LEARNING MADE EASY

2nd Edition

Tableau®

for dummies®
A Wiley Brand

Navigate the user interface

Connect to multiple sources of data

Create and share custom visualizations

Jack Hyman
Author of *Microsoft Power BI For Dummies*

Tableau®

2nd Edition

by Jack Hyman

for dummies®
A Wiley Brand

Tableau® For Dummies®, 2nd Edition

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

Published simultaneously in Canada

Library of Congress Control Number: 2023937361

ISBN: 978-1-119-68458-9 (pbk)

ISBN 978-1-119-68462-6 (ebk); ISBN 978-1-119-68459-6 (ebk)

Contents at a Glance

Table of Contents

Introduction

Maybe you've picked up this book because you want to know about business intelligence and data analytics from the perspective of an enterprise vendor. Or perhaps you want to get into the proverbial weeds of learning how to use an enterprise-class business intelligence platform to tell stories using data visualization techniques. As you read *Tableau For Dummies,* 2nd Edition, you may just develop a zest for both.

Tableau is a French word that means to present a graphic description or a representation, which may include artistic groupings, arrangements, or scenes. Translating this idea into the technology platform, the Tableau platform allows users to take their datasets and create competing data visualizations, reports, and key performance indicators (KPIs) to help tell a story.

Tableau For Dummies, 2nd Edition, offers a look at all the versions of Tableau, but the central focus is on working with Tableau Desktop, Tableau Prep, and Tableau Cloud. Whether you are a novice learning how to create robust data solutions, an analyst looking to explore your datasets, or an end user wanting to consume a story told by Tableau through data visualizations, this book will give you a great start on working with Tableau.

About This Book

If you purchased this book from an online website such as Amazon or Barnes & Noble, it may interest you to know that these retailers use data analytics tools, including Tableau, to evaluate user behavior and customer purchasing patterns. Deciding to buy this book instead of the 250+ other books and counting on Tableau authored over the past five years may be the result of a pervasive pattern. Or perhaps it comes down to what retail market you live in when searching for this book. Ah, the beauty of big data!

In this Tableau primer, you find out how to use Tableau to delve into various questions and scenarios. Examples in this book use publicly accessible datasets found on the website Kaggle.com, an open-source data science community owned by Google, and usaspending.gov, the official source for all spending records in the United States Government.

Tableau For Dummies, 2nd Edition, has been completely revamped since the first edition of 2015. Since the acquisition of Tableau by Salesforce in 2019, many new application features have been added to the product family. Remember that what you see and read in this book reflects the 2022 edition of Tableau, not older versions.

This book doesn't get deep into the programmatic weeds, nor does it show you everything about all the product offerings of Tableau. Instead, this book focuses on understanding and using Tableau Desktop, Tableau Prep, and Tableau Cloud — what I call the foundation for all things Tableau.

If you've used other best-in-breed business intelligence and data analytics solutions, you'll see that Tableau keeps in line with commonly desired features available industry wide. However, Tableau offers many unique nuances that this book helps you uncover. Some concepts are easy to grasp, whereas others take time and patience, even for the most experienced data scientist.

As you read the book, I point to the specific websites for those areas that get a bit hairy. If vendors such as Tableau change the URL to the website, don't fret! Run a Google search for using the heading of whatever section of the book you're currently reading as a guide to finding the latest resources. I'm confident that you'll often find the updated link on the first page of the results.

Also, the many step-by-step exercises in the book use easily obtained, existing datasets to demonstrate the concepts. And you don't have to start at Chapter 1 and work your way through each chapter to move on to the next (most of the time).

Now, I have some good and bad news for you. Getting a free copy of Tableau is easy for a short while. Tableau offers a free trial of the foundational solutions we cover in this book for 14 days. If you need to take a spin for a few more days, don't hesitate to ask for a trial extension. Instructions are abundantly available online to achieve this. Be warned that you won't get all the bells and whistles needed to complete this book until you purchase an actual license, though.

To get a handle on everything covered in *Tableau For Dummies,* 2nd Edition, you'll need access to (a minimum) one Tableau Creator license, which costs about $900 per year. Yes, there are cheaper options, such as a Tableau Explorer or Tableau Viewer license. The only way you'll get access to all the tools covered in the book (especially Desktop and Prep) is by acquiring a Creator subscription.

Tableau offers a free edition for students and educators in K-12 or those active in the college classroom reading this book and wanting to get through most of the exercises. The catch: You must have a valid institutional email address and credentials to back the claim. To get your hands on the academic edition, search for the Academic Programs link under the Resources menu at `www.tableau.com`.

Foolish Assumptions

I've written this book with a few assumptions about you in mind:

- **You want to learn about data analytics and business intelligence.** For readers new to business intelligence and data analytics, this book introduces a cornucopia of terminology and high-level concepts in the context of a leading industry tool. By the time you finish this book, I hope you'll feel comfortable enough to consider yourself no longer a newbie.
- **You hope to learn a new enterprise business intelligence and data analytics platform.** In this book, I walk you through the core features you need to know to get Tableau up and running in an enterprise-scale business intelligence operation. I also help you identify those features so that you can grab the crowd's attention.
- **You need a one-stop shop and hands-on Tableau reference:** If you are overwhelmed by Tableau's online documentation or the litany of online training offered, you are not alone. This book synthesizes the key messages across the foundational products into a single reference.

Regardless of your situation, I hope that by acquiring this book, you'll gain tremendous knowledge to help accelerate your professional and personal goals as a data and analytics expert.

Icons Used in This Book

If you've read a *For Dummies* book before, you're probably familiar with the icons, but here's what I've used them for:

The Tip icon offers a few pieces of advice as you work diligently through the book.

This icon points to those "must know" concepts. If you see one of these icons, take a gander because it provides sage advice on dealing with a significant product, feature, or concept.

If you're looking to dig a bit further into the rabbit hole, I point out exactly what to look for and often how to get to the finish line. Take or leave these bits as you see fit.

WARNING

The Warning icon is the equivalent of an "uh-oh; watch out!" Heed the sound advice in these paragraphs; their goal is to keep you from getting stuck in the Tableau trenches.

ON THE WEB

If you are curious and want to learn well beyond the book, the On The Web icon points to detailed references and solutions, often directly on Tableau.com.

Beyond the Book

Every *For Dummies* book has an associated cheat sheet available online, and *Tableau For Dummies*, 2nd Edition, is no different. The cheat sheet offers you essential tips, tricks, and shortcuts to achieve mastery in Tableau quickly. You can locate the cheat sheet by visiting `https://www.dummies.com` and typing **Tableau For Dummies** in the Search box.

Where to Go from Here

Many folks, including myself, tend to focus on a few critical areas when reading a *For Dummies* book. Although the flow of the book makes sense if you read each chapter sequentially, I anticipate that you'll dance around a bit between chapters. And you indeed can, because all the chapters stand on their own. With that, let's rock and roll!

1 Tackling Tableau Basics

IN THIS PART . . .

Grasp the key concepts necessary to become proficient as a data analyst using Tableau.

Learn how to build the appropriate Tableau solution based on the modular approach to delivery for you and your organization.

Understand the full data life cycle and how it fits into various data modeling and visualization approaches.

IN THIS CHAPTER

- Differentiating among business intelligence, data analytics, and data visualization
- Grasping key Tableau terminology
- Discovering Tableau Desktop and Tableau Prep installation prerequisites
- Tackling data selection fundamentals

Chapter 1
Learning Tableau Lingo

There is much hype about data, and the use of business intelligence, data analytics, and data visualization tooling gets plenty of hype as well. Although there are many enterprise business intelligence tools on the market, Tableau stands out among the leaders for being bundled as a single platform for business intelligence, analytics, and visualization.

In this chapter, you start exploring the Tableau landscape by discovering the main Tableau terminology you need to familiarize yourself with regarding business intelligence, data analytics, and data visualization functionality. In addition, you can dip your toes into what it takes to install Tableau applications and the various file-based output types produced depending on the Tableau product.

