

How can I filter a dataset in R using dplyr to only include unique values?

Authored by
stats writer

June 29, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I filter a dataset in R using dplyr to only include unique values?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=159861>

To filter a dataset in R using dplyr to only include unique values, one can use the `distinct()` function. This function removes all duplicate rows from the dataset, leaving only the unique values. It can be applied to specific columns or the entire dataset. This method is useful for data cleaning and analysis, as it allows for a more accurate representation of the data without redundant information. By using dplyr and the `distinct()` function, one can easily filter and extract unique values from a dataset in R.

Filter for Unique Values Using dplyr

You can use the following methods to filter for unique values in a data frame in R using the package:

Method 1: Filter for Unique Values in One Column

```
df %>% distinct(var1)
```

Method 2: Filter for Unique Values in Multiple Columns

```
df %>% distinct(var1, var2)
```

Method 3: Filter for Unique Values in All Columns

```
df %>% distinct()
```

The following examples show how to use each method in practice with the following data frame in R:

```
#create data frame
```

```
df <- data.frame(team=c('A', 'A', 'A', 'A', 'B', 'B', 'B', 'B'),  
points=c(10, 10, 8, 6, 15, 15, 12, 12),  
rebounds=c(8, 8, 4, 3, 10, 11, 7, 7))
```

```
#view data frame
```

```
df
```

```
team points rebounds
```

```
1 A 10 8
```

```
2 A 10 8
```

```
3 A 8 4
```

```
4 A 6 3
```

```
5 B 15 10
```

```
6 B 15 11
```

```
7 B 12 7
```

```
8 B 12 7
```

Example 1: Filter for Unique Values in Column

We can use the following code to filter for unique values in just the team column:

```
library(dplyr)
```

```
#select only unique values in team column
```

```
df %>% distinct(team)
```

```
team
```

```
1 A
```

```
2 B
```

Notice that only the unique values in the team column are returned.

Example 2: Filter for Unique Values in Multiple Columns

We can use the following code to filter for unique values in the team and points columns:

```
library(dplyr)
```

```
#select unique values in team and points columns
```

```
df %>% distinct(team, points)
```

```
team points
```

```
1 A 10
```

```
2 A 8
```

```
3 A 6
```

```
4 B 15
```

```
5 B 12
```

Notice that only the unique values in the team and points columns are returned.

Example 3: Filter for Unique Values in All Columns

We can use the following code to filter for unique values across all columns in the data frame:

```
library(dplyr)
```

```
#select unique values across all columns  
df %>% distinct()
```

```
team points rebounds
```

```
1 A 10 8
```

```
2 A 8 4
```

```
3 A 6 3
```

```
4 B 15 10
```

```
5 B 15 11
```

```
6 B 12 7
```

Notice that the unique values across all three columns are returned.

Note: You can find the complete documentation for the distinct function in dplyr .

Additional Resources

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

ARABPSYCHOLOGY.COM