A boxplot, also known as a box-and-whisker plot, is a graphical representation that displays the dispersion (spread) of a dataset. It shows the minimum, first quartile (Q1), median (Q2), third quartile (Q3), and maximum of the data. It also highlights outliers.

Example:

Suppose we have the following dataset representing the test scores of 10 students:

$$\text{Scores} = [55, 60, 65, 70, 75, 80, 85, 90, 92, 95]$$

Step 1: Organize the Data

First, arrange the data in ascending order (already done here):

$$[55, 60, 65, 70, 75, 80, 85, 90, 92, 95]$$

Step 2: Calculate the Quartiles

- **Q1 (First Quartile):** The median of the lower half of the data (excluding the overall median if the number of data points is odd).
- **Q2 (Median):** The middle value of the dataset.
- **Q3 (Third Quartile):** The median of the upper half of the data.

Calculating Q1:

- Lower half: $[55, 60, 65, 70, 75]$
- Median of lower half (Q1): 65 (middle value)

Calculating Q2 (Median):

- Full dataset: $[55, 60, 65, 70, 75, 80, 85, 90, 92, 95]$
- Median (Q2): $\frac{75+80}{2} = 77.5$

Calculating Q3:

- Upper half: $[80, 85, 90, 92, 95]$
- Median of upper half (Q3): 90 (middle value)

Step 3: Identify the Minimum and Maximum Values

- **Minimum:** 55
- **Maximum:** 95

Step 4: Calculate the Interquartile Range (IQR)

The IQR measures the spread of the middle 50% of the data:

$$\text{IQR} = Q3 - Q1 = 90 - 65 = 25$$

Step 5: Identify Outliers

Outliers are data points that fall below $Q1 - 1.5 \times \text{IQR}$ or above $Q3 + 1.5 \times \text{IQR}$.

- **Lower Bound for Outliers:**

  $$Q1 - 1.5 \times \text{IQR} = 65 - 1.5 \times 25 = 65 - 37.5 = 27.5$$

  (No data points below 27.5, so no lower outliers)

- **Upper Bound for Outliers:**

  $$Q3 + 1.5 \times \text{IQR} = 90 + 1.5 \times 25 = 90 + 37.5 = 127.5$$

  (No data points above 127.5, so no upper outliers)

Step 6: Create the Boxplot

The boxplot would look like this:

- **Minimum:** 55
- **Q1:** 65
- **Median (Q2):** 77.5
- **Q3:** 90
- **Maximum:** 95

Interpretation:

- **Box (Q1 to Q3):** Represents the middle 50% of the data (interquartile range). In this case, from 65 to 90.
- **Whiskers:** Extend from the minimum to Q1 and from Q3 to the maximum. In this case, from 55 to 65 and from 90 to 95.
- **Median Line:** The line inside the box represents the median (77.5).
- **Outliers:** None in this dataset.

+