33.1 Tables and Charts

There are many different ways of displaying data, and knowing the best way to display different types makes things much easier when it comes to interpreting it.

Two-Way Tables

Learning Objective — Spec Ref S2:
Display and interpret data in two-way tables.

Prior Knowledge Check:
Be able to write one number as a percentage of another. See p.59.

Two-way tables are used to show the frequencies for **two different variables** — e.g. the hair colour and eye colour of school pupils. The **rows** represent the **categories** for one variable (e.g. 'brown hair', 'blonde hair', etc.) and the **columns** represent the **categories** for the other variable (e.g. 'blue eyes', 'brown eyes', etc.). Each **cell** then shows the number of items in a particular row AND a particular column — e.g. the number of pupils with brown hair AND blue eyes. **Row** and **column totals** show the **total** number of items in each **category** and the bottom-right cell shows the **overall** total.

A complete two-way table can be used to work out the **proportions** of groups with **particular characteristics**. You can find proportions of the **whole group** (e.g. the percentage of pupils with blue eyes and brown hair in the whole school) or proportions within a **category** (e.g. the percentage of blue-eyed pupils that have brown hair). Be careful with which **totals** you use — you need the **overall total** if you're looking at the **whole group**, or the **row/column totals** if you're looking within a **category**.

Example 1

This table shows how students in a class travel to school.

	Walk	Bus	Car	Total
Boys	8	7		19
Girls	6		2	
Total		12		

a) Complete the table.

Add frequencies to find row/column totals. Subtract from row/column totals to find other frequencies.

	Walk	Bus	Car	Total
Boys	8	7	19 – 8 – 7 = **4**	19
Girls	6	12 – 7 = **5**	2	6 + 5 + 2 = **13**
Total	8 + 6 = **14**	12	4 + 2 = **6**	19 + 13 = **32**

b) How many girls take the bus to school?

Read off the value in the 'Girls' row and the 'Bus' column.

5 girls take the bus to school

c) What percentage of students take the bus to school?

Divide the total of the 'Bus' column by the overall total.

$\frac{12}{32} \times 100 = \mathbf{37.5\%}$

Exercise 1

Q1 This two-way table gives information about the colours of the vehicles in a car park.

a) Copy and complete the table.

b) How many motorbikes were blue?

c) What percentage of vehicles (to 3 s.f.) were:
(i) cars? (ii) vans? (iii) red?

	Red	Black	Blue	White	Total
Cars	8	7	4		22
Vans		2	1	10	
Motorbikes	2	1		2	
Total	12		6		

Q2 A group of schoolchildren were asked if they have been ice skating before. Copy the two-way table below, then use the following information to complete it.

PROBLEM SOLVING

- 20 girls were asked in total.
- Half as many boys as girls were asked.
- 18 of the children have been ice skating.
- One-fifth of the girls haven't been ice skating.

	Have been	Haven't been	Total
Boys			
Girls			
Total			

Q3 This two-way table shows the heights in centimetres of 50 people.

	$h < 160$	$160 \leq h < 170$	$170 \leq h < 180$	$180 \leq h < 190$	$190 \leq h$	Total
Women	4		11		0	24
Men		2	6		5	
Total	4	9				50

a) Copy and complete the table.

b) How many of the people are smaller than 170 cm?

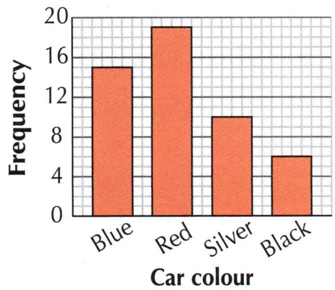
c) What fraction of the women are at least 180 cm tall?

d) Comment on the difference between the heights of the men and the women.

Bar Charts

Learning Objective — Spec Ref S2:
Display and interpret data in bar charts.

Bar charts show the **number** (or frequency) of items in **different categories**. They're used for **discrete data** (see p.406). Each bar represents a different category — the bars **shouldn't touch** because the categories are **distinct**.



Multiple sets of data can be displayed on the **same bar chart** — e.g. data for boys and girls or children and adults.

- **Dual bar charts** have two bars per category — one for each data set.
- **Composite bar charts** have single bars split into different sections for each data set.

You can easily read the **mode** (see Section 32) from a bar chart — it's the category with the greatest frequency, so it's shown by the **tallest bar**.

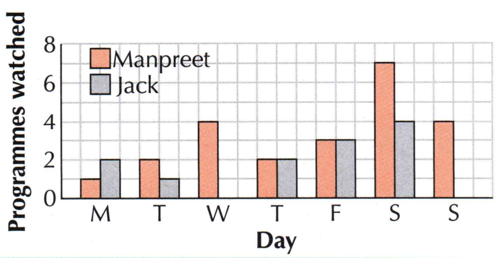
Tip: The bars could be replaced by lines, and drawn either vertically or horizontally.

Example 2

Manpreet and Jack recorded how many TV programmes they watched each day for a week. Their results are shown in the table to the right. Draw a dual bar chart to display this information.

Day	M	T	W	T	F	S	S
No. watched by Manpreet	1	2	4	2	3	7	4
No. watched by Jack	2	1	0	2	3	4	0

1. Each day should have two bars — one for Manpreet and one for Jack.
2. The height of each bar is the frequency.
3. Use different colours for Manpreet and Jack's bars — and include a key to show which is which.

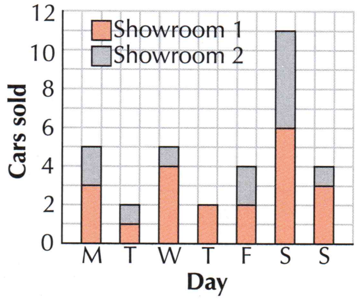


Example 3

The number of cars sold at two car showrooms over a week is shown in the table to the right. Draw a composite bar chart to display this information.

Day	M	T	W	T	F	S	S
No. sold at Showroom 1	3	1	4	2	2	6	3
No. sold at Showroom 2	2	1	1	0	2	5	1

1. Start with Showroom 1 — the height of the orange bar is the frequency for Showroom 1.
2. Add grey bars onto the top of the orange ones. The height of each grey bar should be the frequency for Showroom 2. Make sure you include a key to show which colour is for which showroom.
3. The total height of the composite bar (the orange and grey bars combined) is the total frequency for both showrooms.



Exercise 2

Q1 The eye colour of 50 students is shown in the table.

Eye colour	Blue	Brown	Green	Other
No. of males	8	7	5	4
No. of females	7	12	5	2

a) Draw a dual bar chart to display the data.

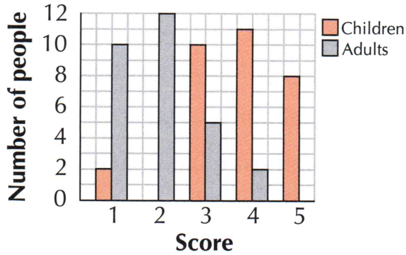
b) Draw a composite bar chart to display the data.

c) Which chart is best for comparing males and females?

d) What is the modal eye colour?

e) Which chart was easier to use to find the mode?

Q2 A group of children and adults were asked to rate a music magazine on a scale from 1 to 5. This dual bar chart shows the results.



a) How many children gave the magazine a score of 1?

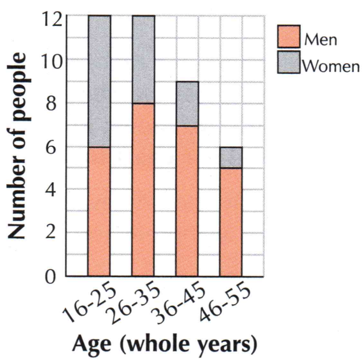
b) How many adults rated the magazine?

c) Which score did no children give?

d) Which score was given by twice as many children as adults?

e) Make one comment about the scores given by the adults compared to those given by the children.

Q3 This composite bar chart shows the ages in whole years of the members of a small gym.



a) How many members are in the age range 26-35 years?

b) How many members does the gym have in total?

c) What is the modal age range for the female members?

d) How many more men aged 26-35 use the gym than women aged 26-35?

e) Which age range has the greatest difference in numbers of men and women?

Pie Charts

Learning Objective — Spec Ref S2:
Display and interpret data in pie charts.

Like bar charts, **pie charts** show how data is divided into categories, but pie charts show the **proportion** in each category, rather than the actual frequency. The sizes of the **angles** of the sectors are **proportional** to the **frequencies**.

To **draw** a pie chart, you need to find the **angle** that represents a **frequency of 1**. To do this, **divide** 360° (the full circle) by the **total frequency**, which you find by **adding up** the frequencies of each category. **Multiply** this value by the frequency of each category to find the **angle** of the **sector**, then use these angles to draw the pie chart.

Tip: The sum of the sector angles should add up to 360°, so use this to check your working.

Example 4

Jake asked everyone in his class to name their favourite colour. The frequency table on the right shows his results. Draw a pie chart to show his results.

Colour	Red	Green	Blue	Pink
Frequency	12	7	5	6

1. Calculate the total frequency — the total number of people in Jake's class.

   Total frequency = 12 + 7 + 5 + 6 = 30

2. Divide 360° by the total frequency to find the number of degrees needed to represent each person.

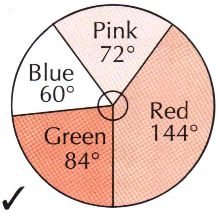
   Each person is represented by 360° ÷ 30 = 12°

3. Multiply each frequency by the number of degrees for one person to find each angle (check that the angles add up to 360°).

   Red: 12 × 12° = **144°**
   Green: 7 × 12° = **84°**
   Blue: 5 × 12° = **60°**
   Pink: 6 × 12° = **72°**

4. Draw the pie chart using the angles you've just calculated to mark out the sectors.

   Check:
   144° + 84° + 60° + 72° = 360° ✓



Exercise 3

Q1 Zofia asked a group of 90 people where they went on holiday last year. Her results are shown in the table. Draw a pie chart to show Zofia's data.

Destination	UK	Europe	USA	Other	Nowhere
Frequency	22	31	8	11	18

Q2 Vicky asked people entering a sports centre what activity they were going to do. 33 were going to play squash, 52 were going to use the gym, 21 were going swimming and 14 had come to play tennis. Draw a pie chart to show this data.

Q3 Pablo surveyed his friends to find out how they travel to school. The table shows his results. Draw a pie chart to represent Pablo's data.

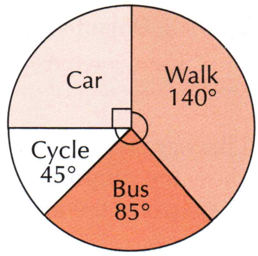
Method of transport	Walking	Bus	Car	Bike
Frequency	36	16	12	16

Q4 Basil used a questionnaire to find out which subject pupils at his school enjoyed the most. His data is shown in the frequency table below. Show this data in a pie chart.

Subject	Maths	Art	PE	English	Science	Other
Frequency	348	297	195	87	108	45

Example 5

A head teacher carried out a survey to find out how pupils travel to school. The pie chart on the right shows the results of the survey.



a) What is the most popular way to travel to school?

This is the sector with the largest angle. **Walking** is the most popular.

b) Which method of transport is twice as common as cycling?

Cycling is represented by a sector with an angle of 45°, so look for a sector with an angle of 2 × 45° = 90° — travelling by car. **Travelling by car** is twice as common as cycling

c) 280 pupils said they walk to school. How many pupils took part in the survey altogether?

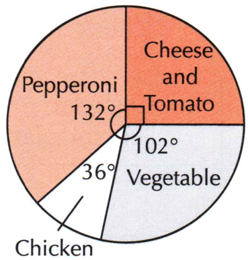
1. Work out how many pupils are represented by 1°. 140° represents 280 pupils. So 1° represents 280 ÷ 140 = 2 pupils.
2. Multiply by 360° to work out how many pupils the whole pie chart represents. This means 360° represents 360 × 2 = **720 pupils**

d) Half the children at another school walk to school. Juan says, "A greater number of pupils at this other school walk to school." Is Juan correct? Explain your answer.

Juan is **not necessarily correct**. Without knowing **how many pupils** are at the other school you can't tell whether more or fewer pupils walk to school — you can **only compare the proportions**.

Exercise 4

Q1 Jennifer asked pupils in her school to name their favourite type of pizza. The pie chart on the right shows the results.



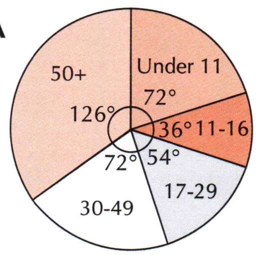
a) Which was the most popular type of pizza?

b) What fraction of the pupils said cheese and tomato was their favourite?

c) Jennifer asked 60 pupils altogether. Calculate the number of pupils who said vegetable was their favourite.

Q2 A librarian carried out a survey of the ages of people using the library. Chart A on the right shows the results.

A

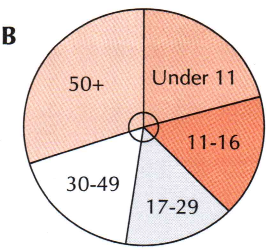


a) There were 18 people aged 17-29 who took part in the survey. How many people took part in the survey altogether?

b) Use your answer to part a) to calculate the number of people in each category.

Leaflets were handed out to persuade more young people to use the library. Another survey was then carried out to find the ages of library users. The results are shown in chart B on the right.

B



c) What fraction of the people in the second survey were aged 11-16?

d) Compare the fraction of people aged 11-16 in the second survey with the fraction of people aged 11-16 in the first survey.

e) Do these pie charts show that there were more people aged 11-16 in the second survey than in the first survey? Explain your answer.

33.2 Stem and Leaf Diagrams

Stem and leaf diagrams are a bit like horizontal bar charts, but the bars are made out of the actual data. They are useful for showing the spread of data visually, whilst still showing each individual data value.

Learning Objective — Spec Ref S2/S4:
Display and interpret data in stem and leaf diagrams.

Prior Knowledge Check:
Be able to find the mode, median, range and interquartile range. See p.416-419.

In **stem and leaf diagrams**, data values are split up into '**stems**' (their first digits(s)) and **leaves** (the remaining digit). So for the value **25**, the stem would be **2** and the leaf would be **5**. The leaves are then **ordered** numerically.
Stem and leaf diagrams always have a **key** — e.g. '2 | 5 means 25'.
Once your data is in a stem and leaf diagram, you can easily find the **mode**, **median**, **range** and **interquartile range**.

Tip: Three figure numbers and decimals can be shown using different keys — e.g. 0 | 3 = 0.3 or 20 | 4 = 204.

Example 1

The marks scored by pupils in a class test are shown here.

56, 52, 82, 65, 76, 82, 57, 63, 69, 73, 58, 81, 73, 52, 73, 71, 67, 59, 63

a) Use this data to draw an ordered stem and leaf diagram.

1. Write down the 'stems' in order in a column — here use the first digit of the marks. The smallest first digit is 5 and the largest is 8, so use 5, 6, 7 and 8.
2. Make a 'leaf' for each data value by writing each second digit next to the correct stem.

5	6	2	7	8	2	9
6	5	3	9	7	3	
7	6	3	3	3	1	
8	2	2	1			

3. Put the leaves in each row in order — from lowest to highest.
4. Remember to add a key. Key: 5 | 2 means 52 marks

5	**2**	**2**	**6**	**7**	**8**	**9**
6	**3**	**3**	**5**	**7**	**9**	
7	**1**	**3**	**3**	**3**	**6**	
8	**1**	**2**	**2**			

b) Use your stem and leaf diagram to find the mode, median and range.

1. Find the mode by looking for the 'leaf' that repeats most often in one of the rows — here, there are three 3s in the third row.
2. There are 19 data values, so the median is the 10th value (p.416).
3. Find the range by subtracting the smallest value from the largest.

Mode = **73 marks**
Median = **67 marks**
Range = 82 – 52 = **30 marks**

Exercise 1

Q1 Draw ordered stem and leaf diagrams for the data sets below, using appropriate keys.

a) 41, 48, 51, 54, 59, 65, 74, 80, 86, 89

b) 36, 41, 15, 39, 41, 12, 15, 27, 17, 31, 24, 26

c) 5.3, 3.1, 4.0, 5.7, 6.0, 5.9, 7.7, 4.4, 3.4

d) 205, 203, 232, 211, 236, 234, 240, 203, 221

Q2 Use this diagram showing 11 pupils' test marks to find:

a) the mode, median, range and interquartile range,

b) the number of pupils who scored between 53 and 63 marks.

5	2	4	5	7	8
6	1	1	8	9	
7	1	7			

Key: 5 | 2 means 52

Back-to-Back Stem and Leaf Diagrams

Learning Objectives — Spec Ref S2/S4:
- Display and interpret data using back-to-back stem and leaf diagrams.
- Compare distributions using back-to-back stem and leaf diagrams.

Prior Knowledge Check: Be able to find the mode, median, range and interquartile range. See p.416-419.

Back-to-back stem and leaf diagrams can be used to display **two sets of data** alongside one other — e.g. they might show data for different genders or age groups. In these diagrams, the stem is in the **centre** and the leaves from the two data sets are placed on **either side** of the stem. This means that one data set has to be read '**backwards**' (the **smaller leaves** are closest to the centre) — so the **key** is really important. It's easy to **compare** the data sets just by looking at the diagram.

Example 2

14 girls and 14 boys completed a puzzle. The times in seconds they took to finish it was recorded. This ordered back-to-back stem and leaf diagram shows the results.

Girls		Boys
	1	4 4 7 9
9 8 8 5 5 3	2	7 9
8 7 4 2 1	3	2 3 8
4 2 2	4	5 6 6 7 9

Key: 2 | 7 for boys means 27
3 | 2 for girls means 23

a) Find the median times for the girls and the boys.

There are 14 data values for each set of data, so the median is between the 7th and 8th values.

Median for girls = $\frac{31+32}{2}$ = **31.5 seconds**

Median for boys = $\frac{32+33}{2}$ = **32.5 seconds**

b) Find the range of times for the girls and the boys.

Subtract the first number from the last.

Range for girls = 44 – 23 = **21 seconds**

Range for boys = 49 – 14 = **35 seconds**

c) Use your answers to make two comparisons between the times for the girls and boys.

Compare the medians and the ranges and interpret what these values show about the times.

The **median** for the boys is **higher** than for the girls, suggesting that boys took **longer on average**.
The **range** for the girls is **smaller** than for the boys, suggesting the girls' times were **more consistent**.

Exercise 2

Q1 Use these two data sets to draw an ordered back-to-back stem and leaf diagram.

18, 48, 38, 29, 41, 28, 33, 24, 12, 37, 32

27, 25, 19, 15, 22, 18, 13, 23, 22, 32, 13

Q2 The heart rates in beats per minute (bpm) of 15 people at rest and after exercise are shown on the right.

a) Find the median of each data set.

b) Find the interquartile range for each data set.

c) What conclusions can you draw from your answers?

At rest		After exercise
8 7 6 4 3 2 2	6	5 8 8 9
9 8 6 3 2 2	7	4 5 7 7 8
4 1	8	5 6 7
	9	1 3 7

Key: 6 | 5 after exercise means 65 bpm
2 | 6 at rest means 62 bpm

Q3 The data on the right shows average daily temperatures (in °C) in Dundee and in London during the same 12-day period.

a) Draw a back-to-back stem and leaf diagram to show the data.

b) By calculating the median and range for each set of data, compare the temperatures in the two places.

Dundee	12	19	6	17	23	4
	3	1	15	5	2	3

London	13	21	12	18	24	9
	4	7	15	12	11	16

33.3 Frequency Polygons

Frequency polygons are a visual representation of how the frequency changes throughout a data set.

Learning Objective — Spec Ref S2:
Display and interpret data in frequency polygons.

Prior Knowledge Check:
Understand grouped frequency tables — see p.422.

Frequency polygons are used to show the data from a **grouped frequency table**. To draw a frequency polygon, you plot the **frequency** on the **vertical axis** against the **midpoint** of the group on the **horizontal axis** — see p.422 for how to find the midpoint of a group. Frequency polygons are used for **continuous** data (see p.406), so the horizontal axis has a **continuous scale**. You can **compare** the **distributions** of two or more data sets by drawing a frequency polygon for each on the **same axes** and comparing their **shapes**.

Example 1

Amin recorded the speeds of the cars that passed outside his house one day. The results are shown in the grouped frequency table on the right. Draw a frequency polygon to represent this information.

Speed (s) in mph	Frequency
$25 \le s < 30$	1
$30 \le s < 35$	5
$35 \le s < 40$	12
$40 \le s < 45$	16
$45 \le s < 50$	9
$50 \le s < 55$	2

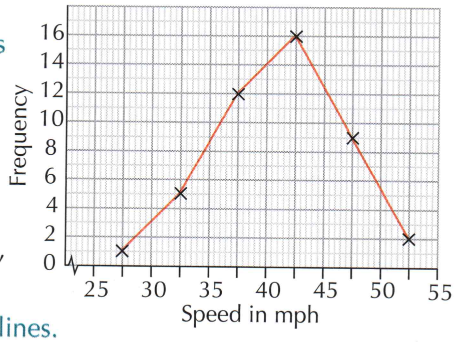
1. Find the midpoint of each class by adding the endpoints and dividing by 2.

   E.g. $\frac{25+30}{2} = 27.5$ mph

2. For each class, plot the frequency at the midpoint.

   (27.5, 1), (32.5, 5), (37.5, 12), (42.5, 16), (47.5, 9), (52.5, 2)

3. Join your points with straight lines.



Exercise 1

Q1 Some students recorded the total floor area of local supermarkets for a project. Their results are shown in the table.

Draw a frequency polygon to show this information.

Floor area (a) in 1000s of square feet	Frequency
$9 \le a < 13$	3
$13 \le a < 14$	7
$14 \le a < 15$	5
$15 \le a < 16$	2

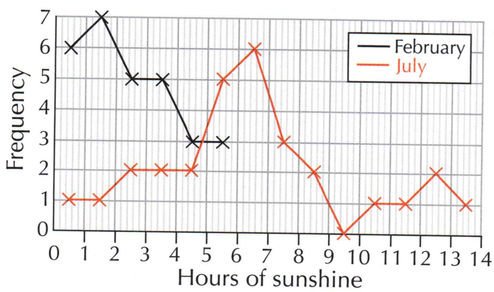
Q2 Emma recorded the number of hours of sunshine on each day from 1st to 29th February 2016 and each day from 1st to 29th July 2016. She used classes with a width of 1 hour. The frequency polygons below show her results.

a) How many days with 4 or more hours of sunshine were there in February?

b) What is the modal class for February?

c) What is the modal class for July?

d) Use your answers to parts b) and c) to compare the daily hours of sunshine in February and July. Compare the spread of the data for the two months.



33.4 Histograms

Histograms may look a lot like bar charts, but they're a bit more complicated to draw and interpret.

Histograms

Learning Objective — Spec Ref S3:
Represent data on a histogram.

Prior Knowledge Check: Understand grouped frequency tables — see p.422.

Histograms are used to show **grouped continuous data** — so the scale on the horizontal axis is **continuous**, and there are **no gaps** between the bars. The vertical axis shows **frequency density** (**not** frequency) — to calculate the frequency density, use the formula:

$$\text{Frequency density} = \frac{\text{Frequency}}{\text{Class width}}$$

This means that the **frequency** of data values is shown by the **area of each bar**, not the height. You can find this by rearranging the formula above:

Area = Frequency = Frequency density × class width

Example 1

Draw a histogram to represent the information about the height, h, of 30 adults shown in the table.

1. Find the class widths for each group.

   E.g. $155 \leq h < 165$: $165 - 155 = 10$
   $165 \leq h < 170$: $170 - 165 = 5$
   ⋮
   $185 \leq h < 200$: $200 - 185 = 15$

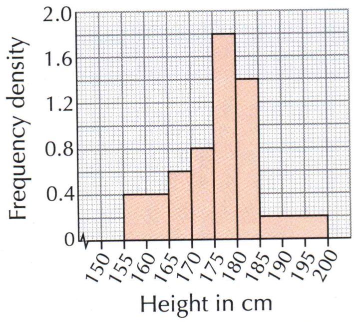
2. Add a 'frequency density' column to the table and use the formula to work out the value for each class.

Height (h) in cm	Frequency	Class Width	Frequency Density
$155 \leq h < 165$	4	10	$4 \div 10 = \mathbf{0.4}$
$165 \leq h < 170$	3	5	$3 \div 5 = \mathbf{0.6}$
$170 \leq h < 175$	4	5	$4 \div 5 = \mathbf{0.8}$
$175 \leq h < 180$	9	5	$9 \div 5 = \mathbf{1.8}$
$180 \leq h < 185$	7	5	$7 \div 5 = \mathbf{1.4}$
$185 \leq h < 200$	3	15	$3 \div 15 = \mathbf{0.2}$

3. Draw axes with a continuous scale for height along the bottom and frequency density up the side.

4. Draw a bar for each class with a height equal to the frequency density. Then check that the area equals the frequency.

   E.g. area of the first bar:
   frequency density × class width = $0.4 \times 10 = 4$ ✓



Exercise 1

Q1 Copy and complete the grouped frequency tables below.

a)

Height (h) in cm	Frequency	Frequency Density
$0 < h \leq 5$	4	
$5 < h \leq 10$	6	
$10 < h \leq 15$	3	
$15 < h \leq 20$	2	

b)

Height (h) in cm	Frequency	Frequency Density
$10 < h \leq 20$	5	
$20 < h \leq 25$	15	
$25 < h \leq 30$	12	
$30 < h \leq 40$	8	

Q2 The tables below show some information about the volume of tea drunk each day by a group of people. Copy and complete each table and use it to draw a histogram to represent the information.

a)

Volume (v) in ml	Frequency	Frequency Density
$0 \le v < 500$	50	
$500 \le v < 1000$	75	
$1000 \le v < 1500$	70	
$1500 \le v < 2000$	55	

b)

Volume (v) in ml	Frequency	Frequency Density
$0 \le v < 300$	30	
$300 \le v < 600$	15	
$600 \le v < 900$	24	
$900 \le v < 1500$	42	

Q3 The table on the right shows some information about the masses (m) in kg of 22 pigs on a free range farm. Draw a histogram to represent this information.

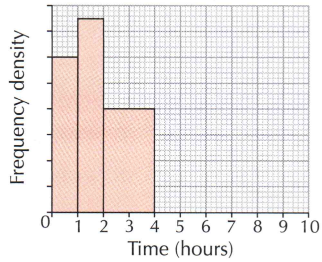
Mass (m) in kg	Frequency
$10 \le m < 20$	5
$20 \le m < 30$	7
$30 \le m < 50$	10

Q4 This incomplete table and histogram show information about the length of time (in hours) that some people spend watching the TV programme 'Celebrity Ironing Challenge' each week.

PROBLEM SOLVING

a) Copy the histogram and use the information given to label the vertical axis.

Time (t) in hours	Frequency
$0 \le t < 1$	
$1 \le t < 2$	
$2 \le t < 4$	16
$4 \le t < 6$	14
$6 \le t < 10$	12



b) Use the information given to fill in the gaps in the table and histogram.

Interpreting Histograms

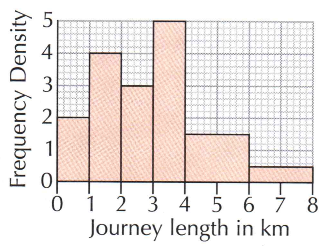
Learning Objective — Spec Ref S3/S4:
Use histograms to estimate and interpret data.

Prior Knowledge Check: Be able to find averages for grouped data — see p.422.

You can use histograms to **estimate** things about the data — you have to **assume** that the data is **spread evenly** through each interval. For example, the **number of data values** in a particular interval can be found using the **areas of the bars** — so if you wanted to find the number of values between 2 and 2.5 but the class was $2 \le x < 3$, you just divide the **frequency** (i.e. **area**) of the whole bar by 2. You can estimate values for the **mean** or **median** by working out the **frequency** of each class (class width × frequency density) and following the methods shown for grouped frequency tables on p.422.

Example 2

Some students were asked the length of their journey to university. The histogram on the right shows the information.



a) Estimate the number of students with a journey of 1.5 km or less.

1. Work out the frequency of journeys between 0 and 1 km by finding the area of the first bar.
   Frequency = class width × frequency density
   = 1 × 2 = 2

2. Work out the frequency of journeys between 1 and 2 km and divide by 2 to find the frequency of journeys between 1 and 1.5 km.
   Frequency of $1 < j \le 2$ group
   = 1 × 4 = 4
   Frequency of $1 < j \le 1.5$
   ≈ 4 ÷ 2 = 2

3. Add the values for $0 < j \le 1$ and $1 < j \le 1.5$ together.
   2 + 2 = **4 students**

Tip: You can only estimate here as you don't know the actual data values.

b) Estimate the mean journey length for the students.

1. Draw a grouped frequency table using the classes shown on the histogram.
2. Work out the frequency (class width × frequency density) and midpoint for each class, then multiply them together.
3. Estimate the mean by dividing the total of the 'midpoint × frequency' column by the total frequency.

Journey (j) in km	Frequency	Midpoint	Midpoint × frequency
$0 < j \le 1$	$2 \times 1 = 2$	0.5	1
$1 < j \le 2$	$4 \times 1 = 4$	1.5	6
$2 < j \le 3$	$3 \times 1 = 3$	2.5	7.5
$3 < j \le 4$	$5 \times 1 = 5$	3.5	17.5
$4 < j \le 6$	$1.5 \times 2 = 3$	5	15
$6 < j \le 8$	$0.5 \times 2 = 1$	7	7
Totals	18		54

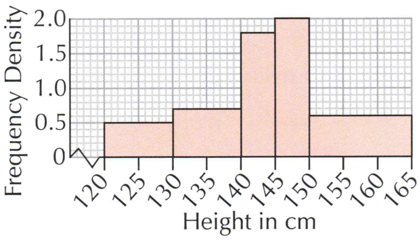
Estimated mean = 54 ÷ 18
= **3 km**

Tip: You can't tell from the histogram if the inequality signs should be < or ≤ so the classes could be e.g. $0 \le j < 1$. You just need to make sure there are no overlaps (e.g. you couldn't have $0 < j \le 1$ and $1 \le j \le 2$ since 1 is covered twice).

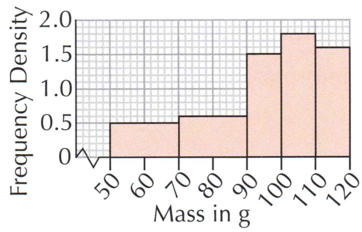
Exercise 2

Q1 For each of the histograms below, work out the frequency represented by each bar.

a)



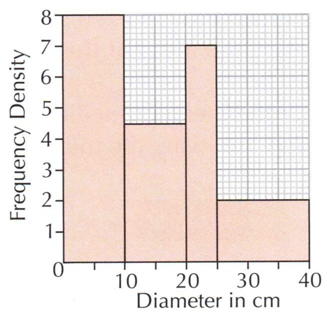
b)



Q2 For the histogram in Q1a) above, estimate the number of data values in the following intervals.

a) 150-155 cm b) 150-160 cm c) 145-160 cm d) 120-126 cm

Q3 The histogram on the right represents the diameters in centimetres of all the cakes made at a bakery one day.



a) Estimate the number of cakes with a diameter of less than 16 cm.

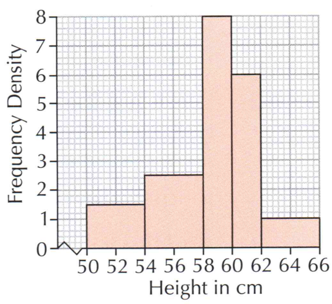
b) Estimate the number of cakes with a diameter in the range 20-30 cm.

c) Work out the total number of cakes made.

d) Calculate an estimate of the mean diameter of the cakes.

e) Explain what assumption you have made for part d) and suggest a problem with this assumption.

Q4 The histogram on the right represents the heights in centimetres of all the penguins at a wildlife park.



a) Estimate the number of penguins with a height of less than 56 cm.

b) Work out the total number of penguins at the wildlife park.

c) 24 of the penguins have a height of less than H cm. Estimate the value of H. PROBLEM SOLVING

d) Use your answers to b) and c) to write down an estimate of the median height. PROBLEM SOLVING

33.5 Cumulative Frequency Diagrams

Cumulative frequency means adding up as you go along — it is the total frequency 'so far' and increases as you go through a data set.

Learning Objectives — Spec Ref S3/S4:

- Display and interpret data on cumulative frequency diagrams.
- Use them to estimate the median, quartiles and the interquartile range.

Prior Knowledge Check: Be able to find the median, quartiles and the interquartile range — see p.416 and p.419.

Cumulative frequency is the **running total** of frequencies for a set of data. You find the cumulative frequency by **adding** the frequency of a class to the **sum** of the frequencies of **all the classes** that came before it. For example, if the frequencies were 1, 5, 4, 9..., the cumulative frequencies would be **1**, 1 + 5 = **6**, 6 + 4 = **10**, 10 + 9 = **19**... You can add an **extra column** to your **frequency table** to keep track of the cumulative frequency — the **final entry** in the cumulative frequency column will be the **total frequency** for all the data.

To draw a **cumulative frequency diagram** for grouped data, the cumulative frequency goes on the **vertical axis** and the horizontal axis has a continuous scale. Plot the points at the **upper limit** of each class (e.g. if the class $5 \le x < 10$ had a cumulative frequency of 15, you'd plot the point (10, 15)). Then draw an **S-shaped curve** (or line) through the points. There are two types of cumulative frequency diagrams:

- **Cumulative frequency curves** — where the points are joined with a smooth curve.
- **Cumulative frequency polygons** — where the points are joined with straight lines instead of a curve.

You can use cumulative frequency diagrams to find the **median** and **interquartile range** of a data set. To find the **median, divide** the **total frequency** by 2, draw a horizontal line at this point and read off from the **horizontal axis**. To find the **lower** and **upper quartiles**, divide the total frequency by 4 (and multiply by 3 for the upper quartile), draw horizontal lines at these points and read off from the horizontal axis. You can then calculate the **interquartile range** by **subtracting** the lower quartile from the upper quartile.

Tip: For grouped data, you can only estimate the median and interquartile range because you don't know the actual data values.

Example 1

The table below shows the heights of a set of plants.

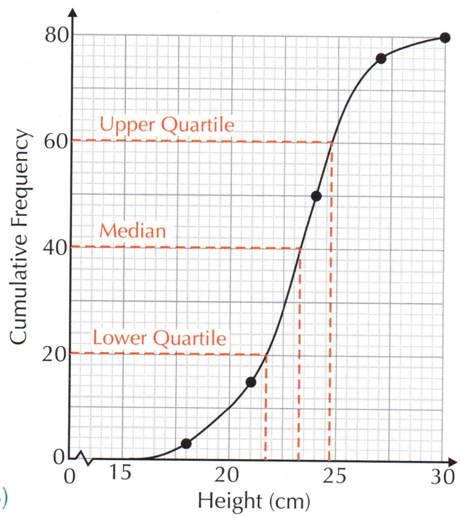
a) Complete the cumulative frequency column for the table.

Just add up the frequencies, working down the table.

Height, h (cm)	Frequency	Cumulative Frequency
$15 < h \le 18$	3	**3**
$18 < h \le 21$	12	3 + 12 = **15**
$21 < h \le 24$	35	15 + 35 = **50**
$24 < h \le 27$	26	50 + 26 = **76**
$27 < h \le 30$	4	76 + 4 = **80**

b) (i) Draw a cumulative frequency diagram for the data.

Cumulative frequency goes on the vertical axis and you need a continuous scale between 15 and 30 on the horizontal axis. Plot the cumulative frequency for each group against the upper limit of the group — so you want to plot the points (18, 3), (21, 15), (24, 50), (27, 76) and (30, 80). Then join the points with a smooth curve.



(ii) Estimate the median and interquartile range for the data.

1. Median — read off the plant height that corresponds to a cumulative frequency of 40 (half of 80).

Median = **23.25 cm**

2. Interquartile range (IQR) — read off the plant height that corresponds to 60 (= ¾ of 80) for the upper quartile and 20 (= ¼ of 80) for the lower quartile. Then find the difference.

Upper quartile = 24.75
Lower quartile = 21.75
IQR = 24.75 – 21.75 = **3 cm**

Exercise 1

Q1 The cumulative frequency graph below shows the monthly earnings for a group of 16-year-olds.

a) How many 16-year-olds took part in the survey?

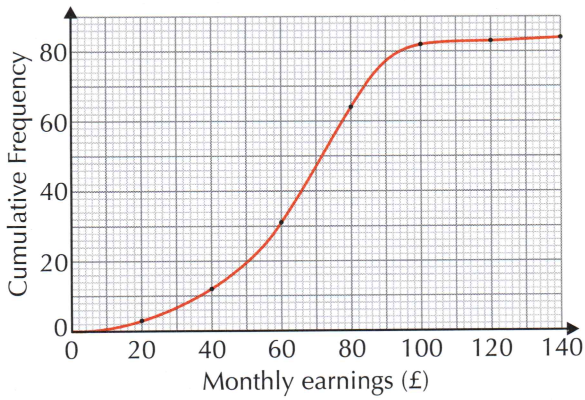
b) Estimate how many earned less than £20.

c) Estimate how many earned less than £80.

d) Estimate how many earned between £40 and £100.

e) Estimate the median earnings for a 16-year-old in the survey.

f) Estimate the lower quartile, the upper quartile and the interquartile range for this data.



Q2 This table shows Year 11 marks in a Maths test.

a) Copy the table and complete the cumulative frequency column.

b) Draw a cumulative frequency curve for the marks.

c) Estimate the median mark.

d) Estimate the upper and lower quartiles and interquartile range for the data.

e) Pupils who achieved fewer than 45 marks have to resit the test. Estimate how many pupils will resit.

f) Pupils who achieved more than 55 marks will sit the higher tier exam. Estimate how many will be entered for the higher tier. PROBLEM SOLVING

Marks, m	Frequency	Cumulative Frequency
$0 < m \leq 10$	0	
$10 < m \leq 20$	2	
$20 < m \leq 30$	4	
$30 < m \leq 40$	5	
$40 < m \leq 50$	19	
$50 < m \leq 60$	33	
$60 < m \leq 70$	43	
$70 < m \leq 80$	10	
$80 < m \leq 90$	3	
$90 < m \leq 100$	1	

Q3 Some students were asked to pour out a sample of sand that they estimated would have a mass of 25 g. The table shows a summary of the actual masses of their samples.

a) Draw a cumulative frequency curve for this data.

b) Estimate the median and interquartile range.

c) Estimate how many students' samples were within 10 g of the median.

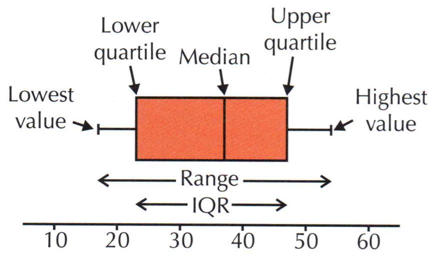
d) Comment on how good the students' estimates were.

Mass, m, in g	Frequency
$m \leq 5$	2
$5 < m \leq 10$	3
$10 < m \leq 15$	4
$15 < m \leq 20$	7
$20 < m \leq 25$	25
$25 < m \leq 30$	51
$30 < m \leq 35$	31
$35 < m \leq 40$	9
$40 < m \leq 45$	5
$45 < m \leq 50$	3

Box Plots

Learning Objective — Spec Ref S4:
Use box plots to interpret and compare distributions.

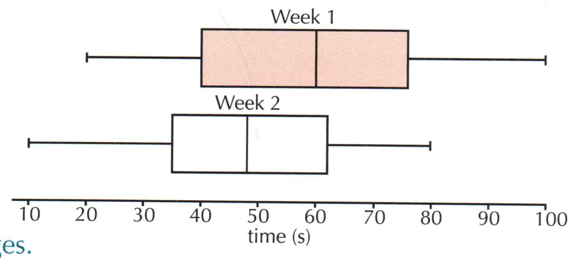
Box plots provide a useful summary of the **distribution of data** in a data set. The **lower quartile**, **median** and **upper quartile** are shown as **vertical lines**, with the lower and upper quartiles forming the **ends** of the **box** — so the **width** of the box is the **interquartile range**. The **maximum** and **minimum** values are also plotted, and joined to the box with **horizontal lines** — you can use these to find the **range**.



By plotting two box plots on the **same scale**, you can easily **compare** the two distributions.

Example 2

The box plots on the right show the distribution of times taken to answer calls at a call centre over two weeks. The call centre manager says, 'The times taken to answer calls in Week 1 were greater but more consistent than those in Week 2'. Do you agree? Use the data to explain your answer.



Compare the medians, quartiles and interquartile ranges.

Week 1: $Q_1 = 40$, Median = 60, $Q_3 = 76$, IQR = 36
Week 2: $Q_1 = 35$, Median = 48, $Q_3 = 62$, IQR = 27

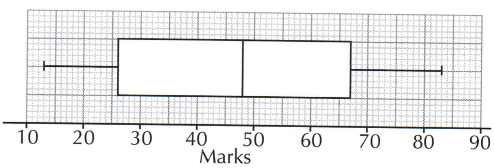
All values are greater for Week 1 than the equivalent values for Week 2, so the data supports the first part of the statement. However, the **interquartile range for Week 2 is smaller** than for Week 1, so the times taken to answer calls were more consistent during Week 2.

Tip: You could use the minimum/maximum values and the range instead of the IQR — in this case it'd lead you to the same conclusions.

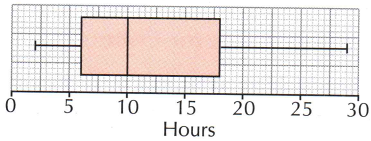
Exercise 2

Q1 For each of the following box plots, find:

a)



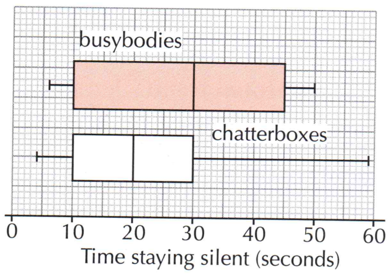
b)



(i) the median,
(ii) the upper quartile,
(iii) the lower quartile,
(iv) the interquartile range,
(v) the lowest value,
(vi) the highest value.

Q2 The box plots on the right show the distributions of the times a group of busybodies and a group of chatterboxes were able to stay silent.

A busybody says, 'The busybodies were definitely better at staying silent than the chatterboxes'. Do you agree? Use the data to support your answer.



33.6 Time Series

Some line graphs can be used to show how things change over time — these are known as time series graphs. They're useful for spotting repeating patterns or trends in data.

Learning Objective — Spec Ref S2:
Display and interpret time series on a line graph.

Time series graphs are used to display data collected at **regular intervals** over time — e.g. every day for a week, every month for a year or over several years. As time is a **continuous variable**, time series are plotted on **line graphs** and the plotted coordinates are joined with **straight lines**.

You can use time series to find **trends** in the data over time. Look for:

- **Seasonality** — a basic pattern that is **repeated** regularly over time. For example, the average monthly temperatures will follow a similar pattern year on year, or quarterly sales may follow the same pattern over time. The time taken for a pattern to repeat itself (measured from **peak-to-peak** or **trough-to-trough**) is called the **period**.
- An **overall trend** — where the data values **generally** get **smaller** or **larger** over time (ignoring seasonal patterns). For example, the price of a weekly shop may steadily increase or decrease. **Trend lines** can be added to times series to better **illustrate** overall trends.

Tip: Seasonality doesn't have to match the actual seasons — e.g. tide levels show seasonality over the course of a day.

You can plot **multiple** time series on the same set of axes to **compare** how different data sets change over time — this is called a **comparative time series** graph.

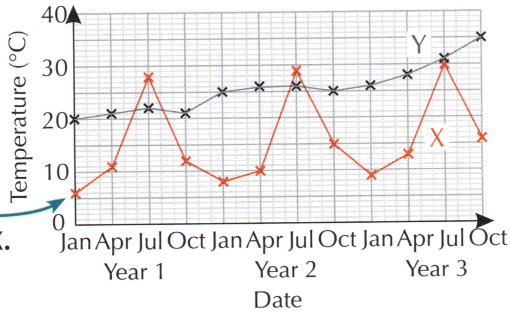
Example 1

The temperature in two countries is recorded once every three months, on the first day of January, April, July and October, for three years. The data for Country X is recorded in the table and the data for Country Y is shown on the time series graph below.

Date	Year 1				Year 2				Year 3			
	Jan	Apr	Jul	Oct	Jan	Apr	Jul	Oct	Jan	Apr	Jul	Oct
Temp. (°C)	6	11	28	12	8	10	29	15	9	13	30	16

a) Draw the time series for Country X on the same axes.

1. Plot the points on the axes. The date is plotted on the horizontal axis and the temperature is plotted on the vertical axis.
2. Join the plotted points with straight lines and label the line.



b) Comment on the trend of the temperatures in Country X.

The time series for Country X shows seasonality.

There is a **seasonal trend**. Temperatures are **coldest in January** and **warmest in July** every year.

c) Comment on the trend of the temperatures in Country Y.

There is an overall trend — the temperatures are generally increasing over time (but there is no seasonal trend).

Temperatures in Country Y are steadily **increasing** with time.

Exercise 1

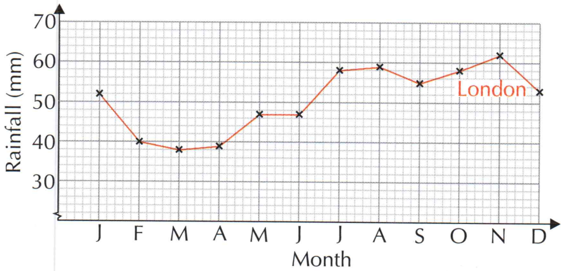
Q1 The table below shows the fastest time in seconds a professional athlete took to run 100 m every day for two weeks leading up to a major competition.

Day	1	2	3	4	5	6	7
Time in seconds	10.8	10.5	10.5	10.4	10.5	10.6	10.5

Day	8	9	10	11	12	13	14
Time in seconds	10.3	10.3	10.2	10.4	10.2	10.1	10.1

Draw a time series graph to show this data and describe the overall trend in the athlete's times.

Q2 The graph shows the average rainfall in London. The average rainfall for Glasgow is shown below.



Month	Jan	Feb	Mar	Apr	May	Jun
Rainfall (mm)	111	85	69	67	63	70

Month	Jul	Aug	Sep	Oct	Nov	Dec
Rainfall (mm)	97	93	102	119	106	127

a) Copy the time series graph for London and draw on the data for Glasgow.

b) Write two sentences to compare the rainfall in the two cities.

c) Find the range of each set of data. PROBLEM SOLVING

Q3 A shop records its sales (in £) of two different brands over an 8-week period in the table below.

Week	1	2	3	4	5	6	7	8
'Impact' brand	520	365	815	960	985	1245	1505	1820
'On Trend' brand	840	795	830	925	960	875	905	965

Draw a comparative time series graph of the data and comment on the sales trends of the two brands.

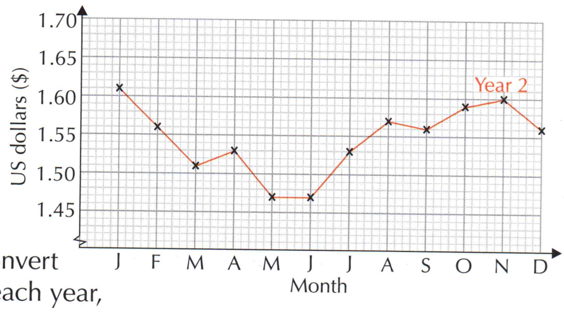
Q4 A bank converts pounds sterling (£) into US dollars ($). The table below shows the highest conversion rate from £ to $ offered each month in Year 1.
The graph shows the highest conversion rate offered each month for Year 2.

Month (Year 1)	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
US dollars ($)	1.45	1.44	1.42	1.47	1.54	1.63	1.64	1.65	1.63	1.62	1.66	1.62

a) Copy the original graph, and add a line showing the data from the table.

b) What was the highest value of £1 in dollars in Year 2?

c) What was the highest value of £1 in dollars in Year 1? When did it occur?



d) A businesswoman always uses this bank to convert her money. She travels to New York in June each year, and changes £500 into dollars each time. Assuming she gets the highest possible rate for June each year, how many more dollars does she get for £500 in Year 1 than in Year 2?

33.7 Scatter Graphs

Scatter graphs are used to show how closely two variables are related to one another — this is known as correlation. A line of best fit can be used to show correlation and estimate values of either variable.

Correlation

Learning Objectives — Spec Ref S6:
- Draw and interpret scatter graphs.
- Recognise and describe correlation.

A **scatter graph** shows **two variables** plotted against each other, e.g. height and weight or temperature and BBQ sales. To **draw** a scatter graph, first decide **which variable** should go on **which axis** — the one that you think **depends** on the other should go on the **vertical** axis. Then **plot** your data as points (x, y), where x is the variable on the horizontal axis and y is the variable on the vertical axis.

If two variables are **related** to each other then they are **correlated**.
Variables can be **positively correlated**, **negatively correlated** or **not correlated** at all.

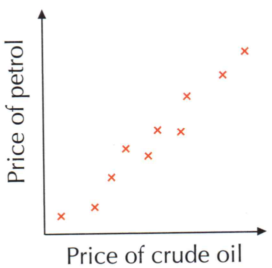
- **Positive correlation** means that both variables **increase and decrease together** — e.g. the number of ice creams sold and the average daily temperature are likely to be positively correlated, as people buy more ice creams when the weather is hot and fewer ice creams when it is cold. The **points** on the scatter graph will look like a line sloping **upward** from left to right.
- **Negative correlation** means that as one variable **increases**, the other **decreases** — e.g. the number of woolly hats sold and the average daily temperature are likely to be negatively correlated, as people buy fewer woolly hats when the weather is hot and more woolly hats when it is cold. The **points** on the graph will look like a line sloping **downward** from left to right.
- **No correlation** means that there is **no linear relationship** between the variables — e.g. the number of newspapers sold is unlikely to change with the average daily temperature, as people buy newspapers regardless of the temperature. The **points** on the graph will look **randomly scattered**.

If two variables are correlated it **doesn't necessarily** mean that one **causes** the other. There could be a **third factor** affecting both, or it could just be a **coincidence**. For example, the number of people wearing **shorts** and the number of **ice creams** sold are likely to be **positively correlated** — but this doesn't mean that people wearing shorts **cause** ice creams to be sold. Here, both variables are affected by the **weather**.

Example 1

Describe the relationship shown by each of the graphs.

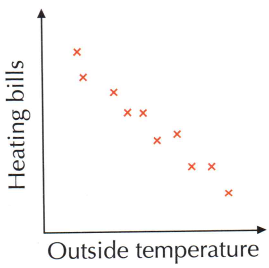
a)



As the price of crude oil increases, the price of petrol also increases, so this is...

Positive correlation

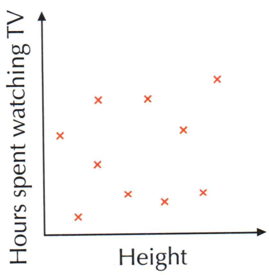
b)



As the temperature outside increases, heating bills decrease, so this is...

Negative correlation

c)



The height of the viewer seems to have no connection to the hours spent watching TV, so this is...

No correlation

Exercise 1

Q1 The outside temperature and the number of ice creams sold in a cafe were recorded for six days. The results are shown in the table on the right.

Temp (°C)	28	25	26	21	23	29
Ice creams sold	30	22	27	5	13	33

a) Use the data from the table to plot a scatter graph.

b) Describe the correlation between the outside temperature and the number of ice creams sold.

Q2 Ten children of different ages were asked how many baby teeth they still had.

Age (years)	5	6	8	7	9	7	10	6	8	9
Baby teeth	20	17	11	15	7	17	5	19	13	8

a) Use the data to plot a scatter graph.

b) Describe the relationship between the age of the children and the number of baby teeth they have.

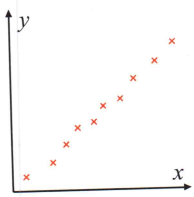
You can describe the **strength** of the correlation as well.
The **closer** the points are to forming a **straight line**, the **stronger** the correlation.

- If most of your points are close to a **straight line**, then you have **strong correlation.**
- If your points are spread **loosely around** a straight line, then you have **moderate correlation.**
- If your points **don't line up** nicely but you can still see that there is a **relationship** between the two variables, then you have **weak correlation.**

Example 2

Describe the strength and type of correlation shown by each of the scatter graphs.

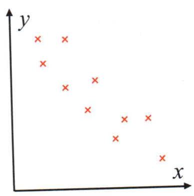
a)



The points form an upward slope fairly close to a straight line, so this is...

Strong positive correlation

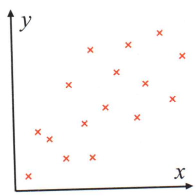
b)



The points form a downward slope loosely around a straight line, so this is...

Moderate negative correlation

c)



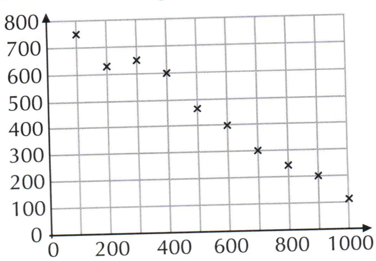
The points form an upward slope, but do not lie close to a straight line, so this is...

Weak positive correlation

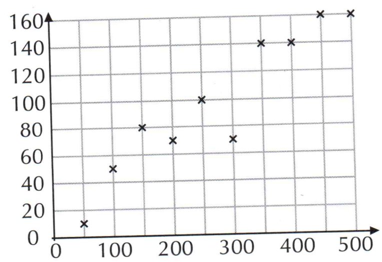
Exercise 2

Q1 Describe the strength and type of correlation shown by each of the scatter graphs below.

a)



b)



Q2 Jeremy measured the height and shoe size of 10 people.

Height (cm)	165	159	173	186	176	172	181	169	179	194
Shoe size	6	5	8	9	8.5	7	8	6	8	11

a) Use his data to plot a scatter graph of shoe size (on the y-axis) against height in cm (on the x-axis).

b) Is there evidence to show that height and shoe size are correlated? Explain your answer.

Lines of Best Fit

Learning Objectives — Spec Ref S6:

- Draw a line of best fit and use it to estimate and predict data values.
- Be able to recognise outliers.

If two variables are correlated, then you can draw a **line of best fit** on their scatter graph. This is a **straight line** that passes through the **middle of the points** with a roughly **equal number** on either side.

Outliers are points that **don't fit** the general pattern of the rest of the data. They can **move** your line of best fit away from other values, so are usually **ignored** when drawing your line. Outliers can sometimes be caused by **errors** in the data, but not always — they can just be **unusually high** or **low values**.

You can use a line of best fit to **predict values** for one variable when you know the value of the other. All you have to do is **draw a line** from the value you're given to the line of best fit, then **read off** the value for the variable on the other axis. Using your line to predict values **within** the range of data you have is known as **interpolation**, and should be **fairly reliable**. Using your line to predict values **outside** the range of the data is known as **extrapolation** and can be **unreliable** because you don't know that the pattern continues outside the data range.

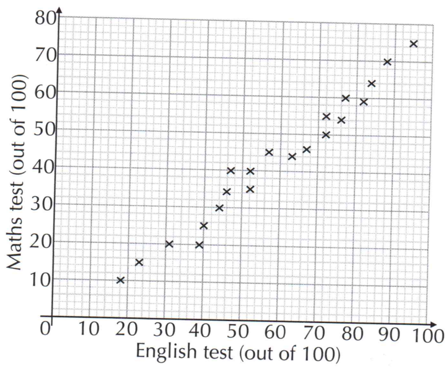
Example 3

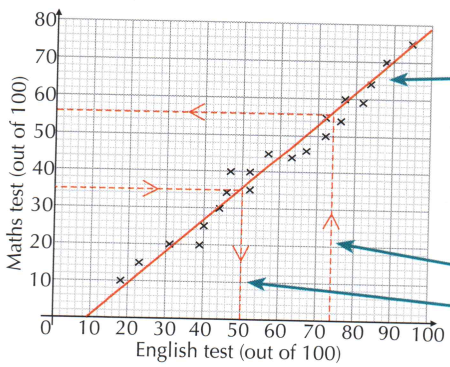
The scatter graph shows pupils' marks on a Maths test plotted against their marks on an English test.

a) Draw a line of best fit on the graph.

b) Jimmy was ill on the day of the Maths test. If he scored 74 in his English test, predict what his Maths mark would have been.

c) Elena was ill on the day of the English test. If she scored 35 on her Maths test, predict what her English result would have been.





a) Draw a straight line through the middle of the points — there should be about the same number of points on either side of the line.

You can now use the line of best fit to predict Jimmy's and Elena's results:

b) Predicted Maths mark for Jimmy = **56**

c) Predicted English mark for Elena = **50**

Exercise 3

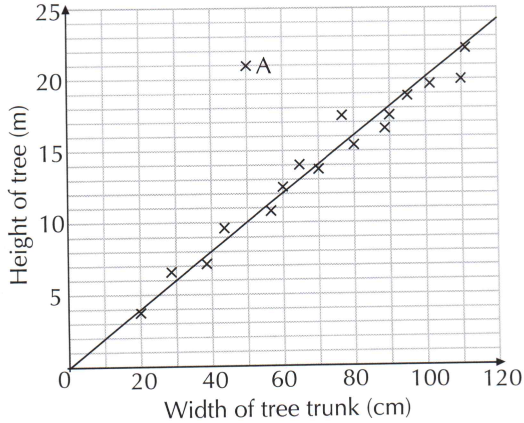
Q1 The graph on the right shows the height of a number of trees plotted against the width of their trunks.

a) Describe the strength and type of correlation between the width of the trunks and the height of the trees.

b) Point A does not fit the correlation pattern shown by the rest of the data. Suggest a reason for this.

c) Use the graph to predict the width of a tree's trunk if it is 13 m tall.

d) A tree has grown into power lines, meaning that measuring its height would be dangerous. If the trunk is 100 cm wide, predict the tree's height.



Q2 The outside midday temperature and the number of drinks sold by two vending machines were recorded over a 10-day period. The results are shown in these tables.

Temperature (°C)	14	29	23	19	22	31	33	18	27	21
Drinks sold — Machine 1	6	24	16	13	15	28	31	20	22	14
Drinks sold — Machine 2	7	25	18	15	17	32	35	14	24	17

For each machine:

a) Draw a scatter graph of drinks sold against temperature.

b) Draw a line of best fit for the data.

c) Circle any outliers and suggest a reason for them.

d) Predict the number of drinks that would be sold if the outside midday temperature was 25 °C.

e) Explain why it might not be appropriate to use your line of best fit to estimate the number of drinks sold if the outside midday temperature was 3 °C.

Lucas says, "An increase in temperature causes more vending machine drinks to be sold."

f) Do you agree with Lucas's statement? Explain your answer.

Q3 Craig measures the leg length of 10 members of a running club. He then times how long it takes each member to run 100 m. His results are shown below.

Club member	1	2	3	4	5	6	7	8	9	10
Leg length (in cm)	60	65	75	90	80	69	96	76.5	85	66
Time taken to run 100 m (in s)	16.9	12.8	15.6	13.5	14.3	15.4	13.0	14.8	14.4	16.3

a) Plot a scatter graph of the time taken against leg length.

b) Describe the relationship shown by the scatter graph.

c) Circle any outliers and draw a line of best fit for the data.

d) Predict how long it would take for a club member with a leg length of 87 cm to run 100 m.

e) (i) Estimate how long it would take a club member with a leg length of 100 cm to run 100 m.

(ii) Is your answer to part (i) a reliable estimate? Explain your answer.

33.8 Appropriate Representation of Data

Don't automatically trust statistical diagrams — they can be misleading...

Learning Objectives — Spec Ref S2:
- Choose an appropriate diagram for a data set.
- Understand how statistical diagrams can be misleading.

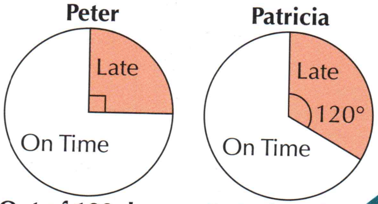
When **choosing** which diagram to use, you need to think about which one will best **represent** your data — for example, although you could use a pie chart to represent continuous grouped data, it would be much more appropriate to show this type of data on a histogram or cumulative frequency diagram. Using the **wrong type of diagram** can make it **difficult** or **impossible** to **interpret** the data and draw **conclusions**.

Data can also be **deliberately** misrepresented to guide you towards a **misleading conclusion**, or to **hide raw data** that presents a negative view. For example, a company could **adjust the scale** on a line graph to make it look like their sales are increasing by more than they actually are. They could also **hide** raw data by using a **pie chart** to show proportions, rather than actual values — for example, if a company had 50 negative reviews out of 1000, using a pie chart to display this data would make the amount of negative reviews seem very small, when there are actually quite a lot.

Example 1

Criticise the diagrams below.

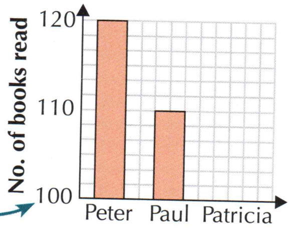
Peter **Patricia**



Out of 100 days **Out of 30 days**

Remember that pie charts show proportion — not actual data values. Here, notice the total number of days is different.

Look at the scale on bar charts to make sure that nothing has been cut off.

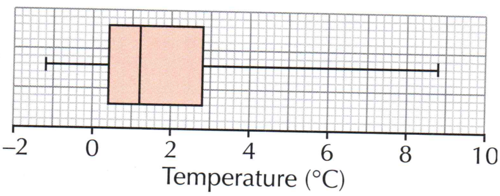


It looks like Patricia is late more than Peter but it's only true that the **proportion** of days she was late is higher — not the actual number of days.

It looks like Paul has read half as many books as Peter, and that Patricia has read none — but that's just because the scale **starts at 100**.

Exercise 1

Q1 This box plot shows information about the temperature inside a fridge, measured at hourly intervals. Explain whether you think a box plot is a good diagram to use for this data. If not, suggest an alternative diagram.



Score (s)	Frequency
$0 < s \leq 20$	13
$20 < s \leq 30$	35
$30 < s \leq 40$	15
$40 < s \leq 50$	3
$50 < s \leq 55$	20
$55 < s \leq 60$	35

Q2 The grouped frequency table to the left shows the scores hit per dart by players during a darts tournament. Would a pie chart be a good way to represent this data? Explain your answer. If not, suggest an alternative diagram.

Review Exercise

Q1 Ten people competed in a quiz. The scores for the first two rounds are shown in the table on the right.

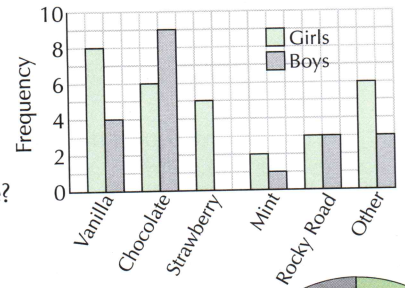
Player	A	B	C	D	E	F	G	H	I	J
Round 1	12	19	6	11	16	15	18	13	12	8
Round 2	9	16	1	8	15	11	13	10	7	4

a) Using this raw data, draw and complete a two-way table to show the frequencies of scores in the following categories for each round: 0-3, 4-7, 8-11, 12-15 and 16-20 points.

b) What percentage of people scored 12 or more in round 1?

c) The questions in one of the rounds were easier than in the other round. Use your two-way table to suggest which round was easier and explain your answer.

Q2 The children at a youth club were asked to name their favourite flavour of ice cream. The dual bar chart on the right shows the results.

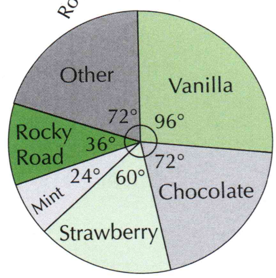


a) How many girls were asked altogether?

b) How many more boys than girls chose chocolate?

c) What is the modal flavour for the girls?

The pie chart on the right shows the same data for the girls.



d) Work out the number of degrees used to represent one girl.

e) Draw a pie chart to show the data for the boys.

f) Rocky road was chosen by the same number of girls and boys. Explain why the sectors representing rocky road in each pie chart are different sizes.

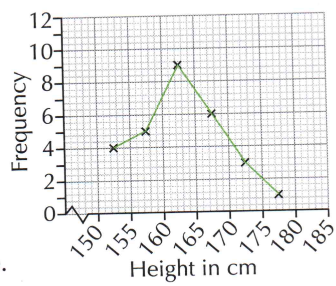
Q3 The data below shows the ages of people queuing in a post office at 10 am and at 3 pm.

10 am:	65	48	51	27	29	35	58	51	54	60	59
3 pm:	15	23	32	31	35	22	23	18	27		

a) Draw a back-to-back stem and leaf diagram to show the data.

b) Find and compare the median age of the people queuing at each time.

Q4 The frequency polygon shows information on the heights of members of a gymnastics club. The heights were put into groups with 5 cm intervals.



a) Create and fill in a grouped frequency table using the data from the frequency polygon.

b) What is the modal class for the gymnasts' heights?

c) How many gymnasts are there in the club?

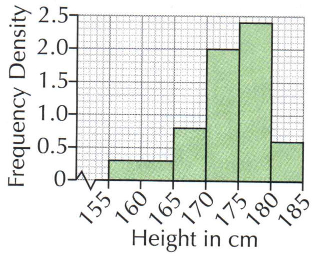
d) Gymnasts who have a height of 160 cm or greater use higher bar apparatus than everyone else. Estimate the percentage of members who use the higher bar apparatus to 1 decimal place.

Q5 The histogram below shows information about the heights of all the members of a netball club.

a) Copy this grouped frequency table and use the histogram to fill in the frequencies.

b) Estimate the number of netball players who are taller than 172 cm.

Height (h) in cm	Frequency
$155 \le h < 165$	
$165 \le h < 170$	
$170 \le h < 175$	
$175 \le h < 180$	
$180 \le h < 185$	

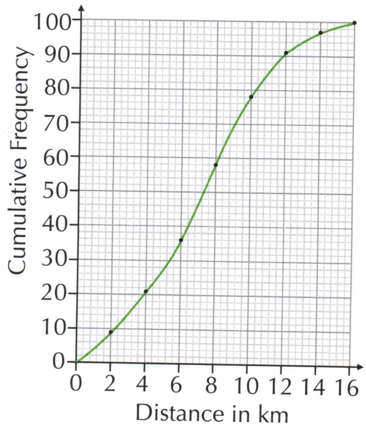


Q6 The cumulative frequency graph on the right shows the distances that 100 school children travel to school.

a) Use the graph to estimate the median.

b) Use the graph to estimate the interquartile range.

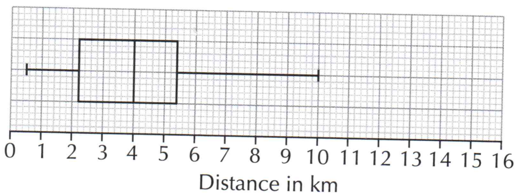
c) Children who live more than 4 km away get the school bus. How many children get the school bus?



The box plot below shows the distribution of the distances that 100 children at a second school travel to school.

d) Copy the box plot for the second school and draw a box plot for the first school above it. Use 1 km as the shortest distance travelled and 15.5 km as the longest distance travelled.

e) Use your box plots to compare the distances travelled by the children at the two schools.



Q7 The table shows the number of components made by a machine each hour for 10 hours.

Hour	1	2	3	4	5	6	7	8	9	10
No. of components	148	151	150	150	149	150	147	142	136	131

a) Draw a time series graph to show this data.

b) The machine is designed to produce 150 components per hour. Draw a line on your graph to show the target number of components for each hour.

c) A factory worker thinks the machine should be checked for faults. Do you agree? Use the data to explain your answer.

Q8 Anton wants to buy a particular model of car. The table below shows the cost of seven of these cars that are for sale, as well as their mileage.

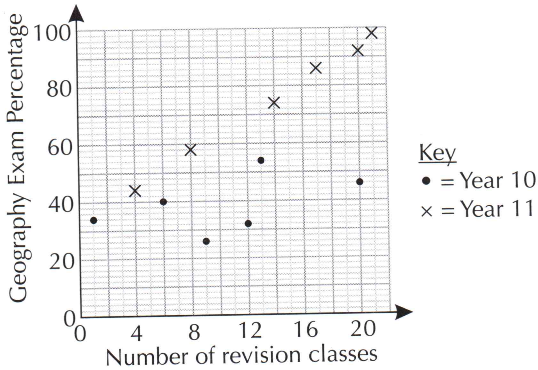
Mileage (×1000)	5	20	10	12	5	25	27
Cost (£)	3500	2000	3000	2500	3900	1000	500

a) Draw a scatter graph to show this data and add a line of best fit.

b) Predict the cost of a car of the same model with a mileage of 15 000 miles.

Exam-Style Questions

Q1 The scatter diagram shows the percentage obtained in a geography exam and the number of lunchtime revision classes attended by 12 students from Years 10 and 11.



a) Compare the correlations between the percentage obtained and the number of revision classes attended by Year 11 and Year 10 students.

[1 mark]

b) Duane was due to sit the exam in Year 11 but was absent. He says:

"I didn't attend any revision classes but I can see from the diagram that I would have got over 20%."

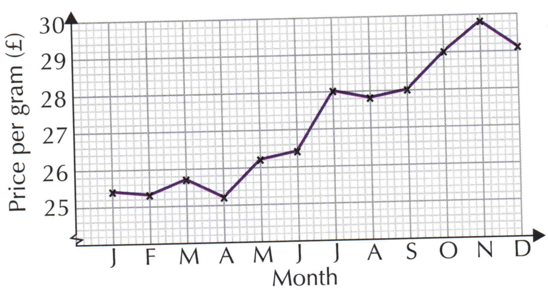
Is Duane correct? Refer to the diagram to explain your answer.

[1 mark]

Q2 The resting pulse rate in beats per minute (bpm) was measured for a group of athletes. The lowest rate was 44 bpm. 25% of the athletes had a rate of at least 71 bpm. The range was 32 bpm. The interquartile range was 19 bpm. The median was equal to the mean of the lower and upper quartiles. Draw a box plot to show this information.

[3 marks]

Q3 At the end of each month for a year, Midas recorded the price of gold per gram to the nearest £0.10. He displayed the data he recorded in the time series graph below.



a) Describe the overall trend in gold prices over the course of the year.

[1 mark]

b) Midas owns a 12 kg gold bar. According to the graph, what is the maximum amount of money he could have made from selling the gold bar this year?

[2 marks]

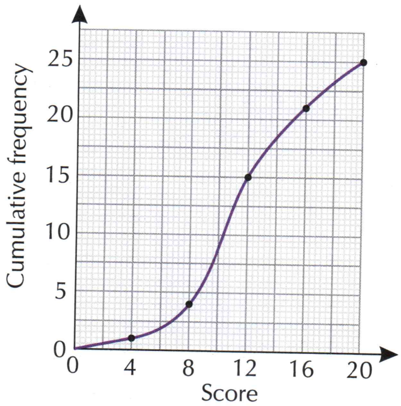
Q4 Participants in a circus skills workshop were asked to juggle three balls for as long as possible. The table on the right shows some information about how long they were able to do this for.

Draw a histogram to represent this information.

[3 marks]

Time (t) in seconds	Frequency
$0 \le t < 4$	18
$4 \le t < 8$	12
$8 \le t < 12$	6
$12 \le t < 20$	4
$20 \le t < 30$	1

Q5 Five judges each marked the bands in a competition out of 20. The mean of these five marks is the score that each band received. The scores are represented on the cumulative frequency graph below.



Score	Frequency
$0 < s \le 4$	1
$4 < s \le 8$	
$8 < s \le 12$	
$12 < s \le 16$	
$16 < s \le 20$	

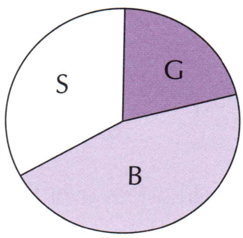
a) Copy and complete the frequency table above to show the bands' scores.

[2 marks]

b) Bands that scored over 14 went through to the next round of the competition. Estimate the percentage of bands that went through to the next round.

[2 marks]

Q6 Six classes competed against each other at a sports day. First, second and third places in each event won gold (G), silver (S) and bronze (B) medals. Students in Class 11R won a total of 24 medals, which are shown on the pie chart below.



drawn accurately

Students in class 11S won a total of 18 medals. Class 11S won the same number of gold medals as Class 11R and Class 11S won the same proportion of silver medals as Class 11R. If a pie chart was drawn for the medals won by Class 11S, work out the size of the angle of the sector for bronze.

[4 marks]