

How can I subset a data frame in R based on factor levels?

Authored by
stats writer

June 26, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I subset a data frame in R based on factor levels?*.

PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=153601>

Subsetting a data frame in R based on factor levels refers to selecting specific rows or columns from a data frame based on the levels of a categorical variable, known as a factor. This can be accomplished using the `subset()` function, which allows for filtering the data frame by specifying the factor level(s) of interest. This method is useful for organizing and analyzing data sets with categorical variables and allows for efficient data manipulation and analysis. It is a common technique used in statistical analysis and data science projects.

Subset Data Frame by Factor Levels in R

You can use one of the following methods to subset a data frame by factor levels in R:

Method 1: Subset by One Factor Level

```
#subset rows where team is equal to 'B'df_sub <- df
```

Method 2: Subset by Multiple Factor Levels

```
#subset rows where team is equal to 'A' or 'C'  
df_sub <- df
```

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

```
#create data frame  
df <- data.frame(team=factor(c('A', 'A', 'B', 'B', 'B', 'C')),  
points=c(22, 35, 19, 15, 29, 23))
```

```
#view data frame
```

```
df
```

```
team points
```

```
1 A 22
```

```
2 A 35
```

```
3 B 19
```

```
4 B 15
```

```
5 B 29
```

```
6 C 23
```

```
Method 1: Subset by One Factor Level
```

The following code shows how to create a new data frame that subsets by the rows where the value in the team column is equal to 'B':

```
#subset rows where team is equal to 'B'
```

```
df_sub <- df
```

```
#view updated data frame
```

```
df_sub
```

```
team points
```

```
3 B 19
```

```
4 B 15
```

5 B 29

Notice that the new data frame only contains rows where the value in the team column is equal to 'B'.

Example 2: Subset by Multiple Factor Levels

The following code shows how to create a new data frame that subsets by the rows where the value in the team column is equal to 'A' or 'C':

```
#subset rows where team is equal to 'A' or 'C'
```

```
df_sub <- df
```

```
#view updated data frame
```

```
df_sub
```

```
team points
```

```
1 A 22
```

```
2 A 35
```

```
6 C 23
```

Notice that the new data frame only contains rows where the value in the team column is equal to 'A' or 'C'.

Using this syntax, you can include as many factor

levels as you'd like in the vector following the `%in%` operator to subset by even more factor levels.

Related:

ARABPSYCHOLOGY.COM