CHAPTER

1

Dashboard Basics

In This Chapter

- Determining When to Use a Dashboard
- Establishing User Requirements
- Assembling the Data
- Building the Dashboard
- Formatting the Dashboard

Dashboards have never been more popular. We have more data available to us all the time and better visualization tools than ever before. At its core, a dashboard is a collection of charts. But it's much more than that. If you put some charts on a page, you would technically have a dashboard, but perhaps not a very good one. Creating a good dashboard takes some preparation, knowledge, and skill. In this chapter, I introduce you to dashboards and the concepts, skills, and best practices you'll need to create them.

Determining When to Use a Dashboard

Dashboards are used to present data. Data can be thought to be at various stages: raw, aggregated, analyzed, and presented. The stage your data is in depends on where it comes from and what you plan to do with it. There are many levels of aggregation and an infinite number of ways to analyze or present data. For example, an invoice is an aggregation of invoice lines and a sales report is an aggregation of invoices. Relative to an invoice, the invoice lines are raw data but relative to the sales report, the invoices are the raw data. Figure 1.1 shows data in its various stages.

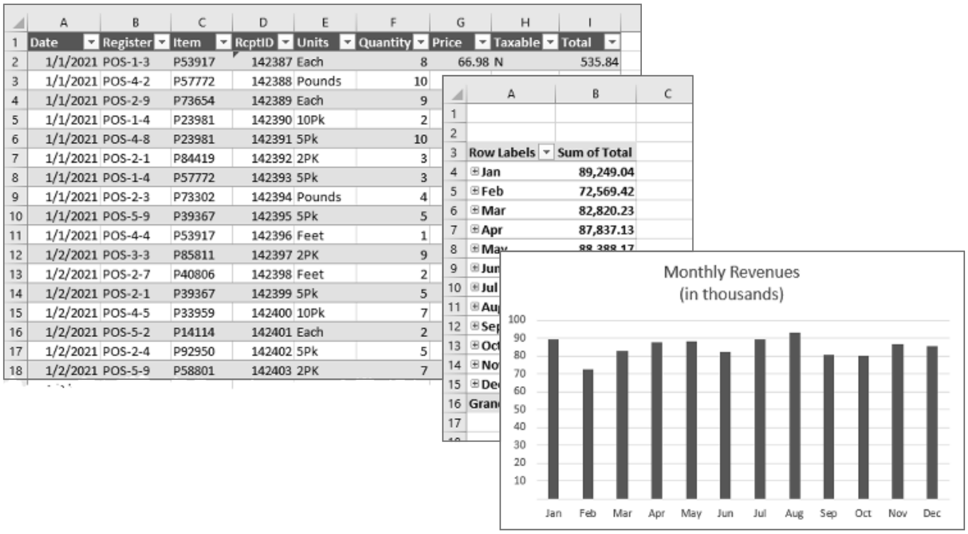


Figure 1.1: Data shown raw, aggregated, and analyzed and presented

Raw data is data that hasn't been processed. It can be transactions that come out of an accounting system, sales information from a point of sale, or readings from a measuring device like tank levels or temperatures. If you're starting with raw data, you will have to do some aggregating and possibly some analyzing before it's ready for a dashboard.

NOTE **A workbook containing the charts in the figures for this chapter is named `Chapter1Figures.xlsx` and can be found on this book's companion website at `www.wiley.com/go/datavizwithexcel/`.**

Aggregated data has been grouped and summarized in some way. A report of units produced by month sums the units produced each week or each day. And that may be a sum of units produced by shifts for a day. In many cases, dashboard builders start with aggregated data.

Dashboards tell a story about the underlying data. Analyzing data is determining what stories the data tells and which of those stories is worth telling. Analyzing is more than just drawing conclusions from the data. It's also understanding the nature of the data and what questions the data raises. It's common during data analysis to have to take a step back and aggregate the data in a different way.

Finally, there's the presentation stage, where dashboards live. The dashboard building process can start at any stage. If you get the source data from a data analyst, the story to tell may have already been determined and it's just a matter of presenting that story in an effective way. Conversely, if you start with raw data, you'll need to first aggregate and then analyze the data to make those determinations.

Dashboards are constantly evolving. At one time they were only static visuals telling one story. Now, dashboards include self-service business intelligence (BI) tools that either tell multiple stories or allow the users to find the meaning in the data themselves. With Microsoft's Power BI tool and its integration into Excel with Power Pivot and Power Query, self-service BI is becoming more mainstream and accessible.

CROSS-REFERENCE **Power Pivot, Power Query, and Power BI are introduced in Chapter 3, "Organizing Data for Dashboards."**

What Is a Dashboard?

A dashboard is one or more visual elements that tell a story about related data. A report that aggregates data isn't a dashboard because it's not telling a story. That's typically called a report or table, although these terms are often used to mean the same thing. For our purposes, a dashboard must contain visual elements and not just a list of data.

The story is the most important aspect of a dashboard. It comes from analyzing the data to determine what's important about it. Key performance indicators (KPIs) are commonly displayed on dashboards. KPIs are ready-made stories for your dashboard to tell. I briefly discuss KPIs in the next section. A common pitfall in dashboard building is to start with a conclusion. The person requesting the dashboard may have an agenda or preconceived notion of what that data should say. But the data should drive the story, not the other way around. Try to reframe the conclusion as a question. If someone wants you to create a dashboard that shows that sales decreased because of bad weather, you can turn that into a question like "How does average daily temperature correlate with daily sales?" or "How much do we sell on rainy days vs. sunny days?"

The underlying data on a dashboard is related, but how it's related depends on who's looking at it. A member of the Human Resources department's dashboard might use data related to employee retention like hiring rate, firing rate, layoffs, voluntary terminations, and retirements. The human resources manager may have a dashboard that's a level above, such as more aggregated employee retention data along with payroll costs and benefit engagement. The person in charge of all administration in a company would look at human resources data next to finance, accounting, and legal data. At the top level of a company, data from administration, operations, and research and development is related.

Key Performance Indicators

How KPIs are determined and what makes a good one is well beyond the scope of this book. An organization's leaders will develop KPIs based on what they know about the organization. If you're running a for-profit business, net income is an important measurement and you don't have to analyze the data to know that it's something you'll want to look at. KPIs are unique to each organization, but similar organizations will have similar KPIs. Finance departments are interested in net income, free cash flow, and working capital. And manufacturers are interested in units produced and line utilization.

Establishing User Requirements

Don't start building a dashboard until you have a plan. Just like building a house, if you start without a plan, you may have to tear it down and start over. To make your plan, start by finding out what the end users need. There are at least three users you'll want to talk to before you begin: the person requesting the dashboard, the person providing the data, and the end user. All these users may be one person, and that person may even be you.

Get as much detail as you can from the person requesting the dashboard. If they have a general idea of what they want, now is the time to probe for details to get a clear picture. As I mentioned in the previous section, the requester might be starting with a conclusion in mind. Try to turn that conclusion into a question or series of questions so you're on the same page.

Questions about the source data are sometimes overlooked but shouldn't be. Find out where the data is coming from and if it's already been aggregated or analyzed. Depending on your project, you may want to try to get the data in as raw a form as possible in case you have to change direction once you get started. It's a lot easier to aggregate raw data in a different way but almost impossible to disaggregate it.

Determine if the data is coming from inside or outside of the organization, who maintains it, and how often it's updated. Financial data from an accounting system may only be available monthly or quarterly. Other types of data, like data from a point of sale, may be able to be queried in real time.

If you don't have the data you need, your dashboard project might turn into two projects: a data-collection project and a dashboard project. You may find that not only is the data not readily available, it doesn't exist at all. If the organization doesn't track defects from the production line, there may be no way to get historical data. In that case, you could set up a system to start tracking the data you need, which would delay how quickly a dashboard could be created. Having this conversation early in the project helps set the expectations of all the stakeholders.

Types of End Users

You can divide end users by how they intend to use the dashboard to get a better understanding of how to construct it. Monitors use dashboards to see the state of an organization or project at a given time. You use your car's dashboard to monitor speed, fuel levels, and trouble alerts. Deciders use dashboards to determine if they should take one action, another action, or no action at all. A production manager might use a dashboard of sales and line utilization data to determine if a third shift is necessary.

Planners are people at the highest levels of an organization that determine the direction of the organization. They are looking at broader trends, and the actions they take are more policy based. Planners might look at operating results by division to determine how to allocate resources for the next five years. Presenters use dashboards to present information. A dashboard presented at a shareholder meeting may be used to simply give shareholders information they don't see day to day.

There is a lot of overlap in these categories of users. Someone monitoring a project, for example, will certainly take action if the information dictates it. And shareholders might change the direction of a company by changing management if they don't like the information presented. Know your audience, and in some cases your audience's audience, so you can provide the right level of information.

Determine how often the dashboard will be created. If you're creating a dashboard to show the effects of the Olympics on a city's finances, you'll probably only do that one time. For one-off dashboards, you don't have to be as concerned about data maintenance and how efficiently you can build it.

For periodic or real-time dashboards, make sure the data availability lines up with how often you will be publishing. Also, invest more of your time in automation for dashboards you will be publishing more. Real-time dashboards have to be fully automated. Dashboards you publish annually can be less so.

Document the dashboard-building process from the start. You don't need fancy software for this, just a text document or a spreadsheet. Record the information about the users and the data that you've already discovered. Also document how the data moves through the process from raw, possibly through a database, and into a spreadsheet. Even if it's a dashboard that will be created often, or fully automated, document the data flow so you or someone else can troubleshoot problems when they arise. Imagine yourself re-creating the dashboard a year from now, after you've forgotten all the details, and try to provide answers to the questions you would have.

Plan on iterating through the dashboard design process. When you get a usable draft, send it to the stakeholders for input. Continue to send iterations for input throughout the process. You can continue to work on the next stage while you wait for input, but if there are problems, you'll save yourself a lot of rework compared to handing over a final design that's not right.

Finally, plan to review the results after the dashboard is complete. If you email a dashboard weekly, set up a reminder to check back with the end users in a few months to make sure it's still meeting their needs. This is particularly important if you are spending time creating the ongoing dashboard. In one case, I was creating a monthly dashboard and the publishing mechanism had failed. I didn't realize that it didn't get published for several weeks, but nobody asked where it was. It turned out that the needs of the end users had changed, and we stopped publishing that dashboard, saving time every month.

Assembling the Data

For Excel users, working with data is the best part of dashboard building. As you are no doubt aware, Excel provides great tools for data aggregation and manipulation, including hundreds of formulas and data tables and tools for sorting data, splitting data, and removing duplicates. In the following sections, I discuss some of Excel's tools for manipulating data.

REFERENCE **If the data driving your dashboard originates outside Excel, the first task will be to import the data. I discuss working with external data in Chapter 3.**

PivotTables

One of Excel's most powerful tools is the *PivotTable*. A PivotTable is a report that filters, sorts, and summarizes your data based on conditions that you set. PivotTables are interactive in that you can drag and drop fields into the appropriate PivotTable Fields areas to quickly change how your data is summarized.

REFERENCE **Charts made from PivotTables are called PivotCharts. PivotCharts have interactive functions that other charts don't have. I discuss PivotCharts in more detail in Chapter 9, "Visualizing Changes over Time."**

Anything you can do with a PivotTable you can do with worksheet formulas. A PivotTable will summarize your data in much less time than it takes to write the formulas to do the same thing. If you want to make changes to a PivotTable by adding, removing, or rearranging fields, it updates in a fraction of a second. Rewriting formulas for changes takes much longer. Figure 1.2 shows a portion of some raw data and Figure 1.3 shows one possible PivotTable you can create from the data.

	A	B	C	D	E
1	State	Region	Month	Qtr	Sales
2	California	West	Jan	Qtr-1	1,118
3	California	West	Feb	Qtr-1	1,960
4	California	West	Mar	Qtr-1	1,252
5	California	West	Apr	Qtr-2	1,271
6	California	West	May	Qtr-2	1,557
7	California	West	Jun	Qtr-2	1,679
8	Washington	West	Jan	Qtr-1	1,247
9	Washington	West	Feb	Qtr-1	1,238
10	Washington	West	Mar	Qtr-1	1,028
11	Washington	West	Apr	Qtr-2	1,345
12	Washington	West	May	Qtr-2	1,784
13	Washington	West	Jun	Qtr-2	1,574
14	Oregon	West	Jan	Qtr-1	1,460
15	Oregon	West	Feb	Qtr-1	1,954
16	Oregon	West	Mar	Qtr-1	1,726
17	Oregon	West	Apr	Qtr-2	1,461
18	Oregon	West	May	Qtr-2	1,764
19	Oregon	West	Jun	Qtr-2	1,144

Figure 1.2: Raw data for a PivotTable

Sum of Sales	Column Labels		
Row Labels	Qtr-1	Qtr-2	Grand Total
Central	**21,025**	**22,176**	**43,201**
Illinois	4,243	3,623	7,866
Kansas	4,776	4,007	8,783
Kentucky	3,342	4,626	7,968
Missouri	4,669	4,556	9,225
Oklahoma	3,995	5,364	9,359
East	**17,267**	**17,864**	**35,131**
Florida	4,722	4,630	9,352
Massachusetts	3,442	4,155	7,597
New Jersey	4,365	4,316	8,681
New York	4,738	4,763	9,501
West	**16,778**	**18,242**	**35,020**
Arizona	3,795	4,663	8,458
California	4,330	4,507	8,837
Oregon	5,140	4,369	9,509
Washington	3,513	4,703	8,216
Grand Total	**55,070**	**58,282**	**113,352**

Figure 1.3: A PivotTable

The PivotTable in Figure 1.3 was made by dragging the Qtr field to the Columns area, the Region and State fields to the Rows area, and the Sales field to the Values area. By default, Excel sums the Sales field because its data is numeric. For non-numeric data, Excel defaults to counting the data. Figure 1.4 shows the PivotTable Fields task pane with these conditions.

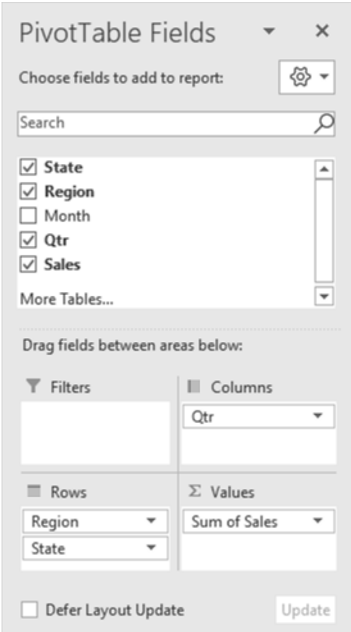


Figure 1.4: The PivotTable Fields task pane

The areas of the PivotTable Fields task pane are as follows:

- **Tools (represented by the gear icon)**: This section allows you to change the layout of the task pane and sort the field names. If you have a lot of fields, you may prefer a layout that shows more fields.
- **Search**: By typing in this text box, you limit the field list to only those fields that match or partially match the text you type.
- **Filters**: If you drag a field into this area, Excel places a drop-down box above your PivotTable. You can select one or more values from the drop-down to limit what data the PivotTable displays.
- **Columns**: Fields in this area are displayed across the top of the PivotTable. You can put multiple fields in the Columns area to nest them. That is, fields that are higher up in the Columns area will span multiple columns and the cells below it will only show values that are related to the higher fields. Figure 1.5 shows the Month field nested below the Qtr field. The values Jan, Feb, and Mar are the only values related to Qtr-1, so they're the only values shown nested beneath it.

	Column Labels								
	⊟Qtr-1			Qtr-1 Total	⊟Qtr-2			Qtr-2 Total	Grand Total
	Jan	Feb	Mar		Apr	May	Jun		
Sum of Sales	18,579	19,194	17,297	55,070	20,754	18,771	18,757	58,282	113,352

Figure 1.5: Nested fields in the Columns area

- **Rows**: Fields in this area are displayed down the left side of the PivotTable. As with the Columns area, you can nest fields to display them in a hierarchy. The PivotTable shown in Figure 1.3 shows the State field nested beneath the Region field.
- **Values**: Fields in this area are aggregated in the body of the PivotTable. Sum and Count are the most commonly used aggregators. But there are several more ways to aggregate, including Average, Min, and Max. The aggregation occurs at the intersection of the Column and Row fields. For example, where Qtr-1 and Missouri intersect, the PivotTable in Figure 1.3 sums the sales for only those rows that contain a State of Missouri and a Qtr of Qtr-1.

Excel adds grand totals and, if there are nested fields in the Columns or Rows areas, subtotals. These totals can be shown or hidden depending on how you want to display the data. To turn off the grand totals, select any part of the PivotTable and choose Options from the PivotTable Tools Analyze tab of the Ribbon to show the PivotTable Options dialog box shown in Figure 1.6. Uncheck the check boxes labeled Show Grand Totals For Rows and Show Grand Totals For Columns.

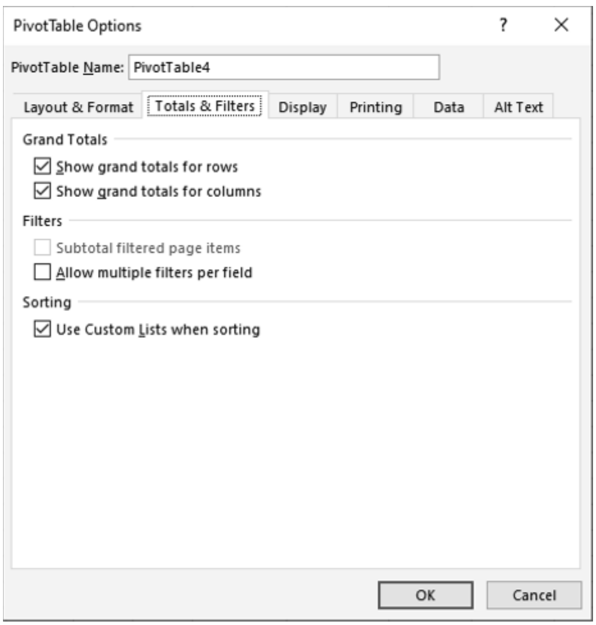


Figure 1.6: The PivotTable Options dialog box

To hide the subtotals for nested fields, right-click one of the values in the field and choose Field Settings from the context menu. On the Field Settings dialog box, shown in Figure 1.7, choose one of the following option buttons.

- **Automatic:** Displays subtotals with the same aggregation (sum, count, etc.) as the child field.
- **None:** Hides subtotals for the field.
- **Custom:** Displays subtotals with the same or a different aggregation than the Child field.

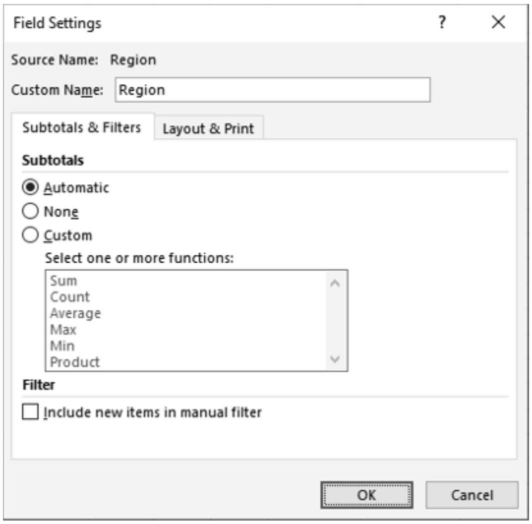


Figure 1.7: The Field Settings dialog box

Figure 1.8 shows a PivotTable with both grand totals and subtotals hidden.

Sum of Sales	Column Labels	
Row Labels	Qtr-1	Qtr-2
⊟Central		
Illinois	4,243	3,623
Kansas	4,776	4,007
Kentucky	3,342	4,626
Missouri	4,669	4,556
Oklahoma	3,995	5,364
⊟East		
Florida	4,722	4,630
Massachusetts	3,442	4,155
New Jersey	4,365	4,316
New York	4,738	4,763
⊟West		
Arizona	3,795	4,663
California	4,330	4,507
Oregon	5,140	4,369
Washington	3,513	4,703

Figure 1.8: A PivotTable with no grand totals or subtotals

The GETPIVOTDATA Worksheet Function

PivotTables are dynamic by nature. That means they can change size and the location of the data you want, making it difficult to refer to their cells in formulas. Fortunately, Excel provides the GETPIVOTDATA worksheet function so that you can be sure you always point to the correct data.

For example, if you wanted to use California's first quarter sales in a formula, you can use this GETPIVOTDATA function:

```
=GETPIVOTDATA("Sales",$I$3,"State","California","Region","West","Qtr","
Qtr-1")
```

If more data is added to the source data and the PivotTable is refreshed so the cell that contains California's first quarter sales moves, the formula still points to the right place. The first argument of the function refers to the field in the Values area that you want to return. The next argument, I3, is the top-left cell of the PivotTable that contains the data. The last six arguments are pairs of values that determine what intersection of columns and rows to use.

The argument pairs can be in any order, but within the pairs, the field name must come first followed by the value for that field. The pair of arguments "State" and "California" tell the function to use the row where California appears on the State field. As the function evaluates the remaining pairs of arguments, it narrows down to the value to return.

You can also use GETPIVOTDATA to return totals. For example, to return the grand total of the Qtr-1 column, use this function:

```
=GETPIVOTDATA("Sales",$I$3,"Qtr","Qtr-1")
```

Or to return the grand total of the entire PivotTable, use this function:

```
=GETPIVOTDATA("Sales",$I$3)
```

The GETPIVOTDATA function is tedious to create manually. Fortunately, Excel will build the function for you when you click inside a PivotTable. Simply type an equal sign and click a cell inside the body of the PivotTable to have Excel build the function with all the proper arguments. If you want to refer to a cell in a PivotTable and avoid the GETPIVOTDATA function, you can simply type the cell address in your formula. To turn off the default behavior of using GETPIVOTDATA, go to File ➪ Options ➪ Formulas and uncheck the Use GetPivotData Functions For PivotTable References check box.

WARNING **Unlike formulas, a PivotTable does not automatically display changes made to data after the PivotTable is created. Click the Refresh button on the PivotTable Tools Analyze tab of the Ribbon to update the PivotTable for any changes.**

Worksheet Functions

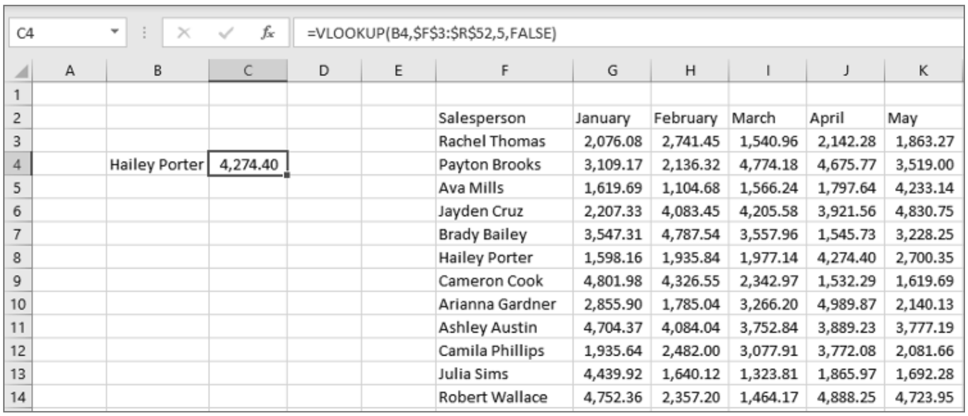
Excel has hundreds of worksheet functions to help you manipulate and stage data for a dashboard. In practice, you won't use most of them. But there are some worksheet functions that show up more often than not in dashboarding projects. In the following sections, I'll discuss some of the more common ones.

The VLOOKUP Function

The VLOOKUP function finds a value in the first column of a range and returns a value in another column on the same row as the found value. The syntax for VLOOKUP is as follows:

```
VLOOKUP(lookup_value, table_array, col_index_num, range_lookup)
```

Figure 1.9 shows an example of the VLOOKUP function.



Salesperson	January	February	March	April	May
Rachel Thomas	2,076.08	2,741.45	1,540.96	2,142.28	1,863.27
Payton Brooks	3,109.17	2,136.32	4,774.18	4,675.77	3,519.00
Ava Mills	1,619.69	1,104.68	1,566.24	1,797.64	4,233.14
Jayden Cruz	2,207.33	4,083.45	4,205.58	3,921.56	4,830.75
Brady Bailey	3,547.31	4,787.54	3,557.96	1,545.73	3,228.25
Hailey Porter	1,598.16	1,935.84	1,977.14	4,274.40	2,700.35
Cameron Cook	4,801.98	4,326.55	2,342.97	1,532.29	1,619.69
Arianna Gardner	2,855.90	1,785.04	3,266.20	4,989.87	2,140.13
Ashley Austin	4,704.37	4,084.04	3,752.84	3,889.23	3,777.19
Camila Phillips	1,935.64	2,482.00	3,077.91	3,772.08	2,081.66
Julia Sims	4,439.92	1,640.12	1,323.81	1,865.97	1,692.28
Robert Wallace	4,752.36	2,357.20	1,464.17	4,888.25	4,723.95

Figure 1.9: The VLOOKUP function

In this example, there is a table of commissions by salesperson and month in the range F3:R52. The formula finds the April commission for the salesperson in cell B4. The salesperson's name must be in the first column of the range. The formula looks down column F until it finds the salesperson's name, then returns the value in the fifth column.

The range_lookup argument is set to FALSE in this example because you want the formula to only return a value if it finds an exact match. If you set range_lookup to TRUE, Excel expects the data in the first column to be sorted and will return a close match if there isn't an exact one. There are situations where returning an approximate match is useful, but they are rare, and most of the time you'll want the last argument to be FALSE.

The XLOOKUP Function

The XLOOKUP function is a relatively new addition to the Excel worksheet function lineup. It works like VLOOKUP but doesn't require the value to find to be in the first column. Here is the syntax for XLOOKUP:

```
XLOOKUP(lookup_value, lookup_array, return_array, if_not_found,
match_mode, search_mode)
```

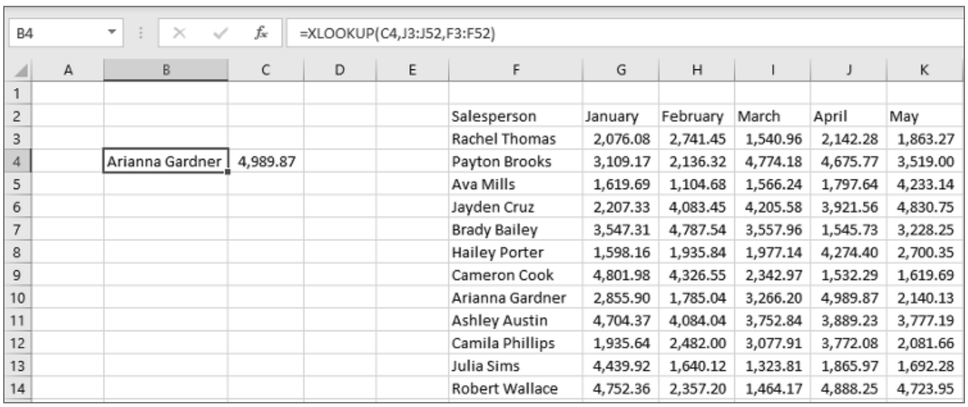
The first argument is the same for XLOOKUP and VLOOKUP. Instead of supplying the entire range and assuming the lookup value is in the first column, you supply a lookup_array (the column to search) and a return_array (the column with the value to return).

The XLOOKUP function also includes the if_not_found argument that lets you specify a value to return if the lookup value isn't in the list. If you don't specify this argument and the value isn't found, XLOOKUP returns the #N/A error value just like VLOOKUP.

The match_mode argument is like the range_lookup argument in VLOOKUP. Instead of TRUE or FALSE, you supply a number to tell Excel whether you want an exact match or one of the approximate match options. The exact match option, 0, is the default.

The last argument, search_mode, allows you to tell Excel how to search for the value. If you have a large data set, the search_mode argument can help speed up the search. For most data sets, it can be ignored.

Figure 1.10 shows an XLOOKUP formula to find the name of the salesperson with the highest commission in April.



B4 =XLOOKUP(C4,J3:J52,F3:F52)

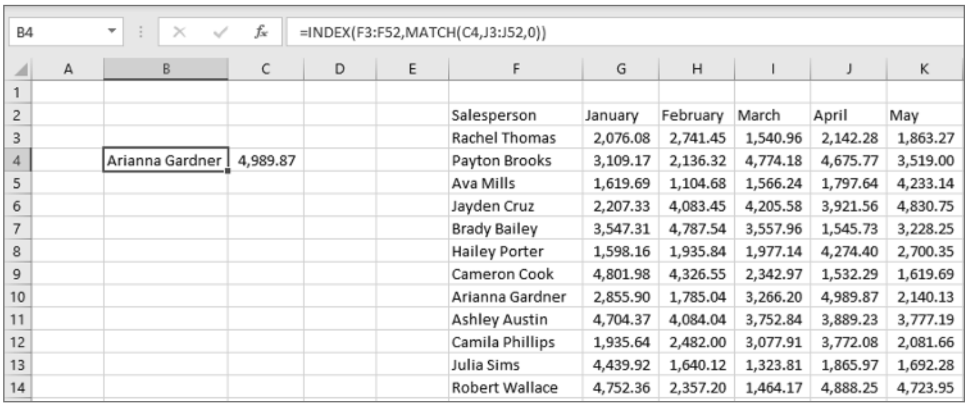
	A	B	C	D	E	F	G	H	I	J	K
1											
2						Salesperson	January	February	March	April	May
3						Rachel Thomas	2,076.08	2,741.45	1,540.96	2,142.28	1,863.27
4		Arianna Gardner	4,989.87			Payton Brooks	3,109.17	2,136.32	4,774.18	4,675.77	3,519.00
5						Ava Mills	1,619.69	1,104.68	1,566.24	1,797.64	4,233.14
6						Jayden Cruz	2,207.33	4,083.45	4,205.58	3,921.56	4,830.75
7						Brady Bailey	3,547.31	4,787.54	3,557.96	1,545.73	3,228.25
8						Hailey Porter	1,598.16	1,935.84	1,977.14	4,274.40	2,700.35
9						Cameron Cook	4,801.98	4,326.55	2,342.97	1,532.29	1,619.69
10						Arianna Gardner	2,855.90	1,785.04	3,266.20	4,989.87	2,140.13
11						Ashley Austin	4,704.37	4,084.04	3,752.84	3,889.23	3,777.19
12						Camila Phillips	1,935.64	2,482.00	3,077.91	3,772.08	2,081.66
13						Julia Sims	4,439.92	1,640.12	1,323.81	1,865.97	1,692.28
14						Robert Wallace	4,752.36	2,357.20	1,464.17	4,888.25	4,723.95

Figure 1.10: The XLOOKUP function

Instead of looking for a value in the first column, this formula looks for a value in the fifth column and returns the corresponding value from the first column. The formula in cell C4, `=MAX(J3:J52)`, finds the largest value in column J. The XLOOKUP function looks down column J for that value and returns the value in column F on the same row. Because you know the value exists and you want an exact match, you don't need to supply any of the other arguments to the function.

The INDEX and MATCH Functions

If you're using a version of Excel that doesn't include XLOOKUP, you can create the same result using a combination of the INDEX function and the MATCH function. The INDEX function returns a value from a range based on the row number and column number you provide. For a single column range, you only have to provide the row number. The MATCH function part of the formula computes which row INDEX uses. Figure 1.11 shows a formula to return the same values as the XLOOKUP function in the previous example.



B4 =INDEX(F3:F52,MATCH(C4,J3:J52,0))

	A	B	C	D	E	F	G	H	I	J	K
1											
2						Salesperson	January	February	March	April	May
3						Rachel Thomas	2,076.08	2,741.45	1,540.96	2,142.28	1,863.27
4		Arianna Gardner	4,989.87			Payton Brooks	3,109.17	2,136.32	4,774.18	4,675.77	3,519.00
5						Ava Mills	1,619.69	1,104.68	1,566.24	1,797.64	4,233.14
6						Jayden Cruz	2,207.33	4,083.45	4,205.58	3,921.56	4,830.75
7						Brady Bailey	3,547.31	4,787.54	3,557.96	1,545.73	3,228.25
8						Hailey Porter	1,598.16	1,935.84	1,977.14	4,274.40	2,700.35
9						Cameron Cook	4,801.98	4,326.55	2,342.97	1,532.29	1,619.69
10						Arianna Gardner	2,855.90	1,785.04	3,266.20	4,989.87	2,140.13
11						Ashley Austin	4,704.37	4,084.04	3,752.84	3,889.23	3,777.19
12						Camila Phillips	1,935.64	2,482.00	3,077.91	3,772.08	2,081.66
13						Julia Sims	4,439.92	1,640.12	1,323.81	1,865.97	1,692.28
14						Robert Wallace	4,752.36	2,357.20	1,464.17	4,888.25	4,723.95

Figure 1.11: The INDEX and MATCH functions

The syntax for the MATCH function is

```
MATCH(lookup_value, lookup_array, match_type)
```

The MATCH function works like VLOOKUP except that there is no column to return. It simply returns the position of lookup_value in the lookup_array list of values. The last argument of MATCH, 0, looks for an exact match like when the last argument of VLOOKUP is FALSE. In this example, MATCH returns the value 8 because the value in C4 is the eighth value in the range J3:J52.

Once Excel computes the value returned by MATCH, the INDEX function is evaluated like

```
=INDEX(F3:F52,8)
```

The syntax for the INDEX function is

```
INDEX(reference, row_number, column_number)
```

Because the reference argument is a single column, F3:F52 in this example, you only need to supply the row_number argument.

The SUMPRODUCT Function

The SUMPRODUCT function is one of Excel's most powerful worksheet functions. It's designed to multiply two or more ranges together and sum the results to a single number. A simple example of SUMPRODUCT is shown in Figure 1.12.

C7 =SUMPRODUCT(A2:A5,B2:B5)

	A	B	C	D	E	F	G
1	Quantity	Price					
2	2	9.07					
3	3	4.46					
4	10	8.43					
5	9	4.53					
6							
7	Invoice Total:		156.59				
8							

Figure 1.12: The SUMPRODUCT function

In this example there are quantities, and for each quantity, a price. SUMPRODUCT multiplies each quantity by its corresponding price and sums all of the results. It's the same as putting in the formula `=A2*B2` and filling down in column C and then using `=SUM(C2:C5)`, but it does it in one formula.

The SUMPRODUCT function works like an array formula, discussed in the next section, and that's where its real power lies. With SUMPRODUCT, you can filter out values by multiplying them by 0, so they have no effect on the final sum. Figure 1.13 shows an example of using SUMPRODUCT to only sum totals for one product in one month.

L6 =SUMPRODUCT(($C$2:$C$3651=K6)*(MONTH($A$2:$A$3651)=4)*($I$2:$I$3651))

	A	B	C	D	E	F	G	H	I	J	K	L
1	Date	Register	Item	RcptID	Units	Quantity	Price	Taxable	Total			
2	1/1/2021	POS-1-3	P53917	142387	Each	8	66.98	N	535.84			
3	1/1/2021	POS-4-2	P57772	142388	Pounds	10	75.43	N	754.30			
4	1/1/2021	POS-2-9	P73654	142389	Each	9	93.00	Y	837.00			
5	1/1/2021	POS-1-4	P23981	142390	10Pk	2	51.03	N	102.06			
6	1/1/2021	POS-4-8	P23981	142391	5Pk	10	57.56	N	575.60		P23981	5,000.32
7	1/1/2021	POS-2-1	P84419	142392	2PK	3	44.53	N	133.59			
8	1/1/2021	POS-1-4	P57772	142393	5Pk	3	59.39	Y	178.17			
9	1/1/2021	POS-2-3	P73302	142394	Pounds	4	69.06	N	276.24			
10	1/1/2021	POS-5-9	P39367	142395	5Pk	5	61.18	N	305.90			
11	1/1/2021	POS-4-4	P53917	142396	Feet	1	19.67	N	19.67			

Figure 1.13: Filter values with SUMPRODUCT

This SUMPRODUCT example consists of three sections: two comparison sections and one result section. The magic of SUMPRODUCT is that when a comparison returns TRUE, multiplying it treats TRUE as the number 1, and when a comparison returns FALSE, multiplying it treats FALSE as the number 0. The first section, `($C$2:$C$3651=K6)`, compares every value in column C to the value in cell K6 and returns a TRUE or FALSE. The second section, `(MONTH($A$2:$A$3651)=4)`, compares the month of every date in column A to 4 and returns a TRUE or FALSE.

The last section, `($I$2:$I$3651)`, doesn't do any comparison. Instead, it simply returns the long list of numbers from column I. Each of these sections is evaluated to an array of 3,650 values. The values in the arrays for the first two sections are either TRUE or FALSE and the values in the array of the last section are the values in column I.

Next, SUMPRODUCT multiplies each line of the arrays together. For the first items in the arrays, which correspond to row 2 of the data, SUMPRODUCT multiplies FALSE * FALSE * 535.84 because the value in C2 is not equal to the value in K6 and the month in A2 isn't 4. Since FALSE evaluates to 0 and multiplying any number by 0 equals 0, the total for that item is 0.

For the data in row 6, the item does match, but the month still does not. SUMPRODUCT evaluates that row as TRUE * FALSE * 575.60. The TRUE value is evaluated as a 1, but since the second section is FALSE, thus 0, the whole row still evaluates to 0. Only when the formula gets down to row 924 do both the item and the month match. When it evaluates that row, it multiplies TRUE * TRUE * 47.12. The TRUE values evaluate to 1 and that row adds 47.12 to the final sum.

You can see that if any condition is FALSE, Excel inserts a 0 into the multiplication and makes the whole row 0. Only if every condition is TRUE is a value other than 0 added to the total. It doesn't matter if you have 2 conditions or 20, if they're all TRUE, Excel keeps multiplying by 1, which has no effect on the final section.

In this way, SUMPRODUCT is like a PivotTable. In a PivotTable, Excel aggregates the values where all the row fields intersect all the column fields. The value of a PivotTable over SUMPRODUCT is that it's far easier to change a PivotTable than to rewrite a bunch of SUMPRODUCT formulas. But the value of formulas is that they recalculate instantly when the data changes, whereas you must refresh a PivotTable.

Another similarity between SUMPRODUCT and PivotTables is that you can aggregate values by counting them instead of summing them. To count values using SUMPRODUCT, only include comparison sections in the formulas. For example, to count all the rows with a specific product in April, use the following formula:

```
=SUMPRODUCT(($C$2:$C$3651=K7)*(MONTH($A$2:$A$3651)=4))
```

This formula is like the preceding example except that the section representing column I is omitted. Now when both conditions are TRUE, Excel multiplies 1 * 1, and the result, 1, is included in the sum. There are 18 rows in the example data that have P23981 as the item and the date is in April. For each of those rows, 1 is added to the total and the formula returns 18.

Another technique you can use with SUMPRODUCT is adding comparison sections together that refer to the same column. For example, if you want to sum the total for two products, P23981 and P73302, that occur in April, use the following formula:

```
=SUMPRODUCT((($C$2:$C$3651="P23981")+($C$2:$C$3651="P73302"))
*(MONTH($A$2:$A$3651)=4)*($I$2:$I$3651))
```

In this example, there are two conditional sections that refer to column C. They can never both be TRUE because any row in column C can only contain one item. By adding them together, Excel will add 0 + 0, 0 + 1, or 1 + 0 and will return the total if column C is either product (or 0 if it's neither). Because Excel multiplies before it adds, you have to enclose these conditional sections in their own sets of parentheses to force Excel to perform that evaluation first. Only use this technique when the comparison sections refer to the same column. If you add together comparisons for different columns, and they're both TRUE, it would return a 2 and your result would be wrong.

Array Formulas

The SUMPRODUCT function in the previous section is a special kind of array function. It only sums the results and can't aggregate them in any other way. Array formulas use the exact same calculation as SUMPRODUCT but you can choose how to aggregate the filtered values. With array formulas, you have to commit the formula after you enter it using Ctrl+Shift+Enter, not just Enter as you do with SUMPRODUCT or other worksheet functions.

Using the same data set from the SUMPRODUCT examples, if you want to return the largest total for product P23981 in April, you would use an array formula and the MAX function, as shown below and in Figure 1.14:

```
=MAX(($C$2:$C$3651=K6)*(MONTH($A$2:$A$3651)=4)*($I$2:$I$3651))
```

The formula bar in Figure 1.14 shows curly braces around the formula. You don't type those curly braces. Excel inserts them automatically when you commit the formula using Ctrl+Shift+Enter instead of just Enter. If your array formula isn't returning the correct result, the first thing to check is that the curly braces are there. If they are not, press F2 to edit the cell and recommit it with Ctrl+Shift+Enter.

L6 {=MAX(($C$2:$C$3651=K6)*(MONTH($A$2:$A$3651)=4)*($I$2:$I$3651))}

	A	B	C	D	E	F	G	H	I	J	K	L
1	Date	Register	Item	RcptID	Units	Quantity	Price	Taxable	Total			
2	1/1/2021	POS-1-3	P53917	142387	Each	8	66.98	N	535.84			
3	1/1/2021	POS-4-2	P57772	142388	Pounds	10	75.43	N	754.30			
4	1/1/2021	POS-2-9	P73654	142389	Each	9	93.00	Y	837.00			
5	1/1/2021	POS-1-4	P23981	142390	10Pk	2	51.03	N	102.06			
6	1/1/2021	POS-4-8	P23981	142391	5Pk	10	57.56	N	575.60		P23981	797.70
7	1/1/2021	POS-2-1	P84419	142392	2PK	3	44.53	N	133.59			
8	1/1/2021	POS-1-4	P57772	142393	5Pk	3	59.39	Y	178.17			
9	1/1/2021	POS-2-3	P73302	142394	Pounds	4	69.06	N	276.24			

Figure 1.14: An array formula to find the largest value

Just like SUMPRODUCT, this array formula contains two conditional sections and one values section. After Excel multiplies the three sections together, again treating TRUE as 1 and FALSE as 0, instead of summing the result, it finds the maximum value of the results. In fact, if you were to change the MAX to SUM in that formula, you would get the same result as the SUMPRODUCT formula.

To find the smallest value from a list of filtered values requires a slightly different technique. Since the values that you filter out evaluate to 0 prior to being aggregated, the smallest value from a list of positive values would always equal 0. Fortunately, the MIN function ignores text values and we can use that to only include nonzero values. The following formula returns the smallest value for item P23981 in April:

```
=MIN(IF(($C$2:$C$3651=K15)*(MONTH($A$2:$A$3651)=4)=0,"",$I$2:$I$3651))
```

The two conditional sections are places in the conditional argument of an IF function. If when multiplied together they equal 0, an empty string (two double quotes) is inserted in the filtered values, otherwise the value from column I is inserted. Since MIN ignores strings, it will ignore all the empty string values and only return the smallest value where the conditions are TRUE.

Tables

Tables are special areas in a spreadsheet that hold data. They're meant to be like database tables, but they only share some of the same properties. The two main advantages to using tables are that cell formats and formulas are consistent for all the rows, and references to the table grow and shrink automatically as the table grows and shrinks. Figure 1.15 shows a small table and a column chart created from that table.

If you change the number of rows or columns in the table, the chart will adjust automatically. The chart isn't simply referencing the range A1:C6, it's referencing the special area with a name like Table1 and is aware when that area grows or shrinks.

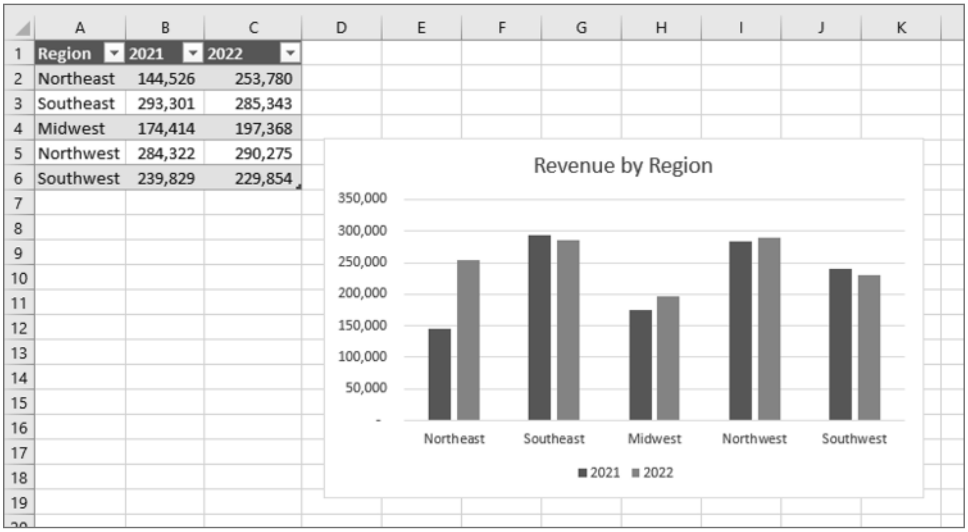


Figure 1.15: A column chart based on a table

The first step is to convert a normal Excel range into a table. To do that, select the range and choose Insert ➪ Table from the Ribbon to show the Create Table dialog in Figure 1.16.

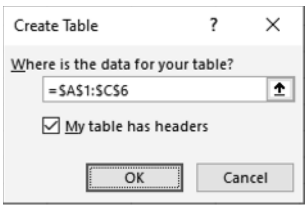
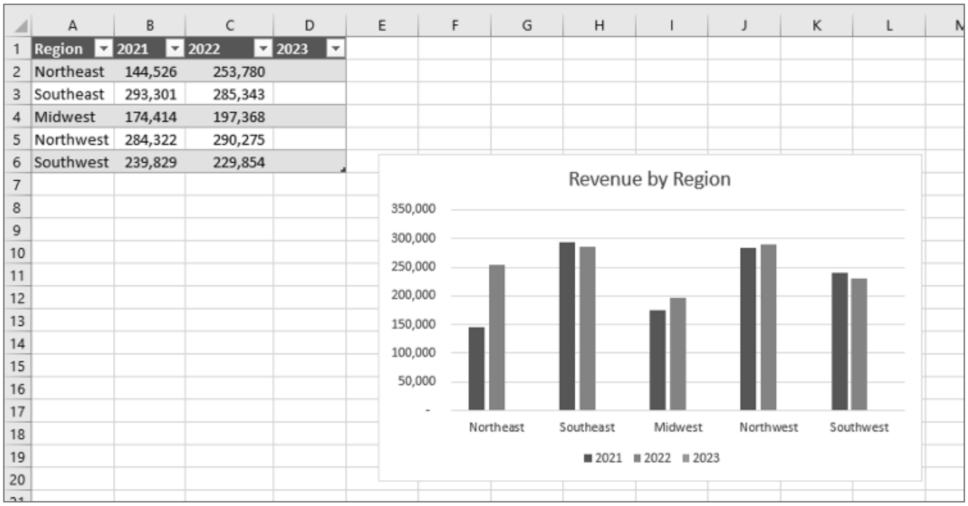


Figure 1.16: The Create Table dialog

This dialog confirms the range you want to convert and lets you specify if the first row of your range contains headers. Click OK to finish creating the table. Next, select the table and choose Insert ➪ Recommended Charts from the Ribbon and choose the Clustered Column option to insert the chart. You can change the chart's title as I have, but it's not necessary for this example.

Now imagine a scenario where you want to add the next year to this chart and you want to split Texas out of the Southwest region and make it its own region. With a normal Excel range, you would add a new series for the new year and adjust the existing series to point to the larger range. But because you're using a table, you can simply add the data and the chart will adjust.

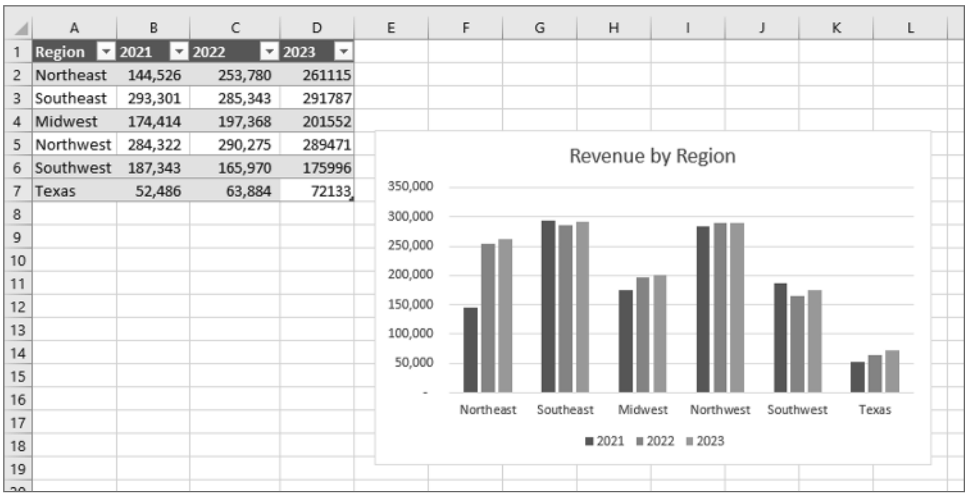
To add the new year, enter **2023** in cell D1. Figure 1.17 shows that Excel has increased the table columns for the new range and has already added a new series to the chart.



Region	2021	2022	2023
Northeast	144,526	253,780	
Southeast	293,301	285,343	
Midwest	174,414	197,368	
Northwest	284,322	290,275	
Southwest	239,829	229,854	

Figure 1.17: The table expands for new data.

Next, enter **Texas** into cell A7. Excel will increase the table rows and you can enter values for the new year and split the Southwest region. Figure 1.18 shows the new table size and the resulting chart.



Region	2021	2022	2023
Northeast	144,526	253,780	261115
Southeast	293,301	285,343	291787
Midwest	174,414	197,368	201552
Northwest	284,322	290,275	289471
Southwest	187,343	165,970	175996
Texas	52,486	63,884	72133

Figure 1.18: The chart adjusts for new data.

Excel has automatically added a new clustered column to the chart for the new region and the 2023 series is populated with the data entered in the table. The new data entered for Texas for existing years is automatically formatted with the same number format as the other data in those columns. Because 2023 is a new column, it does not have existing formatting to apply and you must format that column separately. Nothing on the chart was changed. You only have to add new data to the table to achieve these results.

Structured Table Referencing

The chart in the previous example adjusted as the data in the table changed. When you refer to tables in formulas using *structured table referencing*, your formulas will automatically account for changes in the table too. For these examples, the name of the table was changed to tblRevenue from Excel's default table name. To change the name of a table, select any part of the table and choose Table Tools Design ➪ Table Name from the Ribbon and type the table's new name.

To create a formula using structured table referencing, select an unused cell in the worksheet, enter the start of a formula like **=SUM(**, and point to the range B2:B7. If you enter data in a cell that's adjacent to the table, Excel will assume you want to expand the table, so choose a cell away from the table. By default, Excel will convert your formula to structured table referencing like the formula below:

```
=SUM(tblRevenue[2021])
```

Instead of referring to the range B2:B7, Excel converted the argument to tblRevenue[2021], referring to the column labeled 2021 in the table named tblRevenue. If you're staging data for a dashboard, you can use structured table referencing to make sure your formulas always cover all the data.

You don't have to point to the data to get structured table referencing. You can enter that same formula manually. Start entering a formula by typing **=SUM(tbl** into a cell. Figure 1.19 shows the drop-down box Excel displays showing all your choices that contain those letters.

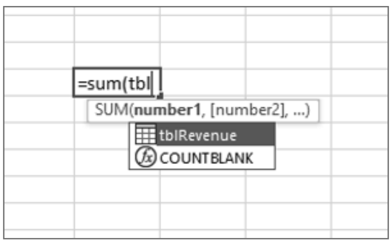


Figure 1.19: Excel's formula entry drop-down

If you have more than one table in your workbook and you choose to name them all starting with the same letters, like *tbl*, this method gives you a handy list of tables to choose from. In this example, there is only one table and it is selected because it's the first one on the list. From here, you can press the Tab key to complete the table name without having to type it all. After the table name, type an open bracket, [, to see a list of options related to that table. Figure 1.20 shows the list of options for tblRevenue.

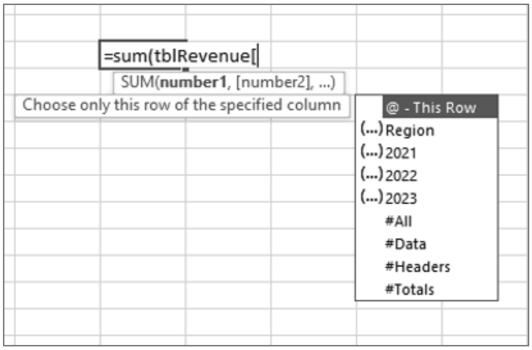


Figure 1.20: Options for table formula completion

The first option is the @ symbol. When you're using formulas from within a table, the @ symbol tells Excel to return the value from the same row as the formula. The reference [@2021] returns the value from the column named 2021 and that's on the same row. Contrast that to the reference [2021], which returns the entire column.

The next four options are the column names in the table. Excel is aware of the structure of the table and autogenerates named ranges for the columns. The last four options are references to special areas in the table. The #All reference refers to the entire table including headers and totals. The #Data reference is the same as #All except it excludes headers and totals. The #Headers and #Totals references refer only to their respective areas. You can arrow down to the option you want and press Tab to complete the entry.

For example, if you arrow down to the #Data reference, press Tab, and enter a closing bracket and closing parenthesis, you would get the following formula that sums all the data in the table:

```
=SUM(tblRevenue[#Data])
```

Text to Columns

When you're dealing with data that originated outside of Excel, sometimes the data is all lumped together instead of nicely laid out in columns. Excel's Text to Columns feature can help you manipulate data into a usable format. Figure 1.21 shows data pasted into column A of a worksheet. It is columnar data separated by commas but is all in one cell.

To separate the data, follow these steps:

1. Select the data and choose Data ➪ Text to Columns from the Ribbon.
2. From the Convert Text to Columns Wizard, shown in Figure 1.22, select the Delimited option and click Next.

3. Unselect the Tab delimiter and select the Comma delimiter. Excel shows a preview of how it split the data, as shown in Figure 1.23. Click Next.
4. On the last step, you can format each column. Excel will generally convert dates and numbers successfully. If you have text that contains numbers with leading zeros, Excel will remove the zeros unless you format that column as Text. Figure 1.24 shows the data split into separate columns.

Figure 1.21: Data that is lumped together in a cell

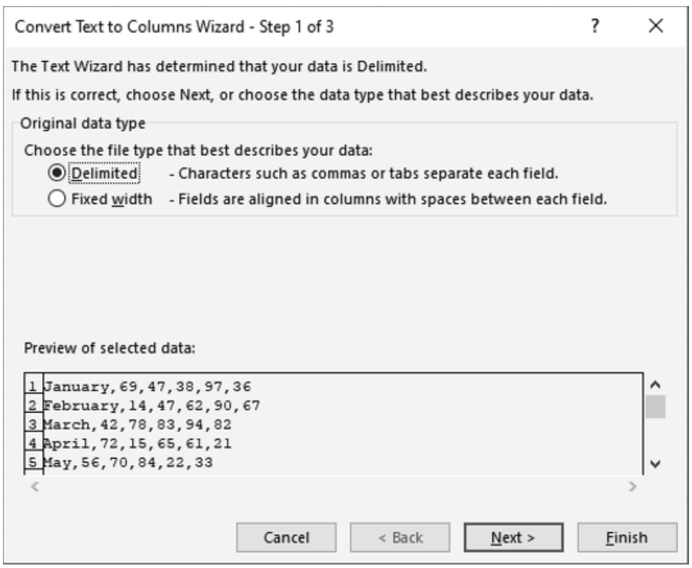


Figure 1.22: The Convert Text to Columns Wizard

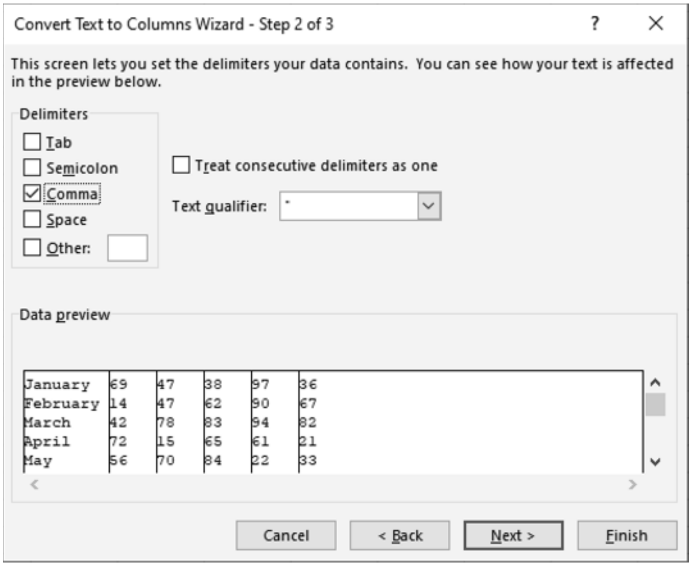


Figure 1.23: Excel splits the data on the delimiter.

	A	B	C	D	E	F
1	January	69	47	38	97	36
2	February	14	47	62	90	67
3	March	42	78	83	94	82
4	April	72	15	65	61	21
5	May	56	70	84	22	33
6	June	37	38	97	39	34
7	July	18	96	76	91	88
8	August	13	59	79	43	80
9	September	70	36	33	77	69
10	October	29	58	25	99	87
11	November	56	83	88	39	24
12	December	93	99	52	33	97

Figure 1.24: Data split into separate columns

Removing Duplicates

Another common problem when dealing with raw data is duplicate records. Excel's Remove Duplicates tool is a quick and easy way to delete duplicate records in place without having to use PivotTables or tedious formulas. The data shown in Figure 1.25 has some duplicate records.

To remove the duplicates, select the data and click Remove Duplicates from the Data tab on the Ribbon. This will show the Remove Duplicates dialog, shown in Figure 1.26.

	A	B	C	D
1	Invoice	Date	Customer	Amount
2	IN001000	12/1/2021	Atlantic Northern	847.97
3	IN001000	12/1/2021	Atlantic Northern	847.97
4	IN001003	12/21/2021	Big T Burgers and Fries	763.62
5	IN001003	12/21/2021	Big T Burgers and Fries	763.62
6	IN001006	12/25/2021	Sixty Second Avenue	516.08
7	IN001007	12/4/2021	Mainway Toys	706.85
8	IN001008	12/19/2021	Carrys Candles	726.68
9	IN001013	12/2/2021	The Legitimate Businessmens Club	666.51
10	IN001014	12/14/2021	Zevo Toys	780.58
11	IN001016	12/15/2021	North Central Positronics	370.02
12	IN001022	12/25/2021	Fake Brothers	142.65
13	IN001022	12/25/2021	Fake Brothers	142.65
14	IN001022	12/25/2021	Fake Brothers	142.65
15	IN001023	12/14/2021	General Services Corporation	931.21
16	IN001024	12/3/2021	General Products	918.33
17	IN001026	12/6/2021	Taco Grande	303.72
18	IN001032	12/10/2021	Videlectrix	939.56
19	IN001036	12/15/2021	Wernham Hogg	909.72
20	IN001037	12/30/2021	Sample, inc	563.13
21	IN001040	12/24/2021	Taco Grande	472.36
22	IN001040	12/24/2021	Taco Grande	472.36
23	IN001042	12/7/2021	Roxxon	909.71
24	IN001046	12/1/2021	LuthorCorp	121.00
25	IN001048	12/15/2021	Videlectrix	897.77

Figure 1.25: Data with duplicate records

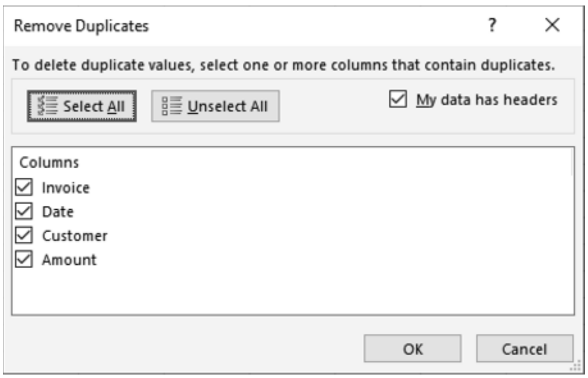


Figure 1.26: The Remove Duplicates dialog

The dialog shows all the columns from the data, and you can select which of those columns will determine what's a duplicate. With invoices, like in this example, all the data will be duplicated, so choose all the columns. For other types of data, maybe only one or two of the columns will determine a duplicate. To avoid deleting data you shouldn't, try to use as many columns as possible. When you click OK, Excel displays the message box in Figure 1.27 showing how many rows will be deleted.

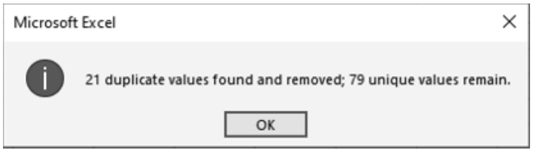


Figure 1.27: Excel displays the number of deleted values.

Click OK to dismiss the message box and you're left with only unique records. You can undo this action if you determine you made a mistake.

Building the Dashboard

If you're a typical Excel user, like me, you probably relish working with numbers, tables, and PivotTables. You may not be as comfortable designing visual elements. Fortunately, there are lot of great examples to draw from, and you probably already see them in news articles or on the websites you visit. Dashboarding has been around long enough that the community has developed some visualization principles to help you create great dashboards. Once you learn these principles, you'll likely become a critic of every dashboard you see!

The first rule is to stay on message. If you've done the planning steps, you already have a good idea of what data you have available and what the end user's needs are. Make every element of your dashboard further the story your dashboard tells. Can you tell the story and meet the user's needs with one simple chart? If so, don't clutter the dashboard with other elements. There is no minimum number of visual elements you need to make a dashboard.

To that end, don't let the availability of data determine which elements to include on the dashboard. It's very likely you'll have more data available than you need. When you have a data set that's perfectly aggregated and requires almost no work to drive a chart, it's tempting to include it because of the low cost of creation. But there's an even greater cost if it detracts from the dashboard's overall message.

That's not to say that every element on your dashboard must be a showstopper. On every dashboard there are more important and less important elements. Elements that support the story, but aren't necessarily critical, can still be included if they don't add clutter. There's a balance between providing as much information as possible and keeping the dashboard clean and simple that you will learn with experience.

Organizing Elements

There's a lot of information and opinions about how to design for documents with multiple elements, particularly in the area of web page design. I think there are three main principles that cover most of it: size, color, and placement.

The largest element will get the most attention. Make the element with most important information larger than the others. It doesn't have to be so large that it dominates the page, but its size should be an indication of its importance. Consider the white space around an element as part of its size. Figure 1.28 shows a dashboard with three elements, one of which is larger to indicate its importance over the others.

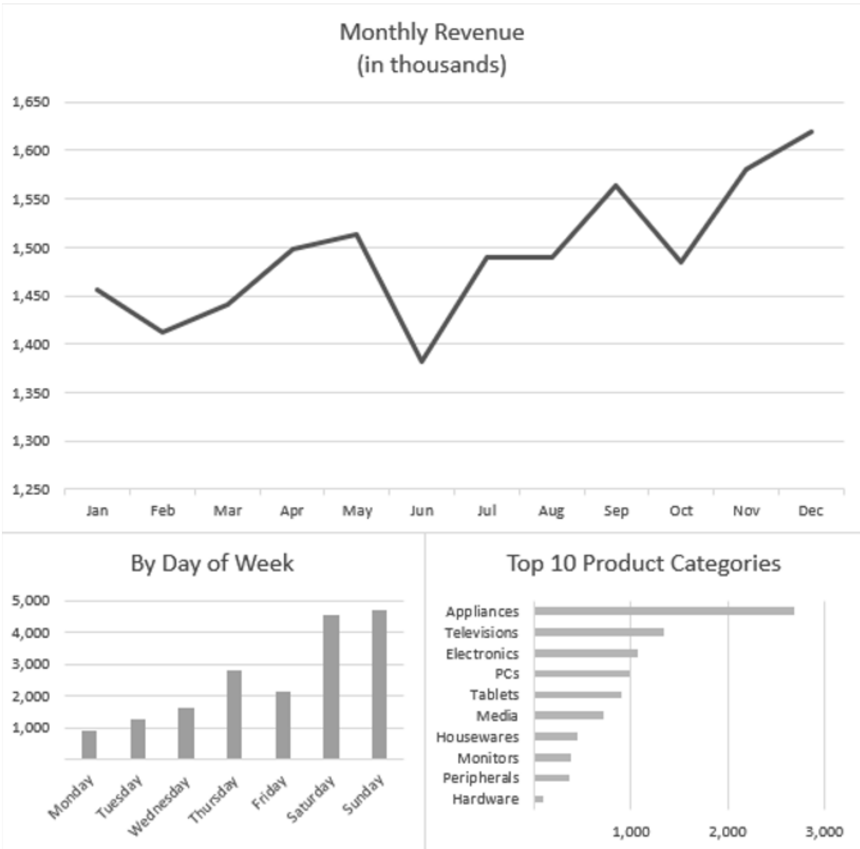


Figure 1.28: A dashboard with different sized elements

Elements with bright and contrasting colors will be perceived as more important than light, consistently colored elements. If every element on your dashboard has bright, contrasting colors, it will look more cluttered than the same layout with only one such element. I'm not advocating one garishly colored chart with a bunch of gray ones. It only takes a subtle difference in color to draw a reader's attention. In every case, you'll want to limit the total number of colors you use and the ones that are there should complement each other. But a slight variation can make an element stand out.

Generally, the most important element should be at the top left of the page and the least important at the bottom right. Think of someone reading a few lines of text. They start at the top left and read across, then drop down to the next line and read across from there. Whatever's at the top left will get the reader's attention first. Another design pattern is to place the most important element in the center and surround it with less important elements. This pattern seems most effective when one element is overwhelmingly more important than the others.

Varying Elements

A dashboard typically contains several different chart types. If every element is a line chart, the reader has to work harder to interpret the information. There is a case for having the same chart type for all the elements. If the data is the same for every element except for one dimension, like sales over time, but each chart is for a different region, then using the same chart type will help the reader make the connection that the data is similar. However, when the data varies by more dimensions, like units vs. dollars and time scale, different chart types are warranted. Figure 1.29 shows a dashboard where all eight charts are the same type, indicating that the data is the same except for one dimension.

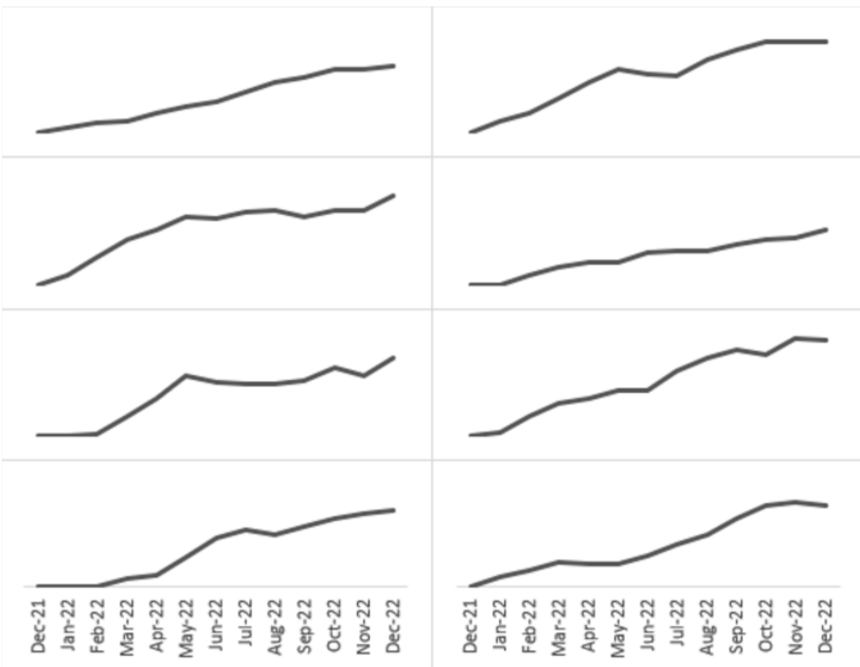


Figure 1.29: A dashboard of the same chart type

As you'll see in Part III, "Tell a Story with Visualization," the chart type you select should match the data and the story you're telling. But there is some overlap. Line charts are great for showing changes over time, but a column chart can be just as effective if the time is broken into discrete periods, like months.

That doesn't mean that every element has to be a different chart type. It just means that if you do use the same chart type for different data, don't place them right next to each other and make them the same size and color. And like the warning against including data simply because it's available, don't include a chart type because it's the hot, new thing or you just learned how to make it. Ensure that the chart types fit the data and further the story.

Showing Trends

If you want to show a trend, you can add a trendline to a chart. Some charts already show the trend and the trendline doesn't add any information. Figure 1.30 shows a line chart with a dashed trendline. The trend is easy to see without the added trendline, so it should be omitted.

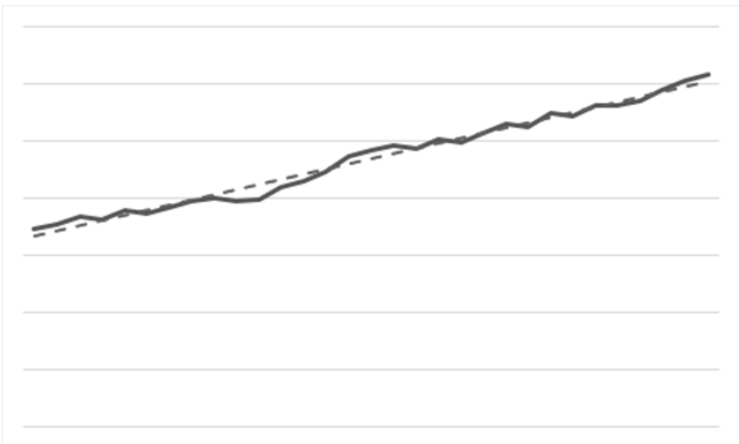


Figure 1.30: A trendline that adds little information

In other cases, the trendline is helpful. If the data has variations that make the trend hard to see, and showing the trend is part of the story you're telling, then include a trendline. Figure 1.31 shows a line chart of daily sales. The periodic dips for the decreased sales over the weekends make it harder to spot the trend. In this case, a trendline helps the reader interpret the data more easily.

To add a trendline, right-click on a series and choose Add Trendline. By default, a linear trendline is added to the chart and the Format Trendline task pane is displayed, as shown in Figure 1.32.

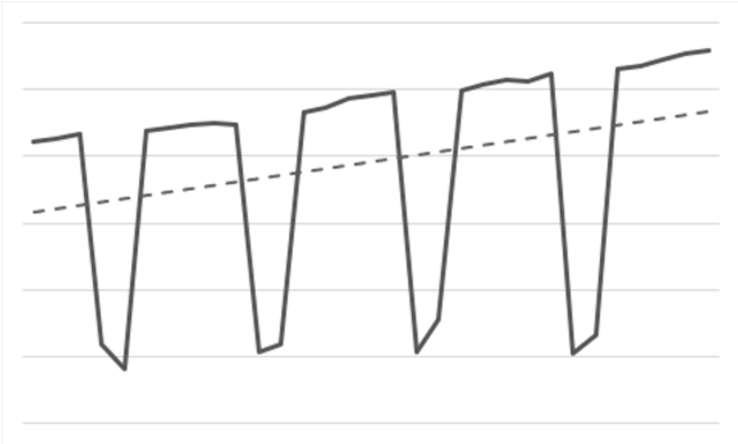


Figure 1.31: A trendline helps when data varies.

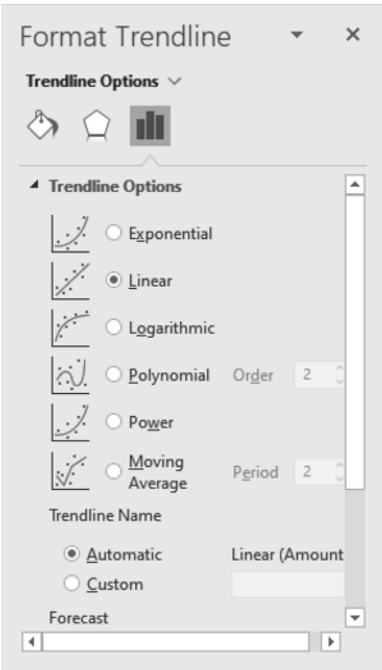


Figure 1.32: The Format Trendline task pane

There are several options for the type of trendline to display. Choose a trendline type that fits your data:

- **Exponential:** Best for data sets that rise or fall at increasingly higher rates.
- **Linear:** Best for simple, linear data sets.

- **Logarithmic:** Best for data sets that rise or fall quickly, then level out.
- **Polynomial:** Best for data sets that fluctuate.
- **Power:** Best for data sets that rise or fall at a constant rate.
- **Moving Average:** Best for data sets that have regular fluctuations that you want to smooth out.

Formatting the Dashboard

Formatting can be the difference between a good dashboard and a great one. Fortunately, Excel provides a lot of formatting options for most of the chart types it offers. Excel also provides formatting options that don't add much value and can take away from the message of the chart.

As with overall dashboard design, keep your formatting clean and simple. Overformatting distracts the user from interpreting the data and leads to less effective charts. Figure 1.33 shows a chart with a lot of formatting. Figure 1.34 shows the same chart with a clean and simple design. While the first chart shows how powerful Excel's charting engine is, the second chart showcases the data and the message.

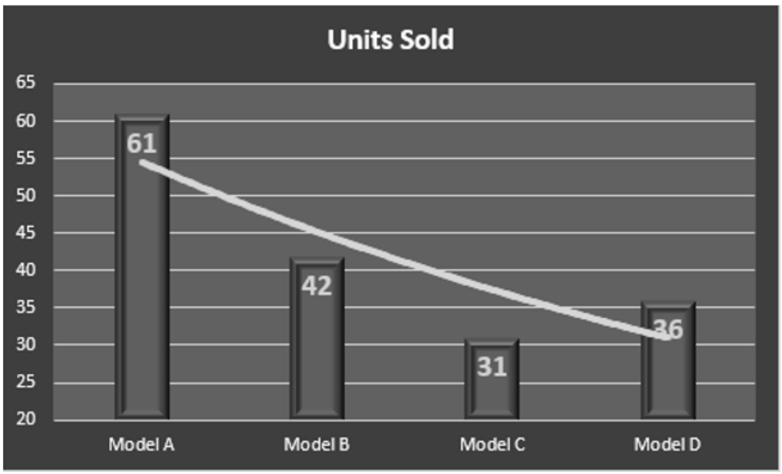


Figure 1.33: Too much formatting detracts from the chart's message.

The worst offender of Excel's formatting options is 3D. Applying 3D formatting almost always distorts the visual elements so they no longer represent the underlying data. With 3D column charts, the front row of columns has a larger area and the back row is partially obscured. 3D charts look cool, but they make it difficult for the reader to interpret the data accurately and the message may be lost.

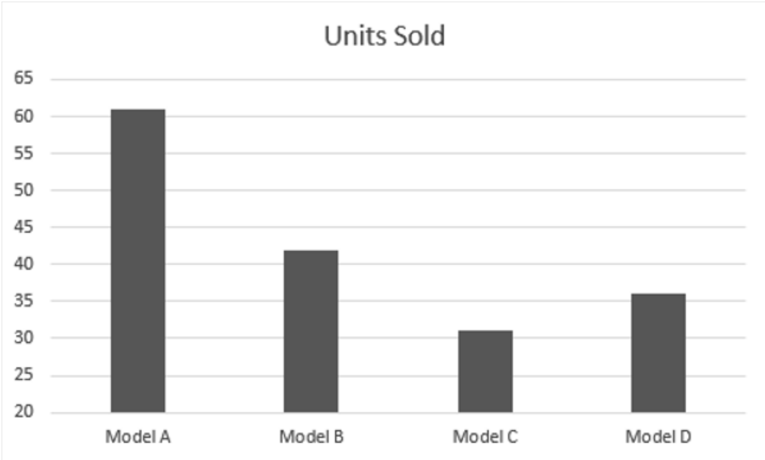


Figure 1.34: Clean and simple formatting keeps the message front and center.

Here are some other chart formatting options to avoid or use with caution:

- **Background color:** White or very light colors are acceptable backgrounds. Black and other dark colors lessen the contrast between the chart elements that represent data and the background. The background never represents data and should not be emphasized.
- **Foreground color:** Don't include too many colors in your chart elements. Use color for a specific purpose, like to call attention to an important data point. Excel's color pallet, shown in Figure 1.35, makes it easy to keep your colors consistent by offering different levels of light and dark. Excel's theme gallery, shown in Figure 1.36, provides different pallets of colors that are complementary.

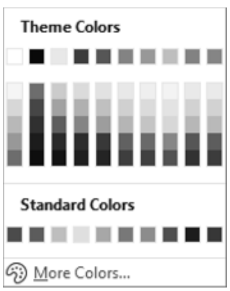


Figure 1.35: Excel's color pallet

- **Gradients:** Gradient fills change the color of a chart element from very dark to very light. This offers no value and should be avoided.

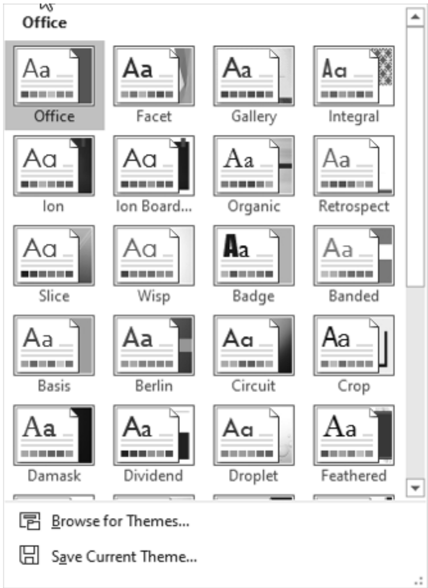


Figure 1.36: Excel's theme gallery

- **Pictures:** Picture fill colors a chart element with a repeating image. It not only adds no value but is distracting and decreases the effectiveness of the chart. Picture fill should also be avoided.
- **Effects:** The Format Data Series task pane has an Effects tab for most chart elements. It includes formatting options like Shadow, Glow, and Bevels. If used judiciously, some of these options can add depth to a chart that can be visually appealing. But there is a tendency to overuse these options and distract from the chart's message. Generally, you should avoid this tab unless you have a compelling reason to use it.
- **Borders:** Borders are lines that surround the plot area of a chart. They are almost never necessary. The axes already provide a visual boundary for the plot area. If you hide the axes because they aren't necessary, the border probably isn't necessary either. If you decide to use a border, make the color lighter so the data elements stand out.
- **Gridlines:** Excel includes gridlines by default on most chart types. Often, they don't add any value to the chart. If there is an important data point that you want to draw attention to and it's not near the axis, consider using a data label or color for that data point. If you choose to use gridlines, make sure they are light enough that they don't draw attention from the data elements.

- **Data labels:** If every data point has a data label, then you've created a table of data that's hard to read. Use data labels to highlight important data points, like the minimum or maximum value. Data labels can also be used to explain unexpected values. Figure 1.37 shows a chart that uses a data label to explain an unusual dip in the output for a manufacturing plant.

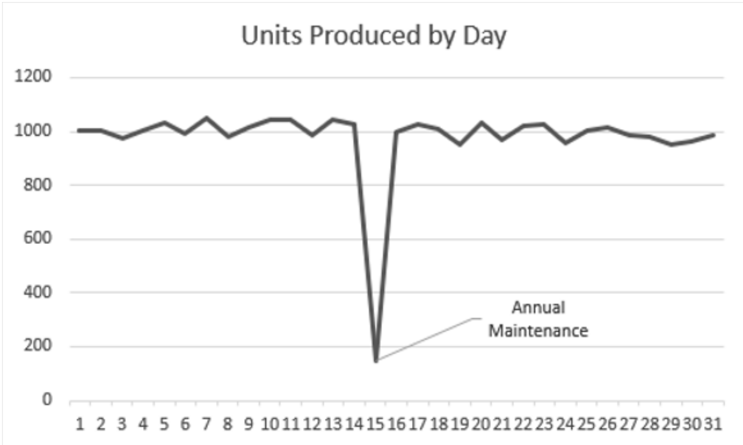


Figure 1.37: Data labels can be used to highlight unusual variations.

- **Axes:** In most cases, the axes provide context and scale to the chart. In some cases, such as variance charts, the value of the axes is diminished. Don't include an axis just because Excel adds them by default. Think about how they contribute to the message and help the reader interpret the data when you choose to show or hide them.

Avoid adding clip art to your charts. Your chart is already a picture. You don't need to add another picture to it. It can only detract from the data and message.

Number Formats

Your charts will almost always contain numbers, either in the axes or in data labels. Excel will use the formatting of the source data by default and that may not be what you want. Always use commas to make the numbers easier to read.

Format the numbers in your chart to the proper level of significance. If you're showing revenue numbers in the tens of millions, you don't need to show the pennies. Excel provides a robust number formatting system that allows you to format almost any way you want. Try to show the lowest number of digits possible without losing any data. For example, show the number 17,483,262 as 17.4 and use the title or axis label to tell the user that the numbers are in millions. The chart shown in Figure 1.38 uses this technique in the value axis.

Figure 1.38: Formatted large numbers in the value axis

To format large numbers in an axis, use the Number section of the Format Axis task pane, shown in Figure 1.39. The zeros indicate where a digit is displayed. Where there are no digits to the right of a comma, the triplet representing hundreds or thousands is completely hidden.

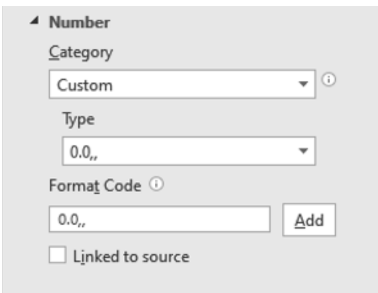


Figure 1.39: The Number section of the Format Axis task pane

Table 1.1 shows some ways that you can format large numbers.

Table 1.1: Custom formats for large numbers

TO DISPLAY THIS NUMBER. . .	USE THIS CUSTOM FORMAT
17.5	0,,.0
17	0,,
$17.5	$0.0,,
$17	$0,,

Continues

Table 1-1 (*continued*)

TO DISPLAY THIS NUMBER. . .	USE THIS CUSTOM FORMAT
17.5M	0.0,,"M"
17M	0,,"M"
$17.5M	$0,,.0"M"
$17M	$0,,"M"
17,483K	0,000,"K"
$17,483K	$0,000,"K"

You can use dollar signs for money and letters, like K for thousands, M for millions, and B for billions, to show the scale of the numbers. Only use these options if you have a few numbers on your chart, like if you've highlighted a couple of important data points. For axes, don't use dollar signs or letters. Instead, use the axis label or the chart's title to show the scale.