Logistic Regression (Classification)

# Logistic Regression (Classification)

## BAN502

### Danielle Harris

# Libraries

library(tidyverse)

## ── Attaching packages ────────────────────────────────────────── tidyverse 1.2.1 ──

## ✔ ggplot2 3.1.0 ✔ purrr 0.2.5  
## ✔ tibble 1.4.2 ✔ dplyr 0.7.7  
## ✔ tidyr 0.8.2 ✔ stringr 1.3.1  
## ✔ readr 1.1.1 ✔ forcats 0.3.0

## ── Conflicts ───────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

library(MASS)

##   
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':  
##   
## select

library(caret)

## Loading required package: lattice

##   
## Attaching package: 'caret'

## The following object is masked from 'package:purrr':  
##   
## lift

library(ROCR)

## Loading required package: gplots

## Warning: package 'gplots' was built under R version 3.5.2

##   
## Attaching package: 'gplots'

## The following object is masked from 'package:stats':  
##   
## lowess

library(e1071)

## Warning: package 'e1071' was built under R version 3.5.2

parole= read\_csv("parole.csv")

## Parsed with column specification:  
## cols(  
## male = col\_integer(),  
## race = col\_integer(),  
## age = col\_double(),  
## state = col\_integer(),  
## time.served = col\_double(),  
## max.sentence = col\_integer(),  
## multiple.offenses = col\_integer(),  
## crime = col\_integer(),  
## violator = col\_integer()  
## )

parole = parole %>% mutate(male= as.factor(male)) %>%   
 mutate (male= fct\_recode(male, "female"="0", "male"= "1")) %>%   
mutate(race= as.factor(race)) %>% mutate(race=fct\_recode(race, "white" = "1", "other"= "2")) %>%   
mutate(age= as.factor(age)) %>%  
mutate(state= as.factor(state)) %>% mutate(state= fct\_recode(state, "Kentucky"="2", "Louisiana"= "3", "Virginia"="4", "other"= "1")) %>%  
mutate(crime= as.factor(crime)) %>%  
mutate(crime= fct\_recode(crime, "larceny"= "2", "drug-related"="3", "driving-related"="4", "other"= "1")) %>%  
mutate(multiple.offenses= as.factor(multiple.offenses)) %>% mutate(multiple.offenses=fct\_recode(multiple.offenses, "yes"="1", "no"= "0")) %>%  
mutate(violator=as.factor(violator)) %>% mutate (violator= fct\_recode(violator, "yes"= "1", "no"="0"))

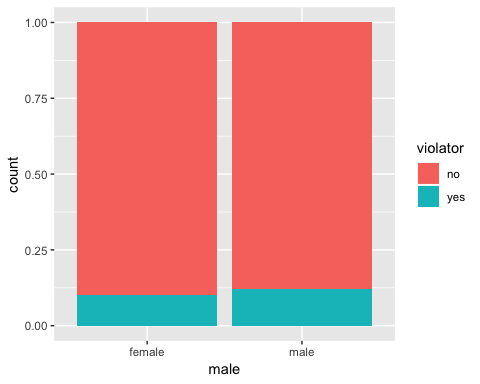
# Task 1

set.seed(12345)  
train.rows= createDataPartition(y=parole$violator, p= 0.7, list=FALSE)  
train=parole[train.rows,]  
test=parole[-train.rows,]

# Task 2

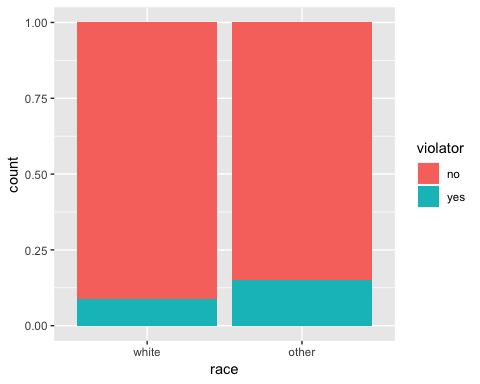
## The bar graphs will display which variables most significantly predict if a parolee will violate their parole.

ggplot(train, aes(x=male, fill=violator)) + geom\_bar(position="fill")



# Although not very significant, a male parolee is slightly more likely to violate his parole.

ggplot(train, aes(x=race, fill=violator)) + geom\_bar(position="fill")



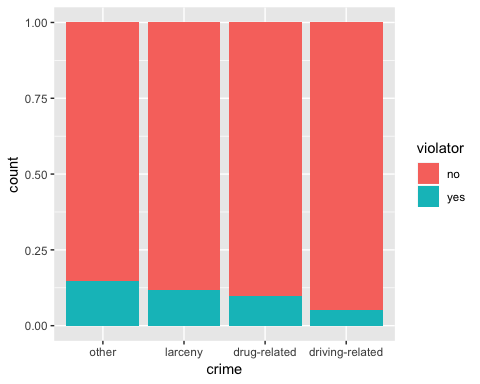
# Although not very significant, a parolee with a race other than white is more likely to violate their parole.

ggplot(train, aes(x=state, fill=violator)) + geom\_bar(position="fill")



