STATISTICS

TOPICS

- Average and Spread
  - Mean
  - Median
  - Mode
  - Range
- Standard Deviation
- Comparing averages and spread
- Quartiles and Boxplots
  - Lower and upper quartiles
  - Interquartile and semi interquartile range
  - Drawing and reading from boxplots
- Plotting and reading from scattergraphs

MEANS, MEDIANS, MODES AND RANGE

STATISTICS

WHAT IS AVERAGE AND SPREAD?

1

When we find the mean, median or mode we are finding an average for a set of data. This average tells us about the tendency of the data, or gives us a value that represents data as a whole

2

When we calculate the range we are finding out about the **spread**, or how consistent the data is

EXAMPLE 1: BASIC CALCULATIONS

Find the MMMR of the set of data:

201 123 159 162 201 123 159 160 212 135 157 169 212 130 178 160

$$\text{Mean} = \frac{2641}{16} = 165.0625$$

Median = 123 123 130 135 157 159 159 160 | 160 162 169 178 201 201 212 212

= $\underline{\underline{160}}$

Mode = 123, 159, 160, 201, 212

Range = High − Low = 212 − 123 = 89

EXAMPLE 2: CHOOSING AN AVERAGE

A class did a short test. The scores are given below.

11, 14, 56, 56, 57, 61, 62, 64, 64

Which measure of average is most representative of the class results. Explain your answer.

Mode = 56 and 64 – the mode has two values so not a good representation.

Median = 57 – this average represents the upper values well but not 11 and 14.

Mean = 49.44 – the mean is lower than the majority of scores so is not the representative of the list overall.

Although it does not demonstrate that there were scores as low as 11 and 14, the **median** is probably most representative overall of the list.

EXAMPLE 3: MISSING DATA

The mean average weight of 5 people is 64kg.

Four of those people weighed 45kg, 78kg, 63kg and 91kg.

What was the weight of the fifth person?

$$\text{Mean} = \frac{\text{total}}{\text{no.}}$$

$$\frac{45 + 78 + 63 + 91 + ?}{5} = 64$$

$$45 + 78 + 63 + 91 + ? = 320$$

$$? = 320 - 45 - 78 - 63 - 91$$

The missing person's weight is 43 kg.

STARTER

a) Find the mean, median and mode and range of the data below.

162, 264, 114, 156, 164, 211, 157, 161, 256,

114, 156, 157, 161, 162, 164, 211, 256, 264

mean = $\frac{1645}{9} = \underline{182.8}$

median = $\underline{162}$

mode = no mode

range = $264 - 114 = \underline{150}$

COMPLETING 'HOMEWORK'

- You should have completed the page on averages (pg232-233)

- If you have not done this I will be expecting to see it tomorrow:
- Q1a, 2a, 3a, 4a, 5, 6
- Q11-12 on pg234 should be attempted by those aiming for A/B

STANDARD DEVIATION

STATISTICS

WHAT IS STANDARD DEVIATION?

Standard deviation is a measure of **spread**. It tells us how far numbers in a set of data differ from the mean.

Imagine two sets of data: 5, 110, 215, 320, 425 and 213, 214, 215, 216, 217. Both of these data sets have the same mean average of 215.

What do you notice about the consistency of the two data sets?

Whilst the average of the data sets is the same, the **standard deviation** will be different to represent the consistency of values.

Companies or researchers may use SD to test whether products they are manufacturing or selling fit the description used.

THE FORMULA

- This is the formula to calculate standard deviation.

each value

total

mean

'sum of'

'x bar'

$$SD = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}$$

number of values

SETTING UP THE TABLE

We use a SD table to calculate the numerator of the formula.

Example: Find the standard deviation of 5, 110, 215, 320, 425

mean = 215 = $\bar{x}$

list your values

subtract the mean from column 1

square column 2

x	$x - \bar{x}$	$(x-\bar{x})^2$
5	5 − 215 = −210	44100 ← 210^2
110	110 − 215 = −105	11025 ← 105^2
215	215 − 215 = 0	0
320	320 − 215 = 105	11025
425	425 − 215 = 210	44100
		110250

add 3rd column

$$SD = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$$

total of 3rd column

no. in the list

$$= \sqrt{\frac{110250}{5-1}}$$

$$= \sqrt{\frac{110250}{4}} = 166.0195\ldots$$

$$= \underline{\underline{166.02}}$$

TASK

- Page 242 Q3 – Q7
- You must finish these in your spare time if you do not finish them in class.

QUARTILES AND IQR

STATISTICS

QUARTILES

- If the median is halfway through a list, then the quartiles show the quarter and three-quarter points.

> To find the quartiles we first have to find the median. Then we look for the middle point of each half to get the lower and upper quartiles.

Example:

6 4 8 12 9 5 6 14

$$\underbrace{4 \quad 5 \quad 6 \quad 6}_{\text{lower quartile } 5.5} \quad \underset{\text{median} = 7}{\uparrow} \quad \underbrace{8 \quad 9 \quad 12 \quad 14}_{\text{upper quartile } 10.5}$$

ANOTHER…(COPY THIS ONE DOWN)

1a) The heights of 9 people in Town A were recorded.

Find the median and quartiles of the heights.

131.2 149.1 134.3 156.4 174.5 182.3 159.0 163.4 190.6

131.2 134.3 149.1 156.4 159.0 163.4 174.5 182.3 190.6

Lower Quartile = 141. ~~3~~ 7

median 159.0

Upper quartile = 178.4

149.1 − 134.3 = 14.8
14.8 ÷ 2 = 7.4
134.3 + 7.4 = 141. ~~3~~ 7

182.3 − 174.5 = 7.8
7.8 ÷ 2 = 3.9
174.5 + 3.9 = 178.4

(SEMI) INTERQUARTILE RANGE

The interquartile range is the difference between the LQ and the UQ.

The semi-interquartile range is half of the IQR.

1b) Find the IQR and the SIQR of the heights from 1a).

$$IQR = UQ - LQ$$
$$= 178.4 - 141.7$$
$$= 36.7$$

$$\begin{array}{r} 17\overset{7}{\cancel{8}}.{}^{1}4 \\ -\ 141.7 \\ \hline 36.7 \end{array}$$

$$SIQR = 36.7 \div 2$$
$$= 18.35$$

$$2\overline{)36.7}$$

COMPARISON

When you have to make **comments** remember:

Comparing two **means/medians** is comparing the **average.**

Comparing the **standard deviation/range** is comparing the **variation.**

1c) A group of people from Town B had a median of 152.5cm and an IQR of 21.6cm.

Town A
med : 159.0
IQR : 36.7

Make two valid statements comparing Town A and Town B.

- The average height of Town A is taller than Town B as 159.0 > 152.5.
- The variation of heights in Town A is bigger than Town B as 36.7 > 21.2.

TASK

- Leckie and leckie textbook
- Page 102 Ex 7B

Exercise 7B

1 **a** 6, 3, 6, 8, 5

b 38.5, 29, 38.5, 46.5, 17.5

c 5.3, 2.95, 5.3, 7.8, 4.85

2 measurements in cm

a **(i)** 135

(ii) 124, 135, 165

(iii) 41

b **(i)** 42

(ii) 38, 42, 52

(iii) 14

c Indian ears very much smaller and less variation.

3 measurements in £

a **(i)** 149000

(ii) 128000, 149000, 163000

(iii) 35000

b **(i)** 118000

(ii) 99000, 118000, 136000

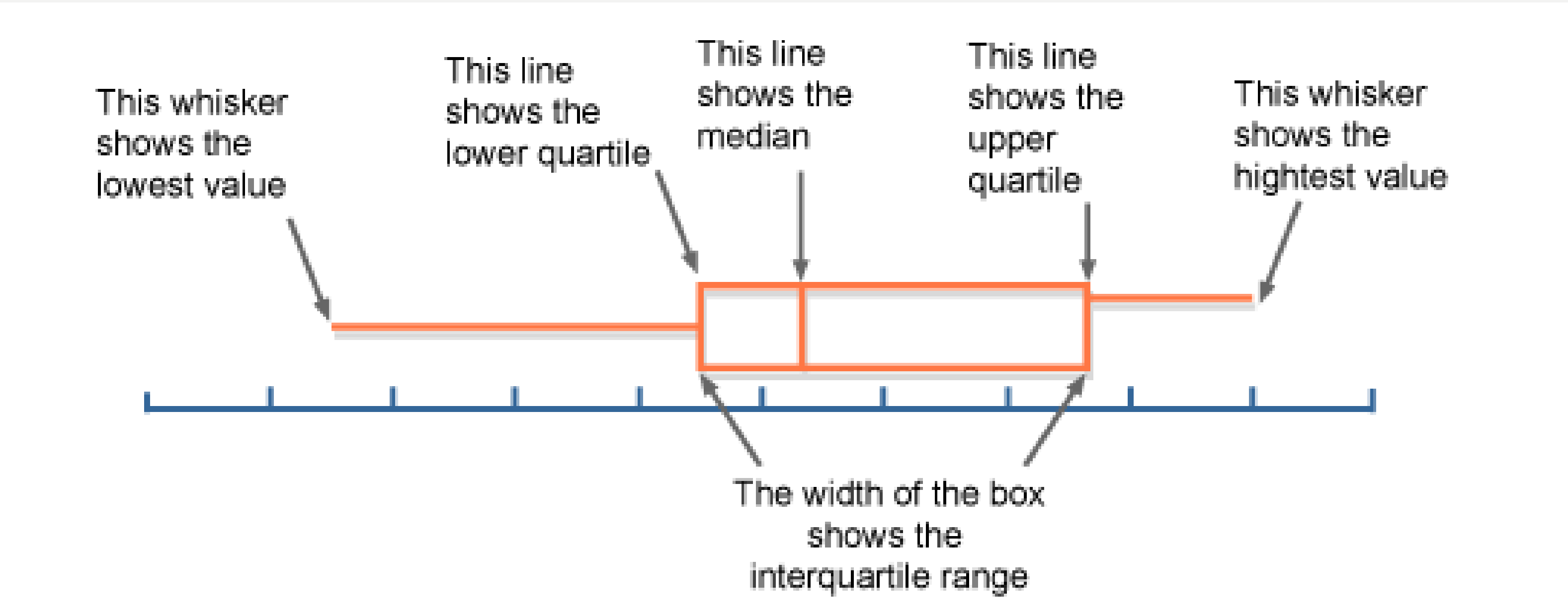
(iii) 37000

BOX PLOTS

STATISTICS

BOX PLOTS

- A box plot represents a summary of data so we can see where the majority of the data lies.
- It is sometimes called a ‘five figure summary’
- It shows us the range, median and interquartile range of the data without having to analyse the values themselves.



DRAWING BOX PLOTS

- We need a scale, and we need some data….

9, 5, 8, 11, 10, 5, 10, 9, 10, 6, 8, 4, 11, 12, 5, 5

* 4, 5, 5, 5, 5, 5, 6, 8, 9, 9, 10, 10, 10, 11, 11, 12

8.5

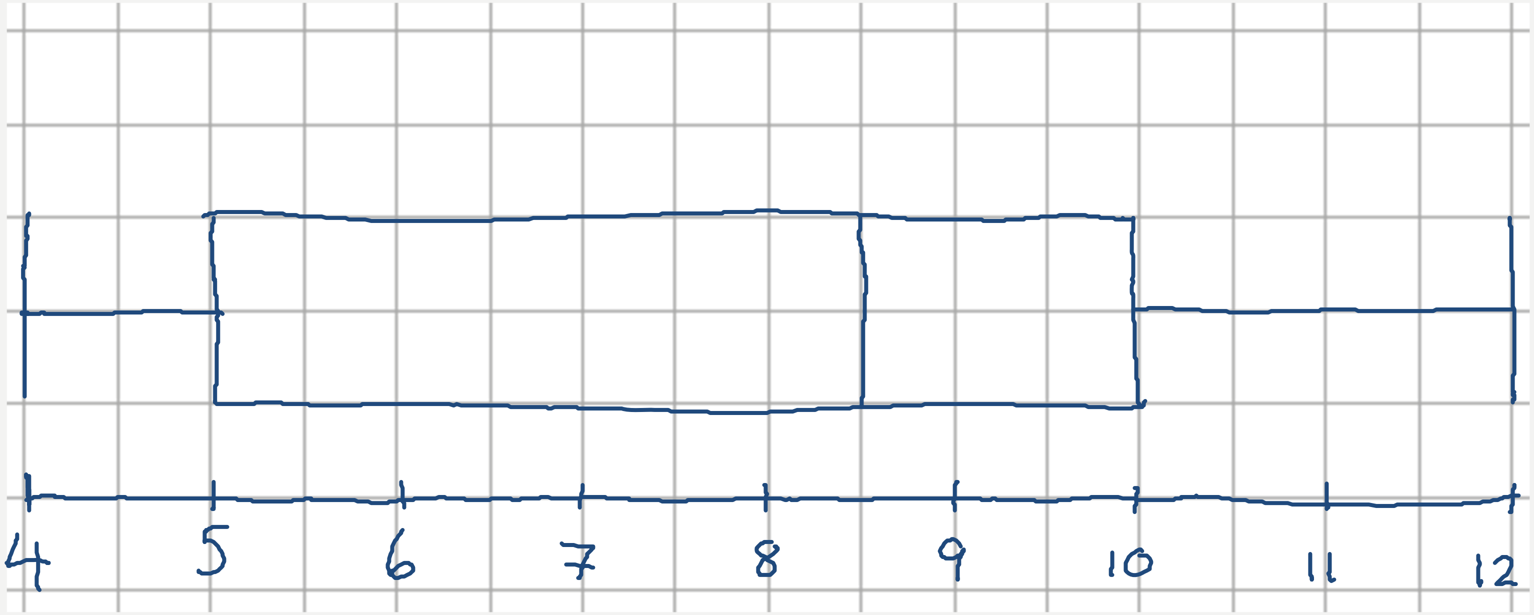
Lowest = 4

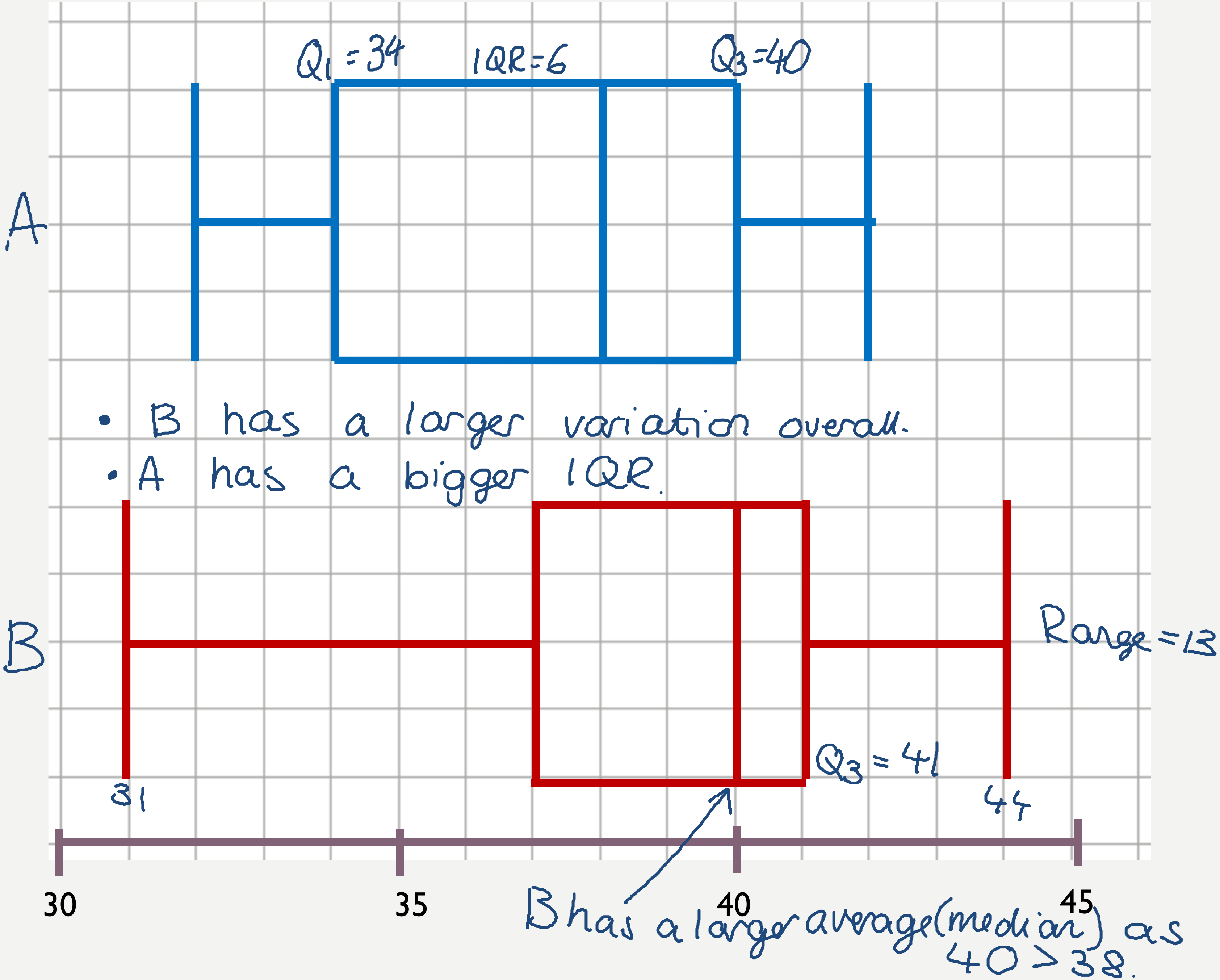
Lower quartile (Q_1) = 5

Median (Q_2) = 8.5

Upper quartile (Q_3) = 10

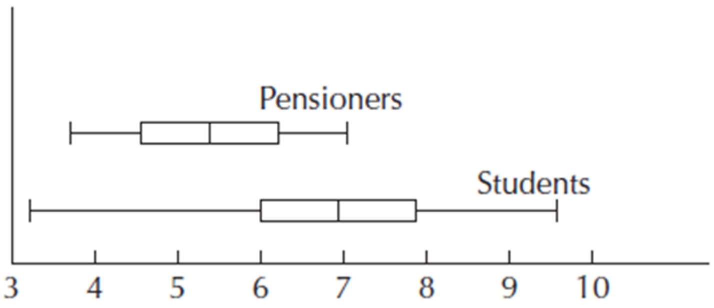
