How can I subset a data frame in R based on a list of values?

Authored by

stats writer

June 26, 2024

Subsetting a data frame in R refers to selecting specific rows or columns from a larger data frame based on certain criteria. One way to subset a data frame is by using a list of values. This involves creating a list of values that correspond to the rows or columns that you want to extract from the data frame. This list can then be used as a filter to select only the desired rows or columns from the data frame. By using this method, you can easily customize your data frame to only include the information that is relevant to your analysis. This can be useful for organizing and manipulating large datasets in R.

Subset Data Frame by List of Values in R

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

Method 1: Use Base R

df_new <- df

Method 2: Use dplyr

library(dplyr)

df_new <- filter(df, my_column %in% vals)

Method 3: Use data.table

library(data.table)

df_new <- setDT(df, key='my_column')

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=c('A', 'B', 'B', 'B', 'C', 'C', 'C', 'D'),
points=c(12, 22, 35, 34, 20, 28, 30, 18),
assists=c(4, 10, 11, 12, 12, 8, 6, 10))

#view data frame
df

team points assists
1 A 12 4
2 B 22 10
3 B 35 11
4 B 34 12
5 C 20 12
6 C 28 8
7 C 30 6
8 D 18 10
```

Method 1: Subset Data Frame by List of Values in Base R

The following code shows how to subset the data frame to only contain rows that have a value of 'A' or 'C' in the

team column:

#define values to subset by
vals <- c('A', 'C')

#subset data frame to only contain rows where team is 'A' or 'C'
df_new <- df

#view results
df_new

team points assists
1 A 12 4
5 C 20 12
6 C 28 8
7 C 30 6

The resulting data frame only contains rows that have a value of 'A' or 'C' in the team column.

Note that we used functions from base R in this example so we didn't have to load any extra packages.

Method 2: Subset Data Frame by List of Values in dplyr

The following code shows how to subset the data frame to only contain rows that have a value of 'A' or 'C' in the team column by using the filter() function from the dplyr package:

```
library(dplyr)

#define values to subset by
vals <- c('A', 'C')

#subset data frame to only contain rows where team is 'A' or 'C'
df_new <- filter(df, team %in% vals)

#view results
df_new

team points assists
1 A 12 4
5 C 20 12
6 C 28 8
7 C 30 6
```

Method 3: Subset Data Frame by List of Values in data.table

The following code shows how to subset the data frame

to only contain rows that have a value of 'A' or 'C' in the team column by using functions from the data.table package:

```
library(data.table)

#define values to subset by
vals <- c('A', 'C')

#subset data frame to only contain rows where team is 'A' or 'C'
df_new <- setDT(df, key='team')

#view results
df_new

team points assists
1: A 12 4
2: C 20 12
3: C 28 8
4: C 30 6
```

The resulting data frame only contains rows that have a value of 'A' or 'C' in the team column.

Related: