

Learning to Add a Regression Line to a Scatterplot in Google Sheets

Authored by
Mohammed loot

November 11, 2025

RECOMMENDED CITATION

Mohammed loot (2025). *Learning to Add a Regression Line to a Scatterplot in Google Sheets*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=17135>

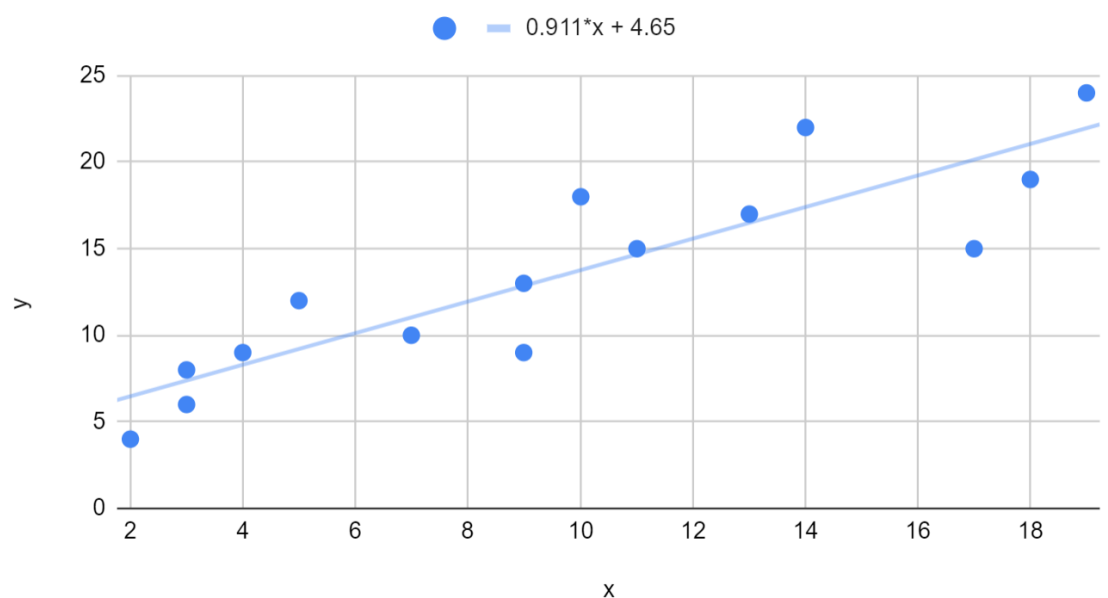
Understanding Regression Lines in Data Visualization

A [regression line](#), often referred to as a [trendline](#), is a fundamental tool in statistical analysis and data visualization. It serves to visually represent the line that mathematically best "fits" a given [dataset](#), illustrating the relationship or correlation between two numerical variables. This crucial line helps analysts summarize the linear relationship between the independent variable (X) and the dependent variable (Y) and is essential for making predictions based on the observed data pattern.

The mathematical process used by software like [Google Sheets](#) to derive this line typically involves the method of least squares. This technique minimizes the sum of the squared vertical distances between the actual plotted data points and the line itself, ensuring the resulting line is the most representative linear summary possible. The resulting visualization provides a simple, yet powerful, summary of potentially complex data, immediately indicating whether the relationship is positive, negative, or negligible.

This comprehensive, step-by-step tutorial will demonstrate precisely how to add this critical analytical element--the regression line--to a [scatterplot](#) within Google Sheets. We will use a standard dataset to illustrate the entire process, which culminates in the easily interpretable chart shown below:

y vs. x



Let's jump in and start structuring our data!

Step 1: Preparing and Entering the Dataset

Before any visualization can begin, the data must be accurately organized. In nearly all statistical models involving two variables, it is standard procedure to arrange the raw data into corresponding columns: one column dedicated to the independent variable (X) and the other to the dependent variable (Y). The independent variable is the predictor and will be plotted on the horizontal axis, while the dependent variable is the outcome and will be plotted on the vertical axis.

For the purpose of this example, we will input the required fifteen paired observations into our Google Sheet. It is paramount that the data is clean, free of textual errors, and correctly recognized as numerical values to ensure the subsequent charting and statistical calculations are accurate. We will use columns A and B for our input.

	A	B	C	D
1	x	y		
2	2	4		
3	3	8		
4	3	6		
5	4	9		
6	5	12		
7	7	10		
8	9	9		
9	9	13		
10	10	18		
11	11	15		
12	13	17		
13	14	22		
14	17	15		
15	18	19		
16	19	24		
17				

By carefully entering the data, we establish the foundation necessary for calculating the least squares [regression line](#), guaranteeing that the final visualization accurately reflects the empirical relationship present in the raw data.

Step 2: Inserting and Formatting the Scatterplot

Once the data entry is complete, the next objective is to visualize the distribution of the data points

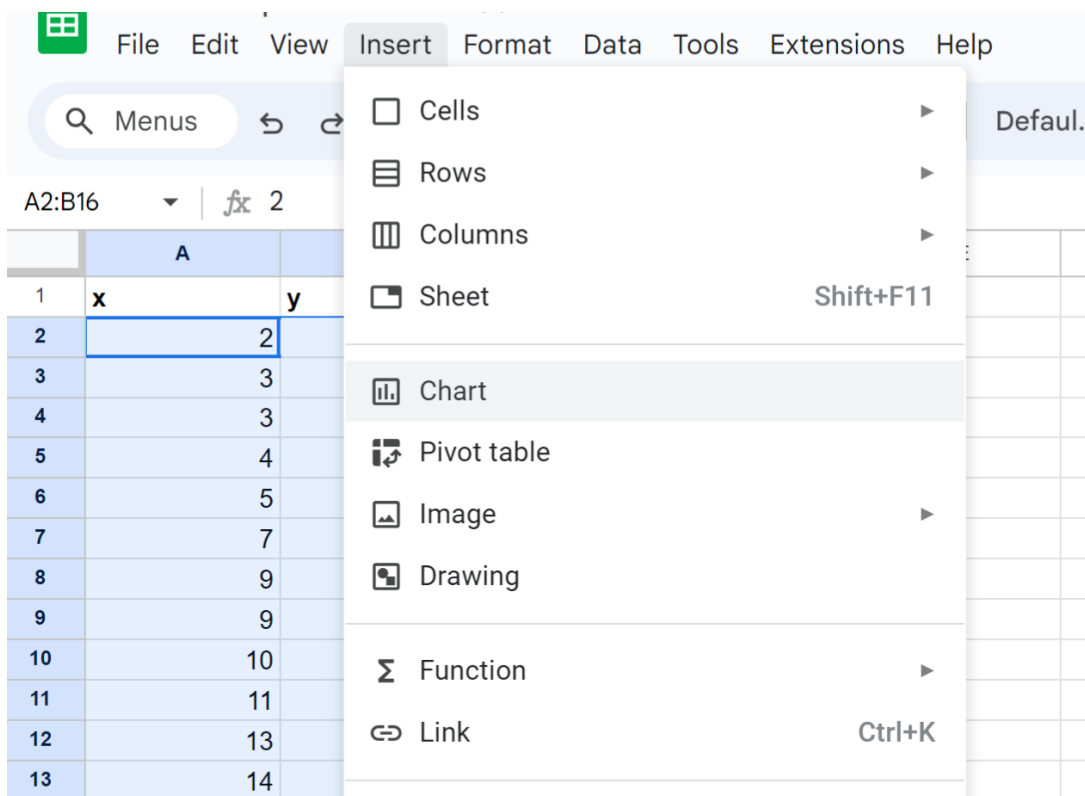
using a [scatterplot](#). This visualization is essential because it provides immediate visual confirmation regarding the shape and direction of the relationship, guiding us to confirm that a linear model (as opposed to an exponential or polynomial one) is the most appropriate fit.

To insert the chart into your [Google Sheets](#) document, follow these sequential steps:

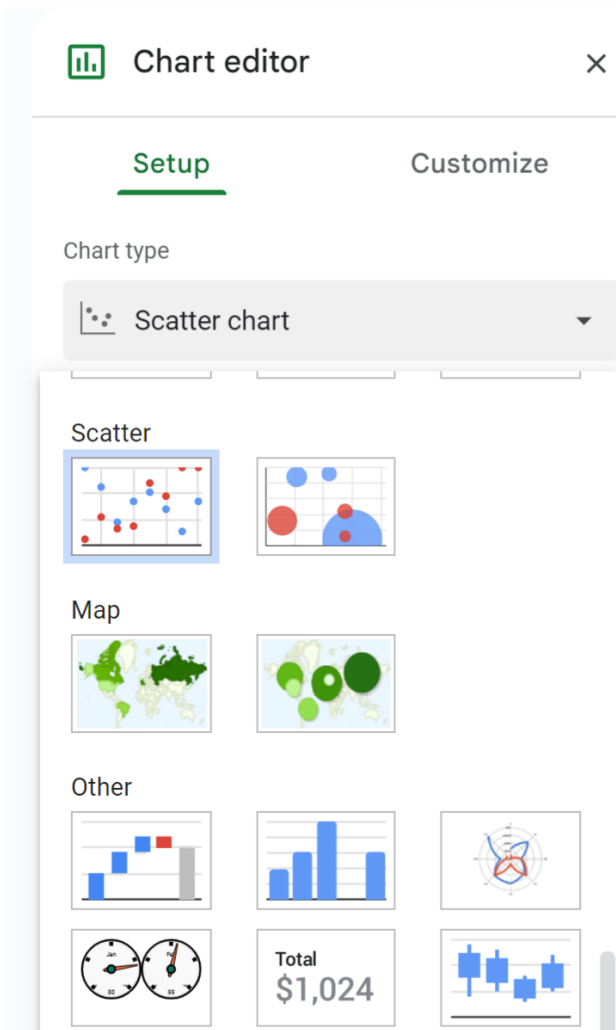
Identify and highlight the complete cell range containing the numerical data, which in this demonstration is **A2:B16**.

Navigate to the primary menu ribbon at the top of the interface and click the **Insert** tab.

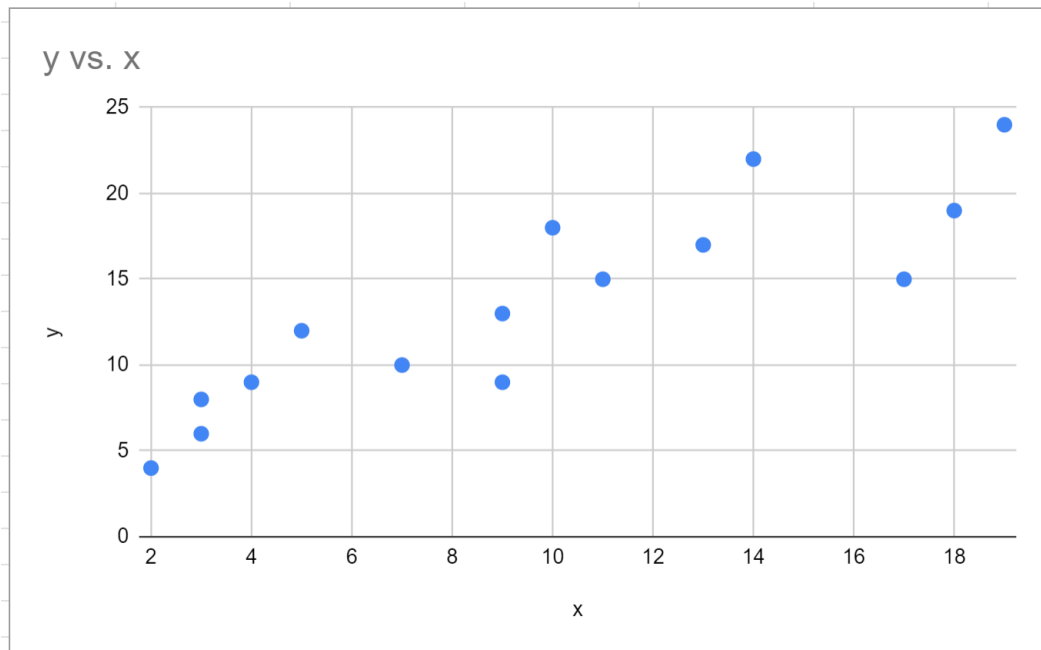
From the options presented in the dropdown menu, select **Chart**. This action will launch the powerful **Chart editor** panel on the right side of your browser window.



The **Chart editor** may default to a different chart type. To correct this, ensure you are on the **Setup** tab, locate the **Chart type** setting, and scroll down the extensive list of options until you find and select **Scatter chart**. This conversion yields the required point-based visualization.



After these steps, the initial [scatterplot](#) will appear, showing the raw distribution of the data points ready for analytical enhancement:



Step 3: Integrating the Linear Regression Line

Once the scatterplot is successfully generated, the final step involves overlaying the statistical model onto the visual data. This is achieved through the customization features available in the Google Sheets **Chart editor**.

To add the [regression line](#), click on the **Customize** tab located within the **Chart editor** panel. This tab is where all aesthetic and analytical enhancements are configured.

Within the **Customize** menu, locate and expand the section labeled **Series**. This section manages the visual properties of the data points and related analytical additions. Scroll down through the Series options until you find the checkbox specifically labeled **Trendline**. Checking this box immediately instructs Google Sheets to calculate the line of best fit using the least squares criterion and display it on the chart.

Chart editor ×

Setup **Customize**

Format data point **Add**

Error bars

Data labels

Trendline

Type: Linear

Line color: Blue

Line opacity: 40%

Line thickness: 2px

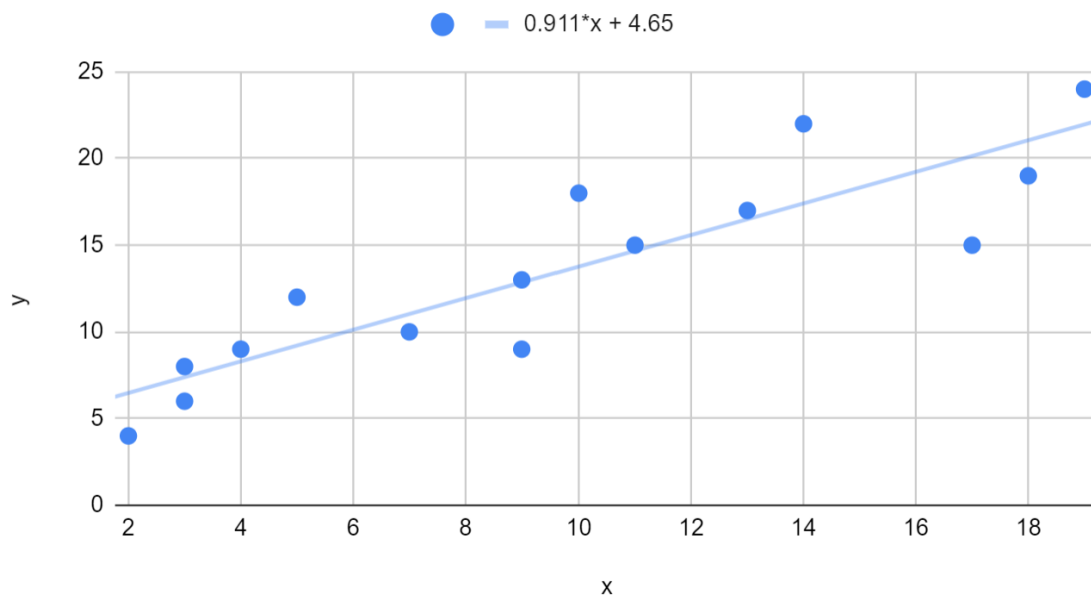
Label: Use Equation

Show R²

Crucially, under the **Trendline** settings, you can also select the option to display the **Label** as **Use Equation**. This feature automatically calculates and displays the [equation of the regression line](#) directly on the chart, transforming the visualization into a complete analytical report.

A regression line will now be added to the [scatterplot](#), along with the corresponding algebraic equation, resulting in the final, enhanced visualization:

y vs. x



Interpreting the Predictive Regression Equation

For the specific data utilized in this tutorial, the calculated linear regression equation displayed above the chart is:

$$y = 0.911x + 4.65$$

This formula perfectly aligns with the general linear model structure, $Y = B_0 + B_1X$, where B_0 is the y-intercept (4.65) and B_1 is the slope coefficient (0.911). A thorough understanding of these components is vital for extracting meaningful statistical insights from the model.

The interpretation of these two key statistical parameters is as follows:

The Slope (0.911): The slope represents the estimated change in the dependent variable (Y) that is expected to occur for every one-unit increase in the independent variable (X). Because the value is positive (0.911), it signifies a positive correlation: as X increases, Y is expected to increase. We can state that for each additional one unit increase in the x variable, the average increase in the y variable is predicted to be **0.911**.

The Y-Intercept (4.65): The intercept (B_0) defines the point where the regression line crosses the Y-axis. Statistically, it represents the predicted average value of the y variable when the x variable is set to zero. Thus, based on this model, when the x variable is equal to zero, the average value for the y variable is estimated to be **4.65**.

The most powerful application of this derived relationship is its ability to perform prediction or interpolation. We can substitute any relevant value of x (within the range of our observed [dataset](#)) into the equation to forecast the corresponding expected value of y .

For example, if we wish to predict the expected value for y when the independent variable x is equal to 15, the calculation proceeds as follows:

$$y = 0.911 * (15) + 4.65 = \mathbf{18.315}$$

Therefore, the expected outcome for Y , given the observed linear trend in the data when X is 15, is approximately 18.315.

Additional Resources for Google Sheets Proficiency

While this tutorial focused on the linear model, it is worth noting that [Google Sheets](#) also provides options for fitting non-linear trendlines, such as exponential, polynomial, and logarithmic models, depending on the characteristics of your data. Selecting the correct [trendline](#) type is vital for ensuring your statistical model accurately reflects the underlying phenomenon.

To help you further enhance your data visualization and analytical skills within the platform, the following tutorials explain how to perform other common and powerful tasks in Google Sheets:

[How to Calculate Correlation in Google Sheets](#)

[How to Calculate Standard Deviation in Google Sheets](#)

[How to Create a Histogram in Google Sheets](#)