

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

Authored by
stats writer

December 5, 2025

RECOMMENDED CITATION

stats writer (2025). *How to Create Side-by-Side Boxplots in Excel: A Step-by-Step Guide*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=105938>

The ability to visualize and compare data distributions is paramount in effective statistical analysis. Creating side-by-side boxplots in Microsoft **Excel** provides an efficient method for achieving this comparative visualization. The process, while automated within recent Excel versions, requires careful data preparation and specific navigation of the charting tools.

To initiate the creation of these plots, the analyst must first ensure the datasets are correctly organized in columns. The core process involves selecting the target data range, navigating to the **Insert** tab on the ribbon, and selecting the dedicated **Box and Whiskers** chart type. Unlike creating a single plot, generating a side-by-side comparison leverages Excel's understanding of data organized across multiple columns, automatically treating each column as a distinct variable for plotting.

This comprehensive guide details the precise steps required to generate, customize, and effectively interpret comparative boxplots, turning raw data into actionable visual insights. Understanding the underlying statistical concepts, particularly the five-number summary, is crucial for meaningful interpretation.

Introduction to Boxplots and Comparative Analysis

A box plot, often referred to as a box-and-whisker plot, stands as an essential tool in exploratory data analysis. Its primary function is to graphically summarize the distribution of a numerical dataset, making the identification of central tendency, spread, and potential outliers straightforward. Unlike histograms, boxplots are particularly effective when the goal is to compare the shape and location of multiple distributions simultaneously.

The structure of the box plot is meticulously designed to display the five-number summary of the dataset, providing a robust, non-parametric view of the data's statistical properties. This summary provides a foundational understanding of data dispersion without requiring assumptions about the underlying data distribution.

When creating a side-by-side visualization, we are essentially placing the visual summaries of two or more independent datasets next to each other. This spatial arrangement allows for immediate qualitative comparisons regarding skewness, the presence of extreme values, and differences in the interquartile range across the groups being studied. The following detailed example demonstrates the workflow within **Excel**.

Understanding the Five-Number Summary

The cornerstone of every boxplot is the comprehensive set of five descriptive statistics that define its shape. These values delineate the range and concentration of the data points, enabling a visual

assessment of the distribution. These critical components are:

Minimum: The smallest observation in the dataset, excluding any defined outliers. This marks the end of the lower whisker.

First Quartile (Q1): Also known as the 25th percentile. This value represents the point below which 25% of the data falls. The box begins here.

Median (Q2): The central value of the dataset, or the 50th percentile. This is represented by the line dividing the box and indicates the central tendency.

Third Quartile (Q3): Also known as the 75th percentile. This value marks the point below which 75% of the data falls. The box ends here.

Maximum: The largest observation in the dataset, excluding any defined outliers. This marks the end of the upper whisker.

By illustrating these five values, a boxplot provides a powerful summary of the distribution of values in a given dataset. Specifically, the length of the box (the distance between Q1 and Q3, known as the Interquartile Range or IQR) visually represents the spread of the central 50% of the data.

The whiskers extend to capture the full range of data, providing context for the extremes, while points lying outside the whiskers are typically highlighted as potential outliers, calculated by Excel based on 1.5 times the Interquartile Range beyond the nearest quartile.

Preparing Your Data for Comparison

Before initiating the chart creation process, meticulous attention must be paid to how the data is structured within the Excel spreadsheet. For Excel to recognize and plot variables as distinct, comparable groups in a side-by-side format, the data must be arranged vertically in adjacent columns. Each column should represent one dataset or variable for comparison.

Crucially, ensure that the columns containing your numerical data are clearly labeled with descriptive headers. These headers will automatically be used by Excel as the labels for the resulting boxplots, making the final visualization immediately understandable. While the length of the datasets does not need to be uniform, having consistent variable names is highly recommended for clarity and professionalism.

For this tutorial, we will utilize three distinct datasets, each representing a separate experimental condition or group (Dataset 1, Dataset 2, and Dataset 3). This setup is ideal for demonstrating the comparative power of the side-by-side boxplot structure for analyzing differences in central tendency and dispersion.

Step 1: Entering the Datasets into Excel

The foundational step involves accurately populating the spreadsheet with the values intended for analysis. We will designate columns A, B, and C for our three datasets. It is good practice to include row 1 for the column headers, ensuring proper labeling of the chart series later on.

Let's proceed by entering the values for our three hypothetical datasets into cells A1 through C21, mirroring the structure required for Excel's plotting function:

	A	B	C	D	E	F	G
1	Dataset 1	Dataset 2	Dataset 3				
2	4	9	13				
3	5	10	14				
4	5	11	14				
5	6	11	15				
6	7	11	15				
7	12	12	16				
8	13	13	17				
9	14	13	19				
10	14	14	22				
11	15	14	23				
12	16	15	23				
13	17	15	24				
14	17	16	26				
15	19	17	27				
16	20	18	29				
17	24	19	29				
18	25	22	30				
19	25	22	30				
20	28	22	31				
21	29	23	33				
22							
23							
24							
25							

Ensure that there are no empty rows within the dataset range, as this might disrupt Excel's automatic data series detection, potentially leading to incorrect chart generation or misinterpretation of the data boundaries. Consistent formatting of numerical values will also prevent charting errors.

Step 2: Initiating the Box & Whisker Chart

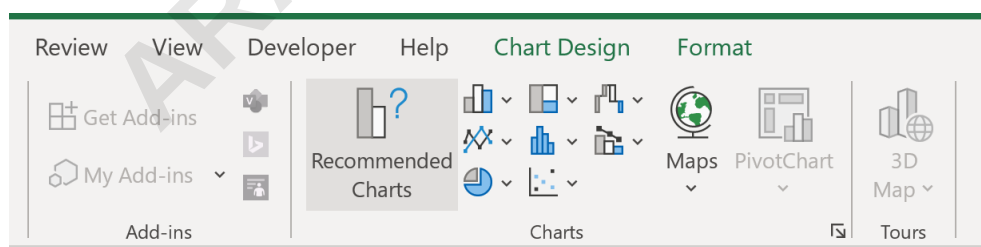
With the data correctly structured, the next phase is navigating the Excel interface to select and generate the visualization. This process is straightforward in modern versions of Excel (2016 and later), which include the specialized **Box & Whisker** chart type specifically designed for statistical summaries.

First, execute a precise selection of the data. Highlight the entire range encompassing all data

points and their corresponding headers. In this example, you must highlight cells **A1:C21**. This selection informs Excel exactly which variables (columns) should be compared.

	A	B	C	D	E	F	G	H
1	Dataset 1	Dataset 2	Dataset 3					
2	4	9	13					
3	5	10	14					
4	5	11	14					
5	6	11	15					
6	7	11	15					
7	12	12	16					
8	13	13	17					
9	14	13	19					
10	14	14	22					
11	15	14	23					
12	16	15	23					
13	17	15	24					
14	17	16	26					
15	19	17	27					
16	20	18	29					
17	24	19	29					
18	25	22	30					
19	25	22	30					
20	28	22	31					
21	29	23	33					
22								
23								
24								
25								
26								

After the selection is confirmed, direct your attention to the main menu ribbon. Click the **Insert** tab located at the top of the Excel window. This action reveals the gallery of available chart options. Next, within the Charts group, click **Recommended Charts**.

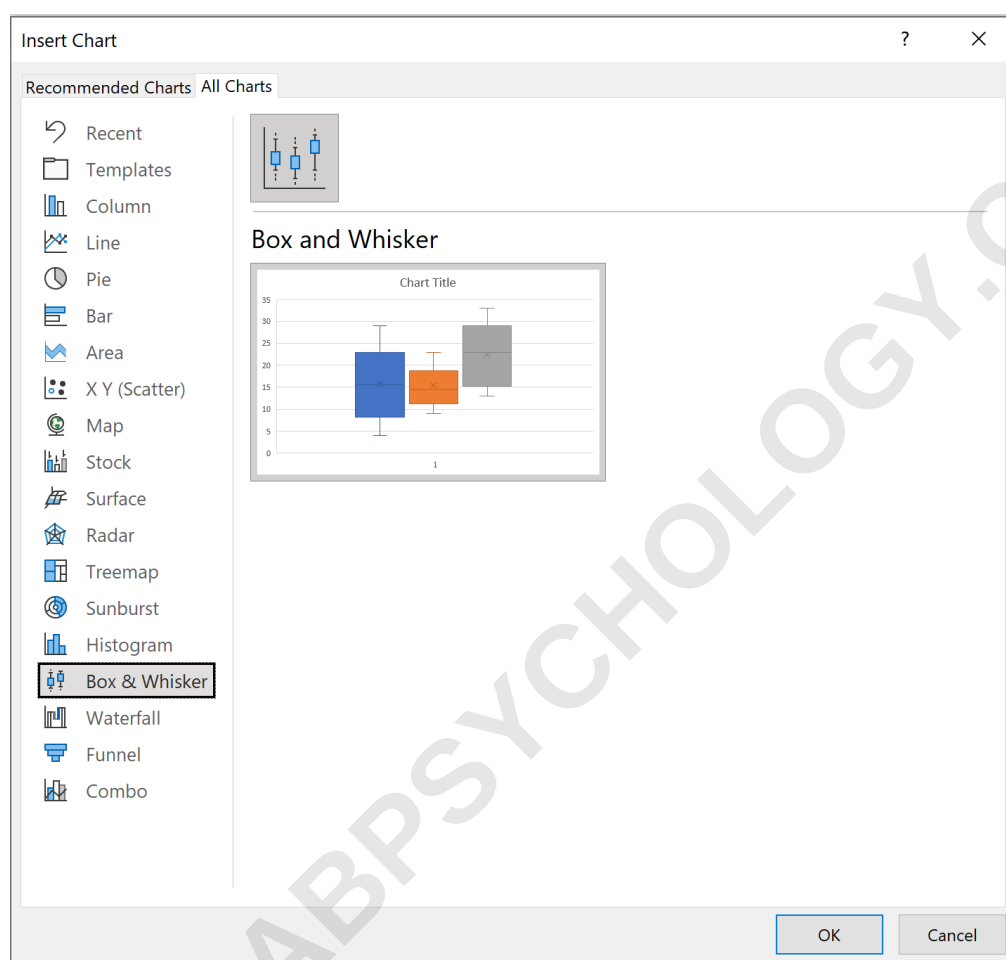


Step 3: Generating the Side-by-Side Visualization

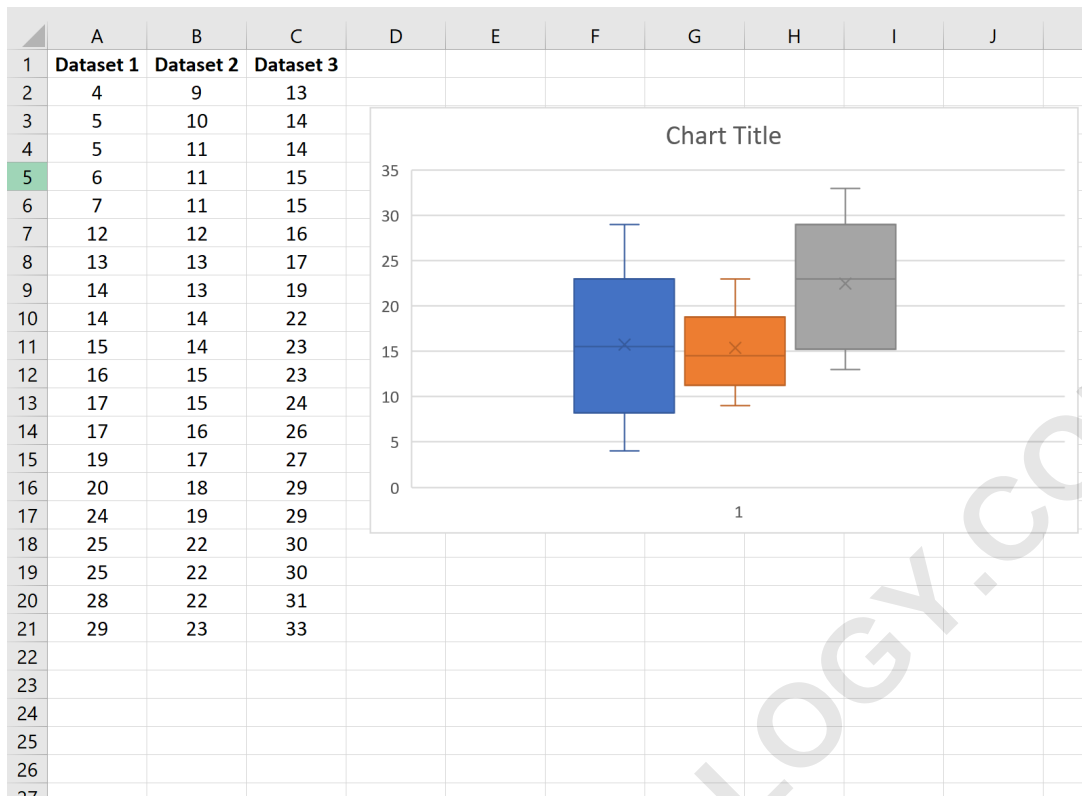
Inside the **Recommended Charts** dialogue box, navigate to the **All Charts** tab. Scroll down the list of chart categories, locate, and select **Box & Whisker**. Excel will generate a preview of the

chart, showing the three distinct datasets plotted side-by-side based on the selected range.

