

Chapter 4 – Understanding and Comparing Distributions

Section 4.1

1. Load factors, 2017.

The distribution of domestic load factors and the distribution of international load factors are both unimodal and skewed to the left. The distribution of international load factors may contain a low outlier. Because the distributions are skewed, the median and IQR are the appropriate measures of center and spread. The medians are very close, which tell us that typical international and domestic load factors are about the same. The IQRs show a bit more variability in the domestic load factors.

2. Load factors, 2017 by season.

The distribution of Spring/Summer load factors and the distribution of Fall/Winter load factors are both unimodal and skewed to the left. Load factors in the Spring/Summer period vary less than load factors in the Fall/Winter period, but are generally higher. The center of the distribution of Fall/Winter load factors is around 77.5, while the center of the distribution of Spring/Summer load factors is around 82.

3. Fuel economy.

If the Mitsubishi i-MiEV is removed, the standard deviation of fuel economy ratings would be much lower, dropping from 11.87 miles per gallon to 4.6 miles per gallon. The IQR would be affected very little, if at all.

4. Prisons 2014.

- a) The tallest bars on the histogram are where the vertical lines on the boxplot are closest together.
- b) The boxplot indicates a skewed distribution when the vertical lines on one side of the median are closer together than the vertical lines on the other side. The histogram indicates a skewed distribution when the bars on one side of the histogram are generally taller than the bars on the other side of the histogram.
- c) The histogram shows clusters and gaps. The boxplot does not show these.
- d) The boxplot shows the quartiles and the median, as well as showing outliers. These cannot be determined from the histogram.

5. Load factors 2017 by month.

Load factors are generally highest and least variable in the summer months (June–August). They are lower and more variable in the winter and spring. Several months have low outliers, and April has a very low outlier that is the minimum of all the data. (The low value is for April of 2003. At that time, there was concern over the deadly SARS virus, which may have depressed international travel.)

6. Load factors 2017 by year.

Load factors have increased since 2002. In recent years the growth has leveled off and the annual variation has decreased.

Section 4.2

7. Extraordinary months.

We are aware of a possible explanation for the depressed air travel that has nothing to do with general monthly or overall annual patterns. If we want to understand these patterns, it might be best to set these months aside.

8. Load factors 2017 by year again.

Outliers are context dependent. Nevertheless, April 2003 shows up in both views of the data. It looks like international travel was severely affected in that month.

38 Part I Exploring and Understanding Data

Section 4.3

9. Load factors 2017 over time.

- a) After a period of little change in 2000–2001, load factors have been increasing steadily.
- b) No. These are percentages, so they can't exceed 100%. Moreover, we have no reason to believe this trend will continue.

10. Load factors over time, a second look.

- a) With this smoother, the seasonal pattern is evident. Exercise 3 showed the pattern in detail with higher load factors in the summer months.
- b) Yes, we can expect this pattern to persist because it reflects seasonal effects such as summer vacation time that will probably continue, although the increase can't continue for very long because these are percentages.

11. Exoplanets.

It is difficult to summarize data with a distribution this skewed. The extremely large values will dominate any summary or description.

12. Exoplanets re-expressed.

- a) Yes, this re-expressed scale is better for understanding these distances. The log scale provides a nearly symmetric distribution, and points out that the sun was included in the data, probably accidentally.
- b) The sun should not be included in data about extra-solar planets.

Chapter Exercises

13. In the news.

Answers will vary.

14. Groups on the Internet.

Answers will vary.

15. Cost of living 2016, selected cities.

- a) Cappuccino is the most expensive commodity, in general. The first quartile of cappuccino prices is higher than all the water prices. The middle 50% of cappuccino prices is less variable than the middle 50% of egg prices. The third quartile and maximum value of cappuccino prices are both higher than those for egg prices.
- b) We can't say anything about price comparisons in individual cities from this display. The minimum price of a cappuccino is lower than the minimum price of a carton of eggs. It could be the case that those two prices are in the same city, and cappuccino costs less than eggs.

16. Cost of living 2016, more cities.

- a) A cappuccino is the most expensive, in general. The median price of a cappuccino is at the third quartile for a carton of eggs and above all the water prices except the high outliers.
- b) No, some of the water prices are high outliers and are more expensive than most of the egg prices.
- c) Yes, some cartons of eggs are more expensive than a cappuccino. Several high outliers of egg prices are above all of the cappuccino prices. We have no idea, however, whether or not this occurs in the same city, or if these prices come from different cities.

17. Rock concert deaths.

- a) The histogram and boxplot of the distribution of “crowd crush” victims’ ages both show the distribution is essentially symmetric, but very slightly skewed to the right with two high outliers at 36 and 48. Most victims are between the ages of 16 and 24.
- b) This histogram shows that there may have been two modes in the distribution of ages of “crowd crush” victims, one at 18–20 years of age and another at 22–24 years of age. Boxplots, in general, can show symmetry and skewness, but not features of shape like bimodality or uniformity. Most victims were between 16 and 24 years of age.
- c) The median is the better measure of center, since the distribution of ages has outliers. Median is more resistant to outliers than the mean.
- d) The IQR is a better measure of spread, since the distribution of ages has outliers. IQR is more resistant to outliers than the standard deviation.

18. Slalom times 2018.

- a) Both distributions show that the distribution of slalom times is skewed to the high end. The histogram and boxplot of the distribution of Men’s Giant Slalom times both show that a typical time of around 175 seconds, that the range of slalom times is about 55 seconds.
- b) Since the distribution of slalom times is skewed, and contains possible outliers, the median is the better summary of center.
- c) In the presence of skewness and possible outliers, we’d prefer the IQR to the standard deviation as a measure of spread.

19. Sugar in cereals.

- a) The maximum sugar content is approximately 60% and the minimum sugar content is approximately 1%, so the range of sugar contents is about $60 - 1 = 59\%$.
- b) The distribution of sugar content of cereals is bimodal, with modes centered around 5% and 45% sugar by weight.
- c) Some cereals are healthy, low-sugar brands, and others are very sugary.
- d) Yes. The minimum sugar content in the children’s cereals is about 35% and the maximum sugar content of adult cereals is only 34%.
- e) The range of sugar contents is about the same for the two types of cereals, approximately 28%, but the IQR is larger for the adult cereals. This is an indication of more variability in the sugar content of the middle 50% of adult cereals.

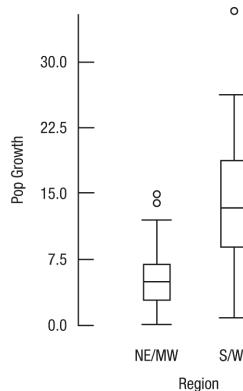
20. Tendon transfers.

- a) The distribution of pushing strength scores is unimodal and symmetric.
- b) The maximum pushing strength score is 4 and the minimum is 1, so the range is a score of $4 - 1 = 3$.
- c) The histogram does not show that the results of the two procedures typically resulted in different strengths.
- d) The distribution of biceps transfer strength scores had a higher median than the distribution of deltoid transfer strength scores.
- e) The biceps transfer was not always the best. The highest strength score in the deltoid transfer was higher than the lowest 25% of biceps transfer strength scores.
- f) The deltoid transfer produced more consistent strength scores. The IQR is much smaller for this group, even though the ranges have approximately the same value of 2.

40 Part I Exploring and Understanding Data

21. Population growth 2010, by region.

- a) Comparative boxplots are shown below.



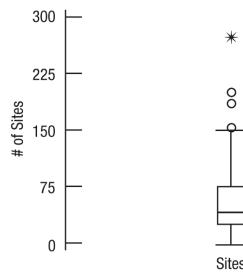
- b) Growth rates in NE/MW states are tightly clustered near 5%, but with two high outliers. S/W states are more variable, with a median near 13% and two outliers as well. The distributions are fairly symmetric.

