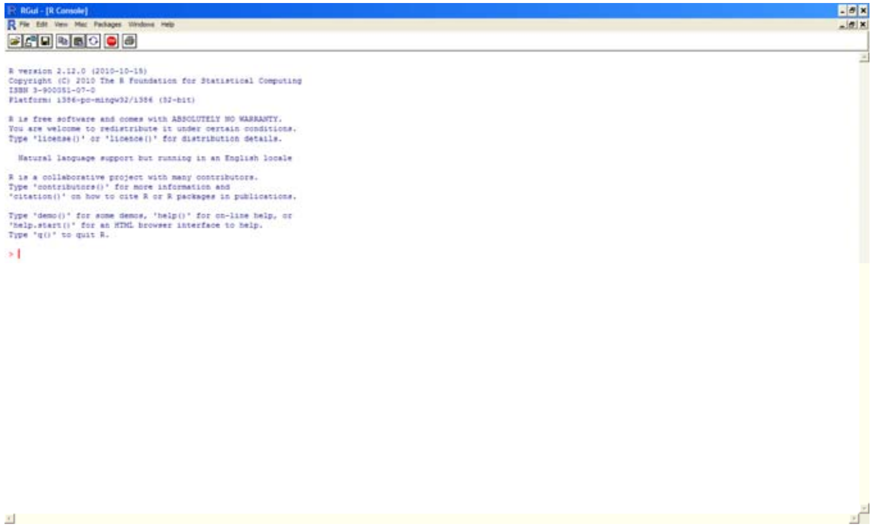
Presentation outline

1. Introduction
2. Examples
3. Help
4. More examples
5. Objects
6. Operators
7. Vectors and arrays
8. Graphics
9. Last but not least
10. Bibliography

This presentation will be available online at
http://itia.ntua.gr/en/docinfo/1230/

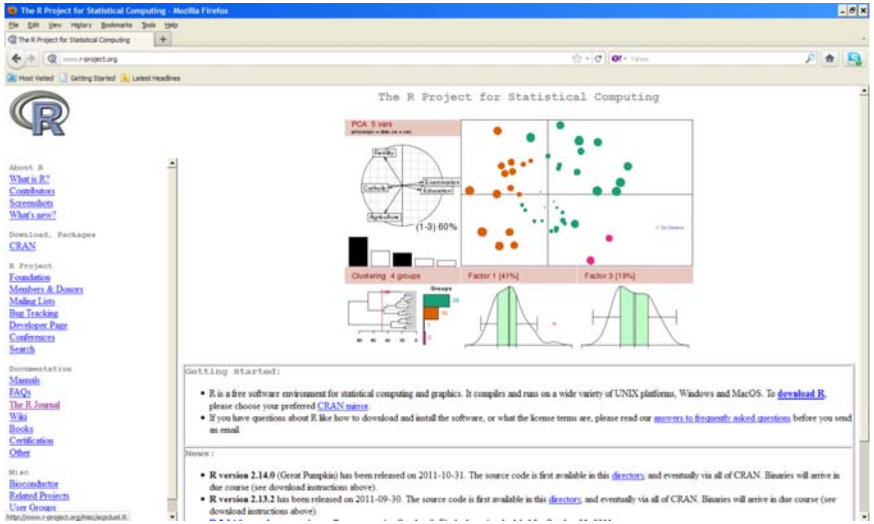
1.1 Introduction

- The presentation is about R under windows
- R is a system for statistical analyses and graphics.
- R is both a software and a language.
- Its development and distribution are carried out by several statisticians known as the R Development Core Team



```
R version 2.12.0 (2010-10-15)
Copyright (C) 2010 The R Foundation for Statistical Computing
ISBN 3-900051-07-0
Platform: i386-pc-mingw32/i386 (32-bit)

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.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

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

> 
```

R workspace



The R Project for Statistical Computing

About R
What is R?
Contributors
Screenshots
What's new?

Download, Packages
CRAN

R Project
Foundation
Members & Donors
Mailing Lists
Bug Tracking
Developer Page
Conferences
Search

Documentation
Manuals
FAQs
The R Journal
Wiki
Books
Certification
Other

Misc
Bioconductor
Related Projects
User Groups

Getting Started:

- R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To **download R**, please choose your preferred CRAN mirror.
- If you have questions about R like how to download and install the software, or what the license terms are, please read our answers to frequently asked questions before you send an email.

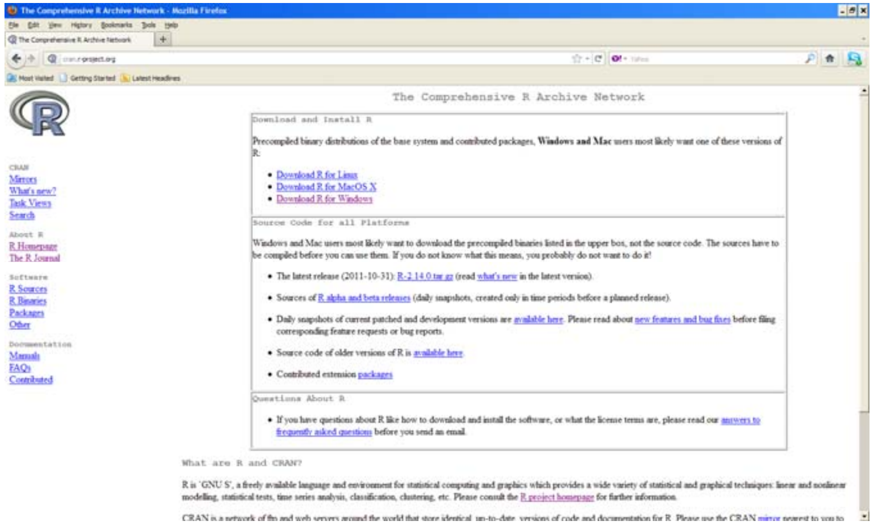
News:

- **R version 2.14.0** (Great Pumpkin) has been released on 2011-10-31. The source code is first available in this directory, and eventually via all of CRAN. Binaries will arrive in due course (see download instructions above).
- **R version 2.13.2** has been released on 2011-09-30. The source code is first available in this directory, and eventually via all of CRAN. Binaries will arrive in due course (see download instructions above).

The R project for statistical computing

1.2 Introduction

- R is freely distributed
- The files needed to install R are distributed from the internet site of the Comprehensive R Archive Network (CRAN).
- R journal is the refereed journal for the R project


The Comprehensive R Archive Network - Mozilla Firefox
File Edit View History Bookmarks Tools Help
The Comprehensive R Archive Network
cran.r-project.org
Most Visited Getting Started Latest Headlines

The Comprehensive R Archive Network

CRAN
Mirrors
What's new?
Task Views
Search

About R
R Homepage
The R Journal

Software
R Sources
R Binaries
Packages
Other

Documentation
Manuals
FAQs
Contributed

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- Download R for Linux
- Download R for MacOS X
- Download R for Windows

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2011-10-31): R-2.14.0.tar.gz (read what's new in the latest version).
- Sources of R alpha and beta releases (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are available here. Please read about new features and bug fixes before filing corresponding feature requests or bug reports.
- Source code of older versions of R is available here.
- Contributed extension packages

Questions About R

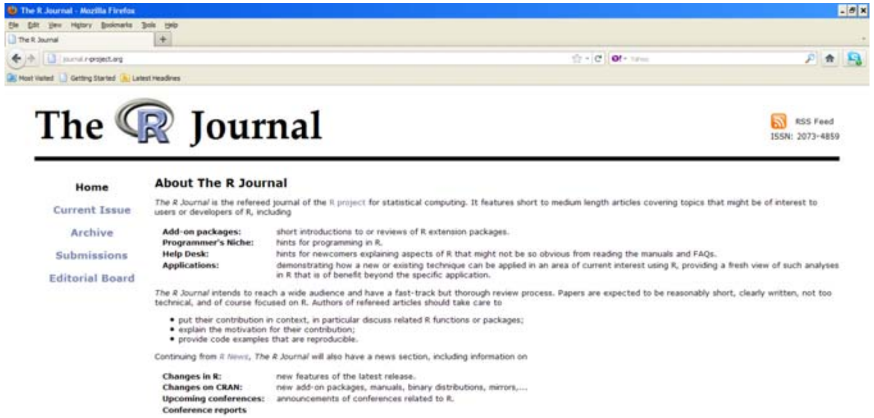
- If you have questions about R like how to download and install the software, or what the license terms are, please read our answers to frequently asked questions before you send an email.

What are R and CRAN?

R is 'GNU S', a freely available language and environment for statistical computing and graphics which provides a wide variety of statistical and graphical techniques: linear and nonlinear modelling, statistical tests, time series analysis, classification, clustering, etc. Please consult the R project homepage for further information.

CRAN is a network of ftp and web servers around the world that store identical, up-to-date, versions of code and documentation for R. Please use the CRAN mirror nearest to you to

The Comprehensive
R Archive Network


The R Journal - Mozilla Firefox
File Edit View History Bookmarks Tools Help
The R Journal
journal.r-project.org
Most Visited Getting Started Latest Headlines
RSS Feed

The R Journal

ISSN: 2073-4859

Home
Current Issue
Archive
Submissions
Editorial Board

About The R Journal

The R Journal is the refereed journal of the R project for statistical computing. It features short to medium length articles covering topics that might be of interest to users or developers of R, including

Add-on packages:	short introductions to or reviews of R extension packages.
Programmer's Niche:	hints for programming in R.
Help Desk:	hints for newcomers explaining aspects of R that might not be so obvious from reading the manuals and FAQs.
Applications:	demonstrating how a new or existing technique can be applied in an area of current interest using R, providing a fresh view of such analyses in R that is of benefit beyond the specific application.

The R Journal intends to reach a wide audience and have a fast-track but thorough review process. Papers are expected to be reasonably short, clearly written, not too technical, and of course focused on R. Authors of refereed articles should take care to

- put their contribution in context, in particular discuss related R functions or packages;
- explain the motivation for their contribution;
- provide code examples that are reproducible.

