

Contents

List of Figures	xi
About the Companion Website	xii
1 Introduction	1
1.1 Why Program? Why Python?	1
1.2 Course Overview and Aims	4
1.3 A Brief Note on the Exercises	5
1.4 Conventions Used in this Book	6
1.5 Installing Python	6
1.5.1 Installing on Windows	6
1.5.2 Installing on the Mac	7
1.5.3 Installing on Linux	8
1.6 Introduction to the Command Line/Console/Terminal	8
1.6.1 Activating the Command Line on Windows	9
1.6.2 Activating the Command Line on the Mac or Linux	9
1.7 Editors and IDEs	10
1.8 Installing and Setting Up WingIDE Personal	10
1.9 Discussions	11
2 Programming Basics I	15
2.1 Statements, Functions, and Variables	15
2.2 Data Types – Overview	17
2.3 Simple Data Types	18
2.3.1 Strings	18
2.3.2 Numbers	20
2.3.3 Binary Switches/Values	21
2.4 Operators – Overview	21
2.4.1 String Operators	21
2.4.2 Mathematical Operators	22
2.4.3 Logical Operators	24
2.5 Creating Scripts/Programs	25
2.6 Commenting Your Code	26
2.7 Discussions	28

3	Programming Basics II	33
3.1	Compound Data Types	33
3.2	Lists	35
3.3	Simple Interaction with Programs and Users	37
3.4	Problem Solving and Damage Control	38
3.4.1	Getting Help from Your IDE	38
3.4.2	Using the Debugger	39
3.5	Control Structures	40
3.5.1	Conditional Statements	41
3.5.2	Loops	42
3.5.3	while Loops	43
3.5.4	for Loops	44
3.5.5	Discussions	45
4	Intermediate String Processing	53
4.1	Understanding Strings	53
4.2	Cleaning Up Strings	54
4.3	Working with Sequences	55
4.3.1	Overview	55
4.3.2	Slice Syntax	56
4.4	More on Tuples	57
4.5	'Concatenating' Strings More Efficiently	59
4.6	Formatting Output	60
4.6.1	Using the % Operator	60
4.6.2	The format Method	61
4.6.3	f-Strings	61
4.6.4	Formatting Options	62
4.7	Handling Case	62
4.8	Discussions	63
5	Working with Stored Data	71
5.1	Understanding and Navigating File Systems	71
5.1.1	Showing Folder Contents	72
5.1.2	Navigating and Creating Folders	74
5.1.3	Relative Paths	75
5.2	Stored Data	76
5.3	Opening and Closing Files	76
5.3.1	File Opening Modes	77
5.3.2	File Access Options	77
5.4	Reading File Contents	78
5.5	Error Handling	79
5.6	Writing to Files	82
5.7	Working with Folders and Paths	83
5.7.1	The os Module	83
5.7.2	The Path Object of the libpath Module	84
5.8	Discussions	86

6	Recognising and Working with Language Patterns	93
6.1	The re Module	93
6.2	General Syntax	94
6.3	Understanding and Working with the Match Object	94
6.4	Character Classes	96
6.5	Quantification	97
6.6	Masking and Using Special Characters	98
6.7	Regex Error Handling	98
6.8	Anchors, Groups and Alternation	99
6.9	Constraining Results Further	101
6.10	Compilation Flags	101
6.11	Discussions	102
7	Developing Modular Programs	109
7.1	Modularity	109
7.2	Dictionaries	109
7.3	User-defined Functions	111
7.4	Understanding Modules	112
7.5	Documenting Your Module	115
7.6	Installing External Modules	116
7.7	Classes and Objects	117
	7.7.1 Methods	118
	7.7.2 Class Schema	118
7.8	Testing Modules	119
7.9	Discussions	120
8	Word Lists, Frequencies and Ordering	129
8.1	Introduction to Word and Frequency Lists	129
8.2	Generating Word Lists	129
8.3	Sorting Basics	130
8.4	Generating Basic Word Frequency Lists	131
8.5	Lambda Functions	132
8.6	Discussions	134
9	Interacting with Data and Users Through GUIs	143
9.1	Graphical User Interfaces	143
9.2	PyQt Basics	144
	9.2.1 The General Approach to Designing GUI-based Programs	144
	9.2.2 Useful PyQt Widgets	145
	9.2.3 A Minimal PyQt Program	146
	9.2.4 Deriving from a Main Window	148
	9.2.5 Working with Layouts	148
	9.2.6 Defining Widgets and Assigning Layouts	150
	9.2.7 Widget Properties, Methods and Signals	150
	9.2.8 Adding Interactive Functionality	152

9.3	Designing More Advanced GUIs	153
9.3.1	Actions	153
9.3.2	Creating Menus, Tool and Status Bars	153
9.3.3	Working with Files and Folder in PyQt	155
9.4	Discussions	159
10	Web Data and Annotations	171
10.1	Markup Languages	171
10.2	Brief Intro to HTML	172
10.3	Using the <code>urllib.request</code> Module	174
10.4	Extracting Text from Web Pages	177
10.5	List and Dictionary Comprehension	178
10.6	Brief Intro to XML	179
10.7	Complex Regex Replacements Using Functions	182
10.8	Brief Intro to the TEI Scheme	182
10.8.1	The Header	183
10.8.2	The Text Body	184
10.9	Discussions	188
11	Basic Visualisation	201
11.1	Using Matplotlib for Basic Visualisation	201
11.2	Creating Word Clouds	207
11.3	Filtering Frequency Data Through Stop-Words	208
11.4	Working with Relative Frequencies	210
11.5	Comparing Frequency Data Visually	212
11.6	Discussions	216
12	Conclusion	227
	Appendix – Program Code	231
	Index	273