How can I select the first row within each group using dplyr?

Authored by

stats writer

April 20, 2024

The process of selecting the first row within each group using dplyr involves using the group_by() function to group the data by a specific variable, followed by the slice() function to extract the first row from each group. This allows for easy manipulation and analysis of data within specific groups.

Select the First Row by Group Using dplyr

Often you may want to select the first row in each group using the **dplyr** package in R. You can use the following basic syntax to do so:

```
df %>%
group_by(group_var) %>%
arrange(values_var) %>%
filter(row_number()==1)
```

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

Example: Select the First Row by Group in R

Suppose we have the following dataset in R:

```
#create dataset
df <- data.frame(team=c('A', 'A', 'A', 'B', 'B', 'B', 'C', 'C', 'C', 'C'),
points=c(4, 9, 7, 7, 6, 13, 8, 8, 4, 17))

#view dataset
```

```
df

team points
1 A 4
2 A 9
3 A 7
4 B 7
5 B 6
6 B 13
7 C 8
8 C 8
9 C 4
10 C 17
```

The following code shows how to use the dplyr package to select the first row by group in R:

```
library(dplyr)

df %>%
group_by(team) %>%
arrange(points) %>%
filter(row_number()==1)

# A tibble: 3 x 2
```

```
# Groups: team
team points

1 A 4
2 C 4
3 B 6
```

By default, arrange() sorts the values in ascending order but we can easily sort the values in descending order instead:

```
df %>%
group_by(team) %>%
arrange(desc(points)) %>%
filter(row_number()==1)

# A tibble: 3 x 2
# Groups: team
team points

1 C 17
2 B 13
3 A 9
```

Note that you can easily modify this code to select the

nth row by each group. Simply change row_number() == n.

For example, if you'd like to select the 2nd row by group, you can use the following syntax:

```
df %>%
group_by(team) %>%
arrange(desc(points)) %>%
filter(row_number()==2)
```

Or you could use the following syntax to select the last row by group:

```
df %>%
group_by(team) %>%
arrange(desc(points)) %>%
filter(row_number()==n())
```


How to Arrange Rows in R

How to Count Observations by Group in R

How to Find the Maximum Value by Group in R