



ACADGILD

SESSION 4: FOUNDATIONAL R PROGRAMMING-II

Assignment 1

Table of Contents

1.Introduction	3
2.Objective.....	3
3.Prerequisites	3
4.Associated Data Files	3
5.Problem Statement	3
6.Expected Output.....	3
7.Approximate Time to Complete Task.....	3

1. Introduction

This assignment will help you understand the concepts learnt in the session.

2. Objective

This assignment will test your skills on foundational R Programming- writing functions.

3. Prerequisites

Not applicable.

4. Associated Data Files

Not applicable.

5. Problem Statement

1.

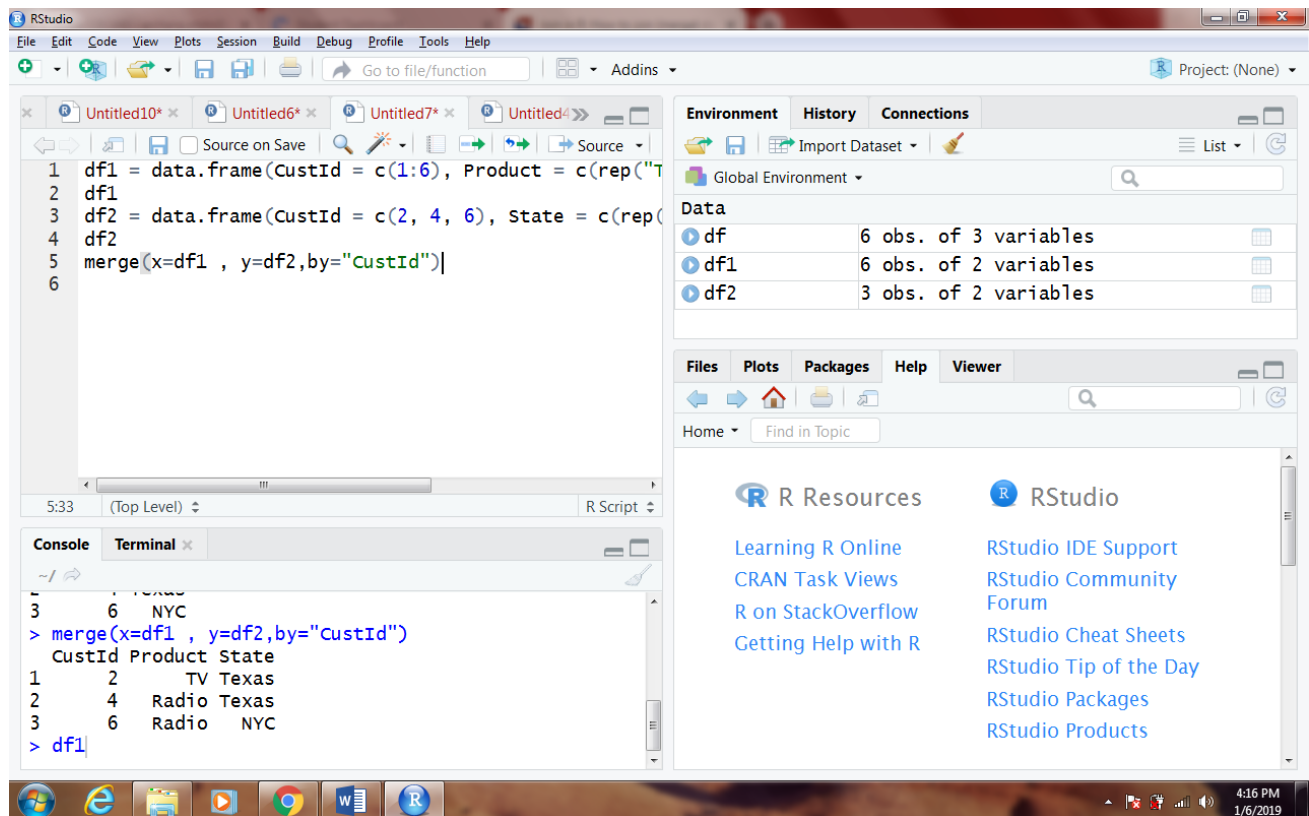
```
df1 = data.frame(CustId = c(1:6), Product = c(rep("TV", 3), rep("Radio", 3)))  
df2 = data.frame(CustId = c(2, 4, 6), State = c(rep("Texas", 2), rep("NYC", 1)))  
df1 #left table  
df2 #right table
```

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

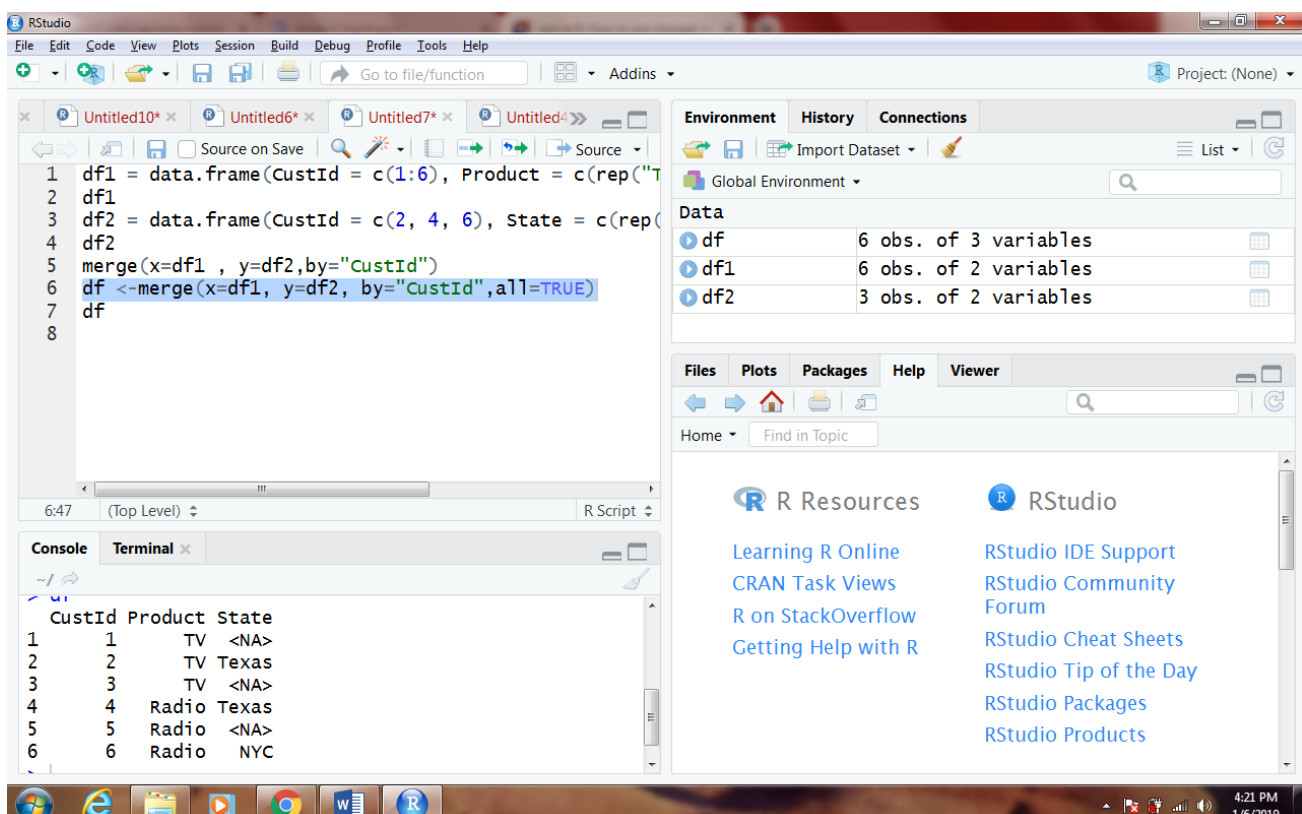
- Return only the rows in which the left table have match.
- Returns all rows from both tables, join records from the left which have matching keys in the right table.
- Return all rows from the left table, and any rows with matching keys from the right table.
- Return all rows from the right table, and any rows with matching keys from the left table.

1.

```
df1 = data.frame(CustId = c(1:6), Product = c(rep("TV", 3), rep("Radio", 3)))  
df1  
df2 = data.frame(CustId = c(2, 4, 6), State = c(rep("Texas", 2), rep("NYC", 1)))  
df2  
merge(x=df1 , y=df2,by="CustId")
```



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



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

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 df1
3 df2 = data.frame(CustId = c(2, 4, 6), State = c(rep("Texas", 2), rep("NYC", 1)))
4 df2
5 merge(x=df1 , y=df2,by="CustId")
6 df <-merge(x=df1, y=df2, by="CustId",all=TRUE)
7 df
8 df <-merge(x=df1 ,y=df2, by="CustId",all.x=TRUE)
9 df
10
```

The console output shows the result of the merge operation:

```
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
```

The Environment pane shows the following data objects:

Object	Dimensions
df	6 obs. of 3 variables
df1	6 obs. of 2 variables
df2	3 obs. of 2 variables

3.

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


df

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 df1
3 df2 = data.frame(CustId = c(2, 4, 6), State = c(rep("Texas", 2), rep("NYC", 1)))
4 df2
5 merge(x=df1 , y=df2,by="CustId")
6 df <-merge(x=df1, y=df2, by="CustId",all=TRUE)
7 df
8 df <-merge(x=df1 ,y=df2, by="CustId",all.x=TRUE)
9 df
10 df <-merge(x=df1 ,y=df2 ,by="CustId",all.y=TRUE)
11 df
12
```

The console output shows the result of the merge operation:

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

The Environment pane shows the following data objects:

Object	Dimensions
df	3 obs. of 3 variables
df1	6 obs. of 2 variables
df2	3 obs. of 2 variables

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

- Return a long format of the datasets without matching key.
- Keep only observations in df1 that match in df2.
- Drop all observations in df1 that match in df2.

```
df <- merge(x=df1, y=df2, by=NULL)
```

```
df
```

The screenshot shows the RStudio IDE interface. The script editor on the left contains the following R code:

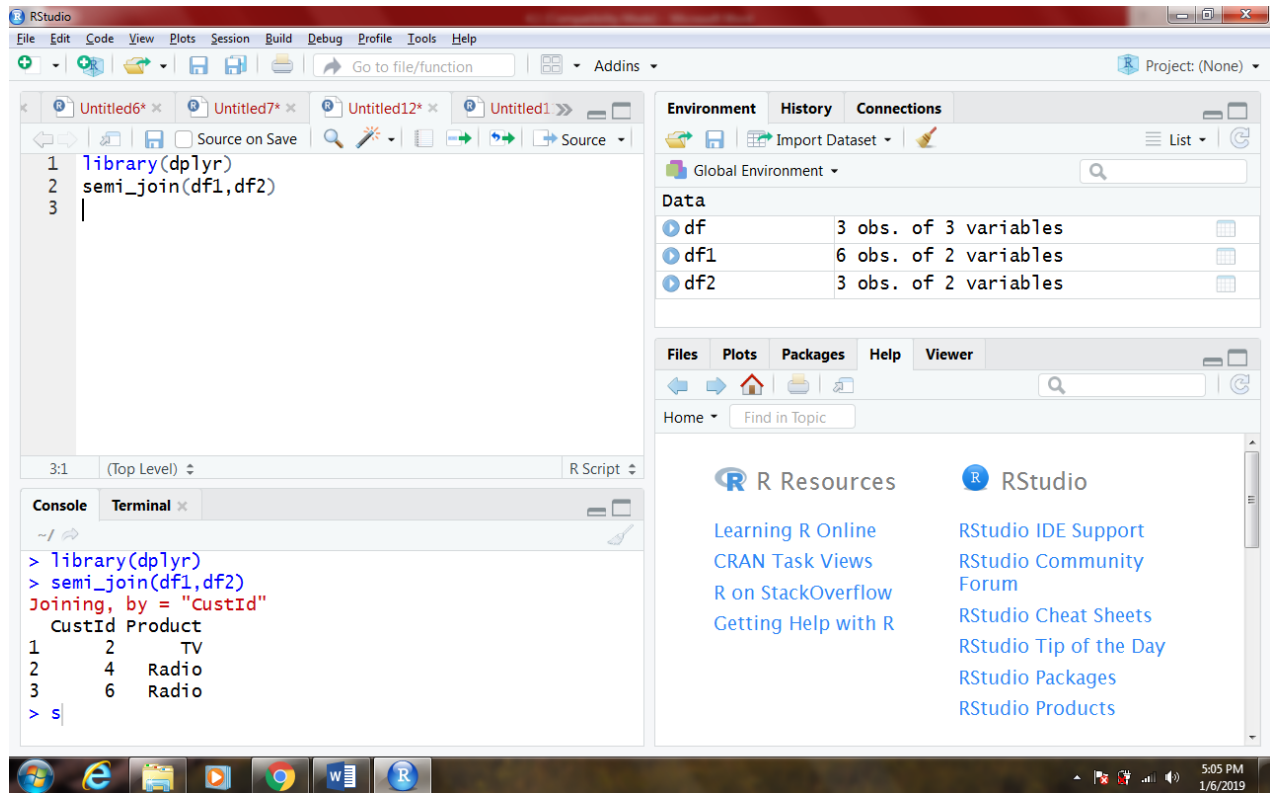
```
1 df <- merge(x=df1, y=df2, by=NULL)
2 df
```

The console on the bottom left displays the output of the code, showing a data frame with 18 rows and 4 columns: CustId.x, Product, CustId.y, and State. The data is as follows:

	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

The Environment pane on the right shows the Global Environment with three data frames: df (18 obs. of 4 variables), df1 (6 obs. of 2 variables), and df2 (3 obs. of 2 variables). The bottom right pane shows the RStudio help page with links to R Resources and RStudio support.

```
library(dplyr)
semi_join(df1,df2)
```



```
anti_join(df1,df2)
```

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins Project: (None)

Untitled6* x Untitled7* x Untitled12* x Untitled1 >>

```
1 library(dplyr)
2 semi_join(df1,df2)
3 anti_join(df1,df2)
4
```

3:1 (Top Level) R Script

Console Terminal

```
3 6 Radio
> anti_join(df1,df2)
Joining, by = "CustId"
  CustId Product
1      1      TV
2      3      TV
3      5      Radio
> s
```

Environment History Connections

Global Environment

Data

df	3 obs. of 3 variables
df1	6 obs. of 2 variables
df2	3 obs. of 2 variables

Files Plots Packages Help Viewer

Home Find in Topic

R Resources RStudio

- Learning R Online
- CRAN Task Views
- R on StackOverflow
- Getting Help with R
- RStudio IDE Support
- RStudio Community Forum
- RStudio Cheat Sheets
- RStudio Tip of the Day
- RStudio Packages
- RStudio Products

5:06 PM 1/6/2019

Data Analytics

6. Expected Format

1. R file should be submitted where applicable.
2. R file should be in PDF or in .r format
3. Proper screenshots of the outputs should be submitted as well
4. The r codes, if submitted in any other format, will be subjected to deduction in marks

Note: Your solution will not be entertained if it is any other format, e.g., .zip, .doc, .rtf etc.

7. Approximate Time to Complete Task

20 mins.