

Chapter 07

Analyzing Data in Tables

Once the data has been transferred into a spreadsheet program, the specialized features of a spreadsheet program can be utilized to analyze the data and transform it into information.

Why Do I Need to Know This?

One of the dramatic changes that has occurred in the last forty years is the use of computers. Almost every job and many of our leisure activities now involve computers. Computers only know how to work with numbers but they can be very useful if you understand their capabilities, their limitations, and how to interact with them effectively. To understand those capabilities and limitations, you need to know some of the basics about how computers store numbers and how you can get those numbers out of them and into a form that humans can understand.

1 Analyzing Data in a Spreadsheet

Learning Objectives

1. Describe the parallel between the use of QBE to write SQL in Access and the use of filters in Excel. [7.1.1]
2. Describe the similarity between criteria in Access and filters in Excel. [7.1.2]
3. Identify uses of sorting for analyzing data. [7.1.3]

More people are familiar with spreadsheets like Microsoft Excel than with database programs like Microsoft Access so they prefer to use Excel to analyze data if possible.

Database management systems (DBMS) can store data in many related tables and the

SQL language has many options for retrieving the data. A DBMS is very fast at finding and retrieving individual records. For example, if an online store wanted to retrieve a customer credit card number and the shopping history associated with it, a database management system would be used. A spreadsheet can be used effectively to analyze the data to identify trends using charts, find extreme values by sorting, or to restrict the records shown using criteria. Excel has some of the same features as a DBMS but they often have different names. Applying a criteria to limit the rows displayed is called **filtering** in Excel.

Consider an example of how Excel can analyze data for a local business that is considering installation of a wind turbine but is concerned about the likelihood that a tornado would destroy it. The manager goes to a website that has a history of tornadoes, downloads the csv file, and then imports it into Excel. Excel recognizes the data as a table and formats the rows with alternating colors and places sorting and filtering buttons next to each field name, as shown in Figure 7.1. The table is sorted by the Index field that was its primary key field in the database.

Field names with sorting and filtering options

	A	B	C	D	G	H	I	J
1	Index	Segment	Date	State1	Fujita	Fatalities	Injuries	County1
2	19500103.29.1	Summary	1/3/1950	Missouri	3	0	3	Madison
3	19500103.29.1	State	1/3/1950	Missouri	3	0	3	St. Louis
4	19500103.29.1	State	1/3/1950	Illinois	3	0	0	Madison
5	19500103.17.2	Summary	1/3/1950	Illinois	3	0	3	Montgomery
6	19500103.39.1	Summary	1/3/1950	Ohio	1	0	1	Van Wert
7	19510503.26.1	Summary	5/3/1951	Michigan	2	0	0	Livingston
8	19500125.29.2	Summary	1/25/1950	Missouri	2	0	5	Iron
9	19500125.17.3	Summary	1/25/1950	Illinois	2	0	0	Kankakee
10	19500126.48.1	Summary	1/26/1950	Texas	2	0	2	Brooks
11	19500211.48.2	Summary	2/11/1950	Texas	2	0	0	Brazoria
12	19530511.48.9	Summary	5/11/1953	Texas	5	114	597	McLennan

Symbol on button indicates sorting

Colors indicate an Excel table

Figure 7.1. Data table in Excel.

Filtering

Limiting the records that display is called filtering. This is the same as applying a criterion in a database. Excel provides an interface that creates the program code used to filter the data that is similar to the way Access uses the query by example (QBE) to create SQL code. The Excel method uses menus which are not as flexible but they are easier for most people to use, especially if they already know the underlying principles of using AND and OR logic with criteria. To limit the records displayed to just the tornadoes that occurred in Michigan, the criteria are selected from a menu, as shown in Figure 7.2.

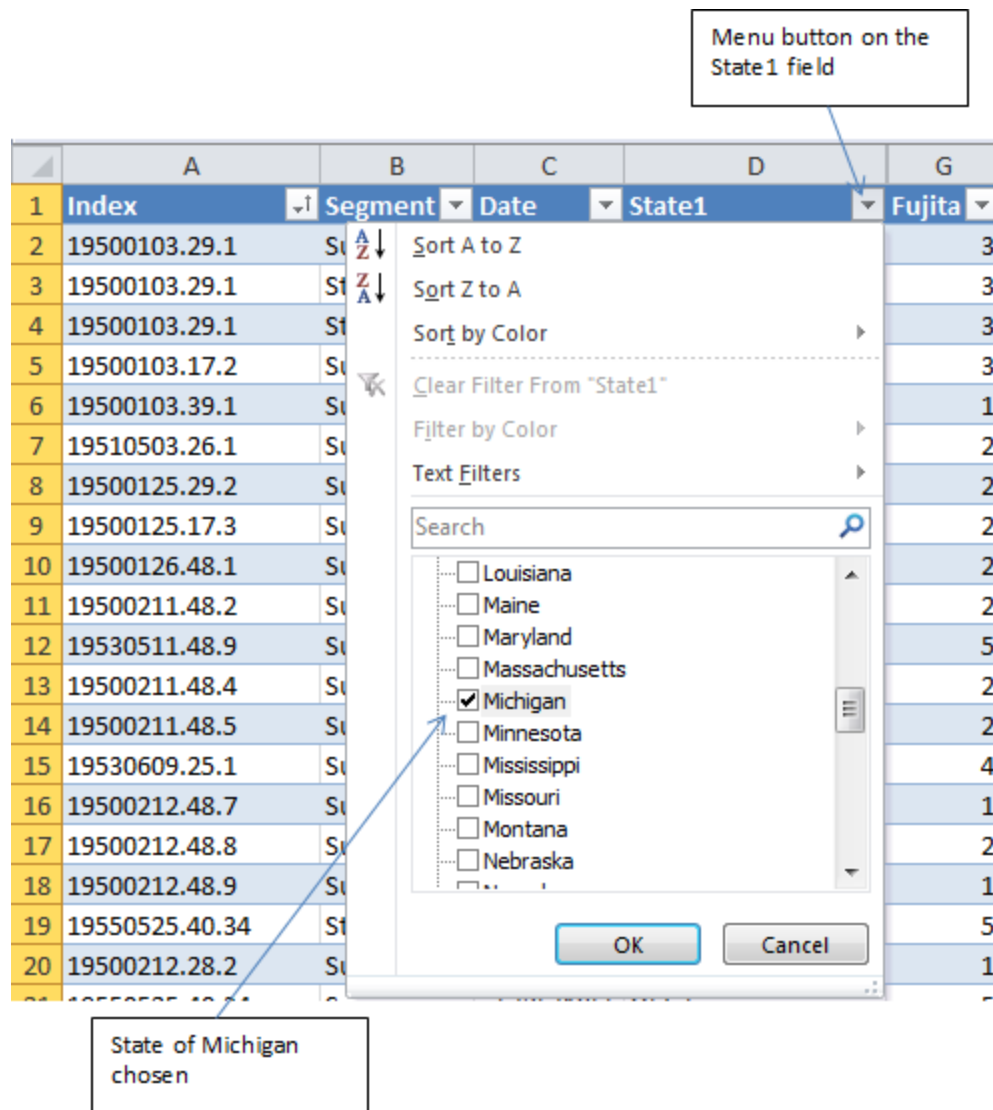


Figure 7.2. Filtering an Excel table.

If the filter for the county field is used, it adds a second criterion that is used with the AND function to restrict the records to the tornadoes that were in Michigan and in Washtenaw county. The original table has more than 54,000 records. By applying these two criteria, the list is reduced to a few records that can be examined visually, as shown in Figure 7.3.

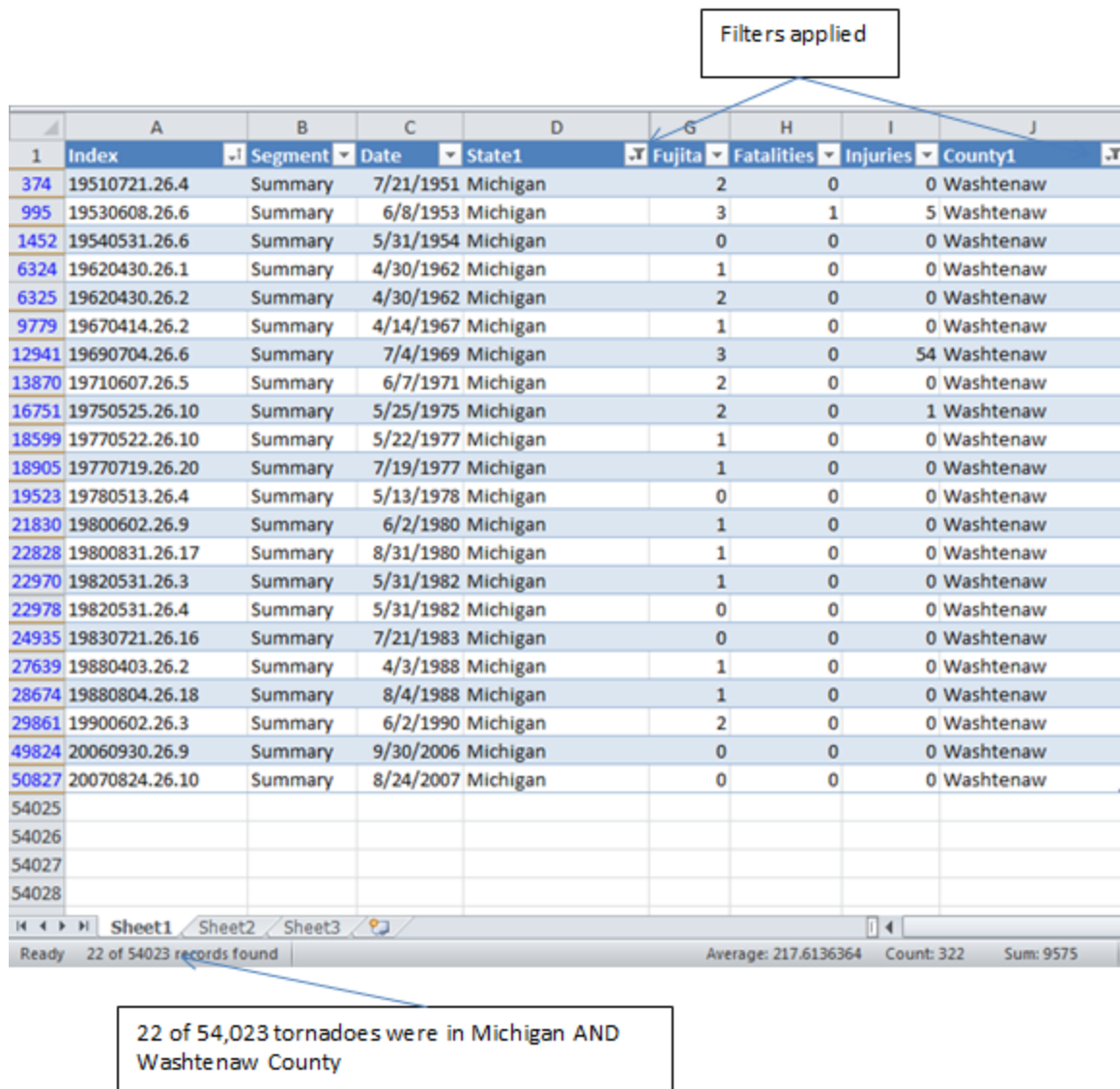


Figure 7.3. Applying two filters is like using two criteria with the AND function.

The records can be filtered using comparison operators that are specified in a dialog box. For example if the user is only interested in tornadoes that have occurred since January 1, 2000, the Date Filters options take the user to a Custom AutoFilter dialog box as shown in Figure 7.4.

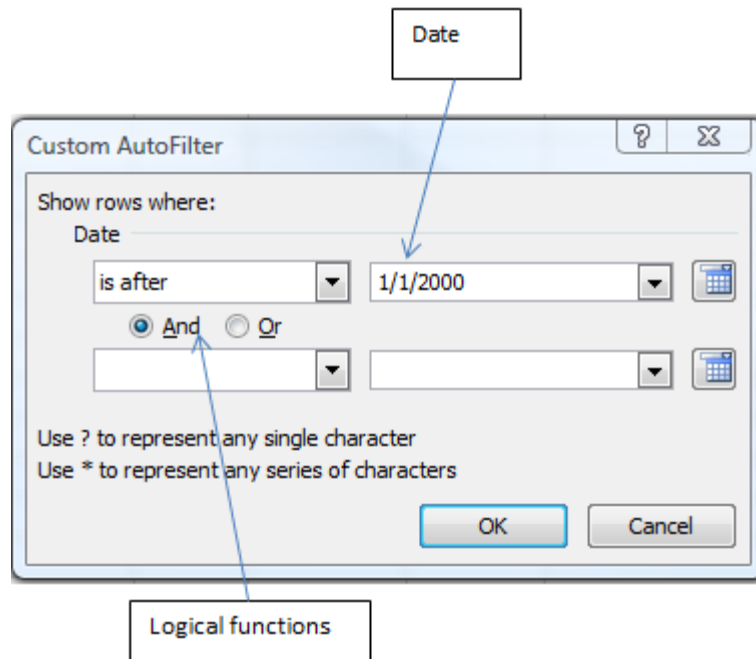


Figure 7.4. Filter by date.

The analysis shows that there have been only two tornadoes in the county since January 1, 2000 up to the latest entry in the database which is in 2009 and the manager concludes that tornadoes do not pose an unacceptable risk to the project.

Sorting

The smallest or largest values in a field can be identified by sorting the list to bring the smallest or largest values to the top. To find which tornado has killed the most people, the table can be sorted in decreasing order on the values in the Fatalities field, as shown in Figure 7.5.

Menu button on the Fatalities field

	A	B	C	D	G	H	I	J
1	Index	Segment	Date	State1	Fujita	Fatalities	Injuries	County1
2	19500103.29.1	Summary	1/3/1950	Sort Smallest to Largest			3	Madison
3	19500103.29.1	State	1/3/1950	Sort Largest to Smallest			3	St. Louis
4	19500103.29.1	State	1/3/1950	Sort by Color			0	Madison
5	19500103.17.2	Summary	1/3/1950	Clear Filter From "Fatalities"			3	Montgomery
6	19500103.39.1	Summary	1/3/1950	Filter by Color			1	Van Wert
7	19510503.26.1	Summary	5/3/1951				0	Livingston

Figure 7.5. Sort by fatalities from largest to smallest.

The sorted records show which tornado killed the most people during the years covered by the data (1950-2009), as shown in Figure 7.6.

	A	B	C	D	G	H	I	J
1	Index	Segment	Date	State1	Fujita	Fatalities	Injuries	County1
2	19530608.26.10	Summary	6/8/1953	Michigan	5	116	844	Genesee
3	19530511.48.9	Summary	5/11/1953	Texas	5	114	597	McLennan
4	19530609.25.1	Summary	6/9/1953	Massachusetts	4	90	1228	Worcester
5	19550525.40.34	State	5/25/1955	Kansas	5	80	273	Sumner
6	19550525.40.34	Summary	5/25/1955	Oklahoma	5	80	273	Sumner

Figure 7.6. Tornado in Michigan has most fatalities.

Key Takeaways

- QBE is a visual aid in describing criteria and their logical functions so they can be translated into the SQL language. Similarly, the filters in Excel are used to describe criteria for use with Excel. [7.1.1]
- Filters in Excel fulfill the same role as criteria in Access. [7.1.2]
- Extreme values like the largest or smallest value in a column can be brought to the top of a column by sorting the column in ascending or descending order. [7.1.3]

2 Analyzing Tables in a Spreadsheet

Learning Objectives

1. Identify the characteristics of a Pivot Table and describe their uses as analysis tools. [7.2.1]

2. Identify the characteristics of a pivot chart and describe their uses as analysis tools. [7.2.2]

Pivot Tables

Advanced spreadsheet programs like Excel have the ability to summarize and report on several fields at once in a device called a **Pivot Table**.

Tip: The word *table* can have different meanings depending on the context in which it is used;

- 1) In a database, table refers to the rectangular array of data organized into records and fields.
- 2) In Excel, a table is a block of cells where each column is a type of data, or
- 3) the term table can refer to a summary of data where one or more of the numeric fields is summed, counted or otherwise calculated.

This type of table has category labels on the rows and columns and the cells of the table are calculated values based on a third field. Other fields can be used to filter the records included in the Pivot Table. Because there are many fields from which to choose in most database tables, it is most effective to start with an objective. For example, a manager at the consulting firm mentioned previously would like to know how many hours of work have been billed by each employee for each project. The manager has the data in an Excel table that was exported from the company's database into Excel, as shown in Figure 7.7.

	A	B	C	D	E	F
1	Last Name	Specialty	Hours Worked	Date	Project Name	Client Name
2	Adams	Mechanical Engineer	1	4/3/2010	Window replacement	Rapid Prototyping Group
3	Adams	Mechanical Engineer	3	4/5/2010	Window replacement	Rapid Prototyping Group
4	Adams	Mechanical Engineer	2	4/6/2010	Ground water heat pump	Multi-Variable Response Systems
5	Baker	Civil Engineer	4	4/6/2010	Ground water heat pump	Multi-Variable Response Systems
6	Adams	Mechanical Engineer	4	4/6/2010	Window replacement	Rapid Prototyping Group
7	Baker	Civil Engineer	3	4/7/2010	Ground water heat pump	Multi-Variable Response Systems
8	Adams	Mechanical Engineer	1	4/10/2010	Window replacement	Rapid Prototyping Group
9	Baker	Civil Engineer	1	4/11/2010	Ground water heat pump	Multi-Variable Response Systems

Figure 7.7. Table of hours worked by person and project.

The next step is to identify the field that contains the data that needs to be summarized and how it will be summarized. In this case, the values in the Hours Worked field will be summed. The two fields that will be used as categories will be Last Name and Project Name. In Excel, the process begins by inserting a Pivot Table. After the three fields are selected, the table and dialog box appear as shown in Figure 7.8.

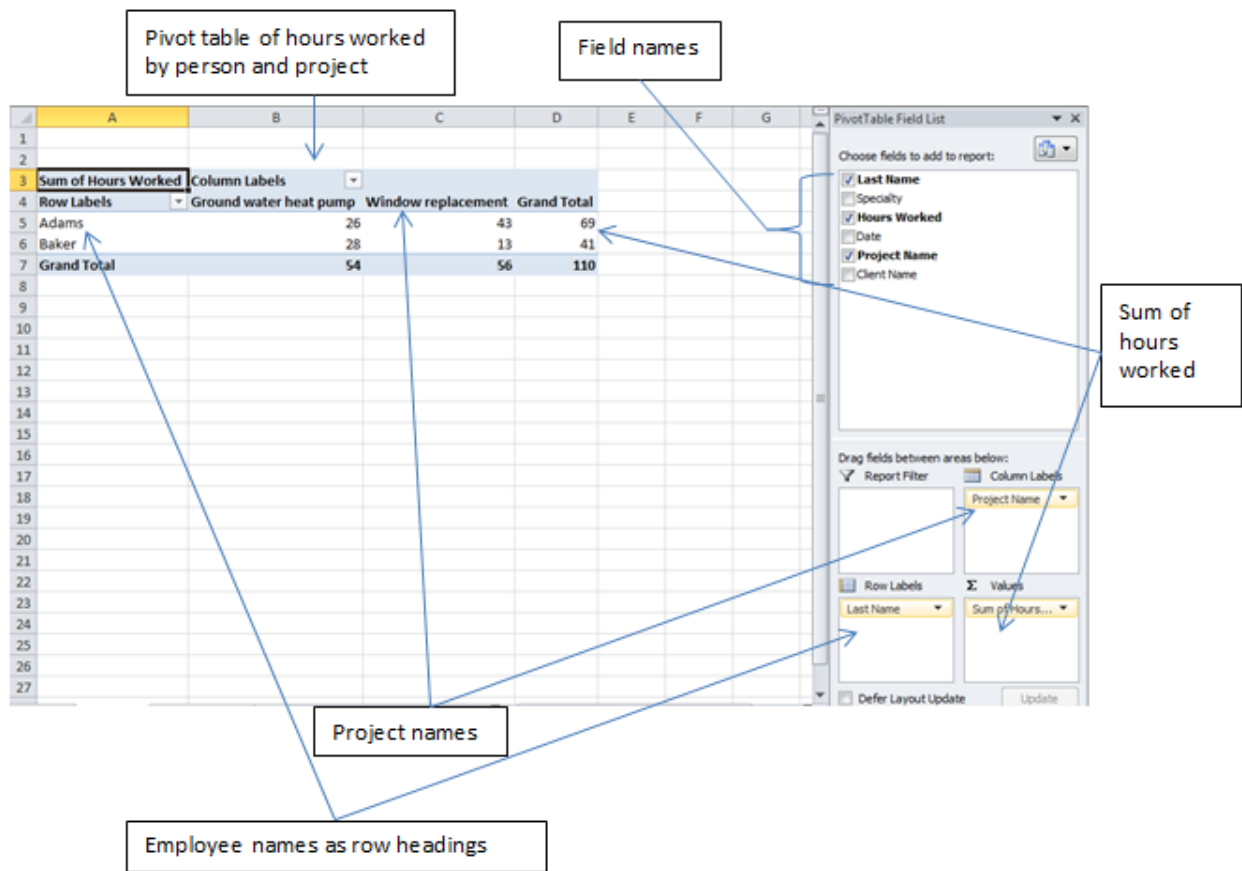


Figure 7.8. Pivot table summarizes hours worked by two different categories.

The fields used for the row or column headings can be switched by changing the field names in the dialog box, as shown in Figure 7.9. This ability to change which fields are used for the vertical or horizontal labels is where this type of table gets its name—Pivot Table.

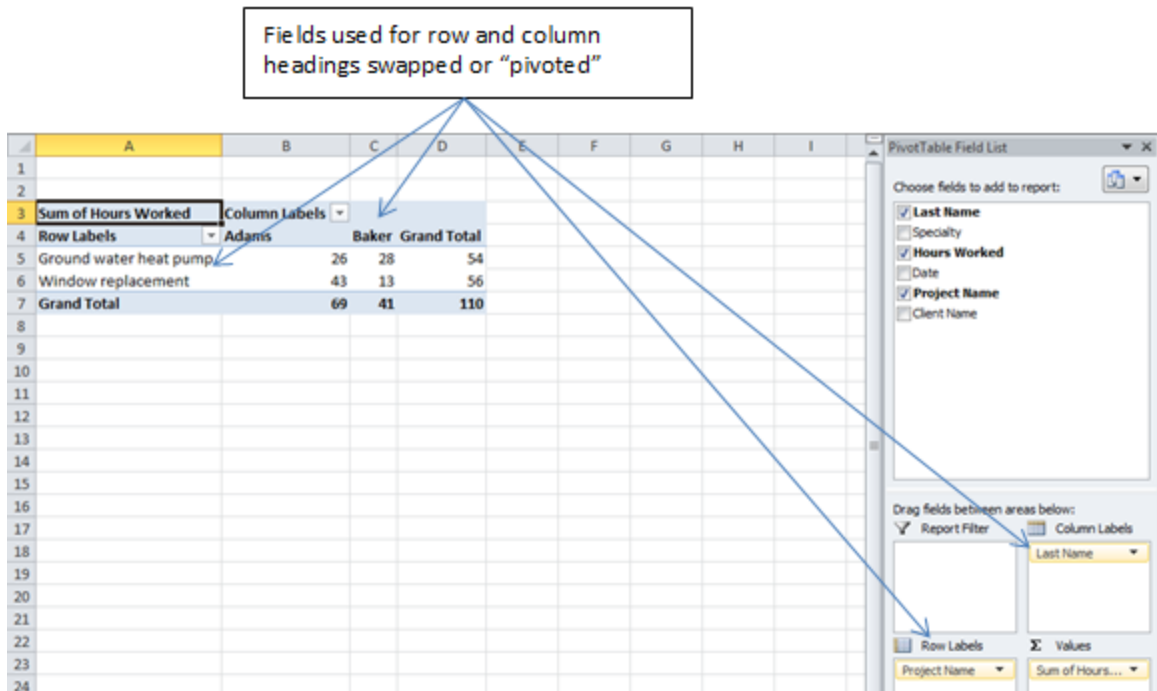


Figure 7.9. Table “pivoted” by switching the use of category fields.

The other fields can be used as subcategories or as filters. For example, the data field could be added and then a criterion applied that limits the work to the first week of the month, as shown in Figure 7.10.

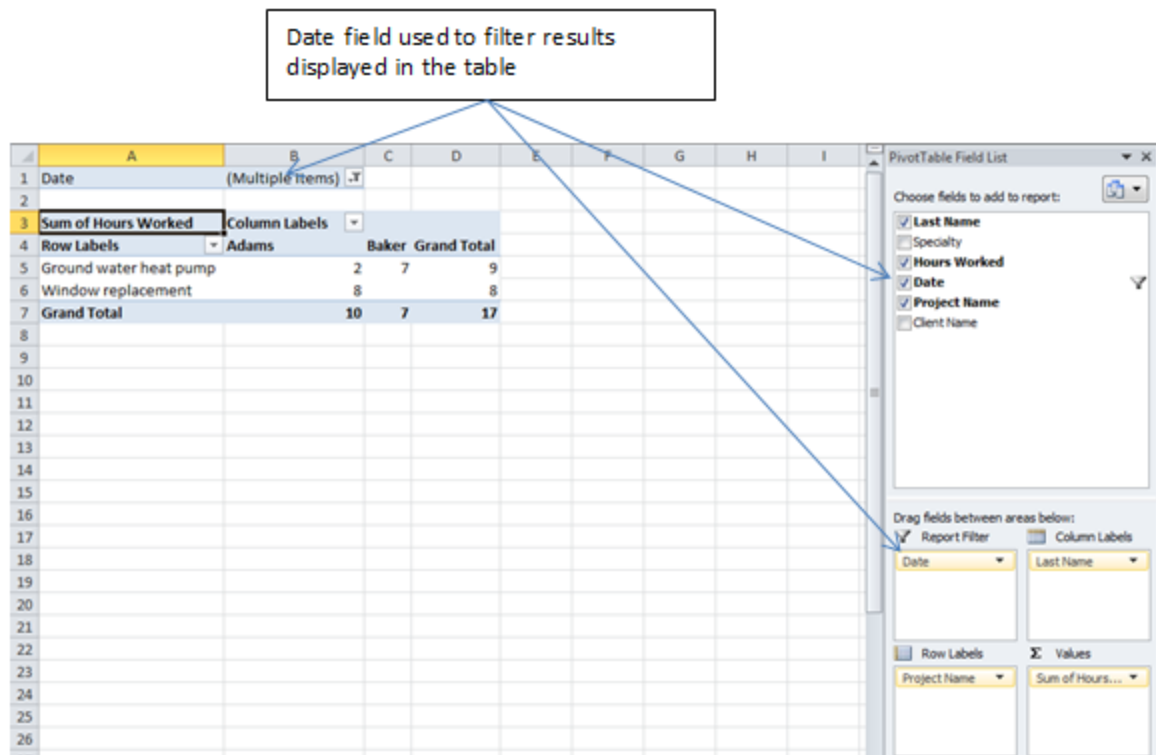


Figure 7.10. Date field used as a filter.

Pivot Table Charts

A similar feature is the **Pivot Chart** which is a chart that is based on the Pivot Table.

Instead of row and column categories, it uses the X-axis labels and the legend labels. The same Pivot Table filtering options are available on other fields such as the date field. For example, a column chart showing the sum of hours worked by two employees could be shown with the project categories as X-axis labels and the employee names in the legend as shown in Figure 7.11.

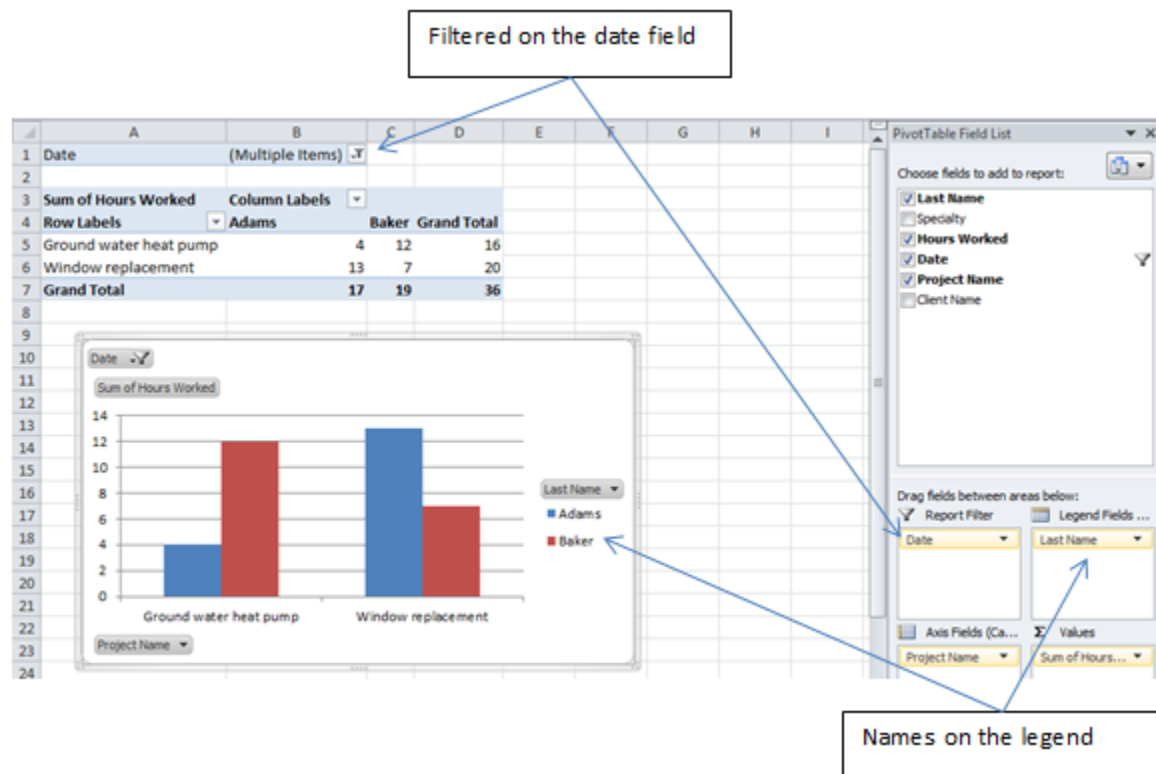


Figure 7.11. Pivot chart filtered by date using the name field for a legend and project name for X-axis labels.

Like a Pivot Table, the legend and X-axis labels can be switched by dragging the field names, as shown in Figure 7.12.

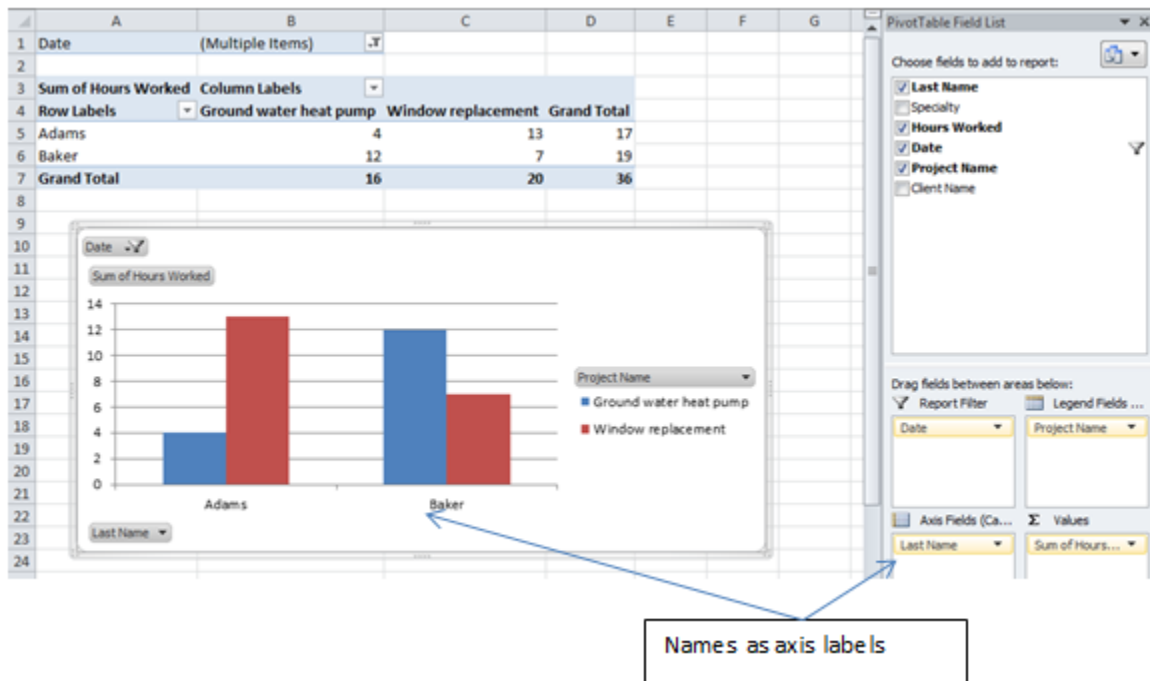


Figure 7.12. Pivot chart with legend and axis labels switched

Pivot Tables and Charts are interactive. Notice that the Pivot Table changed to match the change that was made to the Pivot Chart. Because Pivot Tables and Pivot Charts have so many options, it is necessary to begin with an idea of what kind of table or chart should be produced and then work toward that objective.

Key Takeaways

- In a Pivot Table, two fields can be chosen as the row headings and column headings. A third field that contains values that can be summarized provides the summary numbers in the table's cells. Other fields can be used as filters. [7.2.1]
- In a Pivot Chart two fields can be chosen as the X-axis labels and legend. A third field that contains values that can be summarized provides the summary numbers that are charted on the vertical axis. Other fields can be used as filters. [7.2.2]

3 Analyzing Text in Excel

Learning Objectives

1. Parsing text to extract data using LEFT, RIGHT, or MID functions. [7.3.1]
2. Sorting and filtering a table. [7.3.2]
3. Hiding columns. [7.3.3]
4. Selecting data to chart from non-adjacent areas. [7.3.4]

There is a wealth of data that is available on the web and in other electronic formats. To make sense of the data, you manipulate the data to turn it into useful information that can answer questions. Transferring data into Excel is the first step. The next step is to manipulate the data to extract information from it.

Analyzing wind data embedded in text

Data can be stored within longer phrases of text or in the wrong order. In this section, you extract numbers from text and sort the results so it can be charted.

1. Start Excel. Open Ch07Analyze.xlsx, and then save it on your computer or portable storage device as Ch07AnalyzeStudentName, using your own name for Student Name.
2. If a Security Warning displays on the Warning bar, click the Options button, click enable this content, and then click OK.
3. Click the Wind sheet tab. In cell D2, type **Time of next observation:**
4. Refer to the date and time displayed in cell A2. Mentally, round the time up to the next whole hour when the next observation will be made, and then type that date and time in cell E2 followed by a space and either AM or PM.
5. Apply the date and time format that displays the date and the time in the same cell. For example, if the date and time of the last observation was 10/06//2010 5:50am, you would type 10/06/2010 6:00 AM, as shown in Figure 7.13.

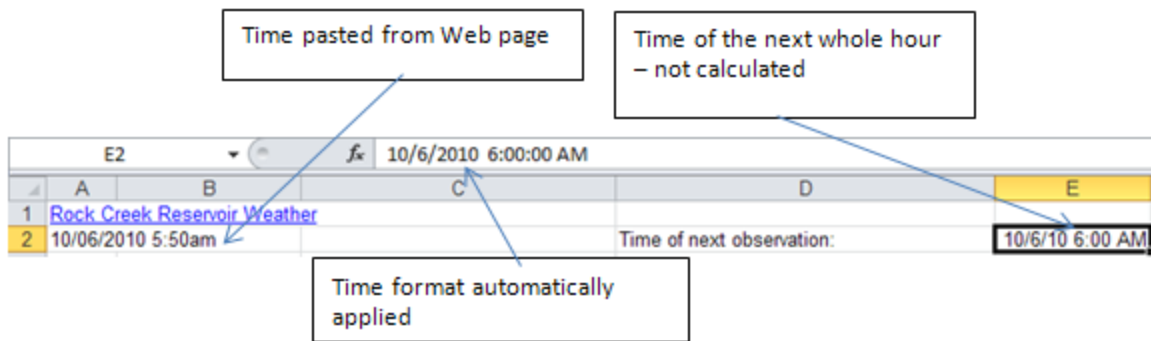


Figure 7.13. Four tables in a database.

5. With cell E2 selected, on the Home tab, in the Number group, at the right end of the box, click the arrow and then click Time. The display is formatted to display the time without the date.

6. In cell E4, type **Time** and then press Enter.

7. In cell E5, type **=RIGHT(B5,2)** and then, on the Formula bar, click the Enter button.

Observe that this function picks two characters from the right side of the text in cell B5. The value, 01, is displayed as text that is aligned at the left of the cell as shown in Figure 7.14.

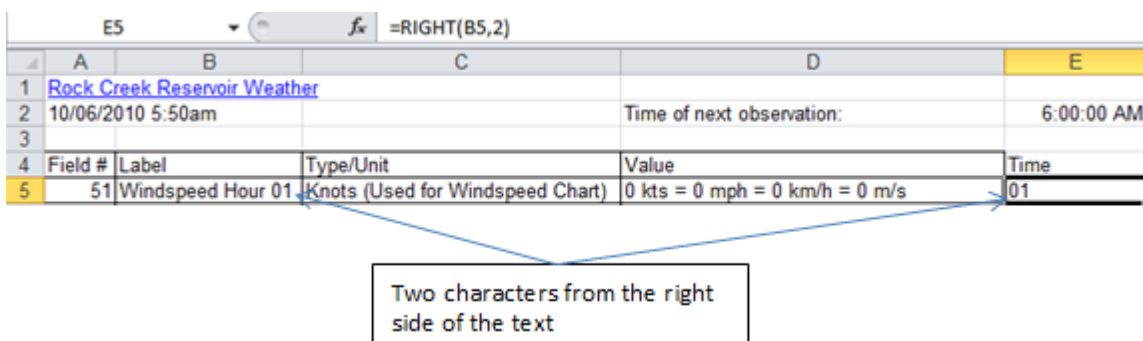


Figure 7.14. RIGHT function.

Tip: Some databases are not well designed for data analysis. In this example, numbers such as the hour and wind speed are part of text. To break something down into component parts for analysis is called parsing. To parse a line of text to extract the useful data, three functions are commonly used. They are RIGHT, LEFT, and MID. The arguments that accompany these functions specify the location of the cell that contains the text and how many characters to extract.

8. On the formula bar, edit the formula to nest the RIGHT function within the VALUE function, and then on the formula bar, click the Enter button. The value is now recognized as the number one, as shown in Figure 7.15.

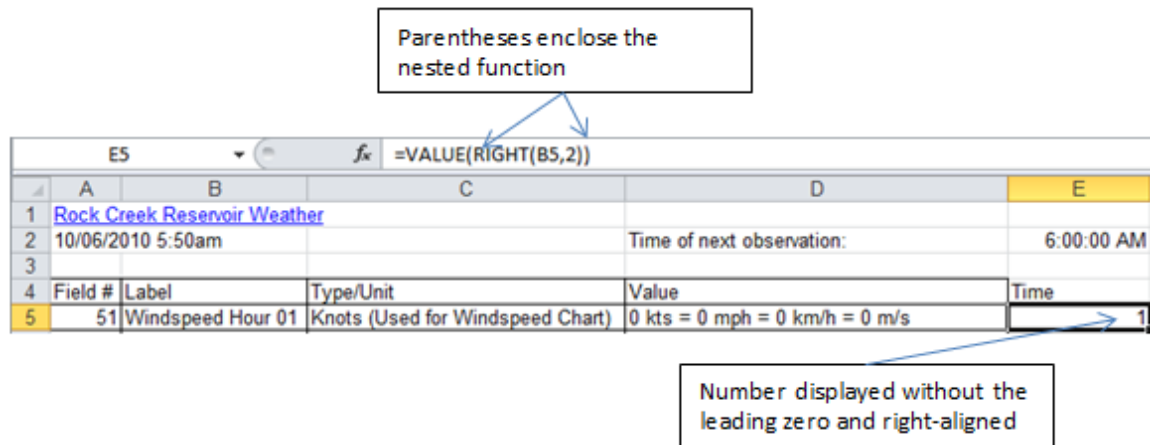


Figure 7.15. RIGHT function nested within the VALUE function.

9. On the formula bar, edit the formula to subtract the hours divided by 24 from the time in E2. Alternatively, type `=E2-VALUE(RIGHT(B5,2))/24` and then click the Enter button. The formula calculates the time that is one hour before the next observation, which will be the time when this wind speed was measured, as shown in Figure 7.16.

Tip: Recall that dates and times are stored as sequential days after a reference date and that time is a fraction of a day. One hour is 1/24 of a day. The absolute cell reference is used so this formula can be filled.

Absolute reference to the time of the next observation in cell E2

E5					fx	= \$E\$2-VALUE(RIGHT(B5,2))/24
	A	B	C	D	E	
1	Rock Creek Reservoir Weather					
2	10/06/2010 5:50am			Time of next observation:	6:00:00 AM	
3						
4	Field #	Label	Type/Unit	Value	Time	
5	51	Windspeed Hour 01	Knots (Used for Windspeed Chart)	0 kts = 0 mph = 0 km/h = 0 m/s	5:00:00 AM	

Dividing by 24 converts hours into fractions of a day

Figure 7.16. Formula calculates time of previous observation.

10. In cell E5, locate the fill handle at the lower right corner, and then double-click it.

The formula is filled into the cells adjacent to the rows of data.

11. In cell F4, type **Speed (mph)** and then press Enter. Observe the text in column D that contains the wind speed; it shows the speed in knots, miles per hour, kilometers per hour, and meters per second.

Tip: Computers are very fast at performing calculations. It is a waste of space to store a number that can be calculated from other numbers in the table. In this example, writing the same speed in four units of measure makes it more difficult to use the data instead of easier. The speed in miles per hour is not in a consistent location within each cell. Instead of extracting the speed in miles per hour, it will be easier to extract the speed in knots from the first two characters and then convert it to mph.

12. In cell F5, type **=VALUE(LEFT(D5,2))** and then click the Enter button on the formula bar. An assumption was made to pick the first two characters because it was unlikely that a wind speed of more than 99 knots would occur. To convert knots to mph we can multiply by 1.15.

13. In cell F5, edit the formula to multiply by 1.15, ***1.15**, or type

=VALUE(LEFT(D5,2))*1.15

14. With cell F5 selected, double-click the fill handle to fill the formula down the

column.

15. Select cells E4:F24. On the Insert tab, in the Charts group, click the Column button and then click the 2-D Clustered Column option.

16. Use skills that you have practiced to delete the legend, add the label **Miles per Hour** to the vertical axis, and the title **Wind Speed at Rock Creek**. Refer to Figure 7.17.

Observe that the time labels on the horizontal axis indicate progressively earlier times. Recall that earlier times and dates are less than later times or dates because of the way they are stored as sequential numbers.

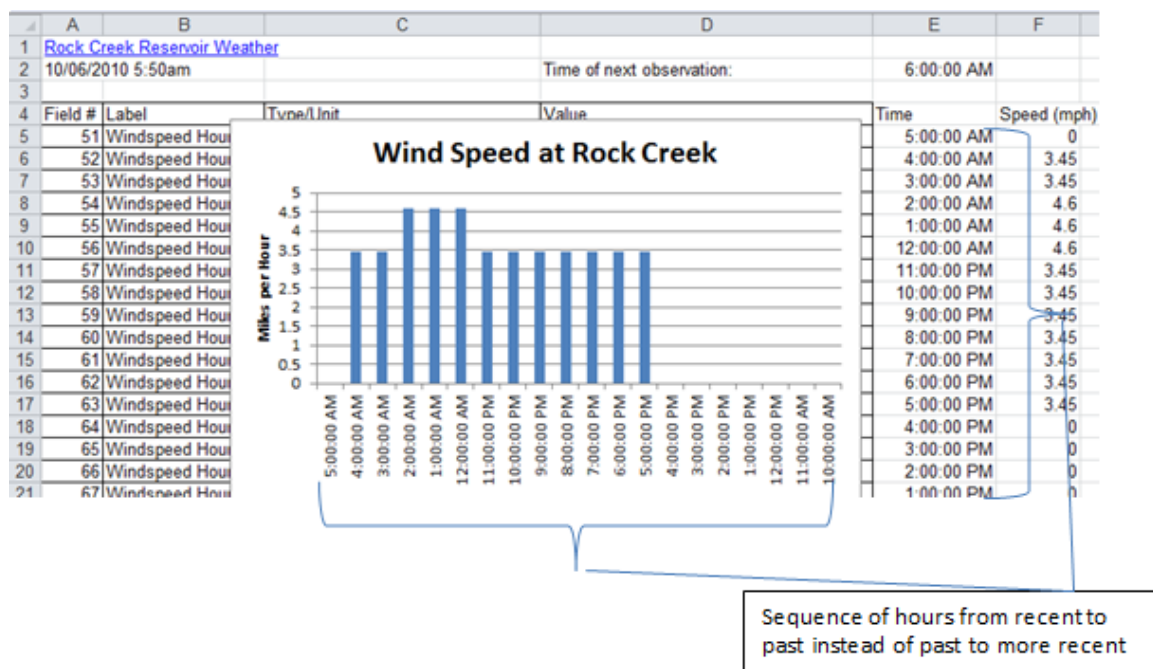


Figure 7.17. Chart of wind speed.

17. Click cell E5. On the Data tab, in the Sort and Filter group, click the Smallest to Largest button. Observe that the table is sorted so that time progresses from top to bottom in the column and from left to right on the chart, as shown in Figure 7.18. Position the chart as shown in the figure.

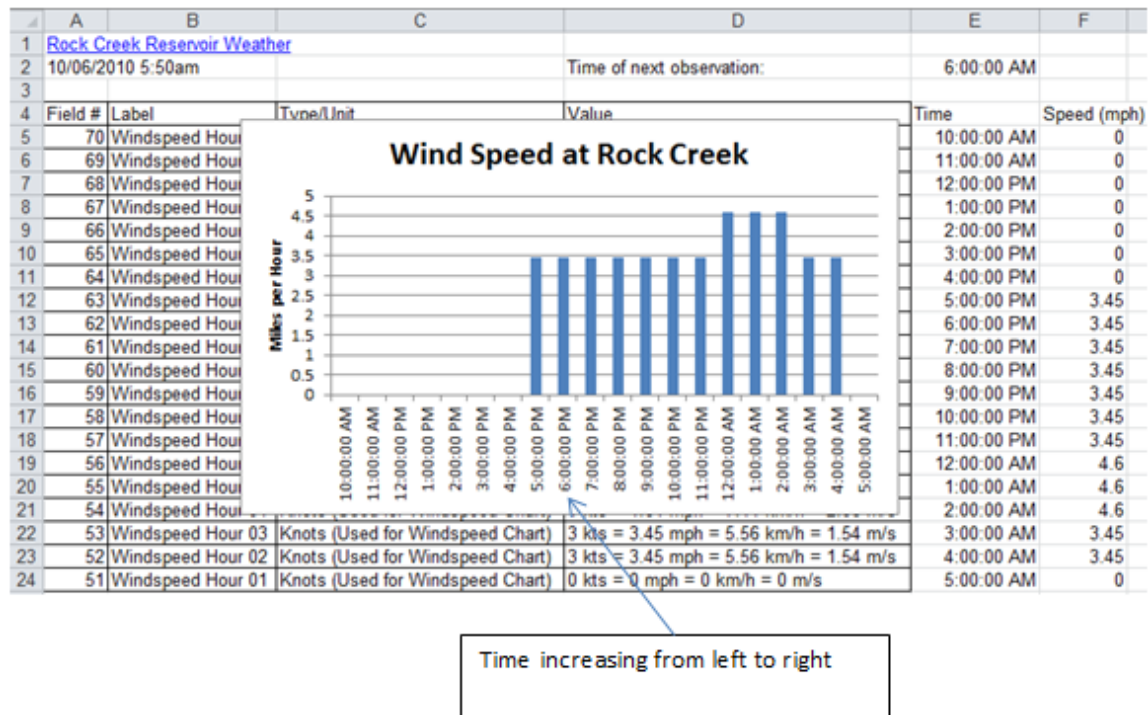


Figure 7.18. Chart of wind speed with time increasing from left to right.

18. Note that it is now relatively easy to look at the chart and observe the speed of the wind and how much it has varied in the past twenty hours. If you also know that a wind turbine usually does not operate below eight miles per hour, you could draw conclusions about when power would have been available from this site. This is an example of how data can be analyzed to create information.

19. Save your work.

Extracting Text and Filtering

Imported data may use codes that have to be deciphered before the data can be filtered to find the information that you want to examine. To do this will parse the code, filter the data, and then hide columns to look at the most recent data. In the next example you will extract information from the Michigan Gasoline data to determine the total BTUs of motor gasoline used in Michigan.

1. Click the MichGasoline sheet tab. This is data that was imported from the U.S. Energy

Information Administration web site.

2. Right-click the row 2 heading to select the row and from the shortcut menu click Insert. This inserts a new blank row 2 between the title in cell A1 and the column headings in row 3. The blank row is inserted to separate the data from the title so that Excel will be recognized the data as a table.
3. Point to the column C heading and when you see the black downward pointing select arrow, click and drag to the right to select columns C, D, and E. With the columns selected, right-click the column heading to display the shortcut menu, as shown in Figure 7.19.

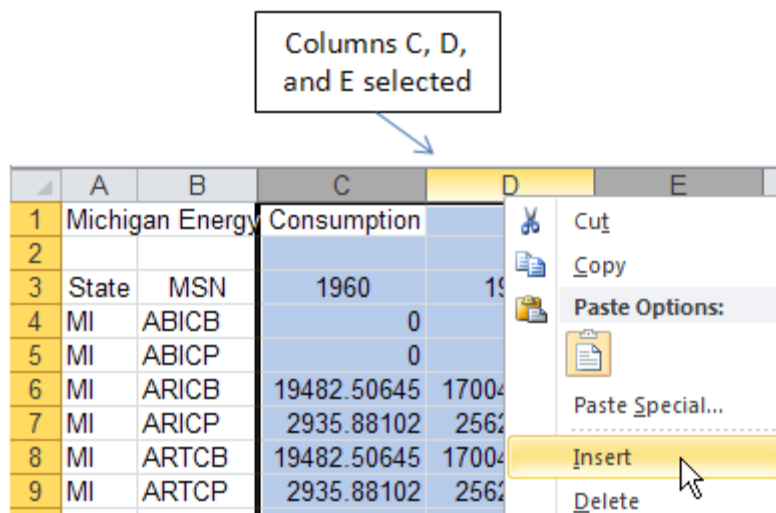


Figure 7.19. Select multiple columns to insert new columns.

4. On the shortcut menu, click Insert. In cell C3 type **Source**, in cell D3 type **Sector** and in cell E3 type **Data Type**.
5. Click the MSN Descriptions sheet tab. Read the information at the top about the five-character codes and how they define the MSN code. Examine the Units used for each code. Notice that different units of measure are used for different codes. This makes the data in the table difficult to compare if the units of measure are different. To make

sense of the data it is necessary to decipher the code and identify the unit that you want to examine.

6. Click the MichGasoline sheet tab. In cell C4 type `=Left (B4, 2)` and then press Enter.

This extracts the left most characters from cell B4—AB. The first two characters describe the energy source

7. In cell D4 type `=MID(B4,3,2)` and then press Enter. This extracts the middle two characters—IC—beginning at the third position in the string. The third and fourth characters describe the energy sector.

8. In cell E4 type `=Right (B4, 1)` and then press Enter. This extracts one character from the right—B—which describes the type of data. You will use these three fields—source, sector and data type—to analyze the data.
9. Select cells C4:E4. Point to the fill handle in the lower right corner of cell E4 and double-click to fill the formulas down the columns.
10. Click in a single cell in the data to deselect the columns. On the Data tab, in the Sort & Filter group, click the Filter button. This adds filter buttons in row 3 next to the column labels.
11. Click the filter arrow in cell C3—Source. Click the Select All check box to remove all of the check marks from the list, and then scroll down the list and click MG—Motor Gasoline. Click OK.
12. Click the filter arrow in cell D4—Sector. Click the Select all arrow and then select TC for Total Consumption. Click OK. This filters the data to two rows. The first is for billions of BTUs, and the second row is for thousands of barrels.
13. Click the filter arrow in cell E4. Use the skills you just practice to limit the display to data type B.
14. Click the column F heading to select column F. Use the horizontal scroll bar to scroll to the right until column AS displays. Hold down the Shift key and click the column AS heading (1999). This selects the columns from F to AS. Right-click on a selected column heading and then click Hide from the shortcut menu. The years from 1960 through 1999 are hidden. The remaining data shows the motor gasoline consumption in Michigan from 2000 to 2008, as shown in Figure 7.20.

Filters applied to source, sector, and data type

Columns F through AS are hidden

	A	B	C	D	E	AT	AU	AV	AW	AX	AY	AZ	BA	BB
1	Michigan	Energy	Consumption											
2														
3	St	MSN	Source	Sect	Data Ty	2000	2001	2002	2003	2004	2005	2006	2007	2008
142	MI	MGTCB	MG	TC	B	615612.661	622447.6801	634049.1786	619729.8222	620413.5338	623987.2424	616275.8617	605713.4164	581336.8972

Figure 7.20. Data is filtered and columns are hidden.

15. Select cells AT142:BB142 and format with the Comma Style and no decimals.
16. Notice that the labels are in row 3, and the filter results are in row 142. Because there is a gap in the rows between the labels and the data, each row has to be selected separate as non-adjacent rows so that you can chart the information. Select cells AT3:BB3, hold down the Ctrl key and select cells AT142:BB142.
17. Insert a 2-D Clustered Column chart. The charting program interprets the years as data to be charted. To fix this, on the Design tab, in the Data group, click Select Data. On the left side of the dialog box select Series 1 and click the Remove button. On the right side of the dialog box, click the Edit button. When the Axis Labels dialog box displays, drag cells AT3:BB3 to select the years as the axis labels. Click OK two times.
18. Notice that the vertical axis begins at 550,000. Right-click the vertical axis and select Format Axis from the shortcut menu. In the Format Axis dialog box, with Axis Options selected in the left panel, on the right side of the dialog box, change the Minimum to Fixed and type 0 in the box as shown in Figure 7.21.

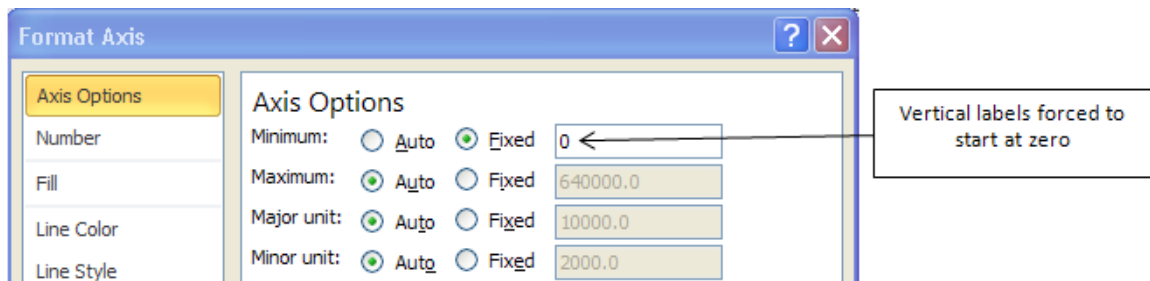


Figure 7.21. Axis minimum is changed to zero.

19. Click Close. Using the techniques you have practiced previously, delete the legend, add a vertical axis title that reads **BTU (Billions)**, and a chart title that reads **Motor Gasoline Energy Use in Michigan**. Drag the chart under the data as shown in Figure 7.22.

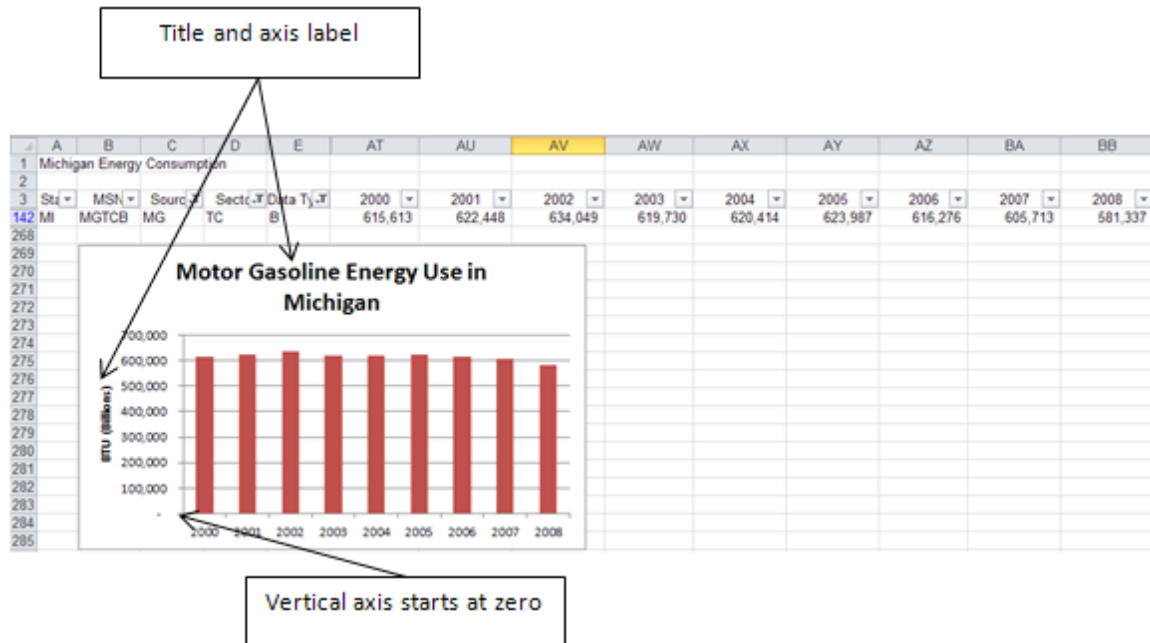


Figure 7.22. Data condensed to an informational chart about gasoline use in Michigan.

20. Save your work.

Key Takeaways

- Data can be parsed by using the LEFT, RIGHT and MID functions to separate data into columns and then the new columns can be sorted, filtered, and charted to find particular information. [7.3.1]
- On the Data tab, use the sort and filtering tools to sort data on one or more fields and limit the records shown to those that meet particular conditions. [7.3.2]
- To hide extraneous columns select the column headings, right-click the selected column

headings and then click Hide. Alternatively, on the Home tab, in the Cells group, click Format and then under Visibility click Hide & Unhide. [7.3.3]

- In data that has been collapsed, if the row numbers are non-continuous, use the .Ctrl key to select the data to chart. [7.3.4]

4 Analyzing Tables of Data in Excel

Learning Objectives

1. Use subtotals to create collapsible outlines. [7.4.1]
2. Analyze three variables with a PivotTable. [7.4.2]
3. Analyze three variables with a PivotChart. [7.4.3]
4. Filtering on additional variables in a PivotTable. [7.4.4]

In this section, you group and subtotal data using the sorting and subtotal features in Excel.

Subtotaling groups of records

Data can be grouped and subtotaled to summarize data by counting the number of records in a group, or summing a numerical field. To do this the data first needs to be sorted by the field that you want to subtotal so that all of the records for that group are together.

1. Click the Projects sheet tab. This data is formatted as a table with filter arrows at the top of each column. To group data so you can subtotal the data by a particular field in the table, it is first necessary to remove the table formatting. To preserve the original table, you will copy it to another worksheet.
2. Right-click the Projects sheet tab, and then from the shortcut menu click Move or Copy.
3. In the Move or Copy dialog box, verify that the *To book* box displays the name of your currently active file. In the *Before sheet* box click (move to end). Click the *Create a*

copy check box and then compare your screen with Figure 7.23.

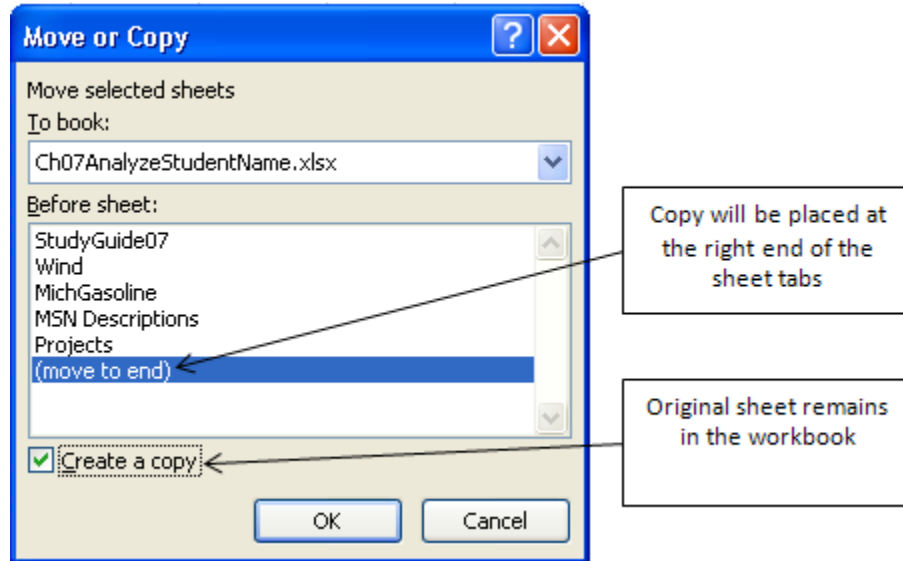


Figure 7.23. Copy of the Projects will be placed at the end of the other worksheets.

4. Click OK. The new worksheet is named Projects(2) and is inserted after all the other worksheets. It becomes the active worksheet.
5. Click a cell in the table of data. Click the Table Tools Design tab. In the Tools group, click the Convert to Range button. In the Warning dialog box that tells you that this action will remove the query definition from the sheet and convert it to a normal range, click OK to continue. The filter arrows are removed but the striped formatting remains on the table.
6. In the upper left corner of the worksheet, above the row headings and to the left of the column headings, click the Select All button. The entire worksheet is selected.
7. On the Home tab, in the Styles group, scroll the list of styles to find the Normal style, as shown in Figure 7.24.

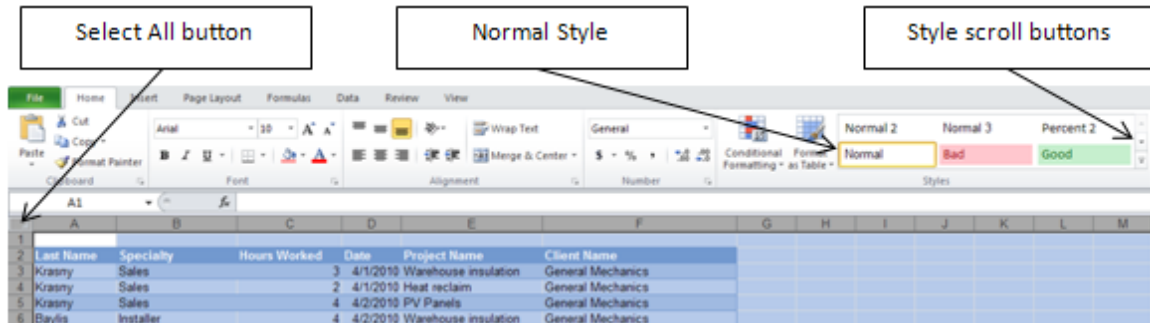


Figure 7.24. To remove table formats change the cell style to Normal.

8. Click the Normal style. The stripped formatting is removed.
9. Click in cell A1 and type **Hours Worked per Client** and press Enter. Select cells A1:F1 and then on the Home tab, in the Alignment group click the Merge and Center button. In the Font group, click the Increase Font Size button three times until the Font Size button displays 14.
10. Click in a cell in column F. On the Home tab, in the Editing group, click the Sort & Filter button and then click Sort A to Z. This sorts the entire group of data by the Client name. Excel recognized the words in row 2 as column labels and kept those at the top of each column and sorted the rest of the data. Arbor Welding appears at the top of the list.
11. On the Data tab, in the Outline group, click the Subtotal button. In the displayed Subtotal dialog box, click the list arrow in the *At each change in* box and click Client Name. In the *Use function* box, click the list arrow and then click SUM. In the *Add subtotal to* box, click Hours Worked, and then remove the check mark next to Client Name. Compare your results with Figure 7.25.

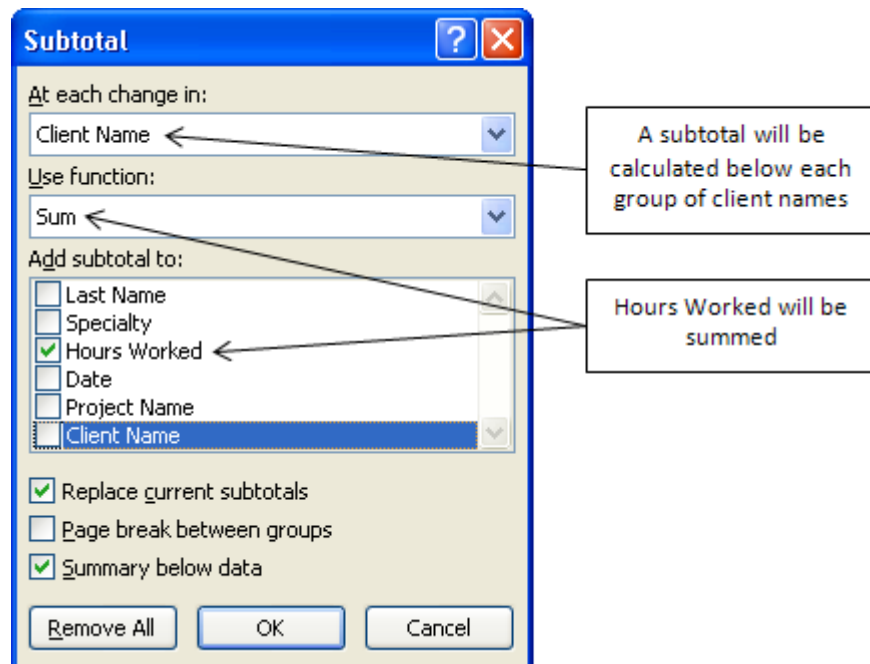


Figure 7.25. Hours Worked will be summed for each change in Client Name.

12. Click OK. A total line is inserted under each group of rows that have the same client name and the number of hours worked is summed in column C. Notice that an outline format has been added to the left side of the data and the data shows all of the records and totals. The 1, 2, and 3 at the top of the outline area can be used to collapse the list to show only the subtotal (2) or only the grand total (1).
13. At the top of the outline click 1 to display only the Grand Total—245 hours worked. Click the 2 to display the subtotals by Client Name and compare your results with Figure 7.26.

Click to expand or collapse display

	A	B	C	D	E	F	G
1							
2	Last Name	Specialty	Hours Worked	Date	Project Name	Client Name	
15			41			Arbor Welding Total	
32			38			First United Church Total	
59			78			General Mechanics Total	
64			14			IBS Medical Group Total	
70			14			Johnstone Building Supply Total	
77			14			Multi-Variable Response Systems Total	
88			31			Peter Bullard Total	
95			15			Rapid Prototyping Group Total	
96			245			Grand Total	

Figure 7.26. Data collapsed to subtotals of hours worked for each client.

14. Double-click the Projects (2) sheet tab and type **HoursWorked** and then press Enter to rename this sheet. Save your work.

In the next section you will create a PivotTable and related PivotChart that will tell you how many hours selected staff positions have worked on various types of projects.

Creating a PivotTable

PivotTables and PivotCharts enable you to consolidate data and look at the same information in different ways. You can manipulate the data to answer questions such as how much was sold in each product category by the each sales region in the company. It is a powerful tool when you have large quantities of data to transform into information. To begin, you need to have an idea of the question that you want to answer. You use one or two fields as row labels, one as a column label, and then typically a numeric field as the value field that will be summarized at the intersection between the row and column labels. You can add a fourth field to further filter the results. In our project database, you will create a PivotTable to determine how many hours are worked for different types of projects by the professional staff positions.

1. Click the Projects sheet tab and click in a cell in the table of data. On the Insert tab, in the Tables group, click the PivotTable button.
2. In the Create PivotTable dialog box, the table range box displays the name of this table

of data that was imported from Access. You can place the new PivotTable on its own worksheet, or place it in the Existing Worksheet. Under *Choose where you want the PivotTable report to be placed*, click the *Existing Worksheet* option button. In the Location box type H6 as the cell location for the new table, and then click OK.

3. Use the horizontal scroll bar to scroll to the right until column H is the first column displayed in the window. A PivotTable placeholder displays beginning in cell H6 and the PivotTable Field List pane displays on the right side of the screen as shown in Figure 7.27

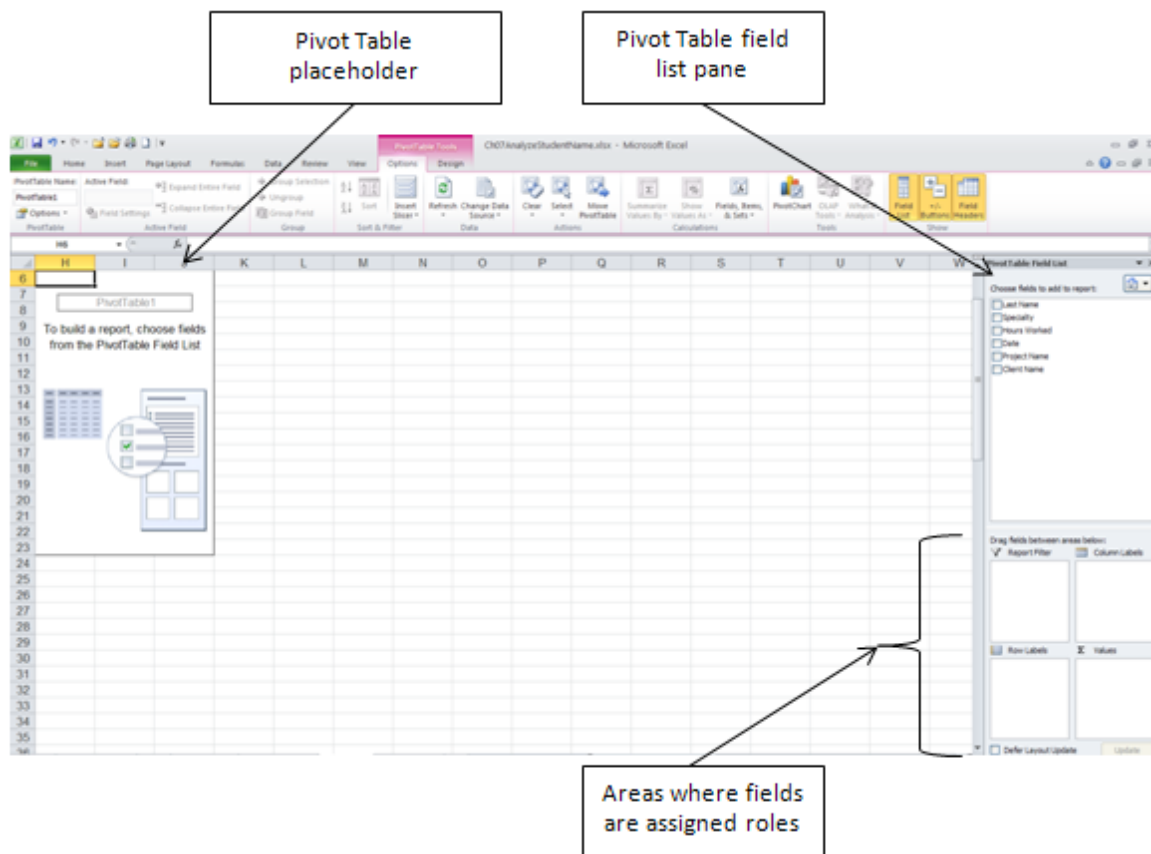


Figure 7.27. PivotTable Field List is used to position fields on the PivotTable.

4. In the PivotTable Field List, click the check box next to the Last Name field. This places the Last Name field in the Row Labels area at the bottom of the Pivot Table

pane displays the last names as row labels in the PivotTable area and.

Tip: You can add fields to a PivotTable by clicking the check box next to the field name, or by dragging a field name to the in the lower part of the PivotTable pane to the box where you want it to display.

5. In the PivotTable Field List, click Project Name and drag it to the Column Labels box in the lower part of the pane. This adds the Project Names as column Labels in the worksheet.
6. Drag the field name Hours Worked to the Values box and compare your screen to Figure 7.28.

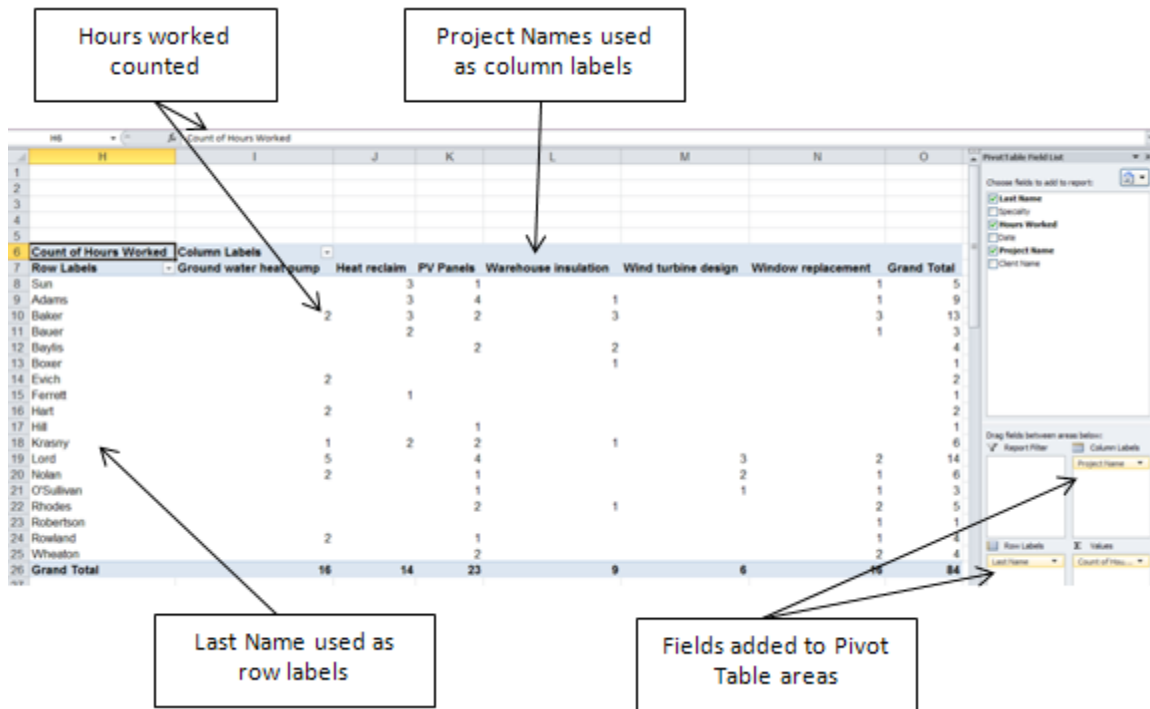


Figure 7.28. Fields are added to the PivotTable.

7. In the PivotTable pane, in the Values box, click the Count of Hours field and then from the shortcut menu click Value Field settings. By default, the Hours Worked field was counted rather than summed. In the Value Fields Settings dialog box you can change how the data is summarized. You can also use this dialog box to format the numbers if necessary.
8. In the Value Field Settings dialog box, under Summarize value field by, click Sum and then click OK. The hours worked display totals by Last Name in column O and totals by Project Name in row 26.
9. In the PivotTable Field List, click the check box next to the Specialty field. This adds this field as a row label under the Last Name field. In the Row Labels area of the PivotTable pane, drag the Specialty field up so that it displays in the Row Labels box above the Last Name field and then compare your screen to Figure 7.29.

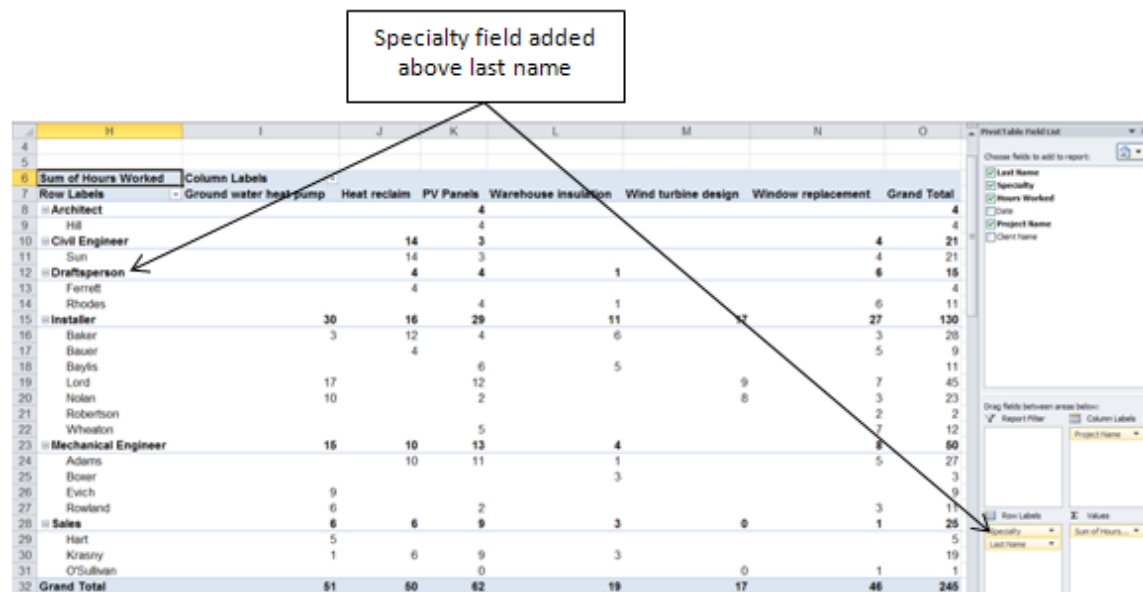


Figure 7.29. Two fields can be used as row labels which creates subtotal.

10. In the PivotTable pane, in the Row Labels box, drag the Specialty field to the Report Filter box. Fields can be added to the Report Filter area which can then be used to filter or limit the overall report.
11. In the Columns Labels box, drag Project Name to the Row Labels box, and then drag Last Name to the Column Labels box. This is how you pivot a report by switching the placement of the column and row labels.
12. Click the PivotTables Tools/Options tab and examine the tools that are available to manipulate the PivotTable. In the Show group, click the Field Headers button to make it inactive. This removes the Row Labels and Column Labels list arrows from the PivotTable.

Next you create a PivotChart that will chart the hours worked by employee for various projects.

Creating a PivotChart

A PivotChart is based on the information in a PivotTable and is interactive with the table. If

you change the table, the chart will be update. You can also make changes to the chart which will update the information displayed in the table.

1. On the PivotTables Tools/Options tab, in the Tools group, click the PivotChart button. In the Insert Chart dialog box, under column, click the first option—Clustered Column—and then click OK. A column chart is added to the worksheet, as shown in Figure 7.30.

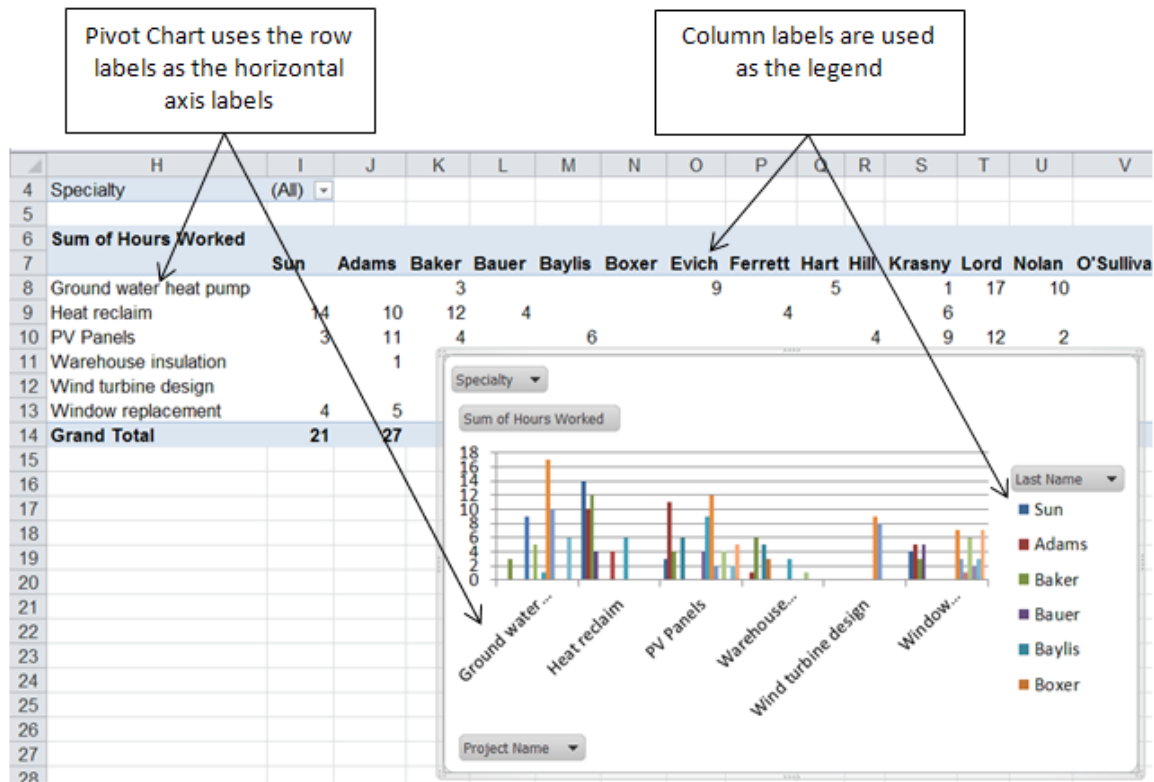


Figure 7.30. A PivotChart is interactive with the PivotTable it is based upon.

2. On the PivotTable, click the Specialty list arrow in cell I4. On the list of specialties, click *Select Multiple Items*, and then click to remove the check marks next to Installer and Sales. Click OK. The chart is changed to reflect the change in the PivotTable.
3. If necessary, click the chart to make it active, and then on the PivotChart Tools / Design tab, in the Type group, click the Change Chart Type button. Under Column, click the second button—Stacked Column—and then click OK.
4. In the Chart Layout group, move the pointer onto the first layout option. A screen tip

displays indicating that this is Layout 1. Click the Layout 1 button. On the chart, click Chart Title and then type **Hours Worked by Project by Selected Staff** and then press Enter.

5. On the PivotChart Tools / Layout tab, in the Labels group, click the Axis Titles and then click Primary Vertical Axis Title. Click Rotated Title and then type **Hours** for the vertical axis.
6. Drag the chart to the left and position it under the PivotTable.
7. Click the PivotTable to make it active. Click the PivotTable Tools / Options tab. In the Show group, click the Field Headers to turn them back on. Click the Row Labels list arrow and then click Warehouse Insulation to remove this from the list. Click OK. Both the table and the chart are updated.
8. On the PivotTable Tools / Design tab, in the PivotTable Styles group point to several different designs to see how they display on the PivotTable. As you point to a design a screen tip displays the name of the design. Scroll down the examples and locate Pivot Style Medium 23, and then click it. Compare your results with Figure 7.31.

Key Takeaways

- Sort data by the field that you want to subtotal and then on the Data tab, in the Outline group click the Subtotal button and select the field you want to subtotal by, the field you want to subtotal and the placement of the subtotal. Expand the outlined data to display all the details or collapse it to show only the subtotal by group, or the grand total of all records. [7.4.1]
- Create a PivotTable by identifying the range of data to use and then drag fields to the Row Label, Column Label and Value boxes on the PivotTable Field List pane. Drag fields between these boxes to pivot the table and change the view of the data. [7.4.2]
- On the PivotTable Options tab, in the Tools group, click the PivotChart button and select the type of chart to create. Changes made to the PivotTable affect the PivotChart and vice versa. Format the PivotChart as you would any other chart. [7.4.3]
- Add fields to the PivotTable Report Filter box and use this field to limit the entire report and chart to particular data. Filters can be applied to the Report field, Row Labels or Column Labels. [7.4.4]

Key Terms

Database management system (DBMS)

Software and hardware designed to facilitate storage and retrieval of data..

Filtering

Restricting the display of records based on criteria.

Parse

To extract useful data from a line of text by using functions that locate the text within a string of characters.

Pivot Chart

Interactive chart of summary data where the user can switch the labels of horizontal chart axis and legend easily.

Pivot Table

Interactive table of summary data where the user can switch the labels of the rows and columns easily.

Last edited on January 11, 2011