STATISTICS

(EASIEST PORTION)

Central Tendencies and spread → 1) Mode, Median, Mean, Quartiles and range.

Representation of Data.
2) HISTOGRAMS & FREQ. POLYGONS
3) CUMULATIVE FREQUENCY.
4) STEM & LEAF AND BOX AND WHISKER DIAGRAMS.

MODE - MEDIAN - MEAN - QUARTILES - RANGE.

1) RAW DATA (NO FREQUENCY TABLE)

10	8	7	6	8	12	9	11
10	10	6	11	10	7	6	8

Total = n = 16

[1] MODE: (MOST OCCURING VALUE) = 10

[2] MEDIAN: (MIDDLE VALUE)

STEP1: ARRANGE IN INCREASING ORDER

6, 6, 6, 7, 7, 8, 8, 8, 9, 10, 10, 10, 10, 11, 11, 12

Median = $\frac{8+9}{2}$ = 8.5

3 MEAN : (AVERAGE) = $\frac{\text{SUM}}{\text{NO}}$

$$\frac{6+6+6+7+7+8+8+8+9+10+10+10+10+11+11+12}{16}$$

Mean = ☐

4 RANGE = Max − Min

= 12 − 6

= 6

Q: Ages of some friends are as follows.

5, 8, 11, 11, 4, 6

(i) mode = 11

(ii) Median = Step1: Arrange. 4, 5, 6, 8, 11, 11

$= \frac{6+8}{2}$

Median = 7

(iii) Mean = $\frac{5+8+11+11+4+6}{6}$ = ☐

(iv) Range = Max − Min = 11 − 4 = 7

MEDIAN AND QUARTILES

STEP 1: ARRANGE in increasing order

Median = $\left(\frac{n+1}{2}\right)^{th}$ term

Upper Quartile = $\frac{3(n+1)^{th}}{4}$ term.

Lower Quartile = $\frac{(n+1)^{th}}{4}$ term.

Q: 5, 4, 7, 6, 8, 2, 4, 8, 3 $\quad n=9$

Step 1: Arrange 2, 3, 4, 4, 5, 6, 7, 8, 8

(3 → 2nd term; 4 → 3rd term; 5 → 5th term; 7 → 7th term; 8 → 8th term)

Median = $\left(\frac{n+1}{2}\right)^{th}$ term = $\left(\frac{9+1}{2}\right)^{th}$ term = 5^{th} term. → 5

Upper Quartile = $\frac{3(n+1)^{th}}{4}$ term = $\frac{3(9+1)^{th}}{4}$ term = 7.5^{th} term

(7^{th} and 8^{th})

$\frac{7+8}{2} = 7.5 \rightarrow UQ$

Lower Quartile = $\frac{(n+1)^{th}}{4}$ term = $\frac{(9+1)^{th}}{4} = 2.5^{th}$ term.

(2^{nd} and 3^{rd})

$\frac{3+4}{2} = 3.5 \rightarrow LQ$

INTERQUARTILE RANGE = UQ − LQ

$7.5 - 3.5 = 4$

MEDIAN:

- WE CAN USE CUTTING METHOD ONLY IF RAW DATA
- IF TABLE IS GIVEN, YOU CANNOT USE THIS METHOD.

FOR THAT LET'S LEARN A NEW METHOD TO APPLY TO TABLE DATA LATER.

Using NEW METHOD TO FIND MEDIAN.

Q: 13, 8, 4, 5, 12, 8, 18, 11 $\quad n = 8$

Find Median:

Step1: Arrange 4, 5, 8, 8, 11, 12, 13, 18 (4th, 5th)

Step2: $\text{Median} = \left(\frac{n+1}{2}\right)^{th}\text{term} = \left(\frac{8+1}{2}\right)^{th}\text{term} = 4.5^{th}\text{term}$

(TOOTH)

(4^{th} term, 5^{th} term)

$$\frac{8 + 11}{2}$$

$$\text{Median} = 9.5$$

→ ON RAW DATA WE IDEALLY USE METHOD 1 (CUTTING) TO FIND MEDIAN.

→ FORMULA IS USED TO FIND MEDIAN ON TABLE.

2 TABLE DATA (SINGLE VALUE TABLE)

x/marks	f	fx = f x x
0	20	20 x 0 = 0
1	10	1 x 10 = 10
2	30	2 x 30 = 60
	n = 60	Σfx = 70
	Σf = 60	

(31st, 30th term)

Total

Be careful! Do not write 30 here

(a) MODE = (Most frequent value) = 2

(b) MEDIAN =

$$\text{Median} = \left(\frac{n+1}{2}\right)^{th}\text{term} = \left(\frac{60+1}{2}\right)^{th}\text{term} = 30.5^{th}\text{term}$$

30th term, 31st term.

$$= \frac{1 + 2}{2}$$

MEDIAN = 1.5

(c) MEAN: FOR TABLE: 1) New column = fx

2) $\text{Mean} = \frac{\Sigma fx}{\Sigma f}$ (Total / Total)

Σ = Sigma ⇒ It means to add/Total

$$\text{Mean} = \frac{\Sigma fx}{\Sigma f} = \frac{70}{60} = \frac{7}{6} = 1.1667.$$

(d) RANGE: Max − Min

2 − 0 = 2.

[3] TABLE DATA (RANGE TABLE)

Add ÷ 2

$a < x < b$ ↓ x is between a and b.

Mid value of x	x/marks	f	fx
$\frac{5+15}{2} = 10$	$5 < x \leq 15$	20	$20 \times 10 = 200$
$\frac{15+35}{2} = 25$	$15 < x \leq 35$	30	$30 \times 25 = 750$
$\frac{35+40}{2} = 37.5$	$35 < x \leq 40$	10	$10 \times 37.5 = 375$
		$n = \Sigma f = 60$	$\Sigma fx = 1325$

31st, 30th

(a) Modal class (MODE) = $15 < x \leq 35$

(b) Median class = $\left(\frac{n+1}{2}\right)^{th}$ term = $\left(\frac{60+1}{2}\right)^{th}$ term = 30.5^{th} term

30^{th} term, 31^{st} term.

Median Class	$15 < x \leq 35$

NOTE: In case of Range Table, Median always lies on the same floor. There is no need to take average.

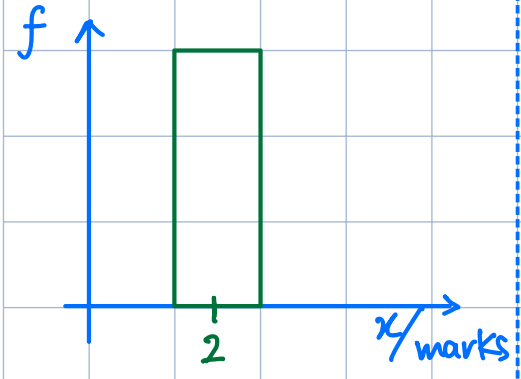
(c) MEAN: New column: ① fx ② Mid value of x

$$\text{Mean} = \frac{\Sigma fx}{\Sigma f} = \frac{1325}{60} = 22.083$$

(d) RANGE: Max − Min = 40 − 5 = 35.

HISTOGRAMS

SINGLE VALUE TABLE



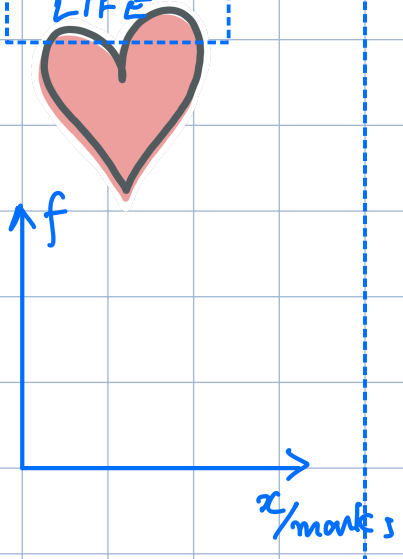
Note: The number must be in middle of bar.

