EXAMPLE 1.5 WATCH THAT CAFFEINE!

The U.S. Food and Drug Administration limits the amount of caffeine in a 12-ounce can of carbonated beverage to 72 milligrams (mg). Data on the caffeine content of popular soft drinks are provided in Table 1.1. How does the caffeine content of these drinks compare to the USFDA's limit?

TABLE 1.1 Caffeine content (in milligrams) for an 8-ounce serving of popular soft drinks

Brand	Caffeine (mg per 8-oz. serving)	Brand	Caffeine (mg per 8-oz. serving)
A&W Cream Soda	20	IBC Cherry Cola	16
Barq's root beer	15	Kick	38
Cherry Coca-Cola	23	KMX	36
Cherry RC Cola	29	Mello Yello	35
Coca-Cola Classic	23	Mountain Dew	37
Diet A&W Cream Soda	15	Mr. Pibb	27
Diet Cherry Coca-Cola	23	Nehi Wild Red Soda	33
Diet Coke	31	Pepsi One	37
Diet Dr. Pepper	28	Pepsi-Cola	25
Diet Mello Yello	35	RC Edge	47
Diet Mountain Dew	37	Red Flash	27
Diet Mr. Pibb	27	Royal Crown Cola	29
Diet Pepsi-Cola	24	Ruby Red Squirt	26
Diet Ruby Red Squirt	26	Sun Drop Cherry	43
Diet Sun Drop	47	Sun Drop Regular	43
Diet Sunkist Orange Soda	28	Sunkist Orange Soda	28
Diet Wild Cherry Pepsi	24	Surge	35
Dr. Nehi	28	TAB	31
Dr. Pepper	28	Wild Cherry Pepsi	25

Source: National Soft Drink Association, 1999.

The caffeine levels spread from 15 to 47 milligrams for these soft drinks. You could make a dotplot for these data, but a stemplot might be preferable due to the large spread.

How to construct a stemplot:

Step 1: Separate each observation into a *stem* consisting of all but the rightmost digit and a *leaf*, the final digit. A&W Cream Soda has 20 milligrams of caffeine per 8-ounce serving. The number 2 is the stem and 0 is the leaf.

Step 2: Write the stems vertically in increasing order from top to bottom, and draw a vertical line to the right of the stems. Go through the data, writing each leaf to the right of its stem and spacing the leaves equally.

```
1 | 5 5 6
2 | 0 3 9 3 3 8 7 4 6 8 4 8 8 7 5 7 9 6 8 5
3 | 1 5 7 8 6 5 7 3 7 5 1
4 | 7 7 3 3
```

Step 3: Write the stems again, and rearrange the leaves in increasing order out from the stem.

Step 4: Title your graph and add a key describing what the stems and leaves represent. Figure 1.4(a) shows the completed stemplot.

What *shape* does this distribution have? It is difficult to tell with so few stems. We can get a better picture of the caffeine content in soft drinks by "splitting stems." In Figure 1.4(a), the values from 10 to 19 milligrams are placed on the "1" stem. Figure 1.4(b) shows another stemplot of the same data. This time, values having leaves 0 through 4 are placed on one stem, while values ending in 5 through 9 are placed on another stem.

Now the bimodal (two-peaked) *shape* of the distribution is clear. Most soft drinks seem to have between 25 and 29 milligrams or between 35 and 38 milligrams of caffeine per 8-ounce serving. The center of the distribution is 28 milligrams per 8-ounce serving. At first glance, it looks like none of these soft drinks even comes close to the USFDA's caffeine limit of 72 milligrams per 12-ounce serving. Be careful! The values in the stemplot are given in milligrams per 8-ounce serving. Two soft drinks have caffeine levels of 47 milligrams per 8-ounce serving. A 12-ounce serving of these beverages would have 1.5(47) = 70.5 milligrams of caffeine. Always check the units of measurement!

CAFFEINE CONTENT (MG) PER 8-OUNCE SERVING OF VARIOUS SOFT DRINKS

```
1 | 5 5 6
2 | 0 3 3 3 4 4 5 5 6 6 7 7 7 8 8 8 8 8 9 9
3 | 1 1 3 5 5 5 6 7 7 7 8
4 | 3 3 7 7
```

(a)

Key:
3|5 means the soft drink contains 35 mg of caffeine per 8-ounce serving.

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

(b)

Key:
2|8 means the soft drink contains 28 mg of caffeine per 8-ounce serving.

FIGURE 1.4 Two stemplots showing the caffeine content (mg) of various soft drinks. Figure 1.4(b) improves on the stemplot of Figure 1.4(a) by splitting stems.

Here are a few tips for you to consider when you want to construct a stemplot:

- Whenever you split stems, be sure that each stem is assigned an equal number of possible leaf digits.
- There is no magic number of stems to use. Too few stems will result in a skyscraper-shaped plot, while too many stems will yield a very flat "pancake" graph.

- Five stems is a good minimum.
- You can get more flexibility by *rounding* the data so that the final digit after rounding is suitable as a leaf. Do this when the data have too many digits.

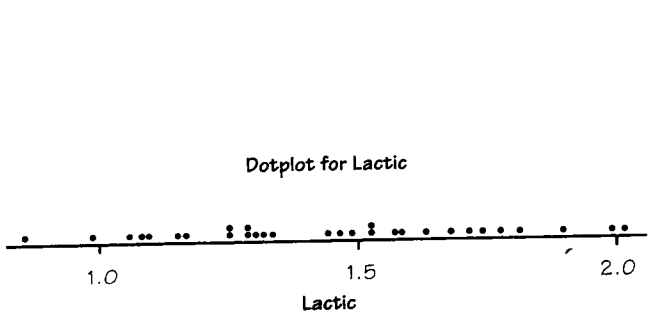
The chief advantages of dotplots and stemplots are that they are easy to construct and they display the actual data values (unless we round). Neither will work well with large data sets. Most statistical software packages will make dotplots and stemplots for you. That will allow you to spend more time making sense of the data.

TECHNOLOGY TOOLBOX *Interpreting computer output*

As cheddar cheese matures, a variety of chemical processes take place. The taste of mature cheese is related to the concentration of several chemicals in the final product. In a study of cheddar cheese from the Latrobe Valley of Victoria, Australia, samples of cheese were analyzed for their chemical composition. The final concentrations of lactic acid in the 30 samples, as a multiple of their initial concentrations, are given below.[6]

A dotplot and a stemplot from the Minitab statistical software package are shown in Figure 1.5. The dots in the dotplot are so spread out that the distribution seems to have no distinct shape. The stemplot does a better job of summarizing the data.

0.86	1.53	1.57	1.81	0.99	1.09	1.29	1.78	1.29	1.58
1.68	1.90	1.06	1.30	1.52	1.74	1.16	1.49	1.63	1.99
1.15	1.33	1.44	2.01	1.31	1.46	1.72	1.25	1.08	1.25




```
Stem-and-leaf of Lactic      N = 30
Leaf Unit = 0.010

 1    8   6
 2    9   9
 5   10   689
 7   11   56
11   12   5599
14   13   013
(3)  14   469
13   15   2378
 9   16   38
 7   17   248
 4   18   1
 3   19   09
 1   20   1
```

FIGURE 1.5 Minitab dotplot and stemplot for cheese data.

TECHNOLOGY TOOLBOX *Interpreting computer output (continued)*

Notice how the data are recorded in the stemplot. The "leaf unit" is 0.01, which tells us that the stems are given in tenths and the leaves are given in hundredths. We can see that the *spread* of the lactic acid concentrations is from 0.86 to 2.01. Where is the *center* of the distribution? Minitab counts the number of observations from the bottom up and from the top down and lists those counts to the left of the stemplot. Since there are 30 observations, the "middle value" would fall between the 15th and 16th data values from either end—at 1.45. The (3) to the far left of this stem is Minitab's way of marking the location of the "middle value." So a typical sample of mature cheese has 1.45 times as much lactic acid as it did initially. The distribution is roughly symmetrical in *shape*. There appear to be no *outliers*.

EXERCISES

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	1
Qatar	0	India	1
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.

1.8 ARE YOU DRIVING A GAS GUZZLER? Table 1.3 displays the highway gas mileage for 32 model year 2000 midsize cars.

TABLE 1.3 Highway gas mileage for model year 2000 midsize cars

Model	MPG	Model	MPG
Acura 3.5RL	24	Lexus GS300	24
Audi A6 Quattro	24	Lexus LS400	25
BMW 740I Sport M	21	Lincoln-Mercury LS	25
Buick Regal	29	Lincoln-Mercury Sable	28
Cadillac Catera	24	Mazda 626	28
Cadillac Eldorado	28	Mercedes-Benz E320	30
Chevrolet Lumina	30	Mercedes-Benz E430	24
Chrysler Cirrus	28	Mitsubishi Diamante	25
Dodge Stratus	28	Mitsubishi Galant	28
Honda Accord	29	Nissan Maxima	28
Hyundai Sonata	28	Oldsmobile Intrigue	28
Infiniti I30	28	Saab 9-3	26
Infiniti Q45	23	Saturn LS	32
Jaguar Vanden Plas	24	Toyota Camry	30
Jaguar S/C	21	Volkswagon Passat	29
Jaguar X200	26	Volvo S70	27

(a) Make a dotplot of these data.

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

1.9 MICHIGAN COLLEGE TUITIONS There are 81 colleges and universities in Michigan. Their tuition and fees for the 1999 to 2000 school year run from $1260 at Kalamazoo Valley Community College to $19,258 at Kalamazoo College. Figure 1.6 (next page) shows a stemplot of the tuition charges.

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

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

1.10 DRP TEST SCORES There are many ways to measure the reading ability of children. One frequently used test is the Degree of Reading Power (DRP). In a research study on third-grade students, the DRP was administered to 44 students.[7] Their scores were:

40	26	39	14	42	18	25	43	46	27	19
47	19	26	35	34	15	44	40	38	31	46
52	25	35	35	33	29	34	41	49	28	52
47	35	48	22	33	41	51	27	14	54	45

Display these data graphically. Write a paragraph describing the distribution of DRP scores.

```
 1 | 3555555566666677777888899999
 2 | 8
 3 | 155666899
 4 | 0125
 5 | 01
 6 | 133466
 7 | 014567
 8 | 169
 9 | 7
10 | 139
11 | 13678
12 | 39
13 | 24457
14 | 39
15 | 16
16 | 0
17 |
18 | 2
19 | 3
```

FIGURE 1.6 Stemplot of the Michigan tuition and fee data, for Exercise 1.9.

1.11 SHOPPING SPREE! A marketing consultant observed 50 consecutive shoppers at a supermarket. One variable of interest was how much each shopper spent in the sto Here are the data (in dollars), arranged in increasing order:

3.11	8.88	9.26	10.81	12.69	13.78	15.23	15.62	17.00	17.39
18.36	18.43	19.27	19.50	19.54	20.16	20.59	22.22	23.04	24.47
24.58	25.13	26.24	26.26	27.65	28.06	28.08	28.38	32.03	34.98
36.37	38.64	39.16	41.02	42.97	44.08	44.67	45.40	46.69	48.65
50.39	52.75	54.80	59.07	61.22	70.32	82.70	85.76	86.37	93.34

(a) Round each amount to the nearest dollar. Then make a stemplot using tens of dollars as the stem and dollars as the leaves.

(b) Make another stemplot of the data by splitting stems. Which of the plots shows the shape of the distribution better?

(c) Describe the shape, center, and spread of the distribution. Write a few sentences describing the amount of money spent by shoppers at this supermarket.

Displaying quantitative variables: histograms

Quantitative variables often take many values. A graph of the distribution is clearer if nearby values are grouped together. The most common graph of the distribution of one quantitative variable is a **histogram.**

EXAMPLE 1.6 PRESIDENTIAL AGES AT INAUGURATION

How old are presidents at their inaugurations? Was Bill Clinton, at age 46, unusually young? Table 1.4 gives the data, the ages of all U.S presidents when they took office.

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		

How to make a histogram:

Step 1: Divide the range of the data into classes of equal width. Count the number of observations in each class. The data in Table 1.4 range from 42 to 69, so we choose as our classes

$$40 \le \text{president's age at inauguration} < 45$$
$$45 \le \text{president's age at inauguration} < 50$$
$$\vdots$$
$$65 \le \text{president's age at inauguration} < 70$$

Be sure to specify the classes precisely so that each observation falls into exactly one class. Martin Van Buren, who was age 54 at the time of his inauguration, would fall into the third class interval. Grover Cleveland, who was age 55, would be placed in the fourth class interval.

Here are the counts:

Class	Count
40–44	2
45–49	6
50–54	13
55–59	12
60–64	7
65–69	3

Step 2: Label and scale your axes and title your graph. Label the horizontal axis "Age at inauguration" and the vertical axis "Number of presidents." For the classes we chose, we should scale the horizontal axis from 40 to 70, with tick marks 5 apart. The vertical axis contains the scale of counts and should range from 0 to at least 13.

Step 3: Draw a bar that represents the count in each class. The base of a bar should cover its class, and the bar height is the class count. Leave no horizontal space between the bars (unless a class is empty, so that its bar has height 0). Figure 1.7 shows the completed histogram.

Graphing note: It is common to add a "break-in-scale" symbol (//) on an axis that does not start at 0, like the horizontal axis in this example.

Interpretation:

Center: It appears that the typical age of a new president is about 55 years, because 55 is near the center of the histogram.

Spread: As the histogram in Figure 1.7 shows, there is a good deal of variation in the ages at which presidents take office. Teddy Roosevelt was the youngest, at age 42, and Ronald Reagan, at age 69, was the oldest.

Shape: The distribution is roughly symmetric and has a single peak (unimodal).

Outliers: There appear to be no outliers.

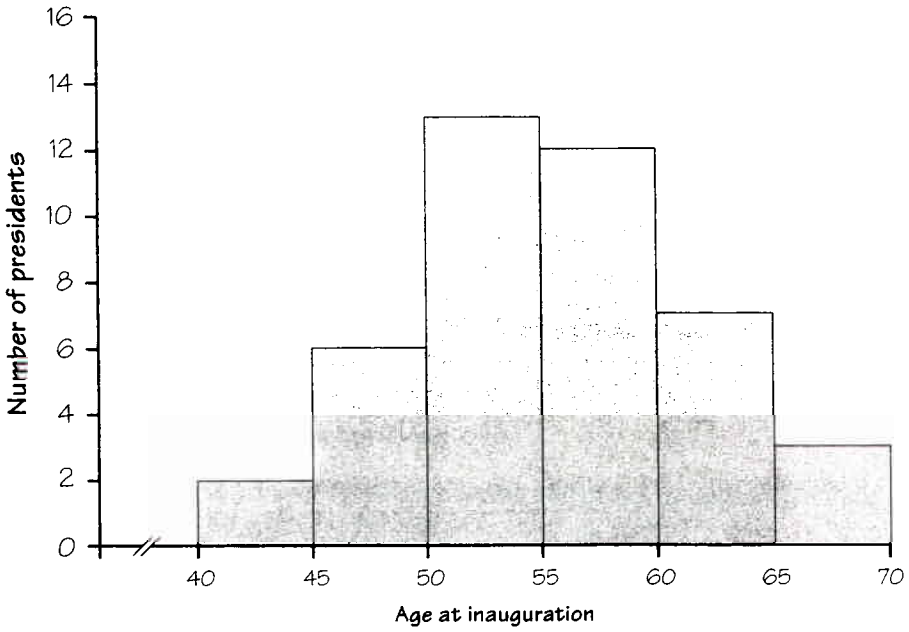


FIGURE 1.7 The distribution of the ages of presidents at their inaugurations, from Table 1.4.