How can I use the Gather function in R, and what are some examples of its application?

Authored by

stats writer

May 1, 2024

The Gather function in R is a useful tool for reshaping data frames by consolidating multiple columns into key-value pairs. This function allows users to easily transform wide data sets into long formats, making it easier to analyze and manipulate data. Some examples of its applications include converting survey data with multiple response options into a tidy format, merging multiple data sets with different column names, and creating visualizations with grouped data. Overall, the Gather function is a powerful tool for data manipulation and organization in R.

Use Gather Function in R (With Examples)

The gather() function from the package can be used to "gather" a key-value pair across multiple columns.

This function uses the following basic syntax:

gather(data, key value, ...)

where:

data: Name of the data frame
key: Name of the key column to create
value: Name of the value column to create
... : Specify which columns to gather from

The following examples show how to use this function in practice.

Example 1: Gather Values From Two Columns

Suppose we have the following data frame in R:

```
#create data frame
df <- data.frame(player=c('A', 'B', 'C', 'D'),
year1=c(12, 15, 19, 19),
year2=c(22, 29, 18, 12))

#view data frame
df

player year1 year2
1 A 12 22
2 B 15 29
3 C 19 18
4 D 19 12
```

We can use the gather() function to create two new columns called "year" and "points" as follows:

```
library(tidyr)
#gather data from columns 2 and 3
gather(df, key="year", value="points", 2:3)

player year points
1 A year1 12
2 B year1 15
3 C year1 19
```

4 D year1 19

5 A year2 22

6 B year2 29

7 C year2 18

8 D year2 12

Example 2: Gather Values From More Than Two Columns

Suppose we have the following data frame in R:

```
#create data frame
df2 <- data.frame(player=c('A', 'B', 'C', 'D'),
year1=c(12, 15, 19, 19),
year2=c(22, 29, 18, 12),
year3=c(17, 17, 22, 25))

#view data frame
df2
```

player year1 year2 year3

1 A 12 22 17

2 B 15 29 17

3 C 19 18 22

4 D 19 12 25

We can use the gather() function to "gather" the values from columns 2, 3, and 4 into two new columns called "year" and "points" as follows:

```
library(tidyr)
#gather data from columns 2, 3, and 4
gather(df, key="year", value="points", 2:4)
```

```
player year points
1 A year1 12
2 B year1 15
3 C year1 19
4 D year1 19
5 A year2 22
6 B year2 29
7 C year2 18
8 D year2 12
9 A year3 17
10 B year3 17
11 C year3 22
12 D year3 25
```

Every column is a variable.

Every row is an observation.

Every cell is a single value.

The tidyr package uses four core functions to create tidy data:

1. The function.

2. The gather() function.

3. The function.

4. The function.

If you can master these four functions, you will be able to create "tidy" data from any data frame.