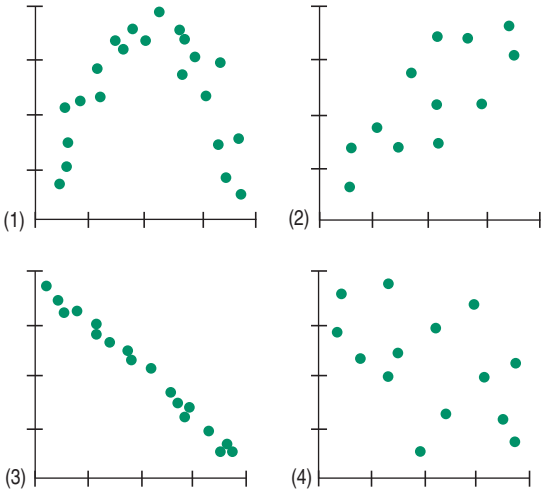
EXERCISES



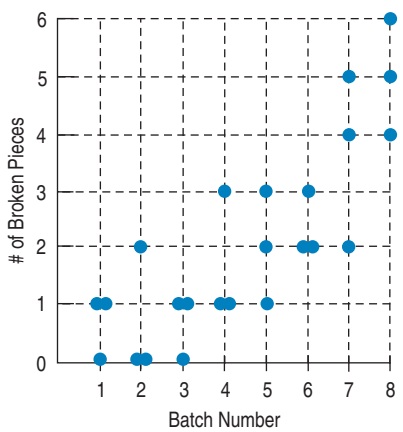
3. Association. Suppose you were to collect data for each pair of variables. You want to make a scatterplot. Which variable would you use as the explanatory variable and which as the response variable? Why? What would you expect to see in the scatterplot? Discuss the likely direction, form, and strength.

a) When climbing mountains: altitude, temperature
b) For each week: ice cream cone sales, air-conditioner sales
c) People: age, grip strength
d) Drivers: blood alcohol level, reaction time

5. Scatterplots. Which of the scatterplots at the top of the next column show

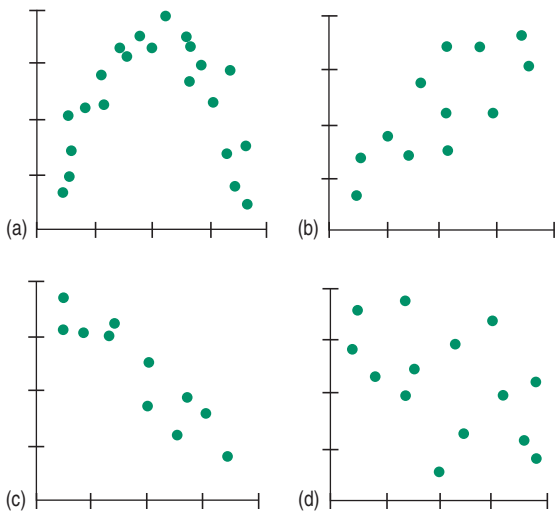
a) little or no association?
b) a negative association?
c) a linear association?
d) a moderately strong association?
e) a very strong association?

9. Firing pottery. A ceramics factory can fire eight large batches of pottery a day. Sometimes a few of the pieces break in the process. In order to understand the problem better, the factory records the number of broken pieces in each batch for 3 days and then creates the scatterplot shown.

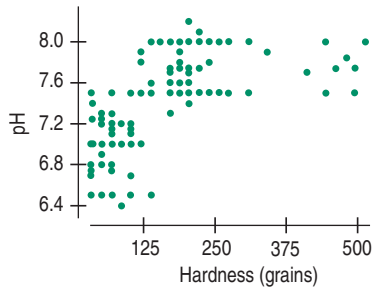


a) Make a histogram showing the distribution of the number of broken pieces in the 24 batches of pottery examined.
b) Describe the distribution as shown in the histogram. What feature of the problem is more apparent in the histogram than in the scatterplot?
c) What aspect of the company's problem is more apparent in the scatterplot?

11. Matching. Here are several scatterplots. The calculated correlations are −0.923, −0.487, 0.006, and 0.777. Which is which?



17. Hard water. In a study of streams in the Adirondack Mountains, the following relationship was found between the water's pH and its hardness (measured in grains):



Is it appropriate to summarize the strength of association with a correlation? Explain.

21. Prediction units. The errors in predicting hurricane tracks (examined in this chapter) were given in nautical miles. An ordinary mile is 0.86898 nautical miles. Most people living on the Gulf Coast of the United States would prefer to know the prediction errors in miles rather than nautical miles. Explain why converting the errors to miles would not change the correlation between *Prediction Error* and *Year.*

23. Correlation errors. Your Economics instructor assigns your class to investigate factors associated with the gross domestic product (*GDP*) of nations. Each student examines a different factor (such as *Life Expectancy, Literacy Rate,* etc.) for a few countries and reports to the class. Apparently, some of your classmates do not understand Statistics very well because you know several of their conclusions are incorrect. Explain the mistakes in their statements below.

a) "My very low correlation of −0.772 shows that there is almost no association between *GDP* and *Infant Mortality Rate.*"

b) "There was a correlation of 0.44 between *GDP* and *Continent.*"

25. Height and reading. A researcher studies children in elementary school and finds a strong positive linear association between height and reading scores.

a) Does this mean that taller children are generally better readers?

b) What might explain the strong correlation?

27. Correlation conclusions I. The correlation between *Age* and *Income* as measured on 100 people is $r = 0.75$. Explain whether or not each of these possible conclusions is justified:

a) When *Age* increases, *Income* increases as well.

b) The form of the relationship between *Age* and *Income* is straight.

c) There are no outliers in the scatterplot of *Income* vs. *Age.*

d) Whether we measure *Age* in years or months, the correlation will still be 0.75.

29. Baldness and heart disease. Medical researchers followed 1435 middle-aged men for a period of 5 years, measuring the amount of *Baldness* present (none = 1, little = 2, some = 3, much = 4, extreme = 5) and presence of *Heart Disease* (No = 0, Yes = 1). They found a correlation of 0.089 between the two variables. Comment on their conclusion that this shows that baldness is not a possible cause of heart disease.

Day 2

Vehicle	Horsepower	Highway Gas Mileage (mpg)
Audi A4	200	32
BMW 328	230	30
Buick LaCrosse	200	30
Chevy Cobalt	148	32
Chevy TrailBlazer	291	22
Ford Expedition	300	20
GMC Yukon	295	21
Honda Civic	140	40
Honda Accord	166	34
Hyundai Elantra	138	36
Lexus IS 350	306	28
Lincoln Navigator	300	18
Mazda Tribute	212	25
Toyota Camry	158	34
Volkswagen Beetle	150	30

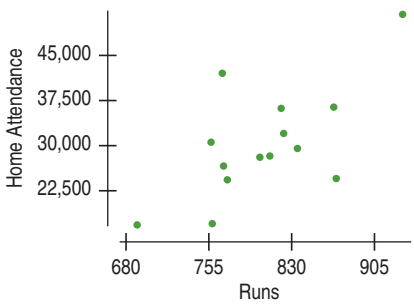
a) Make a scatterplot for these data.
b) Describe the direction, form, and strength of the plot.
c) Find the correlation between horsepower and miles per gallon.
d) Write a few sentences telling what the plot says about fuel economy.

T **33. Fuel economy 2007.** Here are advertised horsepower ratings and expected gas mileage for several 2007 vehicles. (http://www.kbb.com/KBB/ReviewsAndRatings)

35. Burgers. Fast food is often considered unhealthy because much of it is high in both fat and sodium. But are the two related? Here are the fat and sodium contents of several brands of burgers. Analyze the association between fat content and sodium.

Fat (g)	19	31	34	35	39	39	43
Sodium (mg)	920	1500	1310	860	1180	940	1260

37. Attendance 2006. American League baseball games are played under the designated hitter rule, meaning that pitchers, often weak hitters, do not come to bat. Baseball owners believe that the designated hitter rule means more runs scored, which in turn means higher attendance. Is there evidence that more fans attend games if the teams score more runs? Data collected from American League games during the 2006 season indicate a correlation of 0.667 between runs scored and the number of people at the game. (http://mlb.mlb.com)



a) Does the scatterplot indicate that it's appropriate to calculate a correlation? Explain.
b) Describe the association between attendance and runs scored.
c) Does this association prove that the owners are right that more fans will come to games if the teams score more runs?

39. Thrills. People who responded to a July 2004 Discovery Channel poll named the 10 best roller coasters in the United States. The table below shows the length of the initial drop (in feet) and the duration of the ride (in seconds). What do these data indicate about the height of a roller coaster and the length of the ride you can expect?

Roller Coaster	State	Drop (ft)	Duration (sec)
Incredible Hulk	FL	105	135
Millennium Force	OH	300	105
Goliath	CA	255	180
Nitro	NJ	215	240
Magnum XL-2000	OH	195	120
The Beast	OH	141	65
Son of Beast	OH	214	140
Thunderbolt	PA	95	90
Ghost Rider	CA	108	160
Raven	IN	86	90