

How do you create and interpret pairs plots in R?

Authored by
stats writer

April 18, 2024

RECOMMENDED CITATION

stats writer (2024). *How do you create and interpret pairs plots in R?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=136599>

Pairs plots in R are graphical representations of the pairwise relationships between multiple variables in a dataset. These plots are created using the R programming language and provide a quick and easy way to visualize the relationships between different variables. To create a pairs plot, the R user can use the "pairs" function, which takes in a data frame as an input and produces a grid of scatter plots, with each variable plotted against every other variable.

Interpreting pairs plots involves analyzing the patterns and trends in the scatter plots. The plots can help identify any linear or non-linear relationships between the variables, as well as any outliers or clusters of data points. The diagonal plots in a pairs plot show the distribution of each individual variable, providing information about its range and spread. Pairs plots are useful for gaining insights into the relationships between variables and can aid in identifying potential correlations or trends in the data. They are a valuable tool for exploratory data analysis and can assist in making data-driven decisions.

Create and Interpret Pairs Plots in R

A pairs plot is a matrix of scatterplots that lets you understand the pairwise relationship between different variables in a dataset.

Fortunately it's easy to create a pairs plot in R by using the `pairs()` function. This tutorial provides several examples of how to use this function in practice.

Example 1: Pairs Plot of All Variables

The following code illustrates how to create a basic pairs plot for all variables in a data frame in R:

```
#make this example reproducible  
set.seed(0)
```

```
#create data frame
```

```
var1 <- rnorm(1000)
```

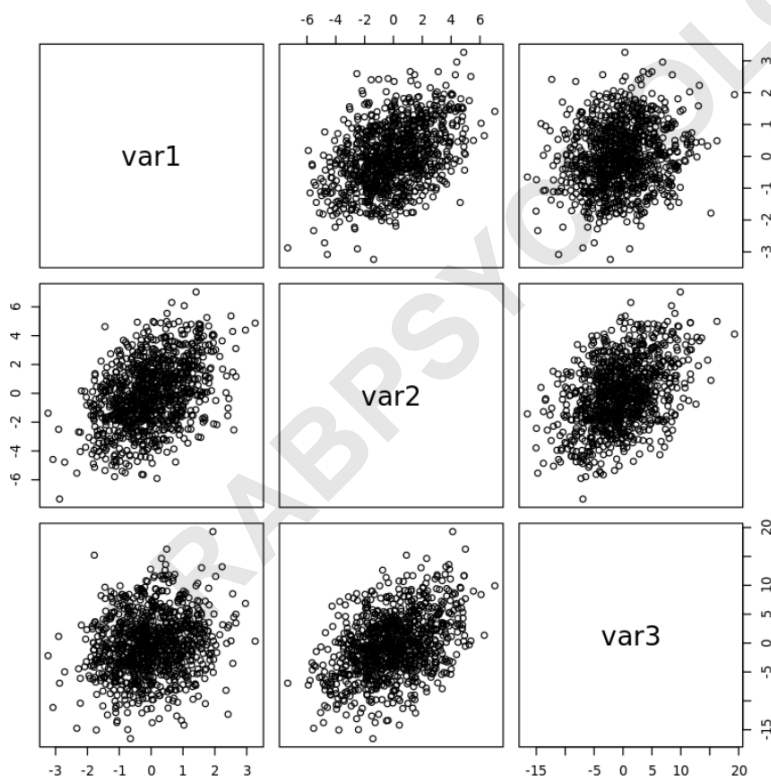
```
var2 <- var1 + rnorm(1000, 0, 2)
```

```
var3 <- var2 - rnorm(1000, 0, 5)
```

```
df <- data.frame(var1, var2, var3)
```

```
#create pairs plot
```

```
pairs(df)
```



The way to interpret the matrix is as follows:

The variable names are shown along the diagonals

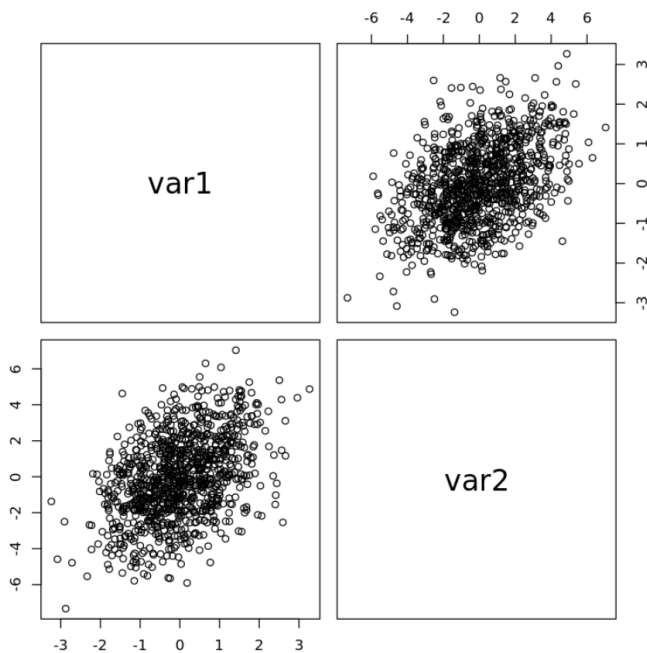
boxes. All other boxes display a scatterplot of the relationship between each pairwise combination of variables. For example, the box in the top right corner of the matrix displays a scatterplot of values for var1 and var3. The box in the middle left displays a scatterplot of values for var1 and var2, and so on.

This single plot gives us an idea of the relationship between each pair of variables in our dataset. For example, var1 and var2 seem to be positively correlated while var1 and var3 seem to have little to no correlation.

Example 2: Pairs Plot of Specific Variables

The following code illustrates how to create a basic pairs plot for just the first two variables in a dataset:

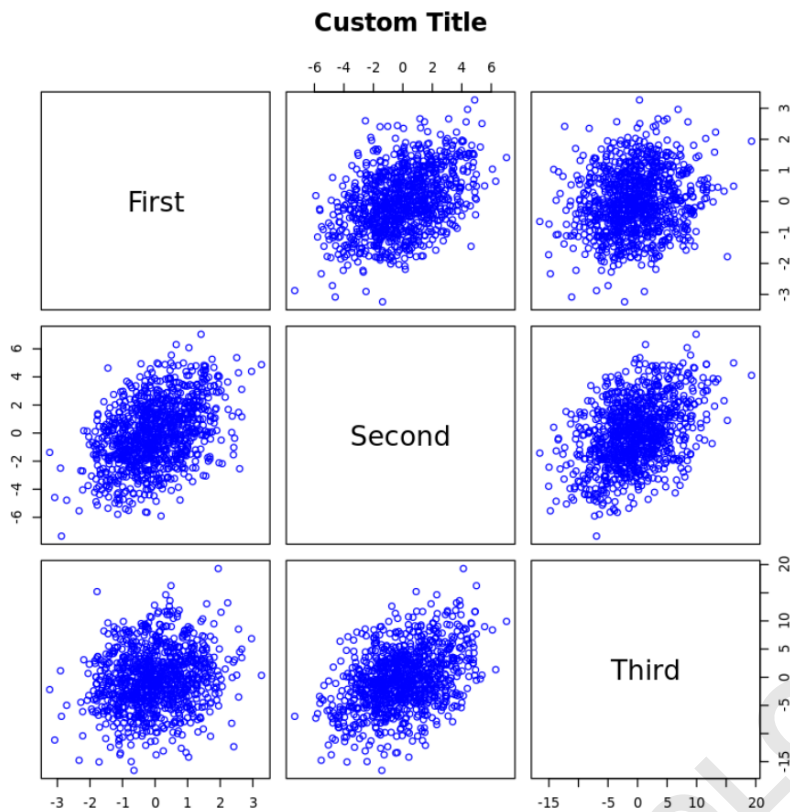
```
#create pairs plot for var1 and var2 only  
pairs(df)
```



Example 3: Modify the Aesthetics of a Pairs Plot

The following code illustrates how to modify the aesthetics of a pairs plot, including the title, the color, and the labels:

```
pairs(df,  
col = 'blue', #modify color  
labels = c('First', 'Second', 'Third'), #modify labels  
main = 'Custom Title') #modify title
```



Example 4: Obtaining Correlations with ggpairs

You can also obtain the Pearson correlation coefficient between variables by using the `ggpairs()` function from the `GGally` library. The following code illustrates how to use this function:

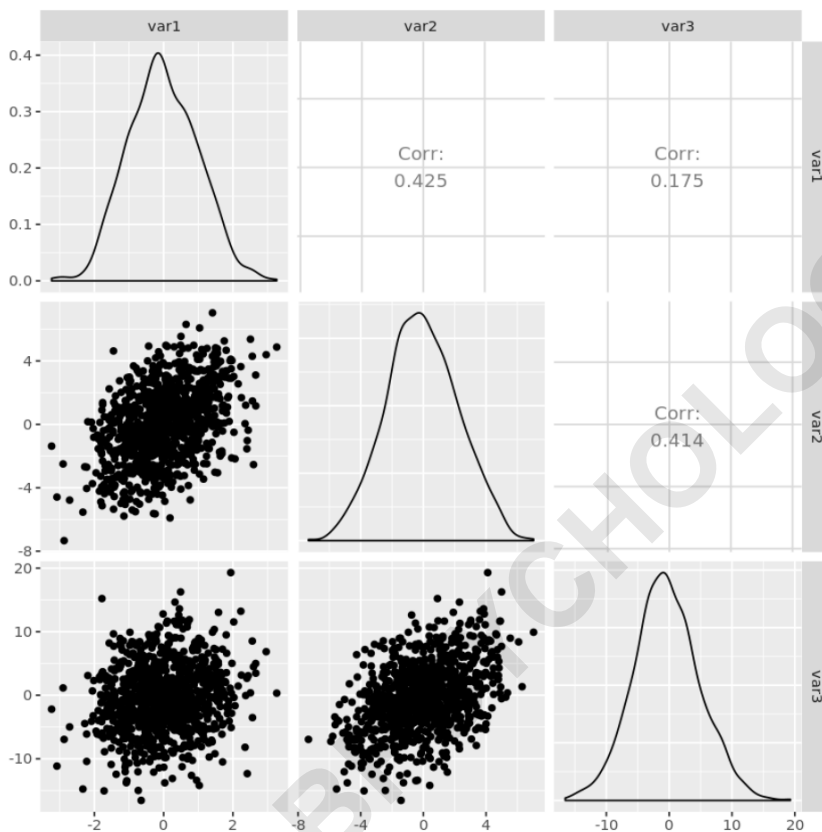
```
#install necessary libraries
install.packages('ggplot2')
install.packages('GGally')
```

```
#load libraries
library(ggplot2)
```

```
library(GGally)
```

```
#create pairs plot
```

```
ggpairs(df)
```



The way to interpret this matrix is as follows:

The variable names are displayed on the outer edges of the matrix. The boxes along the diagonals display the density plot for each variable. The boxes in the lower left corner display the scatterplot between each variable. The boxes in the upper right corner display the

Pearson correlation coefficient between each variable. For example, the correlation between var1 and var2 is 0.425.

The benefit of using `ggpairs()` over the base R function `pairs()` is that you can obtain more information about the variables. Specifically, you can see the correlation coefficient between each pairwise combination of variables as well as a density plot for each individual variable.

You can find the complete documentation for the `ggpairs()` function [here](#).