

Table of Contents

Welcome to the R Cookbook, 2nd Edition.....	xi
1. Getting Started and Getting Help.....	1
1.1 Downloading and Installing R.....	2
1.2 Installing RStudio.....	4
1.3 Starting RStudio.....	5
1.4 Entering Commands.....	7
1.5 Exiting from RStudio.....	9
1.6 Interrupting R.....	11
1.7 Viewing the Supplied Documentation.....	12
1.8 Getting Help on a Function.....	14
1.9 Searching the Supplied Documentation.....	16
1.10 Getting Help on a Package.....	17
1.11 Searching the Web for Help.....	18
1.12 Finding Relevant Functions and Packages.....	21
1.13 Searching the Mailing Lists.....	22
1.14 Submitting Questions to Stack Overflow or Elsewhere in the Community.....	23
2. Some Basics.....	27
2.1 Printing Something to the Screen.....	27
2.2 Setting Variables.....	29
2.3 Listing Variables.....	31
2.4 Deleting Variables.....	32
2.5 Creating a Vector.....	34
2.6 Computing Basic Statistics.....	35
2.7 Creating Sequences.....	38
2.8 Comparing Vectors.....	39
2.9 Selecting Vector Elements.....	41

2.10 Performing Vector Arithmetic	44
2.11 Getting Operator Precedence Right	46
2.12 Typing Less and Accomplishing More	47
2.13 Creating a Pipeline of Function Calls	49
2.14 Avoiding Some Common Mistakes	52
3. Navigating the Software.	59
3.1 Getting and Setting the Working Directory	59
3.2 Creating a New RStudio Project	60
3.3 Saving Your Workspace	63
3.4 Viewing Your Command History	64
3.5 Saving the Result of the Previous Command	65
3.6 Displaying Loaded Packages via the Search Path	66
3.7 Viewing the List of Installed Packages	68
3.8 Accessing the Functions in a Package	69
3.9 Accessing Built-in Datasets	70
3.10 Installing Packages from CRAN	72
3.11 Installing a Package from GitHub	73
3.12 Setting or Changing a Default CRAN Mirror	74
3.13 Running a Script	76
3.14 Running a Batch Script	77
3.15 Locating the R Home Directory	79
3.16 Customizing R Startup	81
3.17 Using R and RStudio in the Cloud	84
4. Input and Output.	87
4.1 Entering Data from the Keyboard	87
4.2 Printing Fewer Digits (or More Digits)	88
4.3 Redirecting Output to a File	90
4.4 Listing Files	91
4.5 Dealing with “Cannot Open File” in Windows	94
4.6 Reading Fixed-Width Records	94
4.7 Reading Tabular Data Files	97
4.8 Reading from CSV Files	101
4.9 Writing to CSV Files	103
4.10 Reading Tabular or CSV Data from the Web	104
4.11 Reading Data from Excel	105
4.12 Writing a Data Frame to Excel	107
4.13 Reading Data from a SAS File	109
4.14 Reading Data from HTML Tables	111
4.15 Reading Files with a Complex Structure	113
4.16 Reading from MySQL Databases	118

4.17 Accessing a Database with dbplyr	121
4.18 Saving and Transporting Objects	123
5. Data Structures	127
5.1 Appending Data to a Vector	135
5.2 Inserting Data into a Vector	136
5.3 Understanding the Recycling Rule	137
5.4 Creating a Factor (Categorical Variable)	139
5.5 Combining Multiple Vectors into One Vector and a Factor	140
5.6 Creating a List	142
5.7 Selecting List Elements by Position	144
5.8 Selecting List Elements by Name	145
5.9 Building a Name/Value Association List	147
5.10 Removing an Element from a List	149
5.11 Flattening a List into a Vector	150
5.12 Removing NULL Elements from a List	151
5.13 Removing List Elements Using a Condition	152
5.14 Initializing a Matrix	153
5.15 Performing Matrix Operations	155
5.16 Giving Descriptive Names to the Rows and Columns of a Matrix	156
5.17 Selecting One Row or Column from a Matrix	157
5.18 Initializing a Data Frame from Column Data	158
5.19 Initializing a Data Frame from Row Data	160
5.20 Appending Rows to a Data Frame	162
5.21 Selecting Data Frame Columns by Position	165
5.22 Selecting Data Frame Columns by Name	168
5.23 Changing the Names of Data Frame Columns	170
5.24 Removing NAs from a Data Frame	171
5.25 Excluding Columns by Name	172
5.26 Combining Two Data Frames	173
5.27 Merging Data Frames by Common Column	174
5.28 Converting One Atomic Value into Another	177
5.29 Converting One Structured Data Type into Another	178
6. Data Transformations	181
6.1 Applying a Function to Each List Element	182
6.2 Applying a Function to Every Row of a Data Frame	184
6.3 Applying a Function to Every Row of a Matrix	185
6.4 Applying a Function to Every Column	186
6.5 Applying a Function to Parallel Vectors or Lists	188
6.6 Applying a Function to Groups of Data	191
6.7 Creating a New Column Based on Some Condition	192

