Model 04

2024-07-29

#Importing the necessary libraries

```
library(dplyr)
```

```
##
## 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(tidyr)
library(prettyR)
library(dplyr)
library(caret)
```

```
## Warning: package 'caret' was built under R version 4.3.3
```

```
## Loading required package: ggplot2
```

```
## Warning: package 'ggplot2' was built under R version 4.3.3
```

```
## Loading required package: lattice
```

```
library(rpart)
library(partykit)
```

```
## Warning: package 'partykit' was built under R version 4.3.3
```

```
## Loading required package: grid
```

```
## Loading required package: libcoin
```

```
## Warning: package 'libcoin' was built under R version 4.3.3
```

```
## Loading required package: mvtnorm
```

```
## Warning: package 'mvtnorm' was built under R version 4.3.3
```

```
library(prettyR)
```

Loading the data file from Wave 2 interviews to calculate the BMI

```
load("34921-0001-Data.rda")

da34921.0001 <- da34921.0001 %>%
  mutate(
         OBESITY = case_when(
           ((WEIGHT)/(HEIGHT*HEIGHT) * 703) >= 30.000 ~ 1,
           ((WEIGHT)/(HEIGHT*HEIGHT) * 703) < 30.000 ~ 0
         ))

obesity <- da34921.0001 %>% select(ID, OBESITY)
head(obesity)

##       ID OBESITY
## 1 100005       0
## 2 100033       1
## 3 100067       0
## 4 100080       1
## 5 100149       1
## 6 100154       0
```

Loading and Processing the Independent Social Network Variables to calculate Bridge from WAVE 1.

```
load("20541-0001-Data.rda")
load("20541-0004-Data.rda")


da20541.0001 <- da20541.0001 %>%
  select (ID, HEARN_RECODE, GENDER, AGE, RACE_RECODE, ETHGRP, COMBUILD,
CONDITNS_6, CONDITNS_7,CONDITNS_4,CONDITNS_1,CONDITNS_5,EXERCISE,NOTEAT,
FLTDEP, DEGREE_RECODE, HISPANIC, MARITLST,JOBSTAT_1, PHYSHLTH, MNTLHLTH,
ATNDSERV )

names(da20541.0001)[names(da20541.0001) == "CONDITNS_6"] <- "HYPERTENSION"
names(da20541.0001)[names(da20541.0001) == "CONDITNS_7"] <- "DIABETES"
names(da20541.0001)[names(da20541.0001) == "CONDITNS_4"] <- "ASTHMA"
names(da20541.0001)[names(da20541.0001) == "CONDITNS_1"] <- "ARTHRITIS"
names(da20541.0001)[names(da20541.0001) == "CONDITNS_5"] <- "STROKE"

da20541.0001 <- da20541.0001 %>%
  mutate(DEGREE_RECODE = if_else(DEGREE_RECODE == "(-2) don't know", NA,
DEGREE_RECODE),
         HEARN_RECODE = if_else(HEARN_RECODE == "(-2) don't know", NA,
HEARN_RECODE),
         RACE_RECODE = if_else(RACE_RECODE == "(-2) don't know", NA,
RACE_RECODE))

head(da20541.0001)
```

```
##       ID       HEARN_RECODE     GENDER AGE         RACE_RECODE
## 1 100005 (4) 100k or higher (2) female  62 (1) white/caucasian
## 2 100033  (2) 25,000-49,999 (2) female  79 (1) white/caucasian
## 3 100080  (3) 50,000-99,999   (1) male  60 (1) white/caucasian
## 4 100154  (2) 25,000-49,999 (2) female  78 (1) white/caucasian
## 5 100203               <NA> (2) female  61 (1) white/caucasian
## 6 100359  (3) 50,000-99,999   (1) male  75 (1) white/caucasian
##                    ETHGRP          COMBUILD HYPERTENSION DIABETES ASTHMA
## 1               (1) white       (3) average      (1) yes   (0) no (0) no
## 2               (1) white (4) above average      (1) yes   (0) no (0) no
## 3               (1) white       (3) average      (1) yes  (1) yes (0) no
## 4               (1) white       (3) average      (1) yes   (0) no (0) no
## 5 (3) hispanic, non-black       (3) average      (1) yes  (1) yes (0) no
## 6               (1) white       (3) average      (1) yes   (0) no (0) no
##   ARTHRITIS STROKE EXERCISE                          NOTEAT
## 1   (1) yes (0) no   (0) no (1) rarely or none of the time
## 2   (1) yes (0) no   (0) no (1) rarely or none of the time
## 3   (1) yes (0) no   (0) no (1) rarely or none of the time
## 4   (1) yes (0) no   (0) no           (2) some of the time
## 5    (0) no (0) no   (0) no           (2) some of the time
## 6   (1) yes (0) no   (0) no (1) rarely or none of the time
##                            FLTDEP                       DEGREE_RECODE
HISPANIC
## 1 (1) rarely or none of the time                         (5) masters   (0)
no
## 2 (1) rarely or none of the time (2) high school diploma/equivalency   (0)
no
## 3 (1) rarely or none of the time (2) high school diploma/equivalency   (0)
no
## 4           (2) some of the time (2) high school diploma/equivalency   (0)
no
## 5           (2) some of the time                           (1) none  (1)
yes
## 6           (2) some of the time (2) high school diploma/equivalency   (0)
no
##      MARITLST JOBSTAT_1      PHYSHLTH      MNTLHLTH
ATNDSERV
## 1 (1) married   (1) yes (4) very good (4) very good  (3) several times a
year
## 2 (5) widowed    (0) no (4) very good (4) very good (1) less than once a
year
## 3 (1) married   (1) yes      (3) good (5) excellent            (5) every
week
## 4 (1) married    (0) no      (3) good      (3) good  (6) several times a
week
## 5 (5) widowed   (1) yes      (1) poor      (2) fair                 (0)
never
## 6 (1) married    (0) no      (2) fair      (3) good  (6) several times a
week
```

```r
nrow(da20541.0001)
```

```
## [1] 3005
```

```r
da20541.0004 <- da20541.0004 %>%
  group_by(ID) %>%
  filter(n() > 2) %>%
  ungroup()

da20541.0004 <- da20541.0004 %>%
  pivot_longer(
    cols = starts_with("TALKFREQ"),
    names_to = "TALKFREQ",
    values_to = "FREQ"
  )

da20541.0004 <- da20541.0004 %>%
  group_by(ID) %>%
  summarize(
    BRIDGE = if_else(any(FREQ == '(0) have never spoken to each other', na.rm
= TRUE), 1, 0),
    HEALTHDISCUSSIONS = if_else(any(HEALTHTALK == '(3) very likely', na.rm =
TRUE), 1, 0),
    LIVEALONE = if_else(any(LIVEWITH == '(1) yes -- lives in the same
household', na.rm = TRUE), 0,1))

head(da20541.0004)
```

```
## # A tibble: 6 × 4
##   ID     BRIDGE HEALTHDISCUSSIONS LIVEALONE
##   <fct>   <dbl>             <dbl>     <dbl>
## 1 100005      1                 1         0
## 2 100033      0                 1         0
## 3 100080      1                 1         0
## 4 100154      1                 1         0
## 5 100203      0                 1         0
## 6 100359      0                 1         0
```

```r
nrow(da20541.0004)
```

```
## [1] 2522
```

```r
modeldata <- da20541.0001 %>%
  left_join(da20541.0004, by = "ID")

modeldata <- modeldata %>%
  left_join(obesity, by = "ID")

modeldata<- na.omit(modeldata)
modeldata <- modeldata %>% select(-ID)
```

```r
modeldata$BRIDGE <- as.factor(modeldata$BRIDGE)
modeldata$HEALTHDISCUSSIONS <- as.factor(modeldata$HEALTHDISCUSSIONS)
modeldata$LIVEALONE <- as.factor(modeldata$LIVEALONE)
modeldata$OBESITY <- as.factor(modeldata$OBESITY)
head(modeldata)
```

```
##         HEARN_RECODE     GENDER AGE         RACE_RECODE    ETHGRP
## 1 (4) 100k or higher (2) female  62 (1) white/caucasian (1) white
## 2  (2) 25,000-49,999 (2) female  79 (1) white/caucasian (1) white
## 3  (3) 50,000-99,999   (1) male  60 (1) white/caucasian (1) white
## 4  (2) 25,000-49,999 (2) female  78 (1) white/caucasian (1) white
## 7  (2) 25,000-49,999   (1) male  80 (1) white/caucasian (1) white
## 9  (3) 50,000-99,999 (2) female  59 (1) white/caucasian (1) white
##            COMBUILD HYPERTENSION DIABETES ASTHMA ARTHRITIS STROKE EXERCISE
## 1       (3) average      (1) yes   (0) no (0) no   (1) yes (0) no   (0) no
## 2 (4) above average      (1) yes   (0) no (0) no   (1) yes (0) no   (0) no
## 3       (3) average      (1) yes  (1) yes (0) no   (1) yes (0) no   (0) no
## 4       (3) average      (1) yes   (0) no (0) no   (1) yes (0) no   (0) no
## 7 (4) above average      (1) yes   (0) no (0) no    (0) no (0) no   (0) no
## 9       (3) average      (1) yes   (0) no (0) no    (0) no (0) no   (0) no
##                            NOTEAT                          FLTDEP
## 1 (1) rarely or none of the time (1) rarely or none of the time
## 2 (1) rarely or none of the time (1) rarely or none of the time
## 3 (1) rarely or none of the time (1) rarely or none of the time
## 4           (2) some of the time           (2) some of the time
## 7 (1) rarely or none of the time (1) rarely or none of the time
## 9 (1) rarely or none of the time (1) rarely or none of the time
##                         DEGREE_RECODE HISPANIC    MARITLST JOBSTAT_1
## 1                         (5) masters   (0) no (1) married   (1) yes
## 2 (2) high school diploma/equivalency   (0) no (5) widowed    (0) no
## 3 (2) high school diploma/equivalency   (0) no (1) married   (1) yes
## 4 (2) high school diploma/equivalency   (0) no (1) married    (0) no
## 7 (2) high school diploma/equivalency   (0) no (5) widowed    (0) no
## 9 (2) high school diploma/equivalency   (0) no (1) married   (1) yes
##        PHYSHLTH      MNTLHLTH                       ATNDSERV BRIDGE
## 1 (4) very good (4) very good       (3) several times a year      1
## 2 (4) very good (4) very good      (1) less than once a year      0
## 3      (3) good (5) excellent                (5) every week      1
## 4      (3) good      (3) good       (6) several times a week      1
## 7      (3) good      (3) good                (5) every week      0
## 9 (4) very good (4) very good (2) about once or twice a year      1
##   HEALTHDISCUSSIONS LIVEALONE OBESITY
## 1                 1         0       0
## 2                 1         0       1
## 3                 1         0       1
## 4                 1         0       0
## 7                 1         1       0
## 9                 1         0       0
```

Creating Data Partition for 70% Training Data and 30% Testing Data

```
library(rpart)
library(caret)

set.seed(19032023)

index <- createDataPartition(modeldata$OBESITY,
                             p=0.7,
                             list=FALSE,
                             times = 1
                           )

modeldata.train <- modeldata[index,]
modeldata.test <- modeldata[-index,]

nrow(modeldata.train)

## [1] 875

nrow(modeldata.test)

## [1] 374
```

Applying Logistic Regression on to find the association between Bridge and Obesity.

```
model.lr <- glm(OBESITY ~ ., data = modeldata.train, family = "binomial")

summary.lr <- summary(model.lr)
```

p-value for Bridge variable

```
print(summary.lr)

##
## Call:
## glm(formula = OBESITY ~ ., family = "binomial", data = modeldata.train)
##
## Coefficients: (1 not defined because of singularities)
##
Estimate
## (Intercept)
0.855674
## HEARN_RECODE(1) 0-24,999
-0.138948
## HEARN_RECODE(2) 25,000-49,999
-0.642000
## HEARN_RECODE(3) 50,000-99,999
```

```
-0.007764
## HEARN_RECODE(4) 100k or higher
-0.404381
## GENDER(2) female
-0.179268
## AGE
-0.060938
## RACE_RECODE(2) black/african american
0.457448
## RACE_RECODE(3) asian, pacific islander, american indian or alaskan native
-0.587417
## ETHGRP(2) black
NA
## ETHGRP(3) hispanic, non-black
-13.280666
## ETHGRP(4) other
0.318138
## COMBUILD(2) below average
0.570431
## COMBUILD(3) average
1.190760
## COMBUILD(4) above average
0.794813
## COMBUILD(5) far above average
1.335705
## HYPERTENSION(1) yes
0.699643
## DIABETES(1) yes
0.774167
## ASTHMA(1) yes
0.461655
## ARTHRITIS(1) yes
0.485375
## STROKE(1) yes
-0.255597
## EXERCISE(1) yes
0.118071
## NOTEAT(2) some of the time
0.195676
## NOTEAT(3) occasionally
0.007752
## NOTEAT(4) most of the time
-1.471371
## FLTDEP(2) some of the time
0.009499
## FLTDEP(3) occasionally
0.173227
## FLTDEP(4) most of the time
-0.643105
## DEGREE_RECODE(2) high school diploma/equivalency
```

```
0.356742
## DEGREE_RECODE(3) associates
-0.108598
## DEGREE_RECODE(4) bachelors
-0.094170
## DEGREE_RECODE(5) masters
-0.417242
## DEGREE_RECODE(6) law, md or phd
0.210618
## HISPANIC(1) yes
13.856662
## MARITLST(2) living with a partner
0.220220
## MARITLST(3) separated
-0.815975
## MARITLST(4) divorced
0.178953
## MARITLST(5) widowed
0.353827
## MARITLST(6) never married
0.417238
## JOBSTAT_1(1) yes
0.369579
## PHYSHLTH(2) fair
0.198072
## PHYSHLTH(3) good
-0.203859
## PHYSHLTH(4) very good
-0.648715
## PHYSHLTH(5) excellent
-0.890077
## MNTLHLTH(2) fair
-0.744250
## MNTLHLTH(3) good
-0.574006
## MNTLHLTH(4) very good
-0.209373
## MNTLHLTH(5) excellent
-0.447808
## ATNDSERV(1) less than once a year
-0.273375
## ATNDSERV(2) about once or twice a year
-0.533960
## ATNDSERV(3) several times a year
0.482579
## ATNDSERV(4) about once a month
-0.044166
## ATNDSERV(5) every week
-0.268687
## ATNDSERV(6) several times a week
```

```
-0.174500
## BRIDGE1
-0.303202
## HEALTHDISCUSSIONS1
1.914274
## LIVEALONE1
0.005291
##
Std. Error
## (Intercept)
1.984841
## HEARN_RECODE(1) 0-24,999
0.310656
## HEARN_RECODE(2) 25,000-49,999
0.300880
## HEARN_RECODE(3) 50,000-99,999
0.296451
## HEARN_RECODE(4) 100k or higher
0.372787
## GENDER(2) female
0.181651
## AGE
0.013459
## RACE_RECODE(2) black/african american
0.244359
## RACE_RECODE(3) asian, pacific islander, american indian or alaskan native
0.538124
## ETHGRP(2) black
NA
## ETHGRP(3) hispanic, non-black
367.721128
## ETHGRP(4) other
0.878281
## COMBUILD(2) below average
0.944963
## COMBUILD(3) average
0.877985
## COMBUILD(4) above average
0.886092
## COMBUILD(5) far above average
0.927117
## HYPERTENSION(1) yes
0.172001
## DIABETES(1) yes
0.211538
## ASTHMA(1) yes
0.277647
## ARTHRITIS(1) yes
0.171699
## STROKE(1) yes
```

```
0.335890
## EXERCISE(1) yes
0.390855
## NOTEAT(2) some of the time
0.260326
## NOTEAT(3) occasionally
0.313968
## NOTEAT(4) most of the time
0.674064
## FLTDEP(2) some of the time
0.239706
## FLTDEP(3) occasionally
0.280538
## FLTDEP(4) most of the time
0.676887
## DEGREE_RECODE(2) high school diploma/equivalency
0.250091
## DEGREE_RECODE(3) associates
0.287020
## DEGREE_RECODE(4) bachelors
0.315612
## DEGREE_RECODE(5) masters
0.374522
## DEGREE_RECODE(6) law, md or phd
0.558010
## HISPANIC(1) yes
367.720961
## MARITLST(2) living with a partner
0.567623
## MARITLST(3) separated
0.840749
## MARITLST(4) divorced
0.359010
## MARITLST(5) widowed
0.340944
## MARITLST(6) never married
0.572834
## JOBSTAT_1(1) yes
0.186624
## PHYSHLTH(2) fair
0.459150
## PHYSHLTH(3) good
0.457222
## PHYSHLTH(4) very good
0.469655
## PHYSHLTH(5) excellent
0.529214
## MNTLHLTH(2) fair
1.039685
## MNTLHLTH(3) good
```

```
1.038536
## MNTLHLTH(4) very good
1.041877
## MNTLHLTH(5) excellent
1.050288
## ATNDSERV(1) less than once a year
0.488157
## ATNDSERV(2) about once or twice a year
0.344471
## ATNDSERV(3) several times a year
0.311633
## ATNDSERV(4) about once a month
0.346150
## ATNDSERV(5) every week
0.251633
## ATNDSERV(6) several times a week
0.312994
## BRIDGE1
0.168059
## HEALTHDISCUSSIONS1
1.122846
## LIVEALONE1
0.303274
##
z value
## (Intercept)
0.431
## HEARN_RECODE(1) 0-24,999
-0.447
## HEARN_RECODE(2) 25,000-49,999
-2.134
## HEARN_RECODE(3) 50,000-99,999
-0.026
## HEARN_RECODE(4) 100k or higher
-1.085
## GENDER(2) female
-0.987
## AGE
-4.528
## RACE_RECODE(2) black/african american
1.872
## RACE_RECODE(3) asian, pacific islander, american indian or alaskan native
-1.092
## ETHGRP(2) black
NA
## ETHGRP(3) hispanic, non-black
-0.036
## ETHGRP(4) other
0.362
## COMBUILD(2) below average
```

```
0.604
## COMBUILD(3) average
1.356
## COMBUILD(4) above average
0.897
## COMBUILD(5) far above average
1.441
## HYPERTENSION(1) yes
4.068
## DIABETES(1) yes
3.660
## ASTHMA(1) yes
1.663
## ARTHRITIS(1) yes
2.827
## STROKE(1) yes
-0.761
## EXERCISE(1) yes
0.302
## NOTEAT(2) some of the time
0.752
## NOTEAT(3) occasionally
0.025
## NOTEAT(4) most of the time
-2.183
## FLTDEP(2) some of the time
0.040
## FLTDEP(3) occasionally
0.617
## FLTDEP(4) most of the time
-0.950
## DEGREE_RECODE(2) high school diploma/equivalency
1.426
## DEGREE_RECODE(3) associates
-0.378
## DEGREE_RECODE(4) bachelors
-0.298
## DEGREE_RECODE(5) masters
-1.114
## DEGREE_RECODE(6) law, md or phd
0.377
## HISPANIC(1) yes
0.038
## MARITLST(2) living with a partner
0.388
## MARITLST(3) separated
-0.971
## MARITLST(4) divorced
0.498
## MARITLST(5) widowed
```

```
1.038
## MARITLST(6) never married
0.728
## JOBSTAT_1(1) yes
1.980
## PHYSHLTH(2) fair
0.431
## PHYSHLTH(3) good
-0.446
## PHYSHLTH(4) very good
-1.381
## PHYSHLTH(5) excellent
-1.682
## MNTLHLTH(2) fair
-0.716
## MNTLHLTH(3) good
-0.553
## MNTLHLTH(4) very good
-0.201
## MNTLHLTH(5) excellent
-0.426
## ATNDSERV(1) less than once a year
-0.560
## ATNDSERV(2) about once or twice a year
-1.550
## ATNDSERV(3) several times a year
1.549
## ATNDSERV(4) about once a month
-0.128
## ATNDSERV(5) every week
-1.068
## ATNDSERV(6) several times a week
-0.558
## BRIDGE1
-1.804
## HEALTHDISCUSSIONS1
1.705
## LIVEALONE1
0.017
##
Pr(>|z|)
## (Intercept)
0.666392
## HEARN_RECODE(1) 0-24,999
0.654678
## HEARN_RECODE(2) 25,000-49,999
0.032864
## HEARN_RECODE(3) 50,000-99,999
0.979105
## HEARN_RECODE(4) 100k or higher
```

```
0.278032
## GENDER(2) female
0.323703
## AGE
5.97e-06
## RACE_RECODE(2) black/african american
0.061202
## RACE_RECODE(3) asian, pacific islander, american indian or alaskan native
0.275008
## ETHGRP(2) black
NA
## ETHGRP(3) hispanic, non-black
0.971190
## ETHGRP(4) other
0.717182
## COMBUILD(2) below average
0.546074
## COMBUILD(3) average
0.175022
## COMBUILD(4) above average
0.369726
## COMBUILD(5) far above average
0.149667
## HYPERTENSION(1) yes
4.75e-05
## DIABETES(1) yes
0.000253
## ASTHMA(1) yes
0.096364
## ARTHRITIS(1) yes
0.004700
## STROKE(1) yes
0.446684
## EXERCISE(1) yes
0.762587
## NOTEAT(2) some of the time
0.452257
## NOTEAT(3) occasionally
0.980302
## NOTEAT(4) most of the time
0.029048
## FLTDEP(2) some of the time
0.968391
## FLTDEP(3) occasionally
0.536917
## FLTDEP(4) most of the time
0.342065
## DEGREE_RECODE(2) high school diploma/equivalency
0.153739
## DEGREE_RECODE(3) associates
```

```
0.705160
## DEGREE_RECODE(4) bachelors
0.765419
## DEGREE_RECODE(5) masters
0.265252
## DEGREE_RECODE(6) law, md or phd
0.705843
## HISPANIC(1) yes
0.969941
## MARITLST(2) living with a partner
0.698039
## MARITLST(3) separated
0.331781
## MARITLST(4) divorced
0.618158
## MARITLST(5) widowed
0.299370
## MARITLST(6) never married
0.466384
## JOBSTAT_1(1) yes
0.047666
## PHYSHLTH(2) fair
0.666186
## PHYSHLTH(3) good
0.655695
## PHYSHLTH(4) very good
0.167200
## PHYSHLTH(5) excellent
0.092591
## MNTLHLTH(2) fair
0.474089
## MNTLHLTH(3) good
0.580464
## MNTLHLTH(4) very good
0.840732
## MNTLHLTH(5) excellent
0.669840
## ATNDSERV(1) less than once a year
0.575469
## ATNDSERV(2) about once or twice a year
0.121120
## ATNDSERV(3) several times a year
0.121490
## ATNDSERV(4) about once a month
0.898472
## ATNDSERV(5) every week
0.285624
## ATNDSERV(6) several times a week
0.577172
## BRIDGE1
```

```
0.071210
## HEALTHDISCUSSIONS1
0.088224
## LIVEALONE1
0.986080
##
## (Intercept)
## HEARN_RECODE(1) 0-24,999
## HEARN_RECODE(2) 25,000-49,999
*
## HEARN_RECODE(3) 50,000-99,999
## HEARN_RECODE(4) 100k or higher
## GENDER(2) female
## AGE
***
## RACE_RECODE(2) black/african american
.
## RACE_RECODE(3) asian, pacific islander, american indian or alaskan native
## ETHGRP(2) black
## ETHGRP(3) hispanic, non-black
## ETHGRP(4) other
## COMBUILD(2) below average
## COMBUILD(3) average
## COMBUILD(4) above average
## COMBUILD(5) far above average
## HYPERTENSION(1) yes
***
## DIABETES(1) yes
***
## ASTHMA(1) yes
.
## ARTHRITIS(1) yes
**
## STROKE(1) yes
## EXERCISE(1) yes
## NOTEAT(2) some of the time
## NOTEAT(3) occasionally
## NOTEAT(4) most of the time
*
## FLTDEP(2) some of the time
## FLTDEP(3) occasionally
## FLTDEP(4) most of the time
## DEGREE_RECODE(2) high school diploma/equivalency
## DEGREE_RECODE(3) associates
## DEGREE_RECODE(4) bachelors
## DEGREE_RECODE(5) masters
## DEGREE_RECODE(6) law, md or phd
## HISPANIC(1) yes
## MARITLST(2) living with a partner
## MARITLST(3) separated
```

```
## MARITLST(4) divorced
## MARITLST(5) widowed
## MARITLST(6) never married
## JOBSTAT_1(1) yes
*
## PHYSHLTH(2) fair
## PHYSHLTH(3) good
## PHYSHLTH(4) very good
## PHYSHLTH(5) excellent
.
## MNTLHLTH(2) fair
## MNTLHLTH(3) good
## MNTLHLTH(4) very good
## MNTLHLTH(5) excellent
## ATNDSERV(1) less than once a year
## ATNDSERV(2) about once or twice a year
## ATNDSERV(3) several times a year
## ATNDSERV(4) about once a month
## ATNDSERV(5) every week
## ATNDSERV(6) several times a week
## BRIDGE1
.
## HEALTHDISCUSSIONS1
.
## LIVEALONE1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1167.35  on 874  degrees of freedom
## Residual deviance:  967.19  on 819  degrees of freedom
## AIC: 1079.2
##
## Number of Fisher Scoring iterations: 12
```

Odds Ratio and 95% Confidence Interval

```
odds_ratio <- exp(coef(model.lr)["BRIDGE1"])
print(odds_ratio)
```

```
##   BRIDGE1
## 0.7384502
```

```
conf_int <- exp(confint(model.lr, "BRIDGE1"))
```

```
## Waiting for profiling to be done...
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

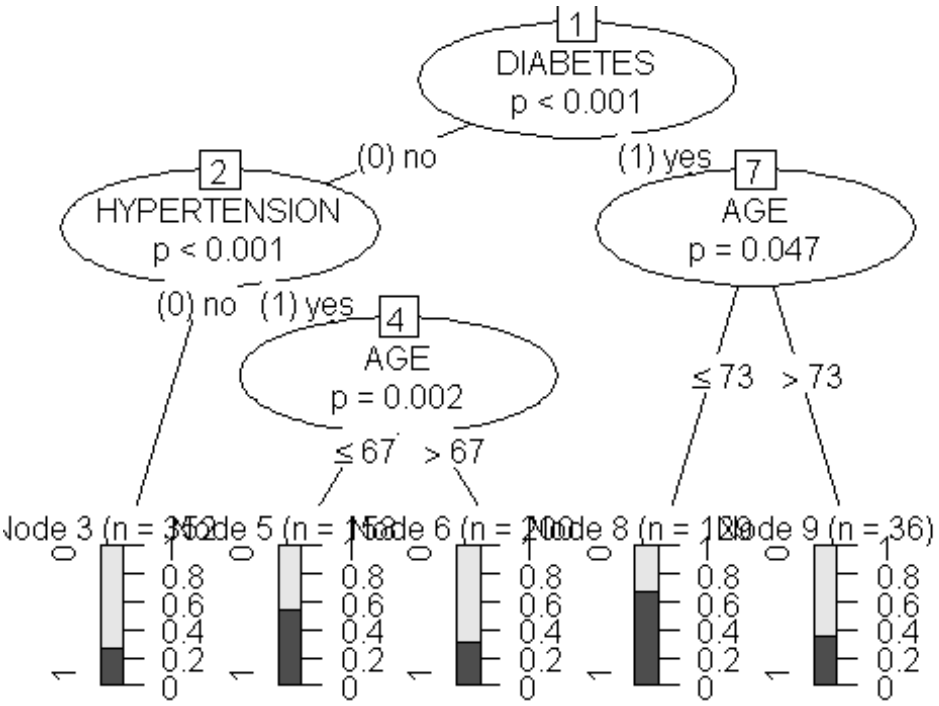
```
print(conf_int)
```

