



Community Experience Distilled

# Python Data Analysis

Learn how to apply powerful data analysis techniques with popular open source Python modules

Ivan Idris

**[PACKT]** open source\*  
PUBLISHING community experience distilled

# Table of Contents

---

[Python Data Analysis](#)

[Credits](#)

[About the Author](#)

[About the Reviewers](#)

[www.PacktPub.com](#)

[Support files, eBooks, discount offers, and more](#)

[Why subscribe?](#)

[Free access for Packt account holders](#)

[Preface](#)

[What this book covers](#)

[What you need for this book](#)

[Who this book is for](#)

[Conventions](#)

[Reader feedback](#)

[Customer support](#)

[Downloading the example code](#)

[Errata](#)

[Piracy](#)

[Questions](#)

[1. Getting Started with Python Libraries](#)

[Software used in this book](#)

[Installing software and setup](#)

[On Windows](#)

[On Linux](#)

[On Mac OS X](#)

[Building NumPy, SciPy, matplotlib, and IPython from source](#)

[Installing with setuptools](#)

[NumPy arrays](#)

[A simple application](#)

[Using IPython as a shell](#)

[Reading manual pages](#)

[IPython notebooks](#)

[Where to find help and references](#)

[Summary](#)

[2. NumPy Arrays](#)

[The NumPy array object](#)

[The advantages of NumPy arrays](#)

[Creating a multidimensional array](#)

[Selecting NumPy array elements](#)

[NumPy numerical types](#)

[Data type objects](#)

[Character codes](#)

[The dtype constructors](#)

[The dtype attributes](#)

[One-dimensional slicing and indexing](#)

[Manipulating array shapes](#)

[Stacking arrays](#)

[Splitting NumPy arrays](#)

[NumPy array attributes](#)

[Converting arrays](#)

[Creating array views and copies](#)

[Fancy indexing](#)

[Indexing with a list of locations](#)

[Indexing NumPy arrays with Booleans](#)

[Broadcasting NumPy arrays](#)

[Summary](#)

### [3. Statistics and Linear Algebra](#)

[NumPy and SciPy modules](#)

[Basic descriptive statistics with NumPy](#)

[Linear algebra with NumPy](#)

[Inverting matrices with NumPy](#)

[Solving linear systems with NumPy](#)

[Finding eigenvalues and eigenvectors with NumPy](#)

[NumPy random numbers](#)

[Gambling with the binomial distribution](#)

[Sampling the normal distribution](#)

[Performing a normality test with SciPy](#)

[Creating a NumPy-masked array](#)

[Disregarding negative and extreme values](#)

[Summary](#)

### [4. pandas Primer](#)

[Installing and exploring pandas](#)

[pandas DataFrames](#)

[pandas Series](#)

[Querying data in pandas](#)

[Statistics with pandas DataFrames](#)

[Data aggregation with pandas DataFrames](#)

[Concatenating and appending DataFrames](#)

[Joining DataFrames](#)

[Handling missing values](#)

[Dealing with dates](#)

[Pivot tables](#)

[Remote data access](#)

[Summary](#)

### [5. Retrieving, Processing, and Storing Data](#)

[Writing CSV files with NumPy and pandas](#)

[Comparing the NumPy .npy binary format and pickling pandas DataFrames](#)

[Storing data with PyTables](#)

---

[Reading and writing pandas DataFrames to HDF5 stores](#)

[Reading and writing to Excel with pandas](#)

[Using REST web services and JSON](#)

[Reading and writing JSON with pandas](#)

[Parsing RSS and Atom feeds](#)

[Parsing HTML with Beautiful Soup](#)

[Summary](#)

## [6. Data Visualization](#)

[matplotlib subpackages](#)

[Basic matplotlib plots](#)

[Logarithmic plots](#)

[Scatter plots](#)

[Legends and annotations](#)

[Three-dimensional plots](#)

[Plotting in pandas](#)

