STAT 331: Final Project

Tian Shi

Xinyi Sheng

Kevin Huang

1. Summary

The study aims to investigate the relationship between birth weight and some variables related to pregnant mothers about air quality during pregnancy. The sample collects data from 1000 births between 2003 and 2009. In this study, 6 factors measuring air quality were are selected and analyzed based on sex of babies. This project performed the statistical analysis according to the following steps: we started by making a summary based on the level of exposure to air pollutants that the mother was exposed to during pregnancy. Comparing the differences in the effects of air pollution in infants with different sex, found the factors influencing birth weight. Then, we analyzed the relationship between pollutants and birth weight with multiple regressions, and the effects on birth weight were studied when six factors were synergistic or antagonistic . By analyzing the models we have built and examining with cross-validation, we concluded that there was a slight negative relationship between air quality and birth weight. Moreover, the impact of air quality on newborn boys was smaller than girls. However, the relationship between most measures of air quality and birthweight were not statistically significant.

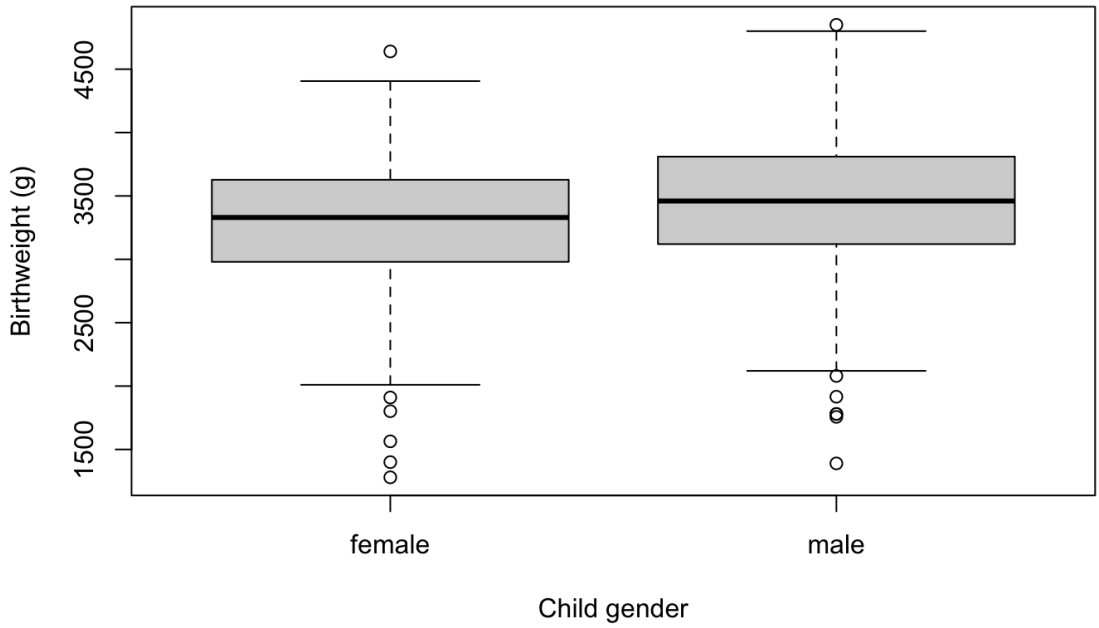
2. Objective

Smoky haze from recent wildfires raises Canadians' awareness to air pollution. Further, air quality has been examined as an important factor for personal wellness.The relationship between air quality and birth weight of newborn babies is meaningful to be analyzed. In this study, we included measures of air quality including pollutants like nitrogen dioxide (NO_2), particulate matter ($PM_{2.5}$ and PM_{10}), maternal smoking, as well as humidity. Then focus on their impact on birthweight of newborn babies with different sex. Thus, the study was aimed to investigate the impact of air quality on newborn babies, and identify if air quality influences male and female babies equally.

3. Exploratory Data Analysis

3.1 Data summary

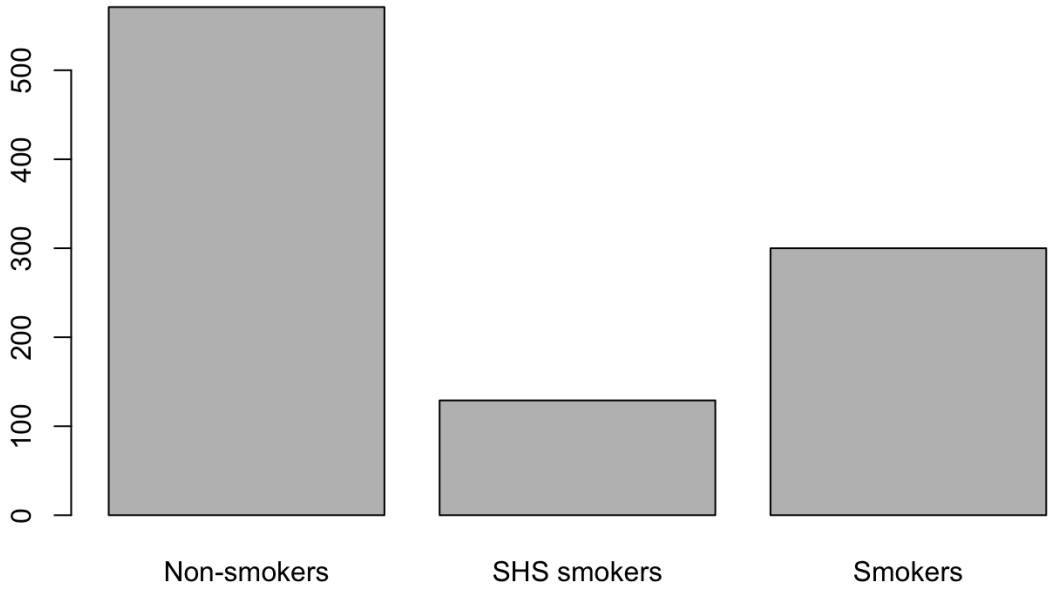
First of all, the report focused on whether pregnancy in an environment of air pollution, and baby gender can affect the birth weight of newborns. Thus, we started to analyze the impact of following variables associated with air quality and child weight at birth in grams (e3_bw): child sex (e3_sex_None), humidity average during pregnancy (h_humidity_preg_None), nitrogen dixoxide value during pregnancy (h_no2_ratio_preg_Log,), PM10 value during pregnancy(h_pm10_ratio_preg_None), PM2.5 value during pregnancy (h_pm25_ratio_preg_None), the mean number of maternal active tobacco smoke during pregnancy daily (e3_asmokcigd_p_None), and amount of cotinine in mothers (hs_cotinine_mcat_None) (see Appendix 1).



The shape and spread of the distributions of babies with different sex were very similar. For the males and the females, the center line in the box, which corresponds to the median, divides both the box and the whiskers approximately in half which indicates that both distributions are roughly symmetric about the median.

Besides, the median birth weight for females is approximately 3300 grams, while the median birth weight for males is approximately 3400 grams. For females, q(0.25) = 3000, q(0.75) = 3600, IQR = 3600−3000 = 600, and range = 4400−2000 = 2200. For males, q(0.25) = 3100, q(0.75) = 3750, IQR = 3750−3100 = 650, and range = 4900−2100 = 2800. The IQR and range for females are slightly similar to the IQR and

range for males. From the box plot and data, it is visible that boys are generally much heavier than girls; and weight for boys is more spread out.

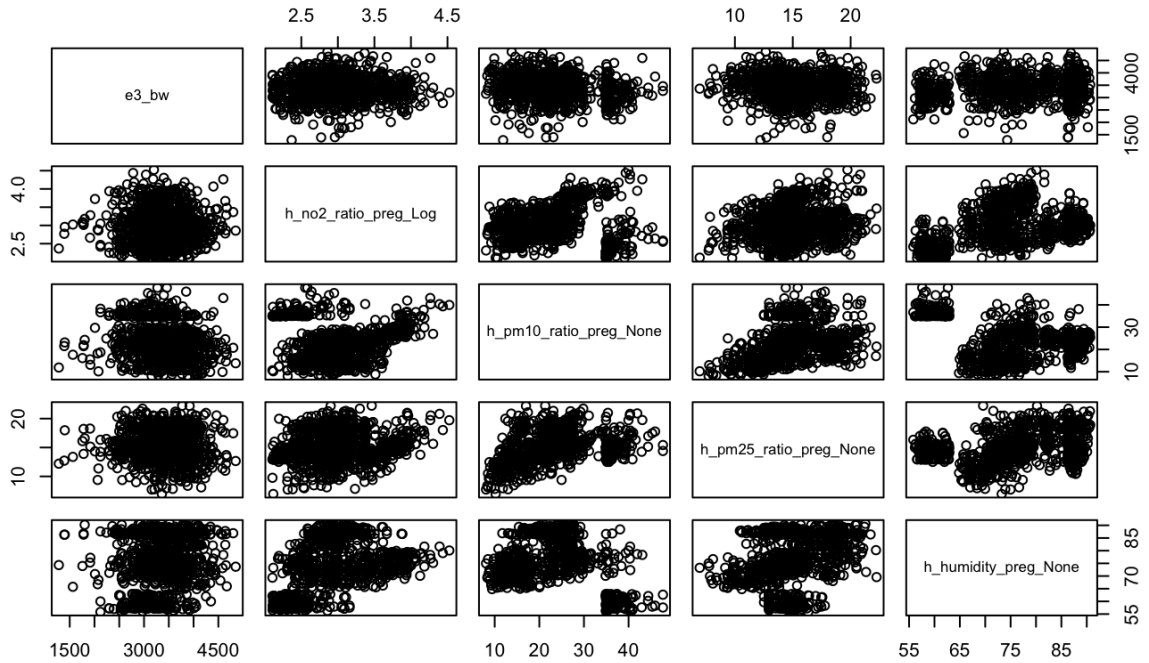


According to the figure above, the vast majority of mothers chose not to smoke during pregnancy for the sake of their children's health. At the same time, there was a small number of mothers exposed to second-hand smoke.

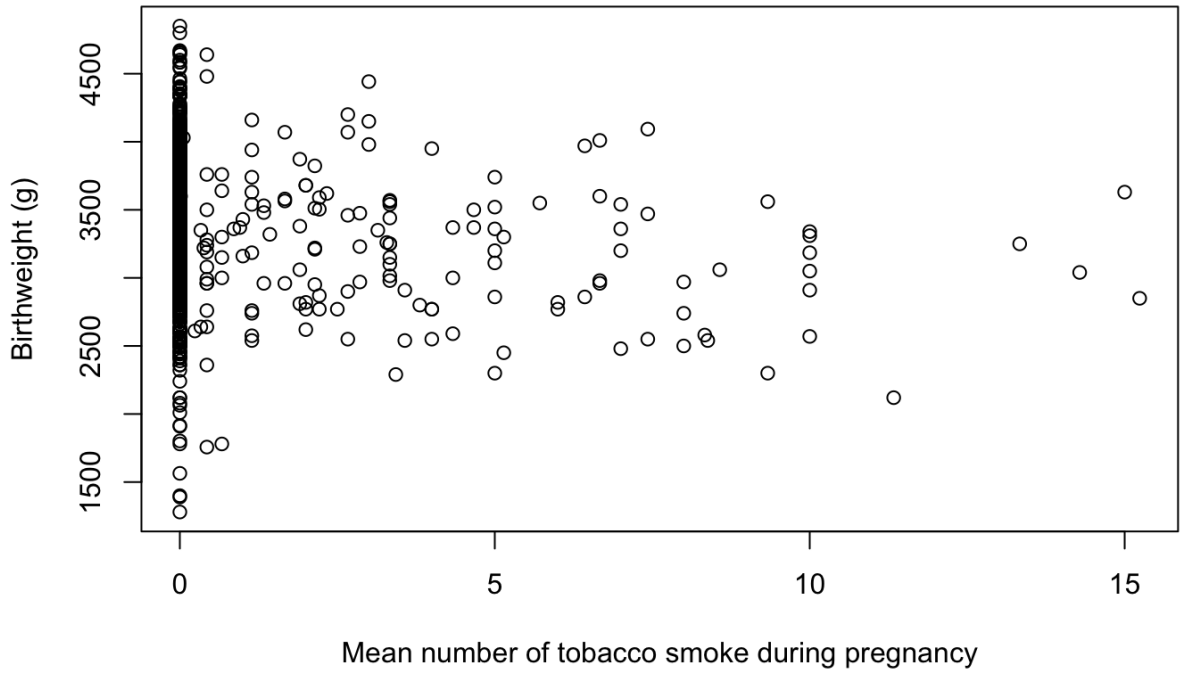
Beyond that, we noticed that the two variables sex (e3_sex_None) and cotinine (hs_cotinine_mcat_None) are categorical variables. Hence, we converted them to factors for subsequent calculations (see Appendix 3).

3.2 Data analysis

To begin with, we set up a pair plot for testing which variables are related to the birth weight (see Appendix 2). From the figure below, it turned out that there existed a weak and likely unimportant correlation (i.e. no obvious relationship) between NO2, PM2.5, and birth weight. What's more, there was a small positive correlation between humidity average and birth weight since the correlation was approximately 0.1. Besides, the correlation between PM10 and birth weight was negative.



Meanwhile, we noticed that the birth weight was not significantly related to the amount of tobacco that the mother smokes on average.



4. Methods

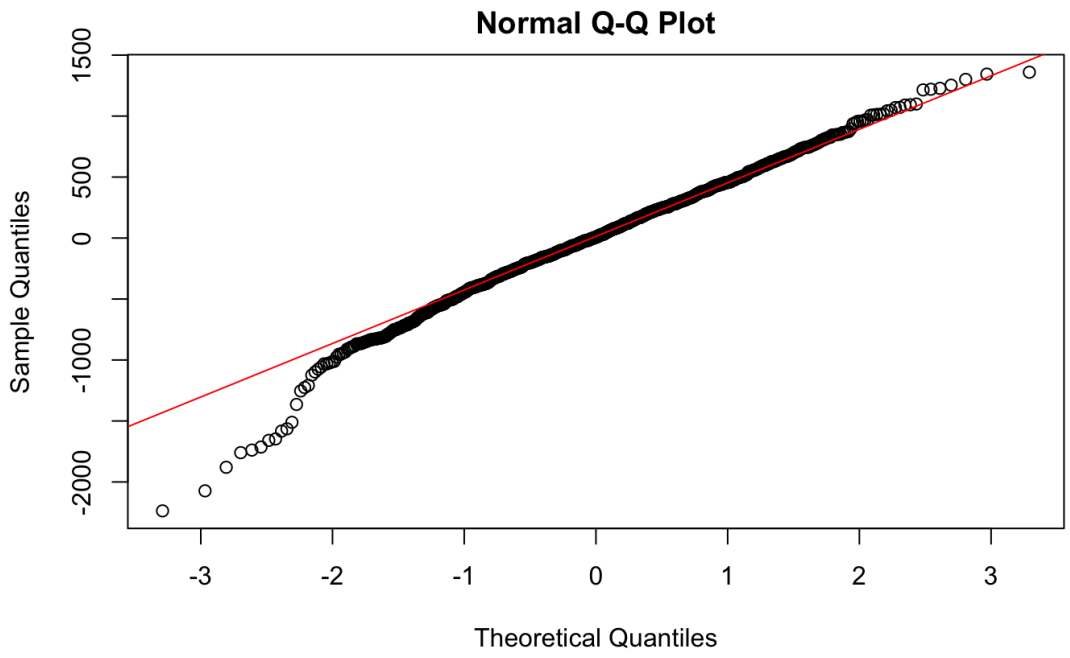
4.1 Model Selection

We chose to use backward elimination and compared it with stepwise model selection.The benefit of backward selection is this approach starts with a full model, which includes all the variables. The variables are then removed from the full model until all remaining variables are considered to relate to the results significantly. It is easy to determine which factor should be removed from each process step. In addition, several factors are combined while using the backwards elimination, indicating that they may have better predictive accuracy.

Therefore, we constructed the minimal and full models for backwards elimination (see Appendix 5-7). We knew that the mean amount of cotinine and the mean number of cigarettes that mothers took had no apparent contribution to babies' birth weight. Then, when we built up the full model, we didn't associate these two variables with gender (see Appendix 6). Furthermore, we generated a model through running the stepwise selection. It turned out that backward and stepwise selection produced the same model.

According to the model produced from both selection processes, we ended up with birth weight associated with gender, PM10 value, mean maternal smoking and amount of cotinine in mothers as the final model (see Appendix 9).

Beyond that, we plotted the QQ-plot for the final model (see Appendix 10). Clearly, the points lay reasonably along a straight line, which meant that it was reasonable to assume a normal model for the dataset. It was also worth mentioning that babies with extremely low birth weight were not much affected by air quality.

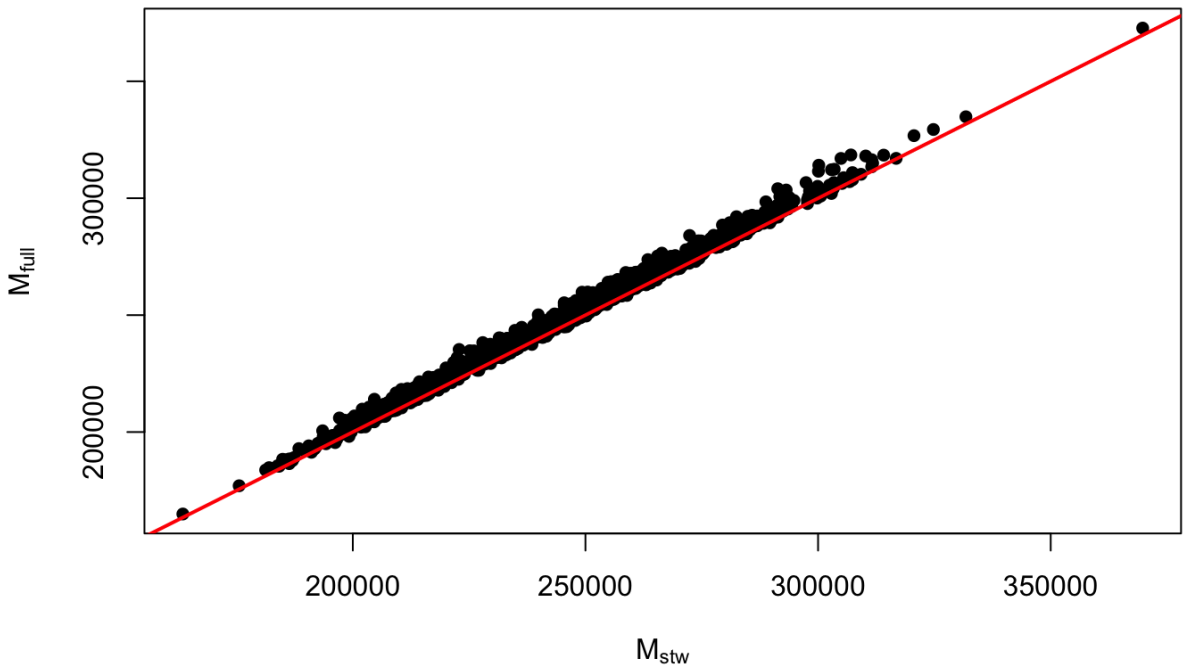


Additionally, we constructed two different models based on child sex (see Appendix 11). From the comparison of two models, air quality had a slightly smaller impact on boys than girls. Also, the p-value of hs_cotinine_mcat_None was greater than 0.05, then we concluded that the interaction was not statistically significant. As a result, for both male and female babies, the maternal smoking during pregnancy (smoking herself or secondhand smoke) had a minor effect on the child's birth weight.

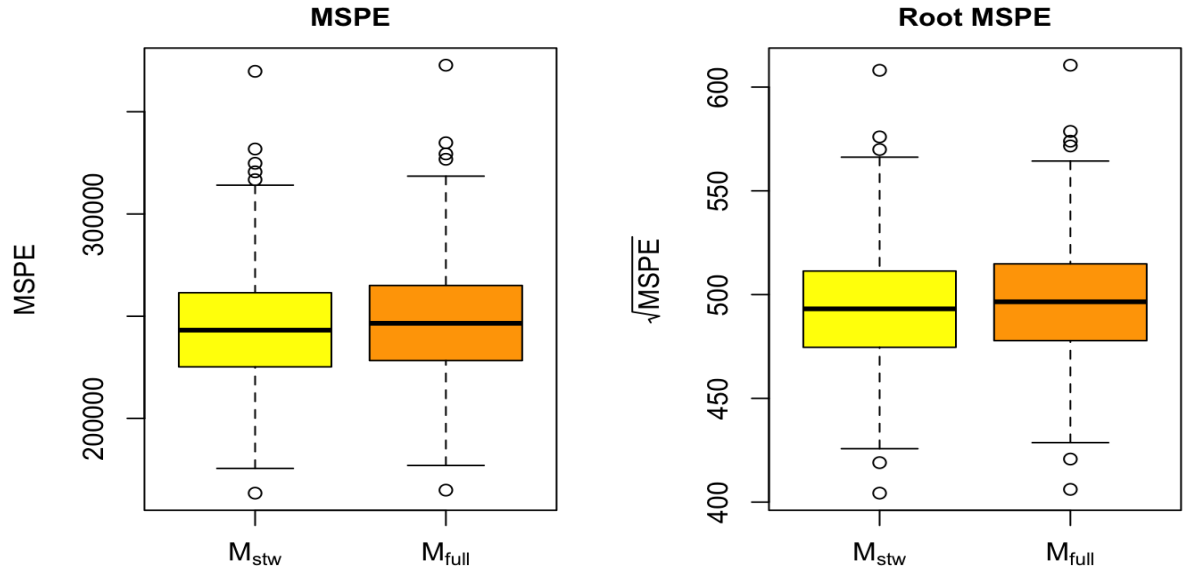
4.2 Cross-validation of model

In order to measure the predictive accuracy of above models, cross validation was used to compare the models (see Appendix 12). Out of the 1000 data entries in the observation, 800 of them were set to be the training set, and the remaining 200 formed the test set.

From random subsets cross-validation tests, the graph produced from prediction of the training set showed that the stepwise model had fewer prediction errors than the full model. Clearly, the stepwise model was more accurate in prediction than the original full model.



However, by looking at the mean squared prediction error (MSPE), there was not much difference between these two models. From leave-one-out cross validation, the MSPE of stepwise model and full model were 243003.8 and 245799.3 respectively. The stepwise model was still a better prediction of the outcome.



Similarly, the mean squared prediction error was used to compare the stepwise model and full model produced from cross validation. The leave-one-out cross validation MSPE of the full model was 244677.4, slightly higher than the stepwise model. Therefore, the stepwise model was chosen to be our final model.

5. Results

The study looked at the relationship between six measures of air quality during pregnancy and the birth weight of newborns, and found different effects on babies with different gender. After automatic stepwise model selection with Akaike information criterion (AIC), we found that exposure levels of PM10 and smoke had a negative relationship. Baby boys were a bit less impacted by bad air quality than girls. Overall, the association between air quality and birthweight was not strong.

6. Discussion

There are several factors limiting our study and influencing the accuracy of findings. Air quality, as a measure of air pollution, which counts the mixture of particles and gases in air. However, we only have data available on a small number of factors influencing air quality; namely, humidity, nitrogen dioxide, PM10, PM2.5 and smoking. There are a lot of pollutants from air that are not included in the data set, like sulfur dioxide, carbon monoxide, ozone, etc. Their effects on birth weight are not considered in this study.

Particles in air (mainly PM10) could carry various pathogens that cause communicable diseases. That could be the reason why PM10 has much more impact on birth weight compared to all other air pollutants. However, the effect of airborne transmission from PM10 varies from region to region and differs from time to time. The relationship between communicable diseases which are airborne transitive and birth weight of boys and girls are unknown as well.

Furthermore, air pollution could lead to miscarrage and stillbirth, which both are not taking into consideration in this study.

This study showed that some focus needs to be placed on recognizing the negative impact bad air quality has on birth weight. Though they did not cause a serious problem on birth weight, it is still recommended for future mothers to install an air purifier, change air filters frequently and choose a less polluted community. It will not only give a heavier and healthier baby, but also boost parents' wellness.