Continuing from *R News*, *The R Journal* will also have a news section, including information on

Changes in R:	new features of the latest release.
Changes on CRAN:	new add-on packages, manuals, binary distributions, mirrors,...
Upcoming conferences:	announcements of conferences related to R.
Conference reports	

R journal

1.3 Introduction

- R has many functions for statistical analyses and graphics
- The R language allows the user to combine in a single program different statistical functions
- The results from a statistical analysis are displayed on the screen
- A prominent feature of R is its flexibility
- R stores these results in an object

- For example a function that computes the ratio [(standard deviation)/mean] of a sample

```
ratiosm <- function(x) {
sd(x)/mean(x)
}
```

- sd(x) is a function, mean(x) is another function and ratiosm(x) is the new function

1.4 Introduction

- R is an interpreted language, not a compiled one, meaning that all commands typed on the keyboard are directly executed without requiring to build a complete program like in most computer languages (C, Fortran, Pascal, . . .)
- R's syntax is very simple and intuitive
- When R is running, variables, data, functions, results, etc, are stored in the active memory of the computer in the form of objects which have a name
- The user can do actions on these objects with operators (arithmetic, logical, comparison, . . .) and functions (which are themselves objects)
- Pay attention: Everything in R is an object. Even to define a vector we use an object

2.1 Examples

- assign operator: <-

```
> n <- 10
> n
[1] 10
```

this is equivalent to the following:

```
> n = 10
> 10 -> n
```

- R is case sensitive

```
> a <- 5
> A <- 10
> a
[1] 5
> A
[1] 10
```

2.2 Examples

- sum of two numbers: +

```
> n <- 10 + 2
> n
[1] 12
```

- just the sum of two numbers:

```
> 5 + 3
[1] 8
```

- some more examples of assignment

```
> name <- “Carmen”; n1 <- 3
> name
[1] “Carmen”
> n1
[1] 3
```

2.3 Examples

- How can I make a list of the workspace objects?: ls

```
> ls()
[1] “n” “n1” “name”
```

Be careful, ls is a function. Here it has not any arguments.

- Now we give it an argument:

```
> ls(pat = “m”)
[1] “name”
```

- Another list function: ls.str

```
> ls.str()
n : num 10
n1 : num 3
name : chr “Carmen”
```

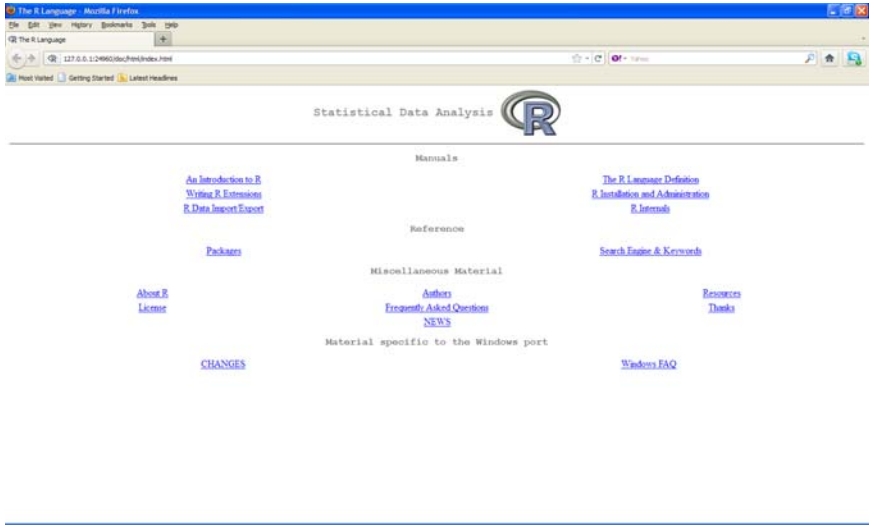
2.4 Examples

- How to remove all objects?: rm

```
> rm(list=ls())
```

3. Help

- R html-help
- It also works on the R environment


The R Language - Mozilla Firefox
File Edit View History Bookmarks Tools Help
The R Language
127.0.0.1:24960/doc/html/index.html
Most Visited Getting Started Latest Headlines

Statistical Data Analysis

Manuals

An Introduction to R
Writing R Extensions
R Data Import/Export

The R Language Definition
R Installation and Administration
R Internals

Reference

Packages

Search Engine & Keywords

Miscellaneous Material

About R
License

Authors
Frequently Asked Questions
NEWS

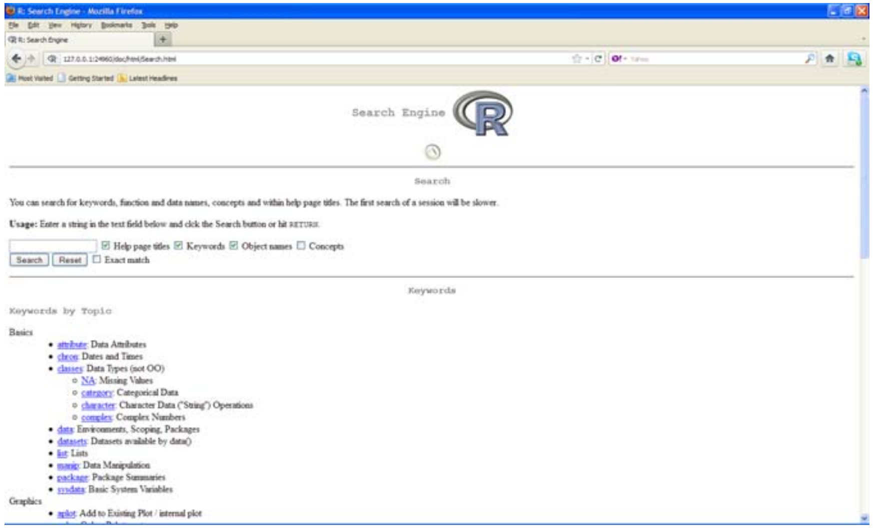
Resources
Thanks

Material specific to the Windows port

CHANGES

Windows FAQ

R html-help


R: Search Engine - Mozilla Firefox
File Edit View History Bookmarks Tools Help
R: Search Engine
127.0.0.1:24960/doc/html/Search.html
Most Visited Getting Started Latest Headlines

Search Engine

Search

You can search for keywords, function and data names, concepts and within help page titles. The first search of a session will be slower.

Usage: Enter a string in the text field below and click the Search button or hit RETURN.

☑ Help page titles ☑ Keywords ☑ Object names □ Concepts
Search Reset □ Exact match

Keywords

Keywords by Topic

Basics

- attribute: Data Attributes
- chron: Dates and Times
- classes: Data Types (not OO)
  - NA: Missing Values
  - category: Categorical Data
  - character: Character Data ("String") Operations
  - complex: Complex Numbers
- data: Environments, Scoping, Packages
- datasets: Datasets available by data()
- list: Lists
- manip: Data Manipulation
- package: Package Summaries
- sysdata: Basic System Variables

Graphics

- aplot: Add to Existing Plot / internal plot

search engine

4.1 More examples

- Object attributes?: mode, length

```
> x <- 1
> mode(x)
[1] “numeric”
> length(x)
[1] 1
```

- More attributes:

```
> A <- “Gomp” ; compar <- TRUE ; z <- 1i
> mode(A)
[1] “character”
> mode(compar)
[1] “logical”
> mode(z)
[1] “complex”
```

4.2 More examples

- Big numbers:

```
> N <- 2.1e23
> N
[1] “2.1e+23”
> x <- 5/0
> x
[1] “Inf”
> exp(-x)
[1] 0
```

5.1 Objects

- Objects representing the data:

- vector
- factor
- array
- matrix
- data frame
- ts
- list

- In this lecture we work with vectors and matrices

5.2 Objects

- Vector example:

```
> rm(list=ls())
> x <-  1:5
> x
[1] 1 2 3 4 5
> 1:10-1
[1] 0 1 2 3 4 5 6 7 8 9
> 1:(10-1)
[1] 1 2 3 4 5 6 7 8 9
> seq(1,5,0.5)
[1] 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0
> seq(length=9,from=1,to=5)
[1] 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0
> c(1 , 1.5 , 2 , 2.5 , 3 , 3.5 , 4 , 4.5 , 5)
[1] 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0
> rep(1,5)
[1] 1 1 1 1 1
> sequence(4:6)
[1] 1 2 3 4 1 2 3 4 5 1 2 3 4 5 6
```

5.3 Objects

- Normal distribution:

```
> a <- rnorm(1000, mean=0, sd=1)
```

Simulates 1000 numbers from a standard normal distribution

```
> hist(a)
```

See the histogram of a

```
> a <- rnorm(1000)
```

Simulates 1000 numbers from a standard normal distribution again. mean = 0 and sd = 1 are the default argument values of the function rnorm

```
> hist(a)
> a <- rnorm(1000, mean = 10, sd = 5)
```

Simulates from a normal distribution. See the histogram again.

```
> hist(a)
```

- In rnorm r denotes random numbers from a normal distribution. There are also the density function dnorm, the cumulative probability function pnorm and the value of the q quantile qnorm.

5.4 Objects

- Probability distributions in R:

- beta	beta
- binomial	binom
- Cauchy	caushy
- chi-squared	chisq
- exponential	exp
- F	f
- gamma	gamma
- geometric	geom
- hypergeometric	hyper
- log-normal	lnorm
- logistic	logis

and other distributions

6.1 Operators

- Arithmetic operators

- +
- -
- *
- /
- ^

- Logical operators

- <
- >
- <=
- >=
- == (for exact equality)
- != (for inequality)

6.2 Operators

- Some operations

```
> x<- 0.5 ; y <- 0.7
> 0 < x & x < 1
[1] TRUE
> 0 < x < 1 #wrong expression
Error: unexpected ‘x’ in “0 < x <“
> x<- 1:3 ; y <- 1:3
> x == y
[1] TRUE TRUE TRUE
> identical(x,y)
[1] TRUE
```

7.1 Vectors and arrays

- Some operations

```
> x <- 1:5
> x[3]
[1] 3
> x[3] <- 20
> x
[1] 1 2 20 4 5
> i <- c(1,3)
> x[i]
[1] 1 20
> x <- matrix(1:6, 2, 3)
> x
     [,1] [,2] [,3]
[1,]    1    3    5
[2,]    2    4    6
> x[,3] <- 21:22
> x
     [,1] [,2] [,3]
[1,]    1    3   21
[2,]    2    4   22
```

7.2 Vectors and arrays

- Some operations

```
> x <- 1:10
> x[x>=5] <- 20
> x
[1] 1 2 3 4 20 20 20 20 20 20
> x <- 1:4
> y <- rep(1,4)
> z <- x + y
> z
[1] 2 3 4 5
> x <- 1:4
> y <- 1:2
> z <- x + y #y is repeated until it reaches the length of x
> z
[1] 2 4 4 6
> x <- 1:4
> a <- 10
> z <- a * x #a is repeated until it reaches the length of x
> z
[1] 10 20 30 40
```

7.3 Vectors and arrays

- Some operations

```
> x <- rnorm(1000)
> sum(x)
[1] -40.19532
> prod(x)
[1] -3.135068e-264
> mean(x)
[1] -0.04019532
> var(x)
[1] 1.015683
> m1 <- matrix(1,nr=2,nc=2)
> m2 <- matrix(2,nr=2,nc=2)
> cbind(m1,m2)
     [,1] [,2] [,3] [,4]
[1,]    1    1    2    2
[2,]    1    1    2    2
```

7.4 Vectors and arrays

- Some operations

```
> rbind(m1,m2)
      [,1] [,2]
[1,]    1    1
[2,]    1    1
[3,]    2    2
[4.]    2    2
> rbind(m1,m2) %*% cbind(m1,m2) #matrix product
      [,1] [,2] [,3] [,4]
[1,]    2    2    4    4
[2,]    2    2    4    4
[3,]    4    4    8    8
[4.]    4    4    8    8
> cbind(m1,m2) %*% rbind(m1,m2) #matrix product
      [,1] [,2]
[1,]   10   10
[2,]   10   10
```

7.5 Vectors and arrays

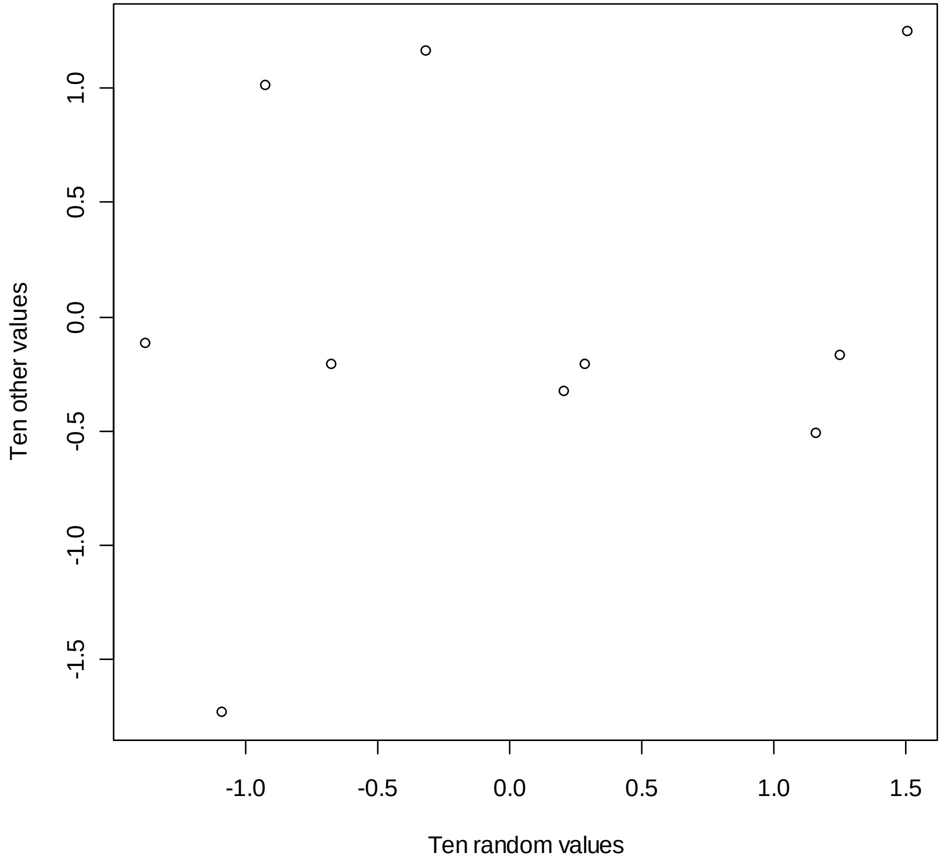
- Some operations

```
> diag(m1)
[1]   1    1
> diag(m1) <- 10
> m1
     [,1] [,2]
[1,]  10    1
[2,]   1   10
```

8. Graphics

- Some operations

```
> x <- rnorm(10)
> y <- rnorm(10)
> plot(x, y, xlab = “Ten random values”, ylab = “Ten random values”)
```



9.1 Last but not least

- Programming. Here again is our first function, loaded in the workspace

```
> ratiosm <- function(x) {
sd(x)/mean(x)
}
```

Now see the use of this function

```
> ratiosm(x)
[1] 445.7792
```

Remember that mean(x) was equal to 0.00239885, and sd(x) was equal to 1.043072.

- Another option is using the usual if, for, while, etc. But try to avoid them. The program becomes very slow.
- The apply function can solve a lot of problems, when trying to avoid the usual syntax.

```
> apply(m1,1,mean)
[1] 5.5 5.5
```

Again be careful. Try to use the available R functions, not loops and other programming syntax whenever possible. Your program will run faster.

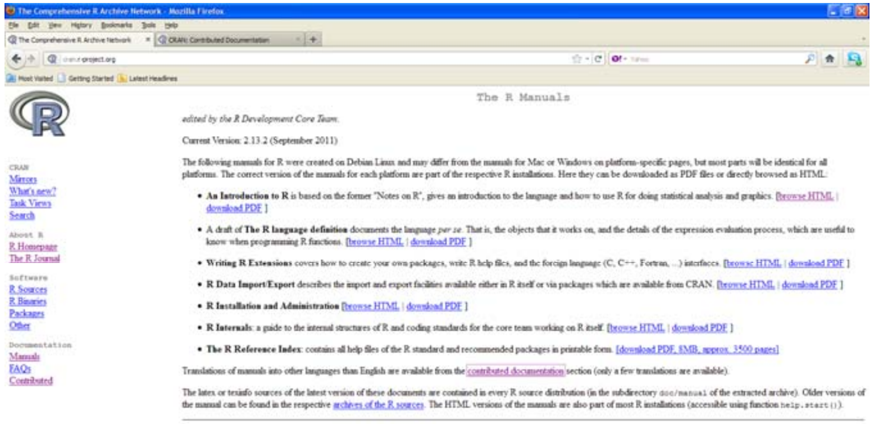
- Another possibility is the ability to interact with C, Fortran Matlab and other programs.

9.2 Last but not least

- Literature for R

There are a lot of manuals free to download. Some recommended for beginners

- An introduction to R. (Installed with the software)
- R for beginners (by Emmanuel Paradis)



The R Manuals

edited by the R Development Core Team.

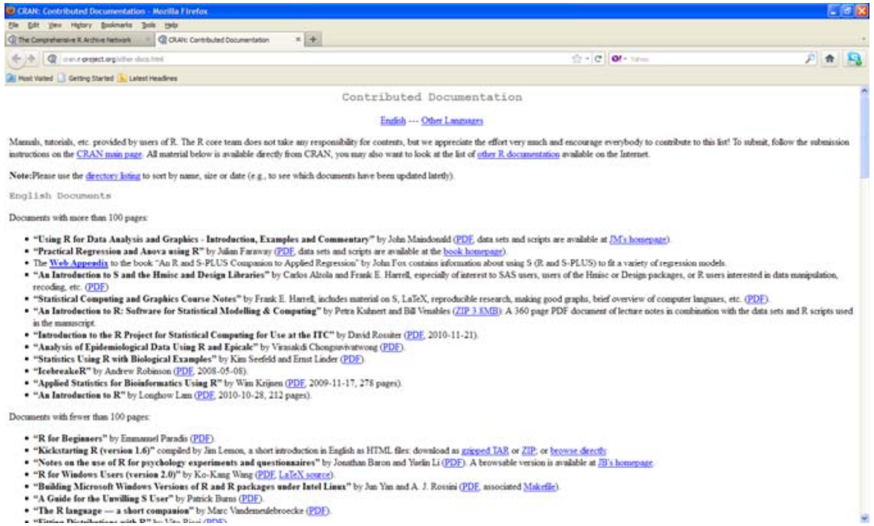
Current Version: 2.13.2 (September 2011)

The following manuals for R were created on Debian Linux and may differ from the manuals for Mac or Windows on platform-specific pages, but most parts will be identical for all platforms. The correct version of the manuals for each platform are part of the respective R installations. Here they can be downloaded as PDF files or directly browsed as HTML:

- **An Introduction to R** is based on the former "Notes on R", gives an introduction to the language and how to use R for doing statistical analysis and graphics. [browse HTML | download PDF]
- A draft of **The R language definition** documents the language *per se*. That is, the objects that it works on, and the details of the expression evaluation process, which are useful to know when programming R functions. [browse HTML | download PDF]
- **Writing R Extensions** covers how to create your own packages, write R help files, and the foreign language (C, C++, Fortran, ...) interfaces. [browse HTML | download PDF]
- **R Data Import/Export** describes the import and export facilities available either in R itself or via packages which are available from CRAN. [browse HTML | download PDF]
- **R Installation and Administration** [browse HTML | download PDF]
- **R Internals**: a guide to the internal structures of R and coding standards for the core team working on R itself. [browse HTML | download PDF]
- **The R Reference Index**: contains all help files of the R standard and recommended packages in printable form. [download PDF, 8MB, approx. 3500 pages]

Translations of manuals into other languages than English are available from the contributed documentation section (only a few translations are available).

