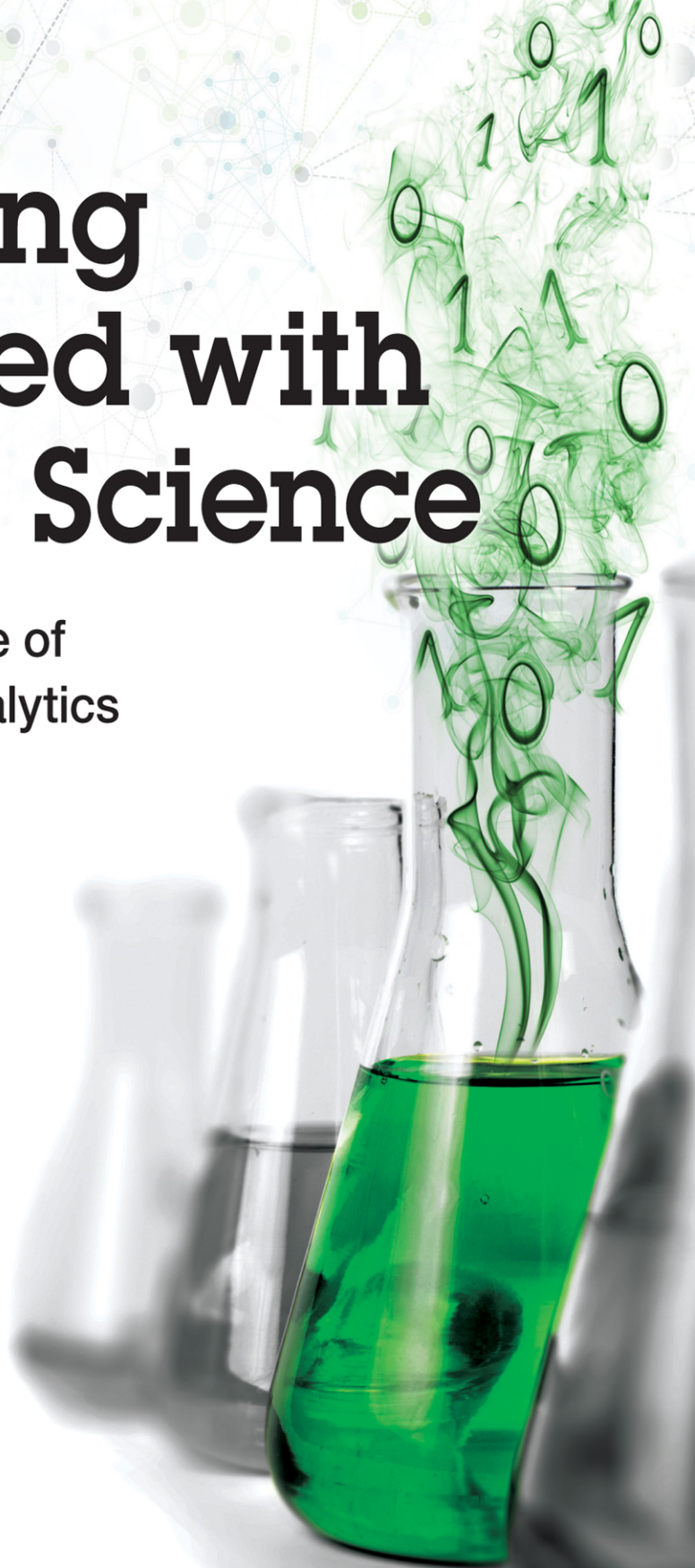


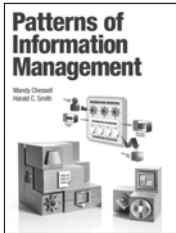
Getting Started with Data Science

Making Sense of
Data with Analytics

Murtaza Haider



Related Books of Interest



Patterns of Information Management

By Mandy Chessell and Harald C. Smith
ISBN: 978-0-13-315550-1

Use Best Practice Patterns to Understand and Architect Manageable, Efficient Information Supply Chains That Help You Leverage All Your Data and Knowledge

Building on the analogy of a supply chain, Mandy Chessell and Harald Smith explain how information can be transformed, enriched, reconciled, redistributed, and utilized in even the most complex environments. Through a realistic, end-to-end case study, they help you blend overlapping information management, SOA, and BPM technologies that are often viewed as competitive.

Using this book's patterns, you can integrate all levels of your architecture—from holistic, enterprise, system-level views down to low-level design elements. You can fully address key non-functional requirements such as the amount, quality, and pace of incoming data. Above all, you can create an IT landscape that is coherent, interconnected, efficient, effective, and manageable.



