

# How to Create and Interpret Stem-and-Leaf Plots

Authored by  
**stats writer**

January 21, 2026

## RECOMMENDED CITATION

stats writer (2026). *How to Create and Interpret Stem-and-Leaf Plots*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=126841>

The stem-and-leaf plot, often referred to as a stem plot, is an inventive tool utilized in exploratory data analysis to visualize quantitative numerical data. Unlike a traditional histogram, this plot retains the original data values while organizing them based on their magnitude. It serves as a semi-graphical method where the data points are separated into two distinct components: the stem (usually the leading digit or digits) and the leaf (usually the final digit). This powerful technique was developed by statistician John Tukey in the 1970s.

The primary benefit of employing a stem-and-leaf plot is the ability to rapidly assess the shape and spread of a dataset. By examining the structure, one can quickly determine key statistical measures such as the range, locate the median, identify concentrations of data, and spot potential outliers. These plots are particularly useful when dealing with small to moderately sized datasets, offering a clear snapshot of the underlying distribution of values.

A **stem-and-leaf plot** displays data by systematically splitting each recorded value in a dataset into two parts: the representative "stem" and the corresponding "leaf." The organization of these components allows for immediate visualization of the frequency and density of the data points.

This comprehensive guide will thoroughly explain the methodology required to accurately create, interpret, and analyze various types of stem-and-leaf plots, using detailed, step-by-step examples.

## Fundamentals of the Stem and Leaf Structure

The core concept behind the stem-and-leaf plot is the partitioning of each observed value. Typically, the stem represents the higher place value digits, while the leaf represents the lowest place value digit. This convention ensures that the plot maintains the integrity of the original numbers while providing a graphical representation of the frequency.

The stem is generally positioned vertically, acting as a category organizer, similar to the bins in a histogram. The leaves extend horizontally from the stem, and their collective length visually indicates the frequency of data points within that specific category or range defined by the stem. The plot must always include a key or legend to clarify what the stem and leaf represent, especially when dealing with decimals or varied scaling.

For instance, if we are plotting the value 42, the stem might be 4, and the leaf would be 2. If the value is 105, the stem might be 10, and the leaf would be 5. This flexibility in defining the stem is essential when dealing with datasets that span a wide range of values or require differing levels of detail in their graphical representation.

## How to Construct a Stem-and-leaf Plot: Example 1 (Two-Digit Data)

The following example demonstrates the fundamental process of creating a stem-and-leaf plot

using a simple dataset consisting of two-digit integers. This process involves sorting the data and then systematically separating each number into its designated stem and leaf components.

Suppose we are analyzing the scores from a short quiz and have the following numerical records:

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

Before plotting, it is standard practice to arrange the data in ascending order. Since the provided data is already ordered, we proceed directly to the separation process. The resulting stem-and-leaf plot visually organizes the frequency of scores across the tens places:

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

Here is a detailed breakdown of the construction methodology:

### **Step 1: Define and Separate the Stem and Leaf Components.**

For this dataset, which consists entirely of two-digit numbers, the "**stem**" is defined as the first digit (the tens place), and the "**leaf**" is the second digit (the units place). We must systematically process every value:

The "**stem**" of each value (the leading digit):

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

The "**leaf**" of each value (the trailing digit):

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

### **Step 2: Construct the Plot.**

Place the collected "stems" along the vertical axis in ascending order. Next, arrange the corresponding "leaves" horizontally next to their stem, also in ascending order. Note that any stem value present between the minimum and maximum observed stems must be included, even if it has no leaves:

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

## Handling Complex Data: Example 2 (Three-Digit Data)

When dealing with larger numerical data points, the definition of the stem must be adjusted to ensure the plot remains manageable and informative. Using the first two digits as the stem often provides the best visual balance.

Consider the following dataset, which represents measured heights in centimeters:

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

Since these values are concentrated in the 100s, using the first two digits as the stem is the most appropriate approach. The resulting stem-and-leaf plot is structured as follows:

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

Let us examine the steps taken to achieve this representation:

### Step 1: Re-defining the Stem and Leaf.

In this case, since the values in this dataset have three digits, the "**stem**" is defined as the first two digits (the hundreds and tens place values):

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

The "**leaf**" of each value is the last digit:

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

### Step 2: Constructing the Final Plot.

The stems are plotted vertically, including the missing stem '14' to ensure proper representation of the scale. The leaves are appended in increasing order:

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

## Interpreting a Stem-and-Leaf Plot for Statistical Insight

Once constructed, the stem-and-leaf plot is not merely an organized list; it is a powerful tool for visual data analysis. Interpretation involves examining the shape, center, and spread of the data, which collectively reveal the underlying distribution.

The length of the leaf rows reveals the frequency (or count) for that specific stem range. Longer rows indicate higher concentrations of data. Furthermore, by tilting the plot 90 degrees to the left, the visualization resembles a histogram, allowing for immediate assessment of skewness, symmetry, and modality. Identifying gaps in the data (empty stems) or isolated extreme values (very short rows at the ends) is also straightforward.

Suppose we analyze a plot detailing the number of ice cream cones that Marie sold at her ice cream shop during each of the past 14 days:

```
7 | 3 3 3 7 9
8 | 4 5 8 8
9 | 1 2 5 8 9
```

**Question 1:** *What was the maximum number of cones that Marie sold on any given day?*

Answer: By looking at the highest stem ('9') and the largest leaf attached to it ('9'), we reconstruct the highest value: 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?*

Answer: We count all leaves associated with stems 8 and higher. Stem 8 has four leaves, and Stem 9 has five leaves, totaling nine days.

```
7 | 3 3 3 7 9
8 | 4 5 8 8
9 | 1 2 5 8 9
```

(During these nine days, she sold 84, 85, 88, 88, 91, 92, 95, 98, and 99 cones.)

## Advanced Measures: Finding the Median, Range, and Outliers

The ordered nature of the stem-and-leaf plot makes calculating key statistical measures exceptionally easy, often requiring less effort than analyzing raw, unsorted data. These plots are excellent tools for preliminary data screening before moving to more complex analyses.

To find the **median**, which is the middle value of the ordered dataset, one simply counts the total number of leaves ( $N$ ). If  $N$  is odd, the median is the value corresponding to the  $(N+1)/2$  position. If  $N$  is even, the median is the average of the two central values ( $N/2$  and  $N/2 + 1$  positions). Since the leaves are already ordered horizontally and the stems are ordered vertically, finding this position is a direct counting exercise on the plot itself.

The **range** is the simplest measure of spread, calculated as the difference between the maximum and minimum values. These two extreme values are instantly identifiable: the minimum is the value formed by the lowest stem and its first leaf, and the maximum is the value formed by the highest stem and its last leaf. For the ice cream data (73 to 99), the range is  $99 - 73 = 26$  cones.

Identifying **outliers** is another crucial application. An outlier appears as a solitary leaf or a very short row separated by significant empty stems from the main body of the data. While this identification is qualitative without calculating formal fence values, the visual gap provides a strong indication that an extreme value warrants further investigation.

## Stem-and-Leaf Plots versus Histograms

While the stem-and-leaf plot provides a graphical summary similar to a histogram, its unique construction offers distinct advantages. Both tools effectively visualize the distribution shape and central tendency of a continuous numerical data set.

The primary advantage of the stem-and-leaf plot is its **data preservation**. Unlike a histogram, where data is grouped into bins and the individual values are lost, the stem-and-leaf plot retains every original numerical entry. This preservation is invaluable during the exploratory phase of analysis when raw data points are critical for verification and detailed observation.

However, histograms are superior when dealing with very large datasets, as a stem-and-leaf plot can become unwieldy and impractical due to the sheer volume of leaves. Histograms also allow for greater flexibility in controlling the bin width, which can be adjusted to smooth out variation and reveal underlying patterns more clearly than the fixed structure of a stem-and-leaf plot.

## Practical Applications and Advanced Variations

The stem-and-leaf plot is a versatile technique used across various fields for quick data assessment:

**Education:** Quickly mapping student test scores to identify if grades cluster around a certain average or if there are significant outliers.

**Quality Control:** Analyzing manufacturing defect rates or product specifications to ensure consistency and compliance.

**Environmental Science:** Plotting daily temperature readings or pollution levels to observe seasonal trends and rapid changes.

One notable variation of the standard plot is the **Back-to-Back Stem-and-Leaf Plot**. This design is specifically engineered to compare two related datasets side-by-side using a common central stem. The leaves for the first dataset extend to the left of the stem, while the leaves for the second dataset extend to the right. This visual comparison tool is extremely effective for contrasting metrics, such as comparing the test scores of two different classes or the performance metrics of two teams.

Furthermore, when a dataset is tightly clustered, analysts might employ a **Split Stem** technique. This involves repeating each stem value, where the first iteration holds leaves 0 through 4, and the second iteration holds leaves 5 through 9. This effectively doubles the number of categories, stretching the visualization vertically to reveal finer details in the distribution shape that might otherwise be obscured.

## Conclusion and Further Exploration

The stem-and-leaf plot remains a cornerstone of exploratory data analysis due to its simplicity and its capacity to display both the graphical shape and the precise numerical values of a dataset. While perhaps overlooked in favor of computerized graphics, its efficiency for manual calculation and rapid preliminary data assessment ensures its continued relevance in statistical education and basic data reporting.

[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](#)