TEB2164 Introduction to Data Science
Lab 6a- Create, display, access and modify using data frame

Learning Outcomes

At the end of the session, you will be able to:

- Creating data frame and displaying values from data frame
- Accessing values in data frame and expanding row and column of data frame

Activity

1. Creating data frame
   1.1. Create the data frame
   - Write and run the following in R. Make your conclusion about the code:

```
emp.data <- data.frame(
  emp_id = c (1:5),
  emp_name = c("Rick","Dan","Michelle","Ryan","Gary"),
  salary = c(623.3,515.2,611.0,729.0,843.25)
)
```

   1.2. Creating a data frame from matrix and renaming the column
   - Write and run the following in R. Make your conclusion about the code:

```
my_matrix <- matrix(c(1, 2, 3, 4, 5, 6), nrow=2)
print(my_matrix)
df_from_matrix <- data.frame(my_matrix)
print(df_from_matrix)
names(df_from_matrix) <- c('col_1', 'col_2', 'col_3')
print(df_from_matrix)
```

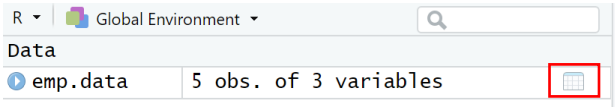
   1.3. Creating a data frame from list of vectors
   - Write and run the following in R. Make your conclusion about the code:

```
my_list <- list(rating=1:4,
                animal=c('koala', 'hedgehog', 'sloth', 'panda'),
                country=c('Australia', 'Italy', 'Peru', 'China'),
                avg_sleep_hours=c(21, 18, 17, 10))
print(my_list)
super_sleepers <- data.frame(my_list)
print(super_sleepers)
```

2. Displaying the data frame
   2.1. Display the values from the data frame
   - Write and run the following in R.

```
print(emp.data)
```

   - Or click the 'table icon' below. Identify the R code written in R Console.



   - Make your conclusion about the difference in both ways of displaying the data frame.

2.2. Display the structures of the data frame

- Write and run the following in R. Make your conclusion about the code:

```
str(emp.data)
```

2.3. Display the number of rows and columns

- Write and run the following in R. Make your conclusion about the code:

```
print(summary(emp.data))
```

2.4. Display the summary of the data frame

- Write and run the following in R. Make your conclusion about the code:

```
print(dim(emp.data))
print(ncol(emp.data))
print(nrow(emp.data))
```

3. Accessing the data frame

3.1. Extract specific column from a data frame using column name or indexing

- Write and run the following in R. Make your conclusion about the code:

```
print(emp.data$salary)
print(emp.data[["salary"]])
print(emp.data[3])
```

3.2. Extract the first two rows and then all columns

- Write and run the following in R. Make your conclusion about the code:

```
result <- emp.data[1:2,]
print(result)
```

4. Modifying the data frame rows and columns

4.1. Adding new column

- Add the column vector using a new column name or cbind(). Write and run the following in R. Make your conclusion about the code:

```
# Add the "dept" column.
emp.data$dept <- c("IT","Operations","IT","HR","Finance")
v <- emp.data
print(v)

# Create new data frame and add the "start_date" column.
emp.newdata<- cbind(emp.data,
start_date = as.Date(c("2012-01-01", "2013-09-23", "2014-11-15",
"2014-05-11","2015-03-27")))
```

4.2. Adding new 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 then use the rbind() function.

```
# Create the second data frame
emp.newdata <- data.frame(
  emp_id = c (6:8),
  emp_name = c("Rasmi","Pranab","Tusar"),
  salary = c(578.0,722.5,632.8),
  dept = c("IT","Operations","Finance")
)

# Bind the two data frames.
emp.finaldata <- rbind(emp.data,emp.newdata)
print(emp.finaldata)
```

4.3. Remove Rows and Columns

- Use the c() function to remove rows and columns in a Data Frame. Write and run the following in R. Make your conclusion about the code:

```
# Remove the second row and first column
result <- emp.data[-c(2), -c(1)]
print(result)
```

5. Extra: Create table below as data frame in R. Modify the structure of the data frame by adding more rows and columns. Challenge yourself to add various datatype (string, int, date, etc.) into the table.

friend_id	friend_name	location
1	Sachin	Kolkata
2	Sourav	Delhi
3	Dravid	Bangalore
4	Sehwag	Hyderabad
5	Dhoni	Chennai