

Please Note: this is the text version, with most of the larger graphics removed. To download the full PDF version go to: <http://www.edu-sol.co.uk/sample.asp>

GCSE Maths

Foundation and Higher Tiers p 1

Revision of Bar Charts

A Qualitative Bar Chart p 2
 Qualitative and Quantitative Data p 2
 Qualitative and Quantitative Bar Charts p 3
 The x-axis on a Quantitative Bar Chart p 3

Introduction to Histograms

Continuous Data p 4
 Example of a Histogram p 4
 Grouping p 5
 Class and Class Interval p 6
 Example 1 – Grouping Values p 7
 Exercise 1 – Grouping p 8
 Exercise 2 – Histogram p 8
 Summary p 9

Drawing Frequency Diagrams in the Exam

Coursework p 10
 About Exam Questions p 10
 Exercise 3 – Draw a Histogram p 10

GCSE Mathematics

This booklet is for the GCSE Maths Higher tier. It may be more suited to those, possibly adult students, who cannot attend formal classes or who do not have time in class to cover the Higher tier material. It could be useful for non-specialist teachers, who may need more detail than you find in a typical textbook.

Statistics, with its reliance on sophisticated charts, particularly the histogram, can be a problem given that some students have limited experience of traditional x-y graphs. My experience with mature students is that tackling the intricacies of histograms and frequency density, with only a vague recollection of graph plotting, is a challenge.

Fortunately, in the GCSE Maths specification, the emphasis on continuous data and the assumption that measurements are 100% accurate make histograms more ‘digestible’. The treatment can progress logically from bar charts, for discrete data, to grouped, continuous data and hence the histogram. Histograms for grouped, *discrete* data, with the extra complications that involves, is not required. Perhaps though, students should at least be aware that histograms are used with large ranges of discrete data.

The booklet starts with a review of bar charts. The Foundation tier does not require you to draw or interpret a histogram and does not expect any knowledge of frequency density. The sections on qualitative and quantitative bar charts are relevant to Foundation. Students are likely to have covered grouped data already, and created grouped tables from raw data. The booklet introduces grouping as a natural requirement of the need to chart continuous data. In the introduction to histograms the classes are equal width, and frequency, not frequency density, is used. This should make histograms more accessible. Frequency density, required for the Higher tier, is covered in a later section. Page 10 (Frequency Diagrams in the Exam) may be more relevant after the section on Frequency Polygons.

Copyright. You are permitted to download and print copies of this document for educational use only. You are not permitted to alter, store (in any medium), transmit, adapt or change the content in any way. Any rights not expressly granted in these conditions are hereby reserved.

Frequency Polygons

Introduction p 11
 Using Excel p 12
 Advantages of a Frequency Polygon p 14
 Exercise 4 – Frequency Polygon p 14
 Summary p 15
 Exam Question – Frequency Diagram p 16

Frequency Density

Definition p 18
 The Problem with Frequency p 18
 Frequency and Frequency Density p 19
 The Effect of Class Width p 20
 The Area of a Histogram p 21
 Modal Class p 22
 Summary p 22
 Exam Questions on Frequency Density p 23
 Exercise 5 – Reading a Histogram p 23
 Exam Question – Histogram p 24
 Answers to Exercises p 25

Appendices

A Explanation of Frequency Density p 27
 B Variable Class Width p 28
 C Is a Histogram a Proper Graph? p 30

Revision of Bar Charts

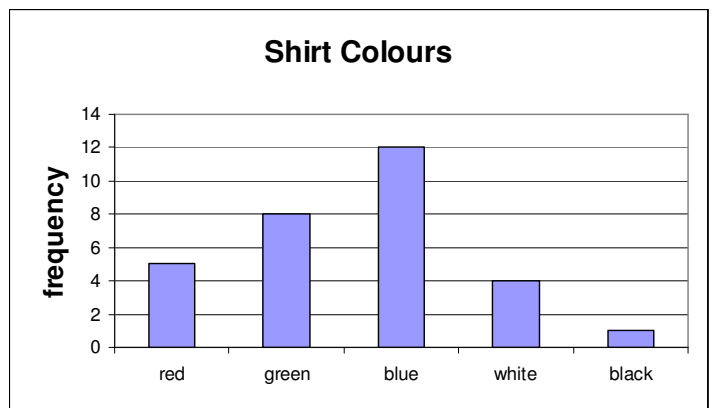
A Qualitative Bar Chart

A bar chart of the colours of shirts, worn by children in a class, shows the number of children wearing the different colours. You could have 5 wearing red, 8 with green, 12 with blue, 4 with white and so on. The colours are called *categories*. The number of shirts of each colour is called the *frequency*. The frequency of green shirts is eight. Those numbers, the frequencies of the different categories, are called a *frequency distribution*. A table of the categories and frequencies is called a frequency distribution table or, more simply, a *frequency table*.

A *frequency diagram* is a chart of the frequencies of the different categories. A bar chart is a frequency diagram. It is a graphical display of the frequency distribution. A frequency diagram is easier to interpret than a table of the frequencies. It is easier to see a pattern. Using suitable scales, labelling and shading will make the diagram more readable.

In the shirt example the categories are the colours. The bars could be in any order – either an ascending or descending order of height may be preferred. The frequency scale starts at zero. The divisions on the frequency axis are equally spaced.

You will meet two more types of frequency diagram in this booklet, both based on a bar chart: histograms and frequency polygons.



Qualitative and Quantitative Data

Data can be *qualitative* or *quantitative*. ‘Quantitative data’ means that the values of the data can be counted or measured, such as the number of kittens in a litter or the weight of sacks of flour. The number of kittens can be counted, so that is **discrete** quantitative data. You can have 2 kittens or 3 kittens but nothing in between. The weight of a sack can be measured. It can have any value within a range and so weight is **continuous**. Here’s a small sample of continuous data, the weights, in kilograms, of six small children: 20, 24.5, 28.1, 30, 32.5, 35. The word ‘continuous’, when applied to data, does not mean that every possible value must be present in the sample. None of the six children weighs 26.8kg, but it is possible to weigh that amount, or any value within a range. Therefore weight is continuous.

Qualitative data are neither measurements nor counts. Examples of qualitative categories are colour (e.g. of shirts) and names (e.g. of products). The colour red is not a measurement. An instrument can *measure* a colour in the sense that it can measure the wavelength of the colour and then identify the colour as red. The word ‘red’ is not the measurement – it is just a name and so is qualitative.

The qualitative bar chart is the simplest type of frequency diagram. Frequency diagrams of quantitative data, particularly continuous data, are more complicated. In the next section we compare a qualitative with a quantitative bar chart.

Comparison of Qualitative and Quantitative Bar Charts

In this qualitative bar chart on pet ownership there are 5 categories of animal. The frequency is the number of pets in each category. You can see that 15 of the animals are dogs and only 2 are pet rats. The order is alphabetic from left to right.

The next bar chart is quantitative. It shows the number of people living in thirty apartments. None of them is empty.

The horizontal axis (the x-axis) shows the occupancy. The range is from 1 to 5 people. The occupancy is quantitative. It is discrete, because a number of people is an integer. You can't have 1.5 people.

The frequency is the count of apartments for each level of occupancy. For example, 10 of the apartments are each occupied by one person only, so the frequency is 10. How many are occupied by 5 people?

Bar Chart: Pet Ownership

Bar Chart: Occupancy of Apartments

In a question you could be asked to work out the total number of apartments. You add together the frequencies (the heights) of the bars. The total frequency is:

$$10 + 7 + 8 + 4 + 1 = \mathbf{30 \text{ apartments}}$$

How many people in total live in apartments occupied by 3 people? There are 8 apartments occupied by 3 people per apartment, so the total occupancy is $8 \times 3 = 24$ people.

Calculations like these are best completed in a frequency table:

Occupancy x	1	2	3	4	5	Totals
Frequency f	10	7	8	4	1	30
Total Occupancy $f \times x$	10	14	24	16	5	69

The x-axis on a Quantitative Bar Chart

A quantitative bar chart is used for discrete data only. The x-axis looks like a scale. It shows all the possible values in numerical order. In the Occupancy chart, the bars are in ascending order of occupancy from left to right. The x-axis is labelled with the integers 0 to 6. The labels are equally spaced because the occupancy increases in equal amounts. If there are no apartments with, for example, 3 occupants there is no bar, but you still label 3 on the axis. The 0 and 6 labels can be omitted because they are outside the data range. The bars are positioned so that the middle of each bar is aligned with the value on the x-axis.

The width of the bars on a bar chart has no mathematical significance. In fact, a maths view is that the bars on a quantitative bar chart should be vertical lines, with no width. The appearance of the bars has more to do with style than maths. A coloured bar with width simply looks better than a thin, vertical line.

Introduction to Histograms

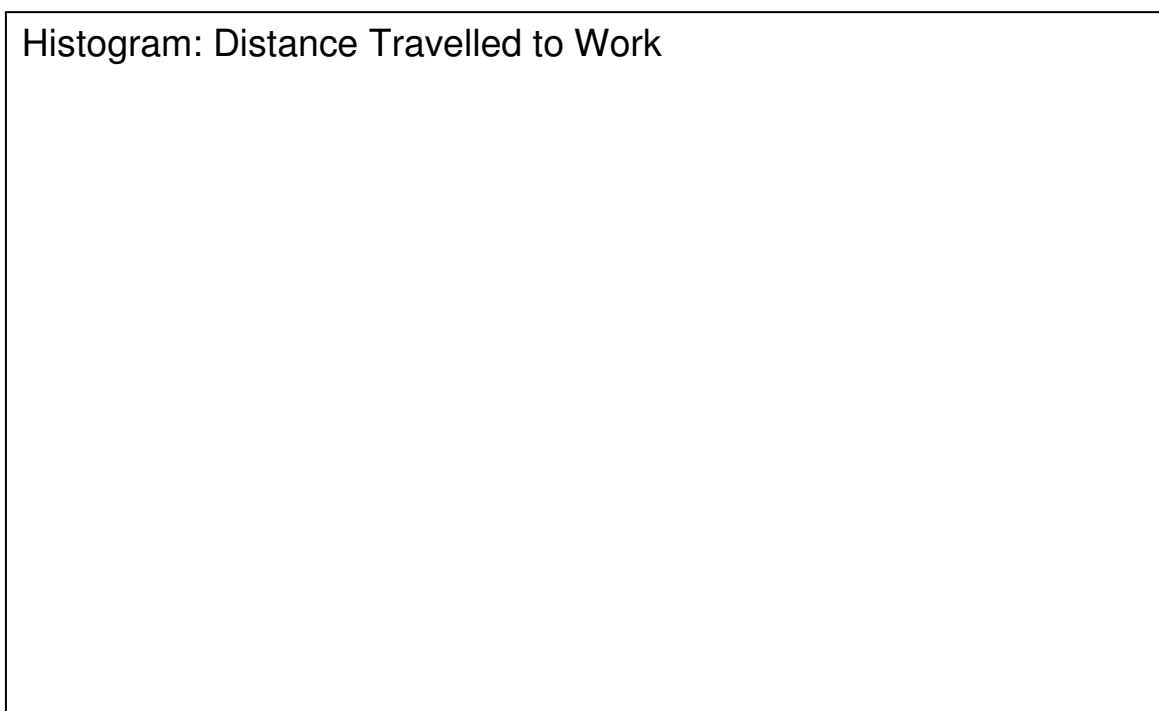
Continuous Data

It is fair to say and it may be an understatement: histograms can be confusing. To avoid initial confusion, we'll start by describing a histogram, instead of explaining in detail.

A **histogram** is a frequency diagram and is similar to a quantitative bar chart. In the simpler type of histogram that we will look at first, the vertical y-axis shows the frequency, just like a bar chart. The x-axis is a numerical scale with equally spaced divisions. The example histogram below has an x-axis from 0 to 14miles. A histogram is used for continuous data, such as length, time and weight. Therefore the x-scale is continuous. Continuous data cannot be represented on a bar chart. The choice of scale divisions, as with any graph, depends on the data range. In this example 1mile was chosen. For a smaller range, such as 0 to 5miles, you could use half-mile divisions on the x-axis.

The bars on a histogram are drawn touching each other – shown in the example histogram. The bars touch *because* the data are continuous. There are no gaps between the data's possible values and so there are no gaps between the bars. Compare this with a bar chart – where the gaps remind us that the data are discrete. The bars on a histogram are often called rectangles. An important thing about histograms is that the data are grouped. To represent continuous data on a chart, the data must be grouped, explained on the next page.

Example of a Histogram



This is the main example of a histogram used throughout this booklet. The x-axis shows the distance employees travel to work. The frequency is the number of employees. No one travels less than 1mile or more than 13miles – the range is from 1 to 13miles.

The group of sixty people, who travel from 1mile up to 3miles, is represented by a rectangle that is 2miles wide and 60 people high. The vertical sides of the rectangle line up with 1mile and 3miles on the x-axis. The next group extends from 3 to 5miles. This is explained fully in the section on **Grouping** (next page).

