

Contents

Part I Introduction

1	The Role of Data Wrangling	3
2	Introduction to R	7
2.1	Open Source	7
2.2	Flexibility	8
2.3	Community	9
3	The Basics	11
3.1	Installing R and RStudio	11
3.2	Understanding the Console	13
3.2.1	Script Editor	13
3.2.2	Workspace Environment	13
3.2.3	Console	15
3.2.4	Misc. Displays	15
3.2.5	Workspace Options and Shortcuts	15
3.3	Getting Help	16
3.3.1	General Help	16
3.3.2	Getting Help on Functions	16
3.3.3	Getting Help from the Web	17
3.4	Working with Packages	17
3.4.1	Installing Packages	18
3.4.2	Loading Packages	18
3.4.3	Getting Help on Packages	19
3.4.4	Useful Packages	19
3.5	Assignment and Evaluation	19
3.6	R as a Calculator	21
3.6.1	Vectorization	22

3.7	Styling Guide	24
3.7.1	Notation and Naming	24
3.7.2	Organization	25
3.7.3	Syntax	26

Part II Working with Different Types of Data in R

4	Dealing with Numbers	31
4.1	Integer vs. Double	31
4.1.1	Creating Integer and Double Vectors	31
4.1.2	Converting Between Integer and Double Values	32
4.2	Generating Sequence of Non-random Numbers	32
4.2.1	Specifying Numbers Within a Sequence	32
4.2.2	Generating Regular Sequences	33
4.3	Generating Sequence of Random Numbers	33
4.3.1	Uniform Numbers	34
4.3.2	Normal Distribution Numbers	34
4.3.3	Binomial Distribution Numbers	35
4.3.4	Poisson Distribution Numbers	36
4.3.5	Exponential Distribution Numbers	36
4.3.6	Gamma Distribution Numbers	37
4.4	Setting the Seed for Reproducible Random Numbers	37
4.5	Comparing Numeric Values	37
4.5.1	Comparison Operators	38
4.5.2	Exact Equality	39
4.5.3	Floating Point Comparison	39
4.6	Rounding Numbers	39
5	Dealing with Character Strings	41
5.1	Character String Basics	41
5.1.1	Creating Strings	41
5.1.2	Converting to Strings	42
5.1.3	Printing Strings	43
5.1.4	Counting String Elements and Characters	45
5.2	String Manipulation with Base R	46
5.2.1	Case Conversion	46
5.2.2	Simple Character Replacement	46
5.2.3	String Abbreviations	47
5.2.4	Extract/Replace Substrings	47
5.3	String Manipulation with <code>stringr</code>	49
5.3.1	Basic Operations	49
5.3.2	Duplicate Characters Within a String	51
5.3.3	Remove Leading and Trailing Whitespace	51
5.3.4	Pad a String with Whitespace	52

5.4	Set Operatons for Character Strings	52
5.4.1	Set Union	52
5.4.2	Set Intersection.....	52
5.4.3	Identifying Different Elements	53
5.4.4	Testing for Element Equality	53
5.4.5	Testing for <i>Exact</i> Equality	53
5.4.6	Identifying If Elements Are Contained in a String	54
5.4.7	Sorting a String	54
6	Dealing with Regular Expressions.....	55
6.1	Regex Syntax	55
6.1.1	Metacharacters	56
6.1.2	Sequences.....	56
6.1.3	Character Classes	57
6.1.4	POSIX Character Classes	58
6.1.5	Quantifiers.....	59
6.2	Regex Functions.....	60
6.2.1	Main Regex Functions in R	60
6.2.2	Regex Functions in <code>stringr</code>	63
6.3	Additional Resources	66
7	Dealing with Factors	67
7.1	Creating, Converting and Inspecting Factors.....	67
7.2	Ordering Levels	68
7.3	Revalue Levels.....	69
7.4	Dropping Levels.....	69
8	Dealing with Dates	71
8.1	Getting Current Date and Time.....	71
8.2	Converting Strings to Dates	72
8.2.1	Convert Strings to Dates	72
8.2.2	Create Dates by Merging Data.....	73
8.3	Extract and Manipulate Parts of Dates.....	73
8.4	Creating Date Sequences	75
8.5	Calculations with Dates	76
8.6	Dealing with Time Zones and Daylight Savings	77
8.7	Additional Resources	78
Part III Managing Data Structures in R		
9	Data Structure Basics	81
9.1	Identifying the Structure	81
9.2	Attributes.....	82

10	Managing Vectors.....	85
10.1	Creating Vectors.....	85
10.2	Adding On To Vectors.....	86
10.3	Adding Attributes to Vectors.....	87
10.4	Subsetting Vectors.....	88
10.4.1	Subsetting with Positive Integers.....	88
10.4.2	Subsetting with Negative Integers.....	88
10.4.3	Subsetting with Logical Values.....	89
10.4.4	Subsetting with Names.....	89
10.4.5	Simplifying vs. Preserving.....	89
11	Managing Lists.....	91
11.1	Creating Lists.....	91
11.2	Adding On To Lists.....	92
11.3	Adding Attributes to Lists.....	93
11.4	Subsetting Lists.....	95
11.4.1	Subset List and Preserve Output as a List.....	95
11.4.2	Subset List and Simplify Output.....	96
11.4.3	Subset List to Get Elements Out of a List.....	96
11.4.4	Subset List with a Nested List.....	96
12	Managing Matrices.....	99
12.1	Creating Matrices.....	99
12.2	Adding On To Matrices.....	100
12.3	Adding Attributes to Matrices.....	101
12.4	Subsetting Matrices.....	103
13	Managing Data Frames.....	105
13.1	Creating Data Frames.....	105
13.2	Adding On To Data Frames.....	107
13.3	Adding Attributes to Data Frames.....	109
13.4	Subsetting Data Frames.....	111
14	Dealing with Missing Values.....	113
14.1	Testing for Missing Values.....	113
14.2	Recoding Missing Values.....	114
14.3	Excluding Missing Values.....	114
 Part IV Importing, Scraping, and Exporting Data with R		
15	Importing Data.....	119
15.1	Reading Data from Text Files.....	119
15.1.1	Base R Functions.....	119
15.1.2	readr Package.....	122
15.2	Reading Data from Excel Files.....	123
15.2.1	xlsx Package.....	123
15.2.2	readxl Package.....	125

15.3	Load Data from Saved R Object File.....	127
15.4	Additional Resources.....	127
16	Scraping Data.....	129
16.1	Importing Tabular and Excel Files Stored Online	129
16.2	Scraping HTML Text	134
16.2.1	Scraping HTML Nodes	135
16.2.2	Scraping Specific HTML Nodes	139
16.2.3	Cleaning Up	141
16.3	Scraping HTML Table Data.....	143
16.3.1	Scraping HTML Tables with rvest	143
16.3.2	Scraping HTML Tables with XML.....	146
16.4	Working with APIs.....	150
16.4.1	Prerequisites?	150
16.4.2	Existing API Packages	151
16.4.3	httr for All Things Else.....	158
16.5	Additional Resources.....	162
17	Exporting Data.....	163
17.1	Writing Data to Text Files.....	163
17.1.1	Base R Functions.....	163
17.1.2	readr Package	164
17.2	Writing Data to Excel Files.....	165
17.2.1	xlsx Package.....	165
17.2.2	rexcel Package.....	167
17.3	Saving Data as an R Object File	169
17.4	Additional Resources.....	169

Part V Creating Efficient and Readable Code in R

18	Functions.....	173
18.1	Function Components	173
18.2	Arguments.....	174
18.3	Scoping Rules	175
18.4	Lazy Evaluation	177
18.5	Returning Multiple Outputs from a Function	177
18.6	Dealing with Invalid Parameters.....	178
18.7	Saving and Sourcing Functions	179
18.8	Additional Resources.....	181
19	Loop Control Statements.....	183
19.1	Basic Control Statements (i.e. if, for, while, etc.).....	183
19.1.1	if Statement	183
19.1.2	if...else Statement.....	184
19.1.3	for Loop.....	186
19.1.4	while Loop	187
19.1.5	repeat Loop.....	189

19.1.6	break Function to Exit a Loop.....	189
19.1.7	next Function to Skip an Iteration in a Loop.....	190
19.2	Apply Family	190
19.2.1	apply() for Matrices and Data Frames	191
19.2.2	lapply() for Lists...Output as a List	192
19.2.3	sapply() for Lists...Output Simplified	193
19.2.4	tapply() for Vectors	194
19.3	Other Useful “Loop-Like” Functions	195
19.4	Additional Resources	197
20	Simplify Your Code with %>%	199
20.1	Pipe (%>%) Operator.....	199
20.1.1	Nested Option.....	200
20.1.2	Multiple Object Option	200
20.1.3	%>% Option.....	201
20.2	Additional Functions.....	203
20.3	Additional Pipe Operators.....	204
20.4	Additional Resources	207
 Part VI Shaping and Transforming Your Data with R		
21	Reshaping Your Data with tidyr	211
21.1	Making Wide Data long	212
21.2	Making Long Data wide	213
21.3	Splitting a Single Column into Multiple Columns	213
21.4	Combining Multiple Columns into a Single Column	214
21.5	Additional tidyr Functions.....	215
21.6	Sequencing Your tidyr Operations.....	217
21.7	Additional Resources	218
22	Transforming Your Data with dplyr.....	219
22.1	Selecting Variables of Interest.....	220
22.2	Filtering Rows.....	221
22.3	Grouping Data by Categorical Variables	222
22.4	Performing Summary Statistics on Variables.....	223
22.5	Arranging Variables by Value	225
22.6	Joining Data Sets.....	226
22.7	Creating New Variables	228
22.8	Additional Resources	232
Index.....		233

Data Wrangling with R

Boehmke, B.

2016, XII, 238 p. 24 illus., 10 illus. in color., Softcover

ISBN: 978-3-319-45598-3