Bivariate data, correlation, scatter diagrams and line of best fit

1. Recognise different types of correlation of two sets of data
2. Draw scatter diagrams to show correlation
3. Use line of the best fit to make predictions

Sometimes we need to examine two sets of data to see if there is any relation between them.

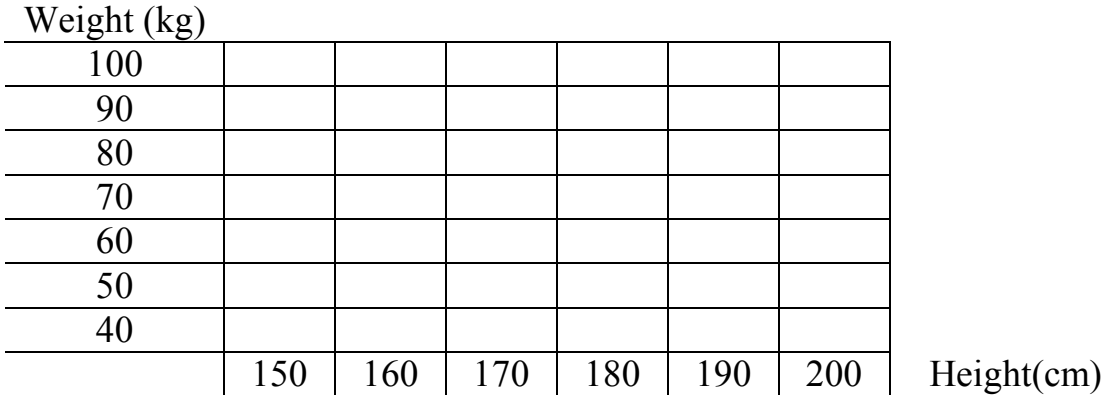
Working Example 1:

Maria has weighed and measured the people in her maths group. The two sets of data are shown in this table. The height and weight are called **variables**. Data like this are called bivariate ('bi' for 'two')

	Alice	Brian	Carl	Davina	Elaine	Gary	Howard
Weight (kg)	49	65	82	60	65	94	88
Height (cm)	156	176	183	153	163	192	180

Draw a scatter diagram of the information.

Considering each pair of values as a different pair of coordinates then the following scatter diagram is produced.

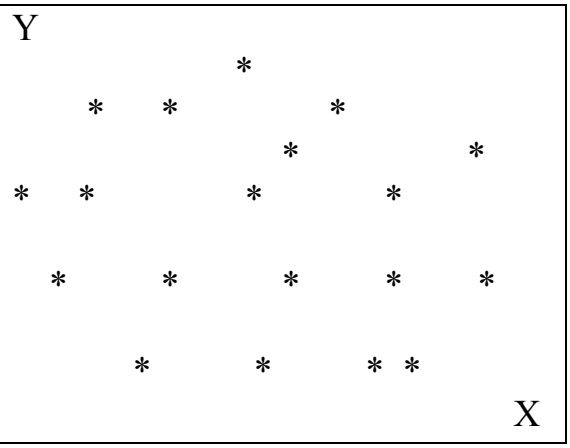


- The points fall in an upward-slopping diagonal band: the taller people are generally heavier. We can say that height and weight are **positively correlated.**

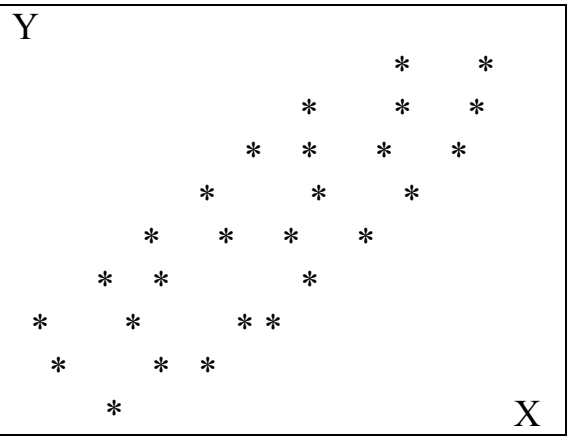
- This scatter diagram shows a **strong correlation** because the points are close to a line.

Scatter graphs are used to show whether there is any relationship or correlation between two sets of data (two variables). You need to know the following description of such relationships or correlation.

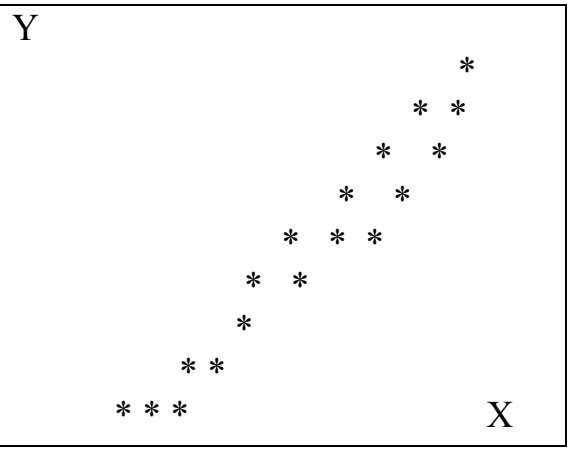
Little or no apparent correlation

The points are scattered randomly over the graph indicating little or no correlation between the two variables.	
What data would you expect to show this type of correlation? • Marks on chemistry exam and marks on art exam • What kinds of fruit people prefer and their shoe sizes	

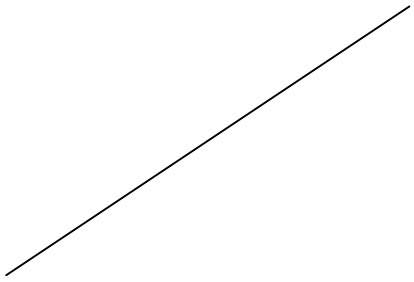
Moderate correlation

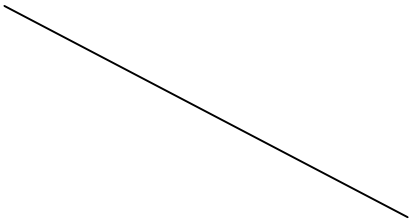
The points lie close to a straight line indicating moderate correlation between the two variables (the closer the points to the straight line then the stronger the correlation).	
What data would you expect to show this type of correlation? • Distance travelled against cost per mile	

Strong correlation

The points lie along a straight line indicating a strong correlation between the two variables.	
What data would you expect to show this type of correlation? • Length against weight of fish • Marks on a Physics exam and marks on a Maths exam • Rain falling and the number of people outdoors	

Correlation can also be described as **positive** or **negative correlation.**

Positive or direct correlation*:* as one quantity increases, the other one is increases (the slope on the graph is upwards)	Y  X
• Power of engine and the maximum speed of a car	

Negative or inverse correlation: as one quantity increases, the other one decreases (a downward slope)	Y  X
• Thickness of insulation and heat loss through the roof • The number of sunny days and the number of people visiting an aquarium at a seaside resort	

Line of best fit

If the points on a scatter diagram lie in a narrow band (strong or moderate correlation) then a line can be drawn to approximate the relationship and this line can be used to predict/ estimate the value of one variable, given the value of the other.

In the most cases a line of best fit can be drawn 'by eye'. Make sure that roughly an equal numbers of points lie above and below the line.

Using ***Example 1*** draw the line of the best fit. Use the line of best fit to estimate the weight of Fatima whose height is 170cm.

For more accurate work the line of best fit should go through mean point of each set of data.

Mean weight = (49 +65+82+60+65+94+88)/7 = 71kg875g = 72kg

Mean height = (166+176+183+153+163+192+180)/7 = 173 cm

Exercise 1

The table lists the weights of ten books and the number of pages in each one.

Number of pages	85	150	100	120	90	140	137	105	115	160
Weight (g)	165	325	200	250	180	285	250	170	230	340

1) Draw a scatter diagram to show the information the table.
2) Describe the correlation between the number of pages in these books and their weights.
3) Draw a line of best fit on the scatter diagram.
4) Use the line to estimate:
   the number of pages in a book of weight 270 g
   the weight of a book with 130 pages

Elena Worsfold