Chapter 1

Introduction and Getting Started with Tableau

Tableau was created to empower people to analyze their data regardless of the level of their technical know-how. At the core of Tableau is VizQL, an innovative visual query language that translates mouse inputs such as drag-and-drop into database queries. This allows the user to quickly find insights in their data and to share the results with others.

Crucially, it is not necessary to know what you are looking for or how you want to present your findings. Instead, with Tableau, you can immerse yourself in data. Through visual analysis, you will be able to unearth patterns and relationships in your data that you

might not have known existed. In this regard, Tableau is different from other tools, which often require you to know beforehand in what form you want to display your data.

The purpose of this chapter is to introduce you to the different products that make up the Tableau application suite, the Tableau user interface, and to how Tableau processes your data. We will also introduce the sample dataset that is used throughout this book and provide a first glimpse of the possibilities that Tableau gives you for creating data visualizations.

By the end of this chapter you will be able to:

- Install Tableau on your computer.
- Identify data that is suitable for analysis.
- Create your first data visualization in Tableau.

THE ADVANTAGES OF A MODERN ANALYTICS PLATFORM

The first thing you typically do in Tableau is to connect to a dataset. The data can come from simple files, databases, data cubes, data warehouses, Hadoop clusters, or even different cloud services such as Google Analytics. Next, you interact with the Tableau interface to query your data visually and to display the results in various types of charts and maps. Then, you can collate the individual charts in a dashboard in order to put them into the right context.

Finally, depending on the product used, there are different options for communicating the results with others, from sending individual workbooks, to embedding interactive dashboards, to sharing them on social media. Tableau helps you with both the analysis as well as the communication of results, by providing capabilities such as the creation and sharing of explanatory diagrams, data stories, and interactive dashboards (see Figure 1.1).

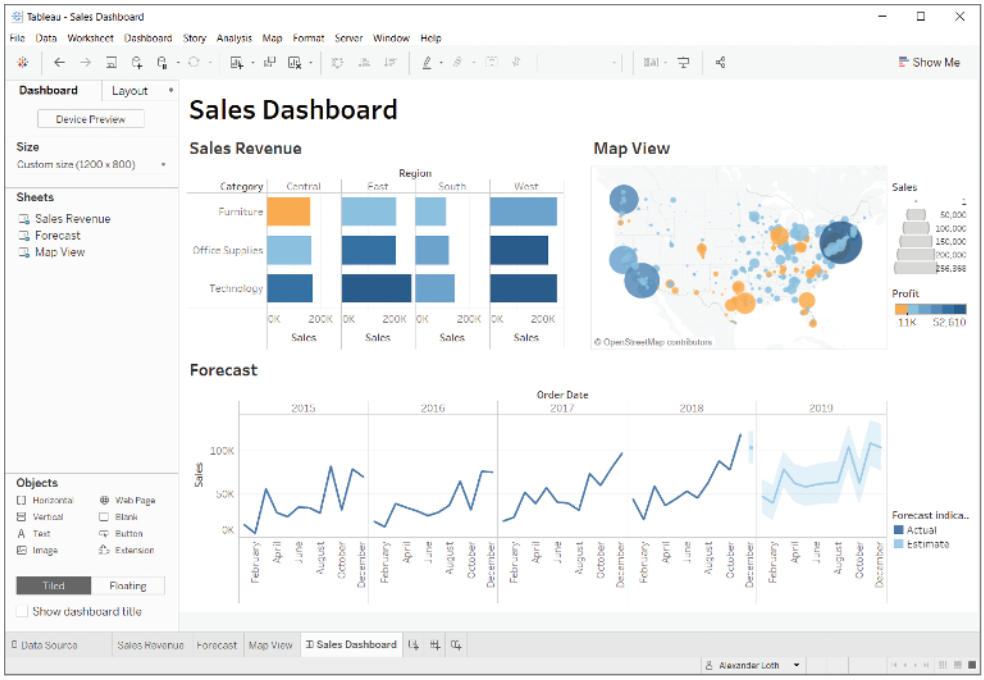

Figure 1.1 An interactive sales dashboard. We will build this in Chapter 8.

MY PERSONAL TABLEAU STORY

I first came across Tableau in 2009, when I was writing my thesis at CERN, the European Organization for Nuclear Research in Geneva. I was exploring the landscape of available tools for the visualization and communication of data because I was not happy with the clunky, inflexible solutions that were commonly used back then.

Like most of my colleagues at CERN, I spent a lot of time aggregating data in Python, a popular universal programming language, only to then visualize it in another tool, the command-line tool GnuPlot. It was a struggle to keep all the scripts well maintained, and even small changes required a lot of time and effort.

When new data came in, the scripts had to be re-run. The resulting visualizations were, of course, static and didn't offer any interactivity to the end user. And the software packages I used had a lot of dependencies that had to be resolved every time a new version became available.

When I eventually learned about Tableau, I was amazed by the ease of use of the graphical interface and the possibility of being able to interact with my data directly.

Every time I dropped another measure or dimension onto the canvas, I got new insights from my data. What used to take me hours could now be done in minutes, and it was fun, to boot! The interactivity of the resulting dashboards and the ability to have them automatically refresh when the underlying data changed sealed the deal for me. I was a fan. I still feel as passionately about Tableau today as I did back then, and I hope to be able to impart some of that enthusiasm to the readers of this book.

THE TABLEAU APPLICATION SUITE

Some readers may have bought this book because they already have one or more Tableau products installed on their machine and would like to jump right in and learn how to use them. But for those who are not so familiar with the different Tableau products, here is a quick overview:

Tableau Desktop Tableau Desktop is an application for Windows and Mac, appreciated by both analysts and business users. In Tableau Desktop, you can connect to flat files (such as Excel and CSV files) and save your workbooks to your local hard drive. To tap into an organization's IT infrastructure, you can also use Tableau Desktop to connect to a host of different database solutions, and you can share your workbooks via Tableau Server or the cloud-based Tableau Online.

Tableau Prep Tableau Prep is the latest addition to the Tableau product suite and is designed to help you prepare your data before you analyze it in Tableau Desktop. The visual interface allows you to quickly merge differently formatted datasets, clean the data, and unify the level of aggregation. Tableau Prep fits seamlessly into your analysis workflow.

Tableau Server Tableau Server is a platform for data analysis and is used by small family-run businesses and large Fortune 500 companies alike. It is intended for the organization-wide provision of data visualizations and dashboards that can be viewed in a browser and are frequently embedded into the organization's intranet.

Tableau Online Tableau Online is a Tableau-hosted solution for storing and deploying dashboards. It provides similar functionality to Tableau Server but is a cloud-based service. No purchase and maintenance of server hardware is necessary here.

Tableau Public Tableau Public is a hosting service for the publication of data visualizations to the web. It is used by newsrooms and bloggers but also by companies, research institutes, governmental bodies, and non-governmental organizations that aim to get their data stories into the public eye. The interactive visualizations can be viewed in the browser directly on the Tableau Public platform, or they can be embedded into blogs and websites.

Tableau Reader Tableau Reader is a free desktop application that allows you to open and interact with Tableau workbook files that have been created in Tableau Desktop. However, it is not possible to make any changes to the visualizations in Tableau Reader.

NOTE The figures throughout this book show Tableau Desktop version 2019.1, unless stated otherwise. The web-edit screen of Tableau Server and Tableau Online contains a number of features that you might recognize from Tableau Desktop. But the functionality of the browser-based products is still limited when it comes to creating new visualizations and dashboards. Therefore, I advise you to install Tableau Desktop on your machine, especially if you are still new to Tableau. The following section will provide more information about the system requirements and the installation process of Tableau Desktop.

INSTALLING TABLEAU DESKTOP

Installing Tableau Desktop is a simple process and takes only a few minutes. Therefore, this will be a very brief section.

System Requirements for Tableau Desktop

Before installing Tableau Desktop, be sure your machine meets the necessary requirements for this application. Tableau Desktop is available for Windows and Mac.

These are the official minimum requirements for a Windows installation:

- Microsoft Windows 7 or later (64-bit)
- Microsoft Server 2008 R2 or later
- Intel Pentium 4 or AMD Opteron processor or later
- 2 GB RAM
- At least 1.5 GB of free hard disk space

These are the official minimum requirements for a Mac installation:

- iMac/MacBook 2009 or later
- OS X 10.10 or later
- At least 1.5 GB of free hard disk space

Should you wish to work with large datasets, I recommend the following additional specifications:

- Latest service pack or update for your operating system
- Intel Core i3/i5/i7/i9 or AMD FX processor or later
- At least 8 GB RAM
- Solid-state drive (SSD) with at least 20 GB of free space
- Full-HD resolution (1920 × 1080 pixels) or higher with 32-bit color depth

Downloading and Installing Tableau Desktop

If you don't already have Tableau Desktop installed on your machine, use this link to download the latest trial version: https://www.tableau.com/products/desktop.

Make sure you are logged in to your machine as administrator and that you have the rights to install software on the machine. Run the installer as you normally would, given your operating system:

On a Windows Machine Open the setup (EXE) file, and accept any safety prompts from your OS.

On a Mac Open the image (DMG) file, and double-click the installation package (PKG) file to start the installation.

Follow the prompts during the setup process. Changes to the installation path or similar changes usually are not required.

Registering and Activating Tableau Desktop

Once the installation process is completed, open Tableau Desktop. A registration form will appear, which you can use to register and activate your Tableau Desktop installation using the product key.

If you do not have a product key for Tableau Desktop yet, you can test it for free for two full weeks. You will be able to use all the features of the software during this trial period.

DATA PREPARATION

Before starting any analysis, one of the key issues to consider is whether the data is in the right format.

Datasets can be structured in different ways. In most cases, datasets have one of two shapes:

- *Wide* tables, with many columns. Often these are summary tables containing aggregated measures (such as pivot tables in Excel). Here, some preprocessing of the data may be necessary.
- *Long* tables, most of the time without aggregations and with each row containing one data point.

Crosstab Reports with Wide Tables

A common mistake made by new Tableau users is attempting to connect to a fully formatted Excel report that already shows data aggregations (see Figure 1.2). Fair enough, since the claim is that you can use Tableau to visualize all sorts of data. But you will quickly realize that this won't get you very far, because you won't be able to create many different types of visualizations.

This scenario is not unusual, and it can be a stumbling block when learning how to analyze data in Tableau. Instead, the recommendation is to work with the unaggregated raw data, if you can get access to it. This will show the items broken down to the smallest units: for example, one data point per row, with each column being a unique field in the data. If that is not possible, follow the recommendations in the next section to prepare your data prior to analysis.

	A	B	C	D	E
1	Temperature Measurements				
2		Month	Morning (6:00)	Noon (12:00)	Evening (18:00)
3	*Seattle*				
4		April	5	17	11
5		May	9	20	15
6		June	12	25	18
7		**Average**	**8.7**	**20.7**	**14.7**
8	*New York*				
9		April	4	12	9
10		May	8	18	13
11		June	11	21	16
12		**Average**	**7.7**	**17**	**12.7**
13	**Average across all**		**8.2**	**18.9**	**13.7**

Figure 1.2 Formatted and aggregated report showing averages across several data points.

Preparing Your Data for Analysis

Using the example presented in the previous section, you can do the following to get the data into better shape:

- Remove the introductory text (Temperature Measurements).
- Put the hierarchical headers (Seattle, New York) in a new, separate column (Location).
- Pivot the data from the wide format, with Morning, Noon, Evening in the headers, to a *long* format, with this information about the time of day in a new column (named Time of Day). Use the full date (and time, if necessary)—for example, 01.04.2018 06:00—instead of just stating the month.
- Ensure that numbers are formatted as such and not as text.
- Remove any summary rows and columns (Average and Average Across All).
- Remove any empty rows.
- Make sure each column has a meaningful heading. For example, ensure that the date column has the heading Time Stamp and the temperature column has the title Temperature.

Long Tables Suitable for Analysis

If you followed the data-preparation steps in the previous section, you have transformed your wide crosstab into a long rows-based table that is ready for analysis in Tableau (see Figure 1.3).

	A	B	C	D
1	Location	Time Stamp	Time of Day	Temperature
2	Seattle	01.04.2018 06:00	Morning	5
3	Seattle	01.04.2018 12:00	Noon	7
4	Seattle	01.04.2018 18:00	Evening	11
5	Seattle	01.05.2018 06:00	Morning	9
6	Seattle	01.05.2018 12:00	Noon	20
7	Seattle	01.05.2018 18:00	Evening	15
8	Seattle	01.06.2018 06:00	Morning	12
9	Seattle	01.06.2018 12:00	Noon	15
10	Seattle	01.06.2018 18:00	Evening	18
11	New York	01.04.2018 06:00	Morning	4
12	New York	01.04.2018 12:00	Noon	12
13	New York	01.04.2018 18:00	Evening	9
14	New York	01.05.2018 06:00	Morning	8
15	New York	01.05.2018 12:00	Noon	18
16	New York	01.05.2018 18:00	Evening	13
17	New York	01.06.2018 06:00	Morning	11
18	New York	01.06.2018 12:00	Noon	21
19	New York	01.06.2018 18:00	Evening	16

Figure 1.3 Long table without any aggregations that is suitable for analysis in Tableau.

As you can see, every row contains exactly one temperature record with the exact time stamp. It doesn't contain any aggregations, such as averages. If you are interested in the average value, you can always calculate it in Tableau later. What's more, you can adjust the level of aggregation—the rows across which the average is to be calculated—corresponding to the questions you are asking.

THE SAMPLE DATASET

In the previous section, you saw what requirements a dataset should meet so that you can easily use it in Tableau. From here on, I will be using a sample dataset that can be found in the `Documents` directory in Tableau. It is already in the correct format and contains the sales numbers of a fictitious company called Superstore. The dataset has the following file name: `Sample – Superstore.xls`.

Finding the Dataset

The file location can vary depending on the installed version of Tableau, the operating system, and your language settings. On my Windows 10 machine, for instance, it is

```
C:\Users\<User name>\Documents\My Tableau Repository\Datasources\2019.1\en_US-EU
```

In Windows Explorer, it looks like Figure 1.4.

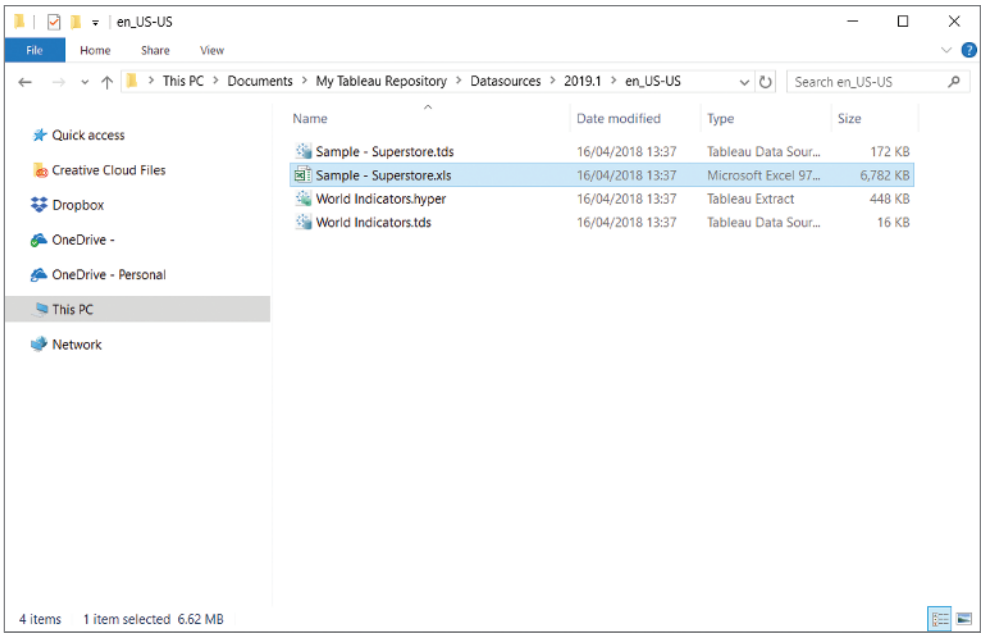

Figure 1.4 Folder with the sample dataset in Windows Explorer.

NOTE You might have noticed that Windows Explorer displays the file location differently (starting with `This PC`, instead of `C:\`).

Understanding the Data

Taking a brief glance at the data in Excel, you can see from the tabs at the bottom that the file contains three different sheets: Orders, People, and Returns (see Figure 1.5). While this dataset is a fictitious example, you could easily imagine seeing this type of sales data in a real-world setting.

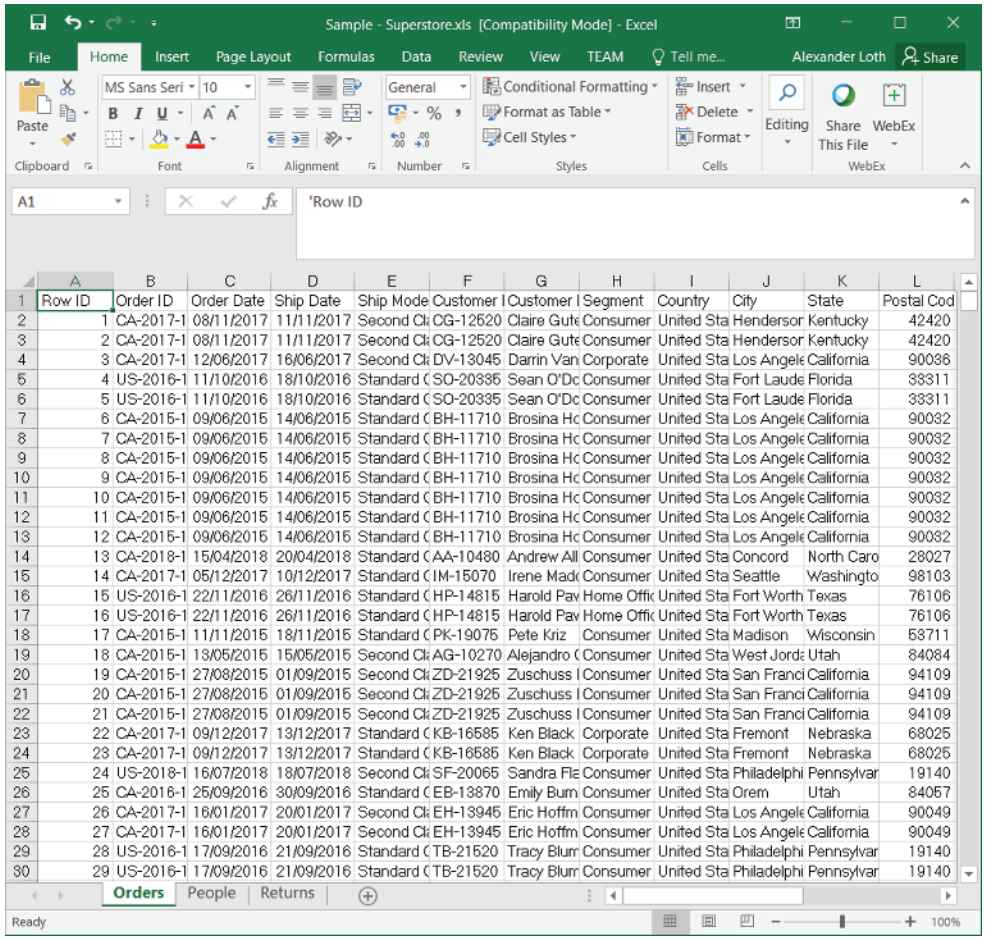

Figure 1.5 Data structure of the sample dataset in Excel.

The three tables are relational—they complement each other—and together provide information about individual sales transactions. In addition, there are no summary aggregations; each row contains one record.

Opening the Excel File Containing the Sample Dataset

When you launch Tableau Desktop, the first thing you will see is the start screen with different data types (see Figure 1.6). In the Connect panel on the left, you can select a file or a server as a data source.

Select the Microsoft Excel option under the To A File heading. Then find the `Sample – Superstore.xls` file in the `Documents` folder. Click Open to use this file as a data source in Tableau.

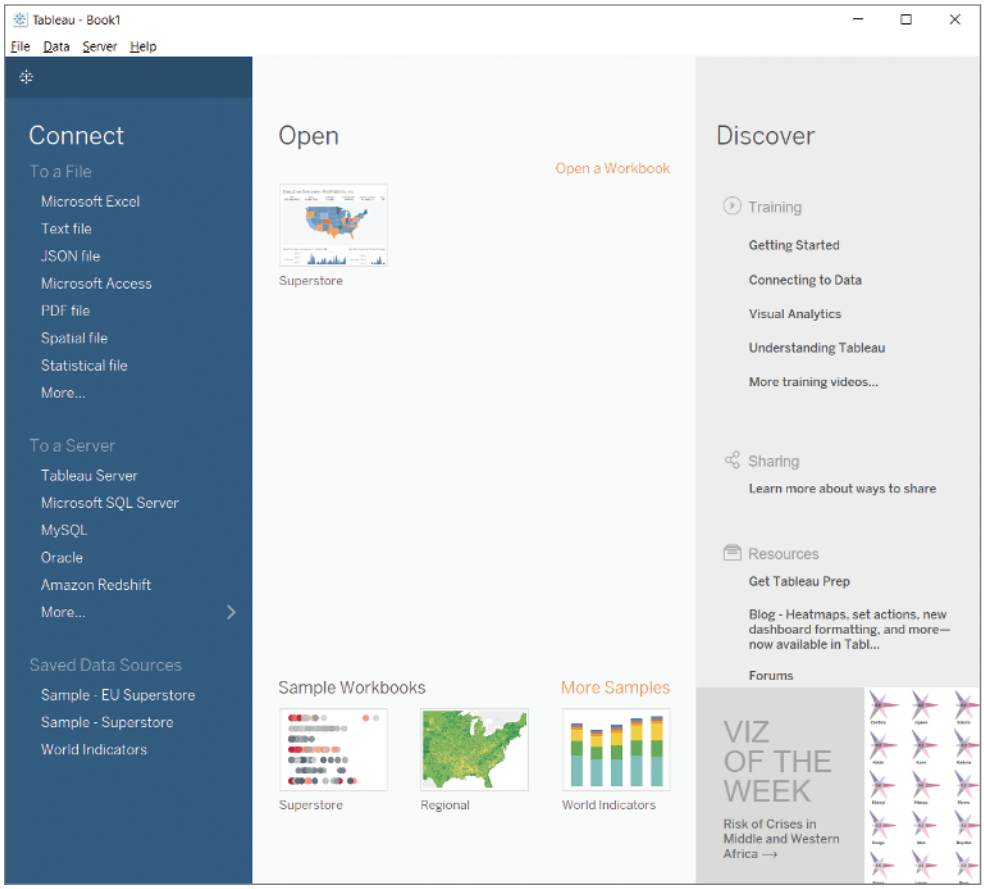

Figure 1.6 Tableau Desktop start screen.

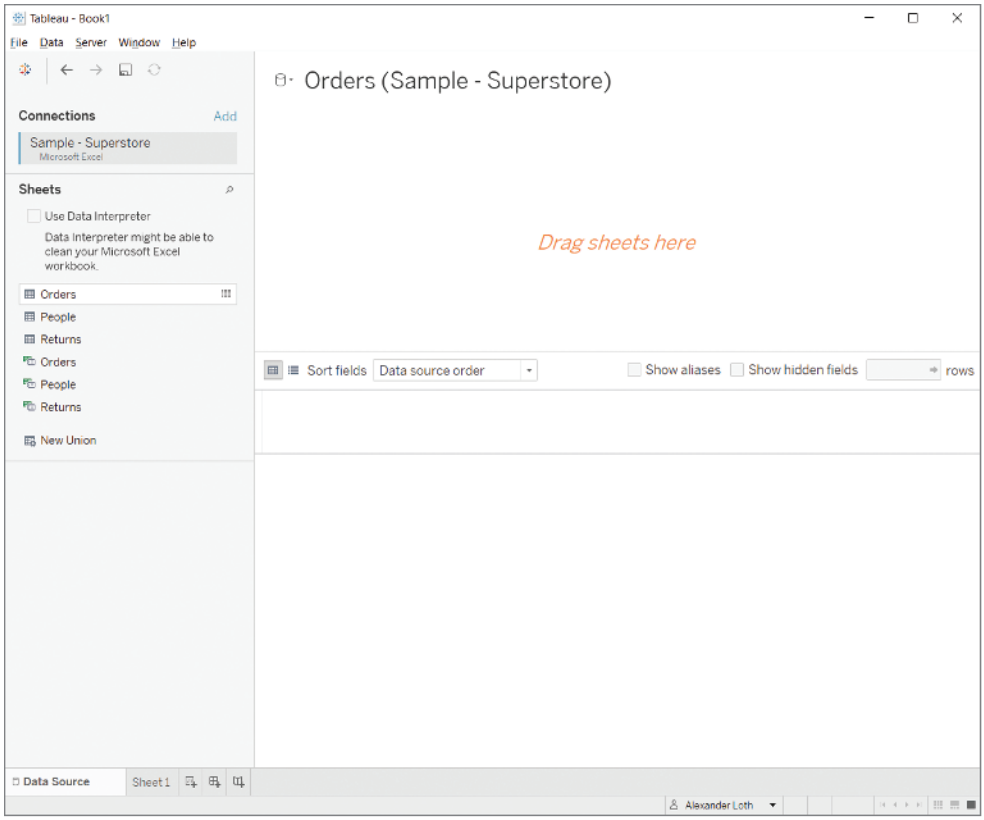

Figure 1.7 The data source view with the Excel file open and the names of the three Excel sheets shown on the left.

Tableau will then switch to the Data Source pane and, on the left-hand side, list the names of the three sheets contained in the Excel file (see Figure 1.7).

Drag and drop the `Orders` sheet onto the white space in the top half of the screen, as shown in Figure 1.8.

In the bottom half of the screen, you will now see a preview of the data. Finally, click Sheet 1 in the tabs bar at the bottom of the window, to create your first Tableau worksheet.

NOTE The examples in the following chapters of this book will all use this Superstore sample dataset, unless stated otherwise.

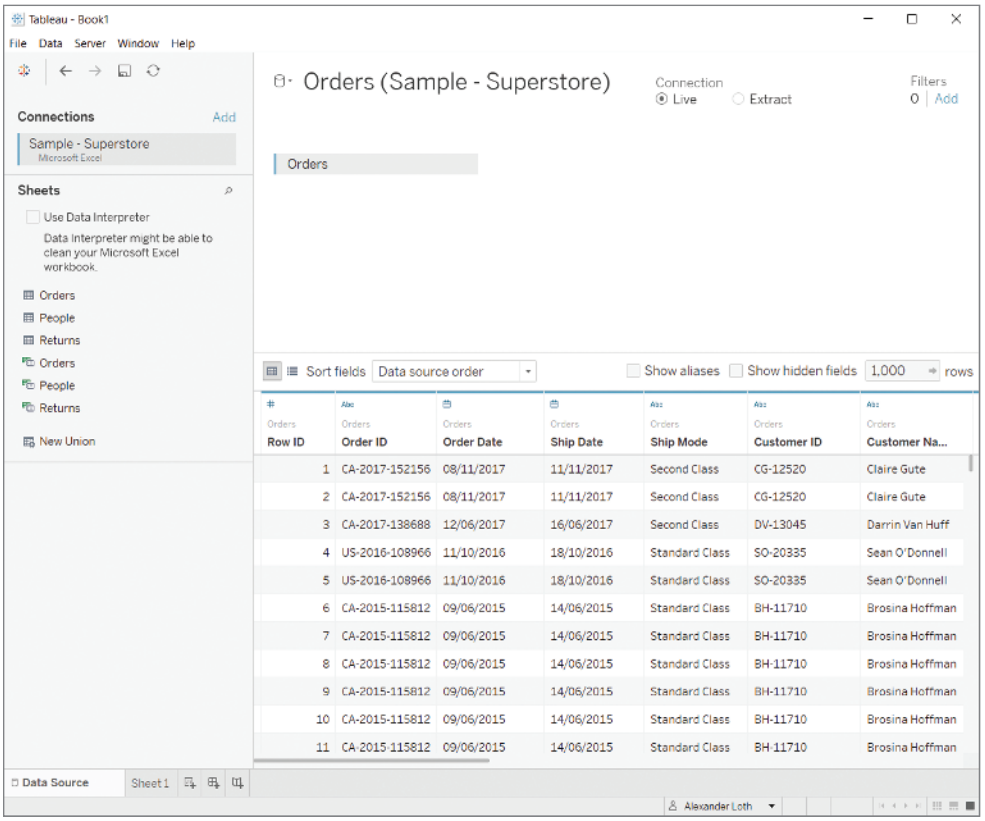

Figure 1.8 Preview of the data from the `Orders` sheet.

THE TABLEAU WORKSPACE

Now that you have connected to a data source and opened your first worksheet, it's time to become familiar with the Tableau interface. Figure 1.9 shows 15 numbered portions of the interface. The still-blank canvas (1), as it is called, includes the title Sheet 1 (2). On the left, you will find the Data pane (3) (more on that in the section "The Data Pane"). The tab next to it opens the Analytics pane (see Chapter 7 for more details).

In Tableau, most interactions are achieved by dragging and dropping items onto the canvas. This makes interacting with Tableau easy and intuitive. Both dimensions (4) (including hierarchical dimensions [5]) and measures (6) can be moved directly onto it. Alternatively, they can be placed on the Columns (7) and Rows (8) shelves, in order to add them to your visualization.

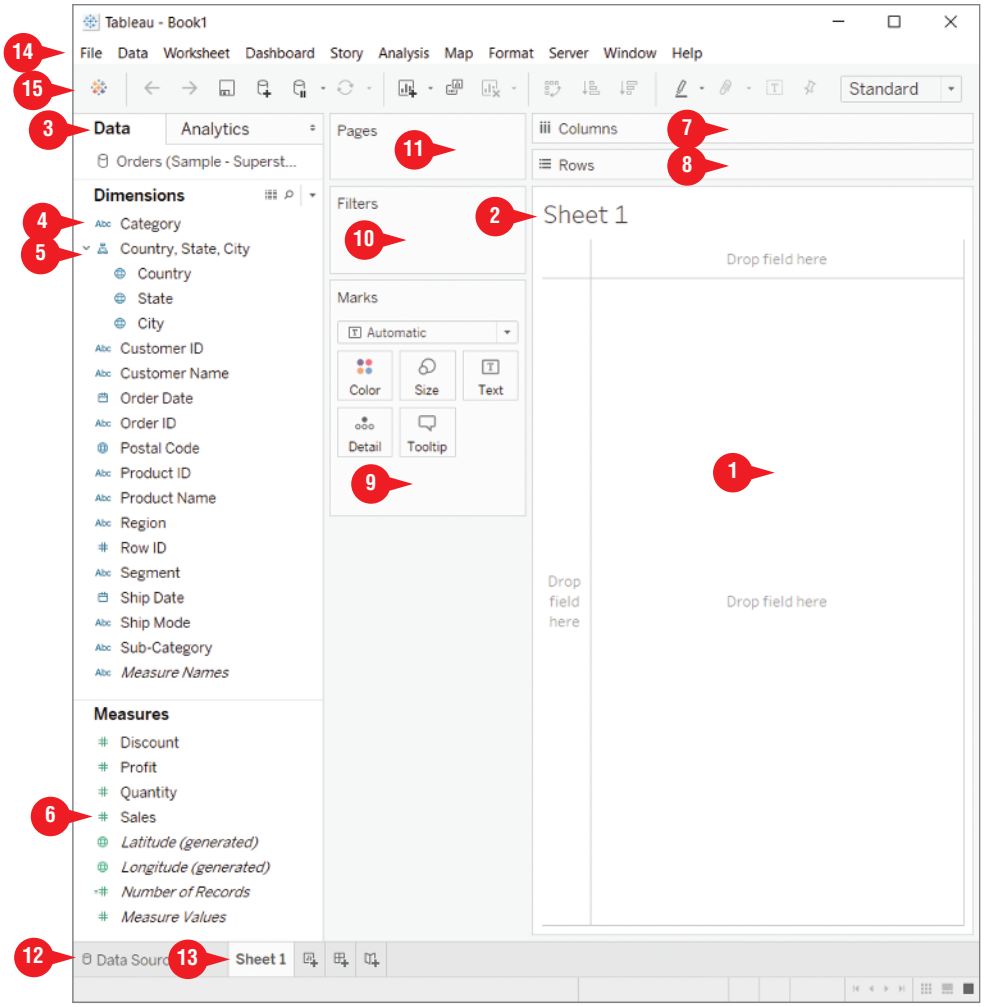

Figure 1.9 The Tableau workspace with a yet-to-be-filled canvas.

Fields from the Data pane can also be placed onto the Marks (9), Filters (10), and Pages (11) cards: for example, to change the color of marks or to only display marks for a filtered-out subset of the data. We will take a closer look at many combinations of these throughout the remainder of the book.

The tabs bar at the bottom of the screen allows you to go back to the data source editor (12) and to toggle between your different worksheets (13), each containing a single visualization. With the three buttons to the right of the tabs, you can open additional worksheets, new dashboards, and stories, respectively.

At the top of the screen, you can find the menu bar (14) (discussed in more detail in the section "The Menu Bar"). And directly under that is the toolbar (15), with three important buttons:

- **The Tableau icon:** This brings you back to the start screen, where, among other things, you can add additional data sources.
- **Undo:** This allows you to go back a step so you can safely try out different ideas. You can go back as many steps as you like.
- **Redo:** This allows you to restore any undone actions.

The Menu Bar

Even though most of the work in Tableau can be achieved by directly interacting with items using the mouse, there is also a menu bar at the top that lets you access additional features and settings. Let's take a closer look at some of the particularly useful entries:

File Menu The File menu contains the key functions Open, Save, and Save As (see also the section "Saving, Opening, and Sharing Your Workbooks"). The Print To PDF menu item allows you to export your worksheets and dashboards as PDF files. With the Repository Location option, you can look up and change the default location for Tableau files on your machine. With Export As Version, you can create workbooks for colleagues who might still be using an older version of Tableau Desktop.

Data Menu Here, the Insert function is especially interesting, as it presents a quick, ad hoc way to add a data table—for example, from a website. Simply select and copy the table in the original document, and click Insert in Tableau. This will add the data to your workbook as a new data source. (You will learn more about working with data sources in Chapter 2.)

Worksheet Menu With Export, you can take your data out of Tableau by creating an image, a database file, or an Excel crosstab. Duplicate As Crosstab, on the other hand, opens a new worksheet in Tableau, showing a crosstab view of the data used in your visualization.

Dashboard Menu Dashboard actions that add interactivity to dashboards are set up and tweaked by clicking Actions. You will learn more about filter, highlight, and URL actions in Chapter 8.

Story Menu The Story menu entry lets you create a story from your worksheets and dashboards. In a story, content is arranged sequentially for presentation and enriched with annotations.

Analysis Menu With this menu, you can create and edit calculated fields (see Chapter 4). Here, you will also find options for tweaking table layouts as well as for showing grand totals, forecasts, and trend lines (see Chapter 7).

Map menu In the Map menu, you can choose between different background maps. The Offline option is particularly useful when you have no Internet connection and would like to access the built-in cartographic material. You can find more on maps in Chapter 6.

Format Menu In this menu, you can set the font, alignment, shading, and other formatting options. In addition, you can set the overall workbook design and adjust the cell size.

Server Menu Use this menu for sharing your dashboard via Tableau Online, Tableau Server, or Tableau Public (more on these options in Chapter 9). With the Create User Filter submenu, you can set audience-specific filters that grant specific users or user groups (which have been defined in Tableau Online or Tableau Server) access to selected subsets of the data.

Window Menu Use the Presentation Mode option to use the full screen for your dashboard.

Help Menu Via this menu, you have access to the Tableau online help, training videos, and sample workbooks. Use the Start Performance Recording option in the Settings And Performance submenu to analyze the processing time of your dashboard.

The Data Pane

The Data pane is divided into measures and dimensions. You control what visualizations you want to display by adding different combinations of measures and dimensions to the canvas.

Measures

Measures are numeric variables. By adding a measure to the view, you decide which values from your dataset to visualize. By default, Tableau automatically applies an aggregation function such as `SUM` or `AVG` (the arithmetic mean) to measures. That way, you can, for instance, show the sum or the average of a sales discount across different transactions.

Measures typically (but not always) come with green symbols, which represent continuous variables.

Dimensions

Dimensions are descriptive, categorical variables. With dimensions, you can decide how to group the aggregated values of the used measures. For instance, the sum of sales revenue (a measure) could be broken down by country, product category, or both (i.e. two different dimensions).

Typically, dimensions come with blue symbols in Tableau, which represent discrete variables.

TIP If Tableau has erroneously added a measure to the Dimensions section of the Data pane, you can simply drag it into the Measures section with the mouse, and vice versa.

WORKING WITH MEASURES AND DIMENSIONS

Now that you are familiar with the Tableau interface and know where to find measures and dimensions in the Data pane, you can finally get started with your first data visualization!

Visualizing a First Measure

Often it makes sense to start a visualization by adding a measure to the view. In the Superstore workbook, sales revenue is a good candidate.

To do this, drag the Sales measure onto the left side of the canvas, to the vertical area labeled Drop Field Here. As you can see in Figure 1.10, the area will be highlighted in green when you move a measure there.

Alternatively, you can drop the measure onto the Rows shelf above the canvas. The result is the same: you see the total sales revenue of all the records in the dataset.

TIP Do you want to see things even faster? In addition to using drag-and-drop, you can just double-click measures and dimensions in the Data pane. It does take a little getting used to, though, as Tableau automatically decides where to place the fields on the canvas, depending on the context.

Breaking Down a Measure Based on a Dimension

Having visualized your first measure, you can now break it down using one of the dimensions. You might want to learn, for example, how the sales numbers break down by product category.

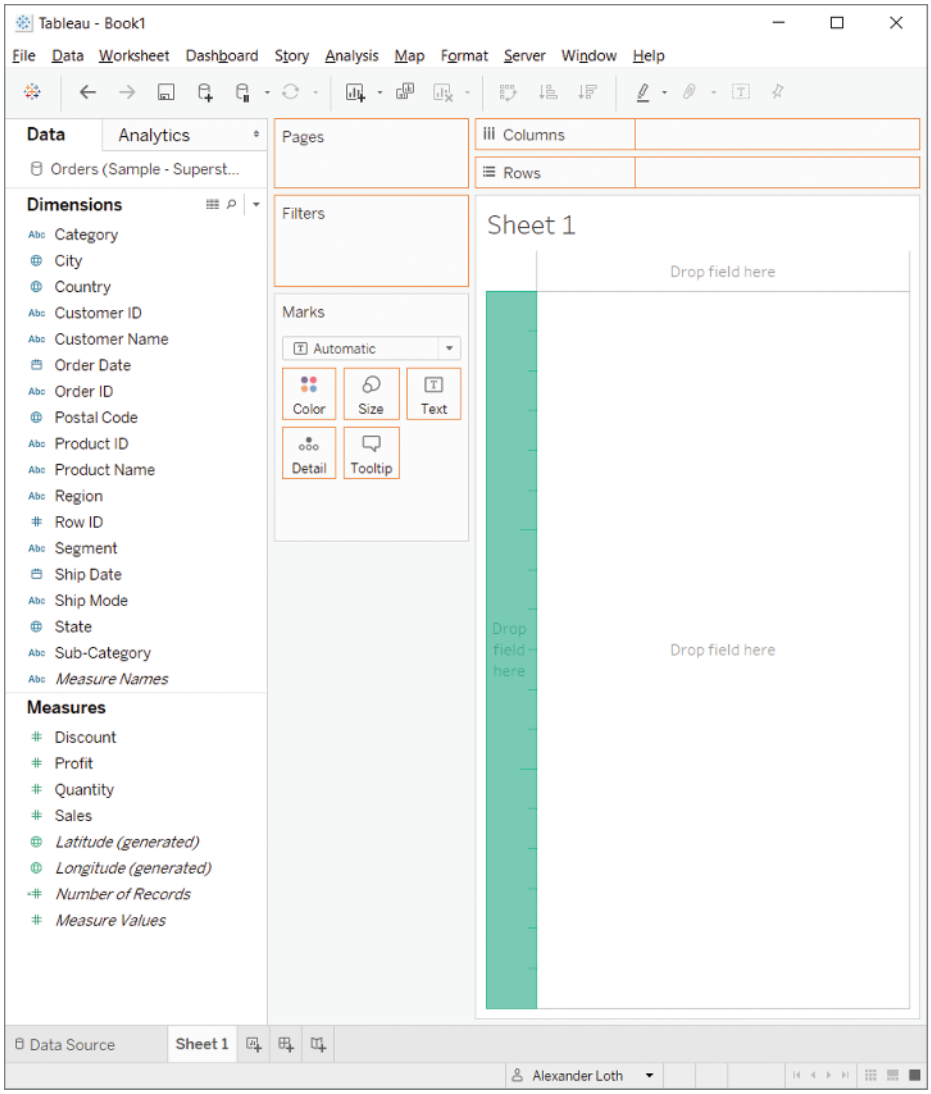

Figure 1.10 A green highlight appears when you drag a measure onto this section of the canvas.

To answer that question, hold down the mouse button and drag the Category dimension onto the upper edge of the bar chart. As you can see in Figure 1.11, a blue highlight appears as you move the dimension there.

Alternatively, you can place the dimension onto the Columns shelf above the canvas. Again, the result is the same: you see the total sales revenue broken down by product category, in accordance with the dataset.

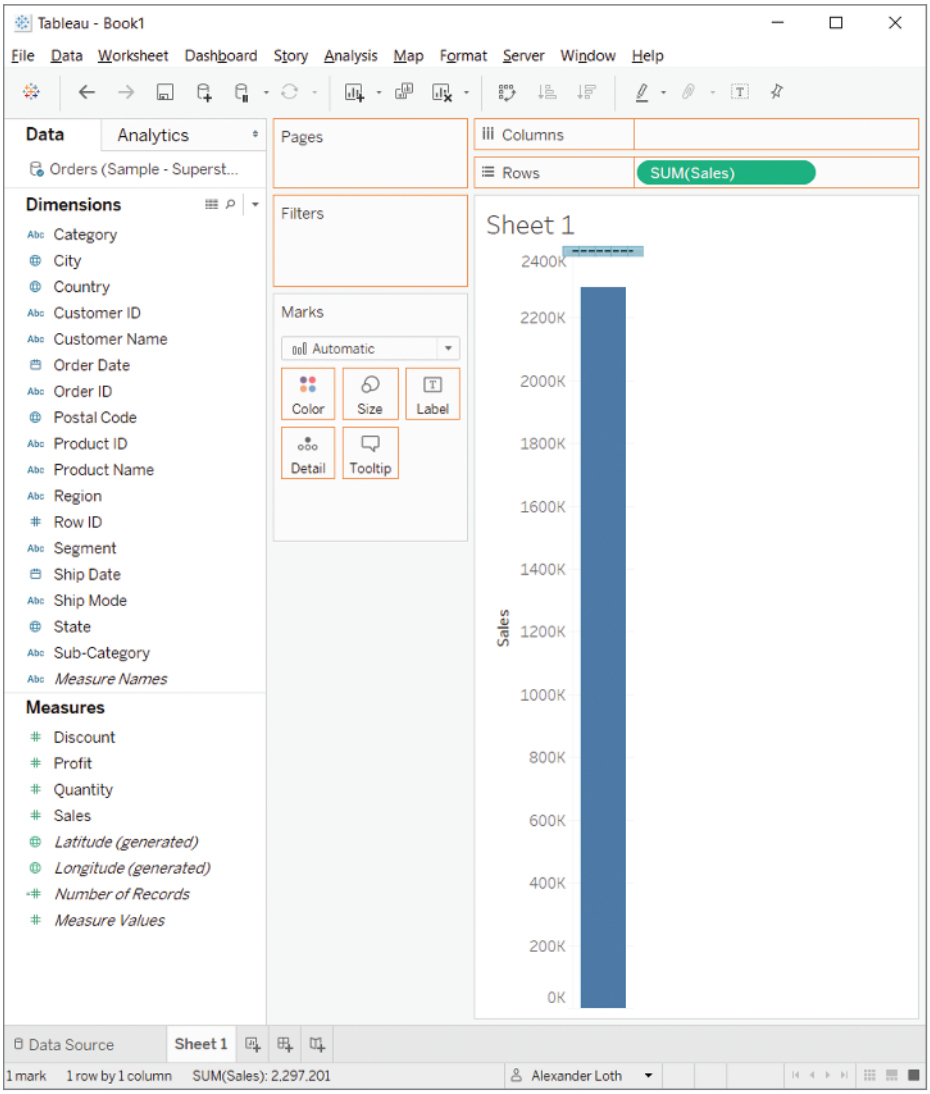

Figure 1.11 A blue highlight appears when you drag a dimension onto this part of the canvas.

WORKING WITH MARKS

After the Rows and Columns shelves, the next-most-important area is the Marks card with its various symbols. You can add dimensions and measures here, too, to further style your chart.

Among other things, you can control the color, size, form, and labeling of the marks displayed in your visualization.

Working with Color

Probably the most-used feature of the Marks card is the Color field. To try this out, drag the Segment dimension onto Color, as demonstrated in Figure 1.12.

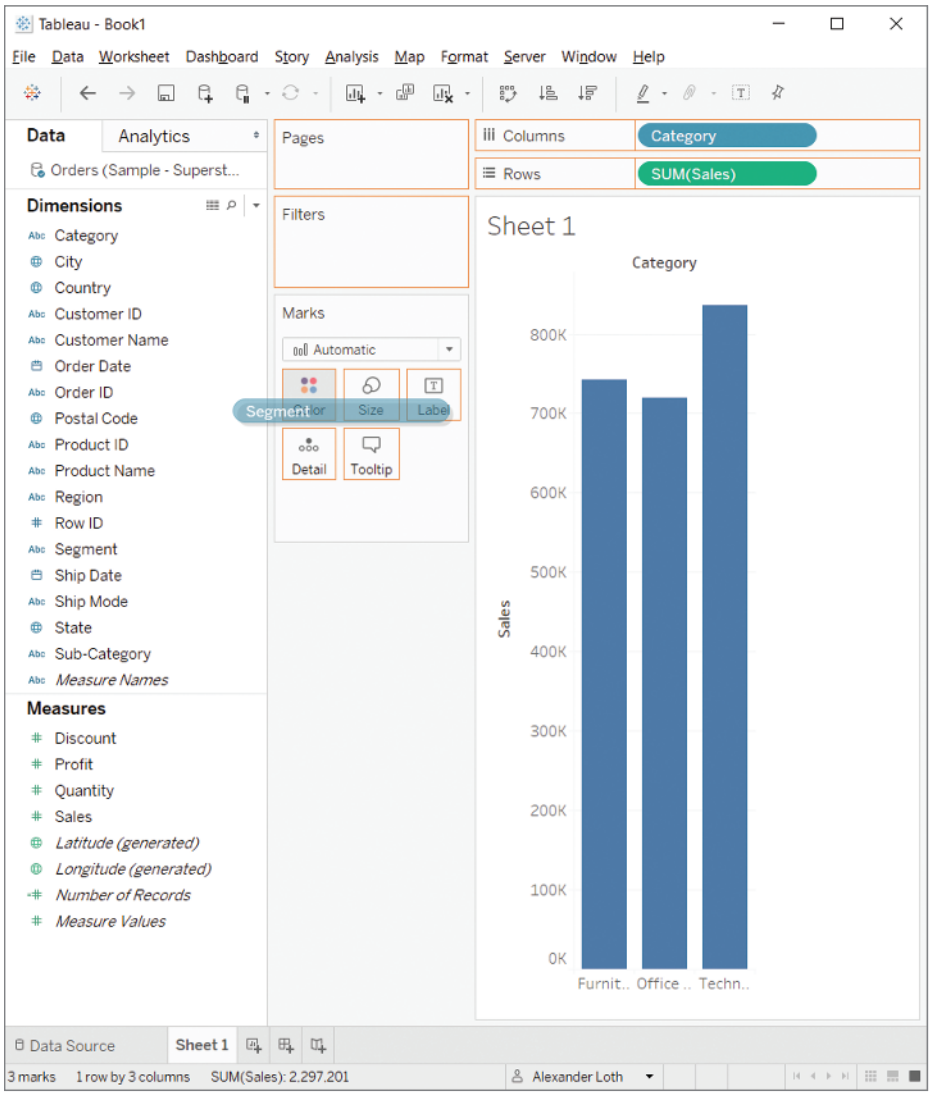

Figure 1.12 Dragging the Segment dimension onto Color.