The latex or texinfo sources of the latest version of these documents are contained in every R source distribution (in the subdirectory `doc/manual` of the extracted archive). Older versions of the manual can be found in the respective archives of the R sources. The HTML versions of the manuals are also part of most R installations (accessible using function `help.start()`).


CRAN: Contributed Documentation - Mozilla Firefox
File Edit View History Bookmarks Tools Help
The Comprehensive R Archive Network
CRAN: Contributed Documentation
cran.r-project.org/other-docs.html
Most Visited
Getting Started
Latest Headlines

Contributed Documentation

English --- Other Languages

Manuals, tutorials, etc. provided by users of R. The R core team does not take any responsibility for contents, but we appreciate the effort very much and encourage everybody to contribute to this list! To submit, follow the submission instructions on the CRAN main page. All material below is available directly from CRAN, you may also want to look at the list of other R documentation available on the Internet.

Note: Please use the directory listing to sort by name, size or date (e.g., to see which documents have been updated lately).

English Documents

Documents with more than 100 pages:

- **"Using R for Data Analysis and Graphics - Introduction, Examples and Commentary"** by John Maindonald (PDF, data sets and scripts are available at JM's homepage).
- **"Practical Regression and Anova using R"** by Julian Faraway (PDF, data sets and scripts are available at the book homepage).
- The **Web Appendix** to the book "An R and S-PLUS Companion to Applied Regression" by John Fox contains information about using S (R and S-PLUS) to fit a variety of regression models.
- **"An Introduction to S and the Hmisc and Design Libraries"** by Carlos Alzola and Frank E. Harrell, especially of interest to SAS users, users of the Hmisc or Design packages, or R users interested in data manipulation, recoding, etc. (PDF)
- **"Statistical Computing and Graphics Course Notes"** by Frank E. Harrell, includes material on S, LaTeX, reproducible research, making good graphs, brief overview of computer languaes, etc. (PDF).
- **"An Introduction to R: Software for Statistical Modelling & Computing"** by Petra Kuhnert and Bill Venables (ZIP 3.8MB): A 360 page PDF document of lecture notes in combination with the data sets and R scripts used in the manuscript.
- **"Introduction to the R Project for Statistical Computing for Use at the ITC"** by David Rossiter (PDF, 2010-11-21).
- **"Analysis of Epidemiological Data Using R and Epicalc"** by Virasakdi Chongsuvivatwong (PDF).
- **"Statistics Using R with Biological Examples"** by Kim Seefeld and Ernst Linder (PDF).
- **"IcebreakeR"** by Andrew Robinson (PDF, 2008-05-08).
- **"Applied Statistics for Bioinformatics Using R"** by Wim Krijnen (PDF, 2009-11-17, 278 pages).
- **"An Introduction to R"** by Longhow Lam (PDF, 2010-10-28, 212 pages).

Documents with fewer than 100 pages:

- **"R for Beginners"** by Emmanuel Paradis (PDF).
- **"Kickstarting R (version 1.6)"** compiled by Jim Lemon, a short introduction in English as HTML files: download as gzipped TAR or ZIP, or browse directly.
- **"Notes on the use of R for psychology experiments and questionnaires"** by Jonathan Baron and Yuelin Li (PDF). A browsable version is available at JB's homepage.
- **"R for Windows Users (version 2.0)"** by Ko-Kang Wang (PDF, LaTeX source).
- **"Building Microsoft Windows Versions of R and R packages under Intel Linux"** by Jun Yan and A. J. Rossini (PDF, associated Makefile).
- **"A Guide for the Unwilling S User"** by Patrick Burns (PDF).
- **"The R language — a short companion"** by Marc Vandemeulebroecke (PDF).
- **"Fitting Distributions with R"** by Vito Ricci (PDF).

The R-manuals

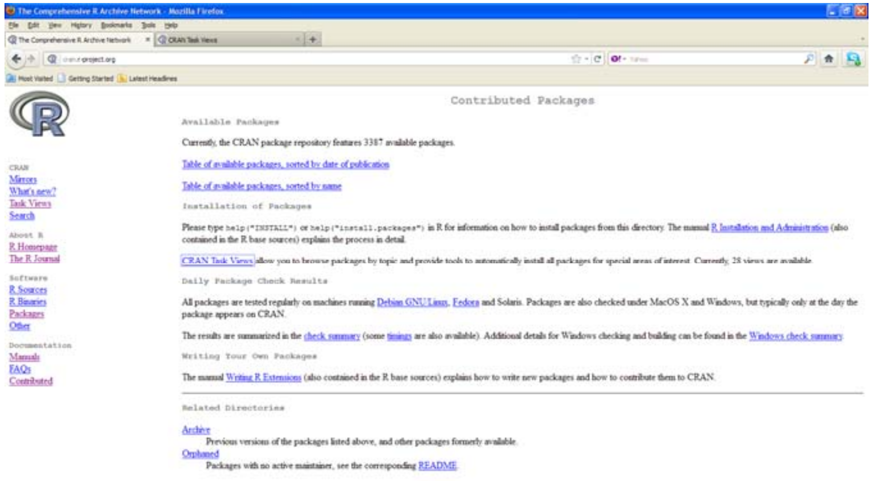
Contributed documenation

9.3 Last but not least

- R-packages

Some basic packages are already installed. Some of them are loaded when the software starts. For the others you have to load them yourself if you need them.

