Bivariate Analysis

Code

#Cind820

Hide

```r
data <- read.csv(file = \C:/Users/16133/OneDrive/Desktop/CIND820/Data sets/PRSA_Data_Shunyi.csv\
                     , sep = \
```

Hide

```r
str(datab)
```


```
<!-- rnb-source-end -->

<!-- rnb-output-begin eyJkYXRhIjoiJ2RhdGEuZnJhbWUnOlx0MzUwNjQgb2JzLiBvZiAgMTUgdmFyaWFibGVzOlxuICQgeWVhciA6IG51bSAgMjAxMyAyMDEzIDIwMTMgMjAxMyAyMDEzIC4uLlxuICQgbW9udGg6IG51bSAgMyAzIDMgMyAzIDMgMyAzIDMgMyAuLi5cbiAkIGRheSAgOiBudW0gIDEgMSAxIDEgMSAxIDEgMSAxIDEgLi4uXG4gJCBob3VyIDogbnVtICAwIDEgMiAzIDQgNSA2IDcgOCA5IC4uLlxuICQgUE0yLjU6IG51bSAgMyAxMiAxNCAxMiAxMiAxMSAxMiAxMyA4IDMgLi4uXG4gJCBQTTEwIDogbnVtICA2IDEyIDE0IDEyIDEyIDExIDEyIDEzIDggNiAuLi5cbiAkIFNPMiAgOiBudW0gIDMgMyA1IDMgMyAzIDMgMyAzIDMgLi4uXG4gJCBOTzIgIDogbnVtICA4IDcgNyA1IDM3IDcgOSAyMyAxOSAyMSAuLi5cbiAkIENPICAgOiBudW0gIDMwMCAzMDAgMjAwIDgwMCAyMDAgMjAwIDMwMCAzMDAgNDAwIDQwMCAuLi5cbiAkIE8zICAgOiBudW0gIDQ0IDQ3IDIyIDQzIDExIDQ1IDc0IDU5IDY2IDYwIC4uLlxuICQgVEVNUCA6IG51bSAgLTAuOSAtMS4xIC0xLjcgLTIuMSAtMi40IC0yLjggLTQgLTIuNCAtMSAwIC4uLlxuICQgUFJFUyA6IG51bSAgMTAyNiAxMDI2IDEwMjYgMTAyNyAxMDI4IC4uLlxuICQgREVXUCA6IG51bSAgLTIwLjUgLTIxLjMgLTIzIC0yMy4zIC0yMi45IC0yMi4xIC0yMS4yIC0yMS4zIC0yMS44IC0yMi45IC4uLlxuICQgUkFJTiA6IG51bSAgMCAwIDAgMCAwIDAgMCAwIDAgMCAuLi5cbiAkIFdTUE0gOiBudW0gIDkuMyA5LjQgOC42IDYuNiA0LjUgMS43IDEuNiAxLjcgMi43IDAuOCAuLi5cbiJ9 -->
```


‘data.frame’: 35064 obs. of 15 variables: $ year : num 2013 2013 2013 2013 2013 … $ month: num 3 3 3 3 3 3 3 3 3 3 … $ day : num 1 1 1 1 1 1 1 1 1 1 … $ hour : num 0 1 2 3 4 5 6 7 8 9 … $ PM2.5: num 3 12 14 12 12 11 12 13 8 3 … $ PM10 : num 6 12 14 12 12 11 12 13 8 6 … $ SO2 : num 3 3 5 3 3 3 3 3 3 3 … $ NO2 : num 8 7 7 5 37 7 9 23 19 21 … $ CO : num 300 300 200 800 200 200 300 300 400 400 … $ O3 : num 44 47 22 43 11 45 74 59 66 60 … $ TEMP : num -0.9 -1.1 -1.7 -2.1 -2.4 -2.8 -4 -2.4 -1 0 … $ PRES : num 1026 1026 1026 1027 1028 … $ DEWP : num -20.5 -21.3 -23 -23.3 -22.9 -22.1 -21.2 -21.3 -21.8 -22.9 … $ RAIN : num 0 0 0 0 0 0 0 0 0 0 … $ WSPM : num 9.3 9.4 8.6 6.6 4.5 1.7 1.6 1.7 2.7 0.8 …

```
<!-- rnb-output-end -->

<!-- rnb-chunk-end -->


<!-- rnb-chunk-begin -->

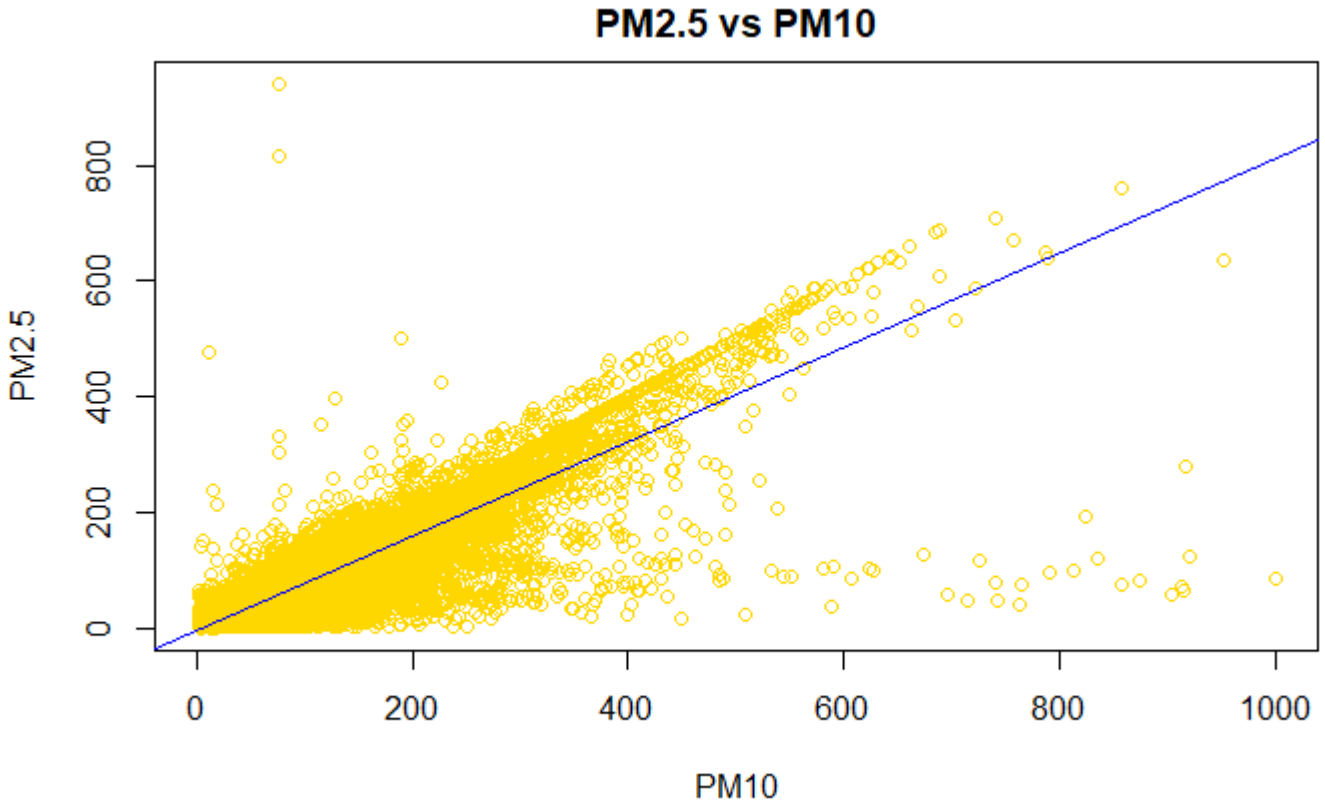

<!-- rnb-source-begin eyJkYXRhIjoiYGBgclxuI2xvYWQgbmVjZXNzYXJ5IGxpYnJhcmllc1xubGlicmFyeShnZ3Bsb3QyKVxubGlicmFyeShnZ3B1YnIpXG
4jaW5zdGFsbC5wYWNrYWdlcyhcInZpcmlkaXNcIilcbmxpYnJhcnkoXCJ2aXJpZGlzXCIpXG5gYGAifQ== -->

