



Whats' beyond Concerto: An introduction to the R package *catR*

Session 2:

Overview of R software and *catR* package

The Psychometrics Centre, Cambridge, June 10th, 2014

Outline:

1. The R software
2. R GUI and RStudio
3. R language
4. R packages

1. The R software

R was created in 1997

Similar to the S language (used in e.g., S-PLUS)...

... but is **open source** and freely accessible

Network of developpers and researchers in R

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

Specific meetings and conferences (**useR!**)

Regular **updates** and personnal improvements:

- first version R-0.49 released April 23, 1997
- last version R-3.1.0 released April 10, 2014

Increasing litterature about R and its applications

1. The R software

R can handle a variety of tasks:

- Descriptive statistics
- Basic and advanced figures
- Basic and advanced statistical methods (ANOVA, PCA, factor analysis, GLM, bootstrap, neural networks, time series, ...)
- Easy data generation processes and design of simulation studies
- Interaction with other statistical software (e.g. WinBUGS)
- Easy loading and management of data sets from Excel, SAS, SPSS, ...
- Personal implementation and optimization possible (through e.g. packages)

1. The R software

Not all these features will be looked at today!

Some useful **references** and reading:

- Spector, P. (2008). *Data manipulation with R*. New York: Springer.
- Zuur, A. F., Ieno, E. N., & Meesters, E. H. W. G. (2009). *A beginner's guide to R*. New York: Springer.
- ... and many other nice books from the **UseR!** series

Check also my slides from former workshop on R (Trier, Germany, 2013):

<http://hdl.handle.net/2268/147096>

2. R GUI and RStudio

If you install R from the **CRAN** (Comprehensive R Archive Network) website:

