

How can I split a data frame in R and what are some examples of how to do so?

Authored by
stats writer

July 2, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I split a data frame in R and what are some examples of how to do so?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=165497>

Splitting a data frame in R refers to the process of dividing a large data frame into smaller subsets based on certain criteria. This can be useful for analyzing specific portions of the data or for creating separate datasets for different purposes. To split a data frame in R, one can use the "split" function which takes in the data frame as well as the variable or factor to split by. For example, if we have a data frame containing information about students such as their grades, age, and gender, we can split the data frame by gender using the code "split(df, df\$gender)". This will create two new data frames, one for male students and one for female students.

Another way to split a data frame is using the "subset" function, which allows us to create subsets based on specific conditions. For instance, we can create a subset of students who are above the age of 18 by using the code "subset(df, age > 18)". This will create a new data frame with only the information of students who are above 18 years old.

In summary, splitting a data frame in R allows for better organization and analysis of data by creating smaller, more manageable subsets. It can be done using the "split" or "subset" function, depending on the desired criteria for splitting the data.

Split a Data Frame in R (With Examples)

You can use one of the following three methods to split a data frame into several smaller data frames in R:

Method 1: Split Data Frame Manually Based on Row Values

```
#define first n rows to include in first data frame
```

```
n <- 4
```

```
#split data frame into two smaller data frames
```

```
df1 <- df
```

```
df2 <- df
```

Method 2: Split Data Frame into n Equal-Sized Data Frames

```
#define number of data frames to split into
```

```
n <- 3
```

```
#split data frame into n equal-sized data frames
```

```
split(df, factor(sort(rank(row.names(df))%%n)))
```

Method 3: Split Data Frame Based on Column Value

```
#split data frame based on particular column value
```

```
df1 <- df
```

```
df2 <- df
```

The following examples show how to use each method in practice with the following data frame:

```
#create data frame
```

```
df <- data.frame(ID=c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12),
```

```
sales=c(7, 8, 8, 7, 9, 7, 8, 9, 3, 3, 14, 10),
```

```
leads=c(0, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 0))
```

```
#view data frame
```

```
df
```

ID sales leads

1 1 7 0

2 2 8 0

3 3 8 1

4 4 7 1

5 5 9 0

6 6 7 1

7 7 8 1

8 8 9 0

9 9 3 1

10 10 3 0

11 11 14 1

12 12 10 0

Method 1: Split Data Frame Manually Based on Row Values

The following code shows how to split a data frame into two smaller data frames where the first one contains rows 1 through 4 and the second contains rows 5 through the last row:

```
#define row to split on
```

```
n <- 4
```

```
#split into two data frames
```

```
df1 <- df
```

```
df2 <- df
```

```
#view resulting data frames
```

```
df1
```

```
ID sales leads
```

```
1 1 7 0
```

```
2 2 8 0
```

```
3 3 8 1
```

```
4 4 7 1
```

```
df2
```

```
ID sales leads
```

```
5 5 9 0
```

```
6 6 7 1
```

```
7 7 8 1
```

```
8 8 9 0
```

```
9 9 3 1
```

```
10 10 3 0
```

```
11 11 14 1
```

```
12 12 10 0
```

Method 2: Split Data Frame into n Equal-Sized Data Frames

The following code shows how to split a data frame into n equal-sized data frames:

```
#define number of data frames to split into
```

```
n <- 3
```

```
#split data frame into n equal-sized data frames
```

```
split(df, factor(sort(rank(row.names(df))%%n)))
```

```
$`0`
```

```
ID sales leads
```

```
1 1 7 0
```

```
2 2 8 0
```

```
3 3 8 1
```

```
4 4 7 1
```

```
$`1`
```

```
ID sales leads
```

```
5 5 9 0
```

```
6 6 7 1
```

```
7 7 8 1
```

```
8 8 9 0
```

```
$`2`
```

ID sales leads

9 9 3 1

10 10 3 0

11 11 14 1

12 12 10 0

The result is three data frames of equal size.

Method 3: Split Data Frame Based on Column Value

```
#split data frame based on particular column value
```

```
df1 <- df
```

```
df2 <- df
```

```
#view resulting data frames
```

```
df1
```

ID sales leads

1 1 7 0

2 2 8 0

5 5 9 0

8 8 9 0

10 10 3 0

12 12 10 0

```
df2
```

ID sales leads

3 3 8 1

4 4 7 1

6 6 7 1

7 7 8 1

9 9 3 1

11 11 14 1

Note that df1 contains all rows where 'leads' was equal to zero in the original data frame and df2 contains all rows where 'leads' was equal to one in the original data frame.

Additional Resources