

Chapter 3

Using Charts for Quantitative Analysis

A **chart** is a visual representation of a group of related numbers using lines, bars, columns or other shapes.

The charting process begins by organizing the data in a spreadsheet in such a way that it has meaning and its categories of data can be easily recognized by the program.

Why Do I Need to Know This?

Quantitative reasoning is used to help people make decisions. Those decisions can be based on recognizing the need for change based on a trend that a chart makes clear. Another chart might make it apparent where to begin when addressing a problem by identifying which portion is the largest. You may need to decide where to cut a budget, when to take a risk on the stock market, or which fund to invest in for your retirement. Charts will help if you know how to read them. You may need to create charts to communicate quantitative data to other people at a meeting that could range in importance from a local community fundraiser to a critical presentation at your job. Whether using charts created by others or making your own, charts are a great way to represent the relative size of numbers and perceive trends.

1 Organizing Data for Charting

Learning Objectives

1. Identify principles of data organization for charting purposes. [3.1.1]
2. Convert measurements to the same units for comparison. [3.1.2]
3. Use relative and absolute cell references, and then fill formulas in tables. [3.1.3]
4. Fill the SUM function to calculate the sum of multiple columns. [3.1.4]

Excel has extensive charting capability. Before these features can be used effectively, the data must be organized with charting in mind. When ranges of data are selected to chart, the ranges must be simple

rectangles. Use the following principles to organize data in a worksheet so the program can represent it with a chart:

- Numbers that make up a group should be in a row or column without spaces between them
- Labels should be in the same row or column as the data
- If you compare two or more ranges of data, they should be arranged in either rows or columns of the same length.

For example, a family decided to start tracking the number of miles they were driving and recorded the data in the following table:

	A	B	C
1	<i>Log of Transportation Miles</i>		
2	<i>Type</i>	<i>Miles</i>	<i>Totals</i>
3	Vacation		
4	Cabin	250	
5	Grand parents	650	
6	Mountains	325	
7		Total	1,225
8	Work		
9	Dad	220	
10	Mom	340	
11	Pizza delivery	830	
12		Total	1,390
13	Other		
14	Local errands	145	
15	Visiting Friends	65	
16		Total	210

Figure 3.1. Table of data.

If you want to make a chart that compared the three types of vacation mileage, you could select them as a group and create a chart as shown in Figure 3.2.

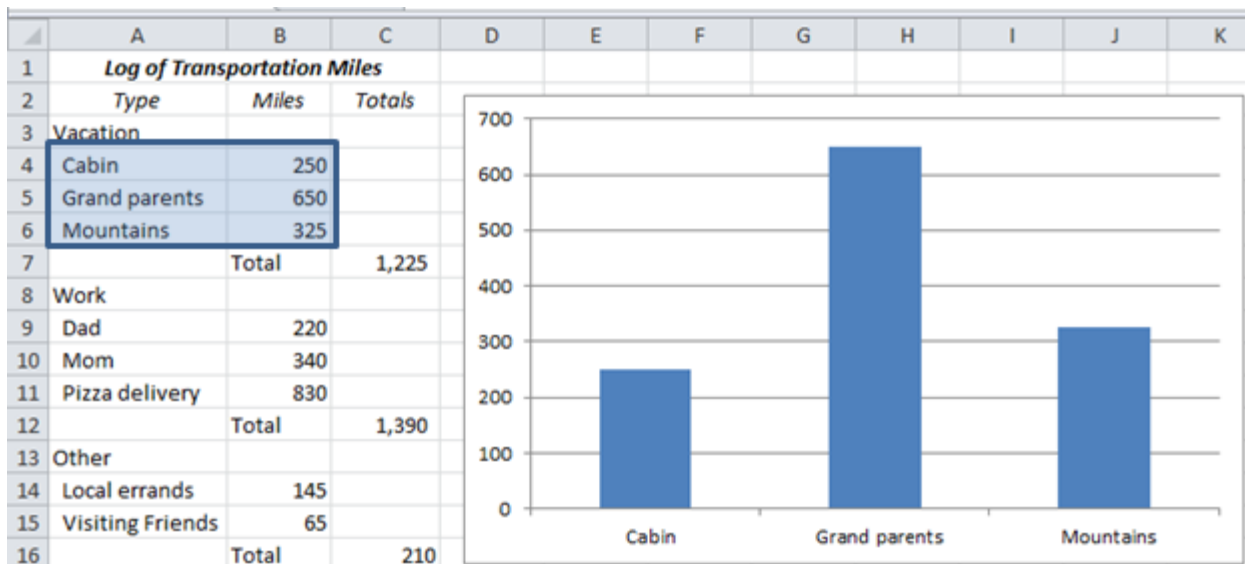


Figure 3.2. Chart of three items.

However, if you wanted to chart the totals of each category it would be difficult because the names of the category types; Vacation, Work, and Other; are not next to their respective totals. If you try to select each cell individually, the program thinks you have six separate categories to chart and you get a chart like the one shown in Figure 3.3.

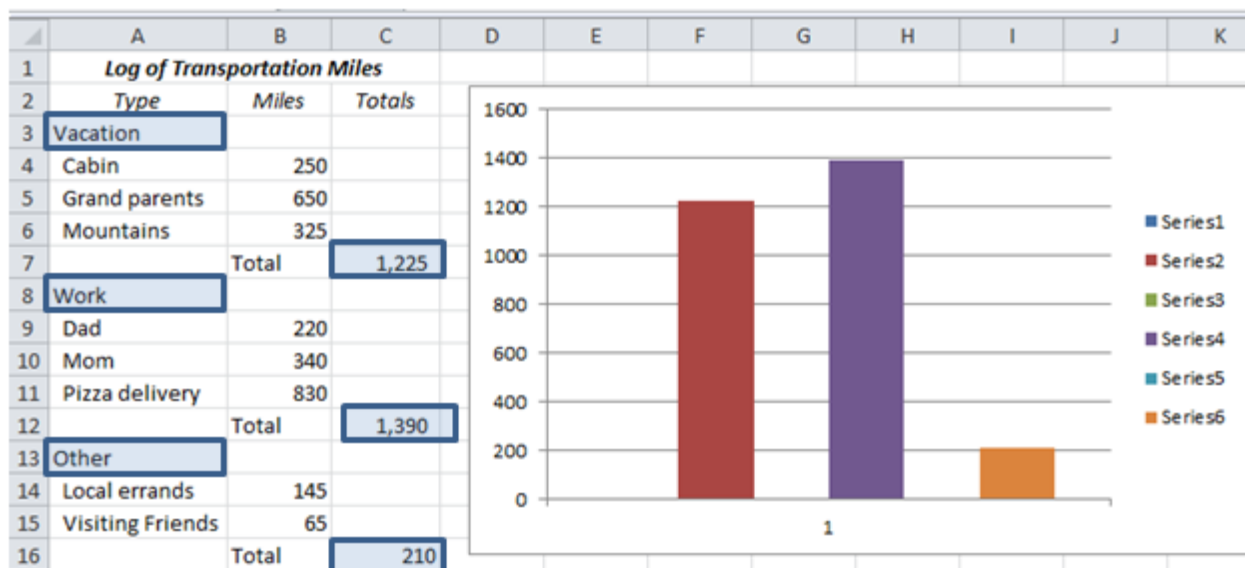


Figure 3.3. Chart of three totals that are not positioned conveniently.

An improvement in the design would be to move the totals for each category next to the category

label, as shown in Figure 3.4.

	A	B	C
1	Log of Transportation Miles		
2	Type	Miles	Totals
3	Vacation		1,225
4	Cabin	250	
5	Grand parents	650	
6	Mountains	325	
7	Work		1,390
8	Dad	220	
9	Mom	340	
10	Pizza delivery	830	
11	Other		210
12	Local errands	145	
13	Visiting Friends	65	

Figure 3.4. Totals in the same row at the label.

The problem with this design is the empty space between the category name and the total. If you select these three rows and try to chart the totals in this layout, you get the chart with columns of zero height as shown in Figure 3.5.

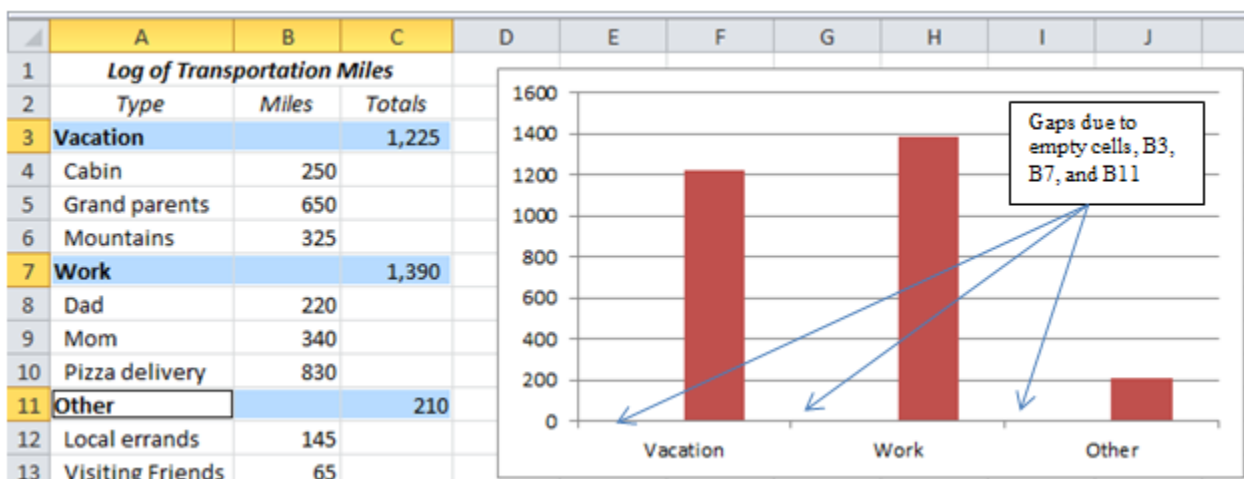


Figure 3.5. Empty cells in the ranges to be charted produce columns of zero height

One way to get around this problem is to collect the data into another area of the worksheet or workbook where it is arranged for convenient charting. Instead of simply typing the data into those cells, it's

a good idea to use formulas that refer to the original data so that if you change the original figures, the chart changes with it. In Figure 3.6, locate cell B17 and observe that it is a simple formula that references cell C11. Similar formulas were entered in cells B15 and B16 to reference the total miles for vacation and work.

	A	B	C
1	Log of Transportation Miles		
2	<i>Type</i>	<i>Miles</i>	<i>Totals</i>
3	Vacation		1,225
4	Cabin	250	
5	Grand parents	650	
6	Mountains	325	
7	Work		1,390
8	Dad	220	
9	Mom	340	
10	Pizza delivery	830	
11	Other		210
12	Local errands	145	
13	Visiting Friends	65	
14			
15	Vacation	1,225	
16	Work	1,390	
17	Other	=C11	

Figure 3.6. Simple formulas can be used to transfer values to a simple table for charting

The most flexible layout for charting is a very simple database style where categories have their own column, even though it looks redundant, as shown in Figure 3.7.

	A	B	C
1	Category	Item	Miles
2	Vacation	Cabin	250
3	Vacation	Grand parents	650
4	Vacation	Mountains	325
5	Work	Dad	220
6	Work	Mom	340
7	Work	Pizza delivery job	830
8	Other	Local errands	145
9	Other	Visiting friends	65

Figure 3.7. Row for each item with same types of data in columns

Another advantage of this type of design is that if you select two levels of labels with the number, both can display in certain types of charts like the column chart, as shown in Figure 3.8.

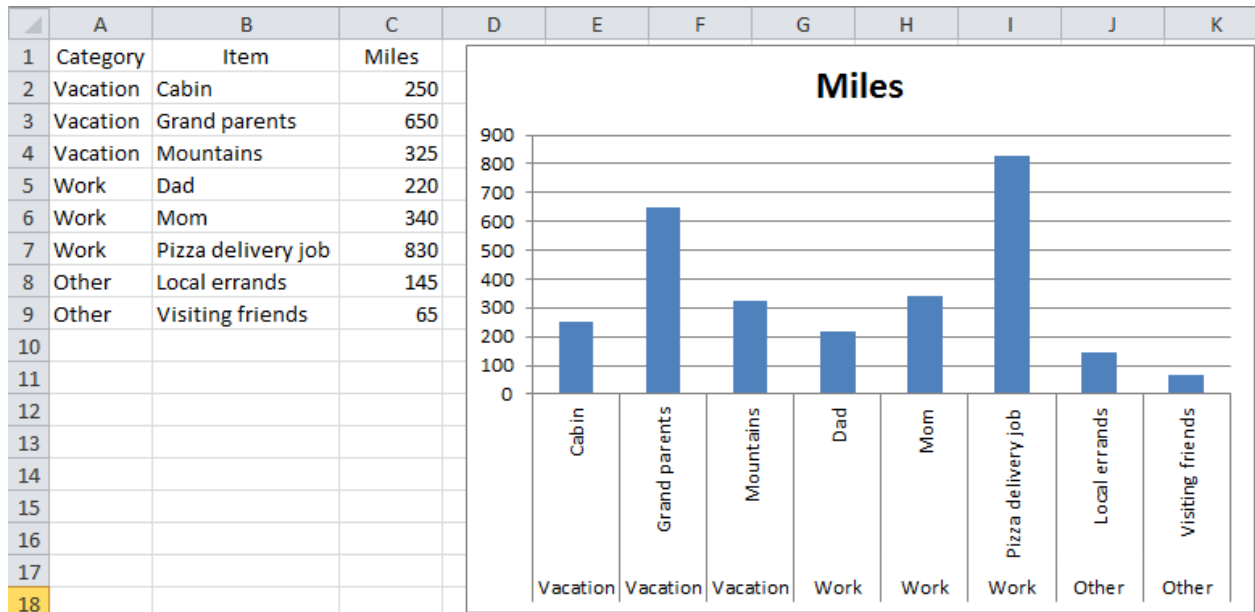


Figure 3.8. Two levels of data labels

This type of layout lends itself to using subtotals which you will learn in a later unit.

In this example, the layout shown in Figure 3.8 is easier to chart but the layout in Figure 3.1 looks better. You can have both, by creating a worksheet that looks good for printing and another that works better for charting. Use formulas in one of the two worksheets that refer to the corresponding cell in the other sheet so you can change the data in one place and it will be updated in the other worksheet.

Charting Home Energy Use

It is much easier to make comparisons between rows of data if the data is represented in a chart. In some cases, the data must be converted to the same units to make the comparison meaningful. The resulting data should still be arranged in a simple table so that it can be charted conveniently.

1. Start Excel. Navigate to the location of files for this class, and then open Ch03Charts.xlsx and then save the file as Ch03ChartsStudentName using your own name in place of StudentName.
2. At the bottom of the workbook, click the UtilityBills sheet tab, if necessary to select it. Examine the table

of data in rows 6 through 10, as shown in Figure 3.9.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Utility Bills												
2	Unit of Fuel	Units	Conversion Factor										
3	CCF of Natural Gas	1 CCF				KWH per CCF							
4	Facecord of wood	BTU				KWH per face cord							
5													
6		Dec-08	Jan-09	Feb-09	Mar-09	Apr-09	May-09	Jun-09	Jul-09	Aug-09	Sep-09	Oct-09	Nov-09
7	Electricity (kWh)	1,461	922	1,105	951	903	647	706	803	689	704	1,042	1,035
8	Hot Water (kWh)	308	205	274	266	395	153	281	295	238	235	313	244
9	Natural Gas (CCF)	278	547	35	172	115	22	5	1	13	10	134	141
10	Wood (face Cords)			5									

Figure 3.9. Table of utility bills

- Observe that the data is arranged in a simple table to make it convenient to chart. Observe that the labels in row 6 are months and the intervals are the same. Observe that four types of energy are represented using three different units of measure; kWh, CCF, and face cords.

An underlying assumption that is common to most chart types is that the units of measure are the same for each type of data. Before you chart the data, use the skills you practiced previously to convert them to the same units. In this instance, kWh will be used.

- Use the skills you practiced previously to locate the energy conversion factor for Hundreds of Cubic Feet of natural gas (CCF) to kilowatt-hours, and then paste it into cell D3. Remove the spaces and adjust the display to show one decimal place. Add a comment to indicate the source of the conversion factor.
- In cell E3, type `=B3*D3` to calculate the KWH in a CCF of natural gas.
- Use the skills you practiced previously to search the Internet and find the energy content for a face cord of wood measured in BTUs. Type a suitable factor into cell B4. Adjust the display to show zero decimal places and use commas as separators. Because the heating value of firewood is given as a range, add a comment to B4 in which you explain why you chose the particular value used in this cell and provide the Web address for the source.
- Use the skills you practiced previously to find the conversion factor for BTU to KHW. Paste the conversion factor into cell D4. Add a comment and provide the Web address of the source.
- Use the skills you practiced previously to fill the formula in E3 downward into cell E4. Format E3 and E4

to display the numbers with commas and one decimal place.

9. On the UtilityBills sheet, click cell B17. Type **=B9*E3** and then press the F4 key on the keyboard.


Observe that the F4 key adds the dollar signs automatically to the adjacent cell reference to make it an absolute reference. Alternatively, you could type the dollar signs. The formula uses a relative reference to the cell that is seven rows above (B9) and an absolute reference to cell E3, as shown in Figure 3.10.

	A	B	C	D	E
1					
2	Unit of Fuel		Units	Conversion Factor	
3	CCF of Natural Gas	1	CCF	30.2	30.2
4	Facecord of wood	7,000,000	BTU	0.000293071	2,051.5
5					
6		Dec-08	Jan-09	Feb-09	Mar-09
7	Electricity (kWh)	1,461	922	1,105	951
8	Hot Water (kWh)	308	205	274	266
9	Natural Gas (CCF)	278	547	35	172
10	Wood (face Cords)			5	
11					
12					
13		Dec-08	Jan-09	Feb-09	Mar-09
14	Electricity	1,461	922	1,105	951
15	Hot Water	308	205	274	266
16	Natural Gas	=B9*\$E\$3			

Relative
reference

Absolute
reference

Figure 3.10. Relative and absolute reference

10. On the Formula bar, click the Enter button, . Using the Enter button has the advantage of leaving the selection on the cell.
11. With cell B16 selected, drag the fill handle to the right to cell M19 to fill the formula into cells C16:M16.
While cells B16:M16 are still selected, format the cells to display zero decimal places. Each cell in row 16 is the product of the value in the corresponding cell in row 9 and the conversion factor in E3.

12. In cell B17, enter a formula that multiplies the value in B10 by the value in cell E4. Make the reference to cell E4 absolute, and then fill the formula across to cell M17. Format the cells to display zero decimal places.
13. Click cell B18. On the Home tab, in the Editing group, click the SUM button, Σ . Confirm that the range is B14:B17, and then click the Sum button a second time. Fill this formula to the right to cells C18:M18.
14. While B18:M18 is selected, apply a border with a single top line and a double bottom line. Compare your worksheet to Figure 3.11.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Utility Bills												
2	Unit of Fuel	Units	Conversion Factor										
3	CCF of Natural Gas	1 CCF	30.2	30.2	KWH per CCF								
4	Facecord of wood	7,000,000 BTU	0.000293071	2,051.5	KWH per face cord								
5													
6		Dec-08	Jan-09	Feb-09	Mar-09	Apr-09	May-09	Jun-09	Jul-09	Aug-09	Sep-09	Oct-09	Nov-09
7	Electricity (kWh)	1,461	922	1,105	951	903	647	706	803	689	704	1,042	1,035
8	Hot Water (kWh)	308	205	274	266	395	153	281	295	238	235	313	244
9	Natural Gas (CCF)	278	547	35	172	115	22	5	1	13	10	134	141
10	Wood (face Cords)			5									
11													
12	Utility Bills (KWH)												
13		Dec-08	Jan-09	Feb-09	Mar-09	Apr-09	May-09	Jun-09	Jul-09	Aug-09	Sep-09	Oct-09	Nov-09
14	Electricity	1,461	922	1,105	951	903	647	706	803	689	704	1,042	1,035
15	Hot Water	308	205	274	266	395	153	281	295	238	235	313	244
16	Natural Gas	8396	16519	1057	5194	3473	664	151	30	393	302	4047	4258
17	Wood	0	0	10257	0	0	0	0	0	0	0	0	0
18	Total	10,165	17,646	12,693	6,411	4,771	1,464	1,138	1,128	1,320	1,241	5,402	5,537

Figure 3.11. Data arranged in a basic table that uses the same units of measure

Key Takeaways

- For charting purposes, a simple table works best with labels in one row or column and the numbers next to them.[3.1.1]
- Multiply units of measure by appropriate conversion factors so they are all the same unit of measure. Document the source of the conversion factor and any assumptions using comments. [3.1.2]
- Design the formulas using relative and absolute cell references so they can be copied or filled to make it faster to complete the table and to reduce errors that might be caused by typing the same formula repeatedly. [3.1.3]

- Click the sigma that denotes the SUM function in the Editing group on the Home tab to sum the cells in the table. Confirm that the marquee encloses the intended range and then click the sigma again. Fill the function across the table using the fill handle. [3.1.4]

2 Choosing the Appropriate Chart

Learning Objectives

1. Identify chart types and their strengths. [3.2.1]
2. Identify assumptions associated with using each chart type. [3.2.2]
3. Identify best practices for designing charts. [3.2.3]

Charts can be used to change data into information. For example, a sequence of numbers can produce a line chart that is rising to convey that the subject of the chart is increasing with the passage of time. Each type of chart uses assumptions that are shared with the viewers. In this section, the most popular types of charts are associated with their uses and underlying assumptions.

Types of Charts

The most common charts are line, column, and pie.

Line Chart

A **line chart** is used to show a relationship between pairs of data where one number is a measure of an independent variable like time and the other is a dependent variable such as distance or population. The assumptions underlying a line chart are:

- The values along the horizontal axis increase from left to right
- The intervals between the values on the horizontal axis are equal

Scatter Chart

A scatter chart like the one shown in Figure 3.12 is similar to a line chart. It is used to show pairs of data where one number is a measure of an independent variable like time and the other is a dependent variable

such as distance or population. Lines can be added that connect the data points or trend lines can be added to compare the data to functions that can be written as formulas. The assumptions underlying a scatter chart are the same as a line chart.

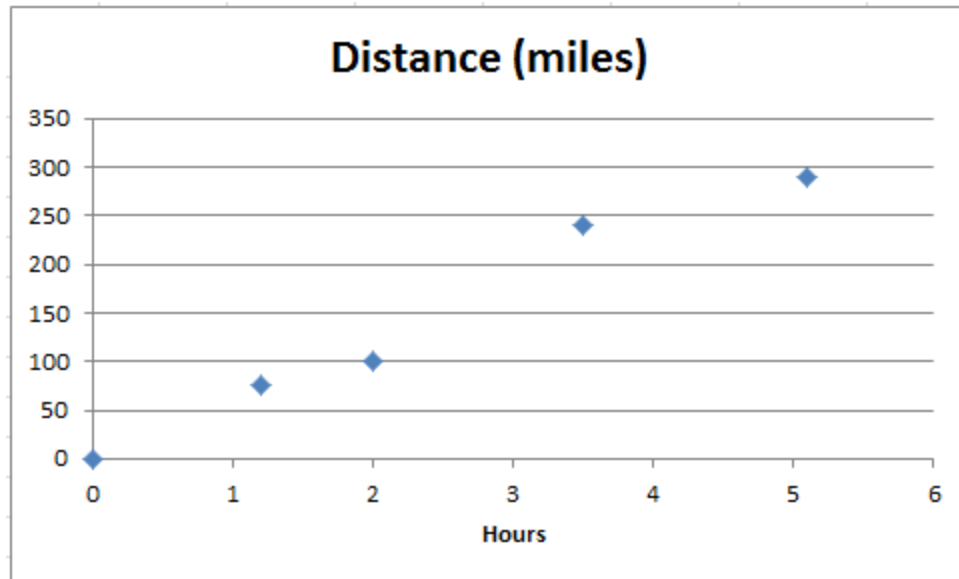


Figure 3.12. Scatter chart

Column Chart

A **column chart** like the one shown in Figure 3.13 displays a number that is associated with a category. The category can be a number like a dress size or a label. The height of the column is measured on the vertical axis. The assumption underlying a column chart is that all of the columns represent values that use the same measurement units

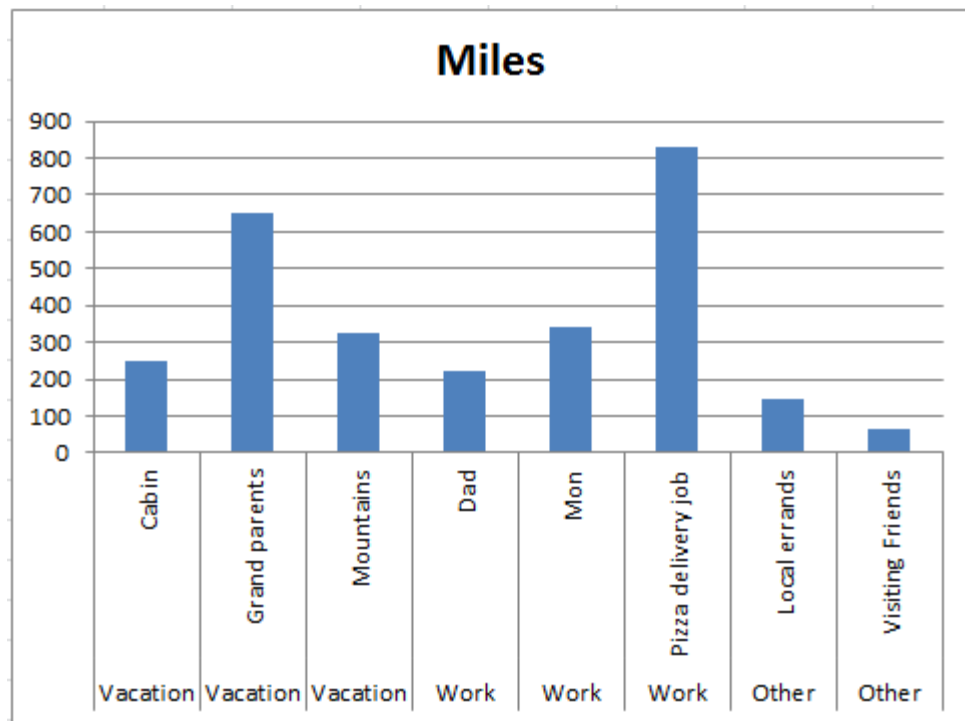


Figure 3.13. Column chart

Column charts can also be used like line charts to show a trend and a relationship between a dependent variable and an independent variable, as shown in Figure 3.14. If the horizontal axis is divided into units of time, they must use equal intervals and they must be in chronological order from left to right.

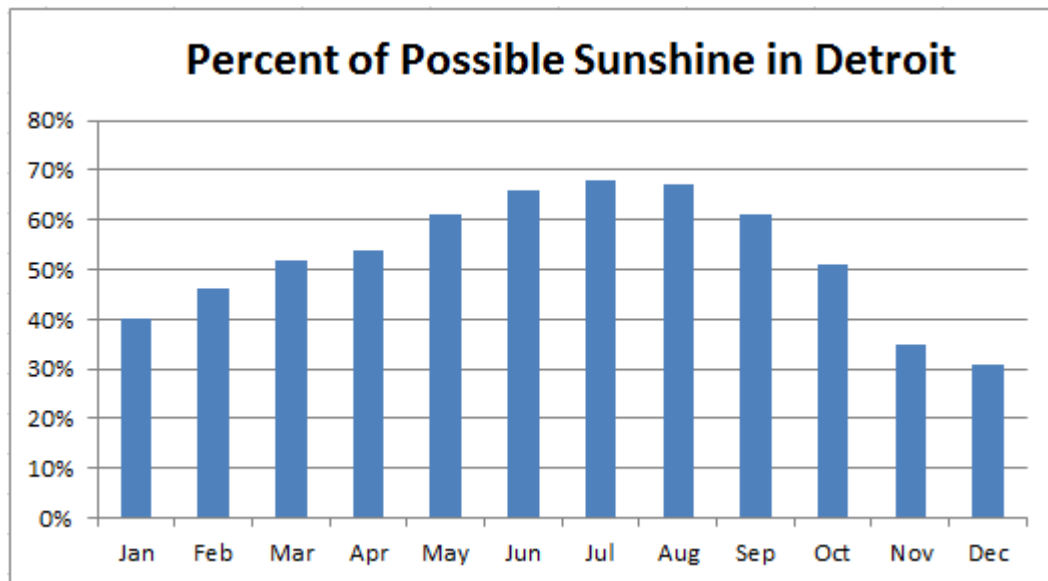


Figure 3.14. Column chart

Bar Chart

A **bar chart** like the one shown in Figure 3.15 displays a number that is associated with a category. The category can be a number like a grade of gasoline or a label. The length of the column is measured on the horizontal axis. The assumption underlying a column chart is that all of the bars represent values that use the same measurement units and the categories are related to the topic of the chart.

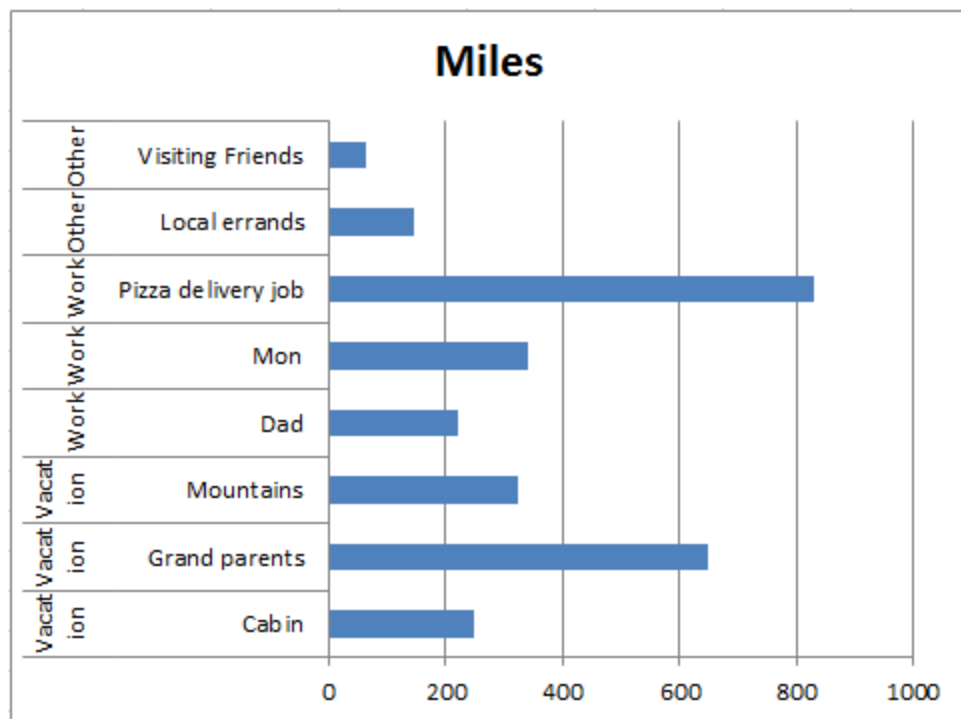


Figure 3.15. Bar chart

Pie Chart

A **pie chart** displays how each part contributes to the whole, as shown in Figure 3.16. The assumptions underlying a pie chart are:

- The numbers are part of a group
- All of the numbers in the group are included and they add up to 100% of the group

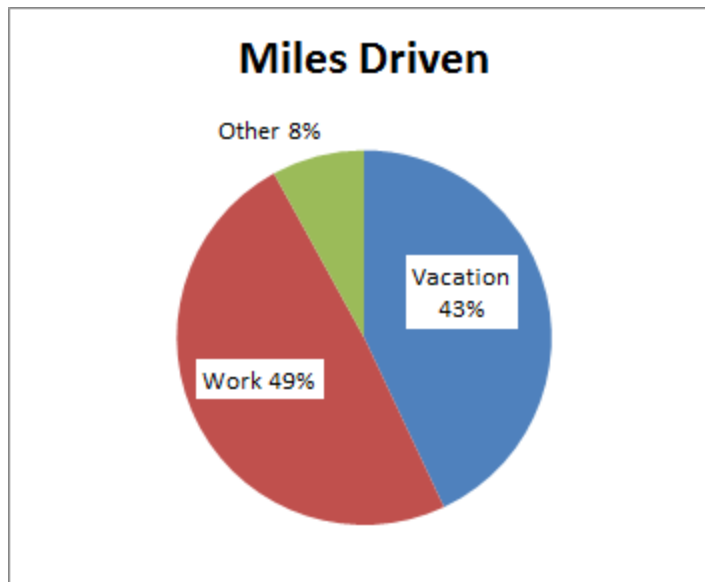


Figure 3.16. Pie chart

Labeling Charts

The purpose of a chart is to make it easier to analyze the data upon which it is based. Line charts help identify trends, column and bar charts make comparisons easier, and pie charts show how each part contributes to the whole.

Other tips for effective charting are:

- Label the horizontal and vertical axis to explain what the number represent
- Always show time units on the bottom axis of the chart, increasing from left to right
- If the labels on the bottom axis are numbers that show an increase from left to right, the interval between each one should be the same.
- Use simple, short labels for column headings because they are used as chart labels and will fit better on the chart axis
- Label the axes if necessary and pick a descriptive title
- If you use numbers such as years for labels, Excel may mistake them for data in which case you have to manually define the label ranges.
- Don't expect your audience to interpret the chart the way you do, even if you think it's obvious. Tell them what the chart shows by using a more descriptive title or a callout. If you think the chart shows that sales are increasing, say it on the chart.

Key Takeaways

- Line charts used for trends in time, column and bar charts for comparisons between categories, pie chart for contributions to the whole, scatter chart for data points and trend lines, histograms for frequency distributions, Pareto charts for identifying the most frequent occurrences and what percentage of the problem they represent. [3.2.1]
- Line and scatter charts assume the values shown on the vertical axis depend on the values on the horizontal axis and that the intervals on the horizontal axis are uniform. Bar and column charts assume the values use the same measurement units. Pie charts assume that the parts add up to a whole. Histograms assume that numbers on the boundary of two ranges are counted with the lower range. [3.2.2]
- Use small words as labels, label the axis, use the title and callouts to express your interpretation of the chart. [3.2.3]

3 Create Charts to Display Patterns, Relationships, and Anomalies

Learning Objectives

1. Create a column chart. [3.3.1]
2. Add labels, titles, and legends to explain the chart [3.3.2]
3. Change chart types from column to line. [3.3.3]

In this section, you create charts and explore how they can be used to display and recognize patterns and trends, relationships, and unusual occurrences.

Charting a trend or pattern

Column charts and line charts are useful for recognizing a trend or pattern that varies with time. The table of data in the UtilityBills worksheet is organized by month and the months increase from left to right in equal increments. This type of worksheet organization is suitable for showing trends. To show a trend by month, you select the range of months and a range of corresponding data. There are some tricks and guidelines you


should know about when you choose ranges to chart. In this section you will learn how to choose the ranges to make column charts that show a trend.

1. On the UtilityBills sheet, select A13:M14, as shown in Figure 3.17.

12	Utility Bills (KWH)												
13		Dec-08	Jan-09	Feb-09	Mar-09	Apr-09	May-09	Jun-09	Jul-09	Aug-09	Sep-09	Oct-09	Nov-09
14	Electricity	1,461	922	1,105	951	903	647	706	803	689	704	1,042	1,035

Figure 3.17. Months and electricity selected

In general, select the labels at the beginning of the rows and at the top of columns with the numbers. The charting program can recognize sequences of time, text labels, and numbers and it often places them in the chart appropriately.

2. On the Insert tab, in the Charts group, click the Column button, . A selection of column charts displays, as shown in Figure 3.18.

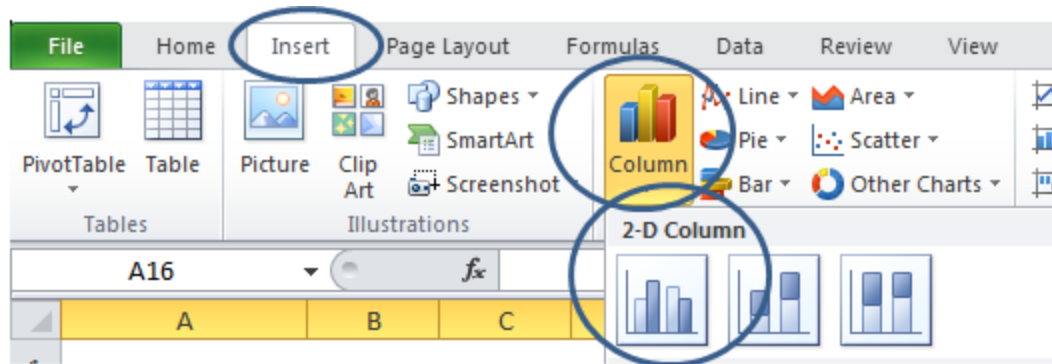


Figure 3.18. Column chart options

3. On the menu of column chart options, under 2-D Column, click the first option—Clustered Column.

A column chart is inserted on the worksheet, as shown in Figure 3.19.

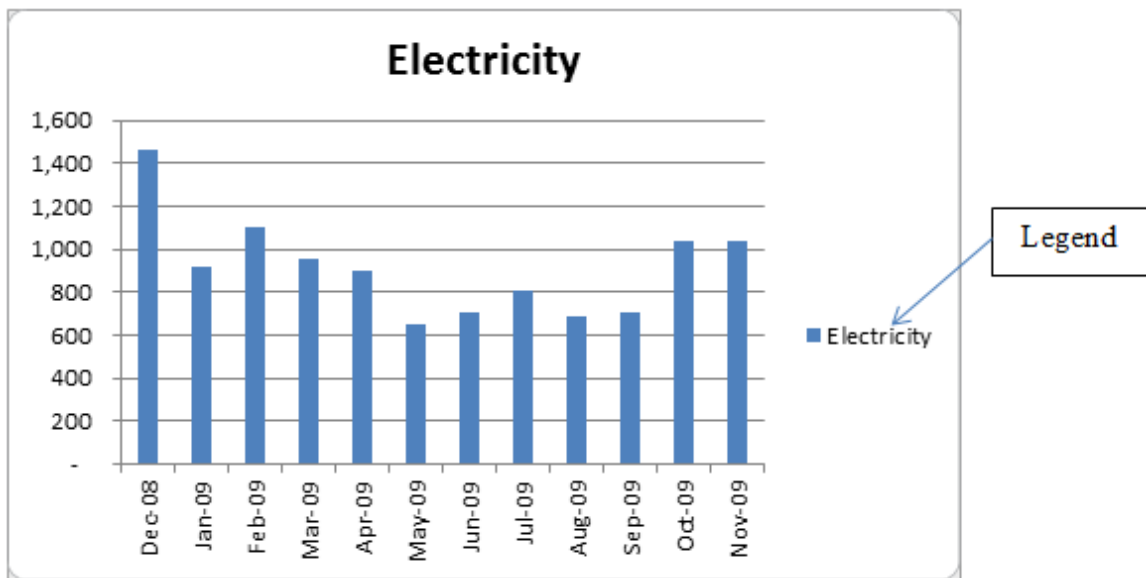


Figure 3.19. 2-D column chart inserted

4. Observe the following elements of the chart:
 - the program correctly identified the sequence of months as the independent variable that should be arranged along the horizontal (X) axis
 - the size and intervals of the vertical axis are calculated automatically based on the largest number in the selected range of numbers
 - columns are displayed whose height represents the numerical value and a color is assigned to the columns
 - a legend is displayed to explain the meaning of the color
 - a title is displayed based on the label used for the row of numbers.
5. Because there is only one set of colored columns in this chart, the legend is not needed. Click the legend to select it as shown in Figure 3.20.

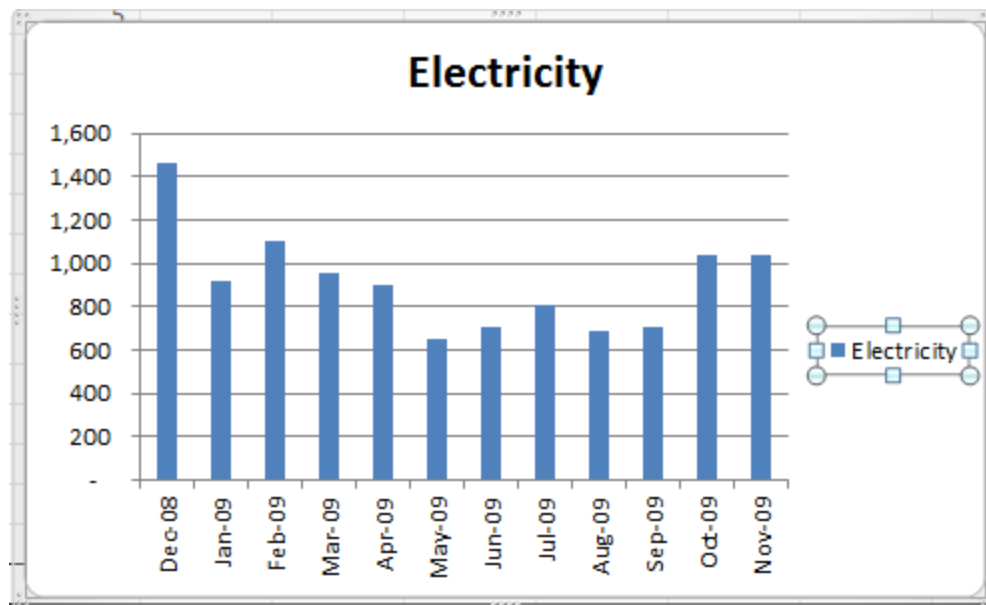


Figure 3.20. Legend selected

Charts have many parts that can be selected together or separately. When selecting a part of a chart, notice which part of the chart is surrounded by a border with circles and squares. The circles are sizing handles in the corners that can change the size of the selected item both horizontally and vertically and the squares on the sides can change the size in only one direction.

6. On the keyboard, press the Delete key. The legend is deleted and the chart resizes.
7. Look at the chart and consider what the bars mean. Observe that the use of electricity seems to be higher in the winter months and lower during the summer.

Use the title to communicate the interpretation of the chart that you wish to convey to the audience for whom the chart is intended.

8. Click the title to select it. Type **Lower Electricity Use in Summer** and then press Enter. As you type, the words display in the formula bar.
9. On the Chart Tools Layout tab, in the Labels group, click the Axis Titles button, and then move the pointer onto Primary Vertical Axis, as shown in Figure 3.21.

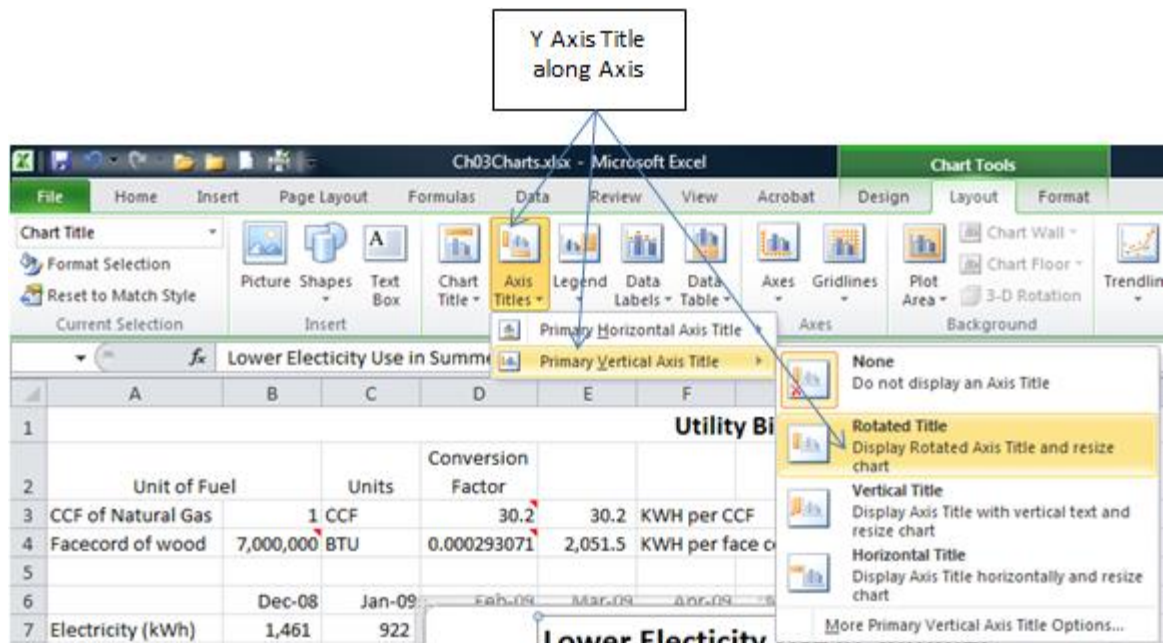


Figure 3.21. Add a title to the vertical axis

10. On the menu, click *Rotated Title*. Type **kWh** and then press Enter. Click a cell outside the chart to deselect the chart.
11. At the bottom right of the screen, use the Zoom function to shrink the display, if necessary to display rows 1 through 35.
12. Move the pointer onto the chart where it becomes a four-headed arrow. Drag the chart below row 21. Compare your chart with Figure 3.22.

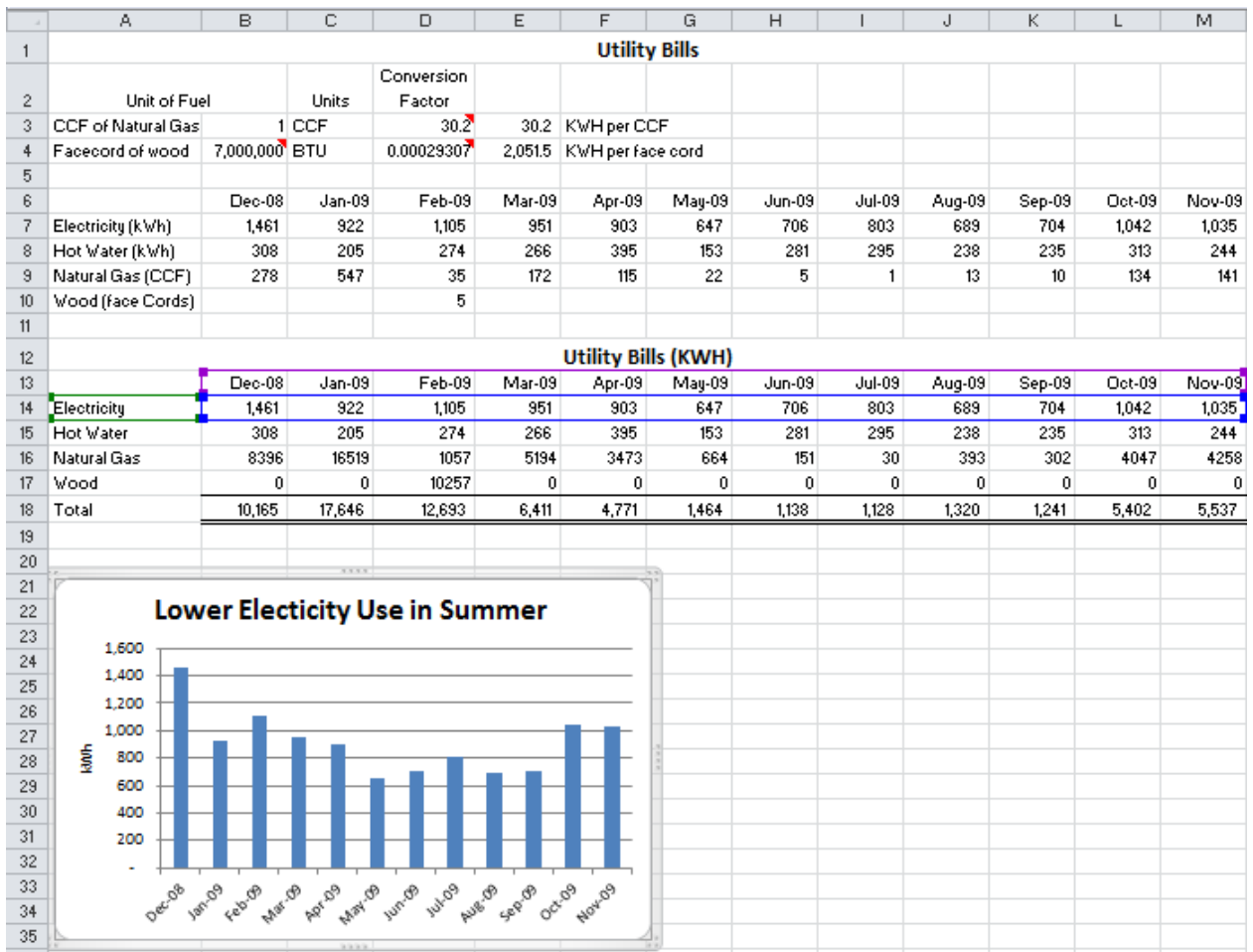


Figure 3.22. Place chart below the data.

13. With the chart selected, on the Chart Tools Design tab, in the Type group, click Change Chart Type.
14. In the Change Chart Type dialog box, in the left pane, click Line. In the row of line charts, click the fourth option, Line with Markers, as shown in Figure 3.23.

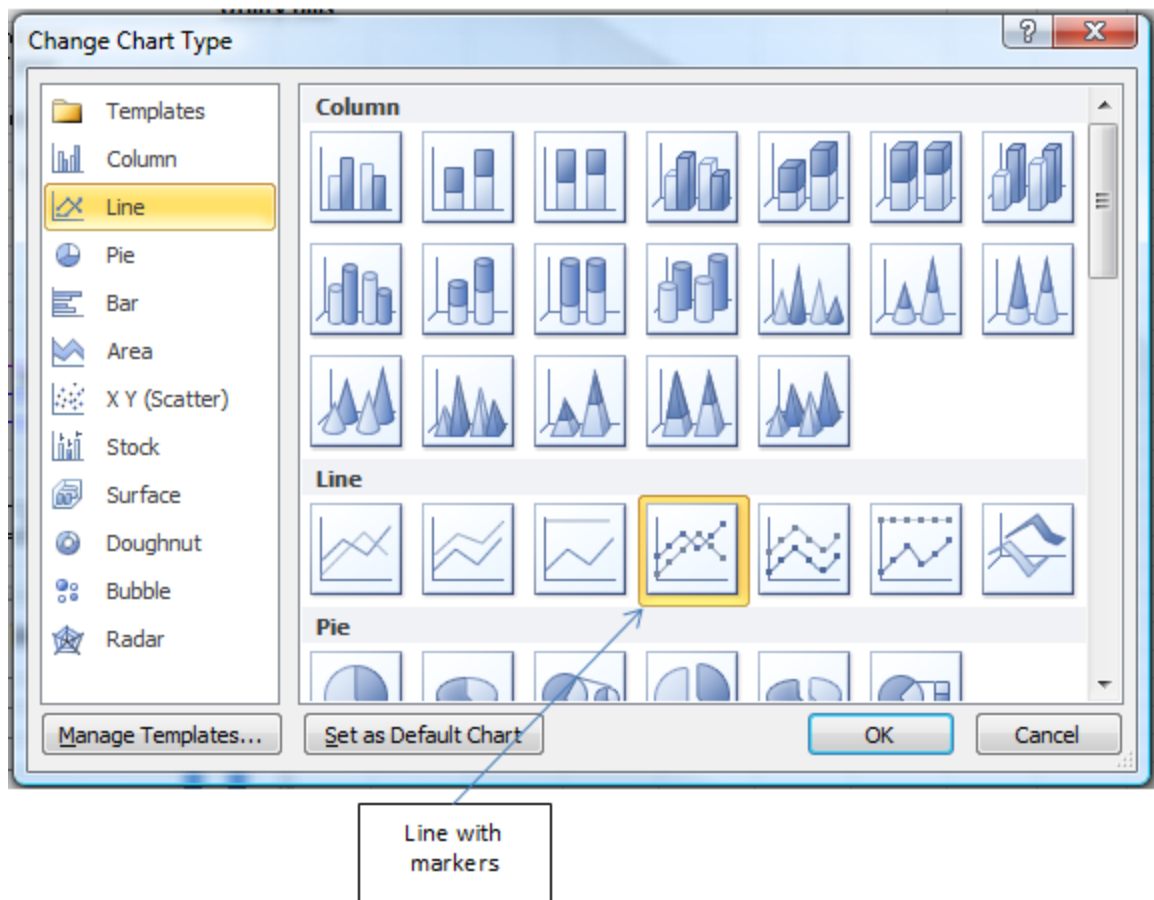


Figure 3.23. Change chart type to a line chart.

15. Click OK. The column chart changes to a line chart where each data point is indicated by a small diamond marker as shown in Figure 3.24.

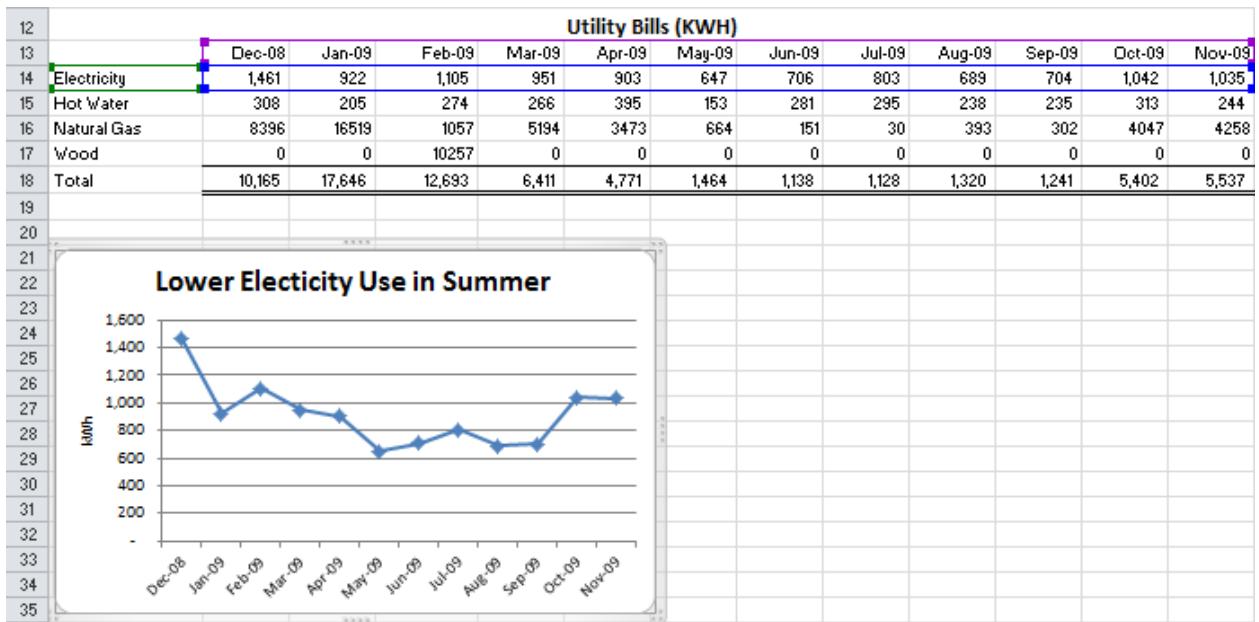


Figure 3.23. Line chart shows pattern of electricity use.

16. Save the workbook but leave it open.

Key Takeaways

- Select the data and the adjacent labels. On the Insert tab, in the Charts group, click the Column button. For other types of charts, click the appropriate button. [3.3.1]
- With the chart selected, on the Chart Tools/Layout tab, in the Labels group, use the Axis Titles, Chart title, or Legend buttons to add a label. Alternatively, if the legend or title already exists on the chart, click the legend or label to edit it. [3.3.2]
- With the chart selected, on the Chart Tools/Design tab, in the Type group, click Change Chart Type, and then select the new chart type. [3.3.3]

4 Chart Related Data

Learning Objectives

1. Assign independent variable to the horizontal axis with uniform intervals. [3.4.1]
2. Infer information from the shape of the line. [3.4.2]
3. Interpret the slope of the line. [3.4.3]

4. Estimate values between and beyond known data points using a trend line. [3.4.4]
5. Create a scatter chart with a trend line. [3.4.5]

Line charts show the relationship between pairs of measurements. For example if you were considering where in the country a photocell array would be most effective, you might compare how cloudy it is in each area. Figure 3.25 shows a line chart of the percentage of possible sunshine for each month in Detroit, Michigan and Phoenix, Arizona. (United States Climate 2007) <[Link](#)>.

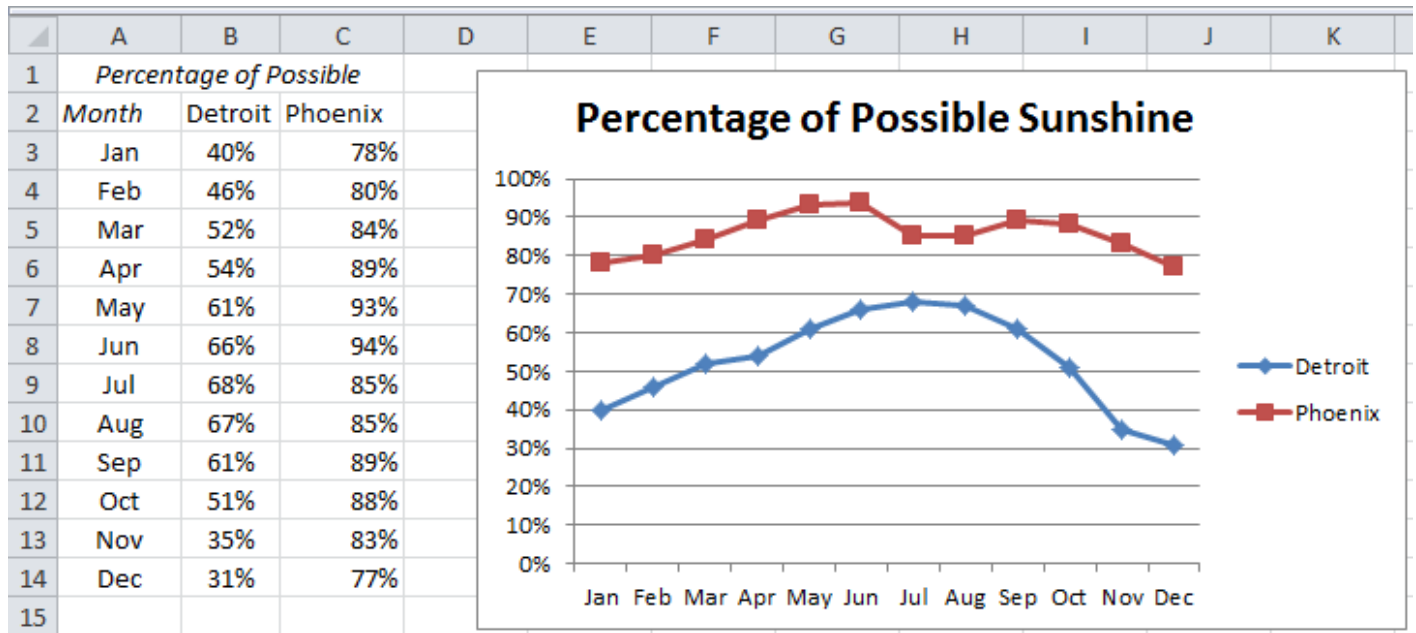


Figure 3.25. Sunshine

Line charts show the relationship between an **independent variable** such as time and another variable that depends on it, which is called the **dependent variable**. The independent variable is always displayed along the horizontal axis of the chart. The values along the horizontal axis must increase from left to right and the intervals must be uniform. Charts show patterns that are more easily recognized as shapes or a trend than a sequence of numbers.

Inspection of the chart in Figure 3.25 could lead to the inference that it is less cloudy in Detroit in the summer months and the month with the most clouds is December. Inferences like these are based on

assumptions. In this case, the assumption is that clouds were the cause of less than 100% of the possible sunshine from reaching the ground. In some cities, it could be smog so this assumption is not always correct. By comparing the two lines, you could infer that Phoenix has about twice the percentage of possible sunshine during the winter months.

If the lines appear to be consistent shapes like straight lines or smooth curves, the relationship between the independent and dependent variables might be a **function** that could be represented by a formula.

For example, Figure 3.26 shows a table of time and distances. They represent measurements that were taken during a trip in an automobile. Notice that the interval of hours is not equal. There is a three-hour interval between the last two pairs of data.

	A	B
1	<i>Time (hrs)</i>	<i>Distance (miles)</i>
2	0	0
3	1	60
4	2	120
5	5	300

Figure 3.26. Table of time and distance measurements for a car traveling at 60 mph

In this example, time is the independent variable. The dependent variable (distance) is measured along the vertical axis. Mathematicians call the horizontal axis the **X-axis** and the vertical axis the **Y-axis**.

If this table was charted using a line chart, it would look like the chart in Figure 3.27.

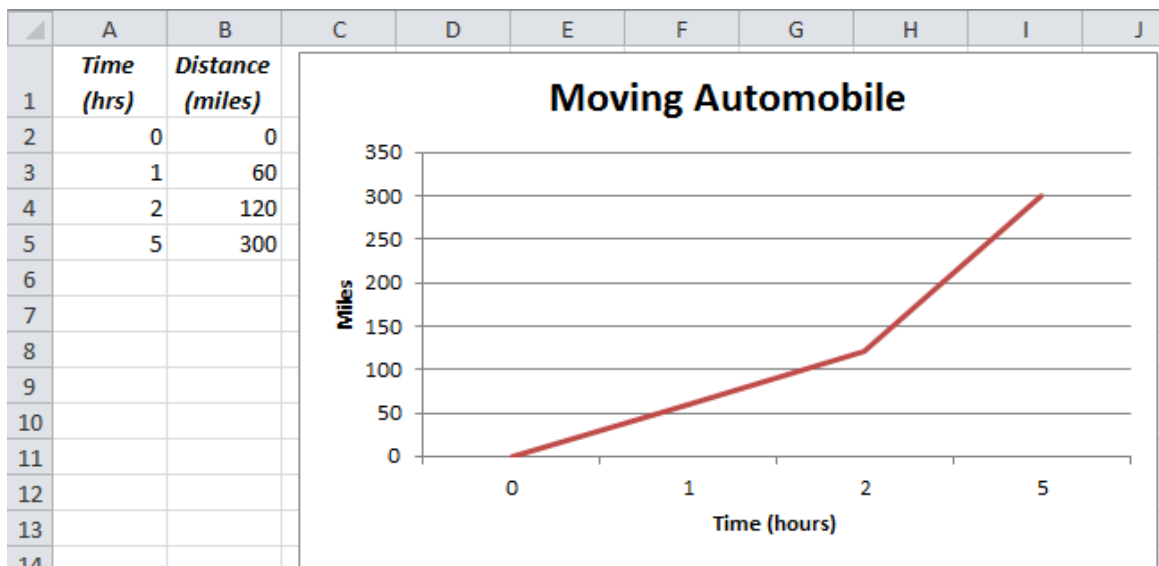


Figure 3.27. Line chart of data with unequal intervals of time

The chart in Figure 3.27 gives the mistaken impression that the relationship between time and distance changed after two hours. Line charts should always use equal intervals on the horizontal axis to avoid this type of misunderstanding. If the data you have does not have equal intervals between the measurements of the independent variables, Excel has a special type of line chart called a **scatter chart** that will display the intervals equally as shown in Figure 3.28.