The Business of IT How to Improve Service and Lower Costs

By Robert Ryan and Tim Raducha-Grace
ISBN: 978-0-13-700061-6

Drive More Business Value from IT...and Bridge the Gap Between IT and Business Leadership

IT organizations have achieved outstanding technological maturity, but many have been slower to adopt world-class business practices. This book provides IT and business executives with methods to achieve greater business discipline throughout IT, collaborate more effectively, sharpen focus on the customer, and drive greater value from IT investment. Drawing on their experience consulting with leading IT organizations, Robert Ryan and Tim Raducha-Grace help IT leaders make sense of alternative ways to improve IT service and lower cost, including ITIL, IT financial management, balanced scorecards, and business cases. You'll learn how to choose the best approaches to improve IT business practices for your environment and use these practices to improve service quality, reduce costs, and drive top-line revenue growth.

Getting Started with Data Science: Making Sense of Data with Analytics

Table of Contents

Contents

Preface

Chapter 1 The Bazaar of Storytellers

Data Science: The Sexiest Job in the 21st Century

Storytelling at Google and Walmart

Getting Started with Data Science

Do We Need Another Book on Analytics?

Repeat, Repeat, Repeat, and Simplify

Chapters Structure and Features

Analytics Software Used

What Makes Someone a Data Scientist?

Existential Angst of a Data Scientist

Data Scientists: Rarer Than Unicorns

Beyond the Big Data Hype

Big Data: Beyond Cheerleading

Big Data Hubris

Leading by Miles

Predicting Pregnancies, Missing Abortions

Whats Beyond This Book?

Summary

Endnotes

Chapter 2 Data in the 24/7 Connected World

Table of Contents

The Liberated Data: The Open Data

The Caged Data

Big Data Is Big News

Its Not the Size of Big Data; Its What You Do with It

Free Data as in Free Lunch

FRED

Quandl

U.S. Census Bureau and Other National Statistical Agencies

Search-Based Internet Data

Google Trends

Google Correlate

Survey Data

PEW Surveys

ICPSR

Summary

Endnotes

Chapter 3 The Deliverable

The Final Deliverable

What Is the Research Question?

What Answers Are Needed?

How Have Others Researched the Same Question in the Past?

What Information Do You Need to Answer the Question?

What Analytical Techniques/Methods Do You Need?

The Narrative

The Report Structure

Have You Done Your Job as a Writer?

Building Narratives with Data

Big Data, Big Analytics, Big Opportunity

Table of Contents

Urban Transport and Housing Challenges

Human Development in South Asia

The Big Move

Summary

Endnotes

Chapter 4 Serving Tables

2014: The Year of Soccer and Brazil

Using Percentages Is Better Than Using Raw Numbers

Data Cleaning

Weighted Data

Cross Tabulations

Going Beyond the Basics in Tables

Seeing Whether Beauty Pays

Data Set

What Determines Teaching Evaluations?

Does Beauty Affect Teaching Evaluations?

Putting It All on (in) a Table

Generating Output with Stata

Summary Statistics Using Built-In Stata

Using Descriptive Statistics

Weighted Statistics

Correlation Matrix

Reproducing the Results for the Hamermesh and Parker Paper

Statistical Analysis Using Custom Tables

Summary

Endnotes

Chapter 5 Graphic Details

Telling Stories with Figures

Table of Contents

Data Types

Teaching Ratings

The Congested Lives in Big Cities

Summary

Endnotes

Chapter 6 Hypothetically Speaking

Random Numbers and Probability Distributions.

Casino Royale: Roll the Dice

Normal Distribution

The Student Who Taught Everyone Else

Statistical Distributions in Action

Z-Transformation

Probability of Getting a High or Low Course Evaluation

Probabilities with Standard Normal Table

Hypothetically Yours

Consistently Better or Happenstance

Mean and Not So Mean Differences

Handling Rejections

The Mean and Kind Differences

Comparing a Sample Mean When the Population SD Is Known

Left Tail Between the Legs

Comparing Means with Unknown Population SD

Comparing Two Means with Unequal Variances

Comparing Two Means with Equal Variances

Worked-Out Examples of Hypothesis Testing

Best BuyApple Store Comparison

Assuming Equal Variances

Exercises for Comparison of Means

Table of Contents

Regression for Hypothesis Testing

Analysis of Variance

Significantly Correlated

Summary

Endnotes

Chapter 7 Why Tall Parents Dont Have Even Taller Children

The Department of Obvious Conclusions

Why Regress?

Introducing Regression Models

All Else Being Equal

Holding Other Factors Constant

Spuriously Correlated

A Step-By-Step Approach to Regression

Learning to Speak Regression

The Math Behind Regression

Ordinary Least Squares Method

Regression in Action

This Just In: Bigger Homes Sell for More

Does Beauty Pay? Ask the Students

Survey Data, Weights, and Independence of Observations

What Determines Household Spending on Alcohol and Food

What Influences Household Spending on Food?

Advanced Topics

Homoskedasticity

Multicollinearity

Summary

Endnotes

Table of Contents

Chapter 8 To Be or Not to Be

To Smoke or Not to Smoke: That Is the Question

Binary Outcomes

Binary Dependent Variables

Lets Question the Decision to Smoke or Not

Smoking Data Set

Exploratory Data Analysis

What Makes People Smoke: Asking Regression for Answers

Ordinary Least Squares Regression

Interpreting Models at the Margins

The Logit Model

Interpreting Odds in a Logit Model

Probit Model

Interpreting the Probit Model

Using Zelig for Estimation and Post-Estimation Strategies

Estimating Logit Models for Grouped Data

Using SPSS to Explore the Smoking Data Set

Regression Analysis in SPSS

Estimating Logit and Probit Models in SPSS

Summary

Endnotes

Chapter 9 Categorically Speaking About Categorical Data

What Is Categorical Data?

Analyzing Categorical Data

Econometric Models of Binomial Data

Estimation of Binary Logit Models

Odds Ratio

Log of Odds Ratio

Table of Contents

Interpreting Binary Logit Models

Statistical Inference of Binary Logit Models

How I Met Your Mother? Analyzing Survey Data

A Blind Date with the Pew Online Dating Data Set

Demographics of Affection

High-Techies

Romancing the Internet

Dating Models

Multinomial Logit Models

Interpreting Multinomial Logit Models

Choosing an Online Dating Service

Pew Phone Type Model

Why Some Women Work Full-Time and Others Don't

Conditional Logit Models

Random Utility Model

Independence From Irrelevant Alternatives

Interpretation of Conditional Logit Models

Estimating Logit Models in SPSS

Summary

Endnotes

Chapter 10 Spatial Data Analytics

Fundamentals of GIS

GIS Platforms

Freeware GIS

GIS Data Structure

GIS Applications in Business Research

Retail Research

Hospitality and Tourism Research

Table of Contents

Lifestyle Data: Consumer Health Profiling

Competitor Location Analysis

Market Segmentation

Spatial Analysis of Urban Challenges

The Hard Truths About Public Transit in North America

Toronto Is a City Divided into the Haves, Will Haves, and Have Nots

Income Disparities in Urban Canada

Where Is Toronto's Missing Middle Class? It Has Suburbanized Out
of Toronto

Adding Spatial Analytics to Data Science

Race and Space in Chicago

Developing Research Questions

Race, Space, and Poverty

Race, Space, and Commuting

Regression with Spatial Lags

Summary

Endnotes

Chapter 11 Doing Serious Time with Time Series

Introducing Time Series Data and How to Visualize It

How Is Time Series Data Different?

Starting with Basic Regression Models

What Is Wrong with Using OLS Models for Time Series Data?

Newey-West Standard Errors

Regressing Prices with Robust Standard Errors

Time Series Econometrics

Stationary Time Series

Autocorrelation Function (ACF)

Partial Autocorrelation Function (PCF)

Table of Contents

White Noise Tests

Augmented Dickey Fuller Test

Econometric Models for Time Series Data

Correlation Diagnostics

Invertible Time Series and Lag Operators

The ARMA Model

ARIMA Models

Distributed Lag and VAR Models

Applying Time Series Tools to Housing Construction

Macro-Economic and Socio-Demographic Variables Influencing Housing Starts

Estimating Time Series Models to Forecast New Housing Construction

OLS Models

Distributed Lag Model

Out-of-Sample Forecasting with Vector Autoregressive Models

ARIMA Models

Summary

Endnotes

Chapter 12 Data Mining for Gold

Can Cheating on Your Spouse Kill You?

Are Cheating Men Alpha Males?

UnFair Comments: New Evidence Critiques Fairs Research

Data Mining: An Introduction

Seven Steps Down the Data Mine

Establishing Data Mining Goals

Selecting Data

Preprocessing Data

Table of Contents

Transforming Data

Storing Data

Mining Data

Evaluating Mining Results

Rattle Your Data

What Does Religiosity Have to Do with Extramarital Affairs?

The Principal Components of an Extramarital Affair

Will It Rain Tomorrow? Using PCA For Weather Forecasting

Do Men Have More Affairs Than Females?.

Two Kinds of People: Those Who Have Affairs, and Those Who Dont

Models to Mine Data with Rattle

Summary

Endnotes

Index