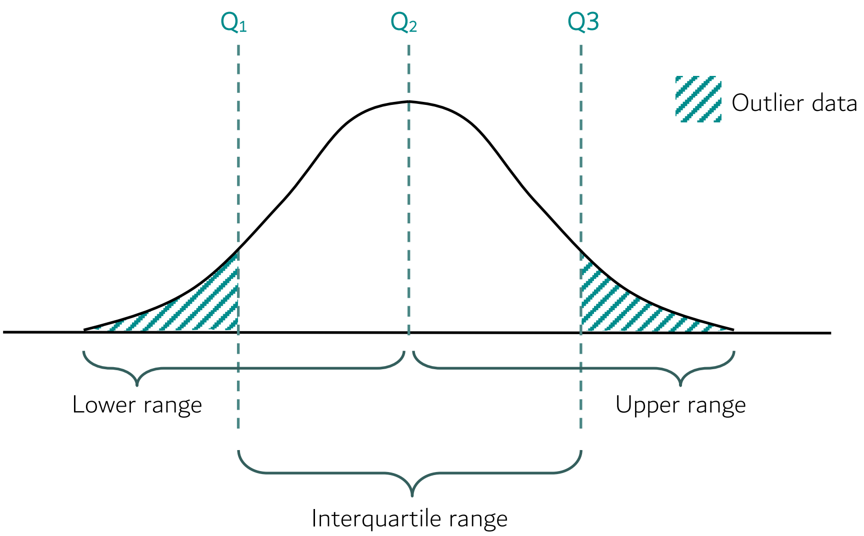
Interquartile Ranges

The Interquartile Range of some data is a measure of its spread or dispersion when placed in numerical order. It works by removing outlying data (that which lies at either extreme of the range) and is therefore viewed as a truer representation of the spread.



Outlying data is seen as that which sits below the lower quartile and above the upper quartile. These values are based on the median values of different ranges. The median value is the middle value when the data is placed in numerical order.

Q_1 (the lower quartile) is the median value of the lower range.

Q_2 is the median value of the whole range of data.

Q_3 (the upper quartile) is the median value of the upper range.

As a single figure, the interquartile range means little so, in general, two interquartile ranges are compared. There are many examples of geographical research which could look at the interquartile range of a spread of data:

- Demography - the **age range of a population** in a particular year.
- Weather - the spread of **maximum temperature data** over a month.
- Tourism - the **distance people are willing to travel** to see a particular site.
- Coasts - the **size of beach sediment** in a certain location.
- Settlements - the range of **favourability scores** around a regeneration site.
- Ecology - the spread of **carbon content** amongst trees in an area of woodland.
- Transport - the **volume of traffic** during different time periods of the day.
- Development - the range of **GDP values** observed in a region or set of countries.

How to carry out an interquartile range calculation:

For this example, a geographical researcher wishes to compare the environmental quality survey (EQS) scores given by members of the public to two different sites - one being a housing development from the 1960s and the other from a development built in the late 2000s. The researcher feels that the uniformity of the more recent development will have created a smaller range of EQS scores than the 1960s development which is more varied in character. This created the following hypothesis:

> "The housing development from the 1960s will have a larger interquartile range than that for the housing development from the 2000s."

The researcher started by placing the EQS scores in numerical order for each housing development (going from smallest to largest).

1960s	5	6	11	12	14	14	18	19	19	24	26	27	27	27	31	32	32	35	39	40	41	43	44	45	46
2000s	5	5	9	13	13	15	17	17	21	25	26	29	32	33	35	35	38	38	39	39	40	42	42	42	43

This allowed them to calculate the median (Q_2) value for each housing development. In this example the number of values is odd, therefore the Q_2 figure will be an actual value. If the total number of values is even, the Q_2 figure will lie between two real values.

1960s	5	6	11	12	14	14	18	19	19	24	26	27	**27**	27	31	32	32	35	39	40	41	43	44	45	46
2000s	5	5	9	13	13	15	17	17	21	25	26	29	**32**	33	35	35	38	38	39	39	40	42	42	42	43

The lower and upper ranges are then isolated and the each quartile is calculated by taking the number of values in that range (n) and adding 1 before dividing by 2. The result is the **position** of the Q_1 or Q_3 value in the range of data.

$$Q_1 \text{ or } Q_3 = \frac{n + 1}{2}$$

$$= \frac{13}{2}$$

$$= 6.5^{th} \text{ position}$$

We look for the number value that occupies the equivalent of the 6.5^{th} position in the ordered data for the lower range. In this case, it lies between two values

1960s	5	6	11	12	14	14	18	19	19	24	26	27
2000s	5	5	9	13	13	15	17	17	21	25	26	29

6.5^{th} position

Q_1 (1960s) = 16

Q_1 (2000s) = 16

We then repeat the exercise for the ordered data in the upper range.

1960s	27	31	32	32	35	39	40	41	43	44	45	46
2000s	33	35	35	38	38	39	39	40	42	42	42	43

6.5^{th} position

Q_1 (1960s) = 39.5

Q_1 (2000s) = 39

The interquartile range (IQR) for each development is calculated by subtracting the Q_1 from the Q_3.

IQR (1960s) = $Q_3 - Q_1$

= 39.5 - 16

= 23.5

IQR (2000s) = $Q_3 - Q_1$

= 39 - 16

= 23

In this case, it can be seen that the interquartile range for the environmental quality survey scores are very similar for both age housing developments, though the 1960s development has a marginally larger range. Therefore, the researcher can accept their hypothesis but should acknowledge that the difference between the two sites is marginal.

