

# How can I use ggplot2 to plot a regression line by group?

Authored by  
**stats writer**

April 24, 2024

## RECOMMENDED CITATION

stats writer (2024). *How can I use ggplot2 to plot a regression line by group?*.  
PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=138924>

GGplot2 is a powerful data visualization tool that allows for the creation of high-quality graphs and charts. One useful feature of GGplot2 is the ability to plot regression lines by group. This means that you can visually represent the relationship between two variables while also considering the differences between different categories or groups within your data. To do this, you can use the "facet\_wrap" function in GGplot2 to separate your data into groups, and then use the "geom\_smooth" function to add a regression line for each group. This allows for a more in-depth analysis and understanding of the relationship between variables within specific groups.

## Plot a Regression Line by Group with ggplot2

We can use the following syntax to plot a regression line by group using the R visualization package :

```
ggplot(df, aes(x = x_variable, y = y_variable, color =  
group_variable)) +  
geom_point() +  
geom_smooth(method = "lm", fill = NA)
```

This tutorial provides a quick example of how to use this function in practice.

Example: Plot Regression Lines by Group with ggplot2

Suppose we have the following dataset that shows the following three variables for 15 different students:

Number of hours studied  
Exam score received  
Study technique used (either A, B, or C)

**#create dataset**

```
df <- data.frame(hours=c(1, 2, 3, 3, 4, 1, 2, 2, 3, 4, 1, 2, 3,
4, 4),
score=c(84, 86, 85, 87, 94, 74, 76, 75, 77, 79, 65, 67, 69,
72, 80),
technique=rep(c('A', 'B', 'C'), each=5))#view dataset
df
```

