

Final_Project

Team 4

2024-10-09

```
##Here is where we load necessary libraries

library(dplyr) #For data manipulation

## 
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
## 
##     filter, lag

## The following objects are masked from 'package:base':
## 
##     intersect, setdiff, setequal, union

library(MASS) #For Stepwise regression

## 
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
## 
##     select

library(glmnet) #For Lasso regression

## Loading required package: Matrix

## Loaded glmnet 4.1-8

library(pROC) #For ROC analysis

## Type 'citation("pROC")' for a citation.

## 
## Attaching package: 'pROC'

## The following objects are masked from 'package:stats':
## 
##     cov, smooth, var
```

```

library(car)

## Loading required package: carData

##
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
##      recode

```

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```

##Read in the diabetes dataset from the specified file path
diabetes_data <- read.csv("/Users/rahulsingh/Downloads/DSC 423/Group Project/data set/diabetes_binary_ho

##Explore the dataset

##Display the feature names and first six rows of the dataset as a quick review to familiarize ourselves
head(diabetes_data)

```

```

##   Diabetes_binary HighBP HighChol CholCheck BMI Smoker Stroke
## 1              0     1       1        1  40     1     0
## 2              0     0       0        0  25     1     0
## 3              0     1       1        1  28     0     0
## 4              0     1       0        1  27     0     0
## 5              0     1       1        1  24     0     0
## 6              0     1       1        1  25     1     0
##   HeartDiseaseorAttack PhysActivity Fruits Veggies HvyAlcoholConsump
## 1                  0          0     0     1             0
## 2                  0          1     0     0             0
## 3                  0          0     1     0             0
## 4                  0          1     1     1             0
## 5                  0          1     1     1             0
## 6                  0          1     1     1             0
##   AnyHealthcare NoDocbcCost GenHlth MentHlth PhysHlth DiffWalk Sex Age
## 1              1          0     5    18     15     1   0   9
## 2              0          1     3     0     0     0   0   7
## 3              1          1     5    30     30     1   0   9
## 4              1          0     2     0     0     0   0  11
## 5              1          0     2     3     0     0   0  11
## 6              1          0     2     0     2     0   1  10
##   Education Income
## 1        4     3
## 2        6     1
## 3        4     8

```

```
## 4      3      6
## 5      5      4
## 6      6      8
```

```
##Retrieve the dimensions of our data set (number of rows and columns)
dim(diabetes_data)
```

```
## [1] 253680     22
```

```
##Review the structure of the data, including data types for each variable
str(diabetes_data)
```

```
## 'data.frame': 253680 obs. of 22 variables:
##   $ Diabetes_binary : num  0 0 0 0 0 0 0 0 1 0 ...
##   $ HighBP          : num  1 0 1 1 1 1 1 1 0 ...
##   $ HighChol         : num  1 0 1 0 1 1 0 1 1 0 ...
##   $ CholCheck        : num  1 0 1 1 1 1 1 1 1 1 ...
##   $ BMI              : num  40 25 28 27 24 25 25 30 25 30 24 ...
##   $ Smoker            : num  1 1 0 0 0 1 1 1 1 0 ...
##   $ Stroke            : num  0 0 0 0 0 0 0 0 0 0 ...
##   $ HeartDiseaseorAttack: num  0 0 0 0 0 0 0 0 1 0 ...
##   $ PhysActivity       : num  0 1 0 1 1 1 0 1 0 0 ...
##   $ Fruits             : num  0 0 1 1 1 1 0 0 1 0 ...
##   $ Veggies            : num  1 0 0 1 1 1 0 1 1 1 ...
##   $ HvyAlcoholConsump  : num  0 0 0 0 0 0 0 0 0 0 ...
##   $ AnyHealthcare       : num  1 0 1 1 1 1 1 1 1 1 ...
##   $ NoDocbcCost         : num  0 1 1 0 0 0 0 0 0 0 ...
##   $ GenHlth             : num  5 3 5 2 2 2 3 3 5 2 ...
##   $ MentHlth            : num  18 0 30 0 3 0 0 0 30 0 ...
##   $ PhysHlth             : num  15 0 30 0 0 2 14 0 30 0 ...
##   $ DiffWalk            : num  1 0 1 0 0 0 0 1 1 0 ...
##   $ Sex                 : num  0 0 0 0 0 1 0 0 0 1 ...
##   $ Age                 : num  9 7 9 11 11 10 9 11 9 8 ...
##   $ Education            : num  4 6 4 3 5 6 6 4 5 4 ...
##   $ Income               : num  3 1 8 6 4 8 7 4 1 3 ...
```

```
##Create a summary of all variables in the dataset to see key statistics
summary(diabetes_data)
```

```
## Diabetes_binary      HighBP      HighChol      CholCheck
## Min.   :0.0000  Min.   :0.000  Min.   :0.0000  Min.   :0.0000
## 1st Qu.:0.0000  1st Qu.:0.000  1st Qu.:0.0000  1st Qu.:1.0000
## Median :0.0000  Median :0.000  Median :0.0000  Median :1.0000
## Mean   :0.1393  Mean   :0.429  Mean   :0.4241  Mean   :0.9627
## 3rd Qu.:0.0000  3rd Qu.:1.000  3rd Qu.:1.0000  3rd Qu.:1.0000
## Max.   :1.0000  Max.   :1.000  Max.   :1.0000  Max.   :1.0000
## 
##           BMI          Smoker      Stroke      HeartDiseaseorAttack
## Min.   :12.00    Min.   :0.0000  Min.   :0.00000  Min.   :0.00000
## 1st Qu.:24.00    1st Qu.:0.0000  1st Qu.:0.00000  1st Qu.:0.00000
## Median :27.00    Median :0.0000  Median :0.00000  Median :0.00000
## Mean   :28.38    Mean   :0.4432  Mean   :0.04057  Mean   :0.09419
## 3rd Qu.:31.00    3rd Qu.:1.0000  3rd Qu.:0.00000  3rd Qu.:0.00000
```

```

##  Max.   :98.00   Max.   :1.0000   Max.   :1.00000   Max.   :1.00000
##  PhysActivity      Fruits       Veggies      HvyAlcoholConsump
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:1.0000  1st Qu.:0.0000  1st Qu.:1.0000  1st Qu.:0.0000
##  Median :1.0000  Median :1.0000  Median :1.0000  Median :0.0000
##  Mean   :0.7565  Mean   :0.6343  Mean   :0.8114  Mean   :0.0562
##  3rd Qu.:1.0000  3rd Qu.:1.0000  3rd Qu.:1.0000  3rd Qu.:0.0000
##  Max.   :1.0000  Max.   :1.0000  Max.   :1.0000  Max.   :1.0000
##  AnyHealthcare    NoDocbcCost  GenHlth      MenthHlth
##  Min.   :0.0000  Min.   :0.00000  Min.   :1.000  Min.   : 0.000
##  1st Qu.:1.0000  1st Qu.:0.00000 1st Qu.:2.000  1st Qu.: 0.000
##  Median :1.0000  Median :0.00000  Median :2.000  Median : 0.000
##  Mean   :0.9511  Mean   :0.08418  Mean   :2.511  Mean   : 3.185
##  3rd Qu.:1.0000  3rd Qu.:0.00000 3rd Qu.:3.000  3rd Qu.: 2.000
##  Max.   :1.0000  Max.   :1.00000  Max.   :5.000  Max.   :30.000
##  PhysHlth        DiffWalk     Sex          Age
##  Min.   : 0.000  Min.   :0.0000  Min.   :0.0000  Min.   : 1.000
##  1st Qu.: 0.000  1st Qu.:0.0000  1st Qu.:0.0000  1st Qu.: 6.000
##  Median : 0.000  Median :0.0000  Median :0.0000  Median : 8.000
##  Mean   : 4.242  Mean   :0.1682  Mean   :0.4403  Mean   : 8.032
##  3rd Qu.: 3.000  3rd Qu.:0.0000  3rd Qu.:1.0000  3rd Qu.:10.000
##  Max.   :30.000  Max.   :1.0000  Max.   :1.0000  Max.   :13.000
##  Education       Income
##  Min.   :1.00  Min.   :1.000
##  1st Qu.:4.00  1st Qu.:5.000
##  Median :5.00  Median :7.000
##  Mean   :5.05  Mean   :6.054
##  3rd Qu.:6.00  3rd Qu.:8.000
##  Max.   :6.00  Max.   :8.000

```

```

##Checking all columns in the data set to see if any have NA values. This data set is complete and has no missing values.
colSums(is.na(diabetes_data))

```

```

##  Diabetes_binary      HighBP      HighChol
##  0                      0                      0
##  CholCheck            BMI        Smoker
##  0                      0                      0
##  Stroke   HeartDiseaseorAttack  PhysActivity
##  0                      0                      0
##  Fruits       Veggies      HvyAlcoholConsump
##  0                      0                      0
##  AnyHealthcare  NoDocbcCost  GenHlth
##  0                      0                      0
##  MenthHlth      PhysHlth      DiffWalk
##  0                      0                      0
##  Sex           Age          Education
##  0                      0                      0
##  Income
##  0

```

```

##Creating box plots of each independent variable to look at the spread of data for each variable. A significant outlier is present in the age variable.

```

```

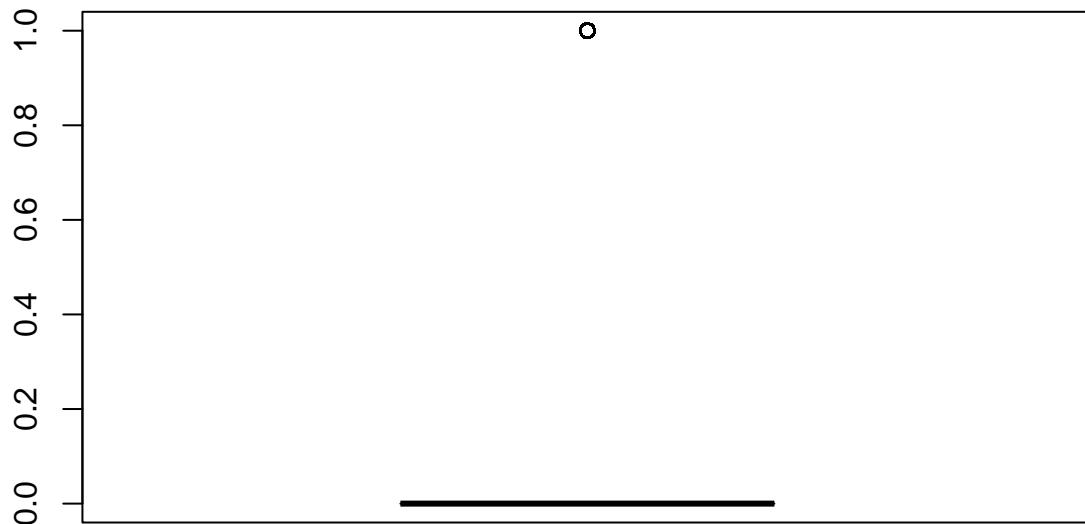
##Create a vector of colors using the rainbow function

```

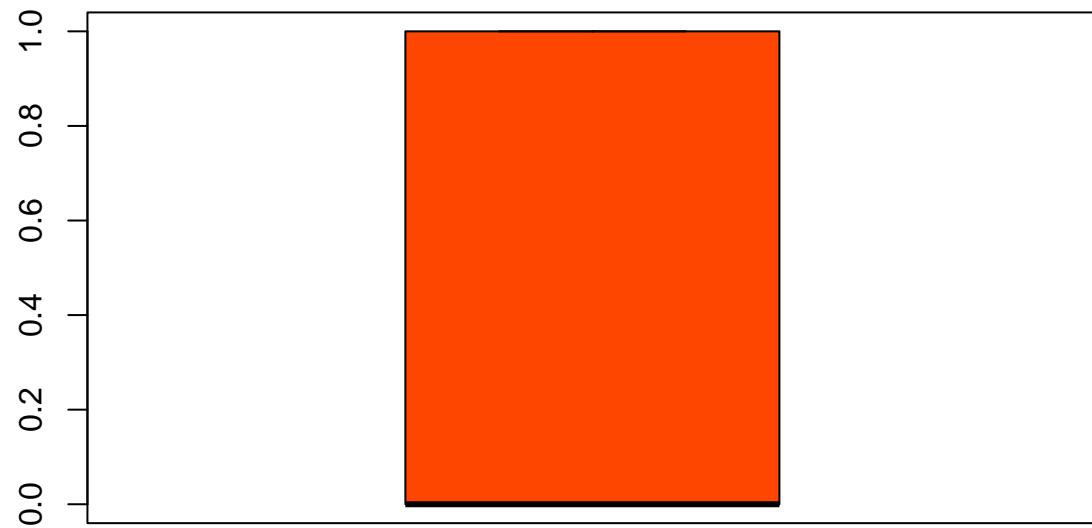
```
colors <- rainbow(length(names(diabetes_data)))

