

Alain F. Zuur • Elena N. Ieno
Erik H.W.G. Meesters

A Beginner's Guide to R

^c*n.* Springer

Contents

Preface.	vii
Acknowledgements.	ix
1 Introduction	1
1.1 What Is R?	1
1.2 Downloading and Installing R	2
1.3 An Initial Impression	4
1.4 Script Code	7
1.4.1 The Art of Programming	7
1.4.2 Documenting Script Code	8
1.5 Graphing Facilities in R	10
1.6 Editors	12
1.7 Help Files and Newsgroups	13
1.8 Packages	16
1.8.1 Packages Included with the Base Installation	16
1.8.2 Packages Not Included with the Base Installation	17
1.9 General Issues in R	19
1.9.1 Quitting R and Setting the Working Directory.	21
1.10 A History and a Literature Overview.	22
1.10.1 A Short-Historical Overview of R	22
1.10.2 Books on R and Books Using R	22
1.11 Using This Book	24
1.11.1 If You Are an Instructor.	25
1.11.2 If You Are an Interested Reader with Limited R Experience	25
1.11.3 If You Are an R Expert	25
1.11.4 If You Are Afraid of R	25
1.12 Citing R and Citing Packages.	26
1.13 Which R Functions Did We Learn?	27

2	Getting Data into R	29
2.1	First Steps in R	29
2.1.1	Typing in Small Datasets	29
2.1.2	Concatenating Data with the <code>c</code> Function	31
2.1.3	Combining Variables with the <code>c</code> , <code>cbind</code> , and <code>rbind</code> Functions	34
2.1.4	Combining Data with the <code>vector</code> Function*	39
2.1.5	Combining Data Using a Matrix*	39
2.1.6	Combining Data with the <code>data.frame</code> Function	42
2.1.7	Combining Data Using the <code>list</code> Function*	43
2.2	Importing Data	46
2.2.1	Importing Excel Data	47
2.2.2	Accessing Data from Other Statistical Packages**	51
2.2.3	Accessing a Database***	52
2.3	Which R Functions Did We Learn?	54
2.4	Exercises	54
3	Accessing Variables and Managing Subsets of Data	57
3.1	Accessing Variables from a Data Frame	57
3.1.1	The <code>str</code> Function	59
3.1.2	The Data Argument in a Function	60
3.1.3	The <code>\$</code> Sign	61
3.1.4	The <code>attach</code> Function	62
3.2	Accessing Subsets of Data	63
3.2.1	Sorting the Data	66
3.3	Combining Two Datasets with a Common Identifier	67
3.4	Exporting Data	69
3.5	Recoding Categorical Variables	71
3.6	Which R Functions Did We Learn?	74
3.7	Exercises	74
4	Simple Functions	77
4.1	The <code>tapply</code> Function	77
4.1.1	Calculating the Mean Per Transect	78
4.1.2	Calculating the Mean Per Transect More Efficiently	79
4.2	The <code>sapply</code> and <code>lapply</code> Functions	80
4.3	The <code>summary</code> Function	81
4.4	The <code>table</code> Function	82
4.5	Which R Functions Did We Learn?	84
4.6	Exercises	84
5	An Introduction to Basic Plotting Tools	85
5.1	The <code>plot</code> Function	85
5.2	Symbols, Colours, and Sizes	88
5.2.1	Changing Plotting Characters	88

5.2.2	Changing the Colour of Plotting Symbols	92
5.2.3	Altering the Size of Plotting Symbols	93
5.3	Adding a Smoothing Line	95
5.4	Which R Functions Did We Learn?	97
5.5	Exercises	97
6	Loops and Functions	99
6.1	Introduction to Loops	99
6.2	Loops	101
6.2.1	Be the Architect of Your Code	102
6.2.2	Step 1: Importing the Data	102
6.2.3	Steps 2 and 3: Making the Scatterplot and Adding Labels	103
6.2.4	Step 4: Designing General Code	104
6.2.5	Step 5: Saving the Graph	105
6.2.6	Step 6: Constructing the Loop	107
6.3	Functions	108
6.3.1	Zeros and NAs	108
6.3.2	Technical Information	10
6.3.3	A Second Example: Zeros and NAs	11
6.3.4	A Function with Multiple Arguments	13
6.3.5	Foolproof Functions	15
6.4	More on Functions and the if Statement	17
6.4.1	Playing the Architect Again	18
6.4.2	Step 1: Importing and Assessing the Data	18
6.4.3	Step 2: Total Abundance per Site	19
6.4.4	Step 3: Richness per Site	20
6.4.5	Step 4: Shannon Index per Site	21
6.4.6	Step 5: Combining Code	22
6.4.7	Step 6: Putting the Code into a Function	122
6.5	Which R Functions Did We Learn?	125
6.6	Exercises	125
7	Graphing Tools	127
7.1	The Pie Chart	127
7.1.1	Pie Chart Showing Avian Influenza Data	127
7.1.2	The par Function	130
7.2	The Bar Chart and Strip Chart	131
7.2.1	The Bar Chart Using the Avian Influenza Data	131
7.2.2	A Bar Chart Showing Mean Values with Standard Deviations	133
7.2.3	The Strip Chart for the Benthic Data	135
7.3	Boxplot	137
7.3.1	Boxplots Showing the Owl Data	137
7.3.2	Boxplots Showing the Benthic Data	140

7.4	Cleveland Dotplots	141
7.4.1	Adding the Mean to a Cleveland Dotplot	143 •
7.5	Revisiting the plot Function	145
7.5.1	The Generic plot Function	145
7.5.2	More Options for the plot Function	146
7.5.3	Adding Extra Points, Text, and Lines	148
7.5.4	Using type = "n".	149
7.5.5	Legends	150
7.5.6	Identifying Points	152
7.5.7	Changing Fonts and Font Size*	153
7.5.8	Adding Special Characters	153
7.5.9	Other Useful Functions	154
7.6	The Pairplot	155
7.6.1	Panel Functions	156
7.7	The Coplot	157
7.7.1	A Coplot with a Single Conditioning Variable	157
7.7.2	The Coplot with Two Conditioning Variables	161
7.7.3	Jazzing Up the Coplot*	162
7.8	Combining Types of Plots*	164
7.9	Which R Functions Did We Learn?	166
7.10	Exercises	167
	An Introduction to the Lattice Package	169
8.1	High-Level Lattice Functions	169
8.2	Multipanel Scatterplots: xyplot	170
8.3	Multipanel Boxplots: bwplot	173
8.4	Multipanel Cleveland Dotplots: dotplot	174
8.5	Multipanel Histograms: histogram	176
8.6	Panel Functions	177
8.6.1	First Panel Function Example	177
8.6.2	Second Panel Function Example	179
8.6.3	Third Panel Function Example*	181
8.7	3-D Scatterplots and Surface and Contour Plots	184
8.8	Frequently Asked Questions	185
8.8.1	How to Change the Panel Order?	186
8.8.2	How to Change Axes Limits and Tick Marks?	188
8.8.3	Multiple Graph Lines in a Single Panel	189
8.8.4	Plotting from Within a Loop*	190
8.8.5	Updating a Plot	191
8.9	Where to Go from Here?	191
8.10	Which R Functions Did We Learn?	192
8.11	Exercises	192

9 Common R Mistakes195

9.1 Problems Importing Data195

9.1.1 Errors in the Source File.195

9.1.2 Decimal Point or Comma Separation . . .195

9.1.3 Directory Names.197

9.2 Attach Misery.197

9.2.1 Entering the Same attach Command Twice.197

9.2.2 Attaching Two Data Frames Containing the Same
Variable Names.198

9.2.3 Attaching a Data Frame and Demo Data199

9.2.4 Making Changes to a Data Frame After Applying the
attach Function.200

9.3 Non-attach Misery.201

9.4 The Log of Zero.202

9.5 Miscellaneous Errors.203

9.5.1 The Difference Between 1 and 1.203

9.5.2 The ;Colour of 0.203

9.6 Mistakenly,Saved the R Workspace.204

References.207

Index.211