Group 3 Business Analytics Project

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(gam)
```

```
## Loading required package: splines
```

```
## Loading required package: foreach
```

```
## Loaded gam 1.16
```

```
library(readr)
Churn_Train <- read_csv("Churn_Train.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   state = col_character(),
##   account_length = col_integer(),
##   area_code = col_character(),
##   international_plan = col_character(),
##   voice_mail_plan = col_character(),
##   number_vmail_messages = col_integer(),
##   total_day_calls = col_integer(),
##   total_eve_calls = col_integer(),
##   total_night_calls = col_integer(),
##   total_intl_calls = col_integer(),
##   number_customer_service_calls = col_integer(),
##   churn = col_character()
## )
```

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

```
summary(Churn_Train)
```

```
##     state           account_length     area_code
##  Length:3333        Min.   :-209.00   Length:3333
##  Class :character   1st Qu.:  72.00   Class :character
##  Mode  :character   Median : 100.00   Mode  :character
##                     Mean   :  97.32
##                     3rd Qu.: 127.00
##                     Max.   : 243.00
##                     NA's   :501
##  international_plan voice_mail_plan    number_vmail_messages
##  Length:3333        Length:3333        Min.   :-10.000
##  Class :character   Class :character   1st Qu.:  0.000
##  Mode  :character   Mode  :character   Median :  0.000
##                                        Mean   :  7.333
##                                        3rd Qu.: 16.000
##                                        Max.   : 51.000
##                                        NA's   :200
##  total_day_minutes total_day_calls total_day_charge total_eve_minutes
##  Min.   :   0.0    Min.   :  0.0   Min.   : 0.00    Min.   :   0.0
##  1st Qu.: 149.3    1st Qu.: 87.0   1st Qu.:24.45    1st Qu.: 170.5
##  Median : 190.5    Median :101.0   Median :30.65    Median : 209.9
##  Mean   : 418.9    Mean   :100.3   Mean   :30.63    Mean   : 324.3
##  3rd Qu.: 237.8    3rd Qu.:114.0   3rd Qu.:36.84    3rd Qu.: 257.6
##  Max.   :2185.1    Max.   :165.0   Max.   :59.64    Max.   :1244.2
##  NA's   :200       NA's   :200     NA's   :200      NA's   :301
##  total_eve_calls total_eve_charge total_night_minutes total_night_calls
##  Min.   :  0.0   Min.   : 0.00    Min.   : 23.2       Min.   : 33.0
##  1st Qu.: 87.0   1st Qu.:14.14    1st Qu.:167.3       1st Qu.: 87.0
##  Median :100.0   Median :17.09    Median :201.4       Median :100.0
##  Mean   :100.1   Mean   :17.08    Mean   :201.2       Mean   :100.1
##  3rd Qu.:114.0   3rd Qu.:20.00    3rd Qu.:235.3       3rd Qu.:113.0
##  Max.   :170.0   Max.   :30.91    Max.   :395.0       Max.   :175.0
##  NA's   :200     NA's   :200      NA's   :200
##  total_night_charge total_intl_minutes total_intl_calls total_intl_charge
##  Min.   : 1.040     Min.   : 0.00      Min.   : 0.00    Min.   :0.000
##  1st Qu.: 7.530     1st Qu.: 8.50      1st Qu.: 3.00    1st Qu.:2.300
##  Median : 9.060     Median :10.30      Median : 4.00    Median :2.780
##  Mean   : 9.054     Mean   :10.23      Mean   : 4.47    Mean   :2.762
##  3rd Qu.:10.590     3rd Qu.:12.10      3rd Qu.: 6.00    3rd Qu.:3.270
##  Max.   :17.770     Max.   :20.00      Max.   :20.00    Max.   :5.400
##  NA's   :200        NA's   :200        NA's   :301      NA's   :200
##  number_customer_service_calls    churn
##  Min.   :0.000                 Length:3333
##  1st Qu.:1.000                 Class :character
##  Median :1.000                 Mode  :character
##  Mean   :1.561
##  3rd Qu.:2.000
##  Max.   :9.000
##  NA's   :200
```

