Final Project STAC51

Alex Cheng, Zaamin Rattansi, Jacob Temple, & Jeffrey Wong

4/10/2023

Introduction

The purpose of this research is to study the direct and indirect relationships of several factors related to credit risk. This will be done through a thorough analysis of the variables that have an effect on the credit risk of an individual. Variables that are directly related and variables that have less importance were included to see a larger range of relationships possible and to really understand what what characteristics of a persons finacials have the greatest effect on their credit risk.

Background

The data we have analyzed and studied is taken from the Beuth University of applied sciences Berlin. The data was taken to analyze the importance of different factors that affect credit risk. This data was taken in the 70's so it may be outdated, for exampple, one variable is the presence of a landline. The objective of the data collected is to build a model to predict a person's credit risk. Through the several variables collected, the model will be able to predict if someone has high or low credit risk.

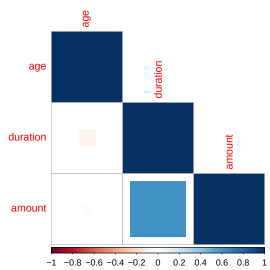
Study Goal

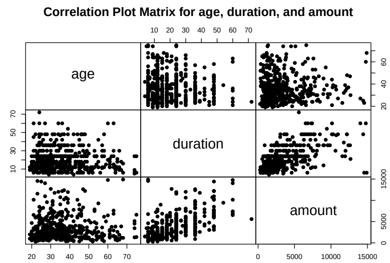
To analyze what variables cause an increase or decrease in credit risk to learn more about what is required to have low credit risk.

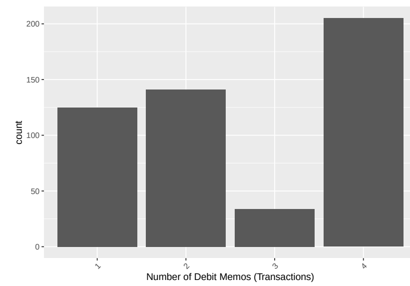
Data Preparation

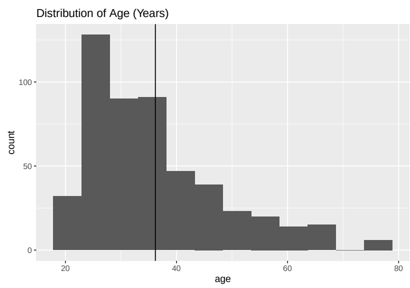
We split our data into 2 halves for cross validation. This is to prevent over-fitting of our final model.

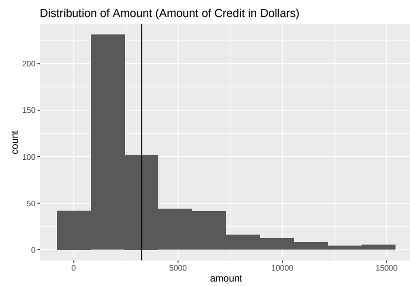
Description and Visualization of Data

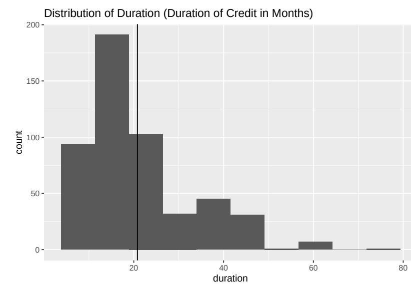












Model Building

All Terms with Interaction

Fit 1 regresses on all terms with interaction with the exception a selection of interactions with the `foreign_worker` variable due to Null values.

This initial saturated fit has an AIC of 573.2395461 and residual deviance 161.2395461.

Stepwise Regression

We completed a stepwise regression of Fit 1 to eliminate insignificant explanatory variables and increase the AIC.

Binomial Logistic

Fit 2 is a binomial logistic regression on the terms that remained from the previous stepwise regression.

Fit 2 has an AIC of 503.2176239 and residual deviance 307.2176239, making it a better fit than the initial saturated fit with regard to the AIC.

Binomial Probit

Fit 3 is a binomial probit regression of the terms that remained from the previous stepwise regression.

Fit 3 has an AIC of 502.9423147 and residual deviance 306.9423147, making it a marginally better fit than the initial saturated fit and Fit 2 with regard to the AIC and residual deviance.

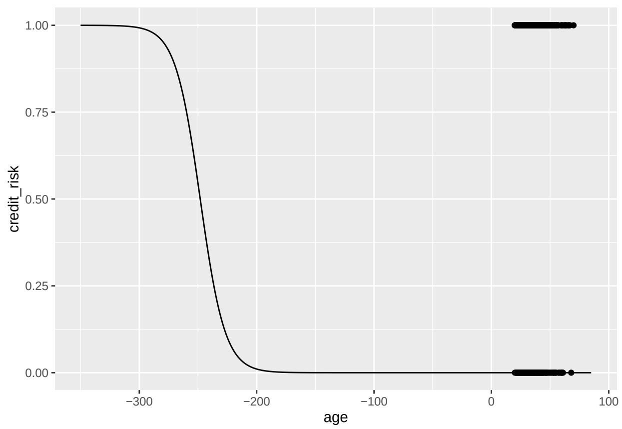
Significant Values Only ($\alpha \leq 0.05$)

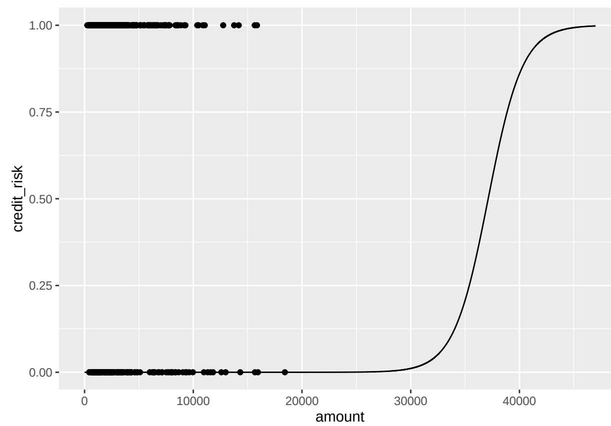
We see that Fit 3, binomial probit, has a lower AIC and is therefore a better fit than Fit 2, binomial logistic. To attempt to further improve Fit 3, we remove the insignificant values from Fit 3 to get Fit 4.

We see that the AIC in Fit 4, 518.2316563 is higher than that in Fit 3, which means Fit 3 is a better fit. This makes sense, since since we ran `stepwise(direction="both")` on Fit 1 to reduce the AIC.

Diagnoses

Predicted Probabilities





These are the predicted probability curves of age and amount. The age graph is decreasing, so as one gets older they are more likely to be at a higher risk borrower. However, we see that as the amount borrowed increases, one is at a lower credit risk.

Goodness of Fit

```
##
##  Hosmer and Lemeshow goodness of fit (GOF) test
##
## data:  fit2$y, fitted(fit2)
## X-squared = 5.7133, df = 8, p-value = 0.6793
```

```
##
##  Hosmer and Lemeshow goodness of fit (GOF) test
##
## data:  fit3$y, fitted(fit3)
## X-squared = 7.616, df = 8, p-value = 0.4718
```

The p-value for the goodness of fit test of Fit 2 is greater than 0.05 so we fail to reject H_0, however, we reject H_0 for Fit 3. We conclude that Fit 2 fits the data well.

Sensitivity and Specificity

Classification Tables:

```
##    Predicted_2
## y     0   1
##   0 139   9
##   1 210 147
```

```
##    Predicted_3
## y     0   1
##   0 139   9
##   1 211 146
```

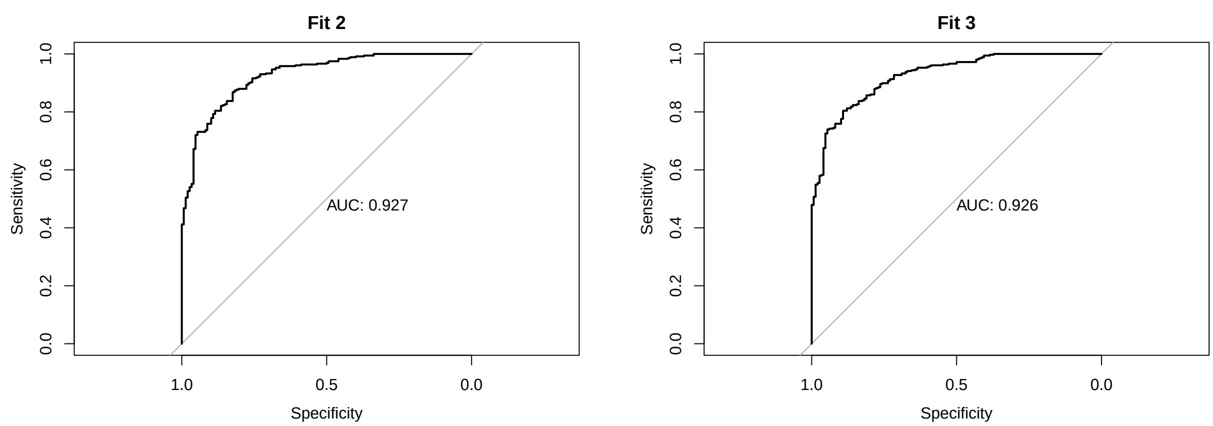
Higher sensitivity and specificity indicate a more accurate model. For Fit 2, we have:

Sensitivity = 139/(139 + 9) = 0.9391892

Specificity = 210/(147 + 210) = 0.5882353

We have similar results for Fit 3 as well. We have high sensitivity and average specificity, thus we can conclude we have a fairly good model.

ROC Curves



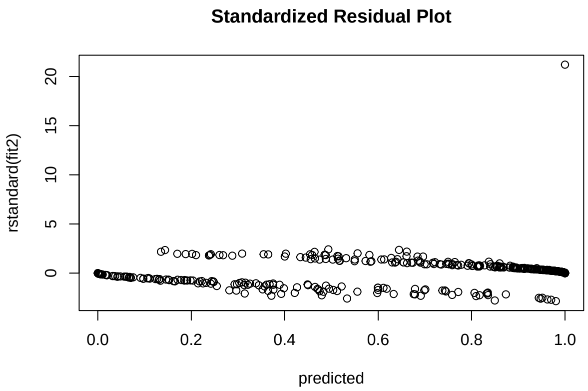
The area under a ROC curve is the concordance index c, which estimates the probability that the predictions and the outcomes are concordant, which means that the observations with the larger y also has the larger $\hat{\pi}$.

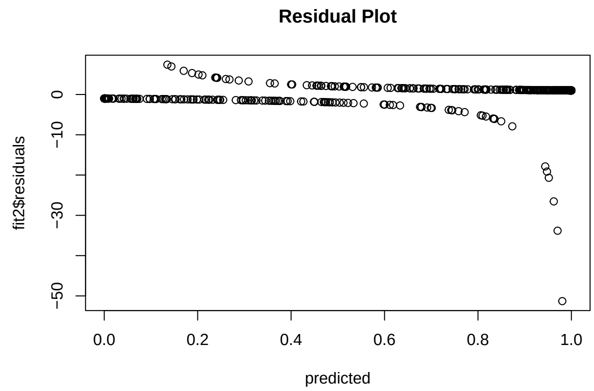
The larger the concordance index is, the better. The area under the curve for both Fit 2 and Fit 3 are very good, with value over 0.9, and hence the predictions are very close to the actual outcomes.

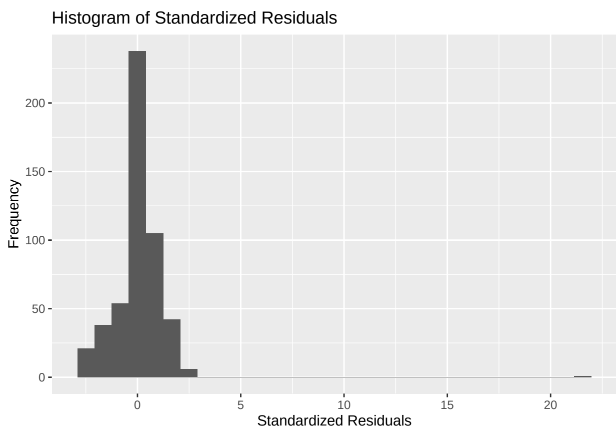
Residual Analysis

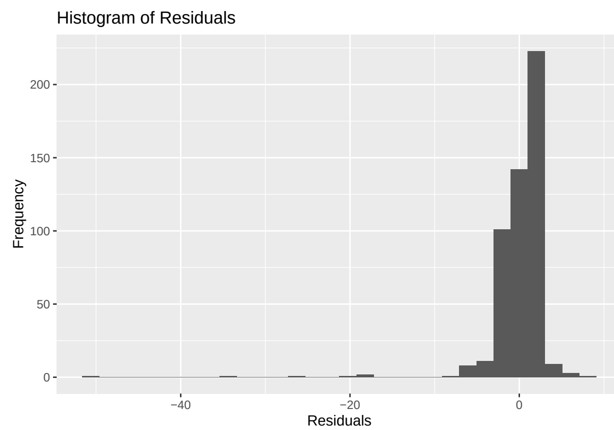
For Fit 2, we have Studentized Deviance Residual 307.2176239 with degrees of freedom 407

Below are the residuals plotted against the predicted values:









The residuals are most frequently valued around 0, with a few outliers with values less than -10 which should be removed from the data.

Conclusion

Our study resulted in few expected and unexpected findings. The variables we used allowed us to build an accurate and useful model to enable us to conclude which variables have the most and least effect on credit risk. The variable that stood out the most is `job`, the quality of the debtors job. This variable causes a huge dip in credit risk when the debtor has a high quality job where as credit risk was high when the debtor had a low quality job, this was as expected. A variable that ended up having little effect on credit risk was the number of credits variable. It was expected that this would have a significant effect on credit risk but the number of credit accounts an individual has, resulted in little change to their respective credit risk.

Final Model

Our final model is the following logistic regression:

```
##
## Call:
## glm(formula = paste("credit_risk ~", terms2, sep = " "), family = binomial(link = "logit"),
##     data = data_build)
##
## Coefficients:
##                                                Estimate Std. Error z value
```

```
## (Intercept)                                 -3.952e+01  1.453e+01  -2.719
## property                                    -3.806e+00  2.252e+00  -1.690
## housing                                      4.251e+00  2.790e+00   1.524
## number_credits                              -2.526e-01  3.424e+00  -0.074
## job                                          7.044e+00  2.644e+00   2.664
## people_liable                                1.254e+01  4.844e+00   2.589
## telephone                                    2.370e+00  3.721e+00   0.637
## credit_history                              -9.564e-01  1.668e+00  -0.573
## purpose                                      5.056e+01  1.355e+03   0.037
## amount                                       1.088e-03  7.740e-04   1.405
## savings                                     -1.262e+00  1.101e+00  -1.146
## employment_duration                          2.416e+00  1.143e+00   2.114
## installment_rate                            -6.220e-01  1.255e+00  -0.495
## personal_status_sex                          3.690e+00  1.755e+00   2.102
## other_debtors                                5.239e+00  3.672e+00   1.427
## present_residence                            1.207e+00  2.044e+00   0.590
## age                                         -2.034e-01  2.098e-01  -0.969
## other_installment_plans                      3.120e+00  2.592e+00   1.204
## status                                       8.113e-01  1.349e+00   0.601
## foreign_worker                              -1.290e+00  1.973e+00  -0.654
## property:number_credits                     -2.355e-01  4.765e-01  -0.494
## property:job                                 1.569e+00  4.140e-01   3.791
## property:people_liable                       1.588e+00  7.504e-01   2.116
## property:telephone                          -1.377e-01  4.212e-01  -0.327
## property:credit_history                      6.939e-01  2.616e-01   2.652
## property:amount                             -8.397e-05  8.059e-05  -1.042
## property:installment_rate                   -2.413e-01  2.117e-01  -1.140
## property:personal_status_sex                -4.891e-01  2.824e-01  -1.732
## property:other_installment_plans            -6.298e-01  3.079e-01  -2.045
## property:status                              7.360e-02  1.640e-01   0.449
## housing:number_credits                       1.316e+00  8.254e-01   1.594
## housing:job                                 -1.296e+00  7.333e-01  -1.768
## housing:credit_history                      -1.696e+00  4.720e-01  -3.594
## housing:installment_rate                     4.458e-01  3.822e-01   1.166
## number_credits:job                          -1.665e+00  5.551e-01  -3.000
## number_credits:people_liable                -1.670e+00  1.033e+00  -1.616
## number_credits:purpose                      -1.381e-01  1.221e-01  -1.131
## number_credits:employment_duration           5.944e-01  3.052e-01   1.948
## number_credits:installment_rate              8.243e-01  3.545e-01   2.325
## number_credits:other_installment_plans       9.104e-01  6.329e-01   1.438
## job:people_liable                           -1.973e+00  8.926e-01  -2.210
## job:purpose                                 -1.942e-01  1.068e-01  -1.818
## job:amount                                  -1.868e-04  1.030e-04  -1.814
## job:present_residence                       -5.936e-01  3.184e-01  -1.864
## job:age                                      4.787e-02  3.080e-02   1.554
## people_liable:telephone                     -1.203e+00  1.270e+00  -0.947
## people_liable:credit_history                 4.377e-01  4.989e-01   0.877
## people_liable:amount                        -1.172e-06  1.911e-04  -0.006
## people_liable:present_residence             -1.212e+00  7.620e-01  -1.591
## people_liable:age                            1.729e-02  7.004e-02   0.247
## people_liable:other_installment_plans       -7.251e-01  7.546e-01  -0.961
## people_liable:status                         1.713e-01  4.524e-01   0.379
## telephone:credit_history                     8.546e-01  4.377e-01   1.953
## telephone:amount                            -4.079e-04  1.442e-04  -2.829
```

```
## telephone:savings                              2.341e-01  2.741e-01   0.854
## telephone:employment_duration                 -7.064e-01  3.781e-01  -1.868
## telephone:other_debtors                        9.341e-01  9.334e-01   1.001
## telephone:age                                  8.648e-02  4.983e-02   1.735
## credit_history:installment_rate               -6.473e-01  2.104e-01  -3.076
## credit_history:other_debtors                   6.173e-01  4.648e-01   1.328
## credit_history:present_residence               2.881e-01  1.788e-01   1.611
## credit_history:other_installment_plans         5.209e-01  2.776e-01   1.876
## purpose:savings                                1.338e-01  6.155e-02   2.174
## purpose:age                                   -1.138e-02  6.301e-03  -1.806
## purpose:other_installment_plans               -4.201e-01  1.306e-01  -3.216
## amount:savings                                -2.653e-06  4.137e-05  -0.064
## amount:employment_duration                    -1.085e-05  6.175e-05  -0.176
## amount:installment_rate                        4.048e-05  6.596e-05   0.614
## amount:personal_status_sex                     1.063e-04  1.140e-04   0.932
## amount:present_residence                       5.091e-05  8.005e-05   0.636
## amount:age                                    -1.526e-05  9.395e-06  -1.624
## amount:other_installment_plans                 2.817e-05  1.068e-04   0.264
## savings:other_debtors                          6.506e-01  8.176e-01   0.796
## savings:present_residence                      3.885e-01  1.173e-01   3.312
## savings:age                                   -8.458e-03  1.174e-02  -0.720
## savings:other_installment_plans               -4.065e-01  1.617e-01  -2.514
## employment_duration:other_debtors             -1.501e+00  5.150e-01  -2.914
## employment_duration:present_residence          1.050e-01  1.555e-01   0.675
## employment_duration:other_installment_plans -3.440e-01  2.511e-01  -1.370
## employment_duration:status                     2.078e-01  1.409e-01   1.475
## installment_rate:personal_status_sex           2.112e-01  2.426e-01   0.871
## installment_rate:other_debtors                 8.613e-02  4.528e-01   0.190
## installment_rate:present_residence             4.586e-02  1.800e-01   0.255
## installment_rate:age                          -3.462e-02  2.056e-02  -1.684
## installment_rate:status                        2.466e-01  1.421e-01   1.736
## personal_status_sex:age                       -3.017e-02  2.325e-02  -1.297
## personal_status_sex:other_installment_plans -1.054e+00  4.199e-01  -2.511
## other_debtors:age                              1.089e-01  6.394e-02   1.703
## other_debtors:status                          -1.029e-01  3.494e-01  -0.294
## present_residence:age                         -2.978e-02  1.866e-02  -1.596
## present_residence:other_installment_plans      6.574e-01  2.875e-01   2.287
## age:other_installment_plans                    7.556e-02  3.148e-02   2.401
## purpose:foreign_worker                        -2.423e+01  6.774e+02  -0.036
## people_liable:other_debtors                   -2.571e+00  1.418e+00  -1.813
## property:other_debtors                        -1.276e+00  5.340e-01  -2.390
## job:status                                    -4.851e-01  2.715e-01  -1.787
## savings:personal_status_sex                    2.564e-01  1.672e-01   1.533
## telephone:other_installment_plans             -7.846e-01  7.197e-01  -1.090
##                                               Pr(>|z|)
## (Intercept)                                   0.006539 **
## property                                      0.090983 .
## housing                                       0.127524
## number_credits                                0.941202
## job                                           0.007727 **
## people_liable                                 0.009632 **
## telephone                                     0.524140
## credit_history                                0.566421
## purpose                                       0.970230
```

```
## amount                                      0.159909
## savings                                     0.251748
## employment_duration                         0.034476 *
## installment_rate                            0.620306
## personal_status_sex                         0.035559 *
## other_debtors                               0.153627
## present_residence                           0.554890
## age                                         0.332316
## other_installment_plans                     0.228640
## status                                      0.547572
## foreign_worker                              0.513271
## property:number_credits                     0.621047
## property:job                                0.000150 ***
## property:people_liable                      0.034369 *
## property:telephone                          0.743668
## property:credit_history                     0.007994 **
## property:amount                             0.297401
## property:installment_rate                   0.254371
## property:personal_status_sex                0.083324 .
## property:other_installment_plans            0.040832 *
## property:status                             0.653704
## housing:number_credits                      0.110916
## housing:job                                 0.077080 .
## housing:credit_history                      0.000326 ***
## housing:installment_rate                    0.243435
## number_credits:job                          0.002704 **
## number_credits:people_liable                0.106024
## number_credits:purpose                      0.258198
## number_credits:employment_duration          0.051434 .
## number_credits:installment_rate             0.020066 *
## number_credits:other_installment_plans      0.150301
## job:people_liable                           0.027089 *
## job:purpose                                 0.069068 .
## job:amount                                  0.069692 .
## job:present_residence                       0.062304 .
## job:age                                     0.120150
## people_liable:telephone                     0.343693
## people_liable:credit_history                0.380347
## people_liable:amount                        0.995107
## people_liable:present_residence             0.111622
## people_liable:age                           0.805016
## people_liable:other_installment_plans       0.336560
## people_liable:status                        0.705044
## telephone:credit_history                    0.050844 .
## telephone:amount                            0.004673 **
## telephone:savings                           0.393061
## telephone:employment_duration               0.061704 .
## telephone:other_debtors                     0.316937
## telephone:age                               0.082696 .
## credit_history:installment_rate             0.002099 **
## credit_history:other_debtors                0.184192
## credit_history:present_residence            0.107164
## credit_history:other_installment_plans      0.060599 .
## purpose:savings                             0.029679 *
```

```
## purpose:age                               0.070874 .
## purpose:other_installment_plans           0.001299 **
## amount:savings                            0.948860
## amount:employment_duration                0.860529
## amount:installment_rate                   0.539441
## amount:personal_status_sex                0.351217
## amount:present_residence                  0.524817
## amount:age                                0.104374
## amount:other_installment_plans            0.792026
## savings:other_debtors                     0.426178
## savings:present_residence                 0.000925 ***
## savings:age                               0.471222
## savings:other_installment_plans           0.011945 *
## employment_duration:other_debtors         0.003567 **
## employment_duration:present_residence     0.499818
## employment_duration:other_installment_plans 0.170652
## employment_duration:status                0.140309
## installment_rate:personal_status_sex      0.383973
## installment_rate:other_debtors            0.849134
## installment_rate:present_residence        0.798952
## installment_rate:age                      0.092221 .
## installment_rate:status                   0.082597 .
## personal_status_sex:age                   0.194460
## personal_status_sex:other_installment_plans 0.012048 *
## other_debtors:age                         0.088569 .
## other_debtors:status                      0.768411
## present_residence:age                     0.110525
## present_residence:other_installment_plans  0.022214 *
## age:other_installment_plans               0.016369 *
## purpose:foreign_worker                    0.971464
## people_liable:other_debtors               0.069759 .
## property:other_debtors                    0.016827 *
## job:status                                0.073951 .
## savings:personal_status_sex               0.125234
## telephone:other_installment_plans         0.275640
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 610.93  on 504  degrees of freedom
## Residual deviance: 307.22  on 407  degrees of freedom
## AIC: 503.22
##
## Number of Fisher Scoring iterations: 16
```