22. Camp sites.

- a) The distribution of the number of campsites in public parks in Vermont is strongly skewed to the right, so median and IQR are appropriate measures of center and spread.
- b) $IQR = Q3 - Q1 = 78 - 28 = 50$.
Upper Fence: $Q3 + 1.5(IQR) = 78 + 1.5(50) = 78 + 75 = 153$
Lower Fence: $Q1 - 1.5(IQR) = 28 - 1.5(50) = 28 - 75 = -47$, which is well below 0 campsites.

There are 3 parks with greater than 180 campsites. These are definitely outliers. There are 2 parks with between 150 and 160 campsites each. These may be outliers as well.

- c) A boxplot of the distribution of number of campsites is shown below.



- d) The distribution is unimodal with a strong skew to the right. There are several outliers past the $1.5 \times IQR$ upper fence of 153 camp sites. The median number of camp sites is 43.5 sites. The mean is 62.8 sites. The mean is larger than the median because it has been influenced by the strong skew and the outliers.

23. Hospital stays.

- a) The histograms of male and female hospital stay durations would be easier to compare if they were constructed with the same scale, perhaps from 0 to 20 days.
- b) The distribution of hospital stays for men is skewed to the right, with many men having very short stays of about 1 or 2 days. The distribution tapers off to a maximum stay of approximately 25 days. The distribution of hospital stays for women is skewed to the right, with a mode at approximately 5 days, and tapering off to a maximum stay of approximately 22 days. Typically, hospital stays for women are longer than those for men.
- c) The peak in the distribution of women's hospital stays can be explained by childbirth. This time in the hospital increases the length of a typical stay for women, and not for men.

24. Deaths 2014.

- a) The distributions of ages at death of black Americans and white Americans are both skewed unimodal and skewed to the left, towards the lower ages of death. The White distribution is strongly skewed to the left with a small peak at the low end that is infant mortality. The Black distribution is also skewed to the left, although it has three peaks: one at infancy, then at 60 to 64, and another at 85 and older.
- b) The distributions of death ages differ mainly in their spreads and centers. The age at death for black Americans is more variable than the age at death for white Americans. A greater proportion of black Americans die between the ages of 25 and 74 than do white Americans, while a greater proportion of white Americans dies at ages older than 74 than do black Americans. The distribution of death ages of black Americans shows a higher infant mortality rate than the distribution of death ages of white Americans.
- c) The interval widths are not consistent. Most of the bars are 5 years wide, but the first bar is 1 year wide, the second bar is 4 years wide, and the last bar has width “85 and older”.

25. Women’s basketball.

- a) Both girls have a median score of about 17 points per game, but Scyrine is much more consistent. Her IQR is about 2 points, while Alexandra’s is over 10.
- b) If the coach wants a consistent performer, she should take Scyrine. She’ll almost certainly deliver somewhere between 15 and 20 points. But, if she wants to take a chance and needs a “big game”, she should take Alexandra. Alex scores over 24 points about a quarter of the time. On the other hand, she scores under 11 points about as often.

26. Gas prices 2018.

- a) Gas prices rose during the first nine years of the century, then dropped precipitously in 2008. They then resumed their rise, peaking in 2012. After that they have fallen and then recovered somewhat. The year 2008 was the most variable. Variability was small in 2017 and 2018.
- b) 2008 had both the greatest range and the largest IQR. The years 2014 and 2015 also show large variation—possibly less clear because it is spread across two years.

27. Marriage age 2016.

The distribution of marriage age of U.S. men is skewed right, with a typical man (as measured by the median) first marrying at around 24 years old. The middle 50% of male marriage ages is between about 23 and 26 years. For U.S. women, the distribution of marriage age is also skewed right, with median of around 21 years. The middle 50% of female marriage age is between about 20 and 23 years. When comparing the two distributions, the most striking feature is that the distributions are nearly identical in spread, but have different centers. Females typically seem to marry about 2 years younger than males. In fact, between 50% and 75% of the women marry at a younger age than any man.

