

# How to Scale Only Numeric Columns in R

Tutorial from [DataScienceTut - How to scale only numeric columns by 'Jim'](#)

Using dplyr from the Tidyverse and the *scale()* function.

## Create data frame

```
df <- data.frame(Team=c('P1', 'P2', 'P3', 'P4', 'P5'),  
                 points=c(2, 3, 7, 22, 8),  
                 value=c(27, 39, 49, 82, 54))
```

df

	Team	points	value
1	P1	2	27
2	P2	3	39
3	P3	7	49
4	P4	22	82
5	P5	8	54

## Scaling

In this example, only scaling is performed on the value. The mean is subtracted and divided by the standard deviation.

This standardizes the value but does not shift its center.

- The mean is subtracted and divided by the Standard Deviation  $\frac{value - \mu}{SD}$

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

	Team	points	value
1	P1	-0.79813157	-1.1284228
2	P2	-0.67342351	-0.5447558
3	P3	-0.17459128	-0.0583667
4	P4	1.69602958	1.5467175
5	P5	-0.04988322	0.1848279

**Center - Does not involve scaling.**

```
centered <- df %>% mutate(across(where(is.numeric), scale, center=TRUE, scale=FALSE))
centered
```

	Team	points	value
1	P1	-6.4	-23.2
2	P2	-5.4	-11.2
3	P3	-1.4	-1.2
4	P4	13.6	31.8
5	P5	-0.4	3.8

**Centering can be helpful in various analytical scenarios. Here are a few reasons why centering can be beneficial:**

**Interpretation:** Centering allows you to interpret the effects of predictors or variables relative to their mean. By centering variables, you can interpret the regression coefficients as the change in the response variable associated with a one-unit change in the predictor variable when all other predictors are held constant at their means.

**Avoiding collinearity:** Centering variables can help mitigate issues related to collinearity or multicollinearity in regression analysis. When predictors are highly correlated, centering can help reduce the collinearity by making the predictors less correlated with each other.

**Stability and convergence:** In some numerical optimization procedures, centering variables can improve stability and convergence of the algorithm. Centering helps prevent large numbers from dominating the optimization process, making the computation more stable and reliable.

**Visualization:** Centering variables can aid in visualization by placing the center of the data at a meaningful reference point (such as zero) on the plot. This can make it easier to interpret and compare data points.

While centering is not always necessary or appropriate for every analysis, it can be advantageous in specific contexts to improve interpretation, address collinearity, enhance numerical stability, and facilitate visualization. The decision to center variables depends on the specific analytical objectives and the nature of the data being analyzed.