Quick Find

Search

Stems, Leaves, Boxes and Whiskers

- Representation of Data

    -
    - Back to Representation of Data
    - Histograms and cumulative frequency

Revise quicker now!

It is a very good idea to make sure you are fully competent with the Topic **The Basics** before revising this section.

There are many ways to present data. In your GCSE course you will have already met plenty of these. For example, there are bar charts, pie charts, frequency diagrams, frequency polygons, stem-and-leaf diagrams, cumulative frequency diagrams and histograms.

These are basic methods but still should be remembered for AS or A2 level. Instead of being told which diagram to draw at this level, you will be expected to **choose** the diagram you think best fits the data.

Stem-and-leaf diagrams

A stem-and-leaf diagram is one way of taking raw data and presenting it in an easy and quick way to help spot patterns in the spread of data.

A stem-and-leaf diagram is one way of taking raw data and presenting it inan easy and quick way to help spot patterns in the spread of data.They are best used when we have a relatively **small** set of data andwant to find the **median** or **quartiles**.

Example:

The following raw data was taken from the speeds (in mph) of 30 cars passing a particular point on a quiet country road:

32 24 35 56 22 29 35 44 43 52
22 35 30 44 42 36 34 38 42 39
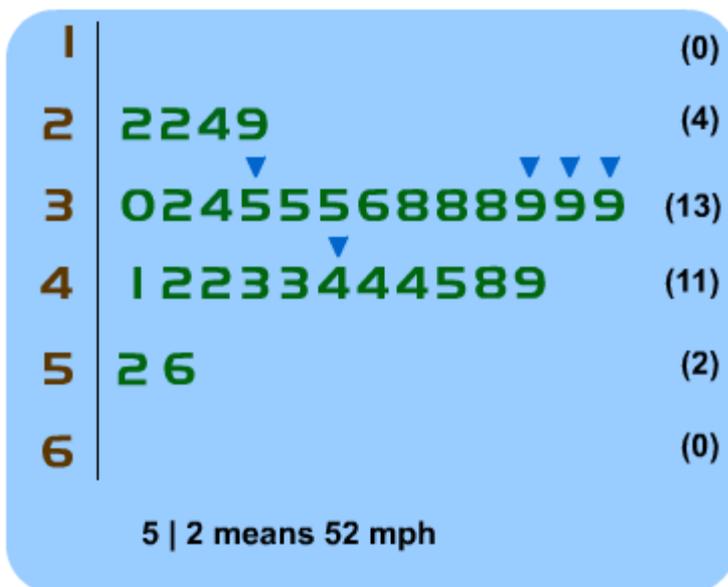49 41 44 45 39 38 43 38 48 39

First we start with the stem - in this case, we use the '**ten**' digits from the data given...
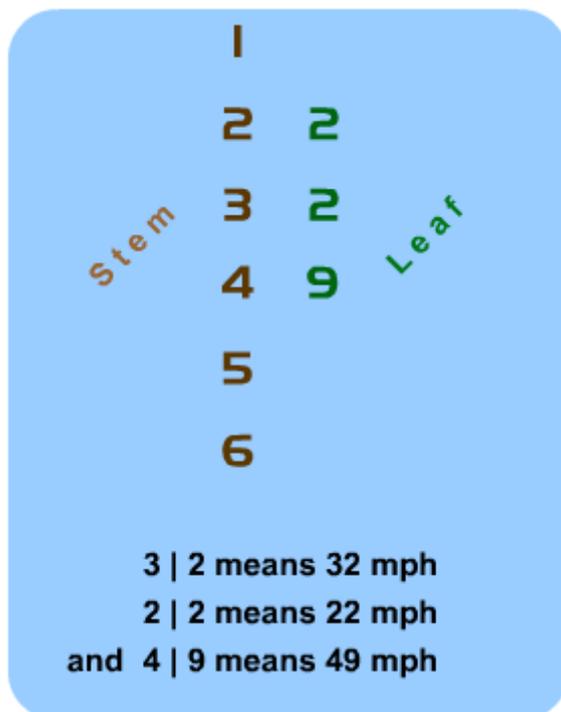
We then fill in the raw data one piece at a time to get the leaves of the diagram. Reading down from our raw data starting with 32 then 22 then 49 etc...until all 30 pieces of data are included in the diagram.

```
1 |
2 | 2429
3 | 2550968548899
4 | 91445234328
5 | 62
6 |

3 | 2 means 32 mph
```
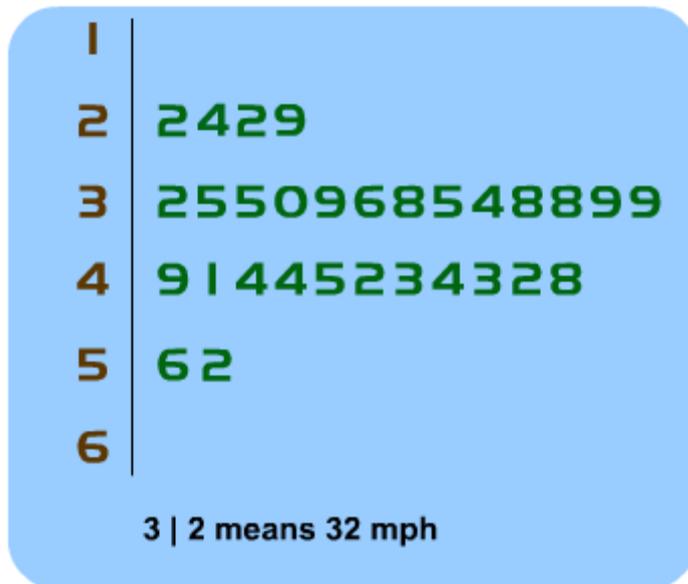
However, it is best to order our diagram by re-writing it with the leaves in numerical order, shown below:

```
1 |                        (0)
2 | 2249                   (4)
3 | 0245556888999         (13)
4 | 12233444589           (11)
5 | 26                     (2)
6 |                        (0)

5 | 2 means 52 mph
```

The numbers in brackets at the end of each leaf tells us the number of leaves belonging to each stem.

Note the similarity to a bar chart (just rotate 90^0 anti-clockwise).

Once we have completed our diagram we can then analyse the data.

If you're unsure how this is done you may need to refer to the Topic *The Basics*.

As the data is now in numerical order it will be easy to calculate the median value and inter-quartile range.

Median: There are 30 pieces of data so our median will be halfway between the 15th and 16th pieces of data. Counting along the leaves.

The 15th is 39 and the 16th is 39 shown ringed on the diagram.

Hence the median is 39mph.

Quartiles: Lower quartile is 35mph
Upper quartile is 44mph
Hence: Inter-quartile range = 44 - 35 = 9mph

Box-and-whisker plots (or boxplots)

These are very basic diagrams used to highlight the quartiles and median to give a quick and clear way of presenting the spread of the data.

If you don't remember how to calculate any of these then please refer back to *The Basics*.

There are 5 points to remember when drawing a Box-and-whisker plot.

They are:

1. The 'box' part is drawn from the lower quartile to the upper quartile. The median is then drawn within the box.
2. The 'box' shows the inter-quartile range, which houses the middle half of the data.
3. The 'whiskers' are then drawn to the lowest and highest values of the data.

Boxplots can be drawn either vertically (as above) or horizontally.

Example:

