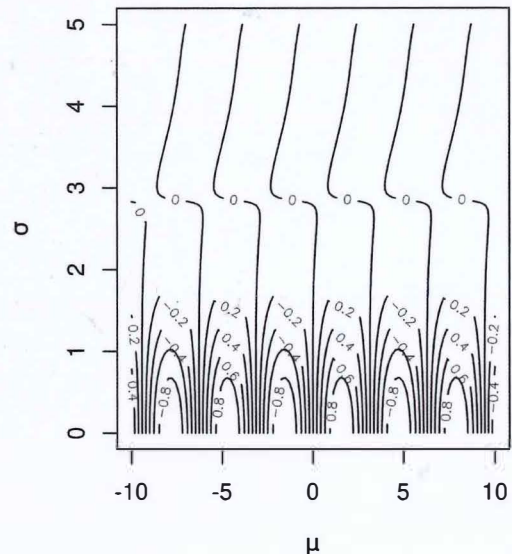
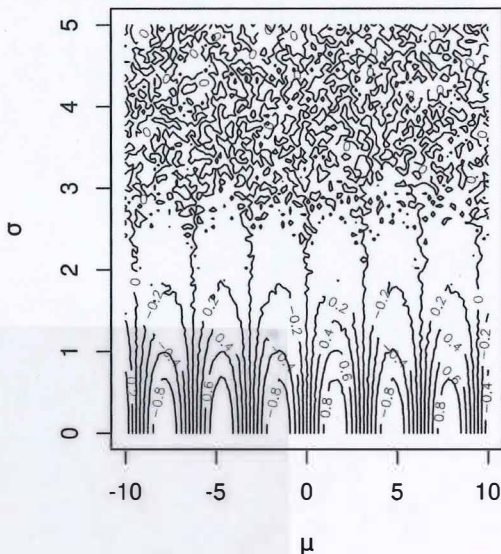


An Introduction to R and Python For Data Analysis

A Side-By-Side Approach



Taylor R. Brown

สำนักหอสมุด มหาวิทยาลัยเชียงใหม่

b16707163
012579075
12268802X

An Introduction to R and Python For Data Analysis

A Side-By-Side Approach

Taylor R. Brown



CRC Press

Taylor & Francis Group

Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A CHAPMAN & HALL BOOK

Contents

List of Figures	xiii
Welcome	xv
Preface	xvii
I Introducing the Basics	1
1 Introduction	3
1.1 Hello World in R	3
1.2 Hello World in Python	6
1.3 Getting Help	9
1.3.1 Reading Documentation	9
1.3.2 Understanding File Paths	9
2 Basic Types	11
2.1 Basic Types in Python	11
2.1.1 Type Conversions in Python	12
2.2 Basic Types in R	13
2.2.1 Type Conversions in R	13
2.2.2 R's Simplification	14
2.3 Exercises	14
2.3.1 R Questions	14
2.3.2 Python Questions	15
3 R vectors versus Numpy Arrays and Pandas' Series	17
3.1 Overview of R	17
3.2 Overview of Python	18
3.3 Vectorization in R	19
3.4 Vectorization in Python	20
3.5 Indexing Vectors in R	24
3.6 Indexing Numpy arrays	24
3.7 Indexing Pandas' Series	25
3.8 Some Gotchas	27

3.8.1	Shallow versus Deep Copies	27
3.8.2	How R and Python Handle Missing Values	30
3.9	An Introduction to Regular Expressions	32
3.9.1	Literal Characters versus Metacharacters	33
3.9.2	The Trouble with Backslashes: Escape Sequences	35
3.9.3	More Examples of Using Regular Expressions	36
3.10	Exercises	39
3.10.1	R Questions	39
3.10.2	Python Questions	42
4	Numpy ndarrays versus R's Matrix and Array Types	47
4.1	Numpy ndarrays in Python	47
4.2	The Matrix and Array Classes in R	48
4.3	Exercises	52
4.3.1	R Questions	52
4.3.2	Python Questions	54
5	R's lists versus Python's lists and dicts	59
5.1	lists in R	59
5.2	lists in Python	60
5.3	Dictionaries in Python	61
5.4	Exercises	63
5.4.1	R Questions	63
5.4.2	Python Questions	63
6	Functions	65
6.1	Defining R Functions	65
6.2	Defining Python Functions	66
6.3	More Details on R's User-Defined Functions	67
6.4	More Details on Python's User-Defined Functions	69
6.5	Function Scope in R	70
6.6	Function Scope in Python	73
6.7	Modifying a Function's Arguments	75
6.7.1	Passing by Value in R	75
6.7.2	Passing by Assignment in Python	76
6.8	Accessing and Modifying Captured Variables	79
6.8.1	Accessing Captured Variables in R	79
6.8.2	Accessing Captured Variables in Python	81
6.8.3	Modifying Captured Variables in R	81
6.8.4	Modifying Captured Variables in Python	82
6.9	Exercises	83
6.9.1	R Questions	83
6.9.2	Python Questions	84

7	Categorical Data	87
7.1	factors in R	87
7.2	Two Options for Categorical Data in Pandas	90
7.3	Exercises	94
7.3.1	R Questions	94
7.3.2	Python Questions	94
8	Data Frames	97
8.1	Data Frames in R	97
8.2	Data Frames in Python	100
8.3	Exercises	105
8.3.1	R Questions	105
8.3.2	Python Questions	107
II	Common Tasks and Patterns	111
9	Input and Output	113
9.1	General Input Considerations	113
9.2	Reading in Text Files with R	116
9.3	Reading in Text Files with Pandas	118
9.4	Saving Data in R	121
9.4.1	Writing Out Tabular Plain Text Data in R	122
9.4.2	Serialization in R	122
9.5	Saving Data in Python	124
9.5.1	Writing Out Tabular Plain Text Data in Python	124
9.5.2	Serialization in Python	124
9.6	Exercises	125
9.6.1	R Questions	125
9.6.2	Python Questions	126
10	Using Third-Party Code	127
10.1	Installing Packages in R	127
10.2	Installing Packages in Python	127
10.3	Loading Packages in R	128
10.4	Loading Packages in Python	130
10.4.1	Importing Examples	132
10.5	Exercises	134
11	Control Flow	137
11.1	Conditional Logic	137
11.2	Loops	138
11.3	Exercises	142

11.3.1	R Questions	142
11.3.2	Python Questions	145
12	Reshaping and Combining Data Sets	149
12.1	Ordering and Sorting Data	149
12.2	Stacking Data Sets and Placing Them Shoulder to Shoulder	151
12.3	Merging or Joining Data Sets	153
12.4	Long versus Wide Data	156
12.4.1	Long versus Wide in R	156
12.4.2	Long versus Wide in Python	159
12.5	Exercises	161
12.5.1	R Questions	161
12.5.2	Python Questions	164
13	Visualization	167
13.1	Base R Plotting	167
13.2	Plotting with <code>ggplot2</code>	170
13.3	Plotting with <code>Matplotlib</code>	175
13.4	Plotting with <code>Pandas</code>	179
13.5	Exercises	180
13.5.1	R Questions	180
13.5.2	Python Questions	181
III	Programming Styles	183
14	An Introduction to Object-Oriented Programming	185
14.1	OOP in Python	186
14.1.1	Overview	186
14.1.2	A First Example	187
14.1.3	Adding Inheritance	190
14.1.4	Adding in Composition	192
14.2	OOP in R	194
14.2.1	S3 Objects: The Big Picture	194
14.2.2	Using S3 Objects	195
14.2.3	Creating S3 Objects	199
14.2.4	S4 Objects: The Big Picture	200
14.2.5	Using S4 Objects	200
14.2.6	Creating S4 Objects	201
14.2.7	Reference Classes: The Big Picture	204
14.2.8	Creating Reference Classes	204
14.2.9	Creating R6 Classes	205
14.3	Exercises	206

14.3.1 Python Questions	206
14.3.2 R Questions	209

15 An Introduction to Functional Programming 213

15.1 Functions as Function Inputs in R	215
15.1.1 sapply() and vapply()	215
15.1.2 lapply()	216
15.1.3 apply()	217
15.1.4 tapply()	219
15.1.5 mapply()	220
15.1.6 Reduce() and do.call()	221
15.2 Functions as Function Inputs in Python	223
15.2.1 Functions as Function Inputs in Base Python	223
15.2.2 Functions as Function Inputs in Numpy	224
15.2.3 Functional Methods in Pandas	225
15.3 Functions as Function Outputs in R	228
15.4 Functions as Function Outputs in Python	233
15.4.1 Writing Our Own Decorators	234
15.5 Exercises	236
15.5.1 Python Questions	236
15.5.2 R Questions	238

Bibliography 241

Index 245