How can I find the column with the maximum value for each row in R?

Authored by

stats writer

June 25, 2024

To find the column with the maximum value for each row in R, you can use the "which.max" function which returns the index of the maximum value in a given vector. This function can be applied to each row of a data frame using the "apply" function with the argument "MARGIN = 1" to specify row-wise operations. This will result in a vector containing the column index of the maximum value for each row. Alternatively, you can use the "max.col" function which returns the column index of the maximum value for each row in a data frame. Both methods will allow you to efficiently find the column with the maximum value for each row in R.

R: Find Column with Max Value for Each Row

You can use the following syntax to find the column with the max value for each row in a data frame in R:

df$max_col <- colnames(df)

Note that the argument ties.method='first' specifies that the first max column should be returned if there are multiple columns with a max value in a given row.

Other values you can provide to this argument include random and last, if you'd like to return a random max column or the last max column instead.

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

Example: Find Column with Max Value for Each Row in R

Suppose we have the following data frame in R that

contains information about the number of points scored by six different basketball players during three games:

```
#create data frame
df <- data.frame(game1=c(23, 20, 14, 12, 19, 15),
game2=c(9, 10, 11, 13, 13, 15),
game3=c(29, 11, 22, 19, 14, 15))

#view data frame
df
```

```
game1 game2 game3
1 23 9 29
2 20 10 11
3 14 11 22
4 12 13 19
5 19 13 14
6 15 15 15
```


Suppose we would like to create a new column that contains the name of the column with the max value in each row of the data frame.

We can use the following syntax to do so:

#create new column that contains column with max value for each row
df$max_col <- colnames(df)

#view updated data frame
df

game1 game2 game3 max_col
1 23 9 29 game3
2 20 10 11 game1
3 14 11 22 game3
4 12 13 19 game3
5 19 13 14 game1
6 15 15 15 game1

The new column called max_col contains the name of the column with the max value in each row.

For example:

In the first row, game3 contained the max value.In the second row, game1 contained the max value.In the third row, game3 contained the max value.

And so on.

Note that each column in the last row has the same value.

Since we specified ties.method='first' in the max.col() function, the code returned game1 as the column with the max value since this is the first max column.