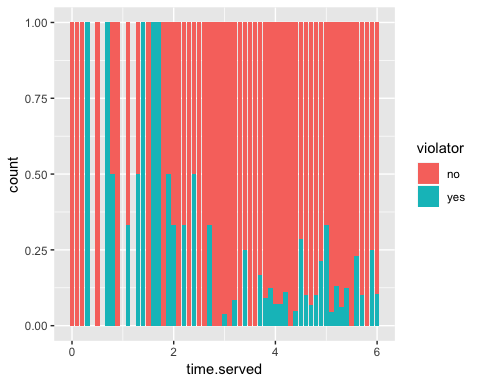
# A parolee from Lousiana is more likely to violate their parole than any parolee from another state.

ggplot(train, aes(x=crime, fill=violator)) + geom\_bar(position="fill")



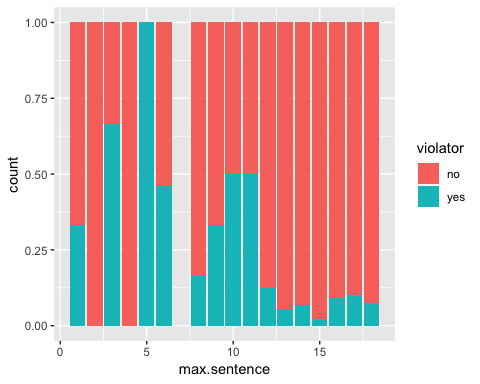
# There seems to be little prediction of a parolee violating their parole based on the crime committed.

ggplot(train, aes(x=time.served, fill=violator)) + geom\_bar(position="fill")



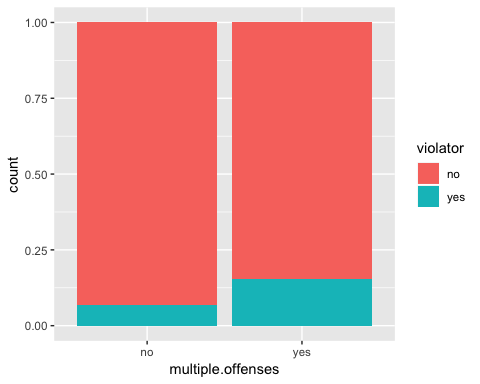
# A parolee is more likely to violate their parole between years 0 and 2.

ggplot(train, aes(x=max.sentence, fill=violator)) + geom\_bar(position="fill")



# A parolee with a max sentence of 5 years is most likely to violate their parole.

ggplot(train, aes(x=multiple.offenses, fill=violator)) + geom\_bar(position="fill")



# A parolee with multiple offenses is more likely to violate their parole than a parolee with only one offense.

# Task 3

mod1 = glm(violator ~ time.served , train, family = "binomial")  
summary(mod1)

##   
## Call:  
## glm(formula = violator ~ time.served, family = "binomial", data = train)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.7419 -0.5343 -0.4721 -0.4253 2.2579   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.1495 0.4411 -2.606 0.00915 \*\*  
## time.served -0.2197 0.1079 -2.035 0.04183 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 340.04 on 472 degrees of freedom  
## Residual deviance: 336.02 on 471 degrees of freedom  
## AIC: 340.02  
##   
## Number of Fisher Scoring iterations: 4

# The variable that seems to predict whether or not a parolee will violate their parole is time served. The p-value is less than 0.05, proving to be a strong and true model.

allmod = glm(violator ~ male + race + age + state + time.served + max.sentence + multiple.offenses + crime, parole, family = "binomial")

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

summary(allmod)

##   
## Call:  
## glm(formula = violator ~ male + race + age + state + time.served +   
## max.sentence + multiple.offenses + crime, family = "binomial",   
## data = parole)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.64760 -0.00003 -0.00002 0.00000 2.84152   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -3.105e+01 2.764e+04 -0.001 0.99910   
## malemale 4.028e-01 1.039e+00 0.388 0.69836   
## raceother 8.184e-01 8.821e-01 0.928 0.35355   
## age18.5 8.158e+00 5.556e+04 0.000 0.99988   
## age18.7 2.929e+01 2.764e+04 0.001 0.99915   
## age18.8 5.820e+00 4.141e+04 0.000 0.99989   
## age19 8.766e+00 5.556e+04 0.000 0.99987   
## age19.1 3.932e+00 3.598e+04 0.000 0.99991   
## age19.2 3.613e+00 3.354e+04 0.000 0.99991   
## age19.3 1.042e+01 5.556e+04 0.000 0.99985   
## age19.4 1.210e+00 3.305e+04 0.000 0.99997   
## age19.5 5.359e+00 4.388e+04 0.000 0.99990   
## age19.6 7.271e+00 3.600e+04 0.000 0.99984   
## age19.7 6.621e+00 4.386e+04 0.000 0.99988   
## age19.9 2.465e+01 2.764e+04 0.001 0.99929   
## age20 1.162e+00 3.654e+04 0.000 0.99997   
## age20.2 2.076e+01 2.764e+04 0.001 0.99940   
## age20.3 1.616e+01 2.764e+04 0.001 0.99953   
## age20.4 8.462e+00 5.556e+04 0.000 0.99988   
## age20.5 2.472e+01 2.764e+04 0.001 0.99929   
## age20.6 2.850e+01 2.764e+04 0.001 0.99918   
## age20.7 -1.802e+00 3.173e+04 0.000 0.99995   
## age20.8 7.938e+00 4.315e+04 0.000 0.99985   
## age20.9 4.127e+00 5.556e+04 0.000 0.99994   
## age21 2.179e+00 4.362e+04 0.000 0.99996   
## age21.1 4.287e+00 4.049e+04 0.000 0.99992   
## age21.2 -6.165e+00 3.264e+04 0.000 0.99985   
## age21.3 8.457e+00 5.556e+04 0.000 0.99988   
## age21.4 3.325e+00 3.432e+04 0.000 0.99992   
## age21.5 6.658e+00 4.341e+04 0.000 0.99988   
## age21.6 2.388e+01 2.764e+04 0.001 0.99931   
## age21.7 2.179e+01 2.765e+04 0.001 0.99937   
## age21.8 5.466e+00 4.388e+04 0.000 0.99990   
## age21.9 -6.428e+00 3.082e+04 0.000 0.99983   
## age22 5.212e+00 3.572e+04 0.000 0.99988   
## age22.1 8.802e+00 4.349e+04 0.000 0.99984   
## age22.2 8.071e+00 5.556e+04 0.000 0.99988   
## age22.3 4.374e+00 3.922e+04 0.000 0.99991   
## age22.4 2.588e+01 2.764e+04 0.001 0.99925   
## age22.5 -4.155e-01 3.946e+04 0.000 0.99999   
## age22.6 5.202e+00 3.810e+04 0.000 0.99989   
## age22.8 2.578e+01 2.764e+04 0.001 0.99926   
## age22.9 7.261e+00 5.556e+04 0.000 0.99990   
## age23 -8.493e+00 3.042e+04 0.000 0.99978   
## age23.1 7.410e+00 4.127e+04 0.000 0.99986   
## age23.2 1.586e+00 3.135e+04 0.000 0.99996   
## age23.3 2.040e+01 2.764e+04 0.001 0.99941   
## age23.4 5.148e+00 3.744e+04 0.000 0.99989   
## age23.6 2.051e+01 2.764e+04 0.001 0.99941   
## age23.7 2.325e+01 2.764e+04 0.001 0.99933   
## age23.8 6.986e+00 4.384e+04 0.000 0.99987   
## age24 -1.011e+00 5.556e+04 0.000 0.99999   
## age24.2 2.102e+01 2.764e+04 0.001 0.99939   
## age24.3 -6.479e+00 3.830e+04 0.000 0.99987   
## age24.4 7.647e+00 4.383e+04 0.000 0.99986   
## age24.5 5.684e+00 3.720e+04 0.000 0.99988   
## age24.6 6.119e+00 5.556e+04 0.000 0.99991   
## age24.7 1.088e+01 5.556e+04 0.000 0.99984   
## age24.8 9.366e+00 4.349e+04 0.000 0.99983   
## age24.9 -3.569e+00 3.449e+04 0.000 0.99992   
## age25 6.037e+00 3.852e+04 0.000 0.99987   
## age25.1 1.760e+00 3.469e+04 0.000 0.99996   
## age25.2 8.674e+00 5.556e+04 0.000 0.99988   
## age25.3 2.437e+01 2.764e+04 0.001 0.99930   
## age25.4 7.142e+00 5.556e+04 0.000 0.99990   
## age25.5 -6.112e+00 5.556e+04 0.000 0.99991   
## age25.6 2.710e+01 2.764e+04 0.001 0.99922   
## age25.7 -3.906e+00 3.253e+04 0.000 0.99990   
## age25.8 2.104e+01 2.764e+04 0.001 0.99939   
## age25.9 5.172e-01 3.475e+04 0.000 0.99999   
## age26 4.679e+01 5.556e+04 0.001 0.99933   
## age26.3 5.055e+00 4.388e+04 0.000 0.99991   
## age26.4 -4.979e-01 4.360e+04 0.000 0.99999   
## age26.5 8.215e-01 3.908e+04 0.000 0.99998   
## age26.6 1.767e+00 5.556e+04 0.000 0.99997   
## age26.8 2.859e+01 2.764e+04 0.001 0.99917   
## age26.9 3.052e+00 3.417e+04 0.000 0.99993   
## age27 6.031e+00 4.197e+04 0.000 0.99989   
## age27.1 8.048e+00 4.251e+04 0.000 0.99985   
## age27.2 -3.893e-02 3.797e+04 0.000 1.00000   
## age27.3 -8.729e+00 5.556e+04 0.000 0.99987   
## age27.4 5.111e+01 5.556e+04 0.001 0.99927   
## age27.5 2.834e+01 2.764e+04 0.001 0.99918   
## age27.6 6.905e+00 4.209e+04 0.000 0.99987   
## age27.7 3.908e+00 3.949e+04 0.000 0.99992   
## age27.8 7.679e+00 3.784e+04 0.000 0.99984   
## age27.9 8.885e+00 5.556e+04 0.000 0.99987   
## age28 -6.731e+00 3.753e+04 0.000 0.99986   
## age28.1 1.679e+01 2.764e+04 0.001 0.99952   
## age28.2 6.816e+00 3.598e+04 0.000 0.99985   
## age28.3 4.100e+00 4.169e+04 0.000 0.99992   
## age28.4 1.986e+01 2.764e+04 0.001 0.99943   
## age28.5 2.916e+01 2.764e+04 0.001 0.99916   
## age28.7 8.115e+00 3.494e+04 0.000 0.99981   
## age28.8 2.199e+01 2.764e+04 0.001 0.99937   
## age28.9 2.475e+01 2.764e+04 0.001 0.99929   
## age29 7.101e+00 5.556e+04 0.000 0.99990   
## age29.1 4.350e+00 4.263e+04 0.000 0.99992   
## age29.2 6.461e+00 4.361e+04 0.000 0.99988   
## age29.5 2.680e+01 2.764e+04 0.001 0.99923   
## age29.6 8.634e+00 4.230e+04 0.000 0.99984   
## age29.7 8.407e+00 3.829e+04 0.000 0.99982   
## age29.9 2.881e+01 2.764e+04 0.001 0.99917   
## age30 1.829e+00 3.585e+04 0.000 0.99996   
## age30.1 5.417e+00 4.114e+04 0.000 0.99989   
## age30.2 2.948e+00 5.556e+04 0.000 0.99996   
## age30.3 8.782e+00 5.556e+04 0.000 0.99987   
## age30.4 1.118e+01 5.556e+04 0.000 0.99984   
## age30.7 2.276e+01 2.764e+04 0.001 0.99934   
## age30.8 2.324e+01 2.764e+04 0.001 0.99933   
## age31 2.499e+01 2.764e+04 0.001 0.99928   
## age31.1 -2.017e+00 3.905e+04 0.000 0.99996   
## age31.2 -5.712e-01 3.532e+04 0.000 0.99999   
## age31.3 6.543e+00 5.556e+04 0.000 0.99991   
## age31.4 1.841e+01 2.764e+04 0.001 0.99947   
## age31.5 2.317e+01 2.764e+04 0.001 0.99933   
## age31.6 8.252e+00 5.556e+04 0.000 0.99988   
## age31.7 6.110e+00 5.556e+04 0.000 0.99991   
## age31.8 8.425e+00 3.807e+04 0.000 0.99982   
## age32 6.201e+00 4.239e+04 0.000 0.99988   
## age32.1 9.188e+00 3.717e+04 0.000 0.99980   
## age32.2 2.564e+01 2.764e+04 0.001 0.99926   
## age32.3 4.200e+00 5.556e+04 0.000 0.99994   
## age32.4 7.386e+00 3.508e+04 0.000 0.99983   
## age32.5 5.158e+00 5.556e+04 0.000 0.99993   
## age32.6 5.926e+00 5.556e+04 0.000 0.99991   
## age32.7 2.648e+00 4.388e+04 0.000 0.99995   
## age32.8 2.404e+01 2.764e+04 0.001 0.99931   
## age32.9 7.420e+00 3.541e+04 0.000 0.99983   
## age33 8.887e+00 4.076e+04 0.000 0.99983   
## age33.2 3.332e+00 3.374e+04 0.000 0.99992   
## age33.3 5.095e+01 5.556e+04 0.001 0.99927   
## age33.4 8.671e+00 5.556e+04 0.000 0.99988   
## age33.5 2.837e+01 2.764e+04 0.001 0.99918   
## age33.6 7.948e+00 4.303e+04 0.000 0.99985   
## age33.7 2.453e+01 2.764e+04 0.001 0.99929   
## age33.8 6.920e+00 5.556e+04 0.000 0.99990   
## age33.9 2.144e+00 3.437e+04 0.000 0.99995   
## age34 1.118e+01 4.254e+04 0.000 0.99979   
## age34.1 4.138e+01 5.556e+04 0.001 0.99941   
## age34.2 2.652e+01 2.764e+04 0.001 0.99923   
## age34.3 6.747e+00 4.316e+04 0.000 0.99988   
## age34.4 6.221e+00 5.556e+04 0.000 0.99991   
## age34.5 -6.885e-01 3.827e+04 0.000 0.99999   
## age34.6 7.129e+00 5.556e+04 0.000 0.99990   
## age34.7 9.567e+00 5.556e+04 0.000 0.99986   
## age34.8 4.335e+00 5.556e+04 0.000 0.99994   
## age34.9 2.736e+01 2.764e+04 0.001 0.99921   
## age35 -2.477e+00 3.473e+04 0.000 0.99994   
## age35.1 3.341e+00 4.281e+04 0.000 0.99994   
## age35.2 -1.887e+00 5.556e+04 0.000 0.99997   
## age35.3 7.331e+00 5.556e+04 0.000 0.99989   
## age35.4 7.361e+00 3.717e+04 0.000 0.99984   
## age35.5 -4.056e+00 5.556e+04 0.000 0.99994   
## age35.6 2.252e+00 3.575e+04 0.000 0.99995   
## age35.8 -4.544e+00 3.703e+04 0.000 0.99990   
## age35.9 -1.146e+00 3.810e+04 0.000 0.99998   
## age36 8.615e+00 5.556e+04 0.000 0.99988   
## age36.1 8.820e+00 4.349e+04 0.000 0.99984   
## age36.2 2.808e+00 3.502e+04 0.000 0.99994   
## age36.3 4.714e+00 3.363e+04 0.000 0.99989   
## age36.4 6.694e+00 5.556e+04 0.000 0.99990   
## age36.5 2.605e+01 2.764e+04 0.001 0.99925   
## age36.6 7.245e+00 5.556e+04 0.000 0.99990   
## age36.7 1.589e+00 3.519e+04 0.000 0.99996   
## age36.8 9.002e+00 5.556e+04 0.000 0.99987   
## age37 -1.188e+00 5.556e+04 0.000 0.99998   
## age37.2 2.581e+01 2.764e+04 0.001 0.99926   
## age37.3 2.619e+01 2.764e+04 0.001 0.99924   
## age37.4 2.909e+01 2.764e+04 0.001 0.99916   
## age37.5 6.436e+00 3.689e+04 0.000 0.99986   
## age37.6 7.858e+00 5.556e+04 0.000 0.99989   
## age37.8 1.727e+00 3.926e+04 0.000 0.99996   
## age38 7.485e+00 3.807e+04 0.000 0.99984   
## age38.1 4.380e+00 3.944e+04 0.000 0.99991   
## age38.2 8.593e+00 5.556e+04 0.000 0.99988   
## age38.3 3.004e+01 2.764e+04 0.001 0.99913   
## age38.4 -9.730e-03 3.272e+04 0.000 1.00000   
## age38.5 8.916e+00 4.319e+04 0.000 0.99984   
## age38.6 8.249e+00 3.850e+04 0.000 0.99983   
## age38.7 2.588e+01 2.764e+04 0.001 0.99925   
## age38.8 1.263e+00 5.556e+04 0.000 0.99998   
## age38.9 5.433e+00 4.006e+04 0.000 0.99989   
## age39 4.402e+00 3.648e+04 0.000 0.99990   
## age39.1 3.915e-02 4.061e+04 0.000 1.00000   
## age39.2 2.717e+01 2.764e+04 0.001 0.99922   
## age39.4 4.952e+00 5.556e+04 0.000 0.99993   
## age39.5 1.170e+01 5.556e+04 0.000 0.99983   
## age39.6 2.396e+00 3.896e+04 0.000 0.99995   
## age39.7 3.012e+01 2.764e+04 0.001 0.99913   
## age39.8 2.675e+01 2.764e+04 0.001 0.99923   
## age39.9 4.895e+00 5.556e+04 0.000 0.99993   
## age40 9.652e+00 5.556e+04 0.000 0.99986   
## age40.1 6.160e+00 3.683e+04 0.000 0.99987   
## age40.3 2.644e+00 3.231e+04 0.000 0.99993   
## age40.4 5.241e+00 4.010e+04 0.000 0.99990   
## age40.6 1.070e+01 5.556e+04 0.000 0.99985   
## age40.8 9.223e+00 5.556e+04 0.000 0.99987   
## age40.9 8.032e+00 4.293e+04 0.000 0.99985   
## age41 4.648e+00 5.556e+04 0.000 0.99993   
## age41.1 2.705e+01 2.764e+04 0.001 0.99922   
## age41.2 -1.671e+00 3.511e+04 0.000 0.99996   
## age41.3 2.679e+01 2.764e+04 0.001 0.99923   
## age41.4 2.821e+01 2.764e+04 0.001 0.99919   
## age41.6 6.731e+00 5.556e+04 0.000 0.99990   
## age41.7 2.636e+01 2.764e+04 0.001 0.99924   
## age41.9 5.250e+00 4.091e+04 0.000 0.99990   
## age42 4.473e+00 3.551e+04 0.000 0.99990   
## age42.1 3.215e+01 5.556e+04 0.001 0.99954   
## age42.3 4.677e+00 3.790e+04 0.000 0.99990   
## age42.4 5.752e+00 3.781e+04 0.000 0.99988   
## age42.5 7.104e+00 3.712e+04 0.000 0.99985   
## age42.6 6.322e-01 5.556e+04 0.000 0.99999   
## age42.8 7.956e+00 5.556e+04 0.000 0.99989   
## age43 6.319e+00 3.649e+04 0.000 0.99986   
## age43.1 8.766e+00 5.556e+04 0.000 0.99987   
## age43.2 -3.749e+00 3.825e+04 0.000 0.99992   
## age43.3 1.573e+00 4.102e+04 0.000 0.99997   
## age43.4 8.901e-01 3.526e+04 0.000 0.99998   
## age43.5 4.892e+00 3.485e+04 0.000 0.99989   
## age43.6 2.117e+01 2.764e+04 0.001 0.99939   
## age43.7 5.740e+00 5.556e+04 0.000 0.99992   
## age43.8 1.191e+00 3.242e+04 0.000 0.99997   
## age44 6.958e+00 3.871e+04 0.000 0.99986   
## age44.1 2.938e+01 2.764e+04 0.001 0.99915   
## age44.2 7.948e+00 5.556e+04 0.000 0.99989   
## age44.3 9.121e+00 4.099e+04 0.000 0.99982   
## age44.4 4.933e+01 5.556e+04 0.001 0.99929   
## age44.5 5.604e+00 5.556e+04 0.000 0.99992   
## age44.6 4.268e-01 5.556e+04 0.000 0.99999   
## age44.7 2.228e+01 2.764e+04 0.001 0.99936   
## age44.8 3.800e+00 4.095e+04 0.000 0.99993   
## age44.9 2.504e+01 2.764e+04 0.001 0.99928   
## age45 2.351e+01 2.764e+04 0.001 0.99932   
## age45.1 3.714e+00 3.647e+04 0.000 0.99992   
## age45.4 6.337e+00 3.622e+04 0.000 0.99986   
## age45.5 5.925e+00 5.556e+04 0.000 0.99991   
## age45.6 7.678e+00 4.291e+04 0.000 0.99986   
## age45.8 4.150e+01 5.556e+04 0.001 0.99940   
## age45.9 5.261e+00 4.388e+04 0.000 0.99990   
## age46 7.897e+00 4.380e+04 0.000 0.99986   
## age46.1 7.956e-01 3.941e+04 0.000 0.99998   
## age46.2 7.788e+00 5.556e+04 0.000 0.99989   
## age46.3 3.121e+00 3.635e+04 0.000 0.99993   
## age46.4 7.820e+00 4.260e+04 0.000 0.99985   
## age46.5 3.639e+01 5.556e+04 0.001 0.99948   
## age46.6 1.031e+01 5.556e+04 0.000 0.99985   
## age46.7 -6.935e+00 3.241e+04 0.000 0.99983   
## age46.8 9.348e+00 5.556e+04 0.000 0.99987   
## age46.9 7.105e+00 4.229e+04 0.000 0.99987   
## age47 2.979e+01 2.764e+04 0.001 0.99914   
## age47.1 6.740e+00 3.795e+04 0.000 0.99986   
## age47.2 6.797e+00 4.106e+04 0.000 0.99987   
## age47.3 7.553e+00 5.556e+04 0.000 0.99989   
## age47.5 3.119e+00 3.652e+04 0.000 0.99993   
## age47.7 -9.085e-01 5.556e+04 0.000 0.99999   
## age47.8 6.207e+00 3.842e+04 0.000 0.99987   
## age48 3.622e+00 5.556e+04 0.000 0.99995   
## age48.2 2.709e+01 2.764e+04 0.001 0.99922   
## age48.4 4.958e+01 5.556e+04 0.001 0.99929   
## age48.5 2.635e+01 2.764e+04 0.001 0.99924   
## age48.7 6.989e+00 3.823e+04 0.000 0.99985   
## age48.8 2.386e+01 2.764e+04 0.001 0.99931   
## age48.9 3.308e+00 5.556e+04 0.000 0.99995   
## age49 8.759e+00 4.298e+04 0.000 0.99984   
## age49.3 4.242e+01 5.556e+04 0.001 0.99939   
## age49.9 -5.178e+00 4.032e+04 0.000 0.99990   
## age50.1 6.420e+00 4.312e+04 0.000 0.99988   
## age50.2 1.012e+01 4.111e+04 0.000 0.99980   
## age50.5 5.569e+00 5.556e+04 0.000 0.99992   
## age50.6 4.270e+00 5.556e+04 0.000 0.99994   
## age50.9 4.268e-01 5.556e+04 0.000 0.99999   
## age51 9.845e+00 5.556e+04 0.000 0.99986   
## age51.1 7.165e-01 3.323e+04 0.000 0.99998   
## age51.2 2.018e+00 5.556e+04 0.000 0.99997   
## age51.3 6.818e+00 5.556e+04 0.000 0.99990   
## age51.4 4.406e+01 5.556e+04 0.001 0.99937   
## age51.7 9.180e+00 5.556e+04 0.000 0.99987   
## age51.8 1.847e+00 3.687e+04 0.000 0.99996   
## age52.1 1.240e+00 5.556e+04 0.000 0.99998   
## age52.5 4.466e+01 5.556e+04 0.001 0.99936   
## age52.6 7.002e+00 5.556e+04 0.000 0.99990   
## age53 -1.386e-01 4.138e+04 0.000 1.00000   
## age53.5 1.240e-01 3.611e+04 0.000 1.00000   
## age53.8 8.166e+00 5.556e+04 0.000 0.99988   
## age53.9 2.214e-01 5.556e+04 0.000 1.00000   
## age54.1 7.553e+00 5.556e+04 0.000 0.99989   
## age54.4 2.958e+00 4.029e+04 0.000 0.99994   
## age54.5 2.492e+00 5.556e+04 0.000 0.99996   
## age54.8 4.360e+00 5.556e+04 0.000 0.99994   
## age54.9 3.970e+00 5.556e+04 0.000 0.99994   
## age55 1.054e+00 4.322e+04 0.000 0.99998   
## age55.7 8.585e+00 5.556e+04 0.000 0.99988   
## age56.4 2.638e+00 5.556e+04 0.000 0.99996   
## age56.5 1.050e+01 5.556e+04 0.000 0.99985   
## age56.8 5.158e+00 4.388e+04 0.000 0.99991   
## age57.5 5.157e+00 4.388e+04 0.000 0.99991   
## age58.5 3.049e+00 5.556e+04 0.000 0.99996   
## age59.4 3.632e+00 4.126e+04 0.000 0.99993   
## age61.4 3.241e-01 5.556e+04 0.000 1.00000   
## age61.6 3.306e+00 5.556e+04 0.000 0.99995   
## age63.4 6.038e+00 5.556e+04 0.000 0.99991   
## age65.1 9.064e+00 5.556e+04 0.000 0.99987   
## age67 8.375e+00 5.556e+04 0.000 0.99988   
## stateKentucky 8.711e-01 1.070e+00 0.814 0.41544   
## stateLouisiana 7.621e+00 2.103e+00 3.624 0.00029 \*\*\*  
## stateVirginia -7.404e+00 1.459e+00 -5.073 3.91e-07 \*\*\*  
## time.served -1.027e+00 3.736e-01 -2.749 0.00598 \*\*   
## max.sentence 6.082e-01 2.150e-01 2.829 0.00468 \*\*   
## multiple.offensesyes 2.505e+00 9.669e-01 2.591 0.00956 \*\*   
## crimelarceny 1.151e+00 1.148e+00 1.003 0.31588   
## crimedrug-related -7.236e-01 9.377e-01 -0.772 0.44033   
## crimedriving-related 1.230e+00 1.197e+00 1.027 0.30435   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 483.267 on 674 degrees of freedom  
## Residual deviance: 83.336 on 362 degrees of freedom  
## AIC: 709.34  
##   
## Number of Fisher Scoring iterations: 21