```
##     2.5 %    97.5 %
## 0.5307437 1.0262219

predicted.prob.lr <- predict(model.lr, modeldata.test, type = "response")
predicted.obesity.lr <- ifelse(predicted.prob.lr > 0.5, 1, 0)

actual.obesity.lr <- modeldata.test$OBESITY
conf.matrix.lr <- table(Predicted = predicted.obesity.lr, Actual =
actual.obesity.lr)

print(conf.matrix.lr)

##          Actual
## Predicted   0   1
##         0 178  86
##         1  52  58

confusionMatrix(factor(predicted.obesity.lr), factor(modeldata.test$OBESITY),
positive = as.character(1))

## Confusion Matrix and Statistics
##
##           Reference
## Prediction   0   1
##          0 178  86
##          1  52  58
##
##                Accuracy : 0.631
##                  95% CI : (0.5799, 0.68)
##     No Information Rate : 0.615
##     P-Value [Acc > NIR] : 0.280390
##
##                   Kappa : 0.1849
##
##  Mcnemar's Test P-Value : 0.004967
##
##             Sensitivity : 0.4028
##             Specificity : 0.7739
##          Pos Pred Value : 0.5273
##          Neg Pred Value : 0.6742
##              Prevalence : 0.3850
##          Detection Rate : 0.1551
##    Detection Prevalence : 0.2941
##       Balanced Accuracy : 0.5883
##
##        'Positive' Class : 1
##
```

Decision Tree

Conditional Inference Tree implementation using ctree

```
set.seed(123)

model.dt <- ctree(OBESITY ~ .,
                                data=modeldata.train)
plot(model.dt)
```




```
predictions.dt <- predict(model.dt, modeldata.test)

confusionMatrix(predictions.dt, modeldata.test$OBESITY)
```

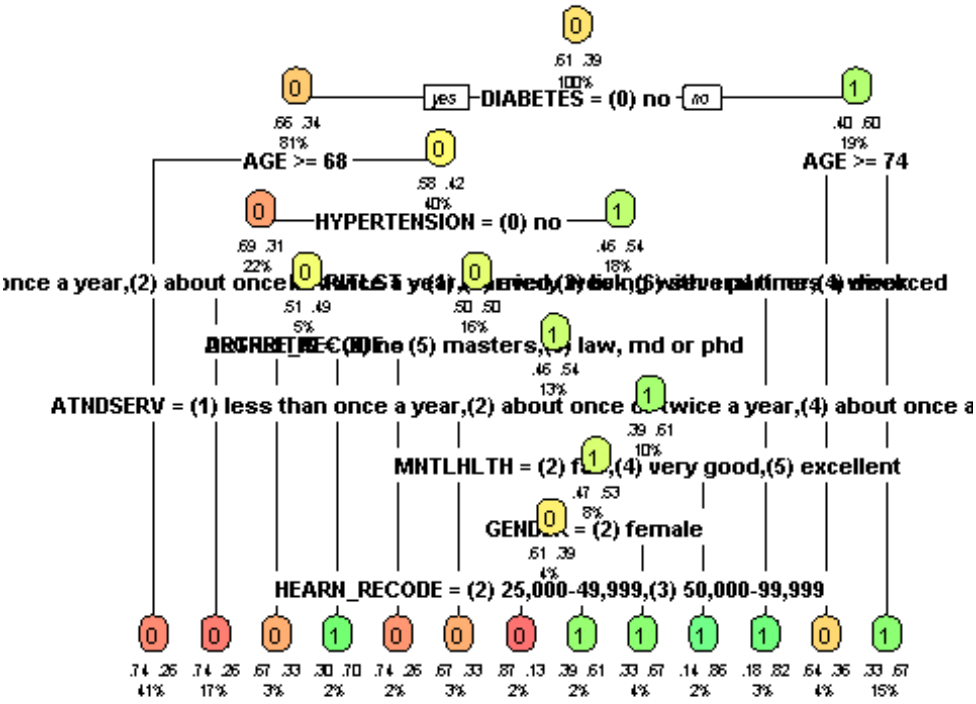
```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   0   1
##          0 176  77
##          1  54  67
##
##                Accuracy : 0.6497
##                  95% CI : (0.599, 0.6981)
##     No Information Rate : 0.615
##     P-Value [Acc > NIR] : 0.09147
##
```

```
##                   Kappa : 0.2376
##
##  Mcnemar's Test P-Value : 0.05459
##
##             Sensitivity : 0.7652
##             Specificity : 0.4653
##          Pos Pred Value : 0.6957
##          Neg Pred Value : 0.5537
##              Prevalence : 0.6150
##          Detection Rate : 0.4706
##    Detection Prevalence : 0.6765
##       Balanced Accuracy : 0.6152
##
##        'Positive' Class : 0
##
```

Classification and Regression Tree implementation using rpart

```
rpart.tree <- rpart(OBESITY ~ ., data = modeldata.train, parms  = list(split
= "information"))
rpart.tree

## n= 875
##
## node), split, n, loss, yval, (yprob)
##       * denotes terminal node
##
##   1) root 875 338 0 (0.6137143 0.3862857)
##     2) DIABETES=(0) no 710 239 0 (0.6633803 0.3366197)
##       4) AGE>=67.5 356  92 0 (0.7415730 0.2584270) *
##       5) AGE< 67.5 354 147 0 (0.5847458 0.4152542)
##        10) HYPERTENSION=(0) no 196  61 0 (0.6887755 0.3112245)
##          20) ATNDSERV=(0) never,(1) less than once a year,(2) about once
or twice a year,(5) every week,(6) several times a week 149  38 0 (0.7449664
0.2550336) *
##          21) ATNDSERV=(3) several times a year,(4) about once a month 47
23 0 (0.5106383 0.4893617)
##            42) ARTHRITIS=(0) no 27   9 0 (0.6666667 0.3333333) *
##            43) ARTHRITIS=(1) yes 20   6 1 (0.3000000 0.7000000) *
##        11) HYPERTENSION=(1) yes 158  72 1 (0.4556962 0.5443038)
##          22) MARITLST=(1) married,(2) living with a partner,(4) divorced
136  68 0 (0.5000000 0.5000000)
##            44) DEGREE_RECODE=(5) masters,(6) law, md or phd 19   5 0
(0.7368421 0.2631579) *
##            45) DEGREE_RECODE=(1) none,(2) high school
diploma/equivalency,(3) associates,(4) bachelors 117  54 1 (0.4615385
0.5384615)
##              90) ATNDSERV=(1) less than once a year,(2) about once or
twice a year,(4) about once a month 30  10 0 (0.6666667 0.3333333) *
##              91) ATNDSERV=(0) never,(3) several times a year,(5) every
```

```
week,(6) several times a week 87  34 1 (0.3908046 0.6091954)
##               182) MNTLHLTH=(2) fair,(4) very good,(5) excellent 66  31 1
(0.4696970 0.5303030)
##                 364) GENDER=(2) female 33  13 0 (0.6060606 0.3939394)
##                   728) HEARN_RECODE=(2) 25,000-49,999,(3) 50,000-99,999 15
2 0 (0.8666667 0.1333333) *
##                   729) HEARN_RECODE=(-1) refused,(1) 0-24,999,(4) 100k or
higher 18   7 1 (0.3888889 0.6111111) *
##                 365) GENDER=(1) male 33  11 1 (0.3333333 0.6666667) *
##               183) MNTLHLTH=(1) poor,(3) good 21   3 1 (0.1428571
0.8571429) *
##          23) MARITLST=(3) separated,(5) widowed,(6) never married 22   4 1
(0.1818182 0.8181818) *
##     3) DIABETES=(1) yes 165  66 1 (0.4000000 0.6000000)
##       6) AGE>=73.5 36  13 0 (0.6388889 0.3611111) *
##       7) AGE< 73.5 129  43 1 (0.3333333 0.6666667) *

library(rpart.plot)

## Warning: package 'rpart.plot' was built under R version 4.3.3

rpart.plot(
  rpart.tree,
  type = 2,
  extra = 104,
  under = TRUE,
  cex = 0.6,
  tweak = 1.1,
  box.palette = "RdYlGn",
  compress = TRUE
)

## Warning: cex and tweak both specified, applying both
```




```
library(dplyr)

importances <- varImp(rpart.tree) %>%
  arrange(desc(Overall))

importances
```

```
##                     Overall
## HYPERTENSION      37.6950341
## PHYSHLTH          36.5821021
## AGE               33.0132858
## ETHGRP            22.6198697
## DIABETES          19.1033086
## HEARN_RECODE      17.5535001
## DEGREE_RECODE     16.5356375
## ARTHRITIS         15.2116095
## ATNDSERV          14.0298152
## RACE_RECODE        9.6494528
## ASTHMA             7.0687866
## MNTLHLTH           6.6646176
## NOTEAT             6.1790473
## COMBUILD           4.2217617
## MARITLST           4.1971057
## GENDER             2.4956975
## STROKE             1.6369455
## JOBSTAT_1          0.4343071
## EXERCISE           0.0000000
```

```
## FLTDEP              0.0000000
## HISPANIC            0.0000000
## BRIDGE              0.0000000
## HEALTHDISCUSSIONS   0.0000000
## LIVEALONE           0.0000000
```

s