

# How to Create and Interpret Box Plots in SPSS: A Step-by-Step Guide

Authored by  
**stats writer**

March 14, 2026

## RECOMMENDED CITATION

stats writer (2026). *How to Create and Interpret Box Plots in SPSS: A Step-by-Step Guide*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=135682>

The **box plot**, also known as a box-and-whisker plot, serves as one of the most powerful tools in **Descriptive Statistics** for visualizing the distribution of numerical data. By utilizing **SPSS**, researchers and data analysts can quickly generate these plots to gain a comprehensive overview of a dataset's central tendency, spread, and potential anomalies. This graphical representation is particularly valued for its ability to condense a large volume of data into a compact form that highlights critical statistical milestones without the clutter of individual data points.

In the professional research environment, interpreting a **box plot** involves more than just looking at a rectangle; it requires a deep understanding of how the data is partitioned. **SPSS** facilitates this by automating the calculation of the **five-number summary**, which includes the minimum, the first **quartile**, the **median**, the third **quartile**, and the maximum. This summary provides a snapshot of the data's range and concentration, allowing for immediate insights into the symmetry or skewness of the underlying distribution.

Overall, the use of **SPSS** for generating box plots streamlines the process of data cleaning and exploratory analysis. Whether you are comparing test scores across different classrooms or evaluating the performance of athletes, the **box plot** offers a standardized method to communicate findings effectively. This tutorial will guide you through the intricate steps of creating these visualizations and, more importantly, how to derive meaningful conclusions from the resulting charts.

## The Statistical Foundation of the Five-Number Summary

Before diving into the software mechanics, it is essential to understand the mathematical framework that supports every **box plot**. The **five-number summary** is a collection of descriptive statistics that provide information about a dataset. The minimum and maximum represent the boundaries of the data, excluding **outliers**, while the **median** marks the 50th percentile, effectively splitting the dataset into two equal halves. This allows the researcher to see where the "center" of the data truly lies regardless of extreme values.

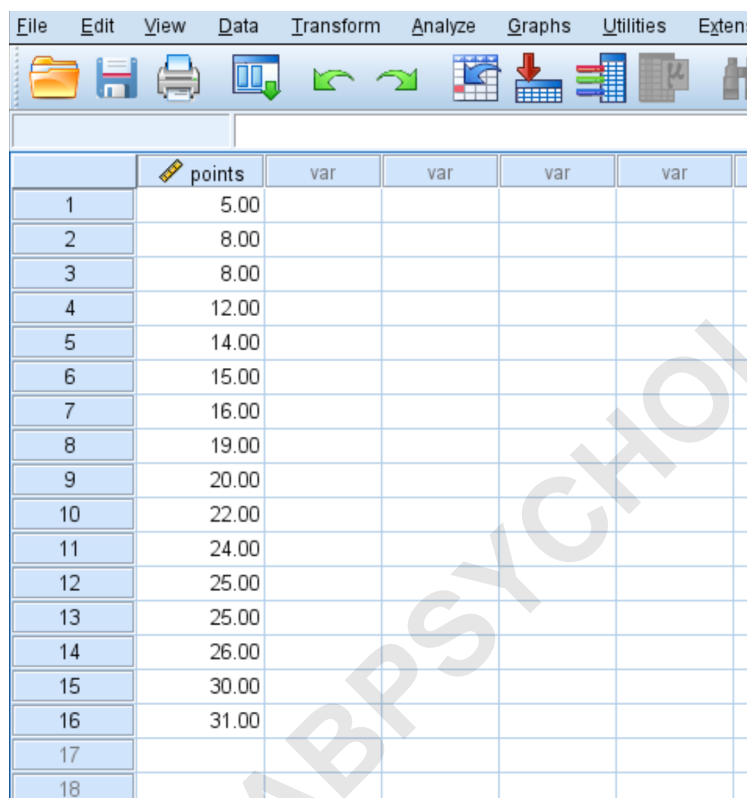
The quartiles, specifically the first **quartile** (Q1) and the third **quartile** (Q3), define the "box" in the plot. The first **quartile** represents the 25th percentile, meaning 25% of the data falls below this value. Conversely, the third **quartile** represents the 75th percentile. The distance between these two quartiles is known as the **interquartile range** (IQR), which contains the middle 50% of the data. This range is a robust measure of statistical dispersion and is less sensitive to **outliers** than the standard deviation.

Understanding these components is vital because they dictate the shape of the **box plot**. A long box indicates that the middle 50% of the data is widely spread, while a short box suggests that the data points are closely clustered around the **median**. Similarly, the "whiskers" that extend from the box to the minimum and maximum values provide a visual cue for the overall range of the dataset,

helping analysts identify how far the data stretches beyond the central concentration.

## Initial Data Setup and Preparation in SPSS

To begin our practical application, let us consider a dataset representing the average points scored per game by 16 basketball players on a specific team. Proper data entry is the first step toward accurate visualization in **SPSS**. Each player's score should be entered into a single column, which we will label as our dependent variable. Ensuring that the variable measurement level is set to "Scale" in the Variable View tab is crucial, as box plots are designed for continuous numerical data.



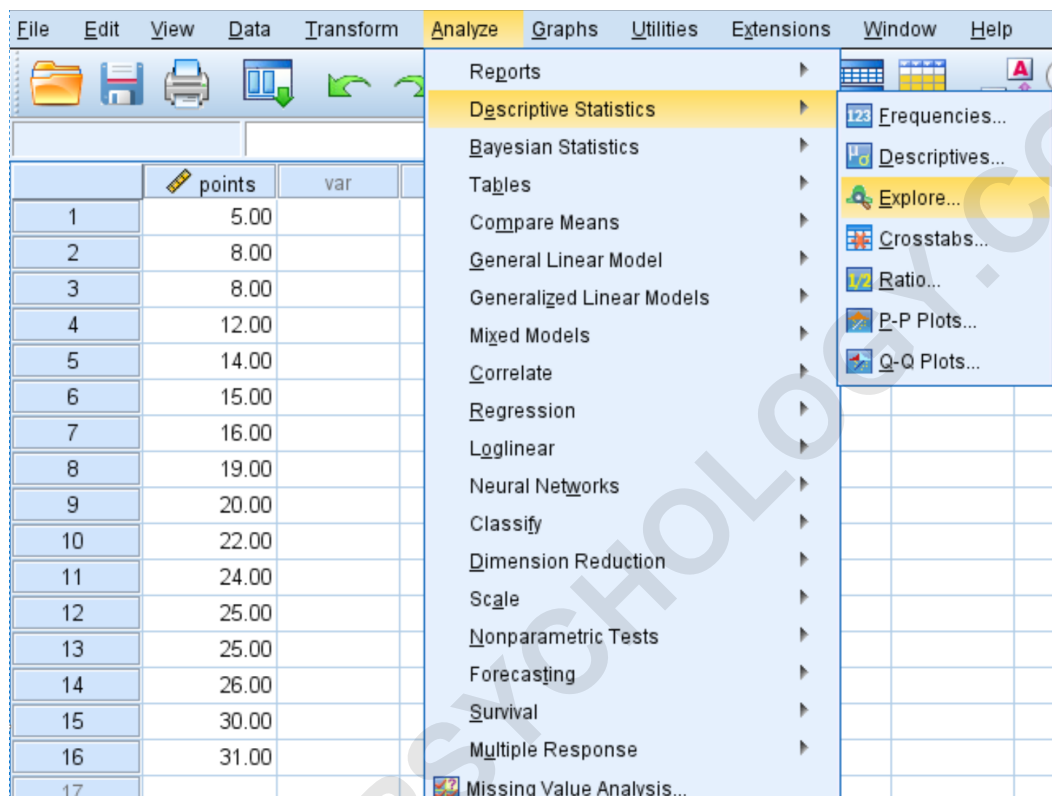
	points	var	var	var	var
1	5.00				
2	8.00				
3	8.00				
4	12.00				
5	14.00				
6	15.00				
7	16.00				
8	19.00				
9	20.00				
10	22.00				
11	24.00				
12	25.00				
13	25.00				
14	26.00				
15	30.00				
16	31.00				
17					
18					

Once the data is accurately entered, the next phase involves navigating the **SPSS** interface to access the visualization tools. While there are multiple ways to generate graphs in **SPSS**, using the "Explore" function is often the most efficient method for box plots. This is because the Explore command provides a comprehensive set of **Descriptive Statistics** alongside the chart, allowing you to verify the numerical values that correspond to the visual elements of the plot.

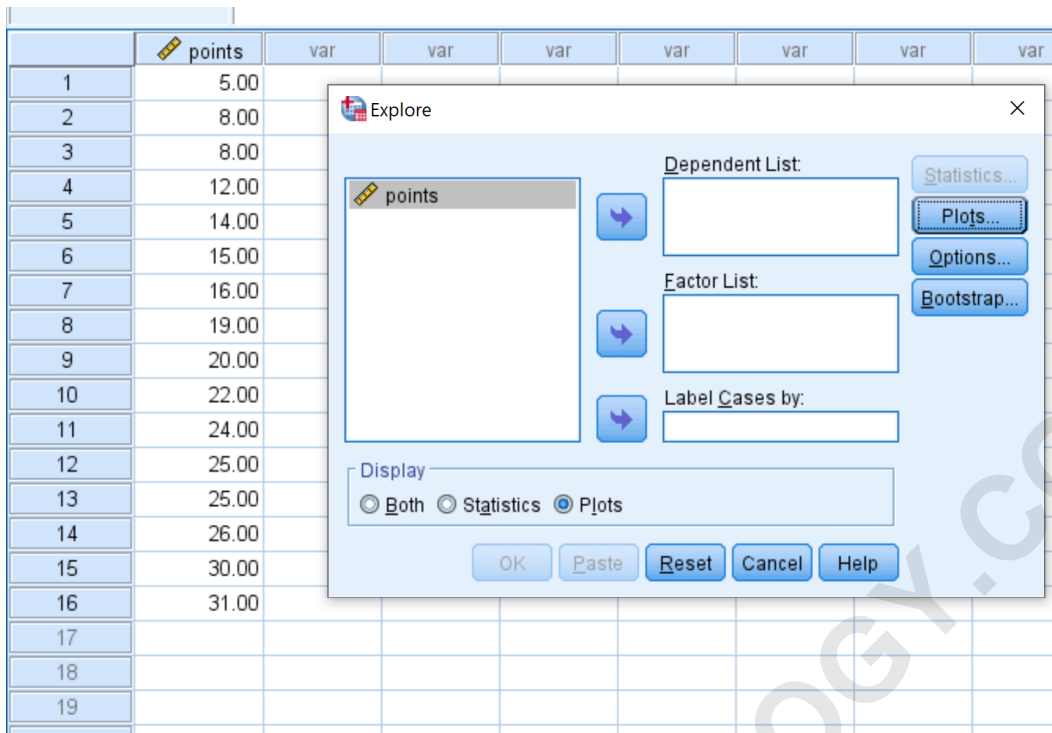
The reliability of your **box plot** depends entirely on the quality of the input data. Before generating the plot, it is a best practice to scan the data for obvious entry errors. For instance, if a basketball player's score is recorded as 500 instead of 50, this error will manifest as an extreme **outlier**, skewing the entire plot and leading to incorrect interpretations. Spending a few moments on data verification ensures that the subsequent visual analysis remains valid and professional.

## Step-by-Step Procedure for Creating a Single Box Plot

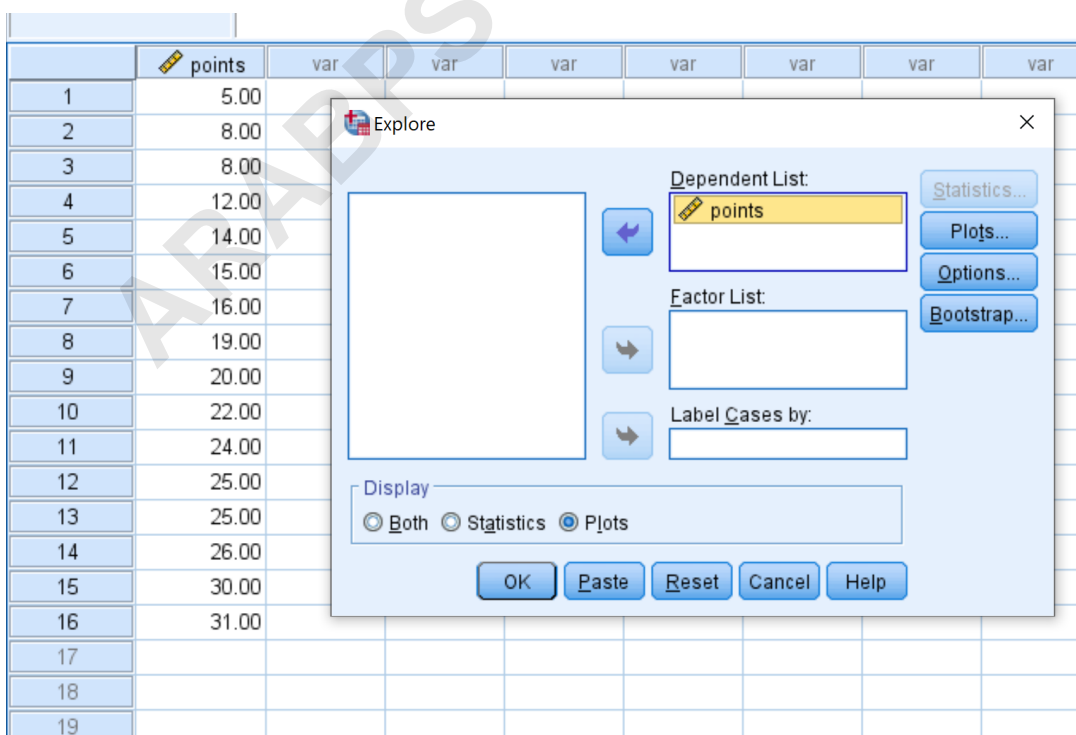
To create a **box plot** for a single variable in **SPSS**, follow a structured workflow. First, navigate to the top menu bar and click on the **Analyze** tab. From the resulting dropdown menu, hover over **Descriptive Statistics** and select the **Explore** option. This specific command is ideal for generating detailed summaries and visual distributions of your scale variables.



Upon selecting Explore, a new dialog window will appear. You will see a list of your variables on the left side. Locate the variable you wish to analyze--in this case, "points"--and click the arrow button to move it into the box labeled **Dependent List**. The Dependent List is where you place the variables you want to describe or visualize. It is important to ensure that only scale variables are placed here for box plot generation.

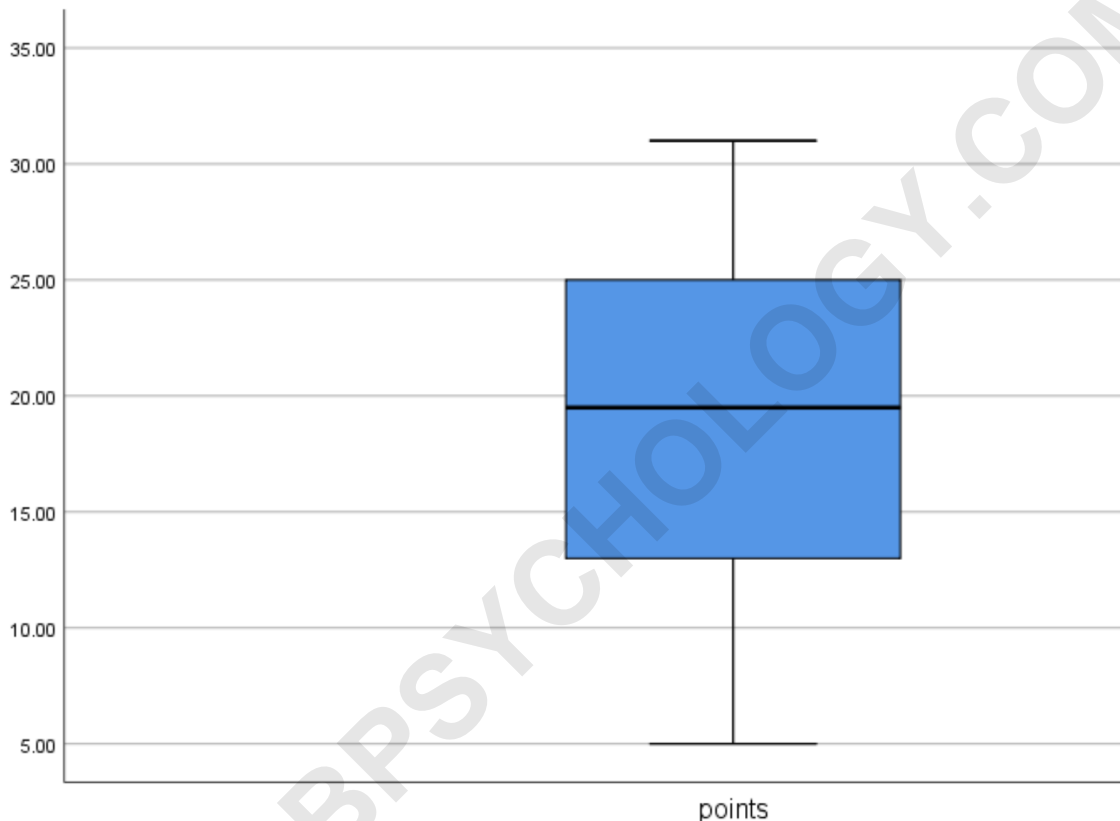


Before finalizing the procedure, look at the **Display** section at the bottom left of the Explore window. You have the option to display "Both" (statistics and plots), "Statistics," or "Plots" only. To focus specifically on the visualization, you can select **Plots**. Once your selections are made, click the **OK** button. **SPSS** will then process the data and generate the output in the Viewer window.

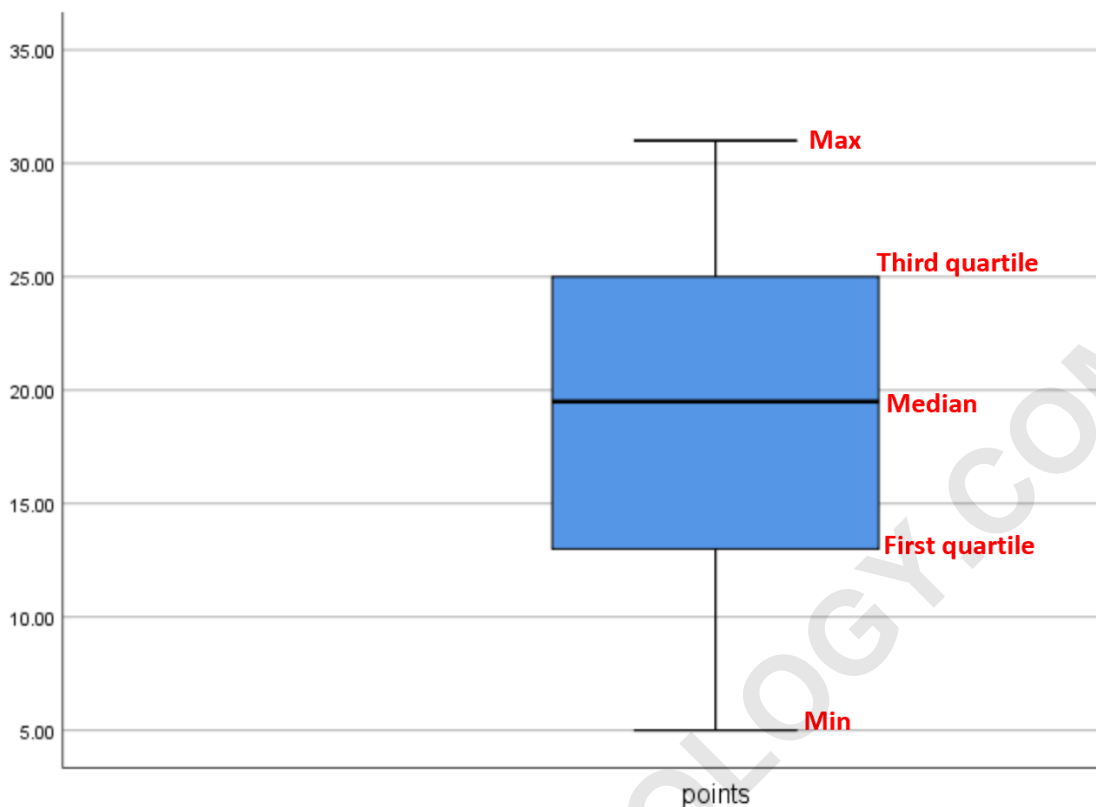


## Decoding the Visual Elements of the Output

Once the **SPSS** output viewer opens, you will be presented with the generated **box plot**. Interpreting this chart requires a systematic approach to each of its components. The thick horizontal line inside the box is the **median** of the dataset. This line represents the middle value; if the **median** is not centered within the box, it indicates that the data is skewed. For example, a **median** closer to the bottom of the box suggests a positive skew.



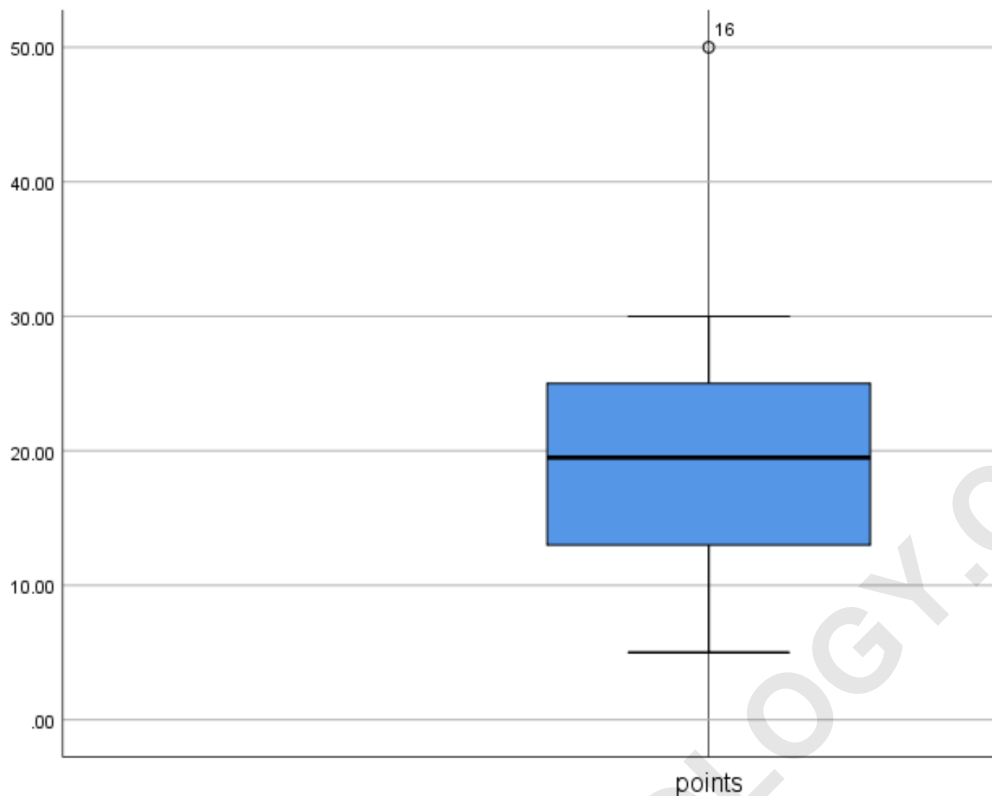
The top and bottom boundaries of the box represent the third and first **quartiles**, respectively. The vertical height of the box itself is the **interquartile range**. Analysts look at this height to understand the volatility of the data; a taller box indicates a higher degree of variability among the middle 50% of observations. The "whiskers" extending above and below the box reach out to the most extreme data points that are not considered **outliers**.



By examining the relationship between the whiskers and the box, you can determine the overall spread of the dataset. If the whiskers are of unequal length, it provides further evidence of data asymmetry. This visual summary is exceptionally useful for comparing the distribution of your data against a theoretical normal distribution, allowing you to make informed decisions about which statistical tests (parametric or non-parametric) might be appropriate for further analysis.

### Advanced Analysis: Identifying and Managing Outliers

One of the most critical functions of a **box plot** is its ability to highlight **outliers**. In **SPSS**, **outliers** are typically defined as values that fall more than 1.5 times the **interquartile range** away from either the first or third **quartile**. These are displayed as small circles or asterisks beyond the whiskers. The presence of these points can significantly impact the mean and standard deviation, making the box plot an essential tool for identifying data points that may need further investigation.



If your **box plot** reveals **outliers**, you must decide how to handle them. The first step is always to verify the data entry. An **outlier** is often the result of a simple typo or a measurement error. If the value is indeed an error, it should be corrected or removed. However, if the **outlier** is a legitimate data point, it may represent a rare but important phenomenon that warrants its own analysis rather than being discarded.

When dealing with legitimate **outliers** that threaten to skew your overall results, you have several sophisticated options:

**Verification:** Ensure the **outlier** is not a data entry error by cross-referencing with the original data source.

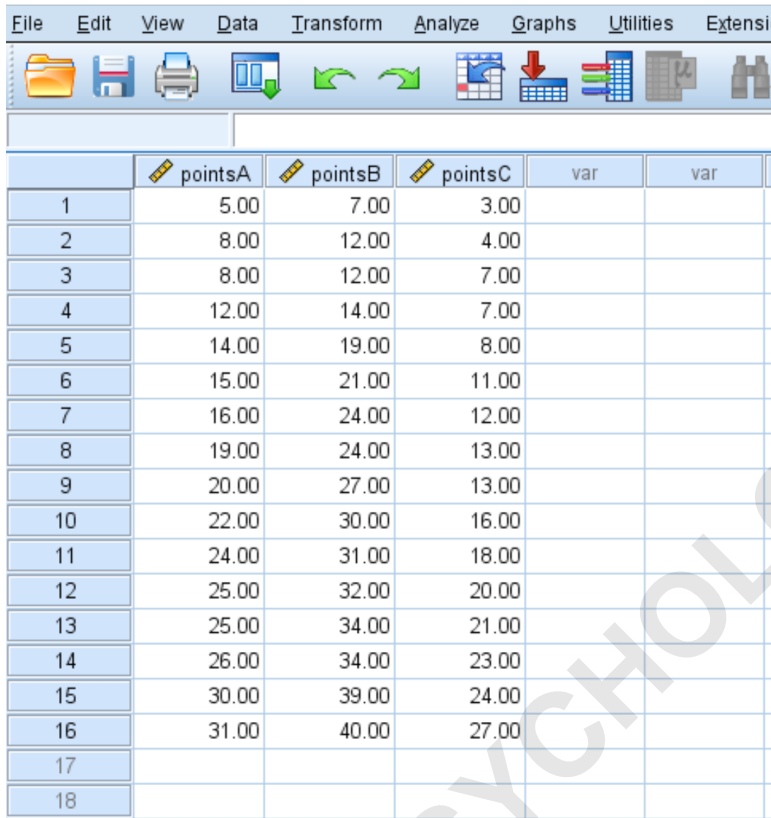
**Winsorization:** Instead of removing the point, you might assign it a new value, such as the value of the nearest non-outlier, to minimize its influence.

**Exclusion:** If the **outlier** is a true anomaly that does not represent the population, you may choose to remove it from the analysis, provided you document this decision transparently in your research report.

## Comparative Visualizations: Creating Multiple Box Plots

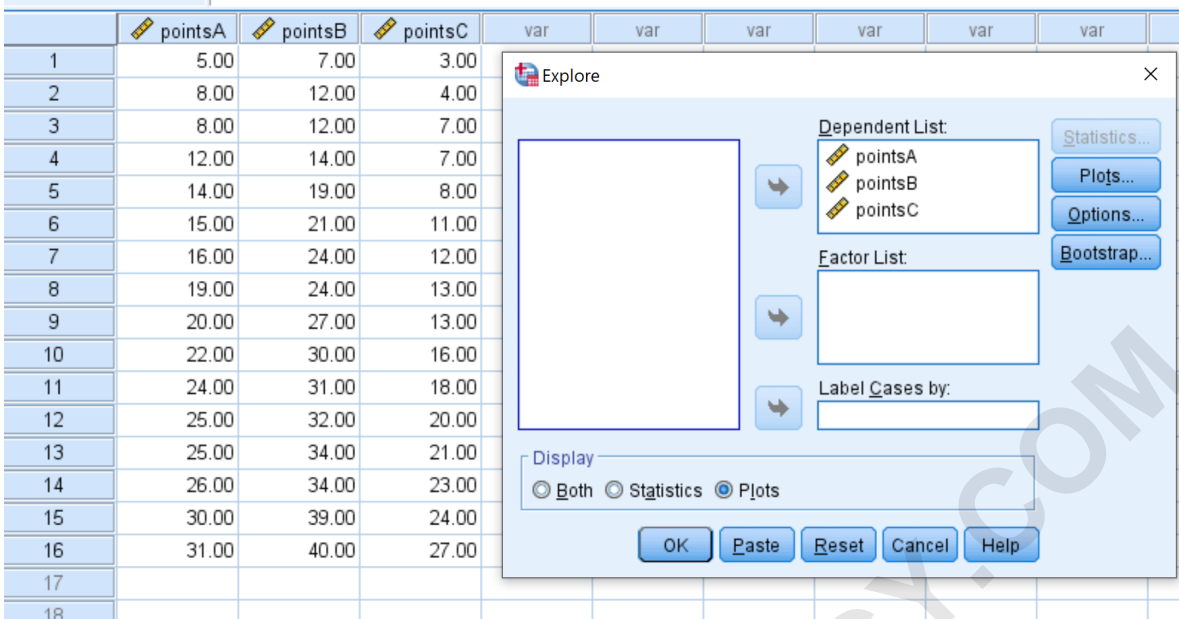
While analyzing a single variable is useful, the true power of the **box plot** is often realized when

comparing multiple groups side-by-side. Suppose you have data on points scored by players from three different teams. **SPSS** allows you to generate multiple box plots within the same chart, providing an immediate visual comparison of the **median**, **interquartile range**, and overall spread for each group.

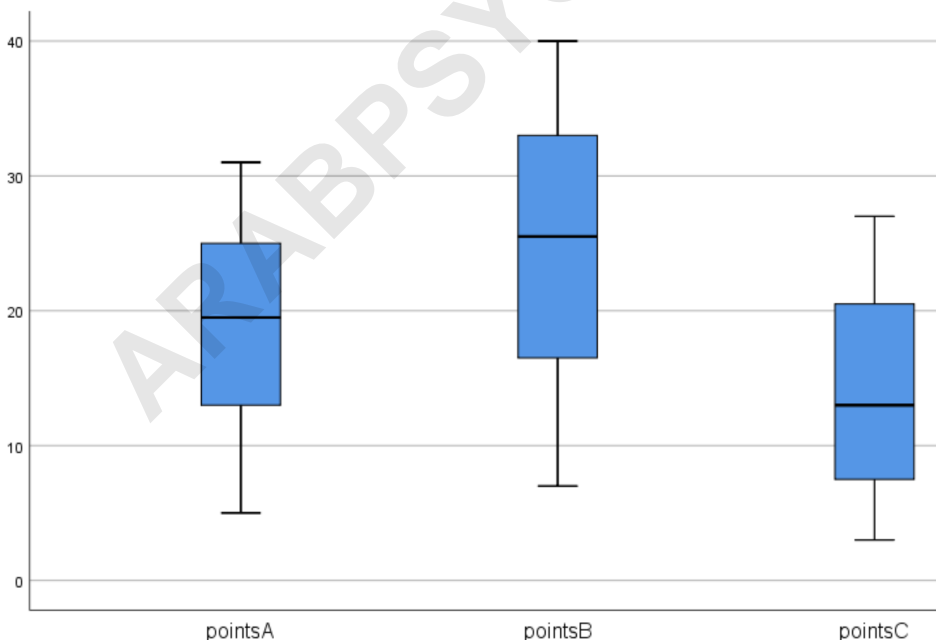


	pointsA	pointsB	pointsC	var	var
1	5.00	7.00	3.00		
2	8.00	12.00	4.00		
3	8.00	12.00	7.00		
4	12.00	14.00	7.00		
5	14.00	19.00	8.00		
6	15.00	21.00	11.00		
7	16.00	24.00	12.00		
8	19.00	24.00	13.00		
9	20.00	27.00	13.00		
10	22.00	30.00	16.00		
11	24.00	31.00	18.00		
12	25.00	32.00	20.00		
13	25.00	34.00	21.00		
14	26.00	34.00	23.00		
15	30.00	39.00	24.00		
16	31.00	40.00	27.00		
17					
18					

To create these comparative plots, return to the **Explore** dialog box. Instead of dragging just one variable into the **Dependent List**, you can select multiple scale variables. Alternatively, if your data is organized by a grouping variable (such as "Team Name"), you would place the numerical variable in the **Dependent List** and the grouping variable in the **Factor List**. This tells **SPSS** to split the box plots based on the categories within that factor.



Once you click **OK**, the output will display the box plots adjacent to one another. This layout is invaluable for identifying trends across groups. For instance, you might notice that while Team B has a higher **median** score than Team A, it also has a much larger **interquartile range**, suggesting that Team B's performance is more inconsistent. Such insights are much harder to glean from raw numbers alone.



## Synthesizing Insights from Comparative Box Plots

When viewing multiple box plots, several key observations can be made to drive data-driven decision-making. First, look at the relative positions of the **median** lines. In our basketball example, if Team B's **median** is significantly higher than Team C's, it suggests a general performance advantage. However, the **median** only tells half the story; we must also consider the dispersion.

The variation in the data is clearly visible through the length of the boxes. A longer box for Team B indicates that there is a wider variation in the points scored per game among its players compared to Team A or C. High variation can be a sign of a team with a few star players and several low scorers, whereas a compact box suggests a more balanced team where all players contribute similar point totals. These nuances are vital for coaches or analysts looking to optimize performance.

Finally, identify the extreme values at the ends of the whiskers. The player with the highest points per game across the entire dataset can be identified by the top whisker (or **outlier**) of the respective team. By synthesizing the **median**, **interquartile range**, and extremes, box plots provide a multi-dimensional view of the data that is both comprehensive and easy to communicate to non-technical stakeholders.

## Conclusion and Best Practices for Data Visualization

Mastering the **box plot** in **SPSS** is a fundamental skill for any data analyst. These plots offer a unique blend of **Descriptive Statistics** and visual clarity, making them indispensable for exploratory data analysis. By following the structured steps of data preparation, execution, and interpretation, you can ensure that your statistical findings are both accurate and impactful.

Always remember that the goal of a **box plot** is to simplify complexity without losing the essential characteristics of the data. Use them to check for **outliers**, compare distributions, and verify assumptions before moving on to more complex inferential statistics. With the robust tools provided by **SPSS**, creating professional-grade box plots becomes a seamless part of your research workflow.

In summary, whether you are analyzing a single variable or comparing several groups, the **box plot** remains one of the most reliable methods for visualizing the **five-number summary**. By integrating these visualizations into your regular data processing habits, you will enhance your ability to spot patterns, identify errors, and communicate the story behind your data with greater confidence and precision.