```
#Removed 200 lines that contain a majority of NA values
Churn_Train1 <- as.data.frame(Churn_Train[!is.na(Churn_Train$number_vmail_messages),])
#Replace the remaining NA values with mean of the columns
Churn <-  Churn_Train1 %>%
  na.gam.replace()
#Make the churn column a factor data type
Churn[, 'churn'] <- as.factor(Churn[, 'churn'])
#Turn the negative values in the account_length and number_vmail_messages into positive values
Churn[, 'account_length'] <- abs(Churn[, 'account_length'])
Churn[, 'number_vmail_messages'] <- abs(Churn[, 'number_vmail_messages'])
summary(Churn)
```

```
##     state           account_length    area_code         international_plan
##  Length:3133        Min.   :  1.00   Length:3133        Length:3133
##  Class :character   1st Qu.: 76.00   Class :character   Class :character
##  Mode  :character   Median : 97.32   Mode  :character   Mode  :character
##                     Mean   :100.50
##                     3rd Qu.:124.00
##                     Max.   :243.00
##  voice_mail_plan    number_vmail_messages total_day_minutes
##  Length:3133        Min.   : 0.000        Min.   :   0.0
##  Class :character   1st Qu.: 0.000        1st Qu.: 149.3
##  Mode  :character   Median : 0.000        Median : 190.5
##                     Mean   : 8.056        Mean   : 418.9
##                     3rd Qu.:16.000        3rd Qu.: 237.8
##                     Max.   :51.000        Max.   :2185.1
##  total_day_calls total_day_charge total_eve_minutes total_eve_calls
##  Min.   :  0.0   Min.   : 0.00    Min.   :   0.0    Min.   :  0.0
##  1st Qu.: 87.0   1st Qu.:24.45    1st Qu.: 172.1    1st Qu.: 87.0
##  Median :101.0   Median :30.65    Median : 211.7    Median :100.0
##  Mean   :100.3   Mean   :30.63    Mean   : 324.3    Mean   :100.1
##  3rd Qu.:114.0   3rd Qu.:36.84    3rd Qu.: 264.7    3rd Qu.:114.0
##  Max.   :165.0   Max.   :59.64    Max.   :1244.2    Max.   :170.0
##  total_eve_charge total_night_minutes total_night_calls total_night_charge
##  Min.   : 0.00    Min.   : 23.2       Min.   : 33.0     Min.   : 1.040
##  1st Qu.:14.14    1st Qu.:167.3       1st Qu.: 87.0     1st Qu.: 7.530
##  Median :17.09    Median :201.4       Median :100.0     Median : 9.060
##  Mean   :17.08    Mean   :201.2       Mean   :100.1     Mean   : 9.054
##  3rd Qu.:20.00    3rd Qu.:235.3       3rd Qu.:114.0     3rd Qu.:10.590
##  Max.   :30.91    Max.   :395.0       Max.   :175.0     Max.   :17.770
##  total_intl_minutes total_intl_calls total_intl_charge
##  Min.   : 0.00      Min.   : 0.00    Min.   :0.000
##  1st Qu.: 8.50      1st Qu.: 3.00    1st Qu.:2.300
##  Median :10.30      Median : 4.00    Median :2.780
##  Mean   :10.23      Mean   : 4.47    Mean   :2.762
##  3rd Qu.:12.10      3rd Qu.: 6.00    3rd Qu.:3.270
##  Max.   :20.00      Max.   :20.00    Max.   :5.400
##  number_customer_service_calls churn
##  Min.   :0.000                 no :2676
##  1st Qu.:1.000                 yes: 457
##  Median :1.000
##  Mean   :1.561
##  3rd Qu.:2.000
##  Max.   :9.000
```

```
library(pROC)
```

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

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

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

```
model <- glm(formula = churn ~ ., family="binomial", data=Churn)
summary(model)
```

```
##
## Call:
## glm(formula = churn ~ ., family = "binomial", data = Churn)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9039  -0.5081  -0.3106  -0.1648   3.0895
##
## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                 -9.4471995  0.9989458  -9.457  < 2e-16 ***
## stateAL                      0.3911365  0.7655604   0.511  0.60941
## stateAR                      0.8162181  0.7639613   1.068  0.28534
## stateAZ                      0.1127551  0.8475101   0.133  0.89416
## stateCA                      1.9723430  0.7916823   2.491  0.01273 *
## stateCO                      0.4787175  0.7752567   0.617  0.53691
## stateCT                      1.0816963  0.7286717   1.484  0.13768
## stateDC                      0.8175349  0.8117119   1.007  0.31385
## stateDE                      0.6975276  0.7525338   0.927  0.35398
## stateFL                      0.4387212  0.7800572   0.562  0.57383
## stateGA                      0.7276099  0.7843528   0.928  0.35359
## stateHI                     -0.0919787  0.8984942  -0.102  0.91846
## stateIA                      0.1948852  0.9088458   0.214  0.83021
## stateID                      0.7588835  0.7610772   0.997  0.31871
## stateIL                     -0.1405240  0.8379987  -0.168  0.86683
## stateIN                      0.4528995  0.7566573   0.599  0.54947
## stateKS                      1.0579782  0.7331881   1.443  0.14903
## stateKY                      0.6822305  0.7850798   0.869  0.38485
## stateLA                      0.6582476  0.8411341   0.783  0.43388
## stateMA                      1.0956810  0.7508275   1.459  0.14448
## stateMD                      1.1615773  0.7194015   1.615  0.10639
## stateME                      1.2862431  0.7382799   1.742  0.08147 .
## stateMI                      1.4876047  0.7214728   2.062  0.03922 *
## stateMN                      1.1997394  0.7183840   1.670  0.09491 .
## stateMO                      0.5675962  0.8009720   0.709  0.47855
## stateMS                      1.2686664  0.7369622   1.721  0.08516 .
## stateMT                      1.7631968  0.7233573   2.438  0.01479 *
## stateNC                      0.6420020  0.7582680   0.847  0.39718
## stateND                      0.1925709  0.7984972   0.241  0.80943
## stateNE                      0.2152010  0.8404369   0.256  0.79791
## stateNH                      1.1045932  0.7782999   1.419  0.15583
## stateNJ                      1.5506865  0.7161856   2.165  0.03037 *
## stateNM                      0.5575979  0.7954387   0.701  0.48331
## stateNV                      1.2765115  0.7291431   1.751  0.08000 .
## stateNY                      1.0911297  0.7292181   1.496  0.13458
## stateOH                      0.5532740  0.7582834   0.730  0.46561
## stateOK                      0.7228998  0.7698285   0.939  0.34771
## stateOR                      0.6718900  0.7470107   0.899  0.36842
## statePA                      1.2028390  0.7833174   1.536  0.12464
## stateRI                     -0.2755070  0.8458325  -0.326  0.74463
## stateSC                      1.7288901  0.7505058   2.304  0.02124 *
## stateSD                      0.8652566  0.7632888   1.134  0.25697
## stateTN                      0.3857261  0.8279508   0.466  0.64130
```

```
## stateTX                         1.6753270  0.7114979   2.355  0.01854 *
## stateUT                         1.2250202  0.7456569   1.643  0.10041
## stateVA                        -0.4087997  0.8579790  -0.476  0.63374
## stateVT                         0.0666644  0.8013926   0.083  0.93370
## stateWA                         1.4927534  0.7319348   2.039  0.04140 *
## stateWI                         0.3191137  0.7860244   0.406  0.68475
## stateWV                         0.6086066  0.7368888   0.826  0.40885
## stateWY                         0.2418391  0.7636715   0.317  0.75149
## account_length                  0.0001484  0.0015649   0.095  0.92443
## area_codearea_code_415         -0.0304658  0.1455332  -0.209  0.83418
## area_codearea_code_510         -0.0977078  0.1689150  -0.578  0.56296
## international_planyes           2.1942438  0.1591012  13.791  < 2e-16 ***
## voice_mail_planyes             -1.2136356  0.4387638  -2.766  0.00567 **
## number_vmail_messages           0.0068628  0.0145504   0.472  0.63717
## total_day_minutes              -0.0014743  0.0010804  -1.365  0.17235
## total_day_calls                 0.0032519  0.0029666   1.096  0.27301
## total_day_charge                0.0847134  0.0088452   9.577  < 2e-16 ***
## total_eve_minutes               0.0026047  0.0021395   1.217  0.22345
## total_eve_calls                 0.0003895  0.0029721   0.131  0.89574
## total_eve_charge                0.0663964  0.0277251   2.395  0.01663 *
## total_night_minutes             0.0801288  0.9278029   0.086  0.93118
## total_night_calls              -0.0007419  0.0030162  -0.246  0.80570
## total_night_charge             -1.6906729 20.6168583  -0.082  0.93464
## total_intl_minutes             -2.6724013  5.6932294  -0.469  0.63878
## total_intl_calls               -0.0829101  0.0267874  -3.095  0.00197 **
## total_intl_charge              10.1953319 21.0846419   0.484  0.62871
## number_customer_service_calls  0.5410994  0.0420861  12.857  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2603.3  on 3132  degrees of freedom
## Residual deviance: 1956.3  on 3063  degrees of freedom
## AIC: 2096.3
##
## Number of Fisher Scoring iterations: 6
```

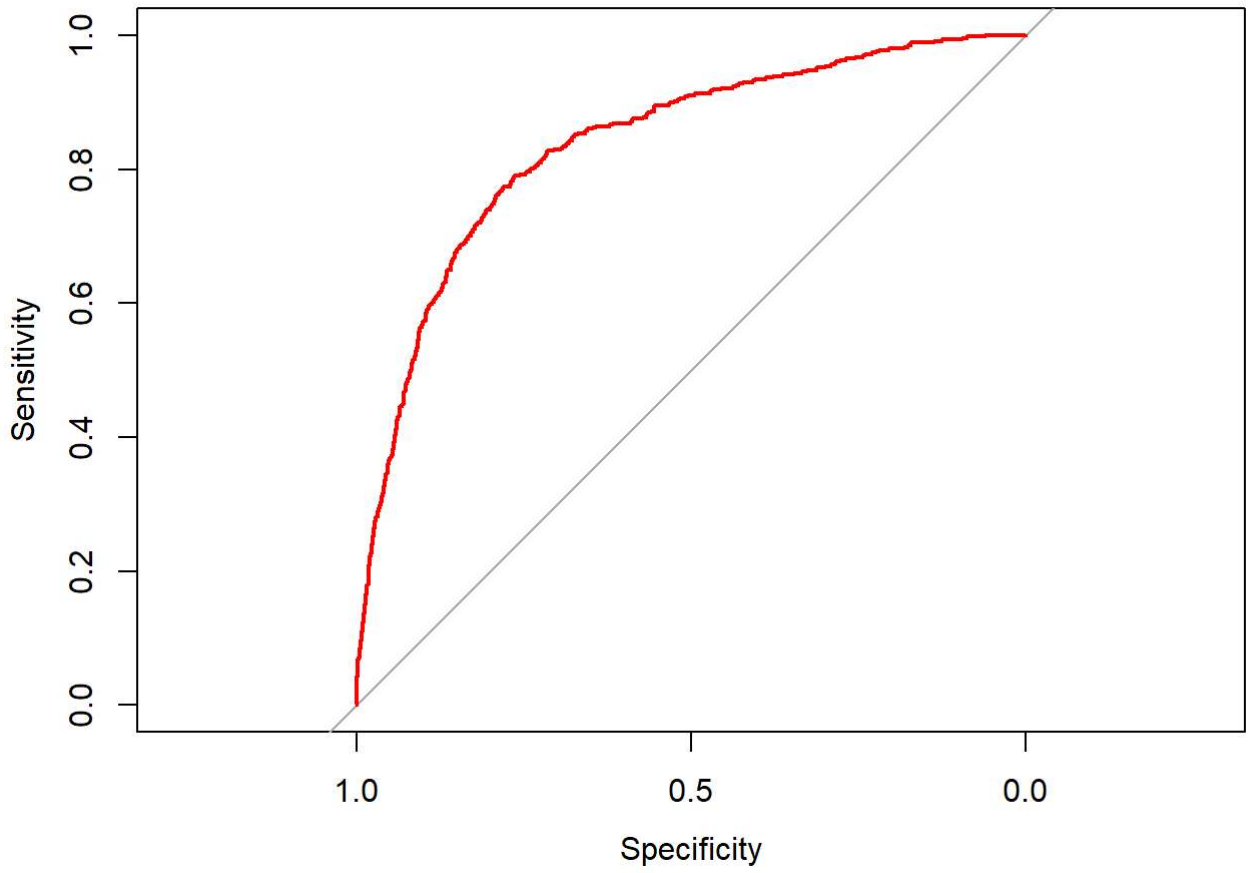
```
predict_model <- predict(model, newdata=Churn, type = 'response')

probability <- as.numeric(predict_model>.55)

levels(predict_model) <- list(no='FALSE',yes='TRUE')
table(Predicted=probability, True=Churn$churn)
```

```
##          True
## Predicted   no  yes
##         0 2626  361
##         1   50   96
```

```
roc(Churn$churn,predict_model)
```

```
##
## Call:
## roc.default(response = Churn$churn, predictor = predict_model)
##
## Data: predict_model in 2676 controls (Churn$churn no) < 457 cases (Churn$churn yes).
## Area under the curve: 0.8377
```

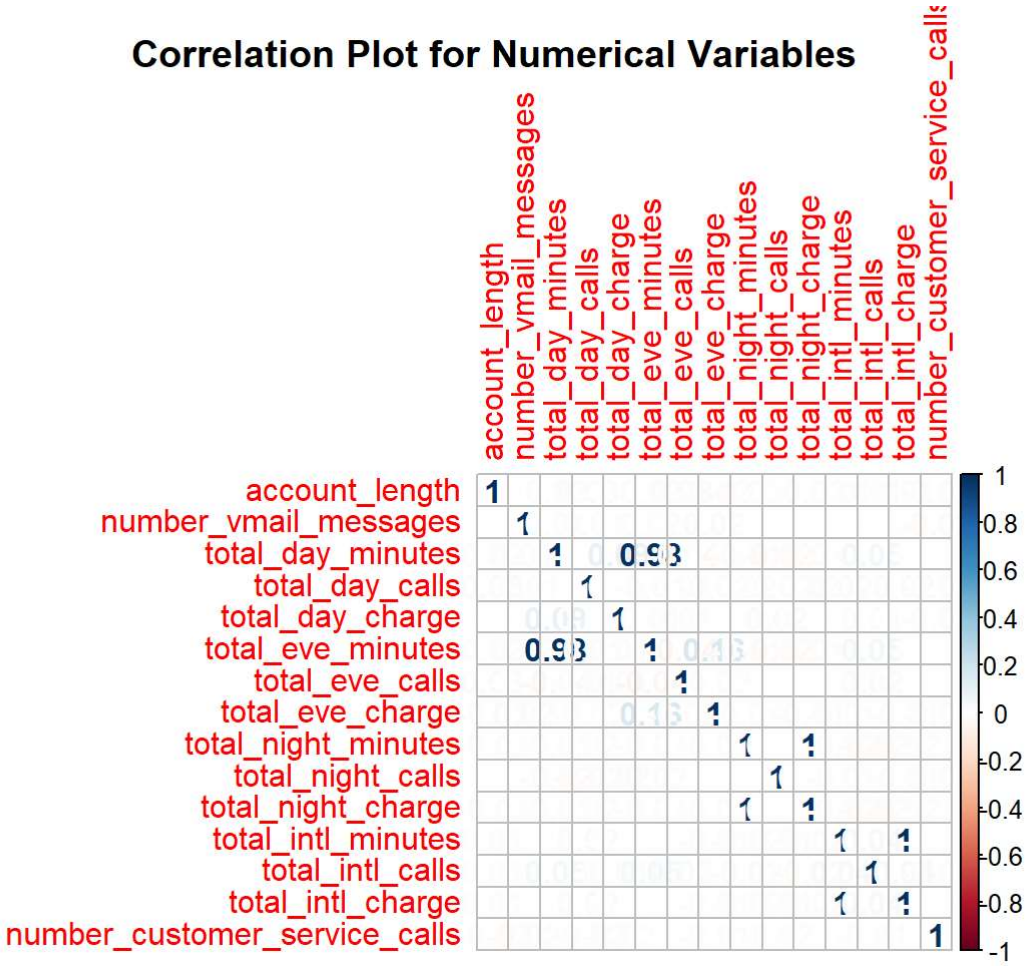
```
plot(roc(Churn$churn,predict_model),col='red',lwd=2)
```




