

Doing Data Science Straight Talk From The Frontline

Praise for previous editions: "Gandrud has written a great outline of how a fully reproducible research project should look from start to finish, with brief explanations of each tool that he uses along the way... Advanced undergraduate students in mathematics, statistics, and similar fields as well as students just beginning their graduate studies would benefit the most from reading this book. Many more experienced R users or second-year graduate students might find themselves thinking, 'I wish I'd read this book at the start of my studies, when I was first learning R!'...This book could be used as the main text for a class on reproducible research ..." (The American Statistician) Reproducible Research with R and R Studio, Third Edition brings together the skills and tools needed for doing and presenting computational research. Using straightforward examples, the book takes you through an entire reproducible research workflow. This practical workflow enables you to gather and analyze data as well as dynamically present results in print and on the web. Supplementary materials and example are available on the author's website. New to the Third Edition Updated package recommendations, examples, URLs, and removed technologies no longer in regular use. More advanced R Markdown (and less LaTeX) in discussions of markup languages and examples. Stronger focus on reproducible working directory tools. Updated discussion of cloud storage services and persistent reproducible material citation. Added discussion of Jupyter notebooks and reproducible practices in industry. Examples of data manipulation with Tidyverse tibbles (in addition to standard data frames) and pivot_longer() and pivot_wider() functions for pivoting data. Features Incorporates the most important advances that have been developed since the editions were published Describes a complete reproducible research workflow, from data gathering to the presentation of results Shows how to automatically generate tables and figures using R Includes instructions on formatting a presentation document via markup languages Discusses cloud storage and versioning services, particularly Github Explains how to use Unix-like shell programs for working with large research projects

Discover how data science can help you gain in-depth insight into your business - the easy way! Jobs in data science abound, but few people have the data science skills needed to fill these increasingly important roles. Data Science For Dummies is the perfect starting point for IT professionals and students who want a quick primer on all areas of the expansive data science space. With a focus on business cases, the book explores topics in big data, data science, and data engineering, and how these three areas are combined to produce tremendous value. If you want to pick-up the skills you need to begin a new career or initiate a new project, reading this book will help you understand what technologies, programming languages, and mathematical methods on which to focus. While this book serves as a wildly fantastic guide through the broad, sometimes intimidating field of big data and data science, it is not an instruction manual for hands-on implementation. Here's what to expect: Provides a background in big data and data engineering before moving on to data science and how it's applied to generate value Includes coverage of big data frameworks like Hadoop, MapReduce, Spark, MPP platforms, and NoSQL Explains machine learning and many of its algorithms as well as artificial intelligence and the evolution of the Internet of Things Details data visualization techniques that can be used to showcase, summarize, and communicate the data insights you generate It's a big, big data world out there—let Data Science For Dummies help you harness its power and gain a competitive edge for your organization.

"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"--

Written by renowned data science experts Foster Provost and Tom Fawcett, Data Science for Business introduces the fundamental principles of data science, and walks you through the "data-analytic thinking" necessary for extracting useful knowledge and business value from the data you collect. This guide also helps you understand the many data-mining techniques in use today. Based on an MBA course Provost has taught at New York University over the past ten years, Data Science for Business provides examples of real-world business problems to illustrate these principles. You'll not only learn how to improve communication between business stakeholders and data scientists, but also how participate intelligently in your company's data science projects. You'll also discover how to think data-analytically, and fully appreciate how data science methods can support business decision-making. Understand how data science fits in your organization—and how you can use it for competitive advantage Treat data as a business asset that requires careful investment if you're to gain real value Approach business problems data-analytically, using the data-mining process to gather good data in the most appropriate way Learn general concepts for actually extracting knowledge from data Apply data science principles when interviewing data science job candidates

If you know how to program, you have the skills to turn data into knowledge, using tools of probability and statistics. This concise introduction shows you how to perform statistical analysis computationally, rather than mathematically, with programs written in Python. By working with a single case study throughout this thoroughly revised book, you'll learn the entire process of exploratory data analysis—from collecting data and generating statistics to identifying patterns and testing hypotheses. You'll explore distributions, rules of probability, visualization, and many other tools and concepts. New chapters on regression, time series analysis, survival analysis, and analytic methods will enrich your discoveries. Develop an understanding of probability and statistics by writing and testing code Run experiments to test statistical behavior, such as generating samples from several distributions Use simulations to understand concepts that are hard to grasp mathematically Import data from most sources with Python, rather than rely on data that's cleaned and formatted for statistics tools Use statistical inference to answer questions about real-world data

Data Science gets thrown around in the press like it's magic. Major retailers are predicting everything from when their customers are pregnant to when they want a new pair of Chuck Taylors. It's a brave new world where seemingly meaningless data can be transformed into valuable insight to drive smart business decisions. But how does one exactly do data science? Do you have to hire one of these priests of the dark arts, the "data scientist," to extract this gold from your data? Nope. Data science is little more than using straight-forward steps to process raw data into actionable insight. And in DataSmart, author and data scientist John Foreman will show you how that's done within the familiar environment of a spreadsheet. Why a spreadsheet? It's comfortable! You get to look at the data every step of the way, building confidence as you learn the tricks of the trade. Plus, spreadsheets are a vendor-neutral place to learn data science without the hype. But don't let the Excel sheets fool you. This is a book for those serious about learning the analytic techniques, the math and the magic, behind big data. Each chapter will cover a different technique in a spreadsheet so you can follow along: Mathematical optimization, including non-linear programming and genetic algorithms Clustering via k-means, spherical k-means, and graph modularity Data mining in graphs, such as outlier detection Supervised AI through logistic regression, ensemble models, and bag-of-words models Forecasting, seasonal adjustments, and prediction intervals through monte carlo simulation Moving from spreadsheets into the R programming language You get your hands dirty as you work alongside John through each technique. But never fear, the topics are readily applicable and the author laces humor throughout. You'll even learn what a dead squirrel has to do with optimization modeling, which you no doubt are dying to know.

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.

Written for statisticians, computer scientists, geographers, research and applied scientists, and others interested in visualizing data, this book presents a unique foundation for producing almost every quantitative graphic found in scientific journals, newspapers, statistical packages, and data visualization systems. It was designed for a distributed computing environment, with special attention given to conserving computer code and system resources. While the tangible result of this work is a Java production graphics library, the text focuses on the deep structures involved in producing quantitative graphics from data. It investigates the rules that underlie pie charts, bar charts, scatterplots, function plots, maps, mosaics, and radar charts. These rules are abstracted from the work of Bertin, Cleveland, Kosslyn, MacEachren, Pinker, Tufte, Tukey, Tobler, and other theorists of quantitative graphics.

A new way of thinking about data science and data ethics that is informed by the ideas of intersectional feminism. Today, data science is a form of power. It has been used to expose injustice, improve health outcomes, and topple governments. But it has also been used to discriminate, police, and surveil. This potential for good, on the one hand, and harm, on the other, makes it essential to ask: Data science by whom? Data science for whom? Data science with whose interests in mind? The narratives around big data and data science are overwhelmingly white, male, and techno-heroic. In *Data Feminism*, Catherine D'Ignazio and Lauren Klein present a new way of thinking about data science and data ethics—one that is informed by intersectional feminist thought. Illustrating data feminism in action, D'Ignazio and Klein show how challenges to the male/female binary can help challenge other hierarchical (and empirically wrong) classification systems. They explain how, for example, an understanding of emotion can expand our ideas about effective data visualization, and how the concept of invisible labor can expose the significant human efforts required by our automated systems. And they show why the data never, ever “speak for themselves.” *Data Feminism* offers strategies for data scientists seeking to learn how feminism can help them work toward justice, and for feminists who want to focus their efforts on the growing field of data science. But *Data Feminism* is about much more than gender. It is about power, about who has it and who doesn't, and about how those differentials of power can be challenged and changed. 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.

Equal parts mail art, data visualization, and affectionate correspondence, *Dear Data* celebrates "the infinitesimal, incomplete, imperfect, yet exquisitely human details of life," in the words of Maria Popova (Brain Pickings), who introduces this charming and graphically powerful book. For one year, Giorgia Lupi, an Italian living in New York, and Stefanie Posavec, an American in London, mapped the particulars of their daily lives as a series of hand-drawn postcards they exchanged via mail weekly—small portraits as full of emotion as they are data, both mundane and magical. *Dear Data* reproduces in pinpoint detail the full year's set of cards, front and back, providing a remarkable portrait of two artists connected by their attention to the details of their lives—including complaints, distractions, phone addictions, physical contact, and desires. These details illuminate the lives of two remarkable young women and also inspire us to map our own lives, including specific suggestions on what data to draw and how. A captivating and unique book for designers, artists, correspondents, friends, and lovers everywhere.

