TMC 204
Statistical Data Analysis with R
Unit 4
Manipulating Objects Part 4

Presented By : Aditya Joshi

Asst. Professor

Department of Computer Application

Graphic Era Deemed to be University

10-04-2020

R Data Reshaping

data reshaping is totally dependent on a data frame in R

R data reshaping is all about changing the way in which data is organized into rows and columns.

Most of the time data processing in R is done by taking the input data as a data frame.

Also, extracting data from the rows and columns of a data frame is an easy task but there are situations when we need the data frame in a format that is different from the format in which we received it.

In R, it has many functions to split, merge and change the rows to columns in a data frame.

Joining Columns and Rows in a Data Frame

We use vectors to create a data frame using the cbind()function.

1. cbind()

We use cbind() function to combine vector, matrix or data frame by columns.

cbind(x1,x2,…)

x1,x2: vector, matrix, data frames

2. rbind()

We use rbind() function to combine vector, matrix or data frame by rows.

rbind(x1,x2,…)

x1,x2: vector, matrix, data frames

3. melt()

We use melt() function in R to convert an object into a molten data frame. It takes input in the form of a wide format and stacks multiple columns into a single column of the data. The melt() function has the following attributes –

```
melt(data, …, na.rm = FALSE, value.name = “value”)
```

data – The input data that is to be melted.

…. – Arguments that are passed to or from.

na.rm – Used for converting explicit missings into implicit missings.

value.name – Used for storing values in variables.

In the following example, make use of the mtcars data and apply melt() function to the id variables – ‘gears’ and ‘carbs’ and the measured variables – ‘mpg’, ‘cyl’, ‘disp’, ‘hp’. We use this melt function to melt the mtcars data frame.

```
library(reshape)
library(datasets)
str(mtcars)
molted =
melt(mtcars,id.vars=c("gear","carb"),measured.vars=c("mpg","cyl","disp","hp"))
str(molted)
molted[sample(nrow(molted),10),]
```

4. dcast()

Once you have a molten dataset with you, it is ready to be cast or reshaped. We will construct the original dataset using the dcast() function. The dcast() function:

```
> head(dcast(molted,gear+carb~variable, length))
```

There are three arguments in dcast():

data – The data attribute taken in the molten data frame.

formula – The formula specifies how the data is to be cast. The formula is present in the form of x_variable ~ y_variable, but there can be multiple variables present.

fun.aggregate – We use this function if there is data aggregation due to implementation of the casting formula. (example – length(), mean() and sum()).

What if we use only one of the variables gear or carb in dcast()?

```
> dcast(molted,gear~variable,mean)
```

We can also perform a transpose operation on this as follows:

```
> dcast(molted,variable~gear,mean)
```

We can also avail . (dot) which does not signify any variable:

```
> dcast(molted,variable~.,mean)
```

We can also perform:

```
> dcast(molted,carb~.,mean)
```

Margins, that are known as column totals can be created by specifying an attribute ‘margin’ and setting it to TRUE.

```
> dcast(molted,variable~gear,mean,margins=TRUE)
```