## PART B

## **ASSIGNMENT 2**

```
> airquality
    Ozone Solar.R Wind Temp Month Day
               190
1234567
       41
                    7.4
                           67
                                   5555555555
                                       1
2
3
       36
               118
                    8.0
                           72
       12
               149 12.6
                           74
                                       4
       18
               313 11.5
                           62
                NA 14.3
                           56
                                       5
6
7
       NA
                NA 14.9
       28
                           66
       23
               299
                   8.6
                           65
8
                                       8
       19
                99 13.8
                           59
9
        8
                19 20.1
                           61
                                       9
10
                                      10
       NA
               194
                   8.6
                           69
 summary(airquality)
     Ozone
                       Solar.R
                                          Wind
                                                             Temp
                           : 7.0
                                                               :56.00
 Min.
            1.00
                   Min.
                                     Min.
                                            : 1.700
                                                       Min.
                                     1st Qu.: 7.400
Median : 9.700
                   1st Qu.:115.8
 1st Qu.: 18.00
                                                       1st Qu.:72.00
 Median : 31.50
                   Median:205.0
                                                       Median :79.00
                          :185.9
 Mean
        : 42.13
                   Mean
                                     Mean
                                            : 9.958
                                                       Mean
                                                               :77.88
                                                        3rd Qu.:85.00
 3rd Qu.: 63.25
                   3rd Qu.:258.8
                                     3rd Qu.:11.500
        :168.00
                   Max.
                          :334.0
                                     Max.
                                            :20.700
                                                       Max.
                                                               :97.00
 Max.
 NA's
        :37
                   NA's
                           : 7
                       Day
     Month
                          : 1.0
 Min.
        :5.000
                  Min.
 1st Qu.:6.000
                  1st Qu.: 8.0
 Median :7.000
                  Median:16.0
 Mean
        :6.993
                  Mean
                          :15.8
 3rd Qu.:8.000
                  3rd Qu.:23.0
 Max.
        :9.000
                  Max.
                          :31.0
                                Data Cleaning
> airquality$0zone=ifelse(is.na(airquality$0zone),median(airquality$0zone,na.
rm=TRUE), airquality$0zone)
> summary(airquality)
                       Solar.R
     Ozone
                                          Wind
                                                             Temp
           1.00
                                                               :56.00
                          : 7.0
                                            : 1.700
 Min.
                                     Min.
                                                       Min.
                   Min.
                                     1st Qu.: 7.400
                                                        1st Qu.:72.00
 1st Qu.: 21.00
                   1st Qu.:115.8
Median : 31.50
Mean : 39.56
                   Median :205.0
                                     Median: 9.700
                                                       Median :79.00
                           :185.9
                                             : 9.958
                                                               :77.88
                   Mean
                                     Mean
                                                       Mean
 3rd Qu.: 46.00
                                     3rd Qu.:11.500
                   3rd Qu.:258.8
                                                        3rd Qu.:85.00
        :168.00
                   Max.
                          :334.0
                                            :20.700
                                                               :97.00
 Max.
                                     Max.
                                                       Max.
                   NA's
                           :7
     Month
                       Day
        :5.000
                  Min.
 1st Qu.:6.000
                  1st Qu.: 8.0
 Median :7.000
                  Median:16.0
        :6.993
                        :15.8
 Mean
                  Mean
 3rd Qu.:8.000
                  3rd Qu.:23.0
 Max.
        :9.000
                  Max.
                          :31.0
> airquality$Solar.R=ifelse(is.na(airquality$Solar.R),median(airquality$Solar
.R,na.rm=TRUE),airquality$Solar.R)
> summary(airquality)
                       Solar.R
     Ozone
                                          Wind
                                                             Temp
```

: 1.700

Min.

:56.00

Min.

: 7.0

Min.

Min.

: 1.00

```
1st Qu.: 21.00
                        1st Qu.:120.0
                                              1st Qu.: 7.400
                                                                      1st Qu.:72.00
                                                                     Median :79.00
 Median : 31.50
                        Median :205.0
                                              Median : 9.700
           : 39.56
                                  :186.8
                                                        : 9.958
 Mean
                        Mean
                                              Mean
                                                                     Mean
                                                                               :77.88
 3rd Qu.: 46.00
                        3rd Qu.:256.0
                                              3rd Qu.:11.500
                                                                      3rd Qu.:85.00
           :168.00
                                                                               :97.00
 Max.
                        Max.
                                  :334.0
                                              Max.
                                                        :20.700
                                                                     Max.
      Month
                              Day
           :5.000
                       Min.
 Min.
                       1st Qu.: 8.0
 1st Qu.:6.000
                       Median:16.0
 Median :7.000
                              :15.8
           :6.993
 Mean
                       Mean
 3rd Qu.:8.000
                       3rd Qu.:23.0
 Max.
           :9.000
                       Max.
                                :31.0
> head(airquality)
   Ozone Solar.R Wind Temp Month Day
                190
    41.0
                       7.4
                               67
                                              2
    36.0
                118
                       8.0
                               72
                                         5
5
5
5
    12.0
                149 12.6
                                              4
    18.0
                313 11.5
                               62
5
                                              5
    31.5
                205 14.3
                               56
                                              6
    28.0
                205 14.9
                               66
                                Data Transformation
> breaks=c(0,50,100,150,200,250,300,350)
> breaks
[1]
        0 50 100 150 200 250 300 350
   airquality$Solar.R=cut(airquality$Solar.R,breaks=breaks,include.lowest=TRUE
             rquality)
Solar.R Wind Temp Month Day
L50,200] 7.4 67 5 1
  head(airquality)
   Ozone
    41.0 (150,200]
                                            5
5
5
5
    36.0 (100,150] 8.0
12.0 (100,150] 12.6
18.0 (300,350] 11.5
31.5 (200,250] 14.3
                                                 3
                                  74
                                                 4
                                  62
                                            5
                                                 5
                                  56
    28.0 (200,250] 14.9
                                  66
air=airquality
air=airquairty
> air$Month=gsub(13,"May",air$Month)
> air$Month=gsub(14,"June",air$Month)
> air$Month=gsub(15,"July",air$Month)
> air$Month=gsub(16,"August",air$Month)
> air$Month=gsub(17,"Sept",air$Month)
  head(air)
              Solar.R Wind Temp Month Day
   Ozone
    41.0 (150,200]
                                                 1
                         7.4
                                  67
                                                 2
                                  72
    36.0 (100,150]
                          8.0
                                            5
5
5
5
5
    12.0 (100,150]
18.0 (300,350]
31.5 (200,250]
28.0 (200,250]
                        12.6
                                  74
                                  62
                                                 4
                        11.5
```

5 6

14.3

14.9

56

66

## Data model Building

```
> data("airquality")
  str(airquality)
'data.frame':
                  153 obs. of 6 variables:
41 36 12 18 NA 28 23 19 8 NA ...
 $ Ozone
           : int
                    190 118 149 313 NA NA 299 99 19 194
   Solar.R: int
                    7.4 8 12.6 11.5 14.3 14.9 8.6 13.8 20.1 8.6 ... 67 72 74 62 56 66 65 59 61 69 ...
   Wind
              num
   Temp
              int
                    5 5 5 5 5 5 5 5 5 5 5 ...
1 2 3 4 5 6 7 8 9 10 ...
   Month
              int
 $ Day
              int
> head(airquality)
  Ozone Solar R Wind Temp Month Day
      41
              190
                    7.4
                            67
                                         2
2
                    8.0
                            72
                                    5
      36
              118
3
                                         3
      12
              149 12.6
                            74
                                    5
5
5
4
      18
               313 11.5
                            62
                                         4
5
                                         5
      NA
               NA 14.3
                            56
               NA 14.9
                                         6
6
      28
                            66
  air=airquality
  head(air)
  Ozone Solar.R Wind Temp Month Day
1
              190
                                         1
      41
                    7.4
                            67
2
                                    5
      36
                    8.0
                            72
              118
3
      12
              149 12.6
                                    5
                                         3
                            74
                                    5
4
                                         4
      18
               313 11.5
                            62
                                         5
5
      NA
               NA 14.3
                            56
                                         6
6
               NA 14.9
      28
                            66
  summary(air)
      Ozone
                         Solar.R
                                               Wind
                                                                    Temp
                             : 7.0
                                                                      :56.00
                                                 : 1.700
 Min.
             1.00
                     Min.
                                         Min.
                                                              Min.
 1st Qu.: 18.00
                      1st Qu.:115.8
                                         1st Qu.: 7.400
                                                              1st Qu.:72.00
                                         Median: 9.700
 Median : 31.50
                     Median :205.0
                                                              Median :79.00
          : 42.13
                              :185.9
                                                  : 9.958
                     Mean
                                         Mean
                                                              Mean
                                                                      :77.88
 Mean
 3rd Qu.: 63.25
                      3rd Qu.:258.8
                                         3rd Qu.:11.500
                                                              3rd Qu.:85.00
          :168.00
                              :334.0
                                                  :20.700
                                                                      :97.00
 Max.
                     Max.
                                         Max.
                                                              Max.
                     NA's
 NA's
          :37
      Month
                          Day
         :5.000
 Min.
                    Min.
                               1.0
 1st Qu.:6.000
                    1st Qu.: 8.0
 Median :7.000
                    Median:16.0
 Mean
          :6.993
                    Mean
                            :15.8
 3rd Qu.:8.000
                    3rd Qu.:23.0
          :9.000
                             :31.0
 Max.
                    Max.
> for(i in 1:nrow(air)){
  if(is.na(air[i,"Ozone"])){
  air[i,"Ozone"]<-mean(air[which(air[,"Month"]==air[i,"month"]),"Ozone"],na.rm</pre>
=TRUE)
 #input monthly mean in Solar.R
if(is.na(air[i,"Solar.R"])){
air[i,"Solar.R"]<-mean(air[which(air[,"Month"]==air[i,"Month"]),"Solar.R"],n</pre>
a.rm=TRUE)
> summary(air)
     Ozone
                         Solar.R
                                               Wind
                                                                    Temp
                                                 : 1.700
                                                                      :56.00
             1.00
                     Min.
                              :144.6
                                         Min.
                                                              Min.
 1st Qu.: 18.00
                      1st Qu.:163.4
                                         1st Qu.: 7.400
                                                              1st Qu.:72.00
 Median : 31.50
                     Median :179.5
                                         Median : 9.700
                                                              Median :79.00
                     Mean :179.8
                                                : 9.958
         : 42.13
                                         Mean
                                                              Mean :77.88
 Mean
 3rd Qu.: 63.25
                      3rd Qu.:188.0
                                         3rd Qu.:11.500
                                                              3rd Qu.:85.00
```

```
:168.00
                   Max.
                           :218.0
                                     Max.
                                             :20.700
                                                                :97.00
 Max.
                                                        Max.
 NA's
         :37
     Month
                        Day
                          : 1.0
 Min.
         :5.000
                  Min.
 1st Qu.:6.000
                  1st Qu.: 8.0
 Median :7.000
                  Median:16.0
 Mean
        :6.993
                  Mean :15.8
 3rd Qu.:8.000
                  3rd Qu.:23.0
 Max.
        :9.000
                  Max.
                          :31.0
> normalize<-function(x){</pre>
+ return(x-min(x)/(max(x)-min(x)))
+ }
> air<-normalize(air)</pre>
> summary(air)
     Ozone
                   Solar.R
                                     Wind
                                                     Temp
                                                                   Month
                Min. : NA
                                               Min.
 Min.
        : NA
                                Min.
                                       : NA
                                                      : NA
                                                              Min.
                                                                     : NA
 1st Qu.: NA
                1st Qu.: NA
                                               1st Qu.: NA
                                                              1st Qu.: NA
                                1st Qu.: NA
 Median : NA
                Median: NA
                                Median: NA
                                               Median: NA
                                                              Median: NA
                Mean
                      :NaN
                                Mean : NaN
                                               Mean
                                                      :NaN
                                                              Mean
 Mean
        :NaN
                                                                    :NaN
 3rd Qu.: NA
                3rd Qu.: NA
                                3rd Qu.: NA
                                               3rd Qu.: NA
                                                              3rd Qu.: NA
                                                       : NA
         : NA
                       : NA
                                       : NA
                                               Max.
                                                              Max.
                                                                      : NA
 Max.
                Max.
                                Max.
 NA's
         :153
                NA's
                        :153
                                NA's
                                       :153
                                               NA's
                                                       :153
                                                              NA's
                                                                      :153
      Day
 Min.
         : NA
 1st Qu.: NA
 Median: NA
 Mean
        :NaN
 3rd Qu.: NA
 Max.
        : NA
         :153
 NA's
> str(air)
                153 obs. of 6 variables:
 data.frame':
 $ Ozone : num NA ...
 $ Solar.R: num
                  NA NA NA NA NA NA NA NA NA ...
 $ Wind
          : num
                  NA NA NA NA NA NA NA NA NA ...
 $ Temp
           : num
                  NA NA NA NA NA NA NA NA NA ...
 $ Month
          : num NA NA NA NA NA NA NA NA NA ...
 $ Day
           : num NA NA NA NA NA NA NA NA NA ...
> x=air[,"Ozone"]
> y=air[,"Solar.R"]
> plot(x~y)
> plot(y~x)
> model=lm(y~x)
> x=airquality[,"Ozone"]
> y=airquality[,"Solar.R"]
  plot(x~y)
> model1=1m(y~x)
> model1
call:
lm(formula = y \sim x)
Coefficients:
(Intercept)
                   0.9542
   144.6306
> abline(model1,col="pink")
> abline(model1,col="Blue")
```