

SESSION 4: FOUNDATIONAL R PROGRAMMING-II

Assignment 1

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`

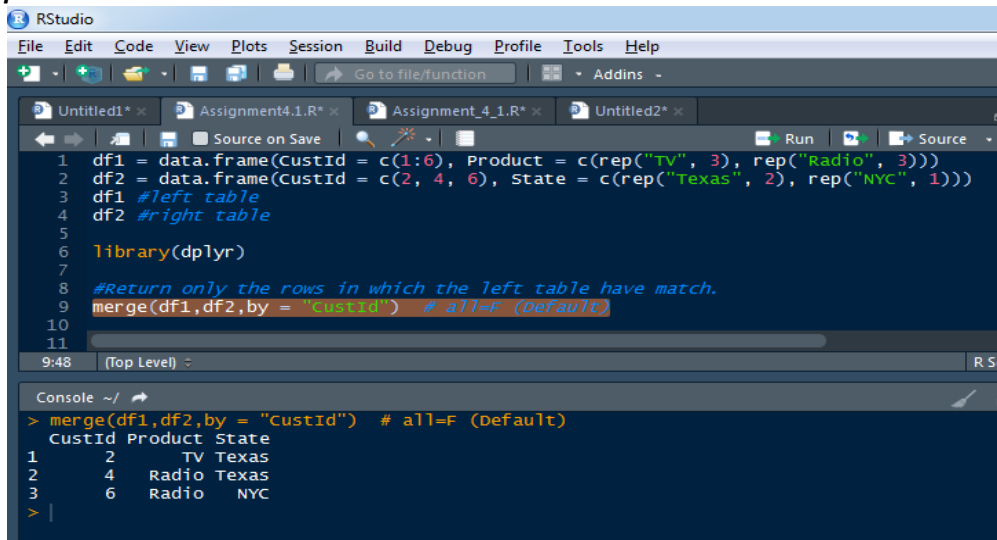
For the above given data frames and tables perform the following operations:

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

Answer:

`merge(df1,df2,by = "CustId") #all=F (Default)`

Output:



The screenshot shows the RStudio interface. The script editor contains the following code:

```
1 df1 = data.frame(CustId = c(1:6), Product = c(rep("TV", 3), rep("Radio", 3)))
2 df2 = data.frame(CustId = c(2, 4, 6), State = c(rep("Texas", 2), rep("NYC", 1)))
3 df1 #left table
4 df2 #right table
5
6 library(dplyr)
7
8 #Return only the rows in which the left table have match.
9 merge(df1,df2,by = "CustId") # all=F (Default)
10
11
```

The console output shows the result of the merge operation:

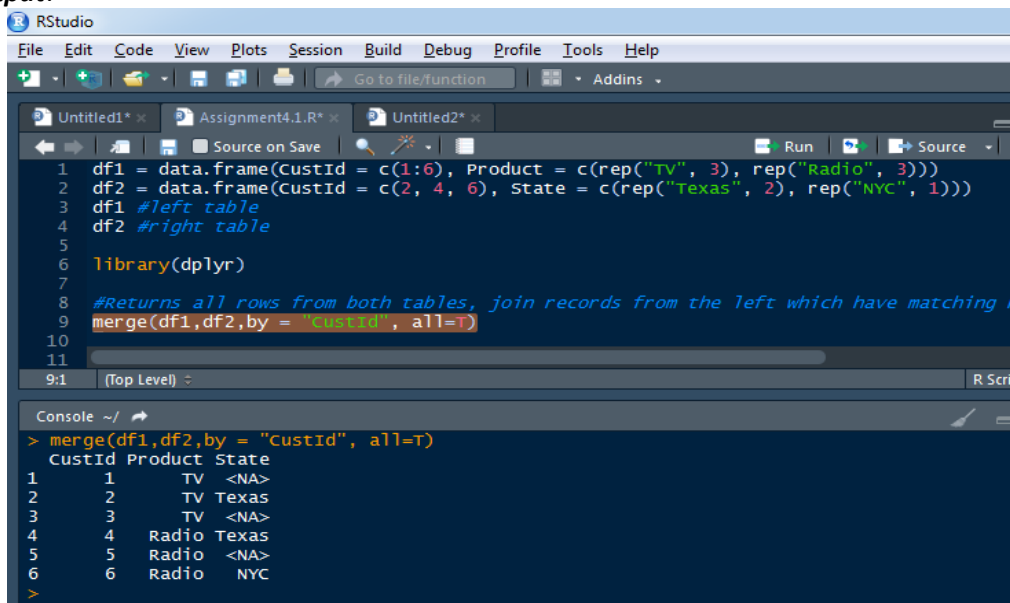
```
> merge(df1,df2,by = "CustId") # all=F (Default)
  CustId Product State
1      2      TV  Texas
2      4    Radio  Texas
3      6    Radio   NYC
```

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

Answer:

`merge(df1,df2,by = "CustId", all=T)`

Output:



The screenshot shows the RStudio interface. The script editor contains the following code:

```
1 df1 = data.frame(CustId = c(1:6), Product = c(rep("TV", 3), rep("Radio", 3)))
2 df2 = data.frame(CustId = c(2, 4, 6), State = c(rep("Texas", 2), rep("NYC", 1)))
3 df1 #left table
4 df2 #right table
5
6 library(dplyr)
7
8 #Returns all rows from both tables, join records from the left which have matching k
9 merge(df1,df2,by = "CustId", all=T)
10
11
```

The console output shows the result of the merge operation:

```
> merge(df1,df2,by = "CustId", all=T)
  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.

Answer:

`merge(df1,df2,by = "CustId", all.x=T)`

Output:

```

1 df1 = data.frame(CustId = c(1:6), Product = c(rep("TV", 3), rep("Radio", 3)))
2 df2 = data.frame(CustId = c(2, 4, 6), State = c(rep("Texas", 2), rep("NYC", 1)))
3 df1 #left table
4 df2 #right table
5
6 library(dplyr)
7
8 #Return all rows from the left table, and any rows with matching keys from the right
9 merge(df1,df2,by = "CustId", all.x=T)
10
11

```

9:1 (Top Level) R Script

```

> merge(df1,df2,by = "CustId", all.x=T)
  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.

Answer:

`merge(df1,df2,by = "CustId", all.y=T)`

Output:

```

1 df1 = data.frame(CustId = c(1:6), Product = c(rep("TV", 3), rep("Radio", 3)))
2 df2 = data.frame(CustId = c(2, 4, 6), State = c(rep("Texas", 2), rep("NYC", 1)))
3 df1 #left table
4 df2 #right table
5
6 library(dplyr)
7
8 #Return all rows from the right table, and any rows with matching keys from the left
9 merge(df1,df2,by = "CustId", all.y=T)
10
11

```

9:38 (Top Level) R Script

```

> merge(df1,df2,by = "CustId", all.y=T)
  CustId Product State
1      2      TV Texas
2      4      Radio Texas
3      6      Radio  NYC

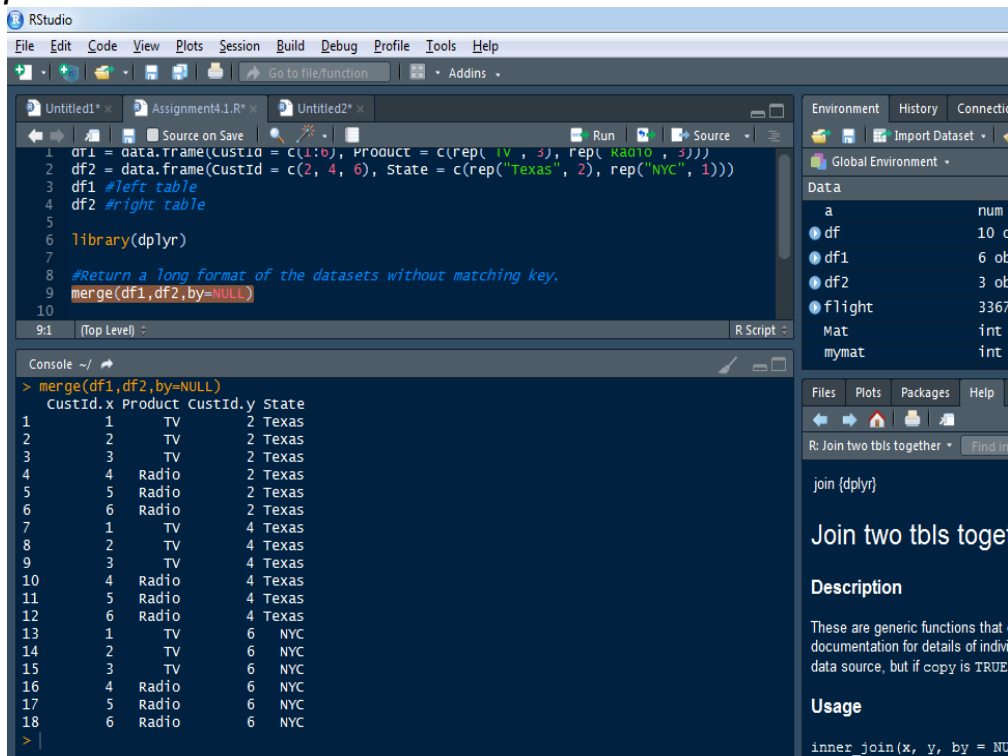
```

2. Perform the below operations on above given data frames and tables:
- Return a long format of the datasets without matching key.

Answer:

`merge(df1,df2, by=NULL)`

Output:



```
1 df1 = data.frame(CustId = c(1:6), Product = c(rep("TV", 3), rep("Radio", 3)))
2 df2 = data.frame(CustId = c(2, 4, 6), State = c(rep("Texas", 2), rep("NYC", 1)))
3 df1 #left table
4 df2 #right table
5
6 library(dplyr)
7
8 #Return a long format of the datasets without matching key.
9 merge(df1,df2,by=NULL)
10
```

Console ~/

```
> merge(df1,df2,by=NULL)
  CustId.x Product CustId.y State
1        1      TV        2  Texas
2        2      TV        2  Texas
3        3      TV        2  Texas
4        4    Radio        2  Texas
5        5    Radio        2  Texas
6        6    Radio        2  Texas
7        1      TV        4  Texas
8        2      TV        4  Texas
9        3      TV        4  Texas
10       4    Radio        4  Texas
11       5    Radio        4  Texas
12       6    Radio        4  Texas
13       1      TV        6  NYC
14       2      TV        6  NYC
15       3      TV        6  NYC
16       4    Radio        6  NYC
17       5    Radio        6  NYC
18       6    Radio        6  NYC
```

Environment History Connections

Global Environment

Data

	num
a	
df	10 o
df1	6 ob
df2	3 ob
flight	3367
Mat	int
mymat	int

Files Plots Packages Help

R: Join two tbls together

join (dplyr)

Join two tbls together

Description

These are generic functions that... documentation for details of indivi... data source, but if copy is TRUE

Usage

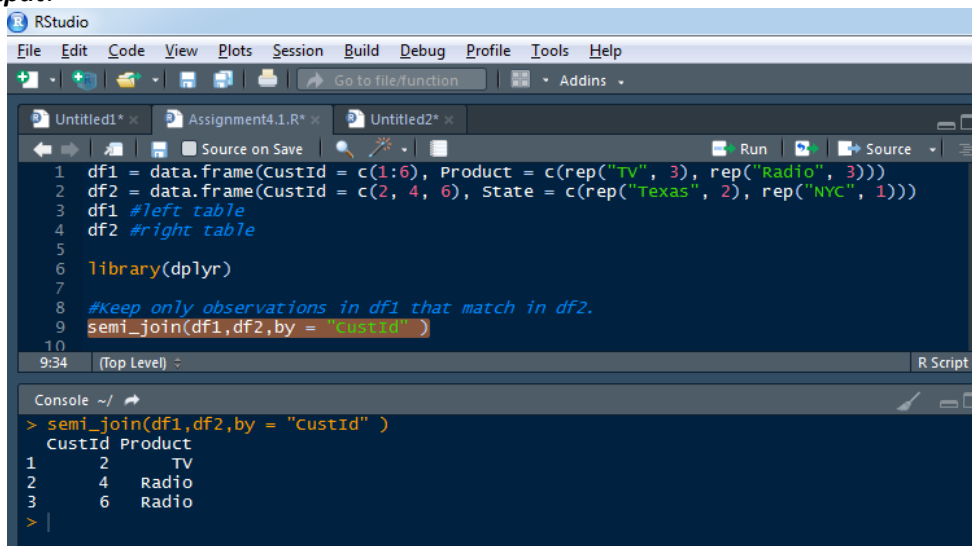
inner_join(x, y, by = NU

- Keep only observations in df1 that match in df2.

Answer:

`semi_join(df1,df2,by = "CustId")`

Output:



```
1 df1 = data.frame(CustId = c(1:6), Product = c(rep("TV", 3), rep("Radio", 3)))
2 df2 = data.frame(CustId = c(2, 4, 6), State = c(rep("Texas", 2), rep("NYC", 1)))
3 df1 #left table
4 df2 #right table
5
6 library(dplyr)
7
8 #Keep only observations in df1 that match in df2.
9 semi_join(df1,df2,by = "CustId")
10
```

Console ~/

```
> semi_join(df1,df2,by = "CustId" )
  CustId Product
1       2      TV
2       4    Radio
3       6    Radio
```

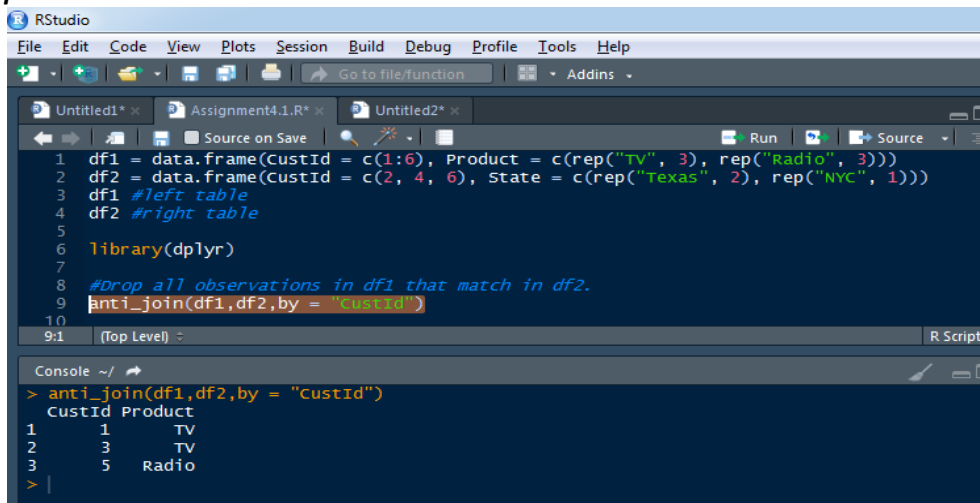
Environment History Connections

- Drop all observations in df1 that match in df2.

Answer:

`anti_join(df1,df2,by = "CustId")`

Output:



The screenshot shows the RStudio interface. The source editor contains the following R code:

```
1 df1 = data.frame(CustId = c(1:6), Product = c(rep("TV", 3), rep("Radio", 3)))
2 df2 = data.frame(CustId = c(2, 4, 6), State = c(rep("Texas", 2), rep("NYC", 1)))
3 df1 #left table
4 df2 #right table
5
6 library(dplyr)
7
8 #Drop all observations in df1 that match in df2.
9 anti_join(df1,df2,by = "CustId")
10
```

The console shows the output of the `anti_join` function:

```
> anti_join(df1,df2,by = "CustId")
  CustId Product
1      1      TV
2      3      TV
3      5     Radio
```