

SECTION 7

Before you go any further, go back to page 20 of Numeracy Outcome 1 pack, and remind yourself of what a histogram looks like and what makes it different from an ordinary bar chart.

A histogram is a quite specialised type of chart. You would not draw one if an ordinary bar chart would do, unless you were making a statistical analysis of some data.

The horizontal scale must be a continuous scale, so if you are asked to illustrate a table like this one, say,

Year	1990	1991	1992	1993	1994	1995
Number of cars sold	1346	1587	1904	1850	1897	2015

you would draw an ordinary bar chart because the 'year' scale is not continuous. The table shows totals for complete, discrete years, with no smooth flow between them.

Example 7a

Draw a histogram of this data which shows the reaction times (in seconds) of a group of 70 students taking part in a psychology experiment.

Reaction time (seconds)	5-10	10-15	15-20	20-25	25-30	30-35	35-40
Frequency (no. of students)	8	12	25	13	7	4	1

The first class interval (note the technical jargon) in the table is marked 5-10 seconds, and the next class interval is marked 10-15 seconds. You will notice that the first group ends on the same number as that on which the second group begins (i.e. 10).

[Another way of specifying the intervals is '5 under 10' and '10 under 15' etc.]

So in which group does a student who scored 'exactly' 10 seconds go? By convention, the first group will have all those (8) students whose time was between 5.0000 seconds and 9.9999 seconds. The next group contains those 12 students whose time was between 10.0000 seconds and 14.9999 seconds. In other words the second number which defines the interval is exclusive, whereas the first number is inclusive.

We call the numbers 5, 10, 15, etc. the class boundaries and it is these class boundaries which we mark on the horizontal axis as a continuous, regular scale.

The bars are then placed so that the edges of each bar line up with its numerical class boundaries on the horizontal axis. Like this for the first one:

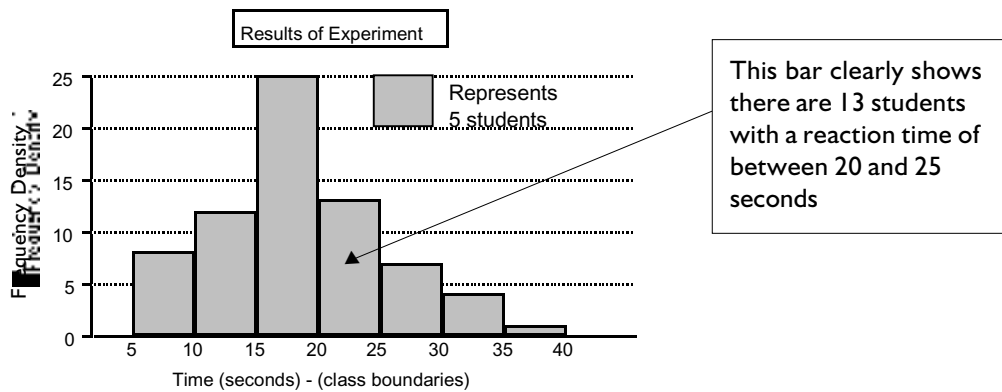
8

Again by convention, the first bar is not placed against the vertical axis but a bar-width (or so) space is left at the start.

5 10

But what is the most important difference of all? It is that the *area* is proportional to the frequency. When all the intervals are of the same width this does not matter very much, although the vertical axis must start at 0 as a result (just as with any bar chart). When the intervals are different, it's a new ball game !

Here is the completed histogram:



Notice the all-important area key, telling us what area of the graph corresponds to how many students.

Example 7b

Draw a histogram of this data, which shows the weights of various parcels in a mailing office:

Weight of parcel (kg)	10-15	15-20	20-25	25-30	30-35	35-40
Frequency (no. of parcels)	27	35	16	10	8	3

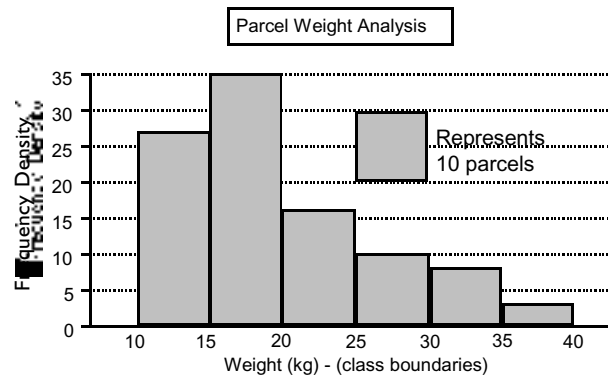
Solution:

Again, we plot the boundaries (10, 15, 20, etc.) on the horizontal axis, with the bars' edges lined up with them.

Notice again the appearance of an area key. This shows the area of the chart which represents a particular frequency. It doesn't matter how you show the area, rectangle or square, or what size, so long as its area corresponds correctly to the chart.

Note also the label on the vertical axis. Strictly speaking you do NOT write 'frequency' here, although you will see this done (wrongly) even in text books. You either write nothing at all or, if you must, 'frequency density' as shown.

Here is the histogram.



And finally, what you've been waiting for, a histogram with unequal class intervals!

Suppose you have 12 people aged 20-25, and also 12 people aged 25-35, and also 12 people aged 35-50. You notice right away that the intervals are not equal.

- 20-25 is 5 years wide, and is the narrowest
- 25-35 is 10 years wide, i.e. double the width of the narrowest
- 35-50 is 15 years wide, i.e. three times the width of the narrowest

But remember that the scale has to be regular, and your column edges line up with the boundaries 20, 25, 35 and 50. And since the frequencies are all 12, the bars will all be 12 high, right? WRONG!

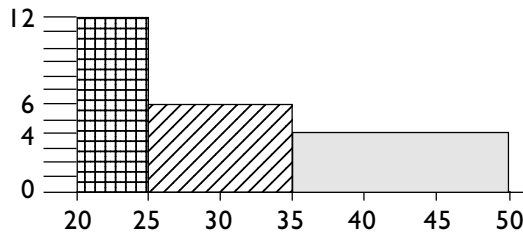
Look and see what happens if we do this:

Remember that it is the AREA which is proportional to the frequency. The first rectangle represents 12 people. The second rectangle is clearly double the area and so represents 24 people. The last one is clearly three times the area of the first and so represents 36 people. But we know that each of the three is supposed to represent 12! So what do we do about it?

To keep the area the same,

- if we double the width, we halve the height
- if we treble the width, we take one third of the height

and so on. So a correct picture based on the AREA is as follows:



Although this chart may be a bit harder to understand (remember it is a specialised one, and not one that most people have ready access to) it is now correct in that the areas of the three rectangles are equal.

Example 7c

This table shows the number of houses a solicitor has for sale, split into various price brackets. Draw a histogram to show this data.

Price Range (£)	20-40k	40-60k	60-100k	100-140k	140-200k
No. of Houses	15	27	50	35	22

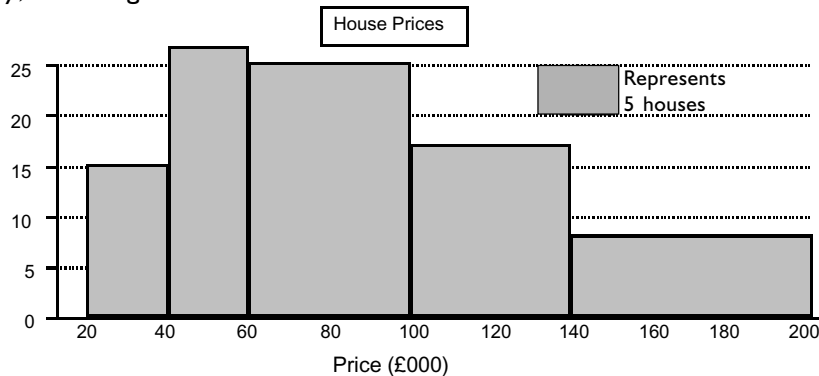
Solution:

Right away you see that the intervals are not equal. You take the narrowest interval (usually the first one) as your basic interval and base all others on this one.

It is easy enough to do the next bit in your head, but some people prefer to make a table and see all their calculations in black and white:

Price Range	Frequency	Class width	Column Height
20-40k	15	20 = narrowest	15
40-60k	27	20 = narrowest	27
60-100k	50	40 = 20 x 2	50/2 = 25
100-140k	35	40 = 20 x 2	35/2 = 17.5
140k-200k	22	60 = 20 x 3	22/3 = 7.33

and, finally, the histogram:



Did you notice that the vertical axis only has to go as high as 27 now, not as high as the highest frequency in the table (50) ?

Always check the height of your highest column before you start drawing and scaling the axes!



Draw a histogram to represent each of the following sets of data.

1. This table gives the output by machine operators in a factory over a period of time:

Units of Output	Under 300	300-310	310-320	320-340	340-380	380+
No. of Operators	7	12	19	16	10	8

NOTE: The first interval is defined 'under 300'. This is called an open-ended interval. In order to draw the histogram at all we make it the same width as the one next to it, i.e. 10 wide. So the axis will start at 290.

Similarly the last one, defined '380+' will end on 420 to make it the same width as the adjacent one.

2. This table shows the weekly amount spent on shopping by a group of families who took part in a family expenditure survey:

Amount spent (£)	20-30	30-40	40-60	60-80	80-100	100-150
Number of Families	10	15	30	50	40	40

3. This table shows the monthly expenditure on food of a sample of 220 households:

Monthly Expenditure (£)	Less than 100	100-150	150-200	200-300	300-450	450+
Frequency	15	25	30	80	55	15