```r
#load necessary libraries
library(ggplot2)
library(ggpubr)
#install.packages("viridis")
library("viridis")
```

Hide

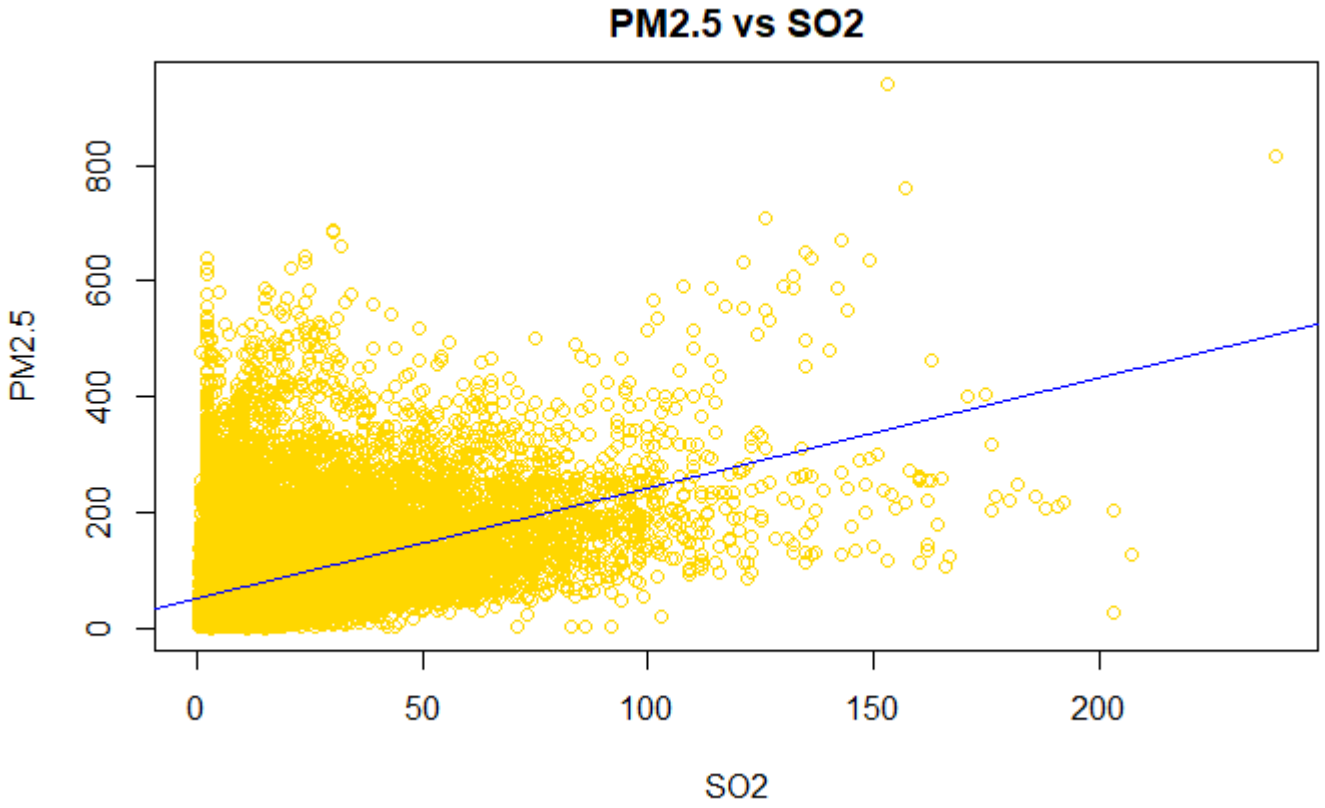
```
#plot PM2.5 vs PM10
#matches with strong positive correlation
plot(datab$PM10,datab$PM2.5, ylab = "PM2.5", xlab = "PM10", col = "gold", main= "PM2.5 vs PM10")

