MathsHUBS Resources

National Centre for Excellence in the Teaching of Mathematics

Mastery Professional Development

5 Statistics and probability

5.1 Statistical representations and measures

Guidance document | Key Stage 3

Making connections

The NCETM has identified a set of six 'mathematical themes' within Key Stage 3 mathematics that bring together a group of 'core concepts'.

The fifth of these themes is *Statistics and probability*, which covers the following interconnected core concepts:

5.1 **Statistical representations and measures**

5.2 Statistical analysis

5.3 Probability

This guidance document breaks down core concept *5.1 Statistical representations and measures* into two statements of knowledge, skills and understanding:

5.1.1 Understand and calculate accurately measures of central tendency and spread

5.1.2 Construct accurately statistical representations

Then, for each of these statements of knowledge, skills and understanding we offer a set of key ideas to help guide teacher planning.

Please note that these materials are principally for professional development purposes. Unlike a textbook scheme they are not designed to be directly lifted and used as teaching materials. The materials can support teachers to develop their subject and pedagogical knowledge and so help to improve mathematics teaching in combination with other high-quality resources, such as textbooks.

Overview

At Key Stage 2, students encountered the concept of central tendency and learnt how to calculate the (arithmetic) mean. At Key Stage 3, they will develop their knowledge of calculating measures of central tendency to include the mode and median, will work with grouped data and be introduced to a measure of spread in statistics: range. This will enable students to engage in more sophisticated data analysis.

While calculating measures of central tendency accurately and efficiently is important, this should not be the dominant aspect of the learning and teaching in this core concept. It is vital that students have a sense of what the measures of central tendency are actually measuring, and engage in activities which prompt questions, such as:

- How can we use measures of central tendency to compare sets of data?
- What do these measures tell us? For example, '*On average, who has the most pocket money: class A or class B?*'
- How do these measures change when particular data points change? For example, '*When considering the average wage in a company, what difference does it make to the various measures when the company director's salary is added in, or removed?*'

Students will construct scatter graphs for the first time, building on the representations covered at Key Stage 2 (bar charts, pie charts and pictograms). Constructing pie charts at Key Stage 3 will involve students making connections with angles, fractions and percentages and using rulers, protractors and angle measurers. Again, while the accurate construction of such diagrams is important in order to communicate findings clearly, it is also necessary for students to think about when a particular statistical diagram is appropriate and what each type of diagram is communicating about the data. Engagement in a range of real-life, contextual problems that require the collection, analysis and representation of data will be an important part of students' study in this area.

This core concept forms the foundations for interpretation and reasoning in core concept *5.2 Statistical analysis* and learning about further statistics in Key Stage 4 and beyond.

Prior learning

Before beginning to teach *Statistical representations and measures* at Key Stage 3, students should already have a secure understanding of the following from previous study:

Key stage	Learning outcome
Upper Key Stage 2	• Calculate and interpret the mean as an average. • Draw given angles and measure them in degrees (°). • Interpret and construct pie charts and line graphs and use these to solve problems.

You may find it useful to speak to your partner schools to see how the above has been covered and the language used.

You can find further details regarding prior learning in the following segment of the NCETM primary mastery professional development materials[1]:

- Year 6: 2.26 Mean average and equal shares

Checking prior learning

The following activities from the NCETM primary assessment materials[2] offer useful ideas for assessment, which you can use in your classes to check whether prior learning is secure:

Reference	Activity
Year 6 page 38	*Ten pupils take part in some races on Sports Day, and the following times are recorded.* *Time to run 100 m (seconds): 23, 21, 21, 20, 21, 22, 24, 23, 22, 20.* *Time to run 100 m holding an egg and spoon (seconds): 45, 47, 49, 43, 44, 46, 78, 46, 44, 48.* *Time to run 100 m in a three-legged race (seconds): 50, 83, 79, 48, 53, 52, 85, 81, 49, 84.* *Calculate the mean average of the times recorded in each race.* *For each race, do you think that the mean average of the times would give a useful summary of the ten individual times?* *Explain your decision.*

Key vocabulary

Term	Definition
(arithmetic) mean	The sum of a set of numbers, or quantities, divided by the number of terms in the set. Example: The arithmetic mean of 5, 6, 14, 15 and 45 is (5 + 6 + 14 + 15 + 45) ÷ 5 i.e. 17.
bivariate data	Data that compares the values of two variables by pairing each value of one of the variables with a value of the other.
measure of central tendency	In statistics, a measure of how the values of a particular variable are located in terms of the values collected for a particular sample, or for the relevant population as a whole. In school mathematics up to Key Stage 4, there are three important measures of central tendency: the arithmetic mean, the median and the mode. These are all statistical averages and often one is more useful than another, depending on the spread of the values under consideration.

measures of dispersion	Measures of dispersion describe how much a set of data is spread out or dispersed. They include the range (where you subtract the lowest score from the highest score) and the standard deviation, which calculates the spread of values around the mean.
median	The middle number or value when all values in a set of data are arranged in ascending order. Example: The median of 5, 6, 14, 15 and 45 is 14. When there is an even number of values, the arithmetic mean of the two middle values is calculated. Example: The median of 5, 6, 7, 8, 14 and 45 is (7 + 8) ÷ 2 i.e. 7.5 The median is one example of an average.
mode	The most commonly occurring value or class with the largest frequency. Example: The mode of this set of data: 2, 3, 3, 3, 4, 4, 5, 5, 6, 7, 8 is 3. Some sets of data may have more than one mode.
range	A measure of spread in statistics. The difference between the greatest value and the least value in a set of numerical data.
scatter graph	A graph on which paired observations are plotted and which may indicate a relationship between the variables. Example: The heights of a number of people could be plotted against their arm span measurements. If height is roughly related to arm span, the points that are plotted will tend to lie along a line.

Collaborative planning

Below we break down each of the two statements within *Statistical representations and measures* into a set of key ideas to support more detailed discussion and planning within your department. You may choose to break them down differently depending on the needs of your students and timetabling; however, we hope that our suggestions help you and your colleagues to focus your teaching on the key points and avoid conflating too many ideas.

Please note: We make no suggestion that each key idea represents a lesson. Rather, the 'fine-grained' distinctions we offer are intended to help you think about the learning journey irrespective of the number of lessons taught. Not all key ideas are equal in length and the amount of classroom time required for them to be mastered will vary, but each is a noteworthy contribution to the statement of knowledge, skills and understanding with which it is associated.

The following letters draw attention to particular features:

D Suggested opportunities for **deepening** students' understanding through encouraging mathematical thinking.

L Examples of shared use of **language** that can help students to understand the structure of the mathematics. For example, sentences that all students might say together and be encouraged to use individually in their talk and their thinking to support their understanding (for example, *'The smaller the denominator, the bigger the fraction.'*).

R Suggestions for use of **representations** that support students in developing conceptual understanding as well as procedural fluency.

V Examples of the use of **variation** to draw students' attention to the important points and help them to see the mathematical structures and relationships.

PD Suggestions of questions and prompts that you can use to support a **professional development** session.

For selected key ideas, marked with an asterisk (*), we exemplify the common difficulties and misconceptions that students might have and include elements of what teaching for mastery may look like. We provide examples of possible student tasks and teaching approaches, together with suggestions and prompts to support professional development and collaborative planning. You can find these at the end of the set of key ideas.

Key ideas

5.1.1 Understand and calculate accurately measures of central tendency and spread

Students will calculate statistical measures of central tendency (mean, median and mode) and spread (range). Students should appreciate how these values, which summarise a set of data in some way, are affected by extra data being added to the whole data set and how such values can be found by comparing averages before and after the inclusion of additional data.

5.1.1.1* Understand what the mean is measuring, how it is measuring it and calculate the mean from data presented in a range of different ways

5.1.1.2 Understand what the median is measuring, how it is measuring it and find the median from data presented in a range of different ways

5.1.1.3* Understand what the mode is measuring, how it is measuring it and identify the mode from data presented in a range of different ways

5.1.1.4 Understand what the range is measuring, how it is measuring it and calculate the range from data presented in a range of different ways

5.1.2 Construct accurately statistical representations

Students will construct all of the Key Stage 3 statistical representations, including representing bivariate data in scatter graphs. They should appreciate the difference between a frequency-based chart (such as a bar chart or pictogram) and a proportion-based chart (such as a pie chart). Teaching should encourage students to think about when one type of chart is more appropriate than another.

5.1.2.1 Construct bar charts from data presented in a number of different ways

5.1.2.2* Construct pie charts from data presented in a number of different ways

5.1.2.3 Construct pictograms from data presented in a number of different ways

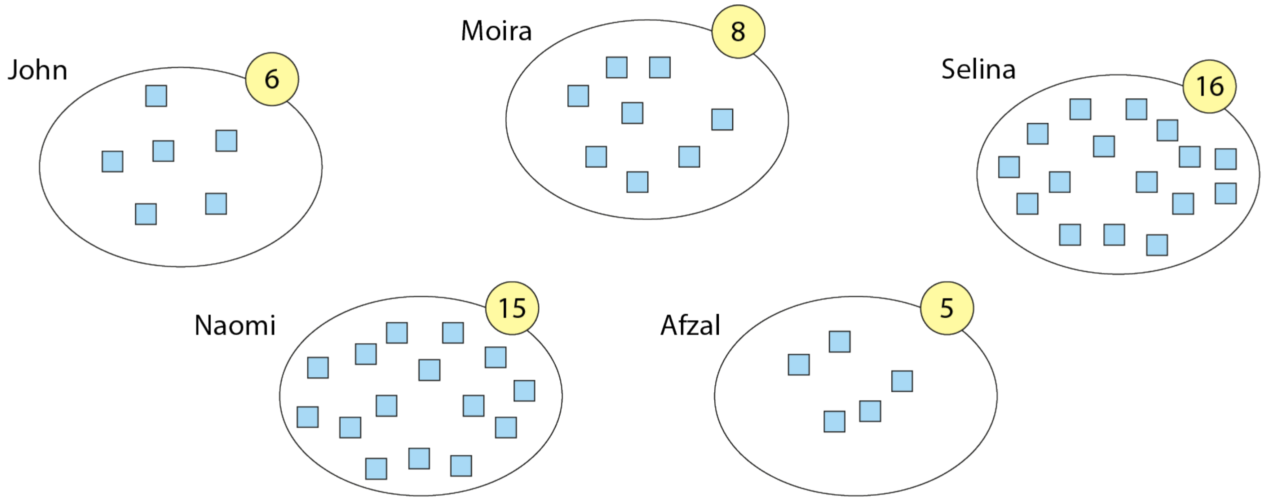
5.1.2.4 Construct scatter graphs from data presented in a number of different ways

Exemplified key ideas

5.1.1.1 Understand what the mean is measuring, how it is measuring it and calculate the mean from data presented in a range of different ways

Common difficulties and misconceptions
Students may know the mean only as a calculation to perform without understanding what the calculation is measuring.

R Various representations could be used to support students to develop a conceptual understanding of the method for finding the mean. For example, you could use counters, place-value counters, Dienes or multi-link cubes to support the question, *'If five students have a different amount each, how much would they have if, without adding or removing any, they all had the same amount?'*



Students may begin by taking counters from one student and giving them to another until all students have the same amount. Careful questioning on how this might be achieved more efficiently, coupled with a change of example where the numbers involved are much larger, can result in students becoming aware of the fact that the total (6 + 8 + 16 + 15 + 5) needs to be distributed among five people.

When working with grouped data, errors often arise from students not fully understanding that the same values are represented many times in a frequency table.

In many cases, asking students to write out the full data set will help them to appreciate what is represented and how they might calculate with the data.

What students need to understand	Guidance, discussion points and prompts

What students need to understand

Calculate the mean from a list of data values.

Example 1:

James wants to improve his diet. For a fortnight he records the number of portions of vegetables he eats each day. What is the mean daily number of portions of vegetables he eats?

M	T	W	Th	F	Sa	Su
5	3	6	2	2	1	2
5	4	4	2	3	1	3

Guidance, discussion points and prompts

Calculation of the mean requires three operations: the sum of the data values, the count of the number of data and then the division of the sum by the count.

L Students could be encouraged to verbalise what different rows or columns represent, to help them correctly identify the values needed for each step of the calculation. Furthermore, you could encourage students to use the words 'sum' and 'count' to distinguish between the two concepts and calculations.

Example 2:

James also records how many portions of protein he eats each day for ten days. What is the mean number of protein portions he eats in that time period?

2, 3, 1, 1, 2, 0, 2, 1, 1, 2

V The use of data values of zero, such as in *Example 2*, will draw students' attention to what values must be included in the count of the number of data.

Example 3:

Find the mean:

a) 6, 6, 6, 6

b) 6, 7, 5, 6

c) 12, 0, 0, 12

d) 4, 5, 7, 8

V The values in *Example 3* have been chosen to draw students' attention to the features of the mean calculation and what it is measuring. Specific questions that may prompt students' thinking are:

- In part a): '*Is a calculation necessary? Why is the mean 6?*'
- In part b): '*What has changed and what has stayed the same between parts a) and b)?*'
- In part c): '*Why must the zeros be included in the calculation? What would the mean be if they were left out?*'

D To deepen students' understanding, consider asking questions, such as, *'How do the numbers in part d) relate to the numbers in part a)?', 'If you start with all values equivalent to the mean (as in part a)) and then add a value to one number and subtract the same value from another number, will the mean remain the same? Why?'*

Apply knowledge of the mean to arithmetic problems.

Example 4:

Reena times her walk from the bus stop each day for six days. The timings for the first five days are 16, 14, 20, 11 and 17 minutes. On the sixth day Reena calculates that her mean walking time for the six days is 15 minutes. How long did it take for her to walk from the bus stop on the sixth day?

D Offering problems such as *Example 4*, where the answer (the mean) is given and students have to figure out one of the missing values, forces students to work with inverse operations, think more deeply and reason about the mean as a concept. For example, students should reason from the statement '… her mean walking time for the six days is 15 minutes' that the sum of her walking times must be 6×15 minutes.

Example 5:

The mean of eight data values is six. Another piece of data is included and the mean is now seven. What is the value of the ninth piece of data?

D This is another example, similar to *Example 4*, where inverse operations need to be used. Such exercises are less about a student proving they can calculate the mean and more about thinking deeply about what is meant by the mean.

Correctly calculate the mean from a frequency table.

Example 6:

Students were asked to calculate the mean number of goals scored in a set of matches, as recorded in this table.

Goals	Matches
1	3
2	5
4	2
5	3
7	2

a) *One student described their thinking, 'I calculated 50 divided by five because the total number of goals is 50 and there are five items.' What has this student misunderstood?*

b) *Another student stated, 'There are 19 goals and 15 matches, so I worked out the goals divided by matches, 19 divided by 15'.*
Do you agree with this student's method?

c) *What advice would you give the students to ensure they calculate the mean from a frequency table correctly in future?*

V Problems involving frequency tables draw students' attention to what values must be included in the sum and count of data values. In a table such as the one in *Example 6*, students should realise that, in order to calculate the total number of goals scored, they must first multiply the number of goals by the number of matches.

PD How can your teaching support students in understanding strategies such as creating a third column of frequency multiplied by data value? You could revisit how frequency tables are compiled by listing the data that is shown in the frequency table. In *Example 6*, list: one goal, one goal, one goal, two goals, two goals, two goals, two goals, two goals… and ask, '*How many goals are scored altogether?*'

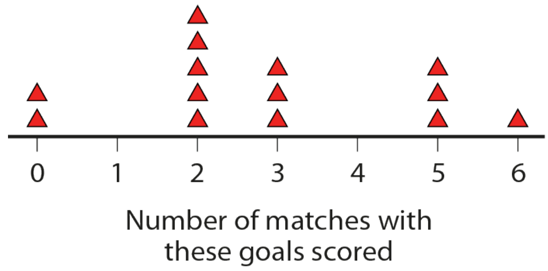
Example 7:

Students were asked to calculate the mean number of goals scored in another set of matches, as recorded in this table.

Goals	Matches
0	2
2	5
3	3
5	3
6	1

Some students used this calculation: 40 ÷ 12. What have they misunderstood?

R A number line representation of the data is a powerful visualisation that could help students move from a list of data values to a frequency table.



Example 8:

Bhanu records how long she can hold the 'plank' position for. What is her mean hold time?

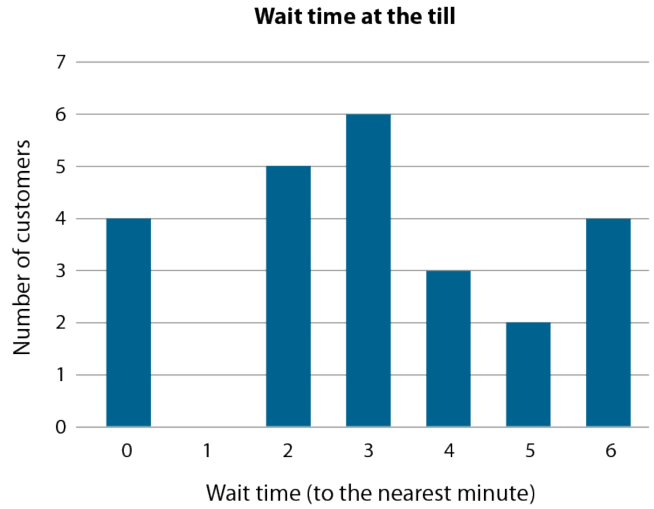
25 s	30 s	45 s	1 min
4	6	5	2

V *Example 8* involves mixed units, drawing students' attention to the need to use a single unit.

Calculate the mean of data represented in a bar chart.

Example 9:

This bar chart shows the length of time people spent waiting at a self-service till (to the nearest minute). What is the mean wait time?



R Students should be encouraged to work with a variety of different representations of data from which to calculate the mean.

When working with bar charts, students should be encouraged to think about what the list of data points would look like and to reason that the vertical axis is indicating the frequency of different wait times.

5.1.1.3 Understand what the mode is measuring, how it is measuring it and identify the mode from data presented in a range of different ways

Common difficulties and misconceptions

Students may not understand what feature of a set of data they are trying to find when identifying the mode. The aim is to summarise the data by finding the most frequently occurring item. Focusing students' attention on the question '*What are we trying to find out about this set of data?*' will help to alleviate any confusion.

Some sets of data may have multiple data items that occur the most frequently and students may wish to say that a data set has three or more modes. It is important for students to recognise that this is not helping to get a representative or summarising value for the set and, hence, the mode may be an inappropriate average.

This key idea provides an opportunity for teachers and students to refer to a wide range of statistical representations, both familiar and unfamiliar.

What students need to understand

Identify the mode from qualitative data.

Example 1:

The following neighbourhood policing statistics show the types of crime reported in one street in September 2018. What type of reported crime was the mode?

Crime types
Anti-social behaviour (1)
Bicycle theft (1)
Burglary (0)
Criminal damage and arson (0)
Drugs (0)
Other crime (0)
Other theft (0)
Possession of weapons (0)
Public order (4)
Robbery (0)
Shoplifting (10)
Theft from the person (0)
Vehicle crime (1)
Violence and sexual offences (1)

Source: Data.Police.UK
Public sector information licensed under the Open Government Licence v3.0

Guidance, discussion points and prompts

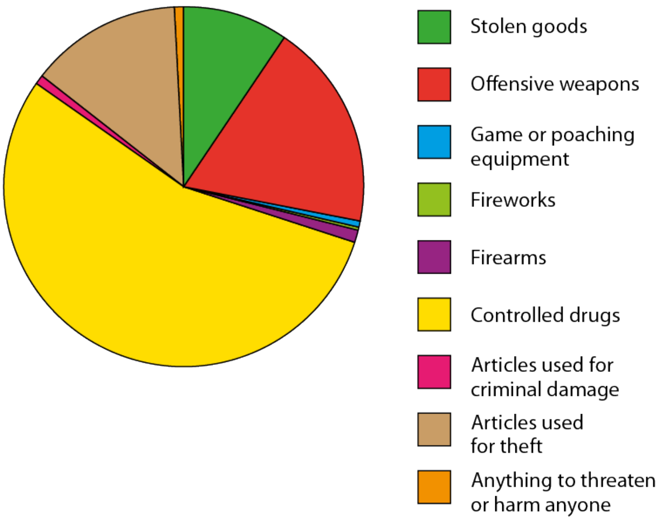
Students may confuse frequencies with data values. For example, in *Example 1*, students may report that '0' is the mode, rather than 'shoplifting'. It is important that students realise that the mode is the most frequent piece of data and not the frequency itself.

PD The Royal Statistical Society recommends that statistical content be drawn from real-life examples. *Example 1* is drawn from actual UK police data. What are the strengths and limitations of using such data? What should you be mindful of if you choose to use it?

Example 2:

This pie chart also refers to neighbourhood policing statistics. What was the modal reason for Humberside Police carrying out a stop and search between April 2018 and September 2018?

Percentage of stop and searches grouped by the reason for the stop, between April 2018 and September 2018.



Source: Data.Police.UK
Public sector information licensed under the Open Government Licence v3.0

L In *Example 2*, students need to appreciate that 'modal' is an adjective and that 'mode' is a noun.

D *Example 2* provides an opportunity for students to deepen their understanding by identifying the mode from a pie chart. They could generalise that the mode in a pie chart is its largest sector.

Example 3:

This list refers to crimes reported in a hypothetical neighbourhood in one month. Change one thing so that 'anti-social behaviour' becomes the mode.

Crime types
Anti-social behaviour (4)
Bicycle theft (2)
Burglary (1)
Criminal damage (5)
Drugs (3)
Possession of weapons (3)
Vehicle crime (0)

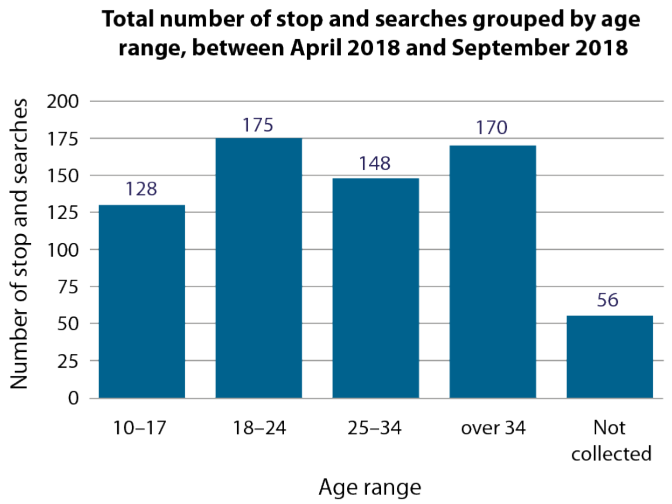
In *Example 3*, ensure that students are not confusing frequency values with data values. They should understand that the mode is 'criminal damage', not '5'.

D Students may suggest that the frequency for 'anti-social behaviour' needs to become greater than five. However, they should also consider reducing the number of reports of 'criminal damage' to less than four.

Identify the modal value from a grouped frequency representation.

Example 4:

This bar chart refers to neighbourhood policing statistics. Identify the modal age group that Humberside Police carried out a stop and search on between April 2018 and September 2018.



