

Lesson

6-7A

Describing Patterns
in Data

Vocabulary

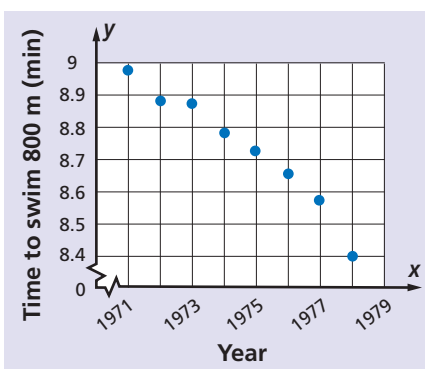
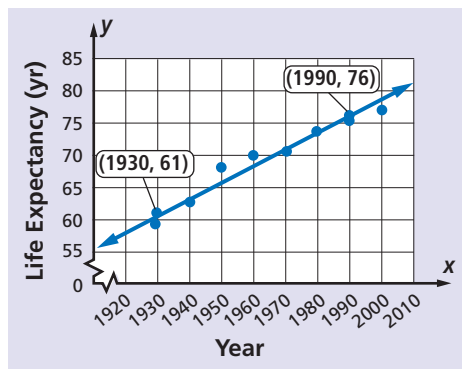
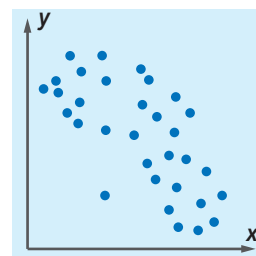
linear association
positive association
negative association
nonlinear association
outlier

► **BIG IDEA** Some patterns in numerical data can be seen by using a scatterplot.

Linear Associations

Scatterplots can be used to investigate patterns of association between two quantities. When the data points are close to being linear, the data represent a **linear association**. The more the data points tend to cluster around the line of best fit (whether the line of best fit is drawn or not), the stronger the relationship between the two variables. The graphs below both show a strong linear association. In contrast, the graph at the right shows a weak linear association because the data points do not tend to cluster very close to the line of best fit.

The two scatterplots below were in Lesson 6-7. The scatterplot below left appeared in Activity 1. In general, as the last 100 years have progressed (increasing values of x), life expectancies tended to increase (increasing values of y). Question 5 asked you to make the scatterplot below right. Notice that in later years (increasing values of x), the world record swimming times decrease (decreasing values of y).

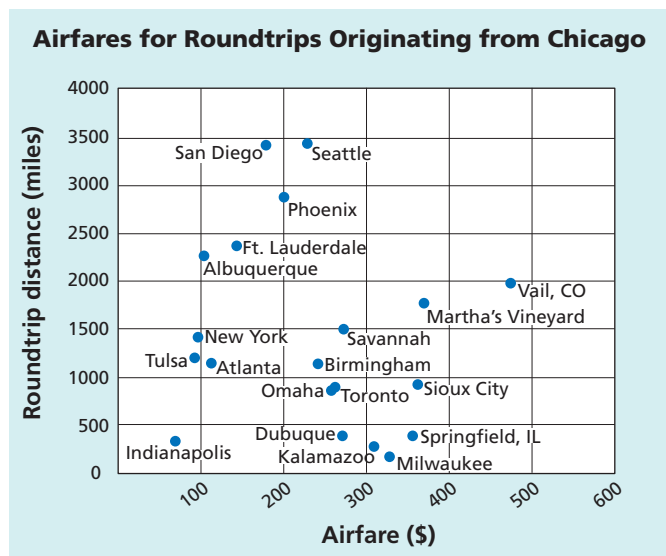


Patterns of association between two variables can be categorized as follows:

- There is a **positive association** when greater values of one variable tend to be associated with greater values of the other variable. The graph of life expectancy above shows a positive association. The line of best fit has a positive slope.

- There is a **negative association** when greater values of one variable tend to be accompanied by lesser values of the other variable. The graph of world swimming records on the previous page shows a negative association. The line of best fit would have a negative slope.

Sometimes the values of one variable are not associated with values of the other variable. When the data points in a scatterplot show no pattern, there is no association.



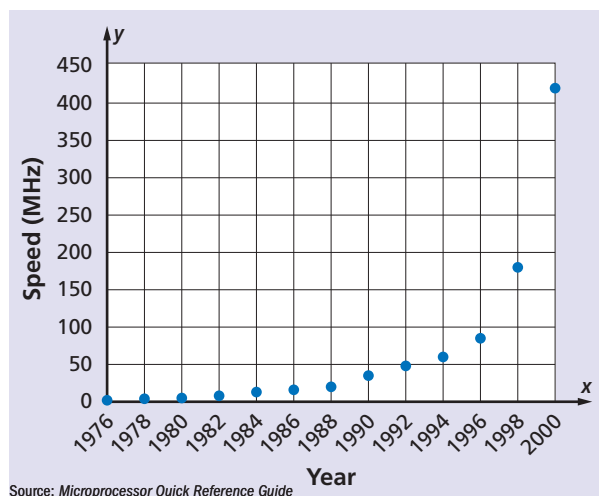
Source: <http://www.farecompare.com/maps/myAirportCategoryMap.html?departure=CHI&category=Domestic&sort=name>

The scatterplot above shows selected airfares for roundtrips originating from Chicago. There seems to be no pattern between the distance traveled and the airfare. Therefore, there is no association between the variables.

STOP QY1

Nonlinear Associations

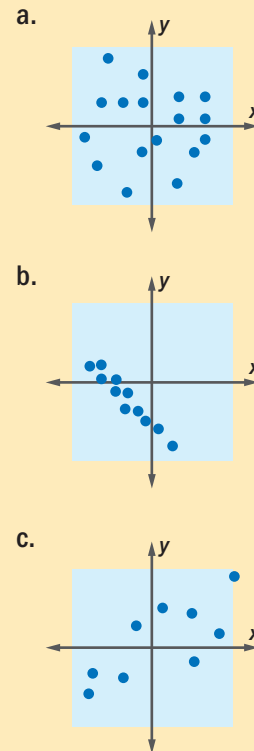
Sometimes the data in a scatterplot show a pattern that is *not* linear. The graph at the right shows a **nonlinear association** because the relationship between the variables is better modeled by a curve than a line. The graph shows the average processing speed of a computer (measured in megahertz) and the year it was made. You will be analyzing this graph and others similar to it in Lesson 7-4.



Source: Microprocessor Quick Reference Guide

► QY1

For each scatterplot, tell whether there is a positive association, a negative association, or no association. If there is an association, tell if the association is strong or weak.

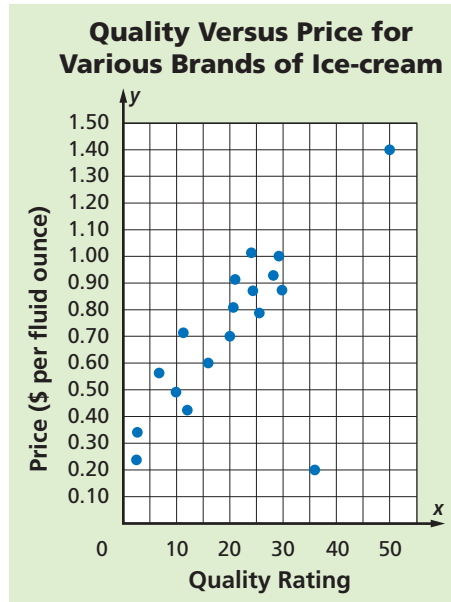


STOP QY2

Outliers

A scatterplot not only provides information about a possible relationship between two variables, but it also can show if there are any data points that appear to fall outside the overall pattern of data. Such a data point is called an **outlier**. You can use visual inspection to determine an outlier from a scatterplot.

In the scatterplot at the right, the data point located at about (36, 0.20) is an outlier because it has a very high quality rating compared to other brands at a similar price.



STOP QY3

GUIDED

Example

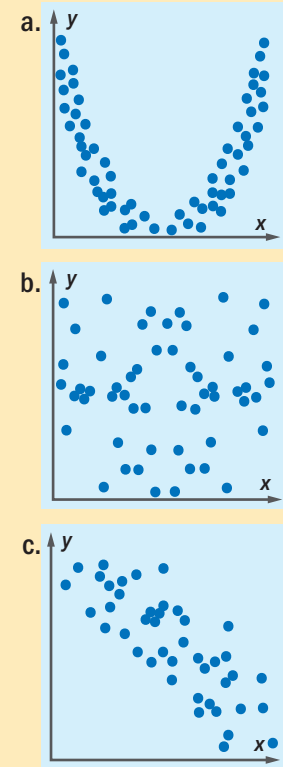
- a. Make a scatterplot based on the data below. Graph Production along the x-axis and graph Consumption along the y-axis.

Production (millions of barrels)	Consumption (millions of barrels)	Country
4.172	1.809	Iran
3.289	2.511	Canada
3.001	2.078	Mexico
2.798	0.435	United Arab Emirates
2.572	2.640	Brazil
2.472	0.740	Venezuela
2.399	0.687	Iraq
0.133	4.363	Japan
0.879	2.980	India
1.502	1.699	United Kingdom

- b. Categorize the association between the variables as positive association, negative association, or no association.
- c. If there is an association, is it strong or is it weak? Explain.

QY2

For each scatterplot, tell whether there is a linear association, a nonlinear association, or no association.



QY3

In the graph about ice-cream ratings, the data point (50, 1.40) has much greater x- and y-values than the other data points. Why is it not an outlier?

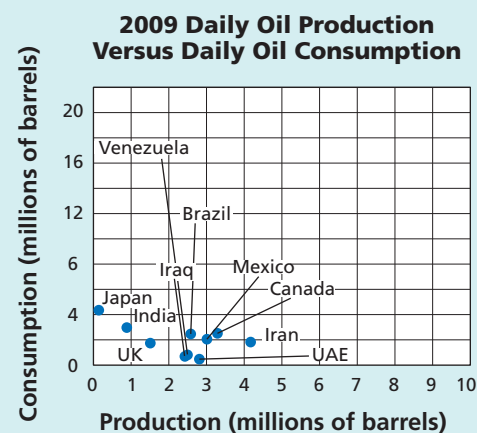
- d. Now graph the following additional data points in your scatterplot.

2009 Daily Oil Production Versus Daily Oil Consumption		
Production (millions of barrels)	Consumption (millions of barrels)	Country
9.932	2.85	Russia
9.764	2.43	Saudi Arabia
9.056	18.69	United States
2.383	13.68	European Union

- e. Explain how these additional data points affect your answers to Parts b and c.

Solution

- The scatterplot is shown at the right.
- Greater values of x tend to be associated with _____ values of y , so **there is (a) _____ association.**
- The data points are _____ around the line of best fit, so **this is a _____ association.**
- Graph the four additional data points on your graph.
- Describe what has happened to the pattern of data points.



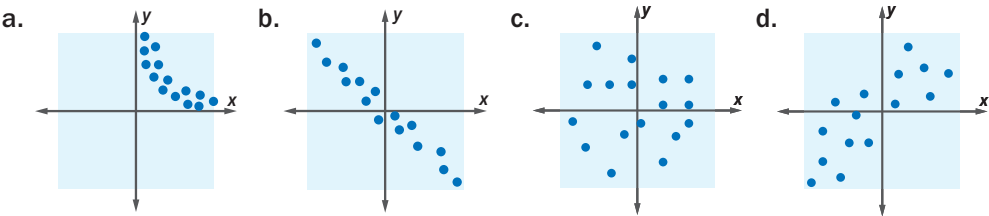
Questions

COVERING THE IDEAS

In 1–4, Fill in the Blank

- When greater values of one variable tend to be associated with lesser values of the other variable, there is a _____ association.
- When greater values of one variable tend to be associated with greater values of the other variable, there is a _____ association.
- When the data points in a linear association cluster tightly around the line of best fit, there is a _____ association.
- When there is a positive association, the line of best fit has a _____ slope.
- Explain why there is no association between the two variables in the graph “Airtfares for Roundtrips Originating from Chicago” in the lesson.
- Refer to the graph of average computer processing speed.
 - Find the approximate rate of speed increase from 1976 to 1988.
 - Repeat Part a for the years from 1996 to 2000.
 - What do your answers from Parts a and b indicate about the relationship between year and computer clock speed?

- 7. Refer to the graph you made of oil production versus consumption. Name four outliers and explain why they can be classified as such.
- 8. For each scatterplot, tell if there is a linear association, a nonlinear association, or no association. If there is a linear association, tell if it is positive or negative, and if it is strong or weak.



9. True or False

- a. The x - and y -values of an outlier’s coordinates are both much greater than those of the other data points.
- b. The x - and y -values of an outlier’s coordinates are both much less than those of the other data points.
- c. The data point of an outlier does not fit the overall pattern of the other data points in the distribution.

APPLYING THE MATHEMATICS

- 10. Refer to the scatterplot of “Airfares for Roundtrips Originating from Chicago” in the lesson.
 - a. Which destination provides the most miles for the least airfare?
 - b. Which destination provides the fewest miles for the greatest airfare?
- 11. The table at the right shows the number of books read during the summer for people of various ages.
 - a. Give the dimensions of an appropriate scatterplot window on your graphing utility.
 - b. Draw a scatterplot with Age along the x -axis and Number of Books along the y -axis.
 - c. For these data, is there an association between age and the number of books read during the summer? Explain your answer.

Number of Books Read During the Summer	
Age	Number of Books
5	15
5	3
8	10
9	11
10	6
11	8
11	21
13	15
14	5
18	0
18	18
19	13
22	6
25	9
25	1
30	6
35	13
50	4
60	3
60	7
70	12
90	3

12. Use a graphing utility to make a scatterplot based on the Expected Life Span data at the right. Is the association linear, nonlinear, or is there no association?
13. A student was asked to estimate the number of calories in one serving of some of her favorite foods. Then the student was asked to research the actual number of calories in those foods. The results are shown in the table below.

Food	Calories	
	Student Estimate	Actual
Apple juice	120	115
Bacon	385	110
Beef, ground	160	230
Blueberry pie	800	380
Cheese pizza	230	290
Chicken, fried drumstick	200	195
Chocolate chip cookie	150	185
Corn flakes	85	110
Egg, poached	100	75
Kiwi	60	45
Low-fat yogurt	90	145
Margarine	70	100
Marshmallows	20	90
Peach	50	35
Pecan halves	115	720
Potato chips	115	105
Provolone cheese	85	100
Red salmon, baked	195	140
Spinach, raw	75	10
Waffle, home recipe	130	245
Wheat flour, sifted	40	420
White bread	145	65

- Make a scatterplot for the data. Graph “Estimated Calories” along the x -axis; graph “Actual Calories” along the y -axis.
- Categorize the association between the variables as positive association, negative association, or no association.
- If there is an association, is it strong or is it weak? Explain.
- Give the approximate ordered pairs of any outliers.

Current Age	Expected Life Span
1	78.4
10	78.5
20	78.8
30	79.3
40	79.9
50	80.9
60	82.5
70	85.1
80	89.1

Source: World Almanac and Book of Facts, 2008

QY ANSWERS

- No association
 - Strong negative association
 - Weak positive association
- Nonlinear association
 - No association
 - Linear association
- Even though (50, 1.40) is far away from other points, it still follows the overall pattern of the data.