

# How can I select unique rows in a data frame in R?

Authored by  
**stats writer**

June 27, 2024

## RECOMMENDED CITATION

stats writer (2024). *How can I select unique rows in a data frame in R?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=154864>

In R, a data frame is a commonly used data structure that consists of rows and columns. In some cases, it may be necessary to select only the unique rows in a data frame, meaning that rows with duplicate values are omitted. This can be accomplished by using the function "unique()" which returns a new data frame with only the unique rows from the original data frame. This process can be useful in data analysis and manipulation, as it allows for the removal of duplicate observations and a more accurate representation of the data. By using this function, individuals can easily select and work with unique rows in a data frame in R.

## Select Unique Rows in a Data Frame in R

You can use the following methods to select unique rows from a data frame in R:

### Method 1: Select Unique Rows Across All Columns

```
library(dplyr)
```

```
df %>% distinct()
```

### Method 2: Select Unique Rows Based on One Column

```
library(dplyr)df %>% distinct(column1, .keep_all=TRUE)
```

### Method 3: Select Unique Rows Based on Multiple Columns

```
library(dplyr)
```

```
df %>% distinct(column1, column2, .keep_all=TRUE)
```

This tutorial explains how to use each method in practice with the following data frame:

```
#create data frame
```

```
df <- data.frame(team=c('A', 'A', 'A', 'A', 'B', 'B', 'B', 'B'),  
position=c('G', 'G', 'F', 'F', 'G', 'G', 'F', 'F'),  
points=c(10, 10, 8, 14, 15, 15, 17, 17))
```

```
#view data frame
```

```
df
```

```
team position points
```

```
1 A G 10
```

```
2 A G 10
```

```
3 A F 8
```

```
4 A F 14
```

```
5 B G 15
```

```
6 B G 15
```

```
7 B F 17
```

```
8 B F 17
```

**Example 1: Select Unique Rows Across All Columns**

The following code shows how to select rows that have unique values across all columns in the data frame:

```
library(dplyr)
```

```
#select rows with unique values across all columns
```

```
df %>% distinct()
```

```
team position points
```

```
1 A G 10
```

```
2 A F 8
```

```
3 A F 14
```

```
4 B G 15
```

```
5 B F 17
```

We can see that there are five unique rows in the data frame.

**Note:** When duplicate rows are encountered, only the first unique row is kept.

**Example 2: Select Unique Rows Based on One Column**

The following code shows how to select unique rows based on the team column only.

```
library(dplyr)
```

```
#select rows with unique values based on team column  
only
```

```
df %>% distinct(team, .keep_all=TRUE)
```

```
team position points
```

```
1 A G 10
```

```
2 B G 15
```

**Note:** The argument `.keep_all=TRUE` tells R to keep all other columns in the output.

**Example 3: Select Unique Rows Based on Multiple Columns**

The following code shows how to select unique rows based on the team and position columns only.

```
library(dplyr)
```

```
#select rows with unique values based on team and  
position columns only
```

```
df %>% distinct(team, position, .keep_all=TRUE)
```

```
team position points
```

```
1 A G 10
```

**2 A F 8**

**3 B G 15**

**4 B F 17**

**Four rows are returned, since there are four unique combinations of values across the team and position columns.**

ARABPSYCHOLOGY.COM