

How to Create Side-by-Side Boxplots in SPSS: A Step-by-Step Guide

Authored by
mohammed looti

January 6, 2026

RECOMMENDED CITATION

mohammed looti (2026). *How to Create Side-by-Side Boxplots in SPSS: A Step-by-Step Guide*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=124734>

Creating boxplots, particularly in a side-by-side format, is a fundamental analytical step for visually comparing the data distribution across distinct groups. This tutorial provides a comprehensive guide to generating these visualizations within SPSS (Statistical Package for the Social Sciences).

The modern method in **SPSS** typically utilizes the **Chart Builder** interface. To begin, ensure your relevant dataset is open in the program. Navigate to the **Graphs** menu and select **Chart Builder**. Within the **Gallery** tab, locate and choose the **Boxplot** visualization type. For a side-by-side comparison, the process involves carefully mapping the variables: drag the quantitative variable (the data you wish to analyze, e.g., scores) to the Y-Axis or **Variable** slot, and drag the categorical grouping variable (e.g., teams, gender) to the X-Axis or **Category Axis** slot. If using the **Chart Builder**, you might also utilize the **Set color by** option for aesthetic differentiation, though the **Category Axis** usually handles the grouping for side-by-side plots.

Once the variables are correctly positioned, clicking **OK** generates the initial visualization. However, the utility of this chart extends beyond its basic creation. **SPSS** provides powerful customization features through the **Chart Editor** window, which automatically opens upon double-clicking the generated graph. Here, analysts can refine elements such as axis labels, titles, colors, outlier markers, and overall layout to meet specific publication or presentation standards, ensuring the visual output is both accurate and highly professional.

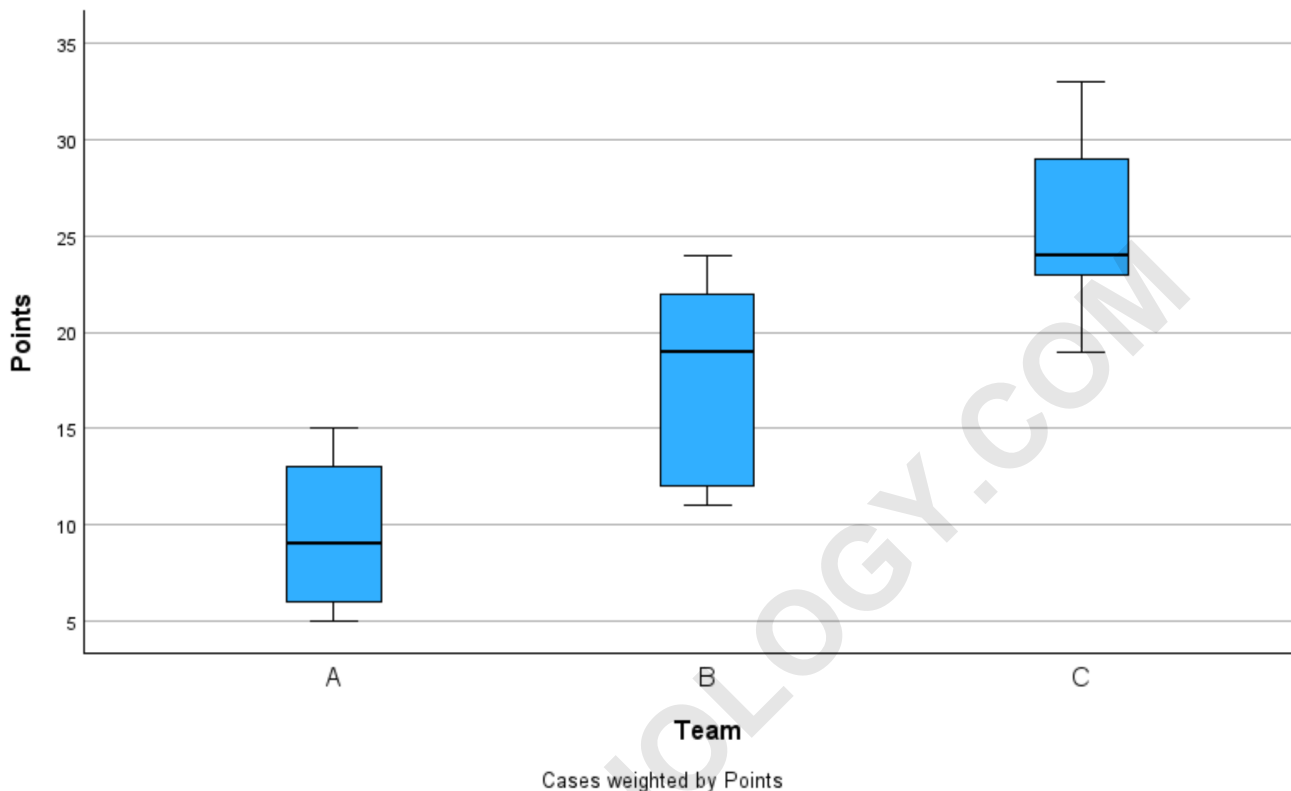
1. Introduction to Side-by-Side Boxplots

The utilization of **side-by-side boxplots** is a cornerstone technique in exploratory data analysis. These charts, also known as box-and-whisker plots, provide a concise visual summary of five key statistics for a dataset: the minimum value, the first quartile (Q1), the median (Q2), the third quartile (Q3), and the maximum value. When multiple boxplots are placed adjacent to one another--hence "side-by-side"--they allow for immediate, direct comparisons between the central tendencies, variances, and skewness of two or more distinct subgroups within the larger dataset.

This visualization is particularly effective for identifying group differences, spotting potential outliers, and assessing the general shape of the distribution without the complexity of viewing numerous individual histograms. For researchers and analysts working with **SPSS**, generating these comparative plots is an essential step before moving into formal inferential statistical testing, as it helps validate assumptions and guide model selection. The ability to quickly grasp distributional characteristics across groups is invaluable for drawing preliminary conclusions about the data.

The goal of this tutorial is to guide you precisely through the steps necessary to produce clear, comparative visualizations, such as the example shown below. This specific example, which we will replicate step-by-step, contrasts the distribution of points across three different teams,

showcasing how powerful these simple plots can be for complex group comparisons.



2. Why Boxplots Are Essential for Data Distribution Analysis

While standard descriptive statistics (mean, standard deviation) are useful, they often fail to capture the full picture of data distribution, especially when the data is skewed or contains anomalies. Boxplots excel at summarizing complex distributional properties into an easily digestible graphical format. They clearly delineate the **Interquartile Range (IQR)**, which represents the middle 50% of the data--a robust measure of spread that is less sensitive to extreme values than the standard deviation.

When generating side-by-side plots in **SPSS**, the vertical alignment immediately highlights differences in **central tendency**. If the boxes are positioned at different heights on the Y-axis, it suggests meaningful variation in the group medians. Furthermore, comparing the lengths of the boxes (the IQR) instantly reveals differences in data variability, indicating which groups are more spread out versus those that are tightly clustered. The lengths of the whiskers also provide insights into the overall range and the presence of potential outliers, which are typically marked individually beyond the whisker boundaries.

For researchers utilizing **SPSS** in fields such as psychology, economics, or biology, understanding these visual cues is crucial. A boxplot provides rapid diagnostic feedback. For instance, if one

group's box is noticeably skewed (the median line is not centered within the box), it alerts the analyst to potential non-normality, which might necessitate the use of non-parametric tests or data transformations before proceeding with inferential statistics. Thus, generating this visual first is a mandatory step in responsible statistical reporting.

3. Setting Up the Example Dataset in SPSS

To illustrate the process of creating comparative boxplots, we will use a hypothetical dataset tracking basketball player performance. The data structure required for this type of visualization must include at least two variables: one continuous quantitative variable (the metric being measured) and one categorical grouping variable (the defining characteristic of the groups being compared). In our example, the quantitative variable is **Points Scored**, and the categorical grouping variable is **Team Identity**.

Suppose our raw data, loaded into the **SPSS Data View**, looks exactly like the representation below. Note the clear distinction between the two variable types: **Points** contains numerical data suitable for measuring central tendency and spread, while **Team** contains categorical data (A, B, or C) necessary for creating the distinct groupings on the category axis.

	Team	Points	var	var	var
1	A	5			
2	A	5			
3	A	6			
4	A	6			
5	A	8			
6	A	9			
7	A	13			
8	A	15			
9	B	11			
10	B	11			
11	B	12			
12	B	14			
13	B	15			
14	B	19			
15	B	22			
16	B	24			
17	C	19			
18	C	23			
19	C	23			
20	C	23			
21	C	24			
22	C	26			
23	C	29			
24	C	33			
25					
26					
27					

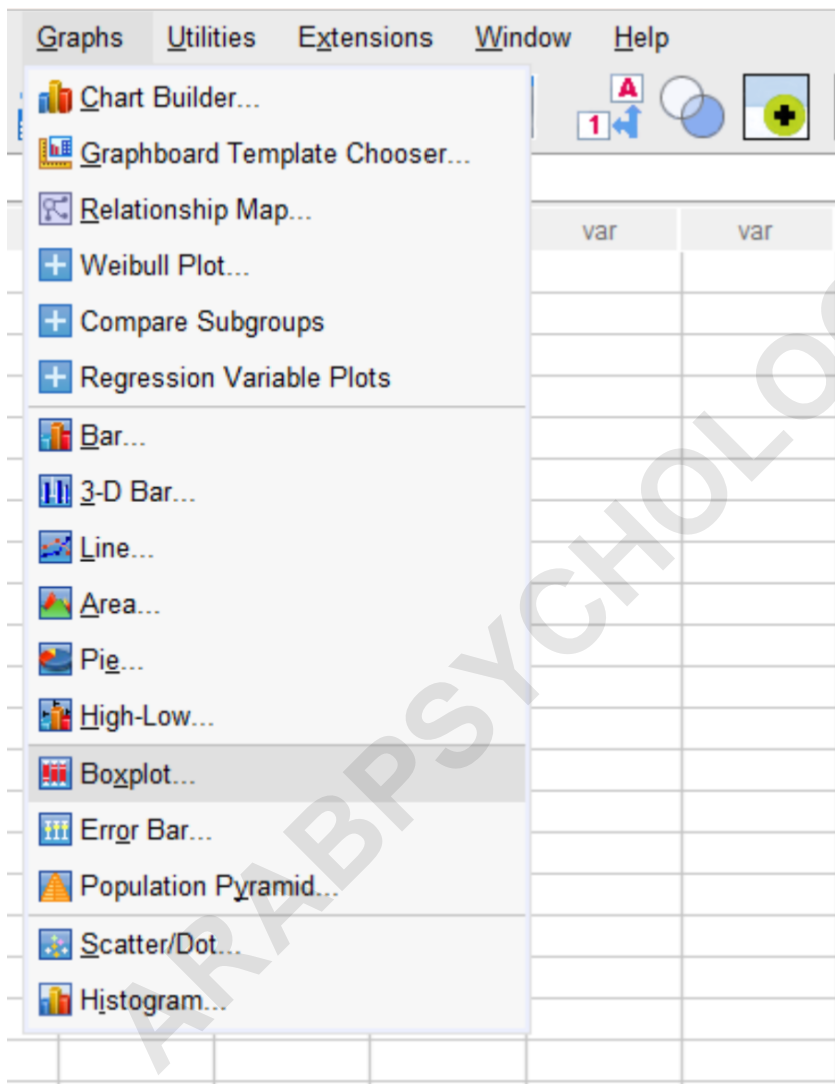
Ensuring the correct definition of variable types in the **Variable View** tab of **SPSS** is a critical precursor to generating accurate charts. The **Points** variable should be set as numeric, and ideally, the **Team** variable should be defined as a nominal or ordinal categorical variable. Although **SPSS** is flexible, properly defining these roles minimizes potential errors during the chart creation process, particularly when utilizing older or legacy dialogs, which can sometimes be less forgiving than the modern **Chart Builder** interface.

4. Generating Boxplots Using the Legacy Dialogs (Detailed Procedure)

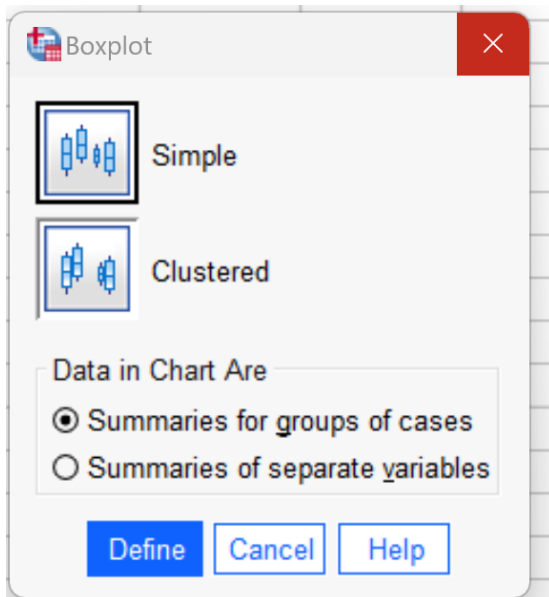
While the **Chart Builder** is the current standard, many long-time **SPSS** users and specific institutional workflows still rely on the **Legacy Dialogs** for generating charts, as these offer a quick,

direct route to common visualizations. Our example utilizes this classic method, which is highly efficient for simple side-by-side boxplots.

The initial procedural step requires navigating the primary menu structure. Click the **Graphs** tab located at the top of the **SPSS** window. From the subsequent dropdown menu, scroll down and select **Legacy Dialogs**, and then choose **Boxplot**. This action opens the dedicated Boxplot dialog box, prompting the user to select the appropriate chart type for the analysis.



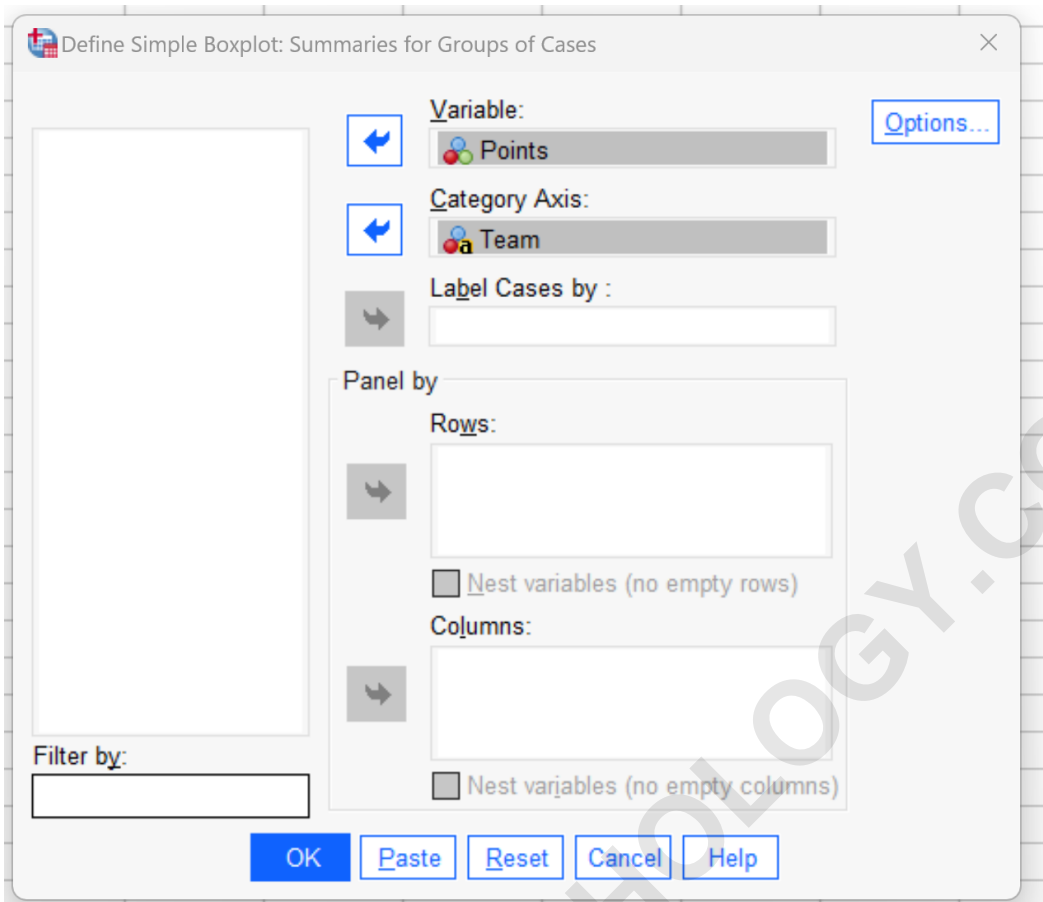
In the initial Boxplot dialog window that appears, you must specify the exact configuration needed. Since we are comparing the distribution of a single continuous variable across multiple groups, we select the **Simple** icon under the "Data in Chart Are" section. Crucially, underneath this, ensure you select **Summaries for groups of cases**. This tells **SPSS** that the chart should summarize the score variable based on the levels of the categorical variable. After confirming these selections, click the **Define** button to proceed to the variable assignment stage.



5. Defining Variables for Comparative Visualization

The "Define Simple Boxplot: Summaries for Groups of Cases" window is where the mapping of variables takes place, determining which variable constitutes the data points and which variable dictates the grouping. It is imperative that the variables are placed in their correct panels to ensure the generation of accurate side-by-side plots.

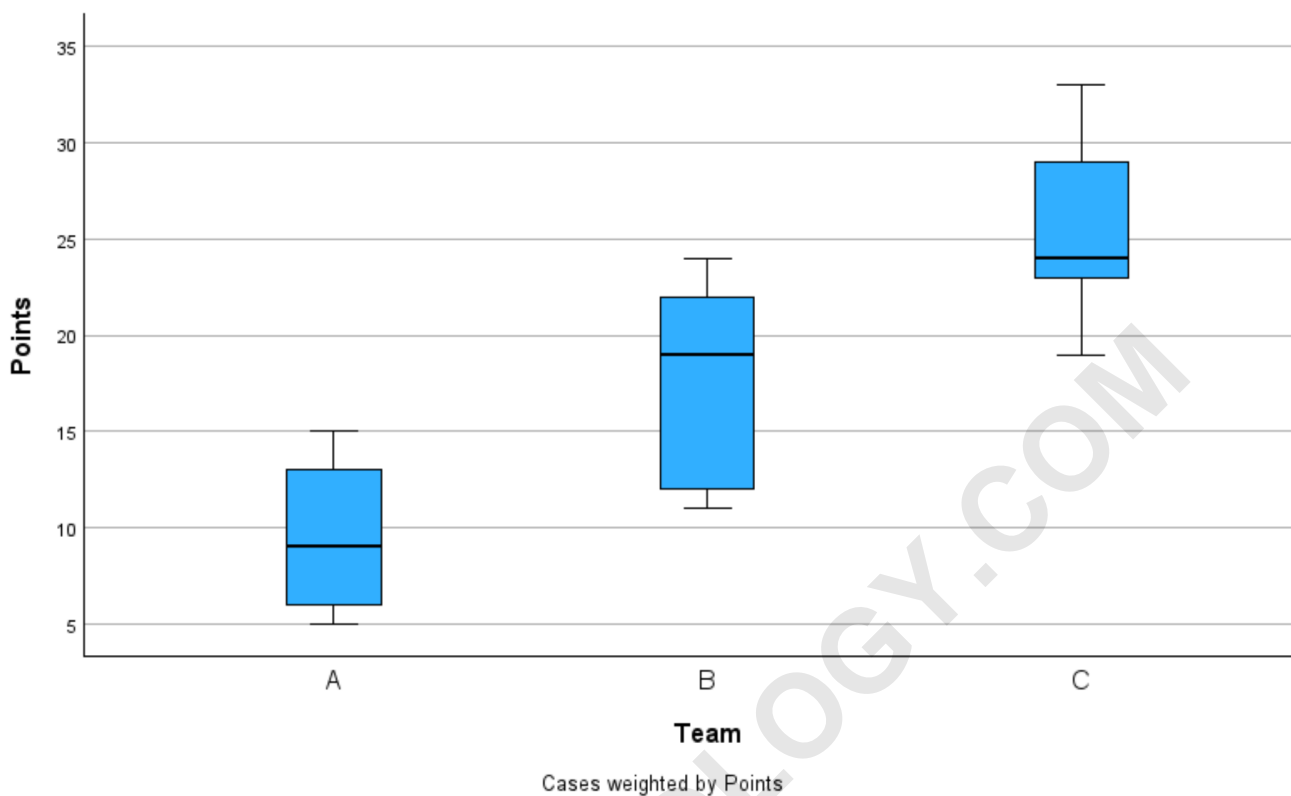
First, identify the dependent variable, which is the metric being measured--in this case, **Points**. Drag this variable from the variable list on the left and place it into the **Variable** panel. This defines the vertical axis of the final chart. Second, identify the independent or grouping variable, which is **Team**. Drag this variable into the **Category Axis** panel. This placement instructs **SPSS** to create a separate boxplot for each unique level (A, B, C) of the **Team** variable along the horizontal axis.



Before clicking **OK**, analysts should briefly review the optional settings available in this dialog. Options such as **Titles** allow for the immediate insertion of meaningful chart titles and footnotes, improving the readability of the output. While the default settings usually suffice for the basic structure, configuring titles at this stage saves subsequent effort in the **Chart Editor**. Once satisfied with the variable assignments and optional parameters, click **OK** to execute the command and generate the desired visualization in the **SPSS Output Viewer**.

6. Interpreting the Generated Side-by-Side Boxplots

Upon execution, the following side-by-side boxplots are generated, providing a clear visual representation of the distribution of points scored across the three basketball teams. The X-axis clearly delineates the categorical group (Team A, Team B, Team C), while the Y-axis represents the quantitative measurement (Points Scored).



The interpretation of these plots is multi-faceted and highly informative. By observing the position of the horizontal line within each box, which denotes the median (the 50th percentile), we can immediately draw conclusions about the center of the score distributions. Specifically, we observe that **Team C** exhibits the highest median points per player, positioning its box significantly higher on the Y-axis than the others. Conversely, **Team A** demonstrates the lowest median score, indicating a group whose typical performance lags behind the others. The relative height difference between the medians of Team A and Team B appears minimal, but Team C is clearly performing at a statistically higher central level.

Furthermore, examining the spread, defined by the length of the box, reveals differences in variability. The length of the box represents the Interquartile Range (IQR), which is the difference between the 75th percentile (top of the box) and the 25th percentile (bottom of the box). In this comparison, **Team B** clearly possesses the largest IQR, signifying that the scores of its players are the most dispersed or variable. This suggests a less consistent performance within Team B compared to Team A or Team C, which show tighter clustering around their respective medians. Analyzing these elements together--central tendency, spread, and potential outliers (if present outside the whiskers)--allows for a robust and visually compelling summary of group differences in performance.

7. Advanced Customization and Refinement in the Chart Editor

While the generated output from the Legacy Dialogs in **SPSS** is statistically accurate, it often requires aesthetic refinement for publication or professional presentation. This refinement is achieved within the **Chart Editor**. To access this powerful tool, simply double-click the generated boxplot in the **SPSS Output Viewer**. The **Chart Editor** provides granular control over virtually every visual element of the graph.

Key areas for customization often include the refinement of axis labels and titles. For instance, the Y-axis title might need to be changed from the generic variable name "Points" to the more descriptive "Points Scored per Player." Furthermore, managing **outliers** is a crucial aspect of boxplot customization. **SPSS** often flags outliers (usually marked as circles or asterisks) that fall more than 1.5 times the Interquartile Range (IQR) below Q1 or above Q3. Within the Chart Editor, analysts can adjust how these markers appear, or even choose to suppress them if the focus is strictly on the quartiles.

For high-quality output, specific formatting adjustments are highly recommended: adjusting font styles and sizes to match document standards, modifying fill colors for visual distinction between teams, and cleaning up unnecessary grid lines. Utilizing the **Properties** window within the **Chart Editor** allows users to fine-tune the scale and range of the axes, ensuring the visual differences between the group medians are emphasized effectively. These advanced steps transform a standard statistical graphic into a polished, interpretive figure.

8. Alternative Tasks and Further Resources

Mastering the creation of side-by-side boxplots is just one step in leveraging the full potential of **SPSS** for statistical analysis. Once comfortable with basic comparative visualizations, researchers often move on to related tasks that build upon these fundamental skills.

Some of the most common related procedures include generating split boxplots (where a single variable is split by two or more categorical factors), creating error bar charts for visualizing confidence intervals, and ultimately, performing inferential tests such as ANOVA or Kruskal-Wallis H test, which statistically confirm the visual differences observed in the boxplots. The boxplot visualization serves as the initial, non-parametric check on group differences.

The following list provides references to other essential tutorials covering common tasks within the **SPSS** environment, ensuring a continuous path toward statistical proficiency:

How to calculate descriptive statistics for different groups in SPSS.

Performing a one-way ANOVA to test for mean differences after visualization.

Conducting the Kruskal-Wallis H Test when distributions are non-normal.