

AP Statistics

Solutions to Packet 1

Exploring Data

Displaying Distributions with Graphs
Describing Distributions with Numbers

1.1 **FUEL-EFFICIENT CARS** Here is a small part of a data set that describes the fuel economy (in miles per gallon) of 1998 model motor vehicles.

Make and Model	Vehicle Type	Transmission Type	Number of Cylinders	City MPG	Highway MPG
:					
BMW 3181	Subcompact	Automatic	4	22	31
BMW 3181	Subcompact	Manual	4	23	32
Buick Century	Midsized	Automatic	6	20	29
Chevrolet Blazer	Four-wheel drive	Automatic	6	16	30
:					

(a) What are the individuals in this data set?

The individuals are vehicles (or “cars”)

(b) For each individual, what variables are given? Which of these variables are categorical and which are quantitative?

The variables are: vehicle type (categorical), transmission type (categorical), number of cylinders (quantitative), city MPG (quantitative), and highway MPG (quantitative).

1.2 **MEDICAL STUDY VARIABLES** Data from a medical study contain values of many variables for each of the people who were the subjects of the study. Which of the following variables are categorical and which are quantitative?

- (a) Gender (female or male) **categorical**
- (b) Age (years) **quantitative**
- (c) Race (Asian, black, white, or other) **categorical**
- (d) Smoker (yes or no) **categorical**
- (e) Systolic blood pressure (millimeters of mercury) **quantitative**
- (f) Level of calcium in the blood (micrograms per milliliter) **quantitative**

1.3 You want to compare the “size” of several statistics textbooks. Describe at least three possible numerical variables that describe the “size” of a book. In what units would you measure each variable? **Possible answers (units):**

- Number of pages (pages)
- Number of chapters (chapters)
- Number of words (words)
- Weight or mass (pounds, ounces, kilograms . . .)
- Height and/or width and/or thickness (inches, centimeters . . .)
- Volume (cubic inches, cubic centimeters . . .)

1.4 Popular magazines often rank cities in terms of how desirable it is to live and work in each city. Describe five variables that you would measure for each city if you were designing such a study. Give reasons for your choices. **Possible answers include:**

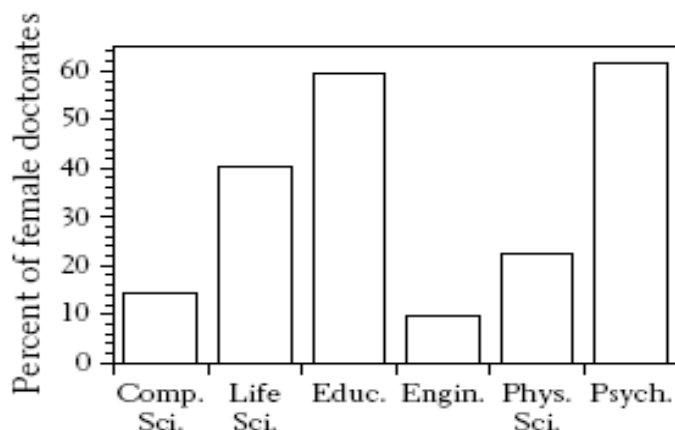
unemployment rate, average (mean or median) income, quality/availability of public transportation, number of entertainment and cultural events, housing costs, crime statistics, population, population density, number of automobiles, various measures of air quality, commuting times (or other measures of traffic), parking availability, taxes, quality of schools.

1.5 **FEMALE DOCTORATES** Here are data on the percent of females among people earning doctorates in 1994 in several fields of study.

Computer science	15.4%	Life sciences	40.7%
Education	60.8%	Physical sciences	21.7%
Engineering	11.1%	Psychology	62.2%

(a) Present these data in a well-labeled bar graph.

The bars are given in the same order as the data in the table—the most obvious way—but that is not necessary (since the variable is nominal, not ordinal)



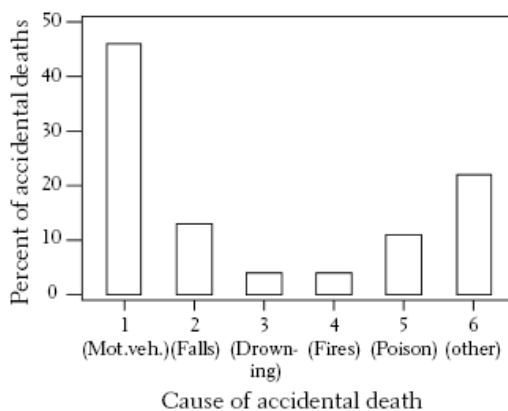
(b) Would it also be correct to use a pie chart to display these data? If so, construct the pie chart. If not, explain why not.

A pie chart would not be appropriate, since the different entries in the table do not represent parts of a single whole.

1.6 **ACCIDENTAL DEATHS** In 1997 there were 92,353 deaths from accidents in the United States. Among these were 42,340 deaths from motor vehicle accidents, 11,858 from falls, 10,163 from poisoning, 4051 from drowning, and 3601 from fires.

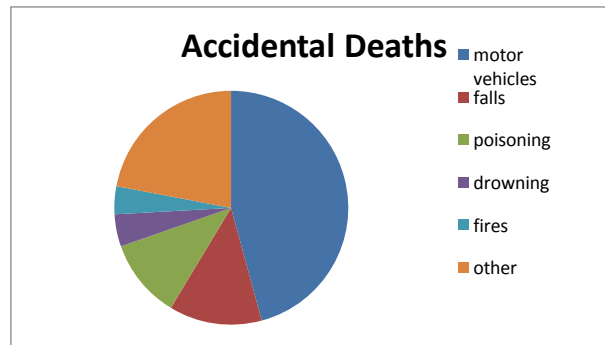
(a) Find the percent of accidental deaths from each of the causes, rounded to the nearest percent. What percent of deaths were due to other causes? (b) Make a well-labeled bar graph of the distribution of causes of accidental deaths. Be sure to include an “other causes” bar.

Motor Vehicles = 46 %
 Falls = 13 %
 Drowning = 4 %
 Fires = 4 %
 Poisoning = 11 %
 Other causes = 22 %



(c) Would it also be correct to use a pie-chart to display these data? If so, construct the pie-chart. If not, explain why not.

A pie chart could also be used, since the categories represent parts of a whole (all accidental deaths).



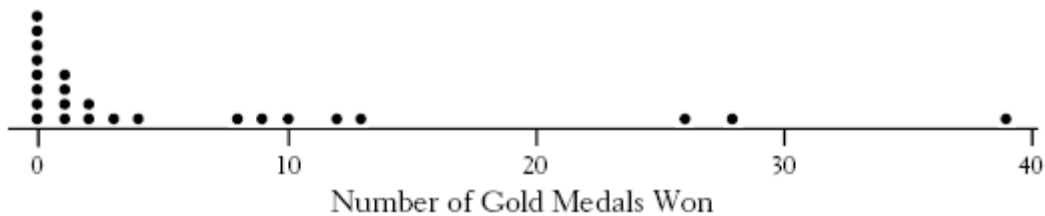
1.7 **OLYMPIC GOLD** Athletes like Cathy Freeman, Rulon Gardner, Ian Thorpe, Marion Jones and Jenny Thompson captured public attention by winning gold medals in the 2000 Summer Olympic Games in Sydney, Australia. Table 1.2 displays the total number of gold medals won by several countries in the 2000 Summer Olympics.

TABLE 1.2 Gold medals won by selected countries in the 2000 Summer Olympics

Country	Gold medals	Country	Gold medals
Sri Lanka	0	Netherlands	12
Qatar	0	India	0
Vietnam	0	Georgia	0
Great Britain	28	Kyrgyzstan	0
Norway	10	Costa Rica	0
Romania	26	Brazil	0
Switzerland	9	Uzbekistan	1
Armenia	0	Thailand	1
Kuwait	0	Denmark	2
Bahamas	1	Latvia	1
Kenya	2	Czech Republic	2
Trinidad and Tobago	0	Hungary	8
Greece	13	Sweden	4
Mozambique	1	Uruguay	0
Kazakhstan	3	United States	39

Source: BBC Olympics Web site.

Make a dotplot to display these data. Describe the distribution of number of gold medals won.



The distribution has a peak at 0 and a long right tail. There are eight outliers, with the most severe being 26, 28, and 39. The spread is 0 to 39 and the center is 1.

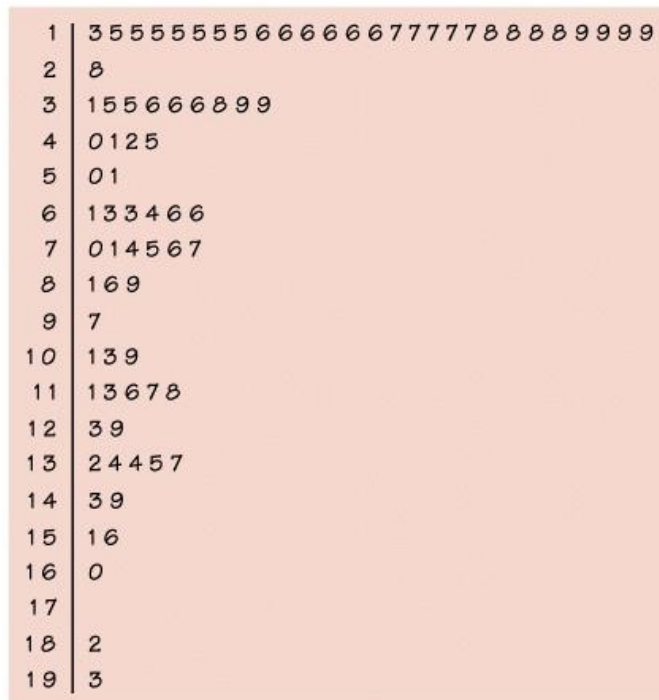
1.9 **MICHIGAN COLLEGE TUITIONS** There are 91 colleges and universities in Michigan. Their tuitions and fees for the 1999 to 2000 school year run from \$1260 at Kalamazoo Valley Community College to \$19,258 at Kalamazoo College. See the stemplot below.

(a) What do the stems and leaves represent in the stemplot? Have the data been rounded?

Stems = thousands, leaves = hundreds. The data have been rounded to the nearest \$100.

(b) Describe the shape, center, and spread of the tuition distribution. Are there any outliers?

The distribution is skewed strongly to the right, with a peak at the 1 stem. The spread is approximately 18,000 (\$1300 to \$19,300). The center is 45 (\approx \$4500). The observations 182 (\approx \$18,200) and 193 (\approx \$19,300) appear to be outliers.

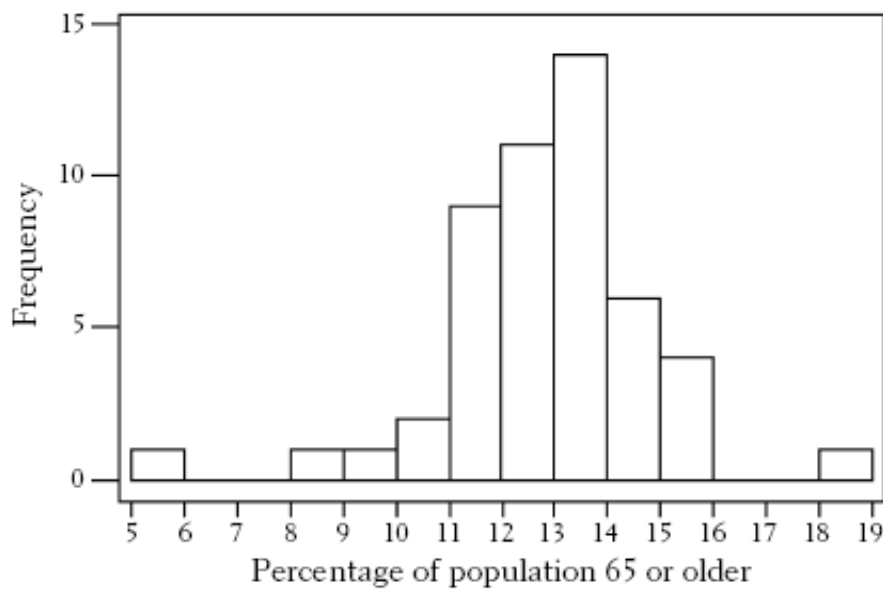


1.12 **WHERE DO OLDER FOLKS LIVE?** The table below gives the percentage of residents aged 65 or older in each of the 50 states.

Percent of the population in each state aged 65 or older

State	Percent	State	Percent	State	Percent
Alabama	13.1	Louisiana	11.5	Ohio	13.4
Alaska	5.5	Maine	14.1	Oklahoma	13.4
Arizona	13.2	Maryland	11.5	Oregon	13.2
Arkansas	14.3	Massachusetts	14.0	Pennsylvania	15.9
California	11.1	Michigan	12.5	Rhode Island	15.6
Colorado	10.1	Minnesota	12.3	South Carolina	12.2
Connecticut	14.3	Mississippi	12.2	South Dakota	14.3
Delaware	13.0	Missouri	13.7	Tennessee	12.5
Florida	18.3	Montana	13.3	Texas	10.1
Georgia	9.9	Nebraska	13.8	Utah	8.8
Hawaii	13.3	Nevada	11.5	Vermont	12.3
Idaho	11.3	New Hampshire	12.0	Virginia	11.3
Illinois	12.4	New Jersey	13.6	Washington	11.5
Indiana	12.5	New Mexico	11.4	West Virginia	15.2
Iowa	15.1	New York	13.3	Wisconsin	13.2
Kansas	13.5	North Carolina	12.5	Wyoming	11.5
Kentucky	12.5	North Dakota	14.4		

(a) Construct a histogram to display these data. Record your class intervals and counts.

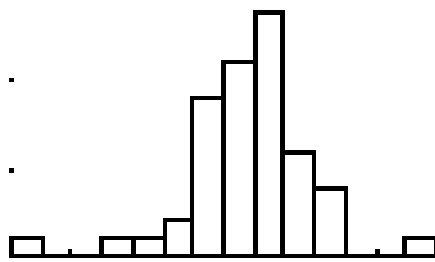


Percent	Freq.
5.0-5.9	1
6.0-6.9	0
7.0-7.9	0
8.0-8.9	1
9.0-9.9	1
10.0-10.9	2
11.0-11.9	9
12.0-12.9	11
13.0-13.9	14
14.0-14.9	6
15.0-15.9	4
16.0-16.9	0
17.0-17.9	0
18.0-18.9	1
Total	50

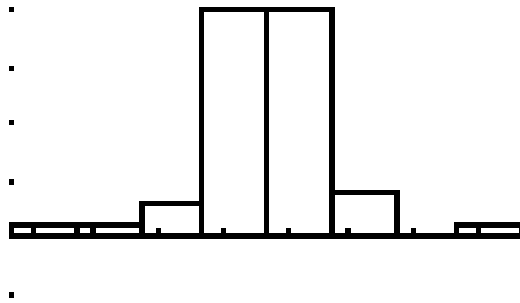
(b) Describe the distribution of people aged 65 and over in the states.

The distribution is slightly skewed to the left with a peak at the class 13.0–13.9. There is one outlier in each tail of the distribution.

(c) Enter the data into your calculator's statistics list editor. Make a histogram using a window that matches your histogram from part (a). Copy the calculator histogram and mark the scales on your paper.



(d) Use the calculator's zoom feature to generate a histogram. Copy this histogram onto your paper and mark the scales.

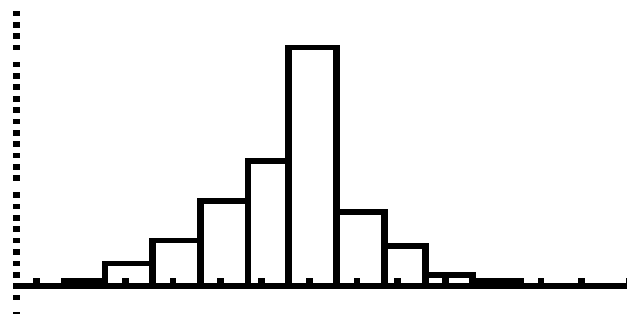
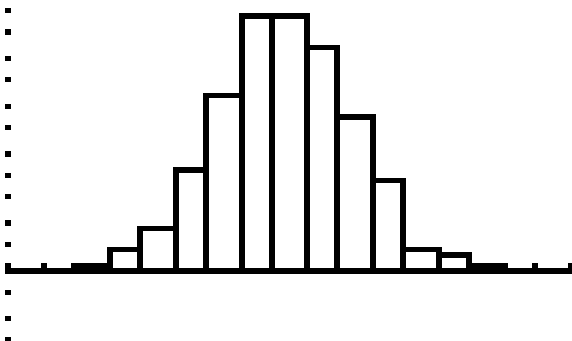


1.15 **CHEST OUT SOLDIER** In 1846, a published paper provided chest measurements (in inches) of 5738 Scottish militiamen. This table displays the data in summary form.

Chest measurements (inches) of 5738 Scottish militiamen.

Chest size	Count	Chest size	Count
33	3	41	934
34	18	42	658
35	81	43	370
36	185	44	92
37	420	45	50
38	749	46	21
39	1073	47	4
40	1079	48	1

(a) Use your graphing calculator to make a histogram of data presented in summary form like the chest measurements of Scottish militiamen. Store chest size into L1 and corresponding counts into L2. Use window X[32, 49] by Y[-300, 1100]. Graph below. (Try using your calculator's Zoom Stat command. What happens to the histogram?)

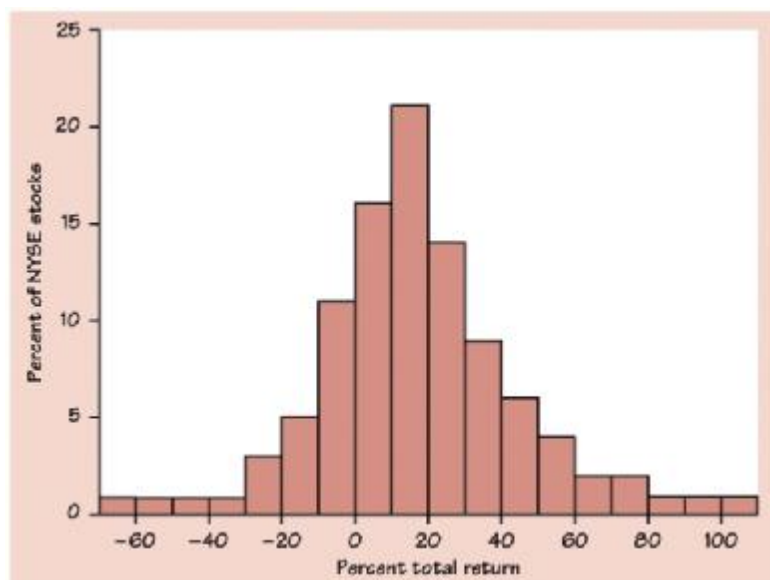


(b) Describe the shape, center, and spread of the chest measurements distribution. Why might this information be useful?

The distribution is symmetric with a peak at class (chest size) 40. The center is also located at 40. The spread is 15 (33 to 48). Assuming that the sample is representative of all members of the population, the distribution would provide a useful guide to those making clothing for the militiamen.

From the frequency table, it is easy to estimate the percentage of all militiamen who have a certain chest size. The production of uniforms can reflect this distribution.

1.16 **STOCK RETURNS** The total return on a stock is the change in its market price plus any dividend payments made. Total return is usually expressed as a percent of the beginning price. The histogram below shows the distribution of total returns for all 1528 stocks listed on the New York State Exchange in one year. This is a histogram of the percents in each class rather than a histogram of counts.



The distribution of percent total return for all New York Stock Exchange common stocks in one year.

(a) Describe the overall shape of the distribution of total returns.

Roughly symmetric, though it might be viewed as SLIGHTLY skewed to the right.

(b) What is the approximate center of this distribution? (For now, take the center to be the value with roughly half the stocks having lower returns and half having higher returns.)

About 15%. (39% of the stocks had a total return less than 10%, while 60% had a return less than 20%. This places the center of the distribution somewhere between 10% and 20%.)

(c) Approximately what were the smallest and largest total returns? (This describes the spread of the distribution)

The smallest return was between -70% and -60%, while the largest was between 100% and 110%.

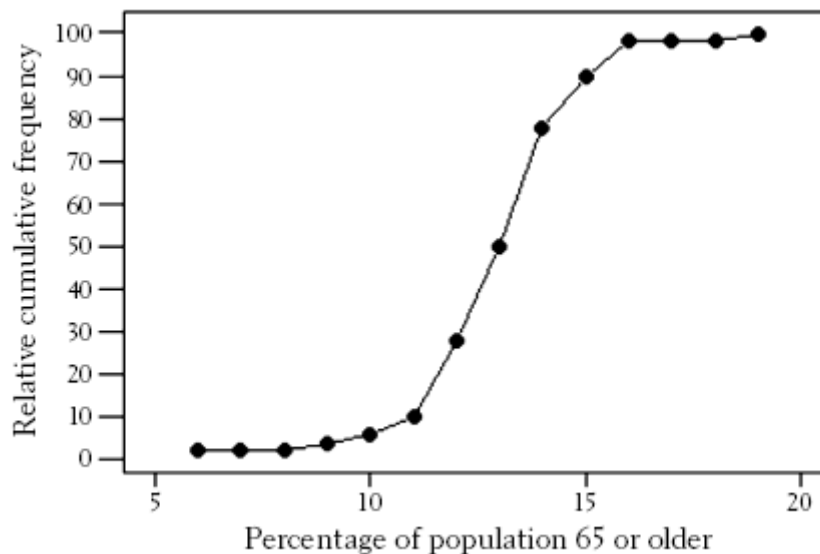
(d) A return less than zero means that an owner of the stock lost money. About what percent of all stocks lost money?

23% (1 + 1 + 1 + 1 + 3 + 5 + 11).

1.19 **OLDER FOLKS, II** In Exercise 1.12, you constructed a histogram of the percentage of people aged 65 or older in each state.

(a) Construct a relative cumulative frequency graph (ogive) for these data.

Percent	Cumulative frequency	Relative cumulative frequency	Percent	Cumulative frequency	Relative cumulative frequency
5.0-5.9	1	2%	12.0-12.9	25	50%
6.0-6.9	1	2%	13.0-13.9	39	78%
7.0-7.9	1	2%	14.0-14.9	45	90%
8.0-8.9	2	4%	15.0-15.9	49	98%
9.0-9.9	3	6%	16.0-16.9	49	98%
10.0-10.9	5	10%	17.0-17.9	49	98%
11.0-11.9	14	28%	18.0-18.9	50	100%



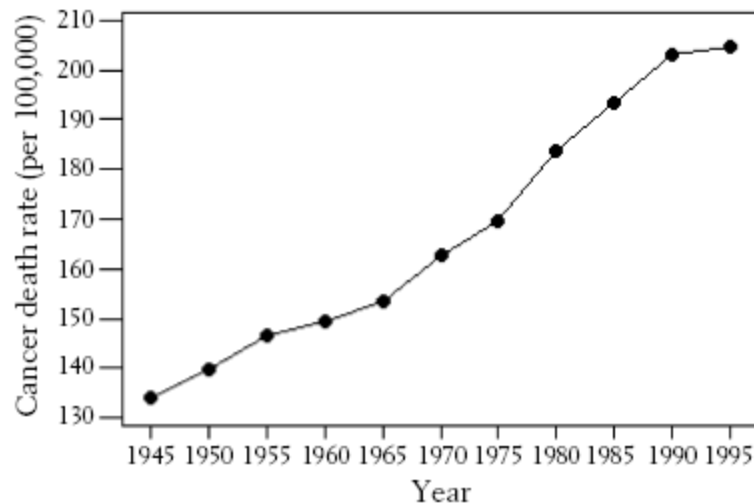
(b) Use your ogive to answer the following questions:

- In what percentage of states was the percentage of “65 or older” less than 15%? = 90% since the point (15, 90) lies on the ogive
- What is the 40th percentile of this distribution, and what does it tell us? = 12.4 %, since the horizontal line drawn from 40% on the vertical axis intersects the ogive at a point whose horizontal coordinate is approximately 12.4% Less than 40% of states have 12.4% or less of their population aged 65 or older.
- What percentile is associated with Missouri? About the 75th percentile.

1.21 **CANCER DEATHS** Here are data on the rate of deaths from cancer (deaths per 100,000 people) in the United States over the 50-year period from 1945 to 1995:

Year:	1945	1950	1955	1960	1965	1970	1975	1980	1985	1990	1995
Deaths:	134.0	139.8	146.5	149.2	153.5	162.8	169.7	183.9	193.3	203.2	204.7

(a) Construct a time plot for these data. Describe what you see in a few sentences.

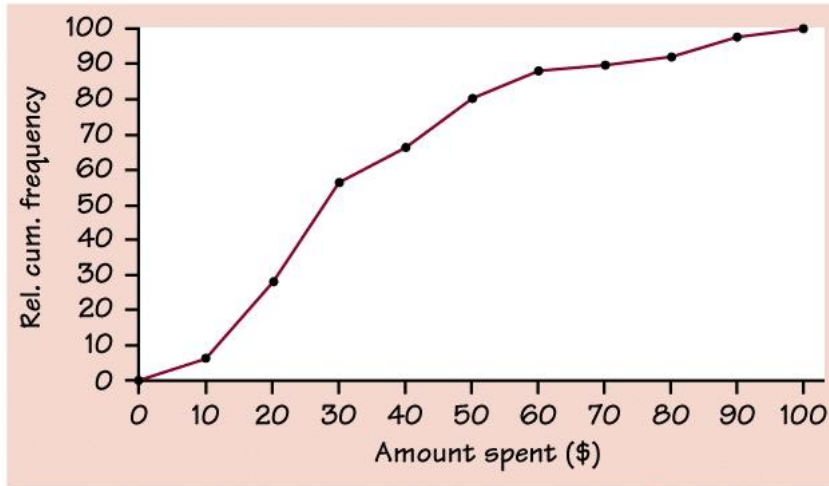


The cancer death rate has risen steadily from 1945 to 1995, with the largest increase occurring in the period 1975–1980.

(b) Do these data suggest that we have made no progress in treating cancer? Explain.

No, the slower rate of increase during the period 1990–1995 suggests that some progress was made during that time (at least in terms of treating the disease effectively). Other things to remember include that we diagnose more accurately now. Life span is longer and the elderly make up a higher proportion of the population. However, we have yet to see a decrease in the death rate, indicating that much work remains to be done in terms of actively preventing the disease.

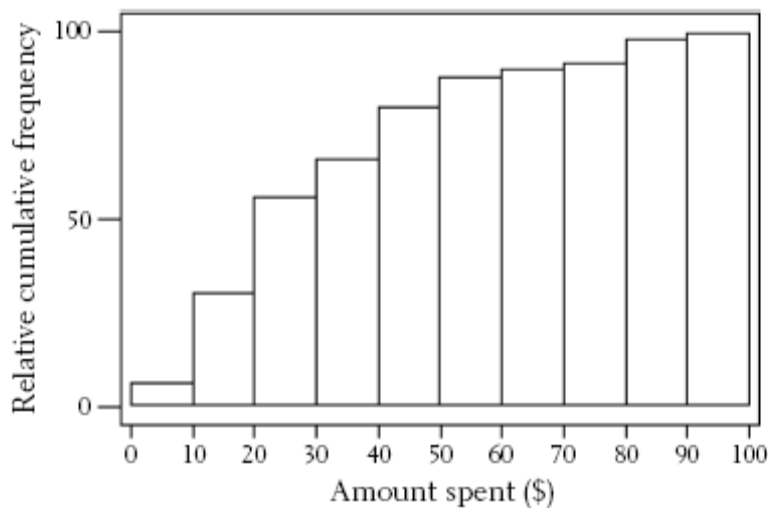
1.20 **SHOPPING SPREE, II** The figure below is an ogive of the amount spent by grocery shoppers.



(a) Estimate the center of this distribution. Explain your method. **The center corresponds to the 50th percentile. Draw a horizontal line from the value 50 on the vertical axis and determine the point on the ogive where the line intersects the ogive. Then draw a vertical line from this point to the horizontal axis. The line intersects the axis at approximately \$28. Thus, \$28 is the estimate of the center.**

(b) At what percentile would the shopper who spent \$17.00 fall? **The 20th percentile**

(c) Draw the histogram that corresponds to the ogive.

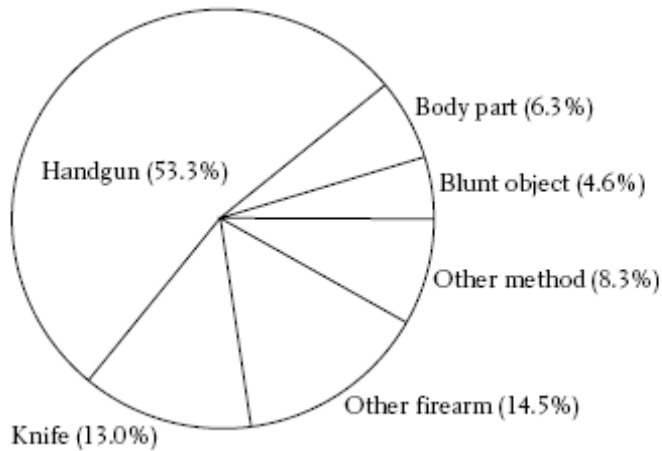


1.23 **GENDER EFFECTS IN VOTING** Political party preference in the United States depends in part on the age, income, and gender of the voter. A political scientist selects a large sample of registered voters. For each voter, she records gender, age, household income, and whether they voted for the Democratic or Republican candidate in the last congressional election. Of these 4 variables, which are categorical and which are quantitative?

Gender, party voted for: Categorical

Age, income: Quantitative

1.25 **MURDER WEAPONS** The 1999 *Statistical Abstract of the United States* reports FBI data on murders for 1997. In that year, 53.3% of all murders were committed with handguns, 14.5% with other firearms, 13.0% with knives, 6.3% with a part of the body (usually the hands or feet), and 4.6% with blunt objects. Make a graph to display these data. Do you need an “other methods” category?



An “Other Methods” category is needed because the sum of the percentages for the other categories is less than 100%.



1.31 Joey's first 14 quiz grades in a marking period were:

86 84 91 75 78 80 74 87 76 96 82 90 98 93

(a) Use the formula to calculate the mean. $n = 14$, $\sum x = 1190$. The mean is $\bar{x} = \frac{\sum x}{n} = \frac{1190}{14} = 85$.

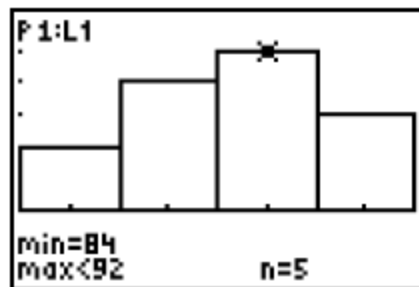
(b) Suppose Joey has an unexcused absence for the fifteenth quiz and he receives a score of zero. Determine his final quiz average. What property of the mean does this situation illustrate? Write a sentence about the effect of the zero on Joey's quiz average that mentions this property.

If the 15th score is 0, then $n = 15$, $\sum x = 1190$, and the new mean is $\bar{x} = \frac{\sum x}{n} = \frac{1190}{15} = 79.3$

The fact that this value of \bar{x} is less than 85 indicates the nonresistance property of \bar{x} . The extremely low outlier at 0 pulled the mean below 85.

(c) What kind of plot would best show Joey's distribution of grades? Assume an 8-point grading scale (A: 93 to 100, B: 85 to 92, etc.) Make an appropriate plot.

Given a rather small data set like this one, a stem plot would normally be preferable. But since we are interested in letter grades in this case, perhaps the histogram would be most informative. Here is a histogram, with the widths of the bars specified to correspond to letter grades: D (68–75), C (76–83), B (84–91), and A (92–100). Both plots show a fairly balanced or symmetric distribution, with the histogram suggesting a slight skewness to the left. (Note that the mean and the median are the same (85))



1.32 **SSHA SCORES** The Survey of Study Habits and Attitudes (SSHA) is a psychological test that evaluates college students' motivation, study habits, and attitudes toward school. A private college gives the SSHA to a sample of 18 of its incoming first-year women students. Their scores were:

154 109 137 115 152 140 154 178 101
103 126 126 137 165 165 129 200 148

(a) Make a stemplot of these data. The overall shape of the distribution is irregular, as often happens when only a few observations are available. Are there any potential outliers? About where is the center of the distribution (the score with half the scores above it and half below)? What is the spread of the scores (ignoring any outliers)?

```

10 | 139
11 | 5
12 | 669
13 | 77
14 | 08
15 | 244
16 | 55
17 | 8
18 |
19 |
20 | 0

```

200 is a potential outlier.
The center is approximately 140.
The spread (excluding 200) is 77.

(b) Use one-vbl stats on your calculator to find the mean and the median of the distribution. Which is larger: the mean or the median? Explain why.

The mean $\bar{x} = 2539/18 = 141.058$. The median = average of ninth and tenth scores = 138.5.
The mean is larger than the median because of the outlier at 200, which pulls the mean towards the long right tail of the distribution.

1.33 Suppose a major league baseball team's mean yearly salary for a player is \$1.2 million, and that the team has 25 players on its active roster. What is the team's annual payroll for players?

Since the mean = \$1.2 million and the number of players on the team is $n = 25$, the team's annual payroll is $(\$1.2 \text{ million}) (25) = \30 million

If you knew only the median salary, would you be able to answer that question? Why or why not?

If you knew only the median salary, you would not be able to calculate the total payroll because you cannot determine the sum of all 25 values from the median. You can only do so when the arithmetic average of the values is provided.

1.35 **U. S. INCOMES** The distribution of individual incomes in the United States is strongly skewed to the right. In 1997, the mean and median incomes of the top 1% of Americans were \$330,000 and \$675,000. Which of these numbers is the mean and which is the median? Explain your reasoning.

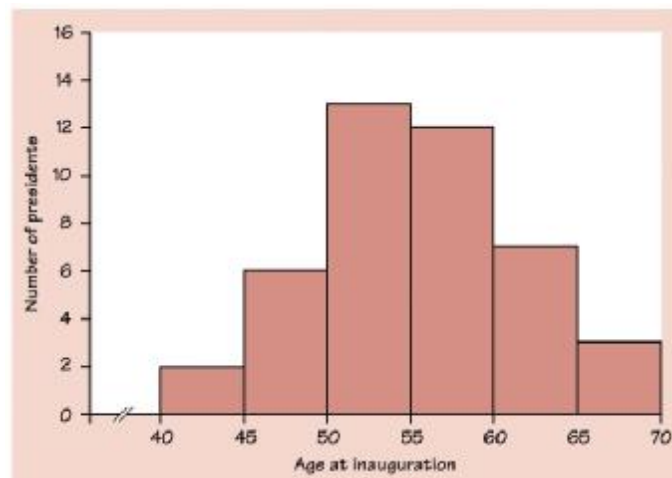
Mean = \$675,000; median = \$330,000. The mean is nonresistant to the effects of the extremely high incomes in the right tail of the distribution. It will therefore be larger than the median.

1.37 **HOW OLD ARE PRESIDENTS?** We return to the data on presidential ages.

TABLE 1.4 Ages of the presidents at inauguration

President	Age	President	Age	President	Age
Washington	57	Lincoln	52	Hoover	54
J. Adams	61	A. Johnson	56	F. D. Roosevelt	51
Jefferson	57	Grant	46	Truman	60
Madison	57	Hayes	54	Eisenhower	61
Monroe	58	Garfield	49	Kennedy	43
J. Q. Adams	57	Arthur	51	L. B. Johnson	55
Jackson	61	Cleveland	47	Nixon	56
Van Buren	54	B. Harrison	55	Ford	61
W. H. Harrison	68	Cleveland	55	Carter	52
Tyler	51	McKinley	54	Reagan	69
Polk	49	T. Roosevelt	42	G. Bush	64
Taylor	64	Taft	51	Clinton	46
Fillmore	50	Wilson	56	G. W. Bush	54
Pierce	48	Harding	55		
Buchanan	65	Coolidge	51		

Here is a histogram of the age data:



(a) From the shape of the histogram, do you expect the mean to be much less than the median, about the same as the median, or much greater than the median? Explain.

The mean and median should be approximately equal since the distribution is roughly symmetric.

(b) Find the five-number summary and verify your expectation from (a).

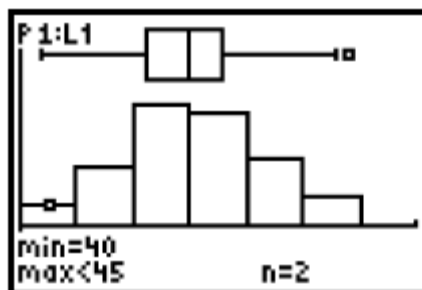
Five-number summary: 42, 51, 55, 58, 69 $\bar{x} = 2357/43 = 54.8$

As expected, median and are very similar.

(c) What is the range of the middle half of the ages of new presidents?

Between Q1 and Q3: 51 to 58.

(d) Construct by hand a (modified) boxplot of the ages of the presidents.



(e) On your calculator, define Plot 1 to be a histogram using the list named PREZ. Define Plot 2 to be a modified boxplot. Use your calculator's zoom command to generate a graph. To remove the overlap, adjust your viewing window $Y[-6, 22]$. Then graph. Use TRACE to inspect values. Press the up and down cursor keys to toggle between plots. Is there an outlier? If so, what is it?

The point 69 is an outlier; this is Ronald Reagan's age on inauguration day. W. H. Harrison was 68, but that is not an outlier according to the $1.5 \times \text{IQR}$ test.

1.38 Is the interquartile range a resistant measure of spread? _____ Give an example of a small data set that supports your answer.

Yes, IQR is resistant.

Take the data set 1, 2, 3, 4, 5, 6, 7, 8 as an example. In this case the median = 4.5, $Q1 = 2.5$, $Q3 = 6.5$, and $\text{IQR} = 4$. Changing any "extreme" value (that is, any value outside the interval between $Q1$ and $Q3$) will have no effect on the IQR. For example, if 8 is changed to 88, the IQR will still equal 4.

1.39 **SHOPPING SPREE, III** The figure below displays computer output for the data on amount spent by grocery shoppers. The first box is from DataDesk. The second box is from Minitab.

DataDesk

Summary of spending	
No Selector	
Percentile	25
Count	50
Mean	34.7022
Median	27.8550
StdDev	21.6974
Min	3.11000
Max	93.3400
Lower ith %tile	19.2700
Upper ith %tile	45.4000

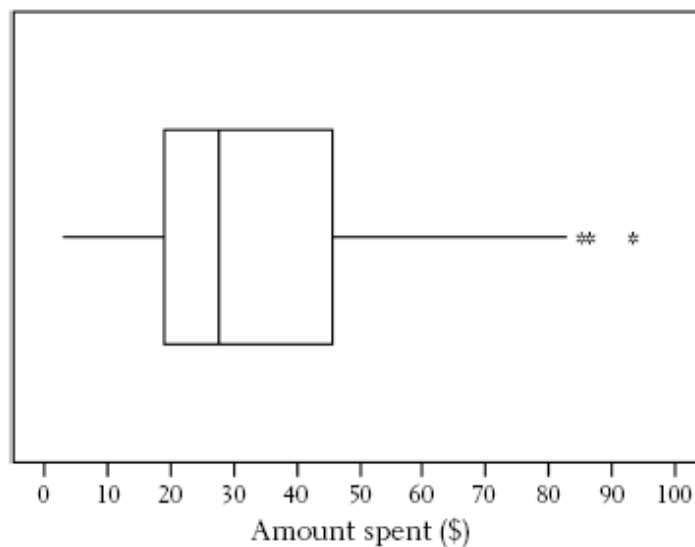
Minitab

Descriptive Statistics						
Variable	N	Mean	Median	TrMean	StDev	SEMean
spending	50	34.70	27.85	32.92	21.70	3.07
Variable	Min	Max	Q1	Q3		
spending	3.11	93.34	19.06	45.72		

(a) Find the total amount spent by shoppers.

$\bar{x} = 34.7022$ and $n = 50$: Thus, total amt. spent = $(34.7022)(50) = \$1735.11$.

(b) Make a boxplot from the computer output. Make sure to check for outliers!!



The boxplot indicates the presence of several outliers. According to the 1.5 (IQR) rule, these outliers are 85.76, 86.37, and 93.34.

1.40 **PHOSPHATE LEVELS** The level of various substances in the blood influences our health. Here are measurements of the level of phosphate in the blood of a patient, in milligrams of phosphate per deciliter of blood, made on 6 consecutive visits to a clinic:

5.6 5.2 4.6 4.9 5.7 6.4

A graph of only 6 observations gives little information, so we proceed to compute the mean and standard deviation.

(a) Calculate the mean from the definition. (show your work)

$$\bar{x} = \frac{32.4}{6} = 5.4$$

(b) Calculate the standard deviation from the definition. (show your work)

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{(5.6-5.4)^2 + (5.2-5.4)^2 + (4.6-5.4)^2 + (4.9-5.4)^2 + (5.7-5.4)^2 + (6.4-5.4)^2}{5} = 0.412$$

$$s = \sqrt{s^2} = 0.6419$$

(c) Now enter the data into your calculator to obtain \bar{x} and s . Do the results agree with your hand calculations? **YESJ**

1.43 This is a standard deviation contest. You must choose four numbers from the whole numbers 0 to 10, with repeats allowed.

(a) Choose four numbers that have the smallest possible standard deviation. **5, 5, 5, 5**

(b) Choose four numbers that have the largest possible standard deviation. **0, 0, 10, 10**

(c) Is more than one choice possible in either (a) or (b)? Explain. **For (a), any set of four identical numbers will have $s = 0$. For (b), the answer is unique; here is a rough description of why. We want to maximize the “spread-out”-ness of the numbers (that is what standard deviation measures), so 0 and 10 seem to be reasonable choices based on that idea. We also want to make each individual squared deviation $(x_1 - \bar{x})^2, (x_2 - \bar{x})^2, (x_3 - \bar{x})^2,$ and $(x_4 - \bar{x})^2$ as large as possible. If we choose 0, 10, 10, 10—or 10, 0, 0, 0—we make the first squared deviation (7.5^2) , but the other three are only (2.5^2) . Our best choice is two at each extreme, which makes all four squared deviations equal to 5^2**

1.44 **COCKROACHES!** Maria measures the lengths of 5 cockroaches that she finds at school. Here are her results (in inches):

1.4 2.2 1.1 1.6 1.2

(a) Find the mean and standard deviation of Maria’s measurements.

$$\bar{x} = 1.5 \text{ in}, \quad s = 0.436 \text{ in}$$

(b) Maria’s science teacher is upset to discover that she has measured the cockroach lengths in inches rather than centimeters. (There are 2.54 cm in 1 inch.) She gives Maria two minutes to report the mean and standard deviation of the 5 cockroaches in centimeters. Maria succeeded. Will you?

To obtain \bar{x} and s in centimeters, multiply the results in inches by 2.54:

$$\bar{x} = 3.81 \text{ cm}, \quad s = 1.107 \text{ cm.}$$

(c) Considering the 5 cockroaches that Maria found as a sample from the population of all cockroaches at her school, what would you estimate as the average length of the population of all cockroaches? How sure of your estimate are you? **The average cockroach length can be estimated as the mean length of the five sampled cockroaches: that is, 1.5 inches. This is, however, a questionable estimate, because the sample is so small.**

1.45 **RAISING TEACHERS' PAY** A school system employs teachers at salaries between \$30,000 and \$60,000. The teachers' union and the school board are negotiating the form of next year's increase in the salary schedule. Suppose that every teacher is given a flat \$1000 raise.

- (a) How much will the mean salary raise? __\$1000__ the median salary? __\$1000__
- (b) Will a flat \$1000 raise increase the spread as measured by the difference between the quartiles?
No. Each quartile will increase by \$1000, thus the difference $Q3 - Q1$ will remain the same.
- (c) Will a flat \$1000 raise increase the spread as measured by the standard deviation of the salaries?
No. The standard deviation remains unchanged when the same amount is added to each data value

HW #7 54, 63, 64, 67, 68, 69

1.54 **\bar{x} AND s ARE NOT ENOUGH** The mean \bar{x} and standard deviation s measure center and spread but are not a complete description of a distribution. Data sets with different shapes can have the same mean and standard deviation. To demonstrate this fact, use your calculator to find \bar{x} and s for the following two small data sets. Then make a stemplot of each and comment on the shape of each distribution.

Data A: 9.14 8.14 8.74 8.77 9.26 8.10 6.13 3.10 9.13 7.26 4.74
 Data B: 6.58 5.76 7.71 8.84 8.47 7.04 5.25 5.56 7.91 6.89 12.50

Set A	Set B
3 1	5 257
4 7	6 58
5	7 079
6 1	8 48
7 2	9
8 1177	10
9 112	11
	12 5

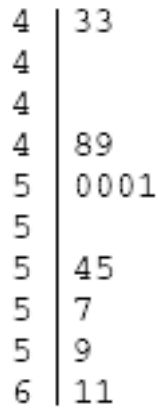
The means and standard deviations are basically the same. For Set A, $\bar{x} = 7.501$ and $s \approx 2.032$, while for Set B, $\bar{x} = 7.501$ and $s \approx 2.031$.

Set A is skewed to the left, while Set B has a high outlier.

1.63 **PRESIDENTIAL ELECTIONS** Here are the percents of the popular vote won by the successful candidate in each of the presidential elections from 1948 to 2000:

Year:	1948	1952	1956	1960	1964	1968	1972	1976	1980	1984	1988	1992	1996	2000
Percent:	49.6	55.1	57.4	49.7	61.1	43.4	60.7	50.1	50.7	58.8	53.9	43.2	49.2	47.9

(a) Make a stemplot of the winners' percents. (Round to whole numbers and use split stems)

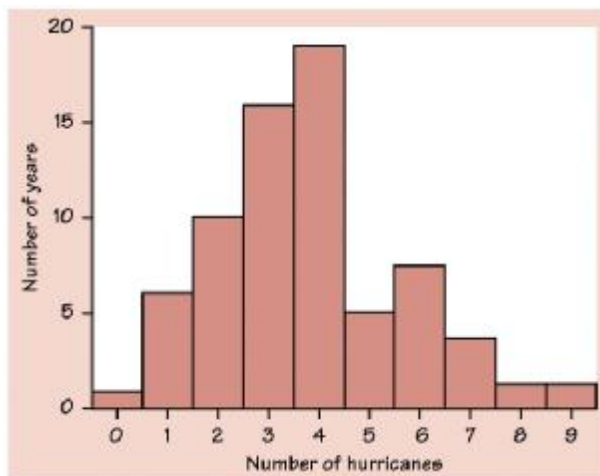


(b) What is the median percent of the vote won by the successful candidate in presidential elections? (Work with the unrounded data.) **Median = 50.4**

(c) Call an election a landslide if the winner's percent falls at or above the 3rd quartile. Find the 3rd quartile. Which elections were landslides?

$Q_3 = 57.4$. Landslides occurred in 1956, 1964, 1972, and 1984.

1.64 **HURRICANES** The histogram below shows the number of hurricanes reaching the east coast of the United States each year over a 70-year period. Give a brief description of the overall shape of this distribution. About where does the center of the distribution lie?



Slightly skewed to the right, centered at 4.

1.67 **WAL-MART STOCK** The rate of return on a stock is its change in price plus any dividends paid. Rate of return is usually measured in percent of the starting value. We have data on the monthly rates of return for the stock of Wal-Mart stores for the years 1973 to 1991, the first 19 years Wal-Mart was listed on the New York Stock Exchange. There are 228 observations.

The figure below displays output from statistical software that describes the distribution of these data. The stems in the stemplot are the tens digits of the percent returns. The leaves are the ones digits. The stemplot uses split stems to give a better display. The software gives high and low outliers separately from the stemplot rather than spreading out the stemplot to include them.

(a) Give the five-number summary for monthly returns on Wal-Mart stock.

Min = -34.04, Q1 = -2.95, Med = 3.47, Q3 = 8.45, Max = 58.68.

```

Mean = 3.064
Standard deviation = 11.49

N = 228   Median = 3.4691
Quartiles = -2.950258, 8.4511

Decimal point is 1 place to the right of the colon

Low:  -34.04255  -31.25000  -27.06271  -26.61290

-1 : 985
-1 : 444443322222110000
-0 : 99998877766666665555
-0 : 44444444333333322222222222111111100
 0 : 000001111111111122222233333334444444
 0 : 5555555555555555566666666666777777888888888899999
 1 : 00000000011111111122233334444
 1 : 55566667889
 2 : 011334

High:  32.01923  41.80531  42.05607  57.89474  58.67769

```

(b) Describe in words the main features of the distribution.

The distribution is fairly symmetric, with a single peak in the high single digits (5 to 9). There are no gaps, but four “low” outliers and five “high” outliers are listed separately.

(c) If you had \$1000 worth of Wal-Mart stock at the beginning of the best month during these 19 years, how much would your stock be worth at the end of the month? If you had \$1000 worth of stock at the beginning of the worst month, how much would your stock be worth at the end of the month?

58.68% of \$1000 is \$586.60. The stock is worth \$1586.50 at the end of the best month. In the worst month, the stock lost $1000(.3404) = \$340.40$, so the \$1000 decreased in worth to $1000 - 340.40 = \$659.60$.

(d) Find the interquartile range (IQR) for the Wal-Mart data. Are there any outliers according to the $1.5 \times \text{IQR}$ criterion? Does it appear to you that the software uses this criterion in choosing which observations to report separately as outliers?

$$\text{IQR} = Q3 - Q1 = 8.45 - (-2.95) = 11.4$$

$$1.5 \times \text{IQR} = 17.1$$

$$Q1 - (1.5 \times \text{IQR}) = -2.95 - 17.1 = -20.05$$

$$Q3 + (1.5 \times \text{IQR}) = 8.45 + 17.1 = 25.55$$

The four “low” and five “high” values are all outliers according to the criterion. It does appear that SPLUS uses the $1.5 \times \text{IQR}$ criterion to identify outliers.

1.68 A study of jury awards in civil cases (such as injury, product liability, and medical malpractice) in Chicago showed that the median award was about \$8000. But the mean award was \$69,000. Explain how this great difference between the two measures of center can occur.

The difference in the mean and median indicates that the distribution of awards is skewed sharply to the right—that is, there are some very large awards.

1.69 You want to measure the average speed of vehicles on the interstate highway on which you are driving. You adjust your speed until the number of vehicles passing you equals the number you are passing. Have you found the mean speed or the median speed of vehicles on the highway? Explain your answer.

The median—half are traveling faster than you, and half are traveling slower. (Actually, you have found a median—it could be that a whole range of speeds, say from 56 mph to 58 mph, might satisfy this condition.)