

YEAR 8 - REASONING WITH DATA... The data handling cycle

@whisto_maths

What do I need to be able to do?

By the end of this unit you should be able to:

- Set up a statistical enquiry
- Design and criticise questionnaires
- Draw and interpret multiple bar charts
- Draw and interpret line graphs
- Represent and interpret grouped quantitative data
- Find and interpret the range
- Compare distributions

Keywords

Hypothesis: an idea or question you want to test

Sampling: the group of things you want to use to check your hypothesis

Primary Data: data you collect yourself

Secondary Data: data you source from elsewhere e.g the internet/ newspapers/ local statistics

Discrete Data: numerical data that can only take set values

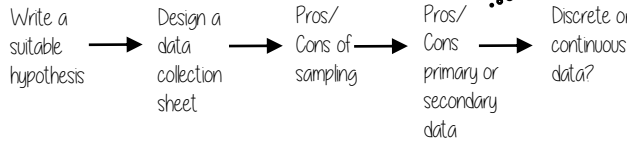
Continuous Data: numerical data that has an infinite number of values (often seen with height, distance, time)

Spread: the distance/ how spread out/ variation of data

Average: a measure of central tendency – or the typical value of all the data together

Proportion: numerical relationship that compares two things

Set up a statistical enquiry



Features of a data collection sheet

Data Title	Tally	Frequency
Grouped or ungrouped categories		Total number of that group observed

Design and criticise a questionnaire

The Question - be clear with the question - don't be too leading/ judgemental

e.g How much pocket money do you get a week?

Responses - do you want closed or open responses? - do any options overlap? - Have you an option for all responses?

Zero option → £0 £0.01- £2 £2.01- £4 more than £4 ← More option

NOTE: For responses about continuous data include inequalities $< x \leq$

Pictograms, bar and line charts

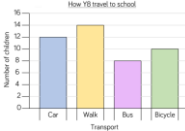
Pictogram

Language	
French	4 circles
Spanish	3 circles
German	1 circle

○ = 4 people

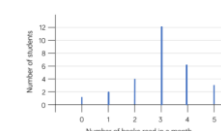
- Need to remember a key
- Visually able to identify mode

Bar Chart



- Gaps between the bars
- Clearly labelled axes
- Scale for the axes
- Title for the bar chart
- Discrete Data

Line Chart



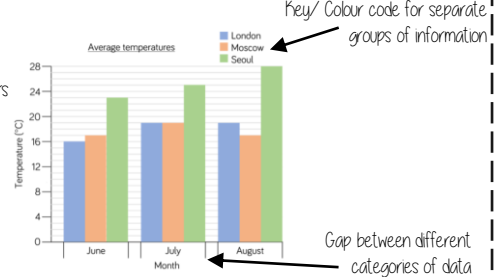
- Gaps between the lines
- Clearly labelled axes
- Scale for the axes
- Discrete Data

Represents quantitative data

Multiple Bar chart

Compares multiple groups of data

- Clearly labelled axes
- Scale for axes
- Comparable data bars drawn next to each other



Draw and interpret Pie Charts

Draw and interpret Pie Charts

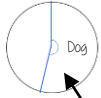
Type of pet	Dog	Cat	Hamster
Frequency	32	25	3

There were 60 people asked in this survey (Total frequency)

$\frac{32}{60}$ "32 out of 60 people had a dog"

This fraction of the 360 degrees represents dogs

$$\frac{32}{60} \times 360 = 192^\circ$$



Use a protractor to draw This is 192°

Multiple method

As 60 goes into 360 - 6 times
Each frequency can be multiplied by 6 to find the degrees (proportion of 360)

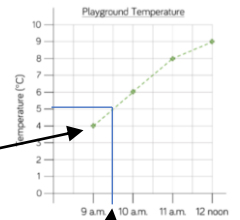
Represents quantitative, discrete data

Draw and interpret line graphs

- Commonly used to show changing over time
- The points are the recorded information and the lines join the points

Line graphs do not need to start from 0

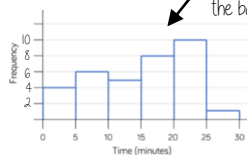
More than one piece of data can be plotted on the same graph to compare data



It is possible to make estimates from the line e.g temperature at 9:30am is 5°C

Grouped quantitative data

Time (minutes)	Frequency
$0 \leq t < 5$	4
$5 \leq t < 10$	6
$10 \leq t < 15$	5
$15 \leq t < 20$	8
$20 \leq t < 25$	10
$25 \leq t < 30$	1



This is a frequency diagram There are no gaps between the bars

Grouping the data is useful if there is a large spread of data to begin with

The use of inequalities shows that this will be a frequency diagram

"More than or equal to 25 and less than 30 minutes"

Find and interpret the range

The range is a measure of **spread**

A smaller range means there is less variation in the results - it is more consistent data

A range of 0 means all the data is the same value

Shop 1 has the smallest range - this indicates it has a more consistent flow of customers each week.

Difference between the biggest and smallest values



Range of customers = $25 - 22 = 3$ (Shop 1)

YEAR 8 - REASONING WITH DATA...

Measures of location

@whisto_maths

What do I need to be able to do?

By the end of this unit you should be able to:

- Understand and use mean, median and mode
- Choose the most appropriate average
- Identify outliers
- Compare distributions using averages and range

Keywords

Spread: the distance/ how spread out/ variation of data

Average: a measure of central tendency – or the typical value of all the data together

Total: all the data added together

Frequency: the number of times the data values occur

Represent: something that shows the value of another

Outlier: a value that stands apart from the data set

Consistent: a set of data that is similar and doesn't change very much

Mean, Median, Mode

The Mean

A measure of average to find the central tendency... a typical value that represents the data

24, 8, 4, 11, 8

Find the sum of the data (add the values) 55

Divide the overall total by how many pieces of data you have $55 \div 5$

Mean = 11

The Median

The value in the center (in the middle) of the data

24, 8, 4, 11, 8

Put the data in order

4, 8, 8, 11, 24

Find the value in the middle

4, 8, 8, 11, 24

Median = 8

NOTE: If there is no single middle value find the mean of the two numbers left

The Mode (The modal value)

This is the number OR the item that occurs the most (it does not have to be numerical)

24, 8, 4, 11, 8

This can still be easier if the data is ordered first

4, 8, 8, 11, 24

Mode = 8

Choosing the appropriate average

The average should be a representative of the data set – so it should be compared to the set as a whole - to check if it is an appropriate average

Here are the weekly wages of a small firm

£240 £240 £240 £240 £240
£260 £260 £300 £350 £700

Which average best represents the weekly wage?

The Mean = £307

The Median = £250

The Mode = £240

Put the data back into context

Mean/Median – too high (most of this company earn £240)

Mode is the best average that represents this wage

It is likely that the salaries above £240 are more senior staff members – their salary doesn't represent the average weekly wage of the majority of employees

Identify outliers

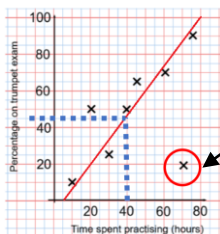
Outliers are values that stand well apart from the rest of the data

Outliers can have a big impact on range and mean. They have less impact on the median and the mode

Sometimes it is best to not use an outlier in calculations

Height in cm
152 150 142 158 182 151 153 149 156 160 151 144

Where an outlier is identified try to give it some context. This is likely to be a taller member of the group. Could it be an older student or a teacher?



Outliers can also be identified graphically e.g. on scatter graphs

Comparing distributions

Comparisons should include a statement of average and central tendency, as well as a statement about spread and consistency

Here are the number of runs scored last month by Lucy and James in cricket matches

Lucy: 45, 32, 37, 41, 48, 35

James: 60, 90, 41, 23, 14, 23

Lucy

Mean: 39.6 (1dp), Median: 38, Mode: no mode, Range: 16

James

Mean: 41.8 (1dp), Median: 32, Mode: 23, Range: 76

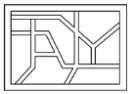
James has two extreme values that have a big impact on the range

"James is less consistent than Lucy because his scores have a greater range. Lucy performed better on average because her scores have a similar mean and a higher median"

YEAR 8 - PROPORTIONAL REASONING...

Ratio and Scale

@whisto_maths



What do I need to be able to do?

By the end of this unit you should be able to:

- Simplify any given ratio
- Share an amount in a given ratio
- Solve ratio problems given a part

Solutions should be modelled, explained and solved

Keywords

Ratio: a statement of how two numbers compare

Equal Parts: all parts in the same proportion, or a whole shared equally

Proportion: a statement that links two ratios

Order: to place a number in a determined sequence

Part: a section of a whole

Equivalent: of equal value

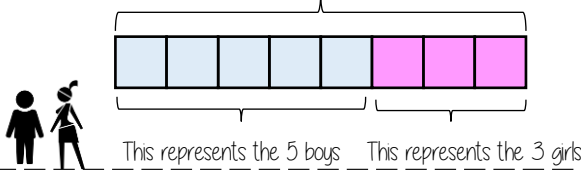
Factors: integers that multiply together to get the original value

Scale: the comparison of something drawn to its actual size.

