How can I aggregate multiple columns in R, and can you provide examples?

Authored by

stats writer

April 25, 2024

Aggregating multiple columns in R refers to the process of combining or summarizing data from different columns into a single column. This can be useful for analyzing and understanding patterns or trends in large datasets. One way to aggregate columns in R is by using the aggregate() function, which allows for the creation of new columns based on the values in existing columns. For example, if we have a dataset with columns for sales by month and by product, we can use the aggregate() function to calculate the total sales for each product over the entire year. Other functions such as sum(), mean(), and count() can also be used for aggregation purposes. Overall, aggregating multiple columns in R can help simplify and streamline data analysis and provide valuable insights.

Aggregate Multiple Columns in R (With Examples)

We can use the aggregate() function in R to produce summary statistics for one or more variables in a data frame.

This function uses the following basic syntax:

aggregate(sum_var ~ group_var, data = df, FUN = mean)

where:

sum_var: The variable to summarizegroup_var: The variable to group bydata: The name of the data frameFUN: The summary statistic to compute

This tutorial provides several examples of how to use this function to aggregate one or more columns at once in R, using the following data frame as an example:

```
#create data frame
df <- data.frame(team=c('A', 'A', 'A', 'B', 'B', 'B', 'C', 'C'),
conf=c('E', 'E', 'W', 'W', 'W', 'W', 'W', 'W'),
points=c(1, 3, 3, 4, 5, 7, 7, 9),
rebounds=c(7, 7, 8, 3, 2, 7, 14, 13))

#view data frame
df

team conf points rebounds
1 A E 1 7
2 A E 3 7
3 A W 3 8
4 B W 4 3
5 B W 5 2
6 B W 7 7
7 C W 7 14
8 C W 9 13
```


Example 1: Summarize One Variable & Group by One Variable

The following code shows how to find the mean points scored, grouped by team:

```
#find mean points scored, grouped by team
```

```
aggregate(points ~ team, data = df, FUN = mean, na.rm = TRUE)

  team   points
1    A 2.333333
2    B 5.333333
3    C 8.000000
```

Example 2: Summarize One Variable & Group by Multiple Variables

The following code shows how to find the mean points scored, grouped by team and conference:

```
#find mean points scored, grouped by team and conference
aggregate(points ~ team + conf, data = df, FUN = mean, na.rm = TRUE)

  team conf   points
1    A    E 2.000000
2    A    W 3.000000
3    B    W 5.333333
4    C    W 8.000000
```

Example 3: Summarize Multiple Variables & Group by One Variable

The following code shows how to find the mean points and the mean rebounds, grouped by team:

```
#find mean points scored, grouped by team and conference
aggregate(cbind(points,rebounds) ~ team, data = df, FUN = mean, na.rm = TRUE)
```

```
team points rebounds
1 A 2.333333 7.333333
2 B 5.333333 4.000000
3 C 8.000000 13.500000
```

Example 4: Summarize Multiple Variables & Group by Multiple Variables

```
#find mean points scored, grouped by team and conference
aggregate(cbind(points,rebounds) ~ team + conf, data = df, FUN = mean, na.rm = TRUE)
```

```
team conf points rebounds
1 A E 2.000000 7.0
2 A W 3.000000 8.0
3 B W 5.333333 4.0
4 C W 8.000000 13.5
```