

Learning to Create and Interpret Side-by-Side Boxplots in SPSS

Authored by
Mohammed looti

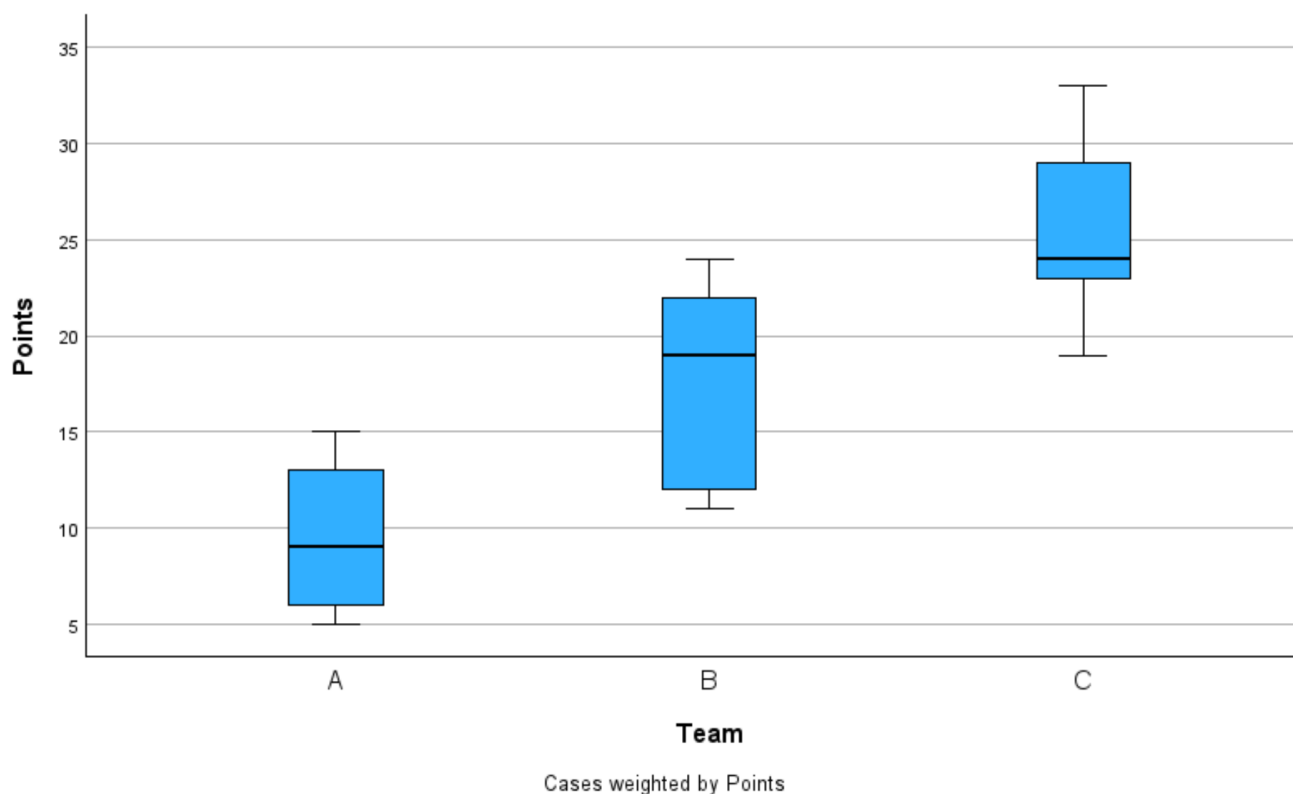
November 12, 2025

RECOMMENDED CITATION

Mohammed looti (2025). *Learning to Create and Interpret Side-by-Side Boxplots in SPSS*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=18268>

Visualizing the spread and central tendency of complex data across various subgroups is a fundamental requirement in modern [statistical analysis](#). To achieve this clarity, researchers rely heavily on graphical tools. One of the most effective and succinct instruments available is the [side-by-side boxplot](#) (also known as the box-and-whisker plot). These visual representations efficiently encapsulate the entire [data distribution](#), enabling analysts to quickly compare key metrics--such as the median, the overall variability, and the existence of extreme values or outliers--across multiple categories or experimental conditions simultaneously.

