Cambridge OL

Mathematics

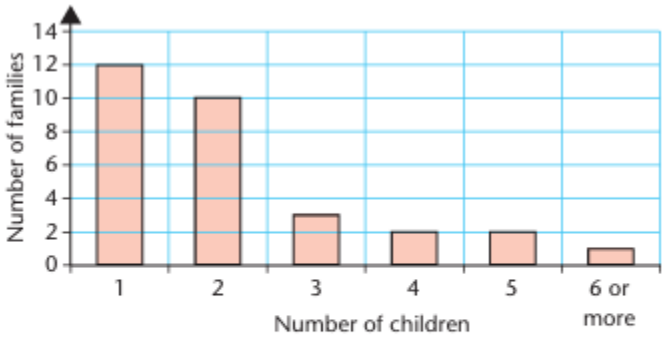
CODE: (4024)

Chapter 46

STATISTICAL DIAGRAMS

Bar charts and pictograms

Bar charts and pictograms can both be used to represent data that is divided into categories. This bar chart shows the number of children in each of the 30 families of Class 11S.



You can find the number of families with different numbers of children from the bar chart.

There are 10 families with 2 children.

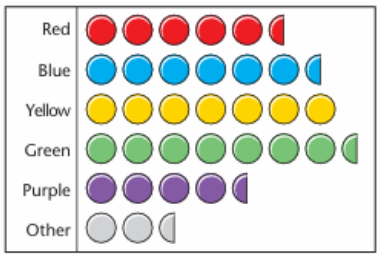
There is 1 family with 6 or more children.

Example 46.1

Question

This pictogram shows people's favourite colour for their mp3 player.

● represents 2 people.



a What does ◐ represent?

b How many people chose blue?

c How many people altogether were asked about their favourite colour of mp3 player?

Solution

a ● represents 2 people.

◐ is half a symbol, so it represents 1 person.

b There are $6\frac{1}{2}$ symbols for blue.

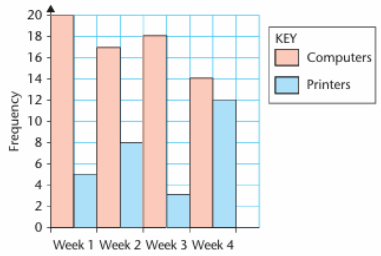
Each symbol represents 2 people, so 13 people chose blue.

c There are a total of $33\frac{1}{2}$ symbols in the pictogram, so 67 people were asked altogether.

Question

A shop records the numbers of computers and printers it sells in four weeks.

The information is shown in this dual bar chart.



a In which week did the shop sell the most printers?

b Work out how many computers were sold altogether in the four weeks.

c Work out how many more printers were sold in Week 4 than in Week 1.

d In which week did the shop sell the greatest total number of computers and printers?

Solution

a Week 4

b 69

c 7

d Week 4

There are other forms of bar charts, dual (side-by-side) bar charts and composite (stacked) bar charts.

Pie charts

A pie chart is another way to display data from a frequency table. The angle of the sector is proportional to the frequency. The sectors show the proportions not the actual numbers.

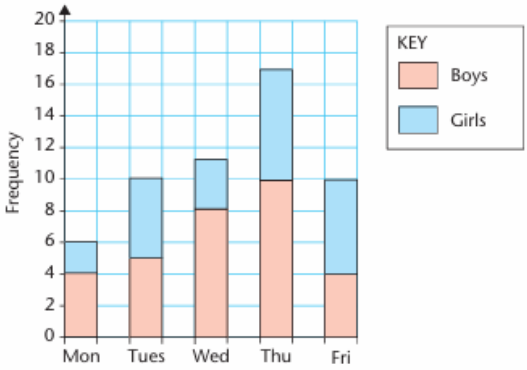
A pie chart shows proportions. You can write these as fractions or percentages.

To find out the number of data items in each group, you need to know the total number of data items.

Example 46.3

Question

The composite bar chart shows the numbers of boys and girls in one class absent from school one week.



a How many absences were there altogether in the week?
b On which day were more girls than boys absent?
c On which day was the largest number of girls absent?
d How many more boys were absent on Thursday than on Tuesday?

Solution

a 54
b Friday
c Thursday
d 5

Example 46.4

Question

The table shows data from a survey about which mathematics topic students preferred.

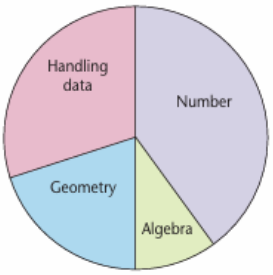
Draw a pie chart using this data.

Topic	Frequency
Number	8
Algebra	2
Geometry	4
Handling data	6

Solution

To draw a pie chart for this data, you need to work out the angle for each sector.

Topic	Frequency	Calculation	Angle
Number	8	$\frac{8}{20} \times 360$	144°
Algebra	2	$\frac{2}{20} \times 360$	36°
Geometry	4	$\frac{4}{20} \times 360$	72°
Handling data	6	$\frac{6}{20} \times 360$	108°
Total	20		360°



Note

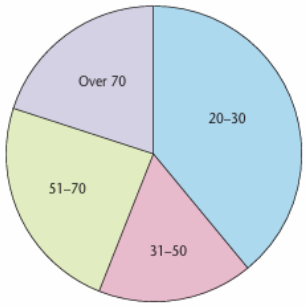
You should check that the angles add up to 360° and measure all sectors accurately when you draw the pie chart.

Example 46.5

Question

This pie chart shows the age groups of the people who entered a shop one morning.

a What was the most common age group?
b What fraction of the people were aged over 70?
c A total of 180 people entered the shop.
How many people were aged 51–70?



Solution

a The most common age group has the largest sector angle.
The most common age group was 20–30.
b The sector angle for over 70 is 72°.
The fraction is $\frac{72}{360} = \frac{1}{5}$.
c The sector angle for 51–70 is 86°.
$\frac{86}{360} \times 180 = 43$
There were 43 people aged 51–70.

Scatter diagrams

A scatter diagram is used to find out whether there is a correlation, or relationship, between two sets of data.

The data is presented as pairs of values, each of which is plotted as a coordinate point on a graph.

Here are some examples of what a scatter diagram could look like and how it might be interpreted.

Strong positive correlation

Here, one quantity increases as the other increases.

This is called **positive correlation**.

When the points are closely in line, the correlation is strong.

Weak positive correlation

These points also display positive correlation.

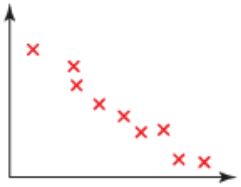
The points are more scattered so the correlation is weak.

Strong negative correlation

Here, one quantity decreases as the other increases.

This is called **negative correlation**.

Again, the points are closely in line so the correlation is strong.



Weak negative correlation

These points also display negative correlation.

The points are more scattered so the correlation is weak.

No correlation

When the points are totally scattered and there is no clear pattern, there is no correlation between the two quantities.

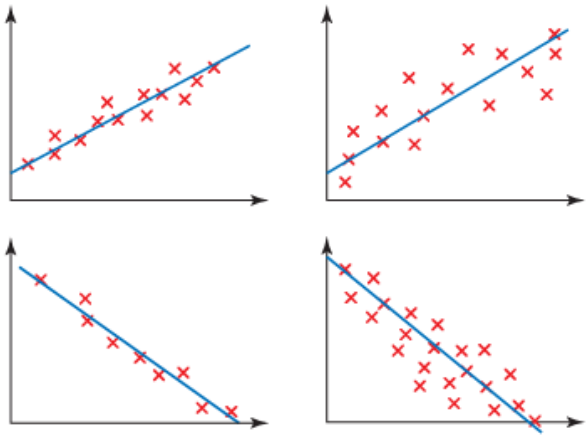
Line of best fit

If a scatter diagram shows correlation, you can draw a line of best fit on it.

This line goes in the same general direction as the points and has approximately the same number of points on each side of the line.

The line of best fit should extend across the full data set.

You cannot draw a line of best fit on a scatter diagram with no correlation.



You can use a line of best fit to estimate a value when only one of the pair of quantities is known. An estimate should only be made if the given value is within the range of the data used in the scatter diagram. If it is outside this range, the estimate will be unreliable

Example 46.6

Question

The table shows the heights and weights of 12 newborn giraffes.

Height (cm)	150	152	155	158	158	160	163	165	170	175	178	180
Weight (kg)	56	62	63	57	64	62	65	66	65	70	66	67

a Draw a scatter diagram to show this data.

