

How can I calculate the percentage by group in R, with an example?

Authored by
stats writer

June 28, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I calculate the percentage by group in R, with an example?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=155886>

In order to calculate the percentage by group in R, you can use the "group_by" and "summarise" functions from the dplyr package. First, use the "group_by" function to group your data by a specific variable. Then, use the "summarise" function to calculate the percentage for each group by dividing the count of each group by the total count of the data. For example, if you have a dataset with sales data for different regions, you can use the above steps to calculate the percentage of sales for each region. This will give you a clear understanding of the distribution of sales across different regions.

Calculate Percentage by Group in R (With Example)

You can use the following syntax to calculate a percentage by group in R:

```
library(dplyr)
```

```
df %>%
```

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

```
  mutate(percent = value_var/sum(value_var))
```

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

Example: Calculate Percentage by Group in R

Suppose we have the following data frame that shows the points scored by basketball players on various teams:

```
#create data frame
```

```
df <- data.frame(team=c('A', 'A', 'A', 'A', 'A', 'B', 'B', 'B',  
'B', 'B'),  
points=c(12, 29, 34, 14, 10, 11, 7, 36, 34, 22))
```

```
#view data frame
```

```
df
```

```
team points
```

```
1 A 12
```

```
2 A 29
```

```
3 A 34
```

```
4 A 14
```

```
5 A 10
```

```
6 B 11
```

```
7 B 7
```

```
8 B 36
```

```
9 B 34
```

```
10 B 22
```

We can use the following code to create a new column in the data frame that shows the percentage of total points scored, grouped by team:

```
library(dplyr)
```

```
#calculate percentage of points scored, grouped by  
team
```

```
df %>%
```

```
group_by(team) %>%
```

```
mutate(percent = points/sum(points))
```

```
# A tibble: 10 x 3
```

```
# Groups: team
```

```
team points percent
```

```
1 A 12 0.121
```

```
2 A 29 0.293
```

```
3 A 34 0.343
```

```
4 A 14 0.141
```

```
5 A 10 0.101
```

```
6 B 11 0.1
```

```
7 B 7 0.0636
```

```
8 B 36 0.327
```

```
9 B 34 0.309
```

```
10 B 22 0.2
```

The percent column shows the percentage of total points scored by that player within their team.

For example, players on team A scored a total of 99

points.

Thus, the player in the first row of the data frame who scored 12 points scored a total of $12/99 = 12.12\%$ of the total points for team A.

Similarly, the player in the second row of the data frame who scored 29 points scored a total of $29/99 = 29.29\%$ of the total points for team A.

And so on.

Additional Resources

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