

Chapter 6

Encoding and Retrieving Data

Numbers are stored in computers using special codes. Text and dates are converted into numbers and then stored. When the data is retrieved from the computer, those codes are deciphered and turned into numbers, text, and dates that we understand. In this chapter, you learn how this is done and how those codes influence the way data is stored, retrieved, and interpreted.

Why Do I Need to Know This?

One of the dramatic changes that occurred in the last forty years is the ubiquitous use of computers. Almost every job and many of our leisure activities now involve computers. Computers only know how to work with numbers and they excel at quantitative tasks but they have limitations. Almost every job requires workers to 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, how you can get those numbers out of them, and how they are transformed back into a form that humans can understand.

1 Storing Numbers in Computers

Learning Objectives

1. Identify the reason for using computers with the 1890 census. [6.1.1]
2. Distinguish between numbers and numerals and give examples. [6.1.2]
3. Distinguish between data and information. [6.1.3]
4. Convert numbers between systems using the Calculator program [6.1.4]
5. Describe how the ASCII standard encodes numbers and characters. [6.1.5]
6. Describe the UTF code and its relationship to the ASCII code. [6.1.6]
7. Identify the method used to encode dates and time in personal computers. [6.1.7]
8. Identify the components of a data table. [6.1.8]

9. Describe the characteristic of a key field and how key fields are used to form relations between tables. [6.1.9]

The U.S. census of 1880 asked 26 questions, as shown in Figure 6.1. (Minnesota Population Center N.D.)

Schedule 1. Free Inhabitants in _____, in the County of _____, State of _____,
enumerated by me, on the _____ day of _____, 1880.
_____, Enumerator.
Enumeration district No. _____.

In cities: Name of street; house number

1. Dwelling houses numbered in order of visitation.
2. Families numbered in order of visitation.
3. The name of every person whose place of abode on the 1st day of June, 1880, was in this family.

Personal description:

4. Color - White, W; black, B; Mulatto, Mu; Chinese, C; Indian, I.
5. Sex - Males (M), females (F)
6. Age at last birthday prior to June 1, 1880. If under 1 year, give months in fractions, thus, 3/12.
7. If born within the census year, give the month.
8. Relationship of each person to the head of this family - whether wife, son, daughter, servant, boarder, or other.

Civil condition:

9. Single.
10. Married.
11. Widowed; divorced.
12. Married during census year.

Occupation:

13. Profession, occupation, or trade of each person, male or female.
14. Number of months this person has been unemployed during the census year.

Health:

15. Is the person (on the day of the enumerator's visit) sick or temporarily disabled, so as to be unable to attend to ordinary business or duties? Is so, what is the sickness or disability?
16. Blind.
17. Deaf and dumb.
18. Idiotic.
19. Insane.
20. Maimed, crippled, bedridden, or otherwise disabled.

Education:

21. Attended school within the census year.
22. Can not read.
23. Can not write.

Nativity:

24. Place of birth of this person, naming State or Territory of United States, or the country, if of foreign birth.
25. Place of birth of the father of this person, naming State or Territory of United States, or the country, if of foreign birth
26. Place of birth of the mother of this person, naming State or Territory of United States, or the country, if of foreign birth

Figure 6.1. Census questions of 1880.

In 1880, the population of the U.S. was about 50 million people and the census was recorded on paper forms. Simple facts are called **data**. The data contained in those millions of forms had to be counted and organized to turn the data into **information**—data that have been given meaning—so government officials

could use the information to make decisions. Tabulating the forms by hand took seven years. The delay was so long that the data were out of date by the time it became available and they projected that the census of 1890 would take more than ten years to tabulate.

Herman Hollerith provided a solution to the problem by inventing a system where the answers on each form were transferred to a card as holes. The holes on the card represented the answers to the census questions. The answers could be counted by passing stacks of cards through a counter like the one shown in Figure 6.2. (IBM N.D.)

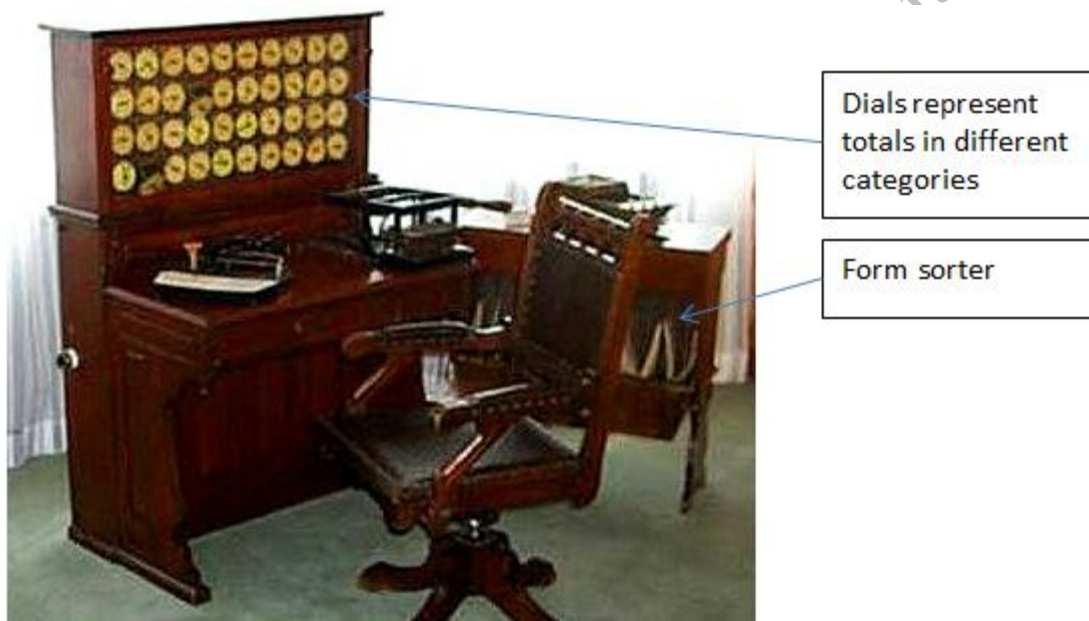


Figure 6.2. Hollerith tabulator and sorter.

The card and counting system reduced the tabulation time to only one year. Hollerith went on to found International Business Machines (IBM)—one of the largest computer companies in the world. Since that time, machines have been used to store and tabulate numbers using codes to represent the numbers and text.

Numbering Systems

A symbol or mark that represents a number is called a **numeral**. For example, measurement of natural gas uses Roman numerals. The first C in CCF is the Roman numeral C for 100. Similarly, the gas industry uses

the Roman numeral M for thousand so MCF means *thousand cubic feet* (not million). Most people count using a number system that has ten numerals from 0 to 9. The number of individual numerals indicates the **base**. The system that uses ten numerals to represent the numbers from zero to nine is the **base 10 (decimal)** system. For example, the numerals in the number 3,245 in base 10 mean $3 \cdot 1000 + 2 \cdot 100 + 4 \cdot 10 + 5 \cdot 1$ as shown in Figure 6.3.

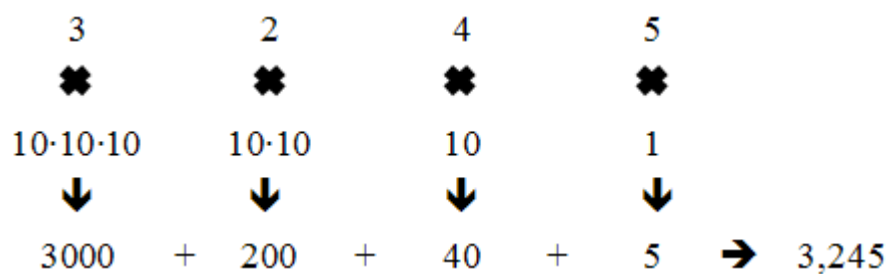


Figure 6.3. Base 10 numbering system.

In other circumstances, different numbering systems are more appropriate. Scientists have discovered that the instructions for making life are coded into DNA molecules where the smallest unit of data is represented by the pairings of four small molecules that are represented by the letters A, T, G, and C. There are only four ways these molecules combine. They are AT, TA, CG, and GC. This implies that biological life itself is digital and it uses base four! These combinations of molecules are used to specify which amino acid to use when building proteins. In the spring of 2010, a team led by Craig Venter used a computer to program the DNA for a new form of synthetic life. (Venter 2010)

Tip: Each DNA molecule has millions of pairs of AT, TA, CG and GC. Understanding, studying and modifying genetic materials requires storing enormous amounts of data in computers. A new field called bioinformatics has great potential for students who are interested in biology and data management.

Mechanical and electrical devices are much simpler than biological life. These devices are usually just switched on or off. A numbering system with only two numerals, 0 and 1, works well with mechanical and electrical switches. Because there are only two numerals this system is a **base 2** system, which is also called a **binary** numbering system. A single numeral in a binary number is called a **bit**.

Tip: Data transmission rates are measured in bits per second (bps). Your Internet connection's transmission speed may be as low as 36 kilobits per second if you use a dial-up telephone modem or between one and six megabits per second for high-speed home connections.

Computer processing chips are assembled from billions of tiny electrical switches so computers work well with binary numbers. For numbers larger than one, more numerals are used where their position indicates a multiple of the base. For example, the numerals in the number 1101 in base 2 mean $1 \cdot 8 + 1 \cdot 4 + 0 \cdot 2 + 1 \cdot 1$ as shown in Figure 6.4.

$$\begin{array}{ccccccc}
 1 & 1 & 0 & 1 & & & \\
 \times & \times & \times & \times & & & \\
 2 \cdot 2 \cdot 2 & 2 \cdot 2 & 2 & 1 & & & \\
 \downarrow & \downarrow & \downarrow & \downarrow & & & \\
 \text{In Base 10: } 8 & + & 4 & + & 0 & + & 1 \rightarrow 13
 \end{array}$$

Figure 6.4. Base 2 (binary) numbering system.

It takes a lot of numerals to represent a large number in binary. Computer programmers use a more compact system with sixteen numerals that can be easily converted to binary numbers. The first ten numerals are 0 through 9 and first six letters of the alphabet are used as numerals to represent the numbers ten through fifteen. This system is the **base 16** or **hexadecimal** system. For example, the numerals in the number 2A9D in base 16 mean $2 \cdot (16 \cdot 16 \cdot 16) + 10 \cdot (16 \cdot 16) + 9 \cdot 16 + 13 \cdot 1$ as shown in Figure 6.5

$$\begin{array}{ccccccc}
 2 & A & 9 & D & & & \\
 \times & \times & \times & \times & & & \\
 16 \cdot 16 \cdot 16 & 16 \cdot 16 & 16 & 1 & & & \\
 \downarrow & \downarrow & \downarrow & \downarrow & & & \\
 \text{In Base 10: } 8192 & + & 2560 & + & 144 & + & 13 \rightarrow 10,909
 \end{array}$$

Figure 6.5. Base 16 (hexadecimal) numbering system.

One numeral of a hexadecimal number can directly replace four numerals of a binary number because they both represent numbers from zero to fifteen.

Converting Between Decimal, Binary, and Hexadecimal Systems

The Windows operating system includes numerous small programs such as the Calculator which is found under Accessories, as shown in Figure 6.6.

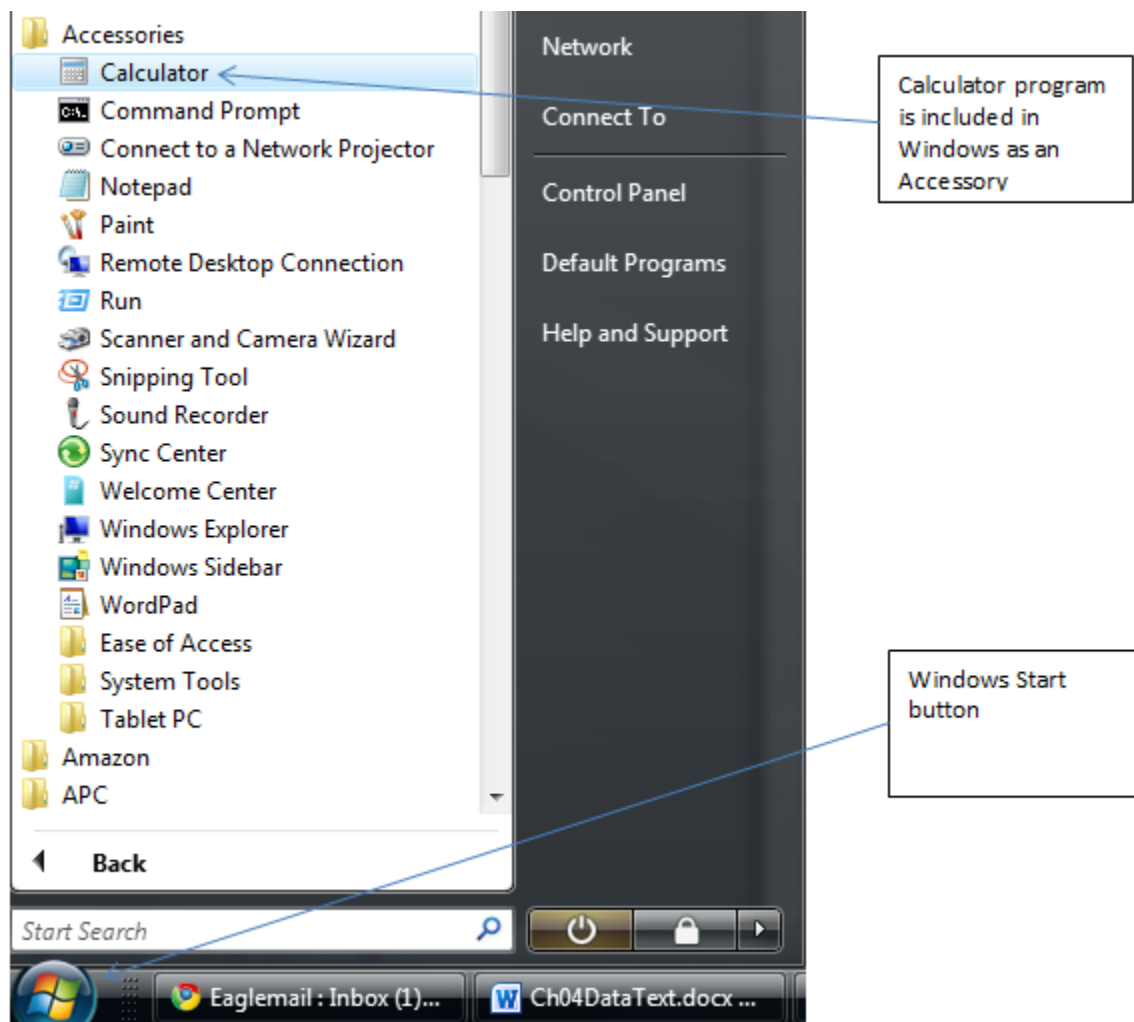


Figure 6.6. Calculator program in Windows.

In the Scientific view, the calculator program has the option of displaying numbers in Hexadecimal (Hex), Decimal (Dec), Octal (Oct-base 8), and Binary (Bin), as shown in Figure 6.7.

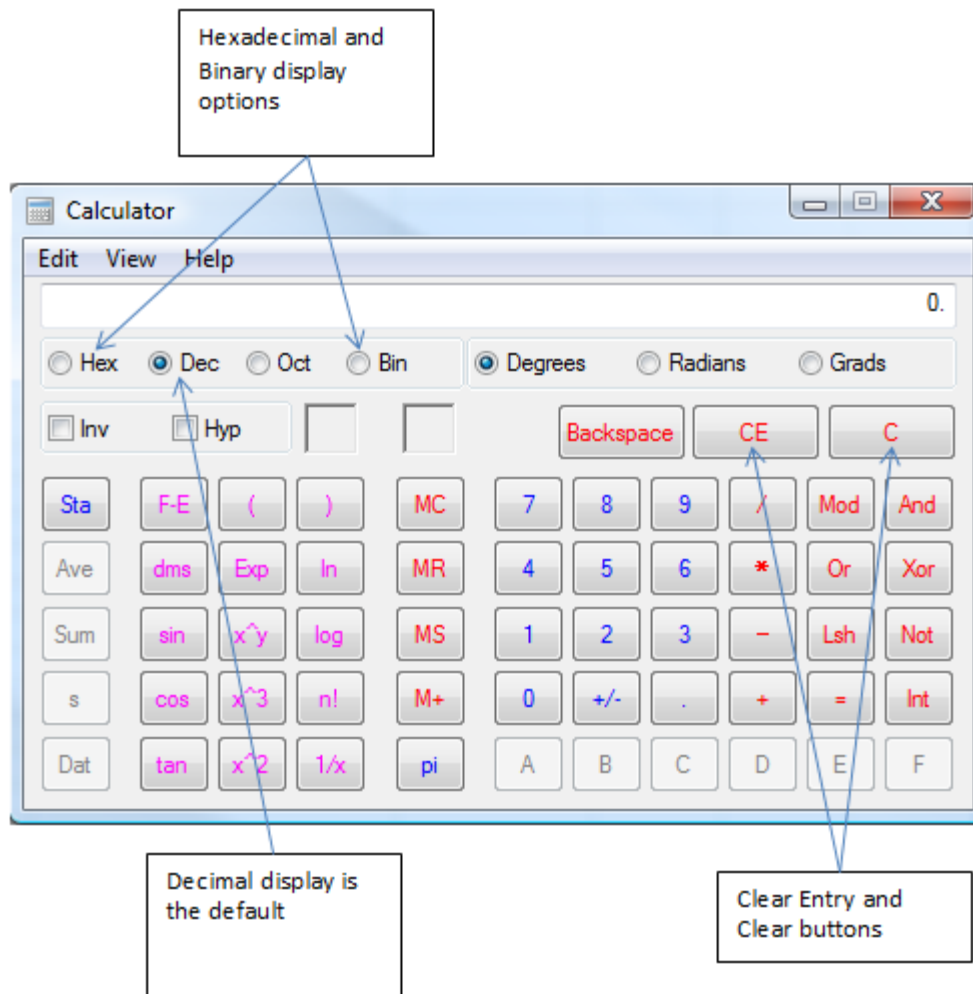


Figure 6.7. Numbering system options.

The calculator can be used to convert numbers between display systems, as shown in Figure 6.8.

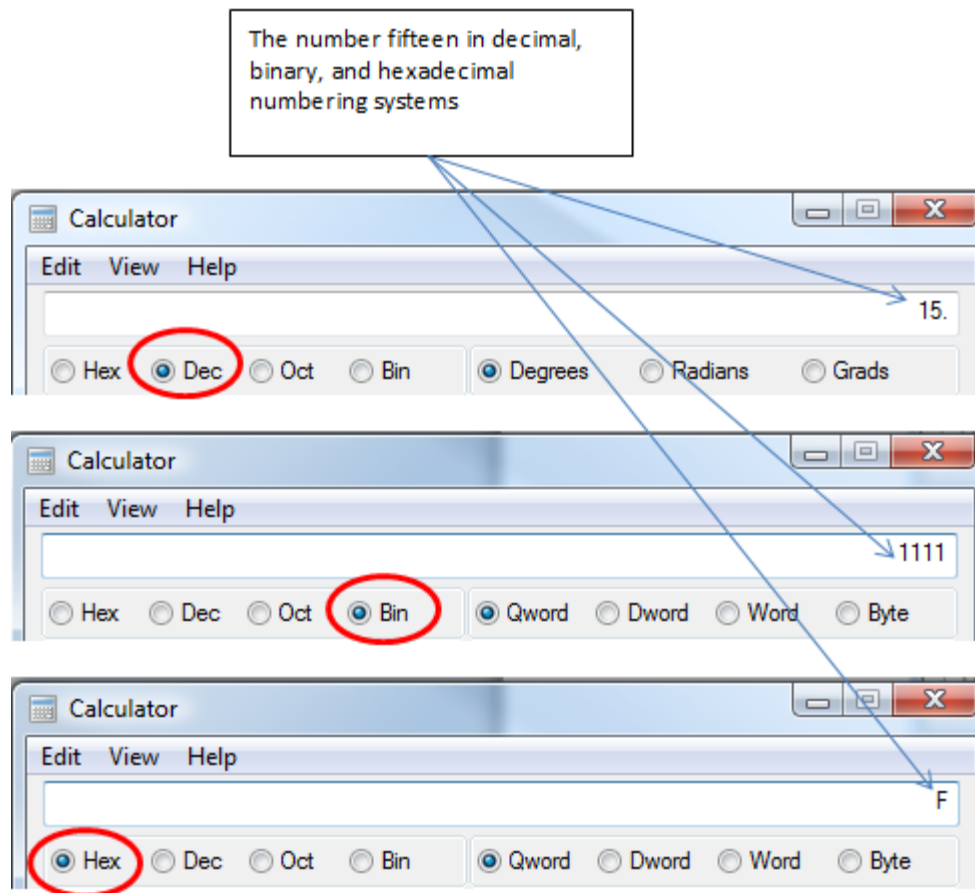


Figure 6.8. Same number displayed in three different numbering systems.

Codes for Letters and Characters

To store words or other characters in a computer, numbers are assigned to each character or letter. A standard was developed called the **American Standard Code for Information Interchange (ASCII)** (pronounced ask-key-2) where symbols, numerals, and characters were arranged in a table that had sixteen columns and each cell in the table could be represented by a pair of hexadecimal numbers that indicate the row and column. For example, a capital letter K is in row 4, column B so its code number is 4B, as shown in Figure 6.9.

ASCII Code Chart																
	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	HT	LF	VT	FF	CR	SO	SI
1	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS	RS	US
2		!	"	#	\$	%	&	'	()	*	+	,	-	.	/
3	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
6	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7	p	q	r	s	t	u	v	w	x	y	z	{		}	~	DEL

Figure 6.9. Letter K is in row 4, column B. Its code number in hexadecimal is 4B

There are several concepts built into this system that affect how we interact with data on computers.

They are:

- NULL (meaning empty), zero, a blank (space), and the capital letter O are four different codes
- The blank character has a lower number (20) than any of the other characters
- Capital and lower case letters have different codes
- A two-numeral hexadecimal number translates into an eight-numeral binary number. A group of eight binary numbers is called a byte. Each letter or character is identified by one byte of binary data.

Tip: Data file size is measured in multiples of bytes. For example, a picture from a digital camera might be 3 megabytes or a new piece of software might be 500 megabytes. Because data transmission rates are usually given in bits per second you must multiply the file size by 8 before you divide it by the transmission rate to find out how many seconds it will take to download a certain file.

Even though some characters or symbols appear to be similar when printed—such as a zero and the letter O—they are completely different to a computer. This is also true of the blank space—character 20—and the NULL (empty) —character 00. This is important to understand when working with database or spreadsheet files.

Sorting words is a matter of sorting the numbers that represent the letters in the words, beginning with the first letter in the word. The code number for a blank was chosen to be a lower value than the first letter in

the alphabet intentionally so that blanks would appear at the top of a list. Blanks between words make them easier to read and to identify as separate entities but they present problems to some computers when used in file names. Recall that Excel replaces blanks with underscore characters when it converts labels to cell names.

Microsoft Windows operating system is programmed to recognize blanks as acceptable characters between words in a file name while other operating systems might not. Blanks at the end of words do not cause a problem with sorting but can be a significant problem when trying to match the word. Many people habitually type a space after each word, but because the blank is not visible the difference is often hard for us to notice. Words with a space at the end of the word will be treated differently by computer programs than words that do not have a space following the word. Again this is most often an issue with database and spreadsheet files.

To a computer, the capital letter and lowercase letter are completely different which is why capitalization matters when entering a password and using a mixture of upper and lower case letters in a password makes it harder to break.

The early ASCII code was based on the English language and the table of codes only had eight rows. To accommodate other languages, the standard was expanded. The expanded version is the **Unicode Transformation Format (UTF)**. The first eight rows of the table of codes in **UTF-8** are the same as the original ASCII code, which makes it backwardly compatible with the older ASCII format. To store data in other languages such as Chinese and Japanese, codes with additional bytes of binary digits are used. For example, the UTF code for water in Chinese, 水, uses two bytes; 0110,1100,0011,0100 in binary or four numerals in hexadecimal, 6C34.

Tip: Long binary numbers are separated with commas every four digits. Each group of four binary numerals can be converted to one hexadecimal numeral. For example, 1100 is C in hexadecimal. Both represent the number twelve.

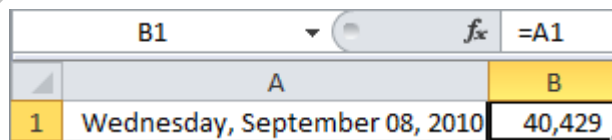
Dates and Times

Personal computers store dates as sequential numbers that begin at some reference date such as the turn of the century. The reference date is an arbitrary choice. One of the first popular spreadsheet programs for personal computers, Lotus 1-2-3, used January 1, 1900 as day 1.

Tip: The earth rotates on its axis (one day) about $365 \frac{1}{4}$ times for each revolution around the sun (one year). To correct for the fractional day, one day is added to the end of February every four years, called a leap year. The number of days per year is slightly less than 365.25 so this method is corrected by only including every fourth century year as a leap year. The year 1900 was not a leap year and February only had 28 days that year.

Unfortunately, the programmer who wrote the code assumed that 1900 was a leap year and wrote the code so that day 60 was February 29, 1900 when it should be March 1. To maintain backward compatibility with Lotus 1-2-3, the programmers of Microsoft Excel intentionally kept this error. The programmers for the Macintosh version of Excel, chose to use the first leap year of the century, 1904, as the beginning. Consequently, the sequential numbers used for the same date will be stored differently in Excel for Macintosh and Excel for Windows. The difference rarely matters unless dates are copied and pasted between spreadsheets that use the two different systems.

An example in Microsoft Excel for Windows demonstrates the use of numbers to store dates. A date is typed into cell A1. In B1, a formula refers to the content of cell A1 but B1 is formatted to display a number using a comma, as shown in Figure 6.10.



	B1	f_x	=A1
	A		B
1	Wednesday, September 08, 2010		40,429

Figure 6.10. Date is stored as the number of days since Jan 1, 1900.

This system makes it easy to find the number of days between two dates by simply subtracting the earlier date from the later date, as shown in Figure 6.11.

B3		f_x	=B2-B1
	A	B	
1	Begin	Wednesday, September 08, 2010	
2	End	Friday, December 17, 2010	
3	Days	100	

Formula to calculate the number of days between dates

Figure 6.11. Days between dates found by subtraction.

The time of day is a decimal fraction of a day. For example, 6 AM on Sept 27, 2010 is represented by 40,448.25. Because each day has 24 hours, 6 AM is one-quarter of a day or 0.25 as shown in Figure 6.12.

Custom Date and Time format

	A	B
1	Date	Number
2	9/27/10 6:00 AM	40,448.25

6 hours is .25 of a day

Figure 6.12. Days between dates found by subtraction.

Other computer systems store dates in different ways. In the mid-1900s, computer memory space was expensive so programmers used a system that reserved two digits for the month, day, and year. For example, September 8, 2010 would be stored as three, two-digit numbers; 09, 08, and 10. To find the number of days between two dates in this system required a program that could interpret the three numbers. The programmers who wrote the programs in the 1970s did not expect that their programs would be in use decades later when the century changed and the year 00 would come after the year 99. There was fear that computer programs that managed banking and credit card accounts would fail and many hours were spent checking and fixing old computer code. This was known as the Y2K crisis for Year 2000. Fortunately, the problems were mostly avoided and now those systems use four digits for the year, assuming that their code will not be in use when the year changes from 9999 to 10,000.

Data Tables

A common method of storing data in computers is to use a table structure where each row is about one item or person and each column is a category.

Data about people are usually stored using this method. For example, the employees of a consulting engineering firm that does installations of alternative energy systems are shown in Figure 6.13.

Fields

	A	B	C	D	E	F	G	H	I	J	K	L
1	Employee ID	Title	First Name	Middle Initial	Last Name	Street Address	City	State	Postal Code	Home Phone Number	Specialty	SSN
2	EID002145	Mrs. Pat	E.	Rowland	4589 Stadium	Ann Arbor	MI	48103	(734) 555-1231	Mechanical Engineer	247-80-7889	
3	EID002146	Mr. Archie	I.	Baylis	342 Michigan	Ypsilanti	MI	48197	(734) 555-1244	Civil Engineer	345-24-7635	
4	EID002147	Mr. John	W.	Baker	122 Sill	Ypsilanti	MI	48197	(734) 555-1246	Civil Engineer	380-55-5520	
5	EID002148	Mr. Wayne	M.	Robertson	1356 Washtenaw	Ypsilanti	MI	48197	(734) 555-1247	Civil Engineer	837-45-6274	
6	EID002149	Mr. Mike	I.	O'Sullivan	345 Michigan	Milan	MI	48169	(734) 555-1533	Draftsperson	690-98-9583	

Records

Figure 6.13. Table of employee data.

Observe that each row of the table is about one individual. Rows in a data table are called **records**. Each column is a category or type of data. Types of data in a data table are called **fields**. One of the fields typically contains a denominator number that is unique for each record. The field is called the **key field**.

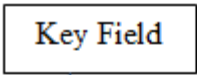
This company has a data table of its clients, as shown in Figure 6.14 where the client code is the key field.

The diagram shows a data table with 4 columns and 4 rows. A box labeled 'Key Field' has an arrow pointing to the 'Client Code' column (column A).

	A	B	C	D
1	Client Code	Client Name	Phone	Contact Person
2	GM-Assy	General Mechanics	313-555-1212	Ralph
3	MVRS	Multi-Variable Response Systems	484-555-2345	Juan
4	RPG	Rapid Prototyping Group	734-555-2211	John

Figure 6.14. Table of clients.

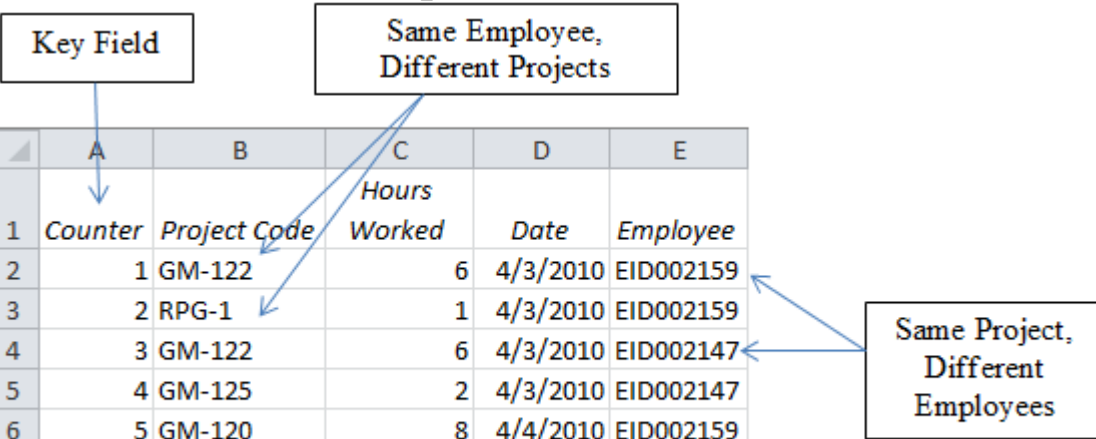
It has another table of projects where the project code is the key field, as shown in Figure 6.15.



	A	B	C
1	Project Code	Project Name	Client Code
2	GM-120	Office renovation	GM-Assy
3	GM-122	Warehouse	GM-Assy
4	GM-125	Paint factory	GM-Assy
5	MVRS-2	Warehouse	MVRS
6	RPG-1	Office annex	RPG

Figure 6.15. Table of projects.

Another table shows how much work each employee does on each project. It has fields for the employee identification number and the project code. Because an employee could work on several different projects and each project could have several employees working on it, neither the project code nor employee ID field is suitable for use as a key field because there are duplicate values in each field. Instead, a counter is used as the key field, as shown in Figure 6.16.



	A	B	C	D	E
	Counter	Project Code	Hours Worked	Date	Employee
1	1	GM-122	6	4/3/2010	EID002159
2	2	RPG-1	1	4/3/2010	EID002159
3	3	GM-122	6	4/3/2010	EID002147
4	4	GM-125	2	4/3/2010	EID002147
5	5	GM-120	8	4/4/2010	EID002159
6	6	GM-122	4	4/4/2010	EID002147
7	7	GM-125	4	4/4/2010	EID002147

Figure 6.16. Table of hours worked by each employee on each project.

When several tables are involved that are related to each other, the combination is called a **relational**

database. Excel can connect a few tables in a limited way. When there are several tables as described in this example, special purpose software called a **database management system (DBMS)** is needed. The database management system program provided as part of Microsoft Office Professional is Microsoft Access. If the Excel tables are imported into Access, they can be connected using the key fields, as shown in Figure 6.17.

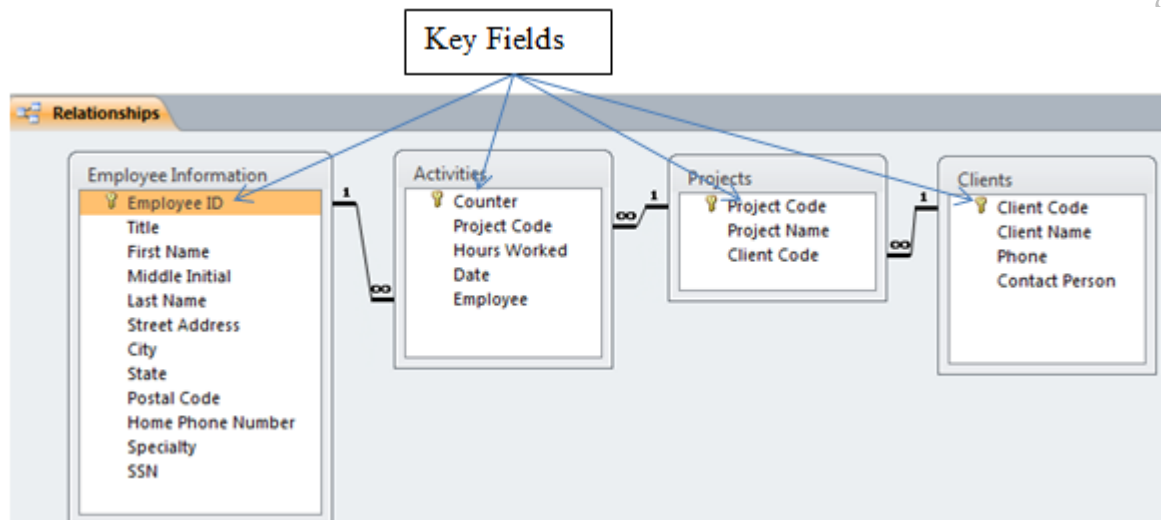


Figure 6.17. Relationships in a relational database.

Key Takeaways

- Tabulating the data from the 1880 census took seven years. They needed a faster way to count the responses in each category. [6.1.1]
- Numerals are the symbols used to represent numbers. Examples are letters in Roman numerals such as C for 100 and M for 1000. In Hexadecimal, the letter A represents the number ten. [6.1.2]
- Data are the raw facts. Information is what is derived from the data such as trends and patterns. [6.1.3]
- The binary system uses two numerals, 0 and 1, and each position in the number is an increasing multiple of two. The decimal system uses ten numerals, 0 through 9, and each position in the

number is an increasing multiple of ten. The hexadecimal system uses sixteen numerals, 0 through 9 and A through F. Each position in the number is an increasing multiple of sixteen. The Scientific View of the Calculator Accessory can convert numbers from one system to another.

[6.1.4]

- ASCII assigns a two-digit hexadecimal code to characters, numbers, and common symbols in the English language. There is a different code for capital and lower-case letters and the blank is a lower number than other letters so words sort correctly. [6.1.5]
- The UTF-8 code is backwardly compatible to the ASCII code but it is based on international standards so it can represent characters in non-English languages. [6.1.6]
- In personal computers, time is stored as a sequential number from a reference date such as Jan 1, 1900 (Windows) or Jan 1, 1904 (Macintosh). Hours are decimal fractions of a day, e.g. 6 AM is 0.25 days. Intervals between dates can be calculated by simply subtracting one date from the other in programs like Excel that run on personal computers. [6.1.7]
- A data table consists of rows called records that are about one thing. Each column in the table is a category of data. The categories are also called fields. [6.1.8]
- The field that contains a unique value for each record can be the key field. The key field is used to create a relationship with another table where that field might not be the key field for the purpose of gathering data from different tables to create information. [6.1.9]

2 Retrieving and Transferring Data

Learning Objectives

1. Identify the relationship between Structured Query Language and Query by Example. [6.2.1]
2. Identify and apply comparison operators for numbers, dates, and text. [6.2.2]
3. Identify and apply logic functions AND, OR, NOT, IS NULL, and IS NOT NULL. [6.2.3]

4. Identify the structure of a comma separated values data file. [6.2.4]
5. Identify techniques for exporting data tables and queries from Access. [6.2.5]

To select a few records and fields from millions that might be stored on the computer's disk drive, a user needs to ask for them in a particular way called a **query**. A query uses a set of formal commands that tell the computer what records to look for and which fields to display. The commands and the rules for using the commands make up a computer language called the **structured query language (SQL)**. Writing queries in this language takes training and is usually done by specialists. Programs like Microsoft Access provide a shortcut. Microsoft Access uses a system called **query by example (QBE)** that takes advantage of familiar Windows functions like drag-and-drop to create a query. It is used here to show how queries are structured and how a computer program interpretes the instructions. For example, the six fields in the grid shown in Figure 6.18 were placed in the QBE grid by dragging them from the field lists above.

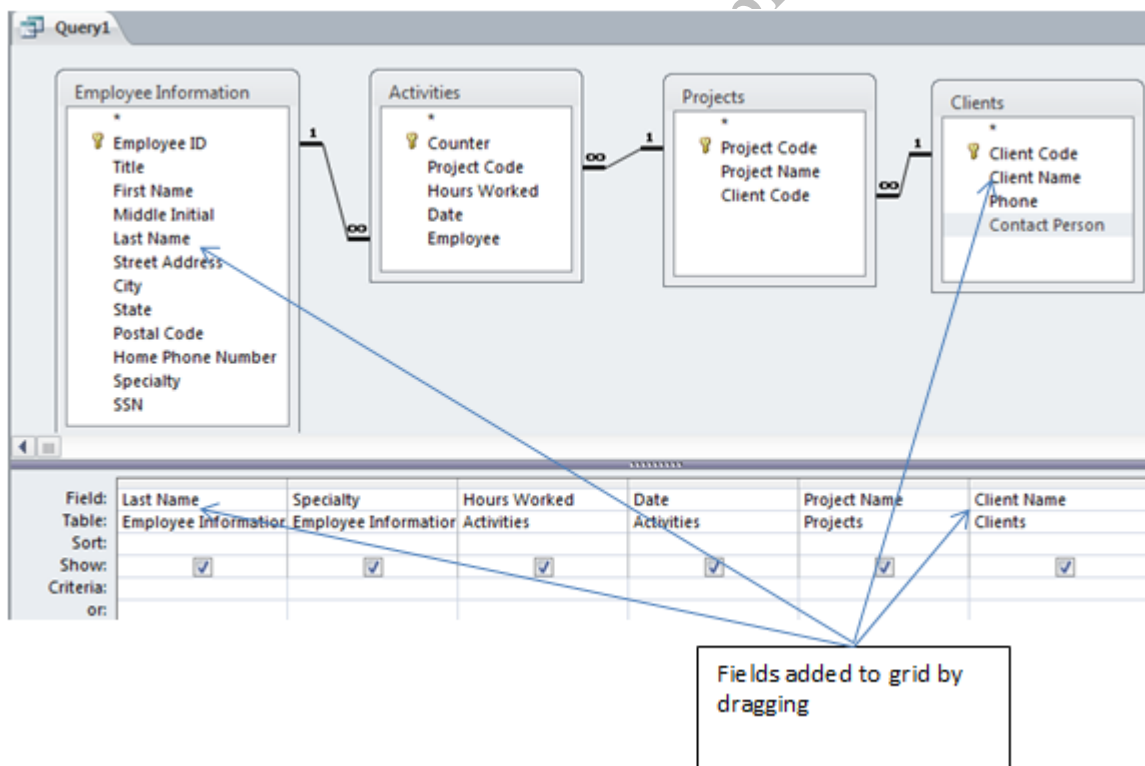


Figure 6.18. Field names from tables added to the query by example grid.

The Access program translates this grid into the SQL language as shown in Figure 6.19.

Query1

```
SELECT [Employee Information].[Last Name], [Employee Information].[Specialty], Activities.[Hours Worked], Activities.[Date], Projects.[Project Name], Clients.[Client Name]
FROM (Clients INNER JOIN Projects ON Clients.[Client Code] = Projects.[Client Code]) INNER JOIN ([Employee Information] INNER JOIN Activities ON [Employee
Information].[Employee ID] = Activities.[Employee]) ON Projects.[Project Code] = Activities.[Project Code];
```

Figure 6.19. The SQL code that will select all the records but only show the six chosen fields.

When this query is sent to the database management software that controls the computer's memory and data retrieval system, it returns all the records but only displays the six chosen fields, as shown in Figure 6.20.

Query1

Last Name	Specialty	Hours Worked	Date	Project Name	Client Name
Adams	Mechanical Engineer	8	4/4/2010	PV Panels	General Mechanics
Adams	Mechanical Engineer	4	4/5/2010	PV Panels	General Mechanics
Adams	Mechanical Engineer	2	4/6/2010	PV Panels	General Mechanics
Adams	Mechanical Engineer	5	4/7/2010	PV Panels	General Mechanics
Adams	Mechanical Engineer	6	4/10/2010	PV Panels	General Mechanics
Adams	Mechanical Engineer	7	4/11/2010	PV Panels	General Mechanics
Adams	Mechanical Engineer	6	4/12/2010	PV Panels	General Mechanics
Adams	Mechanical Engineer	7	4/13/2010	PV Panels	General Mechanics
Adams	Mechanical Engineer	4	4/14/2010	PV Panels	General Mechanics
Adams	Mechanical Engineer	8	4/17/2010	PV Panels	General Mechanics
Adams	Mechanical Engineer	8	4/18/2010	PV Panels	General Mechanics
Adams	Mechanical Engineer	2	4/19/2010	PV Panels	General Mechanics
Adams	Mechanical Engineer	2	4/20/2010	PV Panels	General Mechanics
Adams	Mechanical Engineer	1	4/27/2010	PV Panels	General Mechanics
Adams	Mechanical Engineer	6	4/3/2010	Warehouse insulation	General Mechanics
Baker	Civil Engineer	6	4/3/2010	Warehouse insulation	General Mechanics

Figure 6.20. All the records displayed with the six fields that were selected.

Comparison Operators

To limit the records that are displayed, restrictions called **criteria** are added to the grid. One type of criterion is a numerical comparison. Comparison operators are used to help limit the records that display. A list of the comparison operators and their meanings are shown in Figure 6.21.

Symbol	Meaning
=	equal to
<	Less than
<=	Less than or equal to
>	Greater than
>=	Greater than or equal to
<>	not equal

Figure 6.21. Comparison operators.

For example, the company has a policy of charging for an hour's time for any part of an hour. If a manager wants to see the records where the *Hours Worked* field contains values of less than 1, the criterion would be *<1* as shown in Figure 6.22.

The image shows a database query design grid with the following fields: Last Name, Specialty, Hours Worked, Date, Project Name, and Client Name. The table is 'Employee Information', the sort is 'Employee Information', and the show is 'Activities'. The criteria for 'Hours Worked' is '<1'. A callout box points to this criterion with the text 'Criterion for hours worked, less than 1'.

Field:	Last Name	Specialty	Hours Worked	Date	Project Name	Client Name
Table:	Employee Information	Employee Information	Activities	Activities	Projects	Clients
Sort:						
Show:	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Criteria:			<1			
or:						

Criterion for hours worked, less than 1

Figure 6.22. Restricting the records to those where the value in the Hours Worked field is less than 1.

Because dates are stored as numbers, we can say that an earlier date is less than a later date. For example, if we enter the criterion *< 4/8/2010* in the Date field, it would return all of the records for work done before (but not including) April 8, 2010. Because the ASCII and UTF codes assign numbers to the letters in the same sequence as the alphabet, we can use numerical comparison operators with words. If we want to write a criterion that restricted the records to workers whose last names came before Baker, the criterion *<Baker* under the Last Name field would work.

Logic Functions

Criteria can be combined using logic functions. Five commonly used logic functions are described in Figure 6.23.

Logic Function	Description
AND	Both are true
OR	Either are true
NOT	Criterion is not true
IS NULL	Empty field (not the same as zero)
IS NOT NULL	Field is not empty

Figure 6.23. Commonly used logic functions.

In the Access QBE grid, criteria can be written for each field. They are combined using the **AND** function when they are translated into the query language. For example, if we wanted to restrict the records that meet the criteria of Hours Worked greater than 4 and the client name to General Mechanics they would both be written in the QBE grid, as shown in Figure 6.24.

Field:	Last Name	Specialty	Hours Worked	Date	Project Name	Client Name
Table:	Employee Information	Employee Information	Activities	Activities	Projects	Clients
Sort:						
Show:	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Criteria:			>4			= "General Mechanics"

Greater than 4 AND General Mechanics

Figure 6.24. Two criteria.

When this is translated into SQL, the AND function is used as shown in Figure 6.25.

```

SELECT [Employee Information].[Last Name], [Employee Information].[Specialty], Activities.[Hours Worked], Activities.[Date], Projects.[Project Name], Clients.[Client Name]
FROM ([Clients INNER JOIN Projects ON Clients.[Client Code] = Projects.[Client Code]) INNER JOIN ([Employee Information] INNER JOIN Activities ON [Employee Information].[Employee ID] = Activities.[Employee ID] ON Projects.[Project Code] = Activities.[Project Code])
WHERE (([Activities].[Hours Worked])>4) AND ([Clients].[Client Name]="General Mechanics");
  
```

Figure 6.25. The AND function requires that both criteria are true.

If the record should be displayed if either of two conditions is true, the **OR** function is used. For example, if you wanted to see the records where the date was before 4/7/2010 or after 4/21/2010 the two criteria are placed on separate lines in the QBE grid, as shown in Figure 6.26. The pound signs are added by the program to indicate a date field.

Field:	Last Name	Specialty	Hours Worked	Date	Project Name	Client Name
Table:	Employee Information	Employee Information	Activities	Activities	Projects	Clients
Sort:						
Show:	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
riter:				< #4/7/2010#		
or:				> #4/21/2010#		

Figure 6.26. The OR function displays the record if either criterion is true.

When this is translated into SQL, the OR function is used as shown in Figure 6.27.

```

Project Work
SELECT [Employee Information].[Last Name], [Employee Information].[Specialty], Activities.[Hours Worked], Activities.[Date], Projects.[Project Name], Clients.[Client Name]
FROM (Clients INNER JOIN Projects ON Clients.[Client Code] = Projects.[Client Code]) INNER JOIN ([Employee Information] INNER JOIN Activities ON [Employee
Information].[Employee ID] = Activities.[Employee ID]) ON Projects.[Project Code] = Activities.[Project Code]
WHERE (((Activities.Date) < #4/7/2010#) OR ((Activities.Date) > #4/21/2010#));

```

Figure 6.27. The OR function used in the SQL language.

When all of the records should be displayed except those that match a criteria, the **NOT** function is used. For example, if all of the records except those for the General Mechanics client, the criteria would be NOT “General Mechanics”.

Recall that there is a difference between zero, blank, and NULL. The **IS NULL** function is useful for checking to see if any fields were left empty. To display records where no value was entered for the Hours Worked field, the criterion under Hours Worked would be IS NULL. Similarly, to find records where a value is present in the field, the **IS NOT NULL** criterion is used.

Exporting Data from a Database Program

Computers usually store data as a sequence of very tiny spots on a spinning disk or on arrays of tiny transistors. On disks, two different types of spots are used to represent zeros or ones. This coding method makes it possible to represent binary numbers on a physical disk. These binary numbers can be interpreted using ASCII or UTF codes to determine what words, dates, or numbers they represent. The data is stored in a continuous sequence of numbers instead of a rectangular table.

Tip: On a hard disk, the surface is coated with an iron-oxide that can be magnetized. A spot can be magnetized up or down to indicate a one or zero. On an optical disc, a spot is “burned” on the disc by a laser at high power. A low power laser will reflect from a spot if it has not been burned (1) or reflect poorly if it has (0). (Flexible plastic media are

usually called discs spelled with a c while rigid metal media are called disks, spelled with a k.)

Tables of data can be reconstructed from the continuous sequence of number by placing a special character between the data in each field and another special character to mark the end of each record. One of the most common methods is to use a comma character to separate fields and a **Line Feed (LF)** or **Carriage Return (CR)** character to mark the end of a record. This method of coding tables of data is called **comma separated values (csv)**. Spreadsheet programs like Excel can convert this type of data into a table. On many websites that contain tables of data, there is an option to download the data in csv format, as shown in Figure 6.28.

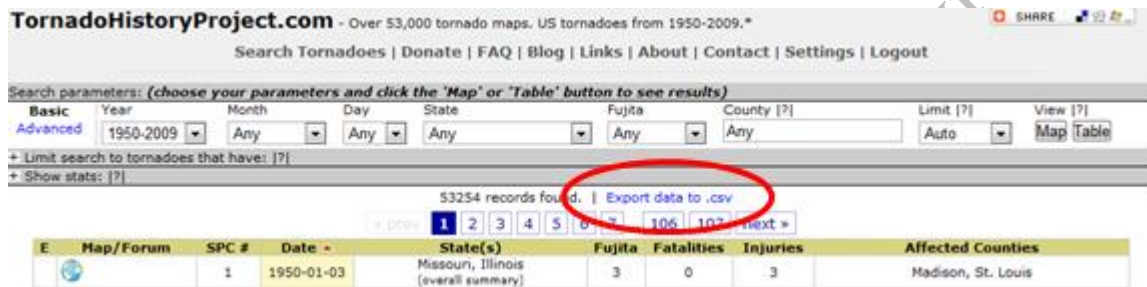


Figure 6.28. Export data in comma separated values format.

Excel can recreate the data table from the csv file. Alternatively, tables and the data assembled in queries can be exported from the Access program into several formats including an Excel file, as shown in Figure 6.29.

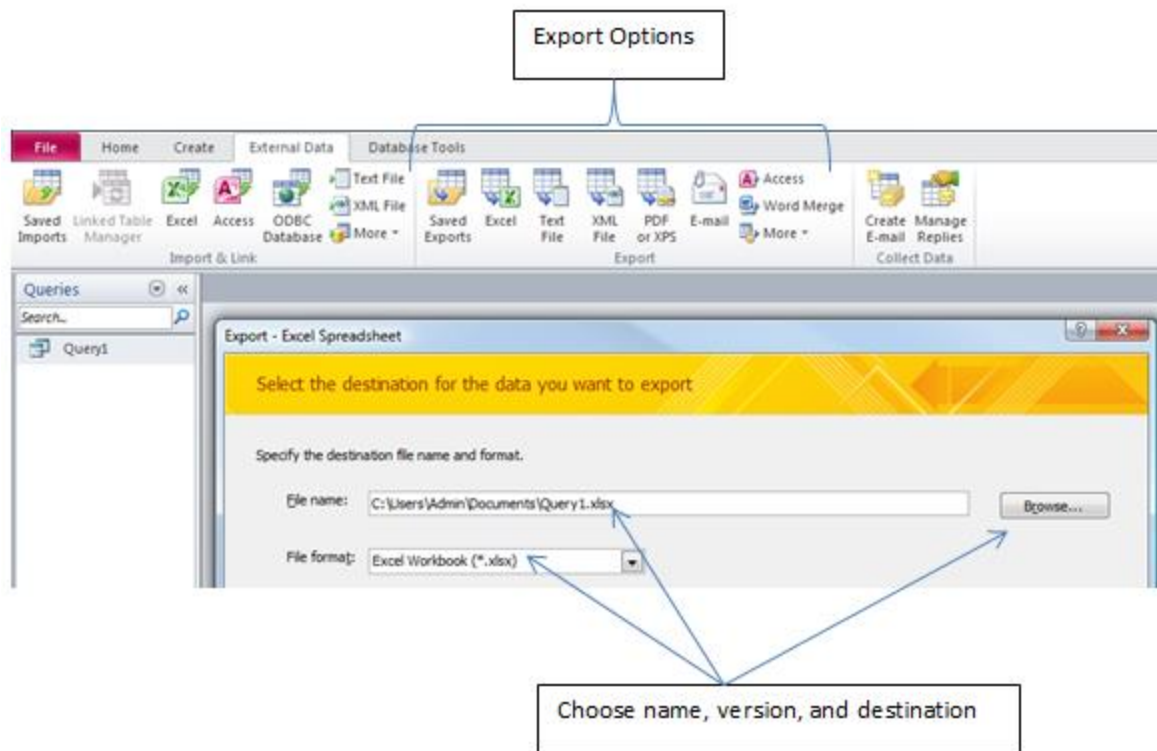


Figure 6.29. Access query exported as an Excel workbook.

Key Takeaways

- The structured query language (SQL) is a programming language that is used to retrieve data from a database management system. The query by example is a more user friendly interface used in Access where the user can drag fields into a grid and type criteria that will be translated into SQL commands: [6.2.1]
- The comparison operators and their meanings are: [6.2.2]
 - = equal to
 - < less than
 - <= less than or equal to
 - > greater than
 - >= greater than or equal to
 - <> not equal
- The logic functions and their meanings are: [6.2.3]

- AND both criteria must be true
- OR either criteria is true
- NOT the opposite of true or false
- IS NULL the field is empty
- IS NOT NULL the field has data in it
- A comma separated file uses commas to separate the fields and a line feed or carriage return character to mark the end of a record. [6.2.4]
- Use the External Data/Export option to translate the table or query result into an Excel file or other file format. [6.2.5]

3 Importing Data from the Web

Learning Objectives

1. Copy data from Web pages into Excel using Paste Special options. [6.3.1]
2. Importing tables from Web pages. [6.3.2]
3. Importing text files. [6.3.3]
4. Moving or copying worksheets between workbooks. [6.3.4]
5. Importing tables and queries from Access. [6.3.5]

In this section, you practice transferring data from different sources into Excel. The instructions assume that your computer is connected to the Internet and that you are using Internet Explorer as your Web browser.

Copying from a Web Page

In many cases, data is available on a website. If it can be selected, it can usually be copied and then pasted into a spreadsheet.

1. Start Excel. Navigate to the folder where the files for this class are stored, and then open Ch06Data. Save the file on your computer or memory device as **Ch06DataStudentName.xlsx**
2. Start a Web browser and go to <http://rockcreekweather.com/wxclientraw.php> This is a website that displays weather data from a site in Oregon, as shown in Figure 6.30.

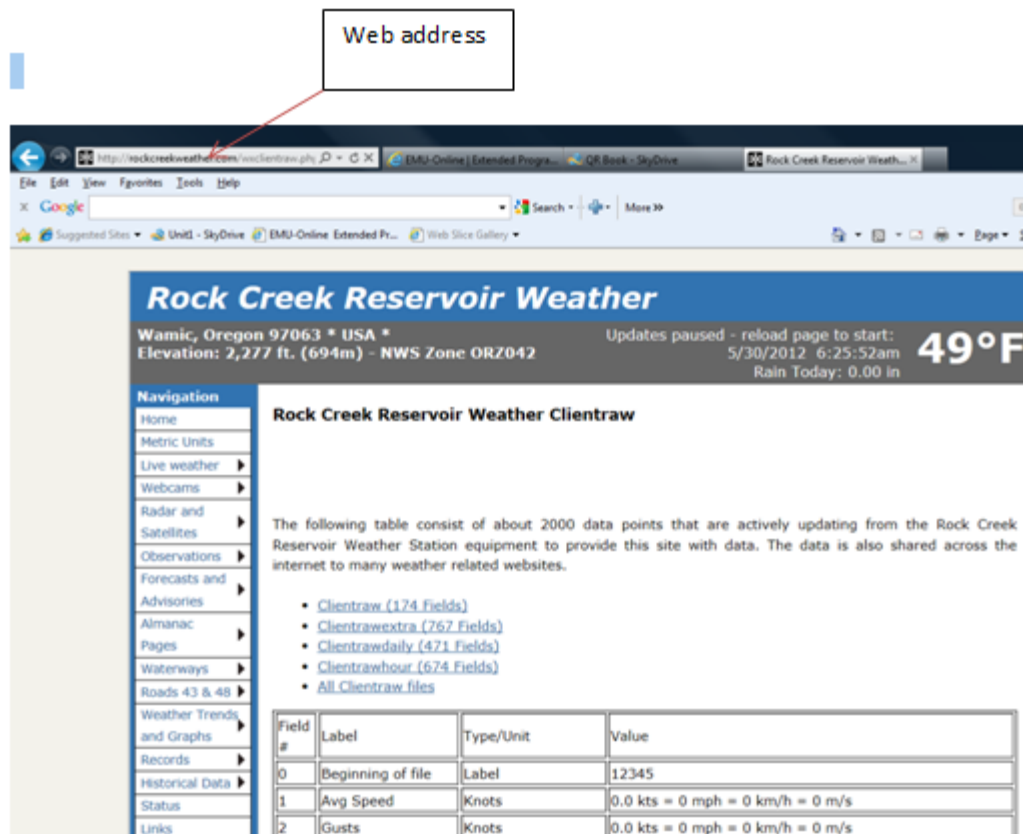


Figure 6.30. Website with current weather data.

3. Drag across the title—Rock Creek Reservoir Weather—to select it. Press and hold the Ctrl key and then press C to copy the selected title. Switch to the workbook. At the bottom of the workbook window, click the Wind sheet tab.
4. In the Wind worksheet, click cell A1. On the Home tab, in the Clipboard group, under the Paste button, click the arrow. On the menu, click Paste Special. The Paste Special dialog box displays. Click the HTML option, as shown in Figure 6.31.

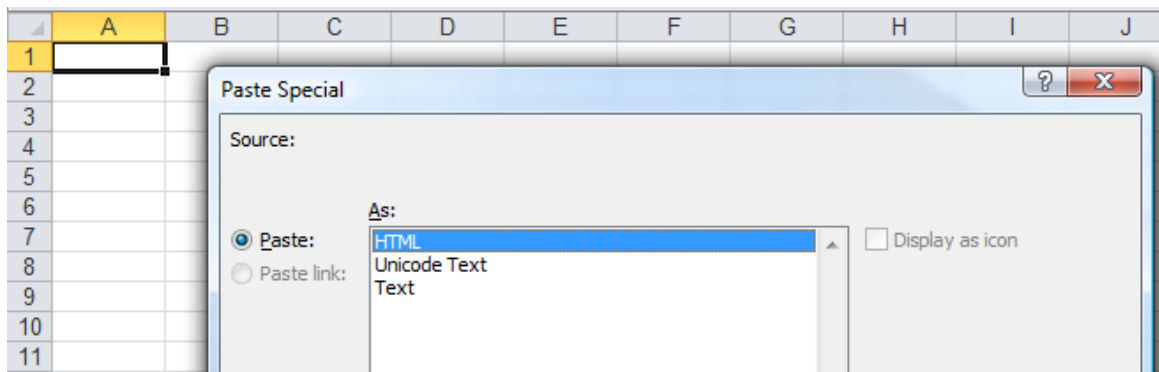


Figure 6.31. Special paste options.

Tip: The HTML option will retain some of the formatting used to display the data on the webpage. The Unicode Text option refers to the UTF-8 or UTF-16 codes that might be useful if the data is in a foreign language. The Text option uses the simple ASCII codes for text and numbers without formatting.

5. At the bottom of the Paste Special dialog box, click OK. The title of the webpage is pasted into the cell as a hyperlink back to the webpage.

Tip: If you use a browser other than Internet Explorer, this may not work as described. If you have problems, open Internet Explorer and try again.

6. Switch back to the webpage. Below the title, select the date and time. Use the method you practiced previously to copy the date and time and paste it into cell A2 in the Wind worksheet.
7. Switch back to the webpage. Select the first row of the table that begins with *Field #* and ends with *Value*. Use the method you practiced previously to copy the first row of column headings and paste it as HTML into cell A4 in the Wind worksheet.
8. In the Wind worksheet, drag across column headings A through D to select all four headings. Right click one of the selected columns. On the shortcut menu, click Column Width. In the Column Width dialog box, type 40, and then click OK.

Tip: It is easier to manage the row heights if the text does not wrap in narrow columns when it is first pasted into the worksheet. To prevent the wrapping, widen the columns enough to accept the longest string of text before pasting.

9. Switch back to the webpage. Scroll down the page to locate the rows that contain the wind speeds that begin about row 51.

10. Drag the entire row of data for *Windspeed Hour 01* to select it. Confirm that all of the data in the row is selected, as shown in Figure 6.32. The wind speed data in your table will be different.

All of the data in the row is selected

51	Windspeed Hour 01	Knots (Used for Windspeed Chart)	6 kts = 6.91 mph = 11.12 km/h = 3.09 m/s
52	Windspeed Hour 02	Knots (Used for Windspeed Chart)	6 kts = 6.91 mph = 11.12 km/h = 3.09 m/s

Figure 6.32. First row of data selected.

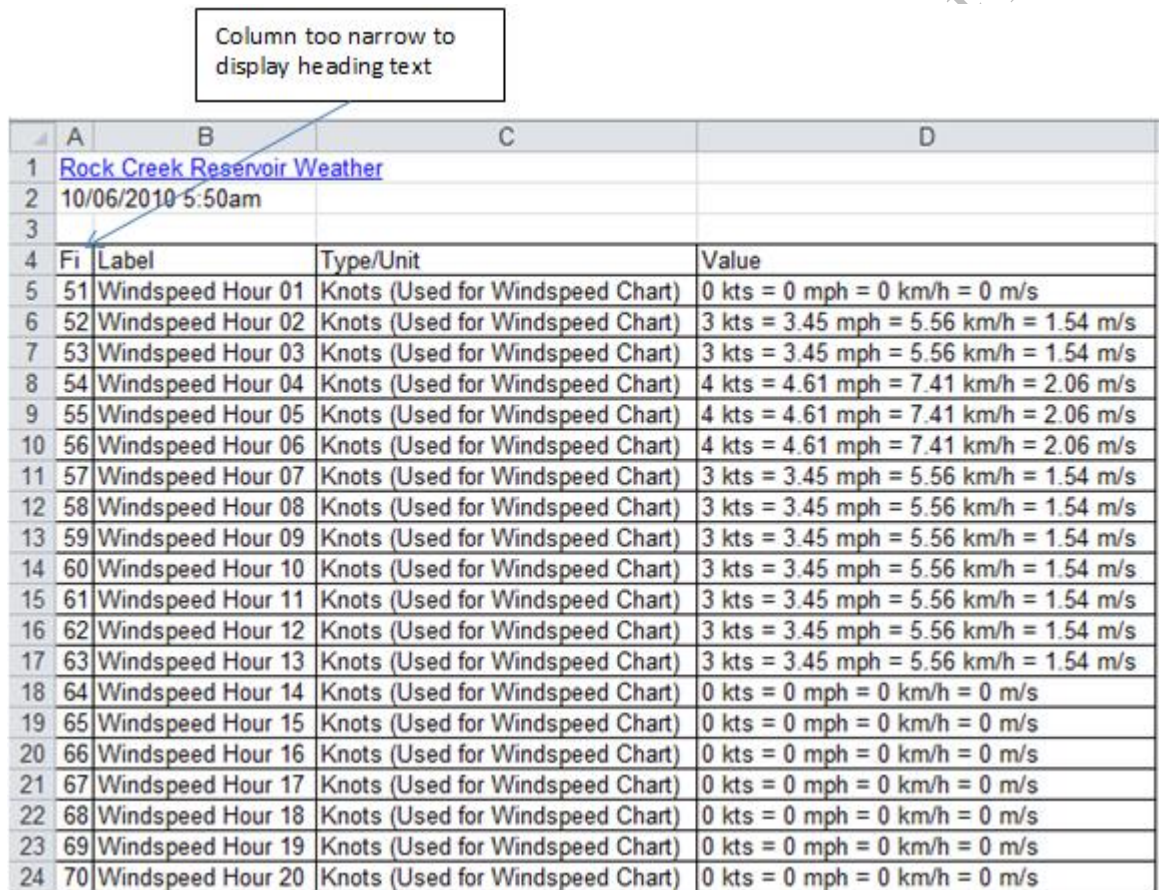
11. Scroll to downward to the row for *Windspeed Hour 20*. Move the pointer to the end of the row. Hold down the Shift key, and then click at the end of the row. The rows that contain the data for wind speeds from hour 1 through hour 20 are selected, as shown in Figure 6.33.

Wind speed rows selected

62	Windspeed Hour 12	Knots (Used for Windspeed Chart)	10 kts = 11.52 mph = 18.53 km/h = 5.15 m/s
63	Windspeed Hour 13	Knots (Used for Windspeed Chart)	10 kts = 11.52 mph = 18.53 km/h = 5.15 m/s
64	Windspeed Hour 14	Knots (Used for Windspeed Chart)	7 kts = 8.06 mph = 12.97 km/h = 3.6 m/s
65	Windspeed Hour 15	Knots (Used for Windspeed Chart)	7 kts = 8.06 mph = 12.97 km/h = 3.6 m/s
66	Windspeed Hour 16	Knots (Used for Windspeed Chart)	7 kts = 8.06 mph = 12.97 km/h = 3.6 m/s
67	Windspeed Hour 17	Knots (Used for Windspeed Chart)	7 kts = 8.06 mph = 12.97 km/h = 3.6 m/s
68	Windspeed Hour 18	Knots (Used for Windspeed Chart)	7 kts = 8.06 mph = 12.97 km/h = 3.6 m/s
69	Windspeed Hour 19	Knots (Used for Windspeed Chart)	8 kts = 9.21 mph = 14.83 km/h = 4.12 m/s
70	Windspeed Hour 20	Knots (Used for Windspeed Chart)	8 kts = 9.21 mph = 14.83 km/h = 4.12 m/s
71	Max Wind Gust	Knots (Used for Windspeed Chart)	12.2 kts = 14.05 mph = 22.61 km/h = 6.28 m/s

Figure 6.33. Portion of the data for wind speeds is selected.

12. With the rows that show the wind speed for hours 1 through 20 selected, press Ctrl key + C to copy the data. Alternatively, from the browser's toolbar, click the Copy button.
13. Switch to the Wind worksheet. Paste the data into cell A5 as HTML.
14. With the data still selected from pasting, on the Home tab, in the Cells group, click Format. On the Format menu, click *AutoFit Column Width*. The text is transferred and formatted as shown in Figure 6.34. Observe that width of column A is too narrow to display the heading in cell A4.



Column too narrow to display heading text

	A	B	C	D
1	Rock Creek Reservoir Weather			
2	10/06/2010 5:50am			
3				
4	Fi	Label	Type/Unit	Value
5	51	Windspeed Hour 01	Knots (Used for Windspeed Chart)	0 kts = 0 mph = 0 km/h = 0 m/s
6	52	Windspeed Hour 02	Knots (Used for Windspeed Chart)	3 kts = 3.45 mph = 5.56 km/h = 1.54 m/s
7	53	Windspeed Hour 03	Knots (Used for Windspeed Chart)	3 kts = 3.45 mph = 5.56 km/h = 1.54 m/s
8	54	Windspeed Hour 04	Knots (Used for Windspeed Chart)	4 kts = 4.61 mph = 7.41 km/h = 2.06 m/s
9	55	Windspeed Hour 05	Knots (Used for Windspeed Chart)	4 kts = 4.61 mph = 7.41 km/h = 2.06 m/s
10	56	Windspeed Hour 06	Knots (Used for Windspeed Chart)	4 kts = 4.61 mph = 7.41 km/h = 2.06 m/s
11	57	Windspeed Hour 07	Knots (Used for Windspeed Chart)	3 kts = 3.45 mph = 5.56 km/h = 1.54 m/s
12	58	Windspeed Hour 08	Knots (Used for Windspeed Chart)	3 kts = 3.45 mph = 5.56 km/h = 1.54 m/s
13	59	Windspeed Hour 09	Knots (Used for Windspeed Chart)	3 kts = 3.45 mph = 5.56 km/h = 1.54 m/s
14	60	Windspeed Hour 10	Knots (Used for Windspeed Chart)	3 kts = 3.45 mph = 5.56 km/h = 1.54 m/s
15	61	Windspeed Hour 11	Knots (Used for Windspeed Chart)	3 kts = 3.45 mph = 5.56 km/h = 1.54 m/s
16	62	Windspeed Hour 12	Knots (Used for Windspeed Chart)	3 kts = 3.45 mph = 5.56 km/h = 1.54 m/s
17	63	Windspeed Hour 13	Knots (Used for Windspeed Chart)	3 kts = 3.45 mph = 5.56 km/h = 1.54 m/s
18	64	Windspeed Hour 14	Knots (Used for Windspeed Chart)	0 kts = 0 mph = 0 km/h = 0 m/s
19	65	Windspeed Hour 15	Knots (Used for Windspeed Chart)	0 kts = 0 mph = 0 km/h = 0 m/s
20	66	Windspeed Hour 16	Knots (Used for Windspeed Chart)	0 kts = 0 mph = 0 km/h = 0 m/s
21	67	Windspeed Hour 17	Knots (Used for Windspeed Chart)	0 kts = 0 mph = 0 km/h = 0 m/s
22	68	Windspeed Hour 18	Knots (Used for Windspeed Chart)	0 kts = 0 mph = 0 km/h = 0 m/s
23	69	Windspeed Hour 19	Knots (Used for Windspeed Chart)	0 kts = 0 mph = 0 km/h = 0 m/s
24	70	Windspeed Hour 20	Knots (Used for Windspeed Chart)	0 kts = 0 mph = 0 km/h = 0 m/s

Figure 6.34. Table of data from a Web page is pasted into Excel.

15. Point to the line between column headings A and B, drag the line to the right until column A is wide enough to display the heading in cell A4, as shown in Figure 6.35.

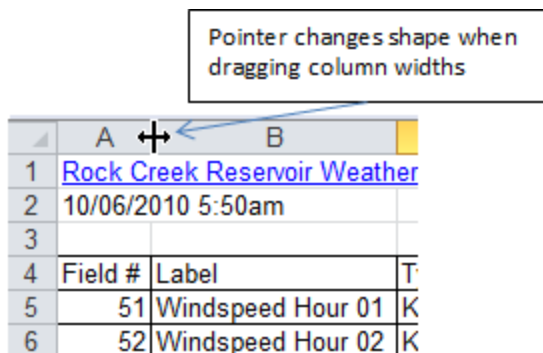


Figure 6.35. Widening column A manually.

16. The data is transferred and labeled appropriately, as shown in Figure 6.36. Save the workbook.

	A	B	C	D
1	Rock Creek Reservoir Weather			
2	10/06/2010 5:50am			
3				
4	Field #	Label	Type/Unit	Value
5	51	Windspeed Hour 01	Knots (Used for Windspeed Chart)	0 kts = 0 mph = 0 km/h = 0 m/s
6	52	Windspeed Hour 02	Knots (Used for Windspeed Chart)	3 kts = 3.45 mph = 5.56 km/h = 1.54 m/s
7	53	Windspeed Hour 03	Knots (Used for Windspeed Chart)	3 kts = 3.45 mph = 5.56 km/h = 1.54 m/s
8	54	Windspeed Hour 04	Knots (Used for Windspeed Chart)	4 kts = 4.61 mph = 7.41 km/h = 2.06 m/s
9	55	Windspeed Hour 05	Knots (Used for Windspeed Chart)	4 kts = 4.61 mph = 7.41 km/h = 2.06 m/s
10	56	Windspeed Hour 06	Knots (Used for Windspeed Chart)	4 kts = 4.61 mph = 7.41 km/h = 2.06 m/s
11	57	Windspeed Hour 07	Knots (Used for Windspeed Chart)	3 kts = 3.45 mph = 5.56 km/h = 1.54 m/s
12	58	Windspeed Hour 08	Knots (Used for Windspeed Chart)	3 kts = 3.45 mph = 5.56 km/h = 1.54 m/s
13	59	Windspeed Hour 09	Knots (Used for Windspeed Chart)	3 kts = 3.45 mph = 5.56 km/h = 1.54 m/s
14	60	Windspeed Hour 10	Knots (Used for Windspeed Chart)	3 kts = 3.45 mph = 5.56 km/h = 1.54 m/s
15	61	Windspeed Hour 11	Knots (Used for Windspeed Chart)	3 kts = 3.45 mph = 5.56 km/h = 1.54 m/s
16	62	Windspeed Hour 12	Knots (Used for Windspeed Chart)	3 kts = 3.45 mph = 5.56 km/h = 1.54 m/s
17	63	Windspeed Hour 13	Knots (Used for Windspeed Chart)	3 kts = 3.45 mph = 5.56 km/h = 1.54 m/s
18	64	Windspeed Hour 14	Knots (Used for Windspeed Chart)	0 kts = 0 mph = 0 km/h = 0 m/s
19	65	Windspeed Hour 15	Knots (Used for Windspeed Chart)	0 kts = 0 mph = 0 km/h = 0 m/s
20	66	Windspeed Hour 16	Knots (Used for Windspeed Chart)	0 kts = 0 mph = 0 km/h = 0 m/s
21	67	Windspeed Hour 17	Knots (Used for Windspeed Chart)	0 kts = 0 mph = 0 km/h = 0 m/s
22	68	Windspeed Hour 18	Knots (Used for Windspeed Chart)	0 kts = 0 mph = 0 km/h = 0 m/s
23	69	Windspeed Hour 19	Knots (Used for Windspeed Chart)	0 kts = 0 mph = 0 km/h = 0 m/s
24	70	Windspeed Hour 20	Knots (Used for Windspeed Chart)	0 kts = 0 mph = 0 km/h = 0 m/s

Figure 6.36. Wind speed data transferred from the webpage to a worksheet.

Importing a Table from a Web Page

Data can be formatted to appear as a table in a Web page. The entire table can be selected and imported into Excel.

1. In the Ch06DataStudentName workbook, at the bottom of the workbook, click the NatGas sheet tab.

2. In cell A1, type **U.S. Residential Price of Natural Gas (\$ per thousand cubic feet)**, and then press Enter.
3. On the Data tab, click *Get External Data*. The menu of options displays, as shown on Figure 6.37.

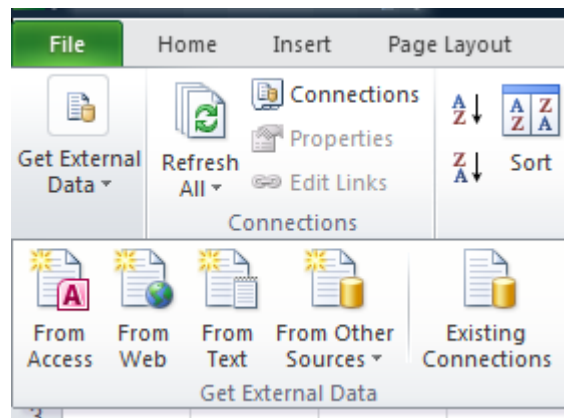


Figure 6.37. Options for importing external data.

Tip: If you have a wider screen, the menu of choices may appear on the ribbon instead of the Get External Data button.

4. Click *From Web*. A web browser window opens. In the Address box, type <http://www.eia.doe.gov/dnav/ng/hist/n3010us3m.htm> and then press Enter. A webpage with a table of natural gas prices displays. Each section that can be imported as a separate table is indicated with a small arrow, as shown in Figure 6.38.



Figure 6.38. Table of data on a webpage.

- Refer to Figure 6.38. Click the small yellow arrow near top of the web page, and then click the Import

button.

6. In the Import Data dialog box, confirm that the data will be imported to the Existing worksheet in cell A2, and then click OK. The data is transferred into the spreadsheet.

7. Delete rows 2:11 and adjust the column widths to fit the labels that now display in row 2, as shown in Figure 6.39.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	U.S. Residential Price of Natural Gas (\$ per thousand cubic feet)												
2	Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
3	1973	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
4	1974	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
5													
6	1975	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
7	1976	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
8	1977	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
9	1978	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
10	1979	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
11													
12	1980	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
13	1981	3.94	3.99	4.06	4.11	4.29	4.3	4.32	4.3	4.47	4.5	4.53	4.55
14	1982	4.65	4.69	4.78	4.86	5.17	5.2	5.23	5.23	5.41	5.66	5.68	5.74
15	1983	5.86	5.87	6	6.06	6.22	6.2	6.21	6.18	6.19	6.7	6.3	5.94
16	1984	5.78	5.84	5.92	5.96	6.27	6.76	7.11	7.23	7.17	6.8	6.31	6.05

Figure 6.39. Table of data imported from the webpage.

Transferring Data as Text

Data can be transferred as a string of text where the fields in the table are separated by a text character such as a comma. This was a common practice in the early years of computing and older data files are often available in the comma separated values (csv) format. The Excel program can reconstruct the table of data if the characters used to separate the fields and rows are identified. The most common method is to use commas to separate the fields in each row. In this section, you export a file from a website as a csv file and then import it into a worksheet.

1. Open a Web browser, in the Address box, type http://www.eia.doe.gov/states/_seds.html and then press Enter. This web site includes energy information by category and by state as shown in Figure 6.40. Some

of the older data is available in older formats.

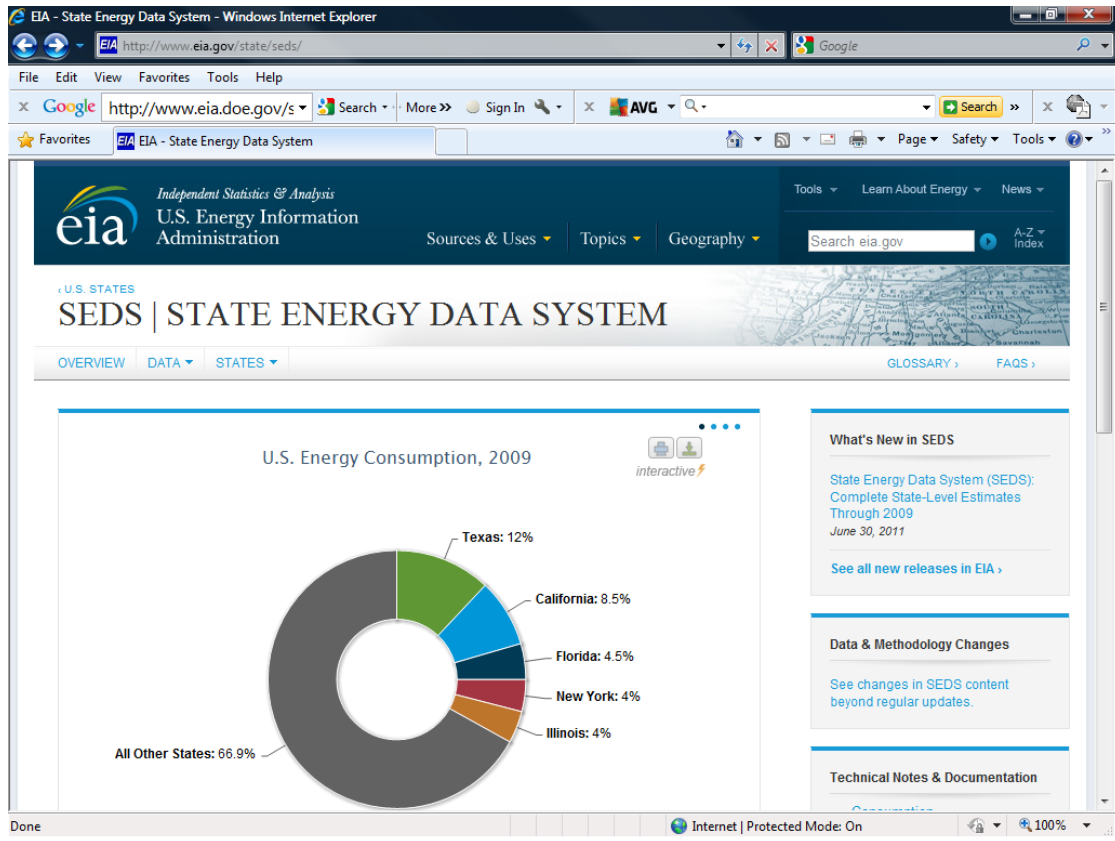


Figure 6.40. Department of Energy web site for State energy data.

Tip: In the early days of computing, the underscore character was often used instead of a space to make long names more readable. Sites that refer to older data often have underscores in the Web address like the one in the previous step. The later convention of underlining a Web address to indicate a hyperlink caused confusion because the underline obscured the underscore character and made it look like a space, e.g.
http://www.eia.doe.gov/states/_seds.html.

2. On the tab near the top, point to States and then click Michigan. On the Michigan page click Consumption Estimates, 1960-2009 and notice that this is available in csv format as shown in Figure 6.41.

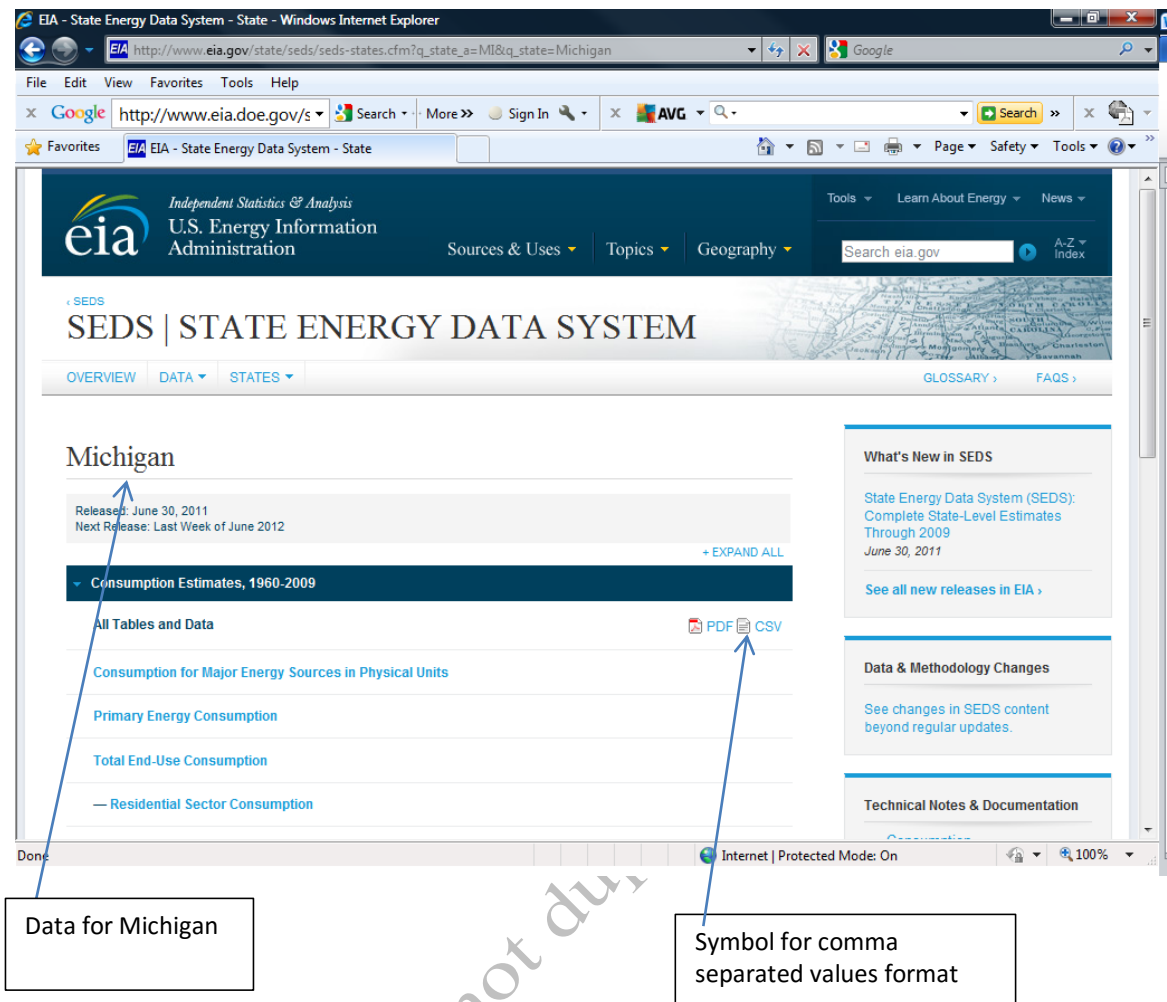



Figure 6.41. Michigan consumption data available in csv format.

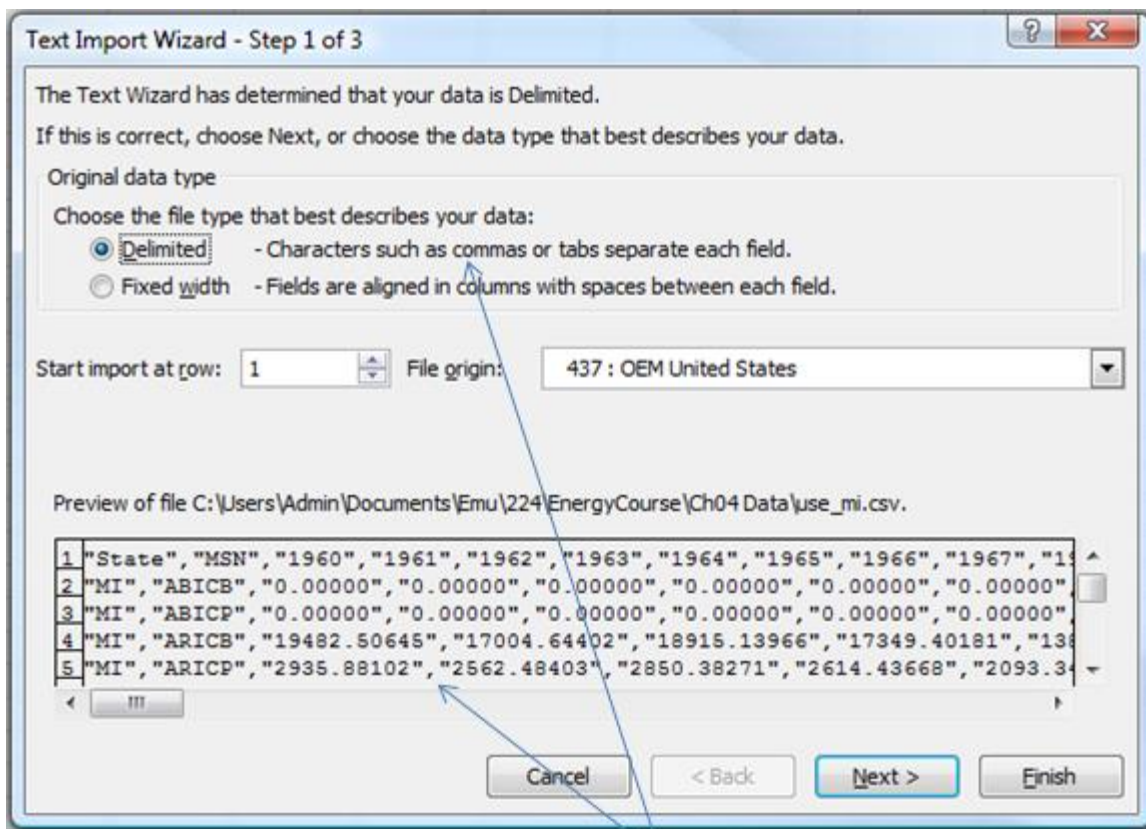
Tip: Another option for displaying data is the use of **portable document format (PDF)** format. This method allows for a display of data tables that are intended to be viewed with their formatting but not easily edited. PDF files are viewable with a program called a **reader** that is available from Adobe.com at no charge. Spreadsheets and other documents can be saved in PDF format from the Save As menu in Excel and Word. This format is useful if the objective is to share data in a formatted document that should not be editable by anyone else.

3. Under *Consumption, 1960-2009*, to the right of *All Tables and Data*, click the CSV button, . The File Download dialog box displays.
4. In the File Download dialog box, click the Save button. In the Save As dialog box, navigate to a folder on your computer or memory storage device where you store your work for this class, and then save the file

in that folder. Observe that the name of the file is use_mi.csv. Do not open the file. Close the dialog box after the file has been saved, but leave the browser open to this site.

5. Switch to the Ch06DataStudentName workbook, and then click the Mich sheet tab.
6. Click cell A2 to leave room for a title.
7. On the Data tab, click the *Get External Data* button, and then click From Text.
8. In the *Import Text File* dialog box, navigate to the folder where use_mi.csv is stored. Select it, and then click the Import button. The *Text Import Wizard* dialog box displays, as shown in Figure 6.42.

Tip: A wizard is a program that assists with a process. The Text Import Wizard asks questions about the text file to determine what character is used to separate the fields. When a character is used to separate fields, it is called a **delimiter**. In some early databases where memory storage was very limited, delimiters were not used. Instead, the same section of the row was reserved for each field called a **fixed width**. To import data from that type of database, a key to the width of each field is required. The program also asks what type of data is in each field so that it can be formatted correctly.



Commas used to separate fields

Figure 6.42. First step of the Text Import Wizard.

9. Confirm that *Delimited* is selected, and then click the Next button. The second dialog displays. Observe that the Tab character is the default choice and that a preview of the data is shown in the lower portion of the box.
10. Under Delimiters, next to the Tab option, click the check box to deselect it. Next to Comma, click the check box to select it as the delimiter. Observe that the program correctly identifies how the data should be separated into columns in the Data preview window, as shown in Figure 6.43.

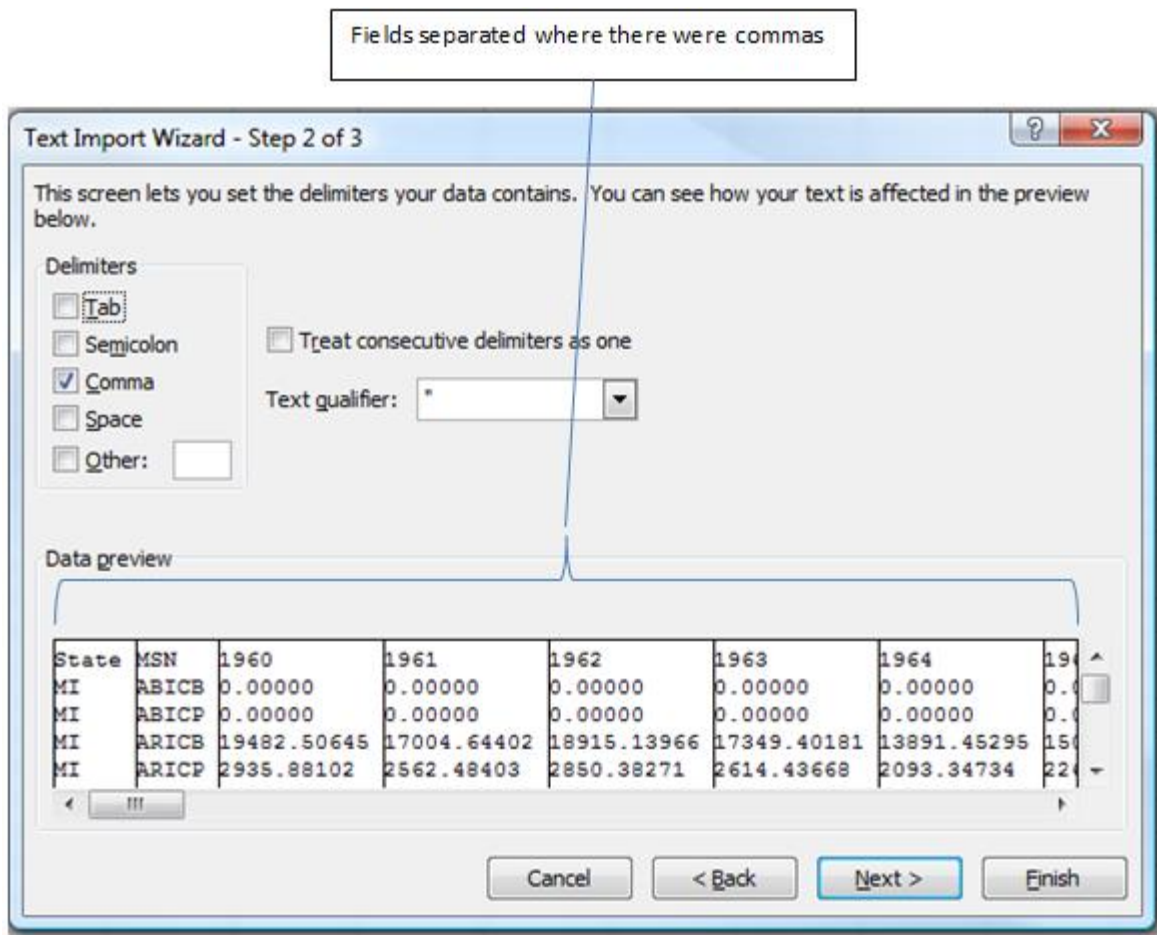


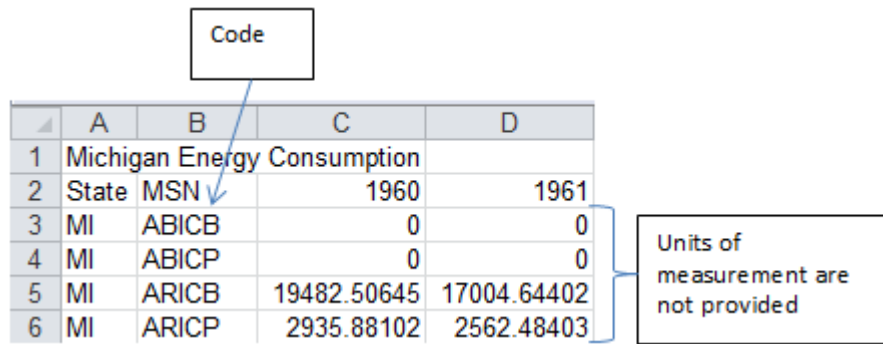
Figure 6.43. Second step of the Text Import Wizard.

11. Click the Next button. Under *Column data format* confirm that General is selected for each column.

Tip: The General format attempts to recognize numbers and dates and format them accordingly. Dates must be handled carefully. In the U.S., dates are usually abbreviated using a month/day/year sequence. It is more logical to represent dates using an order that increases or decreases from left to right such as day/month/year or year/month/day. In Europe, the day/month/year format is commonly used. Computer programmers prefer to use year/month/day because it is easier to sort. It is usually a good idea to scroll across the columns of data to see if there are fields with dates in them. Individual years such as those in the first row of this table do not need a date format.

12. Click the Finish button. Confirm that the data will be imported into the existing worksheet beginning at cell A2, and then click OK.
13. In cell A1, type **Michigan Energy Consumption** and then press Enter.
14. Observe that the numbers do not have unit labels and there is an unexplained code in column B, as shown

in Figure 6.44. Save the workbook.



	A	B	C	D
1	Michigan Energy Consumption			
2	State MSN		1960	1961
3	MI	ABICB	0	0
4	MI	ABICP	0	0
5	MI	ARICB	19482.50645	17004.64402
6	MI	ARICP	2935.88102	2562.48403

Figure 6.44. Imported data is not labeled.

Transferring Data as an Excel File

Many providers of data offer to export the data as an Excel worksheet. Excel switched file formats beginning with Excel 2007 for Windows. The old format used the file extension .xls and the newer version uses .xlsx. Because the newer versions of the Excel program can read the older file formats, the data files are usually provide in .xls format that can be read by almost any spreadsheet program. The xls extension is often used to denote the file type.

1. Switch back to the Web browser that is still displaying the State Energy Data System (SEDS) for the state of Michigan.
2. On the right, under *Technical Notes, Documentation, Codes, and Descriptions*, click *Data File Documentation* as shown in Figure 6.45.

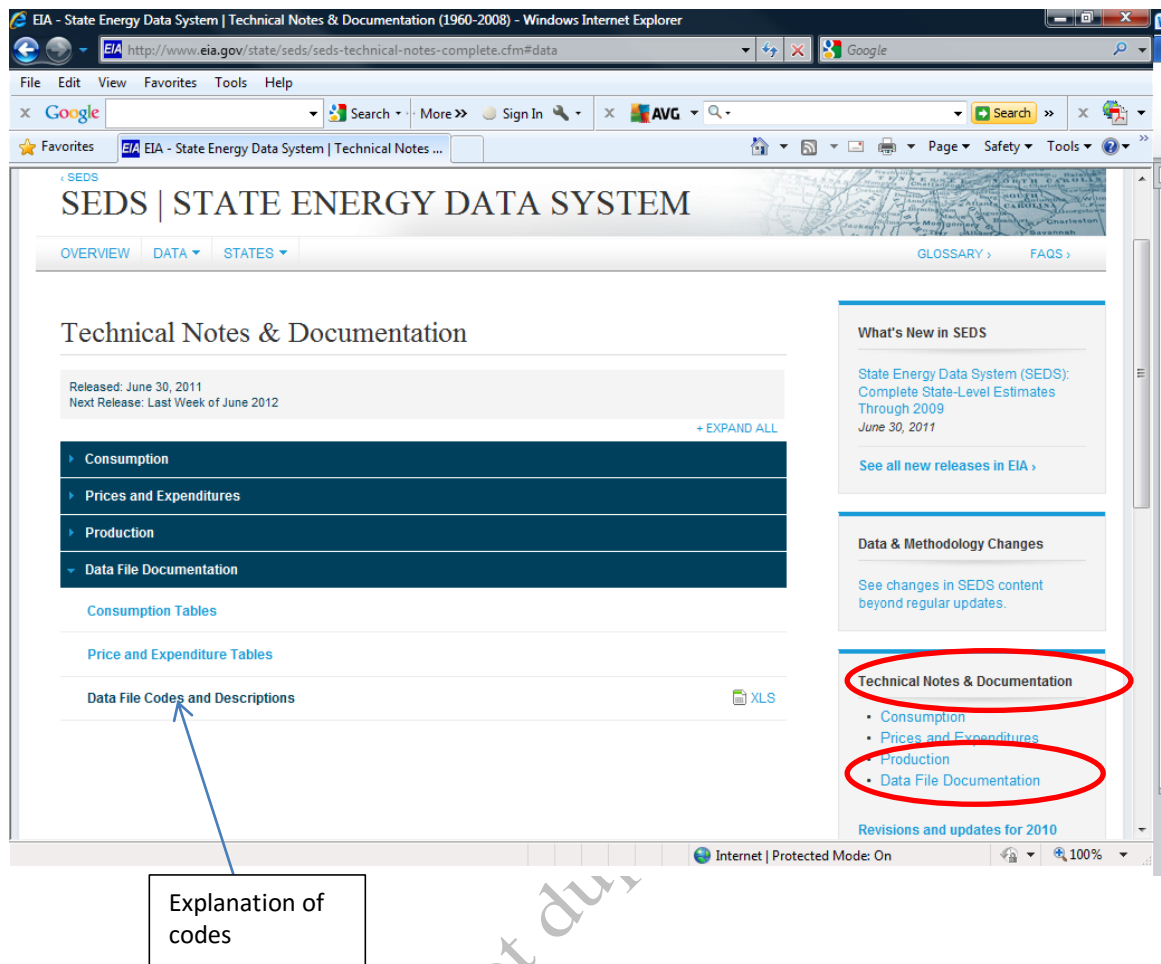



Figure 6.45. Link to download code definitions.

3. To the right of *Data File Codes and Descriptions*, click the Excel file button, .
4. Save the file to your computer or portable memory device.
5. Switch to Ch06DataStudentName. On the File tab, click Open. Navigate to the folder where you saved the file, *Data File Codes and Descriptions*.
6. Select the file and open it. Observe that the file is a workbook with three worksheets. On the Protection toolbar, click Enable editing. Click the *MSN Descriptions* sheet tab. The sheet has a table in which the codes are explained, as shown in Figure 6.46.

Internal structure of the codes

Units of measure for each code

**State Energy Data System
Data Series Names (MSN) and Descriptions**

The MSNs are five-character codes, most of which are structured as follows:
 1st & 2nd characters - describe an energy source (for example, NG for natural gas, MG for motor gasoline)
 3rd & 4th characters - describe an energy sector or an energy activity (for example, RC for residential consumption, PR for production)
 5th character - describes a type of data (for example, P for data in physical unit, B for data in billion Btu)

MSN	Description	Unit
ABICB	Aviation gasoline blending components consumed by the industrial sector.	Billion Btu
ABICP	Aviation gasoline blending components consumed by the industrial sector.	Thousand barrels
ARICB	Asphalt and road oil consumed by the industrial sector.	Billion Btu
ARICD	Asphalt and road oil price in the industrial sector.	Dollars per million Btu
ARICP	Asphalt and road oil consumed by the industrial sector.	Thousand barrels
ARICV	Asphalt and road oil expenditures in the industrial sector.	Million dollars
ARTCB	Asphalt and road oil total consumption.	Billion Btu
ARTCD	Asphalt and road oil average price, all sectors.	Dollars per million Btu
ARTCP	Asphalt and road oil total consumption.	Thousand barrels
ARTCV	Asphalt and road oil total expenditures.	Million dollars
AVACB	Aviation gasoline consumed by the transportation sector.	Billion Btu
AVACD	Aviation gasoline price in the transportation sector.	Dollars per million Btu
AVACP	Aviation gasoline consumed by the transportation sector.	Thousand barrels
AVACV	Aviation gasoline expenditures in the transportation sector.	Million dollars
AVTCB	Aviation gasoline total consumption.	Billion Btu
AVTCD	Aviation gasoline average price, all sectors.	Dollars per million Btu
AVTCP	Aviation gasoline total consumption.	Thousand barrels
AVTCV	Aviation gasoline total expenditures.	Million dollars
BMTCB	Biomass total consumption	Billion Btu
CCEXB	Coal coke exported from the United States.	Billion Btu
CCEXD	Coal coke exports average price, United States.	Dollars per million Btu
CCEXP	Coal coke exported from the United States.	Thousand short tons
CCEXV	Coal coke exports expenditures, United States.	Million dollars
CCIMB	Coal coke imported into the United States.	Billion Btu
CCIMD	Coal coke imports average price, United States.	Dollars per million Btu
CCIMP	Coal coke imported into the United States.	Thousand short tons

Sheet tabs

Codes used in column B of the Energy data table

Figure 6.46. Definition of codes.

- At the bottom of the screen, right-click the MSN Descriptions sheet tab.
- On the shortcut menu, click *Move or Copy*. The Move or Copy dialog box displays, as shown in Figure 6.47.

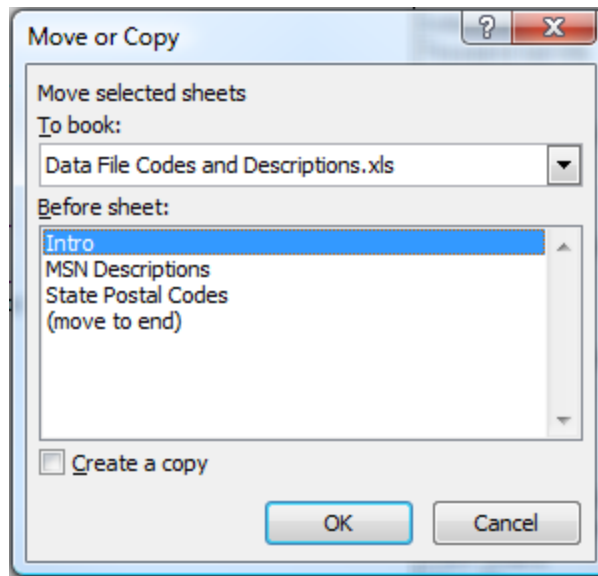


Figure 6.47. Move or Copy dialog box.

9. Under *To book*, at the right of the file name box, click the arrow. From the list of currently open Excel workbooks, click Ch06DataStudentName. A list of the worksheets in Ch06DataStudentName is displayed.
10. In the *Before sheet* box, click *Projects*, and then next to *Create a Copy* click the check box. Compare your settings with those shown in Figure 6.48.

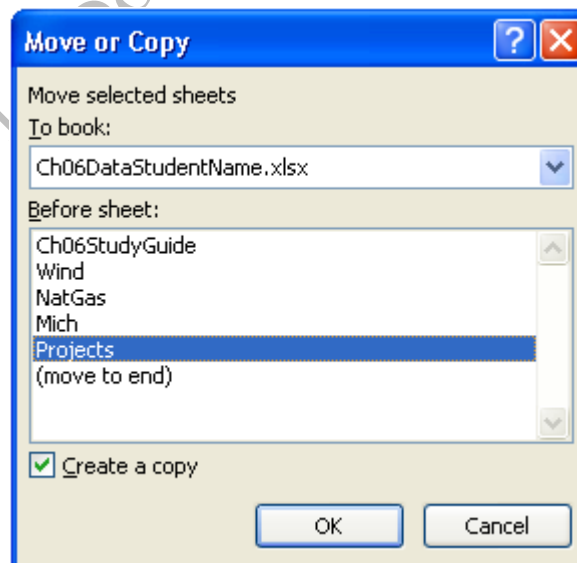


Figure 6.48. Settings for moving sheet into another workbook.

11. Click OK. The worksheet is copied into Ch06DataStudentName and placed after the Mich sheet, as shown in Figure 6.49.

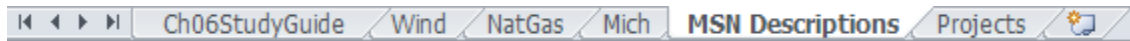


Figure 6.49. Worksheet copied from one workbook into another.

12. Switch back to the *Data File Codes and Descriptions* workbook, and then close it. Save the C06DataStudentName workbook.

Importing Data from an Access File

Microsoft Office is a **suite** of programs that are designed to work together by exchanging data easily between programs. Excel can import data directly from Access files. If the desired data is in a single table in Access, the table can be imported as a worksheet. If the desired data is distributed across several related tables, a **view** can be created by writing a query that specifies the fields and criteria. The view looks like a table but it is a temporary collection of data from the tables. Excel can import either tables or views into worksheets.

1. In Ch06DataStudentName, click the Projects sheet tab. Click cell A2.
2. On the Data tab, in the Get External Data group, click From Access.
3. Navigate to the folder where files for this book are stored, click Ch06Contractor.accdb, and then click the Open button. The Select a Table dialog box displays, as shown in Figure 6.50.

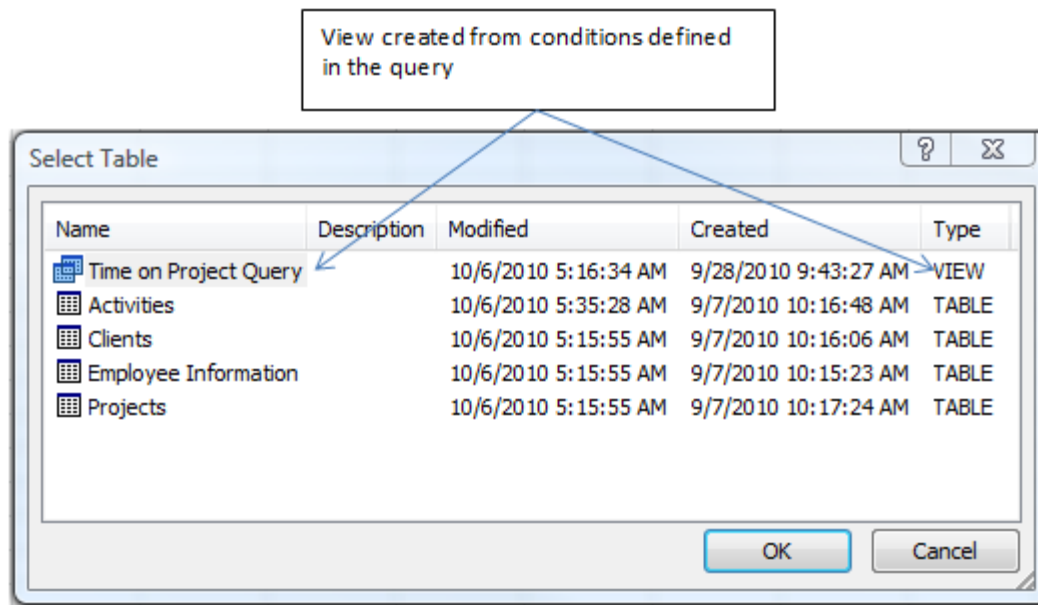


Figure 6.50. List of tables and views from an Access file.

4. Click *Time on Project Query* and then click OK. Confirm that the data will be brought in as a table in the existing worksheet beginning in cell A2, and then click OK.
5. The data is copied from the Access file into the *Projects* worksheet and formatted as an Excel table, as shown in Figure 6.51.

Data formatted as an Excel table

	A	B	C	D	E	F
1						
2	Last Name	Specialty	Hours Worked	Date	Project Name	Client Name
3	Krasny	Sales	3	4/1/2010	Warehouse insulation	General Mechanics
4	Krasny	Sales	2	4/1/2010	Heat reclaim	General Mechanics
5	Krasny	Sales	4	4/2/2010	PV Panels	General Mechanics
6	Baylis	Installer	4	4/2/2010	Warehouse insulation	General Mechanics
7	Krasny	Sales	4	4/2/2010	Heat reclaim	General Mechanics
8	Hill	Architect	4	4/5/2010	PV Panels	General Mechanics

Figure 6.51. Query in Access imported as an Excel table into the *Projects* worksheet.

Tip: An **Excel table** is similar to a table in Access but instead of using a separate query to filter and sort the data, an Excel table has a drop down menu arrow at the top of each column that can display a menu of options for filtering the rows displayed based on the content of the field in that column.

6. In cell A1 type **Hours Worked by Employee** and then save the workbook. Leave the workbook open for

the last part of the assignment.

Key Takeaways

- Data copied from the Web can be pasted into Excel using the Paste Special feature to control how the data is copied to retain or remove formatting and links. Adjust column widths to fully display the data. [6.3.1]
- Data that is formatted as a table on a Web page is identified by a small arrow in the corner of the table and can be imported into Excel using the Get External Data button. [6.3.2]
- Data on the Web or in other files may be stored as text files that can be downloaded as a comma separated values file (CSV) where a specific character is used to separate fields. Use the Text Import Wizard to identify the field (columns) and the character that is used to separate one field from the next. Files may also be imported as an Excel file. [6.3.3]
- Worksheets can be copied or moved between workbooks or within workbooks by right-clicking a sheet name and choosing the Move or Copy command. [6.3.4]
- Tables or queries can be imported into Excel from Access. When data is imported from Access it is formatted as an Excel table with filter arrows and table formatting.[6.3.5]

4 Simplifying Tables for Charting

Learning Objectives

1. Sort data to remove gaps and empty records. [6.4.1]
2. Create charts using non-adjacent rows or columns. [6.4.2]

Tables of data often include empty rows to make them easier to read on a website but those gaps interfere with charting and should be removed. In this section, a table of data is sorted to remove empty rows and then you learn how to select non-adjacent columns for charting.

1. In Ch06DataStudentName, click the NatGas sheet tab. Observe that blank rows were inserted every five years to make the table easier to read and that NA was placed in cells where no data had been collected.
2. Drag the row headings from 3 through 12 to select them. Right-click on one of the rows and on the shortcut menu click Delete. The rows with no data are deleted and the remaining data moves up.

- Click cell A2. Scroll down and locate the bottom row. Press and hold the Shift key and then click cell M38. The table of data is selected without the title in the first row.
- On the Data tab, click the Sort button. In the Sort dialog box, confirm that there is a check in the box next to *My data has headers* to indicate that the first row of the selected range should not be moved.
- Next to the Sort by box, click the arrow, and then click Year. In the Order box confirm that *Smallest to Largest* is displayed. Click OK. The data is sorted by year without moving the labels in row 2, as shown in Figure 6.52. Observe that by sorting the table by the year, the empty rows were moved to the bottom. Recall that data is easier to chart if it is arranged in compact simple tables without spaces.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	U.S. Residential Price of Natural Gas (\$ per thousand cubic feet)												
2	Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
3	1981	3.94	3.99	4.06	4.11	4.29	4.3	4.32	4.3	4.47	4.5	4.53	4.55
4	1982	4.65	4.69	4.78	4.86	5.17	5.2	5.23	5.23	5.41	5.66	5.68	5.74
5	1983	5.86	5.87	6	6.06	6.22	6.2	6.21	6.18	6.19	6.7	6.3	5.94
6	1984	5.78	5.84	5.92	5.96	6.27	6.76	7.11	7.23	7.17	6.8	6.31	6.05
7	1985	5.97	5.86	5.99	6.11	6.59	6.96	7.07	7.21	7.06	6.5	6.13	5.7
8	1986	5.63	5.67	5.71	5.89	6.18	6.67	6.84	6.94	6.83	6.38	5.66	5.28
9	1987	5.3	5.34	5.36	5.46	5.98	6.55	6.78	6.84	6.64	5.85	5.42	5.13
10	1988	5.08	5.09	5.18	5.35	5.87	6.5	6.74	6.92	6.79	5.95	5.56	5.39

Figure 6.52. Data sorted to remove empty rows.

Tip: The sorting program in Excel places null or empty cells at the bottom of sorted lists.

- To create a simple chart that shows the trend of prices from 1981 to 2010 only one month needs to be charted. Select cells A2:A32 to select the years.
- Press and hold the CTRL key and then select cells H2:H32. The control key is used to select a second range of cells that is not adjacent to the first range. Both ranges have to be the same length, as shown in Figure 6.53.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	U.S. Residential Price of Natural Gas (\$ per thousand cubic feet)												
2	Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
3	1981	3.94	3.99	4.06	4.11	4.29	4.3	4.32	4.3	4.47	4.5	4.53	4.55
4	1982	4.65	4.69	4.78	4.86	5.17	5.2	5.23	5.23	5.41	5.66	5.68	5.74
5	1983	5.86	5.87	6	6.06	6.22	6.2	6.21	6.18	6.19	6.7	6.3	5.94
6	1984	5.78	5.84	5.92	5.96	6.27	6.76	7.11	7.23	7.17	6.8	6.31	6.05
7	1985	5.97	5.86	5.99	6.11	6.59	6.96	7.07	7.21	7.06	6.5	6.13	5.7
8	1986	5.63	5.67	5.71	5.89	6.18	6.67	6.84	6.94	6.83	6.38	5.66	5.28
9	1987	5.3	5.34	5.36	5.46	5.98	6.55	6.78	6.84	6.64	5.85	5.42	5.13
10	1988	5.08	5.09	5.18	5.35	5.87	6.5	6.74	6.92	6.79	5.95	5.56	5.39
11	1989	5.41	5.38	5.45	5.54	5.93	6.58	6.92	7.07	6.8	6.06	5.56	5.3
12	1990	5.43	5.65	5.6	5.64	6	6.56	7.04	7.08	6.9	6.14	5.69	5.62
13	1991	5.54	5.56	5.6	5.9	6.28	6.97	7.23	7.36	6.92	6.2	5.51	5.51
14	1992	5.53	5.54	5.5	5.62	6.15	6.84	7.27	7.45	7.15	6.52	6.02	5.74
15	1993	5.73	5.73	5.67	6.02	6.78	7.37	7.86	8.13	7.75	6.79	6.17	6.07
16	1994	5.93	6.04	6.3	6.6	6.84	7.66	8.1	8.22	7.84	6.86	6.27	6.06
17	1995	5.85	5.76	5.84	6.06	6.54	7.49	7.82	8.13	7.73	6.62	5.61	5.54
18	1996	5.64	5.82	5.93	6.27	6.84	7.83	8.64	8.73	7.99	7.05	6.37	6.47
19	1997	6.74	6.79	6.52	6.53	6.83	8.3	8.78	8.99	8.84	7.69	6.86	6.54
20	1998	6.41	6.41	6.29	6.81	7.7	8.51	8.53	9.25	8.96	7.6	6.58	6.34
21	1999	6	6.29	6.06	6.44	7.3	8.2	8.83	9.14	8.63	7.56	7.15	6.51
22	2000	6.37	6.54	6.91	7.19	8.26	9.5	10.32	10.37	10.1	9.44	8.58	8.56
23	2001	10.12	10.26	9.85	10.16	11.14	11.58	11.22	10.89	10.17	8.24	7.98	7.3
24	2002	7.38	7.23	7.1	7.66	8.54	9.58	10.31	10.44	10.23	8.61	7.99	7.87
25	2003	8.18	8.58	9.77	10.18	10.79	12.08	12.75	12.84	12.31	10.64	9.77	9.51
26	2004	9.71	9.85	10.03	10.54	11.63	13.08	13.54	13.74	13.31	11.69	11.44	11.09
27	2005	10.9	10.87	10.84	11.88	12.74	13.79	14.86	15.51	16.56	16.44	15.64	14.6
28	2006	14.92	13.98	13.17	13.27	14.41	15.07	15.72	16.18	15.71	12.51	12.45	12.53
29	2007	12.17	12.13	12.81	13.31	14.69	16.28	16.71	16.71	16.03	14.57	13.04	12.34
30	2008	12.24	12.58	13.13	14.49	16.31	18.82	20.68	20.08	18.36	15.49	13.82	12.84
31	2009	12.38	12.14	11.81	11.55	12.49	13.83	14.82	15.13	14.34	11.62	11.27	10.3
32	2010	10.48	10.58	10.86	11.78	12.73	14.25	15.45					

Figure 6.53. Ranges selected for charting that are not adjacent.

8. On the Insert tab, in the Charts group, click the Column button and then click the 2-D Clustered Column option. The program attempts to chart both ranges, as shown in Figure 6.54.

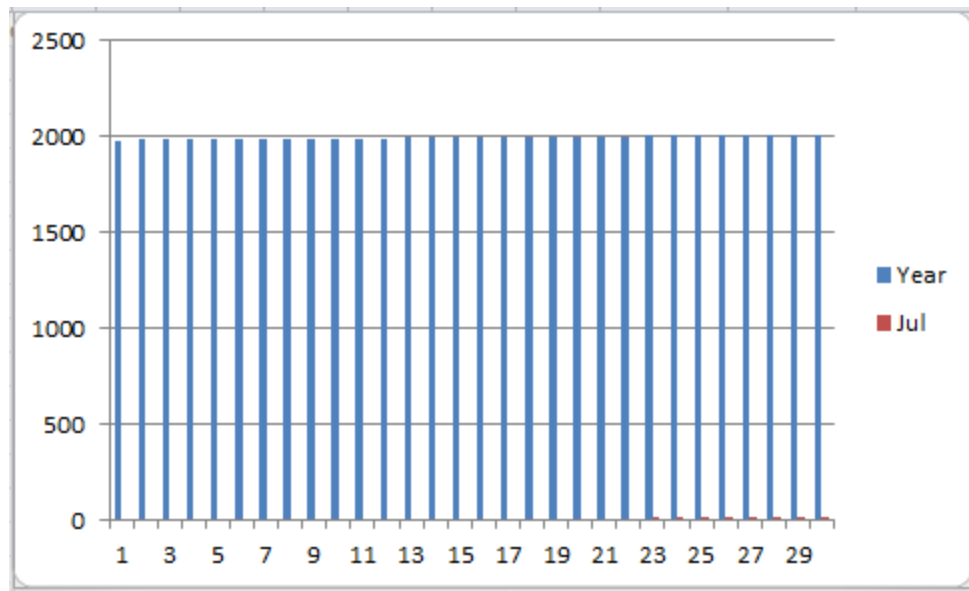


Figure 6.54. The ranges are charted as if they were both dependent.

9. On the Chart Tools/Design tab, in the Data group, click the Select Data button.
10. In the *Select Data Source* dialog box, under *Legend Entries* click Year and then click the Remove button.
11. Under *Horizontal (Category) Axis Labels* click the Edit button.
12. With the Axis Labels dialog box displayed, drag cells A3:A32 and then click OK. In the *Select Data Source* dialog box, click OK.
13. Use the charting skills you have practiced to remove the legend, add the label \$ per thousand cubic feet to the vertical axis and Natural Gas Prices in July as a chart title, as shown in Figure 6.55.

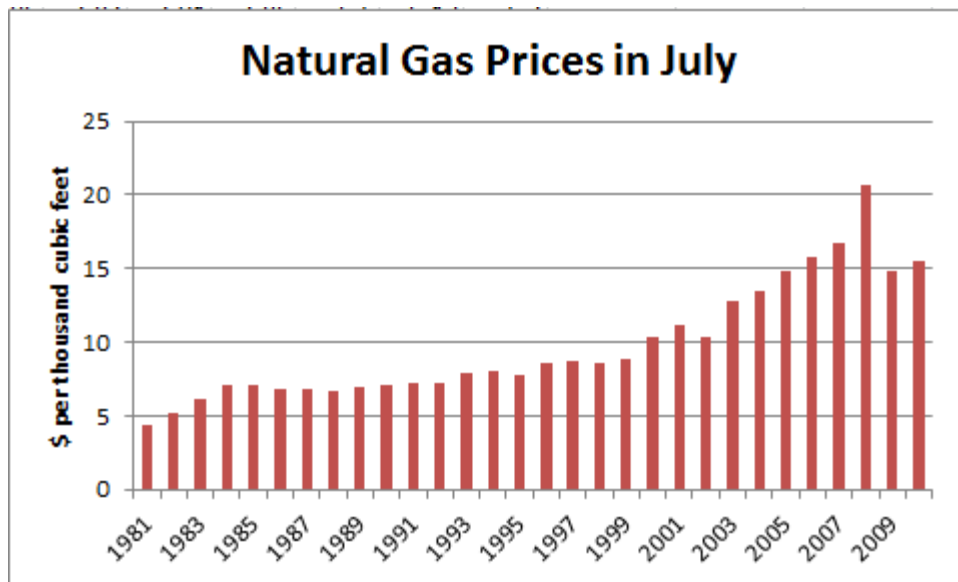


Figure 6.55. Chart of prices for one month.

14. Observe that it is now easier to see the trend in price and to spot the anomaly in 2008 which turns this data into information. Move the chart to the left of the data beginning in cell N2.
15. Save the workbook, and then close it.

Key Takeaways

- To move empty rows to the bottom of the table, sort the table in ascending order: [6.4.1]
- To select non-adjacent rows or columns for charting, select the first range of cells, hold the CTRL key, and then select the other ranges. The ranges must be parallel and the same length [6.4.2]

Key Terms

ASCII

American Standard Code for Information Interchange refers to a code that represents numbers and letters as two-digit hexadecimal numbers.

AND

Logical function that requires that two criteria are both true.

Base

Number used to determine the positional value of a numeral.

Base 2 (binary)

Numbering system that uses two numerals to represent numbers from zero to one.

Base 10 (decimal)

Numbering system that uses ten numerals to represent numbers from zero to nine.

Base 16 (hexadecimal)

Numbering system that uses sixteen numerals to represent numbers from zero to fifteen. The letters A through F are used as numerals for numbers ten through fifteen.

Binary

Base 2 numbering system that uses two numerals, 0 and 1.

Bit

One numeral of a binary system.

Carriage Return (CR)

Character in the ASCII and UTF-8 code that was initially used to force the carriage of an early typing device to move to the beginning of the next line of text.

Comma separated values (csv)

Method of separating fields with commas used when transferring data between computers.

Criterion

Condition that must be met to qualify. Plural is criteria.

Data

Facts such as numbers, dates, or names.

Database management system (DBMS)

Collection of programs that allow users to modify and recall data from storage.

Delimiter

Character used to separate fields in linear string of text that contains the data from a table.

Excel table

Range of cells with menu options for each column.

Fields

Columns in a data table that represent categories of data.

Fixed width

Fields are defined by their position in the string of text that represents the data from a table.

Information

Data that have been given meaning.

IS NOT NULL

Function that qualifies a record if the field has some content.

IS NULL

Function that qualifies a record if the field is empty.

Key field

Column in a data table that has a unique value for each row.

Line Feed (LF)

Character in the ASCII and UTF-8 code that indicates the end of a line of text.

NOT

Function that qualifies all records that do not match a criterion.

Numeral

Symbol used to represent a number.

OR

Logical function that requires that at least one of two criteria is true.

Portable Document Format (PDF)

Method of displaying documents with original formatting that may be viewed but not easily edited.

Query

Structured request to the database management system that can be used to retrieve records that match the criteria provided.

Query by Example (QBE)

User interface in Access that uses a grid to associate fields with criteria that are then translated into SQL commands.

Reader

Program that displays PDF files.

Records

Rows in a data table that contain facts about one person, place, event or other individual item.

Relational database

Collection of data tables that can be joined using key fields.

Structured Query Language (SQL)

Computer language used to request records from a database management system.

Suite

Collection of programs that are designed to transfer data between them easily

Unicode Transformation Format (UTF)

International standard for coding.

UTF-8

International standard for coding that is similar to the old ASCII code.

View

Collection of records and fields from several tables whose selection is based on a set of criteria.

Last edited on May 30, 2012