30 Graphical Representations of Data

Visualization techniques are ways of creating and manipulating graphical representations of data. We use these representations in order to gain better insight and understanding of the problem we are studying - pictures can convey an overall message much better than a list of numbers.

In this section we describe some graphical presentations of data.

Line or Dot Plots

Line plots are graphical representations of numerical data. A **line plot** is a number line with x's placed above specific numbers to show their frequency. By the **frequency** of a number we mean the number of occurrence of that number. Line plots are used to represent one group of data with fewer than 50 values.

Example 30.1

Suppose thirty people live in an apartment building. These are the following ages:

$$
\begin{array}{cccccccccc}
58 & 30 & 37 & 36 & 34 & 49 & 35 & 40 & 47 & 47 \\
39 & 54 & 47 & 48 & 54 & 50 & 35 & 40 & 38 & 47 \\
48 & 34 & 40 & 46 & 49 & 47 & 35 & 48 & 47 & 46
\end{array}
$$

Make a line plot of the ages.

Solution.

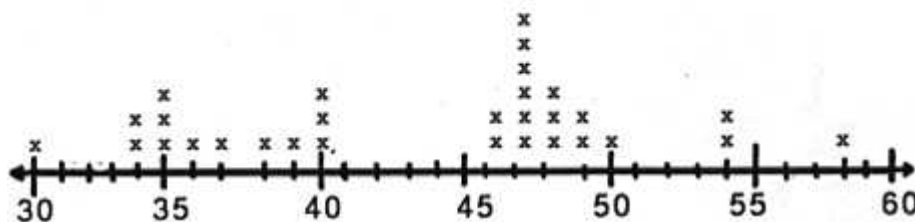The line plot is given in Figure 30.1

Figure 30.1

This graph shows all the ages of the people who live in the apartment building. It shows the youngest person is 30, and the oldest is 58. Most people in

1

the building are over 46 years of age. The most common age is 47.■

Line plots allow several features of the data to become more obvious. For example, outliers, clusters, and gaps are apparent.

- **Outliers** are data points whose values are significantly larger or smaller than other values, such as the ages of 30, and 58.
- **Clusters** are isolated groups of points, such as the ages of 46 through 50.
- **Gaps** are large spaces between points, such as 41 and 45.

Practice Problems

Problem 30.1
Following are the ages of the 30 students at Washington High School who participated in the city track meet. Draw a dot plot to represent these data.

$$
\begin{array}{cccccccccc}
10 & 10 & 11 & 10 & 13 & 8 & 10 & 13 & 14 & 9 \\
14 & 13 & 10 & 14 & 11 & 9 & 13 & 10 & 11 & 12 \\
11 & 12 & 14 & 13 & 12 & 8 & 13 & 14 & 9 & 14
\end{array}
$$

Problem 30.2
The heights (in inches) of the players on a professional basketball team are 70, 72, 75, 77, 78, 78, 80, 81,81,82, and 83. Make a line plot of the heights.

Problem 30.3
Draw a Dot Plot for the following data set.

$$
\begin{array}{ccccccc}
50 & 35 & 70 & 55 & 50 & 30 & 40 \\
65 & 50 & 75 & 60 & 45 & 35 & 75 \\
60 & 55 & 55 & 50 & 40 & 55 & 50
\end{array}
$$

Stem and Leaf Plots
Another type of graph is the **stem-and-leaf plot**. It is closely related to the line plot except that the number line is usually vertical, and digits are used instead of x's. To illustrate the method, consider the following scores which twenty students got in a history test:

$$
\begin{array}{cccccccccc}
69 & 84 & 52 & 93 & 61 & 74 & 79 & 65 & 88 & 63 \\
57 & 64 & 67 & 72 & 74 & 55 & 82 & 61 & 68 & 77
\end{array}
$$

We divide each data value into two parts. The left group is called a **stem** and the remaining group of digits on the right is called a **leaf**. We display horizontal rows of leaves attached to a vertical column of stems. we can construct the following table

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

where the stems are the ten digits of the scores and the leaves are the one digits.

The disadvantage of the stem-and-leaf plots is that data must be grouped according to place value. What if one wants to use different groupings? In this case histograms, to be discussed below, are more suited.

If you are comparing two sets of data, you can use a **back-to-back stem-and-leaf plot** where the leaves of sets are listed on either side of the stem as shown in the table below.

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

where the stems represent the tens digits of a science test scores and the leaves represent the ones digits.

Practice Problems

Problem 30.4
Given below the scores of a class of 26 fourth graders.

```
64  82  85  99  96  81  97  80  81  80  84  87  98
75  86  88  82  78  81  86  80  50  84  88  83  82
```

Make a stem-and-leaf display of the scores.

Problem 30.5

Each morning, a teacher quizzed his class with 20 geography questions. The class marked them together and everyone kept a record of their personal scores. As the year passed, each student tried to improve his or her quiz marks. Every day, Elliot recorded his quiz marks on a stem and leaf plot. This is what his marks looked like plotted out:

$$
\begin{array}{c|ccccccccccc}
0 & 3 & 6 & 5 \\
1 & 0 & 1 & 4 & 3 & 5 & 6 & 5 & 6 & 8 & 9 & 7 & 9 \\
2 & 0 & 0 & 0 & 0
\end{array}
$$

What is his most common score on the geography quizzes? What is his highest score? His lowest score? Are most of Elliot's scores in the 10s, 20s or under 10?

Problem 30.6

A teacher asked 10 of her students how many books they had read in the last 12 months. Their answers were as follows:

$$12, 23, 19, 6, 10, 7, 15, 25, 21, 12$$

Prepare a stem and leaf plot for these data.

Problem 30.7

Make a back-to-back stem and leaf plot for the following test scores:

Class 1:

100	96	93	92	92	92	90	90
89	89	85	82	79	75	74	73
73	73	70	69	68	68	65	61
35							

Class 2:

79	85	56	79	84	64	44	57
69	85	65	81	73	51	61	67
71	89	69	77	82	75	89	92
74	70	75	88	46			

Frequency Distributions and Histograms

When we deal with large sets of data, a good overall picture and sufficient information can be often conveyed by distributing the data into a number of

4

classes or **class intervals** and to determine the number of elements belonging to each class, called **class frequency.** For instance, the following table shows some test scores from a math class.

65	91	85	76	85	87	79	93
82	75	100	70	88	78	83	59
87	69	89	54	74	89	83	80
94	67	77	92	82	70	94	84
96	98	46	70	90	96	88	72

It's hard to get a feel for this data in this format because it is unorganized. To construct a frequency distribution,

- Compute the class width $CW = \frac{Largest\ data\ value - smallest\ data\ value}{Desirable\ number\ of\ classes}$.

- Round CW to the next highest whole number so that the classes cover the whole data.

Thus, if we want to have 6 class intervals then $CW = \frac{100-46}{6} = 9$. The low number in each class is called the **lower class limit**, and the high number is called the **upper class limit.**

With the above information we can construct the following table called **frequency distribution**.

Class	Frequency
41-50	1
51-60	2
61-70	6
71-80	8
81-90	14
91-100	9

Once frequency distributions are constructed, it is usually advisable to present them graphically. The most common form of graphical representation is the **histogram**.

In a histogram, each of the classes in the frequency distribution is represented by a vertical bar whose height is the class frequency of the interval. The horizontal endpoints of each vertical bar correspond to the class endpoints.

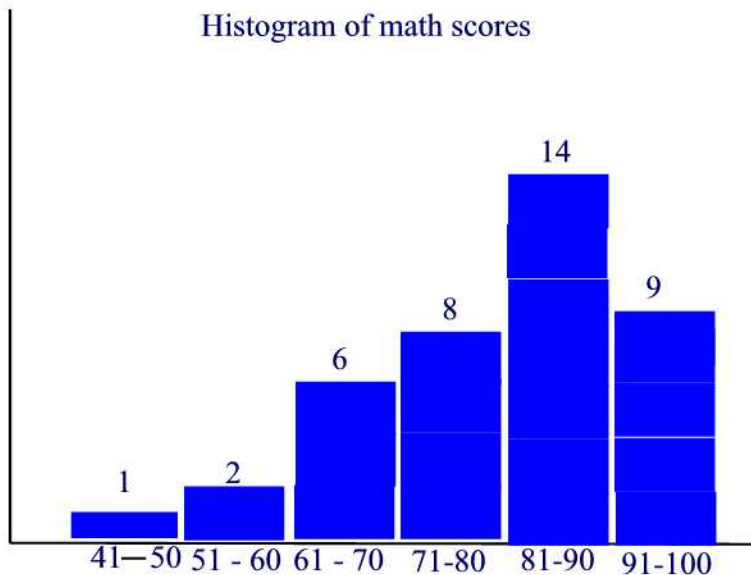A histogram of the math scores is given in Figure 30.2

Figure 30.2

One advantage to the stem-and-leaf plot over the histogram is that the stem-and-leaf plot displays not only the frequency for each interval, but also displays all of the individual values within that interval.

Practice Problems

Problem 30.8
Suppose a sample of 38 female university students was asked their weights in pounds. This was actually done, with the following results:

130	108	135	120	97	110
130	112	123	117	170	124
120	133	87	130	160	128
110	135	115	127	102	130
89	135	87	135	115	110
105	130	115	100	125	120
120	120				

(a) Suppose we want 9 class intervals. Find CW.
(b) Construct a frequency distribution.
(c) Construct the corresponding histogram.

6

Problem 30.9
The table below shows the response times of calls for police service measured in minutes.

34	10	4	3	9	18	4
3	14	8	15	19	24	9
36	5	7	13	17	22	27
3	6	11	16	21	26	31
32	38	40	30	47	53	14
6	12	18	23	28	33	
3	4	62	24	35	54	
15	6	13	19	3	4	
4	20	5	4	5	5	
10	25	7	7	42	44	

Construct a frequency distribution and the corresponding histogram.

Problem 30.10
A nutritionist is interested in knowing the percent of calories from fat which Americans intake on a daily basis. To study this, the nutritionist randomly selects 25 Americans and evaluates the percent of calories from fat consumed in a typical day. The results of the study are as follows

34%	18%	33%	25%	30%
42%	40%	33%	39%	40%
45%	35%	45%	25%	27%
23%	32%	33%	47%	23%
27%	32%	30%	28%	36%

Construct a frequency distribution and the corresponding histogram.

Bar Graphs
Bar Graphs, similar to histograms, are often useful in conveying information about categorical data where the horizontal scale represents some nonnumerical attribute. In a bar graph, the bars are nonoverlapping rectangles of equal width and they are equally spaced. The bars can be vertical or horizontal. The length of a bar represents the quantity we wish to compare.

Example 30.2
The areas of the various continents of the world (in millions of square miles)

are as follows:11.7 for Africa; 10.4 for Asia; 1.9 for Europe; 9.4 for North America; 3.3 Oceania; 6.9 South America; 7.9 Soviet Union. Draw a bar chart representing the above data and where the bars are horizontal.

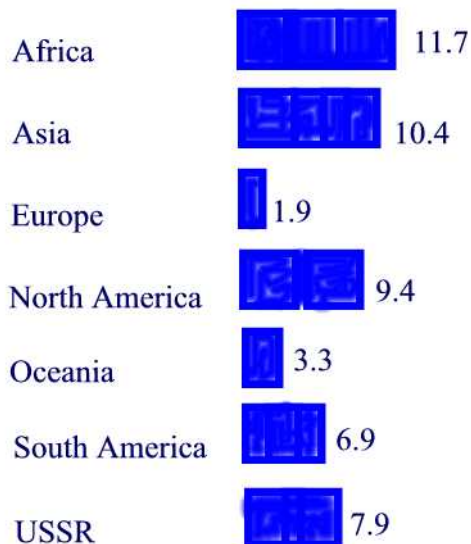**Solution.**
The bar graph is shown in Figure 30.3.

Figure 30.3: Areas (in millions of square miles) of the
various continents of the world

A double bar graph is similar to a regular bar graph, but gives 2 pieces of information for each item on the vertical axis, rather than just 1. The bar chart in Figure 30.4 shows the weight in kilograms of some fruit sold on two different days by a local market. This lets us compare the sales of each fruit over a 2 day period, not just the sales of one fruit compared to another. We can see that the sales of star fruit and apples stayed most nearly the same. The sales of oranges increased from day 1 to day 2 by 10 kilograms. The same amount of apples and oranges was sold on the second day.
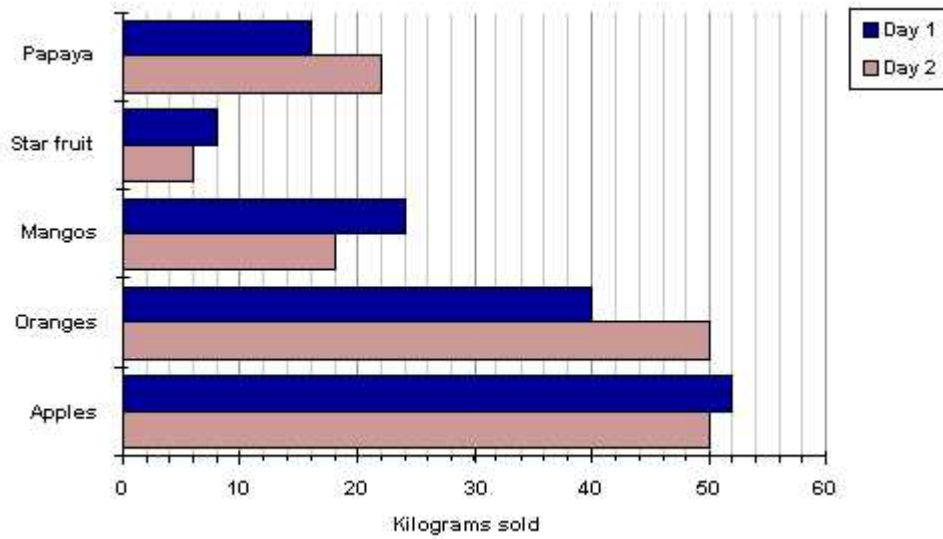
Figure 30.4

Example 30.3

The figures for total population, male and female population of the UK at decade intervals since 1959 are given below:

Year	Total UK Resident Population	Male	Female
1959	51,956,000	25,043,000	26,913,000
1969	55,461,000	26,908,000	28,553,000
1979	56,240,000	27,373,000	28,867,000
1989	57,365,000	27,988,000	29,377,000
1999	59,501,000	29,299,000	30,202,000

Construct a bar chart representing the data.

Solution.

The bar graph is shown in Figure 30.5.
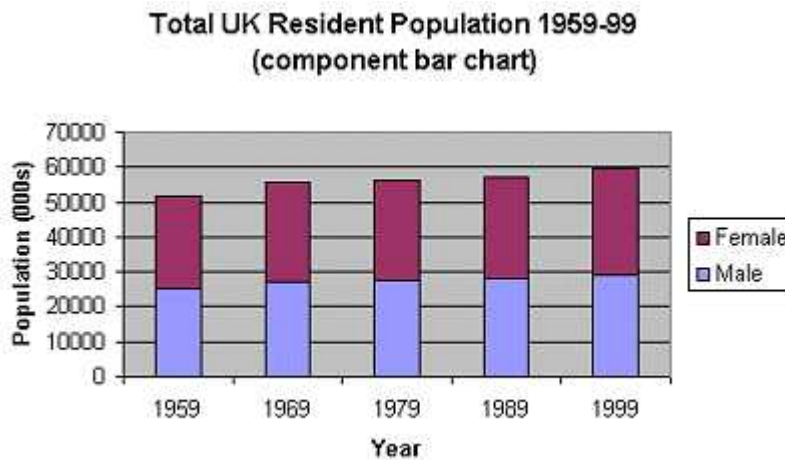
Figure 30.5

Line Graphs

A **Line graph** (or **time series plot**)is particularly appropriate for representing data that vary continuously. A line graph typically shows the trend of a variable over time. To construct a time series plot, we put time on the horizontal scale and the variable being measured on the vertical scale and then we connect the points using line segments.

Example 30.4

The population (in millions) of the US for the years 1860-1950 is as follows: 31.4 in 1860; 39.8 in 1870; 50.2 in 1880; 62.9 in 1890; 76.0 in 1900; 92.0 in 1910; 105.7 in 1920; 122.8 in 1930; 131.7 in 1940; and 151.1 in 1950. Make a time plot showing this information.