**hours score technique**

1 1 84 A

2 2 86 A

3 3 85 A

4 3 87 A

5 4 94 A

6 1 74 B

7 2 76 B

8 2 75 B

9 3 77 B

10 4 79 B

11 1 65 C

12 2 67 C

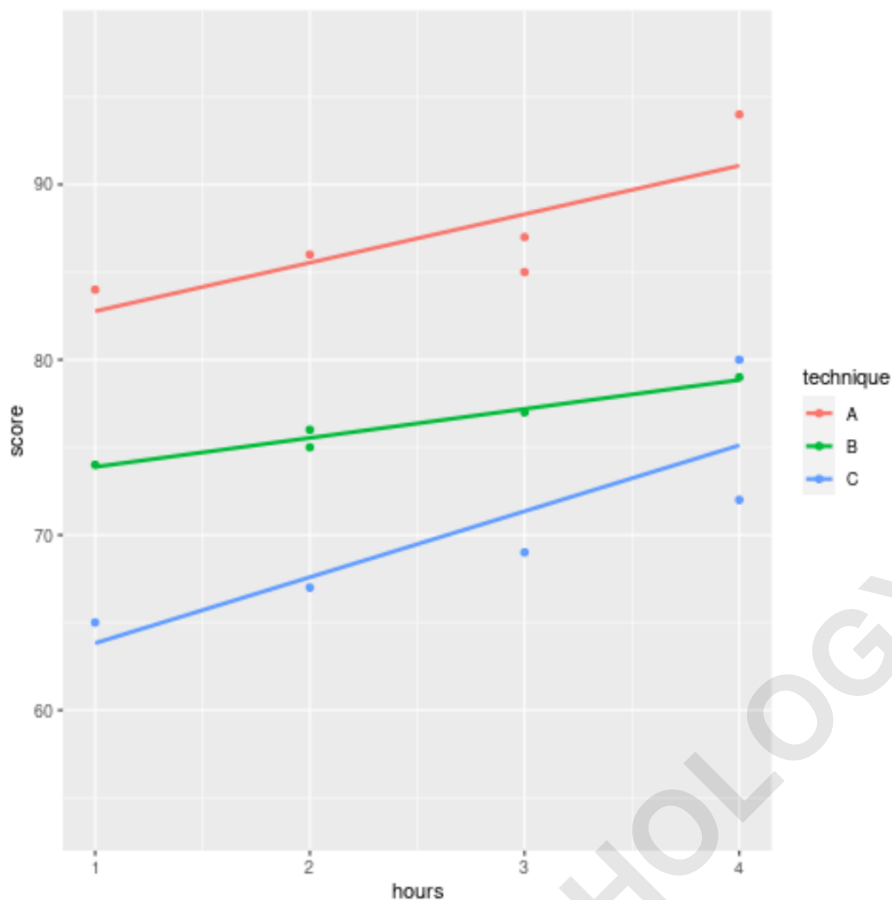
13 3 69 C

14 4 72 C

15 4 80 C

The following code shows how to plot a regression line that captures the relationship between hours studied and exam score received for each of the three study techniques:

```
#load ggplot2  
library(ggplot2)  
#create regression lines for all three groups  
ggplot(df, aes(x = hours, y = score, color = technique)) +  
geom_point() +  
geom_smooth(method = "lm", fill = NA)
```

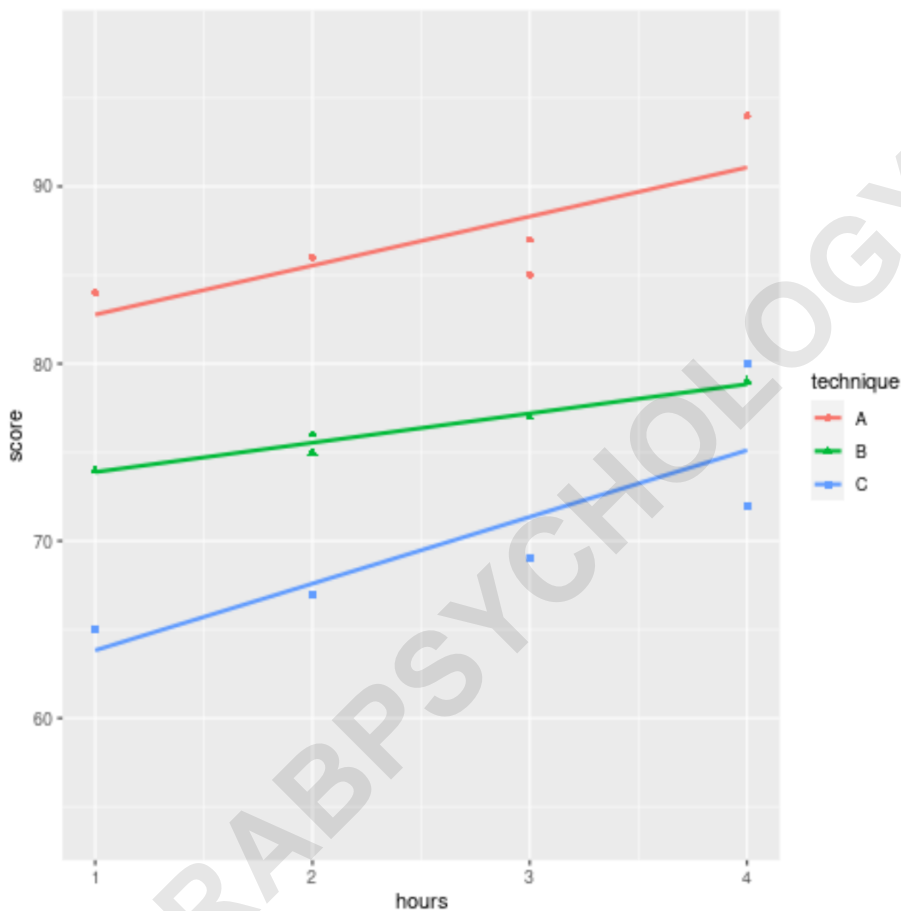


**Note that in `geom_smooth()` we used `method = 'lm'` to specify a linear trend.**

**We could also use other smoothing methods like `"glm"`, `"loess"`, or `"gam"` to capture nonlinear trends in the data. You can find the full documentation for `geom_smooth()` .**

**Note that we could also use different shapes to display the exam scores for each of the three groups:**

```
ggplot(df, aes(x = hours, y = score, color = technique,  
shape = technique)) +  
geom_point() +  
geom_smooth(method = "lm", fill = NA)
```



You can find more ggplot2 tutorials .