How to make a boxplot

BHCC Mat-181

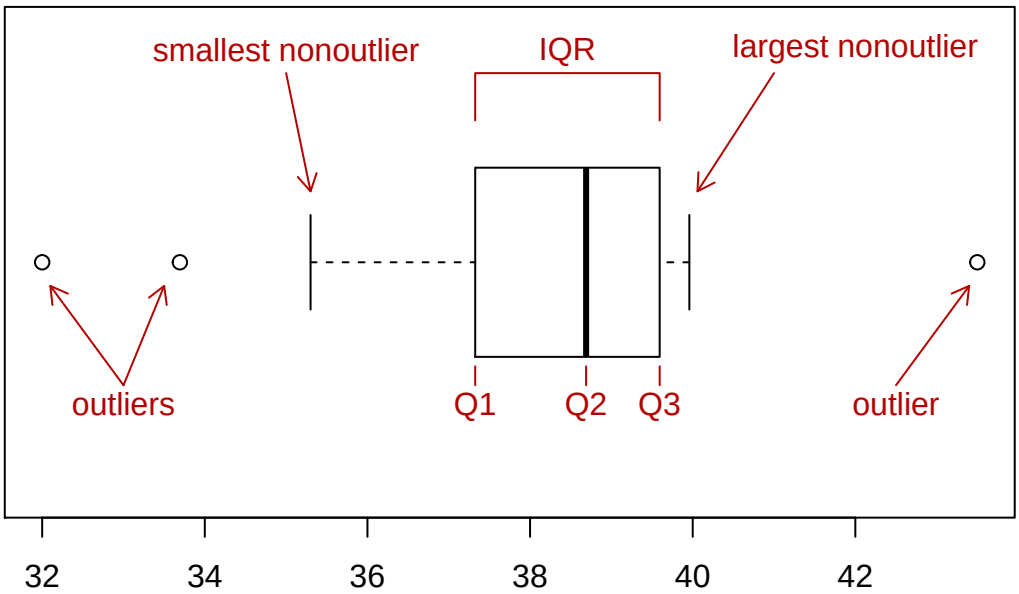
Quick overview of a boxplot

A boxplot's main feature is a box that indicates the values at the 25th, 50th, and 75th percentiles. The value at the 25th percentile is called Q1, the value at the 50th percentile is M or Q2, and the value at the 75th percentile is Q3.

The inter-quartile range (IQR) is the difference between Q3 and Q1. This IQR is a robust **measure of spread**.

Using IQR, we define outliers as values smaller than $(\text{Q1} - 1.5 \times \text{IQR})$ and values larger than $(\text{Q3} + 1.5 \times \text{IQR})$. We indicate outliers with dots.

With the remaining (nonoutlier) values, we use whiskers to indicate the smallest and largest nonoutliers.



However, different definitions of percentile lead to different boxplots (but all methods converge as sample size gets large).

Our method: nearest-rank method

Let ℓ be percentile. Let i be an index. And let n be the sample size. In the nearest-rank method, we determine the index by **rounding up** the product of n and ℓ.

$$i = \lceil n\ell \rceil$$

For example, consider the following data.

140.4	140.53	141.34	141.56	142.06	142.99	146.87	147.71	151.14	152.32	168.17

The sample size, n, is 11. We determine the indeces and values of Q1, Q2, and Q3.

Quartile	Formula for i	i	x
Q1	$\lceil 0.25 \times 11 \rceil$	3	141.34
Q2	$\lceil 0.5 \times 11 \rceil$	6	142.99
Q3	$\lceil 0.75 \times 11 \rceil$	9	151.14

We determine the IQR.

$$\begin{aligned} \text{IQR} &= \text{Q3} - \text{Q1} \\ &= 151.14 - 141.34 \\ &= 9.8 \end{aligned}$$

We determine the outlier boundaries.

$$\begin{aligned} \text{lower boundary} &= \text{Q1} - 1.5 \times \text{IQR} \\ &= 141.34 - 1.5 \times 9.8 \\ &= 126.64 \end{aligned}$$

$$\begin{aligned} \text{upper boundary} &= \text{Q3} + 1.5 \times \text{IQR} \\ &= 151.14 + 1.5 \times 9.8 \\ &= 165.84 \end{aligned}$$

We see there is one value outside the boundaries.

$$\text{outliers} = \{168.17\}$$

We can identify the ends of the whiskers: 140.4 and 152.32. We plot the boxplot.