emptymod = glm(violator ~1, parole, family = "binomial")   
summary(emptymod)

##   
## Call:  
## glm(formula = violator ~ 1, family = "binomial", data = parole)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.4956 -0.4956 -0.4956 -0.4956 2.0775   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -2.0352 0.1204 -16.9 <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: 483.27 on 674 degrees of freedom  
## Residual deviance: 483.27 on 674 degrees of freedom  
## AIC: 485.27  
##   
## Number of Fisher Scoring iterations: 4

forwardmod = stepAIC(emptymod, direction = "forward", scope=list(upper=allmod,lower=emptymod),  
 trace = TRUE)

## Start: AIC=485.27  
## violator ~ 1  
##   
## Df Deviance AIC  
## + state 3 382.89 390.89  
## + max.sentence 1 465.68 469.68  
## + multiple.offenses 1 475.81 479.81  
## + time.served 1 477.05 481.05  
## + race 1 479.56 483.56  
## <none> 483.27 485.27  
## + male 1 483.17 487.17  
## + crime 3 480.48 488.48  
## + age 301 231.44 835.44  
##   
## Step: AIC=390.89  
## violator ~ state  
##   
## Df Deviance AIC  
## + multiple.offenses 1 358.69 368.69  
## + race 1 376.71 386.71  
## <none> 382.89 390.89  
## + time.served 1 381.65 391.65  
## + max.sentence 1 381.93 391.93  
## + male 1 382.16 392.16  
## + crime 3 380.87 394.87  
## + age 301 117.95 727.95  
##   
## Step: AIC=368.69  
## violator ~ state + multiple.offenses

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

## Df Deviance AIC  
## + race 1 353.26 365.26  
## <none> 358.69 368.69  
## + max.sentence 1 356.73 368.73  
## + time.served 1 358.02 370.02  
## + male 1 358.04 370.04  
## + crime 3 357.47 373.47  
## + age 301 108.18 720.18  
##   
## Step: AIC=365.26  
## violator ~ state + multiple.offenses + race

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

## Df Deviance AIC  
## <none> 353.26 365.26  
## + max.sentence 1 351.62 365.62  
## + time.served 1 352.43 366.43  
## + male 1 352.71 366.71  
## + crime 3 351.81 369.81  
## + age 301 107.40 721.40

summary(forwardmod)

##   
## Call:  
## glm(formula = violator ~ state + multiple.offenses + race, family = "binomial",   
## data = parole)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.4012 -0.4051 -0.2604 -0.1801 2.8739   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -2.50359 0.30055 -8.330 < 2e-16 \*\*\*  
## stateKentucky 0.04449 0.39449 0.113 0.9102   
## stateLouisiana 0.75016 0.39147 1.916 0.0553 .   
## stateVirginia -3.12945 0.51147 -6.119 9.44e-10 \*\*\*  
## multiple.offensesyes 1.51964 0.32027 4.745 2.09e-06 \*\*\*  
## raceother 0.74594 0.31828 2.344 0.0191 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 483.27 on 674 degrees of freedom  
## Residual deviance: 353.26 on 669 degrees of freedom  
## AIC: 365.26  
##   
## Number of Fisher Scoring iterations: 6

