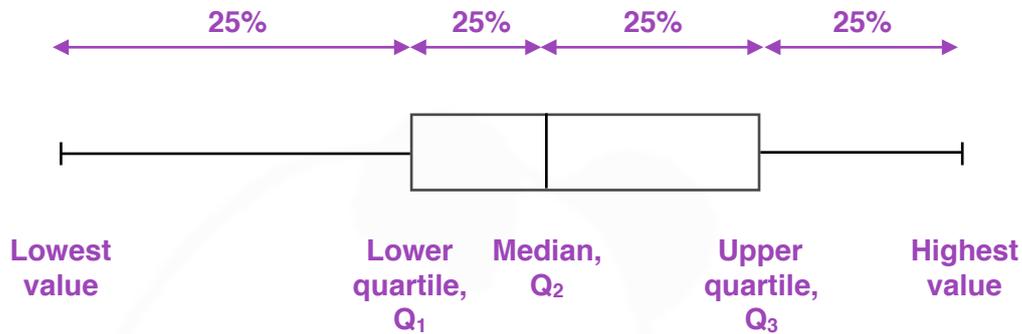## Box Plots

Notes

A box plot is a simple diagram that displays key information.

A box plot should be drawn accurately and the end points of the lines and boxes must have the numbers written on them. You might find it useful to draw a scale under your diagram.

N.B. If the data is grouped (i.e. $10 \leq h < 20$), the lowest and highest values are not known. In which case, the *highest and lowest possible* values are used
The mid-interval values are never used with cumulative frequency

E.g. 1 Draw a box plot for the heights of the set of plants. (E.g. 1 from cumulative frequency lesson).

Working: Median ≈ 23.1 cm   (line in the box)
Lower quartile, $Q_1 \approx 21.5$ (LHS of the box)
Upper quartile, $Q_3 \approx 25$ (RHS of the box)
The first interval is $15 \leq h < 18$ so the lowest possible value is 15 — this value becomes the start of the line
The final interval is $27 \leq h < 30$ so the highest possible value is 30 — this value becomes the end of the line
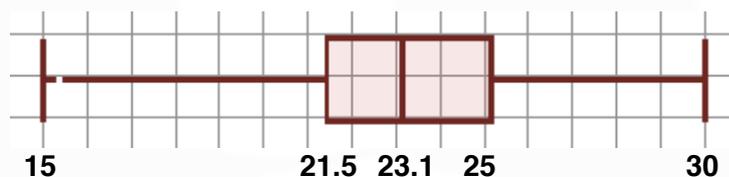
E.g. 2 Draw a box plot for the marks of the students in the maths test. (E.g. 1 from cumulative frequency lesson).

Working: Median ≈ 59 cm   (line in the box)
Lower quartile, $Q_1 \approx 50$ (LHS of the box)
Upper quartile, $Q_3 \approx 66$ (RHS of the box)
The first interval is $0 \leq h < 10$ so the lowest possible value is 0 — this value becomes the start of the line
The final interval is $90 \leq h < 100$ so the highest possible value is 100 — this value becomes the end of the line

Comparing data

When comparing data we need to make two comments

1. State which group has the highest *central tendency* — i.e. the highest *median* (or mean)
2. State which group has the least *spread* — so use the *interquartile range* (or range if the interquartile ranges are equal)

Higher median ⇒ *higher average* — this is usually a good thing (e.g. test scores) but could also be bad (e.g. golf scores — the least number number of shots is best)

Lower IQR (or range) ⇒ *lower spread* — this is usually a good thing because it means greater consistency

N.B. You must use numbers to support your statements

E.g. 3 Compare the following data for the English test scores of two classes:

Class	Lowest score	Lower quartile	Median	Upper quartile	Highest score
10X	36	53	65	74	93
10Y	38	48	70	76	87

Working: *1st comment = central tendency*
Class 10Y did better on the English because their median is better than the one of 10X (70 vs 65)
2nd comment = spread
However, class 10X's results were more consistent because their interquartile range is lower than 10Y's (21 vs. 28)

Video: **Drawing and reading box plots**

Solutions to Starter and E.g.s

E.g. 4 Do example on p483 in comparing data.
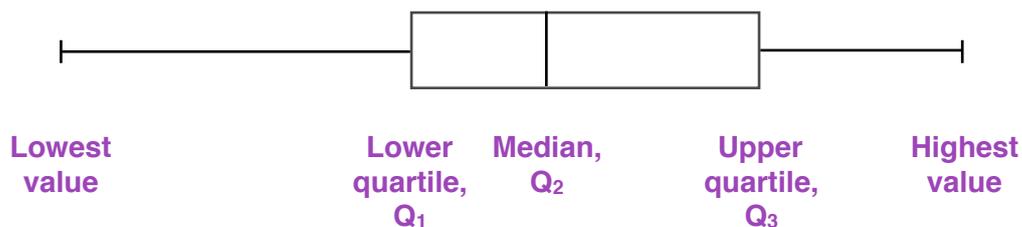Further examples
CGP p 349 Ex 2 Qu 1-2

Exercise

9-1 class textbook:	p483 E14.3 Qu 1-6
A*-G class textbook:	p439 M14.8 Qu 1-6
9-1 homework book:	p166 E14.3 Qu 1-5
A*-G homework book:	p123 M14.8 Qu 1-5

Summary

Lowest value	**Lower quartile, Q₁**　**Median, Q₂**	**Upper quartile, Q₃**　**Highest value**

A box plot should be drawn accurately with numbers written on the the end points of the lines and boxes.
For grouped data (i.e. $10 \leq h < 20$), the *highest and lowest possible* values are used as the end points.

Comparing data:
Two comments are needed, with numbers to support the statements:
1. State which group has the highest *central tendency* — i.e. the highest *median* (or mean)
2. State which group has the least *spread* — so use the *interquartile range* (or range if the interquartile ranges are equal)

Higher median ⇒ higher average — this is usually a good thing (e.g. test scores) but could also be bad (e.g. golf scores — the least number number of shots is best)

Lower IQR (or range) ⇒ lower spread — this is usually a good thing because it means greater consistency