# *Quartiles*

For a more detailed description of the spread, we divide the dataset into smaller parts.

**Percentiles**: divide the dataset into 100 parts, to score in the 90$^{th}$ percentile, indicates 90% of the scores were less than or equal to your score i.e. you scored in the top 10%.

**Deciles**: divide the dataset into 10 parts

**Quartiles**: divide the dataset into 4 parts

**First Quartile, $Q_1$ (lower quartile)**

Another name for the 25$^{th}$ percentile. The first quartile divides the ordered data such that 25% of the scores are at or below this value.

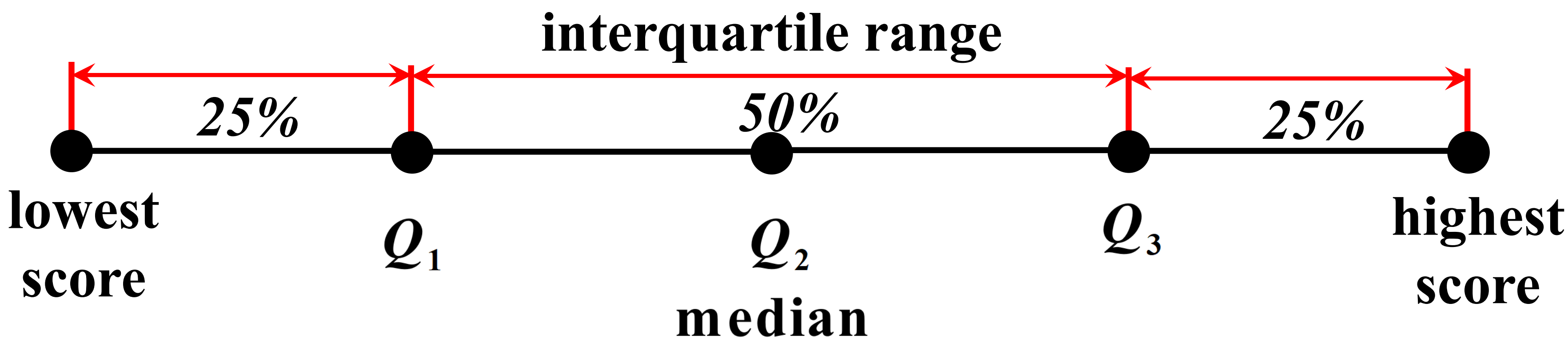
**Second Quartile, $Q_2$ (median)**

**Third Quartile, $Q_3$ (upper quartile)**

Another name for the 75$^{th}$ percentile. The third quartile divides the ordered data such that 75% of the scores are at or below this value.

# *Interquartile Range*

The interquartile range is a measure of spread that covers 50% of the data




$$\boxed{IQR = Q_3 - Q_1}$$

**Calculating the interquartile range – odd number of scores**

1    $\boxed{2 \quad 2}$    3    5    7    $\boxed{8 \quad 9}$    10

$\boldsymbol{Q_1 = 2}$    $\boldsymbol{Q_2}$    $\boldsymbol{Q_3 = 8.5}$

1. Omit the median
2. $Q_1$ is the median of the left hand side
3. $Q_3$ is the median of the right hand side

$$IQR = Q_3 - Q_1$$
$$= 6.5$$

## Calculating the interquartile range – even number of scores

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 6 | 7 | 8 | 9 | 9 | 10 | 10 | 11 | 13 | 15 |

$Q_1$ $\qquad$ $Q_2 = 9.5$ $\qquad$ $Q_3$

1. Separate the data into two halves
2. $Q_1$ is the median of the left hand side
3. $Q_3$ is the median of the right hand side

$$IQR = Q_3 - Q_1$$
$$= 11 - 8$$
$$= 3$$

## Five-number summary

$\boxed{\text{minimum}\ ,\ Q_1\ ,\ \text{median}\ ,\ Q_3\ ,\ \text{maximum}}$

The quartiles along with the minimum and maximum score make up the five-number summary.