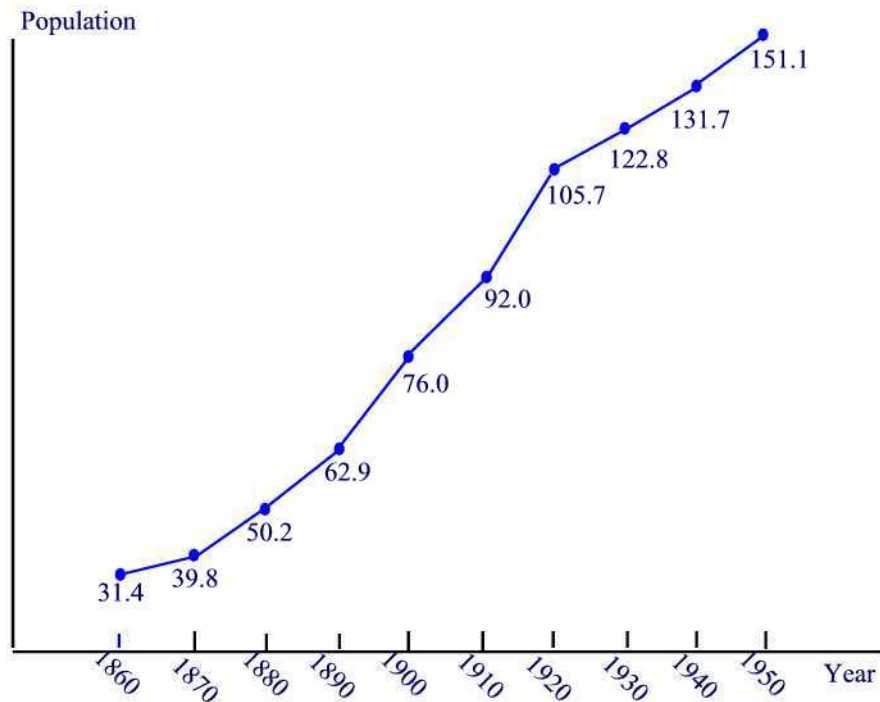**Solution.**

The line graph is shown in Figure 30.6.

Figure 30.6: US population (in millions) from 1860 - 1950

Practice Problems

Problem 30.11
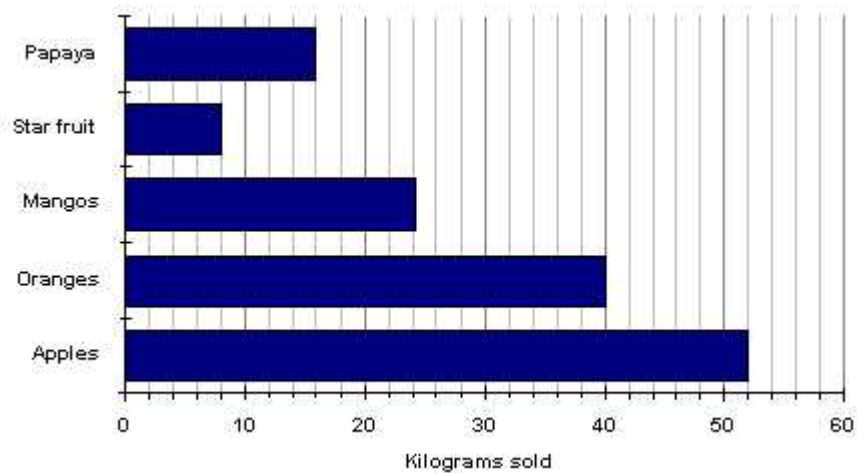Given are several gasoline vehicles and their fuel consumption averages.

Buick	27 mpg
BMW	28 mpg
Honda Civic	35 mpg
Geo	46 mpg
Neon	38 mpg
Land Rover	16 mpg

(a) Draw a bar graph to represent these data.
(b) Which model gets the least miles per gallon? the most?

Problem 30.12
The bar chart below shows the weight in kilograms of some fruit sold one

day by a local market.

How many kg of apples were sold? How many kg of oranges were sold?

Problem 30.13
The figures for total population at decade intervals since 1959 are given below:

$Year$	$Total\ UK\ Resident\ Population$
1959	$51,956,000$
1969	$55,461,000$
1979	$56,240,000$
1989	$57,365,000$
1999	$59,501,000$

Construct a bar chart for this data.

Problem 30.14
The following data gives the number of murder victims in the U.S in 1978 classified by the type of weapon used on them: Gun, $11,910$; cutting/stabbing, $3,526$; blunt object, 896; strangulation/beating, $1,422$; arson, 255; all others 705. Construct a bar chart for this data. Use vertical bars.

Circle Graphs or Pie Charts

Another type of graph used to represent data is the circle graph. A **circle graph** or **pie chart**, consists of a circular region partitioned into disjoint sections, with each section representing a part or percentage of a whole. To construct a pie chart we first convert the distribution into a percentage distribution. Then, since a complete circle corresponds to 360 degrees, we obtain the central angles of the various sectors by multiplying the percentages by 3.6. We illustrate this method in the next example.

Example 30.5

A survey of 1000 adults uncovered some interesting housekeeping secrets. When unexpected company comes, where do we hide the mess? The survey showed that 68% of the respondents toss their mess in the closet, 23% shove things under the bed, 6% put things in the bath tub, and 3% put the mess in the freezer. Make a circle graph to display this information.

Solution.

We first find the central angle corresponding to each case:

$$
\begin{array}{rrcl}
in\ closet & 68 \times 3.6 & = & 244.8 \\
under\ bed & 23 \times 3.6 & = & 82.8 \\
in\ bathtub & 6 \times 3.6 & = & 21.6 \\
in\ freezer & 3 \times 3.6 & = & 10.8
\end{array}
$$

Note that

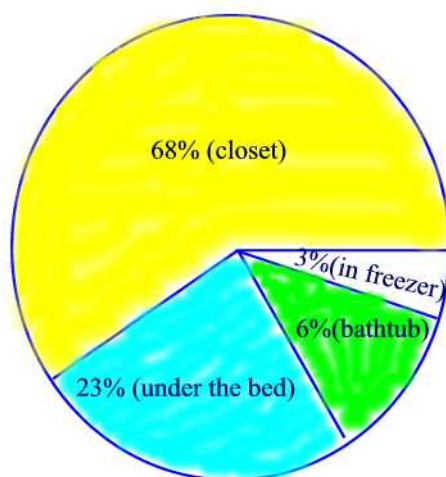$$244.8 + 82.8 + 21.6 + 10.8 = 360.$$

The pie chart is given in Figure 30.7.

Figure 30.7: Where to hide the mess?

13

Practice Problems

Problem 30.15
The table below shows the ingredients used to make a sausage and mushroom pizza.

Ingredient	%
Sausage	7.5
Cheese	25
Crust	50
Tomato Sause	12.5
Mushroom	5

Plot a pie chart for the data.

Problem 30.16
A newly qualified teacher was given the following information about the ethnic origins of the pupils in a class.

Ethnic origin	No. of pupils
White	12
Indian	7
Black African	2
Pakistani	3
Bangladeshi	6
TOTAL	30

Plot a pie chart representing the data.

Problem 30.17
The following table represents a survey of people's favorite ice cream flavor

Flovor	Number of people
Vanilla	21.0%
Chocolate	33.0%
Strawberry	12.0%
Raspberry	4.0%
Peach	7.0%
Neopolitan	17.0%
Other	6.0%

Plot a pie chart to representing the data.

Problem 30.18
In the United States, approximately 45% of the population has blood type O; 40% type A; 11% type B; and 4% type AB. Illustrate this distribution of blood types with a pie chart.

Pictographs
One type of graph seen in newspapers and magazines is a **pictograph**. In a pictograph, a symbol or icon is used to represent a quantity of items. A pictograph needs a title to describe what is being presented and how the data are classified as well as the time period and the source of the data. Example of a pictograph is given in Figure 30.8.
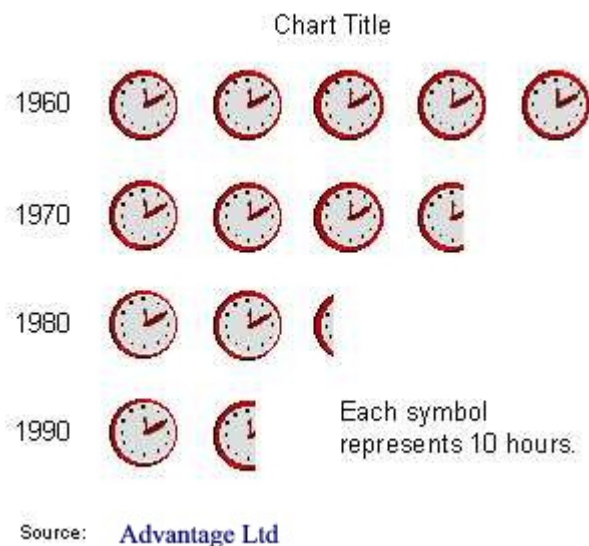

Figure 30.8

A disadvantage of a pictograph is that it is hard to quantify partial icons.

Practice Problems

Problem 30.19

Make a pictograph to represent the data in the following table. Use to represent 10 glasses of lemonade.

Day	Frequency
Monday	15
Tuesday	20
Wednesday	30
Thursday	5
Friday	10

Problem 30.20
The following pictograph shows the approximate number of people who speak the six common languages on earth.
(a) About how many people speak Spanish?
(b) About how many people speak English?
(c) About how many more people speak Mandarin than Arabic?
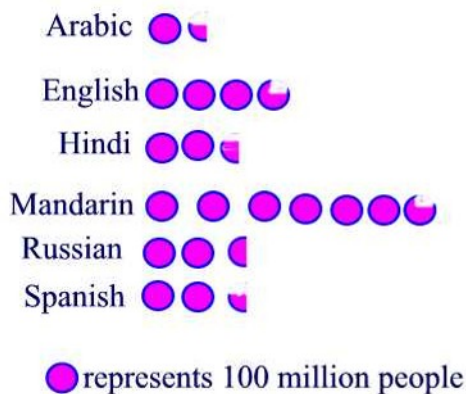
Problem 30.21
Ten people were surveyed about their favorite pets and the result is shown in the table below.

Pet	Frequency
Dog	2
Cat	5
Hamster	3

Make a pictograph for the following table of data. Let 🐾 stand for 2 votes.

Scatterplots
A relationship between two sets of data is sometimes determined by using

a **scatterplot**. Let's consider the question of whether studying longer for a test will lead to better scores. A collection of data is given below

Study Hours	3	5	2	6	7	1	2	7	1	7
Score	80	90	75	80	90	50	65	85	40	100

Based on these data, a scatterplot has been prepared and is given in Figure 30.9.(Remember when making a scatterplot, do NOT connect the dots.)
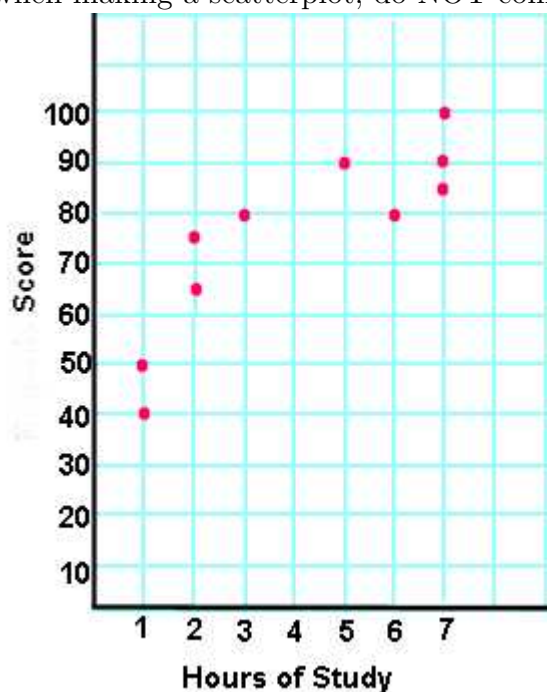

Figure 30.9

The data displayed on the graph resembles a line rising from left to right. Since the slope of the line is positive, there is a **positive correlation** between the two sets of data. This means that according to this set of data, the longer I study, the better grade I will get on my exam score.

If the slope of the line had been negative (falling from left to right), a **negative correlation** would exist. Under a negative correlation, the longer I study, the worse grade I would get on my exam.

If the plot on the graph is scattered in such a way that it does not approximate a line (it does not appear to rise or fall), there is **no correlation** between the sets of data. No correlation means that the data just doesn't

show if studying longer has any affect on my exam score.

Practice Problems

Problem 30.22
Coach Lewis kept track of the basketball team's jumping records for a 10-year period, as follows:

Year	'93	'94	'95	'96	'97	'98	'99	'00	'01	'02
Record(nearest in)	65	67	67	68	70	74	77	78	80	81

(a) Draw a scatterplot for the data.
(b) What kind of correlation is there for these data?

Problem 30.23
The gas tank of a National Motors Titan holds 20 gallons of gas. The following data are collected during a week.

Fuel in tank (gal.)	20	18	16	14	12	10
Dist. traveled (mi)	0	75	157	229	306	379

(a) Draw a scatterplot for the data.
(b) What kind of correlation is there for these data?