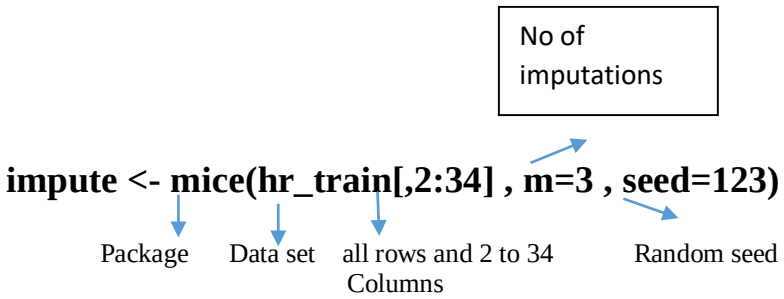
IMPUTING THE MISSING VALUE

- A simplified approach to impute missing data is with MICE package
- The mice package in R, helps you imputing missing values with probable data values. These probable values are drawn from a distribution specifically designed for each missing data point.
- In this HR analytics train data set using summary () we will try to get the detailed summary of the data set.
- From this we get to know the variable Nursing Room has 1712 missing values and we want to impute that missing values as we cannot delete the entire column.
- Using the mice package it provides the function md.pattern() which helps to understand the missing pattern.
- In this process VIM package is also used to visualize the missing data in that variable.
- The function md.pairs() gives the detailed view of all data points in four format.
  $rr-It gives totally how many data points are observed in each variables
  $rm- It gives the table of observed and missing values from variables
  $mr- It gives the table if missing and observed values from variables
  $mm- It gives the table if missing and missing values from variables
- Now using the mice() function which takes care of the imputing function in the Nursing room variable.




```
iter imp variable
  1   1  Nursing_room
  1   2  Nursing_room
  1   3  Nursing_room
  2   1  Nursing_room
  2   2  Nursing_room
  2   3  Nursing_room
  3   1  Nursing_room
  3   2  Nursing_room
  3   3  Nursing_room
  4   1  Nursing_room
  4   2  Nursing_room
  4   3  Nursing_room
  5   1  Nursing_room
  5   2  Nursing_room
  5   3  Nursing_room
```

When we run the above code we get 5 iterations and for each iterations we get 3 imputations as mentioned in the code.

print(impute)

```
Class: mids
Number of multiple imputations:  3
Imputation methods:

   Age                         Gender                Marital_status
    ""                           ""                            ""
Nursing_room
"pmm"
```

From the above we can see there is no method is used for remaining variable except the Nursing Room which has NA value so “pmm”-predictive mean matching method has been used to predict the NA values in that variable.

imputeimpNursing_room

This gives 3 imputed values of Nursing Room values from this we can take any one imputed value for further work.

```
   1 2 3
1    1 1 1
2    1 1 1
3    1 1 1
4    1 1 1
6    1 1 1
8    1 1 1
9    1 1 1
11   1 1 0
12   0 0 0
14   0 0 0
16   1 1 1
17   1 1 1
18   1 1 1
19   1 1 1
20   1 1 1
21   1 1 1
22   1 1 1
24   1 1 1
25   1 1 1
27   1 1 1
28   0 0 1
29   1 1 1
30   0 0 0
```