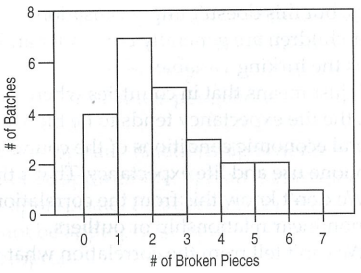
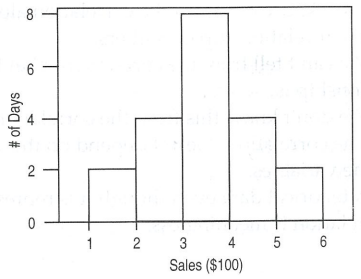
CHAPTER 7

1. a) Weight in ounces: explanatory; Weight in grams: response. (Could be other way around.) To predict the weight in grams based on ounces. Scatterplot: positive, straight, strong (perfectly linear relationship).
   b) Circumference: explanatory. Weight: response. To predict the weight based on the circumference. Scatterplot: positive, linear, moderately strong.
   c) Shoe size: explanatory; GPA: response. To try to predict GPA from shoe size. Scatterplot: no direction, no form, very weak.
   d) Miles driven: explanatory; Gallons remaining: response. To predict the gallons remaining in the tank based on the miles driven since filling up. Scatterplot: negative, straight, moderate.
2. a) Price: explanatory; Number sold: response. To predict the number sold based on the price. Scatterplot: negative, straight, moderate.
   b) Depth: explanatory; Water pressure: response. To predict water pressure based on depth. Scatterplot: positive, straight, strong.
   c) Visibility: explanatory; Depth: response. To predict depth based on visibility (although predicting visibility based on depth is also possible). Scatterplot: negative, straight(?), weak to moderate.
   d) Weight: explanatory; Reading score: response. To predict reading test scores based on weight. Scatterplot: positive, possibly straight, moderate.
3. a) Altitude: explanatory; Temperature: response. (Other way around possible as well.) To predict the temperature based on the altitude. Scatterplot: negative, possibly straight, weak to moderate.
   b) Ice cream cone sales: explanatory. Air-conditioner sales: response—although the other direction would work as well. To predict one from the other. Scatterplot: positive, straight, moderate.
   c) Age: explanatory; Grip strength: response. To predict the grip strength based on age. Scatterplot: curved down, moderate. Very young and elderly would have grip strength less than that of adults.
   d) Reaction time: explanatory; Blood alcohol level: response. To predict blood alcohol level from reaction time test. (Other way around is possible.) Scatterplot: positive, nonlinear, moderately strong.
4. a) Time: explanatory; Cost: response. To predict cost based on time. Scatterplot: positive, straight, strong.
   b) Time delay: explanatory; Distance: response. To predict the distance from the lightning based on the time delay of the thunder. Scatterplot: positive, straight, strong.
   c) Brightness: explanatory; Distance: response. To predict distance based on apparent brightness. Scatterplot: negative, curved, moderate.
   d) Weight of car: explanatory; Age of owner: response. To predict the age of the owner based on the weight of the car. (Or other way around.) Scatterplot: no direction, no shape, very weak.
5. a) None b) 3 and 4 c) 2, 3, and 4
   d) 1 and 2 e) 3 and possibly 1
6. a) 1 b) 4 c) 2 and 4 d) 3 e) 2 and 4
7. There seems to be a very weak—or possibly no—relation between brain size and performance IQ.
8. Nonlinear form. Moderately strong. The rate has not been constant; the rate of increase from the beginning to about 1950 is steeper than from 1950 to the present. One winner in the early 1890s was quite slow.
9. a)



   b) Unimodal, skewed to the right. The skew.
   c) The positive, somewhat linear relation between batch number and broken pieces.
10. a)



b) Positive linear relation, so sales are increasing.
c) Unimodal, symmetric sales; average sales around $350.

11. a) 0.006 b) 0.777 c) −0.923 d) −0.487

12. a) −0.977 b) 0.736 c) 0.951 d) −0.021

13. There may be an association, but not a correlation unless the variables are quantitative. There could be a correlation between average number of hours of TV watched per week per person and number of crimes committed per year. Even if there is a relationship, it doesn't mean one causes the other.

