

How do I extract the string after a specific character in R?

Authored by
stats writer

June 25, 2024

RECOMMENDED CITATION

stats writer (2024). *How do I extract the string after a specific character in R?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=151856>

The process of extracting a specific string after a designated character in R involves using the built-in functions and operators in the language. This can be achieved by using the "strsplit" function to split the original string into multiple substrings, and then selecting the desired substring using indexing or regular expressions. Other methods such as the "substring" function or the "gsub" function can also be utilized for this purpose. This process allows for the isolation and extraction of a specific part of a string, based on the location of a designated character within the original string.

Extract String After Specific Character in R

You can use the following methods to extract a string after a specific character in R:

Method 1: Extract String After Specific Characters Using Base R

```
sub('.*the', "", my_string)
```

Method 2: Extract String After Specific Characters Using stringr

```
library(stringr)
```

```
str_replace(my_string, '(.*?)the(.*?)', '1')
```

Both of these examples extract the string after the pattern "the" within my_string.

The following examples show how to use each method in practice with the following data frame:

```
#create data frame
```

```
df <- data.frame(team=c('theMavs', 'theHeat', 'theNets',  
'theRockets'),  
points=c(114, 135, 119, 140))
```

```
#view data frame
```

```
df
```

```
team points
```

```
1 theMavs 114
```

```
2 theHeat 135
```

```
3 theNets 119
```

```
4 theRockets 140
```

Example 1: Extract String After Specific Characters Using Base R

The following code shows how to extract the string after "the" for each row in the team column of the data frame:

```
#create new column that extracts string after "the" in  
team column
```

```
df$team_name <- sub('.*the', "", df$team)
```

```
#view updated data frame
```

```
df
```

```
team points team_name
```

```
1 theMavs 114 Mavs
```

```
2 theHeat 135 Heat
```

```
3 theNets 119 Nets
```

```
4 theRockets 140 Rockets
```

Notice that the new column called `team_name` contains the string after "the" for each row in the `team` column of the data frame.

Related:

Example 2: Extract String After Specific Characters Using `stringr` Package

The following code shows how to extract the string after "the" for each row in the `team` column of the data frame by using the `str_replace()` function from the `stringr` package in R:

```
library(stringr)
```

#create new column that extracts string after "the" in team column

```
df$team_name <- str_replace(df$team, '(.*?)the(.*?)', '1')
```

#view updated data frame

```
df
```

```
team points team_name
```

```
1 team Mavs pro 114 Mavs
```

```
2 team Heat pro 135 Heat
```

```
3 team Nets pro 119 Nets
```

Notice that the new column called team_name contains the string after "the" for each row in the team column of the data frame.