Highest Value = 12


4
5
6
7
8
9
10
11
12



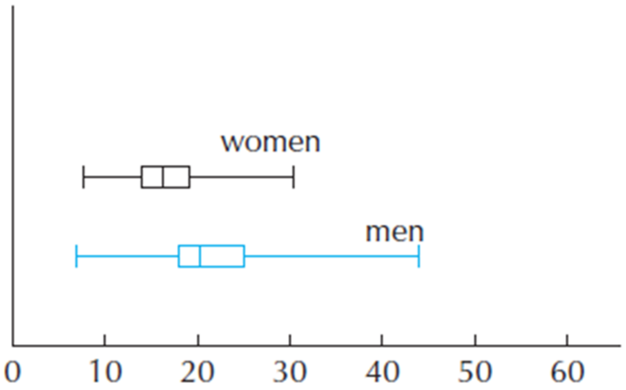
A
B
$Q_1 = 34$
IQR = 6
$Q_3 = 40$
• B has a larger variation overall.
• A has a bigger IQR.
Range = 13
$Q_3 = 41$
31
44
30
35
40
45
B has a larger average (median) as 40 > 38.

2 **a**



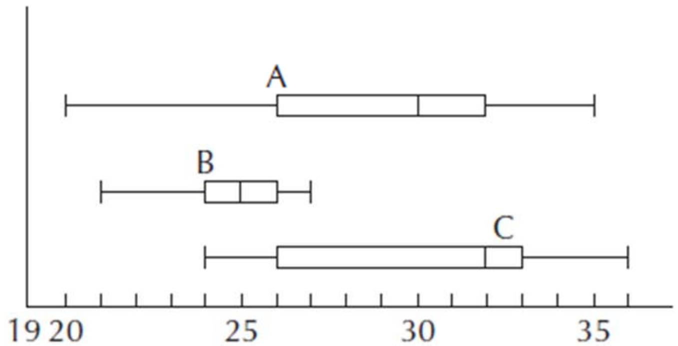
b Much wider range in students. On average pensioners better.

3 **a**



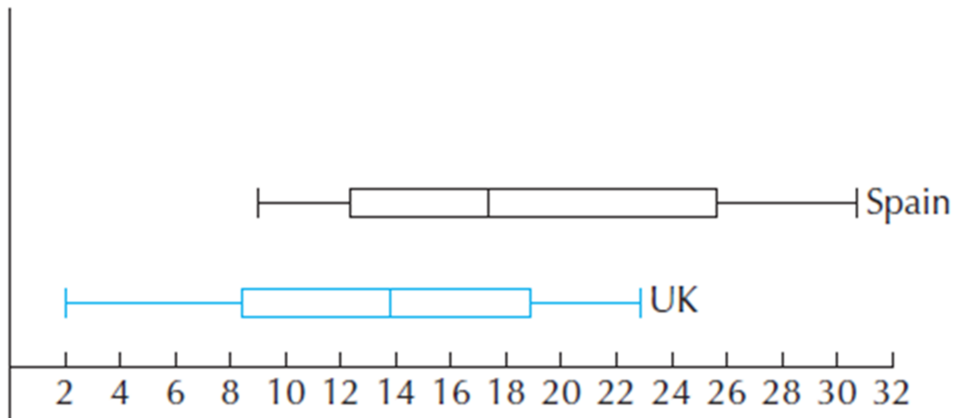
b Men wider range and on average higher. IQR greater.

4 **a**



b C most effective. C has greater IQR and much higher median.

5 **a**



b Spanish tempertures are generally higher, greater IQR.

6 **a** Friday lowest = 27, Q1 = 36, Q2 = 46, Q3 = 51, highest = 61

Saturday lowest = 21, Q1 = 29, Q2 = 43, Q3 = 60, highest = 78, SIQR = 15.5

b 7.5, 15.5

c Median Friday time was slightly higher. Saturday SIQR greater.

7 Although Dr Ball has a higher median waiting time (7 mins vs 5 mins), there are much less extremes. Dr Ball is best as you are never waiting longer than 10 minutes.

8 Town A has greater extremes in temp but median is much lower. Town B better for holiday.

SCATTERGRAPHS

STATISTICS

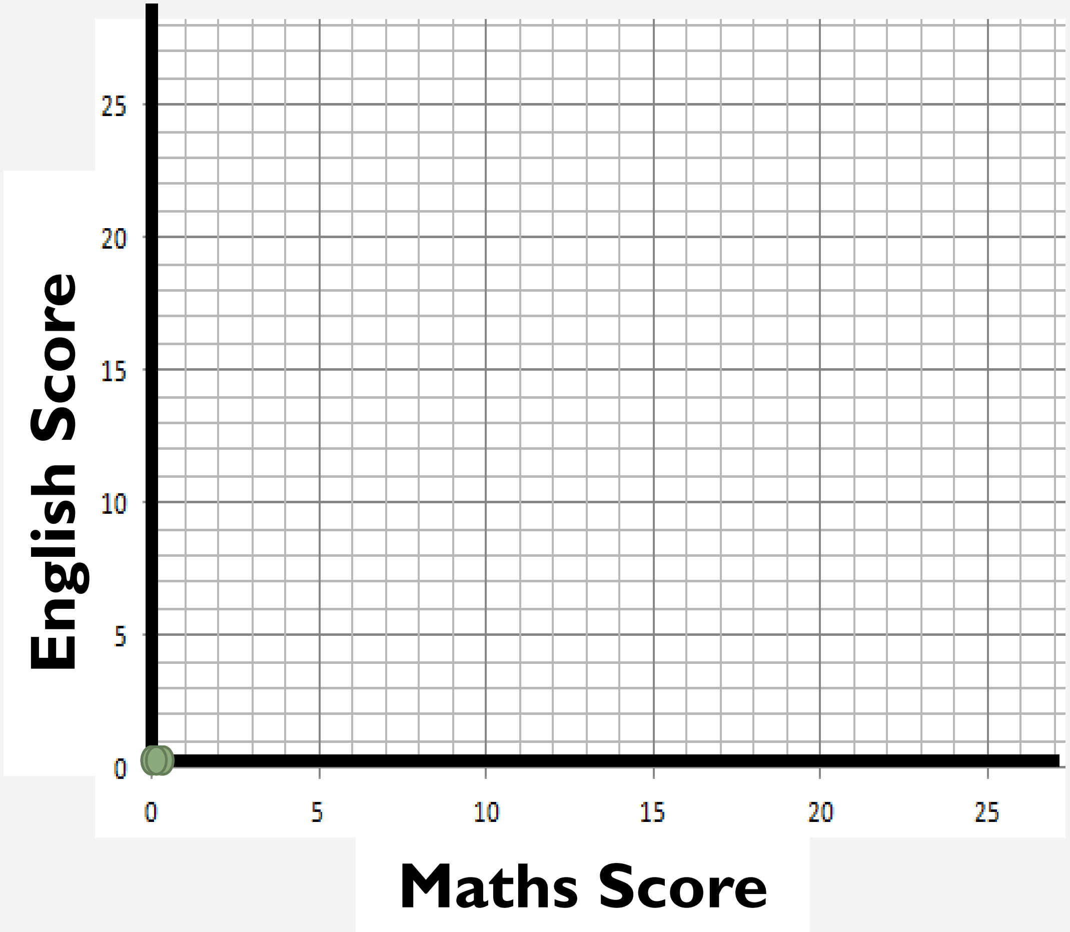
YOU SHOULD ALREADY KNOW…

- How to plot points on a scattergraph
- How to draw a line of best fit

You may already know…

- How to use the line of best fit to estimate one variable given the other
- Different types of correlation

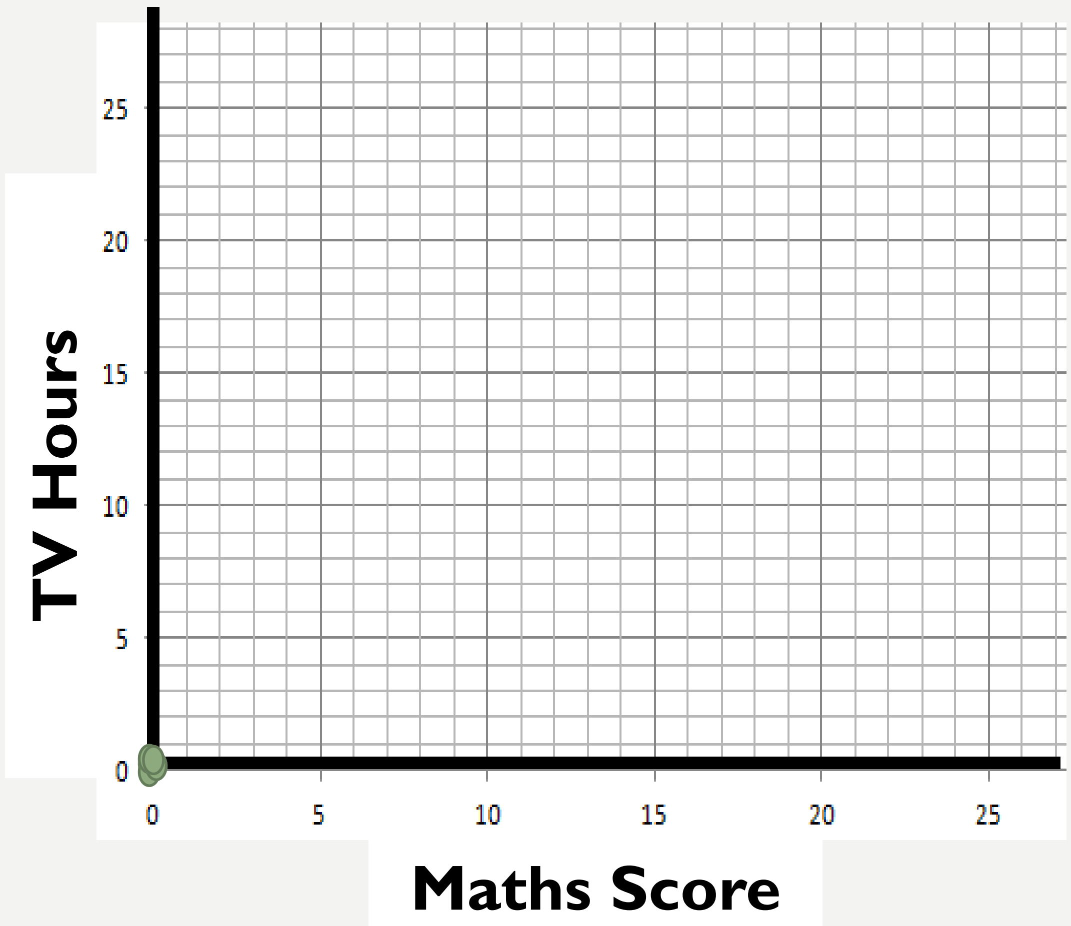
PLOTTING POINTS



Maths score	English Score
10	14
15	22
18	15
25	28
7	6

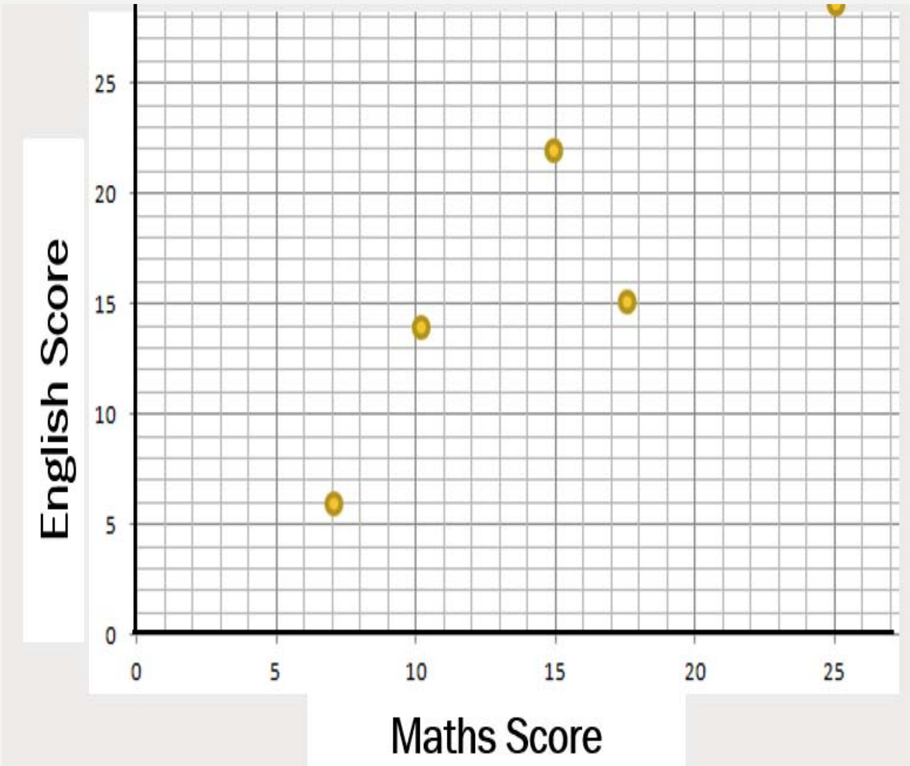
LINE OF BEST FIT

- Same number of points on each side of the line
- Use one or two middle points as a guide
- It has to be straight!
- Doesn't have to go through (0,0)



Maths score	Hours watching TV
7	27
10	20
15	15
18	7
25	9

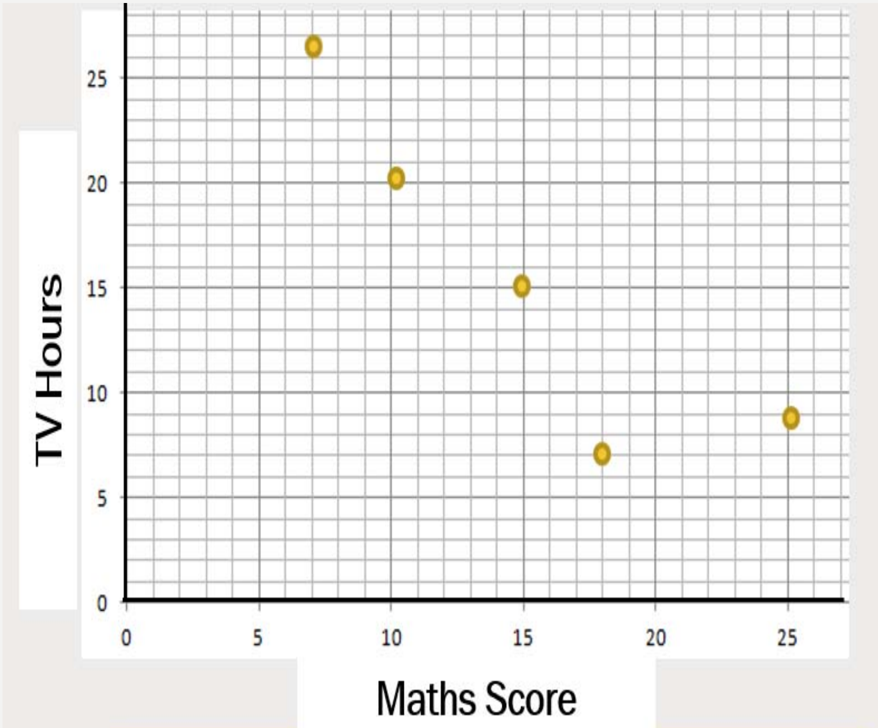
CORRELATION



Positive correlation
As one variable increases so does the other.

Example: Height and shoe size

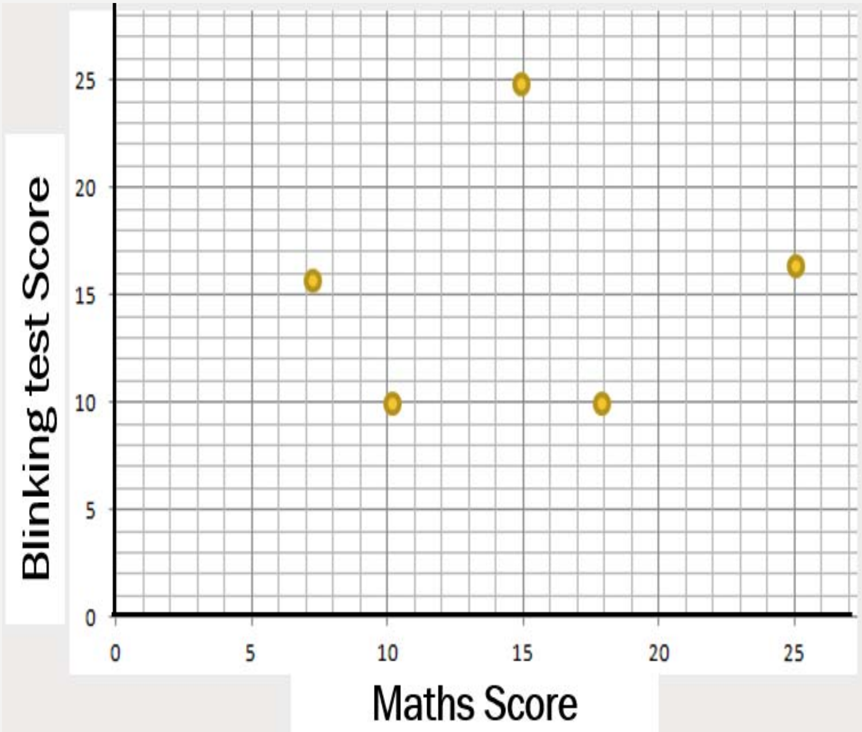
CORRELATION



Negative correlation
As one variable increase the other decreases

Example: Hours of watching tv against hours of exercise

CORRELATION



No correlation
There is no correlation between the variables

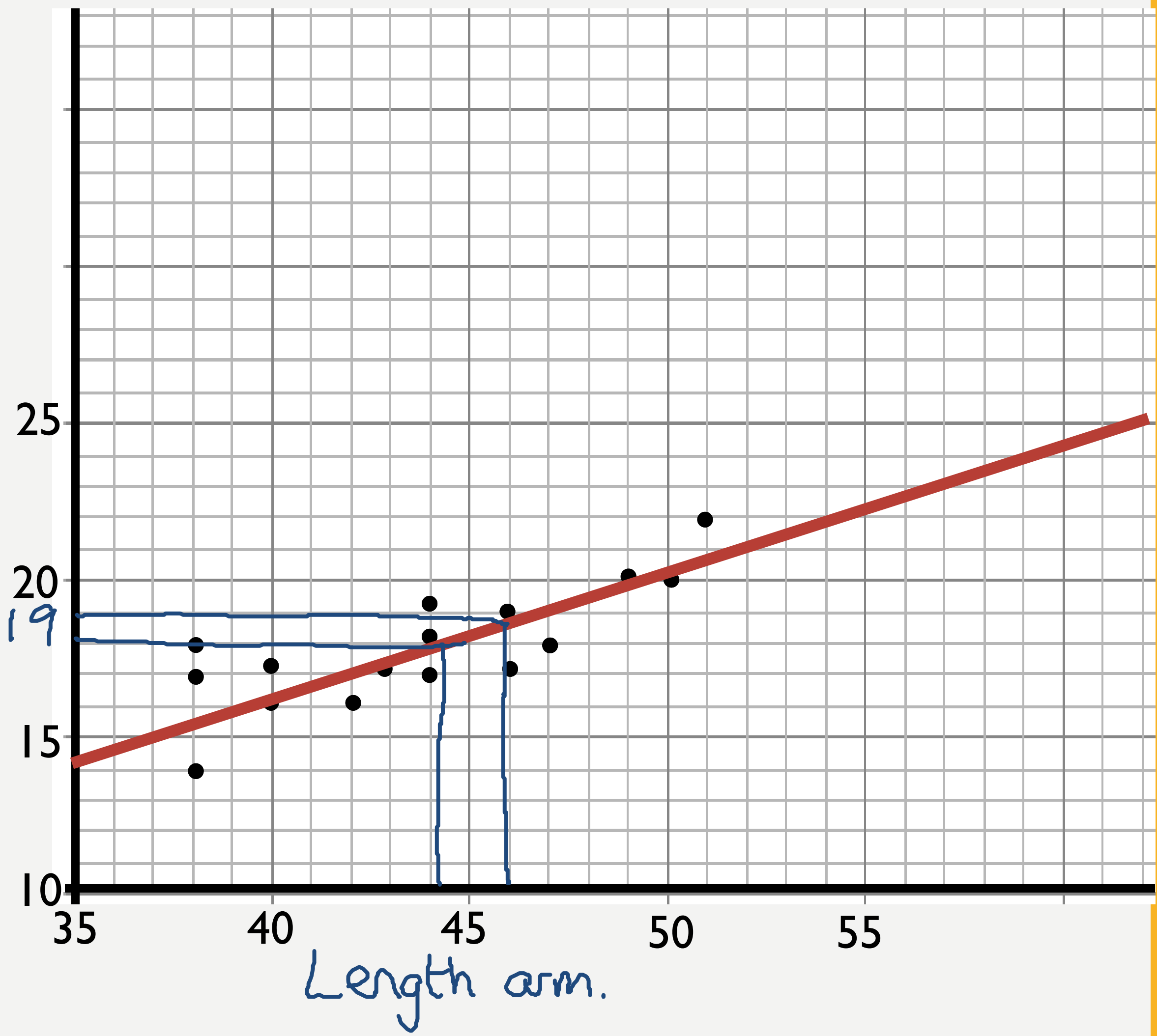
Example: Height and IQ

CORRELATION

- You can also have **strong** and **weak** correlations if the points are very tightly grouped.

- **Task 1**
- **Q1 and Q2**
- **Page 117 L&L**

Length of arm (cm)	Length of hand (cm)
46	19
38	17
43	~~25~~ 17
47	18
51	22
49	20
40	16
44	~~20~~ 19
42	16
44	17
37	18
46	17
50	20
44	18
38	14
40	17



A person with arm length of 46 would be expected to have a hand length of 19cm.

A person with hand length of 18 would be expected to have a arm length of 44cm.