e.g. the five-number summaries for our two examples would be

1 , 2 , 5 , 8.5 , 10

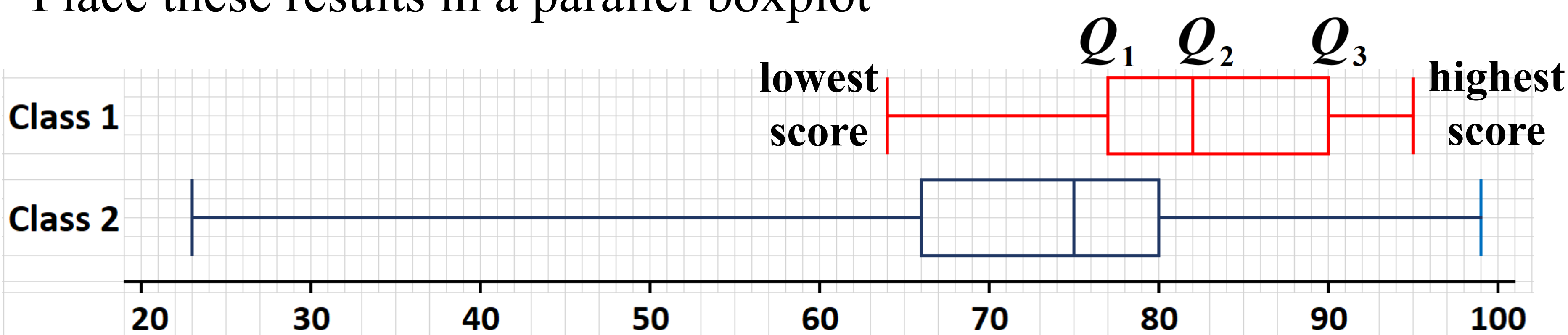
and 6 , 8 , 9.5 , 11 , 15

# Box-and-whisker Plots

A box-and-whisker plot (box plot) is a graphical way of displaying data using the five-number summary.
These plots are useful for comparing different sets of data in a parallel box plot.
e.g. the results of a major exam for two particular classes are given below

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Class 1: | 64 | 66 | 68 | 70 | 75 | 77 | 78 | 79 | 81 | 82 | 82 |
| | 82 | 85 | 89 | 89 | 90 | 93 | 93 | 93 | 94 | 94 | |
| Class 2: | 23 | 37 | 60 | 64 | 65 | 66 | 70 | 71 | 72 | 72 | 75 |
| | 75 | 76 | 76 | 78 | 80 | 80 | 90 | 92 | 95 | 99 | |

Place these results in a parallel boxplot



# Outliers

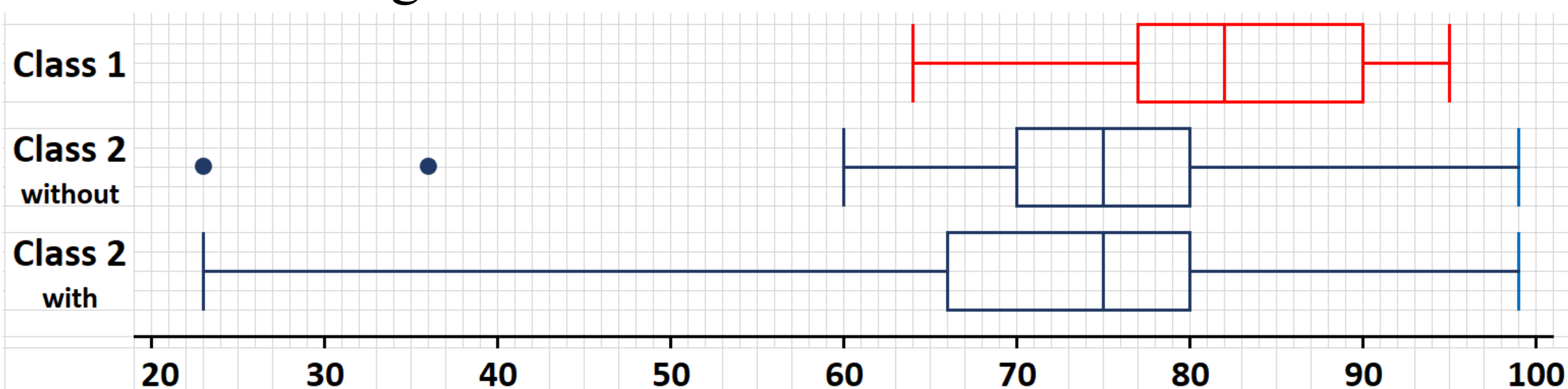
An outlier is an unusual observation. It lies at an abnormal distance from the rest of the data.

A common practice for identifying outliers is;

$$\boxed{\text{outlier} < Q_1 - 1.5 \times IQR \ \text{ or } \ > Q_3 + 1.5 \times IQR}$$

Careful attention must be paid to outliers and their inclusion/exclusion, due to the influence on the shape of the distribution and their effect on the value of other statistics

If outliers are excluded from the dataset, their existence should always be acknowledged




$\bar{x} = 82.1$
$s = 9.37$

$\bar{x} = 76.6$
$s = 10.48$

$\bar{x} = 72.1$
$s = 17.07$

**Exercise 15C; 1bd, 2bdfg, 3bdfh, 4, 5, 6, 7, 8**