15. Do running times of American movies differ somehow from running times of French movies? The author investigated this question by randomly selecting 25 recent movies of each type, resulting in the following running times:

Am: 94 90 95 93 128 95 125 91 104 116 162 102 90 110 92 113 116 90 97 103 95 120 109 91 138

Fr: 123 116 90 158 122 119 125 90 96 94 137 102 105 106 95 125 122 103 96 111 81 113 128 93 92

Construct a *comparative* stem-and-leaf display by listing stems in the middle of your paper and then placing the Am leaves out to the left and the Fr leaves out to the right. Then comment on interesting features of the display.

16. The article cited in Example 1.2 also gave the accompanying strength observations for cylinders:

6.1 5.8 7.8 7.1 7.2 9.2 6.6 8.3 7.0 8.3
7.8 8.1 7.4 8.5 8.9 9.8 9.7 14.1 12.6 11.2

a. Construct a comparative stem-and-leaf display (see the previous exercise) of the beam and cylinder data, and then answer the questions in parts (b)–(d) of Exercise 10 for the observations on cylinders.

b. In what ways are the two sides of the display similar? Are there any obvious differences between the beam observations and the cylinder observations?

c. Construct a dotplot of the cylinder data.

17. Temperature transducers of a certain type are shipped in batches of 50. A sample of 60 batches was selected, and the number of transducers in each batch not conforming to design specifications was determined, resulting in the following data:

2 1 2 4 0 1 3 2 0 5 3 3 1 3 2 4 7 0 2 3
0 4 2 1 3 1 1 3 4 1 2 3 2 2 8 4 5 1 3 1
5 0 2 3 2 1 0 6 4 2 1 6 0 3 3 3 6 1 2 3

a. Determine frequencies and relative frequencies for the observed values of x = number of nonconforming transducers in a batch.

b. What proportion of batches in the sample have at most five nonconforming transducers? What proportion have fewer than five? What proportion have at least five nonconforming units?

c. Draw a histogram of the data using relative frequency on the vertical scale, and comment on its features.

18. In a study of author productivity ("Lotka's Test," *Collection Mgmt.,* 1982: 111–118), a large number of authors were classified according to the number of articles they had published during a certain period. The results were presented in the accompanying frequency distribution:

Number of papers	1	2	3	4	5	6	7	8
Frequency	784	204	127	50	33	28	19	19

Number of papers	9	10	11	12	13	14	15	16	17
Frequency	6	7	6	7	4	4	5	3	3

a. Construct a histogram corresponding to this frequency distribution. What is the most interesting feature of the shape of the distribution?

b. What proportion of these authors published at least five papers? At least ten papers? More than ten papers?

c. Suppose the five 15s, three 16s, and three 17s had been lumped into a single category displayed as "≥15." Would you be able to draw a histogram? Explain.

d. Suppose that instead of the values 15, 16, and 17 being listed separately, they had been combined into a 15–17 category with frequency 11. Would you be able to draw a histogram? Explain.

19. The number of contaminating particles on a silicon wafer prior to a certain rinsing process was determined for each wafer in a sample of size 100, resulting in the following frequencies:

Number of particles	0	1	2	3	4	5	6	7
Frequency	1	2	3	12	11	15	18	10
Number of particles	8	9	10	11	12	13	14	
Frequency	12	4	5	3	1	2	1	

a. What proportion of the sampled wafers had at least one particle? At least five particles?

b. What proportion of the sampled wafers had between five and ten particles, inclusive? Strictly between five and ten particles?

c. Draw a histogram using relative frequency on the vertical axis. How would you describe the shape of the histogram?

20. The article "Determination of Most Representative Subdivision" (*J. of Energy Engr.,* 1993: 43–55) gave data on various characteristics of subdivisions that could be used in deciding whether to provide electrical power using overhead lines or underground lines. Here are the values of the variable x = total length of streets within a subdivision:

1280 5320 4390 2100 1240 3060 4770
1050 360 3330 3380 340 1000 960
1320 530 3350 540 3870 1250 2400
960 1120 2120 450 2250 2320 2400
3150 5700 5220 500 1850 2460 5850
2700 2730 1670 100 5770 3150 1890
510 240 396 1419 2109

a. Construct a stem-and-leaf display using the thousands digit as the stem and the hundreds digit as the leaf, and comment on the various features of the display.

b. Construct a histogram using class boundaries 0, 1000, 2000, 3000, 4000, 5000, and 6000. What proportion of subdivisions have total length less than 2000? Between 2000 and 4000? How would you describe the shape of the histogram?

21. The article cited in Exercise 20 also gave the following values of the variables y = number of culs-de-sac and z = number of intersections:

y	1	0	1	0	0	2	0	1	1	1	2	1	0	0	1	1	0	1	1
z	1	8	6	1	1	5	3	0	0	4	4	0	0	1	2	1	4	0	4

y	1	1	0	0	0	1	1	2	0	1	2	2	1	1	0	2	1	1	0
z	0	3	0	1	1	0	1	3	2	4	6	6	0	1	1	8	3	3	5
y	1	5	0	3	0	1	1	0	0										
z	0	5	2	3	1	0	0	0	3										

a. Construct a histogram for the y data. What proportion of these subdivisions had no culs-de-sac? At least one cul-de-sac?

b. Construct a histogram for the z data. What proportion of these subdivisions had at most five intersections? Fewer than five intersections?

22. How does the speed of a runner vary over the course of a marathon (a distance of 42.195 km)? Consider determining both the time to run the first 5 km and the time to run between the 35-km and 40-km points, and then subtracting the former time from the latter time. A positive value of this difference corresponds to a runner slowing down toward the end of the race. The accompanying histogram is based on times of runners who participated in several different Japanese marathons ("Factors Affecting Runners' Marathon Performance," *Chance,* Fall, 1993: 24–30).

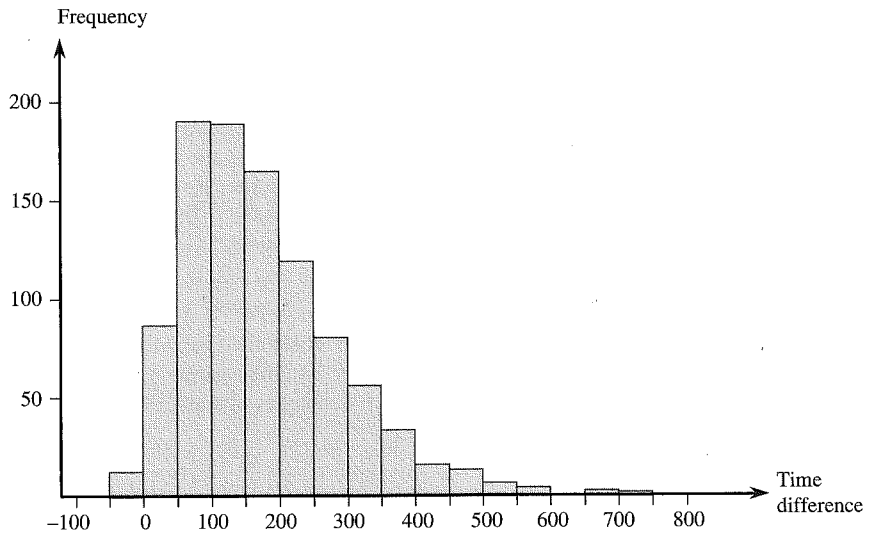
What are some interesting features of this histogram? What is a typical difference value? Roughly what proportion of the runners ran the late distance more quickly than the early distance?

23. The article "Statistical Modeling of the Time Course of Tantrum Anger" (*Annals of Applied Stats,* 2009: 1013–1034) discussed how anger intensity in children's tantrums could be related to tantrum duration as well as behavioral indicators such as shouting, stamping, and pushing or pulling. The following frequency distribution was given (and also the corresponding histogram):

0−<2:	136	2−<4:	92	4−<11:	71
11−<20:	26	20−<30:	7	30−<40:	3

Draw the histogram and then comment on any interesting features.

Histogram for Exercise 22



24. The accompanying data set consists of observations on shear strength (lb) of ultrasonic spot welds made on a certain type of alclad sheet. Construct a relative frequency histogram based on ten equal-width classes with boundaries 4000, 4200, ... [The histogram will agree with the one in "Comparison of Properties of Joints Prepared by Ultrasonic Welding and Other Means" (*J. of Aircraft,* 1983: 552–556).] Comment on its features.

5434	4948	4521	4570	4990	5702	5241
5112	5015	4659	4806	4637	5670	4381
4820	5043	4886	4599	5288	5299	4848
5378	5260	5055	5828	5218	4859	4780
5027	5008	4609	4772	5133	5095	4618
4848	5089	5518	5333	5164	5342	5069
4755	4925	5001	4803	4951	5679	5256
5207	5621	4918	5138	4786	4500	5461
5049	4974	4592	4173	5296	4965	5170
4740	5173	4568	5653	5078	4900	4968
5248	5245	4723	5275	5419	5205	4452
5227	5555	5388	5498	4681	5076	4774
4931	4493	5309	5582	4308	4823	4417
5364	5640	5069	5188	5764	5273	5042
5189	4986					

25. A transformation of data values by means of some mathematical function, such as $\sqrt{x}$ or $1/x$, can often yield a set of numbers that has "nicer" statistical properties than the original data. In particular, it may be possible to find a function for which the histogram of transformed values is more symmetric (or, even better, more like a bell-shaped curve) than the original data. As an example, the article "Time Lapse Cinematographic Analysis of Beryllium–Lung Fibroblast Interactions" (*Environ. Research,* 1983: 34–43) reported the results of experiments designed to study the behavior of certain individual cells that had been exposed to beryllium. An important characteristic of such an individual cell is its interdivision time (IDT). IDTs were determined for a large number of cells, both in exposed

(treatment) and unexposed (control) conditions. The authors of the article used a logarithmic transformation, that is, transformed value = log(original value). Consider the following representative IDT data:

IDT	$\log_{10}$(IDT)	IDT	$\log_{10}$(IDT)	IDT	$\log_{10}$(IDT)
28.1	1.45	60.1	1.78	21.0	1.32
31.2	1.49	23.7	1.37	22.3	1.35
13.7	1.14	18.6	1.27	15.5	1.19
46.0	1.66	21.4	1.33	36.3	1.56
25.8	1.41	26.6	1.42	19.1	1.28
16.8	1.23	26.2	1.42	38.4	1.58
34.8	1.54	32.0	1.51	72.8	1.86
62.3	1.79	43.5	1.64	48.9	1.69
28.0	1.45	17.4	1.24	21.4	1.33
17.9	1.25	38.8	1.59	20.7	1.32
19.5	1.29	30.6	1.49	57.3	1.76
21.1	1.32	55.6	1.75	40.9	1.61
31.9	1.50	25.5	1.41		
28.9	1.46	52.1	1.72		

Use class intervals $10-<20, 20-<30, \ldots$ to construct a histogram of the original data. Use intervals $1.1-<1.2, 1.2-<1.3, \ldots$ to do the same for the transformed data. What is the effect of the transformation?

26. Automated electron backscattered diffraction is now being used in the study of fracture phenomena. The following information on misorientation angle (degrees) was extracted from the article "Observations on the Faceted Initiation Site in the Dwell-Fatigue Tested Ti-6242 Alloy: Crystallographic Orientation and Size Effects (*Metallurgical and Materials Trans.*, 2006: 1507–1518).

Class:	0−<5	5−<10	10−<15	15−<20
Rel freq:	.177	.166	.175	.136
Class:	20−<30	30−<40	40−<60	60−<90
Rel freq:	.194	.078	.044	.030

a. Is it true that more than 50% of the sampled angles are smaller than 15°, as asserted in the paper?
b. What proportion of the sampled angles are at least 30°?
c. Roughly what proportion of angles are between 10° and 25°?
d. Construct a histogram and comment on any interesting features.

27. The paper "Study on the Life Distribution of Microdrills" (*J. of Engr. Manufacture*, 2002: 301–305) reported the following observations, listed in increasing order, on drill lifetime (number of holes that a drill machines before it breaks) when holes were drilled in a certain brass alloy.

11	14	20	23	31	36	39	44	47	50
59	61	65	67	68	71	74	76	78	79
81	84	85	89	91	93	96	99	101	104
105	105	112	118	123	136	139	141	148	158
161	168	184	206	248	263	289	322	388	513

a. Why can a frequency distribution not be based on the class intervals 0–50, 50–100, 100–150, and so on?
b. Construct a frequency distribution and histogram of the data using class boundaries 0, 50, 100, . . . , and then comment on interesting characteristics.
c. Construct a frequency distribution and histogram of the natural logarithms of the lifetime observations, and comment on interesting characteristics.
d. What proportion of the lifetime observations in this sample are less than 100? What proportion of the observations are at least 200?

28. Human measurements provide a rich area of application for statistical methods. The article "A Longitudinal Study of the Development of Elementary School Children's Private Speech" (*Merrill-Palmer Q.*, 1990: 443–463) reported on a study of children talking to themselves (private speech). It was thought that private speech would be related to IQ, because IQ is supposed to measure mental maturity, and it was known that private speech decreases as students progress through the primary grades. The study included 33 students whose first-grade IQ scores are given here:

82 96 99 102 103 103 106 107 108 108 108 108
109 110 110 111 113 113 113 113 115 115 118 118
119 121 122 122 127 132 136 140 146

Describe the data and comment on any interesting features.

29. Consider the following data on types of health complaint (J = joint swelling, F = fatigue, B = back pain, M = muscle weakness, C = coughing, N = nose running/irritation, O = other) made by tree planters. Obtain frequencies and relative frequencies for the various categories, and draw a histogram. (The data is consistent with percentages given in the article "Physiological Effects of Work Stress and Pesticide Exposure in Tree Planting by British Columbia Silviculture Workers," *Ergonomics*, 1993: 951–961.)

O O N J C F B B F O J O O M
O F F O O N O N J F J B O C
J O J J F N O B M O J M O B
O F J O O B N C O O O M B F
J O F N

30. A **Pareto diagram** is a variation of a histogram for categorical data resulting from a quality control study. Each category represents a different type of product nonconformity or production problem. The categories are ordered so that the one with the largest frequency appears on the far left, then the category with the second largest frequency, and so on. Suppose the following information on nonconformities in circuit packs is obtained: failed component, 126; incorrect component, 210; insufficient solder, 67; excess solder, 54; missing component, 131. Construct a Pareto diagram.

31. The **cumulative frequency** and cumulative relative frequency for a particular class interval are the sum of frequencies and relative frequencies, respectively, for that interval and all intervals lying below it. If, for example,

EXERCISES Section 1.3 (33–43)

33. The May 1, 2009 issue of *The Montclarian* reported the following home sale amounts for a sample of homes in Alameda, CA that were sold the previous month (1000s of $):

590 815 575 608 350 1285 408 540 555 679

a. Calculate and interpret the sample mean and median.
b. Suppose the 6th observation had been 985 rather than 1285. How would the mean and median change?
c. Calculate a 20% trimmed mean by first trimming the two smallest and two largest observations.
d. Calculate a 15% trimmed mean.

34. Exposure to microbial products, especially endotoxin, may have an impact on vulnerability to allergic diseases. The article "Dust Sampling Methods for Endotoxin—An Essential, But Underestimated Issue" (*Indoor Air,* 2006: 20–27) considered various issues associated with determining endotoxin concentration. The following data on concentration (EU/mg) in settled dust for one sample of urban homes and another of farm homes was kindly supplied by the authors of the cited article.

U: 6.0 5.0 11.0 33.0 4.0 5.0 80.0 18.0 35.0 17.0 23.0
F: 4.0 14.0 11.0 9.0 9.0 8.0 4.0 20.0 5.0 8.9 21.0 9.2 3.0 2.0 0.3

a. Determine the sample mean for each sample. How do they compare?
b. Determine the sample median for each sample. How do they compare? Why is the median for the urban sample so different from the mean for that sample?
c. Calculate the trimmed mean for each sample by deleting the smallest and largest observation. What are the corresponding trimming percentages? How do the values of these trimmed means compare to the corresponding means and medians?

35. The minimum injection pressure (psi) for injection molding specimens of high amylose corn was determined for eight different specimens (higher pressure corresponds to greater processing difficulty), resulting in the following observations (from "Thermoplastic Starch Blends with a Polyethylene-Co-Vinyl Alcohol: Processability and Physical Properties," *Polymer Engr. and Science,* 1994: 17–23):

15.0 13.0 18.0 14.5 12.0 11.0 8.9 8.0

a. Determine the values of the sample mean, sample median, and 12.5% trimmed mean, and compare these values.
b. By how much could the smallest sample observation, currently 8.0, be increased without affecting the value of the sample median?
c. Suppose we want the values of the sample mean and median when the observations are expressed in kilograms per square inch (ksi) rather than psi. Is it necessary to reexpress each observation in ksi, or can the values calculated in part (a) be used directly? [*Hint:* 1 kg = 2.2 lb.]

36. A sample of 26 offshore oil workers took part in a simulated escape exercise, resulting in the accompanying data on time (sec) to complete the escape ("Oxygen Consumption and Ventilation During Escape from an Offshore Platform," *Ergonomics,* 1997: 281–292):

389 356 359 363 375 424 325 394 402
373 373 370 364 366 364 325 339 393
392 369 374 359 356 403 334 397

a. Construct a stem-and-leaf display of the data. How does it suggest that the sample mean and median will compare?
b. Calculate the values of the sample mean and median. [*Hint:* $\Sigma x_i = 9638$.]
c. By how much could the largest time, currently 424, be increased without affecting the value of the sample median? By how much could this value be decreased without affecting the value of the sample median?
d. What are the values of $\bar{x}$ and $\tilde{x}$ when the observations are reexpressed in minutes?

37. The article "Snow Cover and Temperature Relationships in North America and Eurasia" (*J. Climate and Applied Meteorology,* 1983: 460–469) used statistical techniques to relate the amount of snow cover on each continent to average continental temperature. Data presented there included the following ten observations on October snow cover for Eurasia during the years 1970–1979 (in million km^2):

6.5 12.0 14.9 10.0 10.7 7.9 21.9 12.5 14.5 9.2

What would you report as a representative, or typical, value of October snow cover for this period, and what prompted your choice?

38. Blood pressure values are often reported to the nearest 5 mmHg (100, 105, 110, etc.). Suppose the actual blood pressure values for nine randomly selected individuals are

118.6 127.4 138.4 130.0 113.7 122.0 108.3 131.5 133.2

a. What is the median of the *reported* blood pressure values?
b. Suppose the blood pressure of the second individual is 127.6 rather than 127.4 (a small change in a single value). How does this affect the median of the reported values? What does this say about the sensitivity of the median to rounding or grouping in the data?

39. The propagation of fatigue cracks in various aircraft parts has been the subject of extensive study in recent years. The accompanying data consists of propagation lives (flight hours/10^4) to reach a given crack size in fastener holes intended for use in military aircraft ("Statistical Crack

is stretched out compared with the cancer box ($f_s = 18$ vs. $f_s = 11$), and the positions of the median lines in the two boxes show much more skewness in the middle half of the no-cancer sample than the cancer sample. Outliers are represented by horizontal line segments, and there is no distinction between mild and extreme outliers. ■

EXERCISES Section 1.4 (44–61)

44. The article "Oxygen Consumption During Fire Suppression: Error of Heart Rate Estimation" (*Ergonomics,* 1991: 1469–1474) reported the following data on oxygen consumption (mL/kg/min) for a sample of ten firefighters performing a fire-suppression simulation:

29.5 49.3 30.6 28.2 28.0 26.3 33.9 29.4 23.5 31.6

Compute the following:

a. The sample range
b. The sample variance s^2 from the definition (i.e., by first computing deviations, then squaring them, etc.)
c. The sample standard deviation
d. s^2 using the shortcut method

45. The value of Young's modulus (GPa) was determined for cast plates consisting of certain intermetallic substrates, resulting in the following sample observations ("Strength and Modulus of a Molybdenum-Coated Ti-25Al-10Nb-3U-1Mo Intermetallic," *J. of Materials Engr. and Performance,* 1997: 46–50):

116.4 115.9 114.6 115.2 115.8

a. Calculate $\bar{x}$ and the deviations from the mean.
b. Use the deviations calculated in part (a) to obtain the sample variance and the sample standard deviation.
c. Calculate s^2 by using the computational formula for the numerator S_{xx}.
d. Subtract 100 from each observation to obtain a sample of transformed values. Now calculate the sample variance of these transformed values, and compare it to s^2 for the original data.

46. The accompanying observations on stabilized viscosity (cP) for specimens of a certain grade of asphalt with 18% rubber added are from the article "Viscosity Characteristics of Rubber-Modified Asphalts" (*J. of Materials in Civil Engr.,* 1996: 153–156):

2781 2900 3013 2856 2888

a. What are the values of the sample mean and sample median?
b. Calculate the sample variance using the computational formula. [*Hint:* First subtract a convenient number from each observation.]

47. Calculate and interpret the values of the sample median, sample mean, and sample standard deviation for the following observations on fracture strength (MPa, read from a graph in "Heat-Resistant Active Brazing of Silicon Nitride: Mechanical Evaluation of Braze Joints," *Welding J.,* August, 1997):

87 93 96 98 105 114 128 131 142 168

48. Exercise 34 presented the following data on endotoxin concentration in settled dust both for a sample of urban homes and for a sample of farm homes:

U: 6.0 5.0 11.0 33.0 4.0 5.0 80.0 18.0 35.0 17.0 23.0
F: 4.0 14.0 11.0 9.0 9.0 8.0 4.0 20.0 5.0 8.9 21.0 9.2 3.0 2.0 0.3

a. Determine the value of the sample standard deviation for each sample, interpret these values, and then contrast variability in the two samples. [*Hint:* $\sum x_i = 237.0$ for the urban sample and $= 128.4$ for the farm sample, and $\sum x_i^2 = 10{,}079$ for the urban sample and 1617.94 for the farm sample.]
b. Compute the fourth spread for each sample and compare. Do the fourth spreads convey the same message about variability that the standard deviations do? Explain.
c. The authors of the cited article also provided endotoxin concentrations in dust bag dust:

U: 34.0 49.0 13.0 33.0 24.0 24.0 35.0 104.0 34.0 40.0 38.0 1.0
F: 2.0 64.0 6.0 17.0 35.0 11.0 17.0 13.0 5.0 27.0 23.0 28.0 10.0 13.0 0.2

Construct a comparative boxplot (as did the cited paper) and compare and contrast the four samples.

49. A study of the relationship between age and various visual functions (such as acuity and depth perception) reported the following observations on the area of scleral lamina (mm^2) from human optic nerve heads ("Morphometry of Nerve Fiber Bundle Pores in the Optic Nerve Head of the Human," *Experimental Eye Research,* 1988: 559–568):

2.75 2.62 2.74 3.85 2.34 2.74 3.93 4.21 3.88
4.33 3.46 4.52 2.43 3.65 2.78 3.56 3.01

a. Calculate $\sum x_i$ and $\sum x_i^2$.
b. Use the values calculated in part (a) to compute the sample variance s^2 and then the sample standard deviation s.

50. In 1997 a woman sued a computer keyboard manufacturer, charging that her repetitive stress injuries were caused by the keyboard (*Genessy v. Digital Equipment Corp.*). The injury awarded about $3.5 million for pain and suffering, but the court then set aside that award as being unreasonable

compensation. In making this determination, the court identified a "normative" group of 27 similar cases and specified a reasonable award as one within two standard deviations of the mean of the awards in the 27 cases. The 27 awards were (in $1000s) 37, 60, 75, 115, 135, 140, 149, 150, 238, 290, 340, 410, 600, 750, 750, 750, 1050, 1100, 1139, 1150, 1200, 1200, 1250, 1576, 1700, 1825, and 2000, from which $\sum x_i = 20{,}179$, $\sum x_i^2 = 24{,}657{,}511$. What is the maximum possible amount that could be awarded under the two-standard-deviation rule?

