HW4

Jingshi

2/14/2017

Question 6

a)

$$\hat{\Pr}(Y = A|X_1, X_2) = \frac{e^{\hat{\beta}_0+\hat{\beta}_1 x_1+\hat{\beta}_2 x_2}}{1+e^{\hat{\beta}_0+\hat{\beta}_1 x_1+\hat{\beta}_2 x_2}} = \frac{e^{-6+0.05\times x_1+1\times x_2}}{1+e^{-6+0.05\times x_1+1\times x_2}} = \frac{e^{-6+0.05\times 40+1\times 3.5}}{1+e^{-6+0.05\times 40+1\times 3.5}}$$

Calculated in R:

```
exp(-6+0.05*40+1*3.5)/(1+exp(-6+0.05*40+3.5))
```

```
## [1] 0.3775407
```

As the output shows, the probability is 0.3775407.

b)

When $\hat{Pr}(Y = A|X_2) = 0.5$, $odds = \frac{\hat{Pr}(Y=A|X_2)}{1-\hat{Pr}(Y=A|X_2)} = \frac{0.5}{1-0.5} = 1$.

$log\ odds = log\ 1 = 0 = -6 + 0.05 \times x_1 + 1 \times 3.5$

$=> x_1 = 50$

Therefore, the student needs to study 50 hours.

Question 10

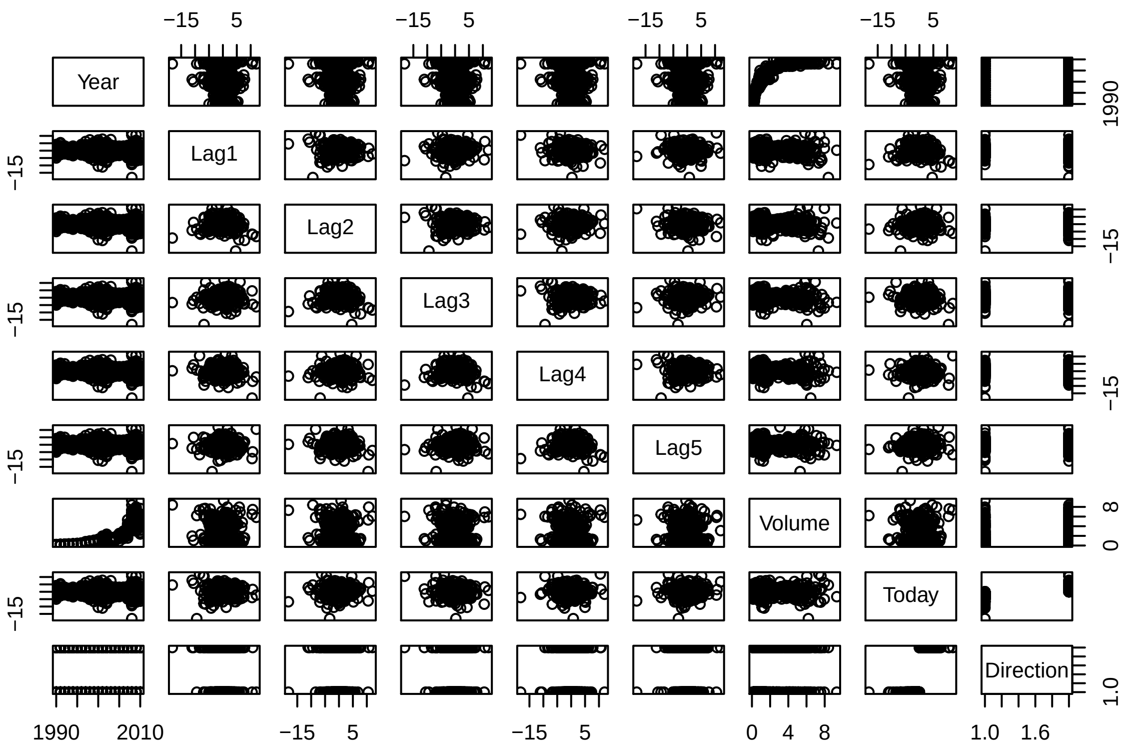
a)

```
library(ISLR)
help(Weekly)
summary(Weekly)
```

```
##       Year           Lag1               Lag2               Lag3
##  Min.   :1990   Min.   :-18.1950   Min.   :-18.1950   Min.   :-18.1950
##  1st Qu.:1995   1st Qu.: -1.1540   1st Qu.: -1.1540   1st Qu.: -1.1580
##  Median :2000   Median :  0.2410   Median :  0.2410   Median :  0.2410
##  Mean   :2000   Mean   :  0.1506   Mean   :  0.1511   Mean   :  0.1472
##  3rd Qu.:2005   3rd Qu.:  1.4050   3rd Qu.:  1.4090   3rd Qu.:  1.4090
##  Max.   :2010   Max.   : 12.0260   Max.   : 12.0260   Max.   : 12.0260
##       Lag4               Lag5              Volume
##  Min.   :-18.1950   Min.   :-18.1950   Min.   :0.08747
##  1st Qu.: -1.1580   1st Qu.: -1.1660   1st Qu.:0.33202
##  Median :  0.2380   Median :  0.2340   Median :1.00268
##  Mean   :  0.1458   Mean   :  0.1399   Mean   :1.57462
##  3rd Qu.:  1.4090   3rd Qu.:  1.4050   3rd Qu.:2.05373
##  Max.   : 12.0260   Max.   : 12.0260   Max.   :9.32821
```

```
##      Today          Direction
##  Min.   :-18.1950   Down:484
##  1st Qu.: -1.1540   Up  :605
##  Median :  0.2410
##  Mean   :  0.1499
##  3rd Qu.:  1.4050
##  Max.   : 12.0260
```

```r
pairs(Weekly)
```




```r
cor(Weekly[,1:8])
```

```
##               Year         Lag1        Lag2        Lag3         Lag4
## Year    1.00000000 -0.032289274 -0.03339001 -0.03000649 -0.031127923
## Lag1   -0.03228927  1.000000000 -0.07485305  0.05863568 -0.071273876
## Lag2   -0.03339001 -0.074853051  1.00000000 -0.07572091  0.058381535
## Lag3   -0.03000649  0.058635682 -0.07572091  1.00000000 -0.075395865
## Lag4   -0.03112792 -0.071273876  0.05838153 -0.07539587  1.000000000
## Lag5   -0.03051910 -0.008183096 -0.07249948  0.06065717 -0.075675027
## Volume  0.84194162 -0.064951313 -0.08551314 -0.06928771 -0.061074617
## Today  -0.03245989 -0.075031842  0.05916672 -0.07124364 -0.007825873
##                Lag5      Volume        Today
## Year   -0.030519101  0.84194162 -0.032459894
## Lag1   -0.008183096 -0.06495131 -0.075031842
## Lag2   -0.072499482 -0.08551314  0.059166717
## Lag3    0.060657175 -0.06928771 -0.071243639
## Lag4   -0.075675027 -0.06107462 -0.007825873
## Lag5    1.000000000 -0.05851741  0.011012698
## Volume -0.058517414  1.00000000 -0.033077783
## Today   0.011012698 -0.03307778  1.000000000
```

```r
#plot(Weekly$Volume~Weekly$Year)
```

According to the output, most of the variales have very small correlations (less than 0.1 or even close to zero). However, the correlation between Year and Volume is noticeably large (0.84194162). As you can see from the scatter plot matrix, there is a strong positive relationship (increasing trend) between Volume and Year.

b)

```r
model.fit<-glm(Direction ~ Lag1+Lag2+Lag3+Lag4+Lag5+Volume, data=Weekly, family=binomial)
summary(model.fit)
```

```
##
## Call:
## glm(formula = Direction ~ Lag1 + Lag2 + Lag3 + Lag4 + Lag5 +
##     Volume, family = binomial, data = Weekly)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6949  -1.2565   0.9913   1.0849   1.4579
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.26686    0.08593   3.106   0.0019 **
## Lag1        -0.04127    0.02641  -1.563   0.1181
## Lag2         0.05844    0.02686   2.175   0.0296 *
## Lag3        -0.01606    0.02666  -0.602   0.5469
## Lag4        -0.02779    0.02646  -1.050   0.2937
## Lag5        -0.01447    0.02638  -0.549   0.5833
## Volume      -0.02274    0.03690  -0.616   0.5377
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1496.2  on 1088  degrees of freedom
## Residual deviance: 1486.4  on 1082  degrees of freedom
## AIC: 1500.4
##
## Number of Fisher Scoring iterations: 4
```

According to the output, the only predictor that appears to be significant is lag2 because it has a p-value (0.0296) lower than $\alpha = 0.05$.

c)

```r
model.predict<-predict(model.fit, type = "response")
table(Weekly$Direction, model.predict>0.5)
```

```
##
##        FALSE TRUE
##   Down    54  430
##   Up      48  557
```

According to the output, the overall fraction of correct predictions (accuracy) is $\frac{54+557}{54+48+430+557} = 0.56107$. We have 48 false negative and 430 false positive (suppose Up is positive and Down is negative). The fraction of false positive is $\frac{430}{430+54} = 0.88843$. The specificity is $\frac{54}{430+54} = 0.11157$. The fraction of false negative is $\frac{48}{48+557} = 0.07934$. The sensitivity is $\frac{557}{557+48} = 0.92066$. The rate of missclassification $= \frac{48+430}{54+48+430+557} = 0.43893$ which indicates that the model does not fit very well.

When the model guesses "up", it has a probability of 0.5643364 ($=\frac{557}{430+557}$) to be correct; when the model guesses "down", it has a probability of 0.5294118 ($=\frac{54}{54+48}$) to be correct.

d)

```
# fit the logistic regression model using a training data period from 1990 to 2008,
# with Lag2 as the only predictor.
model.fit2<-glm(Direction~Lag2,data=Weekly[Weekly$Year<2009,],family = binomial)
summary(model.fit2)
```

```
##
## Call:
## glm(formula = Direction ~ Lag2, family = binomial, data = Weekly[Weekly$Year <
##     2009, ])
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -1.536  -1.264   1.021   1.091   1.368
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.20326    0.06428   3.162  0.00157 **
## Lag2         0.05810    0.02870   2.024  0.04298 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1354.7  on 984  degrees of freedom
## Residual deviance: 1350.5  on 983  degrees of freedom
## AIC: 1354.5
##
## Number of Fisher Scoring iterations: 4
```

```
# make predictions based on the held out data (the data from 2009 and 2010)
model.predict2<-predict(model.fit2, Weekly[Weekly$Year>=2009,],type = "response")
# compute the confusion matrix for the held out data (the data from 2009 and 2010).
table(Weekly$Direction[Weekly$Year>=2009],model.predict2>0.5)
```

```
##
##         FALSE TRUE
##    Down     9   34
##    Up       5   56
```

The overall fraction of correct predictions (accuracy) is $\frac{9+56}{9+34+5+56} = 0.625$ which is higher than the previous model. Its sensitivity is $\frac{56}{5+56} = 0.91803$. Its specificity is $\frac{9}{9+34} = 0.20930$.

