How can the sum be calculated by group in R with examples?

Authored by

stats writer

April 30, 2024

Calculating the sum by group in R refers to the process of finding the total value of a specific variable within different groups or categories in a dataset. This can be achieved through the use of the "group_by" function in the dplyr package. This function allows for the creation of groups based on a selected variable, and then the "summarize" function can be used to calculate the sum for each group. For example, if we have a dataset of sales data with variables such as product category and sales amount, we can use the "group_by" function to group the data by product category and then use "summarize" to calculate the total sales for each category. This allows for a better understanding and comparison of the data within different groups.

Calculate the Sum by Group in R (With Examples)

Often you may want to calculate the sum by group in R. There are three methods you can use to do so:

Method 1: Use base R.

```
aggregate(df$col_to_aggregate,
list(df$col_to_group_by), FUN=sum)
```

Method 2: Use the dplyr() package.

```
library(dplyr)

df %>%
group_by(col_to_group_by) %>%
summarise(Freq = sum(col_to_aggregate))
```

Method 3: Use the data.table package.

```
library(data.table)

dt
```

The following examples show how to use each of these methods in practice.

Method 1: Calculate Sum by Group Using Base R

The following code shows how to use the aggregate() function from base R to calculate the sum of the points scored by team in the following data frame:

```
#create data frame
df <- data.frame(team=c('a', 'a', 'b', 'b', 'b', 'c', 'c'),
pts=c(5, 8, 14, 18, 5, 7, 7),
rebs=c(8, 8, 9, 3, 8, 7, 4))

#view data frame
df

team pts rebs
1 a 5 8
2 a 8 8
3 b 14 9
4 b 18 3
```

5 b 5 8

6 c 7 7

7 c 7 4

```
#find sum of points scored by team
aggregate(df$pts, list(df$team), FUN=sum)
```

```
Group.1 x
1 a 13
2 b 37
3 c 14
```

Method 2: Calculate Sum by Group Using dplyr

The following code shows how to use the group_by() and summarise() functions from the dplyr package to calculate the sum of points scored by team in the following data frame:

```
library(dplyr)

#create data frame
df <- data.frame(team=c('a', 'a', 'b', 'b', 'b', 'c', 'c'),
pts=c(5, 8, 14, 18, 5, 7, 7),
rebs=c(8, 8, 9, 3, 8, 7, 4))
```

```
#find sum of points scored by team df %>%
group_by(team) %>%
summarise(Freq = sum(pts))

# A tibble: 3 x 2
team Freq
<chr> <dbl>
1 a 13
2 b 37
3 c 14
```

Method 3: Calculate Sum by Group Using data.table

The following code shows how to use the data.table package to calculate the sum of points scored by team in the following data frame:

```
library(data.table)

#create data frame
df <- data.frame(team=c('a', 'a', 'b', 'b', 'b', 'c', 'c'),
pts=c(5, 8, 14, 18, 5, 7, 7),
rebs=c(8, 8, 9, 3, 8, 7, 4))

#convert data frame to data table
```

setDT(df)

#find sum of points scored by team df

team sum
1: a 13
2: b 37
3: c 14

Note: If you have an extremely large dataset, the data.table method will work the fastest among the three methods listed here.