What Is Tableau?

Tableau is a business intelligence platform that helps users see and understand their data using highly visual representations. Unlike other enterprise business intelligence platforms, Tableau incorporates business intelligence, data analytics, data science, data mining, and data visualization into a single solution. As a result, its capabilities are considered the broadest and deepest for data evaluation on the market.

In 2019, Salesforce acquired Tableau. At the time, Tableau's focus on data was big but not all-encompassing. It included enterprise data applications, data management and governance, visual analytics, and end-to-end storytelling. As with every other platform on the market, machine learning (ML) and artificial intelligence (AI) have become entrenched in the platform. Salesforce's Einstein AI engine is built into Tableau to help accelerate data analytics predictions, provide a strong recommendation engine, and afford an advanced workflow while touting a low-code development environment.

Tableau is not a single product but is rather a suite of products that includes Tableau Desktop, Tableau Prep, and Tableau Server or Tableau Cloud. Chapter 2 describes the purpose of each in more detail, but in brief, people use Tableau Desktop to create their data models. In contrast, Tableau Prep facilitates data preparation. And when users are ready to collaborate with others, they must publish their outputs from Desktop and Prep to Tableau Server or Tableau Cloud.

Tableau and Business Intelligence

The term *business intelligence* refers to taking in the big picture of an organization's activities and goals, from the collection and analysis of data to the presentation and dissemination of the data using a single platform. A look at the big picture is precisely what you get with Tableau. This best-in-breed platform allows users like yourself to customize views of their data to make data-driven decisions at speed and scale.

Why do folks like you and me need an enterprise business intelligence (BI) solution to organize data? The more data you have, the more difficult it is to dig in and get the information quickly. Making informed decisions requires various capabilities, from data mining to visualizations and analytics. With business intelligence solutions, you get everything under a single umbrella. The key benefits of business intelligence are plentiful, but here are the main ones:


REMEMBER

- **Provide a platform for faster analysis:** BI platforms perform heavy-duty data processing, leading to quick calculations and the creation of stunning visualizations. Assuming that you've connected to your data source and you have already gone about prepping the data with a robust data model, Tableau can accelerate the visualization and analysis process by as much as 100 times as conducting data analysis and business intelligence activities manually, especially when integrating many data sources into a single repository.
- **Create business efficiency and driving decisions:** Leaders can benchmark results with speed and agility when a business intelligence platform offers a holistic view of operations. It's easy to spot opportunities and find those

needle-in-a-haystack moments. Instead of spending hours poring through datasets, users can filter, aggregate, and forecast using Tableau data analytics and visualization options, thereby cutting down the time to make decisions from months, weeks, or days to perhaps even minutes. Talk about saving time!

- **Drive customer and employee experience satisfaction:** What is the worst possible thing for an organization to experience? Sure, most say financial loss. But financial loss results from two factors: lack of customer satisfaction and low employee morale. A primary culprit is the inability of customers and employees to access data quickly; it impacts their entire experience of interacting with the organization, internally and externally. Investing in business intelligence solutions that present a 360-degree view from all data sources can lead to less time worrying about analysis paralysis and more time innovating. The opportunity costs are often measurable in loyalty and, yes, financial rewards.
- **Have data you trust:** When you have many data sources, organizations try to figure out ways to control the disorganized chaos. When you have thousands of Excel or CSV files, a good tactic is to centralize them in a single data repository. But wait a moment: How do you connect the dots — that is, discover the relationships between the data in those files? The answer is to use a business intelligence solution. Relationships exist if the data is like-kind, and you can create potential single-point data sources, hence the use of governed data repositories in a business intelligence platform such as Tableau. Trusted data is not limited to the one-off files; engagement rules apply to relational and nonrelational database stores with tens of millions of records.

Connecting Big Data with Business Intelligence

Make no mistake: The term *big data* is undoubtedly a catch-all buzzword. It pops up a lot in this book. It's meant to encompass five aspects of a business intelligence activity: data volume, data velocity, data veracity, data value, and data variety. Big data brings together *unstructured data* (data with no organized convention), *semi-structured data* (data that has some logical order but isn't necessarily formalized), and *structured data* (data that is formalized or organized). Each of these data types maintains some level of these five attributes:

- **Volume:** The amount of data that exists
- **Velocity:** The speed at which data is generated and moves

- **Veracity:** The quality and accuracy of data available
- **Value:** The credibility, in monetary and nonmonetary terms, that the data provides
- **Variety:** The diversity of data types available within the dataset

Big data is paramount for business intelligence solutions such as Tableau because businesses constantly create more data, practically by the minute. These businesses must keep up with the data deluge. A good business intelligence platform such as Tableau grows with the increasing demands; however, if the data is not maintained, your ability to handle data visualizations and the associated data sources also becomes impaired. Therefore, it's essential to implement good data hygiene and maintenance practices.

Analyzing Data with Tableau

Don't get business intelligence confused with data analytics. Business intelligence platforms use data analytics as a building block to tell the complete story. A data analyst or scientist evaluates the data using the treasure trove of tools built into Tableau, from advanced statistics to predictive analytics or machine learning solutions to identify patterns and trends.

Tableau offers that end-to-end data analytics experience so that the analyst, scientist, and collaborator can complete the entire data life cycle, from gathering, prepping, analyzing, collaborating, and sharing data insights. The big difference between Tableau and its competitors is the self-service nature of the offering, allowing users to ask questions or predict the kind of visualizations the user may require without manually completing the work, thanks to the predictive Einstein AI engine.

Like the three-year-old child asking "Why?" all the time, as you ask more questions and the platform learns, Tableau builds an analysis output while simultaneously learning from the output. The result is an opportunity for the system to understand why something happens and what can happen next. Business intelligence platforms take the resulting models and algorithms and break these results into actionable language insights for data mining, predictive analytics, and statistics. The final product is data analytics, the byproduct of answering a specific question (or set of questions). The collection of questions helps the organization move forward with its business agenda.

Visualizing Data

Raw data that is transformed into useful information can only go so far. Assume for a moment that you were able to aggregate ten data sources whose total record count exceeded 5 million records. As a data analyst, your job was to try to explain to your target audience what the demographics study dataset incorporates among the 5 million records. How easy would that be? It's not simple to articulate unless you can summarize the data cohesively using some data visualization.

Data visualizations are graphical representations of information and data. Suppose you can access visual elements such as charts, graphs, maps, and tables that can concisely synthesize what those millions of records include. In that case, you are effectively using data visualization tools to provide an accessible platform to address trends, patterns, and outliers within data.


TIP

For those who are enamored with big data, the use of data visualization tools helps users analyze massive amounts of data quickly by applying data-driven decisions using graphical representations rather than requiring users to parse through lines of text one by one.

Understanding Key Tableau Terms

Before you begin drinking from the terminology firehose, I want to set the record straight on a few things. Tableau has its own product-specific terminology, but there are also terms you can't escape no matter what business intelligence and data analysis tool you use, whether it's Microsoft Excel, Microsoft Power BI, IBM Cognos, or others. In this section, I review the most critical Tableau-specific terminology, not the entire business intelligence dictionary.

Data source

A *data source* in Tableau comes from anywhere that Tableau can extract, transport, and load relational and nonrelational data. Sources of data used by Tableau are often divided into the four classifications, with some examples of several:

- **Files:** `.csv`, `.txt`, Excel
- **Relational databases:** Oracle, SQL Server, DB2

- **Cloud databases and virtualization platforms:** Microsoft Azure SQL, Google Big Query, Amazon Aurora, Denodo
- **ODBC datastores:** Datastores using ODBC-related connections

Figure 1-1 shows an overview of the abundant number of data sources you can connect to in Tableau Desktop.

A Tableau data source may contain multiple data connections to different databases or files, as described previously. The connection information includes where the data is located, such as the filename and path of the network location, or perhaps details on connecting to the data source, such as the database server name and the authentication credentials. Regardless, many data sources can connect in a single instance of Tableau. Still, categorically, they connect to some file or server connection, whether local or cloud based.

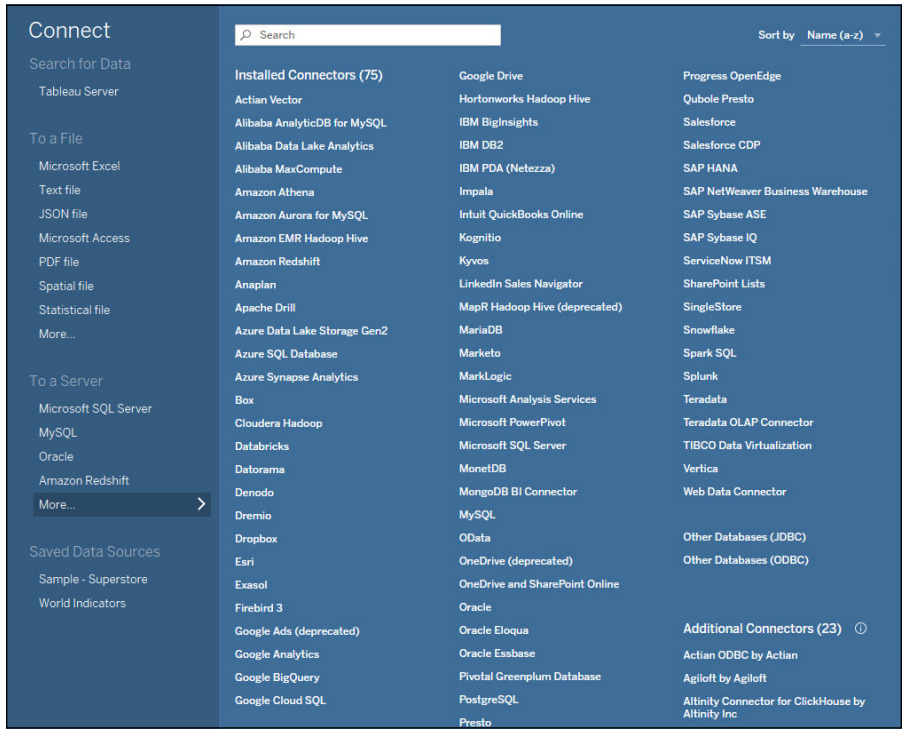


FIGURE 1-1: A sampling of Tableau data sources.

Data type

Going down the data path a bit more, a data field, which is part of a data source (see more details in the next section), must always have a data type. A *data type* reflects whether the field is a number, a type of date, or a string. For example, every area code is an integer (703); a date of birth represents a date (01/01/23); and a state on the U.S map ("Virginia") is a string. Users can identify the data type they are looking for as part of the data field in the Data pane. Each data type also includes one of several icons, including those represented in Figure 1-2. Although the examples are not exhaustive, you see a few common examples of data type icons mapped against their respective data types. The complete list of Tableau data types includes

- Text (string) value
- Date value
- Date & Time value
- Numerical value
- Boolean value (relational data only)
- Geographic value (map data)
- Cluster groups

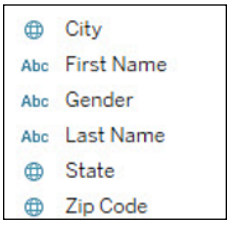
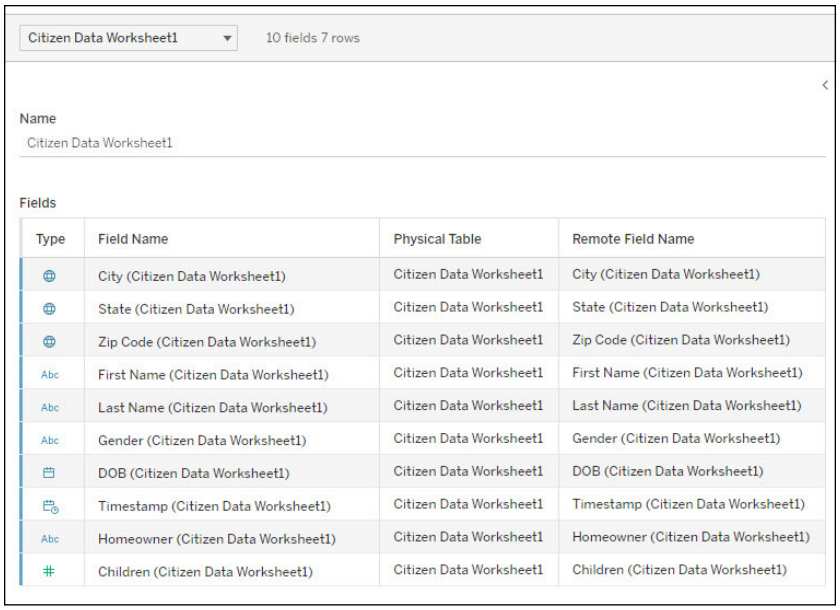


FIGURE 1-2: Examples of data types icons.

Data fields

Every time you connect a data source to Tableau, the connection presents the users with one or more tables from said source. A table includes many data fields composed of a collection of several data types.

As shown in Figure 1-3, data fields are explicitly defined as dimensions or measures as the Tableau database is created. Based on data integrity and quality, Tableau automatically organized the data fields. All data fields containing text date or Boolean values are dimensions by default. On the other hand, fields containing numerical values are measures. The next section talks about how Tableau deals with dimensions and measures.



Type	Field Name	Physical Table	Remote Field Name
🌐	City (Citizen Data Worksheet1)	Citizen Data Worksheet1	City (Citizen Data Worksheet1)
🌐	State (Citizen Data Worksheet1)	Citizen Data Worksheet1	State (Citizen Data Worksheet1)
🌐	Zip Code (Citizen Data Worksheet1)	Citizen Data Worksheet1	Zip Code (Citizen Data Worksheet1)
Abc	First Name (Citizen Data Worksheet1)	Citizen Data Worksheet1	First Name (Citizen Data Worksheet1)
Abc	Last Name (Citizen Data Worksheet1)	Citizen Data Worksheet1	Last Name (Citizen Data Worksheet1)
Abc	Gender (Citizen Data Worksheet1)	Citizen Data Worksheet1	Gender (Citizen Data Worksheet1)
📅	DOB (Citizen Data Worksheet1)	Citizen Data Worksheet1	DOB (Citizen Data Worksheet1)
📅	Timestamp (Citizen Data Worksheet1)	Citizen Data Worksheet1	Timestamp (Citizen Data Worksheet1)
Abc	Homeowner (Citizen Data Worksheet1)	Citizen Data Worksheet1	Homeowner (Citizen Data Worksheet1)
#	Children (Citizen Data Worksheet1)	Citizen Data Worksheet1	Children (Citizen Data Worksheet1)

FIGURE 1-3: Examples of data fields.

Dimensions and measures

In Tableau, dimensions and measures are both data field types. If the field type contains non-numeric data, Tableau references the field as a dimension. Examples include the day of the week, a product category, or geographic data. These variable types don't allow you to complete mathematical equations. Here's an example of an equation with variable types:

State + City / Country = Invalid

All these items are strings because you can't add a state plus a city and divide it by a country to get some magical answer, right?

In Tableau, you can drag each of these fields into a view, which is the part of the Tableau canvas where a visualization is created. Tableau creates headers for each data field. That means you can think of each field as a category, or a dimension

of data. If the dimension of data is placed in a row, the header label is vertically placed. The label is horizontally placed if the dimension is placed in a column. An example of data placed in both rows and columns is displayed in Figure 1-4.

