Logistic Regression on Depression Data

```r
library(mice)
```

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

##
## Attaching package: 'mice'

## The following object is masked from 'package:stats':
##
##     filter

## The following objects are masked from 'package:base':
##
##     cbind, rbind
```

#Loading Data and Exploring the dataset

```r
dep = read.csv('~/Desktop/Education/Winter term/QBS-121/b_depressed.csv', header=T)
head(dep)
```

```
##   Survey_id Ville_id sex Age Married Number_children education_level
## 1       926       91   1  28       1               4              10
## 2       747       57   1  23       1               3               8
## 3      1190      115   1  22       1               3               9
## 4      1065       97   1  27       1               2              10
## 5       806       42   0  59       0               4              10
## 6       483       25   1  35       1               6              10
##   total_members gained_asset durable_asset save_asset living_expenses
## 1             5     28912201      22861940   23399979        26692283
## 2             5     28912201      22861940   23399979        26692283
## 3             5     28912201      22861940   23399979        26692283
## 4             4     52667108      19698904   49647648          397715
## 5             6     82606287      17352654   23399979        80877619
## 6             8     35937466        736707   23399979        30696127
##   other_expenses incoming_salary incoming_own_farm incoming_business
## 1       28203066               0                 0                 0
## 2       28203066               0                 0                 0
## 3       28203066               0                 0                 0
## 4       44042267               0                 1                 0
## 5       74503502               1                 0                 0
## 6       11531066               0                 1                 0
##   incoming_no_business incoming_agricultural farm_expenses labor_primary
## 1                    0              30028818      31363432             0
## 2                    0              30028818      31363432             0
## 3                    0              30028818      31363432             0
## 4                    1              22288055      18751329             0
## 5                    0              53384566      20731006             1
## 6                    1              22688441      18907036             0
##   lasting_investment no_lasting_investmen depressed
```

```
## 1          28411718          28292707          0
## 2          28411718          28292707          1
## 3          28411718          28292707          0
## 4           7781123          69219765          0
## 5          20100562          43419447          0
## 6           4442561          76629095          0
```

```r
str(dep)
```

```
## 'data.frame':    1429 obs. of  23 variables:
##  $ Survey_id          : int  926 747 1190 1065 806 483 849 1386 930 390 ...
##  $ Ville_id           : int  91 57 115 97 42 25 130 72 195 33 ...
##  $ sex                : int  1 1 1 1 0 1 0 1 1 1 ...
##  $ Age                : int  28 23 22 27 59 35 34 21 32 29 ...
##  $ Married            : int  1 1 1 1 0 1 0 1 1 1 ...
##  $ Number_children    : int  4 3 3 2 4 6 1 2 7 4 ...
##  $ education_level    : int  10 8 9 10 10 10 9 10 9 10 ...
##  $ total_members      : int  5 5 5 4 6 8 3 4 9 5 ...
##  $ gained_asset       : int  28912201 28912201 28912201 52667108 82606287 35937466 41303144 1201363
##  $ durable_asset      : int  22861940 22861940 22861940 19698904 17352654 736707 21925041 20323505
##  $ save_asset         : int  23399979 23399979 23399979 49647648 23399979 23399979 23399979 4804610
##  $ living_expenses    : int  26692283 26692283 26692283 397715 80877619 30696127 66730708 80076849
##  $ other_expenses     : int  28203066 28203066 28203066 44042267 74503502 11531066 10890451 5845610
##  $ incoming_salary    : int  0 0 0 0 1 0 0 0 1 0 ...
##  $ incoming_own_farm  : int  0 0 0 1 0 1 0 0 0 0 ...
##  $ incoming_business  : int  0 0 0 0 0 0 0 1 0 0 ...
##  $ incoming_no_business : int  0 0 0 1 0 1 0 0 0 0 ...
##  $ incoming_agricultural: int  30028818 30028818 30028818 22288055 53384566 22688441 26692283 9275569
##  $ farm_expenses      : int  31363432 31363432 31363432 18751329 20731006 18907036 22243569 3697993
##  $ labor_primary      : int  0 0 0 0 1 0 0 0 1 0 ...
##  $ lasting_investment : int  28411718 28411718 28411718 7781123 20100562 4442561 22562288 33922659
##  $ no_lasting_investmen : int  28292707 28292707 28292707 69219765 43419447 76629095 55608922 546001
##  $ depressed          : int  0 1 0 0 0 0 1 0 0 0 ...
```

```r
summary(dep)
```

```
##     Survey_id        Ville_id          sex              Age
##  Min.   :   1   Min.   :  1.00   Min.   :0.0000   Min.   :17.00
##  1st Qu.: 358   1st Qu.: 24.00   1st Qu.:1.0000   1st Qu.:25.00
##  Median : 715   Median : 57.00   Median :1.0000   Median :30.00
##  Mean   : 715   Mean   : 76.29   Mean   :0.9181   Mean   :34.78
##  3rd Qu.:1072   3rd Qu.:105.00   3rd Qu.:1.0000   3rd Qu.:42.00
##  Max.   :1429   Max.   :292.00   Max.   :1.0000   Max.   :91.00
##
##     Married       Number_children  education_level  total_members
##  Min.   :0.0000   Min.   : 0.000   Min.   : 1.000   Min.   : 1.000
##  1st Qu.:1.0000   1st Qu.: 2.000   1st Qu.: 8.000   1st Qu.: 4.000
##  Median :1.0000   Median : 3.000   Median : 9.000   Median : 5.000
##  Mean   :0.7726   Mean   : 2.883   Mean   : 8.687   Mean   : 4.969
##  3rd Qu.:1.0000   3rd Qu.: 4.000   3rd Qu.:10.000   3rd Qu.: 6.000
##  Max.   :1.0000   Max.   :11.000   Max.   :19.000   Max.   :12.000
##
##   gained_asset      durable_asset       save_asset       living_expenses
##  Min.   :  325112   Min.   :  162556   Min.   :  172966   Min.   :  262919
##  1st Qu.:23269824   1st Qu.:19298521   1st Qu.:23399979   1st Qu.:20886711
```

```
##   Median :28912201   Median :22861940   Median :23399979   Median :26692283
##   Mean   :33634478   Mean   :27172957   Mean   :27424708   Mean   :32482566
##   3rd Qu.:37172832   3rd Qu.:26569498   3rd Qu.:23399979   3rd Qu.:38436887
##   Max.   :99127548   Max.   :99615601   Max.   :99926758   Max.   :99295282
##
##   other_expenses     incoming_salary  incoming_own_farm incoming_business
##   Min.   :  172966   Min.   :0.0000   Min.   :0.0000    Min.   :0.0000
##   1st Qu.:20980135   1st Qu.:0.0000   1st Qu.:0.0000    1st Qu.:0.0000
##   Median :28203066   Median :0.0000   Median :0.0000    Median :0.0000
##   Mean   :33666324   Mean   :0.1798   Mean   :0.2519    Mean   :0.1078
##   3rd Qu.:40518887   3rd Qu.:0.0000   3rd Qu.:1.0000    3rd Qu.:0.0000
##   Max.   :99823799   Max.   :1.0000   Max.   :1.0000    Max.   :1.0000
##
##   incoming_no_business incoming_agricultural farm_expenses      labor_primary
##   Min.   :0.0000       Min.   :  325112      Min.   :  271505   Min.   :0.0000
##   1st Qu.:0.0000       1st Qu.:23222287      1st Qu.:22799659   1st Qu.:0.0000
##   Median :0.0000       Median :30028818      Median :31363432   Median :0.0000
##   Mean   :0.2603       Mean   :34510389      Mean   :35491526   Mean   :0.2134
##   3rd Qu.:1.0000       3rd Qu.:40038424      3rd Qu.:43485844   3rd Qu.:0.0000
##   Max.   :1.0000       Max.   :99789095      Max.   :99651194   Max.   :1.0000
##
##   lasting_investment no_lasting_investmen   depressed
##   Min.   :   74292   Min.   :  126312      Min.   :0.0000
##   1st Qu.:20019113   1st Qu.:20642033      1st Qu.:0.0000
##   Median :28411718   Median :28292707      Median :0.0000
##   Mean   :32992215   Mean   :33603851      Mean   :0.1666
##   3rd Qu.:39826862   3rd Qu.:41517625      3rd Qu.:0.0000
##   Max.   :99446667   Max.   :99651194      Max.   :1.0000
##                      NA's   :20
```

```r
dep_cleaned = dep[!is.na(dep$no_lasting_investmen),]
summary(dep_cleaned)
```

```
##     Survey_id         Ville_id          sex              Age
##   Min.   :   1.0   Min.   :  1.00   Min.   :0.0000   Min.   :17.00
##   1st Qu.: 356.0   1st Qu.: 24.00   1st Qu.:1.0000   1st Qu.:25.00
##   Median : 717.0   Median : 57.00   Median :1.0000   Median :31.00
##   Mean   : 715.8   Mean   : 76.48   Mean   :0.9184   Mean   :34.73
##   3rd Qu.:1075.0   3rd Qu.:105.00   3rd Qu.:1.0000   3rd Qu.:42.00
##   Max.   :1429.0   Max.   :292.00   Max.   :1.0000   Max.   :91.00
##     Married        Number_children  education_level   total_members
##   Min.   :0.0000   Min.   : 0.000   Min.   : 1.000   Min.   : 1.000
##   1st Qu.:1.0000   1st Qu.: 2.000   1st Qu.: 8.000   1st Qu.: 4.000
##   Median :1.0000   Median : 3.000   Median : 9.000   Median : 5.000
##   Mean   :0.7743   Mean   : 2.905   Mean   : 8.698   Mean   : 4.996
##   3rd Qu.:1.0000   3rd Qu.: 4.000   3rd Qu.:10.000   3rd Qu.: 6.000
##   Max.   :1.0000   Max.   :11.000   Max.   :19.000   Max.   :12.000
##    gained_asset      durable_asset       save_asset        living_expenses
##   Min.   :  325112   Min.   :  162556   Min.   :  172966   Min.   :  262919
##   1st Qu.:23129761   1st Qu.:19298521   1st Qu.:23399979   1st Qu.:21033521
##   Median :28912201   Median :22861940   Median :23399979   Median :26692283
##   Mean   :33605880   Mean   :27070961   Mean   :27444528   Mean   :32486611
##   3rd Qu.:37172832   3rd Qu.:26345285   3rd Qu.:23399979   3rd Qu.:38703811
##   Max.   :99127548   Max.   :99615601   Max.   :99926758   Max.   :99295282
##   other_expenses     incoming_salary incoming_own_farm incoming_business
```
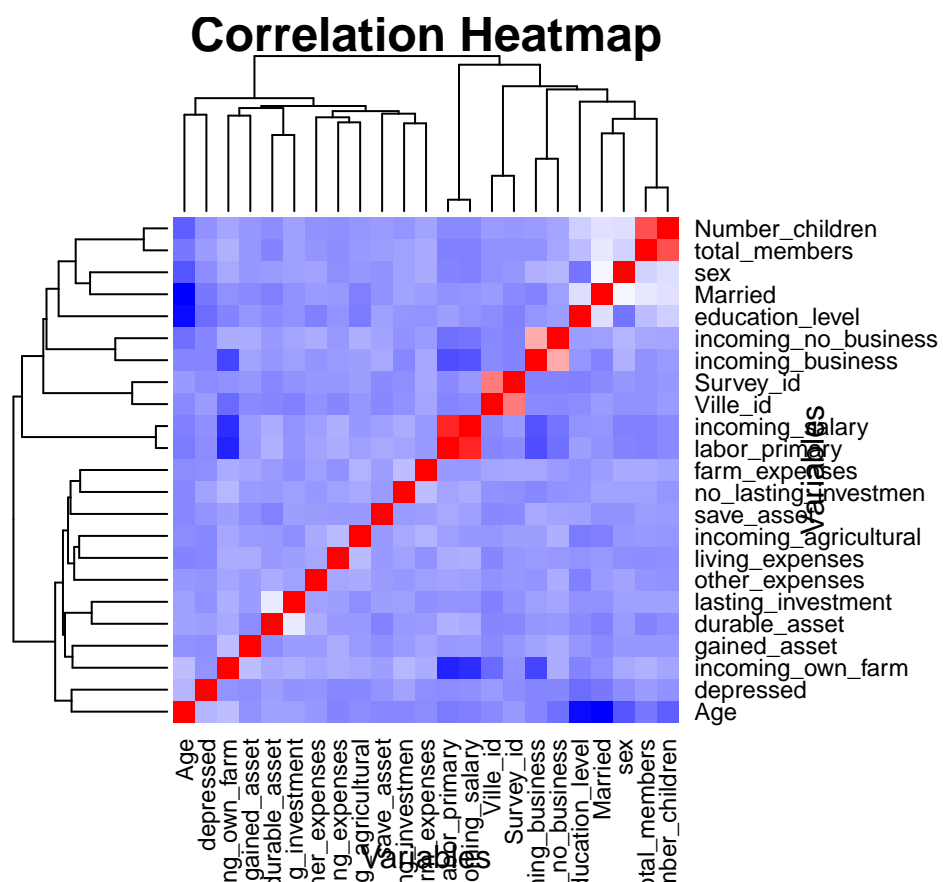
```
##  Min.   :  172966   Min.   :0.000   Min.   :0.0000    Min.    :0.0000
##  1st Qu.:21140289   1st Qu.:0.000   1st Qu.:0.0000    1st Qu.:0.0000
##  Median :28203066   Median :0.000   Median :0.0000    Median :0.0000
##  Mean   :33627939   Mean   :0.176   Mean   :0.2541    Mean    :0.1093
##  3rd Qu.:40278656   3rd Qu.:0.000   3rd Qu.:1.0000    3rd Qu.:0.0000
##  Max.   :99823799   Max.   :1.000   Max.   :1.0000    Max.    :1.0000
##  incoming_no_business incoming_agricultural farm_expenses      labor_primary
##  Min.   :0.000        Min.   :  325112      Min.   :  271505   Min.   :0.0000
##  1st Qu.:0.000        1st Qu.:22955363      1st Qu.:22399275   1st Qu.:0.0000
##  Median :0.000        Median :30028818      Median :31363432   Median :0.0000
##  Mean   :0.264        Mean   :34574003      Mean   :35550122   Mean   :0.2094
##  3rd Qu.:1.000        3rd Qu.:40038425      3rd Qu.:43997784   3rd Qu.:0.0000
##  Max.   :1.000        Max.   :99789095      Max.   :99651194   Max.   :1.0000
##  lasting_investment no_lasting_investmen   depressed
##  Min.   :   74292   Min.   :  126312     Min.   :0.0000
##  1st Qu.:20100562   1st Qu.:20642033     1st Qu.:0.0000
##  Median :28411718   Median :28292707     Median :0.0000
##  Mean   :33006117   Mean   :33603851     Mean   :0.1668
##  3rd Qu.:39784448   3rd Qu.:41517625     3rd Qu.:0.0000
##  Max.   :99446667   Max.   :99651194     Max.   :1.0000
```

Correlation Among Variables

```r
library(ggplot2)

cor_matrix = cor(dep_cleaned)

heatmap(cor_matrix,
        main = "Correlation Heatmap",
        xlab = "Variables",
        ylab = "Variables",
        col = colorRampPalette(c("blue", "white", "red"))(1000),
        scale = "none",
        margins = c(5, 5))
```
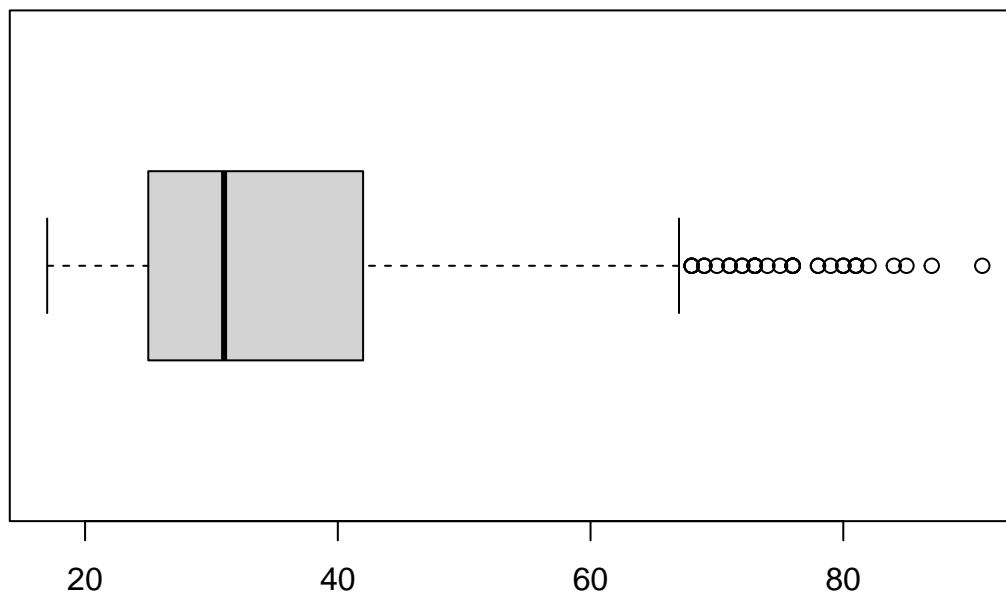
Correlation Heatmap

There is some correlation between incoming_business and education levels.

```r
boxplot(dep_cleaned$Age, horizontal = TRUE)
```



```r
#let's determine what features to be considered

features = glm(depressed~., family=binomial(link="logit"), data=dep_cleaned)
model = step(features, direction = "backward")
```

```
## Start:  AIC=1278.29
## depressed ~ Survey_id + Ville_id + sex + Age + Married + Number_children +
##     education_level + total_members + gained_asset + durable_asset +
##     save_asset + living_expenses + other_expenses + incoming_salary +
##     incoming_own_farm + incoming_business + incoming_no_business +
##     incoming_agricultural + farm_expenses + labor_primary + lasting_investment +
##     no_lasting_investmen
##
##                           Df Deviance    AIC
## - sex                      1   1232.3 1276.3
## - gained_asset             1   1232.3 1276.3
## - incoming_no_business     1   1232.3 1276.3
## - lasting_investment       1   1232.4 1276.4
## - other_expenses           1   1232.4 1276.4
## - Number_children          1   1232.4 1276.4
## - farm_expenses            1   1232.5 1276.5
## - incoming_own_farm        1   1232.6 1276.6
## - incoming_salary          1   1232.6 1276.6
## - living_expenses          1   1232.7 1276.7
## - labor_primary            1   1232.9 1276.9
## - save_asset               1   1232.9 1276.9
## - incoming_agricultural    1   1233.0 1277.0
## - incoming_business        1   1233.1 1277.1
## - Married                  1   1234.2 1278.2
## <none>                         1232.3 1278.3
## - durable_asset            1   1234.5 1278.5
## - Survey_id                1   1235.0 1279.0
## - total_members            1   1235.1 1279.1
## - Age                      1   1235.4 1279.4
## - Ville_id                 1   1235.8 1279.8
## - no_lasting_investmen     1   1236.8 1280.8
## - education_level          1   1238.0 1282.0
##
## Step:  AIC=1276.29
## depressed ~ Survey_id + Ville_id + Age + Married + Number_children +
##     education_level + total_members + gained_asset + durable_asset +
##     save_asset + living_expenses + other_expenses + incoming_salary +
##     incoming_own_farm + incoming_business + incoming_no_business +
##     incoming_agricultural + farm_expenses + labor_primary + lasting_investment +
##     no_lasting_investmen
##
##                           Df Deviance    AIC
## - gained_asset             1   1232.3 1274.3
## - incoming_no_business     1   1232.3 1274.3
## - lasting_investment       1   1232.4 1274.4
## - other_expenses           1   1232.4 1274.4
## - Number_children          1   1232.4 1274.4
## - farm_expenses            1   1232.5 1274.5
## - incoming_own_farm        1   1232.6 1274.6
## - incoming_salary          1   1232.6 1274.6
## - living_expenses          1   1232.7 1274.7
## - labor_primary            1   1232.9 1274.9
## - save_asset               1   1232.9 1274.9
## - incoming_agricultural    1   1233.0 1275.0
```

```
## - incoming_business       1   1233.1 1275.1
## - Married                 1   1234.3 1276.3
## <none>                        1232.3 1276.3
## - durable_asset           1   1234.5 1276.5
## - Survey_id               1   1235.0 1277.0
## - total_members           1   1235.1 1277.1
## - Age                     1   1235.4 1277.4
## - Ville_id                1   1235.8 1277.8
## - no_lasting_investmen    1   1236.8 1278.8
## - education_level         1   1238.4 1280.4
##
## Step:  AIC=1274.3
## depressed ~ Survey_id + Ville_id + Age + Married + Number_children +
##     education_level + total_members + durable_asset + save_asset +
##     living_expenses + other_expenses + incoming_salary + incoming_own_farm +
##     incoming_business + incoming_no_business + incoming_agricultural +
##     farm_expenses + labor_primary + lasting_investment + no_lasting_investmen
##
##                            Df Deviance    AIC
## - incoming_no_business     1   1232.3 1272.3
## - lasting_investment       1   1232.4 1272.4
## - other_expenses           1   1232.4 1272.4
## - Number_children          1   1232.4 1272.4
## - farm_expenses            1   1232.5 1272.5
## - incoming_own_farm        1   1232.6 1272.6
## - incoming_salary          1   1232.6 1272.6
## - living_expenses          1   1232.7 1272.7
## - labor_primary            1   1232.9 1272.9
## - save_asset               1   1232.9 1272.9
## - incoming_agricultural    1   1233.0 1273.0
## - incoming_business        1   1233.1 1273.1
## - Married                  1   1234.3 1274.3
## <none>                         1232.3 1274.3
## - durable_asset            1   1234.5 1274.5
## - Survey_id                1   1235.0 1275.0
## - total_members            1   1235.1 1275.1
## - Age                      1   1235.4 1275.4
## - Ville_id                 1   1235.8 1275.8
## - no_lasting_investmen     1   1236.8 1276.8
## - education_level          1   1238.4 1278.4
##
## Step:  AIC=1272.32
## depressed ~ Survey_id + Ville_id + Age + Married + Number_children +
##     education_level + total_members + durable_asset + save_asset +
##     living_expenses + other_expenses + incoming_salary + incoming_own_farm +
##     incoming_business + incoming_agricultural + farm_expenses +
##     labor_primary + lasting_investment + no_lasting_investmen
##
##                            Df Deviance    AIC
## - other_expenses           1   1232.4 1270.4
## - lasting_investment       1   1232.4 1270.4
## - Number_children          1   1232.5 1270.5
## - farm_expenses            1   1232.5 1270.5
## - incoming_salary          1   1232.6 1270.6
```

```
## - incoming_own_farm       1   1232.6 1270.6
## - living_expenses         1   1232.8 1270.8
## - labor_primary           1   1232.9 1270.9
## - save_asset              1   1232.9 1270.9
## - incoming_agricultural   1   1233.1 1271.1
## - incoming_business       1   1233.8 1271.8
## <none>                        1232.3 1272.3
## - Married                 1   1234.3 1272.3
## - durable_asset           1   1234.6 1272.6
## - Survey_id               1   1235.0 1273.0
## - total_members           1   1235.1 1273.1
## - Age                     1   1235.5 1273.5
## - Ville_id                1   1235.8 1273.8
## - no_lasting_investmen    1   1236.8 1274.8
## - education_level         1   1238.4 1276.4
##
## Step:  AIC=1270.4
## depressed ~ Survey_id + Ville_id + Age + Married + Number_children +
##     education_level + total_members + durable_asset + save_asset +
##     living_expenses + incoming_salary + incoming_own_farm + incoming_business +
##     incoming_agricultural + farm_expenses + labor_primary + lasting_investment +
##     no_lasting_investmen
##
##                           Df Deviance    AIC
## - lasting_investment      1   1232.5 1268.5
## - Number_children         1   1232.5 1268.5
## - farm_expenses           1   1232.6 1268.6
## - incoming_salary         1   1232.7 1268.7
## - incoming_own_farm       1   1232.7 1268.7
## - living_expenses         1   1232.8 1268.8
## - labor_primary           1   1233.0 1269.0
## - save_asset              1   1233.0 1269.0
## - incoming_agricultural   1   1233.1 1269.1
## - incoming_business       1   1233.8 1269.8
## - Married                 1   1234.4 1270.4
## <none>                        1232.4 1270.4
## - durable_asset           1   1234.7 1270.7
## - Survey_id               1   1235.1 1271.1
## - total_members           1   1235.2 1271.2
## - Age                     1   1235.6 1271.6
## - Ville_id                1   1235.9 1271.9
## - no_lasting_investmen    1   1236.9 1272.9
## - education_level         1   1238.5 1274.5
##
## Step:  AIC=1268.48
## depressed ~ Survey_id + Ville_id + Age + Married + Number_children +
##     education_level + total_members + durable_asset + save_asset +
##     living_expenses + incoming_salary + incoming_own_farm + incoming_business +
##     incoming_agricultural + farm_expenses + labor_primary + no_lasting_investmen
##
##                           Df Deviance    AIC
## - Number_children         1   1232.6 1266.6
## - farm_expenses           1   1232.7 1266.7
## - incoming_salary         1   1232.8 1266.8
```

```
## - incoming_own_farm      1   1232.8 1266.8
## - living_expenses        1   1232.9 1266.9
## - labor_primary          1   1233.1 1267.1
## - save_asset             1   1233.1 1267.1
## - incoming_agricultural  1   1233.2 1267.2
## - incoming_business      1   1234.0 1268.0
## - Married                1   1234.5 1268.5
## <none>                       1232.5 1268.5
## - durable_asset          1   1234.7 1268.7
## - Survey_id              1   1235.2 1269.2
## - total_members          1   1235.2 1269.2
## - Age                    1   1235.7 1269.7
## - Ville_id               1   1236.0 1270.0
## - no_lasting_investmen   1   1237.0 1271.0
## - education_level        1   1238.6 1272.6
##
## Step:  AIC=1266.61
## depressed ~ Survey_id + Ville_id + Age + Married + education_level +
##     total_members + durable_asset + save_asset + living_expenses +
##     incoming_salary + incoming_own_farm + incoming_business +
##     incoming_agricultural + farm_expenses + labor_primary + no_lasting_investmen
##
##                           Df Deviance    AIC
## - farm_expenses           1   1232.8 1264.8
## - incoming_salary         1   1232.9 1264.9
## - incoming_own_farm       1   1233.0 1265.0
## - living_expenses         1   1233.0 1265.0
## - labor_primary           1   1233.2 1265.2
## - save_asset              1   1233.2 1265.2
## - incoming_agricultural   1   1233.3 1265.3
## - incoming_business       1   1234.2 1266.2
## <none>                        1232.6 1266.6
## - Married                 1   1234.6 1266.6
## - durable_asset           1   1234.8 1266.8
## - Survey_id               1   1235.3 1267.3
## - Age                     1   1236.0 1268.0
## - Ville_id                1   1236.1 1268.1
## - total_members           1   1237.0 1269.0
## - no_lasting_investmen    1   1237.2 1269.2
## - education_level         1   1238.9 1270.9
##
## Step:  AIC=1264.77
## depressed ~ Survey_id + Ville_id + Age + Married + education_level +
##     total_members + durable_asset + save_asset + living_expenses +
##     incoming_salary + incoming_own_farm + incoming_business +
##     incoming_agricultural + labor_primary + no_lasting_investmen
##
##                           Df Deviance    AIC
## - incoming_salary         1   1233.1 1263.1
## - incoming_own_farm       1   1233.2 1263.2
## - living_expenses         1   1233.2 1263.2
## - save_asset              1   1233.4 1263.4
## - labor_primary           1   1233.4 1263.4
## - incoming_agricultural   1   1233.6 1263.6
```

```
## - incoming_business       1   1234.4 1264.4
## <none>                         1232.8 1264.8
## - Married                  1   1234.8 1264.8
## - durable_asset            1   1235.0 1265.0
## - Survey_id                1   1235.5 1265.5
## - Age                      1   1236.2 1266.2
## - Ville_id                 1   1236.2 1266.2
## - total_members            1   1237.2 1267.2
## - no_lasting_investmen     1   1237.2 1267.2
## - education_level          1   1239.1 1269.1
##
## Step:  AIC=1263.06
## depressed ~ Survey_id + Ville_id + Age + Married + education_level +
##     total_members + durable_asset + save_asset + living_expenses +
##     incoming_own_farm + incoming_business + incoming_agricultural +
##     labor_primary + no_lasting_investmen
##
##                           Df Deviance    AIC
## - incoming_own_farm        1   1233.5 1261.5
## - living_expenses          1   1233.5 1261.5
## - labor_primary            1   1233.5 1261.5
## - save_asset               1   1233.6 1261.6
## - incoming_agricultural    1   1233.9 1261.9
## - incoming_business        1   1234.7 1262.7
## <none>                         1233.1 1263.1
## - Married                  1   1235.1 1263.1
## - durable_asset            1   1235.2 1263.2
## - Survey_id                1   1235.7 1263.7
## - Ville_id                 1   1236.4 1264.4
## - Age                      1   1236.5 1264.5
## - total_members            1   1237.4 1265.4
## - no_lasting_investmen     1   1237.6 1265.6
## - education_level          1   1239.6 1267.6
##
## Step:  AIC=1261.45
## depressed ~ Survey_id + Ville_id + Age + Married + education_level +
##     total_members + durable_asset + save_asset + living_expenses +
##     incoming_business + incoming_agricultural + labor_primary +
##     no_lasting_investmen
##
##                           Df Deviance    AIC
## - labor_primary            1   1233.7 1259.7
## - living_expenses          1   1233.9 1259.9
## - save_asset               1   1233.9 1259.9
## - incoming_agricultural    1   1234.4 1260.4
## - incoming_business        1   1234.8 1260.8
## - durable_asset            1   1235.4 1261.4
## <none>                         1233.5 1261.5
## - Married                  1   1235.5 1261.5
## - Survey_id                1   1236.3 1262.3
## - Age                      1   1236.6 1262.6
## - Ville_id                 1   1237.3 1263.3
## - total_members            1   1237.6 1263.6
## - no_lasting_investmen     1   1237.7 1263.7
```