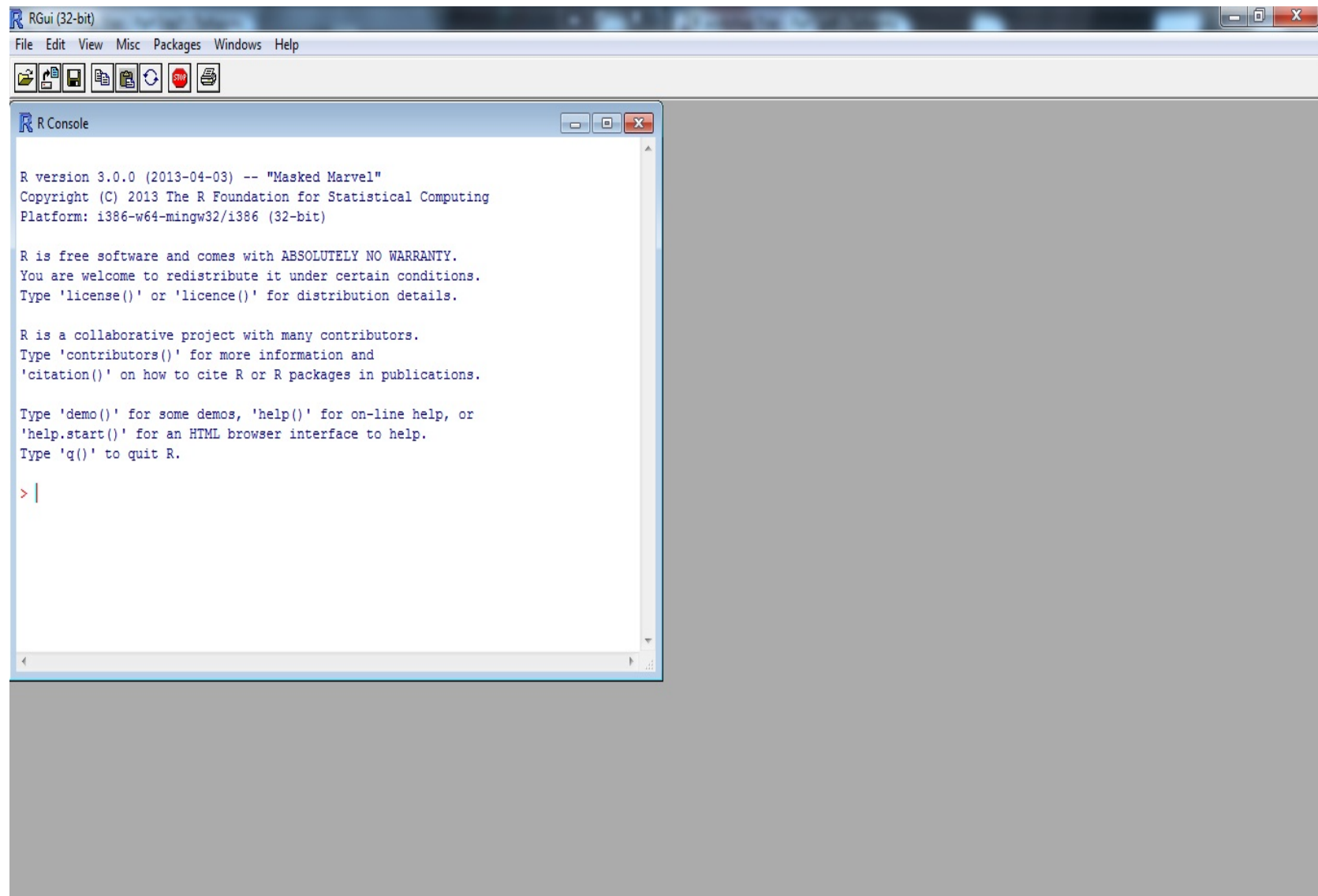
<http://www.r-project.org>

then you will install the **GUI** (graphical user interface) of R

R GUI is the basic interface to program in R...

... so it is not necessarily the most fancy interface :-)

2. R GUI and RStudio



2. R GUI and RStudio

Fortunately (for you) it exists a more user-friendly interface: the **RStudio** software

RStudio can be downloaded and installed freely from

<https://www.rstudio.com/>

RStudio works similarly to R GUI but optimizes windows and display information in a much more enhanced way

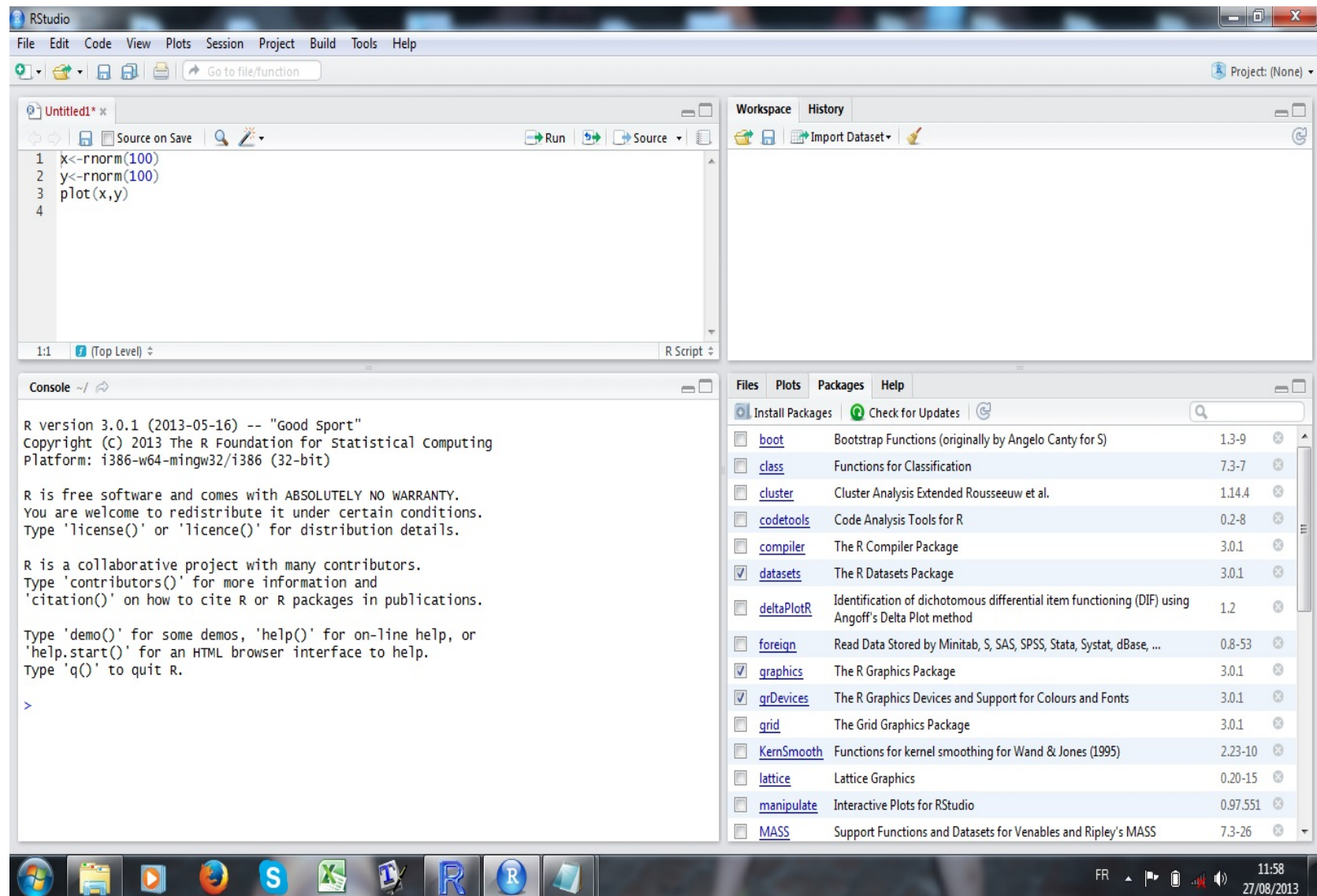
RStudio definitely to be preferred for non-crazy R users...

