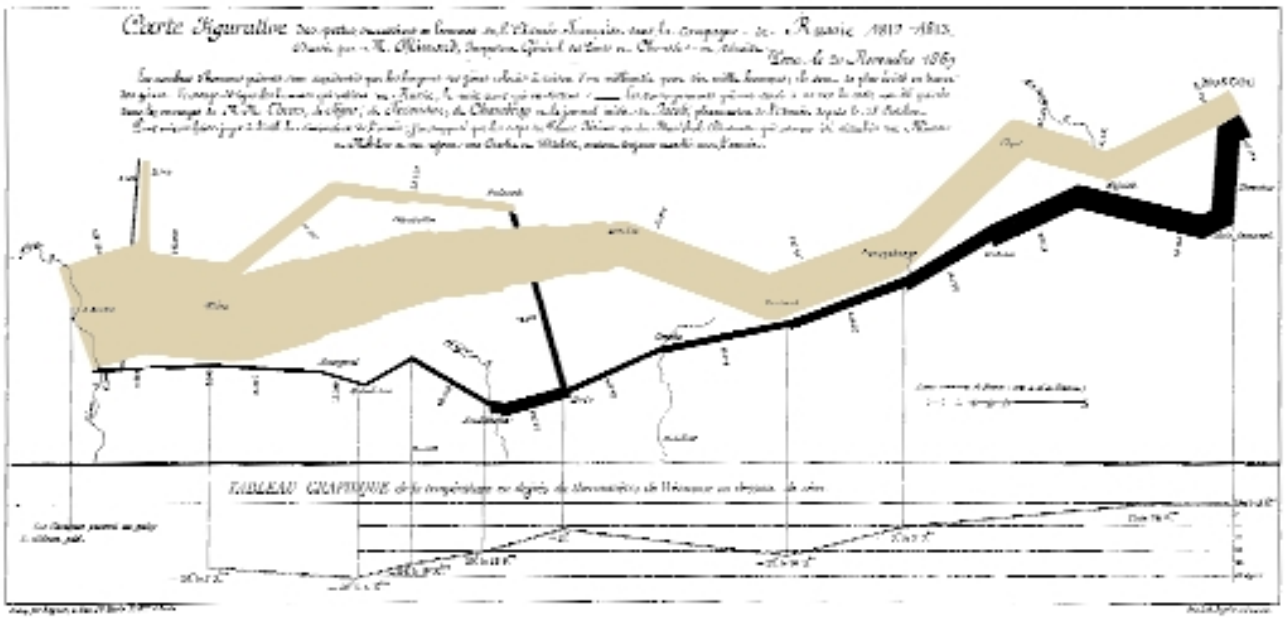
3.1 Graphical Displays of Information

With all the data available today, it is important to be able to filter out the unnecessary data and to present the useful data in an accessible format. One of the best formats for presenting data is a graph or other visual display.

In this section, you will discover some important considerations when choosing, designing, and interpreting a graphic image representing quantitative data, such as Charles Minard's classic depiction of Napoleon's march to Moscow in 1812 and subsequent return (shown below). The thickness of the band represents the size of his army.



DATA TABLES

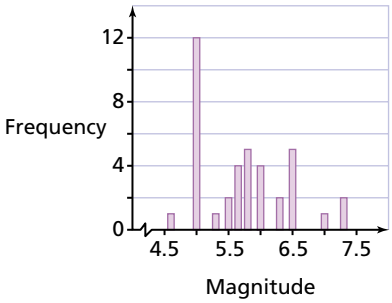
Richter Scale—a logarithmic scale used to measure the magnitude of earthquakes

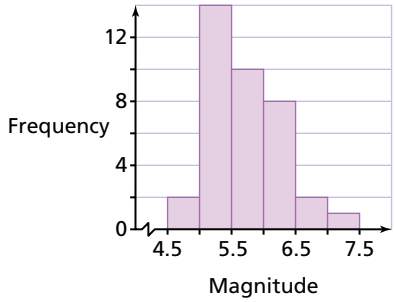
Earthquakes are measured on a scale known as the **Richter Scale**. There data are a sample of earthquake magnitudes in Canada between 1960 and 1965.

5.0	5.0	6.4	5.0	6.0	5.6	6.5	6.5	5.0	5.5
6.4	7.2	5.0	5.7	5.6	5.0	5.0	5.0	5.0	5.7
5.0	7.0	5.5	5.2	4.6	6.3	7.2	6.0	5.4	5.8
6.0	5.7	6.5	5.0	5.7	5.0	5.6	6.0	5.6	6.2

Source: United States Geological Survey, National Earthquake Information Center

The table form makes it difficult to do any analysis beyond determining the mildest and most severe earthquake. To make more sense of the data, you need to sort the magnitudes into groups or classes and count the number of earthquakes that fall into each class. A histogram, unlike a bar graph, often contains continuous data like this, grouped in class (frequency) intervals, displaying how they are spread over a specified range.





bin width—the width of each interval in a histogram

The width of each bar is known as the **bin width**. Different bin widths will produce different results. A small bin width may result in a histogram that does not effectively summarize the distribution (too many small bars). Bin widths should be set so that they are equal, and there should be at least five intervals in the data set so that a representative display is achieved. If a bin width has a frequency of zero, no bar will appear and it no longer looks like a histogram.

Example 1 Creating Effective Histograms

Create an effective histogram of the earthquake data.

Solution 1 *No technology required*

When creating a histogram by hand, determine a bin width that will provide a display that accurately summarizes the distribution of the data.

The range that must be covered is $7.2 - 4.6 = 2.6$, which could be rounded up to 3. Since 6 divides the range nicely, use a bin width of 0.5, as follows:

$$\text{Bin width} = \frac{\text{range}}{\text{number of intervals}}$$
$$= \frac{3.0}{6} \text{ or } 0.5$$

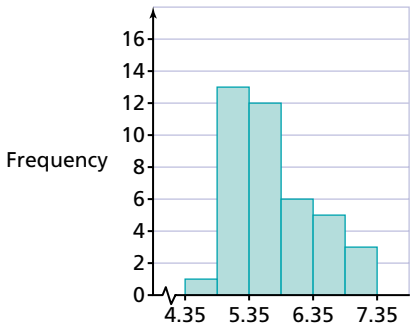
Think about
Bin Widths
Make sure that you do not use intervals that will allow one piece of data to be in two intervals. Why is this?

To ensure no piece of data lies on the borderline between two intervals, make the intervals 4.35–4.85, 4.85–5.35, 5.35–5.85, 5.85–6.35, 6.35–6.85, and 6.85–7.35. Then, create a frequency distribution in which you record the number of pieces of data that falls into each class interval. Draw the resulting histogram.

Frequency Distribution

Class	Tally	Frequency
4.35–4.85	|	1
4.85–5.35	𝍸 𝍸 |||	13
5.35–5.85	𝍸 𝍸 ||	12
5.85–6.35	𝍸 |	6
6.35–6.85	𝍸	5
6.85–7.35	|||	3

Magnitude of Canadian Earthquakes, 1960–1965

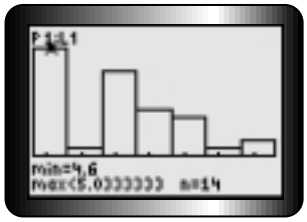


Solution 2 *Using a TI-83 Plus calculator*

Enter the data into $\mathbf{L_1}$ and press 2nd Y= ENTER to set up **Plot 1** as shown. Be sure to select the histogram icon.

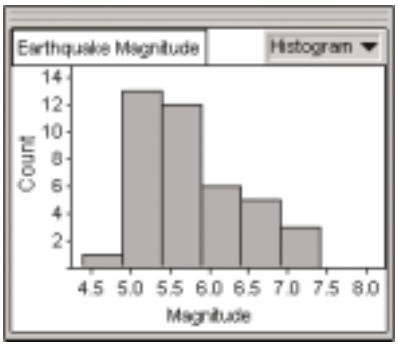
Technolink
See Appendix C.6 on page 405 for more information on creating histograms with a TI-83 Plus calculator.

Press ZOOM and choose **9:ZoomStat**. The calculator determines an appropriate bin width and organizes the data into these class intervals. The TRACE button will allow you to check the intervals and frequencies from the data using the arrow keys.



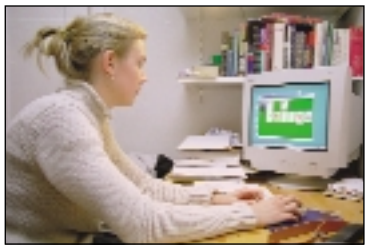
Solution 3 *Using Fathom™*

To create a histogram using Fathom™, enter the earthquake data into a new **Case Table**. Drag a new graph into the Fathom™ document and drag the attribute "magnitude" over to the horizontal axis of the graph. Change the graph type in the upper-right corner to **Histogram**. You can change the bin width by dragging the edges of a bin or by double-clicking on the graph and manually entering the values in the graph information dialogue.



Technolink
See Appendix D.8 on page 422 for more information on creating histograms with Fathom™.

DISTRIBUTION OF DATA



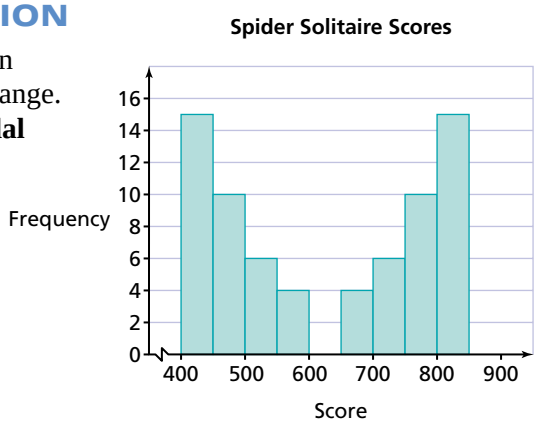
A frequency distribution is categorized by the general shape of its corresponding histogram. Typically, it is described in one of four ways.

Gillian plays a number of different games in her spare time. The following frequency distributions list samples of her scores. Each one of these samples represents a different type of distribution.

U-SHAPED DISTRIBUTION

bimodal—a distribution that has two peaks

A U-shaped distribution occurs when there are peaks at either end of the range. It may also be described as a **bimodal** distribution. The scores from the game of spider solitaire form this type of distribution.

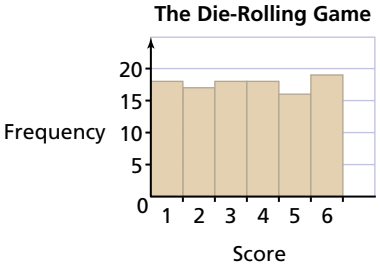


Spider Solitaire

Score	400–449	450–499	500–549	550–599	600–649	650–699	700–749	750–799	800–849
Frequency	15	10	6	4	0	4	6	10	15

UNIFORM DISTRIBUTION

When each outcome has a similar frequency, it is called a uniform distribution. The height of each bar is roughly equal. This is the distribution you would expect from an experiment such as rolling a die.

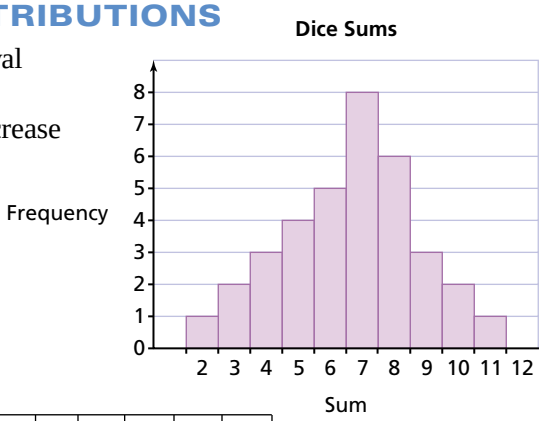


The Die-Rolling Game

Score	1	2	3	4	5	6
Frequency	18	17	18	18	16	19

MOUND-SHAPED DISTRIBUTIONS

In this distribution, there is an interval with the greatest frequency, and the frequencies of all other intervals decrease on either side of that. The frequency distribution then takes on a mound shape. Rolling a pair of dice and recording the sum results in this type of distribution.



Dice Sums

Sum	2	3	4	5	6	7	8	9	10	11	12
Frequency	1	2	3	4	5	8	6	3	2	1	0

symmetric distribution—when the data show a mirror symmetry about the centre

Each of these three distributions is **symmetric**.

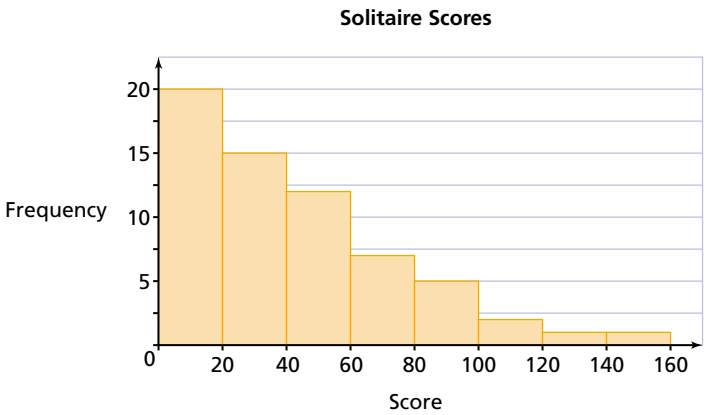
Think about
Skewed Distributions
What does the shape of the distribution say about your likelihood of achieving a high score?

SKEWED DISTRIBUTIONS

In a skewed distribution, the interval or group of intervals that contains the greatest frequencies is near one end of the histogram. As a result, these distributions seem to tail off to the left or right. The direction of the skew is determined by the direction the mean has shifted. This will be discussed in detail in Section 3.2.) Scores from a game of solitaire produce a distribution that is skewed to the right (right-skewed).

Solitaire

Score	0–19	20–39	40–59	60–79	80–99	100–119	120–139	140–159
Frequency	20	15	12	7	5	2	1	1



Think about
Symmetric Distributions
Explain why uniform, mound-shaped, and U-shaped distributions are also referred to as symmetric distributions.

KEY IDEAS

bin width—calculated as $\frac{range}{number\ of\ intervals}$; changing bin width may have a dramatic effect on the interpretation of data

U-shaped distribution—peaks at either end of the range; can be described as bimodal

uniform distribution—each outcome has a similar frequency

mound-shaped distribution—symmetrical about a line passing through the interval with the greatest frequency

symmetric distribution—shows mirror symmetry about the centre (e.g., uniform, U-shaped, and mound-shaped distributions)

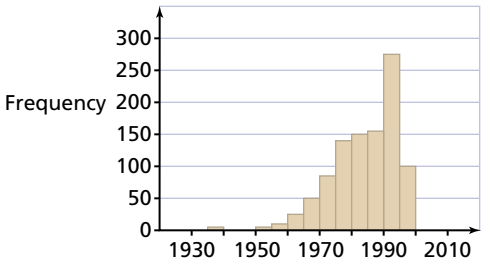
skewed distribution—an assymetrical distribution where the direction denotes skew type (right-skewed, left-skewed)

3.1 Exercises

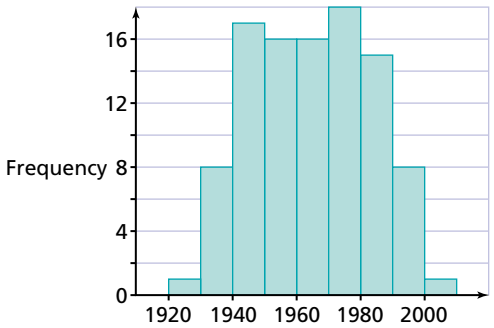
A

1. **Knowledge and Understanding** Describe each of the distributions that follows using the vocabulary of this section.

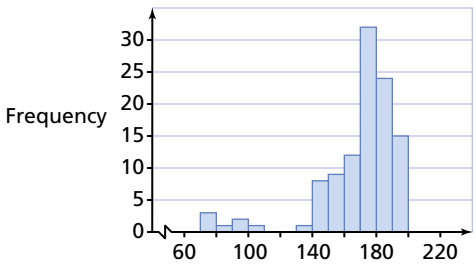
(a)



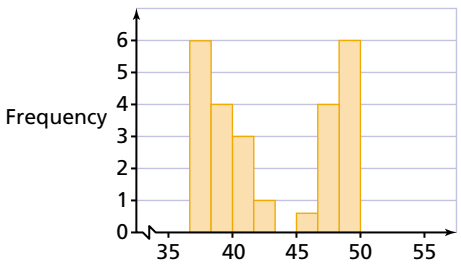
(b)



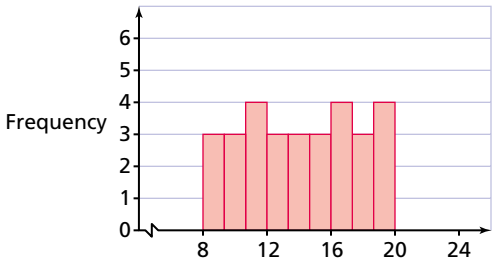
(c)



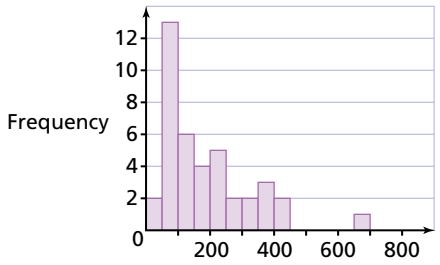
(d)



(e)



(f)



2. Match each of the preceding distributions with one of the scenarios described below.

(a) cost of the "cheap seats" at 30 baseball stadiums
(b) bowling scores
(c) the gestation period in days of various animals
(d) the year shown on a penny
(e) the production year of the American Film Institute's top 100 films
(f) amounts shown on an electric bill

3. **Communication** How can the bin width used in a histogram be used to misrepresent the data? Give examples.

4. **Knowledge and Understanding** Calculate a bin width that would form five uniform intervals for the following data.

(a) 13, 7, 5, 7, 9, 10, 5, 11, 8, 7, 9, 10, 10, 11, 14, 10, 6, 12, 6, 9, 7, 12, 9, 10, 6

(b) 0.59, 0.46, 0.48, 0.52, 0.15, 0.60, 0.86, 0.55, 0.97, 0.86, 0.37, 0.70, 0.60, 0.34, 0.65, 0.94, 0.60, 0.35

(c) 186, 124, 196, 206, 148, 162, 144, 167, 184, 152, 182, 184, 127, 183, 179, 185, 170, 135, 141, 126, 157, 126, 206, 196, 185, 186, 117, 129, 155, 146

(d) 0.8, −0.8, 0.0, 0.8, 0.1, 0.5, 0.9, 0.4, 0.3, 1.1, −0.8, −0.2, −2.2, 0.8, 1.9, −1.9, 0.1, 1.3, 0.9, −0.1, 0.8, 1.6, 1.4, 0.4, −0.8, 0.5, −0.3, −1.0, 1.2, −0.4, −1.1, 0.0, 0.2, −1.7, −0.3, 0.1, −1.6, −1.8, −0.6, −0.9

5. Calculate the starting and ending point for each of the five intervals in Question 4.

B **6.** Use the intervals in Question 5 to create an appropriate histogram.

7. **Application**

(a) Construct an appropriate graphical display of the following information from the Men's Professional Golf Association.

Rank	Name	Events	Earnings ($)
1	Tiger Woods	19	5 687 777
2	Phil Mickelson	23	4 403 883
3	David Toms	28	3 791 595
4	Vijay Singh	26	3 440 829
5	Davis Love III	20	3 169 463
6	Sergio Garcia	18	2 898 635

Rank	Name	Events	Earnings ($)
7	Scott Hoch	24	2 875 319
8	David Duval	20	2 801 760
9	Bob Estes	26	2 795 477
10	Scott Verplank	26	2 783 401
11	Mike Weir	23	2 777 936
12	Chris DiMarco	29	2 595 201

Source: CNN/Sports Illustrated

(b) The same information is given below for the Ladies' Professional Golf Association. Construct a graphic representation of the two tables that allows you to compare the winnings of the top 12 male and top 12 female golfers on tour.

Rank	Name	Events	Earnings ($)
1	Annika Sorenstam	26	2 105 868
2	Se Ri Pak	21	1 623 009
3	Karrie Webb	22	1 535 404
4	Lorie Kane	22	947 489
5	Maria Hjorth	29	848 195
6	Rosie Jones	23	785 010

Rank	Name	Events	Earnings ($)
7	Dottie Pepper	23	776 482
8	Mi Hyun Kim	29	762 363
9	Laura Diaz	27	751 466
10	Catriona Matthew	29	747 970
11	Rachel Teske	27	713 129
12	Wendy Ward	26	686 906

Source: CNN/Sports Illustrated

relative frequency—the percent that an interval represents of the whole population

cumulative frequency—the total of all frequencies up to a certain value of the variable

8. The table to the right gives the number of people by age group in a town of 100 000.

Age	Frequency
$A \leq 5$	12 100
$5 < A \leq 15$	22 300
$15 < A \leq 25$	19 600
$25 < A \leq 35$	16 000
$35 < A \leq 45$	12 200
$45 < A \leq 55$	8 400
$55 < A \leq 65$	5 300
$A > 65$	4 100

(a) Add two columns to the table. In the first, calculate the **relative frequency** by dividing the number of people by the total.

(b) In the second, calculate the **cumulative frequency** by finding the sum of all groups less than or equal to that group (the total so far).

(c) Construct a histogram of the relative frequency in which the bar height is the percent each interval represents of the population.

(d) Construct a histogram of the cumulative frequency.

(e) Use the histogram to describe the distribution.

(f) What effect does the bin size of the first and last groups have on your interpretation of these data?

9. The amounts withdrawn from an ATM are recorded over a single day.

$20	$60	$100	$ 20	$80	$ 40	$50
$20	$30	$ 60	$ 40	$20	$ 40	$60
$40	$60	$110	$ 30	$40	$ 40	$80
$80	$40	$ 20	$100	$40	$ 60	$40
$50	$80	$ 20	$ 60	$80	$200	$40

(a) What bin width gives a good representation of the data?

(b) What do you notice about the larger amounts? How does this affect the graph?

10. Record the age (in years and months) of each of your classmates, and then create a histogram for the data. What observations can you make?

11. Create a bar graph for each city.

	Temperature (°C)											
	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
Belleville, ON	–7.4	–6.5	–0.7	6.7	13.2	18.3	21.4	20.3	15.9	9.4	3.4	–4.1
Hamilton, ON	–5.5	–4.6	0.2	6.9	13.2	18.7	21.7	20.7	16.5	10.1	4.3	–2.2
Kapuskasing, ON	–19.0	–16.0	–9.0	0.5	8.6	13.8	17.0	15.3	10.0	4.2	–4.6	–15.0
Thunder Bay, ON	–15.0	–13.0	–5.6	2.7	9.0	13.9	17.7	16.4	11.2	5.4	–2.6	–11.0
Alert, NT	–31.9	–34.0	–33.0	–25.0	–12.0	–1.0	3.4	1.0	–9.7	–20.0	–27.0	–30.0

Source: Environment Canada

C

12. **Thinking, Inquiry, Problem Solving**

(a) Use scale creatively to make Alert look like Kapuskasing.

(b) Make Hamilton look different from Thunder Bay using scale.

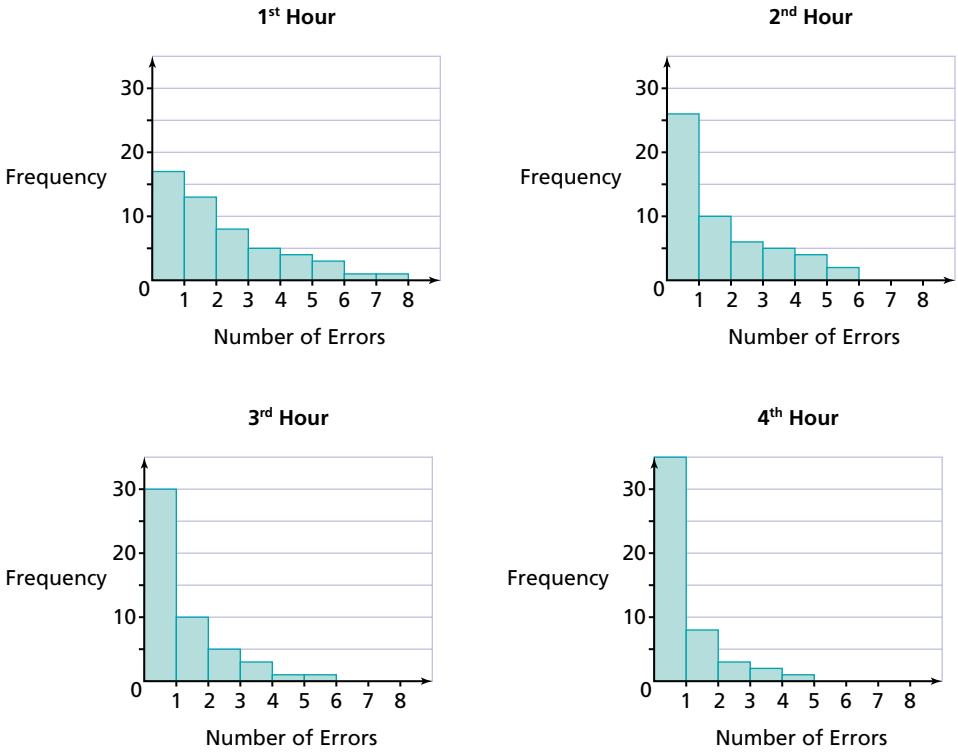
13. Considering the climate data in Question 11, which city has the most moderate climate (i.e., least difference between high and low)? Justify your position with the data.

ADDITIONAL ACHIEVEMENT CHART QUESTIONS

14. **Knowledge and Understanding** Create a bin width that will divide the following data into seven equally sized intervals: 2, 7, 20, 4, 11, 25, 6, 28, 3, 6, 18, 5, 13, 4, 10, 16, 23, 22, 5, 8, 3, 12, 6, 13, 12, 7, 8, 26.

15. **Application** Create a histogram using the data in Question 14.

16. **Thinking, Inquiry, Problem Solving** The following histograms show the number of production errors on vehicles coming off the assembly line during the first, second, third, and fourth hour of the day shift.



Describe the relationship between time and the number of errors. Create one histogram that summarizes this relationship.

17. **Communication** For a histogram to be an accurate display of the distribution of data, it must use intervals of equal width. Explain why this is necessary.

Chapter Problem

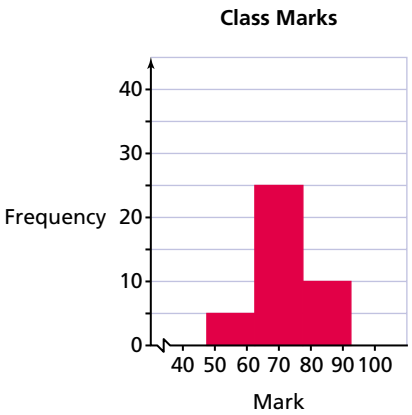
Comparing Marks

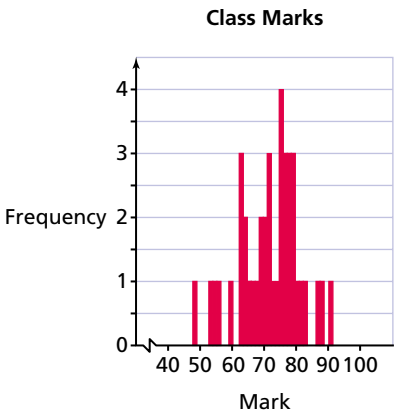
Justin would like to display the information graphically in order to determine whether the average marks of his fellow students are typical of the general population of students applying to university.

Use technology to create a histogram of the data on page 140 that visually displays the distribution of marks.

CP1. What bin width gives a good impression of the performance of Justin's schoolmates?

CP2. Comment on the appropriateness of the histograms that follow.





CP3. Where does Justin's overall average of 82% fall in relation to the rest of his fellow applicants?

CP4. What does this tell Justin about his prospects?