RANGE TABLE

New Column = Class width = difference. $(b-a)$

SAME CLASS WIDTH

EASIEST HISTOGRAM OF YOUR LIFE



DIFFERENT CLASS WIDTH

New column = FREQUENCY DENSITY.

$$FD = \frac{\text{FREQUENCY}}{\text{CLASS WIDTH}}.$$

FREQUENCY POLYGON:

Step1: Draw a histogram.

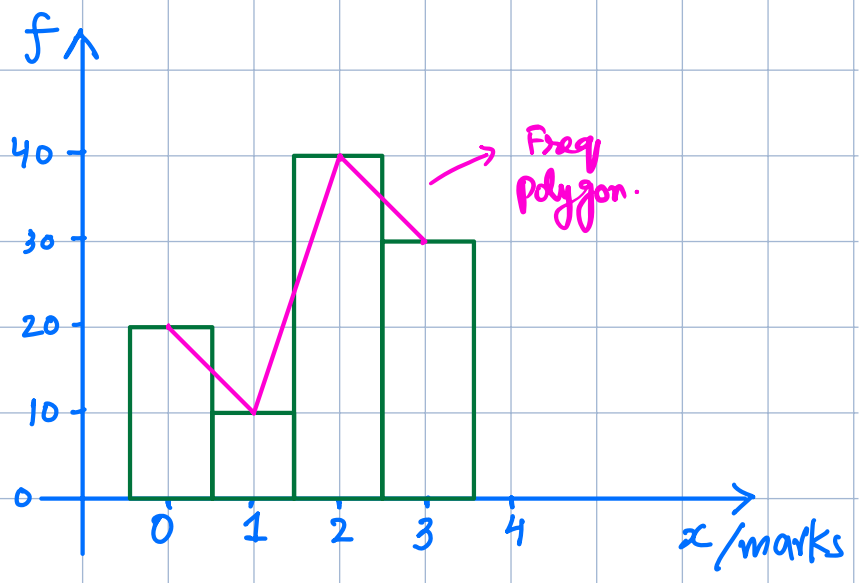
Step2: Join mid points of top of each bar with straight lines

Bar chart ⇒ There is gap between bars

Histogram ⇒ Bars are connected to each other.

1 SINGLE VALUE TABLE

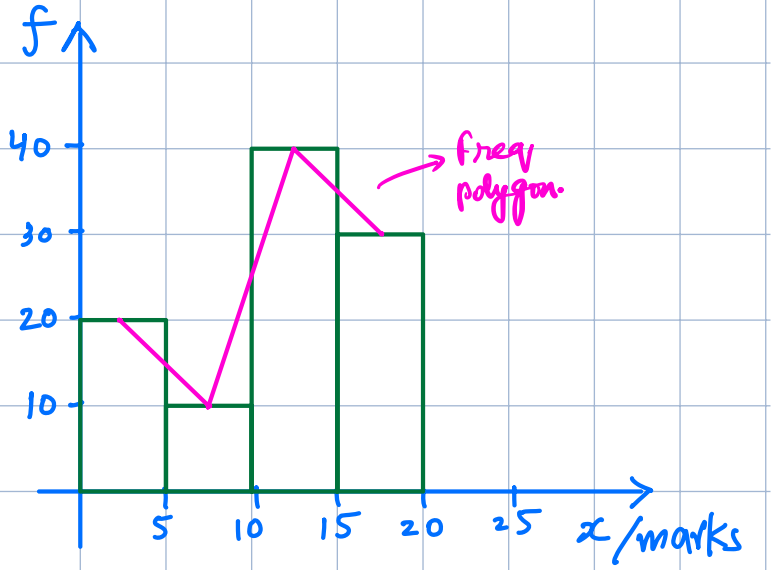
x/marks	f
0	20
1	10
2	40
3	30



Note: On a single value table when we have to draw a bar at zero, we start scale one box ahead.

2 RANGE TABLE (SAME CLASS WIDTH)

Class width (difference)	x-axis: x/marks	y-axis: f
5 − 0 = 5	$0 < x \leq 5$	20
10 − 5 = 5	$5 < x \leq 10$	10
15 − 10 = 5	$10 < x \leq 15$	40
20 − 15 = 5	$15 < x \leq 20$	30



If the class width is same, forget about this column and draw simplest histogram of your life.

Range Table → Mean → Midpoint column
Range Table → Histogram → CW column.

3 RANGE TABLE (DIFFERENT CLASS WIDTH)

x-axis y-axis

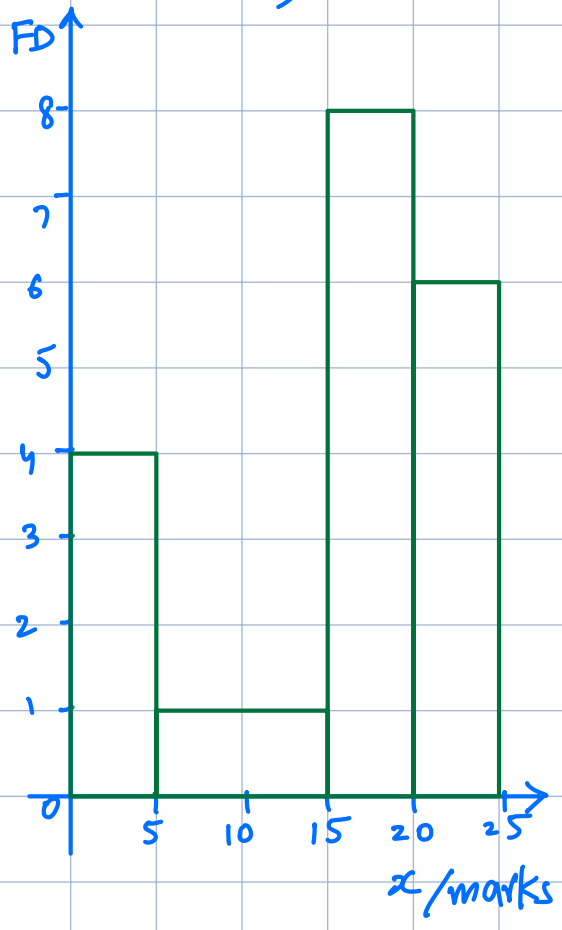
CLASS WIDTH	x/marks	f	FD = f ÷ CW
5-0 = 5	$0 < x \leq 5$	20	20 ÷ 5 = 4
15-5 = 10	$5 < x \leq 15$	10	10 ÷ 10 = 1
20-15 = 5	$15 < x \leq 20$	40	40 ÷ 5 = 8
25-20 = 5	$20 < x \leq 25$	30	30 ÷ 5 = 6

↓

Different class width

↓

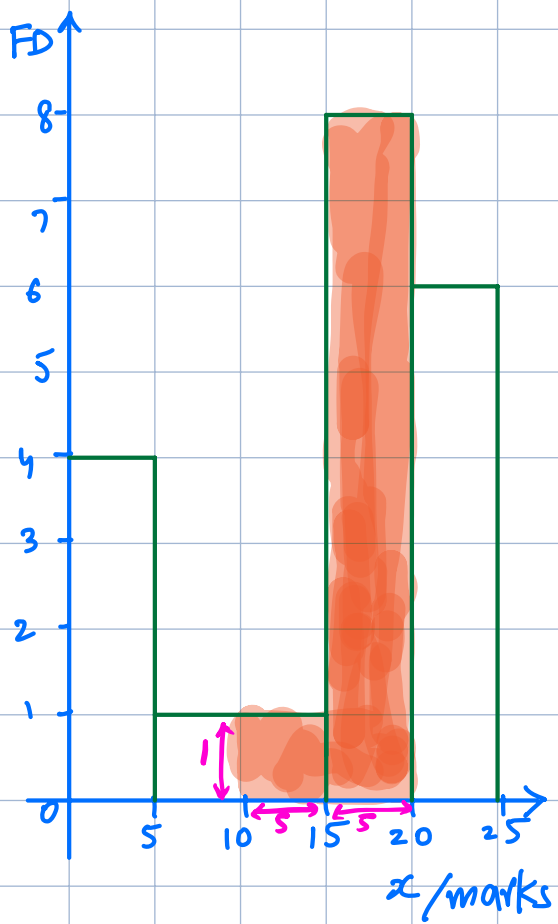
Frequency density.

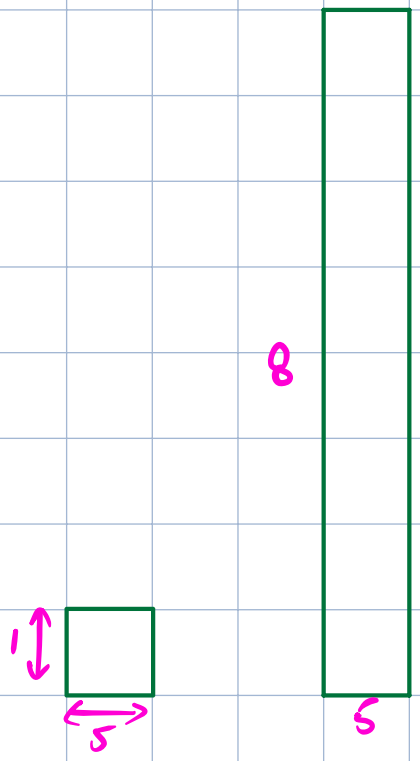


RULE: [AREA UNDER GRAPH] OF [FREQ. DENSITY GRAPH] = [NUMBER OF THINGS (FREQUENCY)]

no. of things

Q. Find number of students who scored between 10 and 20 marks





$(1 \times 5) + (8 \times 5)$

$5 + 40$

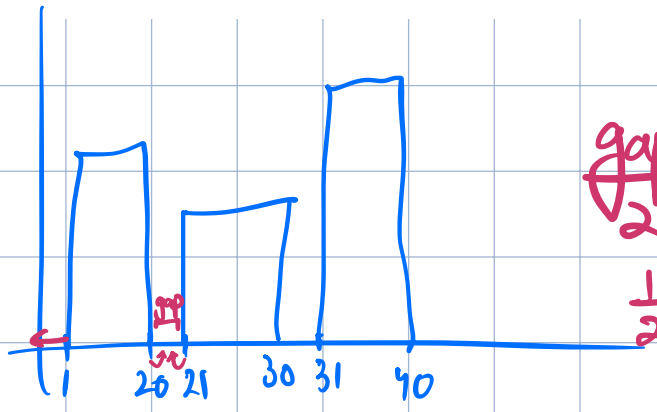
= **45 students**.

29 The following table gives the marks, out of 75, in a pure mathematics examination taken by 234 students.

Marks	1–20	21–30	31–40	41–50	51–60	61–75
Frequency	40	34	56	54	29	21

(i) Draw a histogram on graph paper to represent these results. [5]

(ii) Calculate estimates of the mean mark and the standard deviation. [4]

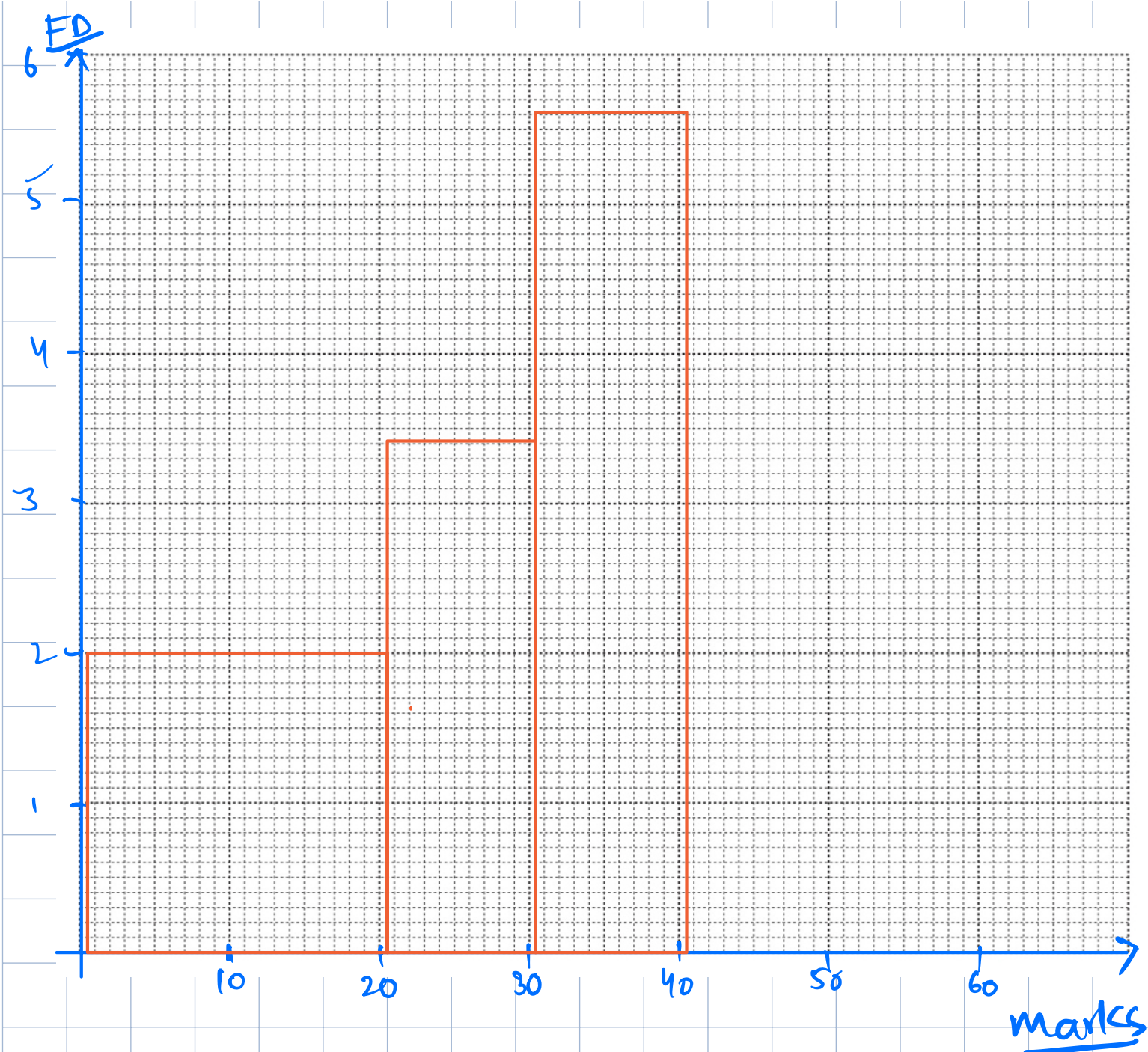


gap = Adjustment.

$\frac{1}{2} = 0.5$

Marks	1–20	21–30	31–40	41–50	51–60	61–75
	-0.5 +0.5	-0.5 +0.5	-0.5 +0.5	-0.5 +0.5	-0.5 +0.5	-0.5 +0.5
Frequency	40	34	56	54	29	21

CW	20	10	10	10	10	15
Marks	0.5–20.5	20.5–30.5	30.5–40.5	40.5–50.5	50.5–60.5	60.5–75.5
Freq	40	34	56	54	29	21
FD (f÷cw)	2	3.4	5.6	5.4	2.9	1.4



MEAN

Midvalue	10·5	25·5	35·5	45·5	55·5	68	
Marks	1–20	21–30	31–40	41–50	51–60	61–75	
Frequency	40	34	56	54	29	21	$=\Sigma f=234$
fx	420	867	1988	2457	1695	1428	$=\Sigma fx=8769.5$

$$\text{Mean} = \frac{8769.5}{234} = \square$$

NOTE: IF INTERVALS ARE DISCONNECTED

1 — 20 21 — 30 31 — 40

For Histograms, first fix the gap. Then find class width. Then proceed to make histogram.

FOR MEAN, SD, VAR FIND THE MIDVALUE OF X DIRECTLY. NO NEED TO FIX THE GAP

CUMULATIVE FREQUENCY (12 MARKS).

(TOTAL)

One day, garage A records the amount of petrol bought by the first 120 customers.
The results are summarised in the table below.

Petrol (k litres)	$0 < k \leqslant 10$	$10 < k \leqslant 20$	$20 < k \leqslant 30$	$30 < k \leqslant 40$	$40 < k \leqslant 50$	$50 < k \leqslant 60$	$60 < k \leqslant 70$	$70 < k \leqslant 80$
Number of customers (f)	9	13	36	30	16	9	5	2

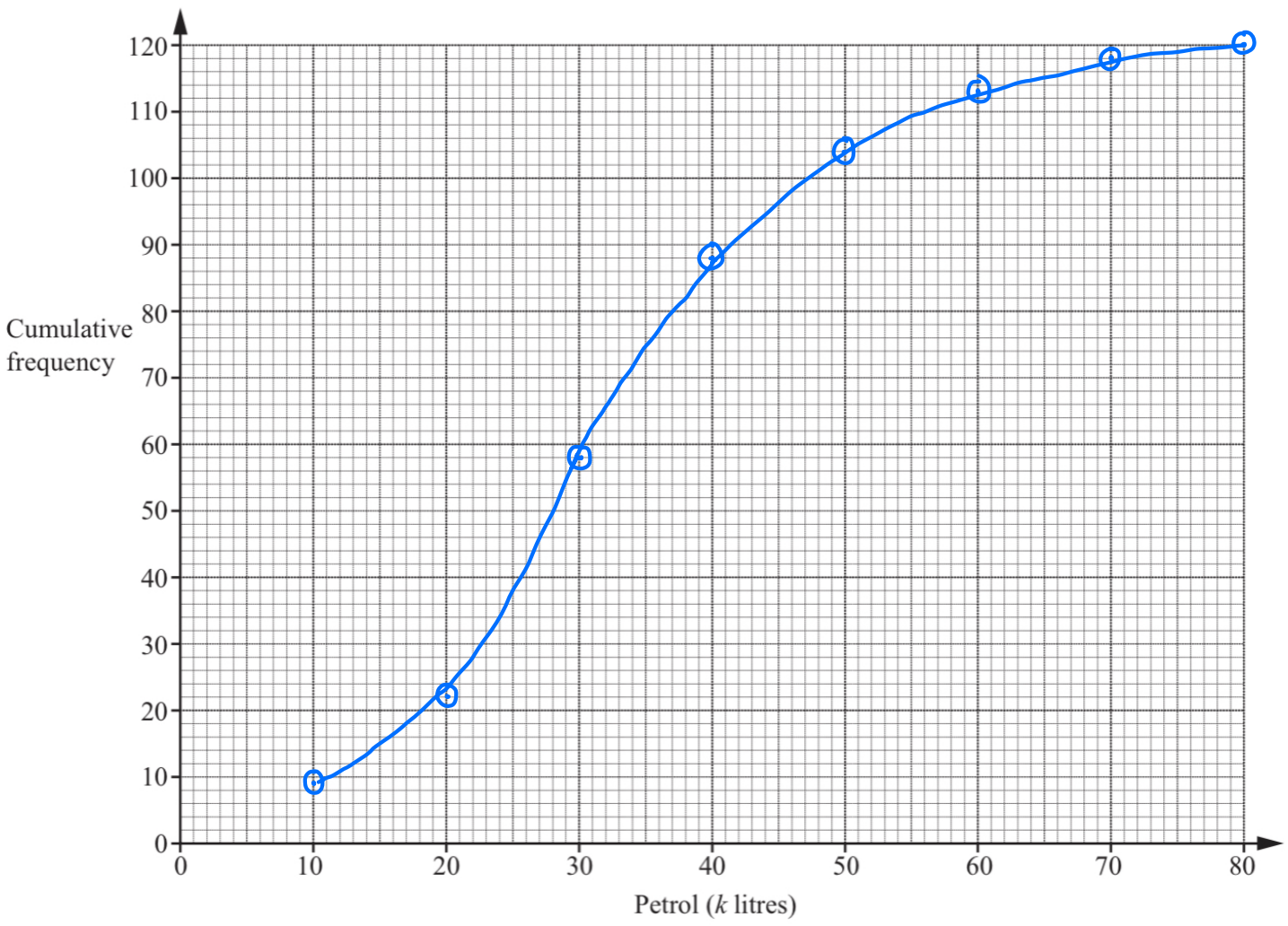
(a) Complete the cumulative frequency table below.

Petrol (k litres) (x)	$k \leqslant 10$	$k \leqslant 20$	$k \leqslant 30$	$k \leqslant 40$	$k \leqslant 50$	$k \leqslant 60$	$k \leqslant 70$	$k \leqslant 80$
Cumulative frequency (TOTAL) (y)	9	22	58	88	104	113	118	120

(10,9) (20,22) (30,58) (40,88) (50,104) (60,113) (70,118) (80,120)

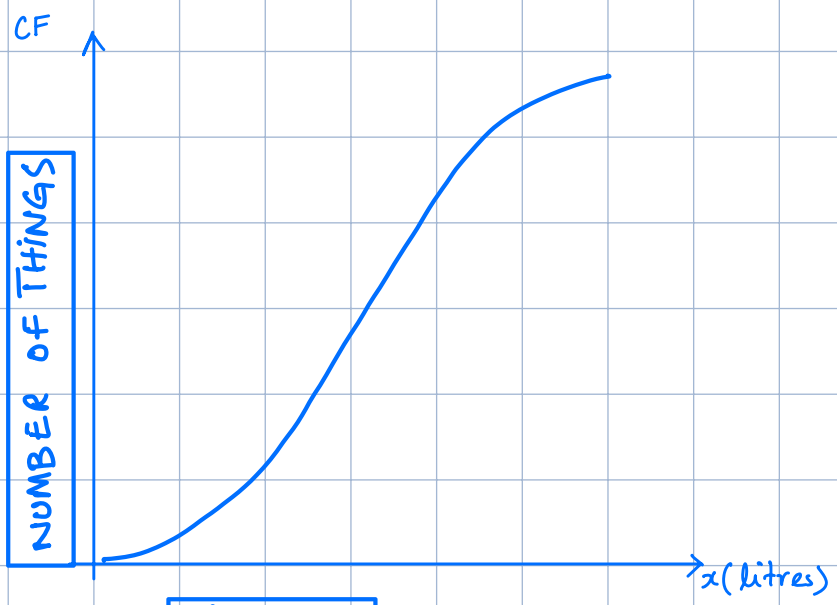
[1]

(b) On the grid below, draw a cumulative frequency curve to represent this data.



[3]

In A Levels S1, you are allowed to join these points with straight lines instead of curve.



MEDIAN

UPPER QUARTILE

LOWER QUARTILE

INTER QUARTILE RANGE

PERCENTILE

only for cumulative frequency

Median $= \frac{n}{2}$

everywhere else,

Median $= \left(\frac{n+1}{2}\right)^{th}$ term.

1) MEDIAN $= \frac{n}{2} = \frac{120}{2} =$ ~~60~~ check - x-axis for Median (M)

$M = 31$

2) UPPER QUARTILE $= \frac{3n}{4} = \frac{3(120)}{4} =$ ~~90~~ check x-axis for UQ.

$UQ = 41$

3) LOWER QUARTILE $= \frac{n}{4} = \frac{120}{4} =$ ~~30~~ check x-axis for LQ

$LQ = 23$

4) INTER QUARTILE RANGE $IQR = UQ - LQ$

$41 - 23 = 18$

Note: Always subtract x-axis values for UQ and LQ.

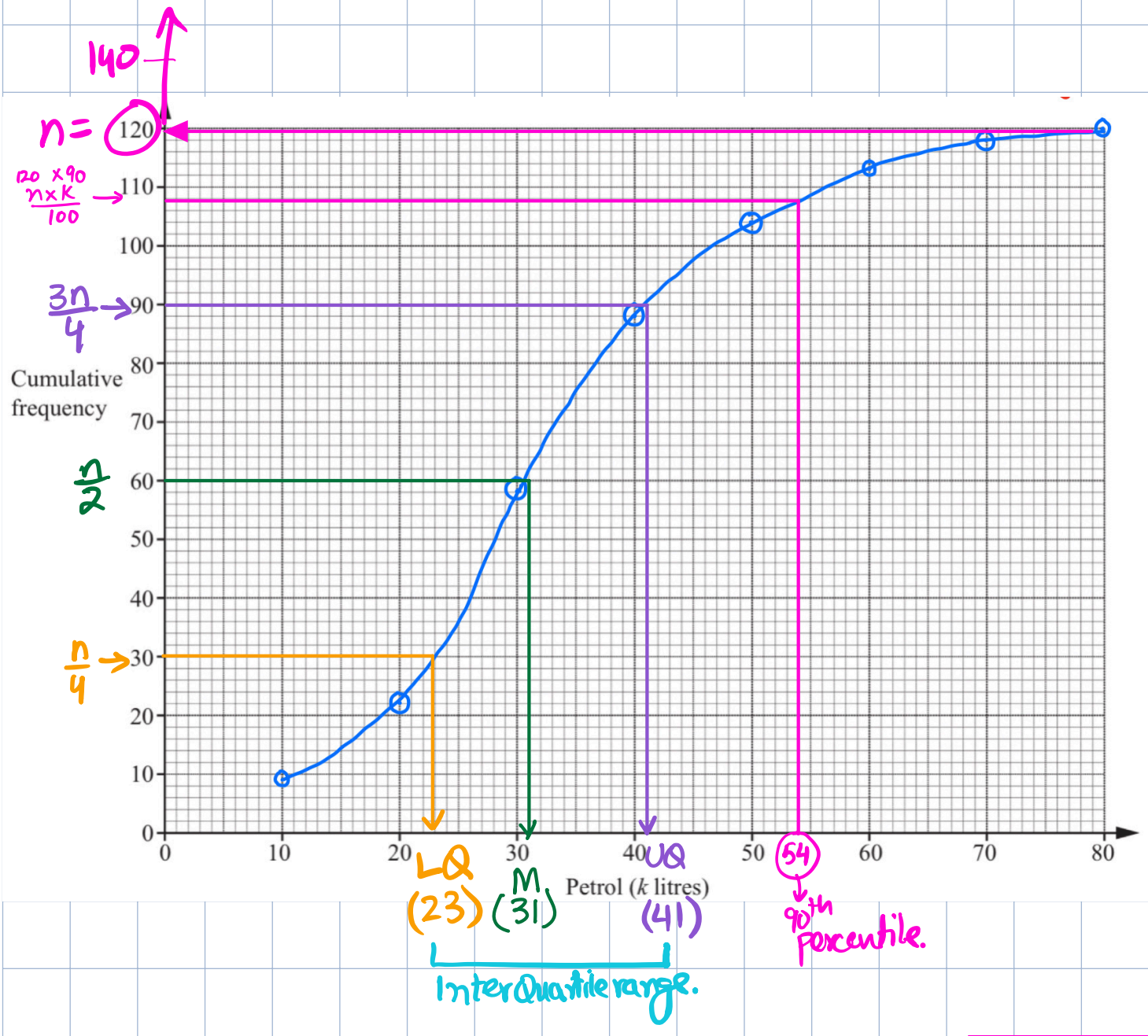
5) **PERCENTILE** (For k^{th} Percentile) = $n \times \frac{k}{100}$

A^* = 90^{th} percentile.

$120 \times \frac{90}{100} = 108$

check - x-axis

90^{th} percentile = 54.

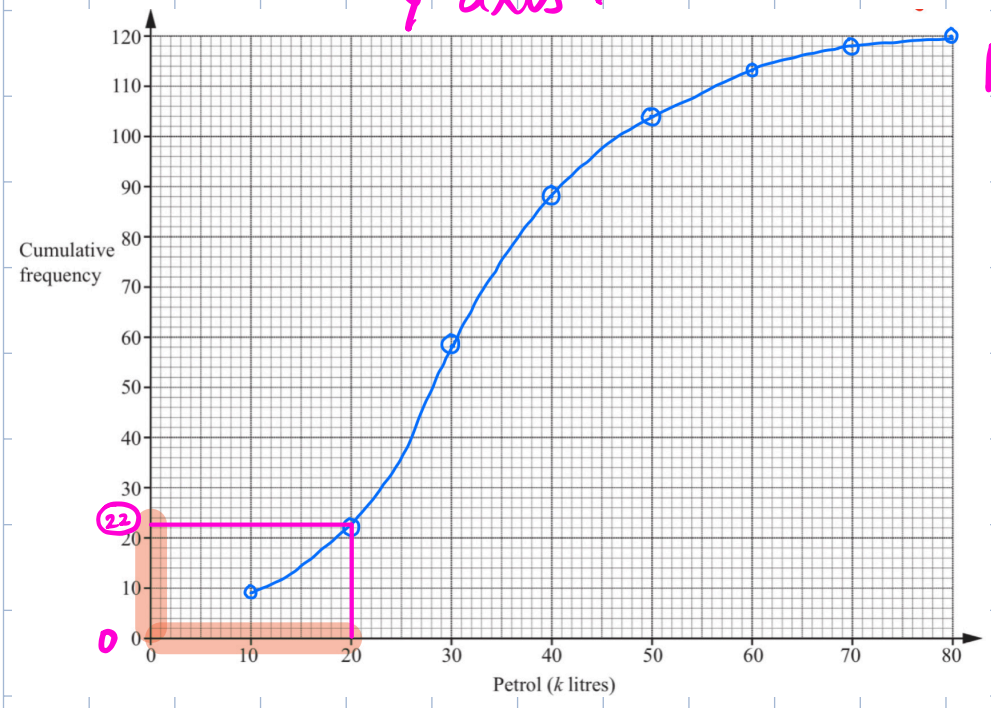


Y-AXIS : NUMBER OF THINGS.

No. of things
- FD graph → Area under graph.
- CF graph → y-axis.

(i) Find number of cars who bought less than 20 litres.

y-axis.



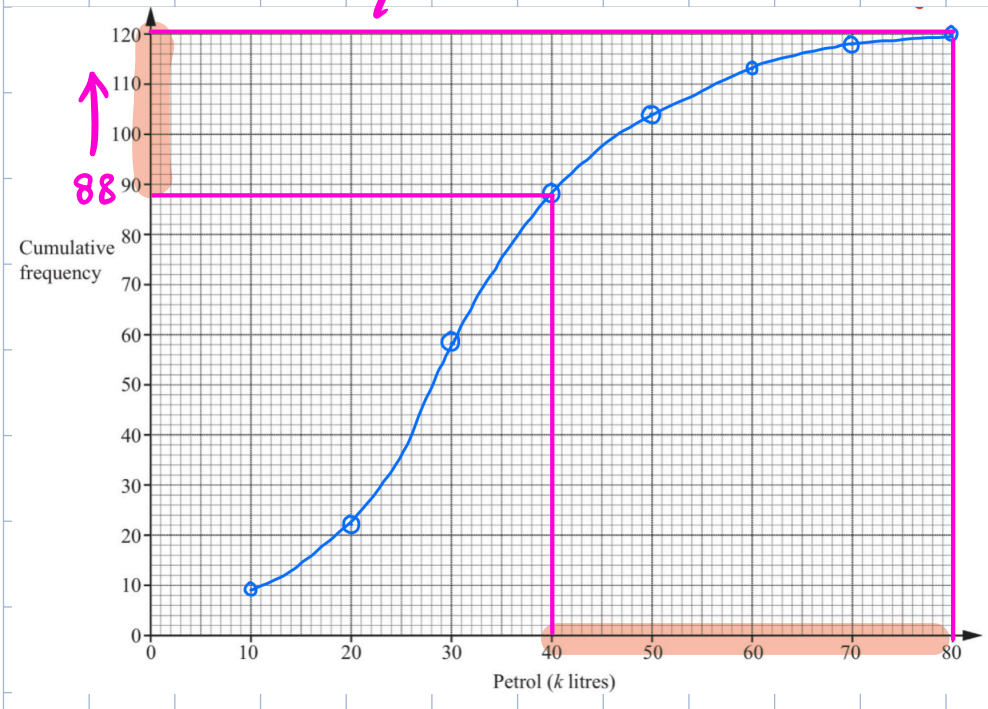
between 0 and 20 litres

22 − 0

22 cars

(ii) Find number of cars who bought more than 40 litres.

y-axis.

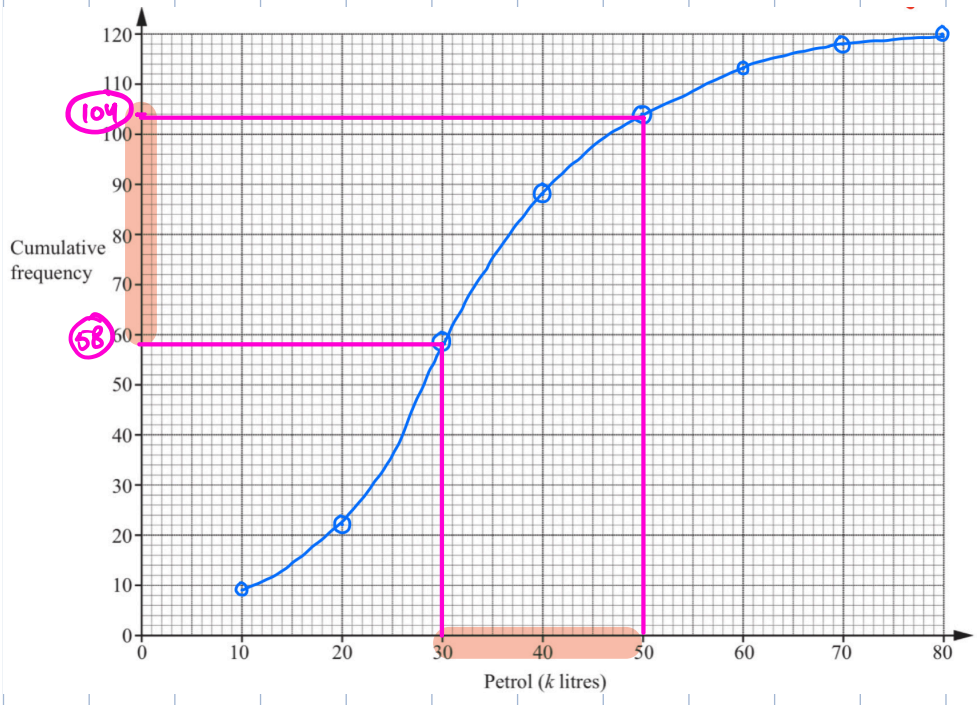


between 40 and 80 litres

120 − 88

= **32 cars**

(iii) Find number of cars who bought between 30 and 50 litres.



$104 - 58$

$= \boxed{66 \text{ cars.}}$

at least 8A's ⟶ more than or equal to 8A's

at most 2A's ⟶ less than or equal to 2A's

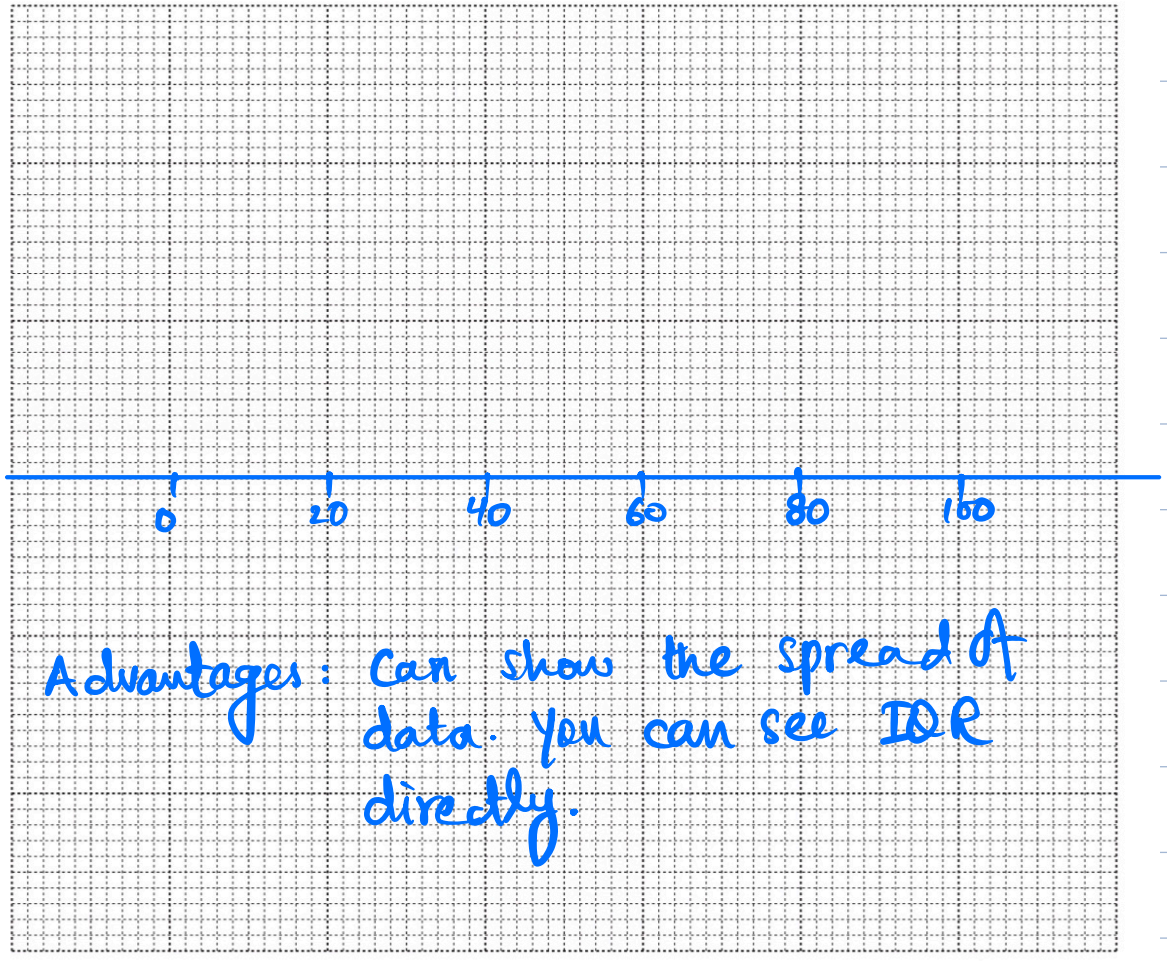
24
− 8 sleep
16
− 6 school
10
− 4 Academy
6

2 : 4

2 Study, 4 Ent.

Box and whisker diagram.

Min = 0, Max = 80, LQ = 23, M = 31, UQ = 41



Advantages: Can show the spread of data. You can see IQR directly.

STEM AND LEAF DIAGRAM

Purpose: Sometimes data has lot of repeats

eg

There is a chance of mistake.

✓ Leaves are always written in ascending order.

Preference of Examiners

1) Stems should be integers.

2) Leaves should be single digit numbers.

36 The weights in kilograms of 11 bags of sugar and 7 bags of flou are as follows.

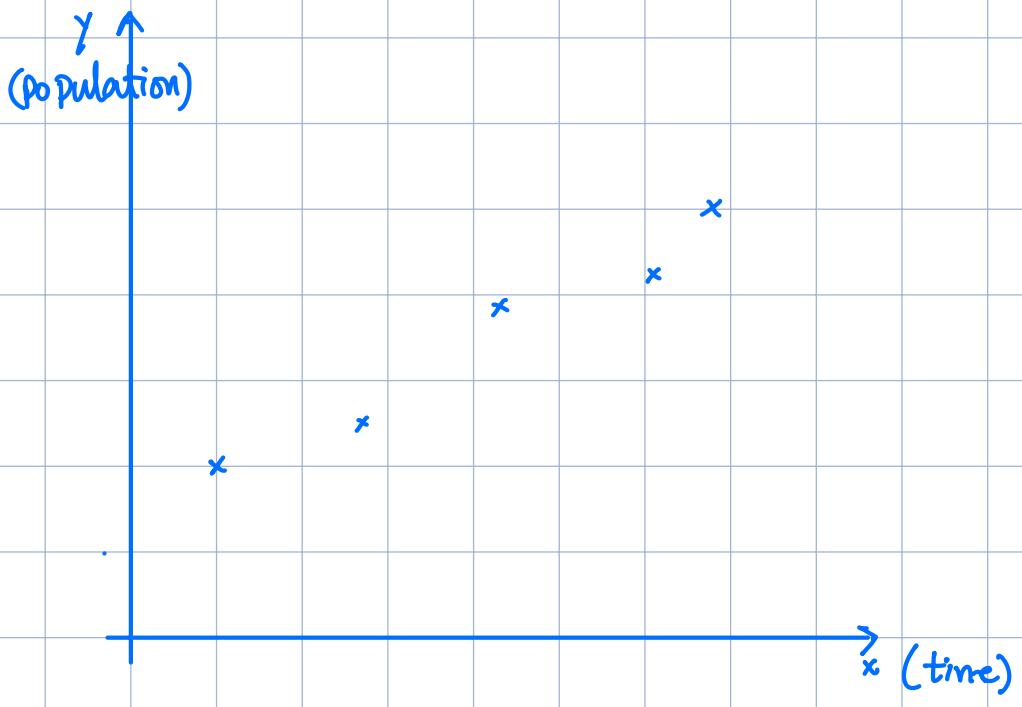
Sugar:	1.961	1.983	2.008	2.014	1.968	1.994	2.011	2.017	1.977	1.984	1.989
Flour:	1.945	1.962	1.949	1.977	1.964	1.941	1.953				

(i) Represent this information on a back-to-back stem-and-leaf diagram with sugar on the left-hand side. [4]

(ii) Find the median and interquartile range of the weights of the bags of sugar. [3]

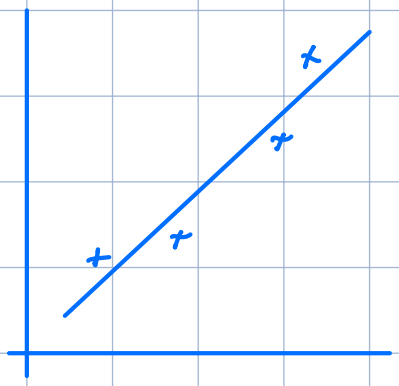
SCATTER PLOTS

(LINE OF BEST FIT) (CORRELATION).

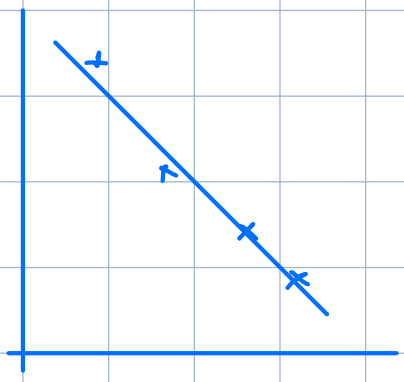


LINE OF BEST FIT:
(TREND LINE)

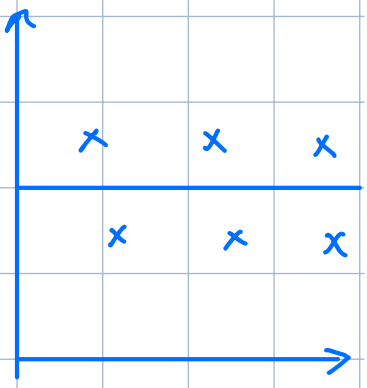
DO's	DONT'S
1) Average line through the points.	1) NEVER FORCE THE LINE THROUGH ORIGIN (0,0)
2) Almost equal balance of points on each side of line.	2) NEVER FORCE THE LINE THROUGH FIRST AND LAST POINT.
3) Line of best fit must represent TREND of points.	3) NEVER FORCE LINE TO TOUCH MAXIMUM POINTS.



POSITIVE CORRELATION



NEGATIVE CORRELATION

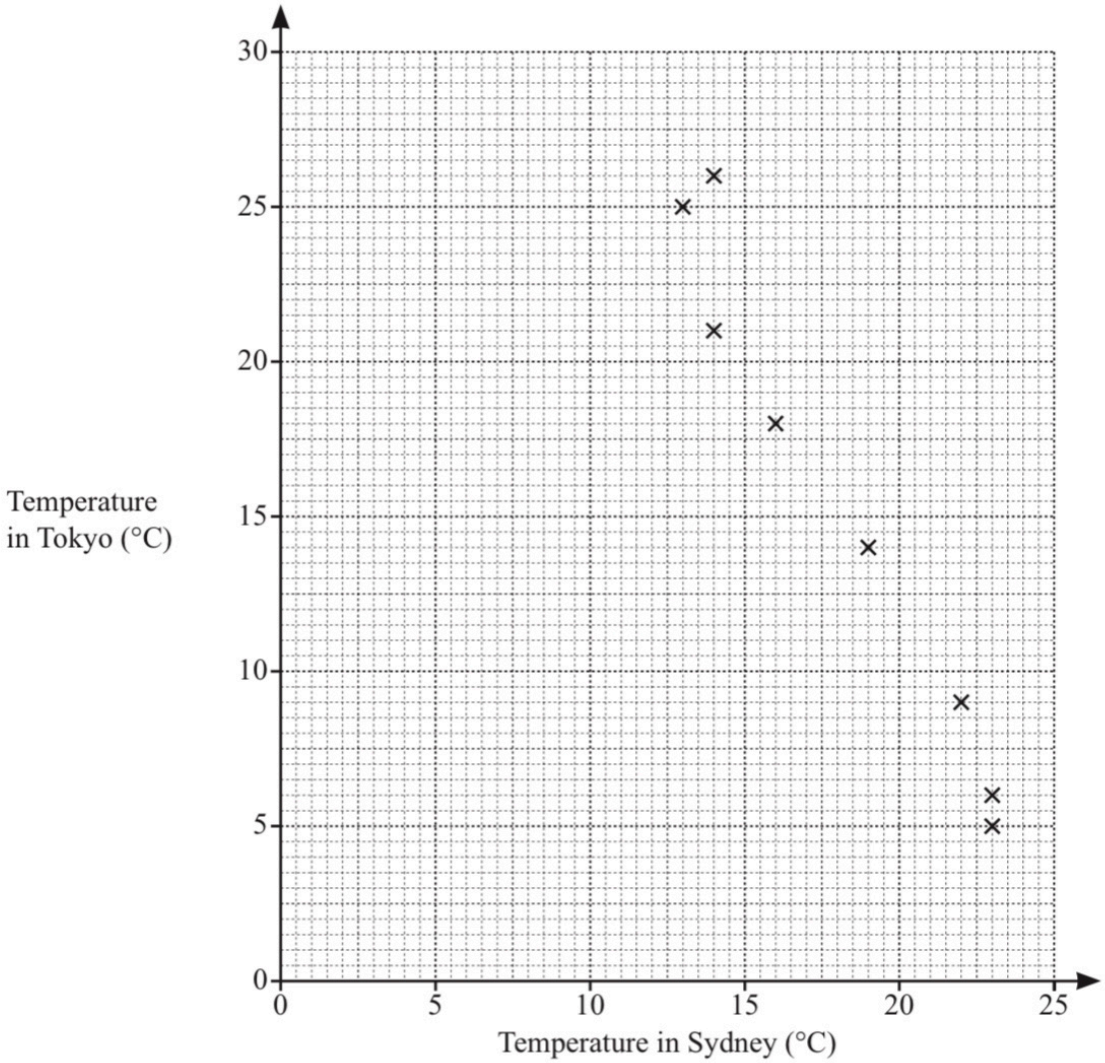


NO (Zero) CORRELATION

2 The table shows the average monthly temperatures (°C) in Tokyo and in Sydney one year.

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Temperature in Sydney (°C)	23	23	22	19	16	14	13	14	16	18	20	21
Temperature in Tokyo (°C)	5	6	9	14	18	21	25	26	23	18	12	8

(a) Complete the scatter diagram.
The first eight points have been plotted for you. [2]



(b) What type of correlation is shown by the scatter diagram?

..... [1]

(c) Draw a line of best fit. [1]

(d) The following year, the average temperature in Sydney during May was 15 °C.

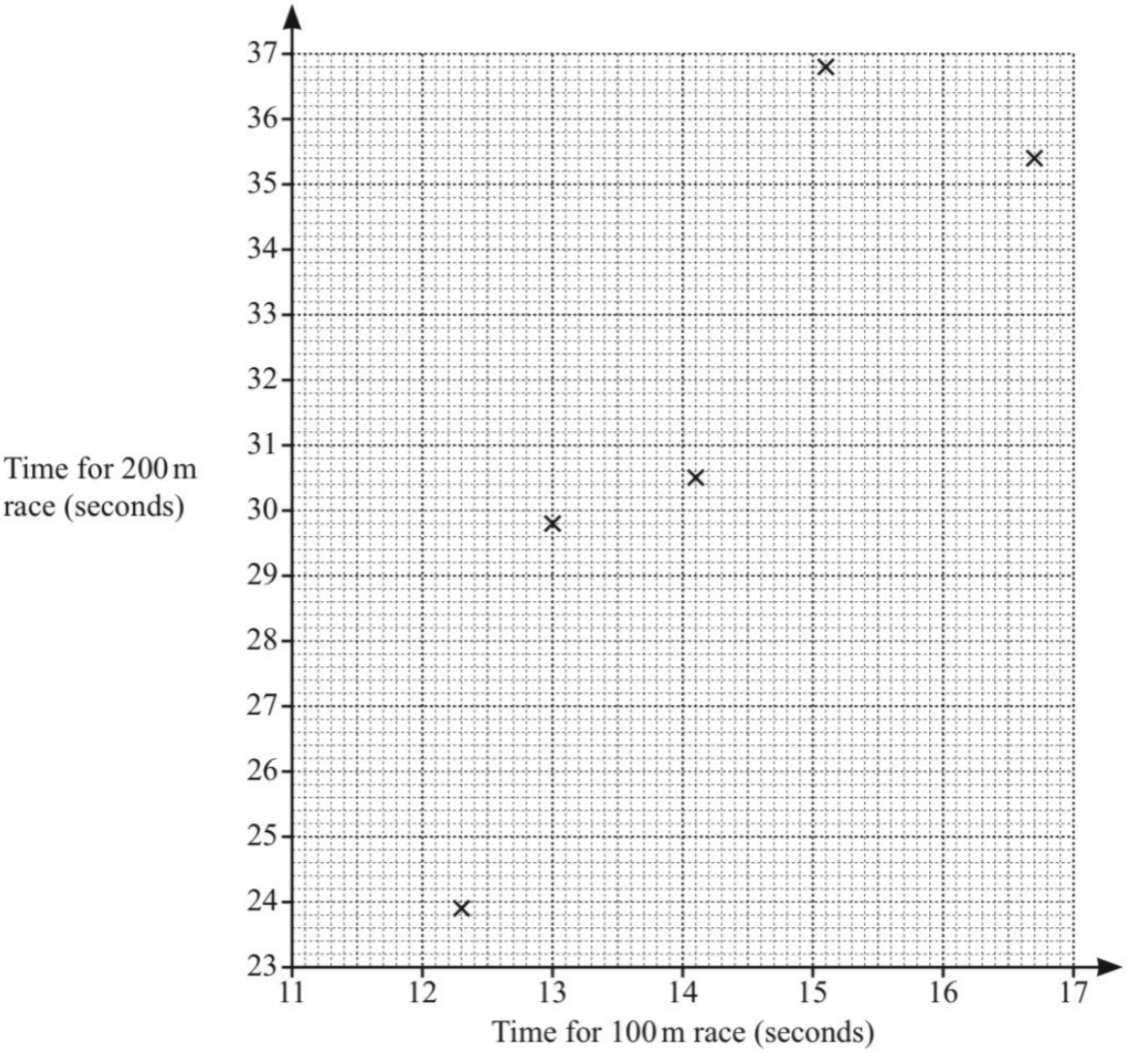
By using your line of best fit, estimate the average temperature in Tokyo that May.

..... °C [1]

2 Ten boys ran in a 100 m race and a 200 m race.
The table below shows their times in seconds.

Time for 100 m race	12.3	14.1	15.1	16.7	13.0	14.7	13.7	12.9	15.2	16.1
Time for 200 m race	23.9	30.5	36.8	35.4	29.8	32.5	28.4	26.1	33.5	36.0

(a) Complete the scatter diagram.
The first five points have been plotted for you.



[2]

(b) What type of correlation is shown in the scatter diagram?

..... [1]

(c) Draw a line of best fit. [1]

(d) Another boy recorded a time of 27.5 s in the 200 m race.

Use your graph to estimate the time it would take him to run 100 m.

..... s [1]

ROOM FOR ERROR FOR PLOTS.

HALF SMALL (TINY) BOX.