14. There may be an association, but not a correlation; type of car is not a quantitative variable.

15. a) Yes. It shows a linear form and no outliers.
b) There is a strong, positive, linear association between drop and speed; the greater the coaster's initial drop, the higher the top speed.

16. a) Yes. The plot shows a positive sign, linear form, and no outliers (though the last point on the right may be a bit out of line).
b) There's a strong, positive association. Experiments that showed a greater mean improvement among patients who took an antidepressant also showed a greater placebo effect.

17. The scatterplot is not linear; correlation is not appropriate.

18. The scatterplot is not linear, it has two high outliers (LA and NY), and there is a cluster of 8 cities that appear to be atypical. Correlation is not appropriate for such a relationship.

19. The correlation may be near 0. We expect nighttime temperatures to be low in January, increase through spring and into the summer months, then decrease again in the fall and winter. The relationship is not linear.

20. The relation might be nonlinear, resulting in a correlation close to 0 even though the association is strong, or there might be an outlier.

21. The correlation coefficient won't change, because it's based on *z*-scores. The *z*-scores of the prediction errors are the same whether they are expressed in nautical miles or miles.

22. The correlation coefficient won't change, because it's based on *z*-scores. The *z*-scores of the prediction errors are not changed by adding or subtracting a constant.

23. a) Assuming the relation is linear, a correlation of −0.772 shows a strong relation in a negative direction.
b) Continent is a categorical variable. Correlation does not apply.

24. a) Correlation cannot be greater than 1. There is an error.
b) Assuming the relation is linear, the strong correlation shows a relation, but it does not show *causality*.

25. a) Actually, yes, taller children will tend to have higher reading scores, but this doesn't imply causation.
b) Older children are generally both taller and are better readers. Age is the lurking variable.

26. a) No. It just means that in countries where cell phone use is high, the life expectancy tends to be high as well.
b) General economic conditions of the country could affect both cell phone use and life expectancy. That's the lurking variable.

27. a) No. We don't know this from the correlation alone. There may be a nonlinear relationship or outliers.
b) No. We can't tell from the correlation what the form of the relationship is.
c) No. We don't know from the correlation coefficient.
d) Yes, the correlation doesn't depend on the units used to measure the variables.

28. a) We don't know this from the correlation alone. There may be a nonlinear relationship or outliers.
b) No. We can't tell from the correlation what the form of the relationship is.
c) No. We don't know this from the correlation coefficient.
d) No. The correlation doesn't depend on the units used to measure the variables.

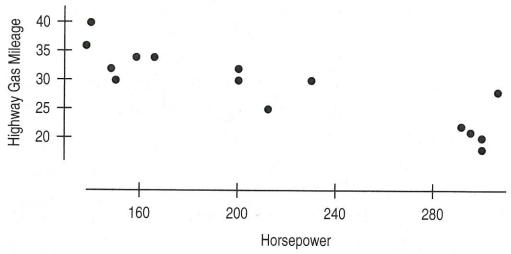
29. This is categorical data even though it is represented by numbers. The correlation is meaningless.

30. The source of data is categorical. The correlation is meaningless.

31. a) The association is positive, moderately strong, and roughly straight, with several states whose HCI seems high for their median income and one state whose HCI appears low given its median income.
b) The correlation would still be 0.65.
c) The correlation wouldn't change.
d) DC would be a moderate outlier whose HCI is high for its median income. It would lower the correlation slightly.
e) No. We can only say that higher median incomes are associated with higher housing costs, but we don't know why. There may be other economic variables at work.

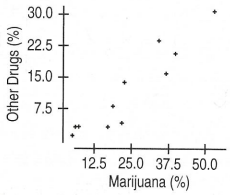
32. a) The association is negative, quite strong, and fairly straight; there are no outliers.
b) The correlation would still be −0.84.
c) The correlation wouldn't change.
d) That year would have a very high mortgage amount for an interest rate that high. That would tend to weaken the correlation (bring it closer to 0).
e) No. We can only say that lower interest rates are associated with larger mortgage amounts, but we don't know why. There may be other economic variables at work.

33. a)



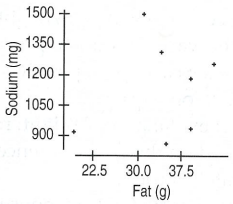
b) Negative, linear, strong. c) −0.869
d) There is a strong linear relation in a negative direction between horsepower and highway gas mileage. Lower fuel efficiency is associated with higher horsepower.

34. a)



b) 0.934
c) There is a strong linear relation in a positive direction between the percent of teens using marijuana and the percent of teens using other drugs.
d) There is no indication of causality. There could be lurking variables.

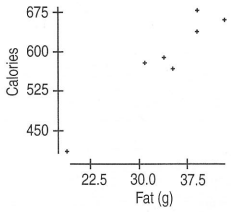
35.



(Plot could have explanatory and predictor variables swapped.) Correlation is 0.199. There does not appear to be a relation between sodium and fat content in burgers, especially without the low-fat, low-sodium item. The correlation of 0.199 shows a weak relationship, even with the outlier included.

36. Correlation is 0.961. There appears to be a strong linear relation in a positive direction. The correlation of 0.961 supports the conclusion

of a strong relation. Even without the outlier at 410 calories and 19 grams of fat, the correlation is 0.837, still strong.

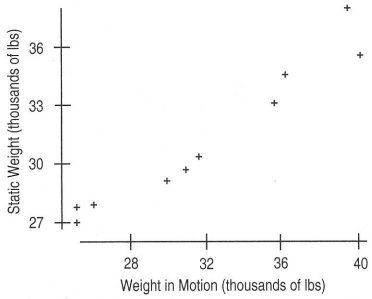


37. a) Yes, the scatterplot appears to be somewhat linear.
b) As the number of runs increases, the attendance also increases.
c) There is a positive association, but it does not *prove* that more fans will come if the number of runs increases. Association does not indicate causality.

38. a) There appears to be a moderately strong positive relationship between the number of wins and home attendance. The points show some significant scattering. The correlation of 0.697 is moderately strong.
b) Winning. The correlation is only slightly higher for wins vs. attendance (0.697) than for runs vs. attendance (0.667).
c) The correlation between runs and wins is 0.605.

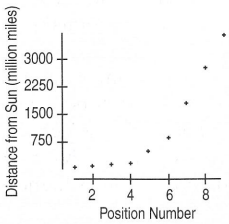
39. A scatterplot shows a generally straight scattered pattern with no outliers. The correlation between *Drop* and *Duration* is 0.35, indicating that rides on coasters with greater initial drops generally last somewhat longer, but the association is weak.

40. a)

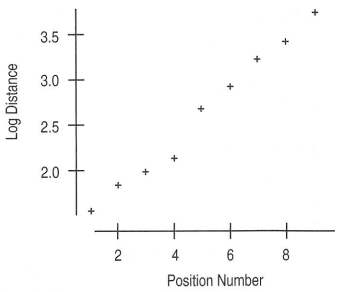


b) Positive, linear, very strong.
c) The new scale is able to predict the static weight fairly well, except possibly at the high end. It may be possible to predict the static weight from the new scale accurately enough to be useful. But the weight-in-motion measurements seem a bit too high.
d) 0.965.
e) The correlation would not change.
f) At the higher end of the weight-in-motion scale, there is one point where the weight in motion is much higher than the static weight. The line does not seem to go through the points where x and y are equal. The new scale may have to be recalibrated.

41. a)

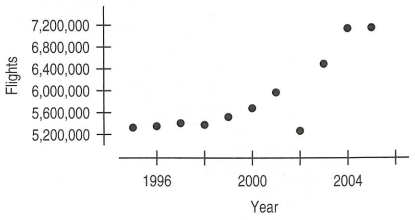


The relation between position and distance is nonlinear, with a positive direction. There is very little scatter from the trend.
b) The relation is not linear.
c)



The relation between position number and log of distance appears to be roughly linear.

42. a) 0.828
b)



The trend is positive and curved. There is a low outlier at 2002.
c) The plot is not straight and has an outlier. Either violation would disqualify the correlation. It isn't unusual for growth in a business to be faster than linear. The outlier in 2002 is due to the drop in airline flights after the 9/11 attacks.