```
library(corrplot)
```

```
## corrplot 0.84 loaded
```

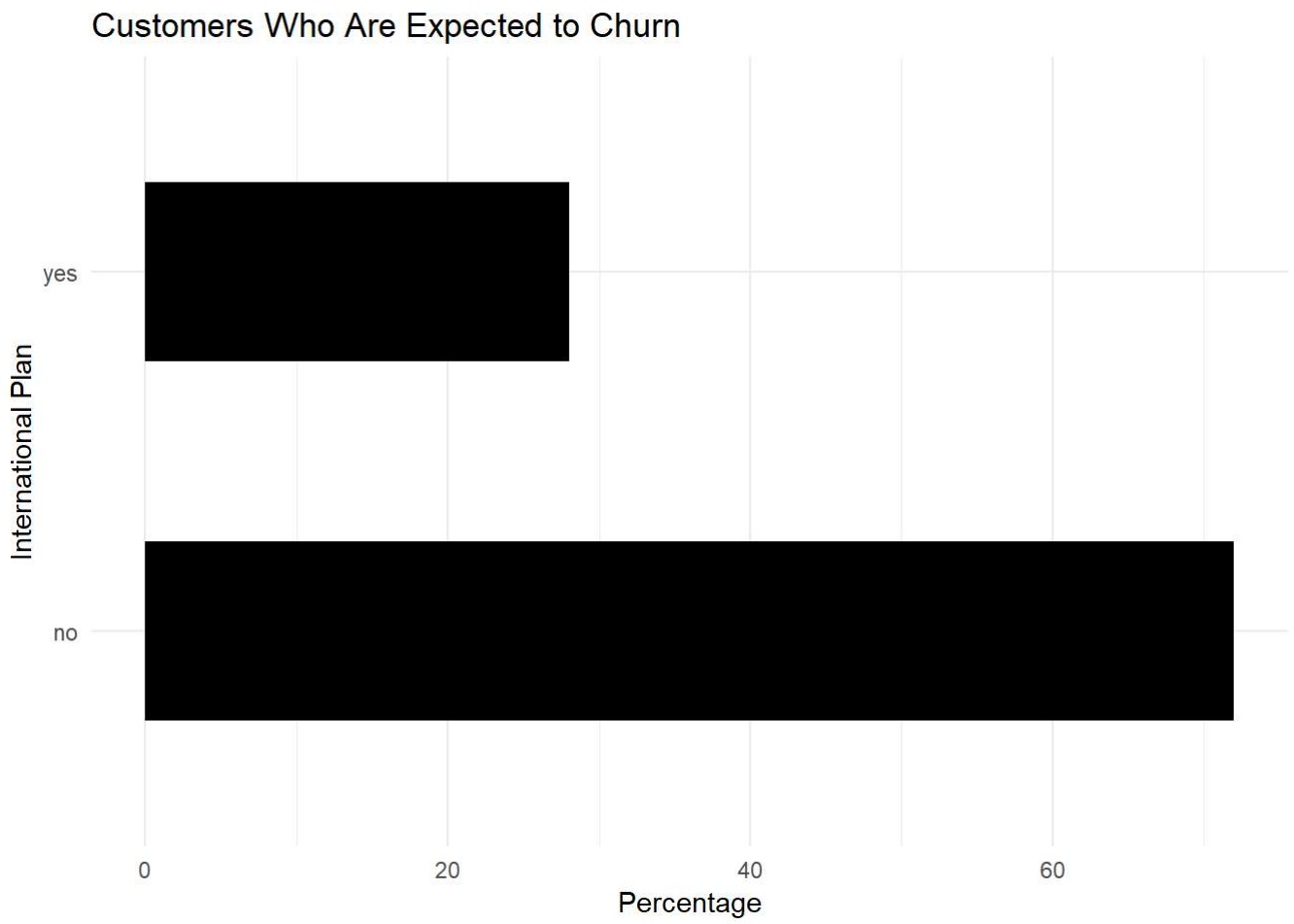
```
numeric.var <- sapply(Churn, is.numeric)
