

# Can I use the “Scale” function in R to only apply to numeric columns in a dataset?

Authored by  
**stats writer**

June 27, 2024

## RECOMMENDED CITATION

stats writer (2024). *Can I use the “Scale” function in R to only apply to numeric columns in a dataset?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=155035>

The "Scale" function in R is a useful tool for standardizing data in a dataset. However, it is important to note that this function can be applied to all columns in a dataset by default. If the intention is to only apply the "Scale" function to numeric columns, additional steps need to be taken. This can be achieved by first identifying the numeric columns in the dataset and then using the "Scale" function specifically on those columns. By doing so, the "Scale" function will only affect the numeric columns and leave the non-numeric columns unchanged.

## Scale Only Numeric Columns in R (With Example)

You can use the following syntax from the package to scale only the numeric columns in a data frame in R:

```
library(dplyr)
```

```
df %>% mutate(across(where(is.numeric), scale))
```

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

Example: Scale Only Numeric Columns Using dplyr

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

```
#create data frame
```

```
df <- data.frame(team=c('A', 'B', 'C', 'D', 'E'),  
points=c(22, 34, 30, 12, 18),  
assists=c(7, 9, 9, 12, 14),
```

```
rebounds=c(5, 10, 10, 8, 8))
```

```
#view data frame
```

```
df
```

```
team points assists rebounds
```

```
1 A 22 7 5
```

```
2 B 34 9 10
```

```
3 C 30 9 10
```

```
4 D 12 12 8
```

```
5 E 18 14 8
```

Suppose we would like to use the function in R to scale only the numeric columns in the data frame.

We can use the following syntax to do so:

```
library(dplyr)
```

```
#scale only the numeric columns in the data frame
```

```
df %>% mutate(across(where(is.numeric), scale))
```

```
team points assists rebounds
```

```
1 A -0.1348400 -1.153200 -1.56144012
```

```
2 B 1.2135598 -0.432450 0.87831007
```

```
3 C 0.7640932 -0.432450 0.87831007
```

```
4 D -1.2585064 0.648675 -0.09759001
5 E -0.5843065 1.369425 -0.09759001
```

Notice that the values in the three numeric columns (points, assists, and rebounds) have been scaled while the team column has remain unchanged.

### Technical Notes

The `scale()` function in R uses the following basic syntax:

```
scale(x, center = TRUE, scale = TRUE)
```

where:

**x:** Name of the object to scale  
**center:** Whether to subtract the mean when scaling. Default is TRUE.  
**scale:** Whether to divide by the standard deviation when scaling. Default is TRUE.

This function uses the following formula to calculate scaled values:

$$x_{scaled} = (x_{original} - \bar{x}) / s$$

**xoriginal: The original x-value  
x?: The sample means:  
The sample standard deviation**

**This is also known as *standardizing* data, which simply converts each original value into a .**

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

ARABPSYCHOLOGY.COM