Prepared By

Handaru Jati, Ph.D

Universitas Negeri Yogyakarta

handaru@uny.ac.id

Chapter 10

Working with Large Data

Chapter Overview

10.1 Introduction

This chapter explains how to work with external data or large data in Excel. We illustrate how to import data from text files, webpages, and databases. For the database discussion, we assume the reader has some previous knowledge about database software, such as Microsoft Access. Knowing how to import data to the spreadsheet environment is an important attribute for developing a DSS. In many real-life applications, large amounts of data will be stored in a separate database; this data has to be imported as input for the model in the DSS. We revisit this topic in Chapter 21 to illustrate how this can be done using VBA. We have developed several DSS applications which use these features, such as the University Information System application.

10.2 Importing Data

Excel offers many tools for working with and analyzing large amounts of data. In many cases, an Excel user may not be given data in a worksheet to manipulate. If this is the case, you can use the Import options in Excel to transfer this data to a worksheet where Excel analysis tools can be applied. There are three sources from which you can import data into Excel: a text file, a webpage, and a database.

10.2.1 Text Files

To import data from a text file, use the *Text Import Wizard*. If you try to open a text file in Excel, this Wizard will appear. There are three simple steps to follow to use the Wizard. The first step is to specify how you want to organize your data into columns for Excel. There are two main options: *Fixed Width* and *Delimited*. If you choose *Fixed Width*, Excel will guess how to separate your data into columns. The second step of the Wizard will then give you a chance to modify this guess. If you choose the *Delimited* option, you specify what character will be used as a separator, for example: commas, spaces, tabs, etc. This character choice would be done in the second step of the Wizard with the *Delimited* option. The third step of the Wizard in both cases is to specify any particular numerical formatting to be applied to the data as it is imported into Excel. Let's take a closer look at the *Text Import Wizard* through an example:

Figure 10.1 presents a text file that records the number of hours ten employees worked for three consecutive weeks. The file records each employee's start date and hours worked for three weeks. This file has been saved as a ".txt" file.

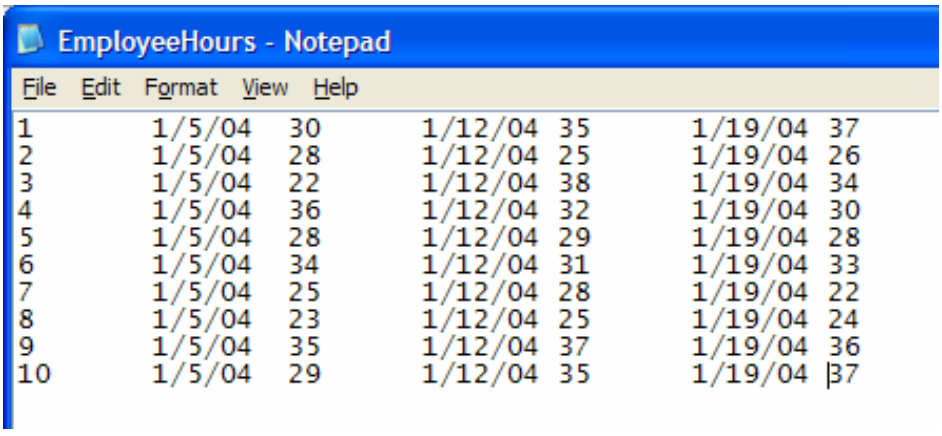


Figure 10.1 Text file of employee hours.

The supervisor of these employees may be interested in calculating the average hours worked per employee and the average hours worked every Monday. To take advantage of Excel tools and functions that will help her accomplish this, she must import this text file into Excel. After ensuring the above data is saved as a text file, we open Excel and choose *File* > *Open* from the menu; then select the above text file. The *Text Import Wizard* will appear (see Figure 10.2).

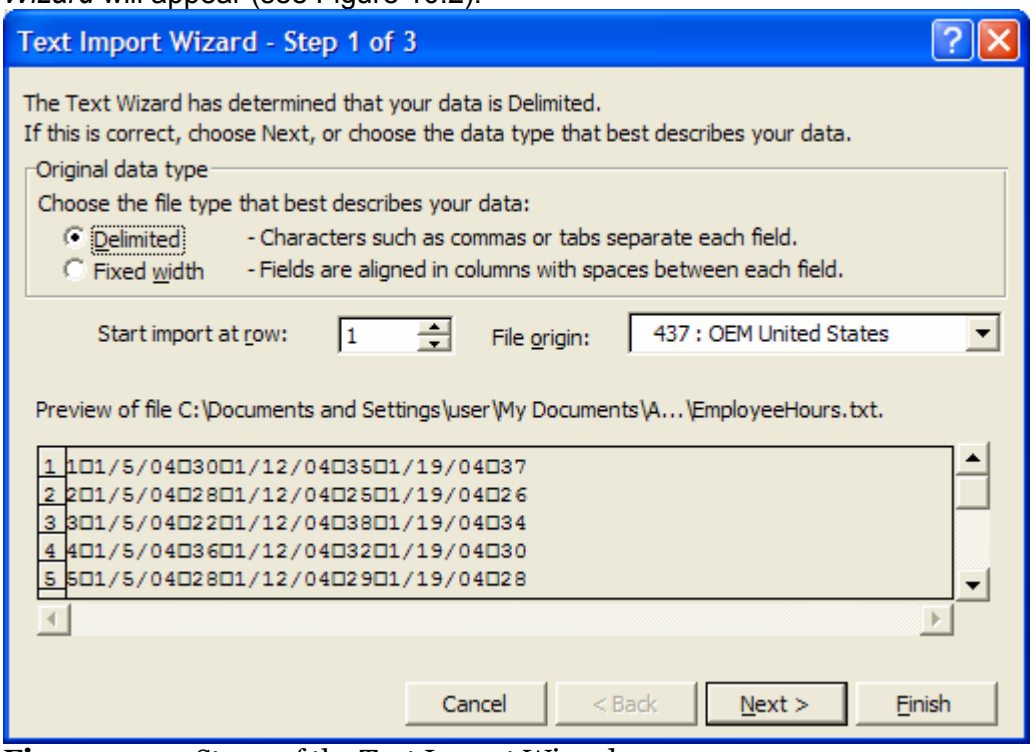


Figure 10.2 Step 1 of the Text Import Wizard.

For the first step of the Wizard, we will choose the *Delimited* option. We notice that tabs separate the text file data. Excel may have guessed this if we had selected the Fixed Width option, but it is still important to ensure that tabs are used (notice that "/" marks are also used in the dates which might be considered as a separator in Excel's guess). In the second step of the Wizard, we select "Tab" as the delimiter (see Figure 10.3). We could try other options to preview the columns that would appear. For example, we could check "Other" and enter the "/" mark in the adjacent space to preview an incorrect separator for the data.

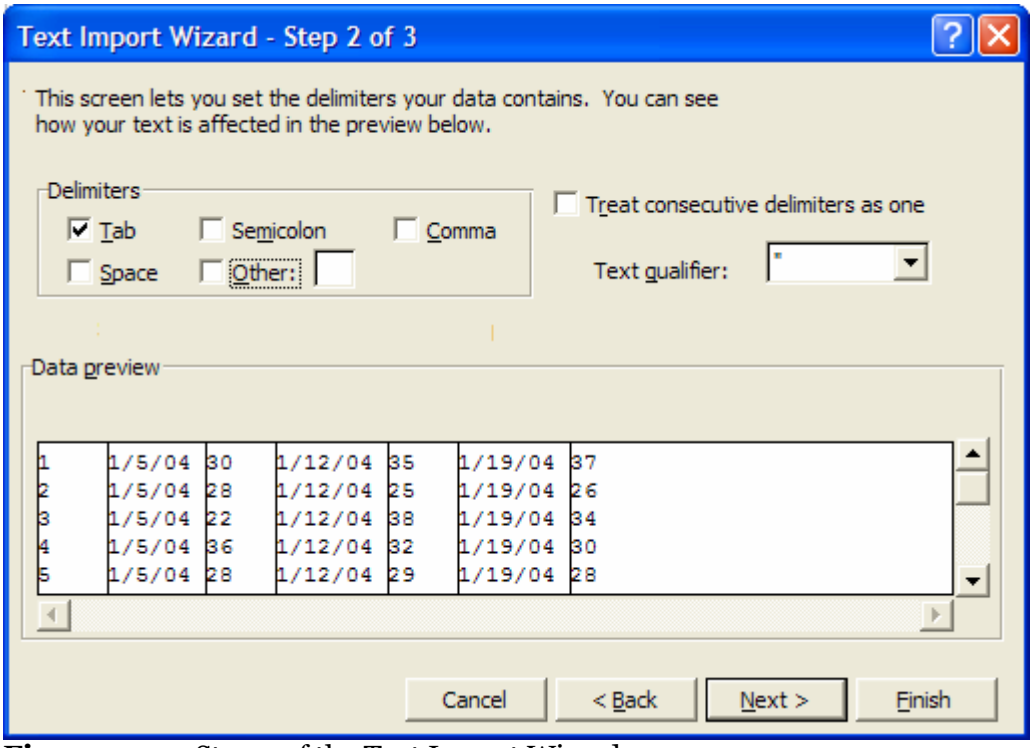


Figure 10.3 Step 2 of the Text Import Wizard.

Next, we go to the third step of the Wizard and specify any numerical formatting for the data (see Figure 10.4). For this text, the second, forth, and sixth columns should be entered as dates. So, we select each of these columns in the preview area and choose "Date: MDY" as the format option. The other columns can be kept in "General" formatting.

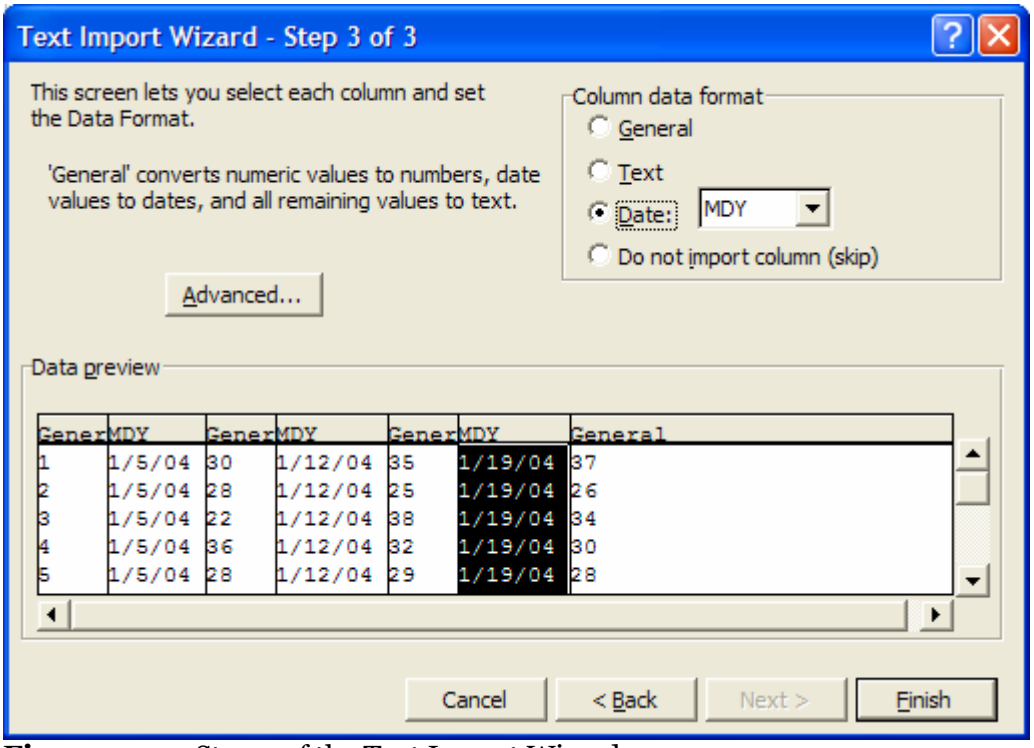


Figure 10.4 Step 3 of the Text Import Wizard.

We are now done with the *Text Import Wizard*; the text data should now be imported into Excel columns, as shown in Figure 10.5.

	A	B	C	D	E	F	G
1	1	1/5/2004	30	1/12/2004	35	1/19/2004	37
2	2	1/5/2004	28	1/12/2004	25	1/19/2004	26
3	3	1/5/2004	22	1/12/2004	38	1/19/2004	34
4	4	1/5/2004	36	1/12/2004	32	1/19/2004	30
5	5	1/5/2004	28	1/12/2004	29	1/19/2004	28
6	6	1/5/2004	34	1/12/2004	31	1/19/2004	33
7	7	1/5/2004	25	1/12/2004	28	1/19/2004	22
8	8	1/5/2004	23	1/12/2004	25	1/19/2004	24
9	9	1/5/2004	35	1/12/2004	37	1/19/2004	36
10	10	1/5/2004	29	1/12/2004	35	1/19/2004	37
11							

Figure 10.5 The text file data imported into Excel columns.

The employer can now use the Excel AVERAGE function to compute the averages that she was looking for (per employee and per date). She can also organize her data by adding column headings and other formatting (see Figure 10.6).

	A	B	C	D	E	F	G	H	I
1	EmployeeID	Date 1	Hours 1	Date 2	Hours 2	Date 3	Hours 3		Average Hours
2	1	1/5/2004	30	1/12/2004	35	1/19/2004	37		34.00
3	2	1/5/2004	28	1/12/2004	25	1/19/2004	26		26.33
4	3	1/5/2004	22	1/12/2004	38	1/19/2004	34		31.33
5	4	1/5/2004	36	1/12/2004	32	1/19/2004	30		32.67
6	5	1/5/2004	28	1/12/2004	29	1/19/2004	28		28.33
7	6	1/5/2004	34	1/12/2004	31	1/19/2004	33		32.67
8	7	1/5/2004	25	1/12/2004	28	1/19/2004	22		25.00
9	8	1/5/2004	23	1/12/2004	25	1/19/2004	24		24.00
10	9	1/5/2004	35	1/12/2004	37	1/19/2004	36		36.00
11	10	1/5/2004	29	1/12/2004	35	1/19/2004	37		33.67
12									
13	Average Hours		29		31.5		30.7		

Figure 10.6 The final organized data with average calculations.

10.2.2 Web Addresses

To import data from a webpage, use the *Data* > *Import External Data* > *New Web Query* option from the menu. In the dialog box that appears, you can enter the web address that contains the data you want to import. There will then be a preview of this website in the window. You can then select what sections of data you want to import; that is, you may not want to import everything on the webpage. The possible data sections will be marked by small arrows. Select the arrows for all the data you are interested in and then click *Import*. Excel will prompt you to select a cell in the worksheet for where you want to place this imported data. Excel will import the data and separate it into columns, as done on the webpage. There is no prompt for how to separate this data, but Excel usually makes a good guess. You can always reorganize your data in the spreadsheet, if desired.

Let's consider an example. The following web address contains current stock quotes and information for MSFT stock (see Figure 10.7), reported by Quicken:

http://www.quicken.com/investments/quotes/?p=MSFT

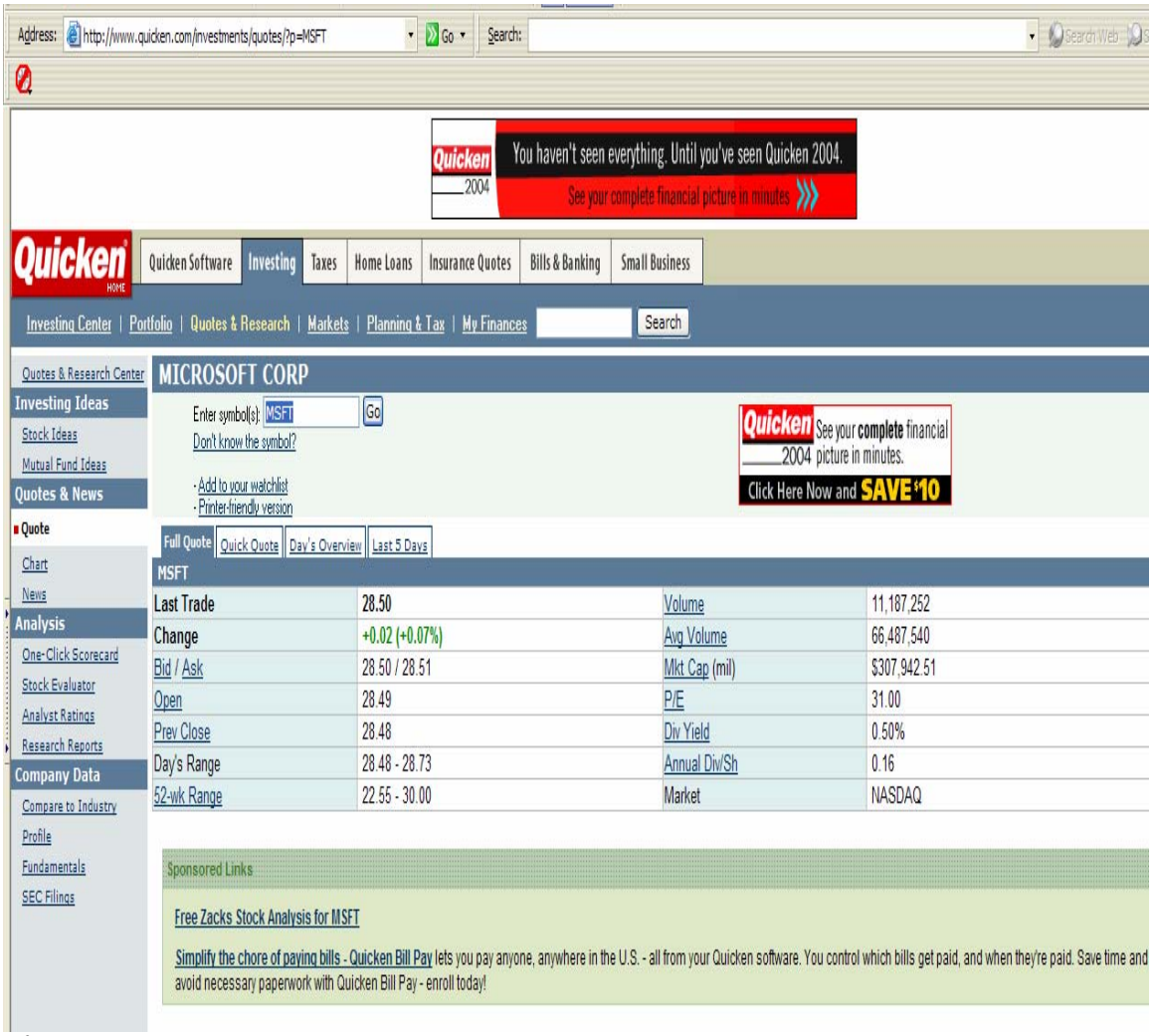


Figure 10.7 Quicken quote on MSFT stock.

Suppose we are interested in maintaining a worksheet with this data in order to compare the company values or to analyze how a possible portfolio of this stock is doing. To import this data, we first copy the web address of this page. Then, we select *Data* > *Import External Data* > *New Web Query* from the menu and simply paste the web address into the "Address" window of the dialog box (see Figure 10.8).

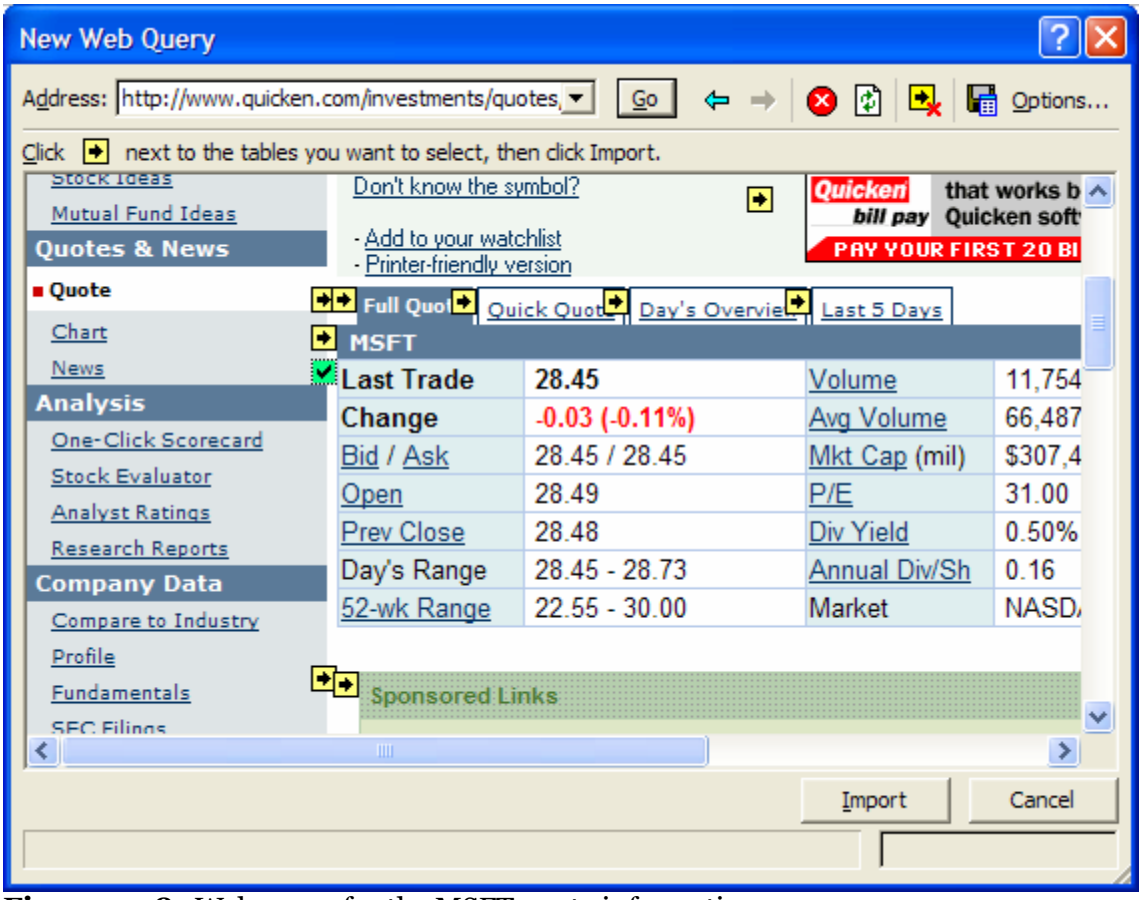


Figure 10.8 Web query for the MSFT quote information.

In the above webpage preview, we can see the small arrows in yellow boxes that Excel has placed on different sections of data. Since we are only interested in the actual stock data, we select the arrow next to "Last Trade." Notice as you place your cursor on this arrow that a square appears around the entire section of data associated with that arrow. After we click the arrow, it will become a check mark, indicating that we wish to import this data. Now that we are done selecting data, we click *Import*. Finally, we must specify where in the worksheet we would like to paste the data (see Figure 10.9).

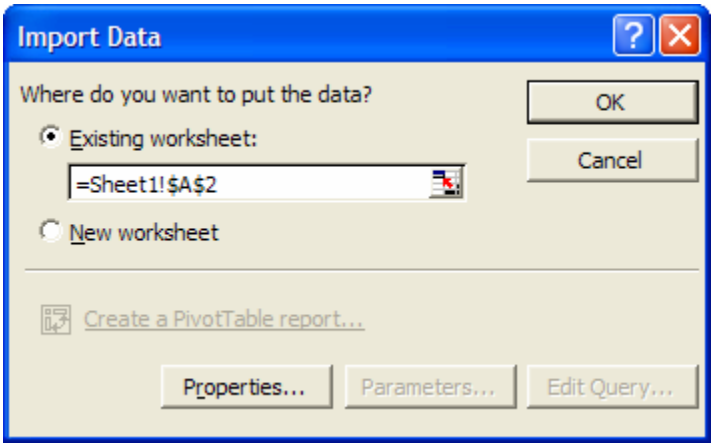


Figure 10.9 Location of the imported data.

The data has now been transferred to the worksheet, as seen in Figure 10.10. Excel has also formed appropriate columns.

	A	B	C	D
1				
2	Last Trade	28.44	Volume	11,944,047
3	Change	-0.04 (-0.14%)	Avg Volume	66,487,540
4	Bid / Ask	28.43 / 28.44	Mkt Cap (mil)	$307,294.22
5	Open	28.49	P/E	31
6	Prev Close	28.48	Div Yield	0.50%
7	Day's Range	28.43 - 28.73	Annual Div/Sh	0.16
8	52-wk Range	22.55 - 30.00	Market	NASDAQ
9				

Figure 10.10 Imported data is separated into columns.

We can now further format this data for future analysis. We have added a title to the table and left a place to enter the date and time when this query was made (see Figure 10.11). We may then keep a record of "Last Trade" values, or others, as we continue to repeat this query in the future. To repeat a query, use the *Refresh Data* option from the *Data* menu. Simply select the range of queried data in the worksheet and select *Data* > *Refresh Data* from the menu. Any changes on the webpage will be reflected on your worksheet.

	A	B	C	D
1	**MSFT Stock (from Quicken)**			
2	Date:	Time:		
3	**Last Trade**	28.44	**Volume**	11,944,047
4	**Change**	-0.04 (-0.14%)	**Avg Volume**	66,487,540
5	**Bid / Ask**	28.43 / 28.44	**Mkt Cap (mil)**	$307,294.22
6	**Open**	28.49	**P/E**	31
7	**Prev Close**	28.48	**Div Yield**	0.50%
8	**Day's Range**	28.43 - 28.73	**Annual Div/Sh**	0.16
9	**52-wk Range**	22.55 - 30.00	**Market**	NASDAQ
10				

Figure 10.11 The web data is organized for analysis.

10.2.3 Databases

To import data from a database, choose *Data* > *Import External Data* from the menu. From here you can browse to find a database file with which you want to work. These files do not need to be other Excel files; they can be any Data Access Object or ActiveX Data Object. The options available to you at this point will all involve an understanding of databases.

To illustrate the concepts in this section, we will use a previously created database in Microsoft Access. This database is titled *Books* and has a short record of a number of books. The fields are: *ISN, Title, Author,* and *Copyright Year*. To work with this data in

Excel, we prepare a new worksheet. First, we choose *Data* > *Import External Data* from the menu; the window in Figure 10.12 will appear.

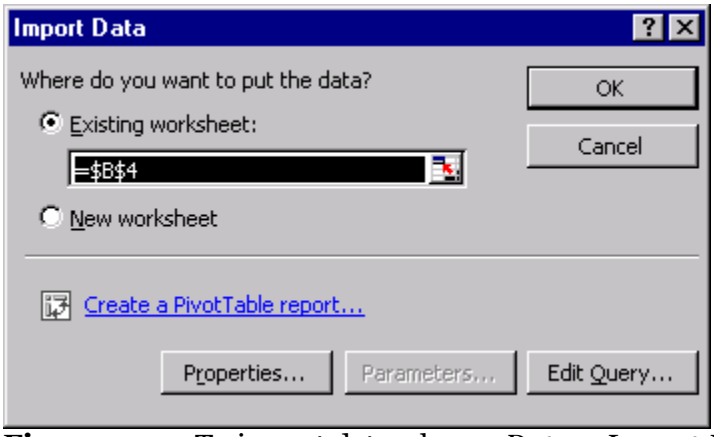


Figure 10.12 To import data, choose *Data* > *Import External Data* from the menu and specify where to place the data.

We then specify where we want the imported data to be located. In this case, we choose a cell in the worksheet, say *B4*, and press *OK*. The imported data will then be copied to the worksheet, as illustrated in Figure 10.13. Notice that an *External Data toolbar* automatically appears next to the data. The first two options on the toolbar are *Edit Query* and *Data Range Properties*. *Edit Query* allows us to modify what part of the database we have imported. A **query** is a search for a particular set of data from the database; it is similar to filtering. The *Data Range Properties* option gives us the opportunity to fine-tune the data before us. We can also update the data by using the  icon. The  icon will ensure that any changes that have been made to the original database are reflected in our Excel worksheet.

	A	B	C	D	E
1	**Importing Data from the Books Database in Access**				
2					
3		**ISN**	**Title**	**Author**	**Copyright Year**
4		1238765645	Programming in VBA	JB White	1999
5		1235679340	Working in Excel	LK Grey	2000
6		1238709643	Engineering Applications	RS Johnson	2001
7		1233457908	Modeling	ID Smith	1999
8		1236543712	Formulations	OB Myers	2000
9		1235485693	Advanced VBA	PW Willson	2000
10					
11			External Data		
12					
13					

Figure 10.13 The data has now been imported into the worksheet in Excel.

A more detailed way to run a query on this database using help from Excel is to run the query before we import the data. To do this, from Excel, choose *Data* > *Import External Data* > *New Database Query*. You will then see the window in Figure 10.14, which configures the data source for the query. Since we are using a database in Microsoft Access, the only change we need to make here is to select *MS Access Database* from the list of *Databases*; then we press *OK*.

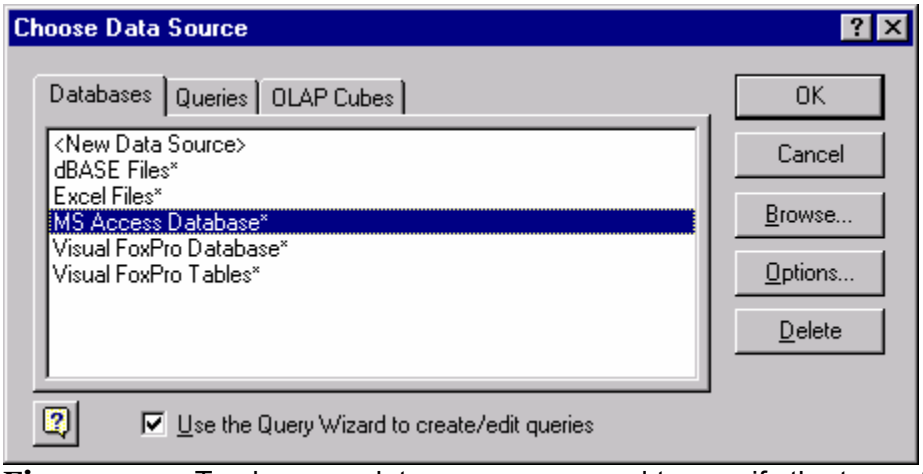


Figure 10.14 To choose a data source, we need to specify the type of database we are using.

The next window will provide a list of files that match this data source type (see Figure 10.15). In the list, we should find the file, b*ooks.mdb;* we select it and press *OK*. Now, we begin to define our query.

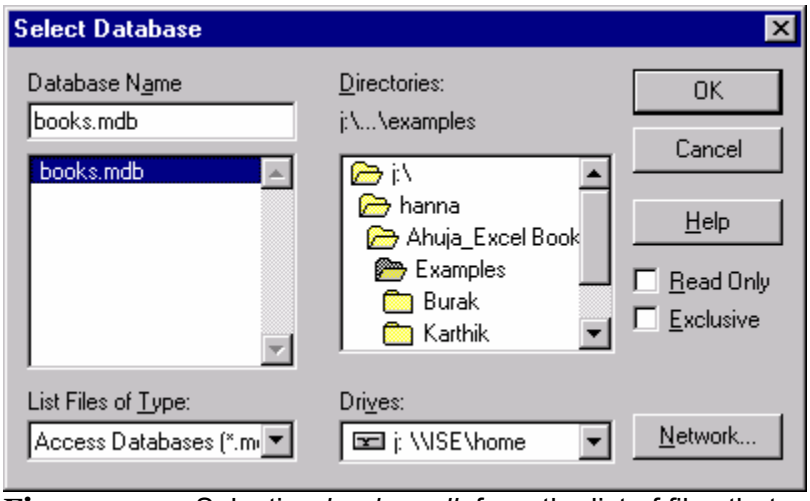


Figure 10.15 Selecting *books.mdb* from the list of files that match our data source.

The first step in defining our query is to specify which columns of data that we want to appear in our final table. We click the + sign next to the table name, *tblbooks,* to see a list of all the column titles. In figure 10.16 we selected the *Title, Author,* and *Copyright Year* columns. We click *Next* to arrive at the next step.

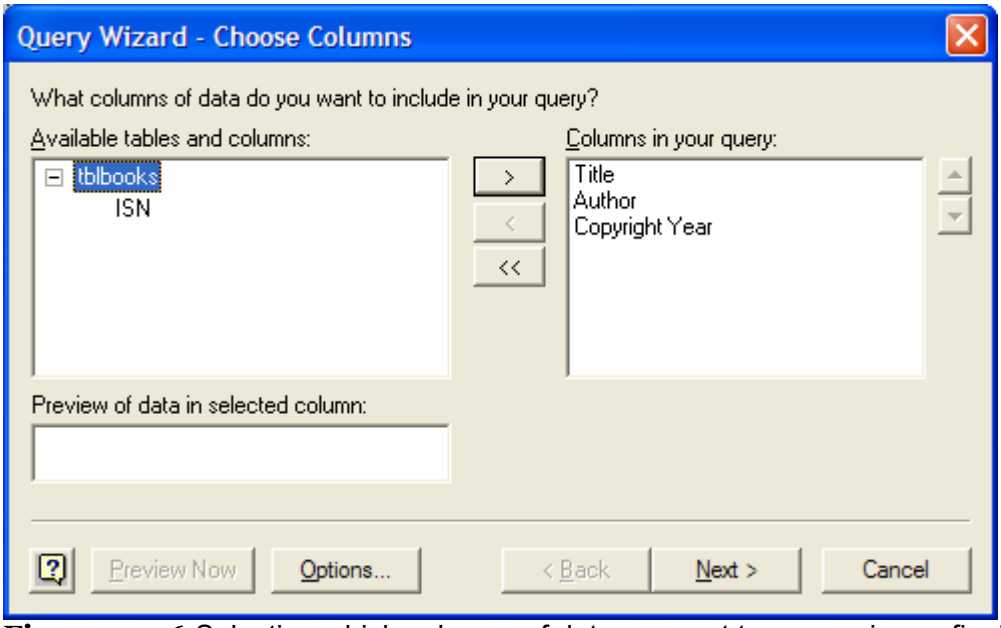


Figure 10.16 Selecting which columns of data we want to appear in our final table.

Now, we specify how we want to filter this data. This is the query definition (see Figure 10.17). First, we select the column title, or field, by which we want to filter our data. Then, we select an equality type and a value from that field. In this case, we decide to filter the data by selecting all entries whose *Copyright Year* is *equal to 2000* and click *Next*. Note that we can have multiple filters among each available field.

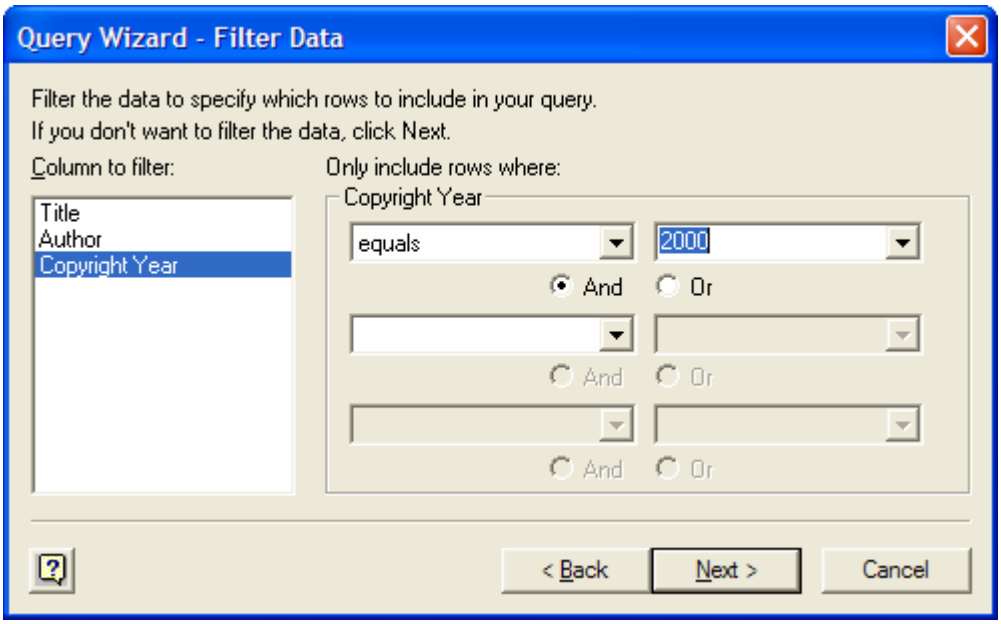


Figure 10.17 Filtering data is how to define a query. We can select one or more fields, equality and a value from the field(s) by which we want to filter the data.

Then, it is necessary that we choose how to sort the data in the final table (see Figure 10.18). We have decided to sort by the *Author* in *Ascending Order*. Again, note that we can have multiple sorts; Section 10.4.1 presents an example of multiple sorting.

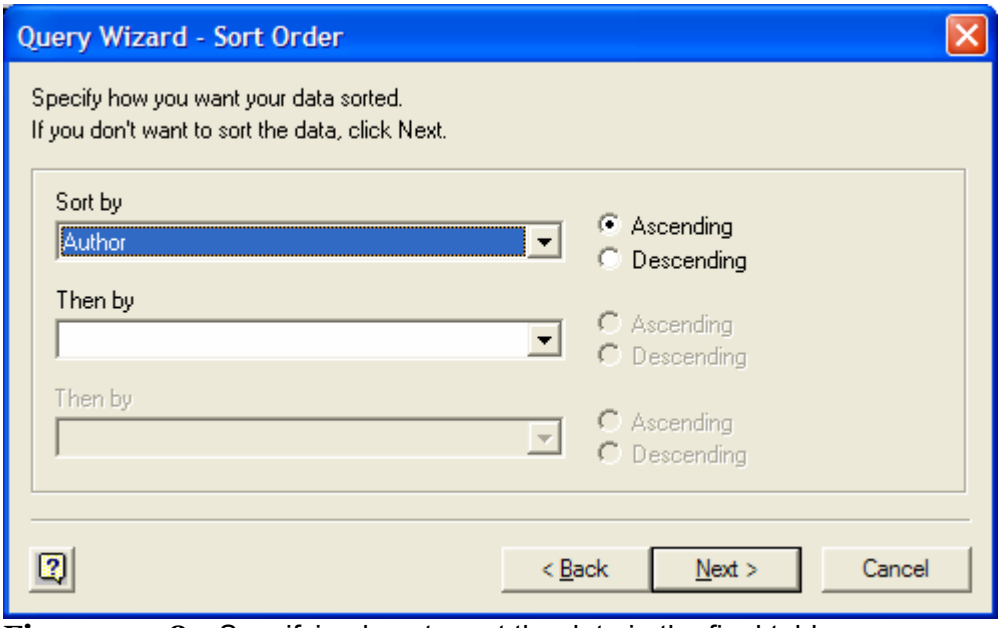


Figure 10.18 Specifying how to sort the data in the final table.

After pressing *Next*, we arrive at the last step in creating our query. This step requires us to decide whether we want to place the data from this query in Excel, view it in its original data source, or create an OLAP cube. The first two options are the most commonly used; we will view our data in Excel (see Figure 10.19). To view the final table, we click *Finish*.

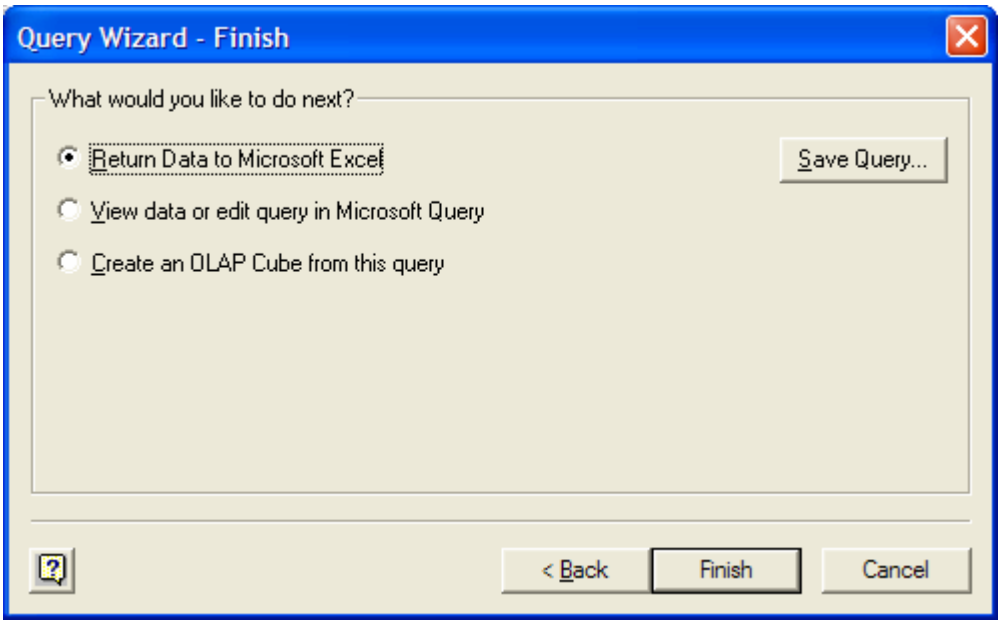


Figure 10.19 Selecting to view the results of the query in Excel.

