

# How to Color a Scatterplot by Value in Excel

Authored by  
**stats writer**

November 18, 2025

## RECOMMENDED CITATION

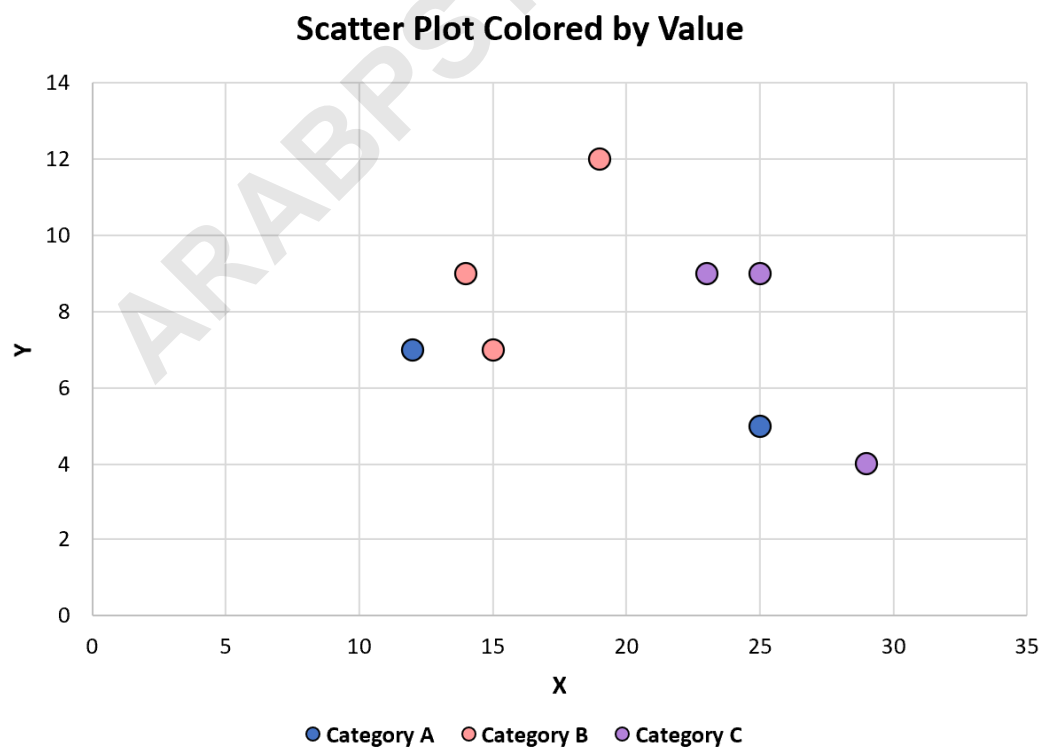
stats writer (2025). *How to Color a Scatterplot by Value in Excel*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=96098>

## Introduction: The Need for Categorical Scatterplots

In advanced data visualization, simply plotting X and Y coordinates is often insufficient. Analysts frequently need to incorporate a third dimension, typically a categorical variable, to understand how different groups interact or cluster within the coordinate system. When working within Excel, achieving a scatterplot where points are distinctly colored based on their category or value is a common and necessary task. This technique transforms a basic visual into a powerful analytical tool, allowing for immediate identification of group segregation or overlap.

The challenge in Excel is that standard scatterplots typically plot only two continuous variables (X and Y) as a single data series. To color points by category, we must engineer the data structure so that Excel treats each category as a separate data series, even though all points share the same X-axis values. This specialized approach ensures that formatting options, such as color and marker type, can be applied uniquely to each group. The final result should look similar to the professional graphic displayed below, where distinct groups are immediately recognizable by color coding.

While Excel does not offer a direct, single-click method for coloring by category, the process involves a clever data restructuring technique combined with a powerful conditional formula. This rigorous step-by-step tutorial will guide you through the exact methodology required to transform raw data into a dynamically colored categorical scatterplot. This method ensures maximum clarity and visual impact for your data presentation.



Fortunately, once the underlying principle of data transformation is understood, implementing this visualization technique in Excel becomes a highly manageable and reproducible process. The subsequent steps detail the precise data manipulation and charting functions necessary to generate this sophisticated visual output.

## Prerequisites: Understanding the Source Data Structure

Before beginning the plotting process, it is essential to understand the structure of the source data. A categorical scatterplot requires at least three columns: one for the X-axis coordinate, one for the Y-axis coordinate, and one for the categorical data (the grouping variable). The grouping variable determines the color of each point. Ensuring that this source data is clean, complete, and correctly labeled is the foundation of any successful visualization effort.

Our objective is to visualize the relationship between two continuous variables, often referred to simply as X and Y, while simultaneously highlighting the influence of the categorical variable. For instance, in a dataset tracking performance metrics (X and Y), the category might represent different regions, experimental treatments, or product types. Each distinct value in the category column requires its own dedicated data series in the final plot.

The technique we employ requires transforming the single category column into multiple columns--one for each unique category. This transformation is crucial because Excel's charting function reads each column (or series) independently, allowing us to assign unique formatting attributes (like color) to each generated series, thereby achieving the desired color-by-value effect. This is a fundamental concept in generating multi-series plots from summarized or grouped data.

## Step 1: Initial Data Entry in Excel

The first operational step involves entering the raw data into an Excel spreadsheet. This foundational dataset must clearly define the X, Y, and Category values for every observation. For this example, we will work with eight pairs of (X, Y) coordinates, each assigned to one of three distinct categories: Category A, Category B, or Category C. Accuracy in this initial entry phase is paramount, as all subsequent steps rely on the correct identification and placement of these variables.

Ensure your columns are clearly labeled (e.g., 'X Value', 'Y Value', 'Category'). While the labels themselves do not affect the formula execution, they are vital for human readability and error checking. Use columns A, B, and C for these initial entries, as illustrated in the image below, to maintain consistency with the subsequent formatting steps.

	A	B	C	D	E	F
1	<b>x</b>	<b>y</b>	<b>category</b>			
2	25	5	A			
3	12	7	A			
4	15	7	B			
5	14	9	B			
6	19	12	B			
7	23	9	C			
8	25	9	C			
9	29	4	C			
10						
11						
12						
13						
14						
15						
16						

Once the data is entered, it serves as the reference point for the calculated columns we are about to create. Note that the X-values (Column A) will remain unchanged throughout this process, as they will define the horizontal position for all data points regardless of their category. The manipulation focuses solely on segregating the Y-values (Column B) based on the Category (Column C).

## Step 2: Structuring Data for Multi-Series Plotting

The critical phase in preparing the data involves creating dedicated columns for each unique category present in the dataset. This restructuring allows Excel to register each group as a distinct series. Begin by identifying all unique values within the 'Category' column (Column C). In our example, these are 'Category A', 'Category B', and 'Category C'.

Enter these unique category names as headings in the row immediately above your data manipulation area. If your raw data occupies cells A1:C9, you should place these new headings starting in cell **D1**, continuing across **E1** and **F1**. These headings will eventually become the names of the three separate data series in the final scatterplot legend.

The principle here is straightforward: for every row, only one of the category columns (D, E, or F) should contain a numerical Y-value, while the others must contain a non-plottable value. This ensures that when Excel processes the data, it only draws a point in the series corresponding to

the observation's actual category. This segregation is achieved efficiently using a conditional logic function within Excel.

### Step 3: Implementing the Conditional Formula (IF and NA)

The segregation of Y-values is performed using a conditional statement, specifically the IF function combined with the `NA()` function. The purpose of this formula is to check if the current observation's category matches the header of the new column. If they match, the formula returns the corresponding Y-value; if they do not match, it returns the `NA()` error.

The `NA()` function is critically important here. When Excel attempts to plot a value that results in the **#N/A** error, it simply skips that point. This effectively creates 'holes' in the data series for points that do not belong to that category, ensuring that only the relevant points are plotted for Category A, Category B, and Category C, respectively.

The required formula, entered first into cell **D2**, utilizes absolute and relative referencing to allow for correct dragging and population across the entire range. The formula structure is designed to compare the category in Column C against the category header in Row 1.