This book provides an introduction to the mathematical and algorithmic foundations of data science, including machine learning, high-dimensional geometry, and analysis of large networks. Topics include the counterintuitive nature of data in high dimensions, important linear algebraic techniques such as singular value decomposition, the theory of random walks and Markov chains, the fundamentals of and important algorithms for machine learning, algorithms and analysis for clustering, probabilistic models for large networks, representation learning including topic modelling and non-negative matrix factorization, wavelets and compressed sensing. Important probabilistic techniques are developed including the law of large numbers, tail inequalities, analysis of random projections, generalization guarantees in machine learning, and moment methods for analysis of phase transitions in large random graphs. Additionally, important structural and complexity measures are discussed such as matrix norms and VC-dimension. This book is suitable for both undergraduate and graduate courses in the design and analysis of algorithms for data.

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.

Data science libraries, frameworks, modules, and toolkits are great for doing data science, but they're also a good way to dive into the discipline without actually understanding data science. In this book, you'll learn how many of the most fundamental data science tools and algorithms work by implementing them from scratch. If you have an aptitude for mathematics and some programming skills, author Joel Grus will help you get comfortable with the math and statistics at the core of data science, and with hacking skills you need to get started as a data scientist. Today's messy glut of data holds answers to questions no one's even thought to ask. This book provides you with the know-how to dig those answers out. Get a crash course in Python Learn the basics of linear algebra, statistics, and probability—and understand how and when they're used in data science Collect, explore, clean, munge, and manipulate data Dive into the fundamentals of machine learning Implement models such as k-nearest Neighbors, Naive Bayes, linear and logistic regression, decision trees, neural networks, and clustering Explore recommender systems, natural language processing, network analysis, MapReduce, and databases SO MANY PEOPLE DREAM OF BECOMING THEIR OWN BOSS OR SUCCEEDING IN THEIR CHOSEN PROFESSION, AND WITH THE RESOURCES AVAILABLE TODAY, MORE ENTREPRENEURS AND PROFESSIONALS ARE ACHIEVING GREAT SUCCESS! HOWEVER, SUCCESS SHOULD BE DEFINED FOR THE LONG TERM, AND AS OPPORTUNITIES START TO GROW, SO DOES THE COMPETITION. Getting your business up and running or starting on your career path is one thing, but have a sustainable business or career is completely another. Many people make the mistake of making plans but having no follow-through. This is where analytics comes in. Don't you wish to have the power to know what your target consumers are thinking? Won't you want to have a preview of what future

trends to expect in the market you are in? Well, this book is just the one you need. This book will teach you, in simple and easy-to-understand terms, how to take advantage of data from your daily operations and make such data a powerful tool that can influence how well your business does over time.

The Data Science Handbook is a curated collection of 25 candid, honest and insightful interviews conducted with some of the world's top data scientists. In this book, you'll hear how the co-creator of the term 'data scientist' thinks about career and personal success. You'll hear from a young woman who created her own data scientist curriculum, subsequently landing her a role in the field. Readers of this book will be left with war stories, wisdom and

Used in Stanford's CS102 Big Data (Spring 2017) course. Want to get started on data science? Our promise: no math added. This book has been written in layman's terms as a gentle introduction to data science and its algorithms. Each algorithm has its own dedicated chapter that explains how it works, and shows an example of a real-world application. To help you grasp key concepts, we stick to intuitive explanations, as well as lots of visuals, all of which are colorblind-friendly. Popular concepts covered include: A/B Testing Anomaly Detection Association Rules Clustering Decision Trees and Random Forests Regression Analysis Social Network Analysis Neural Networks Features: Intuitive explanations and visuals Real-world applications to illustrate each algorithm Point summaries at the end of each chapter Reference sheets comparing the pros and cons of algorithms Glossary list of commonly-used terms With this book, we hope to give you a practical understanding of data science, so that you, too, can leverage its strengths in making better decisions.

Get complete instructions for manipulating, processing, cleaning, and crunching datasets in Python. Updated for Python 3.6, the second edition of this hands-on guide is packed with practical case studies that show you how to solve a broad set of data analysis problems effectively. You'll learn the latest versions of pandas, NumPy, IPython, and Jupyter in the process. Written by Wes McKinney, the creator of the Python pandas project, this book is a practical, modern introduction to data science tools in Python. It's ideal for analysts new to Python and for Python programmers new to data science and scientific computing. Data files and related material are available on GitHub. Use the IPython shell and Jupyter notebook for exploratory computing Learn basic and advanced features in NumPy (Numerical Python) Get started with data analysis tools in the pandas library Use flexible tools to load, clean, transform, merge, and reshape data Create informative visualizations with matplotlib Apply the pandas groupby facility to slice, dice, and summarize datasets Analyze and manipulate regular and irregular time series data Learn how to solve real-world data analysis problems with thorough, detailed examples

A comprehensive overview of data science covering the analytics, programming, and business skills necessary to master the discipline Finding a good data scientist has been likened to hunting for a unicorn: the required combination of technical skills is simply very hard to find in one person. In addition, good data science is not just rote application of trainable skill sets; it requires the ability to think flexibly about all these areas and understand the connections between them. This book provides a crash course in data science, combining all the necessary skills into a unified discipline. Unlike many analytics books, computer science and software engineering are given extensive coverage since they play such a central role in the daily work of a data scientist. The author also describes classic machine learning algorithms, from their mathematical foundations to real-world applications. Visualization tools are reviewed, and their central importance in data science is highlighted. Classical statistics is addressed to help readers think critically about the interpretation of data and its common pitfalls. The clear communication of technical results, which is perhaps the most undertrained of data science skills, is given its own chapter, and all topics are explained in the context of solving real-world data problems. The book also features:

- Extensive sample code and tutorials using Python™ along with its technical libraries
- Core technologies of “Big Data,” including their strengths and limitations and how they can be used to solve real-world problems
- Coverage of the practical realities of the tools, keeping theory to a minimum; however, when theory is presented, it is done in an intuitive way to encourage critical thinking and creativity
- A wide variety of case studies from industry
- Practical advice on the realities of being a data scientist today, including the overall workflow, where time is spent, the types of datasets worked on, and the skill sets needed

The Data Science Handbook is an ideal resource for data analysis methodology and big data software tools. The book is appropriate for people who want to practice data science, but lack the required skill sets. This includes software professionals who need to better understand analytics and statisticians who need to understand software. Modern data science is a unified discipline, and it is presented as such. This book is also an appropriate reference for researchers and entry-level graduate students who need to learn real-world analytics and expand their skill set. FIELD CADY is the data scientist at the Allen Institute for Artificial Intelligence, where he develops tools that use machine learning to mine scientific literature. He has also worked at Google and several Big Data startups. He has a BS in physics and math from Stanford University, and an MS in computer science from Carnegie Mellon.

The fundamental mathematical tools needed to understand machine learning include linear algebra, analytic geometry, matrix decompositions, vector calculus, optimization, probability and statistics. These topics are traditionally taught in disparate courses, making it hard for data science or computer science students, or professionals, to efficiently learn the mathematics. This self-contained textbook bridges the gap between mathematical and machine learning texts, introducing the mathematical concepts with a minimum of prerequisites. It uses these concepts to derive four central machine learning methods: linear regression, principal component analysis, Gaussian mixture models and support vector machines. For students and others with a mathematical background, these derivations provide a starting point to machine learning texts. For those learning the mathematics for the first time, the methods help build intuition and practical experience with applying mathematical concepts. Every chapter includes worked examples and exercises to test understanding. Programming tutorials are offered on the book's web site.

A concise introduction to the emerging field of data science, explaining its evolution, relation to machine learning, current uses, data infrastructure issues, and ethical challenges. The goal of data science is to improve decision making through the analysis of data. Today data science determines the ads we see online, the books and movies that are recommended to us online, which emails are filtered into our spam folders, and even how much we pay for health insurance. This volume in the MIT Press Essential Knowledge series offers a concise introduction to the emerging field of data science, explaining its evolution, current uses, data infrastructure issues, and ethical challenges. It has never been easier for organizations to gather, store, and process data. Use of data science is driven by the rise of big data and social media, the development of high-performance computing, and the emergence of such powerful methods for data analysis and modeling as deep learning. Data science encompasses a set of principles, problem definitions, algorithms, and processes for extracting non-obvious and useful patterns from large datasets. It is closely related to the fields of data mining and machine learning, but broader in scope. This book offers a brief history of the field, introduces fundamental data concepts, and describes the stages in a data science project. It considers data infrastructure and the challenges posed by integrating data from multiple sources, introduces the basics of machine learning, and discusses how to link machine learning expertise with real-world problems. The book also reviews ethical and legal issues, developments in data regulation, and computational approaches to preserving privacy. Finally, it considers the future impact of data science and offers principles for success in data science projects.

Cracking the Data Science Interview is the first book that attempts to capture the essence of data science in a concise, compact, and clean manner. In a Cracking the Coding Interview style, Cracking the Data Science Interview first introduces the relevant concepts, then presents a series of interview questions to help you solidify your understanding and prepare you for your next interview. Topics include: - Necessary Prerequisites (statistics, probability, linear algebra, and computer science) - 18 Big Ideas in Data Science (such as Occam's Razor, Overfitting, Bias/Variance Tradeoff, Cloud Computing, and Curse of Dimensionality) - Data Wrangling (exploratory data analysis, feature engineering, data cleaning and visualization) - Machine Learning Models (such as k-NN, random forests, boosting, neural networks, k-means clustering, PCA, and more) - Reinforcement Learning (Q-Learning and Deep Q-Learning) - Non-Machine Learning Tools (graph theory, ARIMA, linear programming) - Case Studies (a look at what data science means at companies like Amazon and Uber) Maverick holds a bachelor's degree from the College of Engineering at Cornell University in operations research and information engineering (ORIE) and a minor in computer science. He is the author of the popular Data Science Cheatsheet and Data Engineering Cheatsheet on GCP and has previous experience in data science consulting for a Fortune 500 company focusing on fraud analytics.

Publisher's Note: Products purchased from Third Party sellers are not guaranteed by the publisher for quality, authenticity, or access to any online entitlements included with the product. Use machine learning to understand your customers, frame decisions, and drive value The business analytics world has changed, and Data Scientists are taking over. Business Data Science takes you through the steps of using machine learning to implement best-in-class business data science. Whether you are a business leader with a desire to go deep on data, or an engineer who wants to learn how to apply Machine Learning to business problems, you'll find the information, insight, and tools you need to flourish in today's data-driven economy. You'll learn how to: •Use the key building blocks of Machine Learning: sparse regularization, out-of-sample validation, and latent factor and topic modeling•Understand how use ML tools in real world business problems, where causation matters more than correlation•Solve data science programs by scripting in the R programming language Today's business landscape is driven by data and constantly shifting. Companies live and die on their ability to make and implement the right decisions quickly and effectively. Business Data Science is about doing data science right. It's about the exciting things being done around Big Data to run a flourishing business. It's about the precepts, principals, and best practices that you need know for best-in-class business data science.

Statistical methods are a key part 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 This is a tutorial-driven and practical, but well-grounded book showcasing good Machine Learning practices. There will be an emphasis on using existing technologies instead of showing how to write your own implementations of algorithms. This book is a scenario-based, example-driven tutorial. By the end of the book you will have learnt critical aspects of Machine Learning Python projects and experienced the power of ML-based systems by actually working on them. This book primarily targets Python developers who want to learn about and build Machine Learning into their projects, or who want to pro.

"Data is here, it's growing, and it's powerful." Author Cathy O'Neil argues that the right approach to data is skeptical, not cynical—it understands that, while powerful, data science tools often fail. Data is nuanced, and "a really excellent skeptic puts the term 'science' into 'data science.'" The big data revolution shouldn't be dismissed as hype, but current data science tools and models shouldn't be hailed as the end-all-be-all, either.

Doing Data Science Straight Talk from the Frontline "O'Reilly Media, Inc."

