CHAPTER 1

Section 1.1

1.

a. *Houston Chronicle*, *Des Moines Register*, *Chicago Tribune*, *Washington Post*

b. Capital One, Campbell Soup, Merrill Lynch, Pulitzer

c. Bill Jasper, Kay Reinke, Helen Ford, David Menendez

d. 1.78, 2.44, 3.50, 3.04

3.

a. In a sample of 100 phones, what are the chances that more than 20 need service while under warranty? What are the chances than none need service while still under warranty?

b. What proportion of *all* phones of this brand and model will need service within the warranty period?

5.

a. Two variables (at least) were recorded: skin color and hourly wages.

b. Skin color is categorical (with four categories), while hourly wages is quantitative (units: $/hr).

7. **a.** categorical **b.** quantitative **c.** categorical **d.** categorical **e.** categorical

9.

a. No, the relevant conceptual population is all scores of all students who participate in the SI in conjunction with this particular statistics course.

b. The advantage to randomly assigning students to the two groups is that the two groups should then be fairly comparable before the study. If the two groups perform differently in the class, we can reasonably attribute this to the treatments (SI and control). If it were left to students to choose, stronger or more dedicated students might gravitate toward SI, confounding the results.

c. If all students were put in the treatment group there would be no results with which to compare the treatments.

11. One could generate a simple random sample of all single-family homes in the city or a stratified random sample by taking a simple random sample from each of the 10 district neighborhoods. From each of the homes in the sample the necessary variables would be collected. This would be an enumerative study because there exists a finite, identifiable population of objects from which to sample.

13.

a. There could be several explanations for the variability of the measurements. Among them could be measuring error, (due to mechanical or technical changes across measurements), recording error, differences in weather conditions at time of measurements, etc.

b. This could be called a conceptual because there is no sampling frame.

Section 1.2

15.

6L	034	
6H	667899	
7L	00122244	
7H		Stem=Tens
8L	001111122344	Leaf=Ones
8H	5557899	
9L	03	
9H	58	

This display brings out the gap in the data: There are no scores in the high 70s.

17.

a. The stem-and-leaf display appears at the top of the next page.

b. Arguably, a representative crack depth might be around 9-10 μm.

c. This is somewhat subjective, but the display appears quite spread out.

d. No, the distribution is certainly not symmetric. Rather, crack depths appear to be strongly positively skewed.

e. Yes: All of the values 66.5, 76.1, and 81.1 μm appear to be high outliers. (Using an outlier convention described later in the chapter, even the values in the 50s would be considered outliers!)

```
0 | 123333333444444
0 | 55555667788888999
1 | 0000001111224
1 | 5789                  Stem: Tens digit
2 | 0112                  Leaf: Ones digit
2 | 6
3 | 334
3 | 7
4 | 2
4 | 68
5 | 012
5 |
6 |
6 | 6
7 |
7 | 6
8 | 1
```

19.

```
    American |    | French
             |  8 | 1
755543211000 |  9 | 00234566
        9432 | 10 | 2356
        6630 | 11 | 1369
         850 | 12 | 223558
           8 | 13 | 7
             | 14 |
             | 15 | 8          Stem: Tens digit
           2 | 16 |            Leaf: Ones digit
```

The American distribution is positively skewed, but the French distribution is fairly symmetric. Almost half of the American movies are in the 90s, but the French movies are more spread out.

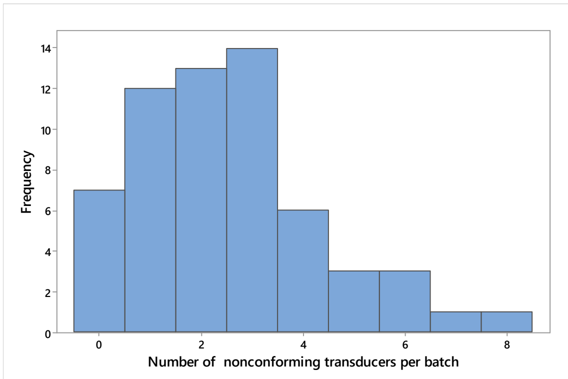
21.

a. *Note: Relative frequencies add to 1.001, not 1, due to rounding.*

Value	Freq.	Rel. Freq. (= Freq. / 60)
0	7	.117
1	12	.200
2	13	.217
3	14	.233
4	6	.100
5	3	.050
6	3	.050
7	1	.017
8	1	.017

b. The number of batches with at most 5 nonconforming items is 7+12+13+14+6+3 = 55, which is a proportion of 55/60 = .917. The proportion of batches with (strictly) fewer than 5 nonconforming items is 52/60 = .867.
Notice that these proportions could also have been computed by using the relative frequencies: e.g., proportion of batches with 5 or fewer nonconforming items = 1 – (.05+.017+.017) = .916; proportion of batches with fewer than 5 nonconforming items = 1 – (.05+.05+.017+.017) = .866.

c. The center of the histogram is somewhere around 2 or 3 and it shows that there is some positive skewness in the data. The histogram also shows that there is a lot of spread/variation in this data.



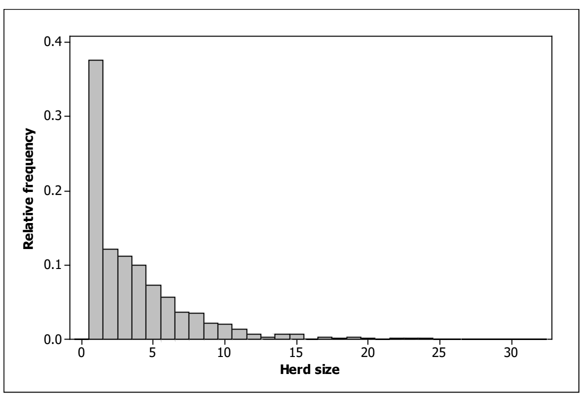
23.

a. 589/1570 = .375.

b. 1 – (589 + 190 + 176 + 157 + 115)/1570 = .218.

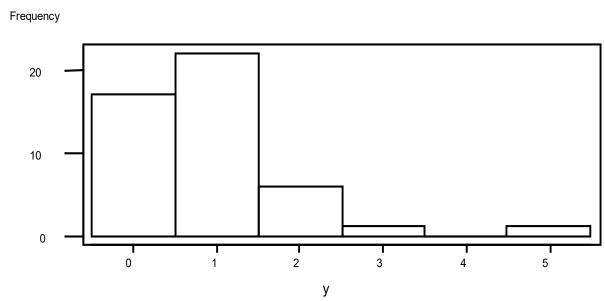
c. (115 + 89 + 57 + 55 + 33 = 31)/1570 = .242.

d. The herd size distribution in the accompanying histogram is extremely positively skewed.

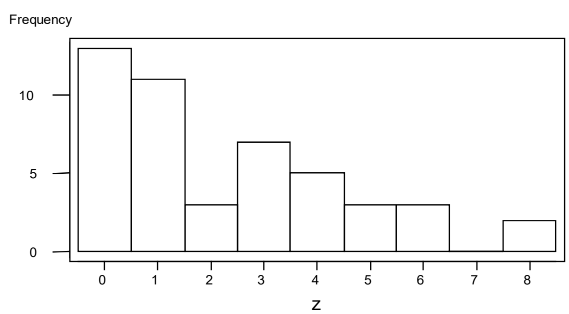


25.

a. A histogram of the y data appears below. From this histogram, the number of subdivisions having no cul-de-sacs (i.e., $y = 0$) is 17/47 = .362, or 36.2%. The proportion having at least one cul-de-sac ($y \geq 1$) is (47 – 17)/47 = 30/47 = .638, or 63.8%. Note that subtracting the number of cul-de-sacs with $y = 0$ from the total, 47, is an easy way to find the number of subdivisions with $y \geq 1$.

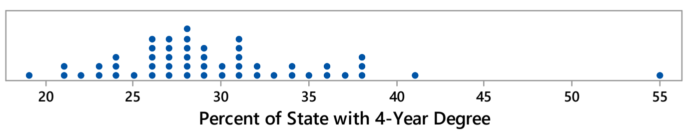


b. A histogram of the z data appears below. From this histogram, the number of subdivisions with at most 5 intersections (i.e., $z \leq 5$) is 42/47 = .894, or 89.4%. The proportion having fewer than 5 intersections ($z < 5$) is 39/47 = .830, or 83.0%.



27.

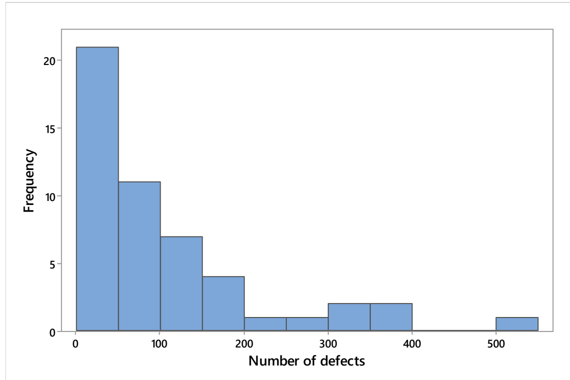
a. The distribution of these by-state values is slightly positively skewed with one extremely high outlier (Washington DC, 54.6%) and two other potential outliers (Massachusetts, 40.5% and West Virginia, 19.2%). The "typical" state percentage appears to be between 25 and 30 percent.



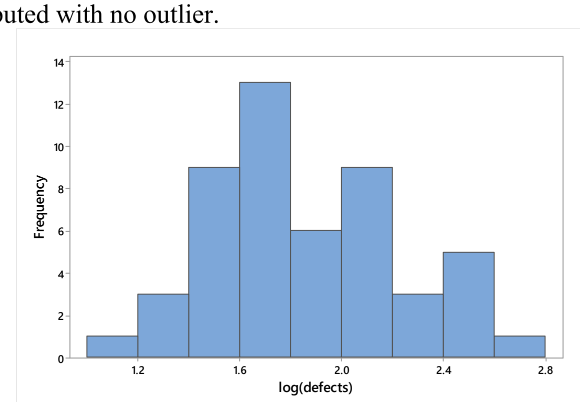
b. No: Since the population sizes of the 50 states + DC are not equal, the mean of these percentages would not equal the overall percentage. (If we knew all 51 population sizes, we could take the appropriate weighted average, effectively re-constructing the total count of people with 4-year degrees and dividing by the total population size.)

29.

a.



b. The transformation substantially changes the shape of the histogram. In particular, while the original variable x = number of defects was strongly positively skewed with an outlier, $\log_{10}(x)$ is reasonably symmetrically distributed with no outlier.



31.

a. 7% of 464 students is roughly (.07)(464) = 32.48, or 32 students. [32/464 = .069, which rounds to .07.]

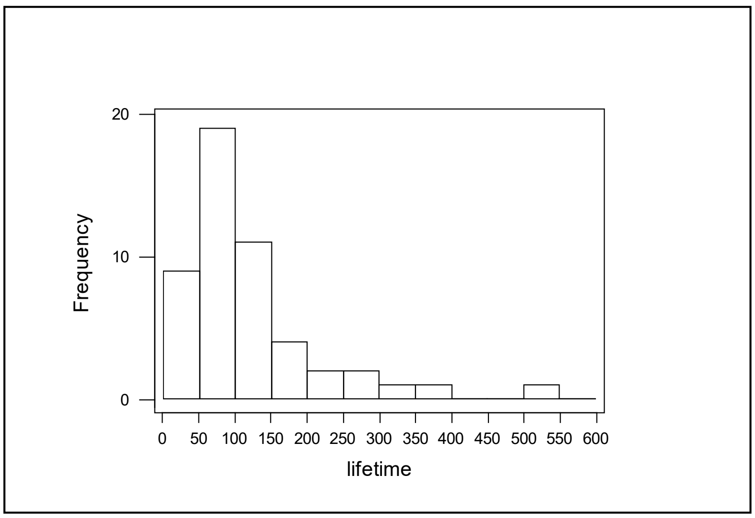
b. 18% + 6% + 5% = 29%.

c. No. Without an upper bound on the last category, we can't even make a density histogram of the data, because we don't know where the last rectangle should end. (*Note:* If we knew that upper bound, say >$5000 means $5001-$10,000, we'd still have to contend with the 7% at exactly $0.)

33.

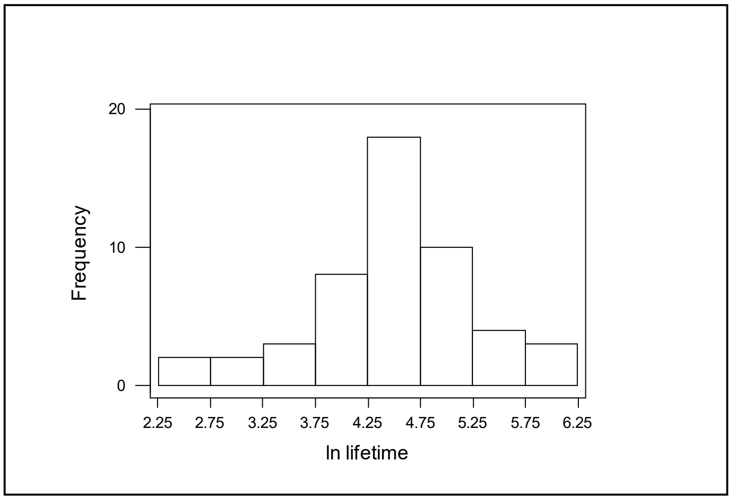
a. The distribution is skewed to the right, or positively skewed. There is a gap in the histogram, and what appears to be an outlier in the 500 – 550 interval.

Class Interval	Frequency	Relative Frequency
0 - 50	9	0.18
50 - 100	19	0.38
100 - 150	11	0.22
150 - 200	4	0.08
200 - 250	2	0.04
250 - 300	2	0.04
300 - 350	1	0.02
350 - 400	1	0.02
400 – 450	0	0.00
450 – 500	0	0.00
500 – 550	1	0.02
	50	1.00



b. The distribution of the natural logs of the original data is much more symmetric than the original.

Class Interval	Frequency	Relative Frequency
2.25 - 2.75	2	0.04
2.75 - 3.25	2	0.04
3.25 - 3.75	3	0.06
3.75 - 4.25	8	0.16
4.25 - 4.75	18	0.36
4.75 - 5.25	10	0.20
5.25 - 5.75	4	0.08
5.75 - 6.25	3	0.06



c. The proportion of lifetime observations in this sample that are less than 100 is .18 + .38 = .56, and the proportion that is at least 200 is .04 + .04 + .02 + .02 + .02 = .14.

35.

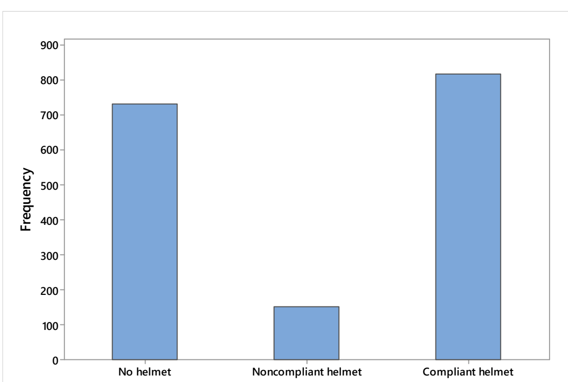
a. The variable here is *helmet status*, a categorical variable. Its possible values are *no helmet, noncompliant helmet,* and *compliant helmet*.

b.

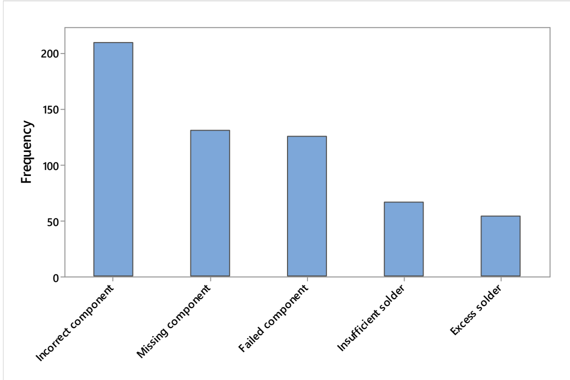
Category	Frequency	Relative Frequency
No helmet	731	.43
Noncompliant helmet	153	.09
Compliant helmet	816	.48
Total	1700	1.00

c. .09 + .48 = .57.

d.



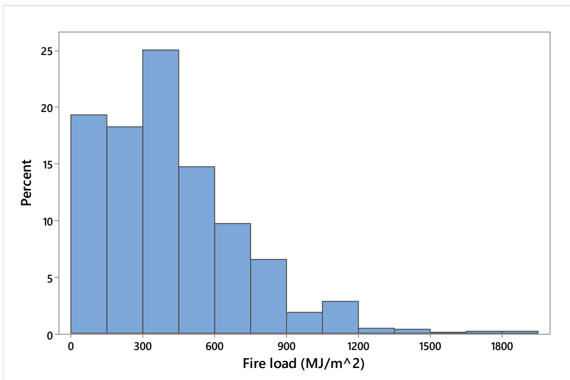
37.



39.

a. The relative frequency distribution is as follows. The relative frequency distribution is almost unimodal and exhibits a large positive skew. The typical middle value is somewhere between 400 and 450, although the skewness makes it difficult to pinpoint more exactly than this.

Class	Rel. Freq.	Class	Rel. Freq.
0-150	.193	1050-1200	.029
150-300	.183	1200-1350	.005
300-450	.251	1350-1500	.004
450-600	.148	1500-1650	.001
600-750	.097	1650-1800	.002
750-900	.066	1800-1950	.002
900-1050	.019		



b. The proportion of the fire loads less than 600 is .193 + .183 + .251 + .148 = .775 (the cumulative proportion for 600). The proportion of loads that are at least 1200 is .005 + .004 + .001 + .002 + .002 = .014 (the opposite of the cumulative proportion for 1200).

c. The proportion of loads between 600 and 1200 is 1 – .775 – .014 = .211.

Section 1.3

41.

a. $\bar{x} = \frac{1}{10}(5+2+\cdots+5+0)$ = 3.5 yards.

b. The two middle values in order are 2 and 2, so $\tilde{x}$ = 2 yards. Todd Gurley's mean rushing gain is artificially increased by the one 16-yard gain, while the median ignores this extreme value.

c. Deleting the 16-yard gain and the 1-yard loss (–1) amounts to trimming 1/10 observations from each end. So, we're talking about the 10% trimmed mean, and the average of the remaining 8 values is $\bar{x}_{\text{tr}(10)}$ = 2.5 yards. As is typically the case, the trimmed mean falls between the median (2 yards) and the mean (3.5 yards).

43.

a. With the one very high outlier (*Wall Street Journal* at over 2.2 million), we anticipate that the mean will be higher than the median.

b. $\bar{x} = \frac{1}{20}(2237601+\cdots+196286)$ = 403,456. In order, the middle two values are 285,129 and 276,445, so $\tilde{x} = \frac{1}{2}(285129+276445)$ = 280,787. Sure enough, the median circulation for the top 20 newspapers is substantially less than the mean, due to the one extremely high outlier.

45. Using software, $\tilde{x} = 92$, $\bar{x}_{\text{tr}(25)} = 95.07$, $\bar{x}_{\text{tr}(10)} = 102.23$, $\bar{x} = 119.3$. The mean is somewhat larger because of positive skewness. Trimming results in a value between the mean and median, and additional trimming gives a value closer to the median.

47.

a. The *reported* values are (in increasing order) 110, 115, 120, 120, 125, 130, 130, 135, and 140. Thus the median of the reported values is 125.

b. 127.6 is reported as 130, so the median is now 130, a very substantial change. When there is rounding or grouping, the median can be highly sensitive to small change.

49. The mean cannot be calculated, because we need the exact value of the two 100+ observations. We can, however, compute median = $\frac{(57+79)}{2} = 68.0$, 20% trimmed mean = 66.2, 30% trimmed mean = 67.5.

51.

a. Manufacturer is a categorical variable.

b. Since Honda is the most frequent manufacturer, arguably Honda is the most representative "value" of this categorical variable.

c. NO. Any numerical coding of these six categories artificially imposes an order on the manufacturers. For instance, sorting alphabetically and sorting by popularity would result in different codings and thus different means and medians. Only the *mode* (i.e., part **b**) makes sense as a representative value.

Section 1.4

53.

a. range = 49.3 – 23.5 = 25.8.

b. From the table below, $\Sigma x = 310.3$, $\overline{x} = 31.03$, $S_{xx} = \Sigma(x_i - \overline{x})^2 = 443.801$, $\Sigma(x_i^2) = 10{,}072.41$.

x_i	$(x_i - \overline{x})$	$(x_i - \overline{x})^2$	x_i^2
29.5	–1.53	2.3409	870.25
49.3	18.27	333.7929	2430.49
30.6	–0.43	0.1849	936.36
28.2	–2.83	8.0089	795.24
28.0	–3.03	9.1809	784.00
26.3	–4.73	22.3729	691.69
33.9	2.87	8.2369	1149.21
29.4	–1.63	2.6569	864.36
23.5	–7.53	56.7009	552.25
31.6	0.57	0.3249	998.56

sample variance $= s^2 = \dfrac{\Sigma(x_i - \overline{x})^2}{n-1} = \dfrac{443.801}{9} = 49.3112$.

c. $s = \sqrt{49.3112} = 7.022$.

d. $s^2 = \dfrac{\Sigma x^2 - (\Sigma x)^2 / n}{n-1} = \dfrac{10{,}072.41 - (310.3)^2 / 10}{9} = 49.3112$.

55.

a. $\overline{x} = \frac{1}{n}\Sigma x_i$ = 14438/5 = 2887.6. The sorted data is: 2781 2856 2888 2900 3013, so the sample median is $\tilde{x}$ = 2888.

b. Subtracting a constant from each observation shifts the data, but does *not* change its sample variance. For example, by subtracting 2700 from each observation we get the values 81, 200, 313, 156, and 188, which are smaller (fewer digits) and easier to work with by hand. The sum of squares of this transformed data is 204210 and its sum is 938, so the computational formula for the variance gives s^2 = [204210 – $(938)^2$/5]/(5 – 1) = 7060.3.

57. Using the computational formula, $S_{xx} = \Sigma x_i^2 - (\Sigma x_i)^2 / n = 3587566 - 9638^2 / 26 =$ 14833.54, so $s^2 = \dfrac{S_{xx}}{n-1} = \dfrac{14833.54}{26-1}$ = 593.34 and s = 24.4. In general, the size of a typical deviation from the sample mean (370.7 sec) is about 24.4 sec. Some observations may deviate from 370.7 by a little more than this, some by less.

59. First, we need $\overline{x} = \frac{1}{n}\sum x_i = \frac{1}{27}(20{,}179) = 747.37$. Then we need the sample standard deviation $s = \sqrt{\frac{24{,}657{,}511 - (20{,}179)^2/27}{27-1}} = 606.89$. The maximum award should be $\overline{x} + 2s = 747.37 + 2(606.89) = 1961.16$, or \$1,961,160. This is quite a bit less than the \$3.5 million that was awarded originally.

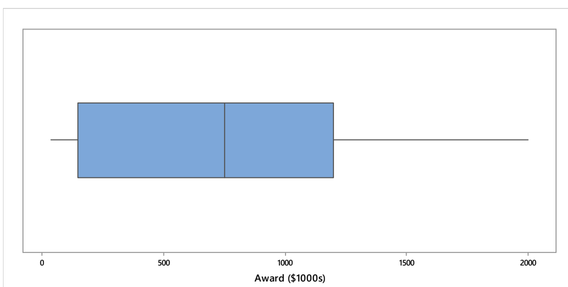
61. Let d denote the fifth deviation. Then $.3 + .9 + 1.0 + 1.3 + d = 0$ or $3.5 + d = 0$, so $d = -3.5$. One sample for which these are the deviations is 3.8, 4.4, 4.5, 4.8, and 0 (obtained by adding 3.5 to each deviation; adding any other number will produce a different sample with the desired property).

63.

a. With $n = 27$ observations, q_1 = median of the 7th and 8th lowest values = (149 + 150)/2 = 149.5, and q_3 = median of the 7th and 8th highest values = (1150 + 1200)/2 = 1175. The iqr = 1175 – 149.5 = 1025.5.

b. Technically, low outliers are not possible here, because q_1 – 1.5iqr < 0 and awards cannot be negative. A high outlier is anything exceeding q_3 + 1.5iqr = 1175 + 1.5(1025.5) = 2713.25, and an extremely high outlier is anything over q_3 + 3iqr = 1175 + 3(1025.5) = 4251.5.

c. The boxplot shows a positively skewed award distribution, with a median award of \$750 thousand and no apparently outliers. (This is consistent with the calculations in part **b**.) *Note*: The box boundaries are not quite at 149.5 and 1175, because software packages calculate quartiles a little differently.

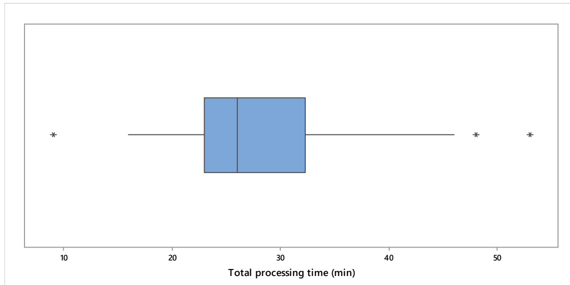


65.

a. The mean is 27.82, the median is 26, and the 5% trimmed mean is 27.38. The mean exceeds the median, in accord with positive skewness. The trimmed mean is between the mean and median, as you would expect.

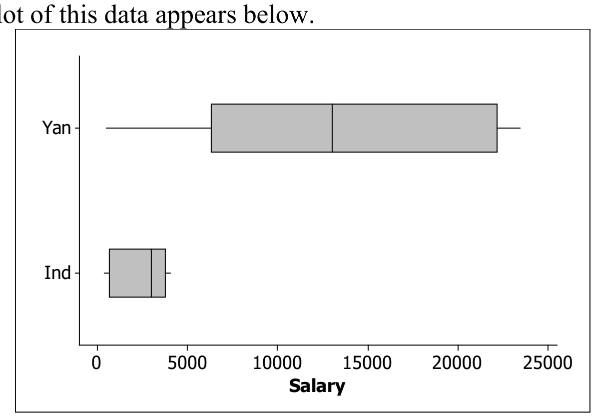
b. From software, the quartiles are roughly 23 and 32, so iqr = 9. Mild outliers are outside 23 – 1.5(9) = 9.5 and 32 + 1.5(9) = 45.5. Extreme outliers are outside 23 – 3(9) = –4 and 32 + 3(9) = 59. Hence, there is one low outlier and there are three high outliers. *Note*: Depending on how the quartiles and iqr are calculated, the observation 46 might or might not be deemed an outlier.

c. The box plot shows two outliers at the high end and one at the low end, but there are no extreme outliers. Because the median is in the lower half of the box, the upper whisker is longer than the lower whisker, and there are two high outliers compared to just one low outlier, the plot suggests positive skewness.



67. The most noticeable feature of the comparative boxplots is that machine 2's sample values have considerably more variation than does machine 1's sample values. However, a typical value, as measured by the median, seems to be about the same for the two machines. The only outlier that exists is from machine 1.

69. A comparative boxplot of this data appears below.



All of the Indian salaries are below the first quartile of Yankee salaries. There is much more variability in the Yankee salaries. Neither team has any outliers.

71. Outliers occur in the 6 a.m. data. The distributions at the other times are fairly symmetric. Variability and the typical values in the data increase a little at the 12 noon and 2 p.m. times. Clearly the 6 a.m. vehicles warrant further investigation!

Supplementary Exercises

73.

a. Each ones place is divided into five sub-intervals: 2.6-2.7, 2.8-2.9, 3.0-3.1, etc.

Males		Females
	2	6
	2	
1	3	0011
	3	22
5444	3	5
776	3	
988	3	8
00	4	
	4	3

Stem: Ones digit
Leaf: Tenths digit

b. Each distribution has some potential outliers, so let's use the medians: $\tilde{x}$ = 3.7 cm for males and 3.15 cm for females, respectively. *Note:* If one distribution were symmetric and the other skewed (or with outliers), it would be preferable to compute the medians for both to allow for an apples-to-apples comparison.

c. Males' aortic root diameters are greater, on average, than females' in this sample (see the medians above). But the women in the sample exhibited much more variability in aortic root diameter than did the men, including some potential high and low outliers.

75.

Flow rate	**Median**	**Lower quartile**	**Upper quartile**	**IQR**	**1.5(IQR)**	**3(IQR)**
125	3.1	2.7	3.8	1.1	1.65	.3
160	4.4	4.2	4.9	0.7	1.05	.1
200	3.8	3.4	4.6	1.2	1.80	3.6

There are no outliers in the three data sets. However, as a comparative boxplot shows, the three data sets differ with respect to their central values (the medians are different) and the data for flow rate 160 is somewhat less variable than the other data sets. Flow rates 125 and 200 also exhibit a small degree of positive skewness.

77.

a. HC data: $\sum_i x_i^2$ = 2618.42 and $\sum_i x_i$ = 96.8, so s^2 = [2618.42 - (96.8)^2/4]/3 = 91.953 and s = 9.59.

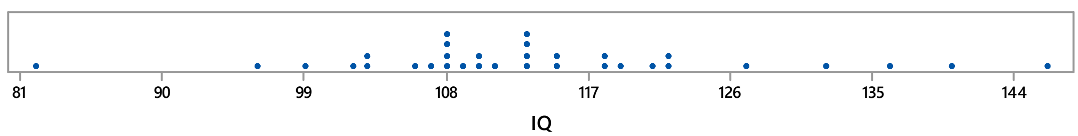
CO data: $\sum_i x_i^2$ = 145645 and $\sum_i x_i$ =735, so s^2 = [145645 - (735)^2/4]/3 = 3529.583 and s = 59.41.

Since the CO data are on a much larger scale, it makes sense that their standard deviation should be larger — standard deviation reflects *absolute* scale.

b. The mean of the HC data is 96.8/4 = 24.2; the mean of the CO data is 735/4 = 183.75. Therefore, the coefficient of variation of the HC data is 9.59/24.2 = .3963, or 39.63%. The coefficient of variation of the CO data is 59.41/183.75 = .3233, or 32.33%. Thus, even though the CO data has a larger standard deviation than does the HC data, it actually exhibits *less* variability (in percentage terms) around its average than does the HC data.

79. $\Sigma x_i = 163.2$. Delete the largest and smallest value, and the 100(1/15)% trimmed mean is $\dfrac{163.2-8.5-15.6}{13} = 10.70$. Delete the next two extreme values, and the 100(2/15)% trimmed mean is $\dfrac{163.2-8.5-8.8-15.6-13.7}{11} = 10.60$. Conveniently, halfway between 1/15 and 2/15 is 1.5/15 = .1 or 10%, so a 10% trimmed mean is halfway between the previous two: $\bar{x}_{\text{tr}(10)} = (10.70 + 10.60)/2 = 10.65$ ppm.

81. As seen in the dotplot below, the IQ distribution for these 33 children is reasonably symmetric, with a mean IQ score of 113.7 and a standard deviation of 12.7. The sample includes three outliers (using the 1.5iqr rule): a low outlier at 82 and two high outliers at 140 and 146.



83.

a. The typical radon level in houses where a child had cancer seems somewhat higher than in no-cancer households. Both distributions are positively skewed. Radon levels of 55, 55, and 85 Bq/m^3 are potential high outliers among the no-cancer households, while an extreme outlier of 210 Bq/m^3 was recorded in one household with a childhood cancer.

1. Cancer		2. No cancer	
9987653	0	33566777889999	
8887666555332111000	1	11111223477	
73322110	2	11449999	
9843	3	389	
5	4		
7	5	55	
	6		
	7		Stem: Tens digit
HI:210	8	5	Leaf: Ones digit

b. With the aid of software, s = 31.7 Bq/m^3 for the cancer households and 17.0 Bq/m^3 for the no-cancer households, suggesting greater variability in the first group. This seemingly contradicts the graph, where the radon distribution on the left appears more concentrated than the one on the right.

c. With the aid of software, iqr = 11.0 for cancer households and 18.0 for non-cancer households. Now the non-cancer households exhibit greater variability in radon levels, which is more consistent with our graph. The culprit here is presumably the extreme value of 210, which greatly influences the standard deviation of the cancer group but has no effect on the iqr of that sample.

85. A table of summary statistics and a stem-and-leaf displayare below. The healthy individuals have higher receptor binding measure on average than the individuals with PTSD. There is also more variation in the healthy individuals' values. The distribution of values for the healthy is reasonably symmetric, while the distribution for the PTSD individuals is negatively skewed.

	PTSD	Healthy
Mean	32.92	52.23
Median	37	51
Std Dev	9.93	14.86
Min	10	23
Max	46	72

	1	0	stem = tens
3	2	058	leaf = ones
9	3	1578899	
7310	4	26	
81	5		
9763	6		
2	7		

87.

a. Mode = .93. It occurs four times in the data set.

b. The *modal category* is the one with the highest (relative) frequency.

89. The measures that are sensitive to outliers are: the mean and the midrange. The mean is sensitive because all values are used in computing it. The midrange is sensitive because it uses only the most extreme values in its computation.

The median, the trimmed mean, and the midquarter are not sensitive to outliers.

The median is the most resistant to outliers because it uses only the middle value (or values) in its computation. The trimmed mean is somewhat resistant to outliers. The larger the trimming percentage, the more resistant the trimmed mean becomes. The midquarter, which uses the quartiles, is reasonably resistant to outliers because both quartiles are resistant to outliers.

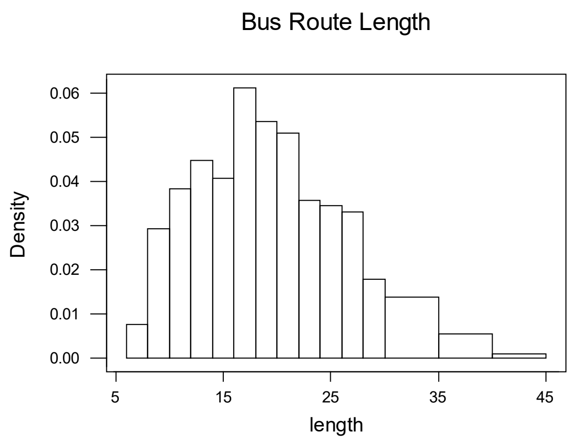
91.

a. Since the constant $\overline{x}$ is subtracted from each x value to obtain each y value, and addition or subtraction of a constant doesn't affect variability, $s_y^2 = s_x^2$ and $s_y = s_x$.

b. Let $a = 1/s$, where s is the sample standard deviation of the x's and also (by part **a**) of the y's. Then $z_i = ay_i \Rightarrow s_z^2 = a^2 s_y^2 = (1/s)^2 s^2 = 1$, and $s_z = 1$. That is, the "standardized" quantities $z_1, \ldots, z_n$ have a sample variance and standard deviation of 1.

93.

a.



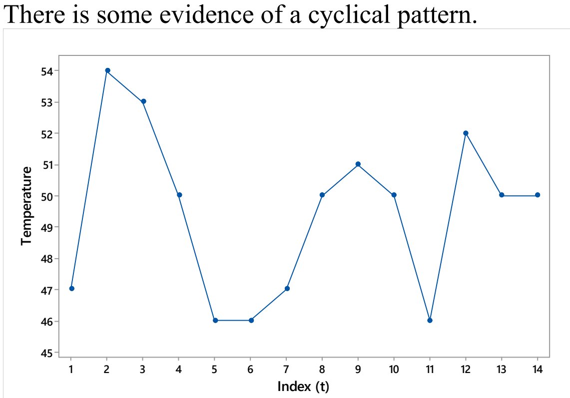
b. Proportion less than 20 = 216/391 = .552. Proportion at least 30 = 40/391 = .102.

c. First compute (.90)(391 + 1) = 352.8. Thus, the 90th percentile should be about the 352nd ordered value. The 351st ordered value lies in the interval 28 - 30. The 352nd ordered value lies in the interval 30 - < 35. There are 27 values in the interval 30 - < 35. We do not know how these values are distributed, however, the smallest value (i.e., the 352nd value in the data set) cannot be smaller than 30. So, the 90th percentile is roughly 30.

d. First compute (.50)(391 + 1) = 196. Thus the median (50th percentile) should be the 196 ordered value. The 174th ordered value lies in the interval 16 -< 18. The next 42 observation lie in the interval 18 -< 20. So, ordered observation 175 to 216 lie in the intervals 18 -< 20. The 196th observation is about in the middle of these. Thus, we would say, the median is roughly 19.

95.

a. There is some evidence of a cyclical pattern.



b. $\bar{x}_2 = .1x_2 + .9\bar{x}_1 = (.1)(54) + (.9)(47) = 47.7; \bar{x}_3 = .1x_3 + .9\bar{x}_2 = (.1)(53) + (.9)(47.7) = 48.23 \approx 48.2;$ etc. As seen below, $\alpha = .1$ gives a smoother series.

t	$\bar{x}_t$ for $\alpha = .1$	$\bar{x}_t$ for $\alpha = .5$
1	47.0	47.0
2	47.7	50.5
3	48.2	51.8
4	48.4	50.9
5	48.2	48.4
6	48.0	47.2
7	47.9	47.1
8	48.1	48.6
9	48.4	49.8
10	48.5	49.9
11	48.3	47.9
12	48.6	50.0
13	48.8	50.0
14	48.9	50.0

c.

$$\begin{aligned}
\bar{x}_t &= \alpha x_t + (1-\alpha)\bar{x}_{t-1} \\
&= \alpha x_t + (1-\alpha)[\alpha x_{t-1} + (1-\alpha)\bar{x}_{t-2}] \\
&= \alpha x_t + \alpha(1-\alpha)x_{t-1} + (1-\alpha)^2[\alpha x_{t-2} + (1-\alpha)\bar{x}_{t-3}] = \cdots \\
&= \alpha x_t + \alpha(1-\alpha)x_{t-1} + \alpha(1-\alpha)^2 x_{t-2} + \ldots + \alpha(1-\alpha)^{t-2}x_2 + (1-\alpha)^{t-1}x_1
\end{aligned}$$

Thus, $\bar{x}_t$ depends on x_t and *all* previous values. As k increases, the coefficient on x_{t-k} decreases (further back in time implies less weight).

d. Not very sensitive, since $(1 - \alpha)^{t-1}$ will be very small.