

LEARNING MADE EASY



Data Analytics & Visualization

ALL-IN-ONE

for
dummies[®]
A Wiley Brand



Jack Hyman et al.



Data Analytics & Visualization

ALL-IN-ONE

**by Jack Hyman; Luca Massaron;
Paul McFedries; John Paul Mueller;
Lillian Pierson; Jonathan Reichental, PhD;
Joseph Schmuller; Alan Simon;
and Allen G. Taylor**

for
dummies[®]
A Wiley Brand

Data Analytics & Visualization All-in-One For Dummies®

Published by: **John Wiley & Sons, Inc.**, 111 River Street, Hoboken, NJ 07030-5774, www.wiley.com

Copyright © 2024 by John Wiley & Sons, Inc., Hoboken, New Jersey

Media and software compilation copyright © 2024 by John Wiley & Sons, Inc. All rights reserved.

Published simultaneously in Canada

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without the prior written permission of the Publisher. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permissions>.

Trademarks: Wiley, For Dummies, the Dummies Man logo, Dummies.com, Making Everything Easier, and related trade dress are trademarks or registered trademarks of John Wiley & Sons, Inc. and may not be used without written permission. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc. is not associated with any product or vendor mentioned in this book.

LIMIT OF LIABILITY/DISCLAIMER OF WARRANTY: THE PUBLISHER AND THE AUTHORS MAKE NO REPRESENTATIONS OR WARRANTIES WITH RESPECT TO THE ACCURACY OR COMPLETENESS OF THE CONTENTS OF THIS WORK AND SPECIFICALLY DISCLAIM ALL WARRANTIES, INCLUDING WITHOUT LIMITATION WARRANTIES OF FITNESS FOR A PARTICULAR PURPOSE. NO WARRANTY MAY BE CREATED OR EXTENDED BY SALES OR PROMOTIONAL MATERIALS. THE ADVICE AND STRATEGIES CONTAINED HEREIN MAY NOT BE SUITABLE FOR EVERY SITUATION. THIS WORK IS SOLD WITH THE UNDERSTANDING THAT THE PUBLISHER IS NOT ENGAGED IN RENDERING LEGAL, ACCOUNTING, OR OTHER PROFESSIONAL SERVICES. IF PROFESSIONAL ASSISTANCE IS REQUIRED, THE SERVICES OF A COMPETENT PROFESSIONAL PERSON SHOULD BE SOUGHT. NEITHER THE PUBLISHER NOR THE AUTHORS SHALL BE LIABLE FOR DAMAGES ARISING HEREFROM. THE FACT THAT AN ORGANIZATION OR WEBSITE IS REFERRED TO IN THIS WORK AS A CITATION AND/OR A POTENTIAL SOURCE OF FURTHER INFORMATION DOES NOT MEAN THAT THE AUTHORS OR THE PUBLISHER ENDORSES THE INFORMATION THE ORGANIZATION OR WEBSITE MAY PROVIDE OR RECOMMENDATIONS IT MAY MAKE. FURTHER, READERS SHOULD BE AWARE THAT INTERNET WEBSITES LISTED IN THIS WORK MAY HAVE CHANGED OR DISAPPEARED BETWEEN WHEN THIS WORK WAS WRITTEN AND WHEN IT IS READ.

For general information on our other products and services, please contact our Customer Care Department within the U.S. at 877-762-2974, outside the U.S. at 317-572-3993, or fax 317-572-4002. For technical support, please visit <https://hub.wiley.com/community/support/dummies>.

Wiley publishes in a variety of print and electronic formats and by print-on-demand. Some material included with standard print versions of this book may not be included in e-books or in print-on-demand. If this book refers to media such as a CD or DVD that is not included in the version you purchased, you may download this material at <http://booksupport.wiley.com>. For more information about Wiley products, visit www.wiley.com.

Library of Congress Control Number: 2024932207

ISBN 978-1-394-24409-6 (pbk); ISBN 978-1-394-24411-9 (ePDF); ISBN 978-1-394-24410-2 (epub)

Contents at a Glance

Introduction	1
Book 1: Learning Data Analytics & Visualizations Foundations	7
CHAPTER 1: Exploring Definitions and Roles	9
CHAPTER 2: Delving into Big Data	19
CHAPTER 3: Understanding Data Lakes	41
CHAPTER 4: Wrapping Your Head Around Data Science	51
CHAPTER 5: Telling Powerful Stories with Data Visualization	81
Book 2: Using Power BI for Data Analytics & Visualization	107
CHAPTER 1: Power BI Foundations	109
CHAPTER 2: The Quick Tour of Power BI	123
CHAPTER 3: Prepping Data for Visualization	141
CHAPTER 4: Tweaking Data for Primetime	167
CHAPTER 5: Designing and Deploying Data Models	183
CHAPTER 6: Tackling Visualization Basics in Power BI	203
CHAPTER 7: Digging into Complex Visualization and Table Data	227
CHAPTER 8: Sharing and Collaborating with Power BI	247
Book 3: Using Tableau for Data Analytics & Visualization	265
CHAPTER 1: Tableau Foundations	267
CHAPTER 2: Connecting Your Data	285
CHAPTER 3: Diving into the Tableau Prep Lifecycle	313
CHAPTER 4: Advanced Data Prep Approaches in Tableau	337
CHAPTER 5: Touring Tableau Desktop	351
CHAPTER 6: Storytelling Foundations in Tableau	371
CHAPTER 7: Visualizing Data in Tableau	391
CHAPTER 8: Collaborating and Publishing with Tableau Cloud	425
Book 4: Extracting Information with SQL	443
CHAPTER 1: SQL Foundations	445
CHAPTER 2: Drilling Down to the SQL Nitty-Gritty	455
CHAPTER 3: Values, Variables, Functions, and Expressions	487
CHAPTER 4: SELECT Statements and Modifying Clauses	513
CHAPTER 5: Tuning Queries	539

CHAPTER 6: Complex Query Design	557
CHAPTER 7: Joining Data Together in SQL	591

Book 5: Performing Statistical Data Analysis & Visualization with R Programming 605

CHAPTER 1: Using Open Source R for Data Science	607
CHAPTER 2: R: What It Does and How It Does It	623
CHAPTER 3: Getting Graphical	651
CHAPTER 4: Kicking It Up a Notch to ggplot2.	671

Book 6: Applying Python Programming to Data Science..... 689

CHAPTER 1: Discovering the Match between Data Science and Python.	691
CHAPTER 2: Using Python for Data Science and Visualization	703
CHAPTER 3: Getting a Crash Course in Matplotlib	721
CHAPTER 4: Visualizing the Data	739

Index.....	761
-------------------	------------

Table of Contents

INTRODUCTION	1
About This Book	1
Foolish Assumptions	3
Icons Used in This Book	3
Beyond the Book	4
Where to Go from Here	4
BOOK 1: LEARNING DATA ANALYTICS & VISUALIZATIONS FOUNDATIONS	7
CHAPTER 1: Exploring Definitions and Roles	9
What Is Data, Really?	10
Working with structured data	10
Looking at unstructured data	11
Adding semi-structured data to the mix	11
Discovering Business Intelligence	12
Understanding Data Analytics	13
Exploring Data Management	15
Diving into Data Analysis	16
Cooking raw data	16
Dealing with data	16
Building data models	17
Performing what-if analysis	17
Visualizing Data	17
CHAPTER 2: Delving into Big Data	19
Identifying the Roles of Data	20
Operations	20
Strategy	21
Decision-making	22
Measuring	23
Monitoring	23
Insight management	24
Reporting	25
Other roles for data	26
Grappling with data volume	28
Handling data velocity	28
Dealing with data variety	29

	What's All the Fuss about Data?	30
	Welcome to the zettabyte era	31
	From data to insight	33
	Identifying Important Data Sources	35
	Role of Big Data in Data Science and Engineering	36
	Defining data science	36
	Defining machine learning engineering	37
	Defining data engineering	38
	Connecting Big Data with Business Intelligence	39
	Analyzing Data with Enterprise Business Intelligence Practices	39
CHAPTER 3:	Understanding Data Lakes	41
	Rock-Solid Water	42
	A Really Great Lake	43
	Expanding the Data Lake	43
	More Than Just the Water	45
	Different Types of Data	45
	Different Water, Different Data	47
	Refilling the Data Lake	48
	Everyone Visits the Data Lake	49
CHAPTER 4:	Wrapping Your Head Around Data Science	51
	Inspecting the Pieces of the Data Science Puzzle	52
	Collecting, querying, and consuming data	53
	Applying mathematical modeling to data science tasks	54
	Deriving insights from statistical methods	55
	Coding, coding, coding — it's just part of the game	55
	Applying data science to a subject area	55
	Choosing the Best Tools for Your Data Science Strategy	57
	Getting a Handle on SQL and Relational Databases	58
	Knowing all about the keys	61
	Investing Some Effort into Database Design	62
	Defining data types	63
	Designing constraints properly	63
	Normalizing your database	64
	Narrowing the Focus with SQL Functions	66
	Making Life Easier with Excel	70
	Using Excel to quickly get to know your data	71
	Reformatting and summarizing with PivotTables	75
	Automating Excel tasks with macros	77
CHAPTER 5:	Telling Powerful Stories with Data Visualization	81
	Data Visualizations: The Big Three	82
	Data storytelling for decision-makers	82

Data showcasing for analysts	83
Designing data art for activists	84
Designing to Meet the Needs of Your Target Audience.	84
Step 1: Brainstorm (All about Eve)	85
Step 2: Define the purpose	86
Step 3: Choose the most functional visualization type for your purpose	86
Picking the Most Appropriate Design Style	87
Inducing a calculating, exacting response	87
Eliciting a strong emotional response	88
Selecting the Appropriate Data Graphic Type.	90
Standard chart graphics	91
Comparative graphics.	94
Statistical plots	95
Topology structures	100
Spatial plots and maps	101
Testing Data Graphics.	103
Adding Context	104
Creating context with data.	105
Creating context with annotations	105
Creating context with graphical elements	105
BOOK 2: USING POWER BI FOR DATA ANALYTICS & VISUALIZATION	107
CHAPTER 1: Power BI Foundations	109
Looking Under the Power BI Hood	109
Posing questions with Power Query.	110
Modeling with Power Pivot	111
Visualizing with Power View.	111
Mapping data with Power Map	111
Interpreting data with Power Q&A	112
Power BI Desktop	112
Power BI Services	112
Knowing Your Power BI Terminology	113
Capacities	113
Workspaces	114
Reports	115
Dashboards	117
Navigation pane.	117
Power BI Products in a Nutshell	118
Introducing the Power BI license options	119
Looking at Desktop versus Services options	120

CHAPTER 2: The Quick Tour of Power BI	123
Power BI Desktop: A Top-Down View	124
Ingesting Data	125
Files or databases?	126
Building data models	128
Analyzing data	129
Creating and publishing items	130
Services: Far and Wide	132
Viewing and editing reports	132
Working with dashboards	136
Collaborating inside Power BI Services	137
Refreshing data	138
CHAPTER 3: Prepping Data for Visualization	141
Getting Data from the Source	142
Managing Data Source Settings	146
Working with Shared versus Local Datasets	147
Storage and Connection Modes	150
Data Sources Oh My!	151
Getting data from Microsoft-based file systems	151
Working with relational data sources	153
Cleansing, Transforming, and Loading Your Data	162
Detecting anomalies and inconsistencies	162
Checking data structures and column properties	163
Data statistics to the rescue	164
CHAPTER 4: Tweaking Data for Primetime	167
Stepping through the Data Lifecycle	167
Resolving Inconsistencies	168
Replacing values	168
Removing rows using Power Query	170
Digging down to the root cause	170
Evaluating and Transforming Column Data Types	171
Finding and creating appropriate keys for joins	171
Shaping your column data to meet Power Query requirements	174
Combining queries	175
Configuring Queries for Data Loading	180
Resolving Errors During Data Import	182
CHAPTER 5: Designing and Deploying Data Models	183
Creating a Data Model Masterpiece	183
Working with Data view and Modeling view	184
Importing queries	186

	Defining data types	188
	Handling formatting and data type properties.	189
	Managing tables	191
	Adding and modifying data to imported, DirectQuery, and composite models.	196
	Managing Relationships	197
	Creating automatic relationships	197
	Creating manual relationships	198
	Deleting relationships.	199
	Arranging Data.	199
	Sorting by and grouping by	200
	Hiding data.	200
	Publishing Data Models	201
CHAPTER 6:	Tackling Visualization Basics in Power BI	203
	Looking at Report Fundamentals and Visualizations	203
	Creating visualizations	204
	Choosing a visualization.	205
	Filtering data	207
	Choosing the Best Visualization for the Job.	209
	Working with Bar charts and Column charts.	209
	Using basic Line charts and Area charts	213
	Combining Line charts and Bar charts	215
	Working with Ribbon charts.	216
	Going with the flow with Waterfall charts	216
	Funneling with Funnel charts.	217
	Scattering with Scatter charts	218
	Sweetening the data using Pie charts and Donut charts.	219
	Branching out with treemaps	219
	Mapping with maps.	221
	Indicating with indicators	222
CHAPTER 7:	Digging into Complex Visualization and Table Data	227
	Dealing with Table-Based and Complex Visualizations	228
	Zeroing in with slicers	228
	Tabling with table visualizations	228
	Combing through data with matrices.	229
	Decomposing with decomposition trees.	230
	Zooming in on key influencers	230
	Using AI Tools to Create Questions and Answers.	231
	Formatting and Configuring Report Visualizations.	232
	Applying conditional formatting	233
	Configuring the report page	233

Exporting reports	235
Perfecting reports for distribution	236
Diving into Dashboards	239
Configuring dashboards	240
Creating a new dashboard	240
Enriching your dashboard with content	242
Pinning reports	245
CHAPTER 8: Sharing and Collaborating with Power BI	247
Working Together in a Workspace	247
Defining the types of workspaces	248
Figuring out the nuts and bolts of workspaces	248
Slicing and Dicing Data	254
Analyzing in Excel	256
Benefiting from Quick Insights	256
Using Usage Metric reports	257
Working with paginated reports	258
Troubleshooting the Use of Data Lineage	258
Datasets, Dataflows, and Lineage	261
Defending Your Data Turf	262

BOOK 3: USING TABLEAU FOR DATA ANALYTICS & VISUALIZATION 265

CHAPTER 1: Tableau Foundations	267
Understanding Key Tableau Terms	268
Data source	268
Data type	269
Data fields	270
Dimensions and measures	270
Continuous versus discrete	272
Filter	273
Aggregation	273
Workbook and worksheet	274
Getting to Know the Tableau Product Line	275
Tableau Desktop	275
Tableau Prep	278
Tableau Server and Tableau Cloud	279
Choosing the Right Version	281
Knowing What Tools You Need in Each Stage of the Data Life Cycle	282
Understanding User Types and Their Capabilities	283
Viewer	283
Explorer	283
Creator	284

CHAPTER 2:	Connecting Your Data	285
	Understanding Data Source Options	286
	Connecting to Data	288
	Making the Desktop or Prep connection	289
	Locating the Server and Online connections	290
	Setting Up and Planning the Data Source	292
	Relating and Combining Data Sources	294
	Working with Data Relationships	296
	Knowing the advantages of relationships	296
	Seeing the disadvantages of relationships	297
	Creating relationships	297
	Editing relationships	299
	Moving tables to create different relationships	299
	Changing the root table of a relationship	300
	Removing tables from a relationship	301
	Joining Data	302
	Understanding join types	302
	Setting up join clauses	303
	Creating a join	304
	Joining fields that contain null values	306
	Blending data from multiple sources	307
	Working with clipboard data	309
CHAPTER 3:	Diving into the Tableau Prep Lifecycle	313
	Dabbling in Data Flows	314
	Connecting the data dots	314
	Going down the data flow pathway	315
	Configuring the data flow	317
	Going with the data flow	318
	Nurturing a flow	324
	Grouping flows	327
	Filtering flows	329
	Saving Prep Data	333
	Automating flows	334
	Crafting published data sources	334
CHAPTER 4:	Advanced Data Prep Approaches in Tableau	337
	Peering into Data Structures	337
	Rows and records	338
	Columns and fields	339
	Categorizing fields	340

	Structuring for Data Visualization	342
	Binning and histograms	343
	Distributions and outliers.	344
	Pivoting with data: Tall versus wide	345
	Normalizing Data.	348
CHAPTER 5:	Touring Tableau Desktop	351
	Getting Hands-On in the Tableau Desktop Workspace	351
	Making Use of the Tableau Desktop Menus	353
	File menu	353
	Data menu	354
	Worksheet menu.	355
	Dashboard menu.	357
	Story menu.	358
	Analysis menu	359
	Map menu	361
	Format menu	362
	Server menu.	363
	Window menu	364
	Help menu	365
	Tooling Around in the Toolbar.	365
	Understanding Sheets versus Workbooks.	369
	Renaming sheets.	370
	Deleting sheets	370
CHAPTER 6:	Storytelling Foundations in Tableau	371
	Working with Dashboards	371
	Configuring the dashboard	372
	Customizing the dashboard.	374
	Adding objects to dashboards.	376
	Creating a Compelling Story	383
	Synthesizing data through a Tableau story	383
	Planning your story to perfection	384
	Surveying the story workspace	385
	Crafting the story.	386
	Formatting the story	389
CHAPTER 7:	Visualizing Data in Tableau	391
	Introducing the Visualizations	392
	The text table	392
	The heat map and highlight table.	394
	Maps with and without symbols	396
	The pie chart	398
	The bar chart	400

	The treemap	401
	Circles and bubbles	402
	The line chart	406
	The area chart	408
	The dual combination chart	411
	The scatter plot	412
	The histogram	414
	The box and whisker plot	415
	The Gantt chart	416
	The bullet chart	418
	Converting a Visualization to a Crosstab	419
	Publishing Visualizations	422
CHAPTER 8:	Collaborating and Publishing with Tableau Cloud	425
	Strolling through the Tableau Cloud Experience	426
	Evaluating Personal Features in Tableau Cloud	430
	Personal Space	430
	Favorites	432
	Recents	434
	Sharing Experiences and Collaborating with Others	435
	Sharing content	435
	Shared with Me	436
	Collections	437
	Explore	439
	Recommendations	440
	BOOK 4: EXTRACTING INFORMATION WITH SQL	443
CHAPTER 1:	SQL Foundations	445
	SQL and the Relational Model	445
	Sets, Relations, Multisets, and Tables	446
	Functional Dependencies	447
	Keys	448
	Views	450
	Users	450
	Privileges	451
	Schemas	451
	Catalogs	452
	Connections, Sessions, and Transactions	452
	Routines	453
	Paths	454

CHAPTER 2: Drilling Down to the SQL Nitty-Gritty	455
Executing SQL Statements.....	455
Interactive SQL.....	456
Challenges to combining SQL with a host language	457
Embedded SQL	457
Module language.....	460
Using Reserved Words Correctly.....	461
SQL's Data Types	461
Exact numerics.....	462
INTEGER	462
SMALLINT	463
BIGINT	463
Approximate numerics.....	464
Character strings.....	466
Binary strings	468
Booleans.....	469
Datetimes.....	469
Intervals	471
XML type.....	471
ROW type	472
Collection types	473
REF types.....	475
User-defined types	475
Handling Null Values.....	478
Applying Constraints.....	479
Column constraints.....	480
Table constraints.....	481
Foreign key constraints.....	483
Assertions.....	484
CHAPTER 3: Values, Variables, Functions, and Expressions	487
Entering Data Values.....	487
Row values have multiple parts.....	488
Identifying values in a column.....	488
Literal values don't change.....	488
Variables vary.....	490
Special variables hold specific values.....	490
Working with Functions	491
Summarizing data with set functions.....	491
Dissecting data with value functions	494
Using Expressions	503
Numeric value expressions	503
String value expressions.....	503
Datetime value expressions.....	504

	Interval value expressions	505
	Boolean value expressions	506
	Array value expressions	506
	Conditional value expressions.	507
	Converting data types with a CAST expression.	510
	Row value expressions	511
CHAPTER 4:	SELECT Statements and Modifying Clauses	513
	Finding Needles in Haystacks with the SELECT Statement	513
	Modifying Clauses	514
	FROM clauses	514
	WHERE clauses	515
	GROUP BY clauses	533
	HAVING clauses	535
	ORDER BY clauses	536
CHAPTER 5:	Tuning Queries	539
	SELECT DISTINCT	540
	Temporary Tables	542
	The ORDER BY Clause	547
	The HAVING Clause	551
	The OR Logical Connective.	555
CHAPTER 6:	Complex Query Design	557
	What Is a Subquery?	557
	What Subqueries Do	558
	Subqueries that return multiple values	558
	Subqueries that return a single value	560
	Quantified subqueries return a single value	563
	Correlated subqueries	566
	Using Subqueries in INSERT, DELETE, and UPDATE Statements	571
	Tuning Considerations for Statements Containing Nested Queries	574
	Tuning Correlated Subqueries.	579
	UNION	584
	UNION ALL	586
	UNION CORRESPONDING	587
	INTERSECT	588
	EXCEPT	590
CHAPTER 7:	Joining Data Together in SQL	591
	JOINS	591
	Cartesian product or cross join	592
	Equi-join	594

Natural join.	596
Condition join.	596
Column-name join.	597
Inner join.	598
Outer join	599
ON versus WHERE.	603
Join Conditions and Clustering Indexes	603

BOOK 5: PERFORMING STATISTICAL DATA ANALYSIS & VISUALIZATION WITH R PROGRAMMING . . . 605

CHAPTER 1: Using Open Source R for Data Science	607
Downloading Open Source R.	608
Comprehending R’s Basic Vocabulary	608
Delving into Functions and Operators	612
Iterating in R.	615
Observing How Objects Work	617
Sorting Out R’s Popular Statistical Analysis Packages	619
Examining Packages for Visualizing, Mapping, and Graphing in R. . .	620
Visualizing R statistics with ggplot2.	620
Analyzing networks with statnet and igraph.	621
Mapping and analyzing spatial point patterns with spatstat . . .	622
CHAPTER 2: R: What It Does and How It Does It	623
The Statistical (and Related) Ideas You Just Have to Know	624
Samples and populations.	624
Variables: Dependent and independent	625
Types of data	626
A little probability	626
Inferential statistics: Testing hypotheses.	628
Null and alternative hypotheses	628
Two types of error	629
Getting R.	630
Getting RStudio	631
A Session with R.	634
The working directory.	634
Getting started.	635
R Functions.	638
User-Defined Functions	639
Comments	640
R Structures	641
Vectors	641
Numerical vectors	642
Matrices	643

	Lists	645
	Data frames	646
	for Loops and if Statements	649
CHAPTER 3:	Getting Graphical	651
	Finding Patterns	651
	Graphing a distribution	652
	Bar-hopping	653
	Slicing the pie	654
	The plot of scatter	654
	Of boxes and whiskers	656
	Doing the Basics: Base R Graphics, That Is	657
	Histograms	657
	Graph features	658
	Bar plots	660
	Pie graphs	662
	Dot charts	662
	Bar plots revisited	663
	Scatter plots	666
	Box plots	669
CHAPTER 4:	Kicking It Up a Notch to ggplot2	671
	Histograms	672
	Bar Plots	675
	Dot Charts	676
	Bar Plots Re-revisited	679
	Scatter Plots	683
	Scatter Plot Matrix	683
	Box Plots	686
BOOK 6: APPLYING PYTHON PROGRAMMING TO DATA SCIENCE		689
CHAPTER 1:	Discovering the Match between Data Science and Python	691
	Creating the Data Science Pipeline	692
	Understanding Python’s Role in Data Science	693
	Considering the shifting profile of data scientists	693
	Working with a multipurpose, simple, and efficient language	694
	Learning to Use Python Fast	695
	Loading data	695
	Training a model	695
	Viewing a result	696

Working with Python	697
Contributing to data science	697
Getting a taste of the language	698
Understanding the need for indentation	698
Using the Python Ecosystem for Data Science	699
Accessing scientific tools using SciPy	699
Performing fundamental scientific computing using NumPy	700
Performing data analysis using pandas	700
Implementing machine learning using Scikit-learn	700
Going for deep learning with Keras and TensorFlow	701
Plotting the data using Matplotlib	701
Creating graphs with NetworkX	702
CHAPTER 2: Using Python for Data Science and Visualization	703
Using Python for Data Science	703
Sorting Out the Various Python Data Types	705
Numbers in Python	706
Strings in Python	706
Lists in Python	707
Tuples in Python	707
Sets in Python	708
Dictionaries in Python	708
Putting Loops to Good Use in Python	708
Having Fun with Functions	709
Keeping Cool with Classes	711
Checking Out Some Useful Python Libraries	713
Saying hello to the NumPy library	714
Getting up close and personal with the SciPy library	716
Bonding with MatPlotLib for data visualization	716
Peeking into the Pandas offering	718
Learning from data with Scikit-learn	719
CHAPTER 3: Getting a Crash Course in Matplotlib	721
Starting with a Graph	722
Defining the plot	722
Drawing multiple lines and plots	723
Saving your work to disk	724
Setting the Axis, Ticks, and Grids	725
Getting the axes	725
Formatting the axes	726
Adding grids	727

Defining the Line Appearance	729
Working with line styles	729
Using colors	730
Adding markers	731
Using Labels, Annotations, and Legends	733
Adding labels	734
Annotating the chart	734
Creating a legend	735
CHAPTER 4: Visualizing the Data	739
Choosing the Right Graph	740
Creating comparisons with bar charts	740
Showing distributions using histograms	741
Depicting groups using boxplots	743
Seeing data patterns using scatterplots	744
Creating Advanced Scatterplots	746
Depicting groups	746
Showing correlations	747
Plotting Time Series	748
Representing time on axes	749
Plotting trends over time	751
Plotting Geographical Data	752
Using an environment in Notebook	753
Using Cartopy to plot geographic data	754
Visualizing Graphs	757
Developing undirected graphs	757
Developing directed graphs	759
INDEX	761

Introduction

Everywhere you go in the business world, you are likely to encounter executives who make decisions driven by tidbits of raw data that together tell a meaningful story. In fact, in our everyday worlds, websites and mobile apps express data using powerful visualizations to explain complex numbers and concepts, not extensive written passages anymore. The phrase “a picture speaks a thousand words” rings true in the world of data analytics and visualization, and for good reason.

Data analytics and visualization allow anyone to turn raw data into meaningful stories and insights. You, as the analyst, act as the detective. Instead of having to solve a mystery with clues, you are provided datasets that, if provided with enough clarity, can answer complex questions using trend and pattern analysis. If you review a dataset enough, you’ll inevitably have an ah-ha moment in your interpretation quest, but if the dataset can be presented visually, you can accelerate your understanding like a racecar going from 0 to 100 miles per hour in seconds.

Data analytics and visualization help you uncover creative ways to showcase data in a manner that is both informative and engaging. Data often starts out as nothing more than a bunch of jumbled numbers; turning those numbers into a story that can influence decisions and drive change is incredibly powerful. Global enterprises rely on folks who have the skills you are about to embark on in this book as a way to determine business strategies, make corporate decisions, and influence change. If you are ready to learn these skills, you are in for a treat with this book.

About This Book

If you’ve picked up this book, you might be on a quest to piece together a whole lot of terms being thrown around in the information economy regarding data, the most precious tool in the information economy. Data is a business asset that sits at the intersection of many disciplines; the resultant product from data can be methodologies, processes, algorithms, and system outputs. To the end user though, the end game is extracting knowledge and insights from the byproducts of data, and taking action upon review.

Book 1 covers the foundational aspects of the data analytics and visualization lifecycle that every user must understand to be proficient as an analytics and visualization savvy. Books 2 and 3 focus on the two leading tools in the enterprise business intelligence market used to perform complex data analytics and visualization tasks; Microsoft Power BI and Tableau. Books 4 through 6 cover the key programming languages used by both proprietary and open-source data analytics and visualization platforms to extract, assess, and visualize data at scale when commercial off-the-shelf enterprise business platforms are unavailable.

This book uses the following technical conventions:

- » Bold text means that you're meant to type the text just as it appears in the book. The exception is when you're working through a steps list: Because each step is bold, the text to type is not bold.
- » Web addresses and programming code appear in monofont. If you're reading a digital version of this book on a device connected to the Internet, note that you can click the web address to visit that website, like this: `www.dummies.com`.
- » For command sequences in software, this book uses the command arrow. Here's an example that uses Microsoft Word: Click the Office button and then choose Page Layout↔Margins↔Narrow to decrease the default margin setting.

If you don't think the book contains any conventions that need to be spelled out in this section, discuss omitting conventions information with your editor.

To make the content more accessible, we divided it into 6 books:

» **Book 1, "Learning Data Analytics & Visualization Foundations."**

Book 1 introduces terms and fundamental concepts. You learn about big data, data lakes, and data science, and you see how you can apply visualization tools to create meaningful stories based on data you collect.

» **Book 2, "Using Power BI for Data Analysis & Visualization."**

Book 2 covers Microsoft Power BI, a data analysis and visualization tool used by many large organizations. This book illustrates how you can use Power BI to make sense of structured, unstructured, and semi-structured data, and develop robust business analytics outputs for your organization.

» **Book 3, “Using Tableau for Data Analysis & Visualization.”**

Book 3 covers Tableau, a data analysis and visualization tool favored by researchers and educational institutions. In this book, you discover how to prepare data and present your findings using Tableau’s storytelling and visualization features. You also see how to collaborate and publish your work with Tableau Cloud.

» **Book 4, “Extracting Information with SQL.”**

Book 4 describes SQL and the relational database model. You discover how SQL is a powerful tool that nonprogrammers can use to write complex queries to get the most out of their data, and more.

» **Book 5, “Performing Statistical Data Analysis & Visualization with R Programming.”**

Book 5 introduces the open-source R programming language. You see how you can use R to perform statistical data analysis, data visualization, and other data science tasks.

» **Book 6, “Applying Python Programming to Data Science.”**

Book 6 describes how Python is used as a data science and visualization tool. The book includes a “crash course” on Matplotlib.

Foolish Assumptions

To get the most out of this book, you need the following:

- » **Access to the Internet:** This may sound a bit obvious. Even with the Desktop client, an Internet connection is required in order to access datasets from the Internet.
- » **A meaningful dataset:** A meaningful dataset includes at least 300 to 400 records containing a minimum of five or six columns’ worth of data.

Icons Used in This Book

Throughout this book, icons in the margins highlight certain types of valuable information that call out for your attention. Here are the icons you’ll encounter and a brief description of each.



BEST
PRACTICE

Best Practice icons highlight points of common knowledge among seasoned professionals in the data industry. If you don't want to look like a complete newbie, follow the well-worn advice described in these paragraphs.



TIP

Tips point out shortcuts or essential suggestions that you can use to do things quicker, faster, and more efficiently.



REMEMBER

Consider these small suggestions that are quite helpful. Remember icons are like signs on the road to suggest a potential better route.



TECHNICAL
STUFF

The Technical Stuff icon marks information of a highly technical nature that you can normally skip over. When appropriate, these paragraphs also suggest specialized resources you may find helpful down the road.



WARNING

The Warning icon makes you aware of a common issue or product challenge many users face. Don't fret, but *do* take note when you see this icon.

Beyond the Book

In addition to the abundance of information and guidance related to data analysis and visualization provided in this book, you get access to even more help and information online at *Dummies.com*. Check out this book's online Cheat Sheet. Just go to www.dummies.com and search for "Data Analysis & Visualization All-in-One For Dummies Cheat Sheet."

Where to Go from Here

The book has three core themes: foundational concepts, tools, and programming languages.

If you want to learn the essential data analytics and visualization concepts, including learning the lingo of the land, head to Book 1.

If you're looking to get up to speed on Microsoft's Enterprise BI tools, head to Book 2. Tableau, a tool used for Enterprise BI but heavily leveraged in communities where data is regulated such as banking, healthcare, insurance, and government, head to Book 3.

The underpinning for data analytics and visualization is SQL, a querying language. To get a crash course on SQL, which is necessary for any proprietary or open-source data analytics and visualization platform, head to Book 4.

Finally, Books 5 and 6 are an introduction to two popular open-source programming languages, R and Python. Both languages can be configured for use with Power BI and Tableau, but are more commonly used with open-source (free) platforms like Jupyter Notebook and Anaconda to conceive data analytics outputs and visualizations. Unlike Power BI and Tableau, open-source tools leveraging programming languages are used in academic settings or by analysts requiring technologies that are data intensive.

1

Learning Data Analytics & Visualizations Foundations

Contents at a Glance

CHAPTER 1: Exploring Definitions and Roles	9
What Is Data, Really?	10
Discovering Business Intelligence	12
Understanding Data Analytics	13
Exploring Data Management	15
Diving into Data Analysis	16
Visualizing Data	17
CHAPTER 2: Delving into Big Data	19
Identifying the Roles of Data	20
What's All the Fuss about Data?	30
Identifying Important Data Sources	35
Role of Big Data in Data Science and Engineering	36
Connecting Big Data with Business Intelligence	39
Analyzing Data with Enterprise Business Intelligence Practices	39
CHAPTER 3: Understanding Data Lakes	41
Rock-Solid Water	42
A Really Great Lake	43
Expanding the Data Lake	43
More Than Just the Water	45
Different Types of Data	45
Different Water, Different Data	47
Refilling the Data Lake	48
Everyone Visits the Data Lake	49
CHAPTER 4: Wrapping Your Head Around Data Science	51
Inspecting the Pieces of the Data Science Puzzle	52
Choosing the Best Tools for Your Data Science Strategy	57
Getting a Handle on SQL and Relational Databases	58
Investing Some Effort into Database Design	62
Narrowing the Focus with SQL Functions	66
Making Life Easier with Excel	70
CHAPTER 5: Telling Powerful Stories with Data Visualization	81
Data Visualizations: The Big Three	82
Designing to Meet the Needs of Your Target Audience	84
Picking the Most Appropriate Design Style	87
Selecting the Appropriate Data Graphic Type	90
Testing Data Graphics	103
Adding Context	104