3) Representations of data

3.1) Outliers

Chapter CONTENTS

Worked example

The scores of 10 students are recorded:

$$1, 8, 10, 9, -7, 21, 11, 10, 35, 0.3$$

An outlier is an observation that falls either

1.5 ×interquartile range above the upper quartile

or

1.5 ×interquartile range below the lower quartile.

Find any outliers.

Your turn

The scores of 10 students are recorded:

$$5, 12, 14, 13, 8, 9, 51, -4, 59, 0.2$$

An outlier is an observation that falls either

1.5 ×interquartile range above the upper quartile

or

1.5 ×interquartile range below the lower quartile.

Find any outliers. $-4, 51, 59$

Worked example

The scores of 10 students are recorded:

$$1, 8, 10, 9, -7, 21, 11, 10, 35, 0.3$$

An outlier is an observation that falls outside ± 2 standard deviations from the mean.

Find any outliers.

Your turn

The scores of 10 students are recorded:

$$5, 12, 14, 13, 8, 9, 51, -4, 59, 0.2$$

An outlier is an observation that falls either

1.5 ×interquartile range above the upper quartile

or

1.5 ×interquartile range below the lower quartile.

Find any outliers. 59

Worked example

Clean this data on ages of people in a group:

$12, 13, 14, 12, 13, 156$

Your turn

Clean this data on ages of people in a group:

$5, 7, 6, 5, 5, 567, 7, 6$

$\bar{x} + 2\sigma = 447.164 \ldots$

$567 \gg 447.164$ and an age of 567 is impossible.

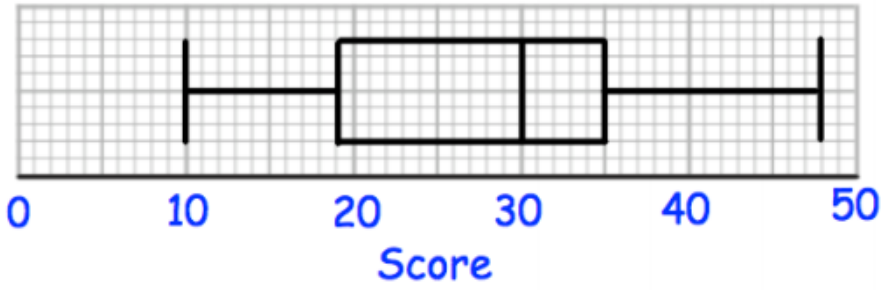
$\therefore$ The clear anomaly of 567 should be removed from the data.

3.2) Box plots

Chapter CONTENTS

Worked example

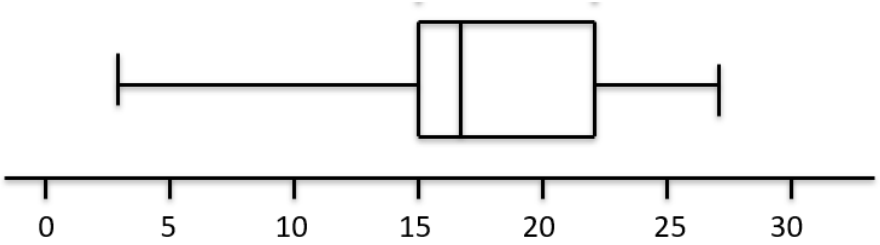
Using the box plot, write down:



a) The minimum
b) The lower quartile
c) The median
d) The upper quartile
e) The maximum
f) The range
g) The interquartile range

Your turn

Using the box plot, write down:



a) The minimum 3
b) The lower quartile 15
c) The median 17
d) The upper quartile 22
e) The maximum 27
f) The range 24
g) The interquartile range 7

Worked example

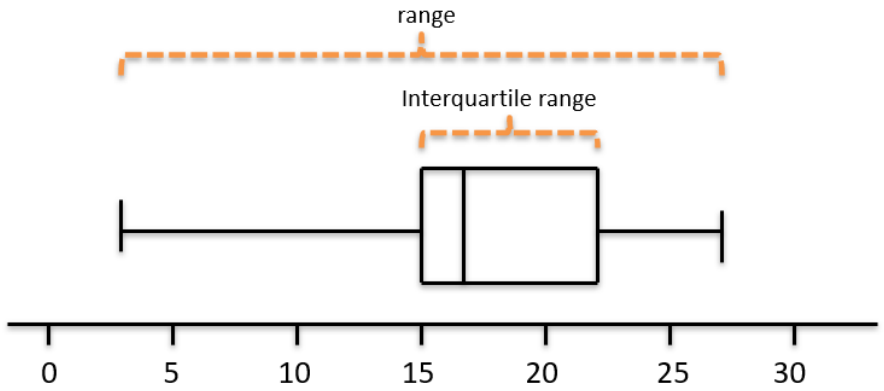
Sketch a box plot given the following data:

Minimum	Lower Quartile	Median	Upper Quartile	Maximum
2	11	18	20	29

Your turn

Sketch a box plot given the following data:

Minimum	Lower Quartile	Median	Upper Quartile	Maximum
3	15	17	22	27



Worked example

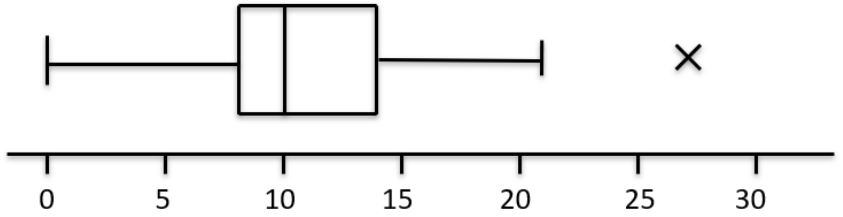
An outlier is an observation that falls either
1.5 ×interquartile range above the upper quartile
or
1.5 ×interquartile range below the lower quartile.
Sketch a box plot for this data, marking any outliers.

Smallest values	Largest values	Lower quartile	Median	Upper quartile
0, 4	22, 26	9	11	15

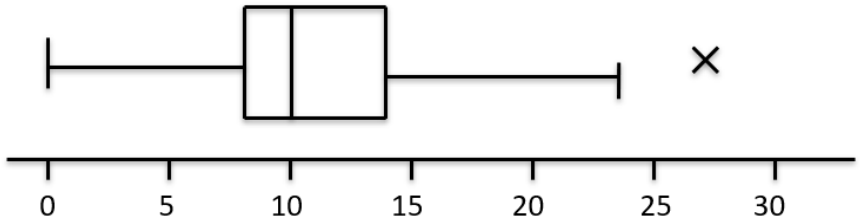
Your turn

An outlier is an observation that falls either
1.5 ×interquartile range above the upper quartile
or
1.5 ×interquartile range below the lower quartile.
Sketch a box plot for this data, marking the outlier boundaries

Smallest values	Largest values	Lower quartile	Median	Upper quartile
0, 3	21, 27	8	10	14

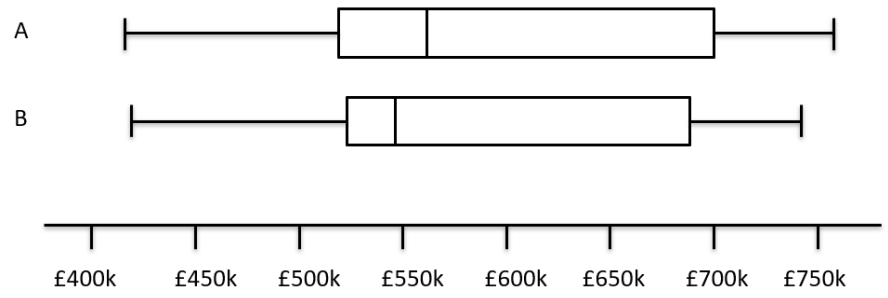


or



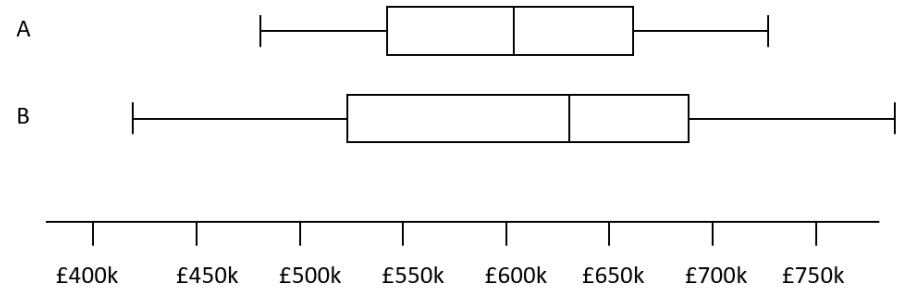
Worked example

Compare the house prices of locations A and B



Your turn

Compare the house prices of locations A and B



- The interquartile range of house prices in B is greater than A.
- The range of house prices in B is greater than A.
- The median house price in Kingston was greater than that in Croydon

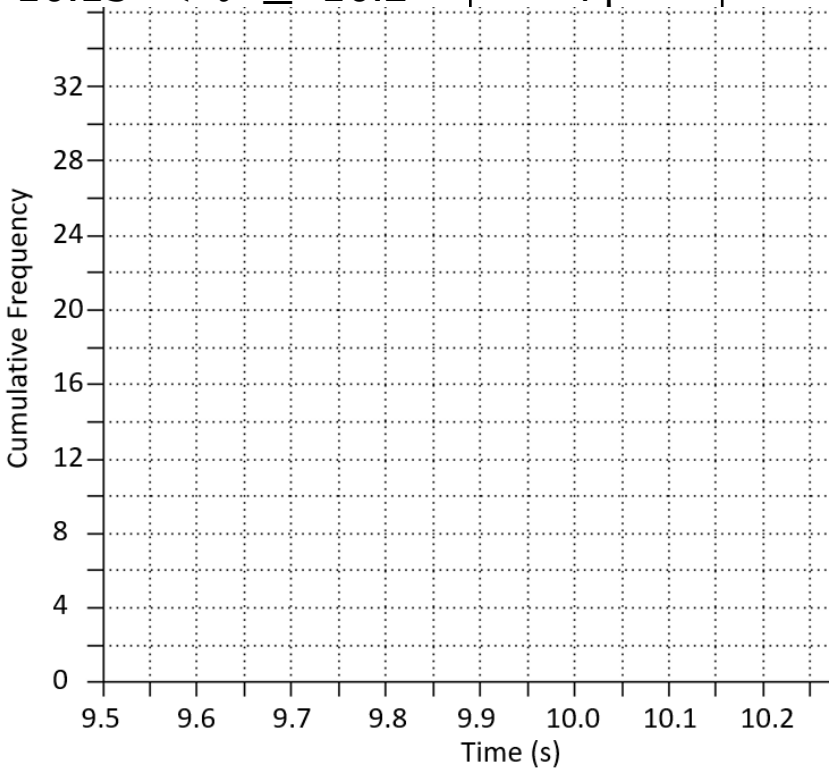
3.3) Cumulative frequency

Chapter CONTENTS

Worked example

Draw a cumulative frequency diagram for the data:

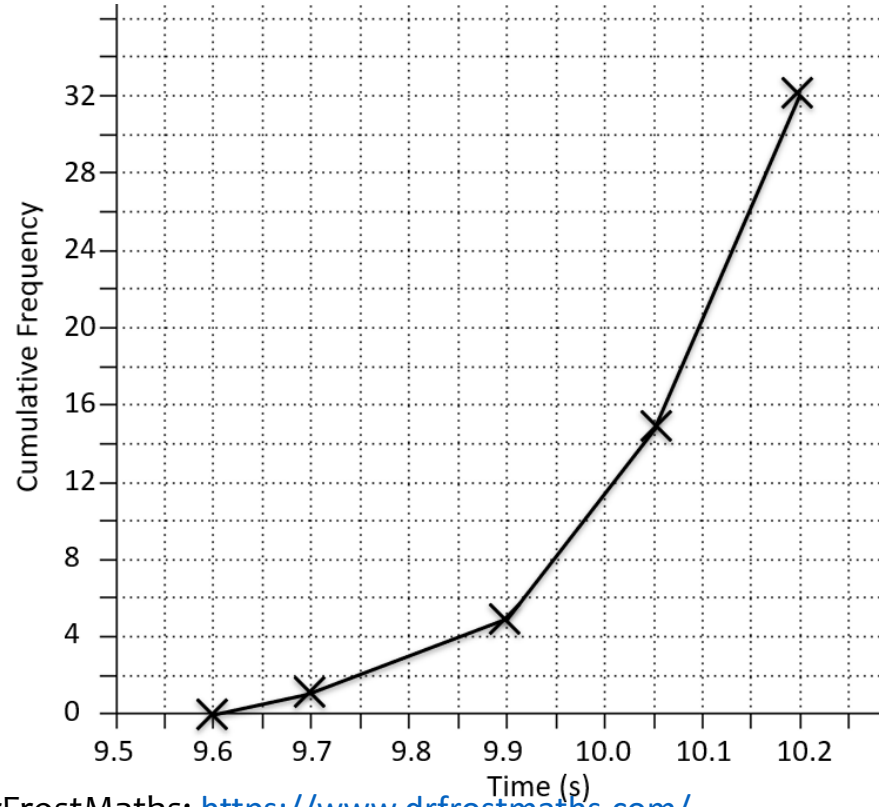
Time (s)	Frequency
$9.6 < t \leq 9.8$	3
$9.8 < t \leq 10.05$	7
$10.05 < t \leq 10.15$	8
$10.15 < t \leq 10.2$	14



Your turn

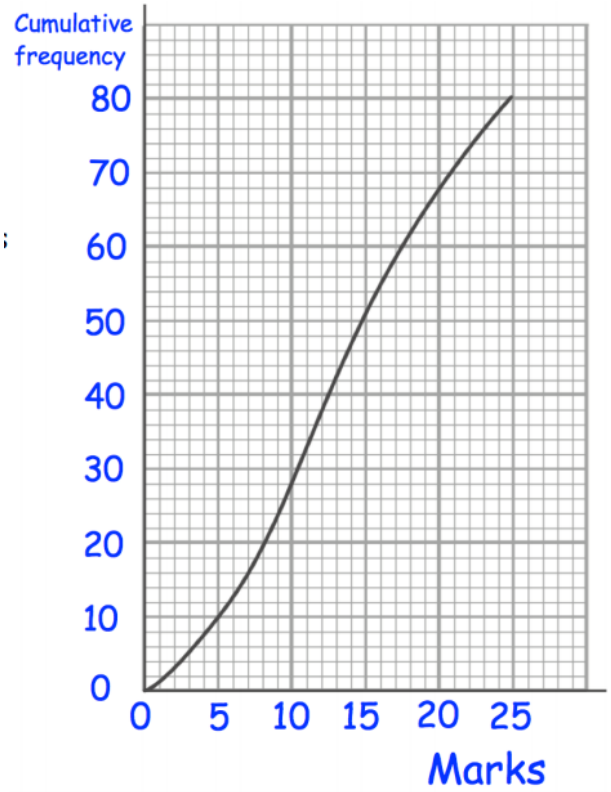
Draw a cumulative frequency diagram for the data:

Time (s)	Frequency
$9.6 < t \leq 9.7$	1
$9.7 < t \leq 9.9$	4
$9.9 < t \leq 10.05$	10
$10.05 < t \leq 10.2$	17



Worked example

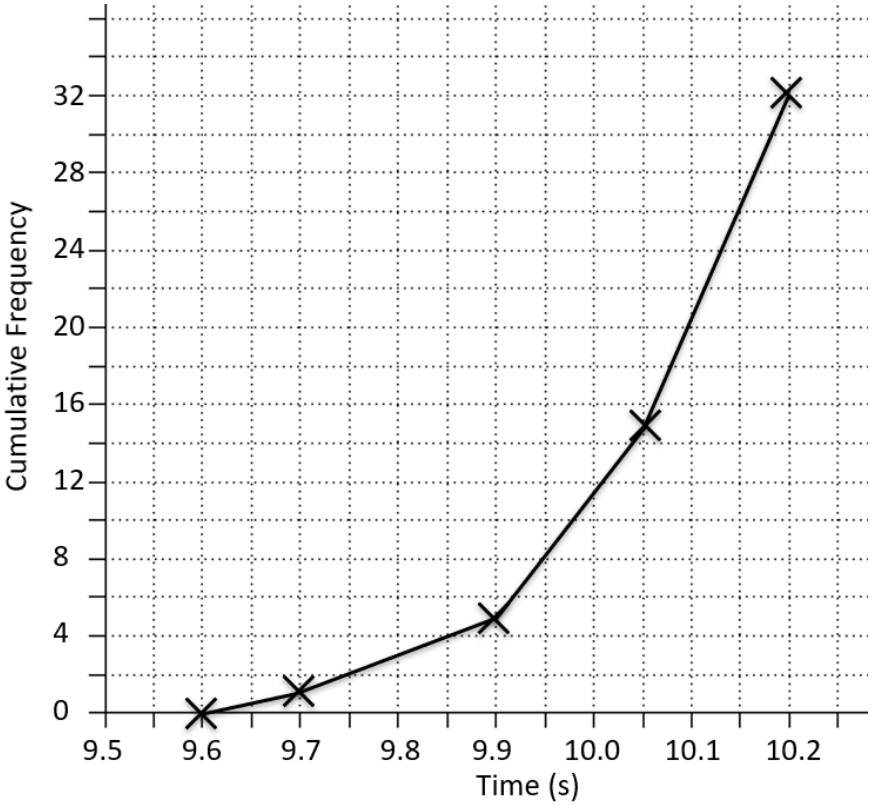
Use the cumulative frequency diagram to estimate the:



a) Lower quartile
b) Median
c) Upper quartile
d) 60th percentile

Your turn

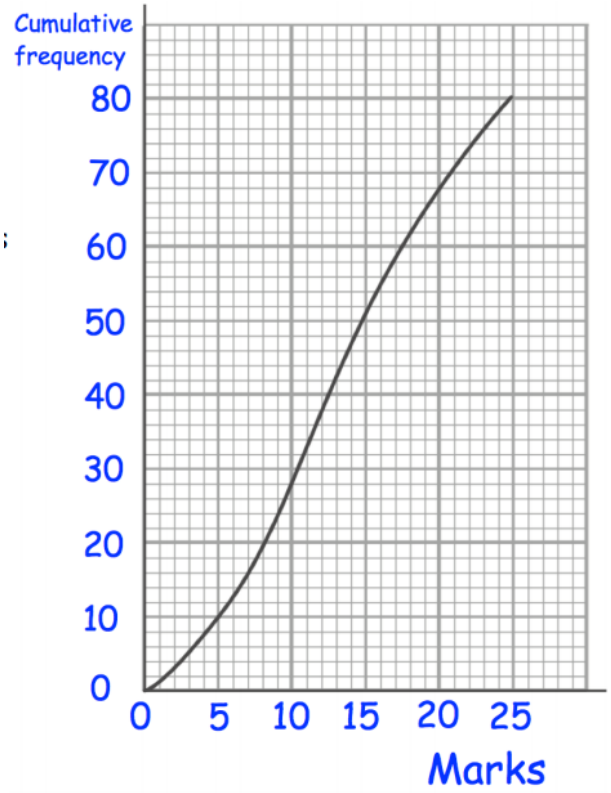
Use the cumulative frequency diagram to estimate the:



a) Lower quartile 9.95 s
b) Median 10.07 s
c) Upper quartile 10.13 s
d) 90th percentile 10.17 s

Worked example

Use the cumulative frequency diagram to estimate the:

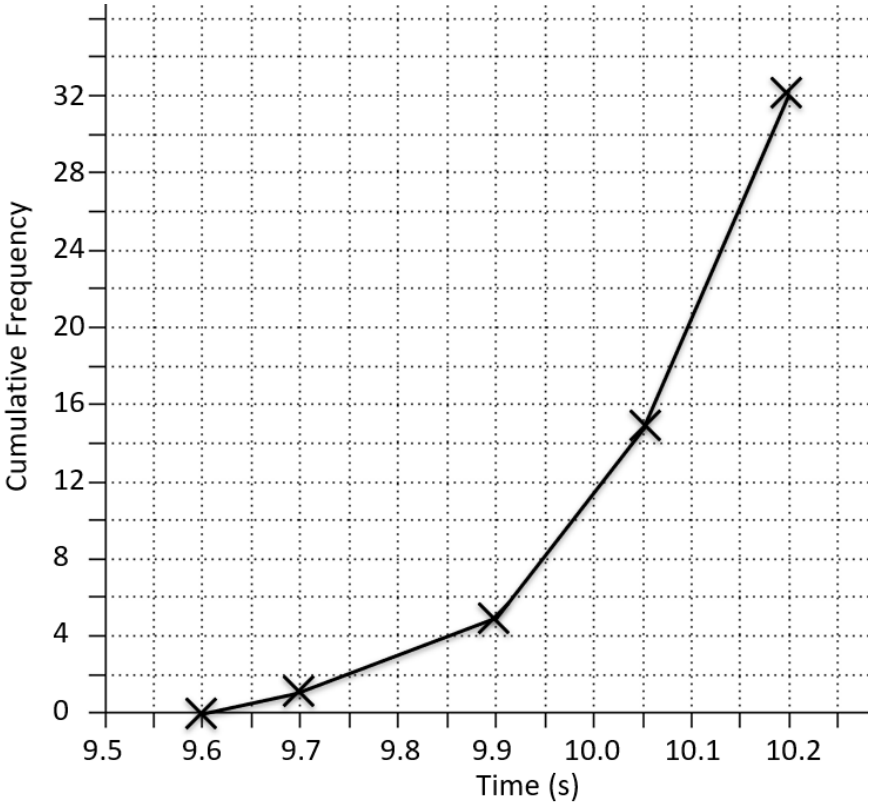


a) Interquartile range

b) 10^{th} – 90^{th} interpercentile range

Your turn

Use the cumulative frequency diagram to estimate the:

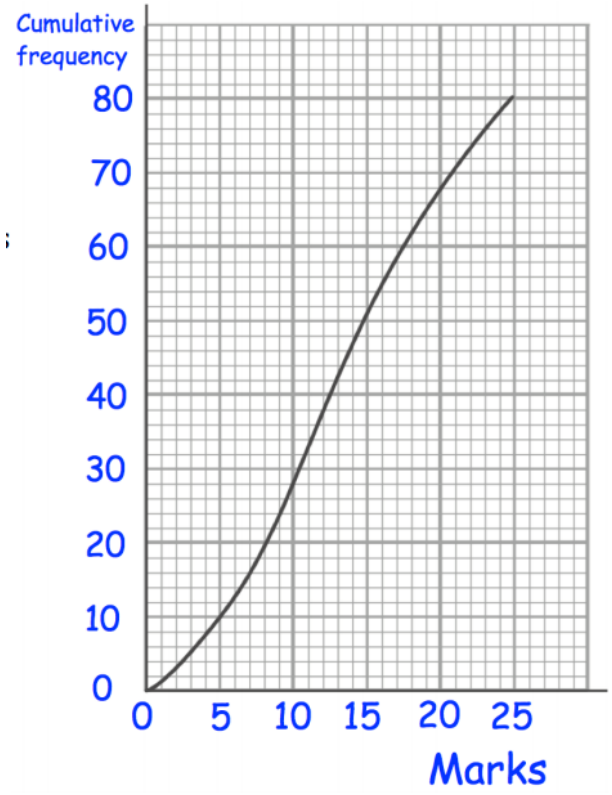


a) Interquartile range 0.18 s

b) 20^{th} – 80^{th} interpercentile range 0.21 s

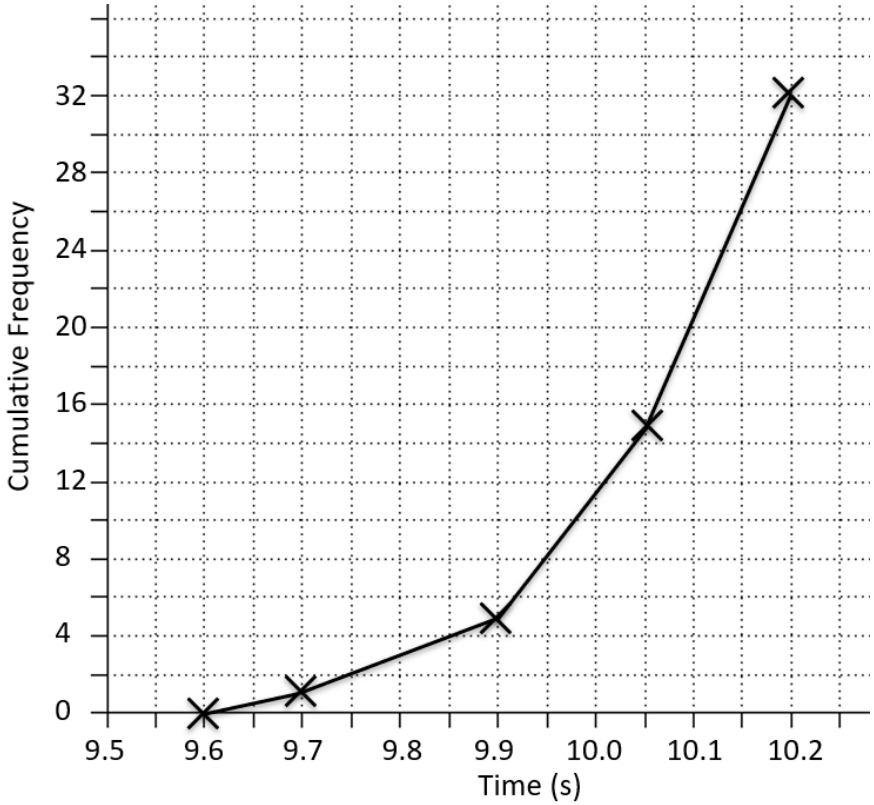
Worked example

Use the cumulative frequency diagram to estimate the number of students who achieved fewer than 23 marks.



Your turn

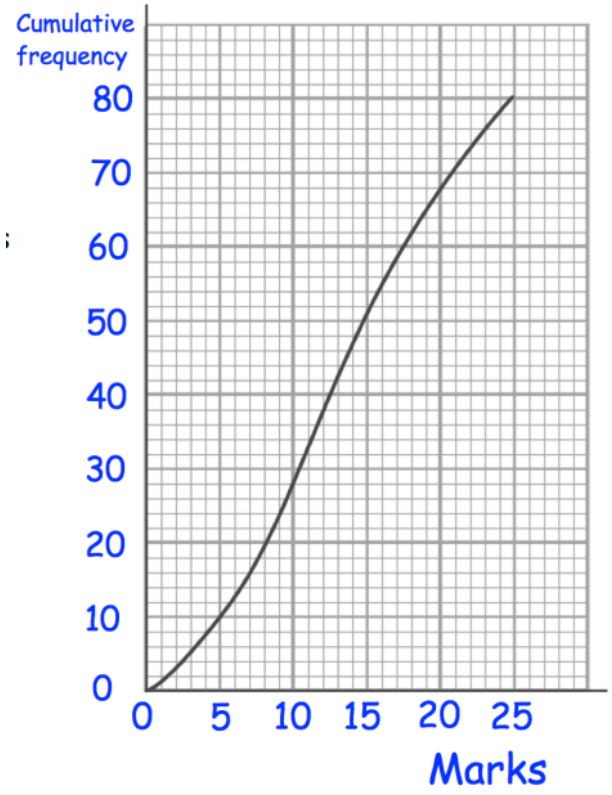
Use the cumulative frequency diagram to estimate the number of runners who had a time less than 10.15 seconds.



26

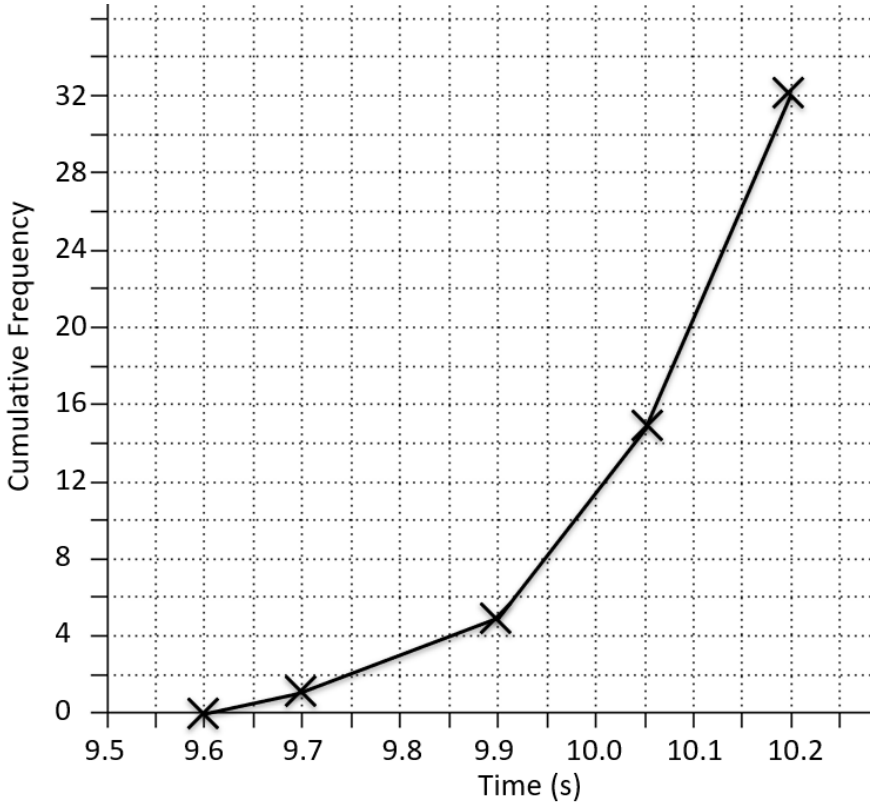
Worked example

Use the cumulative frequency diagram to estimate the number of students who achieved more than 12 marks.



Your turn

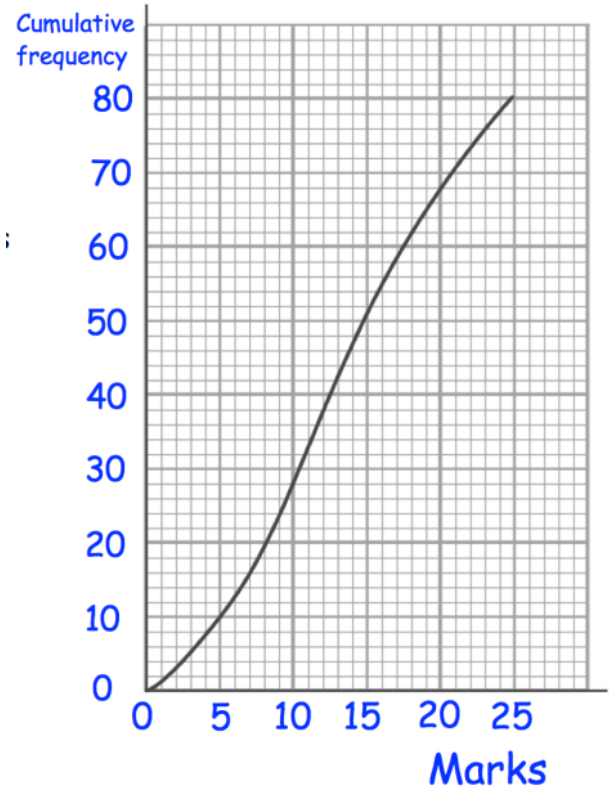
Use the cumulative frequency diagram to estimate the number of runners who had a time greater than 9.95 seconds.



24

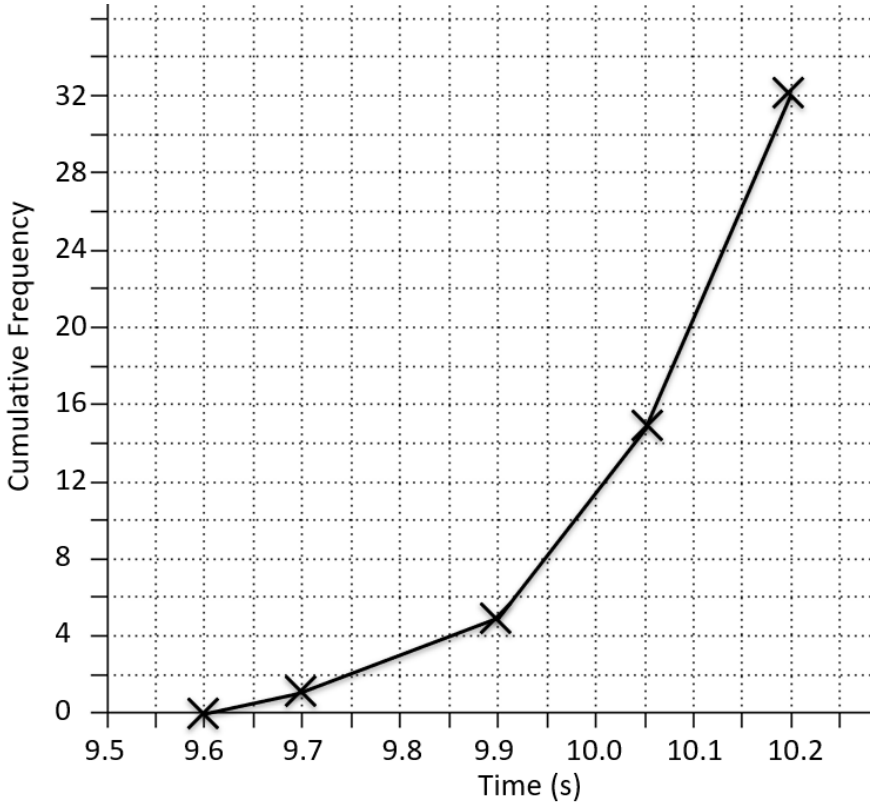
Worked example

Use the cumulative frequency diagram to estimate the number of students who achieved between 7 and 21 marks.



Your turn

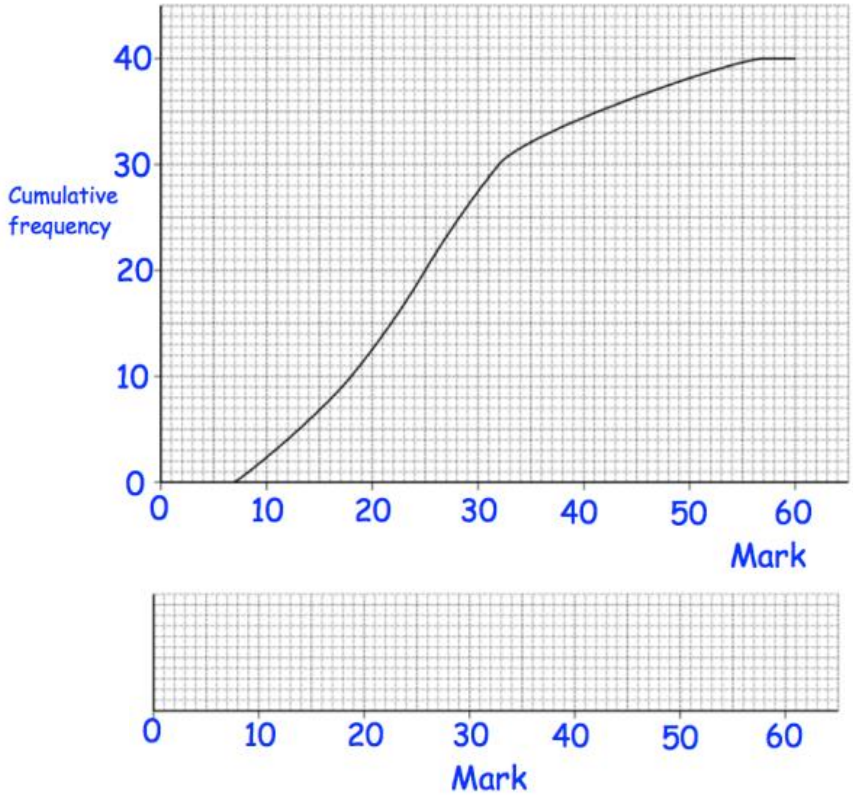
Use the cumulative frequency diagram to estimate the number of runners who had a time between 9.8 and 10 seconds.



8

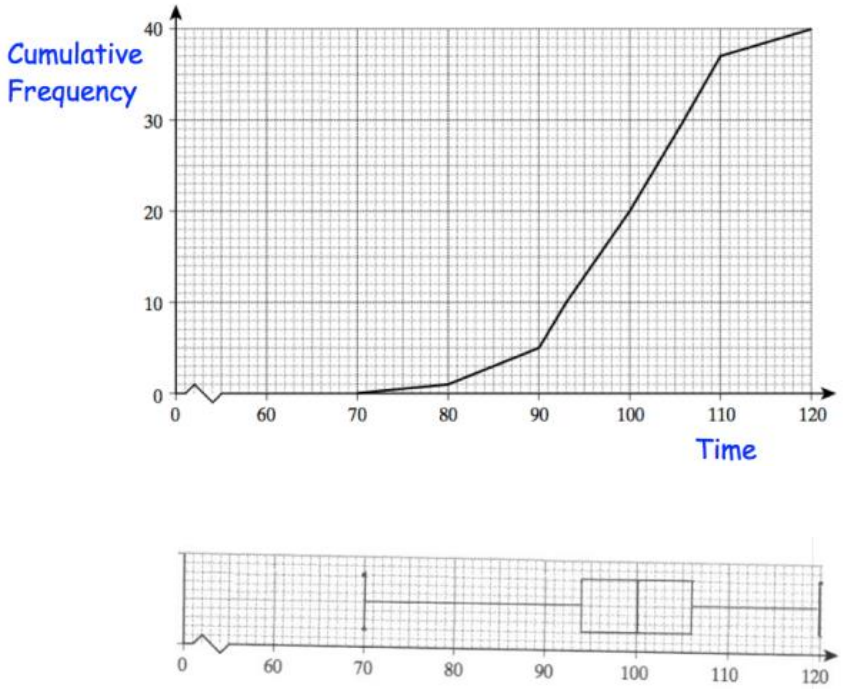
Worked example

Use the cumulative frequency diagram to draw a box plot:



Your turn

Use the cumulative frequency diagram to draw a box plot:



3.4) Histograms

Chapter CONTENTS

Worked example

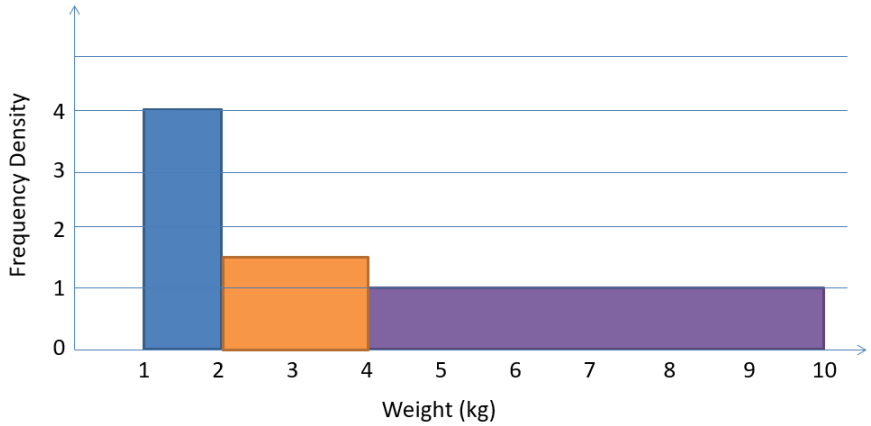
Plot a histogram for the data:

Height, h (nearest cm)	**Frequency**
$1 < h \leq 5$	5
$5 < h \leq 8$	4
$8 < h \leq 9$	3

Your turn

Plot a histogram for the data:

Weight, w (nearest kg)	**Frequency**
$1 \leq w < 2$	4
$2 \leq w < 4$	3
$4 \leq w < 9$	5



Worked example

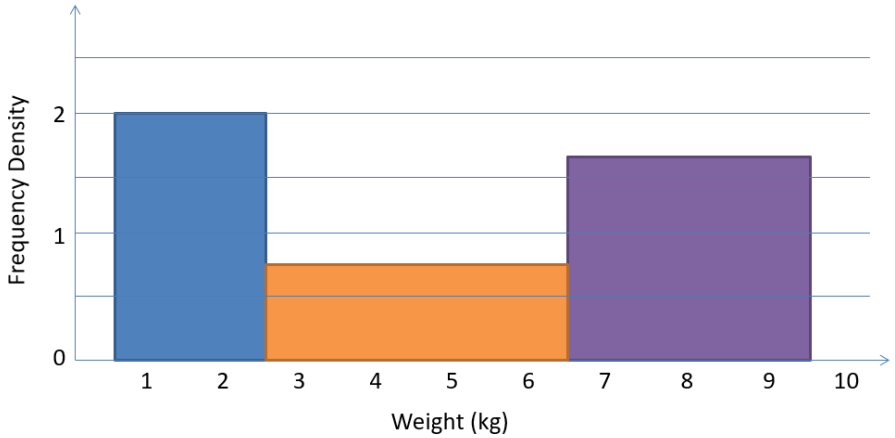
Plot a histogram for the data:

Height (nearest cm)	Frequency
1-4	5
5-7	4
8-9	3

Your turn

Plot a histogram for the data:

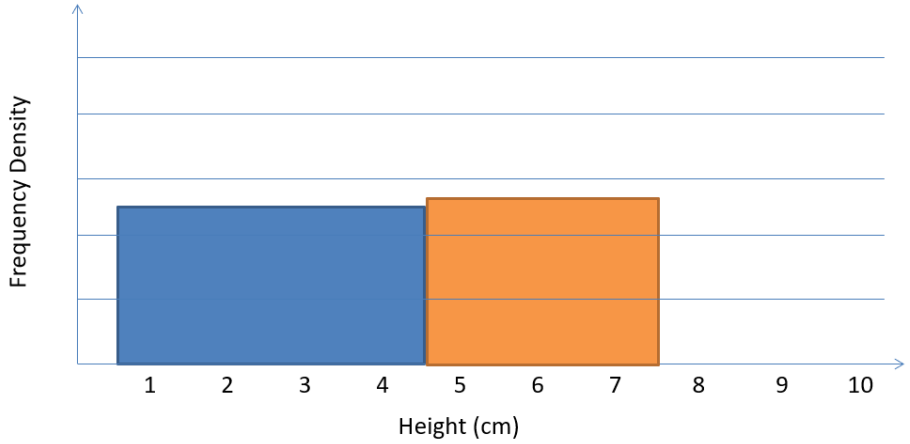
Weight (nearest kg)	Frequency
1-2	4
3-6	3
7-9	5



Worked example

Complete the table and histogram:

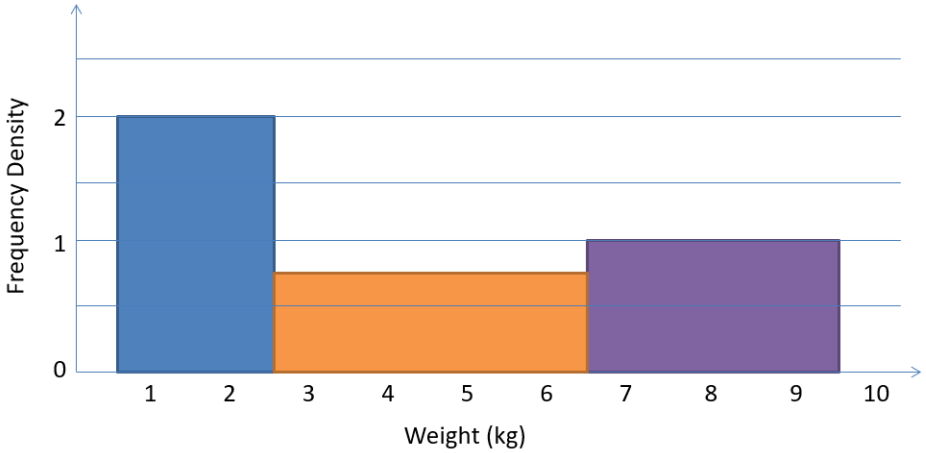
Height (nearest cm)	Frequency
1-4	
5-7	4
8-9	3



Your turn

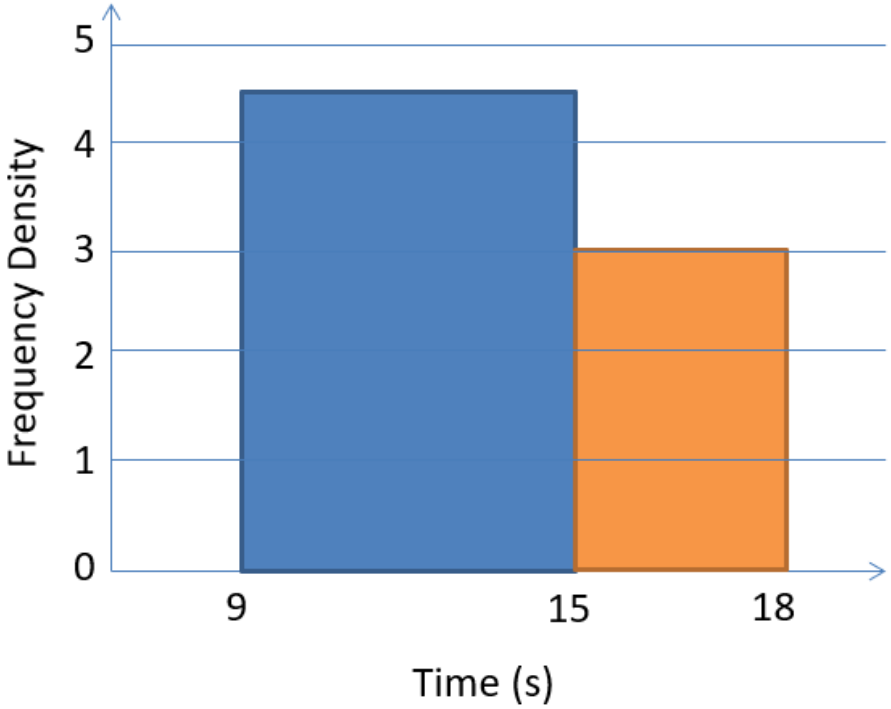
Complete the table and histogram:

Weight (nearest kg)	Frequency
1-2	4
3-6	3
7-9	



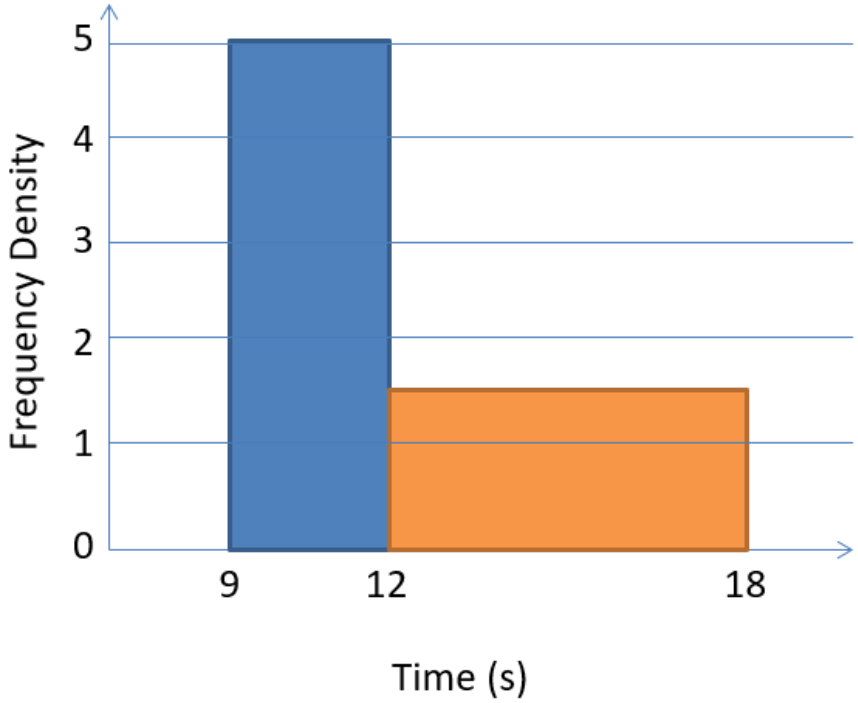
Worked example

There were 54 runners in a 100m race.
The following histogram represents their times.
Determine the number of runners with times below 13 seconds.



Your turn

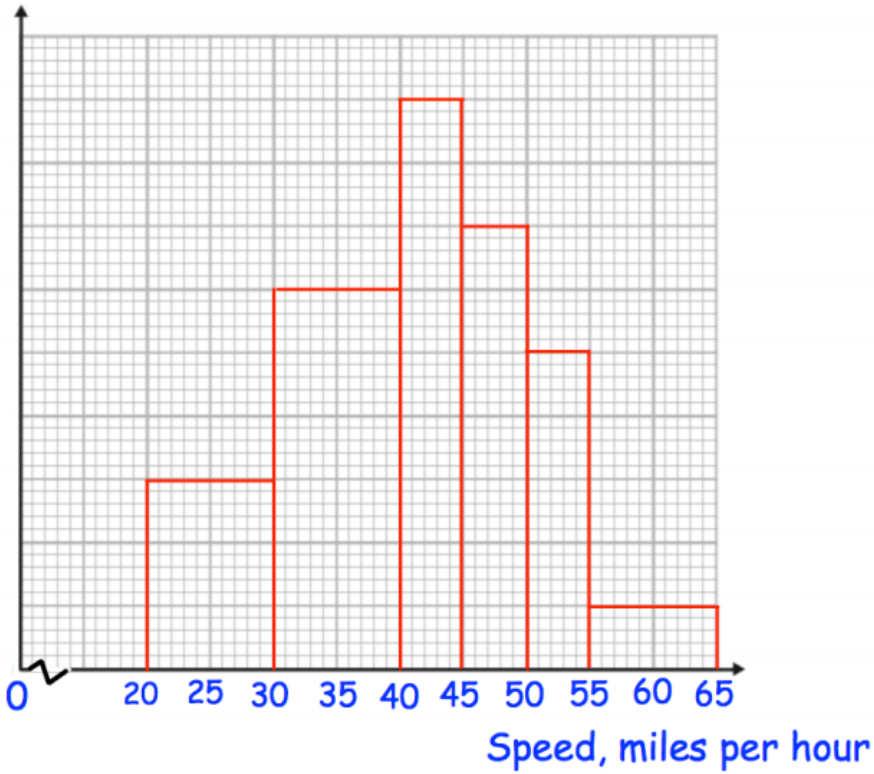
There were 60 runners in a 100m race.
The following histogram represents their times.
Determine the number of runners with times above 14 seconds.



15

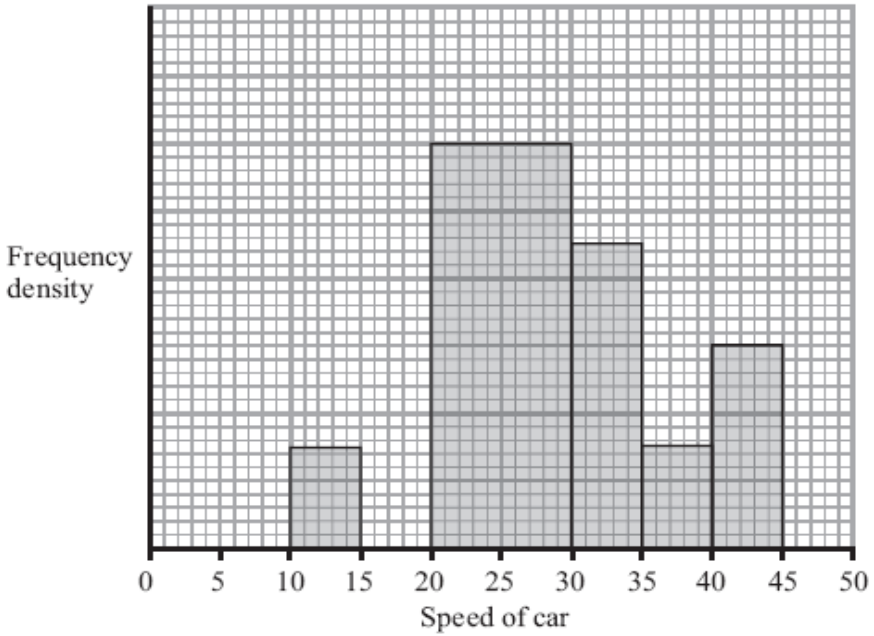
Worked example

The histogram shows the speeds of 82 cars.
Calculate the number of cars that were driving at speeds of at least 50 miles per hour.



Your turn

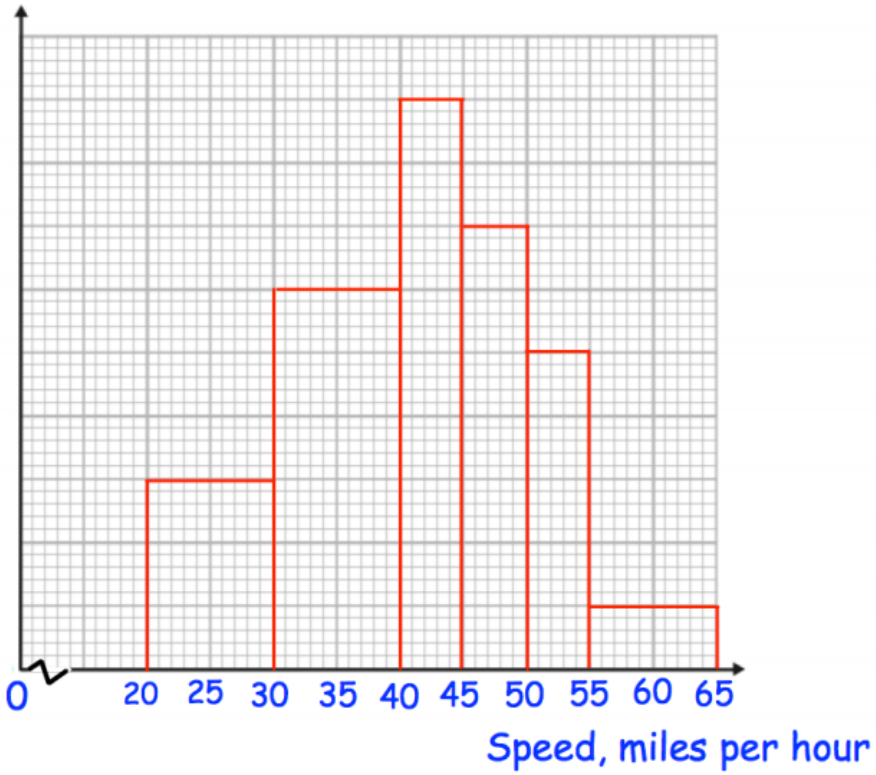
The histogram shows the speeds of 450 cars.
Calculate the number of cars that were driving at speeds of at least 35 miles per hour.



90

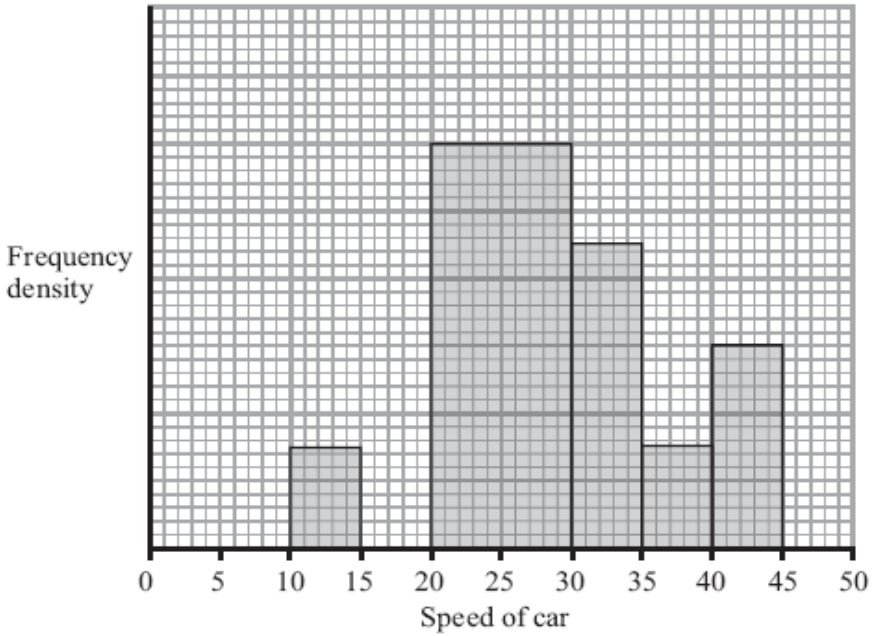
Worked example

The histogram shows the speeds of 82 cars.
Estimate the mean speed.