7. Strings and Dates.....	195
7.1 Getting the Length of a String	197
7.2 Concatenating Strings	198
7.3 Extracting Substrings	199
7.4 Splitting a String According to a Delimiter	200
7.5 Replacing Substrings	201
7.6 Generating All Pairwise Combinations of Strings	202
7.7 Getting the Current Date	204
7.8 Converting a String into a Date	204
7.9 Converting a Date into a String	205
7.10 Converting Year, Month, and Day into a Date	206
7.11 Getting the Julian Date	208
7.12 Extracting the Parts of a Date	208
7.13 Creating a Sequence of Dates	210
8. Probability.....	213
8.1 Counting the Number of Combinations	215
8.2 Generating Combinations	216
8.3 Generating Random Numbers	217
8.4 Generating Reproducible Random Numbers	219
8.5 Generating a Random Sample	220
8.6 Generating Random Sequences	221
8.7 Randomly Permuting a Vector	222
8.8 Calculating Probabilities for Discrete Distributions	223
8.9 Calculating Probabilities for Continuous Distributions	225
8.10 Converting Probabilities to Quantiles	227
8.11 Plotting a Density Function	228
9. General Statistics.....	233
9.1 Summarizing Your Data	235
9.2 Calculating Relative Frequencies	237
9.3 Tabulating Factors and Creating Contingency Tables	238
9.4 Testing Categorical Variables for Independence	239
9.5 Calculating Quantiles (and Quartiles) of a Dataset	240
9.6 Inverting a Quantile	241
9.7 Converting Data to z-Scores	242
9.8 Testing the Mean of a Sample (t-Test)	243
9.9 Forming a Confidence Interval for a Mean	244
9.10 Forming a Confidence Interval for a Median	246
9.11 Testing a Sample Proportion	247
9.12 Forming a Confidence Interval for a Proportion	248
9.13 Testing for Normality	249

9.14 Testing for Runs	250
9.15 Comparing the Means of Two Samples	252
9.16 Comparing the Locations of Two Samples Nonparametrically	254
9.17 Testing a Correlation for Significance	255
9.18 Testing Groups for Equal Proportions	257
9.19 Performing Pairwise Comparisons Between Group Means	258
9.20 Testing Two Samples for the Same Distribution	260
10. Graphics	263
10.1 Creating a Scatter Plot	267
10.2 Adding a Title and Labels	268
10.3 Adding (or Removing) a Grid	270
10.4 Applying a Theme to a ggplot Figure	274
10.5 Creating a Scatter Plot of Multiple Groups	278
10.6 Adding (or Removing) a Legend	280
10.7 Plotting the Regression Line of a Scatter Plot	284
10.8 Plotting All Variables Against All Other Variables	288
10.9 Creating One Scatter Plot for Each Group	290
10.10 Creating a Bar Chart	292
10.11 Adding Confidence Intervals to a Bar Chart	295
10.12 Coloring a Bar Chart	298
10.13 Plotting a Line from x and y Points	300
10.14 Changing the Type, Width, or Color of a Line	302
10.15 Plotting Multiple Datasets	305
10.16 Adding Vertical or Horizontal Lines	307
10.17 Creating a Boxplot	309
10.18 Creating One Boxplot for Each Factor Level	311
10.19 Creating a Histogram	313
10.20 Adding a Density Estimate to a Histogram	315
10.21 Creating a Normal Quantile–Quantile Plot	317
10.22 Creating Other Quantile–Quantile Plots	319
10.23 Plotting a Variable in Multiple Colors	322
10.24 Graphing a Function	325
10.25 Displaying Several Figures on One Page	328
10.26 Writing Your Plot to a File	331
11. Linear Regression and ANOVA	333
11.1 Performing Simple Linear Regression	335
11.2 Performing Multiple Linear Regression	337
11.3 Getting Regression Statistics	339
11.4 Understanding the Regression Summary	342
11.5 Performing Linear Regression Without an Intercept	345

11.6 Regressing Only Variables That Highly Correlate with Your Dependent Variable	347
11.7 Performing Linear Regression with Interaction Terms	350
11.8 Selecting the Best Regression Variables	352
11.9 Regressing on a Subset of Your Data	357
11.10 Using an Expression Inside a Regression Formula	358
11.11 Regressing on a Polynomial	360
11.12 Regressing on Transformed Data	361
11.13 Finding the Best Power Transformation (Box–Cox Procedure)	363
11.14 Forming Confidence Intervals for Regression Coefficients	368
11.15 Plotting Regression Residuals	369
11.16 Diagnosing a Linear Regression	371
11.17 Identifying Influential Observations	375
11.18 Testing Residuals for Autocorrelation (Durbin–Watson Test)	376
11.19 Predicting New Values	378
11.20 Forming Prediction Intervals	379
11.21 Performing One-Way ANOVA	380
11.22 Creating an Interaction Plot	382
11.23 Finding Differences Between Means of Groups	383
11.24 Performing Robust ANOVA (Kruskal–Wallis Test)	386
11.25 Comparing Models by Using ANOVA	388
12. Useful Tricks.....	391
12.1 Peeking at Your Data	391
12.2 Printing the Result of an Assignment	394
12.3 Summing Rows and Columns	394
12.4 Printing Data in Columns	395
12.5 Binning Your Data	396
12.6 Finding the Position of a Particular Value	397
12.7 Selecting Every nth Element of a Vector	398
12.8 Finding Minimums or Maximums	399
12.9 Generating All Combinations of Several Variables	401
12.10 Flattening a Data Frame	402
12.11 Sorting a Data Frame	403
12.12 Stripping Attributes from a Variable	404
12.13 Revealing the Structure of an Object	405
12.14 Timing Your Code	408
12.15 Suppressing Warnings and Error Messages	410
12.16 Taking Function Arguments from a List	411
12.17 Defining Your Own Binary Operators	413
12.18 Suppressing the Startup Message	414
12.19 Getting and Setting Environment Variables	415

12.20 Use Code Sections	416
12.21 Executing R in Parallel Locally	417
12.22 Executing R in Parallel Remotely	420
13. Beyond Basic Numerics and Statistics	425
13.1 Minimizing or Maximizing a Single-Parameter Function	425
13.2 Minimizing or Maximizing a Multiparameter Function	426
13.3 Calculating Eigenvalues and Eigenvectors	428
13.4 Performing Principal Component Analysis	429
13.5 Performing Simple Orthogonal Regression	430
13.6 Finding Clusters in Your Data	433
13.7 Predicting a Binary-Valued Variable (Logistic Regression)	436
13.8 Bootstrapping a Statistic	438
13.9 Factor Analysis	441
14. Time Series Analysis	447
14.1 Representing Time Series Data	449
14.2 Plotting Time Series Data	452
14.3 Extracting the Oldest or Newest Observations	454
14.4 Subsetting a Time Series	456
14.5 Merging Several Time Series	458
14.6 Filling or Padding a Time Series	460
14.7 Lagging a Time Series	463
14.8 Computing Successive Differences	464
14.9 Performing Calculations on Time Series	466
14.10 Computing a Moving Average	467
14.11 Applying a Function by Calendar Period	469
14.12 Applying a Rolling Function	471
14.13 Plotting the Autocorrelation Function	473
14.14 Testing a Time Series for Autocorrelation	475
14.15 Plotting the Partial Autocorrelation Function	476
14.16 Finding Lagged Correlations Between Two Time Series	478
14.17 Detrending a Time Series	479
14.18 Fitting an ARIMA Model	482
14.19 Removing Insignificant ARIMA Coefficients	486
14.20 Running Diagnostics on an ARIMA Model	487
14.21 Making Forecasts from an ARIMA Model	490
14.22 Plotting a Forecast	491
14.23 Testing for Mean Reversion	492
14.24 Smoothing a Time Series	495

15. Simple Programming.....	499
15.1 Choosing Between Two Alternatives: if/else	500
15.2 Iterating with a Loop	502
15.3 Defining a Function	503
15.4 Creating a Local Variable	505
15.5 Choosing Between Multiple Alternatives: switch	506
15.6 Defining Defaults for Function Parameters	507
15.7 Signaling Errors	508
15.8 Protecting Against Errors	509
15.9 Creating an Anonymous Function	510
15.10 Creating a Collection of Reusable Functions	511
15.11 Automatically Reindenting Code	513
16. R Markdown and Publishing.....	515
16.1 Creating a New Document	516
16.2 Adding a Title, Author, or Date	518
16.3 Formatting Document Text	520
16.4 Inserting Document Headings	521
16.5 Inserting a List	521
16.6 Showing Output from R Code	523
16.7 Controlling Which Code and Results Are Shown	525
16.8 Inserting a Plot	526
16.9 Inserting a Table	530
16.10 Inserting a Table of Data	531
16.11 Inserting Math Equations	534
16.12 Generating HTML Output	535
16.13 Generating PDF Output	536
16.14 Generating Microsoft Word Output	539
16.15 Generating Presentation Output	546
16.16 Creating a Parameterized Report	548
16.17 Organizing Your R Markdown Workflow	552
Index.....	555