Thien Pham Final Report

Thien Pham

2024-06-07

NOx Concentration in Polluted Areas

1. Introduction

The goal of the study is to study NOx relationship with other gases in polluted areas and predict NOx concentration level, as NOx is highly linked to air quality, as discussed by study 1, where descending to a certain critical breakpoint in NOx concentration results in the decline in “formation of secondary aerosol.” Another purpose is to attempt to see if NOx concentration levels can be predicted purely through sensors, temperature, and humidity level.

2. Description of Data

Description: The data contains hourly sensor response averages along with gas concentration references from a certified analyzer. This data is collected from a gas multisensor device deployed on the field of an Italian city.

Variable Name	Description	Unit
Date	Date (day/month/year)	NA
Time	Time (hour:minute:second)	NA
CO(GT)	True hourly averaged concentration of Carbon Monoxide	mg/m^3
PT08.S1(CO)	Tin oxide hourly averaged sensor response (CO targeted)	NA
NMHC(GT)	Non-Methane Hydrocarbons concentration	miccrog/m^3
C6H6(GT)	Benzene concentration	microg/m^3
PT08.S2(NMHC)	Titania hourly averaged sensor response (NMHC targeted)	NA
NOx(GT)	Nitrogen Oxides concentration	ppb
PT08.S3(NOx)	Tungsten oxide hourly averaged sensor response (NOx targeted)	NA
NO2(GT)	Nitrogen Dioxide concentration	microg/m^3
PT08.S4(NO2)	Tungsten oxide hourly averaged sensor response (NO2 targeted)	NA
PT08.S5(O3)	Indium oxide hourly averaged sensor response (O3 targeted)	NA
T	Temperature	°C
RH	Relative Humidity	%
AH	Absolute Humidity	NA

3. Exploratory Data Analysis

- Loading in Libraries and Dataset:

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.3.2
```

```
##
## 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
```

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.3.2
```

```
## Warning: package 'ggplot2' was built under R version 4.3.2
```

```
## Warning: package 'tibble' was built under R version 4.3.2
```

```
## Warning: package 'tidyr' was built under R version 4.3.2
```

```
## Warning: package 'readr' was built under R version 4.3.2
```

```
## Warning: package 'purrr' was built under R version 4.3.2
```

```
## Warning: package 'forcats' was built under R version 4.3.2
```

```
## Warning: package 'lubridate' was built under R version 4.3.2
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✓ forcats   1.0.0     ✓ readr     2.1.4
## ✓ ggplot2   3.4.4     ✓ stringr   1.5.0
## ✓ lubridate 1.9.3     ✓ tibble    3.2.1
## ✓ purrr     1.0.2     ✓ tidyr     1.3.0
```

```
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(ggplot2)
library(gridExtra)
```

```
## Warning: package 'gridExtra' was built under R version 4.3.2
```

```
##
## Attaching package: 'gridExtra'
##
## The following object is masked from 'package:dplyr':
##
##     combine
```

```
library(leaps)
```

```
## Warning: package 'leaps' was built under R version 4.3.2
```

```
library(car)
```

```
## Warning: package 'car' was built under R version 4.3.2
```

```
## Loading required package: carData
```

```
## Warning: package 'carData' was built under R version 4.3.2
```

```
##
## Attaching package: 'car'
##
## The following object is masked from 'package:purrr':
##
##     some
##
## The following object is masked from 'package:dplyr':
##
##     recode
```

```
library(corrplot)
```

```
## Warning: package 'corrplot' was built under R version 4.3.3
```

```
## corrplot 0.92 loaded
```

```
data <- read.csv("AirQualityUCI (1).csv", header=T)
```

- Summary Statistics of Response Variable:

```
summary(data$NOx.GT.)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -200.0    50.0   141.0   168.6   284.0  1479.0
```

- Data Cleaning:

```
# Replace -200 values with "NA" so it does not affect mean calculation
data[data == -200] <- NA

# Calculate mean for each feature, exclude NA in mean calcuation
feature_means <- data %>%
  summarise(across(where(is.numeric), ~mean(., na.rm = TRUE)))

# Replace NA with the mean of corresponding feature
cleaned_data <- data %>%
  mutate(across(where(is.numeric), ~ifelse(is.na(.), feature_means[[cur_column()]], .)))
```

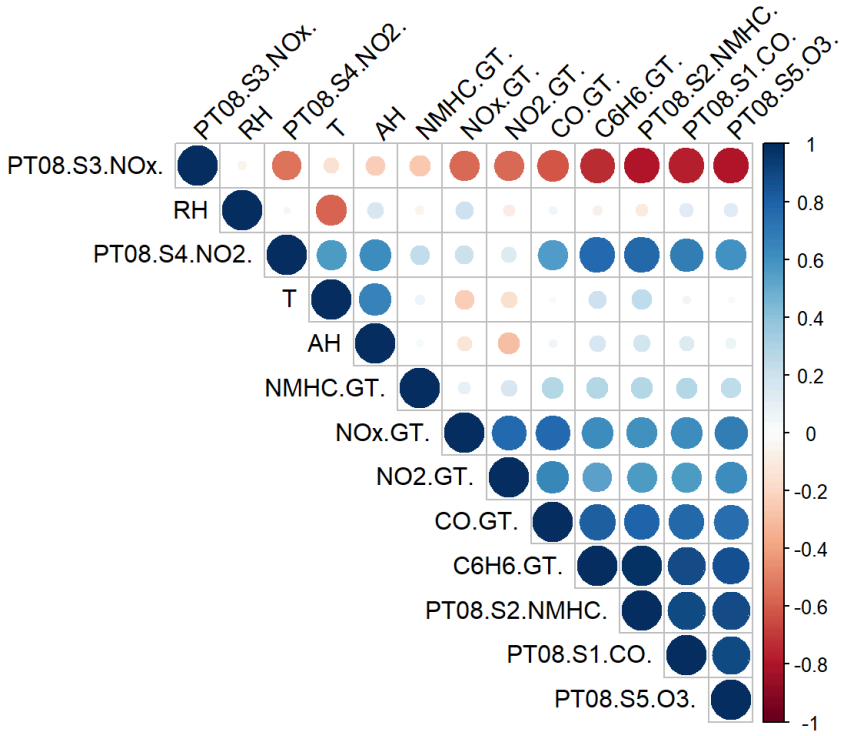
- Summary Statistics of Response Variable after Cleaning:

```
summary(cleaned_data$NOx.GT.)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     2.0   112.0   229.0   246.9   284.0  1479.0
```

- NOx, the response variable, is heavily positively correlated with NO2 concentration, C6H6 concentration, and three sensors. It is also negatively correlated with its corresponding sensor (the sensor for NOx).

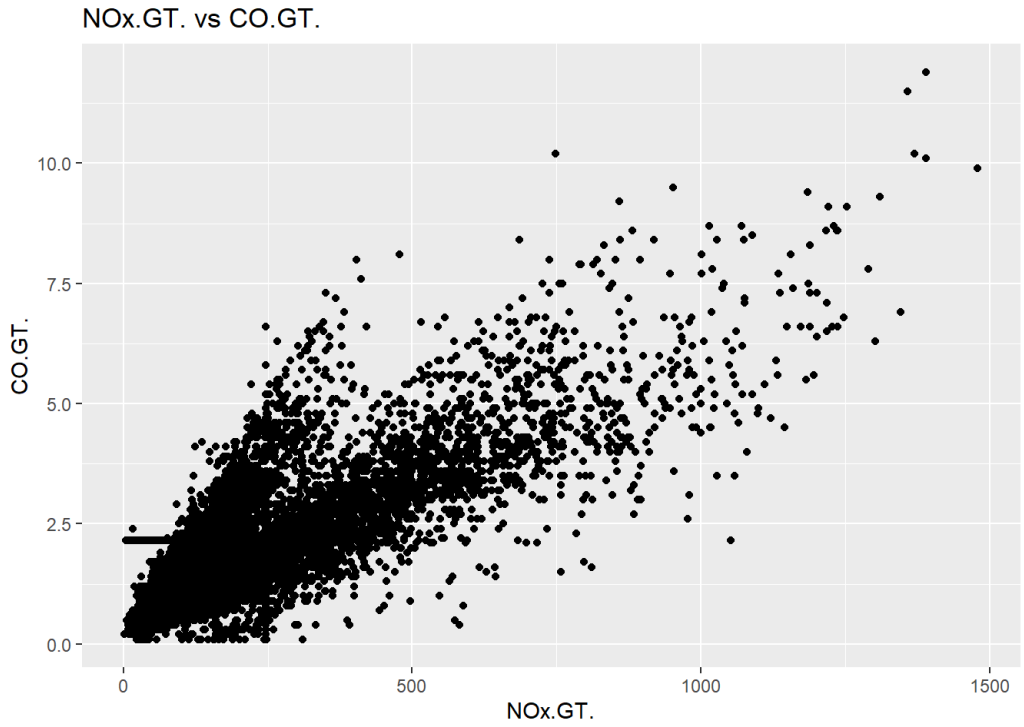
```
cleaned_data_no_time_eda <- select(cleaned_data, -Date, -Time)
cor_matrix <- cor(cleaned_data_no_time_eda)
corrplot(cor_matrix, method = "circle", type = "upper", order = "hclust",
         tl.col = "black", tl.srt = 45)
```

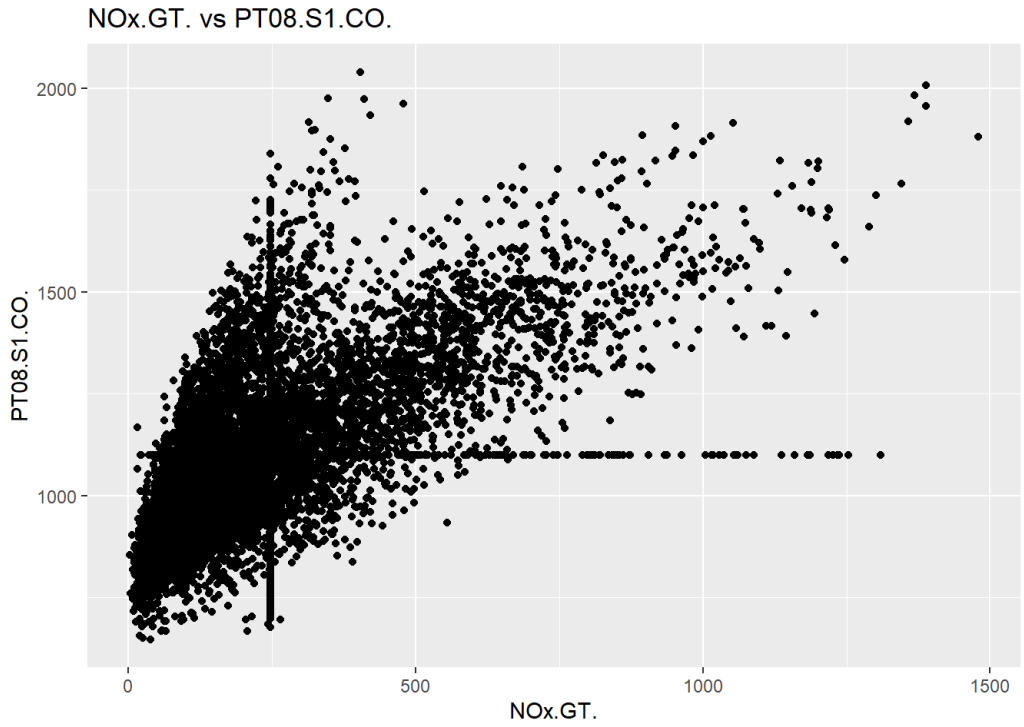


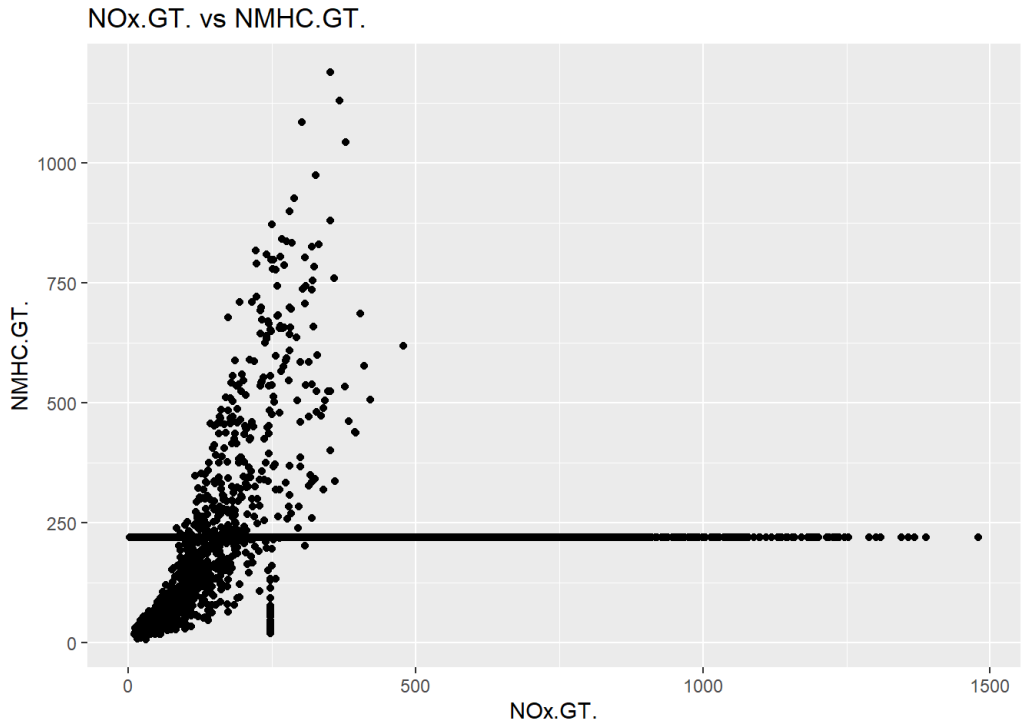
- Scatter Plot of All Variables Versus NOx(response):

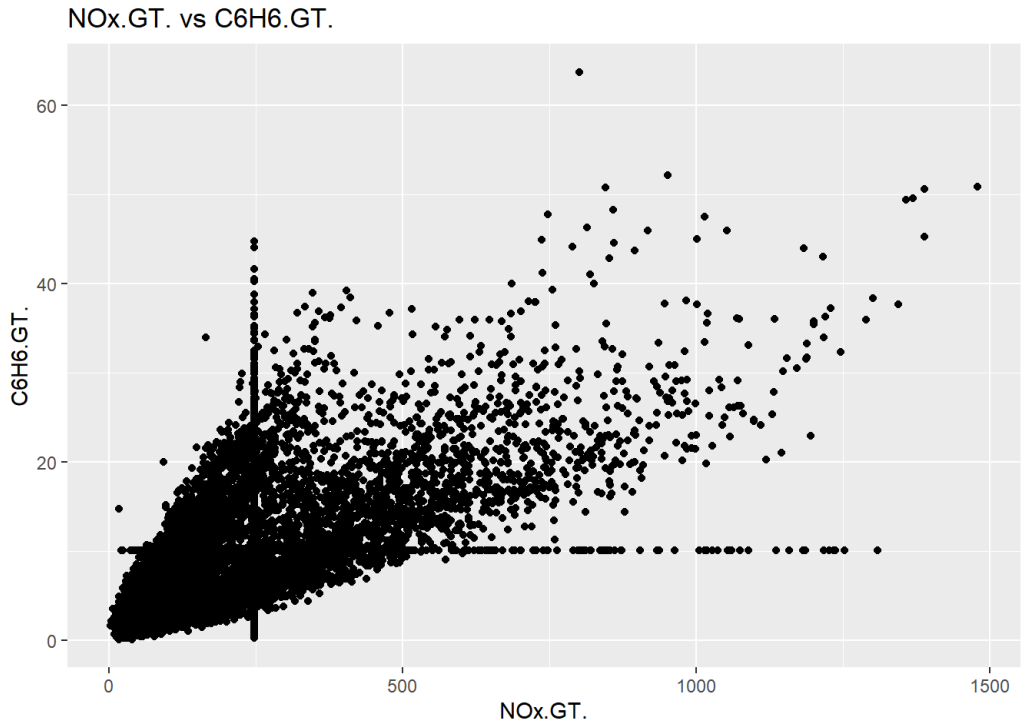
```
variables <- names(cleaned_data_no_time_eda)
for (var in variables) {
  if (var != "NOx.GT.") {
    p <- ggplot(cleaned_data_no_time_eda, aes_string(x="NOx.GT.", y=var)) +
        geom_point() +
        ggtitle(paste("NOx.GT. vs", var)) +
        xlab("NOx.GT.") +
        ylab(var)
      print(p)
  }
}
```

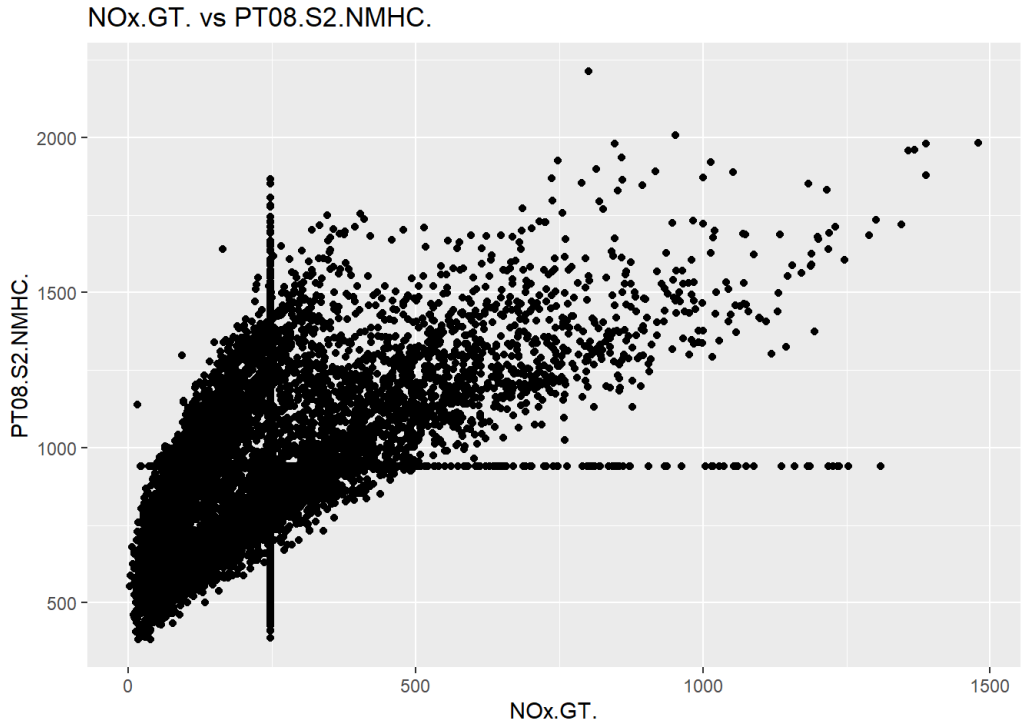
```
## Warning: `aes_string()` was deprecated in ggplot2 3.0.0.
## ℹ Please use tidy evaluation idioms with `aes()`.
## ℹ See also `vignette("ggplot2-in-packages")` for more information.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

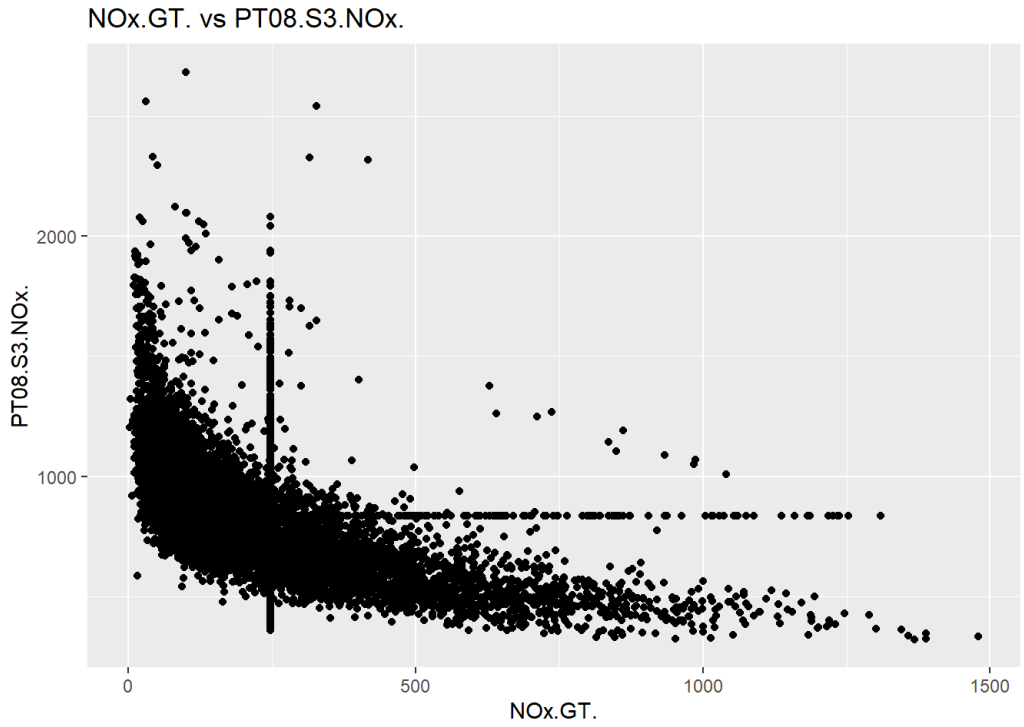

NOx.GT. vs CO.GT.
CO.GT.
10.0
7.5
5.0
2.5
0.0
0
500
1000
1500
NOx.GT.

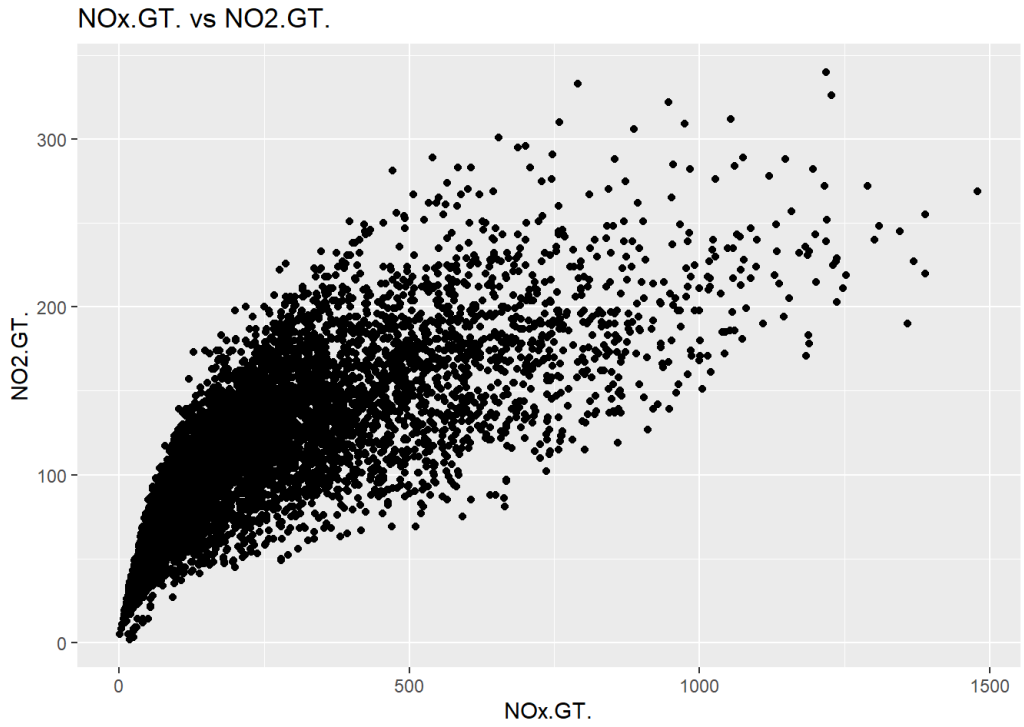

NOx.GT. vs PT08.S1.CO.
PT08.S1.CO.
2000
1500
1000
0
500
1000
1500
NOx.GT.

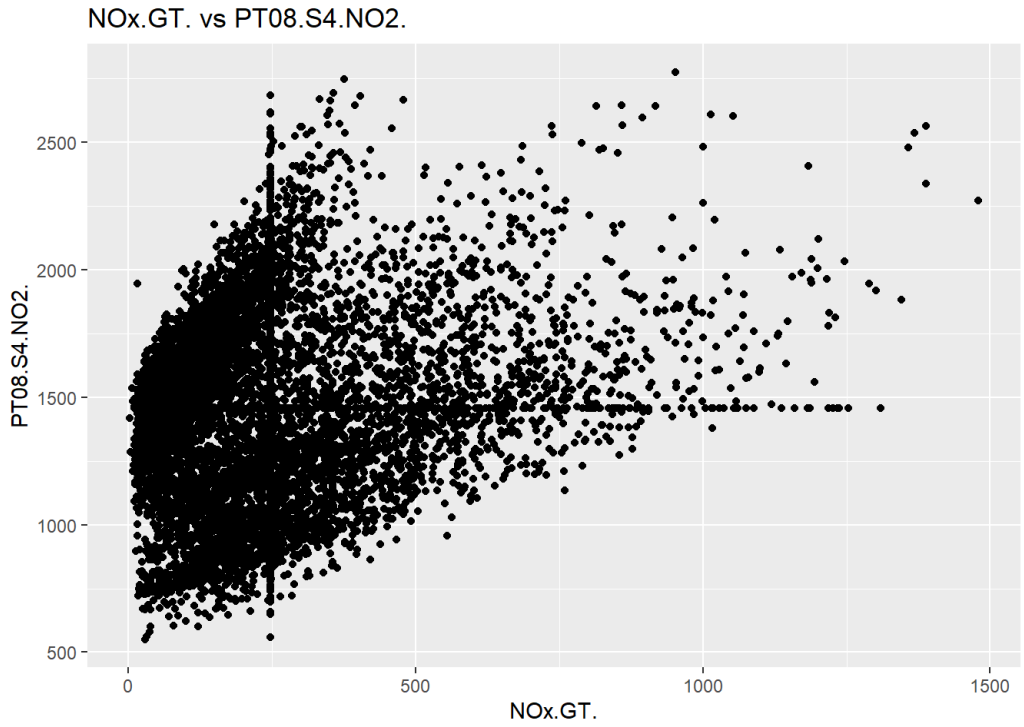

NOx.GT. vs NMHC.GT.
NMHC.GT.
1000
750
500
250
0
0
500
1000
1500
NOx.GT.

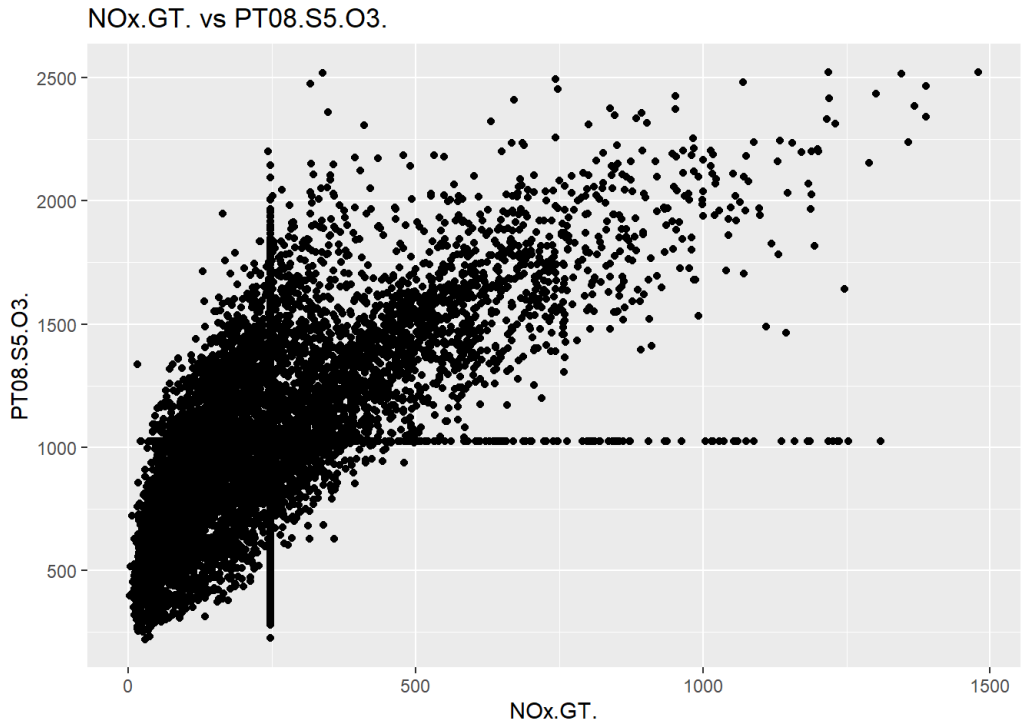

NOx.GT. vs C6H6.GT.
C6H6.GT.
60
40
20
0
0
500
1000
1500
NOx.GT.

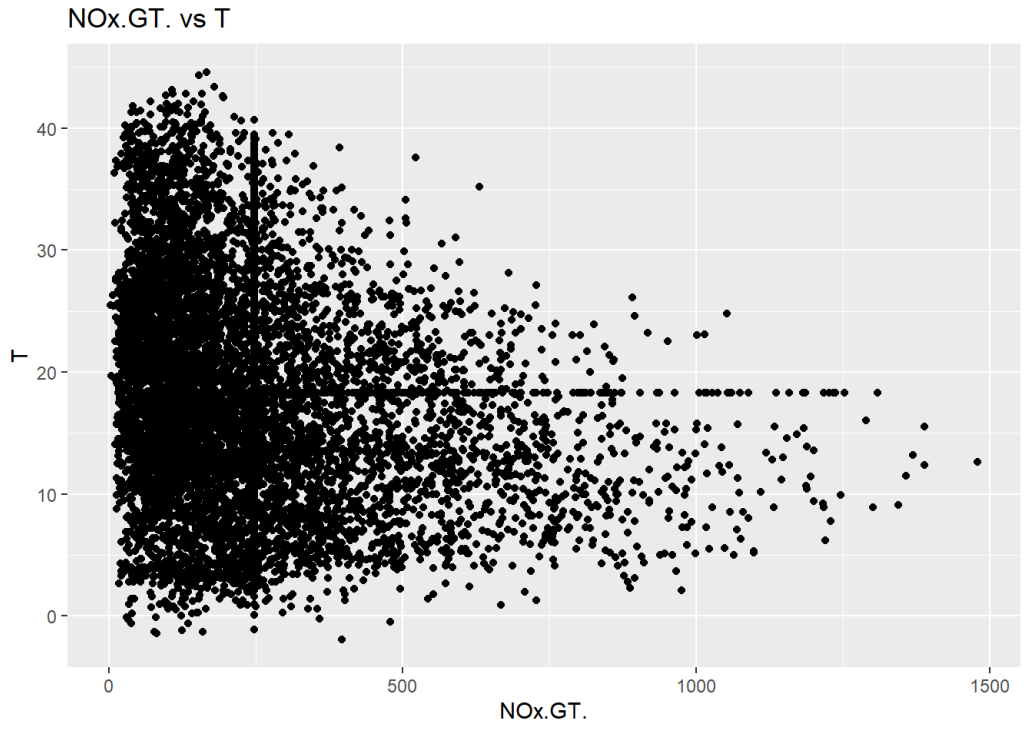

NOx.GT. vs PT08.S2.NMHC.
PT08.S2.NMHC.
2000
1500
1000
500
0
500
1000
1500
NOx.GT.

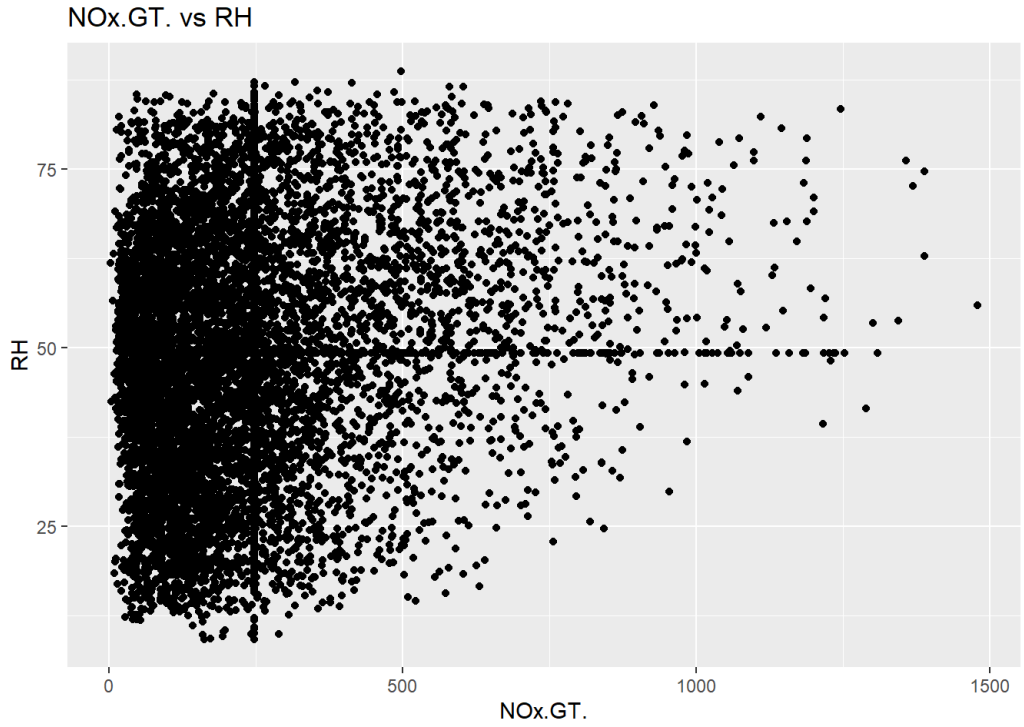

NOx.GT. vs PT08.S3.NOx.
PT08.S3.NOx.
2000
1000
0
500
1000
1500
NOx.GT.

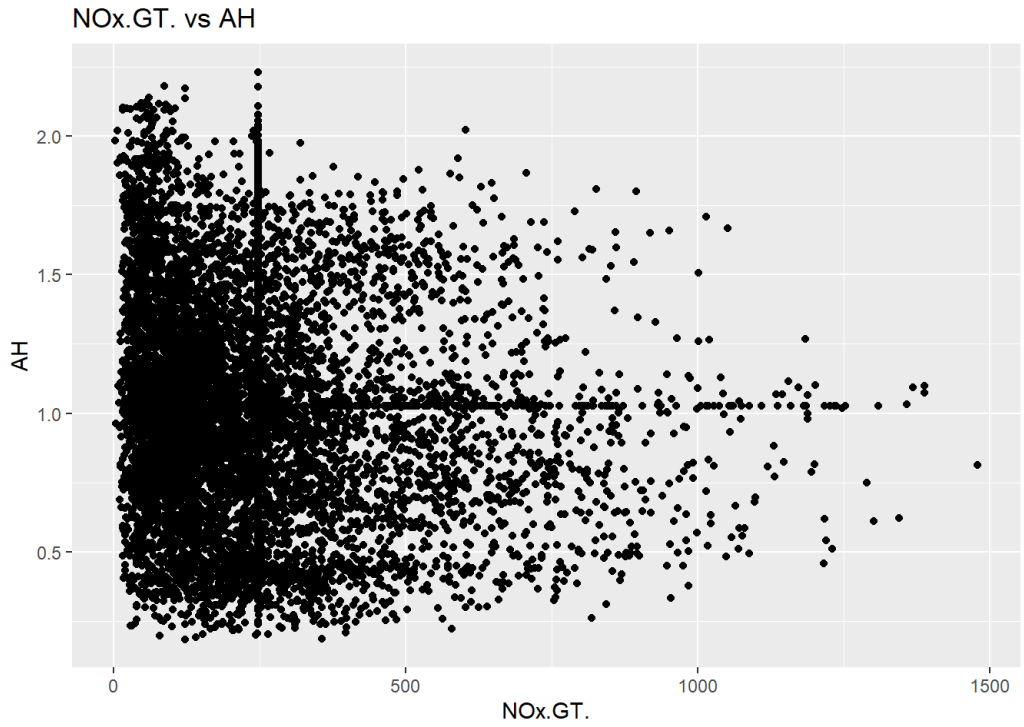

NOx.GT. vs NO2.GT.
NO2.GT.
300
200
100
0
0
500
1000
1500
NOx.GT.


NOx.GT. vs PT08.S4.NO2.
PT08.S4.NO2.
2500
2000
1500
1000
500
0
500
1000
1500
NOx.GT.


NOx.GT. vs PT08.S5.O3.
PT08.S5.O3.
2500
2000
1500
1000
500
0
500
1000
1500
NOx.GT.


NOx.GT. vs T
T
40
30
20
10
0
0
500
1000
1500
NOx.GT.

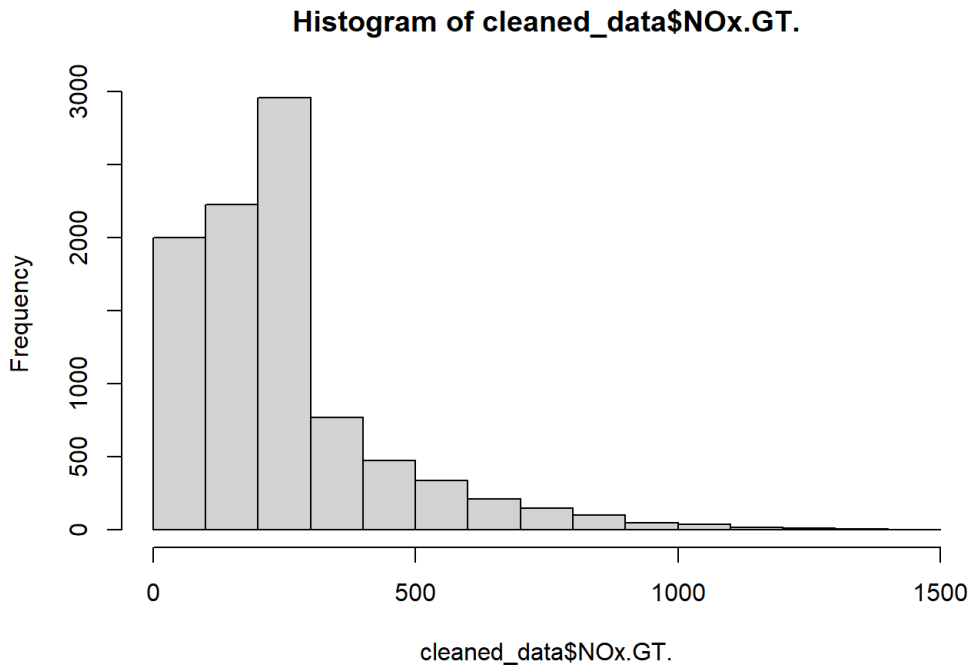




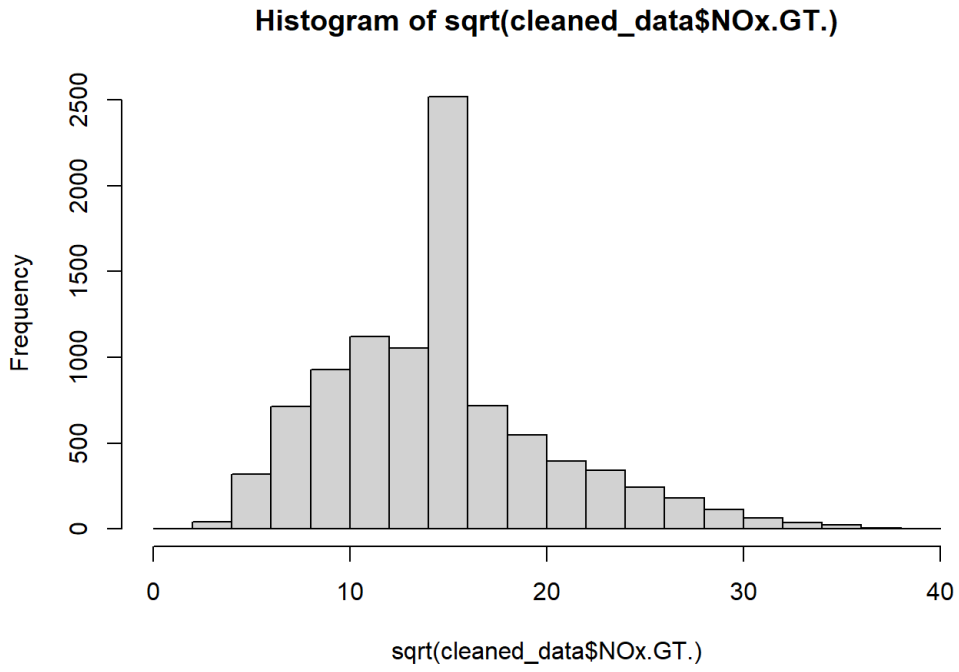
4. Distributions of response variables and statistical models

- Distribution of Response Variable and Transformed Response Variable:

```
hist(cleaned_data$NOx.GT.)
```




```
hist(sqrt(cleaned_data$NOx.GT.))
```



Model Building:

- Removal of Date and Time, and Transformation of Response Variable:

```
cleaned_data$NOx.GT. <- sqrt(cleaned_data$NOx.GT.)
cleaned_data_no_time <- select(cleaned_data, -Date, -Time)
```

- Stepwise Regression, both directions:
    - Null Model: One Feature
    - Full Model: All first-order features

```
null_model <- lm(NOx.GT. ~ PT08.S3.NOx., data = cleaned_data_no_time)
full_model <- lm(NOx.GT.~., data = cleaned_data_no_time)
step_model1 <- step(null_model, scope = list(lower = null_model, upper = full_model), direction = "both", test = "F")
```

```
## Start:  AIC=27876.94
## NOx.GT. ~ PT08.S3.NOx.
##
##                 Df Sum of Sq    RSS   AIC   F value    Pr(>F)
## + NO2.GT.        1     93564  90440 21233 9677.0838 < 2.2e-16 ***
## + CO.GT.         1     63614 120390 23910 4942.6277 < 2.2e-16 ***
## + PT08.S5.O3.    1     37422 146582 25751 2388.0703 < 2.2e-16 ***
## + T              1     33085 150919 26024 2050.5861 < 2.2e-16 ***
## + AH             1     25001 159003 26513 1470.7553 < 2.2e-16 ***
## + PT08.S1.CO.    1     17909 166095 26921 1008.5805 < 2.2e-16 ***
## + C6H6.GT.       1     17558 166446 26941  986.7300 < 2.2e-16 ***
## + PT08.S2.NMHC.  1     13858 170146 27146  761.8616 < 2.2e-16 ***
## + PT08.S4.NO2.   1      7123 176881 27510  376.6928 < 2.2e-16 ***
## + RH             1      6876 177128 27523  363.0905 < 2.2e-16 ***
## + NMHC.GT.       1       164 183840 27871    8.3484  0.003869 **
## <none>                       184004 27877
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=21232.95
## NOx.GT. ~ PT08.S3.NOx. + NO2.GT.
##
##                 Df Sum of Sq    RSS   AIC   F value    Pr(>F)
## + CO.GT.         1     15955  74486 19419 2003.3796 < 2.2e-16 ***
## + RH             1     15924  74516 19423 1998.7399 < 2.2e-16 ***
## + T              1      8652  81789 20294  989.3552 < 2.2e-16 ***
## + PT08.S5.O3.    1      8303  82137 20334  945.5274 < 2.2e-16 ***
## + C6H6.GT.       1      4875  85566 20717  532.8321 < 2.2e-16 ***
## + PT08.S1.CO.    1      4029  86411 20809  436.0844 < 2.2e-16 ***
## + PT08.S2.NMHC.  1      2788  87653 20942  297.4480 < 2.2e-16 ***
## + NMHC.GT.       1       244  90196 21210   25.3257 4.932e-07 ***
## + PT08.S4.NO2.   1       131  90309 21221   13.5288 0.0002362 ***
## + AH             1        21  90419 21233    2.1923 0.1387351
## <none>                       90440 21233
## - NO2.GT.        1     93564 184004 27877 9677.0838 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=19418.92
## NOx.GT. ~ PT08.S3.NOx. + NO2.GT. + CO.GT.
##
##                 Df Sum of Sq    RSS   AIC  F value    Pr(>F)
## + RH             1     13468  61017 17555 2064.260 < 2.2e-16 ***
## + T              1     10174  64312 18047 1479.425 < 2.2e-16 ***
## + PT08.S4.NO2.   1      8280  66205 18318 1169.644 < 2.2e-16 ***
## + NMHC.GT.       1      1584  72901 19220  203.217 < 2.2e-16 ***
## + PT08.S5.O3.    1      1160  73326 19274  147.927 < 2.2e-16 ***
## + PT08.S2.NMHC.  1      1135  73351 19277  144.690 < 2.2e-16 ***
## + C6H6.GT.       1       364  74121 19375   45.937 1.295e-11 ***
## + AH             1       210  74275 19395   26.475 2.723e-07 ***
## + PT08.S1.CO.    1        94  74391 19409   11.829 0.0005856 ***
## <none>                       74486 19419
## - CO.GT.         1     15955  90440 21233 2003.380 < 2.2e-16 ***
## - NO2.GT.        1     45905 120390 23910 5764.146 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=17554.68
## NOx.GT. ~ PT08.S3.NOx. + NO2.GT. + CO.GT. + RH
##
##                 Df Sum of Sq    RSS   AIC   F value    Pr(>F)
## + PT08.S4.NO2.   1      5033  55985 16751  840.5947 < 2.2e-16 ***
## + NMHC.GT.       1       890  60128 17419  138.3366 < 2.2e-16 ***
## + T              1       875  60143 17422  135.9814 < 2.2e-16 ***
## + AH             1       638  60379 17458   98.8834 < 2.2e-16 ***
## + PT08.S1.CO.    1       593  60424 17465   91.7615 < 2.2e-16 ***
## + PT08.S5.O3.    1       187  60830 17528   28.7379 8.484e-08 ***
## + C6H6.GT.       1        98  60920 17542   14.9900 0.0001088 ***
```

```
## + PT08.S2.NMHC.  1        20  60997 17554    3.0564 0.0804515 .
## <none>                       61017 17555
## - RH             1     13468  74486 19419 2064.2599 < 2.2e-16 ***
## - CO.GT.         1     13499  74516 19423 2068.9260 < 2.2e-16 ***
## - NO2.GT.        1     52934 113951 23397 8113.0222 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=16751.23
## NOx.GT. ~ PT08.S3.NOx. + NO2.GT. + CO.GT. + RH + PT08.S4.NO2.
##
##                 Df Sum of Sq   RSS   AIC  F value    Pr(>F)
## + PT08.S2.NMHC.  1    4709.0 51276 15931  858.674 < 2.2e-16 ***
## + C6H6.GT.       1    4127.7 51857 16037  744.235 < 2.2e-16 ***
## + PT08.S5.O3.    1    1323.9 54661 16529  226.459 < 2.2e-16 ***
## + NMHC.GT.       1     848.5 55136 16610  143.885 < 2.2e-16 ***
## + AH             1     698.1 55287 16636  118.056 < 2.2e-16 ***
## + T              1     588.5 55396 16654   99.323 < 2.2e-16 ***
## + PT08.S1.CO.    1      72.2 55912 16741   12.075 0.0005135 ***
## <none>                      55985 16751
## - PT08.S4.NO2.   1    5032.7 61017 17555  840.595 < 2.2e-16 ***
## - RH             1   10220.7 66205 18318 1707.148 < 2.2e-16 ***
## - CO.GT.         1   18524.1 74509 19424 3094.046 < 2.2e-16 ***
## - NO2.GT.        1   27119.6 83104 20445 4529.738 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=15931.11
## NOx.GT. ~ PT08.S3.NOx. + NO2.GT. + CO.GT. + RH + PT08.S4.NO2. +
##     PT08.S2.NMHC.
##
##                 Df Sum of Sq   RSS   AIC   F value  Pr(>F)
## + AH             1    6285.6 44990 14709 1306.1616 < 2e-16 ***
## + T              1    4783.6 46492 15017  961.9205 < 2e-16 ***
## + PT08.S1.CO.    1     877.4 50398 15772  162.7633 < 2e-16 ***
## + NMHC.GT.       1     707.2 50568 15803  130.7440 < 2e-16 ***
## + PT08.S5.O3.    1      19.2 51256 15930    3.4939 0.06163 .
## + C6H6.GT.       1      19.1 51257 15930    3.4793 0.06217 .
## <none>                      51276 15931
## - PT08.S2.NMHC.  1    4709.0 55985 16751  858.6739 < 2e-16 ***
## - CO.GT.         1    8127.1 59403 17306 1481.9547 < 2e-16 ***
## - PT08.S4.NO2.   1    9721.7 60997 17554 1772.7324 < 2e-16 ***
## - RH             1   13517.2 64793 18119 2464.8379 < 2e-16 ***
## - NO2.GT.        1   20910.1 72186 19130 3812.9103 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=14709.44
## NOx.GT. ~ PT08.S3.NOx. + NO2.GT. + CO.GT. + RH + PT08.S4.NO2. +
##     PT08.S2.NMHC. + AH
##
##                 Df Sum of Sq   RSS   AIC   F value    Pr(>F)
## + NMHC.GT.       1     329.0 44661 14643   68.8644 < 2.2e-16 ***
## + T              1     128.8 44861 14685   26.8483 2.247e-07 ***
## + PT08.S1.CO.    1      75.8 44914 14696   15.7755 7.185e-05 ***
## + PT08.S5.O3.    1      31.4 44959 14705    6.5210   0.01068 *
## <none>                      44990 14709
## + C6H6.GT.       1       7.6 44982 14710    1.5722   0.20993
## - AH             1    6285.6 51276 15931 1306.1616 < 2.2e-16 ***
## - CO.GT.         1    8177.1 53167 16270 1699.2068 < 2.2e-16 ***
## - PT08.S2.NMHC.  1   10296.5 55287 16636 2139.6376 < 2.2e-16 ***
## - RH             1   11159.3 56149 16781 2318.9111 < 2.2e-16 ***
## - PT08.S4.NO2.   1   15339.7 60330 17453 3187.6196 < 2.2e-16 ***
## - NO2.GT.        1   24286.2 69276 18747 5046.7183 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=14642.77
```

```
## NOx.GT. ~ PT08.S3.NOx. + NO2.GT. + CO.GT. + RH + PT08.S4.NO2. +
##     PT08.S2.NMHC. + AH + NMHC.GT.
##
##                 Df Sum of Sq   RSS   AIC   F value    Pr(>F)
## + T              1     113.0 44548 14621   23.7086 1.139e-06 ***
## + PT08.S1.CO.    1      65.3 44596 14631   13.6784 0.0002182 ***
## + PT08.S5.O3.    1      26.6 44634 14639    5.5599 0.0183968 *
## + C6H6.GT.       1      14.6 44646 14642    3.0622 0.0801646 .
## <none>                       44661 14643
## - NMHC.GT.       1     329.0 44990 14709   68.8644 < 2.2e-16 ***
## - AH             1    5907.4 50568 15803 1236.4855 < 2.2e-16 ***
## - CO.GT.         1    8493.2 53154 16270 1777.7220 < 2.2e-16 ***
## - PT08.S2.NMHC.  1    9874.8 54536 16510 2066.8873 < 2.2e-16 ***
## - RH             1   10798.8 55460 16667 2260.2921 < 2.2e-16 ***
## - PT08.S4.NO2.   1   14714.1 59375 17305 3079.8095 < 2.2e-16 ***
## - NO2.GT.        1   23668.0 68329 18620 4953.9478 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=14621.06
## NOx.GT. ~ PT08.S3.NOx. + NO2.GT. + CO.GT. + RH + PT08.S4.NO2. +
##     PT08.S2.NMHC. + AH + NMHC.GT. + T
##
##                 Df Sum of Sq   RSS   AIC  F value    Pr(>F)
## + PT08.S1.CO.    1      53.7 44494 14612   11.280 0.0007866 ***
## + PT08.S5.O3.    1      50.1 44498 14612   10.513 0.0011898 **
## + C6H6.GT.       1      35.1 44513 14616    7.375 0.0066257 **
## <none>                       44548 14621
## - T              1     113.0 44661 14643   23.709 1.139e-06 ***
## - NMHC.GT.       1     313.2 44861 14685   65.706 5.892e-16 ***
## - AH             1    1577.8 46126 14945  331.043 < 2.2e-16 ***
## - RH             1    2568.0 47116 15144  538.807 < 2.2e-16 ***
## - CO.GT.         1    8579.8 53128 16267 1800.195 < 2.2e-16 ***
## - PT08.S2.NMHC.  1    9980.3 54528 16511 2094.057 < 2.2e-16 ***
## - PT08.S4.NO2.   1   14510.9 59059 17257 3044.667 < 2.2e-16 ***
## - NO2.GT.        1   23578.3 68126 18594 4947.166 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=14611.78
## NOx.GT. ~ PT08.S3.NOx. + NO2.GT. + CO.GT. + RH + PT08.S4.NO2. +
##     PT08.S2.NMHC. + AH + NMHC.GT. + T + PT08.S1.CO.
##
##                 Df Sum of Sq   RSS   AIC   F value    Pr(>F)
## + PT08.S5.O3.    1     109.5 44385 14591   23.0608 1.594e-06 ***
## + C6H6.GT.       1      41.8 44453 14605    8.7876 0.0030405 **
## <none>                       44494 14612
## - PT08.S1.CO.    1      53.7 44548 14621   11.2800 0.0007866 ***
## - T              1     101.4 44596 14631   21.3065 3.966e-06 ***
## - NMHC.GT.       1     304.6 44799 14674   63.9789 1.407e-15 ***
## - AH             1    1492.5 45987 14918  313.5042 < 2.2e-16 ***
## - RH             1    2619.7 47114 15145  550.2605 < 2.2e-16 ***
## - CO.GT.         1    8624.6 53119 16268 1811.5895 < 2.2e-16 ***
## - PT08.S2.NMHC.  1    9442.3 53937 16411 1983.3559 < 2.2e-16 ***
## - PT08.S4.NO2.   1   13187.8 57682 17039 2770.0811 < 2.2e-16 ***
## - NO2.GT.        1   23614.5 68109 18594 4960.2205 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=14590.71
## NOx.GT. ~ PT08.S3.NOx. + NO2.GT. + CO.GT. + RH + PT08.S4.NO2. +
##     PT08.S2.NMHC. + AH + NMHC.GT. + T + PT08.S1.CO. + PT08.S5.O3.
##
##                 Df Sum of Sq   RSS   AIC   F value    Pr(>F)
## + C6H6.GT.       1      42.4 44342 14584    8.9368  0.002802 **
## <none>                       44385 14591
## - PT08.S5.O3.    1     109.5 44494 14612   23.0608 1.594e-06 ***
## - PT08.S1.CO.    1     113.2 44498 14612   23.8292 1.070e-06 ***
```

```
## - T              1     134.0 44519 14617   28.2223 1.106e-07 ***
## - NMHC.GT.       1     287.0 44672 14649   60.4335 8.415e-15 ***
## - AH             1    1396.4 45781 14879  294.0084 < 2.2e-16 ***
## - RH             1    2622.6 47007 15126  552.1741 < 2.2e-16 ***
## - PT08.S2.NMHC.  1    7464.1 51849 16043 1571.5396 < 2.2e-16 ***
## - CO.GT.         1    8671.2 53056 16258 1825.6868 < 2.2e-16 ***
## - PT08.S4.NO2.   1   12932.1 57317 16981 2722.7889 < 2.2e-16 ***
## - NO2.GT.        1   22528.3 66913 18430 4743.2247 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=14583.77
## NOx.GT. ~ PT08.S3.NOx. + NO2.GT. + CO.GT. + RH + PT08.S4.NO2. +
##     PT08.S2.NMHC. + AH + NMHC.GT. + T + PT08.S1.CO. + PT08.S5.O3. +
##     C6H6.GT.
##
##                 Df Sum of Sq   RSS   AIC   F value    Pr(>F)
## <none>                       44342 14584
## - C6H6.GT.       1      42.4 44385 14591    8.9368  0.002802 **
## - PT08.S5.O3.    1     110.1 44453 14605   23.2087 1.476e-06 ***
## - PT08.S1.CO.    1     122.3 44465 14608   25.7616 3.937e-07 ***
## - T              1     157.9 44500 14615   33.2799 8.234e-09 ***
## - NMHC.GT.       1     297.4 44640 14644   62.6684 2.725e-15 ***
## - AH             1    1275.0 45617 14847  268.6821 < 2.2e-16 ***
## - PT08.S2.NMHC.  1    1620.6 45963 14918  341.5057 < 2.2e-16 ***
## - RH             1    2663.2 47006 15128  561.2101 < 2.2e-16 ***
## - CO.GT.         1    7945.9 52288 16124 1674.3902 < 2.2e-16 ***
## - PT08.S4.NO2.   1   12754.2 57097 16947 2687.6065 < 2.2e-16 ***
## - NO2.GT.        1   22389.5 66732 18406 4718.0100 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(step_model1)
```

