Homework 9

March 27 2022

Demonstrate Interaction using Regression Models and Tree-based Methods using Exposome Data from HELIX

Load .Rdata file and merge into single data frame

Reminder: Merging into a single data frame is optional. Depends upon how you program. This example will assume you've merged everything into a single data frame.

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.8
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1
## Warning: package 'tidyr' was built under R version 4.1.2
## Warning: package 'readr' was built under R version 4.1.2
## Warning: package 'dplyr' was built under R version 4.1.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(caret)
```

```
## Loading required package: lattice
##
## Attaching package: 'caret'
## The following object is masked from 'package:purrr':
##
##     lift
```

```
library(rpart.plot)
```

```
## Loading required package: rpart
```

```
library(randomForest)
```

```
## Warning: package 'randomForest' was built under R version 4.1.2
## randomForest 4.7-1
## Type rfNews() to see new features/changes/bug fixes.
##
## Attaching package: 'randomForest'
```

```
## The following object is masked from 'package:dplyr':
##
##     combine
## The following object is masked from 'package:ggplot2':
##
##     margin
```

```
library(caret)
library(gbm)
```

```
## Loaded gbm 2.1.8
```

```
library(pROC)
```

```
## Type 'citation("pROC")' for a citation.
##
## Attaching package: 'pROC'
## The following objects are masked from 'package:stats':
##
##     cov, smooth, var
```

```
library(rpart)
library(Amelia)
```

```
## Loading required package: Rcpp
## ##
## ## Amelia II: Multiple Imputation
## ## (Version 1.8.0, built: 2021-05-26)
## ## Copyright (C) 2005-2022 James Honaker, Gary King and Matthew Blackwell
## ## Refer to http://gking.harvard.edu/amelia/ for more information
## ##
```

```
#Load data using path of where file is stored
load("./exposome.RData")

#Merge all data frames into a single data frame. FYI, this is just a shortcut by combining baseR with p

studydata<-merge(exposome,phenotype,by="ID") %>% merge(covariates, by="ID")

#Strip off ID Variable
studydata$ID<-NULL

#Partition data for use in demonstration
set.seed(100)
train.indices<-createDataPartition(y=studydata$e3_bw,p=0.7,list=FALSE)
train.data<-studydata[train.indices, ]
test.data<-studydata[-train.indices, ]
```

Step 1: Data Exploration of Training Data

```
#view(train.data)
str(train.data)
```

```
## 'data.frame':    912 obs. of  241 variables:
##  $ h_abs_ratio_preg_Log        : num  0.8967 0.8925 0.7787 0.0891 0.6048 ...
```

```
##  $ h_no2_ratio_preg_Log        : num  2.87 2.98 3.06 3.09 3.85 ...
##  $ h_pm10_ratio_preg_None      : num  25.9 25.9 26.1 15 35.2 ...
##  $ h_pm25_ratio_preg_None      : num  17.4 18.5 18.7 16.4 14.9 ...
##  $ hs_no2_dy_hs_h_Log          : num  2.53 1.93 2.88 1.39 3.2 ...
##  $ hs_no2_wk_hs_h_Log          : num  2.58 2.65 2.59 2.46 3.5 ...
##  $ hs_no2_yr_hs_h_Log          : num  2.61 2.76 2.36 2.4 3.31 ...
##  $ hs_pm10_dy_hs_h_None        : num  22.5 14.1 46.9 29.8 29.8 ...
##  $ hs_pm10_wk_hs_h_None        : num  20.9 29.1 31.5 25.2 24.9 ...
##  $ hs_pm10_yr_hs_h_None        : num  31.4 31.3 27.5 24 24.8 ...
##  $ hs_pm25_dy_hs_h_None        : num  16.95 11.16 28.45 4.62 14.92 ...
##  $ hs_pm25_wk_hs_h_None        : num  17 15.9 21.3 11 13.9 ...
##  $ hs_pm25_yr_hs_h_None        : num  18.4 17.7 16.8 12.7 13.4 ...
##  $ hs_pm25abs_dy_hs_h_Log      : num  0.0974 -0.4304 0.9156 -0.2833 0.9156 ...
##  $ hs_pm25abs_wk_hs_h_Log      : num  0.0712 0.2143 0.7197 -0.1387 -0.138 ...
##  $ hs_pm25abs_yr_hs_h_Log      : num  0.3211 0.2815 0.0987 0.1777 0.2205 ...
##  $ h_accesslines300_preg_dic0  : num  0 0 0 1 1 0 0 0 0 0 ...
##  $ h_accesspoints300_preg_Log  : num  1.96 2.37 1.27 4.53 3.06 ...
##  $ h_builtdens300_preg_Sqrt    : num  405 311 375 565 585 ...
##  $ h_connind300_preg_Sqrt      : num  1.89 6.54 6.26 14.49 18.68 ...
##  $ h_fdensity300_preg_Log      : num  10.3 10.3 10.3 13.8 12.2 ...
##  $ h_frichness300_preg_None    : num  0 0 0 0.2456 0.0877 ...
##  $ h_landuseshan300_preg_None  : num  0.364 0.401 0.288 0.633 0.459 ...
##  $ h_popdens_preg_Sqrt         : num  85 85 85 66.5 96.2 ...
##  $ h_walkability_mean_preg_None: num  0.175 0.2 0.15 0.35 0.275 0.35 0.2 0.225 0.175 0.3 ...
##  $ hs_accesslines300_h_dic0    : num  0 0 0 1 1 0 0 0 0 0 ...
##  $ hs_accesspoints300_h_Log    : num  1.675 2.774 0.577 4.584 3.621 ...
##  $ hs_builtdens300_h_Sqrt      : num  407 383 375 480 210 ...
##  $ hs_connind300_h_Log         : num  4.57 3.75 2.88 5.47 4.82 ...
##  $ hs_fdensity300_h_Log        : num  10.3 10.3 10.3 14 11 ...
##  $ hs_landuseshan300_h_None    : num  0.354 0.321 0.479 0.454 0.51 ...
##  $ hs_popdens_h_Sqrt           : num  84.99 10.25 10.25 66.54 4.56 ...
##  $ hs_walkability_mean_h_None  : num  0.375 0.2 0.25 0.525 0.3 0.375 0.3 0.325 0.275 0.35 ...
##  $ hs_accesslines300_s_dic0    : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ hs_accesspoints300_s_Log    : num  0.577 2.186 2.186 3.285 2.186 ...
##  $ hs_builtdens300_s_Sqrt      : num  385 383.6 366.4 406.3 61.9 ...
##  $ hs_connind300_s_Log         : num  2.37 3.47 3.98 5.65 4.74 ...
##  $ hs_fdensity300_s_Log        : num  10.3 10.3 10.3 12.3 11 ...
##  $ hs_landuseshan300_s_None    : num  0.28 0.368 0.325 0.521 0.521 ...
##  $ hs_popdens_s_Sqrt           : num  84.99 84.99 84.99 25.71 4.56 ...
##  $ h_Absorbance_Log            : num  -0.1351 -0.0577 -0.4372 -0.8705 0.564 ...
##  $ h_Benzene_Log               : num  0.572 0.88 1.379 1.168 0.133 ...
##  $ h_NO2_Log                   : num  4.58 3.37 4.3 2.5 6.92 ...
##  $ h_PM_Log                    : num  2.6 2.33 2.33 1.66 3.62 ...
##  $ h_TEX_Log                   : num  2.53 2.84 2.88 2.67 3.43 ...
##  $ e3_alcpreg_yn_None          : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 2 1 1 ...
##  $ h_bfdur_Ter                 : Factor w/ 3 levels "(0,10.8]","(10.8,34.9]",..: 1 1 3 2 2 3 1 3 1 3
##  $ h_cereal_preg_Ter           : Factor w/ 3 levels "(0,9]","(9,27.3]",..: 1 1 2 3 1 3 2 1 3 2 ...
##  $ h_dairy_preg_Ter            : Factor w/ 3 levels "(0,17.1]","(17.1,27.1]",..: 3 3 3 1 1 3 3 2 1 3
##  $ h_fastfood_preg_Ter         : Factor w/ 3 levels "(0,0.25]","(0.25,0.83]",..: 2 3 3 3 3 2 3 3 3 3
##  $ h_fish_preg_Ter             : Factor w/ 3 levels "(0,1.9]","(1.9,4.1]",..: 3 3 3 1 3 3 1 2 2 2 ..
##  $ h_folic_t1_None             : Factor w/ 2 levels "0","1": 1 1 2 1 1 1 1 1 1 2 ...
##  $ h_fruit_preg_Ter            : Factor w/ 3 levels "(0,0.6]","(0.6,18.2]",..: 2 2 2 2 2 2 2 3 2 2 .
##  $ h_legume_preg_Ter           : Factor w/ 3 levels "(0,0.5]","(0.5,2]",..: 3 3 3 3 2 3 2 3 2 3 ...
##  $ h_meat_preg_Ter             : Factor w/ 3 levels "(0,6.5]","(6.5,10]",..: 2 3 2 3 1 3 1 3 1 2 ...
```

```
##  $ h_pamod_t3_None            : Factor w/ 4 levels "None","Often",..: 4 1 2 2 4 4 3 2 2 4 ...
##  $ h_pavig_t3_None            : Factor w/ 3 levels "High","Low","Medium": 2 2 2 3 3 3 2 2 2 2 ...
##  $ h_veg_preg_Ter             : Factor w/ 3 levels "(0,8.8]","(8.8,16.5]",..: 2 2 2 1 2 2 1 1 2 2 .
##  $ hs_bakery_prod_Ter         : Factor w/ 3 levels "(0,2]","(2,6]",..: 2 1 2 3 2 2 3 3 3 2 ...
##  $ hs_beverages_Ter           : Factor w/ 3 levels "(0,0.132]","(0.132,1]",..: 3 2 2 1 2 1 3 1 1 3
##  $ hs_break_cer_Ter           : Factor w/ 3 levels "(0,1.1]","(1.1,5.5]",..: 1 1 2 3 2 3 2 2 3 3 ..
##  $ hs_caff_drink_Ter          : Factor w/ 2 levels "(0,0.132]","(0.132,Inf]": 1 2 2 1 1 1 2 1 1 2 .
##  $ hs_dairy_Ter               : Factor w/ 3 levels "(0,14.6]","(14.6,25.6]",..: 1 1 1 3 1 3 3 3 2 1
##  $ hs_fastfood_Ter            : Factor w/ 3 levels "(0,0.132]","(0.132,0.5]",..: 2 2 2 2 2 1 2 3 3 3
##  $ hs_KIDMED_None             : num  2 0 1 2 4 5 2 3 3 3 ...
##  $ hs_mvpa_prd_alt_None       : num  47.89 31.83 117.58 -2.03 -7.85 ...
##  $ hs_org_food_Ter            : Factor w/ 3 levels "(0,0.132]","(0.132,1]",..: 2 3 2 2 1 3 2 1 3 2
##  $ hs_pet_cat_r2_None         : Factor w/ 2 levels "0","1": 1 1 1 2 1 1 2 1 1 1 ...
##  $ hs_pet_dog_r2_None         : Factor w/ 2 levels "0","1": 1 1 1 1 2 1 1 1 1 1 ...
##  $ hs_pet_None                : Factor w/ 2 levels "No","Yes": 2 1 1 2 2 2 2 1 2 1 ...
##  $ hs_proc_meat_Ter           : Factor w/ 3 levels "(0,1.5]","(1.5,4]",..: 2 2 3 1 3 3 2 2 1 1 ...
##  $ hs_readymade_Ter           : Factor w/ 3 levels "(0,0.132]","(0.132,0.5]",..: 1 3 3 3 1 1 2 3 1 1
##  $ hs_sd_wk_None              : num  389 523 446 180 454 ...
##  $ hs_total_bread_Ter         : Factor w/ 3 levels "(0,7]","(7,17.5]",..: 1 3 1 3 2 3 1 2 3 2 ...
##  $ hs_total_cereal_Ter        : Factor w/ 3 levels "(0,14.1]","(14.1,23.6]",..: 1 2 2 3 2 3 1 1 3 2
##  $ hs_total_fish_Ter          : Factor w/ 3 levels "(0,1.5]","(1.5,3]",..: 1 1 1 2 3 3 1 2 2 1 ...
##  $ hs_total_fruits_Ter        : Factor w/ 3 levels "(0,7]","(7,14.1]",..: 3 1 1 3 3 3 1 2 3 3 ...
##  $ hs_total_lipids_Ter        : Factor w/ 3 levels "(0,3]","(3,7]",..: 3 3 3 2 3 2 1 3 3 3 ...
##  $ hs_total_meat_Ter          : Factor w/ 3 levels "(0,6]","(6,9]",..: 2 1 3 1 2 3 1 1 1 1 ...
##  $ hs_total_potatoes_Ter      : Factor w/ 3 levels "(0,3]","(3,4]",..: 3 2 2 1 1 3 1 3 1 3 ...
##  $ hs_total_sweets_Ter        : Factor w/ 3 levels "(0,4.1]","(4.1,8.5]",..: 1 3 3 3 1 3 1 2 3 3 ..
##  $ hs_total_veg_Ter           : Factor w/ 3 levels "(0,6]","(6,8.5]",..: 3 2 3 3 1 3 1 3 1 3 ...
##  $ hs_total_yog_Ter           : Factor w/ 3 levels "(0,6]","(6,8.5]",..: 1 1 1 3 1 3 3 2 1 1 ...
##  $ hs_dif_hours_total_None    : num  9.46 9.6 9.97 10.09 10.52 ...
##  $ hs_as_c_Log2               : num  -4.283 -6.43 -7.978 0.632 0.651 ...
##  $ hs_as_m_Log2               : num  -19.5 -9.58 -14.92 1.84 1.29 ...
##  $ hs_cd_c_Log2               : num  -4.14 -4.25 -4.05 -5.26 -3.75 ...
##  $ hs_cd_m_Log2               : num  -3.071 -2.599 -2.966 -0.935 -0.935 ...
##  $ hs_co_c_Log2               : num  1.02 -2.62 -2.56 -3 -1.61 ...
##  $ hs_co_m_Log2               : num  -4.08 -2.41 -2.63 -2.81 -0.55 ...
##  $ hs_cs_c_Log2               : num  0.251 0.202 0.39 0.669 0.903 ...
##  $ hs_cs_m_Log2               : num  0.151 -0.272 0.536 0.714 0.138 ...
##  $ hs_cu_c_Log2               : num  9.71 9.94 9.93 10.19 9.44 ...
##  $ hs_cu_m_Log2               : num  10.22 10.51 10.26 10.09 9.89 ...
##  $ hs_hg_c_Log2               : num  -2.152 -1.3 -0.911 1.48 2.862 ...
##  $ hs_hg_m_Log2               : num  -3.1203 -1.0233 0.0841 2.211 3.2345 ...
##  $ hs_mn_c_Log2               : num  3.46 2.88 3 3.22 2.88 ...
##  $ hs_mn_m_Log2               : num  3.61 4.16 2.88 3.25 3.49 ...
##  $ hs_mo_c_Log2               : num  0.949 1.07 -0.484 -5.866 -0.252 ...
##   [list output truncated]
```

```
#seeing lots of numeric variables with some factor variables
summary(train.data)
```

```
##  h_abs_ratio_preg_Log h_no2_ratio_preg_Log h_pm10_ratio_preg_None
##  Min.   :-0.47756     Min.   :2.105        Min.   : 8.066
##  1st Qu.: 0.09776     1st Qu.:2.671        1st Qu.:17.535
##  Median : 0.30201     Median :2.959        Median :22.796
##  Mean   : 0.39163     Mean   :3.001        Mean   :23.393
##  3rd Qu.: 0.72366     3rd Qu.:3.280        3rd Qu.:27.337
```

```
##  Max.   : 1.70346     Max.   :4.525        Max.   :47.698
##
##  h_pm25_ratio_preg_None hs_no2_dy_hs_h_Log hs_no2_wk_hs_h_Log
##  Min.   : 6.957         Min.   :0.3797     Min.   :1.100
##  1st Qu.:13.237         1st Qu.:2.3209     1st Qu.:2.369
##  Median :14.880         Median :2.9840     Median :3.009
##  Mean   :15.002         Mean   :2.8458     Mean   :2.881
##  3rd Qu.:17.014         3rd Qu.:3.4407     3rd Qu.:3.414
##  Max.   :22.238         Max.   :5.1849     Max.   :4.805
##
##  hs_no2_yr_hs_h_Log hs_pm10_dy_hs_h_None hs_pm10_wk_hs_h_None
##  Min.   :0.8597     Min.   :  2.916      Min.   :  5.838
##  1st Qu.:2.3931     1st Qu.: 17.677      1st Qu.: 19.729
##  Median :3.0517     Median : 22.996      Median : 24.891
##  Mean   :2.9097     Mean   : 26.043      Mean   : 26.290
##  3rd Qu.:3.4093     3rd Qu.: 30.937      3rd Qu.: 32.285
##  Max.   :4.4225     Max.   :157.397      Max.   :184.680
##
##  hs_pm10_yr_hs_h_None hs_pm25_dy_hs_h_None hs_pm25_wk_hs_h_None
##  Min.   :11.50        Min.   : 1.518       Min.   : 3.188
##  1st Qu.:21.52        1st Qu.: 8.065       1st Qu.: 9.391
##  Median :24.75        Median :12.363       Median :12.805
##  Mean   :25.09        Mean   :12.919       Mean   :13.162
##  3rd Qu.:31.46        3rd Qu.:16.319       3rd Qu.:16.147
##  Max.   :46.82        Max.   :58.884       Max.   :67.318
##
##  hs_pm25_yr_hs_h_None hs_pm25abs_dy_hs_h_Log hs_pm25abs_wk_hs_h_Log
##  Min.   : 4.968       Min.   :-1.78220       Min.   :-0.90191
##  1st Qu.:10.421       1st Qu.:-0.26618       1st Qu.:-0.13869
##  Median :13.266       Median : 0.05217       Median : 0.04672
##  Mean   :13.012       Mean   : 0.12830       Mean   : 0.16952
##  3rd Qu.:15.241       3rd Qu.: 0.60535       3rd Qu.: 0.54153
##  Max.   :21.813       Max.   : 2.26537       Max.   : 1.87776
##
##  hs_pm25abs_yr_hs_h_Log h_accesslines300_preg_dic0 h_accesspoints300_preg_Log
##  Min.   :-0.59670       Min.   :0.0000             Min.   :1.270
##  1st Qu.:-0.01809       1st Qu.:0.0000             1st Qu.:2.130
##  Median : 0.17773       Median :0.0000             Median :2.879
##  Mean   : 0.18185       Mean   :0.1952             Mean   :2.669
##  3rd Qu.: 0.31665       3rd Qu.:0.0000             3rd Qu.:3.349
##  Max.   : 1.36495       Max.   :1.0000             Max.   :4.528
##
##  h_builtdens300_preg_Sqrt h_connind300_preg_Sqrt h_fdensity300_preg_Log
##  Min.   : 19.98           Min.   : 1.887         Min.   :10.26
##  1st Qu.:340.04           1st Qu.: 9.983         1st Qu.:10.26
##  Median :401.39           Median :12.935         Median :11.36
##  Mean   :417.31           Mean   :12.786         Mean   :11.61
##  3rd Qu.:501.64           3rd Qu.:16.009         3rd Qu.:12.83
##  Max.   :807.57           Max.   :27.276         Max.   :15.60
##
##  h_frichness300_preg_None h_landuseshan300_preg_None h_popdens_preg_Sqrt
##  Min.   :0.00000          Min.   :0.0000             Min.   :  0.00
##  1st Qu.:0.00000          1st Qu.:0.3400             1st Qu.: 52.28
##  Median :0.03509          Median :0.4234             Median : 74.98
```

```
##  Mean   :0.06517          Mean   :0.4218             Mean   : 76.38
##  3rd Qu.:0.12281          3rd Qu.:0.5092             3rd Qu.: 95.77
##  Max.   :0.42105          Max.   :1.0000             Max.   :261.50
##
##  h_walkability_mean_preg_None hs_accesslines300_h_dic0 hs_accesspoints300_h_Log
##  Min.   :0.1000               Min.   :0.0000           Min.   :0.5771
##  1st Qu.:0.2000               1st Qu.:0.0000           1st Qu.:1.6753
##  Median :0.2500               Median :0.0000           Median :2.7738
##  Mean   :0.2673               Mean   :0.1743           Mean   :2.4104
##  3rd Qu.:0.3250               3rd Qu.:0.0000           3rd Qu.:3.2846
##  Max.   :0.6250               Max.   :1.0000           Max.   :4.5838
##
##  hs_builtdens300_h_Sqrt hs_connind300_h_Log hs_fdensity300_h_Log
##  Min.   : 20.3          Min.   :1.270       Min.   :10.26
##  1st Qu.:304.4          1st Qu.:4.361       1st Qu.:10.26
##  Median :375.1          Median :4.959       Median :10.96
##  Mean   :382.6          Mean   :4.767       Mean   :11.39
##  3rd Qu.:461.8          3rd Qu.:5.364       3rd Qu.:12.34
##  Max.   :805.8          Max.   :6.612       Max.   :14.74
##
##  hs_landuseshan300_h_None hs_popdens_h_Sqrt hs_walkability_mean_h_None
##  Min.   :0.0000           Min.   :  1.732   Min.   :0.1000
##  1st Qu.:0.3138           1st Qu.: 32.163   1st Qu.:0.2750
##  Median :0.4015           Median : 67.356   Median :0.3000
##  Mean   :0.3985           Mean   : 67.730   Mean   :0.3267
##  3rd Qu.:0.5000           3rd Qu.: 84.988   3rd Qu.:0.3750
##  Max.   :0.6619           Max.   :261.500   Max.   :0.6000
##
##  hs_accesslines300_s_dic0 hs_accesspoints300_s_Log hs_builtdens300_s_Sqrt
##  Min.   :0.0000           Min.   :0.5771           Min.   :  6.432
##  1st Qu.:0.0000           1st Qu.:1.6753           1st Qu.:313.276
##  Median :0.0000           Median :2.5225           Median :380.459
##  Mean   :0.1776           Mean   :2.3880           Mean   :399.919
##  3rd Qu.:0.0000           3rd Qu.:3.2846           3rd Qu.:479.125
##  Max.   :1.0000           Max.   :4.0730           Max.   :805.140
##
##  hs_connind300_s_Log hs_fdensity300_s_Log hs_landuseshan300_s_None
##  Min.   :1.270       Min.   :10.26        Min.   :0.08298
##  1st Qu.:4.489       1st Qu.:10.26        1st Qu.:0.34163
##  Median :4.894       Median :11.36        Median :0.44759
##  Mean   :4.773       Mean   :11.55        Mean   :0.43068
##  3rd Qu.:5.381       3rd Qu.:12.57        3rd Qu.:0.53690
##  Max.   :6.451       Max.   :15.25        Max.   :0.72770
##
##  hs_popdens_s_Sqrt h_Absorbance_Log    h_Benzene_Log       h_NO2_Log
##  Min.   :  1.732   Min.   :-0.927367   Min.   :-0.3296   Min.   :1.573
##  1st Qu.: 38.559   1st Qu.:-0.539150   1st Qu.: 0.3078   1st Qu.:2.960
##  Median : 69.264   Median :-0.279289   Median : 0.5594   Median :3.609
##  Mean   : 67.841   Mean   :-0.175179   Mean   : 0.5967   Mean   :3.816
##  3rd Qu.: 84.988   3rd Qu.: 0.006912   3rd Qu.: 0.8490   3rd Qu.:4.576
##  Max.   :210.954   Max.   : 3.404743   Max.   : 1.9502   Max.   :7.093
##
##     h_PM_Log       h_TEX_Log     e3_alcpreg_yn_None      h_bfdur_Ter
##  Min.   :1.549   Min.   :1.943   0:619              (0,10.8]   :353
```

```
##  1st Qu.:2.068   1st Qu.:2.610   1:293             (10.8,34.9]:177
##  Median :2.293   Median :2.980                     (34.9,Inf] :382
##  Mean   :2.426   Mean   :3.012
##  3rd Qu.:2.690   3rd Qu.:3.376
##  Max.   :5.236   Max.   :4.944
##
##   h_cereal_preg_Ter    h_dairy_preg_Ter  h_fastfood_preg_Ter  h_fish_preg_Ter
##  (0,9]     :371     (0,17.1]   :192     (0,0.25]   : 68      (0,1.9]  :241
##  (9,27.3]  :319     (17.1,27.1]:263     (0.25,0.83]:373      (1.9,4.1]:340
##  (27.3,Inf]:222     (27.1,Inf] :457     (0.83,Inf] :471      (4.1,Inf]:331
##
##
##
##
##  h_folic_t1_None   h_fruit_preg_Ter h_legume_preg_Ter h_meat_preg_Ter
##  0:432           (0,0.6]   :  5     (0,0.5]:172       (0,6.5] :296
##  1:480           (0.6,18.2]:643     (0.5,2]:181       (6.5,10]:273
##                  (18.2,Inf]:264     (2,Inf]:559       (10,Inf]:343
##
##
##
##
##    h_pamod_t3_None h_pavig_t3_None    h_veg_preg_Ter hs_bakery_prod_Ter
##  None      : 26    High  : 30      (0,8.8]   :382    (0,2]  :244
##  Often     :332    Low   :654      (8.8,16.5]:331    (2,6]  :297
##  Sometimes :142    Medium:228      (16.5,Inf]:199    (6,Inf]:371
##  Very Often:412
##
##
##
##   hs_beverages_Ter  hs_break_cer_Ter   hs_caff_drink_Ter      hs_dairy_Ter
##  (0,0.132]:238     (0,1.1]  :192     (0,0.132]  :567     (0,14.6]   :245
##  (0.132,1]:318     (1.1,5.5]:361     (0.132,Inf]:345     (14.6,25.6]:316
##  (1,Inf]  :356     (5.5,Inf]:359                         (25.6,Inf] :351
##
##
##
##
##     hs_fastfood_Ter hs_KIDMED_None   hs_mvpa_prd_alt_None  hs_org_food_Ter
##  (0,0.132]  : 97    Min.   :-3.000   Min.   :-27.76       (0,0.132]:297
##  (0.132,0.5]:430    1st Qu.: 2.000   1st Qu.: 23.58       (0.132,1]:283
##  (0.5,Inf]  :385    Median : 3.000   Median : 35.69       (1,Inf]  :332
##                     Mean   : 2.889   Mean   : 38.43
##                     3rd Qu.: 4.000   3rd Qu.: 49.14
##                     Max.   : 9.000   Max.   :134.89
##
##  hs_pet_cat_r2_None hs_pet_dog_r2_None hs_pet_None hs_proc_meat_Ter
##  0:741              0:783              No :573     (0,1.5]:272
##  1:171              1:129              Yes:339     (1.5,4]:321
##                                                    (4,Inf]:319
##
##
##
##
```

```
##     hs_readymade_Ter hs_sd_wk_None      hs_total_bread_Ter  hs_total_cereal_Ter
##  (0,0.132]  :238     Min.   :  3.143   (0,7]     :308      (0,14.1]   :300
##  (0.132,0.5]:198     1st Qu.:158.571   (7,17.5]  :266      (14.1,23.6]:305
##  (0.5,Inf]  :476     Median :210.000   (17.5,Inf]:338      (23.6,Inf] :307
##                      Mean   :235.839
##                      3rd Qu.:280.714
##                      Max.   :887.143
##
##  hs_total_fish_Ter hs_total_fruits_Ter hs_total_lipids_Ter hs_total_meat_Ter
##  (0,1.5]:278       (0,7]     :283      (0,3]  :266         (0,6]  :303
##  (1.5,3]:308       (7,14.1]  :289      (3,7]  :294         (6,9]  :299
##  (3,Inf]:326       (14.1,Inf]:340      (7,Inf]:352         (9,Inf]:310
##
##
##
##
##  hs_total_potatoes_Ter hs_total_sweets_Ter  hs_total_veg_Ter  hs_total_yog_Ter
##  (0,3]  :288           (0,4.1]  :238       (0,6]    :276     (0,6]    :539
##  (3,4]  :286           (4.1,8.5]:361       (6,8.5]  :223     (6,8.5]  :219
##  (4,Inf]:338           (8.5,Inf]:313       (8.5,Inf]:413     (8.5,Inf]:154
##
##
##
##
##  hs_dif_hours_total_None  hs_as_c_Log2       hs_as_m_Log2
##  Min.   : 8.057          Min.   :-15.0124   Min.   :-38.625
##  1st Qu.: 9.758          1st Qu.: -4.0106   1st Qu.: -5.622
##  Median :10.330          Median :  0.4854   Median : -2.615
##  Mean   :10.286          Mean   : -1.0571   Mean   : -3.065
##  3rd Qu.:10.735          3rd Qu.:  1.2403   3rd Qu.:  1.007
##  Max.   :12.852          Max.   :  4.2705   Max.   :  6.493
##
##   hs_cd_c_Log2       hs_cd_m_Log2     hs_co_c_Log2     hs_co_m_Log2
##  Min.   :-10.3945   Min.   :-7.844   Min.   :-5.546   Min.   :-5.184
##  1st Qu.: -4.3990   1st Qu.:-2.711   1st Qu.:-2.708   1st Qu.:-2.506
##  Median : -3.8039   Median :-2.427   Median :-2.423   Median :-2.029
##  Mean   : -3.9424   Mean   :-2.184   Mean   :-2.331   Mean   :-1.713
##  3rd Qu.: -3.3685   3rd Qu.:-1.720   3rd Qu.:-2.047   3rd Qu.:-0.550
##  Max.   :  0.6781   Max.   : 4.802   Max.   : 1.401   Max.   : 2.503
##
##   hs_cs_c_Log2       hs_cs_m_Log2       hs_cu_c_Log2     hs_cu_m_Log2
##  Min.   :-1.45403   Min.   :-1.15843   Min.   : 9.079   Min.   : 9.036
##  1st Qu.: 0.05658   1st Qu.: 0.02857   1st Qu.: 9.676   1st Qu.:10.253
##  Median : 0.46988   Median : 0.36737   Median : 9.830   Median :10.441
##  Mean   : 0.44801   Mean   : 0.47482   Mean   : 9.826   Mean   :10.404
##  3rd Qu.: 0.79285   3rd Qu.: 0.80115   3rd Qu.: 9.966   3rd Qu.:10.534
##  Max.   : 3.06523   Max.   : 2.92220   Max.   :10.482   Max.   :11.116
##
##   hs_hg_c_Log2       hs_hg_m_Log2      hs_mn_c_Log2    hs_mn_m_Log2
##  Min.   :-10.8954   Min.   :-6.3359   Min.   :1.705   Min.   :2.173
##  1st Qu.: -1.2498   1st Qu.:-0.3094   1st Qu.:2.832   1st Qu.:3.295
##  Median : -0.1665   Median : 0.5753   Median :3.123   Median :3.561
##  Mean   : -0.3136   Mean   : 0.5534   Mean   :3.126   Mean   :3.546
##  3rd Qu.:  0.8094   3rd Qu.: 1.5729   3rd Qu.:3.392   3rd Qu.:3.828
```

```
##  Max.   :  3.6554   Max.   : 5.4429   Max.   :4.285   Max.   :5.113
##
##   hs_mo_c_Log2       hs_mo_m_Log2      hs_pb_c_Log2    hs_pb_m_Log2
##  Min.   :-9.23481   Min.   :-2.7179   Min.   :1.084   Min.   :1.384
##  1st Qu.:-0.75025   1st Qu.:-0.9602   1st Qu.:2.709   1st Qu.:2.618
##  Median :-0.39688   Median :-0.7322   Median :3.121   Median :3.191
##  Mean   :-0.29754   Mean   :-0.6864   Mean   :3.123   Mean   :3.206
##  3rd Qu.: 0.04264   3rd Qu.:-0.3622   3rd Qu.:3.485   3rd Qu.:3.807
##  Max.   : 5.12101   Max.   : 6.1334   Max.   :6.087   Max.   :7.547
##
##    hs_tl_cdich_None   hs_tl_mdich_None h_humidity_preg_None
##  Detected  : 72     Detected  : 12     Min.   :56.31
##  Undetected:840     Undetected:900     1st Qu.:70.72
##                                        Median :77.15
##                                        Mean   :76.70
##                                        3rd Qu.:86.62
##                                        Max.   :90.67
##
##  h_pressure_preg_None h_temperature_preg_None hs_hum_mt_hs_h_None
##  Min.   : 974.9       Min.   : 3.120          Min.   :52.05
##  1st Qu.: 980.9       1st Qu.: 8.151          1st Qu.:64.94
##  Median : 983.4       Median :10.158          Median :73.26
##  Mean   : 991.5       Mean   :11.142          Mean   :74.12
##  3rd Qu.:1002.3       3rd Qu.:13.580          3rd Qu.:82.55
##  Max.   :1015.5       Max.   :22.566          Max.   :96.14
##
##  hs_tm_mt_hs_h_None hs_uvdvf_mt_hs_h_None hs_hum_dy_hs_h_None
##  Min.   :-3.477     Min.   :0.0070        Min.   : 26.19
##  1st Qu.: 6.760     1st Qu.:0.2563        1st Qu.: 58.48
##  Median :12.429     Median :1.0073        Median : 71.86
##  Mean   :11.500     Mean   :1.3939        Mean   : 72.58
##  3rd Qu.:16.092     3rd Qu.:2.3274        3rd Qu.: 85.00
##  Max.   :27.271     Max.   :5.1500        Max.   :100.00
##
##  hs_hum_wk_hs_h_None hs_tm_dy_hs_h_None hs_tm_wk_hs_h_None
##  Min.   :48.59       Min.   :-7.900     Min.   :-5.605
##  1st Qu.:63.38       1st Qu.: 6.007     1st Qu.: 6.413
##  Median :73.62       Median :12.037     Median :12.284
##  Mean   :74.14       Mean   :11.291     Mean   :11.308
##  3rd Qu.:84.78       3rd Qu.:16.000     3rd Qu.:16.167
##  Max.   :98.62       Max.   :30.700     Max.   :27.688
##
##  hs_uvdvf_dy_hs_h_None hs_uvdvf_wk_hs_h_None hs_blueyn300_s_None
##  Min.   :0.0000        Min.   :0.001429      0:848
##  1st Qu.:0.2175        1st Qu.:0.232857      1: 64
##  Median :1.0300        Median :1.101429
##  Mean   :1.4280        Mean   :1.424859
##  3rd Qu.:2.3325        3rd Qu.:2.317143
##  Max.   :5.5500        Max.   :5.254286
##
##  h_blueyn300_preg_None h_greenyn300_preg_None h_ndvi100_preg_None
##  0:841                 0:229                  Min.   :0.1074
##  1: 71                 1:683                  1st Qu.:0.2551
##                                               Median :0.4152
```

```
##                                                 Mean   :0.3958
##                                                 3rd Qu.:0.5200
##                                                 Max.   :0.7354
##
##  hs_greenyn300_s_None hs_blueyn300_h_None hs_greenyn300_h_None
##  0:197                0:841               0:193
##  1:715                1: 71               1:719
##
##
##
##
##
##  hs_ndvi100_h_None hs_ndvi100_s_None h_lden_cat_preg_None hs_ln_cat_h_None
##  Min.   :0.09675   Min.   :0.09519   Min.   :33.92        1:341
##  1st Qu.:0.31865   1st Qu.:0.32662   1st Qu.:50.00        2:436
##  Median :0.47920   Median :0.45319   Median :58.51        3: 72
##  Mean   :0.45004   Mean   :0.41855   Mean   :57.20        4: 40
##  3rd Qu.:0.57458   3rd Qu.:0.52496   3rd Qu.:64.36        5: 23
##  Max.   :0.81432   Max.   :0.75681   Max.   :77.40
##
##  hs_lden_cat_s_None hs_dde_cadj_Log2 hs_dde_madj_Log2  hs_ddt_cadj_Log2
##  1:415              Min.   : 1.192   Min.   : 0.8634   Min.   :-15.4250
##  2:190              1st Qu.: 3.543   1st Qu.: 4.3801   1st Qu.: -1.7478
##  3:198              Median : 4.421   Median : 5.5466   Median : -0.4695
##  4: 70              Mean   : 4.656   Mean   : 5.8190   Mean   : -1.5555
##  5: 24              3rd Qu.: 5.500   3rd Qu.: 6.9425   3rd Qu.:  0.7476
##  6: 15              Max.   :11.075   Max.   :10.8937   Max.   :  7.6305
##
##  hs_ddt_madj_Log2   hs_hcb_cadj_Log2 hs_hcb_madj_Log2 hs_pcb118_cadj_Log2
##  Min.   :-14.1418   Min.   :1.440    Min.   :-9.420   Min.   :-6.9507
##  1st Qu.: -0.3734   1st Qu.:2.651    1st Qu.: 2.268   1st Qu.: 0.6026
##  Median :  0.6059   Median :3.070    Median : 2.785   Median : 0.9980
##  Mean   :  0.7535   Mean   :3.180    Mean   : 2.912   Mean   : 1.1057
##  3rd Qu.:  1.3544   3rd Qu.:3.588    3rd Qu.: 3.478   3rd Qu.: 1.5504
##  Max.   :  6.5566   Max.   :6.461    Max.   : 7.357   Max.   : 4.7829
##
##  hs_pcb118_madj_Log2 hs_pcb138_cadj_Log2 hs_pcb138_madj_Log2
##  Min.   :-1.170      Min.   :-9.432      Min.   :-10.187
##  1st Qu.: 0.640      1st Qu.: 1.732      1st Qu.:  1.788
##  Median : 1.052      Median : 2.412      Median :  2.885
##  Mean   : 1.241      Mean   : 2.387      Mean   :  2.846
##  3rd Qu.: 1.790      3rd Qu.: 3.119      3rd Qu.:  3.770
##  Max.   : 5.286      Max.   : 7.746      Max.   :  7.040
##
##  hs_pcb153_cadj_Log2 hs_pcb153_madj_Log2 hs_pcb170_cadj_Log2
##  Min.   :1.207       Min.   :1.110       Min.   :-16.8417
##  1st Qu.:2.863       1st Qu.:2.852       1st Qu.: -0.8820
##  Median :3.496       Median :3.814       Median :  0.1626
##  Mean   :3.557       Mean   :3.863       Mean   : -0.3893
##  3rd Qu.:4.274       3rd Qu.:4.748       3rd Qu.:  1.3776
##  Max.   :7.764       Max.   :7.744       Max.   :  4.7832
##
##  hs_pcb170_madj_Log2 hs_pcb180_cadj_Log2 hs_pcb180_madj_Log2
##  Min.   :-2.0418     Min.   :-11.720     Min.   :-3.887
```

```
##  1st Qu.:-0.3211     1st Qu.:  0.653     1st Qu.: 2.014
##  Median : 0.8727     Median :  1.780     Median : 2.933
##  Mean   : 1.0617     Mean   :  1.723     Mean   : 2.907
##  3rd Qu.: 2.1912     3rd Qu.:  2.981     3rd Qu.: 4.036
##  Max.   : 6.4000     Max.   :  5.878     Max.   : 7.280
##
##  hs_sumPCBs5_cadj_Log2 hs_sumPCBs5_madj_Log2 hs_dep_cadj_Log2
##  Min.   :2.182         Min.   :2.299         Min.   :-11.39697
##  1st Qu.:3.843         1st Qu.:3.967         1st Qu.: -1.20577
##  Median :4.601         Median :4.715         Median :  0.90592
##  Mean   :4.645         Mean   :4.838         Mean   :  0.09785
##  3rd Qu.:5.414         3rd Qu.:5.707         3rd Qu.:  2.29591
##  Max.   :9.277         Max.   :9.341         Max.   :  9.37669
##
##  hs_dep_madj_Log2   hs_detp_cadj_Log2  hs_detp_madj_Log2  hs_dmdtp_cdich_None
##  Min.   :-13.4083   Min.   :-15.4450   Min.   :-28.3791   Detected  :155
##  1st Qu.:  0.9887   1st Qu.: -5.1744   1st Qu.: -4.0972   Undetected:757
##  Median :  1.6288   Median : -3.4476   Median : -0.5251
##  Mean   :  1.6782   Mean   : -2.5073   Mean   : -1.6054
##  3rd Qu.:  2.6686   3rd Qu.:  0.7415   3rd Qu.:  1.0079
##  Max.   :  7.5853   Max.   :  6.2939   Max.   :  5.4700
##
##  hs_dmp_cadj_Log2  hs_dmp_madj_Log2  hs_dmtp_cadj_Log2 hs_dmtp_madj_Log2
##  Min.   :-16.642   Min.   :-17.141   Min.   :-10.645   Min.   :-15.327
##  1st Qu.: -4.951   1st Qu.:  2.011   1st Qu.:  0.288   1st Qu.:  1.046
##  Median : -2.652   Median :  2.731   Median :  1.525   Median :  2.223
##  Mean   : -1.612   Mean   :  2.183   Mean   :  1.091   Mean   :  1.578
##  3rd Qu.:  2.154   3rd Qu.:  3.689   3rd Qu.:  2.675   3rd Qu.:  3.367
##  Max.   :  6.379   Max.   :  7.327   Max.   :  8.664   Max.   :  7.648
##
##  hs_pbde153_cadj_Log2 hs_pbde153_madj_Log2 hs_pbde47_cadj_Log2
##  Min.   :-17.631      Min.   :-15.0030     Min.   :-15.357
##  1st Qu.: -7.909      1st Qu.: -1.7346     1st Qu.: -2.728
##  Median : -2.552      Median : -0.8460     Median : -2.132
##  Mean   : -4.494      Mean   : -1.7038     Mean   : -2.598
##  3rd Qu.: -1.246      3rd Qu.: -0.0321     3rd Qu.: -1.554
##  Max.   :  3.043      Max.   :  6.4338     Max.   :  3.634
##
##  hs_pbde47_madj_Log2 hs_pfhxs_c_Log2   hs_pfhxs_m_Log2    hs_pfna_c_Log2
##  Min.   :-11.5808    Min.   :-8.8953   Min.   :-17.8296   Min.   :-8.1484
##  1st Qu.: -1.6681    1st Qu.:-2.3811   1st Qu.: -1.7262   1st Qu.:-1.7534
##  Median : -0.9383    Median :-1.4718   Median : -0.9284   Median :-1.0525
##  Mean   : -0.7591    Mean   :-1.5906   Mean   : -0.9878   Mean   :-1.0792
##  3rd Qu.:  0.1183    3rd Qu.:-0.7265   3rd Qu.: -0.1574   3rd Qu.:-0.4449
##  Max.   :  5.1183    Max.   : 4.8309   Max.   :  3.7592   Max.   : 2.7050
##
##  hs_pfna_m_Log2      hs_pfoa_c_Log2    hs_pfoa_m_Log2    hs_pfos_c_Log2
##  Min.   :-10.75405   Min.   :-2.2197   Min.   :-5.4760   Min.   :-10.4131
##  1st Qu.: -1.37028   1st Qu.: 0.2688   1st Qu.: 0.3851   1st Qu.:  0.3697
##  Median : -0.60327   Median : 0.6366   Median : 1.1748   Median :  1.0007
##  Mean   : -0.79298   Mean   : 0.6155   Mean   : 1.0324   Mean   :  0.9670
##  3rd Qu.:  0.04944   3rd Qu.: 0.9540   3rd Qu.: 1.7615   3rd Qu.:  1.6747
##  Max.   :  2.56486   Max.   : 2.7352   Max.   : 3.6052   Max.   :  5.0801
##
```

```
##  hs_pfos_m_Log2   hs_pfunda_c_Log2  hs_pfunda_m_Log2    hs_bpa_cadj_Log2
##  Min.   :-1.824   Min.   :-11.453   Min.   :-26.21246   Min.   :-6.072
##  1st Qu.: 1.934   1st Qu.: -5.009   1st Qu.: -3.21222   1st Qu.: 1.266
##  Median : 2.609   Median : -4.065   Median : -2.46861   Median : 2.014
##  Mean   : 2.515   Mean   : -4.224   Mean   : -2.67967   Mean   : 2.129
##  3rd Qu.: 3.202   3rd Qu.: -3.272   3rd Qu.: -1.71446   3rd Qu.: 2.846
##  Max.   : 5.584   Max.   :  0.593   Max.   : -0.04217   Max.   : 7.474
##
##  hs_bpa_madj_Log2   hs_bupa_cadj_Log2 hs_bupa_madj_Log2 hs_etpa_cadj_Log2
##  Min.   :-11.0198   Min.   :-13.940   Min.   :-15.578   Min.   :-6.0647
##  1st Qu.:  0.2552   1st Qu.: -4.350   1st Qu.: -1.237   1st Qu.:-1.1914
##  Median :  1.1363   Median : -3.460   Median :  1.577   Median :-0.5739
##  Mean   :  1.4279   Mean   : -3.532   Mean   :  1.125   Mean   :-0.1661
##  3rd Qu.:  2.2973   3rd Qu.: -2.525   3rd Qu.:  3.792   3rd Qu.: 0.2765
##  Max.   :  6.7361   Max.   :  6.597   Max.   :  7.910   Max.   :10.9895
##
##  hs_etpa_madj_Log2 hs_mepa_cadj_Log2 hs_mepa_madj_Log2 hs_oxbe_cadj_Log2
##  Min.   :-12.119   Min.   :-6.907    Min.   :-0.3096   Min.   :-3.95024
##  1st Qu.:  1.240   1st Qu.: 1.691    1st Qu.: 5.9880   1st Qu.:-0.07006
##  Median :  3.280   Median : 2.705    Median : 7.8549   Median : 1.17068
##  Mean   :  3.385   Mean   : 3.366    Mean   : 7.4038   Mean   : 1.50852
##  3rd Qu.:  5.212   3rd Qu.: 4.562    3rd Qu.: 8.7403   3rd Qu.: 2.85061
##  Max.   : 12.726   Max.   :14.549    Max.   :15.2601   Max.   :12.96310
##
##  hs_oxbe_madj_Log2  hs_prpa_cadj_Log2  hs_prpa_madj_Log2 hs_trcs_cadj_Log2
##  Min.   :-10.5100   Min.   :-12.0208   Min.   :-14.154   Min.   :-4.3599
##  1st Qu.:  0.8008   1st Qu.: -4.3345   1st Qu.:  3.973   1st Qu.:-1.6135
##  Median :  2.5544   Median : -2.2631   Median :  5.775   Median :-0.7975
##  Mean   :  3.0604   Mean   : -1.5893   Mean   :  5.395   Mean   :-0.3496
##  3rd Qu.:  4.7874   3rd Qu.:  0.8664   3rd Qu.:  7.282   3rd Qu.: 0.5012
##  Max.   : 13.6480   Max.   : 10.7801   Max.   : 13.605   Max.   : 9.2782
##
##  hs_trcs_madj_Log2 hs_mbzp_cadj_Log2 hs_mbzp_madj_Log2 hs_mecpp_cadj_Log2
##  Min.   :-3.3986   Min.   :-0.5586   Min.   :-2.369    Min.   : 2.631
##  1st Qu.: 0.5526   1st Qu.: 1.6270   1st Qu.: 1.861    1st Qu.: 4.439
##  Median : 2.7390   Median : 2.3188   Median : 2.838    Median : 5.150
##  Mean   : 3.4690   Mean   : 2.3981   Mean   : 2.956    Mean   : 5.195
##  3rd Qu.: 6.7458   3rd Qu.: 3.0441   3rd Qu.: 4.088    3rd Qu.: 5.909
##  Max.   :10.6909   Max.   : 7.1847   Max.   : 9.304    Max.   :10.628
##
##  hs_mecpp_madj_Log2 hs_mehhp_cadj_Log2 hs_mehhp_madj_Log2 hs_mehp_cadj_Log2
##  Min.   : 2.681     Min.   : 1.820     Min.   :0.8144     Min.   :-1.6330
##  1st Qu.: 4.327     1st Qu.: 3.652     1st Qu.:3.4418     1st Qu.: 0.8224
##  Median : 4.851     Median : 4.357     Median :4.0677     Median : 1.5652
##  Mean   : 5.027     Mean   : 4.404     Mean   :4.1231     Mean   : 1.6053
##  3rd Qu.: 5.594     3rd Qu.: 5.045     3rd Qu.:4.7489     3rd Qu.: 2.3105
##  Max.   :10.411     Max.   :11.130     Max.   :9.9176     Max.   : 8.1407
##
##  hs_mehp_madj_Log2 hs_meohp_cadj_Log2 hs_meohp_madj_Log2 hs_mep_cadj_Log2
##  Min.   :-7.469    Min.   : 1.138     Min.   :-0.0179    Min.   : 2.201
##  1st Qu.: 1.748    1st Qu.: 2.923     1st Qu.: 3.0978    1st Qu.: 4.058
##  Median : 3.006    Median : 3.625     Median : 3.6731    Median : 5.006
##  Mean   : 2.930    Mean   : 3.702     Mean   : 3.7746    Mean   : 5.247
##  3rd Qu.: 3.808    3rd Qu.: 4.356     3rd Qu.: 4.4155    3rd Qu.: 6.205
```

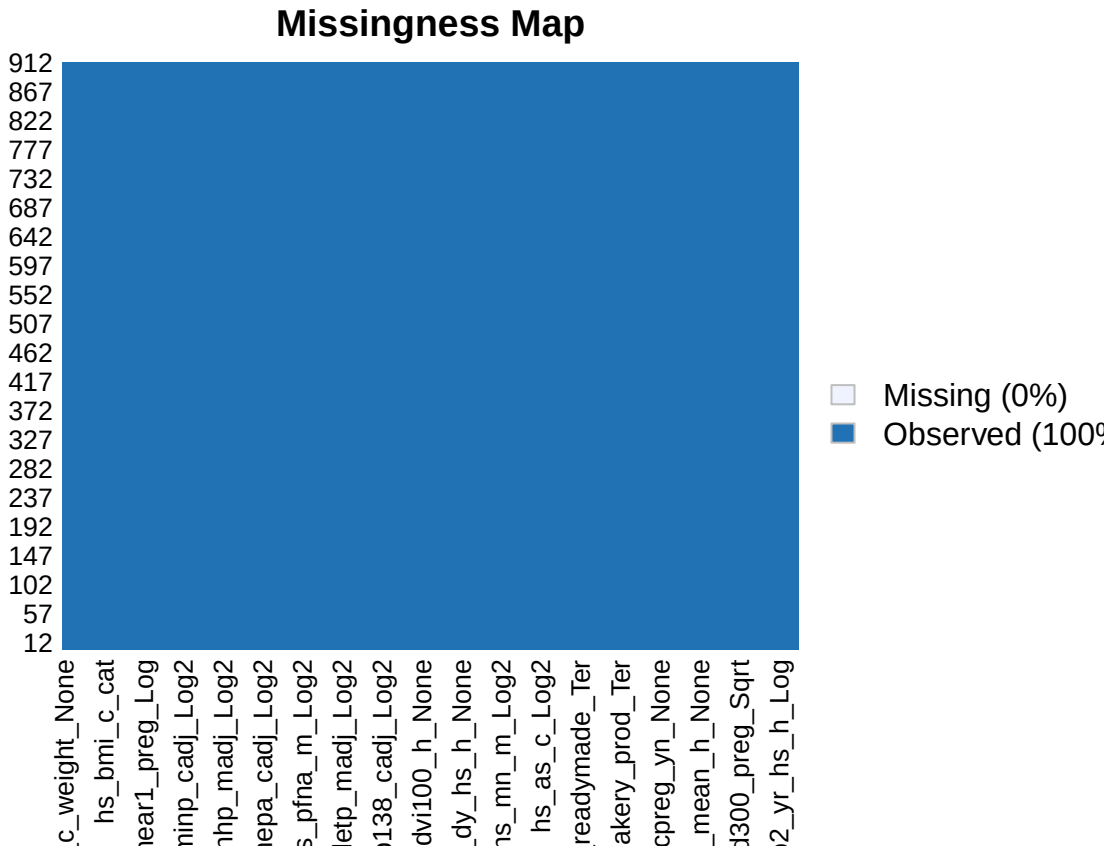
```
##  Max.   : 8.702    Max.   :10.332     Max.   : 9.6122    Max.   :11.540
##
##  hs_mep_madj_Log2 hs_mibp_cadj_Log2 hs_mibp_madj_Log2 hs_mnbp_cadj_Log2
##  Min.   : 3.292   Min.   :2.463     Min.   :2.374     Min.   :1.866
##  1st Qu.: 6.403   1st Qu.:4.727     1st Qu.:4.590     1st Qu.:3.987
##  Median : 7.862   Median :5.448     Median :5.349     Median :4.641
##  Mean   : 7.800   Mean   :5.465     Mean   :5.310     Mean   :4.701
##  3rd Qu.: 8.957   3rd Qu.:6.206     3rd Qu.:5.898     3rd Qu.:5.320
##  Max.   :14.114   Max.   :9.750     Max.   :9.461     Max.   :8.885
##
##  hs_mnbp_madj_Log2 hs_ohminp_cadj_Log2 hs_ohminp_madj_Log2 hs_oxominp_cadj_Log2
##  Min.   : 1.443    Min.   :-0.2821     Min.   :-11.4619    Min.   :-0.9126
##  1st Qu.: 4.186    1st Qu.: 1.6984     1st Qu.: -0.7237    1st Qu.: 0.8919
##  Median : 4.855    Median : 2.4064     Median : -0.2285    Median : 1.5178
##  Mean   : 4.935    Mean   : 2.5786     Mean   : -0.3112    Mean   : 1.6726
##  3rd Qu.: 5.511    3rd Qu.: 3.1363     3rd Qu.:  0.2496    3rd Qu.: 2.2830
##  Max.   :12.654    Max.   : 9.0983     Max.   :  6.0560    Max.   : 9.4093
##
##  hs_oxominp_madj_Log2 hs_sumDEHP_cadj_Log2 hs_sumDEHP_madj_Log2 FAS_cat_None
##  Min.   :-11.55154    Min.   : 2.648       Min.   : 3.211       Low   :105
##  1st Qu.: -0.69643    1st Qu.: 5.247       1st Qu.: 5.218       Middle:340
##  Median : -0.03874    Median : 6.003       Median : 5.826       High  :467
##  Mean   : -0.09366    Mean   : 6.040       Mean   : 6.002
##  3rd Qu.:  0.51914    3rd Qu.: 6.793       3rd Qu.: 6.697
##  Max.   :  5.55327    Max.   :10.052       Max.   :11.691
##
##        hs_contactfam_3cat_num_None hs_hm_pers_None
##  (almost) Daily       :597          Min.   : 1.00
##  Once a week          :276          1st Qu.: 4.00
##  Less than once a week: 39          Median : 4.00
##                                     Mean   : 4.25
##                                     3rd Qu.: 5.00
##                                     Max.   :10.00
##
##           hs_participation_3cat_None e3_asmokcigd_p_None
##  None                    :518          Min.   : 0.0000
##  1 organisation          :256          1st Qu.: 0.0000
##  2 or more organisations:138           Median : 0.0000
##                                        Mean   : 0.5034
##                                        3rd Qu.: 0.0000
##                                        Max.   :15.2381
##
##  hs_cotinine_cdich_None hs_cotinine_mcat_None   hs_globalexp2_None
##  Detected  :149         Non-smokers:544        exposure   :328
##  Undetected:763         SHS smokers:102        no exposure:584
##                         Smokers    :266
##
##
##
##
##  hs_smk_parents_None h_distinvnear1_preg_Log h_trafload_preg_pow1over3
##  both   :103         Min.   :-10.022         Min.   :  0.3458
##  neither:570         1st Qu.: -3.985         1st Qu.: 34.1674
##  one    :239         Median : -3.021         Median : 66.6101
```

```
##                     Mean   : -3.179         Mean   : 74.4569
##                     3rd Qu.: -2.339         3rd Qu.:113.0812
##                     Max.   :  2.794         Max.   :294.2705
##
##  h_trafnear_preg_pow1over3 hs_trafload_h_pow1over3 hs_trafnear_h_pow1over3
##  Min.   : 0.000            Min.   :  0.00          Min.   : 0.000
##  1st Qu.: 7.937            1st Qu.: 77.42          1st Qu.: 8.108
##  Median :12.213            Median :114.87          Median :14.794
##  Mean   :14.967            Mean   :111.45          Mean   :15.876
##  3rd Qu.:21.397            3rd Qu.:136.00          3rd Qu.:22.104
##  Max.   :39.321            Max.   :293.58          Max.   :49.348
##
##  h_bro_preg_Log    h_clf_preg_Log    h_thm_preg_Log       e3_bw
##  Min.   :-2.9759   Min.   :-6.9078   Min.   :-1.600   Min.   :1100
##  1st Qu.:-0.5009   1st Qu.:-0.4809   1st Qu.: 1.879   1st Qu.:3080
##  Median : 1.8387   Median : 2.1043   Median : 2.896   Median :3399
##  Mean   : 1.2031   Mean   : 0.9940   Mean   : 2.704   Mean   :3386
##  3rd Qu.: 2.7488   3rd Qu.: 3.2506   3rd Qu.: 3.839   3rd Qu.:3720
##  Max.   : 4.8916   Max.   : 3.8237   Max.   : 5.031   Max.   :5260
##
##    hs_asthma       hs_zbmi_who      hs_correct_raven   hs_Gen_Tot
##  Min.   :0.0000   Min.   :-3.5800   Min.   : 9.00    Min.   :  0.00
##  1st Qu.:0.0000   1st Qu.:-0.4200   1st Qu.:22.00    1st Qu.: 10.05
##  Median :0.0000   Median : 0.2800   Median :27.00    Median : 21.00
##  Mean   :0.1129   Mean   : 0.3949   Mean   :26.33    Mean   : 24.79
##  3rd Qu.:0.0000   3rd Qu.: 1.1300   3rd Qu.:32.00    3rd Qu.: 34.08
##  Max.   :1.0000   Max.   : 4.7200   Max.   :36.00    Max.   :133.00
##
##  hs_bmi_c_cat h_cohort e3_sex_None  e3_yearbir_None  h_mbmi_None
##  1:  8        1:146    female:421   2003: 39        Min.   :15.88
##  2:633        2:133    male  :491   2004: 69        1st Qu.:21.45
##  3:183        3:150                 2005:164        Median :24.17
##  4: 88        4:155                 2006:179        Mean   :25.12
##               5:194                 2007:181        3rd Qu.:27.48
##               6:134                 2008:271        Max.   :51.42
##                                     2009:  9
##  hs_wgtgain_None  e3_gac_None      h_age_None    h_edumc_None h_native_None
##  Min.   : 0.0    Min.   :28.00   Min.   :16.00   1:125        0:103
##  1st Qu.:10.0    1st Qu.:38.71   1st Qu.:27.50   2:303        1: 53
##  Median :12.0    Median :40.00   Median :31.00   3:484        2:756
##  Mean   :13.6    Mean   :39.58   Mean   :30.69
##  3rd Qu.:18.0    3rd Qu.:40.71   3rd Qu.:34.00
##  Max.   :55.0    Max.   :44.00   Max.   :43.00
##
##  h_parity_None hs_child_age_None hs_c_height_None hs_c_weight_None
##  0:437         Min.   : 5.437    Min.   :1.060    Min.   :16.00
##  1:312         1st Qu.: 6.491    1st Qu.:1.203    1st Qu.:22.70
##  2:163         Median : 7.932    Median :1.280    Median :26.70
##                Mean   : 7.945    Mean   :1.289    Mean   :28.40
##                3rd Qu.: 8.893    3rd Qu.:1.359    3rd Qu.:32.52
##                Max.   :12.068    Max.   :1.685    Max.   :71.10
##
```

```
#view(codebook)
summary(phenotype)
```

```
##        ID            e3_bw        hs_asthma       hs_zbmi_who
##  Min.   :   1   Min.   :1100   Min.   :0.0000   Min.   :-3.5800
##  1st Qu.: 326   1st Qu.:3080   1st Qu.:0.0000   1st Qu.:-0.4000
##  Median : 651   Median :3398   Median :0.0000   Median : 0.2800
##  Mean   : 651   Mean   :3389   Mean   :0.1091   Mean   : 0.4032
##  3rd Qu.: 976   3rd Qu.:3720   3rd Qu.:0.0000   3rd Qu.: 1.1300
##  Max.   :1301   Max.   :5260   Max.   :1.0000   Max.   : 4.7200
##  hs_correct_raven   hs_Gen_Tot     hs_bmi_c_cat
##  Min.   : 9.00    Min.   :  0.00   1: 13
##  1st Qu.:22.00    1st Qu.: 10.00   2:904
##  Median :27.00    Median : 20.00   3:253
##  Mean   :26.29    Mean   : 24.38   4:131
##  3rd Qu.:32.00    3rd Qu.: 33.44
##  Max.   :36.00    Max.   :133.00
```

```
missmap(train.data)
```




```
#observe missing, 0%
```

PM10 value during pregnancy * min: 8.066 * median: 22.796 * mean: 23.393

NO2 value during pregnancy * min: 2.105 * median: 2.959 * mean: 3.001

PM25 value during pregnancy * min: 6.957 * median: 14.880 * mean: 15.002

Landuse Shannon's Evenness Indexat school * min: 0 * median: 0.4015 * mean: 0.3985

Walkability (postnatal) score * min: 0.1 * median: 0.3 * mean: 0.3267

asthma * min: 0 * median: 0 * mean: 0.1091

Step 2: Research Question

Put your Research Question in this section. It can be a prediction question OR it can be a hypothesis-generating question about either combinations of features or interactions between features.

Does consumption of legumes, folic acid supplements, fruit, vegetables, and organic foods during pregnancy predict the BMI category of the child at 6-11 years old?

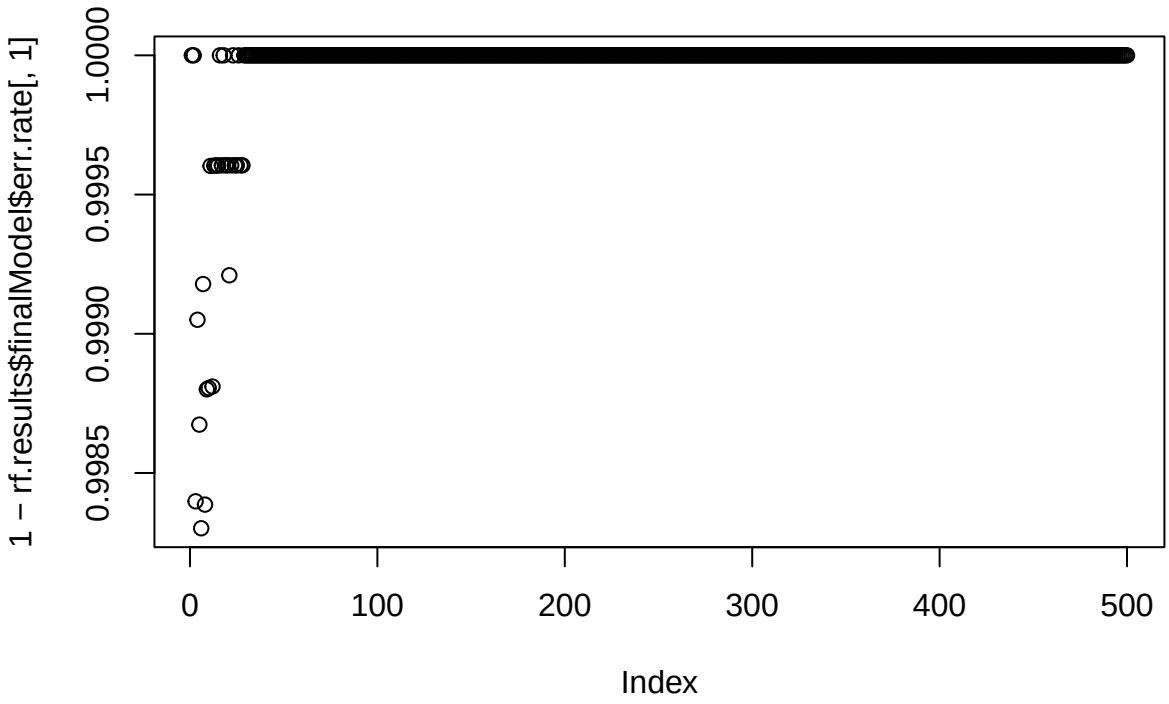
Step 3: Implement pipeline to address research question

You only need to implement a single algorithm to address your research question.Tune hyperparameters to obtain optimal model in training then evaluate in test set.