Question 13

```r
library(MASS)
data(Boston)
help(Boston)
# convert numeric data into binominal: FALSE = 0 and TRUE = 1
Boston$crim01 <- as.numeric(Boston$crim > median(Boston$crim))
# fit a logistic regression with all predictors.
boston.glm<-glm(crim01~. - crim01 -crim, data = Boston, family = binomial)
summary(boston.glm)
```

```
##
## Call:
## glm(formula = crim01 ~ . - crim01 - crim, family = binomial,
##     data = Boston)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3946  -0.1585  -0.0004   0.0023   3.4239
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -34.103704   6.530014  -5.223 1.76e-07 ***
## zn           -0.079918   0.033731  -2.369  0.01782 *
## indus        -0.059389   0.043722  -1.358  0.17436
## chas          0.785327   0.728930   1.077  0.28132
## nox          48.523782   7.396497   6.560 5.37e-11 ***
## rm           -0.425596   0.701104  -0.607  0.54383
## age           0.022172   0.012221   1.814  0.06963 .
## dis           0.691400   0.218308   3.167  0.00154 **
## rad           0.656465   0.152452   4.306 1.66e-05 ***
## tax          -0.006412   0.002689  -2.385  0.01709 *
## ptratio       0.368716   0.122136   3.019  0.00254 **
## black        -0.013524   0.006536  -2.069  0.03853 *
## lstat         0.043862   0.048981   0.895  0.37052
## medv          0.167130   0.066940   2.497  0.01254 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 701.46  on 505  degrees of freedom
## Residual deviance: 211.93  on 492  degrees of freedom
## AIC: 239.93
##
## Number of Fisher Scoring iterations: 9
```

```r
# get predictions
glm.predict<-predict(boston.glm, type = "response")
# compute the confusion matrix
table(Boston$crim01, glm.predict>0.5)
```

```
##
##     FALSE TRUE
```

```
##    0   234   19
##    1    24  229
```

According to the summary of the model, zn, nox, dis, rad, tax, ptratio, black and medv are statistically significant because their p-values are lower than $\alpha = 0.05$.

According to the confusion matrix, the rate of misclassification is $\frac{24+19}{24+19+234+229} = 0.08498$. The accuracy is $\frac{229+234}{24+19+234+229} = 0.91502$ which indicates a good fit. The sensitivity is $\frac{229}{24+229} = 0.90514$. The specificity is $\frac{234}{19+234} = 0.92490$.

```r
# now fit a logistic regression model with another set of predictors
boston.glm<-glm(crim01~zn+nox+dis+rad+tax+ptratio+black+medv, data = Boston, family = binomial)
summary(boston.glm)
```

```
##
## Call:
## glm(formula = crim01 ~ zn + nox + dis + rad + tax + ptratio +
##     black + medv, family = binomial, data = Boston)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4400  -0.1918  -0.0008   0.0025   3.1885
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -28.347709   5.569863  -5.089 3.59e-07 ***
## zn           -0.074499   0.029975  -2.485  0.01294 *
## nox          44.180443   6.289746   7.024 2.15e-12 ***
## dis           0.489849   0.194930   2.513  0.01197 *
## rad           0.692116   0.137842   5.021 5.14e-07 ***
## tax          -0.007448   0.002428  -3.067  0.00216 **
## ptratio       0.272145   0.107311   2.536  0.01121 *
## black        -0.013484   0.006331  -2.130  0.03317 *
## medv          0.087913   0.030787   2.856  0.00430 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 701.46  on 505  degrees of freedom
## Residual deviance: 221.78  on 497  degrees of freedom
## AIC: 239.78
##
## Number of Fisher Scoring iterations: 9
```

```r
# get predictions
glm.predict<-predict(boston.glm, type = "response")
# compute the confusion matrix
table(Boston$crim01, glm.predict>0.5)
```

```
##
##     FALSE TRUE
##   0   229   24
##   1    30  223
```

According to the summary of the model, all the predictors are statistically significant because their p-values are all lower than $\alpha = 0.05$.

According to the confusion matrix, the rate of misclassification is $\frac{30+24}{229+24+30+223} = 0.10672$. The accuracy is $\frac{229+223}{229+24+30+223} = 0.89328$ which indicates a good fit. The sensitivity is $\frac{223}{30+223} = 0.88142$. The specificity is $\frac{229}{24+229} = 0.90514$.

```
# Now fit a LDA model
#library(MASS)
boston.lda<-lda(crim01~zn+nox+dis+rad+tax+ptratio+black+medv, data = Boston)
# get predictions
glm.predict<-predict(boston.lda, type = "response")
# compute the confusion matrix for the lda model
table(Boston$crim01, glm.predict$class)
```

```
##
##       0   1
##   0 247   6
##   1  59 194
```

According to the confusion matrix, the rate of misclassification is $\frac{59+6}{247+6+59+194} = 0.12846$. The accuracy is $\frac{247+194}{247+6+59+194} = 0.87154$ which indicates a good fit. The sensitivity is $\frac{194}{59+194} = 0.76680$. The specificity is $\frac{247}{6+247} = 0.97628$.

```
# Fit a KNN model by first split data into training and testing set.
set.seed(6)
n <- rnorm(nrow(Boston))
test <- n > quantile(n,0.80)
train <- !test
train.X <- cbind(Boston$zn, Boston$indus, Boston$chas)[train, ]
test.X <- cbind(Boston$zn, Boston$indus, Boston$chas)[test, ]
train.crim01 <- Boston$crim01[train]

# Now fit a KNN model with k = 6
library(class)
boston.knn<-knn(train.X, test.X, train.crim01, k = 6)

# compute the confusion matrix for the lda model
table(boston.knn, Boston$crim01[test])
```

```
##
## boston.knn  0  1
##          0 44  2
##          1  2 53
```

According to the confusion matrix, the rate of misclassification is $\frac{2+2}{44+2+2+53} = 0.03960$. The accuracy is $\frac{44+53}{44+2+2+53} = 0.96040$ which indicates a good fit. The sensitivity is $\frac{53}{2+53} = 0.96364$. The specificity is $\frac{44}{2+44} = 0.95652$.

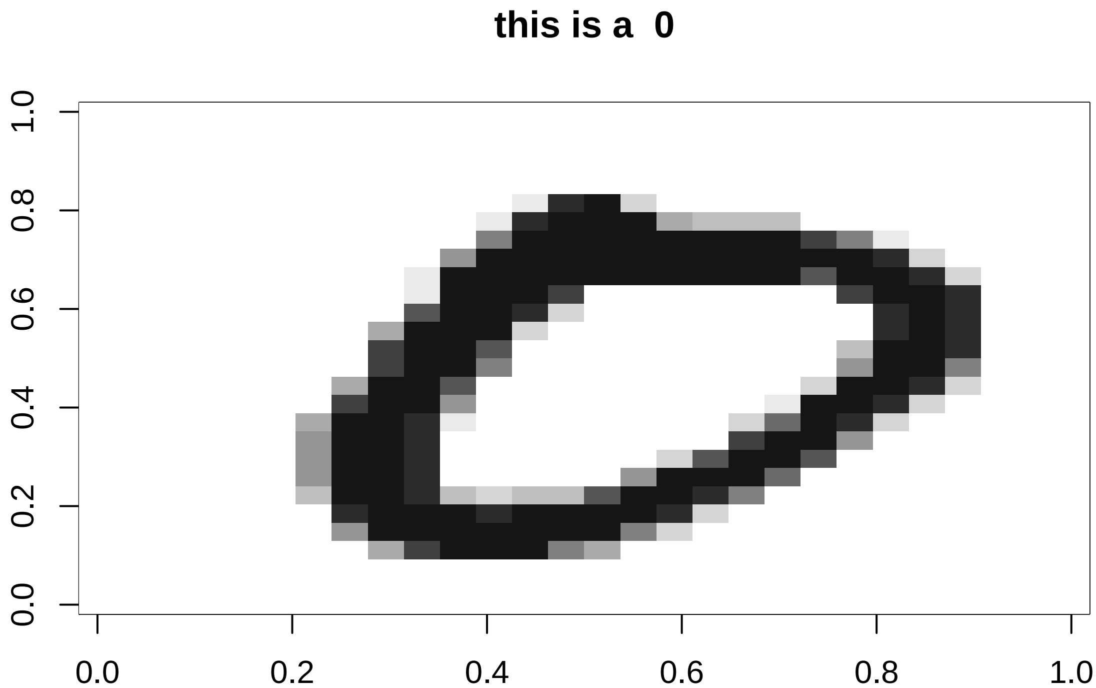
Immage Classification Problem

```
load("mnist_data.rdata")
plot_digit <- function(j){
arr784 <- as.numeric(images_df[j,1:784])
col=gray(12:1/12)
image(matrix(arr784, nrow=28)[,28:1], col=col,
        main = paste("this is a ",images_df$labels[j]))
```

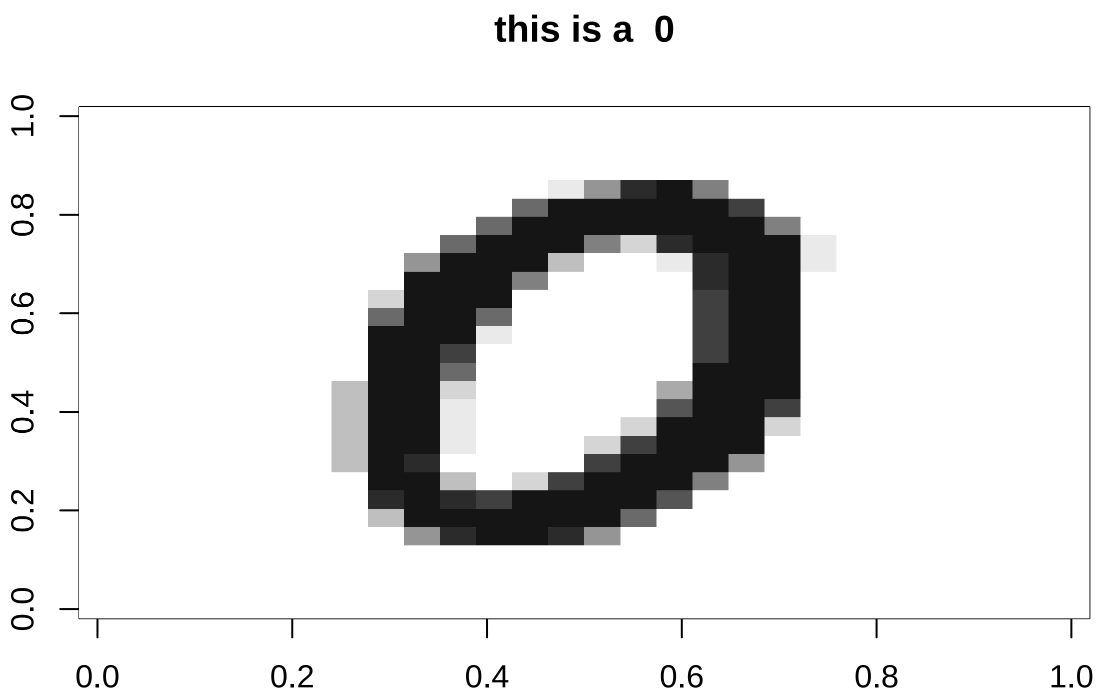
```
}
```

A.

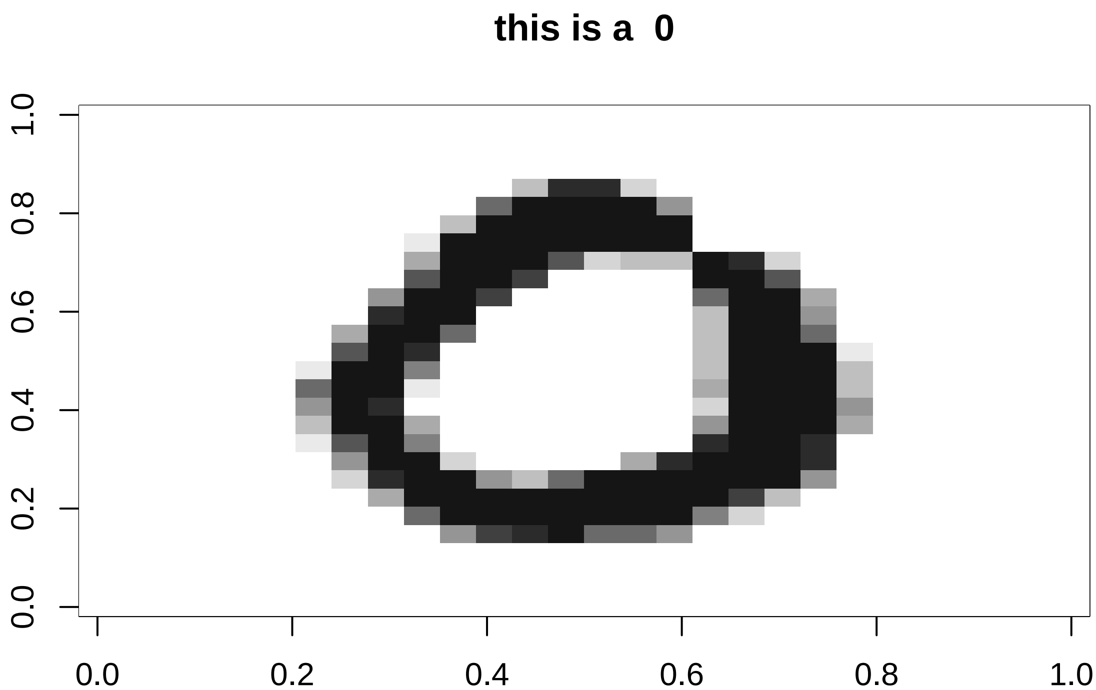
```
# Plot three 0's.
plot_digit(39)
```



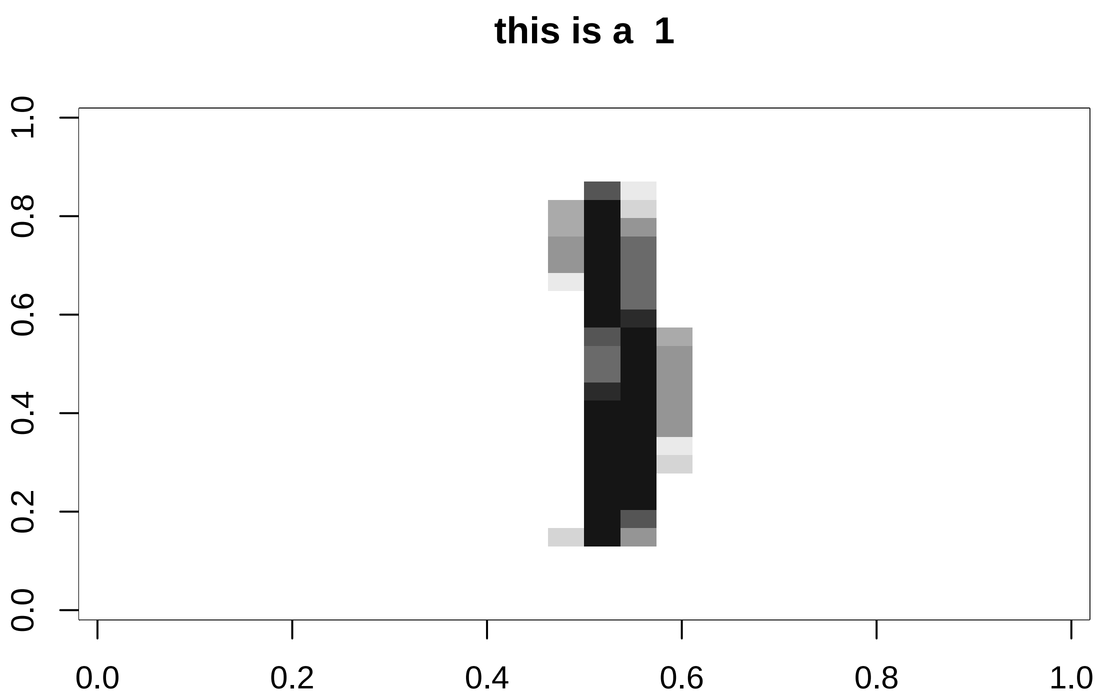

```
plot_digit(49)
```



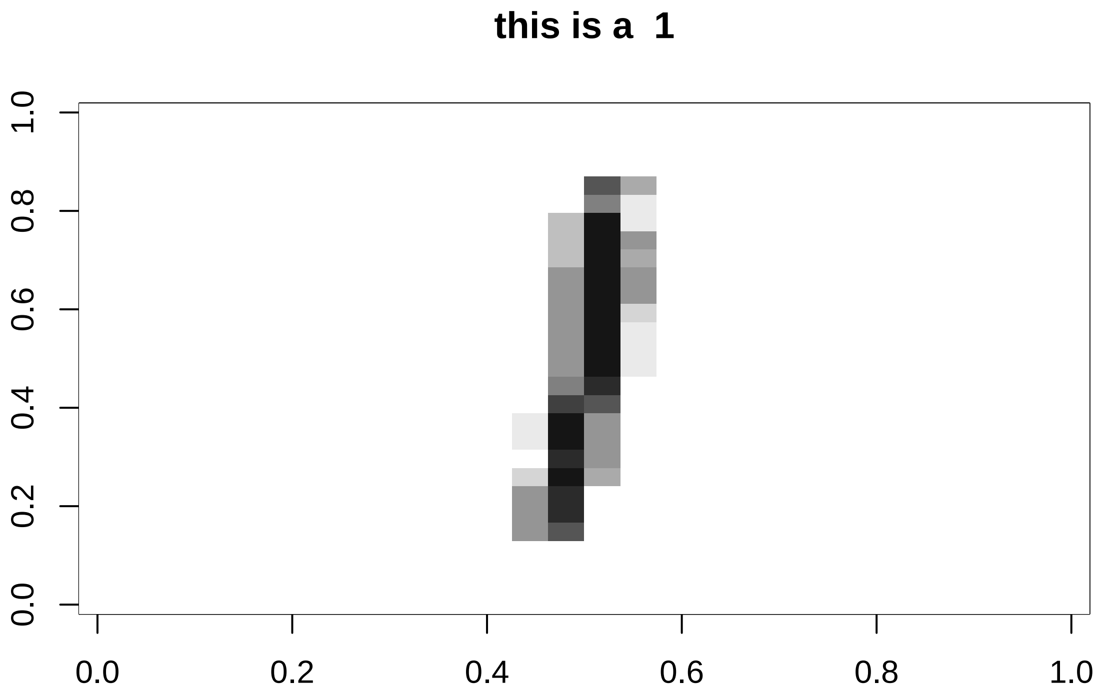
```r
plot_digit(58)
```



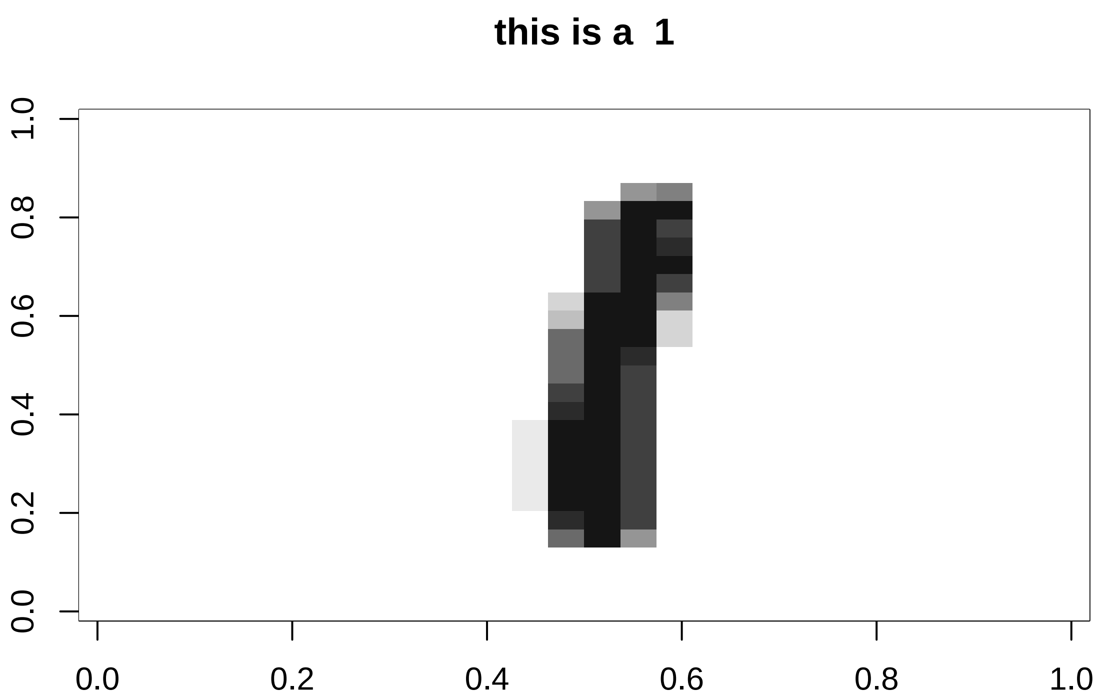

```r
# Plot three 1's.
plot_digit(32)
```




```r
plot_digit(35)
```




```
plot_digit(36)
```



B.

(i)

```
# A function to check whether a pixel has a zero variability
isZero<-function(j){
  res = paste("True, pixel",j,"has a zero variability.")
```

```r
  for (i in 1 : 2115){
    if (images_df[i,j]!=0){
      res = paste("False, the variability of pixel",j,"is not zero.")
      break
    }
  }
  print(res)
}

# Test to see if pixel 1 has a zero variability.
isZero(1)
```

```
## [1] "True, pixel 1 has a zero variability."
```

```r
# Test to see if pixel 2 has a zero variability.
isZero(2)
```

```
## [1] "True, pixel 2 has a zero variability."
```

```r
# Test to see if pixel 3 has a zero variability.
isZero(3)
```

```
## [1] "True, pixel 3 has a zero variability."
```

```r
# Test to see if pixel 4 has a zero variability.
isZero(4)
```

```
## [1] "True, pixel 4 has a zero variability."
```

```r
# Test to see if pixel 5 has a zero variability.
isZero(5)
```

```
## [1] "True, pixel 5 has a zero variability."
```

According to the output, the five features are 1, 2, 3, 4 and 5.

(ii)

```r
# Test to see if pixel 100 has a zero variability.
isZero(100)
```

```
## [1] "False, the variability of pixel 100 is not zero."
```

```r
# Test to see if pixel 101 has a zero variability.
isZero(101)
```

```
## [1] "False, the variability of pixel 101 is not zero."
```

```r
# Test to see if pixel 102 has a zero variability.
isZero(102)
```

```
## [1] "False, the variability of pixel 102 is not zero."
```

```r
# Test to see if pixel 103 has a zero variability.
isZero(103)
```

```
## [1] "False, the variability of pixel 103 is not zero."
```

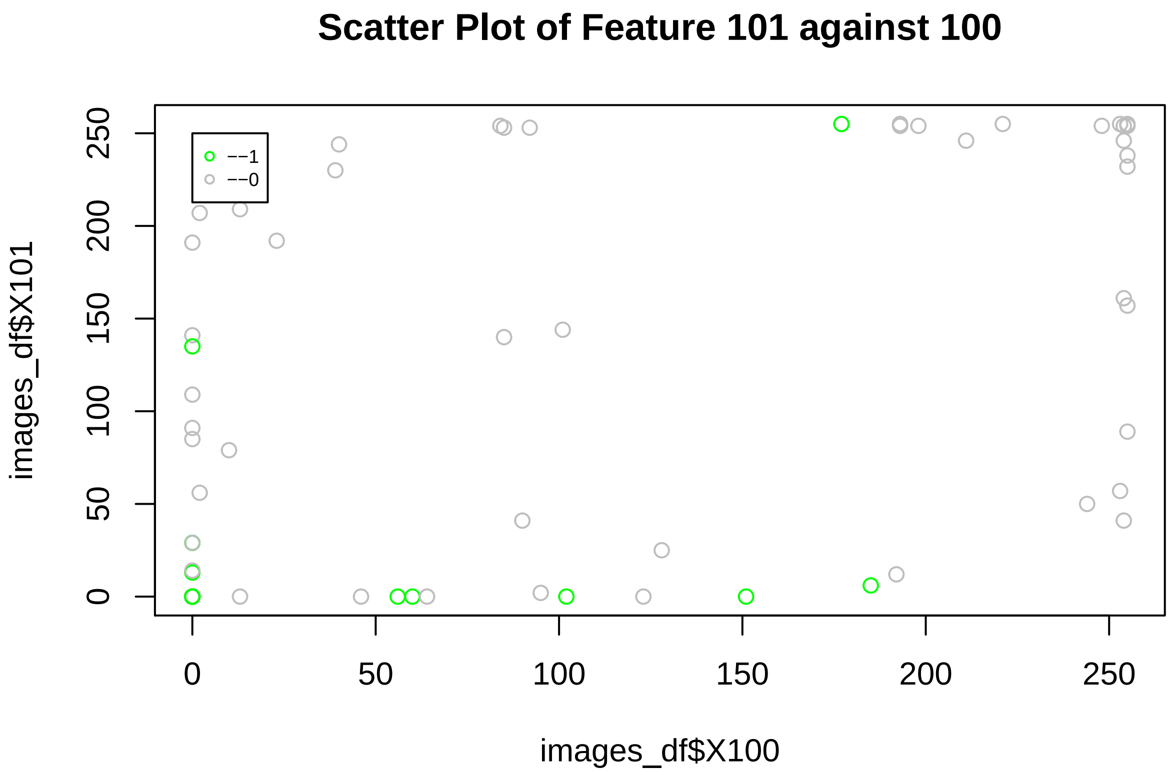
```r
# Test to see if pixel 104 has a zero variability.
isZero(104)
```

```
## [1] "False, the variability of pixel 104 is not zero."
```

According to the output, the five features are 100, 101, 102, 103 and 104.

(iii)

```r
# The two features I pick are 100 and 101.
plot(images_df$X100,images_df$X101,col=ifelse(images_df$labels==1, "green", "grey"),
     main = "Scatter Plot of Feature 101 against 100")
legend(0, 250,pch=c(1,1), col=c("green", "grey"), c("--1", "--0"),cex=.6)
```



C.

```r
#install.packages("pROC")
library(pROC)
```

```
## Warning: package 'pROC' was built under R version 3.3.2

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

##
## Attaching package: 'pROC'

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

```r
# The three pairs that I choose are 100 & 101, 101 & 102, 102 & 103
```

```r
# firstly, fit a logistic regression of labels~X100+X101
c.glm1<-glm(labels~X100+X101, data = images_df, family = binomial)
# take a look at the summary
summary(c.glm1)
```

```
##
## Call:
## glm(formula = labels ~ X100 + X101, family = binomial, data = images_df)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -1.257  -1.257   1.100   1.100   2.847
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.184251   0.044137   4.175 2.99e-05 ***
## X100        -0.004841   0.003784  -1.279  0.20076
## X101        -0.013184   0.004752  -2.774  0.00553 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2920.6  on 2114  degrees of freedom
## Residual deviance: 2879.8  on 2112  degrees of freedom
## AIC: 2885.8
##
## Number of Fisher Scoring iterations: 5
```

```r
# get predictions
glm.predict<-predict(c.glm1, type = "response")

# compute the confusion matrix
table(images_df$labels, glm.predict>0.5)
```

```
##
##     FALSE TRUE
##   0    45  935
##   1     9 1126
```

```r
# compute roc
roc<-roc(images_df$labels, glm.predict)

# comput AUC
auc(roc)
```

```
## Area under the curve: 0.5192
```

According to the summary of the model, the intercept and X101 are statistically significant because their p-values (2.99×10^{-5} and 0.00553 separately) are less than $\alpha = 0.05$. The p-value of X100 is 0.20076 which is much larger than $\alpha = 0.05$ so that X100 is not statistically significant.

According to the confusion matrix, there are 9 false negatives and 935 false positives. The rate of miss-classification is $\frac{935+9}{935+9+45+1126} = 0.44634$. The accuracy is $\frac{45+1126}{935+9+45+1126} = 0.55366$. The sensitivity is $\frac{1126}{1126+9} = 0.99207$. The specificity is $\frac{45}{935+45} = 0.04592$.

The AUC is 0.5192.

```r
# secondly, fit a logistic regression of labels~X101+X102
c.glm2<-glm(labels~X101+X102, data = images_df, family = binomial)
# take a look at the summary
summary(c.glm2)
```

```
##
## Call:
## glm(formula = labels ~ X101 + X102, family = binomial, data = images_df)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -1.291  -1.255   1.102   1.102   2.870
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.1804086  0.0441161   4.089 4.32e-05 ***
## X101        -0.0173091  0.0057676  -3.001  0.00269 **
## X102         0.0009502  0.0036140   0.263  0.79261
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2920.6  on 2114  degrees of freedom
## Residual deviance: 2881.6  on 2112  degrees of freedom
## AIC: 2887.6
##
## Number of Fisher Scoring iterations: 5
```

```r
# get predictions
glm.predict<-predict(c.glm2, type = "response")

# compute the confusion matrix
table(images_df$labels, glm.predict>0.5)
```

```
##
##     FALSE TRUE
##   0    40  940
##   1     4 1131
```

```r
# compute roc
roc<-roc(images_df$labels, glm.predict)

# comput AUC
auc(roc)
```

```
## Area under the curve: 0.5182
```

According to the summary of the model, the intercept and X101 are statistically significant because their p-values (4.32×10^{-5} and 0.00269 separately) are less than $\alpha = 0.05$. The p-value of X102 is 0.79261 which is much larger than $\alpha = 0.05$ so that X102 is not statistically significant.

According to the confusion matrix, there are 4 false negatives and 940 false positives. The rate of miss-classification is $\frac{940+4}{940+4+40+1131} = 0.44634$. The accuracy is $\frac{40+1131}{940+4+40+1131} = 0.55366$. The sensitivity is $\frac{1131}{1131+4} = 0.99648$. The specificity is $\frac{40}{940+40} = 0.04082$.

The AUC is 0.5182.

```r
# thirdly, fit a logistic regression of labels~X102+X103
c.glm3<-glm(labels~X102+X103, data = images_df, family = binomial)
# take a look at the summary
summary(c.glm3)
```

```
##
## Call:
## glm(formula = labels ~ X102 + X103, family = binomial, data = images_df)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -1.252  -1.252   1.105   1.105   2.074
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.173295   0.044061   3.933 8.39e-05 ***
## X102        -0.008884   0.003065  -2.898  0.00375 **
## X103        -0.001711   0.003101  -0.552  0.58096
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2920.6  on 2114  degrees of freedom
## Residual deviance: 2899.4  on 2112  degrees of freedom
## AIC: 2905.4
##
## Number of Fisher Scoring iterations: 4
```

```r
# get predictions
glm.predict<-predict(c.glm3, type = "response")

# compute the confusion matrix
table(images_df$labels, glm.predict>0.5)
```

```
##
##     FALSE TRUE
##   0    38  942
##   1     9 1126
```

```r
# compute roc
roc<-roc(images_df$labels, glm.predict)

# comput AUC
auc(roc)
```

```
## Area under the curve: 0.5155
```

According to the summary of the model, the intercept and X102 are statistically significant because their p-values (8.39×10^{-5} and 0.00375 separately) are less than $\alpha = 0.05$. The p-value of X103 is 0.58096 which is much larger than $\alpha = 0.05$ so that X103 is not statistically significant.

According to the confusion matrix, there are 9 false negatives and 942 false positives. The rate of miss-classification is $\frac{942+9}{942+9+38+1126} = 0.44965$. The accuracy is $\frac{38+1126}{942+9+38+1126} = 0.55035$. The sensitivity is $\frac{1126}{1126+9} = 0.99207$. The specificity is $\frac{38}{942+38} = 0.03878$.

The AUC is 0.5155.

Among the three models the best model is the first model (c.glm1) because it has the lowest rate of misssclassification (0.44634) and the highest AUC value (0.5192).

D.

