

How can I count duplicates in R and what are some examples?

Authored by
stats writer

June 25, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I count duplicates in R and what are some examples?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=152811>

Counting duplicates in R is a process that involves identifying and counting the number of repeated values in a dataset. This can be useful for data analysis and cleaning, as it allows for a better understanding of the data and potential errors. To count duplicates in R, the function "duplicated()" can be used, which returns a logical vector indicating which values are duplicates. This can then be further manipulated using other functions such as "sum()" or "table()" to obtain the total count of duplicates. Some examples of using this function include identifying and removing duplicate entries in a dataset, or identifying common errors such as typos or misspelled words. Overall, counting duplicates in R is a simple yet powerful tool for data management and analysis.

Count Duplicates in R (With Examples)

You can use the following methods to count duplicates in a data frame in R:

Method 1: Count Duplicate Values in One Column

```
sum(duplicated(df$my_column))
```

Method 2: Count Duplicate Rows

```
nrow(df)
```

Method 3: Count Duplicates for Each Unique Row

```
library(dplyr)
```

```
df %>% group_by_all() %>% count
```

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

```
#create data frame
```

```
df = data.frame(team=c('A', 'A', 'A', 'A', 'B', 'B', 'B', 'B'),  
position=c('G', 'G', 'G', 'F', 'G', 'G', 'F', 'F'),  
points=c(5, 5, 8, 10, 5, 7, 10, 10))
```

```
#view data frame
```

```
df
```

```
team position points
```

```
1 A G 5
```

```
2 A G 5
3 A G 8
4 A F 10
5 B G 5
6 B G 7
7 B F 10
8 B F 10
```

Example 1: Count Duplicate Values in One Column

The following code shows how to count the number of duplicate values in the **points** column:

```
#count number of duplicate values in points column  
sum(duplicated(df$points))
```

```
4
```

We can see that there are **4** duplicate values in the **points** column.

Example 2: Count Duplicate Rows

The following code shows how to count the number of duplicate rows in the data frame:

```
#count number of duplicate rowsnrow(df)
```

```
2
```

We can see that there are **2** duplicate rows in the data frame.

```
#display duplicated rowsdf
```

```
team position points  
2 A G 5  
8 B F 10
```

Example 3: Count Duplicates for Each Unique Row

The following code shows how to count the number of duplicates for each unique row in the data

frame:

```
library(dplyr)
```

```
#count number of duplicate rows in data frame
```

```
df %>% group_by_all() %>% count
```

```
# A tibble: 6 x 4
```

```
# Groups: team, position, points
```

```
team position points n
```

```
1 A F 10 1
```

```
2 A G 5 2
```

```
3 A G 8 1
```

```
4 B F 10 2
```

```
5 B G 5 1
```

```
6 B G 7 1
```

The **n** column displays the number of duplicates for each unique row.