corr.matrix <- cor(Churn[,numeric.var])
corrplot(corr.matrix, main="\n\nCorrelation Plot for Numerical Variables", method="number")
```



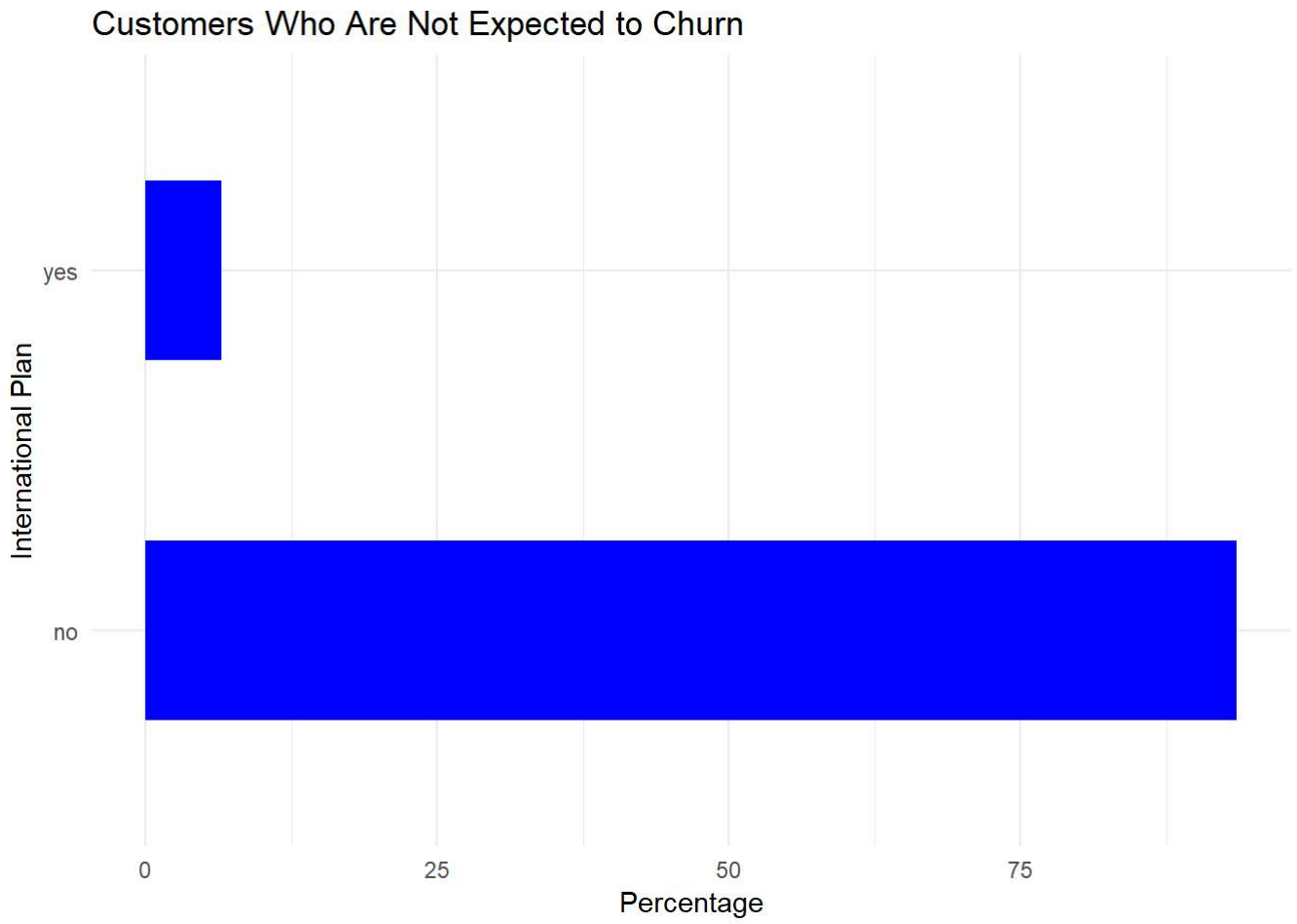

```
library(dplyr)
Churn1 <- Churn %>%
  filter(churn=="yes") %>%
  group_by(state) %>%
  summarise(Total_Churn=n())

Churn2 <- Churn %>%
  group_by(state) %>%
  summarise(Total_Observations=n())

Churn3 <- cbind(Churn1,Churn2[,"Total_Observations"])

Churn_By_State <- Churn3 %>%
  group_by(state) %>%
  summarise(Churn_Ratio=Total_Churn/Total_Observations) %>%
  arrange(desc(Churn_Ratio))
Churn_By_State
```

```
## # A tibble: 51 x 2
##    state Churn_Ratio
##    <chr>       <dbl>
##  1 CA          0.290
##  2 TX          0.261
##  3 NJ          0.25
##  4 MD          0.246
##  5 MI          0.231
##  6 SC          0.226
##  7 NV          0.226
##  8 WA          0.213
##  9 ME          0.207
## 10 MS          0.206
## # ... with 41 more rows
```

```
library(dplyr)
library(ggplot2)
Churn_Yes <- Churn %>%
  filter(churn=="yes")

Churn_No <- Churn %>%
  filter(churn=="no")

intern_churn_yes <- ggplot(Churn_Yes, aes(x=international_plan)) + ggtitle("Customers Who Are Ex
pected to Churn") + xlab("International Plan") +geom_bar(aes(y = 100*(..count..)/sum(..coun
t..)), width = 0.5, fill="black") + ylab("Percentage") + coord_flip() + theme_minimal()
intern_churn_yes
```




```
intern_churn_no <- ggplot(Churn_No, aes(x=international_plan)) + ggtitle("Customers Who Are Not
 Expected to Churn") + xlab("International Plan") + geom_bar(aes(y = 100*(..count..)/sum(..coun
t..)), width = 0.5, fill="blue") + ylab("Percentage") + coord_flip() + theme_minimal()
intern_churn_no
```


Customers Who Are Not Expected to Churn
International Plan
yes
no
0
25
50
75
Percentage