How to use aggregate() function in R?

Authored by

stats writer

July 2, 2024

The aggregate() function in R is a powerful tool for summarizing and organizing data. It allows users to group data by one or more variables and apply a function to each group, providing a comprehensive overview of the data. This function can be used to calculate various statistics such as mean, median, and sum, as well as perform more complex operations such as creating new variables based on the grouped data. The complete guide for using the aggregate() function in R includes step-by-step instructions and examples for its usage, along with explanations of its parameters and potential applications. This guide aims to help users effectively utilize the aggregate() function to manipulate and analyze their data in an efficient and accurate manner.

The Complete Guide: Use the aggregate() Function in R

The aggregate() function in R can be used to calculate summary statistics for a dataset.

This function uses the following basic syntax:

aggregate(x, by, FUN)

where:

x: A variable to aggregate
by: A list of variables to group by
FUN: The summary statistic to compute

The following examples show how to use this function in practice with the following data frame in R:

#create data frame
df <- data.frame(team=c('A', 'A', 'A', 'B', 'B', 'B'),
position=c('G', 'G', 'F', 'G', 'F', 'F'),

```
points=c(99, 90, 86, 88, 95, 99),
assists=c(33, 28, 31, 39, 34, 23),
rebounds=c(30, 28, 24, 24, 28, 33))

#view data frame
df

team position points assists rebounds
1 A G 99 33 30
2 A G 90 28 28
3 A F 86 31 24
4 B G 88 39 24
5 B F 95 34 28
6 B F 99 23 33
```

Example 1: Aggregate Mean by Group

The following code shows how to use the aggregate() function to calculate the mean number of points scored by team:

```
#find mean points by team
aggregate(df$points, by=list(df$team), FUN=mean)

Group.1 x
1 A 91.66667
```

2 B 94.00000

This tells us:

Players on team A scored an average of 91.67 points per game.
Players on team B scored an average of 94 points per game.

Note that you can also change the names of the columns in the output by using the colnames() function:

```
#find mean points by team
agg <- aggregate(df$points, by=list(df$team), FUN=mean)

#rename columns in output
colnames(agg) <- c('Team', 'Mean_Points')

#view output
agg
```

Team Mean_Points
1 A 91.66667
2 B 94.00000

Example 2: Aggregate Count by Group

The following code shows how to use the aggregate() function to count the number of players by team:

```
#count number of players by team
aggregate(df$points, by=list(df$team), FUN=length)
```

```
Group.1 x
1 A 3
2 B 3
```

Team A has 3 players.
Team B has 3 players.

Example 3: Aggregate Sum by Group

The following code shows how to use the aggregate() function to calculate the sum of points scored by each team:

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

```
Group.1 x
1 A 275
2 B 282
```

This tells us:

Team A scored a total of 275 points.
Team B scored a total of 282 points.

Example 4: Aggregate Multiple Columns

The following code shows how to use the aggregate() function to find the mean number of points scored, grouped by team and position:

```
#find mean of points scored, grouped by team and position
aggregate(df$points, by=list(df$team, df$position), FUN=mean)
```

```
  Group.1 Group.2    x
1       A       F 86.0
2       B       F 97.0
3       A       G 94.5
4       B       G 88.0
```

This tells us:

Players in the 'F' position on Team A scored an average of 86 points.

Players in the 'F' position on Team B scored an average of 97 points.

Players in the 'G' position on Team A scored an average of 94.5 points.

Players in the 'G' position on Team B scored an average of 88 points.

Additional Resources