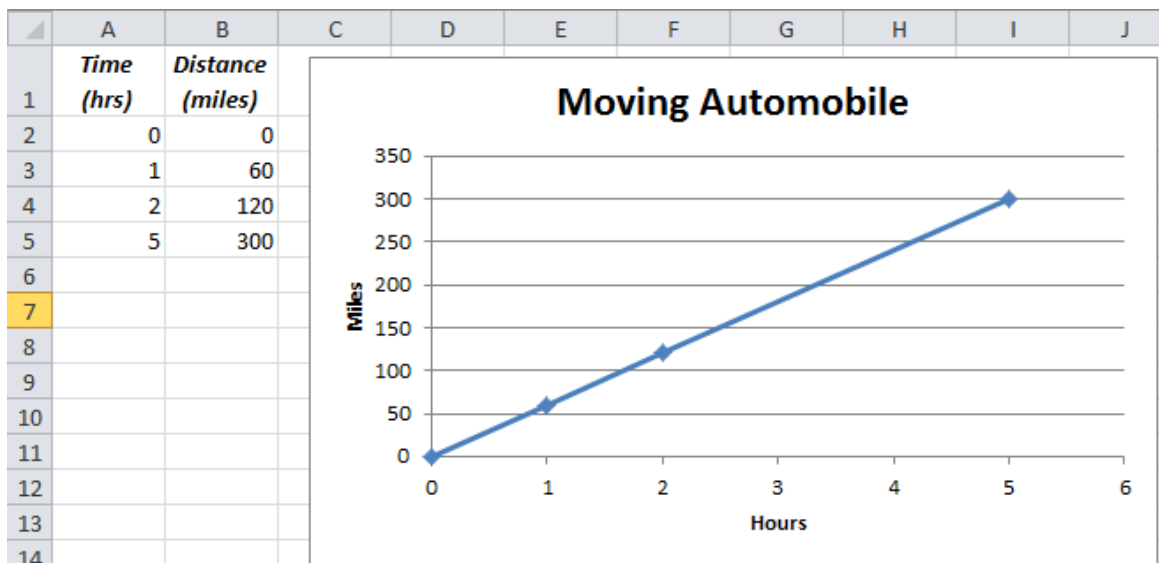


Figure 3.28. Scatter chart forces equal horizontal intervals giving the slope of the line meaning

If the line chart is properly constructed, we see a line that makes a consistent shape. In this example, the ratio of the change in distance to the change in time is a constant that creates a straight line. This ratio is like the pitch of a roof, and is called the **slope**.

Mathematicians often use the **Greek letter delta (Δ)** to represent the phrase “the change in”. They would write this ratio as: $\Delta d / \Delta t$ where d = distance and t = time.

To calculate the slope, pick any two points on the line. Determine the differences in the measurements of the distance and time and find the ratio, as shown in Figure 3.29.

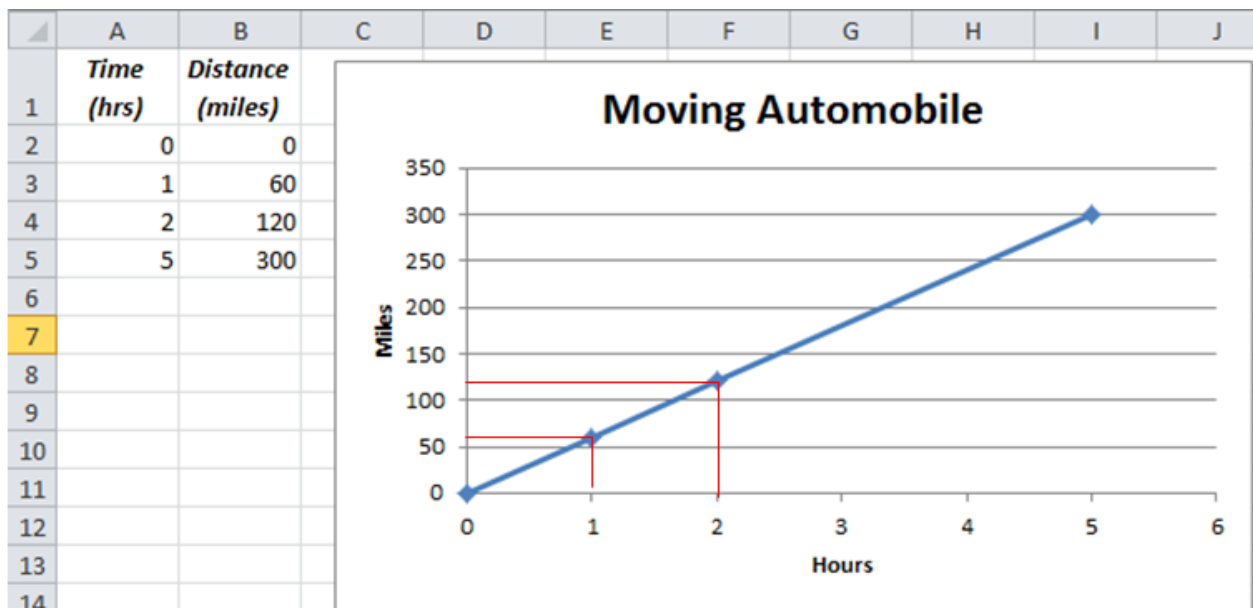


Figure 3.29. Slope is the speed for a distance and time chart

In this example, $\Delta d = 60$ miles and $\Delta t = 1$ hour. The slope = 60 miles/1 hour or 60 miles per hour. For this type of line chart, the slope is also the average speed of the automobile. Because the slope is the same for the entire line, we can infer that the speed was the same throughout the trip. The relationship can be expressed as a formula:

$$\Delta d = \text{slope} \cdot \Delta t.$$

If we infer from this line chart that the relationship between the time and distance was the same throughout the trip, we can estimate the distance at unmeasured times. If we wish to estimate the distance

traveled after three hours, we could locate the spot on the line that is directly above 3 hours on the horizontal axis, and then determine where that would be on the vertical axis, as shown in Figure 3.30.

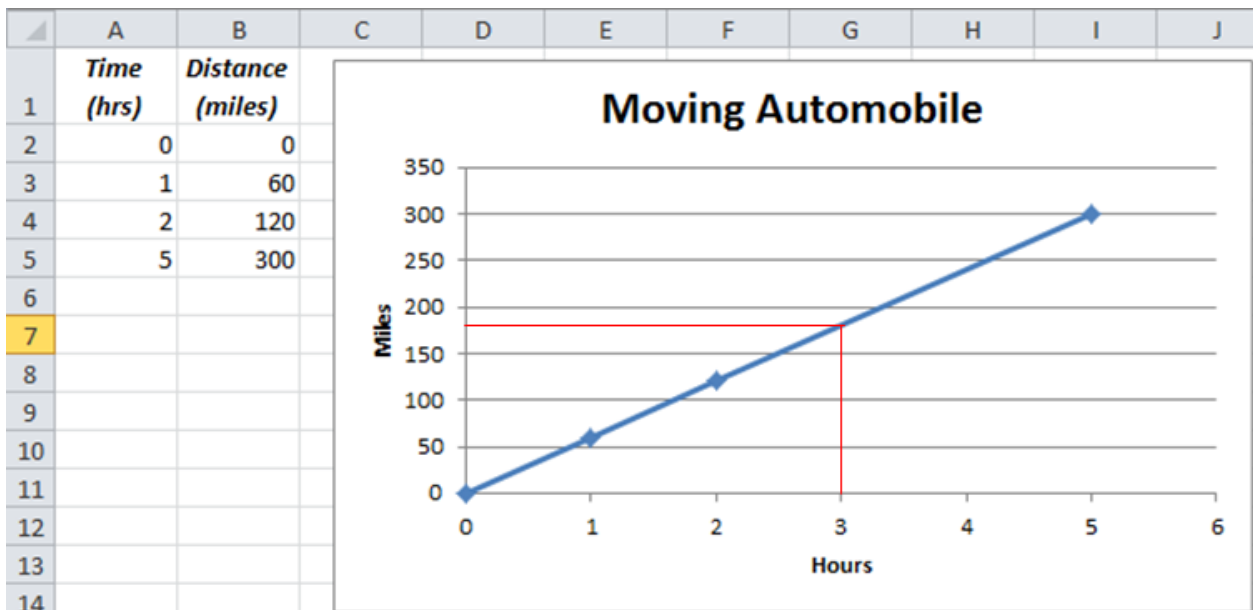


Figure 3.30. Estimating using interpolation

Alternatively, we could calculate the value of the slope, and then apply the formula.

$$\Delta d = \text{slope} \cdot \Delta t$$

$$\Delta d = 60 \text{ miles per hour} \cdot 3 \text{ hours} = 180 \text{ miles.}$$

In this example, the estimate would be close to 180 miles. Estimating values *between* known data points is called **interpolation**.

If we wanted to estimate how far the car would travel after six hours, we would have to assume that the relationship continued to be true beyond the last known data point. One way to perform this estimate is to extend the shape of the line, as shown in Figure 3.31.

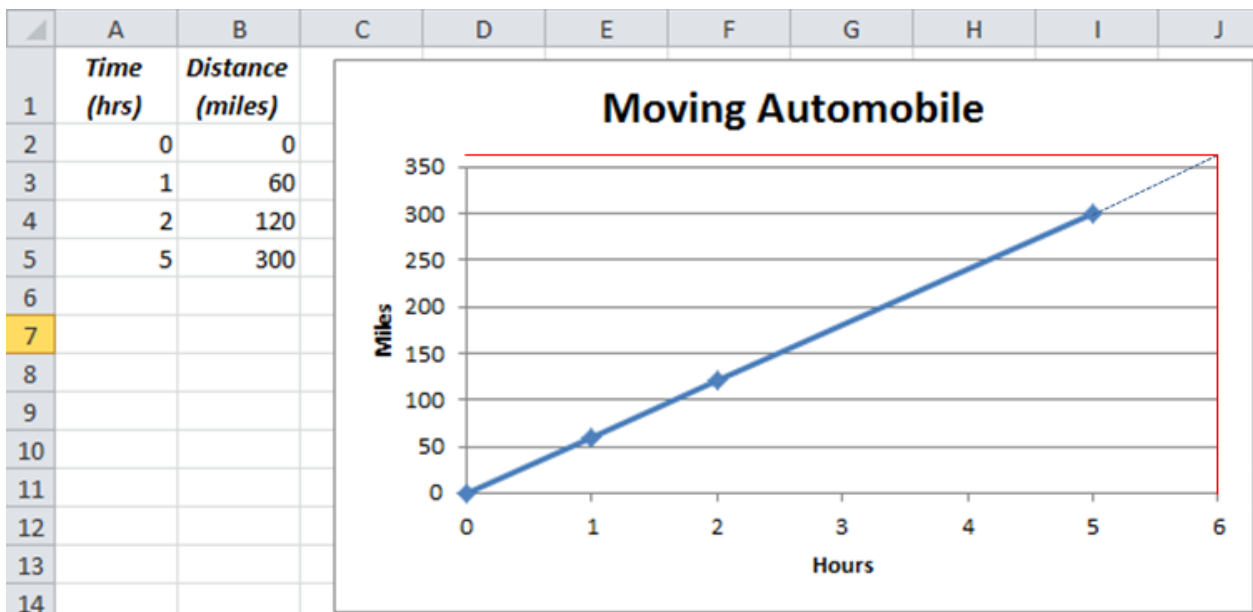


Figure 3.31. Estimating using extrapolation

In this example, the estimate would be close to 360 miles. Alternatively, we could apply the formula.

$$\Delta d = \text{slope} \cdot \Delta t$$

$$\Delta d = 60 \text{ miles per hour} \cdot 6 \text{ hours} = 360 \text{ miles.}$$

Estimating values *beyond* known data points is called **extrapolation**. In general, extrapolation involves assumptions that are more likely to be incorrect. For example, the trip may have ended after five hours. Even though extrapolation uses assumptions that might be incorrect, it is still a useful tool for predicting the future if the users are aware of the assumptions.

Real measurements are usually not exact. Consider the pairs of measurements and the chart shown in Figure 3.32. If you right-click on one of the data points, a shortcut menu appears.

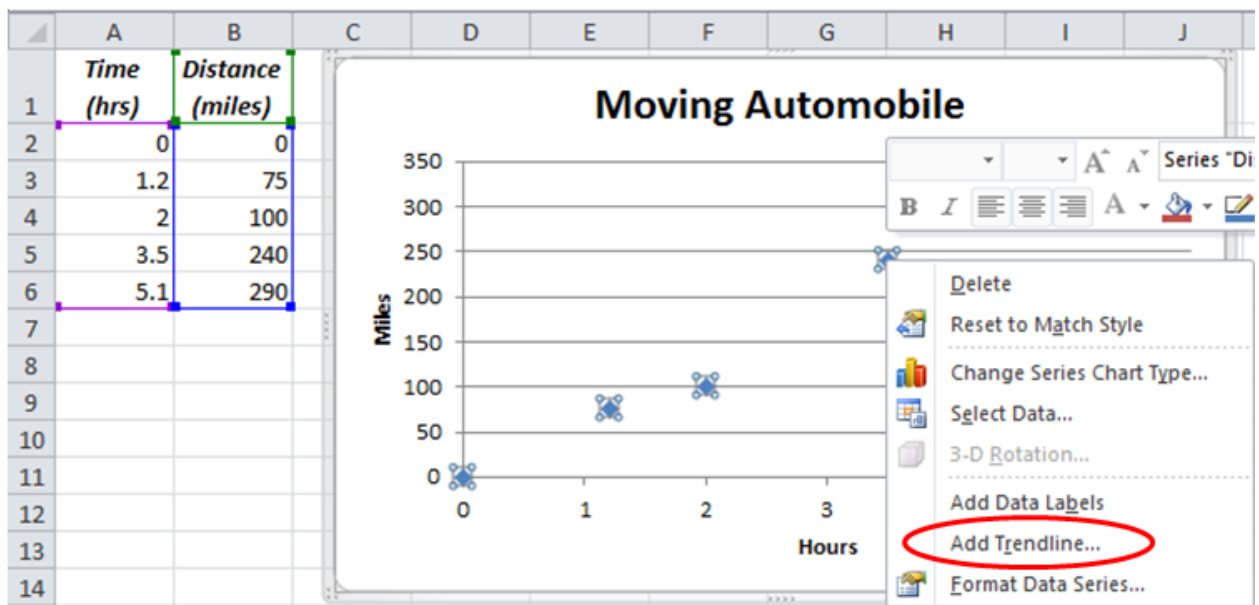


Figure 3.32. A Trendline can be added to data

To activate a program that estimates a function that will be a close fit to the data points, you can click the *Add Trendline* option. The *Format Trendline* dialog box displays, as shown in Figure 3.33. If we wish to use this data to make a prediction of the distance after six hours, the following choices should be made in this dialog box:

- The data appears to lie close to a straight line so the Linear option should be chosen.
- The horizontal measurement units are one hour so we want to forecast one unit into the future.
- We know that the first pair of data is accurate, zero time and zero distance. The point where the line crosses the vertical axis when the horizontal axis value is zero is called the **intercept**. The intercept should be set to zero because we assume that at zero time the distance was also zero.
- To calculate other extrapolations or interpolations, the equation of the line should be displayed.
- Display the R-square value to get an estimate of how well the line fits the data.

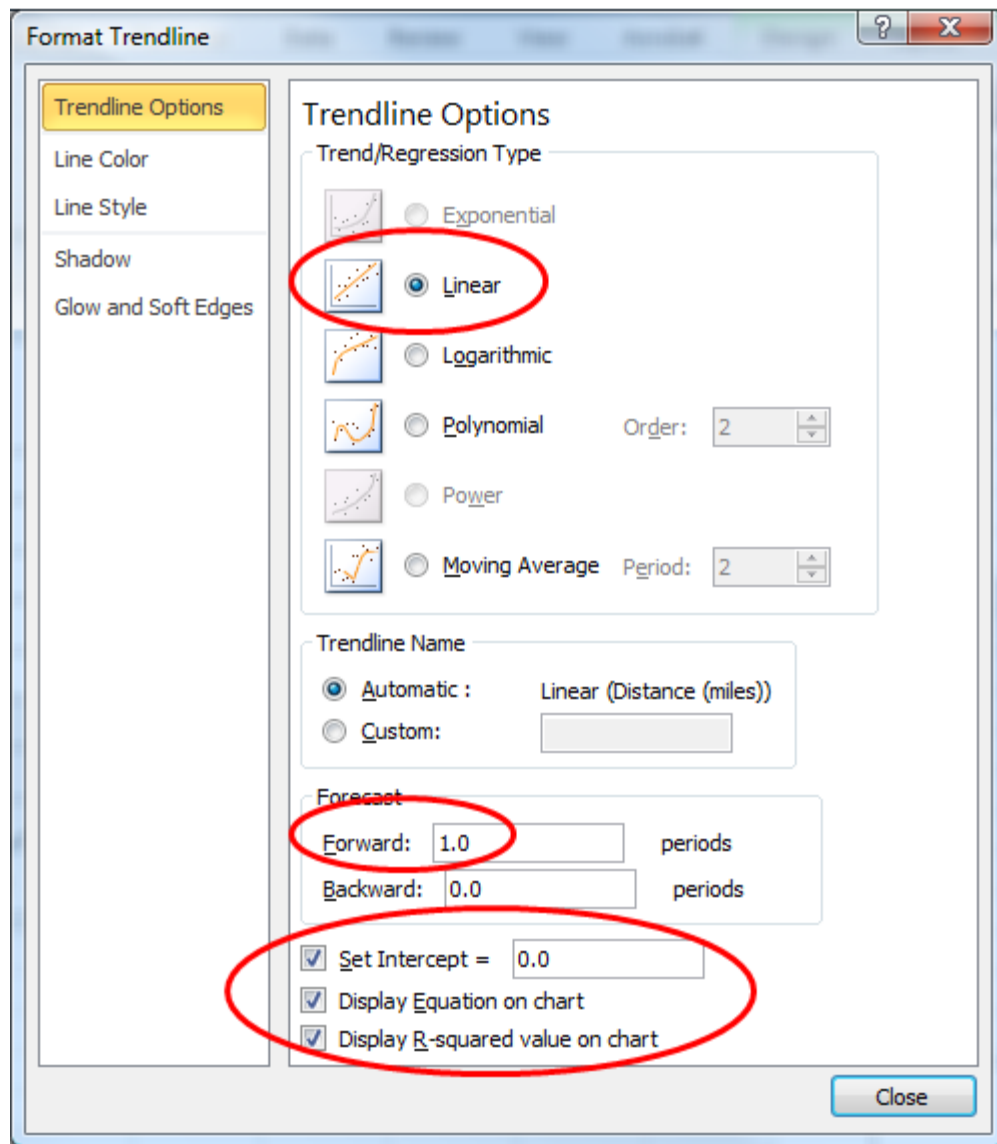


Figure 3.33. Trendline settings

A straight line is placed near the data points that intercepts the vertical axis at zero, as shown in

Figure 3.34.

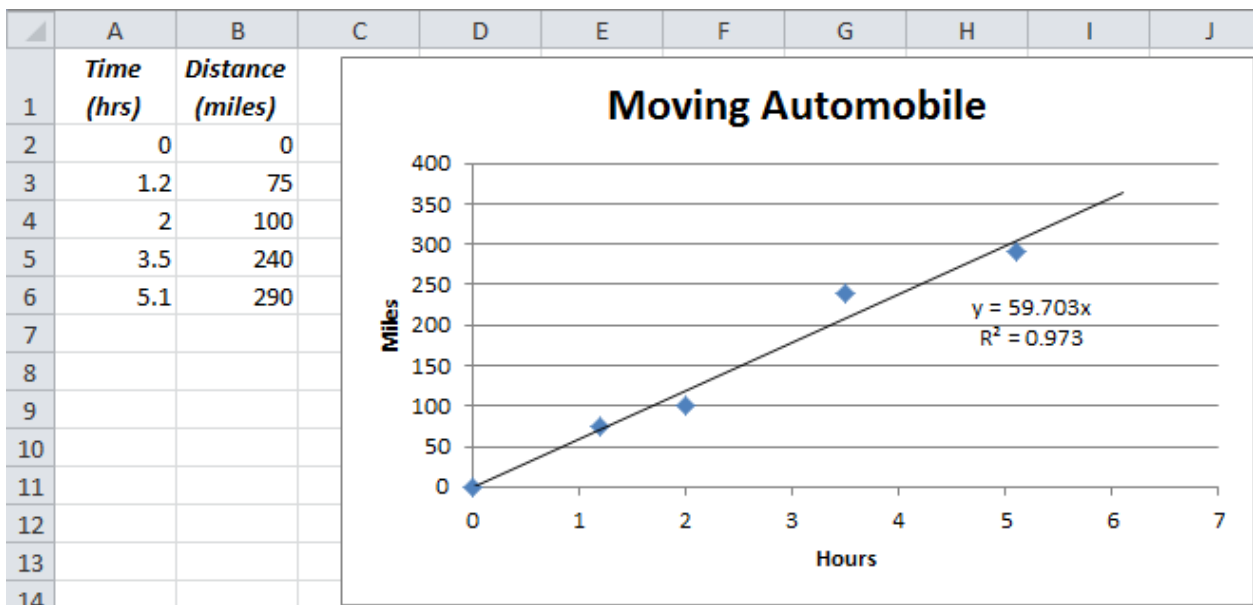


Figure 3.34. Trend line projected one additional hour

Observe that the formula for the line is $y = 59.703x$. Recall that mathematicians call the horizontal axis the X-axis and the vertical axis the Y-axis. The slope of the line is 59.703 which on this chart means the average speed represented by the slope of this line is 59.703 miles per hour. Observe that the line is extrapolated to six hours.

The **R-squared value** is a comparison between the proximity of the data to the trendline and the proximity to a horizontal line at the average value of the data. Consider a new chart of a different set of data points shown in Figure 3.35.

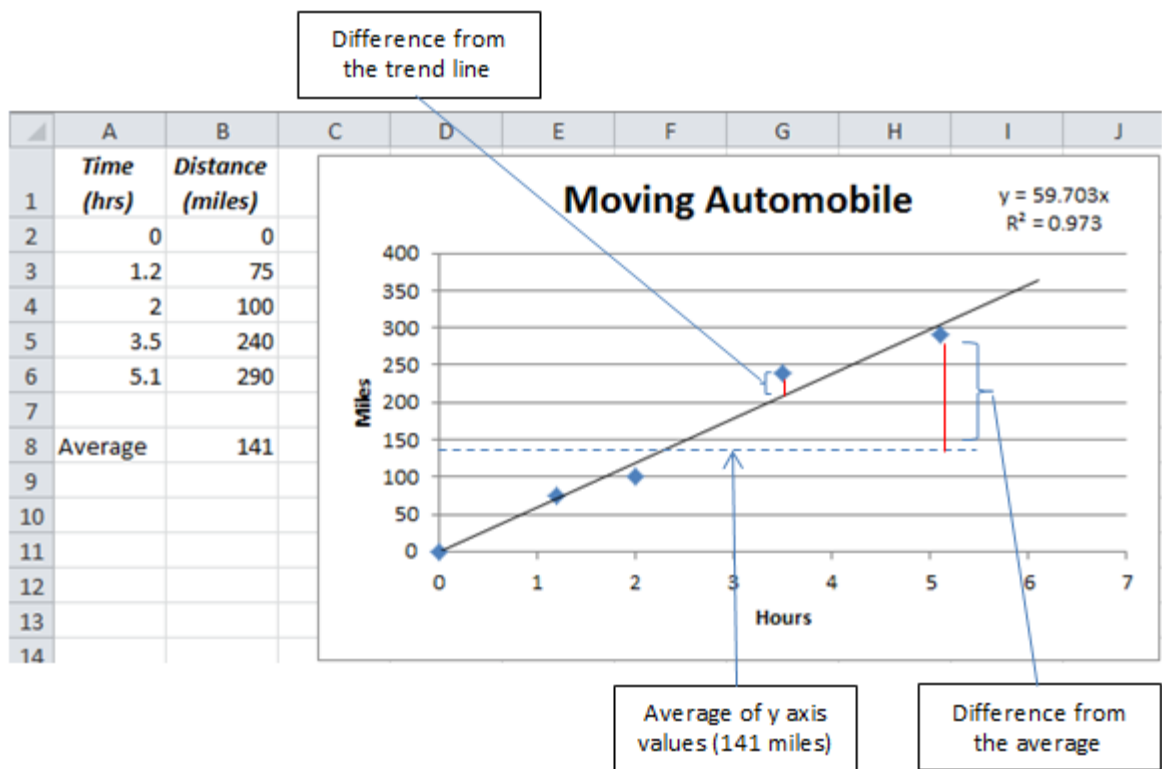


Figure 3.35. Comparing function line to a horizontal line (no dependence)

The average of the five Y values for these five pairs of data is 141 miles and a horizontal line is drawn at that height on the Y-axis. This horizontal line represents the formula $y = 141$. Observe that the value of y is 141 regardless of the value of x which means that the value of y does not depend on x at all. A trend line that has a slope of 59.703 is drawn near the data points. The R^2 value for this line is calculated using the following steps:

1. Find the difference in the vertical direction between the data point and the trend line for each data point.
2. Square each difference (makes all the values positive)
3. Sum the squared numbers (this sum will be small if the data points are close to the line)
4. Find the difference in the vertical direction between the data point and horizontal line at the average for each data point.
5. Square each difference (makes all the values positive)

6. Sum the squared numbers (this sum should be larger than the differences to the trend line)
7. Divide the value from step 3 by the value from step 6 to find the ratio of the sum of squared differences.
8. Subtract the ratio from 1. This is the R^2 value for the trend line.

If the trend line is close to the data points, the sum of squared differences will be small. If the line is sloped, the sum of squared differences to the average will be large. In this example, the ratio of the differences between the data points and the sloped line to the horizontal line is .027. When this number is subtracted from 1, the result is .973 which is the R^2 value shown on the chart.

In general, R^2 values tell you something about how dependent the y values are on the x values and how well the formula of the trend line fits the data. R^2 values range between 0 (the value of y does not depend on the value of x and the trend line is about the same as the horizontal average line) to 1 (the function represented by the trend line is an exact match for the data and the y values depend on the x values).

The closer the R^2 value is to 1, the better the trend line matches the data points, which means it is a more reliable predictor of future results if the underlying relationships and assumptions do not change.

Using a Scatter Chart

Create the scatter chart that shows the relationship between time and distance traveled.

1. In the Ch03ChartsStudentName workbook, click the Driving sheet tab.
2. Select A1:B6. On the Insert tab, in the Charts group, click the Scatter button. On the menu, click the first choice, *Scatter with Only Markers*.
3. Use the skills you previously practiced to delete the legend, change the title to **Moving Automobile**, and add the rotated Y axis label **Miles**. Compare your chart to Figure 3.36.

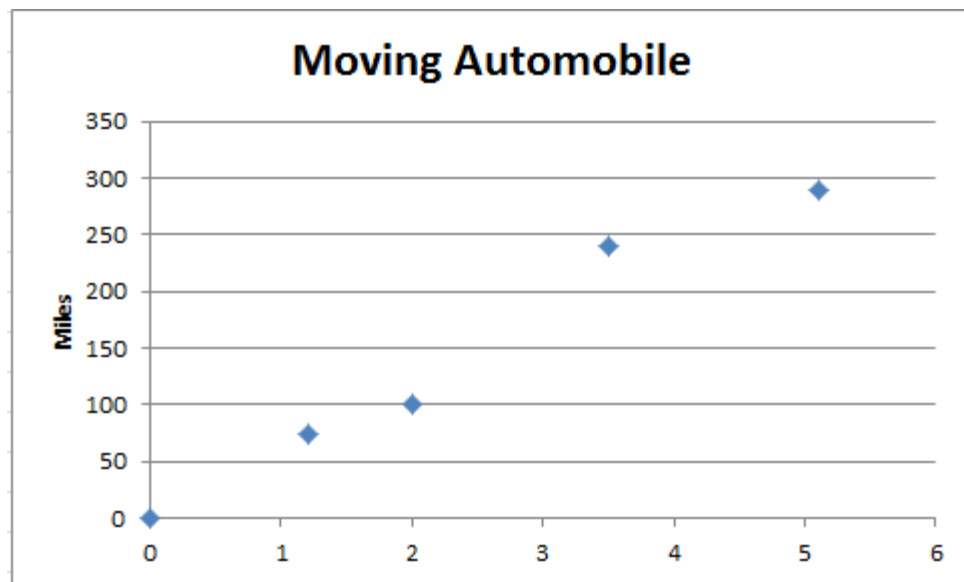


Figure 3.36. Scatter chart with title and Y axis label

4. Move the pointer onto one of the small blue diamonds that represent one of the pairs of data, and then right-click. From the shortcut menu, choose *Add Trendline*.
5. In the Format Trendline dialog box, choose Linear, Forecast forward 1 unit, and then set the intercept to zero. Check *Display Equation on chart* and *Display R-squared value on chart*. Click Close.
6. Click the line formula and R^2 value to select them. Drag them below the trend line to a more open part of the chart. Click outside the chart to deselect it. Compare your Chart with Figure 3.37.

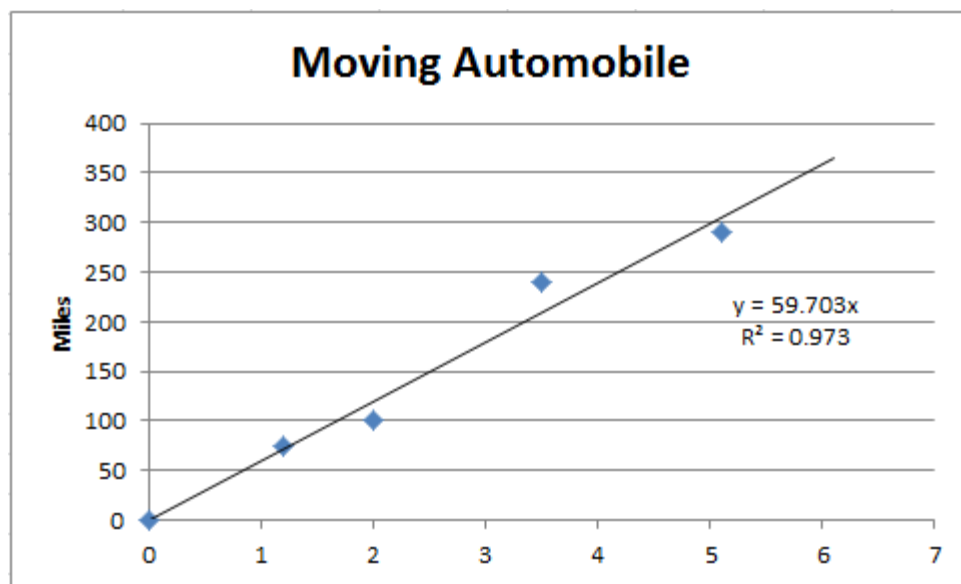


Figure 3.37. Scatter chart with trend line

Key Takeaways

- Line charts display a relationship between an independent variable measured on the horizontal axis and an dependent variable measured on the vertical axis. The intervals of the horizontal axis must be equal. If the data provided does not have equal intervals, use a scatter chart to automatically display equal intervals on the horizontal axis. [3.4.1]
- The shape of the line and how it changes can be used to make logical inferences about the dependent variable such as when maximum or minimum values occur and how two lines compare with each other. [3.4.2]
- The slope of the line is the relationship between the independent and dependent variable. The slope is the ratio of the change in the dependent variable to the change in the independent variable, written as $\Delta y / \Delta x$ where the Greek letter Δ represents the phrase “the change in”. [3.4.3]
- The line chart can be used to estimate values between known data points (interpolation) by finding the point of interest on the line and reading the corresponding measurement from the chart axis. Alternatively, the slope of a straight line can be calculated from the line chart and applied using the formula $\Delta d = \text{slope} \cdot \Delta t$ [3.4.3]
- The line chart can be used to estimate values beyond known data points (extrapolation) by extending the line, finding the point of interest on the extension of the line, and then reading the corresponding measurement from the chart axis. Alternatively, the slope of a straight line can be calculated from the line chart and applied using the formula $\Delta d = \text{slope} \cdot \Delta t$ [3.4.4]
- A trend line can be added that is a close fit to a series of measurements. If you know that the dependent variable is zero when the independent variable is zero, choose to force the line to have a zero intercept. Display the formula and R^2 value to show the slope of the line and the fit. A perfect fit has an R^2 value of 1. If there is no dependent relationship, the R^2 value is zero. [3.4.4]

5 Analyzing Data with Charts in Excel

Learning Objectives

1. Recognize a chart that shows no dependent relationship between x and y values [3.5.1]

2. Use charts to identify anomalies and add callouts [3.5.2]
3. Create, rotate, and label pie charts [3.5.3]
4. Move a chart to a separate worksheet and move the worksheet. [3.5.4]

Charts can be interpreted to change the data upon which they are based into information. In this section, you interpret a trend line, look for unexplained variations in a pattern, and use a pie chart to make comparisons.

Using a trend line to evaluate a possible relationship between two variables

The slope of a trend line and the R^2 value can be used to determine if one variable depends on another. If it does, knowing one of them will allow you to interpolate and extrapolate unknown values, otherwise you know that such estimates are not reliable. In this exercise, you compare the electricity used to heat water and the month of the year.

1. In the Ch03ChartsStudentName workbook, click the UtilityBills sheet tab. Select A13:M14.
2. Insert a scatter chart. Use the first option, *Scatter with Only Markers*.
3. Delete the legend. Add **kWh** as a Y axis label. Observe that the electricity use seems to be lower in the summer months and higher in the winter months. To explore the possible relationship between electricity use and month of the year, you will insert trend lines and evaluate the R^2 values.
4. Right-click one of the data points. On the shortcut menu, click Add Trendline. On the Format Trendline dialog box, choose Linear and check the box next to *Display R-squared value on chart*.
5. Drag the R^2 value to a clear place on the chart where it is easily read. Observe that the value is 0.1859. Recall that low values of R^2 near zero imply that the trend line is not much better than a simple horizontal line at the average. If a relationship exists between the electricity use and the month of the year, it is not a linear (straight line) relationship.
6. Right click the trend line. On the shortcut menu, click *Format Trendline*.
7. In the Format Trendline dialog box, click Polynomial, and then click Close.

8. Observe that the R^2 value is 0.758 which is much closer to 1. This implies that there might be a dependent relationship between the electricity use and the month which would allow cautious use of interpolation and short-term extrapolation. Compare your chart to Figure 3.38.

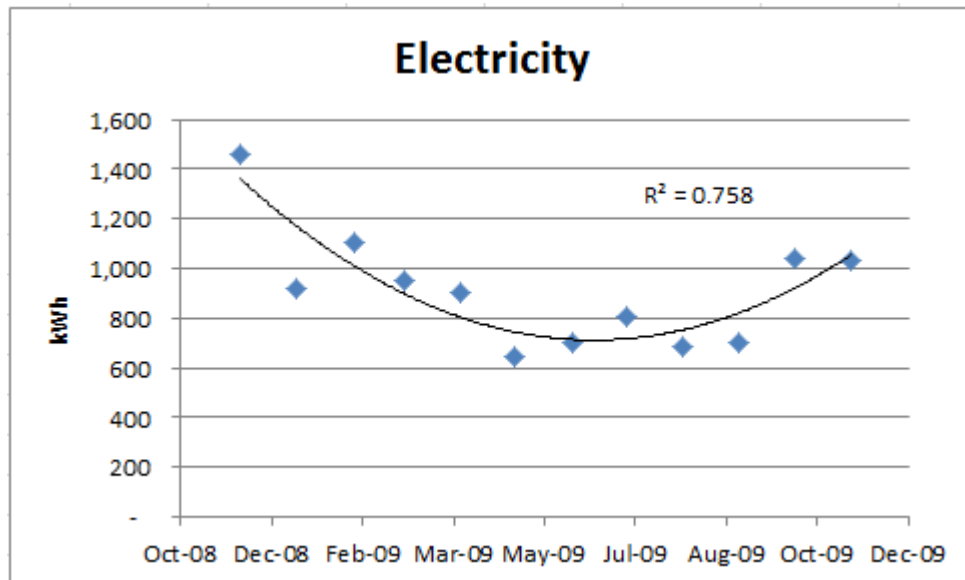


Figure 3.38. Electricity use may be dependent on month but not in a linear relationship

9. On the Utility Bills sheet tab, select A13:M13. Press and hold the CTRL button, and then drag A15:M15. Observe that two ranges are selected that are not adjacent to each other, as shown in Figure 3.39.

	A	B	C	D	E	F	G	H	I	J	K	L	M
12	Utility Bills (KWH)												
13		Dec-08	Jan-09	Feb-09	Mar-09	Apr-09	May-09	Jun-09	Jul-09	Aug-09	Sep-09	Oct-09	Nov-09
14	Electricity	1,461	922	1,105	951	903	647	706	803	689	704	1,042	1,035
15	Hot Water	308	205	274	266	395	153	281	295	238	235	313	244

Figure 3.39. Non-adjacent ranges selected

To select ranges of cells that are not adjacent hold the CTRL key before you drag the second range. The two ranges must be parallel and the same length.

10. Insert a scatter chart using the first option, *Scatter with Only Markers*. Drag the chart below the first chart to row 37.
11. Delete the legend and add a rotated Y axis label **kWh**.

12. Right click one of the data points. Add a trendline. In the dialog box, specify Linear and display the R^2 value. Click the Close button.
13. Observe that the R^2 value is 0.0044 which is very close to zero. This implies that the trend line is not much better than a simple horizontal line drawn at the average of the data.
14. Right-click the trend line and change the shape of the line to a polynomial. Observe that the R^2 value is also very low. Repeat this process for the Exponential, Logarithmic and Power options to confirm that none of them have R^2 values that are much better. From this evidence, you may conclude that the electricity used for heating water does not appear to depend on the month of the year.
15. Change the trendline type back to linear. Change the title to **Hot Water Energy Use Not Dependent on Month**. Compare your chart to Figure 3.40.

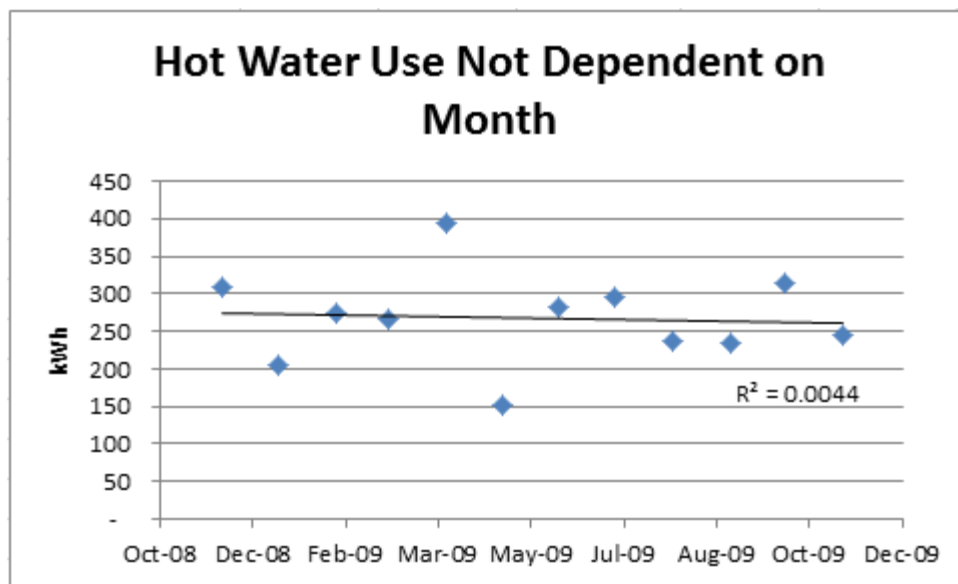


Figure 3.40. Electricity use for hot water does not depend on the month of the year

Spotting and Labeling Anomalies

Sometimes it is the exception to the pattern that is important. Charts make it much easier to recognize data that is far from a trend line or inconsistent with its neighboring data points.

1. On the UtilityBills sheet, select A13:M13 and A16:M16. Insert a 3-D column chart, and then delete

the legend. Drag the chart to row 37 to the right of the hot water chart. Compare your chart to Figure 3.41.

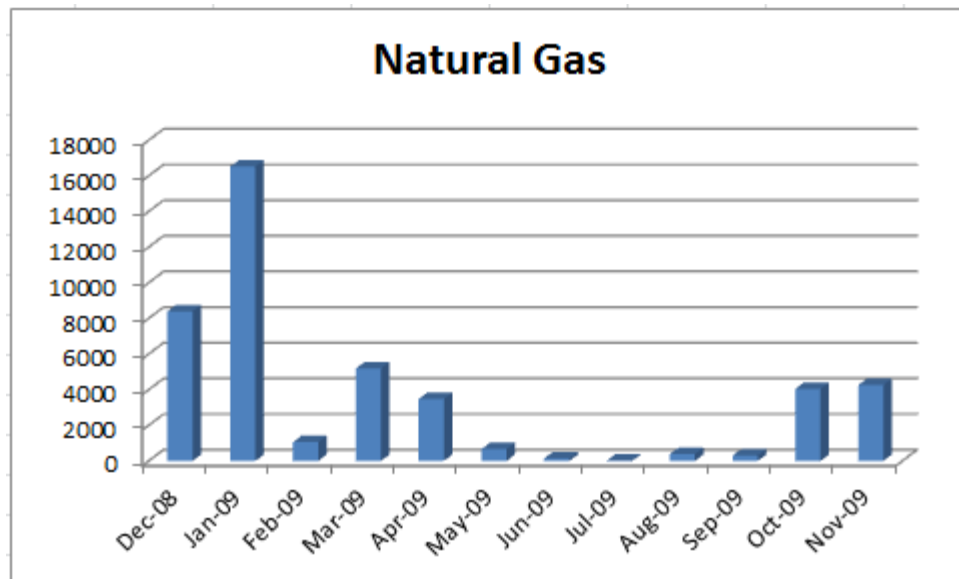


Figure 3.41. Natural gas use by month

2. Observe that the natural gas used for home heating is high in the winter months except for February.

*To identify anomalies you apply your general knowledge of events, human behavior, and science. This type of knowledge is called **world knowledge**. In this example, you know that the month of February is comparable to January for cold weather so it is unexpected that the use of natural gas is so much lower. To call attention to an anomaly, use a callout which is a note on the chart that points to the anomaly.*

3. On the Insert tab, in the Illustrations group, click the Shapes button. A menu of shapes displays, as shown in Figure 3.42

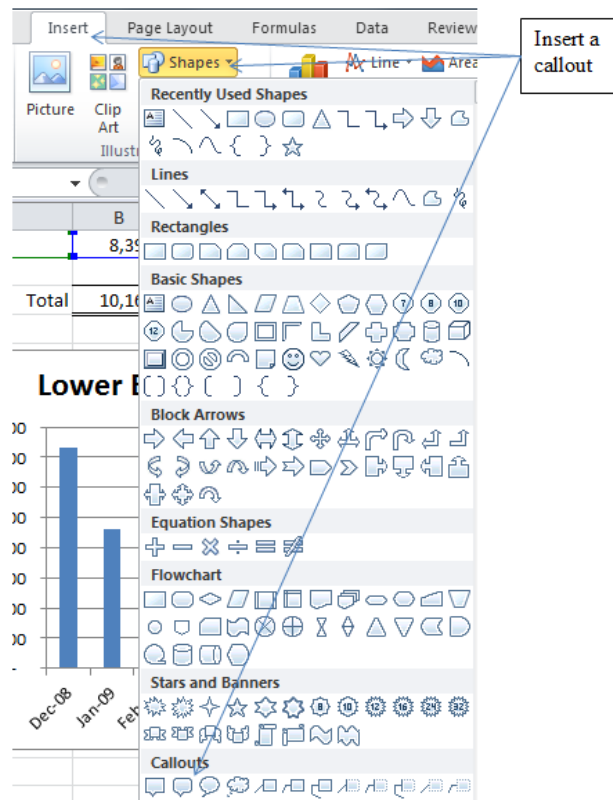


Figure 3.42. Callouts on the Shapes menu

4. Under Callouts, click the second option. On the chart drag a rectangle in the middle of the chart.

The shapes are easy to move and resize. Drag the outline of the shape in the approximate size and location with the knowledge that you will probably resize and relocate it later.

5. Type **Lower than expected gas use in February**
6. Drag the small yellow diamond at the end of the callout point near the column for February. Resize and relocate the callout as needed to display the text without obscuring any of the columns. Refer to Figure 3.43.

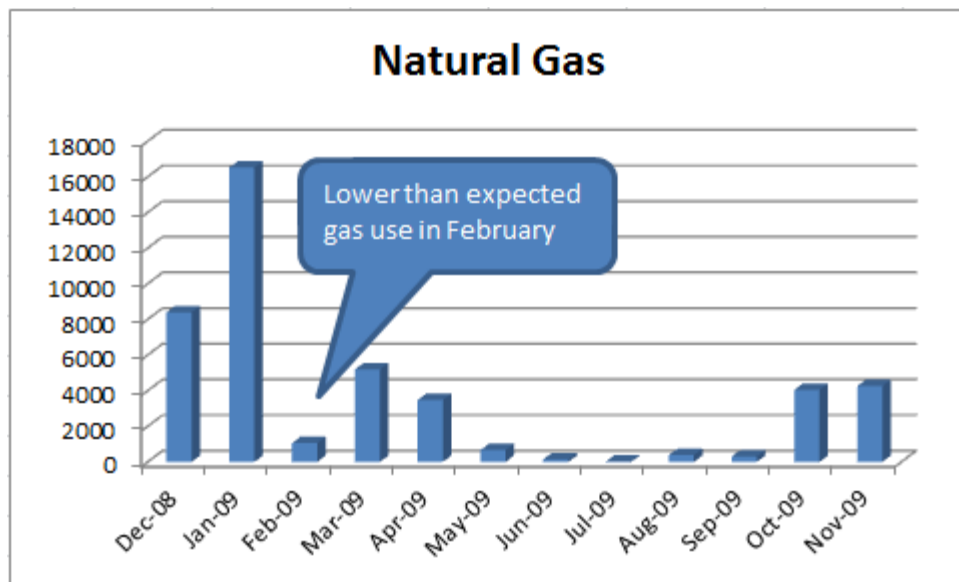


Figure 3.43. Use callouts to point out anomalies

Comparing Categories

Tables of data often have more than one row of data that represents similar categories. Charts can be created with more than one line or more than one set of bars or columns that can be compared side-by-side. The values in each category can be stacked in column or bar charts to show their contribution. In this exercise, you evaluate the contribution of burning wood to the total heating energy by month to explain the anomaly identified in the previous section.

1. Select ranges A13:M13 and A16:M17 to include the energy from fire wood with the natural gas.
2. Insert a 2-D clustered column chart, and then drag it down to row 53, below the hot water chart.
3. Insert a chart title, **Heating Energy** and a rotated Y axis label **kWh**. Observe that the legend performs a valuable function when more than one line, column, or bar is displayed by explaining the color codes used for each type of data, as shown in Figure 3.44.

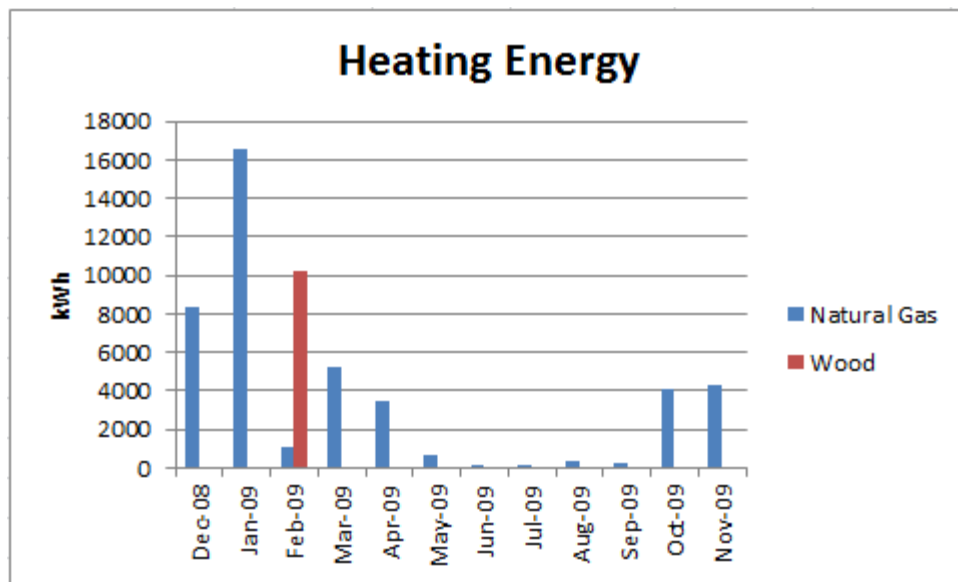


Figure 3.44. Two sets of data for each category

- Observe that by adding the heat energy from burning firewood in February, it could explain the anomaly that was identified in the previous chart of natural gas use.

There are many special chart types from which to choose. Some of them help communicate the interpretation of the chart better than more commonly used types. Do not use unusual chart types just for the sake of variety. People viewing the chart might be distracted from the chart's intended message.

- On the Chart Tools Design tab, in the Type group, click the Change Chart Type button. On the menu, under Column, choose Stacked Column, and then click the OK button. The stacked column chart adds the values for each column and shows how each contributes to the height of the column. Observe that the total heating energy for February is closer to what our world knowledge would lead us to expect, as shown in Figure 3.45.

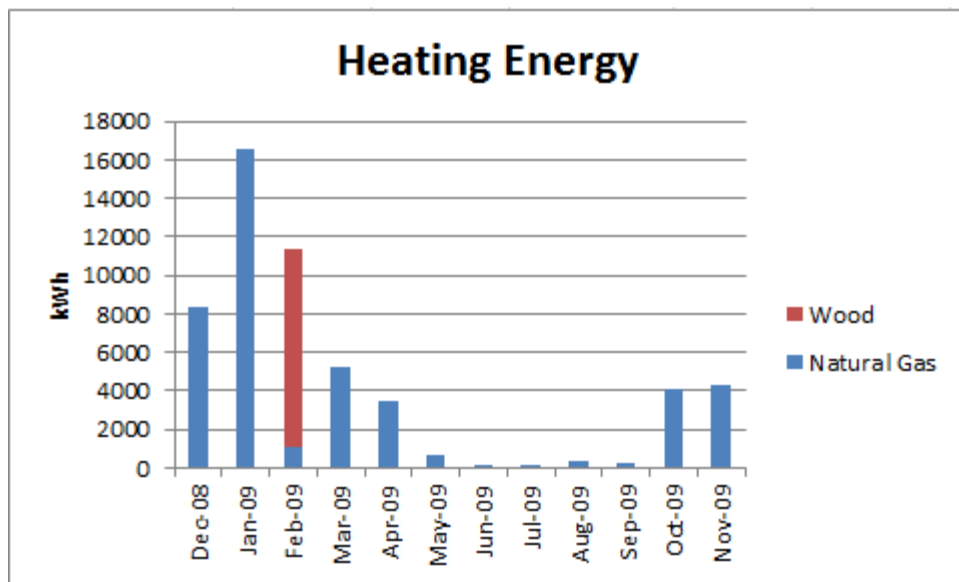


Figure 3.45. Stacked column chart shows contribution of firewood in February

Showing Contributions to a Whole

Column and line charts are used for showing trends. If the data represents numbers that sum to 100% of a category, the percentage of the total of each item can be represented by slices of a circle that represents a pie.

1. On the UtilityBills sheet, click cell N13. Type **Total** and then press Enter.
2. With cell N14 selected, on the Home tab, in the Editing group, click the Sum function button, Σ .

Confirm that the program correctly chose the cells to the left that contain numbers, B14:M14, and then click the button a second time. Fill this function down to cells N15:N17, as shown in Figure 3.46.

N14	=SUM(B14:M14)												
	B	C	D	E	F	G	H	I	J	K	L	M	N
12	Utility Bills (KWH)												
13	Dec-08	Jan-09	Feb-09	Mar-09	Apr-09	May-09	Jun-09	Jul-09	Aug-09	Sep-09	Oct-09	Nov-09	Total
14	1,461	922	1,105	951	903	647	706	803	689	704	1,042	1,035	10,968
15	308	205	274	266	395	153	281	295	238	235	313	244	3,207
16	8396	16519	1057	5194	3473	664	151	30	393	302	4047	4258	44,485
17	0	0	10257	0	0	0	0	0	0	0	0	0	10,257

Figure 3.46. Energy use summed by category for the entire year

3. Select cells A14:A17 and N14:N17. Insert a 3-D Pie chart. Drag the pie chart down to row 53 to the

right of the Heating Energy chart. Add the title **Annual Energy Use** and then compare the chart to Figure 3.47.

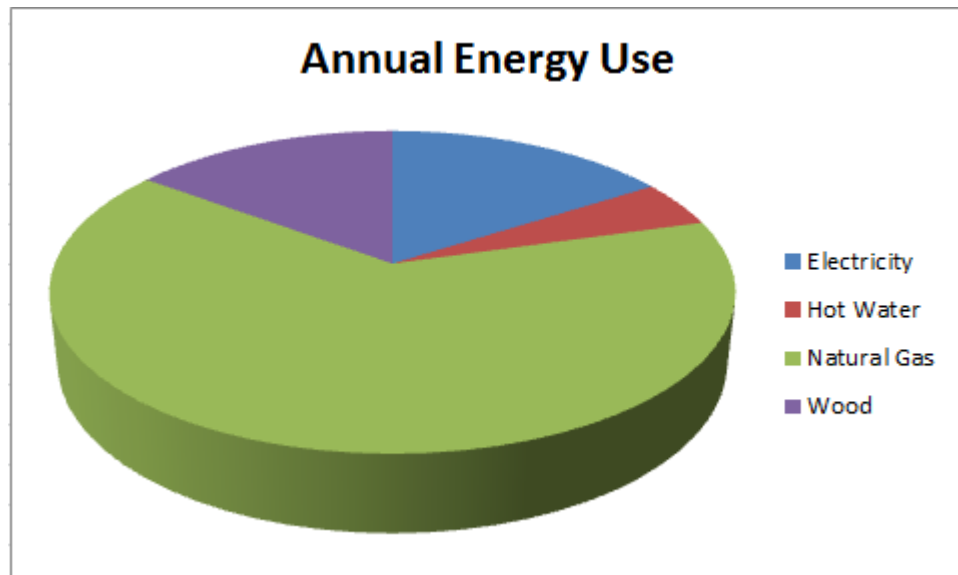


Figure 3.47. Pie chart shows contributions to the whole

Legends have limited usefulness when the chart is printed in black and white because the colors are rendered as shades of gray. For pie charts, it is a better idea to place the labels onto a slice if it is large enough or next to it with a leader line to connect the label to small slices. If there are several small slices, the pie can be rotated so the small slices are on the right side of the chart where there is more room for the labels.

4. On the Design tab, in the Chart Layouts group, click the Layout 1 option. The labels are placed on the pie slices where there is enough room and a percentage is calculated and displayed.
5. On the Layout tab, in the Background group, click the 3-D Rotation button. The Format Chart Area dialog box displays with the 3-D Rotation option selected. Use the spinner arrows to rotate the chart 60 degrees clockwise in the X direction, as shown in Figure 3.48.

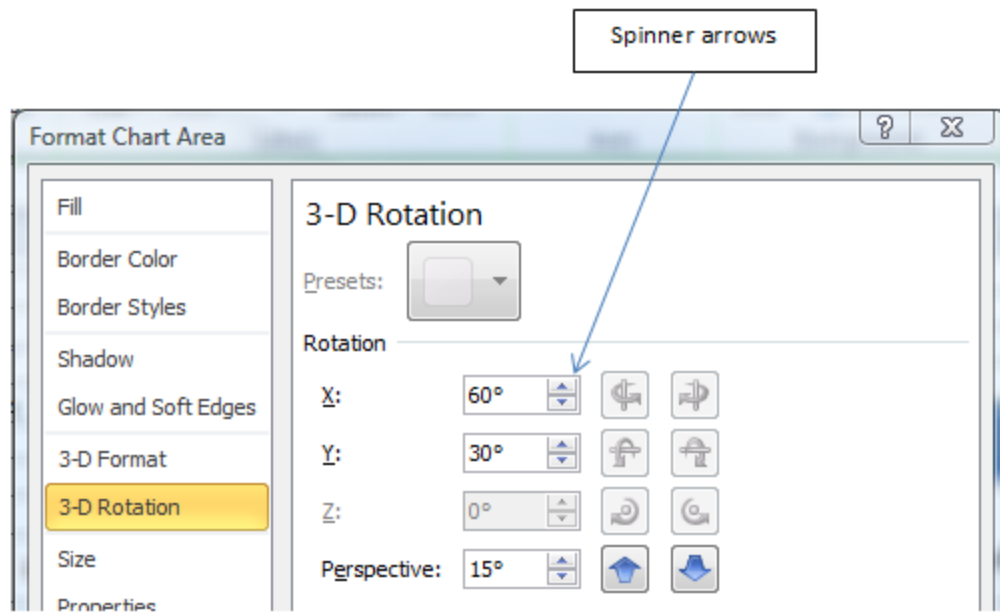


Figure 3.48. Rotate chart to place narrow slices on the right

6. At the bottom of the dialog box, click the Close button.

Use contrasting colors for labels that appear on top of colored chart elements to make them easier to see.

7. Click one of the chart labels to select all of them. Right-click the Wood label to select it.. On the shortcut menu, click Format Data Labels. The Format Data Labels dialog box displays.
8. On the left side of the dialog box, click Fill. Under Fill, click Solid fill. Click the Color button, and then select White. Drag the transparency slide to 25%. Refer to Figure 3.49.

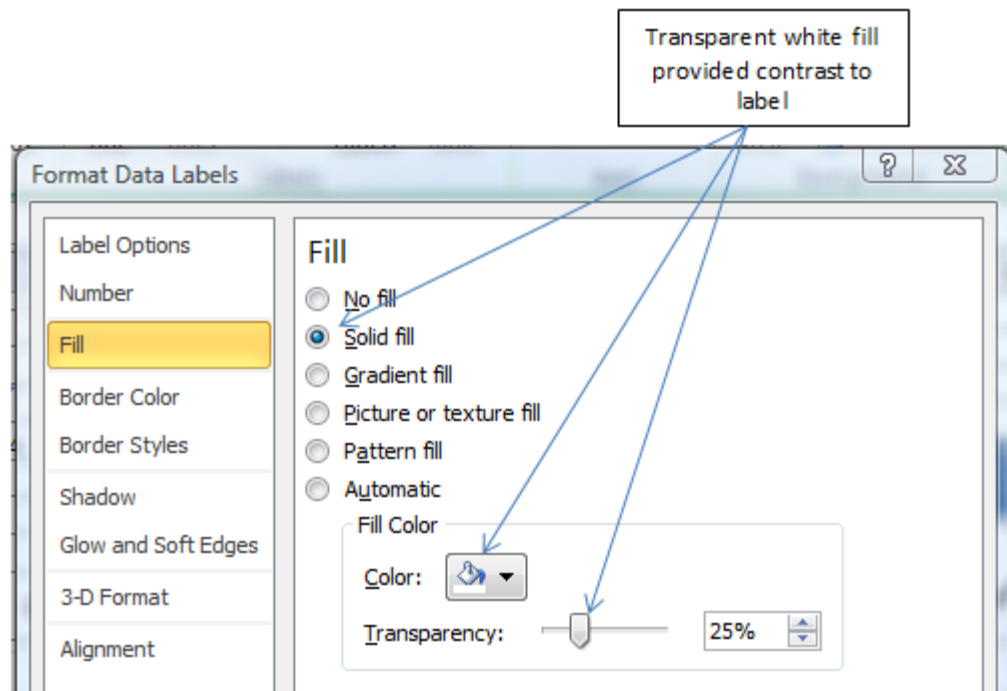


Figure 3.49. Contrast between the label and its background is necessary

9. Click the Close button. With the chart selected, on the Design tab in the Location group, click the Move Chart button. In the Move Chart dialog box, click the New Sheet button and type **Pie**. Click OK. This places the chart on its own Chart sheet.
10. Drag the Pie sheet tab to the right of the UtilityBills sheet tab. Refer to Figure 3.50.

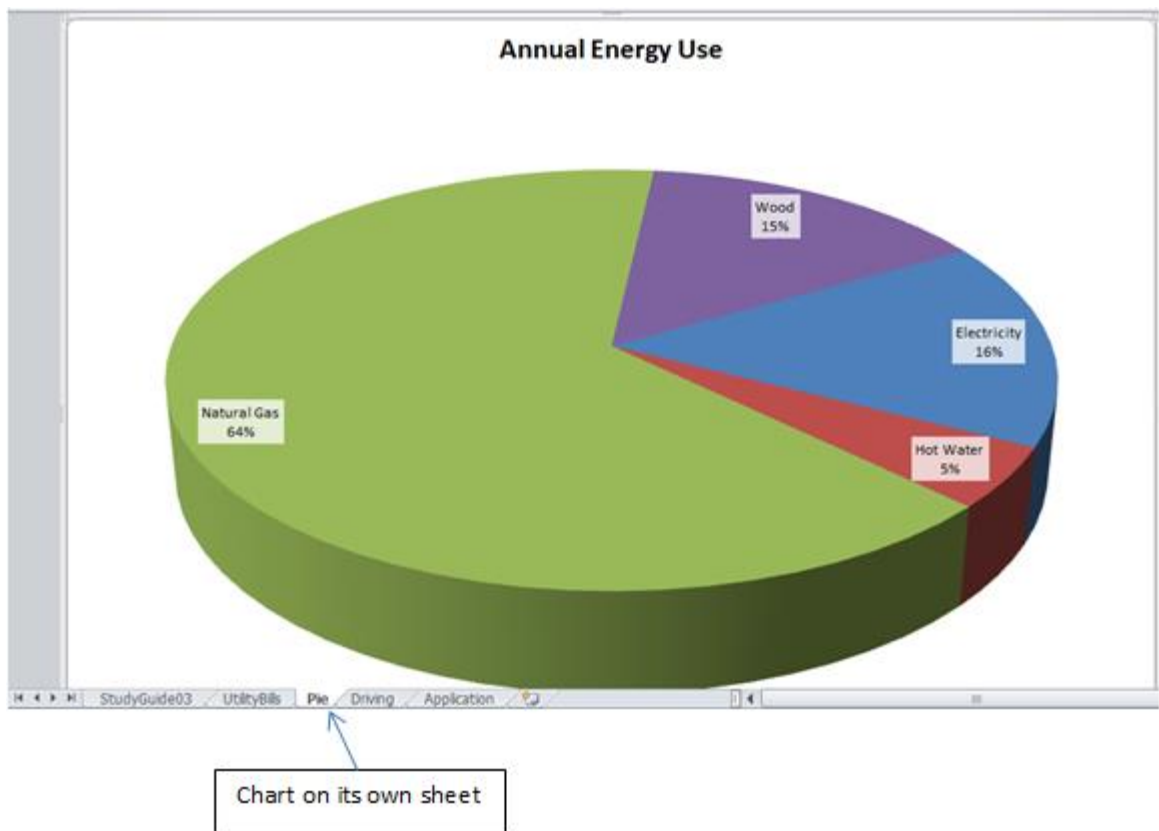


Figure 3.50. Chart transferred to a separate sheet in the workbook

11. Save the workbook and submit it as directed by the instructor.

Key Takeaways

- A chart with a trend line that is horizontal with an R^2 value close to zero indicates that the values charted on the vertical axis do not depend on the values charted on the X axis. [3.5.1]
- Use line or column charts that show trends or patterns of variation where data points that are not consistent with the pattern are easily recognized. Draw attention to the anomalous data by inserting a callout using the Shapes button in the Illustrations group on the Insert tab. [3.5.2]
- Show contributions to the whole using a pie chart. Place the labels on the chart instead of using a legend for most situations, especially if the chart will be printed in black and white. Format the fill color of the label, if necessary, to provide contrast between the font color and the background. [3.5.3]
- Place a chart on a separate worksheet using the Move Chart button in the Location group, on the

Chart Tools/Design tab. Rearrange the sheet tabs by clicking and dragging the name of the sheet tab. [3.5.4]

Key Terms

Bar chart

Displays values in different categories with horizontal rectangles.

Chart

Graphical representation of data.

Column chart

Displays values in different categories with vertical rectangles.

Delta

Greek letter: capital Δ , lower case δ used in equations to mean change in.

Dependent variable

Number whose value is determined by an independent variable.

Extrapolation

Estimating values of a dependent variable beyond other known data points.

Function

Relationship between an independent variable and a dependent variable.

Independent variable

Number used to calculate the value of other variables.

Intercept

The value on the vertical axis where a line crosses it.

Interpolation

Estimating values of a dependent variable between other known data points.

Line chart

Displays relationship between an independent variable and a dependent variable with a line between data points.

Pie chart

Displays how much each data point contributes to the whole.

R-squared (R^2) value

Statistic used to describe the fit of a trend line and the degree of dependence of the y variable on the x variable.

Scatter chart

Displays data points that are located using an independent and dependent variable. The intervals of the horizontal axis are uniform.

Slope

The pitch of a line.

World knowledge

General knowledge of events, human behavior, and science.

X-axis

Horizontal side of a chart.

Y-axis

Vertical side of a chart.

Last edited on December 17, 2010