How can numbers be extracted from strings in R? Can you provide examples?

Authored by

stats writer

June 26, 2024

In R, numbers can be extracted from strings by using various methods such as regular expressions, string manipulation functions, and conversion functions. Regular expressions allow for specific patterns to be defined and searched for within a string. String manipulation functions, such as str_extract() or str_extract_all(), can be used to extract numbers based on certain criteria or conditions. Additionally, conversion functions like as.numeric() can be used to convert a string into a numeric value. For example, the string "I have 3 apples" can be extracted to only include the number "3" by using regular expressions or string manipulation functions. Similarly, the string "5.6 miles" can be converted to the numeric value 5.6 using the as.numeric() function.

Extract Numbers from Strings in R (With Examples)

You can use the following methods to extract numbers from strings in R:

Method 1: Extract Number from String Using Base R

as.numeric(gsub("D", "", df$my_column))

Method 2: Extract Number from String Using readr Package

library(readr)

parse_number(df$my_column)

This tutorial explains how to use each method in practice with the following data frame:

#create data frame

```
df <- data.frame(team=c('A', 'A', 'A', 'B', 'B', 'B'),
position=c('Guard23', 'Guard14', '2Forward',
'Guard25', '6Forward', 'Center99'))

#view data frame
df

team position
1 A Guard23
2 A Guard14
3 A 2Forward
4 B Guard25
5 B 6Forward
6 B Center99
```

Example 1: Extract Number from String Using Base R

The following code shows how to extract the numbers from each string in the position column of the data frame:

```
#extract number from each string in 'position' column
as.numeric(gsub("D", "", df$position))

23 14 2 25 6 99
```

Notice that the numeric values have been extracted from each string in the position column.

Note: The gsub() function simply replaces all non-numbers (D) in a string with a blank space. This has the effect of extracting only the numbers from the string.

If you'd like, you can also store these numeric values in a new column in the data frame:

```
#create new column that contains numbers from each string in 'position' column
df$num <- as.numeric(gsub("D", "", df$position))

#view updated data frame
df

team position num
1 A Guard23 23
2 A Guard14 14
3 A 2Forward 2
4 B Guard25 25
5 B 6Forward 6
6 B Center99 99
```

Example 2: Extract Number from String Using reader Package

The following code shows how to extract the numbers from each string in the position column of the data frame by using the parse_number() function from the package:

```
library(readr)

#extract number from each string in 'position' column
parse_number(df$position)
```

```
23 14 2 25 6 99
```

This matches the results from using the gsub() function in base R.