

How can I drop rows from a dataset that contain a specific string using R?

Authored by
stats writer

April 28, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I drop rows from a dataset that contain a specific string using R?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=140679>

The process of dropping rows from a dataset that contain a specific string using R involves utilizing the "grep" function to search for the specific string within the dataset and then using the "subset" function to remove the rows that contain the string. This can be done by specifying the condition for which rows to be removed within the "subset" function. This method allows for efficient and precise removal of rows containing the specific string from the dataset.

R: Drop Rows that Contain a Specific String

You can use the following syntax to drop rows that contain a certain string in a data frame in R:

```
df
```

This tutorial provides several examples of how to use this syntax in practice with the following data frame in R:

```
#create data frame
```

```
df <- data.frame(team=c('A', 'A', 'A', 'B', 'B', 'C'),  
conference=c('East', 'East', 'East', 'West', 'West', 'East'),  
points=c(11, 8, 10, 6, 6, 5))
```

```
#view data frame
```

```
df
```

```
team conference points
```

```
1 A East 11
```

2 A East 8

3 A East 10

4 B West 6

5 B West 6

6 C East 5

Example 1: Drop Rows that Contain a Specific String

The following code shows how to drop all rows in the data frame that contain 'A' in the team column:

```
df
```

```
team conference points
```

```
4 B West 6
```

```
5 B West 6
```

```
6 C East 5
```

Or we could drop all rows in the data frame that contain 'West' in the conference column:

```
df
```

```
team conference points
```

```
1 A East 11
```

```
2 A East 8
```

3 A East 10

6 C East 5

Example 2: Drop Rows that Contain a String in a List

The following code shows how to drop all rows in the data frame that contain 'A' or 'B' in the team column:

```
df
```

6 C East 5

We could also define a vector of strings and then remove all rows in the data frame that contain any of the strings in the vector in the team column:

```
#define vector of strings  
remove <- c('A', 'B')
```

```
#remove rows that contain any string in the vector in  
the team column
```

```
df
```

6 C East 5

Notice that both methods lead to the same result.