Source: Data.Police.UK
Public sector information licensed under the Open Government Licence v3.0

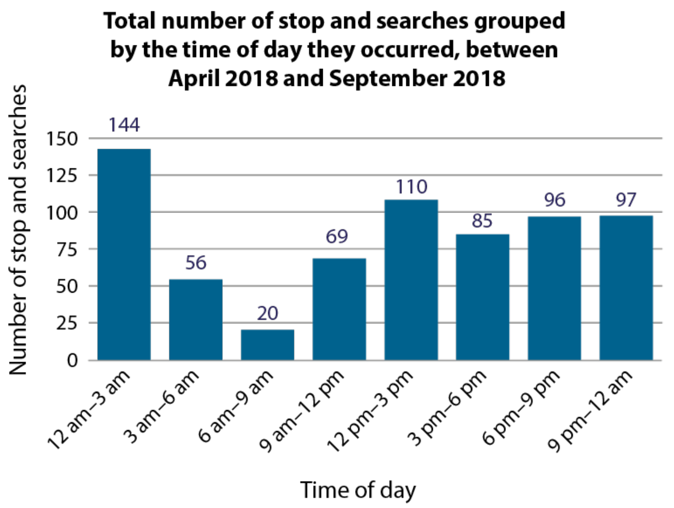
D In *Example 4*, students could be encouraged to generalise that, in a bar chart, the mode is the tallest bar.

You could also extend learning by asking questions, such as, '*What implications does the final bar, labelled "Not collected", have for our confidence in having identified the mode?*'

PD Here, the age group intervals differ. Depending upon how the groups were chosen and how they are used, could it be argued that the modal group is misleading?

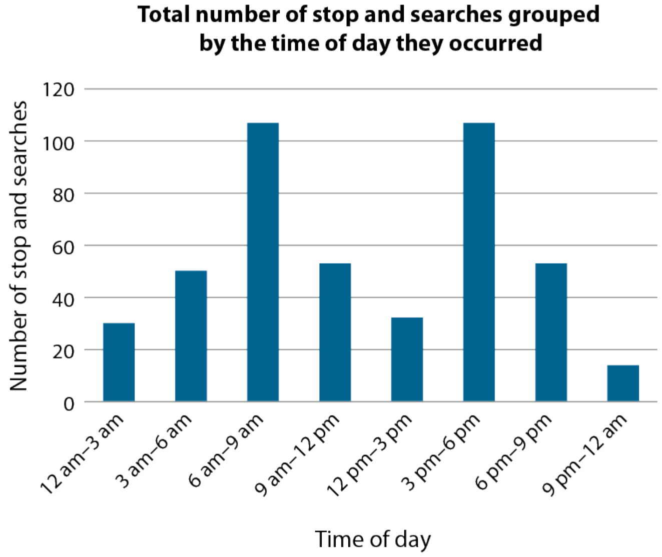
Example 5:

This bar chart refers to neighbourhood policing statistics. Identify the modal time of day that Humberside Police carried out a stop and search between April 2018 and September 2018.



Source: Data.Police.UK
Public sector information licensed under the Open Government Licence v3.0

PD How could you encourage a statistical curiosity with students, using the graph in *Example 5*? For example, why might the modal time interval be between 12 am and 3 am? What neighbourhoods might have a different modal time interval and why?

Work confidently with data sets with no mode. *Example 6:* *This list refers to a hypothetical neighbourhood. What was the modal type of crime reported?* **Crime on or near The Carrs** **8** crimes were reported here in **September 2018**. Anti-social behaviour (0) Bicycle theft (0) Burglary (2) Criminal damage (0) Drugs (2) Possession of weapons (2) Violent crime (2)	**V** *Example 6* is an instance of using a non-example to highlight a characteristic. In this case, the data set is too small to enable the identification of a mode.
Work confidently with data sets that are bimodal. *Example 7:* *This bar chart refers to a hypothetical neighbourhood. Identify the modal time of day that neighbourhood police carried out a stop and search.* 	**V** *Example 7* is an instance of using a non-example to highlight a characteristic. In this case, the data set is best described as bimodal with modes at, perhaps, peak travel times.

5.1.2.2 Construct pie charts from data presented in a number of different ways

Common difficulties and misconceptions

Constructing pie charts may present students with some challenges because it draws upon more than one area of prior learning. Students should have an understanding of multiplicative reasoning, be able to use a calculator and use rulers and angle measurers or protractors to construct lines and angles.

PD How can you ensure that your students are prepared to apply this prior learning to the construction of pie charts?

What students need to understand

Construct pie charts, using the knowledge that angles in a full turn sum to 360°.

Example 1:

Students in a martial arts class were asked how long it takes them to travel to the class. Construct a pie chart to represent these results:

Time taken to travel to the class (nearest minute)	Number of students
15 minutes or less	6
From 16 to 30 minutes	6
From 31 to 45 minutes	12
From 46 to 60 minutes	9
Over an hour	3

Guidance, discussion points and prompts

R In *Example 1*, you could provide pre-drawn circles, with the first radius given, in order to support students in constructing their pie chart with equipment.

Providing circles that are pre-divided helps students connect their understanding of fractions and percentages to how pie charts are a proportional representation.

Example 2:

A class of students were asked about their level of concern towards litter in their community. Construct a pie chart to represent these results:

Level of concern	**Number of students**
Very concerned	3
Somewhat concerned	9
Slightly concerned	12
Not concerned	6

R In *Example 2*, connections can be made between the calculations required to construct the pie chart and students' prior learning of multiplicative reasoning. For example, it may be appropriate for students to use a bar model, such as the one below, to calculate the angle size for each sector.

3	9	12	6

360°

?	?	?	?

R Alternatively, students could use a ratio table:

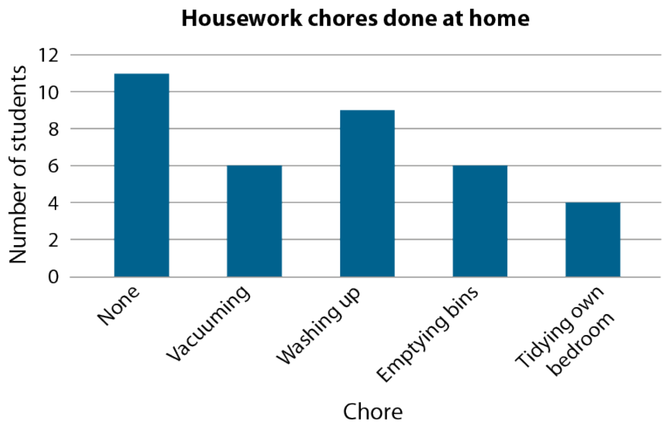
No. of people	30	1	3	9	12	6
Angle size (°)	360°					

V You could use data sets of more than 360 data values to draw students' attention to the use of pie charts as a proportional representation.

Construct pie charts using data taken from other representations.

Example 3:

Students were asked to specify the main housework chore they typically complete at home. Their responses are shown in the bar chart below.



Construct a pie chart to represent this data.

R Offering students representations can be useful to support their understanding. However, even more powerful is to ask students to represent (literally 're-present') a situation themselves.

Example 4:

Students were asked about their attitude to school uniform. Construct a pie chart to represent this data.

Attitude to school uniform	**☺ represents two students**
In favour of a strict uniform policy	☺☺☺☺☺☺
In favour of a dress code	☺☺☺☺☺☺☺☺☺
In favour of no uniform or dress code	☺☺☺☺☺☺☺☺☺☺
Undecided	☺☺☺☺☺

D Offering some data in one form and asking for it to be re-presented in a different one, as in *Examples 3* and *4* here, can support and challenge students to deepen their thinking.

Weblinks

1 NCETM primary mastery professional development materials
https://www.ncetm.org.uk/resources/50639

2 NCETM primary assessment materials
https://www.ncetm.org.uk/resources/46689