Assignment 4.1

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.

• 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.

R file and output screen shot attached. (df_functions.R and joins.png)

```
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

inner_join <- merge(df1, df2, by="CustId")
print(inner_join)

outer_join <- merge(df1, df2, by="CustId", all=TRUE)
print(outer_join)

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

print(left_outer_join)

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

print(right_outer_join)

2. 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**

R file and output screen shot attached. (using_dplyr.R and dplyr_screenshot.png)

```
library(dplyr)
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

#bindrows
bind_value <- bind_rows(df1,df2)
print(bind_value)



#Keep only observations in df1 that match in df2
e <- intersect(df1$CustId,df2$CustId)
#print(e)
matching <- df1[df1$CustId %in% e,]
print(matching)

#Drop all observations in df1 that match in df2
```

```
c <- setdiff(df1$CustId,df2$CustId)
#print(c)
nonmatching <- df1[df1$CustId %in% c,]
print(nonmatching)
```