

How can I select columns in dplyr that do not start with a specific string?

Authored by
stats writer

June 27, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I select columns in dplyr that do not start with a specific string?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=155068>

The process of selecting columns in dplyr that do not begin with a specific string involves using the "select" function and the "starts_with" argument to specify the columns to be excluded. This method allows for efficient and precise selection of columns based on their names, providing a versatile and powerful tool for data manipulation. By using this approach, users can easily filter out unwanted columns and focus on the specific data they need for analysis or visualization.

Select Columns that Do Not Start with String in dplyr

You can use the following functions from the package in R to select columns that do not start with a specific string:

Method 1: Select Columns that Do Not Start with One Specific String

```
df %>%  
select(-starts_with("string1"))
```

Method 2: Select Columns that Do Not Start with One of Several Strings

```
df %>%  
select(-starts_with(c("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(store1_sales=c(12, 10, 14, 19, 22, 25, 29),
store1_returns=c(3, 3, 2, 4, 3, 2, 1),
store2_sales=c(8, 8, 12, 14, 15, 13, 12),
store2_returns=c(1, 2, 2, 1, 2, 1, 3),
promotions=c(0, 1, 1, 1, 0, 0, 1))
```

#view data frame

```
df
```

```
store1_sales    store1_returns    store2_sales
store2_returns promotions
1 12 3 8 1 0
2 10 3 8 2 1
3 14 2 12 2 1
4 19 4 14 1 1
5 22 3 15 2 0
6 25 2 13 1 0
7 29 1 12 3 1
```

Example 1: Select Columns that Do Not Start with One Specific String

The following code shows how to use the `-starts_with()`

function to select only the columns that do not start with "store1" in the data frame:

```
library(dplyr)
```

```
#select all columns that do not start with "store1"
```

```
df %>%
```

```
select(-starts_with("store1"))
```

```
store2_sales store2_returns promotions
```

```
1 8 1 0
```

```
2 8 2 1
```

```
3 12 2 1
```

```
4 14 1 1
```

```
5 15 2 0
```

```
6 13 1 0
```

```
7 12 3 1
```

Notice that the two columns that start with "store1" are not returned.

Example 2: Select Columns that Do Not Start with One of Several Strings

The following code shows how to use the `-starts_with()` function to select only the columns that do not start

with "store1" or "prom" in the data frame:

```
library(dplyr)
```

```
#select all columns that do not start with "store1" or  
"prom"
```

```
df %>%
```

```
select(-starts_with(c("store1", "prom")))
```

```
store2_sales store2_returns
```

```
1 8 1
```

```
2 8 2
```

```
3 12 2
```

```
4 14 1
```

```
5 15 2
```

```
6 13 1
```

```
7 12 3
```

Notice that any columns that start with "store1" or "prom" are not returned.

Note: By default, the `starts_with()` function is case-insensitive. To make the function case-sensitive, use the `ignore.case=FALSE` argument within the function.