

# How can we calculate correlation by group in R?

Authored by  
**stats writer**

June 29, 2024

## RECOMMENDED CITATION

stats writer (2024). *How can we calculate correlation by group in R?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=159346>

Correlation by group in R refers to the process of calculating the relationship between two variables within different subgroups of a dataset. This method allows for the examination of how the correlation between two variables may differ among different groups. In order to calculate correlation by group in R, the data must first be divided into separate groups. Then, the correlation coefficient can be calculated for each group using the appropriate function in R, such as "cor" or "cor.test". This approach allows for a more in-depth analysis of the relationship between variables, as it takes into account the variability within different groups.

## Calculate Correlation By Group in R

**You can use the following basic syntax to calculate the correlation between two variables by group in R:**

```
library(dplyr)
```

```
df %>%
```

```
  group_by(group_var) %>%
```

```
  summarize(cor=cor(var1, var2))
```

**This particular syntax calculates the between var1 and var2, grouped by group\_var.**

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

### **Example: Calculate Correlation By Group in R**

**Suppose we have the following data frame that contains information about basketball players on various teams:**

```
#create data frame
```

```
df <- data.frame(team=c('A', 'A', 'A', 'A', 'B', 'B', 'B', 'B'),  
points=c(18, 22, 19, 14, 14, 11, 20, 28),  
assists=c(2, 7, 9, 3, 12, 10, 14, 21))
```

```
#view data frame
```

```
df
```

```
team points assists
```

```
1 A 18 2
```

```
2 A 22 7
```

```
3 A 19 9
```

```
4 A 14 3
```

```
5 B 14 12
```

```
6 B 11 10
```

```
7 B 20 14
```

```
8 B 28 21
```

We can use the following syntax from the package to calculate the correlation between points and assists, grouped by team:

```
library(dplyr)
```

```
df %>%
```

```
group_by(team) %>%  
summarize(cor=cor(points, assists))
```

```
# A tibble: 2 x 2
```

```
team cor
```

```
1 A 0.603
```

```
2 B 0.982
```

From the output we can see:

The correlation coefficient between points and assists for team A is .603. The correlation coefficient between points and assists for team B is .982.

Since both correlation coefficients are positive, this tells us that the relationship between points and assists for both teams is positive.

Related:

Additional Resources