

Polynomial Regression in Google Sheets: A Step-by-Step Tutorial

Authored by
Mohammed loot

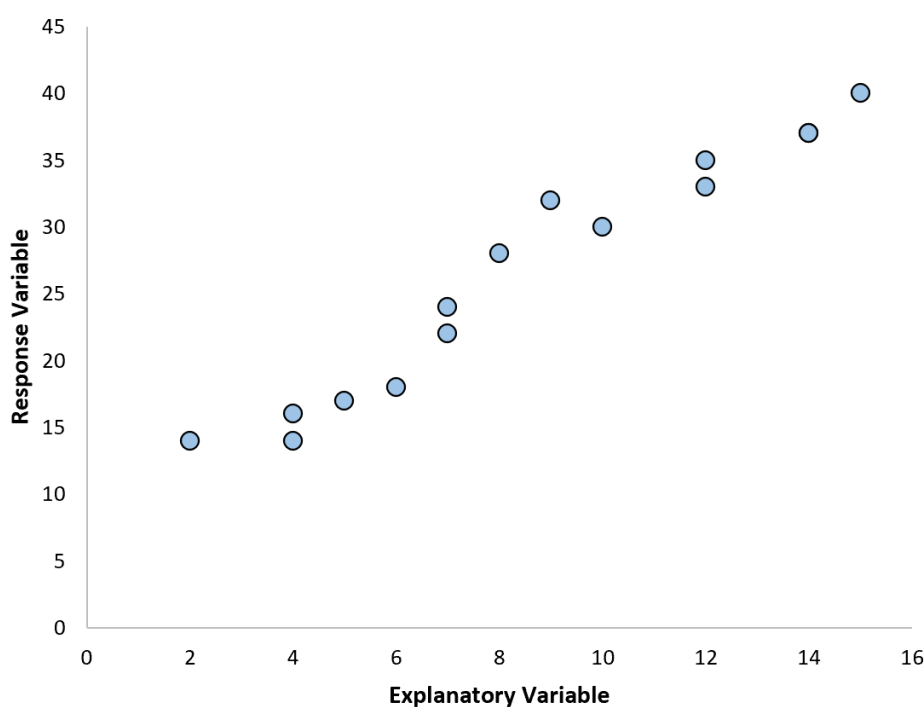
November 5, 2025

RECOMMENDED CITATION

Mohammed loot (2025). *Polynomial Regression in Google Sheets: A Step-by-Step Tutorial*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=10947>

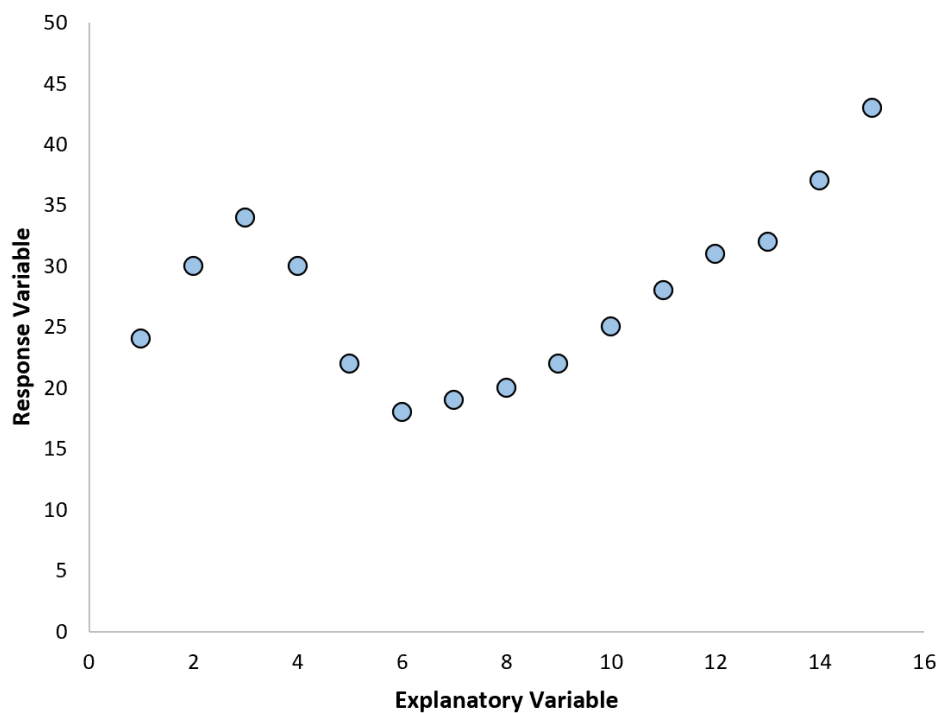
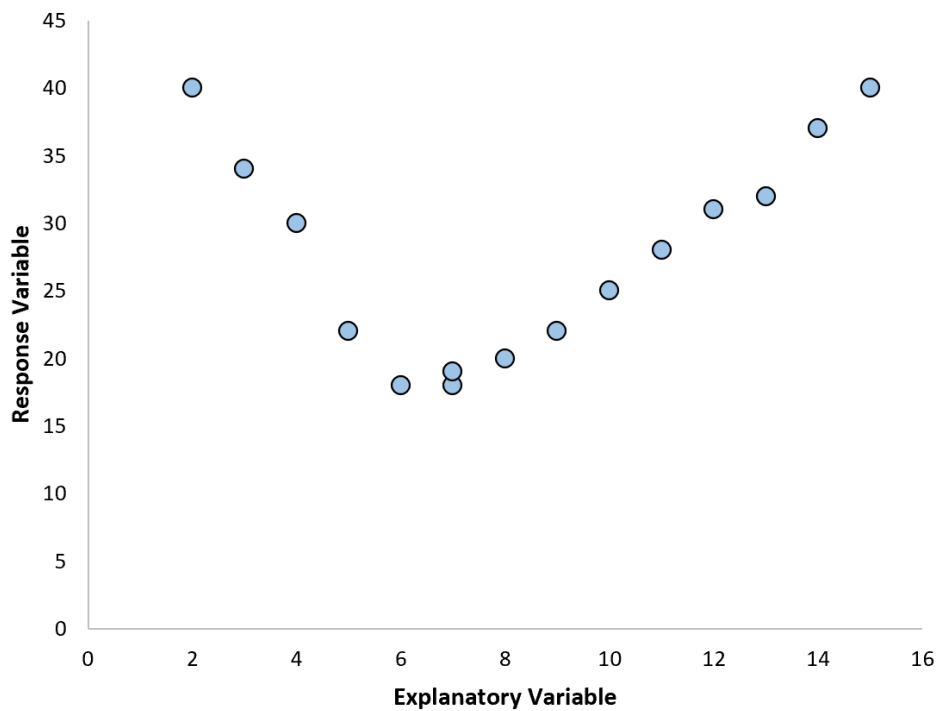
[Regression analysis](#) stands as a cornerstone technique in statistical modeling, allowing researchers and analysts to rigorously quantify the relationship between predictor variables and a corresponding [response variable](#). This analytical framework is essential for developing robust predictive models and understanding the causal impact of input changes on observed outcomes. While powerful, the application requires selecting the model that best represents the underlying data structure.

The most common and straightforward approach is [simple linear regression](#). This model assumes a direct, constant relationship where data points tend to align closely along a single straight line when plotted on a coordinate system. This assumption simplifies interpretation and calculation, making it the preferred method when the relationship is clearly linear, as shown in the visualization below.



However, the complexity inherent in real-world datasets often dictates that relationships are distinctly [nonlinear](#). Standard linear models are fundamentally incapable of accurately capturing the curvature or multiple inflection points present in such data. Attempting to force a straight line onto a curved pattern inevitably leads to poor predictions, inaccurate parameter estimates, and a general failure to model the true relationship.

