

Table of Contents

Preface xv

Part I. R Basics

1. Getting and Installing R 3

 R Versions 3

 Getting and Installing Interactive R Binaries 3

 Windows 4

 Mac OS X 5

 Linux and Unix Systems 5

2. The R User Interface 7

 The R Graphical User Interface 7

 Windows 8

 Mac OS X 8

 Linux and Unix 8

 The R Console 11

 Command-Line Editing 13

 Batch Mode 13

 Using R Inside Microsoft Excel 14

 Other Ways to Run R 15

3. A Short R Tutorial 17

 Basic Operations in R 17

 Functions 19

 Variables 20

 Introduction to Data Structures 22

Objects and Classes	25
Models and Formulas	26
Charts and Graphics	28
Getting Help	32
4. R Packages	35
An Overview of Packages	35
Listing Packages in Local Libraries	36
Loading Packages	38
Loading Packages on Windows and Linux	38
Loading Packages on Mac OS X	38
Exploring Package Repositories	39
Exploring Packages on the Web	40
Finding and Installing Packages Inside R	40
Custom Packages	43
Creating a Package Directory	43
Building the Package	45

Part II. The R Language

5. An Overview of the R Language	49
Expressions	49
Objects	50
Symbols	50
Functions	50
Objects Are Copied in Assignment Statements	52
Everything in R Is an Object	52
Special Values	53
NA	53
Inf and -Inf	53
NaN	54
NULL	54
Coercion	54
The R Interpreter	55
Seeing How R Works	57
6. R Syntax	61
Constants	61
Numeric Vectors	61
Character Vectors	62
Symbols	63
Operators	64
Order of Operations	65
Assignments	67
Expressions	67

Separating Expressions	67
Parentheses	68
Curly Braces	68
Control Structures	69
Conditional Statements	69
Loops	70
Accessing Data Structures	72
Data Structure Operators	73
Indexing by Integer Vector	73
Indexing by Logical Vector	76
Indexing by Name	76
R Code Style Standards	77
7. R Objects	79
Primitive Object Types	79
Vectors	82
Lists	83
Other Objects	84
Matrices	84
Arrays	84
Factors	85
Data Frames	87
Formulas	88
Time Series	89
Shingles	91
Dates and Times	91
Connections	92
Attributes	92
Class	95
8. Symbols and Environments	97
Symbols	97
Working with Environments	98
The Global Environment	99
Environments and Functions	100
Working with the Call Stack	100
Evaluating Functions in Different Environments	101
Adding Objects to an Environment	103
Exceptions	104
Signaling Errors	104
Catching Errors	105
9. Functions	107
The Function Keyword	107
Arguments	107
Return Values	109

Functions As Arguments	109
Anonymous Functions	110
Properties of Functions	111
Argument Order and Named Arguments	113
Side Effects	114
Changes to Other Environments	114
Input/Output	115
Graphics	115
10. Object-Oriented Programming	117
Overview of Object-Oriented Programming in R	118
Key Ideas	118
Implementation Example	119
Object-Oriented Programming in R: S4 Classes	125
Defining Classes	125
New Objects	126
Accessing Slots	126
Working with Objects	127
Creating Coercion Methods	127
Methods	128
Managing Methods	129
Basic Classes	130
More Help	130
Old-School OOP in R: S3	131
S3 Classes	131
S3 Methods	132
Using S3 Classes in S4 Classes	133
Finding Hidden S3 Methods	133
11. High-Performance R	135
Use Built-in Math Functions	135
Use Environments for Lookup Tables	136
Use a Database to Query Large Data Sets	136
Preallocate Memory	137
Monitor How Much Memory You Are Using	137
Monitoring Memory Usage	137
Increasing Memory Limits	138
Cleaning Up Objects	138
Functions for Big Data Sets	139
Parallel Computation with R	139
High-Performance R Binaries	140
Revolution R	140
Building Your Own	141

Part III. Working with Data

12. Saving, Loading, and Editing Data	147
Entering Data Within R	147
Entering Data Using R Commands	147
Using the Edit GUI	148
Saving and Loading R Objects	151
Saving Objects with save	151
Importing Data from External Files	152
Text Files	152
Other Software	161
Exporting Data	161
Importing Data from Databases	162
Export Then Import	162
Database Connection Packages	162
RODBC	163
DBI	173
TSDBI	178
 13. Preparing Data	 179
Combining Data Sets	179
Pasting Together Data Structures	180
Merging Data by Common Fields	183
Transformations	185
Reassigning Variables	185
The Transform Function	185
Applying a Function to Each Element of an Object	186
Binning Data	189
Shingles	189
Cut	190
Combining Objects with a Grouping Variable	191
Subsets	191
Bracket Notation	192
subset Function	192
Random Sampling	193
Summarizing Functions	194
tapply, aggregate	194
Aggregating Tables with rowsum	197
Counting Values	198
Reshaping Data	200
Data Cleaning	205
Finding and Removing Duplicates	206
Sorting	206

14. Graphics	211
An Overview of R Graphics	211
Scatter Plots	212
Plotting Time Series	218
Bar Charts	219
Pie Charts	223
Plotting Categorical Data	224
Three-Dimensional Data	229
Plotting Distributions	237
Box Plots	240
Graphics Devices	243
Customizing Charts	244
Common Arguments to Chart Functions	244
Graphical Parameters	244
Basic Graphics Functions	254
 15. Lattice Graphics	 263
History	263
An Overview of the Lattice Package	264
How Lattice Works	264
A Simple Example	264
Using Lattice Functions	266
Custom Panel Functions	268
High-Level Lattice Plotting Functions	268
Univariate Trellis Plots	269
Bivariate Trellis Plots	293
Trivariate Plots	301
Other Plots	306
Customizing Lattice Graphics	308
Common Arguments to Lattice Functions	308
trellis.skeleton	309
Controlling How Axes Are Drawn	310
Parameters	311
plot.trellis	315
strip.default	316
simpleKey	317
Low-Level Functions	318
Low-Level Graphics Functions	318
Panel Functions	318

Part IV. Statistics with R

16. Analyzing Data	323
Summary Statistics	323
Correlation and Covariance	325

Principal Components Analysis	328
Factor Analysis	332
Bootstrap Resampling	333
17. Probability Distributions	335
Normal Distribution	335
Common Distribution-Type Arguments	338
Distribution Function Families	338
18. Statistical Tests	343
Continuous Data	343
Normal Distribution-Based Tests	344
Distribution-Free Tests	357
Discrete Data	360
Proportion Tests	360
Binomial Tests	361
Tabular Data Tests	362
Distribution-Free Tabular Data Tests	368
19. Power Tests	369
Experimental Design Example	369
t -Test Design	370
Proportion Test Design	371
ANOVA Test Design	372
20. Regression Models	373
Example: A Simple Linear Model	373
Fitting a Model	375
Helper Functions for Specifying the Model	376
Getting Information About a Model	376
Refining the Model	382
Details About the lm Function	382
Assumptions of Least Squares Regression	384
Robust and Resistant Regression	386
Subset Selection and Shrinkage Methods	387
Stepwise Variable Selection	388
Ridge Regression	389
Lasso and Least Angle Regression	390
Principal Components Regression and Partial Least Squares	
Regression	391
Nonlinear Models	392
Generalized Linear Models	392
Nonlinear Least Squares	395
Survival Models	396
Smoothing	401

- Splines 401
- Fitting Polynomial Surfaces 403
- Kernel Smoothing 404
- Machine Learning Algorithms for Regression 405
 - Regression Tree Models 406
 - MARS 418
 - Neural Networks 423
 - Project Pursuit Regression 427
 - Generalized Additive Models 430
 - Support Vector Machines 432
- 21. Classification Models 435**
 - Linear Classification Models 435
 - Logistic Regression 435
 - Linear Discriminant Analysis 440
 - Log-Linear Models 444
 - Machine Learning Algorithms for Classification 445
 - k* Nearest Neighbors 445
 - Classification Tree Models 446
 - Neural Networks 450
 - SVMs 451
 - Random Forests 451
- 22. Machine Learning 453**
 - Market Basket Analysis 453
 - Clustering 458
 - Distance Measures 458
 - Clustering Algorithms 459
- 23. Time Series Analysis 463**
 - Autocorrelation Functions 463
 - Time Series Models 464
- 24. Bioconductor 469**
 - An Example 469
 - Loading Raw Expression Data 470
 - Loading Data from GEO 474
 - Matching Phenotype Data 476
 - Analyzing Expression Data 477
 - Key Bioconductor Packages 481
 - Data Structures 485
 - eSet 485
 - AssayData 487
 - AnnotatedDataFrame 487
 - MIAME 488

Other Classes Used by Bioconductor Packages	489
Where to Go Next	490
Resources Outside Bioconductor	490
Vignettes	490
Courses	491
Books	491
Appendix: R Reference	493
Bibliography	591
Index	593