How can I create a crosstab using dplyr, and can you provide examples?

Authored by

stats writer

May 6, 2024

Cross-tabulation, or crosstab, is a statistical technique used to summarize and analyze data by grouping it into two or more dimensions. This can be achieved using the dplyr package in R, which is a popular tool for data manipulation and analysis. To create a crosstab using dplyr, one can use the `group_by()` function to group the data by the desired dimensions and then use the `summarize()` function to calculate summary statistics. This process can be further customized using various dplyr functions such as `mutate()`, `filter()`, and `arrange()`. Examples of creating crosstabs using dplyr can be found in the package's documentation or through online resources such as tutorials or forums.

Create a Crosstab Using dplyr (With Examples)

You can use the following basic syntax to produce a crosstab using functions from the and packages in R:

```
df %>%group_by(var1, var2) %>%tally() %>%spread(var1, n)
```

The following examples show how to use this syntax in practice.

Example 1: Create Basic Crosstab

Suppose we have 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', 'F', 'C', 'G', 'F', 'F', 'C'),
points=c(7, 7, 8, 11, 13, 15, 19, 13))
```

#view data frame
df

team position points
1 A G 7
2 A G 7
3 A F 8
4 A C 11
5 B G 13
6 B F 15
7 B F 19
8 B C 13

We can use the following syntax to create a crosstab for the 'team' and 'position' variables:

library(dplyr)
library(tidyr)

#produce crosstab df %>% group_by(team, position) %>% tally() %>% spread(team, n)

A tibble: 3 x 3
position A B
1 C 1 1

2 F 1 2
3 G 2 1

Here's how to interpret the values in the crosstab:

There is 1 player who has a position of 'C' and belongs to team 'A'There is 1 player who has a position of 'C' and belongs to team 'B'There is 1 player who has a position of 'F' and belongs to team 'A'There are 2 players who have a position of 'F' and belong to team 'B'There are 2 players who have a position of 'G' and belong to team 'A'There is 1 player who has a position of 'G' and belongs to team 'B'

Note that we can switch the rows and columns of the crosstab by switching the variable used in the spread() function:

```
library(dplyr)
library(tidyr)

#produce crosstab with 'position' along columnsdf %>% group_by(team, position) %>% tally() %>% spread(position, n)
```

 # A tibble: 2 x 4

 # Groups: team

team C F G

1 A 1 1 2

2 B 1 2 1

The following tutorials explain how to perform other common functions in dplyr: