

# How can I extract a string between specific characters in R?

Authored by  
**stats writer**

June 25, 2024

## RECOMMENDED CITATION

stats writer (2024). *How can I extract a string between specific characters in R?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=151854>

To extract a string between specific characters in R, you can use the "str\_extract()" function from the "stringr" package. This function allows you to specify the starting and ending characters, and will return the string that falls between them. This can be useful for extracting specific information from a larger string, such as extracting names or numbers from a larger text. Additionally, you can also use regular expressions with the "str\_extract()" function to extract more complex patterns. Overall, using the "str\_extract()" function in R can efficiently retrieve targeted strings from a larger dataset.

## Extract String Between Specific Characters in R

You can use the following methods to extract a string between specific characters in R:

### Method 1: Extract String Between Specific Characters Using Base R

```
gsub(".*char1 (.+) char2.*", "1", my_string)
```

### Method 2: Extract String Between Specific Characters Using stringr

```
library(stringr)
```

```
str_match(my_string, "char1s*(.*?)s*char2")
```

Both of these examples extract the string between the characters char1 and char2 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('team Mavs pro', 'team Heat  
pro', 'team Nets pro'),  
points=c(114, 135, 119))
```

```
#view data frame
```

```
df
```

```
team points
```

```
1 team Mavs pro 114
```

```
2 team Heat pro 135
```

```
3 team Nets pro 119
```

**Example 1: Extract String Between Specific Characters Using Base R**

The following code shows how to extract the string between the characters *team* and *pro* for each row in the team column of the data frame:

```
#create new column that extracts string between team  
and pro
```

```
df$team_name <- gsub(".*team (.+) pro.*", "1", df$team)
```

```
#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 between the characters *team* and *pro* for each row in the `team` column of the data frame.

**Related:**

**Example 2: Extract String Between Specific Characters Using stringr Package**

The following code shows how to extract the string between the characters *team* and *pro* for each row in the `team` column of the data frame by using the `str_match()` function from the `stringr` package in R:

```
library(stringr)
```

**#create new column that extracts string between team and pro**

```
df$team_name <- str_match(df$team,  
"teams*(.*?)s*pro")
```

**#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 between the characters *team* and *pro* for each row in the team column of the data frame.**

**Note that the str\_match() function returns a matrix in which the first column contains the original strings and the second column contains the substring we're interested in.**

**Thus, we must use to extract only the second column from the matrix returned by the str\_match() function.**