

How to Fix aggregate function missing, defaulting to 'length' in R?

Authored by
stats writer

June 29, 2024

RECOMMENDED CITATION

stats writer (2024). *How to Fix aggregate function missing, defaulting to 'length' in R?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=160065>

The message "Fix in R: aggregate function missing, defaulting to 'length'" occurs when the aggregate function is not specified in a code written in the R programming language. This causes the program to default to using the length function instead. To fix this issue, the user must ensure that the correct aggregate function is specified in the code, or if no specific aggregate function is needed, the code can be modified to use the length function. This message serves as a reminder for users to review their code and ensure that all necessary functions are properly specified.

Fix in R: aggregate function missing, defaulting to 'length'

One error you may encounter when using R is:

Aggregation function missing: defaulting to length

This error occurs when you use the dcast function from the reshape2 package to convert a data frame from a , but more than one value could be placed in the individual cells of the wide data frame.

The following example shows how to fix this error in practice.

How to Reproduce the Error

Suppose we have the following data frame in R that contains information about the sales of various products:

#create data frame

```
df <- data.frame(store=c('A', 'A', 'A', 'A', 'B', 'B', 'B', 'B'),  
promotion=c('Y', 'Y', 'N', 'N', 'Y', 'Y', 'N', 'N'),  
product=c(1, 2, 1, 2, 1, 2, 1, 2),  
sales=c(12, 18, 29, 20, 30, 11, 15, 22))
```

```
#view data frame
```

```
df
```

```
store promotion product sales
```

```
1 A Y 1 12
```

```
2 A Y 2 18
```

```
3 A N 1 29
```

```
4 A N 2 20
```

```
5 B Y 1 30
```

```
6 B Y 2 11
```

```
7 B N 1 15
```

```
8 B N 2 22
```

Now suppose we attempt to use the `dcast` function to convert the data frame from a long to a wide format:

```
library(reshape2)
```

```
#convert data frame to wide format
```

```
df_wide <- dcast(df, store ~ product, value.var="sales")
```

```
#view result
```

```
df_wide
```

```
Aggregation function missing: defaulting to length
```

```
store 1 2
```

```
1 A 2 2
```

```
2 B 2 2
```

Notice that the dcast function works but we receive the warning message of Aggregation function missing.

How to Fix the Error

The reason we receive a warning message is because for each combination of store and product, there are two potential values we could use for sales.

For example, for store A and product 1, the sales value could be 12 or 29.

Thus, the dcast function defaults to using "length" as the aggregate function.

For example, the wide data frame tells us that for store A and product 1, there are a total of 2 sales values.

If you'd instead like to use a different aggregation function, you can use `fun.aggregate`.

For example, we can use the following syntax to calculate the sum of sales by store and product:

```
library(reshape2)
```

```
#convert data frame to wide format
```

```
df_wide <- dcast(df, store ~ product, value.var="sales",  
fun.aggregate=sum)
```

```
#view result
```

```
df_wide
```

```
store 1 2
```

```
1 A 41 38
```

```
2 B 45 33
```

Here's how to interpret the values in the wide data frame:

The sum of sales for store A and product 1 is 41.

The sum of sales for store A and product 2 is 38.

The sum of sales for store B and product 1 is 45.

The sum of sales for store B and product 2 is 33.

Notice that we don't receive any warning message this time because we used the fun.aggregate argument.

Additional Resources

The following tutorials explain how to fix other common errors in R:

ARABPSYCHOLOGY.COM