



The R Project for Statistical Computing



What is R?

“R is a free software environment for statistical computing and graphics.”

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

- Software environment
- Statistical computing
- Graphics

What is R?

- **Software environment**

- Statistical Computing
- Graphics

An interpreted language.

Provides control structures (loops).

Interface with other languages (e.g., C, fortran).

What is R?

- Software environment
- Statistical Computing**
- Graphics

Univariate methods: ANOVA, Linear regression, etc.

Multivariate methods: PCA, Clustering, Multiple regression.

Bayesian tools (MCMC).

Distributions: `rnorm()`, `rbinom()`, etc.

Simulations: Coupling distributions with control loops.

What is R?

- Software environment
- Statistical Computing
- Graphics**

Scatterplots.
Histograms.
Density functions.
Box and whisker plots.
Maps.
3D surfaces.
3D scatterplots.
Heatmaps.

Go to R
and web.

R Background

R is an open source, free version (dialect) of S.

S-plus (www.insightful.com) is a commercial version of S with a GUI (Graphical User Interface).

R “is a GNU project which is similar to the S language and environment which was developed at Bell Laboratories (formerly AT&T, now Lucent Technologies) by John Chambers and colleagues.”



GNU's Not UNIX. The GNU Project was launched in 1984 to develop a complete Unix-like operating system which is free software.

What is R?

R is an interpreted language (e.g., it is compiled on the fly).

Because it is interpreted it is relatively user friendly:

- commands can be sent per line
- commands can be sent as part of a line:
 - Windows GUI text editor.
 - ESS – Emacs Speaks Statistics (UNIX).

This allows for 'trial and error' debugging of code.

Rcmdr – R Commander, a GUI for R!

The screenshot shows the Rcmdr interface with the following components:

- Menu Bar:** File, Edit, Data, Statistics, Graphs, Models, Distributions, Tools, Help
- Data set:** iris (selected)
- Model:** <No active model>
- Script Window:**

```
data(iris, package="datasets")
.Anova <- lm(Petal.Length ~ Species, data=iris)
anova(.Anova)
tapply(iris$Petal.Length, iris$Species, mean, na.rm=TRUE) # means
tapply(iris$Petal.Length, iris$Species, sd, na.rm=TRUE) # std. deviations
tapply(iris$Petal.Length, iris$Species, function(x) sum(!is.na(x))) # counts
remove(.Anova)
```
- Output Window:**

Analysis of Variance Table

Response: Petal.Length

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Species	2	437.10	218.55	1180.2	< 2.2e-16 ***
Residuals	147	27.22	0.19		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
> tapply(iris$Petal.Length, iris$Species, mean, na.rm=TRUE) # means
  setosa versicolor  virginica
  1.462     4.260     5.552

> tapply(iris$Petal.Length, iris$Species, sd, na.rm=TRUE) # std. deviations
  setosa versicolor  virginica
0.1736640 0.4699110 0.5518947

> tapply(iris$Petal.Length, iris$Species, function(x) sum(!is.na(x))) # counts
  setosa versicolor  virginica
    50         50         50
```
- Messages:** (Empty)

R Syntax

R function

Parameters for the function
included in parenthesis.

```
> my.data <- read.table(file="data.txt",  
+                        header=TRUE, sep="\t")
```

Try:
> help(read.table)

R Syntax

**'Object' in which to
store your data**

```
> my.data <- read.table(file="data.txt",  
+                        header=TRUE, sep="\t")
```

**Assignment
operator**

(also see <-; ->; =)

Working directories and slashes

R operates within a 'working directory.'

```
try getwd(), setwd()
```

Unix uses forward slashes '/' therefore R uses forward slashes.

Windows uses backward slashes '\' and must be changed.

R Data Structures

vector - `c(2, 9, 1000, 546, 1, 1, 45)`

matrix – 8 X 7 (numbers only)

data.frame – A matrix that can include non-numbers (e.g., factors)

list – A vector of vectors.

- Most R functions return a list
- Packages that define their own data structures usually define special instances of lists.

Indexing R Data Structures

`my.vector[3]`

Indexes the third element
in the vector.

`my.data.frame[3, 5]`

Indexes the element in
third row and the fifth
column.

`my.list[[4]][5, 8]`

Indexes the fourth
element in the list (here
its a data.frame) and
Indexes the element in
fifth row and the eighth
column.

Functions that create vectors

- > `3:8` #returns (3, 4, 5, 6, 7, 8)
- > `rep(3, times = 4)` #returns (3, 3, 3, 3)
- > `seq(from = 1, to = 2, by = 0.2)`
#returns (1, 1.2, 1.4, 1.6, 1.8, 2.0)

Useful for indexing:

```
> my.data.frame[3:5, 18:22]
```

Indexes elements in rows 3-5 and in columns 18-22.

R Help System

R help is largely command based.

If you don't know the command. . .

`help(command)` – man page style
documentation.

- Searches active libraries only.

`?command` is a synonym for `help(command)`

See Also: section has related commands.

package vignettes are process oriented –
not always present.

`help.start()`

Go to iris example

R Graphics output

In Windows you can copy and paste .emf & .bmp.

```
win.metafile() # Windows specific.
```

```
?Devices
```

```
jpeg() and png()  
postscript()  
pdf()
```

```
dev.cur()
```

```
dev.off() # Turn off the device when done
```

Cairo – R graphics device which includes .tif

R Performance

R is an interpreted language (e.g., it is compiled on the fly).

This makes R inherently slow compared to compiled languages (such as C, C++, and Fortran).

Loops (e.g., 'for') are computationally intensive.

'apply' family of functions:

```
sapply(), lapply(), apply(), mapply(),  
tapply(), rapply()
```

execute a function repeatedly.

Utilize compiled code:

```
?C & ?Fortran
```

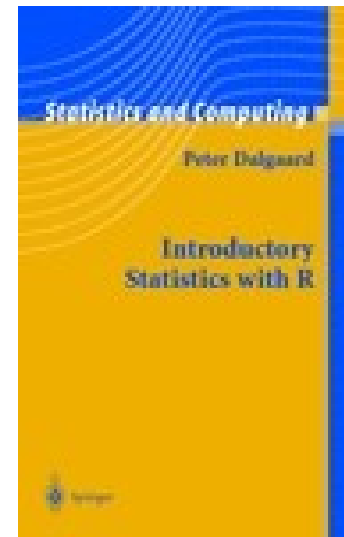
R Support

Get a book (Dalgaard, Peter. 2002. Introductory Statistics with R. Springer)

Google your question – google searches the wikis

Ask a friend!

Join an R user group!



R Packages

More tools!

ade4	# Analysis of ecological data
ape	# Analysis of Phylogenetics and Evolution
Bioconductor	# Bioinformatics (non-CRAN)
Biodiversity	# Ecological community analysis
boot	# Bootstrap functions
bqtl	# Bayesian QTL mapping
Cairo	# R graphics device (supports .tif)
climatol	# Tools for Climatology
colorRamps	# Builds color tables.
Geneland	# Landscape genetics
geometry	# Mesh generation, tessellation
HSAUR	# Handbook of Stat Analysis Using R
kernelPOP	# Spatially explicit population genetics
maps	# Draw Geographic Maps
Rcmdr	# Basic GUI

R Packages

More tools!

R2WinBUGS # Run WinBUGS from R
rJava # Low-level R to Java interface
RMySQL # R interface to the MySQL database
ROracle # Oracle database interface to R
SASxport # Read and write SAS XPORT files
seqinr # Biological sequences retrieval & analysis
shapefiles # Read and write ESRI Shapefiles
spgrass6 # Interface between GRASS GIS and R
spgwr # Geographically weighted regression
vegan # Community Ecology Package

R Resources

R-project for statistical computing
<http://www.r-project.org/>

Bioconductor (microarray stuff):
<http://www.bioconductor.org/>

The R Graphics Gallery:
<http://addictedtor.free.fr/graphiques/>

Emacs Speaks Statistics:
<http://ess.r-project.org/>