Enter the following formula into cell **D2**:

```
=IF($C2=D$1, $B2, NA())
```

Let's break down the referencing logic:

**\$C2**: The column (C) is locked because we always compare against the Category column, but the row (2) is relative so it moves down when dragged vertically.

**D\$1**: The row (1) is locked because we always compare against the Category header in the top row, but the column (D) is relative so it moves right when dragged horizontally (to E1, F1).

**\$B2**: The column (B) is locked, meaning we always pull the Y-value from Column B.

This hybrid referencing structure is essential for accurate calculation across the data matrix.

Once entered into **D2**, you must drag this formula horizontally to cell **F2** (covering Categories A, B, and C). Subsequently, drag the entire row D2:F2 vertically down to the last row of your data (in this case, cell **F9**). This operation automatically populates the matrix, ensuring that only the correct Y-values appear in the category-specific columns. The resulting table should clearly show that for any given row, only one cell in the D-F range contains a number, while the others contain **#N/A**.

D2    :    ✕ ✓ *fx*    =IF(\$C2=D\$1, \$B2, NA())

	A	B	C	D	E	F	G
1	<b>x</b>	<b>y</b>	<b>category</b>	<b>A</b>	<b>B</b>	<b>C</b>	
2	25	5	A	5	#N/A	#N/A	
3	12	7	A	7	#N/A	#N/A	
4	15	7	B	#N/A	7	#N/A	
5	14	9	B	#N/A	9	#N/A	
6	19	12	B	#N/A	12	#N/A	
7	23	9	C	#N/A	#N/A	9	
8	25	9	C	#N/A	#N/A	9	
9	29	4	C	#N/A	#N/A	4	
10							
11							
12							
13							
14							
15							

### Step 4: Generating the Multi-Series Scatterplot

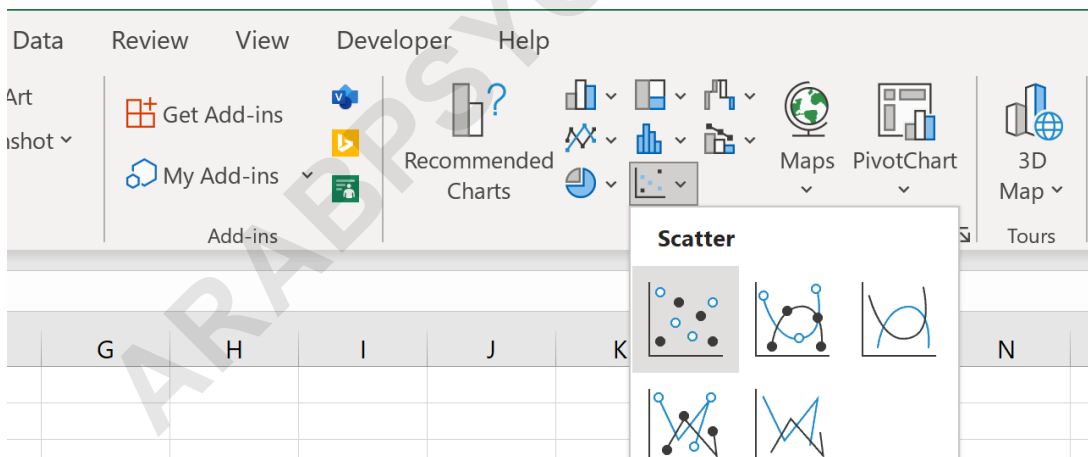
With the data now correctly formatted into independent series, the next step is to instruct Excel to create the scatterplot using the segregated data. This step requires careful selection of non-contiguous ranges, a common requirement when working with complex data layouts in Excel.

Start by selecting the X-values, which are located in the range **A2:A9**. This range will serve as the horizontal axis for all three data series. After selecting the X-values, it is crucial to use the **Ctrl** key (or Command key on Mac) to select the remaining Y-value ranges. Hold **Ctrl** and then highlight the entire block of calculated Y-values, which corresponds to the range **D2:F9**.

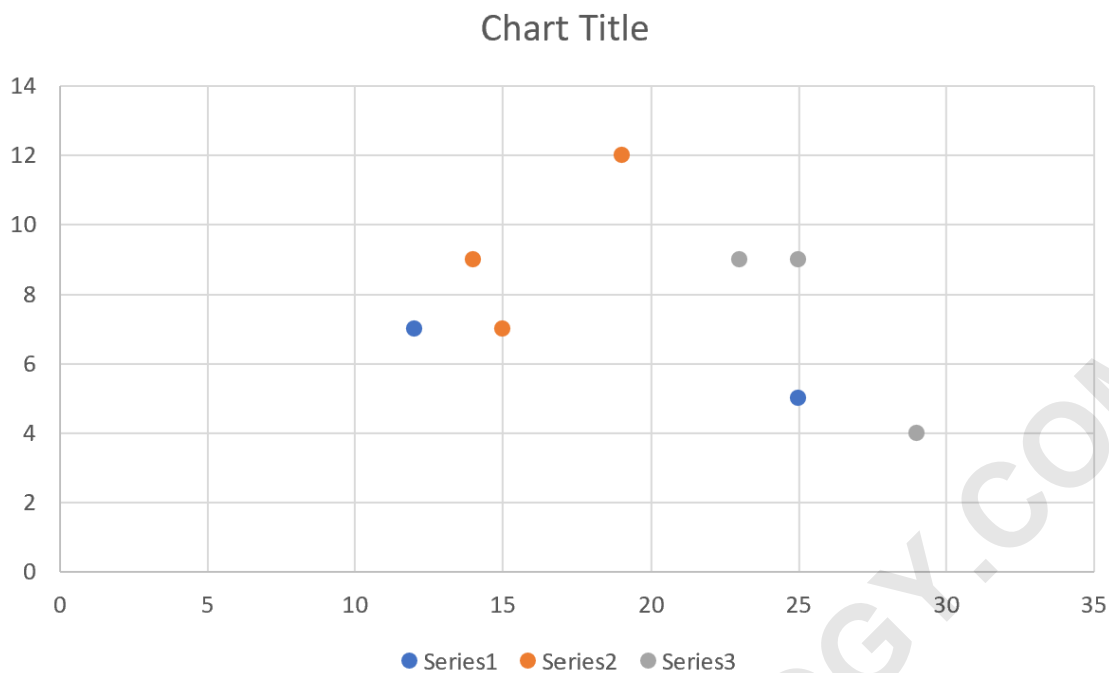
This simultaneous selection of the X-axis range (A2:A9) and the three segmented Y-series (D2:F9) tells Excel that A2:A9 is the shared X-data for the three separate columns in D2:F9. The visual selection process should look like the image provided below, where two distinct, non-adjacent blocks of cells are highlighted.

	A	B	C	D	E	F	G
1	<b>x</b>	<b>y</b>	<b>category</b>	<b>A</b>	<b>B</b>	<b>C</b>	
2	25	5	A	5	#N/A	#N/A	
3	12	7	A	7	#N/A	#N/A	
4	15	7	B	#N/A	7	#N/A	
5	14	9	B	#N/A	9	#N/A	
6	19	12	B	#N/A	12	#N/A	
7	23	9	C	#N/A	#N/A	9	
8	25	9	C	#N/A	#N/A	9	
9	29	4	C	#N/A	#N/A	4	
10							
11							
12							
13							
14							
15							

Once both ranges are highlighted, navigate to the 'Insert' tab in the Excel ribbon, locate the 'Charts' group, and select the 'Scatter' chart type. Choose the option that displays points without lines (usually the first option). Excel will automatically generate the plot, recognizing the three separate columns (D, E, F) as distinct data series, and assigning a unique default color to each one.



The immediate output will be a fully functional scatterplot where each set of (X, Y) coordinates is appropriately displayed, and the points are colored based on the category they were assigned to in Step 3. The legend will automatically reflect the headers from D1, E1, and F1 (Category A, Category B, Category C).



Every coordinate from the original dataset is correctly positioned on the plot, and, crucially, the visual differentiation based on the categorical data is achieved through the use of distinct marker colors for each series.

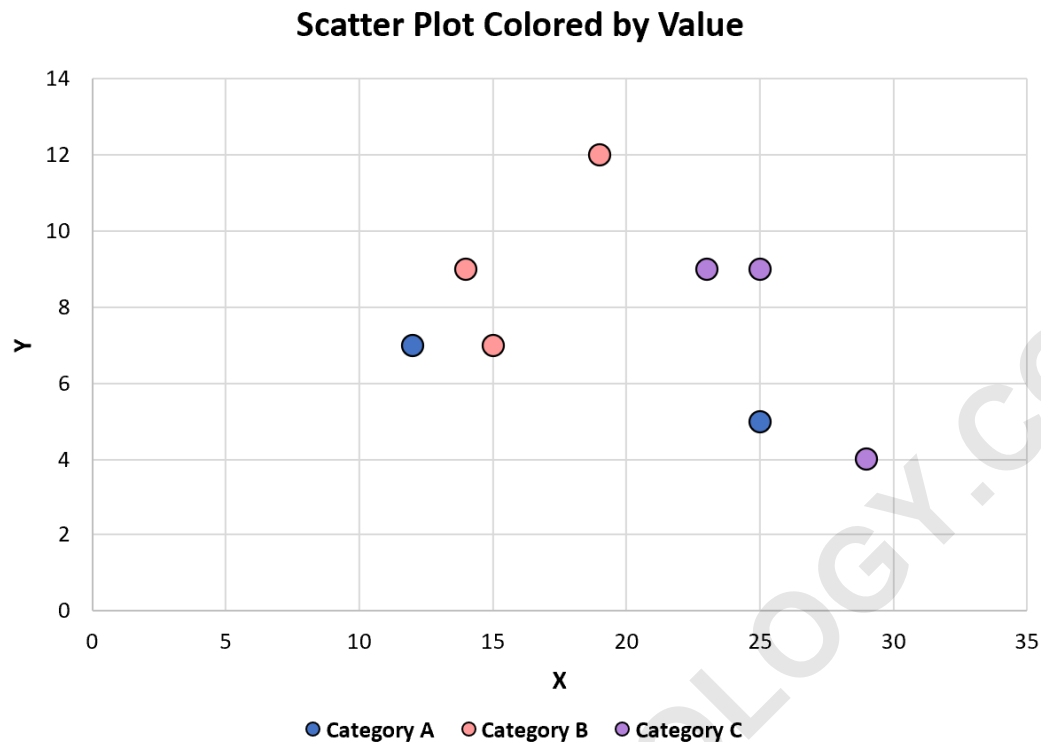
### Step 5: Finalizing Chart Appearance and Aesthetics (Optional)

While the plot is now functionally complete, the final step involves refining the visual aesthetics to enhance readability and professional presentation. This is an optional but highly recommended phase of data visualization. Customizing the chart allows the analyst to align the visual output with specific reporting standards or to simply make the graphic more impactful and easier for the audience to interpret.

Key modifications often include adjusting the colors of the data points, changing the size and style of the markers, and refining the labels and axis titles. To customize a specific category's appearance, select the data series in the chart, right-click, and choose 'Format Data Series'. This pane allows control over fill, line, and marker options. For instance, you might change the colors to a palette that is universally accessible (e.g., avoiding red/green combinations for colorblind viewers) or increasing the marker size for better visibility.

Furthermore, always ensure that the chart title, axis titles, and legend are clear, concise, and accurately reflect the data presented. Removing distracting elements like redundant gridlines or excessive decimal points on axis labels can significantly improve the clarity of the visualization. The final, polished result demonstrates the successful execution of this multi-series technique for

coloring a scatterplot by value in Excel.



## Conclusion: Maximizing Data Visualization

The process outlined above provides a robust and reliable method for solving the common problem of integrating categorical data into a standard Excel scatterplot. By leveraging the conditional logic of the IF function combined with the plot-skipping behavior of the `NA()` function, analysts can effectively segment a single dataset into multiple, color-coded series.

Mastering this data preparation technique opens up possibilities for complex comparative analysis. This structured approach ensures that your visualizations are not only accurate but also highly informative, fulfilling the primary goal of effective data visualization: to communicate patterns and relationships clearly and efficiently. This skill is foundational for anyone producing advanced charts in Microsoft Excel.

The following tutorials explain how to perform other common functions with scatterplots in Excel: