

Create Back to Back Stem-and-Leaf Plots?

Authored by
stats writer

December 10, 2025

RECOMMENDED CITATION

stats writer (2025). *Create Back to Back Stem-and-Leaf Plots?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=107028>

Understanding the Stem-and-Leaf Plot Foundation

The **stem-and-leaf plot**, or stemplot, serves as a powerful descriptive statistical tool within Exploratory Data Analysis (EDA). It is a specialized graph used to display quantitative data in a format that preserves the individual data values while simultaneously providing a visual summary of the data's distribution. Unlike other grouping methods, such as histograms, which sacrifice individual precision for visual grouping, the stem-and-leaf plot allows analysts to see exactly which data points contribute to the shape of the overall distribution.

The underlying mechanism of this plot involves systematically partitioning each data point into two distinct components: the **stem** and the **leaf**. The stem typically comprises the leading digit or digits, representing the magnitude or class interval of the number. The leaf is always the final trailing digit, representing the specific precision of that measurement. This simple, elegant structure makes the stemplot exceptionally useful for summarizing small-to-medium-sized datasets where quick assessment of central tendency, spread, and shape is required.

For example, suppose we are analyzing the following dataset representing test scores from a class of students. The construction process requires defining the rules for splitting the values. If we define the first digit as the "stem" and the second digit as the "leaf," we can visually organize the raw numerical information into a coherent visual display. This organization instantly reveals where the data values are concentrated.

The Structure of a Standard Stem-and-Leaf Plot

To clearly illustrate the concept, let us use a predefined set of scores. Applying the rule (first digit = stem, second digit = leaf), we ensure that the leaves are always arranged in ascending numerical order, moving away from the stem. The stem values themselves are typically listed vertically in ascending order.

Consider the following dataset of values that need to be plotted:

Dataset: 12, 14, 18, 22, 22, 23, 25, 25, 28, 45, 47, 48

If we define the first digit in each value as the "stem" (1, 2, 3, 4) and the second digit as the "leaf," the resulting arrangement provides immediate insight into the grouping of the data points. Notice that there is no data in the 30s range, which results in a gap in the plot, indicating a specific characteristic of the distribution.

Stem	Leaf
1	2 4 8
2	2 2 3 5 5 8
3	
4	5 7 8

Key: 4|7 = 47

This resulting visual structure shows a heavy concentration of scores in the 20s range (Stem 2) and a relatively uniform spread in the teens and late 40s. The plot's primary advantage here is its dual function: it acts as both a frequency tally (the length of the leaf row indicates frequency) and a complete list of all data points.

Introducing the Back-to-Back Stem-and-Leaf Plot: A Comparative Tool

While the standard stem-and-leaf plot is excellent for univariate analysis, its true versatility shines through in its extension: the **back-to-back stem-and-leaf plot**. This specialized adaptation is ingeniously designed for **bivariate comparison**, allowing researchers to display and contrast the distributions of **two distinct datasets** that share a common numerical scale or stem structure. This method eliminates the need to visually compare two separate plots, simplifying the analysis of relative positions and shapes.

The structure of the back-to-back plot involves utilizing a single central column for the stems, which is shared by both datasets. The leaves for the first dataset branch off to the right of the stem, following the standard ascending order (smallest value closest to the stem). Conversely, the leaves for the second dataset branch off to the left of the stem. A crucial convention for the left-side data is that the leaves must be ordered in descending numerical sequence as they move away from the stem, ensuring that the plot maintains a consistent visual representation of ascending values as one reads outward from the center.

This symmetrical design provides an immediate, powerful visual contrast, enabling analysts to quickly identify differences in central tendency, spread, and the overall shape of the data distributions between the two groups. It is especially effective in fields such as sports statistics, health research, and educational assessments where comparing two similar groups is fundamental to drawing conclusions.

Step-by-Step Guide to Creating a Back-to-Back Plot

To illustrate the construction of this comparative tool, we will use an example involving points scored by members of two professional basketball teams. The first step involves ensuring both datasets are sorted in ascending order. This preparation is critical for correctly determining the stem and leaf values and arranging the leaves properly on either side of the plot.

Suppose we have the following two datasets detailing player points scored in a season:

Mavericks (Dataset 1): 2, 4, 8, 12, 12, 12, 15, 19, 23, 25, 31, 35, 38

Lakers (Dataset 2): 6, 6, 7, 12, 13, 15, 16, 20, 22, 24, 28, 30, 31

We must first identify the range of stems needed. Since the scores range from 2 to 38, our stems will be 0, 1, 2, and 3. We then proceed to plot the leaves. The Mavericks' leaves (right side) are written in ascending order away from the stem. The Lakers' leaves (left side) are written in descending order as they move away from the stem to maintain the visual flow of magnitude.

Lakers		Mavericks
7 6 6	0	2 4 8
6 5 3 2	1	2 2 2 5 9
8 4 2 0	2	3 5
1 0	3	1 5 8

As shown in the plot, the points scored by the Mavericks are systematically represented on the right side of the central stem, while the points scored by the Lakers are mirrored on the left. It is important to confirm that the number of individual leaf values shown on each side corresponds precisely to the total number of data points in each original dataset (13 values for the Mavericks and 13 values for the Lakers). This visual confirmation ensures no data points were omitted during the plotting process.