[Lag plots](#)

[Autocorrelation plots](#)

[Plot.ly](#)

[Summary](#)

## [7. Signal Processing and Time Series](#)

[statsmodels subpackages](#)

[Moving averages](#)

[Window functions](#)

[Defining cointegration](#)

[Autocorrelation](#)

[Autoregressive models](#)

[ARMA models](#)

[Generating periodic signals](#)

[Fourier analysis](#)

[Spectral analysis](#)

[Filtering](#)

[Summary](#)

## [8. Working with Databases](#)

[Lightweight access with sqlite3](#)

[Accessing databases from pandas](#)

[SQLAlchemy](#)

[Installing and setting up SQLAlchemy](#)

[Populating a database with SQLAlchemy](#)

[Querying the database with SQLAlchemy](#)

[Pony ORM](#)

[Dataset – databases for lazy people](#)

[PyMongo and MongoDB](#)

[Storing data in Redis](#)

[Apache Cassandra](#)

[Summary](#)

---

## [9. Analyzing Textual Data and Social Media](#)

[Installing NLTK](#)

[Filtering out stopwords, names, and numbers](#)

[The bag-of-words model](#)

[Analyzing word frequencies](#)

[Naive Bayes classification](#)

[Sentiment analysis](#)

[Creating word clouds](#)

[Social network analysis](#)

[Summary](#)

## [10. Predictive Analytics and Machine Learning](#)

[A tour of scikit-learn](#)

[Preprocessing](#)

[Classification with logistic regression](#)

[Classification with support vector machines](#)

[Regression with ElasticNetCV](#)

[Support vector regression](#)

[Clustering with affinity propagation](#)

[Mean Shift](#)

[Genetic algorithms](#)

[Neural networks](#)

[Decision trees](#)

[Summary](#)

## [11. Environments Outside the Python Ecosystem and Cloud Computing](#)

[Exchanging information with MATLAB/Octave](#)

[Installing rpy2](#)

[Interfacing with R](#)

[Sending NumPy arrays to Java](#)

[Integrating SWIG and NumPy](#)

[Integrating Boost and Python](#)

[Using Fortran code through f2py](#)

[Setting up Google App Engine](#)

[Running programs on PythonAnywhere](#)

[Working with Wakari](#)

[Summary](#)

## [12. Performance Tuning, Profiling, and Concurrency](#)

[Profiling the code](#)

[Installing Cython](#)

[Calling C code](#)

[Creating a process pool with multiprocessing](#)

[Speeding up embarrassingly parallel for loops with Joblib](#)

[Comparing Bottleneck to NumPy functions](#)

[Performing MapReduce with Jug](#)

[Installing MPI for Python](#)

[Python Parallel](#)

---

[Summary](#)

[A. Key Concepts](#)

[B. Useful Functions](#)

[matplotlib](#)

[NumPy](#)

[pandas](#)

[Scikit-learn](#)

[SciPy](#)

[scipy.fftpack](#)

[scipy.signal](#)

[scipy.stats](#)

[C. Online Resources](#)

[Index](#)

# Python Data Analysis

---

---

# Python Data Analysis

---

Copyright © 2014 Packt Publishing

All rights reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, without the prior written permission of the publisher, except in the case of brief quotations embedded in critical articles or reviews.

Every effort has been made in the preparation of this book to ensure the accuracy of the information presented. However, the information contained in this book is sold without warranty, either express or implied. Neither the author, nor Packt Publishing, and its dealers and distributors will be held liable for any damages caused or alleged to be caused directly or indirectly by this book.

Packt Publishing has endeavored to provide trademark information about all of the companies and products mentioned in this book by the appropriate use of capitals. However, Packt Publishing cannot guarantee the accuracy of this information.

First published: October 2014

Production reference: 1211014

Published by Packt Publishing Ltd.

Livery Place

35 Livery Street

Birmingham B3 2PB, UK.

ISBN 978-1-78355-335-8

[www.packtpub.com](http://www.packtpub.com)

Cover image by Amy-Lee Winfield (<[abjure@outlook.com](mailto:abjure@outlook.com)>)



# Credits

---

## **Author**

Ivan Idris

## **Reviewers**

Amanda Casari

Thomas A. Dyar

Dr. Hari Shanker Gupta

Puneet Narula

Alan J. Salmoni

## **Commissioning Editor**

Akram Hussain

## **Acquisition Editor**

Owen Roberts

## **Content Development Editor**

Prachi Bisht

## **Technical Editor**

Pankaj Kadam

## **Copy Editors**

Roshni Banerjee

Sarang Chari

Adithi Shetty

## **Project Coordinator**

Shipra Chawhan

## **Proofreaders**

Simran Bhogal

---

Maria Gould

Ameesha Green

## **Indexers**

Hemangini Bari

Mariammal Chettiyar

Rekha Nair

Tejal Soni

## **Graphics**

Sheetal Aute

## **Production Coordinators**

Adonia Jones

Manu Joseph

Komal Ramchandani

## **Cover Work**

Manu Joseph

# About the Author

---

**Ivan Idris** has an MSc degree in Experimental Physics. His graduation thesis had a strong emphasis on Applied Computer Science. After graduating, he worked for several companies as Java developer, data warehouse developer, and QA analyst. His main professional interests are Business Intelligence, Big Data, and Cloud Computing.

Ivan Idris enjoys writing clean, testable code and interesting technical articles. He is the author of *NumPy Beginner's Guide - Second Edition*, *NumPy Cookbook*, and *Learning NumPy Array* all by Packt Publishing. You can find more information and a blog with a few NumPy examples at [ivanidris.net](http://ivanidris.net).

I would like to take this opportunity to thank the reviewers and the team at Packt Publishing for making this book possible. Also, my thanks go to my teachers, professors, and colleagues, who taught me about science and programming. Last but not least, I would like to acknowledge my parents, family, and friends for their support.

# About the Reviewers

---

**Amanda Casari** is currently a data scientist and engineer in the Seattle area. Amanda received her MSEE degree and Certificate of Study in Complex Systems from the University of Vermont and a BS degree in Systems Engineering from the United States Naval Academy. She has more than 10 years of professional experience, ranging from naval officer, analyst, conservation trip leader to integration engineer. Her research interests focus on discovering attributes of natural systems to update and optimize man-made complex networks. Amanda is passionate about making Mathematics and Science approachable to everyone.

I would like to thank my family for supporting our journey and inspiring me during this effort, N. Manukyan for all of her data enthusiasm, C. Stone for creative breakfasts, the Carnation Climbing Club, and P. Nathan for kindly encouraging my myriad interests.

**Thomas A. Dyar** (Tom) is a senior data scientist in the Genomic Sciences group at BD Technologies ([www.bd.com](http://www.bd.com)), Research Triangle Park, North Carolina, where he develops algorithms to process genomic data in a variety of contexts—from targeted panels to whole genomes—for infectious disease and oncology diagnostics applications. His areas of expertise are scientific programming in Java, Python, and R; machine learning, including neural networks and kernel methods; and data analysis and visualization. His primary interests are in conceptualizing and developing large-scale data-driven solutions using Cloud resources.

Tom started his career in software, developing neural networks and expert systems tools for process control in the aerospace and petrochemical industries. He has also worked on distributed virtual environments for stroke rehabilitation at MIT and automated image processing for high-throughput cell biology experiments at BD.

Tom earned his BA degree in Pure & Applied Mathematics from Boston University and is a member of the ACM and IEEE associations.

**Dr. Hari Shanker Gupta** is a senior quantitative research analyst working in the area of algorithmic trading system development. Prior to this, he was a post-doctoral fellow at the Indian Institute of Science (IISc), Bangalore, India. He obtained his PhD in Applied Mathematics and Scientific Computation from IISc. He completed his MSc in Mathematics from Banaras Hindu University (BHU), Varanasi, India. During his MSc, he was awarded four gold medals for outstanding performance at BHU.

