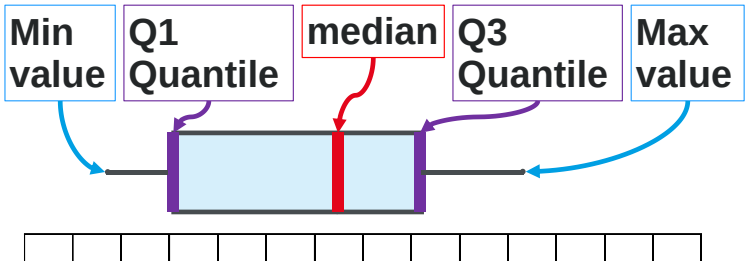
Box & Whiskers Plot
(statistical plot)

Problem
How to visualize a data set's variability?

Difficulty
Easy to use

- A **Box and Whisker Plot** (or **Box Plot**) graphically shows a data set's distribution by highlighting its median, quartiles, and outliers.
- A box plot helps to (1) understand a data set's central tendency, symmetry, spread, and skewness, (2) compare multiple datasets.
- A box plot can be drawn horizontally or vertically.
- The "whiskers" can be defined in multiple ways – a description should accompany the plot.



Collected numerical data → **Creating a Box & Whiskers Plot** → A box and whisker plot

1. Arrange data in ascending order.
2. Calculate the following statistics
   A. Minimum (Q0), median (Q2), maximum (Q4)
   B. Q1 (first quartile) and Q3 (third quartile),
   C. IQR (inter-quartile range) = Q3 – Q1
3. Create a box from Q1 to Q3.
4. Within the box draw a line at the median value.
5. Draw "whiskers" from the box to the minimum and maximum values.
6. Alternatively
   A. Draw "whiskers" from the box to data values close to 1.5*IQR above and below the box.
   B. Plot any outliers (values beyond the whiskers) with a single mark.

Box & Whiskers Plot – Example – 6in6 views

- We counted the number of times each of eight different 6in6 presentations were accessed over a 24-hour period.
- The sorted values were {78, 88, 88, 89, 91, 92, 93, 95, 99}
- The Box & Whickers computations are below.

A. Determine min, max, and median values:

A. Min = Q0 = (0^{th} percentile) = 78
B. Max = Q4 = (100^{th} percentile) = 99
C. Median = Q2 = (50^{th} percentile) = middle value = 91

B. Determine quartiles

A. Q1 = (25^{th} percentile)
= Median of lower half (78, 88, 88, 89) = 88
B. Q3 = (75^{th} percentile)
= Median of upper half (92, 93, 95, 99) = 94
C. Compute IQR = Q3 – Q1 = 94 – 8 = 6

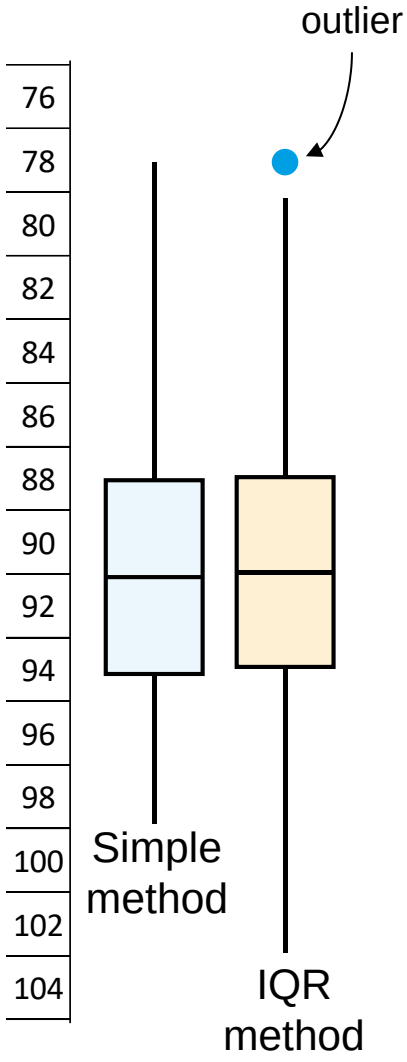
C. Define box: from Q1 to Q3 = from 88 to 94 (for both methods)

D. Define whiskers: use simple method

- Lower whisker = min = 78
- Upper whisker = max = 99
- No outliers: all data points are always within the whiskers

E. Define whiskers: use IQR method

- Lower whisker = Q1 – 1.5*IQR = 88 – 1.5*6 = 79
- Upper whisker = Q3 + 1.5*IQR = 94 + 1.5*6 = 103
- One outlier (i.e., 78 value) is outside of the whiskers



Box plots on the same data using two different methods

Box & Whiskers Plot – Notes

Slide 1

1. A data set can be categorized by its mean and variance. Those values, alone, do not indicate the presence of any asymmetries or extreme values, which a box plot can.
2. A box plot is a non-parametric technique and is robust for exploratory analysis. (It does not assume any specific underlying distribution such as normality.)
3. For the simple method, the plot only depends on 5 values: (median, Q1, Q3, min, max).
4. A box plot gives no insight into the data's underlying distribution; e.g., the data could be bimodal, multimodal, or uniform.
5. A box plot hides individual data points: except for outliers (using the IQR method), which can hide patterns or clusters.
6. A box plot can be misleading
   A. If there are few data points, since the quartiles might not be representative, and outliers can skew perception.
   B. The number of data points for each quartile is unknown; only the boundaries are given.
7. Many SW tools can determine box plots.

Slide 2

1. In this example
   A. the simple method shows there is an asymmetry between the smaller and larger values.
   B. the IQR method shows a single outlier, since one value is unusually small.

Recommended web sites for additional information

- https://datavizcatalogue.com/methods/box_plot.html
- https://asq.org/quality-resources/box-whisker-plot