SESSION 4 – ASSIGNMENT 4.1

Date: 7th Jan 2019

1. df1 = data.frame(CustId = c(1:6), Product = c(rep("TV", 3), rep("Radio", 3)))

   df2 = data.frame(CustId = c(2, 4, 6), State = c(rep("Texas", 2), rep("NYC", 1)))

   df1 #left table

   df2 #right table

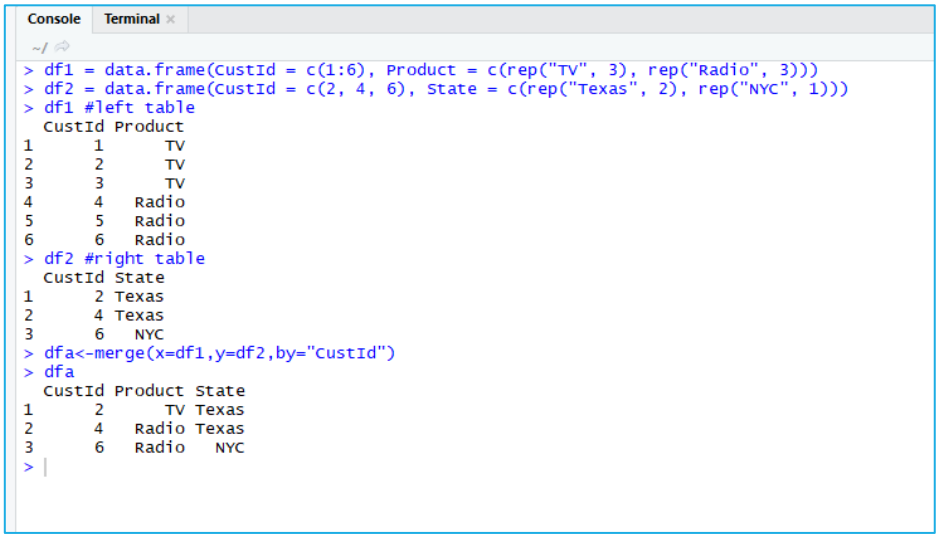
   For the above given data frames and tables perform the following operations:

   - **Return only the rows in which the left table have match.**

     #use of merge function for inner join

     dfa<-merge(x=df1,y=df2,by="CustId")

     dfa



- **Returns all rows from both tables, join records from the left which have matching keys in the right table.**

  #use of merge function for outer join

  df<-merge(x=df1,y=df2,by="CustId",all=TRUE)

  df

```
> #use of merge function
> df<-merge(x=df1,y=df2,by="CustId",all=TRUE)
> df
  CustId Product State
1      1      TV  <NA>
2      2      TV Texas
3      3      TV  <NA>
4      4   Radio Texas
5      5   Radio  <NA>
6      6   Radio   NYC
>
```

- **Return all rows from the left table, and any rows with matching keys from the right table.**

  #left join

  library(dplyr)

  dfb<-left_join(df1,df2)

  dfb

```
Console  Terminal
~/
> df1 = data.frame(CustId = c(1:6), Product = c(rep("TV", 3), rep("Radio", 3)))
> df2 = data.frame(CustId = c(2, 4, 6), State = c(rep("Texas", 2), rep("NYC", 1)))
> df1 #left table
  CustId Product
1      1      TV
2      2      TV
3      3      TV
4      4   Radio
5      5   Radio
6      6   Radio
> df2 #right table
  CustId State
1      2 Texas
2      4 Texas
3      6   NYC
> #left outer join
> library(dplyr)
> dfb<-left_join(df1,df2)
Joining, by = "CustId"
> dfb
  CustId Product State
1      1      TV  <NA>
2      2      TV Texas
3      3      TV  <NA>
4      4   Radio Texas
5      5   Radio  <NA>
6      6   Radio   NYC
>
```

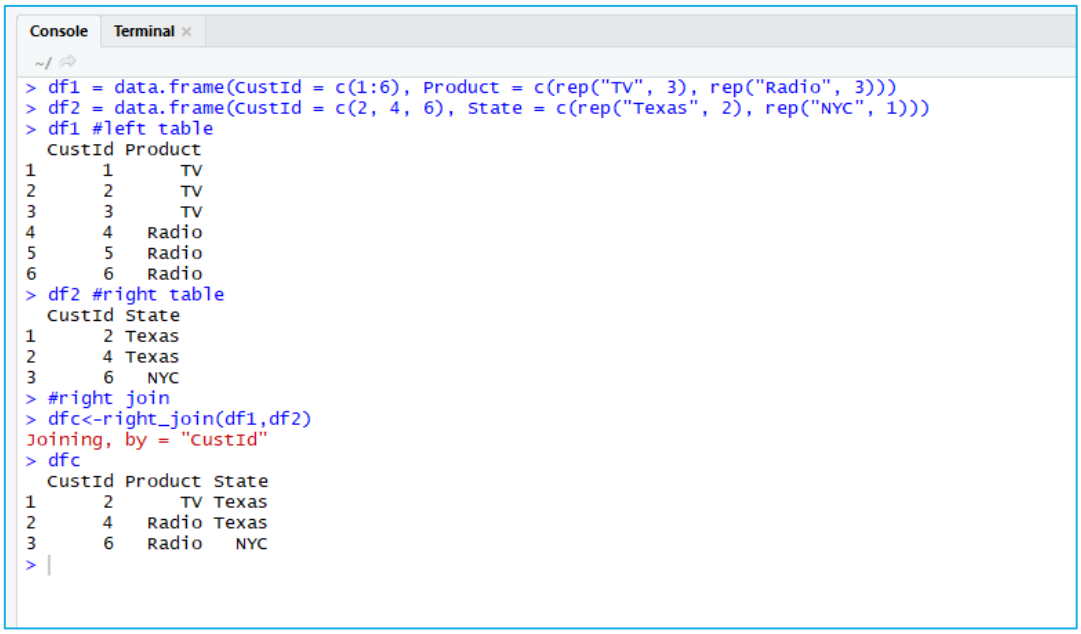
- **Return all rows from the right table, and any rows with matching keys from the left table.**

  #left join

  library(dplyr)

  dfb<-right_join(df1,df2)

  dfb



2. **Perform the below operations on above given data frames and tables:**

- **Return a long format of the datasets without matching key.**

df1 = data.frame(CustId = c(1:6), Product = c(rep("TV", 3), rep("Radio", 3)))

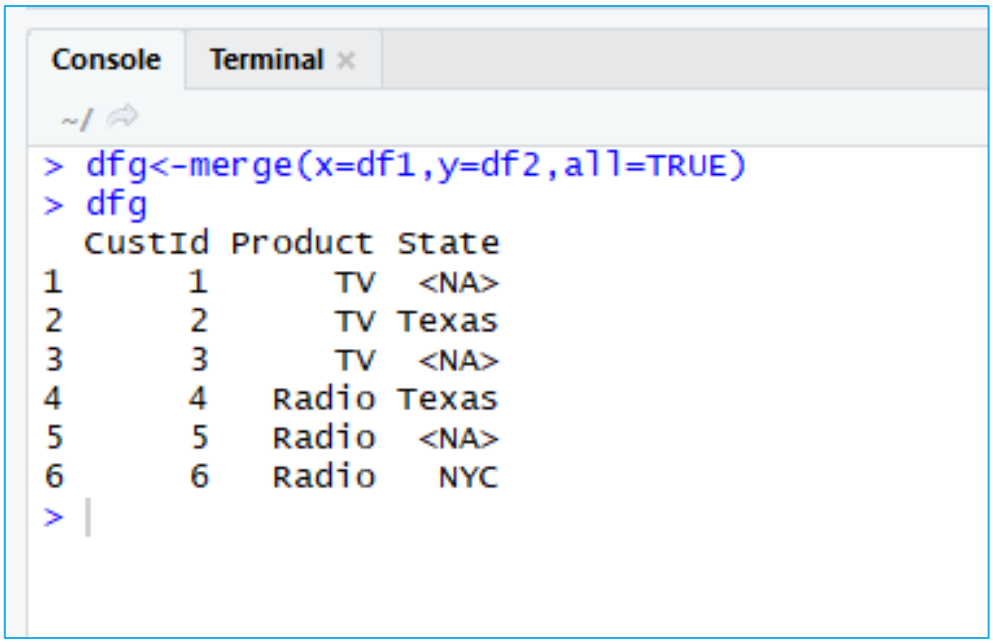
df2 = data.frame(CustId = c(2, 4, 6), State = c(rep("Texas", 2), rep("NYC", 1)))

df1 #left table

df2 #right table

dfg<-merge(x=df1,y=df2,all=TRUE)

dfg



- **Keep only observations in df1 that match in df2.**

df1 = data.frame(CustId = c(1:6), Product = c(rep("TV", 3), rep("Radio", 3)))

df2 = data.frame(CustId = c(2, 4, 6), State = c(rep("Texas", 2), rep("NYC", 1)))

df1 #left table

df2 #right table

semi_join(df1, df2)

```
Console  Terminal
~/
> df1 = data.frame(CustId = c(1:6), Product = c(rep("TV", 3), rep("Radio", 3)))
> df2 = data.frame(CustId = c(2, 4, 6), State = c(rep("Texas", 2), rep("NYC", 1)))
> df1 #left table
  CustId Product
1      1      TV
2      2      TV
3      3      TV
4      4   Radio
5      5   Radio
6      6   Radio
> df2 #right table
  CustId State
1      2 Texas
2      4 Texas
3      6   NYC
> semi_join(df1, df2)
Joining, by = "CustId"
  CustId Product
1      2      TV
2      4   Radio
3      6   Radio
>
```

- **Drop all observations in df1 that match in df2.**

df1 = data.frame(CustId = c(1:6), Product = c(rep("TV", 3), rep("Radio", 3)))

df2 = data.frame(CustId = c(2, 4, 6), State = c(rep("Texas", 2), rep("NYC", 1)))

df1 #left table

df2 #right table

anti_join(df1, df2)

```
Console  Terminal
~/
> df1 = data.frame(CustId = c(1:6), Product = c(rep("TV", 3), rep("Radio", 3)))
> df2 = data.frame(CustId = c(2, 4, 6), State = c(rep("Texas", 2), rep("NYC", 1)))
> df1 #left table
  CustId Product
1      1      TV
2      2      TV
3      3      TV
4      4   Radio
5      5   Radio
6      6   Radio
> df2 #right table
  CustId State
1      2 Texas
2      4 Texas
3      6   NYC
> anti_join(df1,df2)
Joining, by = "CustId"
  CustId Product
1      1      TV
2      3      TV
3      5   Radio
>
```