# Preg_lifestyle

Kazi rahman

2022-11-30

## EDA

You can also embed plots, for example:

```
##                      nan_count
## hs_correct_raven            10
## e3_alcpreg_yn_None           0
## h_cereal_preg_Ter            0
## h_dairy_preg_Ter             0
## h_fastfood_preg_Ter          0
## h_fish_preg_Ter              0
## h_folic_t1_None              0
## h_fruit_preg_Ter             0
## h_legume_preg_Ter            0
## h_meat_preg_Ter              0
## h_pamod_t3_None              0
## h_pavig_t3_None              0
## h_veg_preg_Ter               0
```

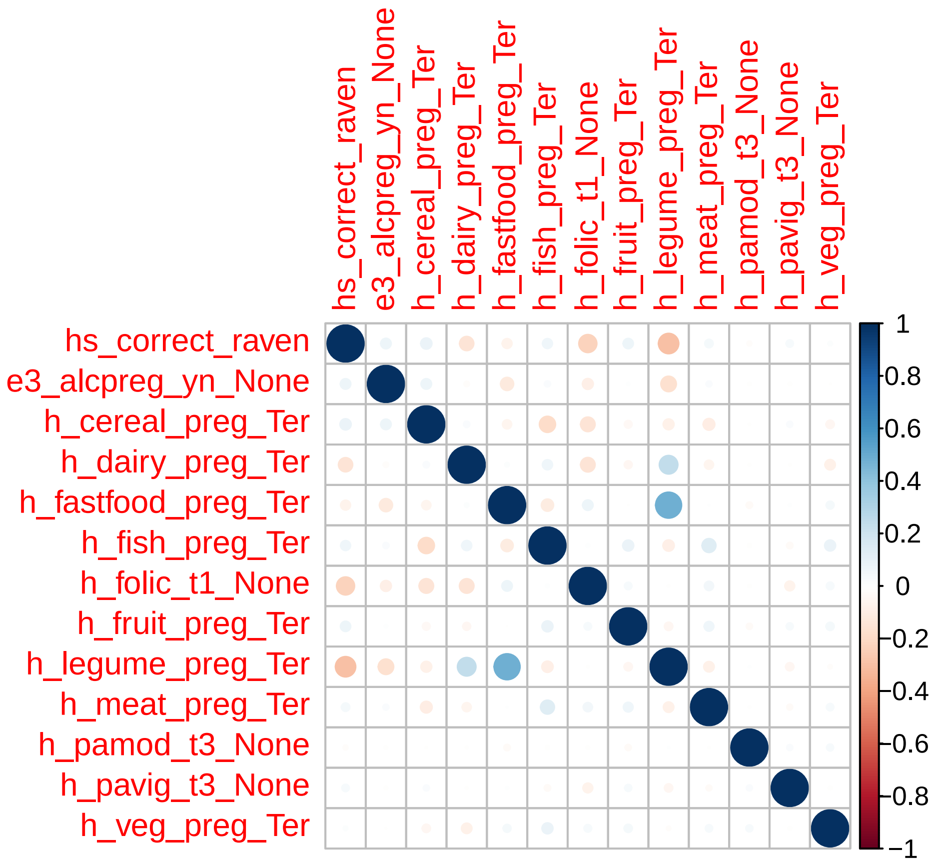
Our covariate of interest or y, has 10 Nan values. The rest of our covariates do not have any Nans which is a good sign. We have established a cleaner dataset now. We can proceed onto visualizing our data to get a better understanding

```
data_matrix<-data.matrix(data)
corr<-cor(data_matrix)
corr
```

```
##                     hs_correct_raven e3_alcpreg_yn_None h_cereal_preg_Ter
## hs_correct_raven          1.00000000       0.0749008262       0.085656914
## e3_alcpreg_yn_None        0.07490083       1.0000000000       0.075155471
## h_cereal_preg_Ter         0.08565691       0.0751554710       1.000000000
## h_dairy_preg_Ter         -0.14085376      -0.0181098988       0.025840810
## h_fastfood_preg_Ter      -0.06631836      -0.1192210952      -0.056468111
## h_fish_preg_Ter           0.06716006       0.0234868613      -0.182609314
## h_folic_t1_None          -0.22997437      -0.0810431040      -0.147859222
## h_fruit_preg_Ter          0.07121111       0.0063250772      -0.039697963
## h_legume_preg_Ter        -0.29503969      -0.1681220825      -0.079638477
## h_meat_preg_Ter           0.04681100       0.0219518432      -0.094071254
## h_pamod_t3_None          -0.01431386      -0.0066895552      -0.002525895
## h_pavig_t3_None           0.03528801      -0.0070849719       0.023055192
## h_veg_preg_Ter            0.01209541      -0.0009826795      -0.046019658
##                     h_dairy_preg_Ter h_fastfood_preg_Ter h_fish_preg_Ter
## hs_correct_raven        -0.140853764       -0.0663183608     0.067160062
## e3_alcpreg_yn_None      -0.018109899       -0.1192210952     0.023486861
## h_cereal_preg_Ter        0.025840810       -0.0564681112    -0.182609314
```

```
## h_dairy_preg_Ter        1.000000000      0.0120431809    0.068881178
## h_fastfood_preg_Ter     0.012043181      1.0000000000   -0.102742225
## h_fish_preg_Ter         0.068881178     -0.1027422245    1.000000000
## h_folic_t1_None        -0.147554489      0.0744555624    0.008306321
## h_fruit_preg_Ter       -0.041246298     -0.0001292932    0.082572352
## h_legume_preg_Ter       0.244208660      0.4872679744   -0.087311749
## h_meat_preg_Ter        -0.055833074      0.0011983712    0.135066218
## h_pamod_t3_None        -0.004063939     -0.0265549757   -0.009149816
## h_pavig_t3_None        -0.003601034      0.0047669426   -0.025855117
## h_veg_preg_Ter         -0.074043678      0.0414468185    0.082153237
##                     h_folic_t1_None h_fruit_preg_Ter h_legume_preg_Ter
## hs_correct_raven       -0.229974369     0.0712111111      -0.295039691
## e3_alcpreg_yn_None     -0.081043104     0.0063250772      -0.168122083
## h_cereal_preg_Ter      -0.147859222    -0.0396979629      -0.079638477
## h_dairy_preg_Ter       -0.147554489    -0.0412462977       0.244208660
## h_fastfood_preg_Ter     0.074455562    -0.0001292932       0.487267974
## h_fish_preg_Ter         0.008306321     0.0825723520      -0.087311749
## h_folic_t1_None         1.000000000     0.0366576251      -0.003012139
## h_fruit_preg_Ter        0.036657625     1.0000000000      -0.046456594
## h_legume_preg_Ter      -0.003012139    -0.0464565945       1.000000000
## h_meat_preg_Ter         0.057253635     0.0640007008      -0.074536692
## h_pamod_t3_None        -0.006410175    -0.0244184871       0.002505104
## h_pavig_t3_None        -0.065188778     0.0322609624      -0.045760495
## h_veg_preg_Ter          0.038318248     0.0445872766      -0.011390236
##                     h_meat_preg_Ter h_pamod_t3_None h_pavig_t3_None
## hs_correct_raven        0.046810999    -0.014313863     0.035288007
## e3_alcpreg_yn_None      0.021951843    -0.006689555    -0.007084972
## h_cereal_preg_Ter      -0.094071254    -0.002525895     0.023055192
## h_dairy_preg_Ter       -0.055833074    -0.004063939    -0.003601034
## h_fastfood_preg_Ter     0.001198371    -0.026554976     0.004766943
## h_fish_preg_Ter         0.135066218    -0.009149816    -0.025855117
## h_folic_t1_None         0.057253635    -0.006410175    -0.065188778
## h_fruit_preg_Ter        0.064000701    -0.024418487     0.032260962
## h_legume_preg_Ter      -0.074536692     0.002505104    -0.045760495
## h_meat_preg_Ter         1.000000000    -0.002599790    -0.021578509
## h_pamod_t3_None        -0.002599790     1.000000000     0.022695158
## h_pavig_t3_None        -0.021578509     0.022695158     1.000000000
## h_veg_preg_Ter          0.035266159     0.031903148    -0.008200237
##                     h_veg_preg_Ter
## hs_correct_raven      0.0120954108
## e3_alcpreg_yn_None   -0.0009826795
## h_cereal_preg_Ter    -0.0460196585
## h_dairy_preg_Ter     -0.0740436780
## h_fastfood_preg_Ter   0.0414468185
## h_fish_preg_Ter       0.0821532369
## h_folic_t1_None       0.0383182475
## h_fruit_preg_Ter      0.0445872766
## h_legume_preg_Ter    -0.0113902361
## h_meat_preg_Ter       0.0352661586
## h_pamod_t3_None       0.0319031481
## h_pavig_t3_None      -0.0082002371
## h_veg_preg_Ter        1.0000000000
```

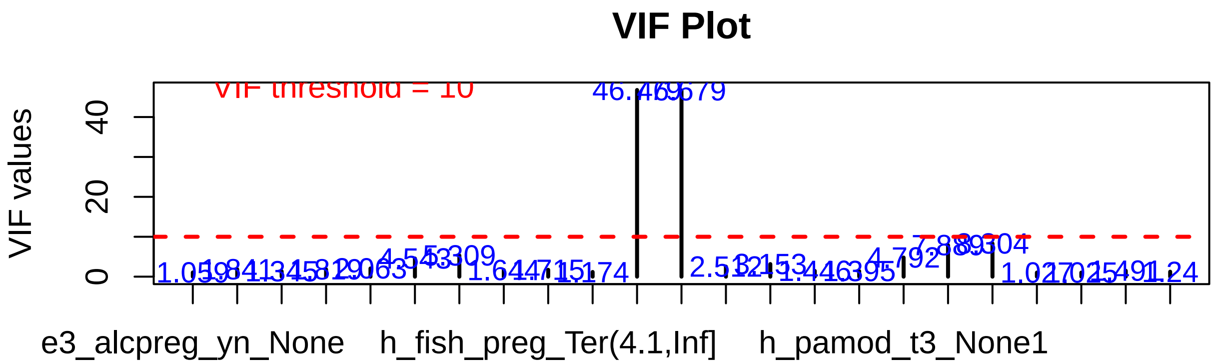
```
corrplot(corr)
```

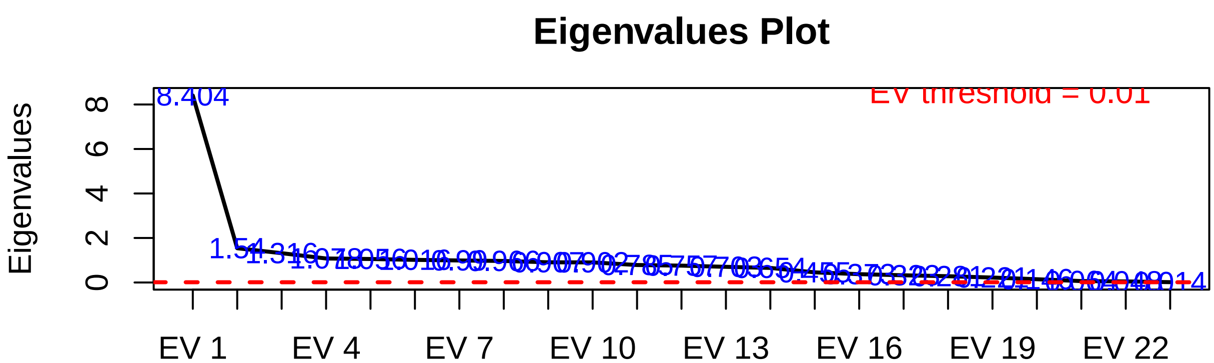



```
vif_values <- vif(M1)
vif_values
```

```
##                         GVIF Df GVIF^(1/(2*Df))
## e3_alcpreg_yn_None  1.058889  1        1.029023
## h_cereal_preg_Ter   1.625913  2        1.129209
## h_dairy_preg_Ter    1.237149  2        1.054643
## h_fastfood_preg_Ter 1.439273  2        1.095307
## h_fish_preg_Ter     1.159154  2        1.037613
## h_folic_t1_None     1.173766  1        1.083405
## h_fruit_preg_Ter    1.107984  2        1.025967
## h_legume_preg_Ter   2.332482  2        1.235818
## h_meat_preg_Ter     1.122058  2        1.029210
## h_pamod_t3_None     1.050353  3        1.008221
## h_pavig_t3_None     1.027930  2        1.006911
## h_veg_preg_Ter      1.286936  2        1.065097
```

```
mc.plot(M1)
```



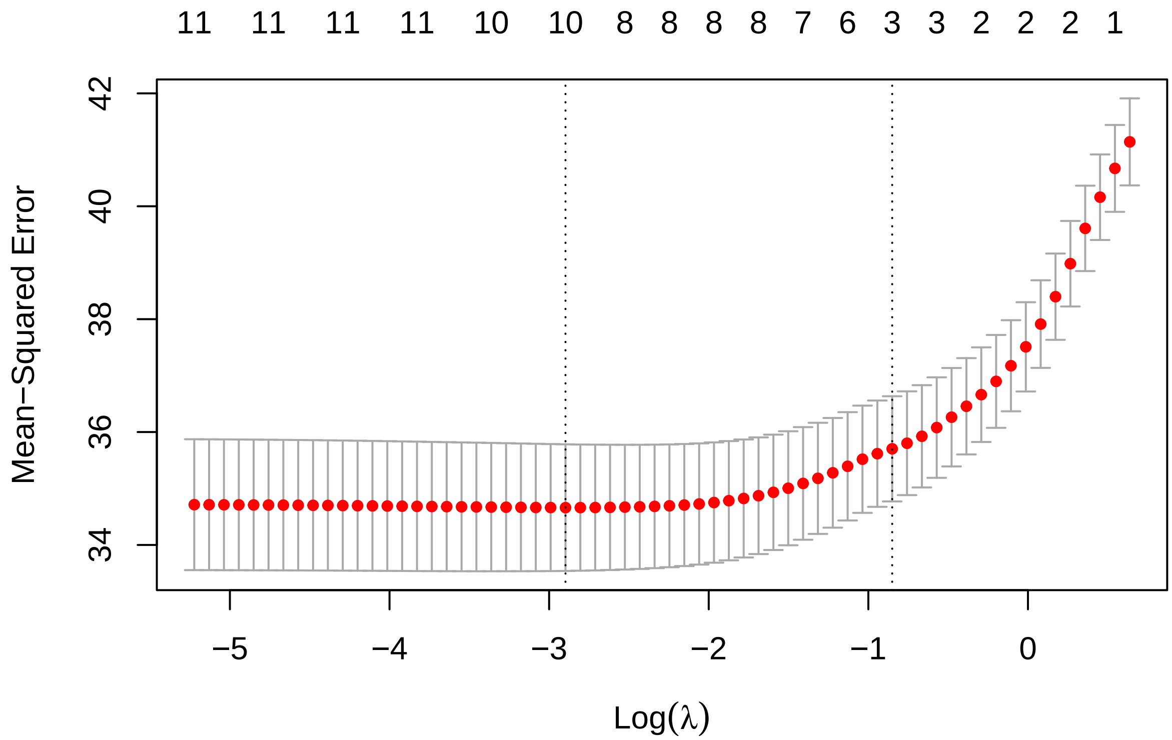

Before proceding to our next step, which will be building the model, we want to perform feature selection and possible test for interactions.

```
x<-data.matrix(data[,c("e3_alcpreg_yn_None" , "h_cereal_preg_Ter" ,
                        "h_dairy_preg_Ter" ,   "h_fastfood_preg_Ter",
                         "h_fish_preg_Ter" ,    "h_folic_t1_None"    ,
                        "h_fruit_preg_Ter"    , "h_legume_preg_Ter"   ,
                        "h_meat_preg_Ter" ,
                        "h_pamod_t3_None" ,    "h_pavig_t3_None"   ,
                        "h_veg_preg_Ter")])

cv_model <- cv.glmnet(x, data$hs_correct_raven, alpha = 1)
best_lambda <- cv_model$lambda.min
best_lambda
```

```
## [1] 0.0551637
```

```
#The lambda value that minimizes the test MSE turns out to be 0.04498289
plot(cv_model)
```




```r
best_model <- glmnet(x, data$hs_correct_raven, alpha = 1, lambda = best_lambda)
coef(best_model)
```

```
## 13 x 1 sparse Matrix of class "dgCMatrix"
##                             s0
## (Intercept)          29.97674697
## e3_alcpreg_yn_None    0.01992317
## h_cereal_preg_Ter     0.28357222
## h_dairy_preg_Ter     -0.79210138
## h_fastfood_preg_Ter   1.00218022
## h_fish_preg_Ter       0.42286780
## h_folic_t1_None      -3.10380331
## h_fruit_preg_Ter      0.69145185
## h_legume_preg_Ter    -2.44085885
## h_meat_preg_Ter       0.14262291
## h_pamod_t3_None      -0.01934883
## h_pavig_t3_None       .
## h_veg_preg_Ter        .
```

```r
test_cov_ind<-which(coef(best_model)==0)
excluding_var<-c()
for (i in test_cov_ind){
  print(names(data[i]))
  excluding_var <- c(excluding_var, names(data[i]))
}
```

```
## [1] "h_pavig_t3_None"
## [1] "h_veg_preg_Ter"
```

```r
print("cov that we may exclude")
```

```
## [1] "cov that we may exclude"
```

```r
print(excluding_var)
```

```
## [1] "h_pavig_t3_None" "h_veg_preg_Ter"
```

```r
DF <- read.csv('preg_data.csv')

df2<-subset(data,select = excluding_var)
df3<-DF[,!names(DF) %in%
      excluding_var]


lass_model<-lm(data$hs_correct_raven~.,data = df3)
M2<-lm(data$hs_correct_raven~.,data = data)
anova(lass_model,M2)
```

```
## Analysis of Variance Table
##
## Model 1: data$hs_correct_raven ~ e3_alcpreg_yn_None + h_cereal_preg_Ter +
##     h_dairy_preg_Ter + h_fastfood_preg_Ter + h_fish_preg_Ter +
##     h_folic_t1_None + h_fruit_preg_Ter + h_legume_preg_Ter +
##     h_meat_preg_Ter + h_pamod_t3_None
## Model 2: data$hs_correct_raven ~ e3_alcpreg_yn_None + h_cereal_preg_Ter +
##     h_dairy_preg_Ter + h_fastfood_preg_Ter + h_fish_preg_Ter +
##     h_folic_t1_None + h_fruit_preg_Ter + h_legume_preg_Ter +
##     h_meat_preg_Ter + h_pamod_t3_None + h_pavig_t3_None + h_veg_preg_Ter
##   Res.Df   RSS Df Sum of Sq      F Pr(>F)
## 1   1281 41738
## 2   1277 41465  4    273.21 2.1035 0.0782 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#cereal and dair
#fish and folic
#fruits and physical activity
```