Collecting data is relatively easy, but turning raw information into something useful requires that you know how to extract precisely what you need. With this insightful book, intermediate to experienced programmers interested in data analysis will learn techniques for working with data in a business environment. You'll learn how to look at data to discover what it contains, how to capture those ideas in conceptual models, and then feed your understanding back into the organization through business plans, metrics dashboards, and other applications. Along the way, you'll experiment with concepts through hands-on workshops at the end of each chapter. Above all, you'll learn how to think about the results you want to achieve -- rather than rely on tools to think for you. Use graphics to describe data with one, two, or dozens of variables Develop conceptual models using back-of-the-envelope calculations, as well as scaling and probability arguments Mine data with computationally intensive methods such as simulation and clustering Make your conclusions understandable through reports, dashboards, and other metrics programs Understand financial calculations, including the time-value of money Use dimensionality reduction techniques or predictive analytics to conquer challenging data analysis situations Become familiar with different open source programming environments for data analysis "Finally, a concise reference for understanding how to conquer piles of data."--Austin King, Senior Web Developer, Mozilla "An indispensable text for aspiring data scientists."--Michael E. Driscoll, CEO/Founder, Dataspora

"Turn yourself into a Data Head. You'll become a more valuable employee and make your organization more successful." Thomas H. Davenport, Research Fellow, Author of Competing on Analytics, Big Data @ Work, and The AI Advantage You've heard the hype around data—now get the facts. In Becoming a Data Head: How to Think, Speak, and Understand Data Science, Statistics, and Machine Learning, award-winning data scientists Alex Gutman and Jordan Goldmeier pull back the curtain on data science and give you the language and tools necessary to talk and think critically about it. You'll learn how to: Think statistically and understand the role variation plays in your life and decision making Speak intelligently and ask the right questions about the statistics and results you encounter in the workplace Understand what's really going on with machine learning, text analytics, deep learning, and artificial intelligence Avoid common pitfalls when working with and interpreting data Becoming a Data Head is a complete guide for data science in the workplace: covering everything from the personalities you'll work with to the math behind the algorithms. The authors have spent years in data trenches and sought to create a fun, approachable, and eminently readable book. Anyone can become a Data Head—an active participant in data science, statistics, and machine learning. Whether you're a business professional, engineer, executive, or aspiring data scientist, this book is for you.

Longlisted for the National Book Award New York Times Bestseller A former Wall Street quant sounds an alarm on the mathematical models that pervade modern life -- and threaten to rip apart our social fabric We live in the age of the algorithm. Increasingly, the decisions that affect our lives--where we go to school, whether we get a car loan, how much we pay for health insurance--are being made not by humans, but by mathematical models. In theory, this should lead to greater fairness: Everyone is judged according to the same rules, and bias is eliminated. But as Cathy O'Neil reveals in this urgent and necessary book, the opposite is true. The models being used today are opaque, unregulated, and uncontrollable, even when they're wrong. Most troubling, they reinforce discrimination: If a poor student can't get a loan because a lending model deems him too risky (by virtue of his zip code), he's then cut off from the kind of education that could pull him out of poverty, and a vicious spiral ensues. Models are propping up the lucky and punishing the downtrodden, creating a "toxic cocktail for democracy." Welcome to the dark side of Big Data. Tracing the arc of a person's life, O'Neil exposes the black box models that shape our future, both as individuals and as a society. These "weapons of math destruction" score teachers and students, sort resumes, grant (or deny) loans, evaluate workers, target voters, set parole, and monitor our health. O'Neil calls on modelers to take more responsibility for their algorithms and on policy makers to regulate their use. But in the end, it's up to us to become more savvy about the models that govern our lives. This important book empowers us to ask the tough questions, uncover the truth, and demand change. -- Longlist for National Book Award (Non-Fiction) -- Goodreads, semi-finalist for the 2016 Goodreads Choice Awards (Science and Technology) -- Kirkus, Best Books of 2016 -- New York Times, 100 Notable Books of 2016 (Non-Fiction) -- The Guardian, Best Books of 2016 -- WBUR's "On Point," Best Books of 2016: Staff Picks -- Boston Globe, Best Books of 2016, Non-Fiction

If you want to outsmart a crook, learn his tricks—Darrell Huff explains exactly how in the classic How to Lie with Statistics. From distorted graphs and biased samples to misleading averages, there are countless statistical dodges that lend cover to anyone with an ax to grind or a product to sell. With abundant examples and illustrations, Darrell Huff's lively and engaging primer clarifies the basic principles of statistics and explains how they're used to present information in honest and not-so-honest ways. Now even more indispensable in our data-driven world than it was when first published, How to Lie with Statistics is the book that generations of readers have relied on to keep from being fooled.

Why can't we think straight about the big issues that face our society? Why are we taken in by the phony arguments of populists and scammers? Where are the philosophers hiding when we need them to tell us what makes sense? They are hiding because they have nothing to say. The airy-fairy answers offered by writers of footnotes to Plato were wrong two thousand years ago, and they are still wrong now. All this time, we should have been listening to a different but equally venerable branch of matter-of-fact philosophy pioneered by the much-maligned philosopher Epicurus. His ideas were suppressed in ancient times as heretical, but the development of the theory of games and decisions makes it timely for those of us who care about science to revive his style of thinking—not just about the world around us but about ourselves as well. The price of transferring our

allegiance to Epicurus and his modern followers is that we can no longer enjoy the luxury of being told what we want to hear. It would be nice if we were really equipped with a hotline to a metaphysical world of transcendental ideals, but the truth is that we are just the flotsam left behind on the beach when the evolutionary tide went out, and we have to get real about what will and will not work for our imperfect species before it is too late. This book is an attempt to point the way. It has no equations and very little jargon; nor does it pull any punches, either in explaining how game theory works or in exposing the follies of famous metaphysicians.

This book is a complete introduction to the power of R for marketing research practitioners. The text describes statistical models from a conceptual point of view with a minimal amount of mathematics, presuming only an introductory knowledge of statistics. Hands-on chapters accelerate the learning curve by asking readers to interact with R from the beginning. Core topics include the R language, basic statistics, linear modeling, and data visualization, which is presented throughout as an integral part of analysis. Later chapters cover more advanced topics yet are intended to be approachable for all analysts. These sections examine logistic regression, customer segmentation, hierarchical linear modeling, market basket analysis, structural equation modeling, and conjoint analysis in R. The text uniquely presents Bayesian models with a minimally complex approach, demonstrating and explaining Bayesian methods alongside traditional analyses for analysis of variance, linear models, and metric and choice-based conjoint analysis. With its emphasis on data visualization, model assessment, and development of statistical intuition, this book provides guidance for any analyst looking to develop or improve skills in R for marketing applications.

Summary You are going to need more than technical knowledge to succeed as a data scientist. Build a Career in Data Science teaches you what school leaves out, from how to land your first job to the lifecycle of a data science project, and even how to become a manager. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the technology What are the keys to a data scientist's long-term success? Blending your technical know-how with the right "soft skills" turns out to be a central ingredient of a rewarding career. About the book Build a Career in Data Science is your guide to landing your first data science job and developing into a valued senior employee. By following clear and simple instructions, you'll learn to craft an amazing resume and ace your interviews. In this demanding, rapidly changing field, it can be challenging to keep projects on track, adapt to company needs, and manage tricky stakeholders. You'll love the insights on how to handle expectations, deal with failures, and plan your career path in the stories from seasoned data scientists included in the book. What's inside Creating a portfolio of data science projects Assessing and negotiating an offer Leaving gracefully and moving up the ladder Interviews with professional data scientists About the reader For readers who want to begin or advance a data science career. About the author Emily Robinson is a data scientist at Warby Parker. Jacqueline Nolis is a data science consultant and mentor. Table of Contents: PART 1 - GETTING STARTED WITH DATA SCIENCE 1. What is data science? 2. Data science companies 3. Getting the skills 4. Building a portfolio PART 2 - FINDING YOUR DATA SCIENCE JOB 5. The search: Identifying the right job for you 6. The application: Résumés and cover letters 7. The interview: What to expect and how to handle it 8. The offer: Knowing what to accept PART 3 - SETTLING INTO DATA SCIENCE 9. The first months on the job 10. Making an effective analysis 11. Deploying a model into production 12. Working with stakeholders PART 4 - GROWING IN YOUR DATA SCIENCE ROLE 13. When your data science project fails 14. Joining the data science community 15. Leaving your job gracefully 16. Moving up the ladder

An engaging introduction to data science that emphasizes critical thinking over statistical techniques An introduction to data science or statistics shouldn't involve proving complex theorems or memorizing obscure terms and formulas, but that is exactly what most introductory quantitative textbooks emphasize. In contrast, Thinking Clearly with Data focuses, first and foremost, on critical thinking and conceptual understanding in order to teach students how to be better consumers and analysts of the kinds of quantitative information and arguments that they will encounter throughout their lives. Among much else, the book teaches how to assess whether an observed relationship in data reflects a genuine relationship in the world and, if so, whether it is causal; how to make the most informative comparisons for answering questions; what questions to ask others who are making arguments using quantitative evidence; which statistics are particularly informative or misleading; how quantitative evidence should and shouldn't influence decision-making; and how to make better decisions by using moral values as well as data. Filled with real-world examples, the book shows how its thinking tools apply to problems in a wide variety of subjects, including elections, civil conflict, crime, terrorism, financial crises, health care, sports, music, and space travel. Above all else, Thinking Clearly with Data demonstrates why, despite the many benefits of our data-driven age, data can never be a substitute for thinking. An ideal textbook for introductory quantitative methods courses in data science, statistics, political science, economics, psychology, sociology, public policy, and other fields Introduces the basic toolkit of data analysis—including sampling, hypothesis testing, Bayesian inference, regression, experiments, instrumental variables, differences in differences, and regression discontinuity Uses real-world examples and data from a wide variety of subjects Includes practice questions and data exercises

From the creator of the popular website Ask a Manager and New York's work-advice columnist comes a witty, practical guide to 200 difficult professional conversations—featuring all-new advice! There's a reason Alison Green has been called "the Dear Abby of the work world." Ten years as a workplace-advice columnist have taught her that people avoid awkward conversations in the office because they simply don't know what to say. Thankfully, Green does—and in this incredibly helpful book, she tackles the tough discussions you may need to have during your career. You'll learn what to say when • coworkers push their work on you—then take credit for it • you accidentally trash-talk someone in an email then hit "reply all" • you're being micromanaged—or not being managed at all • you catch a colleague in a lie • your boss seems unhappy with your work • your cubemate's loud speakerphone is making you homicidal • you got drunk at the holiday party Praise for Ask a Manager "A must-read for anyone who works . . . [Alison Green's] advice boils down to the idea that you should be professional (even when others are not) and that communicating in a straightforward manner with candor and kindness will get you far, no matter where you work."—Booklist (starred review) "The author's friendly, warm, no-nonsense writing is a pleasure to read, and her advice can be widely applied to relationships in all areas of readers' lives. Ideal for anyone new to the job market or new to management, or anyone hoping to improve their work experience."—Library Journal (starred review) "I am a huge fan of Alison Green's Ask a Manager column. This book is even better. It teaches us how to deal with many of the most vexing big and little problems in our workplaces—and to do so with grace, confidence, and a sense of humor."—Robert Sutton, Stanford professor and author of The No Asshole Rule and The Asshole Survival Guide "Ask a Manager is the ultimate playbook for navigating the traditional workforce in a diplomatic but firm way."—Erin Lowry,

author of Broke Millennial: Stop Scraping By and Get Your Financial Life Together

The classic manifesto of the liberated woman, this book explores every facet of a woman's life.

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

JavaScript is the native language of the Internet. Originally created to make web pages more dynamic, it is now used for software projects of all kinds, including scientific visualization and data services. However, most data scientists have little or no experience with JavaScript, and most introductions to the language are written for people who want to build shopping carts rather than share maps of coral reefs. This book will introduce you to JavaScript's power and idiosyncrasies and guide you through the key features of the language and its tools and libraries. The book places equal focus on client- and server-side programming, and shows readers how to create interactive web content, build and test data services, and visualize data in the browser. Topics include: The core features of modern JavaScript Creating templated web pages Making those pages interactive using React Data visualization using Vega-Lite Using Data-Forge to wrangle tabular data Building a data service with Express Unit testing with Mocha All of the material is covered by the Creative Commons Attribution-Noncommercial 4.0 International license (CC-BY-NC-4.0) and is included in the book's companion website at <http://js4ds.org> . Maya Gans is a freelance data scientist and front-end developer by way of quantitative biology. Toby Hodges is a bioinformatician turned community coordinator who works at the European Molecular Biology Laboratory. Greg Wilson co-founded Software Carpentry, and is now part of the education team at RStudio

"This book describes the process of analyzing data. The authors have extensive experience both managing data analysts and conducting their own data analyses, and this book is a distillation of their experience in a format that is applicable to both practitioners and managers in data science."--Leanpub.com.

[Copyright: 500e7f7b1f429b394843cc1a3f49e08f](https://creativecommons.org/licenses/by-nc/4.0/)