7. Reference for R code

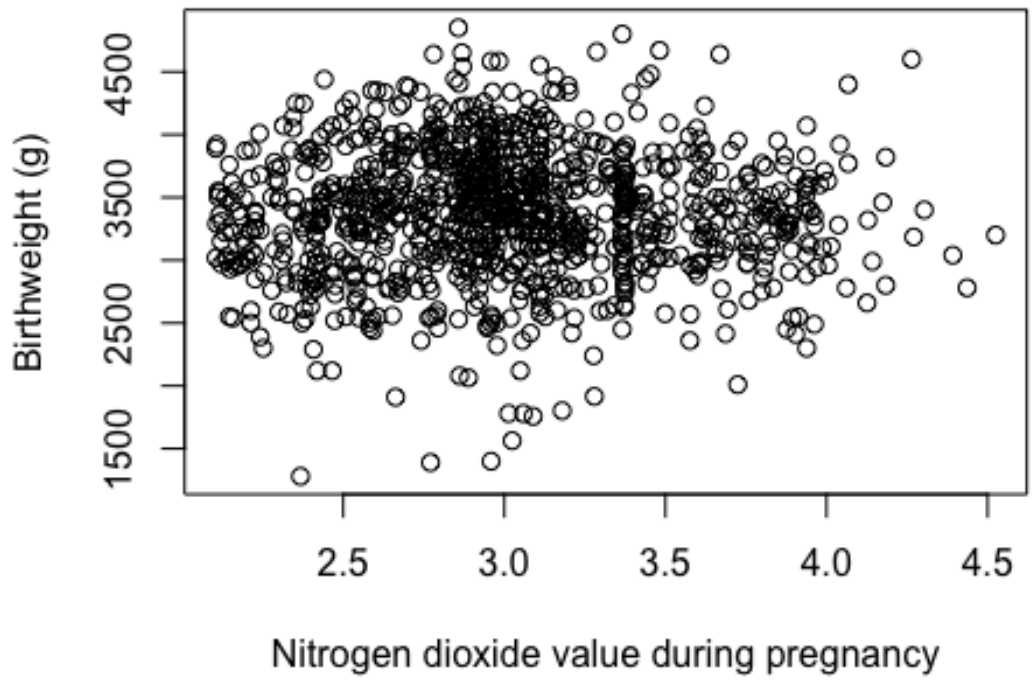
Appendix 1

```
dataset <- get(load("~/Desktop/Project/pollution.Rdata"))
data <- dataset[,c(1,3,4,5,28,69,70,74)]
data <- data.frame(data)
summary(data)
```

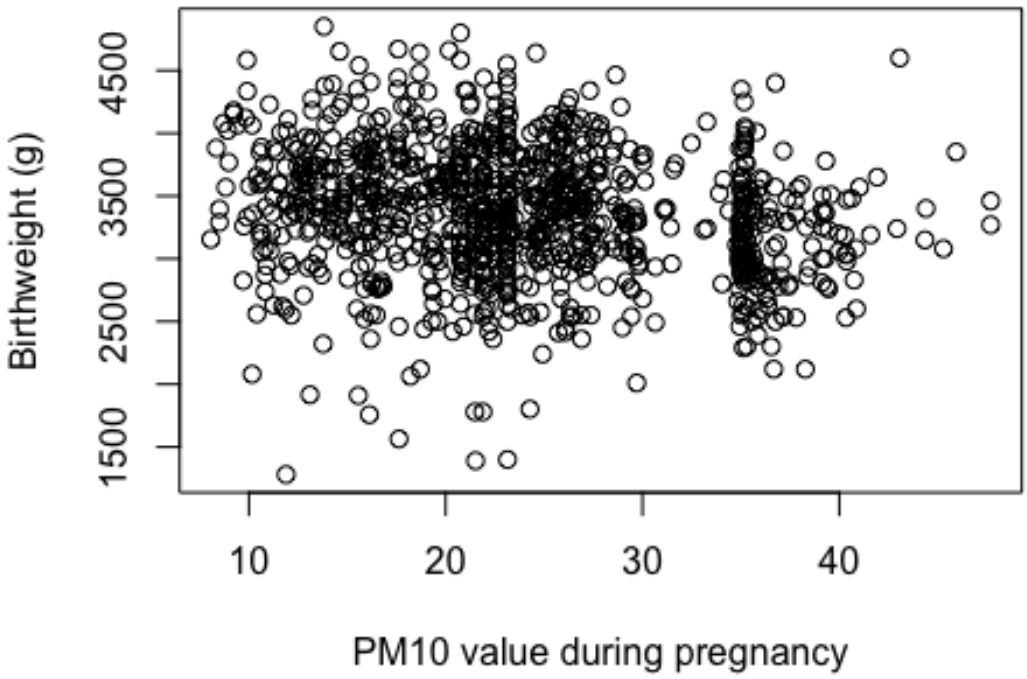
```
##      e3_bw      h_no2_ratio_preg_Log h_pm10_ratio_preg_None
##  Min.   :1280   Min.   :2.105        Min.   : 8.066
##  1st Qu.:3050   1st Qu.:2.663        1st Qu.:17.631
##  Median :3390   Median :2.967        Median :23.119
##  Mean   :3378   Mean   :3.009        Mean   :23.724
##  3rd Qu.:3720   3rd Qu.:3.340        3rd Qu.:27.885
##  Max.   :4850   Max.   :4.525        Max.   :47.698
##  h_pm25_ratio_preg_None h_humidity_preg_None e3_asmokcigd_p_None
##  Min.   : 6.957         Min.   :55.83        Min.   : 0.0000
##  1st Qu.:13.335         1st Qu.:70.44        1st Qu.: 0.0000
##  Median :14.905         Median :77.03        Median : 0.0000
##  Mean   :15.088         Mean   :76.48        Mean   : 0.5545
##  3rd Qu.:17.016         3rd Qu.:86.55        3rd Qu.: 0.0000
##  Max.   :22.238         Max.   :90.67        Max.   :15.2381
##  hs_cotinine_mcat_None e3_sex_None
##  Non-smokers:571       female:471
##  SHS smokers:129       male  :529
##  Smokers    :300
##
##
##
```

Appendix 2

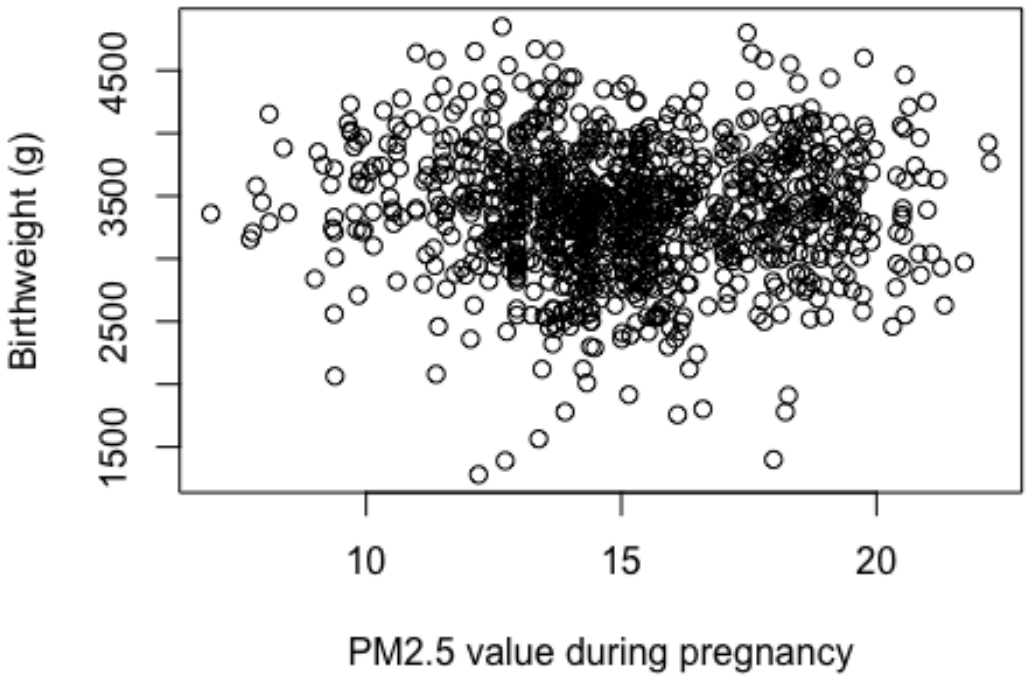
```
# scatter plots
plot(data$h_no2_ratio_preg_Log, data$e3_bw, ylab = "Birthweight (g)",
     xlab = "Nitrogen dioxide value during pregnancy")
```



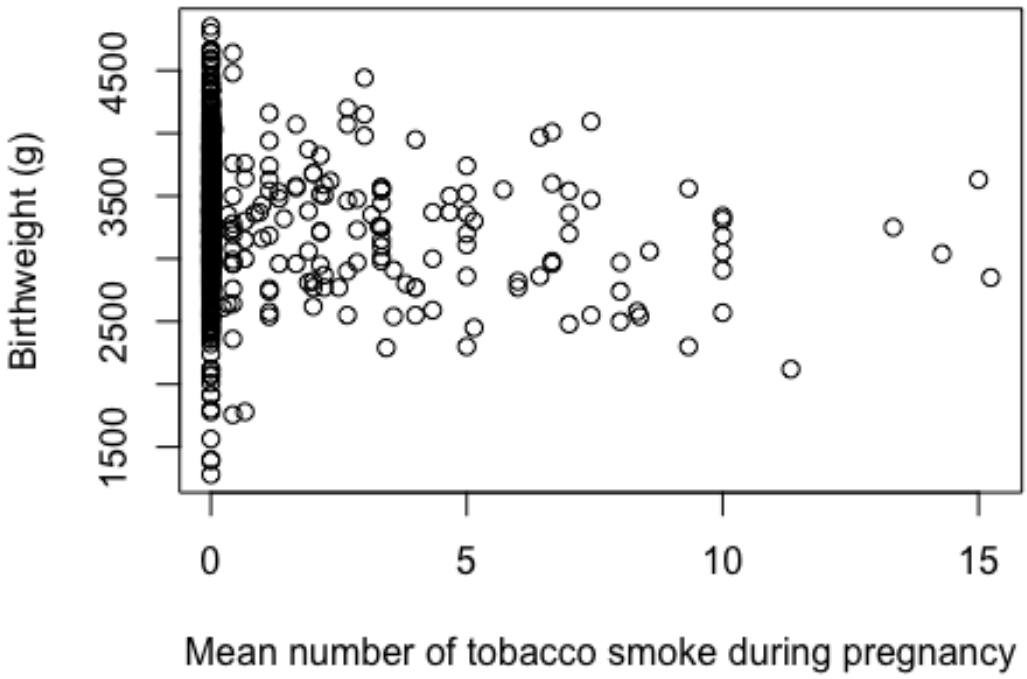

```
plot(data$h_pm10_ratio_preg_None, data$e3_bw, ylab = "Birthweight (g)",
     xlab = "PM10 value during pregnancy")
```



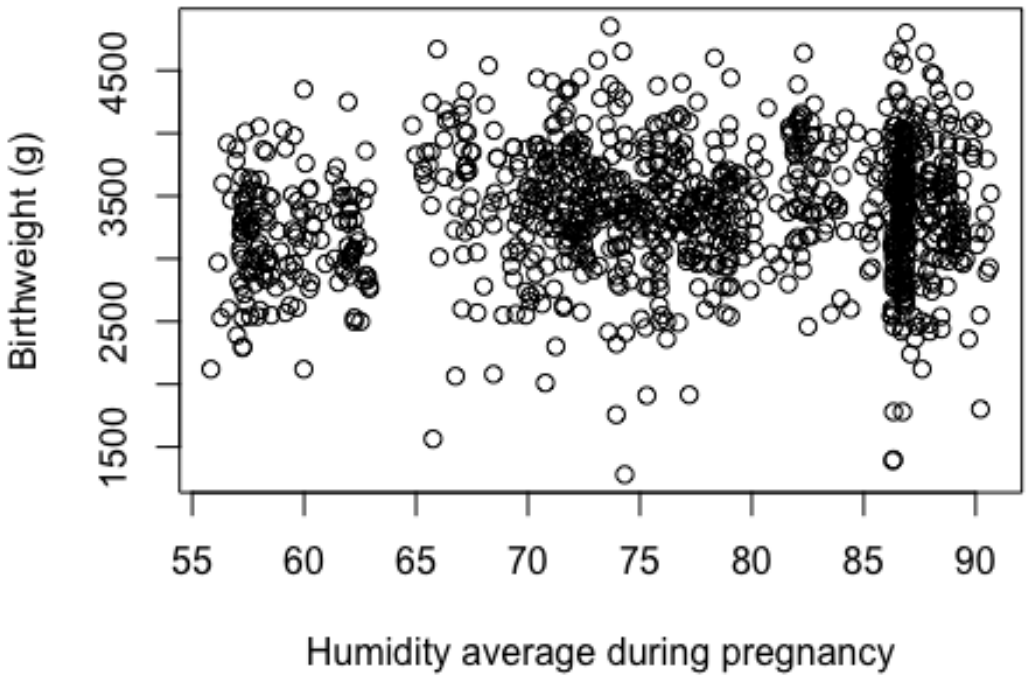

```r
plot(data$h_pm25_ratio_preg_None, data$e3_bw, ylab = "Birthweight (g)",
     xlab = "PM2.5 value during pregnancy")
```




```r
plot(data$e3_asmokcigd_p_None, data$e3_bw, ylab = "Birthweight (g)",
     xlab = "Mean number of tobacco smoke during pregnancy")
```




```r
plot(data$h_humidity_preg_None, data$e3_bw, ylab = "Birthweight (g)",
     xlab = "Humidity average during pregnancy")
```




```
cor(data$h_no2_ratio_preg_Log, data$e3_bw)
```

```
## [1] 0.005153035
```

```
cor(data$h_pm10_ratio_preg_None, data$e3_bw)
```

```
## [1] -0.1842447
```

```
cor(data$h_pm25_ratio_preg_None, data$e3_bw)
```

```
## [1] -0.02244666
```

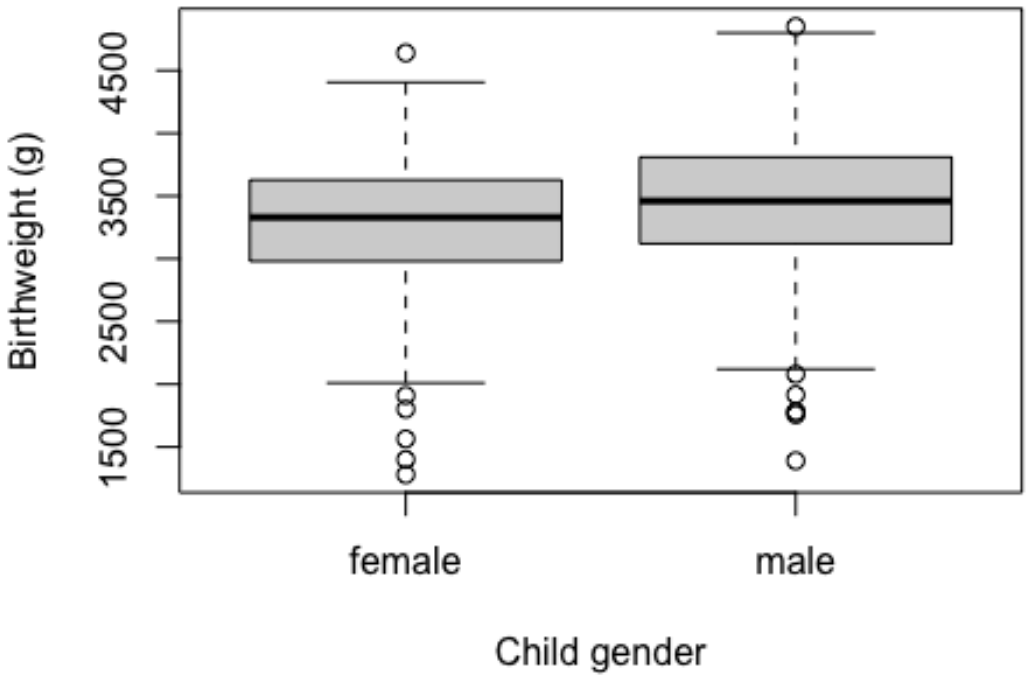
```
cor(data$h_humidity_preg_None, data$e3_bw)
```

```
## [1] 0.1004987
```

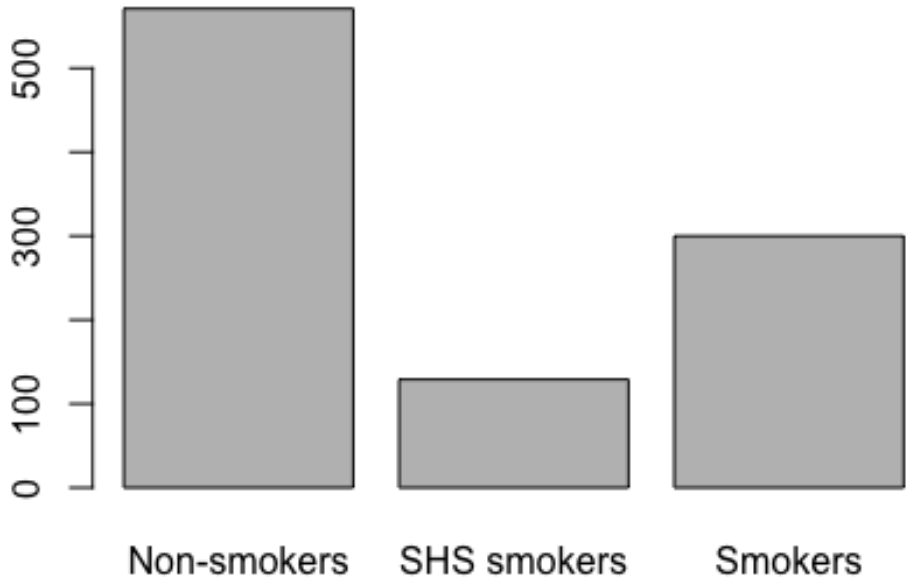
```
cor(data$e3_asmokcigd_p_None, data$e3_bw)
```

```
## [1] -0.1469302
```

```
plot(data$e3_sex_None, data$e3_bw, ylab = "Birthweight (g)",
     xlab = "Child gender")
```




```
barplot(summary(as.factor(data$hs_cotinine_mcat_None)))
```



Appendix 3

```
# convert the categorical variables to numeric
as.factor(data$e3_sex_None)
```

```
##    [1] female female male   male   female male   female male   male   male
##   [11] female female male   female male   female female male   female
female
##   [21] male   female female female male   female male   male   female
female
##   [31] female male   male   male   female female female male   female
female
##   [41] female male   male   male   male   female male   female male   male
##   [51] male   male   male   female male   male   male   female male   male
##   [61] male   female male   female male   female male   male   male
female
##   [71] female female male   female male   female male   male   male
female
##   [81] female female male   male   female female male   male   female male
##   [91] male   female male   female male   male   male   male   female male
##  [101] male   male   female male   female male   female female male   male
##  [111] male   male   female male   male   male   female male   female
```

```
female
##  [121] female female male   female female female female male   male   male
##  [131] male   female male   female male   female male   female male
female
##  [141] male   female male   female female female male   male   female
female
##  [151] male   male   male   female male   male   female male   female male
##  [161] female male   male   male   female male   female male   male   male
##  [171] male   male   female female male   male   female male   male
female
##  [181] female female female male   male   male   female female female male
##  [191] female male   female female male   female male   female male
female
##  [201] female female female female male   male   female female male   male
##  [211] male   male   male   male   male   female male   female male   male
##  [221] male   male   male   male   female male   male   male   male
female
##  [231] male   male   male   female male   male   male   female male
female
##  [241] female female female male   male   male   male   female female
female
##  [251] female male   female female male   male   male   male   female
female
##  [261] female female male   female male   male   female female female male
##  [271] male   female male   female female male   male   male   female male
##  [281] male   male   male   male   female female female male   female male
##  [291] male   female male   male   female female female female female
female
##  [301] male   male   male   male   female male   male   male   male
female
##  [311] male   female male   female female male   female female female male
##  [321] male   male   female male   male   female male   male   male
female
##  [331] male   male   male   male   female male   male   female female male
##  [341] male   female female female female female male   male   female
female
##  [351] female male   female female female female male   female male   male
##  [361] female male   female female male   male   female male   male
female
##  [371] female male   female female male   female female female male
female
##  [381] female male   female male   female male   female female male   male
##  [391] female female female female male   female male   male   male
female
##  [401] female female female male   male   female female male   male
female
##  [411] male   female male   female female female female male   male   male
##  [421] female female female male   female male   male   female female male
##  [431] male   male   male   female female male   male   female female male
##  [441] male   male   male   male   male   male   male   male   male   male
```

```
##  [451] female male   female male   female male   male   female male   male
##  [461] male   male   male   male   female male   female male   female
female
##  [471] female female male   female male   female male   female female male
##  [481] male   male   male   male   male   female male   male   male
female
##  [491] male   male   male   female male   female female male   male   male
##  [501] male   male   female female male   female female male   male   male
##  [511] male   female female male   female female female male   male
female
##  [521] female male   female male   female male   male   male   female male
##  [531] female male   male   female male   male   male   male   male
female
##  [541] male   female female female female male   female male   female
female
##  [551] male   female male   female female female female male   female
female
##  [561] female male   male   female female female male   male   male
female
##  [571] male   male   female female male   female male   male   male
female
##  [581] female male   female male   female female male   female male
female
##  [591] male   female female female female male   female male   female
female
##  [601] male   female female female male   male   female female female male
##  [611] male   female male   male   female male   female male   male   male
##  [621] female male   male   female male   female female female male
female
##  [631] male   female male   male   female male   female male   female
female
##  [641] female male   female male   male   female male   male   male
female
##  [651] male   female male   female male   female male   male   male   male
##  [661] female female female male   male   male   female female female
female
##  [671] male   male   female female male   female male   male   female male
##  [681] male   male   female female male   male   female male   female
female
##  [691] male   male   female male   female female male   female female male
##  [701] male   male   female female female female female female female
female
##  [711] male   female male   female male   female female male   male
female
##  [721] male   female female female male   male   female male   male
female
##  [731] male   female female male   male   female male   female male
female
##  [741] female male   male   female male   male   female male   female male
##  [751] male   male   female male   male   female male   male   female
```

```
female
##  [761] male   male   male   male   female male   male   male   male
female
##  [771] female male   male   male   female female female male   female male
##  [781] female male   male   female male   female male   male   male   male
##  [791] male   male   male   female female female male   female male
female
##  [801] male   female male   male   male   female female female male
female
##  [811] female male   female female female male   female male   female
female
##  [821] male   male   female female female male   male   male   female male
##  [831] female male   male   male   female male   male   female female male
##  [841] male   male   female male   male   male   male   male   female
female
##  [851] female female male   male   female male   male   female female
female
##  [861] male   female female male   female female male   female female
female
##  [871] female male   female female male   male   male   male   male   male
##  [881] female male   female female male   male   male   male   male   male
##  [891] male   female female male   female male   female female male
female
##  [901] female male   male   male   female female male   female male   male
##  [911] female male   male   male   female male   male   male   female
female
##  [921] female male   female female male   male   male   female female
female
##  [931] female female male   female female male   male   male   female
female
##  [941] male   male   male   female male   male   male   male   male   male
##  [951] female female male   female male   female male   male   female
female
##  [961] female male   female female female male   female male   male
female
##  [971] female female female female male   female male   male   male
female
##  [981] female female female male   female male   male   female female male
##  [991] female male   male   female male   male   female female male   male
## Levels: female male
```

```
as.factor(data$hs_cotinine_mcat_None)
```

```
##    [1] Smokers     Non-smokers Non-smokers Non-smokers Non-smokers Smokers
##    [7] Non-smokers SHS smokers Non-smokers Non-smokers Non-smokers SHS
smokers
##   [13] Non-smokers SHS smokers Non-smokers Non-smokers Non-smokers
Non-smokers
##   [19] SHS smokers Non-smokers Non-smokers Non-smokers Non-smokers Smokers
##   [25] Smokers     Smokers     SHS smokers Smokers     Non-smokers
```

```
Non-smokers
##   [31] Non-smokers Smokers     Non-smokers Non-smokers Non-smokers
Non-smokers
##   [37] Non-smokers SHS smokers SHS smokers Non-smokers Smokers     SHS
smokers
##   [43] Non-smokers Smokers     Non-smokers Smokers     Smokers     SHS
smokers
##   [49] Smokers     Non-smokers Non-smokers Non-smokers Non-smokers
Non-smokers
##   [55] Non-smokers Non-smokers Non-smokers SHS smokers Non-smokers Smokers
##   [61] Smokers     Non-smokers Smokers     Non-smokers Non-smokers
Non-smokers
##   [67] Non-smokers Smokers     Non-smokers Smokers     Smokers
Non-smokers
##   [73] Smokers     Smokers     SHS smokers Non-smokers Non-smokers
Non-smokers
##   [79] Non-smokers Smokers     Smokers     Smokers     Non-smokers Smokers
##   [85] Non-smokers Non-smokers Non-smokers Smokers     Non-smokers
Non-smokers
##   [91] Non-smokers Smokers     Smokers     Smokers     Non-smokers Smokers
##   [97] Non-smokers Smokers     Non-smokers Non-smokers Non-smokers Smokers
##  [103] Smokers     Non-smokers Smokers     SHS smokers Smokers
Non-smokers
##  [109] Smokers     Non-smokers Non-smokers SHS smokers Non-smokers Smokers
##  [115] Non-smokers Non-smokers Non-smokers Smokers     Smokers     Smokers
##  [121] Non-smokers Non-smokers Non-smokers Smokers     Non-smokers
Non-smokers
##  [127] Non-smokers Smokers     Smokers     SHS smokers Non-smokers Smokers
##  [133] Non-smokers Non-smokers Non-smokers Smokers     Smokers
Non-smokers
##  [139] SHS smokers Non-smokers Smokers     Smokers     Non-smokers Smokers
##  [145] Smokers     Non-smokers Smokers     Non-smokers Non-smokers Smokers
##  [151] Non-smokers Non-smokers Non-smokers Non-smokers Non-smokers SHS
smokers
##  [157] Non-smokers Smokers     Smokers     SHS smokers Non-smokers Smokers
##  [163] Non-smokers SHS smokers Non-smokers Non-smokers Non-smokers SHS
smokers
##  [169] Non-smokers SHS smokers Non-smokers Smokers     Smokers     Smokers
##  [175] Non-smokers Smokers     Smokers     Non-smokers Smokers
Non-smokers
##  [181] Smokers     SHS smokers Non-smokers Smokers     Non-smokers
Non-smokers
##  [187] Non-smokers Non-smokers Non-smokers Smokers     Non-smokers
Non-smokers
##  [193] Smokers     SHS smokers Smokers     Smokers     Non-smokers
Non-smokers
##  [199] Non-smokers Non-smokers Non-smokers Non-smokers Smokers     SHS
smokers
##  [205] Smokers     Non-smokers Non-smokers Non-smokers Smokers
Non-smokers
```

```
##  [211] Smokers     Non-smokers Non-smokers Non-smokers Non-smokers
Non-smokers
##  [217] SHS smokers Non-smokers Smokers     Non-smokers Non-smokers
Non-smokers
##  [223] Smokers     Non-smokers Smokers     Smokers     Smokers     Smokers
##  [229] Smokers     Smokers     SHS smokers SHS smokers Smokers
Non-smokers
##  [235] SHS smokers SHS smokers Non-smokers Non-smokers SHS smokers
Non-smokers
##  [241] Non-smokers Non-smokers Non-smokers Non-smokers Smokers     Smokers
##  [247] Non-smokers SHS smokers Non-smokers Smokers     Non-smokers Smokers
##  [253] Non-smokers Non-smokers Smokers     Non-smokers Smokers     Smokers
##  [259] Non-smokers Non-smokers Non-smokers Non-smokers Non-smokers Smokers
##  [265] Non-smokers Non-smokers Smokers     Smokers     Smokers
Non-smokers
##  [271] Non-smokers SHS smokers Non-smokers Smokers     SHS smokers
Non-smokers
##  [277] SHS smokers Non-smokers Smokers     SHS smokers Smokers
Non-smokers
##  [283] SHS smokers Smokers     Smokers     Non-smokers Non-smokers SHS
smokers
##  [289] Non-smokers Non-smokers SHS smokers Smokers     Non-smokers Smokers
##  [295] Non-smokers SHS smokers Non-smokers Smokers     Non-smokers Smokers
##  [301] Smokers     Non-smokers Smokers     SHS smokers Non-smokers
Non-smokers
##  [307] Non-smokers SHS smokers Non-smokers SHS smokers Non-smokers Smokers
##  [313] Smokers     Non-smokers Smokers     Smokers     Smokers     SHS
smokers
##  [319] Non-smokers Non-smokers Smokers     Non-smokers SHS smokers
Non-smokers
##  [325] Non-smokers Smokers     Smokers     Smokers     Smokers     SHS
smokers
##  [331] SHS smokers Non-smokers Smokers     Smokers     SHS smokers Smokers
##  [337] Non-smokers Smokers     Non-smokers Non-smokers Non-smokers
Non-smokers
##  [343] Smokers     Non-smokers Non-smokers Smokers     Smokers     Smokers
##  [349] Non-smokers Smokers     SHS smokers Non-smokers Smokers
Non-smokers
##  [355] Non-smokers Non-smokers Smokers     Non-smokers Non-smokers Smokers
##  [361] Non-smokers Smokers     Smokers     Non-smokers Non-smokers SHS
smokers
##  [367] SHS smokers Non-smokers Non-smokers Non-smokers Smokers     SHS
smokers
##  [373] Smokers     Non-smokers Non-smokers SHS smokers Non-smokers
Non-smokers
##  [379] Non-smokers Smokers     Non-smokers Non-smokers Smokers
Non-smokers
##  [385] Smokers     SHS smokers Non-smokers Smokers     Smokers     Smokers
##  [391] Non-smokers Non-smokers Non-smokers SHS smokers Non-smokers Smokers
##  [397] Smokers     Non-smokers Non-smokers Non-smokers Non-smokers Smokers
```

```
##  [403] SHS smokers Smokers     Non-smokers Smokers     Non-smokers SHS
smokers
##  [409] SHS smokers Non-smokers Smokers     Smokers     Smokers
Non-smokers
##  [415] Non-smokers Smokers     Non-smokers Non-smokers Smokers     Smokers
##  [421] Non-smokers Smokers     Non-smokers SHS smokers SHS smokers SHS
smokers
##  [427] Non-smokers Smokers     Non-smokers Smokers     Smokers
Non-smokers
##  [433] Non-smokers Non-smokers Smokers     Smokers     Non-smokers SHS
smokers
##  [439] Non-smokers SHS smokers SHS smokers Smokers     SHS smokers
Non-smokers
##  [445] SHS smokers Smokers     SHS smokers Non-smokers Smokers     Smokers
##  [451] Non-smokers Smokers     Non-smokers Smokers     Smokers     SHS
smokers
##  [457] SHS smokers Non-smokers Smokers     Non-smokers Non-smokers
Non-smokers
##  [463] Non-smokers Smokers     Non-smokers Non-smokers Smokers
Non-smokers
##  [469] Smokers     Smokers     SHS smokers Non-smokers Non-smokers
Non-smokers
##  [475] Non-smokers Non-smokers Smokers     Smokers     Non-smokers Smokers
##  [481] Non-smokers Smokers     Non-smokers Smokers     Non-smokers
Non-smokers
##  [487] Non-smokers Non-smokers Smokers     Smokers     Non-smokers Smokers
##  [493] Smokers     Non-smokers Non-smokers Smokers     Non-smokers
Non-smokers
##  [499] Non-smokers Smokers     Non-smokers Smokers     Non-smokers
Non-smokers
##  [505] Non-smokers Non-smokers Non-smokers Non-smokers Non-smokers
Non-smokers
##  [511] Non-smokers Non-smokers Smokers     Non-smokers Non-smokers
Non-smokers
##  [517] Smokers     Smokers     Non-smokers Smokers     Non-smokers Smokers
##  [523] Smokers     Non-smokers Non-smokers Smokers     Smokers     Smokers
##  [529] Non-smokers Smokers     Non-smokers Non-smokers Non-smokers Smokers
##  [535] Non-smokers Smokers     Non-smokers Non-smokers Non-smokers
Non-smokers
##  [541] Non-smokers Smokers     Non-smokers Non-smokers Non-smokers SHS
smokers
##  [547] Smokers     Non-smokers Smokers     Non-smokers Non-smokers
Non-smokers
##  [553] Non-smokers Non-smokers Non-smokers Smokers     SHS smokers Smokers
##  [559] Non-smokers Non-smokers Non-smokers Non-smokers Non-smokers
Non-smokers
##  [565] Smokers     Smokers     Non-smokers Non-smokers Smokers     SHS
smokers
##  [571] SHS smokers SHS smokers Non-smokers Non-smokers Non-smokers Smokers
##  [577] Non-smokers Non-smokers Non-smokers Non-smokers Non-smokers
```

```
Non-smokers
##  [583] Smokers     Smokers     SHS smokers SHS smokers Non-smokers SHS
smokers
##  [589] Non-smokers Non-smokers Non-smokers Smokers     Smokers
Non-smokers
##  [595] Non-smokers Non-smokers Smokers     Non-smokers Smokers
Non-smokers
##  [601] Non-smokers SHS smokers Smokers     Smokers     Smokers
Non-smokers
##  [607] Smokers     Non-smokers Non-smokers Non-smokers Non-smokers
Non-smokers
##  [613] SHS smokers SHS smokers Non-smokers Non-smokers Smokers     Smokers
##  [619] Non-smokers Non-smokers Smokers     Smokers     Non-smokers
Non-smokers
##  [625] Non-smokers Smokers     Non-smokers SHS smokers Non-smokers
Non-smokers
##  [631] Non-smokers Non-smokers Non-smokers Non-smokers Non-smokers
Non-smokers
##  [637] Smokers     Non-smokers Non-smokers Non-smokers Non-smokers
Non-smokers
##  [643] Non-smokers Smokers     Non-smokers Non-smokers SHS smokers
Non-smokers
##  [649] Non-smokers Smokers     SHS smokers Smokers     Non-smokers
Non-smokers
##  [655] Non-smokers Smokers     Non-smokers Non-smokers Non-smokers
Non-smokers
##  [661] Non-smokers Non-smokers Smokers     Smokers     Non-smokers
Non-smokers
##  [667] Non-smokers Non-smokers Smokers     Non-smokers SHS smokers Smokers
##  [673] Non-smokers Non-smokers Non-smokers SHS smokers Non-smokers SHS
smokers
##  [679] Non-smokers Non-smokers Non-smokers Non-smokers Non-smokers SHS
smokers
##  [685] Non-smokers Non-smokers SHS smokers Smokers     Non-smokers Smokers
##  [691] Smokers     Non-smokers Smokers     Smokers     Non-smokers SHS
smokers
##  [697] Non-smokers Non-smokers SHS smokers Non-smokers Smokers     Smokers
##  [703] Non-smokers SHS smokers Non-smokers Non-smokers Smokers
Non-smokers
##  [709] Non-smokers Smokers     Non-smokers Non-smokers Smokers
Non-smokers
##  [715] Non-smokers Non-smokers Non-smokers Smokers     Smokers
Non-smokers
##  [721] SHS smokers Non-smokers SHS smokers Non-smokers Non-smokers
Non-smokers
##  [727] Smokers     Smokers     Non-smokers SHS smokers SHS smokers
Non-smokers
##  [733] SHS smokers Non-smokers Non-smokers Non-smokers Smokers
Non-smokers
##  [739] Non-smokers Smokers     SHS smokers SHS smokers Non-smokers
```

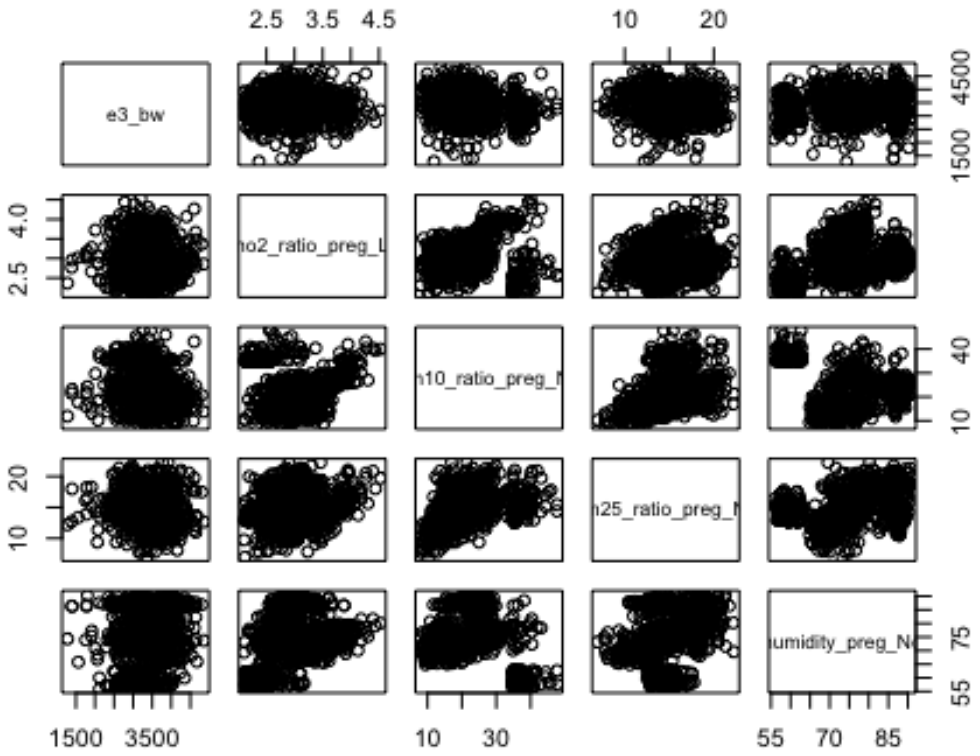
```
Non-smokers
##  [745] Non-smokers Smokers     Smokers     Non-smokers Non-smokers
Non-smokers
##  [751] SHS smokers Non-smokers Non-smokers Non-smokers SHS smokers SHS
smokers
##  [757] Non-smokers Non-smokers Non-smokers Smokers     Non-smokers
Non-smokers
##  [763] Non-smokers SHS smokers Non-smokers Non-smokers Non-smokers
Non-smokers
##  [769] Non-smokers Non-smokers SHS smokers Smokers     Smokers
Non-smokers
##  [775] Non-smokers Non-smokers Non-smokers Non-smokers Non-smokers
Non-smokers
##  [781] Smokers     Non-smokers Non-smokers Non-smokers Non-smokers Smokers
##  [787] Non-smokers Non-smokers SHS smokers Non-smokers Smokers
Non-smokers
##  [793] Smokers     Non-smokers Non-smokers Non-smokers SHS smokers
Non-smokers
##  [799] Non-smokers Non-smokers Non-smokers SHS smokers SHS smokers
Non-smokers
##  [805] Non-smokers Smokers     Non-smokers Non-smokers Non-smokers
Non-smokers
##  [811] Non-smokers Smokers     Smokers     Non-smokers Non-smokers
Non-smokers
##  [817] Non-smokers SHS smokers Smokers     Non-smokers Non-smokers
Non-smokers
##  [823] Non-smokers Non-smokers Smokers     Non-smokers Non-smokers
Non-smokers
##  [829] Non-smokers Non-smokers Non-smokers Non-smokers Non-smokers Smokers
##  [835] Non-smokers Smokers     Non-smokers Non-smokers Smokers     Smokers
##  [841] Non-smokers Smokers     Non-smokers Smokers     SHS smokers
Non-smokers
##  [847] Non-smokers Smokers     Smokers     SHS smokers Non-smokers Smokers
##  [853] Non-smokers SHS smokers Smokers     Non-smokers Smokers
Non-smokers
##  [859] Non-smokers Non-smokers SHS smokers Smokers     Smokers
Non-smokers
##  [865] Non-smokers Non-smokers Smokers     Smokers     SHS smokers
Non-smokers
##  [871] Non-smokers Non-smokers SHS smokers Non-smokers Non-smokers
Non-smokers
##  [877] Non-smokers SHS smokers Non-smokers Non-smokers SHS smokers SHS
smokers
##  [883] Smokers     Non-smokers Non-smokers Smokers     Non-smokers
Non-smokers
##  [889] Smokers     Non-smokers Non-smokers Non-smokers Smokers
Non-smokers
##  [895] Smokers     Non-smokers Non-smokers Smokers     Smokers     Smokers
##  [901] Non-smokers Non-smokers Non-smokers Non-smokers Non-smokers
Non-smokers
```

```
##  [907] Smokers     Non-smokers SHS smokers Smokers     Non-smokers Smokers
##  [913] Smokers     Smokers     Non-smokers SHS smokers Non-smokers
Non-smokers
##  [919] Smokers     Smokers     Non-smokers Non-smokers SHS smokers
Non-smokers
##  [925] Non-smokers Smokers     Non-smokers Non-smokers Smokers     SHS
smokers
##  [931] Non-smokers Non-smokers Non-smokers Smokers     Non-smokers
Non-smokers
##  [937] Non-smokers SHS smokers SHS smokers Non-smokers Non-smokers
Non-smokers
##  [943] Non-smokers SHS smokers Non-smokers Smokers     SHS smokers Smokers
##  [949] Non-smokers Non-smokers Non-smokers Smokers     Non-smokers
Non-smokers
##  [955] Non-smokers Non-smokers Non-smokers SHS smokers Non-smokers
Non-smokers
##  [961] SHS smokers Non-smokers Non-smokers Non-smokers Non-smokers
Non-smokers
##  [967] Non-smokers Non-smokers Smokers     Smokers     Non-smokers Smokers
##  [973] Non-smokers Smokers     Smokers     SHS smokers Non-smokers SHS
smokers
##  [979] Non-smokers Non-smokers Non-smokers Non-smokers Non-smokers Smokers
##  [985] Smokers     Smokers     Non-smokers Non-smokers Smokers     Smokers
##  [991] SHS smokers Non-smokers Non-smokers Smokers     Non-smokers Smokers
##  [997] Smokers     Smokers     Smokers     Non-smokers
## Levels: Non-smokers SHS smokers Smokers
```

3.2 Data analysis

Appendix 4

```
pairs(~ e3_bw + h_no2_ratio_preg_Log + h_pm10_ratio_preg_None +
h_pm25_ratio_preg_None + h_humidity_preg_None, data = data)
```



4.1 Automated Model Selection

Appendix 5

```
# minimal model
M0 <- lm(e3_bw ~ 1, data = data)
summary(M0)

##
## Call:
## lm(formula = e3_bw ~ 1, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2098.48  -328.48    11.52   341.52  1471.52
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3378.5       16.1   209.8   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 509.2 on 999 degrees of freedom
```

Appendix 6

```
# full model
Mfull <- lm(e3_bw ~ as.factor(e3_sex_None) * (h_humidity_preg_None +
                                                h_no2_ratio_preg_Log +
                                                h_pm10_ratio_preg_None +
                                                h_pm25_ratio_preg_None) +
              e3_asmokcigd_p_None +
              as.factor(hs_cotinine_mcat_None), data = data)
summary(Mfull)
```

```
##
## Call:
## lm(formula = e3_bw ~ as.factor(e3_sex_None) * (h_humidity_preg_None +
##     h_no2_ratio_preg_Log + h_pm10_ratio_preg_None +
h_pm25_ratio_preg_None) +
##     e3_asmokcigd_p_None + as.factor(hs_cotinine_mcat_None), data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2131.69  -287.98    15.25   312.99  1315.03
##
## Coefficients:
##                                                  Estimate Std. Error t
value
## (Intercept)                                    3565.45121  241.27167
14.778
## as.factor(e3_sex_None)male                       40.18978  328.99865
0.122
## h_humidity_preg_None                             -1.29133    3.12607
-0.413
## h_no2_ratio_preg_Log                             51.96849   52.01320
0.999
## h_pm10_ratio_preg_None                          -11.60422    3.58713
-3.235
## h_pm25_ratio_preg_None                           -3.50976   10.80055
-0.325
## e3_asmokcigd_p_None                             -37.43169    8.84152
-4.234
## as.factor(hs_cotinine_mcat_None)SHS smokers       10.10250   50.35963
0.201
## as.factor(hs_cotinine_mcat_None)Smokers           90.83369   40.28275
2.255
## as.factor(e3_sex_None)male:h_humidity_preg_None    1.45428    4.17573
0.348
## as.factor(e3_sex_None)male:h_no2_ratio_preg_Log  -58.11887   70.39293
-0.826
```

```
## as.factor(e3_sex_None)male:h_pm10_ratio_preg_None    0.04814    4.80346
0.010
## as.factor(e3_sex_None)male:h_pm25_ratio_preg_None   10.97259   14.20123
0.773
##                                                    Pr(>|t|)
## (Intercept)                                         < 2e-16 ***
## as.factor(e3_sex_None)male                          0.90280
## h_humidity_preg_None                                0.67963
## h_no2_ratio_preg_Log                                0.31797
## h_pm10_ratio_preg_None                              0.00126 **
## h_pm25_ratio_preg_None                              0.74528
## e3_asmokcigd_p_None                                2.51e-05 ***
## as.factor(hs_cotinine_mcat_None)SHS smokers         0.84105
## as.factor(hs_cotinine_mcat_None)Smokers             0.02436 *
## as.factor(e3_sex_None)male:h_humidity_preg_None     0.72771
## as.factor(e3_sex_None)male:h_no2_ratio_preg_Log     0.40921
## as.factor(e3_sex_None)male:h_pm10_ratio_preg_None   0.99201
## as.factor(e3_sex_None)male:h_pm25_ratio_preg_None   0.43991
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 492.8 on 987 degrees of freedom
## Multiple R-squared:  0.07475,    Adjusted R-squared:  0.0635
## F-statistic: 6.645 on 12 and 987 DF,  p-value: 1.492e-11
```