28. Fuel economy and cylinders.

Cars with 4 cylinders generally get better gas mileage than cars with 6 cylinders, which generally get better gas mileage than cars with 8 cylinders. Additionally, the greater the number of cylinders, the more consistent the mileage becomes. 4 cylinder cars typically get between 27 – 33 mpg, 6 cylinder cars typically get between 18 – 22 mpg, and 8 cylinder cars typically get between 16 – 19 mpg. We don’t have enough data to compare 5-cylinder cars.

29. Marriage age 2016 over time.

- a) They are time series because they report values over time. However, the values are not all equally spaced because the early values are reported only every decade, while later values are annual.
- b) Age at first marriage declined in the first part of the 20th century, but has been increasing for both men and women since about 1960. Throughout more than a century of data, men have typically been older at first marriage than women.
- c) The increase in age cannot continue indefinitely. The pattern that men tend to be older than women at first marriage may well continue.

42 Part I Exploring and Understanding Data

30. Ozone.

- a) April had the highest recorded ozone level, approximately 440.
- b) February had the largest IQR of ozone level, approximately 50.
- c) August had the smallest range of ozone levels, approximately 50.
- d) January had a slightly lower median ozone level than June, 340 and 350, respectively, but June's ozone levels were much more consistent.
- e) Strong seasonal pattern with low consistent ozone concentrations in later summer/early fall and high variable concentrations in early spring. The medians follow a cyclic pattern, rising from January to April, then falling to October and rising again from October to December.

31. Test scores.

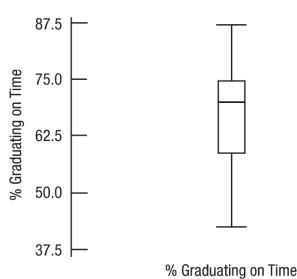
Class A is Class 1. The median is 60, but has less spread than Class B, which is Class 2. Class C is Class 3, since its median is higher, which corresponds to the skew to the left.

32. Eye and hair color.

The graph is not appropriate. Boxplots are for quantitative data, and these are categorical data, although coded as numbers. The numbers used for hair color and eye color are arbitrary, so the boxplot and any accompanying statistics for eye color make no sense.

33. Graduation?

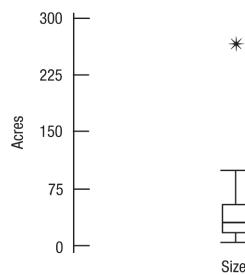
- a) The distribution of the percent of incoming college freshman who graduate on time is probably slightly left skewed. The mean is slightly below the median and the 25th percentile is farther from the median than the 75th percentile
- b) Upper Fence: $Q3 + 1.5(IQR) = 74.75 + 1.5(74.75 - 59.15) = 74.75 + 23.4 = 98.15$
Lower Fence: $Q1 - 1.5(IQR) = 59.15 - 1.5(74.75 - 59.15) = 59.15 - 23.4 = 35.75$
Since the maximum value of the distribution of the percent of incoming freshmen who graduate on time is 87.4% and the upper fence is 98.15%, there are no high outliers. Likewise, since the minimum is 43.2% and the lower fence is 35.75%, there are no low outliers. Since the minimum and maximum percentages are within the fences, all percentages must be within the fences.
- c) A boxplot of the distribution of the percent of incoming freshmen who graduate on time is shown below.



- d) The distribution of the percent of incoming freshmen who graduate on time is roughly symmetric, with mean of approximately 68% of freshmen graduating on time. Universities surveyed had between 43% and 87% of students graduating on time, with the middle 50% of universities reporting between 59% and 75% graduating on time.

34. Vineyards.

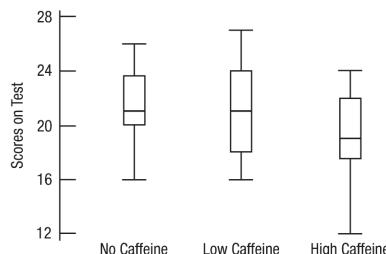
- a) The distribution of size of Finger Lakes vineyards is skewed heavily to the right. The mean size is a great deal higher than the median size.
- b) Upper Fence: $Q3 + 1.5(\text{IQR}) = 55 + 1.5(55 - 18.5) = 55 + 54.75 = 109.75$
 Lower Fence: $Q1 - 1.5(\text{IQR}) = 18.5 - 1.5(55 - 18.5) = 18.5 - 54.75 = -36.25$
- The maximum of 250 acres is well above the upper fence of 109.75 acres. Therefore, there is at least one high outlier, 250 acres. Since the lower fence is negative, there are no low outliers, since it is certainly impossible to have a vineyard with negative size.
- c) The boxplot of the distribution of sizes of Finger Lakes vineyards is shown below. There may be additional outliers, but we are sure that there is at least one, the maximum.



- d) The vineyards range in size from 6 to 250 acres. The median size of the 36 vineyards is 33.5 acres, so half are larger and half are smaller. The middle 50% of the vineyards have sizes between 18.5 and 55 acres. The distribution of sizes is skewed to the right, with at least one outlier.

35. Caffeine.

- a) *Who* – 45 student volunteers; *What* – Level of caffeine consumption and memory test score; *When* – Not specified. *Where* – Not specified. *Why* – The student researchers want to see the possible effects of caffeine on memory. *How* – It appears that the researchers imposed the treatment of level of caffeine consumption in an experiment. However, this point is not clear. Perhaps they allowed the subjects to choose their own level of caffeine.
- b) *Variables* – Caffeine level is a categorical variable with three levels: no caffeine, low caffeine, and high caffeine. Test score is a quantitative variable, measured in number of items recalled correctly.
- c) Comparative boxplots are shown below.

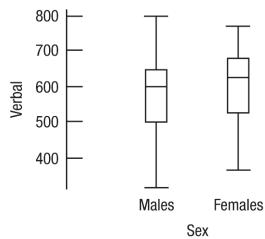


- d) The groups consuming no caffeine and low caffeine had comparable memory test scores. A typical score from these groups was around 21. However, the scores of the group consuming no caffeine were more consistent, with a smaller range and smaller interquartile range than the scores of the group consuming low caffeine. The group consuming high caffeine had lower memory scores in general, with a median score of about 19. No one in the high caffeine group scored above 24, but 25% of each of the other groups scored above 24.

44 *Part I Exploring and Understanding Data*

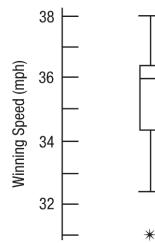
36. SAT scores.

- a) Comparative boxplots comparing the scores of boys and girls SAT scores are shown below.



- b) Females in this graduating class scored slightly higher on the Verbal SAT, with a median of 625, 25 points higher than the median of 600 for the males. Additionally, the females had higher first and third quartiles. The IQR of the males' scores was slightly smaller, than the IQR for the females' scores, indicating a bit more consistency in male scores. However, the overall spread of male scores was greater than that of female scores, with males having both the minimum and maximum score. Both distributions of scores were slightly skewed to the left.

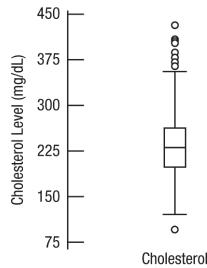
37. Derby speeds 2017.



- e) The distribution of winning speeds in the Kentucky Derby is skewed to the left. The lowest winning speed is just under 31 mph and the fastest speed is about 38 mph. The median speed is approximately 36 mph, and 75% of winning speeds are above 35 mph. Only a few percent of winners have had speeds below 33 mph. The middle 50% of winning speeds are between 35 and 37 mph, for an IQR of 2 mph

38. Framingham cholesterol.

A boxplot for the distribution of cholesterol levels of 1400 men shown below.



The distribution of cholesterol levels is essentially symmetric, with a median near 225. The IQR is about 75 points and the extremes are about 90 and 430.

