Contents

Prologue xvii

Acknowledgments xxxi

Notation xxxv

1 Introduction 1
  1.1 Theory ...................................................... 3
  1.2 Data ......................................................... 4
  1.3 At the End of the Day ....................................... 7

2 Productivity Tools 9
  2.1 Opening a Terminal Window ................................. 9
  2.2 Working on the Command Line ............................... 10
  2.3 Some UNIX Commands ...................................... 10
    2.3.1 Getting Your Bearings ................................. 11
    2.3.2 Learning about Commands .............................. 12
    2.3.3 Seeing What’s There .................................. 12
    2.3.4 Filenames .............................................. 14
    2.3.5 Reserved Characters .................................. 15
    2.3.6 Case Sensitivity ...................................... 16
    2.3.7 Redirecting Output ................................... 17
    2.3.8 Examining the Contents of a File .................... 18
    2.3.9 Comparing Files ...................................... 20
    2.3.10 Pathing ............................................... 20
    2.3.11 Changing Locations .................................. 21
    2.3.12 Creating Directories and Subdirectories ............ 22
CONTENTS

2.14 Archiving Files .................................................. 56
2.15 Version Control ................................................... 57
2.16 Package Managers ............................................... 58
2.17 UNIX File Systems ............................................... 58
2.18 Uniform Resource Identifiers ................................. 60

3 Organizing Data ..................................................... 61
  3.1 Spreadsheet ...................................................... 61
  3.2 Data Modeling .................................................... 63
    3.2.1 Entity-Relationship Model .............................. 64
    3.2.2 Database Normalization ................................. 66
  3.3 Relational Algebra ............................................... 67
  3.4 Basic SQL .......................................................... 75
  3.5 Solved Example ................................................... 82
    3.5.1 Designing Databases and Tables ...................... 83
    3.5.2 Using Excel to Create Tables for SQLite .......... 84
    3.5.3 Creating a SQLite Database and Its Tables ....... 86
    3.5.4 Importing csv Files into SQLite .................... 93
    3.5.5 Querying the RDB ........................................ 94
  3.6 NoSQL ............................................................. 109
    3.6.1 XML .......................................................... 117
    3.6.2 JSON ........................................................ 118
    3.6.3 YAML and BSON ........................................... 119

4 Simple Programming .................................................. 121
  4.1 Python ............................................................ 121
    4.1.1 IDE or Not? ............................................... 123
    4.1.2 Useful Website ............................................ 124
  4.2 Important Concepts in Computer Science ................. 124
  4.3 Basic Grammar .................................................. 125
    4.3.1 Version 2 or 3? ........................................... 125
    4.3.2 Classic Exercise ......................................... 125
    4.3.3 Data Types ................................................. 126
    4.3.4 Python Backslash Characters ........................ 147
    4.3.5 A Digression on Character Sets .................... 147
    4.3.6 Single or Double or Triple Quotes? ............... 149
CONTENTS

4.3.7 Functions ........................................ 151
4.3.8 Input/Output .................................... 154
4.3.9 Loops ............................................. 158
4.3.10 Conditional Statements .......................... 159
4.3.11 While .......................................... 160
4.3.12 Indentation, Whitespaces, and Tabs ............. 161
4.3.13 Exceptions ...................................... 164
4.3.14 Recursion ....................................... 165
4.3.15 Keywords Not Yet Introduced ..................... 166
4.3.16 Modules ......................................... 166
4.3.17 Packages ........................................ 169
4.3.18 Different Ways to Execute Python Scripts ....... 170

4.4 Useful Modules and Packages ...................... 172
4.4.1 copy .............................................. 173
4.4.2 math ............................................. 173
4.4.3 numpy ........................................... 173
4.4.4 matplotlib ....................................... 173
4.4.5 pandas ......................................... 174
4.4.6 scipy ............................................ 174
4.4.7 ipython ......................................... 175
4.4.8 sys ............................................... 175
4.4.9 os ............................................... 175
4.4.10 csv ............................................. 176
4.4.11 json ............................................. 176
4.4.12 sqlite3 ......................................... 177
4.4.13 re ............................................... 177
4.4.14 nltk ............................................. 177
4.4.15 urllib and urllib2; requests ..................... 177
4.4.16 distutils ....................................... 178
4.4.17 f2py ............................................ 178
4.4.18 numba .......................................... 178
4.4.19 xml ............................................. 178
4.4.20 Tkinter ......................................... 179
4.4.21 Abbreviating Module and Package Names ....... 179

4.5 Python Template ..................................... 179
4.6 Design Documents, Flowcharts, and Unit Testing .... 181
## CONTENTS

