

How to Create Added Variable Plots in R: A Step-by-Step Guide

Authored by
stats writer

December 4, 2025

RECOMMENDED CITATION

stats writer (2025). *How to Create Added Variable Plots in R: A Step-by-Step Guide*.
PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=105372>

The field of statistical analysis, particularly within the context of multiple linear regression, relies heavily on diagnostic tools to ensure model validity and interpretability. A cornerstone of these diagnostics available within the R statistical environment is the use of Added Variable Plots (AVPs). These plots offer a sophisticated graphical method designed specifically for assessing the marginal contribution of an individual predictor variable to the model, while simultaneously accounting for the influence of all other variables already included in the regression equation.

Far beyond simple scatter plots, AVPs provide critical visual information regarding the association between a chosen predictor and the response variable. They are indispensable for diagnosing potential issues such as nonlinearity, influential observations, or the need for transformations, all crucial steps in building a robust statistical model. Furthermore, they help analysts determine if a new variable, when added to an existing model, significantly improves the fit.

In the following detailed guide, we will explore the theoretical underpinnings of these visualizations, often alternatively known as Partial Regression Plots, and demonstrate the practical implementation within R, focusing on the powerful tools provided by the car package. Understanding how to generate and correctly interpret these diagrams is fundamental for any serious practitioner utilizing generalized linear models.

The Statistical Purpose of Added Variable Plots

In multiple linear regression, a common challenge is visualizing the unique relationship between the dependent variable and a single predictor. When dealing with complex models involving several correlated independent variables, a standard bivariate plot can be misleading because it ignores the interactions and confounding effects caused by the other predictors. This is precisely where the Added Variable Plot steps in, providing a corrected visualization.

The core concept behind an AVP is that it plots the residuals from two separate regressions. Specifically, the y-axis represents the residuals of the response variable regressed on all other predictors in the model (excluding the predictor of interest). The x-axis represents the residuals of the predictor of interest regressed on all the remaining predictors. Plotting these two sets of residuals together isolates the portion of the variability in the response that is uniquely explained by the predictor in question, thereby 'controlling' for the effects of the other variables.

The resulting slope of the regression line plotted on the AVP is mathematically identical to the coefficient assigned to that predictor in the full multiple regression equation. This dual utility—visualizing marginal contribution while confirming the coefficient magnitude and sign—makes the AVP an essential diagnostic tool for verifying model assumptions and ensuring that the estimated coefficients accurately reflect the partial relationship between the variables.

Identifying Nonlinearity and Outliers

One of the primary benefits of using Added Variable Plots is their effectiveness in detecting deviations from the fundamental assumptions of linear regression. Specifically, they help us diagnose whether the marginal relationship between the predictor and the response is truly linear. If the points on the plot show a distinct curve rather than a tight linear arrangement around the fitted line, it strongly suggests that a non-linear transformation of the predictor variable might be necessary to improve the model fit and satisfy the linearity assumption.

Furthermore, AVPs are exceptionally valuable for spotting influential observations and outliers. Influential points are observations that, if removed, would significantly change the parameters of the regression model. In an AVP, these points often appear far removed from the main cluster of data points, exhibiting both large residuals (vertical distance from the line) and high leverage (horizontal position). The plots often label the two observations with the largest residuals and the two observations with the largest partial leverage, providing the analyst with direct visibility into potential data issues.

Dealing with these outliers is critical, as they can severely skew the estimated regression coefficients and inflate standard errors, leading to incorrect statistical inference. By pinpointing these high-leverage points graphically, the researcher can make informed decisions about whether to investigate data entry errors, perform robust regression, or adjust the modeling strategy entirely.

Implementing Added Variable Plots using the car package

In the R environment, the standard method for generating high-quality Added Variable Plots is through the utilization of the car package, which stands for "Companion to Applied Regression." This package, developed by John Fox and Sanford Weisberg, provides a comprehensive suite of functions for regression analysis diagnostics, significantly enhancing the capabilities of base R.

The specific function required is **avPlots()**. This function takes a fitted linear model object, typically created using the standard **lm()** function in R, and automatically generates one added variable plot for each predictor variable included in the model. This makes the process highly efficient, especially for models containing numerous predictors.

To prepare for its use, the car package must first be installed (if not already present) and then loaded into the current R session. The general syntax is straightforward and involves specifying the fitted model object as the sole argument, as demonstrated in the basic R code provided below:

```
# Load the car package into the R session  
library(car)
```

```
# Fit the multiple linear regression model
model <- lm(y ~ x1 + x2 + ..., data = df)

# Generate the set of added variable plots for the fitted model
avPlots(model)
```

This implementation ensures that the visualizations are generated consistently and accurately, providing the necessary residual plots for comprehensive diagnostic review across all dimensions of the multiple linear regression model.

Step-by-Step Example: Using avPlots() with the mtcars Dataset

To illustrate the practical application of Added Variable Plots, we will utilize a classic dataset built into R: the mtcars dataset. This dataset contains technical specifications and performance metrics for 32 automobiles. Our goal is to model fuel efficiency (miles per gallon, **mpg**) based on several predictors: displacement (**disp**), horsepower (**hp**), and rear axle ratio (**drat**).

First, we must fit the multiple linear regression model using the **lm()** function. Subsequently, we examine the summary output to understand the calculated coefficients and overall model performance metrics, such as R-squared and the residual standard error. This initial step establishes the numerical context that the graphical plots will later confirm and diagnose.

Fit multiple linear regression model predicting mpg

```
model <- lm(mpg ~ disp + hp + drat, data = mtcars)
```

```
# View the statistical summary of the fitted model
summary(model)
```

Call:

```
lm(formula = mpg ~ disp + hp + drat, data = mtcars)
```

Residuals:

```
Min 1Q Median 3Q Max
-5.1225 -1.8454 -0.4456 1.1342 6.4958
```

Coefficients:

```
Estimate Std. Error t value Pr(>|t|)
(Intercept) 19.344293 6.370882 3.036 0.00513 **
disp -0.019232 0.009371 -2.052 0.04960 *
hp -0.031229 0.013345 -2.340 0.02663 *
drat 2.714975 1.487366 1.825 0.07863 .
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.008 on 28 degrees of freedom

Multiple R-squared: 0.775, Adjusted R-squared: 0.7509

F-statistic: 32.15 on 3 and 28 DF, p-value: 3.28e-09

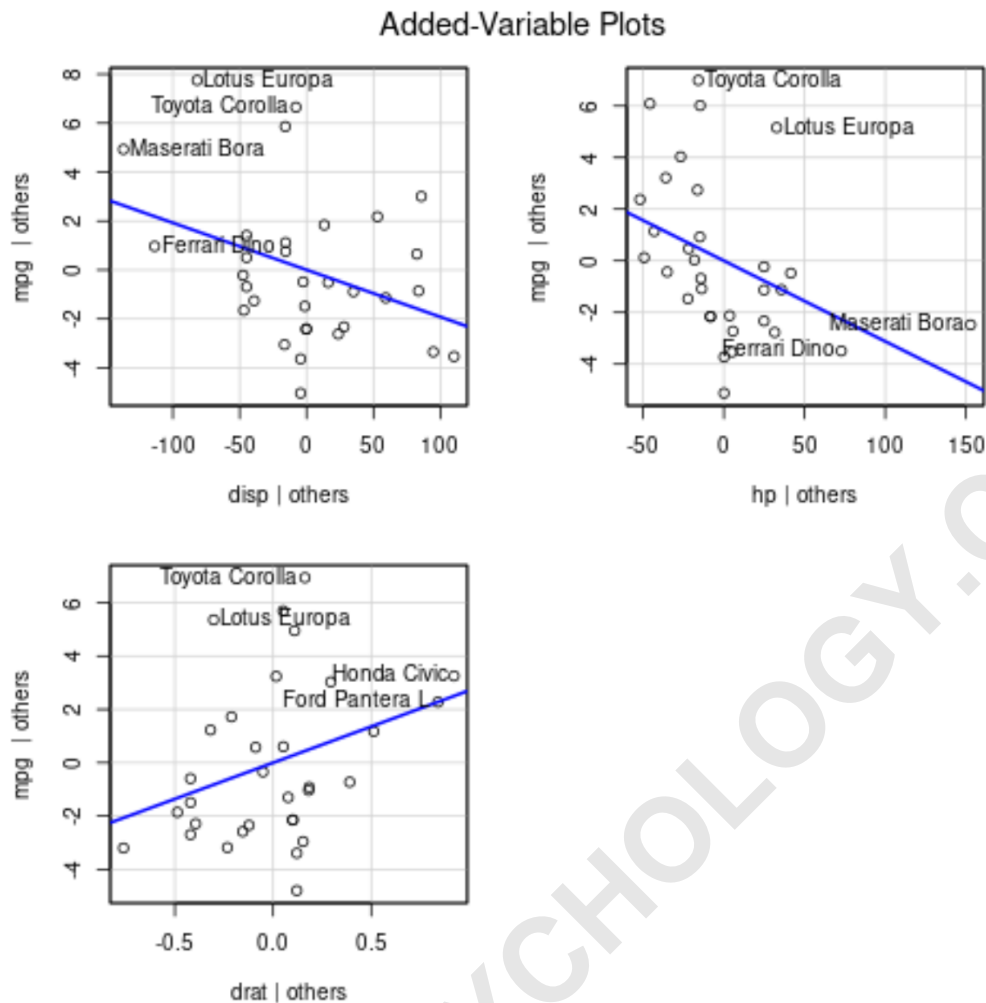
Once the model is fitted, the next logical step is to visualize the partial relationship between the response variable (**mpg**) and each individual predictor. We accomplish this by loading the [car package](#) and calling the **avPlots()** function, passing our fitted **model** object as the argument. This action generates a composite plot showing the partial effects for **disp**, **hp**, and **drat** simultaneously.

Load car package

library(car)

Produce added variable plots for the fitted model

avPlots(model)



Interpreting the Output of Added Variable Plots

Interpreting the visual information presented in Added Variable Plots is essential for a complete regression diagnosis. Each plot consists of two main axes and a fitted regression line, all of which provide specific insights into the model structure and data behavior. The visual geometry of these plots directly reflects the mathematical outcome of the partial regression calculations.

The structure of each plot is defined by the residuals: **The x-axis:** This axis displays the residuals of the predictor variable of interest (e.g., **disp**) regressed against all other predictors (**hp** and **drat**). These residuals represent the variation in that specific predictor that is independent of the other variables in the model. **The y-axis:** This axis displays the residuals of the response variable (**mpg**) regressed against all other predictors. These residuals represent the unexplained variation in the response variable after accounting for the influence of every other variable except the one currently plotted on the x-axis.

The central feature is the blue regression line, which represents the estimated partial linear

relationship between the predictor and the response, having controlled for the influence of the remaining covariates. The slope of this line is, critically, the exact estimated regression coefficient for that predictor from the multiple linear regression model. For instance, if the line slopes downwards sharply, it confirms a strong negative association, consistent with the negative numerical coefficient in the summary output.

A crucial component of the **avPlots()** output is the labeling of specific data points. The points that are labeled in each plot typically represent the two observations with the largest standardized residuals and the two observations with the largest partial leverage. These highlighted points require careful scrutiny, as they are the candidates for being influential outliers that might distort the overall estimation of the regression line. Analysts should cross-reference these labeled observations with the raw mtcars dataset or case identification numbers to understand their specific impact.

The x-axis displays the unique, unexplained portion of the single predictor variable after accounting for all other covariates.

The y-axis displays the residual variation of the response variable, independent of all other predictors in the model.

The blue line illustrates the partial association between the predictor and the response variable, its slope matching the regression coefficient.

Labeled points identify the observations exhibiting the largest residuals and the highest partial leverage, signaling potentially influential data points.

Understanding the Geometry of the Plots and Coefficient Alignment

The powerful congruence between the graphical representation and the numeric statistical summary is a defining characteristic of Added Variable Plots. The visual angle and direction of the fitted regression line in each AVP must perfectly align with the sign and magnitude of the corresponding coefficient derived from the estimated regression equation. This serves as a vital cross-check for the integrity of the model fitting process.

For example, in our **mtcars** analysis, the statistical summary provided the following estimated coefficients for the predictors:

disp: -0.019232

hp: -0.031229

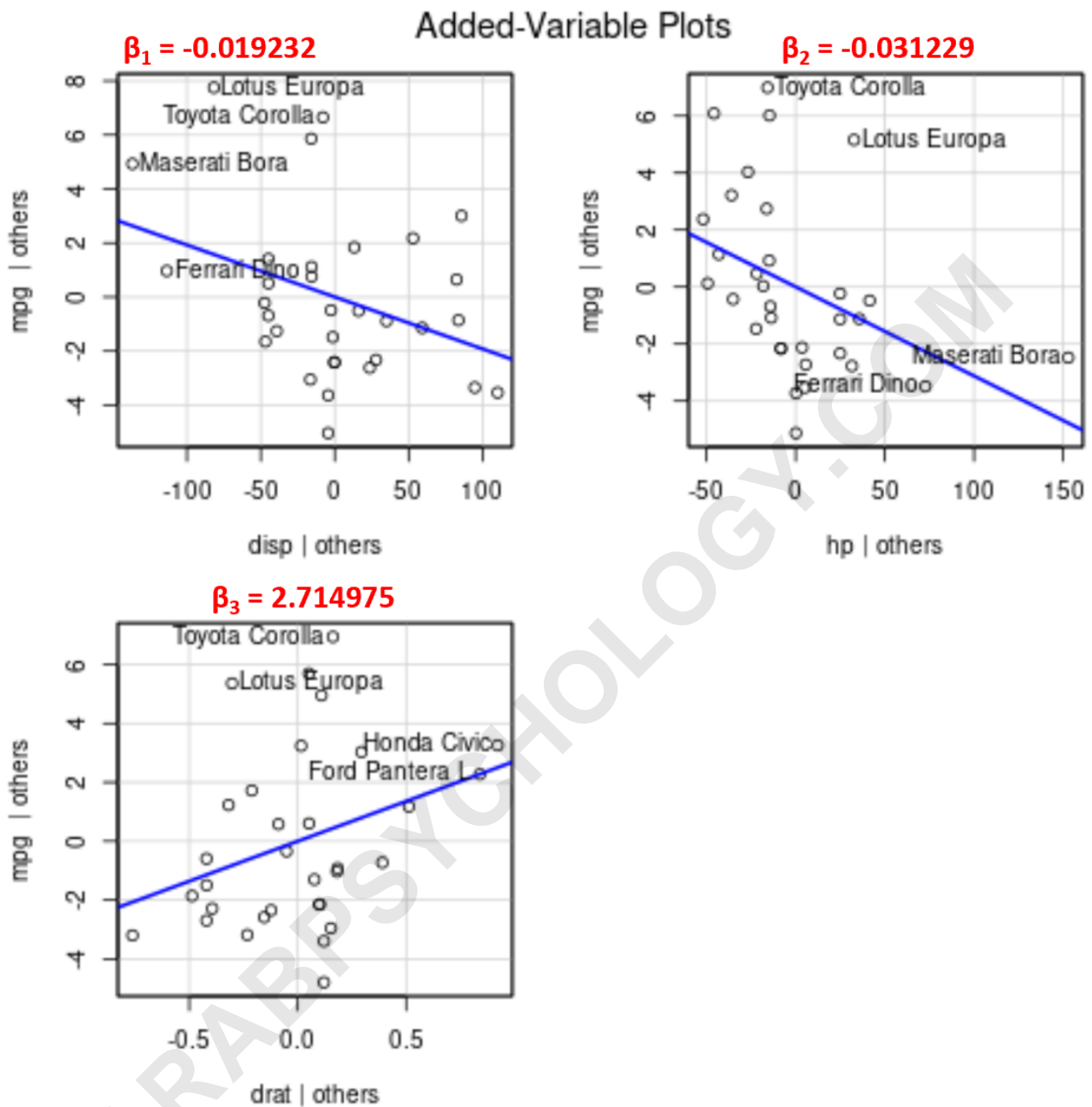
drat: 2.714975

When examining the generated plots (Figure 1), we observe that the lines for **disp** and **hp** both

slope downward, visually confirming their negative partial relationships with **mpg**. Conversely, the line for **drat** slopes upward, clearly indicating its positive partial association. The steepness of the line, while not directly readable as the numerical coefficient, provides an intuitive sense of the strength of the relationship relative to the scale of the partial residuals.

If, for instance, the regression coefficient for **drat** had been positive (as it is, +2.71) but the corresponding AVP line sloped downward, it would immediately signal a fundamental error in the statistical calculation or plotting function. This direct visual validation is what makes AVPs such a reliable diagnostic tool for validating the mathematical estimates within a multiple linear regression framework.

ARABPSYCHOLOGY.COM



Conclusion and Best Practices

The implementation of Added Variable Plots in R via the `avPlots()` function from the `car` package is an indispensable practice for anyone performing sophisticated regression analysis. These plots move beyond the limitations of simple correlation and bivariate scatterplots by graphically representing the true partial relationship between variables, isolated from the noise and confounding effects of other predictors in the model.

By using AVPs, analysts gain the ability to conveniently visualize linearity assumptions, identify

high-leverage observations, and confirm the direction and slope of estimated regression coefficients. This level of rigorous visual diagnostics ensures greater confidence in the resulting model, leading to more accurate interpretations and robust predictive capabilities. Incorporating these plots into every model diagnostic routine is considered a hallmark of high-quality statistical practice.

In summary, while the numerical summary of a linear model provides the precise estimates, the added variable plot offers the essential visual confirmation required to validate the underlying assumptions and detect data anomalies that might otherwise be masked by the complexity of a multivariate analysis. Always leverage these plots to thoroughly understand the marginal contribution of each variable before finalizing your regression model.

ARABPSYCHOLOGY.COM