

# Bivariate data and scatter plots

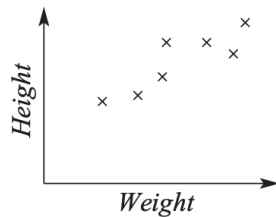
WALT understand and draw scatter plots

Success Criteria I know ....

- When we collect information about two variables in a given context.
- We look for relationships between two variables
- 



When we collect information about two variables in a given context we are collecting bivariate data. As there are two variables involved in bivariate data, we use a number plane to graph the data. These graphs are called scatter plots and are used to show a relationship that may exist between the variables. Scatter plots make it very easy to see the strength of the relationship between the two variables.



- If you feel that a relationship exists, would you expect the second listed variable to increase or to decrease as the first variable increases?
  - a Height of person and Weight of person
  - b Temperature and Life of milk
  - c Length of hair and IQ
  - d Depth of topsoil and Brand of motorcycle
  - e Years of education and Income
  - f Spring rainfall and Crop yield
  - g Size of ship and Cargo capacity
  - h Fuel economy and CD track number
  - i Amount of traffic and Travel time
  - j Cost of 2 litres of milk and Ability to swim
  - k Background noise and Amount of work completed

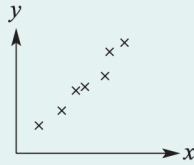
[Check correlation](#) try only level 1

■ Types of correlation:

- The correlation is positive if the  $y$  variable generally increases as the  $x$  variable increases.
- The correlation is negative if the  $y$  variable generally decreases as the  $x$  variable increases.

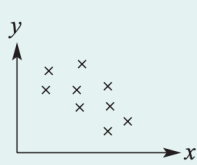
Examples:

Strong positive correlation



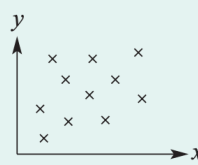
As  $x$  increases,  $y$  clearly increases.

Weak negative correlation



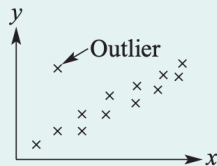
As  $x$  increases,  $y$  generally decreases.

No correlation



As  $x$  increases, there is no particular effect on  $y$ .

- An outlier can clearly be identified as a data point that is isolated from the rest of the data.



- 1 Decide whether it is likely or unlikely that there will be a strong relationship between these pairs of variables.

- a Height of door and width of door
- b Weight of car and fuel consumption
- c Temperature and length of phone calls
- d Colour of flower and strength of perfume
- e Amount of rain and size of vegetables in the vegetable garden

- 2 For each of the following sets of bivariate data with variables  $x$  and  $y$ :

- i draw a scatter plot by hand
- ii decide whether  $y$  generally increases or decreases as  $x$  increases

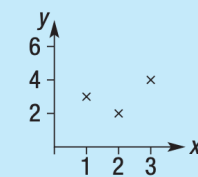
a

$x$	1	2	3	4	5	6	7	8	9	10
$y$	3	2	4	4	5	8	7	9	11	12

b

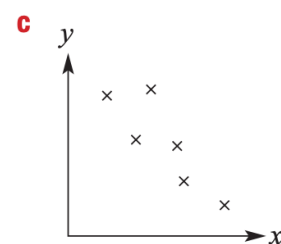
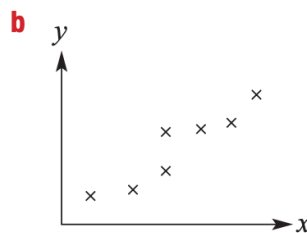
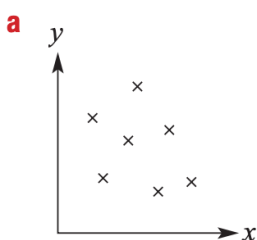
$x$	0.1	0.3	0.5	0.9	1.0	1.1	1.2	1.6	1.8	2.0	2.5
$y$	10	8	8	6	7	7	7	6	4	3	1

On a scatter plot, mark each point of the plot with a  $x$ .



- 3 For these scatterplots, choose two words from those listed below to best describe the correlation between the two variables.

*strong    weak    positive    negative*



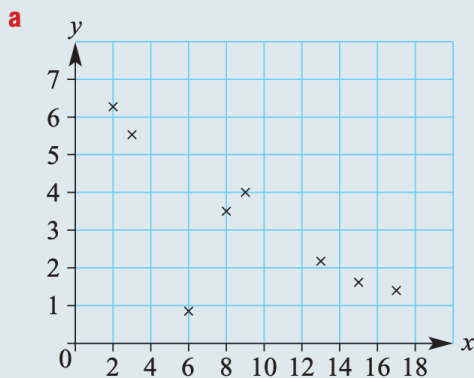
### Example 15 Constructing and interpreting scatter plots

Consider this simple bivariate data set.

$x$	13	9	2	17	3	6	8	15
$y$	2.1	4.0	6.2	1.3	5.5	0.9	3.5	1.6

- Draw a scatter plot for the data.
- Describe the correlation between  $x$  and  $y$  as positive or negative.
- Describe the correlation between  $x$  and  $y$  as strong or weak.
- Identify any outliers.

#### Solution



- Negative correlation
- Strong correlation
- The outlier is (6, 0.9).

#### Explanation

Draw an appropriate scale on each axis by looking at the data:

- $x$  is up to 17
- $y$  is up to 6.2

The scale must be spread evenly on each axis.

Plot each point using a  $\times$  symbol on graph paper.

Looking at the scatterplot, as  $x$  increases  $y$  decreases.

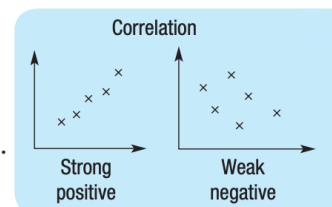
The downwards trend in the data is clearly defined.

This point defies the trend.

- 4** Consider this simple bivariate data set.

$x$	1	2	3	4	5	6	7	8
$y$	1.0	1.1	1.3	1.3	1.4	1.6	1.8	1.0

- Draw a scatter plot for the data.
- Describe the correlation between  $x$  and  $y$  as positive or negative.
- Describe the correlation between  $x$  and  $y$  as strong or weak.
- Identify any outliers.



- 5** Consider this simple bivariate data set.

$x$	14	8	7	10	11	15	6	9	10
$y$	4	2.5	2.5	1.5	1.5	0.5	3	2	2

- Draw a scatter plot for the data.
- Describe the correlation between  $x$  and  $y$  as positive or negative.
- Describe the correlation between  $x$  and  $y$  as strong or weak.
- Identify any outliers.

- 6 By completing scatter plots for each of the following data sets, describe the correlation between  $x$  and  $y$  as 'positive', 'negative' or 'none'.

**a**

$x$	1.1	1.8	1.2	1.3	1.7	1.9	1.6	1.6	1.4	1.0	1.5
$y$	22	12	19	15	10	9	14	13	16	23	16

**b**

$x$	4	3	1	7	8	10	6	9	5	5
$y$	115	105	105	135	145	145	125	140	120	130

**c**

$x$	28	32	16	19	21	24	27	25	30	18
$y$	13	25	22	21	16	9	19	25	15	12

Problem-solving and Reasoning

- 7 A tomato grower experiments with a new organic fertiliser and sets up five separate garden beds: A, B, C, D and E. The grower applies different amounts of fertiliser to each bed and records the diameter of each tomato picked.

The average diameter of a tomato from each garden bed and the corresponding amount of fertiliser are recorded below.

Bed	A	B	C	D	E
Fertiliser (grams per week)	20	25	30	35	40
Average diameter (cm)	6.8	7.4	7.6	6.2	8.5

- a** Draw a scatter plot for the data with 'Diameter' on the vertical axis and 'Fertiliser' on the horizontal axis. Label the points A, B, C, D and E.  
**b** Which garden bed appears to go against the trend?  
**c** According to the given results, would you be confident in saying that the amount of fertiliser fed to tomato plants does affect the size of the tomato produced?



- 8 For common motor vehicles, consider the two variables *Engine size* (cylinder volume) and *Fuel economy* (number of kilometres travelled for every litre of petrol).

- a** Do you expect there to be some relationship between these two variables?  
**b** As the engine size increases, would you expect the fuel economy to increase or decrease?  
**c** The following data was collected for 10 vehicles.

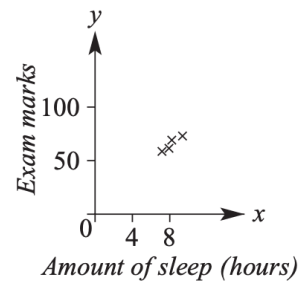
Car	A	B	C	D	E	F	G	H	I	J
Engine size	1.1	1.2	1.2	1.5	1.5	1.8	2.4	3.3	4.2	5.0
Fuel economy	21	18	19	18	17	16	15	20	14	11

- i** Does the data generally support your answers to parts **a** and **b** above?  
**ii** Which car gives a fuel economy reading that does not support the general trend?

- 9 On 14 consecutive days a local council measures the volume of sound heard from a freeway at various points in a local suburb. The volume ( $V$ ) of sound is recorded against the distance ( $d$  m) between the freeway and the point in the suburb.

Distance ( $d$ )	200	350	500	150	1000	850	200	450	750	250	300	1500	700	1250
Volume ( $V$ )	4.3	3.7	2.9	4.5	2.1	2.3	4.4	3.3	2.8	4.1	3.6	1.7	3.0	2.2

- a Draw a scatter plot of  $V$  against  $d$ , plotting  $V$  on the vertical axis and  $d$  on the horizontal axis.
- b Describe the correlation between  $d$  and  $V$  as positive, negative or none.
- c Generally as  $d$  increases, does  $V$  increase or decrease?
- 10 A person presents you with this scatter plot and suggests to you that there is a strong correlation between the amount of sleep and exam marks. What do you suggest is the problem with the person's graph and conclusions?



### Crime rates and police

- 11 A government department is interested in convincing the electorate that a large number of police on patrol leads to lower crime rates. Two separate surveys are completed over a one-week period and the results are listed in this table.

	Area	A	B	C	D	E	F	G
Survey 1	Number of police	15	21	8	14	19	31	17
	Incidence of crime	28	16	36	24	24	19	21
Survey 2	Number of police	12	18	9	12	14	26	21
	Incidence of crime	26	25	20	24	22	23	19

- a By using scatter plots, determine whether or not there is a relationship between the number of police on patrol and the incidence of crime, using the data in:
- i survey 1                                      ii survey 2
- b Which survey results do you think the government will use to make its point? Why?

Number of police will be on the horizontal axis.



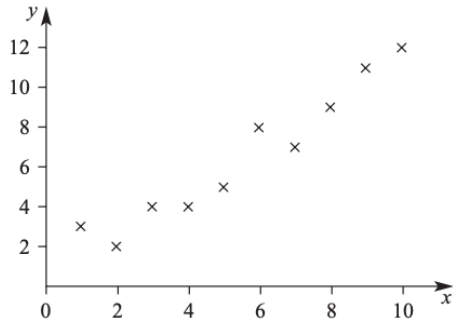
[Plot scatter graphs on transum](#)

Check your answers



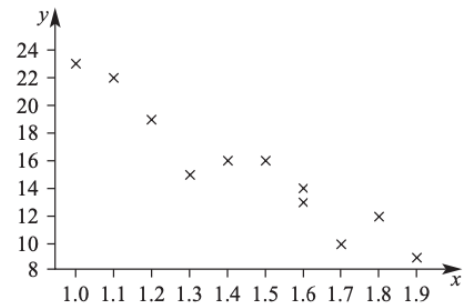
- 1 a Likely    b Likely    c Unlikely  
 d Unlikely    e Likely

2 a i



ii  $y$  generally increases as  $x$  increases.

6 a Negative



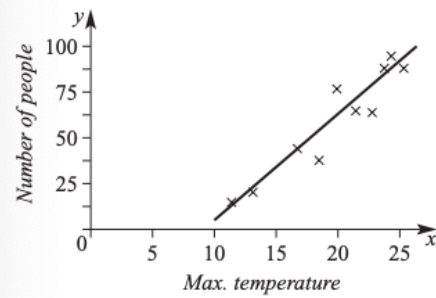
b negative    c as above

d i 13.5    ii 23    iii 9    iv 7

7 a  $\approx 4.5$     b  $\approx 6$     c  $\approx 0.5$     d  $\approx 50$

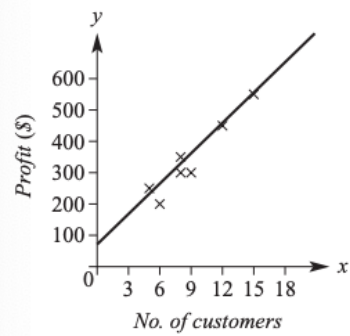
8 a increases

b



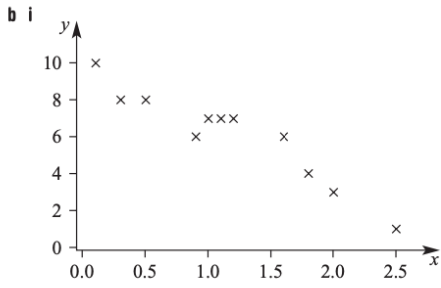
c i 65    ii 15 degrees

9 a, b



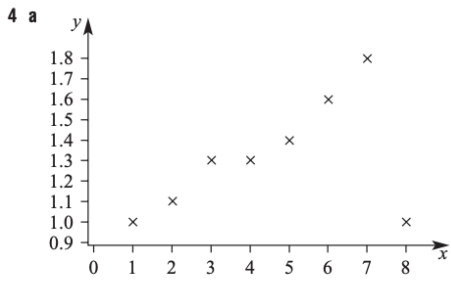
c \$600

d 2

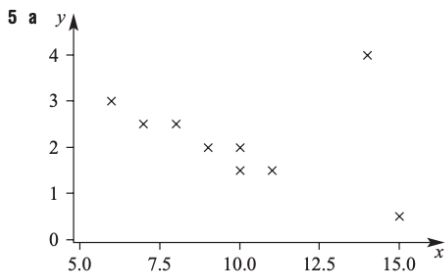


ii  $y$  generally decreases as  $x$  increases.

- 3 a** weak negative  
**b** strong positive  
**c** strong negative

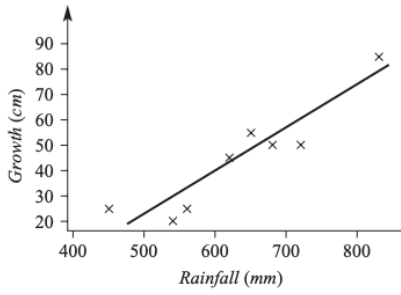


- b** Positive   **c** Strong   **d** (8, 1.0)



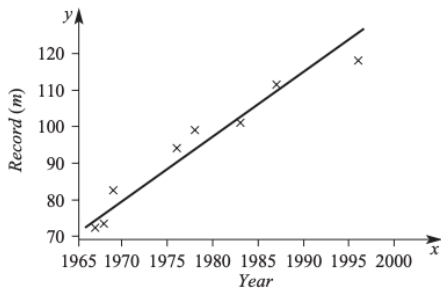
- b** Negative  
**c** Strong  
**d** (14, 4)

10 a, b



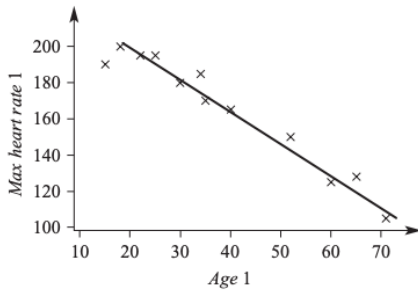
- c i  $\approx 25$  cm      ii  $\approx 85$  cm  
 d i  $\approx 520$  mm      ii  $\approx 720$  mm

11 a, b



- c i 130 m    ii 170 m  
 d No, records are not likely to continue to increase at this rate.

12 a Experiment 1



Experiment 2