Appendix 7

```
# backward selection
system.time({
  Mback <- step(object = Mfull,
                scope = list(lower = M0, upper = Mfull),
                trace = 1,
                direction = "backward")
})
```

```
## Start:  AIC=12413.11
## e3_bw ~ as.factor(e3_sex_None) * (h_humidity_preg_None +
h_no2_ratio_preg_Log +
##     h_pm10_ratio_preg_None + h_pm25_ratio_preg_None) + e3_asmokcigd_p_None
+
##     as.factor(hs_cotinine_mcat_None)
##
##                                                   Df Sum of Sq       RSS
AIC
## - as.factor(e3_sex_None):h_pm10_ratio_preg_None  1        24 239693156
12411
## - as.factor(e3_sex_None):h_humidity_preg_None    1     29456 239722588
12411
## - as.factor(e3_sex_None):h_pm25_ratio_preg_None  1    144979 239838111
12412
```

```
## - as.factor(e3_sex_None):h_no2_ratio_preg_Log    1    165545 239858677
12412
## <none>                                                    239693132
12413
## - as.factor(hs_cotinine_mcat_None)               2   1291350 240984482
12414
## - e3_asmokcigd_p_None                            1   4352751 244045883
12429
##
## Step:  AIC=12411.11
## e3_bw ~ as.factor(e3_sex_None) + h_humidity_preg_None +
h_no2_ratio_preg_Log +
##     h_pm10_ratio_preg_None + h_pm25_ratio_preg_None + e3_asmokcigd_p_None
+
##     as.factor(hs_cotinine_mcat_None) +
as.factor(e3_sex_None):h_humidity_preg_None +
##     as.factor(e3_sex_None):h_no2_ratio_preg_Log +
as.factor(e3_sex_None):h_pm25_ratio_preg_None
##
##                                                 Df Sum of Sq       RSS
AIC
## - as.factor(e3_sex_None):h_humidity_preg_None    1     39710 239732866
12409
## - as.factor(e3_sex_None):h_no2_ratio_preg_Log    1    166921 239860078
12410
## - as.factor(e3_sex_None):h_pm25_ratio_preg_None  1    188858 239882015
12410
## <none>                                                    239693156
12411
## - as.factor(hs_cotinine_mcat_None)               2   1291348 240984505
12412
## - e3_asmokcigd_p_None                            1   4352744 244045901
12427
## - h_pm10_ratio_preg_None                         1   5053817 244746973
12430
##
## Step:  AIC=12409.28
## e3_bw ~ as.factor(e3_sex_None) + h_humidity_preg_None +
h_no2_ratio_preg_Log +
##     h_pm10_ratio_preg_None + h_pm25_ratio_preg_None + e3_asmokcigd_p_None
+
##     as.factor(hs_cotinine_mcat_None) +
as.factor(e3_sex_None):h_no2_ratio_preg_Log +
##     as.factor(e3_sex_None):h_pm25_ratio_preg_None
##
##                                                 Df Sum of Sq       RSS
AIC
## - h_humidity_preg_None                           1     13523 239746389
12407
## - as.factor(e3_sex_None):h_no2_ratio_preg_Log    1    133482 239866348
```

```
12408
## - as.factor(e3_sex_None):h_pm25_ratio_preg_None  1    270351 240003217
12408
## <none>                                                       239732866
12409
## - as.factor(hs_cotinine_mcat_None)               2   1307026 241039893
12411
## - e3_asmokcigd_p_None                            1   4326681 244059547
12425
## - h_pm10_ratio_preg_None                         1   5043318 244776184
12428
##
## Step:  AIC=12407.34
## e3_bw ~ as.factor(e3_sex_None) + h_no2_ratio_preg_Log +
h_pm10_ratio_preg_None +
##     h_pm25_ratio_preg_None + e3_asmokcigd_p_None +
as.factor(hs_cotinine_mcat_None) +
##     as.factor(e3_sex_None):h_no2_ratio_preg_Log +
as.factor(e3_sex_None):h_pm25_ratio_preg_None
##
##                                                 Df Sum of Sq       RSS
AIC
## - as.factor(e3_sex_None):h_no2_ratio_preg_Log    1    134715 239881105
12406
## - as.factor(e3_sex_None):h_pm25_ratio_preg_None  1    274974 240021364
12406
## <none>                                                       239746389
12407
## - as.factor(hs_cotinine_mcat_None)               2   1332133 241078523
12409
## - e3_asmokcigd_p_None                            1   4322869 244069258
12423
## - h_pm10_ratio_preg_None                         1   6708183 246454573
12433
##
## Step:  AIC=12405.9
## e3_bw ~ as.factor(e3_sex_None) + h_no2_ratio_preg_Log +
h_pm10_ratio_preg_None +
##     h_pm25_ratio_preg_None + e3_asmokcigd_p_None +
as.factor(hs_cotinine_mcat_None) +
##     as.factor(e3_sex_None):h_pm25_ratio_preg_None
##
##                                                 Df Sum of Sq       RSS
AIC
## - h_no2_ratio_preg_Log                           1     67017 239948122
12404
## - as.factor(e3_sex_None):h_pm25_ratio_preg_None  1    216130 240097235
12405
## <none>                                                       239881105
12406
```

```
## - as.factor(hs_cotinine_mcat_None)               2   1304140 241185245
12407
## - e3_asmokcigd_p_None                            1   4358871 244239976
12422
## - h_pm10_ratio_preg_None                         1   6818839 246699944
12432
##
## Step:  AIC=12404.18
## e3_bw ~ as.factor(e3_sex_None) + h_pm10_ratio_preg_None +
h_pm25_ratio_preg_None +
##     e3_asmokcigd_p_None + as.factor(hs_cotinine_mcat_None) +
##     as.factor(e3_sex_None):h_pm25_ratio_preg_None
##
##                                                Df Sum of Sq       RSS
AIC
## - as.factor(e3_sex_None):h_pm25_ratio_preg_None  1    208685 240156807
12403
## <none>                                                      239948122
12404
## - as.factor(hs_cotinine_mcat_None)               2   1245721 241193843
12405
## - e3_asmokcigd_p_None                            1   4292934 244241056
12420
## - h_pm10_ratio_preg_None                         1   6869665 246817787
12430
##
## Step:  AIC=12403.05
## e3_bw ~ as.factor(e3_sex_None) + h_pm10_ratio_preg_None +
h_pm25_ratio_preg_None +
##     e3_asmokcigd_p_None + as.factor(hs_cotinine_mcat_None)
##
##                                    Df Sum of Sq       RSS   AIC
## - h_pm25_ratio_preg_None            1     44003 240200810 12401
## <none>                                          240156807 12403
## - as.factor(hs_cotinine_mcat_None)  2   1194677 241351485 12404
## - e3_asmokcigd_p_None               1   4254216 244411023 12419
## - as.factor(e3_sex_None)            1   5083034 245239842 12422
## - h_pm10_ratio_preg_None            1   6905127 247061934 12429
##
## Step:  AIC=12401.23
## e3_bw ~ as.factor(e3_sex_None) + h_pm10_ratio_preg_None +
e3_asmokcigd_p_None +
##     as.factor(hs_cotinine_mcat_None)
##
##                                    Df Sum of Sq       RSS   AIC
## <none>                                          240200810 12401
## - as.factor(hs_cotinine_mcat_None)  2   1429553 241630363 12403
## - e3_asmokcigd_p_None               1   4282393 244483203 12417
## - as.factor(e3_sex_None)            1   5089006 245289817 12420
## - h_pm10_ratio_preg_None            1   7045288 247246098 12428
```

```
##    user  system elapsed
##   0.044   0.002   0.046
```

Appendix 8

```
# stepwise selection
system.time({
  Mstep <- step(object = M0,
                scope = list(lower = M0, upper = Mfull),
                trace = 1,
                direction = "both")
})
```

```
## Start:  AIC=12466.81
## e3_bw ~ 1
##
##                                    Df Sum of Sq       RSS   AIC
## + h_pm10_ratio_preg_None            1   8793991 250263548 12434
## + e3_asmokcigd_p_None               1   5592658 253464881 12447
## + as.factor(e3_sex_None)            1   5308973 253748566 12448
## + h_humidity_preg_None              1   2616477 256441062 12459
## <none>                                          259057540 12467
## + h_pm25_ratio_preg_None            1    130527 258927013 12468
## + as.factor(hs_cotinine_mcat_None)  2    599775 258457764 12468
## + h_no2_ratio_preg_Log              1      6879 259050661 12469
##
## Step:  AIC=12434.27
## e3_bw ~ h_pm10_ratio_preg_None
##
##                                    Df Sum of Sq       RSS   AIC
## + as.factor(e3_sex_None)            1   4987554 245275994 12416
## + e3_asmokcigd_p_None               1   3599179 246664369 12422
## <none>                                          250263548 12434
## + h_humidity_preg_None              1    294013 249969535 12435
## + h_pm25_ratio_preg_None            1    255450 250008098 12435
## + as.factor(hs_cotinine_mcat_None)  2    753021 249510528 12435
## + h_no2_ratio_preg_Log              1      1560 250261988 12436
## - h_pm10_ratio_preg_None            1   8793991 259057540 12467
##
## Step:  AIC=12416.14
## e3_bw ~ h_pm10_ratio_preg_None + as.factor(e3_sex_None)
##
##                                           Df Sum of Sq       RSS
AIC
## + e3_asmokcigd_p_None                      1   3645630 241630363
12403
## <none>                                                 245275994
12416
## + as.factor(hs_cotinine_mcat_None)         2    792791 244483203
12417
```

```
## + h_pm25_ratio_preg_None                             1    251946 245024047
12417
## + h_humidity_preg_None                               1    208459 245067534
12417
## + as.factor(e3_sex_None):h_pm10_ratio_preg_None  1      3162 245272831
12418
## + h_no2_ratio_preg_Log                               1       372 245275621
12418
## - as.factor(e3_sex_None)                             1   4987554 250263548
12434
## - h_pm10_ratio_preg_None                             1   8472573 253748566
12448
##
## Step:  AIC=12403.16
## e3_bw ~ h_pm10_ratio_preg_None + as.factor(e3_sex_None) +
e3_asmokcigd_p_None
##
##                                                 Df Sum of Sq       RSS
AIC
## + as.factor(hs_cotinine_mcat_None)               2   1429553 240200810
12401
## <none>                                                       241630363
12403
## + h_pm25_ratio_preg_None                             1    278879 241351485
12404
## + h_humidity_preg_None                               1    181753 241448610
12404
## + h_no2_ratio_preg_Log                               1     32698 241597666
12405
## + as.factor(e3_sex_None):h_pm10_ratio_preg_None  1       592 241629771
12405
## - e3_asmokcigd_p_None                                1   3645630 245275994
12416
## - as.factor(e3_sex_None)                             1   5034006 246664369
12422
## - h_pm10_ratio_preg_None                             1   6510749 248141113
12428
##
## Step:  AIC=12401.23
## e3_bw ~ h_pm10_ratio_preg_None + as.factor(e3_sex_None) +
e3_asmokcigd_p_None +
##     as.factor(hs_cotinine_mcat_None)
##
##                                                 Df Sum of Sq       RSS
AIC
## <none>                                                       240200810
12401
## + h_no2_ratio_preg_Log                               1     80614 240120196
12403
## + h_pm25_ratio_preg_None                             1     44003 240156807
```

```
12403
## - as.factor(hs_cotinine_mcat_None)               2   1429553 241630363
12403
## + as.factor(e3_sex_None):h_pm10_ratio_preg_None  1      6230 240194580
12403
## + h_humidity_preg_None                           1      3531 240197279
12403
## - e3_asmokcigd_p_None                            1   4282393 244483203
12417
## - as.factor(e3_sex_None)                         1   5089006 245289817
12420
## - h_pm10_ratio_preg_None                         1   7045288 247246098
12428

##    user  system elapsed
##   0.032   0.001   0.033
```

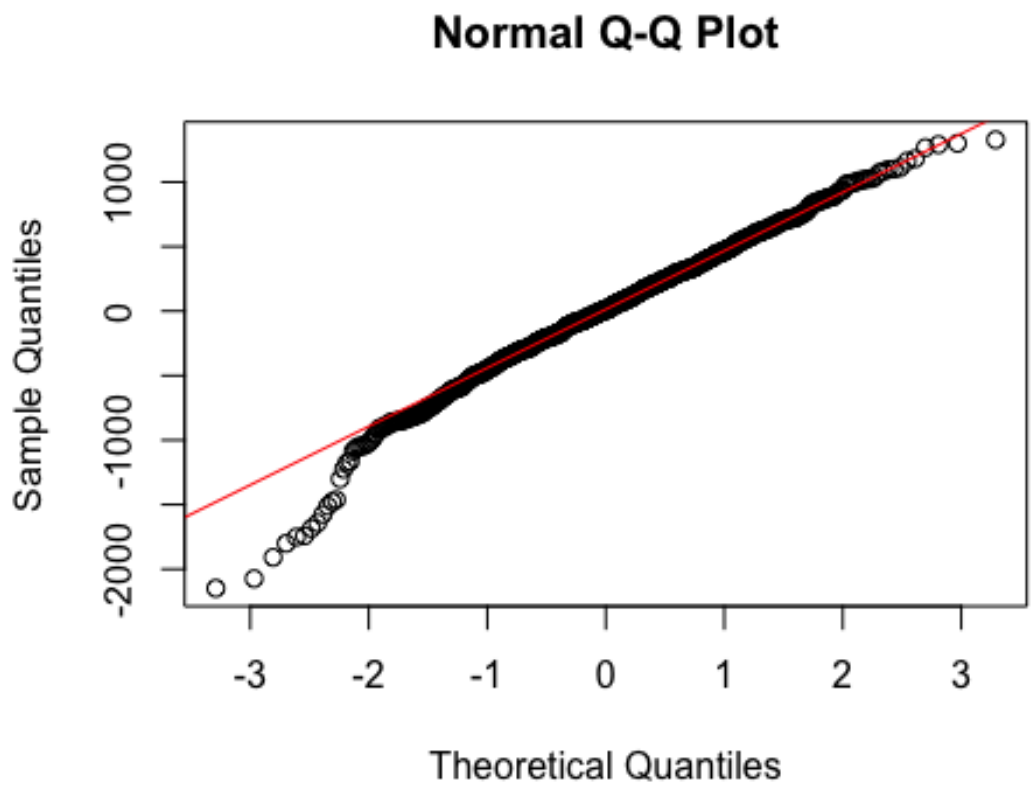
Appendix 9

```
final.model <- lm(e3_bw ~ as.factor(e3_sex_None) + h_pm10_ratio_preg_None +
                    e3_asmokcigd_p_None + as.factor(hs_cotinine_mcat_None),
data = data)
summary(final.model)

##
## Call:
## lm(formula = e3_bw ~ as.factor(e3_sex_None) + h_pm10_ratio_preg_None +
##     e3_asmokcigd_p_None + as.factor(hs_cotinine_mcat_None), data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2148.53  -292.54    11.73   319.50  1327.39
##
## Coefficients:
##                                             Estimate Std. Error t value
## (Intercept)                                 3561.362     51.744  68.827
## as.factor(e3_sex_None)male                   143.015     31.164   4.589
## h_pm10_ratio_preg_None                       -11.170      2.069  -5.400
## e3_asmokcigd_p_None                          -36.652      8.707  -4.210
## as.factor(hs_cotinine_mcat_None)SHS smokers    5.924     49.704   0.119
## as.factor(hs_cotinine_mcat_None)Smokers       86.730     36.923   2.349
##                                             Pr(>|t|)
## (Intercept)                                  < 2e-16 ***
## as.factor(e3_sex_None)male                  5.02e-06 ***
## h_pm10_ratio_preg_None                      8.36e-08 ***
## e3_asmokcigd_p_None                         2.79e-05 ***
## as.factor(hs_cotinine_mcat_None)SHS smokers    0.905
## as.factor(hs_cotinine_mcat_None)Smokers        0.019 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 491.6 on 994 degrees of freedom
## Multiple R-squared:  0.07279,    Adjusted R-squared:  0.06813
## F-statistic: 15.61 on 5 and 994 DF,  p-value: 8.313e-15
```

Appendix 10

```
qqnorm(residuals(final.model))
qqline(residuals(final.model), col = "red")
```



Appendix 11

```
# female model
female.data <- data[data$e3_sex_None == "female",]
female.final.model <- lm(e3_bw ~ h_pm10_ratio_preg_None + e3_asmokcigd_p_None
+
                           as.factor(hs_cotinine_mcat_None), data =
female.data)
summary(female.final.model)

##
## Call:
```

```
## lm(formula = e3_bw ~ h_pm10_ratio_preg_None + e3_asmokcigd_p_None +
##     as.factor(hs_cotinine_mcat_None), data = female.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2154.85  -276.77    10.97   312.91  1249.87
##
## Coefficients:
##                                            Estimate Std. Error t value
## (Intercept)                                3565.483     67.736  52.638
## h_pm10_ratio_preg_None                      -10.985      2.876  -3.820
## e3_asmokcigd_p_None                         -40.590     12.371  -3.281
## as.factor(hs_cotinine_mcat_None)SHS smokers  -67.226     70.896  -0.948
## as.factor(hs_cotinine_mcat_None)Smokers       94.785     52.057   1.821
##                                            Pr(>|t|)
## (Intercept)                                 < 2e-16 ***
## h_pm10_ratio_preg_None                     0.000152 ***
## e3_asmokcigd_p_None                        0.001112 **
## as.factor(hs_cotinine_mcat_None)SHS smokers 0.343504
## as.factor(hs_cotinine_mcat_None)Smokers     0.069277 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 472.1 on 466 degrees of freedom
## Multiple R-squared:  0.06731,    Adjusted R-squared:  0.05931
## F-statistic: 8.408 on 4 and 466 DF,  p-value: 1.482e-06
```

```
# male model
male.data <- data[data$e3_sex_None == "male",]
male.final.model <- lm(e3_bw ~ h_pm10_ratio_preg_None + e3_asmokcigd_p_None +
                         as.factor(hs_cotinine_mcat_None), data =
male.data)
summary(male.final.model)
```

```
##
## Call:
## lm(formula = e3_bw ~ h_pm10_ratio_preg_None + e3_asmokcigd_p_None +
##     as.factor(hs_cotinine_mcat_None), data = male.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2066.82  -295.58    14.99   317.98  1334.33
##
## Coefficients:
##                                            Estimate Std. Error t value
## (Intercept)                                3704.145     69.717  53.131
## h_pm10_ratio_preg_None                      -11.493      2.971  -3.869
## e3_asmokcigd_p_None                         -33.854     12.243  -2.765
## as.factor(hs_cotinine_mcat_None)SHS smokers   65.385     69.692   0.938
## as.factor(hs_cotinine_mcat_None)Smokers       80.555     52.301   1.540
```

```
##                                                Pr(>|t|)
## (Intercept)                                     < 2e-16 ***
## h_pm10_ratio_preg_None                         0.000123 ***
## e3_asmokcigd_p_None                            0.005888 **
## as.factor(hs_cotinine_mcat_None)SHS smokers    0.348580
## as.factor(hs_cotinine_mcat_None)Smokers        0.124112
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 509.1 on 524 degrees of freedom
## Multiple R-squared:  0.04637,    Adjusted R-squared:  0.03909
## F-statistic:  6.37 on 4 and 524 DF,  p-value: 5.199e-05
```

```
# variance model based on child sex
variance.model <- aov(e3_bw ~ e3_sex_None, data = data)
summary(variance.model)
```

```
##              Df    Sum Sq Mean Sq F value  Pr(>F)
## e3_sex_None   1   5308973 5308973   20.88 5.5e-06 ***
## Residuals   998 253748566  254257
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

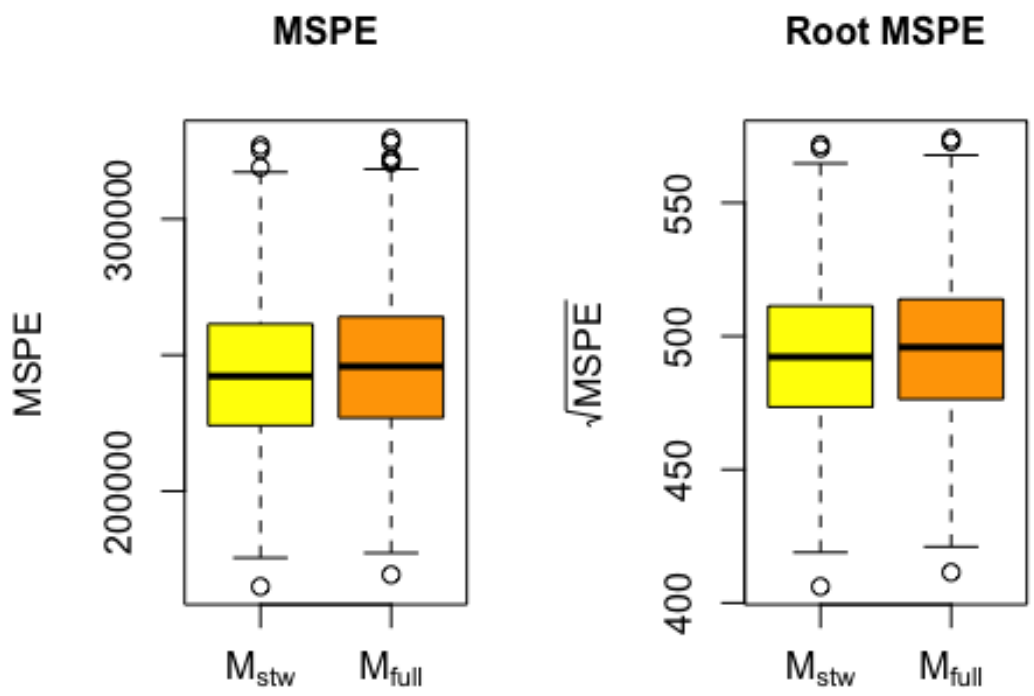
Appendix 12

```
Mstep <- lm(e3_bw ~ as.factor(e3_sex_None) + h_pm10_ratio_preg_None +
                    e3_asmokcigd_p_None + as.factor(hs_cotinine_mcat_None),
data = data)

testdata <- dataset

M1 <- Mstep
M2 <- Mfull
Mnames <- expression(M[stw], M[full])

nreps <- 1e3
ntot <- nrow(testdata)
ntrain <- 800
ntest <- ntot-ntrain

mspe1 <- rep(NA, nreps)
mspe2 <- rep(NA, nreps)

system.time({
  for(ii in 1:nreps) {
    if(ii%%100 == 0) message("ii = ", ii)

    train.ind <- sample(ntot, ntrain)
    M1.cv <- update(M1, subset = train.ind)
```

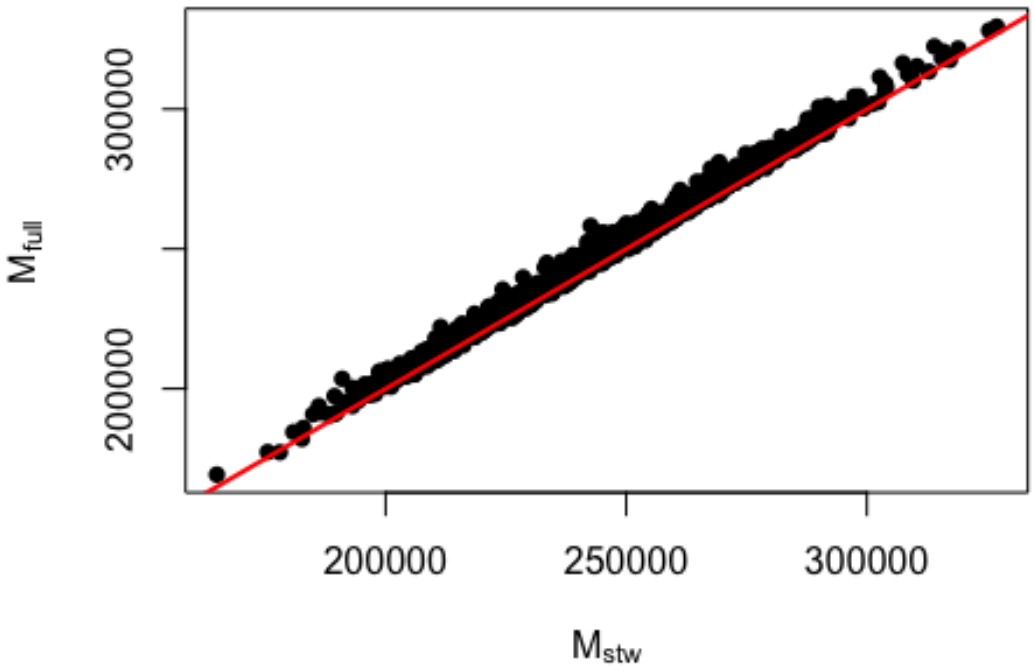
```
    M2.cv <- update(M2, subset = train.ind)

    M1.res <- testdata$e3_bw[-train.ind] -
      predict(M1.cv, newdata = testdata[-train.ind,])
    M2.res <- testdata$e3_bw[-train.ind] -predict(M2.cv, newdata =
testdata[-train.ind,])

    mspe1[ii] <- mean(M1.res^2)
    mspe2[ii] <- mean(M2.res^2)

  }
})
```

```
## ii = 100

## ii = 200

## ii = 300

## ii = 400

## ii = 500

## ii = 600

## ii = 700

## ii = 800

## ii = 900

## ii = 1000

##    user  system elapsed
##   6.044   0.060   6.109
```

```
par(mfrow = c(1,2))
cex <- 1
boxplot(x = list(mspe1, mspe2), names = Mnames,
        main = "MSPE",
        ylab = expression(MSPE),
        col = c("yellow", "orange"),
        cex = cex, cex.lab = cex, cex.axis = cex, cex.main = cex)
boxplot(x = list(sqrt(mspe1), sqrt(mspe2)), names = Mnames,
        main = "Root MSPE",
        ylab = expression(sqrt(MSPE)),
        col = c("yellow", "orange"),
        cex = cex, cex.lab = cex, cex.axis = cex, cex.main = cex)
```




```
par(mfrow=c(1,1))
plot(mspe1, mspe2, pch = 16,
     xlab = Mnames[1], ylab = Mnames[2],
     main = "")
abline(a = 0, b = 1, col= "red", lwd = 2)
```




```
PRESS1 <- resid(M1)/(1-hatvalues(M1))
PRESS2 <- resid(M2)/(1-hatvalues(M2))

#mean square prediction error
mean(PRESS1^2)

## [1] 243003.8

mean(PRESS2^2)

## [1] 245799.3
```