

How can the substring function be used in R?

Authored by
stats writer

June 29, 2024

RECOMMENDED CITATION

stats writer (2024). *How can the substring function be used in R?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=158355>

The substring function in R is used to extract a specific portion of a string based on the given starting and ending positions. This function can be used to manipulate strings by selecting a specific subset of characters, which can be helpful in data cleaning and analysis. The syntax for the substring function is "substring(x, first, last)", where x is the string, first is the starting position, and last is the ending position. Additionally, the substring function can also be used with other string manipulation functions to perform more complex tasks such as replacing characters or concatenating strings. Overall, the substring function is a useful tool for working with strings in R and can aid in various data processing tasks.

Use substring Function in R (4 Examples)

The substring() function in R can be used to extract a substring in a character vector.

This function uses the following syntax:

substring(text, first, last)

where:

text: Name of the character vector
first: The first element to be extracted
last: The last element to be extracted

Also note that the substr() function does the exact same thing, but with slightly different argument names:

substr(text, first, last)

where:

x: Name of the character vector
start: The first element to be extracted
stop: The last element to be extracted

The examples in this tutorial show how to use the `substring()` function in practice with the following data frame in R:

```
#create data frame
df <- data.frame(team=c('Mavericks', 'Hornets',
                        'Rockets', 'Grizzlies'))
```

```
#view data frame
```

```
df
```

```
team
```

```
1 Mavericks
```

```
2 Hornets
```

```
3 Rockets
```

```
4 Grizzlies
```

Example 1: Extract Characters Between Certain Positions

The following code shows how to use the `substring()` function to extract the characters between positions 2 and 5 of the "team" column:

```
#create new column that contains characters between  
positions 2 and 5
```

```
df$between2_5 <- substring(df$team, first=2, last=5)
```

```
#view updated data frame
```

```
df
```

```
team between2_5
```

```
1 Mavericks aver
```

```
2 Hornets orne
```

```
3 Rockets ocke
```

```
4 Grizzlies rizz
```

Notice that the new column contains the characters between positions 2 and 5 of the "team" column.

Example 2: Extract First N Characters

The following code shows how to use the substring() function to extract the first 3 characters of the "team" column:

```
#create new column that contains first 3 characters
```

```
df$first3 <- substring(df$team, first=1, last=3)
```

```
#view updated data frame
```

df

team first3

1 Mavericks Mav

2 Hornets Hor

3 Rockets Roc

4 Grizzlies Gri

Notice that the new column contains the first three characters of the "team" column.

Example 3: Extract Last N Characters

The following code shows how to use the `substring()` function to extract the last 3 characters of the "team" column:

```
#create new column that contains last 3 characters  
df$last3 <- substring(df$team, nchar(df$team)-3+1,  
nchar(df$team))
```

#view updated data frame

df

team last3

1 Mavericks cks

2 Hornets ets

3 Rockets ets

4 Grizzlies ies

Notice that the new column contains the last three characters of the "team" column.

Example 4: Replace a Substring

The following code shows how to use the `substring()` function to replace the first 3 characters of the values in the "team" column with 3 asterisks:

```
#replace first 3 characters with asterisks in team column
```

```
substring(df$team, first=1, last=3) <- "***"
```

```
#view updated data frame
```

```
df
```

```
team
```

```
1 ***ericks
```

```
2 ***nets
```

```
3 ***kets
```

```
4 ***zzlies
```

Notice that the first three characters of each team name has been replaced with asterisks.

Additional Resources

The following tutorials explain how to perform other common operations with strings in R:

[How to Perform Partial String Matching in R](#)

ARABPSYCHOLOGY.COM