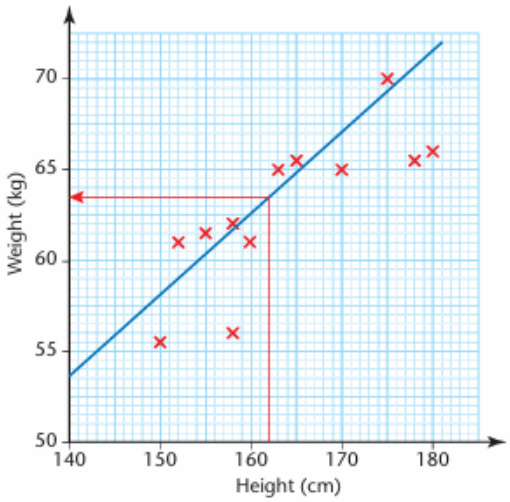
b Comment on the strength and type of correlation between these heights and weights.

c Draw a line of best fit on your scatter diagram.

d Another newborn giraffe is 162 cm tall. Use your line of best fit to estimate its weight.

Solution

a, c



b Weak positive correlation.

d Draw a line up from 162 cm on the height axis to meet your line of best fit.

Now draw a horizontal line and read off the value where it meets the weight axis.

The newborn giraffe's probable weight is about 63 kg.

Cumulative frequency diagrams

A cumulative frequency diagram is a way to display a large set of grouped continuous data. It can be used to estimate the median of the data and to give information about the spread of the data.

This table shows the distribution of the heights of 60 plants.

Height h (cm)	Frequency
$0 < h \leq 10$	1
$10 < h \leq 20$	3
$20 < h \leq 30$	9
$30 < h \leq 40$	25
$40 < h \leq 50$	11
$50 < h \leq 60$	6
$60 < h \leq 70$	3
$70 < h \leq 80$	1
$80 < h \leq 90$	1

The data has been combined in the cumulative frequency table

The cumulative frequency is the total of all the frequencies up to and including the selected group.

Height h (cm)	Cumulative frequency
$h \leq 0$	0
$h \leq 10$	1
$h \leq 20$	4
$h \leq 30$	13
$h \leq 40$	38
$h \leq 50$	49
$h \leq 60$	55
$h \leq 70$	58
$h \leq 80$	59
$h \leq 90$	60

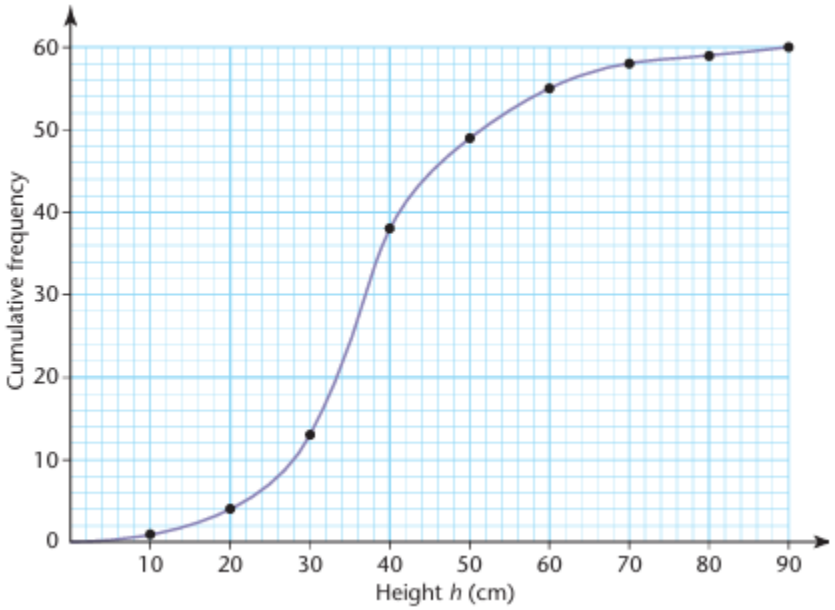
This row is included in the table to show that there are no plants with a height of 0 cm.

This tells you that there are 38 plants that have a height of 40 cm or less.

It is the sum of all the frequencies of the groups up to and including this one.

$1 + 3 + 9 + 25 = 38$

The cumulative frequency values can be used to plot a cumulative frequency diagram. The cumulative frequency values are plotted at the upper value of each interval, so the points plotted are (0, 0), (10, 1), (20, 4) and so on.



Example 46.7

Question

The table shows the distribution of the masses of 100 apples.

Mass m (g)	$50 < m \leq 100$	$100 < m \leq 150$	$150 < m \leq 200$	$200 < m \leq 250$	$250 < m \leq 300$
Frequency	15	18	34	25	8

a Make a cumulative frequency table for the data.

b Draw a cumulative frequency diagram for the data.

c Estimate the number of apples with a mass of

i less than 80 g **ii** greater than 190 g.

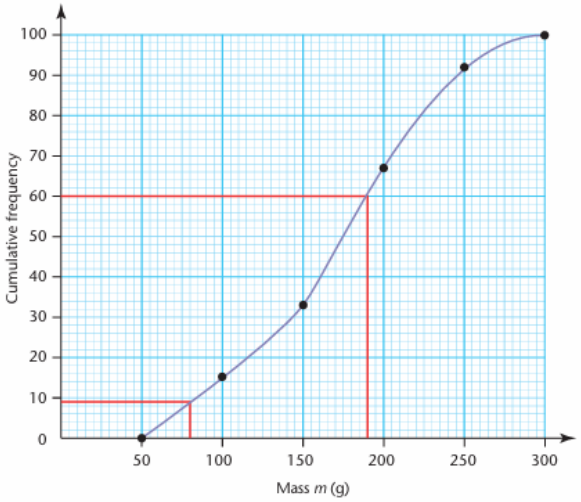
Solution

a

Mass m (g)	$m \leq 50$	$m \leq 100$	$m \leq 150$	$m \leq 200$	$m \leq 250$	$m \leq 300$
Cumulative frequency	0	15	33	67	92	100

There are no apples with a mass less than or equal to 50 g, so this column is added to the cumulative frequency table.

b



c **i** To find an estimate of the number of apples with a mass less than 80 g, read from the graph at this point.

There are 9 apples with a mass less than 80 g.

ii To find an estimate of the number of apples with a mass greater than 190 g, read from the graph at this point.

There are 60 apples with a mass *less than* 190 g.

So there are 100 – 60 = 40 apples with a mass greater than 190 g.

Median, quartiles and percentiles

A cumulative frequency diagram can be used to find the median value of the distribution. Look back at the cumulative frequency table and diagram on page 495.

There were 60 plants, so the median value is the height of the 30th plant. The median can also be called the 50th percentile because 50% of the data has a value less than or equal to that value. The median height is approximately 37 cm.

Note

If you were using a list of values, then for an even number of values the median would be halfway between the middle two values.

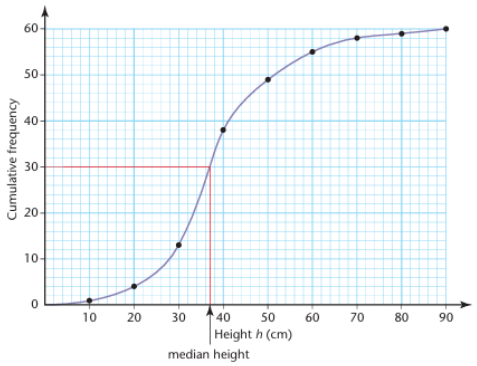
However, when using a cumulative frequency diagram for grouped data, the values of the individual items are not known.

Any results are therefore approximate and it is reasonable to use the 50th percentile.

The lower quartile (LQ) of the data is the 25th percentile. 25% of the data has a value less than or equal to this value.

The lower quartile is the height of the 15th plant. The upper quartile (UQ) of the data is the 75th percentile.

75% of the data has a value less than or equal to this value. The upper quartile is the height of the 45th plant.



The **interquartile range** (IQR) is a measure of the spread of the middle 50% of the data.

The interquartile range is the difference between the upper quartile and the lower quartile. (IQR = UQ – LQ).

Lower quartile = 31 cm

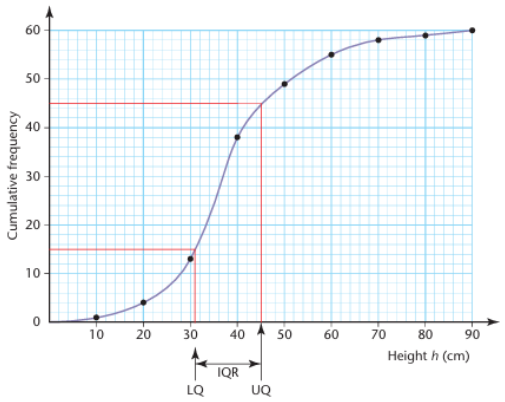
Upper quartile = 45 cm

Interquartile range = 45 – 31 = 14 cm

Other percentiles can be found from the data.

For example, 10% of the data has a value less than or equal to the 10th percentile.

For this data, the 10th percentile is the height of the 6th plant and is 24 cm.



Histograms

A histogram is a frequency diagram that is used to represent continuous data. The area of each bar is used to represent the frequency of the group.

The width of the bar is proportional to the class width. The vertical scale on a histogram is the frequency density.

Example 46.8

Question

For the data in Example 46.7, estimate

a the median

b **i** the lower quartile

ii the upper quartile

iii the interquartile range

c the 90th percentile.

Solution

The required values can be read from the cumulative frequency diagram.

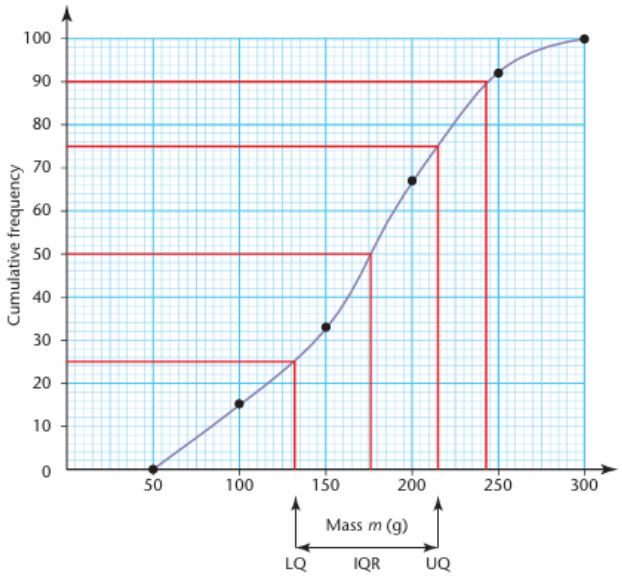
a There are 100 values, so the median is the 50th value.
Median = 175 g

b **i** The LQ is the 25th value.
LQ = 132 g

ii The UQ is the 75th value.
UQ = 215 g

iii IQR = 215 – 132 = 83 g

c 90th percentile = 243 g



$$\text{Frequency density} = \frac{\text{frequency}}{\text{class width}}$$

The units of frequency density depend on the units of the data.

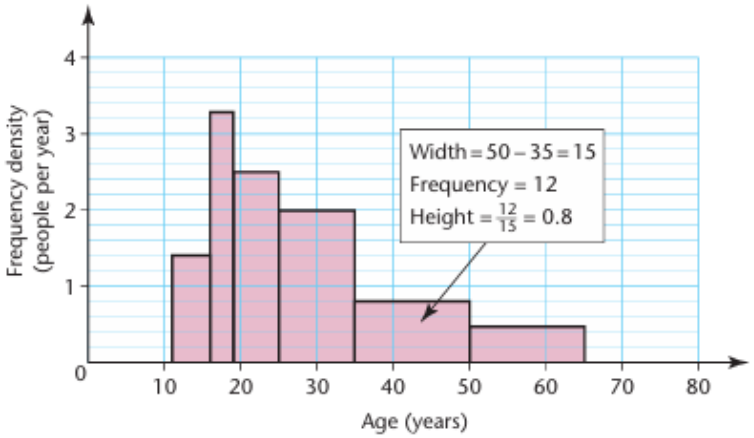
The table shows the distribution of the ages of people in a sports club.

Age in years	Frequency
11–15	7
16–18	10
19–24	15
25–34	20
35–49	12
50–64	7

The histogram below shows this distribution.

The upper boundary for the 11–15 age group is the 16th birthday. So the boundaries of this group are 11 and 16.

This applies similarly to the other groups, so the boundaries of the bars in the histogram are at 11, 16, 19, 25, 35, 50 and 65.



The calculation of the frequency density for the 35–49 group has been added to the histogram.

A similar calculation is done for each group.

For this histogram, the units of frequency density are people per year.

