

# How can I group data by month in R, using an example?

Authored by  
**stats writer**

June 28, 2024

## RECOMMENDED CITATION

stats writer (2024). *How can I group data by month in R, using an example?*.

PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=156411>

Grouping data by month in R allows for the organization and analysis of data based on the time period of one month. This can be useful in identifying trends or patterns within a specific month, or comparing data across multiple months. An example of this would be grouping sales data by month to determine which month had the highest sales. Using the "group\_by" function in the dplyr package, one can easily group the data by month and then perform various calculations or visualizations on the grouped data. This allows for efficient and effective analysis of large datasets.

## Group Data by Month in R (With Example)

You can use the `floor_date()` function from the package in R to quickly group data by month.

This function uses the following basic syntax:

```
library(tidyverse)df %>%  
  group_by(month = lubridate::floor_date(date_column,  
    'month')) %>%  
  summarize(sum = sum(value_column))
```

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

Example: Group Data by Month in R

Suppose we have the following data frame in R that shows the total sales of some item on various dates:

```
#create data frame
```

```
df <- data.frame(date=as.Date(c('1/4/2022', '1/9/2022',  
'2/10/2022', '2/15/2022',  
'3/5/2022', '3/22/2022', '3/27/2022'), '%m/%d/%Y'),  
sales=c(8, 14, 22, 23, 16, 17, 23))
```

```
#view data frame
```

```
df
```

```
date sales
```

```
1 2022-01-04 8
```

```
2 2022-01-09 14
```

```
3 2022-02-10 22
```

```
4 2022-02-15 23
```

```
5 2022-03-05 16
```

```
6 2022-03-22 17
```

```
7 2022-03-27 23
```

We can use the following code to calculate the sum of sales, grouped by month:

```
library(tidyverse)
```

```
#group data by month and sum sales
```

```
df %>%
```

```
group_by(month = lubridate::floor_date(date, 'month'))
```

```
%>%  
summarize(sum_of_sales = sum(sales))
```

```
# A tibble: 3 x 2  
month sum_of_sales
```

```
1 2022-01-01 22  
2 2022-02-01 45  
3 2022-03-01 56
```

From the output we can see:

A total of 22 sales were made in January. A total of 45 sales were made in February. A total of 56 sales were made in March.

We can also aggregate the data using some other metric.

For example, we could calculate the max sales made in one day, grouped by month:

```
library(tidyverse)
```

```
#group data by month and find max sales  
df %>%
```

```
group_by(month = lubridate::floor_date(date, 'month'))  
%>%
```

```
summarize(max_of_sales = max(sales))
```

```
# A tibble: 3 x 2
```

```
month max_of_sales
```

```
1 2022-01-01 14
```

```
2 2022-02-01 23
```

```
3 2022-03-01 23
```

From the output we can see:

The max sales made in one day in January was 14. The max sales made in one day in February was 23. The max sales made in one day in March was 23.

Feel free to use whatever metric you'd like within the `summarize()` function.

Additional Resources

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