

How can I create a plot with a mean line for each group in ggplot2?

Authored by
stats writer

June 25, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I create a plot with a mean line for each group in ggplot2?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=151740>

To create a plot with a mean line for each group in ggplot2, first, organize your data into groups. Then, use the "group" parameter in the aes() function to specify the grouping variable. Next, use the stat_summary() function to calculate the mean for each group and plot it as a line. Finally, use the geom_line() function to add the mean lines to your plot. This will create a visual representation of the mean values for each group in your data.

Plot Mean Line by Group in ggplot2

You can use the following basic syntax to plot a mean line by group in ggplot2:

```
#calculate mean points value by team  
mean_team <- df %>% group_by(team) %>%  
summarise(mean_pts=mean(points))
```

```
#create scatterplot of assists vs points with mean line  
of points by team  
ggplot(df, aes(x=assists, y=points)) +  
geom_point(aes(color=team)) +  
geom_hline(data=mean_team,  
aes(yintercept=mean_pts, col=team))
```

This particular example creates a scatterplot of the variables assists vs. points, then adds a line to represent the mean points value grouped by the team variable.

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

Example: Plot Mean Line by Group in ggplot2

Suppose we have the following data frame in R that contains information about points and assists for basketball players on three different teams:

```
#create data frame
```

```
df <- data.frame(team=rep(c('A', 'B', 'C'), each=5),  
assists=c(2, 4, 4, 5, 6, 6, 7, 7,  
8, 9, 7, 8, 13, 14, 12),  
points=c(8, 8, 9, 9, 10, 9, 12, 13,  
14, 15, 14, 14, 16, 19, 22))
```

```
#view data frame
```

```
df
```

```
team assists points
```

```
1 A 2 8
```

```
2 A 4 8
```

```
3 A 4 9
```

```
4 A 5 9
```

```
5 A 6 10
```

```
6 B 6 9
```

7 B 7 12

8 B 7 13

9 B 8 14

10 B 9 15

11 C 7 14

12 C 8 14

13 C 13 16

14 C 14 19

15 C 12 22

We can use the following code to create a scatterplot of the variables assists vs. points, then add a line to represent the mean points value grouped by the team variable.

```
library(dplyr)
```

```
library(ggplot2)
```

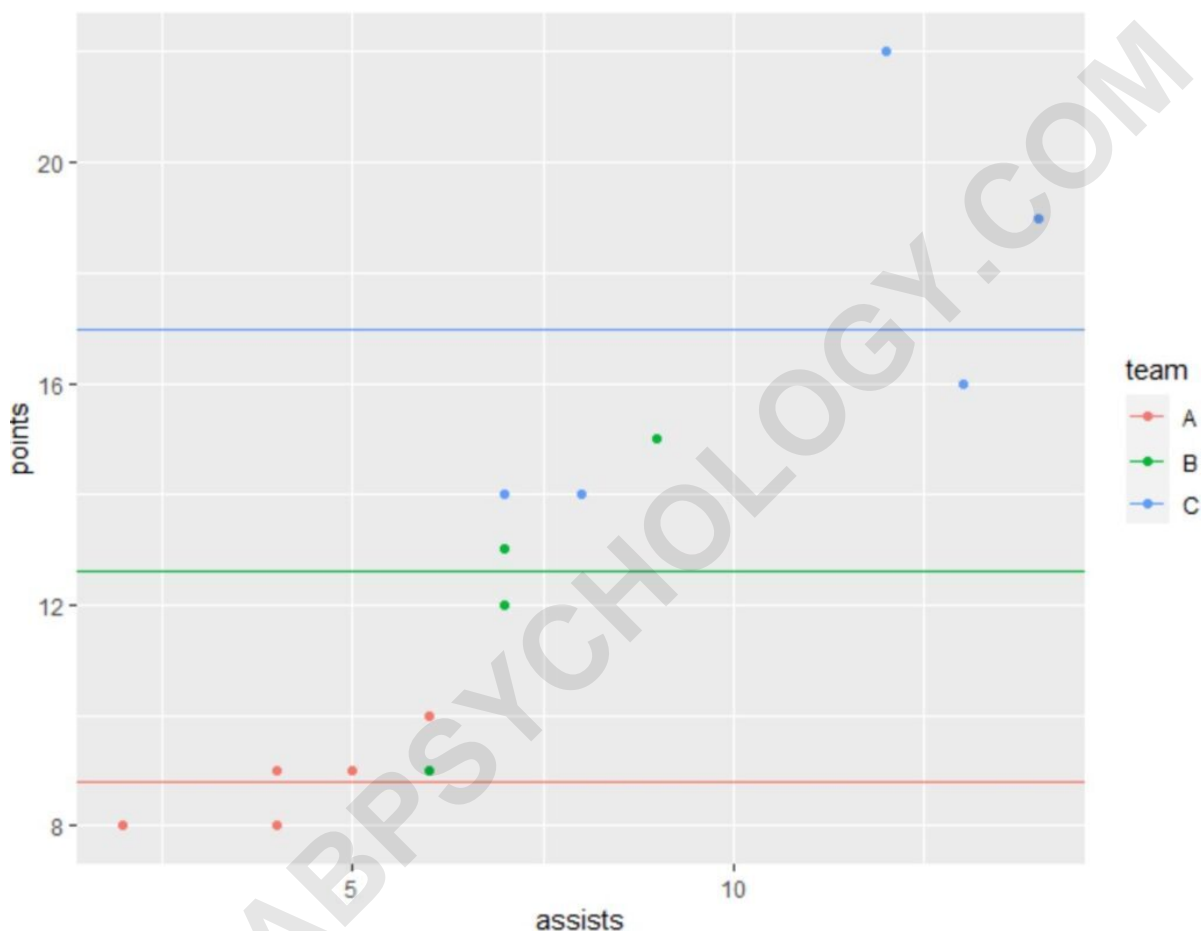
```
#calculate mean points value by team
```

```
mean_team <- df %>% group_by(team) %>%  
summarise(mean_pts=mean(points))
```

```
#create scatterplot of assists vs points with mean line  
of points by team
```

```
ggplot(df, aes(x=assists, y=points)) +
```

```
geom_point(aes(color=team)) +  
geom_hline(data=mean_team,  
aes(yintercept=mean_pts, col=team))
```



The three lines are color-coded to display the mean points value for each team.

We can display the mean_team data frame we created to see the actual mean points values for each team:

```
#view mean points value by team
```

```
mean_team`summarise()` ungrouping output (override  
with `.groups` argument)
```

```
# A tibble: 3 x 2
```

```
team mean_pts
```

```
1 A 8.8
```

```
2 B 12.6
```

```
3 C 17
```

From the output we can see:

The mean points value for players on team A is 8.8
The mean points value for players on team B is 12.6
The mean points value for players on team C is 17

These values match the locations of the lines on the y-axis of the scatterplot that we created.