How can I use the dplyr package in R to filter a dataset by multiple conditions simultaneously?

Authored by

stats writer

July 2, 2024

The dplyr package in R is a useful tool for data manipulation and filtering. It allows users to filter a dataset based on multiple conditions simultaneously, making data analysis more efficient and effective. By using various functions such as filter(), select(), and mutate(), users can easily specify their desired conditions and extract specific subsets of data from a larger dataset. This process can be repeated multiple times, allowing for the creation of complex and precise filters. Additionally, the dplyr package offers a streamlined and user-friendly syntax, making it accessible for both beginners and advanced users. Overall, utilizing the dplyr package in R for filtering data by multiple conditions allows for more efficient and accurate data analysis.

Filter by Multiple Conditions Using dplyr

You can use the following syntax to filter data frames by multiple conditions using the library:

Method 1: Filter by Multiple Conditions Using OR

```
library(dplyr)

df %>%
filter(col1 == 'A' | col2 > 90)
```

Method 2: Filter by Multiple Conditions Using AND

```
library(dplyr)

df %>%
filter(col1 == 'A' & col2 > 90)
```

The following example shows how to use these

methods in practice with the following data frame in R:

```
#create data frame
df <- data.frame(team=c('A', 'A', 'B', 'B', 'C'),
points=c(99, 90, 86, 88, 95),
assists=c(33, 28, 31, 39, 34),
rebounds=c(30, 28, 24, 24, 28))

#view data frame
df

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

Method 1: Filter by Multiple Conditions Using OR

The following code shows how to use the or (|) operator to filter the data frame by rows that meet one of multiple conditions:

```
library(dplyr)
```

```
#filter for rows where team is equal to 'A' or points is greater than 90
df %>%
filter(team == 'A' | points > 90)

team points assists rebounds
1 A 99 33 30
2 A 90 28 28
3 C 95 34 28
```

The only rows returned are those where the team is equal to 'A' or where points is greater than 90.

Note that we can use as many "or" operators as we'd like in the filter function:

```
library(dplyr)

#filter for rows where team is equal to 'A' or 'C' or points is less than 89
df %>%
filter(team == 'A' | team == 'C' | points > 90)

team points assists rebounds
1 A 99 33 30
```

2 A 90 28 28

3 B 86 31 24

4 C 95 34 28

Method 2: Filter by Multiple Conditions Using AND

The following code shows how to use the and (&) operator to filter the data frame by rows that meet several conditions:

```
library(dplyr)

#filter for rows where team is equal to 'A' and points is greater than 90
df %>%
filter(team == 'A' & points > 90)
```

```
team points assists rebounds
1 A 99 33 30
```

Only one row met both conditions in the filter function.

```
library(dplyr)

#filter where team is equal to 'A' and points > 89 and assists < 30
```

```
df %>%
filter(team == 'A' & points > 89 & assists < 30)
```

```
team points assists rebounds
1 A 90 28 28
```

Note: You can find the complete documentation for the dplyr filter() function .

Additional Resources

The following tutorials explain how to perform other common operations in dplyr: