

Concepts to Cover

1. Box and whisker plot
2. Stem and Leaf Diagram
3. Histogram
4. Cumulative Frequency Curve

Thursday

1. Box and whisker plot

Data Required

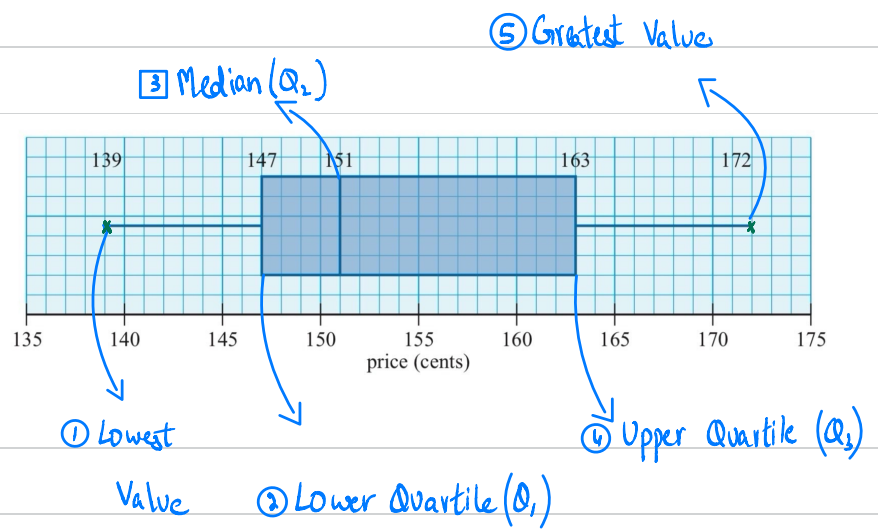
1. Least Value

2. Lower Quartile Q_1

3. Median Q_2

4. Upper Quartile Q_3

5. Greatest value



Pro: A box and whisker plot is a useful way of summarising data and showing the median, upper and lower quartile. (Spread of Data)

Con: Individual or all values cannot be seen from the box and whisker plot

Note: Box and whisker makes it easy to see the spread of the data. The lesser the interquartile range, the more reliable and consistent the values are. (Results are close to one another)

Skewness of Data

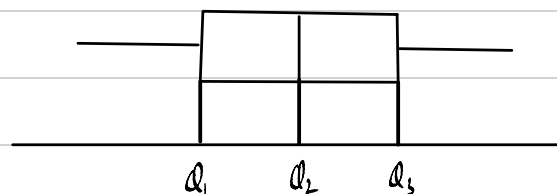
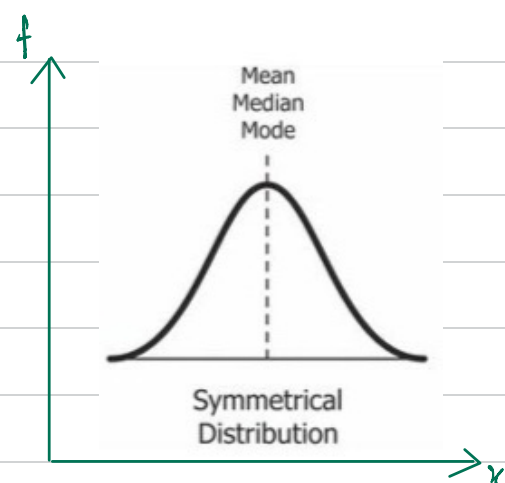
Any data can be three of the following

1. Symmetrical

$$\text{Mean} = \text{Median} = \text{Mode}$$

$$Q_3 - Q_2 = Q_2 - Q_1$$

Heights, Weight, I.Q

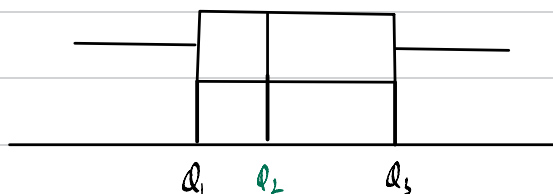
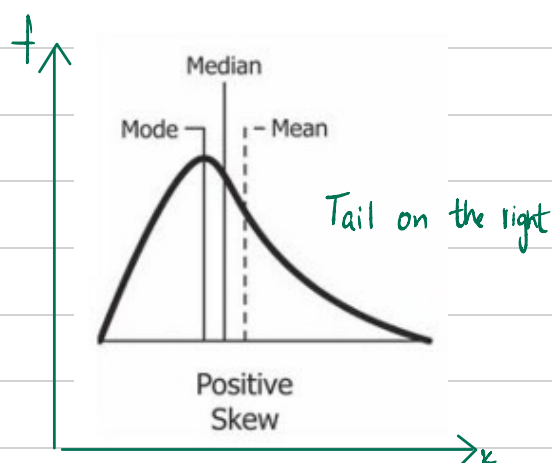


2. Positively Skewed

$$Q_3 - Q_2 > Q_2 - Q_1$$

Exam marks of a difficult test, wealth distribution

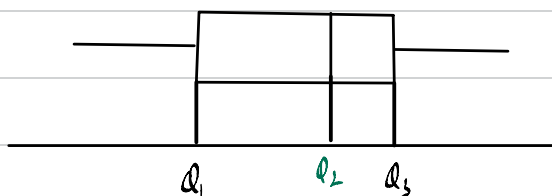
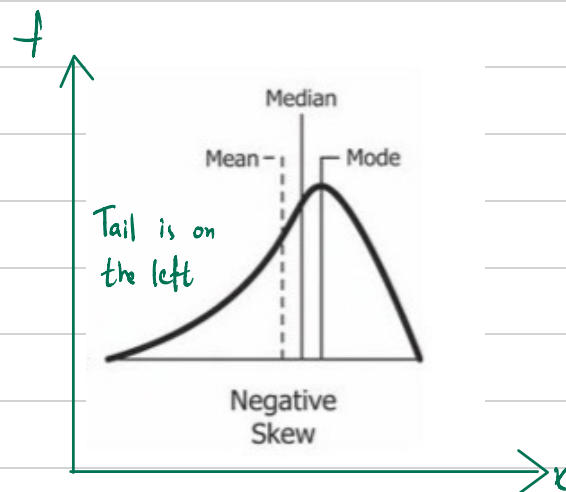
No of pets people own



3. Negatively skewed

Age of retirement, marks of an easy test

$$Q_3 - Q_2 < Q_2 - Q_1$$



2. Stem and leaf diagram

Single

The masses of 20 soccer players were measured to the nearest kilogram and this stem and leaf diagram was produced.

Stem	Leaf	Key
4	6	4 6 = 46 kilograms
5	4 0 0	
5	7 8 9 5	
6	3 0 1 1 3 2	
6	6 8 6 9	
7	4 0	

Back to back

This stem and leaf plot compares the battery life of two different brands of mobile phone.

Brand X Leaf	Stem	Brand Y Leaf	Key
9 4 8 7 2	0	5 8	Brand X 8 2 = 28 hours Brand Y 4 2 = 42 hours
7 8 7 2 3	1	4 7 8 2	
8 4 6 2 7 9 8	2	8 9 7 1 5	
7 2	3	7 2 1 0	
	4	2	
	5	1	

You read the data for Brand X from right to left. The stem is still the tens digit.

A stem and leaf diagram is useful when the data is discrete

Pro: Useful to compare two sets of data

Con: Since we list out all the data, they are unsuitable for large data set

- 3 The back-to-back stem-and-leaf diagram shows the diameters, in cm, of 19 cylindrical pipes produced by each of two companies, A and B.

Company A							Company B					
					4	33	1	2	8			
9	8	3	2	0		34	1	6	8	9	9	
8	7	5	4	1	1	35	1	2	2	3		
		9	6	5	2	36	5	6				
			4	3	1	37	0	3	4			
						38	2	8				

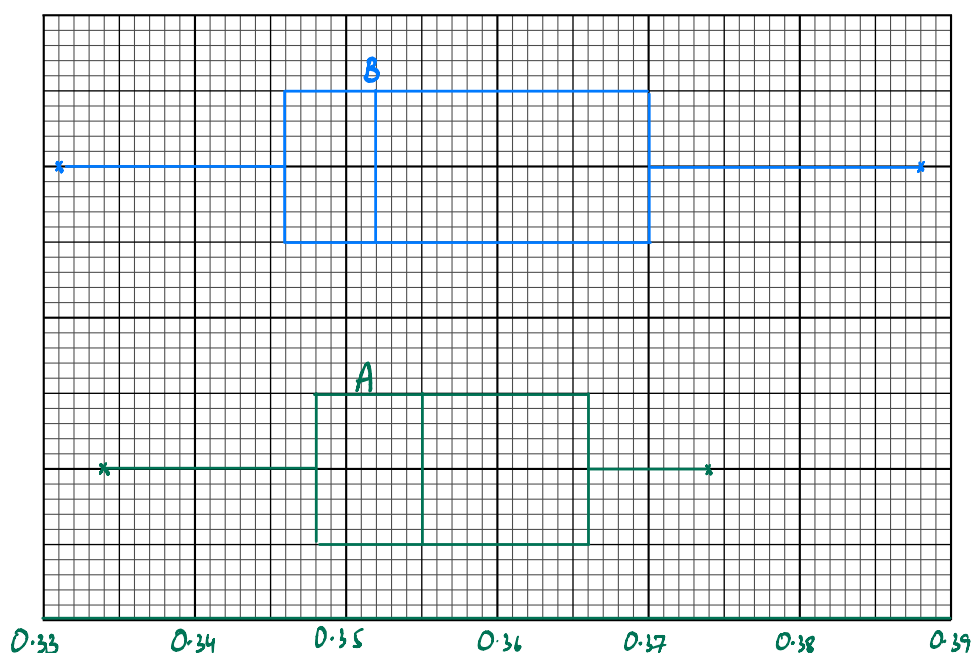
Key: 1 | 35 | 3 means the pipe diameter from company A is 0.351 cm and from company B is 0.353 cm.

- (a) Find the median and interquartile range of the pipes produced by company A. [3]

$$\begin{aligned} \text{Median} &= \frac{19+1}{2} = 10^{\text{th}} \text{ term} \rightarrow 0.355 \\ \text{L.Q.} &= \frac{9+1}{2} = 5^{\text{th}} \text{ term} \rightarrow 0.348 \\ \text{U.Q.} &= \frac{9+1}{2} = 5^{\text{th}} \text{ term} \rightarrow 0.366 \\ \text{I.Q.R.} &= 0.366 - 0.348 = 0.018 \end{aligned}$$

It is given that for the pipes produced by company B the lower quartile, median and upper quartile are 0.346 cm, 0.352 cm and 0.370 cm respectively.

- (b) Draw box-and-whisker plots for companies A and B on the grid below. [3]



- (c) Make one comparison between the diameters of the pipes produced by companies A and B. [1]

- ① A is more consistent than company B. ② B is more positively skewed than A.
③ B has more extreme values.

- 4 The following back-to-back stem-and-leaf diagram shows the cholesterol count for a group of 45 people who exercise daily and for another group of 63 who do not exercise. The figures in brackets show the number of people corresponding to each set of leaves.

People who exercise		People who do not exercise	
(9)	9 8 7 6 4 3 2 2 1	3	1 5 7 7 (4)
(12)	9 8 8 8 7 6 6 5 3 3 2 2	4	2 3 4 4 5 8 (6)
(9)	8 7 7 7 6 5 3 3 1	5	1 2 2 2 3 4 4 5 6 7 8 8 9 (13)
(7)	6 6 6 6 4 3 2	6	1 2 3 3 3 4 5 5 5 7 7 8 9 9 (14)
(3)	8 4 1	7	2 4 5 5 6 6 7 8 8 (9)
(4)	9 5 5 2	8	1 3 3 4 6 7 9 9 9 (9)
(1)	4	9	1 4 5 5 8 (5)
(0)		10	3 3 6 (3)

Key: 2 | 8 | 1 represents a cholesterol count of 8.2 in the group who exercise and 8.1 in the group who do not exercise.

- (i) Give one useful feature of a stem-and-leaf diagram. [1]
- (ii) Find the median and the quartiles of the cholesterol count for the group who do not exercise. [3]

You are given that the lower quartile, median and upper quartile of the cholesterol count for the group who exercise are 4.25, 5.3 and 6.6 respectively.

- (iii) On a single diagram on graph paper, draw two box-and-whisker plots to illustrate the data. [4]

i Useful to compare two data sets.

ii Median = $\frac{63+1}{2} = 32^{\text{nd}} \text{ term} \longrightarrow 6.5$

Lower Quartile = $\frac{31+1}{2} = 16^{\text{th}} \text{ term} \longrightarrow 5.4$

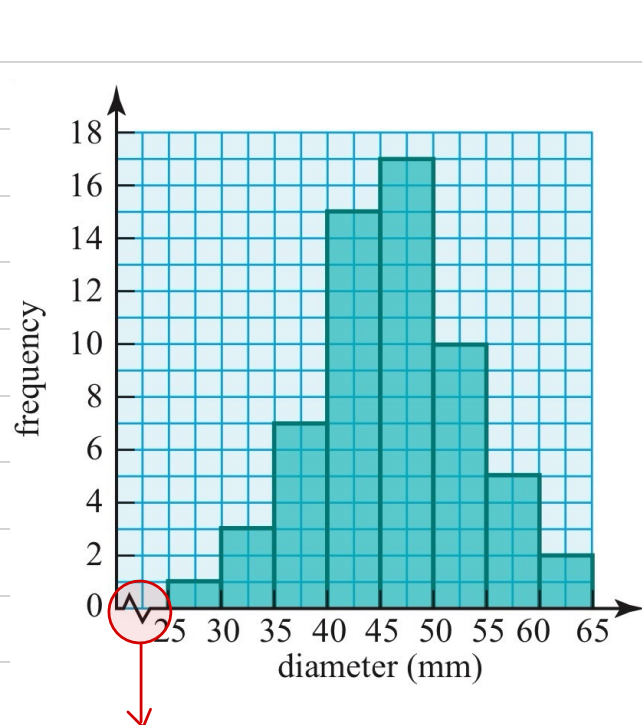
Upper Quartile = $16^{\text{th}} \text{ term} \longrightarrow 8.3$

Histogram

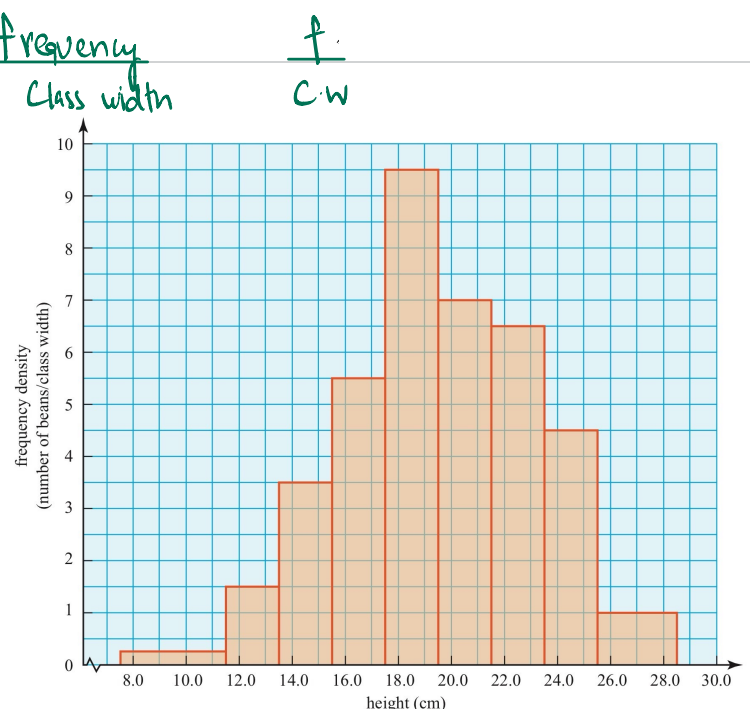
22-01-26

Histograms are used to illustrate continuous data. The columns in a histogram may have different widths and the area of each column is proportional to the frequency. Unlike bar charts, there are no gaps between the columns because where one class ends, the next begins.

1. If the data is not continuous, it is converted into continuous data, by taking the difference in consecutive class intervals and then dividing by two. The result is then added in the value where the preceding class ends and subtracted from where the proceeding class starts
2. If class widths are equal, then frequency is taken on the vertical axis $\rightarrow$ Rarely happens
3. If class widths are unequal then frequency density is taking on the vertical axis $\rightarrow$ Common



Jumper: To skip values b/w 0 & 25



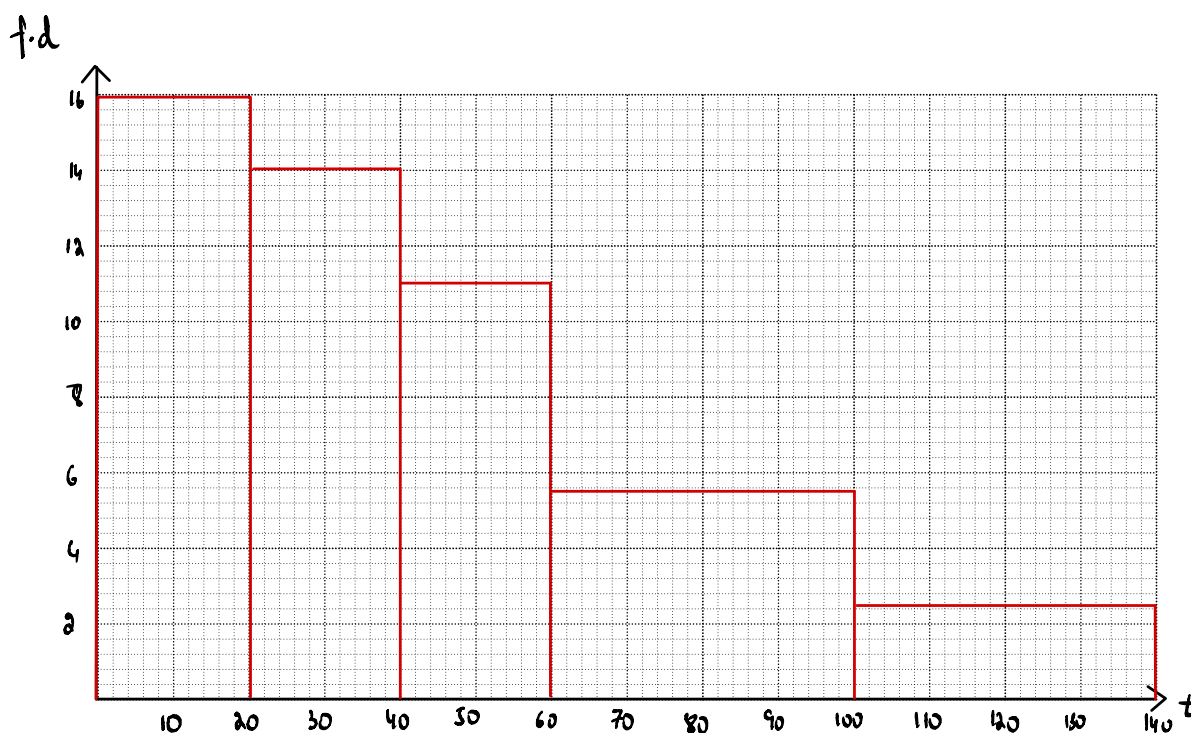
- 4 The times taken, t seconds, by 1140 people to solve a puzzle are summarised in the table.

Time (t seconds)	$0 \leq t < 20$	$20 \leq t < 40$	$40 \leq t < 60$	$60 \leq t < 100$	$100 \leq t < 140$
Number of people	320	280	220	220	100

- (i) On the grid, draw a histogram to illustrate this information.

[4]

$$f.d = \frac{f}{c.w} \quad \frac{320}{20} = 16 \quad \frac{280}{20} = 14 \quad \frac{220}{20} = 11 \quad \frac{220}{40} = 5.5 \quad \frac{100}{40} = 2.5$$



- (ii) Calculate an estimate of the mean of t .

[2]

Time (t seconds)	$0 \leq t < 20$	$20 \leq t < 40$	$40 \leq t < 60$	$60 \leq t < 100$	$100 \leq t < 140$
Number of people	320	280	220	220	100

10 20 50 80 120

$$\text{Mean} = \frac{320 \times 10 + 280 \times 30 + 220 \times 50 + 220 \times 80 + 100 \times 120}{1140}$$

1140

$$\text{Mean} = 45.8$$

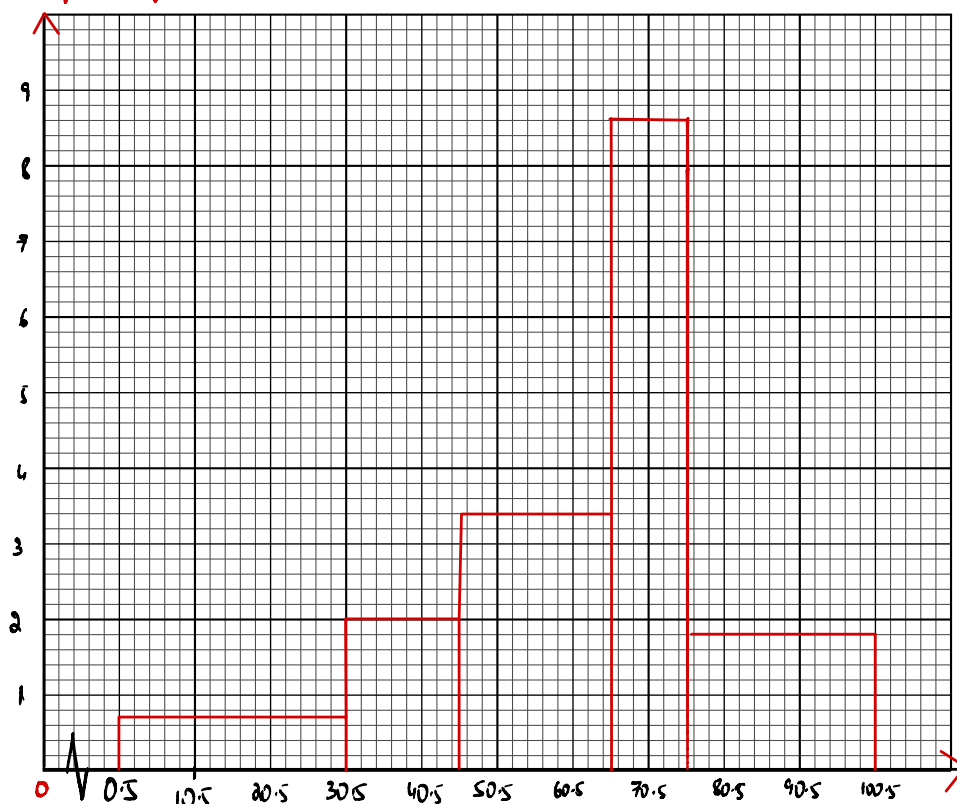
- 3 At a summer camp an arithmetic test is taken by 250 children. The times taken, to the nearest minute, to complete the test were recorded. The results are summarised in the table.

Time taken, in minutes	1 – 30	31 – 45	46 – 65	66 – 75	76 – 100
Frequency	21	30	68	86	45

- (a) Draw a histogram to represent this information.

[4]

Time taken	0.5 – 30.5	30.5 – 45.5	45.5 – 65.5	65.5 – 75.5	75.5 – 100.5
Frequency Density	$\frac{21}{30} = 0.7$	$\frac{30}{15} = 2$	$\frac{68}{20} = 3.4$	$\frac{86}{10} = 8.6$	$\frac{45}{25} = 1.8$



- (b) State which class interval contains the median.

[1]

$$\frac{50 \times 250}{100} = 125$$

66 – 75

3(b)

66 – 75

65.5 – 75.5 ✓

B1

Condone 65.5 – 75.5

- (c) Given that an estimate of the mean time is 61.05 minutes, state what feature of the distribution accounts for the median and the mean being different.

[1]

Data is not symmetrical.

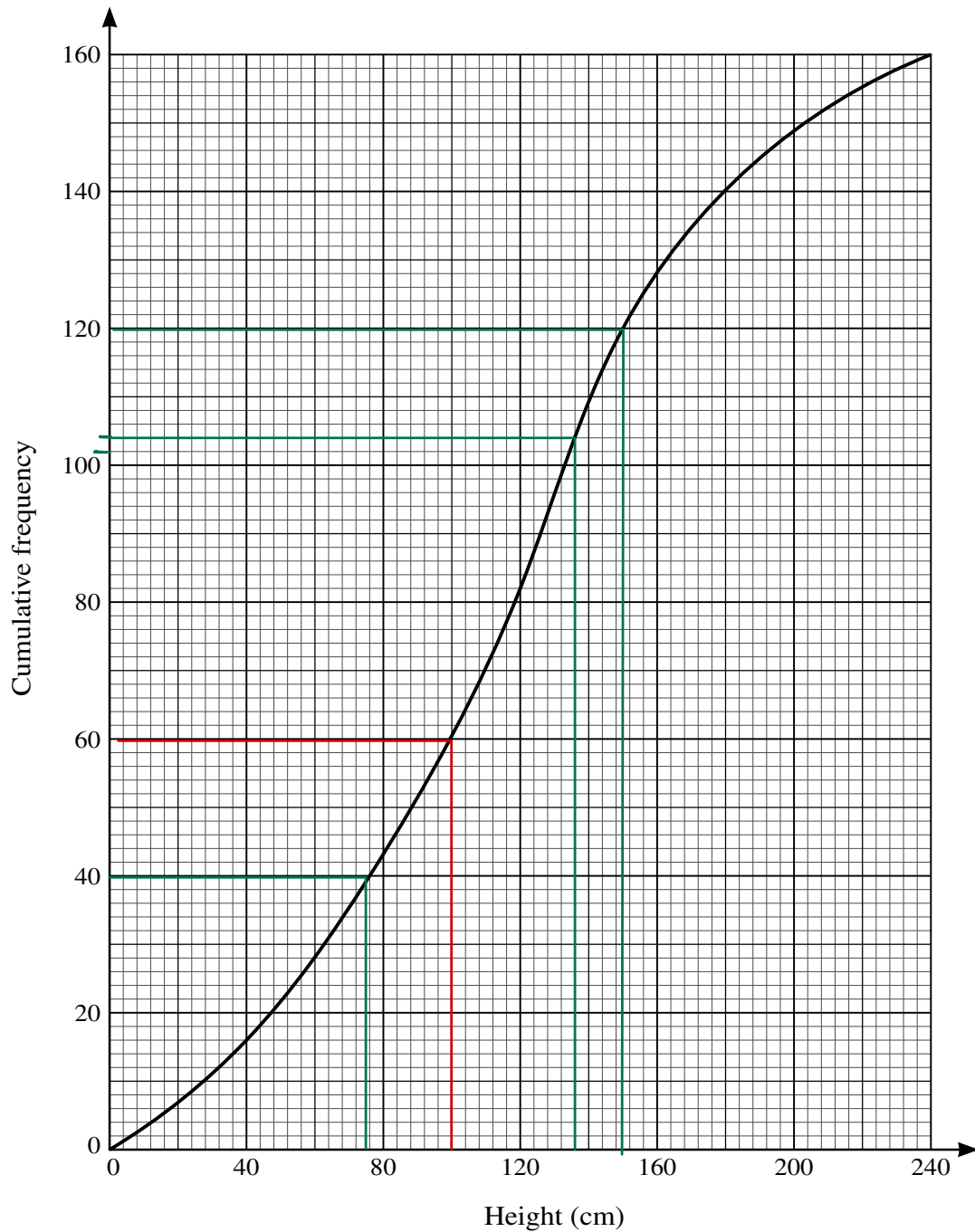
Cumulative Frequency

- A cumulative frequency graph is used with data that has been organised into a grouped frequency table
- The cumulative frequency graph can be used to find estimates of percentiles and quartiles
- As the data is grouped, it is not possible to find the actual values of these statistics but you can find an estimate with the help of a cumulative frequency graph

Less than a value on the horizontal axis $\rightarrow$ Corresponding value on the vertical axis

Greater than a value on the horizontal axis $\rightarrow$ Total - Corresponding value on the vertical axis

- 1 The heights in cm of 160 sunflower plants were measured. The results are summarised on the following cumulative frequency curve.



