What is the difference between grep() and grepl() in R?

Authored by

stats writer

April 18, 2024

Grep() and grepl() are two useful functions in R that are used for pattern matching in strings. While both are used for the same purpose, there are some key differences between them. Grep() searches for a specific pattern in a given string and returns the matching elements, while grepl() returns a logical vector indicating which elements in the string match the given pattern. Additionally, grep() can be used to extract the matched elements from the string, while grepl() is primarily used for testing if a pattern exists in the string. Overall, the main difference between grep() and grepl() is that grep() is used for pattern extraction, while grepl() is used for pattern testing.

Comparing grep() vs. grepl() in R: What's the Difference?

Two functions that people often get mixed up in R are grep() and grepl(). Both functions allow you to see whether a certain pattern exists in a character string, but they return different results:

grepl() returns TRUE when a pattern exists in a character string.grep() returns a vector of indices of the character strings that contain the pattern.

The following example illustrates this difference:

```
#create a vector of data
data <- c('P Guard', 'S Guard', 'S Forward', 'P Forward', 'Center')

grep('Guard', data)
```

1 2

grepl('Guard', data)
TRUE TRUE FALSE FALSE FALSE

The following examples show when you might want to use one of these functions over the other.

When to Use grepl()

1. Filter Rows that Contain a Certain String

One of the most common uses of grepl() is for filtering rows in a data frame that contain a certain string:

```
library(dplyr)

#create data frame
df <- data.frame(player = c('P Guard', 'S Guard', 'S Forward', 'P Forward', 'Center'),
points = c(12, 15, 19, 22, 32),
rebounds = c(5, 7, 7, 12, 11))

#filter rows that contain the string 'Guard' in the player column
df %>% filter(grepl('Guard', player))
```

player points rebounds
1 P Guard 12 5
2 S Guard 15 7

How to Filter Rows that Contain a Certain String Using dplyr

When to Use grep()

1. Select Columns that Contain a Certain String

You can use grep() to select columns in a data frame that contain a certain string:

```
library(dplyr)

#create data frame
df <- data.frame(player = c('P Guard', 'S Guard', 'S Forward', 'P Forward', 'Center'),
points = c(12, 15, 19, 22, 32),
rebounds = c(5, 7, 7, 12, 11))

#select columns that contain the string 'p' in their name
df %>% select(grep('p', colnames(df)))

player points
```

1 P Guard 12

2 S Guard 15

3 S Forward 19

4 P Forward 22

5 Center 32

2. Count the Number of Rows that Contain a Certain String

You can use grep() to count the number of rows in a data frame that contain a certain string:

```
#create data frame
df <- data.frame(player = c('P Guard', 'S Guard', 'S Forward', 'P Forward', 'Center'),
points = c(12, 15, 19, 22, 32),
rebounds = c(5, 7, 7, 12, 11))

#count how many rows contain the string 'Guard' in the player column
length(grep('Guard', df$player))
```

2