Assignment #4 - Nikola Metes

```r
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------------- tidyverse 1.2.1 --
```

```
## v ggplot2 3.1.1       v purrr   0.3.2
## v tibble  2.1.1       v dplyr   0.8.0.1
## v tidyr   0.8.3       v stringr 1.4.0
## v readr   1.3.1       v forcats 0.4.0
```

```
## -- Conflicts ------------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(car)
```

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

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

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

```
## The following object is masked from 'package:purrr':
##
##     some
```

```r
library(pROC)
```

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

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

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

```r
# load the data set
GermanCredit <- read_csv(file.path("Data", "GermanCredit_modified_SP19_001.csv")) %>%
  mutate(good=factor(if_else(Class=="Good",1,0),levels = c(0,1), labels = c("Bad", "Good"))) %>%
  select(-Class)
```

```
## Parsed with column specification:
## cols(
##   .default = col_character(),
##   Duration = col_double(),
##   Amount = col_double(),
##   InstallmentRatePercentage = col_double(),
##   ResidenceDuration = col_double(),
##   Age = col_double(),
##   NumberExistingCredits = col_double(),
##   NumberPeopleMaintenance = col_double(),
##   Telephone = col_double(),
##   ForeignWorker = col_double()
## )

## See spec(...) for full column specifications.
```

```
# Split data into train and test
 set.seed(737900)

# set an index to split the data set
#
# Create the train data frame
s <- sample(nrow(GermanCredit), replace=FALSE, size = .8*nrow(GermanCredit))
trainDF <- GermanCredit[s,]
#
# Create the test data frame
testDF <- GermanCredit[-s,]
#
```

2. Fit a logistic regression model in R

Fit a logistic regression model to predict the Class variable using all of the predictors in trainDF and assign the fitted model to the object logit.fit1.

MISSING CODE

```
str(GermanCredit)
```

```
## Classes 'spec_tbl_df', 'tbl_df', 'tbl' and 'data.frame': 1000 obs. of  21 variables:
##  $ Duration                 : num  6 48 12 42 24 36 24 36 12 30 ...
##  $ Amount                   : num  1169 5951 2096 7882 4870 ...
##  $ InstallmentRatePercentage: num  4 2 2 2 3 2 3 2 2 4 ...
##  $ ResidenceDuration        : num  4 2 3 4 4 4 4 2 4 2 ...
##  $ Age                      : num  67 22 49 45 53 35 53 35 61 28 ...
##  $ NumberExistingCredits    : num  2 1 1 1 2 1 1 1 1 2 ...
##  $ NumberPeopleMaintenance  : num  1 1 2 2 2 2 1 1 1 1 ...
##  $ Telephone                : num  0 1 1 1 1 0 1 0 1 1 ...
##  $ ForeignWorker            : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ Checking                 : chr  "lt.0" "0.to.200" "None" "lt.0" ...
##  $ Credit.History           : chr  "Critical" "PaidDuly" "Critical" "PaidDuly" ...
##  $ Loan.Purpose             : chr  "Radio.Television" "Radio.Television" "Education" "Furniture" ...
##  $ Savings                  : chr  "Unknown" "lt.100" "lt.100" "lt.100" ...
```

```
##  $ Employment.Duration      : chr  "gt.7" "1.to.4" "4.to.7" "4.to.7" ...
##  $ Personal.Status          : chr  "Single" "NotSingle" "Single" "Single" ...
##  $ Other.Debtors            : chr  "None" "None" "None" "Guarantor" ...
##  $ Property                 : chr  "RealEstate" "RealEstate" "RealEstate" "Insurance" ...
##  $ OtherInstallmentPlans    : chr  "None" "None" "None" "None" ...
##  $ Housing                  : chr  "Own" "Own" "Own" "ForFree" ...
##  $ Job.Type                 : chr  "SkilledEmployee" "SkilledEmployee" "UnskilledResident" "SkilledEm
##  $ good                     : Factor w/ 2 levels "Bad","Good": 2 1 2 2 1 2 2 2 2 1 ...
```

```
trainDF %>% select(good) %>% table()
```

```
## .
##  Bad Good
##  234  566
```

```
logit.fit1 <- glm(good~.,family=binomial,data=trainDF)
Anova(logit.fit1)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: good
##                           LR Chisq Df Pr(>Chisq)
## Duration                     6.534  1  0.0105814 *
## Amount                       4.953  1  0.0260525 *
## InstallmentRatePercentage    5.808  1  0.0159524 *
## ResidenceDuration            0.001  1  0.9728939
## Age                          1.292  1  0.2555890
## NumberExistingCredits        3.322  1  0.0683581 .
## NumberPeopleMaintenence      0.671  1  0.4128042
## Telephone                    2.319  1  0.1278036
## ForeignWorker                3.133  1  0.0767030 .
## Checking                    65.952  3  3.139e-14 ***
## Credit.History              21.026  4  0.0003129 ***
## Loan.Purpose                31.333  9  0.0002595 ***
## Savings                     16.870  4  0.0020490 **
## Employment.Duration          2.111  4  0.7153506
## Personal.Status              5.543  3  0.1360603
## Other.Debtors                5.788  2  0.0553489 .
## Property                     3.256  3  0.3537835
## OtherInstallmentPlans        6.546  2  0.0378842 *
## Housing                      2.176  2  0.3368527
## Job.Type                     1.286  3  0.7325137
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(logit.fit1)
```

```
##
## Call:
## glm(formula = good ~ ., family = binomial, data = trainDF)
##
## Deviance Residuals:
```

```
##     Min       1Q   Median       3Q      Max  
## -2.6881  -0.6990   0.3514   0.6981   2.1821  
## 
## Coefficients:
##                                Estimate Std. Error z value Pr(>|z|)    
## (Intercept)                   4.496e+00  1.537e+00   2.926 0.003436 ** 
## Duration                     -2.786e-02  1.092e-02  -2.551 0.010746 *  
## Amount                       -1.089e-04  4.927e-05  -2.211 0.027015 *  
## InstallmentRatePercentage    -2.364e-01  9.911e-02  -2.386 0.017052 *  
## ResidenceDuration            -3.296e-03  9.701e-02  -0.034 0.972894    
## Age                           1.164e-02  1.031e-02   1.129 0.259026    
## NumberExistingCredits        -4.120e-01  2.290e-01  -1.800 0.071903 .  
## NumberPeopleMaintenance      -2.355e-01  2.864e-01  -0.822 0.411004    
## Telephone                    -3.474e-01  2.293e-01  -1.515 0.129816    
## ForeignWorker                -1.062e+00  6.432e-01  -1.652 0.098567 .  
## Checkinggt.200                5.679e-01  4.176e-01   1.360 0.173895    
## Checkinglt.0                 -2.803e-01  2.414e-01  -1.161 0.245617    
## CheckingNone                  1.630e+00  2.701e-01   6.035 1.59e-09 ***
## Credit.HistoryDelay          -4.886e-01  3.782e-01  -1.292 0.196376    
## Credit.HistoryNoCredit.AllPaid -1.356e+00  4.858e-01  -2.791 0.005251 ** 
## Credit.HistoryPaidDuly       -1.002e+00  3.001e-01  -3.338 0.000845 ***
## Credit.HistoryThisBank.AllPaid -1.881e+00  5.170e-01  -3.638 0.000274 ***
## Loan.PurposeDomesticAppliance  -2.924e-01  9.124e-01  -0.320 0.748591    
## Loan.PurposeEducation        -1.155e+00  5.168e-01  -2.235 0.025385 *  
## Loan.PurposeFurniture        -5.517e-02  4.114e-01  -0.134 0.893306    
## Loan.PurposeNewCar           -9.639e-01  3.917e-01  -2.461 0.013852 *  
## Loan.PurposeOther             4.250e-01  8.301e-01   0.512 0.608613    
## Loan.PurposeRadio.Television  -1.927e-01  3.941e-01  -0.489 0.624961    
## Loan.PurposeRepairs          -6.323e-01  6.666e-01  -0.948 0.342876    
## Loan.PurposeRetraining        7.127e-01  1.272e+00   0.560 0.575332    
## Loan.PurposeUsedCar           8.808e-01  5.302e-01   1.661 0.096659 .  
## Savings500.to.1000            3.580e-01  5.667e-01   0.632 0.527499    
## Savingsgt.1000                1.134e+00  6.494e-01   1.746 0.080875 .  
## Savingslt.100                -2.739e-01  3.283e-01  -0.834 0.404137    
## SavingsUnknown                6.971e-01  4.071e-01   1.712 0.086823 .  
## Employment.Duration4.to.7     3.231e-01  3.062e-01   1.055 0.291363    
## Employment.Durationgt.7      -2.233e-04  2.840e-01  -0.001 0.999373    
## Employment.Durationlt.1      -1.582e-01  2.776e-01  -0.570 0.568735    
## Employment.DurationUnemployed  -1.196e-01  4.589e-01  -0.261 0.794360    
## Personal.StatusMarried.Widowed  5.750e-01  5.055e-01   1.137 0.255349    
## Personal.StatusNotSingle      1.202e-01  4.156e-01   0.289 0.772446    
## Personal.StatusSingle         5.955e-01  4.090e-01   1.456 0.145399    
## Other.DebtorsGuarantor        1.033e+00  6.521e-01   1.585 0.113016    
## Other.DebtorsNone            -1.224e-02  4.970e-01  -0.025 0.980360    
## PropertyInsurance            -6.039e-02  2.636e-01  -0.229 0.818813    
## PropertyRealEstate            1.258e-01  2.721e-01   0.463 0.643667    
## PropertyUnknown              -7.174e-01  4.427e-01  -1.621 0.105121    
## OtherInstallmentPlansNone     5.230e-01  2.803e-01   1.866 0.062071 .  
## OtherInstallmentPlansStores  -3.079e-01  4.571e-01  -0.674 0.500558    
## HousingOwn                   -4.041e-01  4.987e-01  -0.810 0.417728    
## HousingRent                  -7.032e-01  5.324e-01  -1.321 0.186578    
## Job.TypeSkilledEmployee       5.843e-02  3.291e-01   0.178 0.859077    
## Job.TypeUnemployedUnskilled   8.167e-01  7.747e-01   1.054 0.291790    
## Job.TypeUnskilledResident     1.643e-01  3.997e-01   0.411 0.681023    
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 967.00  on 799  degrees of freedom
## Residual deviance: 704.44  on 751  degrees of freedom
## AIC: 802.44
##
## Number of Fisher Scoring iterations: 5
```

```r
967-704.44
```

```
## [1] 262.56
```

```r
pchisq(262.56,df=799-751,lower.tail = FALSE)
```

```
## [1] 2.369794e-31
```

a. Comment on the model's Residual deviance as compared to both the degrees of freedom and the Null deviance. Is this a "good" model for the prediction of Class based on these statistics alone?

Residual deviance is less than the Null deviance. We can see a siginificant decrease in the Null deviance. The p value is 2.369794e-31. The model is yes, since the p value is close to zero.

b. Which of the coefficients are most significant?

The coefficients marked with (***) are most significant: CheckingNone Credit.HistoryPaidDuly Credit.HistoryThisBank.AllPaid

```r
table(GermanCredit$Credit.History)
```

```
##
##        Critical           Delay NoCredit.AllPaid        PaidDuly
##             293              88               40             530
## ThisBank.AllPaid
##              49
```

```r
table(GermanCredit$Checking)
```

```
##
## 0.to.200   gt.200     lt.0     None
##      269       63      274      394
```

c. Interpret, in plain english, the Duration and Amount coefficients. How do they effect our prediction of the Class variable.

For every increase in one unit of Duration, the outcome of Good (Class) is multiplied by -2.786e-02. For every increase in one unit of Amount, the outcome of Good (Class) is multiplied by -1.089e-04. Both coeficients decrease the prediction of the Class variable.

d. Interpret, in plain english, the Intercept coefficient of this model. Remember that the Intercept in logistic regression is subject to the same interpretation of factor variables as linear regression.

When all the continuous variables are zero and the factor variables are their reference, then the predicted log odds (4.496e+00) is the intercept.

3.1 Confusion Matrix: Train

```
log.50 <- logit.fit1$fitted.values
log.50[log.50>=0.5] <- 1
log.50[log.50<0.5] <- 0
```

Create factor vectors

```
actual <- trainDF$good
predicted <- factor(log.50, levels = c(0,1), labels = c("Bad", "Good"))
```

```
# Print the confusion matrix
table(actual, predicted)
```

```
##       predicted
## actual Bad Good
##   Bad  125  109
##   Good  57  509
```

```
round(prop.table(table(actual, predicted),1),2)
```

```
##       predicted
## actual  Bad Good
##   Bad  0.53 0.47
##   Good 0.10 0.90
```

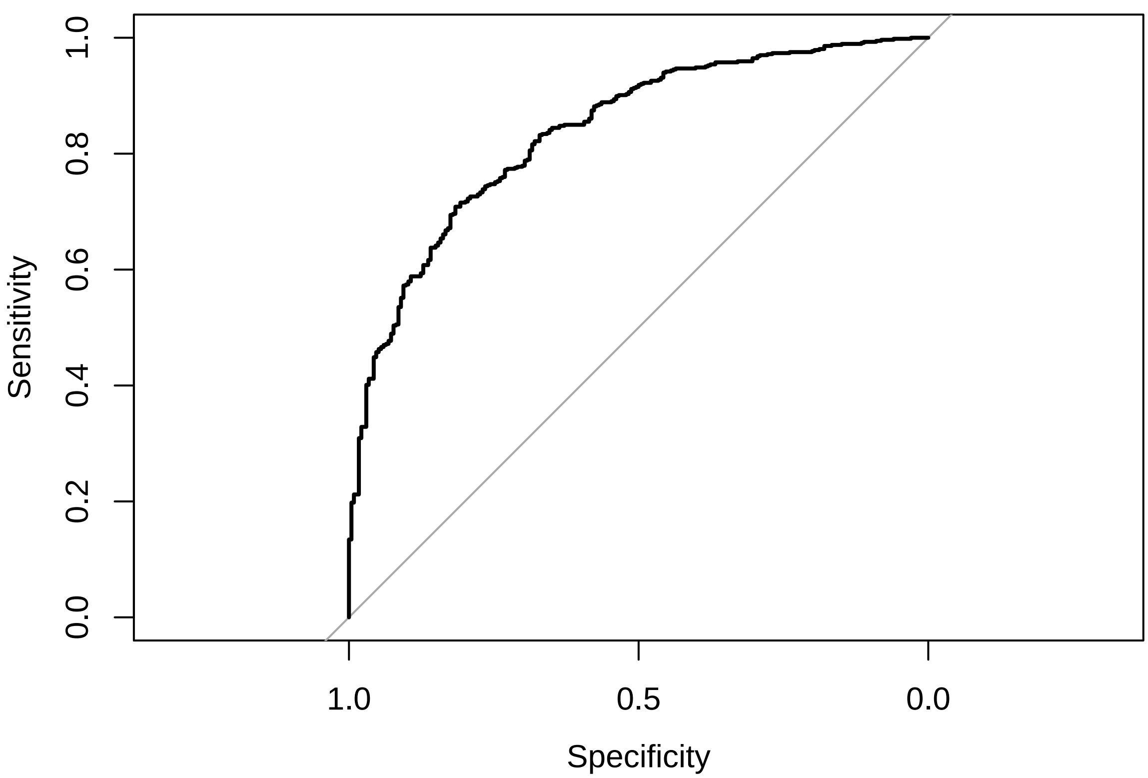
a. What is the specificity and sensitivity of this model on the train data set?

Sensitivity - ability to detect an outcome. In this case that would be to which degree we can predict Good. Specificity - ability to detect when there is not an ability to detect and outcome (i.e. a car accident) can't differentiate between if somebody has it or not, or how many NO items are we going to guess correctly.

b. Is this a good model at a .5 threshold? HINT: Do you think this institution would rather accurately predict cases of Good credit or cases of Bad credit?

They may want to vary that threshold number in order to see if the predicition of Good/Bad credit changes. Given that this model estimates the Good outcome fairly strongly, the threshold could be lowered as our degree of predicition will most likely not change much.

```
plot(roc(trainDF$good, logit.fit1$fitted.values))
```




```
auc(trainDF$good, logit.fit1$fitted.values)
```

```
## Area under the curve: 0.8372
```

Area under the curve: 0.8372

a. What does the above output from the ROC curve tell you about this model?

If you can pair two people and personA scores higher on the model than personB, they should both be rated as the same and/or personA should be rated as good. If it opposite then it would be negativ and personA should be rated as Bad.

b. Does this change your interpretation of this being a good model?

The enlarged area under the ROC curve indicates that the mdodel is good since 0.83 is a pretty good result.

4.1 Confusion Matrix: Test

```
log.test <- predict(logit.fit1, newdata = testDF, type = "response")
log.test[log.test>=0.5] <- 1
log.test[log.test<0.5] <- 0
```

Create factor vectors

```
actual <- testDF$good
predicted <- factor(log.test, levels = c(0,1), labels = c("Bad", "Good"))
```

```
# Print the confusion matrix
table(actual, predicted)
```

```
##       predicted
## actual Bad Good
##   Bad   36   30
##   Good  18  116
```

```
round(prop.table(table(actual, predicted),1),2)
```

```
##       predicted
## actual  Bad Good
##   Bad  0.55 0.45
##   Good 0.13 0.87
```

a. How well did your model perform against the holdout dataset?

The prediction of Good went down down from .90 to .87. This would be classified as TRUE POSITIVES - the measured proportion of actual positives that are correctly identified as such.

5. Improved Model

```
logit.fit1 <- glm(good~.,family=binomial,data=trainDF)
Anova(logit.fit1)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: good
##                           LR Chisq Df Pr(>Chisq)
## Duration                     6.534  1  0.0105814 *
## Amount                       4.953  1  0.0260525 *
## InstallmentRatePercentage    5.808  1  0.0159524 *
## ResidenceDuration            0.001  1  0.9728939
## Age                          1.292  1  0.2555890
## NumberExistingCredits        3.322  1  0.0683581 .
## NumberPeopleMaintenance      0.671  1  0.4128042
## Telephone                    2.319  1  0.1278036
## ForeignWorker                3.133  1  0.0767030 .
## Checking                    65.952  3  3.139e-14 ***
## Credit.History              21.026  4  0.0003129 ***
## Loan.Purpose                31.333  9  0.0002595 ***
## Savings                     16.870  4  0.0020490 **
## Employment.Duration          2.111  4  0.7153506
## Personal.Status              5.543  3  0.1360603
## Other.Debtors                5.788  2  0.0553489 .
## Property                     3.256  3  0.3537835
## OtherInstallmentPlans        6.546  2  0.0378842 *
## Housing                      2.176  2  0.3368527
## Job.Type                     1.286  3  0.7325137
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
logit.fit1 <- trainDF %>%
  select(-ResidenceDuration) %>%
  glm(good~.,family=binomial,data=.)
Anova(logit.fit1)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: good
##                           LR Chisq Df Pr(>Chisq)
## Duration                     6.559  1  0.0104379 *
## Amount                       4.952  1  0.0260554 *
## InstallmentRatePercentage    5.813  1  0.0159095 *
## Age                          1.302  1  0.2538610
## NumberExistingCredits        3.341  1  0.0675706 .
## NumberPeopleMaintenance      0.671  1  0.4127156
## Telephone                    2.330  1  0.1269239
## ForeignWorker                3.141  1  0.0763547 .
## Checking                    66.082  3  2.943e-14 ***
## Credit.History              21.026  4  0.0003130 ***
```

```
## Loan.Purpose                31.334  9  0.0002594 ***
## Savings                     16.881  4  0.0020389 **
## Employment.Duration          2.118  4  0.7141020
## Personal.Status              5.564  3  0.1348337
## Other.Debtors                5.799  2  0.0550630 .
## Property                     3.263  3  0.3528416
## OtherInstallmentPlans        6.547  2  0.0378678 *
## Housing                      2.216  2  0.3302079
## Job.Type                     1.285  3  0.7327659
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
logit.fit1 <- trainDF %>%
  select(-ResidenceDuration,-Age) %>%
  glm(good~.,family=binomial,data=.)
Anova(logit.fit1)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: good
##                           LR Chisq Df Pr(>Chisq)
## Duration                     6.711  1  0.0095808 **
## Amount                       4.985  1  0.0255661 *
## InstallmentRatePercentage    5.754  1  0.0164496 *
## NumberExistingCredits        3.441  1  0.0636128 .
## NumberPeopleMaintenance      0.555  1  0.4564729
## Telephone                    2.649  1  0.1036261
## ForeignWorker                3.257  1  0.0711127 .
## Checking                    66.747  3  2.121e-14 ***
## Credit.History              21.827  4  0.0002170 ***
## Loan.Purpose                30.689  9  0.0003347 ***
## Savings                     17.494  4  0.0015490 **
## Employment.Duration          2.284  4  0.6837069
## Personal.Status              5.372  3  0.1464996
## Other.Debtors                6.161  2  0.0459364 *
## Property                     3.260  3  0.3532993
## OtherInstallmentPlans        6.293  2  0.0430049 *
## Housing                      2.956  2  0.2280999
## Job.Type                     1.326  3  0.7229485
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
logit.fit1 <- testDF %>%
  select (-ResidenceDuration, -Age, -Job.Type) %>%
  glm(good~.,family=binomial,data=.)
Anova(logit.fit1)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: good
##                           LR Chisq Df Pr(>Chisq)
## Duration                    4.4232  1  0.0354545 *
## Amount                      0.9494  1  0.3298826
```

```
## InstallmentRatePercentage  11.2264  1  0.0008064 ***
## NumberExistingCredits       0.1192  1  0.7298574
## NumberPeopleMaintenance     0.0185  1  0.8919388
## Telephone                   0.0970  1  0.7554784
## ForeignWorker               2.7862  1  0.0950774 .
## Checking                   11.2836  3  0.0102873 *
## Credit.History              5.4918  4  0.2404517
## Loan.Purpose               16.7714  8  0.0325791 *
## Savings                     5.1257  4  0.2746414
## Employment.Duration        11.4967  4  0.0215138 *
## Personal.Status             7.5234  3  0.0569602 .
## Other.Debtors               3.4888  2  0.1747454
## Property                    3.7903  3  0.2850205
## OtherInstallmentPlans       6.5854  2  0.0371527 *
## Housing                     4.8559  2  0.0882193 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
logit.fit1 <- testDF %>%
  select (-ResidenceDuration, - Age, -Job.Type, -NumberPeopleMaintenance) %>%
  glm(good~.,family=binomial,data=.)
Anova(logit.fit1)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: good
##                           LR Chisq Df Pr(>Chisq)
## Duration                    4.5176  1  0.0335484 *
## Amount                      0.9337  1  0.3338965
## InstallmentRatePercentage  11.3934  1  0.0007371 ***
## NumberExistingCredits       0.1152  1  0.7343204
## Telephone                   0.0971  1  0.7553075
## ForeignWorker               2.8176  1  0.0932375 .
## Checking                   12.5511  3  0.0057152 **
## Credit.History              5.4775  4  0.2417140
## Loan.Purpose               17.3819  8  0.0263693 *
## Savings                     5.1908  4  0.2682743
## Employment.Duration        11.5037  4  0.0214501 *
## Personal.Status             7.6769  3  0.0531825 .
## Other.Debtors               3.4766  2  0.1758174
## Property                    3.9360  3  0.2684549
## OtherInstallmentPlans       6.5857  2  0.0371469 *
## Housing                     4.8759  2  0.0873417 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
logit.fit1 <- testDF %>%
  select (-ResidenceDuration, -Age, -Job.Type, -NumberPeopleMaintenance, -Telephone) %>%
  glm(good~.,family=binomial,data=.)
Anova(logit.fit1)
```

```
## Analysis of Deviance Table (Type II tests)
##
```

```
## Response: good
##                           LR Chisq Df Pr(>Chisq)
## Duration                    5.3842  1  0.0203197 *
## Amount                      0.8420  1  0.3588356
## InstallmentRatePercentage  11.3921  1  0.0007376 ***
## NumberExistingCredits       0.1271  1  0.7214523
## ForeignWorker               2.8107  1  0.0936387 .
## Checking                   12.8512  3  0.0049697 **
## Credit.History              5.5245  4  0.2375834
## Loan.Purpose               17.2880  8  0.0272459 *
## Savings                     5.4481  4  0.2443313
## Employment.Duration        12.1729  4  0.0161110 *
## Personal.Status             7.5981  3  0.0550910 .
## Other.Debtors               3.6879  2  0.1581883
## Property                    3.8401  3  0.2792551
## OtherInstallmentPlans       6.6240  2  0.0364432 *
## Housing                     4.8504  2  0.0884614 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
logit.fit1 <- testDF %>%
  select (-ResidenceDuration, -Age, -Job.Type, -NumberPeopleMaintenance, -Telephone, -NumberExistingCre
  glm(good~.,family=binomial,data=.)
Anova(logit.fit1)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: good
##                           LR Chisq Df Pr(>Chisq)
## Duration                    5.6130  1  0.0178280 *
## Amount                      0.8302  1  0.3622072
## InstallmentRatePercentage  11.5529  1  0.0006764 ***
## ForeignWorker               2.7999  1  0.0942676 .
## Checking                   12.7527  3  0.0052031 **
## Credit.History              5.6469  4  0.2271166
## Loan.Purpose               17.4097  8  0.0261143 *
## Savings                     5.5308  4  0.2370314
## Employment.Duration        12.0771  4  0.0167872 *
## Personal.Status             7.8709  3  0.0487567 *
## Other.Debtors               3.7012  2  0.1571411
## Property                    3.7478  3  0.2900155
## OtherInstallmentPlans       7.0272  2  0.0297895 *
## Housing                     4.7254  2  0.0941671 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
logit.fit1 <- testDF %>%
  select (-ResidenceDuration, -Age, -Job.Type, -NumberPeopleMaintenance, -Telephone, -NumberExistingCre
  glm(good~.,family=binomial,data=.)
Anova(logit.fit1)
```

```
## Analysis of Deviance Table (Type II tests)
##
```

```
## Response: good
##                            LR Chisq Df Pr(>Chisq)
## Duration                     4.2239  1  0.0398590 *
## Amount                       1.4615  1  0.2266843
## InstallmentRatePercentage   11.0547  1  0.0008846 ***
## ForeignWorker                2.0930  1  0.1479751
## Checking                    13.8502  3  0.0031164 **
## Credit.History               5.4496  4  0.2441965
## Loan.Purpose                18.8995  8  0.0154065 *
## Savings                      5.5444  4  0.2358554
## Employment.Duration         11.5045  4  0.0214424 *
## Personal.Status              7.5935  3  0.0552029 .
## Property                     3.1690  3  0.3662954
## OtherInstallmentPlans        7.9398  2  0.0188756 *
## Housing                      4.8578  2  0.0881319 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
logit.fit1 <- testDF %>%
  select (-ResidenceDuration, -Age, -Job.Type, -NumberPeopleMaintenance, -Telephone, -NumberExistingCre
  glm(good~.,family=binomial,data=.)
Anova(logit.fit1)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: good
##                            LR Chisq Df Pr(>Chisq)
## Duration                     4.5617  1  0.0326948 *
## Amount                       1.8618  1  0.1724219
## InstallmentRatePercentage   11.8197  1  0.0005861 ***
## ForeignWorker                2.6311  1  0.1047876
## Checking                    13.6574  3  0.0034106 **
## Credit.History               4.9755  4  0.2898243
## Loan.Purpose                17.9081  8  0.0219262 *
## Savings                      4.9934  4  0.2879760
## Employment.Duration         10.6035  4  0.0314005 *
## Personal.Status              7.8481  3  0.0492587 *
## OtherInstallmentPlans        8.1150  2  0.0172918 *
## Housing                      4.7659  2  0.0922766 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
logit.fit1 <- testDF %>%
  select (-ResidenceDuration, -Age, -Job.Type, -NumberPeopleMaintenance, -Telephone, -NumberExistingCre
  glm(good~.,family=binomial,data=.)
Anova(logit.fit1)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: good
##                            LR Chisq Df Pr(>Chisq)
## Duration                     6.6640  1  0.0098378 **
## Amount                       1.6628  1  0.1972289
```

```
## InstallmentRatePercentage  11.4632  1  0.0007099 ***
## ForeignWorker               3.3871  1  0.0657079 .
## Checking                   20.4268  3  0.0001385 ***
## Loan.Purpose               16.9365  8  0.0307769 *
## Savings                     4.9373  4  0.2937864
## Employment.Duration        10.6220  4  0.0311568 *
## Personal.Status             8.2914  3  0.0403588 *
## OtherInstallmentPlans       8.1691  2  0.0168308 *
## Housing                     5.6900  2  0.0581341 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
logit.fit1 <- testDF %>%
  select (-ResidenceDuration, -Age, -Job.Type, -NumberPeopleMaintenance, -Telephone, -NumberExistingCre
  glm(good~.,family=binomial,data=.)
Anova(logit.fit1)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: good
##                           LR Chisq Df Pr(>Chisq)
## Duration                    5.2013  1   0.022570 *
## Amount                      1.3982  1   0.237025
## InstallmentRatePercentage  10.1017  1   0.001481 **
## ForeignWorker               5.4644  1   0.019408 *
## Checking                   23.4797  3  3.208e-05 ***
## Loan.Purpose               14.9766  8   0.059603 .
## Employment.Duration        12.7705  4   0.012454 *
## Personal.Status             6.9387  3   0.073878 .
## OtherInstallmentPlans       7.6468  2   0.021853 *
## Housing                     5.8377  2   0.053995 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
logit.fit1 <- testDF %>%
  select (-ResidenceDuration, -Age, -Job.Type, -NumberPeopleMaintenance, -Telephone, -NumberExistingCre
  glm(good~.,family=binomial,data=.)
Anova(logit.fit1)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: good
##                           LR Chisq Df Pr(>Chisq)
## Duration                   14.9283  1  0.0001117 ***
## InstallmentRatePercentage   8.7118  1  0.0031616 **
## ForeignWorker               5.0552  1  0.0245519 *
## Checking                   22.6860  3  4.695e-05 ***
## Loan.Purpose               14.3763  8  0.0724685 .
## Employment.Duration        15.4521  4  0.0038498 **
## Personal.Status             5.7615  3  0.1238092
## OtherInstallmentPlans       7.7193  2  0.0210759 *
## Housing                     7.5569  2  0.0228585 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
logit.fit1 <- testDF %>%
  select (-ResidenceDuration, -Age, -Job.Type, -NumberPeopleMaintenance, -Telephone, -NumberExistingCre
  glm(good~.,family=binomial,data=.)
Anova(logit.fit1)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: good
##                          LR Chisq Df Pr(>Chisq)
## Duration                  14.1649  1  0.0001675 ***
## InstallmentRatePercentage  7.5654  1  0.0059500 **
## ForeignWorker              4.9385  1  0.0262653 *
## Checking                  19.5705  3  0.0002083 ***
## Loan.Purpose              13.8653  8  0.0853450 .
## Employment.Duration       21.1818  4  0.0002914 ***
## OtherInstallmentPlans      8.5885  2  0.0136469 *
## Housing                    8.9350  2  0.0114763 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

What factors increase the probability that a loan will be a good investment for the bank?

Factors that significantly increase the probability that a loan will be good are: Duration, Checking and Employment.Duration.

What factors may indicate that an individual may default on a loan or might be a bad investment for the bank?

Factors that significantly decrease the probability that a loan will be good are those with an extremely low Chisq value: Age, Telephone, NumberExistingCredits…etc. that were systematically stripped out from the improved model.