A cricketer
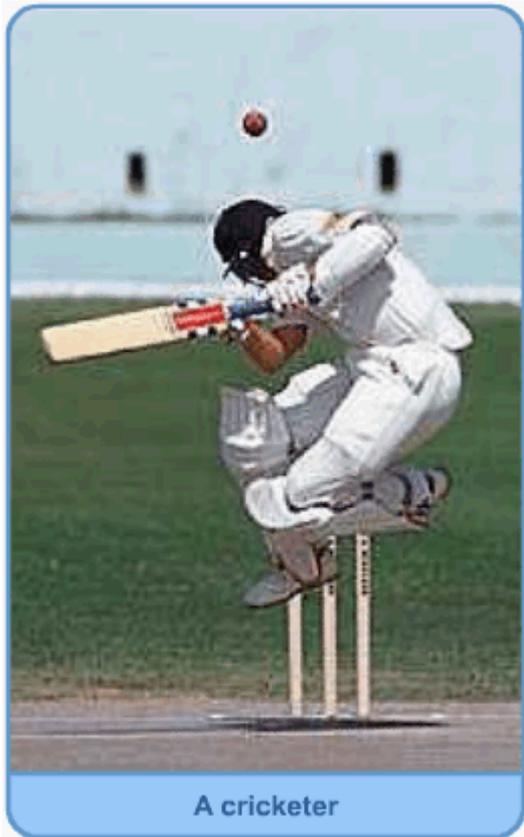
The last 9 scores of a certain English cricket player were:

0 1 3 4 7 9 10 12 15

Before drawing our boxplot we need to calculate the median and quartiles.

Quartiles: lower quartile = 2
upper quartile = 11
highest value = 15

Median = 7

Lowest value = 0

**Hence our boxplot looks like this:**

Like stem-and-leaf diagrams, these diagrams are best used when we have a relatively small set of data or when the data has already been summarised into its median and quartiles.

Skewness

One of the main reasons for drawing a stem-and-leaf or box-and-whisker is to spot quickly any trend in the spread of the data.
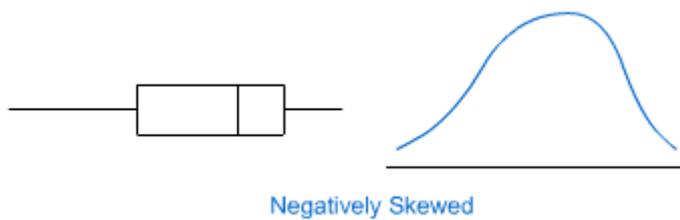
**There are 3 types of distribution we can get:**

1. Symmetrical distribution:
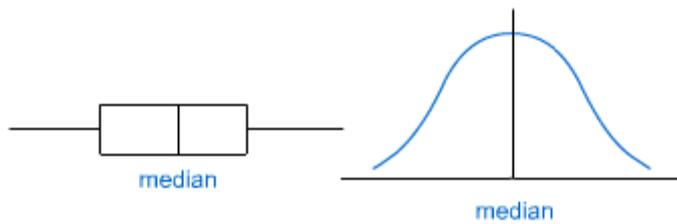
median

median

The median is in the middle exactly halfway from each quartile. This is what we would class as a **normal distribution.**

2. Negatively skewed distribution:

Negatively Skewed

There is a greater proportion of the data at the upper end.

3. Positively skewed distribution:

Positively Skewed
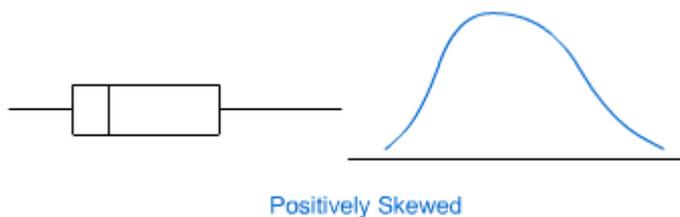
There is a greater proportion of the data at the lower end.

Outliers

Outliers are anomalies of our distribution. These are the values of data that are either a lot larger or smaller than the majority of the rest of the data.

Values of data are usually labelled as outliers if they are more than 1.5 times of the inter-quartile range from either quartile.

Example:

Taking the last 9 scores of a certain English cricket player again of:

Although I used 1.5 as my multiplying factor, this isn't always the case. However, if it differs from this you will be told the multiplying factor within the exam question. 1.5 is most commonly used though!

S-cool 52 Revision Secrets!

Want an easy way to improve your grades? Get 52 proven exam and revision strategies to give you a winning edge in your exams.

[Revise quicker now!](#)

-
- [Back to Representation of Data](#)
- [Histograms and cumulative frequency](#)

Clearing 2014

Email: * [_____]

[Submit]

Revise quicker

- Get revision guides
- Get question banks
- Ask questions
- Make a timetable