... though I will illustrate *catR* with R GUI :-)

2. R GUI and RStudio



2. R GUI and RStudio



3. R language

R has a specific **syntax** and all instructions written in the console will be executed in a **line-by-line** approach

Errors will be mentioned and execution will stop

Basic example: let's write "3+2" in the console:

```
R> 3+2
```

and execute it by pressing ENTER

(Important: **R>** must NOT be written in the console!)

Let's see how R reacts...

3. R language

The following appears in the console:

```
> 3+2  
[1] 5  
> |
```

Surprisingly, R computed $3+2$ and returned value 5 :-)

Preceding symbol `[1]` helps in identifying the component number when a vector is returned

3. R language

Now, let's execute "3+a" in the console

```
R> 3+a
```

Output is

```
> 3+a  
Error: object 'a' not found  
> |
```

Since object **a** was not yet created in R, it doesn't exist and R detects an error

Error **message** is returned, process stops and **nothing** more is done

3. R language

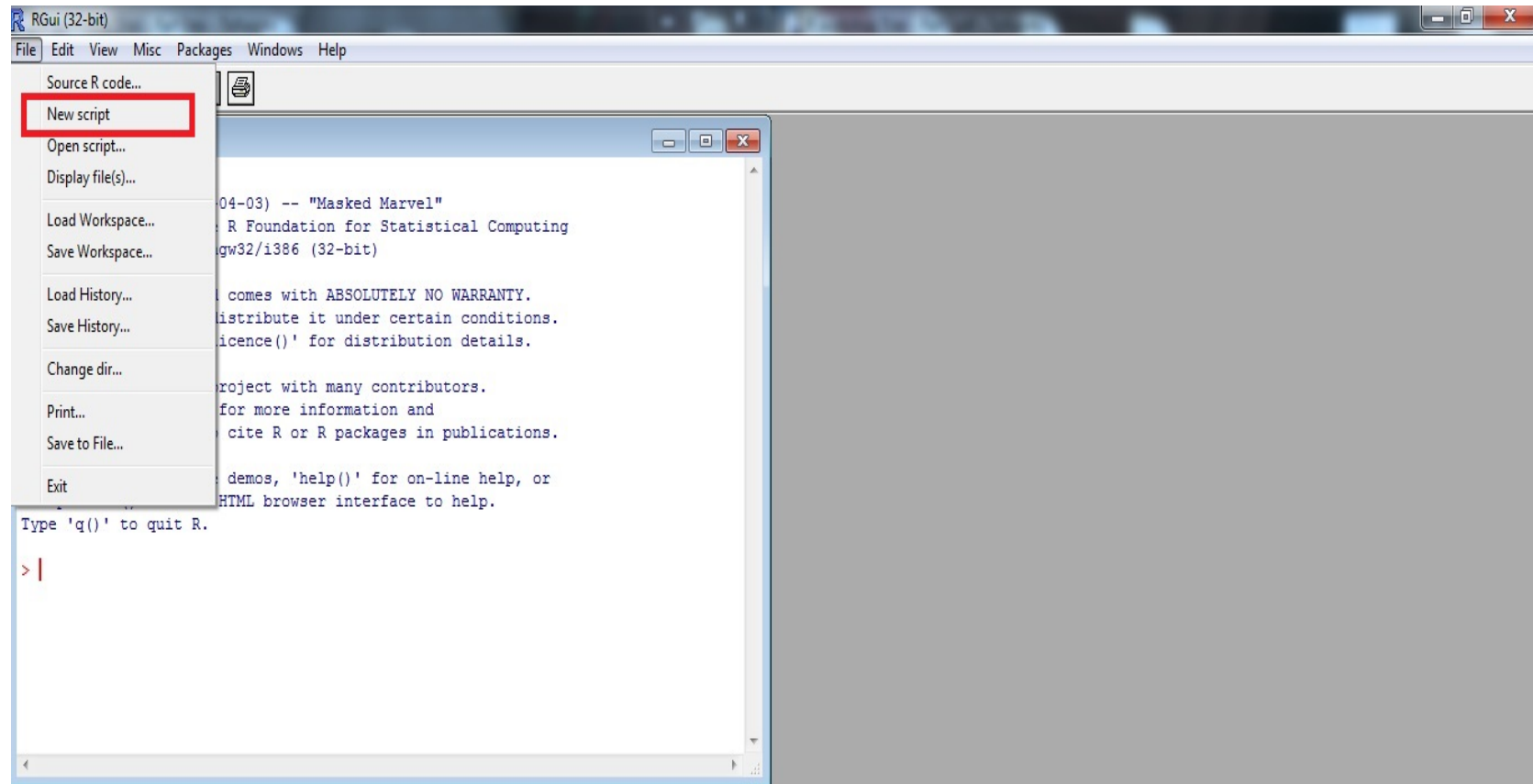
Most often, analyses in R require several successive steps of calculation

One mistake stops the whole process and all steps must be written once again...

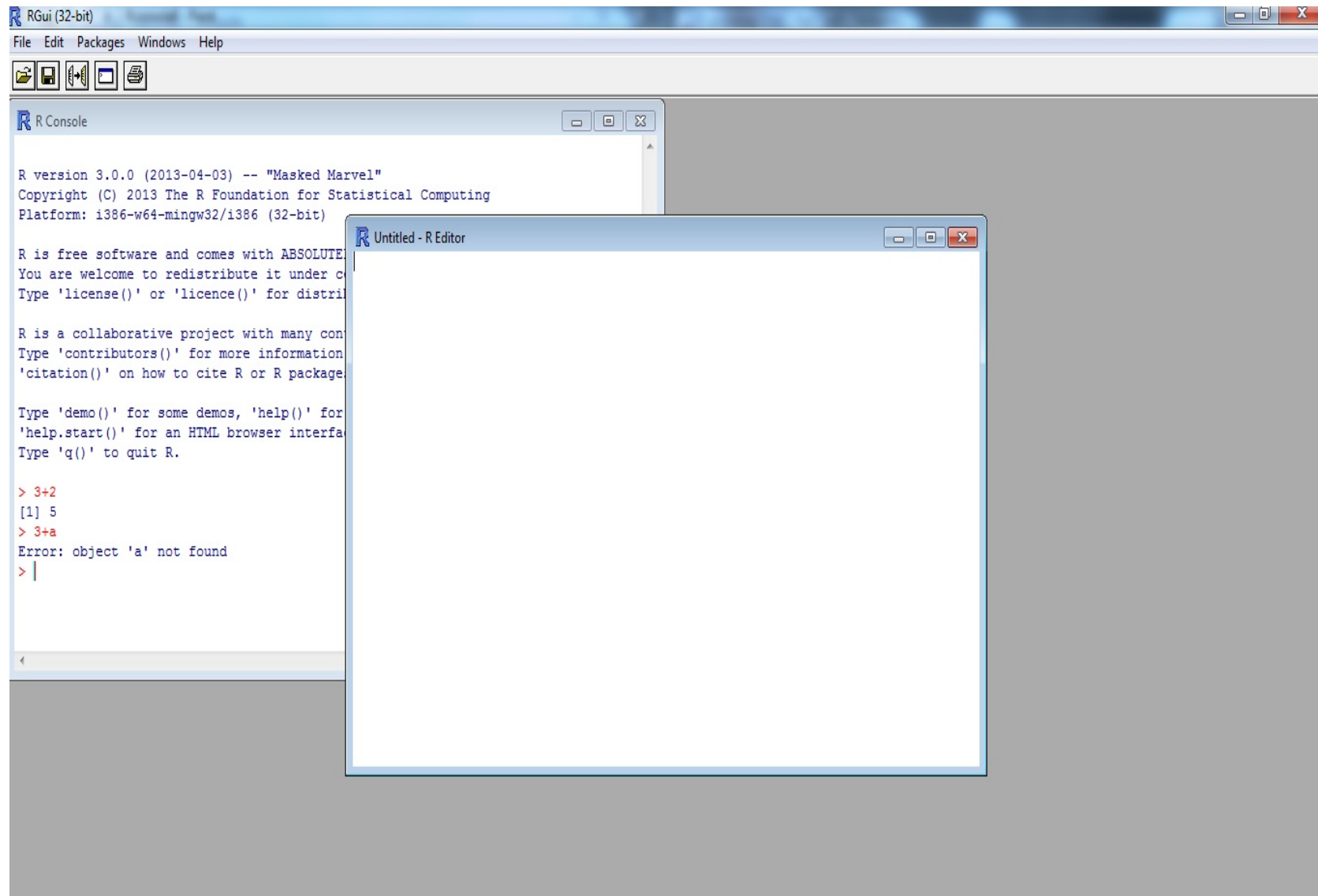
More efficient to write all R code in a **script file** first, then copy-paste (or execute it) in the console

Any text editing software (Word, Notepad etc.) is convenient, but R has an internal script editor (and RStudio opens automatically an R script window)

3. R language



3. R language



3. R language

Within the script window, any R code can be written

Nothing is being executed at once in the console

To **execute** some R code written in the script:

1. select the R code with the mouse
2. Click-right and select “Run line or selection”
(this step can be replaced by **CTRL+R**)

3. R language

R language is based on several elements:

- predefined **functions** that are available internally
- additional functions available in **packages** (from CRAN or elsewhere)
- internal **variables** (to store data or output)
- **vectors** and **matrices**
- **data frames** and **lists**

3. R language

Functions are called by their **name**, and one or several **arguments** must be specified (in parentheses)

Example internal functions are:

- basic calculations: `exp()`, `log()`, `abs()`, `sqrt()`, `cos()`, `sin()`, ...
- graphics: `plot()`, `hist()`, `pie()`, `boxplot()`, ...
- basic statistics: `mean()`, `sd()`, `var()`, `quantile()`, `cor()`, ...
- distributions: `dnorm()`, `dchisq()`, `dpois()`, `dbinom()`, `dt()`, ...
- standard tests: `t.test()`, `chisq.test()`, `ks.test()`, `wilcox.test()`, ...
- advanced statistical models and methods: `glm()`, `prcomp()`, `factanal()`, `anova()`, `ksmooth()`, ...

3. R language

Some functions may not be available by default (see later)

Before going on, two important remarks:

- R makes distinction between lowercase and uppercase letters: `abs()` works, but not `Abs()`!
- Each function in R (either internally or from external package) has a help file with (sometimes) useful information
 - can be accessed by typing first

?

directly followed by the function name, without ()

– for instance,

`R> ?anova`

opens the help file related to `anova()` function

3. R language

Internal variables may be created to store data and provisional or final results

Variable name is up to you (remember that lowercase and uppercase are not the same!)

Storing some value in a variable is done by means of the logical symbol

`<-`

Example: let's write

```
R> x <- 5 + 2
```

R will create an internal variable `x` with value 7 (to see it, write down `x` in the console)

3. R language

Matrices can also be created and handled in R

Matrices are arrays with rows and columns

Elements of a matrix identified by means of two indicators, one for row number, one for column number (in this order)

Consider a $I \times J$ matrix stored in the R variable **mat**

Element (i, j) of **mat** is obtained by

mat[i,j]

i -th row and j -th column of **mat** are obtained by

mat[i,] and **mat[,j]**

3. R language

Matrices should contain only **numeric** values

However, data sets might hold mix of numeric and character / qualitative variables (such as category names)

Handled in R by means of **data frames**

Data frames are **lists** of R objects of different formats, most often vectors of numeric or character values of the **same length**

Display in a “data set” format with column names

Each column is an element of the list (i.e. a vector) and the column name is the variable name

Hint: most often, data sets loaded in R have the data frame format (see later)

4. R packages

R holds a lot of **predefined** functions for basic statistics and graphics

However, R can also be used to develop specific routines... such as CAT applications

Easiest way to get stable and portable R code: creation of an **R package**

Packages hold several functions, with a full description (help file)

Most of them available from **CRAN**:

- 4465 packages on April 27th, 2013
- 5540 packages on May 14th, 2014
- ... how many on June 10th, 2014?

4. R packages

www.r-project.org

The R Project for Statistical Computing

PCA 5 vars
princomp(x = data, cor = cor)

Fertility
Catholic
Agriculture
Examination
Education
(1-3) 60%

Clustering 4 groups

Factor 1 [41%]
Factor 3 [19%]

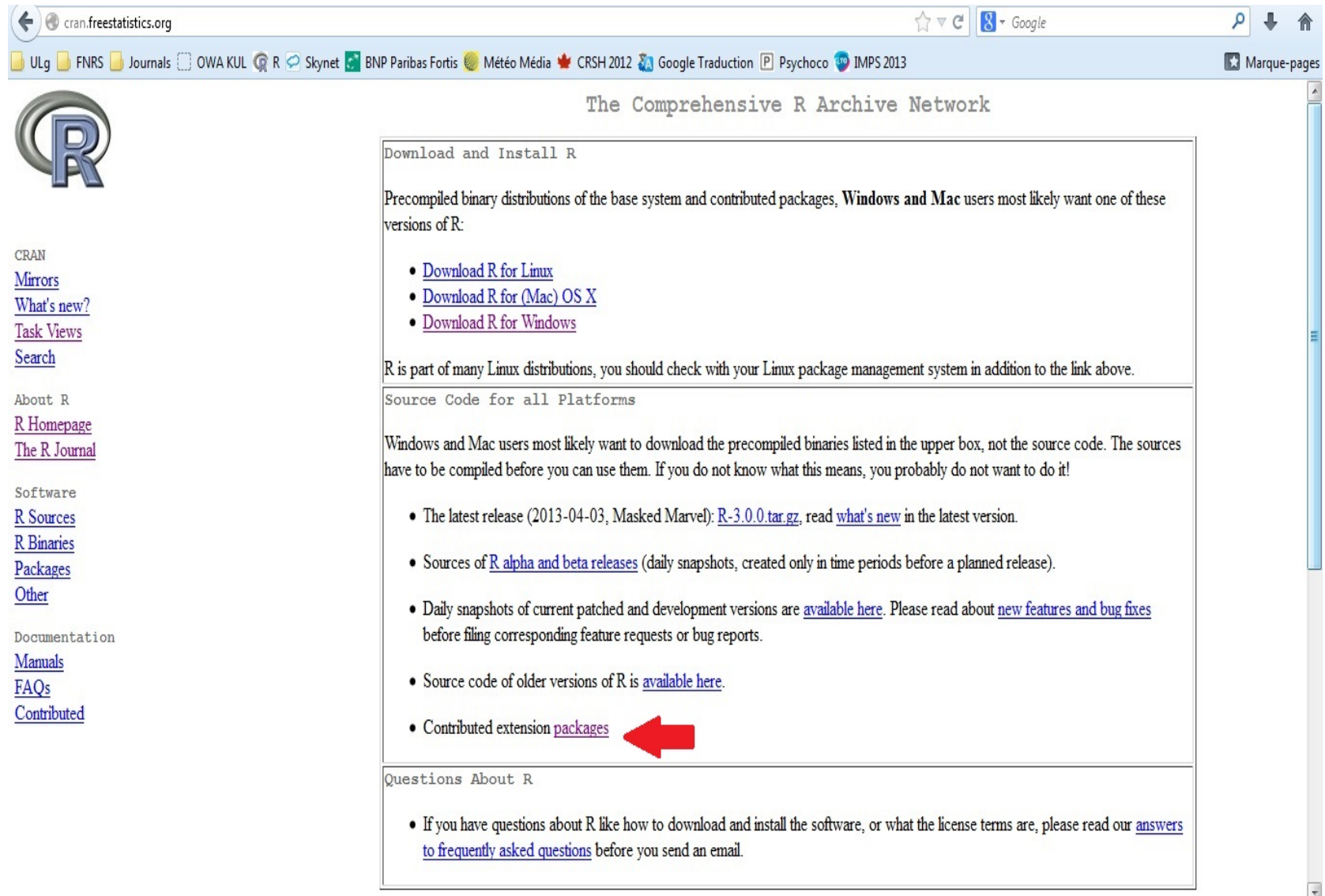
Groups
28
16
1
2

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 :

4. R packages



cran.freeststatistics.org

ULg FNRS Journals OWA KUL R Skynet BNP Paribas Fortis Météo Média CRSH 2012 Google Traduction Psychoco IMPS 2013 Marque-pages

The Comprehensive R Archive Network

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 \(Mac\) OS X](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

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 (2013-04-03, Masked Marvel): [R-3.0.0.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.

Left Sidebar:

CRAN

- [Mirrors](#)
- [What's new?](#)
- [Task Views](#)
- [Search](#)

About R

- [R Homepage](#)
- [The R Journal](#)

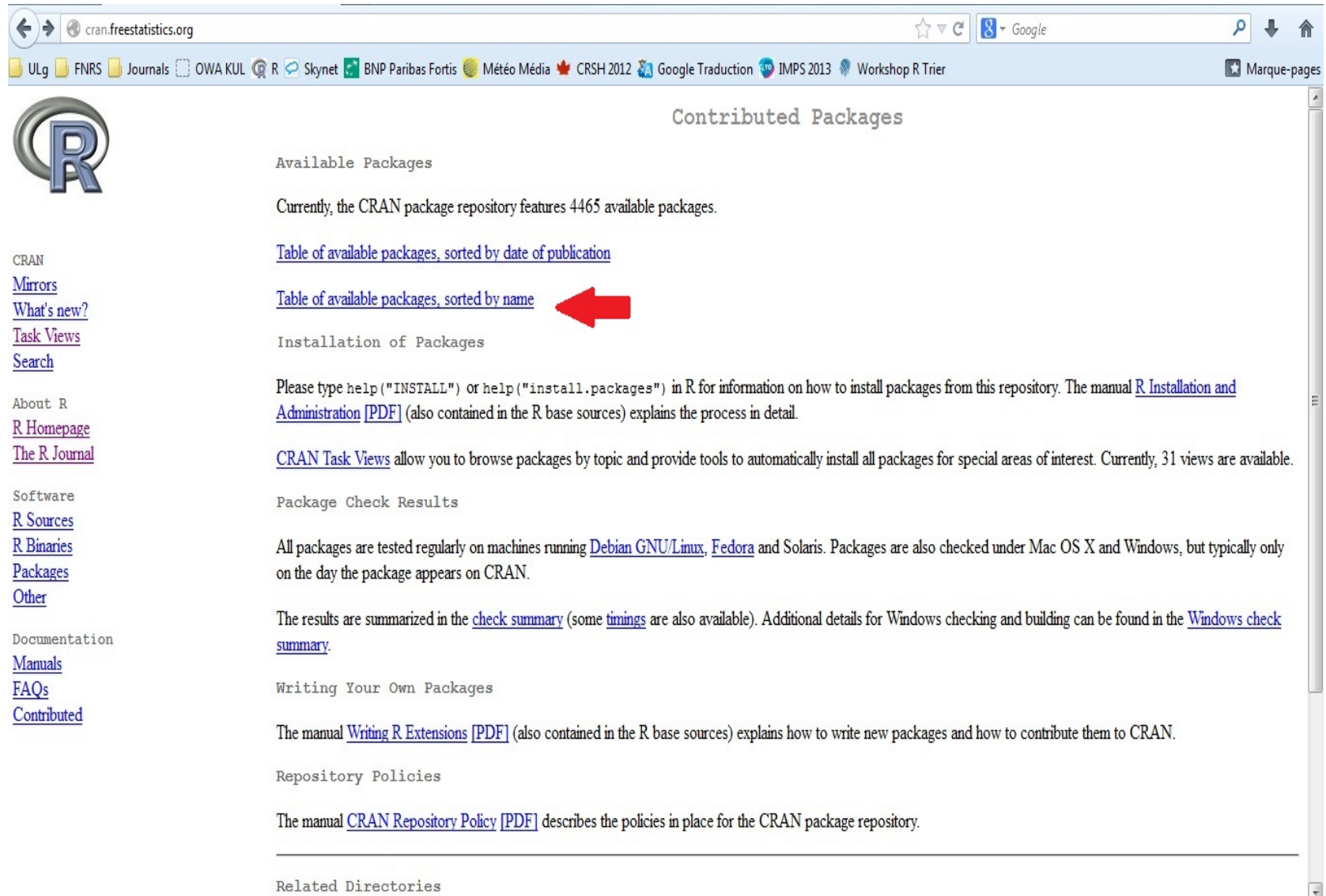
Software

- [R Sources](#)
- [R Binaries](#)
- [Packages](#)
- [Other](#)

Documentation

- [Manuals](#)
- [FAQs](#)
- [Contributed](#)

4. R packages



The screenshot shows the CRAN website interface. The browser's address bar displays 'cran.freeststatistics.org'. The page title is 'Contributed Packages'. The main content area includes sections for 'Available Packages' (stating 4465 packages are available), 'Installation of Packages' (providing instructions on how to install packages), 'Package Check Results' (describing the testing process on various operating systems), and 'Writing Your Own Packages' (explaining how to contribute new packages). A red arrow points to the link 'Table of available packages, sorted by name' under the 'Available Packages' section.

Contributed Packages

Available Packages

Currently, the CRAN package repository features 4465 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 repository. The manual [R Installation and Administration \[PDF\]](#) (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, 31 views are available.

Package Check Results

All packages are tested regularly on machines running [Debian GNU/Linux](#), [Fedora](#) and Solaris. Packages are also checked under Mac OS X and Windows, but typically only on 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 \[PDF\]](#) (also contained in the R base sources) explains how to write new packages and how to contribute them to CRAN.

