How can you create new variables in R using the mutate() function and the case_when() statement?

Authored by

stats writer

April 18, 2024

The mutate() function and the case_when() statement are two useful tools in R for creating new variables. By using the mutate() function, a new variable can be added to an existing data frame by specifying its name and providing a calculation or transformation. Additionally, the case_when() statement allows for the creation of new variables based on specific conditions or logical expressions. By combining these two functions, users can easily create new variables in R that meet their specific needs and enhance their data analysis capabilities.

Create New Variables in R with mutate() and case_when()

Often you may want to create a new variable in a data frame in R based on some condition. Fortunately this is easy to do using the mutate() and case_when() functions from the dplyr package.

This tutorial shows several examples of how to use these functions with the following data frame:

```
#create data frame
df <- data.frame(player = c('a', 'b', 'c', 'd', 'e'),
position = c('G', 'F', 'F', 'G', 'G'),
points = c(12, 15, 19, 22, 32),
rebounds = c(5, 7, 7, 12, 11))

#view data frame
df

player position points rebounds
```

1 a G 12 5

2 b F 15 7

3 c F 19 7

4 d G 22 12

5 e G 32 11

Example 1: Create New Variable Based on One Existing Variable

The following code shows how to create a new variable called 'scorer' based on the value in the points column:

```
library(dplyr)

#define new variable 'scorer' using mutate() and case_when()
df %>%
mutate(scorer = case_when(points < 15 ~ 'low',
points < 25 ~ 'med',
points < 35 ~ 'high'))
```

player position points rebounds scorer

1 a G 12 5 low

2 b F 15 7 med

3 c F 19 7 med

4 d G 22 12 med

5 e G 32 11 high

Example 2: Create New Variable Based on Several Existing Variables

The following code shows how to create a new variable called 'type' based on the value in the player and position column:

```
library(dplyr)

#define new variable 'type' using mutate() and case_when()
df %>%
mutate(type = case_when(player == 'a' | player == 'b' ~ 'starter',
player == 'c' | player == 'd' ~ 'backup',
position == 'G' ~ 'reserve'))
```

player position points rebounds type
1 a G 12 5 starter
2 b F 15 7 starter
3 c F 19 7 backup
4 d G 22 12 backup
5 e G 32 11 reserve

The following code shows how to create a new variable called 'valueAdded' based on the value in the points and rebounds columns:

```
library(dplyr)

#define new variable 'valueAdded' using mutate() and case_when()
df %>%
mutate(valueAdded = case_when(points <= 15 & rebounds <=5 ~ 2,
points <=15 & rebounds > 5 ~ 4,
points < 25 & rebounds < 8 ~ 6,
points < 25 & rebounds > 8 ~ 7,
points >=25 ~ 9))
```

```
player position points rebounds valueAdded
1 a G 12 5 2
2 b F 15 7 4
3 c F 19 7 6
4 d G 22 12 7
5 e G 32 11 9
```

How to Rename Columns in R

How to Remove Columns in R

How to Filter Rows in R