TMC 204
Statistical Data Analysis with R
Unit 4
Manipulating Objects Part 2

Presented By : Aditya Joshi
Asst. Professor
Department of Computer Application
Graphic Era Deemed to be University
06-04-2020

Manipulating Data Frames

```
> emp.data <- data.frame(emp_id = c (1:5), emp_name = c("aditya","santosh","kuldeep","Riya","Shubham"),salary = c(623.3,515.2,611.0,729.0,843.25), start_date = as.Date(c("2012-01-01", "2013-09-23", "2014-11-15", "2014-05-11","2015-03-27")),stringsAsFactors = FALSE)
> emp.data
  emp_id emp_name salary start_date
1      1   aditya 623.30 2012-01-01
2      2  santosh 515.20 2013-09-23
3      3  kuldeep 611.00 2014-11-15
4      4     Riya 729.00 2014-05-11
5      5  Shubham 843.25 2015-03-27
```

Get the Structure of the Data Frame

```
> str(emp.data)
'data.frame':5 obs. of  4 variables:
 $ emp_id    : int  1 2 3 4 5
 $ emp_name  : chr  "aditya" "santosh" "kuldeep" "Riya" ...
 $ salary    : num  623 515 611 729 843
 $ start_date: Date, format: "2012-01-01" ...
```

Summary of Data in Data Frame

The statistical summary and nature of the data can be obtained by applying **summary()** function.

```
> print(summary(emp.data))
    emp_id    emp_name           salary
 Min.   :1   Length:5          Min.   :515.2
 1st Qu.:2   Class :character   1st Qu.:611.0
 Median :3   Mode  :character   Median :623.3
 Mean   :3                      Mean   :664.4
 3rd Qu.:4                      3rd Qu.:729.0
 Max.   :5                      Max.   :843.2
  start_date
 Min.   :2012-01-01
 1st Qu.:2013-09-23
 Median :2014-05-11
 Mean   :2014-01-14
 3rd Qu.:2014-11-15
 Max.   :2015-03-27
```

Extract Data from Data Frame

```
> result <- data.frame(emp.data$emp_name,emp.data$salary)
> print(result)
  emp.data.emp_name emp.data.salary
1           aditya         623.30
2          santosh         515.20
3          kuldeep         611.00
4             Riya         729.00
5          Shubham         843.25
```

Extract first two rows.

```
> result <- emp.data[1:2,]
> print(result)
  emp_id emp_name salary start_date
1      1   aditya  623.3 2012-01-01
2      2  santosh  515.2 2013-09-23
```

Extract 3rd and 5th row with 2nd and 4th column

```
> result <- emp.data[c(3,5),c(2,4)]
> print(result)
  emp_name start_date
3  kuldeep 2014-11-15
5  Shubham 2015-03-27
```

Expand Data Frame

A data frame can be expanded by adding columns and rows.

Add Column

Just add the column vector using a new column name.

```
> emp.data$dept <- c("IT","Operations","IT","HR","Finance")
> print(emp.data)
  emp_id emp_name salary start_date       dept
1      1   aditya 623.30 2012-01-01         IT
2      2  santosh 515.20 2013-09-23 Operations
3      3  kuldeep 611.00 2014-11-15         IT
4      4     Riya 729.00 2014-05-11         HR
5      5  Shubham 843.25 2015-03-27    Finance
```

Add Row

To add more rows permanently to an existing data frame, we need to bring in the new rows in the same structure as the existing data frame and use the rbind() function.

Create first data frame

```
> emp.data <- data.frame(emp_id = c (1:5), emp_name =
c("adi","anand","sunil","Himanshu","simran"),salary =
c(623.3,515.2,611.0,729.0,843.25),start_date = as.Date(c("2012-01-01", "2013-09-23",
"2014-11-15", "2014-05-11", "2015-03-27")),dept
=c("IT","Operations","IT","HR","Finance"),stringsAsFactors = FALSE )

> emp.data
  emp_id emp_name salary start_date       dept
1      1      adi 623.30 2012-01-01         IT
2      2    anand 515.20 2013-09-23 Operations
3      3    sunil 611.00 2014-11-15         IT
4      4 Himanshu 729.00 2014-05-11         HR
5      5   simran 843.25 2015-03-27    Finance
```

Create second data frame

```
> emp.newdata <-  data.frame(emp_id = c (6:8), emp_name = c("Harshit","Pranjal","Ankit"),salary = c(578.0,722.5,632.8), start_date = as.Date(c("2013-05-21","2013-07-30","2014-06-17")),dept = c("IT","Operations","Fianance"),stringsAsFactors = FALSE)

> emp.newdata
  emp_id emp_name salary start_date       dept
1      6  Harshit  578.0 2013-05-21         IT
2      7  Pranjal  722.5 2013-07-30 Operations
3      8    Ankit  632.8 2014-06-17   Fianance
```

Bind final Data

```
> emp.finaldata <- rbind(emp.data,emp.newdata)
> print(emp.finaldata)
  emp_id emp_name salary start_date       dept
1      1      adi 623.30 2012-01-01         IT
2      2    anand 515.20 2013-09-23 Operations
3      3    sunil 611.00 2014-11-15         IT
4      4 Himanshu 729.00 2014-05-11         HR
5      5   simran 843.25 2015-03-27    Finance
6      6  Harshit 578.00 2013-05-21         IT
7      7  Pranjal 722.50 2013-07-30 Operations
8      8    Ankit 632.80 2014-06-17   Fianance
```