Delving into Data

Year 10

#MathsEveryoneCan

White Rose Maths

	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12
Autumn	**Similarity**						**Developing Algebra**					
	Congruence, similarity and enlargement			Trigonometry			Representing solutions of equations and inequalities			Simultaneous equations		
Spring	**Geometry**						**Proportions and Proportional Change**					
	Angles & bearings		Working with circles		Vectors		Ratios & fractions		Percentages and Interest		Probability	
Summer	**Delving into data**						**Using number**					
	Collecting, representing and interpreting data						Non-calculator methods		Types of number and sequences		Indices and Roots	

Summer 1: Delving into Data

This block builds on KS3 work on the collection, representation and use of summary statistics to describe data. Much of the content is familiar, both from previous study within and beyond mathematics (including Geography and Science) and from everyday life. The steps have been chosen to balance consolidation of existing knowledge with extending and deepening, particularly in terms of interpretation of results and evaluating and criticising statistical methods and diagrams. For students following Higher tier, there is additional content relating to continuous data including histograms, cumulative frequency diagrams, box plots and associated measures such as quartiles and the interquartile range. Again the emphasis with these topics should be on interpretation (particularly in making comparisons) and not just construction.
A possible approach to teaching this unit would be project-based, where students collect primary data (or select samples from secondary data) from which they make and test hypotheses, thus giving a purpose to the creation and analysis of the diagrams and measures involved. Looking at data from other subject areas might again be useful here.

National curriculum content covered:

- consolidating subject content from key stage 3:
  - use describe, interpret and compare observed distributions of a single variable through: appropriate graphical representation involving discrete, continuous and grouped data
  - construct and interpret appropriate tables, charts, and diagrams, including frequency tables, bar charts, pie charts, and pictograms for categorical data, and vertical line (or bar) charts for ungrouped and grouped numerical data
  - describe, interpret and compare observed distributions of a single variable through: appropriate graphical representation involving discrete, continuous and grouped data; and appropriate measures of central tendency (mean, mode, median) and spread (range, consideration of outliers)
- infer properties of populations or distributions from a sample, whilst knowing the limitations of sampling
- interpret and construct tables and line graphs for time series data
- **{construct and interpret diagrams for grouped discrete data and continuous data, i.e. histograms with equal and unequal class intervals and cumulative frequency graphs, and know their appropriate use}**
- interpret, analyse and compare the distributions of data sets from univariate empirical distributions through appropriate graphical representation involving discrete, continuous and grouped data, **{including box plots}**
- apply statistics to describe a population
- interpret, analyse and compare the distributions of data sets from univariate empirical distributions through appropriate measures of central tendency (including modal class) and spread **{including quartiles and inter-quartile range}**

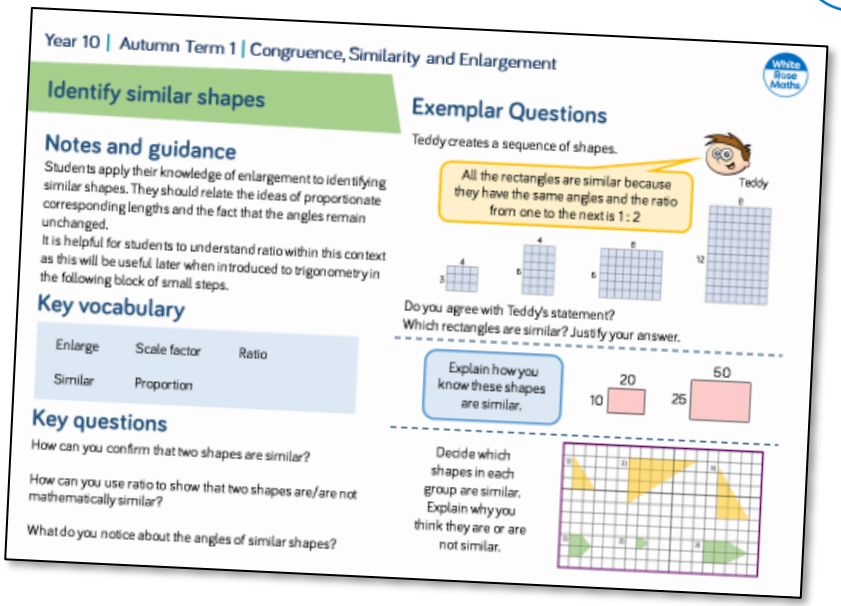
Why Small Steps?

We know that breaking the curriculum down into small manageable steps should help students to understand concepts better. Too often, we have noticed that teachers will try and cover too many concepts at once and this can lead to cognitive overload. We believe it is better to follow a "small steps" approach.

As a result, for each block of content in the scheme of learning we will provide a "small step" breakdown. ***It is not the intention that each small step should last a lesson – some will be a short step within a lesson, some will take longer than a lesson.*** We would encourage teachers to spend the appropriate amount of time on each step for their group, and to teach some of the steps alongside each other if necessary.

What We Provide

- Some ***brief guidance*** notes to help identify key teaching and learning points
- A list of ***key vocabulary*** that we would expect teachers to draw to students' attention when teaching the small step,
- A series of ***key questions*** to incorporate in lessons to aid mathematical thinking.
- A set of questions to help ***exemplify*** the small step concept that needs to be focussed on.



Year 10 | Autumn Term 1 | Congruence, Similarity and Enlargement

Identify similar shapes

Notes and guidance

Students apply their knowledge of enlargement to identifying similar shapes. They should relate the ideas of proportionate corresponding lengths and the fact that the angles remain unchanged.

It is helpful for students to understand ratio within this context as this will be useful later when introduced to trigonometry in the following block of small steps.

Key vocabulary

Enlarge, Scale factor, Ratio, Similar, Proportion

Key questions

How can you confirm that two shapes are similar?

How can you use ratio to show that two shapes are/are not mathematically similar?

What do you notice about the angles of similar shapes?

Exemplar Questions

Teddy creates a sequence of shapes.

All the rectangles are similar because they have the same angles and the ratio from one to the next is 1 : 2

Teddy

Do you agree with Teddy's statement?
Which rectangles are similar? Justify your answer.

Explain how you know these shapes are similar.

10, 20, 25, 50

Decide which shapes in each group are similar. Explain why you think they are or are not similar.

- These include reasoning and problem-solving questions that are fully integrated into the scheme of learning. Depending on the attainment of your students, you many wish to use some or all of these exemplars, which are in approximate order of difficulty. Particularly challenging questions are indicated with the symbol .
- For each block, we also provide ideas for key representations that will be useful for all students.

Some of the small steps are in **bold** and labelled with H to indicate this is Higher tier GCSE content. We would encourage teachers to use these with as many students as possible – if you feel your class can access any particular small step, then please include it in your planning. Steps that review content covered at Key Stage 3 are labelled R.

Delving into data

Small Steps

- Understand populations and samples
- **Construct a stratified sample** H
- Primary and secondary data
- Construct and interpret frequency tables and frequency polygons
- Construct and interpret two-way tables R
- Construct and interpret line and bar charts (including composite bar charts)
- Construct and interpret pie charts R
- Criticise charts and graphs

H denotes Higher Tier GCSE content

R denotes 'review step' – content should have been covered at KS3

Delving into data

Small Steps

Construct histograms	H
Interpret histograms	H
Find and interpret averages from a list	R
Find and interpret averages from a table	R
Construct and interpret time series graphs	R
Construct and interpret stem-and-leaf diagrams	
Construct and interpret cumulative frequency diagrams	H
Use cumulative frequency diagrams to find measures	H

H denotes Higher Tier GCSE content

R denotes 'review step' – content should have been covered at KS3

Delving into data

Small Steps

- **Construct and interpret box plots** H
- Compare distributions using charts and measures
- **Compare distributions using complex charts and measures** H
- Construct and interpret scatter graphs R
- Draw and use a line of best fit R
- Understand extrapolation

H denotes Higher Tier GCSE content

R denotes 'review step' – content should have been covered at KS3

Population and samples

Notes and guidance

Students need to be aware that the 'population' is the whole group being studied rather than, say, the population of a city or country. They should also discuss the merits of random sampling. There is often confusion caused by the colloquial use of the word 'random' to mean haphazard or unexpected, rather than the statistical meaning that each member of the population has an equal chance of being selected.

Key vocabulary

Population	Sample	Representative
Biased	Random	

Key questions

Why do statisticians take samples rather than interview the whole population?

How could the random number generator on your calculator be used to support selecting a random sample?

Exemplar Questions

Jack wants to pick a sample of 5 students from his class to complete a survey. Which of these methods will produce a random sample?

Picking the first five names he thinks of	Picking his five closest friends in the class	Putting all the names in a hat and picking out five

Rosie wants to find out if people in her town are happy with the library service. She asks all of the people who live on her street their opinions.

- Explain why Rosie's survey is not reliable
- Suggest a way to improve Rosie's survey

There are about 80 000 trees in a forest.
A researcher wants to test 1% of the trees for a disease.
Compare these methods of choosing the sample.

Number all the trees and pick 800 numbers at random	Picking the first 800 at the entrance to the forest	Split the forest into 20 regions and pick 40 trees from each region

Alex surveys 5% of the students at her school about a proposed change to the school uniform. She talks to 25 students in Year 10 and 15 students in Year 11

- What is meant by the population in this case?
- What is the size of the population?
- Is Alex's sample likely to be representative? Why or why not?

Stratified samples

H

Notes and guidance

Rather than trying to learn a formula, it is helpful if students approach stratified sampling using proportional reasoning, finding the fraction each group/stratum is of the whole population and assigning the same fraction of the total sample size to each group. It would also be useful to discuss whether stratified samples are necessarily representative e.g. would it be useful to split each group into male/female? etc.

Key vocabulary

Population	Sample	Representative
Biased	Proportion	Stratified

Key questions

What fraction of the whole population is the sample size?
How can you work out this fraction of each group/stratum?

How can we ensure the stratified sample is random?

Exemplar Questions

The table shows the number of students in each year group in a school.
Brett is conducting a survey and decides to interview 10 students from each year group.

Criticise Brett's sampling method and suggest an improvement.

Year	No. of students
7	150
8	180
9	170
10	110
11	90

There are three age groups in a running club.
The table shows the number of people in each group.

Age group	12 - 14	15 - 17	18+
No of members	156	336	108

Dexter wants to survey a sample of the members.
He wants to survey 50 people altogether.
Work out the number of people he should ask from each age group.

House	Merlin	Potter	Gandalf	Glinda
No of students	92	165	83	217

The table shows the number of students in each of the four houses of a school.
In a stratified sample, 7 students are selected from Gandalf house.
How many students should be selected from each of the other houses?

Primary and secondary data

Notes and guidance

This step is probably best covered when discussing sampling, which is the most common way of gathering primary data. It is useful to discuss the pros and cons of each type of data e.g. secondary is much cheaper, but may not be as reliable. The internet is a great source of secondary data which could be useful to exploit to generate the charts and diagrams in the forthcoming steps.

Key vocabulary

Primary	Secondary	Source
Data	Questionnaire	Experiment

Key questions

Why might some secondary sources of data be biased?

Give me an example of a biased question in a questionnaire. How might you improve it?

What columns do you need on a data collection sheet?

Exemplar Questions

Would it be more sensible to use primary or secondary data to investigate these hypotheses/situations? Explain your answers.

The population is generally taller now than they were in the 1970s	Which is the wettest month of the year?
Do taller athletes perform better at the long jump?	A company wants to research the market for a new flavour of crisps
The more you revise, the better your results.	The speed at which you type text messages declines with age.

If you chose primary data, how might you go about collecting this?

Boys have faster reaction times than girls.

Jack wants to collect primary data to test his hypothesis.
Suggest how he may do this.

Eva is collecting data about television and reading habits.
Suggest how her questions could be improved.

- Do you watch a lot of television?
- How much do you spend on books?