```r
# fit a logistic regression of labels~X100+X101+X102+X103+X104
c.glm3<-glm(labels~X100+X101+X102+X103+X104,data=images_df, family = binomial)

# get predictions
glm.predict<-predict(c.glm3, type = "response")

# compute the confusion matrix
table(images_df$labels, glm.predict>0.5)
```

```
##
##     FALSE TRUE
##   0    52  928
##   1    12 1123
```

```r
# compute roc
roc<-roc(images_df$labels, glm.predict)

# comput AUC
auc(roc)
```

```
## Area under the curve: 0.5228
```

According to the confusion matrix, there are 12 false negatives and 928 false positives. The rate of missclassification is $\frac{928+12}{928+12+52+1123} = 0.44444$ which is lower than the best model in C. Additionally, the accuracy is $\frac{52+1123}{928+12+52+1123} = 0.55556$ which is better than the best model in C. The sensitivity is $\frac{1123}{1123+12} = 0.98943$. The specificity is $\frac{52}{928+52} = 0.05306$.

The AUC is 0.5228 which is also better than the best model in C.

E.

```r
# fit a logistic regression with all features
e.glm<-glm(labels~., data = images_df, family = binomial)
```

```
## Warning: glm.fit: algorithm did not converge
```

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

```r
# take a look at the long summary
summary(e.glm)
```

```
##
## Call:
## glm(formula = labels ~ ., family = binomial, data = images_df)
##
## Deviance Residuals:
##        Min          1Q      Median          3Q         Max
## -1.459e-05  -2.110e-08   2.110e-08   4.485e-07   2.148e-05
##
## Coefficients: (286 not defined because of singularities)
```

```
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.903e+00  8.495e+04       0        1
## X1                  NA         NA      NA       NA
## X2                  NA         NA      NA       NA
## X3                  NA         NA      NA       NA
## X4                  NA         NA      NA       NA
## X5                  NA         NA      NA       NA
## X6                  NA         NA      NA       NA
## X7                  NA         NA      NA       NA
## X8                  NA         NA      NA       NA
## X9                  NA         NA      NA       NA
## X10                 NA         NA      NA       NA
## X11                 NA         NA      NA       NA
## X12                 NA         NA      NA       NA
## X13                 NA         NA      NA       NA
## X14                 NA         NA      NA       NA
## X15                 NA         NA      NA       NA
## X16                 NA         NA      NA       NA
## X17                 NA         NA      NA       NA
## X18                 NA         NA      NA       NA
## X19                 NA         NA      NA       NA
## X20                 NA         NA      NA       NA
## X21                 NA         NA      NA       NA
## X22                 NA         NA      NA       NA
## X23                 NA         NA      NA       NA
## X24                 NA         NA      NA       NA
## X25                 NA         NA      NA       NA
## X26                 NA         NA      NA       NA
## X27                 NA         NA      NA       NA
## X28                 NA         NA      NA       NA
## X29                 NA         NA      NA       NA
## X30                 NA         NA      NA       NA
## X31                 NA         NA      NA       NA
## X32                 NA         NA      NA       NA
## X33                 NA         NA      NA       NA
## X34                 NA         NA      NA       NA
## X35                 NA         NA      NA       NA
## X36                 NA         NA      NA       NA
## X37                 NA         NA      NA       NA
## X38                 NA         NA      NA       NA
## X39                 NA         NA      NA       NA
## X40                 NA         NA      NA       NA
## X41                 NA         NA      NA       NA
## X42                 NA         NA      NA       NA
## X43                 NA         NA      NA       NA
## X44                 NA         NA      NA       NA
## X45                 NA         NA      NA       NA
## X46                 NA         NA      NA       NA
## X47                 NA         NA      NA       NA
## X48                 NA         NA      NA       NA
## X49                 NA         NA      NA       NA
## X50                 NA         NA      NA       NA
## X51                 NA         NA      NA       NA
## X52                 NA         NA      NA       NA
```

```
## X53                  NA         NA      NA       NA
## X54                  NA         NA      NA       NA
## X55                  NA         NA      NA       NA
## X56                  NA         NA      NA       NA
## X57                  NA         NA      NA       NA
## X58                  NA         NA      NA       NA
## X59                  NA         NA      NA       NA
## X60                  NA         NA      NA       NA
## X61                  NA         NA      NA       NA
## X62                  NA         NA      NA       NA
## X63                  NA         NA      NA       NA
## X64                  NA         NA      NA       NA
## X65                  NA         NA      NA       NA
## X66                  NA         NA      NA       NA
## X67                  NA         NA      NA       NA
## X68           2.268e-01  2.738e+04       0        1
## X69          -1.449e-01  2.759e+04       0        1
## X70           1.617e-01  2.108e+04       0        1
## X71          -3.743e-02  5.474e+03       0        1
## X72                  NA         NA      NA       NA
## X73                  NA         NA      NA       NA
## X74                  NA         NA      NA       NA
## X75                  NA         NA      NA       NA
## X76           1.101e-01  3.584e+03       0        1
## X77                  NA         NA      NA       NA
## X78                  NA         NA      NA       NA
## X79                  NA         NA      NA       NA
## X80                  NA         NA      NA       NA
## X81                  NA         NA      NA       NA
## X82                  NA         NA      NA       NA
## X83                  NA         NA      NA       NA
## X84                  NA         NA      NA       NA
## X85                  NA         NA      NA       NA
## X86                  NA         NA      NA       NA
## X87                  NA         NA      NA       NA
## X88                  NA         NA      NA       NA
## X89                  NA         NA      NA       NA
## X90                  NA         NA      NA       NA
## X91                  NA         NA      NA       NA
## X92                  NA         NA      NA       NA
## X93                  NA         NA      NA       NA
## X94          -2.887e-01  2.545e+04       0        1
## X95           9.718e-02  6.589e+03       0        1
## X96          -2.445e-02  5.255e+03       0        1
## X97          -4.224e-02  4.146e+03       0        1
## X98           1.605e-02  2.771e+03       0        1
## X99          -6.050e-03  2.318e+03       0        1
## X100          1.542e-02  1.573e+03       0        1
## X101         -2.892e-02  1.363e+03       0        1
## X102         -5.882e-03  1.620e+03       0        1
## X103         -1.043e-02  1.549e+03       0        1
## X104          8.204e-03  1.877e+03       0        1
## X105         -3.367e-02  2.464e+03       0        1
## X106          1.783e-02  3.973e+03       0        1
```

```
## X107         -1.382e-02  5.855e+03       0         1
## X108          1.723e-01  8.893e+03       0         1
## X109         -4.518e-01  1.585e+04       0         1
## X110                 NA         NA      NA        NA
## X111                 NA         NA      NA        NA
## X112                 NA         NA      NA        NA
## X113                 NA         NA      NA        NA
## X114                 NA         NA      NA        NA
## X115                 NA         NA      NA        NA
## X116          2.327e-01  9.938e+03       0         1
## X117                 NA         NA      NA        NA
## X118          3.883e-03  3.805e+03       0         1
## X119          1.310e-01  3.768e+03       0         1
## X120         -1.486e-01  1.244e+04       0         1
## X121         -2.502e-02  1.623e+04       0         1
## X122         -1.921e-02  2.893e+03       0         1
## X123          1.005e-02  1.521e+03       0         1
## X124         -1.560e-02  1.126e+03       0         1
## X125          8.021e-03  7.765e+02       0         1
## X126         -8.455e-03  4.745e+02       0         1
## X127          5.466e-03  4.069e+02       0         1
## X128         -1.039e-03  3.446e+02       0         1
## X129         -9.235e-03  3.776e+02       0         1
## X130          6.547e-03  4.050e+02       0         1
## X131         -5.855e-03  4.964e+02       0         1
## X132         -1.031e-02  4.814e+02       0         1
## X133          8.397e-03  6.157e+02       0         1
## X134         -1.060e-02  8.594e+02       0         1
## X135         -7.675e-03  1.272e+03       0         1
## X136         -3.657e-02  2.345e+03       0         1
## X137          2.493e-02  3.020e+03       0         1
## X138          5.464e+01  6.680e+06       0         1
## X139                 NA         NA      NA        NA
## X140                 NA         NA      NA        NA
## X141                 NA         NA      NA        NA
## X142                 NA         NA      NA        NA
## X143                 NA         NA      NA        NA
## X144                 NA         NA      NA        NA
## X145                 NA         NA      NA        NA
## X146         -6.702e-01  9.708e+04       0         1
## X147          2.848e-01  4.550e+04       0         1
## X148         -2.479e-03  7.953e+03       0         1
## X149          8.590e-02  3.196e+03       0         1
## X150         -3.101e-02  2.127e+03       0         1
## X151          1.307e-02  1.217e+03       0         1
## X152         -7.081e-03  9.184e+02       0         1
## X153         -7.648e-03  6.075e+02       0         1
## X154         -1.312e-03  4.480e+02       0         1
## X155         -5.251e-03  4.288e+02       0         1
## X156          4.349e-03  3.782e+02       0         1
## X157         -1.411e-02  3.980e+02       0         1
## X158          4.773e-03  3.946e+02       0         1
## X159          3.861e-03  4.400e+02       0         1
## X160         -2.225e-03  4.986e+02       0         1
```

