

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

Authored by
Mohammed loot

November 4, 2025

RECOMMENDED CITATION

Mohammed loot (2025). *Learning to Create Side-by-Side Boxplots in Excel: A Step-by-Step Guide*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=10144>

Understanding the Boxplot and the Five-Number Summary

A [boxplot](#), often formally recognized as a box-and-whisker plot, stands as an essential standardized visual tool for summarizing the distribution of quantitative data. This powerful graphical representation is constructed entirely from the dataset's [five-number summary](#), offering immediate insights into data centralization, symmetry (or skewness), and the presence of potential statistical outliers. Consequently, the boxplot is an indispensable component of successful [Exploratory Data Analysis](#).

The core utility of the boxplot is its capacity to distill extensive numerical information into a highly interpretable visual format. It provides a non-parametric view of distribution, allowing analysts to quickly grasp the spread and location of the data without relying on assumptions about the underlying distribution type. Mastery of the five defining metrics is crucial for accurate interpretation and for deriving statistically sound conclusions from the chart.

The foundational [five-number summary](#) that dictates the shape and position of every boxplot is comprised of the following key statistical values:

Minimum Value: This is the smallest observed data point, calculated after excluding any designated outliers. It defines the endpoint of the lower whisker.

First Quartile (Q1): Representing the 25th percentile, this value marks the lower boundary of the box, indicating that 25% of the data falls below this point.

[Median](#) (Q2): Also known as the 50th percentile, the median is the central tendency measure that perfectly divides the dataset into two equal halves. It is marked by a line within the box.

Third Quartile (Q3): Representing the 75th percentile, this value defines the upper boundary of the box, meaning 75% of the data lies below this point.

Maximum Value: This is the largest observed data point, excluding outliers, and determines the extent of the upper whisker.

By visually encoding these five descriptive statistics, the boxplot delivers immediate clarity regarding the data's central location, the density of observations (Interquartile Range, or IQR), and the overall range of values. This efficiency makes it a preferred visualization for comparative statistical work.

The Analytical Advantage of Side-by-Side Comparison

While a solitary boxplot effectively summarizes a single data distribution, its true power in analytical contexts is realized when multiple boxplots are arrayed in a side-by-side configuration. This setup

facilitates a direct, simultaneous comparison of the distributional characteristics of two or more distinct datasets, making differences and similarities instantaneously apparent.

The capacity for comparative analysis is vital across numerous professional disciplines, spanning from rigorous medical trials to industrial quality control and complex financial modeling. When distributions are compared visually, researchers can rapidly identify significant discrepancies in the [median](#), the spread (indicated by the Interquartile Range, defined by the length of the box), and the presence or severity of outliers across different groups.

Such visual comparisons often reveal patterns of performance or inherent characteristics that would otherwise be difficult to detect or quantify solely through reviewing dense statistical tables. For instance, comparing the boxplots of two manufacturing processes can instantly show which process yields results with lower [variance](#) (higher consistency) and which process achieves a higher central value.

The following detailed guide provides a precise, step-by-step methodology for generating these powerful comparative visualizations efficiently using [Microsoft Excel](#). We will concentrate specifically on the necessary data preparation and the utilization of Excel's robust charting tools to produce professional-grade side-by-side boxplots ready for in-depth analysis.

Step 1: Structuring and Inputting Data in Microsoft Excel

The foundational requirement for successfully creating comparative boxplots within Excel is the meticulous and correct organization of the source data. Excel's charting engine mandates that each dataset designated for comparison must reside in its own dedicated column. Crucially, the labels or headings placed at the top of these columns will automatically be utilized as the identifying labels for the corresponding boxplots in the final chart.

To initiate this process, open a clean worksheet in Excel and begin entering the numerical values for the datasets intended for analysis. For the purpose of this practical demonstration, we will input values corresponding to three separate statistical groups, designated as **Dataset 1**, **Dataset 2**, and **Dataset 3**. It is essential to ensure that the data points for each dataset are aligned vertically within their respective columns.

This specific columnar arrangement is not optional; it is the fundamental mechanism by which Excel's chart generation function correctly identifies, segregates, and subsequently plots the individual statistical summaries--the five-number summaries--for each distinct group, enabling the side-by-side comparison.

	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							

Once all data points are accurately and completely entered and the columns are appropriately labeled with descriptive headers, the data preparation phase is complete, and we are ready to proceed to the chart visualization stage.

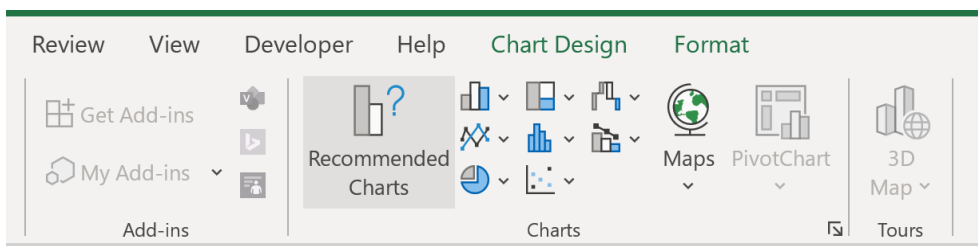
Step 2: Executing Chart Generation via the Insert Tab

With the data correctly structured in the worksheet, the subsequent phase involves leveraging Excel's sophisticated built-in charting capabilities to visualize the distributions simultaneously. This procedure is streamlined and relies on precise selection of the relevant data range prior to chart insertion.

First, the entire data range that is to be charted must be meticulously highlighted. This selection must include all numerical values along with their corresponding column headers. In this specific illustrative example, the user would select cells spanning from A1 through C21. This comprehensive selection explicitly communicates to Excel which data series must be included and compared within the resulting visualization.

	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								

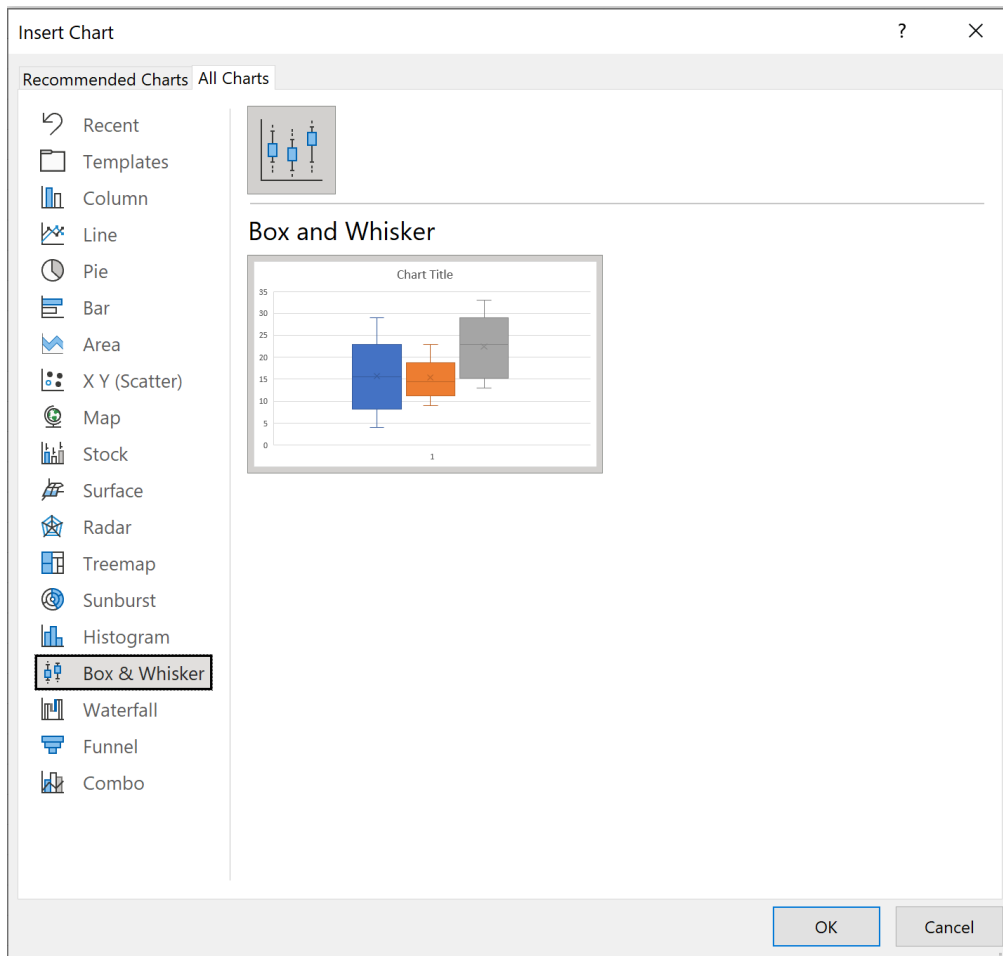
After securing the data selection, navigate to the **Insert** tab, which is prominently located in the primary ribbon interface of Excel. Although the **Box & Whisker** chart option might be immediately visible depending on your Excel version, utilizing the **Recommended Charts** feature often serves as the most reliable and expedient route to accessing the desired graph type.



Within the chart dialogue box that appears, select the **All Charts** tab to reveal the exhaustive list of available chart categories. Scroll down this list and specifically choose the **Box & Whisker** category. Confirm this choice by clicking **OK**.

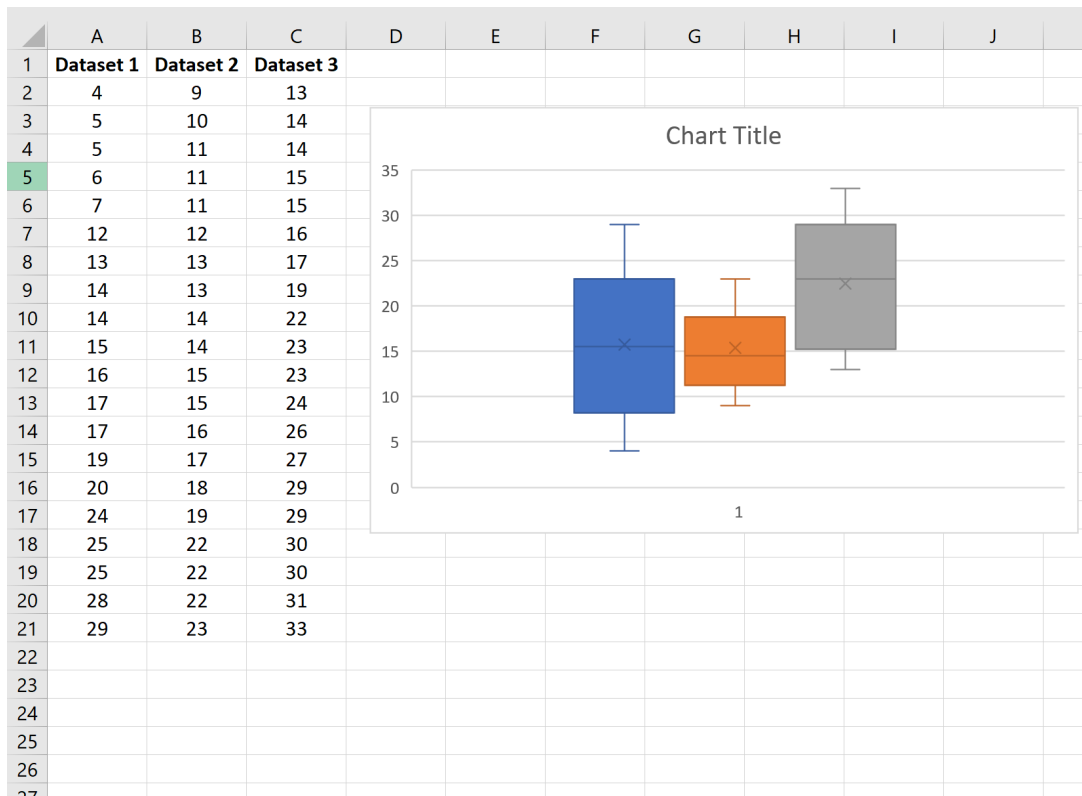
Excel will generate the raw side-by-side boxplots instantly. It automatically executes all necessary statistical calculations--determining the [quartiles](#), the [median](#), and the range--for each individual

dataset based on the highlighted input data.



The resultant preliminary visualization establishes an excellent starting point for detailed comparative analysis. The chart distinctly separates the three distributions, enabling immediate and objective observation of their relative differences in location and spread.

The following image displays the automatically generated initial side-by-side boxplots:

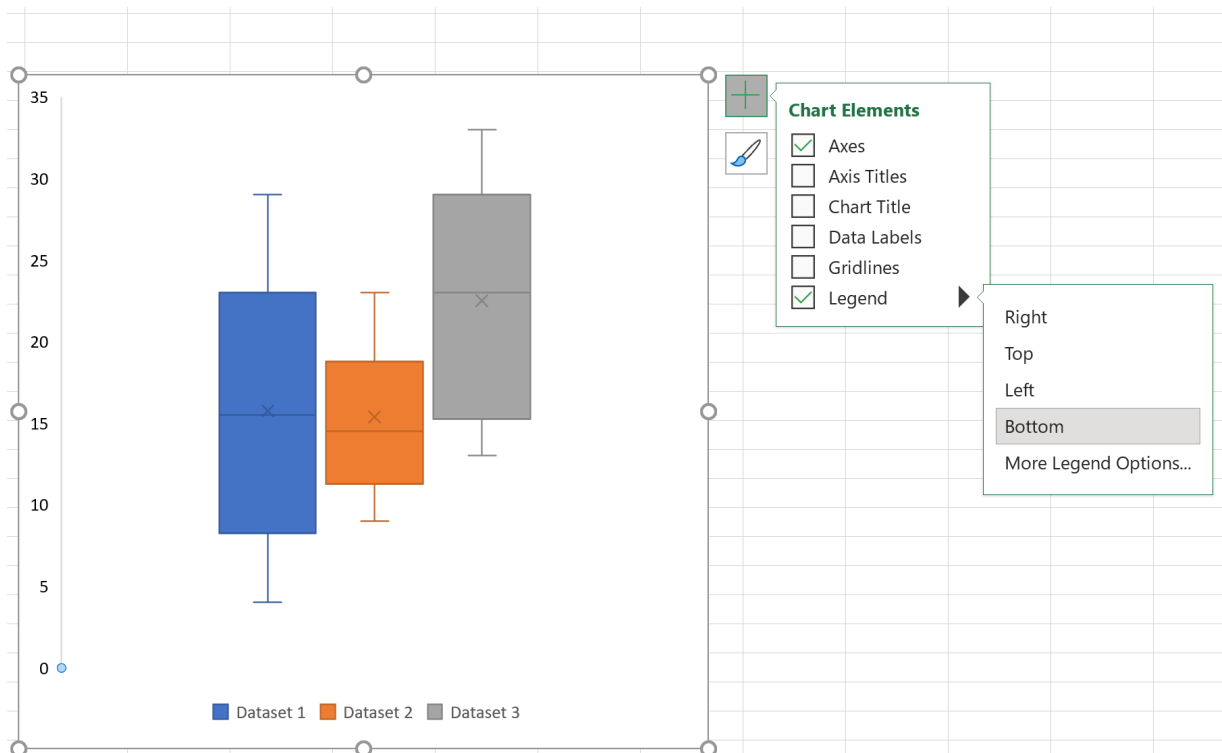


Step 3: Refining Visual Clarity and Aesthetics

Although the boxplots generated in the preceding step are statistically precise, optimizing their visual presentation is a critical step toward achieving enhanced readability and a professional appearance suitable for reporting. Refinement typically focuses on reducing visual noise and ensuring that all necessary interpretative components, such as the chart legend, are clear, present, and optimally positioned.

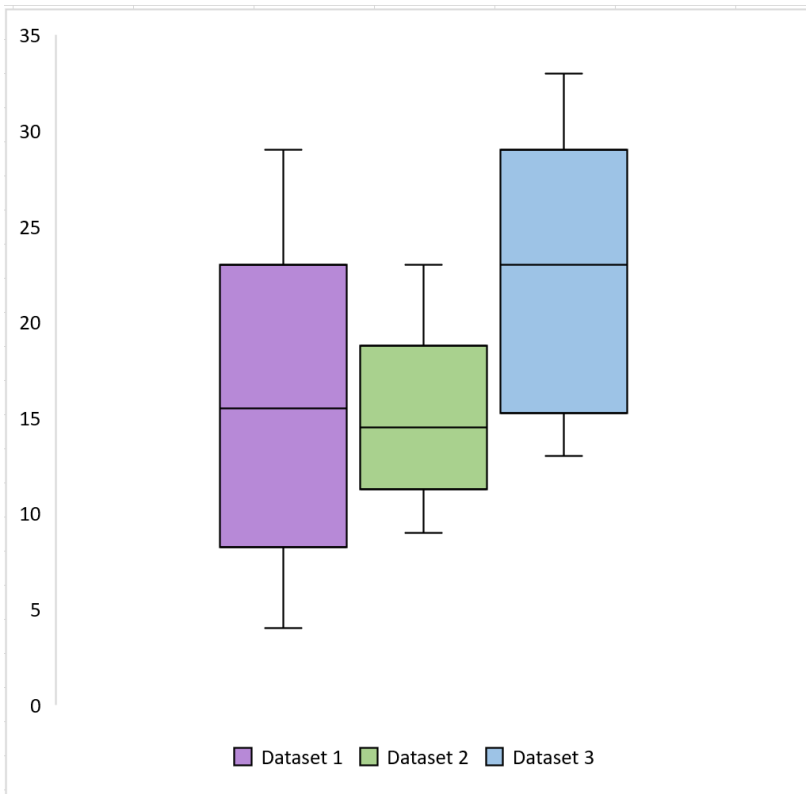
A common and highly recommended practice is the removal of extraneous elements, such as the default background gridlines, which often unnecessarily distract the viewer's attention from the primary data visualization. To remove these, simply click on any of the horizontal gridlines within the main chart area and then press the **Delete** key.

Next, it is essential to ensure the chart is fully self-explanatory by activating the legend, which links the boxplots to their respective dataset labels. To do this, click the green plus sign (+) icon located in the top right corner of the chart boundary. Check the box labeled **Legend** to display the labels corresponding to the datasets. For a balanced and non-obstructive layout, selecting the **Bottom** position option for the legend is generally advised, as it minimizes interference with the primary vertical axis scaling.



Finally, to distinctly differentiate the datasets or to comply with specific organizational branding standards, the appearance of the individual boxplots can be customized. By double-clicking directly on a specific boxplot (for example, the box representing Dataset 1), the formatting pane will appear, granting access to options for modifying fill color, outline weight, and other stylistic elements. Assigning distinct, contrasting colors to each dataset significantly enhances visual contrast and facilitates rapid interpretation of the comparison.

Following these crucial customization steps, the final, polished set of side-by-side boxplots is appropriately prepared for comprehensive statistical reporting and analysis:



Interpreting Distributions: Analyzing Variance and Central Tendency

The core objective of constructing side-by-side boxplots is to facilitate the direct comparison of critical characteristics across multiple data distributions, primarily focusing on measures of spread (or dispersion) and [central tendency](#). The physical dimensions of the box--specifically its length--and the length of its accompanying whiskers provide immediate visual evidence regarding the data's consistency and overall range.

The length of the box itself, which represents the Interquartile Range (IQR, Q3 minus Q1), and the total span defined by the whiskers, serve as direct quantitative indicators of data dispersion. A boxplot exhibiting a significantly longer box and extended whiskers suggests a wider spread among the data points and, consequently, a higher degree of [variance](#) within that specific dataset. Conversely, a short box indicates that the central 50% of the data points are tightly clustered, reflecting low variability and high consistency.

Central tendency is unambiguously represented by the horizontal line drawn within the box, which marks the statistical [median](#) (Q2). By comparing the vertical placement of this median line across the different boxplots, an analyst can swiftly ascertain which dataset exhibits the highest or lowest typical value.

Based on the side-by-side boxplots generated using the example data, we can derive the following

statistically grounded observations regarding the three compared datasets:

Spread Comparison (High Variability): Dataset 1 demonstrates the highest degree of [variance](#) among the three groups. This conclusion is visually supported by the fact that its boxplot is the longest overall, clearly indicating the widest distribution of observed values.

Consistency Comparison (Low Variability): Dataset 2 exhibits the lowest [variance](#). The substantially shorter length of its boxplot suggests that the majority of its data points are much more tightly grouped around the central tendency, indicating superior consistency.

Central Tendency Comparison: Dataset 3 possesses the highest median value. This is definitively confirmed by the horizontal median bar inside its box being positioned highest on the vertical axis relative to the median bars of the other two datasets.

These immediate and intuitive visual comparisons underscore precisely why boxplots are considered an indispensable tool for data professionals who need to communicate complex distributional differences effectively, concisely, and persuasively to non-technical audiences.

Additional Resources for Advanced Analysis

To further deepen your proficiency in statistical concepts and advanced data visualization techniques, we recommend exploring the following related topics:

Advanced statistical methods for the identification, interpretation, and appropriate handling of outliers that are visually displayed in boxplots.

Detailed procedures for calculating the precise [quartile](#) values and the Interquartile Range (IQR) in Excel using specialized statistical formulas (e.g., QUARTILE.EXC and QUARTILE.INC).

A comparative analysis of boxplots versus other essential distributional charts, such as histograms and density plots, emphasizing the situational advantages and disadvantages of each visualization method.