Week 2 Data frames

R Data Frame

R Data Frame is 2-Dimensional table like structure. In a dataframe, row represents a record while columns represent properties of the record. In this tutorial, we shall learn to Access Data of R Data Frame like selecting rows, selecting columns, selecting rows that have a given column value, etc., with Example R Scripts.

We shall look into following items to access meta information and data of an R Data Frame :

- Structure of R Data Frame
- Summary of R Data Frame
- Get Element at (i,j) ith row, jth column
- Extract column(s) of Data Frame
- Add row(s) to R Data Frame
- Add column(s) to R Data Frame
- Delete column(s) of R Data Frame

We shall use an R data frame for examples with columns : name, age, income.

Structure of R Data Frame

Following is the R function used to extract structure of an R Data Frame :

```
str(<data_frame_name>)
```

Example R Script to extract structure of an R Data Frame :

r_df_structure.R - R Script File

```
# Learn R program to extract structure of a data drame

celebrities = data.frame(name = c("Andrew", "Mathew", "Dany", "Philip", "John", "Bing", "Monica"),
                age = c(28, 23, 49, 29, 38, 23, 29),
                income = c(25.2, 10.5, 11, 21.9, 44, 11.5, 45))

# print the structure to output
cat(str(celebrities))
```

Output

```
 $ Rscript r_df_structure.R
'data.frame': 7 obs. of 3 variables:
 $ name : Factor w/ 7 levels "Andrew","Bing",..: 1 5 3 7 4 2 6
 $ age : num 28 23 49 29 38 23 29
 $ income: num 25.2 10.5 11 21.9 44 11.5 45
```

Summary of R Data Frame

Following is the R function used to extract summary of an R Data Frame

```
summary(<data_frame_name>)
```

Example R Script to extract summary of an R Data Frame :

```
r_df_summary.R - R Script File
# Learn R program to extract summary of a data frame

celebrities = data.frame(name = c("Andrew", "Mathew", "Dany", "Philip", "John", "Bing", "Monica"),
                age = c(28, 23, 49, 29, 38, 23, 29),
                income = c(25.2, 10.5, 11, 21.9, 44, 11.5, 45))

# print summary to output
cat("Summary of celebrities data frame : \n", summary(celebrities), "\n")
```

```
Output
 $ Rscript r_df_summary.R
Summary of celebrities data frame :
 Andrew:1 Bing :1 Dany :1 John :1 Mathew:1 Monica:1 Philip:1 Min. :23.00 1st Qu.:25.50 Median :29.00 Mean :31.29 3rd Qu.:3
```

Get element of R Data Frame

Following is the R function used to extract element from ith row, jth column of an R Data Frame

```
<data_frame_name>[c(<comma,separated,row,numbers>),c(<comma,separated,column,numbers>)]
```

Example R Script to extract element from ith row, jth column of an R Data Frame :

```
r_df_get_element.R - R Script File
# Learn R program to get elements of a data frame - Access Data of R Data Frame

celebrities = data.frame(name = c("Andrew", "Mathew", "Dany", "Philip", "John", "Bing", "Monica"),
                age = c(28, 23, 49, 29, 38, 23, 29),
                income = c(25.2, 10.5, 11, 21.9, 44, 11.5, 45))

# get elements from rows(2,5), columns(1,3)
elements = celebrities[c(2,5),c(1,3)]

print(elements)
```

```
Output
 $ Rscript r_df_get_element.R
    name income
2 Mathew 10.5
5 John 44.0
```

Extract column(s) of Data Frame

Following is the R function used to extract some of the columns from a R Data Frame.

```
data.frame(<data_frame_name>$<column_name_1>,<data_frame_name>$<column_name_2>)
```

You may select one or more columns from a data frame. If you are selecting multiple columns, use a comma separated list. Please observe that to select a column, we use <data_frame_name> followed by $ symbol followed by <column_name_1>.

You may write the result to a new Data Frame or overwrite the original data frame.

Example R Script to extract columns (age, income) of R Data Frame (celebrities):

r_df_extract_columns.R - R Script File

```
# Learn R program to extract column(s) of a data drame

celebrities = data.frame(name = c("Andrew", "Mathew", "Dany", "Philip", "John", "Bing", "Monica"),
                age = c(28, 23, 49, 29, 38, 23, 29),
                income = c(25.2, 10.5, 11, 21.9, 44, 11.5, 45))

# extract columns (age, income)
extractedDF = data.frame(celebrities$age, celebrities$income)

# print to output

print("New data frame with two columns extracted from celebrities : ")
print(extractedDF)
```

Output

```
 $ Rscript r_df_extract_columns.R
 [1] "New data frame with two columns extracted from celebrities : "
  celebrities.age celebrities.income
1 28 25.2
2 23 10.5
3 49 11.0
4 29 21.9
5 38 44.0
6 23 11.5
7 29 45.0
```

Add row(s) to R Data Frame

Following R function is used to add more rows to an R Data Frame :

```
resulting_data_frame =rbind(<existing_data_frame_name>,<additional_data_frame_name>)
```

Following is an Example R Script that adds three rows to the existing R data frame, celebrities :

r_df_add_row.R - R Script File

```
# Learn R program to add row(s) to a data frame - Access Data of R Data Frame

celebrities = data.frame(name = c("Andrew", "Mathew", "Dany", "Philip", "John", "Bing", "Monica"),
                age = c(28, 23, 49, 29, 38, 23, 29),
                income = c(25.2, 10.5, 11, 21.9, 44, 11.5, 45))

new_data = data.frame(name = c("Gary", "Lee", "Scofield"),
                age = c(29, 22, 33),
                income = c(21, 5, 31))

# add rows of new_data to celebrities
celebrities = rbind(celebrities, new_data)

# print to output

print("Resulting celebrities data frame with three newly added rows : ")
print(celebrities)
```

Output

```
 $ Rscript r_df_add_row.R
 [1] "Resulting celebrities data frame with three newly added rows : "
       name age income
1 Andrew 28 25.2
2 Mathew 23 10.5
3 Dany 49 11.0
4 Philip 29 21.9
5 John 38 44.0
6 Bing 23 11.5
7 Monica 29 45.0
8 Gary 29 21.0
9 Lee 22 5.0
10 Scofield 33 31.0
```

Add column(s) to R Data Frame

Following R function is used to add more columns to an R Data Frame :

```
resulting_data_frame =rbind(<existing_data_frame_name>,<new_data_frame_name>)
```

Following is an Example R Script that add a column, 'car' to the existing R data frame, celebrities :

r_df_add_column.R - R Script File

```
# Learn R program to add column(s) to R data frame - Access Data of R Data Frame

celebrities = data.frame(name = c("Andrew", "Mathew", "Dany", "Philip", "John", "Bing", "Monica"),
                age = c(28, 23, 49, 29, 38, 23, 29),
                income = c(25.2, 10.5, 11, 21.9, 44, 11.5, 45))

# add column to the celebrities data frame
celebrities$car = c("Audi","Toyota","Bugati","Audi","Agera R","Bugati","Audi")

# print to output

print("Resulting celebrities data frame with newly added column : ")
print(celebrities)
```

Output

```
 $ Rscript r_df_add_column.R
 [1] "Resulting celebrities data frame with newly added column : "
    name age income     car
1 Andrew 28 25.2 Audi
2 Mathew 23 10.5 Toyota
3 Dany 49 11.0 Bugati
4 Philip 29 21.9 Audi
5 John 38 44.0 Agera R
6 Bing 23 11.5 Bugati
7 Monica 29 45.0 Audi
```

Delete column(s) of R Data Frame

To delete a column from R Data Frame, you may select the columns you want to keep using extract columns of data frame and overwrite the existing data frame.