

How can I select columns in R that contain a specific string?

Authored by
stats writer

June 27, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I select columns in R that contain a specific string?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=155066>

The method of selecting columns in R that contain a specific string involves using the "grep" function, which allows for pattern matching within a dataset. This function can be used to search for a desired string within the column names of the dataset and return the columns that contain that string. By specifying the string and the dataset to be searched, the "grep" function can effectively filter and select the desired columns. This method is particularly useful for large datasets with multiple columns, as it allows for a more efficient and accurate way of selecting specific columns based on a given criteria.

Select Columns Containing a Specific String in R

You can use the following functions from the package in R to select columns that contain a specific string:

Method 1: Select Columns that Contain One Specific String

```
df %>%  
select(matches("string1"))
```

Method 2: Select Columns that Contain One of Several Strings

```
df %>%  
select(matches("string1|string2|string3"))
```

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(mavs=c(12, 10, 14, 19, 22, 25, 29),  
cavs=c(22, 41, 14, 15, 15, 19, 22),  
hornets=c(8, 8, 12, 14, 15, 13, 12),  
spurs=c(10, 12, 12, 16, 22, 28, 30),  
nets=c(9, 7, 10, 22, 28, 23, 25))
```

#view data frame

```
df
```

```
mavs cavs hornets spurs nets
```

```
1 12 22 8 10 9
```

```
2 10 41 8 12 7
```

```
3 14 14 12 12 10
```

```
4 19 15 14 16 22
```

```
5 22 15 15 22 28
```

```
6 25 19 13 28 23
```

```
7 29 22 12 30 25
```

Example 1: Select Columns that Contain One Specific String

The following code shows how to use the `matches()` function to select only the columns that contain the string "avs" somewhere in their name:

```
library(dplyr)
```

```
#select all columns that contain "avs" in the name
```

```
df %>%
```

```
select(matches("avs"))
```

```
mavs cavs
```

```
1 12 22
```

```
2 10 41
```

```
3 14 14
```

```
4 19 15
```

```
5 22 15
```

```
6 25 19
```

```
7 29 22
```

Only the columns that contain "avs" in the name are returned.

In this case, "mavs" and "cavs" are the only columns that are returned.

Example 2: Select Columns that Contain One of Several Strings

The following code shows how to use the `matches()` function to select only the columns that contain "avs" or "ets" somewhere in their name:

```
library(dplyr)
```

```
#select all columns that contain "avs" or "ets" in the  
name
```

```
df %>%
```

```
select(matches("avs|ets"))
```

```
mavs cavs hornets nets
```

```
1 12 22 8 9
```

```
2 10 41 8 7
```

```
3 14 14 12 10
```

```
4 19 15 14 22
```

```
5 22 15 15 28
```

```
6 25 19 13 23
```

```
7 29 22 12 25
```

Only the columns that contain "avs" or "ets" in the name are returned.

Note that the vertical bar (|) is the "OR" operator in R.

The following tutorials explain how to perform other common tasks using dplyr: