R

Christiana Kartsonaki

24 October 2016



Figure: Analytics tools used by respondents to the 2015 Rexer Analytics Survey. Each respondent was free to select multiple tools.



Figure: Analytics tools used by respondents to the 2015 Rexer Analytics Survey. Each respondent was free to select multiple tools. Figure: Number of people who follow each software on LinkedIn and Quora.

Source: http://r4stats.com/articles/popularity/

R is freely available from http://cran.r-project.org/.

R is freely available from http://cran.r-project.org/.

Introduction

and the second s		
C R https://cran.r-project.org/	, D → 🚔 C 🕼 The Comprehensive R Arch ×	
CRAN Mirrors What i new2	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	
Search Search	Download R for Windows R is part of many Linux distributions, you should check with your Linux package management	
About R R Homenage	system in addition to the link above.	
The R Journal	Source Code for all Platforms	
Software R. Sinaries Packages Other	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 (Tuesday 2016-06-21, Bug in Your Hair) <u>R-3.3.1 tar.gz</u>, read <u>what's new</u> in the latest version. 	
Documentation Manuals FAOs	 Sources of <u>R alpha and beta releases</u> (daily snapshots, created only in time periods before a planned release). 	
Contributed	 Daily snapshots of current patched and development versions are <u>available here</u>. Please read about <u>new features and bug fixes</u> before filing corresponding feature requests or bug reports. 	
	Source code of older versions of R is <u>available here</u> .	
	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 <u>answers to frequently asked questions</u> before you send an email. 	
What are R and CRAN?		
R is 'GNUS' a freehr available language and environment for statistical computing and graphics which provides a wide \$100% •		

Can use an external text editor (copy and paste code in R console to run it)

- syntax highlighting
- code safe from crashes of R

Alternatively (Windows/Mac)

- text editor of R
- RStudio

Comment your code.

this is a comment

The help file for a particular command or package opens by typing

?function_name

A package can be installed from R by typing

```
install.packages("package_name")
```

Files can be read into R in many formats, the most common ones being .txt and .csv.

The commands with some of their default arguments are

```
read.table(file, header = FALSE, sep = "", row.
    names, col.names)
read.csv(file, header = TRUE, sep = ",", quote="
    \"")
```

The help file (opens by typing ?read.table) contains all the details on these commands.

Datasets from other statistical packages can also be imported in R, see the R Import/Export Manual¹.

¹https://cran.r-project.org/doc/manuals/R-data.html

Working directory

The path of the filename should be typed if it is not in the working directory, e.g.

```
dataset1 <- read.csv("path/filename.csv")</pre>
```

or

dataset1 <- read.table("path/filename.txt")</pre>

Working directory

The path of the filename should be typed if it is not in the working directory, e.g.

```
dataset1 <- read.csv("path/filename.csv")</pre>
```

or

```
dataset1 <- read.table("path/filename.txt")</pre>
```

The working directory can be seen by

getwd()

and set by

```
setwd("path_to_working_directory")
```

In Windows the working directory can be changed from File -> Change dir so that the path does not need to be typed in.

Once the dataset has been read into R, it can easily be modified.

To check as which of the main data types of R the dataset is stored, type class(dataset1)

Once the dataset has been read into R, it can easily be modified.

To check as which of the main data types of R the dataset is stored, type class(dataset1)

If not a dataframe, it can be converted into a dataframe (wherever possible) using

dataset1 <- as.data.frame(dataset1)</pre>

If one wants to keep only columns 1, 2 and 5,

dataset1 <- dataset1[, c(1, 2, 5)]</pre>

If one wants to keep only columns 1, 2 and 5,

dataset1 <- dataset1[, c(1, 2, 5)]

Similarly for keeping only rows from 1 to 50 (inclusive) and from 100 to 150, say,

dataset1 <- dataset1[c(1:50, 100:150),]</pre>

Data manipulation – continued

To see the row names of a dataframe, type

```
row.names(dataset1)
```

Data manipulation – continued

To see the row names of a dataframe, type

```
row.names(dataset1)
```

To change the row names of a dataframe, assign the required value to the corresponding row, or a sequence of values of length equal to the number of rows. For example, if we want the 5th row to have row name 5, we use

```
row.names(dataset1)[5] <- 5</pre>
```

The row names could also be strings, in which case they should be typed in quotation marks, as below.

```
row.names(dataset1)[5] <- "subject 5"</pre>
```

Data manipulation - continued

To see the column names of a dataframe, type

```
names(dataset1)
```

Data manipulation – continued

To see the column names of a dataframe, type

```
names(dataset1)
```

To change the column names of a dataframe, assign the required value to the corresponding column, or a sequence of values of length equal to the number of rows. For example,

```
names(dataset1)[2] <- "sleep_duration"</pre>
```

will give the label sleep_duration to the 2nd column.

Then columns can be indexed by their names instead of their number, if preferred:

```
dataset1[, "sleep_duration"]
```

will be the same as

dataset1[, 2]

The same works for row names, but to select row labelled as 5, the 5 should be in quotation marks; otherwise to select the 5th row, it should be without quotation marks.

To find out the number of rows and columns a dataframe contains, use dim(dataset1)

To use the number of rows as an argument to some other function, you can give the argument as dim(dataset1)[1]. Similarly, for columns it is dim(dataset1)[2].

save(dataset1, file = "data1.RData")

(Such files can be opened in R using load.)

save(dataset1, file = "data1.RData")

(Such files can be opened in R using load.)

Some R objects can also be exported as csv, txt and other file types.

save(dataset1, file = "data1.RData")

(Such files can be opened in R using load.)

Some R objects can also be exported as csv, txt and other file types.

There are also other options for writing objects of other types to text files.

save(dataset1, file = "data1.RData")

(Such files can be opened in R using load.)

Some R objects can also be exported as csv, txt and other file types.

There are also other options for writing objects of other types to text files.

jpeg, pdf and other similar functions are useful for exporting plots. To close the image file, use dev.off().

Types of variables

Basic variable types in R:

- integer
- numeric (double)
- character (string)
- logical (TRUE/FALSE)
- factor (categorical) [also ordered factor]

Types of variables

Basic variable types in R:

- integer
- numeric (double)
- character (string)
- logical (TRUE/FALSE)
- factor (categorical) [also ordered factor]

NA: missing Inf: infinite NaN: usually a number that cannot be defined

Types of objects

R object classes:

- vector
- data frame
- matrix
- list
- table
- array
- ...

Other useful functions

- ls() lists the names of all the objects in your workspace.
- str(object_name) gives some information about the object object_name.
- object_name prints the object object_name.
- rm(object_name) removes the object object_name from your workspace.

Other useful functions

- ls() lists the names of all the objects in your workspace.
- str(object_name) gives some information about the object object_name.
- object_name prints the object object_name.
- rm(object_name) removes the object object_name from your workspace.
- c combines objects into a vector.
- cbind combines columns into a matrix.
- rbind combines rows into a matrix.

Arithmetic operations

+, -, /, *

Arithmetic operations

 $(x^y = x^y)$

Arithmetic operations

$$(x^y = x^y)$$

%% modulo reduction %/% integer division %*% matrix multiplication

Logical operations

== (equal) != (not equal)

Logical operations

- == (equal)
- != (not equal)
- >, <, >=, <=

Logical operations

- == (equal) != (not equal)
- >, <, >=, <=
- ! (not)
- | (or)
- || (or)
- & (and)
- && (and)
- |, &: vectors
 ||, &&: only consider one element

Use R.

"Neurons that fire together, wire together."

Use R.

"Neurons that fire together, wire together."

Later:

Make your own R package.

References



R Development Core Team,

https://cran.r-project.org/manuals.html



📎 Venables, W. N. and Ripley, B. D. (2002). Modern applied statistics with S-PLUS. 4th edition. Springer Science and Business Media.

Ruth Ripley's lectures,

http://portal.stats.ox.ac.uk/userdata/ruth/APTS2013/APTS.html



Venables, W. N., Smith D. M. and the R Core Team (2016). An Introduction to R. https://cran.r-project.org/doc/manuals/R-intro.pdf