NavieBayes

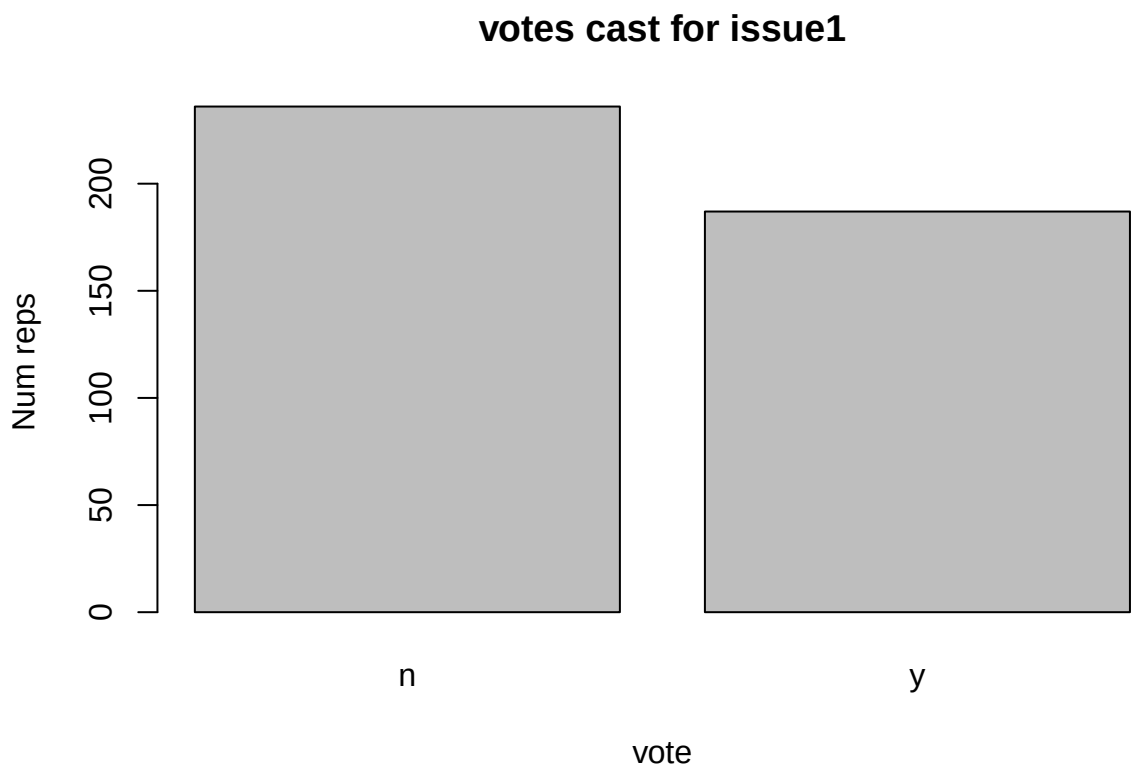
```
library(mlbench)
```

```
## Warning: package 'mlbench' was built under R version 4.0.2
```

```
data(HouseVotes84)
```

Including Plots

You can also embed plots, for example:




```
HouseVotes84$Class
```

```
##   [1] republican republican democrat   democrat   democrat   democrat
##   [7] democrat   republican republican democrat   republican republican
##  [13] democrat   democrat   republican republican democrat   democrat
##  [19] republican democrat   democrat   democrat   democrat   democrat
```

```
##  [25] democrat   democrat   democrat   democrat   republican democrat
##  [31] republican democrat   democrat   republican democrat   republican
##  [37] republican republican republican democrat   democrat   democrat
##  [43] democrat   democrat   democrat   democrat   democrat   democrat
##  [49] democrat   republican democrat   republican democrat   republican
##  [55] democrat   republican republican republican republican republican
##  [61] democrat   republican democrat   democrat   democrat   republican
##  [67] republican republican democrat   democrat   democrat   republican
##  [73] democrat   republican democrat   democrat   democrat   democrat
##  [79] democrat   republican democrat   democrat   republican republican
##  [85] republican democrat   republican republican democrat   republican
##  [91] democrat   democrat   democrat   democrat   democrat   democrat
##  [97] democrat   democrat   democrat   republican democrat   democrat
## [103] democrat   democrat   democrat   democrat   republican republican
## [109] democrat   democrat   democrat   republican democrat   republican
## [115] democrat   democrat   democrat   republican democrat   republican
## [121] republican republican republican republican democrat   republican
## [127] republican democrat   democrat   democrat   democrat   democrat
## [133] republican republican republican republican republican democrat
## [139] democrat   democrat   republican republican republican democrat
## [145] democrat   democrat   republican democrat   republican democrat
## [151] republican democrat   democrat   democrat   republican republican
## [157] republican democrat   republican democrat   democrat   democrat
## [163] democrat   republican democrat   democrat   republican republican
## [169] democrat   democrat   democrat   republican democrat   democrat
## [175] democrat   democrat   republican democrat   democrat   democrat
## [181] democrat   democrat   democrat   democrat   democrat   democrat
## [187] democrat   democrat   republican democrat   republican republican
## [193] democrat   democrat   democrat   republican democrat   republican
## [199] democrat   democrat   democrat   democrat   democrat   democrat
## [205] republican democrat   republican republican democrat   democrat
## [211] democrat   republican democrat   democrat   republican democrat
## [217] democrat   republican democrat   democrat   democrat   democrat
## [223] democrat   republican republican republican democrat   democrat
## [229] republican republican republican republican democrat   republican
## [235] democrat   republican democrat   democrat   democrat   republican
## [241] republican democrat   republican democrat   democrat   democrat
## [247] democrat   republican republican democrat   republican republican
## [253] democrat   republican democrat   democrat   republican republican
## [259] democrat   democrat   democrat   democrat   democrat   democrat
## [265] democrat   democrat   republican republican democrat   democrat
## [271] democrat   democrat   democrat   republican republican democrat
## [277] republican republican republican republican democrat   republican
## [283] republican republican democrat   democrat   democrat   democrat
## [289] democrat   democrat   democrat   democrat   democrat   democrat
## [295] democrat   republican republican democrat   democrat   democrat
## [301] republican democrat   republican republican republican republican
## [307] republican democrat   republican democrat   republican democrat
## [313] democrat   republican republican republican democrat   democrat
## [319] democrat   democrat   democrat   democrat   democrat   democrat
## [325] republican democrat   democrat   republican democrat   democrat
## [331] republican democrat   democrat   democrat   democrat   republican
## [337] democrat   democrat   democrat   republican republican democrat
## [343] democrat   republican democrat   republican republican republican
```

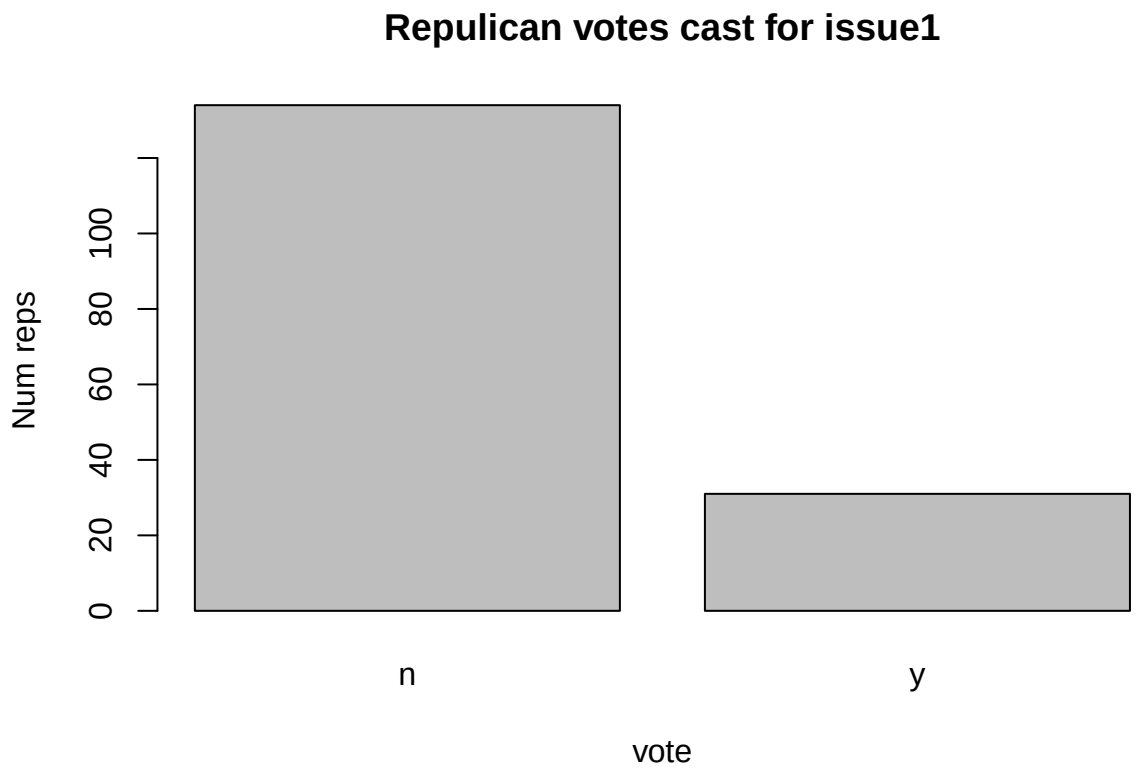
```
## [349] democrat   republican democrat   republican democrat   republican
## [355] democrat   republican republican republican democrat   republican
## [361] democrat   democrat   democrat   republican republican democrat
## [367] democrat   democrat   democrat   republican democrat   democrat
## [373] democrat   democrat   republican democrat   democrat   republican
## [379] republican republican democrat   democrat   democrat   democrat
## [385] democrat   democrat   democrat   democrat   democrat   democrat
## [391] democrat   democrat   republican republican democrat   democrat
## [397] democrat   democrat   democrat   republican republican republican
## [403] republican republican republican republican democrat   democrat
## [409] democrat   republican republican democrat   republican republican
## [415] democrat   democrat   republican democrat   democrat   democrat
## [421] republican democrat   democrat   democrat   democrat   democrat
## [427] democrat   republican democrat   democrat   republican democrat
## [433] republican republican republican
## Levels: democrat republican
```

```
Repub <- HouseVotes84$Class=="republican"
Democrat <- HouseVotes84$Class=="democrat"
Repub
```

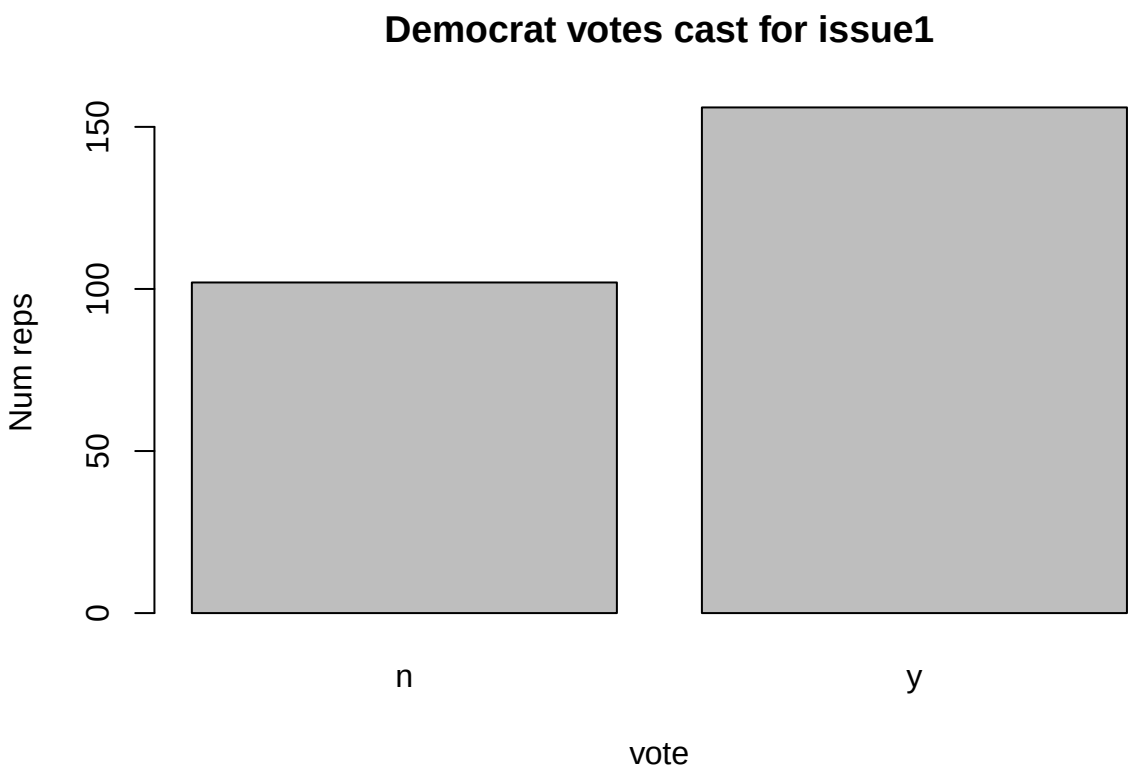
```
##   [1]  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE  TRUE  TRUE
##  [13] FALSE FALSE  TRUE  TRUE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE
##  [25] FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE FALSE FALSE  TRUE FALSE  TRUE
##  [37]  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [49] FALSE  TRUE FALSE  TRUE FALSE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE
##  [61] FALSE  TRUE FALSE FALSE FALSE  TRUE  TRUE  TRUE FALSE FALSE FALSE  TRUE
##  [73] FALSE  TRUE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE  TRUE  TRUE
##  [85]  TRUE FALSE  TRUE  TRUE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE
##  [97] FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE
## [109] FALSE FALSE FALSE  TRUE FALSE  TRUE FALSE FALSE FALSE  TRUE FALSE  TRUE
## [121]  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE
## [133]  TRUE  TRUE  TRUE  TRUE  TRUE FALSE FALSE FALSE  TRUE  TRUE  TRUE FALSE
## [145] FALSE FALSE  TRUE FALSE  TRUE FALSE  TRUE FALSE FALSE FALSE  TRUE  TRUE
## [157]  TRUE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE  TRUE  TRUE
## [169] FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE
## [181] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE  TRUE
## [193] FALSE FALSE FALSE  TRUE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE
## [205]  TRUE FALSE  TRUE  TRUE FALSE FALSE FALSE  TRUE FALSE FALSE  TRUE FALSE
## [217] FALSE  TRUE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE FALSE FALSE
## [229]  TRUE  TRUE  TRUE  TRUE FALSE  TRUE FALSE  TRUE FALSE FALSE FALSE  TRUE
## [241]  TRUE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE  TRUE  TRUE
## [253] FALSE  TRUE FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE
## [265] FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE
## [277]  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE
## [289] FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE
## [301]  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE FALSE  TRUE FALSE
## [313] FALSE  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [325]  TRUE FALSE FALSE  TRUE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE
## [337] FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE  TRUE FALSE  TRUE  TRUE  TRUE
## [349] FALSE  TRUE FALSE  TRUE FALSE  TRUE FALSE  TRUE  TRUE  TRUE FALSE  TRUE
## [361] FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE
## [373] FALSE FALSE  TRUE FALSE FALSE  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE
## [385] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE
```

```
## [397] FALSE FALSE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE FALSE
## [409] FALSE  TRUE  TRUE FALSE  TRUE  TRUE FALSE FALSE  TRUE FALSE FALSE FALSE
## [421]  TRUE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE  TRUE FALSE
## [433]  TRUE  TRUE  TRUE
```

```
plot(as.factor(HouseVotes84[Repub,2]))
title(main="Repulican votes cast for issue1",xlab = "vote",ylab = "Num reps")
```




```
plot(as.factor(HouseVotes84[Democrat,2]))
title(main="Democrat votes cast for issue1",xlab = "vote",ylab = "Num reps")
```




```
na_by_col_class <-function(col,cls){
  return(sum(is.na(HouseVotes84[,col]) & HouseVotes84$Class == cls))
}
na_by_col_class
```

```
## function(col,cls){
##   return(sum(is.na(HouseVotes84[,col]) & HouseVotes84$Class == cls))
## }
```

```
p_y_col_class<- function(col,cls){
  sum_y<-sum(HouseVotes84[,col]=="y"& HouseVotes84$Class == cls,na.rm = TRUE)
  sum_n<-sum(HouseVotes84[,col]=="n"& HouseVotes84$Class == cls,na.rm = TRUE)
  return(sum_y/(sum_y+sum_n))
}
```

```
p_y_col_class(5,"democrat")
```

```
## [1] 0.05405405
```

```
p_y_col_class(5,"republican")
```

```
## [1] 0.9878788
```

```
na_by_col_class(2,"democrat")
```

```
## [1] 9
```

```
na_by_col_class(2,"republican")
```

```
## [1] 3
```

```
#impute the missing values,
#If the republican congresman didn't vote then we are allocating 'y' or 'n' based on if their
#party voted 'y'  or 'n'
for(i in 2:ncol(HouseVotes84)){
  if(sum(is.na(HouseVotes84[,i])>0)){
    c1 <- which(is.na(HouseVotes84[,i])&HouseVotes84$Class == "democrat",arr.ind = TRUE )
    c2 <- which(is.na(HouseVotes84[,i])&HouseVotes84$Class == "repiblican",arr.ind = TRUE )
    HouseVotes84[c1,i] <-
      ifelse(runif(na_by_col_class(i,"democrat"))<p_y_col_class(i,"democrat"),"y","n")
    HouseVotes84[c2,i] <-
      ifelse(runif(na_by_col_class(i,"republican"))<p_y_col_class(i,"republican"),"y","n")
  }
}
```

#divide the data into test and training sets #create new col "train" and assign 1 or 0 in 80/20 proportion via random uniform dist

```
HouseVotes84[,"train"] <- ifelse(runif(nrow(HouseVotes84))<0.80,1,0)
```

#get col number of train /test indicator column (needed later)

```
trainColNum <- grep("train",names(HouseVotes84))
```

#separate training and test sets and remove training column before modeling

```
trainHouseVotes84 <- HouseVotes84[HouseVotes84$train == 1 , -trainColNum]
testHouseVotes84 <- HouseVotes84[HouseVotes84$train == 0 , -trainColNum]
```

#load e1071 library and invoke naiveBayes method

```
library(e1071)
```

```
## Warning: package 'e1071' was built under R version 4.0.2
```

```
nb_model <- naiveBayes(Class~.,data = trainHouseVotes84)
```

#In this notation the dependent variable to be predicted appears on the left hand side of the ~ and the independent variable in right hand side

```
nb_model
```

```
##
## Naive Bayes Classifier for Discrete Predictors
##
## Call:
## naiveBayes.default(x = X, y = Y, laplace = laplace)
##
## A-priori probabilities:
## Y
##   democrat republican
##  0.6133721  0.3866279
##
## Conditional probabilities:
##             V1
## Y                    n         y
##   democrat   0.3886256 0.6113744
##   republican 0.8384615 0.1615385
##
##             V2
## Y                    n         y
##   democrat   0.5213270 0.4786730
##   republican 0.4830508 0.5169492
##
##             V3
## Y                     n          y
##   democrat   0.09952607 0.90047393
##   republican 0.84496124 0.15503876
##
##             V4
## Y                    n         y
##   democrat   0.9478673 0.0521327
##   republican 0.0000000 1.0000000
##
##             V5
## Y                     n          y
##   democrat   0.78199052 0.21800948
##   republican 0.05384615 0.94615385
##
##             V6
## Y                    n         y
##   democrat   0.5308057 0.4691943
##   republican 0.1068702 0.8931298
##
##             V7
## Y                    n         y
##   democrat   0.2227488 0.7772512
##   republican 0.7559055 0.2440945
##
##             V8
## Y                    n         y
##   democrat   0.1753555 0.8246445
##   republican 0.8524590 0.1475410
```

```
##
##             V9
## Y                    n         y
##   democrat   0.2511848 0.7488152
##   republican 0.8846154 0.1153846
##
##             V10
## Y                    n         y
##   democrat   0.5308057 0.4691943
##   republican 0.4461538 0.5538462
##
##             V11
## Y                    n         y
##   democrat   0.5118483 0.4881517
##   republican 0.8629032 0.1370968
##
##             V12
## Y                    n         y
##   democrat   0.8483412 0.1516588
##   republican 0.1393443 0.8606557
##
##             V13
## Y                    n         y
##   democrat   0.7061611 0.2938389
##   republican 0.1370968 0.8629032
##
##             V14
## Y                     n          y
##   democrat   0.66824645 0.33175355
##   republican 0.02380952 0.97619048
##
##             V15
## Y                     n          y
##   democrat   0.38388626 0.61611374
##   republican 0.90243902 0.09756098
##
##             V16
## Y                     n          y
##   democrat   0.07582938 0.92417062
##   republican 0.37719298 0.62280702
```

