

Learning Stem-and-Leaf Plots: A Comprehensive Guide with Examples

Authored by
Mohammed loot

November 9, 2025

RECOMMENDED CITATION

Mohammed loot (2025). *Learning Stem-and-Leaf Plots: A Comprehensive Guide with Examples*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=14544>

A [stem-and-leaf plot](#), frequently called a stem plot, is an ingenious graphical method designed for displaying quantitative [data](#). This visualization technique was a key innovation developed by [John Tukey](#) during the advent of [Exploratory Data Analysis \(EDA\)](#). The method works by systematically partitioning every numerical value within a [dataset](#) into two distinct components: the "stem," which represents the leading digits, and the "leaf," which always represents the trailing single digit.

Unlike more aggregated visualizations, such as the [histogram](#), the stem-and-leaf plot offers a significant advantage: it not only provides a clear visual depiction of the shape and [distribution](#) of the data but also meticulously preserves the original numerical values. This dual functionality makes it an exceptionally powerful tool for preliminary data review. This comprehensive guide will meticulously detail the step-by-step methodology required for constructing these plots and demonstrate effective techniques for interpreting the valuable statistical insights they provide.

Deconstructing the Components: The Stem and the Leaf

To correctly construct a meaningful stem-and-leaf plot, one must first possess a fundamental understanding of the role and structure of its two principal components. The **stem** is the foundational element, representing the class interval or the grouping of the data values. Its composition typically involves the first digit or the leading two digits of the data points, depending entirely on the magnitude and range of the numbers present in the [dataset](#). Crucially, all stems must be arranged vertically and ordered sequentially from the smallest value to the largest.

The **leaf** is the remaining digit after the stem has been accurately identified; by definition, it is always the last single digit of the number. The leaves are placed horizontally alongside their corresponding stem. It is absolutely essential that the leaves for each stem are listed in increasing numerical order, as this ordering is necessary to accurately display the internal structure and [distribution](#) of the data within that interval. Furthermore, for the plot to be statistically complete and interpretable, it must include a clear key or legend that specifies how the stem and the leaf combine to represent the actual full data value.

This visualization technique proves particularly valuable when analyzing small-to-medium sized datasets, generally those containing fewer than 50 observations. For these smaller datasets, the stem plot allows for a rapid, manual assessment of the data's [central tendency](#), spread, and overall shape without sacrificing the inherent precision provided by the raw numerical values.

Step-by-Step Construction: Example 1 (Two-Digit Data)

The following example demonstrates the fundamental process of creating a stem-and-leaf plot using a straightforward dataset composed entirely of two-digit numbers. Mastering this approach provides the necessary foundation for tackling more complex statistical visualizations.

Imagine we are working with the following collection of observations, which have been conveniently sorted in ascending order for ease of processing:

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

The objective is to transform this linear sequence of numbers into a structured plot that visually communicates where the values are concentrated. Below is the final appearance of the structured plot, followed by a detailed explanation of the steps required for its creation:

```
1 | 2 4 8
2 | 2 2 3 5 5 8
3 |
4 | 5 7 8
```

The construction process is broken down into two critical steps that guarantee accuracy and clarity in the resulting visualization:

Step 1: Partition Each Value into a Stem and a Leaf. For two-digit numerical values, the **stem** is defined as the first digit (representing the tens place), and the leaf is defined as the second digit (representing the units place).

The "**stem**" of each value (highlighted in teal) identifies the tens place grouping:

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

The "**leaf**" of each value (highlighted in red) identifies the units place:

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

Step 2: Assemble the Plot. The derived stems are placed along the vertical axis, ordered from the minimum value (1) to the maximum value (4). The corresponding leaves are then carefully aligned horizontally next to their respective stem, ensuring they are also listed in increasing order. It is vital to note the deliberate inclusion of stem 3, which is necessary to represent the gap in the data between the twenties and the forties. Including all consecutive stems ensures the visual distribution accurately reflects the data's true spread.

```
1 | 2 4 8
2 | 2 2 3 5 5 8
3 |
4 | 5 7 8
```

Adapting the Stem: Example 2 (Three-Digit Data)

When confronted with data points that contain three or more digits, the standard definition of the stem must be carefully adapted. This modification is crucial to maintain visual clarity and ensure the plot has an appropriate number of bins. In such instances, the stem will typically encompass the leading two digits, while the leaf must always remain a single, final digit. This necessary adjustment allows for the creation of a meaningful and concise visual representation of the [dataset](#).

Consider the following three-digit dataset, which represents a small collection of measurements:

134, 156, 158, 159, 160, 162, 164

If we were to incorrectly use only the first digit (1) as the stem, the resulting leaves (34, 56, 58, etc.) would be composed of two digits. This violates the fundamental rule that a leaf must only be a single trailing digit, thus rendering the plot structurally invalid and visually useless. Therefore, we must define the stem to include the first two digits (the hundreds and tens place). The resulting plot structure, which effectively groups the data, is shown below:

```
13 | 4
14 |
15 | 6 8 9
16 | 0 2 4
```

Here is the modified construction process detailing the necessary redefinition of the components:

Step 1: Redefine the Stem and Leaf. Since the values are three digits, the revised "**stem**" is defined as the first two digits (the hundreds and tens), and the "**leaf**" remains the final single digit (the units).

The "**stem**" (highlighted in teal) now represents the first two places:

134, 156, 158, 159, 160, 162, 164

The "**leaf**" (highlighted in red) remains the final single digit:

134, 156, 158, 159, 160, 162, 164

Step 2: Plotting the Adjusted Values. The newly defined stems (13, 14, 15, 16) are listed vertically. We again include the empty stem of 14, which visually illustrates the significant gap between the data point 134 and the next cluster starting at 156. The leaves are then correctly plotted and ordered horizontally adjacent to their corresponding stem.

13 | 4

14 |

15 | 6 8 9

16 | 0 2 4

Interpreting Statistical Insights from the Plot

The primary advantage of the stem-and-leaf plot lies in its remarkable ability to quickly reveal the core statistical characteristics of a dataset. These characteristics include the shape of its [distribution](#), measures of [central tendency](#), and the overall range. When interpreting the visual output, it is often helpful to mentally rotate the plot 90 degrees counter-clockwise; the resulting profile formed by the rows of leaves directly resembles a [histogram](#), making patterns like skewness and symmetry immediately apparent.

To illustrate interpretation, suppose we analyze the following stem-and-leaf plot, which systematically documents the number of ice cream cones sold daily by Marie over a period of 14 days:

7 | 3 3 3 7 9

8 | 4 5 8 8

9 | 1 2 5 8 9

In this representation, the stems represent the tens digit (sales in the 70s, 80s, and 90s), and the leaves represent the units digit. The visualization is structured to allow for immediate answers to common statistical inquiries based solely on its visual layout.

Question 1: *What is the highest number of cones Marie sold on any given day?*

To determine the maximum value (the range maximum), we simply examine the last stem (9) and the last leaf (9) in that row. Combining these two components yields the highest recorded sales figure.

Answer: 99 cones

7 | 3 3 3 7 9

8 | 4 5 8 8

9 | 1 2 5 8 9

Question 2: *During how many days did Marie sell more than 80 cones?*

Answering this question involves counting the total number of data points (leaves) that correspond to sales greater than 80. This count must include all leaves associated with stem 8 (since 8 | 4

means 84, which is greater than 80) and all leaves associated with stem 9. The length of the leaf rows immediately provides the [frequency](#) count for those bins.

Answer: nine days

7 | 3 3 3 7 9

8 | 4 5 8 8 (4 observations)

9 | 1 2 5 8 9 (5 observations)

(Summing the leaves in the 8 and 9 rows gives $4 + 5 = 9$ days. The specific sales figures were 84, 85, 88, 88, 91, 92, 95, 98, and 99 cones.)

Strengths and Limitations of Stem Plots

Stem-and-leaf plots maintain their relevance in statistics due to several intrinsic benefits, primarily centered on their superior capability for maintaining the integrity of the original [data](#). Unlike standard [histograms](#), where data points are aggregated into bins and the individual values are permanently lost, the stem plot allows the reader to precisely reconstruct every single observation directly from the visualization. This feature is particularly invaluable in specialized fields such as quality control, auditing, or small-scale scientific research where the exact value of every measurement is paramount.

Furthermore, these plots significantly facilitate the rapid manual calculation of key statistical measures. Since the leaves are naturally ordered when the plot is constructed, it becomes very quick to visually locate the median (the middle value) and calculate the range (maximum value minus minimum value). The plot also simplifies visual comparisons between two related [datasets](#) through the use of back-to-back stem-and-leaf plots, where two sets of leaves are displayed on opposing sides of a shared central stem.

However, stem plots possess notable limitations that restrict their application. They become unwieldy, cumbersome, and difficult to interpret when dealing with extremely large datasets (those with hundreds or thousands of observations), as the rows of leaves grow excessively long and dense. Similarly, if the dataset spans an extremely wide numerical range, the resultant number of stems may become too great, diluting the plot's visual summarizing effect. In such cases, more advanced or highly aggregated visualization methods, such as box plots, specialized frequency tables, or interactive [histograms](#), are typically the preferred tools for [Exploratory Data Analysis](#).

Additional Resources for Data Visualization

For those seeking further exploration and practical application of stem-and-leaf plots and related visualization techniques, consult the following external resources:

[Stem and Leaf Plot Generator](#)

[An Introduction to Back to Back Stem-and-Leaf Plots](#)

[How to Create a Stem-and-Leaf Plot in Excel](#)

[How to Create a Stem-and-Leaf Plot in Python](#)