39. Reading scores.

- a) The highest score for boys was 6, which is higher than the highest score for girls, 5.9.
- b) The range of scores for boys is greater than the range of scores for girls.
 $\text{Range} = \text{Max} - \text{Min}$ $\text{Range(Boys)} = 4$ $\text{Range(Girls)} = 3.1$
- c) The girls had the greater IQR.
 $\text{IQR(Boys)} = Q3 - Q1 = 4.9 - 3.9 = 1$ $\text{IQR(Girls)} = Q3 - Q1 = 5.2 - 3.8 = 1.4$
- d) The distribution of boys' scores is more skewed. The quartiles are not the same distance from the median. In the distribution of girls' scores, Q1 is 0.7 units below the median, while Q3 is 0.7 units above the median.
- e) Overall, the girls did better on the reading test. The median, 4.5, was higher than the median for the boys, 4.3. Additionally, the upper quartile score was higher for girls than boys, 5.2 compared to 4.9. The girls' lower quartile score was slightly lower than the boys' lower quartile score, 3.8 compared to 3.9.
- f) The overall mean is calculated by weighting each mean by the number of students, which is

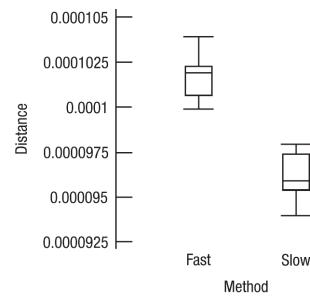
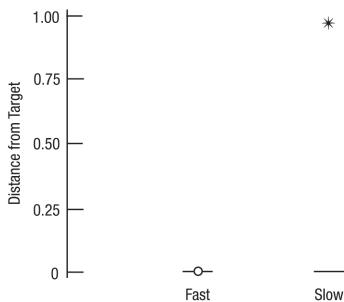
$$\frac{14(4.2) + 11(4.6)}{25} = 4.38.$$

40. Cloud seeding.

- a) The median and IQR, as well as the quartiles are the appropriate summary statistics, since the distribution of amount of rain produced is either skewed to the right or has high outliers. Indication of skewness and outliers can be seen in the comparison of median and mean. The mean amount of rain produced is significantly higher than the median for both seeded and unseeded clouds. Skewness or outliers pulled up the sensitive mean.
- b) There is evidence that that the seeded clouds produced more rain. The median and both quartiles are higher than the corresponding statistics for unseeded clouds. In fact, the median amount of rainfall for seeded clouds is 221.60 acre-feet, about 5 times the median amount for unseeded clouds.

41. Industrial experiment.

Looking at the boxplot below to the left, there is an extreme outlier in the distribution of distances for the slow speed drilling. One hole was drilled almost an inch away from the center of the target! If that distance is correct, the engineers at the computer production plant should investigate the slow speed drilling process closely. It may be plagued by extreme, intermittent inaccuracy. The outlier in the slow speed drilling process is so extreme that no graphical display can display the distribution in a meaningful way while including that outlier. That distance should be removed before looking at a plot of the drilling distances.

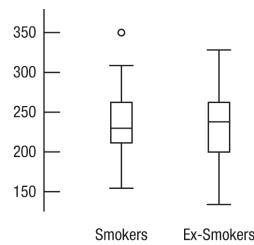


Looking at the boxplot above to the right, with the outlier set aside, we can determine that the slow drilling process is more accurate. The greatest distance from the target for the slow drilling process, 0.000098 inches, is still more accurate than the smallest distance for the fast drilling process, 0.000100 inches.

46 Part I Exploring and Understanding Data

42. Cholesterol and smoking.

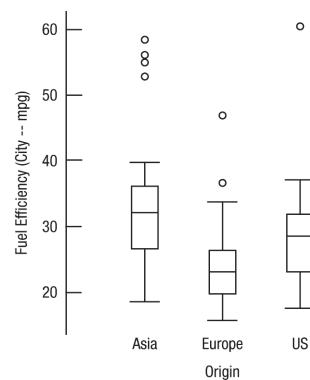
The smoker's distribution has a lower median and is less spread out, but the ex-smokers group has a higher percentage of its values below 200. There is also a high outlier in the smoker's group at 351.



