Homework 3 Question 1

Ronaldlee Ejalu, Keller Abigail, Burns Kathleen

2/20/2021

Load the necessary packages

```
library(GGally, quietly = TRUE) #ggpairs Function

## Warning: package 'GGally' was built under R version 4.0.3

## Warning: package 'ggplot2' was built under R version 4.0.3

## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2

library(ggplot2, quietly = TRUE) #ggplot2 Functions
library(corrplot, quietly = TRUE) #Plot Correlations

## Warning: package 'corrplot' was built under R version 4.0.3

## corrplot 0.84 loaded

library(DescTools, quietly = TRUE) #VIF Function

## Warning: package 'DescTools' was built under R version 4.0.3

library(xlsx, quietly = TRUE)

## Warning: package 'xlsx' was built under R version 4.0.3

library(MASS, quietly = TRUE) # stepAIC
```

###Homework 3, Question 1 ###a Model Building, Multiple Regression by Kate Burns

set to the working directory

```
setwd("C:/Users/rejalu1/OneDrive - Henry Ford Health
System/DSC424/HomeWork3")
```

Load the data file

```
NHANES_DatasetUpdatedv5  <-
read.xlsx(file="../HomeWork3/datasets/NHANES_DatasetUpdatedv5.xlsx"
                    , sheetIndex=1)
```

#create a subset

```
nhanesData <- NHANES_DatasetUpdatedv5[,2:25]
```

#check sample size & number of variables

```
dim(nhanesData)
```

```
## [1] 872  24
```

#show first 6 rows of data

```
head(nhanesData)
```

```
##   diabetes_risk hypertension gluten_free stroke diabetes_relatives
## 1             2            2           2      2                  2
## 2             2            2           1      2                  2
## 3             2            2           2      2                  2
## 4             2            2           1      2                  2
## 5             2            2           2      2                  2
## 6             2            2           2      2                  2
##   moderate_activity diabetes feel_at_risk_diabetes frozen_meals gender age
## 1                30        2                     2            0      2  26
## 2                20        2                     2            2      1  28
## 3               120        2                     2           10      1  35
## 4               120        2                     2            0      1  29
## 5                60        2                     1            1      2  36
## 6                30        2                     2            0      2  57
##   ethnicity education_level annual_income height weight most_weighed
## 1         3               5            15     60    105          120
## 2         5               4             7     69    205          250
## 3         3               3             5     67    165          174
## 4         1               4            12     69    170          220
## 5         1               3             3     69    185          192
## 6         3               3             8     63    217          245
##   sleep_weekdays hours_worked time_outdoors_weekends time_sitting
## 1              8           40                    120          600
## 2              6           40                    120          480
## 3              7           42                    240          300
## 4              8           42                    360          720
## 5              7           21                     30          120
## 6              8           45                     60          120
##   take_away_food poverty_ratio take.away.within.30.days
## 1              3          5.00                       0
## 2              1          2.26                       0
## 3              5          1.74                      21
## 4              7          2.14                       3
## 5             10          0.84                       0
## 6              0          2.27                       0
```

#Show the meta data

```
str(nhanesData)
```

```
## 'data.frame':    872 obs. of  24 variables:
##  $ diabetes_risk           : num  2 2 2 2 2 2 2 2 1 ...
##  $ hypertension            : num  2 2 2 2 2 2 1 2 2 ...
##  $ gluten_free             : num  2 1 2 1 2 2 2 2 1 ...
##  $ stroke                  : num  2 2 2 2 2 2 2 2 2 ...
##  $ diabetes_relatives      : num  2 2 2 2 2 2 1 2 2 1 ...
##  $ moderate_activity       : num  30 20 120 120 60 30 30 20 30 120 ...
##  $ diabetes                : num  2 2 2 2 2 2 2 2 2 ...
##  $ feel_at_risk_diabetes   : num  2 2 2 2 1 2 2 1 2 1 ...
##  $ frozen_meals            : num  0 2 10 0 1 0 0 1 0 0 ...
##  $ gender                  : num  2 1 1 1 2 2 1 1 1 1 ...
##  $ age                     : num  26 28 35 29 36 57 44 35 59 58 ...
##  $ ethnicity               : num  3 5 3 1 1 3 4 5 5 3 ...
##  $ education_level         : num  5 4 3 4 3 3 3 5 5 3 ...
##  $ annual_income           : num  15 7 5 12 3 8 8 15 7 10 ...
##  $ height                  : num  60 69 67 69 69 63 76 68 68 69 ...
##  $ weight                  : num  105 205 165 170 185 217 250 160 163 195
...
##  $ most_weighed            : num  120 250 174 220 192 245 250 160 175 215
...
##  $ sleep_weekdays          : num  8 6 7 8 7 8 8 5 7 7 ...
##  $ hours_worked            : num  40 40 42 42 21 45 40 55 20 50 ...
##  $ time_outdoors_weekends  : num  120 120 240 360 30 60 120 180 120 120
...
##  $ time_sitting            : num  600 480 300 720 120 120 600 660 660 240
...
##  $ take_away_food          : num  3 1 5 7 10 0 2 6 0 7 ...
##  $ poverty_ratio           : num  5 2.26 1.74 2.14 0.84 2.27 1.79 5 1.82
4.45 ...
##  $ take.away.within.30.days: num  0 0 21 3 0 0 2 2 0 0 ...
```

#check for missing values

```
sum(is.na(nhanesData))
```

```
## [1] 0
```

create sets for exploratory analysis

```
diabetes <- nhanesData[,c(1,2,4,5,7,8)]
demo <- nhanesData[,c(10,11,12,13,14,15,16,17,23)]
dietActivity <- nhanesData[,c(3,6,9,18,19,20,21,22,24)]
```

##get sample stats around variables

```
#describe(nhanesData)
```

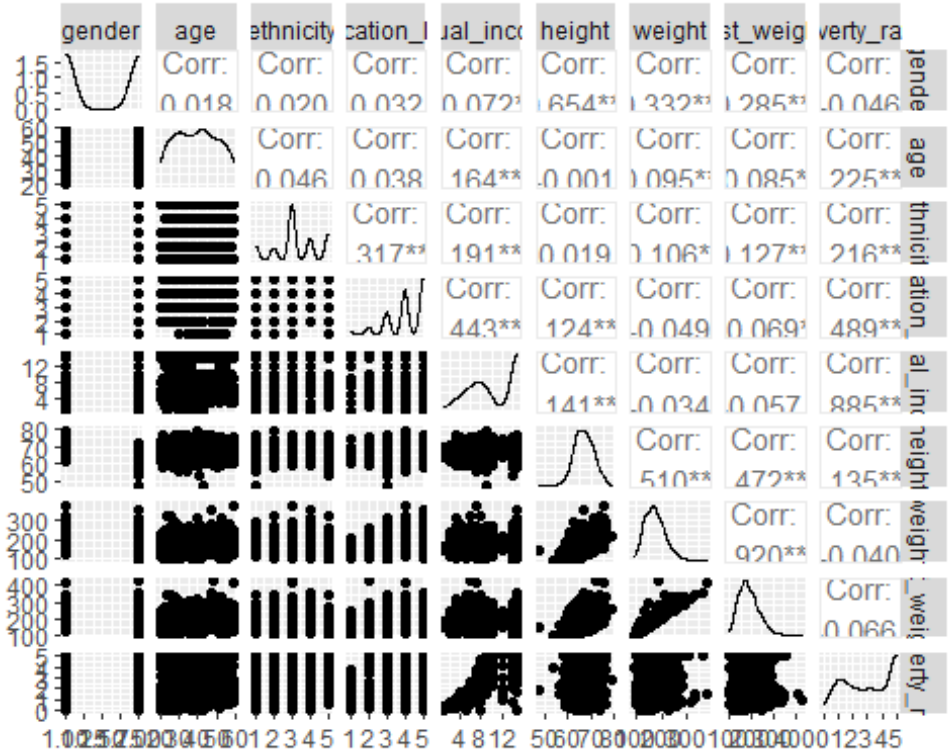
data has been filtered on respondents that answered diabetes risk question:

Have you ever been told by a doctor or other health professional

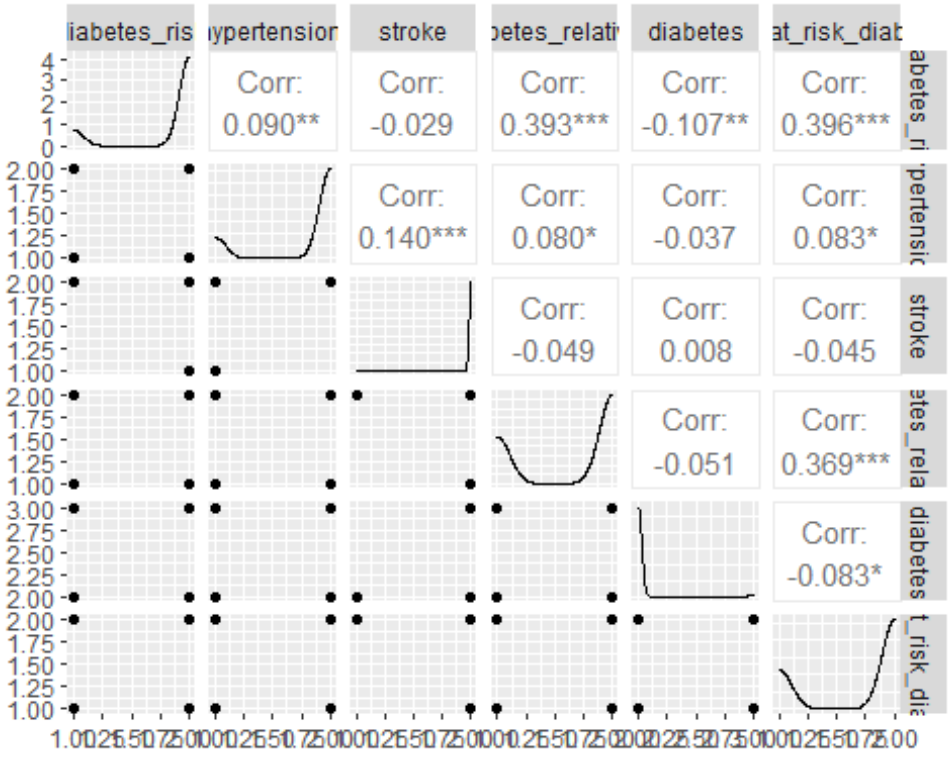
##that you have health conditions or a medical or family history that ##increases your risk for diabetes?average? ####age of respondents is around 39 years old, slightly more females than males, non-hispanic ### white, with some college or AA degree. All respondents work in some capacity. ### In regards to health, they have an average weight around 173 lbs, don't feel they are at risk ### for diabetes or hypertension, are not gluten free, may have had relative with diabetes. ### There dietary habits are around 2 frozen meals in the last 30 days, ate out around 4 times ### in the past week and eaten ready made food from the grocery store 2 times in the past 30 days.

exploratory analysis with graphing

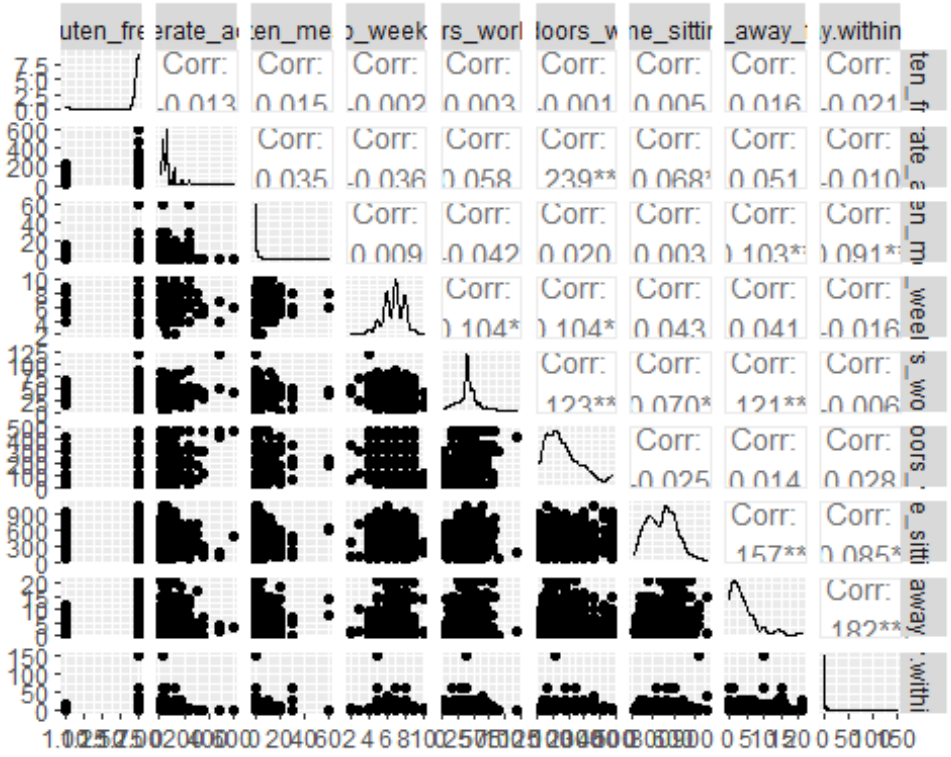
```
ggpairs(demo)
```




```
ggpairs(diabetes)
```




```
ggpairs(dietActivity)
```

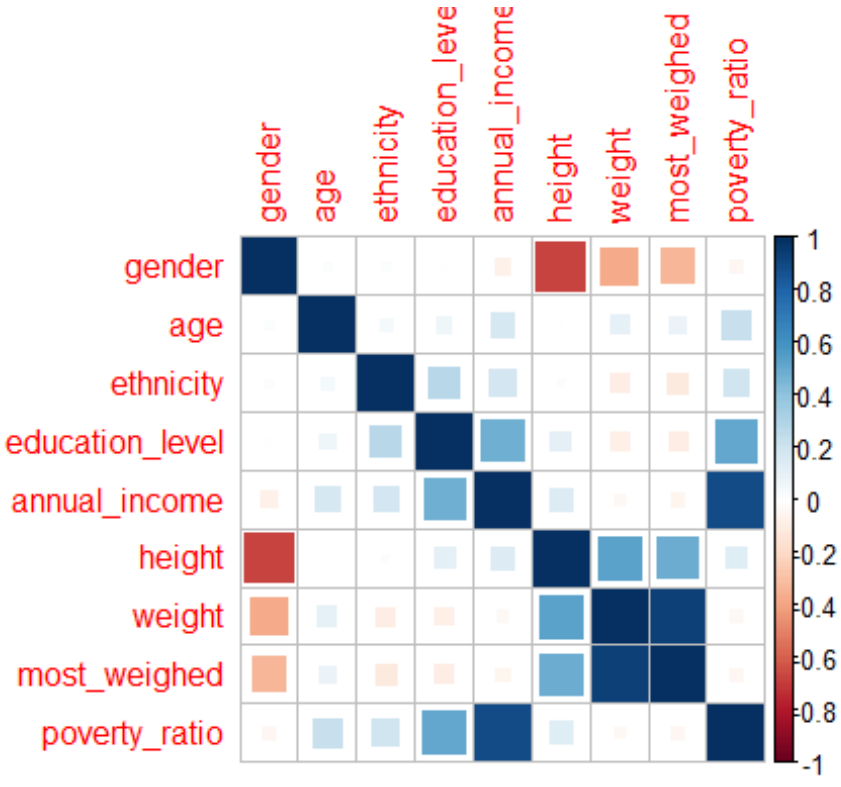


Correlation matrixes

```r
M<-cor(demo, method="spearman")
round(M,2)
```

```
##                  gender  age ethnicity education_level annual_income height
## gender             1.00 0.02      0.02            0.01         -0.07  -0.68
## age                0.02 1.00      0.04            0.07          0.17   0.00
## ethnicity          0.02 0.04      1.00            0.27          0.18   0.01
## education_level    0.01 0.07      0.27            1.00          0.48   0.12
## annual_income     -0.07 0.17      0.18            0.48          1.00   0.14
## height            -0.68 0.00      0.01            0.12          0.14   1.00
## weight            -0.38 0.10     -0.10           -0.09         -0.03   0.53
## most_weighed      -0.34 0.08     -0.12           -0.10         -0.05   0.50
## poverty_ratio     -0.04 0.23      0.19            0.52          0.89   0.13
##                  weight most_weighed poverty_ratio
## gender            -0.38        -0.34         -0.04
## age                0.10         0.08          0.23
## ethnicity         -0.10        -0.12          0.19
## education_level   -0.09        -0.10          0.52
## annual_income     -0.03        -0.05          0.89
## height             0.53         0.50          0.13
## weight             1.00         0.94         -0.03
## most_weighed       0.94         1.00         -0.05
## poverty_ratio     -0.03        -0.05          1.00
```

```r
corrplot(M, method = "square")
```



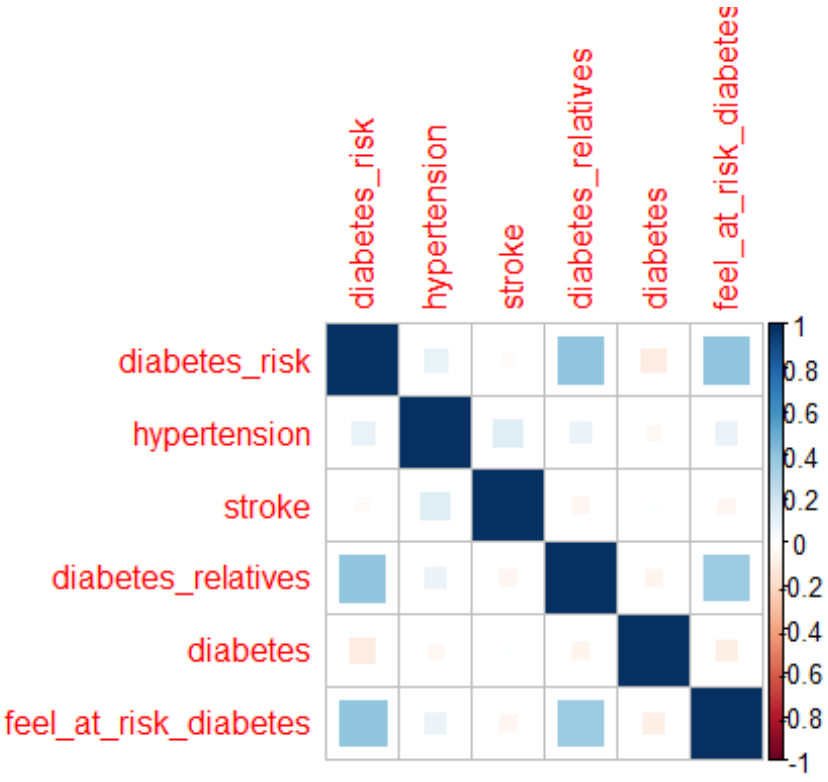
drop most_weighed (17) because highly correlated with weight ## drop annual income (14) because highly correlated with poverty level ratio

#more correlation matrixes

```
M<-cor(diabetes, method="spearman")
round(M,2)
```

```
##                       diabetes_risk hypertension stroke diabetes_relatives
## diabetes_risk                  1.00         0.09  -0.03               0.39
## hypertension                   0.09         1.00   0.14               0.08
## stroke                        -0.03         0.14   1.00              -0.05
## diabetes_relatives             0.39         0.08  -0.05               1.00
## diabetes                      -0.11        -0.04   0.01              -0.05
## feel_at_risk_diabetes          0.40         0.08  -0.05               0.37
##                       diabetes feel_at_risk_diabetes
## diabetes_risk            -0.11                 0.40
## hypertension             -0.04                 0.08
## stroke                    0.01                -0.05
## diabetes_relatives       -0.05                 0.37
## diabetes                  1.00                -0.08
## feel_at_risk_diabetes    -0.08                 1.00
```

```
corrplot(M, method = "square")
```



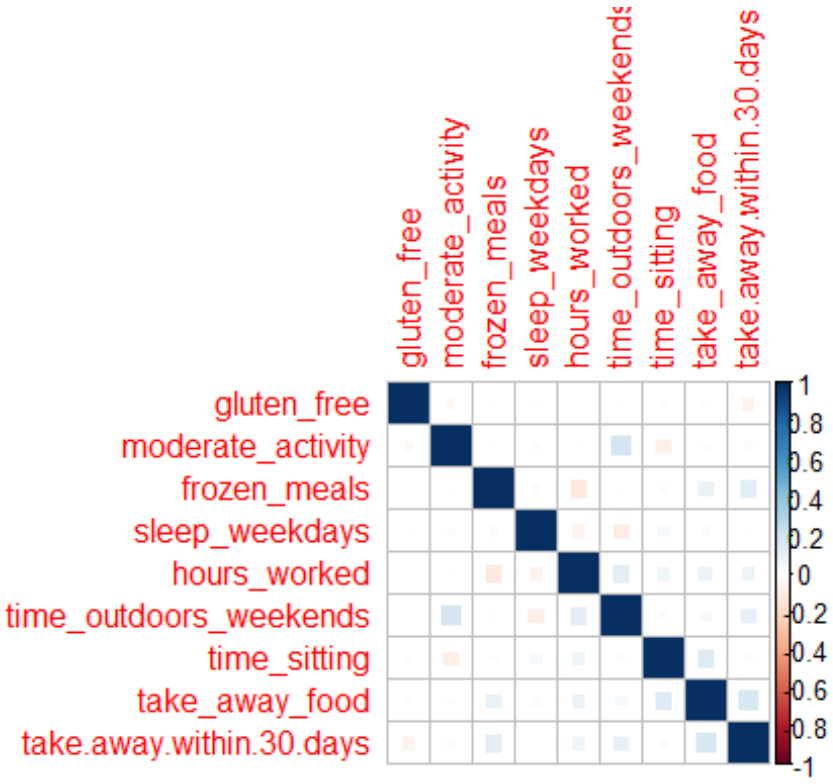
some variables are correlated but doesn't look like there is any multicollinearity

corelation matrix for diety activity

```
M<-cor(dietActivity, method="spearman")
round(M,2)
```

```
##                            gluten_free moderate_activity frozen_meals
## gluten_free                       1.00             -0.03         0.00
## moderate_activity                -0.03              1.00         0.00
## frozen_meals                      0.00              0.00         1.00
## sleep_weekdays                    0.00             -0.01        -0.02
## hours_worked                      0.00              0.01        -0.12
## time_outdoors_weekends            0.00              0.18         0.00
## time_sitting                      0.01             -0.09        -0.02
## take_away_food                    0.02              0.01         0.10
## take.away.within.30.days         -0.07              0.03         0.12
##                            sleep_weekdays hours_worked
time_outdoors_weekends
## gluten_free                          0.00         0.00
0.00
## moderate_activity                   -0.01         0.01
0.18
## frozen_meals                        -0.02        -0.12
0.00
## sleep_weekdays                       1.00        -0.07                 -
0.10
```

```
## hours_worked                         -0.07         1.00
0.12
## time_outdoors_weekends               -0.10         0.12
1.00
## time_sitting                          0.05         0.07                  -
0.02
## take_away_food                        0.03         0.08
0.04
## take.away.within.30.days              0.01         0.08
0.10
##                            time_sitting take_away_food
take.away.within.30.days
## gluten_free                           0.01           0.02                  -
0.07
## moderate_activity                    -0.09           0.01
0.03
## frozen_meals                         -0.02           0.10
0.12
## sleep_weekdays                        0.05           0.03
0.01
## hours_worked                          0.07           0.08
0.08
## time_outdoors_weekends               -0.02           0.04
0.10
## time_sitting                          1.00           0.14
0.02
## take_away_food                        0.14           1.00
0.18
## take.away.within.30.days              0.02           0.18
1.00

corrplot(M, method = "square")
```



low correlations between most variables

##

#create new subset of variables

```
nhanesData2 <- nhanesData[,c(1:13,15, 16, 18:24)]
str(nhanesData2)
```

```
## 'data.frame':    872 obs. of  22 variables:
##  $ diabetes_risk           : num  2 2 2 2 2 2 2 2 2 1 ...
##  $ hypertension            : num  2 2 2 2 2 2 1 2 2 2 ...
##  $ gluten_free             : num  2 1 2 1 2 2 2 2 2 1 ...
##  $ stroke                  : num  2 2 2 2 2 2 2 2 2 2 ...
##  $ diabetes_relatives      : num  2 2 2 2 2 2 1 2 2 1 ...
##  $ moderate_activity       : num  30 20 120 120 60 30 30 20 30 120 ...
##  $ diabetes                : num  2 2 2 2 2 2 2 2 2 2 ...
##  $ feel_at_risk_diabetes   : num  2 2 2 2 1 2 2 1 2 1 ...
##  $ frozen_meals            : num  0 2 10 0 1 0 0 1 0 0 ...
##  $ gender                  : num  2 1 1 1 2 2 1 1 1 1 ...
##  $ age                     : num  26 28 35 29 36 57 44 35 59 58 ...
##  $ ethnicity               : num  3 5 3 1 1 3 4 5 5 3 ...
##  $ education_level         : num  5 4 3 4 3 3 3 5 5 3 ...
##  $ height                  : num  60 69 67 69 69 63 76 68 68 69 ...
##  $ weight                  : num  105 205 165 170 185 217 250 160 163 195
...
##  $ sleep_weekdays          : num  8 6 7 8 7 8 8 5 7 7 ...
##  $ hours_worked            : num  40 40 42 42 21 45 40 55 20 50 ...
##  $ time_outdoors_weekends  : num  120 120 240 360 30 60 120 180 120 120
```

```
...
##  $ time_sitting            : num  600 480 300 720 120 120 600 660 660 240
...
##  $ take_away_food          : num  3 1 5 7 10 0 2 6 0 7 ...
##  $ poverty_ratio           : num  5 2.26 1.74 2.14 0.84 2.27 1.79 5 1.82
4.45 ...
##  $ take.away.within.30.days: num  0 0 21 3 0 0 2 2 0 0 ...
```

#create initial linear regression model

```
model1 <- lm(diabetes_risk ~ ., data=nhanesData2)
model1

##
## Call:
## lm(formula = diabetes_risk ~ ., data = nhanesData2)
##
## Coefficients:
##              (Intercept)              hypertension
gluten_free
##                1.253e+00                 2.518e-02                 1.078e-
01
##                   stroke         diabetes_relatives
moderate_activity
##               -4.961e-04                 2.134e-01                -1.589e-
04
##                 diabetes     feel_at_risk_diabetes
frozen_meals
##               -2.034e-01                 1.991e-01                -4.741e-
03
##                   gender                       age
ethnicity
##               -3.767e-02                 9.405e-04                 1.290e-
02
##          education_level                    height
weight
##               -7.785e-04                 3.243e-03                -8.046e-
04
##           sleep_weekdays              hours_worked
time_outdoors_weekends
##               -4.756e-06                -3.875e-04                -3.675e-
05
##             time_sitting            take_away_food
poverty_ratio
##                2.137e-05                 2.233e-03                -4.651e-
03
## take.away.within.30.days
##                1.916e-04
```

#check VIF

```
VIF(model1)
```

```
##             hypertension             gluten_free                  stroke
##                 1.183361                1.015263                1.047896
##       diabetes_relatives       moderate_activity                diabetes
##                 1.195863                1.100338                1.021360
##    feel_at_risk_diabetes            frozen_meals                  gender
##                 1.238208                1.062640                1.871254
##                      age               ethnicity         education_level
##                 1.244715                1.184650                1.497368
##                   height                  weight           sleep_weekdays
##                 2.316905                1.602446                1.071724
##             hours_worked  time_outdoors_weekends             time_sitting
##                 1.112839                1.130236                1.154435
##           take_away_food           poverty_ratio take.away.within.30.days
##                 1.164569                1.528655                1.055675
```

no multicolinearity

inspect the model

```
summary(model1)
```

```
##
## Call:
## lm(formula = diabetes_risk ~ ., data = nhanesData2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.03144 -0.03426  0.01725  0.19441  0.72311
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)               1.253e+00  4.822e-01   2.598  0.00953 **
## hypertension              2.518e-02  2.933e-02   0.859  0.39084
## gluten_free               1.078e-01  5.851e-02   1.843  0.06574 .
## stroke                   -4.961e-04  1.603e-01  -0.003  0.99753
## diabetes_relatives        2.134e-01  2.438e-02   8.752  < 2e-16 ***
## moderate_activity        -1.589e-04  1.880e-04  -0.845  0.39839
## diabetes                 -2.034e-01  8.826e-02  -2.304  0.02145 *
## feel_at_risk_diabetes     1.991e-01  2.555e-02   7.790 1.95e-14 ***
## frozen_meals             -4.741e-03  2.109e-03  -2.248  0.02486 *
## gender                   -3.767e-02  2.896e-02  -1.301  0.19372
## age                       9.405e-04  1.069e-03   0.880  0.37917
## ethnicity                 1.290e-02  9.702e-03   1.329  0.18413
## education_level          -7.785e-04  1.309e-02  -0.059  0.95261
## height                    3.243e-03  4.071e-03   0.797  0.42589
## weight                   -8.046e-04  3.318e-04  -2.425  0.01551 *
## sleep_weekdays           -4.756e-06  9.336e-03  -0.001  0.99959
## hours_worked             -3.875e-04  8.297e-04  -0.467  0.64058
```

```
## time_outdoors_weekends   -3.675e-05  9.673e-05  -0.380  0.70408
## time_sitting              2.137e-05  5.677e-05   0.377  0.70662
## take_away_food            2.233e-03  2.720e-03   0.821  0.41178
## poverty_ratio            -4.651e-03  7.942e-03  -0.586  0.55832
## take.away.within.30.days  1.916e-04  1.366e-03   0.140  0.88850
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3125 on 850 degrees of freedom
## Multiple R-squared:  0.2541, Adjusted R-squared:  0.2357
## F-statistic: 13.79 on 21 and 850 DF,  p-value: < 2.2e-16
```

diabetes_relatives, diabetes, feel_at_risk_diabetes, frozen_meals, weight

p-value is low, Multiple R-squared: 0.2541,

F - statistic P-value is < < 2.2e-16

#check automatic models

```
null = lm(diabetes_risk ~ 1, data=nhanesData2)
null

##
## Call:
## lm(formula = diabetes_risk ~ 1, data = nhanesData2)
##
## Coefficients:
## (Intercept)
##        1.85

full = lm(diabetes_risk ~ ., data=nhanesData2)
full

##
## Call:
## lm(formula = diabetes_risk ~ ., data = nhanesData2)
##
## Coefficients:
##              (Intercept)              hypertension
gluten_free
##                1.253e+00                 2.518e-02                 1.078e-
01
##                   stroke        diabetes_relatives
moderate_activity
##               -4.961e-04                 2.134e-01                -1.589e-
04
##                 diabetes     feel_at_risk_diabetes
frozen_meals
```

```
##               -2.034e-01                   1.991e-01                  -4.741e-
03
##                   gender                         age
ethnicity
##               -3.767e-02                   9.405e-04                   1.290e-
02
##          education_level                      height
weight
##               -7.785e-04                   3.243e-03                  -8.046e-
04
##           sleep_weekdays                hours_worked
time_outdoors_weekends
##               -4.756e-06                  -3.875e-04                  -3.675e-
05
##             time_sitting             take_away_food
poverty_ratio
##                2.137e-05                   2.233e-03                  -4.651e-
03
## take.away.within.30.days
##                1.916e-04
```

#Forward Regression

```
forwardNhanes = step(null, scope = list(lower=null, upper=full),
direction="forward")

## Start:  AIC=-1792.91
## diabetes_risk ~ 1
##
##                             Df Sum of Sq    RSS     AIC
## + feel_at_risk_diabetes      1   17.4704  93.85 -1939.8
## + diabetes_relatives         1   17.1698  94.15 -1937.0
## + weight                     1    2.2654 109.06 -1808.8
## + diabetes                   1    1.2789 110.04 -1801.0
## + hypertension               1    0.9105 110.41 -1798.1
## + frozen_meals               1    0.6110 110.71 -1795.7
## + gender                     1    0.5404 110.78 -1795.2
## + ethnicity                  1    0.5311 110.79 -1795.1
## + gluten_free                1    0.4212 110.90 -1794.2
## + education_level            1    0.3763 110.94 -1793.9
## <none>                                   111.32 -1792.9
## + height                     1    0.2386 111.08 -1792.8
## + take_away_food             1    0.1980 111.12 -1792.5
## + poverty_ratio              1    0.1925 111.13 -1792.4
## + sleep_weekdays             1    0.1632 111.16 -1792.2
## + moderate_activity          1    0.1128 111.21 -1791.8
## + stroke                     1    0.0907 111.23 -1791.6
## + time_sitting               1    0.0455 111.27 -1791.3
## + hours_worked               1    0.0455 111.27 -1791.3
## + time_outdoors_weekends     1    0.0296 111.29 -1791.1
```

```
## + age                         1    0.0070 111.31 -1791.0
## + take.away.within.30.days  1    0.0009 111.32 -1790.9
##
## Step:  AIC=-1939.77
## diabetes_risk ~ feel_at_risk_diabetes
##
##                             Df Sum of Sq    RSS     AIC
## + diabetes_relatives         1    7.8285 86.021 -2013.7
## + weight                     1    0.6224 93.227 -1943.6
## + diabetes                   1    0.6220 93.228 -1943.6
## + frozen_meals               1    0.4048 93.445 -1941.5
## + hypertension               1    0.3689 93.481 -1941.2
## + gluten_free                1    0.3631 93.486 -1941.2
## <none>                                   93.850 -1939.8
## + ethnicity                  1    0.2146 93.635 -1939.8
## + take_away_food             1    0.1865 93.663 -1939.5
## + education_level            1    0.1784 93.671 -1939.4
## + gender                     1    0.1755 93.674 -1939.4
## + time_sitting               1    0.0788 93.771 -1938.5
## + sleep_weekdays             1    0.0524 93.797 -1938.3
## + moderate_activity          1    0.0495 93.800 -1938.2
## + height                     1    0.0361 93.813 -1938.1
## + poverty_ratio              1    0.0326 93.817 -1938.1
## + hours_worked               1    0.0285 93.821 -1938.0
## + time_outdoors_weekends     1    0.0209 93.829 -1938.0
## + stroke                     1    0.0127 93.837 -1937.9
## + take.away.within.30.days  1    0.0048 93.845 -1937.8
## + age                        1    0.0035 93.846 -1937.8
##
## Step:  AIC=-2013.72
## diabetes_risk ~ feel_at_risk_diabetes + diabetes_relatives
##
##                             Df Sum of Sq    RSS     AIC
## + frozen_meals               1   0.63372 85.387 -2018.2
## + diabetes                   1   0.52965 85.491 -2017.1
## + weight                     1   0.43989 85.581 -2016.2
## + gluten_free                1   0.34398 85.677 -2015.2
## + ethnicity                  1   0.28653 85.735 -2014.6
## + hypertension               1   0.20975 85.811 -2013.8
## <none>                                   86.021 -2013.7
## + gender                     1   0.19074 85.830 -2013.7
## + moderate_activity          1   0.07931 85.942 -2012.5
## + take_away_food             1   0.05111 85.970 -2012.2
## + time_outdoors_weekends     1   0.04561 85.975 -2012.2
## + time_sitting               1   0.04179 85.979 -2012.2
## + height                     1   0.02936 85.992 -2012.0
## + age                        1   0.02569 85.995 -2012.0
## + education_level            1   0.02337 85.998 -2012.0
## + sleep_weekdays             1   0.01775 86.003 -2011.9
## + hours_worked               1   0.00908 86.012 -2011.8
```

```
## + poverty_ratio             1   0.00320 86.018 -2011.8
## + take.away.within.30.days  1   0.00053 86.021 -2011.7
## + stroke                    1   0.00023 86.021 -2011.7
##
## Step:  AIC=-2018.17
## diabetes_risk ~ feel_at_risk_diabetes + diabetes_relatives +
##     frozen_meals
##
##                            Df Sum of Sq    RSS     AIC
## + diabetes                  1   0.53945 84.848 -2021.7
## + weight                    1   0.39630 84.991 -2020.2
## + gluten_free               1   0.35851 85.029 -2019.8
## + ethnicity                 1   0.25700 85.130 -2018.8
## <none>                                  85.387 -2018.2
## + gender                    1   0.18349 85.204 -2018.0
## + hypertension              1   0.16234 85.225 -2017.8
## + take_away_food            1   0.09390 85.293 -2017.1
## + moderate_activity         1   0.06518 85.322 -2016.8
## + time_sitting              1   0.04219 85.345 -2016.6
## + time_outdoors_weekends    1   0.03970 85.348 -2016.6
## + height                    1   0.03530 85.352 -2016.5
## + education_level           1   0.02372 85.364 -2016.4
## + sleep_weekdays            1   0.01957 85.368 -2016.4
## + hours_worked              1   0.01635 85.371 -2016.3
## + take.away.within.30.days  1   0.00964 85.378 -2016.3
## + age                       1   0.00154 85.386 -2016.2
## + poverty_ratio             1   0.00097 85.386 -2016.2
## + stroke                    1   0.00000 85.387 -2016.2
##
## Step:  AIC=-2021.7
## diabetes_risk ~ feel_at_risk_diabetes + diabetes_relatives +
##     frozen_meals + diabetes
##
##                            Df Sum of Sq    RSS     AIC
## + gluten_free               1   0.38060 84.467 -2023.6
## + weight                    1   0.34886 84.499 -2023.3
## + ethnicity                 1   0.27618 84.572 -2022.5
## <none>                                  84.848 -2021.7
## + gender                    1   0.19183 84.656 -2021.7
## + hypertension              1   0.14540 84.702 -2021.2
## + take_away_food            1   0.07811 84.770 -2020.5
## + moderate_activity         1   0.06323 84.785 -2020.3
## + height                    1   0.04345 84.804 -2020.2
## + time_sitting              1   0.04214 84.806 -2020.1
## + time_outdoors_weekends    1   0.04100 84.807 -2020.1
## + education_level           1   0.02083 84.827 -2019.9
## + sleep_weekdays            1   0.01748 84.830 -2019.9
## + hours_worked              1   0.01626 84.832 -2019.9
## + take.away.within.30.days  1   0.01188 84.836 -2019.8
## + age                       1   0.01004 84.838 -2019.8
```

```
## + poverty_ratio              1   0.00321 84.845 -2019.7
## + stroke                     1   0.00002 84.848 -2019.7
##
## Step:  AIC=-2023.62
## diabetes_risk ~ feel_at_risk_diabetes + diabetes_relatives +
##     frozen_meals + diabetes + gluten_free
##
##                             Df Sum of Sq    RSS     AIC
## + weight                     1   0.38095 84.086 -2025.6
## + ethnicity                  1   0.28036 84.187 -2024.5
## <none>                                   84.467 -2023.6
## + gender                     1   0.16449 84.303 -2023.3
## + hypertension               1   0.15061 84.317 -2023.2
## + take_away_food             1   0.07309 84.394 -2022.4
## + moderate_activity          1   0.05915 84.408 -2022.2
## + time_sitting               1   0.04080 84.426 -2022.0
## + time_outdoors_weekends     1   0.04072 84.427 -2022.0
## + height                     1   0.03119 84.436 -2021.9
## + education_level            1   0.02943 84.438 -2021.9
## + sleep_weekdays             1   0.01785 84.449 -2021.8
## + hours_worked               1   0.01685 84.450 -2021.8
## + take.away.within.30.days   1   0.01516 84.452 -2021.8
## + poverty_ratio              1   0.00959 84.458 -2021.7
## + age                        1   0.00622 84.461 -2021.7
## + stroke                     1   0.00015 84.467 -2021.6
##
## Step:  AIC=-2025.56
## diabetes_risk ~ feel_at_risk_diabetes + diabetes_relatives +
##     frozen_meals + diabetes + gluten_free + weight
##
##                             Df Sum of Sq    RSS     AIC
## + gender                     1   0.44188 83.644 -2028.2
## + height                     1   0.35383 83.733 -2027.2
## + ethnicity                  1   0.22219 83.864 -2025.9
## <none>                                   84.086 -2025.6
## + take_away_food             1   0.11596 83.970 -2024.8
## + hypertension               1   0.07524 84.011 -2024.3
## + time_sitting               1   0.05540 84.031 -2024.1
## + moderate_activity          1   0.04751 84.039 -2024.0
## + education_level            1   0.02248 84.064 -2023.8
## + time_outdoors_weekends     1   0.02242 84.064 -2023.8
## + age                        1   0.01848 84.068 -2023.8
## + take.away.within.30.days   1   0.01547 84.071 -2023.7
## + poverty_ratio              1   0.00675 84.080 -2023.6
## + sleep_weekdays             1   0.00273 84.084 -2023.6
## + hours_worked               1   0.00234 84.084 -2023.6
## + stroke                     1   0.00034 84.086 -2023.6
##
## Step:  AIC=-2028.15
## diabetes_risk ~ feel_at_risk_diabetes + diabetes_relatives +
```

```
##     frozen_meals + diabetes + gluten_free + weight + gender
##
##                             Df Sum of Sq    RSS     AIC
## + ethnicity                  1  0.216359 83.428 -2028.4
## <none>                                   83.644 -2028.2
## + moderate_activity          1  0.095799 83.549 -2027.2
## + time_outdoors_weekends     1  0.067284 83.577 -2026.9
## + height                     1  0.064020 83.580 -2026.8
## + hypertension               1  0.059477 83.585 -2026.8
## + take_away_food             1  0.053532 83.591 -2026.7
## + time_sitting               1  0.044279 83.600 -2026.6
## + age                        1  0.030090 83.614 -2026.5
## + education_level            1  0.026977 83.617 -2026.4
## + hours_worked               1  0.025638 83.619 -2026.4
## + take.away.within.30.days   1  0.008960 83.635 -2026.2
## + poverty_ratio              1  0.001943 83.643 -2026.2
## + sleep_weekdays             1  0.000390 83.644 -2026.2
## + stroke                     1  0.000286 83.644 -2026.2
##
## Step:  AIC=-2028.41
## diabetes_risk ~ feel_at_risk_diabetes + diabetes_relatives +
##     frozen_meals + diabetes + gluten_free + weight + gender +
##     ethnicity
##
##                             Df Sum of Sq    RSS     AIC
## <none>                                   83.428 -2028.4
## + moderate_activity          1  0.092675 83.335 -2027.4
## + hypertension               1  0.066264 83.362 -2027.1
## + take_away_food             1  0.050786 83.377 -2026.9
## + time_outdoors_weekends     1  0.049290 83.379 -2026.9
## + height                     1  0.044598 83.384 -2026.9
## + hours_worked               1  0.031115 83.397 -2026.7
## + age                        1  0.022522 83.406 -2026.7
## + time_sitting               1  0.012539 83.416 -2026.5
## + take.away.within.30.days   1  0.006301 83.422 -2026.5
## + poverty_ratio              1  0.003047 83.425 -2026.5
## + sleep_weekdays             1  0.002717 83.425 -2026.4
## + stroke                     1  0.000468 83.428 -2026.4
## + education_level            1  0.000290 83.428 -2026.4
```

```
summary(forwardNhanes)
```

```
##
## Call:
## lm(formula = diabetes_risk ~ feel_at_risk_diabetes + diabetes_relatives +
##     frozen_meals + diabetes + gluten_free + weight + gender +
##     ethnicity, data = nhanesData2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -1.06150 -0.03052  0.01255  0.19793  0.70306
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             1.4966706  0.2284771   6.551 9.84e-11 ***
## feel_at_risk_diabetes   0.2016373  0.0251690   8.011 3.67e-15 ***
## diabetes_relatives      0.2149918  0.0239499   8.977  < 2e-16 ***
## frozen_meals           -0.0049337  0.0020442  -2.413   0.0160 *
## diabetes               -0.2036479  0.0873981  -2.330   0.0200 *
## gluten_free             0.1146462  0.0578983   1.980   0.0480 *
## weight                 -0.0007007  0.0002853  -2.456   0.0143 *
## gender                 -0.0479950  0.0225984  -2.124   0.0340 *
## ethnicity               0.0133756  0.0089408   1.496   0.1350
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3109 on 863 degrees of freedom
## Multiple R-squared:  0.2506, Adjusted R-squared:  0.2436
## F-statistic: 36.06 on 8 and 863 DF,  p-value: < 2.2e-16
```

feel_at_risk_diabetes, diabetes_relatives, frozen_meals, diabetes, gluten_free, weight, gender, ethnicity

##pvalue low, Multiple R-squared: 0.2506,

#Backward Regression

```
backwardNhanes = step(full, direction="backward")
```

```
## Start:  AIC=-2006.6
## diabetes_risk ~ hypertension + gluten_free + stroke + diabetes_relatives +
##     moderate_activity + diabetes + feel_at_risk_diabetes + frozen_meals +
##     gender + age + ethnicity + education_level + height + weight +
##     sleep_weekdays + hours_worked + time_outdoors_weekends +
##     time_sitting + take_away_food + poverty_ratio +
take.away.within.30.days
##
##                            Df Sum of Sq    RSS     AIC
## - sleep_weekdays            1    0.0000 83.029 -2008.6
## - stroke                    1    0.0000 83.029 -2008.6
## - education_level           1    0.0003 83.029 -2008.6
## - take.away.within.30.days  1    0.0019 83.031 -2008.6
## - time_sitting              1    0.0138 83.042 -2008.5
## - time_outdoors_weekends    1    0.0141 83.043 -2008.5
## - hours_worked              1    0.0213 83.050 -2008.4
## - poverty_ratio             1    0.0335 83.062 -2008.2
## - height                    1    0.0620 83.091 -2008.0
## - take_away_food            1    0.0659 83.094 -2007.9
## - moderate_activity         1    0.0697 83.098 -2007.9
## - hypertension              1    0.0720 83.101 -2007.8
```

```
## - age                         1    0.0756 83.104 -2007.8
## - gender                      1    0.1653 83.194 -2006.9
## - ethnicity                   1    0.1726 83.201 -2006.8
## <none>                                    83.029 -2006.6
## - gluten_free                 1    0.3316 83.360 -2005.1
## - frozen_meals                1    0.4935 83.522 -2003.4
## - diabetes                    1    0.5186 83.547 -2003.2
## - weight                      1    0.5745 83.603 -2002.6
## - feel_at_risk_diabetes       1    5.9279 88.957 -1948.5
## - diabetes_relatives          1    7.4824 90.511 -1933.4
##
## Step:  AIC=-2008.6
## diabetes_risk ~ hypertension + gluten_free + stroke + diabetes_relatives +
##     moderate_activity + diabetes + feel_at_risk_diabetes + frozen_meals +
##     gender + age + ethnicity + education_level + height + weight +
##     hours_worked + time_outdoors_weekends + time_sitting + take_away_food
+
##     poverty_ratio + take.away.within.30.days
##
##                              Df Sum of Sq    RSS     AIC
## - stroke                      1    0.0000 83.029 -2010.6
## - education_level             1    0.0003 83.029 -2010.6
## - take.away.within.30.days    1    0.0019 83.031 -2010.6
## - time_sitting                1    0.0139 83.043 -2010.5
## - time_outdoors_weekends      1    0.0142 83.043 -2010.5
## - hours_worked                1    0.0215 83.050 -2010.4
## - poverty_ratio               1    0.0336 83.062 -2010.2
## - height                      1    0.0620 83.091 -2010.0
## - take_away_food              1    0.0659 83.095 -2009.9
## - moderate_activity           1    0.0697 83.098 -2009.9
## - hypertension                1    0.0724 83.101 -2009.8
## - age                         1    0.0758 83.104 -2009.8
## - gender                      1    0.1659 83.195 -2008.9
## - ethnicity                   1    0.1742 83.203 -2008.8
## <none>                                    83.029 -2008.6
## - gluten_free                 1    0.3316 83.360 -2007.1
## - frozen_meals                1    0.4935 83.522 -2005.4
## - diabetes                    1    0.5186 83.547 -2005.2
## - weight                      1    0.5803 83.609 -2004.5
## - feel_at_risk_diabetes       1    5.9280 88.957 -1950.5
## - diabetes_relatives          1    7.4829 90.512 -1935.3
##
## Step:  AIC=-2010.6
## diabetes_risk ~ hypertension + gluten_free + diabetes_relatives +
##     moderate_activity + diabetes + feel_at_risk_diabetes + frozen_meals +
##     gender + age + ethnicity + education_level + height + weight +
##     hours_worked + time_outdoors_weekends + time_sitting + take_away_food
+
##     poverty_ratio + take.away.within.30.days
##
```

```
##                             Df Sum of Sq    RSS     AIC
## - education_level            1    0.0003 83.029 -2012.6
## - take.away.within.30.days   1    0.0019 83.031 -2012.6
## - time_sitting               1    0.0139 83.043 -2012.5
## - time_outdoors_weekends     1    0.0142 83.043 -2012.5
## - hours_worked               1    0.0215 83.050 -2012.4
## - poverty_ratio              1    0.0337 83.062 -2012.2
## - height                     1    0.0620 83.091 -2012.0
## - take_away_food             1    0.0660 83.095 -2011.9
## - moderate_activity          1    0.0698 83.098 -2011.9
## - hypertension               1    0.0736 83.102 -2011.8
## - age                        1    0.0762 83.105 -2011.8
## - gender                     1    0.1659 83.195 -2010.9
## - ethnicity                  1    0.1743 83.203 -2010.8
## <none>                                   83.029 -2010.6
## - gluten_free                1    0.3316 83.360 -2009.1
## - frozen_meals               1    0.4938 83.522 -2007.4
## - diabetes                   1    0.5187 83.547 -2007.2
## - weight                     1    0.5814 83.610 -2006.5
## - feel_at_risk_diabetes      1    5.9330 88.962 -1952.4
## - diabetes_relatives         1    7.5015 90.530 -1937.2
##
## Step:  AIC=-2012.6
## diabetes_risk ~ hypertension + gluten_free + diabetes_relatives +
##     moderate_activity + diabetes + feel_at_risk_diabetes + frozen_meals +
##     gender + age + ethnicity + height + weight + hours_worked +
##     time_outdoors_weekends + time_sitting + take_away_food +
##     poverty_ratio + take.away.within.30.days
##
##                             Df Sum of Sq    RSS     AIC
## - take.away.within.30.days   1    0.0020 83.031 -2014.6
## - time_sitting               1    0.0136 83.043 -2014.5
## - time_outdoors_weekends     1    0.0143 83.043 -2014.5
## - hours_worked               1    0.0212 83.050 -2014.4
## - poverty_ratio              1    0.0438 83.073 -2014.1
## - height                     1    0.0621 83.091 -2013.9
## - take_away_food             1    0.0657 83.095 -2013.9
## - moderate_activity          1    0.0695 83.098 -2013.9
## - hypertension               1    0.0737 83.103 -2013.8
## - age                        1    0.0771 83.106 -2013.8
## - gender                     1    0.1711 83.200 -2012.8
## - ethnicity                  1    0.1802 83.209 -2012.7
## <none>                                   83.029 -2012.6
## - gluten_free                1    0.3322 83.361 -2011.1
## - frozen_meals               1    0.4941 83.523 -2009.4
## - diabetes                   1    0.5184 83.547 -2009.2
## - weight                     1    0.5815 83.610 -2008.5
## - feel_at_risk_diabetes      1    5.9448 88.974 -1954.3
## - diabetes_relatives         1    7.5463 90.575 -1938.7
##
```

```
## Step:  AIC=-2014.57
## diabetes_risk ~ hypertension + gluten_free + diabetes_relatives +
##     moderate_activity + diabetes + feel_at_risk_diabetes + frozen_meals +
##     gender + age + ethnicity + height + weight + hours_worked +
##     time_outdoors_weekends + time_sitting + take_away_food +
##     poverty_ratio
##
##                          Df Sum of Sq    RSS     AIC
## - time_outdoors_weekends  1    0.0139 83.045 -2016.4
## - time_sitting            1    0.0141 83.045 -2016.4
## - hours_worked            1    0.0217 83.053 -2016.3
## - poverty_ratio           1    0.0435 83.074 -2016.1
## - height                  1    0.0616 83.093 -2015.9
## - moderate_activity       1    0.0701 83.101 -2015.8
## - take_away_food          1    0.0714 83.102 -2015.8
## - hypertension            1    0.0734 83.104 -2015.8
## - age                     1    0.0766 83.108 -2015.8
## - gender                  1    0.1722 83.203 -2014.8
## - ethnicity               1    0.1811 83.212 -2014.7
## <none>                                 83.031 -2014.6
## - gluten_free             1    0.3311 83.362 -2013.1
## - frozen_meals            1    0.4922 83.523 -2011.4
## - diabetes                1    0.5172 83.548 -2011.2
## - weight                  1    0.5820 83.613 -2010.5
## - feel_at_risk_diabetes   1    5.9643 88.995 -1956.1
## - diabetes_relatives      1    7.5545 90.585 -1940.6
##
## Step:  AIC=-2016.43
## diabetes_risk ~ hypertension + gluten_free + diabetes_relatives +
##     moderate_activity + diabetes + feel_at_risk_diabetes + frozen_meals +
##     gender + age + ethnicity + height + weight + hours_worked +
##     time_sitting + take_away_food + poverty_ratio
##
##                         Df Sum of Sq    RSS     AIC
## - time_sitting           1    0.0141 83.059 -2018.3
## - hours_worked           1    0.0253 83.070 -2018.2
## - poverty_ratio          1    0.0419 83.087 -2018.0
## - height                 1    0.0600 83.105 -2017.8
## - take_away_food         1    0.0745 83.119 -2017.7
## - hypertension           1    0.0769 83.122 -2017.6
## - age                    1    0.0830 83.128 -2017.6
## - moderate_activity      1    0.0872 83.132 -2017.5
## - gender                 1    0.1650 83.210 -2016.7
## - ethnicity              1    0.1896 83.234 -2016.4
## <none>                               83.045 -2016.4
## - gluten_free            1    0.3321 83.377 -2015.0
## - frozen_meals           1    0.4919 83.537 -2013.3
## - diabetes               1    0.5171 83.562 -2013.0
## - weight                 1    0.5818 83.627 -2012.3
## - feel_at_risk_diabetes  1    5.9727 89.018 -1957.9
```

```
## - diabetes_relatives     1    7.5425 90.587 -1942.6
##
## Step:  AIC=-2018.28
## diabetes_risk ~ hypertension + gluten_free + diabetes_relatives +
##     moderate_activity + diabetes + feel_at_risk_diabetes + frozen_meals +
##     gender + age + ethnicity + height + weight + hours_worked +
##     take_away_food + poverty_ratio
##
##                          Df Sum of Sq    RSS     AIC
## - hours_worked            1    0.0247 83.084 -2020.0
## - poverty_ratio           1    0.0345 83.093 -2019.9
## - height                  1    0.0572 83.116 -2019.7
## - hypertension            1    0.0739 83.133 -2019.5
## - age                     1    0.0785 83.137 -2019.5
## - take_away_food          1    0.0833 83.142 -2019.4
## - moderate_activity       1    0.0923 83.151 -2019.3
## - gender                  1    0.1690 83.228 -2018.5
## <none>                                 83.059 -2018.3
## - ethnicity               1    0.2163 83.275 -2018.0
## - gluten_free             1    0.3346 83.394 -2016.8
## - frozen_meals            1    0.4940 83.553 -2015.1
## - diabetes                1    0.5181 83.577 -2014.9
## - weight                  1    0.5719 83.631 -2014.3
## - feel_at_risk_diabetes   1    5.9602 89.019 -1959.8
## - diabetes_relatives      1    7.5610 90.620 -1944.3
##
## Step:  AIC=-2020.02
## diabetes_risk ~ hypertension + gluten_free + diabetes_relatives +
##     moderate_activity + diabetes + feel_at_risk_diabetes + frozen_meals +
##     gender + age + ethnicity + height + weight + take_away_food +
##     poverty_ratio
##
##                          Df Sum of Sq    RSS     AIC
## - poverty_ratio           1    0.0434 83.127 -2021.6
## - height                  1    0.0552 83.139 -2021.4
## - take_away_food          1    0.0779 83.162 -2021.2
## - age                     1    0.0801 83.164 -2021.2
## - hypertension            1    0.0811 83.165 -2021.2
## - moderate_activity       1    0.0962 83.180 -2021.0
## - gender                  1    0.1577 83.241 -2020.4
## <none>                                 83.084 -2020.0
## - ethnicity               1    0.2162 83.300 -2019.8
## - gluten_free             1    0.3344 83.418 -2018.5
## - frozen_meals            1    0.4834 83.567 -2017.0
## - diabetes                1    0.5162 83.600 -2016.6
## - weight                  1    0.5823 83.666 -2015.9
## - feel_at_risk_diabetes   1    5.9665 89.050 -1961.5
## - diabetes_relatives      1    7.5936 90.677 -1945.8
##
## Step:  AIC=-2021.57
```

```
## diabetes_risk ~ hypertension + gluten_free + diabetes_relatives +
##     moderate_activity + diabetes + feel_at_risk_diabetes + frozen_meals +
##     gender + age + ethnicity + height + weight + take_away_food
##
##                          Df Sum of Sq    RSS     AIC
## - height                  1    0.0410 83.168 -2023.1
## - age                     1    0.0568 83.184 -2023.0
## - take_away_food          1    0.0618 83.189 -2022.9
## - hypertension            1    0.0789 83.206 -2022.7
## - moderate_activity       1    0.0848 83.212 -2022.7
## - gender                  1    0.1671 83.294 -2021.8
## - ethnicity               1    0.1876 83.315 -2021.6
## <none>                              83.127 -2021.6
## - gluten_free             1    0.3583 83.485 -2019.8
## - frozen_meals            1    0.4847 83.612 -2018.5
## - diabetes                1    0.5231 83.650 -2018.1
## - weight                  1    0.5519 83.679 -2017.8
## - feel_at_risk_diabetes   1    5.9618 89.089 -1963.2
## - diabetes_relatives      1    7.5582 90.685 -1947.7
##
## Step:  AIC=-2023.14
## diabetes_risk ~ hypertension + gluten_free + diabetes_relatives +
##     moderate_activity + diabetes + feel_at_risk_diabetes + frozen_meals +
##     gender + age + ethnicity + weight + take_away_food
##
##                          Df Sum of Sq    RSS     AIC
## - age                     1    0.0558 83.224 -2024.5
## - take_away_food          1    0.0681 83.236 -2024.4
## - moderate_activity       1    0.0784 83.246 -2024.3
## - hypertension            1    0.0860 83.254 -2024.2
## <none>                              83.168 -2023.1
## - ethnicity               1    0.2063 83.374 -2023.0
## - gluten_free             1    0.3620 83.530 -2021.3
## - gender                  1    0.4059 83.574 -2020.9
## - frozen_meals            1    0.4810 83.649 -2020.1
## - diabetes                1    0.5218 83.690 -2019.7
## - weight                  1    0.5284 83.696 -2019.6
## - feel_at_risk_diabetes   1    6.1409 89.309 -1963.0
## - diabetes_relatives      1    7.5992 90.767 -1948.9
##
## Step:  AIC=-2024.55
## diabetes_risk ~ hypertension + gluten_free + diabetes_relatives +
##     moderate_activity + diabetes + feel_at_risk_diabetes + frozen_meals +
##     gender + ethnicity + weight + take_away_food
##
##                          Df Sum of Sq    RSS     AIC
## - take_away_food          1    0.0555 83.279 -2026.0
## - hypertension            1    0.0578 83.282 -2025.9
## - moderate_activity       1    0.0852 83.309 -2025.7
## <none>                              83.224 -2024.5
```

```
## - ethnicity              1    0.2168 83.441 -2024.3
## - gluten_free            1    0.3719 83.596 -2022.7
## - gender                 1    0.4001 83.624 -2022.4
## - diabetes               1    0.5009 83.725 -2021.3
## - weight                 1    0.5086 83.733 -2021.2
## - frozen_meals           1    0.5523 83.776 -2020.8
## - feel_at_risk_diabetes  1    6.1366 89.360 -1964.5
## - diabetes_relatives     1    7.6006 90.825 -1950.3
##
## Step:  AIC=-2025.97
## diabetes_risk ~ hypertension + gluten_free + diabetes_relatives +
##     moderate_activity + diabetes + feel_at_risk_diabetes + frozen_meals +
##     gender + ethnicity + weight
##
##                         Df Sum of Sq    RSS     AIC
## - hypertension           1    0.0561 83.335 -2027.4
## - moderate_activity      1    0.0825 83.362 -2027.1
## <none>                              83.279 -2026.0
## - ethnicity              1    0.2196 83.499 -2025.7
## - gluten_free            1    0.3729 83.652 -2024.1
## - gender                 1    0.4636 83.743 -2023.1
## - weight                 1    0.4954 83.775 -2022.8
## - diabetes               1    0.5165 83.796 -2022.6
## - frozen_meals           1    0.5233 83.803 -2022.5
## - feel_at_risk_diabetes  1    6.1090 89.388 -1966.2
## - diabetes_relatives     1    7.7365 91.016 -1950.5
##
## Step:  AIC=-2027.38
## diabetes_risk ~ gluten_free + diabetes_relatives + moderate_activity +
##     diabetes + feel_at_risk_diabetes + frozen_meals + gender +
##     ethnicity + weight
##
##                         Df Sum of Sq    RSS     AIC
## - moderate_activity      1    0.0927 83.428 -2028.4
## <none>                              83.335 -2027.4
## - ethnicity              1    0.2132 83.549 -2027.2
## - gluten_free            1    0.3711 83.707 -2025.5
## - gender                 1    0.4832 83.819 -2024.3
## - diabetes               1    0.5228 83.858 -2023.9
## - frozen_meals           1    0.5463 83.882 -2023.7
## - weight                 1    0.5874 83.923 -2023.3
## - feel_at_risk_diabetes  1    6.1317 89.467 -1967.5
## - diabetes_relatives     1    7.8204 91.156 -1951.2
##
## Step:  AIC=-2028.41
## diabetes_risk ~ gluten_free + diabetes_relatives + diabetes +
##     feel_at_risk_diabetes + frozen_meals + gender + ethnicity +
##     weight
##
##                         Df Sum of Sq    RSS     AIC
```

```
## <none>                                83.428 -2028.4
## - ethnicity              1    0.2164 83.644 -2028.2
## - gluten_free            1    0.3790 83.807 -2026.5
## - gender                 1    0.4361 83.864 -2025.9
## - diabetes               1    0.5249 83.953 -2024.9
## - frozen_meals           1    0.5631 83.991 -2024.5
## - weight                 1    0.5830 84.011 -2024.3
## - feel_at_risk_diabetes  1    6.2046 89.633 -1967.9
## - diabetes_relatives     1    7.7900 91.218 -1952.6
```

```
summary(backwardNhanes)
```

```
##
## Call:
## lm(formula = diabetes_risk ~ gluten_free + diabetes_relatives +
##     diabetes + feel_at_risk_diabetes + frozen_meals + gender +
##     ethnicity + weight, data = nhanesData2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.06150 -0.03052  0.01255  0.19793  0.70306
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             1.4966706  0.2284771   6.551 9.84e-11 ***
## gluten_free             0.1146462  0.0578983   1.980   0.0480 *
## diabetes_relatives      0.2149918  0.0239499   8.977  < 2e-16 ***
## diabetes               -0.2036479  0.0873981  -2.330   0.0200 *
## feel_at_risk_diabetes   0.2016373  0.0251690   8.011 3.67e-15 ***
## frozen_meals           -0.0049337  0.0020442  -2.413   0.0160 *
## gender                 -0.0479950  0.0225984  -2.124   0.0340 *
## ethnicity               0.0133756  0.0089408   1.496   0.1350
## weight                 -0.0007007  0.0002853  -2.456   0.0143 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3109 on 863 degrees of freedom
## Multiple R-squared:  0.2506, Adjusted R-squared:  0.2436
## F-statistic: 36.06 on 8 and 863 DF,  p-value: < 2.2e-16
```

##gluten_free, diabetes_relatives, diabetes, feel_at_risk_diabetes, frozen_meals, gender, ethnicity, weight ## same variables as forward regression just in a different order ##pvalue low, Multiple R-squared: 0.2506

#Stepwise Regression

```
stepNhanes = step(null, scope = list(upper=full), direction="both")
```

```
## Start:  AIC=-1792.91
## diabetes_risk ~ 1
##
```

```
##                              Df Sum of Sq    RSS     AIC
## + feel_at_risk_diabetes       1   17.4704  93.85 -1939.8
## + diabetes_relatives          1   17.1698  94.15 -1937.0
## + weight                      1    2.2654 109.06 -1808.8
## + diabetes                    1    1.2789 110.04 -1801.0
## + hypertension                1    0.9105 110.41 -1798.1
## + frozen_meals                1    0.6110 110.71 -1795.7
## + gender                      1    0.5404 110.78 -1795.2
## + ethnicity                   1    0.5311 110.79 -1795.1
## + gluten_free                 1    0.4212 110.90 -1794.2
## + education_level             1    0.3763 110.94 -1793.9
## <none>                                    111.32 -1792.9
## + height                      1    0.2386 111.08 -1792.8
## + take_away_food              1    0.1980 111.12 -1792.5
## + poverty_ratio               1    0.1925 111.13 -1792.4
## + sleep_weekdays              1    0.1632 111.16 -1792.2
## + moderate_activity           1    0.1128 111.21 -1791.8
## + stroke                      1    0.0907 111.23 -1791.6
## + time_sitting                1    0.0455 111.27 -1791.3
## + hours_worked                1    0.0455 111.27 -1791.3
## + time_outdoors_weekends      1    0.0296 111.29 -1791.1
## + age                         1    0.0070 111.31 -1791.0
## + take.away.within.30.days    1    0.0009 111.32 -1790.9
##
## Step:  AIC=-1939.77
## diabetes_risk ~ feel_at_risk_diabetes
##
##                              Df Sum of Sq     RSS     AIC
## + diabetes_relatives          1    7.8285  86.021 -2013.7
## + weight                      1    0.6224  93.227 -1943.6
## + diabetes                    1    0.6220  93.228 -1943.6
## + frozen_meals                1    0.4048  93.445 -1941.5
## + hypertension                1    0.3689  93.481 -1941.2
## + gluten_free                 1    0.3631  93.486 -1941.2
## <none>                                    93.850 -1939.8
## + ethnicity                   1    0.2146  93.635 -1939.8
## + take_away_food              1    0.1865  93.663 -1939.5
## + education_level             1    0.1784  93.671 -1939.4
## + gender                      1    0.1755  93.674 -1939.4
## + time_sitting                1    0.0788  93.771 -1938.5
## + sleep_weekdays              1    0.0524  93.797 -1938.3
## + moderate_activity           1    0.0495  93.800 -1938.2
## + height                      1    0.0361  93.813 -1938.1
## + poverty_ratio               1    0.0326  93.817 -1938.1
## + hours_worked                1    0.0285  93.821 -1938.0
## + time_outdoors_weekends      1    0.0209  93.829 -1938.0
## + stroke                      1    0.0127  93.837 -1937.9
## + take.away.within.30.days    1    0.0048  93.845 -1937.8
## + age                         1    0.0035  93.846 -1937.8
## - feel_at_risk_diabetes       1   17.4704 111.320 -1792.9
```

```
##
## Step:  AIC=-2013.72
## diabetes_risk ~ feel_at_risk_diabetes + diabetes_relatives
##
##                             Df Sum of Sq    RSS     AIC
## + frozen_meals               1    0.6337 85.387 -2018.2
## + diabetes                   1    0.5296 85.491 -2017.1
## + weight                     1    0.4399 85.581 -2016.2
## + gluten_free                1    0.3440 85.677 -2015.2
## + ethnicity                  1    0.2865 85.735 -2014.6
## + hypertension               1    0.2097 85.811 -2013.8
## <none>                                   86.021 -2013.7
## + gender                     1    0.1907 85.830 -2013.7
## + moderate_activity          1    0.0793 85.942 -2012.5
## + take_away_food             1    0.0511 85.970 -2012.2
## + time_outdoors_weekends     1    0.0456 85.975 -2012.2
## + time_sitting               1    0.0418 85.979 -2012.2
## + height                     1    0.0294 85.992 -2012.0
## + age                        1    0.0257 85.995 -2012.0
## + education_level            1    0.0234 85.998 -2012.0
## + sleep_weekdays             1    0.0177 86.003 -2011.9
## + hours_worked               1    0.0091 86.012 -2011.8
## + poverty_ratio              1    0.0032 86.018 -2011.8
## + take.away.within.30.days   1    0.0005 86.021 -2011.7
## + stroke                     1    0.0002 86.021 -2011.7
## - diabetes_relatives         1    7.8285 93.850 -1939.8
## - feel_at_risk_diabetes      1    8.1291 94.150 -1937.0
##
## Step:  AIC=-2018.17
## diabetes_risk ~ feel_at_risk_diabetes + diabetes_relatives +
##     frozen_meals
##
##                             Df Sum of Sq    RSS     AIC
## + diabetes                   1    0.5394 84.848 -2021.7
## + weight                     1    0.3963 84.991 -2020.2
## + gluten_free                1    0.3585 85.029 -2019.8
## + ethnicity                  1    0.2570 85.130 -2018.8
## <none>                                   85.387 -2018.2
## + gender                     1    0.1835 85.204 -2018.0
## + hypertension               1    0.1623 85.225 -2017.8
## + take_away_food             1    0.0939 85.293 -2017.1
## + moderate_activity          1    0.0652 85.322 -2016.8
## + time_sitting               1    0.0422 85.345 -2016.6
## + time_outdoors_weekends     1    0.0397 85.348 -2016.6
## + height                     1    0.0353 85.352 -2016.5
## + education_level            1    0.0237 85.364 -2016.4
## + sleep_weekdays             1    0.0196 85.368 -2016.4
## + hours_worked               1    0.0163 85.371 -2016.3
## + take.away.within.30.days   1    0.0096 85.378 -2016.3
## + age                        1    0.0015 85.386 -2016.2
```

```
## + poverty_ratio              1    0.0010 85.386 -2016.2
## + stroke                     1    0.0000 85.387 -2016.2
## - frozen_meals               1    0.6337 86.021 -2013.7
## - feel_at_risk_diabetes      1    7.8658 93.253 -1943.3
## - diabetes_relatives         1    8.0574 93.445 -1941.5
##
## Step:  AIC=-2021.7
## diabetes_risk ~ feel_at_risk_diabetes + diabetes_relatives +
##     frozen_meals + diabetes
##
##                             Df Sum of Sq    RSS     AIC
## + gluten_free                1    0.3806 84.467 -2023.6
## + weight                     1    0.3489 84.499 -2023.3
## + ethnicity                  1    0.2762 84.572 -2022.5
## <none>                                   84.848 -2021.7
## + gender                     1    0.1918 84.656 -2021.7
## + hypertension               1    0.1454 84.702 -2021.2
## + take_away_food             1    0.0781 84.770 -2020.5
## + moderate_activity          1    0.0632 84.785 -2020.3
## + height                     1    0.0435 84.804 -2020.2
## + time_sitting               1    0.0421 84.806 -2020.1
## + time_outdoors_weekends     1    0.0410 84.807 -2020.1
## + education_level            1    0.0208 84.827 -2019.9
## + sleep_weekdays             1    0.0175 84.830 -2019.9
## + hours_worked               1    0.0163 84.832 -2019.9
## + take.away.within.30.days   1    0.0119 84.836 -2019.8
## + age                        1    0.0100 84.838 -2019.8
## + poverty_ratio              1    0.0032 84.845 -2019.7
## + stroke                     1    0.0000 84.848 -2019.7
## - diabetes                   1    0.5394 85.387 -2018.2
## - frozen_meals               1    0.6435 85.491 -2017.1
## - feel_at_risk_diabetes      1    7.5468 92.395 -1949.4
## - diabetes_relatives         1    7.9651 92.813 -1945.5
##
## Step:  AIC=-2023.62
## diabetes_risk ~ feel_at_risk_diabetes + diabetes_relatives +
##     frozen_meals + diabetes + gluten_free
##
##                             Df Sum of Sq    RSS     AIC
## + weight                     1    0.3810 84.086 -2025.6
## + ethnicity                  1    0.2804 84.187 -2024.5
## <none>                                   84.467 -2023.6
## + gender                     1    0.1645 84.303 -2023.3
## + hypertension               1    0.1506 84.317 -2023.2
## + take_away_food             1    0.0731 84.394 -2022.4
## + moderate_activity          1    0.0591 84.408 -2022.2
## + time_sitting               1    0.0408 84.426 -2022.0
## + time_outdoors_weekends     1    0.0407 84.427 -2022.0
## + height                     1    0.0312 84.436 -2021.9
## + education_level            1    0.0294 84.438 -2021.9
```

```
## + sleep_weekdays            1    0.0179 84.449 -2021.8
## + hours_worked              1    0.0168 84.450 -2021.8
## + take.away.within.30.days  1    0.0152 84.452 -2021.8
## + poverty_ratio             1    0.0096 84.458 -2021.7
## - gluten_free               1    0.3806 84.848 -2021.7
## + age                       1    0.0062 84.461 -2021.7
## + stroke                    1    0.0001 84.467 -2021.6
## - diabetes                  1    0.5615 85.029 -2019.8
## - frozen_meals              1    0.6588 85.126 -2018.8
## - feel_at_risk_diabetes     1    7.5098 91.977 -1951.3
## - diabetes_relatives        1    7.9461 92.413 -1947.2
##
## Step:  AIC=-2025.56
## diabetes_risk ~ feel_at_risk_diabetes + diabetes_relatives +
##     frozen_meals + diabetes + gluten_free + weight
##
##                            Df Sum of Sq    RSS     AIC
## + gender                    1    0.4419 83.644 -2028.2
## + height                    1    0.3538 83.733 -2027.2
## + ethnicity                 1    0.2222 83.864 -2025.9
## <none>                                  84.086 -2025.6
## + take_away_food            1    0.1160 83.970 -2024.8
## + hypertension              1    0.0752 84.011 -2024.3
## + time_sitting              1    0.0554 84.031 -2024.1
## + moderate_activity         1    0.0475 84.039 -2024.0
## + education_level           1    0.0225 84.064 -2023.8
## + time_outdoors_weekends    1    0.0224 84.064 -2023.8
## + age                       1    0.0185 84.068 -2023.8
## + take.away.within.30.days  1    0.0155 84.071 -2023.7
## + poverty_ratio             1    0.0068 84.080 -2023.6
## - weight                    1    0.3810 84.467 -2023.6
## + sleep_weekdays            1    0.0027 84.084 -2023.6
## + hours_worked              1    0.0023 84.084 -2023.6
## + stroke                    1    0.0003 84.086 -2023.6
## - gluten_free               1    0.4127 84.499 -2023.3
## - diabetes                  1    0.5120 84.598 -2022.3
## - frozen_meals              1    0.6153 84.702 -2021.2
## - feel_at_risk_diabetes     1    6.8969 90.983 -1958.8
## - diabetes_relatives        1    7.7671 91.853 -1950.5
##
## Step:  AIC=-2028.15
## diabetes_risk ~ feel_at_risk_diabetes + diabetes_relatives +
##     frozen_meals + diabetes + gluten_free + weight + gender
##
##                            Df Sum of Sq    RSS     AIC
## + ethnicity                 1    0.2164 83.428 -2028.4
## <none>                                  83.644 -2028.2
## + moderate_activity         1    0.0958 83.549 -2027.2
## + time_outdoors_weekends    1    0.0673 83.577 -2026.9
## + height                    1    0.0640 83.580 -2026.8
```

```
## + hypertension               1    0.0595 83.585 -2026.8
## + take_away_food             1    0.0535 83.591 -2026.7
## + time_sitting               1    0.0443 83.600 -2026.6
## + age                        1    0.0301 83.614 -2026.5
## + education_level            1    0.0270 83.617 -2026.4
## + hours_worked               1    0.0256 83.619 -2026.4
## + take.away.within.30.days   1    0.0090 83.635 -2026.2
## - gluten_free                1    0.3775 84.022 -2026.2
## + poverty_ratio              1    0.0019 83.643 -2026.2
## + sleep_weekdays             1    0.0004 83.644 -2026.2
## + stroke                     1    0.0003 83.644 -2026.2
## - gender                     1    0.4419 84.086 -2025.6
## - diabetes                   1    0.5049 84.149 -2024.9
## - frozen_meals               1    0.5858 84.230 -2024.1
## - weight                     1    0.6583 84.303 -2023.3
## - feel_at_risk_diabetes      1    6.3447 89.989 -1966.4
## - diabetes_relatives         1    7.7239 91.368 -1953.1
##
## Step:  AIC=-2028.41
## diabetes_risk ~ feel_at_risk_diabetes + diabetes_relatives +
##     frozen_meals + diabetes + gluten_free + weight + gender +
##     ethnicity
##
##                            Df Sum of Sq    RSS     AIC
## <none>                                   83.428 -2028.4
## - ethnicity                 1    0.2164 83.644 -2028.2
## + moderate_activity         1    0.0927 83.335 -2027.4
## + hypertension              1    0.0663 83.362 -2027.1
## + take_away_food            1    0.0508 83.377 -2026.9
## + time_outdoors_weekends    1    0.0493 83.379 -2026.9
## + height                    1    0.0446 83.384 -2026.9
## + hours_worked              1    0.0311 83.397 -2026.7
## + age                       1    0.0225 83.406 -2026.7
## + time_sitting              1    0.0125 83.416 -2026.5
## + take.away.within.30.days  1    0.0063 83.422 -2026.5
## - gluten_free               1    0.3790 83.807 -2026.5
## + poverty_ratio             1    0.0030 83.425 -2026.5
## + sleep_weekdays            1    0.0027 83.425 -2026.4
## + stroke                    1    0.0005 83.428 -2026.4
## + education_level           1    0.0003 83.428 -2026.4
## - gender                    1    0.4361 83.864 -2025.9
## - diabetes                  1    0.5249 83.953 -2024.9
## - frozen_meals              1    0.5631 83.991 -2024.5
## - weight                    1    0.5830 84.011 -2024.3
## - feel_at_risk_diabetes     1    6.2046 89.633 -1967.9
## - diabetes_relatives        1    7.7900 91.218 -1952.6
```

```r
summary(stepNhanes)
```

```
##
## Call:
## lm(formula = diabetes_risk ~ feel_at_risk_diabetes + diabetes_relatives +
##     frozen_meals + diabetes + gluten_free + weight + gender +
##     ethnicity, data = nhanesData2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.06150 -0.03052  0.01255  0.19793  0.70306
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             1.4966706  0.2284771   6.551 9.84e-11 ***
## feel_at_risk_diabetes   0.2016373  0.0251690   8.011 3.67e-15 ***
## diabetes_relatives      0.2149918  0.0239499   8.977  < 2e-16 ***
## frozen_meals           -0.0049337  0.0020442  -2.413   0.0160 *
## diabetes               -0.2036479  0.0873981  -2.330   0.0200 *
## gluten_free             0.1146462  0.0578983   1.980   0.0480 *
## weight                 -0.0007007  0.0002853  -2.456   0.0143 *
## gender                 -0.0479950  0.0225984  -2.124   0.0340 *
## ethnicity               0.0133756  0.0089408   1.496   0.1350
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3109 on 863 degrees of freedom
## Multiple R-squared:  0.2506, Adjusted R-squared:  0.2436
## F-statistic: 36.06 on 8 and 863 DF,  p-value: < 2.2e-16
```

same as forward regression

take out ethnicity as it doesnt look to be significant in any of the models

```
model2 <- lm(diabetes_risk ~ feel_at_risk_diabetes + diabetes_relatives
             + frozen_meals + diabetes
             + gluten_free + weight + gender
             , data=nhanesData2
             )
summary(model2)
```

```
##
## Call:
## lm(formula = diabetes_risk ~ feel_at_risk_diabetes + diabetes_relatives +
##     frozen_meals + diabetes + gluten_free + weight + gender,
##     data = nhanesData2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.03954 -0.02917  0.00789  0.19818  0.71222
##
```

```
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             1.5387292  0.2269032   6.781 2.20e-11 ***
## feel_at_risk_diabetes   0.2036190  0.0251521   8.095 1.93e-15 ***
## diabetes_relatives      0.2139945  0.0239578   8.932  < 2e-16 ***
## frozen_meals           -0.0050296  0.0020447  -2.460  0.01410 *
## diabetes               -0.1996492  0.0874198  -2.284  0.02263 *
## gluten_free             0.1144146  0.0579396   1.975  0.04862 *
## weight                 -0.0007412  0.0002842  -2.608  0.00927 **
## gender                 -0.0483127  0.0226136  -2.136  0.03292 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3111 on 864 degrees of freedom
## Multiple R-squared:  0.2486, Adjusted R-squared:  0.2425
## F-statistic: 40.84 on 7 and 864 DF,  p-value: < 2.2e-16
```

Multiple R-squared decreases slightly to 0.2486 but all variables are now significant

what if calculated BMI is a better variable than weight (weight / height)

```
nhanesData2$BMI = nhanesData2$weight / nhanesData2$height
```

remove height and weight because of high correlation with BMI

```
nhanesData3 <- nhanesData2[,c(1:13,16:23)]
```

##re-do models

```
model2 <- lm(diabetes_risk ~ ., data=nhanesData3)
model2

##
## Call:
## lm(formula = diabetes_risk ~ ., data = nhanesData3)
##
## Coefficients:
##              (Intercept)              hypertension
gluten_free
##                1.481e+00                 2.488e-02                  1.088e-
01
##                   stroke         diabetes_relatives
moderate_activity
##               -1.140e-03                 2.134e-01                 -1.579e-
04
##                 diabetes     feel_at_risk_diabetes
frozen_meals
##               -2.037e-01                 1.988e-01                 -4.695e-
03
```

```
##                   gender                      age
ethnicity
##               -4.366e-02                9.503e-04                1.270e-
02
##          education_level            sleep_weekdays
hours_worked
##               -1.352e-04               -2.845e-04               -3.699e-
04
##   time_outdoors_weekends             time_sitting
take_away_food
##               -3.544e-05                2.088e-05                2.299e-
03
##            poverty_ratio  take.away.within.30.days
BMI
##               -4.603e-03                1.818e-04               -5.503e-
02
```

#check VIF

```
VIF(model2)
```

```
##             hypertension              gluten_free                   stroke
##                 1.181943                 1.014055                 1.047666
##       diabetes_relatives        moderate_activity                 diabetes
##                 1.195727                 1.096690                 1.020531
##    feel_at_risk_diabetes             frozen_meals                   gender
##                 1.232883                 1.062433                 1.170617
##                      age                ethnicity          education_level
##                 1.244281                 1.185415                 1.459609
##           sleep_weekdays             hours_worked   time_outdoors_weekends
##                 1.069553                 1.111125                 1.129454
##             time_sitting          take_away_food            poverty_ratio
##                 1.151147                 1.162525                 1.520379
## take.away.within.30.days                      BMI
##                 1.055021                 1.213493
```

no multicolinearity

#inspect the model

```
summary(model2)
```

```
##
## Call:
## lm(formula = diabetes_risk ~ ., data = nhanesData3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.03502 -0.03582  0.01672  0.19431  0.72515
##
## Coefficients:
```

```
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 1.481e+00  4.116e-01   3.598 0.000339 ***
## hypertension                2.488e-02  2.928e-02   0.850 0.395667
## gluten_free                 1.088e-01  5.842e-02   1.862 0.062904 .
## stroke                     -1.140e-03  1.602e-01  -0.007 0.994321
## diabetes_relatives          2.134e-01  2.436e-02   8.760  < 2e-16 ***
## moderate_activity          -1.579e-04  1.876e-04  -0.842 0.400131
## diabetes                   -2.037e-01  8.815e-02  -2.311 0.021057 *
## feel_at_risk_diabetes       1.988e-01  2.547e-02   7.803 1.76e-14 ***
## frozen_meals               -4.695e-03  2.107e-03  -2.228 0.026138 *
## gender                     -4.366e-02  2.289e-02  -1.908 0.056791 .
## age                         9.503e-04  1.068e-03   0.890 0.373703
## ethnicity                   1.270e-02  9.696e-03   1.309 0.190773
## education_level            -1.352e-04  1.292e-02  -0.010 0.991652
## sleep_weekdays             -2.845e-04  9.318e-03  -0.031 0.975651
## hours_worked               -3.699e-04  8.283e-04  -0.447 0.655319
## time_outdoors_weekends     -3.544e-05  9.661e-05  -0.367 0.713824
## time_sitting                2.088e-05  5.663e-05   0.369 0.712466
## take_away_food              2.299e-03  2.715e-03   0.847 0.397382
## poverty_ratio              -4.603e-03  7.913e-03  -0.582 0.560960
## take.away.within.30.days    1.818e-04  1.365e-03   0.133 0.894046
## BMI                        -5.503e-02  2.148e-02  -2.562 0.010583 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3122 on 851 degrees of freedom
## Multiple R-squared:  0.2546, Adjusted R-squared:  0.2371
## F-statistic: 14.54 on 20 and 851 DF,  p-value: < 2.2e-16
```

diabetes relatives, diabetes, feel_at_risk_diabetes, frozen_meals, gender, BMI

pvalue low, Multiple R-squared: 0.2546

#check automatic models

```
null = lm(diabetes_risk ~ 1, data=nhanesData3)
null

##
## Call:
## lm(formula = diabetes_risk ~ 1, data = nhanesData3)
##
## Coefficients:
## (Intercept)
##        1.85

full = lm(diabetes_risk ~ ., data=nhanesData3)
full
```

```
##
## Call:
## lm(formula = diabetes_risk ~ ., data = nhanesData3)
##
## Coefficients:
##              (Intercept)              hypertension
gluten_free
##                1.481e+00                 2.488e-02                  1.088e-
01
##                   stroke        diabetes_relatives
moderate_activity
##               -1.140e-03                 2.134e-01                -1.579e-
04
##                 diabetes     feel_at_risk_diabetes
frozen_meals
##               -2.037e-01                 1.988e-01                -4.695e-
03
##                   gender                       age
ethnicity
##               -4.366e-02                 9.503e-04                 1.270e-
02
##          education_level            sleep_weekdays
hours_worked
##               -1.352e-04                -2.845e-04                -3.699e-
04
##   time_outdoors_weekends              time_sitting
take_away_food
##               -3.544e-05                 2.088e-05                 2.299e-
03
##            poverty_ratio  take.away.within.30.days
BMI
##               -4.603e-03                 1.818e-04                -5.503e-
02
```

#Forward Regression

```
forwardNhanes = step(null, scope = list(lower=null, upper=full),
direction="forward")
```

```
## Start:  AIC=-1792.91
## diabetes_risk ~ 1
##
##                             Df Sum of Sq    RSS     AIC
## + feel_at_risk_diabetes      1   17.4704  93.85 -1939.8
## + diabetes_relatives         1   17.1698  94.15 -1937.0
## + BMI                        1    3.4557 107.86 -1818.4
## + diabetes                   1    1.2789 110.04 -1801.0
## + hypertension               1    0.9105 110.41 -1798.1
## + frozen_meals               1    0.6110 110.71 -1795.7
## + gender                     1    0.5404 110.78 -1795.2
```

```
## + ethnicity                  1    0.5311 110.79 -1795.1
## + gluten_free                1    0.4212 110.90 -1794.2
## + education_level            1    0.3763 110.94 -1793.9
## <none>                                   111.32 -1792.9
## + take_away_food             1    0.1980 111.12 -1792.5
## + poverty_ratio              1    0.1925 111.13 -1792.4
## + sleep_weekdays             1    0.1632 111.16 -1792.2
## + moderate_activity          1    0.1128 111.21 -1791.8
## + stroke                     1    0.0907 111.23 -1791.6
## + time_sitting               1    0.0455 111.27 -1791.3
## + hours_worked               1    0.0455 111.27 -1791.3
## + time_outdoors_weekends     1    0.0296 111.29 -1791.1
## + age                        1    0.0070 111.31 -1791.0
## + take.away.within.30.days   1    0.0009 111.32 -1790.9
##
## Step:  AIC=-1939.77
## diabetes_risk ~ feel_at_risk_diabetes
##
##                            Df Sum of Sq    RSS     AIC
## + diabetes_relatives        1    7.8285 86.021 -2013.7
## + BMI                       1    0.9450 92.905 -1946.6
## + diabetes                  1    0.6220 93.228 -1943.6
## + frozen_meals              1    0.4048 93.445 -1941.5
## + hypertension              1    0.3689 93.481 -1941.2
## + gluten_free               1    0.3631 93.486 -1941.2
## <none>                                   93.850 -1939.8
## + ethnicity                 1    0.2146 93.635 -1939.8
## + take_away_food            1    0.1865 93.663 -1939.5
## + education_level           1    0.1784 93.671 -1939.4
## + gender                    1    0.1755 93.674 -1939.4
## + time_sitting              1    0.0788 93.771 -1938.5
## + sleep_weekdays            1    0.0524 93.797 -1938.3
## + moderate_activity         1    0.0495 93.800 -1938.2
## + poverty_ratio             1    0.0326 93.817 -1938.1
## + hours_worked              1    0.0285 93.821 -1938.0
## + time_outdoors_weekends    1    0.0209 93.829 -1938.0
## + stroke                    1    0.0127 93.837 -1937.9
## + take.away.within.30.days  1    0.0048 93.845 -1937.8
## + age                       1    0.0035 93.846 -1937.8
##
## Step:  AIC=-2013.72
## diabetes_risk ~ feel_at_risk_diabetes + diabetes_relatives
##
##                            Df Sum of Sq    RSS     AIC
## + BMI                       1   0.69110 85.330 -2018.8
## + frozen_meals              1   0.63372 85.387 -2018.2
## + diabetes                  1   0.52965 85.491 -2017.1
## + gluten_free               1   0.34398 85.677 -2015.2
## + ethnicity                 1   0.28653 85.735 -2014.6
## + hypertension              1   0.20975 85.811 -2013.8
```

```
## <none>                                  86.021 -2013.7
## + gender                     1   0.19074 85.830 -2013.7
## + moderate_activity          1   0.07931 85.942 -2012.5
## + take_away_food             1   0.05111 85.970 -2012.2
## + time_outdoors_weekends     1   0.04561 85.975 -2012.2
## + time_sitting               1   0.04179 85.979 -2012.2
## + age                        1   0.02569 85.995 -2012.0
## + education_level            1   0.02337 85.998 -2012.0
## + sleep_weekdays             1   0.01775 86.003 -2011.9
## + hours_worked               1   0.00908 86.012 -2011.8
## + poverty_ratio              1   0.00320 86.018 -2011.8
## + take.away.within.30.days   1   0.00053 86.021 -2011.7
## + stroke                     1   0.00023 86.021 -2011.7
##
## Step:  AIC=-2018.76
## diabetes_risk ~ feel_at_risk_diabetes + diabetes_relatives +
##     BMI
##
##                             Df Sum of Sq    RSS     AIC
## + frozen_meals               1   0.57579 84.754 -2022.7
## + diabetes                   1   0.47588 84.854 -2021.6
## + gender                     1   0.38662 84.943 -2020.7
## + gluten_free                1   0.38068 84.949 -2020.7
## <none>                                   85.330 -2018.8
## + ethnicity                  1   0.19380 85.136 -2018.7
## + hypertension               1   0.08544 85.245 -2017.6
## + take_away_food             1   0.08338 85.247 -2017.6
## + moderate_activity          1   0.07508 85.255 -2017.5
## + age                        1   0.05982 85.270 -2017.4
## + time_sitting               1   0.05817 85.272 -2017.3
## + time_outdoors_weekends     1   0.02784 85.302 -2017.0
## + education_level            1   0.00772 85.322 -2016.8
## + take.away.within.30.days   1   0.00066 85.329 -2016.8
## + sleep_weekdays             1   0.00053 85.329 -2016.8
## + hours_worked               1   0.00032 85.330 -2016.8
## + stroke                     1   0.00013 85.330 -2016.8
## + poverty_ratio              1   0.00001 85.330 -2016.8
##
## Step:  AIC=-2022.66
## diabetes_risk ~ feel_at_risk_diabetes + diabetes_relatives +
##     BMI + frozen_meals
##
##                             Df Sum of Sq    RSS     AIC
## + diabetes                   1   0.48703 84.267 -2025.7
## + gluten_free                1   0.39367 84.360 -2024.7
## + gender                     1   0.36770 84.386 -2024.5
## <none>                                   84.754 -2022.7
## + ethnicity                  1   0.17390 84.580 -2022.5
## + take_away_food             1   0.13198 84.622 -2022.0
## + moderate_activity          1   0.06211 84.692 -2021.3
```

```
## + hypertension               1   0.06025 84.694 -2021.3
## + time_sitting               1   0.05788 84.696 -2021.3
## + time_outdoors_weekends     1   0.02406 84.730 -2020.9
## + age                        1   0.01590 84.738 -2020.8
## + take.away.within.30.days   1   0.00945 84.745 -2020.8
## + education_level            1   0.00841 84.746 -2020.8
## + hours_worked               1   0.00272 84.751 -2020.7
## + sleep_weekdays             1   0.00117 84.753 -2020.7
## + poverty_ratio              1   0.00061 84.754 -2020.7
## + stroke                     1   0.00002 84.754 -2020.7
##
## Step:  AIC=-2025.69
## diabetes_risk ~ feel_at_risk_diabetes + diabetes_relatives +
##     BMI + frozen_meals + diabetes
##
##                             Df Sum of Sq    RSS     AIC
## + gluten_free                1   0.41416 83.853 -2028.0
## + gender                     1   0.37059 83.897 -2027.5
## <none>                                   84.267 -2025.7
## + ethnicity                  1   0.19246 84.075 -2025.7
## + take_away_food             1   0.11248 84.155 -2024.8
## + moderate_activity          1   0.06041 84.207 -2024.3
## + time_sitting               1   0.05714 84.210 -2024.3
## + hypertension               1   0.05346 84.214 -2024.2
## + age                        1   0.03280 84.234 -2024.0
## + time_outdoors_weekends     1   0.02561 84.242 -2024.0
## + take.away.within.30.days   1   0.01156 84.256 -2023.8
## + education_level            1   0.00723 84.260 -2023.8
## + hours_worked               1   0.00303 84.264 -2023.7
## + sleep_weekdays             1   0.00097 84.266 -2023.7
## + stroke                     1   0.00005 84.267 -2023.7
## + poverty_ratio              1   0.00000 84.267 -2023.7
##
## Step:  AIC=-2027.98
## diabetes_risk ~ feel_at_risk_diabetes + diabetes_relatives +
##     BMI + frozen_meals + diabetes + gluten_free
##
##                             Df Sum of Sq    RSS     AIC
## + gender                     1   0.33514 83.518 -2029.5
## + ethnicity                  1   0.19380 83.659 -2028.0
## <none>                                   83.853 -2028.0
## + take_away_food             1   0.10721 83.746 -2027.1
## + moderate_activity          1   0.05618 83.797 -2026.6
## + time_sitting               1   0.05596 83.797 -2026.6
## + hypertension               1   0.05466 83.798 -2026.5
## + age                        1   0.02595 83.827 -2026.2
## + time_outdoors_weekends     1   0.02499 83.828 -2026.2
## + take.away.within.30.days   1   0.01493 83.838 -2026.1
## + education_level            1   0.01248 83.840 -2026.1
## + hours_worked               1   0.00306 83.850 -2026.0
```

```
## + poverty_ratio              1   0.00187 83.851 -2026.0
## + sleep_weekdays             1   0.00088 83.852 -2026.0
## + stroke                     1   0.00023 83.853 -2026.0
##
## Step:  AIC=-2029.48
## diabetes_risk ~ feel_at_risk_diabetes + diabetes_relatives +
##     BMI + frozen_meals + diabetes + gluten_free + gender
##
##                             Df Sum of Sq    RSS     AIC
## + ethnicity                  1  0.194542 83.323 -2029.5
## <none>                                   83.518 -2029.5
## + moderate_activity          1  0.103508 83.414 -2028.6
## + time_outdoors_weekends     1  0.068258 83.450 -2028.2
## + hypertension               1  0.050133 83.468 -2028.0
## + take_away_food             1  0.050041 83.468 -2028.0
## + time_sitting               1  0.044134 83.474 -2027.9
## + age                        1  0.034553 83.483 -2027.8
## + hours_worked               1  0.027257 83.491 -2027.8
## + education_level            1  0.014794 83.503 -2027.6
## + take.away.within.30.days   1  0.009388 83.508 -2027.6
## + sleep_weekdays             1  0.000201 83.518 -2027.5
## + stroke                     1  0.000154 83.518 -2027.5
## + poverty_ratio              1  0.000068 83.518 -2027.5
##
## Step:  AIC=-2029.51
## diabetes_risk ~ feel_at_risk_diabetes + diabetes_relatives +
##     BMI + frozen_meals + diabetes + gluten_free + gender + ethnicity
##
##                             Df Sum of Sq    RSS     AIC
## <none>                                   83.323 -2029.5
## + moderate_activity          1  0.099956 83.223 -2028.6
## + hypertension               1  0.057081 83.266 -2028.1
## + time_outdoors_weekends     1  0.050986 83.272 -2028.0
## + take_away_food             1  0.047617 83.276 -2028.0
## + hours_worked               1  0.032611 83.291 -2027.8
## + age                        1  0.026404 83.297 -2027.8
## + time_sitting               1  0.013607 83.310 -2027.7
## + poverty_ratio              1  0.007171 83.316 -2027.6
## + take.away.within.30.days   1  0.006776 83.317 -2027.6
## + sleep_weekdays             1  0.002107 83.321 -2027.5
## + education_level            1  0.000314 83.323 -2027.5
## + stroke                     1  0.000293 83.323 -2027.5
```

```
summary(forwardNhanes)
```

```
##
## Call:
## lm(formula = diabetes_risk ~ feel_at_risk_diabetes + diabetes_relatives +
##     BMI + frozen_meals + diabetes + gluten_free + gender + ethnicity,
##     data = nhanesData3)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.05545 -0.03108  0.01262  0.19897  0.71201
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)              1.513421   0.228727   6.617 6.44e-11 ***
## feel_at_risk_diabetes    0.199252   0.025253   7.890 9.13e-15 ***
## diabetes_relatives       0.214398   0.023942   8.955  < 2e-16 ***
## BMI                     -0.054801   0.020532  -2.669  0.00775 **
## frozen_meals            -0.004923   0.002043  -2.410  0.01617 *
## diabetes                -0.204281   0.087311  -2.340  0.01953 *
## gluten_free              0.114780   0.057860   1.984  0.04760 *
## gender                  -0.040296   0.021605  -1.865  0.06250 .
## ethnicity                0.012710   0.008954   1.419  0.15612
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3107 on 863 degrees of freedom
## Multiple R-squared:  0.2515, Adjusted R-squared:  0.2446
## F-statistic: 36.25 on 8 and 863 DF,  p-value: < 2.2e-16
```

feel_at_risk_diabetes, diabetes_relatives, BMI, frozen_meals, diabetes, gluten_free, gender, ethnicity

##pvalue low, Multiple R-squared: 0.2515

#Backward Regression

```
backwardNhanes = step(full, direction="backward")
```

```
## Start:  AIC=-2009.19
## diabetes_risk ~ hypertension + gluten_free + stroke + diabetes_relatives +
##     moderate_activity + diabetes + feel_at_risk_diabetes + frozen_meals +
##     gender + age + ethnicity + education_level + sleep_weekdays +
##     hours_worked + time_outdoors_weekends + time_sitting + take_away_food
+
##     poverty_ratio + take.away.within.30.days + BMI
##
##                             Df Sum of Sq    RSS     AIC
## - stroke                     1    0.0000 82.972 -2011.2
## - education_level            1    0.0000 82.972 -2011.2
## - sleep_weekdays             1    0.0001 82.973 -2011.2
## - take.away.within.30.days   1    0.0017 82.974 -2011.2
## - time_outdoors_weekends     1    0.0131 82.986 -2011.0
## - time_sitting               1    0.0133 82.986 -2011.0
## - hours_worked               1    0.0194 82.992 -2011.0
## - poverty_ratio              1    0.0330 83.005 -2010.8
## - moderate_activity          1    0.0691 83.042 -2010.5
```

```
## - take_away_food            1    0.0699 83.042 -2010.5
## - hypertension              1    0.0704 83.043 -2010.5
## - age                       1    0.0772 83.050 -2010.4
## - ethnicity                 1    0.1672 83.140 -2009.4
## <none>                                  82.972 -2009.2
## - gluten_free               1    0.3381 83.311 -2007.6
## - gender                    1    0.3548 83.327 -2007.5
## - frozen_meals              1    0.4840 83.456 -2006.1
## - diabetes                  1    0.5208 83.493 -2005.7
## - BMI                       1    0.6399 83.612 -2004.5
## - feel_at_risk_diabetes     1    5.9370 88.909 -1950.9
## - diabetes_relatives        1    7.4812 90.454 -1935.9
##
## Step:  AIC=-2011.19
## diabetes_risk ~ hypertension + gluten_free + diabetes_relatives +
##     moderate_activity + diabetes + feel_at_risk_diabetes + frozen_meals +
##     gender + age + ethnicity + education_level + sleep_weekdays +
##     hours_worked + time_outdoors_weekends + time_sitting + take_away_food
+
##     poverty_ratio + take.away.within.30.days + BMI
##
##                            Df Sum of Sq    RSS     AIC
## - education_level           1    0.0000 82.972 -2013.2
## - sleep_weekdays            1    0.0001 82.973 -2013.2
## - take.away.within.30.days  1    0.0017 82.974 -2013.2
## - time_outdoors_weekends    1    0.0131 82.986 -2013.0
## - time_sitting              1    0.0133 82.986 -2013.0
## - hours_worked              1    0.0194 82.992 -2013.0
## - poverty_ratio             1    0.0331 83.006 -2012.8
## - moderate_activity         1    0.0691 83.042 -2012.5
## - take_away_food            1    0.0700 83.042 -2012.5
## - hypertension              1    0.0716 83.044 -2012.4
## - age                       1    0.0778 83.050 -2012.4
## - ethnicity                 1    0.1674 83.140 -2011.4
## <none>                                  82.972 -2011.2
## - gluten_free               1    0.3382 83.311 -2009.6
## - gender                    1    0.3548 83.327 -2009.5
## - frozen_meals              1    0.4843 83.457 -2008.1
## - diabetes                  1    0.5209 83.493 -2007.7
## - BMI                       1    0.6407 83.613 -2006.5
## - feel_at_risk_diabetes     1    5.9423 88.915 -1952.9
## - diabetes_relatives        1    7.4998 90.472 -1937.7
##
## Step:  AIC=-2013.19
## diabetes_risk ~ hypertension + gluten_free + diabetes_relatives +
##     moderate_activity + diabetes + feel_at_risk_diabetes + frozen_meals +
##     gender + age + ethnicity + sleep_weekdays + hours_worked +
##     time_outdoors_weekends + time_sitting + take_away_food +
##     poverty_ratio + take.away.within.30.days + BMI
##
```

```
##                             Df Sum of Sq    RSS     AIC
## - sleep_weekdays             1    0.0001 82.973 -2015.2
## - take.away.within.30.days   1    0.0017 82.974 -2015.2
## - time_outdoors_weekends     1    0.0132 82.986 -2015.0
## - time_sitting               1    0.0133 82.986 -2015.0
## - hours_worked               1    0.0194 82.992 -2015.0
## - poverty_ratio              1    0.0411 83.014 -2014.8
## - moderate_activity          1    0.0692 83.042 -2014.5
## - take_away_food             1    0.0700 83.042 -2014.5
## - hypertension               1    0.0716 83.044 -2014.4
## - age                        1    0.0782 83.051 -2014.4
## - ethnicity                  1    0.1763 83.149 -2013.3
## <none>                                   82.972 -2013.2
## - gluten_free                1    0.3383 83.311 -2011.6
## - gender                     1    0.3558 83.328 -2011.5
## - frozen_meals               1    0.4844 83.457 -2010.1
## - diabetes                   1    0.5213 83.494 -2009.7
## - BMI                        1    0.6407 83.613 -2008.5
## - feel_at_risk_diabetes      1    5.9467 88.919 -1954.8
## - diabetes_relatives         1    7.5529 90.525 -1939.2
##
## Step:  AIC=-2015.19
## diabetes_risk ~ hypertension + gluten_free + diabetes_relatives +
##     moderate_activity + diabetes + feel_at_risk_diabetes + frozen_meals +
##     gender + age + ethnicity + hours_worked + time_outdoors_weekends +
##     time_sitting + take_away_food + poverty_ratio +
take.away.within.30.days +
##     BMI
##
##                             Df Sum of Sq    RSS     AIC
## - take.away.within.30.days   1    0.0018 82.974 -2017.2
## - time_outdoors_weekends     1    0.0131 82.986 -2017.0
## - time_sitting               1    0.0133 82.986 -2017.0
## - hours_worked               1    0.0194 82.992 -2017.0
## - poverty_ratio              1    0.0414 83.014 -2016.8
## - moderate_activity          1    0.0691 83.042 -2016.5
## - take_away_food             1    0.0699 83.042 -2016.5
## - hypertension               1    0.0715 83.044 -2016.4
## - age                        1    0.0785 83.051 -2016.4
## - ethnicity                  1    0.1786 83.151 -2015.3
## <none>                                   82.973 -2015.2
## - gluten_free                1    0.3382 83.311 -2013.6
## - gender                     1    0.3561 83.329 -2013.5
## - frozen_meals               1    0.4846 83.457 -2012.1
## - diabetes                   1    0.5213 83.494 -2011.7
## - BMI                        1    0.6480 83.621 -2010.4
## - feel_at_risk_diabetes      1    5.9467 88.919 -1956.8
## - diabetes_relatives         1    7.5535 90.526 -1941.2
##
## Step:  AIC=-2017.17
```

```
## diabetes_risk ~ hypertension + gluten_free + diabetes_relatives +
##     moderate_activity + diabetes + feel_at_risk_diabetes + frozen_meals +
##     gender + age + ethnicity + hours_worked + time_outdoors_weekends +
##     time_sitting + take_away_food + poverty_ratio + BMI
##
##                          Df Sum of Sq    RSS     AIC
## - time_outdoors_weekends  1    0.0127 82.987 -2019.0
## - time_sitting            1    0.0138 82.988 -2019.0
## - hours_worked            1    0.0198 82.994 -2019.0
## - poverty_ratio           1    0.0412 83.016 -2018.7
## - moderate_activity       1    0.0697 83.044 -2018.4
## - hypertension            1    0.0713 83.046 -2018.4
## - take_away_food          1    0.0756 83.050 -2018.4
## - age                     1    0.0781 83.052 -2018.3
## - ethnicity               1    0.1794 83.154 -2017.3
## <none>                                 82.974 -2017.2
## - gluten_free             1    0.3372 83.312 -2015.6
## - gender                  1    0.3567 83.331 -2015.4
## - frozen_meals            1    0.4829 83.457 -2014.1
## - diabetes                1    0.5202 83.495 -2013.7
## - BMI                     1    0.6490 83.623 -2012.4
## - feel_at_risk_diabetes   1    5.9644 88.939 -1958.6
## - diabetes_relatives      1    7.5627 90.537 -1943.1
##
## Step:  AIC=-2019.04
## diabetes_risk ~ hypertension + gluten_free + diabetes_relatives +
##     moderate_activity + diabetes + feel_at_risk_diabetes + frozen_meals +
##     gender + age + ethnicity + hours_worked + time_sitting +
##     take_away_food + poverty_ratio + BMI
##
##                         Df Sum of Sq    RSS     AIC
## - time_sitting           1    0.0138 83.001 -2020.9
## - hours_worked           1    0.0232 83.010 -2020.8
## - poverty_ratio          1    0.0399 83.027 -2020.6
## - hypertension           1    0.0745 83.062 -2020.2
## - take_away_food         1    0.0785 83.066 -2020.2
## - age                    1    0.0844 83.071 -2020.2
## - moderate_activity      1    0.0864 83.073 -2020.1
## - ethnicity              1    0.1874 83.174 -2019.1
## <none>                               82.987 -2019.0
## - gluten_free            1    0.3380 83.325 -2017.5
## - gender                 1    0.3451 83.332 -2017.4
## - frozen_meals           1    0.4828 83.470 -2016.0
## - diabetes               1    0.5201 83.507 -2015.6
## - BMI                    1    0.6507 83.638 -2014.2
## - feel_at_risk_diabetes  1    5.9707 88.958 -1960.5
## - diabetes_relatives     1    7.5513 90.538 -1945.1
##
## Step:  AIC=-2020.89
## diabetes_risk ~ hypertension + gluten_free + diabetes_relatives +
```

```
##     moderate_activity + diabetes + feel_at_risk_diabetes + frozen_meals +
##     gender + age + ethnicity + hours_worked + take_away_food +
##     poverty_ratio + BMI
##
##                          Df Sum of Sq    RSS     AIC
## - hours_worked            1    0.0226 83.024 -2022.7
## - poverty_ratio           1    0.0328 83.034 -2022.5
## - hypertension            1    0.0715 83.072 -2022.1
## - age                     1    0.0800 83.081 -2022.0
## - take_away_food          1    0.0873 83.088 -2022.0
## - moderate_activity       1    0.0917 83.093 -2021.9
## <none>                                 83.001 -2020.9
## - ethnicity               1    0.2134 83.214 -2020.7
## - gluten_free             1    0.3403 83.341 -2019.3
## - gender                  1    0.3479 83.349 -2019.2
## - frozen_meals            1    0.4850 83.486 -2017.8
## - diabetes                1    0.5212 83.522 -2017.4
## - BMI                     1    0.6414 83.642 -2016.2
## - feel_at_risk_diabetes   1    5.9580 88.959 -1962.4
## - diabetes_relatives      1    7.5689 90.570 -1946.8
##
## Step:  AIC=-2022.65
## diabetes_risk ~ hypertension + gluten_free + diabetes_relatives +
##     moderate_activity + diabetes + feel_at_risk_diabetes + frozen_meals +
##     gender + age + ethnicity + take_away_food + poverty_ratio +
##     BMI
##
##                          Df Sum of Sq    RSS     AIC
## - poverty_ratio           1    0.0417 83.065 -2024.2
## - hypertension            1    0.0782 83.102 -2023.8
## - age                     1    0.0817 83.105 -2023.8
## - take_away_food          1    0.0820 83.105 -2023.8
## - moderate_activity       1    0.0956 83.119 -2023.7
## <none>                                 83.024 -2022.7
## - ethnicity               1    0.2130 83.237 -2022.4
## - gender                  1    0.3286 83.352 -2021.2
## - gluten_free             1    0.3399 83.363 -2021.1
## - frozen_meals            1    0.4751 83.499 -2019.7
## - diabetes                1    0.5195 83.543 -2019.2
## - BMI                     1    0.6561 83.680 -2017.8
## - feel_at_risk_diabetes   1    5.9613 88.985 -1964.2
## - diabetes_relatives      1    7.5996 90.623 -1948.3
##
## Step:  AIC=-2024.21
## diabetes_risk ~ hypertension + gluten_free + diabetes_relatives +
##     moderate_activity + diabetes + feel_at_risk_diabetes + frozen_meals +
##     gender + age + ethnicity + take_away_food + BMI
##
##                          Df Sum of Sq    RSS     AIC
## - age                     1    0.0589 83.124 -2025.6
```

```
## - take_away_food         1    0.0650 83.130 -2025.5
## - hypertension           1    0.0759 83.141 -2025.4
## - moderate_activity      1    0.0854 83.151 -2025.3
## - ethnicity              1    0.1844 83.250 -2024.3
## <none>                             83.065 -2024.2
## - gender                 1    0.3170 83.382 -2022.9
## - gluten_free            1    0.3623 83.427 -2022.4
## - frozen_meals           1    0.4773 83.543 -2021.2
## - diabetes               1    0.5267 83.592 -2020.7
## - BMI                    1    0.6312 83.696 -2019.6
## - feel_at_risk_diabetes  1    5.9477 89.013 -1965.9
## - diabetes_relatives     1    7.5640 90.629 -1950.2
##
## Step:  AIC=-2025.6
## diabetes_risk ~ hypertension + gluten_free + diabetes_relatives +
##     moderate_activity + diabetes + feel_at_risk_diabetes + frozen_meals +
##     gender + ethnicity + take_away_food + BMI
##
##                          Df Sum of Sq    RSS     AIC
## - hypertension           1    0.0486 83.173 -2027.1
## - take_away_food         1    0.0524 83.176 -2027.0
## - moderate_activity      1    0.0925 83.217 -2026.6
## <none>                             83.124 -2025.6
## - ethnicity              1    0.1949 83.319 -2025.5
## - gender                 1    0.3140 83.438 -2024.3
## - gluten_free            1    0.3724 83.496 -2023.7
## - diabetes               1    0.5050 83.629 -2022.3
## - frozen_meals           1    0.5500 83.674 -2021.8
## - BMI                    1    0.6084 83.733 -2021.2
## - feel_at_risk_diabetes  1    5.9456 89.070 -1967.3
## - diabetes_relatives     1    7.5659 90.690 -1951.6
##
## Step:  AIC=-2027.09
## diabetes_risk ~ gluten_free + diabetes_relatives + moderate_activity +
##     diabetes + feel_at_risk_diabetes + frozen_meals + gender +
##     ethnicity + take_away_food + BMI
##
##                          Df Sum of Sq    RSS     AIC
## - take_away_food         1    0.0506 83.223 -2028.6
## - moderate_activity      1    0.1030 83.276 -2028.0
## - ethnicity              1    0.1885 83.361 -2027.1
## <none>                             83.173 -2027.1
## - gender                 1    0.3217 83.494 -2025.7
## - gluten_free            1    0.3707 83.543 -2025.2
## - diabetes               1    0.5114 83.684 -2023.7
## - frozen_meals           1    0.5713 83.744 -2023.1
## - BMI                    1    0.7116 83.884 -2021.7
## - feel_at_risk_diabetes  1    5.9583 89.131 -1968.8
## - diabetes_relatives     1    7.6438 90.816 -1952.4
##
```

```
## Step:  AIC=-2028.56
## diabetes_risk ~ gluten_free + diabetes_relatives + moderate_activity +
##     diabetes + feel_at_risk_diabetes + frozen_meals + gender +
##     ethnicity + BMI
##
##                         Df Sum of Sq    RSS     AIC
## - moderate_activity      1    0.1000 83.323 -2029.5
## - ethnicity              1    0.1910 83.414 -2028.6
## <none>                              83.223 -2028.6
## - gluten_free            1    0.3718 83.595 -2026.7
## - gender                 1    0.3823 83.606 -2026.6
## - diabetes               1    0.5261 83.749 -2025.1
## - frozen_meals           1    0.5431 83.766 -2024.9
## - BMI                    1    0.6994 83.923 -2023.3
## - feel_at_risk_diabetes  1    5.9314 89.155 -1970.5
## - diabetes_relatives     1    7.7731 90.996 -1952.7
##
## Step:  AIC=-2029.51
## diabetes_risk ~ gluten_free + diabetes_relatives + diabetes +
##     feel_at_risk_diabetes + frozen_meals + gender + ethnicity +
##     BMI
##
##                         Df Sum of Sq    RSS     AIC
## <none>                              83.323 -2029.5
## - ethnicity              1    0.1945 83.518 -2029.5
## - gender                 1    0.3359 83.659 -2028.0
## - gluten_free            1    0.3799 83.703 -2027.5
## - diabetes               1    0.5285 83.852 -2026.0
## - frozen_meals           1    0.5607 83.884 -2025.7
## - BMI                    1    0.6878 84.011 -2024.3
## - feel_at_risk_diabetes  1    6.0107 89.334 -1970.8
## - diabetes_relatives     1    7.7427 91.066 -1954.0
```

```
summary(backwardNhanes)
```

```
##
## Call:
## lm(formula = diabetes_risk ~ gluten_free + diabetes_relatives +
##     diabetes + feel_at_risk_diabetes + frozen_meals + gender +
##     ethnicity + BMI, data = nhanesData3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.05545 -0.03108  0.01262  0.19897  0.71201
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)             1.513421   0.228727   6.617 6.44e-11 ***
## gluten_free             0.114780   0.057860   1.984  0.04760 *
## diabetes_relatives      0.214398   0.023942   8.955  < 2e-16 ***
```

```
## diabetes               -0.204281   0.087311  -2.340  0.01953 *
## feel_at_risk_diabetes   0.199252   0.025253   7.890 9.13e-15 ***
## frozen_meals           -0.004923   0.002043  -2.410  0.01617 *
## gender                 -0.040296   0.021605  -1.865  0.06250 .
## ethnicity               0.012710   0.008954   1.419  0.15612
## BMI                    -0.054801   0.020532  -2.669  0.00775 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3107 on 863 degrees of freedom
## Multiple R-squared:  0.2515, Adjusted R-squared:  0.2446
## F-statistic: 36.25 on 8 and 863 DF,  p-value: < 2.2e-16
```

gluten_free, diabetes_relatives, diabetes, feel_at_risk_diabetes, frozen_meals, gender, ethnicity, BMI

##pvalue low, Multiple R-squared: 0.2515

take out ethnicity as it doesnt look to be significant in any of the models

```
model3 <- lm(diabetes_risk ~ feel_at_risk_diabetes + diabetes_relatives
             + frozen_meals + diabetes
             + gluten_free + BMI + gender
             , data=nhanesData3
             )
summary(model3)
```

```
##
## Call:
## lm(formula = diabetes_risk ~ feel_at_risk_diabetes + diabetes_relatives +
##     frozen_meals + diabetes + gluten_free + BMI + gender, data =
nhanesData3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.03806 -0.02846  0.00823  0.19806  0.72136
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)             1.555207   0.226958   6.852 1.38e-11 ***
## feel_at_risk_diabetes   0.200913   0.025241   7.960 5.40e-15 ***
## diabetes_relatives      0.213399   0.023945   8.912  < 2e-16 ***
## frozen_meals           -0.005011   0.002043  -2.453  0.01437 *
## diabetes               -0.200460   0.087321  -2.296  0.02193 *
## gluten_free             0.114592   0.057894   1.979  0.04810 *
## BMI                    -0.058154   0.020407  -2.850  0.00448 **
## gender                 -0.040252   0.021617  -1.862  0.06294 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.3109 on 864 degrees of freedom
## Multiple R-squared:  0.2497, Adjusted R-squared:  0.2437
## F-statistic: 41.09 on 7 and 864 DF,  p-value: < 2.2e-16
```

##pvalue low, Multiple R-squared: 0.2497 ## gender is now not significant, remove from model

```
model4 <- lm(diabetes_risk ~ feel_at_risk_diabetes + diabetes_relatives
             + frozen_meals + diabetes
             + gluten_free + BMI
             , data=nhanesData3
             )
summary(model4)
```

```
##
## Call:
## lm(formula = diabetes_risk ~ feel_at_risk_diabetes + diabetes_relatives +
##     frozen_meals + diabetes + gluten_free + BMI, data = nhanesData3)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.0373 -0.0255  0.0088  0.2006  0.6981
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)             1.455255   0.220832   6.590 7.64e-11 ***
## feel_at_risk_diabetes   0.206095   0.025123   8.203 8.42e-16 ***
## diabetes_relatives      0.213629   0.023979   8.909  < 2e-16 ***
## frozen_meals           -0.005092   0.002045  -2.489   0.0130 *
## diabetes               -0.200083   0.087445  -2.288   0.0224 *
## gluten_free             0.119701   0.057912   2.067   0.0390 *
## BMI                    -0.050349   0.020001  -2.517   0.0120 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3114 on 865 degrees of freedom
## Multiple R-squared:  0.2467, Adjusted R-squared:  0.2415
## F-statistic: 47.22 on 6 and 865 DF,  p-value: < 2.2e-16
```

##pvalue low, Multiple R-squared: 0.2467, all variables are significant

```
step <- stepAIC(full, direction = "backward")
```

```
## Start:  AIC=-2009.19
## diabetes_risk ~ hypertension + gluten_free + stroke + diabetes_relatives +
##     moderate_activity + diabetes + feel_at_risk_diabetes + frozen_meals +
##     gender + age + ethnicity + education_level + sleep_weekdays +
##     hours_worked + time_outdoors_weekends + time_sitting + take_away_food
+
##     poverty_ratio + take.away.within.30.days + BMI
```

```
##
##                             Df Sum of Sq    RSS     AIC
## - stroke                     1    0.0000 82.972 -2011.2
## - education_level            1    0.0000 82.972 -2011.2
## - sleep_weekdays             1    0.0001 82.973 -2011.2
## - take.away.within.30.days   1    0.0017 82.974 -2011.2
## - time_outdoors_weekends     1    0.0131 82.986 -2011.0
## - time_sitting               1    0.0133 82.986 -2011.0
## - hours_worked               1    0.0194 82.992 -2011.0
## - poverty_ratio              1    0.0330 83.005 -2010.8
## - moderate_activity          1    0.0691 83.042 -2010.5
## - take_away_food             1    0.0699 83.042 -2010.5
## - hypertension               1    0.0704 83.043 -2010.5
## - age                        1    0.0772 83.050 -2010.4
## - ethnicity                  1    0.1672 83.140 -2009.4
## <none>                                   82.972 -2009.2
## - gluten_free                1    0.3381 83.311 -2007.6
## - gender                     1    0.3548 83.327 -2007.5
## - frozen_meals               1    0.4840 83.456 -2006.1
## - diabetes                   1    0.5208 83.493 -2005.7
## - BMI                        1    0.6399 83.612 -2004.5
## - feel_at_risk_diabetes      1    5.9370 88.909 -1950.9
## - diabetes_relatives         1    7.4812 90.454 -1935.9
##
## Step:  AIC=-2011.19
## diabetes_risk ~ hypertension + gluten_free + diabetes_relatives +
##     moderate_activity + diabetes + feel_at_risk_diabetes + frozen_meals +
##     gender + age + ethnicity + education_level + sleep_weekdays +
##     hours_worked + time_outdoors_weekends + time_sitting + take_away_food
+
##     poverty_ratio + take.away.within.30.days + BMI
##
##                             Df Sum of Sq    RSS     AIC
## - education_level            1    0.0000 82.972 -2013.2
## - sleep_weekdays             1    0.0001 82.973 -2013.2
## - take.away.within.30.days   1    0.0017 82.974 -2013.2
## - time_outdoors_weekends     1    0.0131 82.986 -2013.0
## - time_sitting               1    0.0133 82.986 -2013.0
## - hours_worked               1    0.0194 82.992 -2013.0
## - poverty_ratio              1    0.0331 83.006 -2012.8
## - moderate_activity          1    0.0691 83.042 -2012.5
## - take_away_food             1    0.0700 83.042 -2012.5
## - hypertension               1    0.0716 83.044 -2012.4
## - age                        1    0.0778 83.050 -2012.4
## - ethnicity                  1    0.1674 83.140 -2011.4
## <none>                                   82.972 -2011.2
## - gluten_free                1    0.3382 83.311 -2009.6
## - gender                     1    0.3548 83.327 -2009.5
## - frozen_meals               1    0.4843 83.457 -2008.1
## - diabetes                   1    0.5209 83.493 -2007.7
```

```
## - BMI                         1    0.6407 83.613 -2006.5
## - feel_at_risk_diabetes     1    5.9423 88.915 -1952.9
## - diabetes_relatives        1    7.4998 90.472 -1937.7
##
## Step:  AIC=-2013.19
## diabetes_risk ~ hypertension + gluten_free + diabetes_relatives +
##     moderate_activity + diabetes + feel_at_risk_diabetes + frozen_meals +
##     gender + age + ethnicity + sleep_weekdays + hours_worked +
##     time_outdoors_weekends + time_sitting + take_away_food +
##     poverty_ratio + take.away.within.30.days + BMI
##
##                             Df Sum of Sq    RSS     AIC
## - sleep_weekdays             1    0.0001 82.973 -2015.2
## - take.away.within.30.days  1    0.0017 82.974 -2015.2
## - time_outdoors_weekends    1    0.0132 82.986 -2015.0
## - time_sitting               1    0.0133 82.986 -2015.0
## - hours_worked               1    0.0194 82.992 -2015.0
## - poverty_ratio              1    0.0411 83.014 -2014.8
## - moderate_activity          1    0.0692 83.042 -2014.5
## - take_away_food             1    0.0700 83.042 -2014.5
## - hypertension               1    0.0716 83.044 -2014.4
## - age                        1    0.0782 83.051 -2014.4
## - ethnicity                  1    0.1763 83.149 -2013.3
## <none>                                   82.972 -2013.2
## - gluten_free                1    0.3383 83.311 -2011.6
## - gender                     1    0.3558 83.328 -2011.5
## - frozen_meals               1    0.4844 83.457 -2010.1
## - diabetes                   1    0.5213 83.494 -2009.7
## - BMI                        1    0.6407 83.613 -2008.5
## - feel_at_risk_diabetes     1    5.9467 88.919 -1954.8
## - diabetes_relatives        1    7.5529 90.525 -1939.2
##
## Step:  AIC=-2015.19
## diabetes_risk ~ hypertension + gluten_free + diabetes_relatives +
##     moderate_activity + diabetes + feel_at_risk_diabetes + frozen_meals +
##     gender + age + ethnicity + hours_worked + time_outdoors_weekends +
##     time_sitting + take_away_food + poverty_ratio +
take.away.within.30.days +
##     BMI
##
##                             Df Sum of Sq    RSS     AIC
## - take.away.within.30.days  1    0.0018 82.974 -2017.2
## - time_outdoors_weekends    1    0.0131 82.986 -2017.0
## - time_sitting               1    0.0133 82.986 -2017.0
## - hours_worked               1    0.0194 82.992 -2017.0
## - poverty_ratio              1    0.0414 83.014 -2016.8
## - moderate_activity          1    0.0691 83.042 -2016.5
## - take_away_food             1    0.0699 83.042 -2016.5
## - hypertension               1    0.0715 83.044 -2016.4
## - age                        1    0.0785 83.051 -2016.4
```

```
## - ethnicity                 1    0.1786 83.151 -2015.3
## <none>                                 82.973 -2015.2
## - gluten_free               1    0.3382 83.311 -2013.6
## - gender                    1    0.3561 83.329 -2013.5
## - frozen_meals              1    0.4846 83.457 -2012.1
## - diabetes                  1    0.5213 83.494 -2011.7
## - BMI                       1    0.6480 83.621 -2010.4
## - feel_at_risk_diabetes     1    5.9467 88.919 -1956.8
## - diabetes_relatives        1    7.5535 90.526 -1941.2
##
## Step:  AIC=-2017.17
## diabetes_risk ~ hypertension + gluten_free + diabetes_relatives +
##     moderate_activity + diabetes + feel_at_risk_diabetes + frozen_meals +
##     gender + age + ethnicity + hours_worked + time_outdoors_weekends +
##     time_sitting + take_away_food + poverty_ratio + BMI
##
##                           Df Sum of Sq    RSS     AIC
## - time_outdoors_weekends  1    0.0127 82.987 -2019.0
## - time_sitting            1    0.0138 82.988 -2019.0
## - hours_worked            1    0.0198 82.994 -2019.0
## - poverty_ratio           1    0.0412 83.016 -2018.7
## - moderate_activity       1    0.0697 83.044 -2018.4
## - hypertension            1    0.0713 83.046 -2018.4
## - take_away_food          1    0.0756 83.050 -2018.4
## - age                     1    0.0781 83.052 -2018.3
## - ethnicity               1    0.1794 83.154 -2017.3
## <none>                                 82.974 -2017.2
## - gluten_free             1    0.3372 83.312 -2015.6
## - gender                  1    0.3567 83.331 -2015.4
## - frozen_meals            1    0.4829 83.457 -2014.1
## - diabetes                1    0.5202 83.495 -2013.7
## - BMI                     1    0.6490 83.623 -2012.4
## - feel_at_risk_diabetes   1    5.9644 88.939 -1958.6
## - diabetes_relatives      1    7.5627 90.537 -1943.1
##
## Step:  AIC=-2019.04
## diabetes_risk ~ hypertension + gluten_free + diabetes_relatives +
##     moderate_activity + diabetes + feel_at_risk_diabetes + frozen_meals +
##     gender + age + ethnicity + hours_worked + time_sitting +
##     take_away_food + poverty_ratio + BMI
##
##                          Df Sum of Sq    RSS     AIC
## - time_sitting           1    0.0138 83.001 -2020.9
## - hours_worked           1    0.0232 83.010 -2020.8
## - poverty_ratio          1    0.0399 83.027 -2020.6
## - hypertension           1    0.0745 83.062 -2020.2
## - take_away_food         1    0.0785 83.066 -2020.2
## - age                    1    0.0844 83.071 -2020.2
## - moderate_activity      1    0.0864 83.073 -2020.1
## - ethnicity              1    0.1874 83.174 -2019.1
```

```
## <none>                                82.987 -2019.0
## - gluten_free            1    0.3380 83.325 -2017.5
## - gender                 1    0.3451 83.332 -2017.4
## - frozen_meals           1    0.4828 83.470 -2016.0
## - diabetes               1    0.5201 83.507 -2015.6
## - BMI                    1    0.6507 83.638 -2014.2
## - feel_at_risk_diabetes  1    5.9707 88.958 -1960.5
## - diabetes_relatives     1    7.5513 90.538 -1945.1
##
## Step:  AIC=-2020.89
## diabetes_risk ~ hypertension + gluten_free + diabetes_relatives +
##     moderate_activity + diabetes + feel_at_risk_diabetes + frozen_meals +
##     gender + age + ethnicity + hours_worked + take_away_food +
##     poverty_ratio + BMI
##
##                          Df Sum of Sq    RSS     AIC
## - hours_worked            1    0.0226 83.024 -2022.7
## - poverty_ratio           1    0.0328 83.034 -2022.5
## - hypertension            1    0.0715 83.072 -2022.1
## - age                     1    0.0800 83.081 -2022.0
## - take_away_food          1    0.0873 83.088 -2022.0
## - moderate_activity       1    0.0917 83.093 -2021.9
## <none>                                83.001 -2020.9
## - ethnicity               1    0.2134 83.214 -2020.7
## - gluten_free             1    0.3403 83.341 -2019.3
## - gender                  1    0.3479 83.349 -2019.2
## - frozen_meals            1    0.4850 83.486 -2017.8
## - diabetes                1    0.5212 83.522 -2017.4
## - BMI                     1    0.6414 83.642 -2016.2
## - feel_at_risk_diabetes   1    5.9580 88.959 -1962.4
## - diabetes_relatives      1    7.5689 90.570 -1946.8
##
## Step:  AIC=-2022.65
## diabetes_risk ~ hypertension + gluten_free + diabetes_relatives +
##     moderate_activity + diabetes + feel_at_risk_diabetes + frozen_meals +
##     gender + age + ethnicity + take_away_food + poverty_ratio +
##     BMI
##
##                          Df Sum of Sq    RSS     AIC
## - poverty_ratio           1    0.0417 83.065 -2024.2
## - hypertension            1    0.0782 83.102 -2023.8
## - age                     1    0.0817 83.105 -2023.8
## - take_away_food          1    0.0820 83.105 -2023.8
## - moderate_activity       1    0.0956 83.119 -2023.7
## <none>                                83.024 -2022.7
## - ethnicity               1    0.2130 83.237 -2022.4
## - gender                  1    0.3286 83.352 -2021.2
## - gluten_free             1    0.3399 83.363 -2021.1
## - frozen_meals            1    0.4751 83.499 -2019.7
## - diabetes                1    0.5195 83.543 -2019.2
```

```
## - BMI                     1    0.6561 83.680 -2017.8
## - feel_at_risk_diabetes   1    5.9613 88.985 -1964.2
## - diabetes_relatives      1    7.5996 90.623 -1948.3
##
## Step:  AIC=-2024.21
## diabetes_risk ~ hypertension + gluten_free + diabetes_relatives +
##     moderate_activity + diabetes + feel_at_risk_diabetes + frozen_meals +
##     gender + age + ethnicity + take_away_food + BMI
##
##                          Df Sum of Sq    RSS     AIC
## - age                     1    0.0589 83.124 -2025.6
## - take_away_food          1    0.0650 83.130 -2025.5
## - hypertension            1    0.0759 83.141 -2025.4
## - moderate_activity       1    0.0854 83.151 -2025.3
## - ethnicity               1    0.1844 83.250 -2024.3
## <none>                                 83.065 -2024.2
## - gender                  1    0.3170 83.382 -2022.9
## - gluten_free             1    0.3623 83.427 -2022.4
## - frozen_meals            1    0.4773 83.543 -2021.2
## - diabetes                1    0.5267 83.592 -2020.7
## - BMI                     1    0.6312 83.696 -2019.6
## - feel_at_risk_diabetes   1    5.9477 89.013 -1965.9
## - diabetes_relatives      1    7.5640 90.629 -1950.2
##
## Step:  AIC=-2025.6
## diabetes_risk ~ hypertension + gluten_free + diabetes_relatives +
##     moderate_activity + diabetes + feel_at_risk_diabetes + frozen_meals +
##     gender + ethnicity + take_away_food + BMI
##
##                          Df Sum of Sq    RSS     AIC
## - hypertension            1    0.0486 83.173 -2027.1
## - take_away_food          1    0.0524 83.176 -2027.0
## - moderate_activity       1    0.0925 83.217 -2026.6
## <none>                                 83.124 -2025.6
## - ethnicity               1    0.1949 83.319 -2025.5
## - gender                  1    0.3140 83.438 -2024.3
## - gluten_free             1    0.3724 83.496 -2023.7
## - diabetes                1    0.5050 83.629 -2022.3
## - frozen_meals            1    0.5500 83.674 -2021.8
## - BMI                     1    0.6084 83.733 -2021.2
## - feel_at_risk_diabetes   1    5.9456 89.070 -1967.3
## - diabetes_relatives      1    7.5659 90.690 -1951.6
##
## Step:  AIC=-2027.09
## diabetes_risk ~ gluten_free + diabetes_relatives + moderate_activity +
##     diabetes + feel_at_risk_diabetes + frozen_meals + gender +
##     ethnicity + take_away_food + BMI
##
##                          Df Sum of Sq    RSS     AIC
## - take_away_food          1    0.0506 83.223 -2028.6
```

```
## - moderate_activity      1    0.1030 83.276 -2028.0
## - ethnicity              1    0.1885 83.361 -2027.1
## <none>                                83.173 -2027.1
## - gender                 1    0.3217 83.494 -2025.7
## - gluten_free            1    0.3707 83.543 -2025.2
## - diabetes               1    0.5114 83.684 -2023.7
## - frozen_meals           1    0.5713 83.744 -2023.1
## - BMI                    1    0.7116 83.884 -2021.7
## - feel_at_risk_diabetes  1    5.9583 89.131 -1968.8
## - diabetes_relatives     1    7.6438 90.816 -1952.4
##
## Step:  AIC=-2028.56
## diabetes_risk ~ gluten_free + diabetes_relatives + moderate_activity +
##     diabetes + feel_at_risk_diabetes + frozen_meals + gender +
##     ethnicity + BMI
##
##                          Df Sum of Sq    RSS     AIC
## - moderate_activity      1    0.1000 83.323 -2029.5
## - ethnicity              1    0.1910 83.414 -2028.6
## <none>                                83.223 -2028.6
## - gluten_free            1    0.3718 83.595 -2026.7
## - gender                 1    0.3823 83.606 -2026.6
## - diabetes               1    0.5261 83.749 -2025.1
## - frozen_meals           1    0.5431 83.766 -2024.9
## - BMI                    1    0.6994 83.923 -2023.3
## - feel_at_risk_diabetes  1    5.9314 89.155 -1970.5
## - diabetes_relatives     1    7.7731 90.996 -1952.7
##
## Step:  AIC=-2029.51
## diabetes_risk ~ gluten_free + diabetes_relatives + diabetes +
##     feel_at_risk_diabetes + frozen_meals + gender + ethnicity +
##     BMI
##
##                          Df Sum of Sq    RSS     AIC
## <none>                                83.323 -2029.5
## - ethnicity              1    0.1945 83.518 -2029.5
## - gender                 1    0.3359 83.659 -2028.0
## - gluten_free            1    0.3799 83.703 -2027.5
## - diabetes               1    0.5285 83.852 -2026.0
## - frozen_meals           1    0.5607 83.884 -2025.7
## - BMI                    1    0.6878 84.011 -2024.3
## - feel_at_risk_diabetes  1    6.0107 89.334 -1970.8
## - diabetes_relatives     1    7.7427 91.066 -1954.0
```

```
step$anova
```

```
## Stepwise Model Path
## Analysis of Deviance Table
##
## Initial Model:
```

```
## diabetes_risk ~ hypertension + gluten_free + stroke + diabetes_relatives +
##     moderate_activity + diabetes + feel_at_risk_diabetes + frozen_meals +
##     gender + age + ethnicity + education_level + sleep_weekdays +
##     hours_worked + time_outdoors_weekends + time_sitting + take_away_food
+
##     poverty_ratio + take.away.within.30.days + BMI
##
## Final Model:
## diabetes_risk ~ gluten_free + diabetes_relatives + diabetes +
##     feel_at_risk_diabetes + frozen_meals + gender + ethnicity +
##     BMI
##
##
##                          Step Df     Deviance Resid. Df Resid. Dev
AIC
## 1                                                  851   82.97246 -
2009.189
## 2                    - stroke  1 4.942858e-06       852   82.97246 -
2011.189
## 3           - education_level  1 1.091794e-05       853   82.97247 -
2013.189
## 4            - sleep_weekdays  1 8.951845e-05       854   82.97256 -
2015.188
## 5  - take.away.within.30.days  1 1.769339e-03       855   82.97433 -
2017.169
## 6    - time_outdoors_weekends  1 1.274661e-02       856   82.98708 -
2019.035
## 7              - time_sitting  1 1.378972e-02       857   83.00087 -
2020.890
## 8              - hours_worked  1 2.264919e-02       858   83.02352 -
2022.652
## 9             - poverty_ratio  1 4.167281e-02       859   83.06519 -
2024.215
## 10                      - age  1 5.887685e-02       860   83.12407 -
2025.597
## 11             - hypertension  1 4.864408e-02       861   83.17271 -
2027.087
## 12           - take_away_food  1 5.063528e-02       862   83.22335 -
2028.556
## 13        - moderate_activity  1 9.995600e-02       863   83.32330 -
2029.509
```

#Abby Keller HW3 Problem 1 R code #PCA Analysis

#Load the necessary packages

```r
library(DescTools, quietly = TRUE)
library(Hmisc, quietly = TRUE) #Describe Function
```

```
## Warning: package 'Hmisc' was built under R version 4.0.3

## Warning: package 'survival' was built under R version 4.0.3

## Warning: package 'Formula' was built under R version 4.0.3

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:DescTools':
##
##     %nin%, Label, Mean, Quantile

## The following objects are masked from 'package:base':
##
##     format.pval, units
```

```
library(psych, quietly = TRUE) #Multiple Functions for Statistics and
Multivariate Analysis
```

```
## Warning: package 'psych' was built under R version 4.0.3

##
## Attaching package: 'psych'

## The following object is masked from 'package:Hmisc':
##
##     describe

## The following objects are masked from 'package:DescTools':
##
##     AUC, ICC, SD

## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
```

```
library(GGally, quietly = TRUE) #ggpairs Function
library(ggplot2, quietly = TRUE) #ggplot2 Functions
library(vioplot, quietly = TRUE) #Violin Plot Function
```

```
## Warning: package 'vioplot' was built under R version 4.0.3

## Warning: package 'sm' was built under R version 4.0.3

## Package 'sm', version 2.2-5.6: type help(sm) for summary information

##
## Attaching package: 'sm'

## The following object is masked from 'package:MASS':
##
##     muscle
```

```
## Warning: package 'zoo' was built under R version 4.0.3

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

library(corrplot, quietly = TRUE) #Plot Correlations
library(REdaS, quietly = TRUE) #Bartlett's Test of Sphericity

## Warning: package 'REdaS' was built under R version 4.0.3

library(psych, quietly = TRUE) #PCA/FA functions
library(factoextra, quietly = TRUE) #PCA Visualizations

## Warning: package 'factoextra' was built under R version 4.0.3

## Welcome! Want to learn more? See two factoextra-related books at
https://goo.gl/ve3WBa

library("FactoMineR", quietly = TRUE) #PCA functions

## Warning: package 'FactoMineR' was built under R version 4.0.3

library(ade4, quietly = TRUE) #PCA Visualizations

## Warning: package 'ade4' was built under R version 4.0.3

##
## Attaching package: 'ade4'

## The following object is masked from 'package:FactoMineR':
##
##     reconst

library(RColorBrewer, quietly = TRUE) #for brewer colors
```

Set the working directory

```
#setwd("~/Documents/DePaul MS DS/DePaul Winter 2021/DSC 424/Project_424")
setwd("C:/Users/rejalu1/OneDrive - Henry Ford Health
System/DSC424/HomeWork3/")
```

Read in the nhanes4 dataset from the book by Hair, et. al.

```
nhanes <- read.csv(file="../HomeWork3/datasets/nhanes4.csv"
                   , header=TRUE
                   , sep=",")
```

#Check the meta data

```
str(nhanes)
```

```
## 'data.frame':    872 obs. of  25 variables:
##  $ ï..ID                   : int  73568 73585 73589 73592 73619 73642
73655 73658 73660 73688 ...
##  $ diabetes_risk           : int  1 1 1 1 1 1 1 1 0 ...
##  $ hypertension            : int  1 1 1 1 1 1 0 1 1 1 ...
##  $ gluten_free             : int  1 0 1 0 1 1 1 1 1 0 ...
##  $ stroke                  : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ diabetes_relatives      : int  1 1 1 1 1 1 0 1 1 0 ...
##  $ moderate_activity       : int  30 20 120 120 60 30 30 20 30 120 ...
##  $ diabetes                : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ feel_at_risk_diabetes   : int  1 1 1 1 0 1 1 0 1 0 ...
##  $ frozen_meals            : int  0 2 10 0 1 0 0 1 0 0 ...
##  $ gender                  : int  1 0 0 0 1 1 0 0 0 0 ...
##  $ age                     : int  26 28 35 29 36 57 44 35 59 58 ...
##  $ ethnicity               : int  3 5 3 1 1 3 4 5 5 3 ...
##  $ education_level         : int  5 4 3 4 3 3 3 5 5 3 ...
##  $ annual_income           : int  15 7 5 12 3 8 8 15 7 10 ...
##  $ height                  : int  60 69 67 69 69 63 76 68 68 69 ...
##  $ weight                  : int  105 205 165 170 185 217 250 160 163 195
...
##  $ most_weighed            : int  120 250 174 220 192 245 250 160 175 215
...
##  $ sleep_weekdays          : int  8 6 7 8 7 8 8 5 7 7 ...
##  $ hours_worked            : int  40 40 42 42 21 45 40 55 20 50 ...
##  $ time_outdoors_weekends  : int  120 120 240 360 30 60 120 180 120 120
...
##  $ time_sitting            : int  600 480 300 720 120 120 600 660 660 240
...
##  $ take_away_food          : int  3 1 5 7 10 0 2 6 0 7 ...
##  $ poverty_ratio           : num  5 2.26 1.74 2.14 0.84 2.27 1.79 5 1.82
4.45 ...
##  $ take.away.within.30.days: int  0 0 21 3 0 0 2 2 0 0 ...
```

