

# Preface

We live in an age where data is being collected at unprecedented levels. One can benefit from data only after converting it into useful and actionable knowledge. In the past few years, the methodology of extracting insights from data or “data science” has emerged as a discipline in its own right. Historically, a diverse set of tools have been used for data analysis. The R programming language is increasingly becoming a one-stop solution to data science.

R is an open-source software and can be used on most computing platforms: Windows, Unix/Linux, Mac OS X. The primary reason for the growing popularity of R is due to a vast package library containing implementations for most statistical analysis techniques. On the other hand, R is not an easy programming language to learn due to its esoteric syntax. The documentation for most packages is in the form of reference material, which makes it difficult for programmers without a background in statistics to get started on their own.

## Goal of this Book

In this book, we introduce the readers to some of the useful data science techniques and their implementation with the R programming language. We attempt to strike a balance between the how and the why. We will cover both the how: various data science processes and methodologies, and the why: understanding the motivation and internals of each technique. The goal of this book is to enable the reader to apply similar methodologies to analyze their own datasets.

We aim to provide a tutorial for performing data science with R as opposed to a reference manual. The number of data analysis methods and the corresponding R packages is vast. This book is intended to be a good starting point to begin data analysis. In each chapter, we present a data science methodology with a case study, including a dataset. All the datasets and source code listings are available for download at: [URL](#).

## Prerequisites

This book is intended for the readers who are not familiar with statistical analysis. When covering each data analysis task, we will provide a high-level background of the underlying statistics. Experience with at least one programming language is a prerequisite. This book will be useful for readers who are not familiar with the R programming language. We will review the fundamentals of R in Chapter 2 along the notations and conventions used in this book.

August 2014  
Sunnyvale, California

Dr. Manas A. Pathak

Beginning Data Science with R

Pathak, M.A.

2014, XI, 157 p. 155 illus., 26 illus. in color., Hardcover

ISBN: 978-3-319-12065-2