

# Contents

<b>1</b>	<b>Why R</b>	<b>1</b>
1.1	Reasons for Classifying R as a Complete Analytical Environment	1
1.2	Additional Advantages of R over Other Analytical Packages	2
1.3	Differences Between R as a Statistical Language and R as an Analytical Platform	3
1.4	Costs and Benefits of Using R	3
1.4.1	Business Analytics	3
1.4.2	Data Mining	4
1.4.3	Business Dashboards and Reporting	4
1.5	Using SAS and R Together	5
1.6	Brief Interview: Using R with JMP	5
<b>2</b>	<b>R Infrastructure</b>	<b>9</b>
2.1	Choices in Setting up R for Business Analytics	9
2.1.1	Licensing Choices: Academic, Free, or Enterprise Version of R	9
2.1.2	Operating System Choices	10
2.1.3	Operating System Subchoice: 32- or 64-bit	11
2.1.4	Hardware Choices: Cost-Benefit Tradeoffs for Additional Hardware for R	11
2.1.5	Interface Choices: Command Line Versus GUI. Which GUI Should You Choose as the Default Startup Option?	12
2.1.6	Software Component Choice	13
2.1.7	Additional Software Choices	13
2.2	Downloading and Installing R	13
2.3	Installing R Packages	14
2.4	Starting up Tutorial in R	20
2.5	Types of Data in R	21
2.6	Brief Interview with John Fox, Creator of Rcmdr GUI for R	21

2.7	Summary of Commands Used in This Chapter	23
2.7.1	Packages	23
2.7.2	Functions	23
<b>3</b>	<b>R Interfaces</b>	25
3.1	Interfaces to the R Statistical Language	25
3.2	Basic R	26
3.3	Advantages and Limitations of Graphical User Interfaces to R	27
3.3.1	Advantages of Using GUIs for R	27
3.3.2	Limitations of Using GUIs for R	27
3.4	List of GUI	27
3.4.1	R Commander	28
3.4.2	R Commander E Plugins or Extensions	31
3.5	Summary of R GUIs	34
3.6	Using R from Other Software	34
3.6.1	RExcel: Using R from Microsoft Excel	36
3.7	Web Interfaces to R	37
3.8	Interview: Using R as a Web-Based Application	40
3.9	Cloud Computing to Use R	41
3.9.1	Benefits of R on the Cloud	42
3.9.2	Tutorial: Using Amazon EC2 and R (Linux)	42
3.9.3	Tutorial: Using Amazon EC2 and R (Windows)	44
3.9.4	Installing R on a Basic Linux Amazon EC2 Instance (Which Is Free)	46
3.9.5	Using R Studio on Amazon EC2	46
3.9.6	Running R on the Cloud Using cloudnumbers.com	47
3.10	Google and R	47
3.10.1	Google Style Guide	47
3.10.2	Using R at Google	48
3.10.3	Using Google Services and R Packages	50
3.11	Interview: Using R at Google	51
3.12	Interview: Using R Through Cloud Computing at cloudnumbers.com	53
3.13	Summary of Commands Used in This Chapter	54
3.13.1	Packages	54
<b>4</b>	<b>Manipulating Data</b>	57
4.1	Challenges of Analytical Data Processing	57
4.1.1	Data Formats	57
4.1.2	Data Quality	58
4.1.3	Project Scope	58
4.1.4	Output Results vis-à-vis Stakeholder Expectation Management	58
4.2	Methods for Reading in Smaller Dataset Sizes	59
4.2.1	CSV and Spreadsheets	59

4.2.2	Reading Data from Packages .....	61
4.2.3	Reading Data from Web/APIs .....	61
4.2.4	Missing Value Treatment in R .....	62
4.2.5	Using the as Operator to Change the Structure of Data .....	63
4.3	Some Common Analytical Tasks .....	64
4.3.1	Exploring a Dataset .....	64
4.3.2	Conditional Manipulation of a Dataset .....	65
4.3.3	Merging Data .....	69
4.3.4	Aggregating and Group Processing of a Variable .....	70
4.3.5	Manipulating Text in Data .....	72
4.4	A Simple Analysis Using R .....	73
4.4.1	Input .....	73
4.4.2	Describe Data Structure .....	73
4.4.3	Describe Variable Structure .....	74
4.4.4	Output .....	76
4.5	Comparison of R Graphical User Interfaces for Data Input .....	76
4.6	Using R with Databases and Business Intelligence Systems .....	77
4.6.1	RODBC .....	78
4.6.2	Using MySQL and R .....	78
4.6.3	Using PostGresSQL and R .....	84
4.6.4	Using SQLite and R .....	85
4.6.5	Using JasperDB and R .....	86
4.6.6	Using Pentaho and R .....	87
4.7	Summary of Commands Used in This Chapter .....	88
4.7.1	Packages .....	88
4.7.2	Functions .....	88
4.8	Citations and References .....	89
4.9	Additional Resources .....	89
4.9.1	Methods for Larger Dataset Sizes .....	90
<b>5</b>	<b>Exploring Data .....</b>	<b>103</b>
5.1	Business Metrics .....	103
5.2	Data Visualization .....	104
5.3	Parameters for Graphs .....	105
5.4	Creating Graphs in R .....	106
5.4.1	Basic Graphs .....	106
5.4.2	Summary of Basic Graphs in R .....	118
5.4.3	Advanced Graphs .....	119
5.4.4	Additional Graphs .....	140
5.5	Using ggplot2 for Advanced Graphics in R .....	151
5.6	Interactive Plots .....	152
5.7	Grapher: R GUI for Simple Graphs .....	154
5.7.1	Advantages of Grapher .....	155
5.7.2	Disadvantages of Grapher .....	155

5.8	Deducer: GUI for Advanced Data Visualization .....	156
5.8.1	Advantages of JGR and Deducer .....	157
5.8.2	Disadvantages of Deducer .....	160
5.8.3	Description of Deducer .....	161
5.9	Color Palettes .....	161
5.10	Interview: Hadley Wickham, Author of <i>ggplot2</i> : <i>Elegant Graphics for Data Analysis</i> .....	164
5.11	Summary of Commands Used in This Chapter .....	165
5.11.1	Packages .....	165
5.11.2	Functions .....	166
<b>6</b>	<b>Building Regression Models</b> .....	171
6.1	Linear Regression .....	171
6.2	Logistic Regression .....	172
6.3	Risk Models .....	172
6.4	Scorecards .....	172
6.4.1	Credit Scorecards .....	173
6.4.2	Fraud Models .....	173
6.4.3	Marketing Propensity Models .....	173
6.5	Useful Functions in Building Regression Models in R .....	173
6.6	Using R Cmdr to Build a Regression Model .....	174
6.7	Other Packages for Regression Models .....	188
6.7.1	ROCR for Performance Curves .....	188
6.7.2	rms Package .....	188
6.8	PMML .....	189
6.8.1	Zementis: Amazon EC2 for Scoring .....	189
6.9	Summary of Commands Used in This Chapter .....	190
6.9.1	Packages .....	190
6.9.2	Functions .....	191
<b>7</b>	<b>Data Mining Using R</b> .....	193
7.1	Definition .....	193
7.1.1	Information Ladder .....	194
7.1.2	KDD .....	194
7.1.3	CRISP-DM .....	195
7.1.4	SEMMA .....	196
7.1.5	Four Phases of Data-Driven Projects .....	197
7.1.6	Data Mining Methods .....	198
7.2	Rattle: A GUI for Data Mining in R .....	199
7.2.1	Advantages of Rattle .....	199
7.2.2	Comparative Disadvantages of Using Rattle .....	201
7.2.3	Description of Rattle .....	201
7.3	Interview: Graham Williams, Author of <i>Data Mining with Rattle and R</i> .....	212

