final-test-analysis.R

Shawn

2021-03-27

```
pollutants <- read.csv("C:/Users/Shawn/Downloads/pollutants.csv")
library("car")
```

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

```
library(ggplot2)
library(ggcorrplot)
library(caret)
```

```
## Loading required package: lattice
```

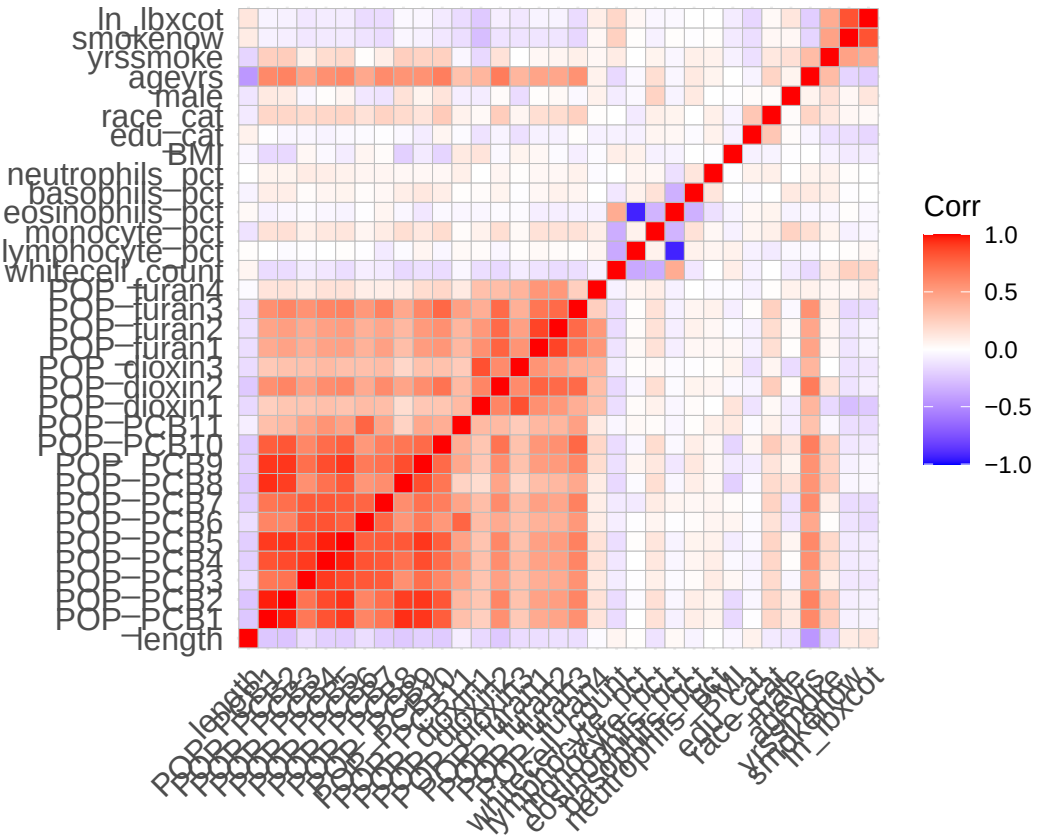
```
library(glmnet)
```

```
## Loading required package: Matrix
```

```
## Loaded glmnet 4.1-1
```

```
#remove the x
pollutants["X"] = NULL

#calculate correlation matrix before removing the multicolinearity covariates
corr_mat = cor(pollutants)
#graph colored corr matrix
ggcorrplot(corr_mat)
```




```
#set the factor
pollutants$edu_cat = as.factor(pollutants$edu_cat)
pollutants$race_cat = as.factor(pollutants$race_cat)
pollutants$male = as.factor(pollutants$male)
pollutants$smokenow = as.factor(pollutants$smokenow)
#fit model
model = lm(length ~ .,data = pollutants)
#summary
summary(model)
```

```
##
## Call:
## lm(formula = length ~ ., data = pollutants)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.5023 -0.1540 -0.0290  0.1224  1.1904
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -5.516e-02  9.700e+00  -0.006   0.9955
## POP_PCB1        -1.604e-06  1.075e-06  -1.492   0.1361
## POP_PCB2         7.240e-07  3.023e-06   0.240   0.8108
## POP_PCB3         1.189e-06  2.157e-06   0.551   0.5816
## POP_PCB4        -1.800e-07  1.026e-06  -0.175   0.8608
## POP_PCB5         1.496e-07  1.070e-06   0.140   0.8889
```

```
## POP_PCB6         2.754e-07  1.059e-06   0.260   0.7949
## POP_PCB7        -5.768e-07  1.207e-06  -0.478   0.6328
## POP_PCB8         1.644e-06  2.447e-06   0.672   0.5021
## POP_PCB9         6.043e-07  2.115e-06   0.286   0.7751
## POP_PCB10        1.181e-03  8.919e-04   1.324   0.1858
## POP_PCB11        3.405e-05  3.079e-04   0.111   0.9120
## POP_dioxin1      2.773e-05  3.056e-04   0.091   0.9277
## POP_dioxin2     -1.732e-04  4.398e-04  -0.394   0.6939
## POP_dioxin3     -1.876e-05  3.027e-05  -0.620   0.5356
## POP_furan1       2.522e-03  3.846e-03   0.656   0.5122
## POP_furan2      -2.915e-04  4.504e-03  -0.065   0.9484
## POP_furan3       4.498e-03  2.762e-03   1.629   0.1038
## POP_furan4      -6.489e-04  9.201e-04  -0.705   0.4808
## whitecell_count -5.233e-03  4.410e-03  -1.186   0.2358
## lymphocyte_pct   1.420e-02  9.698e-02   0.146   0.8836
## monocyte_pct     9.448e-03  9.697e-02   0.097   0.9224
## eosinophils_pct  1.545e-02  9.697e-02   0.159   0.8734
## basophils_pct    1.651e-02  9.706e-02   0.170   0.8650
## neutrophils_pct  2.754e-02  9.816e-02   0.281   0.7791
## BMI             -1.367e-03  1.411e-03  -0.969   0.3329
## edu_cat2         2.439e-02  2.218e-02   1.100   0.2718
## edu_cat3         4.781e-02  2.166e-02   2.207   0.0276 *
## edu_cat4         3.259e-02  2.557e-02   1.275   0.2028
## race_cat2       -2.198e-02  3.267e-02  -0.673   0.5013
## race_cat3        2.464e-02  3.372e-02   0.730   0.4653
## race_cat4       -3.479e-02  2.993e-02  -1.162   0.2455
## male1           -3.947e-02  1.772e-02  -2.227   0.0262 *
## ageyrs          -6.234e-03  7.447e-04  -8.372 2.41e-16 ***
## yrssmoke        -5.325e-04  7.276e-04  -0.732   0.4645
## smokenow1        1.914e-03  3.587e-02   0.053   0.9575
## ln_lbxcot        5.371e-03  3.928e-03   1.367   0.1719
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2221 on 827 degrees of freedom
## Multiple R-squared:  0.2448, Adjusted R-squared:  0.2119
## F-statistic: 7.447 on 36 and 827 DF,  p-value: < 2.2e-16
```

```
#show the VIF
vif(model)
```

```
##                       GVIF Df GVIF^(1/(2*Df))
## POP_PCB1         33.044120  1        5.748401
## POP_PCB2         34.281125  1        5.855009
## POP_PCB3          9.351143  1        3.057964
## POP_PCB4         31.742239  1        5.634025
## POP_PCB5         59.896895  1        7.739308
## POP_PCB6         11.386658  1        3.374412
## POP_PCB7          4.870075  1        2.206825
## POP_PCB8         12.982575  1        3.603134
## POP_PCB9         12.441595  1        3.527264
## POP_PCB10         6.020678  1        2.453707
## POP_PCB11         4.725769  1        2.173883
## POP_dioxin1       5.276251  1        2.297009
## POP_dioxin2       5.413132  1        2.326614
```

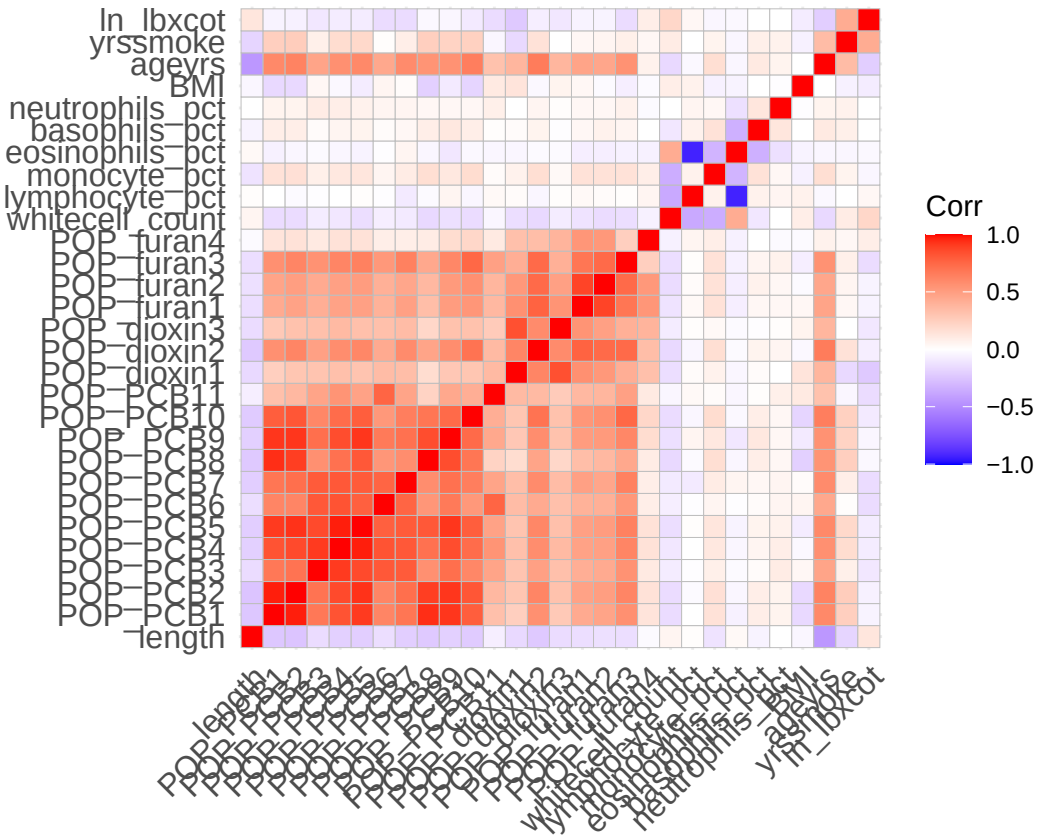
```
## POP_dioxin3         4.398509  1        2.097262
## POP_furan1          6.154213  1        2.480769
## POP_furan2          6.195336  1        2.489043
## POP_furan3          4.464346  1        2.112900
## POP_furan4          1.821809  1        1.349744
## whitecell_count     1.548380  1        1.244339
## lymphocyte_pct  12250.336528  1      110.681238
## monocyte_pct      726.843372  1       26.960033
## eosinophils_pct 15071.561945  1      122.766290
## basophils_pct     867.412798  1       29.451873
## neutrophils_pct    37.984114  1        6.163125
## BMI                 1.263662  1        1.124127
## edu_cat             1.543109  3        1.074978
## race_cat            2.052848  3        1.127352
## male                1.350324  1        1.162034
## ageyrs              3.238631  1        1.799620
## yrssmoke            2.204139  1        1.484634
## smokenow            4.006708  1        2.001676
## ln_lbxcot           3.963407  1        1.990831
```

```
#get set a dataset with no categorical covariates
no_cat = pollutants
no_cat$edu_cat = NULL
no_cat$race_cat = NULL
no_cat$male = NULL
no_cat$smokenow = NULL
#summary of the dataset
summary(no_cat)
```

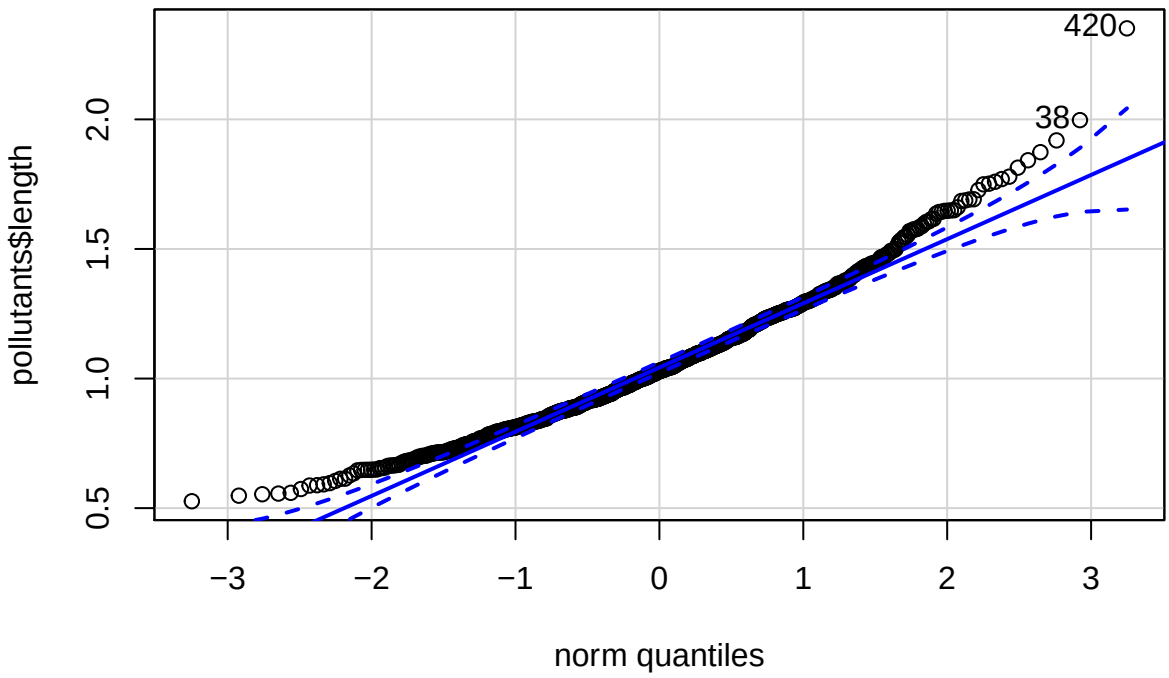
```
##      length          POP_PCB1         POP_PCB2         POP_PCB3
##  Min.   :0.5266   Min.   :  2000   Min.   :  2000   Min.   :  2000
##  1st Qu.:0.8754   1st Qu.:  9975   1st Qu.:  4800   1st Qu.:  3700
##  Median :1.0286   Median : 27600   Median : 11500   Median :  6200
##  Mean   :1.0543   Mean   : 38082   Mean   : 15637   Mean   : 10158
##  3rd Qu.:1.2095   3rd Qu.: 53325   3rd Qu.: 21825   3rd Qu.: 12000
##  Max.   :2.3512   Max.   :572000   Max.   :165000   Max.   :123000
##     POP_PCB4         POP_PCB5         POP_PCB6         POP_PCB7
##  Min.   :  2100   Min.   :  2100   Min.   :  2000   Min.   :  1100
##  1st Qu.: 11475   1st Qu.: 15600   1st Qu.:  4400   1st Qu.:  4000
##  Median : 25550   Median : 36300   Median :  9400   Median :  7450
##  Mean   : 38456   Mean   : 52650   Mean   : 16820   Mean   : 12682
##  3rd Qu.: 50650   3rd Qu.: 68625   3rd Qu.: 19500   3rd Qu.: 15625
##  Max.   :487000   Max.   :708000   Max.   :319000   Max.   :144000
##     POP_PCB8         POP_PCB9        POP_PCB10        POP_PCB11
##  Min.   :  1100   Min.   :  1100   Min.   :  1.70   Min.   :  1.30
##  1st Qu.:  3800   1st Qu.:  3900   1st Qu.:  9.10   1st Qu.: 14.80
##  Median :  6950   Median :  8050   Median : 18.35   Median : 24.50
##  Mean   : 10530   Mean   : 12220   Mean   : 24.49   Mean   : 38.15
##  3rd Qu.: 14425   3rd Qu.: 16025   3rd Qu.: 34.90   3rd Qu.: 42.95
##  Max.   :187000   Max.   :144000   Max.   :172.00   Max.   :845.00
##   POP_dioxin1      POP_dioxin2      POP_dioxin3       POP_furan1
##  Min.   :  1.90   Min.   :  1.40   Min.   :  36.8   Min.   : 1.000
##  1st Qu.: 23.90   1st Qu.: 21.27   1st Qu.: 197.0   1st Qu.: 3.200
##  Median : 41.35   Median : 37.80   Median : 342.5   Median : 5.200
##  Mean   : 57.65   Mean   : 47.81   Mean   : 494.4   Mean   : 6.371
```

```
##  3rd Qu.: 71.62   3rd Qu.: 62.42   3rd Qu.: 603.0   3rd Qu.: 7.700
##  Max.   :760.00   Max.   :281.00   Max.   :8190.0   Max.   :44.400
##    POP_furan2       POP_furan3       POP_furan4     whitecell_count
##  Min.   : 0.800   Min.   : 0.700   Min.   :  0.90   Min.   : 2.300
##  1st Qu.: 2.600   1st Qu.: 2.200   1st Qu.:  6.40   1st Qu.: 5.600
##  Median : 4.200   Median : 5.050   Median :  9.65   Median : 6.900
##  Mean   : 5.390   Mean   : 6.669   Mean   : 11.54   Mean   : 7.191
##  3rd Qu.: 6.825   3rd Qu.: 9.300   3rd Qu.: 14.00   3rd Qu.: 8.300
##  Max.   :33.500   Max.   :38.300   Max.   :234.00   Max.   :20.100
##  lymphocyte_pct   monocyte_pct    eosinophils_pct basophils_pct
##  Min.   : 5.80   Min.   : 1.600   Min.   :21.60   Min.   : 0.000
##  1st Qu.:24.00   1st Qu.: 6.600   1st Qu.:52.35   1st Qu.: 1.500
##  Median :28.95   Median : 7.700   Median :59.30   Median : 2.300
##  Mean   :29.92   Mean   : 7.936   Mean   :58.62   Mean   : 2.903
##  3rd Qu.:35.42   3rd Qu.: 9.100   3rd Qu.:65.22   3rd Qu.: 3.700
##  Max.   :73.40   Max.   :23.800   Max.   :88.10   Max.   :28.200
##  neutrophils_pct       BMI            ageyrs         yrssmoke
##  Min.   :0.0000   Min.   :16.16   Min.   :20.00   Min.   : 0.0
##  1st Qu.:0.4000   1st Qu.:23.88   1st Qu.:34.00   1st Qu.: 0.0
##  Median :0.6000   Median :27.38   Median :46.00   Median : 0.0
##  Mean   :0.6669   Mean   :28.09   Mean   :48.36   Mean   :10.6
##  3rd Qu.:0.8000   3rd Qu.:31.17   3rd Qu.:63.00   3rd Qu.:20.0
##  Max.   :5.5000   Max.   :62.99   Max.   :85.00   Max.   :69.0
##    ln_lbxcot
##  Min.   :-4.5099
##  1st Qu.:-4.0745
##  Median :-2.7334
##  Mean   :-0.9804
##  3rd Qu.: 2.8000
##  Max.   : 6.5848
```

```r
#calculate correlation matrix
corr_matrix = cor(no_cat)
#graph colored corr matrix
ggcorrplot(corr_matrix)
```

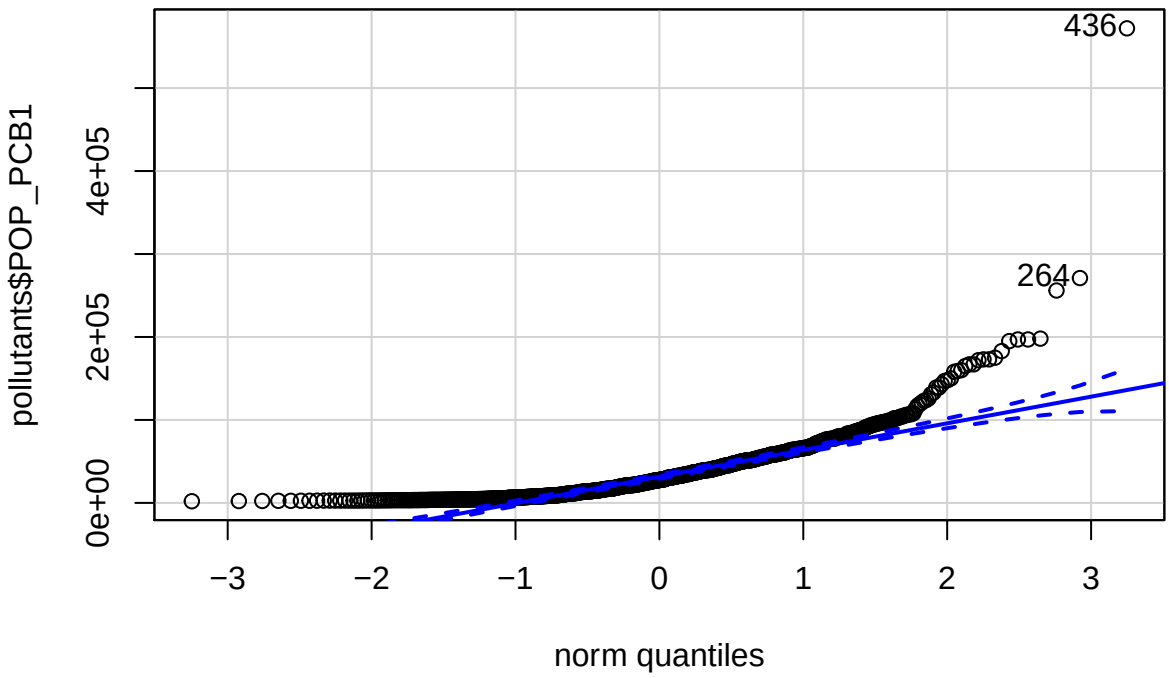



```
#show the qqPlotogram of all covariates
qqPlot(pollutants$length)
```



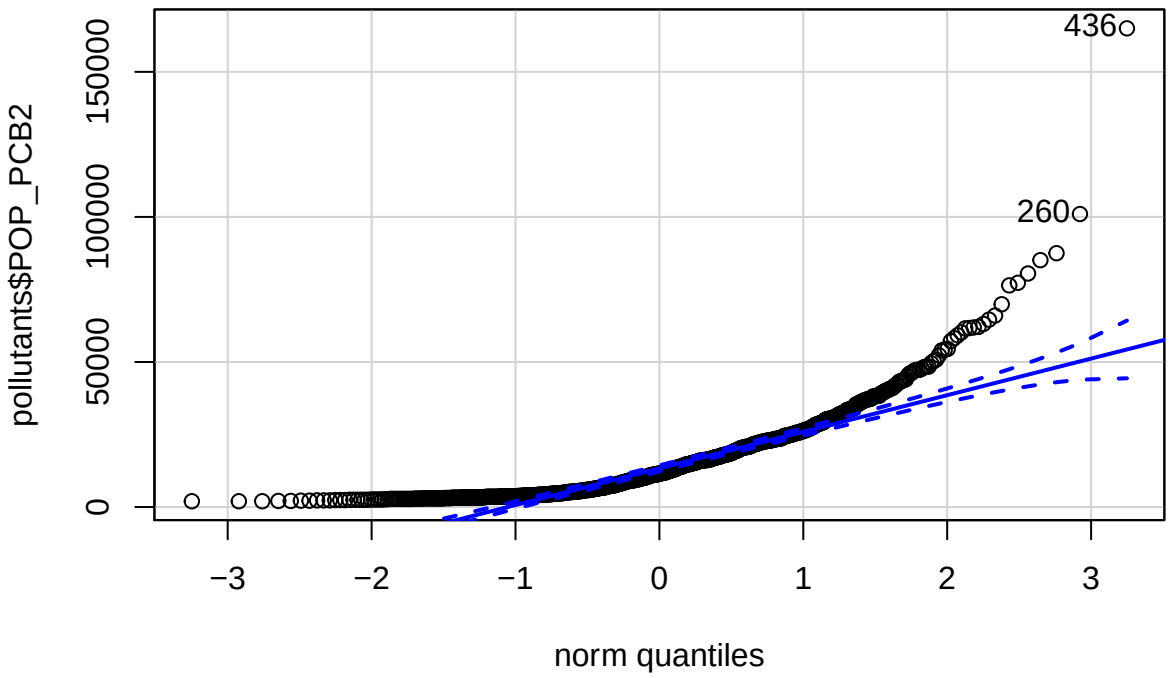

```
## [1] 420  38
```

```
qqPlot(pollutants$POP_PCB1)
```



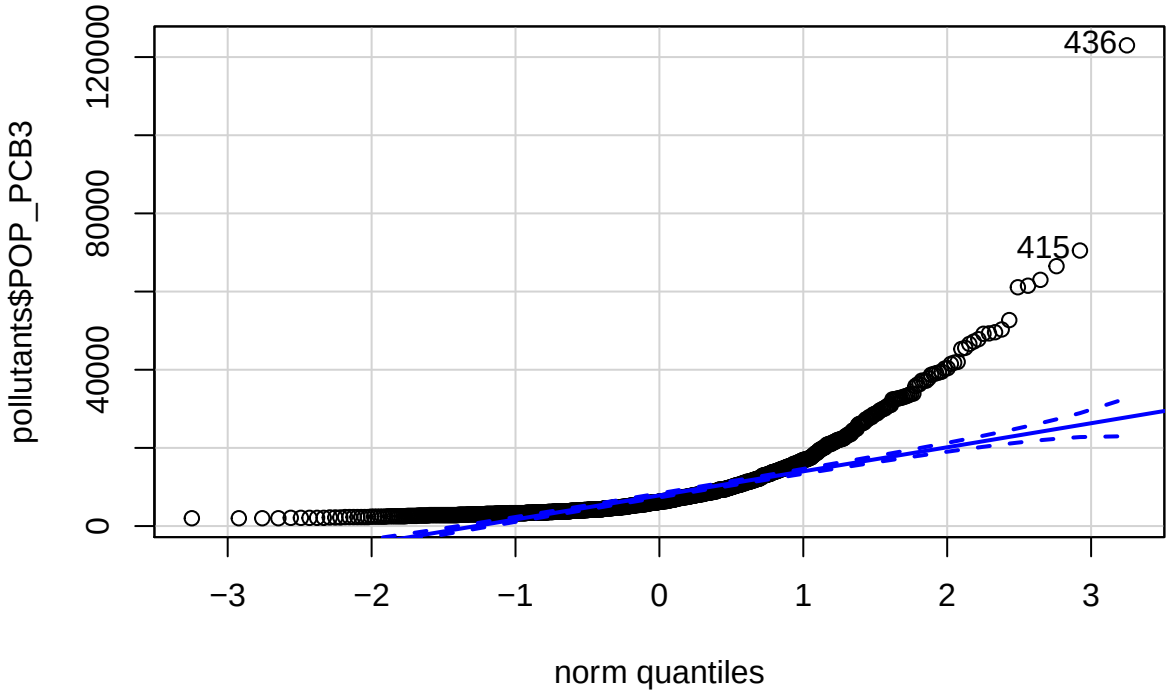

```
## [1] 436 264
```

```
qqPlot(pollutants$POP_PCB2)
```



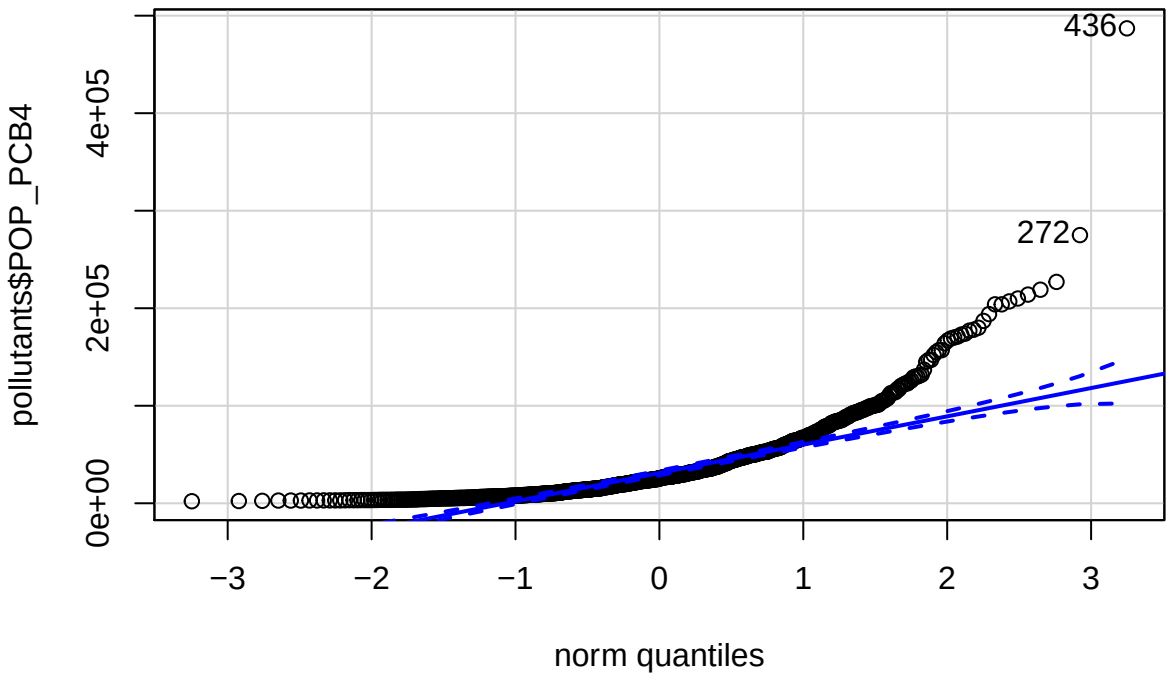

```
## [1] 436 260
```

```
qqPlot(pollutants$POP_PCB3)
```



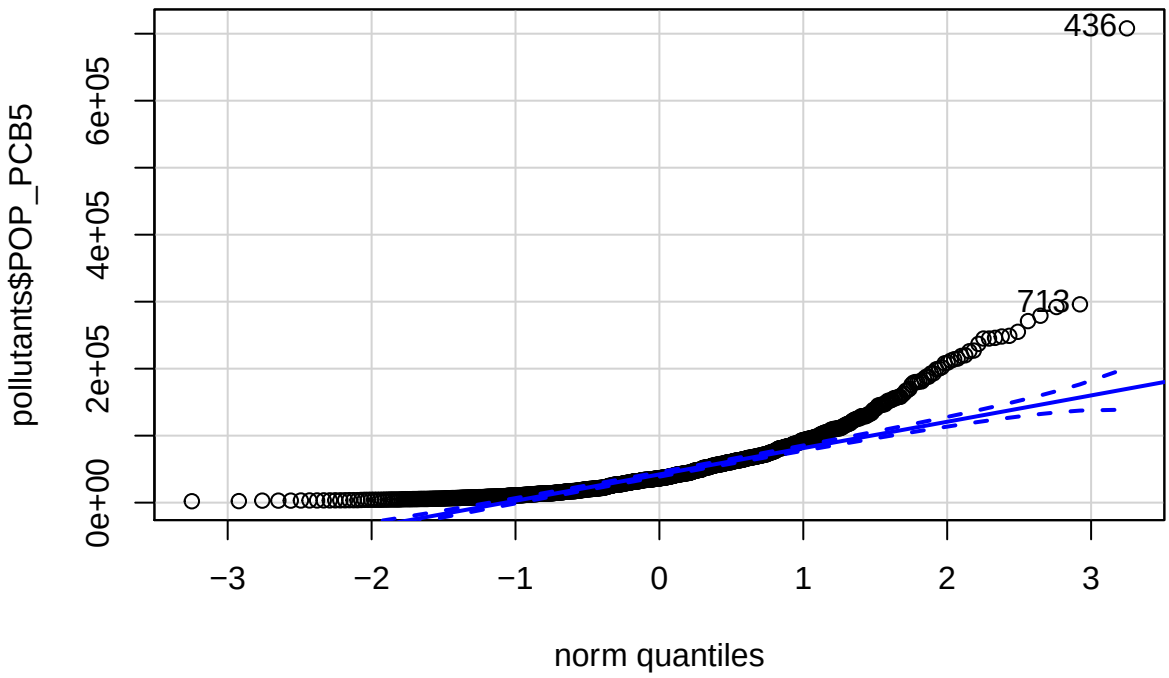

```
## [1] 436 415
```

```
qqPlot(pollutants$POP_PCB4)
```



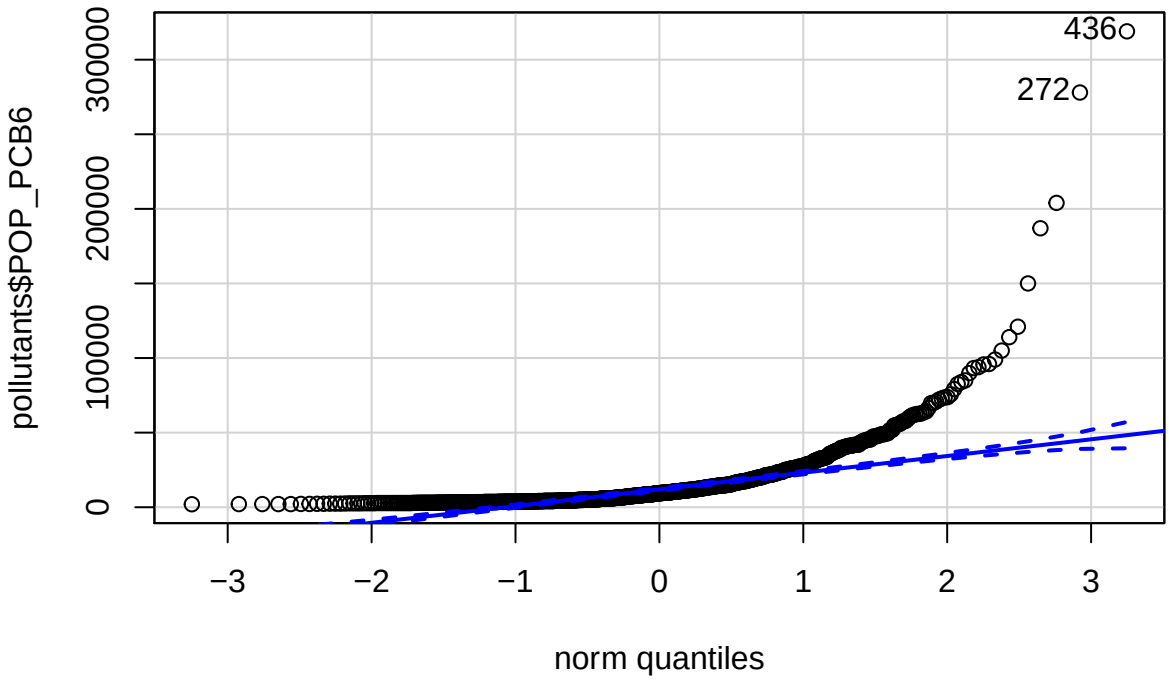

```
## [1] 436 272
```

```
qqPlot(pollutants$POP_PCB5)
```



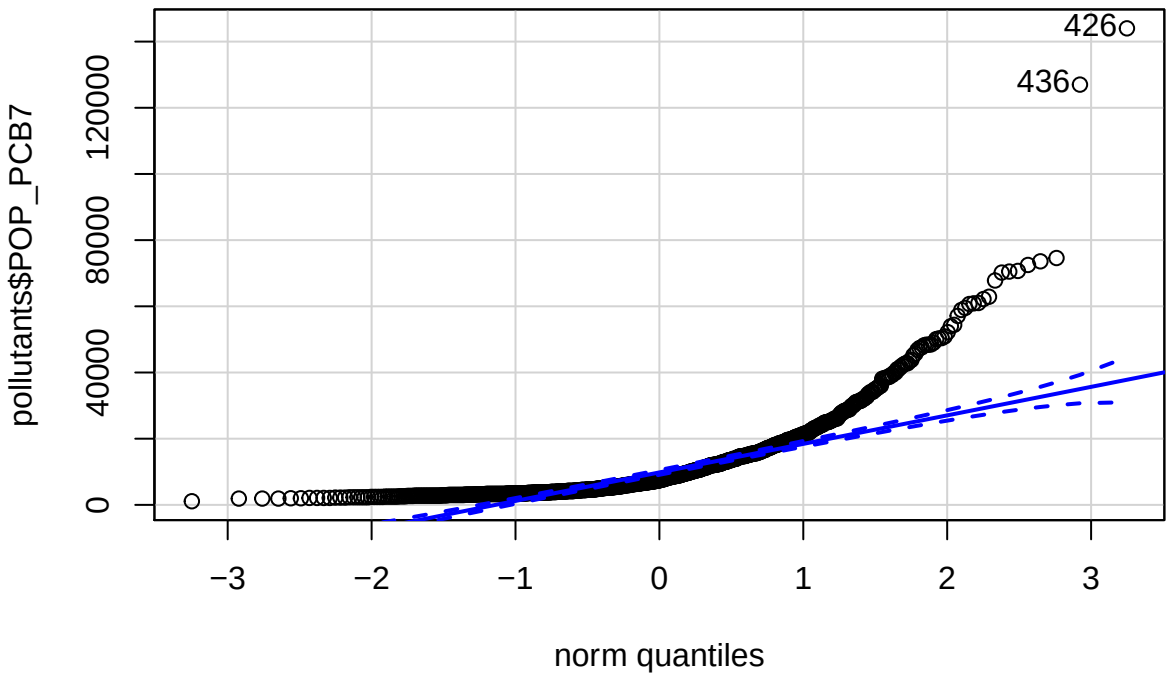

```
## [1] 436 713
```

```
qqPlot(pollutants$POP_PCB6)
```



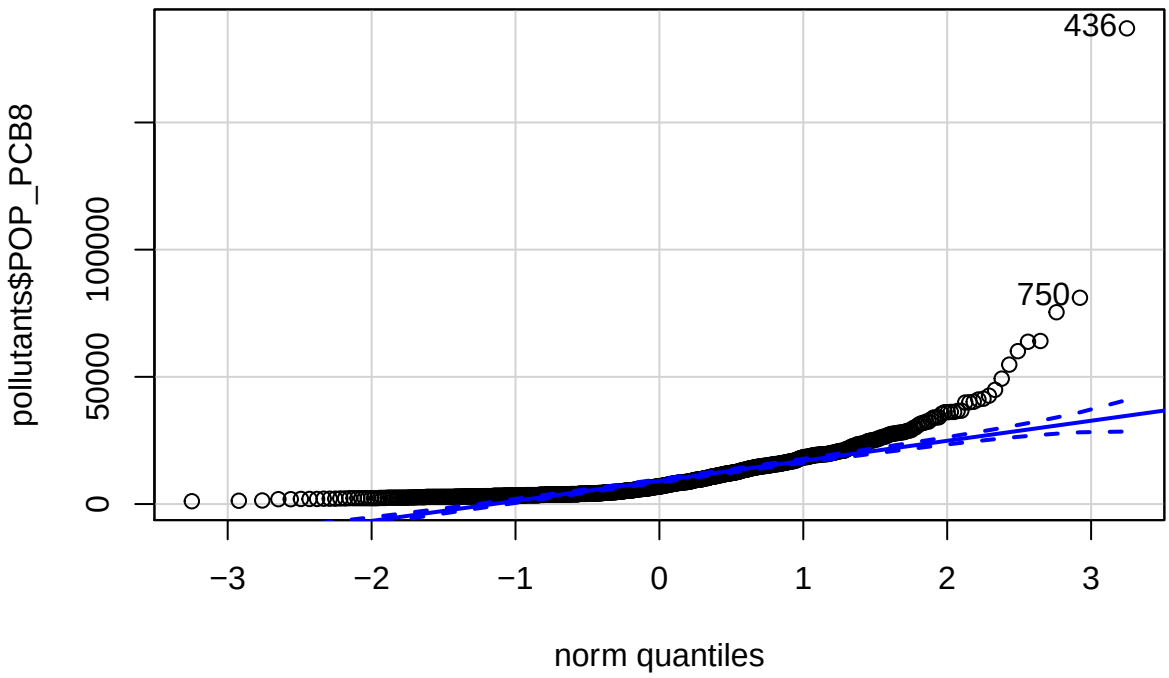

```
## [1] 436 272
```

```
qqPlot(pollutants$POP_PCB7)
```



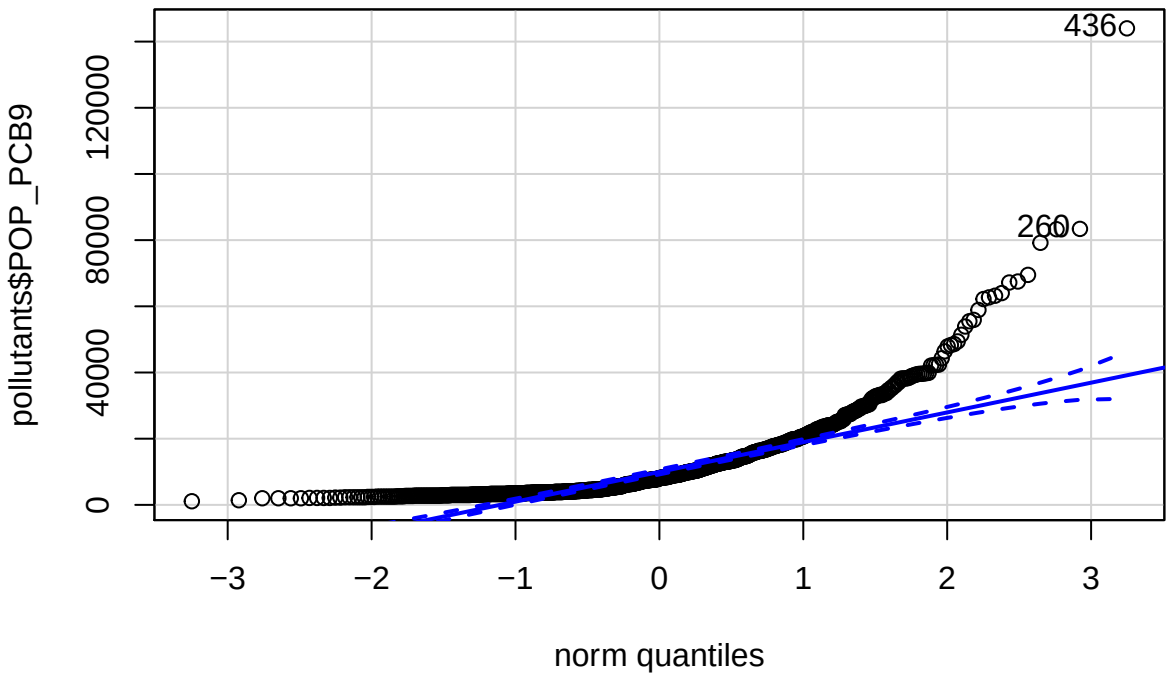

```
## [1] 426 436
```

```
qqPlot(pollutants$POP_PCB8)
```



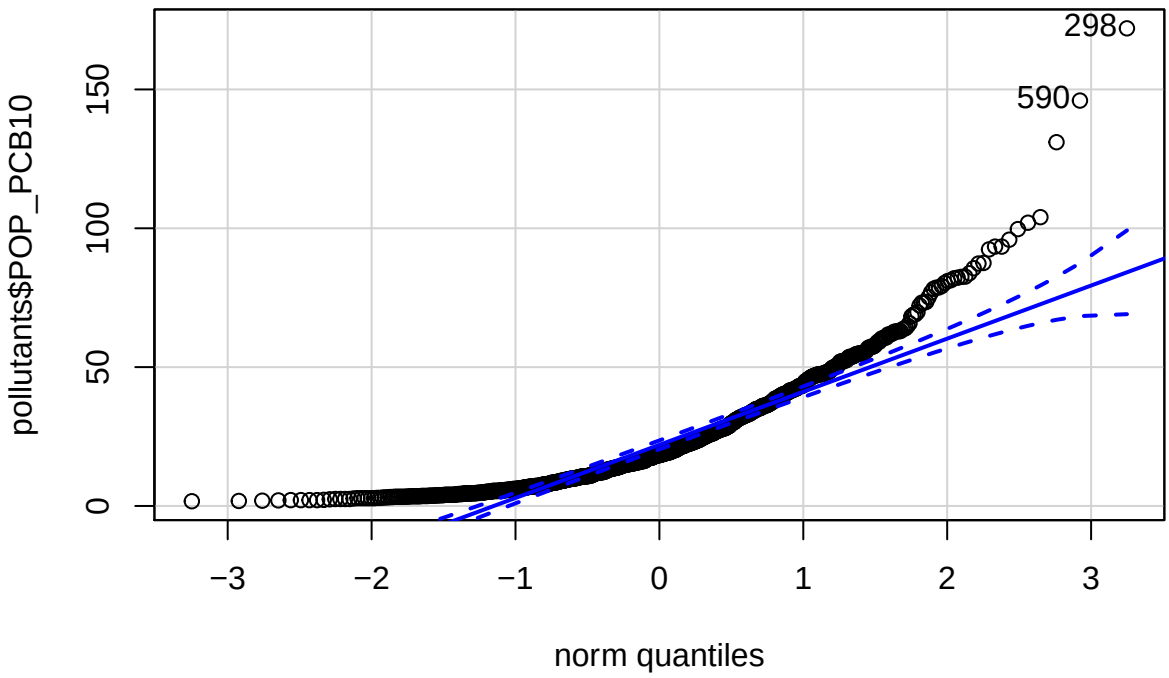

```
## [1] 436 750
```

```
qqPlot(pollutants$POP_PCB9)
```



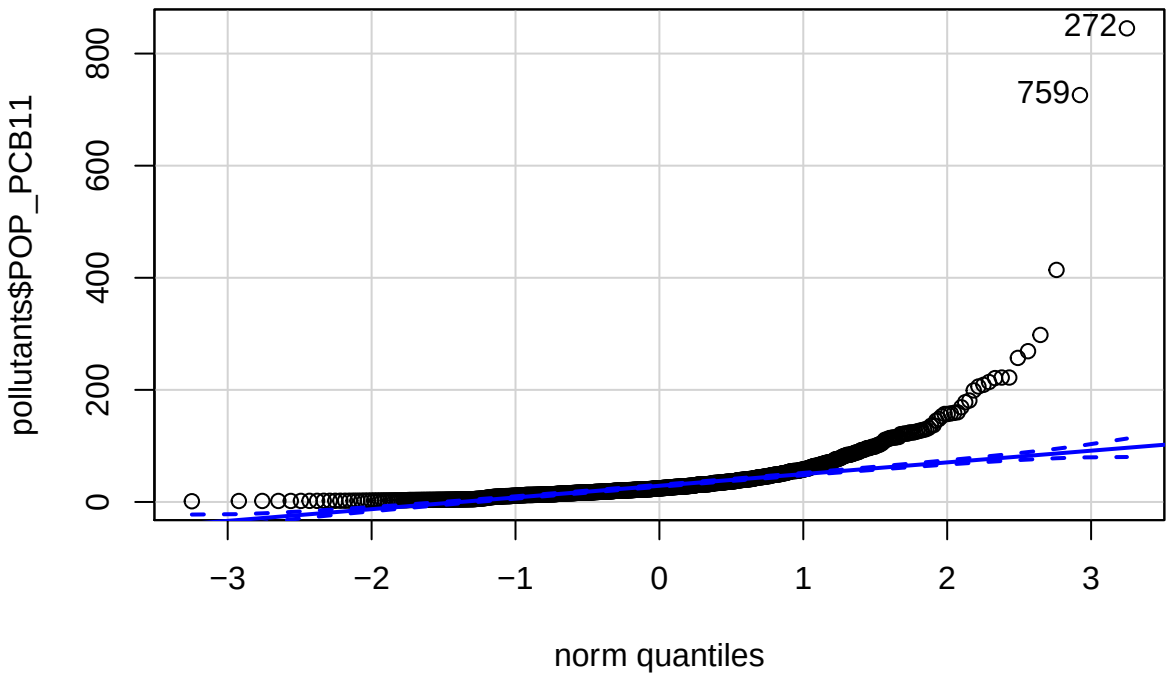

```
## [1] 436 260
```

```
qqPlot(pollutants$POP_PCB10)
```



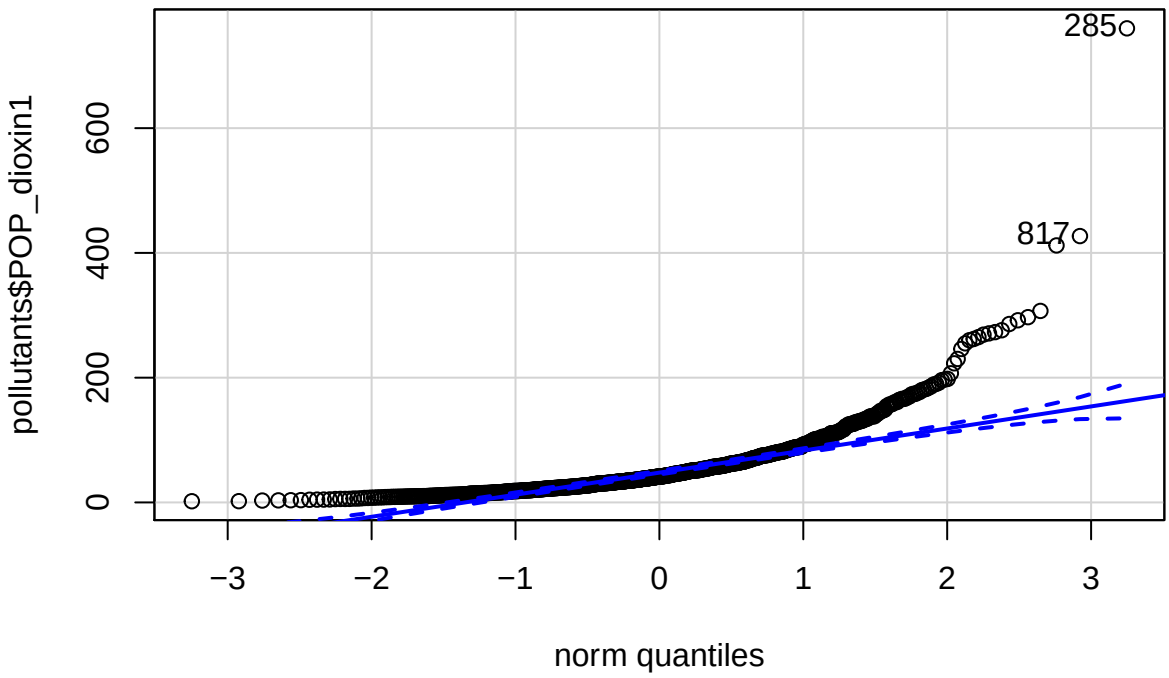

```
## [1] 298 590
```

```
qqPlot(pollutants$POP_PCB11)
```



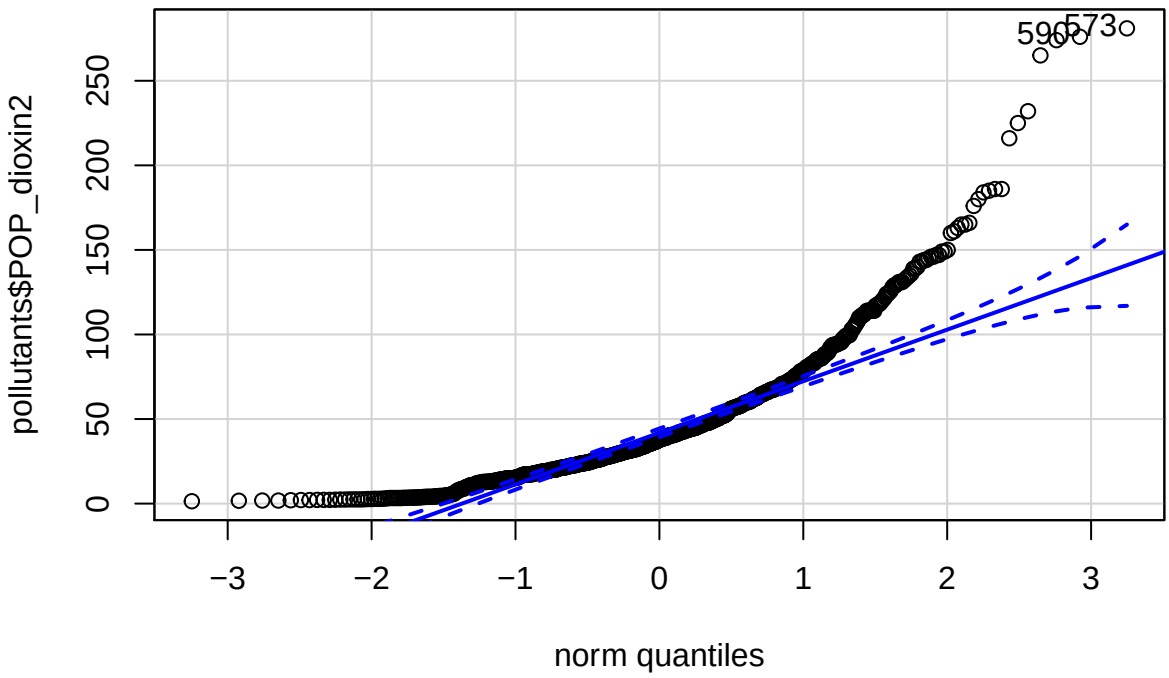

```
## [1] 272 759
```

```
qqPlot(pollutants$POP_dioxin1)
```



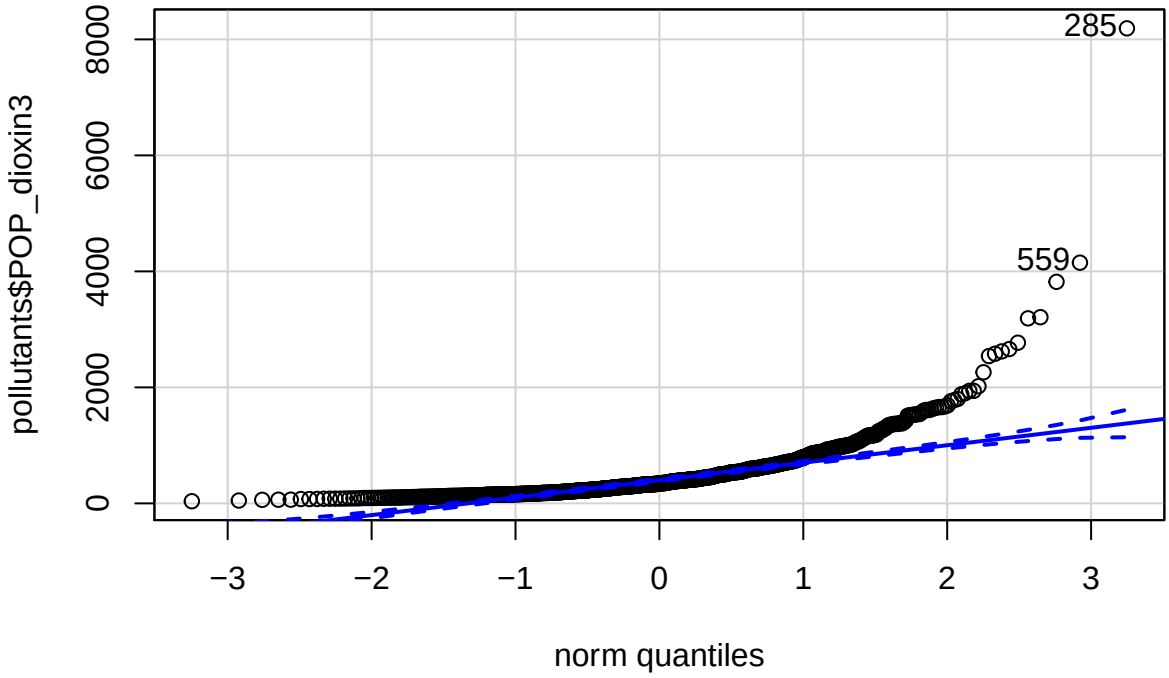

```
## [1] 285 817
```

```
qqPlot(pollutants$POP_dioxin2)
```



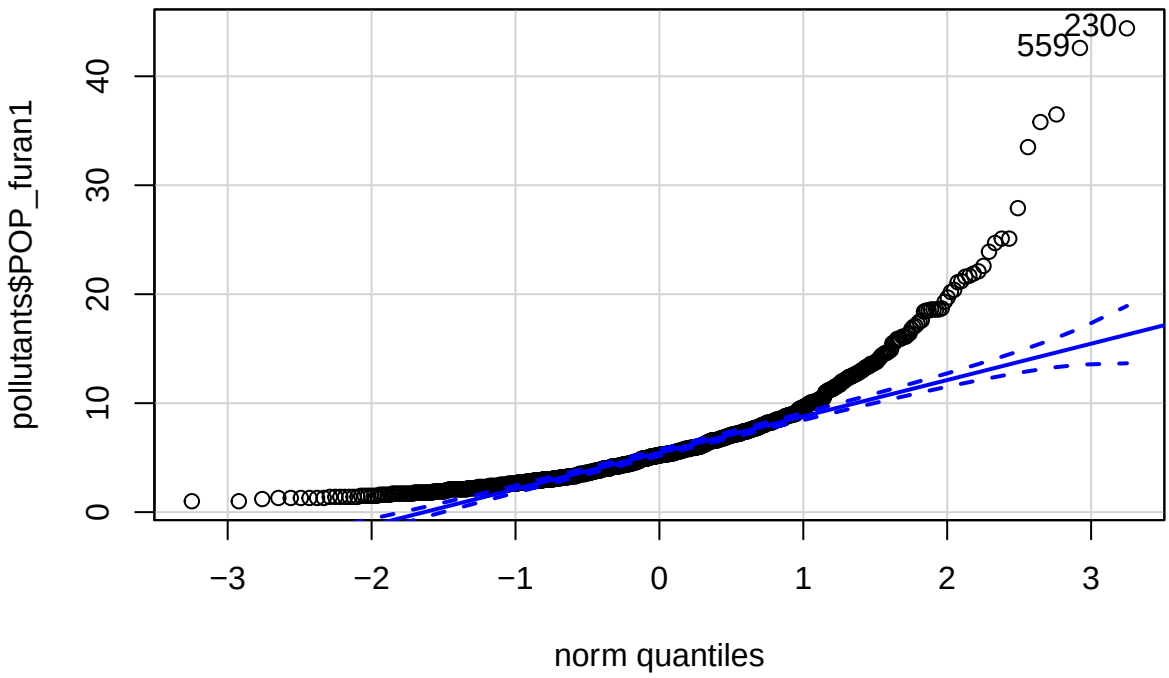

```
## [1] 573 590
```

```
qqPlot(pollutants$POP_dioxin3)
```



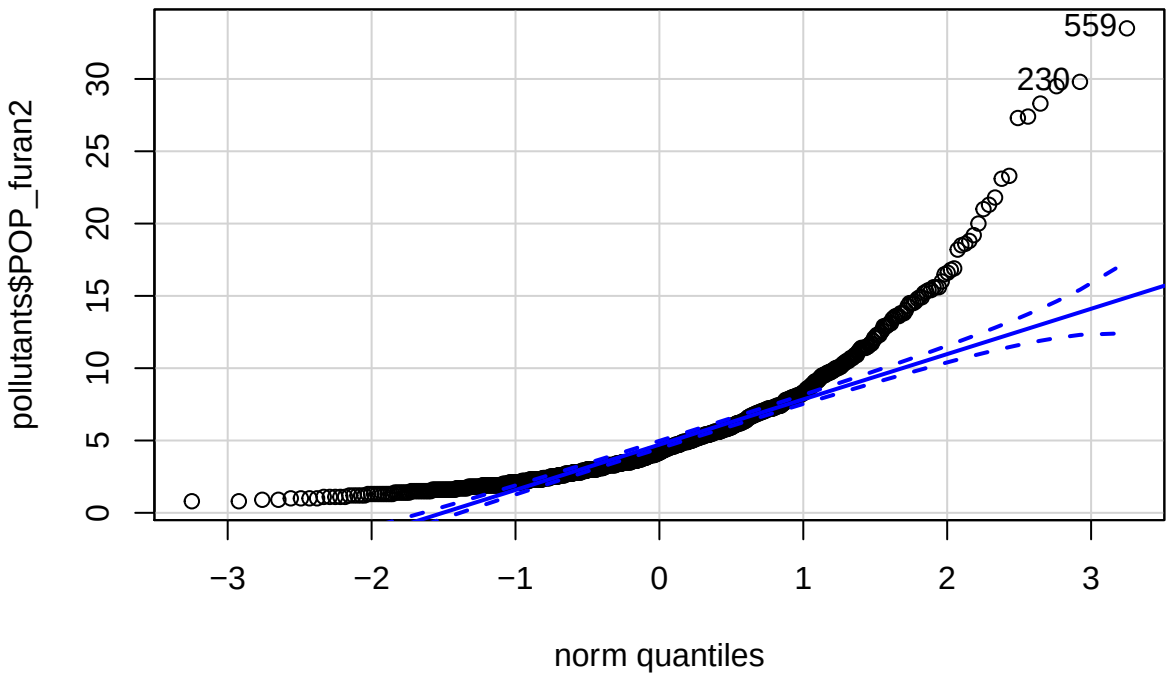

```
## [1] 285 559
```

```
qqPlot(pollutants$POP_furan1)
```



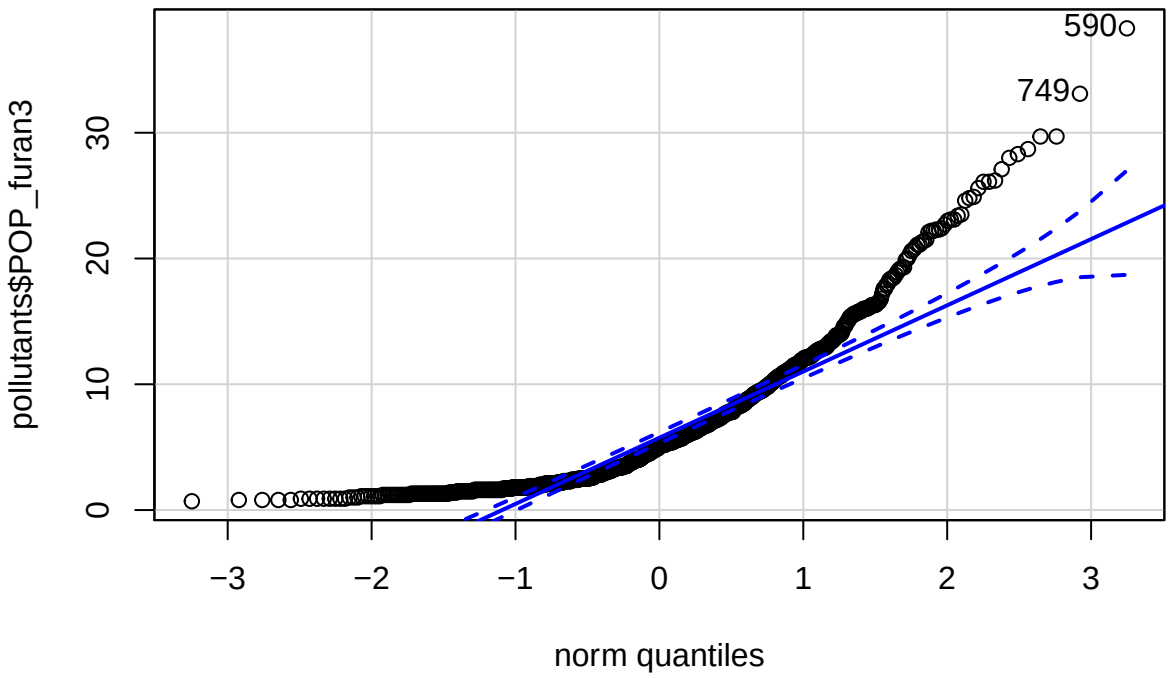

```
## [1] 230 559
```

```
qqPlot(pollutants$POP_furan2)
```



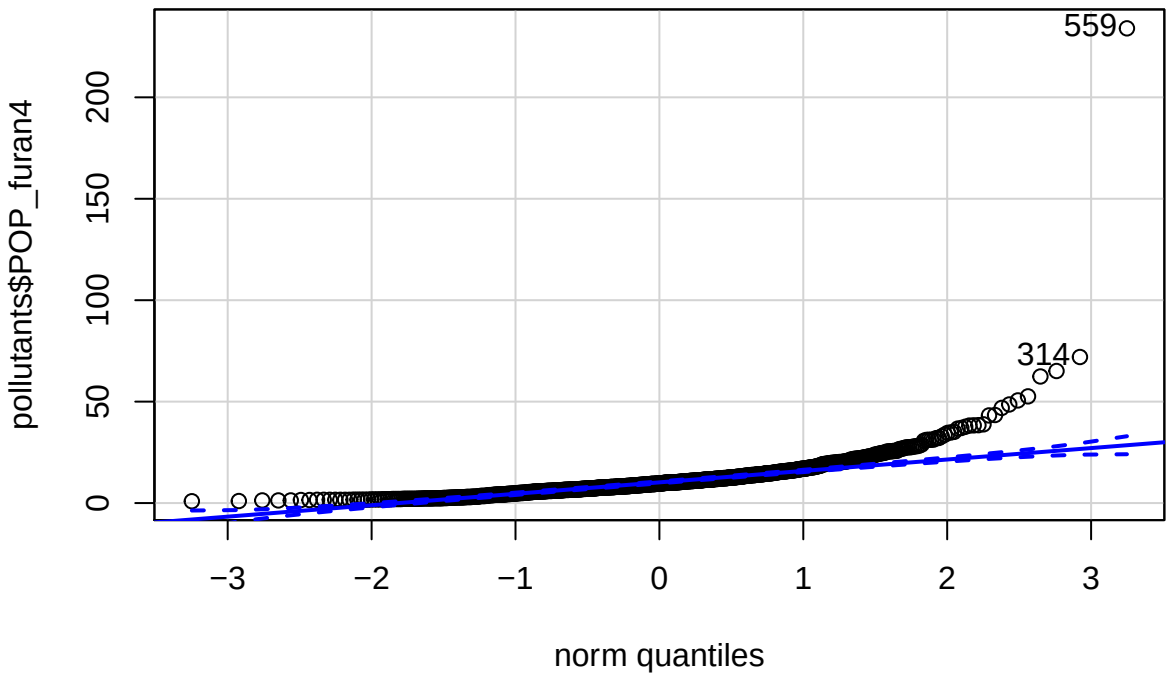

```
## [1] 559 230
```

```
qqPlot(pollutants$POP_furan3)
```



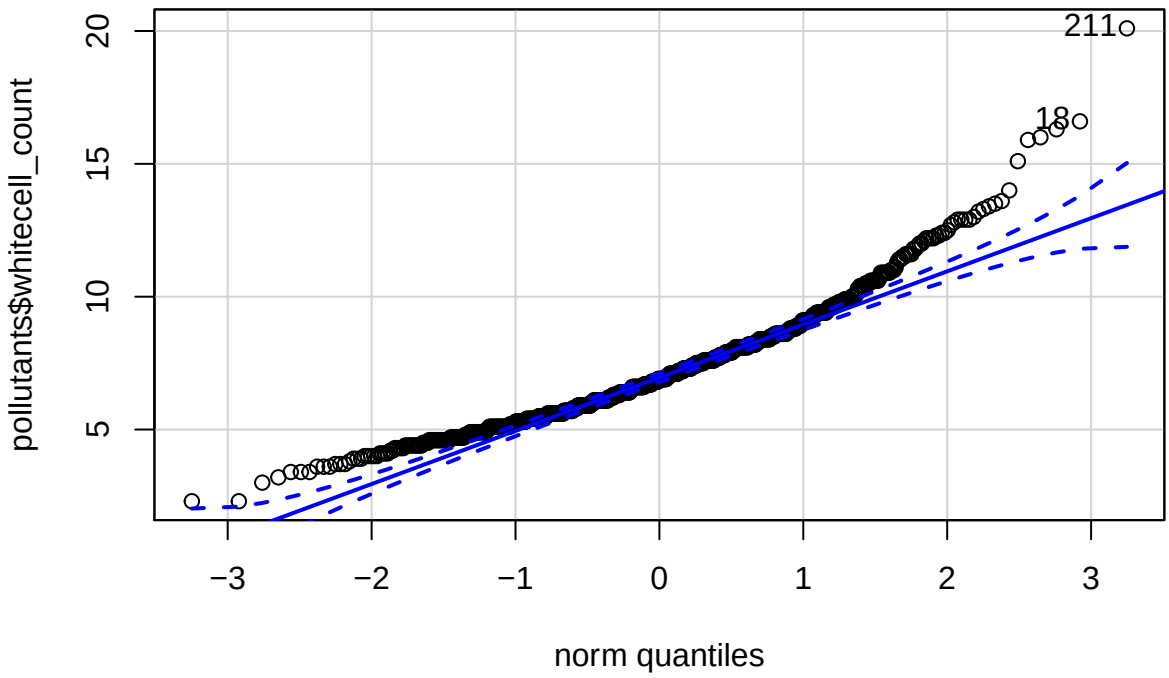

```
## [1] 590 749
```

```
qqPlot(pollutants$POP_furan4)
```



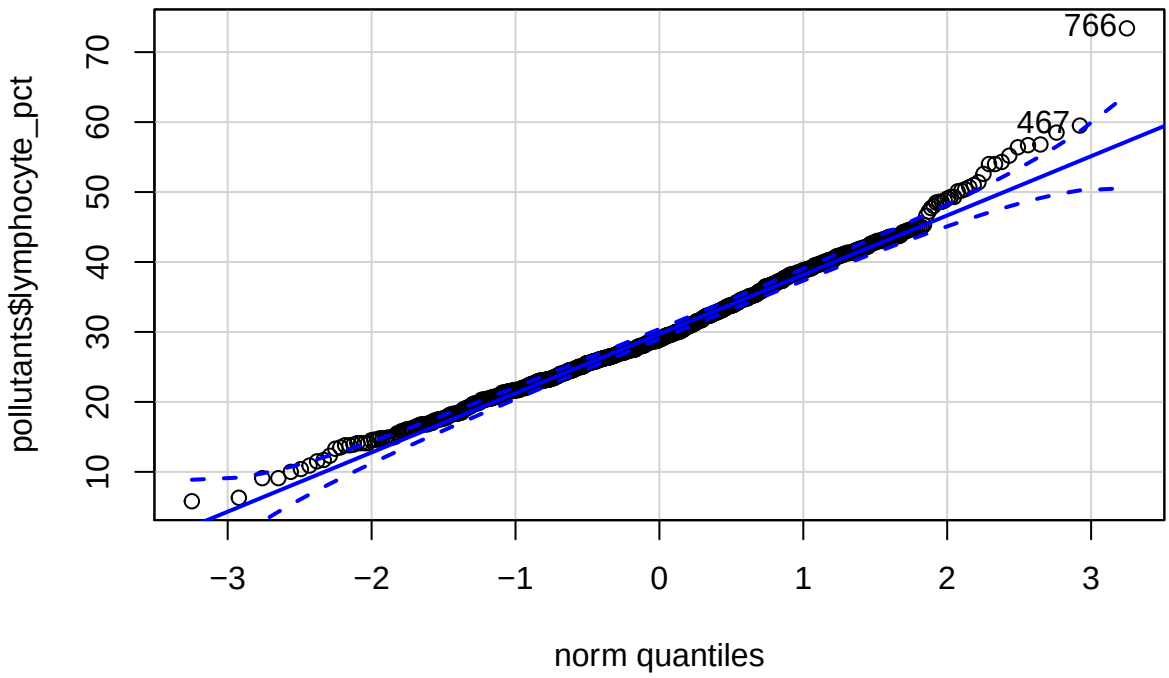

```
## [1] 559 314
```

```
qqPlot(pollutants$whitecell_count)
```



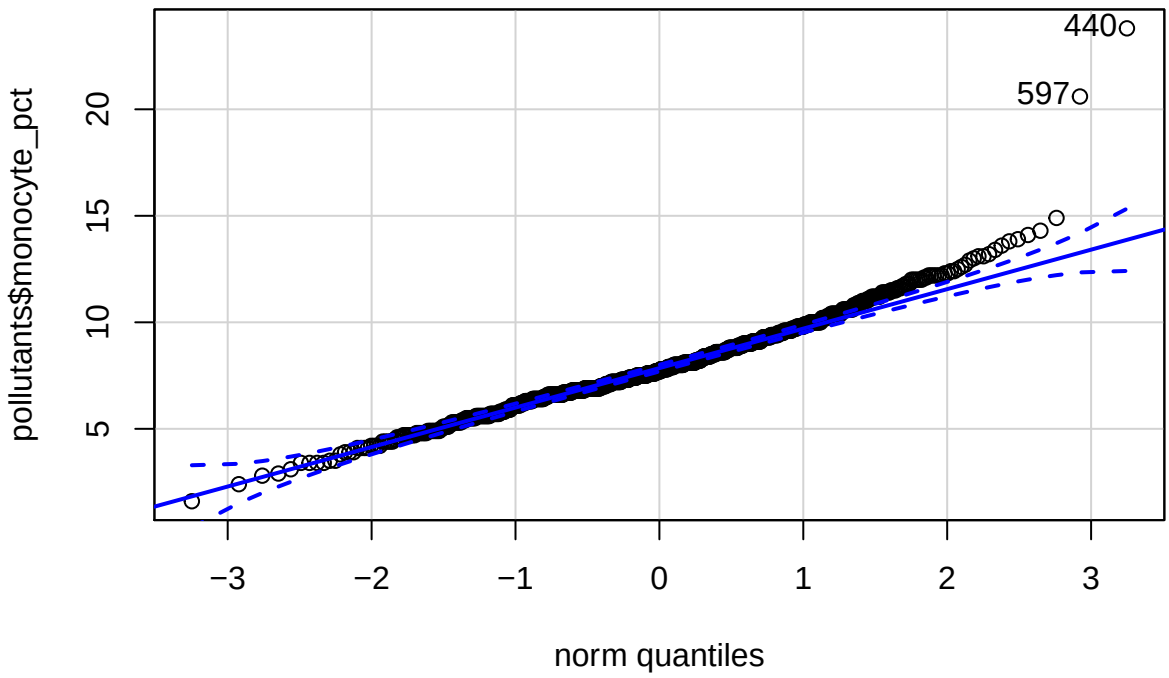

```
## [1] 211  18
```

```
qqPlot(pollutants$lymphocyte_pct)
```



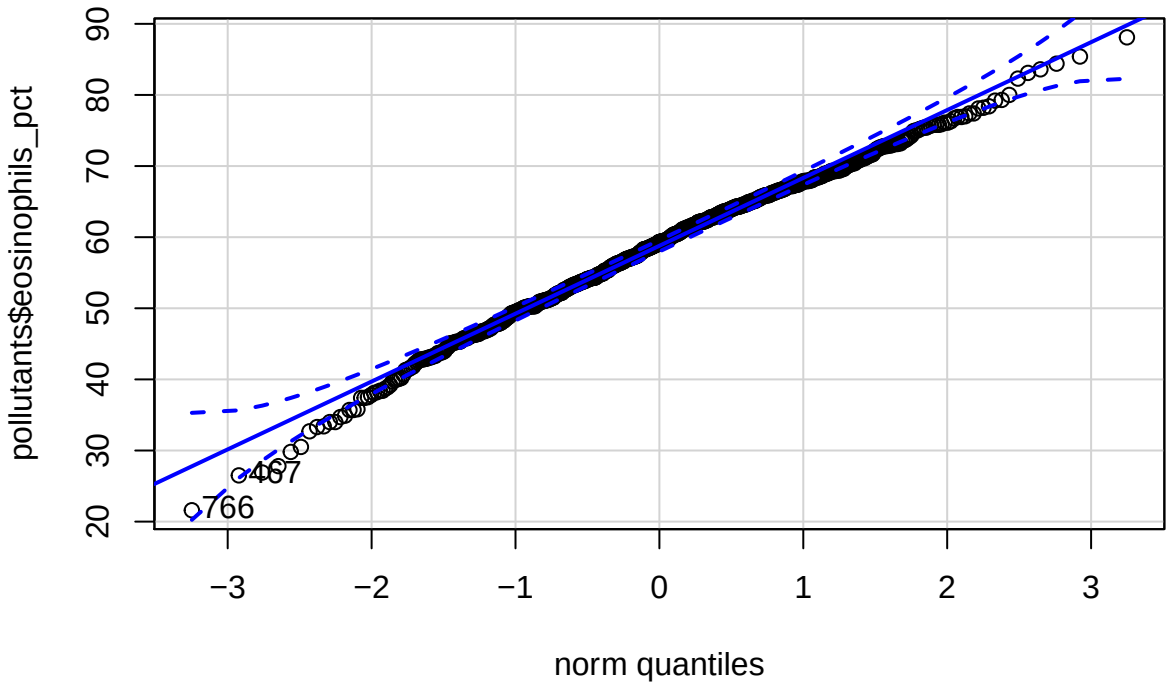

```
## [1] 766 467
```

```
qqPlot(pollutants$monocyte_pct)
```



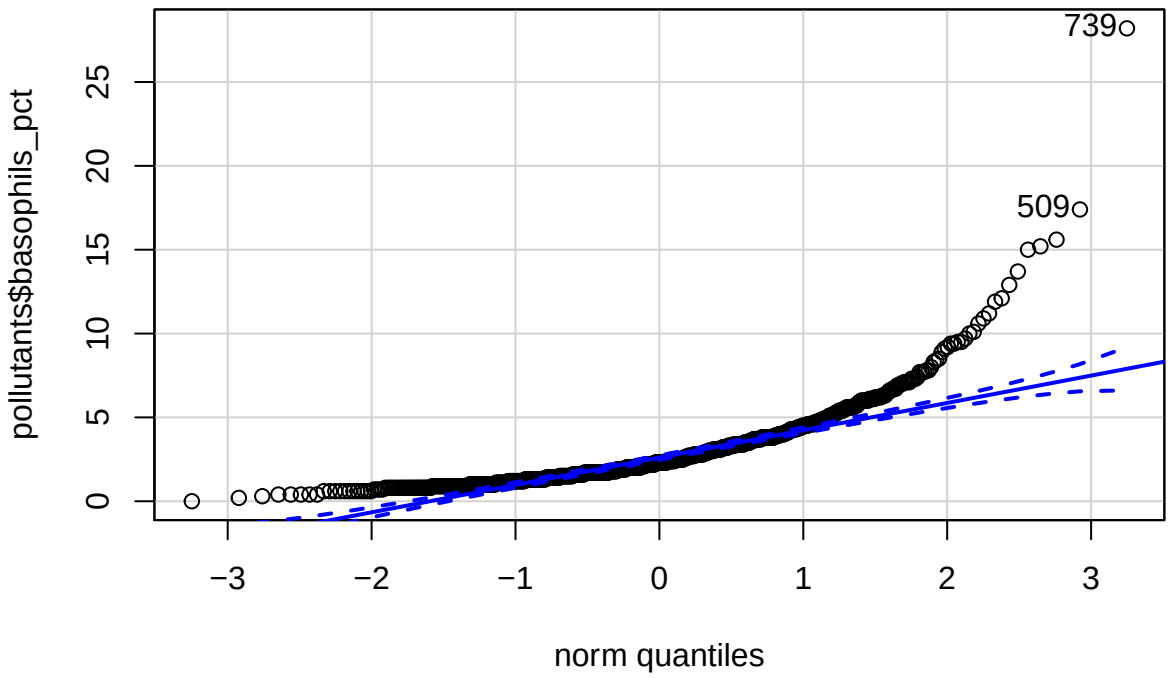

```
## [1] 440 597
```

```
qqPlot(pollutants$eosinophils_pct)
```



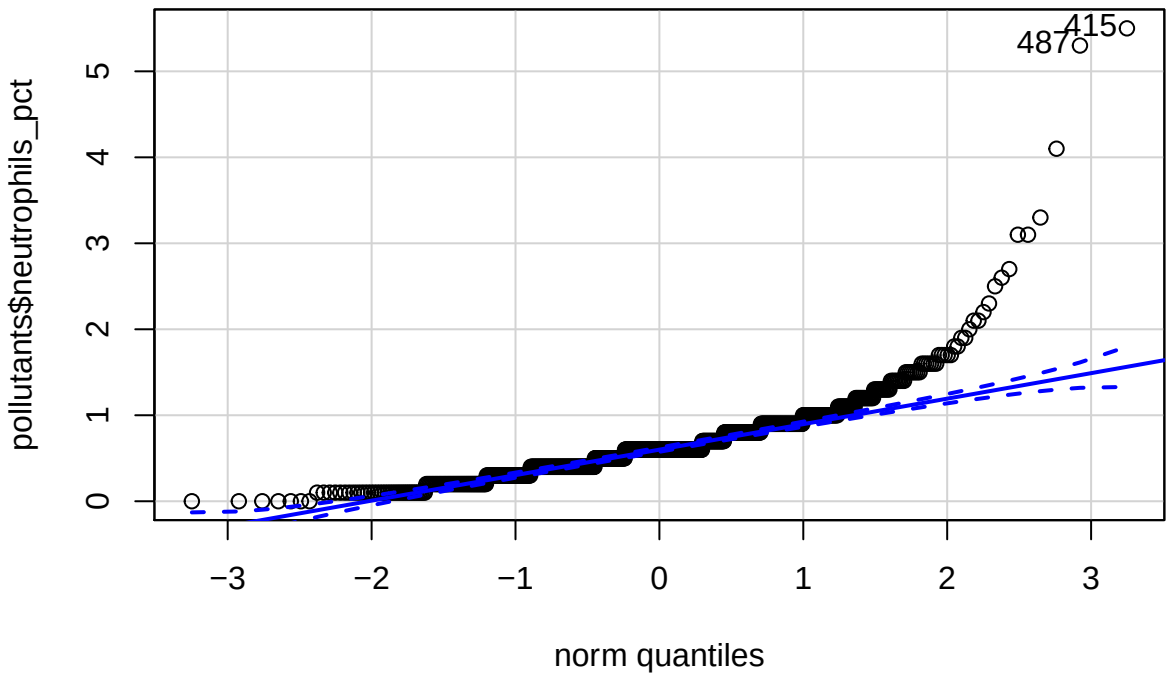

```
## [1] 766 467
```

```
qqPlot(pollutants$basophils_pct)
```



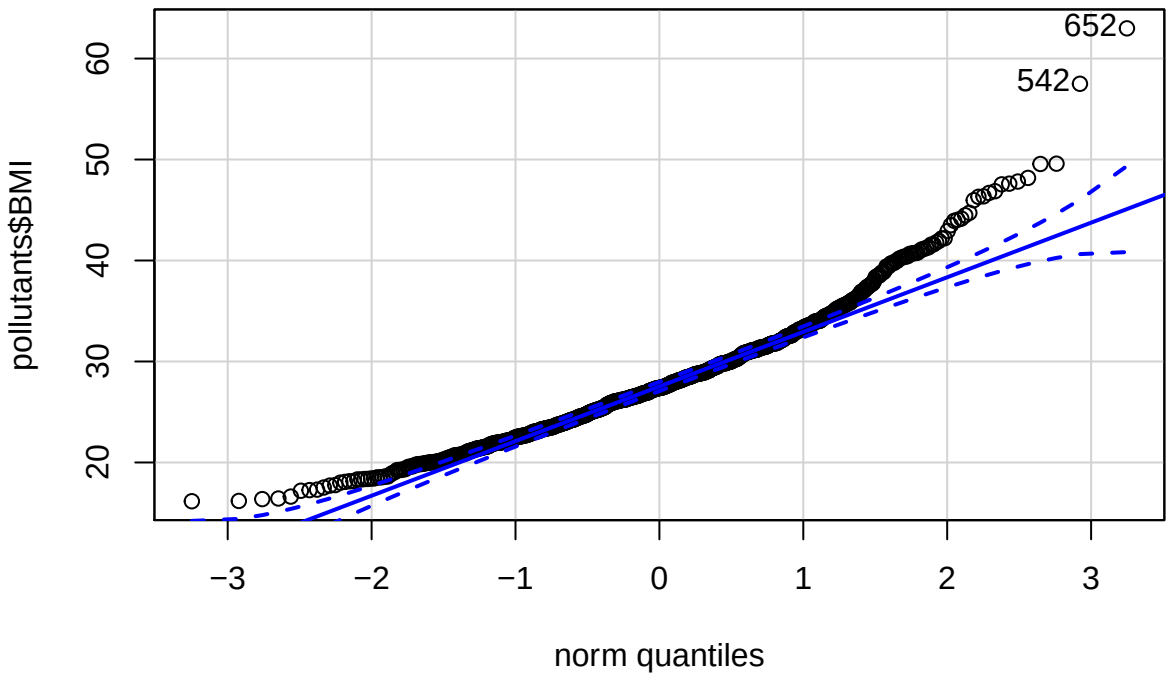

```
## [1] 739 509
```

```
qqPlot(pollutants$neutrophils_pct)
```



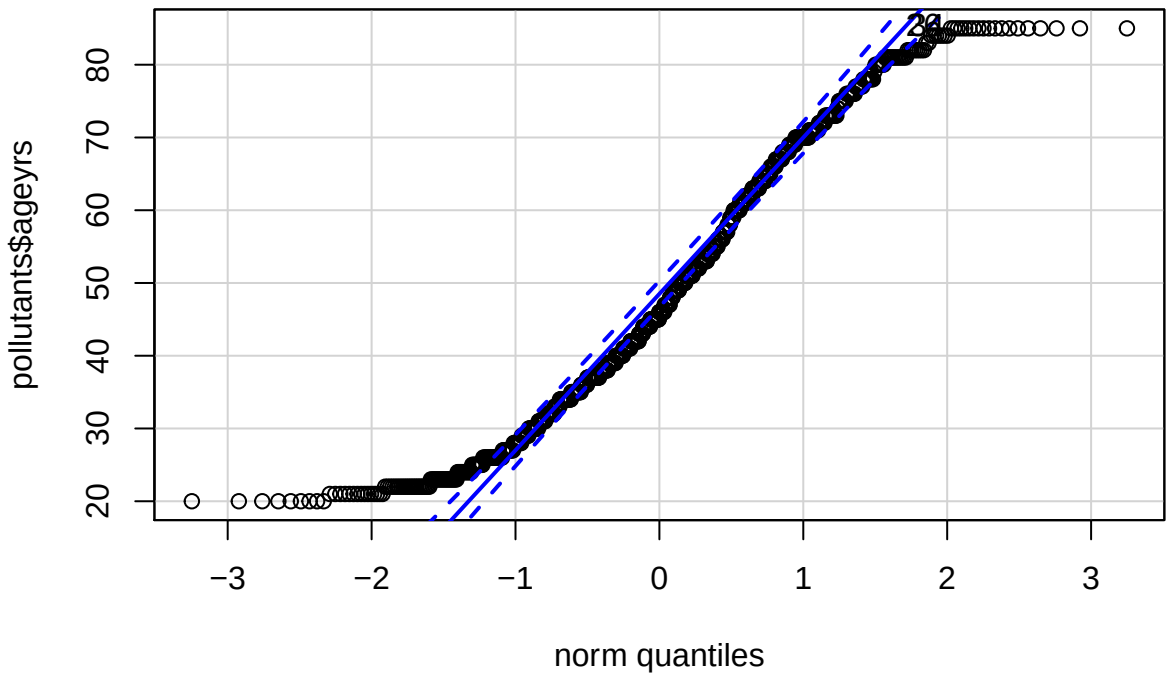

```
## [1] 415 487
```

```
qqPlot(pollutants$BMI)
```



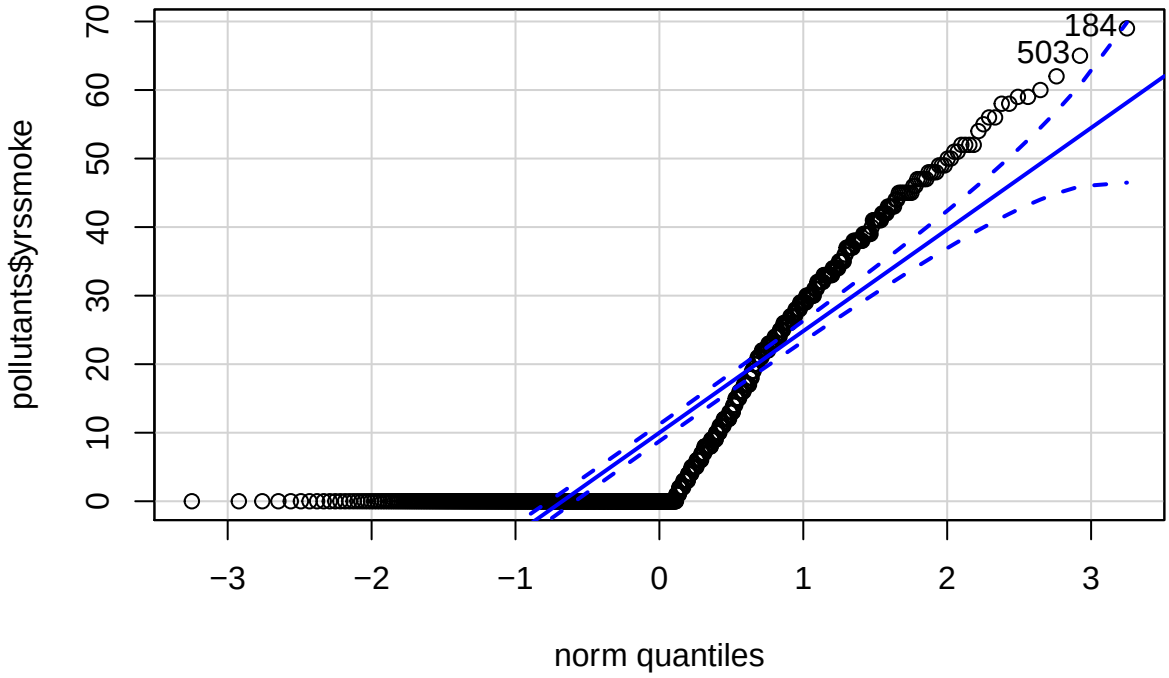

```
## [1] 652 542
```

```
qqPlot(pollutants$ageyrs)
```



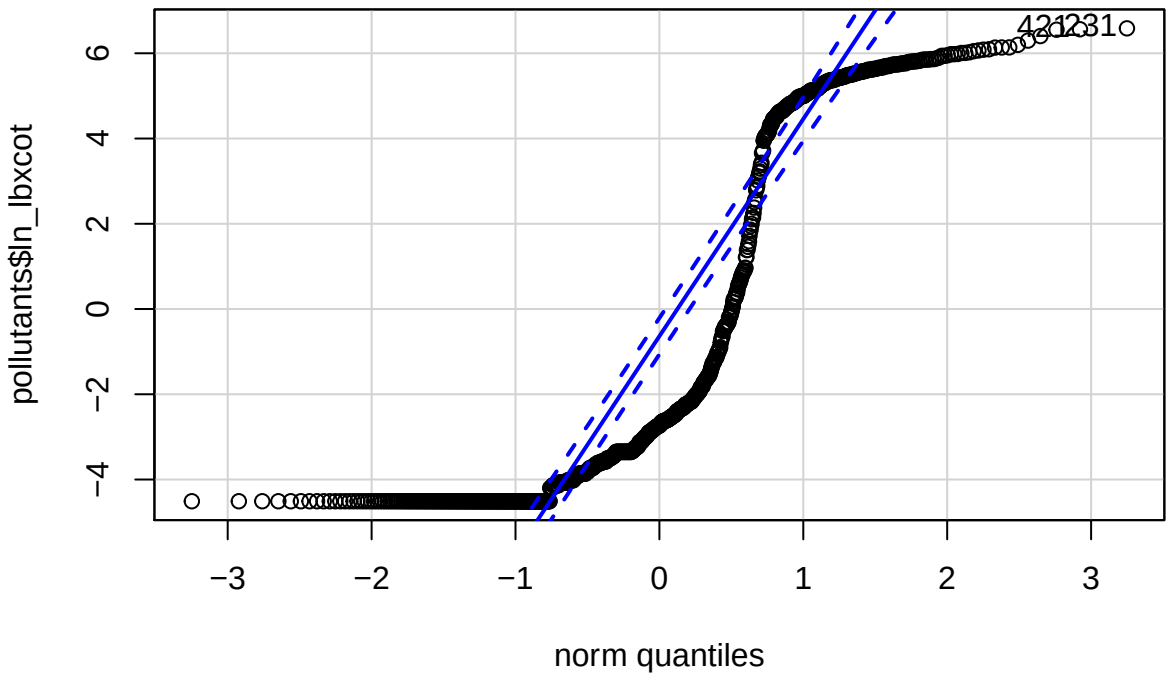

```
## [1] 26 34
```

```
qqPlot(pollutants$yrssmoke)
```



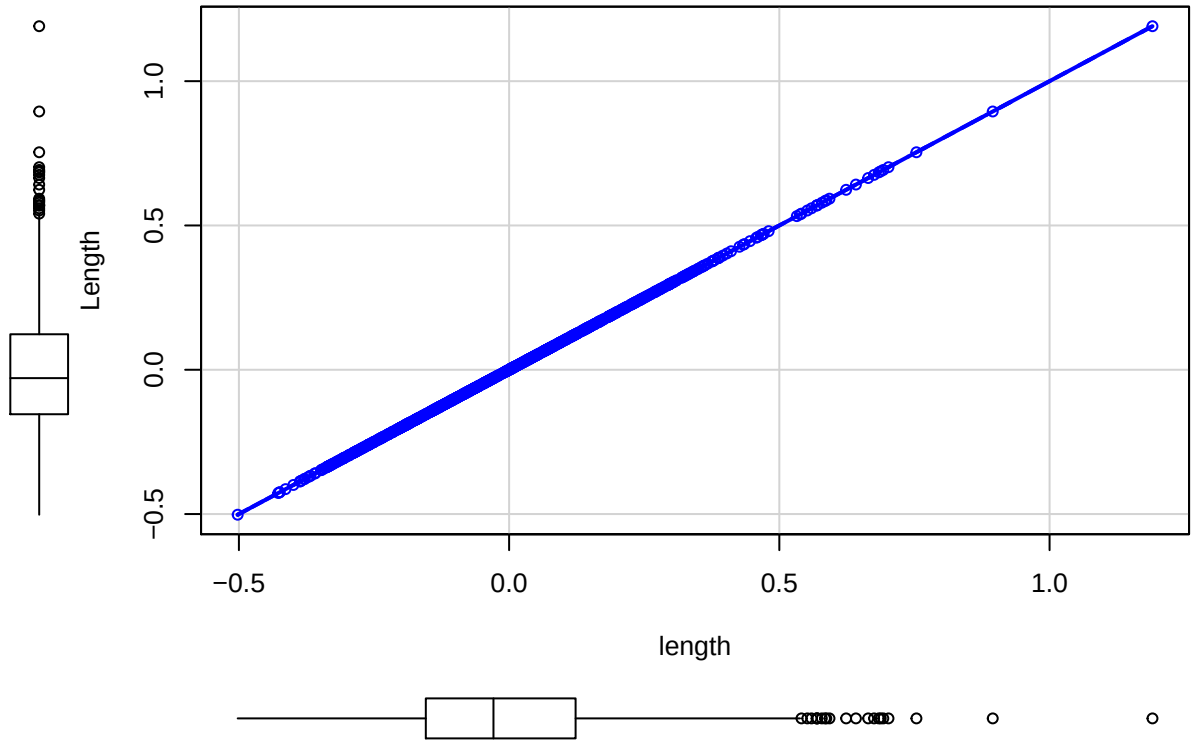

```
## [1] 184 503
```

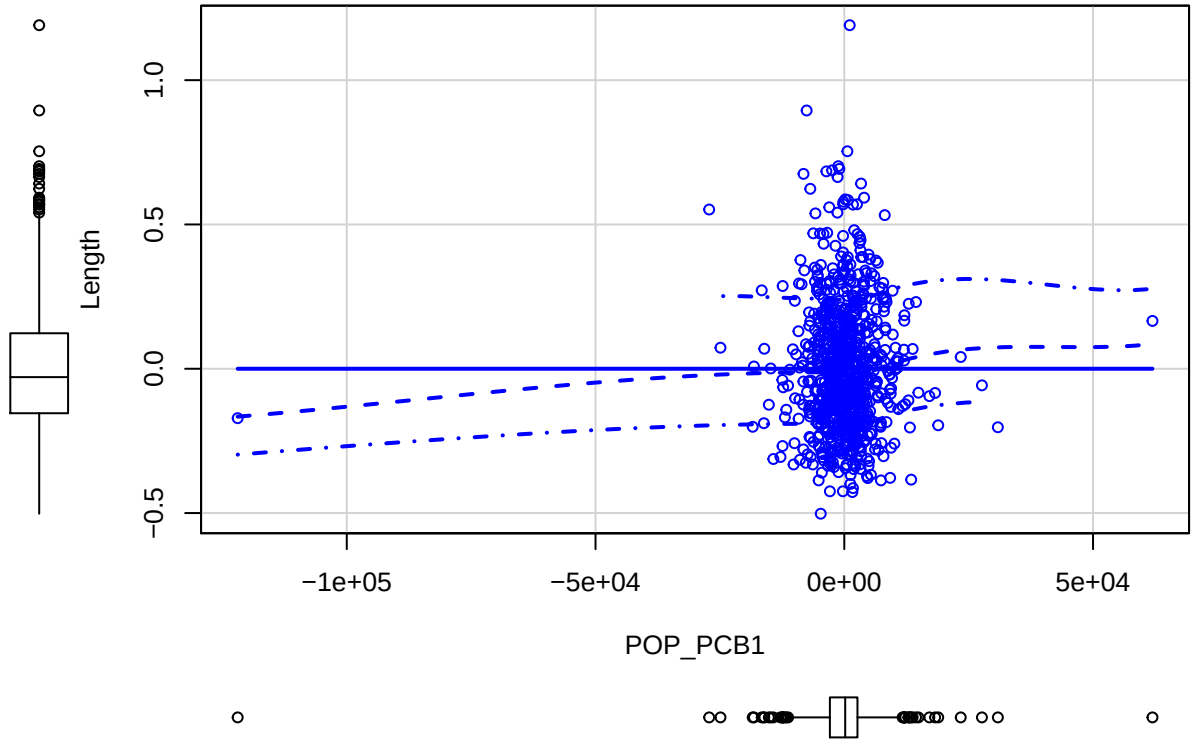
```
qqPlot(pollutants$ln_lbxcot)
```

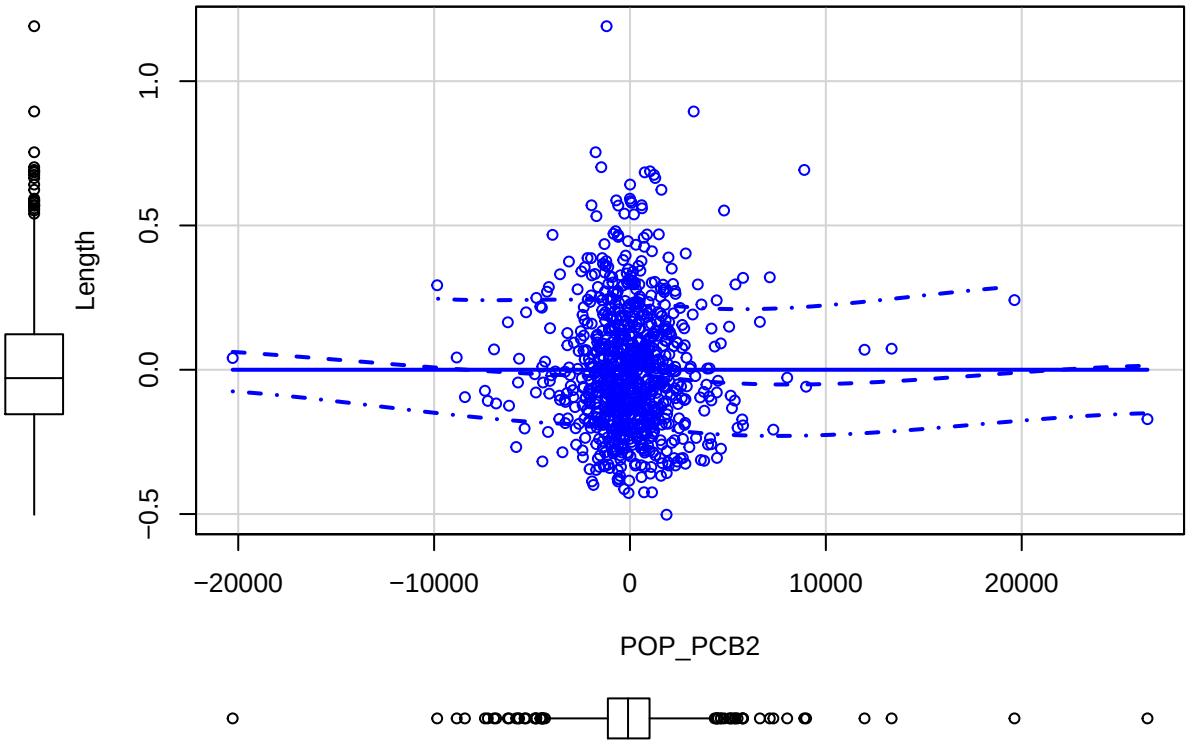


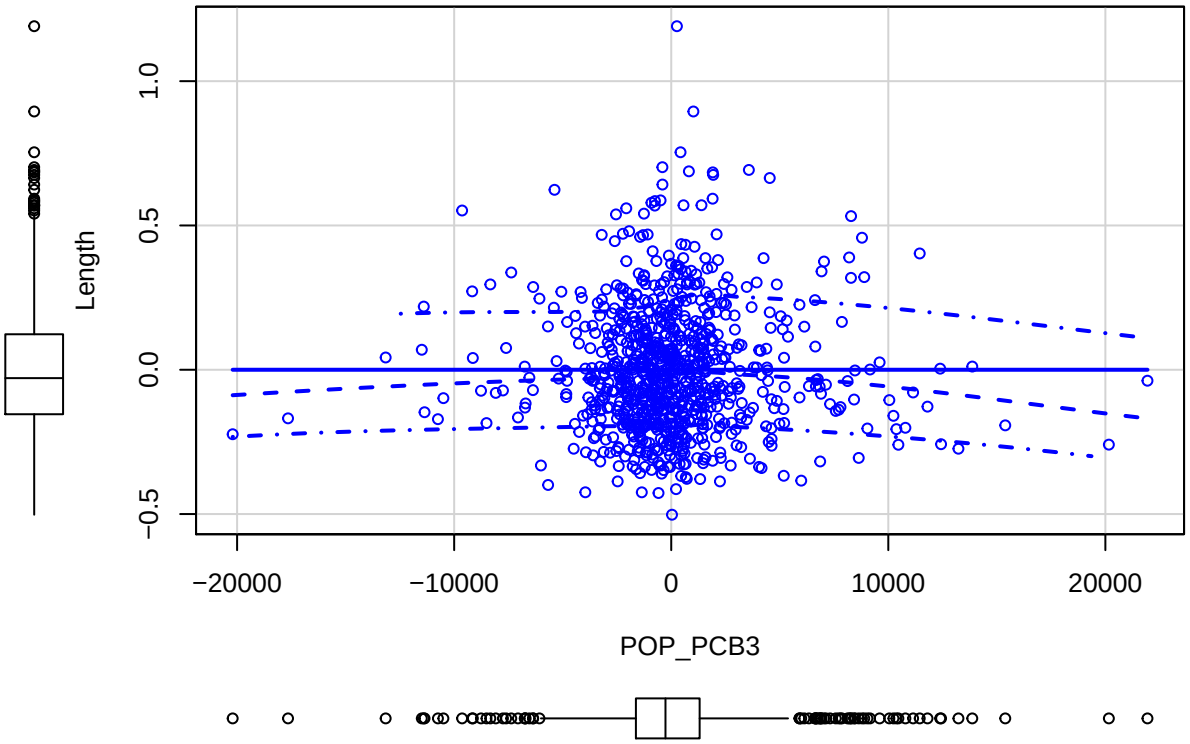

```
## [1] 231 421
```

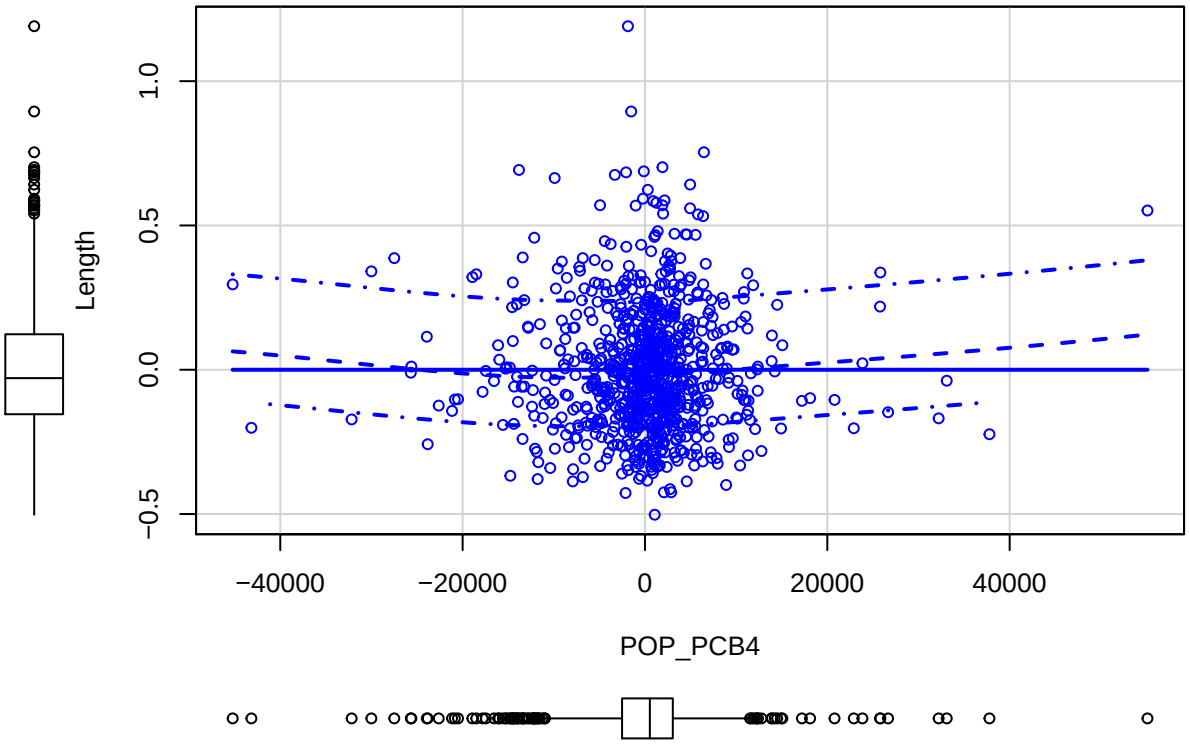
```
#find the linearity (residual y against residual x)
covariates = names(no_cat)
for (name in covariates){
  y_model = lm(paste("length", "~", "."), data = pollutants)
  x_model = lm(paste(name, "~", ".", "- length"), data = pollutants)
  y_resid = resid(y_model)
  x_resid = resid(x_model)
  #linearity
  scatterplot(x_resid, y_resid, ylab = "Length", xlab = name)
}
```

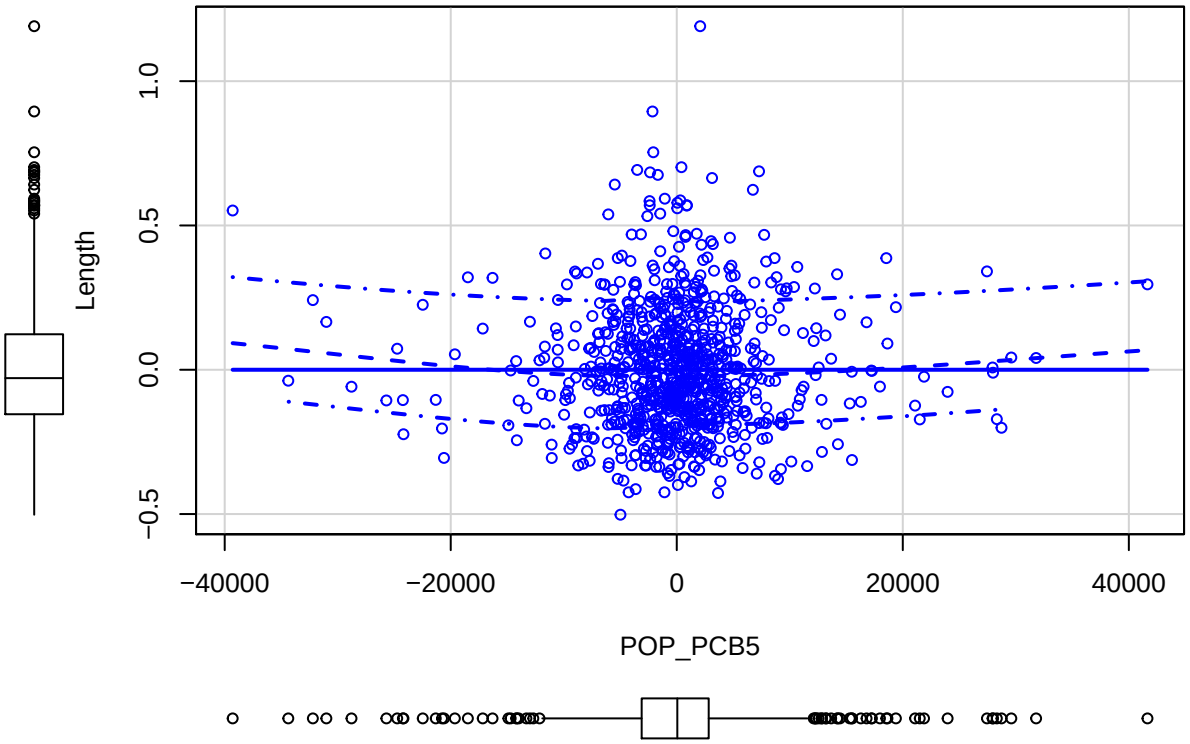

Length
1.0
0.5
0.0
−0.5
−0.5
0.0
0.5
1.0
length

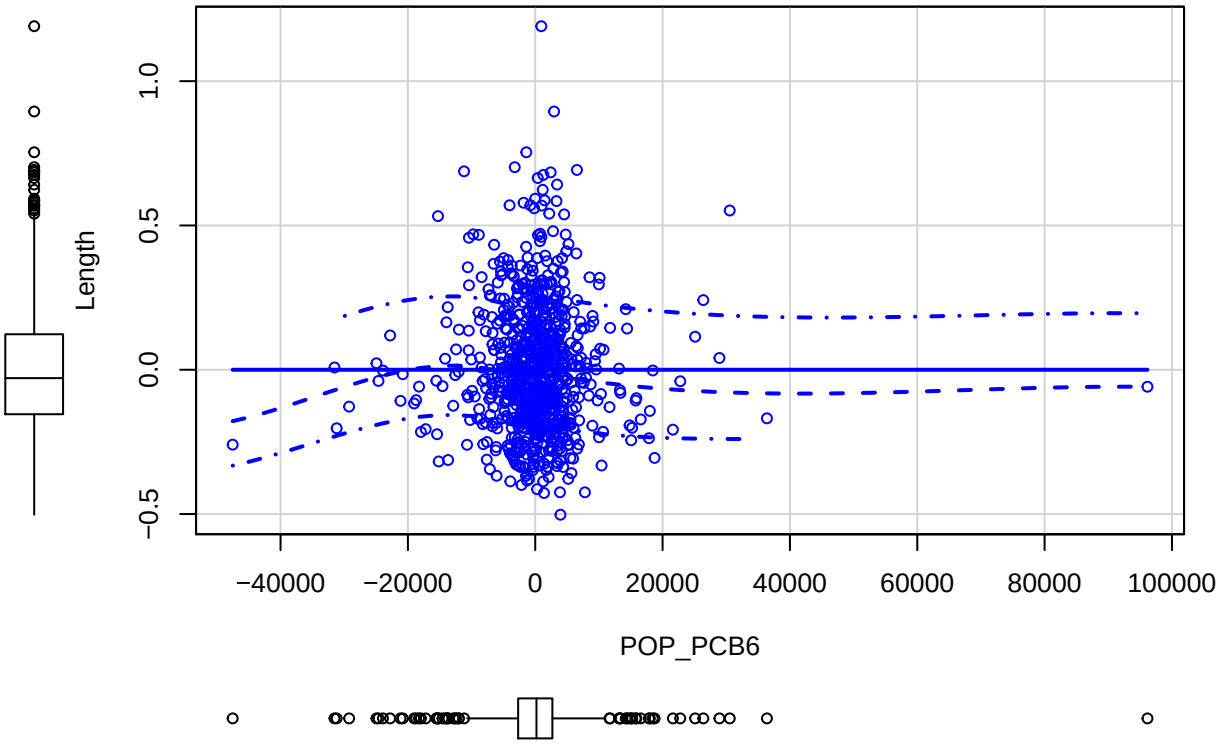

Length
1.0
0.5
0.0
−0.5
−1e+05
−5e+04
0e+00
5e+04
POP_PCB1

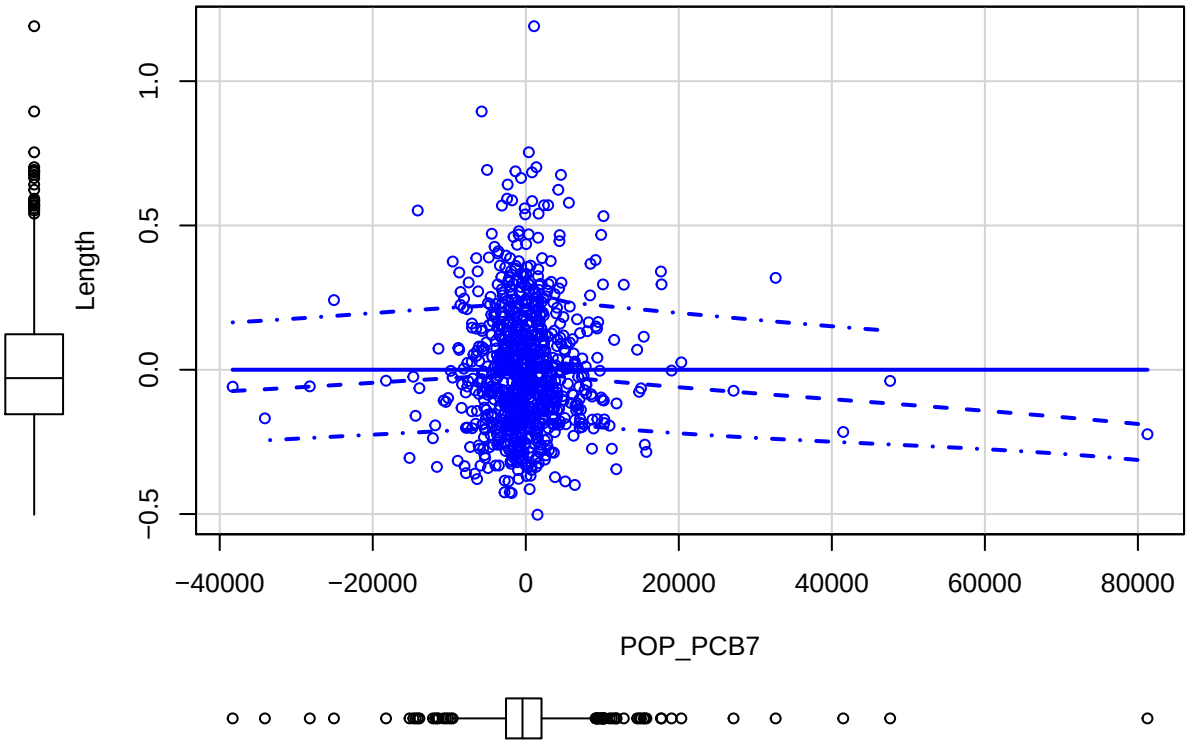

Length
1.0
0.5
0.0
−0.5
−20000
−10000
0
10000
20000
POP_PCB2

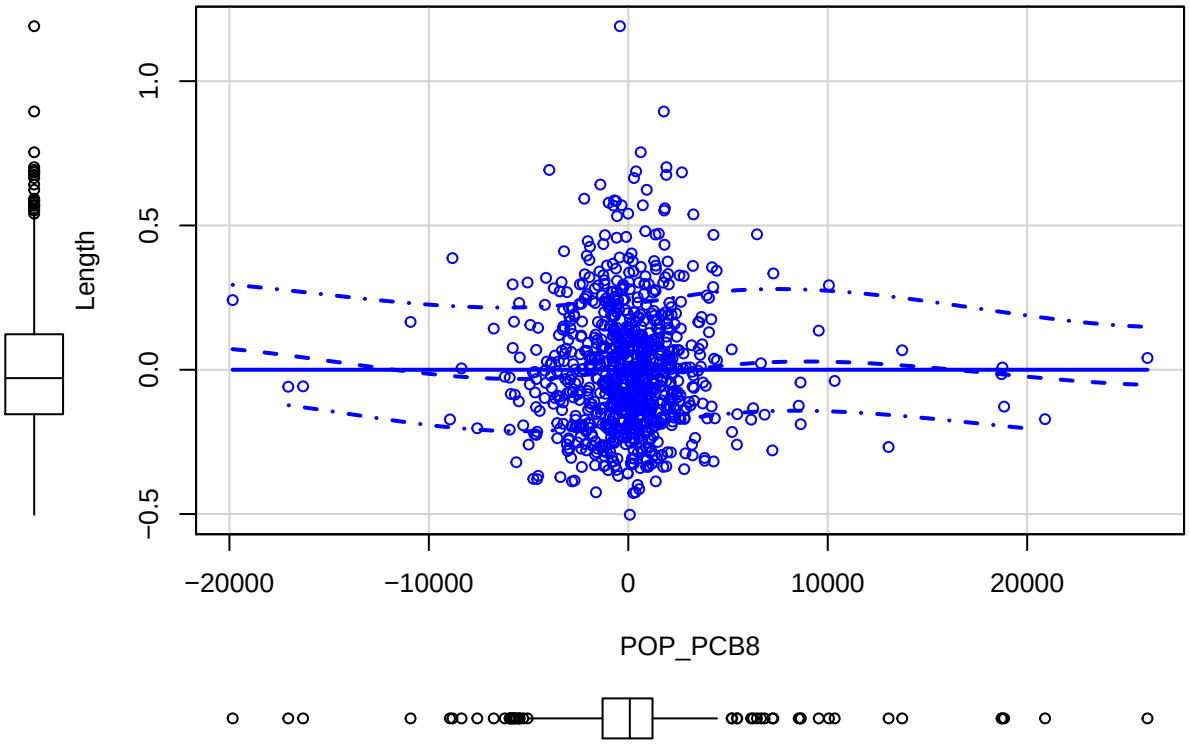

Length
1.0
0.5
0.0
−0.5
−20000
−10000
0
10000
20000
POP_PCB3

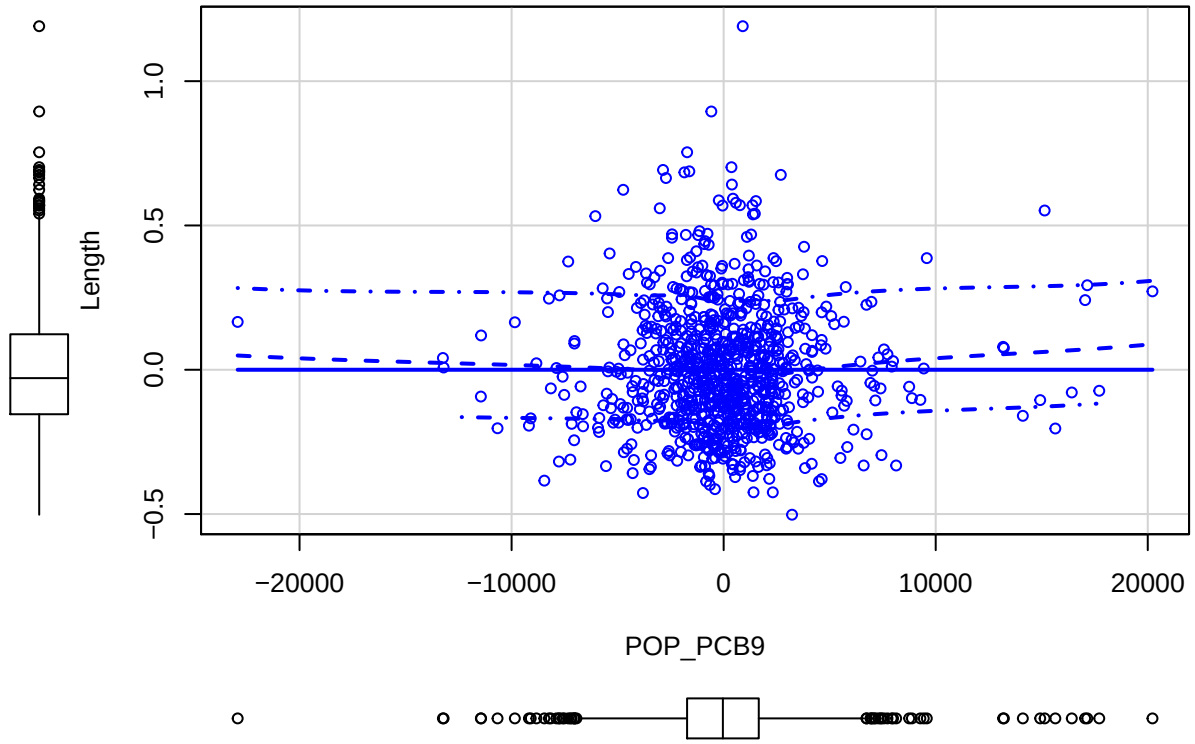

Length
1.0
0.5
0.0
−0.5
−40000
−20000
0
20000
40000
POP_PCB4

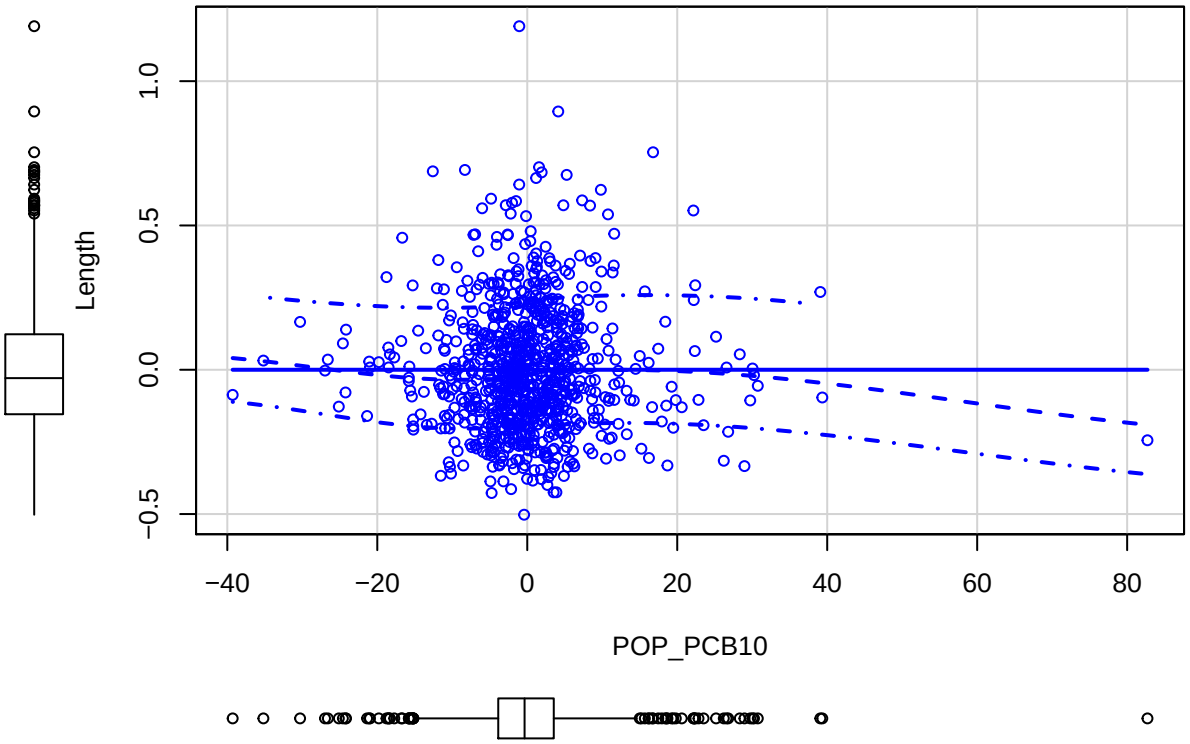

Length
1.0
0.5
0.0
−0.5
−40000
−20000
0
20000
40000
POP_PCB5

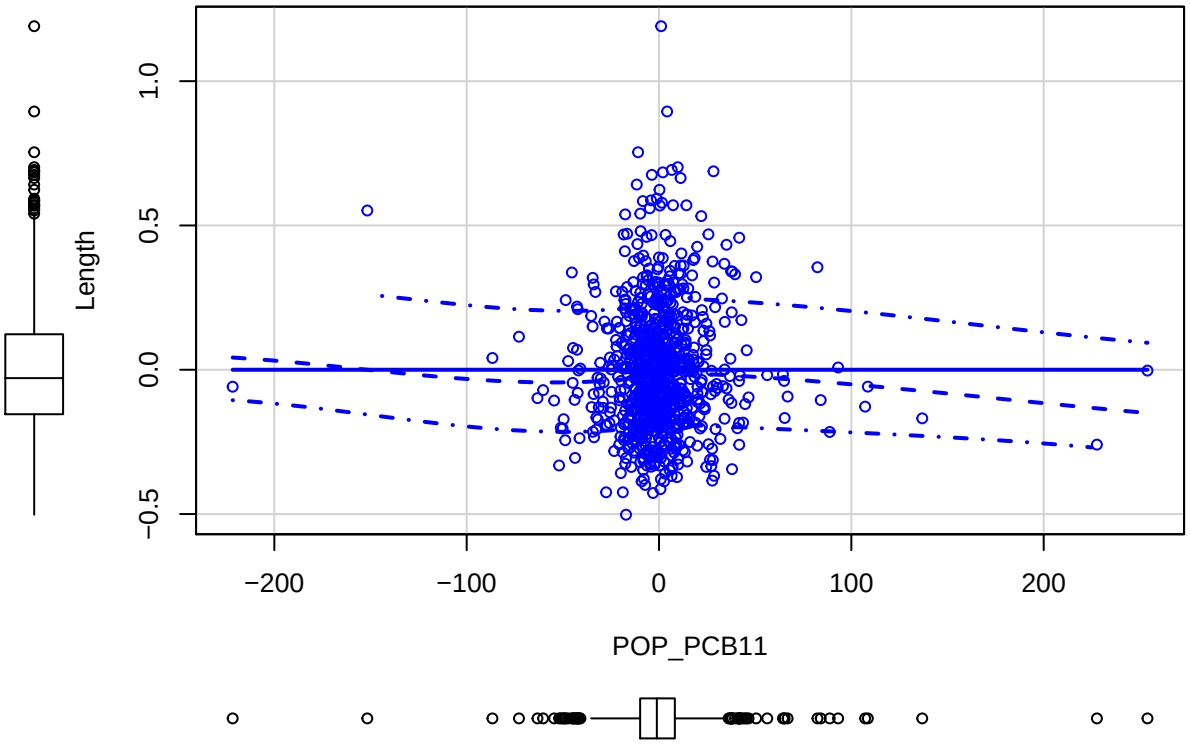

Length
1.0
0.5
0.0
−0.5
−40000
−20000
0
20000
40000
60000
80000
100000
POP_PCB6

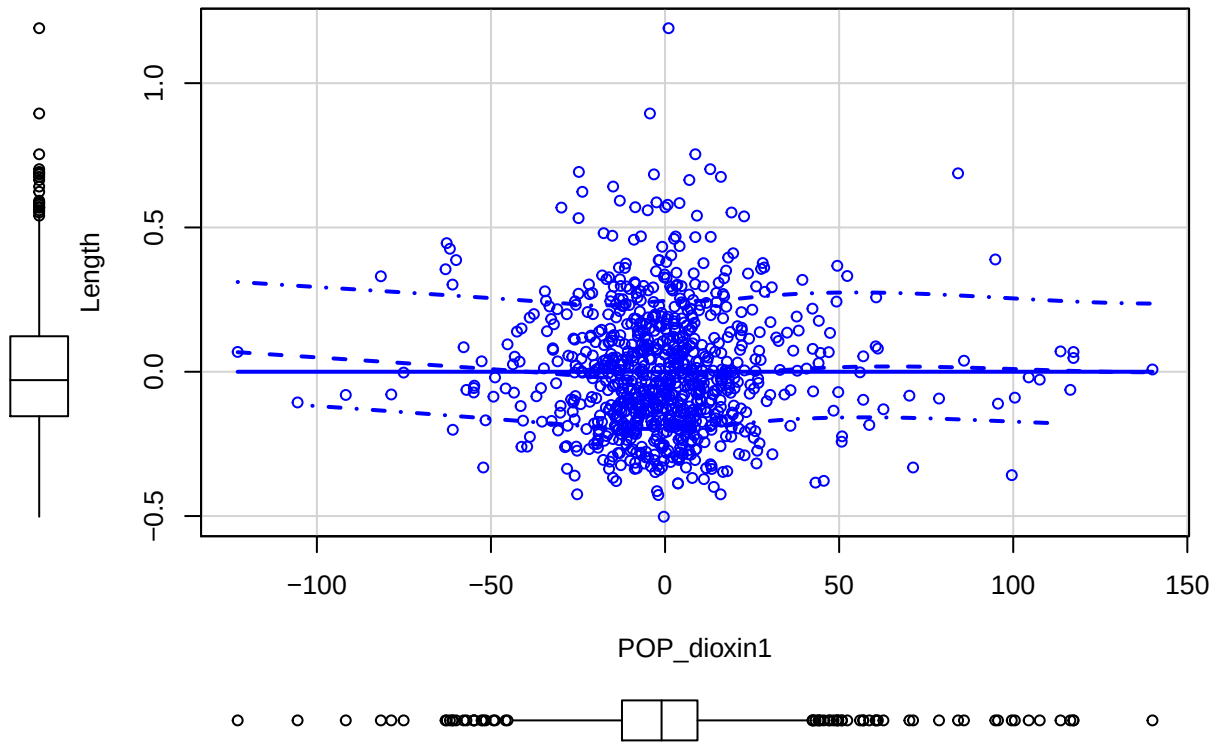

Length
1.0
0.5
0.0
−0.5
−40000
−20000
0
20000
40000
60000
80000
POP_PCB7

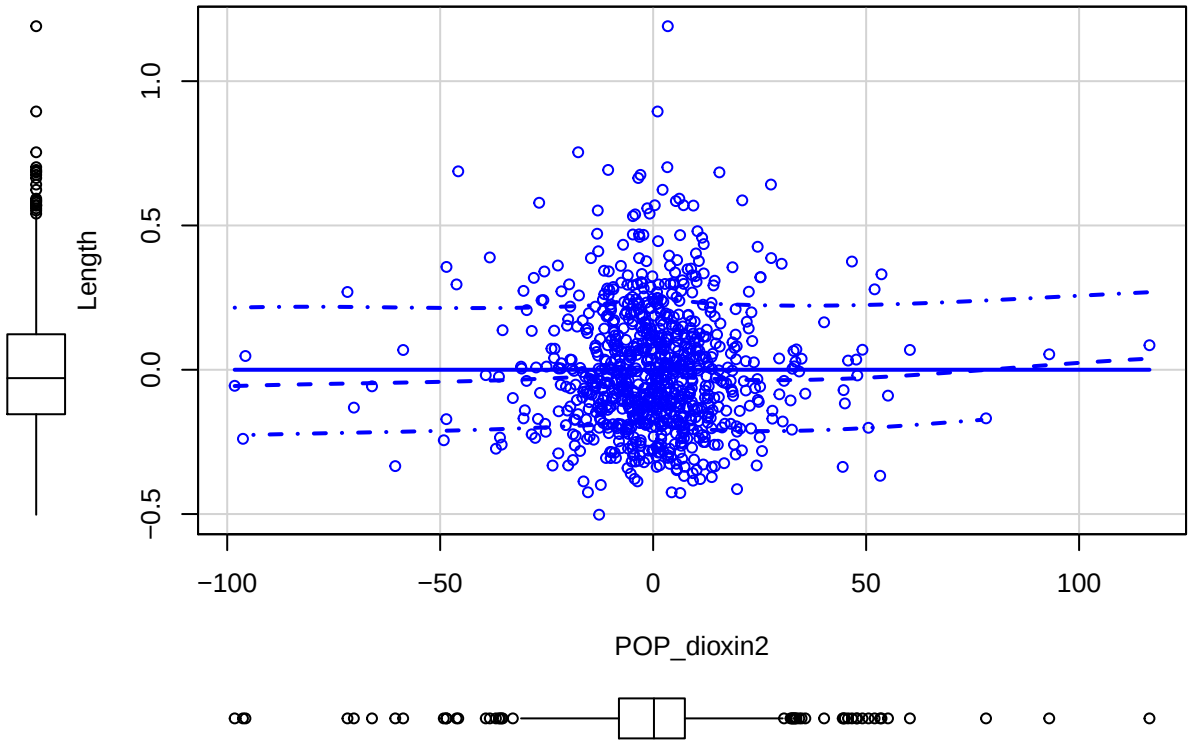

Length
1.0
0.5
0.0
−0.5
−20000
−10000
0
10000
20000
POP_PCB8

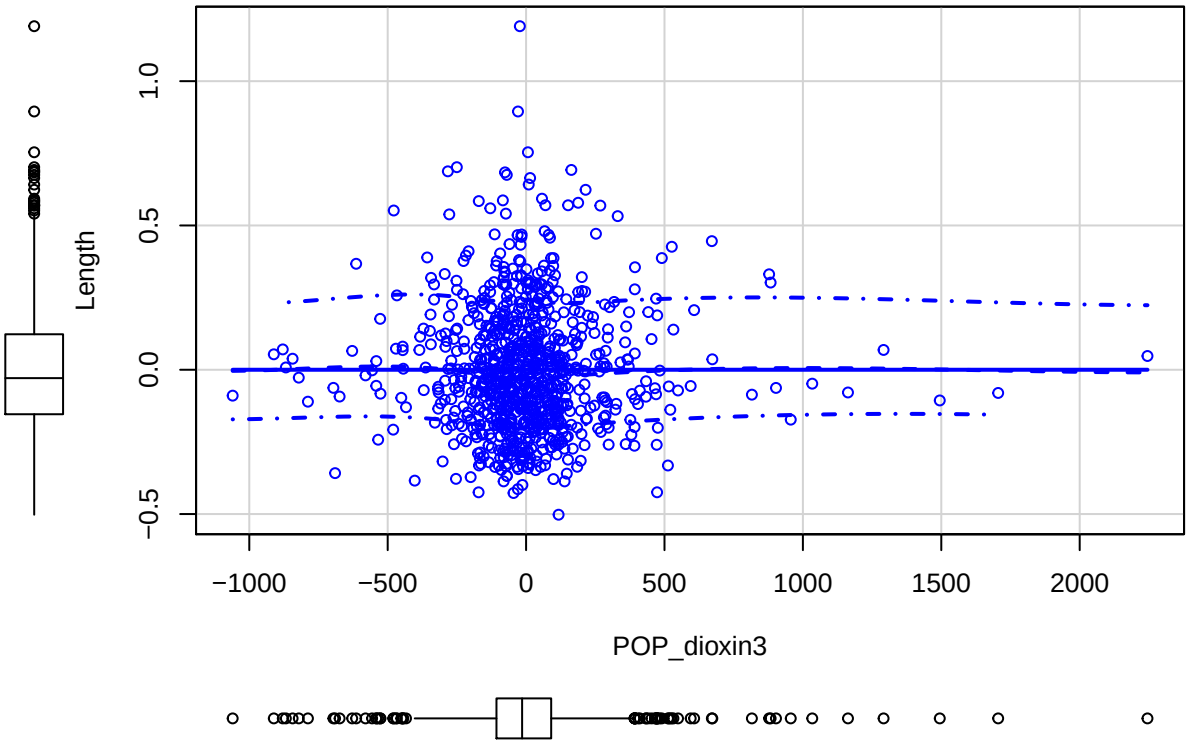

Length
1.0
0.5
0.0
−0.5
−20000
−10000
0
10000
20000
POP_PCB9

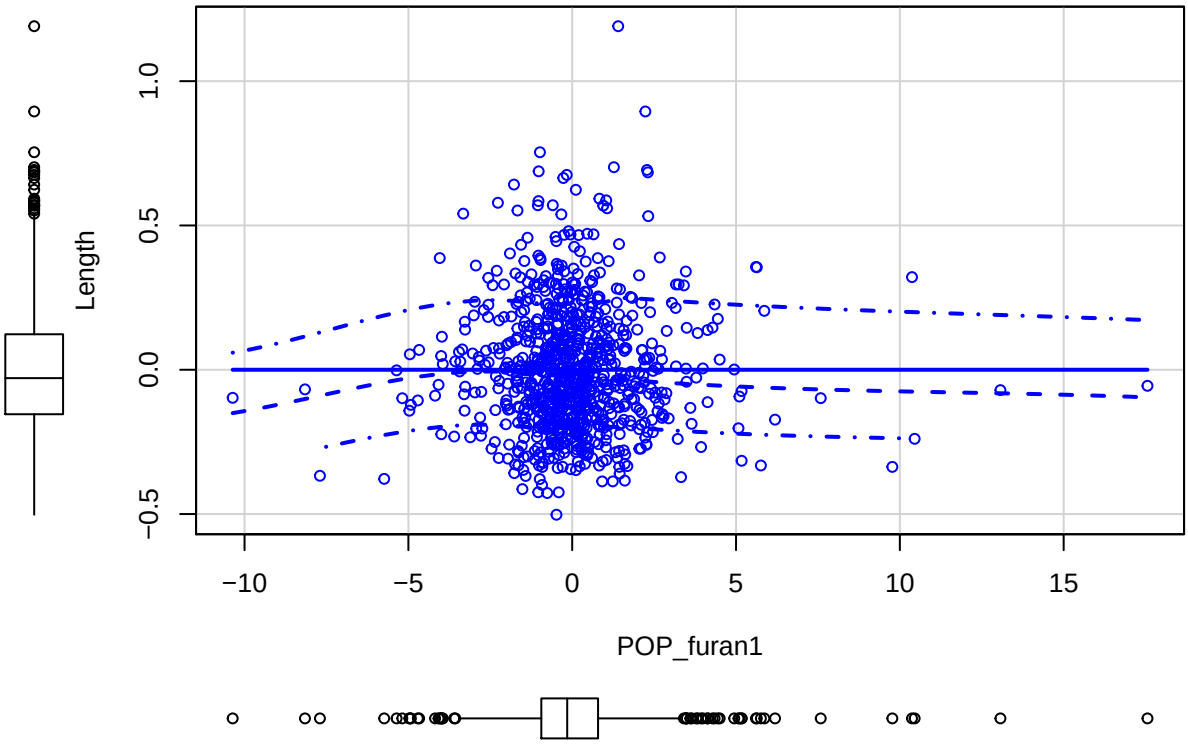

Length
1.0
0.5
0.0
−0.5
−40
−20
0
20
40
60
80
POP_PCB10

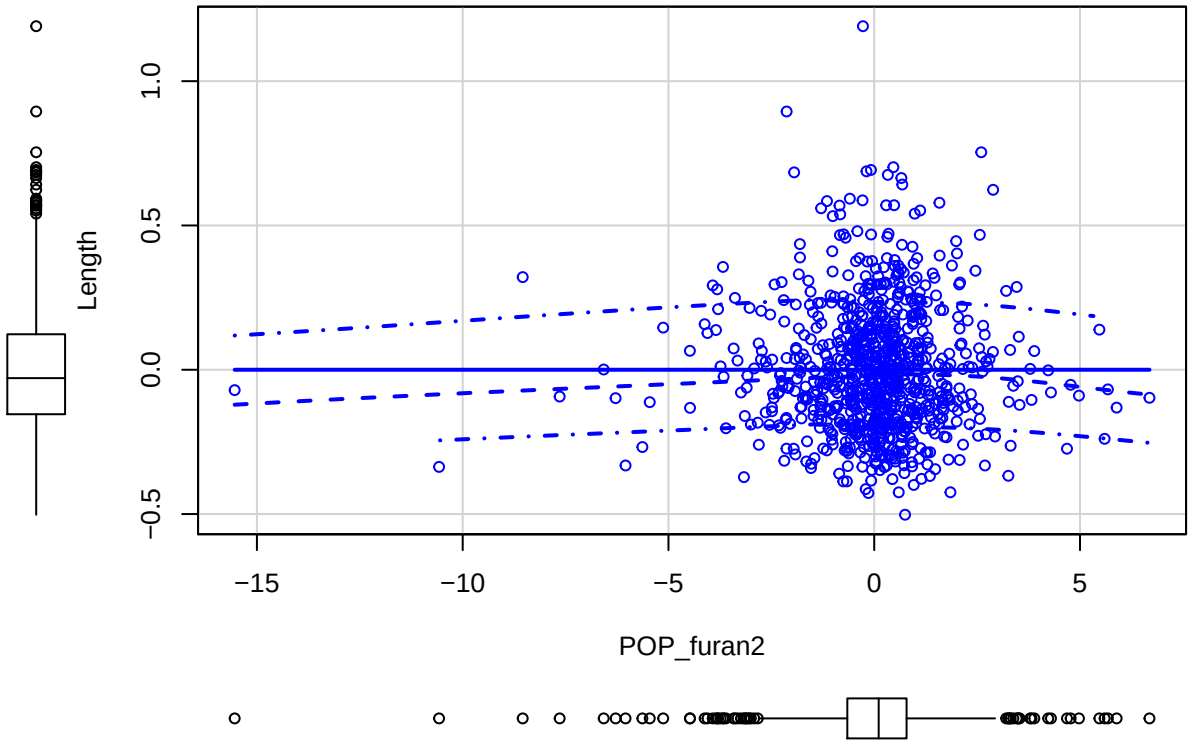

Length
1.0
0.5
0.0
−0.5
−200
−100
0
100
200
POP_PCB11

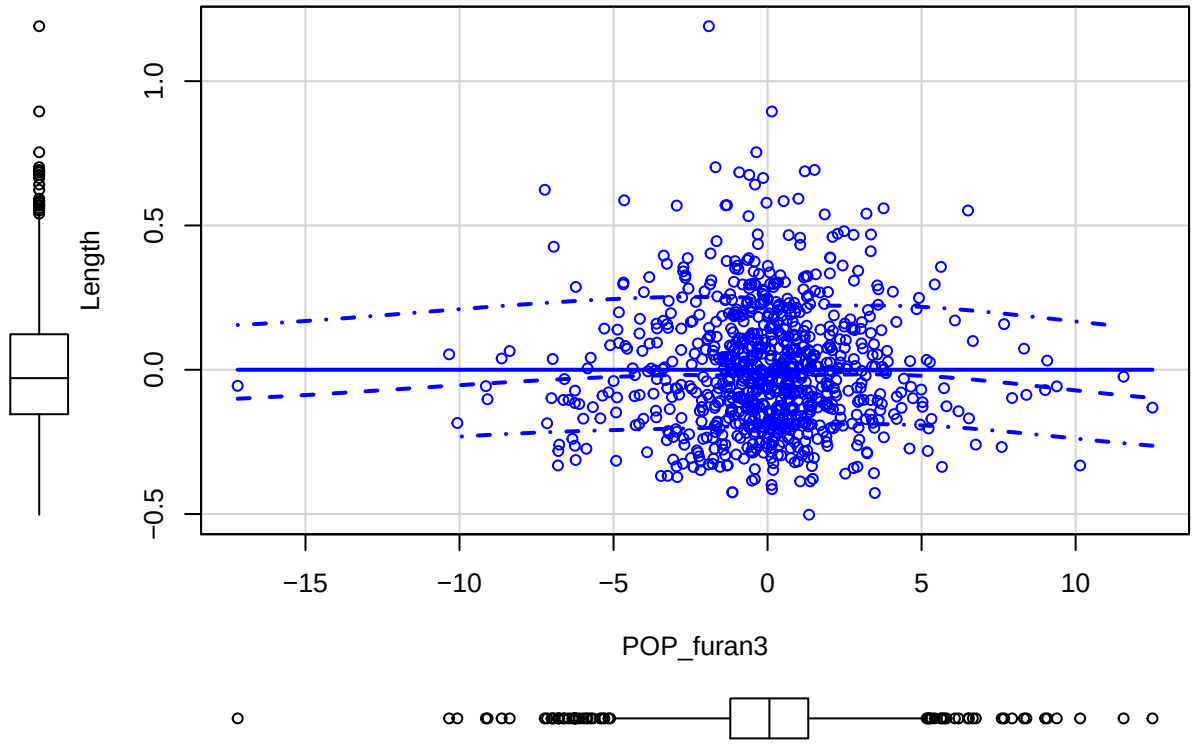

Length
1.0
0.5
0.0
−0.5
−100
−50
0
50
100
150
POP_dioxin1

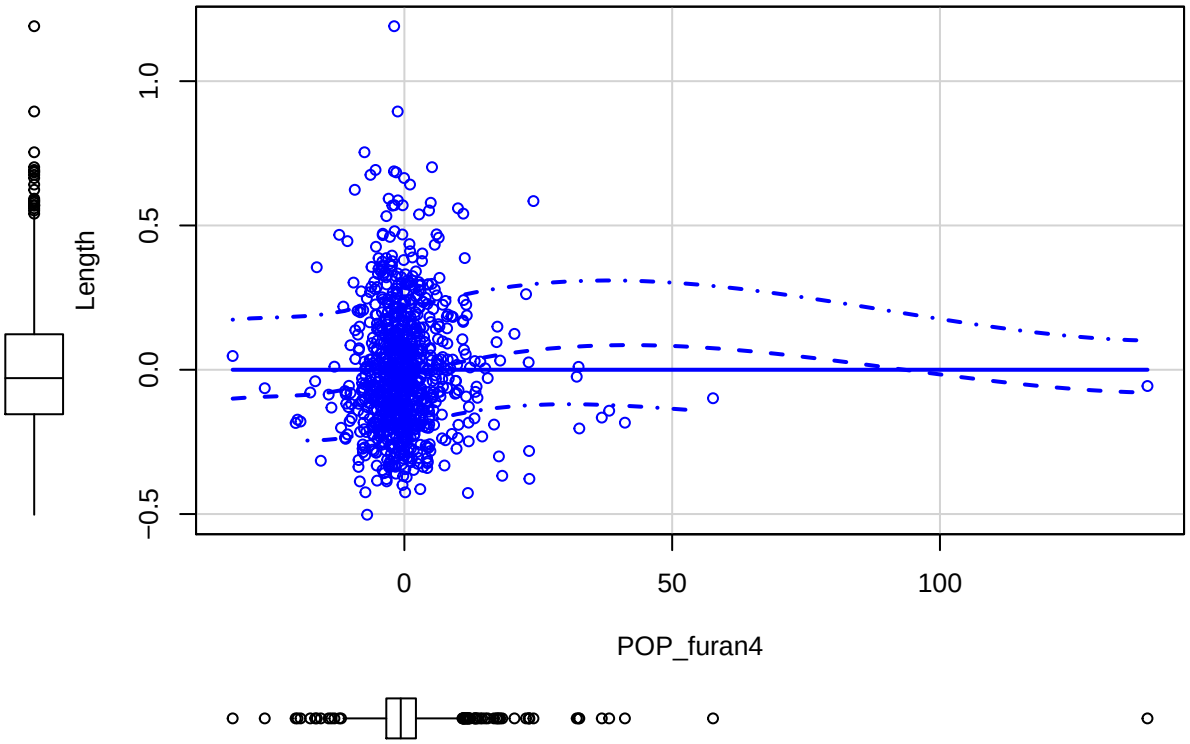

Length
1.0
0.5
0.0
−0.5
−100
−50
0
50
100
POP_dioxin2

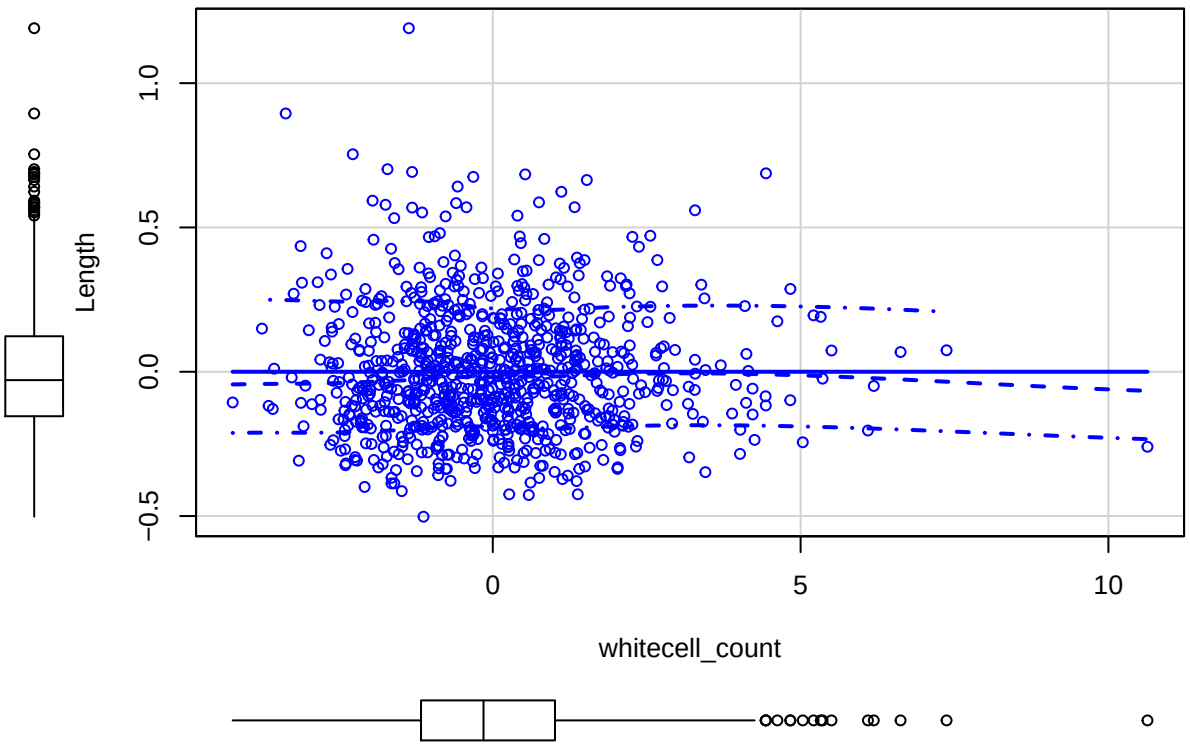

Length
1.0
0.5
0.0
−0.5
−1000
−500
0
500
1000
1500
2000
POP_dioxin3

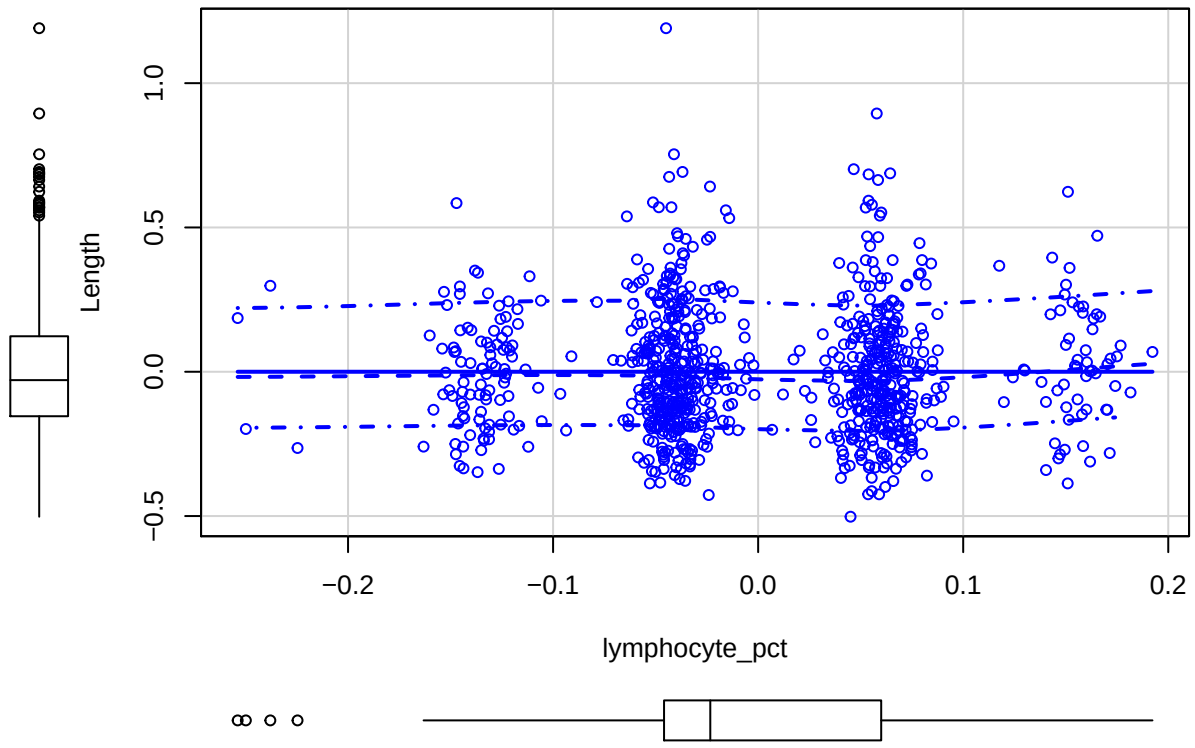

Length
1.0
0.5
0.0
−0.5
−10
−5
0
5
10
15
POP_furan1

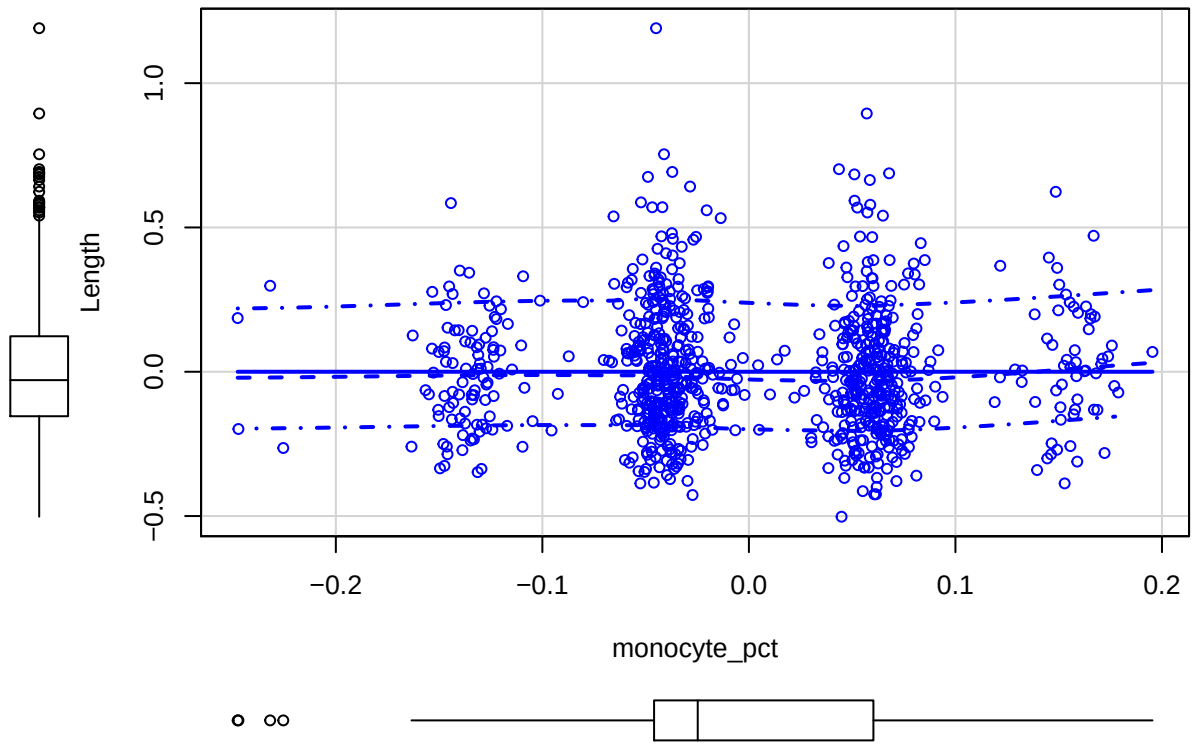

Length
1.0
0.5
0.0
−0.5
−15
−10
−5
0
5
POP_furan2

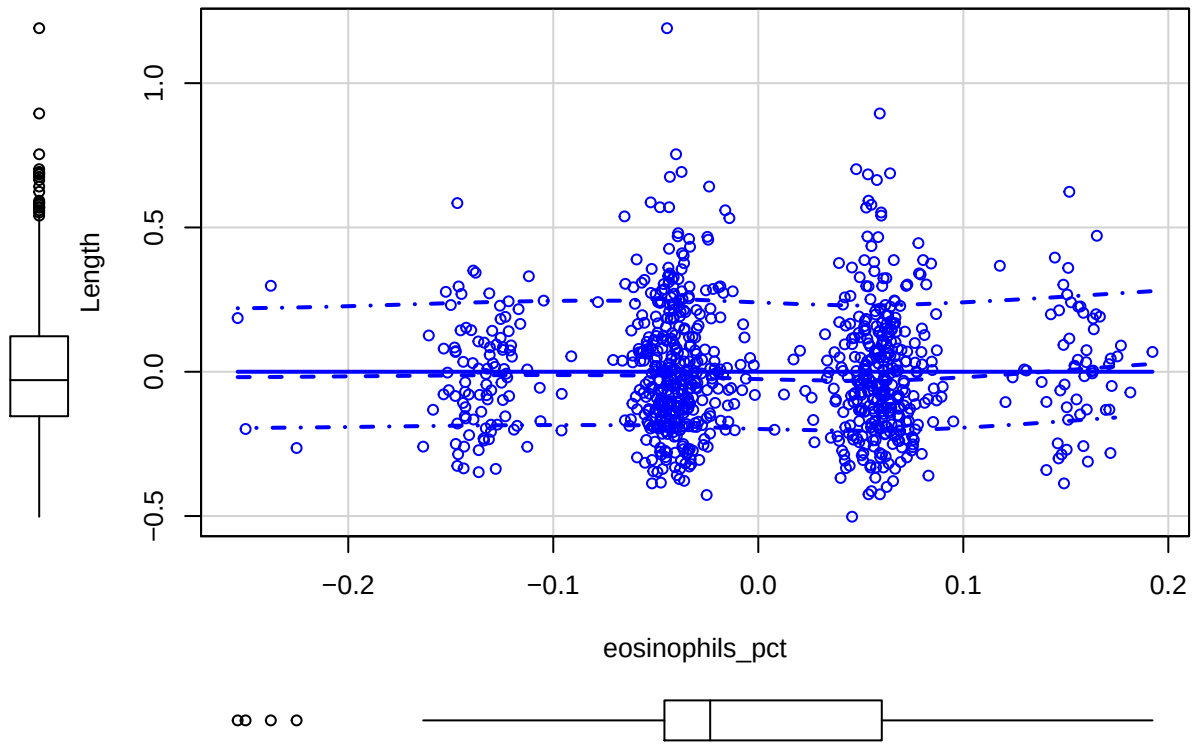

Length
1.0
0.5
0.0
−0.5
−15
−10
−5
0
5
10
POP_furan3

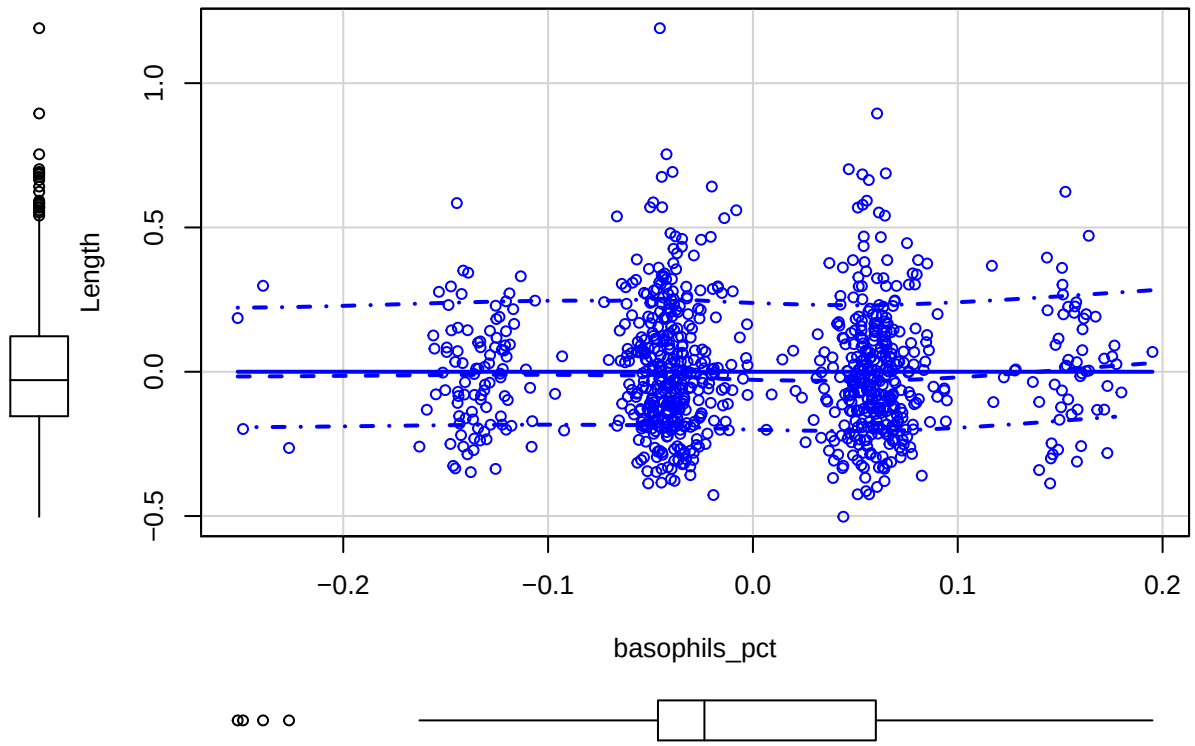

Length
1.0
0.5
0.0
−0.5
0
50
100
POP_furan4

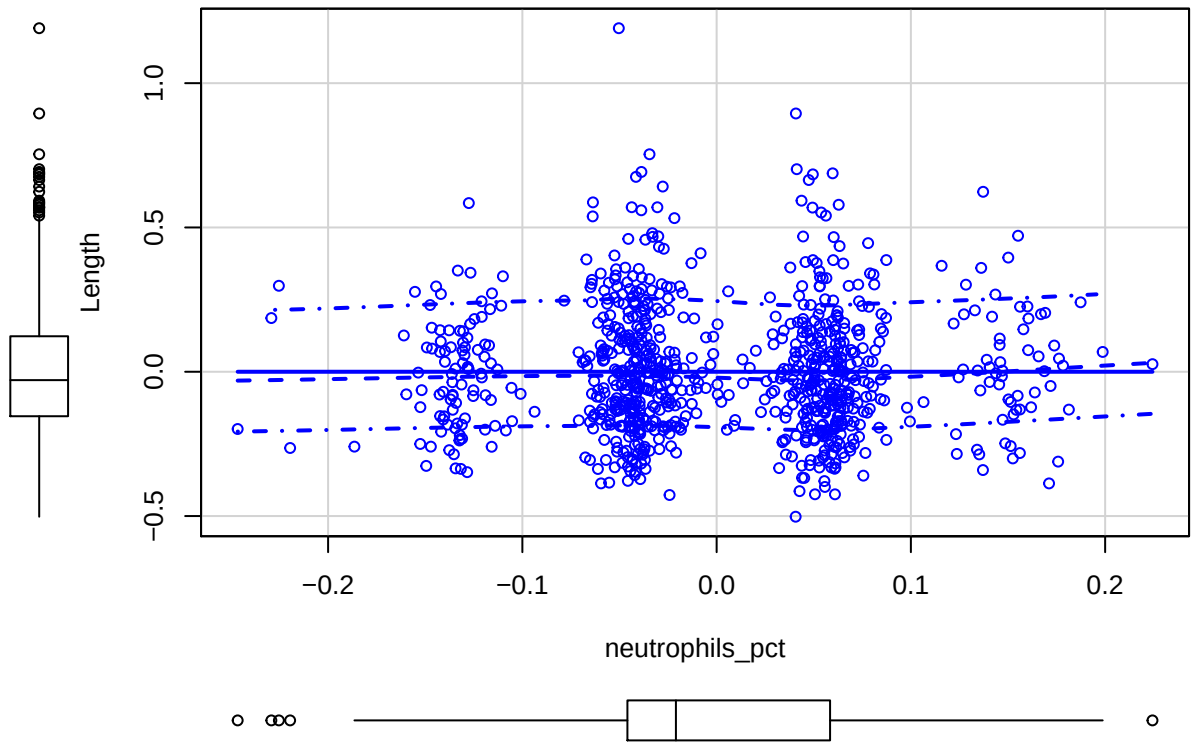

Length
1.0
0.5
0.0
−0.5
0
5
10
whitecell_count

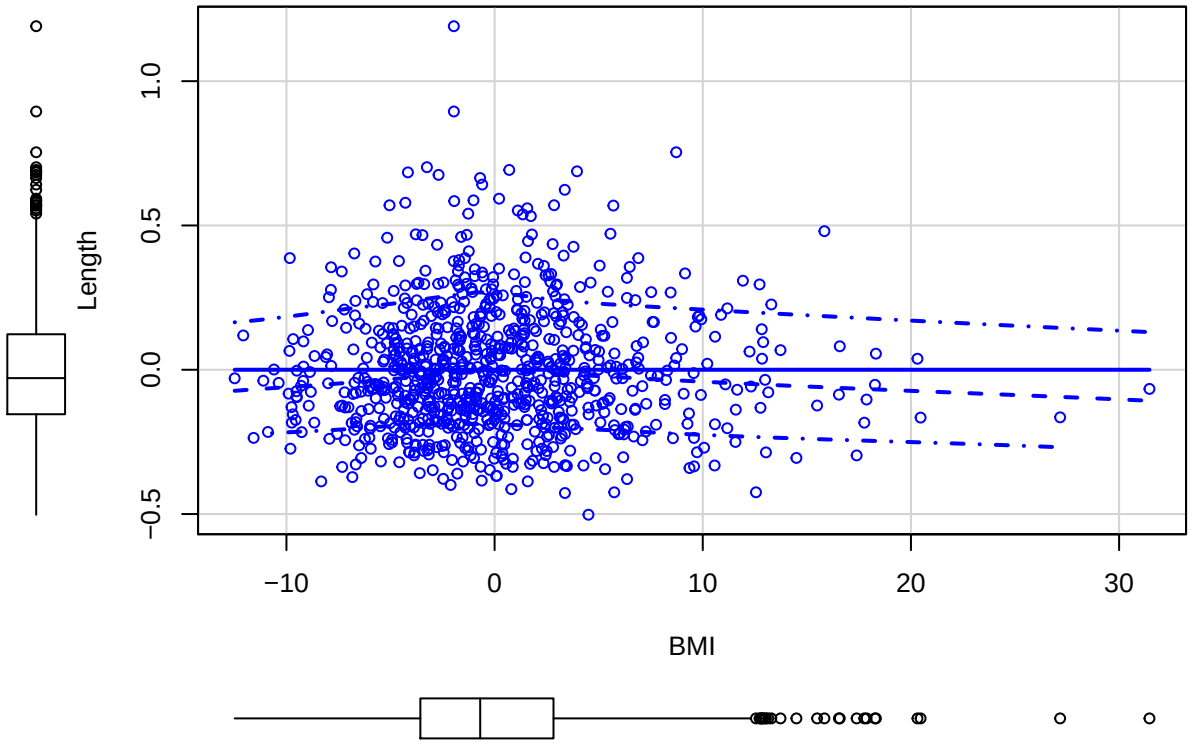

Length
1.0
0.5
0.0
−0.5
−0.2
−0.1
0.0
0.1
0.2
lymphocyte_pct

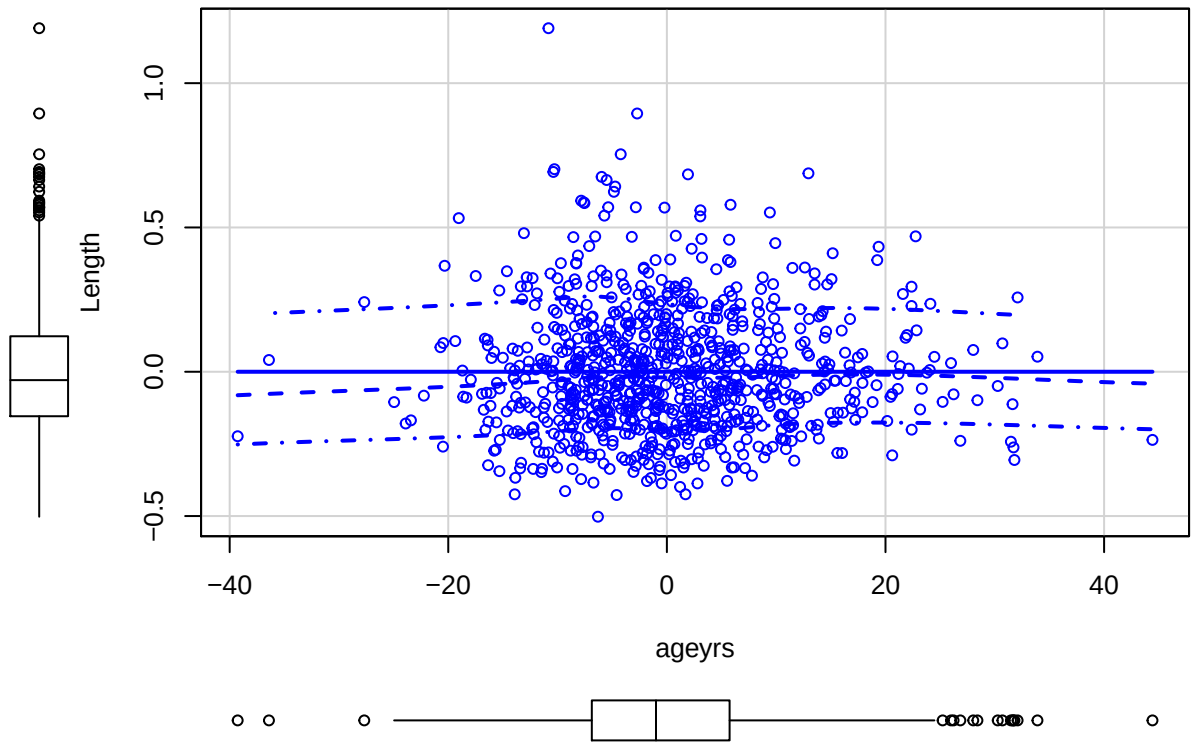

Length
1.0
0.5
0.0
−0.5
−0.2
−0.1
0.0
0.1
0.2
monocyte_pct

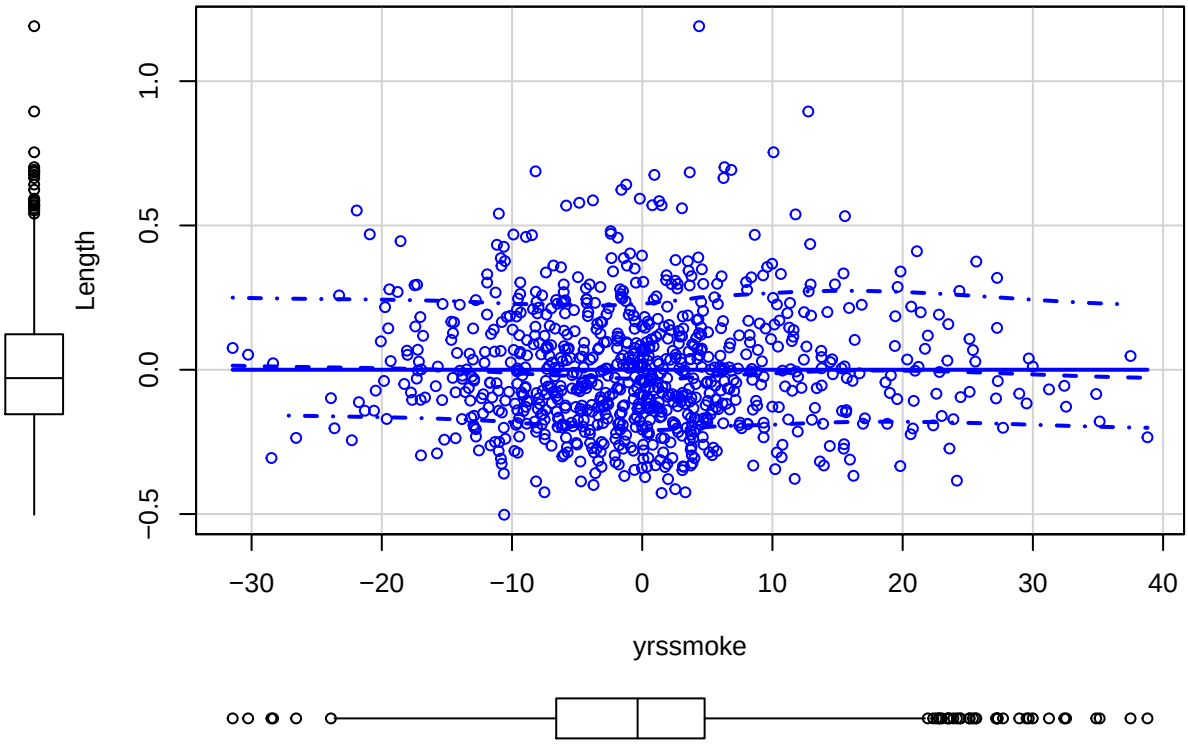

Length
1.0
0.5
0.0
−0.5
−0.2
−0.1
0.0
0.1
0.2
eosinophils_pct

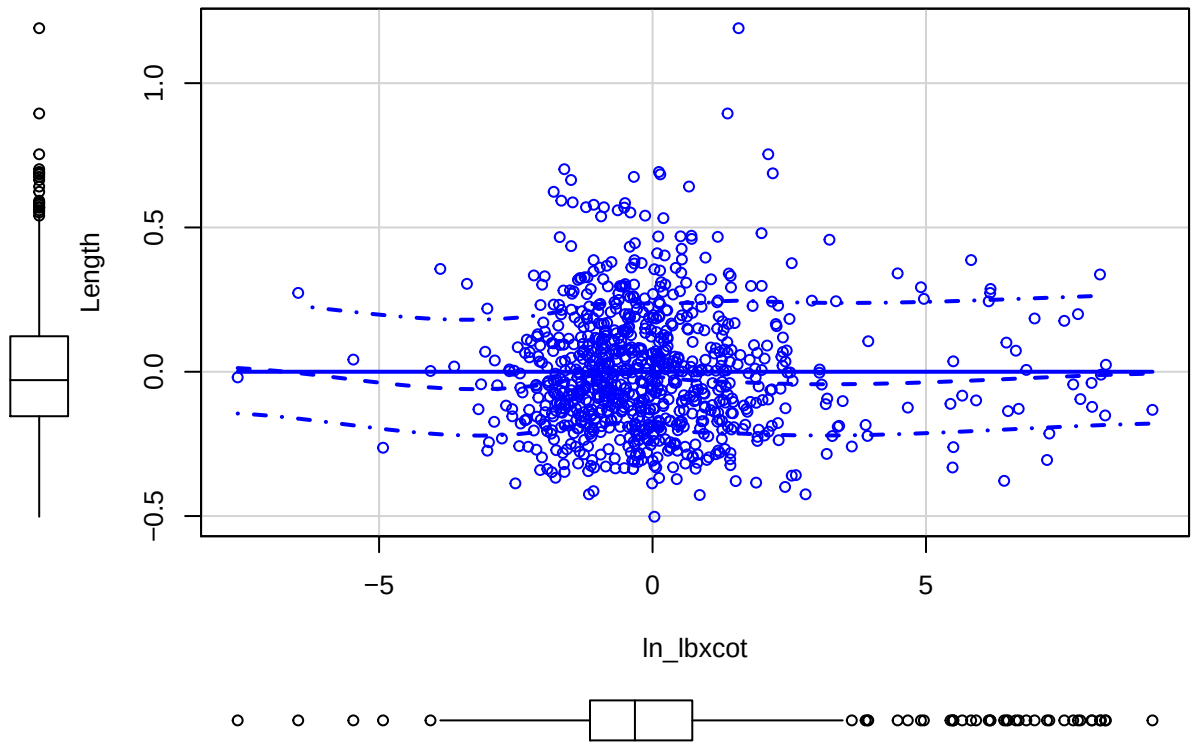

Length
1.0
0.5
0.0
−0.5
−0.2
−0.1
0.0
0.1
0.2
basophils_pct


Length
1.0
0.5
0.0
−0.5
−0.2
−0.1
0.0
0.1
0.2
neutrophils_pct

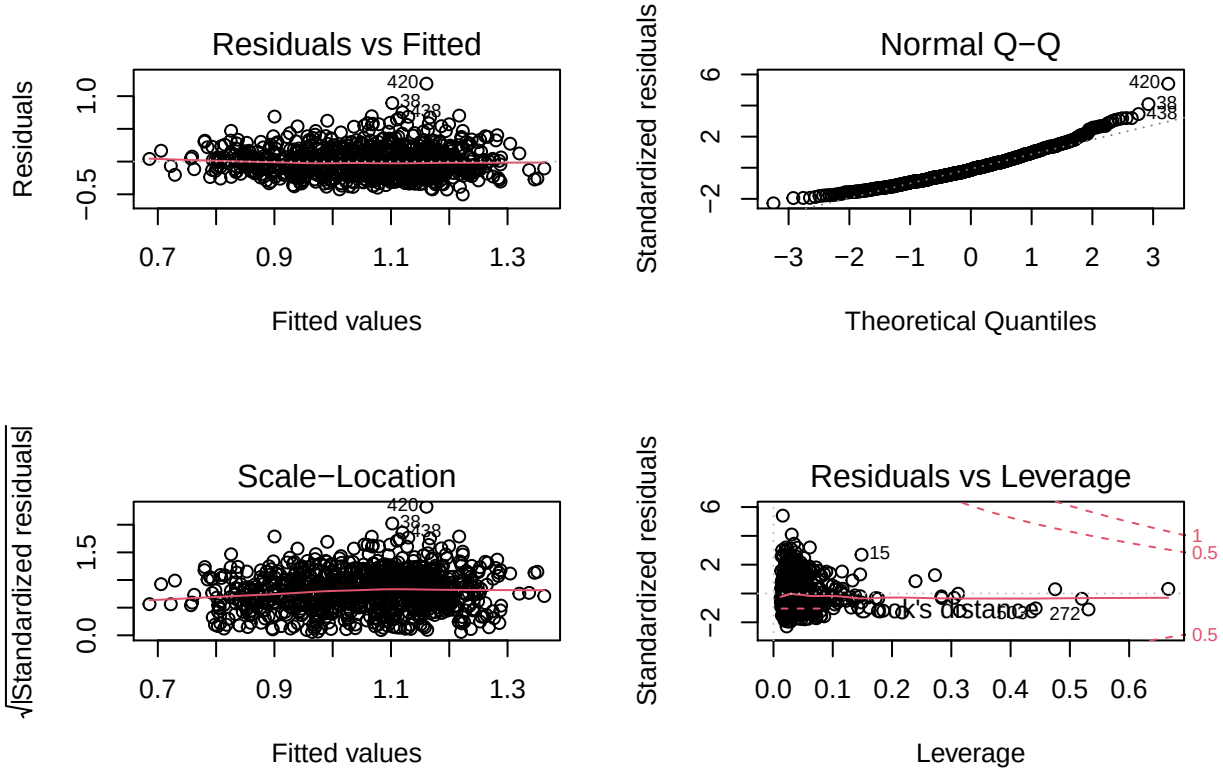

Length
1.0
0.5
0.0
−0.5
−10
0
10
20
30
BMI


Length
1.0
0.5
0.0
−0.5
−40
−20
0
20
40
ageyrs


Length
1.0
0.5
0.0
−0.5
−30
−20
−10
0
10
20
30
40
yrssmoke




```
#find the model fit for homoscedasticity before remove multicolinearity
par(mfrow=c(2,2))
plot(model)
```




```
par(mfrow=c(1,1))

#remove covariates with VIF > 10 below:

#remove highest vif ~ 15047
pollutants$eosinophils_pct = NULL
#fit new model
model = lm(length ~ .,data = pollutants)
#summary
summary(model)
```

```
##
## Call:
## lm(formula = length ~ ., data = pollutants)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.50160 -0.15463 -0.02843  0.12293  1.18974
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.491e+00  8.425e-02  17.693  < 2e-16 ***
## POP_PCB1        -1.606e-06  1.075e-06  -1.494   0.1355
## POP_PCB2         7.294e-07  3.021e-06   0.241   0.8093
## POP_PCB3         1.191e-06  2.156e-06   0.552   0.5809
## POP_PCB4        -1.774e-07  1.025e-06  -0.173   0.8627
```

```
## POP_PCB5         1.403e-07  1.068e-06   0.131   0.8955
## POP_PCB6         2.765e-07  1.058e-06   0.261   0.7939
## POP_PCB7        -5.739e-07  1.206e-06  -0.476   0.6343
## POP_PCB8         1.644e-06  2.446e-06   0.672   0.5016
## POP_PCB9         6.194e-07  2.111e-06   0.293   0.7693
## POP_PCB10        1.191e-03  8.890e-04   1.340   0.1806
## POP_PCB11        3.449e-05  3.078e-04   0.112   0.9108
## POP_dioxin1      3.072e-05  3.048e-04   0.101   0.9198
## POP_dioxin2     -1.744e-04  4.394e-04  -0.397   0.6916
## POP_dioxin3     -1.908e-05  3.018e-05  -0.632   0.5274
## POP_furan1       2.522e-03  3.844e-03   0.656   0.5119
## POP_furan2      -2.799e-04  4.500e-03  -0.062   0.9504
## POP_furan3       4.473e-03  2.756e-03   1.623   0.1049
## POP_furan4      -6.496e-04  9.195e-04  -0.706   0.4801
## whitecell_count -5.261e-03  4.404e-03  -1.195   0.2326
## lymphocyte_pct  -1.255e-03  1.030e-03  -1.219   0.2234
## monocyte_pct    -5.993e-03  4.041e-03  -1.483   0.1384
## basophils_pct    1.051e-03  3.475e-03   0.302   0.7624
## neutrophils_pct  1.213e-02  1.662e-02   0.730   0.4657
## BMI             -1.359e-03  1.409e-03  -0.964   0.3352
## edu_cat2         2.452e-02  2.215e-02   1.107   0.2687
## edu_cat3         4.787e-02  2.164e-02   2.212   0.0272 *
## edu_cat4         3.265e-02  2.555e-02   1.278   0.2016
## race_cat2       -2.195e-02  3.265e-02  -0.672   0.5016
## race_cat3        2.475e-02  3.370e-02   0.735   0.4628
## race_cat4       -3.473e-02  2.991e-02  -1.161   0.2460
## male1           -3.944e-02  1.771e-02  -2.227   0.0262 *
## ageyrs          -6.236e-03  7.442e-04  -8.380 2.25e-16 ***
## yrssmoke        -5.323e-04  7.271e-04  -0.732   0.4643
## smokenow1        2.172e-03  3.581e-02   0.061   0.9516
## ln_lbxcot        5.341e-03  3.922e-03   1.362   0.1736
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.222 on 828 degrees of freedom
## Multiple R-squared:  0.2448, Adjusted R-squared:  0.2129
## F-statistic: 7.668 on 35 and 828 DF,  p-value: < 2.2e-16
```

```
#show the VIF
vif(model)
```

```
##                     GVIF Df GVIF^(1/(2*Df))
## POP_PCB1       33.040956  1        5.748126
## POP_PCB2       34.276806  1        5.854640
## POP_PCB3        9.350969  1        3.057935
## POP_PCB4       31.734341  1        5.633324
## POP_PCB5       59.718357  1        7.727765
## POP_PCB6       11.386137  1        3.374335
## POP_PCB7        4.868942  1        2.206568
## POP_PCB8       12.982530  1        3.603128
## POP_PCB9       12.416573  1        3.523716
## POP_PCB10       5.988829  1        2.447208
## POP_PCB11       4.725385  1        2.173795
## POP_dioxin1     5.256334  1        2.292670
## POP_dioxin2     5.411476  1        2.326258
```

```
## POP_dioxin3      4.378774  1        2.092552
## POP_furan1       6.154213  1        2.480769
## POP_furan2       6.193725  1        2.488720
## POP_furan3       4.450557  1        2.109634
## POP_furan4       1.821773  1        1.349731
## whitecell_count  1.545798  1        1.243301
## lymphocyte_pct   1.382541  1        1.175815
## monocyte_pct     1.263611  1        1.124105
## basophils_pct    1.113199  1        1.055082
## neutrophils_pct  1.090031  1        1.044046
## BMI              1.261934  1        1.123358
## edu_cat          1.541083  3        1.074742
## race_cat         2.051619  3        1.127239
## male             1.350208  1        1.161984
## ageyrs           3.237762  1        1.799378
## yrssmoke         2.204134  1        1.484633
## smokenow         3.998531  1        1.999633
## ln_lbxcot        3.954234  1        1.988526
```

```
#remove highest vif eosinophils_pct ~ 15047 then POP_PCB5 ~ 59.718357
pollutants$POP_PCB5 = NULL
#fit new model
model = lm(length ~ .,data = pollutants)
#summary
summary(model)
```

```
##
## Call:
## lm(formula = length ~ ., data = pollutants)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.50233 -0.15456 -0.02848  0.12242  1.19007
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.491e+00  8.419e-02  17.706   <2e-16 ***
## POP_PCB1        -1.584e-06  1.061e-06  -1.493   0.1359
## POP_PCB2         8.659e-07  2.835e-06   0.305   0.7601
## POP_PCB3         1.229e-06  2.135e-06   0.575   0.5652
## POP_PCB4        -9.537e-08  8.129e-07  -0.117   0.9066
## POP_PCB6         3.209e-07  1.002e-06   0.320   0.7489
## POP_PCB7        -5.874e-07  1.201e-06  -0.489   0.6248
## POP_PCB8         1.574e-06  2.386e-06   0.660   0.5095
## POP_PCB9         7.223e-07  1.960e-06   0.369   0.7125
## POP_PCB10        1.211e-03  8.759e-04   1.382   0.1672
## POP_PCB11        2.293e-05  2.947e-04   0.078   0.9380
## POP_dioxin1      2.999e-05  3.046e-04   0.098   0.9216
## POP_dioxin2     -1.693e-04  4.374e-04  -0.387   0.6989
## POP_dioxin3     -1.949e-05  3.001e-05  -0.650   0.5161
## POP_furan1       2.449e-03  3.801e-03   0.644   0.5196
## POP_furan2      -2.463e-04  4.490e-03  -0.055   0.9563
## POP_furan3       4.477e-03  2.754e-03   1.626   0.1044
## POP_furan4      -6.429e-04  9.176e-04  -0.701   0.4837
## whitecell_count -5.303e-03  4.390e-03  -1.208   0.2274
```

```
## lymphocyte_pct  -1.254e-03  1.029e-03  -1.219   0.2233
## monocyte_pct    -5.978e-03  4.037e-03  -1.481   0.1390
## basophils_pct    1.012e-03  3.460e-03   0.292   0.7700
## neutrophils_pct  1.213e-02  1.661e-02   0.731   0.4653
## BMI             -1.357e-03  1.408e-03  -0.964   0.3356
## edu_cat2         2.450e-02  2.214e-02   1.107   0.2688
## edu_cat3         4.786e-02  2.163e-02   2.213   0.0272 *
## edu_cat4         3.284e-02  2.550e-02   1.288   0.1981
## race_cat2       -2.186e-02  3.263e-02  -0.670   0.5030
## race_cat3        2.503e-02  3.361e-02   0.745   0.4566
## race_cat4       -3.460e-02  2.988e-02  -1.158   0.2472
## male1           -3.945e-02  1.770e-02  -2.228   0.0261 *
## ageyrs          -6.237e-03  7.437e-04  -8.387   <2e-16 ***
## yrssmoke        -5.391e-04  7.248e-04  -0.744   0.4572
## smokenow1        2.165e-03  3.579e-02   0.061   0.9518
## ln_lbxcot        5.346e-03  3.919e-03   1.364   0.1729
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2219 on 829 degrees of freedom
## Multiple R-squared:  0.2448, Adjusted R-squared:  0.2138
## F-statistic: 7.902 on 34 and 829 DF,  p-value: < 2.2e-16
```

```
#show the VIF
vif(model)
```

```
##                      GVIF Df GVIF^(1/(2*Df))
## POP_PCB1        32.255513  1        5.679394
## POP_PCB2        30.220860  1        5.497350
## POP_PCB3         9.182804  1        3.030314
## POP_PCB4        19.965194  1        4.468243
## POP_PCB6        10.226099  1        3.197827
## POP_PCB7         4.833396  1        2.198498
## POP_PCB8        12.363256  1        3.516142
## POP_PCB9        10.707771  1        3.272273
## POP_PCB10        5.820661  1        2.412605
## POP_PCB11        4.338648  1        2.082942
## POP_dioxin1      5.254606  1        2.292293
## POP_dioxin2      5.368810  1        2.317069
## POP_dioxin3      4.332013  1        2.081349
## POP_furan1       6.023932  1        2.454370
## POP_furan2       6.173626  1        2.484678
## POP_furan3       4.450098  1        2.109526
## POP_furan4       1.816182  1        1.347658
## whitecell_count  1.537750  1        1.240060
## lymphocyte_pct   1.382511  1        1.175802
## monocyte_pct     1.262603  1        1.123656
## basophils_pct    1.105130  1        1.051251
## neutrophils_pct  1.090023  1        1.044042
## BMI              1.261809  1        1.123303
## edu_cat          1.533185  3        1.073822
## race_cat         2.042113  3        1.126367
## male             1.350206  1        1.161984
## ageyrs           3.237570  1        1.799325
## yrssmoke         2.192917  1        1.480850
```

```
## smokenow         3.998522  1        1.999630
## ln_lbxcot        3.953781  1        1.988412
```

```
#remove highest vif eosinophils_pct ~ 15047 then POP_PCB5 ~ 59.718357 then POP_PCB1 ~ 32.255513
pollutants$POP_PCB1 = NULL
#fit new model
model = lm(length ~ .,data = pollutants)
#summary
summary(model)
```

```
##
## Call:
## lm(formula = length ~ ., data = pollutants)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.49329 -0.15025 -0.02981  0.12229  1.18766
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.505e+00  8.371e-02  17.977   <2e-16 ***
## POP_PCB2        -1.264e-06  2.451e-06  -0.516   0.6062
## POP_PCB3         1.587e-06  2.123e-06   0.747   0.4551
## POP_PCB4        -2.160e-07  8.094e-07  -0.267   0.7896
## POP_PCB6         1.820e-07  9.987e-07   0.182   0.8555
## POP_PCB7        -6.786e-07  1.200e-06  -0.565   0.5719
## POP_PCB8        -1.708e-07  2.081e-06  -0.082   0.9346
## POP_PCB9        -8.899e-09  1.899e-06  -0.005   0.9963
## POP_PCB10        1.035e-03  8.686e-04   1.192   0.2337
## POP_PCB11        9.381e-05  2.911e-04   0.322   0.7473
## POP_dioxin1      3.693e-05  3.048e-04   0.121   0.9036
## POP_dioxin2     -1.018e-04  4.354e-04  -0.234   0.8152
## POP_dioxin3     -1.942e-05  3.003e-05  -0.647   0.5179
## POP_furan1       2.488e-03  3.804e-03   0.654   0.5132
## POP_furan2      -5.409e-04  4.489e-03  -0.120   0.9041
## POP_furan3       4.497e-03  2.756e-03   1.632   0.1032
## POP_furan4      -6.516e-04  9.182e-04  -0.710   0.4781
## whitecell_count -5.516e-03  4.391e-03  -1.256   0.2094
## lymphocyte_pct  -1.382e-03  1.026e-03  -1.347   0.1784
## monocyte_pct    -5.839e-03  4.039e-03  -1.446   0.1486
## basophils_pct    1.159e-03  3.462e-03   0.335   0.7377
## neutrophils_pct  1.127e-02  1.661e-02   0.678   0.4978
## BMI             -1.358e-03  1.409e-03  -0.964   0.3354
## edu_cat2         2.226e-02  2.211e-02   1.007   0.3143
## edu_cat3         4.353e-02  2.145e-02   2.030   0.0427 *
## edu_cat4         2.960e-02  2.542e-02   1.164   0.2447
## race_cat2       -2.335e-02  3.263e-02  -0.716   0.4744
## race_cat3        2.489e-02  3.363e-02   0.740   0.4596
## race_cat4       -3.529e-02  2.990e-02  -1.181   0.2381
## male1           -3.992e-02  1.771e-02  -2.254   0.0245 *
## ageyrs          -6.298e-03  7.431e-04  -8.476   <2e-16 ***
## yrssmoke        -4.629e-04  7.236e-04  -0.640   0.5225
## smokenow1        7.390e-04  3.580e-02   0.021   0.9835
## ln_lbxcot        5.387e-03  3.922e-03   1.374   0.1699
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.222 on 830 degrees of freedom
## Multiple R-squared:  0.2427, Adjusted R-squared:  0.2126
## F-statistic: 8.062 on 33 and 830 DF,  p-value: < 2.2e-16
```

```
#show the VIF
vif(model)
```

```
##                     GVIF Df GVIF^(1/(2*Df))
## POP_PCB2       22.564616  1        4.750223
## POP_PCB3        9.066898  1        3.011129
## POP_PCB4       19.767850  1        4.446105
## POP_PCB6       10.137933  1        3.184012
## POP_PCB7        4.820891  1        2.195653
## POP_PCB8        9.395135  1        3.065148
## POP_PCB9       10.038785  1        3.168404
## POP_PCB10       5.715628  1        2.390738
## POP_PCB11       4.226036  1        2.055733
## POP_dioxin1     5.253381  1        2.292026
## POP_dioxin2     5.311508  1        2.304671
## POP_dioxin3     4.332003  1        2.081346
## POP_furan1      6.023639  1        2.454310
## POP_furan2      6.161696  1        2.482276
## POP_furan3      4.449999  1        2.109502
## POP_furan4      1.816107  1        1.347630
## whitecell_count 1.536123  1        1.239404
## lymphocyte_pct  1.372891  1        1.171704
## monocyte_pct    1.261930  1        1.123356
## basophils_pct   1.104229  1        1.050823
## neutrophils_pct 1.088688  1        1.043402
## BMI             1.261809  1        1.123303
## edu_cat         1.504890  3        1.070494
## race_cat        2.039232  3        1.126102
## male            1.349779  1        1.161800
## ageyrs          3.227727  1        1.796588
## yrssmoke        2.182035  1        1.477171
## smokenow        3.995671  1        1.998918
## ln_lbxcot       3.953587  1        1.988363
```

```
#remove highest vif eosinophils_pct ~ 15047 then POP_PCB5 ~ 59.718357 then POP_PCB1 ~ 32.255513 then PO
pollutants$POP_PCB2 = NULL
#fit new model
model = lm(length ~ .,data = pollutants)
#summary
summary(model)
```

```
##
## Call:
## lm(formula = length ~ ., data = pollutants)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.49387 -0.15104 -0.02904  0.12181  1.18897
##
## Coefficients:
```

```
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.502e+00  8.354e-02  17.985   <2e-16 ***
## POP_PCB3         1.846e-06  2.062e-06   0.895   0.3709
## POP_PCB4        -4.363e-07  6.872e-07  -0.635   0.5256
## POP_PCB6         3.180e-07  9.628e-07   0.330   0.7413
## POP_PCB7        -7.809e-07  1.183e-06  -0.660   0.5094
## POP_PCB8        -8.255e-07  1.648e-06  -0.501   0.6166
## POP_PCB9        -3.387e-07  1.787e-06  -0.190   0.8497
## POP_PCB10        9.249e-04  8.414e-04   1.099   0.2720
## POP_PCB11        9.491e-05  2.909e-04   0.326   0.7444
## POP_dioxin1      1.933e-05  3.028e-04   0.064   0.9491
## POP_dioxin2     -1.104e-04  4.349e-04  -0.254   0.7997
## POP_dioxin3     -1.856e-05  2.997e-05  -0.619   0.5359
## POP_furan1       2.570e-03  3.799e-03   0.677   0.4989
## POP_furan2      -6.152e-04  4.485e-03  -0.137   0.8909
## POP_furan3       4.443e-03  2.753e-03   1.614   0.1069
## POP_furan4      -6.314e-04  9.170e-04  -0.689   0.4913
## whitecell_count -5.469e-03  4.388e-03  -1.246   0.2130
## lymphocyte_pct  -1.364e-03  1.025e-03  -1.330   0.1838
## monocyte_pct    -5.733e-03  4.032e-03  -1.422   0.1554
## basophils_pct    1.182e-03  3.460e-03   0.342   0.7326
## neutrophils_pct  1.103e-02  1.660e-02   0.664   0.5066
## BMI             -1.332e-03  1.408e-03  -0.946   0.3445
## edu_cat2         2.268e-02  2.208e-02   1.027   0.3046
## edu_cat3         4.375e-02  2.143e-02   2.041   0.0416 *
## edu_cat4         2.929e-02  2.541e-02   1.153   0.2494
## race_cat2       -2.311e-02  3.262e-02  -0.709   0.4788
## race_cat3        2.558e-02  3.359e-02   0.761   0.4466
## race_cat4       -3.492e-02  2.987e-02  -1.169   0.2428
## male1           -3.975e-02  1.770e-02  -2.246   0.0250 *
## ageyrs          -6.308e-03  7.425e-04  -8.495   <2e-16 ***
## yrssmoke        -4.532e-04  7.230e-04  -0.627   0.5310
## smokenow1       -5.910e-04  3.569e-02  -0.017   0.9868
## ln_lbxcot        5.473e-03  3.917e-03   1.397   0.1626
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2219 on 831 degrees of freedom
## Multiple R-squared:  0.2425, Adjusted R-squared:  0.2133
## F-statistic: 8.313 on 32 and 831 DF,  p-value: < 2.2e-16
```

```
#show the VIF
vif(model)
```

```
##                     GVIF Df GVIF^(1/(2*Df))
## POP_PCB3        8.558618  1        2.925512
## POP_PCB4       14.260201  1        3.776268
## POP_PCB6        9.430583  1        3.070925
## POP_PCB7        4.689281  1        2.165475
## POP_PCB8        5.899020  1        2.428790
## POP_PCB9        8.900140  1        2.983310
## POP_PCB10       5.368201  1        2.316938
## POP_PCB11       4.225811  1        2.055678
## POP_dioxin1     5.187456  1        2.277599
## POP_dioxin2     5.303754  1        2.302988
```

```
## POP_dioxin3      4.318541  1        2.078110
## POP_furan1       6.013167  1        2.452176
## POP_furan2       6.155357  1        2.480999
## POP_furan3       4.443638  1        2.107994
## POP_furan4       1.812808  1        1.346405
## whitecell_count  1.535444  1        1.239130
## lymphocyte_pct   1.371183  1        1.170975
## monocyte_pct     1.258625  1        1.121885
## basophils_pct    1.104048  1        1.050737
## neutrophils_pct  1.087848  1        1.042999
## BMI              1.260084  1        1.122535
## edu_cat          1.499070  3        1.069803
## race_cat         2.035520  3        1.125760
## male             1.349340  1        1.161611
## ageyrs           3.225616  1        1.796000
## yrssmoke         2.180551  1        1.476669
## smokenow         3.974936  1        1.993724
## ln_lbxcot        3.946444  1        1.986566
```

```
#remove highest vif eosinophils_pct ~ 15047 then POP_PCB5 ~ 59.718357 then POP_PCB1 ~ 32.255513 then PO
# then POP_PCB4 ~ 14.260201
pollutants$POP_PCB4 = NULL
#fit new model
model = lm(length ~ .,data = pollutants)
#summary
summary(model)
```

```
##
## Call:
## lm(formula = length ~ ., data = pollutants)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.49174 -0.15062 -0.02829  0.12117  1.19031
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.505e+00  8.339e-02  18.048   <2e-16 ***
## POP_PCB3         1.039e-06  1.624e-06   0.640   0.5223
## POP_PCB6         2.008e-07  9.446e-07   0.213   0.8317
## POP_PCB7        -7.992e-07  1.182e-06  -0.676   0.4993
## POP_PCB8        -7.951e-07  1.647e-06  -0.483   0.6294
## POP_PCB9        -7.655e-07  1.655e-06  -0.463   0.6438
## POP_PCB10        8.170e-04  8.238e-04   0.992   0.3216
## POP_PCB11        8.365e-05  2.903e-04   0.288   0.7733
## POP_dioxin1      2.403e-05  3.026e-04   0.079   0.9367
## POP_dioxin2     -1.306e-04  4.336e-04  -0.301   0.7633
## POP_dioxin3     -1.933e-05  2.993e-05  -0.646   0.5185
## POP_furan1       2.682e-03  3.793e-03   0.707   0.4797
## POP_furan2      -5.840e-04  4.483e-03  -0.130   0.8964
## POP_furan3       4.589e-03  2.742e-03   1.673   0.0946 .
## POP_furan4      -6.331e-04  9.166e-04  -0.691   0.4900
## whitecell_count -5.564e-03  4.384e-03  -1.269   0.2047
## lymphocyte_pct  -1.355e-03  1.025e-03  -1.323   0.1863
## monocyte_pct    -5.859e-03  4.025e-03  -1.456   0.1459
```

```
## basophils_pct    1.356e-03  3.448e-03   0.393   0.6941
## neutrophils_pct  1.038e-02  1.656e-02   0.627   0.5311
## BMI             -1.292e-03  1.406e-03  -0.919   0.3585
## edu_cat2         2.258e-02  2.207e-02   1.023   0.3066
## edu_cat3         4.383e-02  2.143e-02   2.046   0.0411 *
## edu_cat4         2.954e-02  2.539e-02   1.163   0.2450
## race_cat2       -2.332e-02  3.260e-02  -0.715   0.4746
## race_cat3        2.375e-02  3.346e-02   0.710   0.4780
## race_cat4       -3.549e-02  2.985e-02  -1.189   0.2348
## male1           -4.034e-02  1.767e-02  -2.282   0.0227 *
## ageyrs          -6.317e-03  7.421e-04  -8.512   <2e-16 ***
## yrssmoke        -5.096e-04  7.173e-04  -0.710   0.4776
## smokenow1        4.148e-04  3.564e-02   0.012   0.9907
## ln_lbxcot        5.455e-03  3.915e-03   1.393   0.1639
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2219 on 832 degrees of freedom
## Multiple R-squared:  0.2421, Adjusted R-squared:  0.2139
## F-statistic: 8.574 on 31 and 832 DF,  p-value: < 2.2e-16
```

```
#show the VIF
vif(model)
```

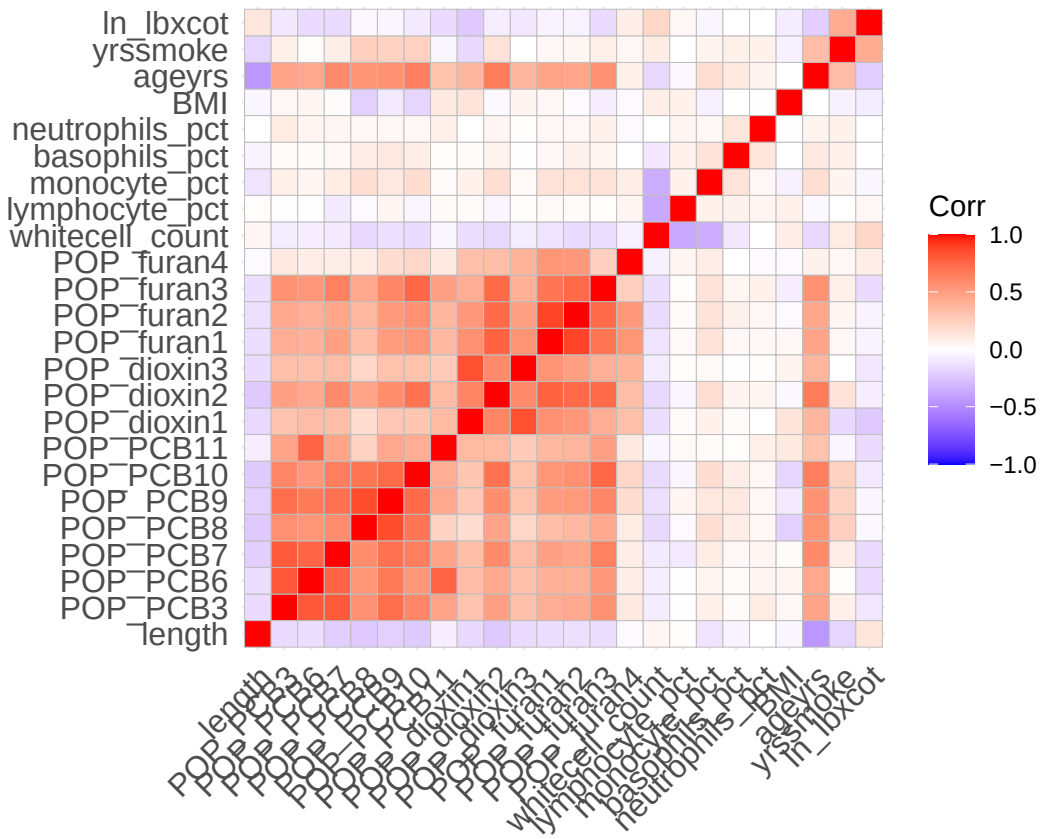
```
##                     GVIF Df GVIF^(1/(2*Df))
## POP_PCB3        5.310340  1        2.304417
## POP_PCB6        9.083828  1        3.013939
## POP_PCB7        4.686485  1        2.164829
## POP_PCB8        5.894052  1        2.427767
## POP_PCB9        7.640480  1        2.764142
## POP_PCB10       5.149483  1        2.269247
## POP_PCB11       4.210120  1        2.051858
## POP_dioxin1     5.184345  1        2.276916
## POP_dioxin2     5.275271  1        2.296796
## POP_dioxin3     4.311410  1        2.076394
## POP_furan1      6.000097  1        2.449509
## POP_furan2      6.154621  1        2.480851
## POP_furan3      4.412739  1        2.100652
## POP_furan4      1.812793  1        1.346400
## whitecell_count 1.533642  1        1.238403
## lymphocyte_pct  1.370966  1        1.170882
## monocyte_pct    1.255543  1        1.120510
## basophils_pct   1.097132  1        1.047441
## neutrophils_pct 1.083675  1        1.040997
## BMI             1.257562  1        1.121411
## edu_cat         1.498239  3        1.069704
## race_cat        2.012804  3        1.123657
## male            1.345703  1        1.160045
## ageyrs          3.224432  1        1.795670
## yrssmoke        2.147610  1        1.465473
## smokenow        3.967106  1        1.991759
## ln_lbxcot       3.946223  1        1.986510
```

```
#get set a dataset with no categorical covariates
no_cat = pollutants
```

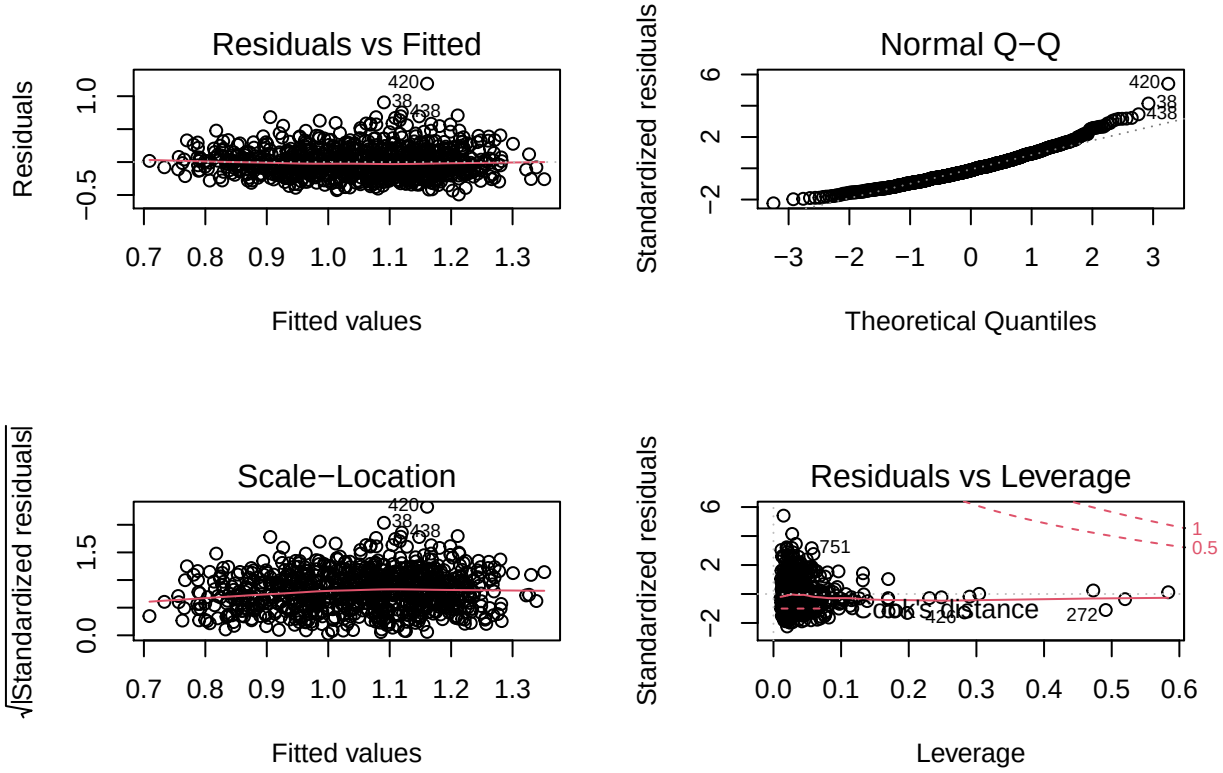
```
no_cat$edu_cat = NULL
no_cat$race_cat = NULL
no_cat$male = NULL
no_cat$smokenow = NULL
#summary of the dataset
summary(no_cat)
```

```
##      length          POP_PCB3         POP_PCB6         POP_PCB7
##  Min.   :0.5266   Min.   :  2000   Min.   :  2000   Min.   :  1100
##  1st Qu.:0.8754   1st Qu.:  3700   1st Qu.:  4400   1st Qu.:  4000
##  Median :1.0286   Median :  6200   Median :  9400   Median :  7450
##  Mean   :1.0543   Mean   : 10158   Mean   : 16820   Mean   : 12682
##  3rd Qu.:1.2095   3rd Qu.: 12000   3rd Qu.: 19500   3rd Qu.: 15625
##  Max.   :2.3512   Max.   :123000   Max.   :319000   Max.   :144000
##     POP_PCB8         POP_PCB9        POP_PCB10        POP_PCB11
##  Min.   :  1100   Min.   :  1100   Min.   :  1.70   Min.   :  1.30
##  1st Qu.:  3800   1st Qu.:  3900   1st Qu.:  9.10   1st Qu.: 14.80
##  Median :  6950   Median :  8050   Median : 18.35   Median : 24.50
##  Mean   : 10530   Mean   : 12220   Mean   : 24.49   Mean   : 38.15
##  3rd Qu.: 14425   3rd Qu.: 16025   3rd Qu.: 34.90   3rd Qu.: 42.95
##  Max.   :187000   Max.   :144000   Max.   :172.00   Max.   :845.00
##   POP_dioxin1      POP_dioxin2      POP_dioxin3       POP_furan1
##  Min.   :  1.90   Min.   :  1.40   Min.   :  36.8   Min.   : 1.000
##  1st Qu.: 23.90   1st Qu.: 21.27   1st Qu.: 197.0   1st Qu.: 3.200
##  Median : 41.35   Median : 37.80   Median : 342.5   Median : 5.200
##  Mean   : 57.65   Mean   : 47.81   Mean   : 494.4   Mean   : 6.371
##  3rd Qu.: 71.62   3rd Qu.: 62.42   3rd Qu.: 603.0   3rd Qu.: 7.700
##  Max.   :760.00   Max.   :281.00   Max.   :8190.0   Max.   :44.400
##    POP_furan2       POP_furan3       POP_furan4     whitecell_count
##  Min.   : 0.800   Min.   : 0.700   Min.   :  0.90   Min.   : 2.300
##  1st Qu.: 2.600   1st Qu.: 2.200   1st Qu.:  6.40   1st Qu.: 5.600
##  Median : 4.200   Median : 5.050   Median :  9.65   Median : 6.900
##  Mean   : 5.390   Mean   : 6.669   Mean   : 11.54   Mean   : 7.191
##  3rd Qu.: 6.825   3rd Qu.: 9.300   3rd Qu.: 14.00   3rd Qu.: 8.300
##  Max.   :33.500   Max.   :38.300   Max.   :234.00   Max.   :20.100
##  lymphocyte_pct   monocyte_pct    basophils_pct    neutrophils_pct
##  Min.   : 5.80   Min.   : 1.600   Min.   : 0.000   Min.   :0.0000
##  1st Qu.:24.00   1st Qu.: 6.600   1st Qu.: 1.500   1st Qu.:0.4000
##  Median :28.95   Median : 7.700   Median : 2.300   Median :0.6000
##  Mean   :29.92   Mean   : 7.936   Mean   : 2.903   Mean   :0.6669
##  3rd Qu.:35.42   3rd Qu.: 9.100   3rd Qu.: 3.700   3rd Qu.:0.8000
##  Max.   :73.40   Max.   :23.800   Max.   :28.200   Max.   :5.5000
##       BMI            ageyrs         yrssmoke      ln_lbxcot
##  Min.   :16.16   Min.   :20.00   Min.   : 0.0   Min.   :-4.5099
##  1st Qu.:23.88   1st Qu.:34.00   1st Qu.: 0.0   1st Qu.:-4.0745
##  Median :27.38   Median :46.00   Median : 0.0   Median :-2.7334
##  Mean   :28.09   Mean   :48.36   Mean   :10.6   Mean   :-0.9804
##  3rd Qu.:31.17   3rd Qu.:63.00   3rd Qu.:20.0   3rd Qu.: 2.8000
##  Max.   :62.99   Max.   :85.00   Max.   :69.0   Max.   : 6.5848
```

```
#calculate correlation matrix
corr_matrix = cor(no_cat)
#graph colored corr matrix
ggcorrplot(corr_matrix)
```




```
#find the model fit for homoscedasticity after removing multicolinearity
par(mfrow=c(2,2))
plot(model)
```




```
par(mfrow=c(1,1))

#set up train and test model
data = model.matrix(length ~ ., data = pollutants)
n = nrow(data)
#set seed for sample and
set.seed(331)
#get index for random train and test index (90% train, 10% test)
train_row = sample(1:n, 0.9*n)
#train set
x_matrix = data[,-1]
y_matrix = pollutants$length

#get the y values
train_y = y_matrix[train_row]
#get the x values
train_x = x_matrix[train_row,]
#get the y values
test_y = y_matrix[-train_row]
#get the x values
test_x = x_matrix[-train_row,]

eval_results <- function(true, predicted, df){
  SSE <- sum((predicted - true)^2)
  SST <- sum((true - mean(true))^2)
  R_square <- 1 - SSE / SST
```

```
  RMSE = sqrt(SSE/nrow(df))


  # Model performance metrics
  data.frame(
    RMSE = RMSE,
    Rsquare = R_square)
}

#k fold lasso
lambdas <- 10^seq(2, -2, by = -.0001)
lasso = cv.glmnet(train_x, train_y, alpha = 1, lambda = lambdas)
best_lam = lasso$lambda.min
best_lam
```

```
## [1] 0.01
```

```
#use lasso with best lambda
best_lasso = glmnet(train_x, train_y, alpha = 1, lambda = best_lam)
#predict with lasso result with training set
pred_train = predict(best_lasso, s = best_lam, newx = train_x)
eval_results(train_y, pred_train, train_x)
```

```
##        RMSE   Rsquare
## 1 0.2206752 0.2146218
```

```
#predict with lasso result with test set
pred_train = predict(best_lasso, s = best_lam, newx = test_x)
eval_results(test_y, pred_train, test_x)
```

```
##        RMSE   Rsquare
## 1 0.2313771 0.2034645
```

```
#training data set
train_data = pollutants[train_row,]
#training model
train_model = lm(length ~ ., data = train_data)

#step wise AIC on training data
step_aic = step(train_model, direction = "both")
```

```
## Start:  AIC=-2308.9
## length ~ POP_PCB3 + POP_PCB6 + POP_PCB7 + POP_PCB8 + POP_PCB9 +
##     POP_PCB10 + POP_PCB11 + POP_dioxin1 + POP_dioxin2 + POP_dioxin3 +
##     POP_furan1 + POP_furan2 + POP_furan3 + POP_furan4 + whitecell_count +
##     lymphocyte_pct + monocyte_pct + basophils_pct + neutrophils_pct +
##     BMI + edu_cat + race_cat + male + ageyrs + yrssmoke + smokenow +
##     ln_lbxcot
##
##                  Df Sum of Sq    RSS     AIC
## - edu_cat         3   0.09431 36.748 -2312.9
## - POP_PCB9        1   0.00001 36.654 -2310.9
## - POP_PCB7        1   0.00009 36.654 -2310.9
## - POP_PCB3        1   0.00023 36.654 -2310.9
## - POP_furan4      1   0.00030 36.654 -2310.9
## - POP_PCB11       1   0.00048 36.654 -2310.9
```

```
## - POP_furan1       1   0.00087 36.654 -2310.9
## - POP_furan2       1   0.00284 36.656 -2310.8
## - POP_dioxin2      1   0.00321 36.657 -2310.8
## - POP_PCB6         1   0.00421 36.658 -2310.8
## - basophils_pct    1   0.00470 36.658 -2310.8
## - yrssmoke         1   0.00501 36.659 -2310.8
## - BMI              1   0.00545 36.659 -2310.8
## - POP_dioxin1      1   0.00584 36.659 -2310.8
## - POP_dioxin3      1   0.01055 36.664 -2310.7
## - smokenow         1   0.01208 36.666 -2310.7
## - neutrophils_pct  1   0.01232 36.666 -2310.6
## - POP_PCB8         1   0.02541 36.679 -2310.4
## - monocyte_pct     1   0.03974 36.693 -2310.1
## - POP_PCB10        1   0.04069 36.694 -2310.0
## - race_cat         3   0.25456 36.908 -2309.5
## <none>                         36.654 -2308.9
## - ln_lbxcot        1   0.12280 36.776 -2308.3
## - lymphocyte_pct   1   0.12699 36.781 -2308.2
## - POP_furan3       1   0.12907 36.783 -2308.2
## - whitecell_count  1   0.12929 36.783 -2308.2
## - male             1   0.30812 36.962 -2304.4
## - ageyrs           1   3.11341 39.767 -2247.6
##
## Step:  AIC=-2312.91
## length ~ POP_PCB3 + POP_PCB6 + POP_PCB7 + POP_PCB8 + POP_PCB9 +
##     POP_PCB10 + POP_PCB11 + POP_dioxin1 + POP_dioxin2 + POP_dioxin3 +
##     POP_furan1 + POP_furan2 + POP_furan3 + POP_furan4 + whitecell_count +
##     lymphocyte_pct + monocyte_pct + basophils_pct + neutrophils_pct +
##     BMI + race_cat + male + ageyrs + yrssmoke + smokenow + ln_lbxcot
##
##                   Df Sum of Sq    RSS     AIC
## - POP_PCB9         1    0.0002 36.748 -2314.9
## - POP_PCB11        1    0.0003 36.748 -2314.9
## - POP_furan4       1    0.0005 36.748 -2314.9
## - POP_PCB7         1    0.0006 36.748 -2314.9
## - POP_furan1       1    0.0007 36.749 -2314.9
## - POP_furan2       1    0.0017 36.750 -2314.9
## - POP_PCB3         1    0.0018 36.750 -2314.9
## - BMI              1    0.0036 36.752 -2314.8
## - POP_dioxin2      1    0.0047 36.753 -2314.8
## - POP_PCB6         1    0.0054 36.753 -2314.8
## - basophils_pct    1    0.0058 36.754 -2314.8
## - yrssmoke         1    0.0074 36.755 -2314.8
## - POP_dioxin1      1    0.0082 36.756 -2314.7
## - POP_dioxin3      1    0.0100 36.758 -2314.7
## - smokenow         1    0.0111 36.759 -2314.7
## - neutrophils_pct  1    0.0156 36.764 -2314.6
## - POP_PCB8         1    0.0248 36.773 -2314.4
## - monocyte_pct     1    0.0378 36.786 -2314.1
## - POP_PCB10        1    0.0481 36.796 -2313.9
## - race_cat         3    0.2452 36.993 -2313.7
## <none>                         36.748 -2312.9
## - ln_lbxcot        1    0.1037 36.852 -2312.7
## - POP_furan3       1    0.1400 36.888 -2311.9
```

```
## - whitecell_count  1    0.1433 36.891 -2311.9
## - lymphocyte_pct   1    0.1443 36.892 -2311.9
## + edu_cat          3    0.0943 36.654 -2308.9
## - male             1    0.3067 37.055 -2308.4
## - ageyrs           1    3.2308 39.979 -2249.4
##
## Step:  AIC=-2314.9
## length ~ POP_PCB3 + POP_PCB6 + POP_PCB7 + POP_PCB8 + POP_PCB10 +
##     POP_PCB11 + POP_dioxin1 + POP_dioxin2 + POP_dioxin3 + POP_furan1 +
##     POP_furan2 + POP_furan3 + POP_furan4 + whitecell_count +
##     lymphocyte_pct + monocyte_pct + basophils_pct + neutrophils_pct +
##     BMI + race_cat + male + ageyrs + yrssmoke + smokenow + ln_lbxcot
##
##                   Df Sum of Sq    RSS     AIC
## - POP_PCB11        1    0.0002 36.748 -2316.9
## - POP_furan4       1    0.0004 36.749 -2316.9
## - POP_PCB7         1    0.0005 36.749 -2316.9
## - POP_furan1       1    0.0007 36.749 -2316.9
## - POP_furan2       1    0.0017 36.750 -2316.9
## - POP_PCB3         1    0.0023 36.750 -2316.8
## - BMI              1    0.0036 36.752 -2316.8
## - POP_dioxin2      1    0.0047 36.753 -2316.8
## - POP_PCB6         1    0.0059 36.754 -2316.8
## - basophils_pct    1    0.0064 36.755 -2316.8
## - yrssmoke         1    0.0079 36.756 -2316.7
## - POP_dioxin1      1    0.0082 36.756 -2316.7
## - POP_dioxin3      1    0.0100 36.758 -2316.7
## - smokenow         1    0.0109 36.759 -2316.7
## - neutrophils_pct  1    0.0162 36.764 -2316.6
## - monocyte_pct     1    0.0376 36.786 -2316.1
## - POP_PCB10        1    0.0479 36.796 -2315.9
## - POP_PCB8         1    0.0581 36.806 -2315.7
## - race_cat         3    0.2507 36.999 -2315.6
## <none>                         36.748 -2314.9
## - ln_lbxcot        1    0.1035 36.852 -2314.7
## - POP_furan3       1    0.1400 36.888 -2313.9
## - whitecell_count  1    0.1446 36.893 -2313.8
## - lymphocyte_pct   1    0.1495 36.898 -2313.8
## + POP_PCB9         1    0.0002 36.748 -2312.9
## + edu_cat          3    0.0945 36.654 -2310.9
## - male             1    0.3065 37.055 -2310.4
## - ageyrs           1    3.2478 39.996 -2251.1
##
## Step:  AIC=-2316.9
## length ~ POP_PCB3 + POP_PCB6 + POP_PCB7 + POP_PCB8 + POP_PCB10 +
##     POP_dioxin1 + POP_dioxin2 + POP_dioxin3 + POP_furan1 + POP_furan2 +
##     POP_furan3 + POP_furan4 + whitecell_count + lymphocyte_pct +
##     monocyte_pct + basophils_pct + neutrophils_pct + BMI + race_cat +
##     male + ageyrs + yrssmoke + smokenow + ln_lbxcot
##
##                   Df Sum of Sq    RSS     AIC
## - POP_PCB7         1    0.0004 36.749 -2318.9
## - POP_furan4       1    0.0004 36.749 -2318.9
## - POP_furan1       1    0.0007 36.749 -2318.9
```

```
## - POP_furan2       1    0.0016 36.750 -2318.9
## - POP_PCB3         1    0.0032 36.751 -2318.8
## - BMI              1    0.0034 36.752 -2318.8
## - POP_dioxin2      1    0.0049 36.753 -2318.8
## - basophils_pct    1    0.0064 36.755 -2318.8
## - yrssmoke         1    0.0078 36.756 -2318.7
## - POP_dioxin1      1    0.0080 36.756 -2318.7
## - POP_dioxin3      1    0.0105 36.759 -2318.7
## - smokenow         1    0.0109 36.759 -2318.7
## - neutrophils_pct  1    0.0169 36.765 -2318.5
## - POP_PCB6         1    0.0185 36.767 -2318.5
## - monocyte_pct     1    0.0377 36.786 -2318.1
## - POP_PCB10        1    0.0543 36.803 -2317.8
## - race_cat         3    0.2516 37.000 -2317.6
## - POP_PCB8         1    0.0689 36.817 -2317.4
## <none>                         36.748 -2316.9
## - ln_lbxcot        1    0.1035 36.852 -2316.7
## - whitecell_count  1    0.1446 36.893 -2315.8
## - POP_furan3       1    0.1450 36.893 -2315.8
## - lymphocyte_pct   1    0.1494 36.898 -2315.7
## + POP_PCB11        1    0.0002 36.748 -2314.9
## + POP_PCB9         1    0.0001 36.748 -2314.9
## + edu_cat          3    0.0942 36.654 -2312.9
## - male             1    0.3065 37.055 -2312.4
## - ageyrs           1    3.2482 39.996 -2253.1
##
## Step:  AIC=-2318.89
## length ~ POP_PCB3 + POP_PCB6 + POP_PCB8 + POP_PCB10 + POP_dioxin1 +
##     POP_dioxin2 + POP_dioxin3 + POP_furan1 + POP_furan2 + POP_furan3 +
##     POP_furan4 + whitecell_count + lymphocyte_pct + monocyte_pct +
##     basophils_pct + neutrophils_pct + BMI + race_cat + male +
##     ageyrs + yrssmoke + smokenow + ln_lbxcot
##
##                   Df Sum of Sq    RSS     AIC
## - POP_furan4       1    0.0005 36.749 -2320.9
## - POP_furan1       1    0.0006 36.749 -2320.9
## - POP_furan2       1    0.0014 36.750 -2320.9
## - POP_PCB3         1    0.0028 36.752 -2320.8
## - BMI              1    0.0033 36.752 -2320.8
## - POP_dioxin2      1    0.0047 36.753 -2320.8
## - basophils_pct    1    0.0064 36.755 -2320.8
## - yrssmoke         1    0.0079 36.757 -2320.7
## - POP_dioxin1      1    0.0083 36.757 -2320.7
## - POP_dioxin3      1    0.0105 36.759 -2320.7
## - smokenow         1    0.0110 36.760 -2320.7
## - neutrophils_pct  1    0.0165 36.765 -2320.5
## - POP_PCB6         1    0.0221 36.771 -2320.4
## - monocyte_pct     1    0.0376 36.786 -2320.1
## - POP_PCB10        1    0.0544 36.803 -2319.7
## - race_cat         3    0.2519 37.001 -2319.6
## - POP_PCB8         1    0.0686 36.817 -2319.4
## <none>                         36.749 -2318.9
## - ln_lbxcot        1    0.1035 36.852 -2318.7
## - whitecell_count  1    0.1460 36.895 -2317.8
```

```
## - POP_furan3       1    0.1532 36.902 -2317.7
## - lymphocyte_pct   1    0.1539 36.903 -2317.6
## + POP_PCB7         1    0.0004 36.748 -2316.9
## + POP_PCB9         1    0.0001 36.749 -2316.9
## + POP_PCB11        1    0.0001 36.749 -2316.9
## + edu_cat          3    0.0946 36.654 -2314.9
## - male             1    0.3262 37.075 -2314.0
## - ageyrs           1    3.3472 40.096 -2253.2
##
## Step:  AIC=-2320.88
## length ~ POP_PCB3 + POP_PCB6 + POP_PCB8 + POP_PCB10 + POP_dioxin1 +
##     POP_dioxin2 + POP_dioxin3 + POP_furan1 + POP_furan2 + POP_furan3 +
##     whitecell_count + lymphocyte_pct + monocyte_pct + basophils_pct +
##     neutrophils_pct + BMI + race_cat + male + ageyrs + yrssmoke +
##     smokenow + ln_lbxcot
##
##                   Df Sum of Sq    RSS     AIC
## - POP_furan1       1    0.0008 36.750 -2322.9
## - POP_furan2       1    0.0011 36.750 -2322.9
## - POP_PCB3         1    0.0028 36.752 -2322.8
## - BMI              1    0.0032 36.752 -2322.8
## - POP_dioxin2      1    0.0045 36.754 -2322.8
## - basophils_pct    1    0.0064 36.756 -2322.7
## - yrssmoke         1    0.0081 36.757 -2322.7
## - POP_dioxin1      1    0.0081 36.757 -2322.7
## - smokenow         1    0.0111 36.760 -2322.6
## - POP_dioxin3      1    0.0121 36.761 -2322.6
## - neutrophils_pct  1    0.0166 36.766 -2322.5
## - POP_PCB6         1    0.0223 36.771 -2322.4
## - monocyte_pct     1    0.0378 36.787 -2322.1
## - POP_PCB10        1    0.0540 36.803 -2321.7
## - race_cat         3    0.2519 37.001 -2321.6
## - POP_PCB8         1    0.0683 36.817 -2321.4
## <none>                         36.749 -2320.9
## - ln_lbxcot        1    0.1040 36.853 -2320.7
## - whitecell_count  1    0.1462 36.895 -2319.8
## - lymphocyte_pct   1    0.1546 36.904 -2319.6
## - POP_furan3       1    0.1617 36.911 -2319.5
## + POP_furan4       1    0.0005 36.749 -2318.9
## + POP_PCB7         1    0.0004 36.749 -2318.9
## + POP_PCB11        1    0.0001 36.749 -2318.9
## + POP_PCB9         1    0.0001 36.749 -2318.9
## + edu_cat          3    0.0947 36.654 -2316.9
## - male             1    0.3303 37.079 -2315.9
## - ageyrs           1    3.4399 40.189 -2253.3
##
## Step:  AIC=-2322.86
## length ~ POP_PCB3 + POP_PCB6 + POP_PCB8 + POP_PCB10 + POP_dioxin1 +
##     POP_dioxin2 + POP_dioxin3 + POP_furan2 + POP_furan3 + whitecell_count +
##     lymphocyte_pct + monocyte_pct + basophils_pct + neutrophils_pct +
##     BMI + race_cat + male + ageyrs + yrssmoke + smokenow + ln_lbxcot
##
##                   Df Sum of Sq    RSS     AIC
## - POP_furan2       1    0.0003 36.750 -2324.9
```

```
## - POP_PCB3         1    0.0027 36.753 -2324.8
## - BMI              1    0.0036 36.754 -2324.8
## - POP_dioxin2      1    0.0058 36.756 -2324.7
## - basophils_pct    1    0.0062 36.756 -2324.7
## - yrssmoke         1    0.0078 36.758 -2324.7
## - POP_dioxin1      1    0.0079 36.758 -2324.7
## - smokenow         1    0.0113 36.761 -2324.6
## - POP_dioxin3      1    0.0131 36.763 -2324.6
## - neutrophils_pct  1    0.0167 36.767 -2324.5
## - POP_PCB6         1    0.0222 36.772 -2324.4
## - monocyte_pct     1    0.0388 36.789 -2324.0
## - POP_PCB10        1    0.0541 36.804 -2323.7
## - race_cat         3    0.2511 37.001 -2323.6
## - POP_PCB8         1    0.0680 36.818 -2323.4
## <none>                         36.750 -2322.9
## - ln_lbxcot        1    0.1040 36.854 -2322.7
## - whitecell_count  1    0.1485 36.898 -2321.7
## - lymphocyte_pct   1    0.1563 36.906 -2321.6
## - POP_furan3       1    0.1616 36.912 -2321.4
## + POP_furan1       1    0.0008 36.749 -2320.9
## + POP_furan4       1    0.0007 36.749 -2320.9
## + POP_PCB7         1    0.0002 36.750 -2320.9
## + POP_PCB9         1    0.0001 36.750 -2320.9
## + POP_PCB11        1    0.0001 36.750 -2320.9
## + edu_cat          3    0.0945 36.655 -2318.9
## - male             1    0.3312 37.081 -2317.9
## - ageyrs           1    3.4460 40.196 -2255.2
##
## Step:  AIC=-2324.86
## length ~ POP_PCB3 + POP_PCB6 + POP_PCB8 + POP_PCB10 + POP_dioxin1 +
##     POP_dioxin2 + POP_dioxin3 + POP_furan3 + whitecell_count +
##     lymphocyte_pct + monocyte_pct + basophils_pct + neutrophils_pct +
##     BMI + race_cat + male + ageyrs + yrssmoke + smokenow + ln_lbxcot
##
##                    Df Sum of Sq    RSS     AIC
## - POP_PCB3          1    0.0027 36.753 -2326.8
## - BMI               1    0.0036 36.754 -2326.8
## - POP_dioxin2       1    0.0055 36.756 -2326.7
## - basophils_pct     1    0.0061 36.756 -2326.7
## - POP_dioxin1       1    0.0077 36.758 -2326.7
## - yrssmoke          1    0.0078 36.758 -2326.7
## - smokenow          1    0.0113 36.762 -2326.6
## - POP_dioxin3       1    0.0131 36.763 -2326.6
## - neutrophils_pct   1    0.0166 36.767 -2326.5
## - POP_PCB6          1    0.0223 36.773 -2326.4
## - monocyte_pct      1    0.0387 36.789 -2326.0
## - POP_PCB10         1    0.0538 36.804 -2325.7
## - race_cat          3    0.2549 37.005 -2325.5
## - POP_PCB8          1    0.0677 36.818 -2325.4
## <none>                         36.750 -2324.9
## - ln_lbxcot         1    0.1045 36.855 -2324.7
## - whitecell_count   1    0.1488 36.899 -2323.7
## - lymphocyte_pct    1    0.1563 36.907 -2323.6
## + POP_furan2        1    0.0003 36.750 -2322.9
```

```
## + POP_furan4       1    0.0002 36.750 -2322.9
## + POP_PCB7         1    0.0002 36.750 -2322.9
## + POP_PCB9         1    0.0001 36.750 -2322.9
## + POP_PCB11        1    0.0001 36.750 -2322.9
## + POP_furan1       1    0.0001 36.750 -2322.9
## - POP_furan3       1    0.2063 36.957 -2322.5
## + edu_cat          3    0.0934 36.657 -2320.8
## - male             1    0.3324 37.083 -2319.9
## - ageyrs           1    3.4652 40.215 -2256.8
##
## Step:  AIC=-2326.8
## length ~ POP_PCB6 + POP_PCB8 + POP_PCB10 + POP_dioxin1 + POP_dioxin2 +
##     POP_dioxin3 + POP_furan3 + whitecell_count + lymphocyte_pct +
##     monocyte_pct + basophils_pct + neutrophils_pct + BMI + race_cat +
##     male + ageyrs + yrssmoke + smokenow + ln_lbxcot
##
##                   Df Sum of Sq    RSS     AIC
## - BMI              1    0.0046 36.758 -2328.7
## - POP_dioxin2      1    0.0056 36.759 -2328.7
## - basophils_pct    1    0.0058 36.759 -2328.7
## - POP_dioxin1      1    0.0070 36.760 -2328.7
## - yrssmoke         1    0.0077 36.761 -2328.6
## - smokenow         1    0.0109 36.764 -2328.6
## - POP_dioxin3      1    0.0142 36.767 -2328.5
## - neutrophils_pct  1    0.0154 36.768 -2328.5
## - POP_PCB6         1    0.0230 36.776 -2328.3
## - monocyte_pct     1    0.0384 36.791 -2328.0
## - POP_PCB10        1    0.0512 36.804 -2327.7
## - race_cat         3    0.2548 37.008 -2327.4
## - POP_PCB8         1    0.0714 36.824 -2327.3
## <none>                         36.753 -2326.8
## - ln_lbxcot        1    0.1033 36.856 -2326.6
## - whitecell_count  1    0.1484 36.901 -2325.7
## - lymphocyte_pct   1    0.1557 36.909 -2325.5
## + POP_PCB3         1    0.0027 36.750 -2324.9
## + POP_PCB11        1    0.0010 36.752 -2324.8
## + POP_PCB9         1    0.0005 36.753 -2324.8
## + POP_furan2       1    0.0004 36.753 -2324.8
## + POP_furan4       1    0.0002 36.753 -2324.8
## + POP_PCB7         1    0.0001 36.753 -2324.8
## + POP_furan1       1    0.0000 36.753 -2324.8
## - POP_furan3       1    0.2048 36.958 -2324.5
## + edu_cat          3    0.0955 36.658 -2322.8
## - male             1    0.3309 37.084 -2321.8
## - ageyrs           1    3.4628 40.216 -2258.8
##
## Step:  AIC=-2328.7
## length ~ POP_PCB6 + POP_PCB8 + POP_PCB10 + POP_dioxin1 + POP_dioxin2 +
##     POP_dioxin3 + POP_furan3 + whitecell_count + lymphocyte_pct +
##     monocyte_pct + basophils_pct + neutrophils_pct + race_cat +
##     male + ageyrs + yrssmoke + smokenow + ln_lbxcot
##
##                   Df Sum of Sq    RSS     AIC
## - POP_dioxin2      1    0.0058 36.763 -2330.6
```

```
## - basophils_pct    1    0.0060 36.764 -2330.6
## - yrssmoke         1    0.0084 36.766 -2330.5
## - POP_dioxin1      1    0.0087 36.766 -2330.5
## - smokenow         1    0.0104 36.768 -2330.5
## - POP_dioxin3      1    0.0129 36.770 -2330.4
## - neutrophils_pct  1    0.0156 36.773 -2330.4
## - POP_PCB6         1    0.0201 36.778 -2330.3
## - monocyte_pct     1    0.0381 36.796 -2329.9
## - POP_PCB10        1    0.0563 36.814 -2329.5
## - race_cat         3    0.2511 37.009 -2329.4
## - POP_PCB8         1    0.0674 36.825 -2329.3
## <none>                         36.758 -2328.7
## - ln_lbxcot        1    0.1042 36.862 -2328.5
## - whitecell_count  1    0.1594 36.917 -2327.3
## - lymphocyte_pct   1    0.1622 36.920 -2327.3
## + BMI              1    0.0046 36.753 -2326.8
## + POP_PCB3         1    0.0037 36.754 -2326.8
## + POP_PCB11        1    0.0009 36.757 -2326.7
## + POP_PCB9         1    0.0007 36.757 -2326.7
## + POP_PCB7         1    0.0004 36.757 -2326.7
## + POP_furan2       1    0.0004 36.757 -2326.7
## + POP_furan1       1    0.0001 36.757 -2326.7
## + POP_furan4       1    0.0001 36.757 -2326.7
## - POP_furan3       1    0.2065 36.964 -2326.3
## + edu_cat          3    0.0938 36.664 -2324.7
## - male             1    0.3399 37.097 -2323.6
## - ageyrs           1    3.4942 40.252 -2260.1
##
## Step:  AIC=-2330.58
## length ~ POP_PCB6 + POP_PCB8 + POP_PCB10 + POP_dioxin1 + POP_dioxin3 +
##     POP_furan3 + whitecell_count + lymphocyte_pct + monocyte_pct +
##     basophils_pct + neutrophils_pct + race_cat + male + ageyrs +
##     yrssmoke + smokenow + ln_lbxcot
##
##                   Df Sum of Sq    RSS     AIC
## - basophils_pct    1    0.0058 36.769 -2332.5
## - yrssmoke         1    0.0077 36.771 -2332.4
## - smokenow         1    0.0103 36.774 -2332.4
## - POP_dioxin3      1    0.0133 36.777 -2332.3
## - neutrophils_pct  1    0.0153 36.779 -2332.3
## - POP_dioxin1      1    0.0159 36.779 -2332.2
## - POP_PCB6         1    0.0238 36.787 -2332.1
## - monocyte_pct     1    0.0392 36.803 -2331.8
## - POP_PCB10        1    0.0511 36.815 -2331.5
## - race_cat         3    0.2463 37.010 -2331.4
## - POP_PCB8         1    0.0678 36.831 -2331.2
## <none>                         36.763 -2330.6
## - ln_lbxcot        1    0.0993 36.863 -2330.5
## - lymphocyte_pct   1    0.1575 36.921 -2329.3
## - whitecell_count  1    0.1579 36.921 -2329.2
## + POP_dioxin2      1    0.0058 36.758 -2328.7
## + BMI              1    0.0048 36.759 -2328.7
## + POP_PCB3         1    0.0038 36.760 -2328.7
## + POP_furan1       1    0.0015 36.762 -2328.6
```

```
## + POP_PCB11        1    0.0013 36.762 -2328.6
## + POP_PCB9         1    0.0010 36.762 -2328.6
## + POP_PCB7         1    0.0007 36.763 -2328.6
## + POP_furan4       1    0.0004 36.763 -2328.6
## + POP_furan2       1    0.0000 36.763 -2328.6
## - POP_furan3       1    0.2111 36.974 -2328.1
## + edu_cat          3    0.0959 36.668 -2326.6
## - male             1    0.3369 37.100 -2325.5
## - ageyrs           1    4.0169 40.780 -2252.0
##
## Step:  AIC=-2332.46
## length ~ POP_PCB6 + POP_PCB8 + POP_PCB10 + POP_dioxin1 + POP_dioxin3 +
##     POP_furan3 + whitecell_count + lymphocyte_pct + monocyte_pct +
##     neutrophils_pct + race_cat + male + ageyrs + yrssmoke + smokenow +
##     ln_lbxcot
##
##                   Df Sum of Sq    RSS     AIC
## - yrssmoke         1    0.0080 36.777 -2334.3
## - smokenow         1    0.0101 36.779 -2334.2
## - POP_dioxin3      1    0.0127 36.782 -2334.2
## - neutrophils_pct  1    0.0134 36.783 -2334.2
## - POP_dioxin1      1    0.0164 36.786 -2334.1
## - POP_PCB6         1    0.0243 36.794 -2333.9
## - monocyte_pct     1    0.0424 36.812 -2333.6
## - POP_PCB10        1    0.0509 36.820 -2333.4
## - race_cat         3    0.2468 37.016 -2333.3
## - POP_PCB8         1    0.0686 36.838 -2333.0
## <none>                         36.769 -2332.5
## - ln_lbxcot        1    0.0990 36.868 -2332.4
## - whitecell_count  1    0.1567 36.926 -2331.2
## - lymphocyte_pct   1    0.1598 36.929 -2331.1
## + basophils_pct    1    0.0058 36.763 -2330.6
## + POP_dioxin2      1    0.0056 36.764 -2330.6
## + BMI              1    0.0051 36.764 -2330.6
## + POP_PCB3         1    0.0035 36.766 -2330.5
## + POP_PCB9         1    0.0018 36.767 -2330.5
## + POP_furan1       1    0.0014 36.768 -2330.5
## + POP_PCB11        1    0.0013 36.768 -2330.5
## + POP_PCB7         1    0.0007 36.769 -2330.5
## + POP_furan4       1    0.0004 36.769 -2330.5
## + POP_furan2       1    0.0000 36.769 -2330.5
## - POP_furan3       1    0.2117 36.981 -2330.0
## + edu_cat          3    0.0971 36.672 -2328.5
## - male             1    0.3410 37.110 -2327.3
## - ageyrs           1    4.0435 40.813 -2253.4
##
## Step:  AIC=-2334.29
## length ~ POP_PCB6 + POP_PCB8 + POP_PCB10 + POP_dioxin1 + POP_dioxin3 +
##     POP_furan3 + whitecell_count + lymphocyte_pct + monocyte_pct +
##     neutrophils_pct + race_cat + male + ageyrs + smokenow + ln_lbxcot
##
##                   Df Sum of Sq    RSS     AIC
## - neutrophils_pct  1    0.0118 36.789 -2336.0
## - POP_dioxin1      1    0.0120 36.789 -2336.0
```

```
## - smokenow         1    0.0157 36.793 -2335.9
## - POP_dioxin3      1    0.0162 36.794 -2335.9
## - POP_PCB6         1    0.0285 36.806 -2335.7
## - monocyte_pct     1    0.0433 36.821 -2335.4
## - POP_PCB10        1    0.0492 36.826 -2335.2
## - race_cat         3    0.2449 37.022 -2335.1
## - POP_PCB8         1    0.0699 36.847 -2334.8
## - ln_lbxcot        1    0.0933 36.871 -2334.3
## <none>                         36.777 -2334.3
## - whitecell_count  1    0.1641 36.941 -2332.8
## - lymphocyte_pct   1    0.1648 36.942 -2332.8
## + yrssmoke         1    0.0080 36.769 -2332.5
## + basophils_pct    1    0.0061 36.771 -2332.4
## + BMI              1    0.0058 36.771 -2332.4
## + POP_dioxin2      1    0.0048 36.772 -2332.4
## + POP_PCB3         1    0.0035 36.774 -2332.4
## + POP_PCB9         1    0.0026 36.775 -2332.3
## + POP_furan1       1    0.0010 36.776 -2332.3
## + POP_PCB11        1    0.0010 36.776 -2332.3
## + POP_furan4       1    0.0005 36.777 -2332.3
## + POP_PCB7         1    0.0005 36.777 -2332.3
## + POP_furan2       1    0.0000 36.777 -2332.3
## - POP_furan3       1    0.2183 36.996 -2331.7
## + edu_cat          3    0.0999 36.677 -2330.4
## - male             1    0.3504 37.128 -2328.9
## - ageyrs           1    5.5320 42.309 -2227.4
##
## Step:  AIC=-2336.04
## length ~ POP_PCB6 + POP_PCB8 + POP_PCB10 + POP_dioxin1 + POP_dioxin3 +
##     POP_furan3 + whitecell_count + lymphocyte_pct + monocyte_pct +
##     race_cat + male + ageyrs + smokenow + ln_lbxcot
##
##                   Df Sum of Sq    RSS     AIC
## - POP_dioxin1      1    0.0126 36.802 -2337.8
## - smokenow         1    0.0156 36.805 -2337.7
## - POP_dioxin3      1    0.0161 36.805 -2337.7
## - POP_PCB6         1    0.0292 36.818 -2337.4
## - monocyte_pct     1    0.0414 36.830 -2337.2
## - POP_PCB10        1    0.0466 36.836 -2337.1
## - race_cat         3    0.2405 37.030 -2337.0
## - POP_PCB8         1    0.0688 36.858 -2336.6
## - ln_lbxcot        1    0.0926 36.882 -2336.1
## <none>                         36.789 -2336.0
## - lymphocyte_pct   1    0.1592 36.948 -2334.7
## - whitecell_count  1    0.1621 36.951 -2334.6
## + neutrophils_pct  1    0.0118 36.777 -2334.3
## + yrssmoke         1    0.0065 36.783 -2334.2
## + BMI              1    0.0059 36.783 -2334.2
## + POP_dioxin2      1    0.0046 36.784 -2334.1
## + basophils_pct    1    0.0042 36.785 -2334.1
## + POP_PCB9         1    0.0029 36.786 -2334.1
## + POP_PCB3         1    0.0023 36.787 -2334.1
## + POP_PCB11        1    0.0016 36.787 -2334.1
## + POP_furan1       1    0.0011 36.788 -2334.1
```

```
## + POP_PCB7         1    0.0007 36.788 -2334.1
## + POP_furan4       1    0.0006 36.788 -2334.1
## + POP_furan2       1    0.0000 36.789 -2334.0
## - POP_furan3       1    0.2245 37.014 -2333.3
## + edu_cat          3    0.1017 36.687 -2332.2
## - male             1    0.3503 37.139 -2330.7
## - ageyrs           1    5.5202 42.309 -2229.4
##
## Step:  AIC=-2337.77
## length ~ POP_PCB6 + POP_PCB8 + POP_PCB10 + POP_dioxin3 + POP_furan3 +
##     whitecell_count + lymphocyte_pct + monocyte_pct + race_cat +
##     male + ageyrs + smokenow + ln_lbxcot
##
##                   Df Sum of Sq    RSS     AIC
## - smokenow         1    0.0117 36.813 -2339.5
## - POP_PCB6         1    0.0248 36.827 -2339.2
## - monocyte_pct     1    0.0427 36.844 -2338.9
## - race_cat         3    0.2400 37.042 -2338.7
## - POP_PCB10        1    0.0504 36.852 -2338.7
## - POP_PCB8         1    0.0638 36.866 -2338.4
## - ln_lbxcot        1    0.0919 36.894 -2337.8
## <none>                         36.802 -2337.8
## - POP_dioxin3      1    0.1400 36.942 -2336.8
## - whitecell_count  1    0.1572 36.959 -2336.5
## - lymphocyte_pct   1    0.1581 36.960 -2336.4
## + POP_dioxin1      1    0.0126 36.789 -2336.0
## + neutrophils_pct  1    0.0124 36.789 -2336.0
## + POP_dioxin2      1    0.0107 36.791 -2336.0
## + BMI              1    0.0084 36.793 -2335.9
## + basophils_pct    1    0.0045 36.797 -2335.9
## + POP_PCB9         1    0.0031 36.799 -2335.8
## + POP_furan1       1    0.0031 36.799 -2335.8
## + yrssmoke         1    0.0026 36.799 -2335.8
## + POP_PCB3         1    0.0017 36.800 -2335.8
## + POP_furan2       1    0.0008 36.801 -2335.8
## + POP_furan4       1    0.0007 36.801 -2335.8
## + POP_PCB11        1    0.0005 36.801 -2335.8
## + POP_PCB7         1    0.0004 36.801 -2335.8
## - POP_furan3       1    0.2141 37.016 -2335.3
## + edu_cat          3    0.1056 36.696 -2334.0
## - male             1    0.3637 37.165 -2332.1
## - ageyrs           1    5.5792 42.381 -2230.1
##
## Step:  AIC=-2339.52
## length ~ POP_PCB6 + POP_PCB8 + POP_PCB10 + POP_dioxin3 + POP_furan3 +
##     whitecell_count + lymphocyte_pct + monocyte_pct + race_cat +
##     male + ageyrs + ln_lbxcot
##
##                   Df Sum of Sq    RSS     AIC
## - POP_PCB6         1    0.0237 36.837 -2341.0
## - monocyte_pct     1    0.0429 36.856 -2340.6
## - POP_PCB10        1    0.0496 36.863 -2340.5
## - POP_PCB8         1    0.0655 36.879 -2340.1
## - race_cat         3    0.2683 37.082 -2339.9
```

```
## <none>                         36.813 -2339.5
## - POP_dioxin3      1    0.1357 36.949 -2338.7
## - ln_lbxcot        1    0.1481 36.962 -2338.4
## - lymphocyte_pct   1    0.1599 36.973 -2338.2
## - whitecell_count  1    0.1658 36.979 -2338.0
## + neutrophils_pct  1    0.0122 36.801 -2337.8
## + smokenow         1    0.0117 36.802 -2337.8
## + POP_dioxin2      1    0.0090 36.804 -2337.7
## + POP_dioxin1      1    0.0087 36.805 -2337.7
## + BMI              1    0.0074 36.806 -2337.7
## + yrssmoke         1    0.0061 36.807 -2337.7
## + basophils_pct    1    0.0044 36.809 -2337.6
## + POP_furan1       1    0.0027 36.811 -2337.6
## + POP_PCB9         1    0.0026 36.811 -2337.6
## + POP_PCB3         1    0.0014 36.812 -2337.6
## + POP_furan4       1    0.0009 36.813 -2337.5
## + POP_furan2       1    0.0006 36.813 -2337.5
## + POP_PCB11        1    0.0004 36.813 -2337.5
## + POP_PCB7         1    0.0003 36.813 -2337.5
## - POP_furan3       1    0.2238 37.037 -2336.8
## + edu_cat          3    0.1044 36.709 -2335.7
## - male             1    0.3532 37.167 -2334.1
## - ageyrs           1    5.6050 42.418 -2231.4
##
## Step:  AIC=-2341.02
## length ~ POP_PCB8 + POP_PCB10 + POP_dioxin3 + POP_furan3 + whitecell_count +
##     lymphocyte_pct + monocyte_pct + race_cat + male + ageyrs +
##     ln_lbxcot
##
##                   Df Sum of Sq    RSS     AIC
## - monocyte_pct     1    0.0457 36.883 -2342.1
## - POP_PCB8         1    0.0469 36.884 -2342.0
## - POP_PCB10        1    0.0500 36.887 -2342.0
## - race_cat         3    0.2690 37.106 -2341.4
## <none>                         36.837 -2341.0
## - POP_dioxin3      1    0.1254 36.962 -2340.4
## - ln_lbxcot        1    0.1386 36.976 -2340.1
## - lymphocyte_pct   1    0.1596 36.997 -2339.7
## - whitecell_count  1    0.1602 36.997 -2339.7
## + POP_PCB6         1    0.0237 36.813 -2339.5
## + POP_PCB11        1    0.0159 36.821 -2339.4
## + neutrophils_pct  1    0.0128 36.824 -2339.3
## + POP_dioxin2      1    0.0116 36.826 -2339.3
## + smokenow         1    0.0106 36.827 -2339.2
## + yrssmoke         1    0.0097 36.827 -2339.2
## + POP_PCB7         1    0.0065 36.831 -2339.2
## + POP_PCB3         1    0.0056 36.832 -2339.1
## + POP_dioxin1      1    0.0054 36.832 -2339.1
## + basophils_pct    1    0.0048 36.832 -2339.1
## + BMI              1    0.0031 36.834 -2339.1
## + POP_furan1       1    0.0027 36.834 -2339.1
## + POP_furan4       1    0.0016 36.835 -2339.1
## + POP_furan2       1    0.0006 36.837 -2339.0
## + POP_PCB9         1    0.0000 36.837 -2339.0
```

```
## - POP_furan3       1    0.2681 37.105 -2337.4
## + edu_cat          3    0.1028 36.734 -2337.2
## - male             1    0.3869 37.224 -2334.9
## - ageyrs           1    5.5829 42.420 -2233.4
##
## Step:  AIC=-2342.06
## length ~ POP_PCB8 + POP_PCB10 + POP_dioxin3 + POP_furan3 + whitecell_count +
##     lymphocyte_pct + race_cat + male + ageyrs + ln_lbxcot
##
##                   Df Sum of Sq    RSS     AIC
## - POP_PCB8         1    0.0493 36.932 -2343.0
## - POP_PCB10        1    0.0503 36.933 -2343.0
## - race_cat         3    0.2769 37.160 -2342.2
## <none>                         36.883 -2342.1
## - POP_dioxin3      1    0.1228 37.006 -2341.5
## - whitecell_count  1    0.1236 37.006 -2341.5
## - ln_lbxcot        1    0.1337 37.017 -2341.2
## + monocyte_pct     1    0.0457 36.837 -2341.0
## - lymphocyte_pct   1    0.1499 37.033 -2340.9
## + POP_PCB6         1    0.0264 36.856 -2340.6
## + POP_PCB11        1    0.0178 36.865 -2340.4
## + POP_dioxin2      1    0.0136 36.869 -2340.3
## + neutrophils_pct  1    0.0108 36.872 -2340.3
## + yrssmoke         1    0.0108 36.872 -2340.3
## + smokenow         1    0.0107 36.872 -2340.3
## + basophils_pct    1    0.0081 36.875 -2340.2
## + POP_PCB3         1    0.0070 36.876 -2340.2
## + POP_PCB7         1    0.0070 36.876 -2340.2
## + POP_dioxin1      1    0.0061 36.877 -2340.2
## + POP_furan1       1    0.0043 36.879 -2340.2
## + BMI              1    0.0029 36.880 -2340.1
## + POP_furan4       1    0.0025 36.880 -2340.1
## + POP_furan2       1    0.0009 36.882 -2340.1
## + POP_PCB9         1    0.0007 36.882 -2340.1
## - POP_furan3       1    0.2588 37.142 -2338.6
## + edu_cat          3    0.1003 36.783 -2338.2
## - male             1    0.4594 37.342 -2334.4
## - ageyrs           1    5.6413 42.524 -2233.5
##
## Step:  AIC=-2343.02
## length ~ POP_PCB10 + POP_dioxin3 + POP_furan3 + whitecell_count +
##     lymphocyte_pct + race_cat + male + ageyrs + ln_lbxcot
##
##                   Df Sum of Sq    RSS     AIC
## - POP_PCB10        1    0.0171 36.949 -2344.7
## - race_cat         3    0.2559 37.188 -2343.7
## <none>                         36.932 -2343.0
## - whitecell_count  1    0.1136 37.046 -2342.6
## - POP_dioxin3      1    0.1181 37.050 -2342.5
## - ln_lbxcot        1    0.1230 37.055 -2342.4
## + POP_PCB8         1    0.0493 36.883 -2342.1
## + monocyte_pct     1    0.0480 36.884 -2342.0
## - lymphocyte_pct   1    0.1442 37.076 -2342.0
## + POP_PCB9         1    0.0208 36.911 -2341.5
```

```
## + POP_PCB11        1    0.0206 36.912 -2341.4
## + smokenow         1    0.0128 36.919 -2341.3
## + yrssmoke         1    0.0113 36.921 -2341.3
## + POP_dioxin2      1    0.0107 36.921 -2341.2
## + neutrophils_pct  1    0.0094 36.923 -2341.2
## + basophils_pct    1    0.0088 36.923 -2341.2
## + POP_PCB6         1    0.0061 36.926 -2341.2
## + POP_dioxin1      1    0.0042 36.928 -2341.1
## + POP_furan1       1    0.0036 36.928 -2341.1
## + POP_furan4       1    0.0017 36.930 -2341.1
## + POP_furan2       1    0.0012 36.931 -2341.1
## + BMI              1    0.0007 36.931 -2341.0
## + POP_PCB3         1    0.0002 36.932 -2341.0
## + POP_PCB7         1    0.0001 36.932 -2341.0
## + edu_cat          3    0.1059 36.826 -2339.2
## - POP_furan3       1    0.3041 37.236 -2338.7
## - male             1    0.4672 37.399 -2335.2
## - ageyrs           1    6.1231 43.055 -2225.8
##
## Step:  AIC=-2344.66
## length ~ POP_dioxin3 + POP_furan3 + whitecell_count + lymphocyte_pct +
##     race_cat + male + ageyrs + ln_lbxcot
##
##                   Df Sum of Sq    RSS     AIC
## - race_cat         3    0.2666 37.216 -2345.1
## <none>                         36.949 -2344.7
## - whitecell_count  1    0.1181 37.067 -2344.2
## - POP_dioxin3      1    0.1199 37.069 -2344.2
## - ln_lbxcot        1    0.1287 37.078 -2344.0
## + monocyte_pct     1    0.0474 36.902 -2343.7
## - lymphocyte_pct   1    0.1533 37.103 -2343.4
## + POP_PCB11        1    0.0243 36.925 -2343.2
## + POP_PCB10        1    0.0171 36.932 -2343.0
## + POP_PCB8         1    0.0161 36.933 -2343.0
## + smokenow         1    0.0115 36.938 -2342.9
## + POP_PCB6         1    0.0105 36.939 -2342.9
## + yrssmoke         1    0.0092 36.940 -2342.9
## + basophils_pct    1    0.0086 36.941 -2342.8
## + neutrophils_pct  1    0.0083 36.941 -2342.8
## + POP_dioxin1      1    0.0064 36.943 -2342.8
## + POP_dioxin2      1    0.0063 36.943 -2342.8
## + POP_furan1       1    0.0043 36.945 -2342.8
## + POP_PCB9         1    0.0041 36.945 -2342.8
## + BMI              1    0.0029 36.946 -2342.7
## + POP_PCB3         1    0.0028 36.946 -2342.7
## + POP_PCB7         1    0.0021 36.947 -2342.7
## + POP_furan2       1    0.0017 36.947 -2342.7
## + POP_furan4       1    0.0011 36.948 -2342.7
## + edu_cat          3    0.1067 36.842 -2340.9
## - male             1    0.4504 37.400 -2337.2
## - POP_furan3       1    0.6221 37.571 -2333.7
## - ageyrs           1    7.0506 44.000 -2211.0
##
## Step:  AIC=-2345.07
```

```
## length ~ POP_dioxin3 + POP_furan3 + whitecell_count + lymphocyte_pct +
##     male + ageyrs + ln_lbxcot
##
##                   Df Sum of Sq    RSS     AIC
## - POP_dioxin3      1    0.0816 37.297 -2345.4
## - lymphocyte_pct   1    0.0922 37.308 -2345.2
## <none>                         37.216 -2345.1
## + race_cat         3    0.2666 36.949 -2344.7
## + monocyte_pct     1    0.0539 37.162 -2344.2
## - whitecell_count  1    0.1489 37.365 -2344.0
## + smokenow         1    0.0384 37.177 -2343.9
## + POP_PCB11        1    0.0303 37.186 -2343.7
## + POP_PCB10        1    0.0279 37.188 -2343.7
## + POP_PCB6         1    0.0169 37.199 -2343.4
## + yrssmoke         1    0.0120 37.204 -2343.3
## + basophils_pct    1    0.0099 37.206 -2343.3
## + POP_PCB3         1    0.0075 37.208 -2343.2
## + POP_dioxin1      1    0.0053 37.211 -2343.2
## + neutrophils_pct  1    0.0046 37.211 -2343.2
## + POP_PCB8         1    0.0037 37.212 -2343.2
## + POP_PCB7         1    0.0028 37.213 -2343.1
## + POP_PCB9         1    0.0025 37.213 -2343.1
## + POP_furan4       1    0.0016 37.214 -2343.1
## - ln_lbxcot        1    0.1906 37.406 -2343.1
## + BMI              1    0.0007 37.215 -2343.1
## + POP_furan2       1    0.0006 37.215 -2343.1
## + POP_dioxin2      1    0.0005 37.215 -2343.1
## + POP_furan1       1    0.0002 37.216 -2343.1
## + edu_cat          3    0.0891 37.127 -2340.9
## - male             1    0.4415 37.657 -2337.9
## - POP_furan3       1    0.6687 37.885 -2333.2
## - ageyrs           1    7.6776 44.893 -2201.3
##
## Step:  AIC=-2345.37
## length ~ POP_furan3 + whitecell_count + lymphocyte_pct + male +
##     ageyrs + ln_lbxcot
##
##                   Df Sum of Sq    RSS     AIC
## - lymphocyte_pct   1    0.0923 37.390 -2345.4
## <none>                         37.297 -2345.4
## + POP_dioxin3      1    0.0816 37.216 -2345.1
## + POP_dioxin1      1    0.0759 37.221 -2345.0
## + monocyte_pct     1    0.0504 37.247 -2344.4
## - whitecell_count  1    0.1427 37.440 -2344.4
## + race_cat         3    0.2284 37.069 -2344.2
## + POP_PCB10        1    0.0305 37.267 -2344.0
## + smokenow         1    0.0299 37.267 -2344.0
## + POP_PCB11        1    0.0232 37.274 -2343.9
## + POP_dioxin2      1    0.0229 37.275 -2343.8
## + POP_furan1       1    0.0107 37.287 -2343.6
## + POP_PCB6         1    0.0105 37.287 -2343.6
## + basophils_pct    1    0.0075 37.290 -2343.5
## + neutrophils_pct  1    0.0063 37.291 -2343.5
## - ln_lbxcot        1    0.1871 37.485 -2343.5
```

```
## + POP_furan4       1    0.0053 37.292 -2343.5
## + yrssmoke         1    0.0053 37.292 -2343.5
## + POP_furan2       1    0.0046 37.293 -2343.5
## + POP_PCB3         1    0.0038 37.294 -2343.4
## + POP_PCB8         1    0.0030 37.294 -2343.4
## + BMI              1    0.0028 37.295 -2343.4
## + POP_PCB7         1    0.0021 37.295 -2343.4
## + POP_PCB9         1    0.0012 37.296 -2343.4
## + edu_cat          3    0.1094 37.188 -2341.7
## - male             1    0.3886 37.686 -2339.3
## - POP_furan3       1    0.5954 37.893 -2335.1
## - ageyrs           1    8.4062 45.704 -2189.4
##
## Step:  AIC=-2345.45
## length ~ POP_furan3 + whitecell_count + male + ageyrs + ln_lbxcot
##
##                   Df Sum of Sq    RSS     AIC
## - whitecell_count  1    0.0786 37.468 -2345.8
## <none>                         37.390 -2345.4
## + lymphocyte_pct   1    0.0923 37.297 -2345.4
## + POP_dioxin3      1    0.0816 37.308 -2345.2
## + POP_dioxin1      1    0.0762 37.313 -2345.0
## + monocyte_pct     1    0.0408 37.349 -2344.3
## + POP_PCB10        1    0.0391 37.351 -2344.3
## - ln_lbxcot        1    0.1574 37.547 -2344.2
## + smokenow         1    0.0283 37.361 -2344.0
## + POP_PCB11        1    0.0195 37.370 -2343.9
## + POP_dioxin2      1    0.0153 37.374 -2343.8
## + POP_furan1       1    0.0118 37.378 -2343.7
## + POP_PCB6         1    0.0113 37.378 -2343.7
## + basophils_pct    1    0.0095 37.380 -2343.7
## + BMI              1    0.0085 37.381 -2343.6
## + yrssmoke         1    0.0080 37.382 -2343.6
## + POP_furan4       1    0.0069 37.383 -2343.6
## + POP_PCB7         1    0.0066 37.383 -2343.6
## + POP_PCB3         1    0.0045 37.385 -2343.6
## + POP_furan2       1    0.0041 37.386 -2343.5
## + neutrophils_pct  1    0.0036 37.386 -2343.5
## + POP_PCB8         1    0.0017 37.388 -2343.5
## + POP_PCB9         1    0.0003 37.389 -2343.5
## + race_cat         3    0.1772 37.213 -2343.1
## + edu_cat          3    0.1282 37.262 -2342.1
## - male             1    0.3672 37.757 -2339.9
## - POP_furan3       1    0.5871 37.977 -2335.3
## - ageyrs           1    8.3216 45.711 -2191.3
##
## Step:  AIC=-2345.82
## length ~ POP_furan3 + male + ageyrs + ln_lbxcot
##
##                   Df Sum of Sq    RSS     AIC
## <none>                         37.468 -2345.8
## + whitecell_count  1    0.0786 37.390 -2345.4
## + POP_dioxin3      1    0.0766 37.392 -2345.4
## - ln_lbxcot        1    0.1246 37.593 -2345.2
```

```
## + POP_dioxin1      1    0.0654 37.403 -2345.2
## + POP_PCB10        1    0.0417 37.427 -2344.7
## + smokenow         1    0.0397 37.429 -2344.7
## + lymphocyte_pct   1    0.0281 37.440 -2344.4
## + POP_PCB11        1    0.0164 37.452 -2344.2
## + race_cat         3    0.2085 37.260 -2344.2
## + BMI              1    0.0151 37.453 -2344.1
## + yrssmoke         1    0.0135 37.455 -2344.1
## + monocyte_pct     1    0.0113 37.457 -2344.1
## + POP_dioxin2      1    0.0108 37.458 -2344.0
## + POP_PCB6         1    0.0099 37.458 -2344.0
## + POP_furan1       1    0.0098 37.459 -2344.0
## + basophils_pct    1    0.0060 37.462 -2343.9
## + POP_PCB7         1    0.0059 37.462 -2343.9
## + POP_furan4       1    0.0052 37.463 -2343.9
## + POP_PCB3         1    0.0041 37.464 -2343.9
## + neutrophils_pct  1    0.0038 37.464 -2343.9
## + POP_furan2       1    0.0019 37.466 -2343.9
## + POP_PCB9         1    0.0009 37.467 -2343.8
## + POP_PCB8         1    0.0003 37.468 -2343.8
## + edu_cat          3    0.1306 37.338 -2342.5
## - male             1    0.3315 37.800 -2341.0
## - POP_furan3       1    0.6214 38.090 -2335.0
## - ageyrs           1    8.2648 45.733 -2192.9
```

```
summary(step_aic)
```

```
##
## Call:
## lm(formula = length ~ POP_furan3 + male + ageyrs + ln_lbxcot,
##     data = train_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.5045 -0.1510 -0.0268  0.1211  1.1946
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.3701058  0.0232028  59.049  < 2e-16 ***
## POP_furan3   0.0058730  0.0016414   3.578 0.000368 ***
## male1       -0.0421775  0.0161389  -2.613 0.009139 **
## ageyrs      -0.0068654  0.0005261 -13.049  < 2e-16 ***
## ln_lbxcot    0.0034173  0.0021327   1.602 0.109501
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2203 on 772 degrees of freedom
## Multiple R-squared:  0.2223, Adjusted R-squared:  0.2183
## F-statistic: 55.17 on 4 and 772 DF,  p-value: < 2.2e-16
```

```
aic_pred = predict(step_aic, newdata = pollutants[-train_row,])

#RMSE AIC
aic_true = pollutants$length[-train_row]
aic_sd = sum((aic_true - aic_pred)^2)
```

```r
msd_aic = aic_sd / length(aic_true)
rmse_aic = sqrt(msd_aic)
rmse_aic
```

```
## [1] 0.2295366
```

```r
#step wise BIC on training data
step_bic = step(train_model, direction = "both", k = log(nrow(train_x)))
```

```
## Start:  AIC=-2159.93
## length ~ POP_PCB3 + POP_PCB6 + POP_PCB7 + POP_PCB8 + POP_PCB9 +
##     POP_PCB10 + POP_PCB11 + POP_dioxin1 + POP_dioxin2 + POP_dioxin3 +
##     POP_furan1 + POP_furan2 + POP_furan3 + POP_furan4 + whitecell_count +
##     lymphocyte_pct + monocyte_pct + basophils_pct + neutrophils_pct +
##     BMI + edu_cat + race_cat + male + ageyrs + yrssmoke + smokenow +
##     ln_lbxcot
##
##                   Df Sum of Sq    RSS     AIC
## - edu_cat          3   0.09431 36.748 -2177.9
## - race_cat         3   0.25456 36.908 -2174.5
## - POP_PCB9         1   0.00001 36.654 -2166.6
## - POP_PCB7         1   0.00009 36.654 -2166.6
## - POP_PCB3         1   0.00023 36.654 -2166.6
## - POP_furan4       1   0.00030 36.654 -2166.6
## - POP_PCB11        1   0.00048 36.654 -2166.6
## - POP_furan1       1   0.00087 36.654 -2166.6
## - POP_furan2       1   0.00284 36.656 -2166.5
## - POP_dioxin2      1   0.00321 36.657 -2166.5
## - POP_PCB6         1   0.00421 36.658 -2166.5
## - basophils_pct    1   0.00470 36.658 -2166.5
## - yrssmoke         1   0.00501 36.659 -2166.5
## - BMI              1   0.00545 36.659 -2166.5
## - POP_dioxin1      1   0.00584 36.659 -2166.5
## - POP_dioxin3      1   0.01055 36.664 -2166.4
## - smokenow         1   0.01208 36.666 -2166.3
## - neutrophils_pct  1   0.01232 36.666 -2166.3
## - POP_PCB8         1   0.02541 36.679 -2166.1
## - monocyte_pct     1   0.03974 36.693 -2165.7
## - POP_PCB10        1   0.04069 36.694 -2165.7
## - ln_lbxcot        1   0.12280 36.776 -2164.0
## - lymphocyte_pct   1   0.12699 36.781 -2163.9
## - POP_furan3       1   0.12907 36.783 -2163.8
## - whitecell_count  1   0.12929 36.783 -2163.8
## - male             1   0.30812 36.962 -2160.1
## <none>                         36.654 -2159.9
## - ageyrs           1   3.11341 39.767 -2103.2
##
## Step:  AIC=-2177.9
## length ~ POP_PCB3 + POP_PCB6 + POP_PCB7 + POP_PCB8 + POP_PCB9 +
##     POP_PCB10 + POP_PCB11 + POP_dioxin1 + POP_dioxin2 + POP_dioxin3 +
##     POP_furan1 + POP_furan2 + POP_furan3 + POP_furan4 + whitecell_count +
##     lymphocyte_pct + monocyte_pct + basophils_pct + neutrophils_pct +
##     BMI + race_cat + male + ageyrs + yrssmoke + smokenow + ln_lbxcot
##
```

```
##                   Df Sum of Sq    RSS     AIC
## - race_cat         3    0.2452 36.993 -2192.7
## - POP_PCB9         1    0.0002 36.748 -2184.6
## - POP_PCB11        1    0.0003 36.748 -2184.6
## - POP_furan4       1    0.0005 36.748 -2184.5
## - POP_PCB7         1    0.0006 36.748 -2184.5
## - POP_furan1       1    0.0007 36.749 -2184.5
## - POP_furan2       1    0.0017 36.750 -2184.5
## - POP_PCB3         1    0.0018 36.750 -2184.5
## - BMI              1    0.0036 36.752 -2184.5
## - POP_dioxin2      1    0.0047 36.753 -2184.4
## - POP_PCB6         1    0.0054 36.753 -2184.4
## - basophils_pct    1    0.0058 36.754 -2184.4
## - yrssmoke         1    0.0074 36.755 -2184.4
## - POP_dioxin1      1    0.0082 36.756 -2184.4
## - POP_dioxin3      1    0.0100 36.758 -2184.3
## - smokenow         1    0.0111 36.759 -2184.3
## - neutrophils_pct  1    0.0156 36.764 -2184.2
## - POP_PCB8         1    0.0248 36.773 -2184.0
## - monocyte_pct     1    0.0378 36.786 -2183.8
## - POP_PCB10        1    0.0481 36.796 -2183.5
## - ln_lbxcot        1    0.1037 36.852 -2182.4
## - POP_furan3       1    0.1400 36.888 -2181.6
## - whitecell_count  1    0.1433 36.891 -2181.5
## - lymphocyte_pct   1    0.1443 36.892 -2181.5
## - male             1    0.3067 37.055 -2178.1
## <none>                         36.748 -2177.9
## + edu_cat          3    0.0943 36.654 -2159.9
## - ageyrs           1    3.2308 39.979 -2119.1
##
## Step:  AIC=-2192.7
## length ~ POP_PCB3 + POP_PCB6 + POP_PCB7 + POP_PCB8 + POP_PCB9 +
##     POP_PCB10 + POP_PCB11 + POP_dioxin1 + POP_dioxin2 + POP_dioxin3 +
##     POP_furan1 + POP_furan2 + POP_furan3 + POP_furan4 + whitecell_count +
##     lymphocyte_pct + monocyte_pct + basophils_pct + neutrophils_pct +
##     BMI + male + ageyrs + yrssmoke + smokenow + ln_lbxcot
##
##                   Df Sum of Sq    RSS     AIC
## - POP_furan1       1    0.0000 36.993 -2199.3
## - POP_PCB11        1    0.0001 36.993 -2199.3
## - POP_furan4       1    0.0005 36.994 -2199.3
## - BMI              1    0.0006 36.994 -2199.3
## - POP_furan2       1    0.0007 36.994 -2199.3
## - POP_PCB7         1    0.0008 36.994 -2199.3
## - POP_PCB3         1    0.0017 36.995 -2199.3
## - POP_dioxin2      1    0.0017 36.995 -2199.3
## - POP_PCB9         1    0.0057 36.999 -2199.2
## - POP_dioxin3      1    0.0064 37.000 -2199.2
## - POP_PCB6         1    0.0069 37.000 -2199.2
## - yrssmoke         1    0.0077 37.001 -2199.2
## - basophils_pct    1    0.0086 37.002 -2199.2
## - neutrophils_pct  1    0.0116 37.005 -2199.1
## - POP_dioxin1      1    0.0124 37.006 -2199.1
## - POP_PCB8         1    0.0318 37.025 -2198.7
```

```
## - smokenow         1    0.0350 37.028 -2198.6
## - monocyte_pct     1    0.0414 37.034 -2198.5
## - POP_PCB10        1    0.0433 37.036 -2198.4
## - lymphocyte_pct   1    0.0989 37.092 -2197.3
## - POP_furan3       1    0.1217 37.115 -2196.8
## - ln_lbxcot        1    0.1736 37.167 -2195.7
## - whitecell_count  1    0.1799 37.173 -2195.6
## <none>                         36.993 -2192.7
## - male             1    0.3340 37.327 -2192.4
## + race_cat         3    0.2452 36.748 -2177.9
## + edu_cat          3    0.0850 36.908 -2174.5
## - ageyrs           1    3.5038 40.497 -2129.0
##
## Step:  AIC=-2199.35
## length ~ POP_PCB3 + POP_PCB6 + POP_PCB7 + POP_PCB8 + POP_PCB9 +
##     POP_PCB10 + POP_PCB11 + POP_dioxin1 + POP_dioxin2 + POP_dioxin3 +
##     POP_furan2 + POP_furan3 + POP_furan4 + whitecell_count +
##     lymphocyte_pct + monocyte_pct + basophils_pct + neutrophils_pct +
##     BMI + male + ageyrs + yrssmoke + smokenow + ln_lbxcot
##
##                   Df Sum of Sq    RSS     AIC
## - POP_PCB11        1    0.0001 36.993 -2206.0
## - POP_furan4       1    0.0005 36.994 -2206.0
## - BMI              1    0.0006 36.994 -2206.0
## - POP_PCB7         1    0.0008 36.994 -2206.0
## - POP_furan2       1    0.0010 36.994 -2206.0
## - POP_PCB3         1    0.0017 36.995 -2206.0
## - POP_dioxin2      1    0.0019 36.995 -2206.0
## - POP_PCB9         1    0.0057 36.999 -2205.9
## - POP_dioxin3      1    0.0065 37.000 -2205.9
## - POP_PCB6         1    0.0069 37.000 -2205.9
## - yrssmoke         1    0.0077 37.001 -2205.8
## - basophils_pct    1    0.0086 37.002 -2205.8
## - neutrophils_pct  1    0.0116 37.005 -2205.8
## - POP_dioxin1      1    0.0124 37.006 -2205.8
## - POP_PCB8         1    0.0319 37.025 -2205.3
## - smokenow         1    0.0350 37.028 -2205.3
## - monocyte_pct     1    0.0416 37.035 -2205.1
## - POP_PCB10        1    0.0434 37.037 -2205.1
## - lymphocyte_pct   1    0.0992 37.092 -2203.9
## - POP_furan3       1    0.1217 37.115 -2203.5
## - ln_lbxcot        1    0.1736 37.167 -2202.4
## - whitecell_count  1    0.1808 37.174 -2202.2
## <none>                         36.993 -2199.3
## - male             1    0.3346 37.328 -2199.0
## + POP_furan1       1    0.0000 36.993 -2192.7
## + race_cat         3    0.2445 36.749 -2184.5
## + edu_cat          3    0.0849 36.908 -2181.2
## - ageyrs           1    3.5092 40.502 -2135.6
##
## Step:  AIC=-2206.01
## length ~ POP_PCB3 + POP_PCB6 + POP_PCB7 + POP_PCB8 + POP_PCB9 +
##     POP_PCB10 + POP_dioxin1 + POP_dioxin2 + POP_dioxin3 + POP_furan2 +
##     POP_furan3 + POP_furan4 + whitecell_count + lymphocyte_pct +
```

```
##     monocyte_pct + basophils_pct + neutrophils_pct + BMI + male +
##     ageyrs + yrssmoke + smokenow + ln_lbxcot
##
##                    Df Sum of Sq    RSS     AIC
## - POP_furan4        1    0.0005 36.994 -2212.7
## - BMI               1    0.0006 36.994 -2212.7
## - POP_furan2        1    0.0009 36.994 -2212.6
## - POP_PCB7          1    0.0010 36.994 -2212.6
## - POP_dioxin2       1    0.0020 36.995 -2212.6
## - POP_PCB3          1    0.0025 36.996 -2212.6
## - POP_PCB9          1    0.0067 37.000 -2212.5
## - POP_dioxin3       1    0.0068 37.000 -2212.5
## - yrssmoke          1    0.0077 37.001 -2212.5
## - basophils_pct     1    0.0088 37.002 -2212.5
## - neutrophils_pct   1    0.0122 37.005 -2212.4
## - POP_dioxin1       1    0.0123 37.006 -2212.4
## - POP_PCB6          1    0.0210 37.014 -2212.2
## - smokenow          1    0.0349 37.028 -2211.9
## - POP_PCB8          1    0.0414 37.035 -2211.8
## - monocyte_pct      1    0.0416 37.035 -2211.8
## - POP_PCB10         1    0.0482 37.041 -2211.7
## - lymphocyte_pct    1    0.0996 37.093 -2210.6
## - POP_furan3        1    0.1257 37.119 -2210.0
## - ln_lbxcot         1    0.1735 37.167 -2209.0
## - whitecell_count   1    0.1811 37.174 -2208.9
## <none>                          36.993 -2206.0
## - male              1    0.3344 37.328 -2205.7
## + POP_PCB11         1    0.0001 36.993 -2199.3
## + POP_furan1        1    0.0000 36.993 -2199.3
## + race_cat          3    0.2444 36.749 -2191.2
## + edu_cat           3    0.0849 36.908 -2187.8
## - ageyrs            1    3.5096 40.503 -2142.2
##
## Step:  AIC=-2212.65
## length ~ POP_PCB3 + POP_PCB6 + POP_PCB7 + POP_PCB8 + POP_PCB9 +
##     POP_PCB10 + POP_dioxin1 + POP_dioxin2 + POP_dioxin3 + POP_furan2 +
##     POP_furan3 + whitecell_count + lymphocyte_pct + monocyte_pct +
##     basophils_pct + neutrophils_pct + BMI + male + ageyrs + yrssmoke +
##     smokenow + ln_lbxcot
##
##                    Df Sum of Sq    RSS     AIC
## - BMI               1    0.0006 36.994 -2219.3
## - POP_PCB7          1    0.0010 36.995 -2219.3
## - POP_dioxin2       1    0.0020 36.996 -2219.3
## - POP_furan2        1    0.0022 36.996 -2219.3
## - POP_PCB3          1    0.0026 36.996 -2219.2
## - POP_dioxin3       1    0.0063 37.000 -2219.2
## - POP_PCB9          1    0.0067 37.000 -2219.2
## - yrssmoke          1    0.0076 37.001 -2219.2
## - basophils_pct     1    0.0088 37.003 -2219.1
## - neutrophils_pct   1    0.0121 37.006 -2219.1
## - POP_dioxin1       1    0.0128 37.007 -2219.0
## - POP_PCB6          1    0.0209 37.015 -2218.9
## - smokenow          1    0.0348 37.029 -2218.6
```

```
## - monocyte_pct     1    0.0413 37.035 -2218.4
## - POP_PCB8         1    0.0416 37.035 -2218.4
## - POP_PCB10        1    0.0499 37.044 -2218.3
## - lymphocyte_pct   1    0.0991 37.093 -2217.2
## - POP_furan3       1    0.1270 37.121 -2216.6
## - ln_lbxcot        1    0.1733 37.167 -2215.7
## - whitecell_count  1    0.1807 37.174 -2215.5
## <none>                         36.994 -2212.7
## - male             1    0.3344 37.328 -2212.3
## + POP_furan4       1    0.0005 36.993 -2206.0
## + POP_PCB11        1    0.0001 36.994 -2206.0
## + POP_furan1       1    0.0000 36.994 -2206.0
## + race_cat         3    0.2442 36.750 -2197.8
## + edu_cat          3    0.0846 36.909 -2194.5
## - ageyrs           1    3.6342 40.628 -2146.5
##
## Step:  AIC=-2219.29
## length ~ POP_PCB3 + POP_PCB6 + POP_PCB7 + POP_PCB8 + POP_PCB9 +
##     POP_PCB10 + POP_dioxin1 + POP_dioxin2 + POP_dioxin3 + POP_furan2 +
##     POP_furan3 + whitecell_count + lymphocyte_pct + monocyte_pct +
##     basophils_pct + neutrophils_pct + male + ageyrs + yrssmoke +
##     smokenow + ln_lbxcot
##
##                   Df Sum of Sq    RSS     AIC
## - POP_PCB7         1    0.0012 36.995 -2225.9
## - POP_dioxin2      1    0.0021 36.996 -2225.9
## - POP_furan2       1    0.0022 36.997 -2225.9
## - POP_PCB3         1    0.0028 36.997 -2225.9
## - POP_dioxin3      1    0.0060 37.000 -2225.8
## - POP_PCB9         1    0.0066 37.001 -2225.8
## - yrssmoke         1    0.0079 37.002 -2225.8
## - basophils_pct    1    0.0090 37.003 -2225.8
## - neutrophils_pct  1    0.0123 37.007 -2225.7
## - POP_dioxin1      1    0.0138 37.008 -2225.7
## - POP_PCB6         1    0.0208 37.015 -2225.5
## - smokenow         1    0.0345 37.029 -2225.2
## - POP_PCB8         1    0.0411 37.035 -2225.1
## - monocyte_pct     1    0.0413 37.036 -2225.1
## - POP_PCB10        1    0.0530 37.047 -2224.8
## - lymphocyte_pct   1    0.1024 37.097 -2223.8
## - POP_furan3       1    0.1281 37.122 -2223.3
## - ln_lbxcot        1    0.1736 37.168 -2222.3
## - whitecell_count  1    0.1871 37.181 -2222.0
## <none>                         36.994 -2219.3
## - male             1    0.3406 37.335 -2218.8
## + BMI              1    0.0006 36.994 -2212.7
## + POP_furan4       1    0.0005 36.994 -2212.7
## + POP_PCB11        1    0.0001 36.994 -2212.6
## + POP_furan1       1    0.0000 36.994 -2212.6
## + race_cat         3    0.2411 36.753 -2204.4
## + edu_cat          3    0.0832 36.911 -2201.1
## - ageyrs           1    3.6431 40.637 -2153.0
##
## Step:  AIC=-2225.92
```

```
## length ~ POP_PCB3 + POP_PCB6 + POP_PCB8 + POP_PCB9 + POP_PCB10 +
##     POP_dioxin1 + POP_dioxin2 + POP_dioxin3 + POP_furan2 + POP_furan3 +
##     whitecell_count + lymphocyte_pct + monocyte_pct + basophils_pct +
##     neutrophils_pct + male + ageyrs + yrssmoke + smokenow + ln_lbxcot
##
##                    Df Sum of Sq    RSS     AIC
## - POP_dioxin2       1    0.0025 36.998 -2232.5
## - POP_furan2        1    0.0027 36.998 -2232.5
## - POP_PCB3          1    0.0054 37.001 -2232.5
## - POP_dioxin3       1    0.0060 37.002 -2232.4
## - POP_PCB9          1    0.0062 37.002 -2232.4
## - yrssmoke          1    0.0076 37.003 -2232.4
## - basophils_pct     1    0.0089 37.004 -2232.4
## - neutrophils_pct   1    0.0128 37.008 -2232.3
## - POP_dioxin1       1    0.0133 37.009 -2232.3
## - POP_PCB6          1    0.0197 37.015 -2232.2
## - smokenow          1    0.0345 37.030 -2231.9
## - POP_PCB8          1    0.0412 37.037 -2231.7
## - monocyte_pct      1    0.0420 37.037 -2231.7
## - POP_PCB10         1    0.0531 37.049 -2231.5
## - lymphocyte_pct    1    0.1013 37.097 -2230.4
## - POP_furan3        1    0.1282 37.124 -2229.9
## - ln_lbxcot         1    0.1741 37.170 -2228.9
## - whitecell_count   1    0.1863 37.182 -2228.7
## <none>                          36.995 -2225.9
## - male              1    0.3471 37.343 -2225.3
## + POP_PCB7          1    0.0012 36.994 -2219.3
## + BMI               1    0.0007 36.995 -2219.3
## + POP_furan4        1    0.0005 36.995 -2219.3
## + POP_PCB11         1    0.0003 36.995 -2219.3
## + POP_furan1        1    0.0000 36.995 -2219.3
## + race_cat          3    0.2421 36.753 -2211.1
## + edu_cat           3    0.0810 36.914 -2207.7
## - ageyrs            1    3.8565 40.852 -2155.5
##
## Step:  AIC=-2232.53
## length ~ POP_PCB3 + POP_PCB6 + POP_PCB8 + POP_PCB9 + POP_PCB10 +
##     POP_dioxin1 + POP_dioxin3 + POP_furan2 + POP_furan3 + whitecell_count +
##     lymphocyte_pct + monocyte_pct + basophils_pct + neutrophils_pct +
##     male + ageyrs + yrssmoke + smokenow + ln_lbxcot
##
##                    Df Sum of Sq    RSS     AIC
## - POP_furan2        1    0.0015 37.000 -2239.2
## - POP_PCB3          1    0.0056 37.004 -2239.1
## - POP_PCB9          1    0.0059 37.004 -2239.1
## - POP_dioxin3       1    0.0063 37.004 -2239.1
## - yrssmoke          1    0.0071 37.005 -2239.0
## - basophils_pct     1    0.0085 37.007 -2239.0
## - neutrophils_pct   1    0.0125 37.011 -2238.9
## - POP_dioxin1       1    0.0181 37.016 -2238.8
## - POP_PCB6          1    0.0218 37.020 -2238.7
## - smokenow          1    0.0342 37.032 -2238.5
## - POP_PCB8          1    0.0409 37.039 -2238.3
## - monocyte_pct      1    0.0428 37.041 -2238.3
```

```
## - POP_PCB10        1    0.0506 37.049 -2238.1
## - lymphocyte_pct   1    0.0993 37.097 -2237.1
## - POP_furan3       1    0.1260 37.124 -2236.5
## - ln_lbxcot        1    0.1717 37.170 -2235.6
## - whitecell_count  1    0.1851 37.183 -2235.3
## <none>                         36.998 -2232.5
## - male             1    0.3446 37.343 -2232.0
## + POP_dioxin2      1    0.0025 36.995 -2225.9
## + POP_PCB7         1    0.0017 36.996 -2225.9
## + BMI              1    0.0008 36.997 -2225.9
## + POP_furan4       1    0.0006 36.997 -2225.9
## + POP_PCB11        1    0.0004 36.998 -2225.9
## + POP_furan1       1    0.0003 36.998 -2225.9
## + race_cat         3    0.2387 36.759 -2217.6
## + edu_cat          3    0.0806 36.917 -2214.3
## - ageyrs           1    4.3790 41.377 -2152.3
##
## Step:  AIC=-2239.15
## length ~ POP_PCB3 + POP_PCB6 + POP_PCB8 + POP_PCB9 + POP_PCB10 +
##     POP_dioxin1 + POP_dioxin3 + POP_furan3 + whitecell_count +
##     lymphocyte_pct + monocyte_pct + basophils_pct + neutrophils_pct +
##     male + ageyrs + yrssmoke + smokenow + ln_lbxcot
##
##                   Df Sum of Sq    RSS     AIC
## - POP_PCB3         1    0.0058 37.005 -2245.7
## - POP_dioxin3      1    0.0061 37.006 -2245.7
## - POP_PCB9         1    0.0071 37.007 -2245.7
## - yrssmoke         1    0.0074 37.007 -2245.7
## - basophils_pct    1    0.0086 37.008 -2245.6
## - neutrophils_pct  1    0.0126 37.012 -2245.5
## - POP_dioxin1      1    0.0168 37.016 -2245.4
## - POP_PCB6         1    0.0213 37.021 -2245.4
## - smokenow         1    0.0344 37.034 -2245.1
## - POP_PCB8         1    0.0423 37.042 -2244.9
## - monocyte_pct     1    0.0423 37.042 -2244.9
## - POP_PCB10        1    0.0499 37.049 -2244.8
## - lymphocyte_pct   1    0.1005 37.100 -2243.7
## - ln_lbxcot        1    0.1763 37.176 -2242.1
## - POP_furan3       1    0.1865 37.186 -2241.9
## - whitecell_count  1    0.1868 37.186 -2241.9
## <none>                         37.000 -2239.2
## - male             1    0.3432 37.343 -2238.6
## + POP_PCB7         1    0.0020 36.998 -2232.5
## + POP_furan4       1    0.0016 36.998 -2232.5
## + POP_furan2       1    0.0015 36.998 -2232.5
## + POP_dioxin2      1    0.0013 36.998 -2232.5
## + BMI              1    0.0008 36.999 -2232.5
## + POP_PCB11        1    0.0004 36.999 -2232.5
## + POP_furan1       1    0.0003 36.999 -2232.5
## + race_cat         3    0.2402 36.759 -2224.2
## + edu_cat          3    0.0803 36.919 -2220.9
## - ageyrs           1    4.3782 41.378 -2158.9
##
## Step:  AIC=-2245.68
```

```
## length ~ POP_PCB6 + POP_PCB8 + POP_PCB9 + POP_PCB10 + POP_dioxin1 +
##     POP_dioxin3 + POP_furan3 + whitecell_count + lymphocyte_pct +
##     monocyte_pct + basophils_pct + neutrophils_pct + male + ageyrs +
##     yrssmoke + smokenow + ln_lbxcot
##
##                    Df Sum of Sq    RSS     AIC
## - POP_PCB9          1    0.0044 37.010 -2252.2
## - yrssmoke          1    0.0070 37.012 -2252.2
## - POP_dioxin3       1    0.0071 37.012 -2252.2
## - basophils_pct     1    0.0075 37.013 -2252.2
## - neutrophils_pct   1    0.0108 37.016 -2252.1
## - POP_dioxin1       1    0.0155 37.021 -2252.0
## - POP_PCB6          1    0.0158 37.021 -2252.0
## - smokenow          1    0.0338 37.039 -2251.6
## - POP_PCB8          1    0.0392 37.045 -2251.5
## - monocyte_pct      1    0.0427 37.048 -2251.4
## - POP_PCB10         1    0.0461 37.051 -2251.4
## - lymphocyte_pct    1    0.0985 37.104 -2250.3
## - ln_lbxcot         1    0.1751 37.180 -2248.7
## - whitecell_count   1    0.1862 37.191 -2248.4
## - POP_furan3        1    0.1866 37.192 -2248.4
## <none>                          37.005 -2245.7
## - male              1    0.3425 37.348 -2245.2
## + POP_PCB3          1    0.0058 37.000 -2239.2
## + POP_PCB7          1    0.0052 37.000 -2239.1
## + POP_PCB11         1    0.0030 37.002 -2239.1
## + POP_furan4        1    0.0019 37.003 -2239.1
## + POP_furan2        1    0.0018 37.004 -2239.1
## + POP_dioxin2       1    0.0014 37.004 -2239.1
## + BMI               1    0.0014 37.004 -2239.1
## + POP_furan1        1    0.0004 37.005 -2239.0
## + race_cat          3    0.2429 36.762 -2230.8
## + edu_cat           3    0.0817 36.924 -2227.4
## - ageyrs            1    4.3823 41.388 -2165.4
##
## Step:  AIC=-2252.25
## length ~ POP_PCB6 + POP_PCB8 + POP_PCB10 + POP_dioxin1 + POP_dioxin3 +
##     POP_furan3 + whitecell_count + lymphocyte_pct + monocyte_pct +
##     basophils_pct + neutrophils_pct + male + ageyrs + yrssmoke +
##     smokenow + ln_lbxcot
##
##                    Df Sum of Sq    RSS     AIC
## - yrssmoke          1    0.0059 37.016 -2258.8
## - basophils_pct     1    0.0062 37.016 -2258.8
## - POP_dioxin3       1    0.0070 37.017 -2258.8
## - neutrophils_pct   1    0.0100 37.020 -2258.7
## - POP_dioxin1       1    0.0148 37.025 -2258.6
## - POP_PCB6          1    0.0259 37.036 -2258.4
## - smokenow          1    0.0366 37.046 -2258.1
## - POP_PCB8          1    0.0446 37.054 -2258.0
## - monocyte_pct      1    0.0460 37.056 -2257.9
## - POP_PCB10         1    0.0519 37.062 -2257.8
## - lymphocyte_pct    1    0.0943 37.104 -2256.9
## - ln_lbxcot         1    0.1825 37.192 -2255.1
```

```
## - whitecell_count  1    0.1837 37.193 -2255.1
## - POP_furan3       1    0.2035 37.213 -2254.6
## <none>                         37.010 -2252.2
## - male             1    0.3467 37.356 -2251.7
## + POP_PCB9         1    0.0044 37.005 -2245.7
## + POP_PCB11        1    0.0037 37.006 -2245.7
## + POP_PCB7         1    0.0034 37.006 -2245.7
## + POP_PCB3         1    0.0032 37.007 -2245.7
## + POP_furan2       1    0.0026 37.007 -2245.7
## + POP_furan4       1    0.0023 37.007 -2245.6
## + POP_furan1       1    0.0011 37.009 -2245.6
## + BMI              1    0.0010 37.009 -2245.6
## + POP_dioxin2      1    0.0010 37.009 -2245.6
## + race_cat         3    0.2463 36.763 -2237.5
## + edu_cat          3    0.0790 36.931 -2233.9
## - ageyrs           1    4.4219 41.432 -2171.2
##
## Step:  AIC=-2258.78
## length ~ POP_PCB6 + POP_PCB8 + POP_PCB10 + POP_dioxin1 + POP_dioxin3 +
##     POP_furan3 + whitecell_count + lymphocyte_pct + monocyte_pct +
##     basophils_pct + neutrophils_pct + male + ageyrs + smokenow +
##     ln_lbxcot
##
##                   Df Sum of Sq    RSS     AIC
## - basophils_pct    1    0.0065 37.022 -2265.3
## - neutrophils_pct  1    0.0089 37.025 -2265.2
## - POP_dioxin3      1    0.0092 37.025 -2265.2
## - POP_dioxin1      1    0.0112 37.027 -2265.2
## - POP_PCB6         1    0.0295 37.045 -2264.8
## - POP_PCB8         1    0.0456 37.061 -2264.5
## - smokenow         1    0.0465 37.062 -2264.5
## - monocyte_pct     1    0.0467 37.062 -2264.4
## - POP_PCB10        1    0.0504 37.066 -2264.4
## - lymphocyte_pct   1    0.0978 37.113 -2263.4
## - ln_lbxcot        1    0.1770 37.193 -2261.7
## - whitecell_count  1    0.1908 37.206 -2261.4
## - POP_furan3       1    0.2093 37.225 -2261.1
## <none>                         37.016 -2258.8
## - male             1    0.3550 37.371 -2258.0
## + yrssmoke         1    0.0059 37.010 -2252.2
## + POP_PCB9         1    0.0033 37.012 -2252.2
## + POP_PCB11        1    0.0033 37.012 -2252.2
## + POP_PCB3         1    0.0032 37.012 -2252.2
## + POP_PCB7         1    0.0031 37.013 -2252.2
## + POP_furan2       1    0.0027 37.013 -2252.2
## + POP_furan4       1    0.0022 37.013 -2252.2
## + POP_furan1       1    0.0014 37.014 -2252.2
## + BMI              1    0.0013 37.014 -2252.2
## + POP_dioxin2      1    0.0008 37.015 -2252.1
## + race_cat         3    0.2445 36.771 -2244.0
## + edu_cat          3    0.0808 36.935 -2240.5
## - ageyrs           1    6.0075 43.023 -2148.6
##
## Step:  AIC=-2265.3
```

```
## length ~ POP_PCB6 + POP_PCB8 + POP_PCB10 + POP_dioxin1 + POP_dioxin3 +
##     POP_furan3 + whitecell_count + lymphocyte_pct + monocyte_pct +
##     neutrophils_pct + male + ageyrs + smokenow + ln_lbxcot
##
##                    Df Sum of Sq    RSS     AIC
## - neutrophils_pct   1    0.0074 37.030 -2271.8
## - POP_dioxin3       1    0.0087 37.031 -2271.8
## - POP_dioxin1       1    0.0116 37.034 -2271.7
## - POP_PCB6          1    0.0303 37.052 -2271.3
## - POP_PCB8          1    0.0462 37.068 -2271.0
## - smokenow          1    0.0466 37.069 -2271.0
## - monocyte_pct      1    0.0503 37.072 -2270.9
## - POP_PCB10         1    0.0503 37.072 -2270.9
## - lymphocyte_pct    1    0.0997 37.122 -2269.9
## - ln_lbxcot         1    0.1766 37.199 -2268.2
## - whitecell_count   1    0.1896 37.212 -2268.0
## - POP_furan3        1    0.2104 37.233 -2267.6
## <none>                          37.022 -2265.3
## - male              1    0.3599 37.382 -2264.4
## + basophils_pct     1    0.0065 37.016 -2258.8
## + yrssmoke          1    0.0062 37.016 -2258.8
## + POP_PCB11         1    0.0033 37.019 -2258.7
## + POP_PCB7          1    0.0029 37.019 -2258.7
## + POP_PCB3          1    0.0028 37.019 -2258.7
## + POP_furan2        1    0.0026 37.020 -2258.7
## + POP_furan4        1    0.0022 37.020 -2258.7
## + POP_PCB9          1    0.0022 37.020 -2258.7
## + POP_furan1        1    0.0016 37.021 -2258.7
## + BMI               1    0.0014 37.021 -2258.7
## + POP_dioxin2       1    0.0006 37.022 -2258.7
## + race_cat          3    0.2449 36.777 -2250.5
## + edu_cat           3    0.0819 36.940 -2247.1
## - ageyrs            1    6.0703 43.092 -2154.0
##
## Step:  AIC=-2271.8
## length ~ POP_PCB6 + POP_PCB8 + POP_PCB10 + POP_dioxin1 + POP_dioxin3 +
##     POP_furan3 + whitecell_count + lymphocyte_pct + monocyte_pct +
##     male + ageyrs + smokenow + ln_lbxcot
##
##                    Df Sum of Sq    RSS     AIC
## - POP_dioxin3       1    0.0086 37.038 -2278.3
## - POP_dioxin1       1    0.0122 37.042 -2278.2
## - POP_PCB6          1    0.0309 37.060 -2277.8
## - POP_PCB8          1    0.0457 37.075 -2277.5
## - smokenow          1    0.0463 37.076 -2277.5
## - monocyte_pct      1    0.0486 37.078 -2277.4
## - POP_PCB10         1    0.0486 37.078 -2277.4
## - lymphocyte_pct    1    0.0967 37.126 -2276.4
## - ln_lbxcot         1    0.1757 37.205 -2274.8
## - whitecell_count   1    0.1877 37.217 -2274.5
## - POP_furan3        1    0.2156 37.245 -2273.9
## <none>                          37.030 -2271.8
## - male              1    0.3604 37.390 -2270.9
## + neutrophils_pct   1    0.0074 37.022 -2265.3
```

```
## + yrssmoke         1    0.0051 37.024 -2265.2
## + basophils_pct    1    0.0050 37.025 -2265.2
## + POP_PCB11        1    0.0041 37.025 -2265.2
## + POP_PCB7         1    0.0032 37.026 -2265.2
## + POP_furan2       1    0.0025 37.027 -2265.2
## + POP_furan4       1    0.0021 37.027 -2265.2
## + POP_PCB3         1    0.0019 37.028 -2265.2
## + POP_PCB9         1    0.0019 37.028 -2265.2
## + BMI              1    0.0015 37.028 -2265.2
## + POP_furan1       1    0.0015 37.028 -2265.2
## + POP_dioxin2      1    0.0005 37.029 -2265.2
## + race_cat         3    0.2405 36.789 -2256.9
## + edu_cat          3    0.0837 36.946 -2253.6
## - ageyrs           1    6.0663 43.096 -2160.6
##
## Step:  AIC=-2278.27
## length ~ POP_PCB6 + POP_PCB8 + POP_PCB10 + POP_dioxin1 + POP_furan3 +
##     whitecell_count + lymphocyte_pct + monocyte_pct + male +
##     ageyrs + smokenow + ln_lbxcot
##
##                   Df Sum of Sq    RSS     AIC
## - POP_PCB6         1    0.0324 37.071 -2284.2
## - POP_PCB10        1    0.0464 37.085 -2283.9
## - monocyte_pct     1    0.0473 37.085 -2283.9
## - POP_PCB8         1    0.0480 37.086 -2283.9
## - smokenow         1    0.0502 37.088 -2283.9
## - lymphocyte_pct   1    0.0969 37.135 -2282.9
## - POP_dioxin1      1    0.1051 37.143 -2282.7
## - ln_lbxcot        1    0.1725 37.211 -2281.3
## - whitecell_count  1    0.1906 37.229 -2280.9
## - POP_furan3       1    0.2171 37.255 -2280.4
## <none>                         37.038 -2278.3
## - male             1    0.3518 37.390 -2277.6
## + POP_dioxin3      1    0.0086 37.030 -2271.8
## + neutrophils_pct  1    0.0073 37.031 -2271.8
## + yrssmoke         1    0.0072 37.031 -2271.8
## + POP_PCB11        1    0.0058 37.032 -2271.7
## + basophils_pct    1    0.0045 37.034 -2271.7
## + POP_PCB7         1    0.0034 37.035 -2271.7
## + POP_PCB3         1    0.0026 37.036 -2271.7
## + POP_furan2       1    0.0023 37.036 -2271.7
## + POP_PCB9         1    0.0017 37.036 -2271.7
## + BMI              1    0.0010 37.037 -2271.6
## + POP_furan1       1    0.0008 37.037 -2271.6
## + POP_dioxin2      1    0.0007 37.037 -2271.6
## + POP_furan4       1    0.0005 37.038 -2271.6
## + race_cat         3    0.2330 36.805 -2263.2
## + edu_cat          3    0.0836 36.955 -2260.1
## - ageyrs           1    6.1217 43.160 -2166.1
##
## Step:  AIC=-2284.25
## length ~ POP_PCB8 + POP_PCB10 + POP_dioxin1 + POP_furan3 + whitecell_count +
##     lymphocyte_pct + monocyte_pct + male + ageyrs + smokenow +
##     ln_lbxcot
```

```
##
##                   Df Sum of Sq    RSS     AIC
## - POP_PCB8         1    0.0278 37.098 -2290.3
## - smokenow         1    0.0463 37.117 -2289.9
## - POP_PCB10        1    0.0475 37.118 -2289.9
## - monocyte_pct     1    0.0509 37.121 -2289.8
## - POP_dioxin1      1    0.0891 37.160 -2289.0
## - lymphocyte_pct   1    0.0963 37.167 -2288.9
## - ln_lbxcot        1    0.1617 37.232 -2287.5
## - whitecell_count  1    0.1832 37.254 -2287.1
## - POP_furan3       1    0.2611 37.332 -2285.4
## <none>                         37.071 -2284.2
## - male             1    0.3902 37.461 -2282.8
## + POP_PCB11        1    0.0331 37.037 -2278.3
## + POP_PCB6         1    0.0324 37.038 -2278.3
## + yrssmoke         1    0.0115 37.059 -2277.8
## + POP_PCB9         1    0.0103 37.060 -2277.8
## + POP_dioxin3      1    0.0101 37.060 -2277.8
## + neutrophils_pct  1    0.0080 37.063 -2277.8
## + POP_PCB3         1    0.0067 37.064 -2277.7
## + basophils_pct    1    0.0050 37.066 -2277.7
## + POP_PCB7         1    0.0039 37.067 -2277.7
## + POP_dioxin2      1    0.0025 37.068 -2277.7
## + POP_furan2       1    0.0019 37.069 -2277.6
## + POP_furan1       1    0.0006 37.070 -2277.6
## + POP_furan4       1    0.0001 37.070 -2277.6
## + BMI              1    0.0000 37.071 -2277.6
## + race_cat         3    0.2341 36.836 -2269.2
## + edu_cat          3    0.0829 36.988 -2266.0
## - ageyrs           1    6.0984 43.169 -2172.6
##
## Step:  AIC=-2290.32
## length ~ POP_PCB10 + POP_dioxin1 + POP_furan3 + whitecell_count +
##     lymphocyte_pct + monocyte_pct + male + ageyrs + smokenow +
##     ln_lbxcot
##
##                   Df Sum of Sq    RSS     AIC
## - POP_PCB10        1    0.0241 37.122 -2296.5
## - smokenow         1    0.0478 37.146 -2296.0
## - monocyte_pct     1    0.0524 37.151 -2295.9
## - POP_dioxin1      1    0.0858 37.184 -2295.2
## - lymphocyte_pct   1    0.0957 37.194 -2295.0
## - ln_lbxcot        1    0.1582 37.257 -2293.7
## - whitecell_count  1    0.1734 37.272 -2293.3
## - POP_furan3       1    0.2930 37.391 -2290.9
## <none>                         37.098 -2290.3
## - male             1    0.3956 37.494 -2288.7
## + POP_PCB11        1    0.0354 37.063 -2284.4
## + POP_PCB8         1    0.0278 37.071 -2284.2
## + POP_PCB6         1    0.0122 37.086 -2283.9
## + POP_dioxin3      1    0.0115 37.087 -2283.9
## + yrssmoke         1    0.0112 37.087 -2283.9
## + neutrophils_pct  1    0.0073 37.091 -2283.8
## + basophils_pct    1    0.0053 37.093 -2283.8
```

```
## + POP_PCB9         1    0.0030 37.095 -2283.7
## + POP_dioxin2      1    0.0022 37.096 -2283.7
## + POP_furan2       1    0.0013 37.097 -2283.7
## + POP_PCB3         1    0.0007 37.098 -2283.7
## + POP_furan1       1    0.0005 37.098 -2283.7
## + BMI              1    0.0003 37.098 -2283.7
## + POP_furan4       1    0.0002 37.098 -2283.7
## + POP_PCB7         1    0.0001 37.098 -2283.7
## + race_cat         3    0.2134 36.885 -2274.8
## + edu_cat          3    0.0852 37.013 -2272.1
## - ageyrs           1    6.4996 43.598 -2171.5
##
## Step:  AIC=-2296.47
## length ~ POP_dioxin1 + POP_furan3 + whitecell_count + lymphocyte_pct +
##     monocyte_pct + male + ageyrs + smokenow + ln_lbxcot
##
##                   Df Sum of Sq    RSS     AIC
## - smokenow         1    0.0469 37.169 -2302.2
## - monocyte_pct     1    0.0513 37.174 -2302.1
## - POP_dioxin1      1    0.0954 37.218 -2301.1
## - lymphocyte_pct   1    0.1038 37.226 -2301.0
## - ln_lbxcot        1    0.1624 37.285 -2299.7
## - whitecell_count  1    0.1806 37.303 -2299.4
## <none>                         37.122 -2296.5
## - male             1    0.3724 37.495 -2295.4
## + POP_PCB11        1    0.0418 37.081 -2290.7
## + POP_PCB10        1    0.0241 37.098 -2290.3
## + POP_PCB6         1    0.0199 37.103 -2290.2
## + yrssmoke         1    0.0096 37.113 -2290.0
## + POP_dioxin3      1    0.0086 37.114 -2290.0
## + neutrophils_pct  1    0.0063 37.116 -2289.9
## + POP_PCB3         1    0.0056 37.117 -2289.9
## + basophils_pct    1    0.0053 37.117 -2289.9
## + POP_PCB8         1    0.0044 37.118 -2289.9
## + POP_PCB7         1    0.0026 37.120 -2289.9
## + POP_furan2       1    0.0012 37.121 -2289.8
## + POP_furan4       1    0.0008 37.122 -2289.8
## + POP_PCB9         1    0.0007 37.122 -2289.8
## + POP_furan1       1    0.0007 37.122 -2289.8
## + BMI              1    0.0002 37.122 -2289.8
## + POP_dioxin2      1    0.0000 37.122 -2289.8
## - POP_furan3       1    0.6682 37.791 -2289.3
## + race_cat         3    0.2238 36.899 -2281.2
## + edu_cat          3    0.0890 37.033 -2278.4
## - ageyrs           1    7.5528 44.675 -2159.2
##
## Step:  AIC=-2302.15
## length ~ POP_dioxin1 + POP_furan3 + whitecell_count + lymphocyte_pct +
##     monocyte_pct + male + ageyrs + ln_lbxcot
##
##                   Df Sum of Sq    RSS     AIC
## - monocyte_pct     1    0.0521 37.221 -2307.7
## - POP_dioxin1      1    0.0776 37.247 -2307.2
## - lymphocyte_pct   1    0.1017 37.271 -2306.7
```

```
## - ln_lbxcot        1    0.1588 37.328 -2305.5
## - whitecell_count  1    0.1989 37.368 -2304.7
## <none>                         37.169 -2302.2
## - male             1    0.3454 37.515 -2301.6
## + smokenow         1    0.0469 37.122 -2296.5
## + POP_PCB11        1    0.0373 37.132 -2296.3
## + POP_PCB10        1    0.0232 37.146 -2296.0
## + yrssmoke         1    0.0216 37.148 -2295.9
## + POP_PCB6         1    0.0166 37.153 -2295.8
## + POP_dioxin3      1    0.0123 37.157 -2295.8
## + neutrophils_pct  1    0.0059 37.163 -2295.6
## + basophils_pct    1    0.0052 37.164 -2295.6
## + POP_PCB8         1    0.0051 37.164 -2295.6
## + POP_PCB3         1    0.0051 37.164 -2295.6
## + POP_PCB7         1    0.0023 37.167 -2295.5
## + POP_furan2       1    0.0018 37.168 -2295.5
## + POP_furan1       1    0.0011 37.168 -2295.5
## + POP_PCB9         1    0.0008 37.169 -2295.5
## + POP_furan4       1    0.0005 37.169 -2295.5
## + BMI              1    0.0000 37.169 -2295.5
## + POP_dioxin2      1    0.0000 37.169 -2295.5
## - POP_furan3       1    0.6904 37.860 -2294.5
## + race_cat         3    0.2528 36.917 -2287.5
## + edu_cat          3    0.0900 37.079 -2284.1
## - ageyrs           1    7.7598 44.929 -2161.5
##
## Step:  AIC=-2307.71
## length ~ POP_dioxin1 + POP_furan3 + whitecell_count + lymphocyte_pct +
##     male + ageyrs + ln_lbxcot
##
##                    Df Sum of Sq    RSS     AIC
## - POP_dioxin1       1    0.0759 37.297 -2312.8
## - lymphocyte_pct    1    0.0920 37.313 -2312.4
## - ln_lbxcot         1    0.1525 37.374 -2311.2
## - whitecell_count   1    0.1557 37.377 -2311.1
## <none>                         37.221 -2307.7
## - male              1    0.4220 37.643 -2305.6
## + monocyte_pct      1    0.0521 37.169 -2302.2
## + smokenow          1    0.0477 37.174 -2302.1
## + POP_PCB11         1    0.0405 37.181 -2301.9
## + yrssmoke          1    0.0235 37.198 -2301.6
## + POP_PCB10         1    0.0221 37.199 -2301.5
## + POP_PCB6          1    0.0183 37.203 -2301.4
## + POP_dioxin3       1    0.0110 37.211 -2301.3
## + basophils_pct     1    0.0088 37.213 -2301.2
## + POP_PCB3          1    0.0060 37.215 -2301.2
## + POP_PCB8          1    0.0060 37.215 -2301.2
## + neutrophils_pct   1    0.0044 37.217 -2301.2
## + POP_PCB7          1    0.0022 37.219 -2301.1
## + POP_PCB9          1    0.0014 37.220 -2301.1
## + POP_furan2        1    0.0014 37.220 -2301.1
## + POP_furan1        1    0.0004 37.221 -2301.1
## + POP_furan4        1    0.0002 37.221 -2301.1
## + POP_dioxin2       1    0.0000 37.221 -2301.1
```

```
## + BMI              1    0.0000 37.221 -2301.1
## - POP_furan3       1    0.6685 37.890 -2300.5
## + race_cat         3    0.2598 36.962 -2293.2
## + edu_cat          3    0.0859 37.136 -2289.5
## - ageyrs           1    7.8878 45.109 -2165.0
##
## Step:  AIC=-2312.78
## length ~ POP_furan3 + whitecell_count + lymphocyte_pct + male +
##     ageyrs + ln_lbxcot
##
##                   Df Sum of Sq    RSS     AIC
## - lymphocyte_pct   1    0.0923 37.390 -2317.5
## - whitecell_count  1    0.1427 37.440 -2316.5
## - ln_lbxcot        1    0.1871 37.485 -2315.6
## <none>                         37.297 -2312.8
## - male             1    0.3886 37.686 -2311.4
## + POP_dioxin3      1    0.0816 37.216 -2307.8
## + POP_dioxin1      1    0.0759 37.221 -2307.7
## + monocyte_pct     1    0.0504 37.247 -2307.2
## - POP_furan3       1    0.5954 37.893 -2307.1
## + POP_PCB10        1    0.0305 37.267 -2306.8
## + smokenow         1    0.0299 37.267 -2306.8
## + POP_PCB11        1    0.0232 37.274 -2306.6
## + POP_dioxin2      1    0.0229 37.275 -2306.6
## + POP_furan1       1    0.0107 37.287 -2306.3
## + POP_PCB6         1    0.0105 37.287 -2306.3
## + basophils_pct    1    0.0075 37.290 -2306.3
## + neutrophils_pct  1    0.0063 37.291 -2306.3
## + POP_furan4       1    0.0053 37.292 -2306.2
## + yrssmoke         1    0.0053 37.292 -2306.2
## + POP_furan2       1    0.0046 37.293 -2306.2
## + POP_PCB3         1    0.0038 37.294 -2306.2
## + POP_PCB8         1    0.0030 37.294 -2306.2
## + BMI              1    0.0028 37.295 -2306.2
## + POP_PCB7         1    0.0021 37.295 -2306.2
## + POP_PCB9         1    0.0012 37.296 -2306.2
## + race_cat         3    0.2284 37.069 -2297.6
## + edu_cat          3    0.1094 37.188 -2295.1
## - ageyrs           1    8.4062 45.704 -2161.5
##
## Step:  AIC=-2317.52
## length ~ POP_furan3 + whitecell_count + male + ageyrs + ln_lbxcot
##
##                   Df Sum of Sq    RSS     AIC
## - whitecell_count  1    0.0786 37.468 -2322.5
## - ln_lbxcot        1    0.1574 37.547 -2320.9
## <none>                         37.390 -2317.5
## - male             1    0.3672 37.757 -2316.6
## + lymphocyte_pct   1    0.0923 37.297 -2312.8
## + POP_dioxin3      1    0.0816 37.308 -2312.6
## + POP_dioxin1      1    0.0762 37.313 -2312.4
## - POP_furan3       1    0.5871 37.977 -2312.1
## + monocyte_pct     1    0.0408 37.349 -2311.7
## + POP_PCB10        1    0.0391 37.351 -2311.7
```

```
## + smokenow         1    0.0283 37.361 -2311.4
## + POP_PCB11        1    0.0195 37.370 -2311.3
## + POP_dioxin2      1    0.0153 37.374 -2311.2
## + POP_furan1       1    0.0118 37.378 -2311.1
## + POP_PCB6         1    0.0113 37.378 -2311.1
## + basophils_pct    1    0.0095 37.380 -2311.1
## + BMI              1    0.0085 37.381 -2311.0
## + yrssmoke         1    0.0080 37.382 -2311.0
## + POP_furan4       1    0.0069 37.383 -2311.0
## + POP_PCB7         1    0.0066 37.383 -2311.0
## + POP_PCB3         1    0.0045 37.385 -2311.0
## + POP_furan2       1    0.0041 37.386 -2310.9
## + neutrophils_pct  1    0.0036 37.386 -2310.9
## + POP_PCB8         1    0.0017 37.388 -2310.9
## + POP_PCB9         1    0.0003 37.389 -2310.9
## + race_cat         3    0.1772 37.213 -2301.2
## + edu_cat          3    0.1282 37.262 -2300.2
## - ageyrs           1    8.3216 45.711 -2168.0
##
## Step:  AIC=-2322.54
## length ~ POP_furan3 + male + ageyrs + ln_lbxcot
##
##                   Df Sum of Sq    RSS     AIC
## - ln_lbxcot        1    0.1246 37.593 -2326.6
## <none>                         37.468 -2322.5
## - male             1    0.3315 37.800 -2322.4
## + whitecell_count  1    0.0786 37.390 -2317.5
## + POP_dioxin3      1    0.0766 37.392 -2317.5
## + POP_dioxin1      1    0.0654 37.403 -2317.2
## + POP_PCB10        1    0.0417 37.427 -2316.8
## + smokenow         1    0.0397 37.429 -2316.7
## + lymphocyte_pct   1    0.0281 37.440 -2316.5
## - POP_furan3       1    0.6214 38.090 -2316.4
## + POP_PCB11        1    0.0164 37.452 -2316.2
## + BMI              1    0.0151 37.453 -2316.2
## + yrssmoke         1    0.0135 37.455 -2316.2
## + monocyte_pct     1    0.0113 37.457 -2316.1
## + POP_dioxin2      1    0.0108 37.458 -2316.1
## + POP_PCB6         1    0.0099 37.458 -2316.1
## + POP_furan1       1    0.0098 37.459 -2316.1
## + basophils_pct    1    0.0060 37.462 -2316.0
## + POP_PCB7         1    0.0059 37.462 -2316.0
## + POP_furan4       1    0.0052 37.463 -2316.0
## + POP_PCB3         1    0.0041 37.464 -2316.0
## + neutrophils_pct  1    0.0038 37.464 -2316.0
## + POP_furan2       1    0.0019 37.466 -2315.9
## + POP_PCB9         1    0.0009 37.467 -2315.9
## + POP_PCB8         1    0.0003 37.468 -2315.9
## + race_cat         3    0.2085 37.260 -2306.9
## + edu_cat          3    0.1306 37.338 -2305.3
## - ageyrs           1    8.2648 45.733 -2174.3
##
## Step:  AIC=-2326.62
## length ~ POP_furan3 + male + ageyrs
```

```
##
##                   Df Sum of Sq    RSS     AIC
## - male             1    0.2906 37.883 -2327.3
## <none>                          37.593 -2326.6
## + ln_lbxcot        1    0.1246 37.468 -2322.5
## + POP_dioxin1      1    0.0939 37.499 -2321.9
## + POP_dioxin3      1    0.0748 37.518 -2321.5
## + POP_PCB10        1    0.0537 37.539 -2321.1
## + whitecell_count  1    0.0458 37.547 -2320.9
## - POP_furan3       1    0.6070 38.200 -2320.8
## + smokenow         1    0.0350 37.558 -2320.7
## + BMI              1    0.0227 37.570 -2320.4
## + lymphocyte_pct   1    0.0224 37.570 -2320.4
## + monocyte_pct     1    0.0134 37.580 -2320.2
## + POP_PCB11        1    0.0104 37.582 -2320.2
## + yrssmoke         1    0.0103 37.583 -2320.2
## + POP_PCB6         1    0.0075 37.585 -2320.1
## + POP_PCB9         1    0.0058 37.587 -2320.1
## + basophils_pct    1    0.0058 37.587 -2320.1
## + POP_PCB7         1    0.0050 37.588 -2320.1
## + POP_PCB3         1    0.0048 37.588 -2320.1
## + POP_furan1       1    0.0045 37.588 -2320.1
## + neutrophils_pct  1    0.0039 37.589 -2320.1
## + POP_dioxin2      1    0.0031 37.590 -2320.0
## + POP_furan4       1    0.0012 37.592 -2320.0
## + POP_PCB8         1    0.0004 37.593 -2320.0
## + POP_furan2       1    0.0002 37.593 -2320.0
## + race_cat         3    0.2544 37.338 -2311.9
## + edu_cat          3    0.0960 37.497 -2308.6
## - ageyrs           1    8.9093 46.502 -2168.0
##
## Step:  AIC=-2327.29
## length ~ POP_furan3 + ageyrs
##
##                   Df Sum of Sq    RSS     AIC
## <none>                          37.883 -2327.3
## + male             1    0.2906 37.593 -2326.6
## + ln_lbxcot        1    0.0837 37.800 -2322.4
## + monocyte_pct     1    0.0585 37.825 -2321.8
## + POP_dioxin1      1    0.0581 37.825 -2321.8
## + POP_dioxin3      1    0.0302 37.853 -2321.3
## + smokenow         1    0.0296 37.854 -2321.2
## + POP_PCB7         1    0.0286 37.855 -2321.2
## + whitecell_count  1    0.0248 37.859 -2321.2
## + POP_PCB6         1    0.0222 37.861 -2321.1
## + lymphocyte_pct   1    0.0222 37.861 -2321.1
## + BMI              1    0.0197 37.864 -2321.0
## + POP_PCB11        1    0.0190 37.864 -2321.0
## + basophils_pct    1    0.0152 37.868 -2320.9
## + POP_PCB10        1    0.0098 37.874 -2320.8
## + POP_PCB3         1    0.0095 37.874 -2320.8
## + POP_furan4       1    0.0065 37.877 -2320.8
## + POP_furan1       1    0.0064 37.877 -2320.8
## + POP_PCB8         1    0.0047 37.879 -2320.7
```

```
## + neutrophils_pct  1    0.0043 37.879 -2320.7
## + POP_dioxin2      1    0.0028 37.881 -2320.7
## + POP_furan2       1    0.0023 37.881 -2320.7
## + yrssmoke         1    0.0014 37.882 -2320.7
## + POP_PCB9         1    0.0012 37.882 -2320.7
## - POP_furan3       1    0.6637 38.547 -2320.4
## + race_cat         3    0.2478 37.636 -2312.4
## + edu_cat          3    0.1056 37.778 -2309.5
## - ageyrs           1    9.2607 47.144 -2164.0
```

```
summary(step_bic)
```

```
##
## Call:
## lm(formula = length ~ POP_furan3 + ageyrs, data = train_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.51364 -0.15642 -0.02609  0.12092  1.16684
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.359962   0.022585  60.215  < 2e-16 ***
## POP_furan3   0.006055   0.001644   3.682 0.000247 ***
## ageyrs      -0.007125   0.000518 -13.755  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2212 on 774 degrees of freedom
## Multiple R-squared:  0.2137, Adjusted R-squared:  0.2116
## F-statistic: 105.2 on 2 and 774 DF,  p-value: < 2.2e-16
```

```
bic_pred = predict(step_bic, newdata = pollutants[-train_row,])

#RMSE AIC
bic_true = pollutants$length[-train_row]
bic_sd = sum((bic_true - bic_pred)^2)
msd_bic = bic_sd / length(bic_true)
rmse_bic = sqrt(msd_bic)
rmse_bic
```

```
## [1] 0.2298015
```