How can I subset a data frame in R?

Authored by

stats writer

May 4, 2024

ARABPSYCHOLOGY.COM

RECOMMENDED CITATION

stats writer (2024). *How can I subset a data frame in R?*. PSYCHOLOGICAL SCALES. Retrieved from https://scales.arabpsychology.com/?p=142566

To subset a data frame in R, you can use the square bracket notation or the subset() function. The square bracket notation allows you to select specific columns or rows based on their position or names. The subset() function takes in a logical expression to filter the data frame. Both methods allow for a subset of the original data frame to be created, making it easier to work with smaller chunks of data.

Subset a Data Frame in R (4 Examples)

You can use the following basic syntax to subset a data frame in R:

df

The following examples show how to use this syntax in practice with the following data frame:

#create data frame
df <- data.frame(team=c('A', 'A', 'B', 'B', 'C', 'C', 'C'),
points=c(77, 81, 89, 83, 99, 92, 97),
assists=c(19, 22, 29, 15, 32, 39, 14))

#view data frame
df

team points assists
1 A 77 19
2 A 81 22

ARABPSYCHOLOGY.COM

3 B 89 29

4 B 83 15

5 C 99 32

6 C 92 39

7 C 97 14

Example 1: Subset Data Frame by Selecting Columns

The following code shows how to subset a data frame by column names:

#select all rows for columns 'team' and 'assists'

df

team assists

1 A 19

2 A 22

3 B 29

4 B 15

5 C 32

6 C 39

7 C 14

We can also subset a data frame by column index values:

ARABPSYCHOLOGY.COM

#select all rows for columns 1 and 3
df

team assists
1 A 19
2 A 22
3 B 29
4 B 15
5 C 32
6 C 39
7 C 14

Example 2: Subset Data Frame by Excluding Columns

The following code shows how to subset a data frame by excluding specific column names:

#define columns to exclude
cols <- names(df) %in% c('points')#exclude points column
df
team assists
1 A 19
2 A 22
3 B 29

ARABPSYCHOLOGY.COM

4 B 15

5 C 32

6 C 39

7 C 14

We can also exclude columns using index values

#exclude column 2

df

team assists

1 A 19

2 A 22

3 B 29

4 B 15

5 C 32

6 C 39

7 C 14

ARABPSYCHOLOGY.COM

Example 3: Subset Data Frame by Selecting Rows

The following code shows how to subset a data frame by specific rows:

#select rows 1, 5, and 7df

team points assists
1 A 77 19
5 C 99 32
7 C 97 14

#select rows 1 through 5df

team points assists
1 A 77 19
2 A 81 22
3 B 89 29
4 B 83 15
5 C 99 32

Example 4: Subset Data Frame Based on Conditions

The following code shows how to use the subset() function to select rows and columns that meet certain conditions:

#select rows where points is greater than 90
subset(df, points > 90)

team points assists
5 C 99 32

ARABPSYCHOLOGY.COM

6 C 92 39
7 C 97 14

We can also use the | ("or") operator to select rows that meet one of several conditions:

#select rows where points is greater than 90 or less than 80
subset(df, points > 90 | points < 80)

team points assists
1 A 77 19
5 C 99 32
6 C 92 39
7 C 97 14

We can also use the & ("and") operator to select rows that meet multiple conditions:

#select rows where points is greater than 90 *and* assists is greater than 30
subset(df, points > 90 & assists > 30)

team points assists
5 C 99 32

ARABPSYCHOLOGY.COM

6 C 92 39

We can also use the select argument to only select certain columns based on a condition:

#select rows where points is greater than 90 and only show 'team' column
subset(df, points > 90, select=c('team'))

team
5 C
6 C
7 C

ARABPSYCHOLOGY.COM