Modified box plots

One value far from the rest stretches a box plot's whisker and makes it look like many values live out there. A modified box plot shows that value as a dot of its own.

WORDS TO KNOW

Five-number summary: the minimum, Q_1, the median, Q_3 and the maximum.

IQR (interquartile range): $Q_3 - Q_1$, the width of the box.

Outlier: a value strictly beyond a fence.

Quartiles: Q_1 and Q_3 are the medians of the lower and upper halves of the ordered data. With an odd count, leave the median out of both halves.

Fences: the two cut-off lines, 1.5 IQR below the box and 1.5 IQR above it.

Modified box plot: a box plot that draws each outlier as a dot and stops the whiskers before them.

KEY IDEAS

1. The fences come from the box

The box is the same as in any box plot. Only the whiskers change. The fences are measured from the quartiles, not from the median, and they are worked out but not drawn.

$$\text{lower fence} = Q_1 - 1.5\,\text{IQR} \qquad \text{upper fence} = Q_3 + 1.5\,\text{IQR}$$

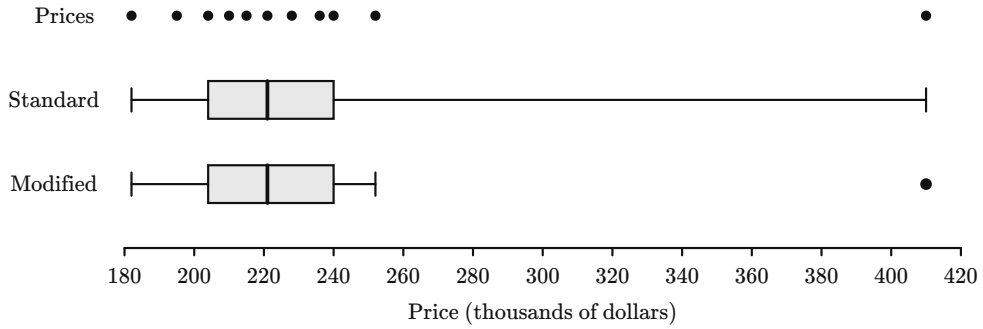
A value **strictly** below the lower fence or **strictly** above the upper fence is an outlier. A value exactly on a fence is not.

2. Dots for outliers, whiskers stop at real values

Each outlier is a dot at its value. Each dot is one value.

A whisker runs from the box to the most extreme value that is **not** an outlier. It stops at that data value, not at the fence.

The picture shows the same house prices both ways. In the standard plot, one house at 410 makes a long whisker that looks like many houses. In the modified plot, that house is a dot. If there are no outliers, the modified plot is the same as the standard one.



3. Reading one back

From the plot you can read Q_1, the median, Q_3, every outlier and the two whisker ends. Between a whisker's end and the next dot there are **no** data values: a value there would either make the whisker reach it or be a dot itself. The plot does not show how many values there are in all.

Free sample from math.wiki for classroom use. Do not post. math.wiki

4. An outlier is a flag, not a mistake

The rule only says "look here". If the record shows a typo, correct the value and find the fences again. If the value is real, it stays in the data and stays a dot.

WATCH OUT

- **Measure the fences from Q_1 and Q_3,** not from the median.
- **A value on a fence is not an outlier.** It must be strictly beyond.
- **Never draw a whisker to the fence.** That would show values that do not exist.
- **Do not delete an outlier just because it is far out.** Only a value shown to be wrong is corrected.

WORKED EXAMPLES

Example 1. Ten phone batteries lasted these numbers of hours: 21, 14, 35, 2, 18, 24, 15, 20, 17, 23. Find the fences, the outliers and where the whiskers end.

1. Put the values in order: 2, 14, 15, 17, 18, 20, 21, 23, 24, 35.
2. With 10 values, the median is the mean of the 5th and 6th values, 18 and 20, which is 19.
3. The lower half 2, 14, 15, 17, 18 gives $Q_1 = 15$. The upper half 20, 21, 23, 24, 35 gives $Q_3 = 23$.
4. IQR = 23 − 15 = 8, and 1.5 × 8 = 12.
$$\text{lower fence} = 15 - 12 = 3 \qquad \text{upper fence} = 23 + 12 = 35$$
5. 2 is below 3, so it is an outlier and gets a dot. 35 sits exactly on the upper fence, so it is not an outlier.
6. The lower whisker stops at the smallest value that is not an outlier, 14. The upper whisker reaches the maximum, 35.

Answer: fences 3 and 35; one outlier, 2; whiskers end at 14 and 35.

Example 2. A modified box plot has a box from 30 to 46 with median 38, whiskers ending at 22 and 60, and dots at 4 and 78. Check that both dots are outliers. Could a value be 65? How many values are above 60?

1. IQR = 46 − 30 = 16, and 1.5 × 16 = 24.
$$\text{lower fence} = 30 - 24 = 6 \qquad \text{upper fence} = 46 + 24 = 70$$
2. 4 < 6 and 78 > 70, so both dots are outliers.
3. 65 is inside the upper fence, so the upper whisker would reach it. But the whisker stops at 60, so no value is 65.
4. Every value above 60 must be a dot, and there is one dot up there.

Answer: both dots are outliers (fences 6 and 70); no value can be 65; exactly one value is above 60, the 78.

Example 3. Seven quiz scores were typed in as 12, 15, 16, 18, 19, 21, 190. The paper record shows the 190 is really 19. Find the outliers before and after the fix.

1. Before: the median is the 4th value, 18. The lower half 12, 15, 16 gives $Q_1 = 15$, and the upper half 19, 21, 190 gives $Q_3 = 21$.
2. IQR = 6 and 1.5 × 6 = 9, so the fences are 15 − 9 = 6 and 21 + 9 = 30. The 190 is an outlier.
3. The rule flags it; the record shows it is a typo. Correct it, do not just drop it. In order: 12, 15, 16, 18, 19, 19, 21.
4. After: the median is still 18, $Q_1 = 15$ and $Q_3 = 19$. IQR = 4 and 1.5 × 4 = 6, so the fences are 9 and 25.

Answer: before the fix, 190 is an outlier; after it, every score is between 9 and 25, so there are no outliers.

Free sample from math.wiki for classroom use. Do not post. math.wiki