After we pick a place in our worksheet to situate the table, the result of the query is transferred to Excel. We can format this data in a table and use it in our worksheet as

needed (refer to Figure 10.20). Notice here that only the books with *Copyright Year* of *2000* are shown in the table and that all of the books are sorted alphabetically by *Author*.

	A	B	C	D
1	**Importing Data from the Books Database in Access**			
2				
3		**Title**	**Author**	**Copyright Year**
4		Working in Excel	LK Grey	2000
5		Formulations	ON Myers	2000
6		Advanced VBA	PW Willson	2000
7				

Figure 10.20 Formatting data in a table to use in our worksheet.

Databases constitute an important part of Excel applications. Involving VBA code in database applications extends your options in organizing your data.

10.3 Exporting Data

After working with a database in an Excel worksheet, you can export data to another workbook or to a database in another program. To export data from Excel to a database, you have to use an Add-In called ***AccessLinks***. This Add-In must be downloaded and installed from the Microsoft webpage: http://www.microsoft.com/downloads. (The only valid database available to export using this Add-In is Microsoft Access.) Once it is installed, simply go to the Data menu option and select from the three exporting methods: ***MS Access Form***, ***MS Access Report***, or ***Convert to MS Access*** (see Figure 10.21).

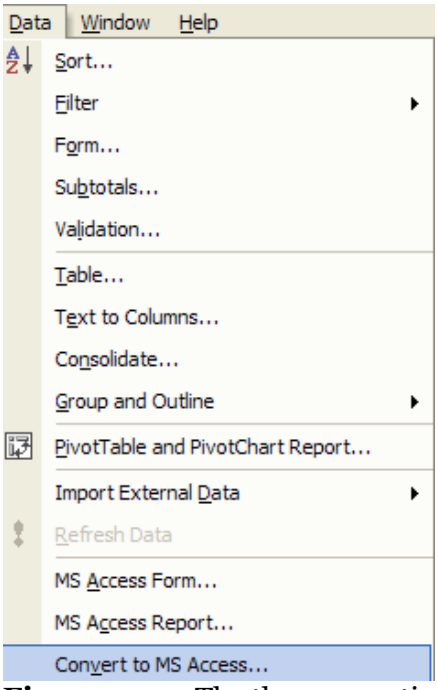


Figure 10.21 The three exporting options available with the *AccessLinks* Add-In appear at the bottom of the *Data* menu.

10.4 Creating Pivot Tables from External Data

When we discussed Pivot Tables in Chapter 6, we mentioned that you can use data not only in your spreadsheet, but also from an external data source. In the first step of the Pivot Table Wizard, we select the type of data that we want to use (see Figure 10.22). To use external data, select *External data source* from the list of options.

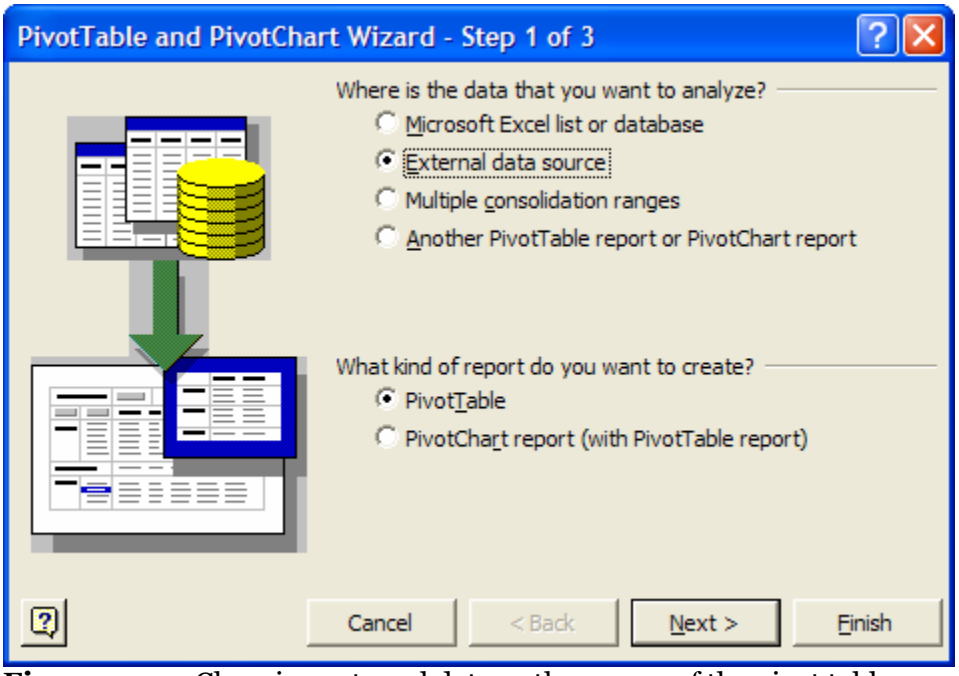


Figure 10.22 Choosing external data as the source of the pivot table.

We will then be prompted to select the actual data (see Figure 10.23). To specify the data type and location, we click *Get Data*.

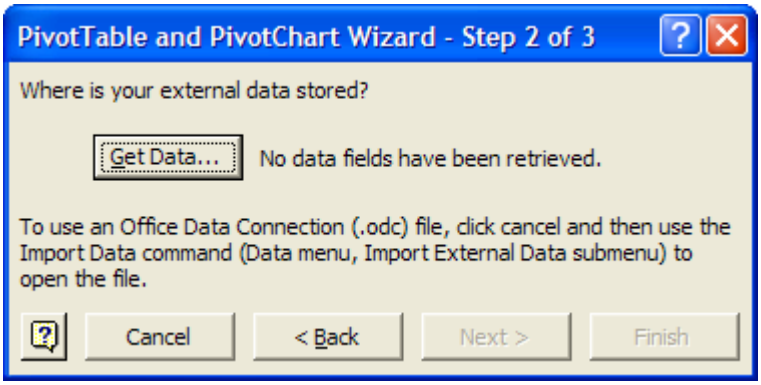


Figure 10.23 Getting the actual data to use for the pivot table.

We will use some example data from a Microsoft Access database file. We select *MS Access Database* from the list of options (see Figure 10.24).

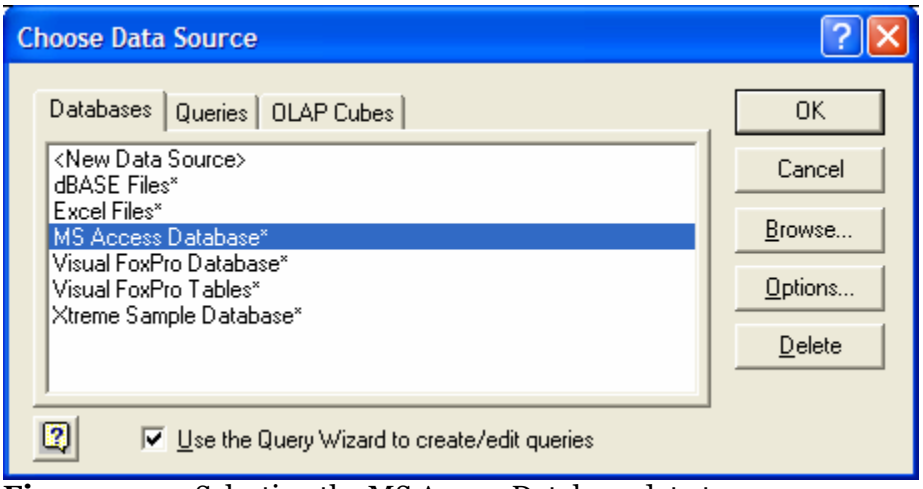


Figure 10.24 Selecting the MS Access Database data type.

We are now prompted to choose the data within our data source to use for the pivot table (see Figure 10.25). In this example database, we have several tables and queries of data for a University System Database. Let's look at the data in the Student Table to make an analysis of the number of students of each nationality and gender by using a pivot table. We select the Student Table from the left window and click the + icon. Then, we select the field names, or column titles, that we are interested in and click the > arrow to move them to the right window. All of the fields listed in the right window will be imported to the pivot table.

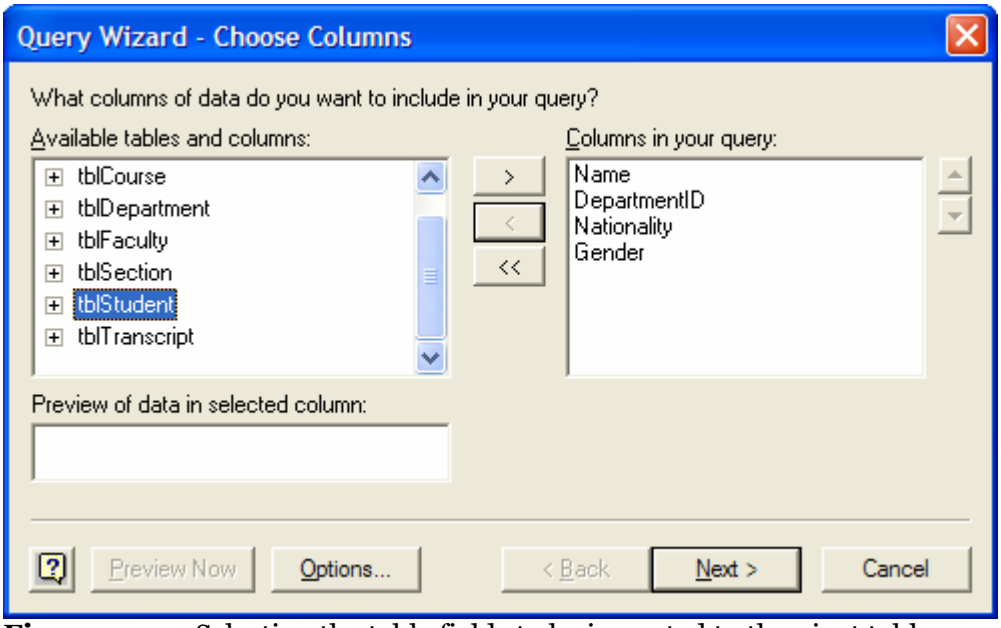


Figure 10.25 Selecting the table fields to be imported to the pivot table.

We are now ready to return to the Pivot Table Wizard. Just as we did in Chapter 6, we specify the layout and location of the pivot table along with any options (see Figure 10.26). Figure 10.27 presents the final pivot table. We can see the number of students per nationality and per gender for each department.

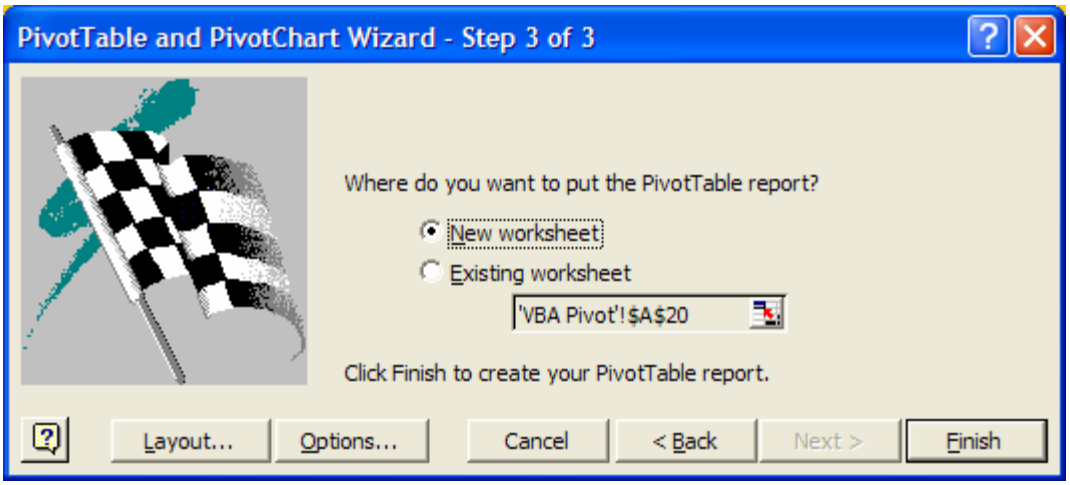


Figure 10.26 The final step of the Pivot Table Wizard.

	A	B	C	D	E	F
1						
2						
3	Count of Name		DepartmentID			
4	Gender	Nationality	CISE	DIS	ISE	Grand Total
5	Female	American			8	8
6		Chinese			4	4
7		Hispanic			4	4
8		Indian		2		2
9	Female Total			2	16	18
10						
11	Male	African American			4	4
12		American		4	4	8
13		Hispanic	2		4	6
14		Russian	2			2
15		Turkish	2		4	6
16	Male Total		6	4	16	26
17						
18	Grand Total		6	6	32	44
19						

Figure 10.27 The final pivot table with the external data.

10.5 Using Excel as a Database

Excel defines any well-defined table of items grouped by similar categories as a *database*. For example, a teaching assistant for a course may develop a detailed table of student records; he could consider this table a database. In this database, he can store the students' names, homework grades, exam scores, and attendance, and he can also calculate their class average in the same table (see Figure 10.28). The basic menu options provide several simple functions that can be performed on Excel databases.

10.5.1 Sorting

The first database function we will discuss is ***sorting***. To sort means to order all entries in a database by a particular ***field***; a field is a category name. (In Figure 10.28, each column heading of the table is a database field name.) In the example shown in Figure 10.28, we see that the names of the students are in no particular order. To place the names in alphabetical order, We can sort the entire database by the field of *Student Name.*

	A	B	C	D	E	F	G	H	I
1									
2									
3		**Course Example**							
4		**Student Name**	**Assignment 1 Grade**	**Assignment 2 Grade**	**Assignment 3 Grade**	**Exam 1 Score**	**Exam 2 Score**	**Attendance (in days absent)**	**Class Average**
5		McDonald, Tom	90.3	88.2	85.3	94.2	94.2	0	91.7
6		Gold, Larry	91.5	87.2	87.2	91.1	95.3	0	91.4
7		Jones, Dave	88.7	86.6	78.7	77.6	80.3	1	81.2
8		Fortatto, Melissa	87.3	85.1	92.9	91.6	97.7	0	92.2
9		Smith, Tracy	85.4	84.7	87.2	84.1	92.5	0	87.3
10		Richards, Amy	90.3	84.6	86.3	86.5	92.6	0	88.6
11		Adams, John	86.6	83.5	83.3	74.0	88.9	1	82.6
12		Edwards, Bill	89.4	83.2	81.8	72.0	80.7	2	79.7
13		Edmonton, Don	93.0	80.3	84.7	88.5	83.0	1	85.9
14		Lexington, Pat	90.0	79.8	89.5	88.8	91.1	0	88.6
15		Mason, Nathan	91.5	79.0	87.4	72.3	90.0	1	83.1
16		Olesman, Daniel	93.7	77.8	80.5	85.4	85.6	0	84.9
17		Nuns, Lisa	89.7	77.2	86.8	73.5	84.2	0	81.1
18		Radcliff, Beth	85.3	76.8	88.1	75.0	81.3	0	80.3
19		Harrison, Sally	87.1	76.1	80.5	81.1	85.2	3	82.4
20		Patterson, Henry	92.4	73.8	86.5	71.3	81.6	1	79.6
21		Gonzalez, Ted	93.0	70.2	80.9	77.3	80.8	1	80.0
22									

Figure 10.28 This database of student records for a course contains eight fields.

To do so, we highlight the entire database, including the field names; this is the range *B4:I21*. Next,we choose *Data* > *Sort* from the menu. We will then see a window similar to Figure 10.29.

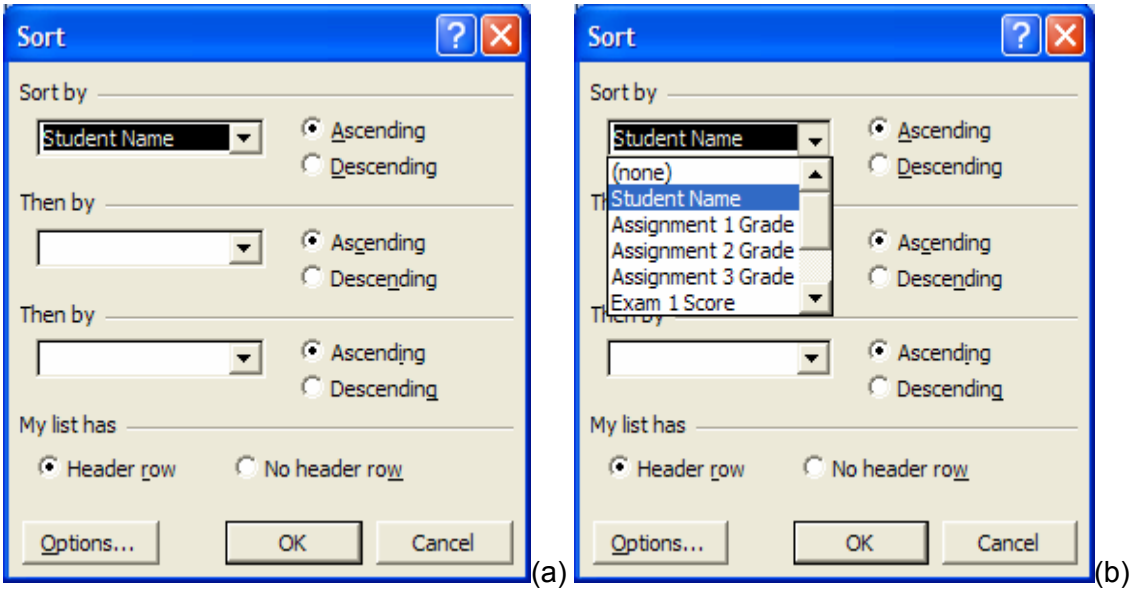


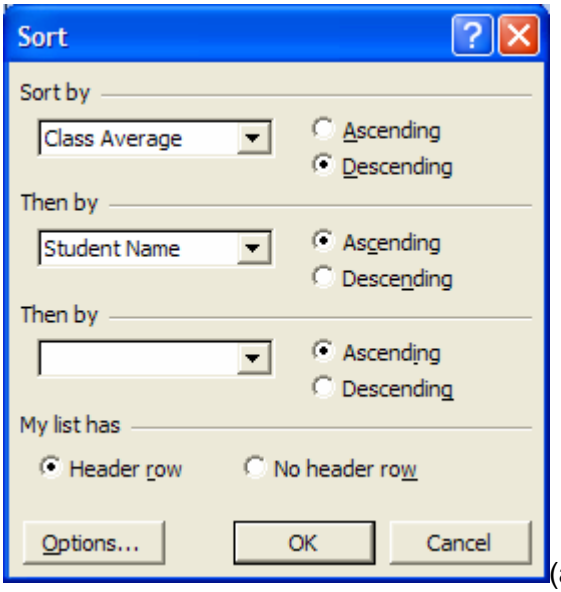
Figure 10.29 (a) The Sort window allows us to choose a field name by which to sort and select *Ascending* or *Descending Order* for the sort. (b) A list of all of the field names.

We can select any field name to sort by; we have chosen *Student Name* in *Ascending Order*. *Ascending Order* will list last names from *A* to *Z*; *Descending Order* would list them in the opposite order. We can see the results of this sort in Figure 10.30. Note that the  and  icons at the top of the window will automatically sort the highlighted database by the first field in *Ascending* and *Descending Order,* respectively.

	A	B	C	D	E	F	G	H	I
1									
2									
3		**Course Example**							
4		**Student Name**	**Assignment 1 Grade**	**Assignment 2 Grade**	**Assignment 3 Grade**	**Exam 1 Score**	**Exam 2 Score**	**Attendance (in days absent)**	**Class Average**
5		Adams, John	86.6	83.5	83.3	74.0	88.9	1	82.6
6		Edmonton, Don	93.0	80.3	84.7	88.5	83.0	1	85.9
7		Edwards, Bill	89.4	83.2	81.8	72.0	80.7	2	79.7
8		Fortatto, Melissa	87.3	85.1	92.9	91.6	97.7	0	92.2
9		Gold, Larry	91.5	87.2	87.2	91.1	95.3	0	91.4
10		Gonzalez, Ted	93.0	70.2	80.9	77.3	80.8	1	80.0
11		Harrison, Sally	87.1	76.1	80.5	81.1	85.2	3	82.4
12		Jones, Dave	88.7	86.6	78.7	77.6	80.3	1	81.2
13		Lexington, Pat	90.0	79.8	89.5	88.8	91.1	0	88.6
14		Mason, Nathan	91.5	79.0	87.4	72.3	90.0	1	83.1
15		McDonald, Tom	90.3	88.2	85.3	94.2	94.2	0	91.7
16		Nuns, Lisa	89.7	77.2	86.8	73.5	84.2	0	81.1
17		Olesman, Daniel	93.7	77.8	80.5	85.4	85.6	0	84.9
18		Patterson, Henry	92.4	73.8	86.5	71.3	81.6	1	79.6
19		Radcliff, Beth	85.3	76.8	88.1	75.0	81.3	0	80.3
20		Richards, Amy	90.3	84.6	86.3	86.5	92.6	0	88.6
21		Smith, Tracy	85.4	84.7	87.2	84.1	92.5	0	87.3
22									

Figure 10.30 The database is now sorted in alphabetical order by *Student Names*.

Note also that in the *Sort* window (Figure 10.29), we can sort by more than one field at a time. For example, in Figure 10.31, we have sorted by *Class Average* in *Descending Order* and then by *Student Name* in *Ascending Order*. Observe that the values for *Class Average* in cells *I8* and *I9* are the same (*88.6*) and the student names are listed alphabetically by last name (*Lexington* and *Richards*). Please note, however, that there is a limit of three fields that you can simultaneously sort by.



(a)

	A	B	C	D	E	F	G	H	I
1									
2									
3		Course Example							
4		Student Name	Assignment 1 Grade	Assignment 2 Grade	Assignment 3 Grade	Exam 1 Score	Exam 2 Score	Attendance (in days absent)	Class Average
5		Fortatto, Melissa	87.3	85.1	92.9	91.6	97.7	0	92.2
6		McDonald, Tom	90.3	88.2	85.3	94.2	94.2	0	91.7
7		Gold, Larry	91.5	87.2	87.2	91.1	95.3	0	91.4
8		Lexington, Pat	90.0	79.8	89.5	88.8	91.1	0	88.6
9		Richards, Amy	90.3	84.6	86.3	86.5	92.6	0	88.6
10		Smith, Tracy	85.4	84.7	87.2	84.1	92.5	0	87.3
11		Edmonton, Don	93.0	80.3	84.7	88.5	83.0	1	85.9
12		Olesman, Daniel	93.7	77.8	80.5	85.4	85.6	0	84.9
13		Mason, Nathan	91.5	79.0	87.4	72.3	90.0	1	83.1
14		Adams, John	86.6	83.5	83.3	74.0	88.9	1	82.6
15		Harrison, Sally	87.1	76.1	80.5	81.1	85.2	3	82.4
16		Jones, Dave	88.7	86.6	78.7	77.6	80.3	1	81.2
17		Nuns, Lisa	89.7	77.2	86.8	73.5	84.2	0	81.1
18		Radcliff, Beth	85.3	76.8	88.1	75.0	81.3	0	80.3
19		Gonzalez, Ted	93.0	70.2	80.9	77.3	80.8	1	80.0
20		Edwards, Bill	89.4	83.2	81.8	72.0	80.7	2	79.7
21		Patterson, Henry	92.4	73.8	86.5	71.3	81.6	1	79.6
22									

(b)

Figure 10.31 First sorting by *Class Average* in *Descending Order* and then by *Student Name* in *Ascending Order*. (b) The database is now sorted.

There is one more option to specify in the *Sort* window; this is the *Header Row* option. It simply informs Excel whether or not the field names are already specified by column names in a *header row*. If there is a *header row*, which we have marked in the above example, then the field names appear in a list as column titles. If, however, we had selected *No header row*, there would only be column names in the field list (see Figure 10.32).

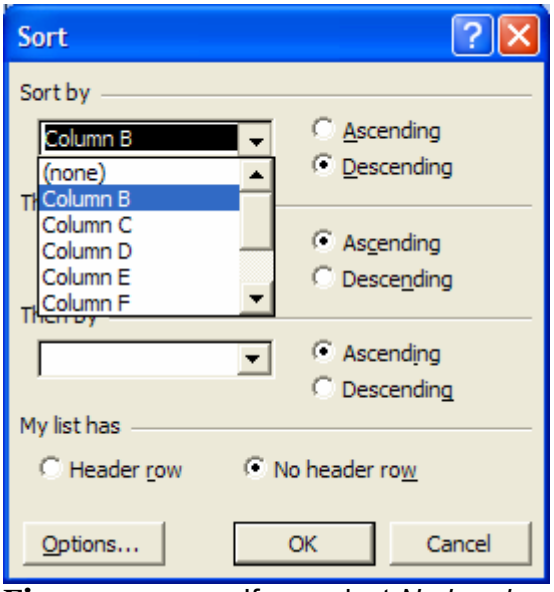


Figure 10.32 If we select *No header row,* the Excel column names are used as field names instead of as column titles in the database.

We can also sort within a row. To do this, we click the *Options* button on the Sort window. The window shown in Figure 10.33 appears. There is an option to determine how the first field we sort by should be sorted; the default value is *Normal*; other options refer to date values. We can also select whether our data is *Case Sensitive* or not. To sort within a row, we will need to change the *Orientation* option from "top to bottom" to "left to right."

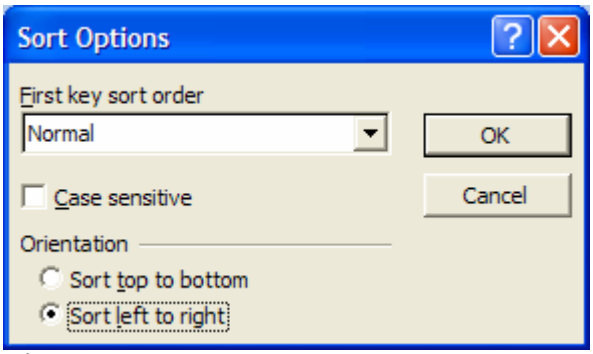


Figure 10.33 Sort Options include sorting from top to bottom or left to right.

Once we press *OK* and return to the main Sort window, wel notice that the field lists show row numbers instead of column numbers (see Figure 10.34). These numbers refer to the row numbering on the spreadsheet; they are not relevant to our table.

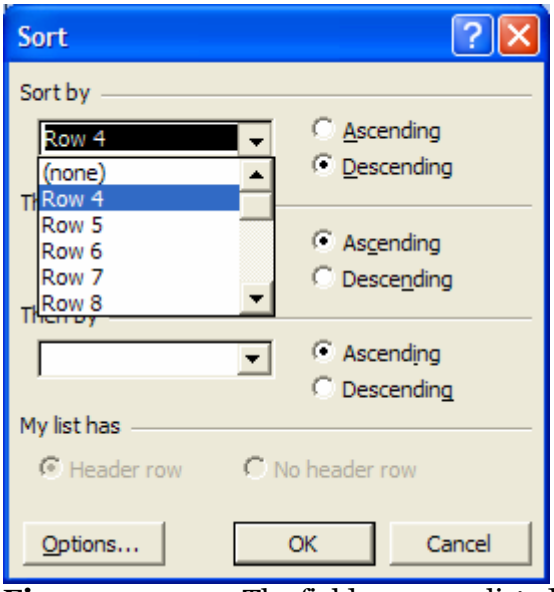


Figure 10.34 The fields are now listed by row number.

Using the student data, let's choose row 4 to sort by since it has the titles of the different grades. If we sort by *Row 4* in *Ascending* order, we will see the table shown in Figure 10.35.

	A	B	C	D	E	F	G	H	I
1									
2									
3		Course Example							
4		Assignment 1 Grade	Assignment 2 Grade	Assignment 3 Grade	Attendance (in days absent)	Class Average	Exam 1 Score	Exam 2 Score	Student Name
5		87.3	85.1	92.9	0	92.2	91.6	97.7	Fortatto, Melissa
6		90.3	88.2	85.3	0	91.7	94.2	94.2	McDonald, Tom
7		91.5	87.2	87.2	0	91.4	91.1	95.3	Gold, Larry
8		90.0	79.8	89.5	0	88.6	88.8	91.1	Lexington, Pat
9		90.3	84.6	86.3	0	88.6	86.5	92.6	Richards, Amy
10		85.4	84.7	87.2	0	87.3	84.1	92.5	Smith, Tracy
11		93.0	80.3	84.7	1	85.9	88.5	83.0	Edmonton, Don
12		93.7	77.8	80.5	0	84.9	85.4	85.6	Olesman, Daniel
13		91.5	79.0	87.4	1	83.1	72.3	90.0	Mason, Nathan
14		86.6	83.5	83.3	1	82.6	74.0	88.9	Adams, John
15		87.1	76.1	80.5	3	82.4	81.1	85.2	Harrison, Sally
16		88.7	86.6	78.7	1	81.2	77.6	80.3	Jones, Dave
17		89.7	77.2	86.8	0	81.1	73.5	84.2	Nuns, Lisa
18		85.3	76.8	88.1	0	80.3	75.0	81.3	Radcliff, Beth
19		93.0	70.2	80.9	1	80.0	77.3	80.8	Gonzalez, Ted
20		89.4	83.2	81.8	2	79.7	72.0	80.7	Edwards, Bill
21		92.4	73.8	86.5	1	79.6	71.3	81.6	Patterson, Henry
22									

Figure 10.35 The table is now sorted by row 4.

Sorting is a useful database function as it allows us to organize our data in multiple ways. It becomes even more valuable as the size of our database grows. It can also function as a search tool. For example, it can help us find an item name in an alphabetical listing.

Summary

Sorting: Ordering all entries in a database by a particular field.

Field: Name of a category by which items are grouped in a database.

10.5.2 Filtering

The next database function we will discuss is ***filtering***. Filtering differs from sorting in that it selects a specified set of data from the database instead of ordering the entire database. Filtering allows you to select a group of entries in a database that is equal to a particular data entry within a field. For example, referring to Figure 10.28, instead of ordering all of the data by sorting by *Class Average*, we could select to view only the row items that have a *Class Average* equal to *88.6*. To do so, we begin by highlighting the entire database, including column titles, cells *B4:I21*. Then,we choose *Data* > *Filter* > *Auto Filter* from the menu. This selection will transform the database by adding drop-down arrows to each field (see Figure 10.36).

	A	B	C	D	E	F	G	H	I
1									
2									
3		Course Example							
4		Student Nam	Assignment 1 Grade	Assignment 2 Grade	Assignment 3 Grade	Exam 1 Score	Exam 2 Score	Attendance (in days absent)	Class Averag
5		Adams, John	86.6	83.5	83.3	74.0	88.9	1	82.6
6		Edmonton, Don	93.0	80.3	84.7	88.5	83.0	1	85.9
7		Edwards, Bill	89.4	83.2	81.8	72.0	80.7	2	79.7
8		Fortatto, Melissa	87.3	85.1	92.9	91.6	97.7	0	92.2
9		Gold, Larry	91.5	87.2	87.2	91.1	95.3	0	91.4
10		Gonzalez, Ted	93.0	70.2	80.9	77.3	80.8	1	80.0
11		Harrison, Sally	87.1	76.1	80.5	81.1	85.2	3	82.4
12		Jones, Dave	88.7	86.6	78.7	77.6	80.3	1	81.2
13		Lexington, Pat	90.0	79.8	89.5	88.8	91.1	0	88.6
14		Mason, Nathan	91.5	79.0	87.4	72.3	90.0	1	83.1
15		McDonald, Tom	90.3	88.2	85.3	94.2	94.2	0	91.7
16		Nuns, Lisa	89.7	77.2	86.8	73.5	84.2	0	81.1
17		Olesman, Daniel	93.7	77.8	80.5	85.4	85.6	0	84.9
18		Patterson, Henry	92.4	73.8	86.5	71.3	81.6	1	79.6
19		Radcliff, Beth	85.3	76.8	88.1	75.0	81.3	0	80.3
20		Richards, Amy	90.3	84.6	86.3	86.5	92.6	0	88.6
21		Smith, Tracy	85.4	84.7	87.2	84.1	92.5	0	87.3

Figure 10.36 The database is auto-filtered with selection options for each field.

We then select the arrow attached to the *Class Average* field. Here, we determine a particular data value, say *88.6*, that will reconstruct the database so that it only displays the rows whose *Class Average* entries match the value chosen (see Figure 10.37). We can again display the entire database by clicking on the option arrow of *Class Average* and selecting *(All)*.

	A	B	C	D	E	F	G	H	I
1									
2									
3		Course Example							
4		Student Nam	Assignment 1 Grade	Assignment 2 Grade	Assignment 3 Grade	Exam 1 Score	Exam 2 Score	Attendance (in days absent)	Class Averag
5		Adams, John	86.6	83.5	83.3	74.0	88.9	1	
6		Edmonton, Don	93.0	80.3	84.7	88.5	83.0	1	
7		Edwards, Bill	89.4	83.2	81.8	72.0	80.7	2	
8		Fortatto, Melissa	87.3	85.1	92.9	91.6	97.7	0	
9		Gold, Larry	91.5	87.2	87.2	91.1	95.3	0	
10		Gonzalez, Ted	93.0	70.2	80.9	77.3	80.8	1	
11		Harrison, Sally	87.1	76.1	80.5	81.1	85.2	3	
12		Jones, Dave	88.7	86.6	78.7	77.6	80.3	1	
13		Lexington, Pat	90.0	79.8	89.5	88.8	91.1	0	
14		Mason, Nathan	91.5	79.0	87.4	72.3	90.0	1	
15		McDonald, Tom	90.3	88.2	85.3	94.2	94.2	0	
16		Nuns, Lisa	89.7	77.2	86.8	73.5	84.2	0	
17		Olesman, Daniel	93.7	77.8	80.5	85.4	85.6	0	
18		Patterson, Henry	92.4	73.8	86.5	71.3	81.6	1	
19		Radcliff, Beth	85.3	76.8	88.1	75.0	81.3	0	
20		Richards, Amy	90.3	84.6	86.3	86.5	92.6	0	88.6
21		Smith, Tracy	85.4	84.7	87.2	84.1	92.5	0	87.3

(All)
(Top 10...)
(Custom...)
79.6
79.7
80.0
80.3
81.1
81.2
82.4
82.6
83.1
84.9
85.9
87.3
88.6
91.4
91.7
92.2

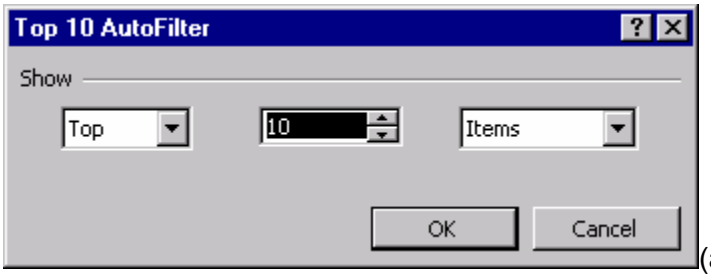
(a)

	A	B	C	D	E	F	G	H	I
1									
2									
3		Course Example							
4		Student Nam	Assignment 1 Grade	Assignment 2 Grade	Assignment 3 Grade	Exam 1 Score	Exam 2 Score	Attendance (in days absent)	Class Averag
13		Lexington, Pat	90.0	79.8	89.5	88.8	91.1	0	88.6
20		Richards, Amy	90.3	84.6	86.3	86.5	92.6	0	88.6
22									

(b)

Figure 10.37 (a) Selecting the drop-down list of values in the *Class Average* field. (b)The database is filtered to only show data entries with a *Class Average* equal to *88.6*.

There are two more options on the filtered drop-down list in each field that we would like to discuss. One is *Top 10* and the other is *Custom*. Selecting *Top 10* allows us to display 10 entries (or another specified amount) or 10 percent (or another specified percent) of the entries on the top or bottom of the database. In Figure 10.38, we chose to show the top 10 entries.



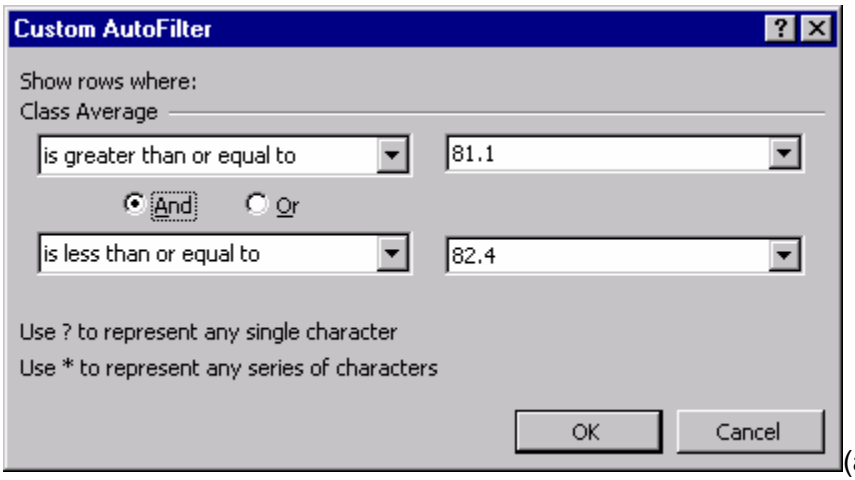
(a)

	A	B	C	D	E	F	G	H	I
1									
2									
3		Course Example							
4		Student Nam	Assignment 1 Grade	Assignment 2 Grade	Assignment 3 Grade	Exam 1 Score	Exam 2 Score	Attendance (in days absent)	Class Averag
5		Adams, John	86.6	83.5	83.3	74.0	88.9	1	82.6
6		Edmonton, Don	93.0	80.3	84.7	88.5	83.0	1	85.9
8		Fortatto, Melissa	87.3	85.1	92.9	91.6	97.7	0	92.2
9		Gold, Larry	91.5	87.2	87.2	91.1	95.3	0	91.4
13		Lexington, Pat	90.0	79.8	89.5	88.8	91.1	0	88.6
14		Mason, Nathan	91.5	79.0	87.4	72.3	90.0	1	83.1
15		McDonald, Tom	90.3	88.2	85.3	94.2	94.2	0	91.7
17		Olesman, Daniel	93.7	77.8	80.5	85.4	85.6	0	84.9
20		Richards, Amy	90.3	84.6	86.3	86.5	92.6	0	88.6
21		Smith, Tracy	85.4	84.7	87.2	84.1	92.5	0	87.3
22									

(b)

Figure 10.38 Choosing the *Top 10* option from the drop-down list. (a) Specifying *Top* or *Bottom*, a certain number, and *items* or *percent*. (b) The database is updated to show only the number of entries specified.

The *Custom* option gives you a set of inequality specifications. You can choose to view entries whose field value is less than, greater than, or equal to a certain value. There is also an *And*/*Or* option which allows you to narrow your search to a particular range of values. For example, in Figure 10.39, we have specified to list only the entries whose *Class Average* is between the values *81.1* and *82.4*.



(a)

	A	B	C	D	E	F	G	H	I
1									
2									
3		Course Example							
4		Student Nam	Assignment 1 Grade	Assignment 2 Grade	Assignment 3 Grade	Exam 1 Score	Exam 2 Score	Attendance (in days absent)	Class Averag
11		Harrison, Sally	87.1	76.1	80.5	81.1	85.2	3	82.4
12		Jones, Dave	88.7	86.6	78.7	77.6	80.3	1	81.2
16		Nuns, Lisa	89.7	77.2	86.8	73.5	84.2	0	81.1
22									

(b)

Figure 10.39 Selecting the *Custom* option. (a) Specifying to show only entries with *Class Average* values greater than or equal to 81.1 and less than or equal to 82.4. (b) The database is updated to show only these entries.