```
set.seed(100)

training.data<-studydata$hs_bmi_c_cat %>%
  createDataPartition(p=0.7, list=F)

train.data2<-studydata[training.data, ]
test.data2<-studydata[-training.data, ]

train.control.df = trainControl(method = "cv", number = 5, sampling = "up")

mtry.df = c(ncol(train.data)-1, sqrt(ncol(train.data2)-1), 0.5*ncol(train.data2)-1)
mtry.grid = expand.grid(.mtry=mtry.df)
rf.results = train(hs_bmi_c_cat ~., data = train.data2, method = "rf", metric = "Accuracy", tuneGrid = m

rf.results$results
```

```
##        mtry  Accuracy     Kappa  AccuracySD    KappaSD
## 1  15.49193 0.8378448 0.5848201 0.020507320 0.06138460
## 2 119.50000 0.9923254 0.9835200 0.004908378 0.01059066
## 3 240.00000 1.0000000 1.0000000 0.000000000 0.00000000
```

```
plot(1-rf.results$finalModel$err.rate[,1])
```




```
varImp(rf.results)
```

```
## rf variable importance
##
##   only 20 most important variables shown (out of 294)
##
##                              Overall
## hs_zbmi_who                100.00000
## hs_c_weight_None             6.10553
## hs_pcb170_cadj_Log2          0.31452
## hs_pfhxs_m_Log2              0.29215
## hs_bakery_prod_Ter(6,Inf]    0.16150
## hs_hcb_cadj_Log2             0.14947
## hs_pbde153_cadj_Log2         0.14350
## hs_child_age_None            0.13318
## hs_trafload_h_pow1over3      0.09820
## hs_pm25_yr_hs_h_None         0.06435
## hs_dmtp_cadj_Log2            0.04944
## hs_pm10_yr_hs_h_None         0.04831
## hs_pm25abs_yr_hs_h_Log       0.04257
## hs_pcb153_madj_Log2          0.04078
## hs_builtdens300_s_Sqrt       0.03441
## hs_pcb180_madj_Log2          0.03353
## hs_landuseshan300_h_None     0.02679
## hs_pfos_c_Log2               0.02191
## hs_pcb180_cadj_Log2          0.01699
## hs_as_c_Log2                 0.01689
```

```
confusionMatrix(rf.results)
```

```
## Cross-Validated (5 fold) Confusion Matrix
##
## (entries are percentual average cell counts across resamples)
```

```
##
##           Reference
## Prediction    1    2    3    4
##          1  1.1  0.0  0.0  0.0
##          2  0.0 69.3  0.0  0.0
##          3  0.0  0.0 19.5  0.0
##          4  0.0  0.0  0.0 10.1
##
##  Accuracy (average) : 1
```

```
#accuracy 1: 0.8433061
#accuracy 2: 0.9967153
#accuracy 3: 0.9989071

rf.test = predict(rf.results, test.data)
confusionMatrix(rf.test, test.data$hs_bmi_c_cat)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   1   2   3   4
##          1   5   0   0   0
##          2   0 271   0   0
##          3   0   0  70   0
##          4   0   0   0  43
##
## Overall Statistics
##
##                Accuracy : 1
##                  95% CI : (0.9906, 1)
##     No Information Rate : 0.6967
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 1
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: 1 Class: 2 Class: 3 Class: 4
## Sensitivity           1.00000   1.0000   1.0000   1.0000
## Specificity           1.00000   1.0000   1.0000   1.0000
## Pos Pred Value        1.00000   1.0000   1.0000   1.0000
## Neg Pred Value        1.00000   1.0000   1.0000   1.0000
## Prevalence            0.01285   0.6967   0.1799   0.1105
## Detection Rate        0.01285   0.6967   0.1799   0.1105
## Detection Prevalence  0.01285   0.6967   0.1799   0.1105
## Balanced Accuracy     1.00000   1.0000   1.0000   1.0000
```

OPTIONAL: Create Models to examine whether two features interact using linear regression

This is a demonstration of code for some interaction analyses.

Note I'm not scaling before running my glm models. If this were a prediction question, I would likely scale so that my coefficients would be interpretable for variable importance. But this is just to show how one codes interaction terms in R using glm. Would be similar if you used within the caret framework. I'm also showing

how you would code interaction terms within an elastic net framework using caret.

You can replace the features here with features from your own research question if you'd like to being exploring interactions using linear regression and elastic net. Model.1 and Model.2.a/b are just linear regression while Model 3 is an elastic net that automates examining interactions.

```
#Model 1: Three features, indoor NO2, building density and walkability metric, in relation to child bir

model.1<-glm(e3_bw~h_NO2_Log+h_builtdens300_preg_Sqrt+h_walkability_mean_preg_None, data=train.data)
summary(model.1)
```

```
##
## Call:
## glm(formula = e3_bw ~ h_NO2_Log + h_builtdens300_preg_Sqrt +
##     h_walkability_mean_preg_None, data = train.data)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2295.46   -304.10      7.59    322.87   1842.79
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 3550.4376    89.0680  39.862  < 2e-16 ***
## h_NO2_Log                     18.8989    16.9342   1.116  0.26471
## h_builtdens300_preg_Sqrt      -0.3659     0.1356  -2.699  0.00709 **
## h_walkability_mean_preg_None -314.2298   254.5668  -1.234  0.21738
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 266091.5)
##
##     Null deviance: 245289824  on 911  degrees of freedom
## Residual deviance: 241611105  on 908  degrees of freedom
## AIC: 13986
##
## Number of Fisher Scoring iterations: 2
```

```
#Model 2a: Including an interaction term between two features
model.2a<-glm(e3_bw~h_NO2_Log+h_builtdens300_preg_Sqrt+h_walkability_mean_preg_None+h_NO2_Log*h_builtde
summary(model.2a)
```

```
##
## Call:
## glm(formula = e3_bw ~ h_NO2_Log + h_builtdens300_preg_Sqrt +
##     h_walkability_mean_preg_None + h_NO2_Log * h_builtdens300_preg_Sqrt,
##     data = train.data)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2293.19   -309.09      6.93    321.83   1840.10
##
## Coefficients:
##                                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       3451.92800  210.50782  16.398   <2e-16 ***
## h_NO2_Log                           44.85385   53.02900   0.846    0.398
## h_builtdens300_preg_Sqrt            -0.13821    0.46124  -0.300    0.765
```

```
## h_walkability_mean_preg_None       -322.94791  255.22836  -1.265    0.206
## h_NO2_Log:h_builtdens300_preg_Sqrt   -0.05712    0.11059  -0.517    0.606
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 266306.6)
##
##     Null deviance: 245289824  on 911  degrees of freedom
## Residual deviance: 241540058  on 907  degrees of freedom
## AIC: 13988
##
## Number of Fisher Scoring iterations: 2
```

```
#Model 2b: Including all combinations of two-way interactions using shortcut in glm
model.2b<-glm(e3_bw~(h_NO2_Log+h_builtdens300_preg_Sqrt+h_NO2_Log+h_walkability_mean_preg_None)^2, data=
summary(model.2b)
```

```
##
## Call:
## glm(formula = e3_bw ~ (h_NO2_Log + h_builtdens300_preg_Sqrt +
##     h_NO2_Log + h_walkability_mean_preg_None)^2, data = train.data)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2270.49   -306.36     14.61    328.29   1844.88
##
## Coefficients:
##                                                    Estimate Std. Error
## (Intercept)                                      3581.13791  321.64922
## h_NO2_Log                                         111.76811   74.15910
## h_builtdens300_preg_Sqrt                           -1.19243    0.66729
## h_walkability_mean_preg_None                     -694.03821 1151.43231
## h_NO2_Log:h_builtdens300_preg_Sqrt                 -0.02168    0.11727
## h_NO2_Log:h_walkability_mean_preg_None           -301.10061  250.65986
## h_builtdens300_preg_Sqrt:h_walkability_mean_preg_None    3.49642    1.76355
##                                                  t value Pr(>|t|)
## (Intercept)                                       11.134   <2e-16 ***
## h_NO2_Log                                          1.507   0.1321
## h_builtdens300_preg_Sqrt                          -1.787   0.0743 .
## h_walkability_mean_preg_None                      -0.603   0.5468
## h_NO2_Log:h_builtdens300_preg_Sqrt                -0.185   0.8534
## h_NO2_Log:h_walkability_mean_preg_None            -1.201   0.2300
## h_builtdens300_preg_Sqrt:h_walkability_mean_preg_None   1.983   0.0477 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 265479)
##
##     Null deviance: 245289824  on 911  degrees of freedom
## Residual deviance: 240258503  on 905  degrees of freedom
## AIC: 13987
##
## Number of Fisher Scoring iterations: 2
```

```r
#Model 3: Using the caret framework to run an elastic-net with interaction terms between all features u
set.seed(100)

model.3<- train(
  e3_bw ~(h_NO2_Log+h_builtdens300_preg_Sqrt+h_NO2_Log+h_walkability_mean_preg_None)^2, data = train.da
  trControl = trainControl("cv", number = 5),
 tuneLength=10
  )
#Print the values of alpha and lambda that gave best prediction
model.3$bestTune
```

```
##    alpha   lambda
## 10   0.1 48.59348
```

```r
#Examine model coeficients for variable importance
coef(model.3$finalModel, model.3$bestTune$lambda)
```

```
## 7 x 1 sparse Matrix of class "dgCMatrix"
##                                                           s1
## (Intercept)                                         3542.26096
## h_NO2_Log                                             11.42824
## h_builtdens300_preg_Sqrt                             -42.76895
## h_walkability_mean_preg_None                         -20.03193
## h_NO2_Log:h_builtdens300_preg_Sqrt                     .
## h_NO2_Log:h_walkability_mean_preg_None                 .
## h_builtdens300_preg_Sqrt:h_walkability_mean_preg_None  .
```

```r
#Predict in test-set
model.3.pred <- model.3 %>% predict(test.data)

# Evaluation metrics and prediction performance
data.frame(
  RMSE = RMSE(model.3.pred, test.data$e3_bw),
  Rsquare = R2(model.3.pred, test.data$e3_bw)
)
```

```
##       RMSE     Rsquare
## 1 486.3721 0.008978641
```