CHAPTER 2

ORGANIZING AND GRAPHING DATA

Opening Example

What future careers interest high school students the most? Information technology? Business/management? Health care? Or some other careers? A sample survey of high school students showed the percentage of students intending to enter each of these fields. In this sample of 1023 high school students, 15% said that they planned to pursue a career in health care. (See Case Study 2-1)

2.1 Raw Data

Definition

Data recorded in the sequence in which they are collected and before they are processed or ranked are called ***raw data***.

Table 2.1 Ages of 50 Students

21	19	24	25	29	34	26	27	37	33
18	20	19	22	19	19	25	22	25	23
25	19	31	19	23	18	23	19	23	26
22	28	21	20	22	22	21	20	19	21
25	23	18	37	27	23	21	25	21	24

Table 2.2 Status of 50 Students

J	F	SO	SE	J	J	SE	J	J	J
F	F	J	F	F	F	SE	SO	SE	J
J	F	SE	SO	SO	F	J	F	SE	SE
SO	SE	J	SO	SO	J	J	SO	F	SO
SE	SE	F	SE	J	SO	F	J	SO	SO

2.2 Organizing and Graphing Qualitative Data

- Frequency Distributions
- Relative Frequency and Percentage Distributions
- Graphical Presentation of Qualitative Data

TABLE 2.3 Types of Employment Students Intend to Engage In

Table 2.3 Type of Employment Students Intend to Engage In

Type of Employment	Number of Students
Private companies/businesses	44
Federal government	16
State/local government	23
Own business	17
	Sum = 100

Variable → Type of Employment

← Frequency column

Category → Federal government

16 ← Frequency

Frequency Distributions

Definition

A ***frequency distribution for qualitative data*** lists all categories and the number of elements that belong to each of the categories.

Example 2-1

A sample of 30 employees from large companies was selected, and these employees were asked how stressful their jobs were. The responses of these employees are recorded below, where *very* represents very stressful, *somewhat* means somewhat stressful, and *none* stands for not stressful at all.

Example 2-1

somewhat	none	somewhat	very	very	none
very	somewhat	somewhat	very	somewhat	somewhat
very	somewhat	none	very	none	somewhat
somewhat	very	somewhat	somewhat	very	none
somewhat	very	very	somewhat	none	somewhat

Construct a frequency distribution table for these data.

Example 2-1: Solution
Table 2.4 Frequency Distribution of Stress on Job

Stress on Job	Tally	Frequency (f)
Very	𝍸 𝍸	10
Somewhat	𝍸 𝍸 ||||	14
None	𝍸 |	6
		Sum = 30

Relative Frequency and Percentage Distributions

Calculating Relative Frequency of a Category

$$\text{Relative frequency of a category} = \frac{\text{Frequency of that category}}{\text{Sum of all frequencies}}$$

Relative Frequency and Percentage Distributions

Calculating Percentage

Percentage = (Relative frequency) · 100

Example 2-2

Determine the relative frequency and percentage for the data in Table 2.4.

Stress on Job	Tally	Frequency (f)
Very	𝍸 𝍸	10
Somewhat	𝍸 𝍸 ||||	14
None	𝍸 |	6
		Sum = 30

Example 2-2: Solution

Table 2.5 Relative Frequency and Percentage Distributions of Stress on Job

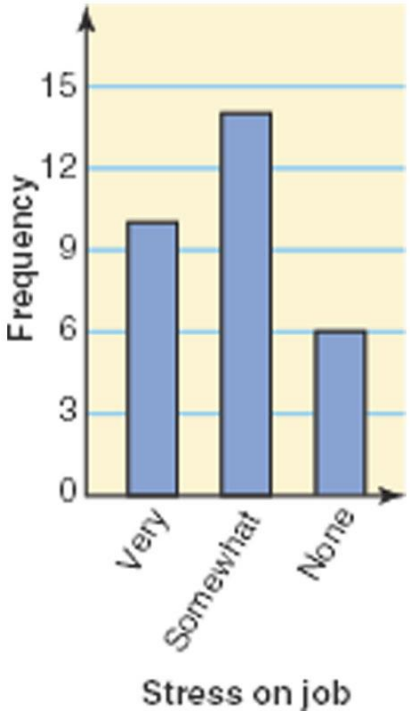
Stress on Job	Relative Frequency	Percentage
Very	10/30 = .333	.333(100) = 33.3
Somewhat	14/30 = .467	.467(100) = 46.7
None	6/30 = .200	.200(100) = 20.0
	Sum = 1.000	Sum = 100

Graphical Presentation of Qualitative Data

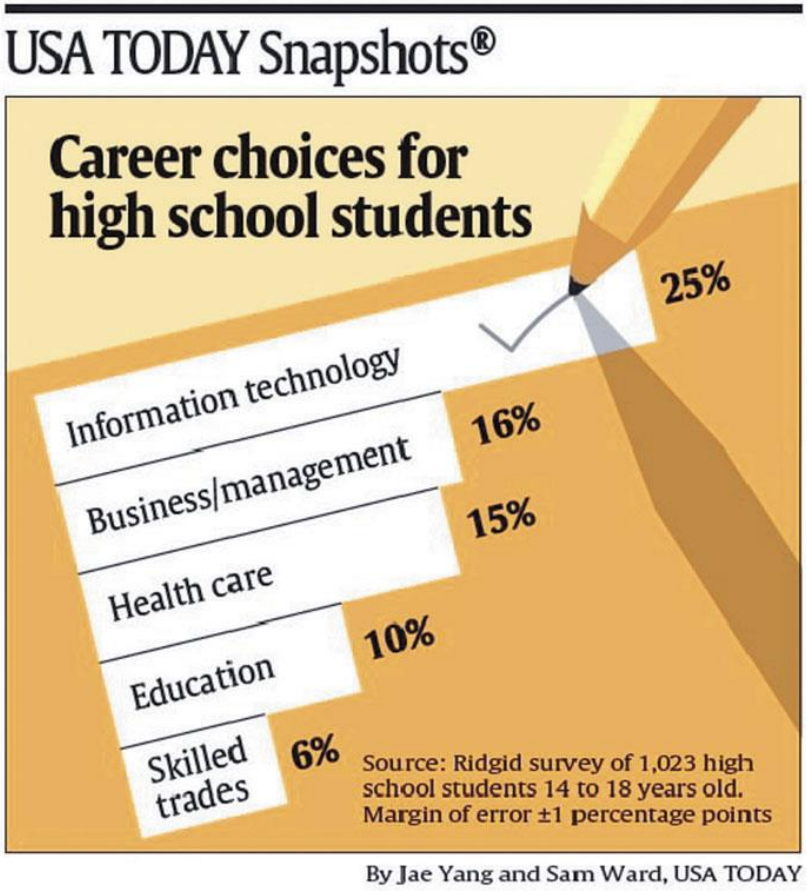
Definition

A graph made of bars whose heights represent the frequencies of respective categories is called a ***bar graph***.

Figure 2.1 Bar graph for the frequency distribution of Table 2.4



Case Study 2-1 Career Choices for High School Students



Graphical Presentation of Qualitative Data

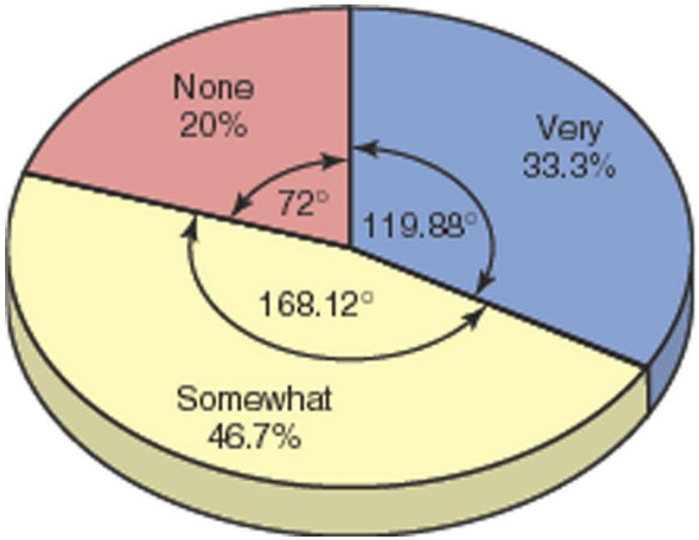
Definition

A circle divided into portions that represent the relative frequencies or percentages of a population or a sample belonging to different categories is called a ***pie chart***.

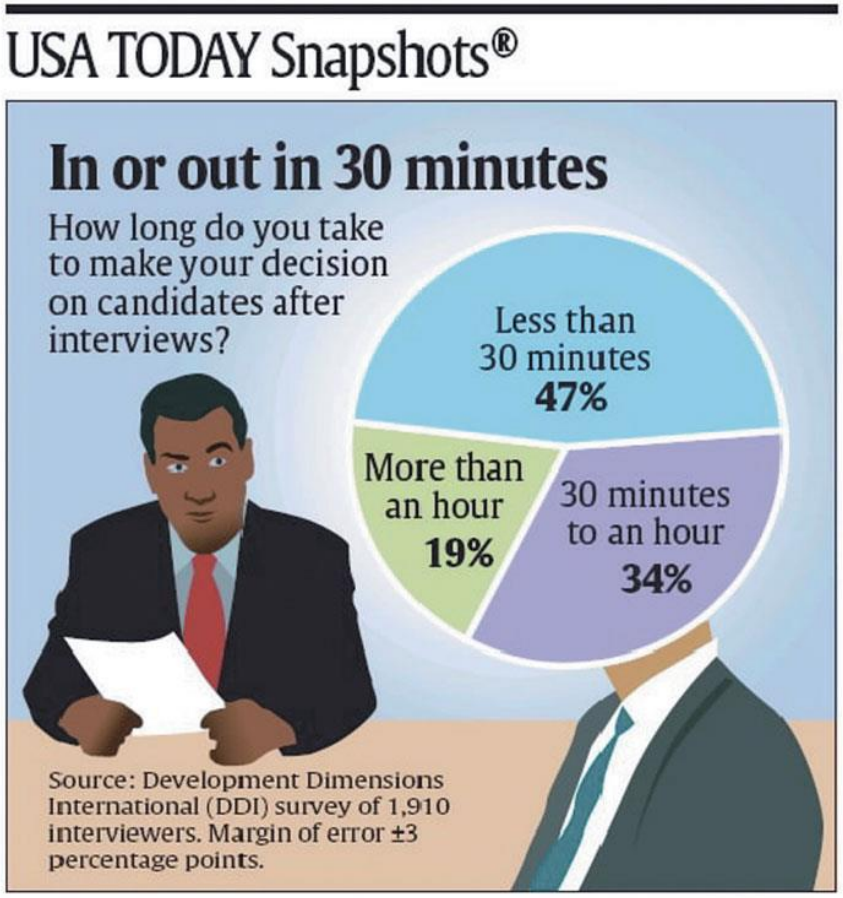
Table 2.6 Calculating Angle Sizes for the Pie Chart

Stress on Job	Relative Frequency	Angle Size
Very	.333	360(.333) = 119.88
Somewhat	.467	360(.467) = 168.12
None	.200	360(.200) = 72.00
	Sum = 1.000	Sum = 360

Figure 2.2 Pie chart for the percentage distribution of Table 2.5.



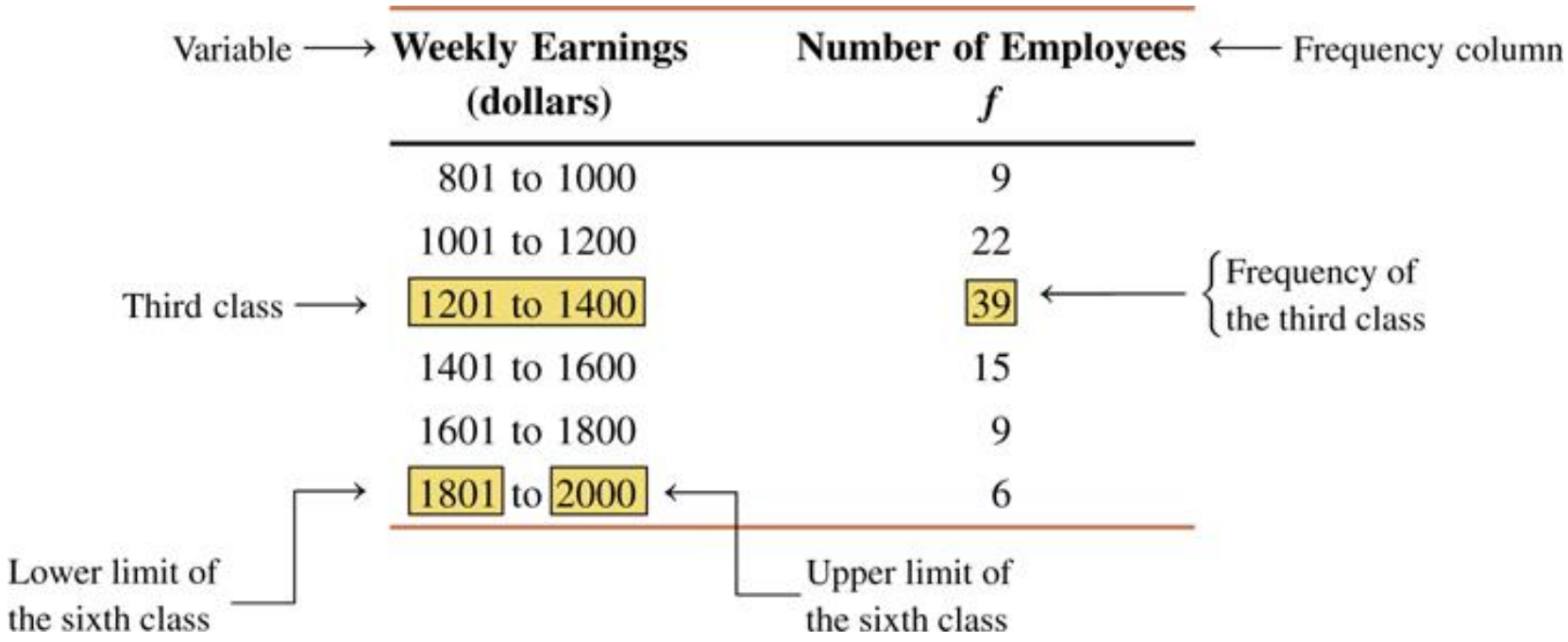
Case Study 2-2 In or Out in 30 Minutes



2.3 Organizing and Graphing Quantitative Data

- Frequency Distributions
- Constructing Frequency Distribution Tables
- Relative and Percentage Distributions
- Graphing Grouped Data

Table 2.7 Weekly Earnings of 100 Employees of a Company



Frequency Distributions

Definition

A ***frequency distribution for quantitative data*** lists all the classes and the number of values that belong to each class. Data presented in the form of a frequency distribution are called ***grouped data***.

Frequency Distributions

Definition

The ***class boundary*** is given by the midpoint of the upper limit of one class and the lower limit of the next class.

Frequency Distributions

Finding Class Width

Class width = Upper boundary – Lower boundary

Frequency Distributions

Calculating Class Midpoint or Mark

$$\text{Class midpoint or mark} = \frac{\text{Lower limit} + \text{Upper limit}}{2}$$

Constructing Frequency Distribution Tables

Calculation of Class Width

$$\text{Approximate class width} = \frac{\text{Largest value} - \text{Smallest value}}{\text{Number of classes}}$$

Table 2.8 Class Boundaries, Class Widths, and Class Midpoints for Table 2.7

Class Limits	Class Boundaries	Class Width	Class Midpoint
801 to 1000	800.5 to less than 1000.5	200	900.5
1001 to 1200	1000.5 to less than 1200.5	200	1100.5
1201 to 1400	1200.5 to less than 1400.5	200	1300.5
1401 to 1600	1400.5 to less than 1600.5	200	1500.5
1601 to 1800	1600.5 to less than 1800.5	200	1700.5
1801 to 2000	1800.5 to less than 2000.5	200	1900.5

Example 2-3

The following data give the total number of iPods® sold by a mail order company on each of 30 days. Construct a frequency distribution table.

8	25	11	15	29	22	10	5	17	21
22	13	26	16	18	12	9	26	20	16
23	14	19	23	20	16	27	16	21	14

Example 2-3: Solution

The minimum value is 5, and the maximum value is 29. Suppose we decide to group these data using five classes of equal width. Then

$$\textbf{Approximate width of each class} = \frac{\mathbf{29 - 5}}{\mathbf{5}} = \mathbf{4.8}$$

Now we round this approximate width to a convenient number, say 5. The lower limit of the first class can be taken as 5 or any number less than 5. Suppose we take 5 as the lower limit of the first class. Then our classes will be

5 – 9, 10 – 14, 15 – 19, 20 – 24, and 25 – 29

Table 2.9 Frequency Distribution for the Data on iPods Sold

iPods Sold	Tally	f
5–9	|||	3
10–14	𝍸 |	6
15–19	𝍸 |||	8
20–24	𝍸 |||	8
25–29	𝍸	5
		$\Sigma f = 30$

Relative Frequency and Percentage Distributions

Calculating Relative Frequency and Percentage

$$\text{Relative frequency of a class} = \frac{\text{Frequency of that class}}{\text{Sum of all frequencies}} = \frac{f}{\sum f}$$

$$\text{Percentage} = (\text{Relative frequency}) \cdot 100$$

Example 2-4

Calculate the relative frequencies and percentages for Table 2.9.

Example 2-4: Solution

Table 2.10 Relative Frequency and Percentage Distributions for Table 2.9

iPods Sold	Class Boundaries	Relative Frequency	Percentage
5–9	4.5 to less than 9.5	3/30 = .100	10.0
10–14	9.5 to less than 14.5	6/30 = .200	20.0
15–19	14.5 to less than 19.5	8/30 = .267	26.7
20–24	19.5 to less than 24.5	8/30 = .267	26.7
25–29	24.5 to less than 29.5	5/30 = .167	16.7
		Sum = 1.001	Sum = 100.1

Graphing Grouped Data

Definition

A ***histogram*** is a graph in which classes are marked on the horizontal axis and the frequencies, relative frequencies, or percentages are marked on the vertical axis. The frequencies, relative frequencies, or percentages are represented by the heights of the bars. In a histogram, the bars are drawn adjacent to each other.

Figure 2.3 Frequency histogram for Table 2.9.

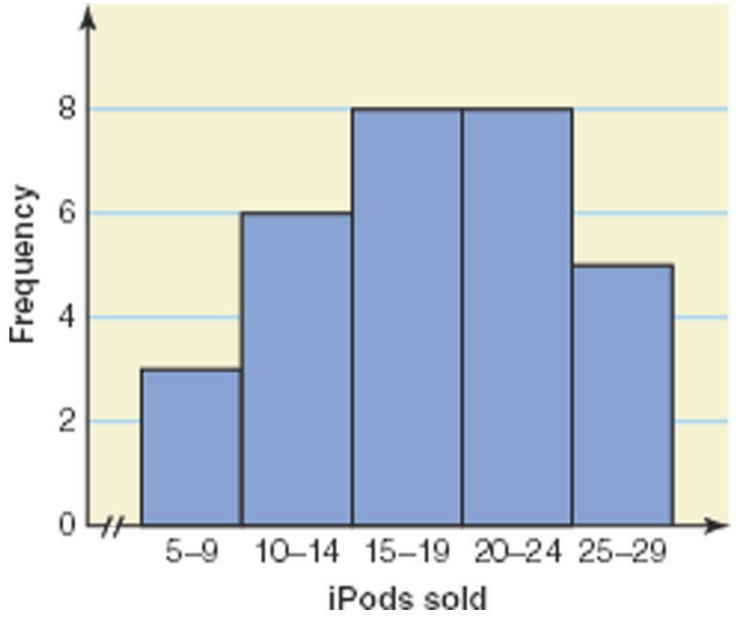
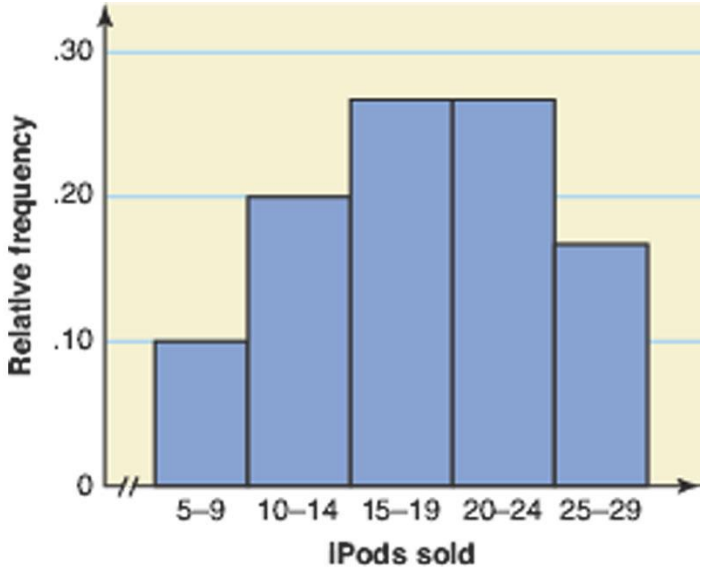
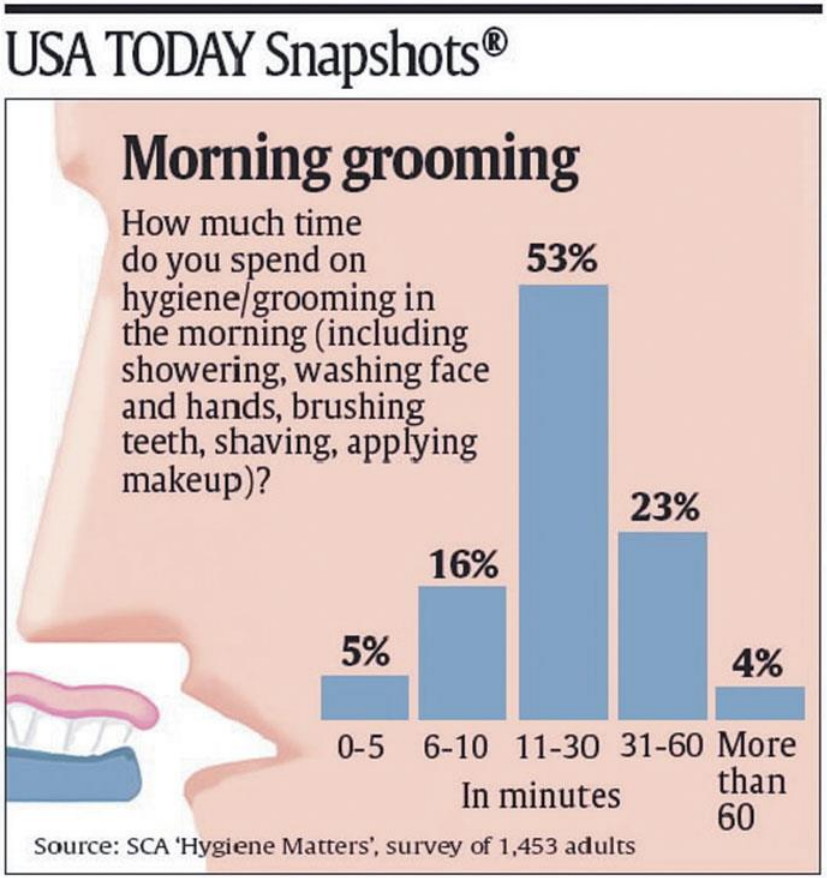


Figure 2.4 Relative frequency histogram for Table 2.10.



Case Study 2-3 Morning Grooming



Graphing Grouped Data

Definition

A graph formed by joining the midpoints of the tops of successive bars in a histogram with straight lines is called a ***polygon***.

Figure 2.5 Frequency polygon for Table 2.9.

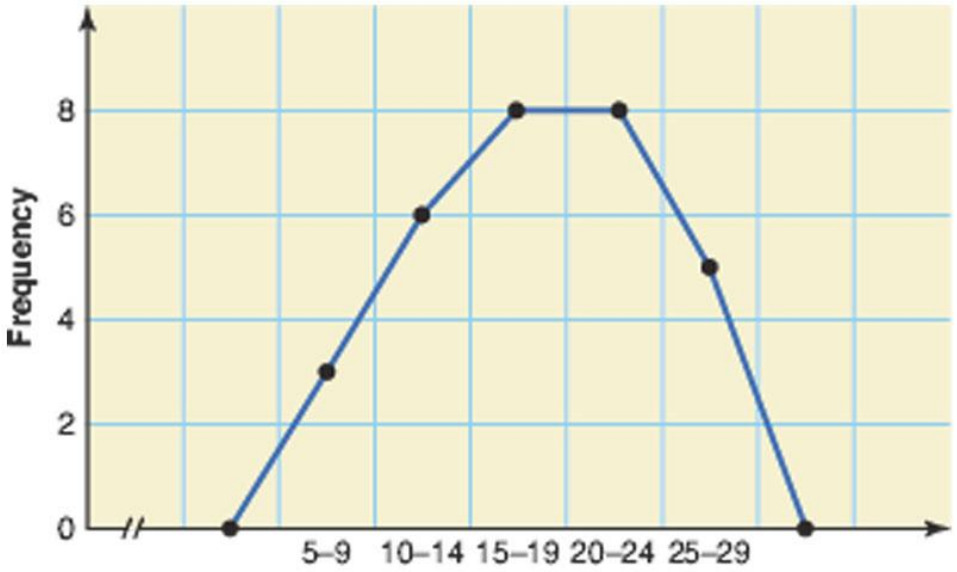
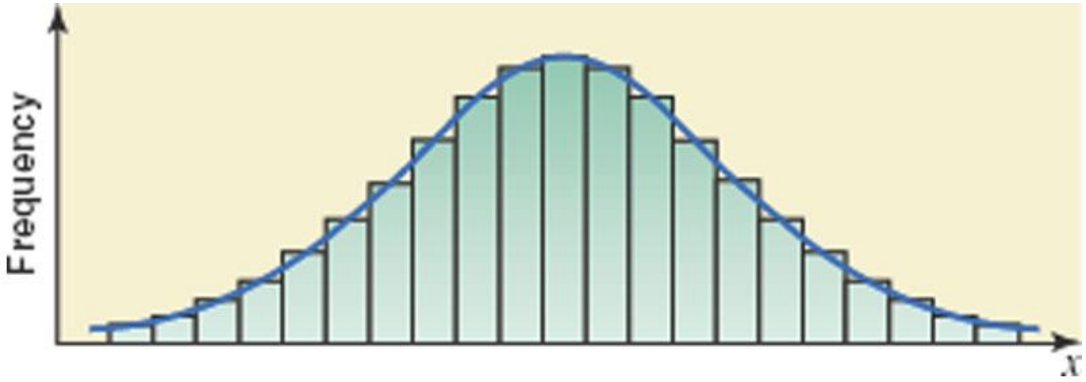


Figure 2.6 Frequency distribution curve.



Example 2-5

On April 1, 2009, the federal tax on a pack of cigarettes was increased from 39¢ to $1.0066, a move that not only was expected to help increase federal revenue, but was also expected to save about 900,000 lives (Time Magazine, April 2009). Table 2.11 shows the total tax (state plus federal) per pack of cigarettes for all 50 states as of April 1, 2009.

Example 2-5

Construct a frequency distribution table. Calculate the relative frequencies and percentages for all classes.

Table 2.11 Total Tax per Pack of Cigarettes

State	Total Tax (in dollars)	State	Total Tax (in dollars)
AL	1.43	MT	2.71
AK	3.01	NE	1.65
AZ	3.01	NV	1.81
AR	2.16	NH	2.34
CA	1.88	NJ	3.58
CO	1.85	NM	1.92
CT	3.01	NY	3.76
DE	2.16	NC	1.36
FL	1.35	ND	1.45
GA	1.38	OH	2.26
HI	3.01	OK	2.04
ID	1.58	OR	2.19
IL	1.99	PA	2.36
IN	2.00	RI	3.47
IA	2.37	SC	1.08
KS	1.80	SD	2.54
KY	1.61	TN	1.63
LA	1.37	TX	2.42
ME	3.01	UT	1.70
MD	3.01	VT	3.00
MA	3.52	VA	1.31
MI	3.01	WA	3.03
MN	2.51	WV	1.56
MS	1.19	WI	2.78
MO	1.18	WY	1.61

Source: Campaign for Tobacco-Free Kids.

Example 2-5: Solution

The minimum value set on cigarette taxes is 1.08, and the maximum value is 3.76. Suppose we decide to group these data using six classes of equal width. Then

$$\textbf{Approximate width of each class} = \frac{\textbf{3.76} - \textbf{1.08}}{\textbf{6}} = \textbf{.45}$$

We round this to a more convenient number, say .50. We can take a lower limit of the first class equal to 1.08 or any number lower than 1.08. If we start the first class at 1, the classes will be written as *1 to less than 1.5*, *1.5 to less than 2.00*, and so on.

Table 2.12 Frequency, Relative Frequency, and Percentage Distributions of the Total Tax on a Pack of Cigarettes

Total Tax (in dollars)	f	Relative Frequency	Percentage
1.00 to less than 1.50	10	.20	20
1.50 to less than 2.00	13	.26	26
2.00 to less than 2.50	10	.20	20
2.50 to less than 3.00	4	.08	8
3.00 to less than 3.50	10	.20	20
3.50 to less than 4.00	3	.06	6
	$\Sigma f = 50$	Sum = 1.00	Sum = 100

Example 2-6

The administration in a large city wanted to know the distribution of vehicles owned by households in that city. A sample of 40 randomly selected households from this city produced the following data on the number of vehicles owned:

5	1	1	2	0	1	1	2	1	1
1	3	3	0	2	5	1	2	3	4
2	1	2	2	1	2	2	1	1	1
4	2	1	1	2	1	1	4	1	3

Construct a frequency distribution table for these data, and draw a bar graph.

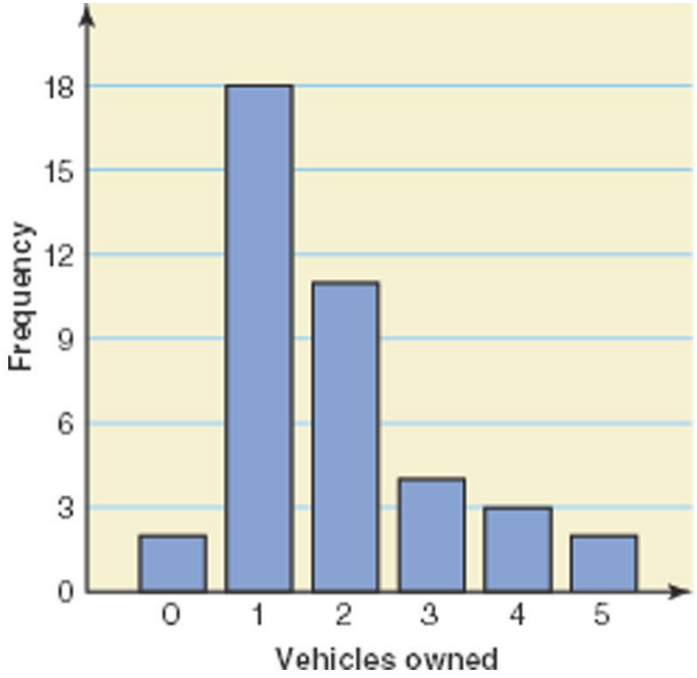
Example 2-6: Solution

Table 2.13 Frequency Distribution of Vehicles Owned

Vehicles Owned	Number of Households (f)
0	2
1	18
2	11
3	4
4	3
5	2
	$\Sigma f = 40$

The observations assume only six distinct values: 0, 1, 2, 3, 4, and 5. Each of these six values is used as a class in the frequency distribution in Table 2.13.

Figure 2.7 Bar graph for Table 2.13.



2.4 Shapes of Histograms

1. Symmetric
2. Skewed
3. Uniform or Rectangular

Figure 2.8 Symmetric histograms.

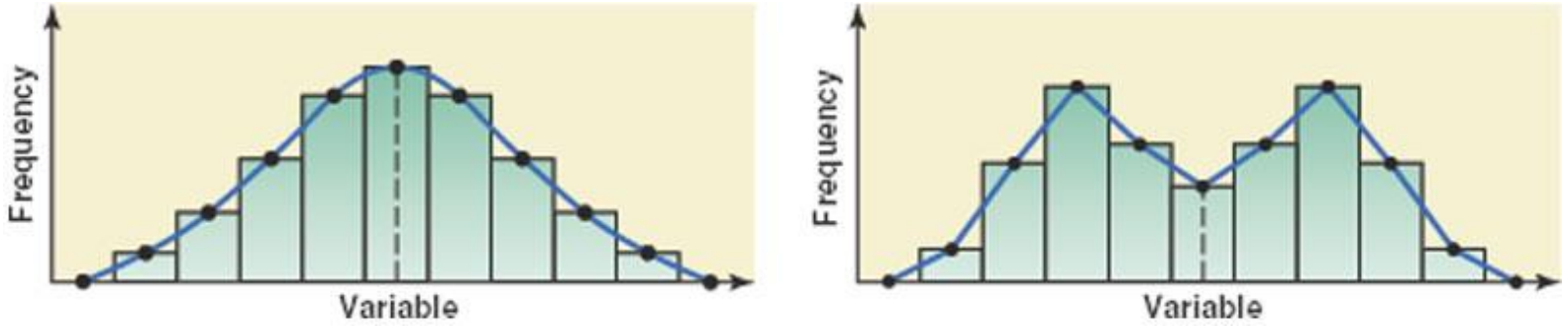


Figure 2.9 (a) A histogram skewed to the right. (b) A histogram skewed to the left.

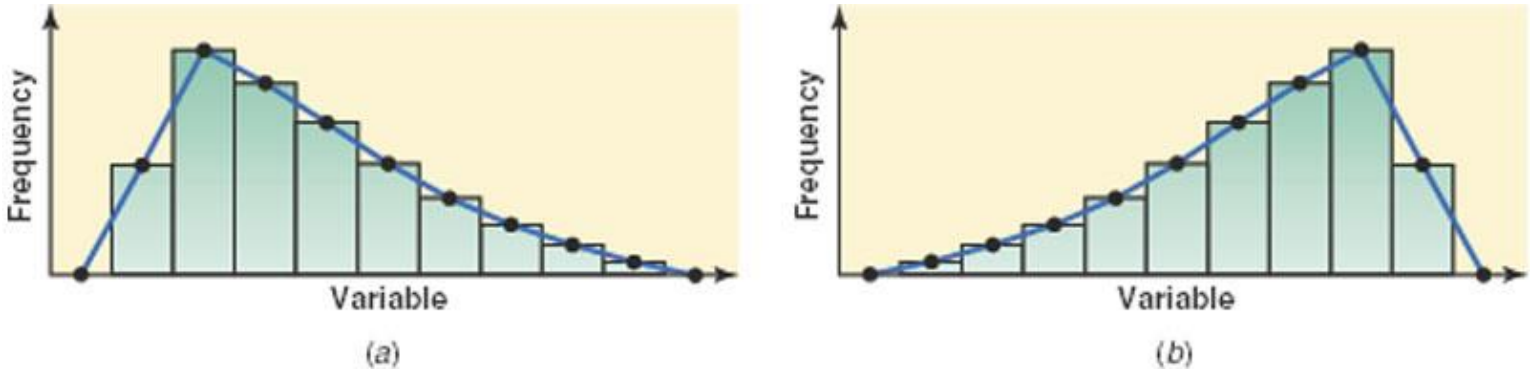


Figure 2.10 A histogram with uniform distribution.

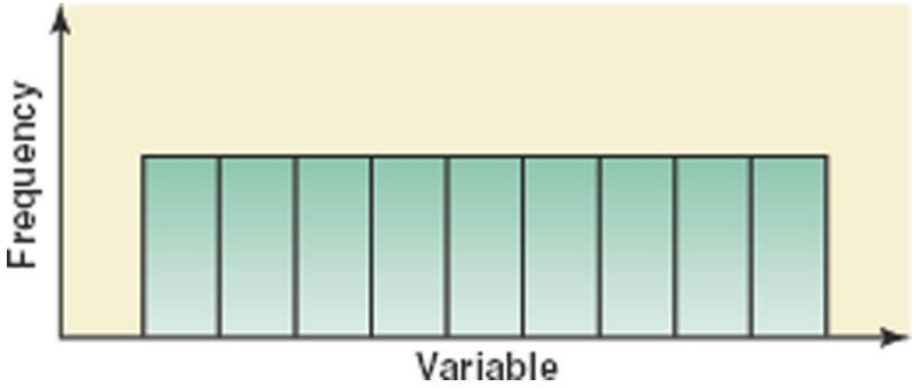
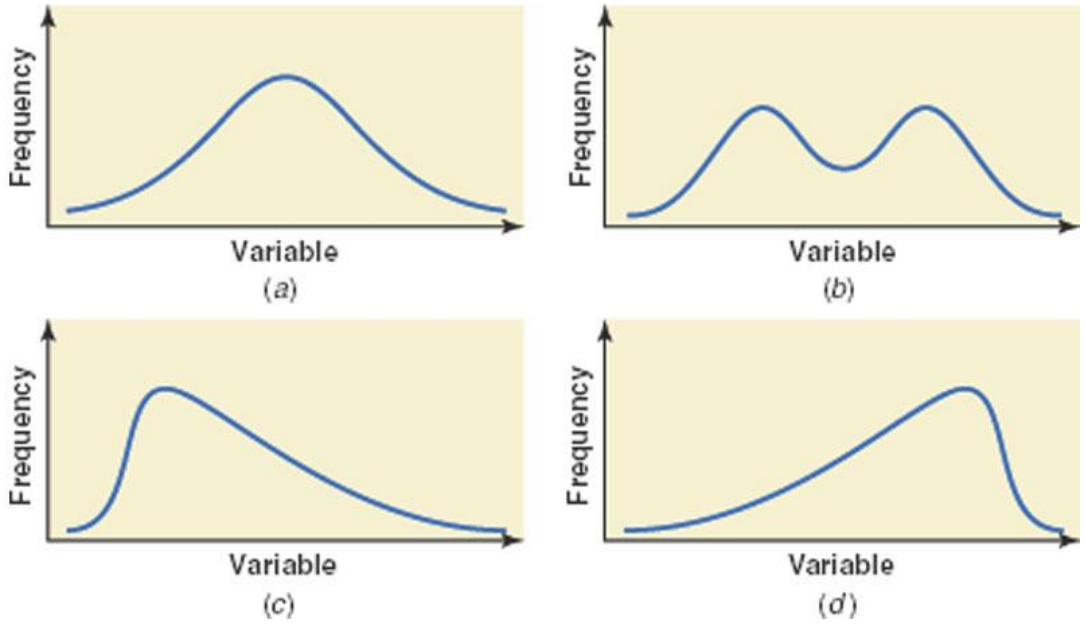


Figure 2.11 (a) and (b) Symmetric frequency curves. (c) Frequency curve skewed to the right. (d) Frequency curve skewed to the left.



2.5 Cumulative Frequency Distributions

Definition

A ***cumulative frequency*** distribution gives the total number of values that fall below the upper boundary of each class.

Example 2-7

Using the frequency distribution of Table 2.9, reproduced here, prepare a cumulative frequency distribution for the number of iPods sold by that company.

iPods Sold	*f*
5–9	3
10–14	6
15–19	8
20–24	8
25–29	5

Example 2-7: Solution

Table 2.14 Cumulative Frequency Distribution of iPods Sold

Class Limits	Class Boundaries	Cumulative Frequency
5–9	4.5 to less than 9.5	3
5–14	4.5 to less than 14.5	3 + 6 = 9
5–19	4.5 to less than 19.5	3 + 6 + 8 = 17
5–24	4.5 to less than 24.5	3 + 6 + 8 + 8 = 25
5–29	4.5 to less than 29.5	3 + 6 + 8 + 8 + 5 = 30

CUMULATIVE FREQUENCY DISTRIBUTIONS

Calculating Cumulative Relative Frequency and Cumulative Percentage

$$\text{Cumulative relative frequency} = \frac{\text{Cumulative frequency of a class}}{\text{Total observations in the data set}}$$

$$\text{Cumulative percentage} = (\text{Cumulative relative frequency}) \cdot 100$$

Table 2.15 Cumulative Relative Frequency and Cumulative Percentage Distributions for iPods Sold

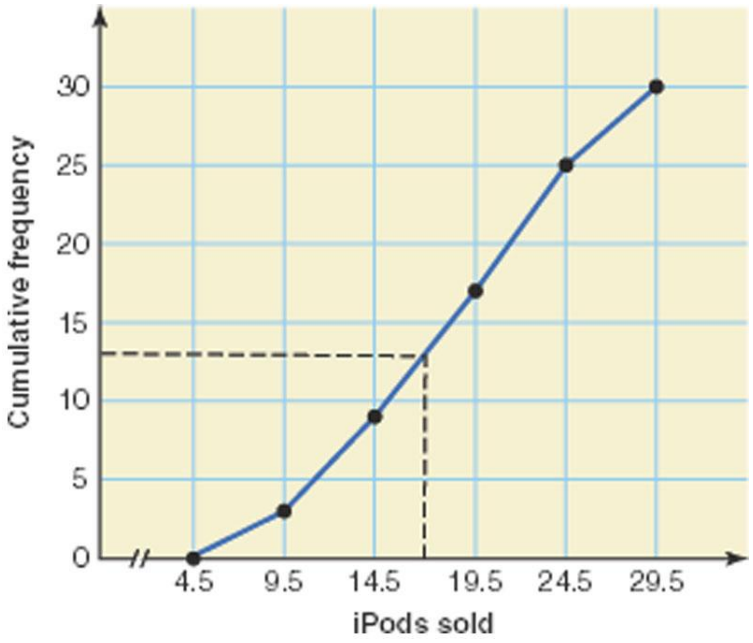
Class Limits	Cumulative Relative Frequency	Cumulative Percentage
5–9	3/30 = .100	10.0
5–14	9/30 = .300	30.0
5–19	17/30 = .567	56.7
5–24	25/30 = .833	83.3
5–29	30/30 = 1.000	100.0

CUMULATIVE FREQUENCY DISTRIBUTIONS

Definition

An ***ogive*** is a curve drawn for the cumulative frequency distribution by joining with straight lines the dots marked above the upper boundaries of classes at heights equal to the cumulative frequencies of respective classes.

Figure 2.12 Ogive for the cumulative frequency distribution of Table 2.14.



2.6 Stem-and-Leaf Displays

Definition

In a ***stem-and-leaf display*** of quantitative data, each value is divided into two portions – a stem and a leaf. The leaves for each stem are shown separately in a display.

Example 2-8

The following are the scores of 30 college students on a statistics test:

75	52	80	96	65	79	71	87	93	95
69	72	81	61	76	86	79	68	50	92
83	84	77	64	71	87	72	92	57	98

Construct a stem-and-leaf display.

Example 2-8: Solution

To construct a stem-and-leaf display for these scores, we split each score into two parts. The first part contains the first digit, which is called the *stem*. The second part contains the second digit, which is called the *leaf*. We observe from the data that the stems for all scores are 5, 6, 7, 8, and 9 because all the scores lie in the range 50 to 98.

Figure 2.13 Stem-and-leaf display.

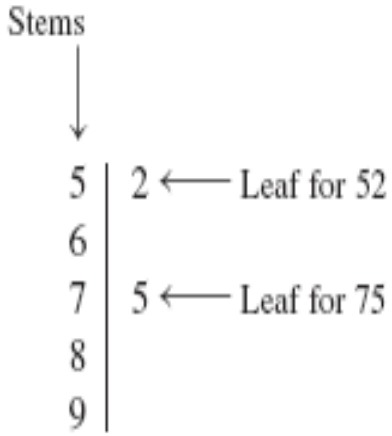


Figure 2.13 Stem-and-leaf display.

Example 2-8: Solution

After we have listed the stems, we read the leaves for all scores and record them next to the corresponding stems on the right side of the vertical line. The complete stem-and-leaf display for scores is shown in Figure 2.14.

Figure 2.14 Stem-and-leaf display of test scores.

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

Figure 2.14 Stem-and-leaf display of test scores.

Example 2-8: Solution

The leaves for each stem of the stem-and-leaf display of Figure 2.14 are *ranked* (in increasing order) and presented in Figure 2.15.

Figure 2.15 Ranked stem-and-leaf display of test scores.

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

Figure 2.15 Ranked stem-and-leaf display of test scores.

Example 2-9

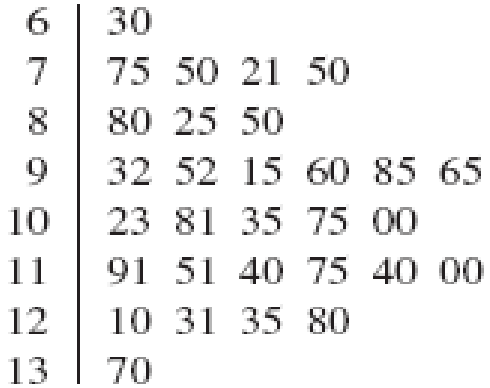
The following data are monthly rents paid by a sample of 30 households selected from a small city.

880	1081	721	1075	1023	775	1235	750	965	960
1210	985	1231	932	850	825	1000	915	1191	1035
1151	630	1175	952	1100	1140	750	1140	1370	1280

Construct a stem-and-leaf display for these data.

Example 2-9: Solution

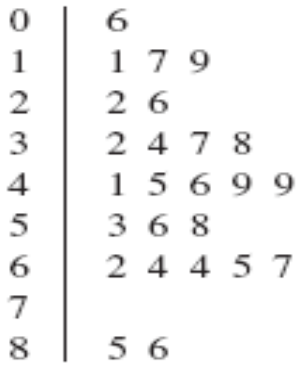
Figure 2.16 Stem-and-leaf display of rents.



6	30
7	75 50 21 50
8	80 25 50
9	32 52 15 60 85 65
10	23 81 35 75 00
11	91 51 40 75 40 00
12	10 31 35 80
13	70

Example 2-10

The following stem-and-leaf display is prepared for the number of hours that 25 students spent working on computers during the last month.



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

Prepare a new stem-and-leaf display by grouping the stems.

Example 2-10: Solution

0–2	6 * 1 7 9 * 2 6
3–5	2 4 7 8 * 1 5 6 9 9 * 3 6 8
6–8	2 4 4 5 7 * * 5 6

Figure 2.17 Grouped stem-and-leaf display.

2.7 Dotplots

Definition

Values that are very small or very large relative to the majority of the values in a data set are called ***outliers or extreme values***.

Example 2-11

Table 2.16 lists the lengths of the longest field goals (in yards) made by all kickers in the American Football Conference (AFC) of the National Football League (NFL) during the 2008 season. Create a dotplot for these data.

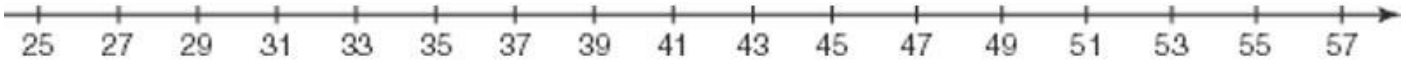
Table 2.16 Distances of Longest Field Goals (in Yards) Made by AFC Kickers During the 2008 NFL Season

Name	Team	Distance	Name	Team	Distance
S. Hauschka	Baltimore	54	C. Barth	Kansas City	45
M. Stover	Baltimore	47	N. Novak	Kansas City	43
R. Lindell	Buffalo	53	D. Carpenter	Miami	50
S. Graham	Cincinnati	45	S. Gostkowski	New England	50
D. Rayner	Cincinnati	26	J. Feely	New York Jets	55
P. Dawson	Cleveland	56	S. Janikowski	Oakland	57
M. Prater	Denver	56	J. Reed	Pittsburgh	53
K. Brown	Houston	53	N. Kaeding	San Diego	57
A. Vinatieri	Indianapolis	52	R. Bironas	Tennessee	51
J. Scobee	Jacksonville	53			

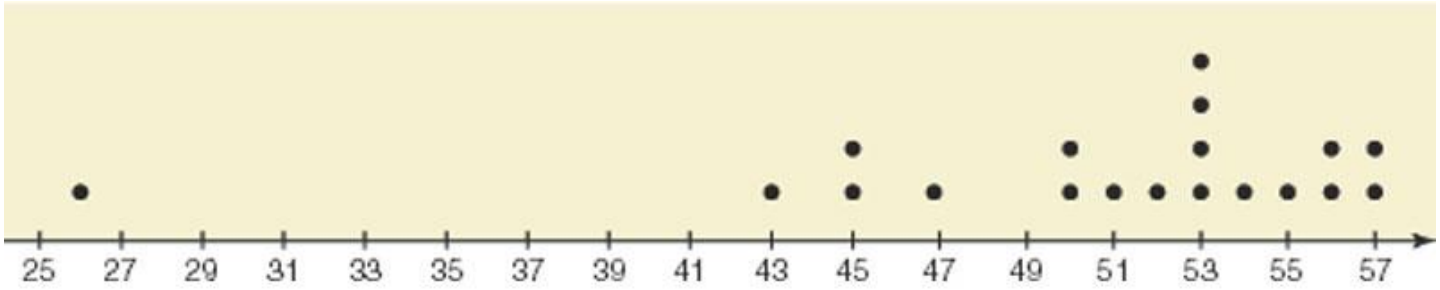
Source: ESPN.com.

Example 2-11: Solution

Step1



Step 2



Example 2-12

Refer to Table 2.16 in Example 2-11, which gives the distances of longest completed field goals for all kickers in the AFC during the 2008 NFL season. Table 2.17 provides the same information for the kickers in the National Football Conference (NFC) of the NFL for the 2008 season. Make dotplots for both sets of data and compare these two dotplots.

Table 2.17 Distances of Longest Field Goals (in Yards) Made by NFC Kickers During the 2008 NFL Season

Name	Team	Distance	Name	Team	Distance
N. Rackers	Arizona	54	T. Mehlhaff	New Orleans	44
J. Elam	Atlanta	50	J. Carney	New York Giants	51
J. Kasay	Carolina	50	L. Tynes	New York Giants	19
R. Gould	Chicago	48	D. Akers	Philadelphia	51
N. Folk	Dallas	52	J. Nedney	San Francisco	53
J. Hanson	Detroit	56	O. Mare	Seattle	51
M. Crosby	Green Bay	53	J. Brown	St. Louis	54
R. Longwell	Minnesota	54	M. Bryant	Tampa Bay	49
M. Gramatica	New Orleans	53	S. Suisham	Washington	50
G. Hartley	New Orleans	47			

Source: ESPN.com.

Example 2-12: Solution

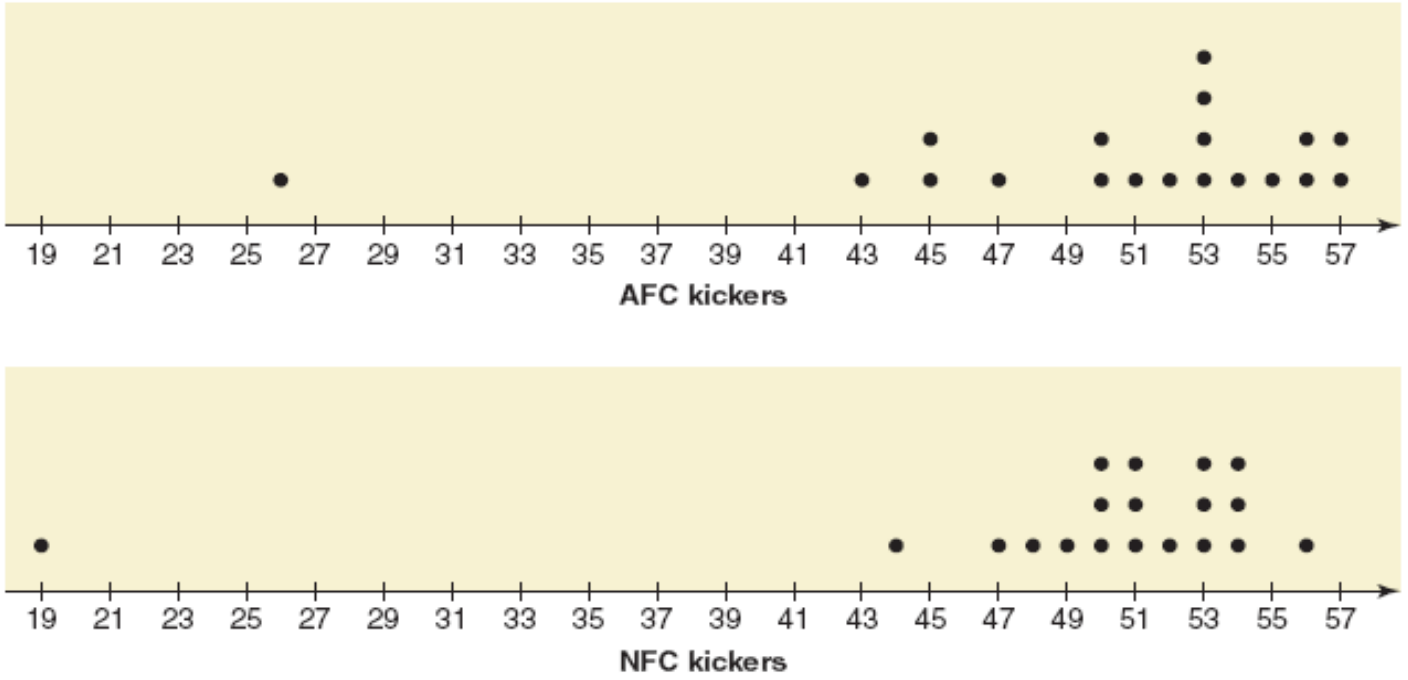


Figure 2.20 Stacked dotplot of longest field goals made by kickers in the AFC and NFC during the 2008 NFL season.

Figure 2.13 Stem-and-leaf display.

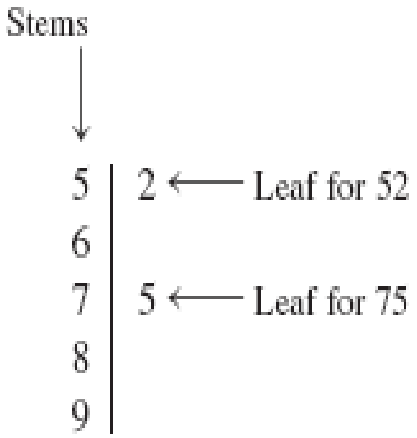


Figure 2.13 Stem-and-leaf display.

TECHNOLOGY INSTRUCTION

Organizing Data

Excel

	A	B	C	D
1	Data	Boundaries	Frequencies	
2				
3	1.1	0	=FREQUENCY(a3:a20,b3:b8)	
4	2.1	5		
5	2.7	10		
6	5.2	15		
7	5.4	20		
8	6.3	25		
9	6.6			
10	8.4			
11	9.3			
12	11			
13	13			
14	13.4			
15	15.1			
16	15.7			
17	17			
18	18.2			
19	19.3			
20	24			