#View the top 6 rows

```
head(nhanes)
```

```
##   ï..ID diabetes_risk hypertension gluten_free stroke diabetes_relatives
## 1 73568             1            1           1      1                  1
## 2 73585             1            1           0      1                  1
## 3 73589             1            1           1      1                  1
## 4 73592             1            1           0      1                  1
## 5 73619             1            1           1      1                  1
## 6 73642             1            1           1      1                  1
##   moderate_activity diabetes feel_at_risk_diabetes frozen_meals gender age
## 1                30        1                     1            0      1  26
## 2                20        1                     1            2      0  28
```

```
## 3               120        1                    1           10      0  35
## 4               120        1                    1            0      0  29
## 5                60        1                    0            1      1  36
## 6                30        1                    1            0      1  57
##   ethnicity education_level annual_income height weight most_weighed
## 1         3               5            15     60    105          120
## 2         5               4             7     69    205          250
## 3         3               3             5     67    165          174
## 4         1               4            12     69    170          220
## 5         1               3             3     69    185          192
## 6         3               3             8     63    217          245
##   sleep_weekdays hours_worked time_outdoors_weekends time_sitting
## 1              8           40                    120          600
## 2              6           40                    120          480
## 3              7           42                    240          300
## 4              8           42                    360          720
## 5              7           21                     30          120
## 6              8           45                     60          120
##   take_away_food poverty_ratio take.away.within.30.days
## 1              3          5.00                       0
## 2              1          2.26                       0
## 3              5          1.74                      21
## 4              7          2.14                       3
## 5             10          0.84                       0
## 6              0          2.27                       0
```

#Check frequencies and missing values for all variables or a specific variable

```
describe(nhanes)
```

```
##                          vars   n     mean      sd   median  trimmed
mad
## ï..ID                       1 872 78757.14 2835.29 78900.50 78780.85
3417.39
## diabetes_risk               2 872     0.85    0.36     1.00     0.94
0.00
## hypertension                3 872     0.81    0.39     1.00     0.89
0.00
## gluten_free                 4 872     0.97    0.18     1.00     1.00
0.00
## stroke                      5 872     1.00    0.07     1.00     1.00
0.00
## diabetes_relatives          6 872     0.66    0.47     1.00     0.70
0.00
## moderate_activity           7 872    64.46   59.07    60.00    53.61
44.48
## diabetes                    8 872     1.03    0.24     1.00     1.00
0.00
## feel_at_risk_diabetes       9 872     0.69    0.46     1.00     0.74
0.00
```

```
## frozen_meals                10 872     2.05    5.18     0.00     0.80
0.00
## gender                      11 872     0.49    0.50     0.00     0.49
0.00
## age                         12 872    39.27   11.05    40.00    39.19
13.34
## ethnicity                   13 872     3.26    1.19     3.00     3.33
1.48
## education_level             14 872     4.08    0.99     4.00     4.21
1.48
## annual_income               15 872    10.39    4.37    10.00    10.73
7.41
## height                      16 872    66.92    3.96    67.00    66.90
4.45
## weight                      17 872   173.76   40.40   170.00   171.02
41.51
## most_weighed                18 872   188.33   46.63   180.00   184.96
44.48
## sleep_weekdays              19 872     6.76    1.17     7.00     6.83
1.48
## hours_worked                20 872    40.37   13.46    40.00    40.51
7.41
## time_outdoors_weekends      21 872   170.43  116.39   120.00   156.19
88.96
## time_sitting                22 872   423.68  200.45   480.00   419.05
177.91
## take_away_food              23 872     4.18    4.20     3.00     3.45
2.97
## poverty_ratio               24 872     3.12    1.65     3.22     3.21
2.64
## take.away.within.30.days    25 872     2.42    7.96     0.00     0.83
0.00
##                            min   max range   skew kurtosis    se
## ï..ID                    73568 83729 10161  -0.06    -1.10 96.02
## diabetes_risk                0     1     1  -1.95     1.82  0.01
## hypertension                 0     1     1  -1.57     0.48  0.01
## gluten_free                  0     1     1  -5.10    24.04  0.01
## stroke                       0     1     1 -14.64   212.51  0.00
## diabetes_relatives           0     1     1  -0.66    -1.57  0.02
## moderate_activity           10   600   590   3.26    16.39  2.00
## diabetes                     1     3     2   7.99    61.94  0.01
## feel_at_risk_diabetes        0     1     1  -0.84    -1.30  0.02
## frozen_meals                 0    60    60   5.29    41.86  0.18
## gender                       0     1     1   0.04    -2.00  0.02
## age                         20    59    39   0.04    -1.11  0.37
## ethnicity                    1     5     4  -0.19    -0.55  0.04
## education_level              1     5     4  -0.97     0.40  0.03
## annual_income                1    15    14  -0.27    -1.40  0.15
## height                      48    79    31  -0.04     0.11  0.13
## weight                      92   380   288   0.81     1.36  1.37
```

```
## most_weighed                98   430   332   0.86     1.41  1.58
## sleep_weekdays               2    10     8  -0.44     0.42  0.04
## hours_worked                 5   120   115   0.18     2.22  0.46
## time_outdoors_weekends      14   480   466   0.98     0.36  3.94
## time_sitting                30  1080  1050   0.19    -0.57  6.79
## take_away_food               0    21    21   1.80     3.51  0.14
## poverty_ratio                0     5     5  -0.20    -1.46  0.06
## take.away.within.30.days     0   150   150   9.86   147.56  0.27
```

#Check for specific values missing

```
sum(is.na(nhanes$x6)) # returns sum of TRUEs, if a specific value of x6
variable is missing

## [1] 0

sum(is.na(nhanes))

## [1] 0
```

split the data set

```
training <- na.omit(nhanes)
training = nhanes[, c(2:25)]
training = nhanes[, c(2:12, 16, 17, 19:25)]
training = nhanes[, c(2:12, 14, 16, 17, 19:25)]
#training = nhanes[, c(10, 12, 16, 17, 19:24)]
```

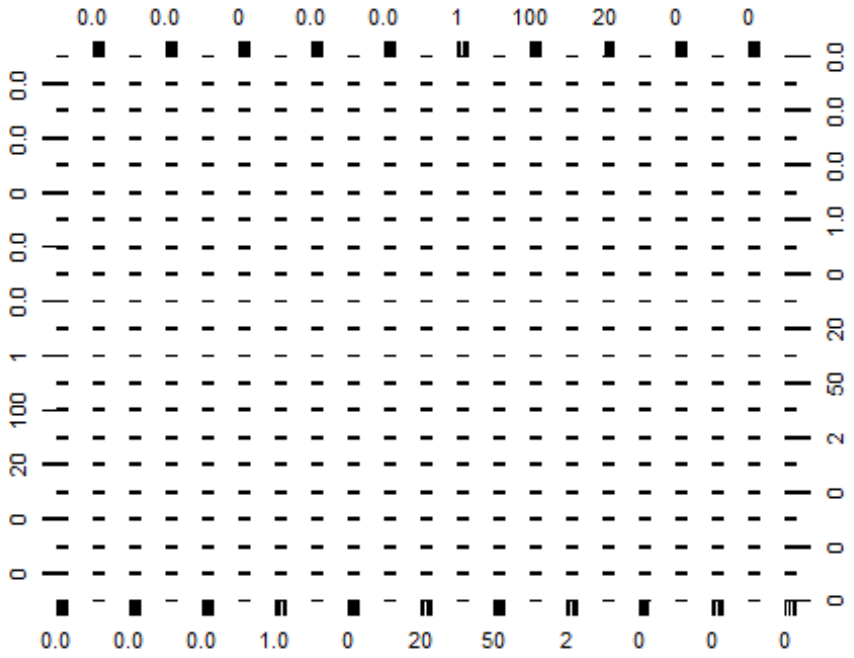
view the top 6 rows

```
head(training)

##   diabetes_risk hypertension gluten_free stroke diabetes_relatives
## 1             1            1           1      1                  1
## 2             1            1           0      1                  1
## 3             1            1           1      1                  1
## 4             1            1           0      1                  1
## 5             1            1           1      1                  1
## 6             1            1           1      1                  1
##   moderate_activity diabetes feel_at_risk_diabetes frozen_meals gender age
## 1                30        1                     1            0      1  26
## 2                20        1                     1            2      0  28
## 3               120        1                     1           10      0  35
## 4               120        1                     1            0      0  29
## 5                60        1                     0            1      1  36
## 6                30        1                     1            0      1  57
##   education_level height weight sleep_weekdays hours_worked
## 1               5     60    105              8           40
## 2               4     69    205              6           40
## 3               3     67    165              7           42
```

```
## 4                4     69    170              8           42
## 5                3     69    185              7           21
## 6                3     63    217              8           45
##   time_outdoors_weekends time_sitting take_away_food poverty_ratio
## 1                    120          600              3          5.00
## 2                    120          480              1          2.26
## 3                    240          300              5          1.74
## 4                    360          720              7          2.14
## 5                     30          120             10          0.84
## 6                     60          120              0          2.27
##   take.away.within.30.days
## 1                        0
## 2                        0
## 3                       21
## 4                        3
## 5                        0
## 6                        0
```

plot scatter plot matrix

```
plot(training)
```



#Check data types

```
str(training)
```

```
## 'data.frame':    872 obs. of  21 variables:
##  $ diabetes_risk           : int  1 1 1 1 1 1 1 1 0 ...
##  $ hypertension            : int  1 1 1 1 1 1 0 1 1 1 ...
##  $ gluten_free             : int  1 0 1 0 1 1 1 1 1 0 ...
##  $ stroke                  : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ diabetes_relatives      : int  1 1 1 1 1 1 0 1 1 0 ...
##  $ moderate_activity       : int  30 20 120 120 60 30 30 20 30 120 ...
##  $ diabetes                : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ feel_at_risk_diabetes   : int  1 1 1 1 0 1 1 0 1 0 ...
##  $ frozen_meals            : int  0 2 10 0 1 0 0 1 0 0 ...
##  $ gender                  : int  1 0 0 0 1 1 0 0 0 0 ...
##  $ age                     : int  26 28 35 29 36 57 44 35 59 58 ...
##  $ education_level         : int  5 4 3 4 3 3 3 5 5 3 ...
##  $ height                  : int  60 69 67 69 69 63 76 68 68 69 ...
##  $ weight                  : int  105 205 165 170 185 217 250 160 163 195
...
##  $ sleep_weekdays          : int  8 6 7 8 7 8 8 5 7 7 ...
##  $ hours_worked            : int  40 40 42 42 21 45 40 55 20 50 ...
##  $ time_outdoors_weekends  : int  120 120 240 360 30 60 120 180 120 120
...
##  $ time_sitting            : int  600 480 300 720 120 120 600 660 660 240
...
##  $ take_away_food          : int  3 1 5 7 10 0 2 6 0 7 ...
##  $ poverty_ratio           : num  5 2.26 1.74 2.14 0.84 2.27 1.79 5 1.82
4.45 ...
##  $ take.away.within.30.days: int  0 0 21 3 0 0 2 2 0 0 ...
```

correlations

```
cor(training)
```

```
##                          diabetes_risk hypertension   gluten_free
stroke
## diabetes_risk             1.000000000  0.090436992  0.0615132218 -
0.0285428067
## hypertension              0.090436992  1.000000000 -0.0113951126
0.1399969066
## gluten_free               0.061513222 -0.011395113  1.0000000000 -
0.0128137048
## stroke                   -0.028542807  0.139996907 -0.0128137048
1.0000000000
## diabetes_relatives        0.392731801  0.080494103  0.0094549842 -
0.0490375107
## moderate_activity        -0.031835550 -0.071961362 -0.0129130166 -
0.0437370394
## diabetes                 -0.107186144 -0.036764358  0.0232209336
0.0083511297
## feel_at_risk_diabetes     0.396154774  0.083479358  0.0111149716 -
0.0450970441
## frozen_meals             -0.074087409 -0.069818003  0.0152227368
```

```
0.0203494126
## gender                     -0.069677113  0.019205510 -0.0542433071 -
0.0014015826
## age                        -0.007915844 -0.256933717  0.0331410204 -
0.0935696239
## education_level             0.058137715  0.007573236 -0.0428094360
0.0395129927
## height                      0.046300803 -0.038801289  0.0536294961
0.0115676008
## weight                     -0.142654851 -0.209344654  0.0411259953
0.0197776476
## sleep_weekdays              0.038292241  0.113132386 -0.0017584288 -
0.0430516962
## hours_worked               -0.020219801 -0.095485329  0.0028256882 -
0.0006652705
## time_outdoors_weekends     -0.016302149 -0.052429020 -0.0008726455
0.0206759969
## time_sitting                0.020223119 -0.051456224  0.0053529889
0.0571595815
## take_away_food              0.042177592 -0.031544236  0.0157896563 -
0.0010566200
## poverty_ratio               0.041587392 -0.019764767 -0.0648712937
0.0490310086
## take.away.within.30.days    0.002877079 -0.018033926 -0.0208501788
0.0035898193
##                          diabetes_relatives moderate_activity
diabetes
## diabetes_risk                   0.392731801      -0.031835550 -
0.1071861442
## hypertension                    0.080494103      -0.071961362 -
0.0367643582
## gluten_free                     0.009454984      -0.012913017
0.0232209336
## stroke                         -0.049037511      -0.043737039
0.0083511297
## diabetes_relatives              1.000000000       0.009568842 -
0.0506855928
## moderate_activity               0.009568842       1.000000000
0.0067329042
## diabetes                       -0.050685593       0.006732904
1.0000000000
## feel_at_risk_diabetes           0.369267734      -0.027135791 -
0.0825251735
## frozen_meals                    0.039744098       0.035277088 -
0.0066889062
## gender                         -0.022163817      -0.135406815 -
0.0069251388
## age                            -0.046230078      -0.039282485
0.0860474862
## education_level                 0.106576676      -0.077169470 -
```

```
0.0189820852
## height                          0.032582968       0.122291702
0.0214838020
## weight                         -0.105673395       0.045252728
0.0679719674
## sleep_weekdays                  0.047275383      -0.035829809 -
0.0147121912
## hours_worked                   -0.028353387       0.058446792
0.0022603991
## time_outdoors_weekends          0.020483716       0.239109169 -
0.0045212160
## time_sitting                    0.019424403      -0.067940765
0.0005736717
## take_away_food                  0.069726673       0.051162178 -
0.0392082121
## poverty_ratio                   0.063959798      -0.082340664
0.0286538545
## take.away.within.30.days       -0.021904478      -0.010243428
0.0125186678
##                          feel_at_risk_diabetes frozen_meals       gender
## diabetes_risk                        0.396154774 -0.074087409 -0.069677113
## hypertension                         0.083479358 -0.069818003  0.019205510
## gluten_free                          0.011114972  0.015222737 -0.054243307
## stroke                              -0.045097044  0.020349413 -0.001401583
## diabetes_relatives                   0.369267734  0.039744098 -0.022163817
## moderate_activity                   -0.027135791  0.035277088 -0.135406815
## diabetes                            -0.082525174 -0.006688906 -0.006925139
## feel_at_risk_diabetes                1.000000000 -0.034889204 -0.075931561
## frozen_meals                        -0.034889204  1.000000000  0.013507463
## gender                              -0.075931561  0.013507463  1.000000000
## age                                 -0.034037281 -0.153016067  0.017538292
## education_level                      0.045806652  0.005306166  0.031733105
## height                               0.071546435  0.018508267 -0.654319198
## weight                              -0.174231487  0.045859507 -0.332371361
## sleep_weekdays                       0.041937838  0.008831964 -0.001365208
## hours_worked                        -0.010670401 -0.041802867 -0.198461173
## time_outdoors_weekends              -0.006576999  0.019592958 -0.177300996
## time_sitting                        -0.016110151  0.003340262 -0.052697811
## take_away_food                       0.003138257  0.102627552 -0.192740111
## poverty_ratio                        0.061879957 -0.031368967 -0.045973080
## take.away.within.30.days             0.023875789  0.090832169 -0.042288767
##                                      age education_level        height
## diabetes_risk            -0.0079158437     0.058137715  0.0463008031
## hypertension             -0.2569337167     0.007573236 -0.0388012886
## gluten_free               0.0331410204    -0.042809436  0.0536294961
## stroke                   -0.0935696239     0.039512993  0.0115676008
## diabetes_relatives       -0.0462300783     0.106576676  0.0325829680
## moderate_activity        -0.0392824853    -0.077169470  0.1222917025
## diabetes                  0.0860474862    -0.018982085  0.0214838020
## feel_at_risk_diabetes    -0.0340372812     0.045806652  0.0715464350
```

```
## frozen_meals              -0.1530160671     0.005306166  0.0185082668
## gender                     0.0175382924     0.031733105 -0.6543191977
## age                        1.0000000000     0.037783357 -0.0009177865
## education_level            0.0377833572     1.000000000  0.1242337371
## height                    -0.0009177865     0.124233737  1.0000000000
## weight                     0.0950790520    -0.048969951  0.5096989759
## sleep_weekdays            -0.0648257873     0.041616202 -0.0526277909
## hours_worked               0.0352481936     0.020136280  0.1861108710
## time_outdoors_weekends    -0.0912503212    -0.057627635  0.1422866669
## time_sitting              -0.0121442337     0.243246800  0.0500421298
## take_away_food            -0.1206496639     0.102931634  0.1838445087
## poverty_ratio              0.2245067391     0.488618938  0.1349542490
## take.away.within.30.days -0.0474925630     0.007608104  0.0177975380
##                                 weight sleep_weekdays  hours_worked
## diabetes_risk             -0.142654851    0.038292241 -0.0202198009
## hypertension              -0.209344654    0.113132386 -0.0954853287
## gluten_free                0.041125995   -0.001758429  0.0028256882
## stroke                     0.019777648   -0.043051696 -0.0006652705
## diabetes_relatives        -0.105673395    0.047275383 -0.0283533868
## moderate_activity          0.045252728   -0.035829809  0.0584467924
## diabetes                   0.067971967   -0.014712191  0.0022603991
## feel_at_risk_diabetes     -0.174231487    0.041937838 -0.0106704015
## frozen_meals               0.045859507    0.008831964 -0.0418028666
## gender                    -0.332371361   -0.001365208 -0.1984611729
## age                        0.095079052   -0.064825787  0.0352481936
## education_level           -0.048969951    0.041616202  0.0201362798
## height                     0.509698976   -0.052627791  0.1861108710
## weight                     1.000000000   -0.138974733  0.1314184231
## sleep_weekdays            -0.138974733    1.000000000 -0.1040991923
## hours_worked               0.131418423   -0.104099192  1.0000000000
## time_outdoors_weekends     0.084074807   -0.103787334  0.1233039916
## time_sitting               0.054325677    0.042603526  0.0696603990
## take_away_food             0.106267394    0.041028361  0.1207308004
## poverty_ratio             -0.039892225    0.050828638  0.1508072756
## take.away.within.30.days   0.003098214   -0.016004933 -0.0061899332
##                          time_outdoors_weekends  time_sitting
take_away_food
## diabetes_risk                     -0.0163021489  0.0202231190
0.042177592
## hypertension                      -0.0524290199 -0.0514562237   -
0.031544236
## gluten_free                       -0.0008726455  0.0053529889
0.015789656
## stroke                             0.0206759969  0.0571595815   -
0.001056620
## diabetes_relatives                 0.0204837157  0.0194244034
0.069726673
## moderate_activity                  0.2391091685 -0.0679407654
0.051162178
## diabetes                          -0.0045212160  0.0005736717   -
```

```
0.039208212
## feel_at_risk_diabetes             -0.0065769993 -0.0161101507
0.003138257
## frozen_meals                       0.0195929583  0.0033402617
0.102627552
## gender                            -0.1773009959 -0.0526978113   -
0.192740111
## age                               -0.0912503212 -0.0121442337   -
0.120649664
## education_level                   -0.0576276353  0.2432467996
0.102931634
## height                             0.1422866669  0.0500421298
0.183844509
## weight                             0.0840748065  0.0543256768
0.106267394
## sleep_weekdays                    -0.1037873345  0.0426035261
0.041028361
## hours_worked                       0.1233039916  0.0696603990
0.120730800
## time_outdoors_weekends             1.0000000000 -0.0247290651
0.013988851
## time_sitting                      -0.0247290651  1.0000000000
0.156956732
## take_away_food                     0.0139888511  0.1569567321
1.000000000
## poverty_ratio                     -0.0689465337  0.2312651439
0.136838530
## take.away.within.30.days           0.0276676684  0.0846926782
0.182265458
##                           poverty_ratio take.away.within.30.days
## diabetes_risk               0.04158739              0.002877079
## hypertension               -0.01976477             -0.018033926
## gluten_free                -0.06487129             -0.020850179
## stroke                      0.04903101              0.003589819
## diabetes_relatives          0.06395980             -0.021904478
## moderate_activity          -0.08234066             -0.010243428
## diabetes                    0.02865385              0.012518668
## feel_at_risk_diabetes       0.06187996              0.023875789
## frozen_meals               -0.03136897              0.090832169
## gender                     -0.04597308             -0.042288767
## age                         0.22450674             -0.047492563
## education_level             0.48861894              0.007608104
## height                      0.13495425              0.017797538
## weight                     -0.03989223              0.003098214
## sleep_weekdays              0.05082864             -0.016004933
## hours_worked                0.15080728             -0.006189933
## time_outdoors_weekends     -0.06894653              0.027667668
## time_sitting                0.23126514              0.084692678
## take_away_food              0.13683853              0.182265458
```

```
## poverty_ratio                1.00000000              0.044973865
## take.away.within.30.days     0.04497386              1.000000000
```

```
cor(training, use = 'pairwise.complete.obs')
```

```
##                           diabetes_risk hypertension   gluten_free
stroke
## diabetes_risk              1.000000000  0.090436992  0.0615132218 -
0.0285428067
## hypertension               0.090436992  1.000000000 -0.0113951126
0.1399969066
## gluten_free                0.061513222 -0.011395113  1.0000000000 -
0.0128137048
## stroke                    -0.028542807  0.139996907 -0.0128137048
1.0000000000
## diabetes_relatives         0.392731801  0.080494103  0.0094549842 -
0.0490375107
## moderate_activity         -0.031835550 -0.071961362 -0.0129130166 -
0.0437370394
## diabetes                  -0.107186144 -0.036764358  0.0232209336
0.0083511297
## feel_at_risk_diabetes      0.396154774  0.083479358  0.0111149716 -
0.0450970441
## frozen_meals              -0.074087409 -0.069818003  0.0152227368
0.0203494126
## gender                    -0.069677113  0.019205510 -0.0542433071 -
0.0014015826
## age                       -0.007915844 -0.256933717  0.0331410204 -
0.0935696239
## education_level            0.058137715  0.007573236 -0.0428094360
0.0395129927
## height                     0.046300803 -0.038801289  0.0536294961
0.0115676008
## weight                    -0.142654851 -0.209344654  0.0411259953
0.0197776476
## sleep_weekdays             0.038292241  0.113132386 -0.0017584288 -
0.0430516962
## hours_worked              -0.020219801 -0.095485329  0.0028256882 -
0.0006652705
## time_outdoors_weekends    -0.016302149 -0.052429020 -0.0008726455
0.0206759969
## time_sitting               0.020223119 -0.051456224  0.0053529889
0.0571595815
## take_away_food             0.042177592 -0.031544236  0.0157896563 -
0.0010566200
## poverty_ratio              0.041587392 -0.019764767 -0.0648712937
0.0490310086
## take.away.within.30.days   0.002877079 -0.018033926 -0.0208501788
0.0035898193
##                           diabetes_relatives moderate_activity
```

```
diabetes
## diabetes_risk                   0.392731801      -0.031835550 -
0.1071861442
## hypertension                    0.080494103      -0.071961362 -
0.0367643582
## gluten_free                     0.009454984      -0.012913017
0.0232209336
## stroke                         -0.049037511      -0.043737039
0.0083511297
## diabetes_relatives              1.000000000       0.009568842 -
0.0506855928
## moderate_activity               0.009568842       1.000000000
0.0067329042
## diabetes                       -0.050685593       0.006732904
1.0000000000
## feel_at_risk_diabetes           0.369267734      -0.027135791 -
0.0825251735
## frozen_meals                    0.039744098       0.035277088 -
0.0066889062
## gender                         -0.022163817      -0.135406815 -
0.0069251388
## age                            -0.046230078      -0.039282485
0.0860474862
## education_level                 0.106576676      -0.077169470 -
0.0189820852
## height                          0.032582968       0.122291702
0.0214838020
## weight                         -0.105673395       0.045252728
0.0679719674
## sleep_weekdays                  0.047275383      -0.035829809 -
0.0147121912
## hours_worked                   -0.028353387       0.058446792
0.0022603991
## time_outdoors_weekends          0.020483716       0.239109169 -
0.0045212160
## time_sitting                    0.019424403      -0.067940765
0.0005736717
## take_away_food                  0.069726673       0.051162178 -
0.0392082121
## poverty_ratio                   0.063959798      -0.082340664
0.0286538545
## take.away.within.30.days       -0.021904478      -0.010243428
0.0125186678
##                          feel_at_risk_diabetes frozen_meals       gender
## diabetes_risk                        0.396154774 -0.074087409 -0.069677113
## hypertension                         0.083479358 -0.069818003  0.019205510
## gluten_free                          0.011114972  0.015222737 -0.054243307
## stroke                              -0.045097044  0.020349413 -0.001401583
## diabetes_relatives                   0.369267734  0.039744098 -0.022163817
## moderate_activity                   -0.027135791  0.035277088 -0.135406815
```

```
## diabetes                          -0.082525174 -0.006688906 -0.006925139
## feel_at_risk_diabetes              1.000000000 -0.034889204 -0.075931561
## frozen_meals                      -0.034889204  1.000000000  0.013507463
## gender                            -0.075931561  0.013507463  1.000000000
## age                               -0.034037281 -0.153016067  0.017538292
## education_level                    0.045806652  0.005306166  0.031733105
## height                             0.071546435  0.018508267 -0.654319198
## weight                            -0.174231487  0.045859507 -0.332371361
## sleep_weekdays                     0.041937838  0.008831964 -0.001365208
## hours_worked                      -0.010670401 -0.041802867 -0.198461173
## time_outdoors_weekends            -0.006576999  0.019592958 -0.177300996
## time_sitting                      -0.016110151  0.003340262 -0.052697811
## take_away_food                     0.003138257  0.102627552 -0.192740111
## poverty_ratio                      0.061879957 -0.031368967 -0.045973080
## take.away.within.30.days           0.023875789  0.090832169 -0.042288767
##                                             age education_level        height
## diabetes_risk             -0.0079158437     0.058137715  0.0463008031
## hypertension              -0.2569337167     0.007573236 -0.0388012886
## gluten_free                0.0331410204    -0.042809436  0.0536294961
## stroke                    -0.0935696239     0.039512993  0.0115676008
## diabetes_relatives        -0.0462300783     0.106576676  0.0325829680
## moderate_activity         -0.0392824853    -0.077169470  0.1222917025
## diabetes                   0.0860474862    -0.018982085  0.0214838020
## feel_at_risk_diabetes     -0.0340372812     0.045806652  0.0715464350
## frozen_meals              -0.1530160671     0.005306166  0.0185082668
## gender                     0.0175382924     0.031733105 -0.6543191977
## age                        1.0000000000     0.037783357 -0.0009177865
## education_level            0.0377833572     1.000000000  0.1242337371
## height                    -0.0009177865     0.124233737  1.0000000000
## weight                     0.0950790520    -0.048969951  0.5096989759
## sleep_weekdays            -0.0648257873     0.041616202 -0.0526277909
## hours_worked               0.0352481936     0.020136280  0.1861108710
## time_outdoors_weekends    -0.0912503212    -0.057627635  0.1422866669
## time_sitting              -0.0121442337     0.243246800  0.0500421298
## take_away_food            -0.1206496639     0.102931634  0.1838445087
## poverty_ratio              0.2245067391     0.488618938  0.1349542490
## take.away.within.30.days -0.0474925630     0.007608104  0.0177975380
##                                   weight sleep_weekdays  hours_worked
## diabetes_risk             -0.142654851    0.038292241 -0.0202198009
## hypertension              -0.209344654    0.113132386 -0.0954853287
## gluten_free                0.041125995   -0.001758429  0.0028256882
## stroke                     0.019777648   -0.043051696 -0.0006652705
## diabetes_relatives        -0.105673395    0.047275383 -0.0283533868
## moderate_activity          0.045252728   -0.035829809  0.0584467924
## diabetes                   0.067971967   -0.014712191  0.0022603991
## feel_at_risk_diabetes     -0.174231487    0.041937838 -0.0106704015
## frozen_meals               0.045859507    0.008831964 -0.0418028666
## gender                    -0.332371361   -0.001365208 -0.1984611729
## age                        0.095079052   -0.064825787  0.0352481936
## education_level           -0.048969951    0.041616202  0.0201362798
```

```
## height                     0.509698976   -0.052627791  0.1861108710
## weight                     1.000000000   -0.138974733  0.1314184231
## sleep_weekdays            -0.138974733    1.000000000 -0.1040991923
## hours_worked               0.131418423   -0.104099192  1.0000000000
## time_outdoors_weekends     0.084074807   -0.103787334  0.1233039916
## time_sitting               0.054325677    0.042603526  0.0696603990
## take_away_food             0.106267394    0.041028361  0.1207308004
## poverty_ratio             -0.039892225    0.050828638  0.1508072756
## take.away.within.30.days   0.003098214   -0.016004933 -0.0061899332
##                          time_outdoors_weekends  time_sitting
take_away_food
## diabetes_risk                      -0.0163021489  0.0202231190
0.042177592
## hypertension                       -0.0524290199 -0.0514562237   -
0.031544236
## gluten_free                        -0.0008726455  0.0053529889
0.015789656
## stroke                              0.0206759969  0.0571595815   -
0.001056620
## diabetes_relatives                  0.0204837157  0.0194244034
0.069726673
## moderate_activity                   0.2391091685 -0.0679407654
0.051162178
## diabetes                           -0.0045212160  0.0005736717   -
0.039208212
## feel_at_risk_diabetes              -0.0065769993 -0.0161101507
0.003138257
## frozen_meals                        0.0195929583  0.0033402617
0.102627552
## gender                             -0.1773009959 -0.0526978113   -
0.192740111
## age                                -0.0912503212 -0.0121442337   -
0.120649664
## education_level                    -0.0576276353  0.2432467996
0.102931634
## height                              0.1422866669  0.0500421298
0.183844509
## weight                              0.0840748065  0.0543256768
0.106267394
## sleep_weekdays                     -0.1037873345  0.0426035261
0.041028361
## hours_worked                        0.1233039916  0.0696603990
0.120730800
## time_outdoors_weekends              1.0000000000 -0.0247290651
0.013988851
## time_sitting                       -0.0247290651  1.0000000000
0.156956732
## take_away_food                      0.0139888511  0.1569567321
1.000000000
## poverty_ratio                      -0.0689465337  0.2312651439
```

```
0.136838530
## take.away.within.30.days           0.0276676684  0.0846926782
0.182265458
##                            poverty_ratio take.away.within.30.days
## diabetes_risk               0.04158739              0.002877079
## hypertension               -0.01976477             -0.018033926
## gluten_free                -0.06487129             -0.020850179
## stroke                      0.04903101              0.003589819
## diabetes_relatives          0.06395980             -0.021904478
## moderate_activity          -0.08234066             -0.010243428
## diabetes                    0.02865385              0.012518668
## feel_at_risk_diabetes       0.06187996              0.023875789
## frozen_meals               -0.03136897              0.090832169
## gender                     -0.04597308             -0.042288767
## age                         0.22450674             -0.047492563
## education_level             0.48861894              0.007608104
## height                      0.13495425              0.017797538
## weight                     -0.03989223              0.003098214
## sleep_weekdays              0.05082864             -0.016004933
## hours_worked                0.15080728             -0.006189933
## time_outdoors_weekends     -0.06894653              0.027667668
## time_sitting                0.23126514              0.084692678
## take_away_food              0.13683853              0.182265458
## poverty_ratio               1.00000000              0.044973865
## take.away.within.30.days    0.04497386              1.000000000
```

Compute the correlation matrix and visualize it

#see http://www.sthda.com/english/wiki/visualize-correlation-matrix-using-correlogram for more options

```
cor.nhanes = cor(training)
cor.nhanes

##                            diabetes_risk hypertension   gluten_free
stroke
## diabetes_risk              1.000000000  0.090436992  0.0615132218 -
0.0285428067
## hypertension               0.090436992  1.000000000 -0.0113951126
0.1399969066
## gluten_free                0.061513222 -0.011395113  1.0000000000 -
0.0128137048
## stroke                    -0.028542807  0.139996907 -0.0128137048
1.0000000000
## diabetes_relatives         0.392731801  0.080494103  0.0094549842 -
0.0490375107
## moderate_activity         -0.031835550 -0.071961362 -0.0129130166 -
0.0437370394
## diabetes                  -0.107186144 -0.036764358  0.0232209336
```

```
0.0083511297
## feel_at_risk_diabetes      0.396154774  0.083479358  0.0111149716 -
0.0450970441
## frozen_meals              -0.074087409 -0.069818003  0.0152227368
0.0203494126
## gender                    -0.069677113  0.019205510 -0.0542433071 -
0.0014015826
## age                       -0.007915844 -0.256933717  0.0331410204 -
0.0935696239
## education_level            0.058137715  0.007573236 -0.0428094360
0.0395129927
## height                     0.046300803 -0.038801289  0.0536294961
0.0115676008
## weight                    -0.142654851 -0.209344654  0.0411259953
0.0197776476
## sleep_weekdays             0.038292241  0.113132386 -0.0017584288 -
0.0430516962
## hours_worked              -0.020219801 -0.095485329  0.0028256882 -
0.0006652705
## time_outdoors_weekends    -0.016302149 -0.052429020 -0.0008726455
0.0206759969
## time_sitting               0.020223119 -0.051456224  0.0053529889
0.0571595815
## take_away_food             0.042177592 -0.031544236  0.0157896563 -
0.0010566200
## poverty_ratio              0.041587392 -0.019764767 -0.0648712937
0.0490310086
## take.away.within.30.days   0.002877079 -0.018033926 -0.0208501788
0.0035898193
##                          diabetes_relatives moderate_activity
diabetes
## diabetes_risk                  0.392731801      -0.031835550 -
0.1071861442
## hypertension                   0.080494103      -0.071961362 -
0.0367643582
## gluten_free                    0.009454984      -0.012913017
0.0232209336
## stroke                        -0.049037511      -0.043737039
0.0083511297
## diabetes_relatives             1.000000000       0.009568842 -
0.0506855928
## moderate_activity              0.009568842       1.000000000
0.0067329042
## diabetes                      -0.050685593       0.006732904
1.0000000000
## feel_at_risk_diabetes          0.369267734      -0.027135791 -
0.0825251735
## frozen_meals                   0.039744098       0.035277088 -
0.0066889062
## gender                        -0.022163817      -0.135406815 -
```

```
0.0069251388
## age                             -0.046230078      -0.039282485
0.0860474862
## education_level                  0.106576676      -0.077169470 -
0.0189820852
## height                           0.032582968       0.122291702
0.0214838020
## weight                          -0.105673395       0.045252728
0.0679719674
## sleep_weekdays                   0.047275383      -0.035829809 -
0.0147121912
## hours_worked                    -0.028353387       0.058446792
0.0022603991
## time_outdoors_weekends           0.020483716       0.239109169 -
0.0045212160
## time_sitting                     0.019424403      -0.067940765
0.0005736717
## take_away_food                   0.069726673       0.051162178 -
0.0392082121
## poverty_ratio                    0.063959798      -0.082340664
0.0286538545
## take.away.within.30.days        -0.021904478      -0.010243428
0.0125186678
##                          feel_at_risk_diabetes frozen_meals       gender
## diabetes_risk                        0.396154774 -0.074087409 -0.069677113
## hypertension                         0.083479358 -0.069818003  0.019205510
## gluten_free                          0.011114972  0.015222737 -0.054243307
## stroke                              -0.045097044  0.020349413 -0.001401583
## diabetes_relatives                   0.369267734  0.039744098 -0.022163817
## moderate_activity                   -0.027135791  0.035277088 -0.135406815
## diabetes                            -0.082525174 -0.006688906 -0.006925139
## feel_at_risk_diabetes                1.000000000 -0.034889204 -0.075931561
## frozen_meals                        -0.034889204  1.000000000  0.013507463
## gender                              -0.075931561  0.013507463  1.000000000
## age                                 -0.034037281 -0.153016067  0.017538292
## education_level                      0.045806652  0.005306166  0.031733105
## height                               0.071546435  0.018508267 -0.654319198
## weight                              -0.174231487  0.045859507 -0.332371361
## sleep_weekdays                       0.041937838  0.008831964 -0.001365208
## hours_worked                        -0.010670401 -0.041802867 -0.198461173
## time_outdoors_weekends              -0.006576999  0.019592958 -0.177300996
## time_sitting                        -0.016110151  0.003340262 -0.052697811
## take_away_food                       0.003138257  0.102627552 -0.192740111
## poverty_ratio                        0.061879957 -0.031368967 -0.045973080
## take.away.within.30.days             0.023875789  0.090832169 -0.042288767
##                                      age education_level        height
## diabetes_risk            -0.0079158437     0.058137715  0.0463008031
## hypertension             -0.2569337167     0.007573236 -0.0388012886
## gluten_free               0.0331410204    -0.042809436  0.0536294961
## stroke                   -0.0935696239     0.039512993  0.0115676008
```

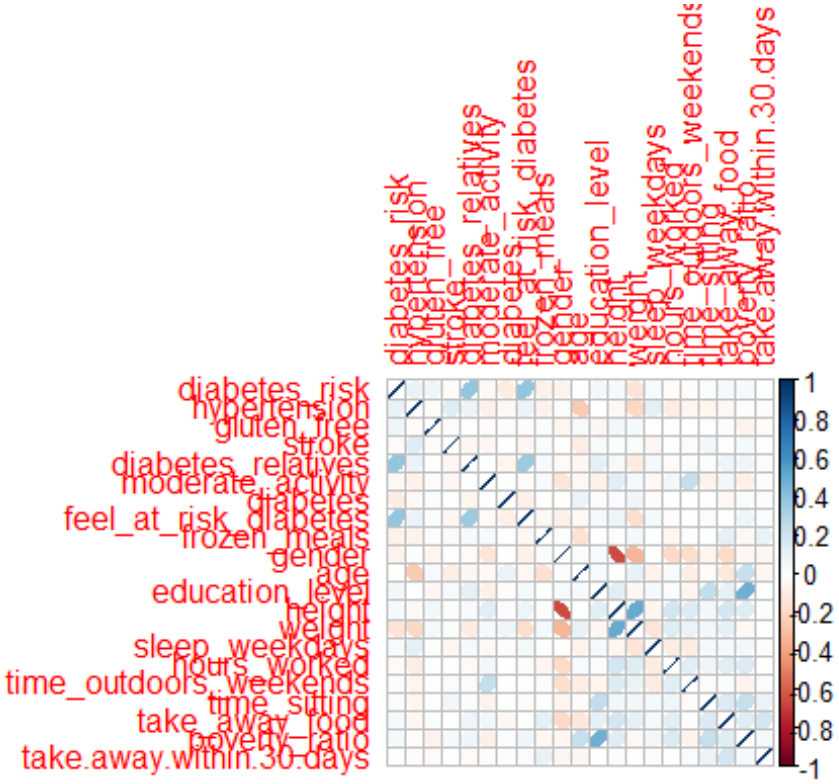
```
## diabetes_relatives       -0.0462300783     0.106576676  0.0325829680
## moderate_activity        -0.0392824853    -0.077169470  0.1222917025
## diabetes                  0.0860474862    -0.018982085  0.0214838020
## feel_at_risk_diabetes    -0.0340372812     0.045806652  0.0715464350
## frozen_meals             -0.1530160671     0.005306166  0.0185082668
## gender                    0.0175382924     0.031733105 -0.6543191977
## age                       1.0000000000     0.037783357 -0.0009177865
## education_level           0.0377833572     1.000000000  0.1242337371
## height                   -0.0009177865     0.124233737  1.0000000000
## weight                    0.0950790520    -0.048969951  0.5096989759
## sleep_weekdays           -0.0648257873     0.041616202 -0.0526277909
## hours_worked              0.0352481936     0.020136280  0.1861108710
## time_outdoors_weekends   -0.0912503212    -0.057627635  0.1422866669
## time_sitting             -0.0121442337     0.243246800  0.0500421298
## take_away_food           -0.1206496639     0.102931634  0.1838445087
## poverty_ratio             0.2245067391     0.488618938  0.1349542490
## take.away.within.30.days -0.0474925630     0.007608104  0.0177975380
##                                weight sleep_weekdays  hours_worked
## diabetes_risk            -0.142654851    0.038292241 -0.0202198009
## hypertension             -0.209344654    0.113132386 -0.0954853287
## gluten_free               0.041125995   -0.001758429  0.0028256882
## stroke                    0.019777648   -0.043051696 -0.0006652705
## diabetes_relatives       -0.105673395    0.047275383 -0.0283533868
## moderate_activity         0.045252728   -0.035829809  0.0584467924
## diabetes                  0.067971967   -0.014712191  0.0022603991
## feel_at_risk_diabetes    -0.174231487    0.041937838 -0.0106704015
## frozen_meals              0.045859507    0.008831964 -0.0418028666
## gender                   -0.332371361   -0.001365208 -0.1984611729
## age                       0.095079052   -0.064825787  0.0352481936
## education_level          -0.048969951    0.041616202  0.0201362798
## height                    0.509698976   -0.052627791  0.1861108710
## weight                    1.000000000   -0.138974733  0.1314184231
## sleep_weekdays           -0.138974733    1.000000000 -0.1040991923
## hours_worked              0.131418423   -0.104099192  1.0000000000
## time_outdoors_weekends    0.084074807   -0.103787334  0.1233039916
## time_sitting              0.054325677    0.042603526  0.0696603990
## take_away_food            0.106267394    0.041028361  0.1207308004
## poverty_ratio            -0.039892225    0.050828638  0.1508072756
## take.away.within.30.days  0.003098214   -0.016004933 -0.0061899332
##                          time_outdoors_weekends  time_sitting
take_away_food
## diabetes_risk                      -0.0163021489  0.0202231190
0.042177592
## hypertension                       -0.0524290199 -0.0514562237   -
0.031544236
## gluten_free                        -0.0008726455  0.0053529889
0.015789656
## stroke                              0.0206759969  0.0571595815   -
0.001056620
## diabetes_relatives                  0.0204837157  0.0194244034
```

```
0.069726673
## moderate_activity                 0.2391091685 -0.0679407654
0.051162178
## diabetes                         -0.0045212160  0.0005736717   -
0.039208212
## feel_at_risk_diabetes            -0.0065769993 -0.0161101507
0.003138257
## frozen_meals                      0.0195929583  0.0033402617
0.102627552
## gender                           -0.1773009959 -0.0526978113   -
0.192740111
## age                              -0.0912503212 -0.0121442337   -
0.120649664
## education_level                  -0.0576276353  0.2432467996
0.102931634
## height                            0.1422866669  0.0500421298
0.183844509
## weight                            0.0840748065  0.0543256768
0.106267394
## sleep_weekdays                   -0.1037873345  0.0426035261
0.041028361
## hours_worked                      0.1233039916  0.0696603990
0.120730800
## time_outdoors_weekends            1.0000000000 -0.0247290651
0.013988851
## time_sitting                     -0.0247290651  1.0000000000
0.156956732
## take_away_food                    0.0139888511  0.1569567321
1.000000000
## poverty_ratio                    -0.0689465337  0.2312651439
0.136838530
## take.away.within.30.days          0.0276676684  0.0846926782
0.182265458
##                          poverty_ratio take.away.within.30.days
## diabetes_risk              0.04158739              0.002877079
## hypertension              -0.01976477             -0.018033926
## gluten_free               -0.06487129             -0.020850179
## stroke                     0.04903101              0.003589819
## diabetes_relatives         0.06395980             -0.021904478
## moderate_activity         -0.08234066             -0.010243428
## diabetes                   0.02865385              0.012518668
## feel_at_risk_diabetes      0.06187996              0.023875789
## frozen_meals              -0.03136897              0.090832169
## gender                    -0.04597308             -0.042288767
## age                        0.22450674             -0.047492563
## education_level            0.48861894              0.007608104
## height                     0.13495425              0.017797538
## weight                    -0.03989223              0.003098214
## sleep_weekdays             0.05082864             -0.016004933
## hours_worked               0.15080728             -0.006189933
```

```
## time_outdoors_weekends     -0.06894653              0.027667668
## time_sitting                0.23126514              0.084692678
## take_away_food              0.13683853              0.182265458
## poverty_ratio               1.00000000              0.044973865
## take.away.within.30.days    0.04497386              1.000000000
```

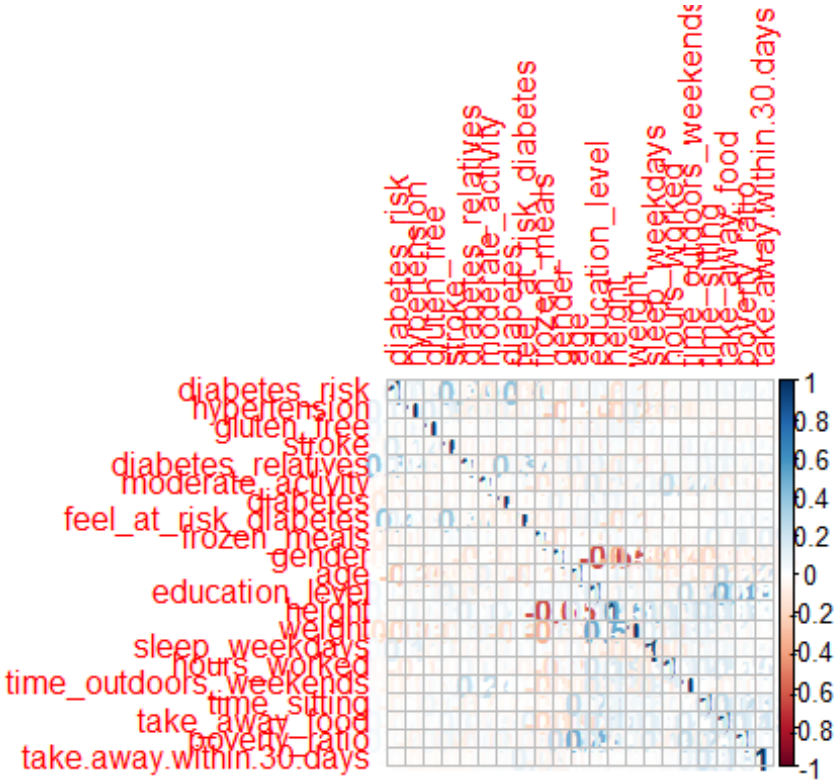
correlation matrix using ellipse method

```
corrplot(cor.nhanes, method="ellipse")
```



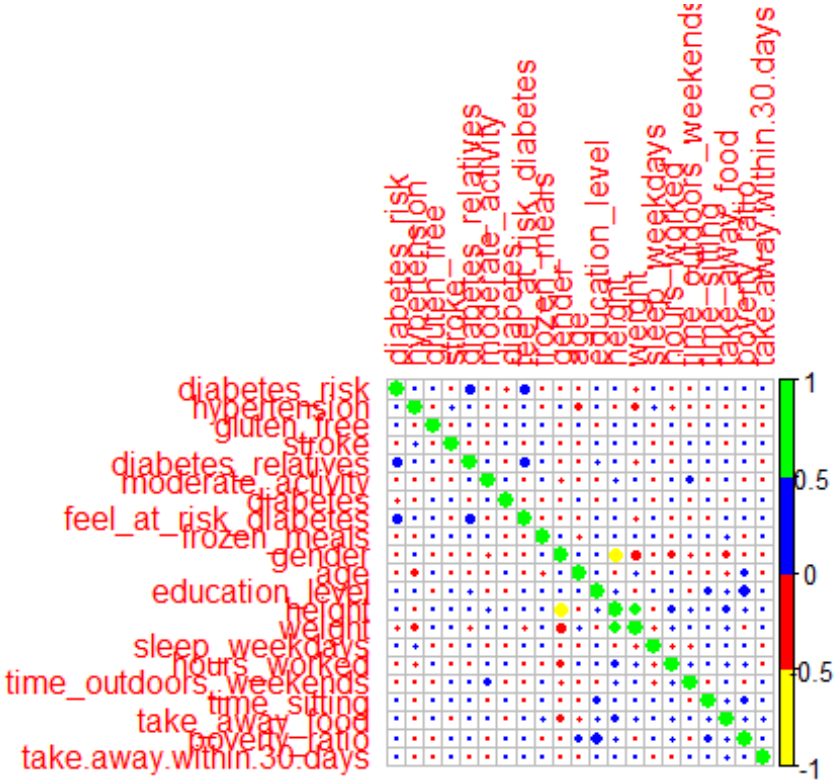
correlation matrix using number method

```
corrplot(cor.nhanes, method="number")
```



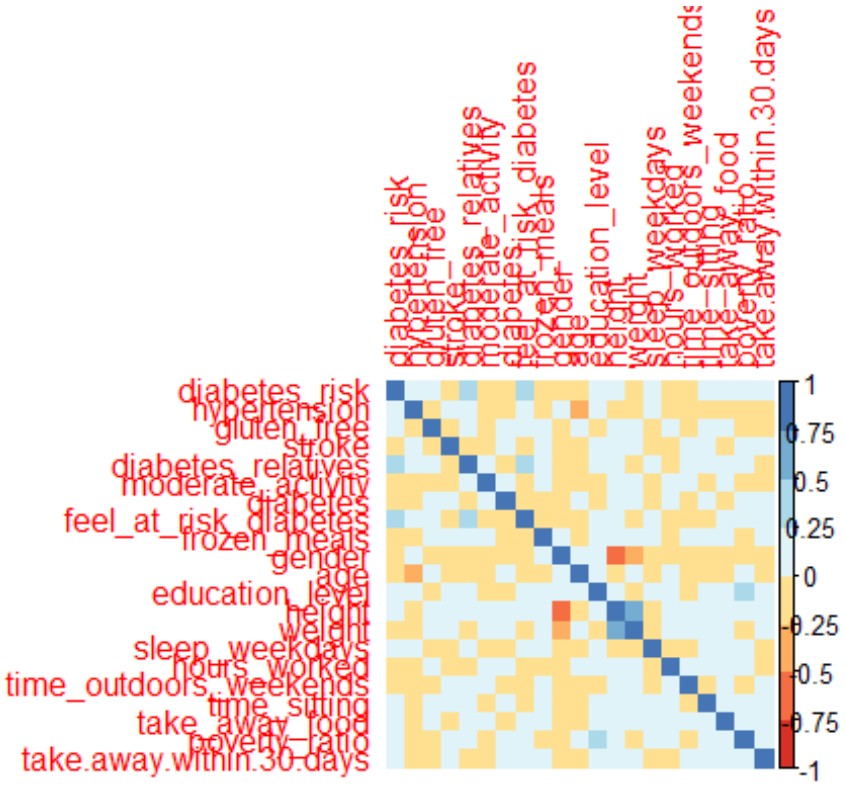
correlation matrix using circle method

```
corrplot(cor.nhanes, method="circle",col=c("yellow", "red","blue","green"))
```



correlation matrix using color method

```
corrplot(cor.nhanes, method="color", col=brewer.pal(n=8, name="RdYlBu"))
```



#initial model

```
model1 <- lm(diabetes ~ ., data=training)
model1

##
## Call:
## lm(formula = diabetes ~ ., data = training)
##
## Coefficients:
##              (Intercept)             diabetes_risk
hypertension
##                8.897e-01                -5.994e-02                -9.823e-
04
##              gluten_free                   stroke
diabetes_relatives
##                3.619e-02                 3.560e-02                 5.986e-
03
##        moderate_activity     feel_at_risk_diabetes
frozen_meals
##                3.300e-05                -2.363e-02                -2.281e-
04
##                   gender                       age
education_level
##                4.442e-04                 1.492e-03                -7.482e-
```

```
03
##                   height                   weight
sleep_weekdays
##                8.753e-04                2.389e-04                2.518e-
04
##             hours_worked    time_outdoors_weekends
time_sitting
##                -1.891e-04               -1.077e-05                2.591e-
06
##           take_away_food             poverty_ratio
take.away.within.30.days
##                -2.356e-03                5.981e-03                7.242e-
04

summary(model1)

##
## Call:
## lm(formula = diabetes ~ ., data = training)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.15533 -0.04873 -0.02412 -0.00203  1.97968
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)               8.897e-01  2.493e-01   3.569 0.000379 ***
## diabetes_risk            -5.994e-02  2.647e-02  -2.265 0.023770 *
## hypertension             -9.823e-04  2.273e-02  -0.043 0.965539
## gluten_free               3.619e-02  4.540e-02   0.797 0.425679
## stroke                    3.560e-02  1.241e-01   0.287 0.774376
## diabetes_relatives        5.986e-03  1.968e-02   0.304 0.761071
## moderate_activity         3.300e-05  1.456e-04   0.227 0.820767
## feel_at_risk_diabetes    -2.363e-02  2.046e-02  -1.155 0.248385
## frozen_meals             -2.281e-04  1.639e-03  -0.139 0.889331
## gender                    4.442e-04  2.246e-02   0.020 0.984228
## age                       1.492e-03  8.269e-04   1.804 0.071563 .
## education_level          -7.482e-03  9.872e-03  -0.758 0.448729
## height                    8.753e-04  3.154e-03   0.278 0.781459
## weight                    2.389e-04  2.564e-04   0.932 0.351838
## sleep_weekdays            2.518e-04  7.201e-03   0.035 0.972116
## hours_worked             -1.891e-04  6.428e-04  -0.294 0.768677
## time_outdoors_weekends   -1.077e-05  7.471e-05  -0.144 0.885368
## time_sitting              2.591e-06  4.346e-05   0.060 0.952471
## take_away_food           -2.356e-03  2.104e-03  -1.120 0.263125
## poverty_ratio             5.981e-03  6.145e-03   0.973 0.330715
## take.away.within.30.days  7.242e-04  1.058e-03   0.685 0.493839
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.2421 on 851 degrees of freedom
## Multiple R-squared:  0.02593,    Adjusted R-squared:  0.003036
## F-statistic: 1.133 on 20 and 851 DF,  p-value: 0.3093
```

#Check VIF

```
VIF(model1)
```

```
##            diabetes_risk             hypertension              gluten_free
##                 1.329944                 1.184333                 1.018481
##                   stroke       diabetes_relatives        moderate_activity
##                 1.046596                 1.297728                 1.099493
##    feel_at_risk_diabetes             frozen_meals                   gender
##                 1.322723                 1.068399                 1.874968
##                      age          education_level                   height
##                 1.241001                 1.417901                 2.316717
##                   weight           sleep_weekdays             hours_worked
##                 1.594677                 1.061999                 1.112884
##   time_outdoors_weekends             time_sitting          take_away_food
##                 1.123208                 1.127544                 1.160735
##            poverty_ratio take.away.within.30.days
##                 1.524619                 1.054395
```

```
library(plotmo)
```

```
## Warning: package 'plotmo' was built under R version 4.0.3
```

```
## Loading required package: plotrix
```

```
## Warning: package 'plotrix' was built under R version 4.0.3
```

```
##
## Attaching package: 'plotrix'
```

```
## The following object is masked from 'package:psych':
##
##     rescale
```

```
## Loading required package: TeachingDemos
```

```
## Warning: package 'TeachingDemos' was built under R version 4.0.3
```

```
##
## Attaching package: 'TeachingDemos'
```

```
## The following objects are masked from 'package:Hmisc':
##
##     cnvrt.coords, subplot
```

```
#plotmo::plotres(lasso)
# the object lasso is missing
```

check if residuals add up to 0

not exactly 0 but looks good (rounding error)

```
#model1$residuals
sum(model1$residuals)
```

```
## [1] -6.82798e-15
```

#check whether or not residuals are independent
shows p-value that the residuals are dependent on each other # 0.08 reject H0 and accept alternative that they are not depednetn on one another # depedneing on alpha. higher than wanted may need to clean up data #load car package

```
library(car)
```

```
## Warning: package 'car' was built under R version 4.0.3
```

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

```
## Warning: package 'carData' was built under R version 4.0.3
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:psych':
##
##     logit
```

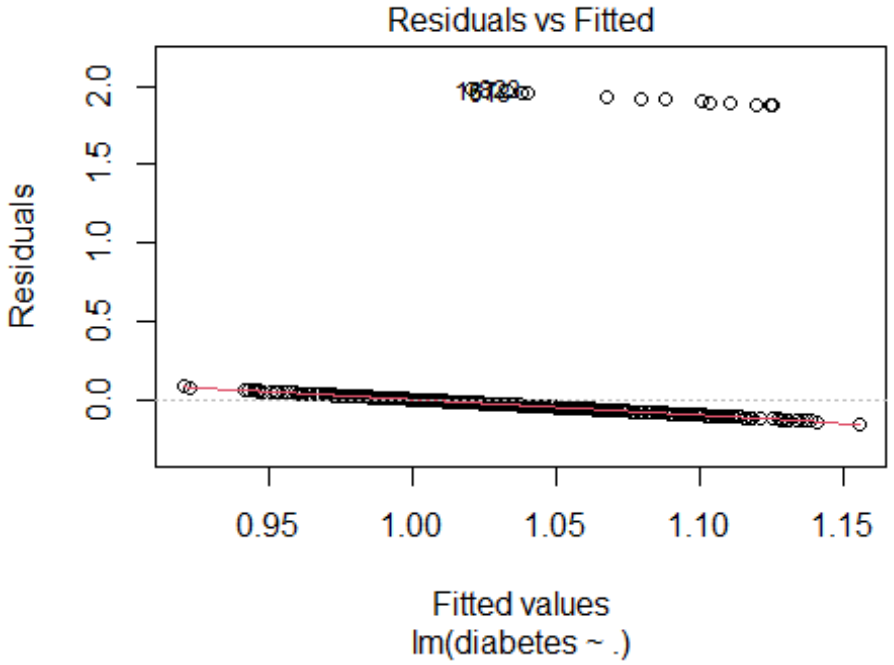
```
## The following object is masked from 'package:DescTools':
##
##     Recode
```

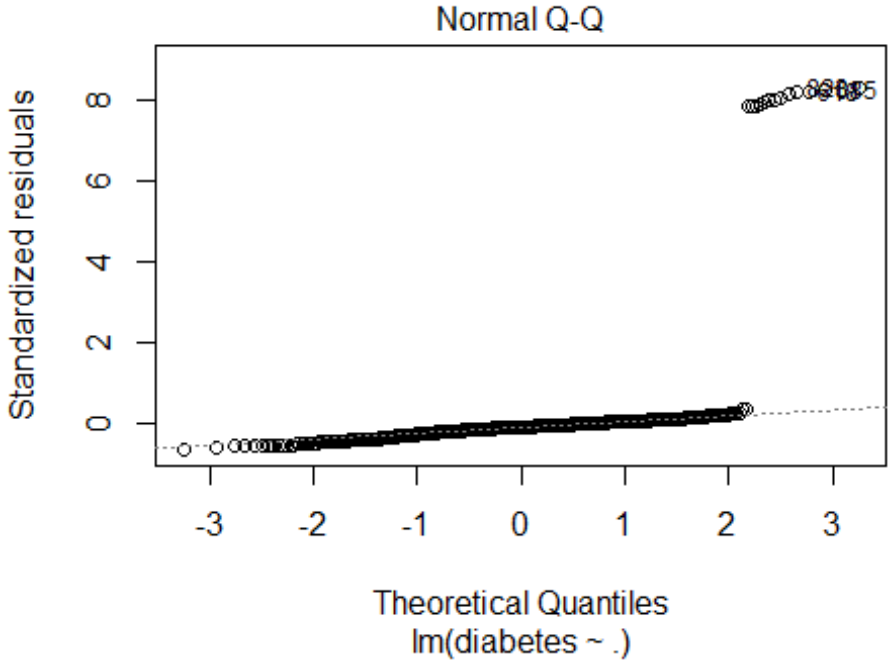
```
durbinWatsonTest(model1)
```

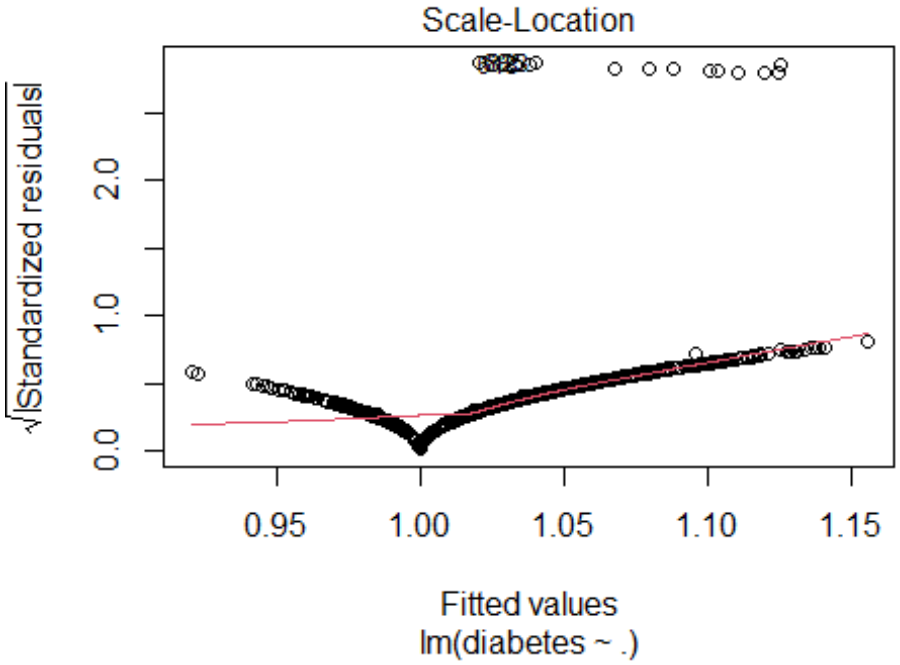
```
##  lag Autocorrelation D-W Statistic p-value
##    1     -0.01220816      2.024404   0.914
##  Alternative hypothesis: rho != 0
```

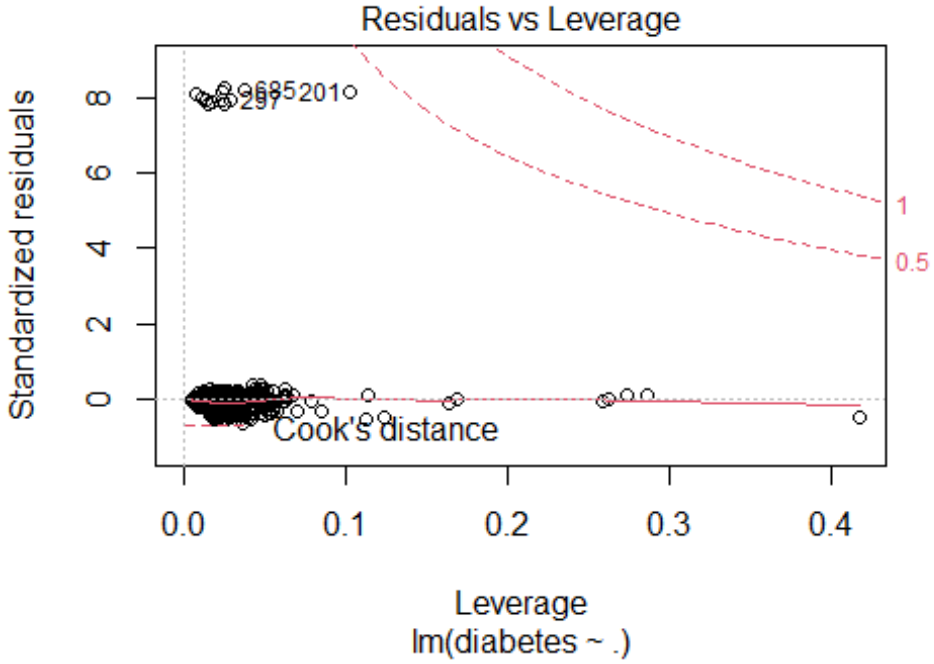
#checking homoscedacticity #plotting model will show a series of plots. first is the fitted values (y) against # residuals, then normal Q-Q plot, fitted values agaisnts squares (standardized # residuals), then leverage againsts standardized residuals (points that have a # lot of leverage, may want to remove)

```
plot(model1)
```


Residuals vs Fitted
Residuals
0.0
0.5
1.0
1.5
2.0
0.95
1.00
1.05
1.10
1.15
Fitted values
lm(diabetes ~ .)


Normal Q-Q
Standardized residuals
0
2
4
6
8
-3
-2
-1
0
1
2
3
Theoretical Quantiles
lm(diabetes ~ .)





#Shapiro-Wilk test (<1,000 sample size)

```
shapiro.test(training$diabetes)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  training$diabetes
## W = 0.096721, p-value < 2.2e-16

library(psych)
describe(training)

##                          vars   n   mean     sd median trimmed    mad min
max
## diabetes_risk               1 872   0.85   0.36   1.00    0.94   0.00   0
1
## hypertension                2 872   0.81   0.39   1.00    0.89   0.00   0
1
## gluten_free                 3 872   0.97   0.18   1.00    1.00   0.00   0
1
## stroke                      4 872   1.00   0.07   1.00    1.00   0.00   0
1
## diabetes_relatives          5 872   0.66   0.47   1.00    0.70   0.00   0
1
## moderate_activity           6 872  64.46  59.07  60.00   53.61  44.48  10
600
## diabetes                    7 872   1.03   0.24   1.00    1.00   0.00   1
3
## feel_at_risk_diabetes       8 872   0.69   0.46   1.00    0.74   0.00   0
1
## frozen_meals                9 872   2.05   5.18   0.00    0.80   0.00   0
60
## gender                     10 872   0.49   0.50   0.00    0.49   0.00   0
1
## age                        11 872  39.27  11.05  40.00   39.19  13.34  20
59
## education_level            12 872   4.08   0.99   4.00    4.21   1.48   1
5
## height                     13 872  66.92   3.96  67.00   66.90   4.45  48
79
## weight                     14 872 173.76  40.40 170.00  171.02  41.51  92
380
## sleep_weekdays             15 872   6.76   1.17   7.00    6.83   1.48   2
10
## hours_worked               16 872  40.37  13.46  40.00   40.51   7.41   5
120
## time_outdoors_weekends     17 872 170.43 116.39 120.00  156.19  88.96  14
480
## time_sitting               18 872 423.68 200.45 480.00  419.05 177.91  30
1080
## take_away_food             19 872   4.18   4.20   3.00    3.45   2.97   0
21
## poverty_ratio              20 872   3.12   1.65   3.22    3.21   2.64   0
```

```
5
## take.away.within.30.days   21 872   2.42   7.96   0.00    0.83   0.00   0
150
##                            range   skew kurtosis   se
## diabetes_risk                  1  -1.95     1.82 0.01
## hypertension                   1  -1.57     0.48 0.01
## gluten_free                    1  -5.10    24.04 0.01
## stroke                         1 -14.64   212.51 0.00
## diabetes_relatives             1  -0.66    -1.57 0.02
## moderate_activity            590   3.26    16.39 2.00
## diabetes                       2   7.99    61.94 0.01
## feel_at_risk_diabetes          1  -0.84    -1.30 0.02
## frozen_meals                  60   5.29    41.86 0.18
## gender                         1   0.04    -2.00 0.02
## age                           39   0.04    -1.11 0.37
## education_level                4  -0.97     0.40 0.03
## height                        31  -0.04     0.11 0.13
## weight                       288   0.81     1.36 1.37
## sleep_weekdays                 8  -0.44     0.42 0.04
## hours_worked                 115   0.18     2.22 0.46
## time_outdoors_weekends       466   0.98     0.36 3.94
## time_sitting                1050   0.19    -0.57 6.79
## take_away_food                21   1.80     3.51 0.14
## poverty_ratio                  5  -0.20    -1.46 0.06
## take.away.within.30.days     150   9.86   147.56 0.27
```

#Check for Multicollinearity with Correlations

```
M<-cor(training, method="spearman")
M

##                            diabetes_risk hypertension   gluten_free
stroke
## diabetes_risk               1.00000000  0.090436992  0.0615132218 -
0.028542807
## hypertension                0.09043699  1.000000000 -0.0113951126
0.139996907
## gluten_free                 0.06151322 -0.011395113  1.0000000000 -
0.012813705
## stroke                     -0.02854281  0.139996907 -0.0128137048
1.000000000
## diabetes_relatives          0.39273180  0.080494103  0.0094549842 -
0.049037511
## moderate_activity           0.02287854 -0.002580419 -0.0312485674 -
0.012988379
## diabetes                   -0.10718614 -0.036764358  0.0232209336
0.008351130
## feel_at_risk_diabetes       0.39615477  0.083479358  0.0111149716 -
0.045097044
## frozen_meals               -0.06052470  0.008622811  0.0019619660
```

```
0.018972733
## gender                       -0.06967711  0.019205510 -0.0542433071 -
0.001401583
## age                          -0.01168395 -0.255692689  0.0318060332 -
0.091725084
## education_level               0.06051259  0.013001408 -0.0426383041
0.020076149
## height                        0.04972194 -0.045484607  0.0587935237
0.014435065
## weight                       -0.14935971 -0.200944019  0.0474780120
0.019455401
## sleep_weekdays                0.03027933  0.114920810  0.0040685882 -
0.022297778
## hours_worked                 -0.01440265 -0.090838394  0.0000000000
0.025453640
## time_outdoors_weekends       -0.02020927 -0.034429111  0.0003044514
0.023061812
## time_sitting                  0.01675713 -0.045670159  0.0101994155
0.056955011
## take_away_food                0.03327236 -0.024963518  0.0168788575
0.025530477
## poverty_ratio                 0.04057254 -0.022894441 -0.0625571800
0.047946425
## take.away.within.30.days     -0.05761859 -0.010581658 -0.0666433560 -
0.020193812
##                           diabetes_relatives moderate_activity
diabetes
## diabetes_risk                    0.392731801       0.022878542 -
0.1071861442
## hypertension                     0.080494103      -0.002580419 -
0.0367643582
## gluten_free                      0.009454984      -0.031248567
0.0232209336
## stroke                          -0.049037511      -0.012988379
0.0083511297
## diabetes_relatives               1.000000000       0.028821447 -
0.0506855928
## moderate_activity                0.028821447       1.000000000 -
0.0676269077
## diabetes                        -0.050685593      -0.067626908
1.0000000000
## feel_at_risk_diabetes            0.369267734       0.025367325 -
0.0825251735
## frozen_meals                     0.026272291      -0.003075780 -
0.0089616915
## gender                          -0.022163817      -0.108895563 -
0.0069251388
## age                             -0.050295961      -0.069537595
0.0856503610
## education_level                  0.112787665      -0.068297589 -
```

```
0.0346027844
## height                          0.029418615       0.100803310
0.0171912442
## weight                         -0.117523226      -0.004752512
0.0463826043
## sleep_weekdays                  0.039140558      -0.012794750
0.0008379748
## hours_worked                   -0.010875355       0.008308572
0.0235590839
## time_outdoors_weekends          0.007973687       0.181254686 -
0.0043309283
## time_sitting                    0.019325702      -0.087756551
0.0026291638
## take_away_food                  0.069750055       0.012316755 -
0.0625085970
## poverty_ratio                   0.068128400      -0.065618032
0.0329806045
## take.away.within.30.days       -0.054519750       0.025194452 -
0.0412578279
##                          feel_at_risk_diabetes frozen_meals       gender
## diabetes_risk                    0.3961547740 -0.060524698 -0.069677113
## hypertension                     0.0834793578  0.008622811  0.019205510
## gluten_free                      0.0111149716  0.001961966 -0.054243307
## stroke                          -0.0450970441  0.018972733 -0.001401583
## diabetes_relatives               0.3692677344  0.026272291 -0.022163817
## moderate_activity                0.0253673249 -0.003075780 -0.108895563
## diabetes                        -0.0825251735 -0.008961691 -0.006925139
## feel_at_risk_diabetes            1.0000000000 -0.073613302 -0.075931561
## frozen_meals                    -0.0736133016  1.000000000  0.049750768
## gender                          -0.0759315613  0.049750768  1.000000000
## age                             -0.0366694196 -0.152693993  0.019902098
## education_level                  0.0629186433  0.013178990  0.006992857
## height                           0.0652878104  0.020402679 -0.679444432
## weight                          -0.1736399363  0.030999144 -0.379199680
## sleep_weekdays                   0.0330335400 -0.024884134  0.010572968
## hours_worked                    -0.0157426269 -0.119191012 -0.215314890
## time_outdoors_weekends           0.0007725219 -0.004092921 -0.171506619
## time_sitting                    -0.0224043620 -0.015489146 -0.055487917
## take_away_food                  -0.0230102595  0.095626120 -0.204912375
## poverty_ratio                    0.0656469578 -0.022337051 -0.044403271
## take.away.within.30.days        -0.0440541471  0.124075691 -0.031992244
##                                        age education_level       height
weight
## diabetes_risk            -0.011683953     0.060512586  0.049721941 -
0.149359711
## hypertension             -0.255692689     0.013001408 -0.045484607 -
0.200944019
## gluten_free               0.031806033    -0.042638304  0.058793524
0.047478012
## stroke                   -0.091725084     0.020076149  0.014435065
```

```
0.019455401
## diabetes_relatives       -0.050295961     0.112787665  0.029418615 -
0.117523226
## moderate_activity        -0.069537595    -0.068297589  0.100803310 -
0.004752512
## diabetes                  0.085650361    -0.034602784  0.017191244
0.046382604
## feel_at_risk_diabetes    -0.036669420     0.062918643  0.065287810 -
0.173639936
## frozen_meals             -0.152693993     0.013178990  0.020402679
0.030999144
## gender                    0.019902098     0.006992857 -0.679444432 -
0.379199680
## age                       1.000000000     0.067624477  0.002005332
0.100244392
## education_level           0.067624477     1.000000000  0.118570625 -
0.089849664
## height                    0.002005332     0.118570625  1.000000000
0.532465796
## weight                    0.100244392    -0.089849664  0.532465796
1.000000000
## sleep_weekdays           -0.070472122     0.058913355 -0.043798421 -
0.125409058
## hours_worked              0.048923495     0.062366964  0.195084095
0.186405220
## time_outdoors_weekends   -0.103727908    -0.072416399  0.163910210
0.104456495
## time_sitting             -0.005680541     0.247850023  0.050355030
0.062273361
## take_away_food           -0.114890602     0.136665450  0.215340282
0.164271273
## poverty_ratio             0.225481225     0.516948571  0.134553372 -
0.033521176
## take.away.within.30.days -0.055797900     0.048381862  0.057383083
0.052730900
##                          sleep_weekdays hours_worked
time_outdoors_weekends
## diabetes_risk              0.0302793331 -0.014402653          -
0.0202092699
## hypertension               0.1149208104 -0.090838394          -
0.0344291109
## gluten_free                0.0040685882  0.000000000
0.0003044514
## stroke                    -0.0222977779  0.025453640
0.0230618121
## diabetes_relatives         0.0391405578 -0.010875355
0.0079736867
## moderate_activity         -0.0127947502  0.008308572
0.1812546864
## diabetes                   0.0008379748  0.023559084          -
```

```
0.0043309283
## feel_at_risk_diabetes      0.0330335400 -0.015742627
0.0007725219
## frozen_meals              -0.0248841339 -0.119191012          -
0.0040929213
## gender                     0.0105729677 -0.215314890          -
0.1715066189
## age                       -0.0704721225  0.048923495          -
0.1037279084
## education_level            0.0589133554  0.062366964          -
0.0724163987
## height                    -0.0437984208  0.195084095
0.1639102103
## weight                    -0.1254090577  0.186405220
0.1044564946
## sleep_weekdays             1.0000000000 -0.067652747          -
0.0958055061
## hours_worked              -0.0676527473  1.000000000
0.1238982414
## time_outdoors_weekends    -0.0958055061  0.123898241
1.0000000000
## time_sitting               0.0473445074  0.065567790          -
0.0196674519
## take_away_food             0.0299312236  0.081010299
0.0397853190
## poverty_ratio              0.0501785549  0.173809971          -
0.0548097095
## take.away.within.30.days   0.0059275083  0.077007736
0.1000427864
##                          time_sitting take_away_food poverty_ratio
## diabetes_risk             0.016757126     0.03327236    0.04057254
## hypertension             -0.045670159    -0.02496352   -0.02289444
## gluten_free               0.010199416     0.01687886   -0.06255718
## stroke                    0.056955011     0.02553048    0.04794643
## diabetes_relatives        0.019325702     0.06975006    0.06812840
## moderate_activity        -0.087756551     0.01231676   -0.06561803
## diabetes                  0.002629164    -0.06250860    0.03298060
## feel_at_risk_diabetes    -0.022404362    -0.02301026    0.06564696
## frozen_meals             -0.015489146     0.09562612   -0.02233705
## gender                   -0.055487917    -0.20491238   -0.04440327
## age                      -0.005680541    -0.11489060    0.22548122
## education_level           0.247850023     0.13666545    0.51694857
## height                    0.050355030     0.21534028    0.13455337
## weight                    0.062273361     0.16427127   -0.03352118
## sleep_weekdays            0.047344507     0.02993122    0.05017855
## hours_worked              0.065567790     0.08101030    0.17380997
## time_outdoors_weekends   -0.019667452     0.03978532   -0.05480971
## time_sitting              1.000000000     0.14095057    0.23408307
## take_away_food            0.140950568     1.00000000    0.17044422
## poverty_ratio             0.234083066     0.17044422    1.00000000
```

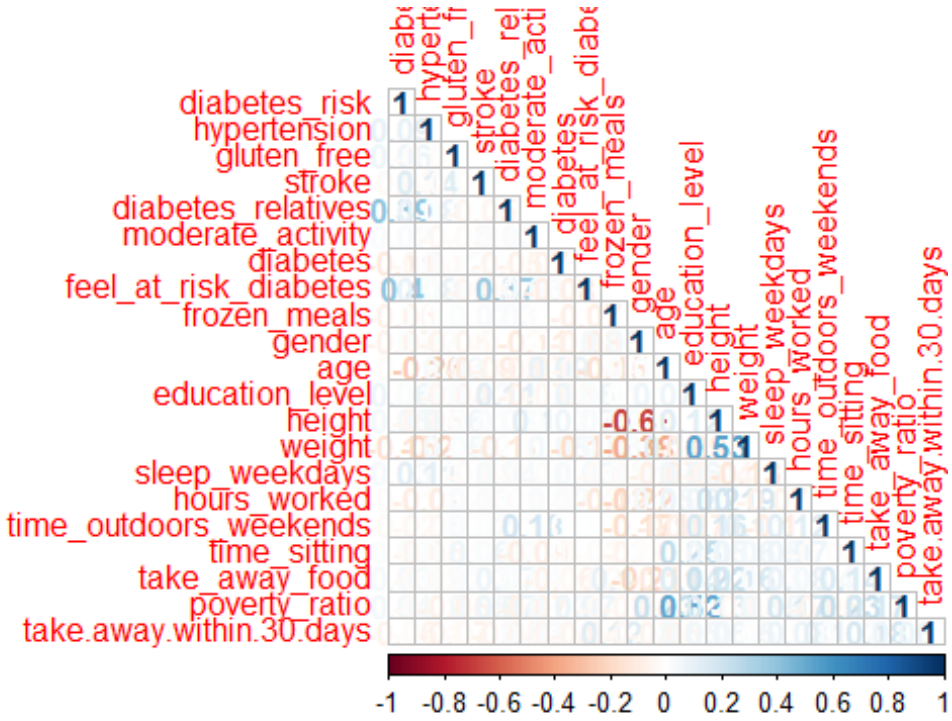
```
## take.away.within.30.days  0.020793923     0.17895033    0.06344348
##                          take.away.within.30.days
## diabetes_risk                        -0.057618592
## hypertension                         -0.010581658
## gluten_free                          -0.066643356
## stroke                               -0.020193812
## diabetes_relatives                   -0.054519750
## moderate_activity                     0.025194452
## diabetes                             -0.041257828
## feel_at_risk_diabetes                -0.044054147
## frozen_meals                          0.124075691
## gender                               -0.031992244
## age                                  -0.055797900
## education_level                       0.048381862
## height                                0.057383083
## weight                                0.052730900
## sleep_weekdays                        0.005927508
## hours_worked                          0.077007736
## time_outdoors_weekends                0.100042786
## time_sitting                          0.020793923
## take_away_food                        0.178950329
## poverty_ratio                         0.063443483
## take.away.within.30.days              1.000000000

round(M,2)

##                          diabetes_risk hypertension gluten_free stroke
## diabetes_risk                     1.00         0.09        0.06  -0.03
## hypertension                      0.09         1.00       -0.01   0.14
## gluten_free                       0.06        -0.01        1.00  -0.01
## stroke                           -0.03         0.14       -0.01   1.00
## diabetes_relatives                0.39         0.08        0.01  -0.05
## moderate_activity                 0.02         0.00       -0.03  -0.01
## diabetes                         -0.11        -0.04        0.02   0.01
## feel_at_risk_diabetes             0.40         0.08        0.01  -0.05
## frozen_meals                     -0.06         0.01        0.00   0.02
## gender                           -0.07         0.02       -0.05   0.00
## age                              -0.01        -0.26        0.03  -0.09
## education_level                   0.06         0.01       -0.04   0.02
## height                            0.05        -0.05        0.06   0.01
## weight                           -0.15        -0.20        0.05   0.02
## sleep_weekdays                    0.03         0.11        0.00  -0.02
## hours_worked                     -0.01        -0.09        0.00   0.03
## time_outdoors_weekends           -0.02        -0.03        0.00   0.02
## time_sitting                      0.02        -0.05        0.01   0.06
## take_away_food                    0.03        -0.02        0.02   0.03
## poverty_ratio                     0.04        -0.02       -0.06   0.05
## take.away.within.30.days         -0.06        -0.01       -0.07  -0.02
##                          diabetes_relatives moderate_activity diabetes
## diabetes_risk                          0.39              0.02    -0.11
```

```
## hypertension                          0.08              0.00    -0.04
## gluten_free                           0.01             -0.03     0.02
## stroke                               -0.05             -0.01     0.01
## diabetes_relatives                    1.00              0.03    -0.05
## moderate_activity                     0.03              1.00    -0.07
## diabetes                             -0.05             -0.07     1.00
## feel_at_risk_diabetes                 0.37              0.03    -0.08
## frozen_meals                          0.03              0.00    -0.01
## gender                               -0.02             -0.11    -0.01
## age                                  -0.05             -0.07     0.09
## education_level                       0.11             -0.07    -0.03
## height                                0.03              0.10     0.02
## weight                               -0.12              0.00     0.05
## sleep_weekdays                        0.04             -0.01     0.00
## hours_worked                         -0.01              0.01     0.02
## time_outdoors_weekends                0.01              0.18     0.00
## time_sitting                          0.02             -0.09     0.00
## take_away_food                        0.07              0.01    -0.06
## poverty_ratio                         0.07             -0.07     0.03
## take.away.within.30.days             -0.05              0.03    -0.04
##                         feel_at_risk_diabetes frozen_meals gender   age
## diabetes_risk                            0.40        -0.06  -0.07 -0.01
## hypertension                             0.08         0.01   0.02 -0.26
## gluten_free                              0.01         0.00  -0.05  0.03
## stroke                                  -0.05         0.02   0.00 -0.09
## diabetes_relatives                       0.37         0.03  -0.02 -0.05
## moderate_activity                        0.03         0.00  -0.11 -0.07
## diabetes                                -0.08        -0.01  -0.01  0.09
## feel_at_risk_diabetes                    1.00        -0.07  -0.08 -0.04
## frozen_meals                            -0.07         1.00   0.05 -0.15
## gender                                  -0.08         0.05   1.00  0.02
## age                                     -0.04        -0.15   0.02  1.00
## education_level                          0.06         0.01   0.01  0.07
## height                                   0.07         0.02  -0.68  0.00
## weight                                  -0.17         0.03  -0.38  0.10
## sleep_weekdays                           0.03        -0.02   0.01 -0.07
## hours_worked                            -0.02        -0.12  -0.22  0.05
## time_outdoors_weekends                   0.00         0.00  -0.17 -0.10
## time_sitting                            -0.02        -0.02  -0.06 -0.01
## take_away_food                          -0.02         0.10  -0.20 -0.11
## poverty_ratio                            0.07        -0.02  -0.04  0.23
## take.away.within.30.days                -0.04         0.12  -0.03 -0.06
##                         education_level height weight sleep_weekdays
## diabetes_risk                      0.06   0.05  -0.15           0.03
## hypertension                       0.01  -0.05  -0.20           0.11
## gluten_free                       -0.04   0.06   0.05           0.00
## stroke                             0.02   0.01   0.02          -0.02
## diabetes_relatives                 0.11   0.03  -0.12           0.04
## moderate_activity                 -0.07   0.10   0.00          -0.01
## diabetes                          -0.03   0.02   0.05           0.00
```

```
## feel_at_risk_diabetes               0.06   0.07  -0.17           0.03
## frozen_meals                        0.01   0.02   0.03          -0.02
## gender                              0.01  -0.68  -0.38           0.01
## age                                 0.07   0.00   0.10          -0.07
## education_level                     1.00   0.12  -0.09           0.06
## height                              0.12   1.00   0.53          -0.04
## weight                             -0.09   0.53   1.00          -0.13
## sleep_weekdays                      0.06  -0.04  -0.13           1.00
## hours_worked                        0.06   0.20   0.19          -0.07
## time_outdoors_weekends             -0.07   0.16   0.10          -0.10
## time_sitting                        0.25   0.05   0.06           0.05
## take_away_food                      0.14   0.22   0.16           0.03
## poverty_ratio                       0.52   0.13  -0.03           0.05
## take.away.within.30.days            0.05   0.06   0.05           0.01
##                          hours_worked time_outdoors_weekends time_sitting
## diabetes_risk                   -0.01                 -0.02         0.02
## hypertension                    -0.09                 -0.03        -0.05
## gluten_free                      0.00                  0.00         0.01
## stroke                           0.03                  0.02         0.06
## diabetes_relatives              -0.01                  0.01         0.02
## moderate_activity                0.01                  0.18        -0.09
## diabetes                         0.02                  0.00         0.00
## feel_at_risk_diabetes           -0.02                  0.00        -0.02
## frozen_meals                    -0.12                  0.00        -0.02
## gender                          -0.22                 -0.17        -0.06
## age                              0.05                 -0.10        -0.01
## education_level                  0.06                 -0.07         0.25
## height                           0.20                  0.16         0.05
## weight                           0.19                  0.10         0.06
## sleep_weekdays                  -0.07                 -0.10         0.05
## hours_worked                     1.00                  0.12         0.07
## time_outdoors_weekends           0.12                  1.00        -0.02
## time_sitting                     0.07                 -0.02         1.00
## take_away_food                   0.08                  0.04         0.14
## poverty_ratio                    0.17                 -0.05         0.23
## take.away.within.30.days         0.08                  0.10         0.02
##                          take_away_food poverty_ratio
take.away.within.30.days
## diabetes_risk                      0.03          0.04                    -
0.06
## hypertension                      -0.02         -0.02                    -
0.01
## gluten_free                        0.02         -0.06                    -
0.07
## stroke                             0.03          0.05                    -
0.02
## diabetes_relatives                 0.07          0.07                    -
0.05
## moderate_activity                  0.01         -0.07
0.03
```

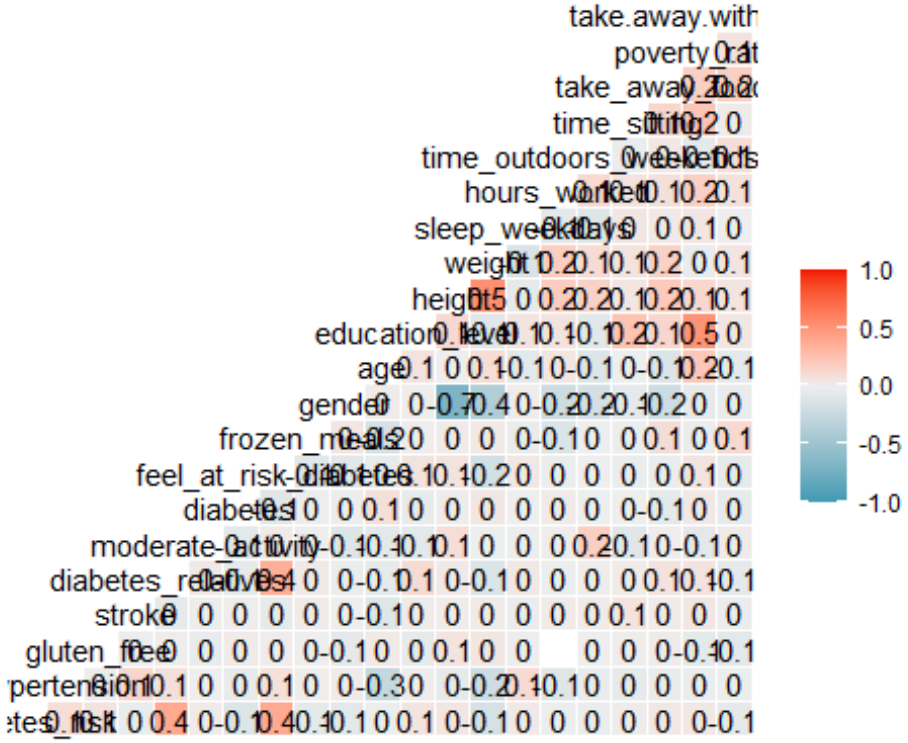
```
## diabetes                          -0.06          0.03                    -
0.04
## feel_at_risk_diabetes             -0.02          0.07                    -
0.04
## frozen_meals                       0.10         -0.02
0.12
## gender                            -0.20         -0.04                    -
0.03
## age                               -0.11          0.23                    -
0.06
## education_level                    0.14          0.52
0.05
## height                             0.22          0.13
0.06
## weight                             0.16         -0.03
0.05
## sleep_weekdays                     0.03          0.05
0.01
## hours_worked                       0.08          0.17
0.08
## time_outdoors_weekends             0.04         -0.05
0.10
## time_sitting                       0.14          0.23
0.02
## take_away_food                     1.00          0.17
0.18
## poverty_ratio                      0.17          1.00
0.06
## take.away.within.30.days           0.18          0.06
1.00
```

```r
corrplot(cor(training,method="spearman"), method = "number", type = "lower")
```

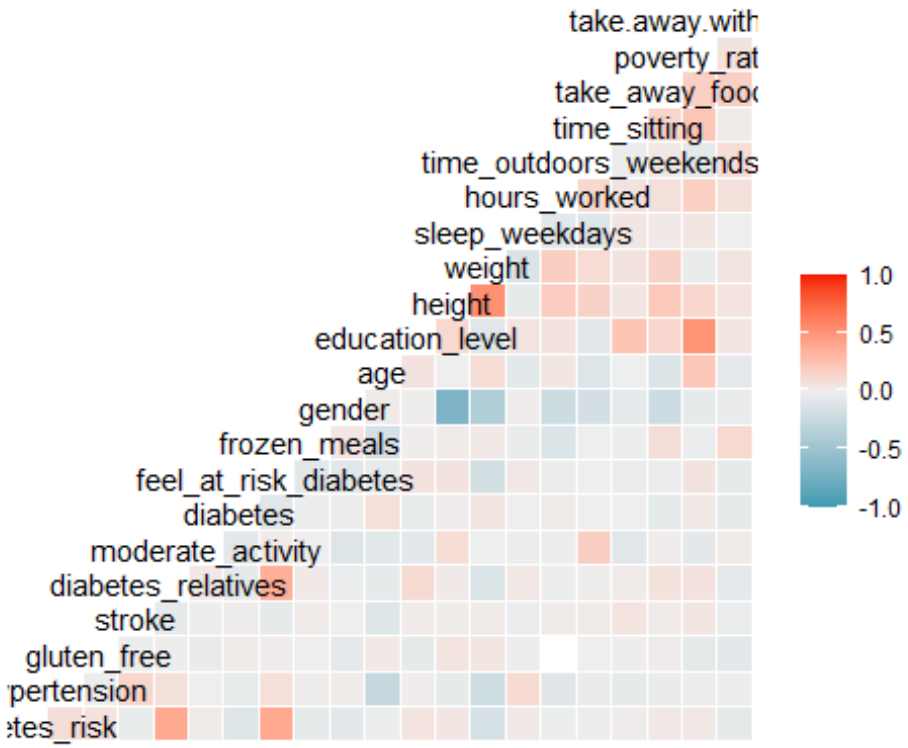


#GGplot Correlation

```
ggcorr(training, method = c("pairwise","spearman"), label=TRUE)
```



```
ggcorr(training, method = c("pairwise","spearman"))
```



PCA_Plot functions - backend code

```
PCA_Plot = function(pcaData)
{
  library(ggplot2)

  theta = seq(0,2*pi,length.out = 100)
  circle = data.frame(x = cos(theta), y = sin(theta))
  p = ggplot(circle,aes(x,y)) + geom_path()

  loadings = data.frame(pcaData$rotation, .names =
row.names(pcaData$rotation))
  p + geom_text(data=loadings, mapping=aes(x = PC1, y = PC2, label = .names,
colour = .names, fontface="bold")) +
    coord_fixed(ratio=1) + labs(x = "PC1", y = "PC2")
}

PCA_Plot_Secondary = function(pcaData)
{
  library(ggplot2)

  theta = seq(0,2*pi,length.out = 100)
```

```
  circle = data.frame(x = cos(theta), y = sin(theta))
  p = ggplot(circle,aes(x,y)) + geom_path()

  loadings = data.frame(pcaData$rotation, .names =
row.names(pcaData$rotation))
  p + geom_text(data=loadings, mapping=aes(x = PC3, y = PC4, label = .names,
colour = .names, fontface="bold")) +
    coord_fixed(ratio=1) + labs(x = "PC3", y = "PC4")
}

PCA_Plot_Psyc = function(pcaData)
{
  library(ggplot2)

  theta = seq(0,2*pi,length.out = 100)
  circle = data.frame(x = cos(theta), y = sin(theta))
  p = ggplot(circle,aes(x,y)) + geom_path()

  loadings = as.data.frame(unclass(pcaData$loadings))
  s = rep(0, ncol(loadings))
  for (i in 1:ncol(loadings))
  {
    s[i] = 0
    for (j in 1:nrow(loadings))
      s[i] = s[i] + loadings[j, i]^2
    s[i] = sqrt(s[i])
  }

  for (i in 1:ncol(loadings))
    loadings[, i] = loadings[, i] / s[i]

  loadings$.names = row.names(loadings)

  p + geom_text(data=loadings, mapping=aes(x = PC1, y = PC2, label = .names,
colour = .names, fontface="bold")) +
    coord_fixed(ratio=1) + labs(x = "PC1", y = "PC2")
}

PCA_Plot_Psyc_Secondary = function(pcaData)
{
  library(ggplot2)

  theta = seq(0,2*pi,length.out = 100)
  circle = data.frame(x = cos(theta), y = sin(theta))
  p = ggplot(circle,aes(x,y)) + geom_path()

  loadings = as.data.frame(unclass(pcaData$loadings))
  s = rep(0, ncol(loadings))
  for (i in 1:ncol(loadings))
```

```
  {
    s[i] = 0
    for (j in 1:nrow(loadings))
      s[i] = s[i] + loadings[j, i]^2
    s[i] = sqrt(s[i])
  }

  for (i in 1:ncol(loadings))
    loadings[, i] = loadings[, i] / s[i]

  loadings$.names = row.names(loadings)

  print(loadings)
  p + geom_text(data=loadings, mapping=aes(x = PC3, y = PC4, label = .names,
colour = .names, fontface="bold")) +
    coord_fixed(ratio=1) + labs(x = "PC3", y = "PC4")
}
```

#PCA/FA
##
##

#Test KMO Sampling Adequacy

```
library(psych)
KMO(training)
```

```
## Kaiser-Meyer-Olkin factor adequacy
## Call: KMO(r = training)
## Overall MSA =  0.62
## MSA for each item =
##            diabetes_risk             hypertension              gluten_free
##                     0.68                     0.58                     0.55
##                   stroke       diabetes_relatives        moderate_activity
##                     0.53                     0.67                     0.65
##                 diabetes    feel_at_risk_diabetes             frozen_meals
##                     0.69                     0.67                     0.50
##                   gender                      age          education_level
##                     0.64                     0.49                     0.57
##                   height                   weight           sleep_weekdays
##                     0.59                     0.65                     0.63
##             hours_worked   time_outdoors_weekends             time_sitting
##                     0.73                     0.67                     0.69
##           take_away_food            poverty_ratio take.away.within.30.days
##                     0.69                     0.57                     0.54
```

#Test Bartlett's Test of Sphericity

```r
library(REdaS)
bart_spher(training)
```

```
##  Bartlett's Test of Sphericity
##
## Call: bart_spher(x = training)
##
##      X2 = 2243.336
##      df = 210
## p-value < 2.22e-16
```

#Test for Reliability Analysis using Cronbach's Alpha

```r
library(psych)
alpha(training, check.keys = TRUE)
```

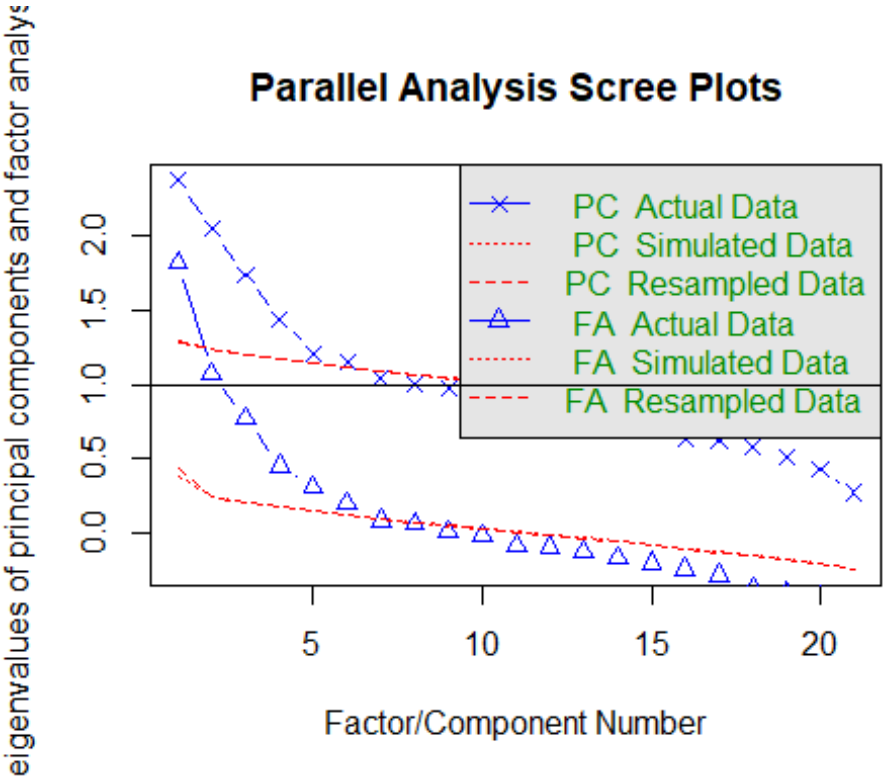
```
## Number of categories should be increased  in order to count frequencies.

## Warning in alpha(training, check.keys = TRUE): Some items were negatively
correlated with total scale and were automatically reversed.
##  This is indicated by a negative sign for the variable name.

##
## Reliability analysis
## Call: alpha(x = training, check.keys = TRUE)
##
##   raw_alpha std.alpha G6(smc) average_r  S/N   ase mean sd median_r
##      0.077      0.49    0.57     0.044 0.97 0.034  356 12     0.02
##
##  lower alpha upper     95% confidence boundaries
## 0.01 0.08 0.14
##
##  Reliability if an item is dropped:
##                          raw_alpha std.alpha G6(smc) average_r  S/N alpha
se
## diabetes_risk-               0.077      0.48    0.55     0.043 0.91
0.034
## hypertension-                0.077      0.47    0.55     0.042 0.89
0.034
## gluten_free                  0.077      0.51    0.58     0.049 1.02
0.034
## stroke                       0.077      0.50    0.58     0.048 1.01
0.034
## diabetes_relatives-          0.077      0.48    0.56     0.045 0.93
0.034
## moderate_activity            0.048      0.49    0.57     0.046 0.97
0.032
## diabetes                     0.077      0.49    0.57     0.046 0.97
0.034
## feel_at_risk_diabetes-       0.077      0.48    0.55     0.043 0.91
0.034
```

```
## frozen_meals                0.076      0.50    0.58     0.048 1.00
0.034
## gender-                     0.076      0.45    0.52     0.040 0.83
0.034
## age                         0.081      0.50    0.57     0.047 0.99
0.034
## education_level             0.076      0.49    0.56     0.046 0.96
0.035
## height                      0.070      0.44    0.50     0.037 0.78
0.035
## weight                      0.040      0.44    0.51     0.037 0.77
0.035
## sleep_weekdays-             0.077      0.49    0.57     0.045 0.95
0.034
## hours_worked                0.061      0.47    0.55     0.042 0.87
0.035
## time_outdoors_weekends      0.031      0.48    0.56     0.045 0.94
0.024
## time_sitting                0.231      0.48    0.56     0.044 0.92
0.023
## take_away_food              0.071      0.47    0.55     0.043 0.90
0.035
## poverty_ratio               0.075      0.47    0.54     0.043 0.90
0.035
## take.away.within.30.days    0.072      0.49    0.57     0.046 0.97
0.035
##                            var.r med.r
## diabetes_risk-           0.0108 0.021
## hypertension-            0.0118 0.019
## gluten_free              0.0122 0.027
## stroke                   0.0121 0.022
## diabetes_relatives-      0.0108 0.020
## moderate_activity        0.0118 0.020
## diabetes                 0.0123 0.021
## feel_at_risk_diabetes-   0.0109 0.020
## frozen_meals             0.0121 0.022
## gender-                  0.0094 0.020
## age                      0.0114 0.020
## education_level          0.0106 0.023
## height                   0.0088 0.020
## weight                   0.0103 0.018
## sleep_weekdays-          0.0122 0.020
## hours_worked             0.0120 0.020
## time_outdoors_weekends   0.0118 0.021
## time_sitting             0.0118 0.020
## take_away_food           0.0116 0.020
## poverty_ratio            0.0105 0.020
## take.away.within.30.days 0.0122 0.022
##
##  Item statistics
```

```
##                              n   raw.r std.r  r.cor  r.drop    mean      sd
## diabetes_risk-             872  0.0257  0.32 0.2650  0.0243 1079.15   0.358
## hypertension-              872  0.1373  0.35 0.2865  0.1357 1079.19   0.393
## gluten_free                872  0.0099  0.16 0.0083  0.0092    0.97   0.182
## stroke                     872  0.0453  0.18 0.0316  0.0450    1.00   0.068
## diabetes_relatives-        872 -0.0058  0.29 0.2167 -0.0077 1079.34   0.475
## moderate_activity          872  0.3013  0.24 0.1282  0.0709   64.46  59.074
## diabetes                   872  0.0163  0.24 0.1082  0.0153    1.03   0.243
## feel_at_risk_diabetes-     872  0.0536  0.32 0.2687  0.0517 1079.31   0.461
## frozen_meals               872  0.0436  0.20 0.0643  0.0231    2.05   5.175
## gender-                    872  0.2343  0.43 0.4530  0.2325 1079.51   0.500
## age                        872 -0.0041  0.21 0.1146 -0.0478   39.27  11.054
## education_level            872  0.1536  0.24 0.1745  0.1498    4.08   0.990
## height                     872  0.2474  0.50 0.5777  0.2327   66.92   3.959
## weight                     872  0.2760  0.51 0.5466  0.1201  173.76  40.405
## sleep_weekdays-            872  0.0580  0.26 0.1452  0.0534 1073.24   1.174
## hours_worked               872  0.2075  0.37 0.2976  0.1557   40.37  13.465
## time_outdoors_weekends     872  0.5164  0.28 0.1831  0.0655  170.43 116.389
## time_sitting               872  0.7858  0.30 0.2115 -0.0111  423.68 200.446
## take_away_food             872  0.1896  0.33 0.2574  0.1735    4.18   4.202
## poverty_ratio              872  0.1570  0.33 0.2941  0.1506    3.12   1.649
## take.away.within.30.days   872  0.1127  0.23 0.1084  0.0815    2.42   7.963
```

#parallel analysis # ****tells you and graphs/suggests how many factors to use

```
library(psych)
comp <- fa.parallel(training)
```



```
## Parallel analysis suggests that the number of factors =  6  and the number
of components =  6

comp

## Call: fa.parallel(x = training)
## Parallel analysis suggests that the number of factors =  6  and the number
of components =  6
##
##  Eigen Values of
##   Original factors Resampled data Simulated data Original components
## 1             1.82           0.44           0.38                 2.37
## 2             1.07           0.25           0.25                 2.05
## 3             0.78           0.21           0.21                 1.74
## 4             0.45           0.18           0.18                 1.44
## 5             0.31           0.15           0.15                 1.21
## 6             0.20           0.13           0.13                 1.15
##   Resampled components Simulated components
## 1                 1.30                  1.28
## 2                 1.24                  1.24
## 3                 1.20                  1.20
## 4                 1.17                  1.17
## 5                 1.14                  1.14
## 6                 1.11                  1.11
```

#Create PCA, scaling the data, gives rotation matrrix

```
p = prcomp(training, center=T, scale=T)
#p = prcomp(responses2)
p

## Standard deviations (1, .., p=21):
##  [1] 1.5398853 1.4304555 1.3174076 1.1982462 1.0999649 1.0741516 1.0216320
##  [8] 0.9999556 0.9893754 0.9859451 0.9357353 0.9167492 0.9003431 0.8852468
## [15] 0.8373727 0.7976577 0.7843861 0.7593787 0.7132246 0.6566862 0.5227533
##
## Rotation (n x k) = (21 x 21):
##                                    PC1          PC2          PC3
PC4
## diabetes_risk            -0.017220368  0.447857170 -0.248209293
0.207978880
## hypertension             -0.157399937  0.180304757 -0.166869536 -
0.303809251
## gluten_free               0.045888222 -0.004914329 -0.090744908
0.100921638
## stroke                    0.008912444  0.005212331  0.054258652 -
0.293407253
## diabetes_relatives       -0.009554807  0.437743818 -0.226861369
0.130837355
## moderate_activity         0.152235463 -0.110301805 -0.241147074 -
0.002150052
```

```
## diabetes                   0.039592162 -0.129999835  0.124036143
0.082879898
## feel_at_risk_diabetes    -0.019045358  0.436726785 -0.258254211
0.192512060
## frozen_meals               0.039345284 -0.033999054 -0.049494294 -
0.345024609
## gender                   -0.483648064 -0.045319640  0.201554673 -
0.008346201
## age                        0.048946093 -0.050586217  0.303076437
0.552935741
## education_level            0.103307956  0.328444979  0.411164748 -
0.075956582
## height                     0.533034096  0.055391460 -0.108813057
0.028247009
## weight                     0.422708440 -0.222522718  0.005676195
0.066737007
## sleep_weekdays           -0.099628197  0.158250670  0.020077839 -
0.166624713
## hours_worked               0.273974790 -0.003272097  0.057190982
0.077214768
## time_outdoors_weekends     0.195225174 -0.089086221 -0.250229900 -
0.042770223
## time_sitting               0.138935523  0.183061527  0.315569077 -
0.196780134
## take_away_food             0.251108342  0.154614597  0.037067580 -
0.334010921
## poverty_ratio              0.158172977  0.309884975  0.472336978
0.040915222
## take.away.within.30.days   0.075689058  0.052733038  0.038244258 -
0.289366632
##                                   PC5          PC6          PC7         PC8
## diabetes_risk            -0.02849167  0.052843588 -0.155451883 -0.02138625
## hypertension              0.49558827 -0.085847997  0.057056250  0.03078400
## gluten_free               0.05498639  0.382411227 -0.252492506  0.23943565
## stroke                    0.39718481 -0.226893817 -0.458866519  0.18431094
## diabetes_relatives       -0.14671347  0.003030195 -0.073923615  0.19750189
## moderate_activity        -0.25342448 -0.370362967  0.339171754  0.24992701
## diabetes                  0.06521336  0.102763137  0.009409529  0.72825840
## feel_at_risk_diabetes    -0.05347183  0.011074192 -0.098777913  0.02034176
## frozen_meals             -0.37798117  0.199621305 -0.105905684  0.30323067
## gender                   -0.20765699 -0.107700807 -0.131066532  0.05009178
## age                      -0.07373931  0.068077831 -0.011882791  0.03042372
## education_level           0.02141892 -0.170612424  0.077207996  0.15947232
## height                    0.23060593  0.120748027  0.073769347  0.02155509
## weight                    0.11049444  0.230636605 -0.107020800 -0.02662396
## sleep_weekdays            0.12088922  0.281190338  0.668712385  0.05893075
## hours_worked             -0.03171206 -0.303816664 -0.088772187 -0.25816092
## time_outdoors_weekends   -0.15945215 -0.488956682  0.008387969  0.18904100
## time_sitting             -0.06657262  0.008351011 -0.081646300 -0.02892368
## take_away_food           -0.24849960  0.155433266  0.041793936 -0.16946501
```

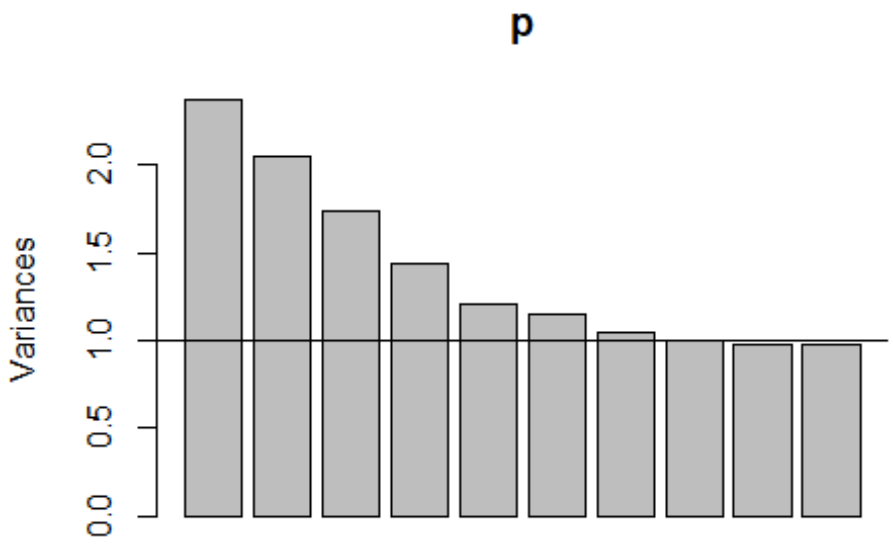
```
## poverty_ratio              0.01628585 -0.179460743  0.082801341  0.09332767
## take.away.within.30.days -0.36765579  0.156225672 -0.214784623 -0.09455427
##                                  PC9         PC10         PC11         PC12
## diabetes_risk              0.046067134 -0.01208746 -0.06037740  0.10262713
## hypertension               0.178564788  0.01109582  0.06390783 -0.20084480
## gluten_free                0.023054895 -0.77540957 -0.09860992 -0.25384209
## stroke                    -0.059097286  0.03640418 -0.18180387 -0.05397513
## diabetes_relatives        -0.131908832  0.09311370  0.11554619  0.15868262
## moderate_activity          0.004504032 -0.14617357 -0.24057957 -0.09946653
## diabetes                   0.445355689  0.18145077  0.32946074  0.20123967
## feel_at_risk_diabetes      0.058869947  0.14074832  0.03394804 -0.01934455
## frozen_meals              -0.493577212  0.12208390  0.24663992 -0.21023390
## gender                    -0.075395654 -0.06611251  0.05181498  0.04102951
## age                        0.079390813  0.05935823 -0.20605947 -0.20393822
## education_level           -0.236251284 -0.02207056 -0.10993236 -0.13746017
## height                    -0.096396529  0.14718424 -0.06878327 -0.07990996
## weight                    -0.186197125  0.16985332 -0.06822525  0.17198276
## sleep_weekdays             0.066068924 -0.13454512 -0.03464555  0.02475105
## hours_worked               0.185069854 -0.25328310  0.66671109 -0.09757071
## time_outdoors_weekends     0.063780571 -0.16693778 -0.19953196  0.08617784
## time_sitting               0.037771368 -0.26330081 -0.16175155  0.70348684
## take_away_food             0.159652425 -0.10306345  0.16225645 -0.02633553
## poverty_ratio             -0.006942839  0.01158744  0.01048632 -0.29856580
## take.away.within.30.days   0.563658533  0.22442896 -0.32093977 -0.24741655
##                                  PC13         PC14         PC15         PC16
## diabetes_risk             -0.106071392  0.079901528 -0.02182960 -0.08003657
## hypertension               0.302656517  0.049898922  0.13231636 -0.47974923
## gluten_free                0.115460745 -0.021700424  0.01036869  0.04147934
## stroke                    -0.612266504  0.112813192 -0.05924233  0.10269426
## diabetes_relatives        -0.028209503  0.118507082 -0.19484683 -0.43822539
## moderate_activity         -0.122479016  0.441153034  0.44712777 -0.04825598
## diabetes                   0.066618670  0.066627876 -0.02681245  0.11566031
## feel_at_risk_diabetes     -0.078558255 -0.161386482  0.21776774  0.48332247
## frozen_meals              -0.117243785 -0.235047614  0.18927707 -0.15680274
## gender                     0.022792537  0.101907925 -0.10568343 -0.07484013
## age                       -0.286586892  0.025501629 -0.03842202 -0.36713067
## education_level            0.321964910  0.004986156 -0.07212821  0.18525869
## height                     0.110691275 -0.040261649  0.07038480  0.05753420
## weight                     0.013602647  0.042230650 -0.09151813 -0.20221067
## sleep_weekdays            -0.462963915 -0.287894867 -0.13371853 -0.02782608
## hours_worked              -0.205128131 -0.176482549  0.17200237 -0.12136761
## time_outdoors_weekends     0.086999143 -0.436759536 -0.50615284 -0.05464141
## time_sitting               0.003141285 -0.099346439  0.34218683 -0.14584907
## take_away_food            -0.073278820  0.553070111 -0.43383670  0.10480485
## poverty_ratio              0.006327403 -0.036926675 -0.01361027 -0.02505809
## take.away.within.30.days   0.025333441 -0.217170009  0.12734926 -0.12228222
##                                  PC17         PC18          PC19
PC20
## diabetes_risk             -0.019310298  0.762863267  0.1946938827
0.06862963
```

```
## hypertension               0.241786125 -0.083718723  0.2607165360 -
0.10637172
## gluten_free               -0.075152305 -0.090732435 -0.0125090228
0.05750504
## stroke                    -0.079169550 -0.034604644 -0.0632166080 -
0.05432680
## diabetes_relatives        -0.303505704 -0.369611979 -0.3662839228
0.03406710
## moderate_activity         -0.116470060 -0.013527462  0.1032305278
0.02856909
## diabetes                   0.009641389  0.097992596  0.0219031592 -
0.02351610
## feel_at_risk_diabetes      0.183533462 -0.430307614  0.3537871574 -
0.05876918
## frozen_meals               0.267133781  0.132728576  0.0570433039 -
0.06772278
## gender                    -0.239846360 -0.095257551  0.5110014112
0.08938977
## age                        0.354106449 -0.082730292  0.0781255931 -
0.36437124
## education_level           -0.318675454  0.106162079  0.0492779905 -
0.52290295
## height                    -0.033790594  0.015220083 -0.0376869756 -
0.02338392
## weight                    -0.299050389 -0.116476780  0.5413397077
0.15717070
## sleep_weekdays            -0.190762548 -0.006004149  0.1210388551 -
0.03183013
## hours_worked              -0.226071815  0.022426289  0.0710929710 -
0.14516181
## time_outdoors_weekends     0.185920628 -0.002839871  0.1327748411 -
0.00527242
## time_sitting               0.215772619 -0.068068170  0.0001921243 -
0.02880464
## take_away_food             0.281800923 -0.075897733  0.1174086706 -
0.10316468
## poverty_ratio              0.129258625 -0.019068963 -0.0067955818
0.70256640
## take.away.within.30.days -0.280436543 -0.011370189 -0.0309955913 -
0.02297522
##                                 PC21
## diabetes_risk             -0.011165033
## hypertension              -0.049423638
## gluten_free               -0.013476727
## stroke                     0.015696687
## diabetes_relatives         0.014069400
## moderate_activity         -0.036858487
## diabetes                  -0.002633246
## feel_at_risk_diabetes     -0.080541571
## frozen_meals              -0.012022287
```

```
## gender                       0.520688616
## age                          0.047815494
## education_level             -0.162768778
## height                       0.747630912
## weight                      -0.346433073
## sleep_weekdays               0.016981735
## hours_worked                 0.012384853
## time_outdoors_weekends       0.016869183
## time_sitting                 0.069276849
## take_away_food               0.023208141
## poverty_ratio               -0.061408042
## take.away.within.30.days     0.010491016
```

#Check Scree Plot

```
plot(p)
abline(1, 0)
```



#Check PCA Summary Information, info of standard deviations and breakdown of varaibles

```
summary(p)
```

```
## Importance of components:
##                           PC1     PC2     PC3     PC4     PC5     PC6
PC7
## Standard deviation     1.5399 1.43046 1.31741 1.19825 1.09996 1.07415
1.0216
## Proportion of Variance 0.1129 0.09744 0.08265 0.06837 0.05762 0.05494
```

```
0.0497
## Cumulative Proportion  0.1129 0.21035 0.29300 0.36137 0.41899 0.47393
0.5236
##                            PC8     PC9    PC10   PC11    PC12   PC13
PC14
## Standard deviation     0.99996 0.98938 0.98595 0.9357 0.91675 0.9003
0.88525
## Proportion of Variance 0.04761 0.04661 0.04629 0.0417 0.04002 0.0386
0.03732
## Cumulative Proportion  0.57125 0.61786 0.66415 0.7058 0.74586 0.7845
0.82178
##                           PC15   PC16   PC17    PC18    PC19    PC20
PC21
## Standard deviation     0.83737 0.7977 0.7844 0.75938 0.71322 0.65669
0.52275
## Proportion of Variance 0.03339 0.0303 0.0293 0.02746 0.02422 0.02054
0.01301
## Cumulative Proportion  0.85517 0.8855 0.9148 0.94223 0.96645 0.98699
1.00000
```

```
print(p)
```

```
## Standard deviations (1, .., p=21):
##  [1] 1.5398853 1.4304555 1.3174076 1.1982462 1.0999649 1.0741516 1.0216320
##  [8] 0.9999556 0.9893754 0.9859451 0.9357353 0.9167492 0.9003431 0.8852468
## [15] 0.8373727 0.7976577 0.7843861 0.7593787 0.7132246 0.6566862 0.5227533
##
## Rotation (n x k) = (21 x 21):
##                                    PC1          PC2          PC3
PC4
## diabetes_risk            -0.017220368  0.447857170 -0.248209293
0.207978880
## hypertension             -0.157399937  0.180304757 -0.166869536 -
0.303809251
## gluten_free               0.045888222 -0.004914329 -0.090744908
0.100921638
## stroke                    0.008912444  0.005212331  0.054258652 -
0.293407253
## diabetes_relatives       -0.009554807  0.437743818 -0.226861369
0.130837355
## moderate_activity         0.152235463 -0.110301805 -0.241147074 -
0.002150052
## diabetes                  0.039592162 -0.129999835  0.124036143
0.082879898
## feel_at_risk_diabetes    -0.019045358  0.436726785 -0.258254211
0.192512060
## frozen_meals              0.039345284 -0.033999054 -0.049494294 -
0.345024609
## gender                   -0.483648064 -0.045319640  0.201554673 -
0.008346201
```

```
## age                          0.048946093 -0.050586217  0.303076437
0.552935741
## education_level              0.103307956  0.328444979  0.411164748 -
0.075956582
## height                       0.533034096  0.055391460 -0.108813057
0.028247009
## weight                       0.422708440 -0.222522718  0.005676195
0.066737007
## sleep_weekdays              -0.099628197  0.158250670  0.020077839 -
0.166624713
## hours_worked                 0.273974790 -0.003272097  0.057190982
0.077214768
## time_outdoors_weekends       0.195225174 -0.089086221 -0.250229900 -
0.042770223
## time_sitting                 0.138935523  0.183061527  0.315569077 -
0.196780134
## take_away_food               0.251108342  0.154614597  0.037067580 -
0.334010921
## poverty_ratio                0.158172977  0.309884975  0.472336978
0.040915222
## take.away.within.30.days     0.075689058  0.052733038  0.038244258 -
0.289366632
##                                     PC5          PC6          PC7         PC8
## diabetes_risk              -0.02849167  0.052843588 -0.155451883 -0.02138625
## hypertension                0.49558827 -0.085847997  0.057056250  0.03078400
## gluten_free                 0.05498639  0.382411227 -0.252492506  0.23943565
## stroke                      0.39718481 -0.226893817 -0.458866519  0.18431094
## diabetes_relatives         -0.14671347  0.003030195 -0.073923615  0.19750189
## moderate_activity          -0.25342448 -0.370362967  0.339171754  0.24992701
## diabetes                    0.06521336  0.102763137  0.009409529  0.72825840
## feel_at_risk_diabetes      -0.05347183  0.011074192 -0.098777913  0.02034176
## frozen_meals               -0.37798117  0.199621305 -0.105905684  0.30323067
## gender                     -0.20765699 -0.107700807 -0.131066532  0.05009178
## age                        -0.07373931  0.068077831 -0.011882791  0.03042372
## education_level             0.02141892 -0.170612424  0.077207996  0.15947232
## height                      0.23060593  0.120748027  0.073769347  0.02155509
## weight                      0.11049444  0.230636605 -0.107020800 -0.02662396
## sleep_weekdays              0.12088922  0.281190338  0.668712385  0.05893075
## hours_worked               -0.03171206 -0.303816664 -0.088772187 -0.25816092
## time_outdoors_weekends     -0.15945215 -0.488956682  0.008387969  0.18904100
## time_sitting               -0.06657262  0.008351011 -0.081646300 -0.02892368
## take_away_food             -0.24849960  0.155433266  0.041793936 -0.16946501
## poverty_ratio               0.01628585 -0.179460743  0.082801341  0.09332767
## take.away.within.30.days -0.36765579  0.156225672 -0.214784623 -0.09455427
##                                    PC9         PC10         PC11         PC12
## diabetes_risk               0.046067134 -0.01208746 -0.06037740  0.10262713
## hypertension                0.178564788  0.01109582  0.06390783 -0.20084480
## gluten_free                 0.023054895 -0.77540957 -0.09860992 -0.25384209
## stroke                     -0.059097286  0.03640418 -0.18180387 -0.05397513
## diabetes_relatives         -0.131908832  0.09311370  0.11554619  0.15868262
```

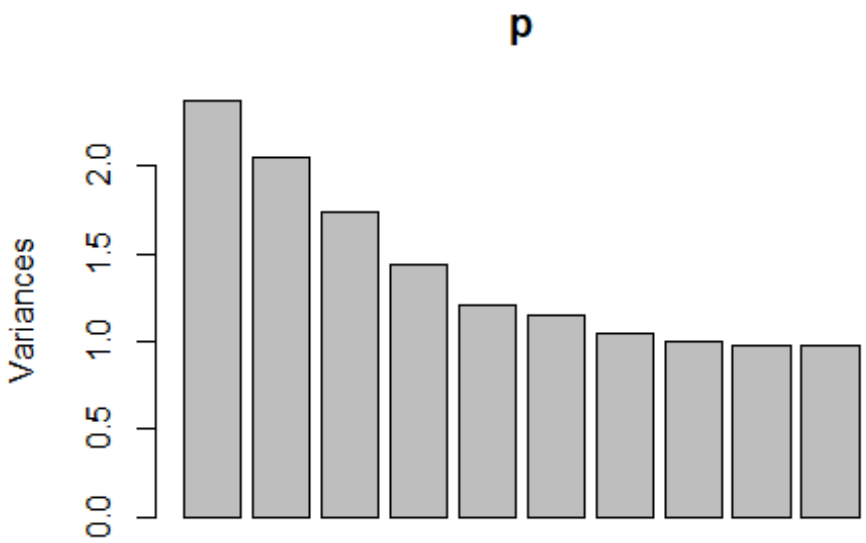
```
## moderate_activity          0.004504032 -0.14617357 -0.24057957 -0.09946653
## diabetes                   0.445355689  0.18145077  0.32946074  0.20123967
## feel_at_risk_diabetes      0.058869947  0.14074832  0.03394804 -0.01934455
## frozen_meals              -0.493577212  0.12208390  0.24663992 -0.21023390
## gender                    -0.075395654 -0.06611251  0.05181498  0.04102951
## age                        0.079390813  0.05935823 -0.20605947 -0.20393822
## education_level           -0.236251284 -0.02207056 -0.10993236 -0.13746017
## height                    -0.096396529  0.14718424 -0.06878327 -0.07990996
## weight                    -0.186197125  0.16985332 -0.06822525  0.17198276
## sleep_weekdays             0.066068924 -0.13454512 -0.03464555  0.02475105
## hours_worked               0.185069854 -0.25328310  0.66671109 -0.09757071
## time_outdoors_weekends     0.063780571 -0.16693778 -0.19953196  0.08617784
## time_sitting               0.037771368 -0.26330081 -0.16175155  0.70348684
## take_away_food             0.159652425 -0.10306345  0.16225645 -0.02633553
## poverty_ratio             -0.006942839  0.01158744  0.01048632 -0.29856580
## take.away.within.30.days   0.563658533  0.22442896 -0.32093977 -0.24741655
##                                   PC13         PC14        PC15        PC16
## diabetes_risk             -0.106071392  0.079901528 -0.02182960 -0.08003657
## hypertension               0.302656517  0.049898922  0.13231636 -0.47974923
## gluten_free                0.115460745 -0.021700424  0.01036869  0.04147934
## stroke                    -0.612266504  0.112813192 -0.05924233  0.10269426
## diabetes_relatives        -0.028209503  0.118507082 -0.19484683 -0.43822539
## moderate_activity         -0.122479016  0.441153034  0.44712777 -0.04825598
## diabetes                   0.066618670  0.066627876 -0.02681245  0.11566031
## feel_at_risk_diabetes     -0.078558255 -0.161386482  0.21776774  0.48332247
## frozen_meals              -0.117243785 -0.235047614  0.18927707 -0.15680274
## gender                     0.022792537  0.101907925 -0.10568343 -0.07484013
## age                       -0.286586892  0.025501629 -0.03842202 -0.36713067
## education_level            0.321964910  0.004986156 -0.07212821  0.18525869
## height                     0.110691275 -0.040261649  0.07038480  0.05753420
## weight                     0.013602647  0.042230650 -0.09151813 -0.20221067
## sleep_weekdays            -0.462963915 -0.287894867 -0.13371853 -0.02782608
## hours_worked              -0.205128131 -0.176482549  0.17200237 -0.12136761
## time_outdoors_weekends     0.086999143 -0.436759536 -0.50615284 -0.05464141
## time_sitting               0.003141285 -0.099346439  0.34218683 -0.14584907
## take_away_food            -0.073278820  0.553070111 -0.43383670  0.10480485
## poverty_ratio              0.006327403 -0.036926675 -0.01361027 -0.02505809
## take.away.within.30.days   0.025333441 -0.217170009  0.12734926 -0.12228222
##                                   PC17         PC18          PC19
PC20
## diabetes_risk             -0.019310298  0.762863267  0.1946938827
0.06862963
## hypertension               0.241786125 -0.083718723  0.2607165360 -
0.10637172
## gluten_free               -0.075152305 -0.090732435 -0.0125090228
0.05750504
## stroke                    -0.079169550 -0.034604644 -0.0632166080 -
0.05432680
## diabetes_relatives        -0.303505704 -0.369611979 -0.3662839228
0.03406710
```

```
## moderate_activity        -0.116470060 -0.013527462  0.1032305278
0.02856909
## diabetes                  0.009641389  0.097992596  0.0219031592 -
0.02351610
## feel_at_risk_diabetes     0.183533462 -0.430307614  0.3537871574 -
0.05876918
## frozen_meals              0.267133781  0.132728576  0.0570433039 -
0.06772278
## gender                   -0.239846360 -0.095257551  0.5110014112
0.08938977
## age                       0.354106449 -0.082730292  0.0781255931 -
0.36437124
## education_level          -0.318675454  0.106162079  0.0492779905 -
0.52290295
## height                   -0.033790594  0.015220083 -0.0376869756 -
0.02338392
## weight                   -0.299050389 -0.116476780  0.5413397077
0.15717070
## sleep_weekdays           -0.190762548 -0.006004149  0.1210388551 -
0.03183013
## hours_worked             -0.226071815  0.022426289  0.0710929710 -
0.14516181
## time_outdoors_weekends    0.185920628 -0.002839871  0.1327748411 -
0.00527242
## time_sitting              0.215772619 -0.068068170  0.0001921243 -
0.02880464
## take_away_food            0.281800923 -0.075897733  0.1174086706 -
0.10316468
## poverty_ratio             0.129258625 -0.019068963 -0.0067955818
0.70256640
## take.away.within.30.days -0.280436543 -0.011370189 -0.0309955913 -
0.02297522
##                                 PC21
## diabetes_risk            -0.011165033
## hypertension             -0.049423638
## gluten_free              -0.013476727
## stroke                    0.015696687
## diabetes_relatives        0.014069400
## moderate_activity        -0.036858487
## diabetes                 -0.002633246
## feel_at_risk_diabetes    -0.080541571
## frozen_meals             -0.012022287
## gender                    0.520688616
## age                       0.047815494
## education_level          -0.162768778
## height                    0.747630912
## weight                   -0.346433073
## sleep_weekdays            0.016981735
## hours_worked              0.012384853
## time_outdoors_weekends    0.016869183
```

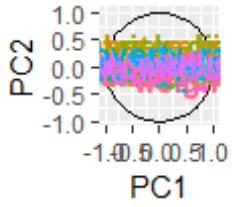
```
## time_sitting               0.069276849
## take_away_food             0.023208141
## poverty_ratio             -0.061408042
## take.away.within.30.days   0.010491016
```

#Check PCA visualizations

```
plot(p) #Scree Plot
```



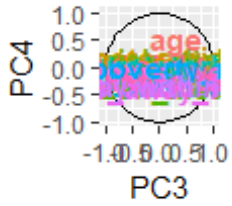

```
PCA_Plot(p) #PCA_plot1
```



.names

a age	a hypertension
a diabetes	a moderate_activity
a diabetes_relatives	a poverty_ratio
a diabetes_risk	a sleep_weekdays
a education_level	a stroke
a feel_at_risk_diabetes	a take.away.within.30.days
a frozen_meals	a take_away_food
a gender	a time_outdoors_weekends
a gluten_free	a time_sitting
a height	a weight
a hours_worked	

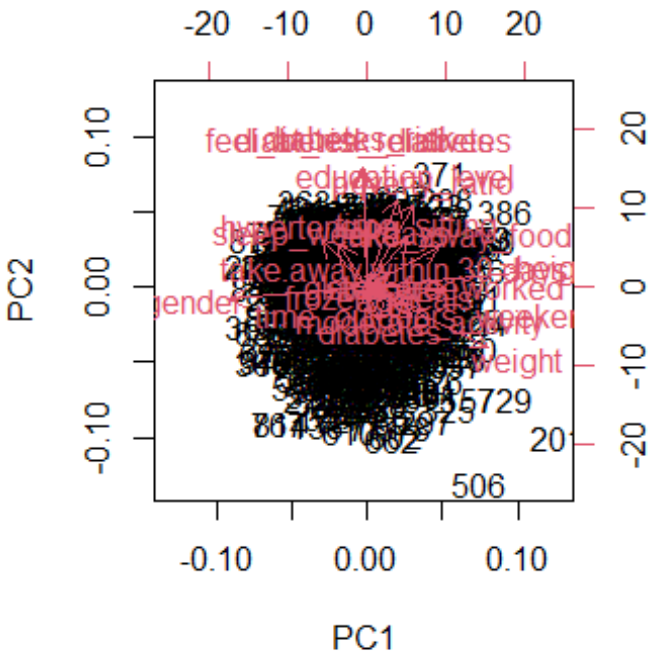
```
PCA_Plot_Secondary(p) #PCA_Plot2
```



.names

a age	a hypertension
a diabetes	a moderate_activity
a diabetes_relatives	a poverty_ratio
a diabetes_risk	a sleep_weekdays
a education_level	a stroke
a feel_at_risk_diabetes	a take.away.within.30.days
a frozen_meals	a take_away_food
a gender	a time_outdoors_weekends
a gluten_free	a time_sitting
a height	a weight
a hours_worked	

```
biplot(p) #Biplot
```



biplot gives data points and vectors at the same time, useful for outliers

```
p2 = psych::principal(training, rotate="varimax", nfactors=6, scores=TRUE)
p2

## Principal Components Analysis
## Call: psych::principal(r = training, nfactors = 6, rotate = "varimax",
##     scores = TRUE)
## Standardized loadings (pattern matrix) based upon correlation matrix
##                             RC1   RC2   RC3   RC5   RC6   RC4    h2   u2
com
## diabetes_risk              0.05  0.75  0.02 -0.02 -0.08 -0.07 0.584 0.42
1.1
## hypertension              -0.05  0.16 -0.04  0.72 -0.19 -0.15 0.612 0.39
1.4
## gluten_free                0.24  0.07 -0.20 -0.14 -0.30  0.03 0.206 0.79
3.4
## stroke                     0.04 -0.18  0.17  0.55  0.03 -0.12 0.379 0.62
1.6
## diabetes_relatives        -0.01  0.72  0.06 -0.03  0.00  0.06 0.532 0.47
1.0
## moderate_activity          0.06  0.06 -0.16 -0.03  0.61  0.11 0.417 0.58
1.2
## diabetes                   0.09 -0.22  0.01 -0.13 -0.09 -0.09 0.092 0.91
2.9
```

```
## feel_at_risk_diabetes     0.03  0.75  0.02 -0.01 -0.03 -0.06 0.564 0.44
1.0
## frozen_meals             -0.04 -0.05 -0.09 -0.02 -0.01  0.62 0.400 0.60
1.1
## gender                   -0.81 -0.11  0.00 -0.08 -0.14 -0.04 0.695 0.30
1.1
## age                       0.04 -0.08  0.22 -0.64 -0.08 -0.38 0.621 0.38
2.0
## education_level          -0.03  0.09  0.75  0.05 -0.04  0.02 0.582 0.42
1.0
## height                    0.87  0.06  0.12  0.04  0.11  0.02 0.783 0.22
1.1
## weight                    0.69 -0.30 -0.05 -0.18  0.04  0.04 0.608 0.39
1.5
## sleep_weekdays           -0.03  0.11  0.03  0.18 -0.40  0.13 0.224 0.78
1.8
## hours_worked              0.25 -0.03  0.25 -0.07  0.40 -0.07 0.300 0.70
2.6
## time_outdoors_weekends    0.11  0.06 -0.09  0.10  0.70  0.05 0.525 0.48
1.2
## time_sitting              0.06 -0.05  0.53  0.03 -0.08  0.24 0.348 0.65
1.5
## take_away_food            0.25  0.08  0.24  0.05  0.01  0.58 0.464 0.54
1.8
## poverty_ratio             0.03  0.06  0.82 -0.08 -0.02 -0.06 0.683 0.32
1.0
## take.away.within.30.days -0.02  0.00  0.08 -0.06 -0.01  0.57 0.334 0.67
1.1
##
##                        RC1  RC2  RC3  RC5  RC6  RC4
## SS loadings           2.11 1.92 1.82 1.38 1.37 1.35
## Proportion Var        0.10 0.09 0.09 0.07 0.07 0.06
## Cumulative Var        0.10 0.19 0.28 0.34 0.41 0.47
## Proportion Explained  0.21 0.19 0.18 0.14 0.14 0.14
## Cumulative Proportion 0.21 0.40 0.59 0.73 0.86 1.00
##
## Mean item complexity =  1.5
## Test of the hypothesis that 6 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.08
##  with the empirical chi square  2096.42  with prob <  0
##
## Fit based upon off diagonal values = 0.57
```

different component loadings

```
print(p2$loadings, cutoff=.4, sort=T)

##
## Loadings:
```

```
##                            RC1    RC2    RC3    RC5    RC6    RC4
## gender                    -0.809
## height                     0.866
## weight                     0.693
## diabetes_risk                     0.754
## diabetes_relatives                0.724
## feel_at_risk_diabetes             0.747
## education_level                          0.754
## time_sitting                             0.526
## poverty_ratio                            0.816
## hypertension                                    0.723
## stroke                                          0.548
## age                                            -0.644
## moderate_activity                                      0.610
## time_outdoors_weekends                                 0.698
## frozen_meals                                                  0.622
## take_away_food                                                0.579
## take.away.within.30.days                                      0.568
## gluten_free
## diabetes
## sleep_weekdays                                        -0.402
## hours_worked                                           0.405
##
##                  RC1   RC2   RC3   RC5   RC6   RC4
## SS loadings    2.11 1.917 1.816 1.384 1.374 1.351
## Proportion Var 0.10 0.091 0.086 0.066 0.065 0.064
## Cumulative Var 0.10 0.192 0.278 0.344 0.410 0.474
```

#PCAs Other Available Information # looks at functions available in the PCA

```
ls(p2)
```

```
##  [1] "Call"         "chi"          "communality"  "complexity"
"criteria"
##  [6] "dof"          "EPVAL"        "factors"      "fit"          "fit.off"
## [11] "fn"           "loadings"     "n.obs"        "null.dof"
"null.model"
## [16] "objective"    "PVAL"         "r.scores"     "R2"
"residual"
## [21] "rms"          "rot.mat"      "rotation"     "scores"
"STATISTIC"
## [26] "Structure"    "uniquenesses" "Vaccounted"   "values"       "weights"
```

looks at values (eigenvalues), commonality of shared varaince, rotated matrix

```
p2$values
```

```
##  [1] 2.3712468 2.0462030 1.7355628 1.4357939 1.2099227 1.1538017 1.0437320
##  [8] 0.9999111 0.9788637 0.9720877 0.8756006 0.8404290 0.8106177 0.7836619
## [15] 0.7011930 0.6362578 0.6152616 0.5766560 0.5086893 0.4312368 0.2732710
```

```
p2$communality
```

```
##            diabetes_risk             hypertension              gluten_free
##                0.5843565                0.6117898                0.2063464
##                   stroke       diabetes_relatives        moderate_activity
##                0.3792287                0.5322643                0.4167547
##                 diabetes    feel_at_risk_diabetes             frozen_meals
##                0.0921919                0.5636995                0.4000463
##                   gender                      age          education_level
##                0.6950370                0.6212407                0.5818760
##                   height                   weight           sleep_weekdays
##                0.7828698                0.6076178                0.2242537
##             hours_worked   time_outdoors_weekends             time_sitting
##                0.2999679                0.5245248                0.3482179
##           take_away_food            poverty_ratio take.away.within.30.days
##                0.4635928                0.6829116                0.3337428
```

```
p2$rot.mat
```

```
##             [,1]         [,2]        [,3]        [,4]       [,5]
[,6]
## [1,]  0.87227124 -0.049458854  0.27354587  0.18516153  0.1206853
0.33618395
## [2,] -0.02316378  0.822962863  0.49957427  0.06806148 -0.1612239 -
0.20492828
## [3,] -0.19240393 -0.484913635  0.77204847 -0.05711768  0.2487411 -
0.25815903
## [4,] -0.07543167 -0.251393970  0.08418285  0.66994361 -0.6886253 -
0.03154677
## [5,] -0.32854072  0.148064506 -0.02255455  0.64385639  0.5906358
0.32592561
## [6,] -0.29657734  0.004763232  0.26822360 -0.30731892 -0.2729733
0.81921668
```

#Calculating scores # HW Q4 part c - calcualte scores

```
scores <- p2$scores
```

#summary scores for each component

```
summary(scores)
```

```
##       RC1                 RC2                 RC3                 RC5
##  Min.   :-2.79834   Min.   :-3.2459   Min.   :-3.4178   Min.   :-8.4179
##  1st Qu.:-0.83808   1st Qu.:-0.6796   1st Qu.:-0.7694   1st Qu.:-0.3919
##  Median : 0.05309   Median : 0.4325   Median : 0.0846   Median : 0.1810
```

```
##  Mean   : 0.00000   Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000
##  3rd Qu.: 0.80043   3rd Qu.: 0.7734   3rd Qu.: 0.8036   3rd Qu.: 0.6585
##  Max.   : 3.06125   Max.   : 2.9813   Max.   : 2.5242   Max.   : 1.9325
##       RC6               RC4
##  Min.   :-2.6278   Min.   :-1.7595
##  1st Qu.:-0.6668   1st Qu.:-0.6116
##  Median :-0.2030   Median :-0.2090
##  Mean   : 0.0000   Mean   : 0.0000
##  3rd Qu.: 0.5544   3rd Qu.: 0.3751
##  Max.   : 6.9368   Max.   : 8.6890
```

```
cor(scores)
```

```
##                RC1           RC2           RC3           RC5           RC6
## RC1  1.000000e+00 -1.393892e-16 -1.570311e-16 -3.168024e-16  4.549982e-16
## RC2 -1.393892e-16  1.000000e+00 -1.184039e-16 -3.369649e-16  1.182636e-16
## RC3 -1.570311e-16 -1.184039e-16  1.000000e+00  8.047279e-17 -6.943235e-16
## RC5 -3.168024e-16 -3.369649e-16  8.047279e-17  1.000000e+00  7.304331e-16
## RC6  4.549982e-16  1.182636e-16 -6.943235e-16  7.304331e-16  1.000000e+00
## RC4  1.401812e-16 -1.034681e-16  2.124222e-17 -1.838211e-16 -9.019769e-16
##                RC4
## RC1  1.401812e-16
## RC2 -1.034681e-16
## RC3  2.124222e-17
## RC5 -1.838211e-16
## RC6 -9.019769e-16
## RC4  1.000000e+00
```