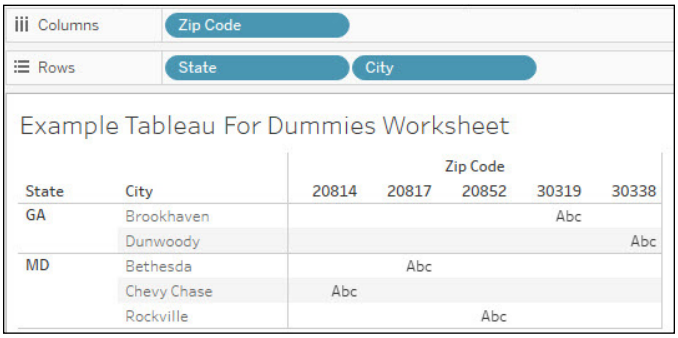


FIGURE 1-4: Rows and column data for dimensions in Tableau.

Measures are numerical data field types. Tableau assumes that these field types are continuous and tags these values by default. Examples of measures include temperature and financial instruments. Unlike independent dimensions, or values that do not rely on other data fields, measures are dependent because they allow you to do the math, as in the following example:

Age (20) + Age (1) / Age (3) = Age (7)

As with dimensions, if you drag a measure into a view, Tableau creates a continuous axis. If a measure is placed in a row, the axis is vertical, whereas a column is horizontal.

In Figure 1-5, you can see that each row (dimension) contains a state, city, and zip code. The column data looks at each value individually and then aggregates the data in the data setup. For example, three individual records in Bethesda, MD 20817 contain children identified. Aggregated, the measure is SUM (3).

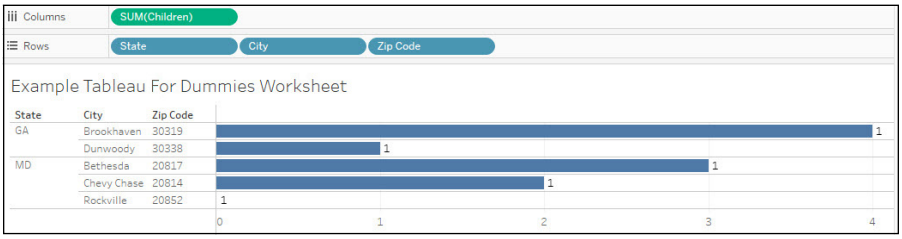


FIGURE 1-5: Rows and column data for measures in Tableau.

Continuous versus discrete

As you'll quickly realize, Tableau separates many concepts based on mathematical reasoning. If a field is based on mathematical representation, Tableau refers to this data as *continuous.* On the other hand, if the data is non-numeric, the data is known as *discrete.*

When it comes to continuous data, you are looking for data that is unbroken, whole, or without interruption. That means data that contains a range of values such as temperature, time, or monetary values. If the data can be added, averaged, or aggregated, and appear as a measure in Tableau, you can almost certainly assume that the value is continuous.

Discrete data is almost always individualized, separate, and unique data. You can have only a particular value. For example, do you have more than one shoe size at a time? Can you be at more than one place at a given location? How many distinct individuals can you claim on a personal tax return? The number 2.39 is not possible; 1, 2, or 3 is more like it.

With discrete data, you have no way to add, average, or aggregate the data points because the values will always be unique by default.

When dragged onto the Tableau View area, discrete data appears as a *blue pill* to form a discrete axis on a chart. Continuous data, on the other hand, appears as a *green pill* to form a continuous axis on a chart.

Filter

The capability for filtering data is one of the essential features of any business intelligence solution associated with big data. Tableau lets a user filter data, whether an individual view contains a few records or an entire data source with millions of records based on dimensions, measures, or values.

As with databases, filtering helps a user see only the data they need based on targeted criteria. When using Tableau filters, you can visualize the data in a readable, actionable format. The real benefit of filtering is to streamline data to limit the number of records for improved performance. An example filter would be to filter all the U.S. states with the word *New.* The result set would return a response of New Hampshire, New Jersey, New Mexico, and New York.

Various filter types are available in Tableau, including the extract, data source, context, dimension, measure, and user filters. You dive into filters a bit more in Chapter 7.

Aggregation

Combining data, also known as *aggregation*, is not uncommon in a business intelligence platform. In Tableau, aggregating measures or dimensions is pervasive. However, aggregation is often numerically focused, meaning focused on the use of measures. Suppose you add a measure to a view. In that case, the aggregation is applied to the specific measure by default, which varies based on context. Read on for an example.

Pretend for a moment that you're the CEO of a Fortune 100 company (think Walmart, Coca-Cola, or Exxon). One of your data analysts prepares a report for you that presents the minimum, maximum, summary, and average number of sales opportunities for a specific product in each region. The scenario would appear as follows (with the bold signifying each data field that is aggregated).

> **Opportunity Value** = 20,000 **products sold** in five **varieties** across 4 **regions** with a customer population of 1,000,000 **households**.

You've now calculated the opportunity value by utilizing the aggregation functions, a way to calculate a set of values and derive a single value.

WARNING

There are limits to what you can aggregate. You can only limit data found in relational data sources. Multidimensional data sources contain data that has already been aggregated, which is impossible to complete. Furthermore, at this time, multidimensional data source aggregation is supported only in the Windows edition of Tableau Desktop.

Workbook and worksheet

Tableau hasn't deviated much from other industry-leading products when it comes to the name of file and formatting conventions. There is a Tableau *workbook*, the main Tableau file, which contains a collection of sheets. The collection of sheets represents the workbook much like that in Microsoft Excel or Microsoft Power BI. In Tableau, a *worksheet* is a single file within a workbook. A worksheet is an element within a dashboard or story.

Although the workbook represents the proverbial catalog of dashboards or stories, the worksheet is a single element or a view. Figure 1-6 represents an example of a single worksheet contained within a Tableau workbook.

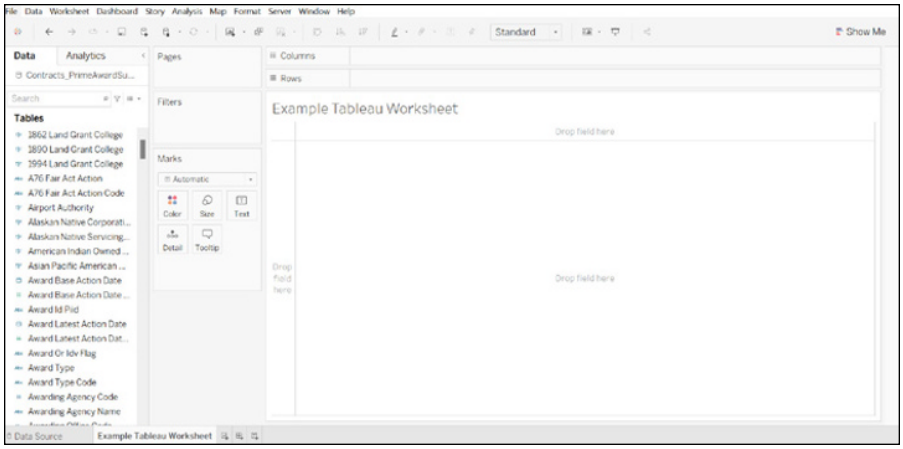


FIGURE 1-6: A Tableau worksheet.

REMEMBER

Here are points to consider when thinking about the use of workbooks and worksheets:

- A Tableau worksheet may contain a single view with many shelves, cards, legends, and analytics panes, which are included as part of a single sidebar on a single page to tell a story.
- When you add many worksheet pages to a workbook, you can generate a *dashboard,* which is a collection of views from many worksheets.
- As you create many worksheets within a workbook, you are compiling a *story,* which is a sequence of worksheets that paint a picture to fuse information.
- Most notably in Tableau Desktop, but also in Tableau Cloud, you can combine views of data by dragging and dropping fields onto the Tableau *shelves,* which are part of a worksheet and which help you create presentations.

In Chapter 7, you take a tour of a worksheet and workbook to see how to collect, organize, and extract data.

Gearing Up for the Tableau Journey

Now that you know the basic Tableau terminology, you can dance your way into installing Tableau. You may be scratching your head, asking "Aren't some instances of Tableau in the cloud?" The answer is yes. Tableau is not a single application; it's a suite of applications. Some Tableau applications still sit on your