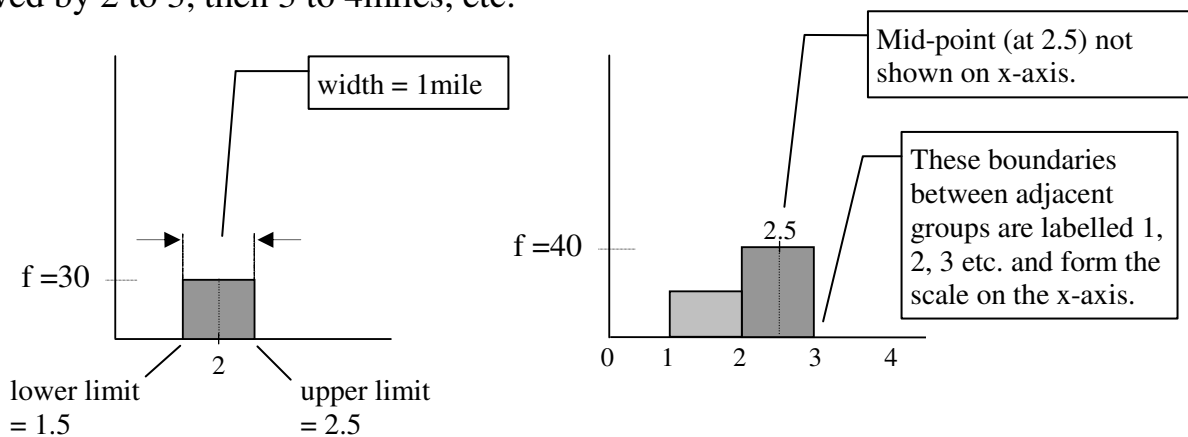
Grouping

A main difference between a bar chart and a histogram is that, with a histogram, the data are grouped. How would you use a bar chart to represent the 'distance travelled' data? A bar chart shows discrete data. Maybe you could count the number who travel 1 mile, then 2 miles, and so on, and plot that as a bar chart, with a bar for each mile.

Distance, however, is a continuous variable, so how would you include those who travel distances in between, such as 1.4 miles? Perhaps you could approximate the distances to the nearest mile. For example, if the distance is 1.4 miles, round it down to 1 mile and, for 1.6 miles, round it to 2 miles. Then use a bar chart with a bar for each mile.

That is similar to what we do. The data between 1.5 and 2.5 miles could be rounded to 2 miles – so that those data are included in the 2 miles bar. A better description is that we lump together all the data from 1.5 to 2.5 miles and place them in a 2 miles group. It is called **grouping**. However, unlike a discrete bar on a bar chart, which represents only one value, a bar on a histogram represents a whole range of values – in this example, from 1.5 to 2.5 miles. On the histogram the bar's width is important. The width of the bar is made to stretch the width of the group, from 1.5 to 2.5 miles, to remind us that the data in the group can be anywhere between 1.5 and 2.5 miles. The bar's height is the count of data items in the group, in this case 30 people. The next group, from 2.5 to 3.5 miles, is also 1 mile wide and is centred on 3 miles.

The left diagram below shows just the 1.5 to 2.5 miles group. The bar represents the 30 people who live between 1.5 and 2.5 miles from work ($f=30$). To simplify, instead of grouping from 1.5 to 2.5 miles, we'll use whole numbers. We can group from 1 to 2, followed by 2 to 3, then 3 to 4 miles, etc.



The right-hand diagram shows the alternative – two groups, from 1 to 2 and from 2 to 3 miles, are shown, with 40 people in the 2nd group. The boundaries between groups line up with divisions on the axis, 1, 2, 3 etc. The mid-point for the 2nd group is 2.5 miles. Unlike a bar chart, it is not necessary to show the mid-point values on the x-axis.

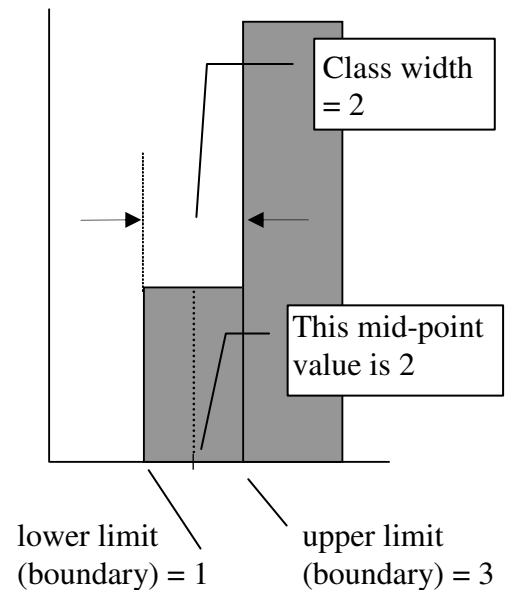
To summarise, in the employees example you *could* place those who travel between 1 and 2 miles in a group that extends from 1 up to 2 miles. The 1.4 miles goes in that group. Those between 2 and 3 miles go in the next group, and so on. Each bar is one mile wide. By choosing groups starting at 1 mile and with a width of 1 mile, the boundaries of the groups and the x-axis divisions coincide. This makes introductory examples of the histogram easier to understand and is often the case in GCSE Maths questions. In general, however, there is no requirement for the boundaries to line up with the scale divisions.

Class and Class Interval

Using 1 mile group widths will give 12 groups for the total range of data, from 1 to 13 miles. You don't have to use a 1 mile width. In the histogram for the employees on page 4, the group width was 2 miles. The first group starts at 1 mile and stretches to 3 miles. This gives altogether 6 groups. A sensible number of groups is no less than 5 and no more than about 15 groups.

The groups are called **classes**. The range of values included within a group is called the **class interval**. The class interval for the first class is from 1 mile up to, **but not including**, 3 miles. 2.9 miles is in the first class but 3 miles is in the next class, from 3 up to, but not including, 5 miles.

The **class limits** are the **boundaries** of a class. The **lower** class limit of the 1 to 3 miles class is 1 mile. The **upper** class limit of that class is 3 miles. The class width is the difference between the upper and lower boundaries: in this case $3 - 1 = 2$ miles. The **mid-point** is the middle value. The mid-point of the 1st class is 2 miles.



In this section you have learned these terms used to describe groups:

class, class width, class interval, lower and upper class limits, class boundary, mid-point.

In the example histogram on 'Distance Travelled to Work', the data are grouped into 6 classes.

- The class limits coincide with the class boundaries. You may see examples where the limits are inside the boundaries.
- The class width is 2 miles, the difference between the upper and lower boundaries.
- The 2nd class interval is 3 miles up to, but not including 5 miles and its mid-point (the mean of 3 and 5) is 4 miles.
- The upper boundary of the 2nd group is 5 miles.
- In GCSE Maths questions, the class boundary values are often integers, such as 10, 20, 30 etc. The boundaries may then coincide with labelled divisions on the x-axis, which could be 5, 10, 15, 20 etc. With a bar chart, however, only the possible data values are labelled on the x-axis and the bars are centred on those labels.

Symbol for the Class Interval

In mathematics, the class interval is written using the inequality symbols $<$ and \leq .

The $<$ means *less than* as in $x < 3$ (x is less than 3).

The \leq means *less than or equal to* as in $1 \leq x$ (1 is less than or equal to x).

The class interval for distance d in the group "from 1 up to, but not including 3" is written:

$$1 \leq d < 3$$

It means that d can take any value between 1 mile and 3 miles, but not including 3. If $d=2.9$ then d is less than 3 and so it belongs in the first class. If $d=3$ it belongs in the next class.

Example 1 – Grouping Values

Here are eight values of a continuous variable. The variable could be any continuous variable such as length or weight or time, and so it is represented by x .

Variable x	5.6	10.02	9.98	10	19.4	21.3	0.5	24.5
--------------------------------	-----	-------	------	----	------	------	-----	------

In the frequency distribution table below, complete column 2. Place each value in its correct class. Record the frequency in column 3. This is an example question and so the answers are already filled in (shown in blue).

Class Interval	Values of x	Frequency (count)
$0 \leq x < 10$	<i>0.5, 5.6, 9.98</i>	<i>3</i>
$10 \leq x < 20$	<i>10, 10.02, 19.4</i>	<i>3</i>
$20 \leq x < 30$	<i>21.3, 24.5</i>	<i>2</i>

An Alternative Symbol for Class Interval

The hyphen symbol is sometimes used to show the class interval as in these examples:

Height h of Children in Centimetres

Class Interval h (cm)	110 –	120 –	130 –	140 –	150 –	160 –
f	10	15	17	20	16	9

In the table the “110 –” represents the class interval from height $h=110$ cm up to, but not including, 120cm. There are 10 children in that group (frequency $f=10$). An alternative is to show 110 – 120 for the first class and then 120 – 130 for the next.

Weight of Children in kg

Class Interval Weight (kg)	15.0 –	15.5 –	16.0 –	16.5 –	17.0 –
---------------------------------------	--------	--------	--------	--------	--------

In this case the first class is any weight from exactly 15.0kg up to, but not including, 15.5kg, which belongs in the next class. 15.49 belongs in the first class. The last class includes weights from exactly 17.0kg up to, but not including, 17.5kg.

Exercise 1 – Grouping a Continuous Variable

Here are eight values of the continuous variable x . Copy and complete the table below. Complete the class interval in the second column in the form $90 \leq x < 100$. Complete the x column by placing each value in its correct group. Complete the frequency f column for each class.

113	125	97	123	104	119	90	120
-----	-----	----	-----	-----	-----	----	-----

Class Interval	Class Interval	Values of x	f
90 –	$90 \leq x < 100$	90, 97	2
100 –			
110 –			
120 –			

Exercise 2 – Histogram

In the example histogram on page 4, the distance travelled is represented by the letter d . In another example we might have a variable h for height or W for weight or a general variable x . The histogram below is about the times taken to run a race, from 7 to 14 minutes. The time is represented by t . There are 24 runners.

There is no point in starting the time-axis (the x -axis) at zero, because the fastest time is 7 minutes. The data are grouped into 1 minute widths, giving seven classes. The first class extends from 7 minutes up to, but not including, 8 minutes ($7 \leq t < 8$). Sensible scale divisions on the x -axis are at 7, 8, 9 etc. and so the class boundaries correspond to scale divisions. For example, the boundary between the 3rd and 4th classes is at 10 minutes, the upper boundary for the 3rd class and the lower boundary for the 4th class.

Histogram: Times to Run a Race

The question is on the next page.

Use the histogram to complete a copy of the table below, giving the frequency and mid-point (the time-value at the middle of the interval) for each class interval. The first row, labelled **Time t** , is the class interval. The 1st column shows that 3 runners ran in times from 7 up to 8 minutes.

Times Taken to Run a Race

Time t (minutes)	$7 \leq t < 8$	$8 \leq t < 9$	$9 \leq t < 10$	$10 \leq t < 11$	$11 \leq t < 12$	$12 \leq t < 13$	$13 \leq t < 14$
Frequency	3						
Mid-point (minutes)	7.5						

Summary

So far we've looked at a simplified form of histogram. The main points are:

- a histogram is a frequency diagram. It is similar to a bar chart but with some important differences.
- the data's values are continuous (the variable is continuous) – any value is possible within the data range.
- the data are grouped into classes.
- the class widths are equal.
- because the variable is continuous, the x-axis is a continuous numerical scale, just like the scale used on a standard x-y graph.
- the width of a bar on a histogram spans the class interval – therefore adjacent bars touch.
- the height of a bar is the frequency (the number of data in the class).

The width of a bar on a histogram corresponds to the class width, which is the range of values included in the class. So far we have simplified and kept the class width constant. The classes and therefore the histogram bars do not have to be a constant width. This complication is dealt with in the section on **Frequency Density**.

Discrete Data

You may come across histograms of discrete data, such as marks in a test. In a GCSE Maths histogram question, the data will almost certainly be continuous or can be treated as continuous. In this booklet we may have implied that grouping data is necessary for continuous data only and, by implication, that a chart of grouped data (a histogram) is only ever used with continuous data. In fact, discrete data has to be grouped when there is a large range of data, otherwise there would be too many bars on the bar chart. The resulting grouped frequency distribution is then plotted as a histogram or a frequency polygon.

Deciding on the class boundaries is a little more complicated than with continuous data. Test marks could be grouped from 0 to 24, 25 to 49, 50 to 74 and 75 to 100. That's too few classes, but it is only an example. The class boundaries have to be midway between the upper mark for one class and the lower mark for the next. For the class 25 to 49, the lower boundary is 24.5 and the upper boundary is 49.5. The class width is 25. This is really beyond GCSE Maths – but at least you should not be surprised if you do see grouped, discrete data used with a histogram.

Frequency Diagrams in the Exam

Coursework

When doing statistics coursework you have to decide about what types of chart to use. If the data are continuous, you have to group the data. Choose a suitable class width to give a sensible number of classes. You should have no less than 5 and no more than about 15 classes on a chart. There is no exact rule. A ‘rule of thumb’ for continuous data is that the number of classes should be about equal to the square root of the number of data values in your sample. If, for example, your sample has 100 items you make about $\sqrt{100} = 10$ classes. Even if the data are discrete, you may still decide to group, but you should consider grouping only if your data range is at least 10. For example, if the data are possible dice scores, from 1 to 6, you should not group. If you are comparing the lengths of sentences in newspapers, the range is large and so you need to group. For example, this sentence contains about fifty characters. Whatever you decide, you have to explain the decision.

About Exam Questions

After 2008, there is no GCSE Maths coursework component. Exam questions on drawing a histogram (Higher tier) have so far been rare. With the end of coursework it is possible that such questions will become more common. You are advised to practise drawing histograms by hand on graph paper.

A question may ask you to draw a “frequency diagram”. You could interpret this as meaning a bar chart, a histogram or a frequency polygon (next section). Be guided by the question – if the data are not grouped (and you are not required to group) then you could draw a simple bar chart. It is more likely that the data are grouped, in which case use a histogram or a frequency polygon.

It is unlikely that you will be expected to draw a histogram of grouped **discrete** data in an exam question. You *are* expected to be able to use a grouped frequency table, e.g. to estimate the mean test mark for a class of students. You use the mid-point values.

An exam question on drawing a histogram will almost certainly provide the table of grouped data and a grid for the chart.

Exercise 3 – Draw a Histogram

Using graph paper, draw a histogram to represent the data given in the following frequency table of continuous data. Be sure to give the graph a title, and to label both axes clearly. Remember that the bars must touch. There is no need to shade the bars.

Table of Weights of Children

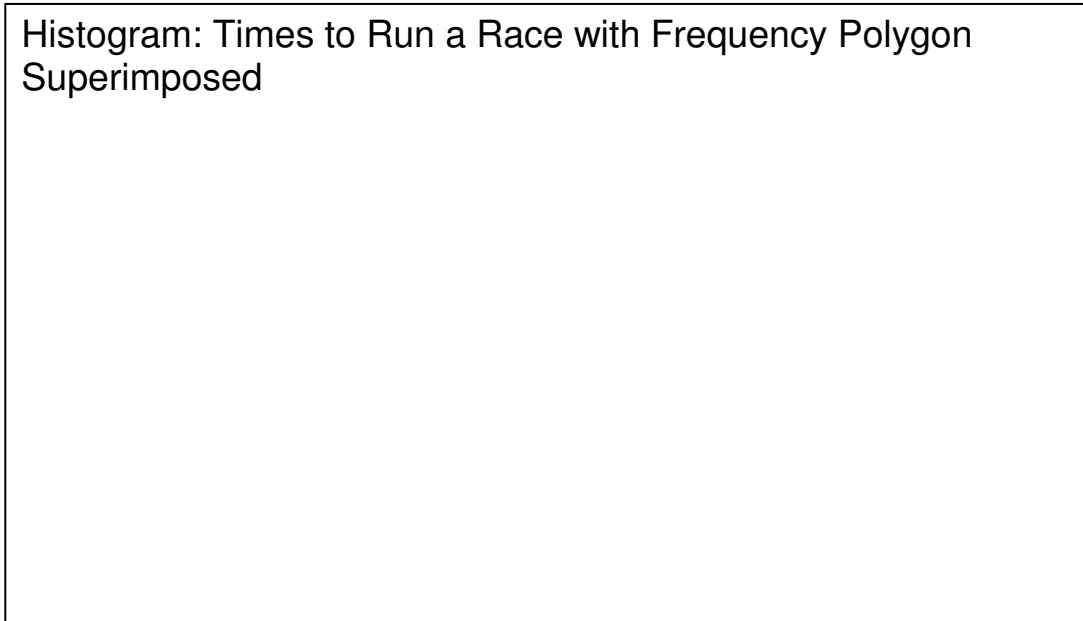
Weight W (kg)	Frequency f (number of children)
$20 \leq w < 25$	15
$25 \leq w < 30$	65
$30 \leq w < 35$	50
$35 \leq w < 40$	20
$40 \leq w < 45$	10

Frequency Polygons

Introduction

A frequency polygon is a frequency diagram of grouped data and is an alternative to a histogram. The bars of the histogram are replaced by straight lines joining the mid-points. We will use the earlier histogram, 'Times for a Race', as an example.

The diagram below shows the histogram and the frequency polygon combined. The mid-points are plotted, shown by '+' . The first point is plotted at $t = 7.5$ minutes, $f = 3$, half way between the class boundary divisions at 7 and 8 minutes. The next point is at 8.5 minutes.



In a question to draw a frequency polygon you are not expected to include the histogram. When you have plotted the points, don't forget to join them with straight lines. Take care to plot the mid-points correctly. You will be given a frequency table listing the frequencies and the class intervals. For example, here's part of the table for 'Times for a Race':

Time t (minutes)	Frequency f
$10 \leq t < 11$	3

Some students mistakenly plot the frequency against one of the class boundaries (10 or 11). You must work out each mid-point. The mid-point of the $10 \leq t < 11$ class is between 10 and 11, at 10.5 minutes, and so the point is plotted at $t = 10.5$ minutes and $f = 3$.

It is conventional on a frequency polygon to extend the polygon by one class interval beyond the first and last mid-points, so that the graph touches the x-axis. In the graph above, a line from the mid-point at $t = 7.5$ minutes is drawn one class interval (in this example 1 minute) to the left, to hit the axis at $t = 6.5$ minutes. The polygon is the shape formed by the line of the x-axis below and the straight lines of the graph above.

In an exam question you may be given a table of grouped data and asked to draw a **frequency diagram**. You can choose to draw a frequency polygon **or** a histogram. Most people find it easier to draw a frequency polygon.

Below we show just the frequency polygon. The mid-point values are plotted (7.5, 8.5 etc). The x-scale was chosen so that the mid-point values and the class boundaries are shown. It resembles a conventional x-y plot in which the frequency is plotted against time.

Frequency Polygon: Times to Run a Race

In an exam question you may be given an empty grid. Take care to label the scales correctly – you may prefer to label the mid-point values, not the class boundaries, on the x-scale, to assist in plotting the points at the correct positions.

Using Excel

Using Excel to create charts may be of interest to the reader. The frequency polygon (above) used the XY (Scatter) chart type. The corresponding Excel histogram (below) uses the Column chart type. The x-axis should be a proper scale, such as 6, 7, 8, 9 etc. Instead the mid-point values are labelled in the style of a bar chart. It is clear, though, that the class boundaries are at 7, 8, 9, 10 etc. The combined histogram and frequency polygon (previous page) used both the Excel Column and the XY (Scatter) chart types, and the x-scale was changed from mid-points to a proper continuous scale.

Histogram: Times to Run a Race

The x-axis on a histogram should be a conventional, continuous scale. The major divisions should be labelled appropriately. For example, if the values range from 2 to 9, a scale from zero to 10 is sensible. For a histogram, where the data may range from 2 to 11, you may decide to start at 2 – this will create a more convenient scale if the width of the grid is 10cm. Of course, as with all graph plotting, the scales you choose for both axes will depend on the size, shape and orientation of the graph paper as well as on the actual data.

The class boundaries in exam questions will often conveniently coincide with the scale divisions. With real data the chances are that the boundaries are not integers and will not coincide with the scale divisions. Do not be tempted to use non-integer boundaries as the basis of the scale. A sensible choice of scale takes priority.

There is no easy way, using Excel, to create a conventional histogram, with a correct scale. The x-axis in the histogram (previous page) can be replaced, in Excel, by the correct x-axis for a histogram (as on page 11). That process is laborious. The histograms (page 19) were created using Excel's XY (Scatter) chart type – even more laborious. The message to the student is, if you wish to chart a true histogram either use specialised software or draw it by hand.

Advantages of a Frequency Polygon

One advantage, from a student's perspective, of a frequency polygon compared with a histogram is that a frequency polygon is easier to draw. The **main** advantage, however, is that you can superimpose two or more frequency polygons on the same axes and make comparisons between the sets of data. The diagram shows two frequency polygons for the marks in two subjects, Maths and English, for the same group of 50 students. In this case it is quite easy to make a comparison between the two sets of marks.

Frequency Polygon: Maths and English Marks Compared

Some students did very well in maths – eight scored more than 80. You can see that the middle range of marks, about 40 to 60, is more common in the English results. The modal group is the same for both subjects ($49.5 \leq \text{mark} < 59.5$)

Data for marks in tests are discrete. Therefore take care, if you get a question on grouped, discrete data, when you work out the mid-points. The marks were grouped into classes from 0 to 9, 10 to 19, etc. Therefore the mid-point values are 4.5, 14.5, 24.5 etc. That's why the unusual scale was chosen – so that the mid-point positions can be read directly from the x-scale. A more standard scale could have been used, such as 0, 10, 20 etc.

Exercise 4 – Draw a Frequency Polygon

The table gives the times to run a race for 25 men and 25 women.

Time (minutes)	$30 \leq t < 40$	$40 \leq t < 50$	$50 \leq t < 60$	$60 \leq t < 70$	$70 \leq t < 80$
Frequency: men	0	4	3	8	10
Frequency: women	2	6	8	6	3

Contd next page

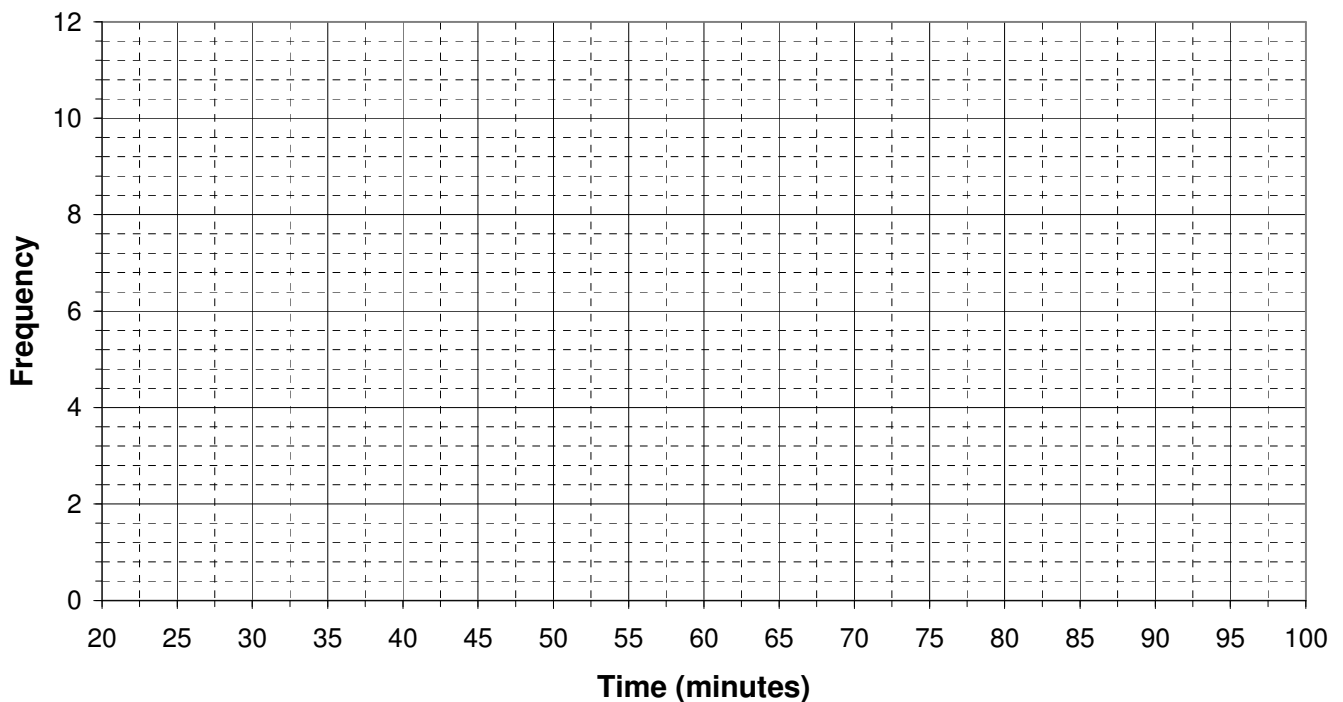
Exercise 4 continued

- a) Is the data continuous or discrete?
- b) On (a copy of) the grid provided below, draw separate frequency polygons for the men and the women. Label the diagram fully.

Note: you may wish to use a copy of this table to record the mid-points before you plot the graph.

Time (minutes)	$30 \leq t < 40$	$40 \leq t < 50$	$50 \leq t < 60$	$60 \leq t < 70$	$70 \leq t < 80$
Frequency: men	0	4	3	8	10
Frequency: women	2	6	8	6	3
Mid-point					

Comparison: Times to Run a Race



Summary

The main points about a frequency polygon are that:

- a frequency polygon is a frequency diagram. It is called a polygon because of its shape.
- the frequencies are plotted against the mid-point values – the plotted points are then joined by straight lines – the result resembles a standard x-y graph.
- it is normally used for grouped data and is an alternative to a histogram.
- the advantage of a frequency polygon is that you can easily compare two sets of related data on the same chart – for example, the marks in a test for 50 men and 50 women.
- not an essential point for GCSE – the area of the polygon equals the total area of the bars on the corresponding histogram. You may be able to spot that on the chart, page 11.

Sample Exam Question – Frequency Diagram

The question is based on AQA GCSE Maths Module 1 Intermediate, March 2005, Section B, question 7.

<http://www.aqa.org.uk/qual/gcse/qp-ms/AQA-33001I-W-QP-MAR05.PDF>

The frequency table shows the prices of over 100 models of TV. Draw a **frequency diagram** to represent the data. (3 marks)

Price £x	Frequency
$150 \leq x < 200$	15
$200 \leq x < 250$	36
$250 \leq x < 300$	62
$300 \leq x < 350$	18
$350 \leq x < 400$	10

The question includes a grid to draw the chart – but the axes are not labelled. The grid and a summary of the examiner’s report are on the next page.

Things to notice about this question:

- The price is represented by x .
- The class intervals are shown as, for example, $150 \leq x < 200$. Don’t be put off if, in a question, the \leq and the $<$ are swapped, as in $150 < x \leq 200$. See the note on the next page.
- The data are continuous and are grouped into five equal classes, width £50.
- The question asks for a frequency diagram. You can choose to draw a histogram or a frequency polygon.
- In this question, label the y-axis **Frequency** and the x-axis **Price (£)**. If you draw a histogram, do **not** use frequency density (Higher tier only). This question expects you to use frequency. It will be clear, from the question, if you are expected to use frequency density. If the classes have varying widths, use frequency density.
- You must decide on suitable scales for both axes, e.g. £100 to £450 on the x-axis.
- You must label the scale divisions, e.g. 100, 150, 200, 250 etc. on the x-axis.

The marks are allocated as follows:

1 Mark

Suitable scales for both axes.
Frequency scale from 0 (origin).
Price scale linear.

1 Mark

Points plotted at correct heights (correct frequency to the nearest $\frac{1}{2}$ division).

1 Mark for histogram

Bars located with class boundaries alongside correct x-values.

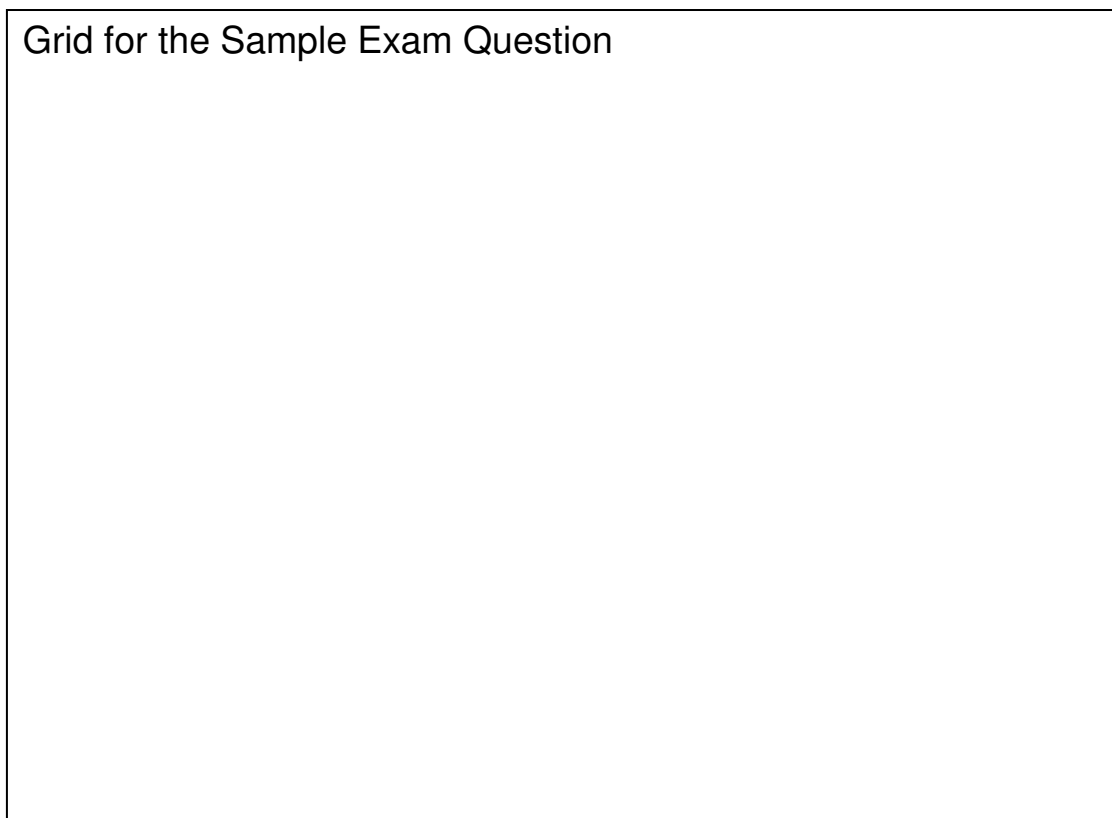
or 1 Mark for frequency polygon

plot mid-points and join points
– best to label mid-point values on x-scale.

Summary of AQA Examiners Report on the Maths Module 1 Intermediate, March 2005, Question 7

The report mentions that a few candidates got the axes the 'wrong way round'. Most correctly used a linear y-axis with the scale divisions equally spaced, but most did not use a continuous x-axis. Many plotted the heights correctly and most made only a single error. Two common errors with **frequency polygons**: didn't join the points; didn't plot the mid-points.

Grid for the Sample Exam Question



Note about Class Boundaries

In an exam question you may see a slight variation on the definition of a class interval. In the sample exam question (previous page) one of the intervals is shown as $150 \leq x < 200$. It means that, for continuous data, values from 150 up to but not including 200 are in the class.

An alternative is to group as in $150 < x \leq 200$. It means that 200 *is* included but 150 is not. Either of these alternatives could be used in a question. It makes no difference to the way you draw the graph of continuous data. In both cases the lower and upper class boundaries are 150 and 200, and the mid-point is 175. You label the axes in the same way. Most of the relevant questions in the exam are about continuous data – so there should be no problem.

Frequency Density

Definition

In the simplified histograms so far, the height of a bar on a histogram equals the frequency. That is OK if the class width equals one. If it is not equal to one then there is a problem (explained below). The solution is to divide the frequency by the class width, to calculate what is called the **frequency density (fd)**. The equation for frequency density is:

$$\text{frequency density fd} = \frac{\text{frequency } f}{\text{width } w}$$

For example, if the frequency is 8 and the width is 2kg, the frequency density is 8 divided by 2 ($8/2 = 4$ per kg). Don't be put off by the "per kg". The data could be the weights of sacks of flour that are grouped into classes. The class width is 2kg. There may be 8 sacks that weigh between 10 and 12 kg. The frequency for that class is 8 sacks in the 2kg width. The frequency density is $8/2 = 4$ sacks per kg width in the 10 to 12 kg class. It measures the average 'concentration' of sacks per kg width, in that particular class.

The Problem with Frequency

Frequency density is better than frequency. Using frequency density, you can make comparisons between data that is grouped into different class widths. The idea is similar to the unit price of a product. Cat biscuits are sold in 2kg bags – nothing to do with the sacks of flour. To compare the price of the biscuits with other brands, you calculate the price per unit weight (per kg). Divide the price by the weight to give the price per kg.

Another similar example is weight (mass) per unit volume, a measure of the 'concentration' of mass called the density of a material. A comparison of the 'heaviness' of materials is possible using density, because you are comparing the weights of equal volumes. In order to compare histograms, you use frequency density. Frequency density is a standardised frequency that allows you to draw and compare histograms using different class widths.

For an example, we'll use the 'Distance Travelled to Work' grouped data that we used in the histogram on page 4. The classes are all the same width ($w = 2$ miles)

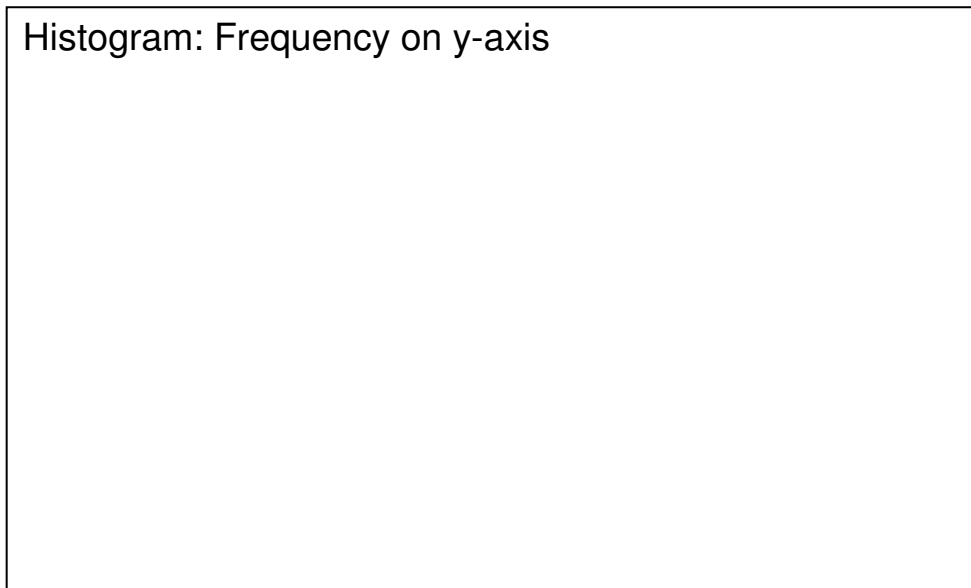
Distance d (miles)	Frequency f	Class Width w (miles)	Frequency Density fd (per mile)
$1 \leq d < 3$	60	2	30
$3 \leq d < 5$	150	2	75
$5 \leq d < 7$	120	2	60
$7 \leq d < 9$	90	2	45
$9 \leq d < 11$	30	2	15
$11 \leq d < 13$	12	2	6

To calculate the frequency density, divide the frequency by the class width. For the 60 people who travel between 1 and 3 miles to work, the frequency density = $60/2 \text{ miles} = 30$ people per mile. That means that, on average, 30 employees live between 1 and 2 miles and 30 live between 2 and 3 miles from their work.

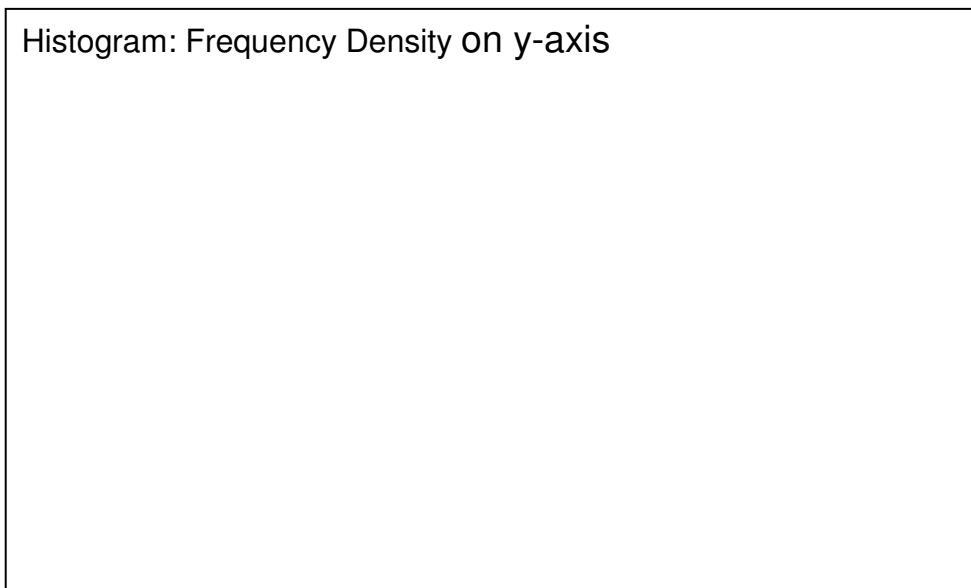
On page 4 the histogram was drawn using the frequency on the y-axis, as shown on the next page. The height of each bar equals the frequency. That works for a bar chart, but you will see, in the next section, that it does not work for a histogram.

Comparison of Frequency with Frequency Density

This is the simplified histogram from page 4, drawn with the height of each bar equal to the frequency. The class width is two miles, there are six classes. The first bar represents the 60 people who live between one and three miles from work.



The correct histogram (below) uses the frequency density for the height of each bar. What effect will that have? The two diagrams appear identical apart from the scales on the y-axis. One difference, as we'll see, is that frequency now relates directly to a bar's area.



Using frequency instead of frequency density on a histogram is not completely wrong. Provided that all the bars are the same width (they are in this example) then the frequency and the frequency density are proportional. Your histogram will then have the correct shape, as you can see. The problem is that, if you plot the frequency, the heights of the bars are not standardised. In this example the bars are too high. This makes two histograms, with different class widths, difficult to compare, as we shall see.

On the next page, to show the effect of changing the class width, the 'distance travelled' data are regrouped into 1mile class widths. Then the two histograms, for the 2mile width and the 1mile width, are compared.

The Effect of Class Width

Two researchers independently analyse the same raw data for the 'distance travelled' histogram. The 1st researcher groups the data into 2mile widths as in the histogram on the previous page. The 2nd uses 1mile widths. So far, there is no problem.

Distance d (miles)	Frequency f	Class Width w (miles)	Frequency Density fd (per mile)
1≤d<2	20	1	20
2≤d<3	40	1	40
3≤d<4	65	1	65
4≤d<5	85	1	85
5≤d<6	70	1	70
6≤d<7	50	1	50
7≤d<8	45	1	45
8≤d<9	45	1	45
9≤d<10	20	1	20
10≤d<11	10	1	10
11≤d<12	4	1	4
12≤d<13	8	1	8

The data as grouped by the second researcher, into 1mile widths. Because the width is 1mile, the frequency and the frequency density are equal. We haven't discussed why they have grouped into different widths. This is just an example. However, it is worth noting that, by adjusting the class width, you can improve the shape of a histogram.

The diagrams below show the 1mile and the 2mile histograms superimposed. Those two histograms should have similar shapes and heights, because they are based on the same raw data. On the 1st diagram, frequency is used on the y-axis. As a result the histograms are difficult to compare, because the 2mile histogram with the blue border is the wrong height. It is, on average, twice as high as the 1mile histogram. That's like concluding that petrol is twice the price, because the price quoted is for two litres, not per litre. The 2nd diagram uses frequency density. Now the similarity of the two histograms is clear.

Comparison of the Histograms

Histogram: Comparison of Histograms for 1mile and 2mile Width

The Area of a Histogram

A typical exam question will ask you to calculate a frequency using a histogram. The method uses the *area* of the histogram. That requires some explanation. The *height* of a bar on a histogram is the frequency density, and not the frequency. How then do we represent frequency on a histogram? The answer is that the frequency is the *area* of the bar.

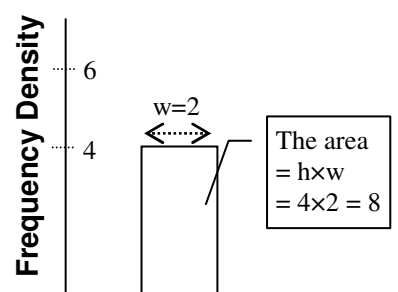
height of bar equals **frequency density**
area of bar equals **frequency**

This is confusing if you are not familiar with ‘area under a graph’. You may know, from speed-time graphs, that the ‘area’ equals the ‘distance travelled’. If so, the good news is that histograms use area and that the maths is similar. Most students, however, meeting histograms for the first time, probably have no experience of the use of area under a graph.

Here’s a simple example to explain the area. We’ll look at just one bar from a histogram. We’ll call it a rectangle because rectangles have an area. The frequency f is 8, the class width is 2. The frequency density is $fd = f/w = 8/2 = 4$.

The height h of the rectangle is the frequency density. $h=4$. Area equals length multiplied by width. In this case, that’s height multiplied by width. The area $A = h \times w = 4 \times 2 = 8$.

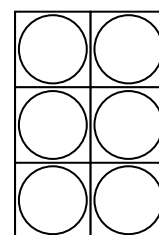
Therefore, the area of the rectangle equals the frequency. This result follows directly from the definition of fd . The height of the rectangle is f divided by w ($h=f/w$), therefore the frequency is the height multiplied by w ($f = h \times w = \text{area}$).



We could reverse the logic. We could start by defining a histogram as a frequency diagram in which the frequency is represented by the area of the bar. It then follows that the height of the bar equals the frequency density ($h = A/w = f/w = fd$). If that definition is used in an introduction it would be very difficult to see the link between a bar chart and a histogram. With the advantage of hindsight, however, that alternative definition may be preferable.

If you are used to calculating area for real things, such as the area of tiles, it may seem odd that an area equals a frequency. Remember, though, that the rectangle just represents a class on a histogram. It is not a real object. It has ‘height’ and ‘width’, but those are not real lengths, measured in metres. The ‘area’ is not a real area.

The area of a rectangle on a histogram is the frequency, which is a count of the number of data items in the class. This example shows a more familiar example of how a count of items can be equated to an area. The items are canned drinks. This rectangle measures 3cm by 2cm. The area = $3 \times 2 = 6\text{cm}^2$. The rectangle could represent a box for storing canned drinks. There are $3 \times 2 = 6$ cans.



Clearly the number of cans behaves like an area – but the number doesn’t actually equal an area in cm^2 . You wouldn’t say that the number of cans in the box is 6cm^2 . Each of the 1cm squares simply *represents* 1 can.

Instead of saying ‘the area of a bar is the frequency’, some people prefer to say that the area of a bar is *proportional* to the frequency. It depends on what you mean by ‘area’. If the bar’s width is 2 (in kg) and the height of the bar is 4 per kg, then the calculation is:

$$\text{frequency} = \text{area} = h \times w = 4 \times 2 = 8$$

The area is 8 and the frequency is 8. If a second bar has twice the area, the area is 16 and so the frequency is 16. However, if you ‘measure’ the area of the bars by counting the squares on the graph paper, the area you get depends on the size of the squares. Suppose you count 1cm squares. The area of the first bar ($f=8$) could be two 1cm squares (area = 2). Now the area and the frequency are not equal, they are proportional. The area of the second bar is double – that’s four 1cm squares. Therefore the frequency is double: $f = 8 \times 2 = 16$.

Therefore remember: if you count squares for the area, the area of a bar is *proportional* to the frequency, and so the ratio of two frequencies equals the ratio of the areas of the corresponding bars. The section [Exam Questions on Frequency Density](#) shows how frequency, frequency density and the area of a histogram are used in exam calculations.

Modal Class

The mode, along with the mean and the median, are three of the important statistics used to summarise a data set. They were explained using examples of ungrouped data. For example, as a very simple reminder, if you have a sample of just three values, 1, 2, and 3, the median (the middle value) is 2. The mean is also 2. There is no mode. You will recall that the mode (also called the modal value) is the value that occurs most often. For the set 1, 2, 3, 3, 5, 6, 7, the mode is 3 (there are two 3s). For a large set of data, once the data are grouped, and without the raw data, there is no way of knowing the modal value. In fact, in that situation, individual values are not important. Knowing the class that contains most data is more useful. That class is called the *modal* class.

For the Employees example (see pages 18-19), 150 employees live between 3 and 5 miles from work ($3 \leq d < 5$). That is the largest group and hence the modal class. Take care with grouped data when the groups are different widths. The modal class is the class with the largest frequency density, **not** necessarily the largest frequency. In the Employees example, the classes are all the same width and so the class interval $3 \leq d < 5$ has the largest frequency and the largest frequency density (see table, page 18).

Summary

The main points about frequency density and histograms are:

- the area of a bar on a histogram is the frequency.
- the ratio of two frequencies equals the ratio of the areas of the corresponding bars.
- the combined area of all the bars equals the total number of data in the sample.
- frequency density is the frequency per unit class width ($fd=f/w$).
- Because the area is the frequency, the height of a bar is the frequency density – therefore, label the y-axis frequency density, not frequency.
- the modal class is the class with the largest frequency density.

See [Appendix A](#), Explanation of Frequency Density, for a visual explanation of frequency density, based on the idea of a histogram bar as a box that holds the data.

We have justified frequency density with mainly practical, not theoretical reasoning. For example “frequency density allows you to use different class widths”. For a mathematical explanation of frequency density, based on a comparison between a histogram and a standard x-y graph, see [Appendix C](#), Is a Histogram a Proper Graph?

Exam Questions on Frequency Density

It will be clear from the question if you are expected to use frequency density. It is almost certain that a Higher tier question, that specifically mentions a histogram, will use frequency density on the y-axis. If the bars (the classes) are the same width, the height of a bar is proportional to the frequency it represents. In a typical question the bars are not all the same width (see the sample exam question, next page). Then you use the areas of the bars, **not** the heights, to compare the frequencies.

Usually the chart's grid consists of 1cm squares. In the example below the histogram shows the weights of items. The two shaded boxes are different widths. Count the 1cm squares. The short shaded box is 2 squares. The tall shaded box is 4 squares. You are told that the number of items (the frequency) in the tall box, from 3.5 to 4kg, is 10. How many items are in the short box, between 1 and 2kg? Compare the areas. The short box is half the area and therefore half the frequency. The frequency for the short box is half of 10. Write that as an equation: $f = \frac{1}{2} \times 10 = 5$. The number is 5.

Histogram: Higher Tier, Different Class Widths - A Typical Question

You may prefer a variation on the calculation. You can start by calculating the number of items in one square. You know that there are 10 items in the 4 squares of the tall box ($f=10$), and so calculate the number of items per 1cm square.

The number of items in 1 square = $10/4 = 2.5$

Therefore, in a box with two squares, the number of items = $2 \times 2.5 = 5$

Exercise 5 – Reading a Histogram

For the histogram above, each 1cm square represents 2.5 items of data. Calculate the number of items that weigh between 4.5 and 6kg.

Sample Exam Question – Histogram

The question is based on the AQA GCSE Maths Module 1 Higher, March 2007, Section B, question 10.

<http://www.aqa.org.uk/qual/gcse/qp-ms/AQA-43001H-W-QP-MAR07.PDF>

The question is typical of the Higher paper. You are given a completed histogram. It is likely, as in this case, that the y-axis has no scale.

Histogram: A Typical Exam Question

The histogram represents the weights of 60 babies and 6 babies weigh from 4 to 4.5kg. Calculate the number of babies weighing less than 3kg.

Things to notice about this question:

- The class width (the width of the bars) is not constant.
- The y-axis is labelled Frequency Density. The question will test your understanding of the link between frequency density and frequency.
- There is no scale on the y-axis.
- Of the alternative solutions below, candidates generally prefer the ‘areas’ method.

Solution

Either use **areas** or **frequency density**.

Areas 6 babies so:

Frequency from 4 to 4.5kg = 6.

From 4 to 4.5kg = 2.4 squares.

Frequency (f) is proportional to number of squares. 6 babies = 2.4 squares.

Therefore $6/2.4 = 2.5$ babies per square.

1 to 3kg is $2+2.4+4.4 = 8.8$ squares.

Babies from 1 to 3kg = $8.8 \times 2.5 = 22$.

Frequency Density 6 babies so:

Frequency from 4 to 4.5kg = 6.

Frequency density = f/w . Width is

0.5kg so $fd = 6/0.5 = 12$ (height of bar)

First 3 bars, from 1 to 3kg. Heights are 5, 12, 22.

The frequencies are ($f=fd \times w$)

$5 \times 1 = 5$, $12 \times 0.5 = 6$, $22 \times 0.5 = 11$.

Babies from 1 to 3kg = $5+6+11 = 22$.

Answer: 22 babies weigh less than 3kg

Answers to Exercises

Exercise 1 – Grouping a Continuous Variable

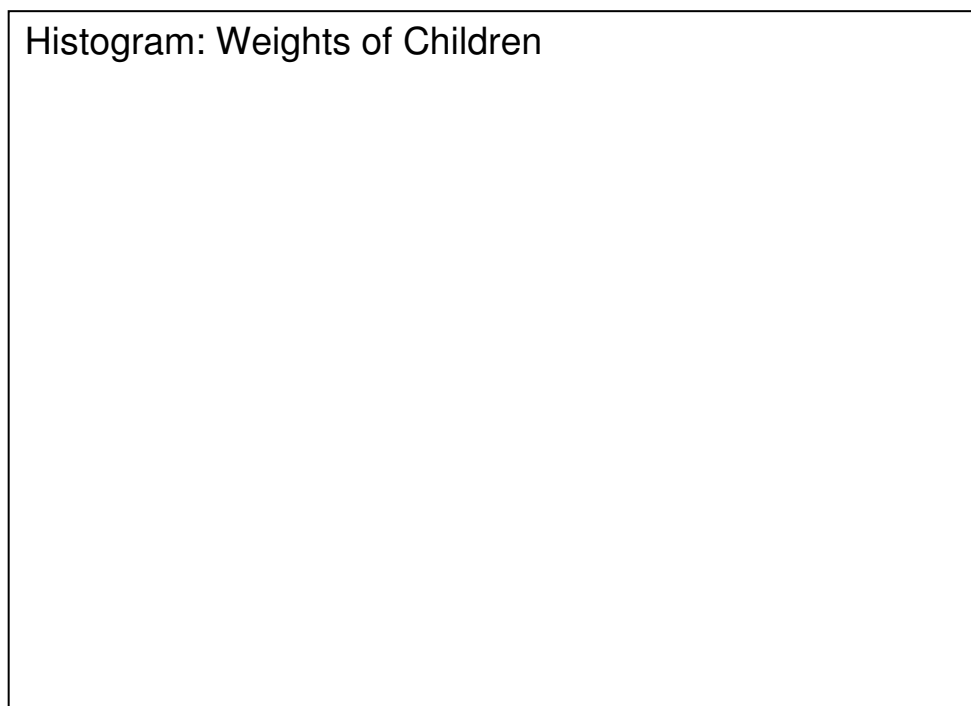
Class Interval	Class Interval	Values of x	f
90 –	$90 \leq x < 100$	90, 97	2
100 –	$100 \leq x < 110$	104	1
110 –	$110 \leq x < 120$	113, 119	2
120 –	$120 \leq x < 130$	120, 123, 125	3

Exercise 2 – Histogram

Times Taken to Run a Race

Time t (minutes)	$7 \leq t < 8$	$8 \leq t < 9$	$9 \leq t < 10$	$10 \leq t < 11$	$11 \leq t < 12$	$12 \leq t < 13$	$13 \leq t < 14$
Frequency	3	5	4	3	6	2	1
Mid-point (minutes)	7.5	8.5	9.5	10.5	11.5	12.5	13.5

Exercise 3 – Draw a Histogram



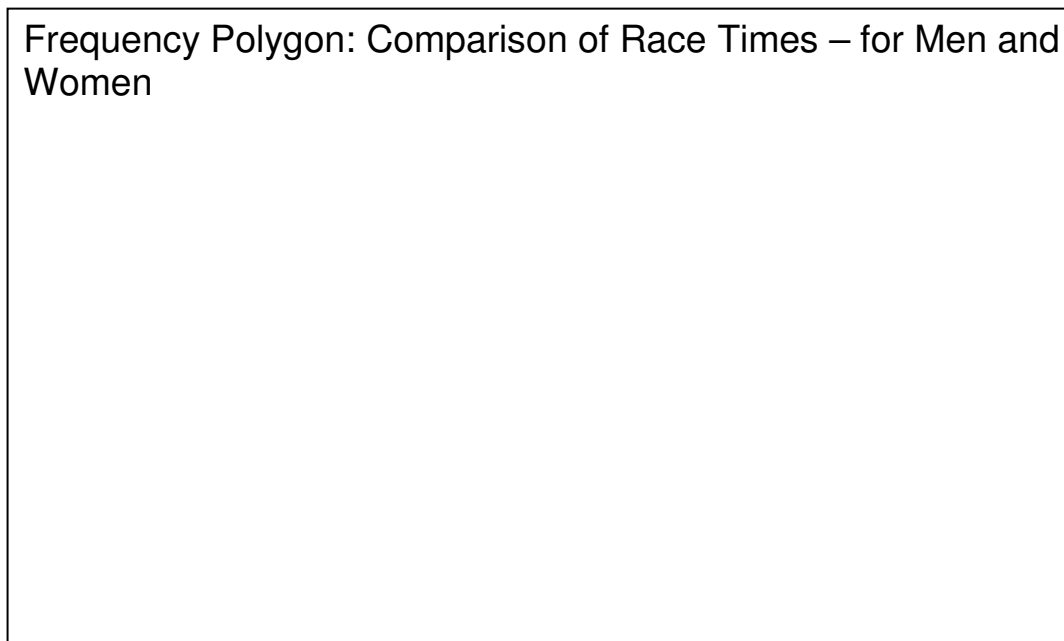
This frequency diagram is based on the table of data given in the question. The frequency is plotted (y-axis). In the Higher tier you are expected to plot frequency density. The class width is 5kg, therefore the frequency density is frequency divided by 5. For example, for the first bar, the frequency density is $15/5 = 3$ per kg.

Exercise 4 – Frequency Polygon

a) Is the data continuous or discrete?

Time is a **continuous** variable.

b) On (a copy of) the grid provided, draw separate frequency polygons for the men and the women. Label the diagram fully.



This frequency diagram is based on the table of data given in the question, shown below with the mid-points included. Check that you plotted your diagram at the mid-points.

Time (minutes)	$30 \leq t < 40$	$40 \leq t < 50$	$50 \leq t < 60$	$60 \leq t < 70$	$70 \leq t < 80$
Frequency: men	0	4	3	8	10
Frequency: women	2	6	8	6	3
Mid-point	35	45	55	65	75

Exercise 5 – Reading a Histogram

The question states:

For the histogram above, each 1cm square represents 2.5 items of data. Calculate the number of items that weigh between 4.5 and 6kg.

The histogram has two classes between 4.5 and 6kg. The areas are 2 and 1.6 cm². The combined area is 3.6 cm².

Therefore number of items = number per square cm \times area in cm².

$$= 2.5 \times 3.6 = \underline{9 \text{ items}}$$

Appendices

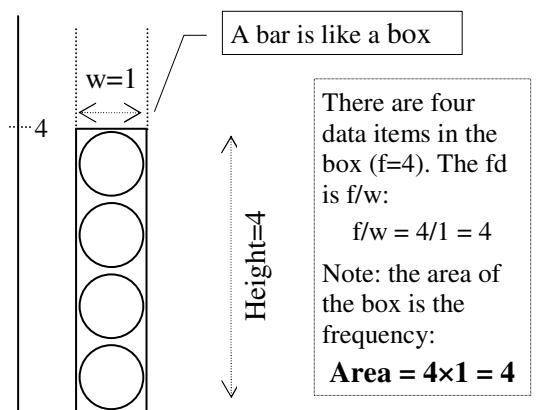
Appendix A – Explanation of Frequency Density

We'll use a wine collection as an example. The wine is grouped according to age into 2year classes (width $w=2$ years). The 10 to 12 year old class contains 8 bottles ($f=8$). There could be 3 bottles between 10 and 11years old, and 5 bottles between 11 and 12years old. The frequency density is $fd = f/w = 8/2 = 4$ per year. What does that mean? The '4 per year' means that the average number of bottles per year, in the 10-12years class, is 4. In this example, there *are* actually 4 bottles in each of those years. Now, imagine that we initially grouped into 1year class widths. The 10-12years class would be divided into two classes, from 10-11years and from 11-12years, with 4 bottles in each.

To understand frequency density, we'll look at the effect of re-grouping those two, 1year classes into the single 10-12years class. The 1st histogram below shows just the 10-11years class containing 4 bottles. Should the y-axis be frequency or frequency density?

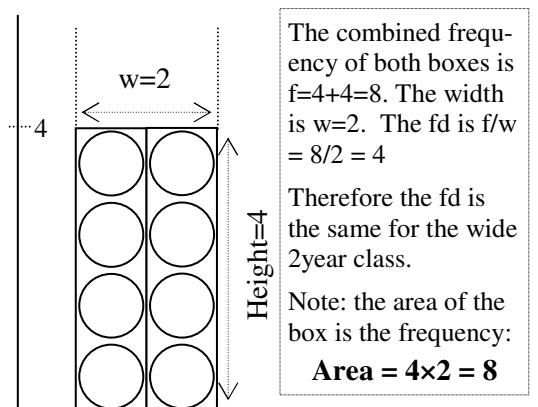
It helps to think of the bar as a tall box. The stack of circles is a pictogram of bottles inside the box. The box is like a wine rack. It contains 4 bottles ($f=4$) and so the height of the box is 4. The class width is 1year, and so the frequency density is $fd=f/w=4/1=4$ per year.

Therefore, for a 1year width, the frequency density equals the frequency and so the y-axis could be labelled frequency or frequency density. The area of the box is the height multiplied by width = $4 \times 1 = 4$.



The 2nd histogram shows the 10-11years class and the next one along, the 11-12years class. There are 4 bottles in each box and so the height of both boxes is 4. There are 8 bottles in total. Suppose we now re-group the data into the 10-12years class that we started with.

You can think of this diagram in two ways. Either it shows the two narrow boxes (classes), each width $w = 1$, or it shows a single, wide box (class), width $w = 2$. The frequency for the wide class is $f = 4+4 = 8$. If the box's *height* is the frequency, you can see that the wide box would have to be 8 bottles high. But the 8 bottles do not stack vertically in the wide box – they stand in two stacks of 4, side by side, to fill the box's width. The area $A = 4 \times 2 = 8$. The height is $A/w = 8/2 = 4$.



The diagram shows that the area of a bar is the frequency. The area of the 1year width is $h \times w = 4 \times 1 = 4$. For the 2year width, the area equals $4 \times 2 = 8$. When you combine the two classes, the areas add. Clearly, the area equals the **frequency**. The height of a bar is **not** the frequency. The **height** ($h = A/w = 8/2 = 4$) equals the **frequency density**.

Therefore a histogram is drawn with bars to represent the classes. The width of a bar equals the class width. The area of the bar is the class frequency. The variable on the y-axis (the height of the bar) is the area divided by width, called the class frequency density.

Appendix B – Variable Class Width

You have seen that if you use frequency on the y-axis of a histogram, there is a problem with different class widths. This isn't just a problem when comparing two histograms that use different class widths. Different widths are commonly used within the same histogram.

For example, in the first example of a histogram, distances travelled to work were grouped into equal 2mile widths. You don't have to use equal widths. You could create classes from 1 to 4miles, then 4 to 6miles, then 6 to 7miles, then 7 to 8miles, then 8 to 10miles and finally 10 to 13miles. The class widths would be 3, 2, 1, 1, 2, and 3miles. The total range is still from 1 to 13miles.

There are a number of reasons why a histogram is drawn with different class widths.

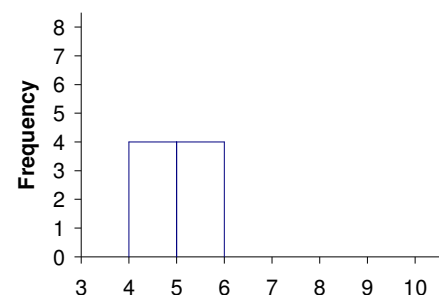
- The outer groups typically contain relatively few data, whereas the central part of a histogram contains most of the data. If you use narrow groups, the height of the outer bars will fluctuate randomly so that a pattern cannot be seen. If the groups are too wide you can miss some of the finer detail of the pattern. To make the shape of a histogram clearer, you choose the class widths carefully – not too narrow and not too wide. For example, you may use narrow groups in the central region and wider groups for the outer regions.
- The data may already be grouped and you do not have the original raw data.
- Sometimes the class widths are pre-determined. This is similar to the previous reason. If the data are about students, you may group according to age, with 1 year class widths. However, the data may be available in natural groupings that are generally not equal width, such as pre-school, early primary, late primary, 11-16yrs, 16-18yrs and adult students.

The definition of a histogram, in which the area, not the height, represents frequency, is better, because then the overall height and shape of the histogram do not depend on the class widths you use. To understand this point you have to think of a histogram as a picture that is characteristic of the data. Both its shape and its height are important. By using frequency for the area of each bar, and frequency density for the height, the histogram retains its recognisable shape and its correct height when the class widths are adjusted.

With different class widths on a histogram, the use of frequency for the height of the bars would be like viewing a histogram through a distorting mirror. The heights of the wider bars grow out of proportion and the shape is not recognisable. The use of frequency density scales the height of each bar according to its class width.

We'll illustrate this with an example similar to Appendix A, where we compared the bars on a histogram to boxes for holding data.

In this histogram we'll start with just the first two bars. They are the same height and width. The height $h=4$ and the width of each bar is $w=1$. Because the width is one, the frequency equals the frequency density. $fd = 4/1 = 4$. You can therefore use frequency or frequency density on the y-axis. Here we have used frequency because the intention is to show the effect on the shape.

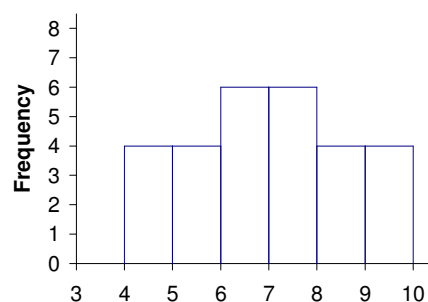


We'll add the rest of the data to the histogram, shown in this table. It doesn't matter what the values represent. We'll call the variable x .

	w	f	$fd = \frac{f}{w}$
$4 \leq x < 5$	1	4	4
$5 \leq x < 6$	1	4	4
$6 \leq x < 7$	1	6	6
$7 \leq x < 8$	1	6	6
$8 \leq x < 9$	1	4	4
$9 \leq x < 10$	1	4	4

There are six classes. The first class is from 4 to 5 ($4 \leq x < 5$). The widths are all equal to one and so again, the frequency densities (fd) equal the frequencies (f).

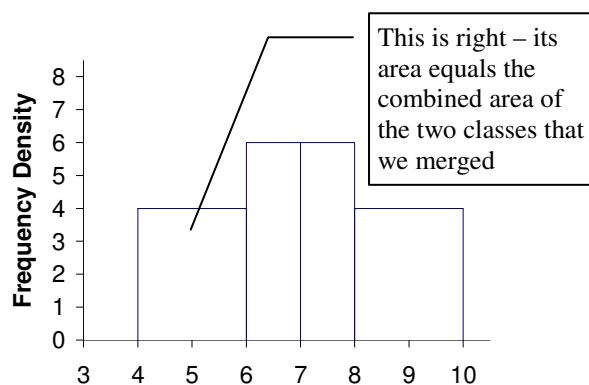
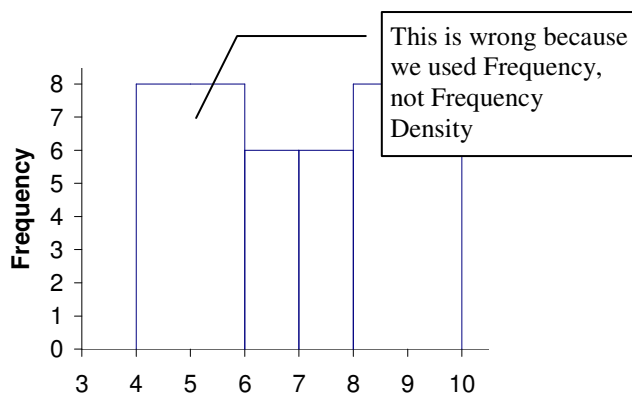
The diagram will be identical if we plot frequency density instead of frequency. The heights of the 1st two boxes are equal, as are the middle two and the last two. That is not necessary but the shape is easy to recognise and the explanation should be easier to follow. Now we'll group differently. We'll merge some of the classes to see the effect on the shape.



The data from the first two classes are merged into a single class from 4 to 6. The frequencies add, e.g. $4+4=8$. Similarly the last two are merged into a class from 8 to 10.

	w	f	$fd = \frac{f}{w}$
$4 \leq x < 6$	2	8	4
$6 \leq x < 7$	1	6	6
$7 \leq x < 8$	1	6	6
$8 \leq x < 10$	2	8	4

The histogram, below left, is now the wrong shape. The two wide bars are taller than the central narrow bars, because we've used frequency for the height of a bar. If the classes are different widths then, to keep the correct shape and height, the histogram must be drawn using frequency density – below right. When classes (bars) are merged their areas add (not their heights), so that the area of the new class equals the total area of the merged classes.



Appendix C – Is a Histogram a Proper Graph?

This section is included for those who have learned about histograms and who can successfully answer GCSE histogram questions, and yet are still left wondering just where the histogram fits in relation to a more familiar, standard x-y graph.

A Standard x-y Graph

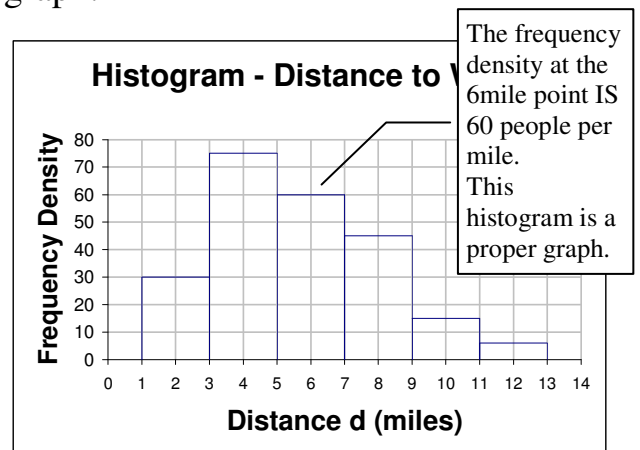
Here's an objection to a histogram from such a student. The histogram in this example uses frequency on the y-axis:

"I don't understand the histogram. Compare it with a typical x-y graph, such as a distance against time graph. For each time value there is a corresponding distance. For a car travelling at a constant 60km per hour, when the time is 30minutes on the x-axis, the distance travelled will be 30km, shown on the y-axis. This means that 30minutes into the journey the car has travelled 30km. What happens on a histogram? The Distance to Work histogram shows frequency and distance. At 6miles from work the histogram shows $f = 120$ people. Does that mean 120 people live 6miles from work? No, it doesn't. The 120 people live between 5 and 7 miles from work. It appears that a histogram does not behave like a normal graph."

Comparison: Histogram and Distance-Time Graph

The student is correct. A histogram that uses frequency is a strange graph. This point is more likely to be raised by those with experience of graphs. It is inevitable that, if you group data and count the frequency in each group, the resulting frequency histogram is going to be different from a more conventional x-y graph.

The matter could rest there except for frequency density. Frequency density is required by the exam specification. Students are expected to understand that a histogram must use frequency density. The student's objection to the frequency histogram is the key to understanding why a histogram uses frequency density. A histogram that uses frequency density, as opposed to a histogram that uses frequency, is essentially a regular x-y graph.



a proper histogram using frequency density is a proper x-y graph

It turns out that a true histogram is very like a standard graph. In any learning process it is important to make connections between apparently unrelated concepts. The histogram is not such a different 'beast' from other graphs after all.

The student's problem with the frequency histogram is that a proper graph, that records a point $y=120$ against $x=6$, would imply that the 120 corresponds to the 6. For example, 120 people successfully competed in a 6mile swim or 120 people live up to 6miles from work. The histogram does not mean that 120 people live exactly 6 miles from work. If that were the intention then it would make no sense. Of course, we know how to interpret the histogram. The 120 people is the number within the 5 to 7miles class. You can only count the number in a range, as in 5 to 7miles. That is precisely the student's point. It is a non-standard graph.

We've said that it makes no sense to talk about the number of people living exactly 6miles from work. That needs some explanation. You can only make a sensible count of continuous data over a range of values (a class). There will be no more than one item at a precise point of a continuous scale, such as a 6miles point. The chance of finding just one item at a particular point is very small. That is why, for continuous data, the data are grouped. For example, you cannot have two sacks of flour that weigh exactly 20kg? The probability is zero. If one sack weighs exactly 20kg the other could weigh, for example, 20.1kg. Another could weigh 20.01kg. In theory, for a continuous variable, there are an infinite number of possible values. The probability that two values are the same is zero. The probability that two people travel exactly 6miles to work is zero.

An easier example to visualise is hawks that hunt along the verges of motorways. There could be fifteen hawks in one 10km stretch. The concentration is $15/10 = 1.5$ per km. That frequency density, 1.5 per km, is an estimate of the hawk concentration at all points along that 10km length. Therefore, although you cannot sensibly measure the frequency at a point, you can measure the concentration (the frequency density) at a point. You could argue that because the value, 1.5 per km, is only an average for the 10km range of points, the 1.5 per km is only an approximate value for points within that stretch. That's true – but how precise can you be if there are at most only about 100 hawks along the entire 100km motorway? In theory, with a large sample of data, you can narrow the range from 10km to 1km, 0.5km or even smaller. If, instead of hawks, you measure the density of wild flowers along the motorway verge, you could count maybe 253 in 10metres. The frequency density of wild flowers is therefore 25,300 per km. Because 10metres is such a small distance, the 25,300 per km density will be an accurate density for any point in that 10metre range. Therefore, in principle, you can measure frequency density at a point.

For the 5 to 7miles class in the distance travelled histogram, if you divide the 120 people by the class width (2miles) you get the frequency density, 60 per mile. Frequency density is an average for each class. The density for the 5 to 7miles class is the average density between 5 and 7 miles and so is the (approximate) frequency density for every point between 5 and 7miles. The density at the 6mile point is about 60 per mile. That is why frequency density is used on a histogram. A histogram that uses density is very like a regular graph.

In principle you can measure and therefore plot an accurate, smooth, standard x-y graph of the frequency density against position along the x-axis. In practice, in statistics it is difficult to measure the precise frequency density. The class width has to be large enough to give a reasonable count of hawks or wild flowers or sacks of flour. That is why rectangles are used on a histogram. The width of a rectangle (the interval) shows us the range of values used to calculate the average frequency density, represented by the height of the rectangle.

Therefore the frequency density is only an average for the class and therefore only an approximation to the precise frequency density at a point within the class. That is the nature of statistics. We will return to this later – first we'll forget statistics and take a close look at traditional graphs.

Comparison of a Histogram with a Standard Graph

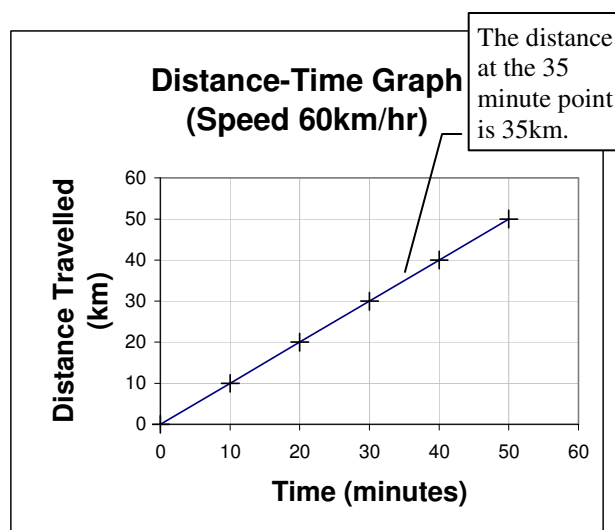
In any subject in which the content develops in logical steps – and maths is a prime example – there is always a problem because of other constraints. Ideally, a student should be thoroughly familiar with a range of basic graphing techniques before starting on histograms. If you are an adult student doing a GCSE Maths the chances are that it is 'years' since you last studied graphs, that you are on a short, one year course and, if you are doing the modular course, you started on the statistics module. That is normally manageable with minimal understanding of graphs, at least unless you are doing the Higher tier and you meet histograms. Even that is 'liveable with'. It is possible to learn the full repertoire of histogram tricks and achieve a creditable grade. Your teacher may point out that the histogram question is worth only a few marks and that you will not need to actually understand the histogram. 'Teaching to the exam' is, of course, a realistic survival strategy that many students will happily endorse.

On a short course there is room for little else. For the moment, let's put time constraints aside. GCSE Maths is primarily a course in mathematics. There is a danger that the statistics content will be presented as a set of utilitarian tools, to be used but not understood. In a maths course, understanding basic concepts is essential. Given the time, how should the histogram be taught? Ideally, it should be rooted in the context of traditional x-y graphs. This booklet has not followed that recipe, because the target students will not have a common, shared experience of graphs. For that reason, we will redress that now and attempt to relate, retrospectively, the histogram to an assumed background in standard x-y graphs. In that situation the histogram can be 'placed in context'.

The graph that we will use as a reference point for the histogram is a particular x-y plot that is studied in a basic maths or science course. It is the distance-time graph and the related speed-time graph. First, we will review the distance-time graph. If you have never met this graph, the detour into graphing a moving object is not a complete distraction, because this type of graph is part of the GCSE Maths specification.

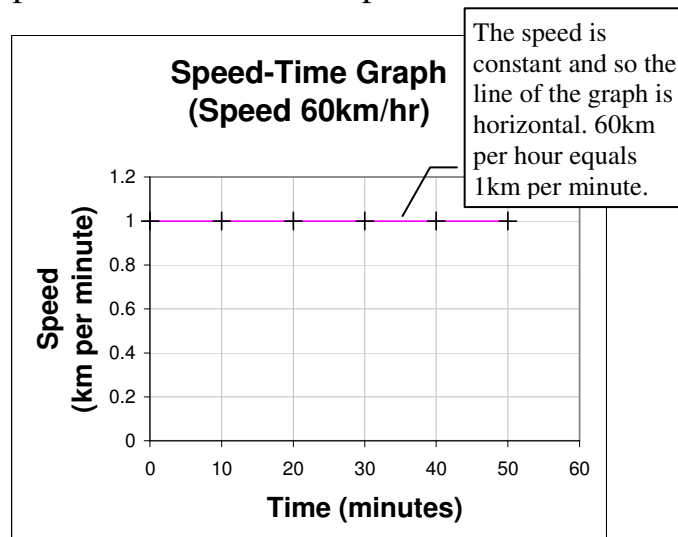
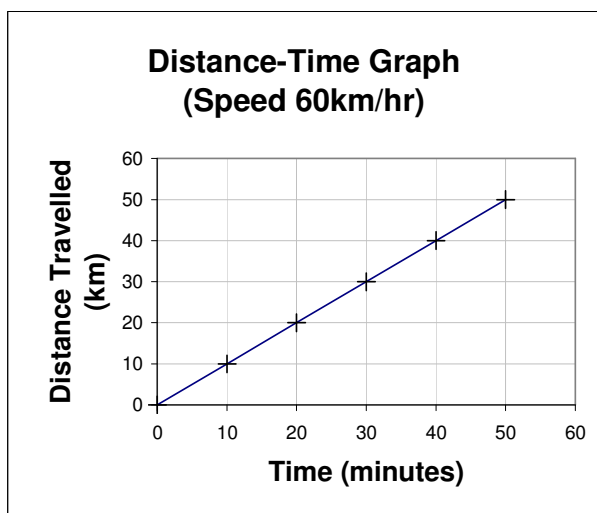
A Distance-Time Graph

We'll use the same graph used earlier as an example of a standard graph. This type of graph should be familiar to most students and it is part of GCSE Maths. The graph shows the distance travelled on the y-axis plotted against time on the x-axis. In this example it represents a car travelling at a constant speed: 60km per hour. The time and the distance are recorded from the start of the journey, where both distance and time are zero. The speed, 60km per hour, is the same as 1km each minute – after 1minute the car travels 1km, after 2minutes a total of 2km, etc.



The line of the distance-time graph is straight because the speed is constant. You may know that the gradient (the slope) of a distance-time graph equals the speed. After 10 minutes the distance is 10km, after 20minutes it now totals 20 km, and so on. For every 10minutes along the x-axis the graph increases 10km. The gradient is the increase in distance divided by the time taken = $10\text{km}/10\text{min} = 1\text{km per minute}$.

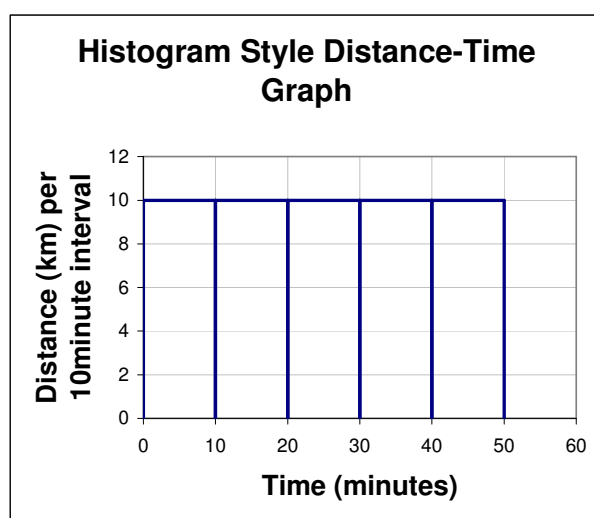
The speed-time graph (below, right) is another way to graph the car's motion. The speed on the y-axis is shown in km per minute because the time axis is in minutes. In this case of constant speed, the line of the speed-time graph is horizontal. The graph shows the speed at all times from 0 to 50minutes. You can see the speed is a constant 1km per minute.



You can use a car's odometer readings and a clock to measure the distances and times and plot the distance-time graph above. To aid the comparison with a histogram, however, we will divide the car journey up into intervals rather like a histogram's data is divided into classes. For the same car journey at 60km per hour, imagine that we divide the journey into consecutive 10minute intervals and measure the distance moved in each interval.

We'll use the x-axis for time, divided into 10minute intervals, and the y-axis for distance similar to the distance-time axes above. Unlike the distance-time graph above we'll record just the distance travelled during each interval on the y-axis. We will call the resulting graph a Histogram Style Distance-Time Graph. It is not actually a histogram.

The 50minute journey is divided into five 10minute intervals. The distance travelled in the first 10minutes is 10km. That is shown by a horizontal line at a height 10km, stretching from 0 to 10minutes. Rectangles are used to emphasise the division into intervals. All the rectangles are 10km high because the distance travelled in each 10minutes is 10km. This is an unusual way to draw a distance-time graph. It is not like the distance time graph above, although you should be able to see how the two graphs are linked. It resembles a histogram.



Each 10minute interval is like a class width on a histogram. The graph resembles a histogram in which *frequency* is plotted on the y-axis. On a frequency histogram, the y-axis shows the number of data items within the width of a rectangle. This distance-time graph is similar: the y-axis shows the distance travelled within the width of a rectangle. We could, by varying the speed of the car, make the heights of the rectangles vary, as they do for a histogram. You will see that this distance-time graph has the same problem as a frequency histogram.

The rectangles on the graph act as dividers. Each rectangle corresponds to a 10minute interval. Here we have removed the vertical sides of the rectangles. That just leaves a horizontal line, formed by the tops of the rectangles. The label on the y-axis states “Distance per 10minute interval”, to remind us that you have to interpret this graph carefully. The y-axis does not show the total distance travelled. It shows just the distance moved in each 10minute interval. The total distance is obtained by adding the separate distances moved during each interval.

Link between a Distance-Time and Speed Time Graph

Now, suppose we halve the time interval from 10 to 5minutes. That is like halving the class width on a histogram from say, 2miles to 1mile. If you halve the class width the frequency will reduce. Because of the nature of statistics, you cannot say that the frequency will exactly halve when the class width is halved. The frequencies halve ‘on average’. A car travelling at a constant speed is more predictable. If you halve the time interval then the distance travelled will halve.

What will happen to the distance-time graph if the time interval is 5minutes? The car is still travelling at 60km per hour. In 5minutes it travels 5km. Therefore, the height of each rectangle reduces from 10km to 5km. The horizontal line of the graph is now at $y = 5\text{km}$.

So, as with a histogram that uses frequency, this alternative distance-time graph has a serious flaw. The height changes if you change the width of the interval. For a histogram we standardised the graph by using frequency density. We need to standardise this imaginary distance-time graph.

Distance per 5minute Interval against Time

Although this distance-time graph has nothing to do with statistics, it has the same problems that we’ve already met when using frequency on a histogram.

We have seen two problems with frequency on a histogram.

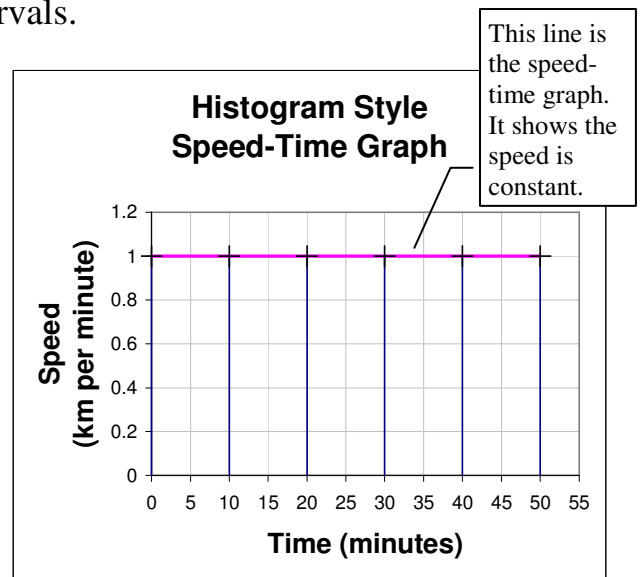
- First, you cannot use a frequency histogram to read the frequency at a particular value of the variable on the x-axis. The graph simply does not work like a normal graph. In the same way, this attempt at a distance-time graph is not a regular graph. Because the distance is divided up into time intervals, like groups in a histogram, the distance on the y-axis is not the total distance travelled at a particular time. To obtain the total distance at a point in time you would have to add together the distances travelled in each interval.
- Second, there is the related problem. The frequency on a histogram depends on the width of the class. Frequency density is like a standardised measure of frequency that is independent of the class width. Both problems are fixed using frequency density.

We need a measurement similar to frequency density to standardise the distance-time graph. Frequency density is frequency divided by class width. We will divide the distance moved (10km) in the interval by the interval width (10minutes) to standardise the distance.

$$\text{distance} / \text{time interval} = 10\text{km} / 10\text{minutes} = 1\text{km per minute}$$

Clearly you get the same result for a 5minute interval. $5\text{km} / 5\text{minutes} = 1\text{km per minute}$. Therefore, if we plot the 'distance divided by time' (the distance travelled per minute) instead of 'distance', you always get the same graph. The graph does not depend on the width of the time interval. The horizontal line of the graph is at a height 1km per minute, whether we use 5minute, 10minute or 1minute intervals.

How should we label the y-axis? We could call it 'distance per minute'. We could call it 'distance density'. After a little thought it should be clear that the quantity 'distance per minute' is simply the speed of the car, measured in km per minute. In a very circular way, we have shown that what we called the 'histogram style graph' of a distance-time relationship is, in fact, simply a standard speed-time graph. The mathematics of this speed-time graph is identical to the mathematics of a histogram plotted using frequency density.



To emphasise the similarity to a histogram we have included the rectangles (the 10minute intervals). The speed-time plot is the horizontal line at height 1km per minute. It is the same speed-time graph shown earlier. At any time along the x-axis you can read the corresponding speed on the y-axis. That may not appear very impressive given that the speed is everywhere constant on this particular graph. Constant speed was used for simplicity. Whether the speed is constant or varying, the graph shows the speed at every point in the journey. For example, the speed at 25minutes is 1km per minute. It is a regular graph. That is not quite the case with a histogram. You can read the frequency density at any point – but because of the limitation of statistics, the frequency density is an average value for each class, so the frequency density at a point is an estimate based on the class average. By collecting more data (a bigger sample) the histogram becomes more accurate.

With a big enough sample or theoretical smoothing, the grainy, low resolution ‘image’ provided by a histogram can be transformed into a high resolution, smooth curve of frequency density – a ‘proper’ graph. The lumpy rectangles are replaced by a smooth curve – part of an advanced maths course. See the bottom of this page for a smoothed histogram from a small data-set of Santa’s Presents.

Does any of this help with GCSE-level statistics? Possibly not, but the aim has been to think a little beyond GCSE and to realise that the histogram is, mathematically, not much different from graphs you may have already met. By comparing frequency in statistics with distance travelled by a moving object and the data variable on the x-axis with the time taken, we have shown that frequency density is similar to speed, the width of a class in statistics is like a time interval and that a histogram is like a speed-time graph.

One further point of similarity between a histogram and a traditional graph is the significance of the area under the line of the graph. You have learned that the area under a frequency density histogram equals the frequency. Compare that with a speed-time graph. For simplicity, we’ll stick with the graph for constant speed.

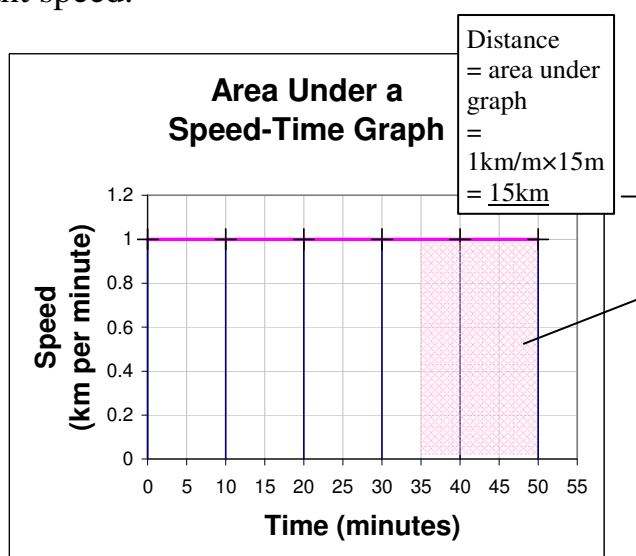
The graph is already conveniently divided up into areas. The first rectangle (0 to 10) is 1km per minute high and 10minutes wide. The area is:

$$h \times w = 1\text{km per minute} \times 10\text{minute} = \underline{10\text{km}}$$

The area is the distance travelled between 0 and 10minutes. This should not be a surprise. The equation for speed is $\text{speed} = \text{distance}/\text{time}$.

$$\text{Therefore distance} = \text{speed} \times \text{time} = 1 \times 10 = \underline{10\text{km}}$$

The area equals the distance for any time interval and for any speed-time graph. The distance travelled in the shaded area is $1 \times 15 = \underline{15\text{km}}$



To summarise, because of the identical mathematics, frequency, frequency density and their representation on a true histogram have a direct parallel with distance moved, speed and their representation on a speed-time graph. Speed is the rate of distance travelled to time taken or, more simply, distance divided by time. Frequency density is the rate of frequency to class width or frequency divided by class width. The reverse of this is that distance is the area under a speed-time graph, just as frequency is the area under a histogram. We used ideas about distance, time and speed to present this parallel between a histogram and a conventional graph, because the speed-time graph is familiar. This topic of graphs, gradients, areas and rates of change is covered in detail in introductory calculus.