##Generate the boxplots
for (i in seq_along(names(diabetes_data))) {
  var <- names(diabetes_data)[i]
  boxplot(diabetes_data[[var]], main = paste("Boxplot of", var), col=colors[i])
}
```

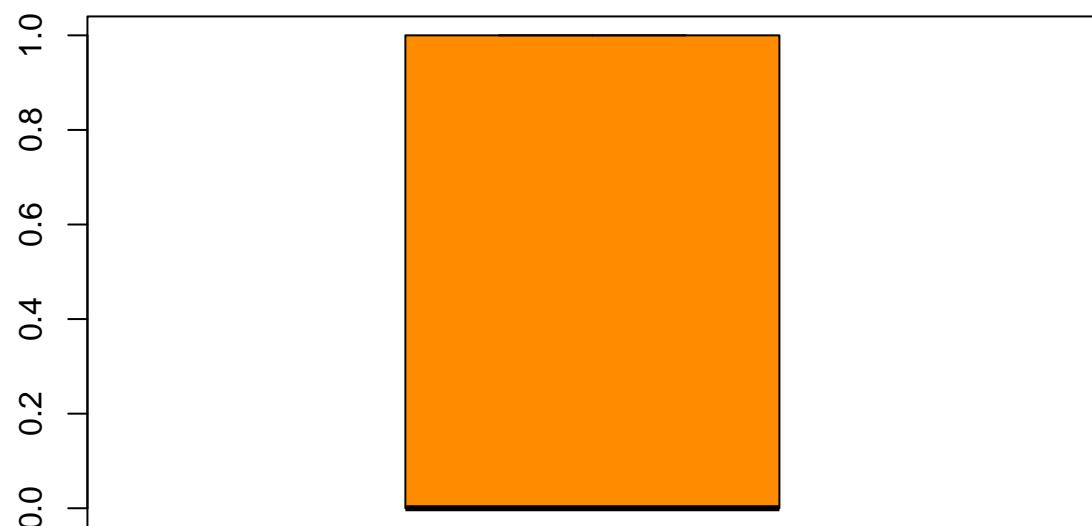
Boxplot of Diabetes_binary



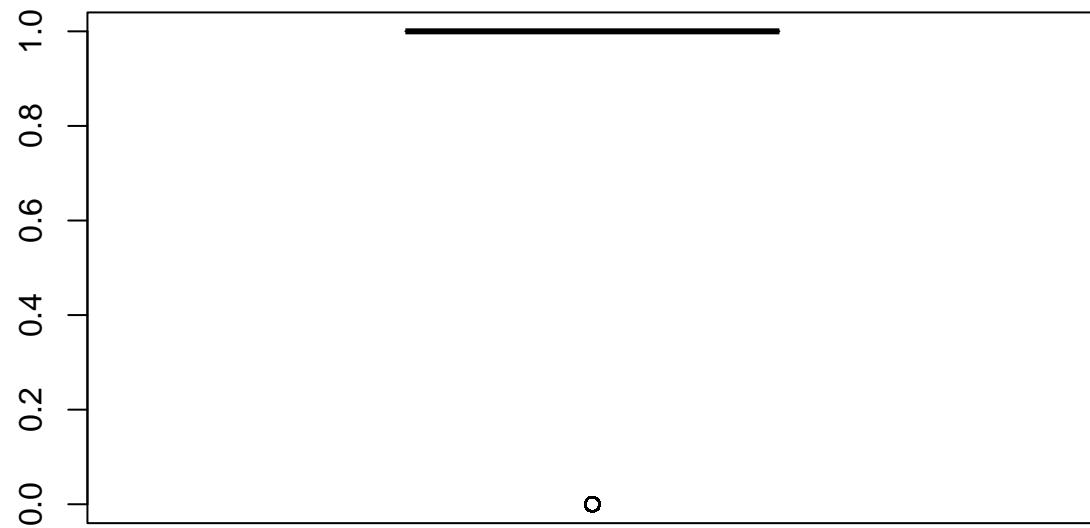
Boxplot of HighBP



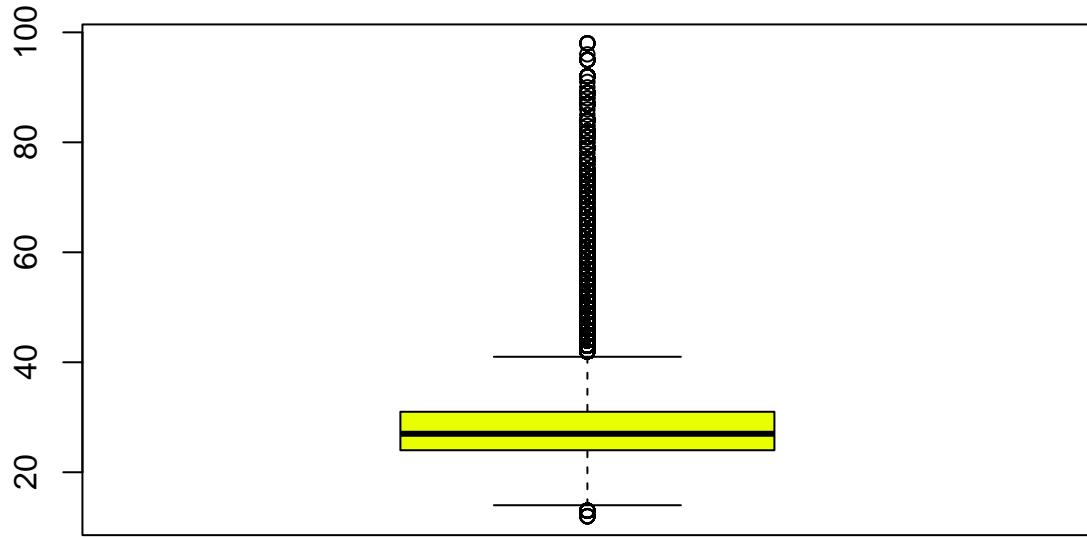
Boxplot of HighChol



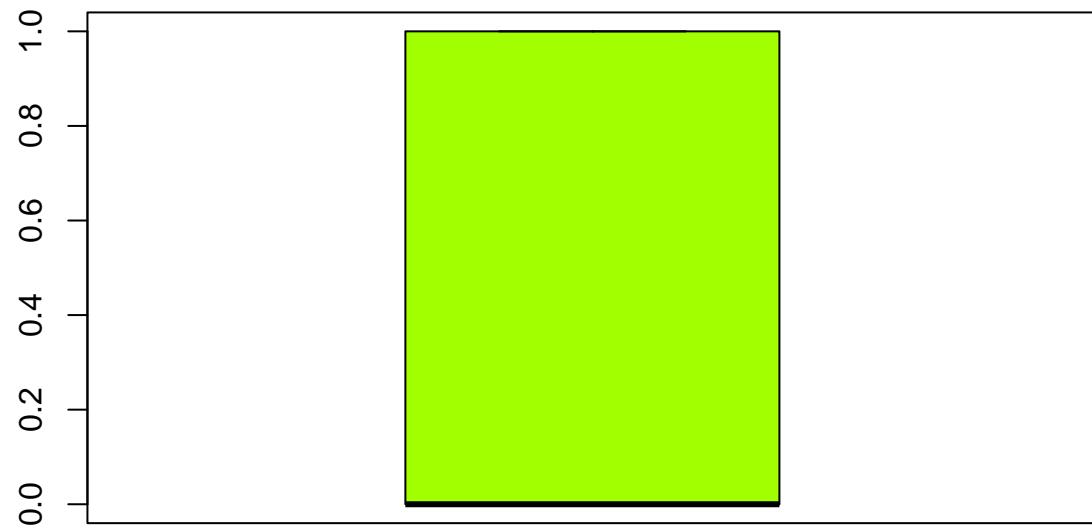
Boxplot of CholCheck



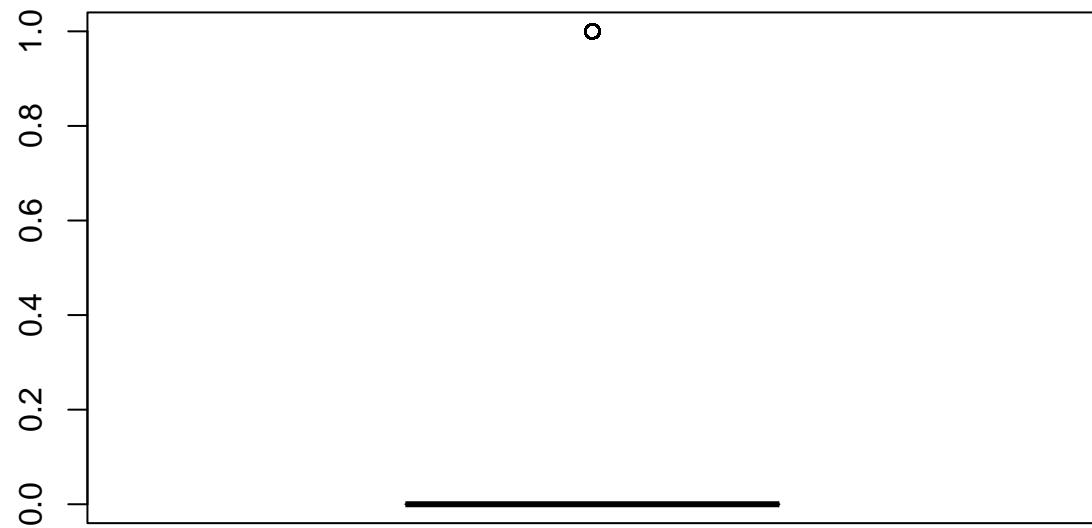
Boxplot of BMI



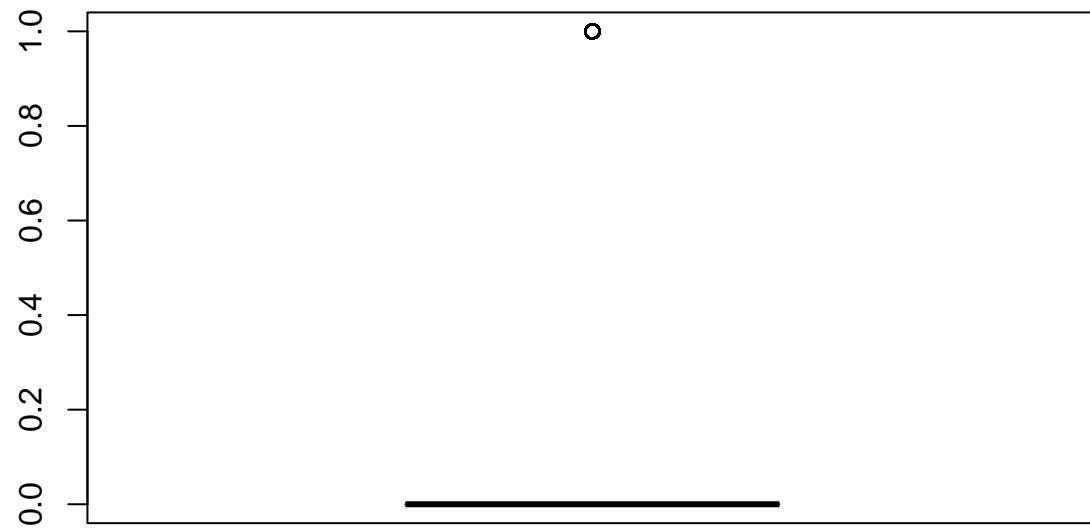
Boxplot of Smoker



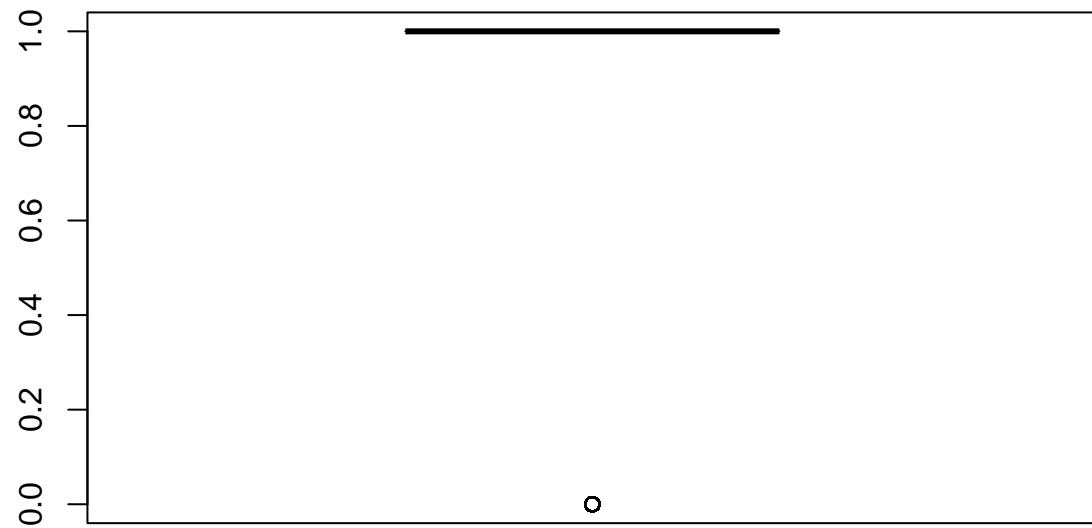
Boxplot of Stroke



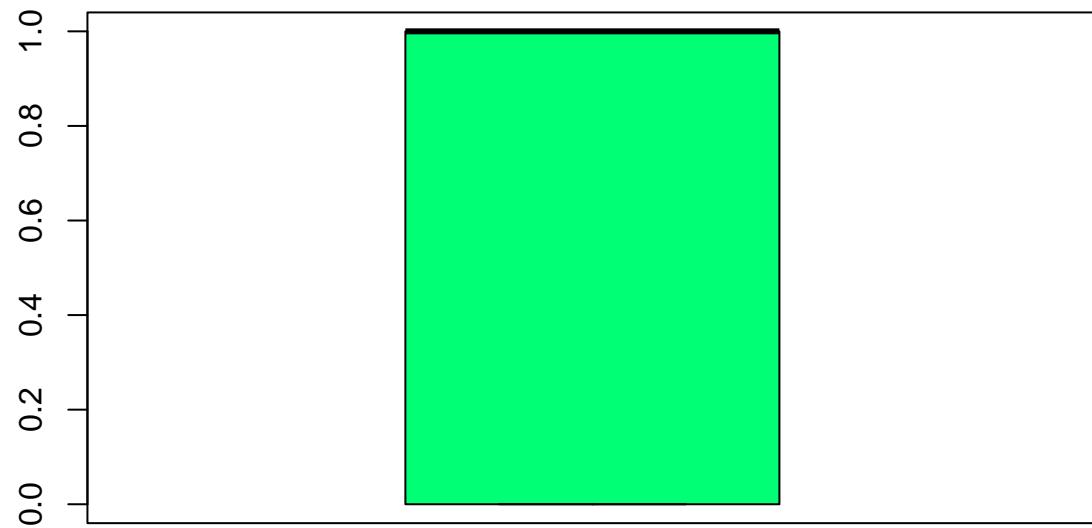
Boxplot of HeartDiseaseorAttack



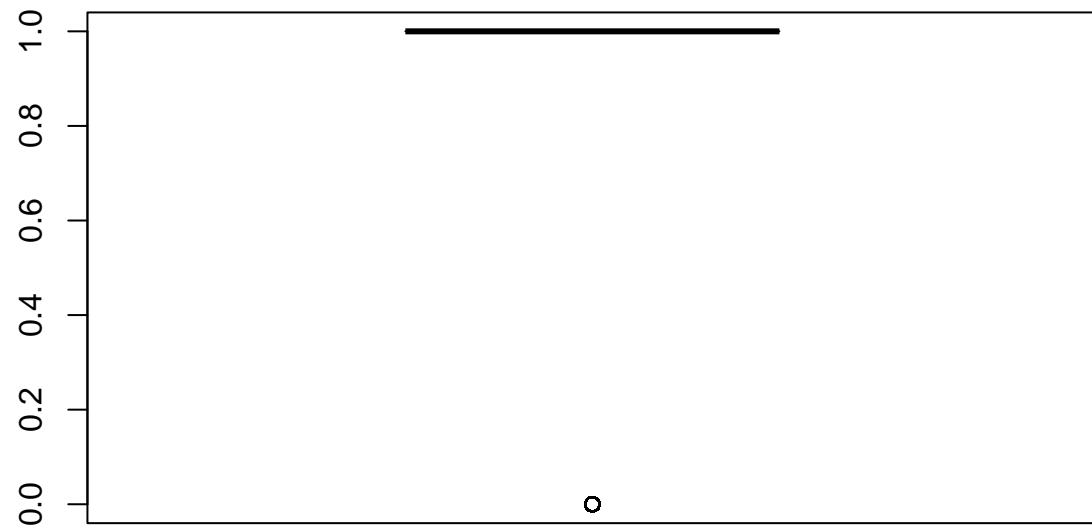
Boxplot of PhysActivity



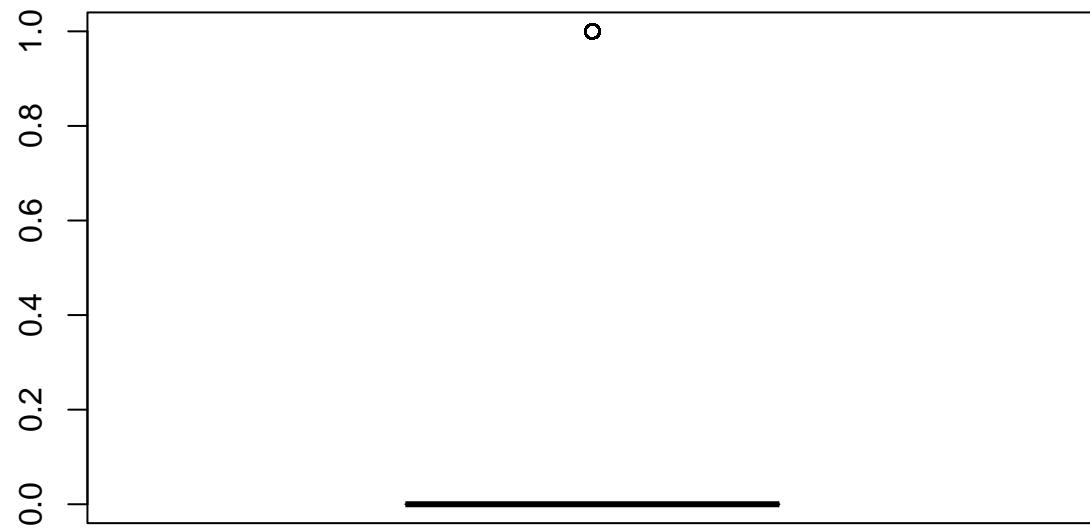
Boxplot of Fruits



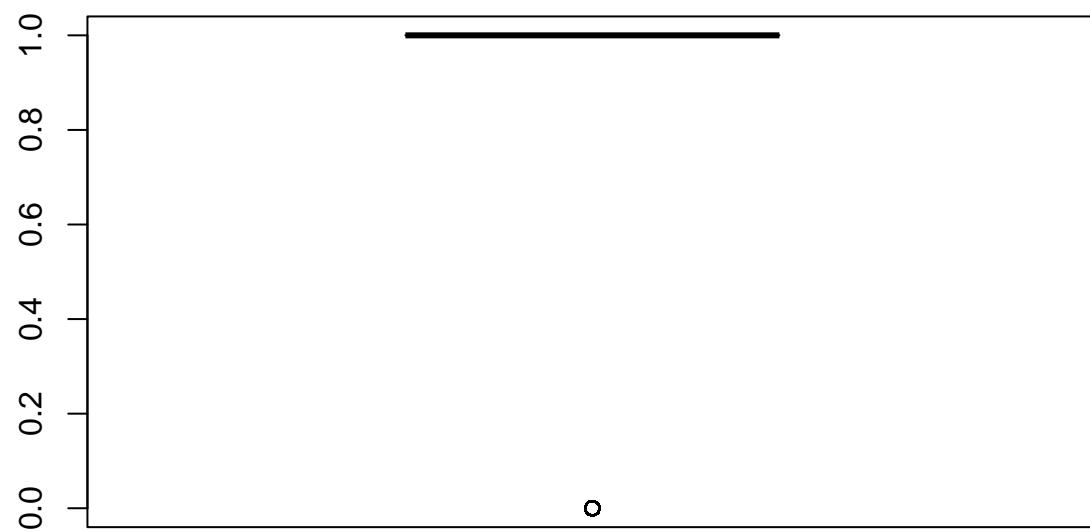
Boxplot of Veggies



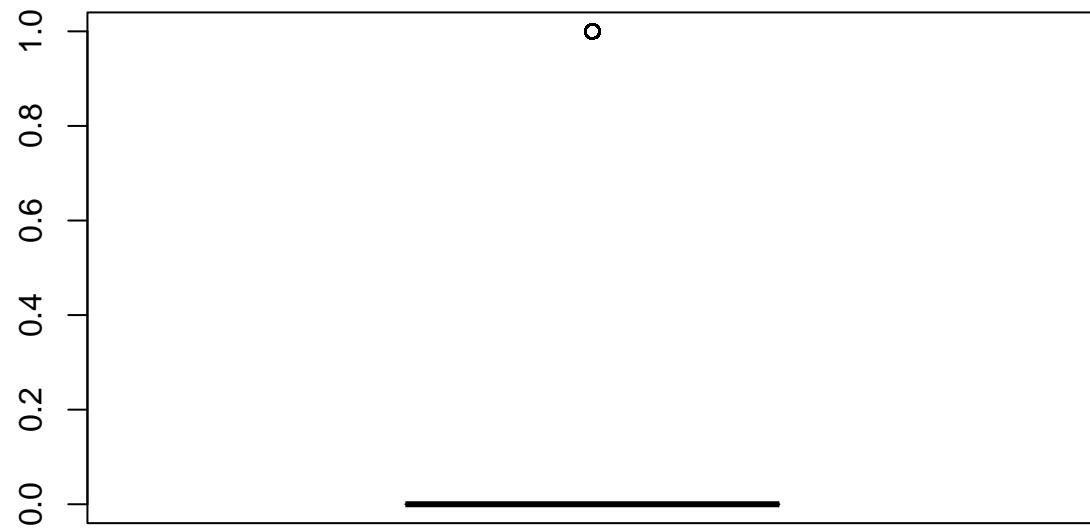
Boxplot of HvyAlcoholConsump



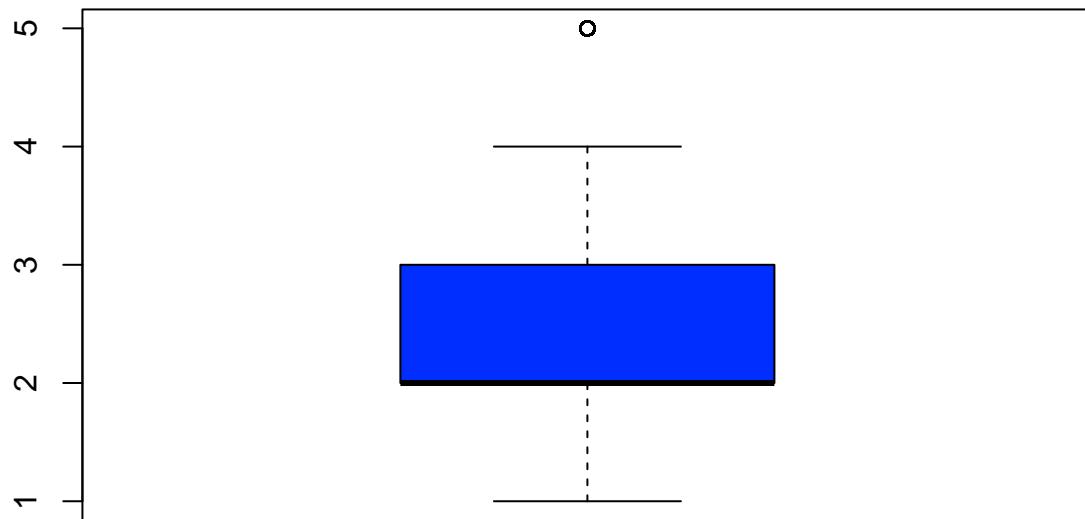
Boxplot of AnyHealthcare



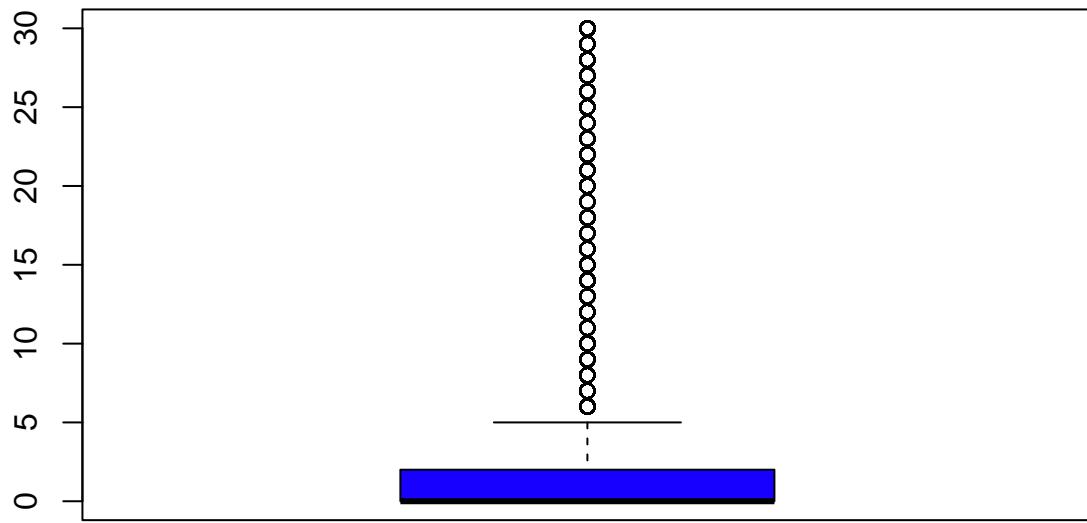
Boxplot of NoDocbcCost



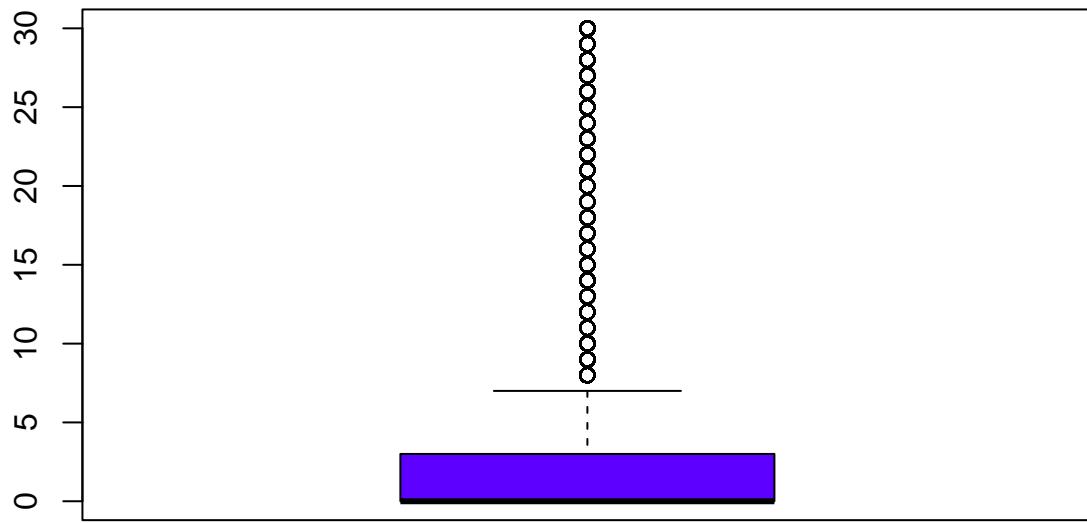
Boxplot of GenHlth



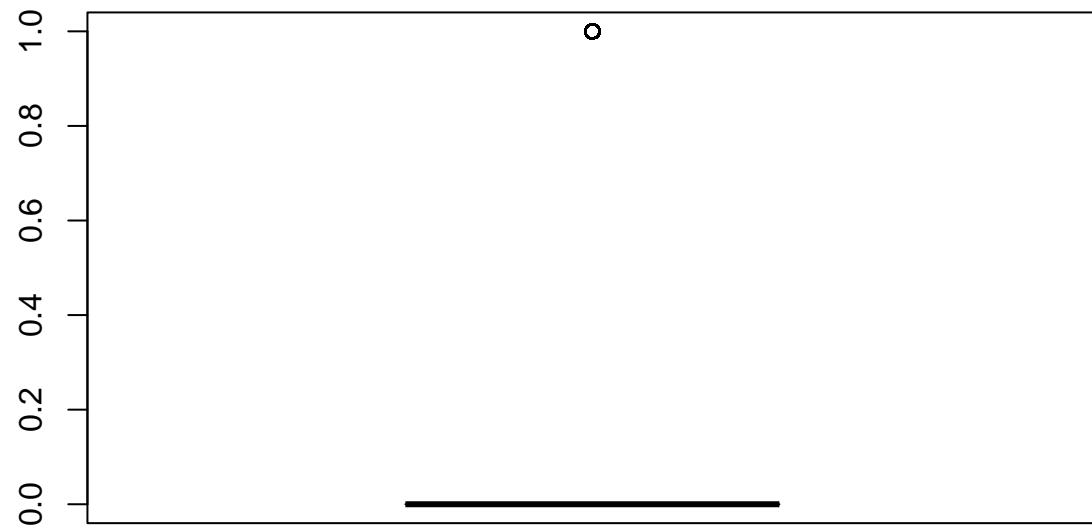
Boxplot of MentHlth



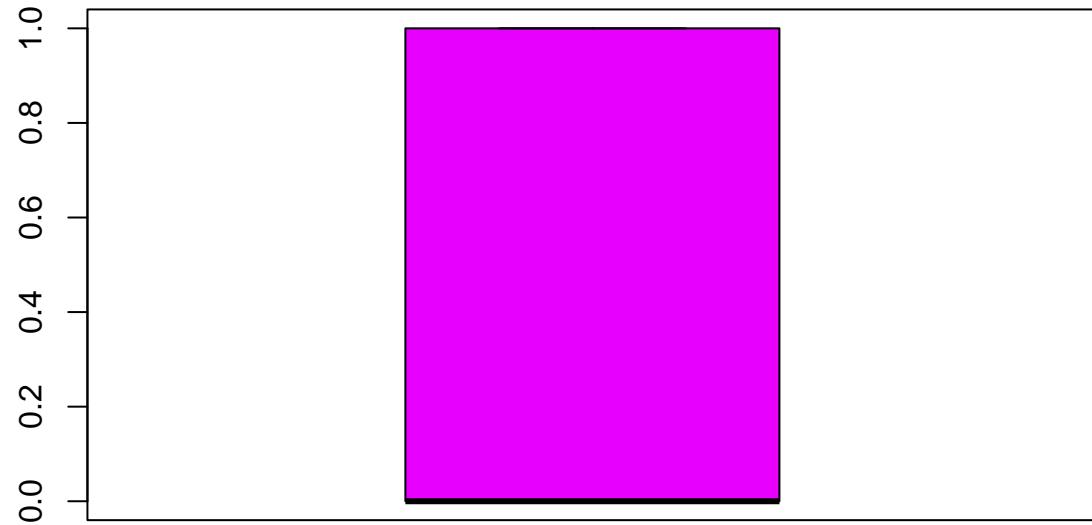
Boxplot of PhysHlth



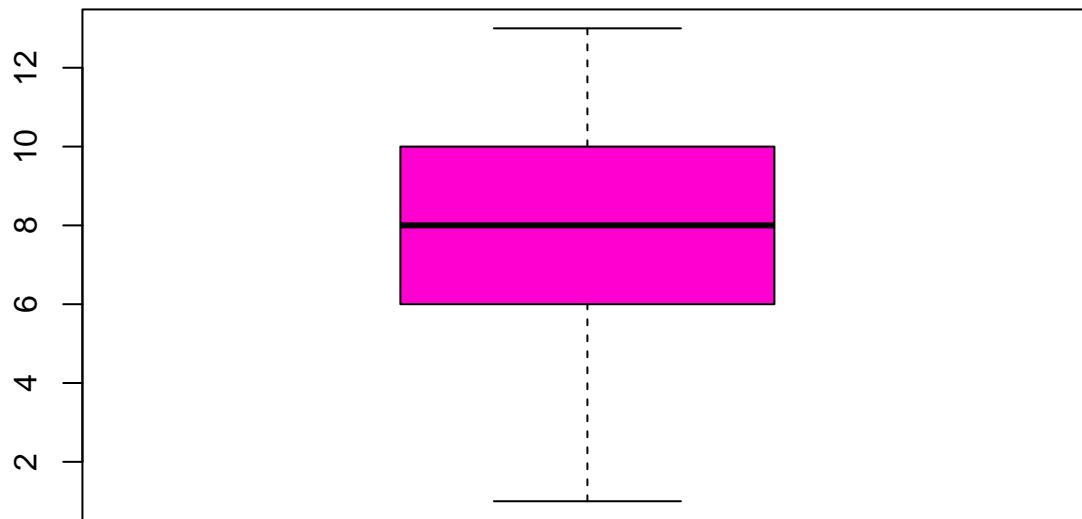
Boxplot of DiffWalk



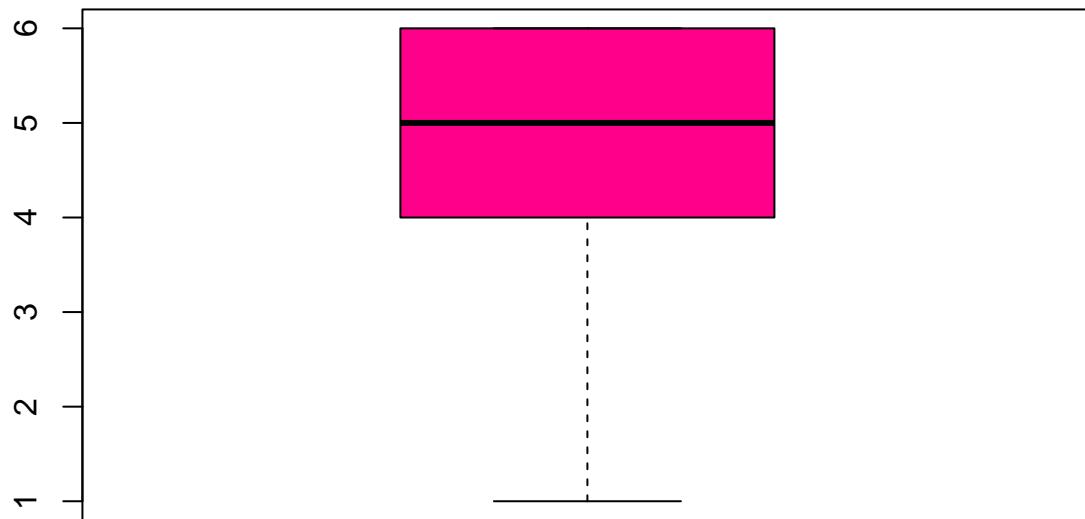
Boxplot of Sex



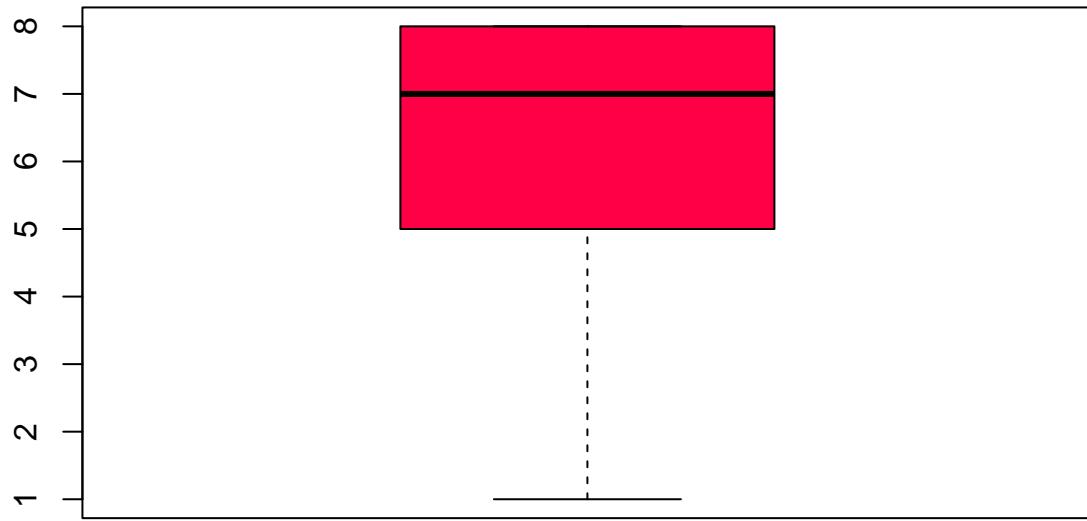
Boxplot of Age



Boxplot of Education

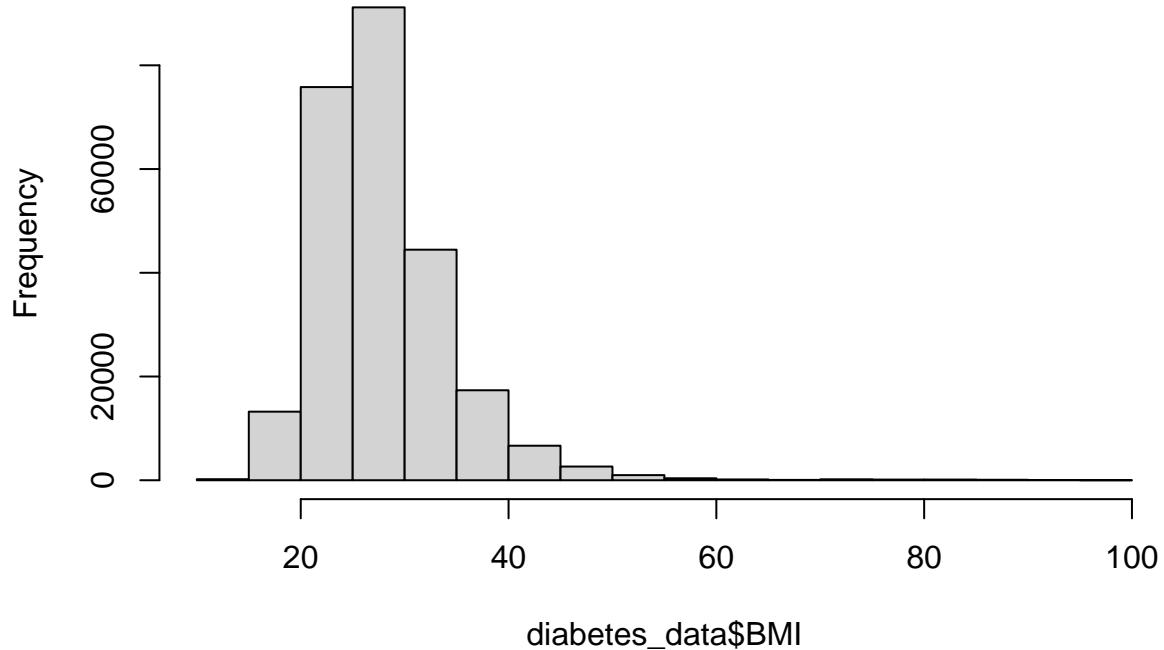


Boxplot of Income



```
##Look at histograms of variables with outliers.  
##Histogram of BMI  
hist(diabetes_data$BMI)
```

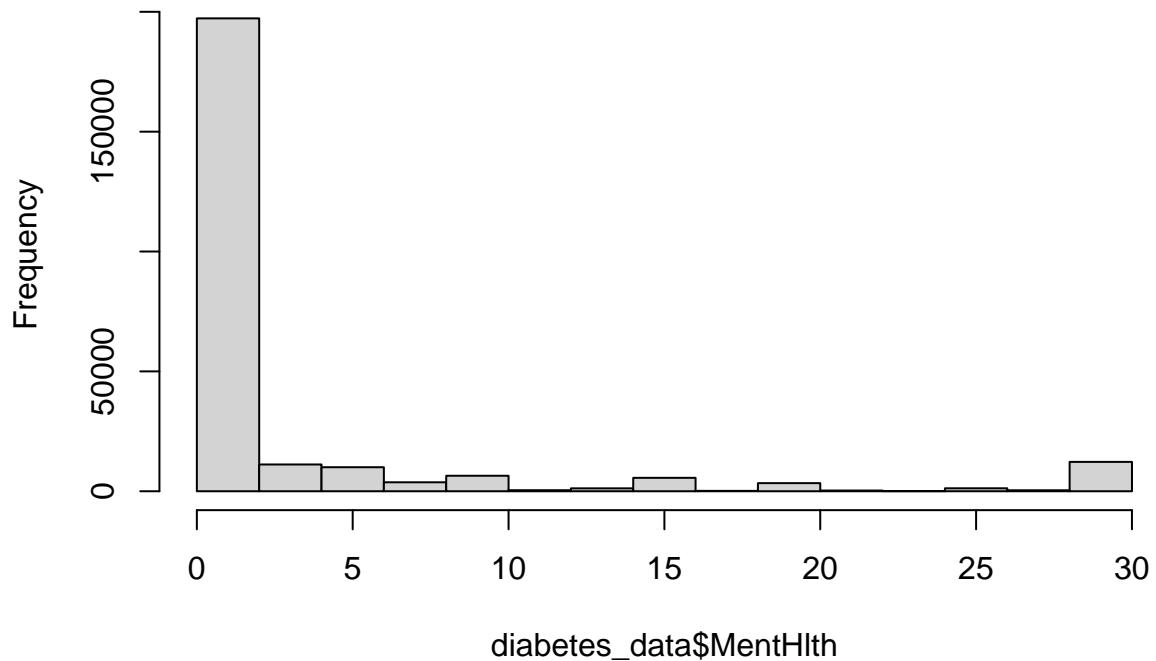
Histogram of diabetes_data\$BMI



```
###This histogram of the BMI data shows that it is right-skewed, with the majority of the data concentrated in the lower BMI range.
```

```
##Histogram of Mental Health
hist(diabetes_data$MentHlth)
```

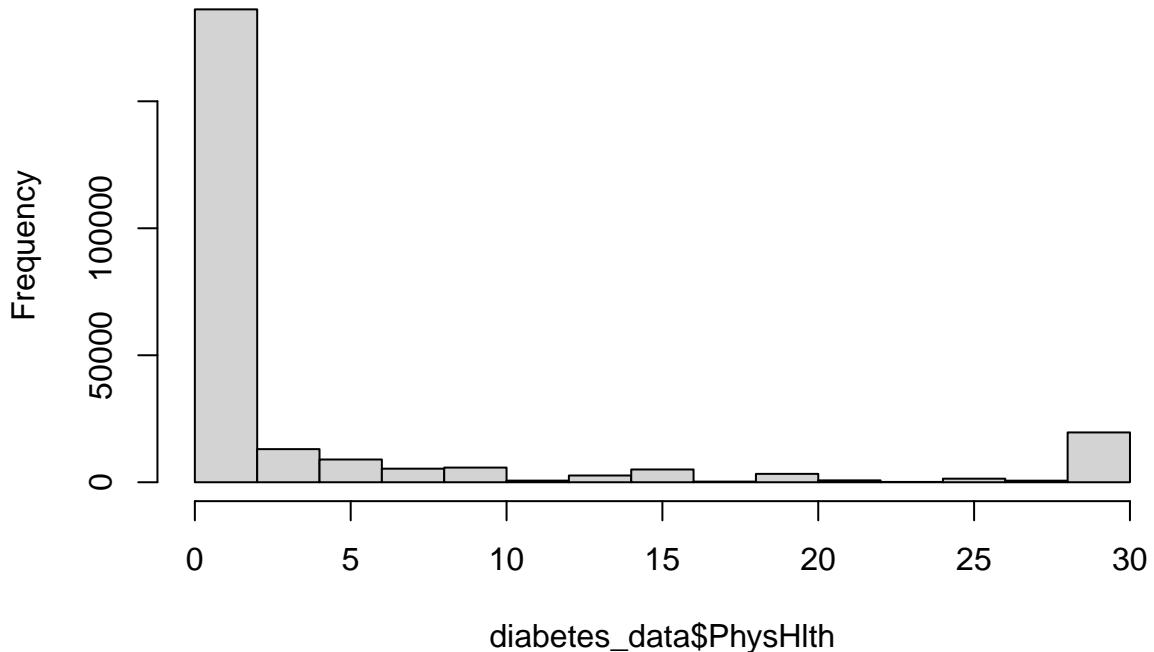
Histogram of diabetes_data\$MentHlth



```
###This histogram is right-skewed.
```

```
##Histogram of Physical Health  
hist(diabetes_data$PhysHlth)
```

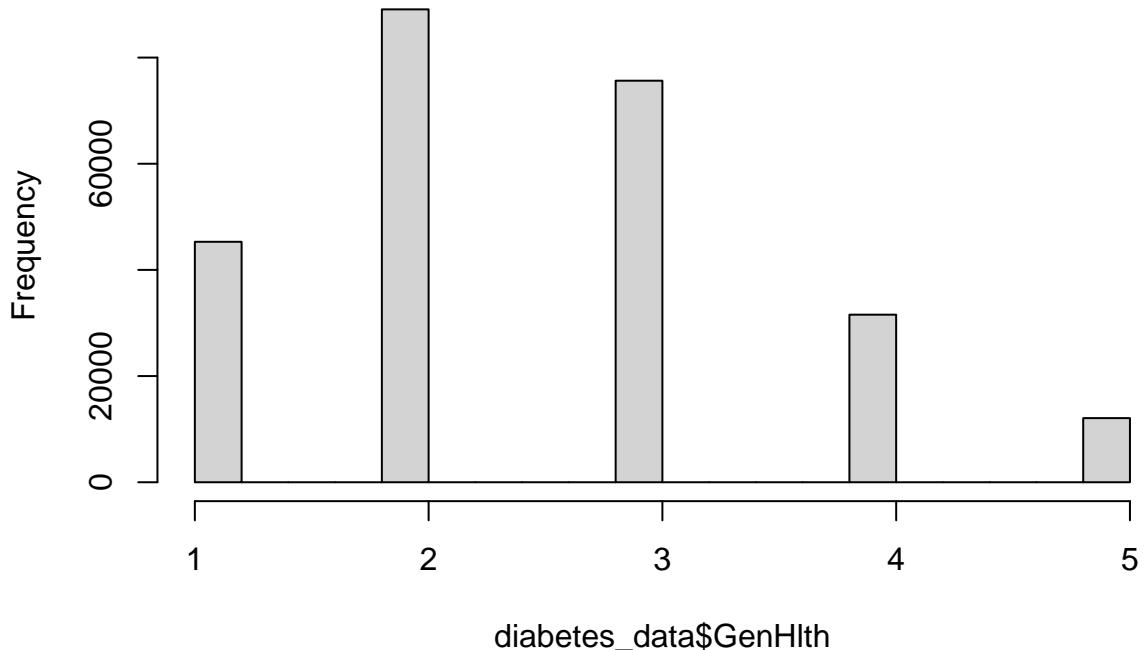
Histogram of diabetes_data\$PhysHlth



```
###This histogram looks almost exactly like the one for Mental Health, also right skewed.
```

```
##Histogram of General Health  
hist(diabetes_data$GenHlth)
```

Histogram of diabetes_data\$GenHlth

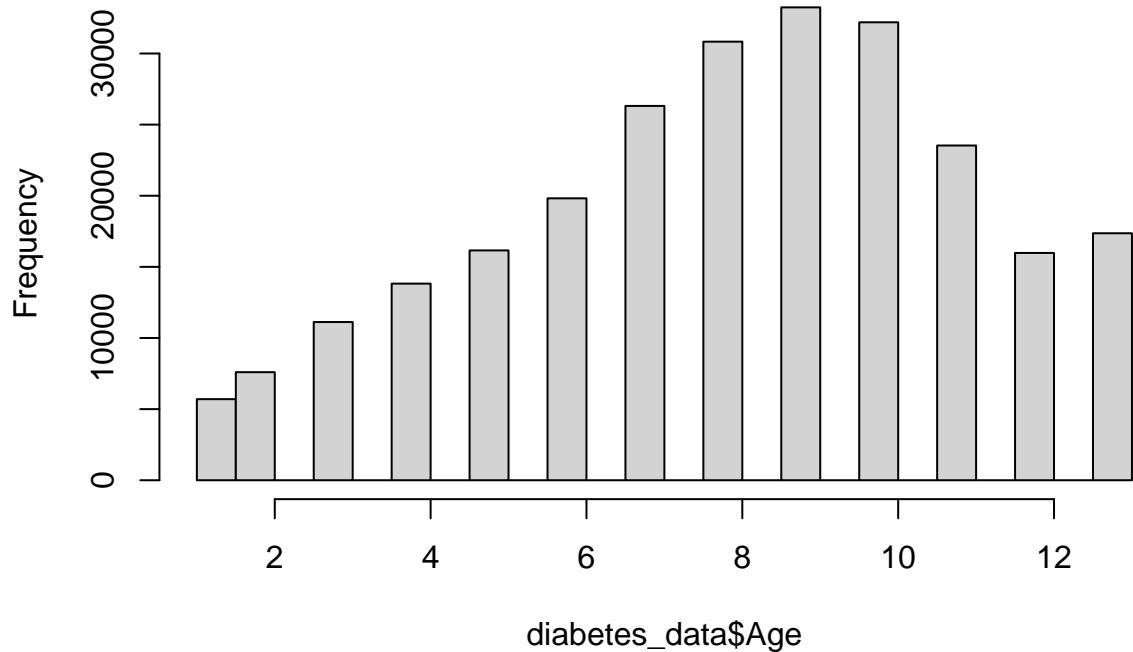


```
###This histogram is right-skewed
```

```
##Histogram of Age
```

```
hist(diabetes_data$Age)
```

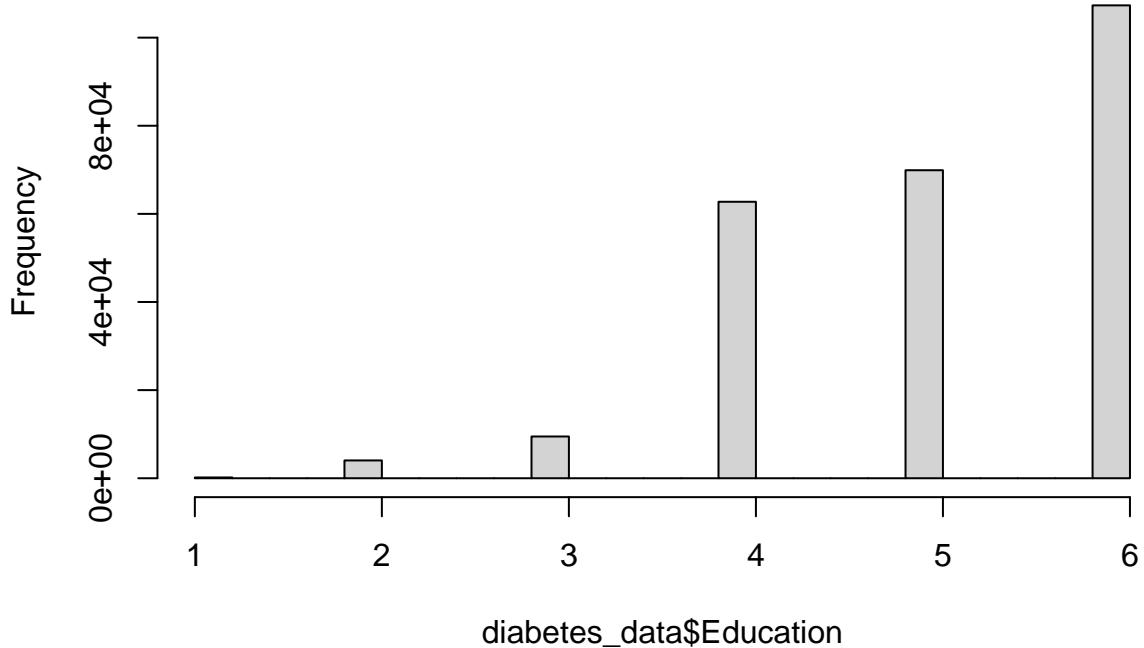
Histogram of diabetes_data\$Age



```
###This histogram is left-skewed
```

```
##Histogram of Education Status  
hist(diabetes_data$Education)
```

Histogram of diabetes_data\$Education



```
###This histogram is left skewed
```

```
## Age Grouping
diabetes_data <- diabetes_data %>%
  mutate(Age_group = case_when(
    Age <= 5 ~ "Young",
    Age > 5 & Age <= 9 ~ "Middle-Aged",
    Age > 9 ~ "Older",
    TRUE ~ NA_character_
  ),
  Education_level = case_when(
    Education %in% c("1", "2") ~ "Low Education",
    Education %in% c("3", "4") ~ "Medium Education",
    Education %in% c("5", "6") ~ "High Education",
    TRUE ~ NA_character_
  ),
  Income_level = case_when(
    Income <= 2 ~ "Low Income",
    Income > 2 & Income <= 4 ~ "Medium Income",
    Income > 4 ~ "High Income",
    TRUE ~ NA_character_
  ),
  GenHealthCategory = case_when(
    GenHlth %in% c(1, 2) ~ "Good Health",
    GenHlth == 3 ~ "Average Health",
    GenHlth %in% c(4, 5) ~ "Poor Health",
    TRUE ~ NA_character_
  )
```

```

        TRUE ~ NA_character_
),
MentHlthCategory = case_when(
    MentHlth <= 5 ~ "Low",
    MentHlth > 5 & MentHlth <= 16 ~ "Moderate",
    MentHlth > 16 ~ "High",
    TRUE ~ NA_character_
),
PhysHlthCategory = case_when(
    PhysHlth <= 5 ~ "Low",
    PhysHlth > 5 & PhysHlth <= 16 ~ "Moderate",
    PhysHlth > 16 ~ "High",
    TRUE ~ NA_character_
))

## Remove the original columns
diabetes_data <- diabetes_data %>%
    dplyr::select(-Age, -Education, -Income, -GenHlth, -MentHlth, -PhysHlth)

##Using this code to convert the columns that are actually of data type factor from numerical columns to factors
diabetes_df <- diabetes_data %>%
    mutate_at(vars(Diabetes_binary, HighBP, HighChol, CholCheck, Smoker, Stroke, HeartDiseaseorAttack, PhysActivity, Fruits, Veggies, HvyAlcoholConsump, AnyHealthcare, NoDocbcCost, DiffWalk, Sex, Age_group, Education_level, Income_level, GenHealthCategory, MentHlthCategory, PhysHlthCategory), as.factor)

##Check to make sure the columns that needed to be converted to type factor were.
str(diabetes_df)

## 'data.frame': 253680 obs. of 22 variables:
## $ Diabetes_binary      : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 2 1 ...
## $ HighBP                 : Factor w/ 2 levels "0","1": 2 1 2 2 2 2 2 2 2 1 ...
## $ HighChol                : Factor w/ 2 levels "0","1": 2 1 2 1 2 2 1 2 2 1 ...
## $ CholCheck                : Factor w/ 2 levels "0","1": 2 1 2 2 2 2 2 2 2 2 ...
## $ BMI                     : num  40 25 28 27 24 25 30 25 30 24 ...
## $ Smoker                  : Factor w/ 2 levels "0","1": 2 2 1 1 1 2 2 2 2 1 ...
## $ Stroke                  : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ HeartDiseaseorAttack: Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 2 1 ...
## $ PhysActivity              : Factor w/ 2 levels "0","1": 1 2 1 2 2 2 1 2 1 1 ...
## $ Fruits                   : Factor w/ 2 levels "0","1": 1 1 2 2 2 2 1 1 2 1 ...
## $ Veggies                  : Factor w/ 2 levels "0","1": 2 1 1 2 2 2 1 2 2 2 ...
## $ HvyAlcoholConsump       : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ AnyHealthcare             : Factor w/ 2 levels "0","1": 2 1 2 2 2 2 2 2 2 2 ...
## $ NoDocbcCost               : Factor w/ 2 levels "0","1": 1 2 2 1 1 1 1 1 1 1 ...
## $ DiffWalk                  : Factor w/ 2 levels "0","1": 2 1 2 1 1 1 1 2 2 1 ...
## $ Sex                      : Factor w/ 2 levels "0","1": 1 1 1 1 1 2 1 1 1 2 ...
## $ Age_group                 : Factor w/ 3 levels "Middle-Aged",...: 1 1 1 2 2 2 1 2 1 1 ...
## $ Education_level            : Factor w/ 3 levels "High Education",...: 3 1 3 3 1 1 1 3 1 3 ...
## $ Income_level                : Factor w/ 3 levels "High Income",...: 3 2 1 1 3 1 1 3 2 3 ...
## $ GenHealthCategory          : Factor w/ 3 levels "Average Health",...: 3 1 3 2 2 2 1 1 3 2 ...
## $ MentHlthCategory            : Factor w/ 3 levels "High","Low","Moderate": 1 2 1 2 2 2 2 1 2 ...
## $ PhysHlthCategory            : Factor w/ 3 levels "High","Low","Moderate": 3 2 1 2 2 2 3 2 1 2 ...

##Define the function to remove outliers based on the IQR method
remove_outliers <- function(data, column) {
    Q1 <- quantile(data[[column]], 0.25, na.rm = TRUE)

```

```

Q3 <- quantile(data[[column]], 0.75, na.rm = TRUE)
IQR_value <- Q3 - Q1
##Filter data to remove outliers in the specific column
data <- data %>% filter(
  data[[column]] >= (Q1 - 1.5 * IQR_value) &
  data[[column]] <= (Q3 + 1.5 * IQR_value)
)
return(data)
}

##Apply the outlier removal function to selected columns only
for (col in c("BMI")) {
  diabetes_df <- remove_outliers(diabetes_df, col)
}

##We set the seed for reproducibility
set.seed(123)

##Extracting 80% of the data to split into train and test data
sample_data <- sample(1:nrow(diabetes_df), .8 * nrow(diabetes_df))

##Split the data into training and testing data sets
train_df <- diabetes_df[sample_data, ]
test_df <- diabetes_df[-sample_data, ]

##Inspect training data frame
str(train_df)

## 'data.frame': 195066 obs. of 22 variables:
## $ Diabetes_binary : Factor w/ 2 levels "0","1": 1 2 1 1 1 1 2 1 1 1 ...
## $ HighBP          : Factor w/ 2 levels "0","1": 2 1 1 2 1 1 2 1 2 2 ...
## $ HighChol        : Factor w/ 2 levels "0","1": 1 2 2 1 1 1 2 1 1 2 ...
## $ CholCheck       : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 2 ...
## $ BMI             : num 19 41 35 23 25 23 40 23 18 33 ...
## $ Smoker          : Factor w/ 2 levels "0","1": 1 2 1 2 1 1 1 1 2 1 ...
## $ Stroke          : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ HeartDiseaseorAttack: Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ PhysActivity    : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 1 2 2 1 ...
## $ Fruits          : Factor w/ 2 levels "0","1": 2 2 2 2 1 1 1 1 2 1 ...
## $ Veggies          : Factor w/ 2 levels "0","1": 2 2 2 2 2 1 2 2 2 1 ...
## $ HvyAlcoholConsump : Factor w/ 2 levels "0","1": 1 1 1 2 1 1 1 1 2 1 ...
## $ AnyHealthcare   : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 2 ...
## $ NoDocbcCost     : Factor w/ 2 levels "0","1": 1 2 1 1 1 1 1 1 2 1 ...
## $ DiffWalk         : Factor w/ 2 levels "0","1": 1 2 1 1 1 1 2 1 1 2 ...
## $ Sex              : Factor w/ 2 levels "0","1": 1 1 1 1 2 1 2 1 1 1 ...
## $ Age_group        : Factor w/ 3 levels "Middle-Aged",...: 2 1 1 2 3 1 3 3 1 2 ...
## $ Education_level : Factor w/ 3 levels "High Education",...: 1 3 1 1 1 1 1 1 1 ...
## $ Income_level     : Factor w/ 3 levels "High Income",...: 3 3 1 3 1 1 1 1 2 ...
## $ GenHealthCategory: Factor w/ 3 levels "Average Health",...: 2 1 2 3 2 2 3 2 1 1 ...
## $ MentHlthCategory : Factor w/ 3 levels "High","Low","Moderate": 2 1 2 2 2 2 1 2 2 2 ...
## $ PhysHlthCategory : Factor w/ 3 levels "High","Low","Moderate": 2 1 2 2 2 2 3 2 2 2 ...

```

```

##Inspect testing data frame
str(test_df)

## 'data.frame': 48767 obs. of 22 variables:
## $ Diabetes_binary : Factor w/ 2 levels "0","1": 1 1 1 2 1 1 1 2 1 1 ...
## $ HighBP          : Factor w/ 2 levels "0","1": 1 2 2 2 1 2 1 2 2 2 ...
## $ HighChol        : Factor w/ 2 levels "0","1": 1 2 2 2 2 2 1 2 2 2 ...
## $ CholCheck       : Factor w/ 2 levels "0","1": 1 2 2 2 2 2 2 2 2 2 ...
## $ BMI             : num 25 24 25 28 33 28 32 37 24 24 ...
## $ Smoker          : Factor w/ 2 levels "0","1": 2 1 2 1 2 2 1 2 2 2 ...
## $ Stroke          : Factor w/ 2 levels "0","1": 1 1 1 1 2 1 1 2 1 1 ...
## $ HeartDiseaseorAttack: Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 2 1 1 ...
## $ PhysActivity    : Factor w/ 2 levels "0","1": 2 2 2 1 2 1 2 1 2 1 ...
## $ Fruits          : Factor w/ 2 levels "0","1": 1 2 2 1 1 2 2 1 1 2 ...
## $ Veggies         : Factor w/ 2 levels "0","1": 1 2 2 2 2 2 2 2 2 2 ...
## $ HvyAlcoholConsump : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ AnyHealthcare   : Factor w/ 2 levels "0","1": 1 2 2 2 2 2 2 2 2 2 ...
## $ NoDocbcCost    : Factor w/ 2 levels "0","1": 2 1 1 1 2 1 1 1 1 1 ...
## $ DiffWalk        : Factor w/ 2 levels "0","1": 1 1 1 2 1 2 1 2 2 1 ...
## $ Sex              : Factor w/ 2 levels "0","1": 1 1 2 1 1 1 1 2 2 2 ...
## $ Age_group       : Factor w/ 3 levels "Middle-Aged",...: 1 2 2 2 3 1 3 2 1 1 ...
## $ Education_level : Factor w/ 3 levels "High Education",...: 1 1 1 3 1 3 1 1 3 3 ...
## $ Income_level    : Factor w/ 3 levels "High Income",...: 2 3 1 1 2 1 1 1 3 2 ...
## $ GenHealthCategory: Factor w/ 3 levels "Average Health",...: 1 2 2 3 3 1 2 3 1 3 ...
## $ MentHlthCategory: Factor w/ 3 levels "High","Low","Moderate": 2 2 2 2 1 3 2 2 2 2 ...
## $ PhysHlthCategory: Factor w/ 3 levels "High","Low","Moderate": 2 2 2 2 1 2 2 2 2 1 ...

##Fit full logistic model
log_model <- glm(Diabetes_binary ~ ., data = train_df, family= binomial)

##Display model summary
summary(log_model)

## Call:
## glm(formula = Diabetes_binary ~ ., family = binomial, data = train_df)
##
## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                -6.564174  0.107259 -61.199 < 2e-16 ***
## HighBP1                     0.729514  0.017098  42.667 < 2e-16 ***
## HighChol1                   0.523102  0.015858  32.988 < 2e-16 ***
## CholCheck1                  1.247285  0.080914  15.415 < 2e-16 ***
## BMI                         0.090959  0.001527  59.553 < 2e-16 ***
## Smoker1                     -0.013202  0.015377 -0.859  0.39059
## Stroke1                     0.222608  0.028865  7.712 1.24e-14 ***
## HeartDiseaseorAttack1      0.270600  0.020593  13.140 < 2e-16 ***
## PhysActivity1               -0.048792  0.016913 -2.885  0.00391 **
## Fruits1                      -0.020796  0.015975 -1.302  0.19298
## Veggies1                     -0.047872  0.018591 -2.575  0.01002 *
## HvyAlcoholConsump1           -0.786925  0.044512 -17.679 < 2e-16 ***
## AnyHealthcare1               0.088325  0.039507  2.236  0.02537 *
## NoDocbcCost1                 0.034701  0.027120  1.280  0.20071
```

```

## DiffWalk1          0.131591  0.019820  6.639 3.15e-11 ***
## Sex1              0.237923  0.015593 15.258 < 2e-16 ***
## Age_groupOlder    0.371400  0.016624 22.341 < 2e-16 ***
## Age_groupYoung    -0.866796  0.030658 -28.273 < 2e-16 ***
## Education_levelLow Education 0.241107  0.046380  5.199 2.01e-07 ***
## Education_levelMedium Education 0.047719  0.016643  2.867  0.00414 **
## Income_levelLow Income     0.285204  0.025179 11.327 < 2e-16 ***
## Income_levelMedium Income     0.190550  0.020597  9.252 < 2e-16 ***
## GenHealthCategoryGood Health -0.822652  0.018817 -43.717 < 2e-16 ***
## GenHealthCategoryPoor Health  0.465167  0.020835 22.327 < 2e-16 ***
## MentHlthCategoryLow        0.023232  0.027993  0.830  0.40658
## MentHlthCategoryModerate   -0.046649  0.036322 -1.284  0.19903
## PhysHlthCategoryLow       0.049577  0.025100  1.975  0.04824 *
## PhysHlthCategoryModerate   0.030407  0.029546  1.029  0.30341
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 151176  on 195065  degrees of freedom
## Residual deviance: 120012  on 195038  degrees of freedom
## AIC: 120068
##
## Number of Fisher Scoring iterations: 6

```

```

##Calculate Variance Inflation Factors (VIF)
vif_values <- vif(log_model)

##Print VIF values to assess multicollinearity
print(vif_values)

```

```

##                               GVIF Df GVIF^(1/(2*Df))
## HighBP           1.138858  1      1.067173
## HighChol         1.073884  1      1.036284
## CholCheck        1.009556  1      1.004766
## BMI              1.084533  1      1.041409
## Smoker            1.071481  1      1.035123
## Stroke            1.072461  1      1.035597
## HeartDiseaseorAttack 1.148379  1      1.071625
## PhysActivity      1.133198  1      1.064518
## Fruits             1.106404  1      1.051857
## Veggies            1.103159  1      1.050314
## HvyAlcoholConsump 1.011891  1      1.005928
## AnyHealthcare      1.092683  1      1.045315
## NoDocbcCost        1.144512  1      1.069819
## DiffWalk           1.462131  1      1.209186
## Sex                1.101062  1      1.049315
## Age_group          1.245464  2      1.056411
## Education_level    1.191266  2      1.044725
## Income_level        1.339944  2      1.075899
## GenHealthCategory   1.684025  2      1.139166
## MentHlthCategory    1.256563  2      1.058756
## PhysHlthCategory    1.645215  2      1.132546

```

```

###VIF values are low and indicate any multicollinearity in this model is minimal and likely would have
##Stepwise Backward Regression for Feature Selection
log_model_full <- glm(Diabetes_binary ~ ., data = train_df, family= binomial)

##Backward stepwise selection
step_model <- stepAIC(log_model_full, direction = "both")

## Start: AIC=120067.6
## Diabetes_binary ~ HighBP + HighChol + CholCheck + BMI + Smoker +
##   Stroke + HeartDiseaseorAttack + PhysActivity + Fruits + Veggies +
##   HvyAlcoholConsump + AnyHealthcare + NoDocbcCost + DiffWalk +
##   Sex + Age_group + Education_level + Income_level + GenHealthCategory +
##   MentHlthCategory + PhysHlthCategory
##
##                                     Df Deviance    AIC
## - Smoker                      1  120012 120066
## - NoDocbcCost                  1  120013 120067
## - Fruits                       1  120013 120067
## - PhysHlthCategory             2  120016 120068
## <none>                         120012 120068
## - MentHlthCategory             2  120018 120070
## - AnyHealthcare                1  120017 120071
## - Veggies                      1  120018 120072
## - PhysActivity                 1  120020 120074
## - Education_level              2  120042 120094
## - DiffWalk                     1  120055 120109
## - Stroke                       1  120070 120124
## - Income_level                 2  120175 120227
## - HeartDiseaseorAttack         1  120181 120235
## - Sex                           1  120244 120298
## - CholCheck                     1  120347 120401
## - HvyAlcoholConsump             1  120384 120438
## - HighChol                      1  121116 121170
## - Age_group                     2  121865 121917
## - HighBP                        1  121897 121951
## - GenHealthCategory              2  123296 123348
## - BMI                           1  123640 123694
##
## Step: AIC=120066.3
## Diabetes_binary ~ HighBP + HighChol + CholCheck + BMI + Stroke +
##   HeartDiseaseorAttack + PhysActivity + Fruits + Veggies +
##   HvyAlcoholConsump + AnyHealthcare + NoDocbcCost + DiffWalk +
##   Sex + Age_group + Education_level + Income_level + GenHealthCategory +
##   MentHlthCategory + PhysHlthCategory
##
##                                     Df Deviance    AIC
## - Fruits                       1  120014 120066
## - NoDocbcCost                  1  120014 120066
## <none>                         120012 120066
## - PhysHlthCategory             2  120016 120066
## + Smoker                       1  120012 120068
## - MentHlthCategory             2  120019 120069

```

