Assignment 10

Chee Kay Cheong

knitr::opts\_chunk$set(message = FALSE, warning = FALSE)  
  
library(tidyverse)  
library(caret)  
library(glmnet)

# Part 1

### Data Exploratory Analyses

# Load data file  
load("./exposome.RData")  
  
# Put all data frames into a list  
df\_list = list(covariates, exposome, phenotype)   
  
# Merge all data frames together  
HW10 = df\_list %>%   
 reduce(full\_join, by = 'ID') %>%   
 select(-ID) %>%   
 mutate(hs\_asthma = as\_factor(hs\_asthma))  
  
# Check to see if there is missing values  
miss.data = HW10[!complete.cases(HW10), ]  
# no missing values in the dataset

The dataset HW10 contains 1301 observations and 241 variables.

# Display rows and columns  
dim(HW10)  
  
# Summarize each variable in dataset  
summary(HW10)

Descriptive measures:

* h\_trafload\_preg\_pow1over3 = Total traffic load of all roads in 100m buffer at pregnancy period (numeric)
  + Min : 0.3458
  + 1st Quartile: 33.6542
  + Median : 66.6101
  + Mean : 75.5390
  + 3rd Quartile: 113.0812
  + Max : 294.2705
* e3\_asmokcigd\_p\_None = maternal active tobacco smoke during pregnancy mean number of cigarette/day (numeric)
  + Min : 0.000
  + 1st Quartile: 0.000
  + Median : 0.000
  + Mean : 0.494
  + 3rd Quartile: 0.000
  + Max : 15.238
* h\_pm10\_ratio\_preg\_None = pm10 value (extrapolated back in time using ratio method) during pregnancy (numeric)
  + Min : 8.066
  + 1st Quartile: 17.535
  + Median : 23.018
  + Mean : 23.504
  + 3rd Quartile: 27.677
  + Max : 47.698
* hs\_pm10\_wk\_hs\_h\_None = pm10 value (extrapolated back in time using ratio method) one week before hs test at home (numeric)
  + Min : 5.838
  + 1st Quartile: 19.142
  + Median : 24.891
  + Mean : 26.409
  + 3rd Quartile: 32.131
  + Max : 211.297
* h\_lden\_cat\_preg\_None = Categorized lden (day, evening, night) at pregnancy period (numeric)
  + Min : 33.92
  + 1st Quartile: 50.00
  + Median : 58.63
  + Mean : 57.47
  + 3rd Quartile: 64.36
  + Max : 77.40
* hs\_asthma = Doctor diagnosed asthma (ever) (factor)
  + 0 (No) : 1159
  + 1 (Yes) : 142

Examine correlation between variables:

# Create correlation matrix  
round(cor(HW10[c('h\_trafload\_preg\_pow1over3', 'e3\_asmokcigd\_p\_None', 'h\_pm10\_ratio\_preg\_None', 'hs\_pm10\_wk\_hs\_h\_None', 'h\_lden\_cat\_preg\_None')]), 2)

## h\_trafload\_preg\_pow1over3 e3\_asmokcigd\_p\_None  
## h\_trafload\_preg\_pow1over3 1.00 0.03  
## e3\_asmokcigd\_p\_None 0.03 1.00  
## h\_pm10\_ratio\_preg\_None 0.31 0.18  
## hs\_pm10\_wk\_hs\_h\_None 0.08 0.13  
## h\_lden\_cat\_preg\_None 0.32 0.12  
## h\_pm10\_ratio\_preg\_None hs\_pm10\_wk\_hs\_h\_None  
## h\_trafload\_preg\_pow1over3 0.31 0.08  
## e3\_asmokcigd\_p\_None 0.18 0.13  
## h\_pm10\_ratio\_preg\_None 1.00 0.49  
## hs\_pm10\_wk\_hs\_h\_None 0.49 1.00  
## h\_lden\_cat\_preg\_None 0.36 0.11  
## h\_lden\_cat\_preg\_None  
## h\_trafload\_preg\_pow1over3 0.32  
## e3\_asmokcigd\_p\_None 0.12  
## h\_pm10\_ratio\_preg\_None 0.36  
## hs\_pm10\_wk\_hs\_h\_None 0.11  
## h\_lden\_cat\_preg\_None 1.00

# Part 2

### Research question

What factors best predict the risk of asthma in children ages between 6-11 years of age from the HELIX cohort?

# Part 3

Since the HW10 dataset contains more than 200 variables, I decided to perform feature selection using **LASSO** to reduce the number of variables in the dataset while retaining as much of the original variation as possible. This can help simplify further analysis (such as random forest, logistic regression, etc.) and improve computational efficiency.

### LASSO

set.seed(123)  
  
asthma\_lasso = HW10  
  
# Find correlated predictors to remove variables that are highly correlated  
# Correlation can only be done with numeric variables, so need to filter out the non-numeric variables  
asthma.numeric =   
 asthma\_lasso %>%   
 select(where(is.numeric))  
  
# Calculate correlations  
correlations = cor(asthma.numeric, use = "complete.obs")  
  
# Find any features that are correlated at 0.7 and above  
high.correlations = findCorrelation(correlations, cutoff = 0.7)   
  
# Remove highly correlated features  
asthma\_lasso = asthma\_lasso[ , -high.correlations]  
  
# Partition data into 70/30 split  
train.index =   
 asthma\_lasso$hs\_asthma %>% createDataPartition(p = 0.7, list = F)  
  
train\_data = asthma\_lasso[train.index, ]  
test\_data = asthma\_lasso[-train.index, ]

set.seed(123)  
  
control.settings = trainControl(method = "cv", number = 10, sampling = "up")  
  
lambda = 10^seq(-3, -1, length = 100)  
  
LASSO = train(hs\_asthma ~ ., data = train\_data, method = "glmnet", trControl = control.settings, preProcess = c("center", "scale"), tuneGrid = expand.grid(alpha = 1, lambda = lambda))  
  
# Output best value of alpha & lambda  
LASSO$finalModel$tuneValue

## alpha lambda  
## 4 1 0.001149757

LASSO$bestTune

## alpha lambda  
## 4 1 0.001149757

LASSO$results

## alpha lambda Accuracy Kappa AccuracySD KappaSD  
## 1 1 0.001000000 0.7620521 0.03766431 0.05592128 0.12298058  
## 2 1 0.001047616 0.7620640 0.03823912 0.05795745 0.12389156  
## 3 1 0.001097499 0.7620640 0.03743512 0.05583502 0.12195199  
## 4 1 0.001149757 0.7620640 0.03743512 0.05583502 0.12195199  
## 5 1 0.001204504 0.7587793 0.03281534 0.05261207 0.11749010  
## 6 1 0.001261857 0.7587793 0.03281534 0.05261207 0.11749010  
## 7 1 0.001321941 0.7587793 0.03281534 0.05261207 0.11749010  
## 8 1 0.001384886 0.7587673 0.03725714 0.05115243 0.12117684  
## 9 1 0.001450829 0.7587673 0.03676052 0.04955361 0.12004220  
## 10 1 0.001519911 0.7587673 0.03676052 0.04955361 0.12004220  
## 11 1 0.001592283 0.7587673 0.03590821 0.04647981 0.11821030  
## 12 1 0.001668101 0.7587673 0.03662814 0.04647981 0.12065954  
## 13 1 0.001747528 0.7576684 0.03511325 0.04545758 0.11930914  
## 14 1 0.001830738 0.7521739 0.03073759 0.04515822 0.12146542  
## 15 1 0.001917910 0.7499761 0.03642578 0.04506770 0.11979035  
## 16 1 0.002009233 0.7477783 0.03508429 0.04632896 0.12089585  
## 17 1 0.002104904 0.7466794 0.03434364 0.04761378 0.12148666  
## 18 1 0.002205131 0.7499642 0.03704795 0.04396016 0.12220085  
## 19 1 0.002310130 0.7488653 0.03621116 0.04477570 0.12256472  
## 20 1 0.002420128 0.7455566 0.04128161 0.04318621 0.11334990  
## 21 1 0.002535364 0.7477664 0.06381094 0.04435320 0.09903199  
## 22 1 0.002656088 0.7477664 0.06381094 0.04435320 0.09903199  
## 23 1 0.002782559 0.7499522 0.07540264 0.04023222 0.10006264  
## 24 1 0.002915053 0.7488653 0.07973207 0.04008711 0.09999918  
## 25 1 0.003053856 0.7466794 0.07069202 0.03759744 0.10033777  
## 26 1 0.003199267 0.7455686 0.07585855 0.03951986 0.11369727  
## 27 1 0.003351603 0.7455805 0.07564124 0.04008576 0.11187366  
## 28 1 0.003511192 0.7433827 0.07440199 0.04283516 0.11426186  
## 29 1 0.003678380 0.7455566 0.07741568 0.04404920 0.11906239  
## 30 1 0.003853529 0.7466794 0.07664015 0.03995238 0.11045232  
## 31 1 0.004037017 0.7477664 0.07775623 0.03955601 0.11114202  
## 32 1 0.004229243 0.7466794 0.07672912 0.04094750 0.11009278  
## 33 1 0.004430621 0.7445055 0.07508395 0.04559568 0.10895016  
## 34 1 0.004641589 0.7466914 0.08269569 0.04800571 0.12245094  
## 35 1 0.004862602 0.7445055 0.07958754 0.04618048 0.11762396  
## 36 1 0.005094138 0.7423077 0.07721261 0.04509264 0.11658549  
## 37 1 0.005336699 0.7390110 0.06774139 0.04134799 0.11144638  
## 38 1 0.005590810 0.7335165 0.06310345 0.04215144 0.11340018  
## 39 1 0.005857021 0.7345915 0.06506486 0.04436845 0.11909345  
## 40 1 0.006135907 0.7367893 0.06738874 0.04440000 0.12067902  
## 41 1 0.006428073 0.7367893 0.07296608 0.04165411 0.10968357  
## 42 1 0.006734151 0.7356785 0.07225320 0.04464391 0.11210385  
## 43 1 0.007054802 0.7356785 0.06621984 0.04670038 0.12120586  
## 44 1 0.007390722 0.7356904 0.06481907 0.04400308 0.11566355  
## 45 1 0.007742637 0.7367893 0.07899324 0.04265715 0.10599273  
## 46 1 0.008111308 0.7367893 0.08415398 0.04137984 0.10612922  
## 47 1 0.008497534 0.7335045 0.08098763 0.04199324 0.10536686  
## 48 1 0.008902151 0.7291328 0.07726159 0.04641400 0.10389287  
## 49 1 0.009326033 0.7280459 0.07725377 0.04484028 0.08852269  
## 50 1 0.009770100 0.7291448 0.08349597 0.04337953 0.08276969  
## 51 1 0.010235310 0.7291448 0.08899246 0.04244147 0.07462338  
## 52 1 0.010722672 0.7247492 0.08609503 0.04683274 0.08672966  
## 53 1 0.011233240 0.7258481 0.09175435 0.04640037 0.08847309  
## 54 1 0.011768120 0.7236622 0.09002407 0.04936004 0.08754807  
## 55 1 0.012328467 0.7225753 0.07950537 0.05504903 0.10190603  
## 56 1 0.012915497 0.7182035 0.07441975 0.05225394 0.09781968  
## 57 1 0.013530478 0.7149068 0.07137308 0.05275691 0.09670576  
## 58 1 0.014174742 0.7159938 0.08206883 0.05290356 0.11074461  
## 59 1 0.014849683 0.7160177 0.08989758 0.04801354 0.11272225  
## 60 1 0.015556761 0.7116340 0.08567419 0.04676996 0.11169541  
## 61 1 0.016297508 0.7072504 0.08223743 0.04814540 0.11248775  
## 62 1 0.017073526 0.7039537 0.08331000 0.04790215 0.11674020  
## 63 1 0.017886495 0.7006570 0.08491565 0.04325977 0.10418947  
## 64 1 0.018738174 0.6995581 0.08882185 0.04454635 0.10980107  
## 65 1 0.019630407 0.6929766 0.08345228 0.04567111 0.11016669  
## 66 1 0.020565123 0.6929646 0.08850297 0.04398825 0.10390739  
## 67 1 0.021544347 0.6896799 0.08522461 0.04180067 0.10116206  
## 68 1 0.022570197 0.6874701 0.09274924 0.04419477 0.10917439  
## 69 1 0.023644894 0.6743192 0.08266869 0.04791376 0.10785315  
## 70 1 0.024770764 0.6677616 0.07823625 0.04747793 0.09295086  
## 71 1 0.025950242 0.6644888 0.07554696 0.04822719 0.09084482  
## 72 1 0.027185882 0.6568204 0.07546261 0.05658045 0.11265751  
## 73 1 0.028480359 0.6590062 0.08581943 0.05577414 0.10710672  
## 74 1 0.029836472 0.6524247 0.07785848 0.05941206 0.10864579  
## 75 1 0.031257158 0.6535475 0.07826422 0.05857914 0.10709337  
## 76 1 0.032745492 0.6502508 0.07677885 0.05730826 0.09909171  
## 77 1 0.034304693 0.6447683 0.07738596 0.05886050 0.09466659  
## 78 1 0.035938137 0.6370640 0.06315366 0.05438447 0.08044683  
## 79 1 0.037649358 0.6326804 0.06733978 0.05177762 0.07526547  
## 80 1 0.039442061 0.6337315 0.09063433 0.05289181 0.07634581  
## 81 1 0.041320124 0.6249881 0.08451257 0.05604167 0.07332067  
## 82 1 0.043287613 0.6205925 0.09279247 0.05878014 0.08580362  
## 83 1 0.045348785 0.6073937 0.08635125 0.05011109 0.07817548  
## 84 1 0.047508102 0.6008003 0.08241943 0.05165033 0.07951898  
## 85 1 0.049770236 0.5963927 0.07957193 0.05497127 0.09422231  
## 86 1 0.052140083 0.5875896 0.07402639 0.05809481 0.09602891  
## 87 1 0.054622772 0.5788223 0.06691510 0.05278435 0.09153980  
## 88 1 0.057223677 0.5667463 0.05967173 0.05212028 0.08069899  
## 89 1 0.059948425 0.5612279 0.05840707 0.05855738 0.07619570  
## 90 1 0.062802914 0.5480769 0.05432290 0.06599173 0.06402579  
## 91 1 0.065793322 0.5304945 0.05020759 0.08408703 0.05033402  
## 92 1 0.068926121 0.5107740 0.05928616 0.09938272 0.06365523  
## 93 1 0.072208090 0.4943024 0.05359158 0.10792276 0.05948882  
## 94 1 0.075646333 0.4712255 0.05197287 0.11791158 0.04008573  
## 95 1 0.079248290 0.4580268 0.06354568 0.12603283 0.03828984  
## 96 1 0.083021757 0.4283564 0.05309763 0.13392187 0.03783152  
## 97 1 0.086974900 0.4108337 0.04871237 0.13468773 0.02806100  
## 98 1 0.091116276 0.3421166 0.03810539 0.09681345 0.02658727  
## 99 1 0.095454846 0.3278667 0.03705385 0.08936116 0.03188781  
## 100 1 0.100000000 0.3103321 0.03246038 0.07025823 0.03399022

confusionMatrix(LASSO)

## Cross-Validated (10 fold) Confusion Matrix   
##   
## (entries are percentual average cell counts across resamples)  
##   
## Reference  
## Prediction 0 1  
## 0 73.8 8.6  
## 1 15.2 2.4  
##   
## Accuracy (average) : 0.7621

coef(LASSO$finalModel, LASSO$bestTune$lambda)

## 259 x 1 sparse Matrix of class "dgCMatrix"  
## s1  
## (Intercept) -1.398201434  
## h\_cohort2 .   
## h\_cohort3 -2.186715123  
## h\_cohort4 0.549602762  
## h\_cohort5 .   
## h\_cohort6 -0.941960434  
## e3\_sex\_Nonemale 0.540060196  
## e3\_yearbir\_None2004 0.023559988  
## e3\_yearbir\_None2005 -0.061508824  
## e3\_yearbir\_None2006 -0.082461288  
## e3\_yearbir\_None2007 -0.090243454  
## e3\_yearbir\_None2008 .   
## e3\_yearbir\_None2009 0.191258857  
## h\_mbmi\_None -0.201547671  
## h\_age\_None -0.553272992  
## h\_edumc\_None2 .   
## h\_edumc\_None3 0.611860890  
## h\_native\_None1 0.103172573  
## h\_native\_None2 0.029810536  
## h\_parity\_None1 0.323186877  
## h\_parity\_None2 0.249303225  
## hs\_child\_age\_None -0.564530238  
## hs\_c\_height\_None 0.301201491  
## h\_no2\_ratio\_preg\_Log -0.720746030  
## hs\_no2\_dy\_hs\_h\_Log -0.061680032  
## hs\_no2\_yr\_hs\_h\_Log .   
## hs\_pm10\_dy\_hs\_h\_None -0.649307277  
## hs\_pm10\_wk\_hs\_h\_None 0.097812543  
## hs\_pm10\_yr\_hs\_h\_None -0.586638919  
## hs\_pm25\_yr\_hs\_h\_None -0.127154938  
## hs\_pm25abs\_dy\_hs\_h\_Log 0.715761570  
## hs\_pm25abs\_yr\_hs\_h\_Log -0.500852614  
## h\_accesslines300\_preg\_dic0 -0.751848812  
## h\_accesspoints300\_preg\_Log 0.142848830  
## h\_builtdens300\_preg\_Sqrt -0.839199772  
## h\_fdensity300\_preg\_Log 0.611961659  
## h\_frichness300\_preg\_None .   
## h\_landuseshan300\_preg\_None -0.270700281  
## h\_popdens\_preg\_Sqrt -0.250170840  
## h\_walkability\_mean\_preg\_None 0.093858261  
## hs\_accesslines300\_h\_dic0 0.837664907  
## hs\_accesspoints300\_h\_Log -0.587775715  
## hs\_builtdens300\_h\_Sqrt -0.096984529  
## hs\_connind300\_h\_Log .   
## hs\_fdensity300\_h\_Log 0.768196824  
## hs\_landuseshan300\_h\_None 0.794717116  
## hs\_popdens\_h\_Sqrt -0.372411711  
## hs\_walkability\_mean\_h\_None .   
## hs\_accesslines300\_s\_dic0 -0.128366645  
## hs\_accesspoints300\_s\_Log 0.095886226  
## hs\_builtdens300\_s\_Sqrt -0.162516265  
## hs\_connind300\_s\_Log 0.209609527  
## hs\_fdensity300\_s\_Log 0.266588273  
## hs\_landuseshan300\_s\_None 0.059931743  
## hs\_popdens\_s\_Sqrt 0.235213685  
## h\_Absorbance\_Log .   
## h\_Benzene\_Log 0.235783059  
## h\_NO2\_Log .   
## h\_PM\_Log -0.070448917  
## h\_TEX\_Log -0.129945099  
## e3\_alcpreg\_yn\_None1 -0.122877971  
## h\_bfdur\_Ter(10.8,34.9] 0.026227307  
## h\_bfdur\_Ter(34.9,Inf] 0.304714391  
## h\_cereal\_preg\_Ter(9,27.3] -0.007644304  
## h\_cereal\_preg\_Ter(27.3,Inf] .   
## h\_dairy\_preg\_Ter(17.1,27.1] 0.397806973  
## h\_dairy\_preg\_Ter(27.1,Inf] -0.396544783  
## h\_fastfood\_preg\_Ter(0.25,0.83] -0.255751772  
## h\_fastfood\_preg\_Ter(0.83,Inf] -0.287381994  
## h\_fish\_preg\_Ter(1.9,4.1] 0.797618839  
## h\_fish\_preg\_Ter(4.1,Inf] 0.204446293  
## h\_folic\_t1\_None1 0.186668519  
## h\_fruit\_preg\_Ter(0.6,18.2] -0.345769605  
## h\_fruit\_preg\_Ter(18.2,Inf] .   
## h\_legume\_preg\_Ter(0.5,2] .   
## h\_legume\_preg\_Ter(2,Inf] -0.592617810  
## h\_meat\_preg\_Ter(6.5,10] 0.010040688  
## h\_meat\_preg\_Ter(10,Inf] 0.681916368  
## h\_pamod\_t3\_NoneOften 0.004043275  
## h\_pamod\_t3\_NoneSometimes 0.326416672  
## h\_pamod\_t3\_NoneVery Often .   
## h\_pavig\_t3\_NoneLow .   
## h\_pavig\_t3\_NoneMedium -0.176830841  
## h\_veg\_preg\_Ter(8.8,16.5] .   
## h\_veg\_preg\_Ter(16.5,Inf] -0.248911747  
## hs\_bakery\_prod\_Ter(2,6] 0.248643887  
## hs\_bakery\_prod\_Ter(6,Inf] 0.934232714  
## hs\_beverages\_Ter(0.132,1] 0.847467773  
## hs\_beverages\_Ter(1,Inf] 0.100829801  
## hs\_break\_cer\_Ter(1.1,5.5] 0.180905354  
## hs\_break\_cer\_Ter(5.5,Inf] 0.487708239  
## hs\_caff\_drink\_Ter(0.132,Inf] .   
## hs\_org\_food\_Ter(0.132,1] -0.479486904  
## hs\_org\_food\_Ter(1,Inf] -0.492686125  
## hs\_pet\_cat\_r2\_None1 -0.861918683  
## hs\_pet\_dog\_r2\_None1 0.017129110  
## hs\_total\_bread\_Ter(7,17.5] .   
## hs\_total\_bread\_Ter(17.5,Inf] .   
## hs\_total\_cereal\_Ter(14.1,23.6] 0.101333751  
## hs\_total\_cereal\_Ter(23.6,Inf] -0.033204354  
## hs\_total\_fish\_Ter(1.5,3] -0.053209683  
## hs\_total\_fish\_Ter(3,Inf] -0.092535700  
## hs\_total\_fruits\_Ter(7,14.1] -0.848865387  
## hs\_total\_fruits\_Ter(14.1,Inf] -0.716160978  
## hs\_total\_lipids\_Ter(3,7] 0.215732146  
## hs\_total\_lipids\_Ter(7,Inf] -0.310190987  
## hs\_total\_meat\_Ter(6,9] 0.035284395  
## hs\_total\_meat\_Ter(9,Inf] 0.356767587  
## hs\_total\_potatoes\_Ter(3,4] -0.311488775  
## hs\_total\_potatoes\_Ter(4,Inf] -0.796502699  
## hs\_total\_sweets\_Ter(4.1,8.5] 0.626097090  
## hs\_total\_sweets\_Ter(8.5,Inf] -0.150697377  
## hs\_total\_veg\_Ter(6,8.5] 0.332605382  
## hs\_total\_veg\_Ter(8.5,Inf] 0.304966442  
## hs\_total\_yog\_Ter(6,8.5] -0.438249318  
## hs\_total\_yog\_Ter(8.5,Inf] 0.062786917  
## hs\_dif\_hours\_total\_None -0.699382342  
## hs\_as\_m\_Log2 0.040301013  
## hs\_cd\_c\_Log2 0.187427057  
## hs\_co\_m\_Log2 -0.260323347  
## hs\_cs\_m\_Log2 -0.139476373  
## hs\_hg\_m\_Log2 .   
## hs\_mn\_c\_Log2 0.154229376  
## hs\_mn\_m\_Log2 0.439752705  
## hs\_mo\_c\_Log2 -0.026141316  
## hs\_mo\_m\_Log2 .   
## hs\_pb\_c\_Log2 0.137914034  
## hs\_pb\_m\_Log2 -0.093839134  
## hs\_tl\_cdich\_NoneUndetected 0.142873015  
## hs\_tl\_mdich\_NoneUndetected -0.025130370  
## h\_humidity\_preg\_None -0.697538968  
## h\_pressure\_preg\_None .   
## h\_temperature\_preg\_None -0.021192044  
## hs\_hum\_mt\_hs\_h\_None -1.301944303  
## hs\_tm\_mt\_hs\_h\_None .   
## hs\_uvdvf\_mt\_hs\_h\_None 0.953006563  
## hs\_hum\_dy\_hs\_h\_None -0.011715063  
## hs\_hum\_wk\_hs\_h\_None 1.873418052  
## hs\_tm\_dy\_hs\_h\_None 0.296434243  
## hs\_tm\_wk\_hs\_h\_None 0.053400382  
## hs\_uvdvf\_dy\_hs\_h\_None -1.490202376  
## hs\_uvdvf\_wk\_hs\_h\_None .   
## hs\_blueyn300\_s\_None1 .   
## h\_blueyn300\_preg\_None1 -0.335982117  
## h\_greenyn300\_preg\_None1 -0.129385979  
## h\_ndvi100\_preg\_None 1.064692963  
## hs\_greenyn300\_s\_None1 .   
## hs\_blueyn300\_h\_None1 0.441178398  
## hs\_greenyn300\_h\_None1 -0.056283219  
## hs\_ndvi100\_h\_None -1.418817113  
## hs\_ndvi100\_s\_None -0.285358266  
## h\_lden\_cat\_preg\_None -0.039873826  
## hs\_ln\_cat\_h\_None2 -0.247678614  
## hs\_ln\_cat\_h\_None3 -0.155403224  
## hs\_ln\_cat\_h\_None4 .   
## hs\_ln\_cat\_h\_None5 0.020147369  
## hs\_lden\_cat\_s\_None2 -0.833208724  
## hs\_lden\_cat\_s\_None3 -0.437793098  
## hs\_lden\_cat\_s\_None4 -0.522113131  
## hs\_lden\_cat\_s\_None5 0.198997253  
## hs\_lden\_cat\_s\_None6 -0.012233323  
## hs\_dde\_madj\_Log2 -0.352326968  
## hs\_ddt\_cadj\_Log2 .   
## hs\_ddt\_madj\_Log2 -0.199857375  
## hs\_hcb\_cadj\_Log2 0.496817830  
## hs\_pcb118\_cadj\_Log2 0.023867321  
## hs\_pcb138\_madj\_Log2 0.818737643  
## hs\_pcb153\_cadj\_Log2 .   
## hs\_pcb170\_cadj\_Log2 0.402198092  
## hs\_pcb170\_madj\_Log2 -0.070567960  
## hs\_pcb180\_cadj\_Log2 -0.246627774  
## hs\_pcb180\_madj\_Log2 -0.505801107  
## hs\_sumPCBs5\_cadj\_Log2 .   
## hs\_sumPCBs5\_madj\_Log2 .   
## hs\_dep\_cadj\_Log2 -0.380195046  
## hs\_dep\_madj\_Log2 0.574586283  
## hs\_detp\_cadj\_Log2 -0.498326115  
## hs\_dmdtp\_cdich\_NoneUndetected -0.156671444  
## hs\_dmp\_cadj\_Log2 .   
## hs\_dmtp\_cadj\_Log2 0.735899490  
## hs\_dmtp\_madj\_Log2 .   
## hs\_pbde153\_cadj\_Log2 -0.684291927  
## hs\_pbde153\_madj\_Log2 .   
## hs\_pbde47\_cadj\_Log2 0.039053064  
## hs\_pbde47\_madj\_Log2 -0.052182212  
## hs\_pfhxs\_c\_Log2 0.125625227  
## hs\_pfna\_c\_Log2 0.137798991  
## hs\_pfna\_m\_Log2 0.462164453  
## hs\_pfoa\_c\_Log2 -0.303901311  
## hs\_pfoa\_m\_Log2 0.641274169  
## hs\_pfos\_c\_Log2 .   
## hs\_pfos\_m\_Log2 -0.642863283  
## hs\_pfunda\_c\_Log2 0.232572628  
## hs\_pfunda\_m\_Log2 -0.097951348  
## hs\_bpa\_cadj\_Log2 0.378737110  
## hs\_bpa\_madj\_Log2 0.290681481  
## hs\_bupa\_cadj\_Log2 0.003572229  
## hs\_bupa\_madj\_Log2 -0.348928892  
## hs\_etpa\_cadj\_Log2 .   
## hs\_etpa\_madj\_Log2 0.133633274  
## hs\_mepa\_cadj\_Log2 0.675123918  
## hs\_mepa\_madj\_Log2 -0.438766135  
## hs\_oxbe\_cadj\_Log2 0.343820801  
## hs\_oxbe\_madj\_Log2 .   
## hs\_prpa\_cadj\_Log2 -1.343636082  
## hs\_prpa\_madj\_Log2 0.494214780  
## hs\_trcs\_cadj\_Log2 0.169454606  
## hs\_trcs\_madj\_Log2 -0.151652656  
## hs\_mbzp\_cadj\_Log2 -0.011177166  
## hs\_mbzp\_madj\_Log2 0.144817269  
## hs\_mecpp\_cadj\_Log2 -0.295767600  
## hs\_mecpp\_madj\_Log2 0.462272783  
## hs\_mehhp\_cadj\_Log2 .   
## hs\_mehhp\_madj\_Log2 -0.024336072  
## hs\_mehp\_cadj\_Log2 0.015485569  
## hs\_mehp\_madj\_Log2 0.482378405  
## hs\_meohp\_cadj\_Log2 .   
## hs\_meohp\_madj\_Log2 .   
## hs\_mep\_cadj\_Log2 0.538689789  
## hs\_mep\_madj\_Log2 0.316182178  
## hs\_mibp\_cadj\_Log2 -0.012110735  
## hs\_mibp\_madj\_Log2 0.836809373  
## hs\_mnbp\_cadj\_Log2 -0.029180989  
## hs\_mnbp\_madj\_Log2 -0.402429506  
## hs\_ohminp\_cadj\_Log2 -0.223070866  
## hs\_ohminp\_madj\_Log2 0.234971330  
## hs\_oxominp\_cadj\_Log2 -0.108363747  
## hs\_oxominp\_madj\_Log2 -0.329685413  
## hs\_sumDEHP\_cadj\_Log2 -0.159500845  
## hs\_sumDEHP\_madj\_Log2 -0.747685761  
## FAS\_cat\_NoneMiddle -0.287937158  
## FAS\_cat\_NoneHigh .   
## hs\_contactfam\_3cat\_num\_NoneOnce a week -0.523938552  
## hs\_contactfam\_3cat\_num\_NoneLess than once a week -0.171488817  
## hs\_hm\_pers\_None -0.040812258  
## hs\_participation\_3cat\_None1 organisation .   
## hs\_participation\_3cat\_None2 or more organisations -0.373623303  
## e3\_asmokcigd\_p\_None 0.509271337  
## hs\_cotinine\_cdich\_NoneUndetected .   
## hs\_cotinine\_mcat\_NoneSHS smokers 0.418784725  
## hs\_cotinine\_mcat\_NoneSmokers .   
## hs\_globalexp2\_Noneno exposure 0.217020123  
## hs\_smk\_parents\_Noneneither -0.119154098  
## hs\_smk\_parents\_Noneone 0.708506437  
## h\_distinvnear1\_preg\_Log -0.023026102  
## h\_trafload\_preg\_pow1over3 -0.014446542  
## h\_trafnear\_preg\_pow1over3 0.024395289  
## hs\_trafload\_h\_pow1over3 0.211514975  
## hs\_trafnear\_h\_pow1over3 .   
## h\_bro\_preg\_Log 0.942242640  
## h\_clf\_preg\_Log -0.757546920  
## h\_thm\_preg\_Log 0.391667783  
## e3\_bw -0.214549197  
## hs\_zbmi\_who 0.408217136  
## hs\_correct\_raven -1.536903807  
## hs\_Gen\_Tot 0.037235029  
## hs\_bmi\_c\_cat2 .   
## hs\_bmi\_c\_cat3 0.188288672  
## hs\_bmi\_c\_cat4 0.136344598

varImp(LASSO)

## glmnet variable importance  
##   
## only 20 most important variables shown (out of 258)  
##   
## Overall  
## h\_cohort3 100.00  
## hs\_hum\_wk\_hs\_h\_None 85.67  
## hs\_correct\_raven 70.28  
## hs\_uvdvf\_dy\_hs\_h\_None 68.15  
## hs\_ndvi100\_h\_None 64.88  
## hs\_prpa\_cadj\_Log2 61.45  
## hs\_hum\_mt\_hs\_h\_None 59.54  
## h\_ndvi100\_preg\_None 48.69  
## hs\_uvdvf\_mt\_hs\_h\_None 43.58  
## h\_bro\_preg\_Log 43.09  
## h\_cohort6 43.08  
## hs\_bakery\_prod\_Ter(6,Inf] 42.72  
## hs\_pet\_cat\_r2\_None1 39.42  
## hs\_total\_fruits\_Ter(7,14.1] 38.82  
## hs\_beverages\_Ter(0.132,1] 38.76  
## h\_builtdens300\_preg\_Sqrt 38.38  
## hs\_accesslines300\_h\_dic0 38.31  
## hs\_mibp\_madj\_Log2 38.27  
## hs\_lden\_cat\_s\_None2 38.10  
## hs\_pcb138\_madj\_Log2 37.44

# Test model  
test\_outcome = predict(LASSO, test\_data)  
  
# Evaluation metric:  
confusionMatrix(test\_outcome, test\_data$hs\_asthma, positive = '1')

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction 0 1  
## 0 271 26  
## 1 76 16  
##   
## Accuracy : 0.7378   
## 95% CI : (0.6911, 0.7808)  
## No Information Rate : 0.892   
## P-Value [Acc > NIR] : 1   
##   
## Kappa : 0.1063   
##   
## Mcnemar's Test P-Value : 1.224e-06   
##   
## Sensitivity : 0.38095   
## Specificity : 0.78098   
## Pos Pred Value : 0.17391   
## Neg Pred Value : 0.91246   
## Prevalence : 0.10797   
## Detection Rate : 0.04113   
## Detection Prevalence : 0.23650   
## Balanced Accuracy : 0.58097   
##   
## 'Positive' Class : 1   
##

Based on the variable importance, we know which variables in the dataset are the most important to predict the outcome (asthma). From there, we can reduce the number of features in our dataset by selecting only the variables as indicated by the variable importance. Then, we can do further analysis using different algorithms such as random forest or logistic regression to predict our outcome of interest.