

How do I create a correlation heatmap in R with an example?

Authored by
stats writer

June 27, 2024

RECOMMENDED CITATION

stats writer (2024). *How do I create a correlation heatmap in R with an example?*.

PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=154778>

A correlation heatmap is a graphical representation of the correlations between different variables in a dataset. It uses color coding to visually display the strength and direction of the correlation between variables. To create a correlation heatmap in R, you can use the "corrplot" package. This package allows you to customize the appearance of the heatmap and also provides options for calculating different types of correlations. For example, you can use the following code to create a correlation heatmap for the "mtcars" dataset:

```
library(corrplot)
data(mtcars)
corrplot(cor(mtcars), method="color")
```

This will generate a heatmap with different colors representing the correlation strength, and the variables on both the x and y axes. This tool can be useful in identifying relationships between variables and identifying patterns in large datasets.

Create a Correlation Heatmap in R (With Example)

You can use the following basic syntax to create a correlation heatmap in R:

```
#calculate correlation between each pairwise
combination of variables
cor_df <- round(cor(df), 2)
```

```
#melt the data frame
melted_cormat <- melt(cor_df)
```

```
#create correlation heatmap
ggplot(data = melted_cormat, aes(x=Var1, y=Var2,
fill=value)) +
geom_tile() +
```

```
geom_text(aes(Var2, Var1, label = value), size = 5) +  
scale_fill_gradient2(low = "blue", high = "red",  
limit = c(-1,1), name="Correlation") +  
theme(axis.title.x = element_blank(),  
axis.title.y = element_blank(),  
panel.background = element_blank())
```

The following example shows how to use this syntax in practice.

Example: Create Correlation Heatmap in R

Suppose we have the following data frame in R that shows various statistics for eight different basketball players:

```
#create data frame
```

```
df <- data.frame(points=c(22, 25, 30, 16, 14, 18, 29, 22),  
assists=c(4, 4, 5, 7, 8, 6, 7, 12),  
rebounds=c(10, 7, 7, 6, 8, 5, 4, 3),  
blocks=c(12, 4, 4, 6, 5, 3, 8, 5))
```

```
#view data frame
```

```
df
```

```
points assists rebounds blocks
```

1 22 4 10 12

2 25 4 7 4

3 30 5 7 4

4 16 7 6 6

5 14 8 8 5

6 18 6 5 3

7 29 7 4 8

8 22 12 3 5

Suppose we would like to create a correlation heatmap to visualize the between each pairwise combination of variables in the data frame.

Before we create the correlation heatmap, we must first calculate the correlation coefficient between each variable using `cor()` and then transform the results into a usable format using the `melt()` function from the `reshape2` package:

```
library(reshape2)
```

```
#calculate correlation coefficients, rounded to 2 decimal places
```

```
cor_df <- round(cor(df), 2)
```

#melt the data frame

```
melted_cor <- melt(cor_df)
```

#view head of melted data frame

```
head(melted_cor)
```

Var1 Var2 value

1 points points 1.00

2 assists points -0.27

3 rebounds points -0.16

4 blocks points 0.10

5 points assists -0.27

6 assists assists 1.00

Next, we can use the `geom_tile()` function from the `ggplot2` package to create correlation heatmap:

```
library(ggplot2)
```

#create correlation heatmap

```
ggplot(data = melted_cor, aes(x=Var1, y=Var2,  
fill=value)) +
```

```
geom_tile() +
```

```
geom_text(aes(Var2, Var1, label = value), size = 5) +
```

```
scale_fill_gradient2(low = "blue", high = "red",
```

```
limit = c(-1,1), name="Correlation") +  
theme(axis.title.x = element_blank(),  
axis.title.y = element_blank(),  
panel.background = element_blank())
```



The result is a correlation heatmap that allows us to visualize the correlation coefficient between each pairwise combination of variables.

In this particular heatmap, the correlation coefficients take on the following colors:

Blue if they are close to -1
White if they are close to

0Red if they are close to 1

Feel free to use whatever colors you'd like for the low and high arguments within the `scale_fill_gradient2()` function.

For example, you could instead use "red" for the low value and "green" for the high value:

```
library(ggplot2)
```

```
#create correlation heatmap
```

```
ggplot(data = melted_cor, aes(x=Var1, y=Var2,  
fill=value)) +  
geom_tile() +  
geom_text(aes(Var2, Var1, label = value), size = 5) +  
scale_fill_gradient2(low = "red", high = "green",  
limit = c(-1,1), name="Correlation") +  
theme(axis.title.x = element_blank(),  
axis.title.y = element_blank(),  
panel.background = element_blank())
```



Note: You can also specify hex color codes to use if you'd like even more control over the exact colors in the correlation heatmap.

The following tutorials explain how to perform other common tasks in ggplot2: