

How can I use the “not in” filter in dplyr to exclude certain values from my data?

Authored by
stats writer

May 6, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I use the “not in” filter in dplyr to exclude certain values from my data?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=143346>

The "not in" filter in dplyr is a useful tool for excluding specific values from a dataset. This filter allows users to specify a list of values that they do not want to include in their data, effectively creating a subset of the original dataset. This feature is particularly helpful when working with large datasets, as it allows for quick and efficient removal of unwanted values. To use the "not in" filter, users simply need to specify the values they want to exclude in the filter function, providing a convenient and intuitive way to clean and manipulate data.

dplyr: Use a "not in" Filter

You can use the following basic syntax in to filter for rows in a data frame that are not in a list of values:

```
df %>%
```

```
filter(!col_name %in% c('value1', 'value2', 'value3', ...))
```

The following examples show how to use this syntax in practice.

Example 1: Filter for Rows that Do Not Contain Value in One Column

Suppose we have the following data frame in R:

```
#create data frame
```

```
df <- data.frame(team=c('A', 'A', 'B', 'B', 'C', 'C', 'D', 'D'),  
position=c('G', 'G', 'F', 'G', 'F', 'C', 'C', 'C'),  
points=c(12, 14, 19, 24, 36, 41, 18, 29))
```

```
#view data frame
```

df

team position points

1 A G 12

2 A G 14

3 B F 19

4 B G 24

5 C F 36

6 C C 41

7 D C 18

8 D C 29

The following syntax shows how to filter for rows where the team name is not equal to 'A' or 'B':

#filter for rows where team name is not 'A' or 'B'

df %>% filter(!team %in% c('A', 'B'))

team position points

1 C F 36

2 C C 41

3 D C 18

4 D C 29

Example 2: Filter for Rows that Do Not Contain Value in Multiple Columns

Suppose we have the following data frame in R:

#create data frame

```
df <- data.frame(team=c('A', 'A', 'B', 'B', 'C', 'C', 'D', 'D'),  
position=c('G', 'G', 'F', 'G', 'F', 'C', 'C', 'C'),  
points=c(12, 14, 19, 24, 36, 41, 18, 29))
```

#view data frame

df

team position points

1 A G 12

2 A G 14

3 B F 19

4 B G 24

5 C F 36

6 C C 41

7 D C 18

8 D C 29

The following syntax shows how to filter for rows where the team name is not equal to 'A' *and* where the position is not equal to 'C':

#filter for rows where team name is not 'A' and position

is not 'C'

```
df %>% filter(!team %in% c('A') & !position %in% c('C'))
```

team position points

1 B F 19

2 B G 24

3 C F 36

The following tutorials explain how to perform other common functions in dplyr: