

# Contents

- Prerequisites: Tooling** **1**
- Python . . . . . 1
- Editor . . . . . 3
- Console (REPL) . . . . . 3
- Using Spyder and Keyboard Shortcuts . . . . . 5
- Anki . . . . . 5
- Remembering What You Learn . . . . . 5
- 
- 1. Introduction** **6**
- The Purpose of Data Analysis . . . . . 6
- What is Data? . . . . . 6
- Example Datasets . . . . . 8
- 100 Game Sample . . . . . 8
- 2018 Season Data . . . . . 10
- What is Analysis? . . . . . 12
- Types of Data Analysis . . . . . 13
- Summary Statistics . . . . . 13
- Modeling . . . . . 14
- High Level Data Analysis Process . . . . . 17
- 1. Collecting Data . . . . . 17
- 2. Storing Data . . . . . 17
- 3. Loading Data . . . . . 17
- 4. Manipulating Data . . . . . 17
- 5. Analyzing Data for Insights . . . . . 18
- Connecting the High Level Analysis Process to the Rest of the Book . . . . . 18
- End of Chapter Exercises . . . . . 19
- 
- 2. Python** **21**
- Introduction to Python Programming . . . . . 21
- How to Read This Chapter . . . . . 21

Important Parts of the Python Standard Library . . . . .	22
Comments . . . . .	22
Variables . . . . .	23
Types . . . . .	24
Interlude: How to Figure Things Out in Python . . . . .	25
Bools . . . . .	28
if statements . . . . .	29
Container Types . . . . .	30
Unpacking . . . . .	32
Loops . . . . .	32
Comprehensions . . . . .	34
Functions . . . . .	37
Libraries are Functions and Types . . . . .	42
The os Library and path . . . . .	42
End of Chapter Exercises . . . . .	43
<b>3. Pandas</b> . . . . .	<b>46</b>
Introduction to Pandas . . . . .	46
Types and Functions . . . . .	46
Things You Can Do with DataFrames . . . . .	47
How to Read This Chapter . . . . .	47
Part 1. DataFrame Basics . . . . .	48
Loading Data . . . . .	48
DataFrame Methods and Attributes . . . . .	49
Working with Subsets of Columns . . . . .	50
Indexing . . . . .	51
Outputting Data . . . . .	55
Exercises . . . . .	56
Part 2. Things you can do with DataFrames . . . . .	57
Introduction . . . . .	57
1. Modify or create new columns of data . . . . .	58
Modifying and creating columns are really the same thing . . . . .	58
Math and number columns . . . . .	59
String columns . . . . .	60
Bool Columns . . . . .	61
Applying Functions to Columns . . . . .	62
Dropping Columns . . . . .	63
Renaming Columns . . . . .	63

---

Missing Data in Columns . . . . .	65
Changing column types . . . . .	67
Review . . . . .	69
Exercises . . . . .	70
2. Use built-in Pandas functions that work on DataFrames . . . . .	71
Axis . . . . .	72
Summary Functions on Boolean Columns . . . . .	74
Other Misc Built-in Summary Functions . . . . .	76
Review . . . . .	77
Exercises . . . . .	78
3. Filter Observations . . . . .	79
loc . . . . .	79
Boolean indexing . . . . .	80
Duplicates . . . . .	82
Combining filtering with changing columns . . . . .	84
Query . . . . .	85
Review . . . . .	86
Exercises . . . . .	87
4. Change Granularity . . . . .	88
Ways of changing granularity . . . . .	88
Grouping . . . . .	88
A note on multilevel indexing . . . . .	91
Stacking and unstacking data . . . . .	92
Review . . . . .	93
Exercises . . . . .	94
5. Combining two or more DataFrames . . . . .	95
1. The columns you're joining on. . . . .	95
2. Whether you're doing a one-to-one, one-to-many, or many-to-many merge . . . . .	97
3. What you do with unmatched observations . . . . .	99
More on pd.merge . . . . .	100
pd.concat() . . . . .	102
Review . . . . .	104
Exercises . . . . .	105
<b>4. SQL</b> . . . . .	<b>106</b>
Introduction to SQL . . . . .	106
How to Read This Chapter . . . . .	106

Databases . . . . .	106
SQL Databases . . . . .	108
A Note on NoSQL . . . . .	108
SQL . . . . .	109
Pandas . . . . .	109
Creating Data . . . . .	109
Queries . . . . .	110
Filtering . . . . .	112
Joining, or Selecting From Multiple Tables . . . . .	114
Misc SQL . . . . .	119
SQL Example — LEFT JOIN, UNION, Subqueries . . . . .	120
End of Chapter Exercises . . . . .	124
<b>5. Web Scraping</b>	<b>125</b>
Introduction to Web Scraping . . . . .	125
Web Scraping . . . . .	125
HTML . . . . .	126
BeautifulSoup . . . . .	128
Baseball Almanac Opening Day History - Web Scraping Example . . . . .	130
<b>6. Distributions and Visualizing them in Python with Seaborn</b>	<b>137</b>
Introduction . . . . .	137
What is a Distribution? . . . . .	138
Summary Stats: Info About Distributions . . . . .	141
Introduction . . . . .	141
Percentiles . . . . .	141
Average = Expected Value . . . . .	142
Dispersion . . . . .	142
Density Plots in Python . . . . .	143
Seaborn . . . . .	143
Plot Options . . . . .	151
<b>7. Intermediate Coding and Next Steps: High Level Strategies</b>	<b>156</b>
Gall’s Law . . . . .	156
Get Quick Feedback . . . . .	157
Use Functions . . . . .	158
DRY: Don’t Repeat Yourself . . . . .	158
Functions Help You Think Less . . . . .	158
Attitude . . . . .	159

Review and Conclusion . . . . .	161
<b>Appendix A: Answers to End of Chapter Exercises</b>	<b>162</b>
1. Introduction . . . . .	162
2. Python . . . . .	164
3.0 Pandas Basics . . . . .	169
3.1 Columns . . . . .	171
3.2 Built-in Functions . . . . .	175
3.3 Filtering . . . . .	177
3.4 Granularity . . . . .	180
3.5 Combining DataFrames . . . . .	183
4. SQL . . . . .	185
<b>Appendix B: Anki</b>	<b>186</b>
Remembering What You Learn . . . . .	186
Installing Anki . . . . .	187
Using Anki with this Book . . . . .	188