```
## X161          1.748e-03  6.230e+02       0        1
## X162         -9.747e-03  9.466e+02       0        1
## X163          6.490e-03  1.209e+03       0        1
## X164          1.845e-03  1.834e+03       0        1
## X165          7.217e-02  4.733e+03       0        1
## X166          4.616e+01  5.516e+06       0        1
## X167                 NA         NA      NA       NA
## X168                 NA         NA      NA       NA
## X169                 NA         NA      NA       NA
## X170                 NA         NA      NA       NA
## X171                 NA         NA      NA       NA
## X172                 NA         NA      NA       NA
## X173          1.730e-01  6.924e+03       0        1
## X174         -1.770e-02  1.618e+04       0        1
## X175         -8.061e-02  1.977e+04       0        1
## X176          1.543e-03  4.157e+03       0        1
## X177         -1.180e-04  2.553e+03       0        1
## X178          1.481e-02  1.338e+03       0        1
## X179         -6.479e-04  1.013e+03       0        1
## X180         -2.236e-02  8.071e+02       0        1
## X181          8.236e-03  6.597e+02       0        1
## X182         -1.329e-02  5.129e+02       0        1
## X183          6.608e-03  4.560e+02       0        1
## X184         -6.164e-03  4.912e+02       0        1
## X185          1.389e-02  4.486e+02       0        1
## X186         -7.114e-03  4.275e+02       0        1
## X187         -1.903e-02  5.159e+02       0        1
## X188          1.149e-02  5.276e+02       0        1
## X189         -7.317e-03  6.661e+02       0        1
## X190          1.421e-02  8.178e+02       0        1
## X191         -4.269e-03  1.007e+03       0        1
## X192         -3.431e-02  2.024e+03       0        1
## X193         -2.843e-02  5.141e+03       0        1
## X194         -3.228e+01  3.858e+06       0        1
## X195                 NA         NA      NA       NA
## X196                 NA         NA      NA       NA
## X197                 NA         NA      NA       NA
## X198                 NA         NA      NA       NA
## X199                 NA         NA      NA       NA
## X200                 NA         NA      NA       NA
## X201         -1.813e+00  5.731e+04       0        1
## X202         -6.315e-02  2.191e+04       0        1
## X203          4.119e-05  4.678e+03       0        1
## X204         -2.200e-02  1.831e+03       0        1
## X205          1.072e-02  1.807e+03       0        1
## X206         -2.026e-02  1.202e+03       0        1
## X207          1.190e-02  9.110e+02       0        1
## X208          7.846e-03  9.129e+02       0        1
## X209          2.299e-02  6.439e+02       0        1
## X210         -5.063e-03  5.254e+02       0        1
## X211          5.123e-03  4.676e+02       0        1
## X212         -1.624e-02  4.395e+02       0        1
## X213         -5.377e-03  4.240e+02       0        1
## X214          6.089e-03  4.800e+02       0        1
```

```
## X215         1.606e-02  5.495e+02       0        1
## X216         1.685e-03  5.490e+02       0        1
## X217        -1.567e-02  6.862e+02       0        1
## X218         8.437e-04  6.538e+02       0        1
## X219        -4.870e-03  1.138e+03       0        1
## X220         1.856e-02  1.856e+03       0        1
## X221         1.320e-02  3.109e+03       0        1
## X222        -1.872e-01  3.054e+04       0        1
## X223                NA         NA      NA       NA
## X224                NA         NA      NA       NA
## X225                NA         NA      NA       NA
## X226                NA         NA      NA       NA
## X227                NA         NA      NA       NA
## X228        -1.127e+00  1.308e+05       0        1
## X229         4.245e-01  5.606e+04       0        1
## X230         5.153e-02  1.215e+04       0        1
## X231        -5.483e-02  2.873e+03       0        1
## X232        -1.927e-02  2.256e+03       0        1
## X233         2.442e-02  1.670e+03       0        1
## X234        -7.122e-04  1.105e+03       0        1
## X235        -1.861e-02  8.892e+02       0        1
## X236        -1.209e-02  7.580e+02       0        1
## X237        -9.207e-03  6.698e+02       0        1
## X238        -2.669e-03  5.480e+02       0        1
## X239        -4.627e-04  4.755e+02       0        1
## X240         1.681e-02  4.739e+02       0        1
## X241        -1.556e-02  4.411e+02       0        1
## X242         6.626e-03  5.112e+02       0        1
## X243        -1.302e-02  5.277e+02       0        1
## X244         2.761e-03  5.984e+02       0        1
## X245         2.154e-03  6.789e+02       0        1
## X246         1.881e-02  7.409e+02       0        1
## X247         5.804e-03  1.066e+03       0        1
## X248        -9.846e-03  1.644e+03       0        1
## X249         8.118e-02  2.763e+03       0        1
## X250         5.398e-02  4.774e+04       0        1
## X251                NA         NA      NA       NA
## X252                NA         NA      NA       NA
## X253                NA         NA      NA       NA
## X254                NA         NA      NA       NA
## X255                NA         NA      NA       NA
## X256                NA         NA      NA       NA
## X257        -1.642e-01  4.502e+04       0        1
## X258         6.284e-02  6.799e+03       0        1
## X259         9.554e-03  2.287e+03       0        1
## X260         7.198e-03  1.824e+03       0        1
## X261         2.171e-04  1.128e+03       0        1
## X262         5.713e-04  9.985e+02       0        1
## X263        -6.334e-04  8.437e+02       0        1
## X264        -5.053e-03  7.826e+02       0        1
## X265        -4.283e-03  6.449e+02       0        1
## X266         5.731e-03  5.081e+02       0        1
## X267         6.443e-03  5.364e+02       0        1
## X268        -2.052e-02  5.033e+02       0        1
```

```
## X269         2.149e-02  5.304e+02       0         1
## X270        -1.288e-02  5.426e+02       0         1
## X271        -3.607e-03  6.441e+02       0         1
## X272         5.289e-03  5.915e+02       0         1
## X273        -8.320e-04  7.914e+02       0         1
## X274        -2.548e-03  8.364e+02       0         1
## X275        -2.557e-02  1.313e+03       0         1
## X276         2.493e-02  1.583e+03       0         1
## X277        -3.301e-02  2.217e+03       0         1
## X278        -1.142e-01  2.046e+04       0         1
## X279                NA         NA      NA        NA
## X280                NA         NA      NA        NA
## X281                NA         NA      NA        NA
## X282                NA         NA      NA        NA
## X283                NA         NA      NA        NA
## X284                NA         NA      NA        NA
## X285         2.296e-01  1.805e+04       0         1
## X286         3.437e-02  3.122e+03       0         1
## X287        -1.209e-02  2.584e+03       0         1
## X288         9.994e-03  1.633e+03       0         1
## X289        -1.717e-02  1.125e+03       0         1
## X290         4.449e-02  1.000e+03       0         1
## X291         1.686e-02  8.725e+02       0         1
## X292        -9.371e-04  9.658e+02       0         1
## X293         1.310e-02  6.485e+02       0         1
## X294         9.793e-03  4.914e+02       0         1
## X295        -3.174e-03  4.903e+02       0         1
## X296         1.989e-02  6.213e+02       0         1
## X297        -2.511e-02  5.415e+02       0         1
## X298         1.071e-02  6.020e+02       0         1
## X299         1.711e-02  7.212e+02       0         1
## X300         2.737e-03  7.331e+02       0         1
## X301        -2.695e-02  9.544e+02       0         1
## X302         2.881e-03  8.902e+02       0         1
## X303         1.322e-02  1.437e+03       0         1
## X304        -4.868e-02  1.531e+03       0         1
## X305        -4.274e-02  1.900e+03       0         1
## X306         7.985e-02  1.182e+04       0         1
## X307                NA         NA      NA        NA
## X308                NA         NA      NA        NA
## X309                NA         NA      NA        NA
## X310                NA         NA      NA        NA
## X311                NA         NA      NA        NA
## X312        -1.018e-01  9.486e+03       0         1
## X313        -1.409e-01  9.436e+03       0         1
## X314         2.630e-02  3.123e+03       0         1
## X315        -2.191e-02  1.910e+03       0         1
## X316         9.742e-03  1.476e+03       0         1
## X317        -3.289e-03  1.025e+03       0         1
## X318        -1.884e-02  1.056e+03       0         1
## X319        -2.912e-02  1.052e+03       0         1
## X320        -2.584e-03  8.053e+02       0         1
## X321        -1.362e-02  7.068e+02       0         1
## X322        -7.740e-03  5.207e+02       0         1
```

```
## X323         1.258e-02  4.996e+02       0        1
## X324        -2.100e-02  6.662e+02       0        1
## X325         1.235e-02  5.196e+02       0        1
## X326         3.430e-03  6.326e+02       0        1
## X327        -2.130e-02  8.298e+02       0        1
## X328        -5.051e-03  8.496e+02       0        1
## X329         1.766e-02  9.275e+02       0        1
## X330         4.007e-03  9.426e+02       0        1
## X331        -3.036e-02  1.288e+03       0        1
## X332         2.314e-03  1.650e+03       0        1
## X333         1.433e-02  1.698e+03       0        1
## X334        -2.464e-02  1.136e+04       0        1
## X335                NA         NA      NA       NA
## X336                NA         NA      NA       NA
## X337                NA         NA      NA       NA
## X338                NA         NA      NA       NA
## X339                NA         NA      NA       NA
## X340        -1.488e-01  5.818e+03       0        1
## X341        -1.892e-02  4.379e+03       0        1
## X342         5.822e-03  2.726e+03       0        1
## X343        -2.256e-02  1.518e+03       0        1
## X344        -9.293e-03  1.082e+03       0        1
## X345         7.306e-03  1.003e+03       0        1
## X346         3.160e-02  1.117e+03       0        1
## X347         3.868e-02  9.489e+02       0        1
## X348        -6.826e-03  8.553e+02       0        1
## X349        -7.193e-03  6.844e+02       0        1
## X350        -1.588e-02  5.680e+02       0        1
## X351         1.725e-02  5.655e+02       0        1
## X352         2.493e-02  7.867e+02       0        1
## X353        -1.664e-02  6.392e+02       0        1
## X354        -2.719e-03  7.623e+02       0        1
## X355        -1.256e-02  9.871e+02       0        1
## X356         2.052e-02  1.028e+03       0        1
## X357        -1.905e-02  1.045e+03       0        1
## X358        -1.465e-02  9.832e+02       0        1
## X359         3.236e-02  1.267e+03       0        1
## X360         2.001e-02  1.891e+03       0        1
## X361        -1.195e-02  1.993e+03       0        1
## X362        -7.502e-03  9.372e+03       0        1
## X363                NA         NA      NA       NA
## X364                NA         NA      NA       NA
## X365                NA         NA      NA       NA
## X366                NA         NA      NA       NA
## X367                NA         NA      NA       NA
## X368                NA         NA      NA       NA
## X369        -8.508e-03  3.469e+03       0        1
## X370         2.577e-03  2.024e+03       0        1
## X371         5.302e-02  1.534e+03       0        1
## X372         1.095e-02  1.422e+03       0        1
## X373        -2.364e-02  1.074e+03       0        1
## X374        -3.081e-02  1.114e+03       0        1
## X375         1.171e-02  1.048e+03       0        1
## X376        -5.214e-03  9.365e+02       0        1
```

```
## X377        1.275e-02  6.310e+02      0        1
## X378        3.057e-03  5.830e+02      0        1
## X379        1.552e-02  6.837e+02      0        1
## X380        1.231e-03  7.947e+02      0        1
## X381        6.598e-03  6.831e+02      0        1
## X382        3.058e-02  8.894e+02      0        1
## X383       -2.637e-03  1.108e+03      0        1
## X384       -2.397e-02  1.150e+03      0        1
## X385       -3.775e-03  1.179e+03      0        1
## X386       -1.452e-02  1.177e+03      0        1
## X387       -3.589e-02  1.203e+03      0        1
## X388       -1.669e-02  1.833e+03      0        1
## X389        9.258e-04  1.863e+03      0        1
## X390        1.788e-02  1.071e+04      0        1
## X391               NA         NA     NA       NA
## X392               NA         NA     NA       NA
## X393               NA         NA     NA       NA
## X394               NA         NA     NA       NA
## X395               NA         NA     NA       NA
## X396               NA         NA     NA       NA
## X397       -2.049e-02  2.791e+03      0        1
## X398       -1.619e-02  1.630e+03      0        1
## X399       -4.268e-02  1.605e+03      0        1
## X400       -4.893e-02  1.474e+03      0        1
## X401        3.861e-03  9.924e+02      0        1
## X402        1.507e-03  1.167e+03      0        1
## X403        1.390e-02  1.006e+03      0        1
## X404       -1.406e-02  9.836e+02      0        1
## X405        2.559e-02  6.603e+02      0        1
## X406       -1.759e-02  6.519e+02      0        1
## X407        4.405e-02  8.187e+02      0        1
## X408       -5.560e-03  7.225e+02      0        1
## X409       -1.923e-04  7.457e+02      0        1
## X410       -3.165e-02  1.241e+03      0        1
## X411       -2.356e-02  1.095e+03      0        1
## X412       -6.082e-03  1.025e+03      0        1
## X413        2.631e-02  1.155e+03      0        1
## X414        2.296e-02  1.109e+03      0        1
## X415        2.547e-02  1.299e+03      0        1
## X416        3.516e-02  1.629e+03      0        1
## X417        1.599e-02  1.537e+03      0        1
## X418       -1.810e-02  1.071e+04      0        1
## X419               NA         NA     NA       NA
## X420               NA         NA     NA       NA
## X421               NA         NA     NA       NA
## X422               NA         NA     NA       NA
## X423               NA         NA     NA       NA
## X424               NA         NA     NA       NA
## X425       -1.622e-02  2.124e+03      0        1
## X426       -9.319e-03  1.336e+03      0        1
## X427        2.856e-02  1.652e+03      0        1
## X428        3.291e-02  1.462e+03      0        1
## X429       -7.305e-03  1.125e+03      0        1
## X430       -2.198e-02  1.097e+03      0        1
```

```
## X431         -9.451e-03  9.749e+02       0        1
## X432          4.065e-03  8.059e+02       0        1
## X433         -2.089e-02  6.751e+02       0        1
## X434          1.607e-02  6.401e+02       0        1
## X435          1.101e-02  8.542e+02       0        1
## X436          7.526e-03  6.035e+02       0        1
## X437          1.220e-03  8.229e+02       0        1
## X438          2.550e-02  1.218e+03       0        1
## X439          1.043e-02  1.240e+03       0        1
## X440          1.739e-02  9.015e+02       0        1
## X441         -1.657e-02  9.392e+02       0        1
## X442         -5.730e-03  1.029e+03       0        1
## X443         -1.471e-02  1.367e+03       0        1
## X444         -2.652e-02  1.556e+03       0        1
## X445         -2.577e-02  1.739e+03       0        1
## X446          1.416e-01  8.654e+03       0        1
## X447         -8.720e-02  4.907e+03       0        1
## X448                 NA          NA      NA       NA
## X449                 NA          NA      NA       NA
## X450          8.321e-02  9.721e+03       0        1
## X451                 NA          NA      NA       NA
## X452          1.945e-01  2.537e+04       0        1
## X453          1.507e-02  2.441e+03       0        1
## X454          6.605e-04  1.369e+03       0        1
## X455         -3.682e-02  1.307e+03       0        1
## X456         -2.334e-03  1.337e+03       0        1
## X457         -1.287e-02  1.077e+03       0        1
## X458         -2.766e-02  1.035e+03       0        1
## X459         -2.352e-02  8.791e+02       0        1
## X460         -8.006e-03  8.339e+02       0        1
## X461          5.935e-03  7.267e+02       0        1
## X462          7.407e-03  6.838e+02       0        1
## X463         -2.640e-03  6.998e+02       0        1
## X464         -3.426e-03  5.826e+02       0        1
## X465          5.419e-03  7.749e+02       0        1
## X466         -4.465e-03  9.199e+02       0        1
## X467         -2.883e-02  9.327e+02       0        1
## X468         -2.110e-02  1.005e+03       0        1
## X469         -2.143e-03  1.001e+03       0        1
## X470          8.483e-03  9.767e+02       0        1
## X471         -7.994e-03  1.225e+03       0        1
## X472          3.032e-02  1.564e+03       0        1
## X473         -3.245e-02  2.450e+03       0        1
## X474         -2.928e-01  1.383e+04       0        1
## X475                 NA          NA      NA       NA
## X476                 NA          NA      NA       NA
## X477                 NA          NA      NA       NA
## X478                 NA          NA      NA       NA
## X479                 NA          NA      NA       NA
## X480         -7.513e-01  1.209e+05       0        1
## X481         -1.419e-02  2.364e+03       0        1
## X482          2.265e-02  1.457e+03       0        1
## X483          1.530e-02  1.438e+03       0        1
## X484         -8.629e-03  1.167e+03       0        1
```

```
## X485          1.927e-02  8.543e+02       0         1
## X486          1.004e-02  8.906e+02       0         1
## X487          6.914e-03  8.488e+02       0         1
## X488          2.825e-02  7.087e+02       0         1
## X489          7.320e-03  6.668e+02       0         1
## X490          3.433e-03  6.912e+02       0         1
## X491         -2.675e-03  6.182e+02       0         1
## X492         -1.920e-02  6.830e+02       0         1
## X493         -1.243e-02  6.128e+02       0         1
## X494          9.029e-04  8.814e+02       0         1
## X495          1.275e-02  9.432e+02       0         1
## X496         -5.568e-04  9.288e+02       0         1
## X497          2.432e-03  1.073e+03       0         1
## X498         -1.410e-02  1.104e+03       0         1
## X499         -2.544e-02  1.361e+03       0         1
## X500         -4.842e-02  1.831e+03       0         1
## X501          5.101e-02  3.108e+03       0         1
## X502          5.790e-02  2.183e+04       0         1
## X503                 NA         NA      NA        NA
## X504                 NA         NA      NA        NA
## X505                 NA         NA      NA        NA
## X506                 NA         NA      NA        NA
## X507                 NA         NA      NA        NA
## X508          3.519e-01  7.262e+04       0         1
## X509         -1.435e-02  2.097e+03       0         1
## X510         -2.209e-02  1.279e+03       0         1
## X511         -1.844e-02  1.414e+03       0         1
## X512         -5.933e-03  1.011e+03       0         1
## X513         -1.704e-03  7.840e+02       0         1
## X514         -1.453e-04  7.348e+02       0         1
## X515         -3.628e-03  7.648e+02       0         1
## X516         -1.462e-02  6.398e+02       0         1
## X517          1.552e-02  6.256e+02       0         1
## X518          2.765e-03  5.958e+02       0         1
## X519          1.128e-02  4.964e+02       0         1
## X520          4.699e-03  6.467e+02       0         1
## X521         -3.542e-03  6.421e+02       0         1
## X522         -1.830e-02  9.086e+02       0         1
## X523         -1.430e-02  8.678e+02       0         1
## X524         -6.651e-03  9.682e+02       0         1
## X525          8.793e-03  1.040e+03       0         1
## X526          6.003e-04  1.213e+03       0         1
## X527          3.405e-02  1.454e+03       0         1
## X528          3.536e-02  2.120e+03       0         1
## X529          2.293e-02  4.675e+03       0         1
## X530          4.012e-01  3.001e+04       0         1
## X531                 NA         NA      NA        NA
## X532                 NA         NA      NA        NA
## X533                 NA         NA      NA        NA
## X534                 NA         NA      NA        NA
## X535                 NA         NA      NA        NA
## X536                 NA         NA      NA        NA
## X537          9.707e-03  2.339e+03       0         1
## X538         -9.948e-03  1.152e+03       0         1
```

```
## X539          2.460e-02  1.103e+03        0        1
## X540          5.787e-03  8.101e+02        0        1
## X541          3.552e-03  7.600e+02        0        1
## X542         -4.964e-03  5.915e+02        0        1
## X543         -7.306e-04  6.910e+02        0        1
## X544          1.438e-02  5.536e+02        0        1
## X545         -2.548e-02  5.199e+02        0        1
## X546          1.457e-02  5.333e+02        0        1
## X547          1.224e-03  5.072e+02        0        1
## X548          7.234e-03  5.858e+02        0        1
## X549          1.163e-02  6.423e+02        0        1
## X550         -3.253e-04  7.430e+02        0        1
## X551          1.120e-03  8.514e+02        0        1
## X552         -3.503e-02  9.732e+02        0        1
## X553         -8.627e-03  1.067e+03        0        1
## X554          1.181e-03  1.281e+03        0        1
## X555         -6.112e-02  2.532e+03        0        1
## X556         -2.638e-02  2.313e+03        0        1
## X557          2.675e-03  7.238e+03        0        1
## X558         -8.343e-01  6.197e+04        0        1
## X559         -7.513e-01  2.476e+04        0        1
## X560                NA         NA       NA       NA
## X561                NA         NA       NA       NA
## X562                NA         NA       NA       NA
## X563                NA         NA       NA       NA
## X564          1.852e-02  2.728e+04        0        1
## X565         -4.219e-02  2.173e+03        0        1
## X566          3.156e-03  1.171e+03        0        1
## X567         -2.890e-02  1.035e+03        0        1
## X568          7.063e-03  7.814e+02        0        1
## X569          5.807e-04  6.395e+02        0        1
## X570         -2.057e-03  5.606e+02        0        1
## X571          1.033e-02  6.471e+02        0        1
## X572         -9.806e-03  5.242e+02        0        1
## X573          1.228e-02  4.759e+02        0        1
## X574         -2.610e-03  5.026e+02        0        1
## X575          4.497e-03  4.572e+02        0        1
## X576         -1.597e-03  5.822e+02        0        1
## X577          6.507e-03  6.073e+02        0        1
## X578         -8.223e-03  7.979e+02        0        1
## X579          8.688e-03  9.107e+02        0        1
## X580          1.640e-02  1.022e+03        0        1
## X581         -6.681e-03  1.303e+03        0        1
## X582         -1.142e-02  2.097e+03        0        1
## X583         -2.152e-02  2.504e+03        0        1
## X584          5.660e-02  4.123e+03        0        1
## X585         -1.501e-02  2.855e+04        0        1
## X586                NA         NA       NA       NA
## X587                NA         NA       NA       NA
## X588                NA         NA       NA       NA
## X589                NA         NA       NA       NA
## X590                NA         NA       NA       NA
## X591                NA         NA       NA       NA
## X592          1.025e-02  6.842e+03        0        1
```

```
## X593          9.289e-02  2.479e+03       0        1
## X594          1.797e-03  1.134e+03       0        1
## X595          4.283e-02  1.026e+03       0        1
## X596          2.983e-03  7.835e+02       0        1
## X597         -1.514e-03  6.272e+02       0        1
## X598         -5.017e-03  6.376e+02       0        1
## X599         -4.652e-03  5.296e+02       0        1
## X600         -6.812e-03  4.817e+02       0        1
## X601          1.238e-02  4.880e+02       0        1
## X602         -1.111e-02  4.823e+02       0        1
## X603          1.230e-02  4.726e+02       0        1
## X604          1.749e-03  5.826e+02       0        1
## X605         -2.944e-03  5.536e+02       0        1
## X606          2.570e-02  7.403e+02       0        1
## X607          2.724e-03  8.623e+02       0        1
## X608         -1.435e-02  1.326e+03       0        1
## X609          7.445e-02  1.338e+03       0        1
## X610          3.745e-02  2.574e+03       0        1
## X611          2.310e-02  3.789e+03       0        1
## X612         -1.015e-01  3.695e+03       0        1
## X613          8.179e-01  4.369e+04       0        1
## X614                 NA         NA      NA       NA
## X615                 NA         NA      NA       NA
## X616                 NA         NA      NA       NA
## X617                 NA         NA      NA       NA
## X618                 NA         NA      NA       NA
## X619                 NA         NA      NA       NA
## X620         -1.333e-01  3.721e+04       0        1
## X621         -1.046e-01  4.095e+03       0        1
## X622         -2.180e-02  1.406e+03       0        1
## X623         -7.354e-03  9.112e+02       0        1
## X624         -5.157e-03  8.384e+02       0        1
## X625          8.268e-03  6.697e+02       0        1
## X626         -2.186e-03  5.907e+02       0        1
## X627         -6.450e-03  4.872e+02       0        1
## X628          5.676e-03  4.733e+02       0        1
## X629         -7.352e-03  4.404e+02       0        1
## X630          1.823e-03  4.325e+02       0        1
## X631          6.319e-04  4.849e+02       0        1
## X632          6.076e-03  5.656e+02       0        1
## X633         -9.979e-03  6.289e+02       0        1
## X634         -1.164e-02  8.703e+02       0        1
## X635          2.198e-02  9.163e+02       0        1
## X636         -1.016e-01  1.802e+03       0        1
## X637          1.340e-02  2.134e+03       0        1
## X638         -3.997e-02  3.491e+03       0        1
## X639          1.852e-02  4.618e+03       0        1
## X640         -4.730e-01  4.905e+04       0        1
## X641                 NA         NA      NA       NA
## X642                 NA         NA      NA       NA
## X643                 NA         NA      NA       NA
## X644                 NA         NA      NA       NA
## X645                 NA         NA      NA       NA
## X646                 NA         NA      NA       NA
```

```
## X647                NA         NA       NA       NA
## X648                NA         NA       NA       NA
## X649         1.343e-01  4.653e+03        0        1
## X650        -3.763e-03  1.802e+03        0        1
## X651         1.746e-02  1.124e+03        0        1
## X652        -1.096e-02  7.784e+02        0        1
## X653         1.070e-05  6.224e+02        0        1
## X654        -7.312e-03  5.404e+02        0        1
## X655         5.874e-03  4.452e+02        0        1
## X656        -3.974e-05  4.299e+02        0        1
## X657        -1.070e-02  3.883e+02        0        1
## X658        -3.786e-03  3.835e+02        0        1
## X659        -2.533e-04  4.127e+02        0        1
## X660        -2.951e-02  4.554e+02        0        1
## X661         2.867e-02  6.193e+02        0        1
## X662        -4.598e-03  8.618e+02        0        1
## X663         7.300e-02  1.856e+03        0        1
## X664        -3.461e-03  2.229e+03        0        1
## X665        -5.124e-02  2.833e+03        0        1
## X666         1.533e-01  1.073e+04        0        1
## X667        -1.106e-01  8.462e+04        0        1
## X668         3.561e+00  4.159e+05        0        1
## X669                NA         NA       NA       NA
## X670                NA         NA       NA       NA
## X671                NA         NA       NA       NA
## X672                NA         NA       NA       NA
## X673                NA         NA       NA       NA
## X674                NA         NA       NA       NA
## X675                NA         NA       NA       NA
## X676                NA         NA       NA       NA
## X677        -1.504e-01  1.731e+04        0        1
## X678         1.815e-03  2.677e+03        0        1
## X679        -2.335e-02  1.231e+03        0        1
## X680         1.079e-02  8.901e+02        0        1
## X681        -1.175e-02  6.469e+02        0        1
## X682         1.779e-02  6.037e+02        0        1
## X683        -2.270e-02  6.139e+02        0        1
## X684         3.281e-03  5.648e+02        0        1
## X685         1.013e-02  5.052e+02        0        1
## X686        -1.065e-02  4.503e+02        0        1
## X687         6.623e-03  4.809e+02        0        1
## X688         4.802e-03  5.484e+02        0        1
## X689        -2.401e-02  7.078e+02        0        1
## X690         1.997e-02  9.664e+02        0        1
## X691        -5.122e-02  1.650e+03        0        1
## X692         1.190e-01  7.667e+03        0        1
## X693        -6.088e-02  5.290e+03        0        1
## X694        -1.967e-01  5.054e+04        0        1
## X695         1.169e-01  7.125e+04        0        1
## X696                NA         NA       NA       NA
## X697                NA         NA       NA       NA
## X698                NA         NA       NA       NA
## X699                NA         NA       NA       NA
## X700                NA         NA       NA       NA
```

```
## X701                 NA         NA       NA       NA
## X702                 NA         NA       NA       NA
## X703                 NA         NA       NA       NA
## X704                 NA         NA       NA       NA
## X705                 NA         NA       NA       NA
## X706                 NA         NA       NA       NA
## X707          1.214e+00  1.529e+05        0        1
## X708         -1.191e+00  5.316e+04        0        1
## X709          5.589e-01  3.128e+04        0        1
## X710          2.752e-01  1.999e+04        0        1
## X711         -9.701e-02  2.121e+04        0        1
## X712         -9.234e-02  1.272e+04        0        1
## X713          6.354e-02  4.142e+03        0        1
## X714          1.547e-01  7.150e+03        0        1
## X715         -1.231e-01  6.146e+03        0        1
## X716          3.333e-02  3.071e+03        0        1
## X717          1.982e-01  1.581e+04        0        1
## X718         -4.929e-01  4.425e+04        0        1
## X719          1.572e+00  3.762e+05        0        1
## X720                 NA         NA       NA       NA
## X721                 NA         NA       NA       NA
## X722                 NA         NA       NA       NA
## X723                 NA         NA       NA       NA
## X724                 NA         NA       NA       NA
## X725                 NA         NA       NA       NA
## X726                 NA         NA       NA       NA
## X727                 NA         NA       NA       NA
## X728                 NA         NA       NA       NA
## X729                 NA         NA       NA       NA
## X730                 NA         NA       NA       NA
## X731                 NA         NA       NA       NA
## X732                 NA         NA       NA       NA
## X733                 NA         NA       NA       NA
## X734                 NA         NA       NA       NA
## X735                 NA         NA       NA       NA
## X736                 NA         NA       NA       NA
## X737                 NA         NA       NA       NA
## X738                 NA         NA       NA       NA
## X739                 NA         NA       NA       NA
## X740                 NA         NA       NA       NA
## X741                 NA         NA       NA       NA
## X742                 NA         NA       NA       NA
## X743                 NA         NA       NA       NA
## X744                 NA         NA       NA       NA
## X745                 NA         NA       NA       NA
## X746                 NA         NA       NA       NA
## X747                 NA         NA       NA       NA
## X748                 NA         NA       NA       NA
## X749                 NA         NA       NA       NA
## X750                 NA         NA       NA       NA
## X751                 NA         NA       NA       NA
## X752                 NA         NA       NA       NA
## X753                 NA         NA       NA       NA
## X754                 NA         NA       NA       NA
```

```
## X755                   NA          NA       NA        NA
## X756                   NA          NA       NA        NA
## X757                   NA          NA       NA        NA
## X758                   NA          NA       NA        NA
## X759                   NA          NA       NA        NA
## X760                   NA          NA       NA        NA
## X761                   NA          NA       NA        NA
## X762                   NA          NA       NA        NA
## X763                   NA          NA       NA        NA
## X764                   NA          NA       NA        NA
## X765                   NA          NA       NA        NA
## X766                   NA          NA       NA        NA
## X767                   NA          NA       NA        NA
## X768                   NA          NA       NA        NA
## X769                   NA          NA       NA        NA
## X770                   NA          NA       NA        NA
## X771                   NA          NA       NA        NA
## X772                   NA          NA       NA        NA
## X773                   NA          NA       NA        NA
## X774                   NA          NA       NA        NA
## X775                   NA          NA       NA        NA
## X776                   NA          NA       NA        NA
## X777                   NA          NA       NA        NA
## X778                   NA          NA       NA        NA
## X779                   NA          NA       NA        NA
## X780                   NA          NA       NA        NA
## X781                   NA          NA       NA        NA
## X782                   NA          NA       NA        NA
## X783                   NA          NA       NA        NA
## X784                   NA          NA       NA        NA
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2.9206e+03  on 2114  degrees of freedom
## Residual deviance: 1.8922e-08  on 1616  degrees of freedom
## AIC: 998
##
## Number of Fisher Scoring iterations: 25
```

The message "glm.fit: algorithm did not convergeglm.fit: fitted probabilities numerically 0 or 1 occurred" implies numerical problems which is caused by too many features or explanatory variables.

The estimated coefficients and standard errors of some features are NA's such as X1, X20, X10 and so on. This is due to that they have zero variability (so they are not important to the model). Meanwhile, other features have very small estimate coefficients, very large standard errors and p-values of 1 (so they are not statistically significant) such as X100, X109, X69 and so on. The approach fails because we can't choose all variables, we have to prune or eliminate those with variability of 0.