An Introduction to Excel Pivot Tables

Pivot Table Introduction

Microsoft introduced Pivot Tables into Excel with version 5. Pivot Tables replaced Excel's older cross-tabulation feature. A Pivot Table lets you display the data contained in a column of an Excel list (database) by means of subtotals (or other calculations) that are defined by another column in the same list. The other calculations might be averages, counts, percentages, standard deviations, and so on.

Why organize list data into a Pivot Table?

Three key reasons for organizing data into a Pivot Table are:

- To summarize the data contained in a lengthy list into a compact format
- To find relationships within the data that are otherwise hard to see because of the amount of detail
- To organize the data into a format that's easy to chart

Below is a simple example of how putting data in a Pivot Table can be useful.

Region	Year	Product	Units Sold
Southeast	1997	Brie	63
Northwest	1998	Gouda	151
Southwest	1998	Brie	87
Northeast	1998	Gouda	128
Southeast	1997	Gouda	72
Northeast	1998	Brie	110
Northwest	1997	Brie	137
Southeast	1998	Gouda	43
Northwest	1997	Brie	108
Northwest	1998	Gouda	49
Southwest	1997	Brie	100
Northeast	1997	Gouda	39

Sum of Units Sold		
Region	Product	Total
Northeast	Brie	110
	Gouda	128
Northwest	Brie	137
	Gouda	151
Southeast	Brie	63
	Gouda	72
Southwest	Brie	87
Grand Total		748

At left in the spreadsheet illustration above is a simple list, or Excel database. Even looking at this simple, short list it's difficult to discern patterns in the data. For example, it takes a bit of study to see that the number of Units Sold in the Northeast region is much greater than the number of Units Sold for the Southwest region. Or to find out that Gouda outsells Brie in the Northwest. Questions of this type that you might have about the data can be answered, but only with some effort.

By contrast, the Pivot Table at the right simplifies and summarizes the data to make relationships and patterns obvious. And, if you had much more data in the list at left (perhaps with many additional entries for each region), you could still achieve a condensed Pivot Table summary the same size as the one at right. The Pivot Table also allows you to include or exclude whatever list data you like. For example, I've excluded the year information in the sample Pivot Table above. I could easily add it back in.

You can easily chart the data organized into a Pivot Table, while to chart the data in the list at left you'd first need to restructure the data and obtain the sum for each region. The Pivot Table simplifies the process because it obtains subtotals automatically and puts them in a range you can immediately use for charting.
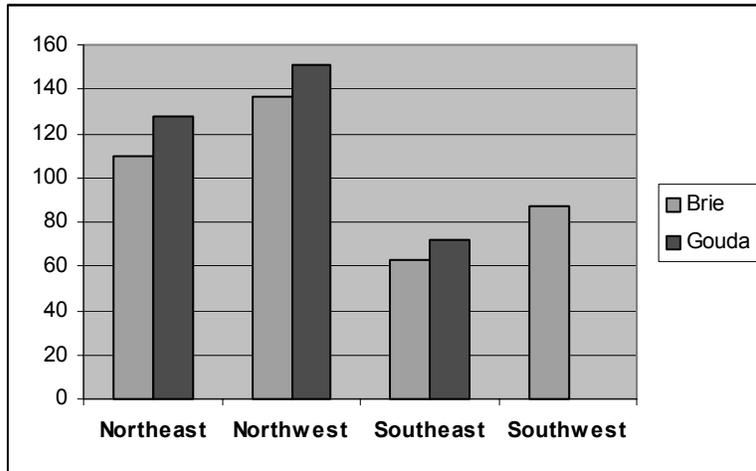

Using the Chart Wizard, it's easy to get a summary, graphical view of your data using the Pivot Table as your tool to organize and summarize the data.