Your turn

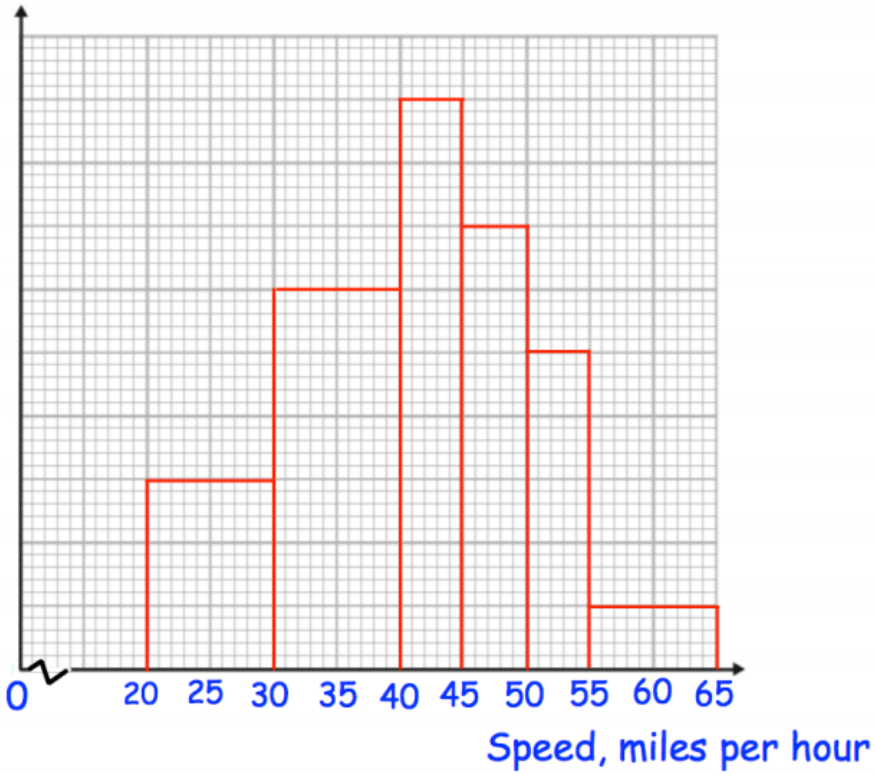
The histogram shows the speeds of 450 cars.
Estimate the mean speed.



28.8 mph

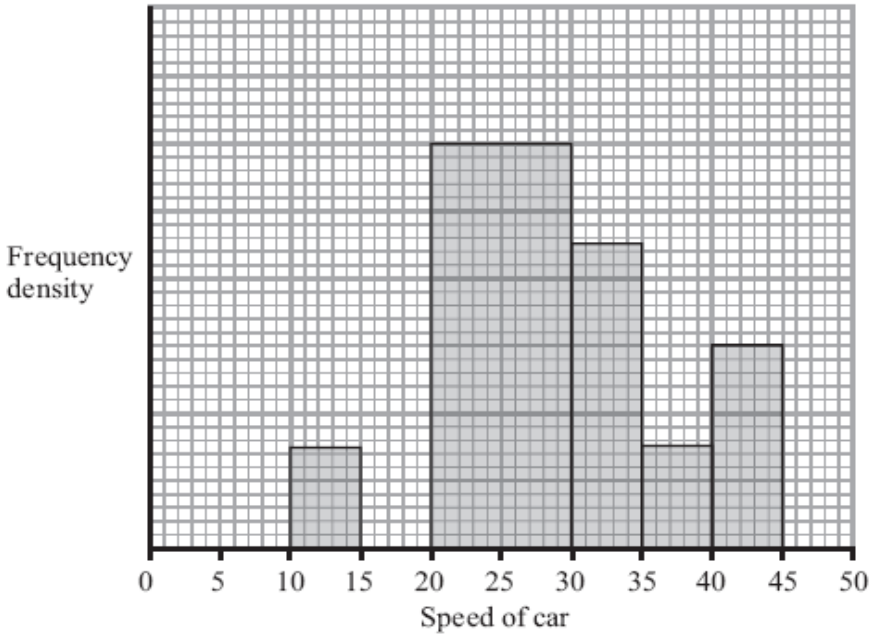
Worked example

The histogram shows the speeds of 82 cars.
Estimate the median speed



Your turn

The histogram shows the speeds of 450 cars.
Estimate the median speed



28.1 mph (3 sf)

Worked example

The frequency table shows some running times.
On a histogram the bar for 0-2 seconds is drawn with width 8cm and height 12cm.
Find the width and height of the bar for 2-6 seconds.

Time (seconds)	**Frequency**
$0 \leq t < 2$	12
$2 \leq t < 6$	3

Your turn

The frequency table shows some running times.
On a histogram the bar for 0-4 seconds is drawn with width 6cm and height 8cm. Find the width and height of the bar for 4-6 seconds.

Time (seconds)	**Frequency**
$0 \leq t < 4$	8
$4 \leq t < 6$	9

Width = 3 cm
Height = 18 cm

Worked example

The variable x was measured to the nearest whole number.
On a histogram the bar representing the $2 - 7$ class has a width of 4 cm and a height of 12 cm.
Find the width and height of the $8 - 10$ class

x	Frequency
$2 - 7$	18
$8 - 10$	6
$12 -$	4

Your turn

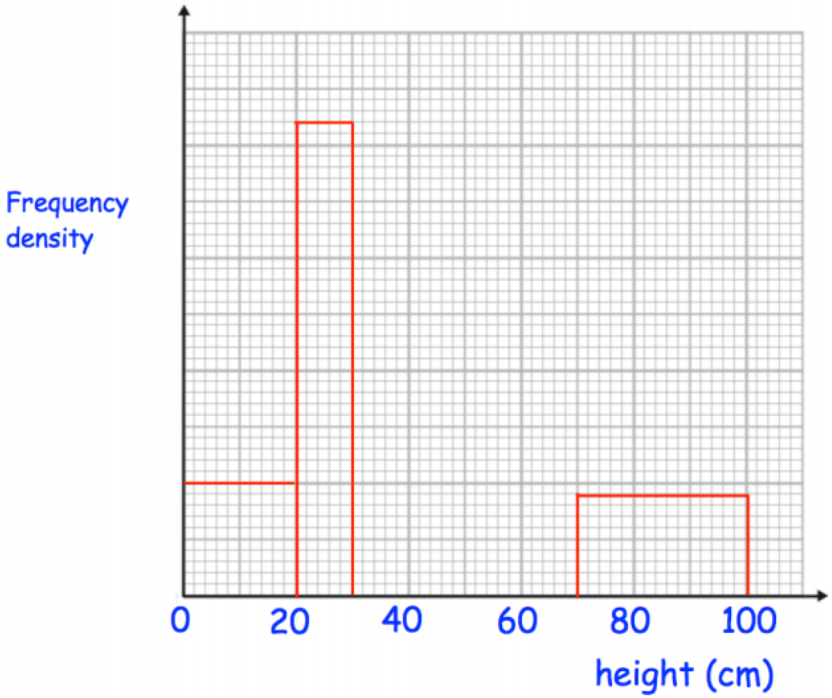
The variable x was measured to the nearest whole number.
On a histogram the bar representing the $10 - 15$ class has a width of 2 cm and a height of 5 cm.
Find the width and height of the $16 - 18$ class

x	Frequency
$10 - 15$	15
$16 - 18$	9
$19 -$	16

Width $= 1$ cm
Height $= 6$ cm

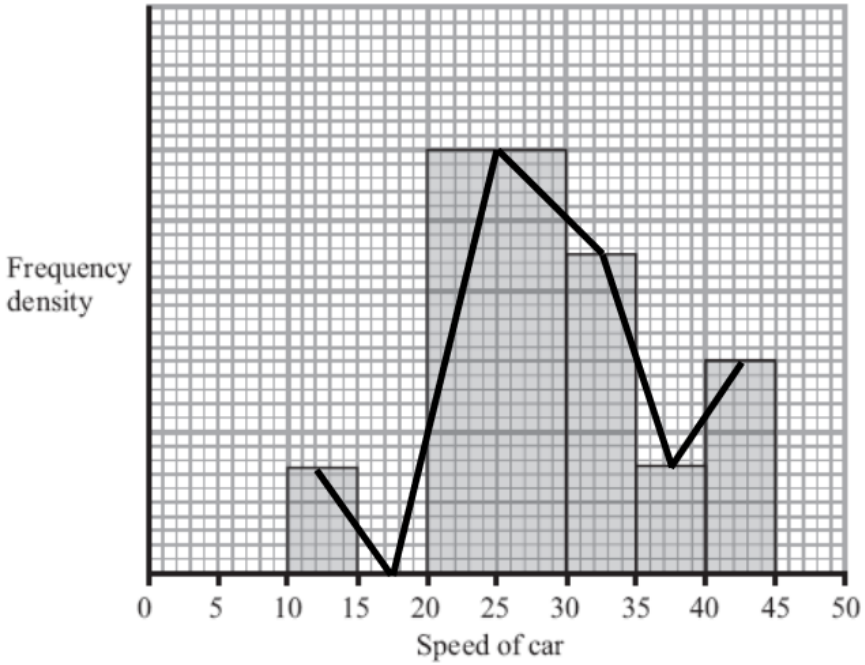
Worked example

Draw a frequency polygon.



Your turn

Draw a frequency polygon.



Worked example

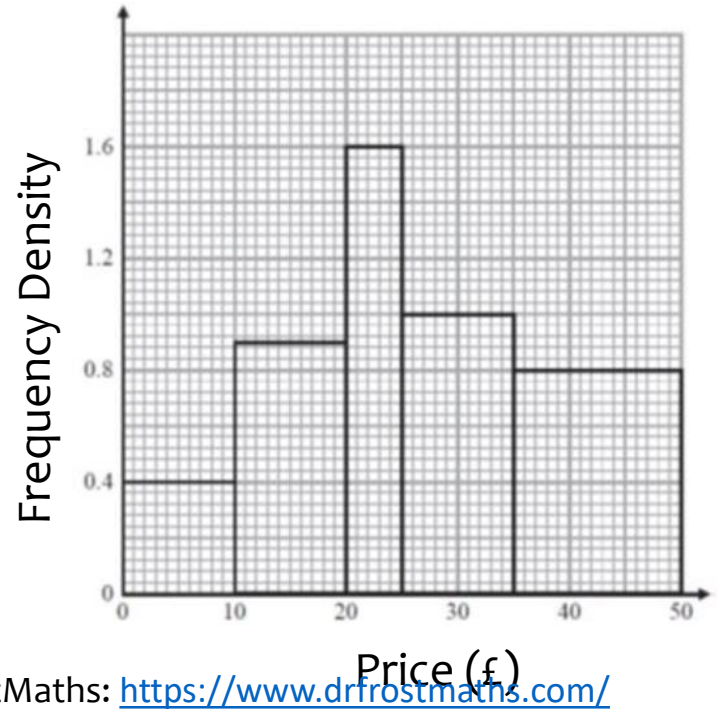
Plot a histogram

Height, x (cm)	Frequency
$140 < x \leq 155$	6
$155 < x \leq 175$	14
$175 < x \leq 185$	6
$185 < x \leq 190$	21

Your turn

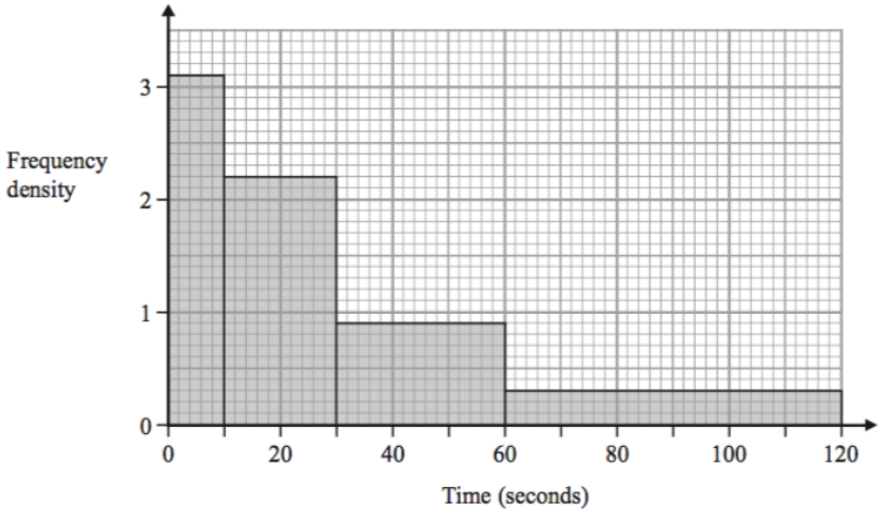
Plot a histogram

Price, y (£)	Frequency
$0 < y \leq 10$	4
$10 < y \leq 20$	9
$20 < y \leq 25$	8
$25 < y \leq 35$	10
$35 < y \leq 50$	12



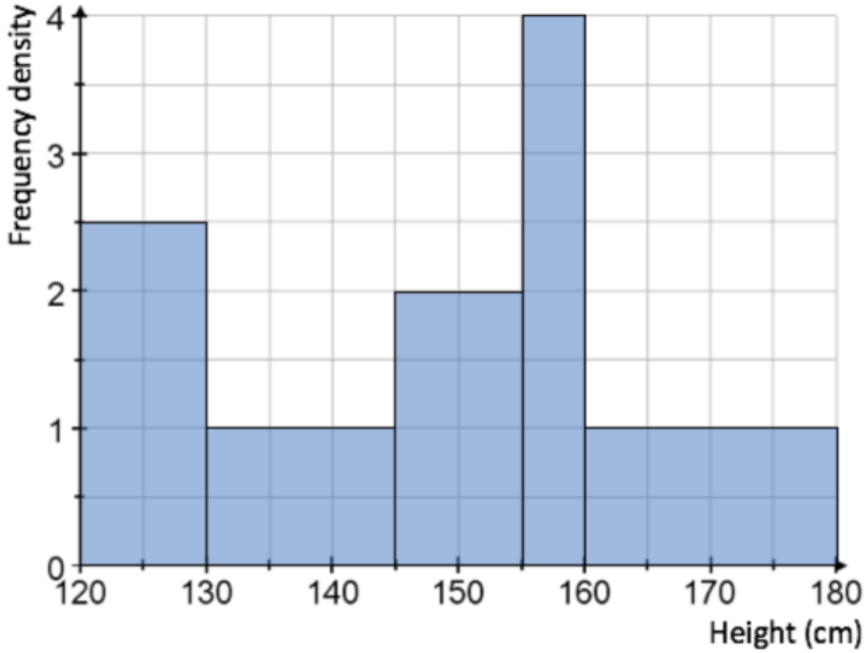
Worked example

Draw a frequency table from the histogram



Your turn

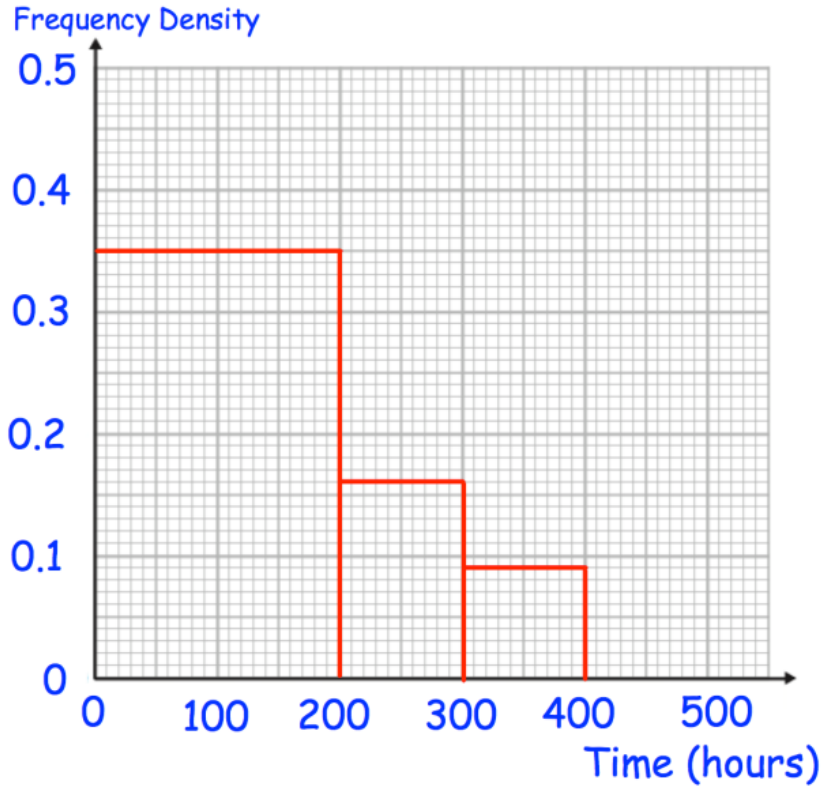
Draw a frequency table from the histogram



Height, y (cm)	Frequency
$120 < y \leq 130$	25
$130 < y \leq 145$	15
$145 < y \leq 155$	20
$155 < y \leq 160$	20
$160 < y \leq 180$	20

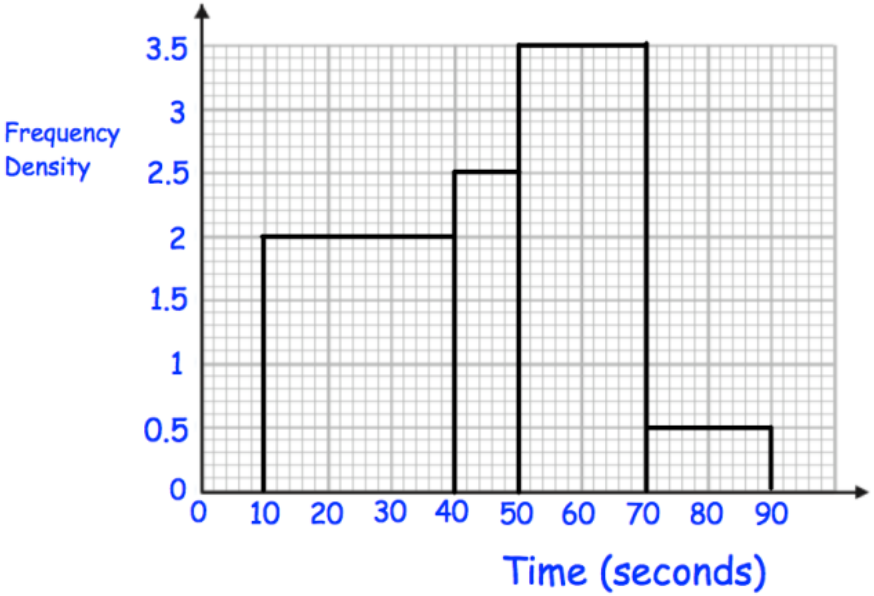
Worked example

Estimate the number of pilots who have flown under 350 hours.



Your turn

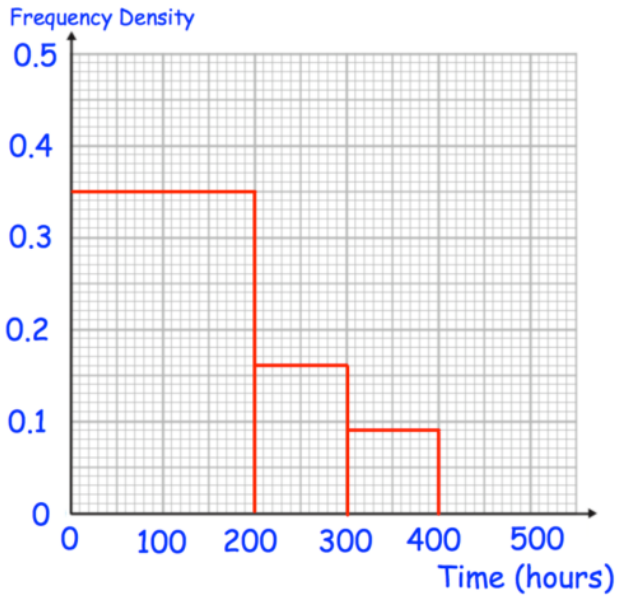
Estimate the number of students who took less than 60 seconds to complete the puzzle



120

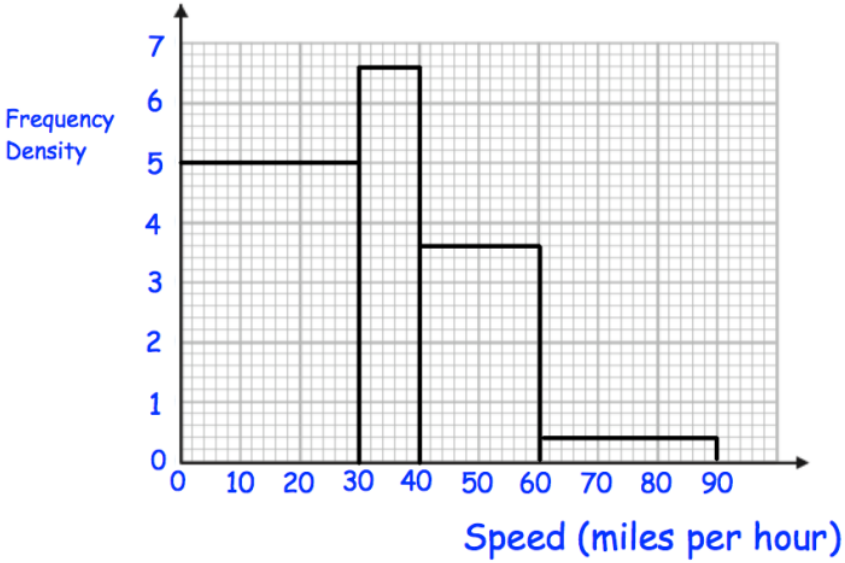
Worked example

Work out the percentage of pilots who have flown under 250 hours.



Your turn

Work out the percentage of cars that were under the speed limit of 60 mph

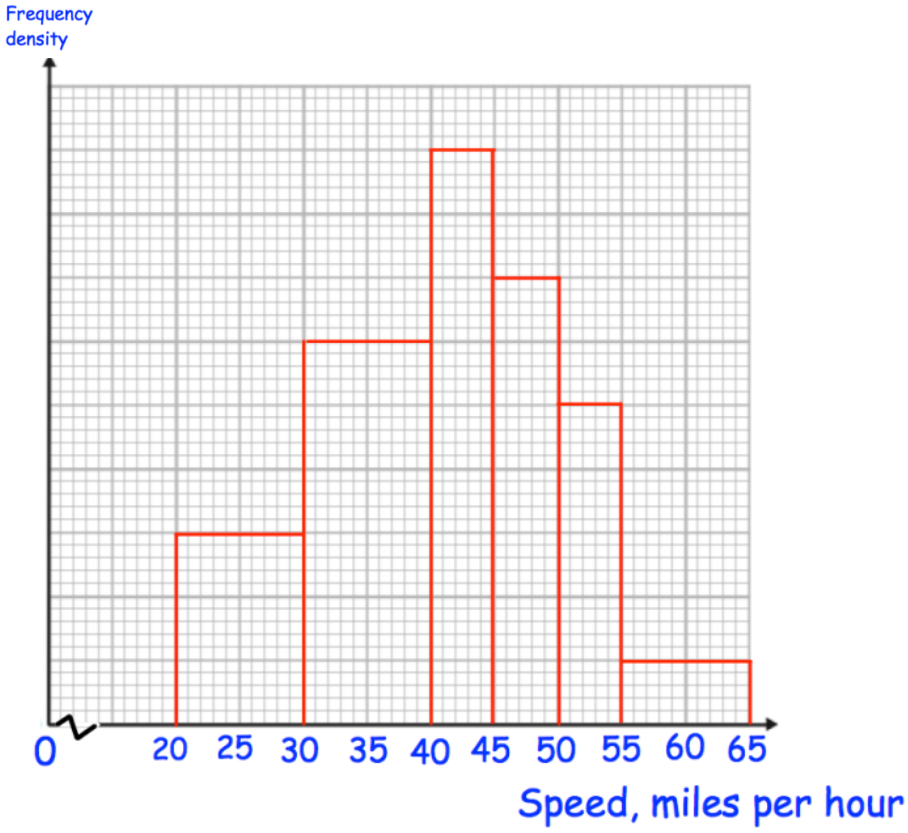


96%

Questions used with permission from Corbettmaths: https://corbettmaths.com/

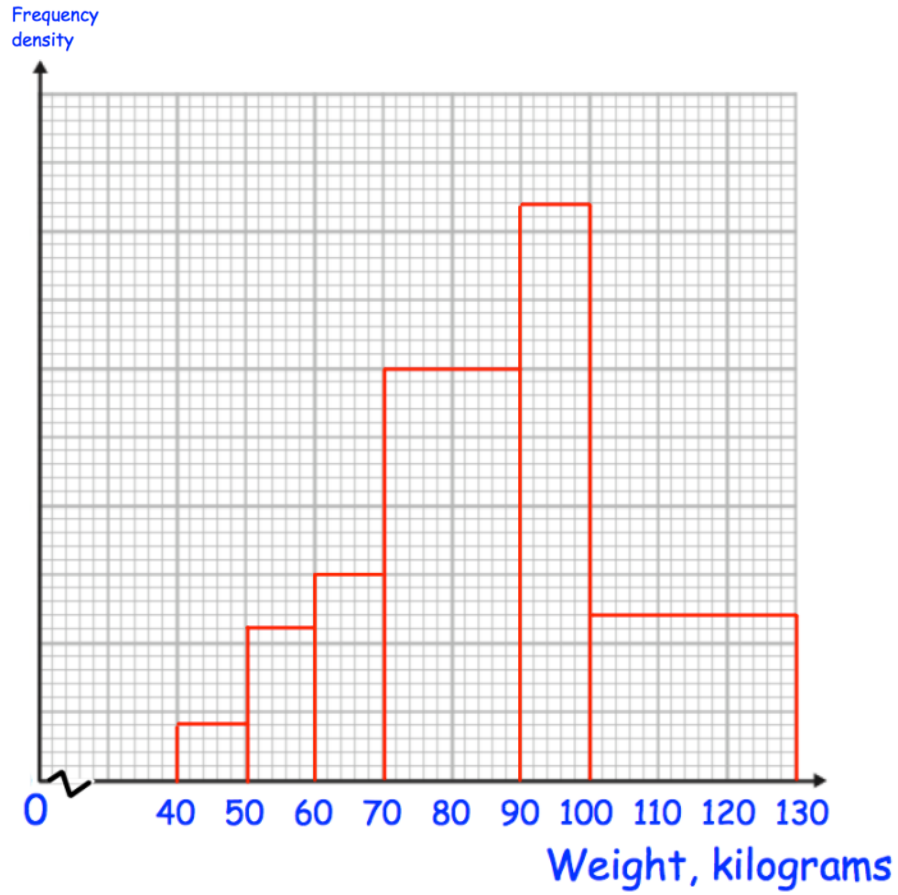
Worked example

There were 82 cars on the road.
14 cars were travelling over 50 mph.
Estimate the number of cars that were travelling between 40 and 49 mph.



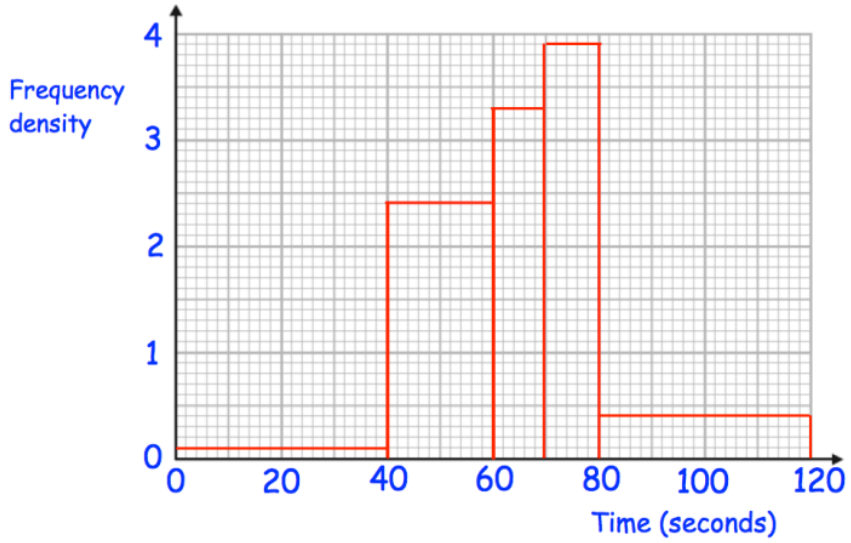
Your turn

There were 504 athletes measured.
45 athletes weigh under 60 kg.
Estimate the number of athletes between 70 and 95 kg.



Worked example

Estimate the median time



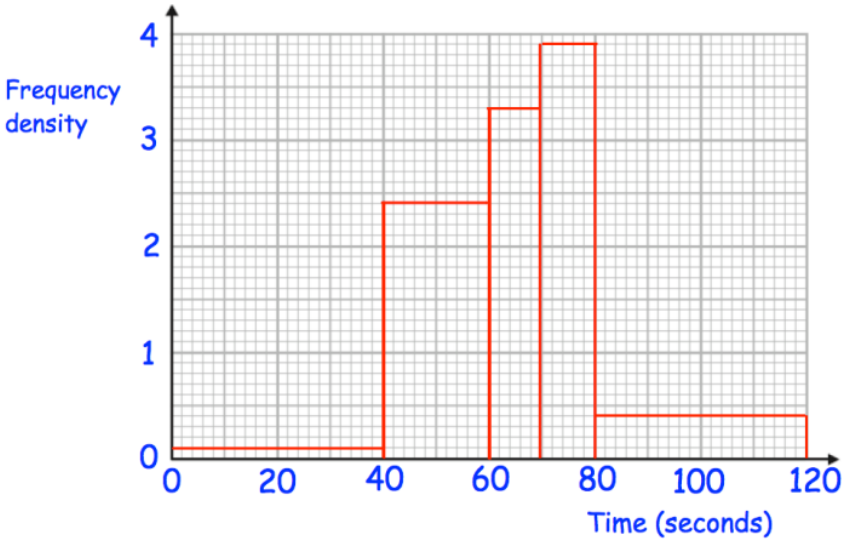
Your turn

Estimate the median weight

14.84 (2 dp)

Worked example

A participant is chosen at random.
What is the probability they took longer than 60 seconds?



Your turn

A participant is chosen at random.
What is the probability they weigh more than $14\ kg$?

$$\frac{14}{25} = 0.56$$

3.5) Comparing data

Chapter CONTENTS

Worked example

From the large data set, the daily mean temperature during June 1987 is recorded at Camborne and Leuchars.
For Camborne, $\sum x = 377.1$ and $\sum x^2 = 4939.45$
For Leuchars, the mean temperature was $10.9\ °C$ with a standard deviation of $2.10\ °C$.
Compare the data for the two locations.

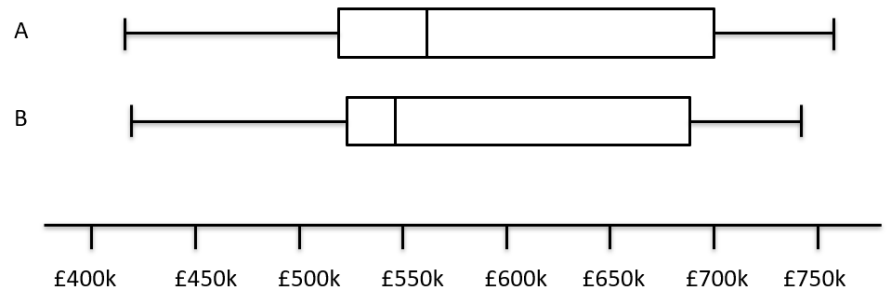
Your turn

From the large data set, the daily mean temperature during August 2015 is recorded at Heathrow and Leeming.
For Heathrow, $\sum x = 562.0$ and $\sum x^2 = 10301.2$
For Leeming, the mean temperature was $15.6\ °C$ with a standard deviation of $2.01\ °C$.
Compare the data for the two locations.

Mean daily temperature in Heathrow $= 18.1\ °C$
Standard deviation in Heathrow $= 1.91\ °C$ (3 sf)
The mean daily temperature in Leeming is lower than in Heathrow.
The spread of temperatures is greater in Leeming than in Heathrow.

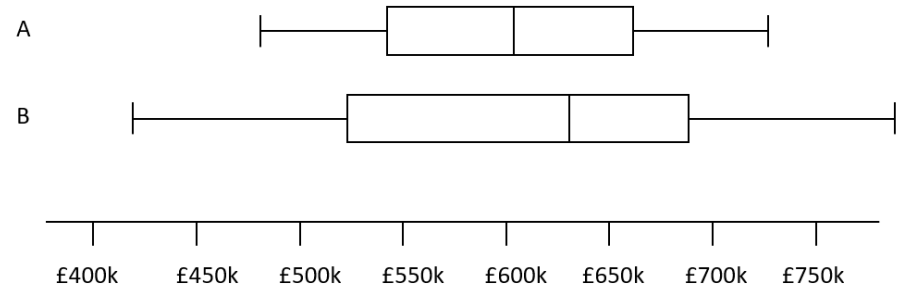
Worked example

Compare the house prices of locations A and B



Your turn

Compare the house prices of locations A and B



- The interquartile range of house prices in B is greater than A.
- The range of house prices in B is greater than A.
- The median house price in Kingston was greater than that in Croydon