Visual Interpretation and Distribution Comparison

The true utility of the back-to-back stem-and-leaf plot lies in its ability to facilitate the comparison of the two data distributions. By viewing the mirrored plot, we can immediately assess how the scoring patterns differ between the Mavericks and the Lakers. We look for characteristics such as

where the data is most heavily clustered (central tendency), how spread out the scores are (variability), and the general shape (symmetry or skewness).

In this specific example, a visual inspection reveals key differences. The Lakers' scores (left side) appear slightly more concentrated in the lower stems (0 and 1) compared to the Mavericks. However, the Lakers also exhibit a greater frequency of high scores in the Stem 2 range (20s). The Mavericks' scores, particularly the leaves in the 10s (Stem 1), show a greater cluster of lower scores (multiple 12s), indicating perhaps a larger group of players scoring consistently, but moderately.

The overall shape of the distributions suggests that the Lakers' data might be slightly less spread out than the Mavericks', especially when looking at the extreme values. The plot provides intuitive, graphical evidence that complements the quantitative analysis derived from calculating statistical measures like the mean and standard deviation. It serves as an initial, critical step in understanding the underlying data patterns before moving to more complex statistical inference.

Key Statistical Measures Derived from the Plot

Beyond visual assessment, the back-to-back stem-and-leaf plot allows for the rapid calculation of several crucial descriptive statistics, as all data points are maintained in an ordered fashion. We can easily determine the range, the mode, and the median directly from the plot without referencing the original raw data lists.

Question 1: What is the range for the number of points scored for each team?

*Recall that the **range** is a simple measure of spread, calculated as the difference between the largest value and the smallest value in the dataset. By reading the outermost leaves on the highest and lowest stems, we quickly find the extremes.*

Range for the Mavericks: The lowest score is 2 (Stem 0, Leaf 2) and the highest is 38 (Stem 3, Leaf 8). Calculation: $38 - 2 = 36$.

Range for the Lakers: The lowest score is 6 (Stem 0, Leaf 6) and the highest is 31 (Stem 3, Leaf 1). Calculation: $31 - 6 = 25$.

The disparity in the range suggests that the Mavericks have a wider spread of individual player scores, indicating greater variability in performance compared to the Lakers.

Question 2: What is the mode for the number of points scored for each team?

*The **mode** is the value that appears most frequently in a dataset. In the stem-and-leaf plot, this corresponds to the leaf value that is repeated the most often on a single stem line.*

Mode for the Mavericks: The value 12 appears three times (Stem 1, Leaves 2, 2, 2). Mode: **12**.

Mode for the Lakers: The value 6 appears twice (Stem 0, Leaves 6, 6). Mode: **6**.

This confirms that for the Mavericks, 12 was the most common score among players, whereas 6 was the most common score for the Lakers.

Question 3: Calculate the median for the number of points scored for each team.

*The **median** represents the middle value in an ordered dataset. Since both datasets contain 13 values (an odd number), the median will be the $(n+1)/2 = 7$ th value. The ordered nature of the stem-and-leaf plot makes locating this value simple by counting the leaves from either end.*

Median for the Mavericks: Counting 7 leaves from the bottom or top of the Mavericks' data set (right side) lands on the value 15 (Stem 1, Leaf 5). Median: **15**.

Median for the Lakers: Counting 7 leaves from the bottom or top of the Lakers' data set (left side) lands on the value 16 (Stem 1, Leaf 6). Median: **16**.

The medians are very close, indicating that the typical scoring output for both teams is nearly identical, despite the differences observed in spread (range) and mode.

Comparative Analysis and Summary of Findings

The back-to-back stem-and-leaf plot is exceptionally useful for addressing direct comparative questions about frequency counts and extreme values. By visually grouping the data, we can quickly answer specific queries related to performance thresholds or outliers, solidifying the analytical insights gained from the descriptive statistics.

Question 4: Which team had more players score 20 or more points?

To answer this, we count the leaves in the stems of 2 and higher (20, 30, etc.) for both teams.

Players who scored 20 or more for the Mavericks: Count leaves on Stem 2 (3) and Stem 3 (3). Total: **6 players** (23, 25, 31, 35, 38).

Players who scored 20 or more for the Lakers: Count leaves on Stem 2 (5) and Stem 3 (2). Total: **7 players** (20, 22, 24, 28, 30, 31).

Based on the leaf counts, the **Lakers** had one more player score 20 or more points during the season, suggesting a slightly deeper pool of high-output players.

Question 5: Which team had the highest-scoring player?

This requires identifying the maximum value for each team by examining the outermost leaf on the highest stem (Stem 3).

Highest scorer for the Mavericks: Found on Stem 3, Leaf 8, representing **38 points**.

Highest scorer for the Lakers: Found on Stem 3, Leaf 1, representing **31 points**.

The **Mavericks** clearly had the highest-scoring individual player, reaching 38 points, which also contributes significantly to their larger overall range.

In summary, the back-to-back stem-and-leaf plot confirms that while both teams had similar median scoring outputs, the Mavericks exhibited a greater degree of variance in player scores (larger range), largely driven by one high-scoring player. Conversely, the Lakers showed a slightly tighter distribution but had a marginally greater number of players achieving scores above the 20-point threshold. This combined visual and numerical analysis highlights the power of this single plot in drawing detailed comparative conclusions.

ARABPSYCHOLOGY.COM