Exploratory Data Analysis Tukey

The approach in this introductory book is that of informal study of the data. Methods range from plotting picture-drawing techniques to rather elaborate numerical summaries. Several of the methods are the original creations of the author, and all can be carried out either with pencil or aided by hand-held calculator. Exploratory Data Analysis Using R provides a classroom-tested introduction to exploratory data analysis (EDA) and introduces the range of "interesting" – good, bad, and ugly – features that can be found in data, and why it is important to find them. It also introduces the mechanics of using R to explore and explain data. The book begins with a detailed overview of data, exploratory analysis, and R, as well as graphics in R. It then explores working with external data, linear regression models, and crafting data stories. The second part of the book focuses on developing R programs, including good programming practices and examples, working with text data, and general predictive models. The book ends with a chapter on "keeping it all together" that includes managing the R installation, managing files, documenting, and an introduction to reproducible computing. The book is designed for both advanced undergraduate, entry-level graduate students, and working professionals with little to no prior exposure to data analysis, modeling, statistics, or programming. it keeps the treatment relatively non-mathematical, even though data analysis is an inherently mathematical subject. Exercises are included at the end of most

chapters, and an instructor's solution manual is available. About the Author: Ronald K. Pearson holds the position of Senior Data Scientist with GeoVera, a property insurance company in Fairfield, California, and he has previously held similar positions in a variety of application areas, including software development, drug safety data analysis, and the analysis of industrial process data. He holds a PhD in Electrical Engineering and Computer Science from the Massachusetts Institute of Technology and has published conference and journal papers on topics ranging from nonlinear dynamic model structure selection to the problems of disguised missing data in predictive modeling. Dr. Pearson has authored or co-authored books including Exploring Data in Engineering, the Sciences, and Medicine (Oxford University Press, 2011) and Nonlinear Digital Filtering with Python. He is also the developer of the DataCamp course on base R graphics and is an author of the datarobot and GoodmanKruskal R packages available from CRAN (the Comprehensive R Archive Network). eBook Version You will receive access to this electronic text via email after using the shopping cart above to complete your purchase.

This collection of essays brings together many of the world's most distinguished statisticians to discuss a wide array of the most important recent developments in data analysis. The book honors John W. Tukey, one of the most influential statisticians of the twentieth century, on the occasion of his eightieth birthday. Contributors, some of them Tukey's former students, use his general theoretical work and his specific contributions to

Exploratory Data Analysis as the point of departure for their papers. They cover topics from "pure" data analysis, such as gaussianizing transformations and regression estimates, and from "applied" subjects, such as the best way to rank the abilities of chess players or to estimate the abundance of birds in a particular area. Tukey may be best known for coining the common computer term "bit," for binary digit, but his broader work has revolutionized the way statisticians think about and analyze sets of data. In a personal interview that opens the book, he reviews these extraordinary contributions and his life with characteristic modesty, humor, and intelligence. The book will be valuable both to researchers and students interested in current theoretical and practical data analysis and as a testament to Tukey's lasting influence. The essays are by Dhammika Amaratunga, David Andrews, David Brillinger, Christopher Field, Leo Goodman, Frank Hampel, John Hartigan, Peter Huber, Mia Hubert, Clifford Hurvich, Karen Kafadar, Colin Mallows, Stephan Morgenthaler, Frederick Mosteller, Ha Nguyen, Elvezio Ronchetti, Peter Rousseeuw, Allan Seheult, Paul Velleman, Maria-Pia Victoria-Feser, and Alessandro Villa. Originally published in 1998. The Princeton Legacy Library uses the latest print-on-demand technology to again make available previously out-of-print books from the distinguished backlist of Princeton University Press. These editions preserve the original texts of these important books while presenting them in durable paperback and hardcover editions. The goal of the Princeton Legacy Library is to vastly increase access to the rich scholarly heritage

found in the thousands of books published by Princeton
University Press since its founding in 1905.
Master scientific computing and perform complex
operations with ease
The Collected Works of John W. Tukey
Principles and Examples
Modern Statistics for Modern Biology
Proceedings of a Forum
Practical Statistics for Data Scientists
Meet John W. Tukey, one of the most consequential statisticians
and original thinkers of the twentieth century. Growing up one
hundred years ago in New Bedford, Massachusetts, a large coastal
town primarily known for its commercial fishing and textile
industries, John Wilder Tukey quickly showed himself to be a child
prodigy. The son of educated parents whose high school classmates
voted them most likely to give birth to a genius, he learned to read
on his own by three years of age, mastered using a hand-crack desk
calculator to speed up arithmetical calculations shortly thereafter,
and was poring through technical journals in the New Bedford Free
Public Library by the time he was a teenager. Homeschooled until
being admitted to Brown University, Tukey majored in chemistry
there--even as he spent countless hours in the university library
compiling lists of statistical techniques on index cards, simply
because he found them interesting and useful. With multiple degrees
in hand, Tukey's next stop was Princeton University, where his
interests shifted to mathematics. After earning a doctorate in
topology, an especially abstract branch of mathematics, Princeton
retained him as a lecturer. But with the United States poised to
enter World War II, Tukey joined the Fire Control Research Office
(FCRO), where he was exposed to a set of life-and-death problems
that bore little resemblance to abstract mathematics: namely,
calculating the trajectories of artillery and ballistics and the motions
of rocket powder, working with stereoscopic height and range

finders, and improving the Boeing B-29 Superfortress bomber. With the stakes never higher, a chance encounter during the war with a fellow polymath and unconventional thinker twenty years his senior set the course for the rest of Tukey's professional life--as well as changing the field of statistics forever. In "Adventures of a Statistician," author Mark Jones Lorenzo chronicles John Tukey's life and times, from his decades spent at Princeton as a teacher and administrator and also at AT&T's Bell Laboratories as a scientific generalist; to his development of the fast Fourier transform (FFT) algorithm, which launched a revolution in digital signal processing; to his innovative ideas in displaying and summarizing data, such as with the intuitive stem-and-leaf plot and the interactive graphics of the PRIM-9 computer system; to his creation of exploratory data analysis, an approach to performing statistics he equated with "detective work"; to his intellectual war with sex researcher Alfred Kinsey over appropriate kinds of statistical sampling; to his productive yet sometimes strained relationships with fellow statisticians such as Ronald Fisher, George Box, and Erich Lehmann; to his enlightening friendship with the legendary physicist Richard Feynman; to his mentoring of dozens of doctoral students, many of whom went on to have highly successful careers in their own right; to his inventive use of language, having coined words like "bit"; to his development of sophisticated mathematical methods to detect underground nuclear explosions; to his groundbreaking work on the jackknife, multiple comparisons, robustness, and many other statistical techniques; and to his accomplishments in health and environmental regulation, U.S. census analysis, election forecasting, and public policy, among a host of other significant and impactful achievements. Nearly a decade in the making, "Adventures of a Statistician" is more than just the complete biography of John W. Tukey, perhaps the most revolutionary applied statistician of the past century. It's also a fascinating intellectual journey through the recent history of statistics as well.

