1.3 The Power of Visualizing Data— Trends

The data in Section 1.2 helped you determine if the issues raised were supported or not supported by the data. Data like these represent the moment in time when they were collected. Once you have identified a pattern at one moment in time, you might find it useful to look at the data over a longer period. Looking at data collected over a longer period of time may show trends and allow you to make predictions about future events. One effective way to predict these events is to create a visual display of the data in the form of a scatter plot.

Example 1 Creating a Scatter Plot

scatter plot—a graphical method of showing the joint distribution of two variables in which each point on the graph represents a pair of values

independent variable—a variable whose values are arbitrarily chosen

dependent variable—a variable whose values depend on the independent variable

Many people (especially insurance agents and driving instructors) are interested in the relationship between the **independent variable** (driver age) and the **dependent variable** (the number of fatal driving accidents per year). The number of fatal accidents over a three-year period and involving 18- to 25-year-olds is given below. Use a scatter plot of the data to determine if there is a relationship.

Age	Number of Fatal Accidents		
	Year 1	Year 2	Year 3
18	8	4	2
19	7	7	6
20	1	7	1
21	6	2	10
22	3	10	4
23	1	9	8
24	10	3	3
25	10	2	9

Solution

Typically, the independent variable is on the horizontal axis and the dependent variable is on the vertical axis. Each piece of data is then plotted as an individual point. A legend is used to differentiate points from one year to the next.

Fatal Accidents by Year
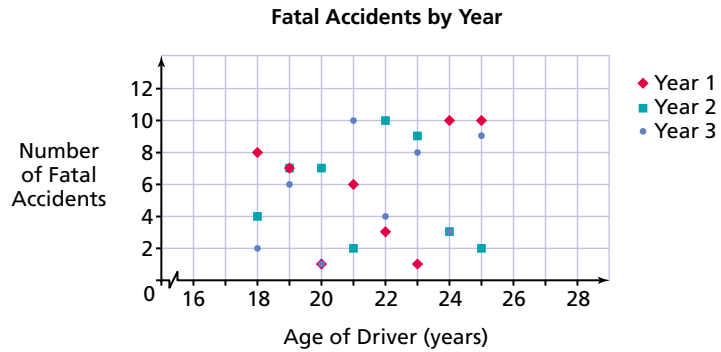
The plotted points appear to be randomly placed; no relationship is clearly evident.

INVESTIGATION: TUITION FEES—WILL I BE ABLE TO AFFORD THEM?

Purpose

To determine whether a table or a graph is more effective in predicting how much one might expect to pay in tuition fees when entering college or university.

Canadian Tuition Consumer Price Index 1975–2000

Year	Relative Cost ($)	Year	Relative Cost ($)
1975	29.6	1988	68.3
1976	29.9	1989	70.2
1977	31.0	1990	75.0
1978	33.4	1991	87.3
1979	34.2	1992	100.0
1980	36.1	1993	111.3
1981	39.6	1994	123.8
1982	41.5	1995	126.5
1983	49.7	1996	140.1
1984	50.2	1997	161.9
1985	52.0	1998	171.4
1986	57.4	1999	184.3
1987	60.1	2000	198.6

Source: Statistics Canada

Procedure

The above table shows a comparison of tuition fees based on assigning $100 as the value of the fees in 1992. The other fees are determined relative to the fees in 1992 using the consumer price index (CPI).

A. Make a scatter plot of the data shown in the table. Describe any patterns that you observe in the graph that you did not notice in the table.

B. In what year(s) does the relative cost increase the most? The least?

trend—a pattern of average behaviour that occurs over time

C. Which of the following words would you use to describe the **trend** of the data: *steady growth*, *steady decline*, *irregular growth*, or *irregular decline*? Justify your choice.

D. Predict the cost of tuition when you enter college or university relative to the cost in 1992.

line of best fit—a straight-line graph that best represents a set of data

E. Draw a **line of best fit**. Was it easier to use the table or the scatter plot to make your prediction? Give reasons for your answer.