4.7 Miscellaneous Topics .................................................. 183  
4.8 Bringing It All Together .............................................. 184  
  4.8.1 Reading and Writing Numeric Data .............................. 185  
  4.8.2 Reading in Mixed Numeric and String Data .................. 189  
  4.8.3 Creating a Histogram and an edf .............................. 192  
  4.8.4 Creating a Figure with \TeX Characters ....................... 197  
  4.8.5 Creating a Scatterplot ........................................ 199  
  4.8.6 Creating a \LaTeX Table ...................................... 202  
  4.8.7 Creating a SQLite Database .................................. 207  
  4.8.8 Downloading Data from the Internet ......................... 209  
  4.8.9 Manipulating Text Using Regular Expressions ............... 210  

5 Analyzing Data ..................................................................... 225  
  5.1 Is Your Answer Right? ................................................. 225  
  5.2 Methods of Sampling Data ........................................... 228  
    5.2.1 Opportunity Sampling ........................................ 229  
    5.2.2 Prospective Sampling ......................................... 230  
    5.2.3 Random Sampling ............................................. 231  
    5.2.4 Choice-Based Sampling ...................................... 231  
  5.3 Useful Data Formats ................................................... 232  
  5.4 R System ....................................................................... 233  
    5.4.1 Getting R and Its Packages .................................. 234  
    5.4.2 RStudio IDE ..................................................... 234  
    5.4.3 Basic R Grammar ............................................... 235  
    5.4.4 Types of R Objects ............................................ 242  
    5.4.5 Reading in Data ............................................... 250  
    5.4.6 Descriptive Statistics .......................................... 253  
    5.4.7 Flow Control and Loops ...................................... 259  
    5.4.8 Figures and Graphs ............................................ 261  
    5.4.9 Regressions ...................................................... 267  
    5.4.10 Batch Scripts .................................................. 272  
  5.5 Useful R Packages ...................................................... 276  
    5.5.1 Reading in Data ................................................ 277  
    5.5.2 Manipulating Data .............................................. 277  
    5.5.3 Plotting Figures ................................................ 282  
    5.5.4 Time-Series Data .............................................. 282
5.5.5 Improving Code Performance ........................................ 283
5.5.6 Estimating Various Models ....................................... 287
5.5.7 Reporting Results .................................................. 288
5.5.8 Other Packages ...................................................... 289
5.6 Connecting R to SQLite ................................................. 289
5.7 Python Library pandas ............................................... 292
5.8 Python or R? ............................................................. 301
5.9 Training, Validation, and Testing ................................. 302
5.9.1 Precision and Recall; ROC Curves ............................. 303
5.10 Fixed-Effect Regressions ............................................. 308
5.10.1 Least-Squares Estimator of $\theta$ ............................ 311
5.10.2 What to Do with It All ........................................... 312

6 Geek Stuff 317
6.1 Hardware ................................................................. 317
6.1.1 What Does It All Mean? ....................................... 323
6.1.2 Raspberry Pi ........................................................ 323
6.2 Algorithmics ............................................................. 325
6.2.1 Analysis and Evaluation ....................................... 327
6.2.2 Sorting Algorithms .............................................. 333
6.2.3 Complexity Classes .............................................. 334
6.2.4 Exploiting Complexity in Computer Security .......... 337
6.2.5 What Does It Mean? ............................................. 338
6.2.6 Further Reading ................................................... 339
6.2.7 Approaches to Algorithmic Design .......................... 339
6.3 Some Programming Paradigms ................................. 345
6.3.1 Imperative Programming ...................................... 345
6.3.2 Procedural Programming ...................................... 346
6.3.3 Declarative Programming ...................................... 347
6.3.4 Object-Oriented Programming ............................... 348
6.3.5 Functional Programming ...................................... 352
6.3.6 Programming Languages and Paradigms .................. 353
6.4 Graph Theory ............................................................ 353
6.4.1 Some Theorems .................................................... 357
7 Numerical Methods

7.1 Round-off and Truncation Errors

7.1.1 Classic Example of Smearing

7.1.2 Summary

7.2 Linear Algebra

7.2.1 Condition Number

7.2.2 Solving a Linear System

7.2.3 Cholesky Decomposition

7.3 Finding the Zero of a Function

7.3.1 Bisection Method

7.3.2 Newton-Raphson Method

7.4 Solving Systems of Nonlinear Equations

7.4.1 Newton-Raphson Method

7.4.2 Jacobi Method

7.4.3 Gauss-Seidel Method

7.4.4 Using the Methods

7.4.5 Solving Nonlinear Equations Using Python

7.5 Unconstrained Optimization

7.5.1 Newton-Raphson Method

7.5.2 Quasi-Newton Methods

7.5.3 Line Search versus Trust Region Methods

7.5.4 Adjusting a Hessian Matrix

7.5.5 Scaling

7.5.6 Gradient Descent

7.5.7 Conjugate Gradient

7.5.8 Stochastic Gradient Descent

7.5.9 Derivative-Free Methods

7.5.10 Numerical Optimization in Python

7.6 Constrained Optimization

7.6.1 Linear Programming

7.6.2 Dual Representation

7.6.3 Quadratic Programming

7.6.4 Convex Optimization

7.6.5 Nonlinear Programming

7.7 Approximation Methods

7.8 Numerical Integration
CONTENTS

7.8.1 Newton-Cotes Formulae ........................................... 450
7.8.2 Monte Carlo Methods ............................................ 452
7.8.3 Quasi-Monte Carlo Methods ................................. 456
7.8.4 Gaussian Quadrature ............................................ 456
7.9 Solving Differential Equations ................................. 465
  7.9.1 Initial- and Boundary-Value Problems ................. 467
  7.9.2 Finite Difference Methods ................................ 468
  7.9.3 Finite Element Methods .................................... 475
7.10 Simulation ............................................................ 477
  7.10.1 Distribution of the cdf ..................................... 477
  7.10.2 Generating Random Numbers ............................. 479
  7.10.3 Pseudo-Random Numbers .................................. 480
  7.10.4 Seeding the PRNG ............................................ 482
  7.10.5 Introducing Dependence .................................... 483
  7.10.6 Antithetic Variates ......................................... 486
  7.10.7 Control Variates ............................................. 487
  7.10.8 Importance Sampling ....................................... 490
  7.10.9 Markov Chain Monte Carlo ............................... 491
7.11 Figures and Graphs ................................................. 496

8 Solved Examples .......................................................... 499
  8.1 Linear Algebra: Portfolio Allocation Problem ............ 499
  8.2 Unconstrained Optimization: Duration Model ........... 504
    8.2.1 Putting Structure on $f_T(t)$ ........................... 504
    8.2.2 Loosening the Structure on $f_T(t)$ ................. 510
    8.2.3 Cox Proportional Hazard Rate Model ............... 512
    8.2.4 Training the Model ....................................... 515
    8.2.5 Putting It All Together ................................ 517
  8.3 Linear Programming: LAD-Lasso Estimator .................. 521
  8.4 Quadratic Programming: Support Vector Machines ......... 527
    8.4.1 Hinge Loss Function ..................................... 531
    8.4.2 Support Vector Machines ................................. 532
    8.4.3 Implementing SVM in Python ............................ 534
    8.4.4 Alternative Solution Strategies ....................... 538
  8.5 Numerical Integration: Gauss-Hermite Quadrature ........ 539
  8.6 Simulation: Demand for Change ............................... 542
## CONTENTS

8.7 Resampling: Quantifying Variability ........................................ 548  
  8.7.1 First-Order Asymptotic Methods ..................................... 549  
  8.7.2 Bootstrap ................................................................. 556  
  8.7.3 Jackknife ................................................................. 561  
  8.7.4 Subsampling ............................................................. 563  
8.8 Makefile: Dealing with Dependencies ..................................... 564  
8.9 Git: Version Control ......................................................... 568  
  8.9.1 Theory ......................................................................... 570  
  8.9.2 Example ....................................................................... 574  

9 Extensions to Python ................................................................. 589  
  9.1 Profiling Python Code ......................................................... 592  
  9.2 C Programming Language .................................................... 593  
    9.2.1 Basic Grammar .......................................................... 594  
  9.3 C Extensions to Python ....................................................... 614  
  9.4 FORTRAN Programming Language ...................................... 617  
    9.4.1 Basic Grammar .......................................................... 617  
  9.5 FORTRAN Extensions to Python ............................................ 628  
  9.6 Numba ........................................................................... 629  

10 Papers and Presentations ......................................................... 631  
  10.1 LATEX ......................................................................... 633  
    10.1.1 Notation ................................................................. 638  
  10.2 BIBTEX ........................................................................ 641  
  10.3 Beamer ......................................................................... 649  
  10.4 Incorporating PGF/TikZ Figures ............................................ 656  
  10.5 Other TEX/LATEX Tricks .................................................... 658  
  10.6 ConTEXt ....................................................................... 658  

11 Final Thoughts ................................................................. 661  
  11.1 Amdahl’s Law ................................................................. 663  
  11.2 MapReduce ................................................................. 664  
  11.3 Summary ................................................................. 668  

Appendices ..................................................................................... 669  

A The Virtual Machine ................................................................. 671
CONTENTS

B Recommended Reading 675

References 681

About the Authors 695

Name Index 697

Subject Index 703