```
summary(nb_model)
```

```
##           Length Class  Mode
## apriori    2     table  numeric
## tables    16     -none- list
## levels     2     -none- character
## isnumeric 16     -none- logical
## call       4     -none- call
```

```
str(nb_model)
```

```
## List of 5
```

```
##  $ apriori  : 'table' int [1:2(1d)] 211 133
##   ..- attr(*, "dimnames")=List of 1
##   .. ..$ Y: chr [1:2] "democrat" "republican"
##  $ tables   :List of 16
##   ..$ V1 : 'table' num [1:2, 1:2] 0.389 0.838 0.611 0.162
##   .. ..- attr(*, "dimnames")=List of 2
##   .. .. ..$ Y : chr [1:2] "democrat" "republican"
##   .. .. ..$ V1: chr [1:2] "n" "y"
##   ..$ V2 : 'table' num [1:2, 1:2] 0.521 0.483 0.479 0.517
##   .. ..- attr(*, "dimnames")=List of 2
##   .. .. ..$ Y : chr [1:2] "democrat" "republican"
##   .. .. ..$ V2: chr [1:2] "n" "y"
##   ..$ V3 : 'table' num [1:2, 1:2] 0.0995 0.845 0.9005 0.155
##   .. ..- attr(*, "dimnames")=List of 2
##   .. .. ..$ Y : chr [1:2] "democrat" "republican"
##   .. .. ..$ V3: chr [1:2] "n" "y"
##   ..$ V4 : 'table' num [1:2, 1:2] 0.9479 0 0.0521 1
##   .. ..- attr(*, "dimnames")=List of 2
##   .. .. ..$ Y : chr [1:2] "democrat" "republican"
##   .. .. ..$ V4: chr [1:2] "n" "y"
##   ..$ V5 : 'table' num [1:2, 1:2] 0.782 0.0538 0.218 0.9462
##   .. ..- attr(*, "dimnames")=List of 2
##   .. .. ..$ Y : chr [1:2] "democrat" "republican"
##   .. .. ..$ V5: chr [1:2] "n" "y"
##   ..$ V6 : 'table' num [1:2, 1:2] 0.531 0.107 0.469 0.893
##   .. ..- attr(*, "dimnames")=List of 2
##   .. .. ..$ Y : chr [1:2] "democrat" "republican"
##   .. .. ..$ V6: chr [1:2] "n" "y"
##   ..$ V7 : 'table' num [1:2, 1:2] 0.223 0.756 0.777 0.244
##   .. ..- attr(*, "dimnames")=List of 2
##   .. .. ..$ Y : chr [1:2] "democrat" "republican"
##   .. .. ..$ V7: chr [1:2] "n" "y"
##   ..$ V8 : 'table' num [1:2, 1:2] 0.175 0.852 0.825 0.148
##   .. ..- attr(*, "dimnames")=List of 2
##   .. .. ..$ Y : chr [1:2] "democrat" "republican"
##   .. .. ..$ V8: chr [1:2] "n" "y"
##   ..$ V9 : 'table' num [1:2, 1:2] 0.251 0.885 0.749 0.115
##   .. ..- attr(*, "dimnames")=List of 2
##   .. .. ..$ Y : chr [1:2] "democrat" "republican"
##   .. .. ..$ V9: chr [1:2] "n" "y"
##   ..$ V10: 'table' num [1:2, 1:2] 0.531 0.446 0.469 0.554
##   .. ..- attr(*, "dimnames")=List of 2
##   .. .. ..$ Y  : chr [1:2] "democrat" "republican"
##   .. .. ..$ V10: chr [1:2] "n" "y"
##   ..$ V11: 'table' num [1:2, 1:2] 0.512 0.863 0.488 0.137
##   .. ..- attr(*, "dimnames")=List of 2
##   .. .. ..$ Y  : chr [1:2] "democrat" "republican"
##   .. .. ..$ V11: chr [1:2] "n" "y"
##   ..$ V12: 'table' num [1:2, 1:2] 0.848 0.139 0.152 0.861
##   .. ..- attr(*, "dimnames")=List of 2
##   .. .. ..$ Y  : chr [1:2] "democrat" "republican"
##   .. .. ..$ V12: chr [1:2] "n" "y"
##   ..$ V13: 'table' num [1:2, 1:2] 0.706 0.137 0.294 0.863
##   .. ..- attr(*, "dimnames")=List of 2
```

```
##   .. .. ..$ Y  : chr [1:2] "democrat" "republican"
##   .. .. ..$ V13: chr [1:2] "n" "y"
##   ..$ V14: 'table' num [1:2, 1:2] 0.6682 0.0238 0.3318 0.9762
##   .. ..- attr(*, "dimnames")=List of 2
##   .. .. ..$ Y  : chr [1:2] "democrat" "republican"
##   .. .. ..$ V14: chr [1:2] "n" "y"
##   ..$ V15: 'table' num [1:2, 1:2] 0.3839 0.9024 0.6161 0.0976
##   .. ..- attr(*, "dimnames")=List of 2
##   .. .. ..$ Y  : chr [1:2] "democrat" "republican"
##   .. .. ..$ V15: chr [1:2] "n" "y"
##   ..$ V16: 'table' num [1:2, 1:2] 0.0758 0.3772 0.9242 0.6228
##   .. ..- attr(*, "dimnames")=List of 2
##   .. .. ..$ Y  : chr [1:2] "democrat" "republican"
##   .. .. ..$ V16: chr [1:2] "n" "y"
##  $ levels   : chr [1:2] "democrat" "republican"
##  $ isnumeric: Named logi [1:16] FALSE FALSE FALSE FALSE FALSE FALSE ...
##   ..- attr(*, "names")= chr [1:16] "V1" "V2" "V3" "V4" ...
##  $ call     : language naiveBayes.default(x = X, y = Y, laplace = laplace)
##  - attr(*, "class")= chr "naiveBayes"
```

#lets test the model

```
nb_test_redict <- predict(nb_model,testHouseVotes84[,-1])
#fraction of correct predictions
mean(nb_test_redict==testHouseVotes84$Class)
```

```
## [1] 0.9450549
```

```
#confusion matrix
table(pred=nb_test_redict ,true= testHouseVotes84$Class)
```

```
##             true
## pred         democrat republican
##   democrat         54          3
##   republican        2         32
```

```
#fraction of correct predictions
mean(nb_test_redict==testHouseVotes84$Class)
```

```
## [1] 0.9450549
```

#function to create run and record model results

```
nb_multiple_runs <- function(train_fraction,n){
  fraction_correct <- rep(NA,n)
  for(i in 1:n){
    HouseVotes84[,'train']<-ifelse(runif(nrow(HouseVotes84))<train_fraction,1,0)
    trainColNum <- grep('train',names(HouseVotes84))
    trainHouseVotes84 <- HouseVotes84[HouseVotes84$train ==1,-trainColNum]
    testHouseVotes84 <- HouseVotes84[HouseVotes84$train ==0,-trainColNum]
    nb_model <-naiveBayes(Class~.,data = trainHouseVotes84)
```

```
   fraction_correct[i] <- mean(nb_test_redict == testHouseVotes84$Class)
  }
  return(fraction_correct)
}
```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.