Discussion Questions

1. Explain how you arrived at your prediction in Step D. Compare your prediction with those of your classmates.

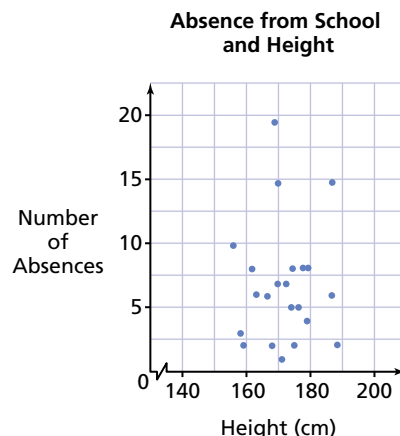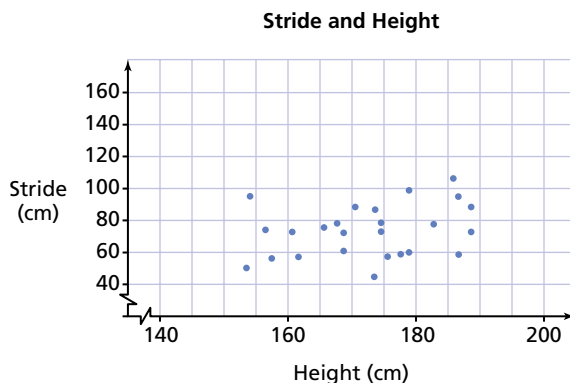**2.** Copy the following *confidence scale* into your notebook.

1	2	3	4	5	6	7	8	9	10
Not Confident									Completely Confident
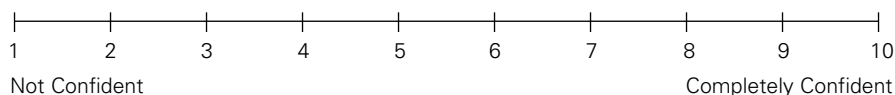
Place a dot on the scale to show your confidence in your prediction. Give reasons for your answer.

3. Do you think you can afford to go to college or university? Why or why not?
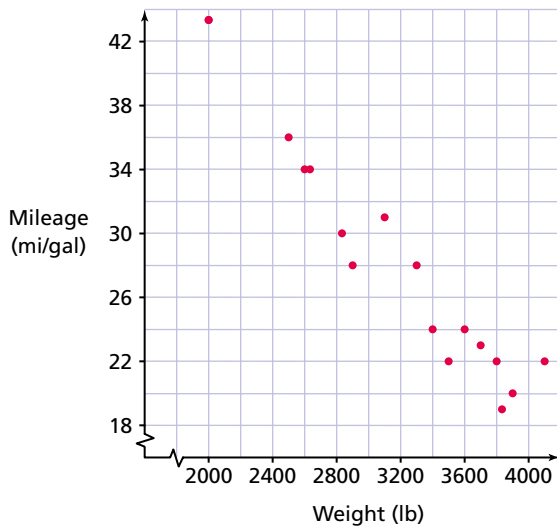
DETERMINING RELATIONSHIPS: CORRELATION

A scatter plot helps reveal a relationship by showing a general trend in the data. The arrangement of points helps determine the type and strength of the relationship. Consider the following scatter plots.
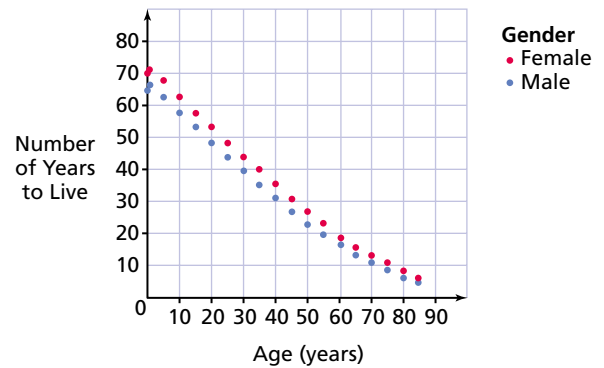
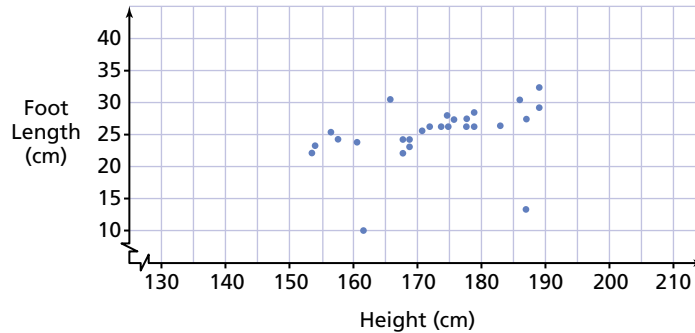Use the graphs to answer the Think about questions on page 34.

Stride and Height

Absence from School and Height

Vehicle Weight and Mileage in 1978

Mileage (mi/gal) vs. Weight (lb)

Life Expectancy

Number of Years to Live vs. Age (years)

Gender
- Female
- Male

Foot Length and Height

Foot Length (cm) vs. Height (cm)

Think about
The Scatter Plots

- Which of the scatter plots indicate the strongest trend? Give reasons for your answers.
- Which of the scatter plots do not indicate a trend? Give reasons for your answer.
- If a line of best fit were drawn on each of the scatter plots that show a trend, describe the slope of each line.

correlation—the apparent relation between two variables

Think about
Negative Correlation

Write a definition for a strong negative correlation and a weak negative correlation.

median–median line—a linear model used to fit a line to a data set. The line is fit only to key points calculated using medians.

There appears to be no relationship between absences and height; the dots are scattered randomly. The strongest relationship appears to be in life expectancy. A very clear trend is evident from the data.

The scatter plot showing a strong trend and having a line of best fit with a positive slope is said to have a strong positive **correlation**. The scatter plot showing a trend that is not strong and that has a line of best fit with a positive slope is said to have a weak positive correlation. The scatter plot showing no trend is said to have no correlation.

Negative correlation occurs in scatter plots where a line of best fit has a negative slope.

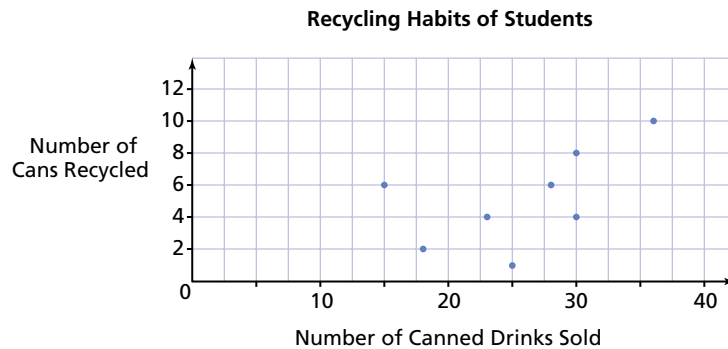## Example 2 The Median–Median Line

The environment club is interested in the relationship between the number of canned drinks sold in the cafeteria and the number of cans that are recycled. The data they collected are listed below.

| Number of Canned Drinks Sold	15	18	23	25	28	30	30	36
Number of Cans Recycled	6	2	4	1	6	8	4	10

(a) Draw a scatter plot for this information.

(b) Observe the trend in the data and identify the type of correlation.

(c) Draw a median–median line for the line of best fit.

(d) Use the line drawn in part (c) to determine the number of cans that will likely be found in the recycling box if 35 canned drinks are sold.

Solution

(a)

Recycling Habits of Students

Number of Cans Recycled

Number of Canned Drinks Sold

(b) The trend has a weak positive correlation because the data points are fairly spread out; yet, they suggest a line with positive slope.

(c) To draw the median–median line, the data are broken up into three vertical sections. As much as possible, each section contains an equal number of data points.

Recycling Habits of Students

3 points   2 points   3 points

Number of Cans Recycled

Number of Canned Drinks Sold
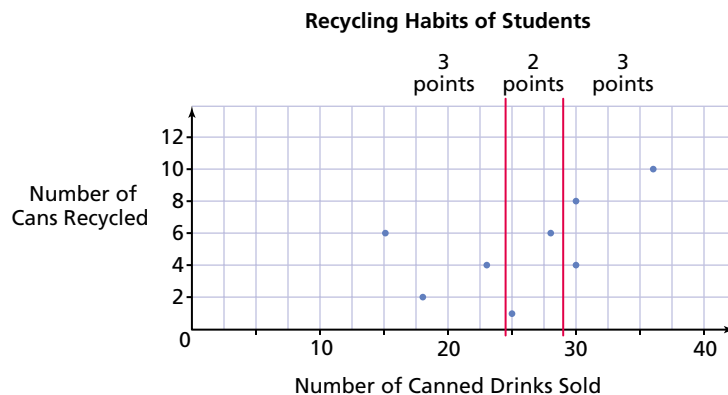
Find the median of the x-coordinates (x-median) and the median of the y-coordinates (y-median) in each section. In the first section, the x-median of 15, 18, and 23 is 18, and the y-median of 6, 2, and 4 is 4. Plot the median point (18, 4). The median point is indicated on the graph that follows by a Δ. Repeat the process for the other two sections. Place the edge of a ruler along the line joining the first and third median points. If the second point is not on this line, slide the ruler about a third of the way toward the second point. Ensure that the slope of the line has not changed. Draw the median–median line along the edge of the ruler.

Recycling Habits of Students
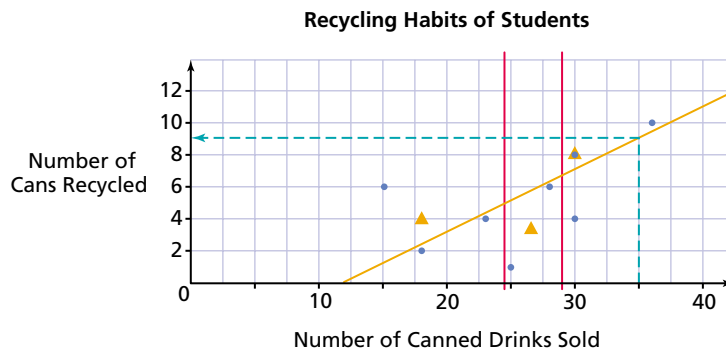
(d) They would expect to find nine cans in the recycling box.

KEY IDEAS

scatter plot—a graphical method of showing the joint distribution of two variables in which each point on the graph represents a pair of values

independent variable—a variable whose value is arbitrarily chosen

dependent variable—a variable whose value depends on the independent variable

trend—a pattern of average behaviour that occurs over time

line of best fit—a straight-line graph that best represents a set of data

correlation—the apparent relation between two variables

strong linear correlation—occurs if the two variables vary at similar rates

weak linear correlation—occurs if the two variables vary at rates that are not similar

positive linear correlation—occurs if the slope of the line of best fit or median–median line is positive

negative linear correlation—occurs if the slope of the line of best fit or median–median line is negative

median–median line—a linear model used to fit a line to a data set; the line is fit only to key points calculated using medians

1.3 Exercises

A

1. **Knowledge and Understanding** Make a scatter plot to display the data in each of the following tables.

(a)

Planet	Time of One Revolution (year)	Mean Distance from the Sun (AU)
Mercury	0.24	0.39
Venus	0.61	0.72
Earth	1.00	1.00
Mars	1.88	1.52
Jupiter	11.86	5.20
Saturn	29.46	9.54
Uranus	84.07	19.19
Neptune	164.82	30.06
Pluto	247.68	39.53

Note: AU $\doteq$ 150 million km

(b) Students posing for yearbook photos have several package options.

Number of Photos	44	30	24	15
Total Cost ($)	18.00	16.00	13.00	10.00

(c) The number of available seats and the average speed of a variety of planes are listed below (the data are taken from Fathom™).

Plane	Seats	Average Speed (km/h)	Plane	Seats	Average Speed (km/h)
B747-400	396	538	B737-400	144	414
B747-100	447	520	MD-80	141	432
L-1011-100/200	310	495	B737-300	131	417
DC-10-10	289	500	DC-9-50	122	374
A300-600	249	473	B737-100/200	113	389
DC-10-30	265	520	DC-9-30	101	385
B767-300ER	214	495	F-100	97	383
B767-200ER	181	488	DC-9-10	71	381
A320-100/200	148	460	B777	292	521
B727-200	150	437	MD-11	253	524

(d) A group of teenagers recorded the length of their forearm span in centimetres and their age in months. The results are shown below.
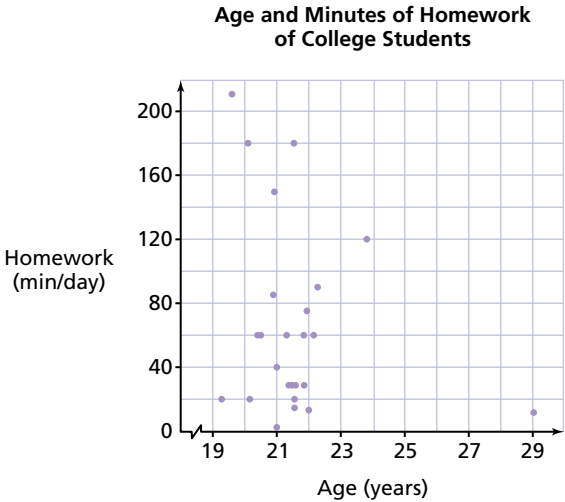
Forearm Span (cm)	20.5	23	27	30	27	24.7	25.5	24	28	26
Age (months)	214	221	237	216	216	210	216	219	209	214

2. Communication For each of the following graphs, state whether or not there appears to be a trend. If there is, state whether the correlation is strong or weak, and positive or negative.
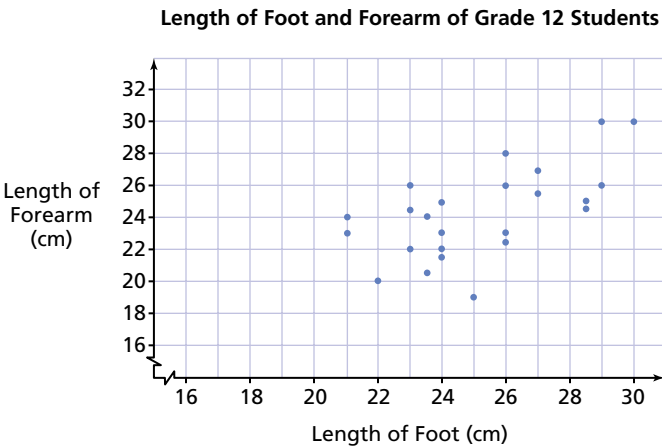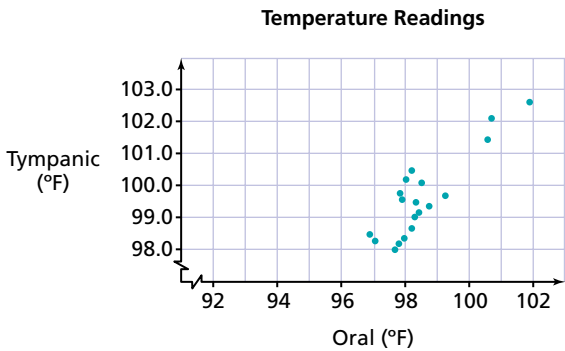
(a)

Age and Sleep Habits of College Students

(b)

Age and Minutes of Homework of College Students

(c)

Length of Foot and Forearm of Grade 12 Students

(d)

Temperature Readings
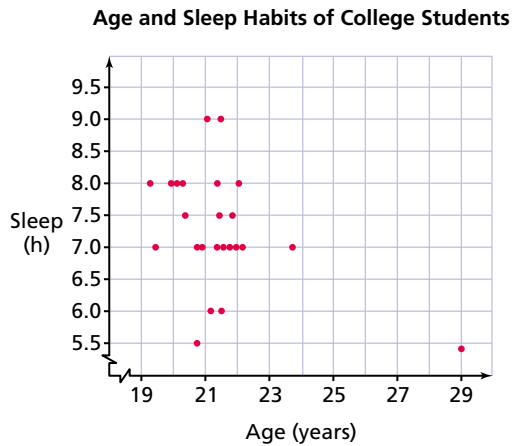
3. **Application** A car manufacturer has tested the stopping distance of a new model of car relative to the speed when the brakes are applied. All testing was done on dry pavement. The results are shown in the table below. The stopping distance is defined as the distance travelled from when the brakes are applied to the time when the vehicle comes to a complete stop.

Speed of Car (km/h)	25	35	45	55	60	70	80	90	100	110
Stopping Distance (m)	10	15	21	27	33	42	54	61	78	103

   (a) Create a scatter plot.
   (b) Is there a trend? Give reasons for your answer.
   (c) Describe the type of correlation that might be appropriate for the data.

4. Every year, students at a local high school collect money for a local charity. They keep track of the number of students who participate, as well as the amount of money that is collected. The information for the past five years is listed in the table below.

Year	Number of Students	Amount Collected ($)
1	130	2250
2	125	2875
3	135	2500
4	147	2300
5	153	2000

   (a) Create a scatter plot for the information in the table.
   (b) Describe the correlation that is observed in the data.

B 5. Attendance at school dances for last year was recorded.

Month	September	October	November	December	February	March
Attendance	250	245	200	260	285	280

   (a) Create a scatter plot for the information in the table.
   (b) Describe any trends that you observe.

Answer Questions 6 through 9 using a scatter plot with a median–median line.

6. The attendance at school hockey games is shown below.

Game	1	2	3	4	5	6	7	8
Attendance	125	111	122	105	100	93	85	72

   Predict the attendance for Game 9. Give reasons for your answer.

7. Twenty members of your class toss a coin for five minutes and record the number of tails tossed. The totals are shown below.

Number of Tosses	Number of Tails	Number of Tosses	Number of Tails
21	15	41	17
36	25	41	14
26	20	52	25
43	18	36	24
37	19	51	25
45	25	50	14
47	26	45	30

If you were to toss the coin 30 times, how many tails would you expect? Give reasons for your answer.

8. The winning women's Olympic long-jump distance is shown in the table below.

Year	Distance (m)	Year	Distance (m)
1948	5.69	1976	6.72
1952	6.24	1980	7.06
1956	6.35	1984	6.96
1960	6.37	1988	7.40
1964	6.76	1992	7.14
1968	6.82	1996	7.12
1972	6.78	2000	6.99

Source: British Broadcasting Corporation (BBC)

If the Olympics had been held in 1944, what might the winning distance have been?

9. A local movie theatre monitors attendance during the first 10 weeks of a movie's showing. The results of one movie are listed below.

Week	1	2	3	4	5	6	7	8	9	10
Attendance	2250	2100	1950	1678	1430	1200	987	731	675	587

If less than 200 people attend a movie, the theatre loses money. How many more weeks will the movie run?

10. **Thinking, Inquiry, Problem Solving** Find a set of data that appears to have a linear correlation. You may find the data in an almanac, newspaper, magazine, or in Appendix A, or on the Internet.
    (a) Record the data in a table.
    (b) Create a scatter plot for the data.
    (c) Describe any trends in the data. Give reasons for your answer.
    (d) Construct the median–median line for the data.
    (e) Write a question that requires the median–median line to make a prediction.
    (f) Exchange your table and question with a partner. Respond to your partner's question.
    (g) Compare your answers to both questions.

ADDITIONAL ACHIEVEMENT CHART QUESTIONS

The table below shows the number of calories a person might use while in-line skating at a comfortable pace.

Time (min)	3	7	12	18	25	35
Energy Used (in calories)	28	70	119	170	241	320

11. **Knowledge and Understanding** Make a scatter plot of the data and construct the median–median line.

12. **Communication** Describe a trend in the data in terms of correlation.

13. **Application** Determine the equation of the median–median line that you constructed.

14. **Thinking, Inquiry, Problem Solving** If a person burned 1000 calories while in-line skating, determine the length of time that she or he skated. How confident are you that your prediction is valid? Give reasons for your answer.

Chapter Problem
Trends in Canada's Population

Use the data given in the chapter problem on page 2 to answer the following questions.

CP7. Create a single scatter plot that illustrates the trends in each age class since 1951.
CP8. For each age class, describe the trends that you see. Where are these trends most visible: in the table or in the graph? Explain.
CP9. For each age class, draw the median–median line on your graph.
CP10. Describe the type of correlation that exists within each age class.
CP11. Provide some possible reasons for the trends that you see.