This comprehensive tutorial is specifically designed to guide you through the exact methodology required to produce high-quality, comparative boxplots using the industry-standard statistical software, [SPSS](#) (Statistical Package for the Social Sciences). By following the steps meticulously outlined in this guide, you will gain the proficiency needed to generate professional visualizations, similar to the example displayed below, significantly enhancing your ability to communicate nuanced statistical findings with precision and clarity.



We will systematically cover the necessary data preparation and structure within [SPSS](#), provide detailed menu navigation instructions, and conclude with crucial considerations for accurately interpreting the resulting comparative chart.

The Power of Comparative Data Visualization

Data visualization is an indispensable element of the initial phase of any project, often referred to as Exploratory Data Analysis (EDA). While conventional numerical summaries, such as the arithmetic mean and standard deviation, provide quantitative metrics, they frequently fail to capture critical nuances of the data, such as underlying skewness, modality, or the presence of subgroups. Boxplots address this deficiency by visually encoding the **five-number summary** of a dataset. When these plots are aligned side-by-side, they transform into an exceptionally powerful instrument for direct, simultaneous group comparison.

The fundamental advantage of employing **side-by-side boxplots** lies in their capacity to simultaneously assess differences in both central location--where the majority of the data points reside--and statistical spread or variability across several distinct groups. Consider a scenario where the performance scores of three experimental groups are being analyzed. A single glance at the boxplots immediately reveals which group exhibits the highest typical score (represented by the **median** line) and which group demonstrates the most consistent or, conversely, the most erratic performance, based on the size of the Interquartile Range (IQR). This holistic visual assessment is essential for forming robust statistical judgments before progressing to more formal inferential statistical testing.

Leveraging the graphical user interface (GUI) of **SPSS** significantly simplifies the process of generating these visualizations. SPSS excels at streamlining the often-complex task of mapping continuous and categorical variables to the appropriate aesthetic elements of a plot. Our methodology will focus on utilizing the Legacy Dialogs feature, found under the Graphs menu, as this provides the most intuitive and reliable pathway for quickly generating these foundational comparative plots. Careful attention must be paid to the correct variable placement, particularly defining the categorical axis, as this step is paramount to producing a meaningful comparative visualization.

Understanding the Boxplot's Anatomy (The Five-Number Summary)

Prior to executing the procedural steps within **SPSS**, it is absolutely essential to internalize what each component of the boxplot signifies. A meaningful and robust interpretation of the final generated graph depends entirely on the analyst's ability to recognize the statistical significance of the central box, the extending whiskers, and any separately plotted data points. Fundamentally, every boxplot is a graphical representation built upon the five critical metrics derived from the distribution of a single variable, commonly referred to as the **five-number summary**.

The horizontal line positioned centrally within the box represents the **Median** (the 50th percentile). The median is considered the most robust measure of central tendency because its value is

minimally sensitive to extreme values or **outliers** within the dataset. The upper and lower boundaries of the box itself correspond to the third quartile (Q3, the 75th percentile) and the first quartile (Q1, the 25th percentile), respectively. Consequently, the vertical length of the box defines the **Interquartile Range (IQR)**, which encompasses the middle 50% of the entire data population. The IQR is a critical metric for statistical dispersion, offering a standardized, resistant method for comparing the inherent variability across different groups.

The lines that extend outwards from the main box, conventionally known as the whiskers, typically stretch to include the minimum and maximum data values that are not classified as outliers. In the standard SPSS configuration, the whiskers usually extend a maximum distance of 1.5 times the **IQR** from the box edges (Q1 and Q3). Any individual data points that fall outside these prescribed whisker limits are then plotted separately, often appearing as small circles or asterisks. These symbols definitively signify potential **outliers**. Identifying and examining these points is crucial, as they may either flag errors during data entry or represent genuinely unusual observations that warrant deeper investigation.

Structuring the Dataset in SPSS for Comparative Analysis

To effectively demonstrate the procedure for generating comparative boxplots, we will use a practical, hypothetical example involving sports analytics. Imagine a basketball coach aiming to compare the seasonal performance of players categorized across three distinct teams: Team A, Team B, and Team C. To successfully perform this side-by-side comparison, the underlying **dataset** within SPSS must adhere to a specific structure, which is mandatory for graphical comparison. Specifically, the data requires two essential variables: a continuous, scale variable (the measurement being compared) and a nominal or ordinal grouping variable (the categories defining the comparison).

Our analytical sample dataset contains the aggregated performance scores for individual players. It must be organized into two fundamental columns:

Points: This serves as the dependent, scale variable, representing the total points scored during the season.

Team: This functions as the independent, **categorical variable**, identifying the specific team the player belongs to (A, B, or C).

This "long format" structure is indispensable for generating accurate side-by-side plots. SPSS requires clarity on which variable defines the vertical distributions that will be plotted (Points) and which variable defines the separate categories for comparison along the horizontal axis (Team). The subsequent image illustrates how this necessary structure appears in the SPSS Data View, listing the observations we are prepared to analyze.

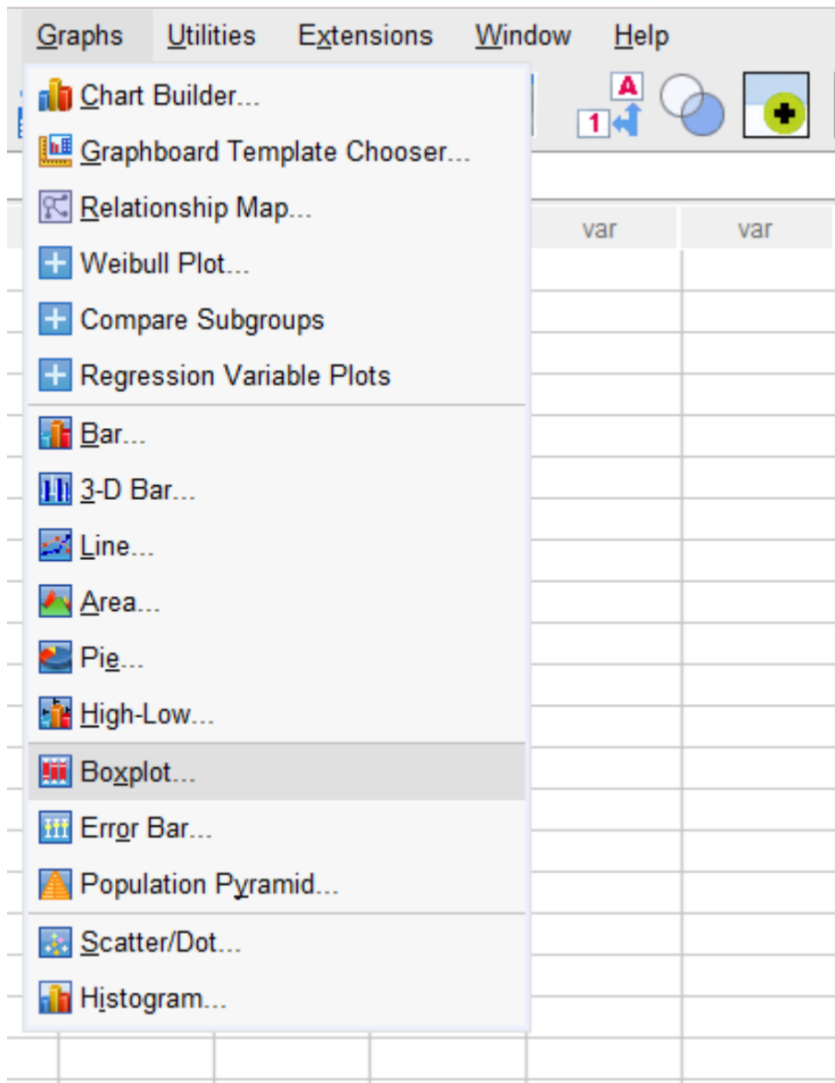
	Team	Points	var	var	var
1	A	5			
2	A	5			
3	A	6			
4	A	6			
5	A	8			
6	A	9			
7	A	13			
8	A	15			
9	B	11			
10	B	11			
11	B	12			
12	B	14			
13	B	15			
14	B	19			
15	B	22			
16	B	24			
17	C	19			
18	C	23			
19	C	23			
20	C	23			
21	C	24			
22	C	26			
23	C	29			
24	C	33			
25					
26					
27					

Our objective remains straightforward: we intend to produce a graphical summary that permits an immediate comparison of the statistical spread and central tendency of the **Points** variable, which is systematically segmented by the **Team** variable. Achieving this clear segmentation is the first step toward highlighting any differences in performance metrics across the three defined groups.

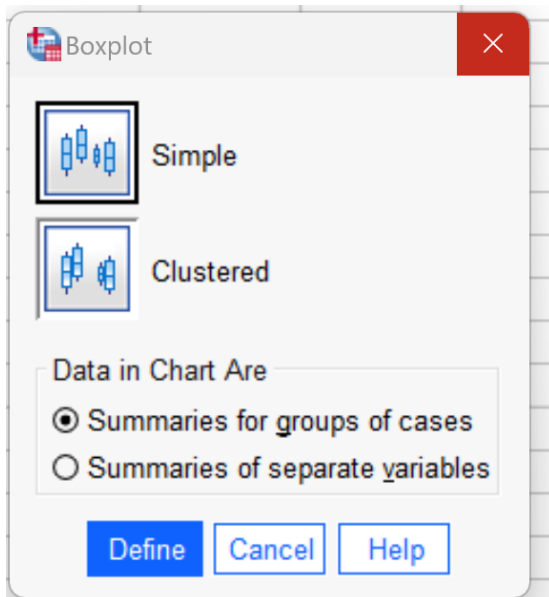
Executing the Boxplot Generation Procedure in SPSS

The process of generating the desired [side-by-side boxplots](#) in **SPSS** is streamlined through the use of the Legacy Dialogs feature, which grants the user precise control over the fundamental plot configuration necessary for comparing multiple distributions. To begin, follow this sequence of navigation steps precisely. First, locate the main menu bar at the top of the SPSS application

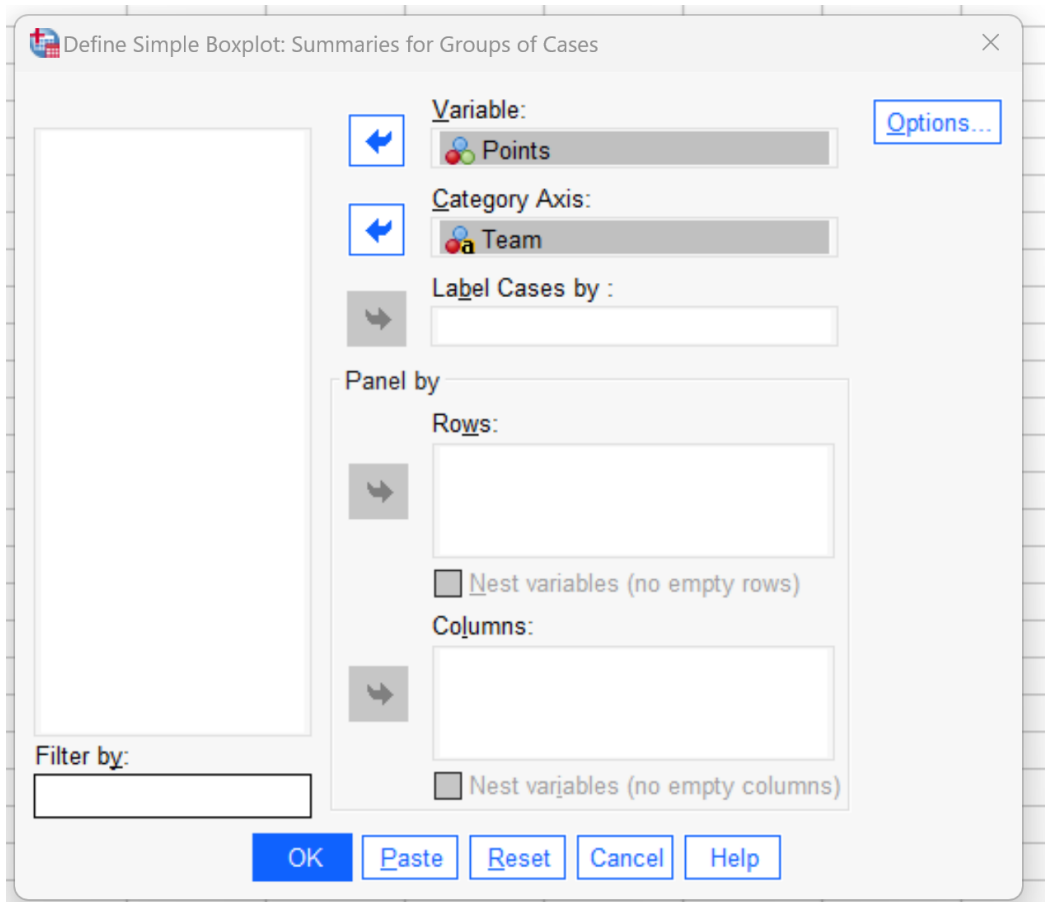
window. Click the **Graphs** tab, then hover your cursor over **Legacy Dialogs**, and finally select **Boxplot** from the resulting cascading dropdown menu. This action initiates the chart creation process, prompting SPSS to ask the user to specify the exact type of boxplot required for the comparison.



In the initial Boxplot window that appears, you must accurately specify the plot architecture. Since our goal is to compare the distribution of a single continuous variable (Points) across three distinct categorical groups (Teams), we require the simplest comparative option. Select the **Simple** icon, which designates one continuous variable for the Y-axis plotting. Crucially, ensure that the radio button option labeled **Summaries for groups of cases** is selected, as we are summarizing the 'Points' variable based on the definitions provided by the 'Team' variable. Once these two selections are confirmed, click the **Define** button to advance to the variable assignment stage.



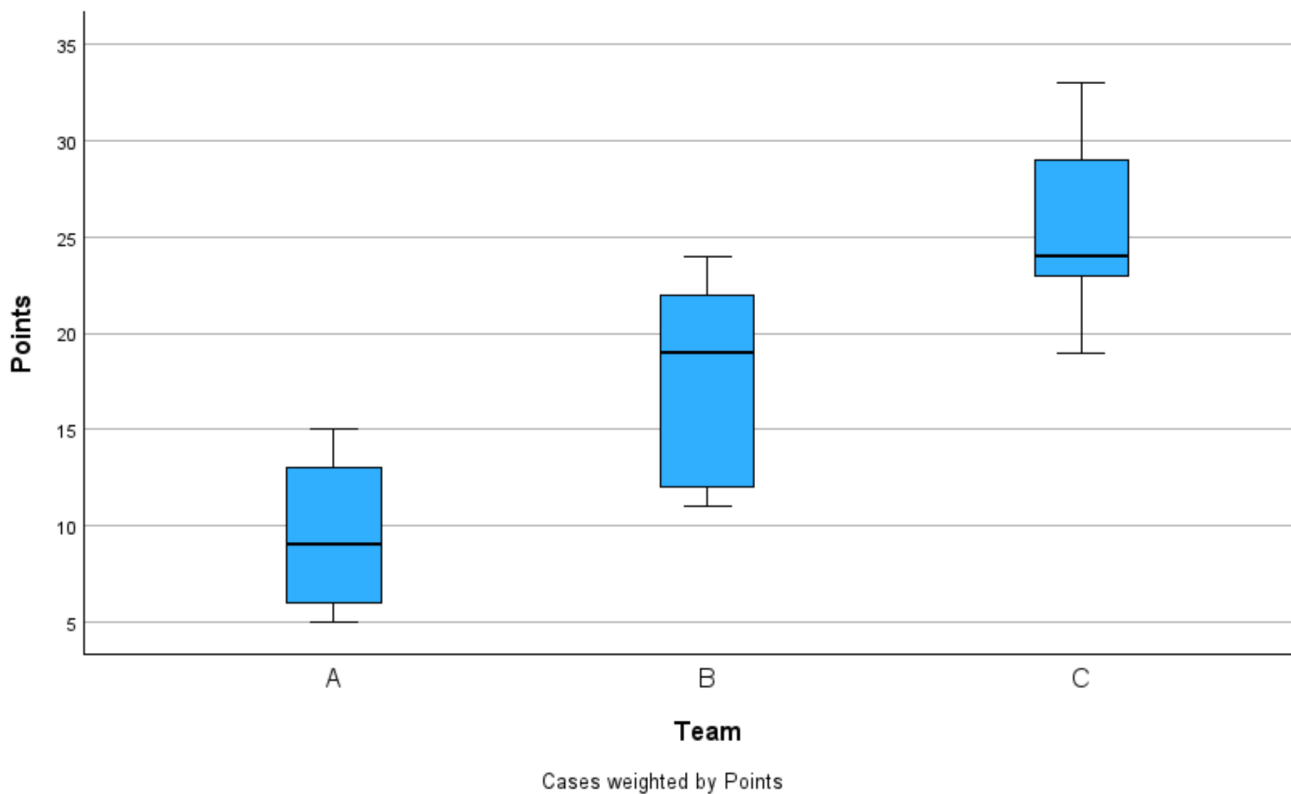
The final mandatory step involves correctly assigning the variables to their respective functional roles within the visualization. In the "Define Simple Boxplot" dialog box, locate the list of your dataset variables. Drag the continuous variable, **Points**, into the **Variable** panel. This action defines the data distribution that will be plotted along the vertical axis. Subsequently, drag the categorical variable, **Team**, into the **Category Axis** panel. This command explicitly instructs SPSS to generate separate boxplots, positioned horizontally, based on each unique value found within the 'Team' column. Once these precise assignments are finalized, click **OK** to command SPSS to generate the output and display the comparative visualization.



Analyzing and Interpreting the Comparative Boxplots

Immediately following the execution of the command, **SPSS** will render the side-by-side boxplots within the Output Viewer. This visualization is the central outcome of our analysis, providing instantaneous insights into the characteristics of each group that are often difficult to glean from raw statistical tables. Structurally, the X-axis represents the grouping or **categorical variable** (Team A, Team B, Team C), while the Y-axis represents the measurement or scale variable (Points Scored).

A primary element of analysis involves comparing the central tendency across groups, which is visually represented by the **median**--the bold horizontal line segment inside the box. By observing the plot, we can immediately establish that Team C exhibits the highest median points per player. This suggests that the typical player performance on Team C surpasses that of the players on the other two teams. Conversely, Team A displays the lowest median score, clearly indicating that its typical player performance is lagging behind its competitors. The vertical height and placement of the boxplot on the Y-axis are a direct proxy for the general performance level of that specific subgroup.



Equally critical is the assessment of statistical variability, which is quantified by the vertical size of the box itself--the **Interquartile Range (IQR)**. A taller or wider box signifies a greater spread of data and, consequently, less consistency in performance within that group. In our running example, Team B clearly possesses the largest IQR, meaning that the middle 50% of its players display the widest range of scores. This finding indicates that while Team B's median score might be reasonably competitive, its overall performance is the least consistent, reflecting a significant disparity between its 25th and 75th percentile players. Team C, despite achieving the highest median score, shows a comparatively tighter IQR, suggesting its player performance is more uniformly clustered around the central value. Furthermore, analysts must vigilantly check for potential **outliers**, which are plotted as distinct symbols outside the whisker boundaries. These points signal potentially exceptional individual scores that deviate substantially from the performance norms of their respective teams.

Conclusion and Expanding Visualization Capabilities

The competency to rapidly generate and accurately interpret **side-by-side boxplots** using **SPSS** is an essential, foundational skill for any professional data analyst. This visualization technique provides a clear, non-parametric overview of group differences, effectively circumventing the inherent vulnerabilities of relying solely on mean comparisons, which can be easily distorted by the presence of extreme values. Based on our analysis, the exercise confirms that Team C

demonstrates the strongest overall central performance, but simultaneously highlights that Team B exhibits the largest degree of internal variability in player scores.

While the simple side-by-side boxplot is an incredibly effective tool for initial comparison, the versatility of **SPSS** extends to offering more sophisticated graphical options. These alternatives include grouped boxplots (where a second categorical variable is nested within the first), as well as specialized visualizations like histograms, scatterplots, and error bar charts. The selection of the most appropriate graph should always be dictated by the specific research question being addressed. Nevertheless, for a direct, initial comparison of central tendency, statistical spread, and the identification of unusual observations across distinct groups, the methodology outlined here remains the benchmark for exploratory [data distribution](#) analysis.

We strongly encourage new users to practice this entire procedure using various datasets to firmly solidify their understanding of correct variable assignment and critical graphical interpretation. Mastery of this specific visualization tool significantly enhances your overall capability to perform robust exploratory analysis and communicate complex data findings effectively.

Additional Resources for SPSS Users

For users seeking to expand their foundational knowledge of [data visualization](#) and advanced statistical testing within the **SPSS** environment, the following tutorials explain how to perform other common and necessary tasks:

Tutorial on Generating Histograms in SPSS.

Guide to Conducting T-Tests for Group Comparisons.

Explanation of Variable Transformation Techniques in SPSS.

How to Identify and Handle [Outliers](#) in Large Datasets.