

How can I remove duplicate rows in R, and could you provide some examples?

Authored by
stats writer

May 6, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I remove duplicate rows in R, and could you provide some examples?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=143405>

In R, duplicate rows in a dataset can be removed using the "unique()" function. This function identifies and removes all duplicate rows, leaving only unique rows in the dataset. It can be applied to both data frames and matrices. For example, if a data frame has multiple rows with the same values, the "unique()" function will keep only one of those rows and remove the rest. This allows for a clean and organized dataset without any redundant information. Furthermore, the "unique()" function can also be used with other functions such as "subset()" to remove duplicate rows based on specific criteria. Overall, using the "unique()" function is an efficient way to handle duplicate rows in R, ensuring accurate and concise data analysis.

Remove Duplicate Rows in R (With Examples)

You can use one of the following two methods to remove duplicate rows from a data frame in R:

Method 1: Use Base R

```
#remove duplicate rows across entire data frame  
df
```

```
#remove duplicate rows across specific columns of  
data frame  
df[, ]
```

Method 2: Use dplyr

```
#remove duplicate rows across entire data frame  
df %>%  
distinct(.keep_all = TRUE)
```

#remove duplicate rows across specific columns of data frame

df %>%

distinct(var1, .keep_all = TRUE)

The following examples show how to use this syntax in practice with the following data frame:

#define data frame

```
df <- data.frame(team=c('A', 'A', 'A', 'B', 'B', 'B'),  
position=c('Guard', 'Guard', 'Forward', 'Guard', 'Center',  
'Center'))
```

#view data frame

df

team position

1 A Guard

2 A Guard

3 A Forward

4 B Guard

5 B Center

6 B Center

Example 1: Remove Duplicate Rows Using Base R

The following code shows how to remove duplicate rows from a data frame using functions from base R:

```
#remove duplicate rows from data frame  
df
```

```
team position
```

```
1 A Guard
```

```
3 A Forward
```

```
4 B Guard
```

```
5 B Center
```

The following code shows how to remove duplicate rows from specific columns of a data frame using base R:

```
#remove rows where there are duplicates in the 'team'  
column  
df), ]
```

```
team position
```

```
1 A Guard
```

```
4 B Guard
```

Example 2: Remove Duplicate Rows Using dplyr

The following code shows how to remove duplicate rows from a data frame using the `distinct()` function from the package:

```
library(dplyr)
```

```
#remove duplicate rows from data frame
```

```
df %>%
```

```
distinct(.keep_all = TRUE)
```

```
team position
```

```
1 A Guard
```

```
2 A Forward
```

```
3 B Guard
```

```
4 B Center
```

Note that the `.keep_all` argument tells R to keep all of the columns from the original data frame.

The following code shows how to use the `distinct()` function to remove duplicate rows from specific columns of a data frame:

```
library(dplyr)
```

#remove duplicate rows from data frame

df %>%

distinct(team, .keep_all = TRUE)

team position

1 A Guard

2 B Guard

The following tutorials explain how to perform other common functions in R: