



Comparing box-and-whisker graphs (CSSI)

Key focus

This objective introduces the CSSI framework for comparing distributions shown as box-and-whisker plots: *Centre* (median, and mean where relevant), *Shape* (left skew, right skew, or symmetric/bell curve), *Spread* (interquartile range and range), and *Interesting* features (outliers, groups, or other unusual points). Students learn to read the five-number summary (minimum, lower quartile, median, upper quartile, maximum) directly off a box plot, then use all four CSSI components together to compare two or more distributions and justify conclusions with reference to specific values rather than visual impressions alone (e.g. “the box is higher” is not, by itself, a full comparison).

Suggested teaching sequence

1. Introduce the acronym CSSI and define each part before showing any plots: Centre (median/mean), Shape (skew/symmetric), Spread (IQR/range), Interesting (outliers/groups).
2. Show a single box plot with the five-number summary labelled. Identify each part: minimum and maximum (whisker ends), lower quartile and upper quartile (box edges), median (line inside the box).
3. Model reading the centre (the median line) and calculating the spread: $IQR = \text{upper quartile} - \text{lower quartile}$; $\text{range} = \text{maximum} - \text{minimum}$.
4. Introduce shape: a box positioned with a long whisker/box section on one side compared to the other indicates skew (right skew: longer/larger upper section; left skew: longer/larger lower section); roughly equal sections on both sides of the median suggest a symmetric distribution.
5. Compare two box plots on the same scale. Model a full comparison: state which has the higher/lower median (centre), which has the greater/lesser spread and what that means (more/less variability), and how their shapes compare.
6. Introduce outliers as points drawn separately beyond the whiskers, and discuss what an outlier might mean for the data (an unusual case, or possibly an error).
7. Practise comparing three or more distributions together, prioritising which comparisons matter most for a given question (e.g. centre for “who did better on average,” spread for “who was more consistent”).
8. Discuss why a full CSSI comparison is more informative than comparing a single statistic (e.g. just the median, or just the mean) alone.

Core examples to model

- Reading a box plot: $\text{min} = 10$, $Q1 = 20$, $\text{median} = 30$, $Q3 = 45$, $\text{max} = 55$. $IQR = 45 - 20 = 25$. $\text{Range} = 55 - 10 = 45$.
- Comparing centre and spread: Class A (median 30, IQR 25) vs. Class B (median 42, IQR 15) – Class B scored higher on average and was more consistent (smaller spread).
- Shape: a box with $\text{min} = 5$, $Q1 = 10$, $\text{median} = 15$, $Q3 = 25$, $\text{max} = 50$ has a much longer upper

section – this is right-skewed, meaning a few unusually high values pull the tail to the right.

- Outlier: a box plot with a ringed point well beyond the upper whisker is an outlier – describe it separately from the main pattern (e.g. “one student scored much higher than the rest of the class”).

Watch for

- Comparing only the box positions (“Class A’s box is higher”) without stating centre, spread, and shape together – a full CSSI comparison always covers more than one component.
- Confusing IQR with range: IQR uses only the middle 50% of the data ($Q3 - Q1$) and is less affected by outliers; range uses the full min-to-max span and can be distorted by a single extreme value.
- Misreading skew direction: right skew means the longer tail/section is on the *right* (higher values), not that more data points sit on the right.
- Treating an outlier as invalid or excludable by default: an outlier should be reported and discussed, not automatically removed – it may be genuine and important.
- Drawing conclusions from centre alone (e.g. “Class A did better”) without checking spread and shape, which can reveal that one class was far more variable or skewed even with a similar median.

Checks for understanding

- Can students state what each letter of CSSI stands for without prompting?
- Can they read the five-number summary and calculate IQR and range from a box plot?
- Can they identify a distribution’s shape (left skew, right skew, symmetric) and justify it using the relative size of the box/whisker sections?
- Can they identify outliers on a box plot and discuss what they might mean?
- Can they compare two or more box plots using centre, shape, spread, and interesting features together?
- Can they explain why comparing distributions fully (not just one statistic) gives a more complete picture?