

How to Easily Create Partial Residual Plots in R with visreg()

Authored by
stats writer

December 1, 2025

RECOMMENDED CITATION

stats writer (2025). *How to Easily Create Partial Residual Plots in R with visreg()*.
PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=103445>

Creating partial residual plots in R is a vital diagnostic step in regression analysis. These plots offer an exceptional visual tool for assessing the validity of the linearity assumption for individual predictor variables within a complex model. While there are several methods available, one primary approach involves using the `visreg()` function found within the dedicated visreg package.

The `visreg()` function is highly versatile, requiring the fitted model object, the specific response variable, the predictor variable of interest, and the desired type of plot as its primary inputs. By plotting the partial residuals, this function isolates and visualizes the marginal effect of a chosen predictor variable on the response variable, effectively stripping away the linear effects of all other covariates in the model. This capability is crucial for identifying potential non-linear patterns or complex curvature in the data that might otherwise be masked in standard scatterplots or overall residual analysis.

The Foundation of Multiple Linear Regression

Multiple linear regression is a fundamental statistical method widely used to quantify and understand the relationship between a dependent variable (the response) and two or more independent variables (the predictors). It provides coefficients that estimate the change in the response variable associated with a one-unit change in a predictor variable, holding all other predictors constant. This methodology forms the bedrock of predictive modeling and causal inference in various fields, including economics, psychology, and engineering.

However, the reliability of a multiple linear regression model hinges on satisfying several key assumptions. One of the most critical of these assumptions is that there exists a strictly **linear relationship** between each predictor variable and the response variable, once the effects of all other predictors in the model have been accounted for. This linearity constraint is fundamental to the interpretation of the regression coefficients.

If this core assumption of linearity is violated--meaning the true relationship is quadratic, exponential, or follows some other non-linear curve--then the parameter estimates derived from the standard linear regression model can be significantly biased, resulting in unreliable inferences and poor predictive performance. It is therefore paramount for practitioners to employ diagnostic tools capable of identifying such discrepancies before relying on the model's output.

Understanding Partial Residual Plots

A **partial residual plot**, often also referred to as a component-plus-residual plot, serves as an excellent graphical diagnostic tool specifically designed to check the assumption of linearity for each predictor variable individually within the context of a multiple regression model. Unlike a simple scatterplot of the response variable versus a single predictor, which ignores the influence of

other variables, the partial residual plot considers the entire model structure.

The plot displays the modified residuals (the vertical distance between the observed response and the predicted response) associated with one predictor variable against the values of that same predictor. More formally, the vertical axis plots the sum of the linear component for the variable of interest and the overall model residuals. This specialized plotting method effectively isolates the relationship between the predictor variable and the response variable, after controlling for the linear effects of all other variables included in the model.

Visually inspecting the scatter of points on a **partial residual plot** allows the analyst to determine if the relationship appears linear or if a curve would be a better fit. The plot typically includes a straight reference line (representing the linear fit) and often a smoothed curve (such as a locally weighted scatterplot smoothing, or LOESS curve) showing the actual trend of the residuals. If the actual trend deviates significantly from the straight reference line, it provides strong visual evidence suggesting that the assumption of linearity has been compromised for that specific predictor.

Step-by-Step Implementation in R

The following example provides a practical demonstration of how to generate and interpret partial residual plots for a fitted multiple linear regression model using R. We will utilize the widely respected **car** package, which provides the convenient `crPlots()` function, a specialized tool for creating these diagnostic visualizations quickly and efficiently across all predictors in the model.

For the purpose of this illustration, we will first simulate a dataset where the underlying relationships are deliberately non-linear for certain predictors, thus creating a scenario where diagnostic checking is necessary. We begin by defining the response and three predictor variables, and then fitting the initial multiple linear regression model in R.

#make this example reproducible

set.seed(0)

#define response variable (linear progression)

y <- c(1:1000)

#define three predictor variables (x2 and x3 are non-linear transformations)

x1 <- c(1:1000)*runif(n=1000)

x2 <- (c(1:1000)*rnorm(n=1000))^2

x3 <- (c(1:1000)*rnorm(n=1000))^3

#fit multiple linear regression model using standard linear terms

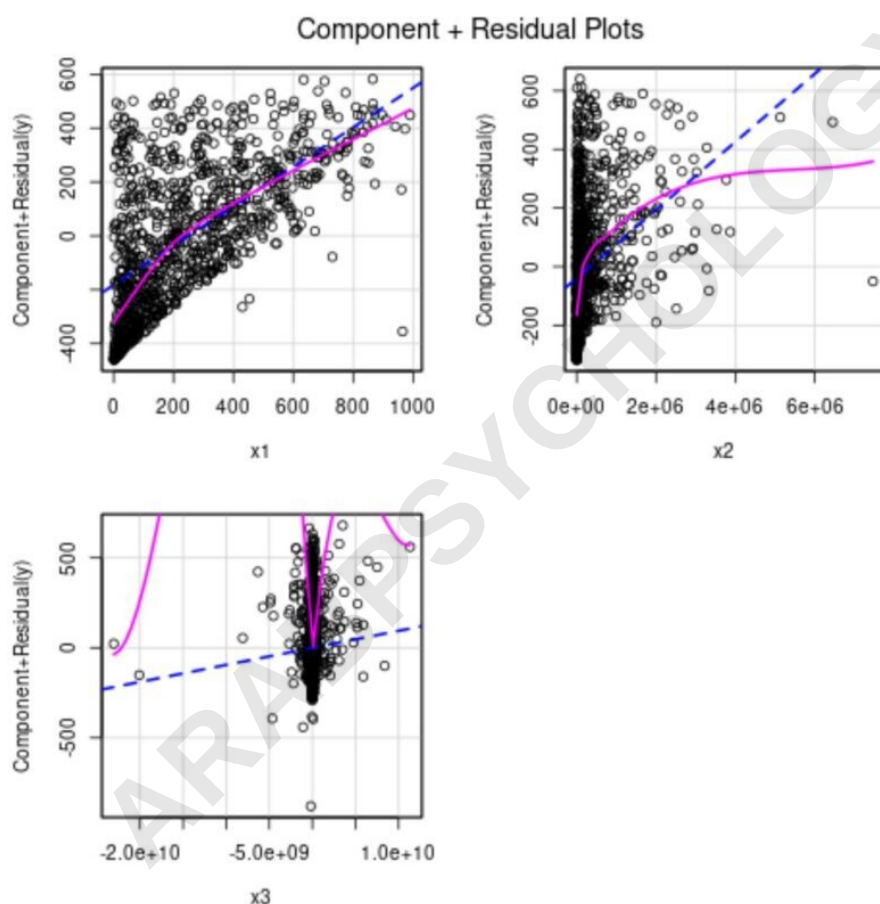
model <- lm(y~x1+x2+x3)

Once the initial model, `model`, has been successfully fitted, we can immediately proceed to generate the partial residual plots. We rely on the `crPlots()` function, which stands for "component-plus-residual plots," sourced from the `car` package (Companion to Applied Regression), a staple package for advanced regression diagnostics in R. This function accepts the fitted model object as its only mandatory argument and produces a panel of plots, one for each predictor variable included in the model.

```
library(car)
```

```
#create partial residual plots for all predictor variables in the model
```

```
crPlots(model)
```



Interpreting the Diagnostic Plots

A careful interpretation of the generated plots is essential for making informed modeling decisions. In the output generated by `crPlots()`, two distinct lines are displayed along with the scattered data points: a straight blue line and a curved pink line. The **straight blue line** represents the expected linear relationship if the assumption of linearity holds true, corresponding exactly to the

estimated coefficient for that predictor variable in the current multiple regression model.

Conversely, the **pink line** (often a smoothed LOESS curve) depicts the actual underlying trend of the partial residuals against the predictor variable. This line visually captures the non-parametric relationship suggested by the data points. If the two lines--the blue linear fit and the pink smoothed curve--are statistically indistinguishable and overlap substantially, it suggests that the linear model accurately captures the relationship, and the assumption of linearity holds for that specific predictor.

However, if the pink smoothed curve deviates significantly from the straight blue line, particularly showing distinct curvature (such as a U-shape, inverted U-shape, or exponential growth), then this constitutes clear evidence of a nonlinear relationship. This indicates that a simple linear term for that predictor is insufficient to model the data accurately, and the model must be adjusted.

Examining the plots generated above for our simulated data reveals clear diagnostic issues. While the plot for x_1 shows the pink line closely following the straight blue line, the partial residual plots for both x_2 and x_3 exhibit profound curvature. Specifically, the relationship appears quadratic or higher-order for x_2 and highly curved for x_3 . This confirms that the initial fitting of the multiple linear regression model violates the critical assumption of linearity for these two predictors, rendering the coefficients for x_2 and x_3 potentially misleading.

Addressing Non-Linearity Through Transformation

When diagnostic plots indicate a violation of the linearity assumption, the most common and effective remedy is to apply a mathematical transformation to the problematic predictor variable(s). The goal of this transformation is to linearize the relationship between the predictor and the response variable. Common transformations include the logarithmic transformation (log), the square root transformation (sqrt), or the reciprocal transformation, depending on the observed pattern of curvature in the partial residual plot.

Based on the highly non-linear patterns observed for x_2 and x_3 , we hypothesize that transforming these variables will correct the linearity violation. For x_2 , which shows signs of a quadratic relationship (as it was generated using a square), we attempt a square root transformation. For x_3 , which exhibits a highly skewed cubic relationship, we apply a more complex combination of cubic root and logarithmic transformations to attempt stabilization and linearization.

We proceed by refitting the regression model using these transformed predictor variables. It is crucial to remember that interpreting the coefficients of the transformed variables requires careful consideration, as they now represent the change in the response associated with the change in the transformed scale (e.g., log of x_3).

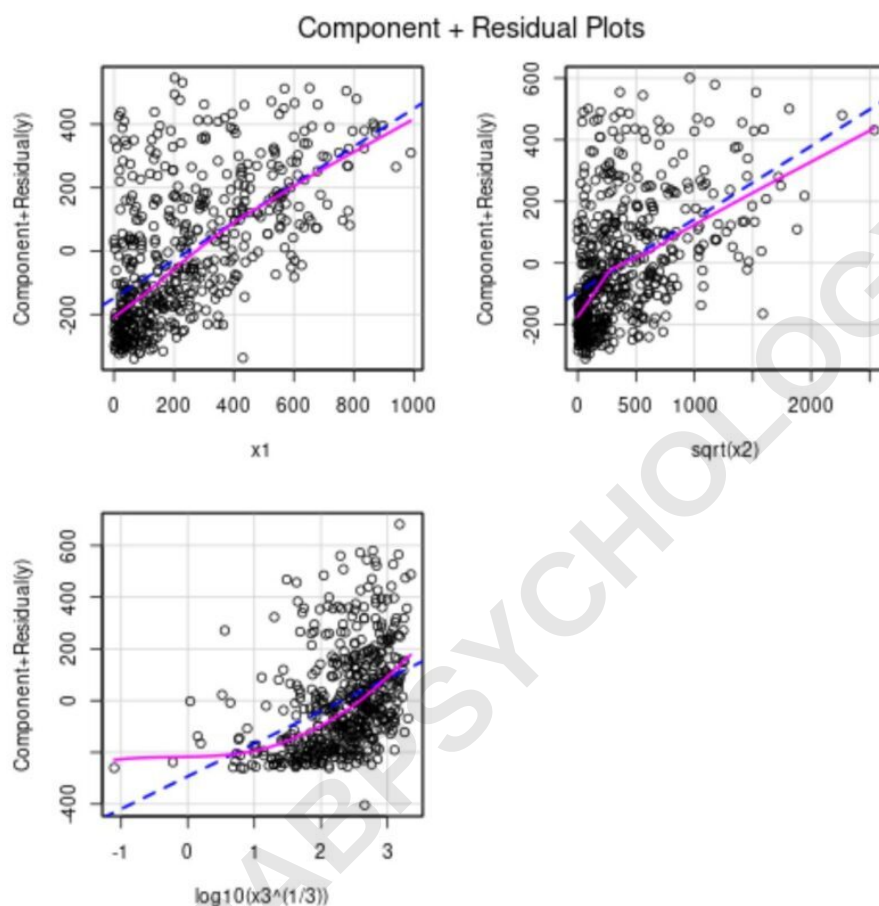
```
library(car)
```

```
#fit new model with transformed predictor variables
```

```
model_transformed <- lm(y~x1+sqrt(x2)+log10(x3^(1/3)))
```

```
#create partial residual plots for new model
```

```
crPlots(model_transformed)
```



Final Model Assessment and Refinement

Analyzing the partial residual plots for the newly transformed model reveals significant improvement. The plot for `sqrt(x2)` now shows a much closer alignment between the pink smoothed curve and the straight blue line, indicating that the square root transformation was successful in achieving linearity for this predictor. The model assumptions are better satisfied for `x2` in the new iteration.

However, the plot for the heavily transformed predictor `log10(x3^(1/3))` still shows some residual non-linearity, particularly towards the extremes of the data range. While the transformation

substantially improved the fit compared to the original model, perfect linearity has not been achieved. In such situations, the analyst has several options:

Try an alternative transformation (e.g., Box-Cox transformation).

Introduce a polynomial term (e.g., x^2) to capture the remaining curve.

Conclude that the variable x_3 is not linearly related to the response variable even after transformation and, if its contribution is not theoretically critical, consider **dropping the variable** from the model altogether to maintain model integrity and interpretability.

Further Statistical Diagnostics in R

Partial residual plots are just one of many crucial diagnostic checks. It is recommended to use them in conjunction with other tools to ensure a robust model.

The following tutorials explain how to create other common plots in R that aid in comprehensive model assessment: