```
> library(xlsx)
> #a. Read the file in Zip format and get it into R
> zipF<- "C:\\Users\\ra306656\\Documents\\AirQualityUCI.zip"
> outDir<-"D:\\Rachit\\Acadgild\\ASSIGNMENTS\\Assignment10.1"
> unzip(zipF,exdir=outDir)
> #file saved in above outDir location.
> data <- read.xlsx("AirQualityUCI.xlsx",1)
> head(data)
        Date                 Time CO.GT. PT08.S1.CO. NMHC.GT.   C6H6.GT.
1 2004-03-10 1899-12-30 18:00:00    2.6     1360.00      150 11.881723
2 2004-03-10 1899-12-30 19:00:00    2.0     1292.25      112  9.397165
3 2004-03-10 1899-12-30 20:00:00    2.2     1402.00       88  8.997817
4 2004-03-10 1899-12-30 21:00:00    2.2     1375.50       80  9.228796
5 2004-03-10 1899-12-30 22:00:00    1.6     1272.25       51  6.518224
6 2004-03-10 1899-12-30 23:00:00    1.2     1197.00       38  4.741012
  PT08.S2.NMHC. NOx.GT. PT08.S3.NOx. NO2.GT. PT08.S4.NO2. PT08.S5.O3.      T
1       1045.50     166      1056.25     113      1692.00     1267.50 13.600
2        954.75     103      1173.75      92      1558.75      972.25 13.300
3        939.25     131      1140.00     114      1554.50     1074.00 11.900
4        948.25     172      1092.00     122      1583.75     1203.25 11.000
5        835.50     131      1205.00     116      1490.00     1110.00 11.150
6        750.25      89      1336.50      96      1393.00      949.25 11.175
      RH         AH NA. NA..1
1 48.875 0.7577538  NA    NA
2 47.700 0.7254874  NA    NA
3 53.975 0.7502391  NA    NA
4 60.000 0.7867125  NA    NA
5 59.575 0.7887942  NA    NA
6 59.175 0.7847717  NA    NA
> #c. Check for missing values in all columns.
> #Missing values are given value -200. Will replace them with NA
> library(naniar)
Warning message:
package 'naniar' was built under R version 3.5.2
> data <- replace_with_na_all(data, condition = ~.x == -200)
> #Visualise missing values
> vis_miss(data)
> #remove columns NMHC.GT. as it as 90 % missing values
> data <- data[-5]
> #remove columns NA. and NA..1 as they have 100% missing values.
> data <- data[-c(15,16)]
> #removing rows 9358 till end as all are NA.
> data <- data[-c(9358:9471),]
> #visualise again
> vis_miss(data)
> # Impute the missing values using appropriate methods
> library(mice)
Loading required package: lattice

Attaching package: 'mice'

The following objects are masked from 'package:base':

    cbind, rbind

> md.pattern(data)
     Date Time PT08.S1.CO. C6H6.GT. PT08.S2.NMHC. PT08.S3.NOx. PT08.S4.NO2.
6941    1    1           1        1             1            1            1
452     1    1           1        1             1            1            1
3       1    1           1        1             1            1            1
400     1    1           1        1             1            1            1
1195    1    1           1        1             1            1            1
317     1    1           0        0             0            0            0
```

```
5       1    1           0        0             0            0            0
13      1    1           0        0             0            0            0
31      1    1           0        0             0            0            0
        0    0         366      366           366          366          366
      PT08.S5.O3.    T  RH  AH NOx.GT. NO2.GT. CO.GT.
6941            1    1   1   1       1       1      1    0
452             1    1   1   1       1       1      0    1
3               1    1   1   1       1       0      1    1
400             1    1   1   1       0       0      1    2
1195            1    1   1   1       0       0      0    3
317             0    0   0   0       1       1      1    9
5               0    0   0   0       1       1      0   10
13              0    0   0   0       0       0      1   11
31              0    0   0   0       0       0      0   12
              366  366 366 366    1639    1642   1683 8258
> str(data)
Classes ‘tbl_df’, ‘tbl’ and 'data.frame':    9357 obs. of  14 variables:
 $ Date         : Date, format: "2004-03-10" "2004-03-10" ...
 $ Time         : POSIXct, format: "1899-12-30 23:21:10" "1899-12-31 00:21:10
" ...
 $ CO.GT.       : num  2.6 2 2.2 2.2 1.6 1.2 1.2 1 0.9 0.6 ...
 $ PT08.S1.CO.  : num  1360 1292 1402 1376 1272 ...
 $ C6H6.GT.     : num  11.88 9.4 9 9.23 6.52 ...
 $ PT08.S2.NMHC.: num  1046 955 939 948 836 ...
 $ NOx.GT.      : num  166 103 131 172 131 89 62 62 45 NA ...
 $ PT08.S3.NOx. : num  1056 1174 1140 1092 1205 ...
 $ NO2.GT.      : num  113 92 114 122 116 96 77 76 60 NA ...
 $ PT08.S4.NO2. : num  1692 1559 1554 1584 1490 ...
 $ PT08.S5.O3.  : num  1268 972 1074 1203 1110 ...
 $ T            : num  13.6 13.3 11.9 11 11.2 ...
 $ RH           : num  48.9 47.7 54 60 59.6 ...
 $ AH           : num  0.758 0.725 0.75 0.787 0.789 ...
```

```
#running the mice function
> data_temp <- mice(data,m=5,maxit=20,seed=500,method = 'cart')

 iter imp variable
  1   1  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  1   2  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  1   3  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  1   4  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  1   5  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  2   1  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  2   2  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  2   3  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  2   4  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  2   5  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  3   1  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
```

```
  3   2  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  3   3  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  3   4  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  3   5  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  4   1  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  4   2  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  4   3  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  4   4  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  4   5  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  5   1  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  5   2  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  5   3  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  5   4  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  5   5  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  6   1  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  6   2  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  6   3  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  6   4  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  6   5  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  7   1  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  7   2  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  7   3  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  7   4  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  7   5  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  8   1  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  8   2  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  8   3  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  8   4  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  8   5  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  9   1  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  9   2  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  9   3  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
```

```
  9   4  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  9   5  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO2
.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  10   1  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  10   2  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  10   3  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  10   4  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  10   5  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  11   1  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  11   2  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  11   3  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  11   4  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  11   5  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  12   1  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  12   2  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  12   3  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  12   4  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  12   5  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  13   1  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  13   2  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  13   3  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  13   4  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  13   5  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  14   1  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  14   2  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  14   3  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  14   4  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  14   5  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  15   1  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  15   2  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  15   3  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  15   4  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  15   5  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
```

```
  16   1  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  16   2  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  16   3  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  16   4  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  16   5  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  17   1  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  17   2  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  17   3  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  17   4  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  17   5  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  18   1  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  18   2  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  18   3  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  18   4  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  18   5  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  19   1  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  19   2  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  19   3  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  19   4  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  19   5  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  20   1  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  20   2  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  20   3  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  20   4  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
  20   5  CO.GT.  PT08.S1.CO.  C6H6.GT.  PT08.S2.NMHC.  NOx.GT.  PT08.S3.NOx.  NO
2.GT.  PT08.S4.NO2.  PT08.S5.O3.  T  RH  AH
> data_complete <- complete(data_temp,1)
> md.pattern(data_complete)
 /\     /\
{  `---'  }
{  O   O  }
==>  V <==  No need for mice. This data set is completely observed.
 \  \|/  /
  `-----'

     Date Time CO.GT. PT08.S1.CO. C6H6.GT. PT08.S2.NMHC. NOx.GT.
9357    1    1      1           1        1             1       1
        0    0      0           0        0             0       0
     PT08.S3.NOx. NO2.GT. PT08.S4.NO2. PT08.S5.O3. T RH AH
9357            1       1           1           1 1  1  1 0
```

```
                0       0            0           0 0  0  0 0

 >

# Create bi-variate analysis for all relationships
> #removing first 2 columns for bi- variate analysis
> data1 <- data_complete[-c(1,2)]
> library(corrplot)
corrplot 0.84 loaded
Warning message:
package 'corrplot' was built under R version 3.5.2
> corrplot(cor(data1))
> cor.test(data1$PT08.S1.CO.,data1$PT08.S2.NMHC.)

        Pearson's product-moment correlation

data:  data1$PT08.S1.CO. and data1$PT08.S2.NMHC.
t = 186.2, df = 9355, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.8830316 0.8916456
sample estimates:
      cor
0.8874161

> str(data_complete)
'data.frame':  9357 obs. of  14 variables:
 $ Date         : Date, format: "2004-03-10" "2004-03-10" ...
 $ Time         : POSIXct, format: "1899-12-30 23:21:10" "1899-12-31 00:21:10
" ...
 $ CO.GT.       : num  2.6 2 2.2 2.2 1.6 1.2 1.2 1 0.9 0.6 ...
 $ PT08.S1.CO.  : num  1360 1292 1402 1376 1272 ...
 $ C6H6.GT.     : num  11.88 9.4 9 9.23 6.52 ...
 $ PT08.S2.NMHC.: num  1046 955 939 948 836 ...
 $ NOx.GT.      : num  166 103 131 172 131 89 62 62 45 25 ...
 $ PT08.S3.NOx. : num  1056 1174 1140 1092 1205 ...
 $ NO2.GT.      : num  113 92 114 122 116 96 77 76 60 33 ...
 $ PT08.S4.NO2. : num  1692 1559 1554 1584 1490 ...
 $ PT08.S5.O3.  : num  1268 972 1074 1203 1110 ...
 $ T            : num  13.6 13.3 11.9 11 11.2 ...
 $ RH           : num  48.9 47.7 54 60 59.6 ...
 $ AH           : num  0.758 0.725 0.75 0.787 0.789 ...
> data_complete_derived <- data_complete
> data_complete_derived$Year<- format(as.Date(data_complete_derived$Date), "%
Y")
> data_complete_derived$Month<- format(as.Date(data_complete_derived$Date), "
%m")
> #cross tabulating to show records in each month.
> table(data_complete_derived$Month)

  01   02   03   04   05   06   07   08   09   10   11   12
 744  672 1254  807  744  720  744  744  720  744  720  744
> #cross tabulating to show records in each year.
> table(data_complete_derived$Year)

2004 2005
7110 2247
> #check for trends and patterns in time series
> #COGT columm
> data3 <-data_complete[,c(1,3)]
```

```
> head(data3)
        Date CO.GT.
1 2004-03-10    2.6
2 2004-03-10    2.0
3 2004-03-10    2.2
4 2004-03-10    2.2
5 2004-03-10    1.6
6 2004-03-10    1.2
> #CONVERTING COGT Values to time series
> COGT <- aggregate(CO.GT. ~ Date, data3, mean)
> COGT1 <-ts(COGT$CO.GT.,start = c(2004,3,10),frequency = 365)
> plot(COGT1)
> # PT08.S1.CO.columm
> data4 <-data_complete[,c(1,4)]
> head(data4)
        Date PT08.S1.CO.
1 2004-03-10     1360.00
2 2004-03-10     1292.25
3 2004-03-10     1402.00
4 2004-03-10     1375.50
5 2004-03-10     1272.25
6 2004-03-10     1197.00
> #computing mean value for each day/date
> PT08.S1.CO <- aggregate(PT08.S1.CO. ~ Date, data4, mean)
> #converting to time series
> PT08.S1.CO1 <-ts(PT08.S1.CO$PT08.S1.CO.,start = c(2004,3,10),frequency = 36
5)
> #plotting
> plot(PT08.S1.CO1)
> # C6H6.GT.columm
> data5 <-data_complete[,c(1,5)]
> head(data5)
        Date  C6H6.GT.
1 2004-03-10 11.881723
2 2004-03-10  9.397165
3 2004-03-10  8.997817
4 2004-03-10  9.228796
5 2004-03-10  6.518224
6 2004-03-10  4.741012
> #computing mean value for each day/date
> C6H6.GT <- aggregate(C6H6.GT. ~ Date, data5, mean)
> #converting to time series
> C6H6.GT1 <-ts(C6H6.GT$C6H6.GT.,start = c(2004,3,10),frequency = 365)
> #plotting
> plot(C6H6.GT1)
> #PT08.S2.NMHC. columm
> data6 <-data_complete[,c(1,6)]
> head(data6)
        Date PT08.S2.NMHC.
1 2004-03-10       1045.50
2 2004-03-10        954.75
3 2004-03-10        939.25
4 2004-03-10        948.25
5 2004-03-10        835.50
6 2004-03-10        750.25
> #computing mean value for each day/date
> PT08.S2.NMHC <- aggregate(PT08.S2.NMHC. ~ Date, data6, mean)
> #converting to time series
> PT08.S2.NMHC1 <-ts(PT08.S2.NMHC$PT08.S2.NMHC.,start = c(2004,3,10),frequenc
y = 365)
> #plotting
> plot(PT08.S2.NMHC1)
>
```

```
>
> #Hence in this manner we can plot for each of the columns/gases
> #Hence in this manner we can plot for each of the columns/gases
> #and identify any seasonal patterns or variance.
> #Hence in this manner we can plot for each of the columns/gases
> #and identify any seasonal patterns or variance.
```