Modern Data Analysis contains the proceedings of a Workshop on

Modern Data Analysis held in Raleigh, North Carolina, on June 2-4, 1980 under the auspices of the United States Army Research Office. The papers review theories and methods of data analysis and cover topics ranging from single and multiple quantile-quantile (Q-Q) plotting procedures to biplot display and pencil-and-paper exploratory data analysis methods. Projection pursuit methods for data analysis are also discussed. Comprised of nine chapters, this book begins with an introduction to styles of data analysis techniques, followed by an analysis of single and multiple Q-Q plotting procedures. Problems involving extreme-value data and the behavior of sample averages are considered. Subsequent chapters deal with the use of smelting in guiding re-expression; geometric data analysis; and influence functions and regression diagnostics. The final chapter examines the use and interpretation of robust analysis of variance for the general non-full-rank linear model. The procedures are described in terms of their mathematical structure, which leads to efficient computational algorithms. This monograph should be of interest to mathematicians and statisticians. This volume of eleven articles compiles important papers by Tukey that examine the intriguing problems inherent in the area of multiple comparisons and provide a useful framework for thinking about them. Each volume in the set is indexed and contains a bibliography. Edited by Guido Deboeck, a leading exponent in the use of computation intelligence methods in finance and economic forecasting, and the originator of SOM, Teuvo Kohonen. An 8-page color section makes this book unique, colorful and exciting to read. Each chapter contains exercises and solutions, perfectly suited to aid self-study. Statistics and Data Analysis for Financial Engineering The Practice of Data Analysis with R examples R for Data Science Mastering Numerical Computing with NumPy

The Biography of John W. Tukey

The new edition of this influential textbook, geared towards graduate or advanced undergraduate students, teaches the statistics necessary for financial engineering. In doing so, it illustrates concepts using financial markets and economic data, R Labs with real-data exercises, and graphical and analytic methods for modeling and diagnosing modeling errors. These methods are critical because financial engineers now have access to enormous quantities of data. To make use of this data, the powerful methods in this book for working with quantitative information, particularly about volatility and risks, are essential. Strengths of this fully-revised edition include major additions to the R code and the advanced topics covered. Individual chapters cover, among other topics, multivariate distributions, copulas, Bayesian computations, risk management, and cointegration. Suggested prerequisites are basic knowledge of statistics and probability, matrices and linear algebra, and calculus. There is an appendix on probability, statistics and linear algebra. Practicing financial engineers will also find this book of interest. Praise for the Second Edition: "The authors present an intuitive and easy-to-read book. ... accompanied by many examples, proposed exercises, good references, and comprehensive appendices that initiate the reader unfamiliar with MATLAB." —Adolfo Alvarez Pinto, International Statistical Review "Practitioners of EDA who use MATLAB will want a copy of this book. ... The authors have done a great service by bringing together so many EDA routines, but their main accomplishment in this dynamic text is providing the understanding and tools to do EDA. —David A Huckaby, MAA Reviews Exploratory Data Analysis (EDA) is an important part of

the data analysis process. The methods presented in this text are ones that should be in the toolkit of every data scientist. As computational sophistication has increased and data sets have grown in size and complexity, EDA has become an even more important process for visualizing and summarizing data before making assumptions to generate hypotheses and models. Exploratory Data Analysis with MATLAB, Third Edition presents EDA methods from a computational perspective and uses numerous examples and applications to show how the methods are used in practice. The authors use MATLAB code, pseudo-code, and algorithm descriptions to illustrate the concepts. The MATLAB code for examples, data sets, and the EDA Toolbox are available for download on the book's website. New to the Third Edition Random projections and estimating local intrinsic dimensionality Deep learning autoencoders and stochastic neighbor embedding Minimum spanning tree and additional cluster validity indices Kernel density estimation Plots for visualizing data distributions, such as beanplots and violin plots A chapter on visualizing categorical data
The Concise Encyclopedia of Statistics presents the essential information about statistical tests, concepts, and analytical methods in language that is accessible to practitioners and students of the vast community using statistics in medicine, engineering, physical science, life science, social science, and business/economics. The reference is alphabetically arranged to provide quick access to the fundamental tools of statistical methodology and biographies of famous statisticians. The more than 500 entries include definitions, history, mathematical details, limitations, examples, references, and further readings. All entries include cross-references

as well as the key citations. The back matter includes a timeline of statistical inventions. This reference will be an enduring resource for locating convenient overviews about this essential field of study.

WILEY-INTERSCIENCE PAPERBACK SERIES The Wiley-Interscience Paperback Series consists ofselected books that have been made more accessible to consumers inan effort to increase global appeal and general circulation. Withthese new unabridged softcover volumes, Wiley hopes to extend thelives of these works by making them available to future generationsof statisticians, mathematicians, and scientists. "Exploring Data Tables, Trends, and Shapes (EDTTS) waswritten as a companion volume to the same editors' book,Understanding Robust and Exploratory Data Analysis (UREDA). WhereasUREDA is a collection of exploratory and resistant methods ofestimation and display, EDTTS goes a step further, describingmultivariate and more complicated techniques . . . I feel that theauthors have made a very significant contribution in the area ofmultivariate nonparametric methods. This book [is] a valuablesource of reference to researchers in the area." —Technometrics "This edited volume . . . provides an important theoretical andphilosophical extension to the currently popular statistical areaof Exploratory Data Analysis, which seeks to reveal structure, orsimple descriptions, in data . . . It is . . . an importantreference volume which any statistical library should considerseriously." —The Statistician This newly available and affordably priced paperback version ofExploring Data Tables, Trends, and Shapes presents majoradvances in exploratory data analysis and robust regression methodsand explains the techniques, relating them to classical methods.The book addresses the role of exploratory and robust techniques inthe overall data-

analytic enterprise, and it also presents newmethods such as fitting by organized comparisons using the squarecombining table and identifying extreme cells in a sizablecontingency table with probabilistic and exploratory approaches.The book features a chapter on using robust regression in lesstechnical language than available elsewhere. Conceptual support foreach technique is also provided.
Understanding Statistics and Probability with Star Wars, LEGO, and Rubber Ducks
With R and GGobi
Exploratory Data Analysis, by John W. Tukey
Exploring Data Tables, Trends, and Shapes
Interactive Graphics for Data Analysis
Modern Data Analysis

Data analytics may seem daunting, but if you're an experienced Excel user, you have a unique head start. With this hands-on guide, intermediate Excel users will gain a solid understanding of analytics and the data stack. By the time you complete this book, you'll be able to conduct exploratory data analysis and hypothesis testing using a programming language. Exploring and testing relationships are core to analytics. By using the tools and frameworks in this book, you'll be well positioned to continue learning more advanced data analysis techniques. Author George Mount, founder and CEO of Stringfest Analytics, demonstrates key statistical concepts with spreadsheets,

then pivots your existing knowledge about data manipulation into R and Python programming. This practical book guides you through: Foundations of analytics in Excel: Use Excel to test relationships between variables and build compelling demonstrations of important concepts in statistics and analytics From Excel to R: Cleanly transfer what you've learned about working with data from Excel to R From Excel to Python: Learn how to pivot your Excel data chops into Python and conduct a complete data analysis
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.*
"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.

Fun guide to learning Bayesian statistics and probability through unusual and illustrative examples. Probability and statistics are increasingly important in a huge range of professions. But many people use data in ways they don't even understand, meaning they aren't getting the most from it. Bayesian Statistics the Fun Way will change that. This book will give you a complete understanding of Bayesian statistics through simple explanations and un-boring examples. Find out the probability of UFOs landing in your garden, how likely Han Solo is to survive a flight through an asteroid shower, how to win an argument about conspiracy theories, and whether a burglary really was a burglary, to name a few examples. By using these off-the-beaten-track examples, the author actually makes learning statistics fun. And you'll learn real skills, like how to: - How to measure your own level of uncertainty in a conclusion or belief - Calculate Bayes

*theorem and understand what it's useful
for - Find the posterior, likelihood, and
prior to check the accuracy of your
conclusions - Calculate distributions to
see the range of your data - Compare
hypotheses and draw reliable conclusions
from them Next time you find yourself with
a sheaf of survey results and no idea what
to do with them, turn to Bayesian
Statistics the Fun Way to get the most
value from your data.*

The Art of Data Science

*Understanding Robust and Exploratory Data
Analysis*

Applications of Exploratory Data Analysis

*Harness Oil and Gas Big Data with
Analytics*

A Primer for Undergraduates

*Convergence and Uniformity in Topology.
(AM-2), Volume 2*

**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 This book trains the next generation of scientists representing different disciplines to leverage the data generated during routine patient care. It formulates a more complete lexicon of evidence-based recommendations and support shared, ethical decision making by doctors with their patients. Diagnostic and therapeutic technologies continue to evolve rapidly, and both individual practitioners and clinical teams face increasingly complex

ethical decisions. Unfortunately, the current state of medical knowledge does not provide the guidance to make the majority of clinical decisions on the basis of evidence. The present research infrastructure is inefficient and frequently produces unreliable results that cannot be replicated. Even randomized controlled trials (RCTs), the traditional gold standards of the research reliability hierarchy, are not without limitations. They can be costly, labor intensive, and slow, and can return results that are seldom generalizable to every patient population. Furthermore, many pertinent but unresolved clinical and medical systems issues do not seem to have attracted the interest of the research enterprise, which has come to focus instead on cellular and molecular investigations and single-agent (e.g., a drug or device) effects. For clinicians, the end result is a bit of a "data desert" when it comes to making decisions. The new research infrastructure proposed in this book will help the medical profession to make ethically sound and well informed decisions for their patients.
A far-reaching course in practical advanced statistics for biologists using R/Bioconductor, data exploration, and

simulation.
This book presents guidelines for the development and evaluation of statistical software designed to ensure minimum acceptable statistical functionality as well as ease of interpretation and use. It consists of the proceedings of a forum that focused on three qualities of statistical software: richness--the availability of layers of output sophistication, guidance--how the package helps a user do an analysis and do it well, and exactness--determining if the output is "correct" and when and how to warn of potential problems.
with Self-Organizing Maps
Handbook of Psychology, Research Methods in Psychology
Interactive and Dynamic Graphics for Data Analysis
The Future of Statistical Software
Visual Explorations in Finance
Adventures of a Statistician

This book serves as an introductory text for exploratory data analysis. It exposes readers and users to a variety of techniques for looking more effectively at data. The emphasis is on general techniques, rather than specific problems.

Stem-and-left displays; Letter-value displays; Boxplots; x-y plotting; Resistant line; Smoothing data; Coded tables; Median polish; Rootograms; Computer graphics; Utility programs; Programming conventions; Minitab implementation; Appendices; Index.

The description for this book, Convergence and Uniformity in

Topology. (AM-2), Volume 2, will be forthcoming.
Originally published: Reading, Mass.: Addison-Wesley Publishing
Company, 1977.
Data Analysis and Prediction Algorithms with R
Exploratory Data Analysis
Exploratory Data Analysis Using R
A Second Course in Statistics
Secondary Analysis of Electronic Health Records
The Concise Encyclopedia of Statistics
Age, Time, and Fertility: Applications of Exploratory Data
Analysis describes change in the age pattern of fertility
that responds to a specific need in making fertility
comparisons across time and place. This book discusses
a modeling process based on Tukey's exploratory data
analysis (EDA) methods, which is proved very effective
in other fields for detecting underlying patterns, even in
flawed data. The first part of this text provides an
introduction to the philosophy and tools of EDA and to
the data analyzed, examining in detail the process of
developing and standardizing the closely fitting, few-
parameter descriptions of demographic change in time
sequence. The rest of the chapters examine the results
and applications of fertility modeling and establish
relations between change in the age pattern of fertility
and level of fertility. This publication is intended for those
interested in the measures and methods of fertility
change that can be applied to demographic data.
An introduction to the underlying principles, central
concepts, and basic techniques for conducting and
understanding exploratory data analysis -- with
numerous social science examples.
Enhance the power of NumPy and start boosting your

scientific computing capabilities Key Features Grasp all aspects of numerical computing and understand NumPy Explore examples to learn exploratory data analysis (EDA), regression, and clustering Access NumPy libraries and use performance benchmarking to select the right tool Book Description NumPy is one of the most important scientific computing libraries available for Python. Mastering Numerical Computing with NumPy teaches you how to achieve expert level competency to perform complex operations, with in-depth coverage of advanced concepts. Beginning with NumPy's arrays and functions, you will familiarize yourself with linear algebra concepts to perform vector and matrix math operations. You will thoroughly understand and practice data processing, exploratory data analysis (EDA), and predictive modeling. You will then move on to working on practical examples which will teach you how to use NumPy statistics in order to explore US housing data and develop a predictive model using simple and multiple linear regression techniques. Once you have got to grips with the basics, you will explore unsupervised learning and clustering algorithms, followed by understanding how to write better NumPy code while keeping advanced considerations in mind. The book also demonstrates the use of different high-performance numerical computing libraries and their relationship with NumPy. You will study how to benchmark the performance of different configurations and choose the best for your system. By the end of this book, you will have become an expert in handling and performing complex data manipulations. What you will learn Perform

vector and matrix operations using NumPy Perform exploratory data analysis (EDA) on US housing data Develop a predictive model using simple and multiple linear regression Understand unsupervised learning and clustering algorithms with practical use cases Write better NumPy code and implement the algorithms from scratch Perform benchmark tests to choose the best configuration for your system Who this book is for Mastering Numerical Computing with NumPy is for you if you are a Python programmer, data analyst, data engineer, or a data science enthusiast, who wants to master the intricacies of NumPy and build solutions for your numeric and scientific computational problems. You are expected to have familiarity with mathematics to get the most out of this book.

This book is about using interactive and dynamic plots on a computer screen as part of data exploration and modeling, both alone and as a partner with static graphics and non-graphical computational methods. The area of int- active and dynamic data visualization emerged within statistics as part of research on exploratory data analysis in the late 1960s, and it remains an active subject of research today, as its use in practice continues to grow. It now makes substantial contributions within computer science as well, as part of the growing ?elds of information visualization and data mining, especially visual data mining. The material in this book includes: • An introduction to data visualization, explaining how it di?ers from other types of visualization. • Adescriptionofourtoolboxofinteractiveanddynamicgraphi calmethods. • An approach for exploring missing values

in data. • An explanation of the use of these tools in
cluster analysis and supervised classi?cation. • An
overview of additional material available on the web. • A
description of the data used in the analyses and
exercises. The book′s examples use the software R and
GGobi. R (Ihaka & Gent-man 1996, RDevelopment
CoreTeam2006) isafreesoftware environment for
statistical computing and graphics; it is most often used
from the command line, provides a wide variety of
statistical methods, and includes high–quality staticgraph
ics.RaroseintheStatisticsDepartmentoftheUniversityofAu-
land and is now developed and maintained by a global
collaborative e?ort.
Visualizing Data
Bayesian Statistics the Fun Way
Fundamentals of Exploratory Analysis of Variance
50 Essential Concepts
Age, Time, and Fertility
Advancing Into Analytics
The analysis of variance is presented as an
exploratory component of data analysis, while
retaining the customary least squares fitting
methods. Balanced data layouts are used to
reveal key ideas and techniques for
exploration. The approach emphasizes both the
individual observations and the separate
parts that the analysis produces. Most
chapters include exercises and the appendices
give selected percentage points of the
Gaussian, t, F chi-squared and studentized
range distributions.
Approaching data analysis; Indication and

indicators; Displays and summaries for
batches; Straightening curves and plots; The
practice of re-expression; Need we re-
express? Hunting out the real uncertainty; A
method of direct assessment; Two-and more-way
tables; Robust and resistant measures;
Standardizing for comparison; Regression for
fitting; Woes of regression coefficients; A
class of mechanisms for fitting; Guided
regression; Examining regression residuals.
Interactive Graphics for Data Analysis:
Principles and Examples discusses exploratory
data analysis (EDA) and how interactive
graphical methods can help gain insights as
well as generate new questions and hypotheses
from datasets.Fundamentals of Interactive
Statistical GraphicsThe first part of the
book summarizes principles and methodology,
demons
Includes established theories and cutting-
edge developments. Presents the work of an
international group of experts. Presents the
nature, origin, implications, an future
course of major unresolved issues in the
area.
Data Analysis and Regression
Import, Tidy, Transform, Visualize, and Model
Data
Essays in Honor of John W. Tukey
Introduction to Data Science
Philosophy and Principles of Data Analysis
1965-1986
Optimize Exploration and Production with Data-
Driven Models

"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"--

Use big data analytics to efficiently drive oil and gas exploration and production Harness Oil and Gas Big Data with Analytics provides a complete view of big data and analytics techniques as they are applied to the oil and gas industry. Including a compendium of specific case studies, the book underscores the acute need for optimization in the oil and gas exploration and production stages and shows how data analytics can provide such optimization. This spans exploration, development, production and rejuvenation of oil and gas assets. The book serves as a guide for fully leveraging data, statistical, and quantitative analysis, exploratory and predictive modeling, and fact-based management to drive decision making in oil and gas operations. This comprehensive resource delves into the three major issues that face the oil and gas industry during the exploration and production stages: Data management, including storing massive quantities of data in a manner conducive to analysis and effectively retrieving, backing up, and purging data Quantification of uncertainty, including a look at the statistical and data analytics methods for making predictions and determining the certainty of those predictions Risk assessment, including predictive analysis of the likelihood that known risks are realized and how to properly deal with unknown risks Covering the major issues facing the oil and gas industry in the exploration and production stages, Harness Big Data with Analytics reveals how to model big data to realize efficiencies and business benefits.

The Wiley Classics Library consists of selected books that have become recognized classics in their respective fields. With these new unabridged and inexpensiveeditions, Wiley hopes to extend the life of these important works by making themavailable to future generations of mathematicians and scientists. Currently available in the Series: T.W. Anderson The Statistical Analysis of Time Series T.S. Arthanari & Yadolah Dodge Mathematical Programming in Statistics Emil Artin Geometric Algebra Norman T. J. Bailey The Elements of Stochastic Processes with Applications to the Natural Sciences Robert G. Bartle The Elements of Integration and Lebesgue Measure George E. P. Box & Norman R. Draper Evolutionary Operation: A Statistical Method for Process Improvement George E. P. Box & George C. Tiao Bayesian Inference in Statistical Analysis R. W. Carter Finite Groups of Lie Type: Conjugacy Classes and Complex Characters R. W. Carter Simple Groups of Lie Type William G. Cochran & Gertrude M. Cox Experimental Designs, Second Edition Richard Courant Differential and Integral Calculus, Volume I Richard Courant Differential and Integral Calculus, Volume II Richard Courant & D. Hilbert Methods of Mathematical Physics, Volume I Richard Courant & D. Hilbert Methods of Mathematical Physics, Volume II D. R. Cox Planning of Experiments Harold S. M. Coxeter Introduction to Geometry, Second Edition Charles W. Curtis & Irving Reiner Representation Theory of Finite Groups and Associative Algebras Charles W. Curtis & Irving Reiner Methods of Representation Theory with Applications to Finite Groups and Orders, Volume I Charles W. Curtis & Irving Reiner Methods of Representation Theory with Applications to Finite Groups and Orders, Volume II Cuthbert Daniel Fitting Equations to Data: Computer

Analysis of Multifactor Data, Second Edition Bruno de Finetti Theory of Probability, Volume I Bruno de Finetti Theory of Probability, Volume 2 W. Edwards Deming Sample Design in Business Research Amos de Shalit & Herman Feshbach Theoretical Nuclear Physics, Volume 1— Nuclear Structure Harold F. Dodge & Harry G. Romig Sampling Inspection Tables: Single and Double Sampling J. L. Doob Stochastic Processes Nelson Dunford & Jacob T. Schwartz Linear Operators, Part One, General Theory Nelson Dunford & Jacob T. Schwartz Linear Operators, Part Two, Spectral Theory—Self Adjoint Operators in Hilbert Space Nelson Dunford & Jacob T. Schwartz Linear Operators, Part Three, Spectral Operators Regina C. Elandt-Johnson & Norman L. Johnson Survival Models and Data Analysis Herman Feshbach Theoretical Nuclear Physics: Nuclear Reactions Joseph L. Fleiss Design and Analysis of Clinical Experiments Bernard Friedman Lectures on Applications-Oriented Mathematics Phillip Griffiths & Joseph Harris Principles of Algebraic Geometry Gerald J. Hahn & Samuel S. Shapiro Statistical Models in Engineering Marshall Hall, Jr. Combinatorial Theory, Second Edition Morris H. Hansen, William N. Hurwitz & William G. Madow Sample Survey Methods and Theory, Volume I—Methods and Applications Morris H. Hansen, William N. Hurwitz & William G. Madow Sample Survey Methods and Theory, Volume II—Theory Peter Henrici Applied and Computational Complex Analysis, Volume 1—Power Series—Integration—Conformal Mapping—Location of Zeros Peter Henrici Applied and Computational Complex Analysis, Volume 2—Special Functions—Integral Transforms—Asymptotics—Continued Fractions Peter Henrici Applied and Computational Complex Analysis, Volume 3—Discrete Fourier Analysis—Cauchy Integrals—Construction of Conformal

Maps—Univalent Functions Peter Hilton & Yel-Chiang Wu A Course in Modern Algebra David C. Hoaglin, Frederick Mosteller & John W. Tukey Understanding Robust and Exploratory Data Analysis Harry Hochstadt Integral Equations Leslie Kish Survey Sampling Shoshichi Kobayashi & Katsumi Nomizu Foundations of Differential Geometry, Volume I Shoshichi Kobayashi & Katsumi Nomizu Foundations of Differential Geometry, Volume 2 Erwin O. Kreyszig Introductory Functional Analysis with Applications William H. Louisell Quantum Statistical Properties of Radiation Rupert G. Miller Jr. Survival Analysis Ali Hasan Nayfeh Introduction to Perturbation Techniques Ali Hasan Nayfeh & Dean T. Mook Nonlinear Oscillations Emanuel Parzen Modern Probability Theory & Its Applications P. M. Prenter Splines and Variational Methods Howard Raiffa & Robert Schlaifer Applied Statistical Decision Theory Walter Rudin Fourier Analysis on Groups Lawrence S. Schulman Techniques and Applications of Path Integration Shayle R. Searle Linear Models I. H. Segel Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems C. L. Siegel Topics in Complex Function Theory, Volume I—Elliptic Functions and Uniformization Theory C. L. Siegel Topics in Complex Function Theory, Volume II—Automorphic and Abelian Integrals C. L. Siegel Topics in Complex Function Theory, Volume III—Abelian Functions and Modular Functions of Several Variables L. Spitzer Physical Processes in the Interstellar Medium J. J. Stoker Differential Geometry J. J. Stoker Water Waves: The Mathematical Theory with Applications J. J. Stoker Nonlinear Vibrations in Mechanical and ElectricalSystems Richard Zallen The Physics of Amorphous Solids Arnold Zellner Introduction to Bayesian Inference in Econometrics

Applications, Basics, and Computing of Exploratory Data
Analysis
Exploratory Data Analysis with MATLAB
Exploratory Data Analysis (Classic Version)
Encyclopedia of Database Systems