7.4	Text Mining Analytics Using R .....	214
7.4.1	Text Mining a Local Document .....	214
7.4.2	Text Mining from the Web and Cleaning Text Data .....	215
7.5	Google Prediction API .....	220
7.6	Data Privacy for Data Miners .....	222
7.7	Summary of Commands Used in This Chapter .....	222
7.7.1	Packages .....	222
7.7.2	Functions .....	223
<b>8</b>	<b>Clustering and Data Segmentation .....</b>	<b>225</b>
8.1	When to Use Data Segmentation and Clustering .....	225
8.2	R Support for Clustering .....	225
8.2.1	Clustering View .....	225
8.2.2	GUI-Based Method for Clustering .....	226
8.3	Using RevoScaleR for Revolution Analytics .....	226
8.4	A GUI Called Playwith .....	227
8.5	Cluster Analysis Using Rattle .....	229
8.6	Summary of Commands Used in This Chapter .....	239
8.6.1	Packages .....	239
8.6.2	Functions .....	239
<b>9</b>	<b>Forecasting and Time Series Models .....</b>	<b>241</b>
9.1	Introduction to Time Series .....	241
9.2	Time Series and Forecasting Methodology .....	241
9.3	Time Series Model Types .....	246
9.4	Handling Date-Time Data .....	247
9.5	Using R Commander GUI with epack Plugin .....	248
9.5.1	Syntax Generated Using R Commander GUI with epack Plugin .....	255
9.6	Summary of Commands Used in This Chapter .....	256
9.6.1	Packages .....	256
9.6.2	Functions .....	256
<b>10</b>	<b>Data Export and Output .....</b>	<b>259</b>
10.1	Summary of Commands Used in This Chapter .....	262
10.1.1	Packages .....	262
10.1.2	Functions .....	262
<b>11</b>	<b>Optimizing R Code .....</b>	<b>263</b>
11.1	Examples of Efficient Coding .....	263
11.2	Customizing R Software Startup .....	265
11.2.1	Where is the R Profile File? .....	265
11.2.2	Modify Settings .....	266
11.3	Code Editors .....	266
11.4	Advantages of Enhanced Code Editors .....	272
11.5	Interview: J.J. Allaire, Creator of R Studio .....	273

11.6	Revolution R Productivity Environment .....	274
11.7	Evaluating Code Efficiency .....	275
11.8	Using system.time to Evaluate Coding Efficiency .....	277
11.9	Using GUIs to Learn and Code R Faster .....	278
11.10	Parallel Programming .....	278
11.11	Using Hardware Solutions .....	279
11.12	Summary of Commands Used in This Chapter .....	279
11.12.1	Packages .....	279
11.12.2	Functions .....	280
<b>12</b>	<b>Additional Training Literature .....</b>	<b>281</b>
12.1	Cran Views .....	281
12.2	Reading Material .....	282
12.3	Other GUIs Used in R .....	283
12.3.1	Red-R: A Dataflow User Interface for R .....	283
12.3.2	RKward .....	284
12.3.3	Komodo Sciviews-K .....	288
12.3.4	PMG (or Poor Man's GUI) .....	289
12.3.5	R Analytic Flow .....	290
12.4	Summary of Commands Used in This Chapter .....	290
12.4.1	Packages .....	290
12.4.2	Functions .....	290
<b>13</b>	<b>Appendix .....</b>	<b>293</b>
13.1	Web Analytics Using R .....	293
13.1.1	Google Analytics with R .....	293
13.2	Social Media Analytics Using R .....	295
13.2.1	Using Facebook Data with R .....	295
13.2.2	Using Twitter Data with R .....	297
13.3	RFM Analytics Using R .....	299
13.4	Propensity Models using R .....	300
13.5	Risk Models in Finance Using R .....	300
13.6	Pharmaceutical Analytics Using R .....	300
13.7	Selected Essays on Analytics by the Author .....	301
13.7.1	What Is Analytics? .....	301
13.7.2	What Are the Basic Business Domains Within Analytics? .....	302
13.8	Reasons a Business Analyst Should Learn R .....	304
13.9	Careers in Analytics .....	305
13.10	Summary of Commands Used in This Chapter .....	307
13.10.1	Packages .....	307
13.10.2	Functions .....	307
	<b>Index .....</b>	<b>309</b>

R for Business Analytics

Ohri, A.

2013, XVIII, 312 p. 206 illus., 160 illus. in color.,

ISBN: 978-1-4614-4343-8