abline(lm(formula = PM2.5 ~ PM10 , data = datab), col = "Blue")
```
```



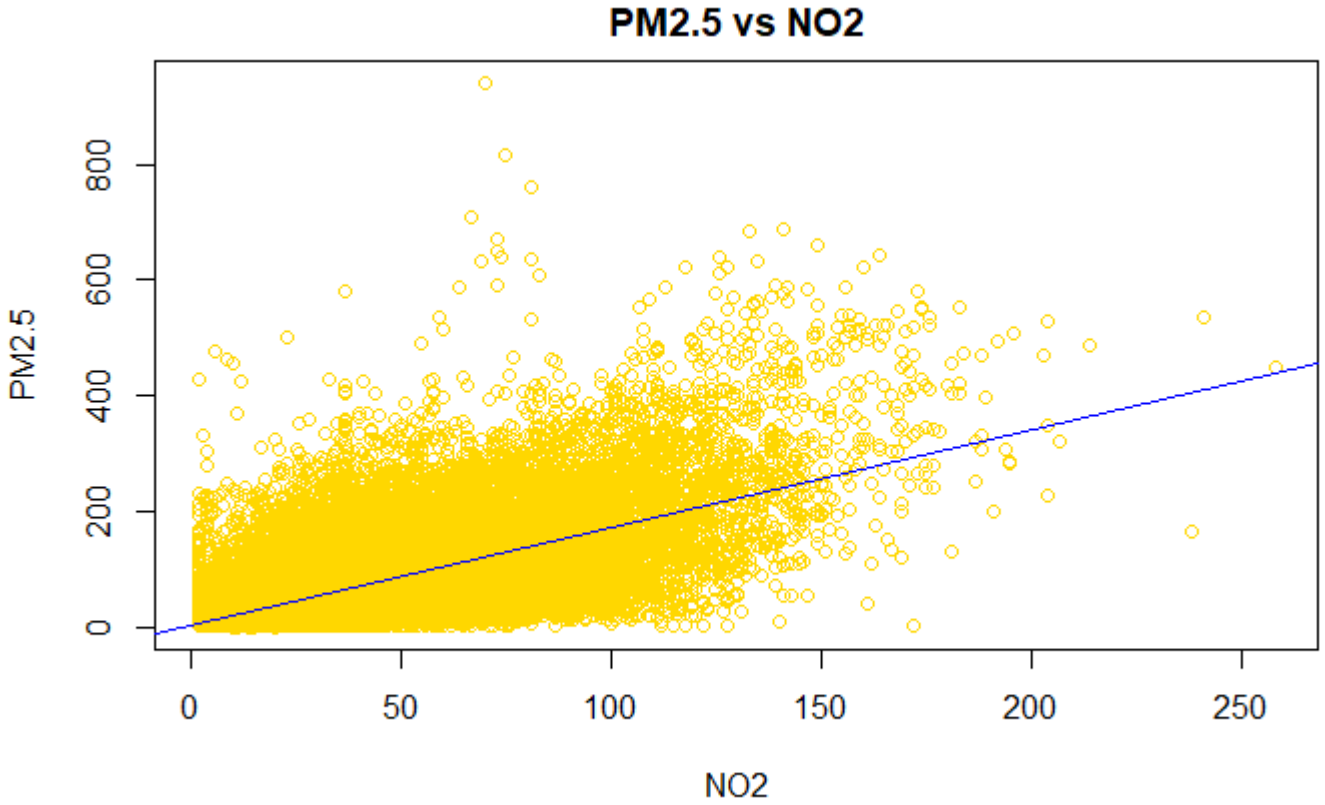
Hide

```
#plot PM2.5 vs So2
#weaker positive correlation with PM2.5
plot(datab$SO2,datab$PM2.5, ylab = "PM2.5", xlab = "SO2",col = "gold", main= "PM2.5 vs SO2")
abline(lm(formula = PM2.5 ~ SO2 , data = datab), col = "Blue")
```



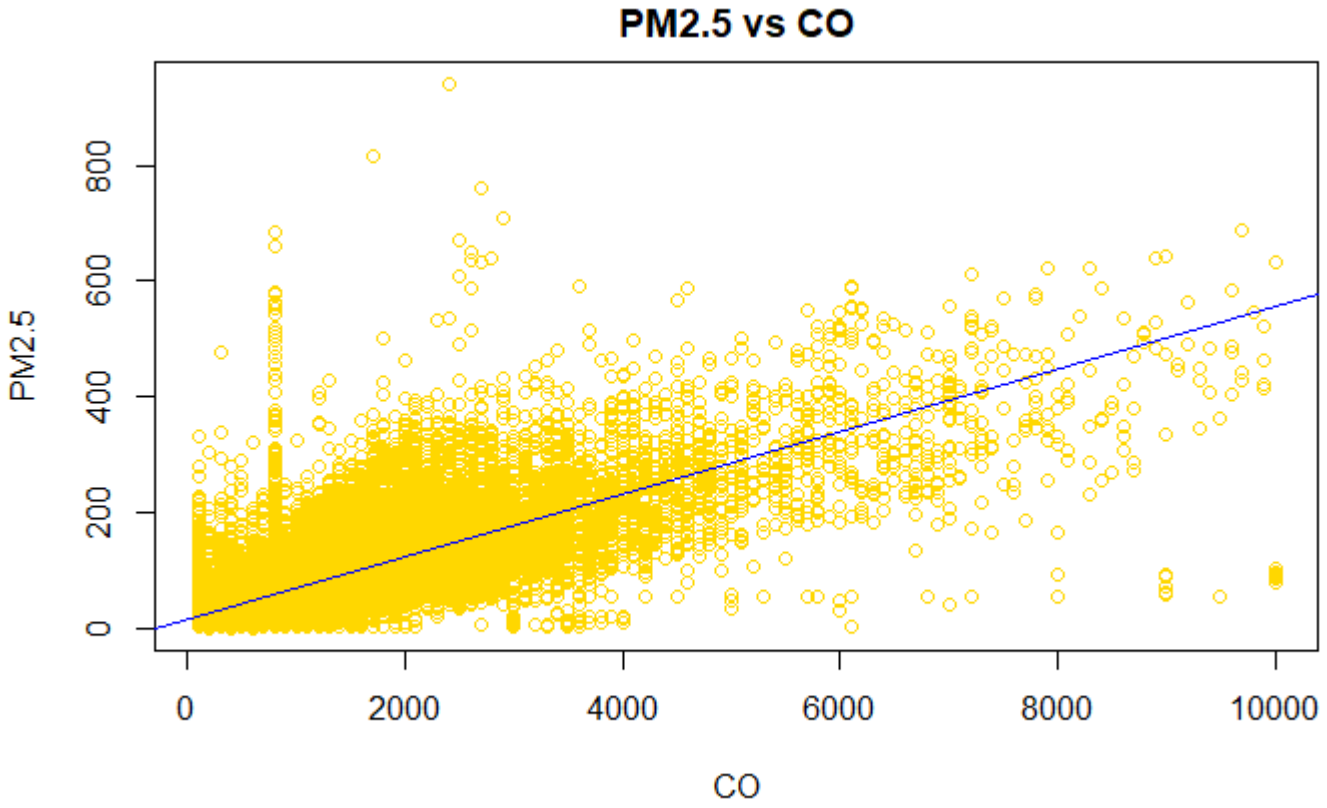
Hide

```
#plot PM2.5 vs NO2
#3rd strongest correlation with PM2.5
plot(datab$NO2,datab$PM2.5, ylab = "PM2.5", xlab = "NO2", col = "gold", main= "PM2.5 vs NO2")
abline(lm(formula = PM2.5 ~ NO2 , data = datab), col = "Blue")
```



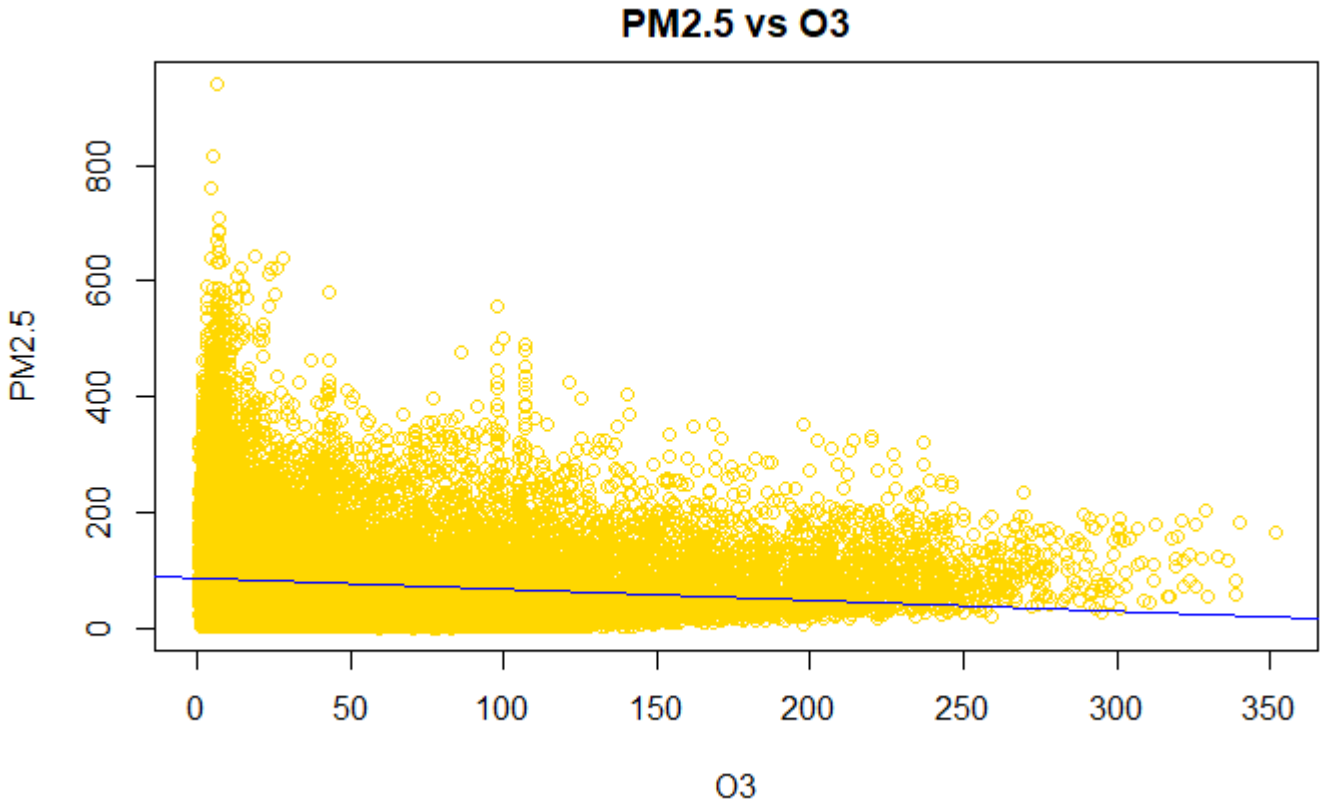
Hide

```
#plot PM2.5 vs CO
#2nd strongest correlation to PM2.5
plot(datab$CO,datab$PM2.5, ylab = "PM2.5", xlab = "CO", col = "gold", main= "PM2.5 vs CO")
abline(lm(formula = PM2.5 ~ CO, data = datab), col = "Blue")
```



Hide

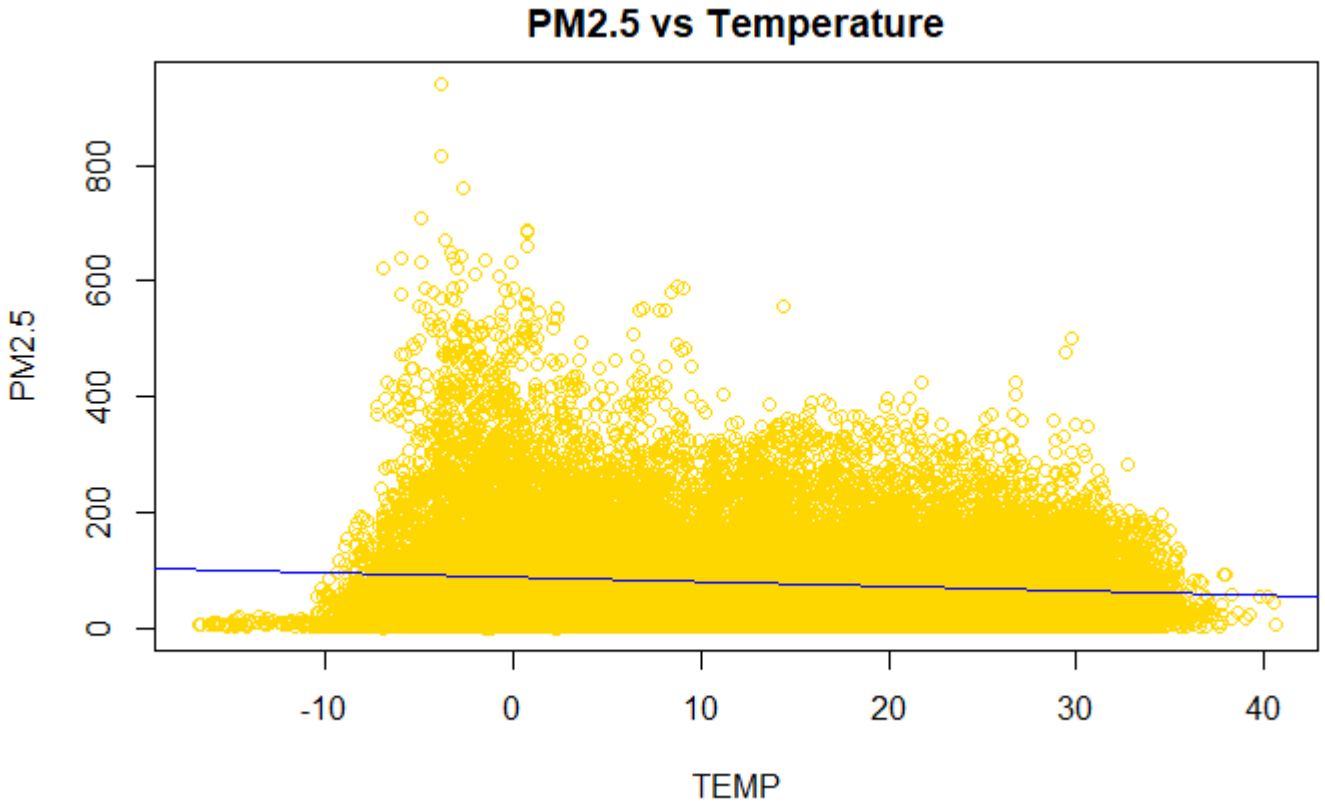
```
#plot PM2.5 vs O3
#higher levels of O3 with lower levels of PM2.5
plot(datab$O3,datab$PM2.5, ylab = "PM2.5", xlab = "O3", col = "gold", main= "PM2.5 vs O3")
abline(lm(formula = PM2.5 ~ O3, data = datab), col = "Blue")
```



Hide

```
#plot PM2.5 vs temp
#matches with strong positive correlation
plot(datab$TEMP,datab$PM2.5, ylab = "PM2.5", xlab = "TEMP", col = "gold", main= "PM2.5 vs Temperature")

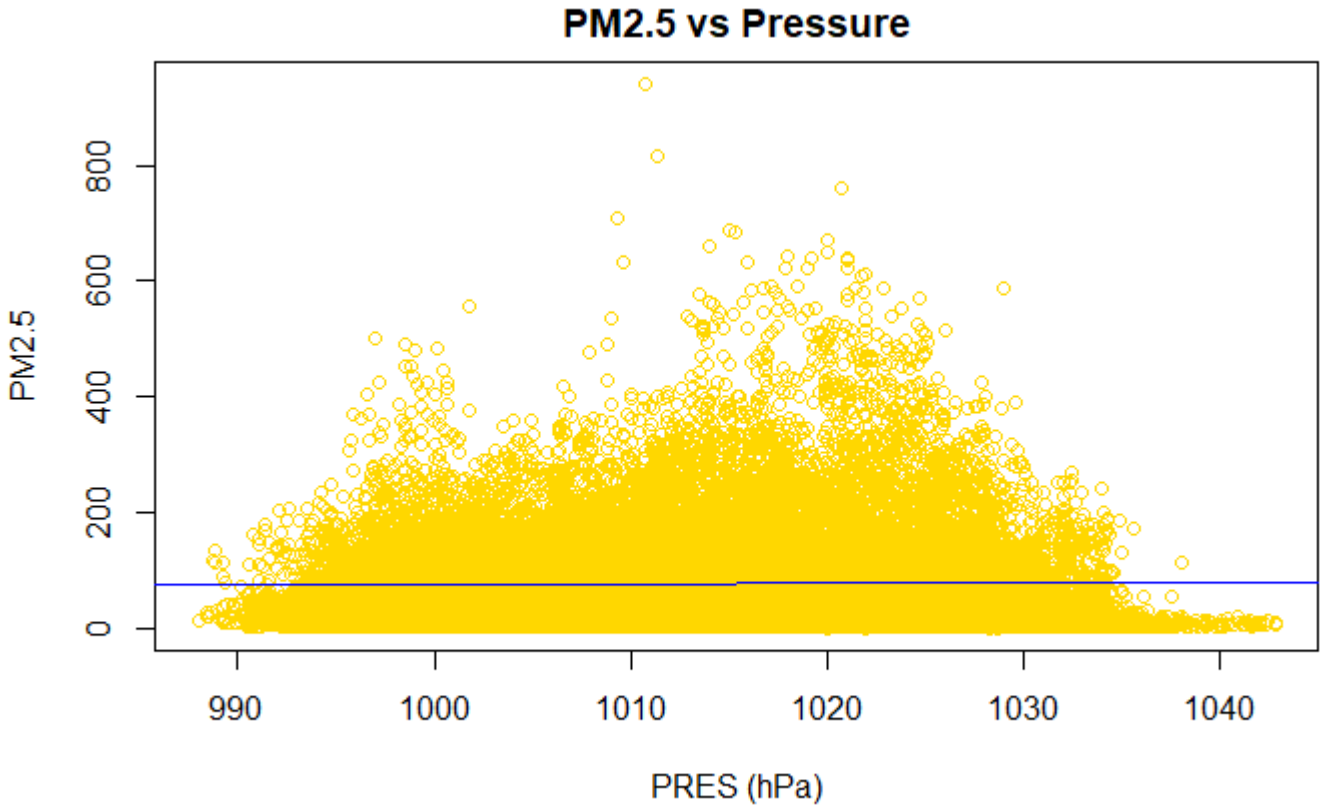

abline(lm(formula = PM2.5 ~ TEMP, data = datab), col = "Blue")
```



Hide

```
#plot PM2.5 vs Pressure
#matches with strong positive correlation
plot(datab$PRES,datab$PM2.5, ylab = "PM2.5", xlab = "PRES (hPa)", col = "gold", main= "PM2.5 vs Pressure")

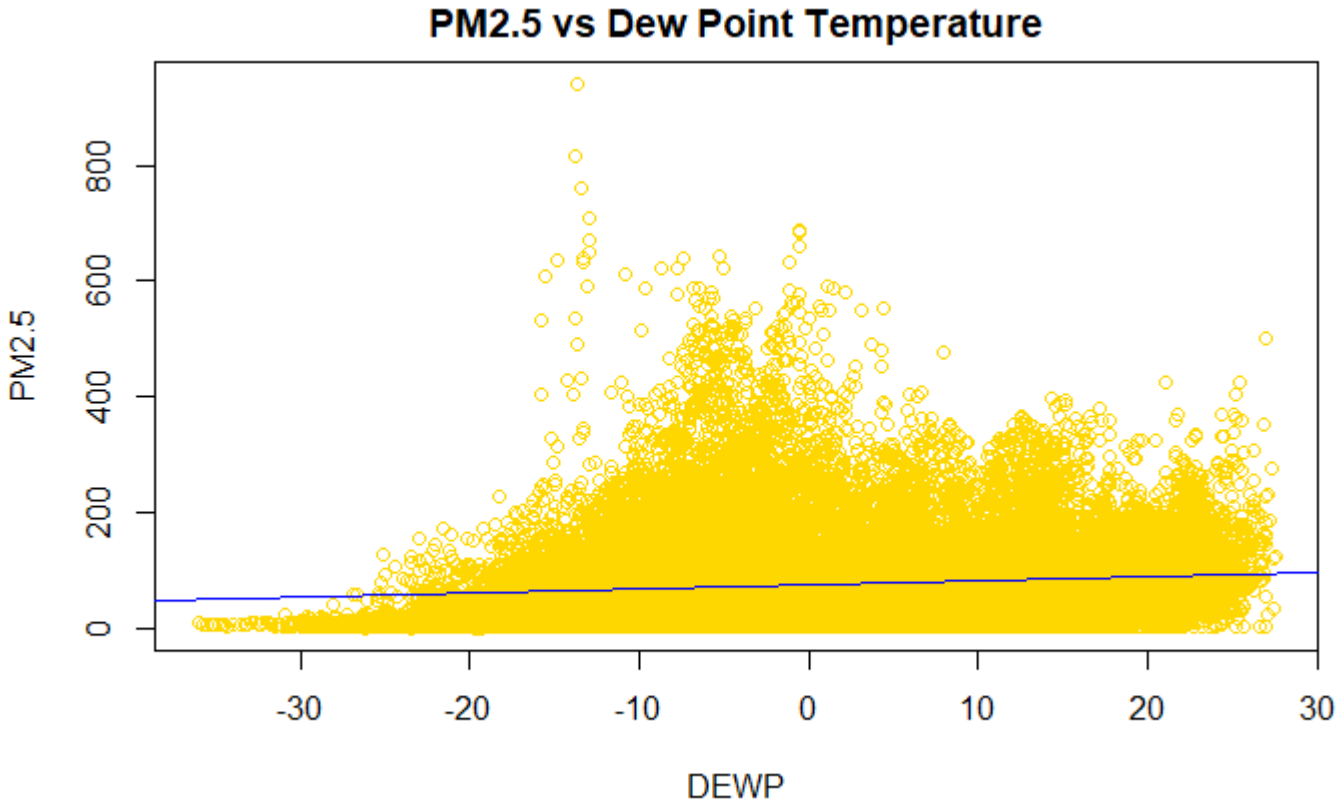

abline(lm(formula = PM2.5 ~ PRES , data = datab), col = "Blue")
```



Hide

```
#plot PM2.5 vs DEWP
#matches with strong positive correlation
plot(datab$DEWP,datab$PM2.5, ylab = "PM2.5", xlab = "DEWP", col = "gold", main= "PM2.5 vs Dew Point Temperature")

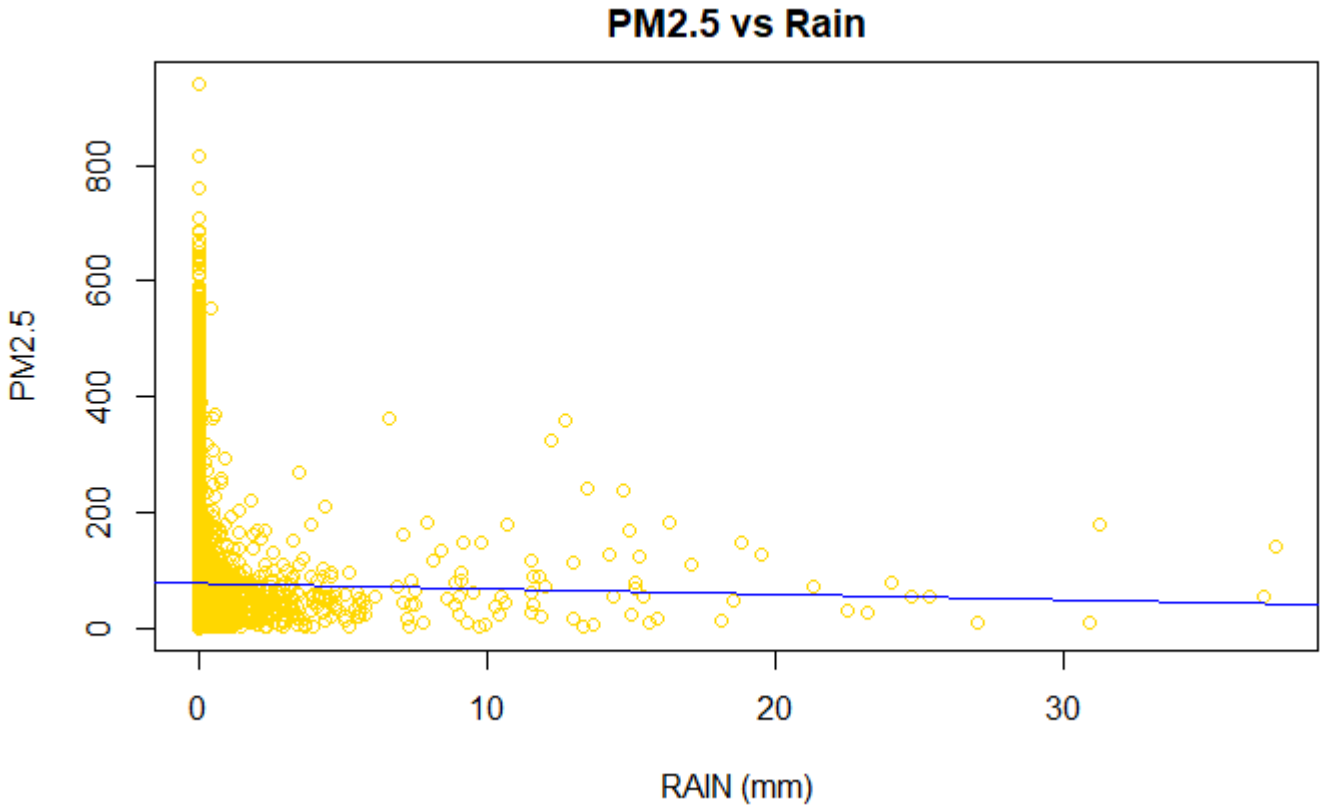

abline(lm(formula = PM2.5 ~ DEWP , data = datab), col="Blue")
```



Hide

```
#plot PM2.5 vs rain
#matches with strong positive correlation
plot(datab$RAIN,datab$PM2.5, ylab = "PM2.5", xlab = "RAIN (mm)", col = "gold", main= "PM2.5 vs Rain")

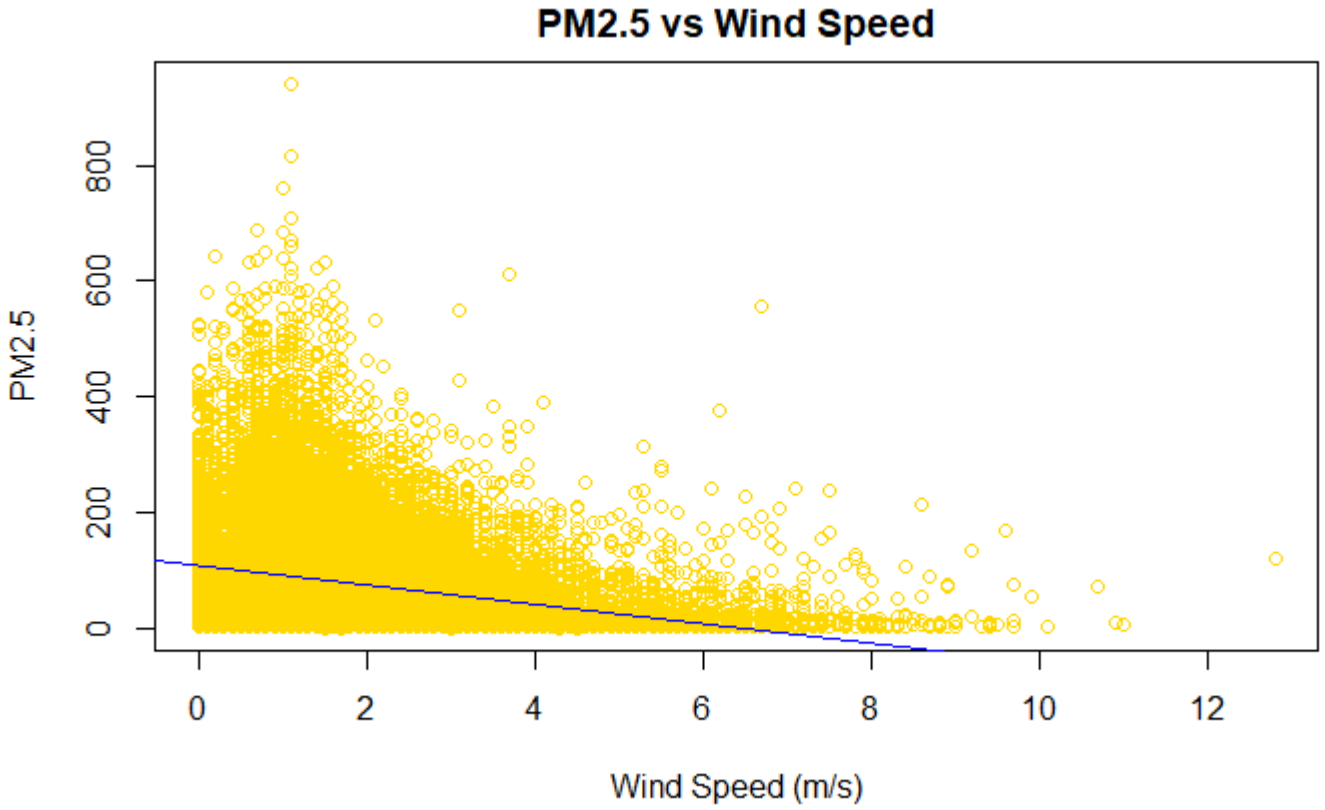
abline(lm(formula = PM2.5 ~ RAIN, data = datab), col="Blue")
```



Hide

```
#plot PM2.5 vs WSPM
#matches with strong positive correlation
plot(datab$WSPM ,datab$PM2.5, ylab = "PM2.5", xlab = "Wind Speed (m/s)", col = "gold", main= "PM2.5 vs Wind Speed")


abline(lm(formula = PM2.5 ~ WSPM, data = datab), col="Blue")
```



Hide

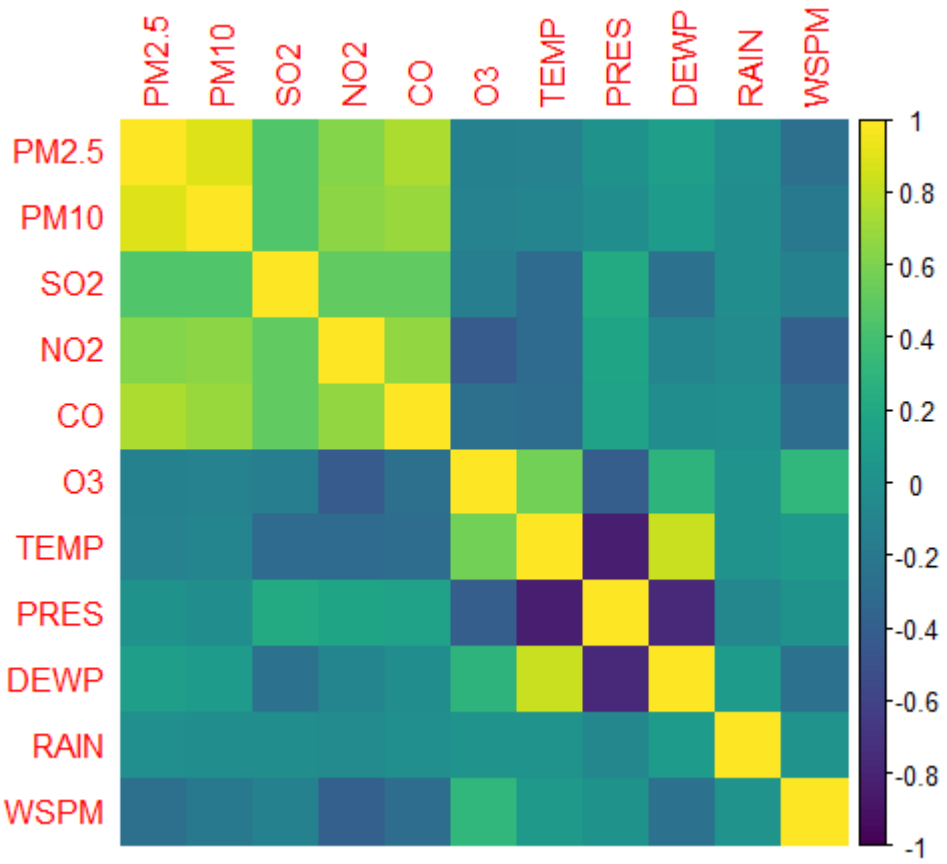
```
#corelation matrix
library(corrplot)
#remove year month day hour
data_c <- datab[,5:15]

corr_data_p <- cor(data_c, method = "pearson")
corr_data_p
```

```
             PM2.5        PM10         SO2         NO2           CO          O3        TEMP         PRES        DEWP
RAIN        WSPM
PM2.5  1.000000000  0.89560172  0.45757037  0.63567288  0.758156257 -0.13155314 -0.11746932  0.001146066  0.11966483 -0.0085
65877 -0.26692466
PM10   0.895601717  1.00000000  0.45695431  0.64594289  0.692973154 -0.11317798 -0.08411335 -0.030123110  0.08778240 -0.0229
43664 -0.19826129
SO2    0.457570372  0.45695431  1.00000000  0.50742828  0.511599532 -0.14002971 -0.30700867  0.212423415 -0.25646925 -0.0387
86876 -0.12218223
NO2    0.635672883  0.64594289  0.50742828  1.00000000  0.663884971 -0.43255481 -0.30586034  0.160063410 -0.08547603 -0.0503
11328 -0.38511026
CO     0.758156257  0.69297315  0.51159953  0.66388497  1.000000000 -0.26205963 -0.29183666  0.153649761 -0.03844737 -0.0073
32498 -0.28305283
O3    -0.131553140 -0.11317798 -0.14002971 -0.43255481 -0.262059628  1.00000000  0.57030331 -0.419318025  0.29235200  0.0229
88793  0.30554509
TEMP  -0.117469320 -0.08411335 -0.30700867 -0.30586034 -0.291836658  0.57030331  1.00000000 -0.825853530  0.82383634  0.0393
24188  0.06319276
PRES   0.001146066 -0.03012311  0.21242341  0.16006341  0.153649761 -0.41931802 -0.82585353  1.000000000 -0.76249285 -0.0639
14775  0.01646404
DEWP   0.119664827  0.08778240 -0.25646925 -0.08547603 -0.038447366  0.29235200  0.82383634 -0.762492846  1.00000000  0.0874
34596 -0.25609572
RAIN  -0.008565877 -0.02294366 -0.03878688 -0.05031133 -0.007332498  0.02298879  0.03932419 -0.063914775  0.08743460  1.0000
00000  0.02894404
WSPM  -0.266924663 -0.19826129 -0.12218223 -0.38511026 -0.283052832  0.30554509  0.06319276  0.016464044 -0.25609572  0.0289
44038  1.00000000
```

Hide

```
corrplot(corr_data_p, method = "color", col = viridis(100))
```



Hide

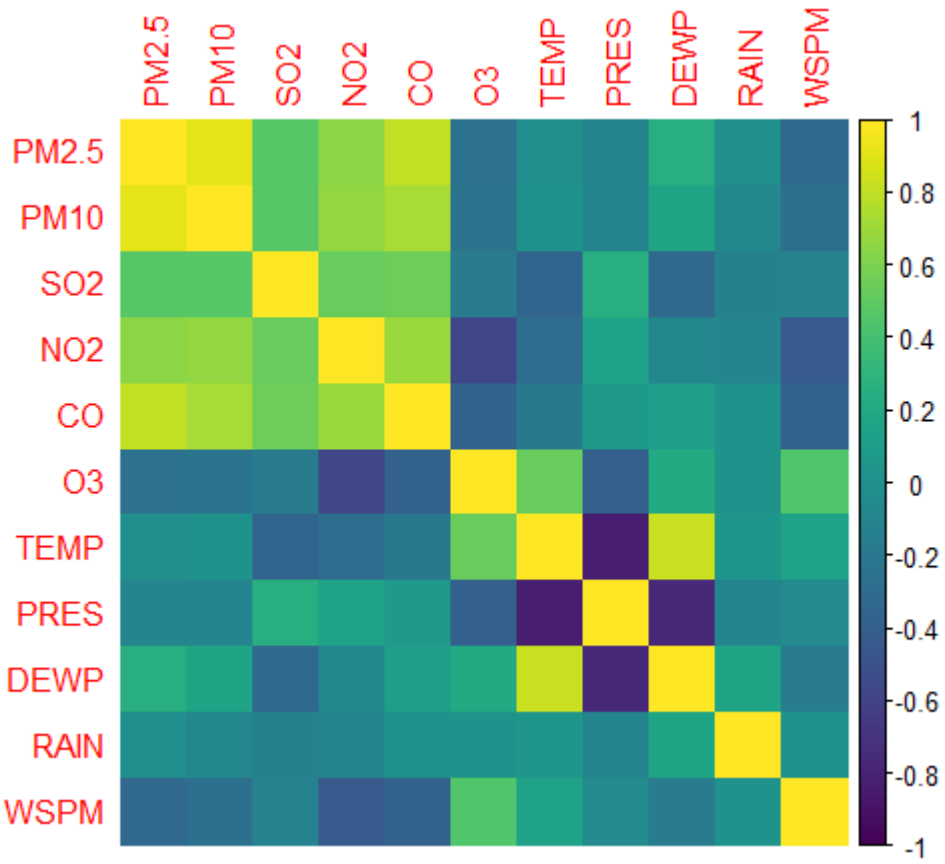
```
corr_data_sp <- cor(data_c, method = "spearman")

corr_data_sp
```

```
             PM2.5         PM10        SO2         NO2          CO           O3         TEMP        PRES        DEWP
RAIN        WSPM
PM2.5  1.000000000  0.907673969  0.4748902  0.65463223  0.80519026 -0.254073853 -0.007406454 -0.08912466  0.24795976 -0.0158
33024 -0.32329398
PM10   0.907673969  1.000000000  0.4745972  0.66950852  0.72234471 -0.235905227  0.001249821 -0.09675595  0.17653217 -0.0674
86441 -0.27108146
SO2    0.474890159  0.474597239  1.0000000  0.53715576  0.54572333 -0.179425558 -0.340386583  0.25840037 -0.32054876 -0.1207
30415 -0.11507547
NO2    0.654632231  0.669508520  0.5371558  1.00000000  0.68635127 -0.561898399 -0.283484397  0.15417133 -0.06728786 -0.0933
59544 -0.43751601
CO     0.805190257  0.722344711  0.5457233  0.68635127  1.00000000 -0.374841769 -0.191761448  0.07921751  0.10066421  0.0119
23750 -0.36046065
O3    -0.254073853 -0.235905227 -0.1794256 -0.56189840 -0.37484177  1.000000000  0.537605630 -0.38934903  0.21317263  0.0079
57887  0.44629919
TEMP  -0.007406454  0.001249821 -0.3403866 -0.28348440 -0.19176145  0.537605630  1.000000000 -0.83118150  0.82114140  0.0441
07185  0.15741244
PRES  -0.089124660 -0.096755954  0.2584004  0.15417133  0.07921751 -0.389349028 -0.831181504  1.00000000 -0.76876947 -0.0864
83656 -0.04223735
DEWP   0.247959764  0.176532174 -0.3205488 -0.06728786  0.10066421  0.213172629  0.821141401 -0.76876947  1.00000000  0.1753
29493 -0.16083797
RAIN  -0.015833024 -0.067486441 -0.1207304 -0.09335954  0.01192375  0.007957887  0.044107185 -0.08648366  0.17532949  1.0000
00000  0.01605394
WSPM  -0.323293984 -0.271081457 -0.1150755 -0.43751601 -0.36046065  0.446299188  0.157412439 -0.04223735 -0.16083797  0.0160
53940  1.00000000
```

Hide

```
corrplot(corr_data_sp, method = "color", col = viridis(100))
```


PM2.5
PM10
SO2
NO2
CO
O3
TEMP
PRES
DEWP
RAIN
WSPM
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1