```
## - education_level         1   1240.0 1266.0
##
## Step:  AIC=1259.66
## depressed ~ Survey_id + Ville_id + Age + Married + education_level +
##     total_members + durable_asset + save_asset + living_expenses +
##     incoming_business + incoming_agricultural + no_lasting_investmen
##
##                          Df Deviance    AIC
## - save_asset              1   1234.1 1258.1
## - living_expenses         1   1234.2 1258.2
## - incoming_agricultural   1   1234.6 1258.6
## - incoming_business       1   1234.8 1258.8
## - durable_asset           1   1235.5 1259.5
## <none>                        1233.7 1259.7
## - Married                 1   1235.7 1259.7
## - Survey_id               1   1236.5 1260.5
## - Age                     1   1236.9 1260.9
## - Ville_id                1   1237.5 1261.5
## - no_lasting_investmen    1   1237.9 1261.9
## - total_members           1   1237.9 1261.9
## - education_level         1   1240.2 1264.2
##
## Step:  AIC=1258.09
## depressed ~ Survey_id + Ville_id + Age + Married + education_level +
##     total_members + durable_asset + living_expenses + incoming_business +
##     incoming_agricultural + no_lasting_investmen
##
##                          Df Deviance    AIC
## - living_expenses         1   1234.6 1256.6
## - incoming_agricultural   1   1235.0 1257.0
## - incoming_business       1   1235.2 1257.2
## - durable_asset           1   1235.9 1257.9
## <none>                        1234.1 1258.1
## - Married                 1   1236.2 1258.2
## - Survey_id               1   1237.0 1259.0
## - Age                     1   1237.3 1259.3
## - Ville_id                1   1237.9 1259.9
## - total_members           1   1238.4 1260.4
## - no_lasting_investmen    1   1238.4 1260.4
## - education_level         1   1240.6 1262.6
##
## Step:  AIC=1256.58
## depressed ~ Survey_id + Ville_id + Age + Married + education_level +
##     total_members + durable_asset + incoming_business + incoming_agricultural +
##     no_lasting_investmen
##
##                          Df Deviance    AIC
## - incoming_agricultural   1   1235.7 1255.7
## - incoming_business       1   1235.7 1255.7
## - durable_asset           1   1236.3 1256.3
## <none>                        1234.6 1256.6
## - Married                 1   1236.7 1256.7
## - Survey_id               1   1237.5 1257.5
## - Age                     1   1237.9 1257.9
```

```
## - Ville_id                1   1238.6 1258.6
## - no_lasting_investmen    1   1238.8 1258.8
## - total_members           1   1238.9 1258.9
## - education_level         1   1241.1 1261.1
##
## Step:  AIC=1255.68
## depressed ~ Survey_id + Ville_id + Age + Married + education_level +
##     total_members + durable_asset + incoming_business + no_lasting_investmen
##
##                          Df Deviance    AIC
## - incoming_business       1   1236.9 1254.9
## - durable_asset           1   1237.4 1255.4
## - Married                 1   1237.6 1255.6
## <none>                        1235.7 1255.7
## - Survey_id               1   1238.8 1256.8
## - Age                     1   1239.3 1257.3
## - no_lasting_investmen    1   1239.6 1257.6
## - Ville_id                1   1239.7 1257.7
## - total_members           1   1239.8 1257.8
## - education_level         1   1241.8 1259.8
##
## Step:  AIC=1254.9
## depressed ~ Survey_id + Ville_id + Age + Married + education_level +
##     total_members + durable_asset + no_lasting_investmen
##
##                          Df Deviance    AIC
## - durable_asset           1   1238.5 1254.5
## - Married                 1   1238.7 1254.7
## <none>                        1236.9 1254.9
## - Survey_id               1   1239.8 1255.8
## - Age                     1   1240.7 1256.7
## - Ville_id                1   1240.8 1256.8
## - no_lasting_investmen    1   1240.9 1256.9
## - total_members           1   1241.0 1257.0
## - education_level         1   1243.1 1259.1
##
## Step:  AIC=1254.51
## depressed ~ Survey_id + Ville_id + Age + Married + education_level +
##     total_members + no_lasting_investmen
##
##                          Df Deviance    AIC
## - Married                 1   1240.3 1254.3
## <none>                        1238.5 1254.5
## - Survey_id               1   1241.5 1255.5
## - Ville_id                1   1242.4 1256.4
## - total_members           1   1242.4 1256.4
## - Age                     1   1242.4 1256.4
## - no_lasting_investmen    1   1242.6 1256.6
## - education_level         1   1244.6 1258.6
##
## Step:  AIC=1254.33
## depressed ~ Survey_id + Ville_id + Age + education_level + total_members +
##     no_lasting_investmen
##
```

```
##                          Df Deviance    AIC
## <none>                      1240.3 1254.3
## - Survey_id             1   1243.2 1255.2
## - total_members         1   1243.3 1255.3
## - Ville_id              1   1244.0 1256.0
## - no_lasting_investmen  1   1244.2 1256.2
## - education_level       1   1246.9 1258.9
## - Age                   1   1247.0 1259.0
```

`summary(model)`

```
##
## Call:
## glm(formula = depressed ~ Survey_id + Ville_id + Age + education_level +
##     total_members + no_lasting_investmen, family = binomial(link = "logit"),
##     data = dep_cleaned)
##
## Coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)          -2.044e+00  4.295e-01  -4.758 1.96e-06 ***
## Survey_id            -4.090e-04  2.440e-04  -1.676  0.09370 .
## Ville_id              2.803e-03  1.454e-03   1.928  0.05382 .
## Age                   1.403e-02  5.353e-03   2.622  0.00875 **
## education_level      -6.626e-02  2.587e-02  -2.561  0.01043 *
## total_members         6.776e-02  3.909e-02   1.734  0.08299 .
## no_lasting_investmen  6.426e-09  3.221e-09   1.995  0.04601 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1270.2  on 1408  degrees of freedom
## Residual deviance: 1240.3  on 1402  degrees of freedom
## AIC: 1254.3
##
## Number of Fisher Scoring iterations: 4
```

Using Backward stepwise modeling, we were able to determine the features that contribute most to AIC, which is goodness of fit of how predictor variable explains the response variable. The variables that have high p-values do not contribute well to AIC and hence are removed improving generalizability of model and preventing overfitting.

Modeling

```
#not including survey-id as it is an identifier col

glm_model = glm(depressed ~ Age + education_level + total_members + no_lasting_investmen + durable_asse

summary(glm_model)
```

```
##
## Call:
## glm(formula = depressed ~ Age + education_level + total_members +
##     no_lasting_investmen + durable_asset, family = binomial(link = "logit"),
```

13

```
##     data = dep_cleaned)
##
## Coefficients:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -2.244e+00  4.233e-01  -5.302 1.14e-07 ***
## Age                 1.343e-02  5.363e-03   2.505   0.0123 *
## education_level    -6.516e-02  2.582e-02  -2.524   0.0116 *
## total_members       6.915e-02  3.885e-02   1.780   0.0751 .
## no_lasting_investmen 6.354e-09 3.206e-09   1.982   0.0475 *
## durable_asset        4.930e-09 3.768e-09   1.309   0.1907
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1270.2  on 1408  degrees of freedom
## Residual deviance: 1242.7  on 1403  degrees of freedom
## AIC: 1254.7
##
## Number of Fisher Scoring iterations: 4
```

We have three variables that are significant, education_level, Age, and no_lasting_investmen, Education_level variable is negatively related to the outcome of depression, as is indicated by the negative estimate of -6.480e-02, whereby people with higher education levels were less depressed than others. Age, and no_lasting_investmen were positively related to influence depression amongst individuals. Higher values of Age led to depression, possibly due to hypertension or something, lasting investment of any sort was another variable that caused depression.

```r
library(pROC)
```

```
## Warning: package 'pROC' was built under R version 4.3.1
```

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

```
##
## Attaching package: 'pROC'
```

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

```r
predicted_probs = predict(glm_model, type = "response")

# Create the ROC curve
roc_curve = roc(depressed ~ predicted_probs, data = dep_cleaned)
```
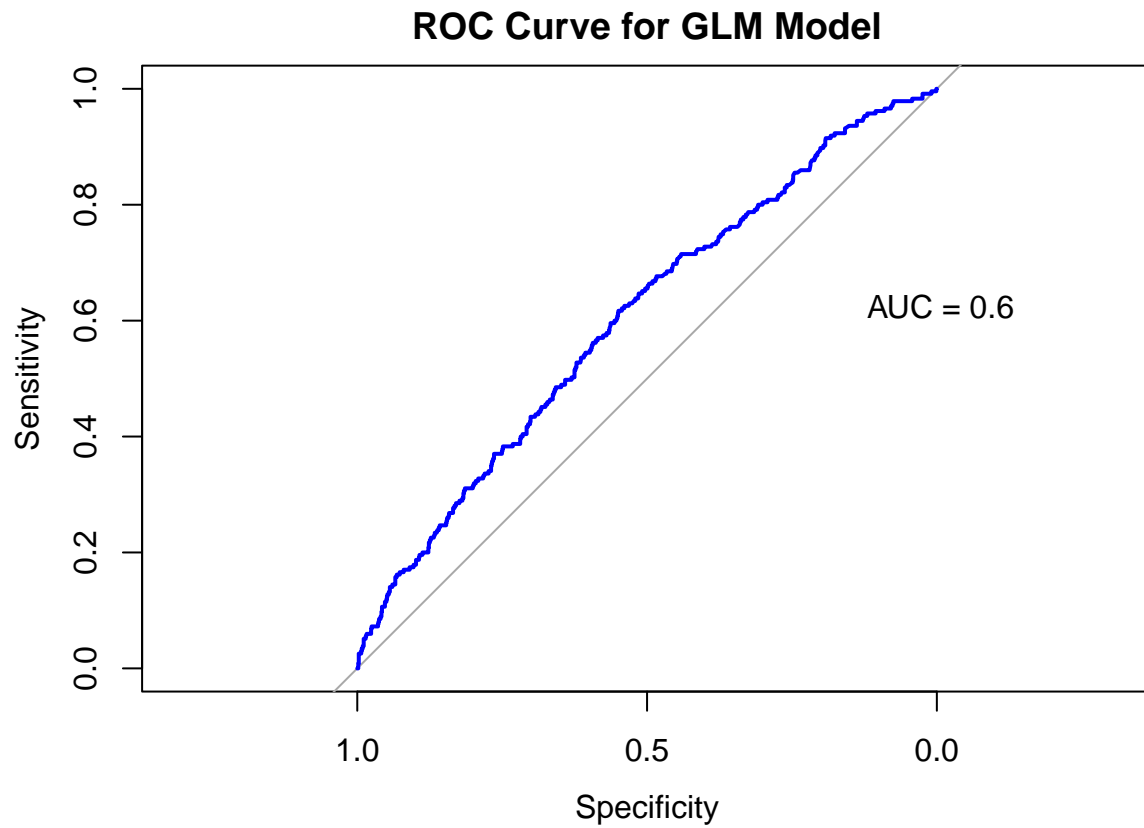
```
## Setting levels: control = 0, case = 1
```

```
## Setting direction: controls < cases
```

```r
# Plot the ROC curve
plot(roc_curve, main = "ROC Curve for GLM Model", col = "blue")
coords = coords(roc_curve, "best", ret = c("threshold", "specificity", "sensitivity"))
text(coords$threshold, coords$sensitivity, labels = paste0("AUC = ", round(auc(roc_curve),2)), pos = 4)
```

ROC Curve for GLM Model


```r
exp(glm_model$coefficients)
```

```
##        (Intercept)               Age       education_level
##          0.1059986         1.0135229             0.9369163
##      total_members no_lasting_investmen         durable_asset
##          1.0715928         1.0000000             1.0000000
```