- (a) Use the graph to estimate the number of plants with heights less than 100 cm.

[1]

60

Corresponding value

- (b) Use the graph to estimate the 65th percentile of the distribution.

[2]

$$\frac{65}{100} \times 160 = 104 \longrightarrow 136$$

- (c) Use the graph to estimate the interquartile range of the heights of these plants.

[2]

$$U.Q : \frac{75}{100} \times 160 = 120 \longrightarrow 150$$

$$L.Q : \frac{25}{100} \times 160 = 40 \longrightarrow 76$$

$$150 - 76 = \boxed{74}$$

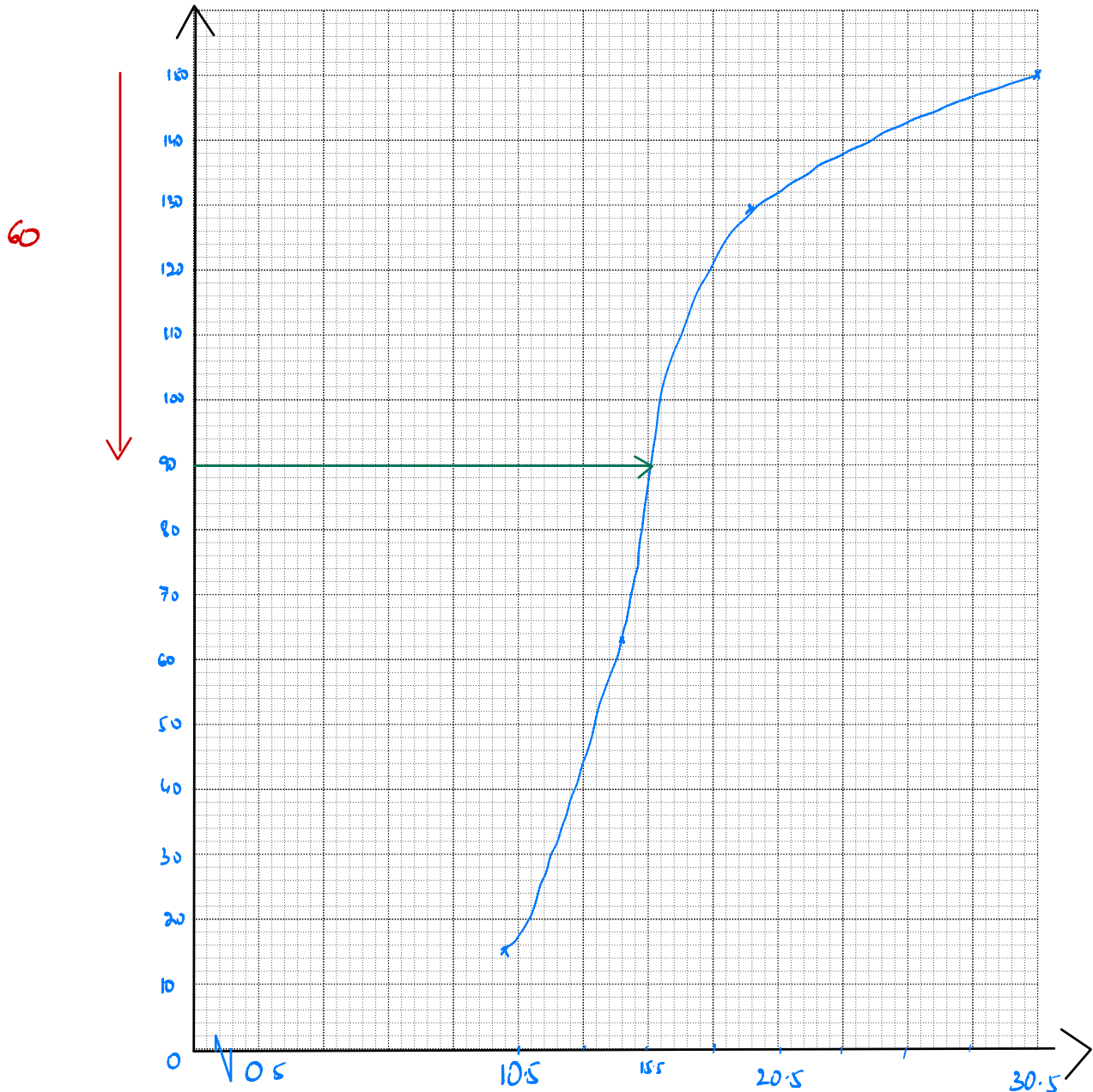
- 7 Helen measures the lengths of 150 fish of a certain species in a large pond. These lengths, correct to the nearest centimetre, are summarised in the following table.

Length (cm)	0 – 9	10 – 14	15 – 19	20 – 30
Frequency	15	48	66	21

- (a) Draw a cumulative frequency graph to illustrate the data.

[4]

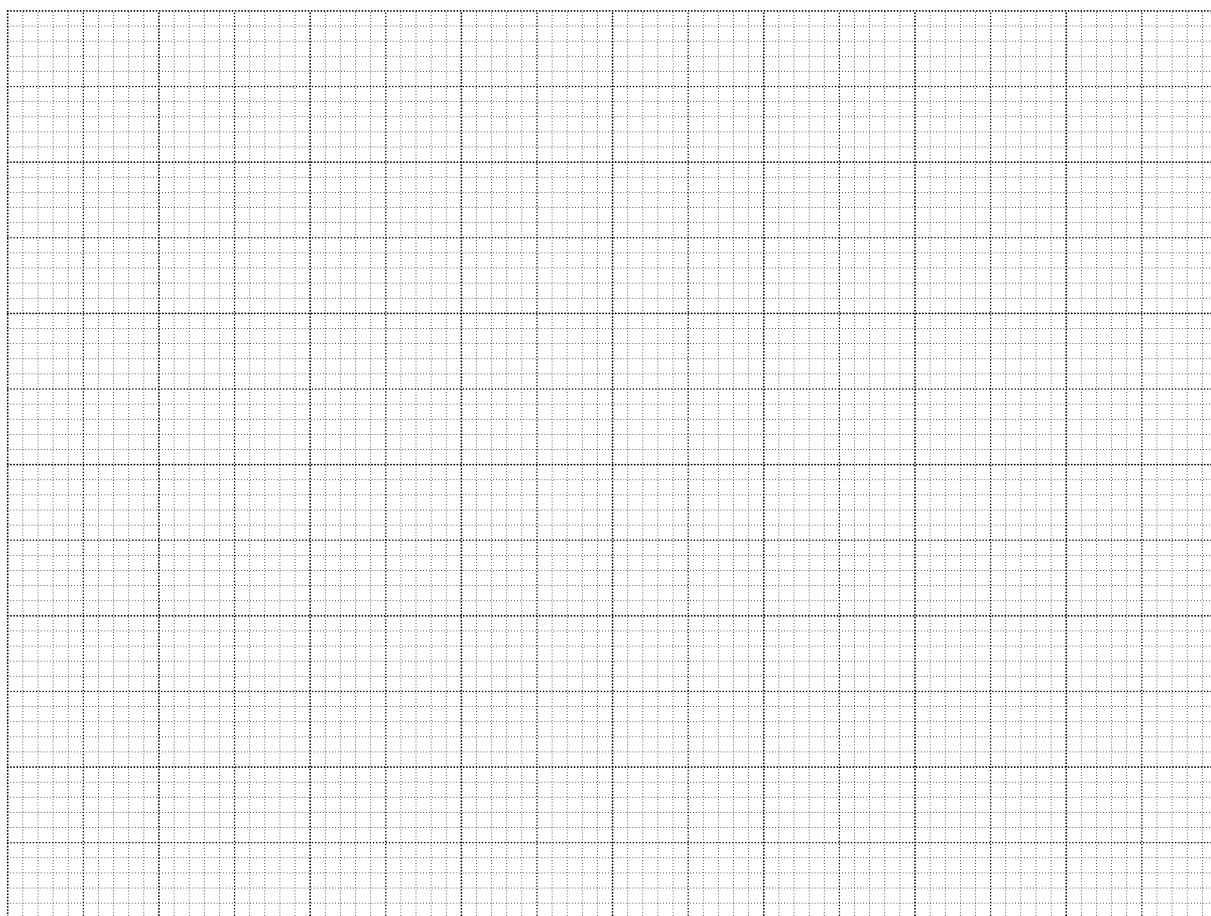
Length	0 – 9.5	9.5 – 14.5	14.5 – 19.5	19.5 – 30.5
C.f	15	63	129	150



- 1 There are 900 students in a certain year-group. An identical puzzle is given to each student and the time taken, t minutes, to complete the puzzle is recorded. These times are summarised in the following frequency table.

Time taken, t minutes	$t \leq 3$	$3 < t \leq 4$	$4 < t \leq 5$	$5 < t \leq 6$	$6 < t \leq 8$	$8 < t \leq 10$	$10 < t \leq 14$
Frequency	120	180	200	160	110	80	50

On the grid, draw a cumulative frequency graph to represent the data. Use your graph to estimate the median time taken by these students to complete the puzzle. [4]



.....

.....

.....

.....