In general, the cholesterol levels of smokers seem to be slightly lower than the cholesterol levels of ex-smokers. Additionally, the cholesterol levels of smokers appear more consistent than cholesterol levels of ex-smokers.

43. MPG.

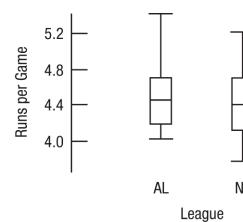
- a) Comparative boxplots are shown below. Comparative dotplots, histograms, or stemplots would also be acceptable.



- b) The distributions of fuel efficiency for all three groups have high outliers, potentially hybrid or electric cars. Asian cars generally have the highest fuel efficiency, while European cars have the lowest. The Asian median is about 32 mpg, European cars around 23 mpg, and U.S. cars 28 mpg.

44. Baseball 2016.

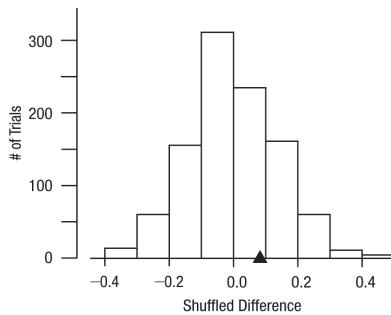
- a) Comparative boxplots are shown below. Back-to-back stem-and-leaf displays, comparative histograms, and dotplots on the same scale would also be acceptable.



- b) Both distributions are roughly symmetric. The American League has a higher center but smaller variability. The Red Sox may be an outlier in the American League (although not extreme enough for the boxplot outlier nomination rule).

44. (continued)

- c) The Red Sox are not nominated as an outlier (they are just below 1.5 IQRs above Q3), but their run production might be viewed as unusually high. The Rockies are well less than 1.5 IQRs above the National League Q3 so, although their run production is high, it's probably not unusually high.
- d) (Answers will vary slightly.) A histogram of the 1000 differences shows that the difference of 0.082 is fairly typical for the means of two randomly assigned groups of 15 teams. In fact, in nearly 55% of the trials, a difference greater than 0.082 was observed. There is no evidence to suggest that such a difference didn't occur by chance.



45. Assets.

- a) The distribution of assets of 79 companies chosen from the *Forbes* list of the nation's top corporations is skewed so heavily to the right that the vast majority of companies have assets represented in the first bar of the histogram, 0 to 10 billion dollars. This makes meaningful discussion of center and spread impossible.
- b) Re-expressing these data by, for example, logs or square roots might help make the distribution more nearly symmetric. Then a meaningful discussion of center might be possible.

46. Music library.

- a) The distribution of the number of songs students had in their digital libraries is extremely skewed to the right. That makes it difficult to determine a center. The typical number of songs in a library is probably in the first bar of the histogram.
- b) Re-expressing these data by, for example, logs or square roots might help make the distribution more nearly symmetric. Then a meaningful discussion of center might be possible.

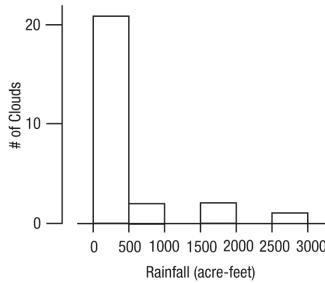
47. Assets again.

- a) The distribution of logarithm of assets is preferable, because it is roughly unimodal and symmetric. The distribution of the square root of assets is still skewed right, with outliers.
- b) If $\sqrt{Assets} = 50$, then the company's assets are approximately $50^2 = \$2500$ million.
- c) If $\log(Assets) = 3$, then the company's assets are approximately $10^3 = \$1000$ million.

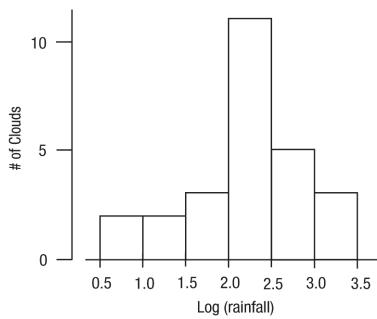
48 Part I Exploring and Understanding Data

48. Rainmakers.

- a) Since one acre-foot is about 320,000 gallons, these numbers are more manageable than gallons.
- b) The distribution of rainfall from 26 clouds seeded with silver iodide is skewed heavily to the right, with the vast majority of clouds producing less than 500 acre-feet of rain. Several clouds produced more, with a maximum of 2745 acre-feet.



- c) The distribution of log (base 10) of rainfall is much more symmetric than the distribution of rainfall. We can see that the center of the distribution is around $\log 2 - \log 2.5$ acre-feet.



- d) Since the re-expressed scale is measured in log (acre-feet), we need to raise 10 to the power of the number on our scale to convert back to acre feet. For example, if a cloud in the new scale has a log (rainfall) of 2.3, we convert back to rainfall as follows: $\log(\text{rainfall}) = 2.3 \Rightarrow \text{rainfall} = 10^{2.3} = 199.5$. The cloud produced 199.5 acre-feet of rain.

49. Stereograms.

- a) The two variables discussed in the description are fusion time and treatment group.
- b) Fusion time is a quantitative variable, measured in seconds. Treatment group is a categorical variable, with subjects either receiving verbal clues only, or visual and verbal clues.
- c) Both groups have distributions that are skewed to the right. Generally, the Visual/Verbal group had shorter fusion times than the No/Verbal group. The median for the Visual/Verbal group was approximately the same as the lower quartile for the No/Verbal group. The No/Verbal Group also had an extreme outlier, with at least one subject whose fusion time was approximately 50 seconds. There is evidence that visual information may reduce fusion time.

50. Stereograms, revisited.

The re-expression using logarithms has a distribution that is more symmetric than the original distribution of fusion times, and the re-expression has no outliers. This symmetry makes it easier to compare the two groups.

51. Cost of living sampled.

- a) If there were no mean difference, we would expect the histogram to be centered at 0.
- b) (Answers may vary.) Because 0 never occurred in 1000 random samples, this provides strong evidence that the difference we saw was not due to chance. We can conclude that there is a real (if small) difference between the mean prices of a cappuccino and a dozen eggs.

52. Stereograms shuffled.

(Answers may vary.) In 1000 random shufflings, a difference as large as the one observed in the data occurred only 16 times. Although it's possible that the observed difference was due to chance, this provides some evidence to suggest that the difference was real.

53. Pizza prices by city.

- a) Pizza prices appear to be both higher on average, and more variable, in Baltimore than in the other three cities. Prices in Chicago may be slightly higher on average than in Dallas and Denver, but the difference is small.
- b) There are low outliers in the distribution of pizza prices in Baltimore and Chicago. There is one high outlier in the distribution of pizza prices in Dallas. These outliers do not affect the overall conclusions reached in the previous part.

54. Pizza price differences.

- a) Baltimore pizza slices cost an average of \$0.23 more, which appears to be a real difference. Random chance never produced a difference bigger than \$0.10.
- b) The outliers on the low end in Baltimore and the high end in Dallas do not affect the conclusion that Baltimore has a larger mean pizza slice price. Baltimore's outliers would pull the mean down, and Dallas' outliers would pull the mean up. If anything, the difference in average pizza slice prices would be greater if the outliers were set aside.
- c) Your simulation will vary. The sample simulation shows that the difference of \$0.05 (\$2.62 for Chicago vs. \$2.57 for Denver) would happen only about 2% of the time by chance. Although it is not a very big difference, it does appear to be a real difference.

Sample computer simulation results:

Event	Count	Proportion of 10000 runs
-0.050705 or below	114	0.0114
0.050705 or above	110	0.0110
Total	224	0.0224