Representing a ratio

"For every 5 boys there are 3 girls"

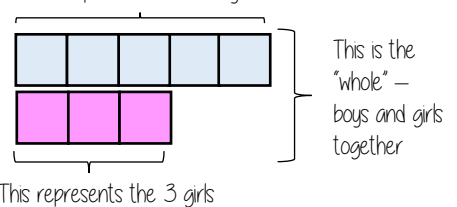
This is the "whole" - boys and girls together



5:3

This represents the 5 boys

Double Number Line



Order is Important

"For every dog there are 2 cats"



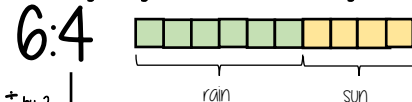
The ratio has to be written in the same order as the information is given

e.g. 2:1 would represent 2 dogs for every 1 cat ✗

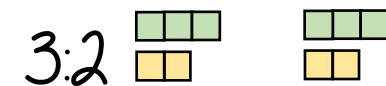
Simplifying a ratio

Cancel down the ratio to its lowest form

"For every 6 days of rain there are 4 days of sun"



+ by 2 ↓



3:2

"For every 3 days of rain there are 2 days of sun" - when this happens twice the ratio becomes 6:4.

Find the biggest common factor that goes into all parts of the ratio

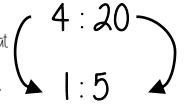
For 6 and 4 the biggest factor (number that multiplies into them is 2)

Ratio In (or n:1)

This is asking you to cancel down until the part indicated represents 1

Show the ratio 4:20 in the ratio of 1:n

The question states that this part has to be 1 unit. Therefore Divide by 4



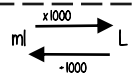
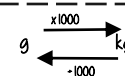
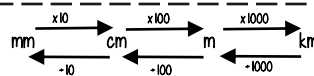
This side has to be divided by 4 too - to keep in proportion

**The n part does not have to be an integer for this type of question

Units are important:

When using a ratio - all parts should be in the same units

Useful Conversions



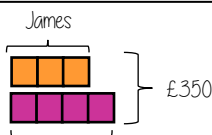
Sharing a whole into a given ratio

James and Lucy share £350 in the ratio 3:4. Work out how much each person earns

Model the Question

James: Lucy

3:4



Lucy

£350 ÷ 7 = £50

□ = one part = £50

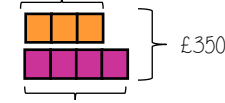
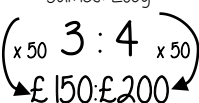
Find the value of one part

Whole: £350
7 parts to share between (3 James, 4 Lucy)

Put back into the question

James: Lucy

James = 3 x £50 = £150



Lucy = 4 x £50 = £200

Finding a value given 1:n (or n:1)

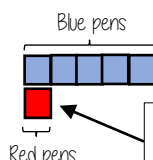
Inside a box are blue and red pens in the ratio 5:1. If there are 10 red pens how many blue pens are there?

Model the Question

Blue: Red

5:1

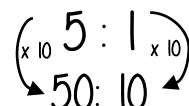
□ = one part = 10 pens



Put back into the question

Blue: Red

Blue pens = 5 x 10 = 50 pens



Red pens = 1 x 10 = 10 pens

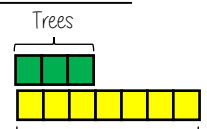
There are 50 Blue Pens



Ratio as a fraction

Trees: Flowers

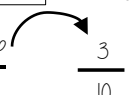
3:7



There are 3 parts for trees

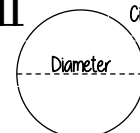
Fraction of trees

Number of parts in group
Total number of parts



Trees parts 3 + Flower parts 7 = 10

π



The ratio of a circles circumference to its diameter

YEAR 8 - REPRESENTATIONS...

Representing Data

@whisto_maths

What do I need to be able to do?

By the end of this unit you should be able to:

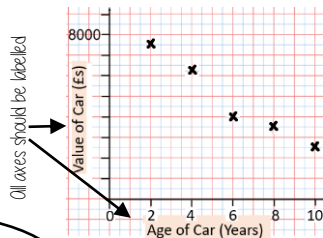
- Draw and interpret scatter graphs
- Describe correlation and relationships
- Identify different types of non-linear relationships
- Design and complete an ungrouped frequency table
- Read and interpret grouped tables (discrete and continuous data)
- Represent data in two way tables

Keywords

- Variable:** a quantity that may change within the context of the problem
- Relationship:** the link between two variables (items) Eg Between sunny days and ice cream sales
- Correlation:** the mathematical definition for the type of relationship.
- Origin:** where two axes meet on a graph
- Line of best fit:** a straight line on a graph that represents the data on a scatter graph
- Outlier:** a point that lies outside the trend of graph
- Quantitative:** numerical data
- Qualitative:** descriptive information, colours, genders, names, emotions etc
- Continuous:** quantitative data that has an infinite number of possible values within its range
- Discrete:** quantitative or qualitative data that only takes certain values
- Frequency:** the number of times a particular data value occurs

Draw and interpret a scatter graph

Age of Car (Years)	2	4	6	8	10
Value of Car (Es)	7500	6250	4000	3500	2500



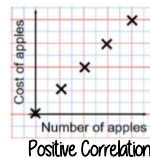
- This data may not be given in size order
- The data forms information pairs for the scatter graph
- Not all data has a relationship

The link between the data can be explained verbally

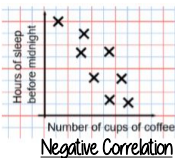
"This scatter graph show as the age of a car increases the value decreases"

The axis should fit all the values on and be equally spread out

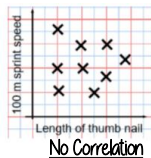
Linear Correlation



As one variable increases so does the other variable



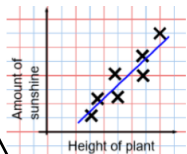
As one variable increases the other variable decreases



There is no relationship between the two variables

The line of best fit

The Line of best fit is used to make estimates about the information in your scatter graph



It is only an estimate because the line is designed to be an average representation of the data

It is always a straight line.

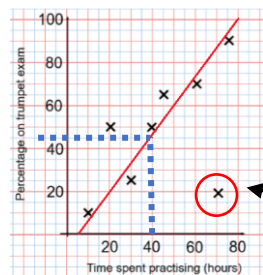
Things to know:

- The line of best fit **DOES NOT** need to go through the origin (The point the axes cross)
- There should be approximately the same number of points above and below the line (It may not go through any points)
- The line extends across the whole graph

Using a line of best fit

Interpolation is using the line of best fit to estimate values inside our data point

e.g 40 hours revising predicts a percentage of 45



Extrapolation is where we use our line of best fit to predict information outside of our data

This is not always useful – in this example you cannot score more than 100%. So revising for longer can not be estimated

This point is an "outlier" it is an outlier because it doesn't fit this model and stands apart from the data

Ungrouped Data

The number of times an event happened

The table shows the number of siblings students have. The answers were
3, 1, 2, 2, 0, 3, 4, 1, 1, 2, 0, 2

2 people had 0 siblings. This means there are 0 siblings to be counted here

Number of siblings	Frequency
0	2
1	3
2	4
3	2
4	1

0 → 0
3 → 3
2 + 2 + 2 + 2 OR 2 x 4 = 8
3 + 3 OR 3 x 2 = 6
4 → 4

2 people have 3 siblings so there are 6 siblings in total

Best represented by discrete data (Not always a number)

OVERALL there are 0 + 3 + 8 + 6 + 4 Siblings = 21 siblings

Grouped Data

If we have a large spread of data it is better to group it. This is so it is easier to look for a trend. Form groups of equal size to make comparison more valid and spread the groups out from the smallest to the largest value.

Cost of TV (£)	Tally	Frequency
101 - 150	THH	7
151 - 200	THH THH	11
201 - 250	THH	5
251 - 300		3

Discrete Data
The groups do not overlap

We do not know the exact value of each item in a group – so an estimate would be used to calculate the overall total (Midpoint)

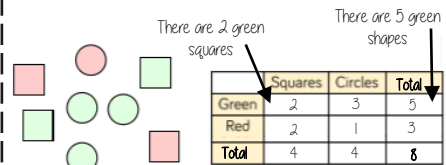
x	Frequency
Weight(g)	
40 < x ≤ 50	1
50 < x ≤ 60	3
60 < x ≤ 70	5

Continuous Data
To make sure all values are included inequalities represent the subgroups

e.g this group includes every weight bigger than 60kg, up to and including 70kg

Representing data in two-way tables

Two-way tables represent discrete information in a visual way that allows you to make conclusions, find probability or find totals of sub groups



Using your two-way table

To find a fraction
e.g What fraction of the items are red? **3 red items**
but **8 items in total** = $\frac{3}{8}$

Interleaving: Use your fraction, decimal percentage, equivalence knowledge