How can I add a count column to a data frame in R?

Authored by

stats writer

June 27, 2024

To add a count column to a data frame in R, the user can utilize the "row_number()" function from the "dplyr" package. This function generates a sequential count for each row in the data frame. The user can then use the "mutate()" function to add this new count column to the original data frame. This process allows for efficient and accurate counting of rows in a data frame, which can be useful in various data analysis tasks.

Add a Count Column to a Data Frame in R

You can use the following basic syntax to add a 'count' column to a data frame in R:

```
df %>%
group_by(var1) %>%
mutate(var1_count = n())
```

This particular syntax adds a column called var1_count to the data frame that contains the count of values in the column called var1.

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

Example: Add Count Column in R

Suppose we have the following data frame in R that contains information about various basketball players:

```
#define data frama
```

```
df <- data.frame(team=c('A', 'A', 'A', 'B', 'B', 'B', 'B', 'B'),
position=c('G', 'F', 'F', 'F', 'G', 'G', 'F', 'F'),
points=c(18, 22, 19, 14, 14, 11, 20, 28))

#view data frame
df

team position points
1 A G 18
2 A F 22
3 A F 19
4 B F 14
5 B G 14
6 B G 11
7 B F 20
8 B F 28
```

We can use the following code to add a column called team_count that contains the count of each team:

```
library(dplyr)

#add column that shows total count of each team
df %>%
group_by(team) %>%
```

mutate(team_count = n())

 # A tibble: 8 x 4

 # Groups: team

team position points team_count

1 A G 18 3

2 A F 22 3

3 A F 19 3

4 B F 14 5

5 B G 14 5

6 B G 11 5

7 B F 20 5

8 B F 28 5

There are 3 rows with a team value of A and 5 rows with a team value of B.

Thus:

For each row where the team is equal to A, the value in the team_count column is 3.For each row where the team is equal to B, the value in the team_count column is 5.

You can also add a 'count' column that groups by multiple variables.

For example, the following code shows how to add a 'count' column that groups by the team and position variables:

```
library(dplyr)

#add column that shows total count of each team and position
df %>%
group_by(team, position) %>%
mutate(team_pos_count = n())

# A tibble: 8 x 4
# Groups: team, position
team position points team_pos_count

1 A G 18 1
2 A F 22 2
3 A F 19 2
4 B F 14 3
5 B G 14 2
6 B G 11 2
```

7 B F 20 3

8 B F 28 3

From the output we can see:

There is 1 row that contains A in the team column and G in the position column.There are 2 rows that contain A in the team column and F in the position column.There are 3 rows that contain B in the team column and F in the position column.There are 2 rows that contain B in the team column and F in the position column.