Confirm that the preview displays the intended comparative structure, with the variable names (Dataset 1, 2, 3) correctly displayed on the horizontal axis. Then, click **OK** to render the chart onto the worksheet. This automated process leverages built-in statistical functions to calculate the minimum, maximum, quartiles, and median for each dataset independently.



The immediate result will be the generation of the side-by-side boxplots, which graphically summarize the three data distributions:

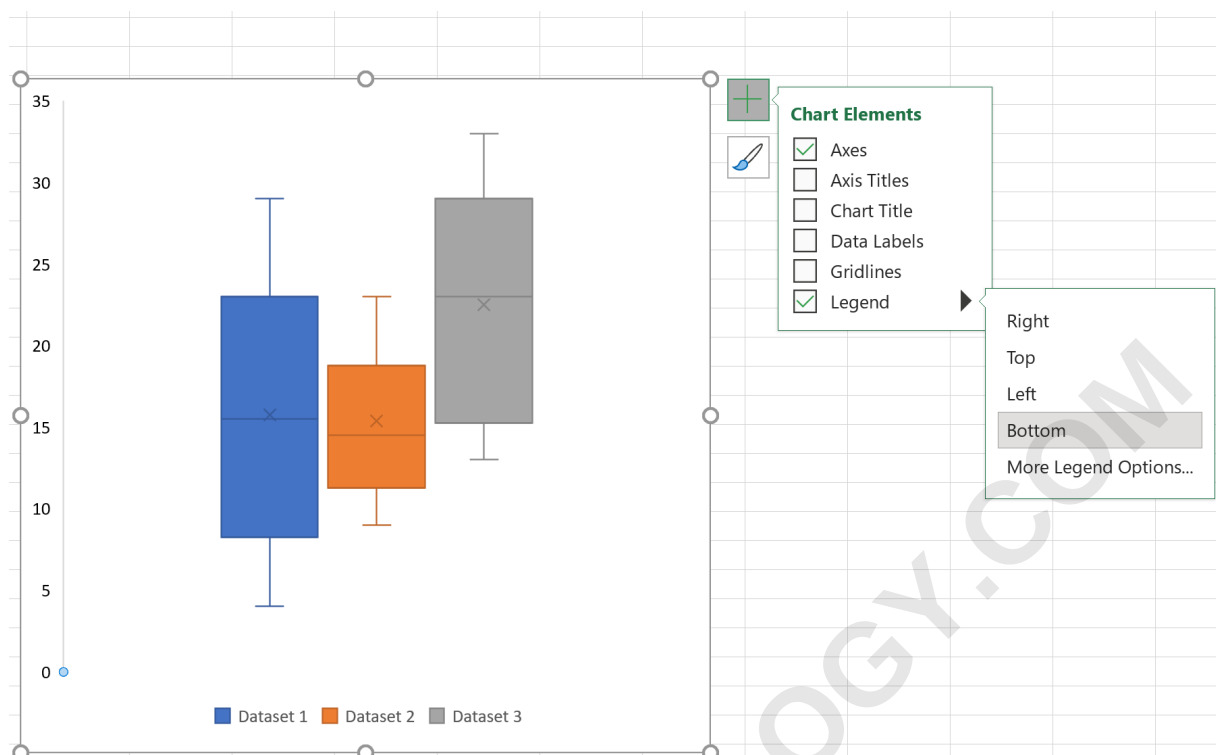


Step 4: Enhancing Chart Readability and Aesthetics

To optimize the visual impact and professional quality of the visualization, customization is necessary. Standard Excel formatting often includes elements that can clutter the visual field, such as default gridlines. Customizing the plot ensures the focus remains squarely on the distribution characteristics.

First, enhance the chart's cleanliness by removing unnecessary visual noise. Click directly on the horizontal gridlines within the plot area, and press the **Delete** key. This simplifies the background, making the box and whiskers stand out more prominently against a clean canvas.