TIP By clicking the Color field, you can choose a different color palette or, when a dimension is used, manually assign different colors to the individual items. When you place a measure onto Color, instead, you can select the colors and intervals of the color gradient.

Adding More Information to Tooltips

A *tooltip* is a little hover box that displays additional information when you point at individual marks in the visualization (see Figure 1.13). This makes your charts interactive—allowing the end user to "touch" the data—and thereby creates real added value versus static charts and PDF reports.

Drag the Profit measure onto the Tooltip field of the Marks card. Now inspect the result: view the tooltips by moving the mouse pointer over different marks in the visualization.

Tooltips add interactivity to your charts without making them more complex. They are a good place to keep secondary information that is not necessary at first glance.

TIP You can add as many measures and dimensions to a tooltip as you like. A simple click on the Tooltip field will open a text editor that allows you to change the appearance of the tooltip as required.

SAVING, OPENING, AND SHARING YOUR WORKBOOKS

Congratulations! If you have followed the previous steps, you have created your first data visualization in Tableau Desktop. Now it is time to save your work, and—if so desired—to share it with others. How to make your work available to various audiences via the different Tableau platforms (Tableau Online, Tableau Server, and Tableau Public) will be covered in Chapter 9. For now, you will save the file locally.

Saving Workbooks

Open the File menu, and click Save As. There are two different file types to choose from in the dialog box that opens:

Tableau Workbook (*.twb) Tableau workbooks contain all the visualizations as well as the metadata. They do not, however, contain the actual data. When you share a Tableau workbook, the recipient will need to have access to the original file or database that you used.

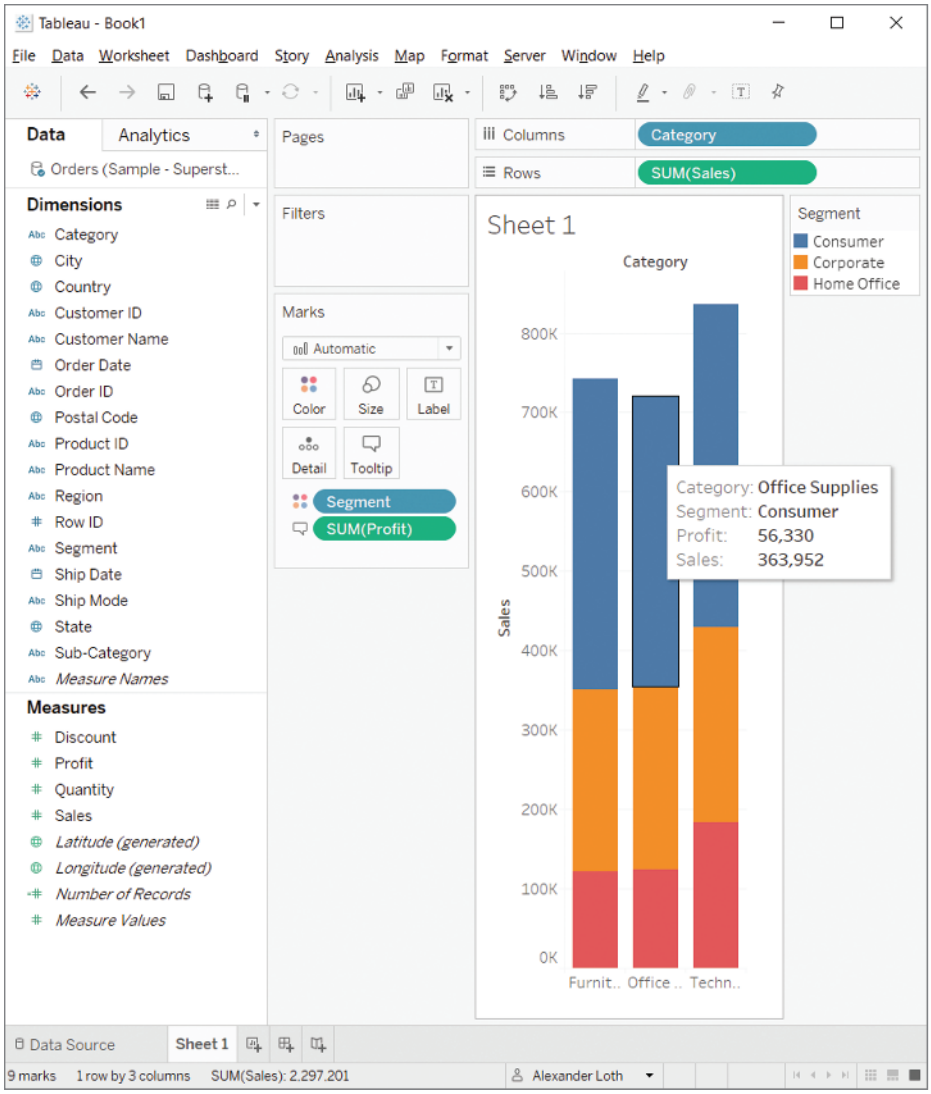

Figure 1.13 Tooltips are displayed when you point at marks.

Tableau Packaged Workbook (*.twbx) A Tableau packaged workbook, on the other hand, contains the actual data in addition to the visualizations and metadata. The data is greatly compressed, thereby reducing the overall file size. When you share a Tableau packaged workbook, the recipient will be able to open and work with your visualizations even without having access to the original data source.

Choose one of the two file types, give the file a name, and click Save.

TIP Are you on the road a lot and frequently find yourself without Wi-Fi? Make a habit of saving your work as Tableau packaged workbooks (`.twbx` files) so you always have your data extracts with you, even if you can't reach your database servers remotely.

Opening Workbooks

To open a previously saved workbook or packaged workbook, go to File and click Open.

Sharing Workbooks with Tableau Reader

Should you wish to share your Tableau packaged workbook with friends, acquaintances, or colleagues who do not have a Tableau Desktop license, they can open the file with the free Tableau Reader. Tableau Reader can be downloaded for Windows and Mac from the following website: https://www.tableau.com/products/reader.

NOTE In Tableau Reader, the interactivity of your visualizations is fully retained. Tooltips, for example, can be displayed as intended. However, neither the data nor the visualizations can be edited or saved in Tableau Reader.