```

## - AnyHealthcare      1 120017 120069
## - Veggies            1 120019 120071
## - PhysActivity       1 120021 120073
## - Education_level    2 120042 120092
## - DiffWalk           1 120056 120108
## - Stroke             1 120071 120123
## - Income_level        2 120175 120225
## - HeartDiseaseorAttack 1 120181 120233
## - Sex                1 120246 120298
## - CholCheck           1 120348 120400
## - HvyAlcoholConsump   1 120390 120442
## - HighChol            1 121116 121168
## - Age_group           2 121867 121917
## - HighBP              1 121898 121950
## - GenHealthCategory   2 123298 123348
## - BMI                1 123651 123703
##
## Step: AIC=120065.9
## Diabetes_binary ~ HighBP + HighChol + CholCheck + BMI + Stroke +
##      HeartDiseaseorAttack + PhysActivity + Veggies + HvyAlcoholConsump +
##      AnyHealthcare + NoDocbcCost + DiffWalk + Sex + Age_group +
##      Education_level + Income_level + GenHealthCategory + MentHlthCategory +
##      PhysHlthCategory
##
##                                     Df Deviance     AIC
## - NoDocbcCost                 1 120016 120066
## <none>                         120014 120066
## - PhysHlthCategory            2 120018 120066
## + Fruits                       1 120012 120066
## + Smoker                       1 120013 120067
## - MentHlthCategory            2 120020 120068
## - AnyHealthcare               1 120019 120069
## - Veggies                      1 120023 120073
## - PhysActivity                1 120023 120073
## - Education_level              2 120044 120092
## - DiffWalk                     1 120057 120107
## - Stroke                       1 120072 120122
## - Income_level                 2 120178 120226
## - HeartDiseaseorAttack         1 120183 120233
## - Sex                          1 120252 120302
## - CholCheck                    1 120349 120399
## - HvyAlcoholConsump            1 120390 120440
## - HighChol                     1 121120 121170
## - Age_group                    2 121868 121916
## - HighBP                       1 121900 121950
## - GenHealthCategory            2 123304 123352
## - BMI                          1 123665 123715
##
## Step: AIC=120065.5
## Diabetes_binary ~ HighBP + HighChol + CholCheck + BMI + Stroke +
##      HeartDiseaseorAttack + PhysActivity + Veggies + HvyAlcoholConsump +
##      AnyHealthcare + DiffWalk + Sex + Age_group + Education_level +
##      Income_level + GenHealthCategory + MentHlthCategory + PhysHlthCategory
##

```

```

##                                     Df Deviance    AIC
## - PhysHlthCategory             2   120019 120065
## <none>                         120016 120066
## + NoDocbcCost                 1   120014 120066
## + Fruits                        1   120014 120066
## + Smoker                        1   120015 120067
## - MentHlthCategory             2   120021 120067
## - AnyHealthcare                1   120020 120068
## - Veggies                       1   120024 120072
## - PhysActivity                  1   120024 120072
## - Education_level               2   120046 120092
## - DiffWalk                      1   120059 120107
## - Stroke                        1   120074 120122
## - Income_level                  2   120184 120230
## - HeartDiseaseorAttack          1   120185 120233
## - Sex                            1   120253 120301
## - CholCheck                      1   120350 120398
## - HvyAlcoholConsump              1   120392 120440
## - HighChol                      1   121123 121171
## - Age_group                      2   121883 121929
## - HighBP                         1   121903 121951
## - GenHealthCategory              2   123317 123363
## - BMI                            1   123667 123715
##
## Step:  AIC=120065.5
## Diabetes_binary ~ HighBP + HighChol + CholCheck + BMI + Stroke +
##      HeartDiseaseorAttack + PhysActivity + Veggies + HvyAlcoholConsump +
##      AnyHealthcare + DiffWalk + Sex + Age_group + Education_level +
##      Income_level + GenHealthCategory + MentHlthCategory
##
##                                     Df Deviance    AIC
## <none>                         120019 120065
## + PhysHlthCategory              2   120016 120066
## + Fruits                        1   120018 120066
## + NoDocbcCost                  1   120018 120066
## + Smoker                        1   120019 120067
## - AnyHealthcare                 1   120024 120068
## - MentHlthCategory              2   120027 120069
## - PhysActivity                  1   120028 120072
## - Veggies                       1   120028 120072
## - Education_level               2   120050 120092
## - DiffWalk                      1   120060 120104
## - Stroke                        1   120077 120121
## - Income_level                  2   120186 120228
## - HeartDiseaseorAttack          1   120188 120232
## - Sex                            1   120257 120301
## - CholCheck                      1   120353 120397
## - HvyAlcoholConsump              1   120395 120439
## - HighChol                      1   121127 121171
## - Age_group                      2   121886 121928
## - HighBP                         1   121908 121952
## - GenHealthCategory              2   123518 123560
## - BMI                            1   123680 123724

```

```

##Display summary of selected model
summary(step_model)

## Call:
## glm(formula = Diabetes_binary ~ HighBP + HighChol + CholCheck +
##      BMI + Stroke + HeartDiseaseorAttack + PhysActivity + Veggies +
##      HvyAlcoholConsump + AnyHealthcare + DiffWalk + Sex + Age_group +
##      Education_level + Income_level + GenHealthCategory + MentHlthCategory,
##      family = binomial, data = train_df)
##
## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                -6.528741  0.104128 -62.699 < 2e-16 ***
## HighBP1                     0.730036  0.017096  42.702 < 2e-16 ***
## HighChol1                  0.523301  0.015841  33.035 < 2e-16 ***
## CholCheck1                 1.243541  0.080863  15.378 < 2e-16 ***
## BMI                         0.091168  0.001524  59.803 < 2e-16 ***
## Stroke1                    0.221141  0.028852   7.665 1.79e-14 ***
## HeartDiseaseorAttack1     0.269000  0.020553  13.088 < 2e-16 ***
## PhysActivity1               -0.048264  0.016807  -2.872  0.00408 **
## Veggies1                   -0.053998  0.018112  -2.981  0.00287 **
## HvyAlcoholConsump1          -0.786926  0.044372 -17.735 < 2e-16 ***
## AnyHealthcare1              0.077618  0.038774   2.002  0.04531 *
## DiffWalk1                  0.121801  0.019165   6.355 2.08e-10 ***
## Sex1                        0.237223  0.015382  15.423 < 2e-16 ***
## Age_groupOlder              0.367657  0.016451  22.348 < 2e-16 ***
## Age_groupYoung              -0.863062  0.030616 -28.190 < 2e-16 ***
## Education_levelLow Education 0.244362  0.046354   5.272 1.35e-07 ***
## Education_levelMedium Education 0.048198  0.016590   2.905  0.00367 **
## Income_levelLow Income       0.286496  0.025086  11.420 < 2e-16 ***
## Income_levelMedium Income    0.192023  0.020528   9.354 < 2e-16 ***
## GenHealthCategoryGood Health -0.821719  0.018763 -43.794 < 2e-16 ***
## GenHealthCategoryPoor Health  0.452569  0.019664  23.015 < 2e-16 ***
## MentHlthCategoryLow          0.031786  0.027259   1.166  0.24359
## MentHlthCategoryModerate    -0.041227  0.036094  -1.142  0.25337
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 151176  on 195065  degrees of freedom
## Residual deviance: 120019  on 195043  degrees of freedom
## AIC: 120065
##
## Number of Fisher Scoring iterations: 6

library(pROC)

##Model Evaluation Function
predict_and_evaluate <- function(model, test_data, lambda = best_lambda) {
  ##Adjust if using glmnet model
  if ("glmnet" %in% class(model)) {

```

```

        x_test <- model.matrix(Diabetes_binary ~ ., data = test_data)[, -1]
        prob.preds <- predict(model, newx = x_test, s = lambda, type = "response")
    } else {
        prob.preds <- predict(model, test_data, type = "response")
    }

    ##Convert probabilities to binary predictions
    resp.preds <- ifelse(prob.preds >= 0.5, 1, 0)
    confusion <- table(test_data$Diabetes_binary, resp.preds)

    ##Calculate Sensitivity and Accuracy
    sensitivity <- confusion[2, 2] / (confusion[2, 1] + confusion[2, 2])
    accuracy <- sum(diag(confusion)) / sum(confusion)

    ##Calculate AUC
    roc_obj <- roc(test_data$Diabetes_binary, as.numeric(prob.preds))
    auc <- auc(roc_obj)

    list(ConfusionMatrix = confusion, Sensitivity = sensitivity, Accuracy = accuracy, AUC = auc)
}

predict_and_evaluate(log_model, test_df, lambda = best_lambda)

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## $ConfusionMatrix
##     resp.preds
##         0      1
##     0 41525    736
##     1  5594    912
##
## $Sensitivity
## [1] 0.1401783
##
## $Accuracy
## [1] 0.8701991
##
## $AUC
## Area under the curve: 0.8224

##Evaluate on test data
step_model_eval <- predict_and_evaluate(step_model, test_df)

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

```

```

##Print evaluation results
print(step_model_eval)

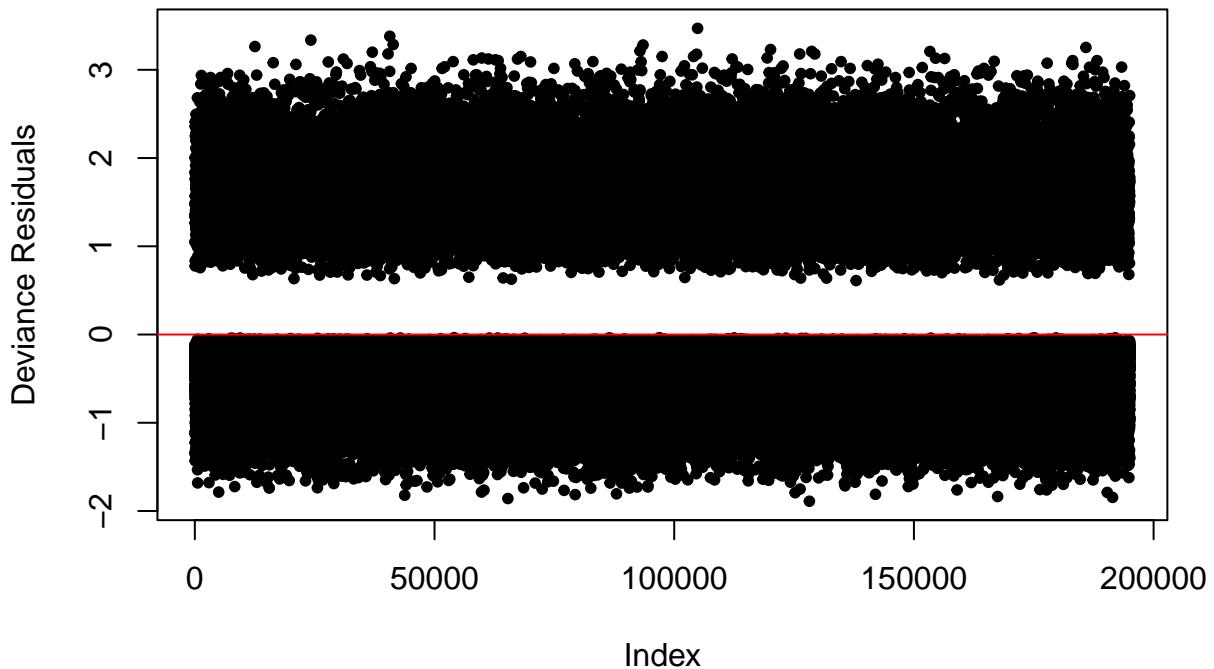
## $ConfusionMatrix
##   resp.preds
##   0    1
## 0 41530   731
## 1  5595   911
##
## $Sensitivity
## [1] 0.1400246
##
## $Accuracy
## [1] 0.8702811
##
## $AUC
## Area under the curve: 0.8223

##Residual and ROC Curve Plotting
##Deviance residuals for stepwise model
residuals_stepwise <- residuals(step_model, type = "deviance")

plot(residuals_stepwise, main = "Deviance Residuals of Stepwise-Selected Model",
      ylab = "Deviance Residuals", xlab = "Index", pch = 20)
abline(h = 0, col = "red")

```

Deviance Residuals of Stepwise-Selected Model

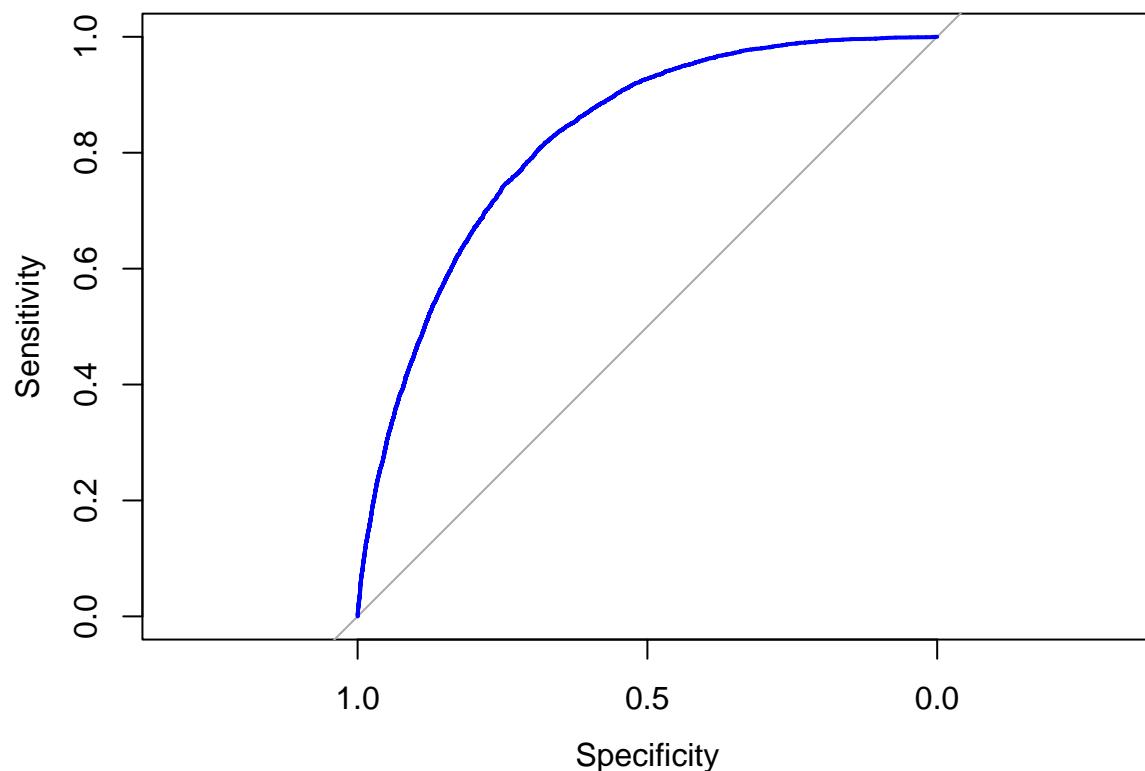


```
##ROC Curve
roc_obj <- roc(test_df$Diabetes_binary, predict(step_model, test_df, type="response"))

## Setting levels: control = 0, case = 1
## Setting direction: controls < cases

plot(roc_obj, col = "blue", main = "ROC Curve for Stepwise Model")
```

ROC Curve for Stepwise Model



```
##Display AUC value
auc(roc_obj)

## Area under the curve: 0.8223

# Fit the full model with all variables
first_order_model_full <- glm(Diabetes_binary ~ Age_group + BMI + MentHlthCategory + PhysHlthCategory +
                               HighBP + HighChol + CholCheck + Smoker + Stroke + HeartDiseaseorAttack +
                               PhysActivity + Fruits + Veggies + HvyAlcoholConsump + AnyHealthcare +
                               NoDocbcCost + GenHealthCategory + DiffWalk + Sex,
                               data = train_df, family = "binomial")

# Perform stepwise selection based on p-values (both forward and backward)
first_order_model_step <- step(first_order_model_full, direction = "both", trace = 0)

# Print the variables included in the final model
cat("Variables used in the First-Order Model after stepwise selection:\n")

## Variables used in the First-Order Model after stepwise selection:

print(names(coef(first_order_model_step)))

## [1] "(Intercept)"                  "Age_groupOlder"
## [3] "Age_groupYoung"               "BMI"
```

```

## [5] "MentHlthCategoryLow"           "MentHlthCategoryModerate"
## [7] "HighBP1"                      "HighChol1"
## [9] "CholCheck1"                   "Stroke1"
## [11] "HeartDiseaseorAttack1"        "PhysActivity1"
## [13] "Fruits1"                      "Veggies1"
## [15] "HvyAlcoholConsump1"           "NoDocbcCost1"
## [17] "GenHealthCategoryGood Health" "GenHealthCategoryPoor Health"
## [19] "DiffWalk1"                     "Sex1"

# Summary of the stepwise model
summary(first_order_model_step)

## 
## Call:
## glm(formula = Diabetes_binary ~ Age_group + BMI + MentHlthCategory +
##      HighBP + HighChol + CholCheck + Stroke + HeartDiseaseorAttack +
##      PhysActivity + Fruits + Veggies + HvyAlcoholConsump + NoDocbcCost +
##      GenHealthCategory + DiffWalk + Sex, family = "binomial",
##      data = train_df)
##
## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                -6.285791  0.098697 -63.687 < 2e-16 ***
## Age_groupOlder              0.385376  0.016448  23.430 < 2e-16 ***
## Age_groupYoung             -0.862931  0.030589 -28.210 < 2e-16 ***
## BMI                         0.090219  0.001523  59.219 < 2e-16 ***
## MentHlthCategoryLow         0.006858  0.027222   0.252 0.801082
## MentHlthCategoryModerate   -0.056840  0.035999  -1.579 0.114352
## HighBP1                      0.739175  0.017067  43.310 < 2e-16 ***
## HighChol1                    0.519133  0.015825  32.805 < 2e-16 ***
## CholCheck1                   1.238796  0.080730  15.345 < 2e-16 ***
## Stroke1                      0.237404  0.028798   8.244 < 2e-16 ***
## HeartDiseaseorAttack1       0.273315  0.020540  13.307 < 2e-16 ***
## PhysActivity1                 -0.057681  0.016802  -3.433 0.000597 ***
## Fruits1                       -0.029568  0.015913  -1.858 0.063159 .
## Veggies1                      -0.076237  0.018438  -4.135 3.55e-05 ***
## HvyAlcoholConsump1            -0.811071  0.044374  -18.278 < 2e-16 ***
## NoDocbcCost1                  0.066835  0.026389   2.533 0.011319 *
## GenHealthCategoryGood Health -0.841313  0.018687  -45.020 < 2e-16 ***
## GenHealthCategoryPoor Health  0.496510  0.019445  25.534 < 2e-16 ***
## DiffWalk1                     0.154906  0.019009   8.149 3.67e-16 ***
## Sex1                          0.205146  0.015256  13.447 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 151176  on 195065  degrees of freedom
## Residual deviance: 120259  on 195046  degrees of freedom
## AIC: 120299
##
## Number of Fisher Scoring iterations: 6

```

```

# Predict on the test set with the selected significant variables
preds_first_order_step <- predict(first_order_model_step, newdata = test_df, type = "response")

# Convert probabilities to binary predictions (threshold 0.5)
resp_first_order_step <- ifelse(preds_first_order_step >= 0.5, 1, 0)

# Confusion matrix for model evaluation
confusion_first_order_step <- table(test_df$Diabetes_binary, resp_first_order_step)
print(confusion_first_order_step)

##      resp_first_order_step
##      0          1
## 0 41541    720
## 1  5607    899

# Sensitivity (True Positive Rate)
sensitivity_first_order_step <- confusion_first_order_step[2, 2] / (confusion_first_order_step[2, 1] + confusion_first_order_step[1, 2])

# Overall accuracy of the model
accuracy_first_order_step <- sum(diag(confusion_first_order_step)) / sum(confusion_first_order_step)

# ROC (Area Under the Curve)
roc_obj_first_order_step <- roc(test_df$Diabetes_binary, as.numeric(preds_first_order_step))

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

auc_first_order_step <- auc(roc_obj_first_order_step)

# Display confusion matrix, sensitivity, accuracy, and AUC
cat("Confusion Matrix:\n")

## Confusion Matrix:

print(confusion_first_order_step)

##      resp_first_order_step
##      0          1
## 0 41541    720
## 1  5607    899

cat("Sensitivity:", sensitivity_first_order_step, "\n")

## Sensitivity: 0.1381801

cat("Accuracy:", accuracy_first_order_step, "\n")

## Accuracy: 0.8702606

```

```

cat("AUC:", auc_first_order_step, "\n")

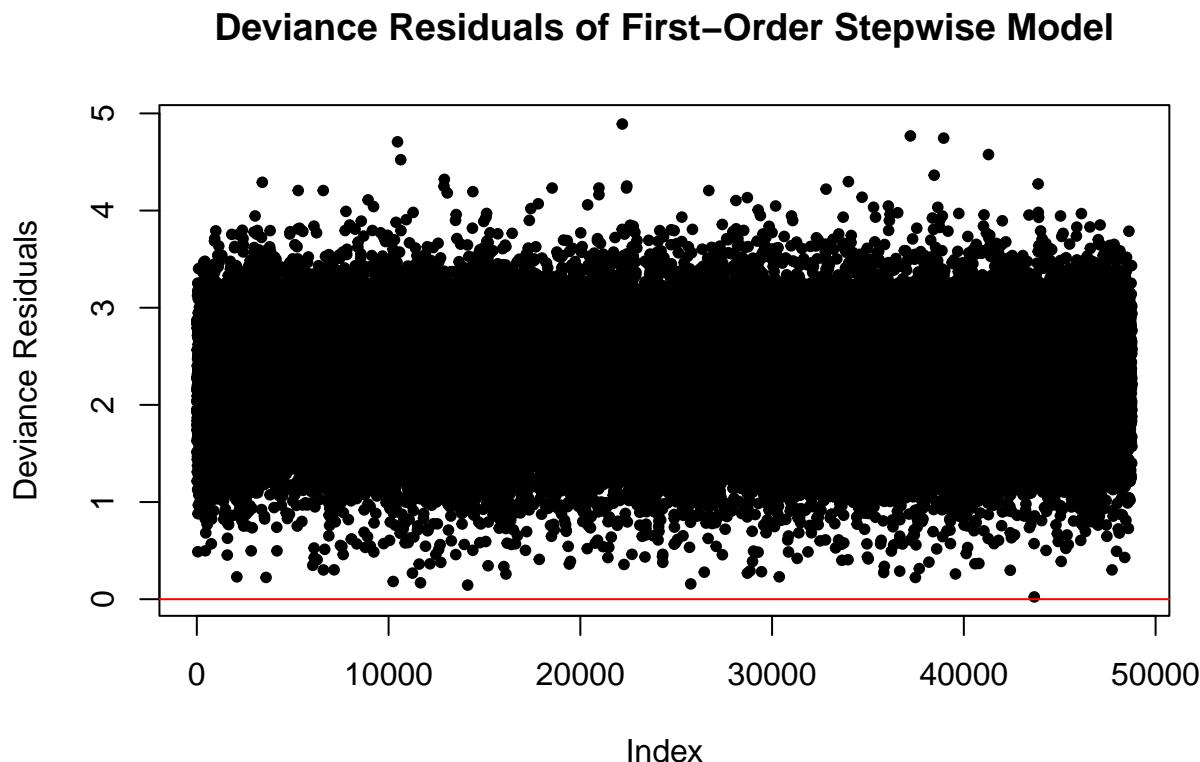
## AUC: 0.8215923

# Deviance residuals plot
deviance_residuals_first_order_step <- sign(as.numeric(test_df$Diabetes_binary) - preds_first_order_step)
deviance_residuals_first_order_step <- deviance_residuals_first_order_step / sqrt(-2 * (as.numeric(test_df$Diabetes_binary) * log(preds_first_order_step) + (1 - as.numeric(test_df$Diabetes_binary)) * log(1 - preds_first_order_step)))

## Warning in sqrt(-2 * (as.numeric(test_df$Diabetes_binary) * log(preds_first_order_step) + (1 - as.numeric(test_df$Diabetes_binary)) * log(1 - preds_first_order_step))) : NaNs produced

plot(deviance_residuals_first_order_step, main = "Deviance Residuals of First-Order Stepwise Model",
      ylab = "Deviance Residuals", xlab = "Index", pch = 20)
abline(h = 0, col = "red")

```

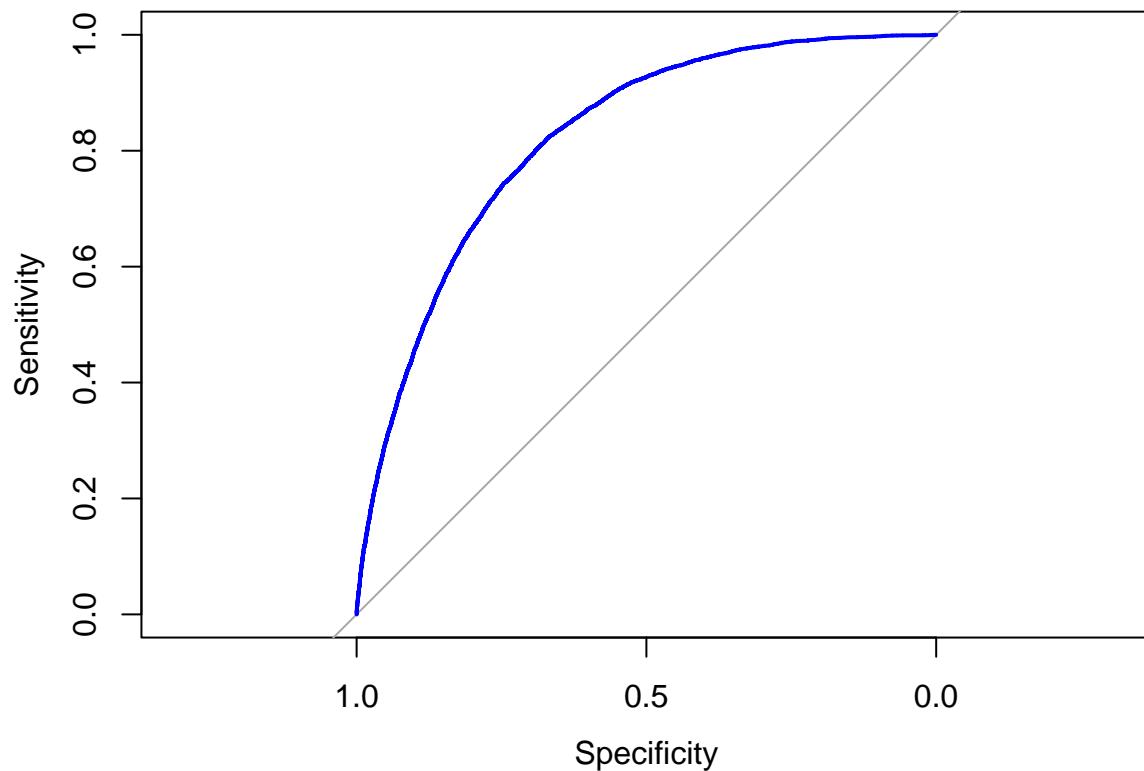


```

# ROC curve
plot(roc_obj_first_order_step, col = "blue", main = "ROC Curve for First-Order Stepwise Model")

```

ROC Curve for First-Order Stepwise Model



```
cat("AUC for First-Order Stepwise Model:", auc_first_order_step, "\n")  
  
## AUC for First-Order Stepwise Model: 0.8215923  
  
# Fit the full second-order model  
second_order_model_full <- glm(Diabetes_binary ~ (Age_group + BMI + MentHlthCategory + PhysHlthCategory +  
Income_level + HighBP + HighChol + CholCheck + Smoker + Stroke + HeartDisease +  
PhysActivity + Fruits + Veggies + HvyAlcoholConsump + AnyHealthcare +  
NoDocbcCost + GenHealthCategory + DiffWalk + Sex,  
data = train_df, family = "binomial")  
  
# Perform stepwise selection based on p-values  
second_order_model_step <- step(second_order_model_full, direction = "both", trace = 0)  
  
# Print the variables included in the final model  
cat("Variables used in the Second-Order Model after stepwise selection:\n")  
  
## Variables used in the Second-Order Model after stepwise selection:  
  
print(names(coef(second_order_model_step)))  
  
## [1] "(Intercept)"  
## [2] "Age_groupOlder"  
## [3] "Age_groupYoung"
```

```

## [4] "BMI"
## [5] "MentHlthCategoryLow"
## [6] "MentHlthCategoryModerate"
## [7] "PhysHlthCategoryLow"
## [8] "PhysHlthCategoryModerate"
## [9] "Education_levelLow Education"
## [10] "Education_levelMedium Education"
## [11] "Income_levelLow Income"
## [12] "Income_levelMedium Income"
## [13] "HighBP1"
## [14] "HighChol1"
## [15] "CholCheck1"
## [16] "Stroke1"
## [17] "HeartDiseaseorAttack1"
## [18] "PhysActivity1"
## [19] "Veggies1"
## [20] "HvyAlcoholConsump1"
## [21] "AnyHealthcare1"
## [22] "GenHealthCategoryGood Health"
## [23] "GenHealthCategoryPoor Health"
## [24] "DiffWalk1"
## [25] "Sex1"
## [26] "Age_groupOlder:PhysHlthCategoryLow"
## [27] "Age_groupYoung:PhysHlthCategoryLow"
## [28] "Age_groupOlder:PhysHlthCategoryModerate"
## [29] "Age_groupYoung:PhysHlthCategoryModerate"
## [30] "BMI:PhysHlthCategoryLow"
## [31] "BMI:PhysHlthCategoryModerate"

# Summary of the stepwise model
summary(second_order_model_step)

## 
## Call:
## glm(formula = Diabetes_binary ~ Age_group + BMI + MentHlthCategory +
##     PhysHlthCategory + Education_level + Income_level + HighBP +
##     HighChol + CholCheck + Stroke + HeartDiseaseorAttack + PhysActivity +
##     Veggies + HvyAlcoholConsump + AnyHealthcare + GenHealthCategory +
##     DiffWalk + Sex + Age_group:PhysHlthCategory + BMI:PhysHlthCategory,
##     family = "binomial", data = train_df)
## 

## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                -6.211237   0.139284 -44.594 < 2e-16
## Age_groupOlder               0.167152   0.037709   4.433 9.30e-06
## Age_groupYoung              -0.648032   0.080496  -8.050 8.25e-16
## BMI                         0.082046   0.003324  24.679 < 2e-16
## MentHlthCategoryLow          0.042501   0.027979   1.519 0.12876
## MentHlthCategoryModerate    -0.031434   0.036282  -0.866 0.38628
## PhysHlthCategoryLow          -0.427164   0.119966  -3.561 0.00037
## PhysHlthCategoryModerate    -0.261619   0.171960  -1.521 0.12816
## Education_levelLow Education      0.246850   0.046290   5.333 9.68e-08
## Education_levelMedium Education     0.047447   0.016583   2.861 0.00422
## Income_levelLow Income           0.278134   0.025107  11.078 < 2e-16

```

```

## Income_levelMedium Income          0.188853  0.020519  9.204 < 2e-16
## HighBP1                      0.725696  0.017093 42.456 < 2e-16
## HighChol1                     0.520908  0.015839 32.887 < 2e-16
## CholCheck1                    1.240734  0.080883 15.340 < 2e-16
## Stroke1                       0.223616  0.028783 7.769 7.90e-15
## HeartDiseaseorAttack1         0.272161  0.020519 13.264 < 2e-16
## PhysActivity1                  -0.052367  0.016824 -3.113 0.00185
## Veggies1                       -0.053444  0.018106 -2.952 0.00316
## HvyAlcoholConsump1              -0.787473  0.044396 -17.738 < 2e-16
## AnyHealthcare1                 0.081686  0.038827  2.104 0.03539
## GenHealthCategoryGood Health   -0.820405  0.018847 -43.529 < 2e-16
## GenHealthCategoryPoor Health   0.465991  0.020822 22.379 < 2e-16
## DiffWalk1                      0.129866  0.019750  6.575 4.85e-11
## Sex1                           0.235643  0.015393 15.309 < 2e-16
## Age_groupOlder:PhysHlthCategoryLow 0.259639  0.041788  6.213 5.19e-10
## Age_groupYoung:PhysHlthCategoryLow -0.254823  0.087652 -2.907 0.00365
## Age_groupOlder:PhysHlthCategoryModerate 0.124341  0.059626  2.085 0.03704
## Age_groupYoung:PhysHlthCategoryModerate -0.119158  0.120654 -0.988 0.32335
## BMI:PhysHlthCategoryLow          0.011890  0.003779  3.147 0.00165
## BMI:PhysHlthCategoryModerate     0.007843  0.005453  1.438 0.15032
##
## (Intercept)                   ***
## Age_groupOlder                  ***
## Age_groupYoung                  ***
## BMI                            ***
## MentHlthCategoryLow             ***
## MentHlthCategoryModerate        ***
## PhysHlthCategoryLow             ***
## PhysHlthCategoryModerate        ***
## Education_levelLow Education    ***
## Education_levelMedium Education  **
## Income_levelLow Income          ***
## Income_levelMedium Income        ***
## HighBP1                        ***
## HighChol1                      ***
## CholCheck1                     ***
## Stroke1                        ***
## HeartDiseaseorAttack1          ***
## PhysActivity1                  **
## Veggies1                       **
## HvyAlcoholConsump1              ***
## AnyHealthcare1                 *
## GenHealthCategoryGood Health   ***
## GenHealthCategoryPoor Health   ***
## DiffWalk1                      ***
## Sex1                           ***
## Age_groupOlder:PhysHlthCategoryLow *** 
## Age_groupYoung:PhysHlthCategoryLow ** 
## Age_groupOlder:PhysHlthCategoryModerate * 
## Age_groupYoung:PhysHlthCategoryModerate ** 
## BMI:PhysHlthCategoryLow          ** 
## BMI:PhysHlthCategoryModerate     ** 
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ',' 1

```

```

##  

## (Dispersion parameter for binomial family taken to be 1)  

##  

## Null deviance: 151176 on 195065 degrees of freedom  

## Residual deviance: 119949 on 195035 degrees of freedom  

## AIC: 120011  

##  

## Number of Fisher Scoring iterations: 6

# Predict on the test set with the selected significant variables
preds_second_order_step <- predict(second_order_model_step, newdata = test_df, type = "response")

# Convert probabilities to binary predictions (threshold 0.5)
resp_second_order_step <- ifelse(preds_second_order_step >= 0.5, 1, 0)

# Confusion matrix for model evaluation
confusion_second_order_step <- table(test_df$Diabetes_binary, resp_second_order_step)
print(confusion_second_order_step)

##      resp_second_order_step
##      0          1
## 0 41536    725
## 1  5611    895

# Sensitivity (True Positive Rate)
sensitivity_second_order_step <- confusion_second_order_step[2, 2] / (confusion_second_order_step[2, 1] + confusion_second_order_step[1, 2])

# Overall accuracy of the model
accuracy_second_order_step <- sum(diag(confusion_second_order_step)) / sum(confusion_second_order_step)

# ROC (Area Under the Curve)
roc_obj_second_order_step <- roc(test_df$Diabetes_binary, as.numeric(preds_second_order_step))

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

auc_second_order_step <- auc(roc_obj_second_order_step)

# Display confusion matrix, sensitivity, accuracy, and AUC
cat("Confusion Matrix:\n")

## Confusion Matrix:

print(confusion_second_order_step)

##      resp_second_order_step
##      0          1
## 0 41536    725
## 1  5611    895

```

```

cat("Sensitivity:", sensitivity_second_order_step, "\n")

## Sensitivity: 0.1375653

cat("Accuracy:", accuracy_second_order_step, "\n")

## Accuracy: 0.8700761

cat("AUC:", auc_second_order_step, "\n")

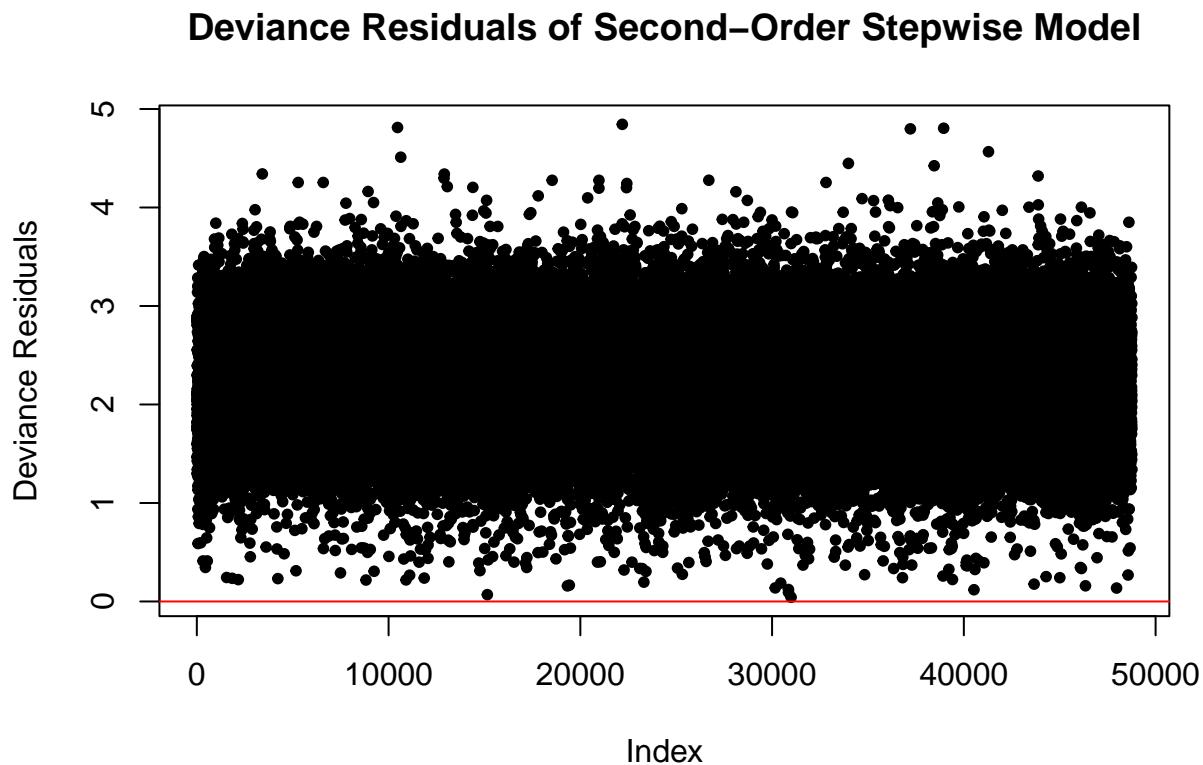
## AUC: 0.8223127

# Deviance residuals plot
deviance_residuals_second_order_step <- sign(as.numeric(test_df$Diabetes_binary) - preds_second_order_step)
deviance_residuals_second_order_step <- sqrt(-2 * (as.numeric(test_df$Diabetes_binary) * log(preds_second_order_step) + (1 - as.numeric(test_df$Diabetes_binary)) * log(1 - preds_second_order_step)))

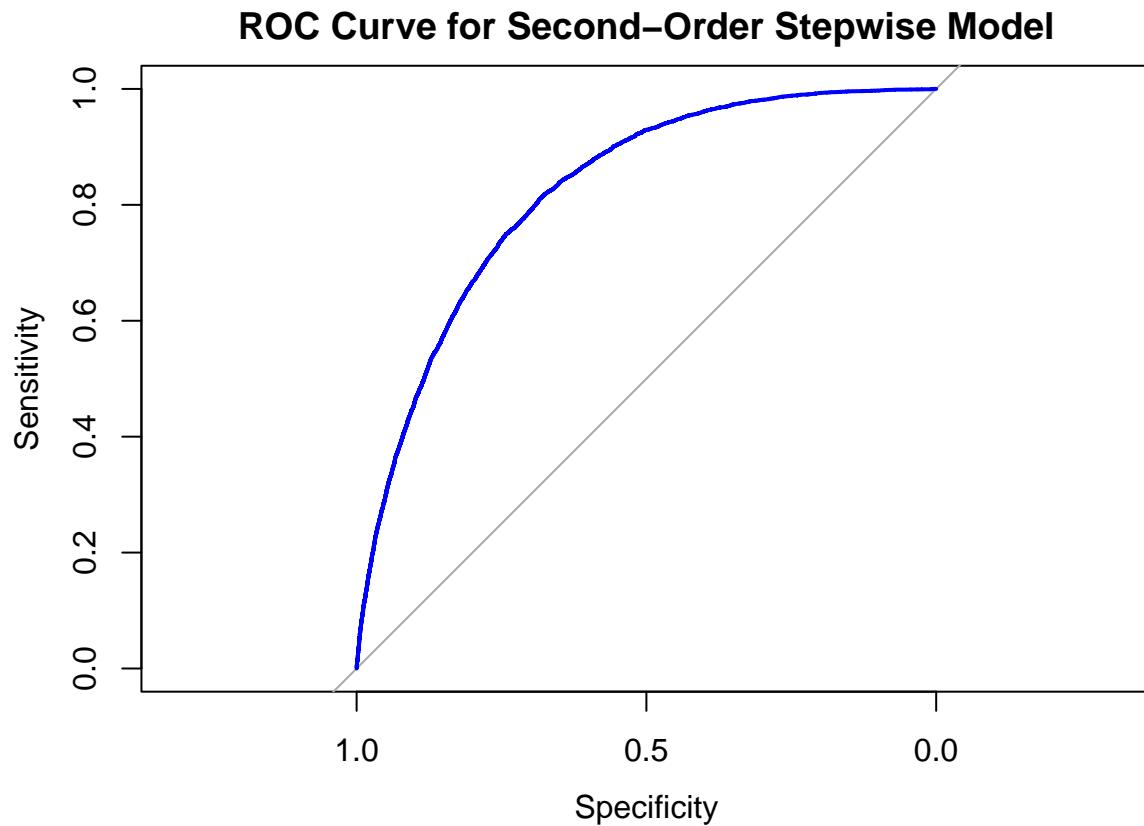
## Warning in sqrt(-2 * (as.numeric(test_df$Diabetes_binary) * log(preds_second_order_step) + (1 - as.numeric(test_df$Diabetes_binary)) * log(1 - preds_second_order_step))) : NaNs produced

plot(deviance_residuals_second_order_step, main = "Deviance Residuals of Second-Order Stepwise Model",
      ylab = "Deviance Residuals", xlab = "Index", pch = 20)
abline(h = 0, col = "red")

```



```
# ROC curve
plot(roc_obj_second_order_step, col = "blue", main = "ROC Curve for Second-Order Stepwise Model")
```



```
cat("AUC for Second-Order Stepwise Model:", auc_second_order_step, "\n")
```

```
## AUC for Second-Order Stepwise Model: 0.8223127
```

```
# Fit the full interaction model
interaction_model_full <- glm(Diabetes_binary ~ (Age_group + BMI + MentHlthCategory + PhysHlthCategory +
                                                 HighBP + HighChol + CholCheck + Smoker + Stroke + Heart +
                                                 PhysActivity + Fruits + Veggies + HvyAlcoholConsump +
                                                 NoDocbcCost + GenHealthCategory + DiffWalk + Sex)^2,
                                 data = train_df, family = "binomial")

# Perform stepwise selection based on p-values
interaction_model_step <- step(interaction_model_full, direction = "forward", trace = 0)

# Print the variables included in the final model
cat("Variables used in the Interaction Model after stepwise selection:\n")
```

```
## Variables used in the Interaction Model after stepwise selection:
```

```

print(names(coef(interaction_model_step)))

## [1] "(Intercept)"
## [2] "Age_groupOlder"
## [3] "Age_groupYoung"
## [4] "BMI"
## [5] "MentHlthCategoryLow"
## [6] "MentHlthCategoryModerate"
## [7] "PhysHlthCategoryLow"
## [8] "PhysHlthCategoryModerate"
## [9] "Education_levelLow Education"
## [10] "Education_levelMedium Education"
## [11] "HighBP1"
## [12] "HighChol1"
## [13] "CholCheck1"
## [14] "Smoker1"
## [15] "Stroke1"
## [16] "HeartDiseaseorAttack1"
## [17] "PhysActivity1"
## [18] "Fruits1"
## [19] "Veggies1"
## [20] "HvyAlcoholConsump1"
## [21] "AnyHealthcare1"
## [22] "NoDocbcCost1"
## [23] "GenHealthCategoryGood Health"
## [24] "GenHealthCategoryPoor Health"
## [25] "DiffWalk1"
## [26] "Sex1"
## [27] "Age_groupOlder:BMI"
## [28] "Age_groupYoung:BMI"
## [29] "Age_groupOlder:MentHlthCategoryLow"
## [30] "Age_groupYoung:MentHlthCategoryLow"
## [31] "Age_groupOlder:MentHlthCategoryModerate"
## [32] "Age_groupYoung:MentHlthCategoryModerate"
## [33] "Age_groupOlder:PhysHlthCategoryLow"
## [34] "Age_groupYoung:PhysHlthCategoryLow"
## [35] "Age_groupOlder:PhysHlthCategoryModerate"
## [36] "Age_groupYoung:PhysHlthCategoryModerate"
## [37] "Age_groupOlder:Education_levelLow Education"
## [38] "Age_groupYoung:Education_levelLow Education"
## [39] "Age_groupOlder:Education_levelMedium Education"
## [40] "Age_groupYoung:Education_levelMedium Education"
## [41] "Age_groupOlder:HighBP1"
## [42] "Age_groupYoung:HighBP1"
## [43] "Age_groupOlder:HighChol1"
## [44] "Age_groupYoung:HighChol1"
## [45] "Age_groupOlder:CholCheck1"
## [46] "Age_groupYoung:CholCheck1"
## [47] "Age_groupOlder:Smoker1"
## [48] "Age_groupYoung:Smoker1"
## [49] "Age_groupOlder:Stroke1"
## [50] "Age_groupYoung:Stroke1"
## [51] "Age_groupOlder:HeartDiseaseorAttack1"

```

```

## [52] "Age_groupYoung:HeartDiseaseorAttack1"
## [53] "Age_groupOlder:PhysActivity1"
## [54] "Age_groupYoung:PhysActivity1"
## [55] "Age_groupOlder:Fruits1"
## [56] "Age_groupYoung:Fruits1"
## [57] "Age_groupOlder:Veggies1"
## [58] "Age_groupYoung:Veggies1"
## [59] "Age_groupOlder:HvyAlcoholConsump1"
## [60] "Age_groupYoung:HvyAlcoholConsump1"
## [61] "Age_groupOlder:AnyHealthcare1"
## [62] "Age_groupYoung:AnyHealthcare1"
## [63] "Age_groupOlder:NoDocbcCost1"
## [64] "Age_groupYoung:NoDocbcCost1"
## [65] "Age_groupOlder:GenHealthCategoryGood Health"
## [66] "Age_groupYoung:GenHealthCategoryGood Health"
## [67] "Age_groupOlder:GenHealthCategoryPoor Health"
## [68] "Age_groupYoung:GenHealthCategoryPoor Health"
## [69] "Age_groupOlder:DiffWalk1"
## [70] "Age_groupYoung:DiffWalk1"
## [71] "Age_groupOlder:Sex1"
## [72] "Age_groupYoung:Sex1"
## [73] "BMI:MentHlthCategoryLow"
## [74] "BMI:MentHlthCategoryModerate"
## [75] "BMI:PhysHlthCategoryLow"
## [76] "BMI:PhysHlthCategoryModerate"
## [77] "BMI:Education_levelLow Education"
## [78] "BMI:Education_levelMedium Education"
## [79] "BMI:HighBP1"
## [80] "BMI:HighChol1"
## [81] "BMI:CholCheck1"
## [82] "BMI:Smoker1"
## [83] "BMI:Stroke1"
## [84] "BMI:HeartDiseaseorAttack1"
## [85] "BMI:PhysActivity1"
## [86] "BMI:Fruits1"
## [87] "BMI:Veggies1"
## [88] "BMI:HvyAlcoholConsump1"
## [89] "BMI:AnyHealthcare1"
## [90] "BMI:NoDocbcCost1"
## [91] "BMI:GenHealthCategoryGood Health"
## [92] "BMI:GenHealthCategoryPoor Health"
## [93] "BMI:DiffWalk1"
## [94] "BMI:Sex1"
## [95] "MentHlthCategoryLow:PhysHlthCategoryLow"
## [96] "MentHlthCategoryModerate:PhysHlthCategoryLow"
## [97] "MentHlthCategoryLow:PhysHlthCategoryModerate"
## [98] "MentHlthCategoryModerate:PhysHlthCategoryModerate"
## [99] "MentHlthCategoryLow:Education_levelLow Education"
## [100] "MentHlthCategoryModerate:Education_levelLow Education"
## [101] "MentHlthCategoryLow:Education_levelMedium Education"
## [102] "MentHlthCategoryModerate:Education_levelMedium Education"
## [103] "MentHlthCategoryLow:HighBP1"
## [104] "MentHlthCategoryModerate:HighBP1"
## [105] "MentHlthCategoryLow:HighChol1"

```

```

## [106] "MentHlthCategoryModerate:HighChol1"
## [107] "MentHlthCategoryLow:CholCheck1"
## [108] "MentHlthCategoryModerate:CholCheck1"
## [109] "MentHlthCategoryLow:Smoker1"
## [110] "MentHlthCategoryModerate:Smoker1"
## [111] "MentHlthCategoryLow:Stroke1"
## [112] "MentHlthCategoryModerate:Stroke1"
## [113] "MentHlthCategoryLow:HeartDiseaseorAttack1"
## [114] "MentHlthCategoryModerate:HeartDiseaseorAttack1"
## [115] "MentHlthCategoryLow:PhysActivity1"
## [116] "MentHlthCategoryModerate:PhysActivity1"
## [117] "MentHlthCategoryLow:Fruits1"
## [118] "MentHlthCategoryModerate:Fruits1"
## [119] "MentHlthCategoryLow:Veggies1"
## [120] "MentHlthCategoryModerate:Veggies1"
## [121] "MentHlthCategoryLow:HvyAlcoholConsump1"
## [122] "MentHlthCategoryModerate:HvyAlcoholConsump1"
## [123] "MentHlthCategoryLow:AnyHealthcare1"
## [124] "MentHlthCategoryModerate:AnyHealthcare1"
## [125] "MentHlthCategoryLow:NoDocbcCost1"
## [126] "MentHlthCategoryModerate:NoDocbcCost1"
## [127] "MentHlthCategoryLow:GenHealthCategoryGood Health"
## [128] "MentHlthCategoryModerate:GenHealthCategoryGood Health"
## [129] "MentHlthCategoryLow:GenHealthCategoryPoor Health"
## [130] "MentHlthCategoryModerate:GenHealthCategoryPoor Health"
## [131] "MentHlthCategoryLow:DiffWalk1"
## [132] "MentHlthCategoryModerate:DiffWalk1"
## [133] "MentHlthCategoryLow:Sex1"
## [134] "MentHlthCategoryModerate:Sex1"
## [135] "PhysHlthCategoryLow:Education_levelLow Education"
## [136] "PhysHlthCategoryModerate:Education_levelLow Education"
## [137] "PhysHlthCategoryLow:Education_levelMedium Education"
## [138] "PhysHlthCategoryModerate:Education_levelMedium Education"
## [139] "PhysHlthCategoryLow:HighBP1"
## [140] "PhysHlthCategoryModerate:HighBP1"
## [141] "PhysHlthCategoryLow:HighChol1"
## [142] "PhysHlthCategoryModerate:HighChol1"
## [143] "PhysHlthCategoryLow:CholCheck1"
## [144] "PhysHlthCategoryModerate:CholCheck1"
## [145] "PhysHlthCategoryLow:Smoker1"
## [146] "PhysHlthCategoryModerate:Smoker1"
## [147] "PhysHlthCategoryLow:Stroke1"
## [148] "PhysHlthCategoryModerate:Stroke1"
## [149] "PhysHlthCategoryLow:HeartDiseaseorAttack1"
## [150] "PhysHlthCategoryModerate:HeartDiseaseorAttack1"
## [151] "PhysHlthCategoryLow:PhysActivity1"
## [152] "PhysHlthCategoryModerate:PhysActivity1"
## [153] "PhysHlthCategoryLow:Fruits1"
## [154] "PhysHlthCategoryModerate:Fruits1"
## [155] "PhysHlthCategoryLow:Veggies1"
## [156] "PhysHlthCategoryModerate:Veggies1"
## [157] "PhysHlthCategoryLow:HvyAlcoholConsump1"
## [158] "PhysHlthCategoryModerate:HvyAlcoholConsump1"
## [159] "PhysHlthCategoryLow:AnyHealthcare1"

```

```

## [160] "PhysHlthCategoryModerate:AnyHealthcare1"
## [161] "PhysHlthCategoryLow:NoDocbcCost1"
## [162] "PhysHlthCategoryModerate:NoDocbcCost1"
## [163] "PhysHlthCategoryLow:GenHealthCategoryGood Health"
## [164] "PhysHlthCategoryModerate:GenHealthCategoryGood Health"
## [165] "PhysHlthCategoryLow:GenHealthCategoryPoor Health"
## [166] "PhysHlthCategoryModerate:GenHealthCategoryPoor Health"
## [167] "PhysHlthCategoryLow:DiffWalk1"
## [168] "PhysHlthCategoryModerate:DiffWalk1"
## [169] "PhysHlthCategoryLow:Sex1"
## [170] "PhysHlthCategoryModerate:Sex1"
## [171] "Education_levelLow Education:HighBP1"
## [172] "Education_levelMedium Education:HighBP1"
## [173] "Education_levelLow Education:HighChol1"
## [174] "Education_levelMedium Education:HighChol1"
## [175] "Education_levelLow Education:CholCheck1"
## [176] "Education_levelMedium Education:CholCheck1"
## [177] "Education_levelLow Education:Smoker1"
## [178] "Education_levelMedium Education:Smoker1"
## [179] "Education_levelLow Education:Stroke1"
## [180] "Education_levelMedium Education:Stroke1"
## [181] "Education_levelLow Education:HeartDiseaseorAttack1"
## [182] "Education_levelMedium Education:HeartDiseaseorAttack1"
## [183] "Education_levelLow Education:PhysActivity1"
## [184] "Education_levelMedium Education:PhysActivity1"
## [185] "Education_levelLow Education:Fruits1"
## [186] "Education_levelMedium Education:Fruits1"
## [187] "Education_levelLow Education:Veggies1"
## [188] "Education_levelMedium Education:Veggies1"
## [189] "Education_levelLow Education:HvyAlcoholConsump1"
## [190] "Education_levelMedium Education:HvyAlcoholConsump1"
## [191] "Education_levelLow Education:AnyHealthcare1"
## [192] "Education_levelMedium Education:AnyHealthcare1"
## [193] "Education_levelLow Education:NoDocbcCost1"
## [194] "Education_levelMedium Education:NoDocbcCost1"
## [195] "Education_levelLow Education:GenHealthCategoryGood Health"
## [196] "Education_levelMedium Education:GenHealthCategoryGood Health"
## [197] "Education_levelLow Education:GenHealthCategoryPoor Health"
## [198] "Education_levelMedium Education:GenHealthCategoryPoor Health"
## [199] "Education_levelLow Education:DiffWalk1"
## [200] "Education_levelMedium Education:DiffWalk1"
## [201] "Education_levelLow Education:Sex1"
## [202] "Education_levelMedium Education:Sex1"
## [203] "HighBP1:HighChol1"
## [204] "HighBP1:CholCheck1"
## [205] "HighBP1:Smoker1"
## [206] "HighBP1:Stroke1"
## [207] "HighBP1:HeartDiseaseorAttack1"
## [208] "HighBP1:PhysActivity1"
## [209] "HighBP1:Fruits1"
## [210] "HighBP1:Veggies1"
## [211] "HighBP1:HvyAlcoholConsump1"
## [212] "HighBP1:AnyHealthcare1"
## [213] "HighBP1:NoDocbcCost1"

```

```

## [214] "HighBP1:GenHealthCategoryGood Health"
## [215] "HighBP1:GenHealthCategoryPoor Health"
## [216] "HighBP1:DiffWalk1"
## [217] "HighBP1:Sex1"
## [218] "HighChol1:CholCheck1"
## [219] "HighChol1:Smoker1"
## [220] "HighChol1:Stroke1"
## [221] "HighChol1:HeartDiseaseorAttack1"
## [222] "HighChol1:PhysActivity1"
## [223] "HighChol1:Fruits1"
## [224] "HighChol1:Veggies1"
## [225] "HighChol1:HvyAlcoholConsump1"
## [226] "HighChol1:AnyHealthcare1"
## [227] "HighChol1:NoDocbcCost1"
## [228] "HighChol1:GenHealthCategoryGood Health"
## [229] "HighChol1:GenHealthCategoryPoor Health"
## [230] "HighChol1:DiffWalk1"
## [231] "HighChol1:Sex1"
## [232] "CholCheck1:Smoker1"
## [233] "CholCheck1:Stroke1"
## [234] "CholCheck1:HeartDiseaseorAttack1"
## [235] "CholCheck1:PhysActivity1"
## [236] "CholCheck1:Fruits1"
## [237] "CholCheck1:Veggies1"
## [238] "CholCheck1:HvyAlcoholConsump1"
## [239] "CholCheck1:AnyHealthcare1"
## [240] "CholCheck1:NoDocbcCost1"
## [241] "CholCheck1:GenHealthCategoryGood Health"
## [242] "CholCheck1:GenHealthCategoryPoor Health"
## [243] "CholCheck1:DiffWalk1"
## [244] "CholCheck1:Sex1"
## [245] "Smoker1:Stroke1"
## [246] "Smoker1:HeartDiseaseorAttack1"
## [247] "Smoker1:PhysActivity1"
## [248] "Smoker1:Fruits1"
## [249] "Smoker1:Veggies1"
## [250] "Smoker1:HvyAlcoholConsump1"
## [251] "Smoker1:AnyHealthcare1"
## [252] "Smoker1:NoDocbcCost1"
## [253] "Smoker1:GenHealthCategoryGood Health"
## [254] "Smoker1:GenHealthCategoryPoor Health"
## [255] "Smoker1:DiffWalk1"
## [256] "Smoker1:Sex1"
## [257] "Stroke1:HeartDiseaseorAttack1"
## [258] "Stroke1:PhysActivity1"
## [259] "Stroke1:Fruits1"
## [260] "Stroke1:Veggies1"
## [261] "Stroke1:HvyAlcoholConsump1"
## [262] "Stroke1:AnyHealthcare1"
## [263] "Stroke1:NoDocbcCost1"
## [264] "Stroke1:GenHealthCategoryGood Health"
## [265] "Stroke1:GenHealthCategoryPoor Health"
## [266] "Stroke1:DiffWalk1"
## [267] "Stroke1:Sex1"

```

```

## [268] "HeartDiseaseorAttack1:PhysActivity1"
## [269] "HeartDiseaseorAttack1:Fruits1"
## [270] "HeartDiseaseorAttack1:Veggies1"
## [271] "HeartDiseaseorAttack1:HvyAlcoholConsump1"
## [272] "HeartDiseaseorAttack1:AnyHealthcare1"
## [273] "HeartDiseaseorAttack1:NoDocbcCost1"
## [274] "HeartDiseaseorAttack1:GenHealthCategoryGood Health"
## [275] "HeartDiseaseorAttack1:GenHealthCategoryPoor Health"
## [276] "HeartDiseaseorAttack1:DiffWalk1"
## [277] "HeartDiseaseorAttack1:Sex1"
## [278] "PhysActivity1:Fruits1"
## [279] "PhysActivity1:Veggies1"
## [280] "PhysActivity1:HvyAlcoholConsump1"
## [281] "PhysActivity1:AnyHealthcare1"
## [282] "PhysActivity1:NoDocbcCost1"
## [283] "PhysActivity1:GenHealthCategoryGood Health"
## [284] "PhysActivity1:GenHealthCategoryPoor Health"
## [285] "PhysActivity1:DiffWalk1"
## [286] "PhysActivity1:Sex1"
## [287] "Fruits1:Veggies1"
## [288] "Fruits1:HvyAlcoholConsump1"
## [289] "Fruits1:AnyHealthcare1"
## [290] "Fruits1:NoDocbcCost1"
## [291] "Fruits1:GenHealthCategoryGood Health"
## [292] "Fruits1:GenHealthCategoryPoor Health"
## [293] "Fruits1:DiffWalk1"
## [294] "Fruits1:Sex1"
## [295] "Veggies1:HvyAlcoholConsump1"
## [296] "Veggies1:AnyHealthcare1"
## [297] "Veggies1:NoDocbcCost1"
## [298] "Veggies1:GenHealthCategoryGood Health"
## [299] "Veggies1:GenHealthCategoryPoor Health"
## [300] "Veggies1:DiffWalk1"
## [301] "Veggies1:Sex1"
## [302] "HvyAlcoholConsump1:AnyHealthcare1"
## [303] "HvyAlcoholConsump1:NoDocbcCost1"
## [304] "HvyAlcoholConsump1:GenHealthCategoryGood Health"
## [305] "HvyAlcoholConsump1:GenHealthCategoryPoor Health"
## [306] "HvyAlcoholConsump1:DiffWalk1"
## [307] "HvyAlcoholConsump1:Sex1"
## [308] "AnyHealthcare1:NoDocbcCost1"
## [309] "AnyHealthcare1:GenHealthCategoryGood Health"
## [310] "AnyHealthcare1:GenHealthCategoryPoor Health"
## [311] "AnyHealthcare1:DiffWalk1"
## [312] "AnyHealthcare1:Sex1"
## [313] "NoDocbcCost1:GenHealthCategoryGood Health"
## [314] "NoDocbcCost1:GenHealthCategoryPoor Health"
## [315] "NoDocbcCost1:DiffWalk1"
## [316] "NoDocbcCost1:Sex1"
## [317] "GenHealthCategoryGood Health:DiffWalk1"
## [318] "GenHealthCategoryPoor Health:DiffWalk1"
## [319] "GenHealthCategoryGood Health:Sex1"
## [320] "GenHealthCategoryPoor Health:Sex1"
## [321] "DiffWalk1:Sex1"

```