What's Required to Construct a Pivot Table?
To create a Pivot Table you need to identify these two elements in your data:

- a *data field*, where the data field is the variable you want to summarize

- a *row* and/or *column field* where the row and/or column fields are the variables that will "control" the data summary

To create the Pivot Table, invoke the Pivot Table Wizard with the menu commands *Data, Pivot Table and Pivot Chart Report*. The Pivot Table Wizard leads you through three steps:
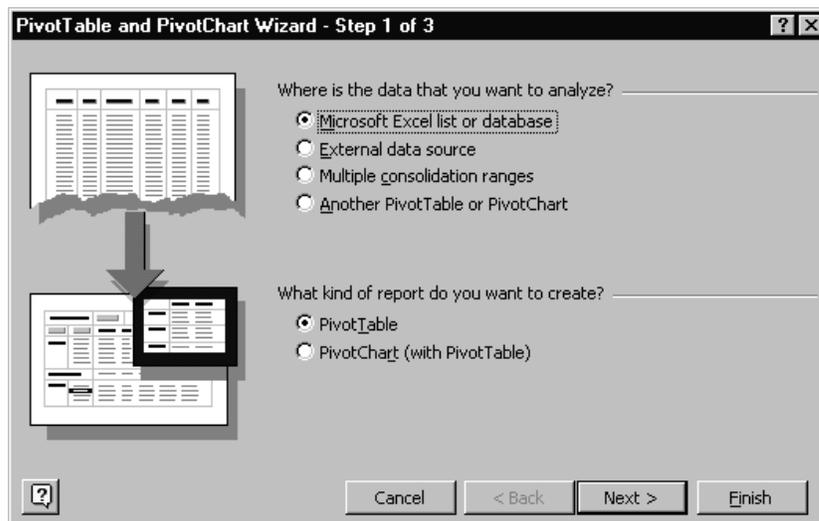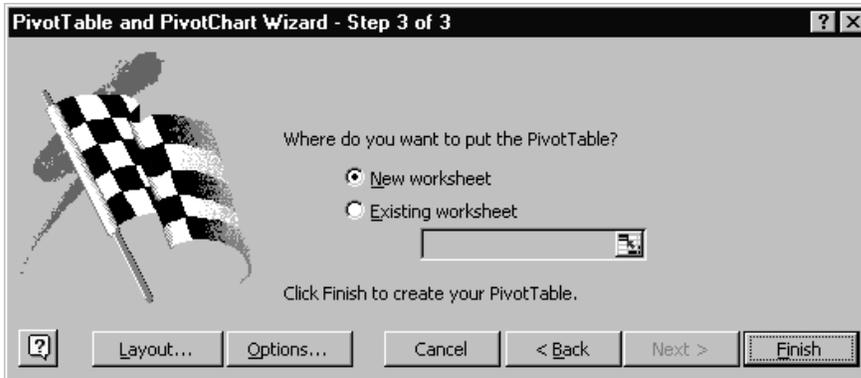

Step 1: Allows you to specify where your data is located and whether you want a chart as well as a table.

Most commonly you'll get your data from an Excel list that's part of the current worksheet.
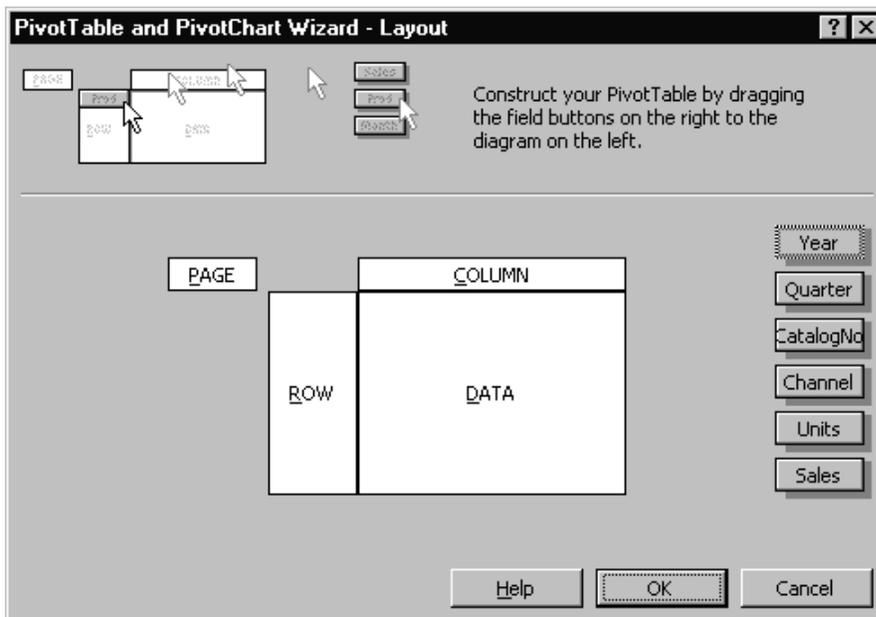
Step 2: Identifies the list range.

If your have the insertion point anywhere in the list when you start the Wizard, Excel defines the list range automatically.

PivotTable and PivotChart Wizard - Step 2 of 3

Where is the data that you want to use?

Range: Database

Browse...

Cancel | < Back | Next > | Finish

PivotTable and PivotChart Wizard - Step 3 of 3

Where do you want to put the PivotTable?

○ New worksheet

○ Existing worksheet

Click Finish to create your PivotTable.

Layout... | Options... | Cancel | < Back | Next > | Finish

Step 3: Creates the Pivot Table using its best guess as to layout.

Users of the previous version of the Excel Pivot Table will recall a step before this one that permits the user to determine the initial Pivot Table layout. That step is still available if one chooses the *Layout* button on Step 3 before clicking *Finish*. (Of course, you can return to the layout view at any time to make changes while working with the completed Pivot Table.)

If you choose the *Layout* button in Step 3 or invoke that view while working with the Pivot Table, you get a dialog that looks like this:

PivotTable and PivotChart Wizard - Layout

Construct your PivotTable by dragging the field buttons on the right to the diagram on the left.

| PAGE | COLUMN |
| ROW | DATA |

Year
Quarter
CatalogNo
Channel
Units
Sales

Help | OK | Cancel

Drag field buttons at right (representing variables from the list) to the layout diagram at left.

You need not use all the field buttons.

3

For example, I've used only four of the six possible field buttons in the layout at right.

You can include multiple field buttons in each of the layout locations.
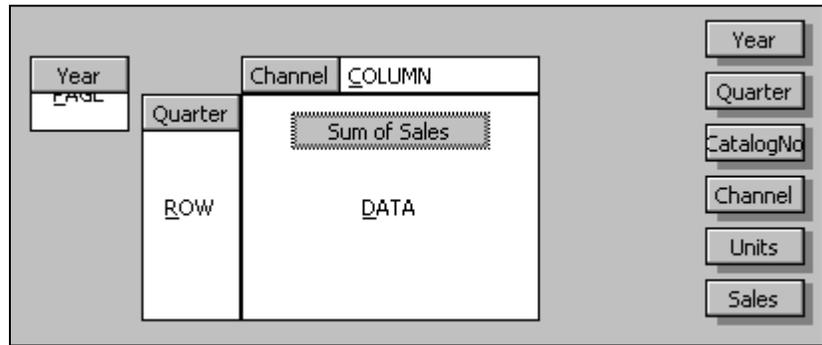
You can choose to locate your new Pivot Table on the sheet with the Excel list or on a new worksheet.

Once you've created your Pivot Table, you can modify it by using the buttons on the Pivot Table toolbar that automatically displays. Two of the buttons are especially key:

Return to layout.   *Edit the selected Pivot Table field.*
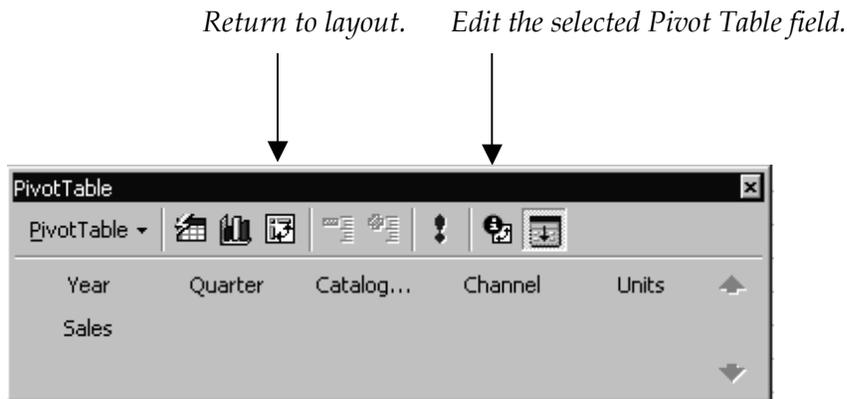
Although you may want to return to the Wizard layout view, you can drag and drop items onto and off of the Pivot Table and move them around in the layout by just using the mouse.

Note that all the field buttons are visible on the toolbar to allow easy dragging onto the Pivot Table, should you want to add more detail.

A drop-down menu (with some redundant options) is available from the first Pivot Table toolbar button on the left.

4
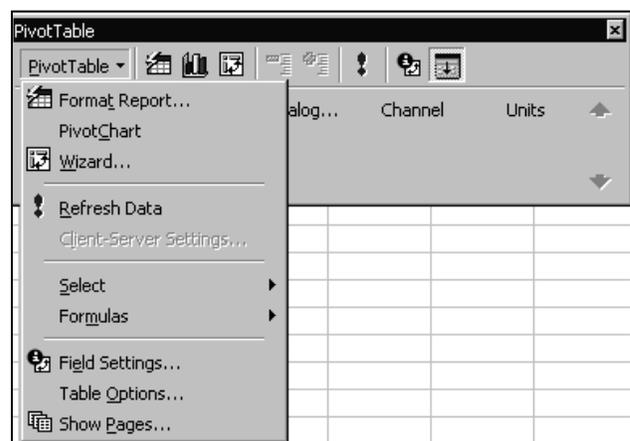
Pivot Table Row and Column Fields

A row field

A row field in a Pivot Table is a variable that takes on different values. For example, a row field might be "Manufacturer" and its values might be "Schwinn", "Cannondale", and "Omega".

The values a variable takes on are sometimes referred to as "items".

In the example below, for each value of the variable "Manufacturer", the Pivot Table displays a summary of the chosen data field in an adjoining column. The data field in this example is "Annual Sales" and the summary function is sum.

Sum of Annual Sales	
Manufacturer	Total
Cannondale	$823
Omega	$435
Schwinn	$500
Grand Total	$1,758

Notice that the Pivot Table uses the label "Sum of Annual Sales" to identify not only the data field (Annual Sales) but also the default summary operation (sum).

A column field

A Pivot Table column field works like a row field. A column field might be the variable "Year" with values ranging from 1995 to 1998. Data beneath each column in the Pivot Table is associated with the year at the head of the column.

Manufacturer	Year	Annual Sales
Schwinn	1995	$500
Cannondale	1995	$823
Omega	1995	$435
Schwinn	1996	$523
Cannondale	1996	$840
Omega	1996	$420
Schwinn	1997	$550
Cannondale	1997	$855
Omega	1997	$418
Schwinn	1998	$585
Cannondale	1998	$889
Omega	1998	$400

Sum of Annual	Year				
Manufacturer	1995	1996	1997	1998	Grand Total
Cannondale	823	840	855	889	3407
Omega	435	420	418	400	1673
Schwinn	500	523	550	585	2158
Grand Total	1758	1783	1823	1874	7238

The basic effect of row and column fields in a Pivot Table is that each value or *item* that the field takes on defines a different row or column. So if a list has a row field that takes on three items (Schwinn, Cannondale, Omega) and a column field that takes on four

items (1995, 1996, 1997, 1998), the Pivot Table has three rows and four columns, and therefore twelve summary cells (exclusive of the cells that hold Grand Totals and labels).

When you define a Pivot Table you *must* identify a data field, or the variable to summarize or Excel displays an error. By contrast, however, you need not define either a row or a column field. In this case, you won't get an error message but you won't get a very meaningful result either. Without a row and/or column "control", Excel will return a simple table since it will have no way to summarize the data.

If you specify a row field only, your Pivot Table will have one column that contains the values of the row field and another column to its right that contains the corresponding summary values for the data field. If you specify a column field only, your Pivot Table will have a top row that holds the values of the column field and a row beneath that holds the corresponding summary values of the data field.

Sum of Annual Sales	
Manufacturer	Total
Cannondale	3407
Omega	1673
Schwinn	2158
Grand Total	7238

Row-only Pivot Table

Sum of Annual Sales	Year				
	1995	1996	1997	1998	Grand Total
Total	1758	1783	1823	1874	7238

Column-only Pivot Table

The Pivot Table gets its name because it's easy to modify, or pivot, the positions of the Pivot Table fields. Imagine a point at the upper left-hand corner of the Pivot Table as the pivot point. You can drag and drop a row field up and to the right, making it a column field. And you can drag and drop a column field down and to the left, making it a row field. As mentioned earlier, although you can return to the Wizard's layout view to accomplish this layout change, it's easier to do it on the fly without using the Wizard.

If you have more than two variables in your list that you want to use as row or column fields you can define subsidiary or *inner* row or column fields.

Sum of Units Sold		
Region	Product	Total
Northeast	Brie	110
	Gouda	128
Northwest	Brie	137
	Gouda	151
Southeast	Brie	63
	Gouda	72
Southwest	Brie	87
Grand Total		748

In the example at left, "Product" is considered the *inner row field* and "Region" is considered the *outer row field*.

The effect of an inner field is to add another level of detail to the Pivot Table. Within one category, or value (such as Northeast) of the outer field, there are two inner field values (Brie and Gouda).

The data field is first summarized by the value of the outer row field. Within that first-level summary, the data field is further summarized by the corresponding values of the inner row field.

Inner and outer *column* fields work the same way as inner and outer *row* fields.

The Pivot Table Data Field

The data field is the variable that the Pivot Table summarizes. For each combination of values in the row and column fields, the data field takes on a different value and this value appears in the Pivot Table's cells.

There are many summarizing calculations available in a Pivot Table. Most often the data to be summarized is numeric; the default summary for numeric data is a *sum*. For example, your Pivot Table might sum Units Sold. If the data to be summarized is text data, not numeric, the default summary must be a *count*. For example, your Pivot Table might count of the number of people working in a department.

Excel provides a variety of ways to summarize data, including Average, Max, Min, Product, StdDev, and so on.
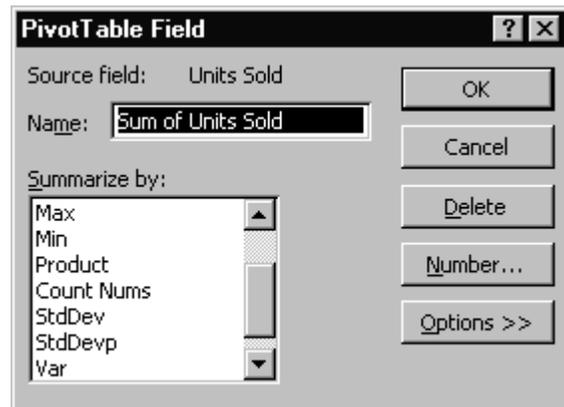
In addition, you can apply certain custom calculations to the data field, including running totals, percent of row, percent of column, and so on. If you need some calculation not provided within the Pivot Table, you can perform that calculation outside the context of the Pivot Table. Then you can include the calculated variable in your list when you start the Pivot Table Wizard. The Pivot Table also has an option to create a calculated field and a calculated item. To explore these options, check Excel's Pivot Table online help.


The Pivot Table Page Field

The page field operates like the row and column fields but provides a third dimension to your data. It allows you to add another variable to your Pivot Table without necessarily viewing all its values at the same time.

In the example at right, the page field is "Year" and is displayed in this illustration showing the three drop-down choices: (All), 1997, or 1998.

You can choose to view "(All)" year values in the Pivot Table, or you can select any year individually and see just that year's data displayed.

To create a "paged" Pivot Table, during the Pivot Table Wizard layout step, drag a button ("Year" in this illustration) to the "Page" element of the layout.
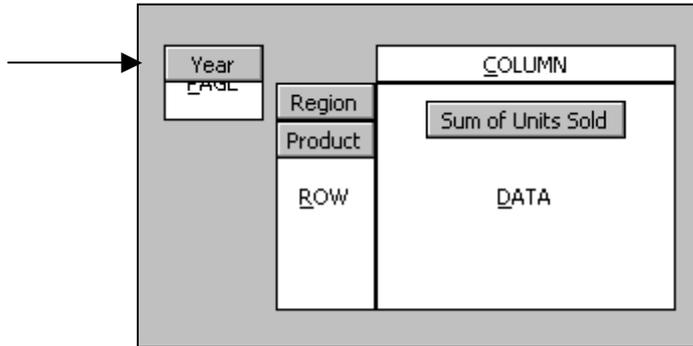
The alternative to a "paged" Pivot Table is to include the "Year" variable as a row or column variable. In the Pivot Table example at right, "Year" is a column variable. Because this Pivot Table layout is less summarized than if Year were a page value, it's a bit harder to read.

Sum of Units Sold		Year		
Region	Product	1997	1998	Grand Total
Northeast	Brie		110	110
	Gouda	39	128	167
Northeast Total		39	238	277
Northwest	Brie	245		245
	Gouda		200	200
Northwest Total		245	200	445
Southeast	Brie	63		63
	Gouda	72	43	115
Southeast Total		135	43	178
Southwest	Brie	100	87	187
Southwest Total		100	87	187
Grand Total		519	568	1087
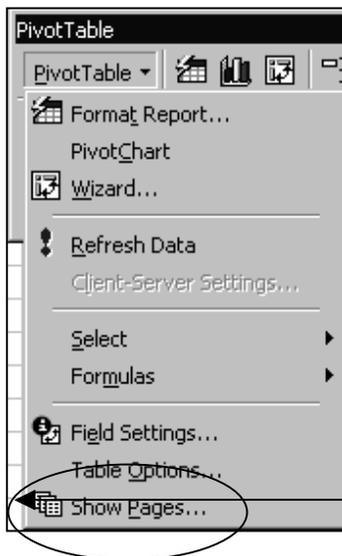
Note that if you include a page field in your Pivot Table you can choose to display the Pivot Table pages on separate worksheets. Click the *Show Pages*" button on the drop-down of the Pivot Table toolbar button to have Excel automatically replicate each page's data on a separate worksheet.

Determining Pivot Table Layout: Discrete vs. Continuous Variables

A *discrete variable* generally has a relatively small number of unique values. Typically, these unique values are different names. Examples of discrete variables are department names, model names, or customer names. Discrete values are most suitable as row and column variables in a Pivot Table, although they can also be used as data fields. If you use a discrete variable as a data field, however, the Pivot Table will only be able to display a summary by count.

A *continuous variable* can take on a large range of values. A variable is termed continuous when its values represent more or less of some quantity. Examples of continuous variables are profit margin, units sold, or daily precipitation. It's generally not a good idea to use a continuous variable as a row or column variable in a Pivot Table because such a variable can take on so many values that the Pivot Table would likely become impossibly large. For example, if you're analyzing income for 300 firms and you use firm income as a column variable in your Pivot Table, you'd be likely to run out of Excel worksheet display space. Normally, you'd use a continuous variable as the data field in a Pivot Table in order to see the sum or average or other summary calculation of its values for different levels of discrete variables. "Grouping" is one method available to reduce the amount of data displayed. Dates used as row or column headers, for example, are often grouped for this reason.

Formatting Data in the Pivot Table

No matter what the format of the data that's the source of your Pivot Table, the initial format of the data in a Pivot Table is "General". After you've created a Pivot Table you can use Excel's normal formatting commands to format cells. To retain formatting when you refresh the data in a PivotTable or change the layout, do the following:

- Click a cell in the PivotTable, and on the Pivot Table toolbar choose the *Pivot Table* button, then *Table Options*. In the "Format options" part of the dialog that displays, make sure the "Preserve formatting" check box is selected.
- Then, before you select the Pivot Table data you want to format, make sure that Pivot Table "Enable Selection" is activated (Choose the *Pivot Table* button on the toolbar, then *Select, Enable Selection*).

Without the steps above, if you make any change to your Pivot Table after you've applied formatting (even if the change is only to refresh the data), your formatting is lost.

A good method for formatting a Pivot Table *data field* is to use the formatting options built into the Pivot Table itself.

To do this:
1. Select any cell in the Pivot Table that represents its data field.
2. Click the Pivot Table Field button on the Pivot Table toolbar. The "Pivot Table Field" dialog displays.
3. Click the *Number* button on this dialog to open the "Format Cells" dialog.

4. Choose one of the available formats, click *OK* to close the "Format Cells" dialog and then *OK* again to close the "Pivot Table Field" dialog.

If you format your Pivot Table in this way, the formatting you select is retained even if you change the layout of the Pivot Table. This formatting is not available for row, column, or page fields, however.
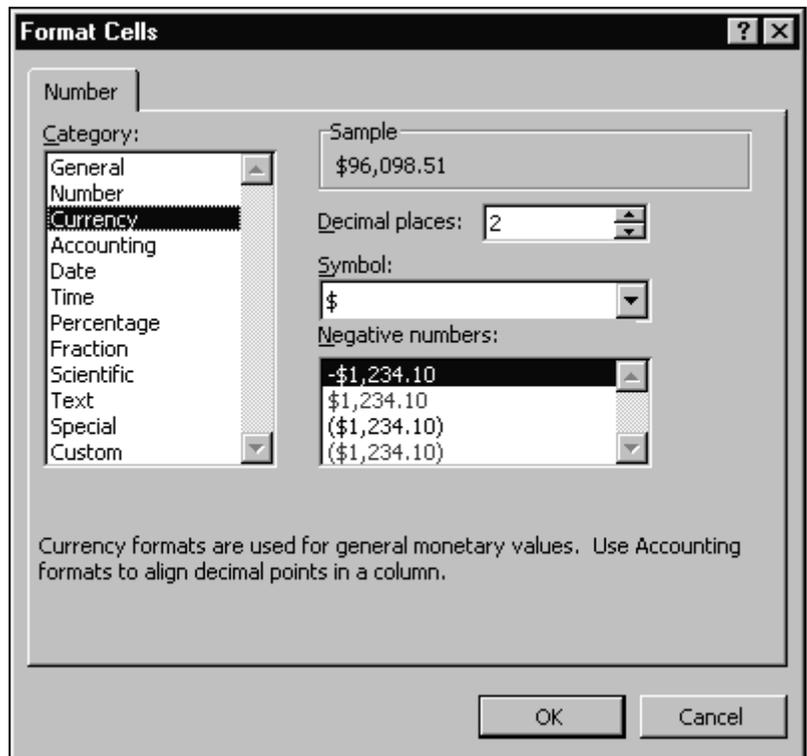
If you want to reformat the *entire* Pivot Table, consider using Excel's *Format, Autoformat* commands, particularly the "table" layouts. Any autoformat choice you make remains with your Pivot Table even if you refresh the data, pivot the table, or add or remove a field.
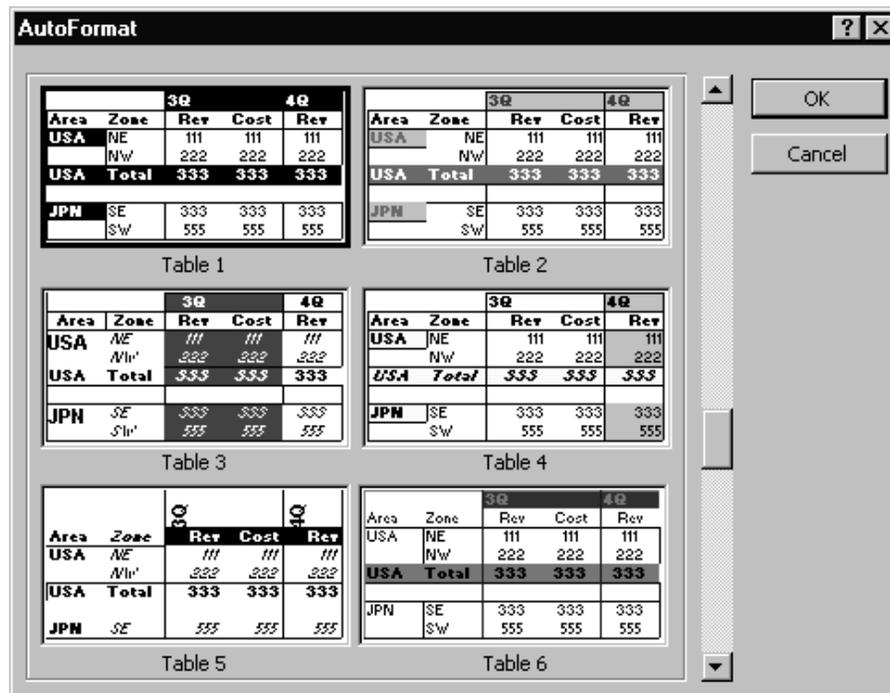
Refreshing Pivot Table Data

A Pivot Table isn't directly linked to its source data. Instead it's linked to a hidden cache that's build from the data source. If you create a Pivot Table from, for example, a worksheet list and then change the data in the list, the Pivot Table does not automatically update. To update the Pivot Table, click any cell within the Table and click the *Refresh Data* button (the red exclamation mark) on the Pivot Table toolbar. Then Excel gets the new data from your changed Excel list. As an alternative you can also right-click a cell in the Pivot Table and choose *Refresh Data* from the pop-up menu that appears.

Sorting Pivot Table Data

Excel usually displays a Pivot Table with field items in sorted order (ascending by rows for a row field and ascending by columns for a column field). However, the original sort order may change as you work with the Pivot Table, particularly if you have multiple row and/or column fields. Or, you may prefer to have your Pivot Table data sorted differently from the default. To sort the Pivot Table, click the button for the field you want to sort.

Year	(All)		
Sum of Units Sold	Product		
Region	Brie	Gouda	Grand Total
Southwest	187		187
Southeast	63	115	178
Northwest	245	200	445
Northeast	110	167	277
Grand Total	605	482	1087

For example, to sort the Region field at left, click the field button for the Region variable. The region values are selected.

Then to sort Region in ascending order invoke Excel's regular *Data, Sort* commands.

Year	(All)		
Sum of Units Sold	Product		
Region	Brie	Gouda	Grand Total
Northeast	110	167	277
Northwest	245	200	445
Southeast	63	115	178
Southwest	187		187
Grand Total	605	482	1087

More About Excel's Pivot Table

The information in this guide is a simple introduction to the power of Excel's Pivot Table. To learn more about using this useful tool, consult a good Excel reference or see Excel's online help.

End