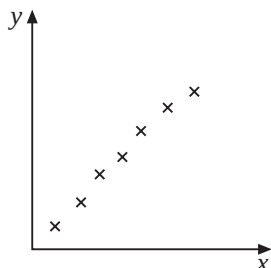
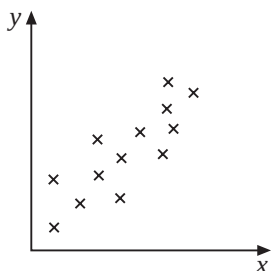


## 8.3 Plotting Scatter Diagrams

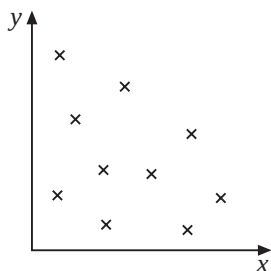
In this section we review plotting scatter diagrams and discuss the different types of correlation that you can expect to see on these diagrams.



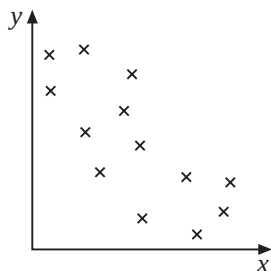
*Strong positive correlation* between  $x$  and  $y$ . The points lie close to a straight line with  $y$  increasing as  $x$  increases.



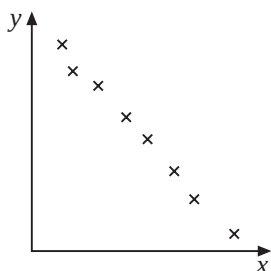
*Weak, positive correlation* between  $x$  and  $y$ . The trend shown is that  $y$  increases as  $x$  increases but the points are not close to a straight line.



*No correlation* between  $x$  and  $y$ ; the points are distributed randomly on the graph.



*Weak, negative correlation* between  $x$  and  $y$ . The trend shown is that  $y$  decreases as  $x$  increases but the points do not lie close to a straight line.



*Strong, negative correlation.* The points lie close to a straight line, with  $y$  decreasing as  $x$  increases.

If the points plotted were all on a straight line we would have *perfect correlation*, but it could be positive or negative as shown in the diagrams above.



### Example 1

The following table lists values of  $x$  and  $y$ .

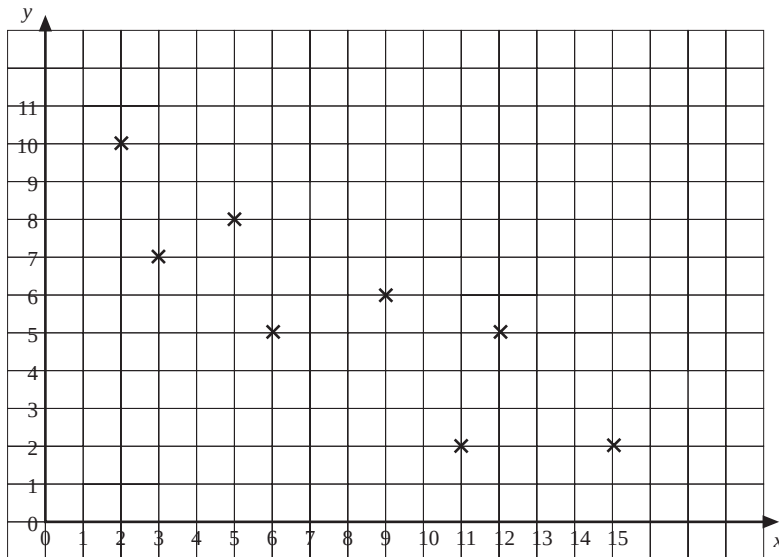
$x$	2	3	5	6	9	11	12	15
$y$	10	7	8	5	6	2	5	2

- Use the data to draw a scatter graph.
- Describe the type of correlation that you observe.



### Solution

- The scatter graph is shown below.



- It shows weak, negative correlation.



### Example 2

What sort of correlation would you expect to find between:

- a person's age and their house number,
- a child's age and their height,
- an adult's age and their height ?



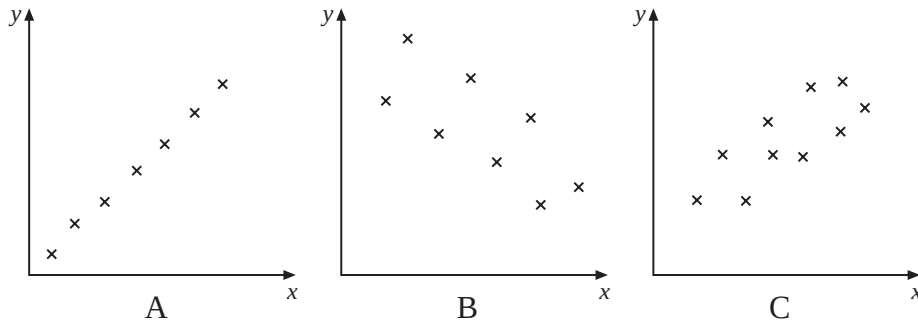
### Solution

- No correlation, because these two quantities are not linked in any way.
- Positive correlation, because children get taller as they get older.
- No correlation, because the height of adults does not change with their age.



## Exercises

1. Consider the following scatter graphs:



- Which graph shows *strong* correlation?
- Which graphs show *positive* correlation?
- Which graph shows *negative* correlation?
- Which graph shows a *weak, positive* correlation?

2. The following table lists values of  $x$  and  $y$ .

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

- Plot a scatter graph for this data.
- Describe the correlation between  $x$  and  $y$ .

3. Copy and complete the table below for 10 people in your class.

<i>House Number</i>										
<i>Day of Month of Birthday</i>										

- Plot a scatter graph for your data.
- Describe the type of correlation that there is between these two quantities.

4. A driver keeps a record of the distance travelled and the amount of fuel in his tank on a long journey.

<i>Distance Travelled (km)</i>	0	50	100	150	200	250	300
<i>Fuel in Tank (litres)</i>	80	73	67	61	52	46	37

- Illustrate this data with a scatter plot.
- Describe the type of correlation that is present.

5. What type of correlation would you expect to find between each of the following quantities:
- Age and pocket money,
  - IQ and height,
  - Price of house and number of bedrooms,
  - Person's height and shoe size ?
6. In a class 10 pupils took a Science test and an English test. Their scores are listed in the following table:

<i>Pupil</i>	A	B	C	D	E	F	G	H	I	J
<i>English Score</i>	2	10	18	4	9	7	18	19	3	10
<i>Science Score</i>	18	12	6	3	11	20	4	17	7	2

- Draw a scatter graph for this data.
  - Describe the correlation between the two scores.
7. Chris carries out an experiment in which he measure the extension of a spring when he hangs different masses on it. The following table lists his results:

<i>Mass (grams)</i>	20	50	100	120	200
<i>Extension (cm)</i>	1.2	3.0	6	7.2	12

- Draw a scatter graph for this data.
  - Describe the correlation between the mass and the extension.
8. Every day Peter picks the ripe tomatoes in his greenhouse. He keeps a record of their mass and the number that he picks. His results are listed in the following table:

<i>Number of Tomatoes Picked</i>	1	3	2	5	8	6	7	4
<i>Total Mass (grams)</i>	40	180	60	270	390	220	420	210

- Draw a scatter graph for this data.
- Describe the correlation between the number of tomatoes picked and their total mass.

9. A competition has 3 different games.  
 (a) Jeff plays 2 of the games.

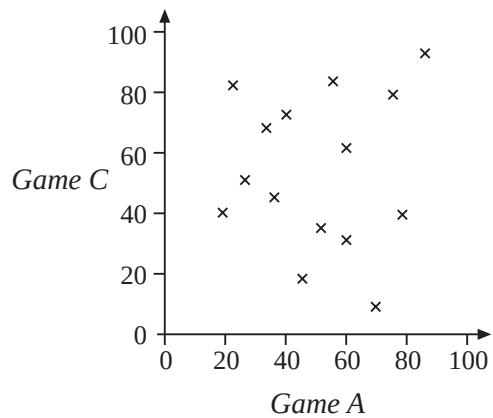
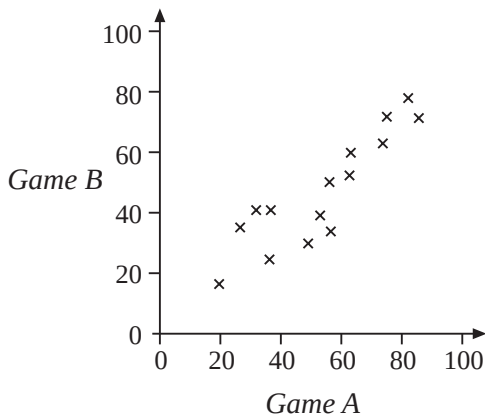
	<i>Game A</i>	<i>Game B</i>	<i>Game C</i>
<i>Score</i>	62	53	

To win, Jeff needs a *mean* score of 60. How many points does he need to score in Game C? Show your working.

- (b) Imran and Nia play the 3 games. Their scores have the *same mean*. The *range* of Imran's score is *twice* the range of Nia's scores. Copy the following table and fill in the missing scores.

<i>Imran's Scores</i>		40	
<i>Nia's Scores</i>	35	40	45

The scatter diagrams show the scores of everyone who plays all 3 games.



- (c) Look at the scatter diagrams. Write down a statement from the table below which most closely describes the *relationship* between the games.

<i>Game A and Game B</i>				
perfect negative relationship	negative relationship	no relationship	positive relationship	perfect positive relationship

<i>Game A and Game C</i>				
perfect negative relationship	negative relationship	no relationship	positive relationship	perfect positive relationship

- (d) What can you tell about the *relationship* between the scores on Game B and the scores on Game C? Write down the statement below which most closely describes the relationship.

Game B and Game C				
perfect negative relationship	negative relationship	no relationship	positive relationship	perfect positive relationship

(KS3/98/Ma/Tier 6-8/P2)

## 8.4 Lines of Best Fit

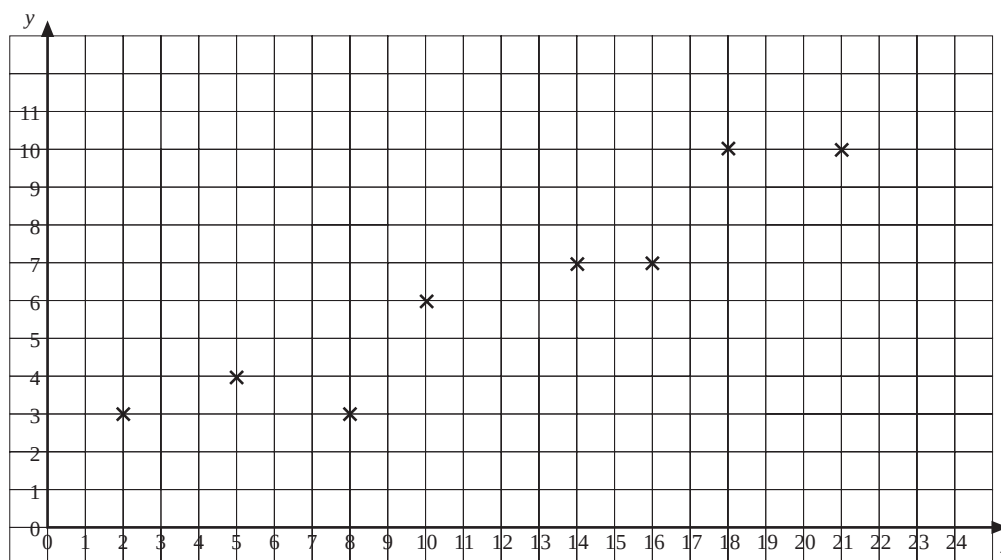
When there reasonable correlation between two variables on a scatter plot, it is possible to draw a *line of best fit*. This line represents the underlying relationship between the two quantities. When drawing a line of best fit the aim is to keep the distances of all the points from the line to a minimum. Sometimes it is helpful to try to keep the number of points above the line the same as the number of points below the line.

Lines of best fit can be used to make predictions. The accuracy and reliability of those predictions will depend on the strength of the correlation between the two variables.



### Example 1

- (a) Draw a line of best fit for the points in the following scatter graph:



- (b) Use the line to predict the value of  $y$  when  $x = 12$ .