

How can I select random rows in R using dplyr?

Authored by
stats writer

June 29, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I select random rows in R using dplyr?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=159883>

The process of selecting random rows in R using dplyr involves utilizing the `sample_n()` function from the dplyr package. This function allows for the random selection of a specified number of rows from a dataset, based on a given percentage or specific number. This method can be useful for creating random samples for statistical analysis or for generating randomized data for testing purposes. By incorporating the dplyr package, this process can be easily implemented and integrated into data manipulation and analysis workflows.

Select Random Rows in R Using dplyr

You can use the following methods to select random rows from a data frame in R using functions from the package:

Method 1: Select Random Number of Rows

```
df %>% sample_n(5)
```

This function randomly selects 5 rows from the data frame.

Method 2: Select Random Fraction of Rows

```
df %>% sample_frac(.25)
```

This function randomly selects 25% of all rows from the data frame.

The following examples show how to use each method

in practice with the following data frame in R:

```
#create data frame
```

```
df <- data.frame(team=c('A', 'B', 'C', 'D', 'E', 'F', 'G', 'H'),  
points=c(10, 10, 8, 6, 15, 15, 12, 12),  
rebounds=c(8, 8, 4, 3, 10, 11, 7, 7))
```

```
#view data frame
```

```
df
```

```
team points rebounds
```

```
1 A 10 8
```

```
2 B 10 8
```

```
3 C 8 4
```

```
4 D 6 3
```

```
5 E 15 10
```

```
6 F 15 11
```

```
7 G 12 7
```

```
8 H 12 7
```

Example 1: Select Random Number of Rows

We can use the following code to randomly select 5 rows from the data frame:

```
library(dplyr)
```

```
#randomly select 5 rows from data frame  
df %>% sample_n(5)
```

```
team points rebounds
```

```
1 F 15 11
```

```
2 A 10 8
```

```
3 D 6 3
```

```
4 G 12 7
```

```
5 B 10 8
```

Notice that five rows are randomly selected from the data frame.

Example 2: Select Random Fraction of Rows

We can use the following code to randomly select 25% of all rows from the data frame:

```
library(dplyr)
```

```
#randomly select 25% of all rows from data frame  
df %>% sample_frac(.25)
```

```
team points rebounds
```

```
1 E 15 10
```

```
2 G 12 7
```

Since the original data frame had 8 total values, 25% of 8 is equal to 2.

Note: You can find the complete documentation for the `sample_n` and `sample_frac` functions in dplyr .

Additional Resources

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

ARABPSYCHOLOGY.COM