

Chapter 3

Show and Tell: Graphing Data

In This Chapter

- ▶ Introducing graphs
 - ▶ Working with Excel's graphics capabilities
 - ▶ Creating graphs for statistical work
-

The visual presentation of data is extremely important in statistics. Visual presentation enables you to discern relationships and trends you might not see if you just look at numbers. Visual presentation helps in another way: It's valuable for presenting ideas to groups and making them understand your point of view.

Graphs come in many varieties. In this chapter, I explore the types of graphs you use in statistics and when it's advisable to use them. I also show you how to use Excel to create those graphs.

Why Use Graphs?

Suppose you have to make a pitch to a Congressional committee about commercial space revenues in the early 1990s.

Which would you rather present? The data in Table 3-1, or the graph in Figure 3-1 that shows the same data? (The data, by the way, are from the U.S. Department of Commerce, via the Statistical Abstract of the U.S.)

<i>Industry</i>	<i>1990</i>	<i>1991</i>	<i>1992</i>	<i>1993</i>	<i>1994</i>
Commercial Satellites Delivered	1,000	1,300	1,300	1,100	1,400
Satellite Services	800	1,200	1,500	1,850	2,330
Satellite Ground Equipment	860	1,300	1,400	1,600	1,970
Commercial Launches	570	380	450	465	580
Remote Sensing Data	155	190	210	250	300
Commercial R&D Infrastructure	0	0	0	30	60
Total	3,385	4,370	4,860	5,295	6,640

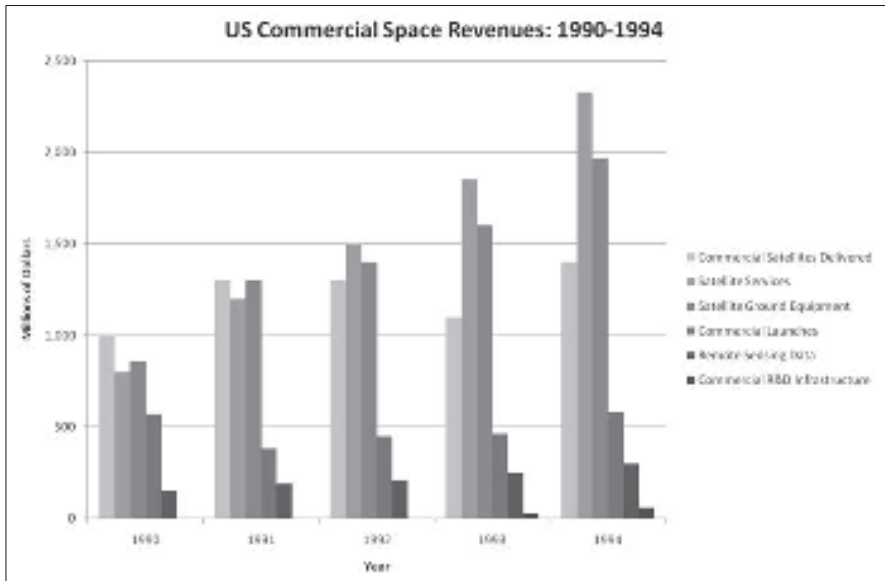


Figure 3-1:
Graphing
the data in
Table 3-1.

Which one would have a greater and more lasting impact? Although the table is certainly informative, most would argue that the graph gets the point across better and more memorably. Eyes that glaze over when looking at numbers often shine brighter when looking at pictures.

The graph shows you trends you might not see as quickly on the table. (Satellite services rose fastest. Commercial launches, not so much.) Bottom line: Tables are good, graphs are better.

Graphs help bring concepts to life that might otherwise be difficult to understand. In fact, I do that throughout the book. I illustrate points by . . . well . . . illustrating points!

Some Fundamentals

Like the graph in Figure 3-1, most graph formats have a horizontal axis and a vertical axis. The *pie graph*, a format I show you later in this chapter, does not. By convention, the horizontal axis is also called the *x-axis* and the vertical axis is also called the *y-axis*.

Also, by convention, what goes on the horizontal axis is called the *independent variable* and what goes on the vertical axis is called the *dependent variable*. One of Excel's graph formats reverses that convention, and I bring that to your attention when I cover it.



Just to give you a heads up, Excel calls that reversed-axis format a *bar graph*. You might have seen the graph in Figure 3-1 referred to as a bar graph. So have I. Excel calls Figure 3-1 a *column graph*, so I use “columns” from here on.

Getting back to “independent” and “dependent,” those terms imply that changes in the vertical direction depend (at least partly) on changes in the horizontal direction.

Another fundamental principle of creating a graph: Don't wear out the viewer's eyes! If you put too much into a graph in the way of information or special effects, you defeat the whole purpose of the graph.

For example, in Figure 3-1 I had to make some choices about filling in the columns. Color-coded columns would have been helpful, but the page you're looking at only shows black, white, and shades of gray.

A lot of graph creation comes with experience, and you just have to use your judgment. In this case, my judgment came into play with the horizontal gridlines. In most graphs, I prefer not to have them. Here, they seem to add

structure and help the viewer figure out the dollar value associated with each column. But then again, that's just my opinion.

Excel's Graphics Capabilities

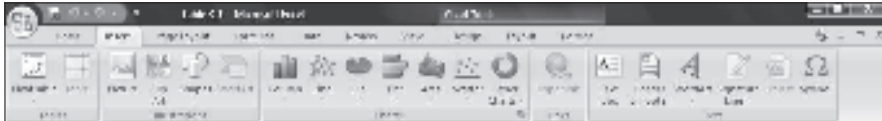
As I mention a couple of paragraphs ago, the graph in Figure 3-1 is a column graph. It's one of many types of graphs you can create with Excel. Of all the graphics possibilities Excel provides, however, only a few are useful for statistical work. Those are the ones I cover in this chapter.

In addition to the column graph, I show you how to create pie graphs, bar graphs, line graphs, and scatter plots. Excel refers to each one as a *chart* rather than a "graph." In this chapter, I use the two terms interchangeably.

Inserting a chart

In Excel 2007, you *insert* a chart into a spreadsheet. This immediately clues you that the chart creation tools are in the Charts area of the Insert tab. (See Figure 3-2.)