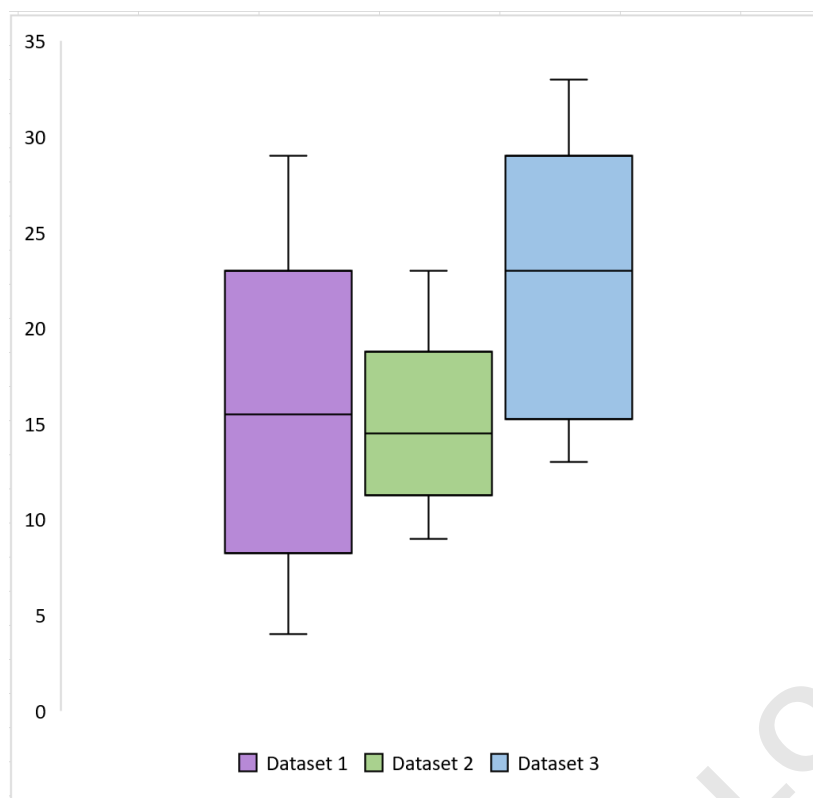
Next, we must incorporate a clear legend for proper identification. Click the green plus symbol (the **Chart Elements** icon) located in the top right corner of the chart boundary. Check the box labeled **Legend** and select **Bottom** as the preferred position.



Lastly, feel free to click on any of the individual boxplots and change the colors to anything that you'd like. Selecting distinct, contrasting colors is particularly useful when comparing many variables, as it aids in immediate differentiation between the groups and improves accessibility.

Interpreting the Comparative Boxplots

Once the customization is complete, the final side-by-side boxplots offer a clear and rich representation of the data. Our refined visualization should appear as follows, ready for analysis:



The primary advantage of this visualization is the ease with which key statistical metrics can be compared across groups. Analysts should focus on three primary dimensions for comparison: central tendency, spread (or variance), and symmetry.

We can derive several critical observations immediately from inspecting the finalized plots:

Assessment of Variability: The overall length of the boxplot (whiskers included) and the length of the box itself (IQR) reflect the dispersion or spread of the data. Dataset 1 exhibits the highest variance (since it is the longest boxplot), indicating a wider range of values compared to the others.

Assessment of Central Tendency: The location of the horizontal line within the box indicates the median. Dataset 3 clearly possesses the highest median value (as indicated by the horizontal bar positioned highest on the y-axis). This suggests that the center of Dataset 3's distribution is greater than the centers of Datasets 1 and 2.

Assessment of Consistency: Dataset 2, being the shortest boxplot both in IQR and overall length, displays the lowest variance, meaning the data points are tightly clustered around the central median.

Advanced Interpretation of Spread and Skewness

The length of the whiskers relative to the box provides additional context on the tails of the distribution. Long whiskers suggest that data points extend far from the central quartiles. In our

example, the large spread of Dataset 1 implies a high degree of heterogeneity within that group, demonstrating that the variance is significantly higher than in the other groups.

In contrast, the very narrow box and short whiskers of Dataset 2 imply a highly homogeneous distribution. This low level of variance is often desirable in controlled studies or quality control applications, indicating consistency across observations. The side-by-side arrangement of these box plots allows for statistical conclusions to be drawn rapidly. For instance, if the boxes of two datasets do not overlap, it is a strong indication that the central tendency (median) of the two distributions is statistically different, prompting further rigorous testing.

Summary and Application of Comparative Analysis

Mastering the creation of side-by-side boxplots in **Excel** transforms raw data summaries into powerful comparative visual tools. This methodology streamlines the analysis of multiple datasets, allowing researchers and analysts to quickly determine differences in spread, central location, and shape.

By following the structured steps--from data preparation and chart generation to meticulous customization and careful interpretation of the five-number summary--users can effectively communicate complex distributional characteristics. This visual approach is indispensable for presenting clear evidence of variance or median disparities between groups and is a foundational technique in data visualization.