51. The article "A Thin-Film Oxygen Uptake Test for the Evaluation of Automotive Crankcase Lubricants" (*Lubric. Engr.*, 1984: 75–83) reported the following data on oxidation-induction time (min) for various commercial oils:

87 103 130 160 180 195 132 145 211 105 145
153 152 138 87 99 93 119 129

a. Calculate the sample variance and standard deviation.
b. If the observations were reexpressed in hours, what would be the resulting values of the sample variance and sample standard deviation? Answer without actually performing the reexpression.

52. The first four deviations from the mean in a sample of $n = 5$ reaction times were .3, .9, 1.0, and 1.3. What is the fifth deviation from the mean? Give a sample for which these are the five deviations from the mean.

53. A **mutual fund** is a professionally managed investment scheme that pools money from many investors and invests in a variety of securities. Growth funds focus primarily on increasing the value of investments, whereas blended funds seek a balance between current income and growth. Here is data on the expense ratio (expenses as a % of assets, from www.morningstar.com) for samples of 20 large-cap balanced funds and 20 large-cap growth funds ("large-cap" refers to the sizes of companies in which the funds invest; the population sizes are 825 and 762, respectively):

Bl	1.03	1.23	1.10	1.64	1.30
	1.27	1.25	0.78	1.05	0.64
	0.94	2.86	1.05	0.75	0.09
	0.79	1.61	1.26	0.93	0.84
Gr	0.52	1.06	1.26	2.17	1.55
	0.99	1.10	1.07	1.81	2.05
	0.91	0.79	1.39	0.62	1.52
	1.02	1.10	1.78	1.01	1.15

a. Calculate and compare the values of $\bar{x}$, $\tilde{x}$, and s for the two types of funds.
b. Construct a comparative boxplot for the two types of funds, and comment on interesting features.

54. Grip is applied to produce normal surface forces that compress the object being gripped. Examples include two people shaking hands, or a nurse squeezing a patient's forearm to stop bleeding. The article "Investigation of Grip Force, Normal Force, Contact Area, Hand Size, and Handle Size for Cylindrical Handles" (*Human Factors*, 2008: 734–744) included the following data on grip strength (N) for a sample of 42 individuals:

16 18 18 26 33 41 54 56 66 68 87 91 95
98 106 109 111 118 127 127 135 145 147 149 151 168
172 183 189 190 200 210 220 229 230 233 238 244 259
294 329 403

a. Construct a stem-and-leaf display based on repeating each stem value twice, and comment on interesting features.
b. Determine the values of the fourths and the fourth-spread.
c. Construct a boxplot based on the five-number summary, and comment on its features.
d. How large or small does an observation have to be to qualify as an outlier? An extreme outlier? Are there any outliers?
e. By how much could the observation 403, currently the largest, be decreased without affecting f_s?

55. Here is a stem-and-leaf display of the escape time data introduced in Exercise 36 of this chapter.

32	55
33	49
34	
35	6699
36	34469
37	03345
38	9
39	2347
40	23
41	
42	4

a. Determine the value of the fourth spread.
b. Are there any outliers in the sample? Any extreme outliers?
c. Construct a boxplot and comment on its features.
d. By how much could the largest observation, currently 424, be decreased without affecting the value of the fourth spread?

56. The following data on distilled alcohol content (%) for a sample of 35 port wines was extracted from the article "A Method for the Estimation of Alcohol in Fortified Wines Using Hydrometer Baumé and Refractometer Brix" (*Amer. J. Enol. Vitic.*, 2006: 486–490). Each value is an average of two duplicate measurements.

16.35 18.85 16.20 17.75 19.58 17.73 22.75 23.78 23.25
19.08 19.62 19.20 20.05 17.85 19.17 19.48 20.00 19.97
17.48 17.15 19.07 19.90 18.68 18.82 19.03 19.45 19.37
19.20 18.00 19.60 19.33 21.22 19.50 15.30 22.25

Use methods from this chapter, including a boxplot that shows outliers, to describe and summarize the data.

57. A sample of 20 glass bottles of a particular type was selected, and the internal pressure strength of each bottle was determined. Consider the following partial sample information:

median = 202.2 lower fourth = 196.0
upper fourth = 216.8

Three smallest observations	125.8	188.1	193.7
Three largest observations	221.3	230.5	250.2

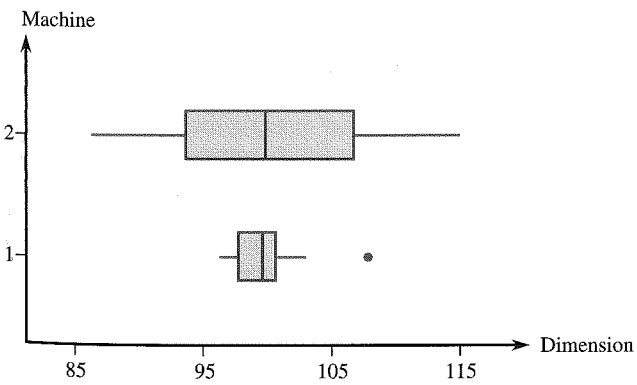
a. Are there any outliers in the sample? Any extreme outliers?
b. Construct a boxplot that shows outliers, and comment on any interesting features.

58. A company utilizes two different machines to manufacture parts of a certain type. During a single shift, a sample of $n = 20$ parts produced by each machine is obtained, and the value of a particular critical dimension for each part is determined. The comparative boxplot at the bottom of this page is constructed from the resulting data. Compare and contrast the two samples.

59. Blood cocaine concentration (mg/L) was determined both for a sample of individuals who had died from cocaine-induced excited delirium (ED) and for a sample of those who had died from a cocaine overdose without excited delirium; survival time for people in both groups was at most 6 hours. The accompanying data was read from a comparative boxplot in the article "Fatal Excited Delirium Following Cocaine Use" (*J. of Forensic Sciences*, 1997: 25–31).

ED	0	0	0	0	.1	.1	.1	.1	.2	.2	.3	.3
	.3	.4	.5	.7	.8	1.0	1.5	2.7	2.8			
	3.5	4.0	8.9	9.2	11.7	21.0						
Non-ED	0	0	0	0	0	.1	.1	.1	.1	.2	.2	.2
	.3	.3	.3	.4	.5	.5	.6	.8	.9	1.0		
	1.2	1.4	1.5	1.7	2.0	3.2	3.5	4.1				
	4.3	4.8	5.0	5.6	5.9	6.0	6.4	7.9				
	8.3	8.7	9.1	9.6	9.9	11.0	11.5					
	12.2	12.7	14.0	16.6	17.8							

Comparative boxplot for Exercise 58



a. Determine the medians, fourths, and fourth spreads for the two samples.
b. Are there any outliers in either sample? Any extreme outliers?
c. Construct a comparative boxplot, and use it as a basis for comparing and contrasting the ED and non-ED samples.

60. Observations on burst strength (lb/in^2) were obtained both for test nozzle closure welds and for production cannister nozzle welds ("Proper Procedures Are the Key to Welding Radioactive Waste Cannisters," *Welding J.*, Aug. 1997: 61–67).

Test	7200	6100	7300	7300	8000	7400
	7300	7300	8000	6700	8300	
Cannister	5250	5625	5900	5900	5700	6050
	5800	6000	5875	6100	5850	6600

Construct a comparative boxplot and comment on interesting features (the cited article did not include such a picture, but the authors commented that they had looked at one).

61. The accompanying comparative boxplot of gasoline vapor coefficients for vehicles in Detroit appeared in the article "Receptor Modeling Approach to VOC Emission Inventory Validation" (*J. of Envir. Engr.*, 1995: 483–490). Discuss any interesting features.

Comparative boxplot for Exercise 61