Besides the basic packages, the CRAN package repository additionaly features approximately 3500 packages, sorted by name, task, date of release or etc. If you need a package, you can install it for free. Other repositories exist also.



Contributed Packages

Available Packages

Currently, the CRAN package repository features 3387 available packages.

Table of available packages, sorted by date of publication

Table of available packages, sorted by name

Installation of Packages

Please type `help("INSTALL")` or `help("install.packages")` in R for information on how to install packages from this directory. The manual R Installation and Administration (also contained in the R base sources) explains the process in detail.

CRAN Task Views allow you to browse packages by topic and provide tools to automatically install all packages for special areas of interest. Currently, 28 views are available.

Daily Package Check Results

All packages are tested regularly on machines running Debian GNU/Linux, Fedora and Solaris. Packages are also checked under MacOS X and Windows, but typically only at the day the package appears on CRAN.

The results are summarized in the check summary (some timings are also available). Additional details for Windows checking and building can be found in the Windows check summary.

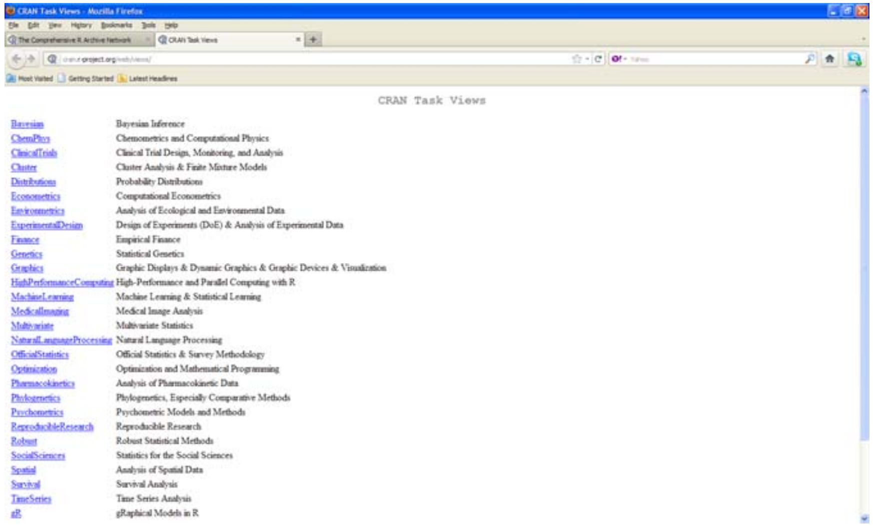
Writing Your Own Packages

The manual Writing R Extensions (also contained in the R base sources) explains how to write new packages and how to contribute them to CRAN.

Related Directories

Archive
Previous versions of the packages listed above, and other packages formerly available.
Orphaned
Packages with no active maintainer, see the corresponding README.

Contributed packages


CRAN Task Views - Mozilla Firefox
File Edit View History Bookmarks Tools Help
The Comprehensive R Archive Network
CRAN Task Views
cran.r-project.org/web/views/
Most Visited
Getting Started
Latest Headlines

CRAN Task Views

Bayesian	Bayesian Inference
ChemPhys	Chemometrics and Computational Physics
ClinicalTrials	Clinical Trial Design, Monitoring, and Analysis
Cluster	Cluster Analysis & Finite Mixture Models
Distributions	Probability Distributions
Econometrics	Computational Econometrics
Environmetrics	Analysis of Ecological and Environmental Data
ExperimentalDesign	Design of Experiments (DoE) & Analysis of Experimental Data
Finance	Empirical Finance
Genetics	Statistical Genetics
Graphics	Graphic Displays & Dynamic Graphics & Graphic Devices & Visualization
HighPerformanceComputing	High-Performance and Parallel Computing with R
MachineLearning	Machine Learning & Statistical Learning
MedicalImaging	Medical Image Analysis
Multivariate	Multivariate Statistics
NaturalLanguageProcessing	Natural Language Processing
OfficialStatistics	Official Statistics & Survey Methodology
Optimization	Optimization and Mathematical Programming
Pharmacokinetics	Analysis of Pharmacokinetic Data
Phylogenetics	Phylogenetics, Especially Comparative Methods
Psychometrics	Psychometric Models and Methods
ReproducibleResearch	Reproducible Research
Robust	Robust Statistical Methods
SocialSciences	Statistics for the Social Sciences
Spatial	Analysis of Spatial Data
Survival	Survival Analysis
TimeSeries	Time Series Analysis
gR	gRaphical Models in R

CRAN Task Views

10. Bibliography

- An introduction to R, (Installed with the software)
- R for beginners, by Emmanuel Paradis
- The friendly beginners’ R course, by Toby Marthews

Σας ευχαριστώ

ขอบคุณ