Learn How to Create a Stem-and-Leaf Plot in SPSS: A Step-by-Step Guide

Authored by

Mohammed looti

November 8, 2025

A Stem-and-leaf plot is a unique and effective statistical graph used in exploratory data analysis. Its fundamental design displays numerical data by partitioning each value in a dataset into two distinct components: a *stem* and a *leaf*. This structure is particularly valuable because it allows researchers to visualize the overall distribution of the data while simultaneously retaining the integrity of the original raw data points. Unlike a histogram, which groups data into bins, the stem-and-leaf plot provides a non-aggregated summary, making it a powerful tool for initial data screening and identifying potential outliers or patterns.

This comprehensive guide provides a detailed walkthrough explaining precisely how to generate, format, and interpret a Stem-and-leaf plot using the industry-standard statistical software, **IBM SPSS Statistics**. We will focus on ensuring every step is clear, from data preparation to the final output interpretation, enabling users to efficiently integrate this visualization method into their analytical workflows.

Understanding the Stem-and-Leaf Plot

The core concept of the Stem-and-leaf plot, developed by statistician John Tukey, centers on dividing each observed numerical value into its significant digits (the stem) and its trailing digit (the leaf). The **stem** typically represents the larger place values, such as the tens, hundreds, or even thousands, depending on the magnitude of the data. Conversely, the **leaf** universally represents the last significant digit, usually the units place, providing the fine detail of the individual observation. This systematic decomposition results in a graphical representation that resembles a histogram turned on its side, but with the added benefit of numerical fidelity.

For instance, if we consider a data point of 24, the stem would be 2 (representing 20), and the leaf would be 4. If the data point were 107, the stem might be 10, and the leaf would be 7. This setup ensures that the length of each row of leaves visually represents the frequency of data points falling within that specific range defined by the stem. The visual display immediately reveals the shape of the data's distribution--whether it is symmetric, skewed, or if it contains multiple peaks (multimodal).

The preservation of raw data is arguably the most significant advantage of this plot type. When reviewing a histogram, one can only estimate the original values within a bin. However, when examining a Stem-and-leaf plot, every single data point used in the analysis is displayed explicitly. This feature is crucial during the early stages of data exploration, allowing researchers to verify the accuracy of the plotted frequencies and easily pinpoint the exact numerical value of any potential outliers that appear in the tail of the distribution.

Why Use a Stem-and-Leaf Plot in SPSS?

While SPSS offers a variety of advanced visualization tools, the Stem-and-leaf plot remains an essential component for sound exploratory data analysis (EDA). It serves as a rapid diagnostic tool to assess the data's fundamental properties before proceeding to complex inferential statistical tests. A quick scan of the plot can reveal essential characteristics, such as the data's central tendency, its spread (variability), and most importantly, its shape. Understanding the shape is critical because many statistical procedures rely on the assumption of normality.

Furthermore, SPSS integrates the generation of the Stem-and-leaf plot directly within the **Explore** function under Descriptive Statistics. This integration means that the plot is generated alongside crucial numerical summaries, including measures of central tendency (mean, median), measures of dispersion (standard deviation, variance, range), and tests for normality. This holistic output allows for an immediate comparison between the visual distribution and the quantitative summary statistics, streamlining the preliminary analysis phase.

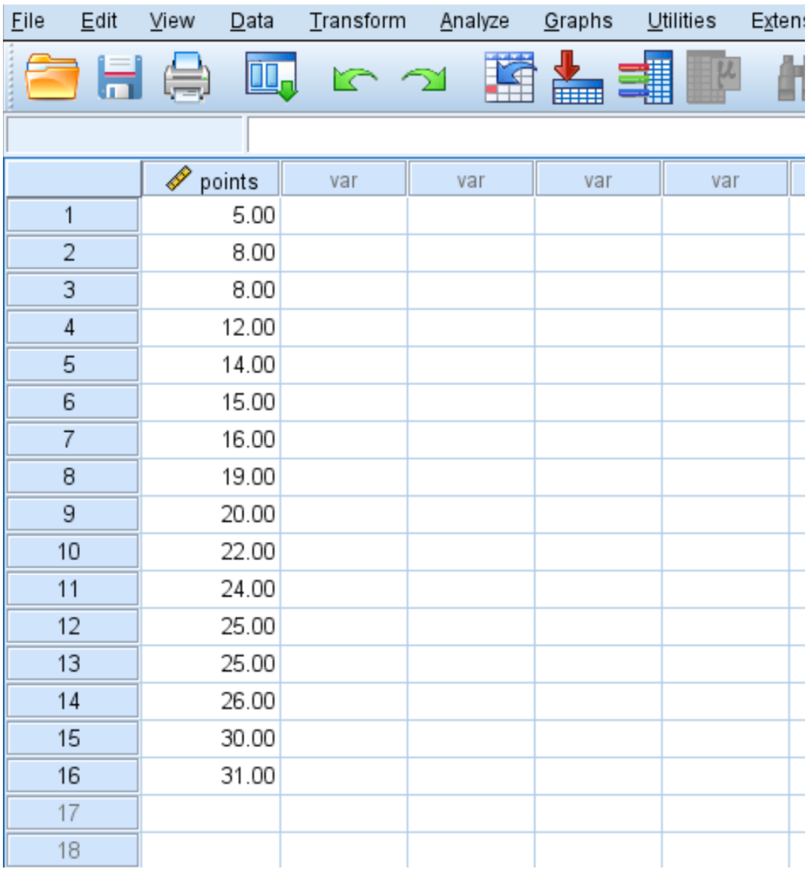
For researchers handling smaller to medium-sized datasets, the Stem-and-leaf plot offers unmatched clarity regarding data structure. Although it can become cumbersome with extremely large datasets, for the common range of sample sizes encountered in social sciences, business, and health research, it provides a transparent view of the data that complex graphical methods sometimes obscure. Its straightforward display makes it easy for non-statisticians to grasp the core characteristics of the data distribution quickly.

Preparing Your Data in SPSS (Using the Example Dataset)

To demonstrate the procedure for generating this plot, we will use a classic example involving performance metrics. Suppose we have collected data on the average points scored per game for a sample of 16 professional basketball players. This dataset represents a continuous, quantitative variable named "points." Before generating the plot, ensure your data is correctly entered into the SPSS Data View, with each row representing a unique player (case) and the scores entered into the designated variable column.

Accurate data entry is the foundational step for any statistical analysis. In our example, the "points" variable contains 16 distinct numerical observations. The scores range from the lowest observed value of 5 points to the highest observed value of 31 points. It is crucial to verify that the variable type is set to "Numeric" within the Variable View tab of SPSS, as graphical procedures rely on the correct classification of data types.

