# How can I use dplyr to summarise data while keeping all columns in the dataset?

Authored by

**stats writer**

June 26, 2024

Dplyr is a powerful R package that allows for efficient data manipulation and summarization. One of its useful features is the ability to summarise data while retaining all columns in the dataset. This means that the summarized data will still contain all the original columns, with the summary values added as new columns. This can be achieved by using the "group_by" and "summarise" functions in dplyr, which allow for grouping the data by a specific variable and calculating summary statistics for each group. By using this method, the original dataset is preserved while still obtaining useful summary information. This can be particularly helpful in analyzing and visualizing complex datasets, as it allows for a more comprehensive understanding of the data.

## dplyr: Summarise Data But Keep All Columns

## When using the summarise() function in , all variables not included in the summarise() or group_by() functions will automatically be dropped.

## However, you can use the mutate() function to summarize data while keeping all of the columns in the data frame.

## The following example shows how to use this function in practice.

### Example: Summarise Data But Keep All Columns Using dplyr

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

## #create data frame
## df <- data.frame(team=rep(c('A', 'B', 'C'), each=3),

**points=c(4, 9, 8, 12, 15, 14, 29, 30, 22),**

**assists=c(3, 3, 2, 5, 8, 10, 4, 5, 12))**

**#view data frame**

**df**

**team points assists**

**1 A 4 3**

**2 A 9 3**

**3 A 8 2**

**4 B 12 5**

**5 B 15 8**

**6 B 14 10**

**7 C 29 4**

**8 C 30 5**

**9 C 22 12**

**We can use the following syntax to summarize the mean points scored by team:**

```
library(dplyr)#summarize mean points values by team
df %>%
group_by(team) %>%
summarise(mean_pts = mean(points))
```

**# A tibble: 3 x 2**

**team mean_pts**

**1 A 7**

**2 B 13.7**

**3 C 27**

**The column called mean_pts displays the mean points scored by each team.**

**From the output we can see:**

**The mean points scored by players on team A is 7.The mean points scored by players on team B is 13.7.The mean points scored by players on team C is 27.**

**However, suppose we would like to keep all other columns from the original data frame.**

**We can use the following syntax with the mutate() function to do so:**

```
library(dplyr)#summarize mean points values by team and keep all columns
df %>%
group_by(team) %>%
```

```
mutate(mean_pts = mean(points)) %>%
ungroup()

# A tibble: 9 x 4
team points assists mean_pts

1 A 4 3 7
2 A 9 3 7
3 A 8 2 7
4 B 12 5 13.7
5 B 15 8 13.7
6 B 14 10 13.7
7 C 29 4 27
8 C 30 5 27
9 C 22 12 27
```

By using the mutate() function, we're able to create a new column called mean_pts that summarizes the mean points scored by team while also keeping all other columns from the original data frame.

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