```

# Summary of the interaction model
summary(interaction_model_step)

## 
## Call:
## glm(formula = Diabetes_binary ~ (Age_group + BMI + MentHlthCategory +
##   PhysHlthCategory + Education_level + HighBP + HighChol +
##   CholCheck + Smoker + Stroke + HeartDiseaseorAttack + PhysActivity +
##   Fruits + Veggies + HvyAlcoholConsump + AnyHealthcare + NoDocbcCost +
##   GenHealthCategory + DiffWalk + Sex)^2, family = "binomial",
##   data = train_df)
##
## Coefficients:
##                               Estimate
## (Intercept)                -6.033e+00
## Age_groupOlder               1.014e+00
## Age_groupYoung              -1.153e+00
## BMI                          6.788e-02
## MentHlthCategoryLow          9.682e-02
## MentHlthCategoryModerate    -7.338e-01
## PhysHlthCategoryLow          2.242e-01
## PhysHlthCategoryModerate    8.741e-02
## Education_levelLow Education 8.529e-01
## Education_levelMedium Education 4.444e-01
## HighBP1                      1.550e+00
## HighChol1                     9.881e-01
## CholCheck1                    8.452e-01
## Smoker1                       -1.167e-01
## Stroke1                       6.392e-01
## HeartDiseaseorAttack1        1.451e+00
## PhysActivity1                 -2.593e-01
## Fruits1                        4.877e-02
## Veggies1                      -2.330e-02
## HvyAlcoholConsump1            -1.313e+00
## AnyHealthcare1                 -9.921e-01
## NoDocbcCost1                  2.325e-01
## GenHealthCategoryGood Health -1.247e+00
## GenHealthCategoryPoor Health  1.101e+00
## DiffWalk1                      2.553e-01
## Sex1                           1.943e-01
## Age_groupOlder:BMI             1.046e-02
## Age_groupYoung:BMI             -7.898e-03
## Age_groupOlder:MentHlthCategoryLow -1.153e-01
## Age_groupYoung:MentHlthCategoryLow 1.512e-01
## Age_groupOlder:MentHlthCategoryModerate -1.123e-01
## Age_groupYoung:MentHlthCategoryModerate 1.181e-01
## Age_groupOlder:PhysHlthCategoryLow 2.895e-02
## Age_groupYoung:PhysHlthCategoryLow -4.372e-02
## Age_groupOlder:PhysHlthCategoryModerate 8.146e-02
## Age_groupYoung:PhysHlthCategoryModerate -3.555e-02
## Age_groupOlder:Education_levelLow Education 1.050e-01
## Age_groupYoung:Education_levelLow Education 2.682e-02
## Age_groupOlder:Education_levelMedium Education 1.070e-02

```

## Age_groupYoung:Education_levelMedium Education	7.879e-02
## Age_groupOlder:HighBP1	-2.220e-01
## Age_groupYoung:HighBP1	-8.955e-02
## Age_groupOlder:HighChol1	-2.739e-01
## Age_groupYoung:HighChol1	3.586e-01
## Age_groupOlder:CholCheck1	-3.756e-01
## Age_groupYoung:CholCheck1	1.716e-01
## Age_groupOlder:Smoker1	-1.047e-03
## Age_groupYoung:Smoker1	-8.661e-03
## Age_groupOlder:Stroke1	-1.890e-01
## Age_groupYoung:Stroke1	-1.808e-01
## Age_groupOlder:HeartDiseaseorAttack1	-1.564e-01
## Age_groupYoung:HeartDiseaseorAttack1	5.596e-02
## Age_groupOlder:PhysActivity1	-7.572e-02
## Age_groupYoung:PhysActivity1	1.933e-02
## Age_groupOlder:Fruits1	-4.862e-02
## Age_groupYoung:Fruits1	2.313e-02
## Age_groupOlder:Veggies1	5.407e-02
## Age_groupYoung:Veggies1	5.677e-02
## Age_groupOlder:HvyAlcoholConsump1	-4.274e-02
## Age_groupYoung:HvyAlcoholConsump1	6.833e-01
## Age_groupOlder:AnyHealthcare1	-2.058e-01
## Age_groupYoung:AnyHealthcare1	1.474e-01
## Age_groupOlder:NoDocbcCost1	9.513e-02
## Age_groupYoung:NoDocbcCost1	1.271e-01
## Age_groupOlder:GenHealthCategoryGood Health	1.991e-01
## Age_groupYoung:GenHealthCategoryGood Health	1.351e-01
## Age_groupOlder:GenHealthCategoryPoor Health	-1.884e-01
## Age_groupYoung:GenHealthCategoryPoor Health	1.789e-01
## Age_groupOlder:DiffWalk1	-7.376e-02
## Age_groupYoung:DiffWalk1	1.827e-01
## Age_groupOlder:Sex1	2.046e-01
## Age_groupYoung:Sex1	-3.038e-01
## BMI:MentHlthCategoryLow	-2.172e-05
## BMI:MentHlthCategoryModerate	5.383e-03
## BMI:PhysHlthCategoryLow	-4.034e-03
## BMI:PhysHlthCategoryModerate	-3.172e-05
## BMI:Education_levelLow Education	-1.974e-02
## BMI:Education_levelMedium Education	-1.113e-02
## BMI:HighBP1	-4.373e-03
## BMI:HighChol1	-6.485e-03
## BMI:CholCheck1	4.693e-03
## BMI:Smoker1	9.711e-03
## BMI:Stroke1	5.995e-03
## BMI:HeartDiseaseorAttack1	1.224e-03
## BMI:PhysActivity1	3.665e-03
## BMI:Fruits1	-2.738e-03
## BMI:Veggies1	6.283e-03
## BMI:HvyAlcoholConsump1	1.056e-02
## BMI:AnyHealthcare1	1.654e-02
## BMI:NoDocbcCost1	-2.872e-05
## BMI:GenHealthCategoryGood Health	1.400e-02
## BMI:GenHealthCategoryPoor Health	-1.556e-03
## BMI:DiffWalk1	-1.002e-02

## BMI:Sex1	-6.325e-03
## MentHlthCategoryLow:PhysHlthCategoryLow	1.323e-01
## MentHlthCategoryModerate:PhysHlthCategoryLow	7.564e-02
## MentHlthCategoryLow:PhysHlthCategoryModerate	9.195e-03
## MentHlthCategoryModerate:PhysHlthCategoryModerate	3.797e-02
## MentHlthCategoryLow:Education_levelLow Education	-6.553e-02
## MentHlthCategoryModerate:Education_levelLow Education	8.257e-02
## MentHlthCategoryLow:Education_levelMedium Education	9.339e-03
## MentHlthCategoryModerate:Education_levelMedium Education	1.114e-01
## MentHlthCategoryLow:HighBP1	-1.160e-02
## MentHlthCategoryModerate:HighBP1	2.487e-02
## MentHlthCategoryLow:HighChol1	-1.522e-01
## MentHlthCategoryModerate:HighChol1	-5.356e-02
## MentHlthCategoryLow:CholCheck1	1.417e-01
## MentHlthCategoryModerate:CholCheck1	3.746e-01
## MentHlthCategoryLow:Smoker1	-2.694e-02
## MentHlthCategoryModerate:Smoker1	-3.806e-02
## MentHlthCategoryLow:Stroke1	-5.147e-02
## MentHlthCategoryModerate:Stroke1	1.275e-01
## MentHlthCategoryLow:HeartDiseaseorAttack1	-5.189e-02
## MentHlthCategoryModerate:HeartDiseaseorAttack1	-1.050e-01
## MentHlthCategoryLow:PhysActivity1	-2.994e-02
## MentHlthCategoryModerate:PhysActivity1	-6.255e-02
## MentHlthCategoryLow:Fruits1	-1.712e-01
## MentHlthCategoryModerate:Fruits1	-1.754e-01
## MentHlthCategoryLow:Veggies1	-1.557e-01
## MentHlthCategoryModerate:Veggies1	-3.678e-03
## MentHlthCategoryLow:HvyAlcoholConsump1	-9.405e-02
## MentHlthCategoryModerate:HvyAlcoholConsump1	4.454e-03
## MentHlthCategoryLow:AnyHealthcare1	1.380e-01
## MentHlthCategoryModerate:AnyHealthcare1	2.293e-01
## MentHlthCategoryLow:NoDocbcCost1	1.906e-02
## MentHlthCategoryModerate:NoDocbcCost1	2.119e-02
## MentHlthCategoryLow:GenHealthCategoryGood Health	-6.829e-02
## MentHlthCategoryModerate:GenHealthCategoryGood Health	-7.430e-02
## MentHlthCategoryLow:GenHealthCategoryPoor Health	2.318e-02
## MentHlthCategoryModerate:GenHealthCategoryPoor Health	3.322e-02
## MentHlthCategoryLow:DiffWalk1	-2.899e-03
## MentHlthCategoryModerate:DiffWalk1	4.806e-02
## MentHlthCategoryLow:Sex1	2.080e-02
## MentHlthCategoryModerate:Sex1	1.239e-01
## PhysHlthCategoryLow:Education_levelLow Education	-1.993e-01
## PhysHlthCategoryModerate:Education_levelLow Education	-9.526e-02
## PhysHlthCategoryLow:Education_levelMedium Education	-9.439e-03
## PhysHlthCategoryModerate:Education_levelMedium Education	3.642e-02
## PhysHlthCategoryLow:HighBP1	-8.794e-02
## PhysHlthCategoryModerate:HighBP1	-8.100e-02
## PhysHlthCategoryLow:HighChol1	8.674e-03
## PhysHlthCategoryModerate:HighChol1	7.197e-02
## PhysHlthCategoryLow:CholCheck1	-1.990e-02
## PhysHlthCategoryModerate:CholCheck1	-2.520e-02
## PhysHlthCategoryLow:Smoker1	-8.961e-02
## PhysHlthCategoryModerate:Smoker1	-9.027e-02
## PhysHlthCategoryLow:Stroke1	-4.002e-02

## PhysHlthCategoryModerate:Stroke1	-6.090e-02
## PhysHlthCategoryLow:HeartDiseaseorAttack1	-3.735e-02
## PhysHlthCategoryModerate:HeartDiseaseorAttack1	-3.514e-02
## PhysHlthCategoryLow:PhysActivity1	2.772e-02
## PhysHlthCategoryModerate:PhysActivity1	-2.572e-04
## PhysHlthCategoryLow:Fruits1	5.624e-02
## PhysHlthCategoryModerate:Fruits1	4.509e-02
## PhysHlthCategoryLow:Veggies1	6.339e-02
## PhysHlthCategoryModerate:Veggies1	7.267e-02
## PhysHlthCategoryLow:HvyAlcoholConsump1	2.273e-01
## PhysHlthCategoryModerate:HvyAlcoholConsump1	5.745e-02
## PhysHlthCategoryLow:AnyHealthcare1	-1.535e-01
## PhysHlthCategoryModerate:AnyHealthcare1	-1.978e-01
## PhysHlthCategoryLow>NoDocbcCost1	-9.730e-02
## PhysHlthCategoryModerate>NoDocbcCost1	1.663e-02
## PhysHlthCategoryLow:GenHealthCategoryGood Health	-2.957e-01
## PhysHlthCategoryModerate:GenHealthCategoryGood Health	-1.996e-01
## PhysHlthCategoryLow:GenHealthCategoryPoor Health	5.431e-02
## PhysHlthCategoryModerate:GenHealthCategoryPoor Health	-7.120e-02
## PhysHlthCategoryLow:DiffWalk1	4.717e-02
## PhysHlthCategoryModerate:DiffWalk1	1.800e-01
## PhysHlthCategoryLow:Sex1	-3.411e-02
## PhysHlthCategoryModerate:Sex1	4.622e-02
## Education_levelLow Education:HighBP1	-1.198e-01
## Education_levelMedium Education:HighBP1	-1.014e-01
## Education_levelLow Education:HighChol1	1.600e-01
## Education_levelMedium Education:HighChol1	-2.575e-02
## Education_levelLow Education:CholCheck1	3.441e-02
## Education_levelMedium Education:CholCheck1	-1.604e-01
## Education_levelLow Education:Smoker1	-4.798e-02
## Education_levelMedium Education:Smoker1	-7.594e-02
## Education_levelLow Education:Stroke1	-3.416e-01
## Education_levelMedium Education:Stroke1	-1.246e-01
## Education_levelLow Education:HeartDiseaseorAttack1	1.736e-01
## Education_levelMedium Education:HeartDiseaseorAttack1	-5.826e-02
## Education_levelLow Education:PhysActivity1	2.457e-02
## Education_levelMedium Education:PhysActivity1	5.312e-02
## Education_levelLow Education:Fruits1	1.270e-01
## Education_levelMedium Education:Fruits1	3.230e-02
## Education_levelLow Education:Veggies1	9.512e-02
## Education_levelMedium Education:Veggies1	5.166e-02
## Education_levelLow Education:HvyAlcoholConsump1	-3.932e-01
## Education_levelMedium Education:HvyAlcoholConsump1	-3.640e-03
## Education_levelLow Education:AnyHealthcare1	7.496e-02
## Education_levelMedium Education:AnyHealthcare1	1.291e-01
## Education_levelLow Education>NoDocbcCost1	1.120e-02
## Education_levelMedium Education>NoDocbcCost1	6.037e-02
## Education_levelLow Education:GenHealthCategoryGood Health	1.170e-01
## Education_levelMedium Education:GenHealthCategoryGood Health	1.413e-01
## Education_levelLow Education:GenHealthCategoryPoor Health	1.705e-02
## Education_levelMedium Education:GenHealthCategoryPoor Health	2.260e-03
## Education_levelLow Education:DiffWalk1	-5.325e-02
## Education_levelMedium Education:DiffWalk1	5.223e-02
## Education_levelLow Education:Sex1	-2.749e-01

## Education_levelMedium Education:Sex1	-4.356e-02
## HighBP1:HighChol1	5.243e-02
## HighBP1:CholCheck1	-4.089e-01
## HighBP1:Smoker1	-5.410e-02
## HighBP1:Stroke1	-1.166e-01
## HighBP1:HeartDiseaseorAttack1	-1.691e-01
## HighBP1:PhysActivity1	2.529e-02
## HighBP1:Fruits1	5.205e-02
## HighBP1:Veggies1	-7.990e-02
## HighBP1:HvyAlcoholConsump1	-1.722e-01
## HighBP1:AnyHealthcare1	-3.153e-02
## HighBP1:NoDocbcCost1	5.588e-02
## HighBP1:GenHealthCategoryGood Health	2.415e-01
## HighBP1:GenHealthCategoryPoor Health	-3.623e-02
## HighBP1:DiffWalk1	-2.695e-02
## HighBP1:Sex1	-1.126e-01
## HighChol1:CholCheck1	5.497e-02
## HighChol1:Smoker1	3.654e-02
## HighChol1:Stroke1	-5.408e-02
## HighChol1:HeartDiseaseorAttack1	-2.578e-01
## HighChol1:PhysActivity1	-4.063e-02
## HighChol1:Fruits1	-5.452e-03
## HighChol1:Veggies1	-2.626e-02
## HighChol1:HvyAlcoholConsump1	8.194e-02
## HighChol1:AnyHealthcare1	1.742e-02
## HighChol1:NoDocbcCost1	8.158e-02
## HighChol1:GenHealthCategoryGood Health	-5.232e-03
## HighChol1:GenHealthCategoryPoor Health	-1.932e-02
## HighChol1:DiffWalk1	3.144e-02
## HighChol1:Sex1	-1.641e-01
## CholCheck1:Smoker1	-1.155e-02
## CholCheck1:Stroke1	3.539e-03
## CholCheck1:HeartDiseaseorAttack1	-6.509e-01
## CholCheck1:PhysActivity1	8.079e-02
## CholCheck1:Fruits1	2.219e-01
## CholCheck1:Veggies1	-1.868e-02
## CholCheck1:HvyAlcoholConsump1	2.325e-01
## CholCheck1:AnyHealthcare1	5.534e-01
## CholCheck1:NoDocbcCost1	-1.861e-01
## CholCheck1:GenHealthCategoryGood Health	2.375e-01
## CholCheck1:GenHealthCategoryPoor Health	-2.599e-01
## CholCheck1:DiffWalk1	5.328e-02
## CholCheck1:Sex1	7.837e-02
## Smoker1:Stroke1	-6.712e-02
## Smoker1:HeartDiseaseorAttack1	-5.041e-03
## Smoker1:PhysActivity1	2.812e-02
## Smoker1:Fruits1	7.197e-02
## Smoker1:Veggies1	1.384e-02
## Smoker1:HvyAlcoholConsump1	7.308e-02
## Smoker1:AnyHealthcare1	-9.279e-02
## Smoker1:NoDocbcCost1	-1.417e-01
## Smoker1:GenHealthCategoryGood Health	8.141e-02
## Smoker1:GenHealthCategoryPoor Health	-7.034e-02
## Smoker1:DiffWalk1	-7.888e-02

## Smoker1:Sex1	3.202e-02
## Stroke1:HeartDiseaseorAttack1	1.396e-01
## Stroke1:PhysActivity1	-1.498e-01
## Stroke1:Fruits1	4.466e-02
## Stroke1:Veggies1	-5.955e-02
## Stroke1:HvyAlcoholConsump1	4.073e-02
## Stroke1:AnyHealthcare1	-9.203e-02
## Stroke1:NoDocbcCost1	-1.529e-01
## Stroke1:GenHealthCategoryGood Health	1.803e-01
## Stroke1:GenHealthCategoryPoor Health	-1.083e-01
## Stroke1:DiffWalk1	7.853e-02
## Stroke1:Sex1	-2.136e-02
## HeartDiseaseorAttack1:PhysActivity1	-7.031e-02
## HeartDiseaseorAttack1:Fruits1	1.751e-02
## HeartDiseaseorAttack1:Veggies1	2.496e-02
## HeartDiseaseorAttack1:HvyAlcoholConsump1	1.095e-01
## HeartDiseaseorAttack1:AnyHealthcare1	-1.014e-01
## HeartDiseaseorAttack1:NoDocbcCost1	-1.100e-01
## HeartDiseaseorAttack1:GenHealthCategoryGood Health	2.641e-01
## HeartDiseaseorAttack1:GenHealthCategoryPoor Health	-1.004e-01
## HeartDiseaseorAttack1:DiffWalk1	2.214e-01
## HeartDiseaseorAttack1:Sex1	-2.845e-02
## PhysActivity1:Fruits1	-5.421e-02
## PhysActivity1:Veggies1	5.454e-02
## PhysActivity1:HvyAlcoholConsump1	-2.150e-01
## PhysActivity1:AnyHealthcare1	8.106e-02
## PhysActivity1:NoDocbcCost1	-1.088e-02
## PhysActivity1:GenHealthCategoryGood Health	-1.139e-01
## PhysActivity1:GenHealthCategoryPoor Health	-5.242e-03
## PhysActivity1:DiffWalk1	4.818e-04
## PhysActivity1:Sex1	3.364e-02
## Fruits1:Veggies1	-1.420e-01
## Fruits1:HvyAlcoholConsump1	1.469e-03
## Fruits1:AnyHealthcare1	4.545e-02
## Fruits1:NoDocbcCost1	1.597e-02
## Fruits1:GenHealthCategoryGood Health	-1.005e-02
## Fruits1:GenHealthCategoryPoor Health	-2.143e-02
## Fruits1:DiffWalk1	-4.105e-02
## Fruits1:Sex1	-1.048e-01
## Veggies1:HvyAlcoholConsump1	7.575e-02
## Veggies1:AnyHealthcare1	-1.122e-01
## Veggies1:NoDocbcCost1	-1.342e-01
## Veggies1:GenHealthCategoryGood Health	-3.978e-02
## Veggies1:GenHealthCategoryPoor Health	6.358e-02
## Veggies1:DiffWalk1	4.374e-03
## Veggies1:Sex1	3.297e-02
## HvyAlcoholConsump1:AnyHealthcare1	-3.410e-01
## HvyAlcoholConsump1:NoDocbcCost1	-1.227e-02
## HvyAlcoholConsump1:GenHealthCategoryGood Health	1.642e-01
## HvyAlcoholConsump1:GenHealthCategoryPoor Health	1.304e-01
## HvyAlcoholConsump1:DiffWalk1	5.789e-02
## HvyAlcoholConsump1:Sex1	2.158e-01
## AnyHealthcare1:NoDocbcCost1	9.119e-02
## AnyHealthcare1:GenHealthCategoryGood Health	-2.171e-01

	Std. Error	z value
## AnyHealthcare1:GenHealthCategoryPoor Health	-5.682e-02	
## AnyHealthcare1:DiffWalk1	9.642e-02	
## AnyHealthcare1:Sex1	1.842e-01	
## NoDocbcCost1:GenHealthCategoryGood Health	2.836e-02	
## NoDocbcCost1:GenHealthCategoryPoor Health	-1.057e-01	
## NoDocbcCost1:DiffWalk1	-5.162e-02	
## NoDocbcCost1:Sex1	1.162e-01	
## GenHealthCategoryGood Health:DiffWalk1	3.019e-01	
## GenHealthCategoryPoor Health:DiffWalk1	-1.460e-01	
## GenHealthCategoryGood Health:Sex1	7.940e-02	
## GenHealthCategoryPoor Health:Sex1	-6.318e-02	
## DiffWalk1:Sex1	1.037e-01	
##		
## (Intercept)	7.505e-01	-8.039
## Age_groupOlder	2.664e-01	3.805
## Age_groupYoung	3.795e-01	-3.039
## BMI	1.929e-02	3.518
## MentHlthCategoryLow	3.393e-01	0.285
## MentHlthCategoryModerate	4.680e-01	-1.568
## PhysHlthCategoryLow	3.477e-01	0.645
## PhysHlthCategoryModerate	3.961e-01	0.221
## Education_levelLow Education	5.894e-01	1.447
## Education_levelMedium Education	2.269e-01	1.958
## HighBP1	2.419e-01	6.410
## HighChol1	2.318e-01	4.262
## CholCheck1	6.571e-01	1.286
## Smoker1	2.279e-01	-0.512
## Stroke1	4.334e-01	1.475
## HeartDiseaseorAttack1	2.979e-01	4.871
## PhysActivity1	2.305e-01	-1.125
## Fruits1	2.300e-01	0.212
## Veggies1	2.570e-01	-0.091
## HvyAlcoholConsump1	5.825e-01	-2.253
## AnyHealthcare1	3.736e-01	-2.656
## NoDocbcCost1	3.040e-01	0.765
## GenHealthCategoryGood Health	3.150e-01	-3.959
## GenHealthCategoryPoor Health	2.866e-01	3.842
## DiffWalk1	2.787e-01	0.916
## Sex1	2.259e-01	0.860
## Age_groupOlder:BMI	3.361e-03	3.113
## Age_groupYoung:BMI	5.945e-03	-1.329
## Age_groupOlder:MentHlthCategoryLow	6.279e-02	-1.836
## Age_groupYoung:MentHlthCategoryLow	1.041e-01	1.452
## Age_groupOlder:MentHlthCategoryModerate	8.236e-02	-1.364
## Age_groupYoung:MentHlthCategoryModerate	1.288e-01	0.917
## Age_groupOlder:PhysHlthCategoryLow	5.518e-02	0.525
## Age_groupYoung:PhysHlthCategoryLow	1.144e-01	-0.382
## Age_groupOlder:PhysHlthCategoryModerate	6.507e-02	1.252
## Age_groupYoung:PhysHlthCategoryModerate	1.296e-01	-0.274
## Age_groupOlder:Education_levelLow Education	1.052e-01	0.999
## Age_groupYoung:Education_levelLow Education	2.168e-01	0.124
## Age_groupOlder:Education_levelMedium Education	3.575e-02	0.299
## Age_groupYoung:Education_levelMedium Education	6.895e-02	1.143
## Age_groupOlder:HighBP1	3.741e-02	-5.934

## Age_groupYoung:HighBP1	6.730e-02	-1.331
## Age_groupOlder:HighChol1	3.511e-02	-7.802
## Age_groupYoung:HighChol1	6.510e-02	5.509
## Age_groupOlder:CholCheck1	2.075e-01	-1.810
## Age_groupYoung:CholCheck1	2.796e-01	0.614
## Age_groupOlder:Smoker1	3.423e-02	-0.031
## Age_groupYoung:Smoker1	6.534e-02	-0.133
## Age_groupOlder:Stroke1	6.517e-02	-2.900
## Age_groupYoung:Stroke1	1.842e-01	-0.982
## Age_groupOlder:HeartDiseaseorAttack1	4.615e-02	-3.389
## Age_groupYoung:HeartDiseaseorAttack1	1.404e-01	0.398
## Age_groupOlder:PhysActivity1	3.737e-02	-2.026
## Age_groupYoung:PhysActivity1	7.307e-02	0.265
## Age_groupOlder:Fruits1	3.504e-02	-1.388
## Age_groupYoung:Fruits1	6.560e-02	0.353
## Age_groupOlder:Veggies1	4.084e-02	1.324
## Age_groupYoung:Veggies1	7.798e-02	0.728
## Age_groupOlder:HvyAlcoholConsump1	1.019e-01	-0.419
## Age_groupYoung:HvyAlcoholConsump1	1.486e-01	4.597
## Age_groupOlder:AnyHealthcare1	1.120e-01	-1.837
## Age_groupYoung:AnyHealthcare1	1.182e-01	1.246
## Age_groupOlder:NoDocbcCost1	6.409e-02	1.484
## Age_groupYoung:NoDocbcCost1	8.756e-02	1.451
## Age_groupOlder:GenHealthCategoryGood Health	4.164e-02	4.781
## Age_groupYoung:GenHealthCategoryGood Health	7.931e-02	1.703
## Age_groupOlder:GenHealthCategoryPoor Health	4.611e-02	-4.086
## Age_groupYoung:GenHealthCategoryPoor Health	8.445e-02	2.119
## Age_groupOlder:DiffWalk1	4.390e-02	-1.680
## Age_groupYoung:DiffWalk1	9.877e-02	1.850
## Age_groupOlder:Sex1	3.431e-02	5.963
## Age_groupYoung:Sex1	6.466e-02	-4.698
## BMI:MentHlthCategoryLow	5.414e-03	-0.004
## BMI:MentHlthCategoryModerate	7.055e-03	0.763
## BMI:PhysHlthCategoryLow	4.850e-03	-0.832
## BMI:PhysHlthCategoryModerate	5.737e-03	-0.006
## BMI:Education_levelLow Education	9.288e-03	-2.126
## BMI:Education_levelMedium Education	3.274e-03	-3.400
## BMI:HighBP1	3.462e-03	-1.263
## BMI:HighChol1	3.221e-03	-2.013
## BMI:CholCheck1	1.639e-02	0.286
## BMI:Smoker1	3.161e-03	3.072
## BMI:Stroke1	5.750e-03	1.043
## BMI:HeartDiseaseorAttack1	4.182e-03	0.293
## BMI:PhysActivity1	3.368e-03	1.088
## BMI:Fruits1	3.260e-03	-0.840
## BMI:Veggies1	3.740e-03	1.680
## BMI:HvyAlcoholConsump1	9.294e-03	1.137
## BMI:AnyHealthcare1	7.768e-03	2.129
## BMI:NoDocbcCost1	5.312e-03	-0.005
## BMI:GenHealthCategoryGood Health	4.026e-03	3.478
## BMI:GenHealthCategoryPoor Health	4.100e-03	-0.379
## BMI:DiffWalk1	3.832e-03	-2.615
## BMI:Sex1	3.178e-03	-1.990
## MentHlthCategoryLow:PhysHlthCategoryLow	7.446e-02	1.777

## MentHlthCategoryModerate:PhysHlthCategoryLow	1.009e-01	0.749
## MentHlthCategoryLow:PhysHlthCategoryModerate	8.287e-02	0.111
## MentHlthCategoryModerate:PhysHlthCategoryModerate	1.039e-01	0.366
## MentHlthCategoryLow:Education_levelLow Education	1.399e-01	-0.468
## MentHlthCategoryModerate:Education_levelLow Education	1.884e-01	0.438
## MentHlthCategoryLow:Education_levelMedium Education	5.910e-02	0.158
## MentHlthCategoryModerate:Education_levelMedium Education	7.708e-02	1.445
## MentHlthCategoryLow:HighBP1	6.489e-02	-0.179
## MentHlthCategoryModerate:HighBP1	8.526e-02	0.292
## MentHlthCategoryLow:HighChol1	6.209e-02	-2.451
## MentHlthCategoryModerate:HighChol1	8.096e-02	-0.662
## MentHlthCategoryLow:CholCheck1	2.609e-01	0.543
## MentHlthCategoryModerate:CholCheck1	3.705e-01	1.011
## MentHlthCategoryLow:Smoker1	5.841e-02	-0.461
## MentHlthCategoryModerate:Smoker1	7.621e-02	-0.499
## MentHlthCategoryLow:Stroke1	9.015e-02	-0.571
## MentHlthCategoryModerate:Stroke1	1.184e-01	1.077
## MentHlthCategoryLow:HeartDiseaseorAttack1	6.998e-02	-0.741
## MentHlthCategoryModerate:HeartDiseaseorAttack1	9.292e-02	-1.130
## MentHlthCategoryLow:PhysActivity1	5.897e-02	-0.508
## MentHlthCategoryModerate:PhysActivity1	7.736e-02	-0.809
## MentHlthCategoryLow:Fruits1	5.913e-02	-2.896
## MentHlthCategoryModerate:Fruits1	7.703e-02	-2.276
## MentHlthCategoryLow:Veggies1	6.608e-02	-2.356
## MentHlthCategoryModerate:Veggies1	8.629e-02	-0.043
## MentHlthCategoryLow:HvyAlcoholConsump1	1.570e-01	-0.599
## MentHlthCategoryModerate:HvyAlcoholConsump1	1.975e-01	0.023
## MentHlthCategoryLow:AnyHealthcare1	1.231e-01	1.121
## MentHlthCategoryModerate:AnyHealthcare1	1.575e-01	1.456
## MentHlthCategoryLow:NoDocbcCost1	7.486e-02	0.255
## MentHlthCategoryModerate:NoDocbcCost1	9.590e-02	0.221
## MentHlthCategoryLow:GenHealthCategoryGood Health	1.056e-01	-0.647
## MentHlthCategoryModerate:GenHealthCategoryGood Health	1.359e-01	-0.547
## MentHlthCategoryLow:GenHealthCategoryPoor Health	7.263e-02	0.319
## MentHlthCategoryModerate:GenHealthCategoryPoor Health	9.534e-02	0.348
## MentHlthCategoryLow:DiffWalk1	6.468e-02	-0.045
## MentHlthCategoryModerate:DiffWalk1	8.481e-02	0.567
## MentHlthCategoryLow:Sex1	5.881e-02	0.354
## MentHlthCategoryModerate:Sex1	7.689e-02	1.612
## PhysHlthCategoryLow:Education_levelLow Education	1.257e-01	-1.586
## PhysHlthCategoryModerate:Education_levelLow Education	1.521e-01	-0.626
## PhysHlthCategoryLow:Education_levelMedium Education	5.207e-02	-0.181
## PhysHlthCategoryModerate:Education_levelMedium Education	6.170e-02	0.590
## PhysHlthCategoryLow:HighBP1	5.885e-02	-1.494
## PhysHlthCategoryModerate:HighBP1	6.968e-02	-1.162
## PhysHlthCategoryLow:HighChol1	5.351e-02	0.162
## PhysHlthCategoryModerate:HighChol1	6.412e-02	1.123
## PhysHlthCategoryLow:CholCheck1	2.804e-01	-0.071
## PhysHlthCategoryModerate:CholCheck1	3.126e-01	-0.081
## PhysHlthCategoryLow:Smoker1	5.133e-02	-1.746
## PhysHlthCategoryModerate:Smoker1	6.094e-02	-1.481
## PhysHlthCategoryLow:Stroke1	8.094e-02	-0.494
## PhysHlthCategoryModerate:Stroke1	9.238e-02	-0.659
## PhysHlthCategoryLow:HeartDiseaseorAttack1	6.027e-02	-0.620

## PhysHlthCategoryModerate:HeartDiseaseorAttack1	7.077e-02	-0.497
## PhysHlthCategoryLow:PhysActivity1	5.170e-02	0.536
## PhysHlthCategoryModerate:PhysActivity1	6.146e-02	-0.004
## PhysHlthCategoryLow:Fruits1	5.268e-02	1.068
## PhysHlthCategoryModerate:Fruits1	6.244e-02	0.722
## PhysHlthCategoryLow:Veggies1	5.903e-02	1.074
## PhysHlthCategoryModerate:Veggies1	7.036e-02	1.033
## PhysHlthCategoryLow:HvyAlcoholConsump1	1.582e-01	1.436
## PhysHlthCategoryModerate:HvyAlcoholConsump1	1.981e-01	0.290
## PhysHlthCategoryLow:AnyHealthcare1	1.272e-01	-1.207
## PhysHlthCategoryModerate:AnyHealthcare1	1.456e-01	-1.358
## PhysHlthCategoryLow:NoDocbcCost1	7.791e-02	-1.249
## PhysHlthCategoryModerate:NoDocbcCost1	8.555e-02	0.194
## PhysHlthCategoryLow:GenHealthCategoryGood Health	9.616e-02	-3.075
## PhysHlthCategoryModerate:GenHealthCategoryGood Health	1.186e-01	-1.683
## PhysHlthCategoryLow:GenHealthCategoryPoor Health	6.013e-02	0.903
## PhysHlthCategoryModerate:GenHealthCategoryPoor Health	7.305e-02	-0.975
## PhysHlthCategoryLow:DiffWalk1	5.315e-02	0.887
## PhysHlthCategoryModerate:DiffWalk1	6.364e-02	2.829
## PhysHlthCategoryLow:Sex1	5.180e-02	-0.658
## PhysHlthCategoryModerate:Sex1	6.190e-02	0.747
## Education_levelLow Education:HighBP1	1.060e-01	-1.130
## Education_levelMedium Education:HighBP1	3.693e-02	-2.746
## Education_levelLow Education:HighChol1	9.931e-02	1.611
## Education_levelMedium Education:HighChol1	3.418e-02	-0.753
## Education_levelLow Education:CholCheck1	4.657e-01	0.074
## Education_levelMedium Education:CholCheck1	1.768e-01	-0.907
## Education_levelLow Education:Smoker1	9.778e-02	-0.491
## Education_levelMedium Education:Smoker1	3.318e-02	-2.289
## Education_levelLow Education:Stroke1	1.539e-01	-2.219
## Education_levelMedium Education:Stroke1	6.032e-02	-2.065
## Education_levelLow Education:HeartDiseaseorAttack1	1.120e-01	1.550
## Education_levelMedium Education:HeartDiseaseorAttack1	4.311e-02	-1.351
## Education_levelLow Education:PhysActivity1	9.483e-02	0.259
## Education_levelMedium Education:PhysActivity1	3.511e-02	1.513
## Education_levelLow Education:Fruits1	9.634e-02	1.318
## Education_levelMedium Education:Fruits1	3.398e-02	0.951
## Education_levelLow Education:Veggies1	1.027e-01	0.926
## Education_levelMedium Education:Veggies1	3.841e-02	1.345
## Education_levelLow Education:HvyAlcoholConsump1	4.085e-01	-0.963
## Education_levelMedium Education:HvyAlcoholConsump1	9.743e-02	-0.037
## Education_levelLow Education:AnyHealthcare1	1.650e-01	0.454
## Education_levelMedium Education:AnyHealthcare1	8.399e-02	1.537
## Education_levelLow Education:NoDocbcCost1	1.309e-01	0.086
## Education_levelMedium Education:NoDocbcCost1	5.686e-02	1.062
## Education_levelLow Education:GenHealthCategoryGood Health	1.706e-01	0.686
## Education_levelMedium Education:GenHealthCategoryGood Health	4.152e-02	3.404
## Education_levelLow Education:GenHealthCategoryPoor Health	1.137e-01	0.150
## Education_levelMedium Education:GenHealthCategoryPoor Health	4.283e-02	0.053
## Education_levelLow Education:DiffWalk1	1.053e-01	-0.506
## Education_levelMedium Education:DiffWalk1	4.058e-02	1.287
## Education_levelLow Education:Sex1	9.824e-02	-2.798
## Education_levelMedium Education:Sex1	3.361e-02	-1.296
## HighBP1:HighChol1	3.453e-02	1.519

## HighBP1:CholCheck1	1.849e-01	-2.212
## HighBP1:Smoker1	3.501e-02	-1.545
## HighBP1:Stroke1	7.366e-02	-1.583
## HighBP1:HeartDiseaseorAttack1	5.089e-02	-3.324
## HighBP1:PhysActivity1	3.896e-02	0.649
## HighBP1:Fruits1	3.628e-02	1.435
## HighBP1:Veggies1	4.262e-02	-1.875
## HighBP1:HvyAlcoholConsump1	1.059e-01	-1.626
## HighBP1:AnyHealthcare1	8.855e-02	-0.356
## HighBP1:NoDocbcCost1	6.243e-02	0.895
## HighBP1:GenHealthCategoryGood Health	4.170e-02	5.793
## HighBP1:GenHealthCategoryPoor Health	4.787e-02	-0.757
## HighBP1:DiffWalk1	4.570e-02	-0.590
## HighBP1:Sex1	3.528e-02	-3.191
## HighChol1:CholCheck1	1.799e-01	0.306
## HighChol1:Smoker1	3.256e-02	1.122
## HighChol1:Stroke1	6.402e-02	-0.845
## HighChol1:HeartDiseaseorAttack1	4.573e-02	-5.638
## HighChol1:PhysActivity1	3.595e-02	-1.130
## HighChol1:Fruits1	3.388e-02	-0.161
## HighChol1:Veggies1	3.936e-02	-0.667
## HighChol1:HvyAlcoholConsump1	9.762e-02	0.839
## HighChol1:AnyHealthcare1	8.454e-02	0.206
## HighChol1:NoDocbcCost1	5.894e-02	1.384
## HighChol1:GenHealthCategoryGood Health	3.924e-02	-0.133
## HighChol1:GenHealthCategoryPoor Health	4.377e-02	-0.441
## HighChol1:DiffWalk1	4.151e-02	0.757
## HighChol1:Sex1	3.286e-02	-4.995
## CholCheck1:Smoker1	1.793e-01	-0.064
## CholCheck1:Stroke1	3.572e-01	0.010
## CholCheck1:HeartDiseaseorAttack1	2.349e-01	-2.771
## CholCheck1:PhysActivity1	1.786e-01	0.452
## CholCheck1:Fruits1	1.797e-01	1.235
## CholCheck1:Veggies1	2.016e-01	-0.093
## CholCheck1:HvyAlcoholConsump1	4.425e-01	0.525
## CholCheck1:AnyHealthcare1	2.295e-01	2.412
## CholCheck1:NoDocbcCost1	2.199e-01	-0.846
## CholCheck1:GenHealthCategoryGood Health	2.403e-01	0.988
## CholCheck1:GenHealthCategoryPoor Health	2.247e-01	-1.157
## CholCheck1:DiffWalk1	2.222e-01	0.240
## CholCheck1:Sex1	1.775e-01	0.442
## Smoker1:Stroke1	5.979e-02	-1.123
## Smoker1:HeartDiseaseorAttack1	4.218e-02	-0.120
## Smoker1:PhysActivity1	3.474e-02	0.809
## Smoker1:Fruits1	3.279e-02	2.195
## Smoker1:Veggies1	3.818e-02	0.362
## Smoker1:HvyAlcoholConsump1	1.012e-01	0.722
## Smoker1:AnyHealthcare1	8.190e-02	-1.133
## Smoker1:NoDocbcCost1	5.589e-02	-2.536
## Smoker1:GenHealthCategoryGood Health	3.849e-02	2.115
## Smoker1:GenHealthCategoryPoor Health	4.217e-02	-1.668
## Smoker1:DiffWalk1	3.996e-02	-1.974
## Smoker1:Sex1	3.185e-02	1.005
## Stroke1:HeartDiseaseorAttack1	6.036e-02	2.313

## Stroke1:PhysActivity1	6.103e-02	-2.454
## Stroke1:Fruits1	6.128e-02	0.729
## Stroke1:Veggies1	6.674e-02	-0.892
## Stroke1:HvyAlcoholConsump1	1.984e-01	0.205
## Stroke1:AnyHealthcare1	1.563e-01	-0.589
## Stroke1:NoDocbcCost1	9.286e-02	-1.647
## Stroke1:GenHealthCategoryGood Health	8.999e-02	2.003
## Stroke1:GenHealthCategoryPoor Health	7.393e-02	-1.465
## Stroke1:DiffWalk1	6.523e-02	1.204
## Stroke1:Sex1	5.987e-02	-0.357
## HeartDiseaseorAttack1:PhysActivity1	4.448e-02	-1.581
## HeartDiseaseorAttack1:Fruits1	4.327e-02	0.405
## HeartDiseaseorAttack1:Veggies1	4.920e-02	0.507
## HeartDiseaseorAttack1:HvyAlcoholConsump1	1.305e-01	0.839
## HeartDiseaseorAttack1:AnyHealthcare1	1.138e-01	-0.891
## HeartDiseaseorAttack1:NoDocbcCost1	7.116e-02	-1.546
## HeartDiseaseorAttack1:GenHealthCategoryGood Health	5.865e-02	4.504
## HeartDiseaseorAttack1:GenHealthCategoryPoor Health	5.222e-02	-1.923
## HeartDiseaseorAttack1:DiffWalk1	4.798e-02	4.615
## HeartDiseaseorAttack1:Sex1	4.306e-02	-0.661
## PhysActivity1:Fruits1	3.517e-02	-1.541
## PhysActivity1:Veggies1	3.915e-02	1.393
## PhysActivity1:HvyAlcoholConsump1	9.961e-02	-2.158
## PhysActivity1:AnyHealthcare1	8.420e-02	0.963
## PhysActivity1:NoDocbcCost1	5.800e-02	-0.188
## PhysActivity1:GenHealthCategoryGood Health	4.537e-02	-2.510
## PhysActivity1:GenHealthCategoryPoor Health	4.367e-02	-0.120
## PhysActivity1:DiffWalk1	4.027e-02	0.012
## PhysActivity1:Sex1	3.500e-02	0.961
## Fruits1:Veggies1	3.743e-02	-3.794
## Fruits1:HvyAlcoholConsump1	9.268e-02	0.016
## Fruits1:AnyHealthcare1	8.374e-02	0.543
## Fruits1:NoDocbcCost1	5.692e-02	0.281
## Fruits1:GenHealthCategoryGood Health	4.035e-02	-0.249
## Fruits1:GenHealthCategoryPoor Health	4.344e-02	-0.493
## Fruits1:DiffWalk1	4.121e-02	-0.996
## Fruits1:Sex1	3.297e-02	-3.180
## Veggies1:HvyAlcoholConsump1	1.174e-01	0.645
## Veggies1:AnyHealthcare1	9.324e-02	-1.204
## Veggies1:NoDocbcCost1	6.359e-02	-2.111
## Veggies1:GenHealthCategoryGood Health	4.831e-02	-0.823
## Veggies1:GenHealthCategoryPoor Health	4.897e-02	1.298
## Veggies1:DiffWalk1	4.633e-02	0.094
## Veggies1:Sex1	3.843e-02	0.858
## HvyAlcoholConsump1:AnyHealthcare1	1.948e-01	-1.750
## HvyAlcoholConsump1:NoDocbcCost1	1.598e-01	-0.077
## HvyAlcoholConsump1:GenHealthCategoryGood Health	1.104e-01	1.488
## HvyAlcoholConsump1:GenHealthCategoryPoor Health	1.231e-01	1.059
## HvyAlcoholConsump1:DiffWalk1	1.206e-01	0.480
## HvyAlcoholConsump1:Sex1	9.563e-02	2.257
## AnyHealthcare1:NoDocbcCost1	9.070e-02	1.005
## AnyHealthcare1:GenHealthCategoryGood Health	1.112e-01	-1.952
## AnyHealthcare1:GenHealthCategoryPoor Health	1.005e-01	-0.565
## AnyHealthcare1:DiffWalk1	1.024e-01	0.941

## AnyHealthcare1:Sex1	8.213e-02	2.242
## NoDocbcCost1:GenHealthCategoryGood Health	8.592e-02	0.330
## NoDocbcCost1:GenHealthCategoryPoor Health	6.858e-02	-1.541
## NoDocbcCost1:DiffWalk1	6.496e-02	-0.795
## NoDocbcCost1:Sex1	5.621e-02	2.067
## GenHealthCategoryGood Health:DiffWalk1	5.752e-02	5.249
## GenHealthCategoryPoor Health:DiffWalk1	4.627e-02	-3.155
## GenHealthCategoryGood Health:Sex1	3.885e-02	2.044
## GenHealthCategoryPoor Health:Sex1	4.281e-02	-1.476
## DiffWalk1:Sex1	4.056e-02	2.558
##	Pr(> z)	
## (Intercept)	9.09e-16	***
## Age_groupOlder	0.000142	***
## Age_groupYoung	0.002371	**
## BMI	0.000434	***
## MentHlthCategoryLow	0.775338	
## MentHlthCategoryModerate	0.116905	
## PhysHlthCategoryLow	0.518993	
## PhysHlthCategoryModerate	0.825366	
## Education_levelLow Education	0.147911	
## Education_levelMedium Education	0.050183	.
## HighBP1	1.46e-10	***
## HighChol1	2.02e-05	***
## CholCheck1	0.198358	
## Smoker1	0.608397	
## Stroke1	0.140203	
## HeartDiseaseorAttack1	1.11e-06	***
## PhysActivity1	0.260618	
## Fruits1	0.832100	
## Veggies1	0.927772	
## HvyAlcoholConsump1	0.024231	*
## AnyHealthcare1	0.007914	**
## NoDocbcCost1	0.444461	
## GenHealthCategoryGood Health	7.53e-05	***
## GenHealthCategoryPoor Health	0.000122	***
## DiffWalk1	0.359586	
## Sex1	0.389683	
## Age_groupOlder:BMI	0.001851	**
## Age_groupYoung:BMI	0.183987	
## Age_groupOlder:MentHlthCategoryLow	0.066347	.
## Age_groupYoung:MentHlthCategoryLow	0.146565	
## Age_groupOlder:MentHlthCategoryModerate	0.172618	
## Age_groupYoung:MentHlthCategoryModerate	0.359100	
## Age_groupOlder:PhysHlthCategoryLow	0.599806	
## Age_groupYoung:PhysHlthCategoryLow	0.702350	
## Age_groupOlder:PhysHlthCategoryModerate	0.210616	
## Age_groupYoung:PhysHlthCategoryModerate	0.783857	
## Age_groupOlder:Education_levelLow Education	0.317977	
## Age_groupYoung:Education_levelLow Education	0.901559	
## Age_groupOlder:Education_levelMedium Education	0.764743	
## Age_groupYoung:Education_levelMedium Education	0.253164	
## Age_groupOlder:HighBP1	2.95e-09	***
## Age_groupYoung:HighBP1	0.183309	
## Age_groupOlder:HighChol1	6.10e-15	***

## Age_groupYoung:HighChol1	3.61e-08 ***
## Age_groupOlder:CholCheck1	0.070253 .
## Age_groupYoung:CholCheck1	0.539233
## Age_groupOlder:Smoker1	0.975599
## Age_groupYoung:Smoker1	0.894557
## Age_groupOlder:Stroke1	0.003736 **
## Age_groupYoung:Stroke1	0.326218
## Age_groupOlder:HeartDiseaseorAttack1	0.000702 ***
## Age_groupYoung:HeartDiseaseorAttack1	0.690274
## Age_groupOlder:PhysActivity1	0.042772 *
## Age_groupYoung:PhysActivity1	0.791309
## Age_groupOlder:Fruits1	0.165286
## Age_groupYoung:Fruits1	0.724409
## Age_groupOlder:Veggies1	0.185568
## Age_groupYoung:Veggies1	0.466593
## Age_groupOlder:HvyAlcoholConsump1	0.674942
## Age_groupYoung:HvyAlcoholConsump1	4.28e-06 ***
## Age_groupOlder:AnyHealthcare1	0.066169 .
## Age_groupYoung:AnyHealthcare1	0.212722
## Age_groupOlder:NoDocbcCost1	0.137710
## Age_groupYoung:NoDocbcCost1	0.146721
## Age_groupOlder:GenHealthCategoryGood Health	1.74e-06 ***
## Age_groupYoung:GenHealthCategoryGood Health	0.088504 .
## Age_groupOlder:GenHealthCategoryPoor Health	4.39e-05 ***
## Age_groupYoung:GenHealthCategoryPoor Health	0.034126 *
## Age_groupOlder:DiffWalk1	0.092884 .
## Age_groupYoung:DiffWalk1	0.064305 .
## Age_groupOlder:Sex1	2.48e-09 ***
## Age_groupYoung:Sex1	2.63e-06 ***
## BMI:MentHlthCategoryLow	0.996799
## BMI:MentHlthCategoryModerate	0.445391
## BMI:PhysHlthCategoryLow	0.405493
## BMI:PhysHlthCategoryModerate	0.995589
## BMI:Education_levelLow Education	0.033528 *
## BMI:Education_levelMedium Education	0.000675 ***
## BMI:HighBP1	0.206532
## BMI:HighChol1	0.044094 *
## BMI:CholCheck1	0.774595
## BMI:Smoker1	0.002128 **
## BMI:Stroke1	0.297072
## BMI:HeartDiseaseorAttack1	0.769715
## BMI:PhysActivity1	0.276468
## BMI:Fruits1	0.401052
## BMI:Veggies1	0.092952 .
## BMI:HvyAlcoholConsump1	0.255663
## BMI:AnyHealthcare1	0.033249 *
## BMI:NoDocbcCost1	0.995685
## BMI:GenHealthCategoryGood Health	0.000506 ***
## BMI:GenHealthCategoryPoor Health	0.704369
## BMI:DiffWalk1	0.008918 **
## BMI:Sex1	0.046591 *
## MentHlthCategoryLow:PhysHlthCategoryLow	0.075544 .
## MentHlthCategoryModerate:PhysHlthCategoryLow	0.453628
## MentHlthCategoryLow:PhysHlthCategoryModerate	0.911656

## MentHlthCategoryModerate:PhysHlthCategoryModerate	0.714732
## MentHlthCategoryLow:Education_levelLow Education	0.639538
## MentHlthCategoryModerate:Education_levelLow Education	0.661109
## MentHlthCategoryLow:Education_levelMedium Education	0.874449
## MentHlthCategoryModerate:Education_levelMedium Education	0.148405
## MentHlthCategoryLow:HighBP1	0.858121
## MentHlthCategoryModerate:HighBP1	0.770508
## MentHlthCategoryLow:HighChol1	0.014227 *
## MentHlthCategoryModerate:HighChol1	0.508267
## MentHlthCategoryLow:CholCheck1	0.587035
## MentHlthCategoryModerate:CholCheck1	0.311904
## MentHlthCategoryLow:Smoker1	0.644656
## MentHlthCategoryModerate:Smoker1	0.617534
## MentHlthCategoryLow:Stroke1	0.568016
## MentHlthCategoryModerate:Stroke1	0.281450
## MentHlthCategoryLow:HeartDiseaseorAttack1	0.458427
## MentHlthCategoryModerate:HeartDiseaseorAttack1	0.258520
## MentHlthCategoryLow:PhysActivity1	0.611620
## MentHlthCategoryModerate:PhysActivity1	0.418762
## MentHlthCategoryLow:Fruits1	0.003781 **
## MentHlthCategoryModerate:Fruits1	0.022827 *
## MentHlthCategoryLow:Veggies1	0.018494 *
## MentHlthCategoryModerate:Veggies1	0.965997
## MentHlthCategoryLow:HvyAlcoholConsump1	0.549201
## MentHlthCategoryModerate:HvyAlcoholConsump1	0.982011
## MentHlthCategoryLow:AnyHealthcare1	0.262349
## MentHlthCategoryModerate:AnyHealthcare1	0.145430
## MentHlthCategoryLow>NoDocbcCost1	0.798993
## MentHlthCategoryModerate>NoDocbcCost1	0.825127
## MentHlthCategoryLow:GenHealthCategoryGood Health	0.517916
## MentHlthCategoryModerate:GenHealthCategoryGood Health	0.584527
## MentHlthCategoryLow:GenHealthCategoryPoor Health	0.749605
## MentHlthCategoryModerate:GenHealthCategoryPoor Health	0.727505
## MentHlthCategoryLow:DiffWalk1	0.964245
## MentHlthCategoryModerate:DiffWalk1	0.570920
## MentHlthCategoryLow:Sex1	0.723559
## MentHlthCategoryModerate:Sex1	0.107022
## PhysHlthCategoryLow:Education_levelLow Education	0.112844
## PhysHlthCategoryModerate:Education_levelLow Education	0.531036
## PhysHlthCategoryLow:Education_levelMedium Education	0.856162
## PhysHlthCategoryModerate:Education_levelMedium Education	0.554962
## PhysHlthCategoryLow:HighBP1	0.135095
## PhysHlthCategoryModerate:HighBP1	0.245045
## PhysHlthCategoryLow:HighChol1	0.871220
## PhysHlthCategoryModerate:HighChol1	0.261633
## PhysHlthCategoryLow:CholCheck1	0.943407
## PhysHlthCategoryModerate:CholCheck1	0.935742
## PhysHlthCategoryLow:Smoker1	0.080859 .
## PhysHlthCategoryModerate:Smoker1	0.138543
## PhysHlthCategoryLow:Stroke1	0.621012
## PhysHlthCategoryModerate:Stroke1	0.509758
## PhysHlthCategoryLow:HeartDiseaseorAttack1	0.535486
## PhysHlthCategoryModerate:HeartDiseaseorAttack1	0.619500
## PhysHlthCategoryLow:PhysActivity1	0.591798

## PhysHlthCategoryModerate:PhysActivity1	0.996661
## PhysHlthCategoryLow:Fruits1	0.285706
## PhysHlthCategoryModerate:Fruits1	0.470251
## PhysHlthCategoryLow:Veggies1	0.282900
## PhysHlthCategoryModerate:Veggies1	0.301711
## PhysHlthCategoryLow:HvyAlcoholConsump1	0.150933
## PhysHlthCategoryModerate:HvyAlcoholConsump1	0.771805
## PhysHlthCategoryLow:AnyHealthcare1	0.227429
## PhysHlthCategoryModerate:AnyHealthcare1	0.174399
## PhysHlthCategoryLow:NoDocbcCost1	0.211726
## PhysHlthCategoryModerate:NoDocbcCost1	0.845878
## PhysHlthCategoryLow:GenHealthCategoryGood Health	0.002108 **
## PhysHlthCategoryModerate:GenHealthCategoryGood Health	0.092377 .
## PhysHlthCategoryLow:GenHealthCategoryPoor Health	0.366401
## PhysHlthCategoryModerate:GenHealthCategoryPoor Health	0.329700
## PhysHlthCategoryLow:DiffWalk1	0.374821
## PhysHlthCategoryModerate:DiffWalk1	0.004669 **
## PhysHlthCategoryLow:Sex1	0.510255
## PhysHlthCategoryModerate:Sex1	0.455235
## Education_levelLow Education:HighBP1	0.258538
## Education_levelMedium Education:HighBP1	0.006034 **
## Education_levelLow Education:HighChol1	0.107157
## Education_levelMedium Education:HighChol1	0.451243
## Education_levelLow Education:CholCheck1	0.941101
## Education_levelMedium Education:CholCheck1	0.364475
## Education_levelLow Education:Smoker1	0.623663
## Education_levelMedium Education:Smoker1	0.022105 *
## Education_levelLow Education:Stroke1	0.026468 *
## Education_levelMedium Education:Stroke1	0.038914 *
## Education_levelLow Education:HeartDiseaseorAttack1	0.121124
## Education_levelMedium Education:HeartDiseaseorAttack1	0.176608
## Education_levelLow Education:PhysActivity1	0.795535
## Education_levelMedium Education:PhysActivity1	0.130247
## Education_levelLow Education:Fruits1	0.187401
## Education_levelMedium Education:Fruits1	0.341688
## Education_levelLow Education:Veggies1	0.354313
## Education_levelMedium Education:Veggies1	0.178649
## Education_levelLow Education:HvyAlcoholConsump1	0.335785
## Education_levelMedium Education:HvyAlcoholConsump1	0.970200
## Education_levelLow Education:AnyHealthcare1	0.649681
## Education_levelMedium Education:AnyHealthcare1	0.124370
## Education_levelLow Education:NoDocbcCost1	0.931813
## Education_levelMedium Education:NoDocbcCost1	0.288347
## Education_levelLow Education:GenHealthCategoryGood Health	0.492853
## Education_levelMedium Education:GenHealthCategoryGood Health	0.000665 ***
## Education_levelLow Education:GenHealthCategoryPoor Health	0.880763
## Education_levelMedium Education:GenHealthCategoryPoor Health	0.957917
## Education_levelLow Education:DiffWalk1	0.613134
## Education_levelMedium Education:DiffWalk1	0.198049
## Education_levelLow Education:Sex1	0.005145 **
## Education_levelMedium Education:Sex1	0.194946
## HighBP1:HighChol1	0.128854
## HighBP1:CholCheck1	0.026990 *
## HighBP1:Smoker1	0.122270

## HighBP1:Stroke1	0.113356
## HighBP1:HeartDiseaseorAttack1	0.000889 ***
## HighBP1:PhysActivity1	0.516266
## HighBP1:Fruits1	0.151390
## HighBP1:Veggies1	0.060813 .
## HighBP1:HvyAlcoholConsump1	0.103847
## HighBP1:AnyHealthcare1	0.721802
## HighBP1:NoDocbcCost1	0.370744
## HighBP1:GenHealthCategoryGood Health	6.92e-09 ***
## HighBP1:GenHealthCategoryPoor Health	0.449165
## HighBP1:DiffWalk1	0.555410
## HighBP1:Sex1	0.001418 **
## HighChol1:CholCheck1	0.759981
## HighChol1:Smoker1	0.261814
## HighChol1:Stroke1	0.398225
## HighChol1:HeartDiseaseorAttack1	1.73e-08 ***
## HighChol1:PhysActivity1	0.258393
## HighChol1:Fruits1	0.872155
## HighChol1:Veggies1	0.504703
## HighChol1:HvyAlcoholConsump1	0.401297
## HighChol1:AnyHealthcare1	0.836713
## HighChol1:NoDocbcCost1	0.166292
## HighChol1:GenHealthCategoryGood Health	0.893946
## HighChol1:GenHealthCategoryPoor Health	0.658997
## HighChol1:DiffWalk1	0.448769
## HighChol1:Sex1	5.90e-07 ***
## CholCheck1:Smoker1	0.948614
## CholCheck1:Stroke1	0.992096
## CholCheck1:HeartDiseaseorAttack1	0.005584 **
## CholCheck1:PhysActivity1	0.651045
## CholCheck1:Fruits1	0.216859
## CholCheck1:Veggies1	0.926169
## CholCheck1:HvyAlcoholConsump1	0.599357
## CholCheck1:AnyHealthcare1	0.015879 *
## CholCheck1:NoDocbcCost1	0.397529
## CholCheck1:GenHealthCategoryGood Health	0.323039
## CholCheck1:GenHealthCategoryPoor Health	0.247372
## CholCheck1:DiffWalk1	0.810482
## CholCheck1:Sex1	0.658798
## Smoker1:Stroke1	0.261632
## Smoker1:HeartDiseaseorAttack1	0.904872
## Smoker1:PhysActivity1	0.418242
## Smoker1:Fruits1	0.028185 *
## Smoker1:Veggies1	0.717046
## Smoker1:HvyAlcoholConsump1	0.470351
## Smoker1:AnyHealthcare1	0.257227
## Smoker1:NoDocbcCost1	0.011216 *
## Smoker1:GenHealthCategoryGood Health	0.034436 *
## Smoker1:GenHealthCategoryPoor Health	0.095355 .
## Smoker1:DiffWalk1	0.048375 *
## Smoker1:Sex1	0.314820
## Stroke1:HeartDiseaseorAttack1	0.020698 *
## Stroke1:PhysActivity1	0.014109 *
## Stroke1:Fruits1	0.466190

## Stroke1:Veggies1	0.372256
## Stroke1:HvyAlcoholConsump1	0.837294
## Stroke1:AnyHealthcare1	0.556107
## Stroke1:NoDocbcCost1	0.099563 .
## Stroke1:GenHealthCategoryGood Health	0.045156 *
## Stroke1:GenHealthCategoryPoor Health	0.142938
## Stroke1:DiffWalk1	0.228600
## Stroke1:Sex1	0.721252
## HeartDiseaseorAttack1:PhysActivity1	0.113980
## HeartDiseaseorAttack1:Fruits1	0.685667
## HeartDiseaseorAttack1:Veggies1	0.611948
## HeartDiseaseorAttack1:HvyAlcoholConsump1	0.401712
## HeartDiseaseorAttack1:AnyHealthcare1	0.372870
## HeartDiseaseorAttack1:NoDocbcCost1	0.122062
## HeartDiseaseorAttack1:GenHealthCategoryGood Health	6.68e-06 ***
## HeartDiseaseorAttack1:GenHealthCategoryPoor Health	0.054429 .
## HeartDiseaseorAttack1:DiffWalk1	3.94e-06 ***
## HeartDiseaseorAttack1:Sex1	0.508724
## PhysActivity1:Fruits1	0.123227
## PhysActivity1:Veggies1	0.163613
## PhysActivity1:HvyAlcoholConsump1	0.030913 *
## PhysActivity1:AnyHealthcare1	0.335688
## PhysActivity1:NoDocbcCost1	0.851255
## PhysActivity1:GenHealthCategoryGood Health	0.012062 *
## PhysActivity1:GenHealthCategoryPoor Health	0.904453
## PhysActivity1:DiffWalk1	0.990454
## PhysActivity1:Sex1	0.336558
## Fruits1:Veggies1	0.000148 ***
## Fruits1:HvyAlcoholConsump1	0.987353
## Fruits1:AnyHealthcare1	0.587314
## Fruits1:NoDocbcCost1	0.779073
## Fruits1:GenHealthCategoryGood Health	0.803241
## Fruits1:GenHealthCategoryPoor Health	0.621839
## Fruits1:DiffWalk1	0.319122
## Fruits1:Sex1	0.001474 **
## Veggies1:HvyAlcoholConsump1	0.518657
## Veggies1:AnyHealthcare1	0.228642
## Veggies1:NoDocbcCost1	0.034795 *
## Veggies1:GenHealthCategoryGood Health	0.410248
## Veggies1:GenHealthCategoryPoor Health	0.194136
## Veggies1:DiffWalk1	0.924778
## Veggies1:Sex1	0.391014
## HvyAlcoholConsump1:AnyHealthcare1	0.080046 .
## HvyAlcoholConsump1:NoDocbcCost1	0.938825
## HvyAlcoholConsump1:GenHealthCategoryGood Health	0.136847
## HvyAlcoholConsump1:GenHealthCategoryPoor Health	0.289375
## HvyAlcoholConsump1:DiffWalk1	0.631167
## HvyAlcoholConsump1:Sex1	0.024017 *
## AnyHealthcare1:NoDocbcCost1	0.314715
## AnyHealthcare1:GenHealthCategoryGood Health	0.050897 .
## AnyHealthcare1:GenHealthCategoryPoor Health	0.571770
## AnyHealthcare1:DiffWalk1	0.346558
## AnyHealthcare1:Sex1	0.024940 *
## NoDocbcCost1:GenHealthCategoryGood Health	0.741348

```

## NoDocbcCost1:GenHealthCategoryPoor Health          0.123306
## NoDocbcCost1:DiffWalk1                           0.426802
## NoDocbcCost1:Sex1                             0.038757 *
## GenHealthCategoryGood Health:DiffWalk1        1.53e-07 ***
## GenHealthCategoryPoor Health:DiffWalk1       0.001606 **
## GenHealthCategoryGood Health:Sex1            0.040962 *
## GenHealthCategoryPoor Health:Sex1           0.139966
## DiffWalk1:Sex1                            0.010530 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ',' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 151176  on 195065  degrees of freedom
## Residual deviance: 118587  on 194745  degrees of freedom
## AIC: 119229
##
## Number of Fisher Scoring iterations: 8

# Predict on the test set with the selected significant variables
preds_interaction_step <- predict(interaction_model_step, newdata = test_df, type = "response")

# Convert probabilities to binary predictions (threshold 0.5)
resp_interaction_step <- ifelse(preds_interaction_step >= 0.5, 1, 0)

# Confusion matrix for model evaluation
confusion_interaction_step <- table(test_df$Diabetes_binary, resp_interaction_step)
print(confusion_interaction_step)

##      resp_interaction_step
##      0          1
## 0 41700    561
## 1  5717    789

# Sensitivity (True Positive Rate)
sensitivity_interaction_step <- confusion_interaction_step[2, 2] / (confusion_interaction_step[2, 1] + confusion_interaction_step[1, 2])

# Overall accuracy of the model
accuracy_interaction_step <- sum(diag(confusion_interaction_step)) / sum(confusion_interaction_step)

# ROC (Area Under the Curve)
roc_obj_interaction_step <- roc(test_df$Diabetes_binary, as.numeric(preds_interaction_step))

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

auc_interaction_step <- auc(roc_obj_interaction_step)

# Display confusion matrix, sensitivity, accuracy, and AUC
cat("Confusion Matrix:\n")

```

```

## Confusion Matrix:

print(confusion_interaction_step)

##      resp_interaction_step
##            0      1
##    0 41700   561
##    1  5717   789

cat("Sensitivity:", sensitivity_interaction_step, "\n")

## Sensitivity: 0.1212727

cat("Accuracy:", accuracy_interaction_step, "\n")

## Accuracy: 0.8712654

cat("AUC:", auc_interaction_step, "\n")

## AUC: 0.824091

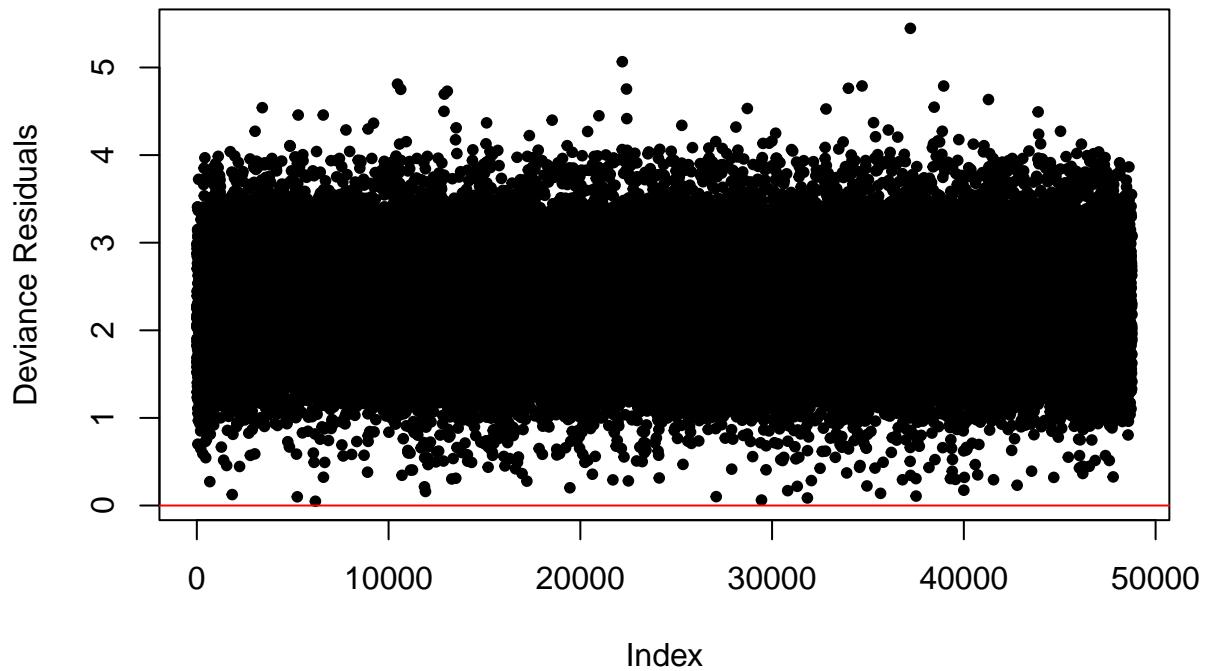
# Deviance residuals plot
deviance_residuals_interaction_step <- sign(as.numeric(test_df$Diabetes_binary) - preds_interaction_step)
deviance_residuals_interaction_step <- sqrt(-2 * (as.numeric(test_df$Diabetes_binary) * log(preds_interaction_step)
(1 - as.numeric(test_df$Diabetes_binary)) * log(1 - pred

## Warning in sqrt(-2 * (as.numeric(test_df$Diabetes_binary) *
## log(preds_interaction_step) + : NaNs produced

plot(deviance_residuals_interaction_step, main = "Deviance Residuals of Interaction Stepwise Model",
     ylab = "Deviance Residuals", xlab = "Index", pch = 20)
abline(h = 0, col = "red")

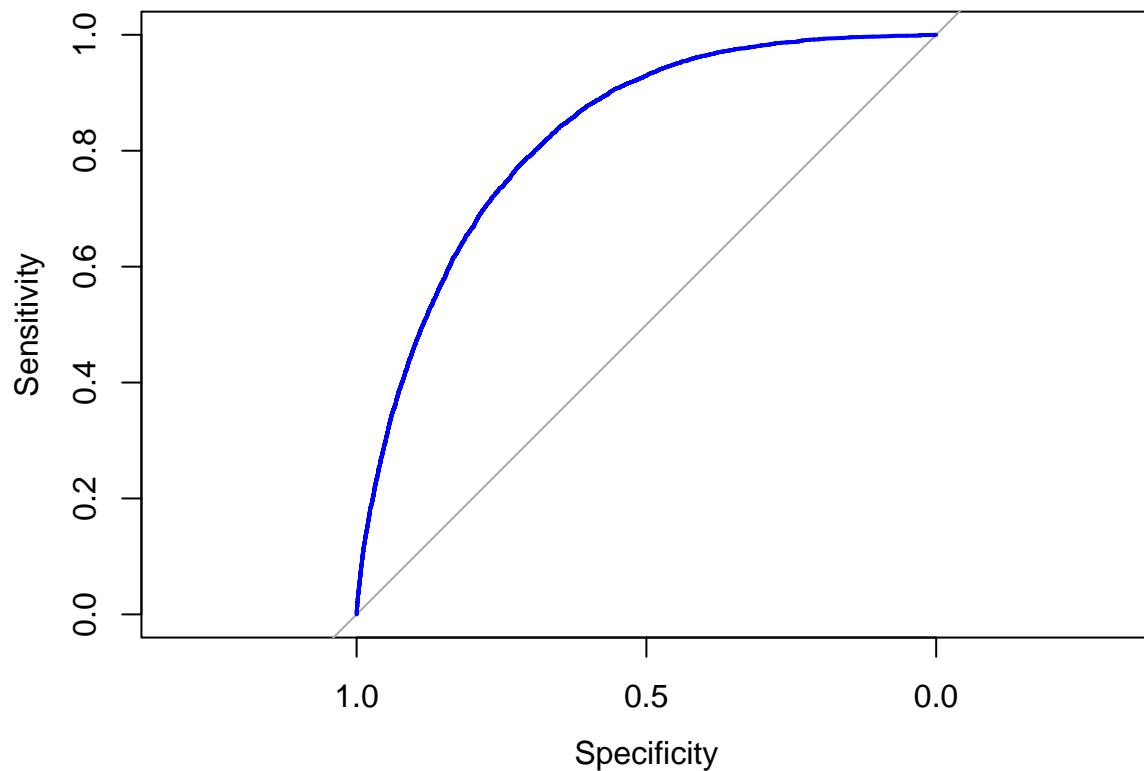
```

Deviance Residuals of Interaction Stepwise Model



```
# ROC curve
plot(roc_obj_interaction_step, col = "blue", main = "ROC Curve for Interaction Stepwise Model")
```

ROC Curve for Interaction Stepwise Model



```
cat("AUC for Interaction Stepwise Model:", auc_interaction_step, "\n")
```

```
## AUC for Interaction Stepwise Model: 0.824091
```