# The AIC tends to decrease in the forward model, making it a good model to follow.

# Task 5

mod1 = glm(violator ~ state + multiple.offenses + race , parole, family = "binomial")  
summary(mod1)

##   
## Call:  
## glm(formula = violator ~ state + multiple.offenses + race, family = "binomial",   
## data = parole)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.4012 -0.4051 -0.2604 -0.1801 2.8739   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -2.50359 0.30055 -8.330 < 2e-16 \*\*\*  
## stateKentucky 0.04449 0.39449 0.113 0.9102   
## stateLouisiana 0.75016 0.39147 1.916 0.0553 .   
## stateVirginia -3.12945 0.51147 -6.119 9.44e-10 \*\*\*  
## multiple.offensesyes 1.51964 0.32027 4.745 2.09e-06 \*\*\*  
## raceother 0.74594 0.31828 2.344 0.0191 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 483.27 on 674 degrees of freedom  
## Residual deviance: 353.26 on 669 degrees of freedom  
## AIC: 365.26  
##   
## Number of Fisher Scoring iterations: 6

# Parolees from the state of Lousiana, with multiple offenses and a race other than white are predicted as more likely to violate their parol.

# Task 6

newdata = data.frame(state= "Louisiana", race = "white", multiple.offenses = "yes")  
predict(forwardmod, newdata, type= "response")

## 1   
## 0.4418174

# There is 44% probability that a white parolee from Lousiana with multiple offenses is likely to violate their parole.

newdata = data.frame(state= "Kentucky", race = "other", multiple.offenses = "no")  
predict(forwardmod, newdata, type= "response")

## 1   
## 0.152755

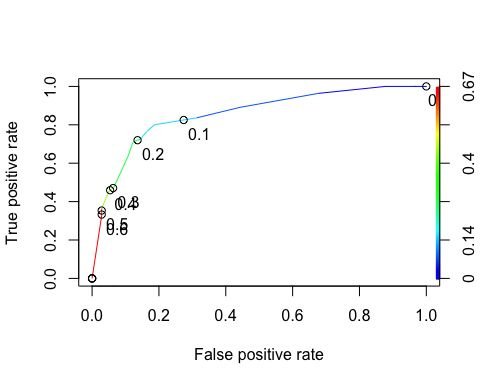
# There is 15% probability that a parolee from the state of Kentucky, with a race other than white and only one offense is likely to violate their parole.

# Task 7

predictions = predict(forwardmod, train, type="response")  
head(predictions)

## 1 2 3 4 5 6   
## 0.07560706 0.14708578 0.07560706 0.14708578 0.14708578 0.07560706

ROCRpred = prediction(predictions, train$violator)  
  
ROCRperf = performance(ROCRpred, "tpr", "fpr")  
plot(ROCRperf, colorize=TRUE, print.cutoffs.at=seq(0,1,by=0.1), text.adj=c(-0.2,1.7))



# Task 8

as.numeric(performance(ROCRpred, "auc")@y.values)

## [1] 0.8540887

opt.cut = function(perf, pred){  
 cut.ind = mapply(FUN=function(x, y, p){  
 d = (x - 0)^2 + (y-1)^2  
 ind = which(d == min(d))  
 c(sensitivity = y[[ind]], specificity = 1-x[[ind]],   
 cutoff = p[[ind]])  
 }, perf@x.values, perf@y.values, pred@cutoffs)  
}  
print(opt.cut(ROCRperf, ROCRpred))

## [,1]  
## sensitivity 0.8000000  
## specificity 0.8133971  
## cutoff 0.1470858

#confusion matrix  
t1 = table(train$violator,predictions > 0.1470858)  
t1

##   
## FALSE TRUE  
## no 350 68  
## yes 13 42

# Accuracy

(t1[1,1]+t1[2,2])/nrow(train)

## [1] 0.8287526

# The accuracy of the model is 82%, the sensitivity is 80% and the specificity is 81%. The implications of incorrectly classifying a parolee could result in unexpected violations. Therefore, precautionary measures should be taken to prevent those violations.

# Task 9

# Trial and error

(t1[1,1]+t1[2,2])/nrow(train)

## [1] 0.8287526

t1 = table(train$violator, predictions > 0.5)  
t1

##   
## FALSE TRUE  
## no 406 12  
## yes 37 18

(t1[1,1]+t1[2,2])/nrow(train)

## [1] 0.8964059

t1 = table(train$violator,predictions > 0.6)  
t1

##   
## FALSE TRUE  
## no 406 12  
## yes 37 18

(t1[1,1]+t1[2,2])/nrow(train)

## [1] 0.8964059

t1 = table(train$violator,predictions > 1)  
t1

##   
## FALSE  
## no 418  
## yes 55

(t1[1])/nrow(train)

## [1] 0.8837209

# Parolees are actually more likely to violate their parole than the first model predicts.