

---

# R For Data Science O'Reilly Media

---

Data Science Using Python and R  
Mastering Shiny  
R Packages  
Data Analysis with Open Source Tools  
R for Data Science  
Hands-On Data Analysis with Pandas  
Doing Data Science  
Python for Data Analysis  
Fundamentals of Data Visualization  
The Book of R  
R Markdown  
Bioinformatics Data Skills  
Data Science from Scratch  
Mastering Spark with R  
Efficient R Programming  
Python for R Users  
R for Data Science  
Data Science Foundations Tools and Techniques  
Big Data Analytics with R  
Data Science for Business  
R Cookbook  
Hands-On Programming with R  
Python and R for the Modern Data Scientist  
Ethics and Data Science  
Learning R  
Beginning Data Science in R  
Practical Statistics for Data Scientists  
R: Data Analysis and Visualization  
R for Data Science Cookbook  
Data Science at the Command Line  
Hands-On Machine Learning with R  
Data Analysis with R  
Data Analysis with R, Second Edition  
R for Data Science  
Simulation for Data Science with R  
Introduction to Data Science  
Introduction to Machine Learning with R  
Text Mining with R

---

## HOLLAND JAYLIN

---

### Data Science Using Python and R No Starch Press

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 Shiny "O'Reilly Media, Inc."

*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.

### R Packages "O'Reilly Media, Inc."

Machine learning is an intimidating subject until you know the fundamentals. If you understand basic coding concepts, this introductory guide will help you gain a solid foundation in machine learning principles. Using the R programming language, you'll first start to learn with regression modelling and then move into more advanced topics such as neural networks and tree-based methods. Finally, you'll delve into the frontier of machine learning, using the caret package in R. Once you develop a familiarity with topics such as the difference between regression and classification models, you'll be able to solve an array of machine learning problems. Author Scott V. Burger provides several examples to help you build a working knowledge of machine learning. Explore machine learning models, algorithms, and data training Understand machine learning algorithms for supervised and unsupervised cases Examine statistical concepts for designing data for use in models Dive into linear regression models used in business and science Use single-layer and multilayer neural networks for calculating outcomes Look at how tree-based models work, including popular decision trees Get a comprehensive view of the machine learning ecosystem in R Explore the powerhouse of tools available in R's caret package *Data Analysis with Open Source Tools* CRC Press With more than 200 practical recipes, this book helps you perform data analysis with R quickly and efficiently. The R language provides everything you need to do statistical work, but its structure can be difficult to master. This collection of concise, task-oriented recipes makes you productive with R immediately, with solutions ranging from basic tasks to input and output, general statistics, graphics, and linear regression. Each recipe addresses a specific problem, with a discussion that explains the solution and offers insight into how it works. If you're a beginner, *R Cookbook* will help get you started. If you're an experienced

data programmer, it will jog your memory and expand your horizons. You'll get the job done faster and learn more about R in the process. Create vectors, handle variables, and perform other basic functions Input and output data Tackle data structures such as matrices, lists, factors, and data frames Work with probability, probability distributions, and random variables Calculate statistics and confidence intervals, and perform statistical tests Create a variety of graphic displays Build statistical models with linear regressions and analysis of variance (ANOVA) Explore advanced statistical techniques, such as finding clusters in your data

"Wonderfully readable, R Cookbook serves not only as a solutions manual of sorts, but as a truly enjoyable way to explore the R language—one practical example at a time."—Jeffrey Ryan, software consultant and R package author  
*R for Data Science* "O'Reilly Media, Inc."

R is the world's most popular language for developing statistical software: Archaeologists use it to track the spread of ancient civilizations, drug companies use it to discover which medications are safe and effective, and actuaries use it to assess financial risks and keep economies running smoothly. The Art of R Programming takes you on a guided tour of software development with R, from basic types and data structures to advanced topics like closures, recursion, and anonymous functions. No statistical knowledge is required, and your programming skills can range from hobbyist to pro. Along the way, you'll learn about functional and object-oriented programming, running mathematical simulations, and rearranging complex data into simpler, more useful formats. You'll also learn to: -Create artful graphs to visualize complex data sets and functions -Write more efficient code using parallel R and vectorization -Interface R with C/C++ and Python for increased speed or functionality -Find new R packages for text analysis, image manipulation, and more -Squash annoying bugs with advanced debugging techniques Whether you're designing aircraft, forecasting the weather, or you just need to tame your data, The Art of R Programming is your guide to harnessing the power of statistical computing.

*Hands-On Data Analysis with Pandas* John Wiley & Sons  
Success in data science depends on the flexible and appropriate use of tools. That includes Python and R, two of the foundational programming languages in the field. This book guides data

scientists from the Python and R communities along the path to becoming bilingual. By recognizing the strengths of both languages, you'll discover new ways to accomplish data science tasks and expand your skill set. Authors Rick Scavetta and Boyan Angelov explain the parallel structures of these languages and highlight where each one excels, whether it's their linguistic features or the powers of their open source ecosystems. You'll learn how to use Python and R together in real-world settings and broaden your job opportunities as a bilingual data scientist. Learn Python and R from the perspective of your current language Understand the strengths and weaknesses of each language Identify use cases where one language is better suited than the other Understand the modern open source ecosystem available for both, including packages, frameworks, and workflows Learn how to integrate R and Python in a single workflow Follow a case study that demonstrates ways to use these languages together  
**Doing Data Science** "O'Reilly Media, Inc."

Get to grips with pandas—a versatile and high-performance Python library for data manipulation, analysis, and discovery Key Features Perform efficient data analysis and manipulation tasks using pandas Apply pandas to different real-world domains using step-by-step demonstrations Get accustomed to using pandas as an effective data exploration tool Book Description Data analysis has become a necessary skill in a variety of positions where knowing how to work with data and extract insights can generate significant value. Hands-On Data Analysis with Pandas will show you how to analyze your data, get started with machine learning, and work effectively with Python libraries often used for data science, such as pandas, NumPy, matplotlib, seaborn, and scikit-learn. Using real-world datasets, you will learn how to use the powerful pandas library to perform data wrangling to reshape, clean, and aggregate your data. Then, you will learn how to conduct exploratory data analysis by calculating summary statistics and visualizing the data to find patterns. In the concluding chapters, you will explore some applications of anomaly detection, regression, clustering, and classification, using scikit-learn, to make predictions based on past data. By the end of this book, you will be equipped with the skills you need to use pandas to ensure the veracity of your data, visualize it for effective decision-making, and reliably reproduce analyses across multiple datasets. What you will learn Understand how data

analysts and scientists gather and analyze data Perform data analysis and data wrangling in Python Combine, group, and aggregate data from multiple sources Create data visualizations with pandas, matplotlib, and seaborn Apply machine learning (ML) algorithms to identify patterns and make predictions Use Python data science libraries to analyze real-world datasets Use pandas to solve common data representation and analysis problems Build Python scripts, modules, and packages for reusable analysis code Who this book is for This book is for data analysts, data science beginners, and Python developers who want to explore each stage of data analysis and scientific computing using a wide range of datasets. You will also find this book useful if you are a data scientist who is looking to implement pandas in machine learning. Working knowledge of Python programming language will be beneficial.

**Python for Data Analysis** O'Reilly Media

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.

[Fundamentals of Data Visualization](#) Packt Publishing Ltd

Use R to turn data into insight, knowledge, and understanding. With this practical book, aspiring data scientists will learn how to do data science with R and RStudio, along with the tidyverse—a collection of R packages designed to work together to make data science fast, fluent, and fun. Even if you have no programming experience, this updated edition will have you doing data science quickly. You'll learn how to import, transform, and visualize your data and communicate the results. And you'll get a complete, big-picture understanding of the data science cycle and the basic tools you need to manage the details. Updated for the latest tidyverse features and best practices, new chapters show you how to get data from spreadsheets, databases, and websites. Exercises help you practice what you've learned along the way. You'll understand how to: Visualize: Create plots for data exploration and communication of results Transform: Discover variable types and the tools to work with them Import: Get data into R and in a form convenient for analysis Program: Learn R tools for solving data problems with greater clarity and ease Communicate: Integrate prose, code, and results with Quarto

[The Book of R](#) Packt Publishing Ltd

Statistical methods are a key part of of data science, yet very few data scientists have any formal statistics training. Courses and books on basic statistics rarely cover the topic from a data science perspective. This practical guide explains how to apply various statistical methods to data science, tells you how to avoid

their misuse, and gives you advice on what's important and what's not. Many data science resources incorporate statistical methods but lack a deeper statistical perspective. If you're familiar with the R programming language, and have some exposure to statistics, this quick reference bridges the gap in an accessible, readable format. With this book, you'll learn: Why exploratory data analysis is a key preliminary step in data science How random sampling can reduce bias and yield a higher quality dataset, even with big data How the principles of experimental design yield definitive answers to questions How to use regression to estimate outcomes and detect anomalies Key classification techniques for predicting which categories a record belongs to Statistical machine learning methods that "learn" from data Unsupervised learning methods for extracting meaning from unlabeled data

*R Markdown* "O'Reilly Media, Inc."

Harness actionable insights from your data with computational statistics and simulations using R About This Book Learn five different simulation techniques (Monte Carlo, Discrete Event Simulation, System Dynamics, Agent-Based Modeling, and Resampling) in-depth using real-world case studies A unique book that teaches you the essential and fundamental concepts in statistical modeling and simulation Who This Book Is For This book is for users who are familiar with computational methods. If you want to learn about the advanced features of R, including the computer-intense Monte-Carlo methods as well as computational tools for statistical simulation, then this book is for you. Good knowledge of R programming is assumed/required. What You Will Learn The book aims to explore advanced R features to simulate data to extract insights from your data. Get to know the advanced features of R including high-performance computing and advanced data manipulation See random number simulation used to simulate distributions, data sets, and populations Simulate close-to-reality populations as the basis for agent-based micro-, model- and design-based simulations Applications to design statistical solutions with R for solving scientific and real world problems Comprehensive coverage of several R statistical packages like boot, simPop, VIM, data.table, dplyr, parallel, StatDA, simecol, simecolModels, deSolve and many more. In Detail Data Science with R aims to teach you how to begin performing data science tasks by taking advantage of Rs powerful

ecosystem of packages. R being the most widely used programming language when used with data science can be a powerful combination to solve complexities involved with varied data sets in the real world. The book will provide a computational and methodological framework for statistical simulation to the users. Through this book, you will get in grips with the software environment R. After getting to know the background of popular methods in the area of computational statistics, you will see some applications in R to better understand the methods as well as gaining experience of working with real-world data and real-world problems. This book helps uncover the large-scale patterns in complex systems where interdependencies and variation are critical. An effective simulation is driven by data generating processes that accurately reflect real physical populations. You will learn how to plan and structure a simulation project to aid in the decision-making process as well as the presentation of results. By the end of this book, you reader will get in touch with the software environment R. After getting background on popular methods in the area, you will see applications in R to better understand the methods as well as to gain experience when working on real-world data and real-world problems. Style and approach This book takes a practical, hands-on approach to explain the statistical computing methods, gives advice on the usage of these methods, and provides computational tools to help you solve common problems in statistical simulation and computer-intense methods.

[Bioinformatics Data Skills](#) No Starch Press

Learn how to program by diving into the R language, and then use your newfound skills to solve practical data science problems. With this book, you'll learn how to load data, assemble and disassemble data objects, navigate R's environment system, write your own functions, and use all of R's programming tools. RStudio Master Instructor Garrett Golemund not only teaches you how to program, but also shows you how to get more from R than just visualizing and modeling data. You'll gain valuable programming skills and support your work as a data scientist at the same time. Work hands-on with three practical data analysis projects based on casino games Store, retrieve, and change data values in your computer's memory Write programs and simulations that outperform those written by typical R users Use R programming tools such as if else statements, for loops, and S3 classes Learn

how to write lightning-fast vectorized R code Take advantage of R's package system and debugging tools Practice and apply R programming concepts as you learn them

[Data Science from Scratch](#) "O'Reilly Media, Inc."

Learn how to perform data analysis with the R language and software environment, even if you have little or no programming experience. With the tutorials in this hands-on guide, you'll learn how to use the essential R tools you need to know to analyze data, including data types and programming concepts. The second half of Learning R shows you real data analysis in action by covering everything from importing data to publishing your results. Each chapter in the book includes a quiz on what you've learned, and concludes with exercises, most of which involve writing R code. Write a simple R program, and discover what the language can do Use data types such as vectors, arrays, lists, data frames, and strings Execute code conditionally or repeatedly with branches and loops Apply R add-on packages, and package your own work for others Learn how to clean data you import from a variety of sources Understand data through visualization and summary statistics Use statistical models to pass quantitative judgments about data and make predictions Learn what to do when things go wrong while writing data analysis code

[Mastering Spark with R](#) "O'Reilly Media, Inc."

If you're like most R users, you have deep knowledge and love for statistics. But as your organization continues to collect huge amounts of data, adding tools such as Apache Spark makes a lot of sense. With this practical book, data scientists and professionals working with large-scale data applications will learn how to use Spark from R to tackle big data and big compute problems. Authors Javier Luraschi, Kevin Kuo, and Edgar Ruiz show you how to use R with Spark to solve different data analysis problems. This book covers relevant data science topics, cluster computing, and issues that should interest even the most advanced users. Analyze, explore, transform, and visualize data in Apache Spark with R Create statistical models to extract information and predict outcomes; automate the process in production-ready workflows Perform analysis and modeling across many machines using distributed computing techniques Use large-scale data from multiple sources and different formats with ease from within Spark Learn about alternative modeling frameworks for graph processing, geospatial analysis, and

genomics at scale Dive into advanced topics including custom transformations, real-time data processing, and creating custom Spark extensions

[Efficient R Programming](#) "O'Reilly Media, Inc."

The Foundational Hands-On Skills You Need to Dive into Data Science "Freeman and Ross have created the definitive resource for new and aspiring data scientists to learn foundational programming skills." -From the foreword by Jared Lander, series editor Using data science techniques, you can transform raw data into actionable insights for domains ranging from urban planning to precision medicine. Programming Skills for Data Science brings together all the foundational skills you need to get started, even if you have no programming or data science experience. Leading instructors Michael Freeman and Joel Ross guide you through installing and configuring the tools you need to solve professional-level data science problems, including the widely used R language and Git version-control system. They explain how to wrangle your data into a form where it can be easily used, analyzed, and visualized so others can see the patterns you've uncovered. Step by step, you'll master powerful R programming techniques and troubleshooting skills for probing data in new ways, and at larger scales. Freeman and Ross teach through practical examples and exercises that can be combined into complete data science projects. Everything's focused on real-world application, so you can quickly start analyzing your own data and getting answers you can act upon. Learn to Install your complete data science environment, including R and RStudio Manage projects efficiently, from version tracking to documentation Host, manage, and collaborate on data science projects with GitHub Master R language fundamentals: syntax, programming concepts, and data structures Load, format, explore, and restructure data for successful analysis Interact with databases and web APIs Master key principles for visualizing data accurately and intuitively Produce engaging, interactive visualizations with ggplot and other R packages Transform analyses into sharable documents and sites with R Markdown Create interactive web data science applications with Shiny Collaborate smoothly as part of a data science team Register your book for convenient access to downloads, updates, and/or corrections as they become available. See inside book for details. [Python for R Users](#) John Wiley & Sons

For many researchers, Python is a first-class tool mainly because of its libraries for storing, manipulating, and gaining insight from data. Several resources exist for individual pieces of this data science stack, but only with the Python Data Science Handbook do you get them all—IPython, NumPy, Pandas, Matplotlib, Scikit-Learn, and other related tools. Working scientists and data crunchers familiar with reading and writing Python code will find this comprehensive desk reference ideal for tackling day-to-day issues: manipulating, transforming, and cleaning data; visualizing different types of data; and using data to build statistical or machine learning models. Quite simply, this is the must-have reference for scientific computing in Python. With this handbook, you'll learn how to use: IPython and Jupyter: provide computational environments for data scientists using Python NumPy: includes the ndarray for efficient storage and manipulation of dense data arrays in Python Pandas: features the DataFrame for efficient storage and manipulation of labeled/columnar data in Python Matplotlib: includes capabilities for a flexible range of data visualizations in Python Scikit-Learn: for efficient and clean Python implementations of the most important and established machine learning algorithms

**R for Data Science** Addison-Wesley Professional Effective visualization is the best way to communicate information from the increasingly large and complex datasets in the natural and social sciences. But with the increasing power of visualization software today, scientists, engineers, and business analysts often have to navigate a bewildering array of visualization choices and options. This practical book takes you through many commonly encountered visualization problems, and it provides guidelines on how to turn large datasets into clear and compelling figures. What visualization type is best for the story you want to tell? How do you make informative figures that are visually pleasing? Author Claus O. Wilke teaches you the elements most critical to successful data visualization. Explore the basic concepts of color as a tool to highlight, distinguish, or represent a value Understand the importance of redundant coding to ensure you provide key information in multiple ways Use the book's visualizations directory, a graphical guide to commonly used types of data visualizations Get extensive examples of good and bad figures Learn how to use figures in a document or report and how employ them effectively to tell a compelling story

*Data Science Foundations Tools and Techniques* CRC Press  
 Over 100 hands-on recipes to effectively solve real-world data problems using the most popular R packages and techniques  
 About This Book Gain insight into how data scientists collect, process, analyze, and visualize data using some of the most popular R packages  
 Understand how to apply useful data analysis techniques in R for real-world applications  
 An easy-to-follow guide to make the life of data scientist easier with the problems faced while performing data analysis  
 Who This Book Is For This book is for those who are already familiar with the basic operation of R, but want to learn how to efficiently and effectively analyze real-world data problems using practical R packages.  
 What You Will Learn Get to know the functional characteristics of R language  
 Extract, transform, and load data from heterogeneous sources  
 Understand how easily R can confront probability and statistics problems  
 Get simple R instructions to quickly organize and manipulate large datasets  
 Create professional data visualizations and interactive reports  
 Predict user purchase behavior by adopting a classification approach  
 Implement data mining techniques to discover items that are frequently purchased together  
 Group similar text documents by using various clustering methods  
 In Detail This cookbook offers a range of data analysis samples in simple and straightforward R code, providing step-by-step resources and time-saving methods to help you solve data problems efficiently. The first section deals with how to create R functions to avoid the unnecessary duplication of code. You will learn how to prepare, process, and perform sophisticated ETL for heterogeneous data sources with R packages. An example of data manipulation is provided, illustrating how to use the “dplyr” and

“data.table” packages to efficiently process larger data structures. We also focus on “ggplot2” and show you how to create advanced figures for data exploration. In addition, you will learn how to build an interactive report using the “ggvis” package. Later chapters offer insight into time series analysis on financial data, while there is detailed information on the hot topic of machine learning, including data classification, regression, clustering, association rule mining, and dimension reduction. By the end of this book, you will understand how to resolve issues and will be able to comfortably offer solutions to problems encountered while performing data analysis. Style and approach This easy-to-follow guide is full of hands-on examples of data analysis with R. Each topic is fully explained beginning with the core concept, followed by step-by-step practical examples, and concluding with detailed explanations of each concept used.  
*Big Data Analytics with R* Packt Publishing Ltd  
 This hands-on guide demonstrates how the flexibility of the command line can help you become a more efficient and productive data scientist. You’ll learn how to combine small, yet powerful, command-line tools to quickly obtain, scrub, explore, and model your data. To get you started—whether you’re on Windows, OS X, or Linux—author Jeroen Janssens introduces the Data Science Toolbox, an easy-to-install virtual environment packed with over 80 command-line tools. Discover why the command line is an agile, scalable, and extensible technology. Even if you’re already comfortable processing data with, say, Python or R, you’ll greatly improve your data science workflow by also leveraging the power of the command line. Obtain data from websites, APIs, databases, and spreadsheets  
 Perform scrub

operations on plain text, CSV, HTML/XML, and JSON  
 Explore data, compute descriptive statistics, and create visualizations  
 Manage your data science workflow using Drake  
 Create reusable tools from one-liners and existing Python or R code  
 Parallelize and distribute data-intensive pipelines using GNU Parallel  
 Model data with dimensionality reduction, clustering, regression, and classification algorithms

*Data Science for Business* "O'Reilly Media, Inc."

Now that people are aware that data can make the difference in an election or a business model, data science as an occupation is gaining ground. But how can you get started working in a wide-ranging, interdisciplinary field that’s so clouded in hype? This insightful book, based on Columbia University’s Introduction to Data Science class, tells you what you need to know. In many of these chapter-long lectures, data scientists from companies such as Google, Microsoft, and eBay share new algorithms, methods, and models by presenting case studies and the code they use. If you’re familiar with linear algebra, probability, and statistics, and have programming experience, this book is an ideal introduction to data science. Topics include: Statistical inference, exploratory data analysis, and the data science process  
 Algorithms  
 Spam filters, Naive Bayes, and data wrangling  
 Logistic regression  
 Financial modeling  
 Recommendation engines and causality  
 Data visualization  
 Social networks and data journalism  
 Data engineering, MapReduce, Pregel, and Hadoop  
 Doing Data Science is collaboration between course instructor Rachel Schutt, Senior VP of Data Science at News Corp, and data science consultant Cathy O’Neil, a senior data scientist at Johnson Research Labs, who attended and blogged about the course.

Related with R For Data Science Oreilly Media:

- Mosaic Threat Assessment Test : [click here](#)