Repository Policies

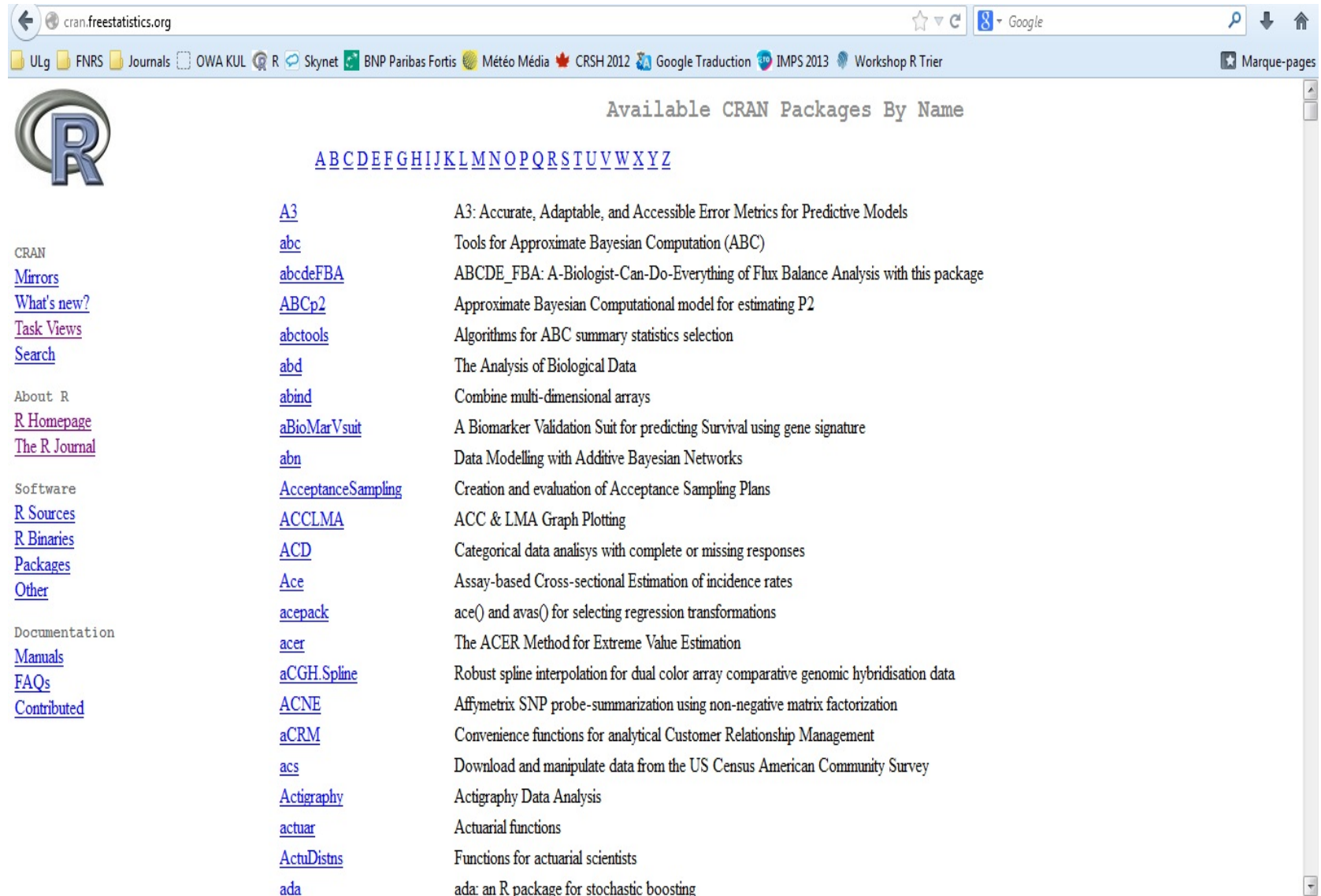
The manual [CRAN Repository Policy \[PDF\]](#) describes the policies in place for the CRAN package repository.

Related Directories

Left Sidebar:

- 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](#)

4. R packages



The screenshot shows the CRAN website (cran.freeststatistics.org) with the R logo on the left. The main content area is titled "Available CRAN Packages By Name" and lists packages alphabetically from A to Z. The left sidebar contains navigation links for CRAN, About R, Software, and Documentation.

Package Name	Description
A3	A3: Accurate, Adaptable, and Accessible Error Metrics for Predictive Models
abc	Tools for Approximate Bayesian Computation (ABC)
abcdeFBA	ABCDE_FBA: A-Biologist-Can-Do-Everything of Flux Balance Analysis with this package
ABCp2	Approximate Bayesian Computational model for estimating P2
abctools	Algorithms for ABC summary statistics selection
abd	The Analysis of Biological Data
abind	Combine multi-dimensional arrays
aBio.MarVsuit	A Biomarker Validation Suit for predicting Survival using gene signature
abn	Data Modelling with Additive Bayesian Networks
AcceptanceSampling	Creation and evaluation of Acceptance Sampling Plans
ACCLMA	ACC & LMA Graph Plotting
ACD	Categorical data analysis with complete or missing responses
Ace	Assay-based Cross-sectional Estimation of incidence rates
acepack	ace() and avas() for selecting regression transformations
acer	The ACER Method for Extreme Value Estimation
aCGH.Spline	Robust spline interpolation for dual color array comparative genomic hybridisation data
ACNE	Affymetrix SNP probe-summarization using non-negative matrix factorization
aCRM	Convenience functions for analytical Customer Relationship Management
acs	Download and manipulate data from the US Census American Community Survey
Actigraphy	Actigraphy Data Analysis
actuar	Actuarial functions
ActuDistns	Functions for actuarial scientists
ada	ada: an R package for stochastic boosting

Left Sidebar Navigation:

- 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](#)

4. R packages

Best way to install an R package: make use of “Install” menu in R GUI / RStudio

Packages must be installed once and loaded with each new session of R

Packages must be re-installed after each update of R itself

To load an R package (for instance *catR*):

```
R> require(catR)
```

This allows R to access to all functions within *catR* package