Hari has published five research papers in reputed journals in the field of Mathematics and Scientific Computation. He has experience working in the areas of Mathematics, Statistics, and Computation. His experience includes working in numerical methods, partial differential equations, mathematical finance, stochastic calculus, data analysis, finite difference, and finite element methods. He is very comfortable with the mathematics software, MATLAB; the statistics programming language, R; Python; and the programming language, C.

He has reviewed the book *Introduction to R for Quantitative Finance*, Packt Publishing.

---

**Puneet Narula** has over 8 years of experience in the Banking and Finance industry, but his aptitude and passion for the technology sector has brought him back into the world of data and analytics. Leaving behind a stable career in banking was a very tough decision, but following his dreams was even more important to him. He completed his MSc degree in Data Analytics from Dublin Institute of Technology in 2013 to enter the world of analytics and data science. Currently, Puneet is working with Web Reservations International as a PPC data analyst.

At Web Reservations International (WRI), Puneet works with massive clickstream data from both direct and affiliate sources. The technologies used for the analysis is a combination of RapidMiner, R, and Python.

I want to thank Silviu Preoteasa for all his support and motivation at all times.

**Alan J. Salmoni** enjoys making sense of data and is the author of Salstat (<http://www.salstat.com>). He has been using Python for data analysis since 2001 and has taught statistics to undergraduates and postgraduates. When not with his family, he spends time generating large statistical models of text for natural language processing.

Alan owns a company, Thought Into Design, which specializes in data analysis and user experience.

I would like to thank my wife, Jell, and my daughter, Louise, for their patience.

## Support files, eBooks, discount offers, and more

You might want to visit [www.PacktPub.com](http://www.PacktPub.com) for support files and downloads related to your book.

Did you know that Packt offers eBook versions of every book published, with PDF and ePub files available? You can upgrade to the eBook version at [www.PacktPub.com](http://www.PacktPub.com) and as a print book customer, you are entitled to a discount on the eBook copy. Get in touch with us at [<service@packtpub.com>](mailto:service@packtpub.com) for more details.

At [www.PacktPub.com](http://www.PacktPub.com), you can also read a collection of free technical articles, sign up for a range of free newsletters and receive exclusive discounts and offers on Packt books and eBooks.



<http://PacktLib.PacktPub.com>

Do you need instant solutions to your IT questions? PacktLib is Packt's online digital book library. Here, you can access, read and search across Packt's entire library of books.

## Why subscribe?

- Fully searchable across every book published by Packt
- Copy and paste, print and bookmark content
- On demand and accessible via web browser

## Free access for Packt account holders

If you have an account with Packt at [www.PacktPub.com](http://www.PacktPub.com), you can use this to access PacktLib today and view nine entirely free books. Simply use your login credentials for immediate access.

# Preface

---

"Data analysis is Python's killer app."

--Unknown

Data analysis has a rich history in the natural, biomedical, and social sciences. You may have heard of *Big Data*. Although, it's hard to give a precise definition of Big Data, we should be aware of its impact on data analysis efforts. Currently, we have the following trends associated with Big Data:

- The world's population continues to grow
- More and more data is collected and stored
- The number of transistors that can be put on a computer chip cannot grow indefinitely
- Governments, scientists, industry, and individuals have a growing need to learn from data

Data analysis has gained popularity lately due to the hype around *Data Science*. Data analysis and Data Science attempt to extract information from data. For that purpose, we use techniques from statistics, machine learning, signal processing, natural language processing, and computer science.

A mind map visualizing Python software that can be used for data analysis can be found at <http://www.xmind.net/m/WvfC/>. The first thing that we should notice is that the Python ecosystem is very mature. It includes famous packages such as NumPy, SciPy, and matplotlib. This should not come as a surprise since Python has been around since 1989. Python is easy to learn and use, less verbose than other programming languages, and very readable. Even if you don't know Python, you can pick up the basics within days, especially if you have experience in another programming language. To enjoy this book, you don't need more than the basics. There are plenty of books, courses, and online tutorials that teach Python.

# What this book covers

---

This book starts as a tutorial on NumPy, SciPy, matplotlib, and pandas. These are open source Python packages useful for numerical work, data wrangling, and visualization. Combined, they can compete with MATLAB, Mathematica, and R. The second half of the book teaches more advanced topics such as signal processing, databases, text analysis, machine learning, interoperability, and performance tuning.

[Chapter 1](#), *Getting Started with Python Libraries*, guides us to achieve a successful installation of the numerical Python software and set it up step by step. Also, we will create a small application.

[Chapter 2](#), *NumPy Arrays*, introduces us to NumPy fundamentals and arrays. By the end of this chapter, we will have basic understanding of NumPy arrays and the associated functions.

[Chapter 3](#), *Statistics and Linear Algebra*, gives a quick overview of linear algebra and statistical functions.

[Chapter 4](#), *pandas Primer*, provides a tutorial on basic pandas functionality where we learn about pandas data structures and operations.

[Chapter 5](#), *Retrieving, Processing, and Storing Data*, explains how to acquire data in various formats and how to clean raw data and store it.

[Chapter 6](#), *Data Visualization*, teaches how to plot data with matplotlib.

[Chapter 7](#), *Signal Processing and Time Series*, contains time series and signal processing examples using sunspot cycles data. The examples mostly use NumPy/SciPy, along with statsmodels in at least one example.

[Chapter 8](#), *Working with Databases*, provides information about various databases (relational and NoSQL) and related APIs.

[Chapter 9](#), *Analyzing Textual Data and Social Media*, analyzes texts for sentiment analysis and topics extraction. A small example is also given of network analysis.

[Chapter 10](#), *Predictive Analytics and Machine Learning*, explains artificial intelligence with weather prediction as a running example and mostly uses scikit-learn. However, some machine learning algorithms are not covered by scikit-learn, so for those, we use other APIs.

[Chapter 11](#), *Environments Outside the Python Ecosystem and Cloud Computing*, gives various examples on how to integrate existing code not written in Python. Also, setup in the Cloud will be demonstrated.

[Chapter 12](#), *Performance Tuning, Profiling, and Concurrency*, gives hints on improving performance with profiling and Cythoning as key techniques. For multicore, distributed



systems, we discuss the relevant frameworks too.

---

[Appendix A](#), *Key Concepts*, serves as a glossary containing short descriptions of key concepts found throughout the book.

[Appendix B](#), *Useful Functions*, gives an overview of functions used in the book.

[Appendix C](#), *Online Resources*, lists links to documentation, forums, articles, and other important information.

# What you need for this book

---

The code examples in this book should work on most modern operating systems. For all chapters, Python 2 and pip is required. To install Python, go to <https://wiki.python.org/moin/BeginnersGuide/Download>. To install pip, go to <http://pip.readthedocs.org/en/latest/installing.html>. Instructions to install software are given throughout the chapters. Most of the time, we need to run the following command with administrator privileges:

```
$ pip install <some software>
```

The following is a list of software used for the examples and versions used for testing purposes:

- NumPy 1.8.1
- SciPy 0.14.0
- matplotlib 1.3.1
- IPython 2.0.0
- pandas Version 0.13.1
- tables 3.1.1
- numexpr 2.4
- openpyxl 2.0.3
- XlsxWriter 0.5.5
- xlrd 0.9.3
- feedparser 5.1.3
- BeautifulSoup 4.3.2
- StatsModels 0.6.0
- SQLAlchemy 0.9.6
- Pony 0.5.1
- dataset 0.5.4
- MongoDB 2.6.3
- PyMongo 2.7.1
- Redis server 2.8.12
- Redis 2.10.1
- Cassandra 2.0.9
- Java 7
- NLTK 2.0.4
- scikit-learn 0.15.0
- NetworkX 1.9
- DEAP 1.0.1
- theano 0.2.0
- Graphviz 2.36.0
- pydot2 1.0.33
- Octave 3.8.0

- R 3.1.1
- rpy2 2.4.2
- JPype 0.5.5.2
- Java 7
- SWIG 3.02
- PCRE 8.35
- Boost 1.56.0
- gfortran 4.9.0
- GAE for Python 2.7
- gprof2dot 2014.08.05
- line\_profiler beta
- Cython 0.20.0
- cytoolz 0.7.0
- Joblib 0.8.2
- Bottleneck 0.8.0
- Jug 0.9.3
- MPI 1.8.1
- mpi4py 1.3.1

Of course, it's not necessary for you to have the same version of the software. Usually, the latest version available should work.

## Note

Some of the software listed are used for a single example; therefore, please check first whether the example is relevant for you before installing the software.

To uninstall Python packages installed with pip, use the following command:

```
$ pip uninstall <some software>
```

# Who this book is for

---

This book is for people with basic knowledge of Python and Mathematics who want to learn how to use Python software to analyze data. We try to keep things simple, but it's not possible to cover all the topics in great detail. It may be useful for you to refresh your knowledge of Mathematics via Khan Academy, Coursera, or Wikipedia.

I would recommend the following books by Packt Publishing for further reading:

- *Building Machine Learning Systems with Python*, Willi Richert and Luis Pedro Coelho (2013)
- *Learning Cython Programming*, Philip Herron (2013)
- *Learning NumPy Array*, Ivan Idris (2014)
- *Learning scikit-learn: Machine Learning in Python*, Raúl Garreta and Guillermo Moncecc (2013)
- *Learning SciPy for Numerical and Scientific Computing*, Francisco J. Blanco-Silva (2013)
- *Matplotlib for Python Developers*, Sandro Tosi (2009)
- *NumPy Beginner's Guide - Second Edition*, Ivan Idris (2013)
- *NumPy Cookbook*, Ivan Idris (2012)
- *Parallel Programming with Python*, Jan Palach (2014)
- *Python Data Visualization Cookbook*, Igor Milovanović (2013)
- *Python for Finance*, Yuxing Yan (2014)
- *Python Text Processing with NLTK 2.0 Cookbook*, Jacob Perkins (2010)

# Conventions

---

In this book, you will find a number of styles of text that distinguish between different kinds of information. Here are some examples of these styles, and an explanation of their meaning.

Code words in text, database table names, folder names, filenames, file extensions, pathnames, dummy URLs, user input, and Twitter handles are shown as follows: "Notice that `numpysum()` does not need a `for` loop."

A block of code is set as follows:

```
def pythonsum(n):
    a = range(n)
    b = range(n)
    c = []

    for i in range(len(a)):
        a[i] = i ** 2
        b[i] = i ** 3
        c.append(a[i] + b[i])

    return c
```

Any command-line input or output is written as follows:

```
$ yum install python-numpy
```

**New terms** and **important words** are shown in bold. Words that you see on the screen, in menus or dialog boxes for example, appear in the text like this: "Click on the **Next** button."

## Note

Warnings or important notes appear in a box like this.

## Tip

Tips and tricks appear like this.

# Reader feedback

---

Feedback from our readers is always welcome. Let us know what you think about this book—what you liked or may have disliked. Reader feedback is important for us to develop titles that you really get the most out of.

To send us general feedback, simply send an e-mail to <[feedback@packtpub.com](mailto:feedback@packtpub.com)>, and mention the book title via the subject of your message.

If there is a topic that you have expertise in and you are interested in either writing or contributing to a book, see our author guide on [www.packtpub.com/authors](http://www.packtpub.com/authors).

# Customer support

---

Now that you are the proud owner of a Packt book, we have a number of things to help you get the most from your purchase.

## Downloading the example code

You can download the example code files for all Packt books you have purchased from your account at <http://www.packtpub.com>. If you purchased this book elsewhere, you can visit <http://www.packtpub.com/support> and register to have the files e-mailed directly to you.

## Errata

Although we have taken every care to ensure the accuracy of our content, mistakes do happen. If you find a mistake in one of our books—maybe a mistake in the text or the code—we would be grateful if you would report this to us. By doing so, you can save other readers from frustration and help us improve subsequent versions of this book. If you find any errata please report them by visiting <http://www.packtpub.com/submit-errata>, selecting your book, clicking on the **errata submission form** link, and entering the details of your errata. Once your errata are verified, your submission will be accepted and the errata will be uploaded on our website, or added to any list of existing errata, under the Errata section of that title. Any existing errata can be viewed by selecting your title from <http://www.packtpub.com/support>.

## Piracy

Piracy of copyright material on the Internet is an ongoing problem across all media. At Packt we take the protection of our copyright and licenses very seriously. If you come across any illegal copies of our works, in any form, on the Internet, please provide us with the location address or website name immediately so that we can pursue a remedy.

Please contact us at [<copyright@packtpub.com>](mailto:copyright@packtpub.com) with a link to the suspected pirated material.

We appreciate your help in protecting our authors, and our ability to bring you valuable content.

## Questions

You can contact us at [<questions@packtpub.com>](mailto:questions@packtpub.com) if you are having a problem with any aspect of the book, and we will do our best to address it.

# Chapter 1. Getting Started with Python Libraries

Let's get started. We can find a mind map describing software that can be used for data analysis at <http://www.xmind.net/m/WvfC/>. Obviously, we can't install all of this software in this chapter. We will install NumPy, SciPy, matplotlib, and IPython on different operating systems and have a look at some simple code that uses NumPy.

**NumPy** is a fundamental Python library that provides numerical arrays and functions.

**SciPy** is a scientific Python library, which supplements and slightly overlaps NumPy. NumPy and SciPy historically shared their code base but were later separated.

**matplotlib** is a plotting library based on NumPy. You can read more about matplotlib in [Chapter 6, Data Visualization](#).

**IPython** provides an architecture for interactive computing. The most notable part of this project is the IPython shell. We will cover the IPython shell later in this chapter.

Installation instructions for the other software we need will be given throughout the book at the appropriate time. At the end of this chapter, you will find pointers on how to find additional information online if you get stuck or are uncertain about the best way to solve problems.

In this chapter, we will cover:

- Installing Python, SciPy, matplotlib, IPython, and NumPy on Windows, Linux, and Macintosh
- Writing a simple application using NumPy arrays
- Getting to know IPython
- Online resources and help

## Software used in this book

The software used in this book is based on Python, so you are required to have Python installed. On some operating systems, Python is already installed. You, however, need to check whether the Python version is compatible with the software version you want to install. There are many implementations of Python, including commercial implementations and distributions. In this book, we will focus on the standard CPython implementation, which is guaranteed to be compatible with NumPy.

### Note



- [click Why We Can't Wait for free](#)
- [C++ Programming: Program Design Including Data Structures \(4th Edition\) here](#)
- [download Khallaas - an A to Z Guide to the Underworld](#)
- [download online Zen Living \(Idiot's Guides\) pdf, azw \(kindle\)](#)
  
- <http://flog.co.id/library/Why-We-Can-t-Wait.pdf>
- <http://www.1973vision.com/?library/Sylvie.pdf>
- <http://rodrigocaporal.com/library/Khallaas---an-A-to-Z-Guide-to-the-Underworld.pdf>
- <http://damianfoster.com/books/The-Haunted-Air--Repairman-Jack--Book-6-.pdf>