Assignment4.1

Aruna

24 November 2018

```
#Assignment 4.1

# 1A. Return only the rows in which the left table have match
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

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) #keep only observations in df1 that match in df2.

## Joining, by = "CustID"

##   CustID Product
## 1      2      TV
## 2      4   Radio
## 3      6   Radio
```

```r
#1.B. Returns all rows from both tables, join records from the left
#   which have a matching Key in the right tables

full_join(df1,df2)
```

```
## Joining, by = "CustID"
```

```
##   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
```

```r
#1.C. Returns all rows from left tables, and any rows with matching key
#   from the right table
#left outer
left_join(df1, df2)
```

```
## Joining, by = "CustID"
```

```
##   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
```

```r
#1.D. Returns all rows from  right tables, and any rows with matching key
#   from the left table
#RIght outer
right_join(df1, df2)
```

```
## Joining, by = "CustID"
```

```
##   CustID Product State
## 1      2      TV Texas
## 2      4   Radio Texas
## 3      6   Radio   NYC
```

```r
# 2. Perform the below operations on above given data frames and tables
#2.A Return a long format  of the datasets without matching key

merge(x = df1, y = df2, by = "CustID", all.x= TRUE)
```

```
##   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

#2.B Keep only observation in df1 that match in df2
semi_join(df1, df2) #keep only observations in df1 that match in df2.

## Joining, by = "CustID"

##   CustID Product
## 1      2      TV
## 2      4   Radio
## 3      6   Radio

#2.C Drop all observation in df1 that match in df2
anti_join(df1, df2) #drops all observations in df1 that match in df2.

## Joining, by = "CustID"

##   CustID Product
## 1      1      TV
## 2      3      TV
## 3      5   Radio
```

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.