Visual inspection using [scatterplots](#) frequently reveals these complex patterns, which may appear as a simple U-shape (quadratic) or include multiple directional shifts (cubic or higher degree). When dealing with such data, analysts must select a more flexible modeling approach.



To address these curved relationships, analysts must transition to [polynomial regression](#). This powerful specialization of [regression analysis](#) extends the basic linear model by incorporating polynomial terms (powers) of the predictor variable. By doing so, the model gains the flexibility to accurately trace the curvature and complex, [nonlinear](#) relationship between the variables, leading

to significantly improved model fit and predictive accuracy.

This comprehensive guide offers a detailed, step-by-step tutorial on implementing and interpreting [polynomial regression](#), utilizing the accessible and powerful data visualization capabilities of [Google Sheets](#).

Understanding Polynomial Regression Fundamentals

Before we begin manipulating data within [Google Sheets](#), it is crucial to establish a conceptual foundation for polynomial models. Although a polynomial relationship generates a curve, the model itself is technically classified as a form of regression that remains linear in its parameters. This distinction is vital: the model achieves its curved fit by utilizing mathematical transformations--specifically, powers (X , X^2 , X^3 , and so on)--of the original predictor variable.

The central controlling factor in a polynomial model is its **degree**, defined by the highest exponent included in the equation. This degree dictates the complexity of the curve the model can accommodate, directly corresponding to the maximum number of bends or inflection points possible. For instance, a degree 2 (quadratic) model can capture a single curve, while a degree 3 (cubic) model can capture two bends, allowing for an S-like shape.

Selecting the optimal degree is perhaps the most critical decision in polynomial modeling. A low degree may result in an underfit model that misses the true pattern, while an excessively high degree risks **overfitting**. Overfitting occurs when the model becomes too complex, fitting the random noise or idiosyncrasies of the specific sample data rather than the underlying population trend. To find the ideal balance between model accuracy and simplicity (parsimony), we rely on both visual confirmation through [scatterplots](#) and quantitative metrics like the [R-squared](#) value.

Step 1: Preparing and Organizing Data in Google Sheets

To successfully execute our [polynomial regression](#), we require a dataset that clearly exhibits a nonlinear relationship. We will utilize a small, simulated dataset comprising ten pairs of observations, representing our predictor variable (X) and our [response variable](#) (Y). Proper initial data preparation is fundamental to all successful statistical modeling.

Begin by opening a new spreadsheet in [Google Sheets](#). Input the designated X and Y values into columns A and B, respectively, ensuring that the input starts from row 2. It is important to label the columns clearly in row 1 (e.g., " X " and " Y ") to ensure accurate data identification during the subsequent charting and analysis phases.

	A	B	C	D	E
1	x	y			
2	2	18			
3	4	14			
4	4	16			
5	5	17			
6	6	18			
7	7	23			
8	7	25			
9	8	28			
10	9	32			
11	12	29			
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					

Once the data is accurately entered and labeled, we can proceed to the visualization stage, which is essential for confirming the presence and general shape of the nonlinear pattern that necessitates polynomial modeling.

Step 2: Visualizing the Relationship via Scatterplot

The visualization of data through a [scatterplot](#) is not merely optional; it is a mandatory diagnostic step when considering any form of [regression analysis](#) that deviates from simple linearity. The scatterplot provides immediate, powerful insight into the functional form of the relationship, confirming whether a curve exists and guiding our initial hypothesis regarding the appropriate polynomial degree.

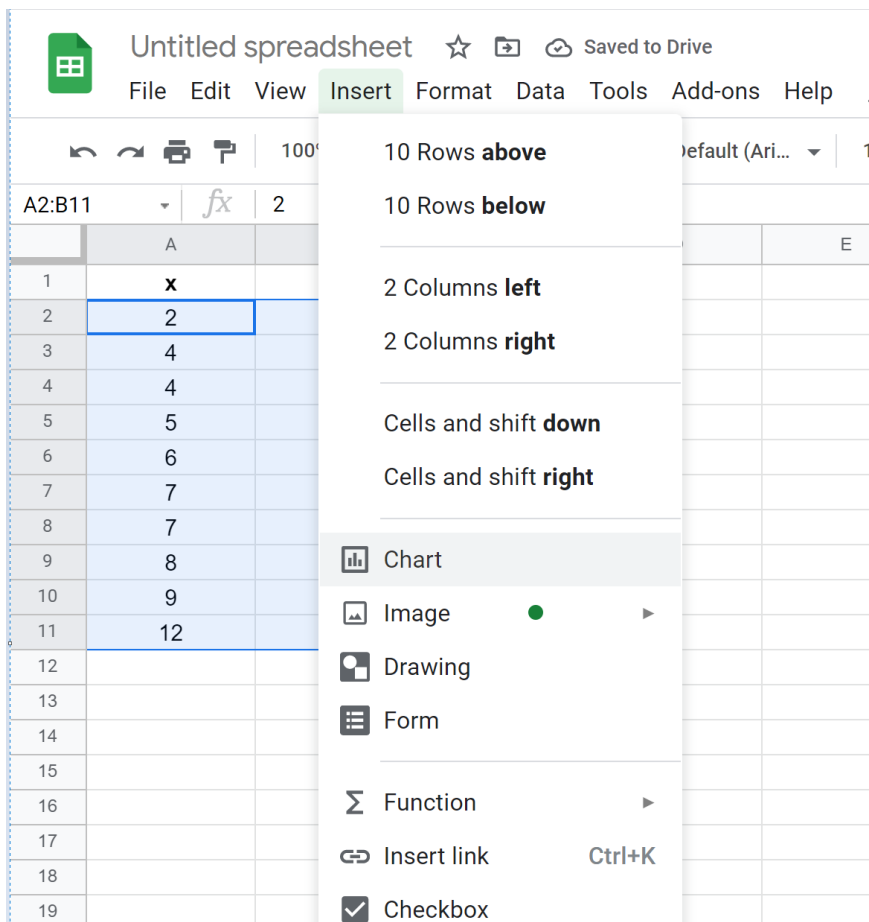
To generate the visualization in [Google Sheets](#), follow these precise steps:

Highlight the entire dataset, specifically the range of cells **A2:B11**, ensuring both the predictor (X) and response (Y) data points are selected.

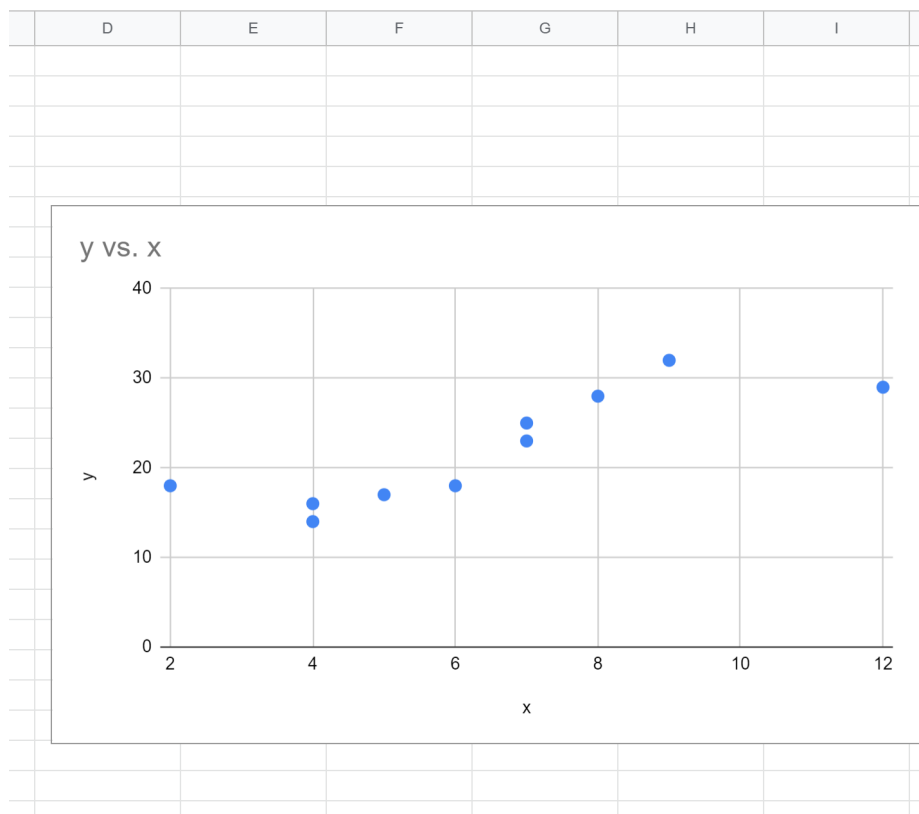
Navigate to the main menu at the top of the screen, select the **Insert** tab, and choose **Chart** from the subsequent dropdown menu.

	A	B	C	D
1	x	y		
2	2	18		
3	4	14		
4	4	16		
5	5	17		
6	6	18		
7	7	23		
8	7	25		
9	8	28		
10	9	32		
11	12	29		
12				
13				
14				
15				
16				
17				
18				
19				

Google Sheets is typically intelligent enough to automatically default to a Scatter Chart when two numerical columns are selected. If it defaults to another chart type, ensure you change the chart type manually within the Chart Editor.

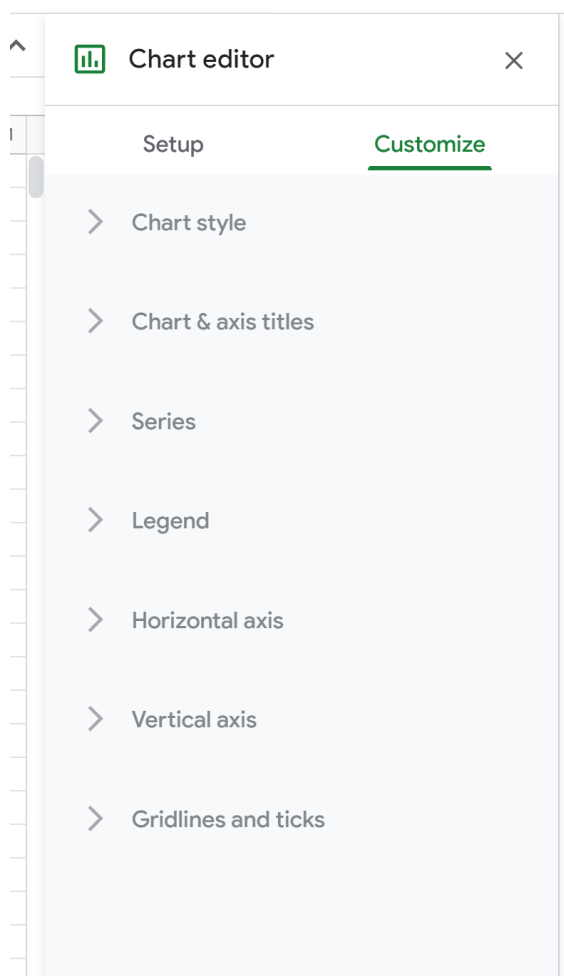


Upon reviewing the generated scatterplot, the non-linear nature of our simulated data becomes immediately obvious. The points exhibit a clear pattern that descends, reaches a minimum, and then ascends again, strongly suggesting that a quadratic (degree 2) or cubic (degree 3) model will be vastly superior to a standard [simple linear regression](#) fit.



Step 3: Determining the Optimal Polynomial Degree

With the scatterplot successfully generated, the next essential step is to modify the chart settings to introduce the [polynomial regression](#) trendline. This process allows us to visually evaluate the fit, estimate the underlying equation, and calculate the critical [R-squared](#) value for comparison. Start by double-clicking anywhere on the chart area to activate the **Chart Editor** panel, which appears on the right side of the sheet.



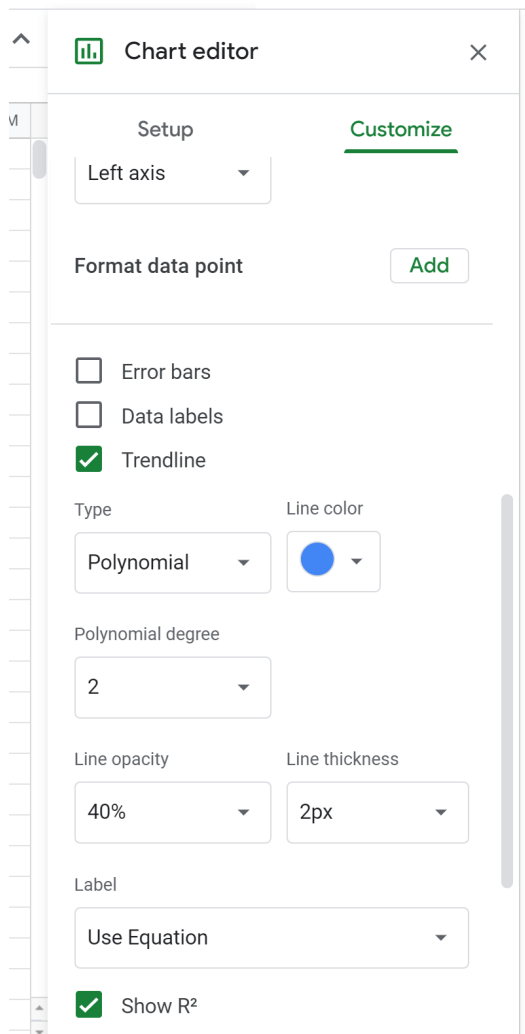
Within the Chart Editor, navigate to the **Customize** tab and perform the following sequence of modifications under the **Series** section:

Scroll down until you locate the **Trendline** option, and check the box to activate it.

Change the default **Type** setting from Linear to **Polynomial**.

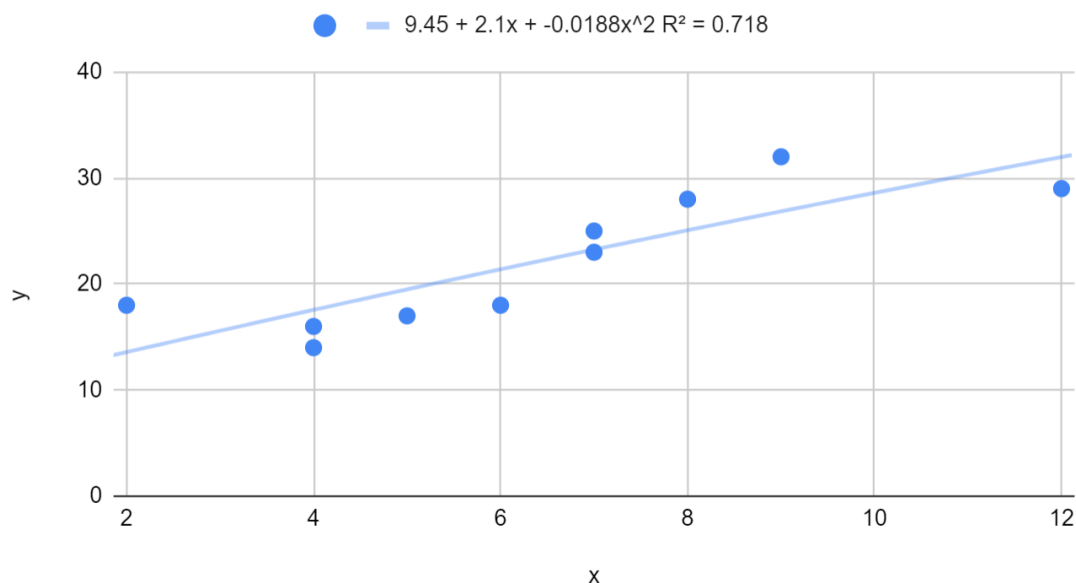
Under the **Label** setting, select **Use Equation**. This step is crucial as it forces the fitted regression formula to be displayed directly on the chart canvas.

Finally, check the box labeled **Show R²**. This coefficient of determination is the primary quantitative metric we will use to assess and compare different model fits.



By default, [Google Sheets](#) initiates the polynomial fit at Degree 2 (a quadratic model). This initial fit generates the corresponding formula and visualization overlaying the scatterplot:

y vs. x

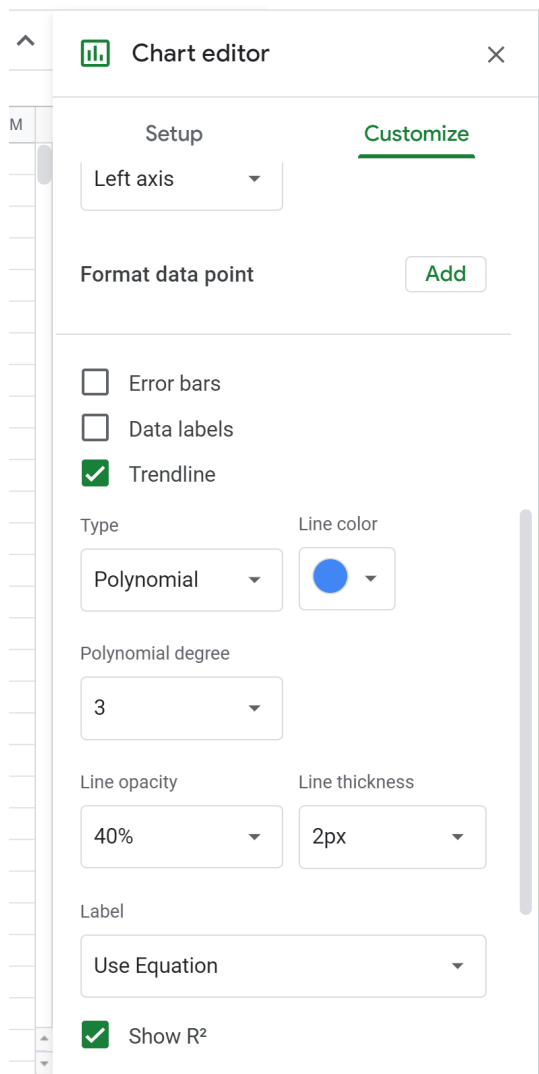


The resulting quadratic [polynomial regression](#) equation for degree 2 is displayed as: $y = 9.45 + 2.1x - 0.0188x^2$. The corresponding [R-squared](#) value for this fit is **0.718**, meaning that approximately 71.8% of the variability in Y is explained by the quadratic function of X.

Step 4: Comparing Fits and Selecting the Optimal Model (Degree 3)

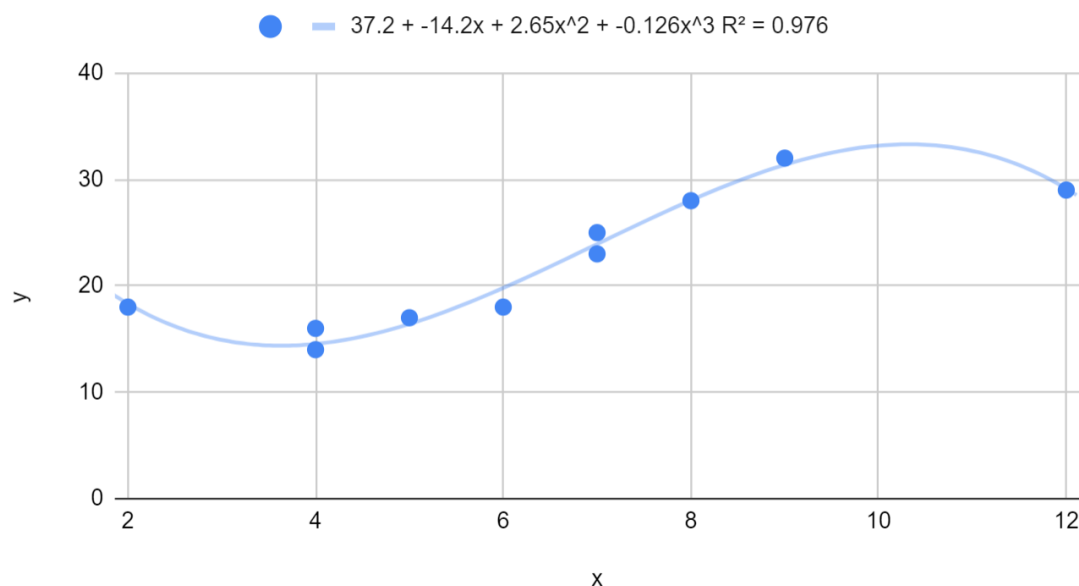
The [R-squared](#) value, or Coefficient of Determination, is the quantitative metric used to gauge the quality of the model fit. It represents the proportion of the variance in the [response variable](#) (Y) that is predictable from the predictor variable(s) (X) within the model. A value closer to 1 (or 100%) indicates a highly effective model that captures a large amount of the data's inherent trend.

While an [R-squared](#) of 0.718 for the quadratic model is respectable, visual inspection confirms that the single curve is still inadequate for tracing the nuanced S-like pattern of the data points. To explore a better fit, we must increase the polynomial degree. Return to the Chart Editor, navigate back to the **Series** tab, locate the **Degree** setting, and change the value from 2 to 3.



Switching the degree to 3 (a cubic model) immediately updates the trendline, formula, and [R-squared](#) calculation on the chart:

y vs. x



The new cubic [polynomial regression](#) equation is now: $y = 37.2 - 14.2x + 2.64x^2 - 0.126x^3$. Most importantly, the [R-squared](#) value experiences a dramatic increase, jumping to **0.976**. This substantial gain confirms that the cubic model is far superior at capturing the complex distribution of the underlying data. Although increasing the degree further (e.g., to degree 4) might marginally increase the R-squared, the gain is minor and often introduces unnecessary complexity, risking overfitting. Therefore, the cubic model (Degree 3) provides the optimal balance of high explanatory power and model parsimony.

Step 5: Utilizing the Fitted Equation for Prediction

The primary operational goal of any [regression analysis](#) is to develop a reliable mathematical tool that can generate predictions. Once the optimal [polynomial regression](#) model has been identified (in our case, the superior cubic model with $R\text{-squared} = 0.976$), we can use its derived equation to forecast the expected value of the [response variable](#) (Y) for any new, hypothetical value of the predictor variable (X).

We will use the established cubic equation:

$$y = 37.2 - 14.2x + 2.64x^2 - 0.126x^3$$

Consider a scenario where we need to determine the expected value for Y if the predictor variable X is equal to 4. We substitute the value 4 into every term of the fitted equation:

$$y = 37.2 - 14.2(4) + 2.64(4)^2 - 0.126(4)^3$$

$$y = 37.2 - 56.8 + 2.64(16) - 0.126(64)$$

$$y = 37.2 - 56.8 + 42.24 - 8.064$$

After performing the necessary arithmetic operations, the final expected value for Y when X = 4 is calculated as **14.576**. This final step showcases the practical utility of performing a rigorous [polynomial regression](#) analysis in [Google Sheets](#), enabling accurate forecasts even when data exhibits complex, nonlinear patterns.

For analysts seeking to deepen their understanding of statistical modeling and data visualization techniques beyond polynomial regression, we offer numerous advanced [Google Sheets](#) tutorials covering complex statistical methods on our resources page.