

## Big Data Analytics With R And Hadoop

Data Analysis with R, Second Edition  
Data Analytics using R  
Modern Data Science with R  
Learning Social Media Analytics with R  
Practical Big Data Analytics  
Big Data Analytics  
Hands-On Data Science with R  
Big Data Analytics with Hadoop 3  
Data Analysis with R  
Business Analytics Using R - A Practical Approach  
Data Science in Education Using R  
Beginning Data Science with R  
Getting Started with Data Science  
An Introduction to Data Science  
Big Data in Medical Image Processing  
Geographical Data Science and Spatial Data Analysis  
Big Data Analytics with R  
Big Data Analytics  
Learning Predictive Analytics with R  
R for Data Science  
Data Science and Predictive Analytics  
Beginning Data Science in R  
Data Mining and Business Analytics with R  
Basketball Data Science  
Data Analysis for the Life Sciences with R  
Using R to Unlock the Value of Big Data: Big Data Analytics with Oracle R Enterprise and Oracle R Connector for Hadoop  
Predictive Analytics For Dummies  
Pentaho for Big Data Analytics  
Data Science and Big Data Analytics  
The Essentials of Data Science: Knowledge Discovery Using R  
Python for R Users  
Data Science Using Python and R  
R for Data Science  
Introduction to Data Science  
Mastering Data Analysis with R  
Exam Prep for: Using R to Unlock the Value of Big Data; Big  
Introduction to Data Science for Social and Policy Research  
Big Data Analytics with R and Hadoop  
Financial Analytics with R  
An Approach for Handling Big Data Analytics with R Using Air Quality Data

### Data Analysis with R, Second Edition

Discover best practices for data analysis and software development in R and start on the path to becoming a fully-fledged data scientist. This book teaches you techniques for both data manipulation and visualization and shows you the best way for developing new software packages for R. Beginning Data Science in R details how data science is a combination of statistics, computational science, and machine learning. You'll see how to efficiently structure and mine data to extract useful patterns and build mathematical models. This requires computational methods and programming, and R is an ideal programming language for this. This book is based on a number of lecture notes for classes the author has taught on data science and statistical programming using the R programming language. Modern data analysis requires computational skills and usually a minimum of programming. What You Will Learn Perform data science and analytics using statistics and the R programming language Visualize and explore data, including working with large data sets found in big data Build an R package Test and check your code Practice version control Profile and optimize your code Who This Book Is For Those with some data science or analytics background, but not necessarily experience with the R programming language.

### Data Analytics using R

Welcome to this book on "Big Data Analytics - Derive Insights" is a beginners guide to big data, mining, manipulation, integration, processing, visualisation and interpretation based on actual case studies in implementing Data Management frameworks across diverse organisations. Big data is a term used collectively for data sets that are voluminous, large and complicated that cannot be processed by

traditional databases and applications. Some of the major challenges are the collection, extraction, mining, integration, analysis, sharing, storage, transfer, loading, querying, updating and visualisation of these data sets from multiple streams and formats. An effective and accurate visualisation of big data will lead to a more informed decision which in turn will result in greater operational efficiency, cost reduction, increased revenues and reduction in risk. Big Data analysis is used to identify business trends, predict accurate supply and demand, prevent diseases, combat crime and growth in revenues. Businessmen, organisations, scientists, hospitals, social media, marketing and governments face difficulties with large data sets in areas including internet search, finance and business trends. This book will give you a basic understanding of how big data works and the process used to get insights from such data.

### **Modern Data Science with R**

"This book introduces you to R, RStudio, and the tidyverse, a collection of R packages designed to work together to make data science fast, fluent, and fun. Suitable for readers with no previous programming experience"--

### **Learning Social Media Analytics with R**

Tap into the realm of social media and unleash the power of analytics for data-driven insights using R About This Book A practical guide written to help leverage the power of the R eco-system to extract, process, analyze, visualize and model social media data Learn about data access, retrieval, cleaning, and curation methods for data originating from various social media platforms. Visualize and analyze data from social media platforms to understand and model complex relationships using various concepts and techniques such as Sentiment Analysis, Topic Modeling, Text Summarization, Recommendation Systems, Social Network Analysis, Classification, and Clustering. Who This Book Is For It is targeted at IT professionals, Data Scientists, Analysts, Developers, Machine Learning Enthusiasts, social media marketers and anyone with a keen interest in data, analytics, and generating insights from social data. Some background experience in R would be helpful, but not necessary, since this book is written keeping in mind, that readers can have varying levels of expertise. What You Will Learn Learn how to tap into data from diverse social media platforms using the R ecosystem Use social media data to formulate and solve real-world problems Analyze user social networks and communities using concepts from graph theory and network analysis Learn to detect opinion and sentiment, extract themes, topics, and trends from unstructured noisy text data from diverse social media channels Understand the art of representing actionable insights with effective visualizations Analyze data from major social media channels such as Twitter, Facebook, Flickr, Foursquare, Github, StackExchange, and so on Learn to leverage popular R packages such as ggplot2, topicmodels, caret, e1071, tm, wordcloud, twittR, Rfacebook, dplyr, reshape2, and many more In Detail The Internet has truly become humongous, especially with the rise of various forms of social media in the last decade, which give users a platform to express themselves and also communicate and collaborate with each other. This book will help the reader to understand the current social media landscape and to learn how analytics can be leveraged to derive insights from it. This data can be analyzed to gain valuable insights into the

behavior and engagement of users, organizations, businesses, and brands. It will help readers frame business problems and solve them using social data. The book will also cover several practical real-world use cases on social media using R and its advanced packages to utilize data science methodologies such as sentiment analysis, topic modeling, text summarization, recommendation systems, social network analysis, classification, and clustering. This will enable readers to learn different hands-on approaches to obtain data from diverse social media sources such as Twitter and Facebook. It will also show readers how to establish detailed workflows to process, visualize, and analyze data to transform social data into actionable insights. **Style and approach** This book follows a step-by-step approach with detailed strategies for understanding, extracting, analyzing, visualizing, and modeling data from several major social network platforms such as Facebook, Twitter, Foursquare, Flickr, Github, and StackExchange. The chapters cover several real-world use cases and leverage data science, machine learning, network analysis, and graph theory concepts along with the R ecosystem, including popular packages such as ggplot2, caret, dplyr, topicmodels, tm, and so on.

### **Practical Big Data Analytics**

Gain sharp insights into your data and solve real-world data science problems with R—from data munging to modeling and visualization **About This Book** Handle your data with precision and care for optimal business intelligence Restructure and transform your data to inform decision-making Packed with practical advice and tips to help you get to grips with data mining **Who This Book Is For** If you are a data scientist or R developer who wants to explore and optimize your use of R's advanced features and tools, this is the book for you. A basic knowledge of R is required, along with an understanding of database logic. **What You Will Learn** Connect to and load data from R's range of powerful databases Successfully fetch and parse structured and unstructured data Transform and restructure your data with efficient R packages Define and build complex statistical models with glm Develop and train machine learning algorithms Visualize social networks and graph data Deploy supervised and unsupervised classification algorithms Discover how to visualize spatial data with R **In Detail** R is an essential language for sharp and successful data analysis. Its numerous features and ease of use make it a powerful way of mining, managing, and interpreting large sets of data. In a world where understanding big data has become key, by mastering R you will be able to deal with your data effectively and efficiently. This book will give you the guidance you need to build and develop your knowledge and expertise. Bridging the gap between theory and practice, this book will help you to understand and use data for a competitive advantage. Beginning with taking you through essential data mining and management tasks such as munging, fetching, cleaning, and restructuring, the book then explores different model designs and the core components of effective analysis. You will then discover how to optimize your use of machine learning algorithms for classification and recommendation systems beside the traditional and more recent statistical methods. **Style and approach** Covering the essential tasks and skills within data science, **Mastering Data Analysis** provides you with solutions to the challenges of data science. Each section gives you a theoretical overview before demonstrating how to put the theory to work with real-world use cases and hands-on examples.

## **Big Data Analytics**

The Essentials of Data Science: Knowledge Discovery Using R presents the concepts of data science through a hands-on approach using free and open source software. It systematically drives an accessible journey through data analysis and machine learning to discover and share knowledge from data. Building on over thirty years' experience in teaching and practising data science, the author encourages a programming-by-example approach to ensure students and practitioners attune to the practise of data science while building their data skills. Proven frameworks are provided as reusable templates. Real world case studies then provide insight for the data scientist to swiftly adapt the templates to new tasks and datasets. The book begins by introducing data science. It then reviews R's capabilities for analysing data by writing computer programs. These programs are developed and explained step by step. From analysing and visualising data, the framework moves on to tried and tested machine learning techniques for predictive modelling and knowledge discovery. Literate programming and a consistent style are a focus throughout the book.

## **Hands-On Data Science with R**

Overview: This book is aimed at undergraduate students of computer science and engineering. The book will be useful companion for IT professionals to data analysts and decision makers responsible for driving strategic initiatives, and management graduates and business analysts, engaged in self-study. This book by Acharya unleashes the power of R as a statistical data analytics and visualization tool and introduces the learners to several data mining algorithms and chart forms / visualizations. It has good emphasis on 'asking the right questions'. Salient Features: - Exhaustive coverage includes installation of R and its package, getting accustomed to R interface and R commands, working with data from disparate data sources (.csv, JSON, XML, RDBMS etc.), getting conversant with classification, clustering, association rule mining, regression, text mining etc. - 12 Case studies namely Insurance Fraud Detection, Customer Insights Analysis, Sales Forecasting, Credit Card Spending by Customer Groups and Helping Retailers Predict In-store Customer Traffic - Pedagogical features: \* 300+ chapter-end and check your progress questions for self-assessment \* 200 Multiple-choice questions \* 10+ hands-on practical exercises \* Exhaustive illustrations

## **Big Data Analytics with Hadoop 3**

The Oracle Press Guide to Big Data Analytics using R Cowritten by members of the Big Data team at Oracle, this Oracle Press book focuses on analyzing data with R while making it scalable using Oracle's R technologies. Using R to Unlock the Value of Big Data provides an introduction to open source R and describes issues with traditional R and database interaction. The book then offers in-depth coverage of Oracle's strategic R offerings: Oracle R Enterprise, Oracle R Distribution, ROacle, and Oracle R Connector for Hadoop. You can practice your new skills using the end-of-chapter exercises.

## **Data Analysis with R**

Get command of your organizational Big Data using the power of data science and analytics Key Features A perfect companion to boost your Big Data storing, processing, analyzing skills to help you take informed business decisions Work with the best tools such as Apache Hadoop, R, Python, and Spark for NoSQL platforms to perform massive online analyses Get expert tips on statistical inference, machine learning, mathematical modeling, and data visualization for Big Data Book Description Big Data analytics relates to the strategies used by organizations to collect, organize and analyze large amounts of data to uncover valuable business insights that otherwise cannot be analyzed through traditional systems. Crafting an enterprise-scale cost-efficient Big Data and machine learning solution to uncover insights and value from your organization's data is a challenge. Today, with hundreds of new Big Data systems, machine learning packages and BI Tools, selecting the right combination of technologies is an even greater challenge. This book will help you do that. With the help of this guide, you will be able to bridge the gap between the theoretical world of technology with the practical ground reality of building corporate Big Data and data science platforms. You will get hands-on exposure to Hadoop and Spark, build machine learning dashboards using R and R Shiny, create web-based apps using NoSQL databases such as MongoDB and even learn how to write R code for neural networks. By the end of the book, you will have a very clear and concrete understanding of what Big Data analytics means, how it drives revenues for organizations, and how you can develop your own Big Data analytics solution using different tools and methods articulated in this book. What you will learn - Get a 360-degree view into the world of Big Data, data science and machine learning - Broad range of technical and business Big Data analytics topics that caters to the interests of the technical experts as well as corporate IT executives - Get hands-on experience with industry-standard Big Data and machine learning tools such as Hadoop, Spark, MongoDB, KDB+ and R - Create production-grade machine learning BI Dashboards using R and R Shiny with step-by-step instructions - Learn how to combine open-source Big Data, machine learning and BI Tools to create low-cost business analytics applications - Understand corporate strategies for successful Big Data and data science projects - Go beyond general-purpose analytics to develop cutting-edge Big Data applications using emerging technologies Who this book is for The book is intended for existing and aspiring Big Data professionals who wish to become the go-to person in their organization when it comes to Big Data architecture, analytics, and governance. While no prior knowledge of Big Data or related technologies is assumed, it will be helpful to have some programming experience.

### **Business Analytics Using R - A Practical Approach**

Get to grips with key data visualization and predictive analytic skills using R About This Book Acquire predictive analytic skills using various tools of R Make predictions about future events by discovering valuable information from data using R Comprehensible guidelines that focus on predictive model design with real-world data Who This Book Is For If you are a statistician, chief information officer, data scientist, ML engineer, ML practitioner, quantitative analyst, and student of machine learning, this is the book for you. You should have basic knowledge of the use of R. Readers without previous experience of programming in R will also be able to use the tools in the book. What You Will Learn Customize R by installing and loading new packages Explore the structure of data using clustering

algorithms Turn unstructured text into ordered data, and acquire knowledge from the data Classify your observations using Naive Bayes, k-NN, and decision trees Reduce the dimensionality of your data using principal component analysis Discover association rules using Apriori Understand how statistical distributions can help retrieve information from data using correlations, linear regression, and multilevel regression Use PMML to deploy the models generated in R In Detail R is statistical software that is used for data analysis. There are two main types of learning from data: unsupervised learning, where the structure of data is extracted automatically; and supervised learning, where a labeled part of the data is used to learn the relationship or scores in a target attribute. As important information is often hidden in a lot of data, R helps to extract that information with its many standard and cutting-edge statistical functions. This book is packed with easy-to-follow guidelines that explain the workings of the many key data mining tools of R, which are used to discover knowledge from your data. You will learn how to perform key predictive analytics tasks using R, such as train and test predictive models for classification and regression tasks, score new data sets and so on. All chapters will guide you in acquiring the skills in a practical way. Most chapters also include a theoretical introduction that will sharpen your understanding of the subject matter and invite you to go further. The book familiarizes you with the most common data mining tools of R, such as k-means, hierarchical regression, linear regression, association rules, principal component analysis, multilevel modeling, k-NN, Naive Bayes, decision trees, and text mining. It also provides a description of visualization techniques using the basic visualization tools of R as well as lattice for visualizing patterns in data organized in groups. This book is invaluable for anyone fascinated by the data mining opportunities offered by GNU R and its packages. Style and approach This is a practical book, which analyzes compelling data about life, health, and death with the help of tutorials. It offers you a useful way of interpreting the data that's specific to this book, but that can also be applied to any other data.

### **Data Science in Education Using R**

Utilize R to uncover hidden patterns in your Big Data About This Book Perform computational analyses on Big Data to generate meaningful results Get a practical knowledge of R programming language while working on Big Data platforms like Hadoop, Spark, H2O and SQL/NoSQL databases, Explore fast, streaming, and scalable data analysis with the most cutting-edge technologies in the market Who This Book Is For This book is intended for Data Analysts, Scientists, Data Engineers, Statisticians, Researchers, who want to integrate R with their current or future Big Data workflows. It is assumed that readers have some experience in data analysis and understanding of data management and algorithmic processing of large quantities of data, however they may lack specific skills related to R. What You Will Learn Learn about current state of Big Data processing using R programming language and its powerful statistical capabilities Deploy Big Data analytics platforms with selected Big Data tools supported by R in a cost-effective and time-saving manner Apply the R language to real-world Big Data problems on a multi-node Hadoop cluster, e.g. electricity consumption across various socio-demographic indicators and bike share scheme usage Explore the compatibility of R with Hadoop, Spark, SQL and NoSQL databases, and H2O platform In Detail Big Data analytics is the process of examining large and complex data sets that often

exceed the computational capabilities. R is a leading programming language of data science, consisting of powerful functions to tackle all problems related to Big Data processing. The book will begin with a brief introduction to the Big Data world and its current industry standards. With introduction to the R language and presenting its development, structure, applications in real world, and its shortcomings. Book will progress towards revision of major R functions for data management and transformations. Readers will be introduced to Cloud based Big Data solutions (e.g. Amazon EC2 instances and Amazon RDS, Microsoft Azure and its HDInsight clusters) and also provide guidance on R connectivity with relational and non-relational databases such as MongoDB and HBase etc. It will further expand to include Big Data tools such as Apache Hadoop ecosystem, HDFS and MapReduce frameworks. Also other R compatible tools such as Apache Spark, its machine learning library Spark MLlib, as well as H2O. Style and approach This book will serve as a practical guide to tackling Big Data problems using R programming language and its statistical environment. Each section of the book will present you with concise and easy-to-follow steps on how to process, transform and analyse large data sets.

### **Beginning Data Science with R**

Explore big data concepts, platforms, analytics, and their applications using the power of Hadoop 3 Key Features Learn Hadoop 3 to build effective big data analytics solutions on-premise and on cloud Integrate Hadoop with other big data tools such as R, Python, Apache Spark, and Apache Flink Exploit big data using Hadoop 3 with real-world examples Book Description Apache Hadoop is the most popular platform for big data processing, and can be combined with a host of other big data tools to build powerful analytics solutions. Big Data Analytics with Hadoop 3 shows you how to do just that, by providing insights into the software as well as its benefits with the help of practical examples. Once you have taken a tour of Hadoop 3's latest features, you will get an overview of HDFS, MapReduce, and YARN, and how they enable faster, more efficient big data processing. You will then move on to learning how to integrate Hadoop with the open source tools, such as Python and R, to analyze and visualize data and perform statistical computing on big data. As you get acquainted with all this, you will explore how to use Hadoop 3 with Apache Spark and Apache Flink for real-time data analytics and stream processing. In addition to this, you will understand how to use Hadoop to build analytics solutions on the cloud and an end-to-end pipeline to perform big data analysis using practical use cases. By the end of this book, you will be well-versed with the analytical capabilities of the Hadoop ecosystem. You will be able to build powerful solutions to perform big data analytics and get insight effortlessly. What you will learn Explore the new features of Hadoop 3 along with HDFS, YARN, and MapReduce Get well-versed with the analytical capabilities of Hadoop ecosystem using practical examples Integrate Hadoop with R and Python for more efficient big data processing Learn to use Hadoop with Apache Spark and Apache Flink for real-time data analytics Set up a Hadoop cluster on AWS cloud Perform big data analytics on AWS using Elastic Map Reduce Who this book is for Big Data Analytics with Hadoop 3 is for you if you are looking to build high-performance analytics solutions for your enterprise or business using Hadoop 3's powerful features, or you're new to big data analytics. A basic understanding of the Java programming language is required.

## **Getting Started with Data Science**

“We live in the age of data. In the last few years, the methodology of extracting insights from data or "data science" has emerged as a discipline in its own right. The R programming language has become one-stop solution for all types of data analysis. The growing popularity of R is due its statistical roots and a vast open source package library. The goal of “Beginning Data Science with R” is to introduce the readers to some of the useful data science techniques and their implementation with the R programming language. The book attempts to strike a balance between the how: specific processes and methodologies, and understanding the why: going over the intuition behind how a particular technique works, so that the reader can apply it to the problem at hand. This book will be useful for readers who are not familiar with statistics and the R programming language.

## **An Introduction to Data Science**

## **Big Data in Medical Image Processing**

Financial Analytics with R sharpens readers' skills in time-series, forecasting, portfolio selection, covariance clustering, prediction, and derivative securities.

## **Geographical Data Science and Spatial Data Analysis**

Master Data Analytics Hands-On by Solving Fascinating Problems You'll Actually Enjoy! Harvard Business Review recently called data science “The Sexiest Job of the 21st Century.” It’s not just sexy: For millions of managers, analysts, and students who need to solve real business problems, it’s indispensable. Unfortunately, there’s been nothing easy about learning data science—until now. Getting Started with Data Science takes its inspiration from worldwide best-sellers like Freakonomics and Malcolm Gladwell’s Outliers: It teaches through a powerful narrative packed with unforgettable stories. Murtaza Haider offers informative, jargon-free coverage of basic theory and technique, backed with plenty of vivid examples and hands-on practice opportunities. Everything’s software and platform agnostic, so you can learn data science whether you work with R, Stata, SPSS, or SAS. Best of all, Haider teaches a crucial skillset most data science books ignore: how to tell powerful stories using graphics and tables. Every chapter is built around real research challenges, so you’ll always know why you’re doing what you’re doing. You’ll master data science by answering fascinating questions, such as: • Are religious individuals more or less likely to have extramarital affairs? • Do attractive professors get better teaching evaluations? • Does the higher price of cigarettes deter smoking? • What determines housing prices more: lot size or the number of bedrooms? • How do teenagers and older people differ in the way they use social media? • Who is more likely to use online dating services? • Why do some purchase iPhones and others Blackberry devices? • Does the presence of children influence a family’s spending on alcohol? For each problem, you’ll walk through defining your question and the answers you’ll need; exploring how others have approached similar challenges; selecting your data and methods; generating



your statistics; organizing your report; and telling your story. Throughout, the focus is squarely on what matters most: transforming data into insights that are clear, accurate, and can be acted upon.

### **Big Data Analytics with R**

Combine business sense, statistics, and computers in a new and intuitive way, thanks to Big Data Predictive analytics is a branch of data mining that helps predict probabilities and trends. Predictive Analytics For Dummies explores the power of predictive analytics and how you can use it to make valuable predictions for your business, or in fields such as advertising, fraud detection, politics, and others. This practical book does not bog you down with loads of mathematical or scientific theory, but instead helps you quickly see how to use the right algorithms and tools to collect and analyze data and apply it to make predictions. Topics include using structured and unstructured data, building models, creating a predictive analysis roadmap, setting realistic goals, budgeting, and much more. Shows readers how to use Big Data and data mining to discover patterns and make predictions for tech-savvy businesses Helps readers see how to shepherd predictive analytics projects through their companies Explains just enough of the science and math, but also focuses on practical issues such as protecting project budgets, making good presentations, and more Covers nuts-and-bolts topics including predictive analytics basics, using structured and unstructured data, data mining, and algorithms and techniques for analyzing data Also covers clustering, association, and statistical models; creating a predictive analytics roadmap; and applying predictions to the web, marketing, finance, health care, and elsewhere Propose, produce, and protect predictive analytics projects through your company with Predictive Analytics For Dummies.

### **Big Data Analytics**

This comprehensive guide provides a step-by-step approach to data collection, cleaning, formatting, and storage, using Python and R.

### **Learning Predictive Analytics with R**

Learn the fundamental aspects of the business statistics, data mining, and machine learning techniques required to understand the huge amount of data generated by your organization. This book explains practical business analytics through examples, covers the steps involved in using it correctly, and shows you the context in which a particular technique does not make sense. Further, Practical Business Analytics using R helps you understand specific issues faced by organizations and how the solutions to these issues can be facilitated by business analytics. This book will discuss and explore the following through examples and case studies: An introduction to R: data management and R functions The architecture, framework, and life cycle of a business analytics project Descriptive analytics using R: descriptive statistics and data cleaning Data mining: classification, association rules, and clustering Predictive analytics: simple regression, multiple regression, and logistic regression This book includes case studies on important business analytic techniques, such as classification,

association, clustering, and regression. The R language is the statistical tool used to demonstrate the concepts throughout the book. What You Will Learn • Write R programs to handle data • Build analytical models and draw useful inferences from them • Discover the basic concepts of data mining and machine learning • Carry out predictive modeling • Define a business issue as an analytical problem Who This Book Is For Beginners who want to understand and learn the fundamentals of analytics using R. Students, managers, executives, strategy and planning professionals, software professionals, and BI/DW professionals.

### **R for Data Science**

#### **Data Science and Predictive Analytics**

A hands-on guide for professionals to perform various data science tasks in R Key Features Explore the popular R packages for data science Use R for efficient data mining, text analytics and feature engineering Become a thorough data science professional with the help of hands-on examples and use-cases in R Book Description R is the most widely used programming language, and when used in association with data science, this powerful combination will solve the complexities involved with unstructured datasets in the real world. This book covers the entire data science ecosystem for aspiring data scientists, right from zero to a level where you are confident enough to get hands-on with real-world data science problems. The book starts with an introduction to data science and introduces readers to popular R libraries for executing data science routine tasks. This book covers all the important processes in data science such as data gathering, cleaning data, and then uncovering patterns from it. You will explore algorithms such as machine learning algorithms, predictive analytical models, and finally deep learning algorithms. You will learn to run the most powerful visualization packages available in R so as to ensure that you can easily derive insights from your data. Towards the end, you will also learn how to integrate R with Spark and Hadoop and perform large-scale data analytics without much complexity. What you will learn Understand the R programming language and its ecosystem of packages for data science Obtain and clean your data before processing Master essential exploratory techniques for summarizing data Examine various machine learning prediction, models Explore the H2O analytics platform in R for deep learning Apply data mining techniques to available datasets Work with interactive visualization packages in R Integrate R with Spark and Hadoop for large-scale data analytics Who this book is for If you are a budding data scientist keen to learn about the popular pandas library, or a Python developer looking to step into the world of data analysis, this book is the ideal resource you need to get started. Some programming experience in Python will be helpful to get the most out of this course

#### **Beginning Data Science in R**

The definitive guide for statisticians and data scientists who understand the advantages of becoming proficient in both R and Python The first book of its kind, Python for R Users: A Data Science Approach makes it easy for R programmers to

code in Python and Python users to program in R. Short on theory and long on actionable analytics, it provides readers with a detailed comparative introduction and overview of both languages and features concise tutorials with command-by-command translations—complete with sample code—of R to Python and Python to R. Following an introduction to both languages, the author cuts to the chase with step-by-step coverage of the full range of pertinent programming features and functions, including data input, data inspection/data quality, data analysis, and data visualization. Statistical modeling, machine learning, and data mining—including supervised and unsupervised data mining methods—are treated in detail, as are time series forecasting, text mining, and natural language processing.

- Features a quick-learning format with concise tutorials and actionable analytics
- Provides command-by-command translations of R to Python and vice versa
- Incorporates Python and R code throughout to make it easier for readers to compare and contrast features in both languages
- Offers numerous comparative examples and applications in both programming languages

Designed for use for practitioners and students that know one language and want to learn the other

- Supplies slides useful for teaching and learning either software on a companion website

Python for R Users: A Data Science Approach is a valuable working resource for computer scientists and data scientists that know R and would like to learn Python or are familiar with Python and want to learn R. It also functions as textbook for students of computer science and statistics. A. Ohri is the founder of Decisionstats.com and currently works as a senior data scientist. He has advised multiple startups in analytics off-shoring, analytics services, and analytics education, as well as using social media to enhance buzz for analytics products. Mr. Ohri's research interests include spreading open source analytics, analyzing social media manipulation with mechanism design, simpler interfaces for cloud computing, investigating climate change and knowledge flows. His other books include R for Business Analytics and R for Cloud Computing.

### **Data Mining and Business Analytics with R**

Over the past decade, Big Data have become ubiquitous in all economic sectors, scientific disciplines, and human activities. They have led to striking technological advances, affecting all human experiences. Our ability to manage, understand, interrogate, and interpret such extremely large, multisource, heterogeneous, incomplete, multiscale, and incongruent data has not kept pace with the rapid increase of the volume, complexity and proliferation of the deluge of digital information. There are three reasons for this shortfall. First, the volume of data is increasing much faster than the corresponding rise of our computational processing power (Kryder's law > Moore's law). Second, traditional discipline-bounds inhibit expeditious progress. Third, our education and training activities have fallen behind the accelerated trend of scientific, information, and communication advances. There are very few rigorous instructional resources, interactive learning materials, and dynamic training environments that support active data science learning. The textbook balances the mathematical foundations with dexterous demonstrations and examples of data, tools, modules and workflows that serve as pillars for the urgently needed bridge to close that supply and demand predictive analytic skills gap. Exposing the enormous opportunities presented by the tsunami of Big data, this textbook aims to identify specific knowledge gaps, educational barriers, and workforce readiness deficiencies.

Specifically, it focuses on the development of a transdisciplinary curriculum integrating modern computational methods, advanced data science techniques, innovative biomedical applications, and impactful health analytics. The content of this graduate-level textbook fills a substantial gap in integrating modern engineering concepts, computational algorithms, mathematical optimization, statistical computing and biomedical inference. Big data analytic techniques and predictive scientific methods demand broad transdisciplinary knowledge, appeal to an extremely wide spectrum of readers/learners, and provide incredible opportunities for engagement throughout the academy, industry, regulatory and funding agencies. The two examples below demonstrate the powerful need for scientific knowledge, computational abilities, interdisciplinary expertise, and modern technologies necessary to achieve desired outcomes (improving human health and optimizing future return on investment). This can only be achieved by appropriately trained teams of researchers who can develop robust decision support systems using modern techniques and effective end-to-end protocols, like the ones described in this textbook.

- A geriatric neurologist is examining a patient complaining of gait imbalance and posture instability. To determine if the patient may suffer from Parkinson's disease, the physician acquires clinical, cognitive, phenotypic, imaging, and genetics data (Big Data). Most clinics and healthcare centers are not equipped with skilled data analytic teams that can wrangle, harmonize and interpret such complex datasets. A learner that completes a course of study using this textbook will have the competency and ability to manage the data, generate a protocol for deriving biomarkers, and provide an actionable decision support system. The results of this protocol will help the physician understand the entire patient dataset and assist in making a holistic evidence-based, data-driven, clinical diagnosis.
- To improve the return on investment for their shareholders, a healthcare manufacturer needs to forecast the demand for their product subject to environmental, demographic, economic, and bio-social sentiment data (Big Data). The organization's data-analytics team is tasked with developing a protocol that identifies, aggregates, harmonizes, models and analyzes these heterogeneous data elements to generate a trend forecast. This system needs to provide an automated, adaptive, scalable, and reliable prediction of the optimal investment, e.g., R&D allocation, that maximizes the company's bottom line. A reader that complete a course of study using this textbook will be able to ingest the observed structured and unstructured data, mathematically represent the data as a computable object, apply appropriate model-based and model-free prediction techniques. The results of these techniques may be used to forecast the expected relation between the company's investment, product supply, general demand of healthcare (providers and patients), and estimate the return on initial investments.

### **Basketball Data Science**

Learn data science by doing data science! Data Science Using Python and R will get you plugged into the world's two most widespread open-source platforms for data science: Python and R. Data science is hot. Bloomberg called data scientist "the hottest job in America." Python and R are the top two open-source data science tools in the world. In Data Science Using Python and R, you will learn step-by-step how to produce hands-on solutions to real-world business problems, using state-of-the-art techniques. Data Science Using Python and R is written for the

general reader with no previous analytics or programming experience. An entire chapter is dedicated to learning the basics of Python and R. Then, each chapter presents step-by-step instructions and walkthroughs for solving data science problems using Python and R. Those with analytics experience will appreciate having a one-stop shop for learning how to do data science using Python and R. Topics covered include data preparation, exploratory data analysis, preparing to model the data, decision trees, model evaluation, misclassification costs, naïve Bayes classification, neural networks, clustering, regression modeling, dimension reduction, and association rules mining. Further, exciting new topics such as random forests and general linear models are also included. The book emphasizes data-driven error costs to enhance profitability, which avoids the common pitfalls that may cost a company millions of dollars. Data Science Using Python and R provides exercises at the end of every chapter, totaling over 500 exercises in the book. Readers will therefore have plenty of opportunity to test their newfound data science skills and expertise. In the Hands-on Analysis exercises, readers are challenged to solve interesting business problems using real-world data sets.

### **Data Analysis for the Life Sciences with R**

Using data from one season of NBA games, Basketball Data Science: With Applications in R is the perfect book for anyone interested in learning and applying data analytics in basketball. Whether assessing the spatial performance of an NBA player's shots or doing an analysis of the impact of high pressure game situations on the probability of scoring, this book discusses a variety of case studies and hands-on examples using a custom R package. The codes are supplied so readers can reproduce the analyses themselves or create their own. Assuming a basic statistical knowledge, Basketball Data Science with R is suitable for students, technicians, coaches, data analysts and applied researchers. Features:

- One of the first books to provide statistical and data mining methods for the growing field of analytics in basketball.
- Presents tools for modelling graphs and figures to visualize the data.
- Includes real world case studies and examples, such as estimations of scoring probability using the Golden State Warriors as a test case.
- Provides the source code and data so readers can do their own analyses on NBA teams and players.

### **Using R to Unlock the Value of Big Data: Big Data Analytics with Oracle R Enterprise and Oracle R Connector for Hadoop**

This book covers several of the statistical concepts and data analytic skills needed to succeed in data-driven life science research. The authors proceed from relatively basic concepts related to computed p-values to advanced topics related to analyzing highthroughput data. They include the R code that performs this analysis and connect the lines of code to the statistical and mathematical concepts explained.

### **Predictive Analytics For Dummies**

Learn, by example, the fundamentals of data analysis as well as several intermediate to advanced methods and techniques ranging from classification and

regression to Bayesian methods and MCMC, which can be put to immediate use.

**Key Features**

- Analyze your data using R – the most powerful statistical programming language
- Learn how to implement applied statistics using practical use-cases
- Use popular R packages to work with unstructured and structured data

**Book Description**

Frequently the tool of choice for academics, R has spread deep into the private sector and can be found in the production pipelines at some of the most advanced and successful enterprises. The power and domain-specificity of R allows the user to express complex analytics easily, quickly, and succinctly. Starting with the basics of R and statistical reasoning, this book dives into advanced predictive analytics, showing how to apply those techniques to real-world data though with real-world examples. Packed with engaging problems and exercises, this book begins with a review of R and its syntax with packages like Rcpp, ggplot2, and dplyr. From there, get to grips with the fundamentals of applied statistics and build on this knowledge to perform sophisticated and powerful analytics. Solve the difficulties relating to performing data analysis in practice and find solutions to working with messy data, large data, communicating results, and facilitating reproducibility. This book is engineered to be an invaluable resource through many stages of anyone's career as a data analyst. What you will learn

- Gain a thorough understanding of statistical reasoning and sampling theory
- Employ hypothesis testing to draw inferences from your data
- Learn Bayesian methods for estimating parameters
- Train regression, classification, and time series models
- Handle missing data gracefully using multiple imputation
- Identify and manage problematic data points
- Learn how to scale your analyses to larger data with Rcpp, data.table, dplyr, and parallelization
- Put best practices into effect to make your job easier and facilitate reproducibility

Who this book is for

Budding data scientists and data analysts who are new to the concept of data analysis, or who want to build efficient analytical models in R will find this book to be useful. No prior exposure to data analysis is needed, although a fundamental understanding of the R programming language is required to get the best out of this book.

### **Pentaho for Big Data Analytics**

Introduction to Data Science: Data Analysis and Prediction Algorithms with R introduces concepts and skills that can help you tackle real-world data analysis challenges. It covers concepts from probability, statistical inference, linear regression, and machine learning. It also helps you develop skills such as R programming, data wrangling, data visualization, predictive algorithm building, file organization with UNIX/Linux shell, version control with Git and GitHub, and reproducible document preparation. This book is a textbook for a first course in data science. No previous knowledge of R is necessary, although some experience with programming may be helpful. The book is divided into six parts: R, data visualization, statistics with R, data wrangling, machine learning, and productivity tools. Each part has several chapters meant to be presented as one lecture. The author uses motivating case studies that realistically mimic a data scientist's experience. He starts by asking specific questions and answers these through data analysis so concepts are learned as a means to answering the questions. Examples of the case studies included are: US murder rates by state, self-reported student heights, trends in world health and economics, the impact of vaccines on infectious disease rates, the financial crisis of 2007-2008, election forecasting, building a baseball team, image processing of hand-written digits, and movie

recommendation systems. The statistical concepts used to answer the case study questions are only briefly introduced, so complementing with a probability and statistics textbook is highly recommended for in-depth understanding of these concepts. If you read and understand the chapters and complete the exercises, you will be prepared to learn the more advanced concepts and skills needed to become an expert.

### **Data Science and Big Data Analytics**

Data Science in Education Using R is the go-to reference for learning data science in the education field. The book answers questions like: What does a data scientist in education do? How do I get started learning R, the popular open-source statistical programming language? And what does a data analysis project in education look like? If you're just getting started with R in an education job, this is the book you'll want with you. This book gets you started with R by teaching the building blocks of programming that you'll use many times in your career. The book takes a "learn by doing" approach and offers eight analysis walkthroughs that show you a data analysis from start to finish, complete with code for you to practice with. The book finishes with how to get involved in the data science community and how to integrate data science in your education job. This book will be an essential resource for education professionals and researchers looking to increase their data analysis skills as part of their professional and academic development.

### **The Essentials of Data Science: Knowledge Discovery Using R**

If you are a data analyst who has a firm grip on some advanced data analysis techniques and wants to learn how to leverage the features of R, this is the book for you. You should have some basic knowledge of the R language and should know about some data science topics.

### **Python for R Users**

Collecting, analyzing, and extracting valuable information from a large amount of data requires easily accessible, robust, computational and analytical tools. Data Mining and Business Analytics with R utilizes the open source software R for the analysis, exploration, and simplification of large high-dimensional data sets. As a result, readers are provided with the needed guidance to model and interpret complicated data and become adept at building powerful models for prediction and classification. Highlighting both underlying concepts and practical computational skills, Data Mining and Business Analytics with R begins with coverage of standard linear regression and the importance of parsimony in statistical modeling. The book includes important topics such as penalty-based variable selection (LASSO); logistic regression; regression and classification trees; clustering; principal components and partial least squares; and the analysis of text and network data. In addition, the book presents:

- A thorough discussion and extensive demonstration of the theory behind the most useful data mining tools
- Illustrations of how to use the outlined concepts in real-world situations
- Readily available additional data sets and related R code allowing readers to apply their own analyses to the discussed materials
-

Numerous exercises to help readers with computing skills and deepen their understanding of the material. Data Mining and Business Analytics with R is an excellent graduate-level textbook for courses on data mining and business analytics. The book is also a valuable reference for practitioners who collect and analyze data in the fields of finance, operations management, marketing, and the information sciences.

### **Data Science Using Python and R**

An Introduction to Data Science by Jeffrey S. Saltz and Jeffrey M. Stanton is an easy-to-read, gentle introduction for people with a wide range of backgrounds into the world of data science. Needing no prior coding experience or a deep understanding of statistics, this book uses the R programming language and RStudio® platform to make data science welcoming and accessible for all learners. After introducing the basics of data science, the book builds on each previous concept to explain R programming from the ground up. Readers will learn essential skills in data science through demonstrations of how to use data to construct models, predict outcomes, and visualize data.

### **R for Data Science**

Data Science and Big Data Analytics is about harnessing the power of data for new insights. The book covers the breadth of activities and methods and tools that Data Scientists use. The content focuses on concepts, principles and practical applications that are applicable to any industry and technology environment, and the learning is supported and explained with examples that you can replicate using open-source software. This book will help you: Become a contributor on a data science team Deploy a structured lifecycle approach to data analytics problems Apply appropriate analytic techniques and tools to analyzing big data Learn how to tell a compelling story with data to drive business action Prepare for EMC Proven Professional Data Science Certification Corresponding data sets are available at [www.wiley.com/go/9781118876138](http://www.wiley.com/go/9781118876138). Get started discovering, analyzing, visualizing, and presenting data in a meaningful way today!

### **Introduction to Data Science**

Load, wrangle, and analyze your data using the world's most powerful statistical programming language About This Book Load, manipulate and analyze data from different sources Gain a deeper understanding of fundamentals of applied statistics A practical guide to performing data analysis in practice Who This Book Is For Whether you are learning data analysis for the first time, or you want to deepen the understanding you already have, this book will prove to be an invaluable resource. If you are looking for a book to bring you all the way through the fundamentals to the application of advanced and effective analytics methodologies, and have some prior programming experience and a mathematical background, then this is for you. What You Will Learn Navigate the R environment Describe and visualize the behavior of data and relationships between data Gain a thorough understanding of statistical reasoning and sampling Employ hypothesis tests to draw inferences from your data Learn Bayesian methods for estimating parameters Perform



regression to predict continuous variables Apply powerful classification methods to predict categorical data Handle missing data gracefully using multiple imputation Identify and manage problematic data points Employ parallelization and Rcpp to scale your analyses to larger data Put best practices into effect to make your job easier and facilitate reproducibility In Detail Frequently the tool of choice for academics, R has spread deep into the private sector and can be found in the production pipelines at some of the most advanced and successful enterprises. The power and domain-specificity of R allows the user to express complex analytics easily, quickly, and succinctly. With over 7,000 user contributed packages, it's easy to find support for the latest and greatest algorithms and techniques. Starting with the basics of R and statistical reasoning, *Data Analysis with R* dives into advanced predictive analytics, showing how to apply those techniques to real-world data though with real-world examples. Packed with engaging problems and exercises, this book begins with a review of R and its syntax. From there, get to grips with the fundamentals of applied statistics and build on this knowledge to perform sophisticated and powerful analytics. Solve the difficulties relating to performing data analysis in practice and find solutions to working with "messy data", large data, communicating results, and facilitating reproducibility. This book is engineered to be an invaluable resource through many stages of anyone's career as a data analyst. Style and approach Learn data analysis using engaging examples and fun exercises, and with a gentle and friendly but comprehensive "learn-by-doing" approach.

### **Mastering Data Analysis with R**

*Big Data Analytics with R and Hadoop* is a tutorial style book that focuses on all the powerful big data tasks that can be achieved by integrating R and Hadoop. This book is ideal for R developers who are looking for a way to perform big data analytics with Hadoop. This book is also aimed at those who know Hadoop and want to build some intelligent applications over Big data with R packages. It would be helpful if readers have basic knowledge of R.

### **Exam Prep for: Using R to Unlock the Value of Big Data; Big**

*Modern Data Science with R* is a comprehensive data science textbook for undergraduates that incorporates statistical and computational thinking to solve real-world problems with data. Rather than focus exclusively on case studies or programming syntax, this book illustrates how statistical programming in the state-of-the-art R/RStudio computing environment can be leveraged to extract meaningful information from a variety of data in the service of addressing compelling statistical questions. Contemporary data science requires a tight integration of knowledge from statistics, computer science, mathematics, and a domain of application. This book will help readers with some background in statistics and modest prior experience with coding develop and practice the appropriate skills to tackle complex data science projects. The book features a number of exercises and has a flexible organization conducive to teaching a variety of semester courses.

### **Introduction to Data Science for Social and Policy Research**

While the term Big Data is open to varying interpretation, it is quite clear that the Volume, Velocity, and Variety (3Vs) of data have impacted every aspect of computational science and its applications. The volume of data is increasing at a phenomenal rate and a majority of it is unstructured. With big data, the volume is so large that processing it using traditional database and software techniques is difficult, if not impossible. The drivers are the ubiquitous sensors, devices, social networks and the all-pervasive web. Scientists are increasingly looking to derive insights from the massive quantity of data to create new knowledge. In common usage, Big Data has come to refer simply to the use of predictive analytics or other certain advanced methods to extract value from data, without any required magnitude thereon. Challenges include analysis, capture, curation, search, sharing, storage, transfer, visualization, and information privacy. While there are challenges, there are huge opportunities emerging in the fields of Machine Learning, Data Mining, Statistics, Human-Computer Interfaces and Distributed Systems to address ways to analyze and reason with this data. The edited volume focuses on the challenges and opportunities posed by "Big Data" in a variety of domains and how statistical techniques and innovative algorithms can help glean insights and accelerate discovery. Big data has the potential to help companies improve operations and make faster, more intelligent decisions. Review of big data research challenges from diverse areas of scientific endeavor Rich perspective on a range of data science issues from leading researchers Insight into the mathematical and statistical theory underlying the computational methods used to address big data analytics problems in a variety of domains

### **Big Data Analytics with R and Hadoop**

The book is a practical guide, full of step-by-step examples that are easy to follow and implement. This book is for developers, system administrators, and business intelligence professionals looking to learn how to get more out of their data through Pentaho. In order to best engage with the examples, some knowledge of Java will be required.

### **Financial Analytics with R**

We are in an age of big data where all of our everyday interactions and transactions generate data. Much of this data is spatial - it is collected some-where - and identifying analytical insight from trends and patterns in these increasing rich digital footprints presents a number of challenges. Whilst other books describe different flavours of Data Analytics in R and other programming languages, there are none that consider Spatial Data (ie the location attached to data), or that consider issues of inference, linking Big Data, Geography, GIS, Mapping and Spatial Analytics. This is a 'learning by doing' text book, building on the previous book by the same authors, An Introduction to R for Spatial Analysis and Mapping. It details the theoretical issues in analyses of Big Spatial Data and developing practical skills in the reader for addressing these with confidence.

### **An Approach for Handling Big Data Analytics with R Using Air Quality Data**

The field of medical imaging seen rapid development over the last two decades and has consequently revolutionized the way in which modern medicine is practiced. Diseases and their symptoms are constantly changing therefore continuous updating is necessary for the data to be relevant. Diseases fall into different categories, even a small difference in symptoms may result in categorising it in a different group altogether. Thus analysing data accurately is of critical importance. This book concentrates on diagnosing diseases like cancer or tumor from different modalities of images. This book is divided into the following domains: Importance of big data in medical imaging, pre-processing, image registration, feature extraction, classification and retrieval. It is further supplemented by the medical analyst for a continuous treatment process. The book provides an automated system that could retrieve images based on user's interest to a point of providing decision support. It will help medical analysts to take informed decisions before planning treatment and surgery. It will also be useful to researchers who are working in problems involved in medical imaging.

[ROMANCE](#) [ACTION & ADVENTURE](#) [MYSTERY & THRILLER](#) [BIOGRAPHIES & HISTORY](#) [CHILDREN'S](#) [YOUNG ADULT](#) [FANTASY](#) [HISTORICAL FICTION](#) [HORROR](#) [LITERARY FICTION](#) [NON-FICTION](#) [SCIENCE FICTION](#)