

## **Beginning Python Visualization: Crafting Visual Transformation Scripts**

**Copyright © 2009 by Shai Vaingast**

All rights reserved. No part of this work may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage or retrieval system, without the prior written permission of the copyright owner and the publisher.

ISBN-13 (pbk): 978-1-4302-1843-2

ISBN-13 (electronic): 978-1-4302-1844-9

Printed and bound in the United States of America 9 8 7 6 5 4 3 2 1

Trademarked names may appear in this book. Rather than use a trademark symbol with every occurrence of a trademarked name, we use the names only in an editorial fashion and to the benefit of the trademark owner, with no intention of infringement of the trademark.

Lead Editors: Frank Pohlmann, Michelle Lowman

Technical Reviewer: C. Titus Brown

Editorial Board: Clay Andres, Steve Anglin, Mark Beckner, Ewan Buckingham, Tony Campbell,

Gary Cornell, Jonathan Gennick, Michelle Lowman, Matthew Moodie, Jeffrey Pepper,

Frank Pohlmann, Ben Renow-Clarke, Dominic Shakeshaft, Matt Wade, Tom Welsh

Project Manager: Kylie Johnston

Copy Editor: Ami Knox

Associate Production Director: Kari Brooks-Copony

Production Editor: Kelly Winkquist

Compositor: Dina Quan

Proofreader: Liz Welch

Indexer: Julie Grady

Artist: April Milne

Cover Designer: Kurt Krames

Manufacturing Director: Tom Debolski

Distributed to the book trade worldwide by Springer-Verlag New York, Inc., 233 Spring Street, 6th Floor, New York, NY 10013. Phone 1-800-SPRINGER, fax 201-348-4505, e-mail [orders-ny@springer-sbm.com](mailto:orders-ny@springer-sbm.com), or visit <http://www.springeronline.com>.

For information on translations, please contact Apress directly at 2855 Telegraph Avenue, Suite 600, Berkeley, CA 94705. Phone 510-549-5930, fax 510-549-5939, e-mail [info@apress.com](mailto:info@apress.com), or visit <http://www.apress.com>.

Apress and friends of ED books may be purchased in bulk for academic, corporate, or promotional use. eBook versions and licenses are also available for most titles. For more information, reference our Special Bulk Sales—eBook Licensing web page at <http://www.apress.com/info/bulksales>.

The information in this book is distributed on an “as is” basis, without warranty. Although every precaution has been taken in the preparation of this work, neither the author(s) nor Apress shall have any liability to any person or entity with respect to any loss or damage caused or alleged to be caused directly or indirectly by the information contained in this work.

The source code for this book is available to readers at <http://www.apress.com>.

# Contents

About the Author .....	xv
About the Technical Reviewer .....	xvi
Acknowledgments .....	xvii
Introduction .....	xviii

<b>CHAPTER 1</b>	<b>Navigating the World of Data Visualization .....</b>	<b>1</b>
	Gathering Data .....	2
	Case Study: GPS Data .....	2
	Scanning Serial Ports .....	3
	Recording GPS Data .....	5
	Data Organization .....	6
	File Format .....	6
	File Naming Conventions .....	7
	Data Location .....	7
	Data Analysis .....	8
	Walking Directories .....	8
	Reading CSV Files .....	9
	Analyzing GPS Data .....	12
	Extracting GPS Data .....	14
	Data Visualization .....	17
	GPS Location Plot .....	18
	Annotating the Graph .....	20
	Velocity Plot .....	22
	Subplots .....	23
	Text .....	23
	Tying It All Together .....	25
	Final Notes and References .....	29

<b>CHAPTER 2</b>	<b>The Environment</b>	31
	Operating Systems	32
	GNU/Linux	32
	Windows	33
	Choosing an Operating System	35
	Then Again, Why Choose? Using Several Operating Systems	36
	The Python Environment	37
	Versions	37
	Python	38
	Python Integrated Development Environments	39
	Scientific Computing	41
	Plotting	42
	Image Processing	43
	Additional Python Packages	43
	Installation Summary	44
	Additional Applications	45
	Editors	45
	A Short List of Text Editors	47
	Spreadsheets	48
	Word Processors	48
	Image Viewers	49
	Version Control Systems	49
	Licensing	51
	Final Notes and References	52
<b>CHAPTER 3</b>	<b>Python for Programmers</b>	53
	What Is Python?	53
	Interactive Python	54
	Invoking Python	54
	Entering Commands	55
	The Interactive Help System	56
	Moving Around	57
	Running Scripts	58
	Data Types	60
	Numbers	60
	Strings	65
	Booleans	67

Data Structures .....	68
Lists .....	69
Tuples .....	72
Dictionaries .....	74
Sets .....	78
Variables .....	80
Statements .....	81
Printing .....	81
User Input .....	84
Comments .....	85
Flow Control .....	85
Some Built-in Functions .....	92
Defining Functions .....	93
Generators .....	94
Generator Expressions .....	95
Object-Oriented Programming .....	96
Modules and Packages .....	97
The import Statement .....	98
Modules Installed in a System .....	99
The dir Statement .....	99
Final Notes and References .....	99

## CHAPTER 4    **Data Organization** ..... 101

File Name Conventions .....	102
Date and Time in a File Name .....	102
Useful File Name Titles .....	104
File Name Extensions .....	104
In Conclusion .....	105
Other Schemes .....	107
File Formats .....	108
CSV File Format .....	109
Binary Files .....	117
Readme Files .....	123
INI Files .....	123
XML .....	125
Other File Formats .....	126
Locating Data Files .....	126
Organization into Directories .....	126
Searching for Files .....	127
Indexing .....	128

Catalogs.....	131
Files vs. a Database.....	133
Final Notes and References.....	134

## CHAPTER 5    **Processing Text Files** ..... 135

Text and Strings.....	136
Splitting Text.....	136
Joining Strings.....	137
Converting Strings to Numbers.....	137
Find and Replace.....	143
Stripping Strings.....	144
String Formatting.....	145
String Conditionals.....	146
More on Strings.....	147
Files.....	147
Opening a File.....	147
Closing a File.....	148
Writing Text.....	148
Reading Text.....	149
Working with Text Files.....	150
Example: Character, Word, and Line Count.....	151
Example: head and tail.....	152
Example: Splitting and Combining Files.....	153
Example: Searching Inside a Text File.....	155
Example: Working with Comments.....	156
Example: Extracting Numbers from a Text File.....	157
CSV Files.....	159
The csv Module.....	159
The csv.reader Object.....	160
The csv.writer Object.....	161
More csv Functionality.....	161
DictReader and DictWriter Objects.....	162
Date and Time.....	163
Time Module.....	164
The struct_time Tuple.....	165
Parsing and Formatting Date and Time.....	165
The Epoch: “Linearizing” the Time Base.....	168
Additional Time and Date Functions.....	173

Regular Expressions .....	173
Regular Expression Patterns .....	173
Special Sequences .....	175
Alternatives .....	175
Ranges .....	175
When to Use Regular Expressions .....	175
Internationalization and Localization .....	176
Locale .....	177
Unicode Strings .....	178
Final Notes and References .....	181

## CHAPTER 6   **Graphs and Plots** .....

The Matplotlib Package .....	183
Interactive Graphs vs. Image Files .....	184
Interactive Graphs .....	185
Saving Graphs to Files .....	187
Plotting Graphs .....	189
Lines and Markers .....	189
Plotting Several Graphs on One Figure .....	191
Line Widths and Marker Sizes .....	192
Colors .....	193
Controlling the Graph .....	194
Axis .....	194
Grid and Ticks .....	195
Subplots .....	196
Erasing the Graph .....	197
Adding Text .....	197
Title .....	198
Axis Labels and Legend .....	198
Text Rendering .....	199
Mathematical Symbols and Expressions .....	200
More Graph Types .....	201
Bar Charts .....	201
Histograms .....	204
Pie Charts .....	206
Logarithmic Plots .....	207
Polar Plots .....	208
Stem Plots .....	209
Additional Graphs .....	210

Getting and Setting Values . . . . .	213
Setting Figure and Axis Parameters . . . . .	215
Patches . . . . .	217
Example: Adding Arrows to a Graph . . . . .	218
Example: Some Other Patches . . . . .	219
Final Notes and References . . . . .	220

## CHAPTER 7    **Math Games** . . . . . 221

Modules math and cmath . . . . .	221
Example: A Newton Fractal . . . . .	224
Module random . . . . .	228
Using random to Solve Probability Questions . . . . .	229
Random Sequences . . . . .	232
Module NumPy . . . . .	233
Array Creation . . . . .	234
Slicing, Indexing, and Reshaping . . . . .	235
N-Dimensional Arrays . . . . .	236
Math Functions . . . . .	239
Array Methods and Properties . . . . .	241
Other Useful Array Functions . . . . .	247
Final Notes and References . . . . .	247

## CHAPTER 8    **Science and Visualization** . . . . . 249

Finding Your Way: Variables and Functions . . . . .	250
SciPy . . . . .	250
Linear Algebra . . . . .	251
Solving a System of Linear Equations . . . . .	251
Vector and Matrix Operations . . . . .	252
Matrix Decomposition . . . . .	253
Additional Linear Algebra Functionality . . . . .	254
Numerical Integration . . . . .	254
More Integration Methods . . . . .	257
Interpolation and Curve Fitting . . . . .	258
Piecewise Linear Interpolation . . . . .	258
Polynomials . . . . .	260
Uses of Polynomials . . . . .	261
Spline Interpolation . . . . .	266

Solving Nonlinear Equations .....	267
Special Functions .....	268
Signal Processing .....	268
Functions where, select, and find .....	269
Functions diff and split .....	273
Waveforms .....	274
Fourier Transform .....	275
Example: FFT of a Sampled Cosine Wave .....	276
Window Functions .....	277
Filtering .....	279
Filter Design .....	279
Example: Heart-Rate Monitor .....	281
Example: Moving Average .....	283
Final Notes and References .....	284

<b>CHAPTER 9</b>	<b>Image Processing .....</b>	<b>285</b>
	Reading, Writing, and Displaying Images .....	286
	Reading Images from File .....	286
	Image Attributes .....	287
	Displaying Images .....	288
	Converting File Formats .....	289
	Image Manipulation .....	291
	Creating New Images .....	291
	Copy and Paste .....	292
	Crop and Resize .....	292
	Rotate .....	293
	Image Annotation .....	294
	Annotating with Geometrical Shapes .....	294
	Text Annotations .....	295
	Image Processing .....	300
	Matrix Representation and Colors .....	300
	Example: Counting Objects (Five Parts) .....	303
	Image Arithmetic .....	312
	Image Filtering .....	315
	Final Notes and References .....	317



<b>CHAPTER 10</b>	<b>Advanced File Processing</b>	319
	Binary Files and Random Access	319
	Example: Reading the Nth Field	321
	Example: Efficient Tail Implementation	322
	Example: Creating a Fixed-Size File	323
	Example: Recording Time-Based Binary Data	323
	Object Serialization	325
	The Pickle Module	325
	Command-Line Parameters	327
	argv	327
	Example: Creating a Fixed-Size File (Stand-Alone Script)	328
	OptParse Module	329
	The FileInput Module	332
	File and Directory Manipulation	333
	Module glob	334
	Additional os Module Functionality	334
	Additional os.path Module Functionality	335
	Module shutil	336
	File Compression	337
	Example: A Compressed tar File	338
	Comparing Files	339
	Module filecmp	339
	Module difflib	341
	Final Notes and References	342
<b>APPENDIX</b>	<b>Additional Source Listing</b>	343
	Nudge Subplots	343
	Magic Square Arrows	345
	Fractal Function Source Code	347
<b>INDEX</b>		349