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). E.g. 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.

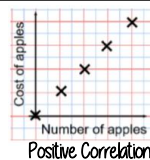| Age of Car (Years)	2	4	6	8	10
Value of Car (£s)	7500	6250	4000	3500	2500
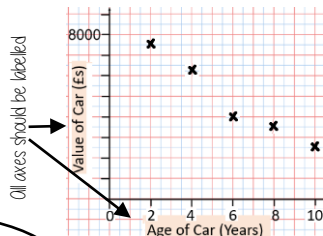
- This data may not be given in size order
- The data forms information pairs for the scatter graph
- Not all data has a relationship

All axes should be labelled

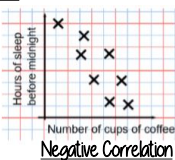*"This scatter graph show as the age of a car increases the value decreases"*

The link between the data can be explained verbally

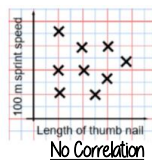The axis should fit all the values on and be equally spread out

Linear Correlation

Positive Correlation — Cost of apples / Number of apples

Negative Correlation — Hours of sleep before midnight / Number of cups of coffee

No Correlation — 100 m sprint speed / Length of thumb nail

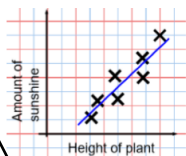| Positive Correlation	Negative Correlation	No Correlation
As one variable increases so does the other variable	As one variable increases the other variable decreases	There is no relationship between the two variables

The line of best fit

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

Amount of sunshine / Height of plant

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

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

It is always a <u>straight line.</u>

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.

Percentage on trumpet exam / Time spent practising (hours)
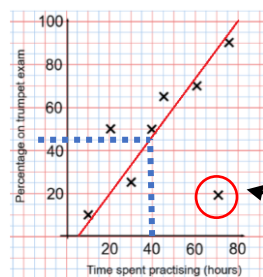
Extrapolation is where we use our line of best fit to predict information outside of our data
This is not always useful – in this example you cannot score more that 100%. So revising for longer can not be estimated

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
3
2 + 2 + 2 + 2 OR 2 x 4 = 8
3 + 3 OR 3 x 2 = 6
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.

Discrete Data — The groups do not overlap

Cost of TV (£)	Tally	Frequency
101 – 150	THL ll	7
151 – 200	THL THL l	11
201 – 250	THL	5
251 - 300	lll	3

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)

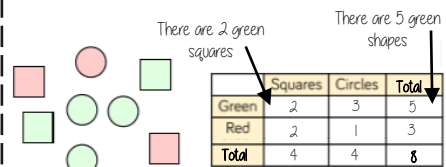**Continuous Data** — To make sure all values are included inequalities represent the subgroups

x Weight(g)	Frequency
$40 < x \le 50$	1
$50 < x \le 60$	3
$60 < x \le 70$	5

e.g this group includes every weight bigger that 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

There are 2 green squares
There are 5 green shapes

	Squares	Circles	Total
Green	2	3	5
Red	2	1	3
Total	4	4	8

There are 8 items in total

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