Figure 3-2:
The Charts
area of the
Insert tab.



In Excel's past a Chart Wizard guided you step by step through chart creation. The chart took shape within the wizard as you worked. You didn't see the full chart on the worksheet (or on a separate page) until you closed the wizard.

Excel 2007 turns the process inside out and simplifies it. Instead of making decisions within the wizard (which is no longer available), you immediately insert a chart into your spreadsheet. Then you use Design and Layout capabilities to modify the chart and make it look just the way you want it.

You follow these steps:

1. **Enter your data into a worksheet.**
2. **Select the data that go into the chart.**

3. In the Charts area of the Insert tab, select the chart type.

When you select a chart type, a box opens that presents a variety of subtypes. Choose one and Excel creates a chart in your worksheet.

4. Modify the chart.

Click on the chart, and Excel adds a Design tab and a Layout tab to the Ribbon. These tabs allow you to make all kinds of changes to your chart.

It's really that simple. The next section shows what I mean.



By the way, here's one more important concept about Excel graphics. In Excel, a chart is *dynamic*. This means that after you create a chart, changing its worksheet data results in an immediate change in the chart.

Becoming a Columnist

In this section, I show you how to create that spiffy graph in Figure 3-1. Follow these steps:

1. Enter your data into a worksheet.

Figure 3-3 shows the data from Table 3-1 entered into a worksheet.

Figure 3-3:
Table 3-1
data entered
into a
worksheet.

	A	B	C	D	E	F
1		1950	1951	1952	1953	1954
2	Commercial Facilities Owned	1,050	1,300	1,300	1,150	1,400
3	Satellite Services	0	1,200	1,500	1,000	2,000
4	Satellite Ground Equipment	850	1,300	1,400	1,600	1,900
5	Commercial Launches	570	300	400	450	500
6	Remote Sensing Data	155	150	210	250	300
7	Commercial R&D Infrastructure	0	0	0	20	50
8	Total	2,325	4,070	4,050	3,250	5,640

2. Select the data that go into the chart.

I selected A1:F7. The selection includes the labels for the axes but doesn't include row G, which holds the column totals.

3. In the Charts area of the Insert tab, select the chart type.

For this example, the chart type is Column. Selecting Insert | Charts | Column opens the gallery in Figure 3-4. Here, you select the specific type of column chart for the data. I selected the first choice in the top row (Clustered Column).

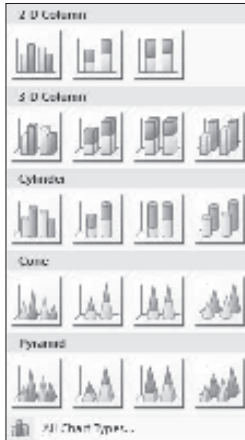


Figure 3-4:
The gallery
for a column
chart.

4. Modify the chart.

Figure 3-5 shows the resulting chart, as well as the Design tab and the Layout tab. As you can see, I have to do some heavyweight modifying. Why? Excel has guessed wrong about how I wanted to design the chart. It looks okay, but it's not. Rather than the years on the x-axis, Excel laid out the industry types. In other words, it interchanged the rows and columns.

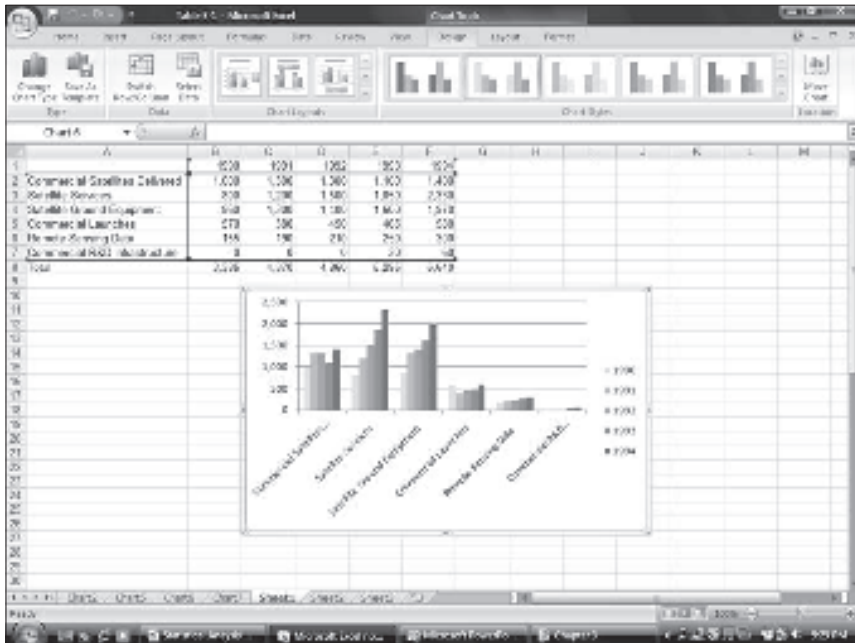


Figure 3-5:
The semi-
finished
graph —
based on a
bad guess
By Excel.

Fortunately, Excel provides a quick fix. Figure 3-5 shows the Design tab selected. In the Design | Data area, the choice on the left is Switch Row/Column. So . . . selecting Design | Data | Switch Row/Column does the trick.

Some work remains. The axes aren't labeled yet, and the graph has no title. Here's where the Layout tab comes into play. Figure 3-6 shows Layout | Axis Titles selected, along with the drop-down menu that allows you to add the title for each axis. Primary Horizontal Axis Title and Primary Vertical Axis Title provide options for laying out the axis titles. Layout | Chart Title does the same for the title of the chart.

Adding the titles finishes things off. The result looks like the chart in Figure 3-1.

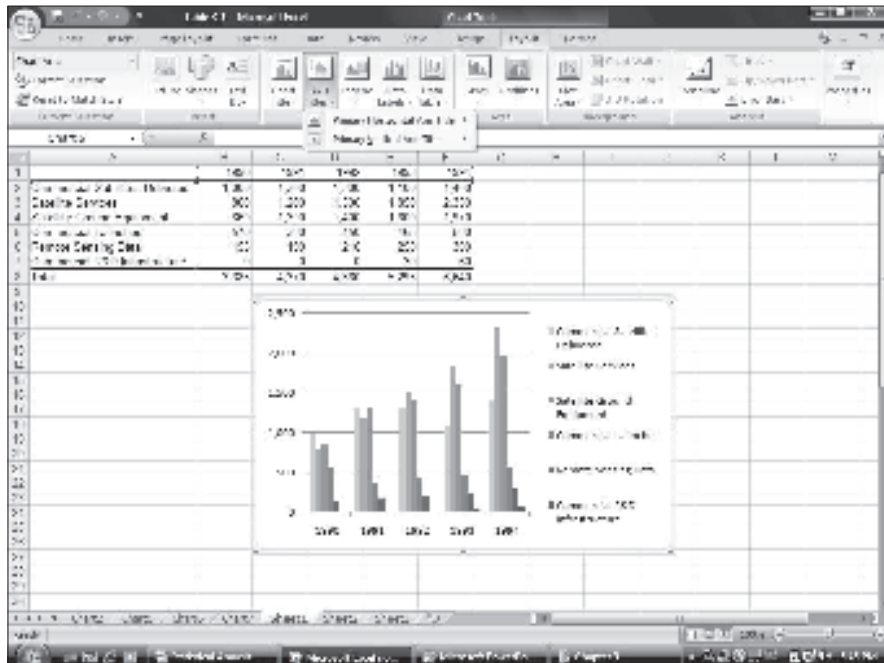


Figure 3-6: The Layout tab enables you to add titles.

Stacking the columns

If I had selected Column's second subtype — Stacked Column — I would have created a set of columns that presents the same information in a slightly different way. Each column represents the total of all the data series at a point on the x-axis. Each column is divided into segments. Each segment's size is proportional to how much it contributes to the total. Figure 3-7 shows this.

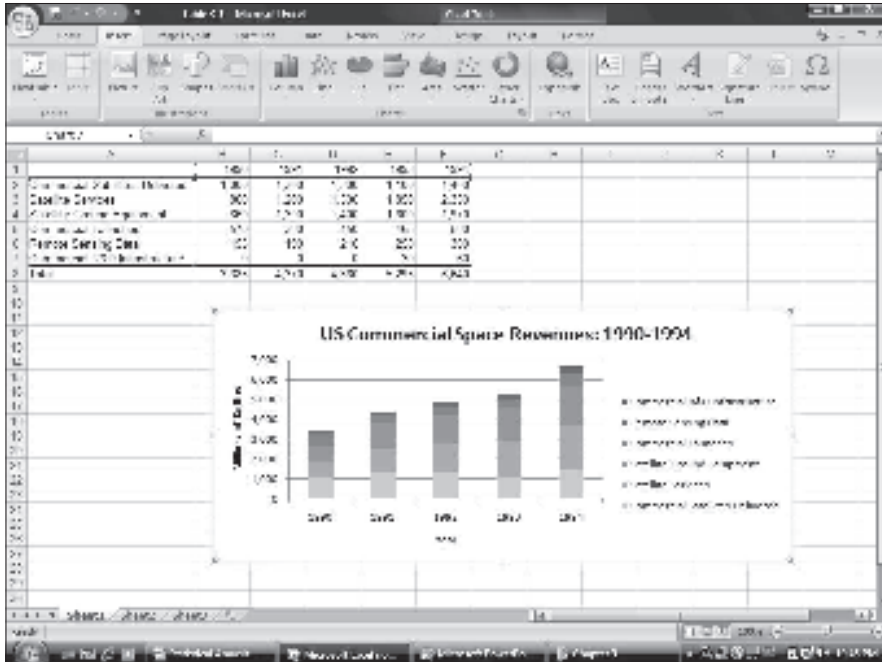


Figure 3-7:
A stacked
column
graph of
the data in
Table 3-1.

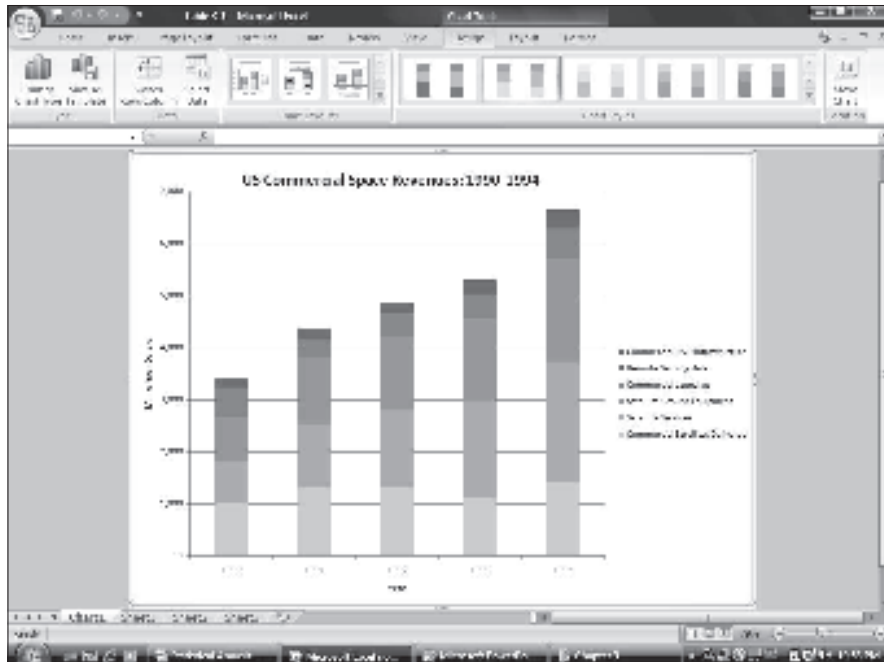
Notice that the data series are in reverse order from the way they're set up in the first column graph. Excel sets them up in this order for the stacked columns, and in the other order for the clustered columns.

I inserted each graph into the worksheet. Excel also allows you to move a graph to a separate page in the workbook. Select Design | Location | Move Chart (it's on the extreme right of the Design tab) to open the Move Chart dialog box. Click the New Sheet radio button to add a worksheet and move the chart there. Figure 3-8 shows how the chart looks on its own page.

In Appendix C, by the way, I show you another use for the stacked column chart.

This is a nice way of showing percentage changes over the course of time. If you just want to focus on percentages in one year, another type of graph is more effective. I discuss it in a moment, but first I want to tell you . . .

Figure 3-8:
The stacked
column
chart on its
own
worksheet.



One more thing

Statisticians often use column graphs to show how frequently something occurs. For example, in a thousand tosses of a pair of dice how many times does a 6 come up? How many tosses result in a 7? The x-axis shows each possible outcome of the dice-tosses, and the heights of the columns represent the frequencies. Whenever the heights represent frequencies, your column graph is a *histogram*.

It's easy enough to use Excel's graphics capabilities to set up a histogram, but Excel makes it easier still. Excel provides a data analysis tool that does everything you need to create a histogram. It's called — believe it or not — Histogram. You provide an array of cells that hold all the data — like the outcomes of many dice-tosses, and an array that holds a list of intervals — like the possible outcomes of the tosses (the numbers 2–12). Histogram goes through the data array, counts the frequencies within each interval, and then draws the column graph. I describe this tool in greater detail in Chapter 7.

Slicing the Pie

On to the next chart type. To show the percentages that make up one total, a pie graph gets the job done effectively.

Suppose you want to focus on the U.S. commercial space revenues in 1994 — that is, on the last column of data in Table 3-1. You'll catch people's attention if you present the data in the form of a pie graph, like the one in Figure 3-9.

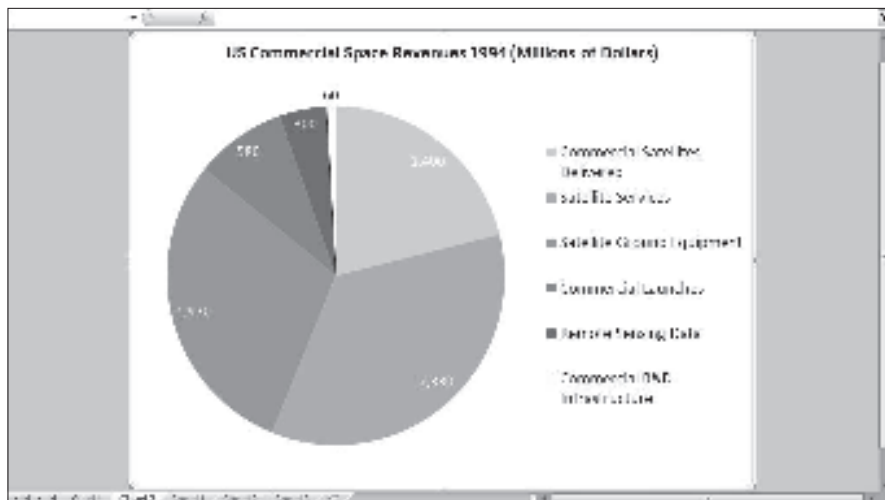


Figure 3-9:
A pie graph
of the last
column
of data in
Table 3-1.

Here's how to create this graph:

- 1. Enter your data into a worksheet.**

Pretty easy, as I've already done this.

- 2. Select the data that go into the chart.**

I want the names in column A and the data in column F. The trick is to select column A (cells A2 through A7) in the usual way and then press and hold the CTRL key. While holding this key, drag the cursor through F2 through F7. Voilà — two nonadjointing columns are selected

- 3. In the Charts area of the Insert tab, select the chart type.**

I selected Insert | Pie and then chose the first subtype.

4. Modify the chart.

Figure 3-10 shows the initial pie chart on its own page. To get it to look like Figure 3-9, I had to do a lot of modifying.

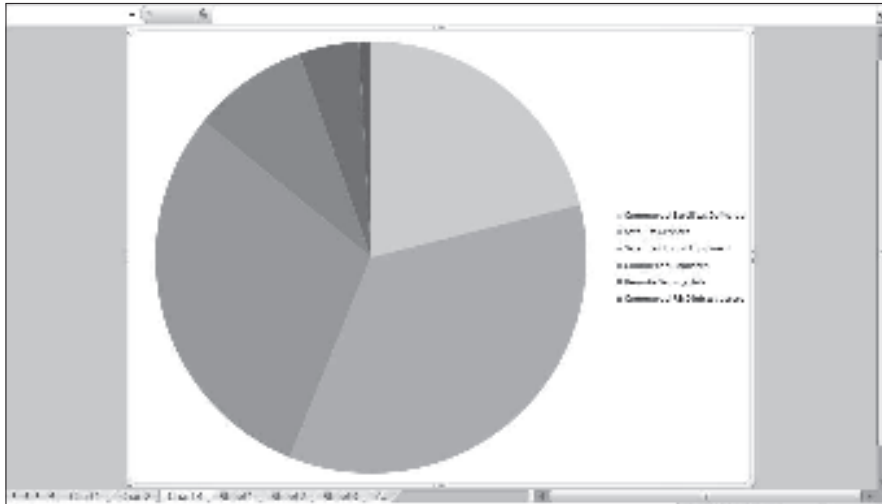
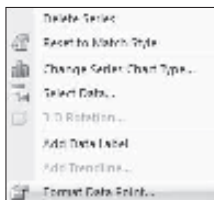


Figure 3-10:
The initial
pie chart
on its own
page.

The little slice filled in black represents 1 percent of the pie (Commercial R&D Infrastructure) and might be hard to see. I changed the fill color from black to white and added a border. How? I clicked on that slice and several slices were selected. Clicking again isolated it. Then I right-clicked to open the menu in Figure 3-11. Choosing Format Data Point opens the Format Data Point dialog box (Figure 3-12). I worked with Fill and Border to change the slice to a white fill with a black border.

Figure 3-11:
Right-
clicking an
isolated pie
chart slice
opens this
menu.



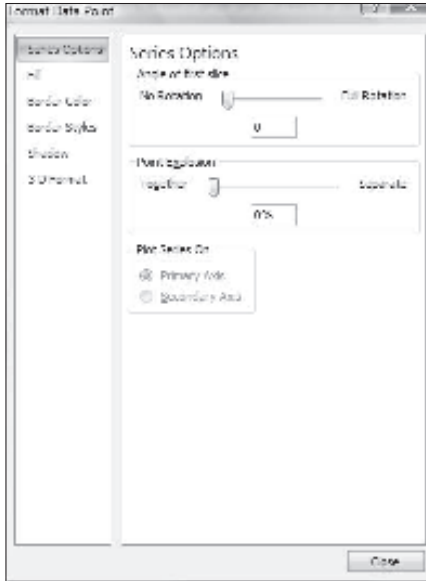


Figure 3-12:
The Format
Data Point
dialog box.

I selected Layout | Data Labels | Best Fit to add the data to each slice.

With the data labels selected, I right-clicked to open a couple of menus (Figure 3-13) that enabled me to manipulate the color and size of the data label font. After I made them all white, the label outside the small slice became invisible, but right-clicking in its area allowed me to reset its font to black. Right-clicking on the legend brings up the same menus for modifying the size of the font in the legend.

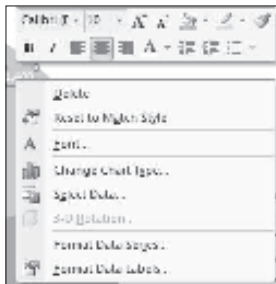


Figure 3-13:
Menus for manipulat-
ing the color and
size of
the data
label font.

Pulling the slices apart

One variant of the pie chart is to explode the slices. I'm not particularly fond of this type of graph, but you might be. In some circumstances, it might come in handy.

One of the nice things about Excel's graphics capabilities is that you can "what-if" to your heart's content. So . . . after I finish creating the pie chart, I can explode it. To do that, I click on the chart and select Design | Change Chart Type. This opens the Change Chart Type dialog box shown in Figure 3-14. Selecting the pie chart subtype that separates the slices (Exploded Pie) creates the chart in Figure 3-15.

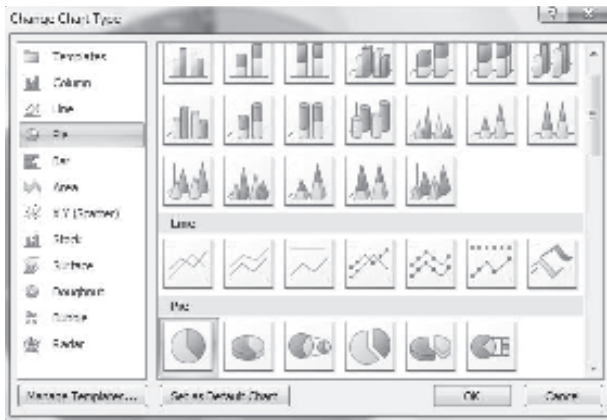


Figure 3-14:
The Change Chart Type dialog box.

Whenever you set up a pie graph — whether intact or exploded — always keep in mind

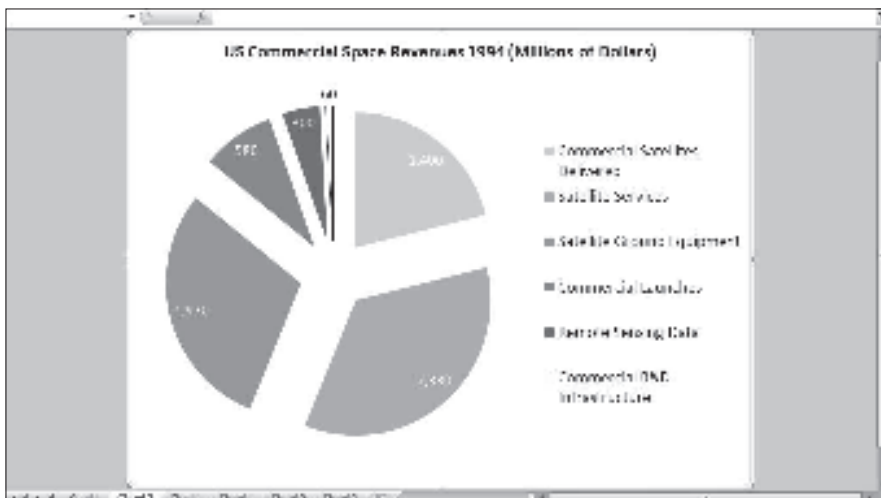


Figure 3-15:
The exploded version of Figure 3-9.

A word from the wise

Social commentator, raconteur, and former baseball player Yogi Berra once went to a restaurant and ordered a whole pizza.

“How many slices should I cut,” asked the waitress, “four or eight?”

“Better make it four,” said Yogi, “I’m not hungry enough to eat eight.”

Yogi’s insightful analysis leads to a useful guideline about pie graphs: They’re more digestible if they have fewer slices. If you cut a pie graph too fine, you’re likely to leave your audience with information overload.

Drawing the Line

In the preceding example, I focused on one column of data from Table 3-1. In this one, I focus on one row. The idea is to trace the progress of one space-related industry across the years 1990–1994. In this example, I graph the revenues from Satellite Services. The final product, shown on its own page, is Figure 3-16.

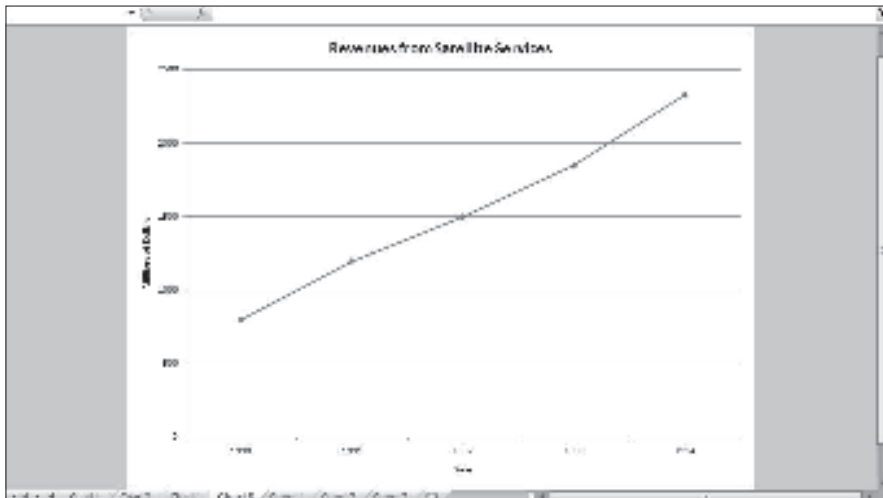


Figure 3-16: A line graph of the second Row of Data in Table 3-1.

A line graph is a good way to show change over time, when you aren’t dealing with too many data series. If you try to graph all six industries on one line graph, it begins to look like spaghetti.

How do you create a graph like Figure 3-16? Follow along:

1. Enter your data into a worksheet.

Once again, already done.

2. Select the data that go into the chart.

For this example, that's cells B3 through F3.

Whoa! Did I forget something? What about that little trick I showed you before where you hold down the CTRL key and select additional cells? Couldn't I do that and select the top row of years for the x-axis?

Nope. Not this time. If I do that, Excel thinks 1990, 1991, 1992, 1993, and 1994 are just another series of data points to plot on the graph. I'll show you another way to put those years on the x-axis.

3. In Charts area of the Insert tab, select the chart type.

This time, it's Insert | Line and then the fourth subtype, Line with Markers. This creates the chart in Figure 3-17.

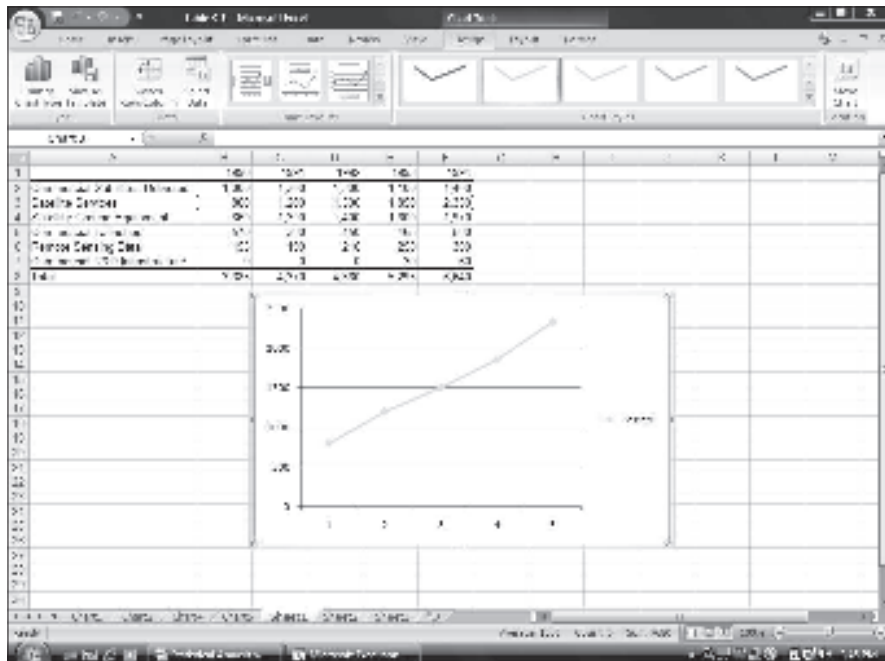


Figure 3-17:
Insert | Line
| Line With
Markers
creates the
line chart.

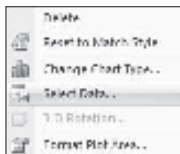
4. Modify the chart.

The line on the chart is a little hard to see. Selecting Design | Chart Styles and the leftmost option makes the line darker.

Next, I added the titles for the chart and for the axes. Clicking the down arrow of Design | Chart Layouts opens a box with 12 possibilities. Selecting the tenth one creates text boxes for the titles on the chart.

I still have to put the years on the x-axis. To do this, I right-clicked inside the chart to open the menu in Figure 3-18.

Figure 3-18:
Right-clicking inside the chart opens this menu.



Clicking Select Data opens the Select Data Source dialog box. (See Figure 3-19.) In the box labeled Horizontal (Category) Axis Labels, clicking the Edit button opens the Axis Labels dialog box (Figure 3-20). A blinking cursor in the Axis label range box shows it's ready for business. Selecting cells B1 through F1 and clicking OK closes this dialog box. Clicking OK closes the Select Source dialog box and puts the years on the x-axis.

Figure 3-19:
The Select Source dialog box.



One more modification completes the graph. The legend (showing “Series 1”) isn’t necessary. Right-click on it and choose Delete from the pop-up menu.

Figure 3-20:
The Axis
Labels
dialog box.



Passing the Bar

Excel's bar chart is a column chart laid on its side. This is the one that reverses the horizontal-vertical convention. Here, the vertical axis holds the independent variable, and it's referred to as the x-axis. The horizontal axis is the y-axis and it tracks the dependent variable.

When would you use the bar graph? This type of graph fits the bill when you want to make a point about reaching a goal, or about the inequities in attaining one.

Table 3-2 shows the data on (what I feel, anyway) is an important social issue. The data, from the U.S. Census Bureau (via the U.S. Statistical Abstract), are for the year 2000. *Percent* means the percentage of children in each income group.

<i>Family Income</i>	<i>Percent</i>
Under \$15,000	7.7
\$15,000-\$19,999	12.9
\$20,000-\$24,999	15.2
\$25,000-\$34,999	21.0
\$35,000-\$49,999	31.8
\$50,000-\$74,999	39.9
Over \$75,000	51.7

The numbers in the table are pretty dramatic. Casting them into a bar chart renders them even more so, as Figure 3-21 shows.

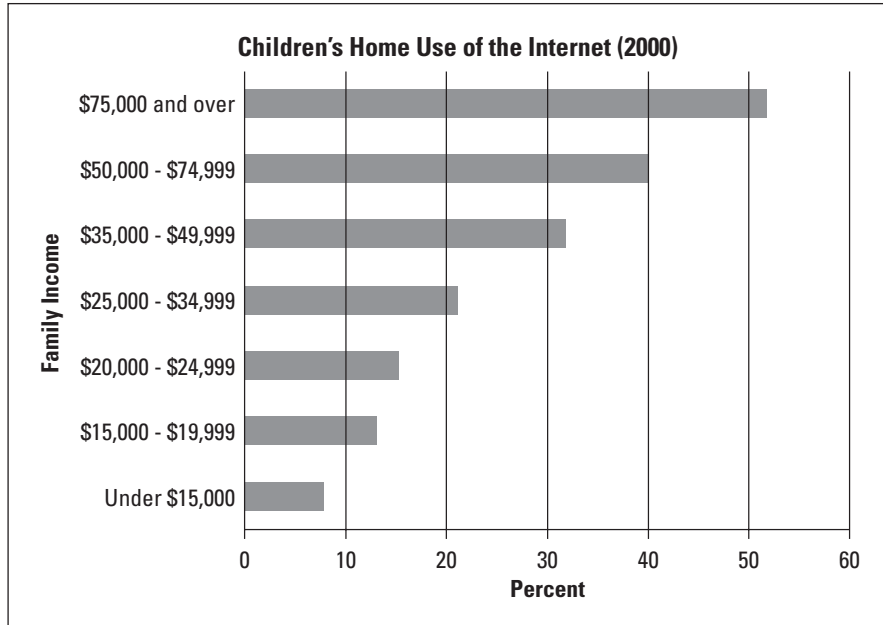


Figure 3-21:
A bar chart
of the data
in Table 3-2.

To create this graph, follow these steps:

1. Enter your data into a worksheet.

Figure 3-22 shows the data entered into a worksheet.

	A	B
1	Family Income	Percent
2	Under \$15,000	7.7
3	\$15,000-\$19,999	12.9
4	\$20,000-\$24,999	15.2
5	\$25,000-\$34,999	21.0
6	\$35,000-\$49,999	31.0
7	\$50,000-\$74,999	38.9
8	\$75,000 and over	51.7
9		

Figure 3-22:
Table 3-2
data in a
worksheet.

2. Select the data that go into the chart.

For this example, the data are cells A1 through B8.

3. In the Charts area of the Insert tab, select the chart type.

I clicked the down arrow for Insert | Bar and selected the first subtype (Clustered Bar). The result appears in Figure 3-23. The title is wrong and the legend is unnecessary, but I deal with that in Step 4.

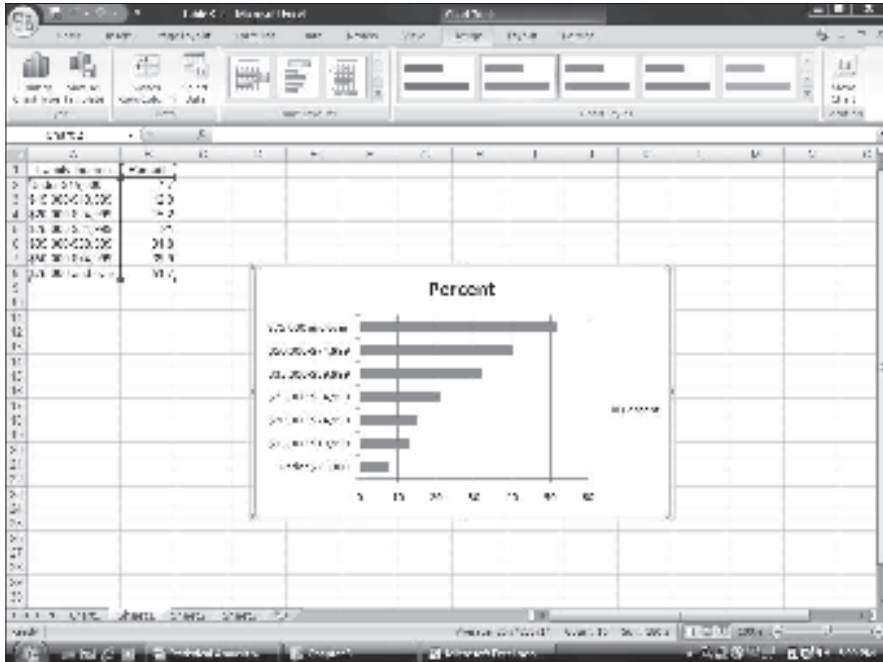


Figure 3-23:
The Initial
Excel bar
chart.

4. Modify the chart.

The first modification is to select a layout. This is a little confusing, because Design | Chart Layouts holds the desired look, rather than a ribbon area within the Layout tab. Go figure. The trick is to select a layout that puts text boxes for the titles onto the graph. The option that does this (Layout 8) is shown selected in Figure 3-24. (The other way to do this is with Layout | Labels as I showed you earlier.)

Figure 3-24:
The layout
for adding
axis titles
and chart
title to the
bar chart.



After entering the labels, I deleted the legend. Then I darkened the bars via Design | Chart Styles. Right-clicking on each label brings up the menu for increasing the font size.

The Plot Thickens

You use an important statistical technique called *linear regression* to determine the relationship between one variable, x , and another variable, y . For more information on linear regression, see Chapter 14.

The basis of the technique is a graph that shows individuals measured on both x and y . The graph represents each individual as a point. Because the points seem to scatter around the graph, the graph is called a *scatterplot*.

Suppose you're trying to find out how well a test of aptitude for sales predicts salespeople's productivity. You administer the test to a sample of salespersons and you tabulate how much money they make in commissions over a two-month period. Each person's pair of scores (test score and commissions) locates him or her within the scatterplot.

To create a scatterplot, follow the steps:

1. Enter your data into a worksheet.

Figure 3-25 shows the entered data.

	A	B	C
1	Salesperson	Aptitude Score	Commissions (in thousands of Dollars)
2	1	14	60
3	2	34	45
4	3	23	38
5	4	46	35
6	5	34	41
7	6	47	50
8	7	51	48
9	8	56	60
10	9	39	48
11	10	45	50
12	11	40	48
13	12	59	48
14	13	39	47
15	14	56	61
16	15	22	38
17	16	58	64
18	17	42	54
19	18	53	50
20	19	39	41
21	20	50	60
22			

Figure 3-25:
Scatterplot
data.

2. Select the data that go into the chart.

Figure 3-25 shows the selected cells — B2 through C21. (Including B1 creates the same chart, but with an incorrect title.) The cells in Column A are just placeholders that organize the data.

3. In the Charts area of the Insert tab, select the chart type.

I chose the first subtype of Insert | Scatter, resulting in the chart shown in Figure 3-26.

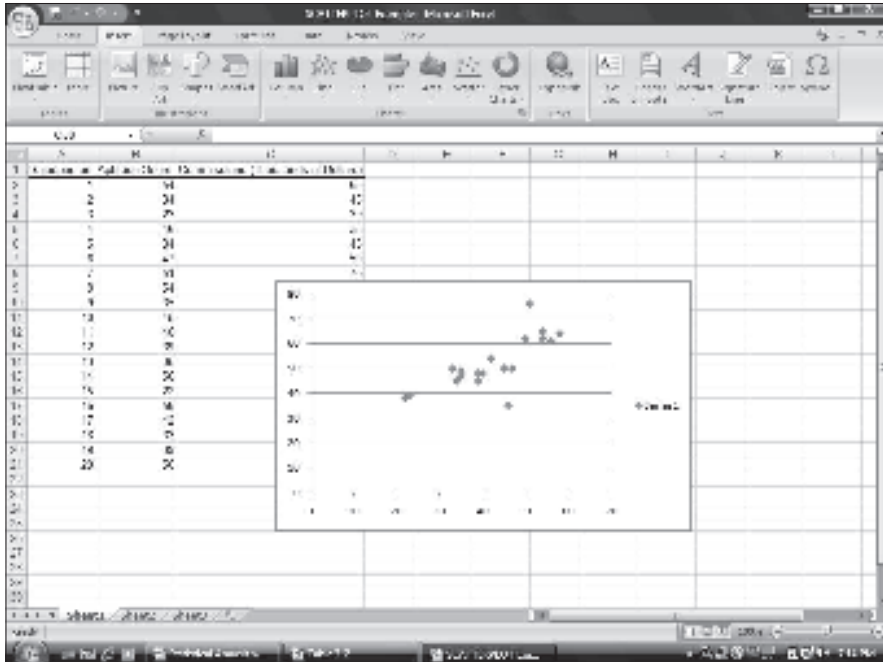


Figure 3-26:
The initial
scatterplot.

4. Modify the chart.

Clicking Design | Chart Layouts and selecting the first layout puts the text box on the chart for each title. I typed the title for the chart and the titles for the axes and then deleted the legend. The result is the scatterplot in Figure 3-27.

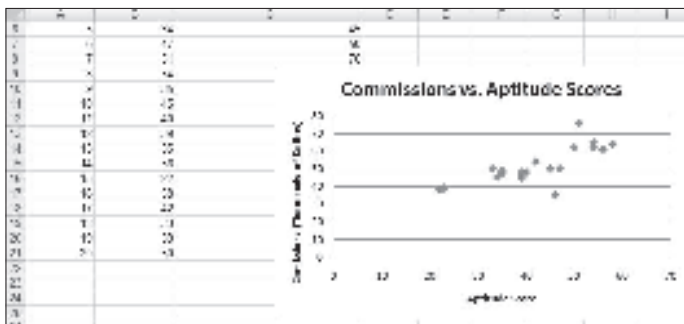
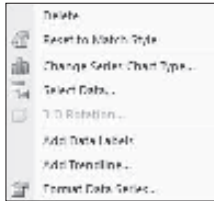


Figure 3-27:
The almost-
finished
scatterplot.

For the other graphs, that would just about do it, but this one's special. Right-clicking any of the points in the scatterplot opens the pop-up menu in Figure 3-28.

Figure 3-28:
Right-clicking any point on the scatterplot opens this menu.



Selecting Add Trendline opens the Format Trendline dialog box. I selected the Linear radio button and clicked the two bottom checkboxes. They're labeled Display Equation on Chart and Display R-Squared Value on Chart (Figure 3-29).

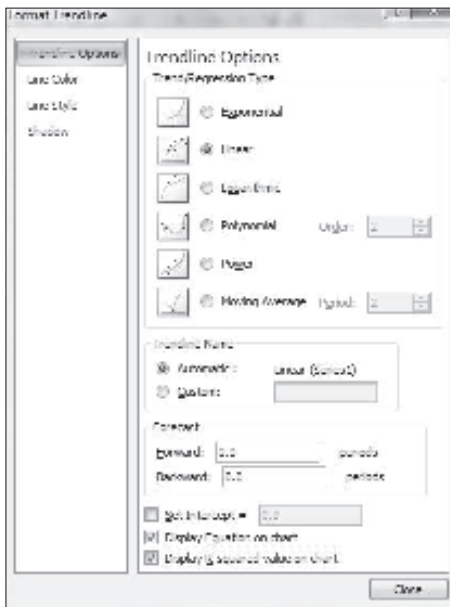


Figure 3-29:
The Format Trendline dialog box.

Clicking OK closes the Format Trendline dialog box. A couple of additional items are now on the scatterplot, as Figure 3-30 shows. A line passes through the points. Excel refers to it as a *trendline*, but it's really called a *regression line*. A couple of equations are there, too. (For clarity, I dragged them from their original locations.) What do they mean? What are those numbers all about?

You'll just have to read Chapter 14 to find out.

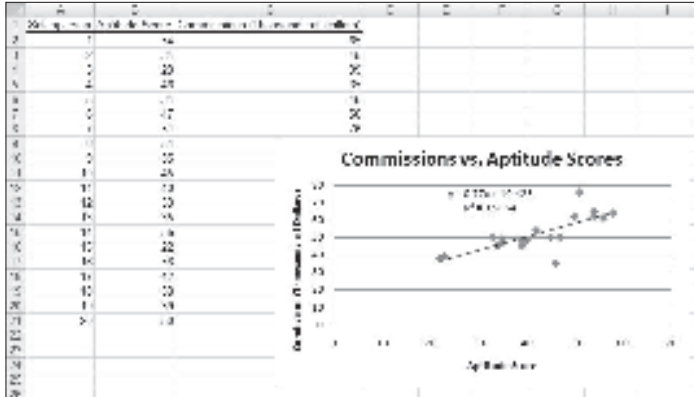


Figure 3-30:
The scatterplot, with additional information.