£1 - £10 ☐ £11 - £20 ☐ £20 - £30 ☐

Frequency tables and polygons

Notes and guidance

Students are familiar with frequency tables for grouped data from KS3, and may recall the idea of the midpoint as used to find the estimate of the mean. Teachers could include grouped frequency diagrams here, linking to the later Higher tier topic of histograms, and explore the similarities and differences between these and frequency polygons.

Key vocabulary

Frequency polygon	Midpoint	Endpoint
Frequency	Class	Interval

Key questions

What's the difference between a midpoint and an endpoint?

Given the endpoints of a class interval, how do you work out the midpoint?

How do you choose the scales for your frequency polygon?

Exemplar Questions

The frequency polygon shows the ages of the employees at a firm.

- How many of the employees are aged less than 45?
- Why does the horizontal axis start at 30?
- Can you use the frequency polygon to work out the age of the oldest employee at the firm?

The table shows the mass in grams of 70 oranges.

Weight (w)	Frequency
$40 < w \leq 50$	7
$50 < w \leq 60$	23
$60 < w \leq 70$	26
$70 < w \leq 80$	12
$80 < w \leq 90$	2

- Draw a frequency polygon to show this information.
- What is the probability that a randomly selected orange weighs over 60 grams?

On the same axis, draw another frequency polygon to show the mass of these 70 apples.

Weight (w)	$50 < w \leq 60$	$60 < w \leq 70$	$70 < w \leq 80$	$80 < w \leq 90$
Frequency	12	26	20	12

What do you notice?

Two-way tables

R

Notes and guidance

Students have worked with two-way tables throughout KS3, so this review step is an opportunity to revisit both extracting and completing information as well as designing tables, looking at more complex tables if appropriate. There are ample opportunities to link to other areas of the curriculum that need revising, including fractions, decimals, percentages, ratios and probability.

Key vocabulary

Table	Row	Column
Total	Difference	

Key questions

How do you decide which categories to use for the rows and columns of the two-way table? Would it make a difference if the rows and columns were swapped?

What ratios can you find out from the two-way table?
What probabilities can you find?

Exemplar Questions

The table shows the ages of students in a class.

- Complete the table.
- What percentage of the whole class is girls?
- A student is picked at random from the class.
  What is the probability the student is a 15 year old boy?

	Age 14	Age 15	Total
Girls	7	12	
Boys		8	
Total			32

The table shows the number of people who correctly identified crisp flavours in a blind taste test.

	Correct	Incorrect	Total
Cheese	25		40
Salt & Vinegar		12	30
Beef			27

Two-thirds of the people correctly identified Beef.
Complete the table.
What proportion correctly identified each of the other two flavours?

---

40 children take part in a school show.
The ratio of boys to girls is 3 : 5
$\frac{3}{5}$ of the children are dancers and the rest are singers.
There are 3 boys who are dancers in the show.

- Represent this information is a two-way table.
- How many more girls are dancers than singers?

Line and bar charts

Notes and guidance

Students will be very familiar with constructing and interpreting bar charts, and should experience them in a variety of forms – vertical, horizontal, lines instead of bars etc. They should also explore multiple and composite bar charts, as in the two exemplar questions, focusing on interpretation and what types of information it is easier to read from one type than the other.

Key vocabulary

Line/Bar chart	Frequency	Difference
Dual/Multiple	Composite	Total

Key questions

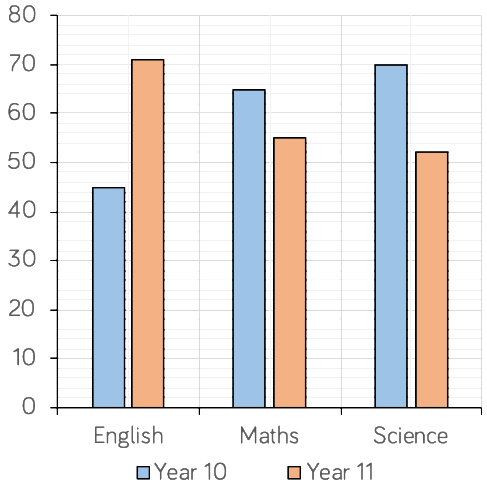
What's the difference between a vertical and a horizontal bar chart?

How thick does a bar have to be? Can you use a line?

What's the difference between a multiple bar chart and a composite bar chart? Which is easier to read?

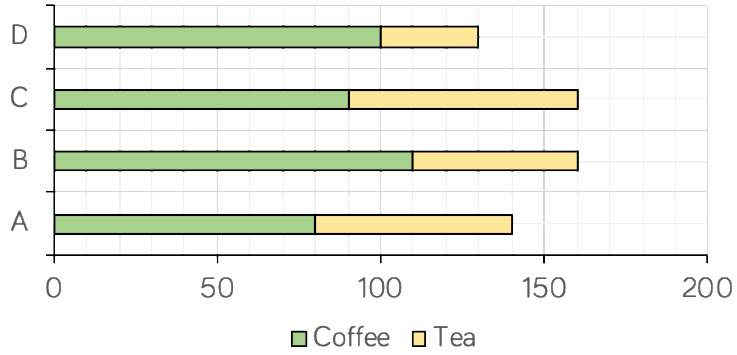
Exemplar Questions

The dual bar chart shows the numbers of students in Y10 and Y11 who chose English, Maths or Science as their favourite subject.



- How many more Year 10s chose Science than Year 11s?
- What other questions can you ask/answer?

The composite bar chart shows the sales of coffee and tea at four branches of a coffee shop between 8 am and 9 am one morning.



How much more coffee than tea was sold in total?
What other questions can you ask/answer?

Pie charts

R

Notes and guidance

Students need to be able to construct pie charts and, equally importantly, need to be able to interpret them. It is useful to look at the proportions in the chart as fractions of 360, as well as percentages and as fractions of the 'whole' that is being represented. It is also worth discussing the pros and cons of using a pie chart rather than a bar chart e.g. proportions of the whole are easier, but comparing parts is not always so.

Key vocabulary

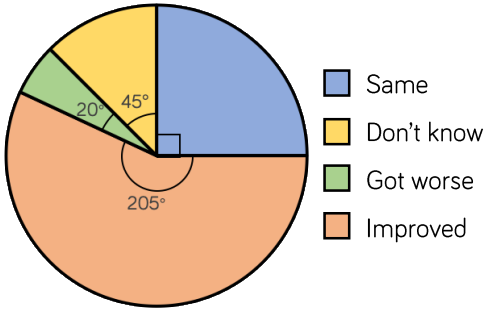
Angle	Sector	Radius
Subtend		

Key questions

If you know the angle of a sector on a pie chart, how can we work out what fraction of the whole this represents?
If you know the proportion of the whole, how can we work out the angle we need for the pie chart?
Does it matter what order we present the data in a pie chart?

Exemplar Questions

The pie chart shows the results of a survey of people by a bus company to find out whether they thought the service had improved.



- What percentage of the people surveyed thought the service had improved?
- 100 people said the service had got worse. What other information can you work out?

Activity	Frequency
Dance	16
Drama	12
Football	21
Music	15
Chess	8

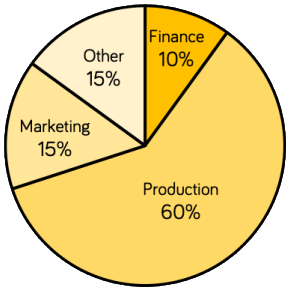
The table shows the after-school activities chosen by a group of students.

Draw a pie chart to represent the information in the table.

What fraction of the students chose music?

The pie chart shows the percentage of employees of a firm who work for each department.

- What angle represents Production?
- 24 people work in the Marketing department.
  How many work in Production or Other?



Criticise charts and graphs

Notes and guidance

Students need to look beyond the superficial criticisms of neatness, labelling of axes and titles to consider the mathematical flaws that charts or graphs may have. In particular, changes in scale, starting the axes from inappropriate points or misuse of scaling may exaggerate or minimise differences. Encourage students to find real-life examples of this – there are plenty available!

Key vocabulary

Scale	Bias	Misleading
Broken axis		

Key questions

What do you notice about the scales of the axis/axes?

What is the difference between the values in ___ and ___? Does the graph show this?

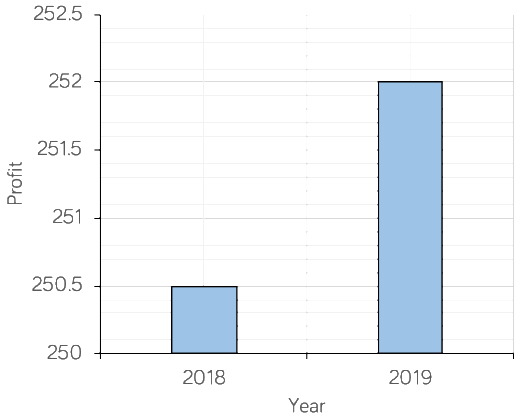
Why might someone want to use a graph to make differences look bigger/smaller?

Exemplar Questions

The pictogram shows the number of heart operations carried out at a hospital over one week.

♥ represents 5 operations

- On which day were most operations performed?
- Comment on the design of the pictogram.



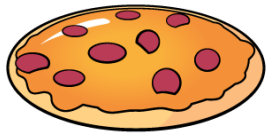
A company shows this bar chart to show the change in profit between 2018 and 2019

Why might this be misleading?

A pizza firm shows the change in average number of pizzas sold per week using the diagram below.
Explain why the diagram might be misleading.

2018
4000

2019
8000



Construct histograms

H

Notes and guidance

This step explores why grouped frequency diagrams are not appropriate for unequal class intervals as using height to represent frequency can be misleading – the first exemplar question addresses this. Once the idea that frequency is proportional to the area of the bar is established, the formula frequency density = frequency ÷ class width is easily established.

Key vocabulary

Histogram	Area	Frequency density
Class interval	Class width	

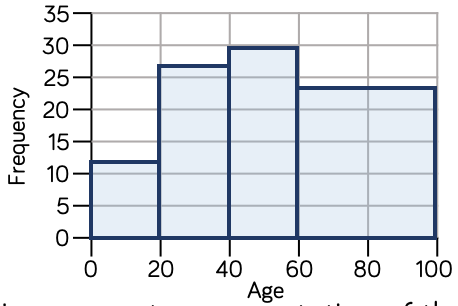
Key questions

Why do we use area to represent frequency rather than height?
What are the main differences between frequency polygons and histograms?
Do you draw histograms using midpoints or end points?

Exemplar Questions

The table and frequency diagram show the ages of people living in a small village.

Age	Number of people
Under 20	12
$20 < w \leq 40$	27
$40 < w \leq 60$	30
$60 < w \leq 100$	24



- Why is the grouped frequency diagram not representative of the data in the table?
- Complete the table and draw a histogram to show the data.

Age	Frequency	Class width	Frequency density
Under 20	12	20	12 ÷ 20 = 0.8
$20 < w \leq 40$	27		
$40 < w \leq 60$	30		
$60 < w \leq 100$	24		

- Explain why the histogram is more representative.

The times, in minutes, taken for a group of people to run a 10 kilometre race are shown in the table.

Time	$40 < w \leq 50$	$50 < w \leq 60$	$60 < w \leq 70$	$70 < w \leq 90$	$90 < w \leq 120$
Frequency	6	24	38	32	24

- Draw a histogram to represent this information.
- Find an estimate for the mean time taken.

Interpret histograms

H

Notes and guidance

As well as knowing how to construct a histogram from scratch, students should also be able to deduce frequencies from a given histogram. Exam questions often give partially completed tables and histograms for students to complete and this is explored in the second exemplar question. Students should also be encouraged to make comments about the shape of distributions.

Key vocabulary

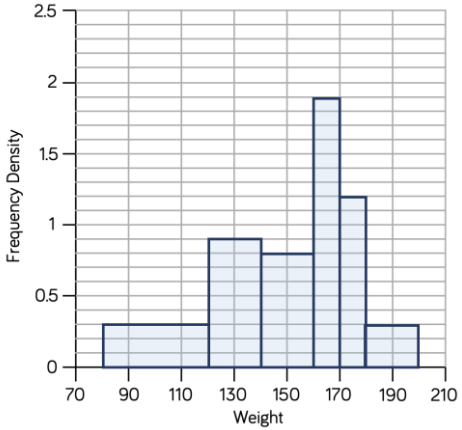
Histogram	Area	Frequency density
Class interval	Class width	Distribution

Key questions

What does the area of each bar represent?
How are frequency and frequency density connected? If we know the class width and the frequency density, how can we work out the frequencies?
What information do we know? What can we find out?

Exemplar Questions

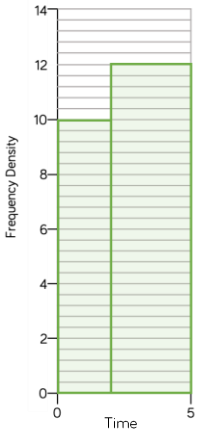
The histogram shows the weights of 87 pieces of fruit.



- Can you tell the modal class of the weights just by looking at the histogram?
- Use the histogram to complete a frequency table, and then use your table to estimate the mean weight of the pieces of fruit.

The table shows information on the amount of time, in minutes, people spend on a phone app per day.

Time (t)	Frequency
$0 < t \leq 2$	20
$2 < t \leq 5$	
$5 < t \leq 10$	20
$10 < t \leq 20$	20
$20 < t \leq 30$	8



- Use the information on the table to complete the histogram.
- Use the histogram to complete the table.

Averages from a list

R

Notes and guidance

Students will have met mean, median and mode several times and at this stage they need to be considering when each one is and isn't appropriate e.g. only the mode is possible with categorical data. Consideration of the effect of removing outliers is a good way to show distortion of the mean and why the median is often preferable.

Key vocabulary

Mean	Median	Mode
Representative	Outlier	Average

Key questions

Do you think any of these values are outliers? Which ones? Why?

What would you do if I asked you to find the average of a set of numbers? Why would you chose that one?

How much does the ... give a sense of the whole set of data?

Exemplar Questions

Compare the means of the sets of numbers.

The first five positive integers	The first five square numbers
The first five multiples of 3	The smallest five factors of 40

The list shows the number of letters received in a group of ten houses on a street one Tuesday.

0, 0, 0, 3, 3, 4, 4, 5, 6, 22

- Find the mean, median and mode of the data.
- Which average is most representative of the data as a whole?
- Why do the other averages not represent the data well.
- Are any of the values outliers? Why or why not?

Here are 10 students' estimates of the size of an angle.

35° 40° 41° 30° 35°
32° 35° 125° 31° 44°

Find the mean, median and mode of the data.
Given that the angle is an acute angle, explain how this would change your calculations and your answers.

Explain whether it is possible to find the mean, median and mode of these expressions.

$3x$ $3x+5$ $4x$ $3x-1$ $3x-7$

Averages from a table

R

Notes and guidance

It is useful for students to look at tables presented both horizontally and vertically when revising this KS3 topic, and then decide which is the best way to set the tables out to find averages. The term 'modal class' will need revisiting, emphasising its relationship to the mode. It is useful to consider where the median lies as an introduction to the Higher tier work on cumulative frequency graphs.

Key vocabulary

Mean	Median	Modal Class
Subtotal	Midpoint	Estimate

Key questions

What does 'modal' mean? Which average does it relate to?
Looking at the table, what is the maximum/minimum value the mean could be? Is the value you've found reasonable?
How do you work out the midpoint of a class interval?

Exemplar Questions

The table shows the number of siblings a group of students have.

Number of siblings	0	1	2	3	4	5
Frequency	7	8	6	4	2	2
Subtotal	0	8	12	12	8	10

Mo and Dora are trying to find the mean number of siblings.
Explain why both their calculations are wrong.

Mean $= \frac{50}{6} = 8.33$

Mean $= \frac{50}{22} = 2.27$

Eva is working out the modal number of siblings.

The mode is 2 because it is the frequency that occurs most.

Do you agree with Eva?

The table shows the weight of some objects.

Weight (w)	Frequency	???	???
$40 < w \leq 50$	6		
$50 < w \leq 60$	15		
$60 < w \leq 70$	14		
$70 < w \leq 80$	2		
Total			

- What is the modal class interval?
- What are the extra columns you need to work out an estimate of the mean of the weights shown in the table?
- Complete the table and find an estimate for the mean.
- Can you work out which class the median will be in?

Time series graphs

R

Notes and guidance

When reviewing these types of line graph, it is worth discussing the meaning (if any) of values between the plotted points e.g. taking a reading between May and June is meaningless and any reading between (say) 9 and 10 am is at best an estimate that assumes linear change between that period. It is also worth discussing seasonal trends and cases where there is no apparent trend.

Key vocabulary

Time Series	Quarter
Trend	

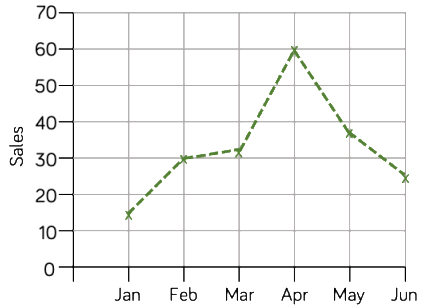
Key questions

Which of the graphs show an upward or downward general trend (or no trend)? Are there any points that don't fit the general pattern?

In this particular graph, does it make sense to read off information from points between the given points? How reliable will these readings be?

Exemplar Questions

The graph shows the sales, in £1000s, of a company in the first six months of a year. The company's target was to have mean sales of at least £35 000



- Did the company meet their target?
- Explain whether it would be sensible to estimate July's sales from the graph.

Draw a graph to represent the greenhouse gas emissions in the UK between 1990 and 2015 shown in the table.

Year	1990	1995	2000	2005	2010	2015
Emissions	794.2	744.3	705.3	681.3	597.1	492.4

Describe the trend.

The table shows the average maximum temperatures, in °C, in Halifax, UK, and Halifax, Canada, for each month of the year.

Month	J	F	M	A	M	J	J	A	S	O	N	D
Halifax, UK	5	5	8	10	15	18	20	19	16	12	8	6
Halifax, Canada	-1	-1	2	8	13	17	22	22	18	13	7	2

On the same axes, draw a pair of graphs to represent the data. Comment on the similarities and differences between your graphs.

Stem and leaf diagrams

Notes and guidance

As with most of the diagrams in the block, interpretation is just as important as construction. When drawing stem and leaf diagrams, students need to take care to keep numbers in line so that the relative lengths of each line are meaningful.
Compare stem and leaf diagrams to horizontal bar charts where all the data is visible, and revisit averages and the range.
Include examples with decimal values e.g. 7|3 means 7.3

Key vocabulary

Stem	Leaf	Median
Range	Modal class	

Key questions

In what way is a stem and leaf diagram similar to a bar chart?
Why do we need a key for a stem and leaf diagram?
How do we work out where the median is in a stem and leaf diagram?

Exemplar Questions

The stem and leaf diagram shows the scores achieved in a test by 21 students.

2	3	7
3	0	2 3 5 5 7
4	1	5 5 5 6 8 9
5	0	6 7 7
6	4	8

Key

4 | 1 means 41

Work out the range of the marks.
Find the median and the mode of the marks.
Put the data into groups 20 to 29, 30 to 39 etc. and find an estimate of the mean mark. Compare your answer to the actual mean.

The back-to-back stem and leaf diagram shows the IQ scores for a group of pensioners and a group of teenagers.

Pensioners				Teenagers				
	9	9	3	9	3			
7	4	4	1	10	2	4	5	8
	8	7	6	11	1	3	6	7
		4	0	12	0	2	2	4 8
			2	13	1	5	5	

Key

1 | 10 | 2
means
101 for pensioners,
102 for teenagers

What was the lowest IQ score in each group?
Find the median score for each group, and state each group's modal class. What else can you find?

Cumulative frequency

H

Notes and guidance

Students can get cumulative frequency polygons mixed up with frequency polygons and it is worth investing time discussing the differences. In particular, when drawing tables for cumulative frequency polygons, it is useful to include an "upper limit" column as well as the cumulative frequency column. Although curved lines are often used to represent cumulative frequencies, these are not an expectation.

Key vocabulary

Cumulative	Frequency	Graph
Polygon	End point	Class

Key questions

What's the difference between a frequency polygon and a cumulative frequency polygon?

Why are cumulative frequency polygons plotted at the upper end points of intervals rather than the midpoints or lower class boundaries?

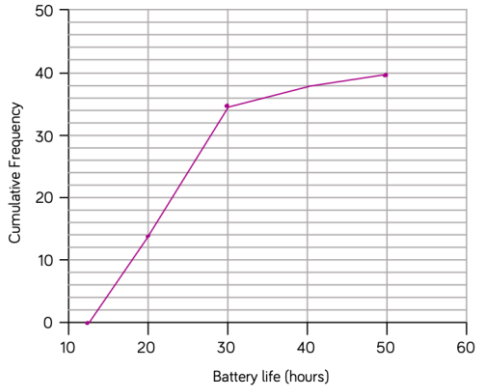
Exemplar Questions

The table shows the heights of 30 plants.

Height (h)	Frequency
$5 \leq h < 10$	4
$10 \leq h < 15$	7
$15 \leq h < 20$	9
$20 \leq h < 30$	6
$30 \leq h < 40$	4

- State the modal class of the heights.
- How many of the plants are less than 10 cm tall?
- How many of the plants are less than 15 cm tall?
- How many of the plants are less than 30 cm tall?
- In which class is the plant with the median height?
- Draw a cumulative frequency graph to represent this information.

The cumulative frequency graph gives information about the battery lives of some mobile phones.



- How many phones had a battery life of less than 20 hours?
- How many phones had a battery life of more than 30 hours?
- Complete the statement: "10 phones had a life of more than..."

Use cumulative frequency

H

Notes and guidance

Building on the previous step, students look at using graphs to find the median and related measures such as the interquartile range. It is good to continue to ask questions such as, "90% of the data are greater/less than…" to ensure students are thinking about the meanings of the values and not just procedures. Drawing lines between the graph and the axes is to be advised rather than estimating by eye!

Key vocabulary

Median	Upper/Lower Quartile	
Interquartile range	Range	Outlier

Key questions

Why can't you tell the range from a grouped frequency table or a cumulative frequency graph?

Why is the value for the median only an estimate?

Given a cumulative frequency graph, how can we deduce a grouped frequency table?

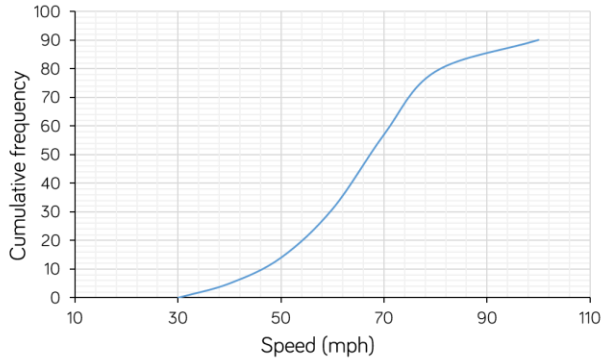
Exemplar Questions

The times taken, in minutes, for a group of people to run a 10 kilometre race are shown in the table.

Time	$40 < w \leq 50$	$50 < w \leq 60$	$60 < w \leq 70$	$70 < w \leq 90$	$90 < w \leq 120$
Frequency	6	24	38	32	24

- Draw a cumulative frequency graph to illustrate the data.
- Use your graph to estimate the median, lower quartile, upper quartile and interquartile range of the times.

The cumulative frequency graph gives some information about the speeds of 100 cars on a motorway.



Where possible, complete the cards. If it is not possible, explain why.

Median speed = _____	80% of the cars are going faster than ___
IQR of the speeds = _____	Range of the speeds = _____

Use the graph to help find an estimate of the mean speed of the cars.

Box plots

H

Notes and guidance

It is useful to tie in the teaching of box plots with that of cumulative frequency polygons, and many exam questions do this. Students need to take particular care with the maximum and minimum values, ensuring the correct end of the intervals has been used. Although many box plots are produced horizontally, some programs produce them vertically, so students should experience both.

Key vocabulary

Median	Upper/Lower Quartile	
Interquartile range	Range	Outlier

Key questions

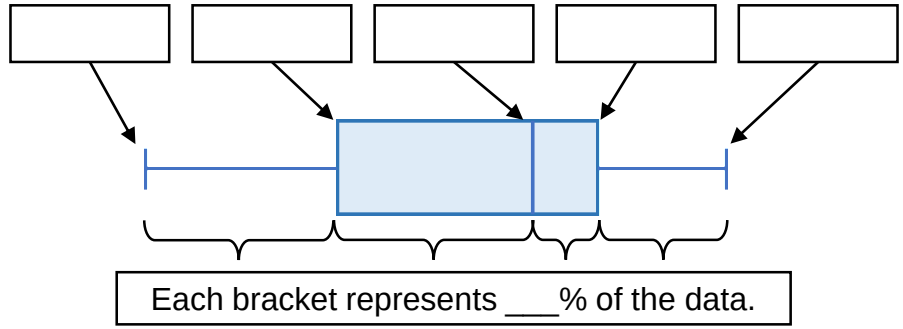
How many parts does a box plot split a distribution of data into?

What information can we read from a box plot?

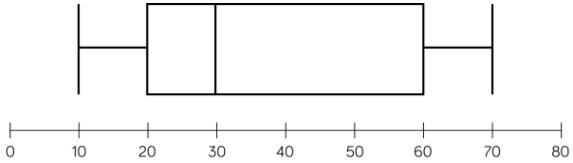
How might we be able to identify an outlier on a box plot?

Exemplar Questions

Complete the cards to describe the features of a box plot.

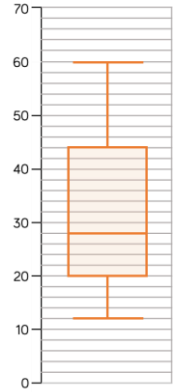


The box plot shows the times in seconds taken by a group of people to do a crossword puzzle.



What can you find out from the box plot?

The box plot shows the distributions of marks scored by students in a test.



- What was the median mark?
- What was the range of the marks?
- Describe the distribution of marks.

Comparing distributions

Notes and guidance

When comparing distributions, students should look at one of the averages and measure of spread; at Foundation level this will always be the range. The average is used as an indicator of overall performance and the range is used to describe the consistency. Students often only look at the average, so looking at data sets where the averages are equal but the ranges differ can be useful.

Key vocabulary

Median	Mean	Spread
Consistent	Range	

Key questions

What does it mean if a data set has a large/small range?
If the averages of two data sets are similar but the ranges are very different, what does that tell you?
How does the range help us decide which ___ is more reliable?

Exemplar Questions

Five students were asked to estimate when one minute had passed. They repeated the experiment five times each, recording in seconds how much time had actually passed.

Jack	55	71	58	64	49
Rosie	47	49	57	42	47
Eva	68	82	77	90	78
Dora	49	47	51	48	47
Teddy	58	65	62	57	64

- Who was the most consistent student?
- Who do you think was the best estimator?
- Justify your answers.

The cards show some information about the times taken for two different bus companies on the same route.

Rider Bus
Mean = 35 minutes
Range = 10 minutes

Speedy Bus
Mean = 32 minutes
Range = 16 minutes

Use the information to compare the performances of the two buses.

Dexter and Dora keep a record of their scores in 8 spelling tests.

Dexter's scores are: 17 18 11 16 15 10 19 12

Dora's mean score is 14 and the range of her scores is 6

Compare Dexter's and Dora's scores in the tests.

Complex comparisons

H

Notes and guidance

Building on the last step, Higher tier students are expected to use box plots to make comparisons between distributions. This means there is a choice of indicators for the spread: the range or the interquartile range, but the average to consider will be the median. It is worth discussing that the range is liable to be affected by outliers. Again, students need to be encouraged to make a comparison of both aspects.

Key vocabulary

Median	Mean	Spread
Consistent	Range	Interquartile range

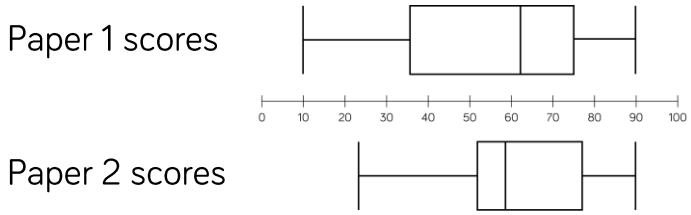
Key questions

Why is just looking at the medians not enough to give us a full picture when comparing two data sets?

What's the difference between the range and the interquartile range? Which of these measures might be affected by an outlier?

Exemplar Questions

The box plots show some information about the marks obtained by a class of students in Paper 1 and Paper 2 of an exam.



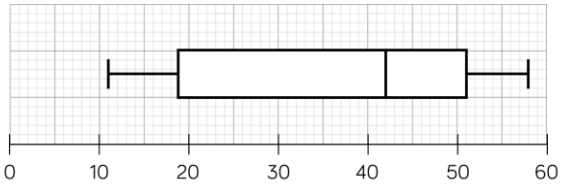
- Estimate the median score of each paper.
- Estimate the interquartile range of scores for each paper
- Use your estimates to compare the students' performances in the two papers.

The table shows some information about the lifetime, in hours, of mobile phone batteries made by company A.

Minimum	Lower Quartile	Median	Upper Quartile	Maximum
30	35	42	46	50

Draw a box plot for this information.

The box plot below shows the distribution of the lifetime, in hours, of mobile phone batteries made by company B.



Compare the distributions of the lifetimes of the batteries of the two companies.

Scatter graphs

R

Notes and guidance

Students will be familiar with correlation from KS3, but this review step is useful to remind them of the vocabulary and to practice choice of scale when plotting points. Where to start and finish axes are also good points for discussion. It is also worth reinforcing that correlation does not imply causality, and that absence of linear relationship does not necessarily mean that the variables are unconnected.

Key vocabulary

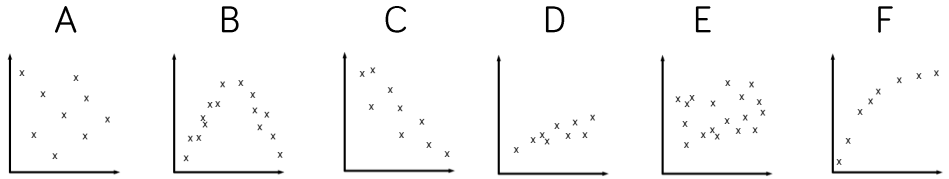
Variable	Relationship	Linear
Positive/negative correlation		Scale

Key questions

How can you tell if a correlation is positive or negative?
Describe the relationship without using the word correlation (e.g. "the taller you are...").
What are the lowest and highest values for each axis?
How do these values help us choose the scales?

Exemplar Questions

Here are six scatter graphs.



Describe the correlation in each of the graphs.

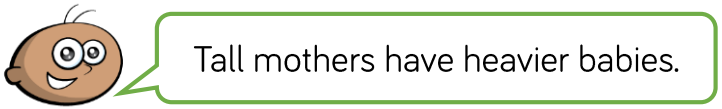
In which graphs do you think there might be a relationship between the variables?

Do you need to have strong correlation between the variables for there to be a relationship?

The table shows the heights in centimetres of 10 new mothers and the weights in kilograms of their babies.

Height	161	172	154	162	158	171	172	160	155	157
Baby's weight	3.9	3.8	4.5	3.1	4.1	3.8	3.6	3.2	2.8	4.5

Plot the information on a scatter graph, thinking carefully about the range and scales of your axes.



Do you agree with Tommy?

Lines of best fit

R

Notes and guidance

When using lines of best fits to make estimates, students should draw lines from/to the axes to make their intention clear and to improve accuracy. They could also link to the last steps, considering whether a line of best fit is appropriate or not, and to the next step considering the range of values over which making estimates is sensible. Be aware that in other subjects students may draw 'curves of best fit'.

Key vocabulary

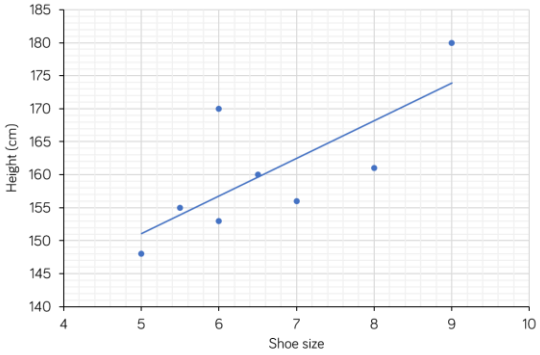
Line of best fit	Origin	Estimate
Correlation	Interpolate	Interpolation

Key questions

Does the line of best fit always have to go through the origin?
Is it possible to have a line of best fit with a negative gradient?
Does a line of best fit need to be straight line?

Exemplar Questions

The graph shows the shoe sizes and heights of a group of people.



Use the line of best fit to estimate the height of a person with:

- shoe size 7
- shoe size 7.5

Use the line of best fit to estimate the shoe size of someone 170 cm tall.
How reliable do you think your answers are? Why?

The table shows the scores of some students in English, Maths and Science tests.

	A	B	C	D	E	F	G	H
Maths	47	51	56	60	63	63	70	76
English	67	55	71	43	57	70	51	56
Science	51	53	67	69	66	70	69	79

On separate axes, draw the graphs of:

- Maths and English scores
- Maths and Science scores
- English and Science scores

Describe the correlation in each case.

On which graph(s) would it be sensible to draw a line of best fit?

Extrapolation

Notes and guidance

This step can be taught alongside lines of best fit, considering when it is and isn't appropriate to extrapolate outside the given data range. This can be demonstrated by looking at examples that give e.g. negative or other impractical answers. Links could be made to Science e.g. considering when relationships may work for certain intervals but not others e.g. length of an extended spring.

Key vocabulary

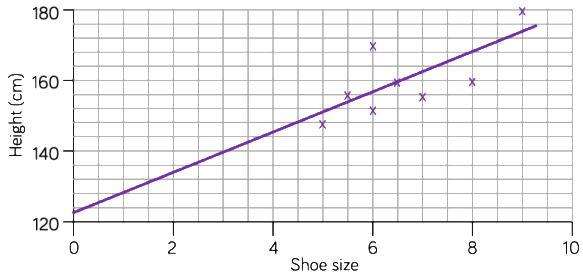
Line of best fit	Origin	Estimate
Correlation	Interpolate	Extrapolate

Key questions

When might it be appropriate to extrapolate beyond the range of values given in a data set? How reliable to do you think the estimates might be?
What difference do outliers make to where we draw the line of best fit?

Exemplar Questions

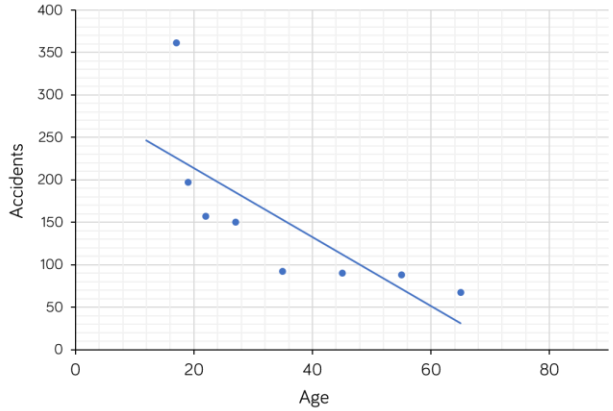
The graph shows the shoe sizes and heights of a group of people.



- Dora uses the line of best fit to estimate the height of someone with shoe size 1. Do you think her answer will be reliable?
- Can the graph be used to estimate the height of someone 1 m tall?

Justify your answers.

The graph shows the number of car accidents per million miles for drivers of different ages.



- Which point is an outlier?
- Why can you not use the graph to estimate the number of accidents for 70 or 80 year olds?