Filtering is useful for locating specific data correlations in a database. You can define a select group of data by choosing a field entry that they have in common. There are many applications of databases for which filtering will be a necessary tool.

10.5.3 DFunctions

There is a group of Excel functions that are meant specifically for working with Excel as a database; we call these functions ***Dfunctions***. These are specific functions designed for use with databases. They include **DSUM**, **DAVERAGE**, **DMIN**, and **DMAX,** which are essentially the same functions as SUM, AVERAGE, MIN and MAX, which we discussed in Chapter 4. Dfunctions differ from the previously described functions because they specify certain criteria before performing the function. That is, the formats of these functions have extra parameters (we use DSUM to illustrate the general format):

=DSUM(database, field, criteria)

The *criteria* parameter must include a field name and a criteria cell. In preparation for using *Dfunctions*, we must add a few rows to our database (see Figure 10.40). These rows contain our criteria and must repeat our field names. We will define the *Dfunction* for this example in a new cell below these criteria. Using the example from Figure 10.28 one more time, let's find the minimum *Exam 2* score for students with a *Class Average* above *85.0*.

	A	B	C	D	E	F	G	H	I
1									
2		**Course Example**							
3		**Student Name**	**Assignment 1 Grade**	**Assignment 2 Grade**	**Assignment 3 Grade**	**Exam 1 Score**	**Exam 2 Score**	**Attendance (in days absent)**	**Class Average**
4		Adams, John	88.4	89.3	81.5	73	84.1	1	81.7
5		Edmonton, Don	88.5	81.7	82.1	85.6	82.9	1	84.2
6		Edwards, Bill	89.2	87.5	82.4	74.6	85.3	2	82.5
7		Fortatto, Melissa	88.6	88	89.1	88.6	91.6	0	89.5
8		Gold, Larry	89.7	82.6	85.8	91.2	91	0	89.1
9		Gonzalez, Ted	89.3	72.1	80.8	73.4	86.3	1	80.2
10		Harrison, Sally	85.8	74.2	86	85.9	85.6	3	84.3
11		Jones, Dave	94.3	85.1	83.4	72	84.8	1	82.1
12		Lexington, Pat	85.1	78.8	89.5	83.4	96.3	0	87.7
13		Mason, Nathan	87.3	72.9	82.1	74.5	80.7	1	78.9
14		McDonald, Tom	91	80.4	86.8	86.1	91.4	0	87.7
15		Nuns, Lisa	94.6	75.6	86.7	70.7	87.3	0	81.7
16		Olesman, Daniel	87.6	73.8	89.1	88.6	82.6	0	84.8
17		Patterson, Henry	90.4	74.1	80.3	73.9	80.5	1	79.0
18		Radcliff, Beth	93.2	72.9	87.2	71.8	86.9	0	81.4
19		Richards, Amy	91.3	80.9	92.9	89.7	94.1	0	90.5
20		Smith, Tracy	93.5	87	86.5	86.9	95.7	0	90.4
21		**Student Name**	**Assignment 1 Grade**	**Assignment 2 Grade**	**Assignment 3 Grade**	**Exam 1 Score**	**Exam 2 Score**	**Attendance (in days absent)**	**Class Average**
22		**Criteria:**						1	> 85
23									

Figure 10.40 Adding criteria rows to the bottom of the database with the repeated field names. The criterion specified in *I22* is associated with the *DMIN* function that finds the minimum *Exam 2* score of students with a *Class Average* above *85*.

To do this, we choose *Insert* > *Function* from the menu, or select the f_x icon from the standard toolbar and find the function category labeled *Database*. We can now view all of the *Dfunctions* and choose **DMIN** from the list (see Figure 10.41). After selecting DMIN, a new window will appear that guides us through defining the function.

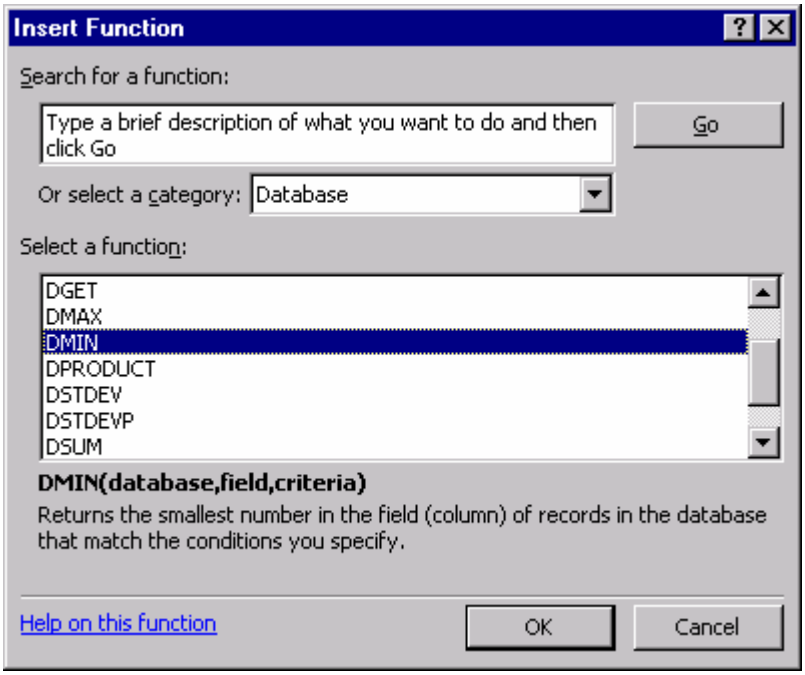


Figure 10.41 In the function window, selecting the *Database* category to view all *Dfunctions*.

Here we begin by defining the *database* parameter, which is the entire database excluding the new criteria rows, cells *B3:I20*. The *field* parameter is the name of the field in which we are searching for a minimum value, in this case, the *Exam 2 Score*. We specify the minimum value by entering the cell with this field name, which is *G3*. (We could also just enter the title of the field here.) Lastly, the *criteria* parameter includes the field name and criteria value cells. In this scenario, we want to find the minimum *Exam 2* score for only the students with a *Class Average* above *85*. Therefore, in the criteria cells, *>85* is the criterion in the *Class Average* field. So for this parameter, we enter the cells *I21:I22.* The final DMIN function is then as follows (also see Figure 10.42):

=DMIN(B3:I23, G3, I21:I22)

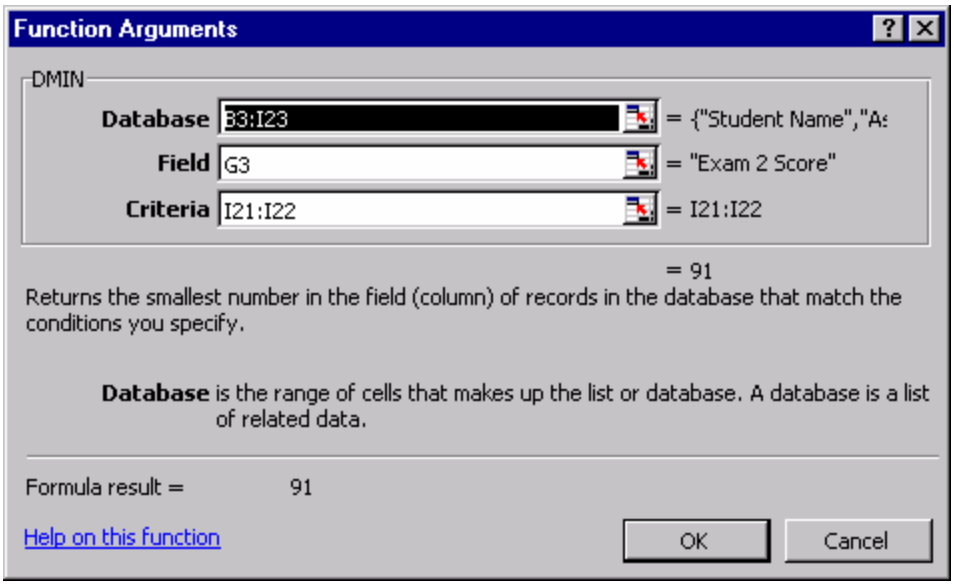


Figure 10.42 Each argument is defined here for the DMIN function. Explanations of each argument are provided below the entry fields.

After entering this *Dfunction,* we will find the minimum *Exam 2 Score* to be *91* (see Figure 10.43(a)).

We have illustrated another scenario that searches for the minimum *Exam 2 Score* for the students with exactly *1* absence. In this case, the number *1* is the criterion for the *Attendance* field. Notice that the DMIN function in this case has the same *database* and *field* parameters but different *criteria* since we are now reviewing a condition in the *Attendance* field instead of the *Class Average* field (see Figure 10.43(b)).

=DMIN(B3:I23, G3, H21:H22)

F26 =DMIN($B$3:$I$20,G3,I21:I22)

	A	B	C	D	E	F	G	H	I
1									
2		**Course Example**							
3		**Student Name**	**Assignment 1 Grade**	**Assignment 2 Grade**	**Assignment 3 Grade**	**Exam 1 Score**	**Exam 2 Score**	**Attendance (in days absent)**	**Class Average**
4		Adams, John	88.4	89.3	81.5	73	84.1	1	81.7
5		Edmonton, Don	88.5	81.7	82.1	85.6	82.9	1	84.2
6		Edwards, Bill	89.2	87.5	82.4	74.6	85.3	2	82.5
7		Fortatto, Melissa	88.6	88	89.1	88.6	91.6	0	89.5
8		Gold, Larry	89.7	82.6	85.8	91.2	91	0	89.1
9		Gonzalez, Ted	89.3	72.1	80.8	73.4	86.3	1	80.2
10		Harrison, Sally	85.8	74.2	86	85.9	85.6	3	84.3
11		Jones, Dave	94.3	85.1	83.4	72	84.8	1	82.1
12		Lexington, Pat	85.1	78.8	89.5	83.4	96.3	0	87.7
13		Mason, Nathan	87.3	72.9	82.1	74.5	80.7	1	78.9
14		McDonald, Tom	91	80.4	86.8	86.1	91.4	0	87.7
15		Nuns, Lisa	94.6	75.6	86.7	70.7	87.3	0	81.7
16		Olesman, Daniel	87.6	73.8	89.1	88.6	82.6	0	84.8
17		Patterson, Henry	90.4	74.1	80.3	73.9	80.5	1	79.0
18		Radcliff, Beth	93.2	72.9	87.2	71.8	86.9	0	81.4
19		Richards, Amy	91.3	80.9	92.9	89.7	94.1	0	90.5
20		Smith, Tracy	93.5	87	86.5	86.9	95.7	0	90.4
21		**Student Name**	**Assignment 1 Grade**	**Assignment 2 Grade**	**Assignment 3 Grade**	**Exam 1 Score**	**Exam 2 Score**	**Attendance (in days absent)**	**Class Average**
22		**Criteria:**						1	> 85
23									
24									
25		**Minimum Exam 2 Score**							
26			**Of Students with A Class Average > 85**			91.0			
27			**Of Students with 1 Absence**			80.5			
28									

(a)

F27 =DMIN($B$3:$I$20,G3, H21:H22)

	A	B	C	D	E	F	G	H	I
1									
2		**Course Example**							
3		**Student Name**	**Assignment 1 Grade**	**Assignment 2 Grade**	**Assignment 3 Grade**	**Exam 1 Score**	**Exam 2 Score**	**Attendance (in days absent)**	**Class Average**
4		Adams, John	88.4	89.3	81.5	73	84.1	1	81.7
5		Edmonton, Don	88.5	81.7	82.1	85.6	82.9	1	84.2
6		Edwards, Bill	89.2	87.5	82.4	74.6	85.3	2	82.5
7		Fortatto, Melissa	88.6	88	89.1	88.6	91.6	0	89.5
8		Gold, Larry	89.7	82.6	85.8	91.2	91	0	89.1
9		Gonzalez, Ted	89.3	72.1	80.8	73.4	86.3	1	80.2
10		Harrison, Sally	85.8	74.2	86	85.9	85.6	3	84.3
11		Jones, Dave	94.3	85.1	83.4	72	84.8	1	82.1
12		Lexington, Pat	85.1	78.8	89.5	83.4	96.3	0	87.7
13		Mason, Nathan	87.3	72.9	82.1	74.5	80.7	1	78.9
14		McDonald, Tom	91	80.4	86.8	86.1	91.4	0	87.7
15		Nuns, Lisa	94.6	75.6	86.7	70.7	87.3	0	81.7
16		Olesman, Daniel	87.6	73.8	89.1	88.6	82.6	0	84.8
17		Patterson, Henry	90.4	74.1	80.3	73.9	80.5	1	79.0
18		Radcliff, Beth	93.2	72.9	87.2	71.8	86.9	0	81.4
19		Richards, Amy	91.3	80.9	92.9	89.7	94.1	0	90.5
20		Smith, Tracy	93.5	87	86.5	86.9	95.7	0	90.4
21		**Student Name**	**Assignment 1 Grade**	**Assignment 2 Grade**	**Assignment 3 Grade**	**Exam 1 Score**	**Exam 2 Score**	**Attendance (in days absent)**	**Class Average**
22		**Criteria:**						1	> 85
23									
24									
25		**Minimum Exam 2 Score**							
26			**Of Students with A Class Average > 85**			91.0			
27			**Of Students with 1 Absence**			80.5			
28									

(b)

Figure 10.43 The DMIN function uses the criteria in the *Class Average* field. (b) The DMIN function uses the criteria in the *Attendance* field.

The other *Dfunctions* can be defined in a similar fashion. After clicking the f_x icon and choosing a *Dfunction* from the *Database* category, the *Function Argument* window will guide you through defining your database and criteria for which you want to perform the selected function. *Dfunctions* are useful for databases as they allow you to consider many options in your search for a particular value.

 **Summary**	**Dfunctions:** Functions with extra criteria specified for database use. **DSUM, DAVERAGE, DMIN, DMAX**

Some other functions that are not *Dfunctions* are also useful to manage database data in Excel. One group of these functions, **COUNT** functions, has four main sub-functions: **COUNT**, **COUNTA**, **COUNTBLANK**, and **COUNTIF**. The **COUNT** function simply counts the number of cells with numerical values in a given range. The format of the function is as follows:

=COUNT(range)

The **COUNTA** function counts all of the cells with data of any kind in a given range of cells. That is, it will count the number of cells with numbers or text, or any other non-blank cell. The format for this function is:

=COUNTA(range)

The **COUNTBLANK** function counts the number of blank cells in a given range. It can be helpful for finding empty entries in a data table. The format for this function is:

=COUNTBLANK(range)

Let's use these three functions on the student data in the previous example. Suppose we want to determine the number of assignments in the record. To use the COUNT function, we must highlight a row of grades, excluding the row of assignment titles. We must exclude the row of assignment titles because the COUNT function only counts cells with numerical values. If we use the COUNT function with the cells *C4:G4*, we find that 5 assignments have been recorded. To count the number of students in the class, we can use the COUNTA function with the first column of data (*B4:B20*); the result is 17 students.

Now suppose that we want to determine the number of students who never had an absence. We can remove the "0"s from the column of absences so that there are blanks instead. Now we can use the COUNTBLANK function with the attendance column (H4:H20); the result is that 9 students have perfect attendance records. See Figure 10.44 for these results.

F31 =COUNTBLANK(H4:H20)

	A	B	C	D	E	F	G	H	I	J
1										
2		Course Example								
3		Student Name	Assignment 1 Grade	Assignment 2 Grade	Assignment 3 Grade	Exam 1 Score	Exam 2 Score	Attendance (in days absent)	Class Average	
4		Adams, John	88.4	89.3	81.5	73	84.1	1	81.7	
5		Edmonton, Don	88.5	81.7	82.1	85.6	82.9	1	84.2	
6		Edwards, Bill	89.2	87.5	82.4	74.6	85.3	2	82.5	
7		Fortatto, Melissa	88.6	88	89.1	88.6	91.6		89.5	
8		Gold, Larry	89.7	82.6	85.8	91.2	91		89.1	
9		Gonzalez, Ted	89.3	72.1	80.8	73.4	86.3	1	80.2	
10		Harrison, Sally	85.8	74.2	86	85.9	85.6	3	84.3	
11		Jones, Dave	94.3	85.1	83.4	72	84.8	1	82.1	
12		Lexington, Pat	85.1	78.8	89.5	83.4	96.3		87.7	
13		Mason, Nathan	87.3	72.9	82.1	74.5	80.7	1	78.9	
14		McDonald, Tom	91	80.4	86.8	86.1	91.4		87.7	
15		Nuns, Lisa	94.6	75.6	86.7	70.7	87.3		81.7	
16		Olesman, Daniel	87.6	73.8	89.1	88.6	82.6		84.8	
17		Patterson, Henry	90.4	74.1	80.3	73.9	80.5	1	79.0	
18		Radcliff, Beth	93.2	72.9	87.2	71.8	86.9		81.4	
19		Richards, Amy	91.3	80.9	92.9	89.7	94.1		90.5	
20		Smith, Tracy	93.5	87	86.5	86.9	95.7		90.4	
21		Student Name	Assignment 1 Grade	Assignment 2 Grade	Assignment 3 Grade	Exam 1 Score	Exam 2 Score	Attendance (in days absent)	Class Average	
22		Criteria:						1	> 85	
23										
24										
28										
29				Number of assignments		5				
30				Number of students		17				
31		Number of students with perfect attendance				9				
32										

Figure 10.44 Using the COUNT, COUNTA, and COUNTBLANK functions.

The **COUNTIF** function is probably the most useful COUNT function. It counts the number of cells in a given range that meet a specified criterion. The format for this function is:

=COUNTIF(range, criterion)

The criterion can contain some helpful characters such as * for a sequence of unknown values or *?* as a single wild card value.

Let's use the COUNTIF function to determine the number of students who have an A (≥ 90), a B+ (≥ 85), or a B (≥ 80) as their final grades. First, we need to find the number of students with an A by using the COUNTIF function as follows:

=COUNTIF(I4:I20, ">=90")

Now, to determine the number of students with a B+, we want to count the grades that are < 90 and ≥ 85. However, since we cannot have multiple criteria with this function, we will first just count the number of students with final averages ≥ 85. This will result in the number of students with an A or a B+. Then, we subtract the previous result from this result to find the number of students with B+'s only. We repeat this process to find students with a B. See Figure 10.45 for the results.

F34 =COUNTIF(I4:I20,">=85")

	A	B	C	D	E	F	G	H	I	J
1										
2		Course Example								
3		Student Name	Assignment 1 Grade	Assignment 2 Grade	Assignment 3 Grade	Exam 1 Score	Exam 2 Score	Attendance (in days absent)	Class Average	
4		Adams, John	88.4	89.3	81.5	73	84.1	1	81.7	
5		Edmonton, Don	88.5	81.7	82.1	85.6	82.9	1	84.2	
6		Edwards, Bill	89.2	87.5	82.4	74.6	85.3	2	82.5	
7		Fortatto, Melissa	88.6	88	89.1	88.6	91.6		89.5	
8		Gold, Larry	89.7	82.6	85.8	91.2	91		89.1	
9		Gonzalez, Ted	89.3	72.1	80.8	73.4	86.3	1	80.2	
10		Harrison, Sally	85.8	74.2	86	85.9	85.6	3	84.3	
11		Jones, Dave	94.3	85.1	83.4	72	84.8	1	82.1	
12		Lexington, Pat	85.1	78.8	89.5	83.4	96.3		87.7	
13		Mason, Nathan	87.3	72.9	82.1	74.5	80.7	1	78.9	
14		McDonald, Tom	91	80.4	86.8	86.1	91.4		87.7	
15		Nuns, Lisa	94.6	75.6	86.7	70.7	87.3		81.7	
16		Olesman, Daniel	87.6	73.8	89.1	88.6	82.6		84.8	
17		Patterson, Henry	90.4	74.1	80.3	73.9	80.5	1	79.0	
18		Radcliff, Beth	93.2	72.9	87.2	71.8	86.9		81.4	
19		Richards, Amy	91.3	80.9	92.9	89.7	94.1		90.5	
20		Smith, Tracy	93.5	87	86.5	86.9	95.7		90.4	
21		Student Name	Assignment 1 Grade	Assignment 2 Grade	Assignment 3 Grade	Exam 1 Score	Exam 2 Score	Attendance (in days absent)	Class Average	
22		Criteria:						1	> 85	
23										
24										
28										
29					Number of assignments	5				
30					Number of students	17				
31					Number of students with perfect attendance	9				
32										
33					Number of students with A	2				
34					Number of students with B+	6	4			
35					Number of students with B	15	9			
36										

Figure 10.45 Using The COUNTIF function with the student data.

Similar to the COUNTIF function is the **SUMIF** function. The SUMIF function will calculate the values in a given range that meet a specified criterion. This function is also similar to the DSUM function that we discussed earlier in this section. The format of the SUMIF function is:

=SUMIF(range, criteria, sum_range)

Even though we use the SUMIF function in the same manner as the DSUM function, it only allows for a single criterion. Therefore, when using Excel as a database, we recommend using the DSUM function instead of the SUMIF function to handle multiple criteria.

We would also like to note that a *Conditional Sum Wizard* Add-In accomplishes the same task as the SUMIF function. After ensuring that it is checked as an Add-In, you can find the *Conditional Sum Wizard* in the *Tools* menu. The Wizard has four steps in which you need to specify the data range, the criteria, how the formula is to be copied, and where the formula will be copied to.

10.5.4 Data Validation

Most database applications have a data validation tool. It enforces a certain format or type of data to be entered by the user for particular input. Excel offers this tool when working with large data. To use data validation, select the cell(s) you want to validate and then click on *Data* > *Validation* from the menu. The window shown below will appear (see Figure 10.46).

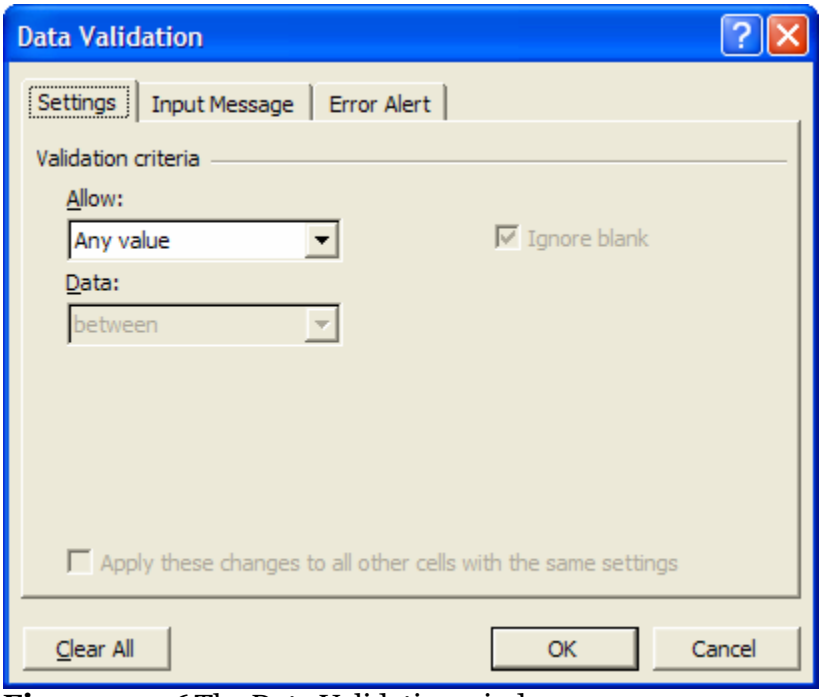


Figure 10.46 The Data Validation window.

There are three main tabs in this window: Settings, Input Message, and Error Alert. In the Settings tab, you specify the criteria for the data that will be entered into the selected cell. As shown in Figure 10.47, there is a list of criteria including Whole Numbers, Dates, Text Lengths, and others.

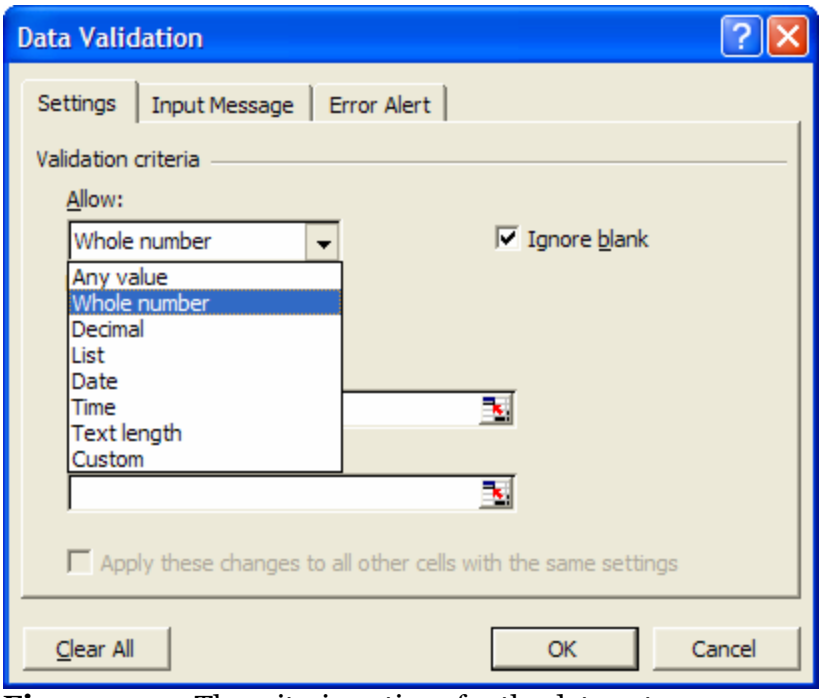


Figure 10.47 The criteria options for the data entry.

Depending on which criteria are selected from the list, you will have a few more options to further specify these criteria. For example, if we choose Whole Number from the list, we then choose from a list of inequalities (greater than, less than or equal to, between, etc) and provide some numerical bounds (see Figure 10.48).

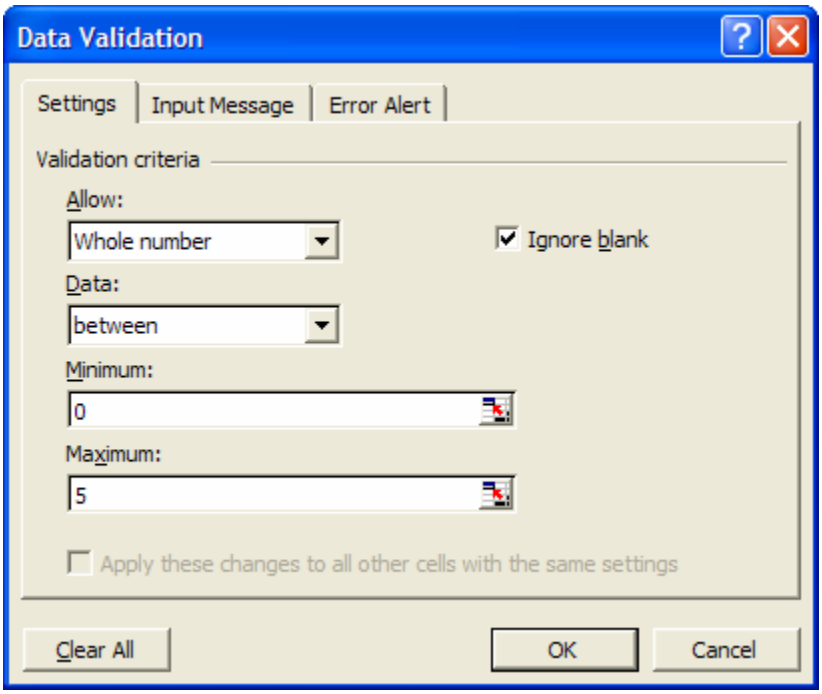


Figure 10.48 Specifying inequality and bounds for the Whole Number criterion.

The Input Message tab allows you to create a comment that will appear next to the cell after it is selected. This message is intended to guide the user to enter the correct data. For example, in Figure 10.49, we have created the message "Please enter a whole number between 0 and 5 only." This message matches the criterion we specified.

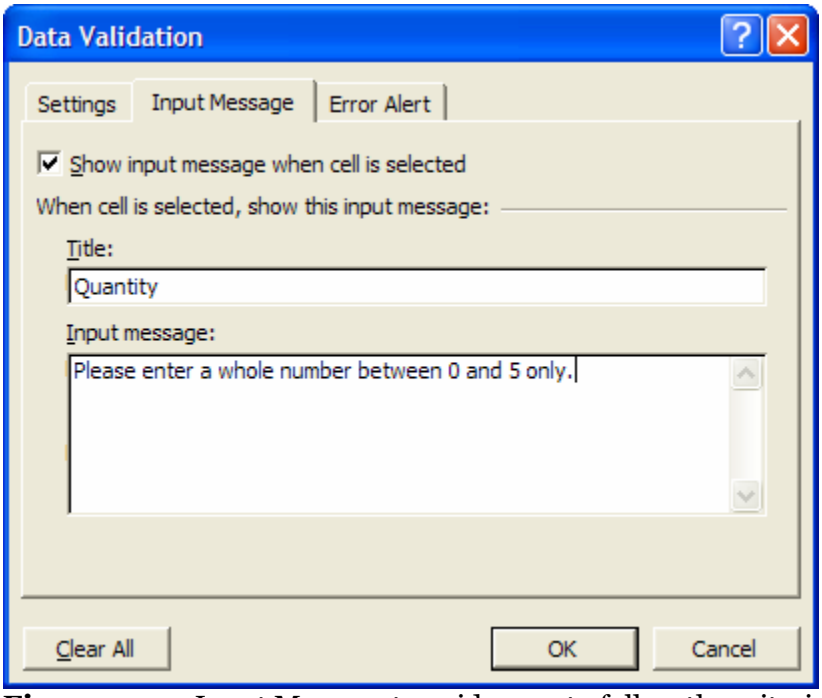


Figure 10.49 Input Message to guide user to follow the criterion.

The Error Alert allows you to display a message to the user after he or she has entered any data incorrectly. You can choose to have an Error, Warning, or Information symbol on this message box also; each symbol has a different set of options for the user in the message box. In Figure 10.50, we have chosen the Warning symbol and have entered the message, "You must enter a whole number between 0 and 5!"

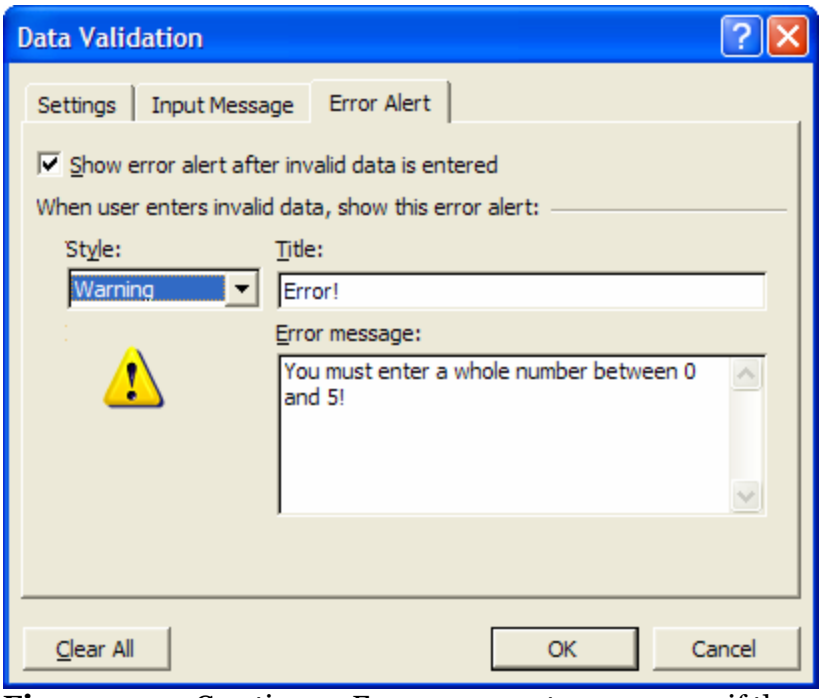


Figure 10.50 Creating an Error message to warn users if they enter any incorrect data.

Let's consider an example. Suppose we have a record of customer orders. For each order, the date, quantity, and price of the order are noted. If a maximum of 5 products can be sold in one order, we can apply the above Whole Number validation to the Quantity column. The Input Message for this validation appears when a cell in the Quantity column is selected (see Figure 10.51). If we enter a number greater than 5 (or less than 0), then the Warning message created in the Error Alert tab will appear (see Figure 10.52).

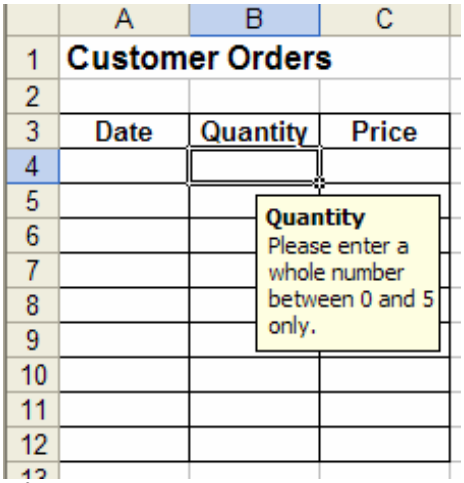


Figure 10.51 When selecting the validated cell, the Input Message appears.

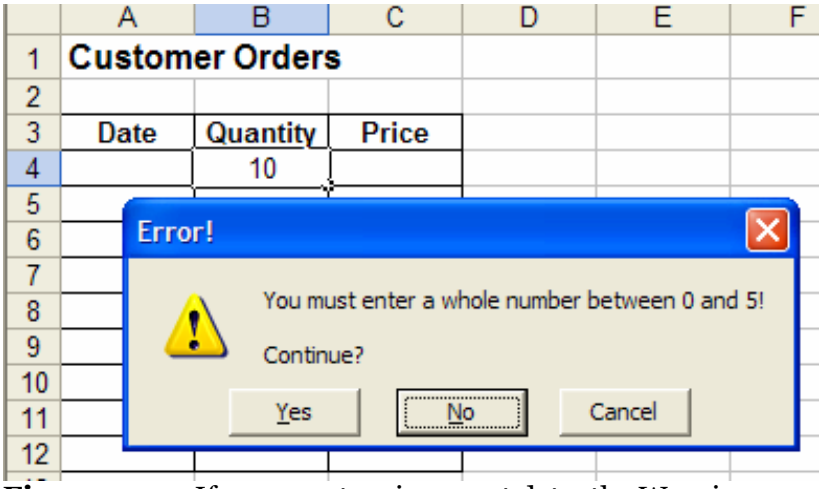


Figure 10.52 If a user enters incorrect data, the Warning message appears.

We could also validate the Date column in this example by choosing the Date criterion. We may, for example, specify that this record is for dates before February 1st only (see Figure 10.53).

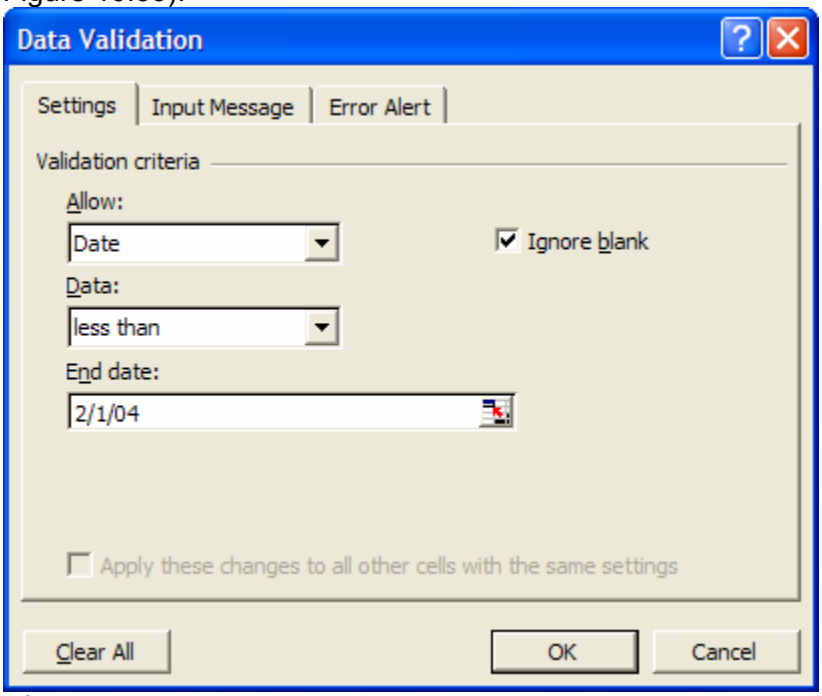


Figure 10.53 Specifying the Date criterion.

This time, let's choose the Information alert type. The Input Message and alert are shown in Figure 10.54 for a selected cell in the Date column.

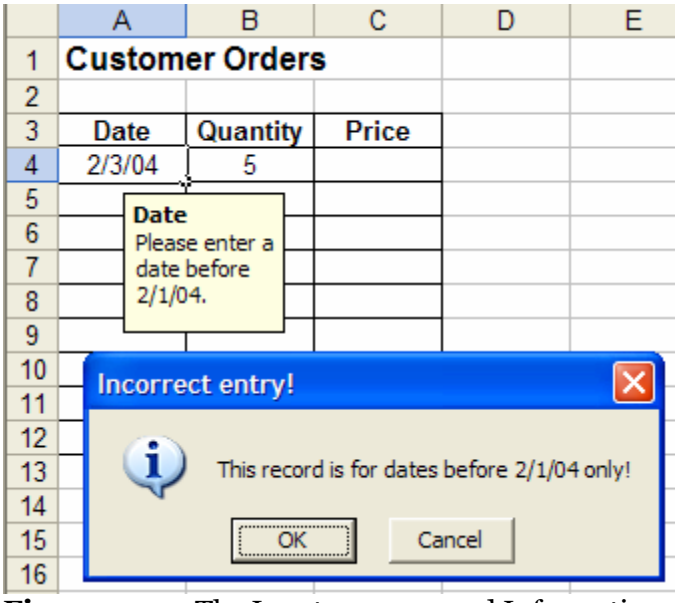


Figure 10.54 The Input message and Information alert.

Another interesting criterion, Custom, is similar to the Formula Is option used in Conditional Formatting (see Chapter 4). With the Custom criterion, the user applies a formula to the first cell in the selected range that is being validated. For example, in Figure 10.55, we are checking if the value entered is a number or not by using the Informational function ISNUMBER. If the result of this formula is false, then the Error Alert message will appear.

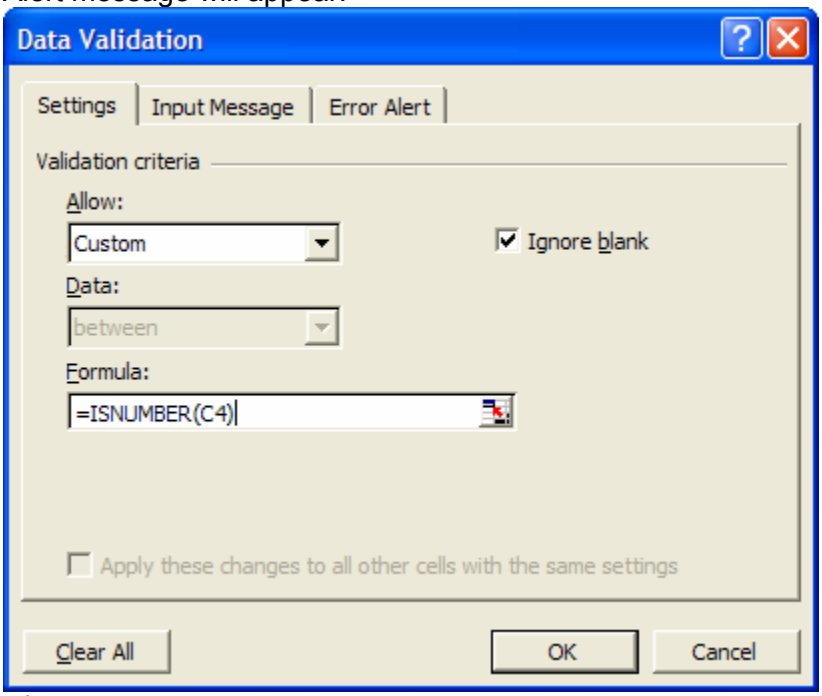


Figure 10.55 The Custom criterion uses a formula to check for errors.

One last unique criterion is List. It allows you to create a list box, or drop-down box, of options for users to choose from as their entry value. We must create a source for this list somewhere in the spreadsheet. For example, if we have an additional column in the

Customer Order table for the Salesperson who took the order, we may want to limit this entry to the list of valid salespersons working for the company. In the spreadsheet below, there is a list of salesperson names. We can then select List from the criteria options and highlight this range of names as the source (see Figure 10.56).

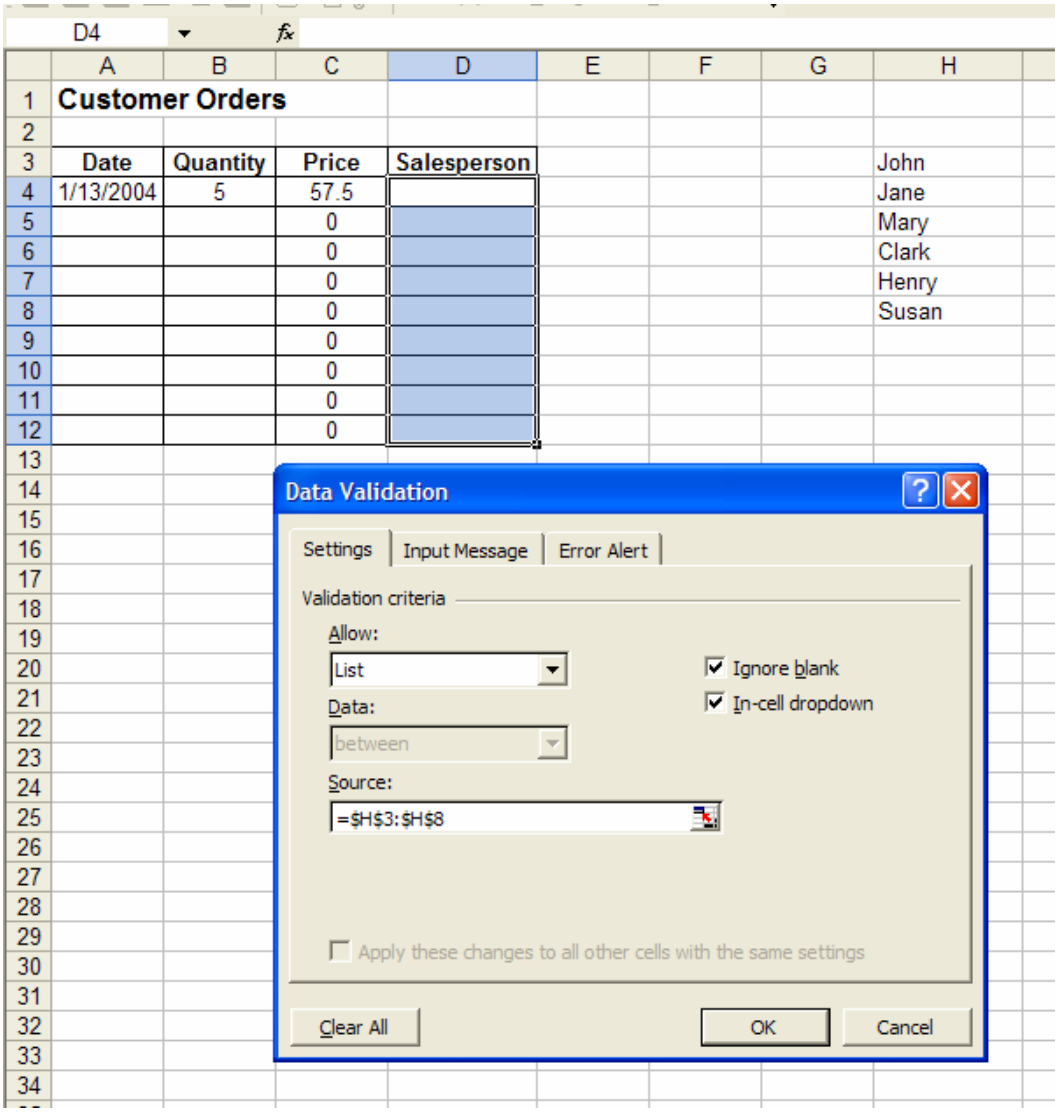


Figure 10.56 The list of names becomes the source for the List criterion.

Then, we can set the Input Message and Error Alert so that the users know they can only enter a salesperson name from the provided list. When they select a cell from the Salesperson column, a drop-down arrow will appear providing them with this list of valid names (see Figure 10.57).

	A	B	C	D	E	F
1	**Customer Orders**					
2						
3	**Date**	**Quantity**	**Price**	**Salesperson**		
4	1/13/2004	5	57.5			
5			0			
6			0			
7			0			
8			0			
9			0			
10			0			
11			0			
12			0			
13						

John
Jane
Mary
Clark
Henry
Susan

Salesperson
Please choose a name from the list.

Figure 10.57 The List criterion provides a list of valid entry values.

10.5.5 Data Consolidation

Another database tool available in Excel, data consolidation, allows you to compare and combine multiple sources of data into a new spreadsheet. Consolidation can also be accomplished with Pivot Tables' "multiple consolidation ranges" option; however, the Data Consolidation tool can sometimes be easier to use. Let's consider an example.

Suppose that we have monthly sales recorded for various products in two different marketing regions (see Figure 10.58). If we want to know the total sales per month (among all regions) for each product, we would want to use data consolidation.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	**Region 1 Sales**												
2													
3	**Product**	**January**	**February**	**March**	**April**	**May**	**June**	**July**	**August**	**September**	**October**	**November**	**December**
4	A	1060	1187	722	1258	1760	377	470	1201	1584	1685	376	1926
5	B	1201	1898	925	1885	1056	1802	984	757	1262	226	444	571
6	C	500	1014	1889	782	688	210	868	1124	1660	501	1163	1297
7	D	1840	1299	820	643	864	888	263	1045	1032	1129	1088	1350
8	E	1009	216	866	593	549	689	1418	1693	1294	622	1032	374
9	F	889	434	413	549	1773	438	647	1673	1083	960	1583	1239
10	G	1452	648	627	755	517	1841	1472	1859	370	1675	537	832
11	H	1989	1602	302	1627	1817	1399	1592	1712	1834	1892	740	477
12	I	450	807	201	1445	1427	1632	526	1075	522	966	746	1183
13	J	1337	1663	802	1186	1272	1797	283	695	1042	1635	859	1452
14	K	900	989	1490	1962	1553	1699	1022	1168	1991	1520	573	1350
15	L	1223	1976	1973	1927	1718	411	1187	279	1828	620	1783	1046
16	M	1751	1655	1359	1893	1766	678	1279	1806	890	1762	403	1899
17	N	1845	1697	1585	445	1292	883	1706	1877	698	1099	1842	538
18	O	753	1020	932	570	1648	1489	1953	562	1863	1255	1745	856
19	P	587	1473	1024	1726	1445	919	218	939	497	1633	891	1441
20													

(a)

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	**Region 2 Sales**												
2													
3	**Product**	**January**	**February**	**March**	**April**	**May**	**June**	**July**	**August**	**September**	**October**	**November**	**December**
4	**A**	266	1281	269	1223	544	1559	1238	585	797	708	837	1196
5	**B**	796	1909	1121	897	1906	1891	1705	1635	240	934	1283	480
6	**C**	1743	670	293	738	1858	1400	730	1360	1514	1821	1931	538
7	**D**	1045	387	1402	567	1331	803	545	1714	375	1703	1160	1010
8	**E**	1548	513	1564	1652	1007	579	656	1905	264	1161	1876	1376
9	**F**	207	444	500	1322	303	1411	485	1700	1422	1690	1916	1358
10	**G**	1099	554	1647	365	1199	1988	306	1815	1329	1397	425	247
11	**H**	1592	1198	617	1334	1750	1563	1041	480	537	1890	422	383
12	**I**	562	796	494	452	998	1053	1608	1844	493	1673	1368	1255
13	**J**	821	1802	815	540	286	1373	1948	984	352	339	398	1029
14	**K**	299	723	1038	882	1051	838	1478	1281	1252	487	1231	1017
15	**L**	1565	436	297	968	566	574	1470	1106	1206	893	1093	1357
16	**M**	946	1004	1100	1728	818	1425	1506	1016	1734	1681	1841	594
17	**N**	366	1270	1902	526	1395	1660	1759	1045	1846	1262	473	1495
18	**O**	1599	1299	490	1830	1181	690	1539	508	1101	1194	1375	1850
19	**P**	1550	320	1995	396	1674	1223	1219	655	863	729	1278	1717

(b)

Figure 10.58 A record of monthly sales per product for two different regions

We begin by creating a new spreadsheet (right-click on a sheet name and choose *Insert* > *Worksheet* from the list of options). In this new sheet, we select *Data* > *Consolidate* from the menu. The window shown below will appear (Figure 10.59).

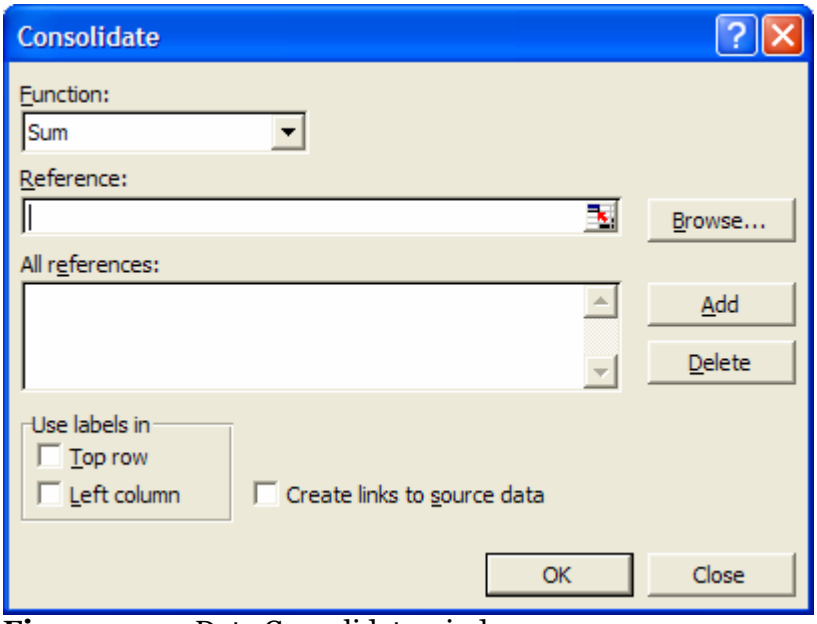


Figure 10.59 Data Consolidate window.

The data consolidation tool provides a list of functions that can be applied to the data as it is consolidated (see Figure 10.60). These functions include Sum, Count, Max, Min, etc. For this example, let's use the Sum function since we want to find the total number of sales in both regions.

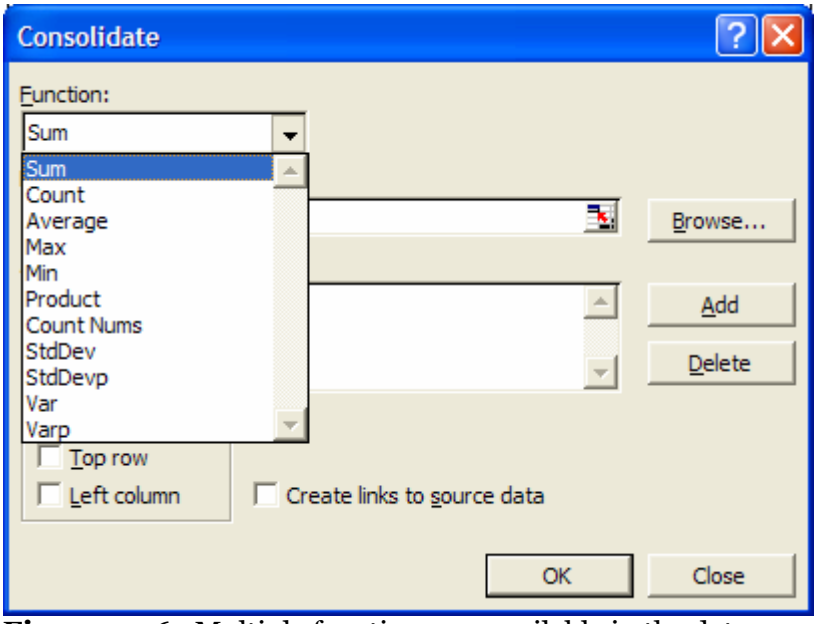


Figure 10.60 Multiple functions are available in the data consolidation tool.

After selecting the function, we must choose the references of the data that we want to consolidate. In this case, we select the entire table from both the Region 1 Sales and Region 2 sales spreadsheets. Next, we select the Top row and Left column check boxes for the location of data labels (since we have selected these labels in the reference as well). We also make sure to check the box "Create Links to source data." This option ensures that any change made to the source data (that is, the original tables) will be reflected automatically in the consolidated table. The completed data consolidation window appears in Figure 10.61.

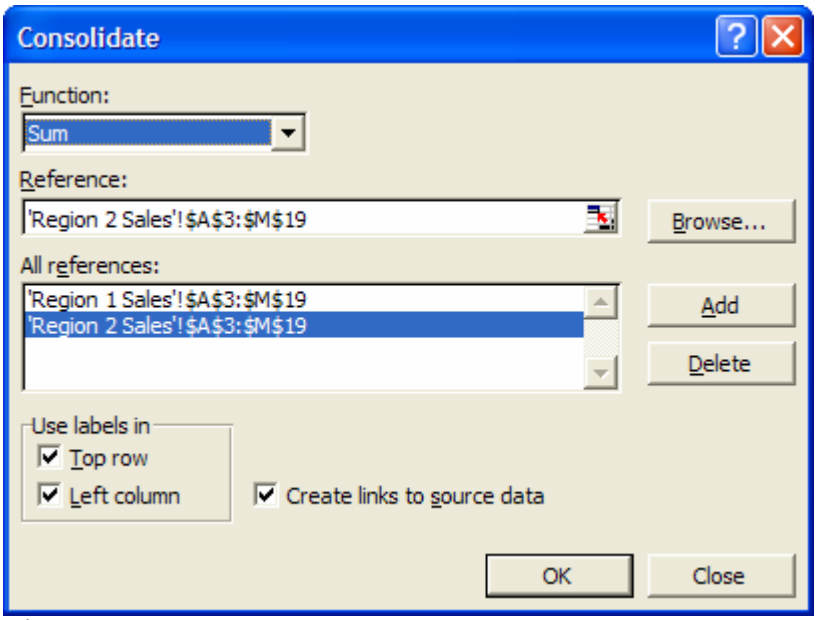


Figure 10.61 The completed data consolidation window.

Figure 10.62 displays the resulting consolidated table. Excel has calculated the sum of the sales of each product per month. Notice that some small + icons appear on the left side of the table. We can click these to expand the hidden rows that show the sources (or the references) of the consolidated data.

		A	B	C	D	E	F	G	H	I	J	K	L	M	N
	1	**Total Sales**													
	2														
	3			**January**	**February**	**March**	**April**	**May**	**June**	**July**	**August**	**September**	**October**	**November**	**December**
+	6	A		1326	2468	991	2481	2304	1936	1708	1786	2381	2393	1213	3122
+	9	B		1997	3807	2046	2782	2962	3693	2689	2392	1502	1160	1727	1051
+	12	C		2243	1684	2182	1520	2546	1610	1598	2484	3174	2322	3094	1835
+	15	D		2885	1686	2222	1210	2195	1691	808	2759	1407	2832	2248	2360
+	18	E		2557	729	2430	2245	1556	1268	2074	3598	1558	1783	2908	1750
+	21	F		1096	878	913	1871	2076	1849	1132	3373	2505	2650	3499	2597
+	24	G		2551	1202	2274	1120	1716	3829	1778	3674	1699	3072	962	1079
+	27	H		3581	2800	919	2961	3567	2962	2633	2192	2371	3782	1162	860
+	30	I		1012	1603	695	1897	2425	2685	2134	2919	1015	2639	2114	2438
+	33	J		2158	3465	1617	1726	1558	3170	2231	1679	1394	1974	1257	2481
+	36	K		1199	1712	2528	2844	2604	2537	2500	2449	3243	2007	1804	2367
+	39	L		2788	2412	2270	2895	2284	985	2657	1385	3034	1513	2876	2403
+	42	M		2697	2659	2459	3621	2584	2103	2785	2822	2624	3443	2244	2493
+	45	N		2211	2967	3487	971	2687	2543	3465	2922	2544	2361	2315	2033
+	48	O		2352	2319	1422	2400	2829	2179	3492	1070	2964	2449	3120	2706
+	51	P		2137	1793	3019	2122	3119	2142	1437	1594	1360	2362	2169	3158
	52														

Figure 10.62 The consolidated data table.

10.6 Summary

- You can import text files, webpage information, and database fields into Excel using the *Import External Data* wizard.
- Queries can be made to data being imported from a database.
- Using the *Access Links Add-In,* data can be exported to a database.
- Pivot Tables can use external data, such as from databases, their source,.
- Sorting and Filtering are helpful tools for organizing large data.
- There are several functions, called *Dfunctions*, that can be used on large data in Excel. There are also some other functions, such as COUNT functions and SUMIF, that can be used with large data.
- *Data Validation* and *Data Consolidation* are database tools available in Excel.

10.7 Exercises

10.7.1 Review Questions

1. Define the terms *sorting* and *field* and discuss how they are related.
2. What are the two ways to *sort* a set of data?
3. How does *filtering* differ from *sorting*?
4. How does selecting Auto Filter affect the display of the spreadsheet?
5. What does the *Top 10* option from a filtered drop-down list allow you to do?
6. What does the *Custom* option from a filtered drop-down list allow you to do?
7. Give an example of an instance when *filtering* data in an Excel database may be necessary.
8. How do *Dfunctions* differ from ordinary functions?

9. List five *Dfunctions* available in Excel.
10. What must be included in the criteria parameter of a *Dfunction* expression?
11. How are *Dfunctions* useful?
12. From an Excel database, where can data be exported to?
13. How can you refer to an external database when using *Dfunctions*?
14. What is a query?
15. How can you run a query before data has been imported from an external database?

10.7.2 Hands-On Exercises

1. The management in a manufacturing company has decided to promote the most efficient of the assembly line workers. To qualify for this promotion a worker should have attended the assembly of at least 330 units of output per week, during the last five weeks; should have attended on average the assembly of at least 400 units of output per week; and finally should have attended the assembly of no more than 6 defective units of output per week in the last five weeks.

    a. Filter the content of the database presented below to determine which workers should earned the promotion.
    b. The most efficient of the workers will further be promoted to team leader. Filter the database of workers being promoted to identify the team leader. Hint: to identify the team leader, find the maximum average output during the five weeks period.

	Unit of Output					Nr of Defects				
Name	Week 1	Week 2	Week 3	Week 4	Week 5	Week 1	Week 2	Week 3	Week 4	Week 5
James	372	336	417	387	446	3	5	7	5	2
Amy	464	411	347	440	479	3	4	6	0	3
Maria	444	313	482	394	432	2	3	1	1	7
Betsy	452	459	339	423	331	7	8	2	8	7
John	409	375	466	488	354	4	0	0	1	4
Dan	481	472	315	328	456	2	8	6	5	5
Jason	475	375	366	404	320	2	4	7	4	7
Cristina	497	344	404	415	473	1	0	6	8	4
Hallagan	348	473	488	306	318	7	6	4	6	2
Emilee	365	447	425	300	358	2	1	6	6	2
George	388	358	341	489	427	3	6	1	3	5
Emma	500	347	387	369	337	6	4	3	2	4
Jack	348	468	473	369	380	8	0	8	8	0
Ashley	493	482	498	351	415	5	4	4	7	4
Colleen	376	389	440	478	420	7	8	6	5	4
Jen	363	425	364	437	345	7	8	6	6	6
Casey	469	446	344	343	408	7	4	3	7	5
Kristin	500	361	448	388	359	5	7	0	5	1
Ben	364	455	442	407	357	0	2	8	3	7
Zaret	368	340	356	345	339	3	4	6	5	3

2. Consider the database presented in hands-on exercise 10.1.
    a. Sort the information by employee name in an ascending order.
    b. Count the number of employees that attended the assembly of more that 400 units of output during week 1.

c. Count the number of employees that attended on average the assembly of at most 5 defective products per week.

3. A quality inspector at a manufacturing plant must periodically take a sample of the products being produced and perform measurements to ensure that the products' dimensions are within certain tolerances. The product passes the inspection if the measurement is within 0.25 cm of the 20.00 cm target value. The inspector obtains the following set of measurements (in centimeters) for one of the samples: {20.18, 19.87, 19.93, 19.99, 20.23, 20.01, 20.17, 19.88, 19.02, 19.96, 20.53}.
   a. Create an Excel database of the inspector's measurements. Using logical functions, add a column to the database that indicates whether each measurement passes or fails the inspection.
   b. Use Dfunctions to compute each of the following values for your database:
      - The number of products that pass the inspection
      - The standard deviation of all products undergoing inspection
      - The standard deviation of all products that pass the inspection
      - The variance of all products undergoing inspection
      - The variance of all products that pass the inspection

4. Another quality inspector in the manufacturing plant mentioned in the previous problem must measure the length, width, and weight of a variety of products. The measurements the inspector obtains are provided in the table below. To meet inspection requirements, each product must be between 1.5 ft and 2.0 ft long, between 1.0 ft and 1.5 ft wide, and weigh between 1.0 lbs and 2.0 lbs. Use Dfunctions to compute the following values for this database:
      - The number of products that pass the inspection
      - The standard deviation and variance of the product lengths
      - The standard deviation and variance of the product widths
      - The standard deviation and variance of the product weights

Sample Nr.	Length	Width	Weight
1	1.65	0.98	1.89
2	1.55	1.01	1.76
3	1.67	1.45	1.23
4	1.48	1.53	0.99
5	2.01	1.36	1.29
6	1.98	1.42	1.56
7	1.78	1.33	2.02
8	1.63	1.23	1.56
9	1.68	1.41	1.76
10	2.02	1.15	1.23
11	1.59	1.37	1.34

5. Consider the student record database presented in Figure 10.28.
   d. Use filtering to display the name, grade and attendance of all the students that never missed a class.
   e. Use filtering to display the name, grade and attendance of all the students that missed three classes.
   f. Calculate the class average for the students that never missed a class.
   g. Calculate the class average for the students that missed three classes. Is there a relation between the number of classes missed and class average.

6. Consider the student record database presented in Figure 10.28.
    1. Use filtering to display the name, grade and attendance of all the students that performed better than average in the 1-st Exam.
    2. Use filtering to display the name, grade and attendance of all the students that performed better than average in the 2-nd Exam. Is there a relation, in terms of grade, between the students that performed better than average in the 1-st and 2-nd exam?

7. A credit card company maintains a database to store information about its cardholders. Cardholders that meet certain requirements are eligible to receive the platinum-level membership that includes a higher credit limit and additional benefits. These requirements are: the cardholder should currently have a minimum credit line of $3,000; the cardholder should have made at most one late payment; the cardholder began using the card prior to year 2000. Use Dfunctions to determine how many of the cardholders listed are eligible for this membership.

Name	Enrollment Date	Credit Limit	Nr. of Late Payments
James	1/3/1995	$3,000	0
Amy	10/3/1996	$2,000	1
Maria	11/30/1996	$4,000	1
Betsy	1/5/1997	$4,000	0
John	9/10/1997	$1,000	3
Dan	10/12/1997	$10,000	4
Jason	1/23/1998	$7,000	2
Cristina	4/5/1998	$6,000	0
Hallagan	10/23/1998	$2,000	1
Emilee	1/3/1999	$1,000	0
George	12/8/2000	$3,000	0
Emma	11/7/2001	$1,000	1
Jack	11/8/2001	$4,000	1
Ashley	12/9/2001	$6,000	0
Colleen	1/10/2002	$7,000	0
Jen	4/19/2002	$9,000	0
Casey	5/5/2002	$500	2
Kristin	7/5/2003	$1,000	3
Ben	4/5/2004	$3,000	1
Zaret	5/5/2004	$5,000	3

8. Read hands-on exercise 10.7. Present the name, enrollment date, credit limit and the number of late payments for the cardholders that are eligible for the platinum-level membership.

9. Using the table below, find the following:

    a. The number of songs not sung by Spears.
    b. The number of songs sung between June 1, 2004 and July 4, 2006.
    c. The number of songs sung by singers with an “e” in their name.

Song Numb.	Singer	Date	Minutes
1	Eminem	5/21/2004	4
2	Eminem	4/15/2004	2
3	Cher	1/28/2005	2
4	Eminem	1/28/2005	4
5	Moore	11/5/2004	2
6	Cher	9/18/2004	4
7	Spears	4/15/2004	3
8	Spears	3/17/2005	3
9	Manilow	1/16/2005	4
10	Eminem	4/10/2005	4
11	Madonna	2/15/2004	3
12	Eminem	1/10/2004	4
13	Springsteen	4/10/2005	2
14	Spears	4/15/2004	4
15	Moore	7/8/2004	3
16	Madonna	6/26/2004	4
17	Spears	5/28/2005	3
18	Mellencamp	7/27/2005	5
19	Spears	9/18/2004	5
20	Madonna	7/8/2004	4

10. The table below presents sales information of a company that sells beauty products in different regions of US. Answer the following questions:

    a. What are the total sales in the Midwest?
    b. What is the total amount of money Ashley made?
    c. What is the number of transactions with a greater than average sales amount?

Name	Date	Product	Units	Dollars	Location
Betsy	4/1/2004	lip gloss	45	$ 137.20	South
Hallagan	3/10/2004	foundation	50	$ 152.01	Midwest
Ashley	2/25/2005	lipstick	9	$ 28.72	Midwest
Hallagan	5/22/2006	lip gloss	55	$ 167.08	West
Zaret	6/17/2004	lip gloss	43	$ 130.60	midwest
Colleen	11/27/2005	eye liner	58	$ 175.99	midwest
Cristina	3/21/2004	eye liner	8	$ 25.80	midwest
Colleen	12/17/2006	lip gloss	72	$ 217.84	midwest
Ashley	7/5/2006	eye liner	75	$ 226.64	south
Betsy	8/7/2006	lip gloss	24	$ 73.50	East
Ashley	11/29/2004	mascara	43	$ 130.84	East
Ashley	11/18/2004	lip gloss	23	$ 71.03	West
Emilee	8/31/2005	lip gloss	49	$ 149.59	West
Hallagan	1/1/2005	eye liner	18	$ 56.47	south
Zaret	9/20/2006	foundation	-8	$ (21.99)	East
Emilee	4/12/2004	mascara	45	$ 137.39	East

Colleen	4/30/2006	mascara	66	$ 199.65	south
Jen	8/31/2005	lip gloss	88	$ 265.19	midwest
Jen	10/27/2004	eye liner	78	$ 236.15	south
Zaret	11/27/2005	lip gloss	57	$ 173.12	midwest

11. The Text option in Data Validation enables you to generate an error message when the number of characters in a cell is not of a desired length. Use the Text option to ensure that each cell in the range A1:A10 will contain at most 5 characters (including blanks).

12. You are entering employee names in the cell range B1:B10. Use Data Validation to ensure that no employee's name is entered more than twice in this column. (Hint: use the COUNTIF function.)

13. Suppose you have asked an assistant to enter values in a debt database. There is a column each for names, phone numbers, and price owed. Use Data Validation to ensure that only text is entered as a name, phone numbers have 10 digits, and price owed is never negative.

14. The URL http://www.baseball-reference.com/b/bondsba01.shtml contains Barry Bonds' major league baseball statistics. Import this data into Excel and perform some analysis of these values.

15. Find a website on foreign exchange rates. Import this data to Excel and save the query. Then run this query again to find the updated exchange rates.

16. Use the following information to create a text file. Import the file to Excel using the appropriate delimiter.

11,039.28	1,536.38	1,920.67
7,760.48	3,859.04	12,844.16
10,659.48	4,916.39	2,041.66
14,345.66	11,897.50	13,500.15

17. Place the following data in a text file and import it into Excel. Do not import the address or zip code columns.

Name	Address	Zip	Phone
John Smith	123 L.Lane	34285	342-6754
Susan Henry	453 J. Street	34245	546-2435
Tom Golfer	298 34th Ave	23445	345-6547
Joe Samual	290 B. Blvd.	34245	637-4235
Tina Tosser	582 23th St.	32456	654-6754
Mac Night	349 Main St.	34256	453-6577

18. Using the information recorded in the table below for the project teams of a class, perform the following actions:

Project Teams	Topics	Team Members				Presentation	Report
1	Capital Budgeting Problem	John	Amy	Ruth	Steve	77	83
2	Reliability Problem	Tom	Joan	Rob	Randy	81	76
3	Quality Control	Sam	Tim	Vic	Sally	80	85
4	Simulation	Susan	Melissa	Mac	Eric	97	93
5	Production Problem	Mary	Jorge	Maggy	Paul	95	100

a. Sort the list by presentation grades in ascending order
b. Sort the list by report grades in descending order
c. Sort the team member names alphabetically

19. Consider the database in hands-on exercise 10.10. Answer the following questions:

d. Which transactions made the top 5 percent of sales (in dollars)?
e. For which transactions the amount sold during 2004 was more than 50 units?
f. Present the lip gloss transactions during the first 6 months of 2004.
g. Identify the eye liner transactions that brought more than the average (dollars) made on eye liners.

20. The database given below presents the sales transactions for different products in two different regions. Perform the following data consolidations:

a. Identify the Max sales (over both regions) for each product in each month.
b. Count the number of sales made (over both regions) for each product during the three months period.

Region 1 Sales					Region 2 Sales				
Day	Product	January	February	March	Day	Product	January	February	March
1	A	205	263	20	1	A	173	1	256
1	B	164	0	146	2	A	208	201	224
1	C	278	177	179	2	B	176	33	350
2	D	156	214	240	3	B	190	249	215
3	D	72	134	48	3	D	162	74	156
4	D	7	256	104	4	D	90	150	170
5	A	141	87	148	7	D	112	284	141
6	A	2	0	135	7	G	154	217	113
7	A	32	47	72	9	G	152	200	275
7	B	7	0	2	10	G	277	183	372
7	E	25	120	171	10	H	131	71	266
8	E	197	90	124	12	F	294	211	249
9	E	221	121	48	13	F	146	125	5
9	A	84	103	134	13	A	115	214	141
15	G	12	250	51	15	F	157	241	73
16	D	23	159	70	15	A	125	227	135
17	E	334	94	28	17	A	314	189	180
18	A	352	290	0	18	C	189	154	101
19	F	45	235	231	19	C	313	182	68
20	G	23	29	903	20	C	389	247	257
21	H	454	0	124	20	E	781	23	0
22	B	632	235	523	22	E	0	892	30
22	C	556	451	230	22	G	20	231	231
24	C	342	123	213	24	H	291	190	56
25	A	688	341	242	25	F	52	200	0
25	D	455	901	267	26	E	193	120	231
25	F	673	129	298	26	D	101	901	0
28	E	757	130	274	27	D	201	0	130
29	H	23	189	136	29	A	304	0	11
30	A	346		0	30	B	109		103
31	B	872		201	30	C	191		31

21. Consider the data presented in hands-on exercise 10.20. Use *Dfunctions* to find the following:

    h. The average amount sold during the first half of March in region 1.
    i. The average amount of product A sold during the second half of January in region 1.
    j. The minimum amount of product B sold during the month of February in region 1.
    k. Total amount sold in January 1st in region 1.

22. A professor saved the results from three different exams he gave in his class, in three tables. Use the Data Consolidation tools in Excel to create the following:
    a. A table that presents for each student the identification number, as well as the grades of the three exams.
    b. A table that presents the average exam grade for each student.
    c. A table that presents the maximum grade received in an exam for each student.
    d. A table that presents the minimum grade received in an exam for each student.

Student ID	Exam 1		Student ID	Exam 2		Student ID	Exam 3
1	90		1	99		1	89
2	89		2	92		2	95
3	78		3	80		3	89
4	99		4	91		4	92
5	82		5	78		5	89

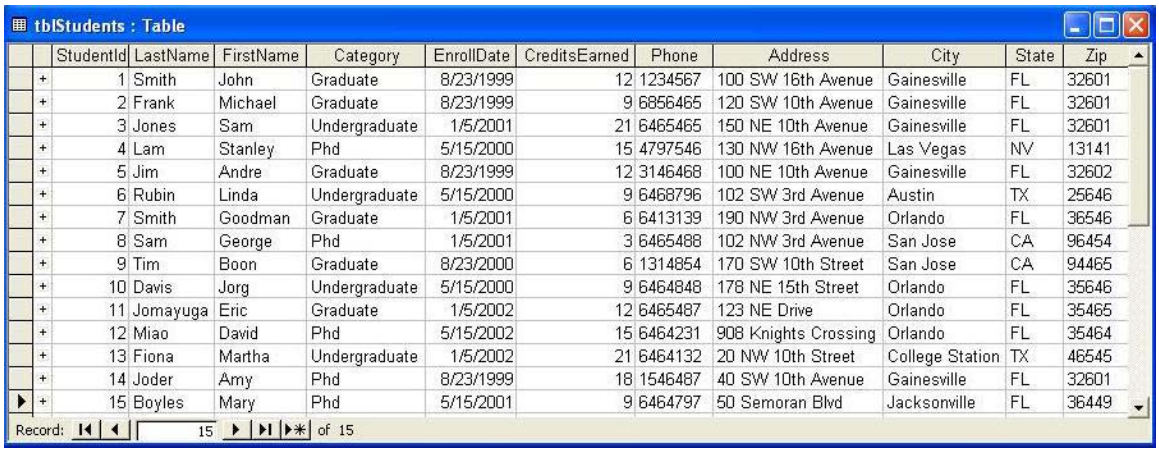
23. A retail store keeps the information about the products carried in the inventory in an excel spreadsheet. This information is presented in the table below. Use Data Validations tools to insure that:

    a. The identification number for each item is a number that has exactly five digits.
    b. Inventory level for each item cannot be negative.
    c. The unit price for each item is at least $150.
    d. A product condition should be either fair, or good, or very good.
    e. The purchase date for the items in the inventory should not be less that 1/1/2000.

ID	Inventory Level	Unit Price	Product Condition	Purchase Date
11111	190	$200	Good	1/1/2000
11112	200	$230	Fair	12/1/2000
11113	213	$250	Very Good	1/4/2002
11114	45	$190	Fair	12/2/2003
11115	120	$179	Good	2/3/2004
11116	135	$189	Good	4/12/2004

24. A University carries the information about the current students in table tblStudents in the School.dbm file in MS Access. The following figure presents the table tblStudents.

    a. Import table tblStudents in Excel.
    b. Display detailed information (Student Id, Last Name, First Name, etc) about the graduate students that have been enrolled no earlier than 1/1/2000.
    c. Display detailed information (Student Id, Last Name, First Name, etc) about the undergraduate students that have earned at least 6 credits.
    d. Display detailed information (Student Id, Last Name, First Name, etc) about the undergraduate students from Florida.

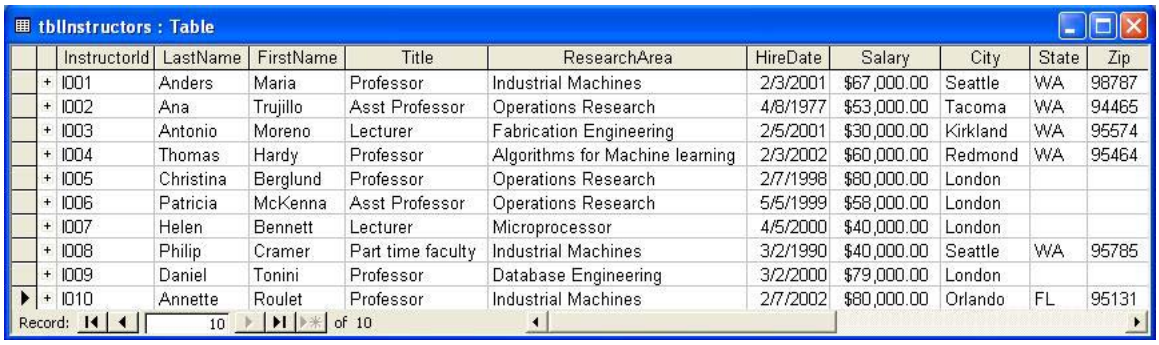


tblStudents : Table

StudentId	LastName	FirstName	Category	EnrollDate	CreditsEarned	Phone	Address	City	State	Zip
1	Smith	John	Graduate	8/23/1999	12	1234567	100 SW 16th Avenue	Gainesville	FL	32601
2	Frank	Michael	Graduate	8/23/1999	9	6856465	120 SW 10th Avenue	Gainesville	FL	32601
3	Jones	Sam	Undergraduate	1/5/2001	21	6465465	150 NE 10th Avenue	Gainesville	FL	32601
4	Lam	Stanley	Phd	5/15/2000	15	4797546	130 NW 16th Avenue	Las Vegas	NV	13141
5	Jim	Andre	Graduate	8/23/1999	12	3146468	100 NE 10th Avenue	Gainesville	FL	32602
6	Rubin	Linda	Undergraduate	5/15/2000	9	6468796	102 SW 3rd Avenue	Austin	TX	25646
7	Smith	Goodman	Graduate	1/5/2001	6	6413139	190 NW 3rd Avenue	Orlando	FL	36546
8	Sam	George	Phd	1/5/2001	3	6465488	102 NW 3rd Avenue	San Jose	CA	96454
9	Tim	Boon	Graduate	8/23/2000	6	1314854	170 SW 10th Street	San Jose	CA	94465
10	Davis	Jorg	Undergraduate	5/15/2000	9	6464848	178 NE 15th Street	Orlando	FL	35646
11	Jomayuga	Eric	Graduate	1/5/2002	12	6465487	123 NE Drive	Orlando	FL	35465
12	Miao	David	Phd	5/15/2002	15	6464231	908 Knights Crossing	Orlando	FL	35464
13	Fiona	Martha	Undergraduate	1/5/2002	21	6464132	20 NW 10th Street	College Station	TX	46545
14	Joder	Amy	Phd	8/23/1999	18	1546487	40 SW 10th Avenue	Gainesville	FL	32601
15	Boyles	Mary	Phd	5/15/2001	9	6464797	50 Semoran Blvd	Jacksonville	FL	36449

Record: 15 of 15

25. A University carries the information about the instructors in table tblInstructors in the School.dbm file in MS Access. The following figure presents the table tblInstructors.



tblInstructors : Table

InstructorId	LastName	FirstName	Title	ResearchArea	HireDate	Salary	City	State	Zip
I001	Anders	Maria	Professor	Industrial Machines	2/3/2001	$67,000.00	Seattle	WA	98787
I002	Ana	Trujillo	Asst Professor	Operations Research	4/8/1977	$53,000.00	Tacoma	WA	94465
I003	Antonio	Moreno	Lecturer	Fabrication Engineering	2/5/2001	$30,000.00	Kirkland	WA	95574
I004	Thomas	Hardy	Professor	Algorithms for Machine learning	2/3/2002	$60,000.00	Redmond	WA	95464
I005	Christina	Berglund	Professor	Operations Research	2/7/1998	$80,000.00	London		
I006	Patricia	McKenna	Asst Professor	Operations Research	5/5/1999	$58,000.00	London		
I007	Helen	Bennett	Lecturer	Microprocessor	4/5/2000	$40,000.00	London		
I008	Philip	Cramer	Part time faculty	Industrial Machines	3/2/1990	$40,000.00	Seattle	WA	95785
I009	Daniel	Tonini	Professor	Database Engineering	3/2/2000	$79,000.00	London		
I010	Annette	Roulet	Professor	Industrial Machines	2/7/2002	$80,000.00	Orlando	FL	95131

Record: 10 of 10

a. Import this table in Excel.
b. The minimum wage for instructors in this University is $40,000. Insure that no instructor is hired for less than $40,000.
c. Display detailed information (Instructor Id, Last Name, First Name, etc) about the instructors that have earned the title Professor.
d. The office of Human resources needs detailed information about the instructors that are not USA citizens. Display detailed information about those instructors.
e. Display detailed information about the instructors that make top 10% of the annual salary.