

How can I count the unique values within a group in R, and what are some examples of this process?

Authored by
stats writer

June 29, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I count the unique values within a group in R, and what are some examples of this process?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=160295>

Counting the unique values within a group in R refers to the process of determining the number of distinct elements in a specific category or subset of data. This can be done by using the "table" or "aggregate" functions in R, which allow for the grouping of data and counting of unique values within each group. For example, if we have a dataset of student grades with categories for different subjects, we can use the "table" function to count the number of unique grades for each subject. Another example would be counting the number of unique customer IDs within each state in a sales dataset using the "aggregate" function. This process is useful for obtaining a better understanding of the distribution of data within different groups, and can help in making informed decisions and analysis.

Count Unique Values by Group in R (With Examples)

You can use the following methods to count the number of unique values by group in R:

Method 1: Using Base R

```
results <- aggregate(data=df, values_var~group_var,  
function(x) length(unique(x)))
```

Method 2: Using dplyr

```
library(dplyr)
```

```
results <- df %>%
```

```
group_by(group_var) %>%
```

```
summarize(count = n_distinct(values_var))
```

Method 3: Using data.table

```
library(data.table)
```

```
df <- data.table(df)
```

```
results <- df
```

Each method returns the exact same result, but the base R method tends to be significantly slower when working with large data frames.

The following examples show how to use each of these methods in practice with the following data frame:

```
#create data frame
```

```
df <- data.frame(team=c('A', 'A', 'A', 'A', 'B', 'B', 'C', 'C',  
'C'),
```

```
points=c(10, 10, 14, 14, 18, 19, 20, 20, 20))
```

```
#view data frame
```

```
df
```

```
team points
```

```
1 A 10
```

```
2 A 10
```

```
3 A 14
```

```
4 A 14
```

5 B 18

6 B 19

7 C 20

8 C 20

9 C 20

Method 1: Count Unique Values by Group Using Base R

The following code shows how to count the number of distinct points values for each team using base R:

```
#count unique points values by team
results <- aggregate(data=df, points~team, function(x)
length(unique(x)))
```

```
#view results
```

```
results
```

```
team points
```

```
1 A 2
```

```
2 B 2
```

```
3 C 1
```

From the output we can see:

There are 2 unique points values for team A. There are 2

unique points values for team B. There is 1 unique points value for team C.

Method 2: Count Unique Values by Group Using dplyr

The following code shows how to count the number of distinct points values for each team using dplyr:

```
library(dplyr)

#count unique points values by team
results <- df %>%
group_by(team) %>%
summarize(count = n_distinct(points))

#view results
results

# A tibble: 3 x 2
  team count
1 A 2
2 B 2
3 C 1
```

Notice that these results match the ones from the base R method.

Method 3: Count Unique Values by Group Using data.table

The following code shows how to count the number of distinct points values for each team using data.table:

```
library(data.table)
```

```
#convert data frame to data table
```

```
df <- data.table(df)
```

```
#count unique points values by team
```

```
results <- df
```

```
#view results
```

```
results
```

```
team count
```

```
1: A 2
```

```
2: B 2
```

```
3: C 1
```

Notice that these results match the ones from the previous two methods.

Additional Resources

The following tutorials explain how to perform other

common operations using dplyr:

ARABPSYCHOLOGY.COM