Note that the bars are not equal in width because the class widths are not equal.

Example 46.9

Question

An airline investigated the ages of passengers flying between London and Johannesburg.

The table shows the findings.

Age A (years)	Frequency
$0 < A \leq 20$	28
$20 < A \leq 30$	36
$30 < A \leq 40$	48
$40 < A \leq 50$	20
$50 < A \leq 70$	30
$70 < A \leq 100$	15

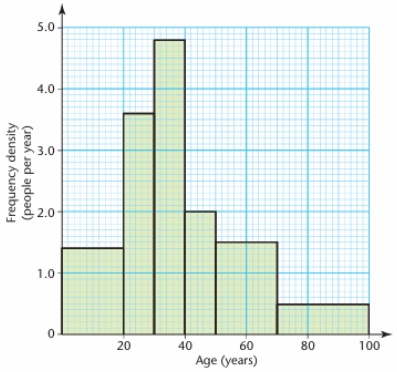
a Draw a histogram to represent this data.

b Estimate the number of passengers aged between 60 and 80.

Solution

a To draw a histogram you must first calculate the frequency density.

Age A (years)	Frequency	Class width	Frequency density (people per year)
$0 < A \leq 20$	28	20	$28 \div 20 = 1.4$
$20 < A \leq 30$	36	10	$36 \div 10 = 3.6$
$30 < A \leq 40$	48	10	$48 \div 10 = 4.8$
$40 < A \leq 50$	20	10	$20 \div 10 = 2$
$50 < A \leq 70$	30	20	$30 \div 20 = 1.5$
$70 < A \leq 100$	15	30	$15 \div 30 = 0.5$



Note

If a range includes parts of a class, then the calculation only gives an estimate, as we do not know how the data is distributed within that group.

b An estimate of the number of people can be found using frequency = frequency density × class width.

The range 60 to 80 contains 10 years from the 50 to 70 group and 10 years from the 70 to 100 group, so add the areas of those two sections of the histogram.

Number of people = 1.5 × 10 + 0.5 × 10 = 20

There are approximately 20 people aged between 60 and 80.

Key points

- Pictograms, bar charts, pie charts and frequency diagrams can all be used to display data.
- You need to know how to extract data from all forms of statistical diagram.
- A scatter diagram can be used to identify a relationship (correlation) between two sets of data.
- Positive correlation is shown by data points going from bottom left to top right (///).
- Negative correlation is shown by data points going from top left to bottom right. (\\\).
- When there is no clear pattern in the points, we say there is no correlation.
- A line of best fit is a straight line drawn in the general direction of the points, with points roughly balanced on both sides.
- The line of best fit can be used to estimate one data value when given another.
- A cumulative frequency is the total of the frequencies up to and including a particular group.
- Cumulative frequencies are plotted at the upper value of each group.
- Points on a cumulative frequency diagram are joined by a curve.
- On a cumulative frequency diagram:
  - the median of the data is the 50th percentile
  - the lower quartile of the data is the 25th percentile
  - the upper quartile of the data is the 75th percentile.
  - interquartile range = upper quartile – lower quartile.
- A histogram is a grouped frequency diagram where the area of each bar represents a frequency.
- The vertical scale of a histogram is the frequency density.
- Frequency density $= \dfrac{\text{frequency}}{\text{class width}}$.

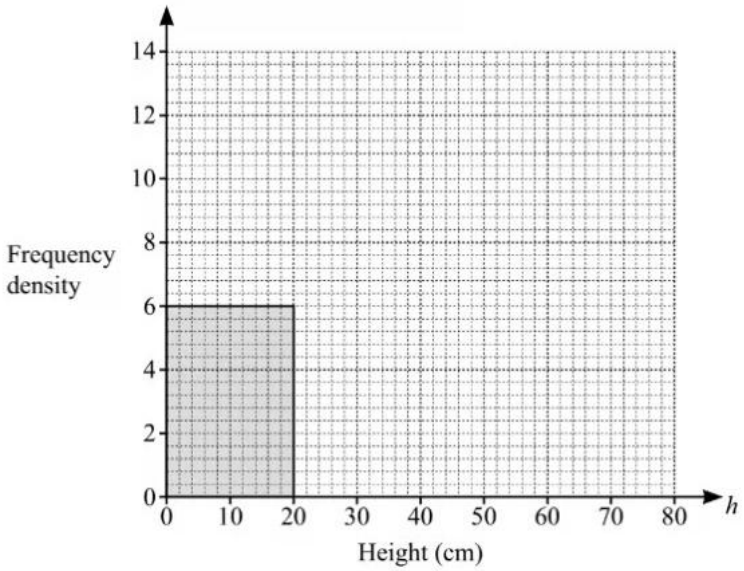
Revision questions

1.

The heights are also shown in the frequency table.

Height (h cm)	$0 < h \leqslant 20$	$20 < h \leqslant 30$	$30 < h \leqslant 40$	$40 < h \leqslant 80$
Frequency	120	80	124	76

Complete the histogram to show this information.



2.

The frequency table shows the times, t minutes, each of 100 children spent exercising in one week.

Time (t minutes)	$0 < t \leq 60$	$60 < t \leq 100$	$100 < t \leq 160$	$160 < t \leq 220$	$220 < t \leq 320$
Frequency	41	24	23	8	4

A histogram is drawn to show the information in the frequency table.
The height of the bar for the interval $60 < t \leq 100$ is 10.8 cm.

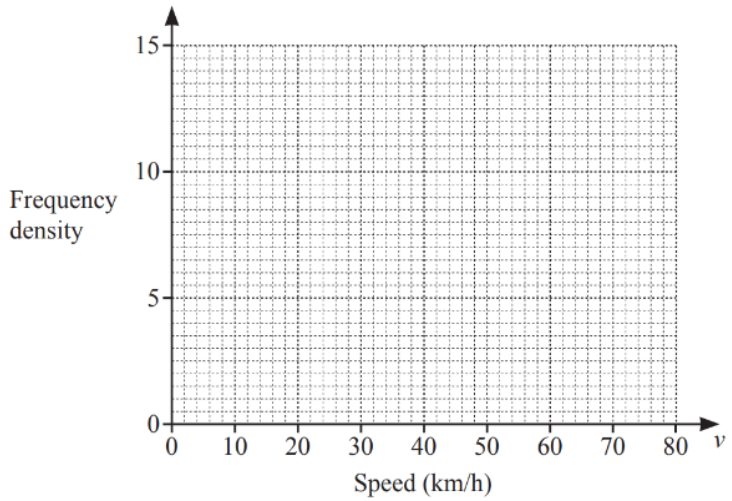
Calculate the height of the bar for the interval $160 < t \leq 220$.

3.

The speed, v km/h, of each of 200 cars passing a building is measured. The frequency table shows the results.

Speed (v km/h)	$0 < v \leq 40$	$40 < v \leq 50$	$50 < v \leq 80$
Frequency	50	120	30

On the grid, draw a histogram to show the information in this table.

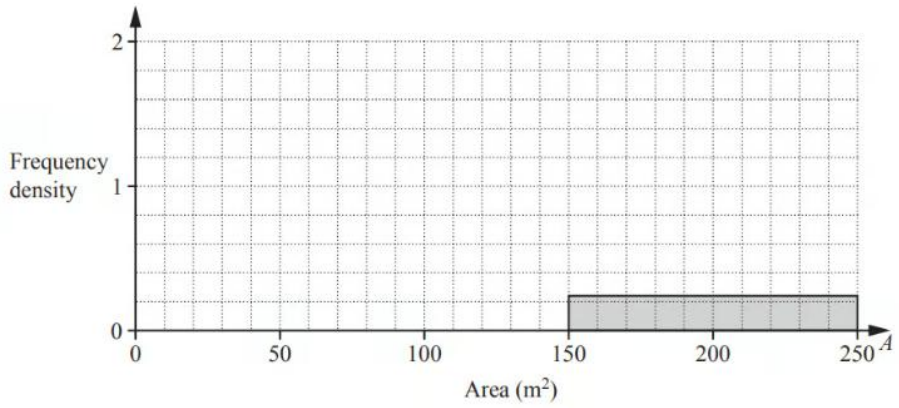


4.

200 students estimate the total area, A m^2, of the windows in a classroom.
The table shows their results.

Area (A m^2)	$20 < A \leq 60$	$60 < A \leq 100$	$100 < A \leq 150$	$150 < A \leq 250$
Frequency	32	64	80	24

Complete the histogram to show the information in the table.



5.

Some students each record the mass, m kg, of their school bag.
Adil wants to draw a histogram to show this information.

Complete the table below.

Mass (m kg)	$0 < m \leq 4$	$4 < m \leq 6$	$6 < m \leq 7$	$7 < m \leq 10$
Frequency	32			42
Height of bar on histogram (cm)	1.6	2	1.2	2.8

6.

A school nurse records the height, h cm, of each of 180 children.
The table shows the information.

Height (h cm)	$60 < h \leq 70$	$70 < h \leq 90$	$90 < h \leq 100$	$100 < h \leq 110$	$110 < h \leq 115$	$115 < h \leq 125$
Frequency	8	26	35	67	28	16

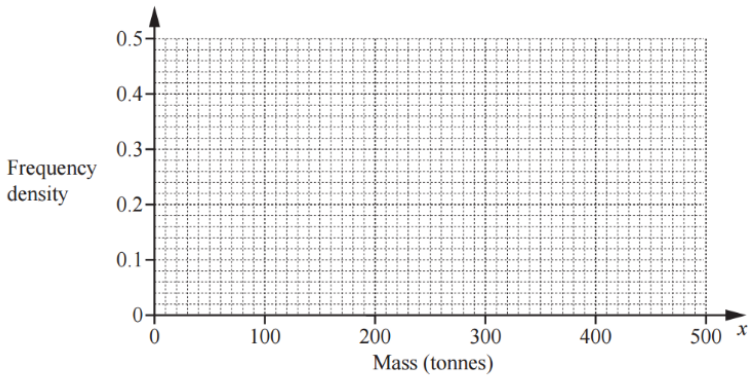
In a histogram showing the information, the height of the bar for the interval $60 < h \leq 70$ is 0.4 cm.

Calculate the height of the bar for each of the following intervals.

7.

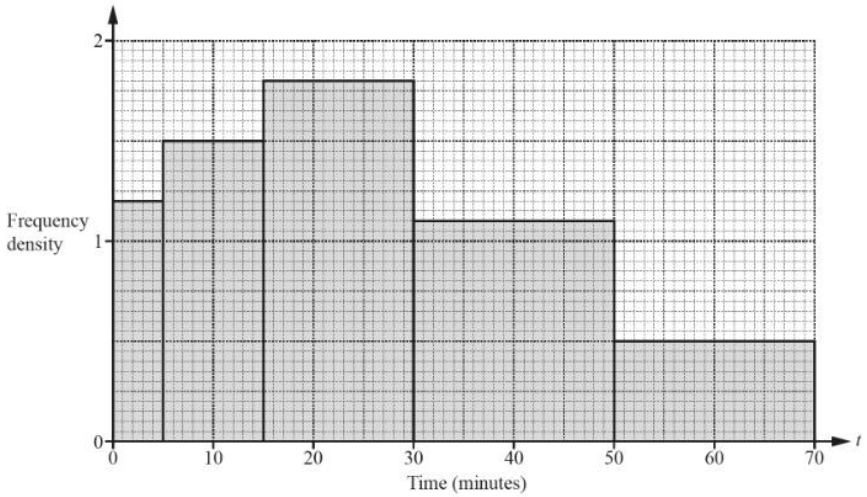
A factory recycles metal.
The mass, x tonnes, of metal is measured each week.
The table shows the results for 52 weeks.

Mass (x tonnes)	$100 < x \leq 200$	$200 < x \leq 250$	$250 < x \leq 300$	$300 < x \leq 500$
Frequency	8	20	12	12



On the grid, draw a histogram to show the information in the table.

8. The histogram shows information about the time, t minutes, spent in a shop by each of 80 people.



Complete the frequency table.

Time (t minutes)	$0 < t \leq 5$	$5 < t \leq 15$	$15 < t \leq 30$	$30 < t \leq 50$	$50 < t \leq 70$
Number of people	6		27		10

9. The height, h cm, of each of 120 plants is measured.
The frequency table shows this information.

Height, h cm	$0 < h \leq 10$	$10 < h \leq 20$	$20 < h \leq 30$	$30 < h \leq 50$
Frequency	2	18	62	38

A histogram is drawn to show the information in the frequency table.
The height of the bar representing the interval $10 < h \leq 20$ is 7.2 cm.

Calculate the height of the bar representing the interval $30 < h \leq 50$.