```
##
## Call:
## lm(formula = NOx.GT. ~ PT08.S3.NOx. + NO2.GT. + CO.GT. + RH +
##     PT08.S4.NO2. + PT08.S2.NMHC. + AH + NMHC.GT. + T + PT08.S1.CO. +
##     PT08.S5.O3. + C6H6.GT., data = cleaned_data_no_time)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.7864  -1.4108  -0.1461   1.2807  10.6809
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.3194251  0.5981440   0.534   0.5933
## PT08.S3.NOx.   0.0018002  0.0002119   8.496  < 2e-16 ***
## NO2.GT.        0.0602139  0.0008766  68.688  < 2e-16 ***
## CO.GT.         1.3981167  0.0341676  40.919  < 2e-16 ***
## RH             0.0878062  0.0037065  23.690  < 2e-16 ***
## PT08.S4.NO2.  -0.0113221  0.0002184 -51.842  < 2e-16 ***
## PT08.S2.NMHC.  0.0127691  0.0006910  18.480  < 2e-16 ***
## AH             3.0042782  0.1832824  16.392  < 2e-16 ***
## NMHC.GT.      -0.0029856  0.0003771  -7.916 2.72e-15 ***
## T              0.0569781  0.0098768   5.769 8.23e-09 ***
## PT08.S1.CO.   -0.0015387  0.0003031  -5.076 3.94e-07 ***
## PT08.S5.O3.    0.0008082  0.0001678   4.818 1.48e-06 ***
## C6H6.GT.       0.0588212  0.0196763   2.989   0.0028 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.178 on 9344 degrees of freedom
## Multiple R-squared:  0.8506, Adjusted R-squared:  0.8504
## F-statistic:  4432 on 12 and 9344 DF,  p-value: < 2.2e-16
```

```
step_model2 <- step(full_model, scope = list(lower = null_model, upper = full_model), direction = "both",test="F")
```

```
## Start:  AIC=14583.77
## NOx.GT. ~ CO.GT. + PT08.S1.CO. + NMHC.GT. + C6H6.GT. + PT08.S2.NMHC. +
##     PT08.S3.NOx. + NO2.GT. + PT08.S4.NO2. + PT08.S5.O3. + T +
##     RH + AH
##
##                 Df Sum of Sq   RSS   AIC   F value    Pr(>F)
## <none>                       44342 14584
## - C6H6.GT.       1      42.4 44385 14591    8.9368  0.002802 **
## - PT08.S5.O3.    1     110.1 44453 14605   23.2087 1.476e-06 ***
## - PT08.S1.CO.    1     122.3 44465 14608   25.7616 3.937e-07 ***
## - T              1     157.9 44500 14615   33.2799 8.234e-09 ***
## - NMHC.GT.       1     297.4 44640 14644   62.6684 2.725e-15 ***
## - AH             1    1275.0 45617 14847  268.6821 < 2.2e-16 ***
## - PT08.S2.NMHC.  1    1620.6 45963 14918  341.5057 < 2.2e-16 ***
## - RH             1    2663.2 47006 15128  561.2101 < 2.2e-16 ***
## - CO.GT.         1    7945.9 52288 16124 1674.3902 < 2.2e-16 ***
## - PT08.S4.NO2.   1   12754.2 57097 16947 2687.6065 < 2.2e-16 ***
## - NO2.GT.        1   22389.5 66732 18406 4718.0100 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(step_model2)
```

```
##
## Call:
## lm(formula = NOx.GT. ~ CO.GT. + PT08.S1.CO. + NMHC.GT. + C6H6.GT. +
##     PT08.S2.NMHC. + PT08.S3.NOx. + NO2.GT. + PT08.S4.NO2. + PT08.S5.O3. +
##     T + RH + AH, data = cleaned_data_no_time)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.7864  -1.4108  -0.1461   1.2807  10.6809
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.3194251  0.5981440   0.534   0.5933
## CO.GT.         1.3981167  0.0341676  40.919  < 2e-16 ***
## PT08.S1.CO.   -0.0015387  0.0003031  -5.076 3.94e-07 ***
## NMHC.GT.      -0.0029856  0.0003771  -7.916 2.72e-15 ***
## C6H6.GT.       0.0588212  0.0196763   2.989   0.0028 **
## PT08.S2.NMHC.  0.0127691  0.0006910  18.480  < 2e-16 ***
## PT08.S3.NOx.   0.0018002  0.0002119   8.496  < 2e-16 ***
## NO2.GT.        0.0602139  0.0008766  68.688  < 2e-16 ***
## PT08.S4.NO2.  -0.0113221  0.0002184 -51.842  < 2e-16 ***
## PT08.S5.O3.    0.0008082  0.0001678   4.818 1.48e-06 ***
## T              0.0569781  0.0098768   5.769 8.23e-09 ***
## RH             0.0878062  0.0037065  23.690  < 2e-16 ***
## AH             3.0042782  0.1832824  16.392  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.178 on 9344 degrees of freedom
## Multiple R-squared:  0.8506, Adjusted R-squared:  0.8504
## F-statistic:  4432 on 12 and 9344 DF,  p-value: < 2.2e-16
```

```
cleaned_data_no_time_all <- regsubsets(NOx.GT. ~ ., data = cleaned_data_no_time, nbest = 1, nvmax = 15)
summary(cleaned_data_no_time_all)$which
```

```
##    (Intercept) CO.GT. PT08.S1.CO. NMHC.GT. C6H6.GT. PT08.S2.NMHC. PT08.S3.NOx.
## 1         TRUE  FALSE       FALSE    FALSE    FALSE         FALSE        FALSE
## 2         TRUE   TRUE       FALSE    FALSE    FALSE         FALSE        FALSE
## 3         TRUE   TRUE       FALSE    FALSE    FALSE         FALSE        FALSE
## 4         TRUE  FALSE       FALSE    FALSE    FALSE          TRUE        FALSE
## 5         TRUE   TRUE       FALSE    FALSE    FALSE          TRUE        FALSE
## 6         TRUE   TRUE       FALSE    FALSE    FALSE          TRUE        FALSE
## 7         TRUE   TRUE       FALSE    FALSE    FALSE          TRUE         TRUE
## 8         TRUE   TRUE       FALSE     TRUE    FALSE          TRUE         TRUE
## 9         TRUE   TRUE       FALSE     TRUE    FALSE          TRUE         TRUE
## 10        TRUE   TRUE        TRUE     TRUE    FALSE          TRUE         TRUE
## 11        TRUE   TRUE        TRUE     TRUE    FALSE          TRUE         TRUE
## 12        TRUE   TRUE        TRUE     TRUE     TRUE          TRUE         TRUE
##    NO2.GT. PT08.S4.NO2. PT08.S5.O3.     T    RH    AH
## 1     TRUE        FALSE       FALSE FALSE FALSE FALSE
## 2     TRUE        FALSE       FALSE FALSE FALSE FALSE
## 3     TRUE        FALSE       FALSE FALSE  TRUE FALSE
## 4     TRUE         TRUE       FALSE FALSE  TRUE FALSE
## 5     TRUE         TRUE       FALSE FALSE  TRUE FALSE
## 6     TRUE         TRUE       FALSE FALSE  TRUE  TRUE
## 7     TRUE         TRUE       FALSE FALSE  TRUE  TRUE
## 8     TRUE         TRUE       FALSE FALSE  TRUE  TRUE
## 9     TRUE         TRUE       FALSE  TRUE  TRUE  TRUE
## 10    TRUE         TRUE       FALSE  TRUE  TRUE  TRUE
## 11    TRUE         TRUE        TRUE  TRUE  TRUE  TRUE
## 12    TRUE         TRUE        TRUE  TRUE  TRUE  TRUE
```

```
summary(cleaned_data_no_time_all)$rsq
```

```
##  [1] 0.6602932 0.7424127 0.7904025 0.7996255 0.8266089 0.8459038 0.8483843
##  [8] 0.8494930 0.8498738 0.8500548 0.8504239 0.8505668
```

```
summary(cleaned_data_no_time_all)$adjr2
```

```
##  [1] 0.6602569 0.7423576 0.7903353 0.7995398 0.8265162 0.8458049 0.8482708
##  [8] 0.8493642 0.8497293 0.8498944 0.8502479 0.8503749
```

- Multicollinearity:

```
model = lm(formula = NOx.GT. ~ CO.GT. + PT08.S1.CO. + NMHC.GT. + C6H6.GT. +
    PT08.S2.NMHC. + PT08.S3.NOx. + NO2.GT. + PT08.S4.NO2. + PT08.S5.O3. +
    T + RH + AH, data = cleaned_data_no_time)
vif(model)
```

```
##        CO.GT.   PT08.S1.CO.      NMHC.GT.      C6H6.GT. PT08.S2.NMHC.
##      3.986495      8.204023      1.143953     40.705228     64.397289
##  PT08.S3.NOx.       NO2.GT.  PT08.S4.NO2.   PT08.S5.O3.             T
##      5.609734      2.922687     10.830157      8.466973     14.415724
##            RH            AH
##      7.804379     10.377074
```

```
# Remove C6H6.GT.
model1 = lm(formula = NOx.GT. ~ CO.GT. + PT08.S1.CO. + NMHC.GT. +
    PT08.S2.NMHC. + PT08.S3.NOx. + NO2.GT. + PT08.S4.NO2. + PT08.S5.O3. +
    T + RH + AH, data = cleaned_data_no_time)
vif(model1)
```

```
##        CO.GT.   PT08.S1.CO.      NMHC.GT. PT08.S2.NMHC.  PT08.S3.NOx.
##      3.776546      8.166078      1.141120     18.081630      4.771462
##       NO2.GT.  PT08.S4.NO2.   PT08.S5.O3.             T            RH
##      2.871233     10.771777      8.466803     13.941246      7.725311
##            AH
##     10.074487
```

```
# Remove Sensor for NMHC
model2 = lm(formula = NOx.GT. ~ CO.GT. + PT08.S1.CO. + NMHC.GT. + PT08.S3.NOx. + NO2.GT. + PT08.S4.NO2. + PT08.S5.O3. +
    T + RH + AH, data = cleaned_data_no_time)
vif(model2)
```

```
##       CO.GT.  PT08.S1.CO.     NMHC.GT. PT08.S3.NOx.      NO2.GT. PT08.S4.NO2.
##     3.424494     8.015084     1.134618     3.757300     2.871107     6.619325
##  PT08.S5.O3.            T           RH           AH
##     7.306677    13.939967     7.543902     9.388079
```

```
# Remove Temp
model3 = lm(formula = NOx.GT. ~ CO.GT. + PT08.S1.CO. + NMHC.GT. + PT08.S3.NOx. + NO2.GT. + PT08.S4.NO2. + PT08.S5.O3. + RH +
AH, data = cleaned_data_no_time)
vif(model3)
```

```
##       CO.GT.  PT08.S1.CO.     NMHC.GT. PT08.S3.NOx.      NO2.GT. PT08.S4.NO2.
##     3.414495     8.014486     1.132275     3.746751     2.863181     6.036886
##  PT08.S5.O3.           RH           AH
##     7.110782     1.379819     3.482655
```

```
# Remove CO sensor
model4 = lm(formula = NOx.GT. ~ CO.GT. + NMHC.GT. + PT08.S3.NOx. + NO2.GT. + PT08.S4.NO2. + PT08.S5.O3. + RH + AH, data = cl
eaned_data_no_time)
vif(model4)
```

```
##       CO.GT.     NMHC.GT. PT08.S3.NOx.      NO2.GT. PT08.S4.NO2.  PT08.S5.O3.
##     3.322005     1.130682     3.478582     2.862893     4.758476     5.617012
##           RH           AH
##     1.307883     3.134334
```

```
# Remove O3 sensor
model5 = lm(formula = NOx.GT. ~ CO.GT. + NMHC.GT. + PT08.S3.NOx. + NO2.GT. + PT08.S4.NO2. + RH + AH, data = cleaned_data_no_
time)
vif(model5)
```

```
##       CO.GT.     NMHC.GT. PT08.S3.NOx.      NO2.GT. PT08.S4.NO2.           RH
##     3.177519     1.125979     2.334402     2.763502     3.630076     1.148650
##           AH
##     2.589095
```

- VIF and Summary of Final Model:

```
summary(model5)
```

```
##
## Call:
## lm(formula = NOx.GT. ~ CO.GT. + NMHC.GT. + PT08.S3.NOx. + NO2.GT. +
##     PT08.S4.NO2. + RH + AH, data = cleaned_data_no_time)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.1562  -1.6343  -0.1493   1.4492  14.9550
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.4505055  0.3288208   28.74   <2e-16 ***
## CO.GT.        1.9671696  0.0338204   58.16   <2e-16 ***
## NMHC.GT.     -0.0047063  0.0004148  -11.35   <2e-16 ***
## PT08.S3.NOx. -0.0038122  0.0001515  -25.16   <2e-16 ***
## NO2.GT.       0.0636361  0.0009451   67.33   <2e-16 ***
## PT08.S4.NO2. -0.0040552  0.0001402  -28.93   <2e-16 ***
## RH            0.0579147  0.0015765   36.74   <2e-16 ***
## AH            1.0297320  0.1015015   10.14   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.415 on 9349 degrees of freedom
## Multiple R-squared:  0.8162, Adjusted R-squared:  0.8161
## F-statistic:  5931 on 7 and 9349 DF,  p-value: < 2.2e-16
```

```
vif(model5)
```

```
##       CO.GT.     NMHC.GT. PT08.S3.NOx.      NO2.GT. PT08.S4.NO2.           RH
##     3.177519     1.125979     2.334402     2.763502     3.630076     1.148650
##           AH
##     2.589095
```

5. Rationale of the fitted model.

Final Model:

$$\sqrt{\mathrm{NOx(GT)}} = 9.4505 + 1.9672 \cdot X_1 - 0.0047 \times \cdot X_2 - 0.0038 \times \cdot X_3 + 0.0636 \times \cdot X_4 - 0.0041 \times \cdot X_5 + 0.0579 \times \cdot X_6 + 1.0297 \times \cdot X$$

Where,

X_1 = $\mathrm{CO(GT)}$: For every unit of increase in Carbon Monoxide concentration, the square root of NOx concentration is increased by 1.9672.

X_2 = $\mathrm{NMHC(GT)}$: For every unit of increase in Non-Methane Hydrocarbons concentration , the square root of NOx concentration is decreased by 0.0047.

X_3 = $\mathrm{PT08.S3(NOx)}$: For every unit of increase in sensor response for NOx, the square root of NOx concentration is decreased by 0.0038.

X_4 = $NO_2\mathrm{(GT)}$: For every unit of increase in Nitrogen Dioxide concentration, the square root of NOx concentration is increased by 0.0636

X_5 = $\mathrm{PT08.S4(NO2)}$: For every unit of increase in sensor response for NO2, the square root of NOx concentration is decreased by 0.0041.

X_6 = RH: For every unit of increase in relative humidity, the square root of NOx concentration is increased by 0.0579

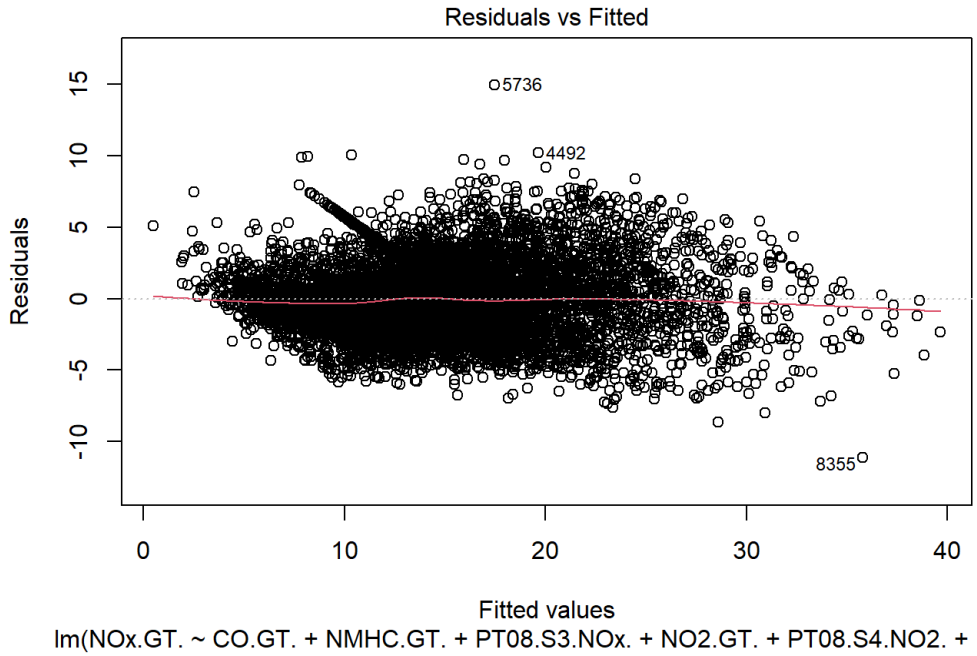
X_7 = AH: For every unit of increase in absolute humidity, the square root of NOx concentration is increased by 1.0297

6. Results of the data analysis, including tables and figures.

Model Constant Variance Assumptions:

- Constant Variance and Normal Residuals Satisfied

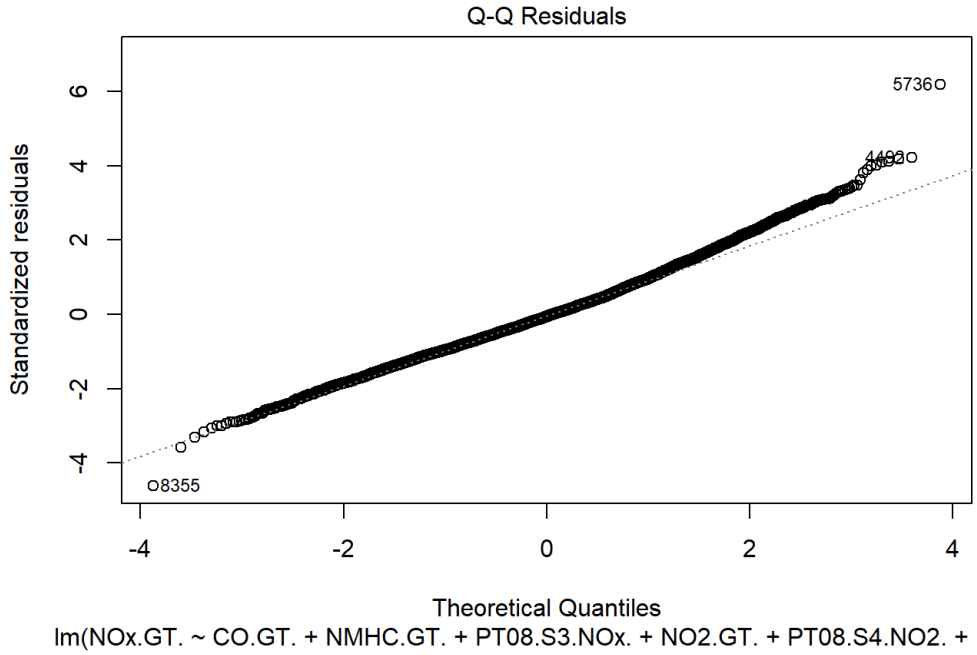
```
plot(model5, which = 1)
```



Normality Assumption:

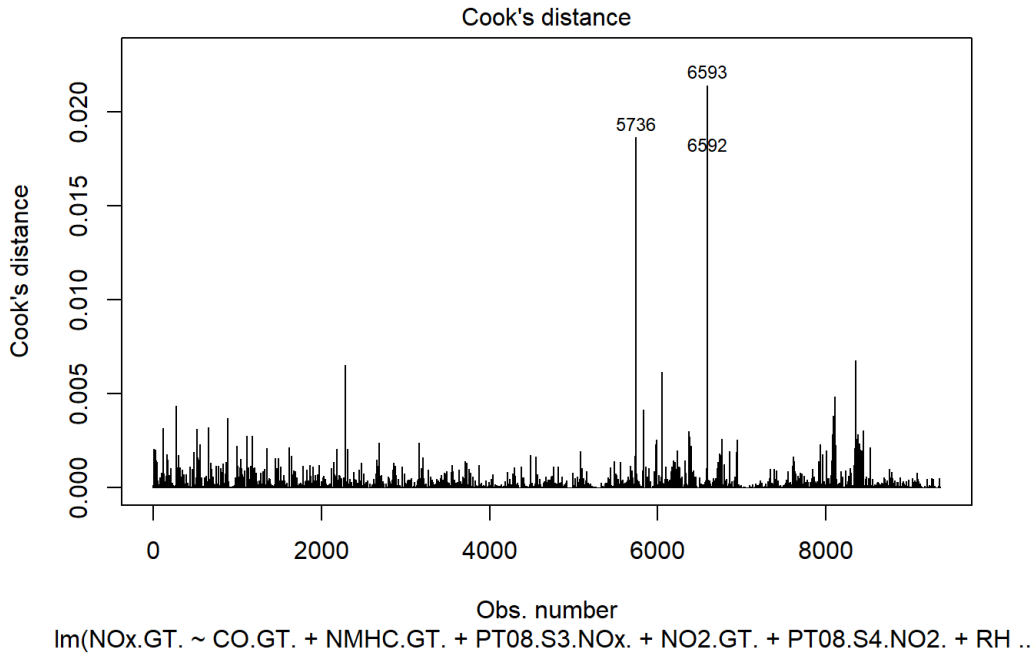
- Normality Assumption Satisfied

```
plot(model5, which = 2)
```



- **Outlier:**

```
plot(model5, which = 4)
```



7. Conclusions and discussion.

The model had a high adjusted R^2 value of 0.8161, meaning that 81% of the variation in NOx concentration is explained by the predictors. The strongest predictors were CO concentration and humidity levels. Predicting NOx concentration relied on three sensors, including itself, the concentration of other particles that contained oxygen, and humidity levels. At one point, it was theorized that NOx's negative correlation with absolute humidity is explained by conditions that existed at higher humidity levels, such as increased increased formation of clouds due to more increased water vapor volume; this can cause the NOx concentration to disperse as a result. This result makes sense, however, NOx is positively correlated with relative humidity, which conflicts with the previous learning. The model is only effective in polluted areas similar to condition from this dataset. It is difficult to generalize the model as the data was taken from highly polluted areas in Italy. The conditions in other areas may affect the model's ability to accurate predict NOx concentration.

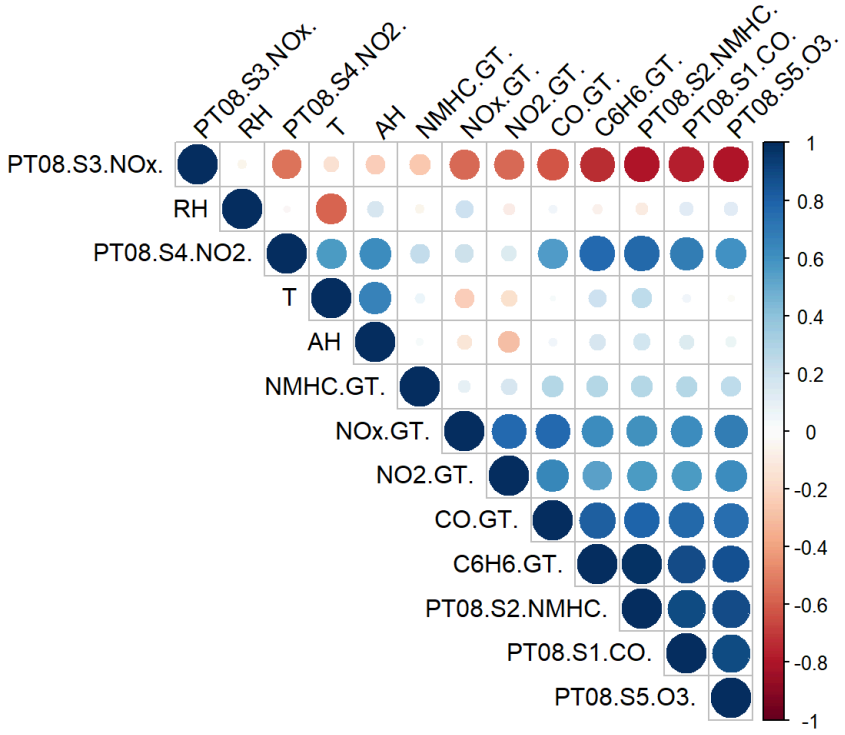
8. Appendix: References and Program code (SAS or R).

- (Study 1): Response of biogenic secondary organic aerosol formation to anthropogenic NOx emission mitigation (https://www.sciencedirect.com/science/article/pii/S004896972402285X)
    - Li, J., Chen, T., Zhang, H., Jia, Y., Chu, Y., Yan, Y., Zhang, H., Ren, Y., Li, H., Hu, J., Wang, W., Chu, B., Ge, M., & He, H. (2024). Nonlinear effect of NO concentration decrease on secondary aerosol formation in the Beijing-Tianjin-Hebei region: Evidence from smog chamber experiments and field observations. *Science of the Total Environment*, *912*, 168333. https://doi.org/10.1016/j.scitotenv.2023.168333 (https://doi.org/10.1016/j.scitotenv.2023.168333)

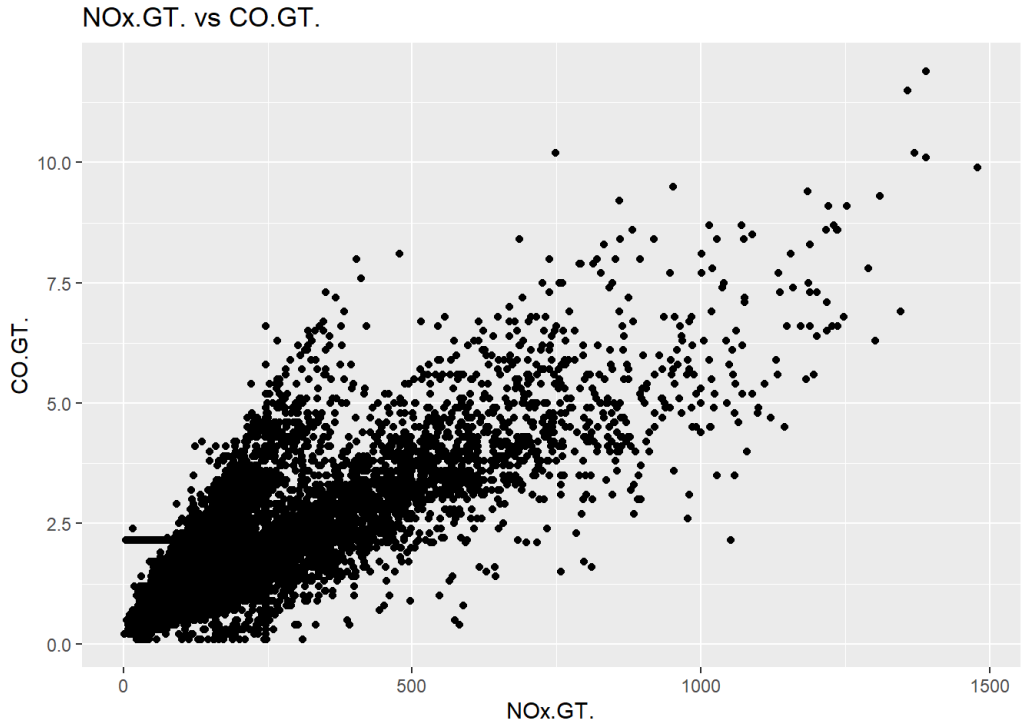

```
library(dplyr)
library(tidyverse)
library(ggplot2)
library(gridExtra)
library(leaps)
library(car)
library(corrplot)

data <- read.csv("AirQualityUCI (1).csv", header=T)

summary(data$NOx.GT.)
```

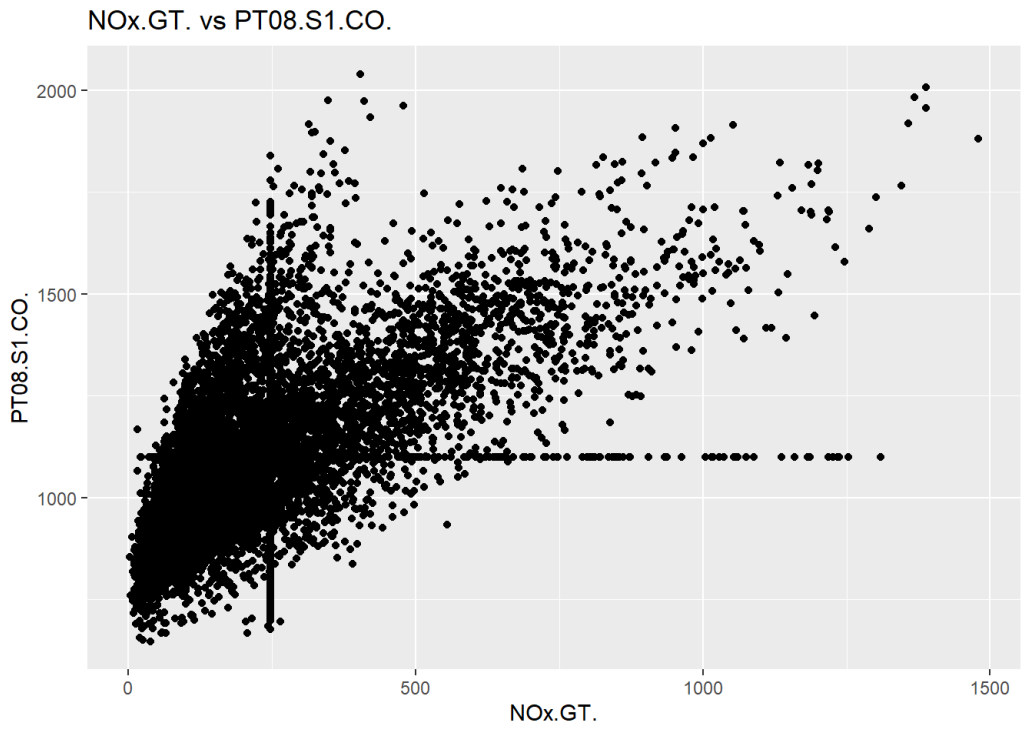
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -200.0    50.0   141.0   168.6   284.0  1479.0
```

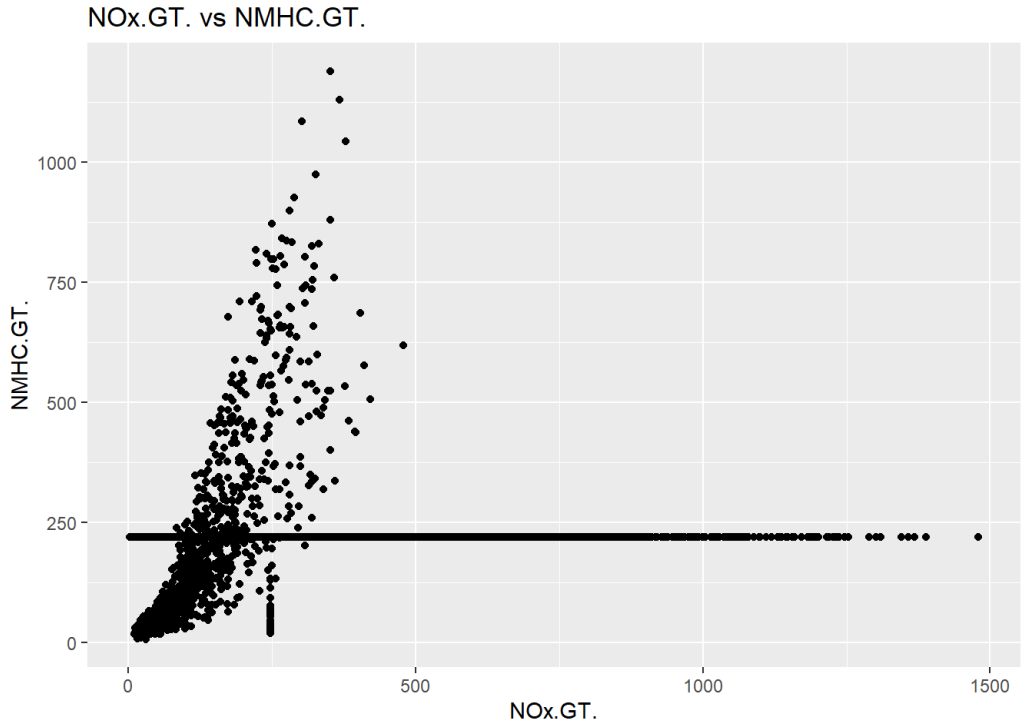
```
# Replace -200 values with "NA" so it does not affect mean calculation
data[data == -200] <- NA

# Calculate mean for each feature, exclude NA in mean calculation
feature_means <- data %>%
  summarise(across(where(is.numeric), ~mean(., na.rm = TRUE)))

# Replace NA with the mean of corresponding feature
cleaned_data <- data %>%
  mutate(across(where(is.numeric), ~ifelse(is.na(.), feature_means[[cur_column()]], .)))

cleaned_data_no_time_eda <- select(cleaned_data, -Date, -Time)
cor_matrix <- cor(cleaned_data_no_time_eda)
corrplot(cor_matrix, method = "circle", type = "upper", order = "hclust",
         tl.col = "black", tl.srt = 45)
```

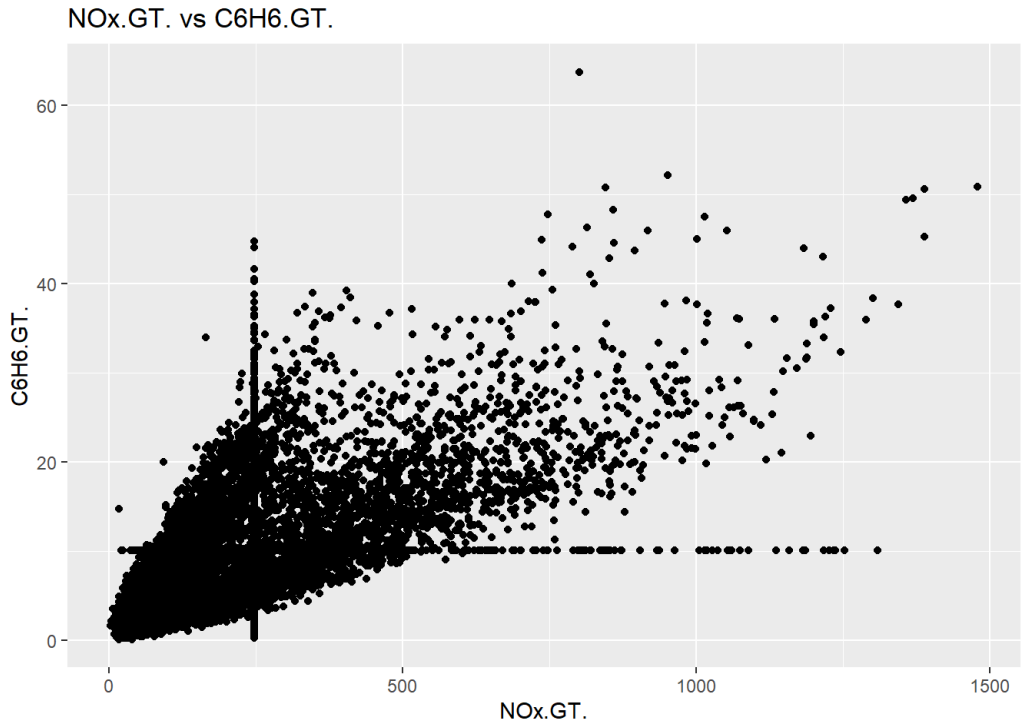


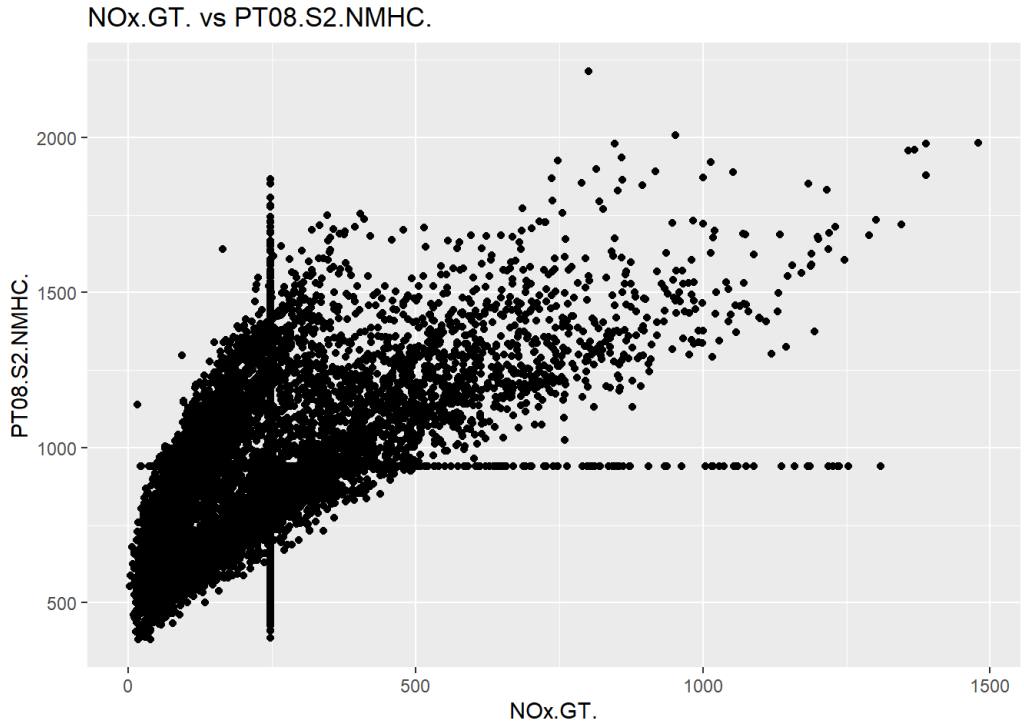

```
# Plot response variable against every other variable
variables <- names(cleaned_data_no_time_eda)
for (var in variables) {
  if (var != "NOx.GT.") {
    p <- ggplot(cleaned_data_no_time_eda, aes_string(x="NOx.GT.", y=var)) +
        geom_point() +
        ggtitle(paste("NOx.GT. vs", var)) +
        xlab("NOx.GT.") +
        ylab(var)
      print(p)
  }
}
```

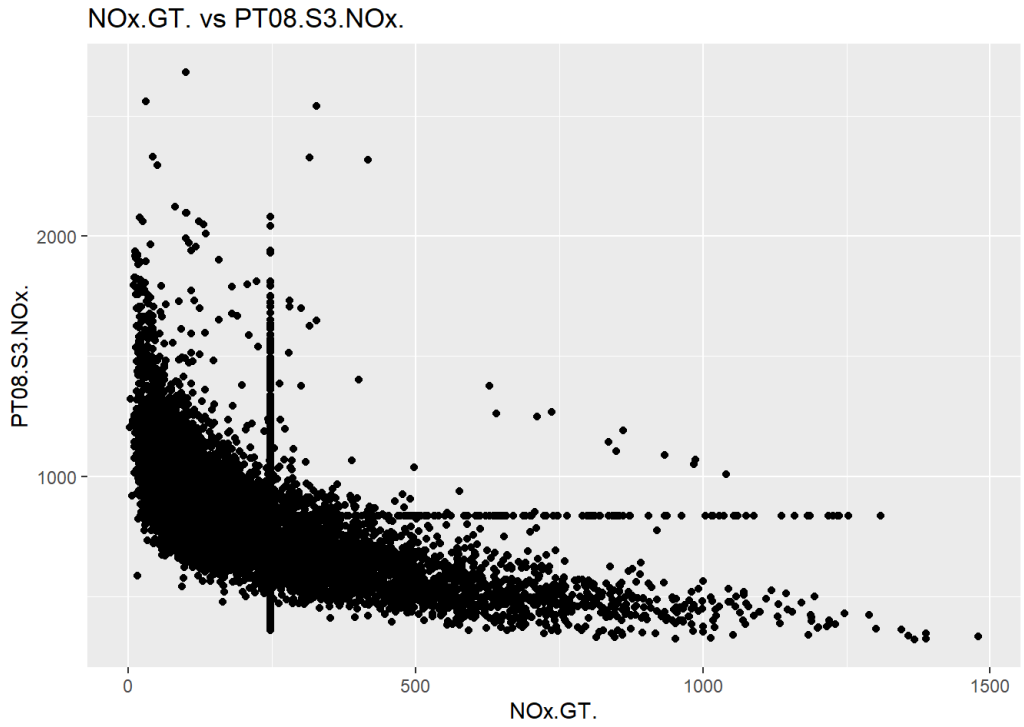

NOx.GT. vs CO.GT.
CO.GT.
10.0
7.5
5.0
2.5
0.0
0
500
1000
1500
NOx.GT.

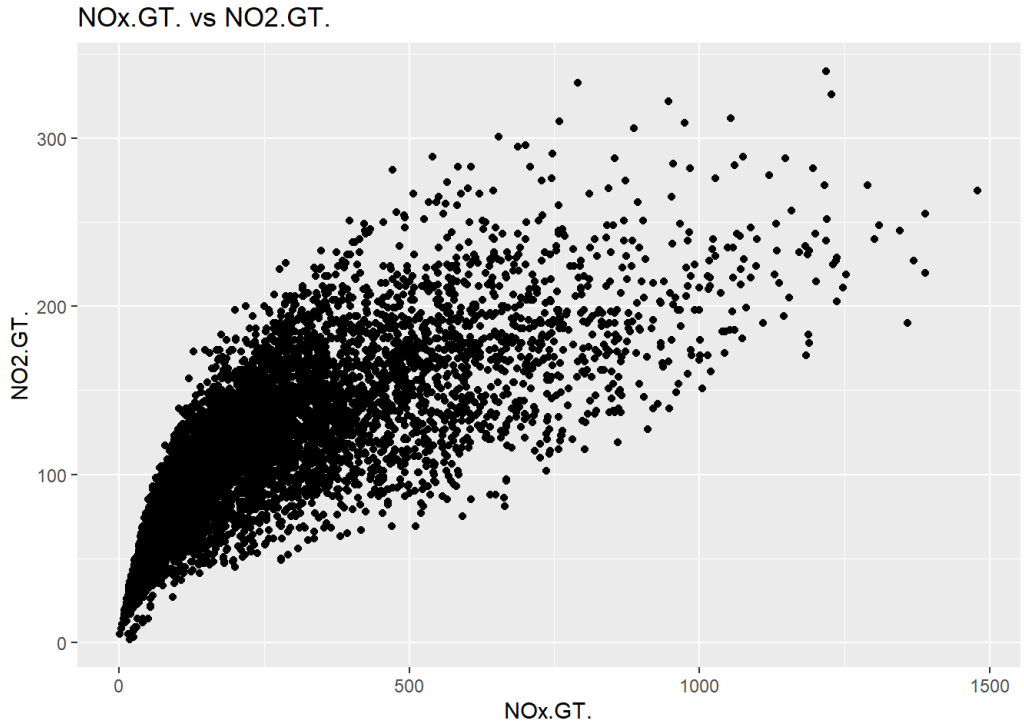

NOx.GT. vs PT08.S1.CO.
PT08.S1.CO.
2000
1500
1000
0
500
1000
1500
NOx.GT.

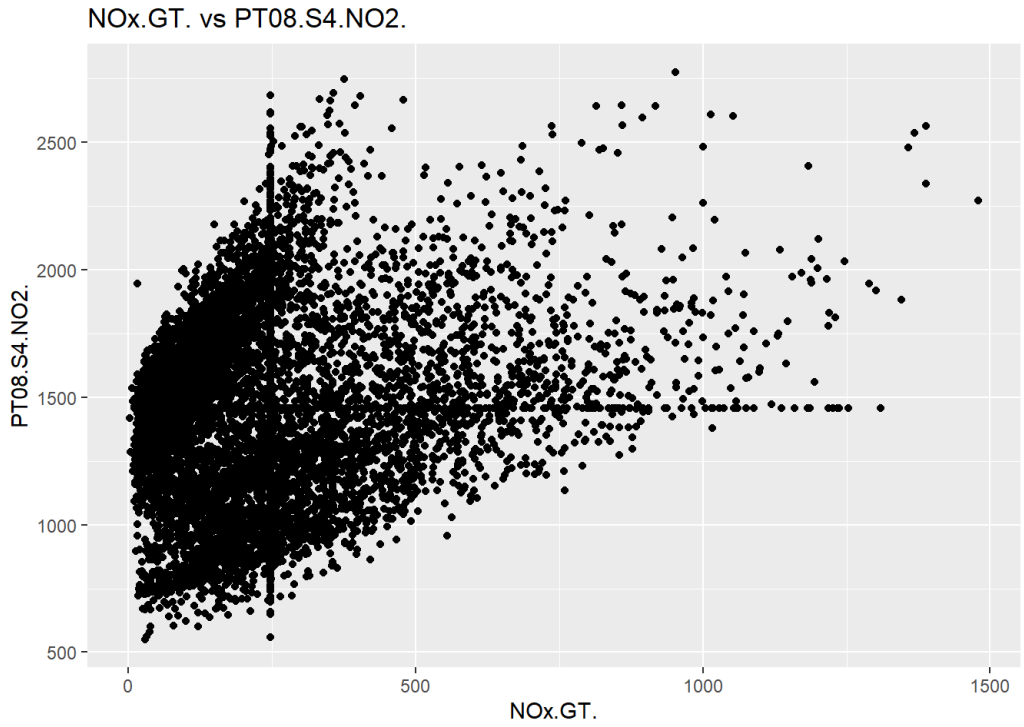

NOx.GT. vs NMHC.GT.
NMHC.GT.
1000
750
500
250
0
0
500
1000
1500
NOx.GT.

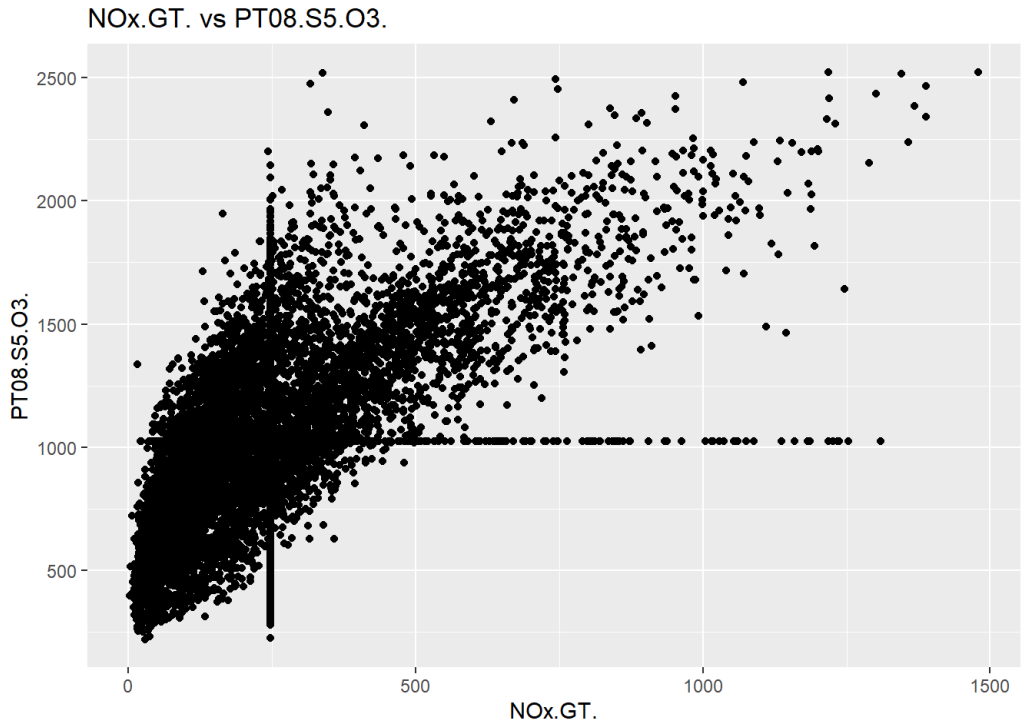

NOx.GT. vs C6H6.GT.
C6H6.GT.
60
40
20
0
0
500
1000
1500
NOx.GT.

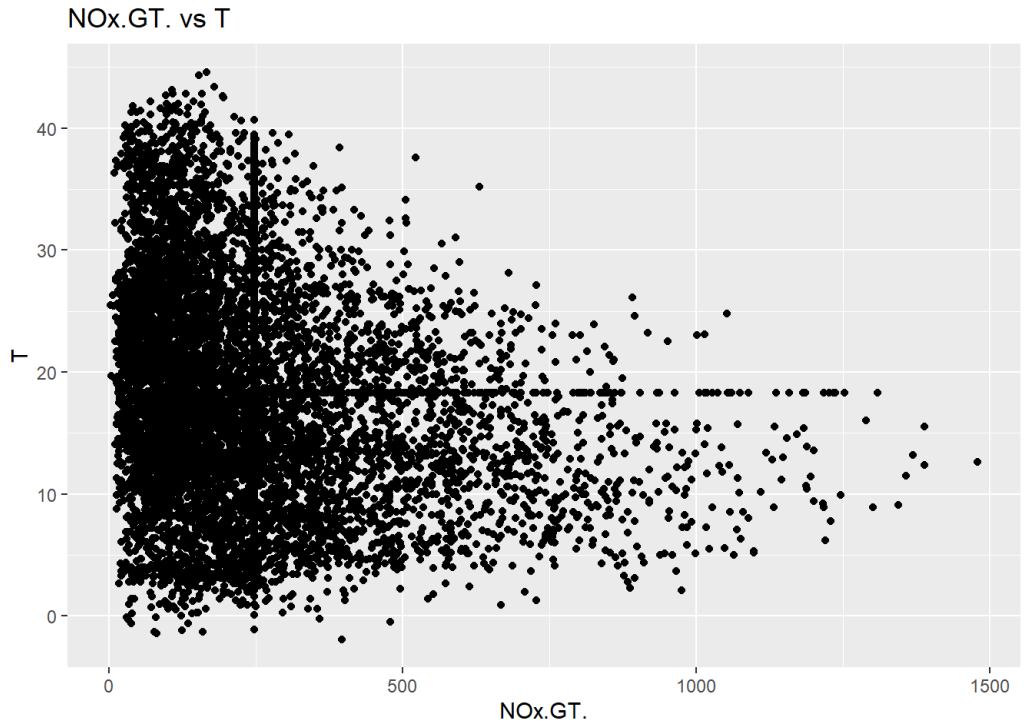

NOx.GT. vs PT08.S2.NMHC.
PT08.S2.NMHC.
2000
1500
1000
500
0
500
1000
1500
NOx.GT.

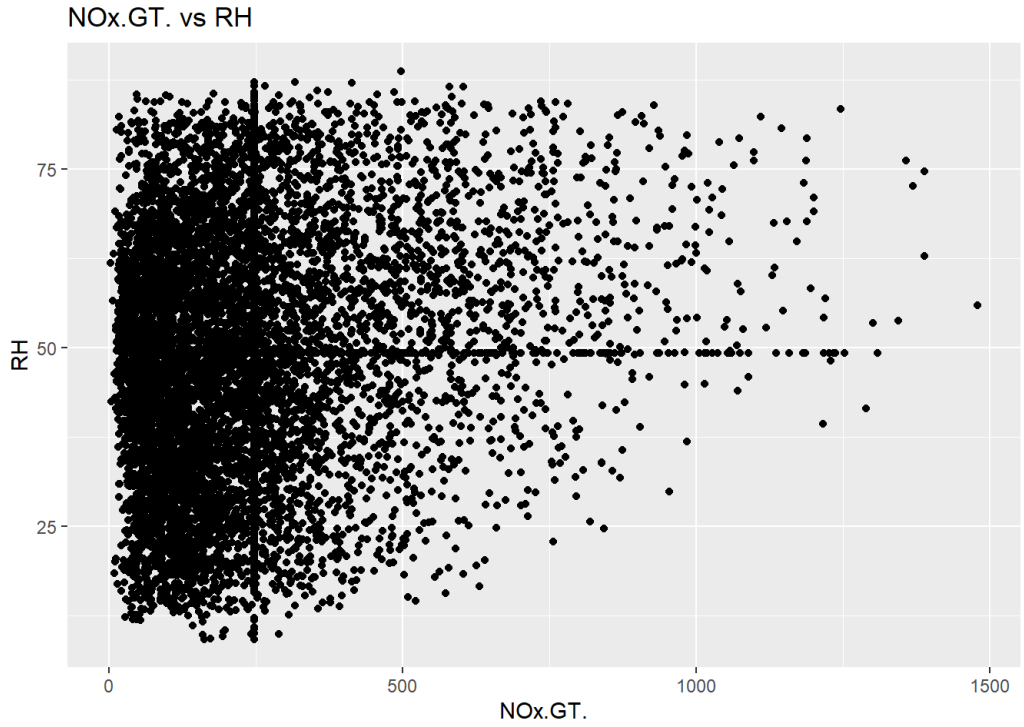

NOx.GT. vs PT08.S3.NOx.
PT08.S3.NOx.
2000
1000
0
500
1000
1500
NOx.GT.

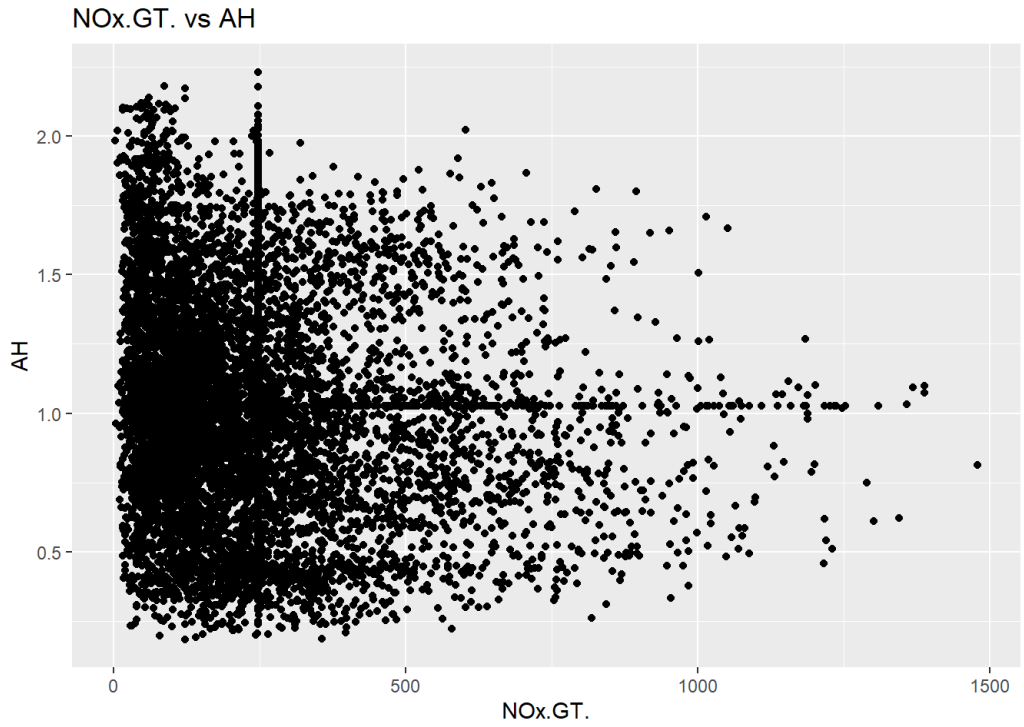

NOx.GT. vs NO2.GT.
NO2.GT.
300
200
100
0
0
500
1000
1500
NOx.GT.


NOx.GT. vs PT08.S4.NO2.
PT08.S4.NO2.
2500
2000
1500
1000
500
0
500
1000
1500
NOx.GT.

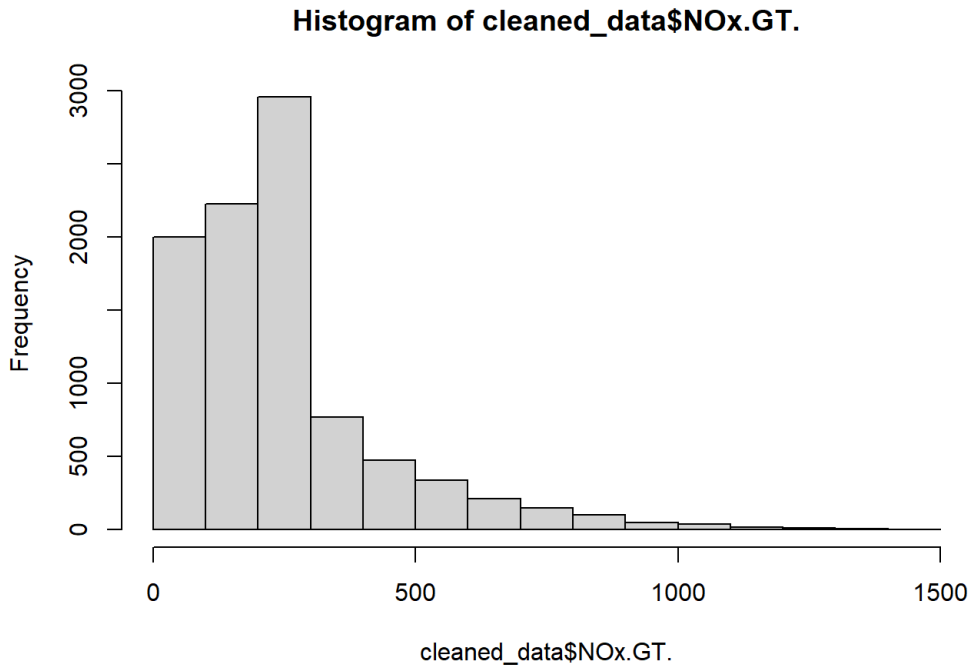

NOx.GT. vs PT08.S5.O3.
PT08.S5.O3.
2500
2000
1500
1000
500
0
500
1000
1500
NOx.GT.


NOx.GT. vs T
T
40
30
20
10
0
0
500
1000
1500
NOx.GT.

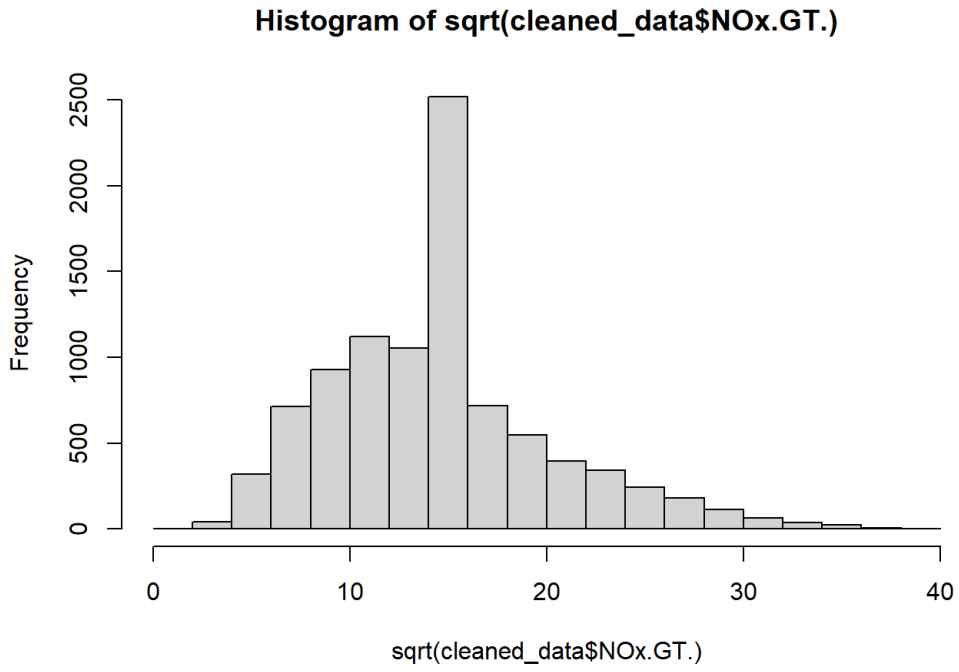





```
hist(cleaned_data$NOx.GT.)
```




```
hist(sqrt(cleaned_data$NOx.GT.))
```




```
cleaned_data$NOx.GT. <- sqrt(cleaned_data$NOx.GT.)
cleaned_data_no_time <- select(cleaned_data, -Date, -Time)

null_model <- lm(NOx.GT. ~ PT08.S3.NOx., data = cleaned_data_no_time)
full_model <- lm(NOx.GT.~., data = cleaned_data_no_time)
step_model1 <- step(null_model, scope = list(lower = null_model, upper = full_model), direction = "both", test = "F")
```

```
## Start:  AIC=27876.94
## NOx.GT. ~ PT08.S3.NOx.
##
##                 Df Sum of Sq    RSS   AIC   F value    Pr(>F)
## + NO2.GT.        1     93564  90440 21233 9677.0838 < 2.2e-16 ***
## + CO.GT.         1     63614 120390 23910 4942.6277 < 2.2e-16 ***
## + PT08.S5.O3.    1     37422 146582 25751 2388.0703 < 2.2e-16 ***
## + T              1     33085 150919 26024 2050.5861 < 2.2e-16 ***
## + AH             1     25001 159003 26513 1470.7553 < 2.2e-16 ***
## + PT08.S1.CO.    1     17909 166095 26921 1008.5805 < 2.2e-16 ***
## + C6H6.GT.       1     17558 166446 26941  986.7300 < 2.2e-16 ***
## + PT08.S2.NMHC.  1     13858 170146 27146  761.8616 < 2.2e-16 ***
## + PT08.S4.NO2.   1      7123 176881 27510  376.6928 < 2.2e-16 ***
## + RH             1      6876 177128 27523  363.0905 < 2.2e-16 ***
## + NMHC.GT.       1       164 183840 27871    8.3484  0.003869 **
## <none>                       184004 27877
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=21232.95
## NOx.GT. ~ PT08.S3.NOx. + NO2.GT.
##
##                 Df Sum of Sq    RSS   AIC   F value    Pr(>F)
## + CO.GT.         1     15955  74486 19419 2003.3796 < 2.2e-16 ***
## + RH             1     15924  74516 19423 1998.7399 < 2.2e-16 ***
## + T              1      8652  81789 20294  989.3552 < 2.2e-16 ***
## + PT08.S5.O3.    1      8303  82137 20334  945.5274 < 2.2e-16 ***
## + C6H6.GT.       1      4875  85566 20717  532.8321 < 2.2e-16 ***
## + PT08.S1.CO.    1      4029  86411 20809  436.0844 < 2.2e-16 ***
## + PT08.S2.NMHC.  1      2788  87653 20942  297.4480 < 2.2e-16 ***
## + NMHC.GT.       1       244  90196 21210   25.3257 4.932e-07 ***
## + PT08.S4.NO2.   1       131  90309 21221   13.5288 0.0002362 ***
## + AH             1        21  90419 21233    2.1923 0.1387351
## <none>                       90440 21233
## - NO2.GT.        1     93564 184004 27877 9677.0838 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=19418.92
## NOx.GT. ~ PT08.S3.NOx. + NO2.GT. + CO.GT.
##
##                 Df Sum of Sq    RSS   AIC  F value    Pr(>F)
## + RH             1     13468  61017 17555 2064.260 < 2.2e-16 ***
## + T              1     10174  64312 18047 1479.425 < 2.2e-16 ***
## + PT08.S4.NO2.   1      8280  66205 18318 1169.644 < 2.2e-16 ***
## + NMHC.GT.       1      1584  72901 19220  203.217 < 2.2e-16 ***
## + PT08.S5.O3.    1      1160  73326 19274  147.927 < 2.2e-16 ***
## + PT08.S2.NMHC.  1      1135  73351 19277  144.690 < 2.2e-16 ***
## + C6H6.GT.       1       364  74121 19375   45.937 1.295e-11 ***
## + AH             1       210  74275 19395   26.475 2.723e-07 ***
## + PT08.S1.CO.    1        94  74391 19409   11.829 0.0005856 ***
## <none>                       74486 19419
## - CO.GT.         1     15955  90440 21233 2003.380 < 2.2e-16 ***
## - NO2.GT.        1     45905 120390 23910 5764.146 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=17554.68
## NOx.GT. ~ PT08.S3.NOx. + NO2.GT. + CO.GT. + RH
##
##                 Df Sum of Sq    RSS   AIC   F value    Pr(>F)
## + PT08.S4.NO2.   1      5033  55985 16751  840.5947 < 2.2e-16 ***
## + NMHC.GT.       1       890  60128 17419  138.3366 < 2.2e-16 ***
## + T              1       875  60143 17422  135.9814 < 2.2e-16 ***
## + AH             1       638  60379 17458   98.8834 < 2.2e-16 ***
## + PT08.S1.CO.    1       593  60424 17465   91.7615 < 2.2e-16 ***
## + PT08.S5.O3.    1       187  60830 17528   28.7379 8.484e-08 ***
## + C6H6.GT.       1        98  60920 17542   14.9900 0.0001088 ***
```

```
## + PT08.S2.NMHC.  1        20  60997 17554    3.0564 0.0804515 .
## <none>                       61017 17555
## - RH             1     13468  74486 19419 2064.2599 < 2.2e-16 ***
## - CO.GT.         1     13499  74516 19423 2068.9260 < 2.2e-16 ***
## - NO2.GT.        1     52934 113951 23397 8113.0222 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=16751.23
## NOx.GT. ~ PT08.S3.NOx. + NO2.GT. + CO.GT. + RH + PT08.S4.NO2.
##
##                 Df Sum of Sq   RSS   AIC  F value    Pr(>F)
## + PT08.S2.NMHC.  1    4709.0 51276 15931  858.674 < 2.2e-16 ***
## + C6H6.GT.       1    4127.7 51857 16037  744.235 < 2.2e-16 ***
## + PT08.S5.O3.    1    1323.9 54661 16529  226.459 < 2.2e-16 ***
## + NMHC.GT.       1     848.5 55136 16610  143.885 < 2.2e-16 ***
## + AH             1     698.1 55287 16636  118.056 < 2.2e-16 ***
## + T              1     588.5 55396 16654   99.323 < 2.2e-16 ***
## + PT08.S1.CO.    1      72.2 55912 16741   12.075 0.0005135 ***
## <none>                      55985 16751
## - PT08.S4.NO2.   1    5032.7 61017 17555  840.595 < 2.2e-16 ***
## - RH             1   10220.7 66205 18318 1707.148 < 2.2e-16 ***
## - CO.GT.         1   18524.1 74509 19424 3094.046 < 2.2e-16 ***
## - NO2.GT.        1   27119.6 83104 20445 4529.738 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=15931.11
## NOx.GT. ~ PT08.S3.NOx. + NO2.GT. + CO.GT. + RH + PT08.S4.NO2. +
##     PT08.S2.NMHC.
##
##                 Df Sum of Sq   RSS   AIC   F value  Pr(>F)
## + AH             1    6285.6 44990 14709 1306.1616 < 2e-16 ***
## + T              1    4783.6 46492 15017  961.9205 < 2e-16 ***
## + PT08.S1.CO.    1     877.4 50398 15772  162.7633 < 2e-16 ***
## + NMHC.GT.       1     707.2 50568 15803  130.7440 < 2e-16 ***
## + PT08.S5.O3.    1      19.2 51256 15930    3.4939 0.06163 .
## + C6H6.GT.       1      19.1 51257 15930    3.4793 0.06217 .
## <none>                      51276 15931
## - PT08.S2.NMHC.  1    4709.0 55985 16751  858.6739 < 2e-16 ***
## - CO.GT.         1    8127.1 59403 17306 1481.9547 < 2e-16 ***
## - PT08.S4.NO2.   1    9721.7 60997 17554 1772.7324 < 2e-16 ***
## - RH             1   13517.2 64793 18119 2464.8379 < 2e-16 ***
## - NO2.GT.        1   20910.1 72186 19130 3812.9103 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=14709.44
## NOx.GT. ~ PT08.S3.NOx. + NO2.GT. + CO.GT. + RH + PT08.S4.NO2. +
##     PT08.S2.NMHC. + AH
##
##                 Df Sum of Sq   RSS   AIC   F value    Pr(>F)
## + NMHC.GT.       1     329.0 44661 14643   68.8644 < 2.2e-16 ***
## + T              1     128.8 44861 14685   26.8483 2.247e-07 ***
## + PT08.S1.CO.    1      75.8 44914 14696   15.7755 7.185e-05 ***
## + PT08.S5.O3.    1      31.4 44959 14705    6.5210   0.01068 *
## <none>                      44990 14709
## + C6H6.GT.       1       7.6 44982 14710    1.5722   0.20993
## - AH             1    6285.6 51276 15931 1306.1616 < 2.2e-16 ***
## - CO.GT.         1    8177.1 53167 16270 1699.2068 < 2.2e-16 ***
## - PT08.S2.NMHC.  1   10296.5 55287 16636 2139.6376 < 2.2e-16 ***
## - RH             1   11159.3 56149 16781 2318.9111 < 2.2e-16 ***
## - PT08.S4.NO2.   1   15339.7 60330 17453 3187.6196 < 2.2e-16 ***
## - NO2.GT.        1   24286.2 69276 18747 5046.7183 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=14642.77
```

```
## NOx.GT. ~ PT08.S3.NOx. + NO2.GT. + CO.GT. + RH + PT08.S4.NO2. +
##     PT08.S2.NMHC. + AH + NMHC.GT.
##
##                 Df Sum of Sq   RSS   AIC   F value    Pr(>F)
## + T              1     113.0 44548 14621   23.7086 1.139e-06 ***
## + PT08.S1.CO.    1      65.3 44596 14631   13.6784 0.0002182 ***
## + PT08.S5.O3.    1      26.6 44634 14639    5.5599 0.0183968 *
## + C6H6.GT.       1      14.6 44646 14642    3.0622 0.0801646 .
## <none>                       44661 14643
## - NMHC.GT.       1     329.0 44990 14709   68.8644 < 2.2e-16 ***
## - AH             1    5907.4 50568 15803 1236.4855 < 2.2e-16 ***
## - CO.GT.         1    8493.2 53154 16270 1777.7220 < 2.2e-16 ***
## - PT08.S2.NMHC.  1    9874.8 54536 16510 2066.8873 < 2.2e-16 ***
## - RH             1   10798.8 55460 16667 2260.2921 < 2.2e-16 ***
## - PT08.S4.NO2.   1   14714.1 59375 17305 3079.8095 < 2.2e-16 ***
## - NO2.GT.        1   23668.0 68329 18620 4953.9478 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=14621.06
## NOx.GT. ~ PT08.S3.NOx. + NO2.GT. + CO.GT. + RH + PT08.S4.NO2. +
##     PT08.S2.NMHC. + AH + NMHC.GT. + T
##
##                 Df Sum of Sq   RSS   AIC  F value    Pr(>F)
## + PT08.S1.CO.    1      53.7 44494 14612   11.280 0.0007866 ***
## + PT08.S5.O3.    1      50.1 44498 14612   10.513 0.0011898 **
## + C6H6.GT.       1      35.1 44513 14616    7.375 0.0066257 **
## <none>                       44548 14621
## - T              1     113.0 44661 14643   23.709 1.139e-06 ***
## - NMHC.GT.       1     313.2 44861 14685   65.706 5.892e-16 ***
## - AH             1    1577.8 46126 14945  331.043 < 2.2e-16 ***
## - RH             1    2568.0 47116 15144  538.807 < 2.2e-16 ***
## - CO.GT.         1    8579.8 53128 16267 1800.195 < 2.2e-16 ***
## - PT08.S2.NMHC.  1    9980.3 54528 16511 2094.057 < 2.2e-16 ***
## - PT08.S4.NO2.   1   14510.9 59059 17257 3044.667 < 2.2e-16 ***
## - NO2.GT.        1   23578.3 68126 18594 4947.166 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=14611.78
## NOx.GT. ~ PT08.S3.NOx. + NO2.GT. + CO.GT. + RH + PT08.S4.NO2. +
##     PT08.S2.NMHC. + AH + NMHC.GT. + T + PT08.S1.CO.
##
##                 Df Sum of Sq   RSS   AIC   F value    Pr(>F)
## + PT08.S5.O3.    1     109.5 44385 14591   23.0608 1.594e-06 ***
## + C6H6.GT.       1      41.8 44453 14605    8.7876 0.0030405 **
## <none>                       44494 14612
## - PT08.S1.CO.    1      53.7 44548 14621   11.2800 0.0007866 ***
## - T              1     101.4 44596 14631   21.3065 3.966e-06 ***
## - NMHC.GT.       1     304.6 44799 14674   63.9789 1.407e-15 ***
## - AH             1    1492.5 45987 14918  313.5042 < 2.2e-16 ***
## - RH             1    2619.7 47114 15145  550.2605 < 2.2e-16 ***
## - CO.GT.         1    8624.6 53119 16268 1811.5895 < 2.2e-16 ***
## - PT08.S2.NMHC.  1    9442.3 53937 16411 1983.3559 < 2.2e-16 ***
## - PT08.S4.NO2.   1   13187.8 57682 17039 2770.0811 < 2.2e-16 ***
## - NO2.GT.        1   23614.5 68109 18594 4960.2205 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=14590.71
## NOx.GT. ~ PT08.S3.NOx. + NO2.GT. + CO.GT. + RH + PT08.S4.NO2. +
##     PT08.S2.NMHC. + AH + NMHC.GT. + T + PT08.S1.CO. + PT08.S5.O3.
##
##                 Df Sum of Sq   RSS   AIC   F value    Pr(>F)
## + C6H6.GT.       1      42.4 44342 14584    8.9368  0.002802 **
## <none>                       44385 14591
## - PT08.S5.O3.    1     109.5 44494 14612   23.0608 1.594e-06 ***
## - PT08.S1.CO.    1     113.2 44498 14612   23.8292 1.070e-06 ***
```

```
## - T              1     134.0 44519 14617   28.2223 1.106e-07 ***
## - NMHC.GT.       1     287.0 44672 14649   60.4335 8.415e-15 ***
## - AH             1    1396.4 45781 14879  294.0084 < 2.2e-16 ***
## - RH             1    2622.6 47007 15126  552.1741 < 2.2e-16 ***
## - PT08.S2.NMHC.  1    7464.1 51849 16043 1571.5396 < 2.2e-16 ***
## - CO.GT.         1    8671.2 53056 16258 1825.6868 < 2.2e-16 ***
## - PT08.S4.NO2.   1   12932.1 57317 16981 2722.7889 < 2.2e-16 ***
## - NO2.GT.        1   22528.3 66913 18430 4743.2247 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=14583.77
## NOx.GT. ~ PT08.S3.NOx. + NO2.GT. + CO.GT. + RH + PT08.S4.NO2. +
##     PT08.S2.NMHC. + AH + NMHC.GT. + T + PT08.S1.CO. + PT08.S5.O3. +
##     C6H6.GT.
##
##                 Df Sum of Sq   RSS   AIC   F value    Pr(>F)
## <none>                       44342 14584
## - C6H6.GT.       1      42.4 44385 14591    8.9368  0.002802 **
## - PT08.S5.O3.    1     110.1 44453 14605   23.2087 1.476e-06 ***
## - PT08.S1.CO.    1     122.3 44465 14608   25.7616 3.937e-07 ***
## - T              1     157.9 44500 14615   33.2799 8.234e-09 ***
## - NMHC.GT.       1     297.4 44640 14644   62.6684 2.725e-15 ***
## - AH             1    1275.0 45617 14847  268.6821 < 2.2e-16 ***
## - PT08.S2.NMHC.  1    1620.6 45963 14918  341.5057 < 2.2e-16 ***
## - RH             1    2663.2 47006 15128  561.2101 < 2.2e-16 ***
## - CO.GT.         1    7945.9 52288 16124 1674.3902 < 2.2e-16 ***
## - PT08.S4.NO2.   1   12754.2 57097 16947 2687.6065 < 2.2e-16 ***
## - NO2.GT.        1   22389.5 66732 18406 4718.0100 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(step_model1)
```

```
##
## Call:
## lm(formula = NOx.GT. ~ PT08.S3.NOx. + NO2.GT. + CO.GT. + RH +
##     PT08.S4.NO2. + PT08.S2.NMHC. + AH + NMHC.GT. + T + PT08.S1.CO. +
##     PT08.S5.O3. + C6H6.GT., data = cleaned_data_no_time)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.7864  -1.4108  -0.1461   1.2807  10.6809
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.3194251  0.5981440   0.534   0.5933
## PT08.S3.NOx.   0.0018002  0.0002119   8.496  < 2e-16 ***
## NO2.GT.        0.0602139  0.0008766  68.688  < 2e-16 ***
## CO.GT.         1.3981167  0.0341676  40.919  < 2e-16 ***
## RH             0.0878062  0.0037065  23.690  < 2e-16 ***
## PT08.S4.NO2.  -0.0113221  0.0002184 -51.842  < 2e-16 ***
## PT08.S2.NMHC.  0.0127691  0.0006910  18.480  < 2e-16 ***
## AH             3.0042782  0.1832824  16.392  < 2e-16 ***
## NMHC.GT.      -0.0029856  0.0003771  -7.916 2.72e-15 ***
## T              0.0569781  0.0098768   5.769 8.23e-09 ***
## PT08.S1.CO.   -0.0015387  0.0003031  -5.076 3.94e-07 ***
## PT08.S5.O3.    0.0008082  0.0001678   4.818 1.48e-06 ***
## C6H6.GT.       0.0588212  0.0196763   2.989   0.0028 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.178 on 9344 degrees of freedom
## Multiple R-squared:  0.8506, Adjusted R-squared:  0.8504
## F-statistic:  4432 on 12 and 9344 DF,  p-value: < 2.2e-16
```

```
step_model2 <- step(full_model, scope = list(lower = null_model, upper = full_model), direction = "both",test="F")
```

```
## Start:  AIC=14583.77
## NOx.GT. ~ CO.GT. + PT08.S1.CO. + NMHC.GT. + C6H6.GT. + PT08.S2.NMHC. +
##     PT08.S3.NOx. + NO2.GT. + PT08.S4.NO2. + PT08.S5.O3. + T +
##     RH + AH
##
##                 Df Sum of Sq   RSS   AIC   F value    Pr(>F)
## <none>                       44342 14584
## - C6H6.GT.       1      42.4 44385 14591    8.9368  0.002802 **
## - PT08.S5.O3.    1     110.1 44453 14605   23.2087 1.476e-06 ***
## - PT08.S1.CO.    1     122.3 44465 14608   25.7616 3.937e-07 ***
## - T              1     157.9 44500 14615   33.2799 8.234e-09 ***
## - NMHC.GT.       1     297.4 44640 14644   62.6684 2.725e-15 ***
## - AH             1    1275.0 45617 14847  268.6821 < 2.2e-16 ***
## - PT08.S2.NMHC.  1    1620.6 45963 14918  341.5057 < 2.2e-16 ***
## - RH             1    2663.2 47006 15128  561.2101 < 2.2e-16 ***
## - CO.GT.         1    7945.9 52288 16124 1674.3902 < 2.2e-16 ***
## - PT08.S4.NO2.   1   12754.2 57097 16947 2687.6065 < 2.2e-16 ***
## - NO2.GT.        1   22389.5 66732 18406 4718.0100 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(step_model2)
```

```
##
## Call:
## lm(formula = NOx.GT. ~ CO.GT. + PT08.S1.CO. + NMHC.GT. + C6H6.GT. +
##     PT08.S2.NMHC. + PT08.S3.NOx. + NO2.GT. + PT08.S4.NO2. + PT08.S5.O3. +
##     T + RH + AH, data = cleaned_data_no_time)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.7864  -1.4108  -0.1461   1.2807  10.6809
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.3194251  0.5981440   0.534   0.5933
## CO.GT.         1.3981167  0.0341676  40.919  < 2e-16 ***
## PT08.S1.CO.   -0.0015387  0.0003031  -5.076 3.94e-07 ***
## NMHC.GT.      -0.0029856  0.0003771  -7.916 2.72e-15 ***
## C6H6.GT.       0.0588212  0.0196763   2.989   0.0028 **
## PT08.S2.NMHC.  0.0127691  0.0006910  18.480  < 2e-16 ***
## PT08.S3.NOx.   0.0018002  0.0002119   8.496  < 2e-16 ***
## NO2.GT.        0.0602139  0.0008766  68.688  < 2e-16 ***
## PT08.S4.NO2.  -0.0113221  0.0002184 -51.842  < 2e-16 ***
## PT08.S5.O3.    0.0008082  0.0001678   4.818 1.48e-06 ***
## T              0.0569781  0.0098768   5.769 8.23e-09 ***
## RH             0.0878062  0.0037065  23.690  < 2e-16 ***
## AH             3.0042782  0.1832824  16.392  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.178 on 9344 degrees of freedom
## Multiple R-squared:  0.8506, Adjusted R-squared:  0.8504
## F-statistic:  4432 on 12 and 9344 DF,  p-value: < 2.2e-16
```

```
cleaned_data_no_time_all <- regsubsets(NOx.GT. ~ ., data = cleaned_data_no_time, nbest = 1, nvmax = 15)
summary(cleaned_data_no_time_all)$which
```

```
##    (Intercept) CO.GT. PT08.S1.CO. NMHC.GT. C6H6.GT. PT08.S2.NMHC. PT08.S3.NOx.
## 1         TRUE  FALSE       FALSE    FALSE    FALSE         FALSE        FALSE
## 2         TRUE   TRUE       FALSE    FALSE    FALSE         FALSE        FALSE
## 3         TRUE   TRUE       FALSE    FALSE    FALSE         FALSE        FALSE
## 4         TRUE  FALSE       FALSE    FALSE    FALSE          TRUE        FALSE
## 5         TRUE   TRUE       FALSE    FALSE    FALSE          TRUE        FALSE
## 6         TRUE   TRUE       FALSE    FALSE    FALSE          TRUE        FALSE
## 7         TRUE   TRUE       FALSE    FALSE    FALSE          TRUE         TRUE
## 8         TRUE   TRUE       FALSE     TRUE    FALSE          TRUE         TRUE
## 9         TRUE   TRUE       FALSE     TRUE    FALSE          TRUE         TRUE
## 10        TRUE   TRUE        TRUE     TRUE    FALSE          TRUE         TRUE
## 11        TRUE   TRUE        TRUE     TRUE    FALSE          TRUE         TRUE
## 12        TRUE   TRUE        TRUE     TRUE     TRUE          TRUE         TRUE
##    NO2.GT. PT08.S4.NO2. PT08.S5.O3.     T    RH    AH
## 1     TRUE        FALSE       FALSE FALSE FALSE FALSE
## 2     TRUE        FALSE       FALSE FALSE FALSE FALSE
## 3     TRUE        FALSE       FALSE FALSE  TRUE FALSE
## 4     TRUE         TRUE       FALSE FALSE  TRUE FALSE
## 5     TRUE         TRUE       FALSE FALSE  TRUE FALSE
## 6     TRUE         TRUE       FALSE FALSE  TRUE  TRUE
## 7     TRUE         TRUE       FALSE FALSE  TRUE  TRUE
## 8     TRUE         TRUE       FALSE FALSE  TRUE  TRUE
## 9     TRUE         TRUE       FALSE  TRUE  TRUE  TRUE
## 10    TRUE         TRUE       FALSE  TRUE  TRUE  TRUE
## 11    TRUE         TRUE        TRUE  TRUE  TRUE  TRUE
## 12    TRUE         TRUE        TRUE  TRUE  TRUE  TRUE
```

```
summary(cleaned_data_no_time_all)$rsq
```

```
##  [1] 0.6602932 0.7424127 0.7904025 0.7996255 0.8266089 0.8459038 0.8483843
##  [8] 0.8494930 0.8498738 0.8500548 0.8504239 0.8505668
```

```
summary(cleaned_data_no_time_all)$adjr2
```

```
##  [1] 0.6602569 0.7423576 0.7903353 0.7995398 0.8265162 0.8458049 0.8482708
##  [8] 0.8493642 0.8497293 0.8498944 0.8502479 0.8503749
```

```
model = lm(formula = NOx.GT. ~ CO.GT. + PT08.S1.CO. + NMHC.GT. + C6H6.GT. +
    PT08.S2.NMHC. + PT08.S3.NOx. + NO2.GT. + PT08.S4.NO2. + PT08.S5.O3. +
    T + RH + AH, data = cleaned_data_no_time)
vif(model)
```

```
##        CO.GT.   PT08.S1.CO.      NMHC.GT.      C6H6.GT. PT08.S2.NMHC.
##      3.986495      8.204023      1.143953     40.705228     64.397289
##  PT08.S3.NOx.       NO2.GT.  PT08.S4.NO2.   PT08.S5.O3.             T
##      5.609734      2.922687     10.830157      8.466973     14.415724
##            RH            AH
##      7.804379     10.377074
```

```
# Remove C6H6.GT.
model1 = lm(formula = NOx.GT. ~ CO.GT. + PT08.S1.CO. + NMHC.GT. +
    PT08.S2.NMHC. + PT08.S3.NOx. + NO2.GT. + PT08.S4.NO2. + PT08.S5.O3. +
    T + RH + AH, data = cleaned_data_no_time)
vif(model1)
```

```
##        CO.GT.   PT08.S1.CO.      NMHC.GT. PT08.S2.NMHC.  PT08.S3.NOx.
##      3.776546      8.166078      1.141120     18.081630      4.771462
##       NO2.GT.  PT08.S4.NO2.   PT08.S5.O3.             T            RH
##      2.871233     10.771777      8.466803     13.941246      7.725311
##            AH
##     10.074487
```

```
# Remove Sensor for NMHC
model2 = lm(formula = NOx.GT. ~ CO.GT. + PT08.S1.CO. + NMHC.GT. + PT08.S3.NOx. + NO2.GT. + PT08.S4.NO2. + PT08.S5.O3. +
    T + RH + AH, data = cleaned_data_no_time)
vif(model2)
```

```
##       CO.GT.  PT08.S1.CO.     NMHC.GT. PT08.S3.NOx.      NO2.GT. PT08.S4.NO2.
##     3.424494     8.015084     1.134618     3.757300     2.871107     6.619325
##  PT08.S5.O3.            T           RH           AH
##     7.306677    13.939967     7.543902     9.388079
```

```
# Remove Temp
model3 = lm(formula = NOx.GT. ~ CO.GT. + PT08.S1.CO. + NMHC.GT. + PT08.S3.NOx. + NO2.GT. + PT08.S4.NO2. + PT08.S5.O3. + RH +
AH, data = cleaned_data_no_time)
vif(model3)
```

```
##       CO.GT.  PT08.S1.CO.     NMHC.GT. PT08.S3.NOx.      NO2.GT. PT08.S4.NO2.
##     3.414495     8.014486     1.132275     3.746751     2.863181     6.036886
##  PT08.S5.O3.           RH           AH
##     7.110782     1.379819     3.482655
```

```
# Remove CO sensor
model4 = lm(formula = NOx.GT. ~ CO.GT. + NMHC.GT. + PT08.S3.NOx. + NO2.GT. + PT08.S4.NO2. + PT08.S5.O3. + RH + AH, data = cl
eaned_data_no_time)
vif(model4)
```

```
##       CO.GT.     NMHC.GT. PT08.S3.NOx.      NO2.GT. PT08.S4.NO2.  PT08.S5.O3.
##     3.322005     1.130682     3.478582     2.862893     4.758476     5.617012
##           RH           AH
##     1.307883     3.134334
```

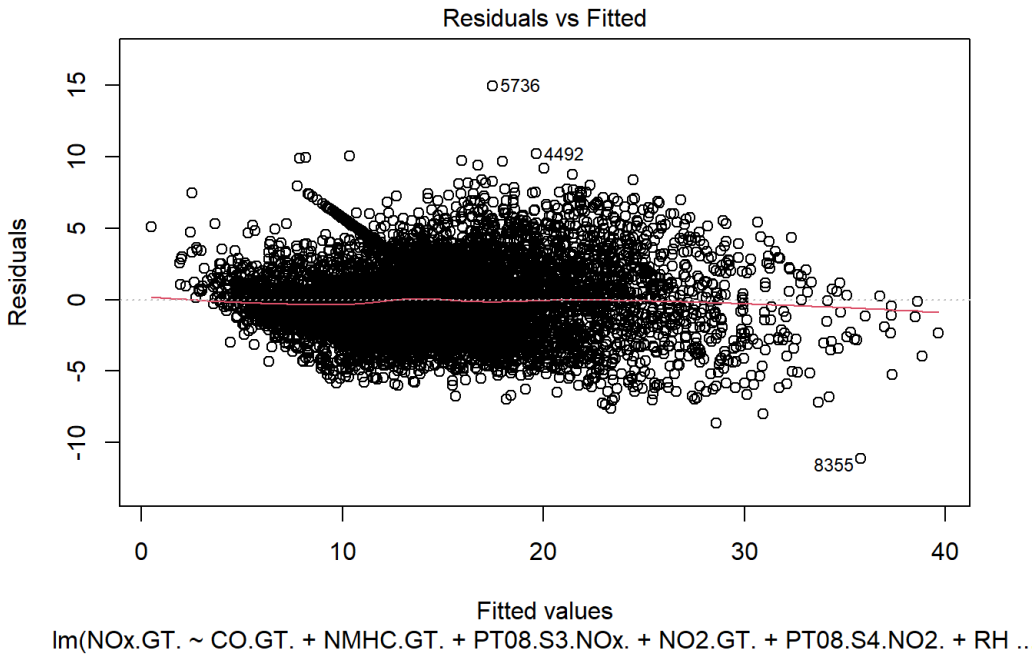
```
# Remove O3 sensor
model5 = lm(formula = NOx.GT. ~ CO.GT. + NMHC.GT. + PT08.S3.NOx. + NO2.GT. + PT08.S4.NO2. + RH + AH, data = cleaned_data_no_
time)
vif(model5)
```

```
##       CO.GT.     NMHC.GT. PT08.S3.NOx.      NO2.GT. PT08.S4.NO2.           RH
##     3.177519     1.125979     2.334402     2.763502     3.630076     1.148650
##           AH
##     2.589095
```

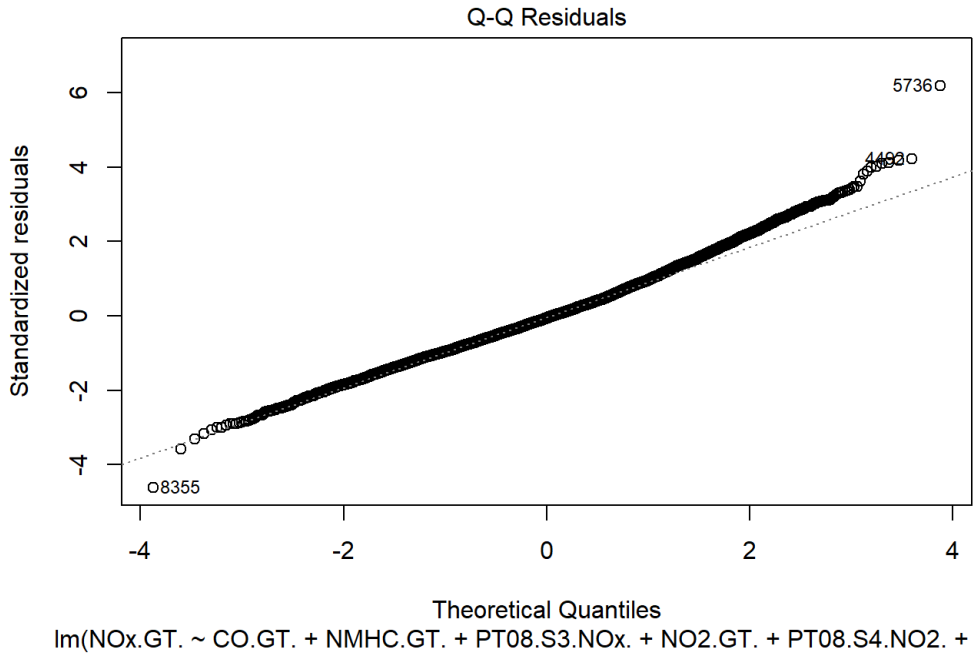
```
summary(model5)
```

```
##
## Call:
## lm(formula = NOx.GT. ~ CO.GT. + NMHC.GT. + PT08.S3.NOx. + NO2.GT. +
##     PT08.S4.NO2. + RH + AH, data = cleaned_data_no_time)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.1562  -1.6343  -0.1493   1.4492  14.9550
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.4505055  0.3288208   28.74   <2e-16 ***
## CO.GT.        1.9671696  0.0338204   58.16   <2e-16 ***
## NMHC.GT.     -0.0047063  0.0004148  -11.35   <2e-16 ***
## PT08.S3.NOx. -0.0038122  0.0001515  -25.16   <2e-16 ***
## NO2.GT.       0.0636361  0.0009451   67.33   <2e-16 ***
## PT08.S4.NO2. -0.0040552  0.0001402  -28.93   <2e-16 ***
## RH            0.0579147  0.0015765   36.74   <2e-16 ***
## AH            1.0297320  0.1015015   10.14   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.415 on 9349 degrees of freedom
## Multiple R-squared:  0.8162, Adjusted R-squared:  0.8161
## F-statistic:  5931 on 7 and 9349 DF,  p-value: < 2.2e-16
```

```
plot(model5, which = 1)
```




```
plot(model5, which = 2)
```




```
plot(model5, which = 4)
```