

# R for Everyone

## Advanced Analytics and Graphics

Jared P. Lander

•Addison-Wesley

Upper Saddle River, NJ • Boston • Indianapolis • San Francisco  
New York • Toronto • Montreal • London • Munich • Paris • Madrid  
Capetown • Sydney • Tokyo • Singapore • Mexico City

# Contents

Foreword xiii

Preface xv

Acknowledgments xix

About the Author xxiii

## **1 Getting R 1**

- 1.1 Downloading R 1
- 1.2 R Version 2
- 1.3 32 bit versus 64-bit 2
- 1.4 Installing 2
- 1.5 RevolutionsCommunity Edition
- 1.6 Conclusion 11

## **2 The R Environment 13**

- 2.1 Command Line Interface 14
- 2.2 RStudio 15
- 2.3 Revolution Analytics RPE 26
- 2.4 Conclusion 27

## **3 R Packages 29**

- 3.1 Installing Packages 29
- 3.2 Loading Packages 32
- 3.3 Building a Package\* 33
- 3.4 Conclusion 33

## **4 Basics of R 35**

- 4.1 Basic Math 35
- 4.2 Variables 36
- 4.3 Data Types 38
- 4.4 Vectors 43
- 4.5 Calling Functions 49
- 4.6 Function Documentation 49
- 4.7 Missing Data 50
- 4.8 Conclusion 51

<b>5</b>	<b>Advanced Data Structures</b>	<b>53</b>
5.1	<code>data, frames</code>	53
5.2	<code>Lists</code>	61
5.3	<code>Matrices</code>	68
5.4	<code>Arrays</code>	71
5.5	Conclusion	72
<b>6</b>	<b>Reading Data into R</b>	<b>73</b>
6.1	Reading CSVs	73
6.2	Excel Data	74
6.3	Reading from Databases	75
6.4	Data from Other Statistical Tools	77
6.5	R Binary Files	77
6.6	Data Included with R	79
6.7	Extract Data from Web Sites	80
6.8	Conclusion	81
<b>7</b>	<b>Statistical Graphics</b>	<b>83</b>
7.1	Base Graphics	83
7.2	<code>ggplot2</code>	86
7.3	Conclusion	98
<b>8</b>	<b>Writing R Functions</b>	<b>99</b>
8.1	Hello, World!	99
8.2	Function Arguments	100
8.3	Return Values	103
8.4	<code>do.call</code>	104
8.5	Conclusion	104
<b>9</b>	<b>Control Statements</b>	<b>105</b>
9.1	<code>if</code> and <code>else</code>	105
9.2	<code>switch</code>	108
9.3	<code>ifelse</code>	109
9.4	Compound Tests	111
9.5	Conclusion	112
<b>10</b>	<b>Loops, the Un-R Way to Iterate</b>	<b>113</b>
10.1	<code>for</code> Loops	113
10.2	<code>while</code> Loops	115

10.3	Controlling Loops	115
10.4	Conclusion	116
<b>11</b>	<b>Group Manipulation</b>	<b>117</b>
11.1	Apply Family	117
11.2	<code>aggregate</code>	120
11.3	<code>plyr</code>	124
11.4	<code>data, table</code>	129
11.5	Conclusion	139
<b>12</b>	<b>Data Reshaping</b>	<b>141</b>
12.1	<code>cbind</code> and <code>rbind</code>	141
12.2	Joins	142
12.3	<code>reshape2</code>	149
12.4	Conclusion	153
<b>13</b>	<b>Manipulating Strings</b>	<b>155</b>
13.1	<code>paste</code>	155
13.2	<code>sprintf</code>	156
13.3	Extracting Text	157
13.4	Regular Expressions	161
13.5	Conclusion	169
<b>14</b>	<b>Probability Distributions</b>	<b>171</b>
14.1	Normal Distribution	171
14.2	Binomial Distribution	176
14.3	Poisson Distribution	182
14.4	Other Distributions	185
14.5	Conclusion	186
<b>15</b>	<b>Basic Statistics</b>	<b>187</b>
15.1	Summary Statistics	187
15.2	Correlation and Covariance	191
15.3	T-Tests	200
15.4	ANOVA	207
15.5	Conclusion	210

<b>16 Linear Models</b>	<b>211</b>
16.1 Simple Linear Regression	211
16.2 Multiple Regression	216
16.3 Conclusion	232
<b>17 Generalized Linear Models</b>	<b>233</b>
17.1 Logistic Regression	233
17.2 Poisson Regression	237
17.3 Other Generalized Linear Models	240
17.4 Survival Analysis	240
17.5 Conclusion	245
<b>18 Model Diagnostics</b>	<b>247</b>
18.1 Residuals	247
18.2 Comparing Models	253
18.3 Cross-Validation	257
18.4 Bootstrap	262
18.5 Stepwise Variable Selection	265
18.6 Conclusion	269
<b>Regularization and Shrinkage</b>	<b>271</b>
19.1 Elastic Net	271
19.2 Bayesian Shrinkage	290
19.3 Conclusion	295
<b>20 Nonlinear Models</b>	<b>297</b>
20.1 Nonlinear Least Squares	297
20.2 Splines	300
20.3 Generalized Additive Models	304
20.4 Decision Trees	310
20.5 Random Forests	312
20.6 Conclusion	313
<b>21 Time Series and Autocorrelation</b>	<b>315</b>
21.1 Autoregressive Moving Average	315
21.2 VAR	322

21.3	GARCH	327
21.4	Conclusion	336

## **22 Clustering 337**

22.1	K-means	337
22.2	PAM	345
22.3	Hierarchical Clustering	352
22.4	Conclusion	357

## **23 Reproducibility, Reports and Slide Shows with knitr 359**

23.1	Installing a @gX Program	359
23.2	ISlqX Primer	360
23.3	Using <b>knitr</b> with ISlq*	362
23.4	Markdown Tips	367
23.5	Using <b>knitr</b> and Markdown	368
23.6	pandoc	369
23.7	Conclusion	371

## **24 Building R Packages 373**

24.1	Folder Structure	373
24.2	Package Files	373
24.3	Package Documentation	380
24.4	Checking, Building and Installing	383
24.5	Submitting to CRAN	384
24.6	C++ Code	384
24.7	Conclusion	390

## **A Real-Life Resources 391**

A.1	Meetups	391
A.2	Stackoverflow	392
A.3	Twitter	393
A.4	Conferences	393
A.5	Web Sites	393
A.6	Documents	394
A.7	Books	394
A.8	Conclusion	394

<b>Glossary</b>	<b>395</b>
<b>List of Figures</b>	<b>409</b>
<b>List of Tables</b>	<b>417</b>
<b>General Index</b>	<b>419</b>
<b>Index of Functions</b>	<b>427</b>
<b>Index of Packages</b>	<b>431</b>
<b>Index of People</b>	<b>433</b>
<b>Data Index</b>	<b>435</b>