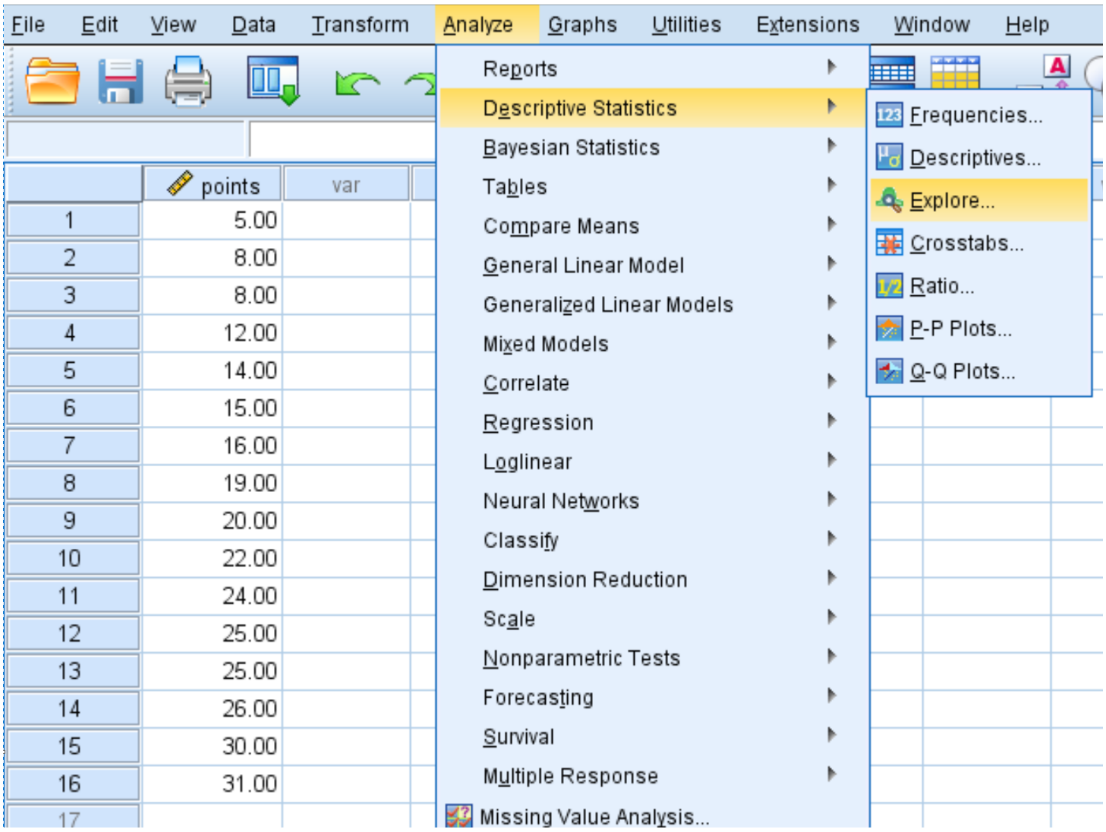
The following image illustrates how the raw data should appear within the SPSS Data Editor before beginning the analysis. Note that the data contains both single-digit and double-digit scores, which will necessitate the plot to define the stem based on the tens digit.



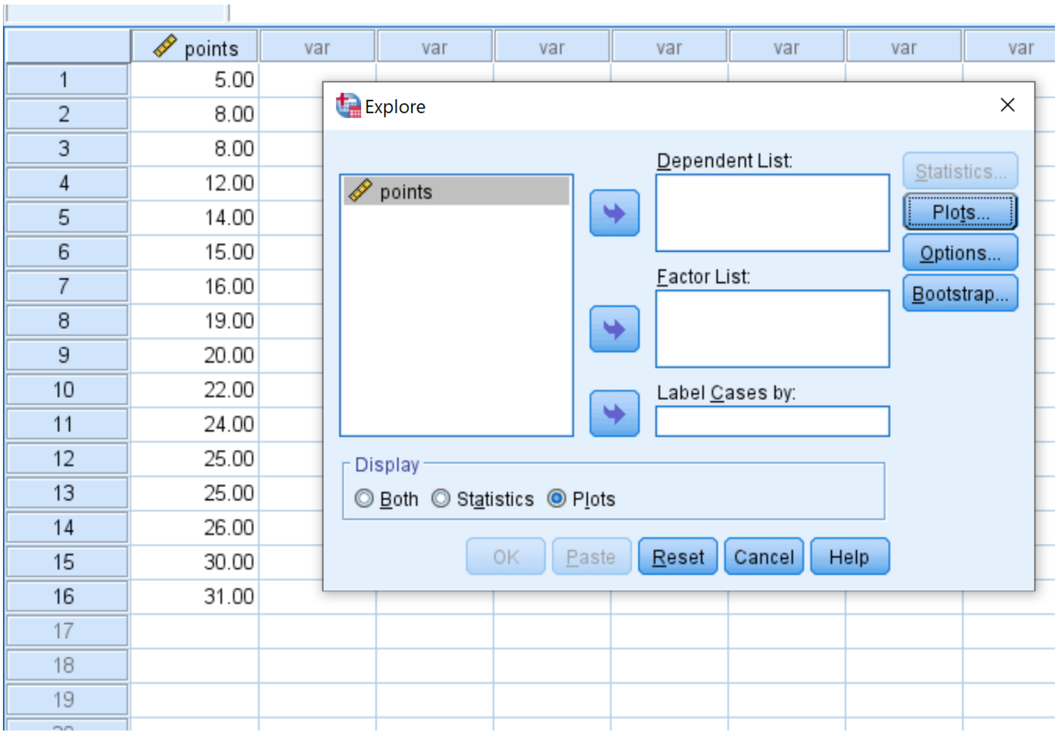
Step-by-Step Procedure for Plot Generation

Generating the Stem-and-leaf plot in SPSS is an intuitive process managed entirely through the graphical user interface. The procedure begins by accessing the primary analysis menu. You must click the **Analyze** tab located at the top toolbar. From the resulting drop-down menu, select **Descriptive Statistics**. Finally, navigate to and click the **Explore** option. This sequence initiates the dedicated dialog box designed for comprehensive data inspection.

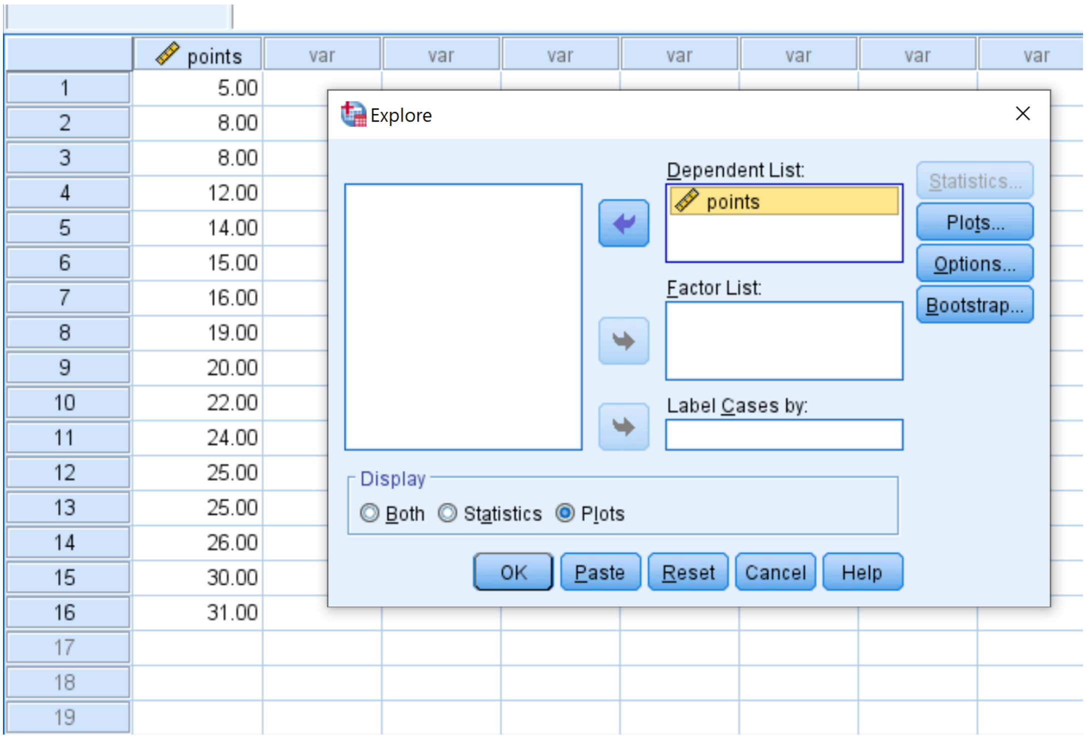
This specific menu path is chosen because the Explore function in SPSS is explicitly designed for detailed Descriptive Statistics and graphical summaries, making it the most appropriate location for generating visualization tools like the Stem-and-leaf plot and the Box plot. The following image highlights the necessary menu navigation required to open the Explore dialog window.



Once the **Explore** window appears, the next step involves specifying the variable you wish to analyze. You must locate the variable **points** in the list on the left side of the dialog box and move it into the field labelled **Dependent List**. The Dependent List is reserved for the quantitative variable(s) whose distribution you intend to examine. For this analysis, we are treating "points" as the sole dependent measure we wish to explore visually.



A critical setting within this dialog box is the **Display** option, located near the bottom left. By default, SPSS often selects **Both** (Statistics and Plots), but it is good practice to verify this selection. To ensure the Stem-and-leaf plot is generated, you must confirm that **Plots** is selected, or that **Both** is selected. If only **Statistics** is selected, the plot will be omitted from the output. After confirming the variable placement and the display setting, click the **Plots** button to further confirm that the "Stem-and-leaf" option is checked, although it is typically selected by default when **Plots** is chosen in the main window.



After ensuring all parameters are correctly configured--the **points** variable is dependent, and the output is set to generate plots--you can finalize the procedure. Click **Continue** in any sub-dialog boxes (like Plots, if you opened it) and then click **OK** in the main Explore dialog box. SPSS will then process the command and open the Output Viewer, displaying the requested descriptive statistics tables and the generated Stem-and-leaf plot.

Interpreting the SPSS Output

The resulting output in the SPSS Viewer will contain the generated Stem-and-leaf plot, which provides a concise visual summary of the basketball players' scores. The plot is structured into three main columns: **Frequency**, **Stem**, and **Leaf**. The **Frequency** column indicates how many data points fall within that particular stem category. The **Stem** column lists the leading digit(s) of the scores, and the **Leaf** column lists the corresponding trailing digits.

The structure of the plot immediately reveals the central tendency and spread of the data. In our example, the scores cluster primarily around the middle stems (1 and 2), indicating that the average points per game for this group of players generally fall within the 10s and 20s. Observing the length of the rows of leaves allows for a rapid assessment of modality and skewness.

points

```
points Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

      .00        0 .
     3.00        0 .  588
     2.00        1 .  24
     3.00        1 .  569
     3.00        2 .  024
     3.00        2 .  556
     2.00        3 .  01

 Stem width:      10.00
 Each leaf:         1 case(s)
```

To correctly interpret the plot, it is essential to understand the scaling provided by SPSS. In this output, the plot key indicates that the stem represents the tens digit and the leaf represents the units digit. For example, focusing on the first row where the stem is '0', the leaf shown is '5'. This translates directly to the smallest score in the dataset: 0 * 10 + 5 = 5 points per game. This low score is visually separated from the main body of the distribution, which begins at the stem '1'.

```
points Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

      .00        0 .
     3.00        0 .  588
     2.00        1 .  24
     3.00        1 .  569
     3.00        2 .  024
     3.00        2 .  556
     2.00        3 .  01

 Stem width:      10.00
 Each leaf:         1 case(s)
```

Conversely, examining the higher end of the distribution provides insight into the maximum scores. The last row shown in the plot has a stem of '3' and a leaf of '1'. This combination corresponds to the highest scoring player in the sample, representing 3 * 10 + 1 = 31 points per game. By reading the leaves horizontally for each stem, we can reconstruct the entire set of 16 scores, confirming the plot's fidelity to the raw data.

```
points Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

      .00        0 .
     3.00        0 .  588
     2.00        1 .  24
     3.00        1 .  569
     3.00        2 .  024
     3.00        2 .  556
     2.00        3 .  01

 Stem width:     10.00
 Each leaf:        1 case(s)
```

Analyzing the Data Distribution

The final step is to leverage the visual information provided by the Stem-and-leaf plot to draw conclusions about the characteristics of the basketball scores. By observing the shape formed by the leaves, we can confirm that the data exhibits a slight positive skew (skewed to the right), as the tail extends further toward the higher values (30s) than the lower values (single digits). The bulk of the scores are concentrated in the '1' and '2' stems, suggesting that the average player scores between 10 and 29 points per game.

We can use the plot to easily identify the mode(s) of the distribution. The stem '2' has the longest row of leaves (a frequency of 5), indicating that scores in the 20s are the most frequent. Specifically, the scores 22 and 23 appear twice each, making them strong candidates for the modal values. This level of detail is immediately accessible from the plot, which complements the aggregated numerical insights provided by the accompanying Descriptive Statistics tables generated by the Explore command.

In conclusion, the Stem-and-leaf plot remains a remarkably useful tool within the SPSS environment for initial data inspection. It provides a clean, informative visualization that successfully bridges the gap between raw numerical data and graphical summaries. Mastering its generation and interpretation ensures that researchers can quickly and accurately assess the underlying distribution of their variables, laying a solid foundation for subsequent, more sophisticated statistical modeling.