

How can a new level be added to a factor in R?

Authored by
stats writer

June 26, 2024

RECOMMENDED CITATION

stats writer (2024). *How can a new level be added to a factor in R?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=153256>

A new level can be added to a factor in R by using the "levels" function and specifying the name of the factor and the new level to be added. This will update the factor's levels and allow for the inclusion of the new level in future data operations. Additionally, the "relevel" function can be used to change the order of the factor levels, with the new level being placed at a desired position. These functions can be helpful in organizing and analyzing data in R, allowing for the incorporation of new levels and improving the accuracy of data analysis.

Add New Level to Factor in R (With Example)

You can use the following basic syntax to add a new level to a factor variable in R:

```
levels(df$my_factor) <- c(levels(df$my_factor),  
'new_level')
```

The following example shows how to use this syntax in practice.

Example: Add New Level to Factor in R

Suppose we have the following data frame in R that shows the number of sales made in different regions for some retail store:

```
#create data frame
```

```
df <- data.frame(region=factor(c('A', 'B', NA, 'D', NA, 'F')),  
sales=c(12, 18, 21, 14, 34, 40))
```

```
#view data frame  
df
```

```
region sales
```

```
1 A 12
```

```
2 B 18
```

```
3 <NA> 21
```

```
4 D 14
```

```
5 <NA> 34
```

```
6 F 40
```

Notice that the region variable is a factor.

To view the levels for this factor, we can use the levels() function:

```
#view factor levels for region  
levels(df$region)
```

```
"A" "B" "D" "F"
```

We can use the following syntax to add a new factor level called "no region":

```
#add factor level called 'no region'
```

```
levels(df$region) <- c(levels(df$region), 'no region')
```

```
#convert each NA to 'no region'
```

```
df$region <- 'no region'
```

```
#view factor levels for region
```

```
levels(df$region)
```

```
"A" "B" "D" "F" "no region"
```

The new level called "no region" has been added as a factor level.

If we'd like, we can use the `table()` function to count the occurrence of each factor level:

```
#view occurrences of each factor level
```

```
table(df$region)
```

```
A B D F no region
```

```
1 1 1 1 2
```

From the output we can see that the new factor level called "no region" occurs twice in the region column of the data frame.