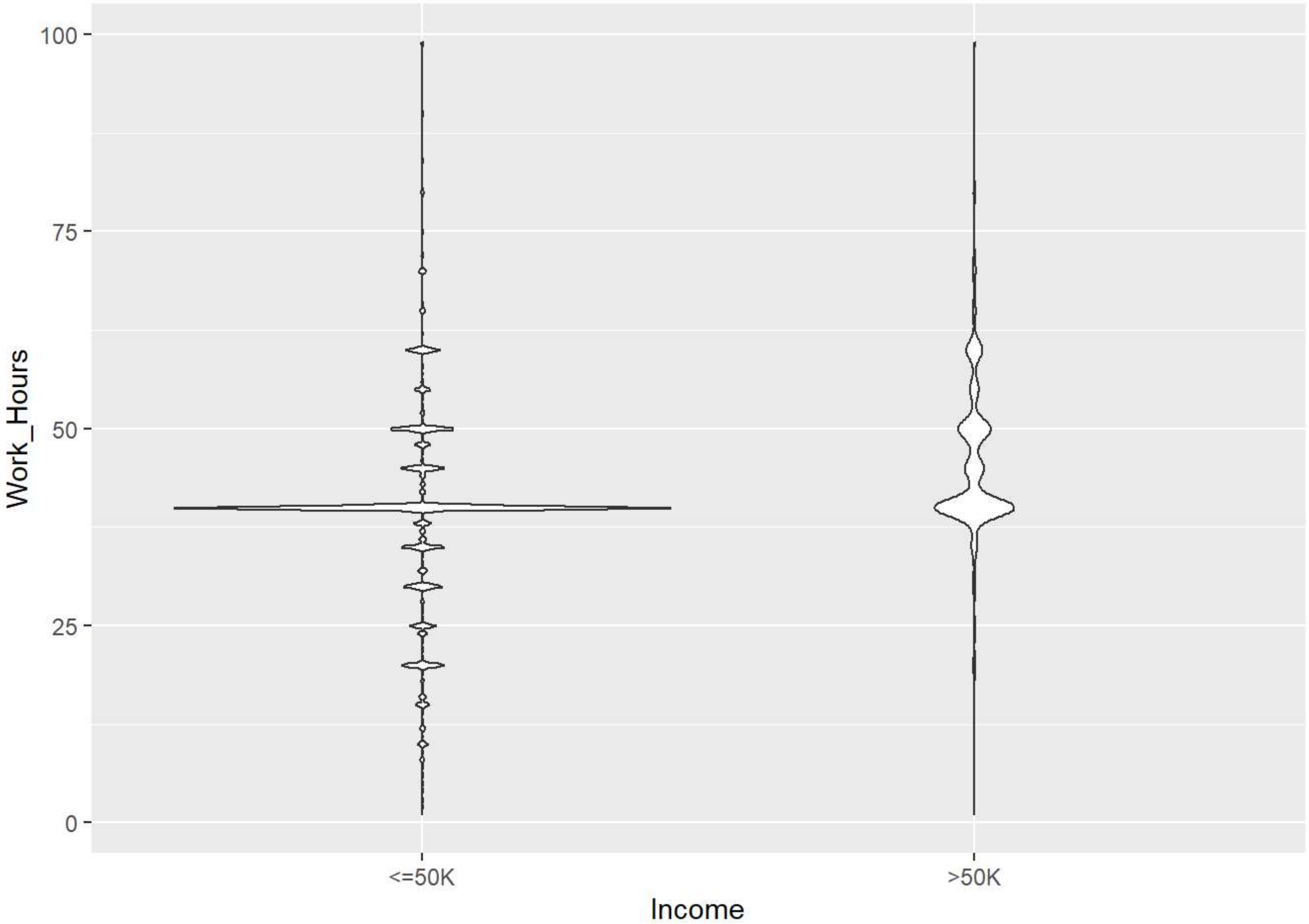
810_Team_8

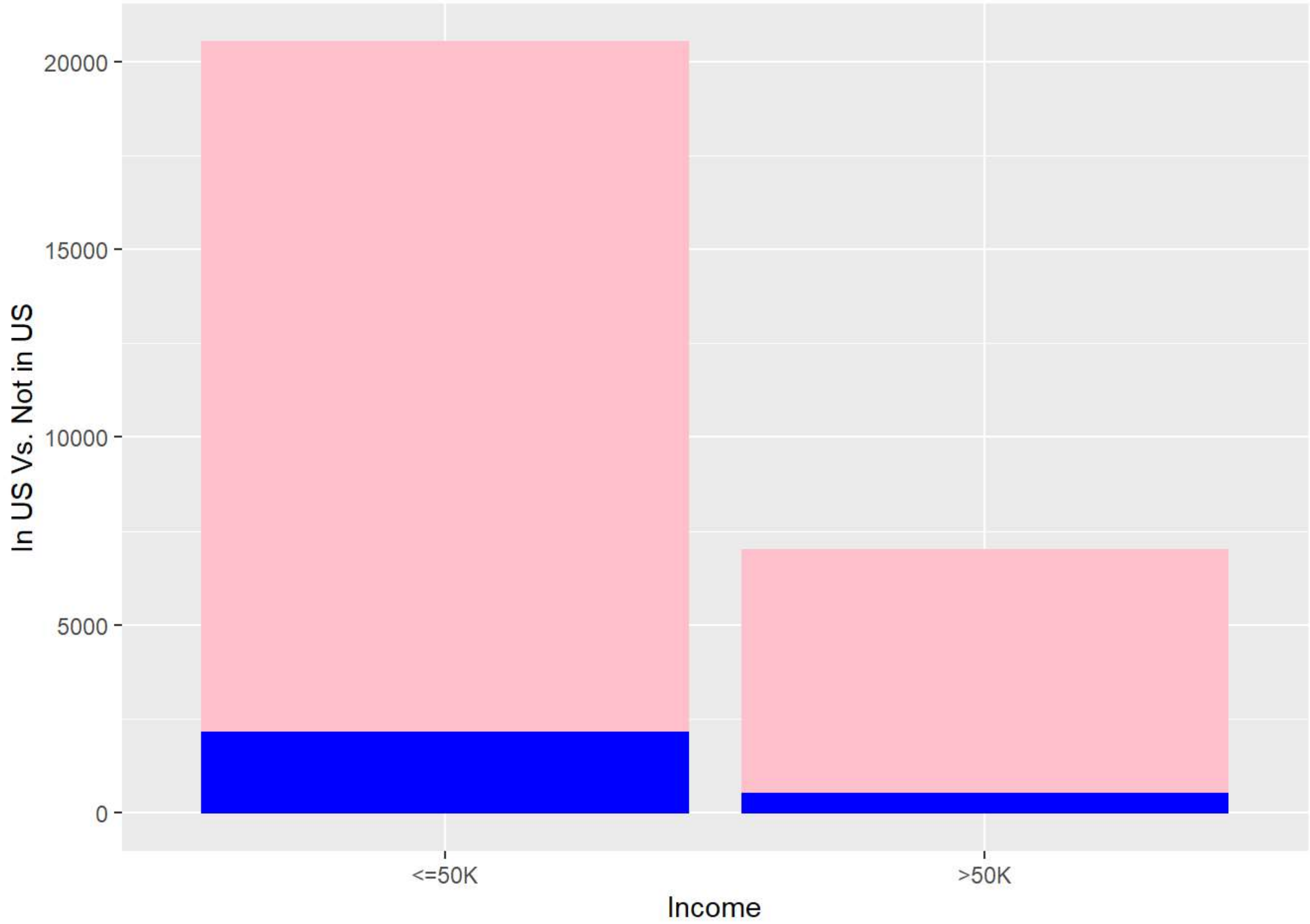
Sihoon Yu (U65129043)

10/13/2021

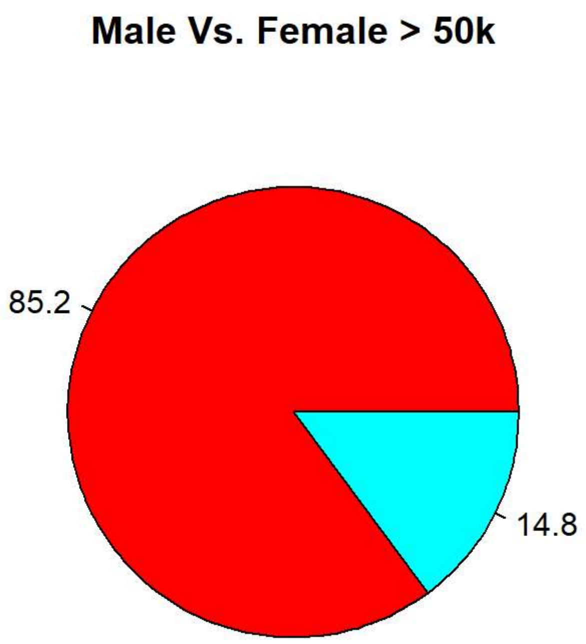
```
df <- fread("adult_revised.csv",stringsAsFactors = TRUE)


#Creating dummy variables, removing the first column and selected columns
dd <- dummy_cols(df, select_columns = c('Employment', 'Education', 'Marital_status','Occupati
on', 'Relationship', 'Race', 'Sex'),
                      remove_first_dummy = TRUE, remove_selected_columns = TRUE)

#Setting US as 1 and others as 0
dd$US<-ifelse(dd$Country == 'United-States',1,0)
dd$Country <- NULL

#rearranges column order so that income is placed in the last row
dd <- dd[,c(1,2,3,4,5,6,8:58,7)]
names(dd) <- make.names(names(dd))

#Creating train and test set and defining formula
id <- createDataPartition(dd$Income, p = 0.8, list = FALSE)
train<-dd[id, ]
test<-dd[-id, ]
formula <- as.formula(Income ~ .)

train2 <- train
test2 <- test
test2$Income <- ifelse(test2$Income == "<=50K", 0, 1)
train2$Income <- ifelse(train2$Income == "<=50K", 0, 1)
y_train <- train2$Income
y_test <- test2$Income
X_train <- model.matrix(as.formula(Income~.),train)[,-1]
X_test <-  model.matrix(as.formula(Income~.),test)[,-1]

#Exploring with the Data

#Violin Graph of Income vs Work Hours
ggplot(data=dd ,mapping=aes(y=Work_Hours,x=Income))+
  geom_violin()
```




```
#Bar Graph of US vs non-US citizens
dd2 <- dd
dd2$INUS<-ifelse(dd2$US == '1',1,0)
dd2$NOTUS<-ifelse(dd2$US == '0',1,0)
ggplot(data=dd2)+
  geom_col(mapping=aes(y=INUS,x=Income),color="pink")+
  geom_col(mapping=aes(y=NOTUS,x=Income),color="blue")+
  ylab("In US Vs. Not in US")
```

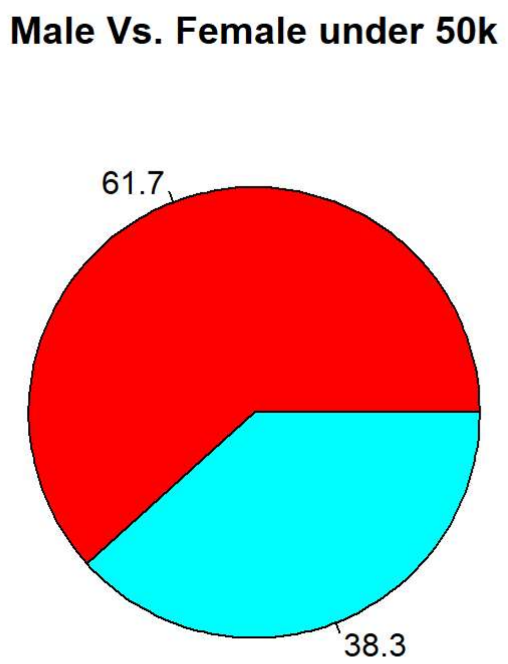



```
#Pie Chart 1

dd2$MFIVEK<-ifelse(dd2$Sex_Male=='1'&dd2$Income=='>50K',1,0)
dd2$FFIVEK<-ifelse(dd2$Sex_Male=='0'&dd2$Income=='>50K',1,0)
dd2$MFIVEK2<-ifelse(dd2$Sex_Male=='1'&dd2$Income=='<=50K',1,0)
dd2$FFIVEK2<-ifelse(dd2$Sex_Male=='0'&dd2$Income=='<=50K',1,0)

greater_50k <- c(sum(dd2$MFIVEK),sum(dd2$FFIVEK))
piepercent <- round(100*greater_50k/sum(greater_50k), 1)
labels <- c('Male > 50k','Female > 50k')
print(pie(greater_50k, labels = piepercent, main = 'Male Vs. Female > 50k', col = rainbow(len
gth(greater_50k))))
```




```
## NULL
```

```
#Pie Chart 2
under_50k <- c(sum(dd2$MFIVEK2),sum(dd2$FFIVEK2))
piepercent <- round(100*under_50k/sum(under_50k), 1)
labels <- c('Male <= 50k','Female <= 50k')
print(pie(under_50k, labels = piepercent, main = 'Male Vs. Female under 50k', col = rainbow(l
ength(under_50k))))
```



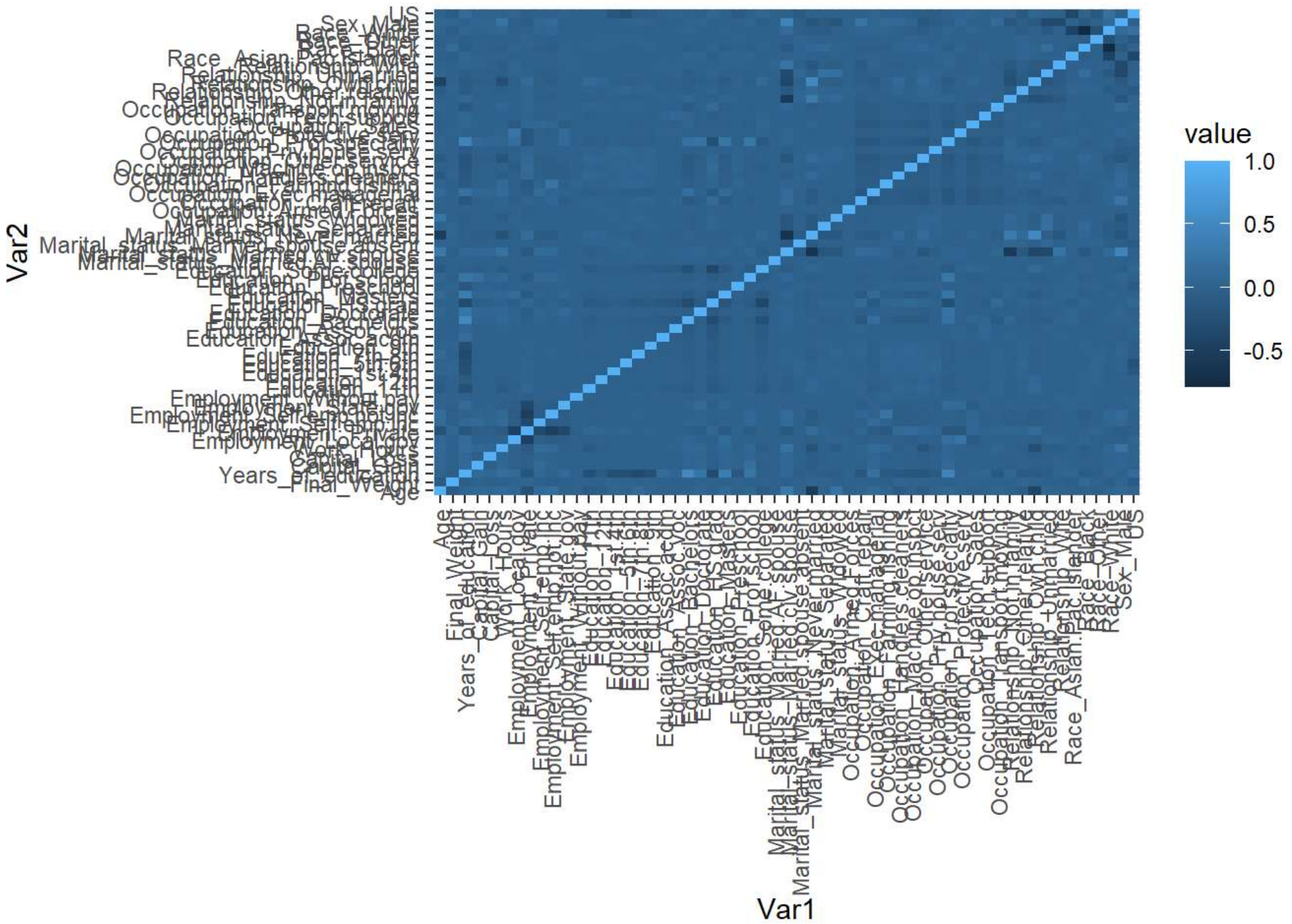

```
## NULL
```

```
#Point Plot with Education on Age
print(ggplot(dd, aes(x = Age, y = Years_of_education, color = Income))+geom_point(position =
"jitter")+
  labs(x = "Age", y = "Years of Education", color = "Income"))
```


```
#Correlation Heatmap
ddf <- dd
ddf$Income <- NULL
cormat <- round(cor(ddf),2)
melted_cormat <- melt(cormat)
```

```
## Warning in melt(cormat): The melt generic in data.table has been passed a matrix
## and will attempt to redirect to the relevant reshape2 method; please note that
## reshape2 is deprecated, and this redirection is now deprecated as well. To
## continue using melt methods from reshape2 while both libraries are attached,
## e.g. melt.list, you can prepend the namespace like reshape2::melt(cormat). In
## the next version, this warning will become an error.
```

```
print(ggplot(data = melted_cormat, aes(x=Var1, y=Var2, fill=value)) +
  geom_tile() + theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)))
```




```
#Simple Linear Regression
fit.lm <- lm(formula, train2)
yhat.train.lm <- predict(fit.lm)
mse.train.lm <- mean((y_train - yhat.train.lm)^2)
print(mse.train.lm)
```

```
## [1] 0.1179064
```

```
#lasso
fit_lasso <- cv.glmnet(X_train, y_train, family = "binomial", alpha = 1, nfolds = 10)
## Predict responses
y_train_hat <-predict(fit_lasso,newx=X_train,type='response')
y_test_hat <-predict(fit_lasso,newx=X_test,type='response')
mse.min <- fit_lasso$cvm[which(fit_lasso$lambda == fit_lasso$lambda.min)]
print(mse.min)
```

```
## [1] 0.6505893
```

```
## Compute MSEs

#ridge
fit_ridge <- cv.glmnet(X_train, y_train, family = "binomial", alpha = 0, nfolds = 10)
## Predict responses
y_train_rid <-predict(fit_ridge,newx=X_train,type='response')
y_test_rid <-predict(fit_ridge,newx=X_test,type='response')
mse.min_rid <- fit_ridge$cvm[which(fit_ridge$lambda == fit_ridge$lambda.min)]
print(mse.min_rid)
```

```
## [1] 0.6759002
```

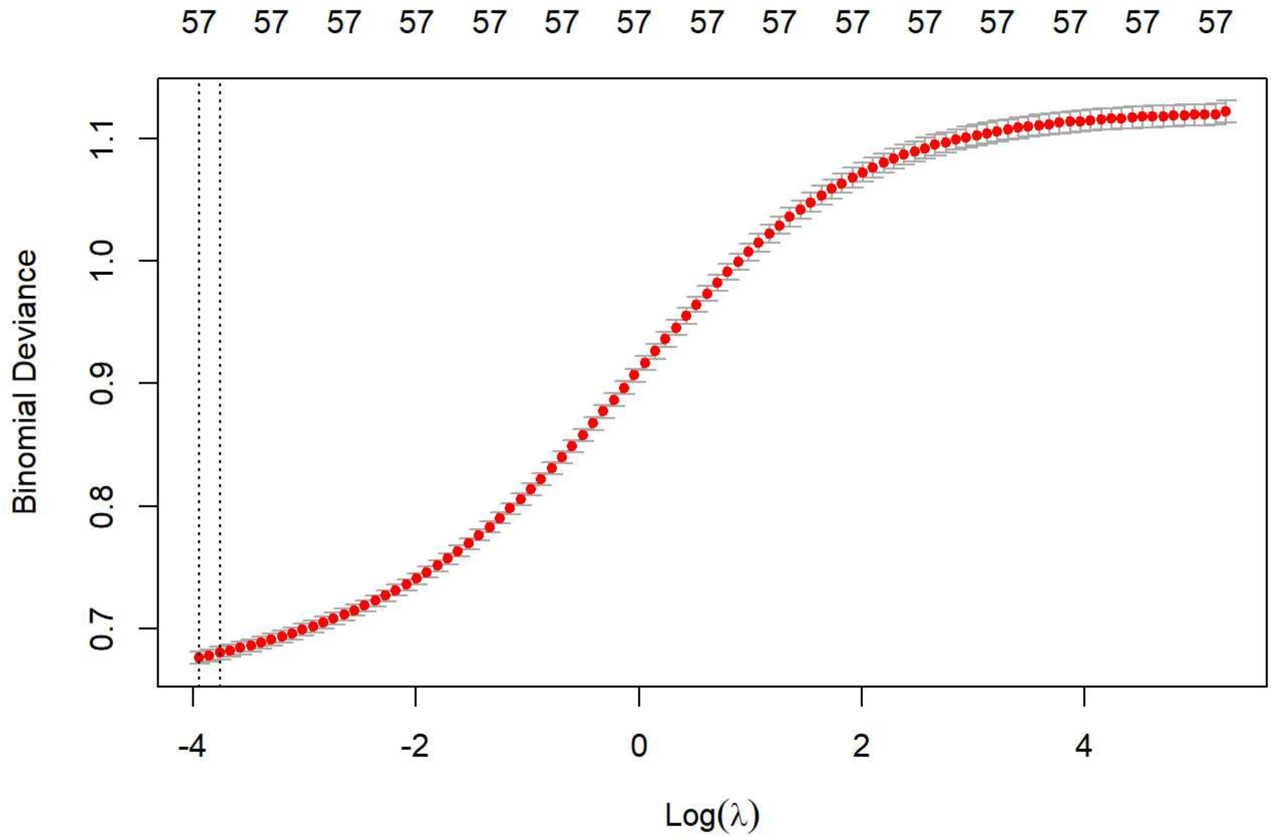
```
#mse_rid <- data.table(lambda = fit_ridge$lambda,mse = mse_train_rid,dataset = "Train")
#mse_rid <- rbind(mse_las, data.table(lambda = fit_ridge$lambda,mse = mse_test_rid,dataset =
 "Test"))

min_mse_rid<-fit_ridge$lambda.min
print(min_mse_rid)
```

```
## [1] 0.01933403
```

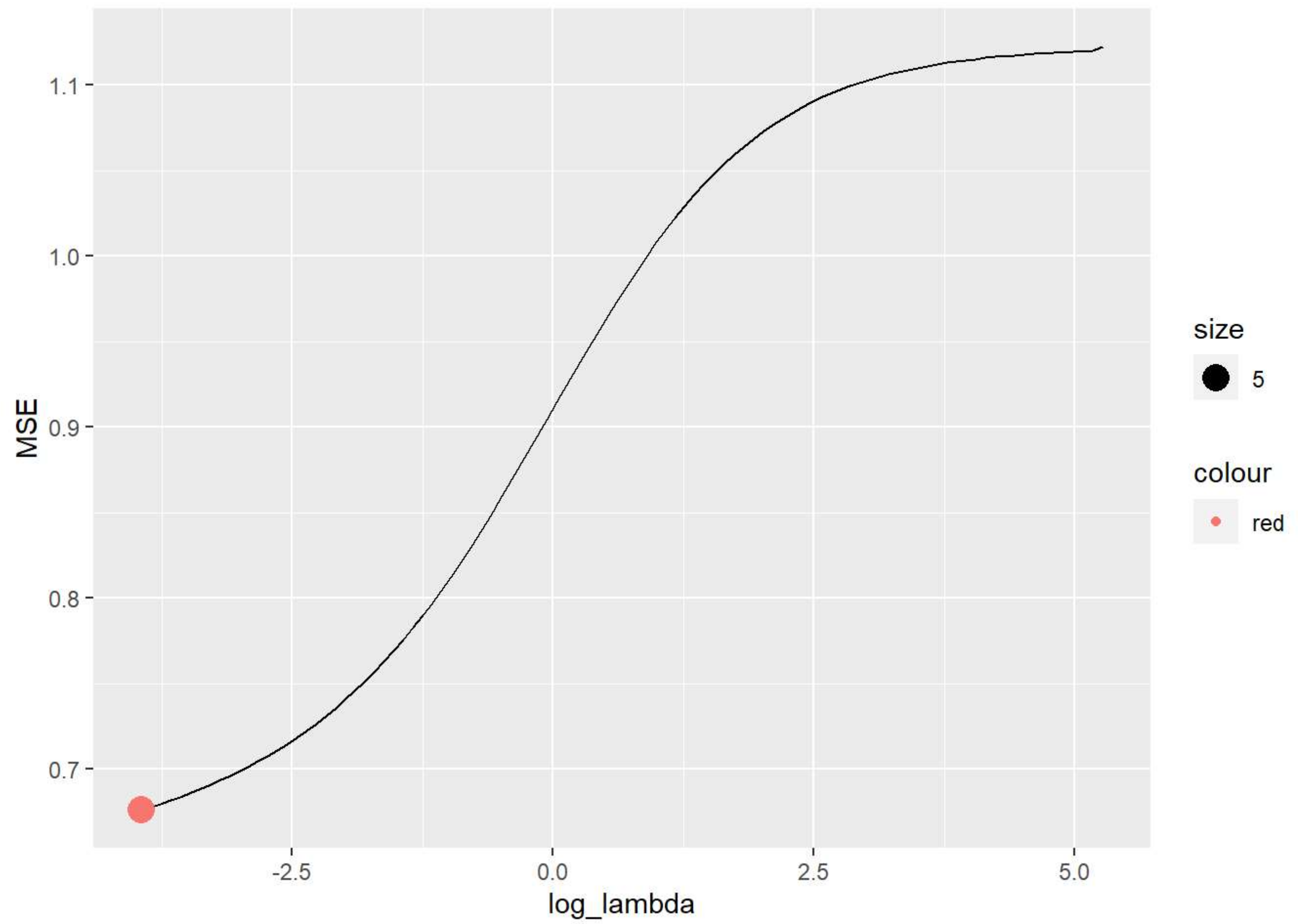
```
print(log(min_mse_rid))
```

```
## [1] -3.945889
```

```
plot(fit_ridge)
```




```
dd_mse_rid <- data.table(log_lambda = log(fit_ridge$lambda), mse = fit_ridge$cvm)
ggplot(dd_mse_rid, aes(log_lambda, mse)) + geom_line() + geom_point(aes(x=log(fit_ridge$lambd
a.min),y=fit_ridge$cvm[which(fit_ridge$lambda == fit_ridge$lambda.min)],colour='red',size=5))
+ scale_y_continuous("MSE")
```



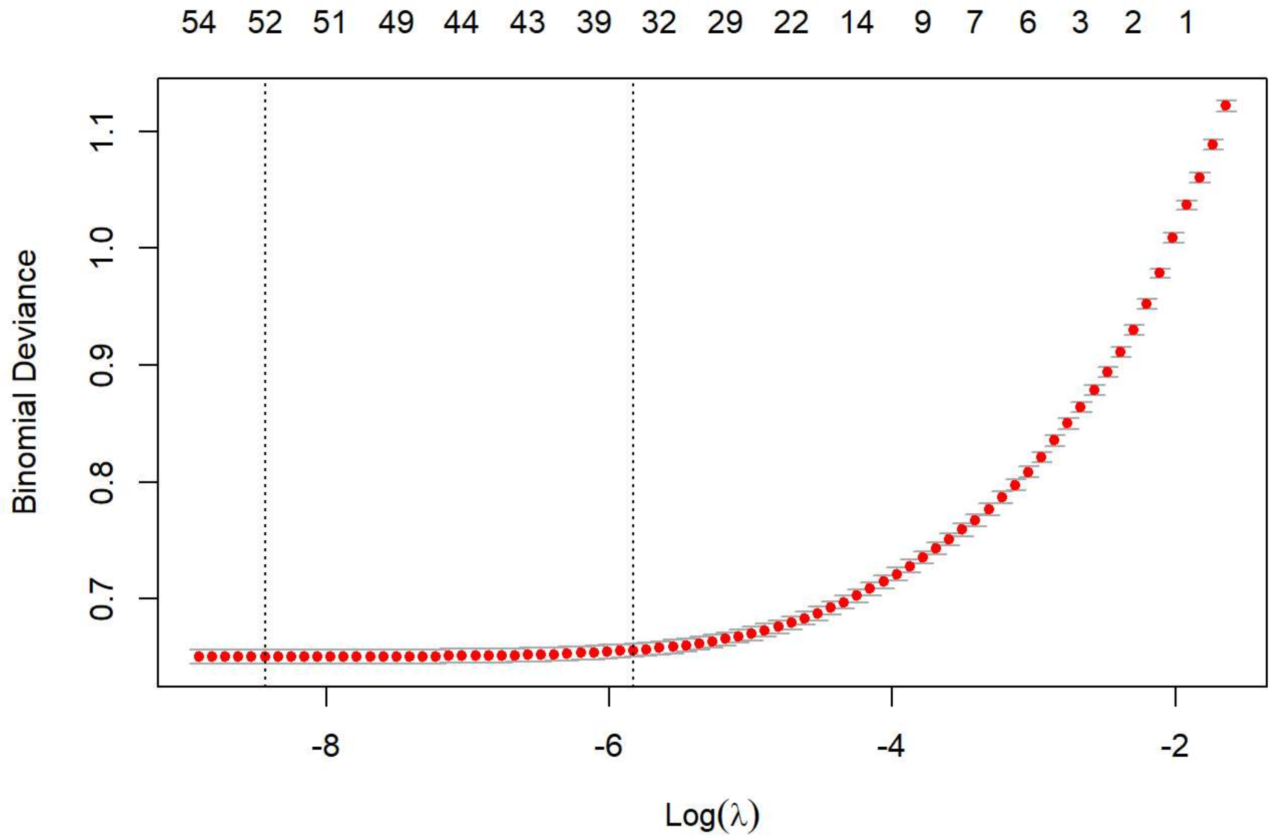

```
###lasso
min_mse_las<-fit_lasso$lambda.min
print(min_mse_las)
```

```
## [1] 0.0002171838
```

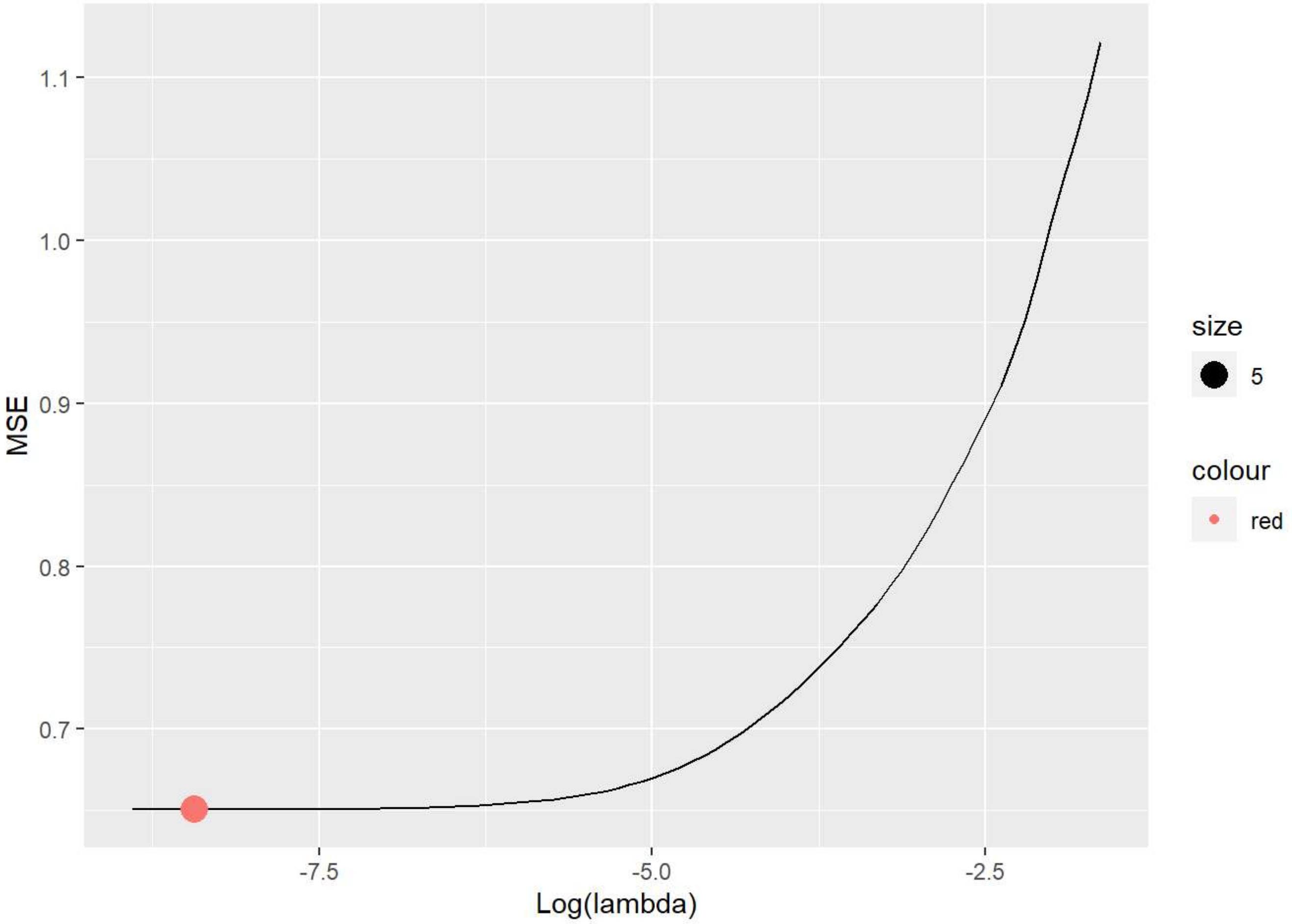
```
print(log(min_mse_las))
```

```
## [1] -8.434767
```

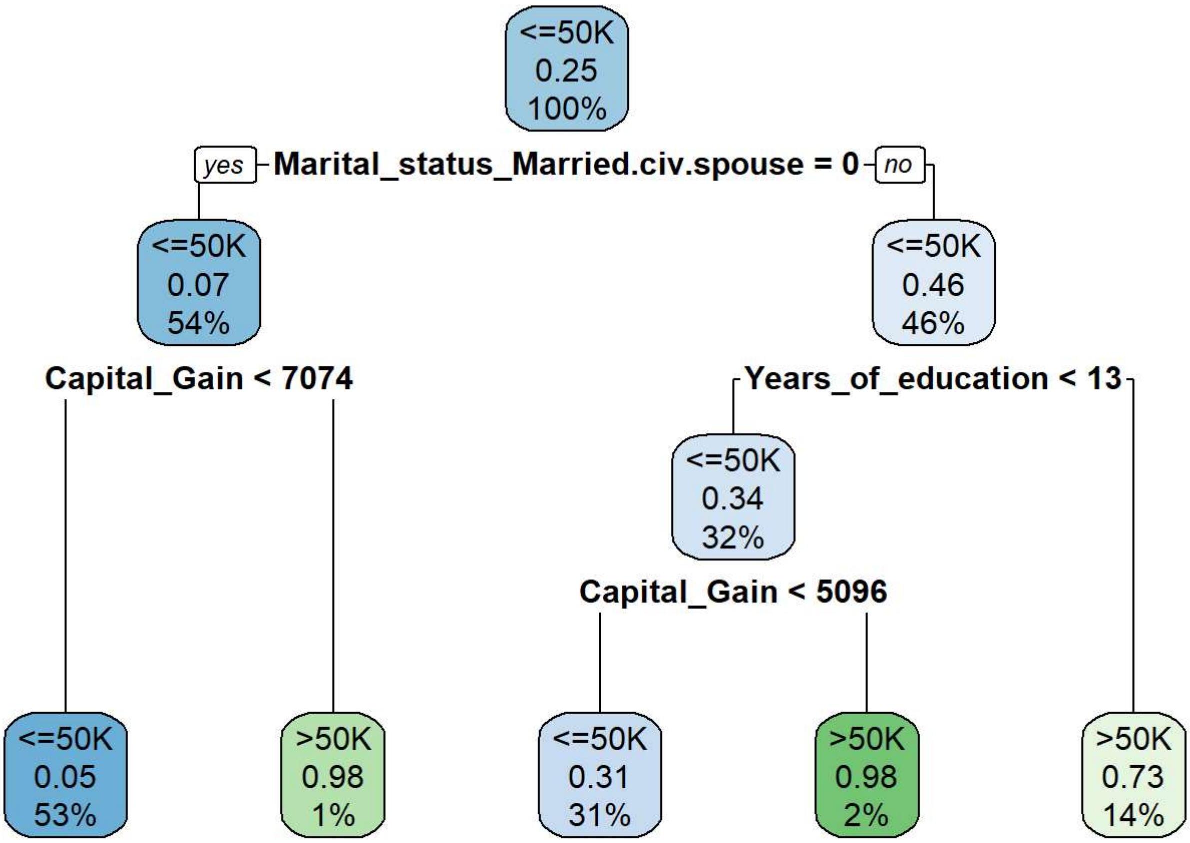
```
#mse_las <- data.table(lambda = fit_lasso$lambda,mse = mse_train_las,dataset = "Train")
#mse_las <- rbind(mse_las, data.table(lambda = fit_lasso$lambda,mse = mse_test_las,dataset =
 "Test"))
plot(fit_lasso)
```




```
#plotting
dd_mse <- data.table(lambda = log(fit_lasso$lambda), mse = fit_lasso$cvm)
ggplot(dd_mse, aes(lambda, mse)) + geom_line() + geom_point(aes(x=log(fit_lasso$lambda.min),y
=fit_lasso$cvm[which(fit_lasso$lambda == fit_lasso$lambda.min)],colour='red',size=5)) + scale
_y_continuous("MSE") + scale_x_continuous("Log(lambda)")
```

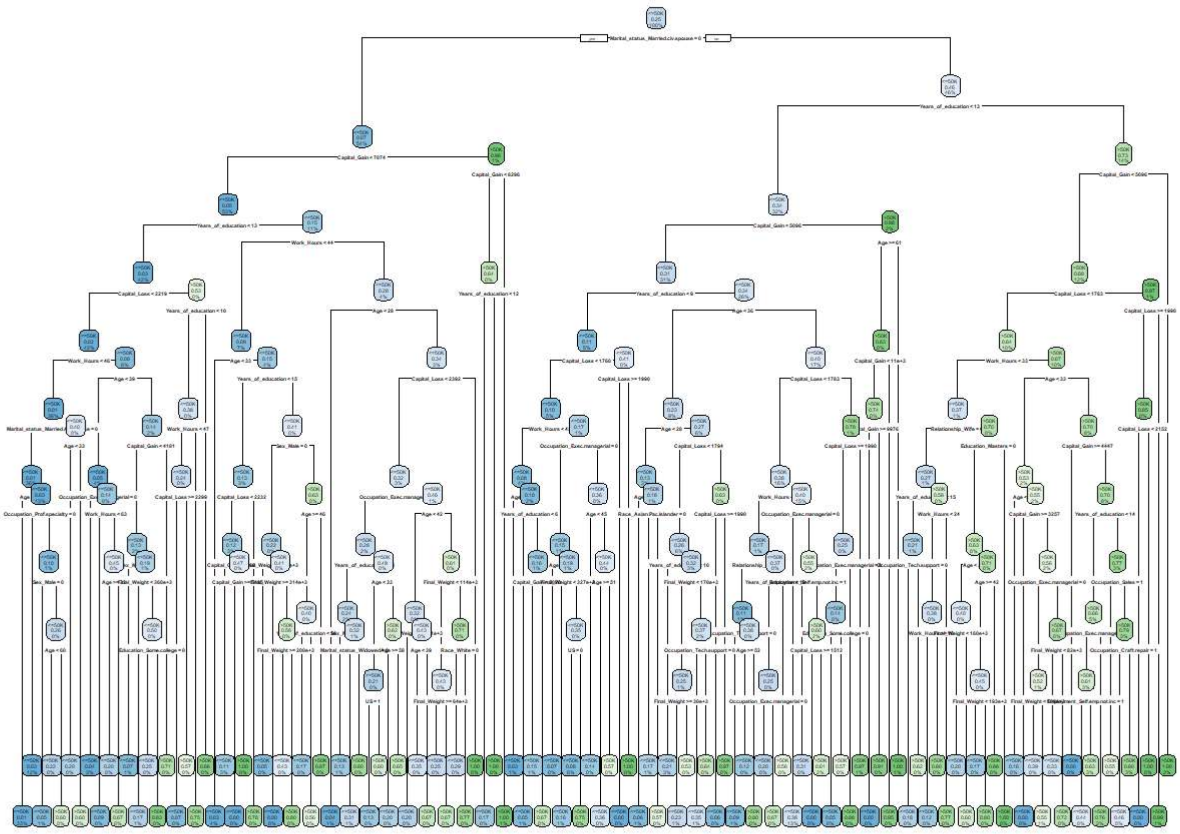



```
#Classification Tree
#making a decision tree
dtm <- rpart(formula, train, method="class", control=rpart.control(cp=0.01,minsplit =0.01, mi
nbucket = 5, maxdepth = 10 ))
rpart.plot(dtm)
```

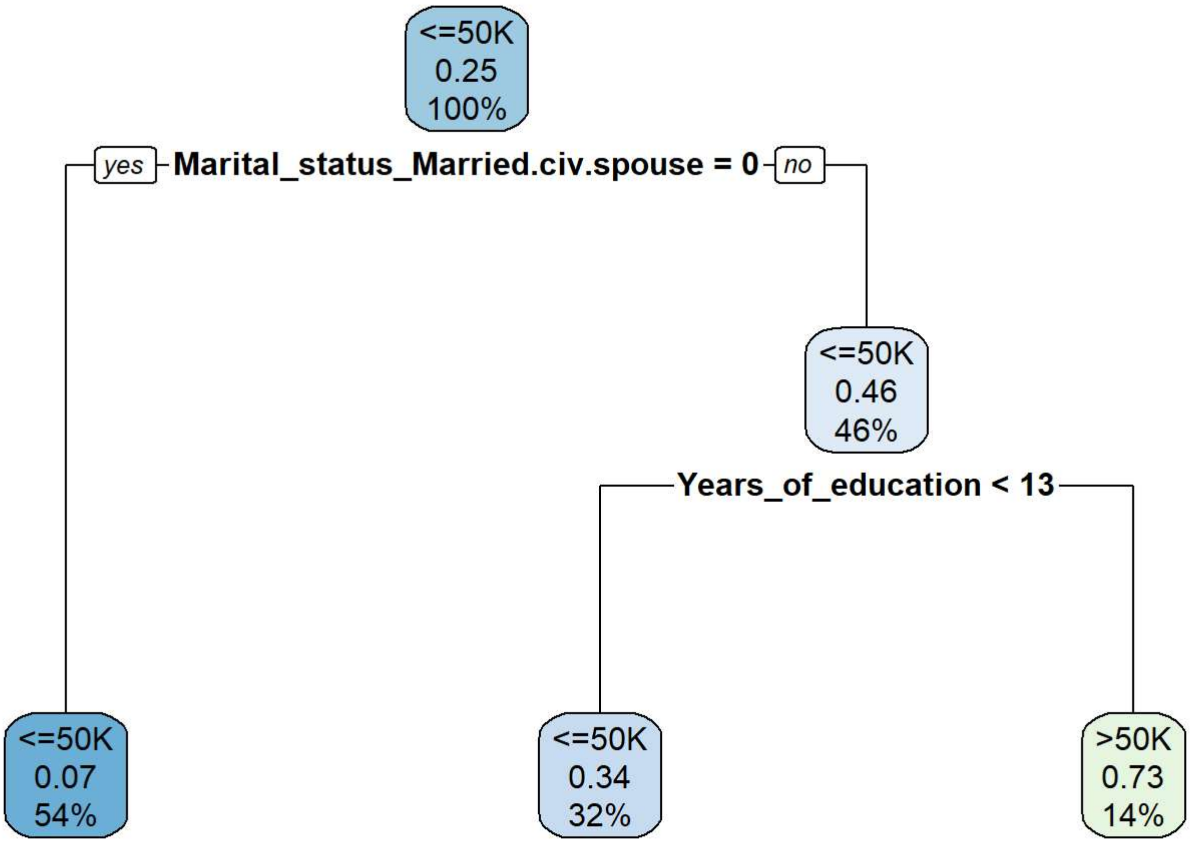



```
#adjustment in cp
#Overfitting
dtm_2 <- rpart(formula, train, method="class",control=rpart.control(cp=0,minsplit =0.01, minb
ucket = 5, maxdepth = 10 ))
rpart.plot(dtm_2)
```

```
## Warning: labs do not fit even at cex 0.15, there may be some overplotting
```



```
#Underfitting
dtm_3 <- rpart(formula, train, method="class",control=rpart.control(cp=0.1,minsplit =0.1, min
bucket = 5, maxdepth = 10 ))
rpart.plot(dtm_3)
```




```
#predicting a result and printing MSE
yhat4.tree <- predict(dtm, train2)
print(mse.tree <- mean((yhat4.tree - y_train) ^ 2))
```

```
## [1] 0.3806214
```

```
#Initialize instance of random forest
#Base Model
randomForest(formula, data = train, do.trace=T, mtry=10, keep.inbag=TRUE)
```

```
## ntree      OOB      1      2
##     1:  18.33% 12.07% 37.17%
##     2:  17.89% 10.99% 38.74%
##     3:  17.75% 10.57% 39.58%
##     4:  17.58%  9.84% 40.97%
##     5:  17.25%  9.52% 40.65%
##     6:  17.16%  9.51% 40.37%
##     7:  16.65%  9.02% 39.79%
##     8:  16.35%  8.76% 39.33%
##     9:  16.26%  8.62% 39.31%
##    10:  16.22%  8.55% 39.37%
##    11:  16.20%  8.53% 39.37%
##    12:  15.90%  8.23% 39.04%
##    13:  15.76%  8.01% 39.13%
##    14:  15.49%  7.78% 38.75%
##    15:  15.36%  7.75% 38.35%
##    16:  15.20%  7.57% 38.27%
##    17:  15.06%  7.53% 37.79%
##    18:  15.14%  7.48% 38.23%
##    19:  15.08%  7.48% 37.99%
##    20:  14.94%  7.33% 37.92%
##    21:  14.80%  7.15% 37.87%
##    22:  14.70%  7.06% 37.76%
##    23:  14.75%  7.02% 38.07%
##    24:  14.77%  6.91% 38.49%
##    25:  14.70%  6.91% 38.19%
##    26:  14.57%  6.79% 38.04%
##    27:  14.61%  6.86% 37.99%
##    28:  14.68%  6.87% 38.24%
##    29:  14.67%  6.89% 38.16%
##    30:  14.61%  6.78% 38.26%
##    31:  14.67%  6.83% 38.32%
##    32:  14.60%  6.79% 38.17%
##    33:  14.56%  6.76% 38.09%
##    34:  14.51%  6.70% 38.07%
##    35:  14.55%  6.74% 38.11%
##    36:  14.54%  6.70% 38.17%
##    37:  14.60%  6.75% 38.29%
##    38:  14.67%  6.80% 38.41%
##    39:  14.70%  6.89% 38.26%
##    40:  14.57%  6.80% 38.04%
##    41:  14.67%  6.80% 38.41%
##    42:  14.59%  6.79% 38.12%
##    43:  14.54%  6.70% 38.21%
##    44:  14.44%  6.64% 37.96%
##    45:  14.40%  6.67% 37.71%
##    46:  14.50%  6.72% 37.96%
##    47:  14.40%  6.64% 37.84%
##    48:  14.39%  6.57% 37.99%
##    49:  14.41%  6.62% 37.92%
##    50:  14.45%  6.64% 38.04%
##    51:  14.45%  6.66% 37.97%
##    52:  14.34%  6.54% 37.87%
##    53:  14.36%  6.62% 37.72%
##    54:  14.40%  6.71% 37.62%
##    55:  14.46%  6.75% 37.72%
##    56:  14.42%  6.65% 37.87%
```

```
##    57:  14.39%  6.63% 37.81%
##    58:  14.41%  6.65% 37.82%
##    59:  14.39%  6.67% 37.67%
##    60:  14.36%  6.62% 37.72%
##    61:  14.33%  6.58% 37.71%
##    62:  14.38%  6.66% 37.69%
##    63:  14.34%  6.59% 37.71%
##    64:  14.34%  6.64% 37.59%
##    65:  14.33%  6.60% 37.67%
##    66:  14.29%  6.51% 37.79%
##    67:  14.19%  6.42% 37.61%
##    68:  14.22%  6.44% 37.69%
##    69:  14.18%  6.50% 37.36%
##    70:  14.22%  6.51% 37.51%
##    71:  14.24%  6.51% 37.57%
##    72:  14.28%  6.53% 37.69%
##    73:  14.28%  6.51% 37.72%
##    74:  14.36%  6.57% 37.86%
##    75:  14.29%  6.57% 37.61%
##    76:  14.28%  6.55% 37.59%
##    77:  14.27%  6.48% 37.79%
##    78:  14.32%  6.54% 37.77%
##    79:  14.35%  6.58% 37.81%
##    80:  14.33%  6.61% 37.62%
##    81:  14.33%  6.55% 37.77%
##    82:  14.30%  6.53% 37.74%
##    83:  14.26%  6.47% 37.77%
##    84:  14.31%  6.54% 37.77%
##    85:  14.31%  6.54% 37.72%
##    86:  14.33%  6.59% 37.69%
##    87:  14.32%  6.56% 37.72%
##    88:  14.30%  6.55% 37.67%
##    89:  14.24%  6.50% 37.61%
##    90:  14.27%  6.53% 37.62%
##    91:  14.21%  6.46% 37.57%
##    92:  14.28%  6.58% 37.52%
##    93:  14.28%  6.60% 37.46%
##    94:  14.29%  6.59% 37.52%
##    95:  14.28%  6.56% 37.57%
##    96:  14.23%  6.55% 37.41%
##    97:  14.25%  6.57% 37.42%
##    98:  14.26%  6.58% 37.41%
##    99:  14.21%  6.51% 37.47%
##   100:  14.19%  6.51% 37.36%
##   101:  14.20%  6.48% 37.47%
##   102:  14.21%  6.52% 37.42%
##   103:  14.17%  6.49% 37.36%
##   104:  14.17%  6.49% 37.36%
##   105:  14.18%  6.54% 37.22%
##   106:  14.17%  6.54% 37.19%
##   107:  14.20%  6.58% 37.17%
##   108:  14.16%  6.57% 37.09%
##   109:  14.18%  6.59% 37.06%
##   110:  14.15%  6.59% 36.96%
##   111:  14.21%  6.62% 37.11%
##   112:  14.17%  6.62% 36.97%
##   113:  14.19%  6.59% 37.12%
##   114:  14.19%  6.60% 37.06%
```

```
##    115:  14.24%  6.65% 37.14%
##    116:  14.22%  6.62% 37.14%
##    117:  14.19%  6.59% 37.09%
##    118:  14.21%  6.63% 37.07%
##    119:  14.22%  6.66% 37.04%
##    120:  14.21%  6.63% 37.07%
##    121:  14.19%  6.62% 37.02%
##    122:  14.15%  6.59% 36.96%
##    123:  14.11%  6.59% 36.81%
##    124:  14.17%  6.63% 36.91%
##    125:  14.17%  6.63% 36.91%
##    126:  14.14%  6.57% 36.99%
##    127:  14.13%  6.55% 36.99%
##    128:  14.14%  6.55% 37.04%
##    129:  14.15%  6.57% 37.01%
##    130:  14.12%  6.55% 36.97%
##    131:  14.13%  6.54% 37.02%
##    132:  14.18%  6.57% 37.12%
##    133:  14.14%  6.55% 37.04%
##    134:  14.09%  6.51% 36.96%
##    135:  14.14%  6.55% 37.04%
##    136:  14.14%  6.55% 37.06%
##    137:  14.20%  6.57% 37.21%
##    138:  14.15%  6.53% 37.12%
##    139:  14.15%  6.53% 37.16%
##    140:  14.16%  6.53% 37.19%
##    141:  14.14%  6.52% 37.14%
##    142:  14.19%  6.53% 37.27%
##    143:  14.14%  6.49% 37.19%
##    144:  14.11%  6.46% 37.21%
##    145:  14.11%  6.50% 37.09%
##    146:  14.12%  6.48% 37.17%
##    147:  14.09%  6.46% 37.09%
##    148:  14.10%  6.46% 37.17%
##    149:  14.07%  6.46% 37.04%
##    150:  14.05%  6.42% 37.07%
##    151:  14.10%  6.49% 37.04%
##    152:  14.04%  6.43% 36.99%
##    153:  14.06%  6.48% 36.94%
##    154:  14.05%  6.47% 36.92%
##    155:  14.03%  6.47% 36.84%
##    156:  14.04%  6.49% 36.81%
##    157:  14.04%  6.48% 36.84%
##    158:  14.00%  6.46% 36.76%
##    159:  14.03%  6.49% 36.77%
##    160:  14.02%  6.44% 36.86%
##    161:  14.05%  6.48% 36.89%
##    162:  13.99%  6.44% 36.79%
##    163:  14.03%  6.47% 36.84%
##    164:  13.99%  6.43% 36.79%
##    165:  14.02%  6.44% 36.87%
##    166:  14.04%  6.46% 36.91%
##    167:  14.04%  6.47% 36.87%
##    168:  14.02%  6.44% 36.89%
##    169:  14.07%  6.48% 36.96%
##    170:  14.06%  6.46% 36.97%
##    171:  14.03%  6.46% 36.87%
##    172:  14.06%  6.49% 36.87%
```

```
##    173:  14.05%  6.48% 36.91%
##    174:  14.05%  6.48% 36.91%
##    175:  14.05%  6.46% 36.96%
##    176:  14.07%  6.49% 36.94%
##    177:  14.04%  6.47% 36.89%
##    178:  14.04%  6.46% 36.89%
##    179:  14.03%  6.46% 36.89%
##    180:  14.03%  6.49% 36.77%
##    181:  14.05%  6.50% 36.84%
##    182:  14.07%  6.47% 37.01%
##    183:  14.05%  6.47% 36.92%
##    184:  14.05%  6.46% 36.97%
##    185:  14.04%  6.42% 37.01%
##    186:  14.05%  6.47% 36.94%
##    187:  14.06%  6.47% 36.96%
##    188:  14.07%  6.46% 37.02%
##    189:  14.04%  6.43% 37.01%
##    190:  14.03%  6.43% 36.97%
##    191:  14.06%  6.46% 37.01%
##    192:  14.05%  6.44% 37.01%
##    193:  14.05%  6.45% 36.97%
##    194:  14.03%  6.47% 36.84%
##    195:  14.06%  6.48% 36.91%
##    196:  14.04%  6.49% 36.82%
##    197:  14.05%  6.50% 36.84%
##    198:  14.04%  6.50% 36.79%
##    199:  14.02%  6.51% 36.67%
##    200:  14.01%  6.49% 36.71%
##    201:  14.03%  6.52% 36.71%
##    202:  14.02%  6.49% 36.74%
##    203:  14.04%  6.51% 36.79%
##    204:  14.05%  6.50% 36.84%
##    205:  14.07%  6.53% 36.82%
##    206:  14.07%  6.52% 36.87%
##    207:  14.06%  6.54% 36.77%
##    208:  14.05%  6.53% 36.72%
##    209:  14.02%  6.52% 36.67%
##    210:  14.03%  6.49% 36.77%
##    211:  14.05%  6.49% 36.86%
##    212:  14.02%  6.49% 36.76%
##    213:  14.06%  6.51% 36.82%
##    214:  14.06%  6.51% 36.84%
##    215:  14.04%  6.52% 36.74%
##    216:  14.06%  6.53% 36.77%
##    217:  14.09%  6.57% 36.77%
##    218:  14.07%  6.55% 36.76%
##    219:  14.09%  6.56% 36.81%
##    220:  14.06%  6.55% 36.74%
##    221:  14.08%  6.55% 36.79%
##    222:  14.06%  6.54% 36.74%
##    223:  14.06%  6.54% 36.72%
##    224:  14.04%  6.54% 36.66%
##    225:  14.03%  6.53% 36.64%
##    226:  14.02%  6.52% 36.66%
##    227:  14.06%  6.54% 36.74%
##    228:  14.06%  6.52% 36.81%
##    229:  14.06%  6.53% 36.76%
##    230:  14.04%  6.51% 36.76%
```

```
##    231:  14.02%  6.48% 36.76%
##    232:  14.04%  6.52% 36.71%
##    233:  14.04%  6.51% 36.77%
##    234:  14.03%  6.50% 36.76%
##    235:  14.04%  6.50% 36.77%
##    236:  14.00%  6.52% 36.57%
##    237:  14.06%  6.55% 36.72%
##    238:  14.03%  6.51% 36.72%
##    239:  14.02%  6.48% 36.76%
##    240:  13.98%  6.47% 36.64%
##    241:  13.99%  6.47% 36.66%
##    242:  14.00%  6.47% 36.72%
##    243:  13.98%  6.48% 36.59%
##    244:  14.02%  6.48% 36.77%
##    245:  14.00%  6.46% 36.76%
##    246:  14.02%  6.47% 36.79%
##    247:  13.99%  6.46% 36.71%
##    248:  13.98%  6.46% 36.69%
##    249:  13.99%  6.46% 36.69%
##    250:  13.99%  6.46% 36.69%
##    251:  13.97%  6.46% 36.62%
##    252:  13.99%  6.47% 36.67%
##    253:  13.97%  6.47% 36.62%
##    254:  13.98%  6.48% 36.59%
##    255:  13.98%  6.47% 36.64%
##    256:  13.95%  6.44% 36.61%
##    257:  13.99%  6.48% 36.66%
##    258:  13.97%  6.46% 36.66%
##    259:  13.98%  6.47% 36.64%
##    260:  14.00%  6.49% 36.66%
##    261:  13.99%  6.47% 36.69%
##    262:  13.97%  6.47% 36.59%
##    263:  13.96%  6.47% 36.56%
##    264:  13.97%  6.46% 36.62%
##    265:  13.97%  6.47% 36.59%
##    266:  13.96%  6.48% 36.54%
##    267:  13.97%  6.46% 36.64%
##    268:  13.97%  6.47% 36.61%
##    269:  13.97%  6.47% 36.59%
##    270:  13.98%  6.48% 36.61%
##    271:  13.97%  6.48% 36.56%
##    272:  13.98%  6.47% 36.66%
##    273:  13.99%  6.50% 36.61%
##    274:  13.98%  6.48% 36.61%
##    275:  13.99%  6.47% 36.69%
##    276:  14.00%  6.48% 36.71%
##    277:  13.98%  6.44% 36.71%
##    278:  14.02%  6.46% 36.81%
##    279:  14.00%  6.46% 36.76%
##    280:  13.97%  6.43% 36.71%
##    281:  13.99%  6.46% 36.72%
##    282:  13.99%  6.45% 36.76%
##    283:  14.00%  6.48% 36.71%
##    284:  13.99%  6.46% 36.74%
##    285:  13.97%  6.46% 36.62%
##    286:  13.98%  6.46% 36.69%
##    287:  14.00%  6.47% 36.72%
##    288:  14.00%  6.46% 36.76%
```

```
##    289:  13.99%  6.46% 36.72%
##    290:  13.98%  6.44% 36.71%
##    291:  14.02%  6.46% 36.82%
##    292:  14.02%  6.48% 36.79%
##    293:  14.02%  6.47% 36.79%
##    294:  14.02%  6.47% 36.79%
##    295:  14.04%  6.47% 36.86%
##    296:  14.02%  6.47% 36.77%
##    297:  14.02%  6.47% 36.82%
##    298:  14.05%  6.49% 36.84%
##    299:  14.04%  6.50% 36.81%
##    300:  14.04%  6.49% 36.82%
##    301:  14.04%  6.52% 36.76%
##    302:  14.06%  6.53% 36.76%
##    303:  14.08%  6.53% 36.86%
##    304:  14.08%  6.53% 36.86%
##    305:  14.07%  6.55% 36.77%
##    306:  14.06%  6.51% 36.84%
##    307:  14.05%  6.52% 36.79%
##    308:  14.06%  6.52% 36.79%
##    309:  14.08%  6.54% 36.81%
##    310:  14.05%  6.52% 36.77%
##    311:  14.04%  6.51% 36.76%
##    312:  14.04%  6.51% 36.79%
##    313:  14.04%  6.49% 36.79%
##    314:  14.05%  6.53% 36.74%
##    315:  14.04%  6.53% 36.72%
##    316:  14.02%  6.52% 36.67%
##    317:  14.05%  6.52% 36.79%
##    318:  14.06%  6.54% 36.72%
##    319:  14.02%  6.49% 36.76%
##    320:  14.04%  6.51% 36.77%
##    321:  14.03%  6.51% 36.72%
##    322:  14.00%  6.46% 36.77%
##    323:  14.02%  6.49% 36.72%
##    324:  14.02%  6.48% 36.76%
##    325:  14.02%  6.49% 36.76%
##    326:  14.01%  6.46% 36.77%
##    327:  14.03%  6.49% 36.79%
##    328:  14.06%  6.51% 36.84%
##    329:  14.06%  6.49% 36.89%
##    330:  14.04%  6.49% 36.82%
##    331:  14.06%  6.50% 36.86%
##    332:  14.05%  6.49% 36.84%
##    333:  14.09%  6.52% 36.92%
##    334:  14.03%  6.49% 36.79%
##    335:  14.04%  6.50% 36.79%
##    336:  14.06%  6.49% 36.91%
##    337:  14.06%  6.51% 36.86%
##    338:  14.04%  6.49% 36.81%
##    339:  14.03%  6.49% 36.77%
##    340:  14.06%  6.52% 36.79%
##    341:  14.04%  6.51% 36.77%
##    342:  14.04%  6.52% 36.71%
##    343:  14.02%  6.49% 36.72%
##    344:  14.01%  6.47% 36.76%
##    345:  14.02%  6.49% 36.72%
##    346:  14.02%  6.48% 36.77%
```

```
##    347:  14.08%  6.55% 36.79%
##    348:  14.01%  6.47% 36.76%
##    349:  14.02%  6.48% 36.77%
##    350:  14.01%  6.47% 36.74%
##    351:  14.02%  6.48% 36.76%
##    352:  14.00%  6.45% 36.79%
##    353:  14.01%  6.48% 36.72%
##    354:  14.02%  6.49% 36.74%
##    355:  14.01%  6.48% 36.71%
##    356:  14.04%  6.48% 36.82%
##    357:  13.99%  6.46% 36.74%
##    358:  14.01%  6.47% 36.76%
##    359:  14.04%  6.51% 36.74%
##    360:  14.07%  6.54% 36.77%
##    361:  14.06%  6.53% 36.81%
##    362:  14.06%  6.54% 36.76%
##    363:  14.06%  6.55% 36.71%
##    364:  14.04%  6.54% 36.69%
##    365:  14.04%  6.54% 36.69%
##    366:  14.04%  6.52% 36.72%
##    367:  14.07%  6.54% 36.79%
##    368:  14.02%  6.52% 36.66%
##    369:  14.04%  6.53% 36.69%
##    370:  14.03%  6.53% 36.67%
##    371:  14.02%  6.52% 36.66%
##    372:  14.02%  6.54% 36.57%
##    373:  14.03%  6.54% 36.64%
##    374:  14.02%  6.51% 36.69%
##    375:  14.02%  6.52% 36.62%
##    376:  14.05%  6.53% 36.74%
##    377:  14.06%  6.55% 36.74%
##    378:  14.04%  6.52% 36.72%
##    379:  14.06%  6.55% 36.72%
##    380:  14.07%  6.57% 36.71%
##    381:  14.07%  6.57% 36.72%
##    382:  14.07%  6.55% 36.74%
##    383:  14.06%  6.53% 36.81%
##    384:  14.06%  6.53% 36.77%
##    385:  14.09%  6.55% 36.86%
##    386:  14.06%  6.54% 36.76%
##    387:  14.06%  6.53% 36.79%
##    388:  14.07%  6.54% 36.79%
##    389:  14.06%  6.54% 36.74%
##    390:  14.07%  6.56% 36.74%
##    391:  14.06%  6.54% 36.74%
##    392:  14.06%  6.55% 36.71%
##    393:  14.04%  6.53% 36.67%
##    394:  14.06%  6.54% 36.74%
##    395:  14.08%  6.55% 36.77%
##    396:  14.08%  6.55% 36.81%
##    397:  14.08%  6.56% 36.76%
##    398:  14.06%  6.54% 36.77%
##    399:  14.10%  6.57% 36.81%
##    400:  14.09%  6.56% 36.82%
##    401:  14.05%  6.54% 36.72%
##    402:  14.05%  6.54% 36.69%
##    403:  14.07%  6.55% 36.77%
##    404:  14.04%  6.52% 36.72%
```

```
##   405:  14.03%  6.54% 36.62%
##   406:  14.04%  6.54% 36.66%
##   407:  14.03%  6.54% 36.62%
##   408:  14.04%  6.54% 36.66%
##   409:  14.02%  6.53% 36.62%
##   410:  14.01%  6.51% 36.66%
##   411:  14.00%  6.51% 36.59%
##   412:  14.02%  6.52% 36.64%
##   413:  14.03%  6.54% 36.62%
##   414:  14.05%  6.54% 36.71%
##   415:  14.05%  6.56% 36.66%
##   416:  14.04%  6.53% 36.71%
##   417:  14.01%  6.53% 36.56%
##   418:  14.04%  6.55% 36.61%
##   419:  14.02%  6.54% 36.61%
##   420:  14.04%  6.54% 36.67%
##   421:  14.05%  6.55% 36.67%
##   422:  14.02%  6.54% 36.56%
##   423:  14.04%  6.56% 36.62%
##   424:  14.04%  6.55% 36.66%
##   425:  14.02%  6.52% 36.62%
##   426:  14.00%  6.52% 36.59%
##   427:  14.00%  6.49% 36.64%
##   428:  14.00%  6.50% 36.62%
##   429:  14.01%  6.54% 36.56%
##   430:  14.01%  6.53% 36.57%
##   431:  14.00%  6.53% 36.52%
##   432:  13.96%  6.50% 36.46%
##   433:  13.97%  6.51% 36.49%
##   434:  13.97%  6.49% 36.51%
##   435:  13.94%  6.48% 36.42%
##   436:  13.97%  6.50% 36.49%
##   437:  13.97%  6.49% 36.52%
##   438:  13.94%  6.46% 36.51%
##   439:  13.96%  6.48% 36.51%
##   440:  13.97%  6.49% 36.56%
##   441:  13.98%  6.49% 36.56%
##   442:  13.95%  6.49% 36.47%
##   443:  13.96%  6.51% 36.44%
##   444:  13.97%  6.51% 36.47%
##   445:  13.97%  6.50% 36.51%
##   446:  13.97%  6.51% 36.46%
##   447:  13.95%  6.52% 36.39%
##   448:  13.95%  6.51% 36.41%
##   449:  13.97%  6.53% 36.44%
##   450:  13.92%  6.49% 36.36%
##   451:  13.93%  6.49% 36.37%
##   452:  13.92%  6.48% 36.36%
##   453:  13.93%  6.48% 36.39%
##   454:  13.90%  6.46% 36.36%
##   455:  13.93%  6.47% 36.42%
##   456:  13.92%  6.47% 36.37%
##   457:  13.92%  6.47% 36.37%
##   458:  13.93%  6.48% 36.42%
##   459:  13.94%  6.48% 36.44%
##   460:  13.94%  6.48% 36.46%
##   461:  13.92%  6.45% 36.46%
##   462:  13.92%  6.44% 36.47%
```

```
##   463:  13.92%  6.46% 36.42%
##   464:  13.91%  6.45% 36.41%
##   465:  13.92%  6.47% 36.41%
##   466:  13.92%  6.45% 36.46%
##   467:  13.90%  6.45% 36.39%
##   468:  13.91%  6.46% 36.39%
##   469:  13.92%  6.44% 36.46%
##   470:  13.90%  6.46% 36.36%
##   471:  13.92%  6.44% 36.47%
##   472:  13.93%  6.44% 36.51%
##   473:  13.93%  6.44% 36.52%
##   474:  13.94%  6.45% 36.52%
##   475:  13.96%  6.47% 36.56%
##   476:  13.94%  6.46% 36.51%
##   477:  13.92%  6.46% 36.46%
##   478:  13.95%  6.47% 36.51%
##   479:  13.92%  6.47% 36.42%
##   480:  13.92%  6.45% 36.47%
##   481:  13.93%  6.46% 36.47%
##   482:  13.94%  6.47% 36.47%
##   483:  13.93%  6.47% 36.44%
##   484:  13.93%  6.47% 36.42%
##   485:  13.95%  6.50% 36.44%
##   486:  13.92%  6.48% 36.39%
##   487:  13.94%  6.48% 36.44%
##   488:  13.91%  6.45% 36.41%
##   489:  13.90%  6.46% 36.37%
##   490:  13.93%  6.48% 36.39%
##   491:  13.92%  6.46% 36.42%
##   492:  13.92%  6.46% 36.44%
##   493:  13.90%  6.46% 36.36%
##   494:  13.92%  6.45% 36.46%
##   495:  13.92%  6.47% 36.42%
##   496:  13.92%  6.46% 36.42%
##   497:  13.90%  6.44% 36.41%
##   498:  13.92%  6.44% 36.51%
##   499:  13.92%  6.44% 36.47%
##   500:  13.91%  6.44% 36.44%
```

```
## 
## Call:
##  randomForest(formula = formula, data = train, do.trace = T, mtry = 10,      keep.inbag = 
TRUE) 
##                Type of random forest: classification
##                      Number of trees: 500
## No. of variables tried at each split: 10
## 
##         OOB estimate of  error rate: 13.91%
## Confusion matrix:
##       <=50K >50K class.error
## <=50K 16957 1167  0.06438976
## >50K   2189 3818  0.36440819
```

```
#Bagging Model, mtry = 57 because the last one is Income
randomForest(formula, data = train2, do.trace=T, mtry=57, keep.inbag=TRUE)
```

```
## Warning in randomForest.default(m, y, ...): The response has five or fewer
## unique values. Are you sure you want to do regression?
```

```
##      |      Out-of-bag   |
## Tree |      MSE  %Var(y) |
##    1 |   0.1773    94.81 |
##    2 |   0.1711    91.49 |
##    3 |   0.1621    86.68 |
##    4 |    0.155    82.89 |
##    5 |   0.1491    79.75 |
##    6 |   0.1437    76.87 |
##    7 |   0.1391    74.42 |
##    8 |   0.1349    72.17 |
##    9 |   0.1312    70.18 |
##   10 |    0.128    68.45 |
##   11 |   0.1256    67.15 |
##   12 |    0.123    65.79 |
##   13 |   0.1213    64.89 |
##   14 |   0.1196    63.95 |
##   15 |   0.1182    63.21 |
##   16 |    0.117    62.59 |
##   17 |   0.1161    62.11 |
##   18 |   0.1151    61.57 |
##   19 |   0.1142    61.10 |
##   20 |   0.1137    60.83 |
##   21 |   0.1128    60.35 |
##   22 |   0.1123    60.07 |
##   23 |   0.1118    59.79 |
##   24 |   0.1112    59.45 |
##   25 |   0.1107    59.23 |
##   26 |   0.1103    58.98 |
##   27 |   0.1099    58.80 |
##   28 |   0.1096    58.64 |
##   29 |   0.1093    58.47 |
##   30 |   0.1093    58.45 |
##   31 |    0.109    58.33 |
##   32 |   0.1087    58.16 |
##   33 |   0.1086    58.08 |
##   34 |   0.1084    57.95 |
##   35 |   0.1082    57.88 |
##   36 |    0.108    57.77 |
##   37 |   0.1079    57.71 |
##   38 |   0.1078    57.64 |
##   39 |   0.1077    57.58 |
##   40 |   0.1075    57.48 |
##   41 |   0.1074    57.45 |
##   42 |   0.1072    57.35 |
##   43 |   0.1072    57.33 |
##   44 |    0.107    57.24 |
##   45 |   0.1069    57.20 |
##   46 |   0.1069    57.15 |
##   47 |   0.1068    57.11 |
##   48 |   0.1067    57.09 |
##   49 |   0.1066    57.01 |
##   50 |   0.1064    56.93 |
##   51 |   0.1063    56.84 |
##   52 |   0.1062    56.79 |
##   53 |    0.106    56.71 |
##   54 |   0.1059    56.64 |
##   55 |   0.1057    56.56 |
```

```
##   56 |   0.1057    56.54 |
##   57 |   0.1056    56.48 |
##   58 |   0.1054    56.40 |
##   59 |   0.1053    56.34 |
##   60 |   0.1053    56.31 |
##   61 |   0.1052    56.27 |
##   62 |   0.1051    56.23 |
##   63 |   0.1051    56.20 |
##   64 |   0.1051    56.19 |
##   65 |    0.105    56.18 |
##   66 |    0.105    56.15 |
##   67 |    0.105    56.16 |
##   68 |    0.105    56.16 |
##   69 |   0.1049    56.13 |
##   70 |   0.1049    56.11 |
##   71 |   0.1049    56.11 |
##   72 |   0.1048    56.03 |
##   73 |   0.1047    56.00 |
##   74 |   0.1047    55.98 |
##   75 |   0.1047    55.98 |
##   76 |   0.1046    55.95 |
##   77 |   0.1047    55.98 |
##   78 |   0.1046    55.96 |
##   79 |   0.1046    55.95 |
##   80 |   0.1046    55.95 |
##   81 |   0.1046    55.95 |
##   82 |   0.1046    55.92 |
##   83 |   0.1045    55.91 |
##   84 |   0.1045    55.92 |
##   85 |   0.1045    55.90 |
##   86 |   0.1045    55.88 |
##   87 |   0.1044    55.85 |
##   88 |   0.1044    55.86 |
##   89 |   0.1044    55.83 |
##   90 |   0.1044    55.82 |
##   91 |   0.1044    55.83 |
##   92 |   0.1043    55.80 |
##   93 |   0.1043    55.78 |
##   94 |   0.1042    55.75 |
##   95 |   0.1042    55.73 |
##   96 |   0.1042    55.72 |
##   97 |   0.1041    55.69 |
##   98 |   0.1041    55.69 |
##   99 |   0.1041    55.69 |
##  100 |   0.1041    55.67 |
##  101 |   0.1041    55.66 |
##  102 |   0.1041    55.67 |
##  103 |   0.1041    55.67 |
##  104 |   0.1041    55.65 |
##  105 |    0.104    55.61 |
##  106 |   0.1039    55.59 |
##  107 |   0.1039    55.58 |
##  108 |   0.1039    55.59 |
##  109 |   0.1039    55.57 |
##  110 |   0.1039    55.55 |
##  111 |   0.1039    55.55 |
##  112 |   0.1038    55.53 |
##  113 |   0.1038    55.52 |
```

```
##  114 |   0.1038    55.53 |
##  115 |   0.1039    55.55 |
##  116 |   0.1039    55.56 |
##  117 |   0.1039    55.55 |
##  118 |   0.1038    55.54 |
##  119 |   0.1039    55.57 |
##  120 |   0.1039    55.55 |
##  121 |   0.1038    55.52 |
##  122 |   0.1038    55.52 |
##  123 |   0.1038    55.52 |
##  124 |   0.1038    55.50 |
##  125 |   0.1037    55.49 |
##  126 |   0.1037    55.47 |
##  127 |   0.1037    55.45 |
##  128 |   0.1037    55.44 |
##  129 |   0.1036    55.43 |
##  130 |   0.1036    55.43 |
##  131 |   0.1036    55.41 |
##  132 |   0.1036    55.39 |
##  133 |   0.1035    55.38 |
##  134 |   0.1035    55.38 |
##  135 |   0.1035    55.37 |
##  136 |   0.1035    55.37 |
##  137 |   0.1035    55.35 |
##  138 |   0.1035    55.33 |
##  139 |   0.1035    55.34 |
##  140 |   0.1034    55.33 |
##  141 |   0.1035    55.34 |
##  142 |   0.1034    55.31 |
##  143 |   0.1034    55.30 |
##  144 |   0.1034    55.29 |
##  145 |   0.1034    55.30 |
##  146 |   0.1034    55.30 |
##  147 |   0.1034    55.29 |
##  148 |   0.1034    55.29 |
##  149 |   0.1034    55.30 |
##  150 |   0.1034    55.30 |
##  151 |   0.1034    55.29 |
##  152 |   0.1034    55.30 |
##  153 |   0.1034    55.30 |
##  154 |   0.1034    55.29 |
##  155 |   0.1034    55.29 |
##  156 |   0.1034    55.29 |
##  157 |   0.1034    55.28 |
##  158 |   0.1033    55.27 |
##  159 |   0.1033    55.27 |
##  160 |   0.1033    55.26 |
##  161 |   0.1033    55.27 |
##  162 |   0.1033    55.25 |
##  163 |   0.1033    55.25 |
##  164 |   0.1033    55.24 |
##  165 |   0.1033    55.23 |
##  166 |   0.1033    55.23 |
##  167 |   0.1032    55.22 |
##  168 |   0.1032    55.22 |
##  169 |   0.1032    55.22 |
##  170 |   0.1033    55.23 |
##  171 |   0.1032    55.22 |
```

```
## 172 |   0.1032   55.22 |
## 173 |   0.1032   55.21 |
## 174 |   0.1032   55.20 |
## 175 |   0.1032   55.20 |
## 176 |   0.1032   55.20 |
## 177 |   0.1032   55.21 |
## 178 |   0.1032   55.21 |
## 179 |   0.1032   55.22 |
## 180 |   0.1033   55.23 |
## 181 |   0.1033   55.23 |
## 182 |   0.1033   55.24 |
## 183 |   0.1033   55.23 |
## 184 |   0.1033   55.23 |
## 185 |   0.1033   55.23 |
## 186 |   0.1032   55.22 |
## 187 |   0.1032   55.21 |
## 188 |   0.1032   55.20 |
## 189 |   0.1032   55.20 |
## 190 |   0.1032   55.21 |
## 191 |   0.1032   55.21 |
## 192 |   0.1032   55.21 |
## 193 |   0.1032   55.22 |
## 194 |   0.1032   55.22 |
## 195 |   0.1032   55.21 |
## 196 |   0.1032   55.21 |
## 197 |   0.1032   55.21 |
## 198 |   0.1032   55.20 |
## 199 |   0.1032   55.20 |
## 200 |   0.1032   55.21 |
## 201 |   0.1032   55.22 |
## 202 |   0.1032   55.21 |
## 203 |   0.1032   55.21 |
## 204 |   0.1032   55.20 |
## 205 |   0.1032   55.19 |
## 206 |   0.1032   55.18 |
## 207 |   0.1032   55.18 |
## 208 |   0.1032   55.18 |
## 209 |   0.1032   55.18 |
## 210 |   0.1032   55.17 |
## 211 |   0.1031   55.16 |
## 212 |   0.1031   55.16 |
## 213 |   0.1031   55.16 |
## 214 |   0.1031   55.16 |
## 215 |   0.1031   55.16 |
## 216 |   0.1032   55.17 |
## 217 |   0.1032   55.18 |
## 218 |   0.1032   55.18 |
## 219 |   0.1032   55.18 |
## 220 |   0.1032   55.19 |
## 221 |   0.1032   55.18 |
## 222 |   0.1031   55.17 |
## 223 |   0.1031   55.16 |
## 224 |   0.1031   55.16 |
## 225 |   0.1031   55.17 |
## 226 |   0.1031   55.15 |
## 227 |   0.1031   55.16 |
## 228 |   0.1032   55.17 |
## 229 |   0.1031   55.17 |
```

```
## 230 | 0.1032 55.18 |
## 231 | 0.1032 55.18 |
## 232 | 0.1032 55.18 |
## 233 | 0.1032 55.17 |
## 234 | 0.1032 55.17 |
## 235 | 0.1032 55.17 |
## 236 | 0.1031 55.17 |
## 237 | 0.1031 55.17 |
## 238 | 0.1031 55.15 |
## 239 | 0.1031 55.15 |
## 240 | 0.1031 55.16 |
## 241 | 0.1031 55.15 |
## 242 | 0.1031 55.16 |
## 243 | 0.1031 55.16 |
## 244 | 0.1031 55.15 |
## 245 | 0.1031 55.14 |
## 246 | 0.1031 55.14 |
## 247 | 0.1031 55.15 |
## 248 | 0.1031 55.15 |
## 249 | 0.1031 55.14 |
## 250 | 0.1031 55.13 |
## 251 | 0.1031 55.14 |
## 252 | 0.1031 55.15 |
## 253 | 0.1031 55.15 |
## 254 | 0.1031 55.15 |
## 255 | 0.1031 55.15 |
## 256 | 0.1031 55.14 |
## 257 | 0.1031 55.15 |
## 258 | 0.1031 55.16 |
## 259 | 0.1031 55.16 |
## 260 | 0.1031 55.17 |
## 261 | 0.1032 55.17 |
## 262 | 0.1032 55.17 |
## 263 | 0.1031 55.16 |
## 264 | 0.1031 55.16 |
## 265 | 0.1031 55.16 |
## 266 | 0.1031 55.16 |
## 267 | 0.1031 55.16 |
## 268 | 0.1031 55.15 |
## 269 | 0.1031 55.15 |
## 270 | 0.1031 55.15 |
## 271 | 0.1031 55.14 |
## 272 | 0.1031 55.15 |
## 273 | 0.1031 55.14 |
## 274 | 0.1031 55.15 |
## 275 | 0.1031 55.15 |
## 276 | 0.1031 55.16 |
## 277 | 0.1031 55.15 |
## 278 | 0.1031 55.16 |
## 279 | 0.1031 55.17 |
## 280 | 0.1031 55.17 |
## 281 | 0.1032 55.17 |
## 282 | 0.1032 55.17 |
## 283 | 0.1032 55.17 |
## 284 | 0.1031 55.17 |
## 285 | 0.1032 55.17 |
## 286 | 0.1032 55.18 |
## 287 | 0.1031 55.16 |
```

```
## 288 |   0.1032   55.17 |
## 289 |   0.1032   55.17 |
## 290 |   0.1031   55.16 |
## 291 |   0.1031   55.16 |
## 292 |   0.1031   55.16 |
## 293 |   0.1031   55.16 |
## 294 |   0.1031   55.15 |
## 295 |   0.1031   55.15 |
## 296 |   0.1031   55.16 |
## 297 |   0.1031   55.15 |
## 298 |   0.1031   55.15 |
## 299 |   0.1031   55.14 |
## 300 |   0.1031   55.14 |
## 301 |   0.1031   55.14 |
## 302 |   0.1031   55.15 |
## 303 |   0.1031   55.16 |
## 304 |   0.1031   55.15 |
## 305 |   0.1031   55.14 |
## 306 |   0.1031   55.13 |
## 307 |   0.1031   55.14 |
## 308 |   0.1031   55.14 |
## 309 |   0.1031   55.14 |
## 310 |   0.1031   55.14 |
## 311 |   0.1031   55.14 |
## 312 |   0.1031   55.14 |
## 313 |   0.1031   55.14 |
## 314 |   0.1031   55.14 |
## 315 |   0.1031   55.13 |
## 316 |   0.1031   55.14 |
## 317 |   0.1031   55.14 |
## 318 |   0.1031   55.13 |
## 319 |   0.1031   55.13 |
## 320 |   0.1031   55.13 |
## 321 |   0.1031   55.12 |
## 322 |   0.1031   55.13 |
## 323 |   0.1031   55.14 |
## 324 |   0.1031   55.14 |
## 325 |   0.1031   55.14 |
## 326 |   0.1031   55.14 |
## 327 |   0.1031   55.13 |
## 328 |   0.1031   55.14 |
## 329 |   0.1031   55.14 |
## 330 |   0.1031   55.13 |
## 331 |   0.1031   55.13 |
## 332 |   0.1031   55.13 |
## 333 |   0.1031   55.12 |
## 334 |    0.103   55.12 |
## 335 |   0.1031   55.12 |
## 336 |   0.1031   55.13 |
## 337 |   0.1031   55.13 |
## 338 |   0.1031   55.13 |
## 339 |   0.1031   55.13 |
## 340 |   0.1031   55.13 |
## 341 |   0.1031   55.13 |
## 342 |   0.1031   55.14 |
## 343 |   0.1031   55.14 |
## 344 |   0.1031   55.13 |
## 345 |   0.1031   55.13 |
```

```
## 346 |  0.1031   55.13 |
## 347 |  0.1031   55.12 |
## 348 |   0.103   55.11 |
## 349 |   0.103   55.12 |
## 350 |   0.103   55.11 |
## 351 |   0.103   55.11 |
## 352 |   0.103   55.11 |
## 353 |   0.103   55.11 |
## 354 |   0.103   55.10 |
## 355 |   0.103   55.10 |
## 356 |   0.103   55.10 |
## 357 |   0.103   55.09 |
## 358 |   0.103   55.09 |
## 359 |   0.103   55.09 |
## 360 |   0.103   55.09 |
## 361 |   0.103   55.08 |
## 362 |   0.103   55.08 |
## 363 |   0.103   55.08 |
## 364 |   0.103   55.08 |
## 365 |   0.103   55.07 |
## 366 |   0.103   55.07 |
## 367 |   0.103   55.07 |
## 368 |   0.103   55.07 |
## 369 |   0.103   55.07 |
## 370 |   0.103   55.08 |
## 371 |   0.103   55.08 |
## 372 |   0.103   55.07 |
## 373 |   0.103   55.07 |
## 374 |   0.103   55.07 |
## 375 |   0.103   55.07 |
## 376 |   0.103   55.07 |
## 377 |   0.103   55.07 |
## 378 |   0.103   55.07 |
## 379 |   0.103   55.08 |
## 380 |   0.103   55.08 |
## 381 |   0.103   55.08 |
## 382 |   0.103   55.07 |
## 383 |   0.103   55.07 |
## 384 |   0.103   55.07 |
## 385 |   0.103   55.08 |
## 386 |   0.103   55.08 |
## 387 |   0.103   55.07 |
## 388 |   0.103   55.08 |
## 389 |   0.103   55.07 |
## 390 |   0.103   55.07 |
## 391 |   0.103   55.07 |
## 392 |   0.103   55.08 |
## 393 |   0.103   55.08 |
## 394 |   0.103   55.09 |
## 395 |   0.103   55.09 |
## 396 |   0.103   55.09 |
## 397 |   0.103   55.09 |
## 398 |   0.103   55.08 |
## 399 |   0.103   55.08 |
## 400 |   0.103   55.08 |
## 401 |   0.103   55.09 |
## 402 |   0.103   55.08 |
## 403 |   0.103   55.08 |
```

```
## 404 |    0.103    55.07 |
## 405 |    0.103    55.07 |
## 406 |    0.103    55.07 |
## 407 |    0.103    55.07 |
## 408 |    0.103    55.07 |
## 409 |    0.103    55.08 |
## 410 |    0.103    55.08 |
## 411 |    0.103    55.08 |
## 412 |    0.103    55.08 |
## 413 |    0.103    55.08 |
## 414 |    0.103    55.08 |
## 415 |    0.103    55.08 |
## 416 |    0.103    55.07 |
## 417 |    0.103    55.07 |
## 418 |    0.103    55.07 |
## 419 |    0.103    55.07 |
## 420 |    0.103    55.07 |
## 421 |    0.103    55.07 |
## 422 |   0.1029    55.06 |
## 423 |    0.103    55.06 |
## 424 |   0.1029    55.06 |
## 425 |   0.1029    55.06 |
## 426 |   0.1029    55.06 |
## 427 |   0.1029    55.06 |
## 428 |   0.1029    55.06 |
## 429 |   0.1029    55.06 |
## 430 |   0.1029    55.06 |
## 431 |    0.103    55.07 |
## 432 |    0.103    55.07 |
## 433 |    0.103    55.06 |
## 434 |    0.103    55.07 |
## 435 |   0.1029    55.06 |
## 436 |   0.1029    55.06 |
## 437 |   0.1029    55.06 |
## 438 |   0.1029    55.06 |
## 439 |   0.1029    55.05 |
## 440 |   0.1029    55.06 |
## 441 |   0.1029    55.06 |
## 442 |   0.1029    55.06 |
## 443 |   0.1029    55.05 |
## 444 |   0.1029    55.05 |
## 445 |   0.1029    55.05 |
## 446 |   0.1029    55.05 |
## 447 |   0.1029    55.05 |
## 448 |   0.1029    55.05 |
## 449 |   0.1029    55.05 |
## 450 |   0.1029    55.05 |
## 451 |   0.1029    55.05 |
## 452 |   0.1029    55.05 |
## 453 |   0.1029    55.05 |
## 454 |   0.1029    55.05 |
## 455 |   0.1029    55.05 |
## 456 |   0.1029    55.05 |
## 457 |   0.1029    55.04 |
## 458 |   0.1029    55.04 |
## 459 |   0.1029    55.03 |
## 460 |   0.1029    55.02 |
## 461 |   0.1029    55.02 |
```

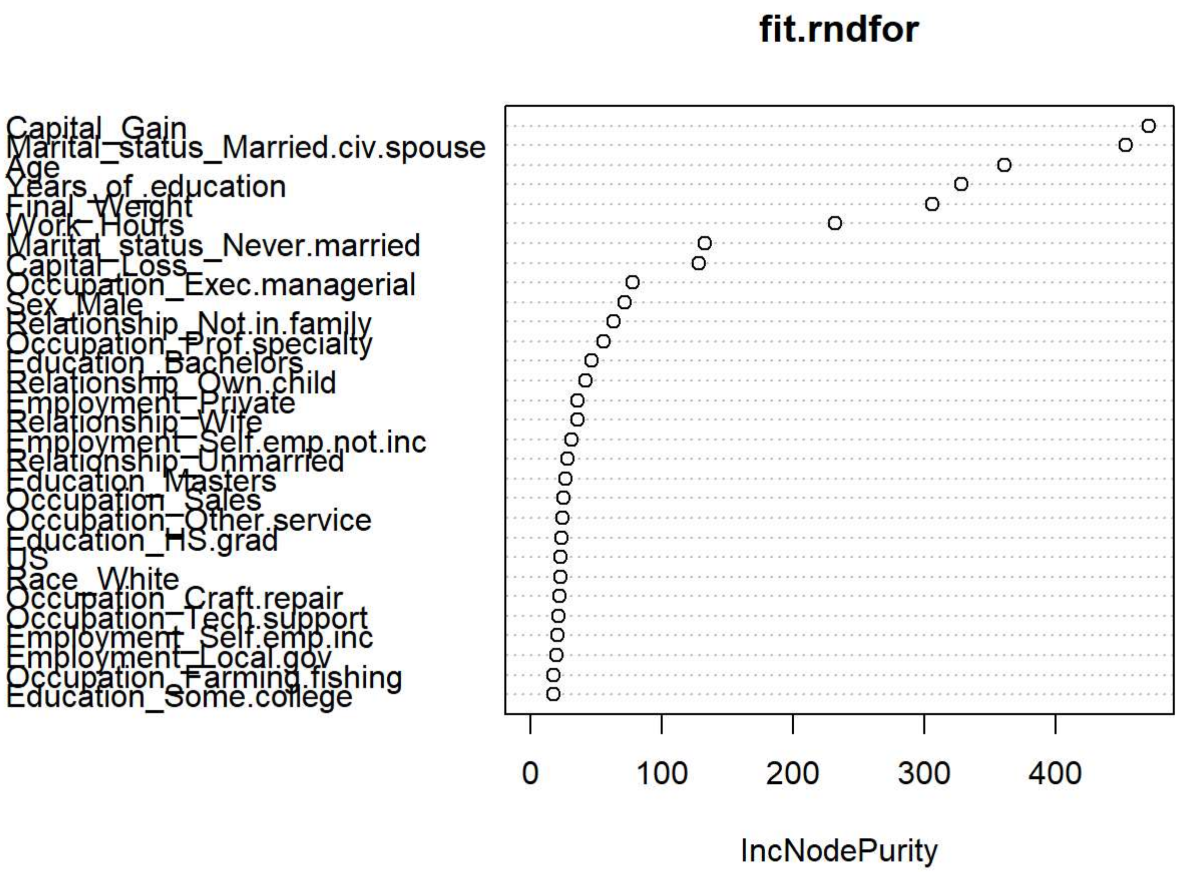
```
##  462 |   0.1029    55.02 |
##  463 |   0.1029    55.01 |
##  464 |   0.1029    55.01 |
##  465 |   0.1029    55.01 |
##  466 |   0.1028    55.01 |
##  467 |   0.1029    55.02 |
##  468 |   0.1029    55.02 |
##  469 |   0.1029    55.02 |
##  470 |   0.1029    55.02 |
##  471 |   0.1029    55.02 |
##  472 |   0.1029    55.02 |
##  473 |   0.1029    55.02 |
##  474 |   0.1028    55.01 |
##  475 |   0.1028    55.01 |
##  476 |   0.1028    55.01 |
##  477 |   0.1028    55.01 |
##  478 |   0.1028    55.01 |
##  479 |   0.1028    55.01 |
##  480 |   0.1029    55.01 |
##  481 |   0.1028    55.01 |
##  482 |   0.1028    55.01 |
##  483 |   0.1028    55.01 |
##  484 |   0.1028    55.00 |
##  485 |   0.1028    55.00 |
##  486 |   0.1028    55.00 |
##  487 |   0.1028    55.00 |
##  488 |   0.1028    55.00 |
##  489 |   0.1028    54.99 |
##  490 |   0.1028    54.99 |
##  491 |   0.1028    54.99 |
##  492 |   0.1028    54.99 |
##  493 |   0.1028    55.00 |
##  494 |   0.1028    55.00 |
##  495 |   0.1028    55.00 |
##  496 |   0.1028    55.00 |
##  497 |   0.1028    55.00 |
##  498 |   0.1028    54.99 |
##  499 |   0.1028    55.00 |
##  500 |   0.1028    54.99 |
```

```
## 
## Call:
##  randomForest(formula = formula, data = train2, do.trace = T,      mtry = 57, keep.inbag = TRUE) 
##                Type of random forest: regression
##                      Number of trees: 500
## No. of variables tried at each split: 57
## 
##           Mean of squared residuals: 0.1028199
##                     % Var explained: 45.01
```

```
#Fitting the model and plotting the varImpPlot
fit.rndfor <- randomForest(formula, data = train2, mtry=10, do.trace=F)
```

```
## Warning in randomForest.default(m, y, ...): The response has five or fewer
## unique values. Are you sure you want to do regression?
```

```
print(varImpPlot(fit.rndfor))
```



```
##                                       IncNodePurity
## Age                                    3.617307e+02
## Final_Weight                           3.066685e+02
## Years_of_education                     3.284605e+02
## Capital_Gain                           4.714479e+02
## Capital_Loss                           1.284757e+02
## Work_Hours                             2.321549e+02
## Employment_Local.gov                   2.014484e+01
## Employment_Private                     3.578570e+01
## Employment_Self.emp.inc                2.105498e+01
## Employment_Self.emp.not.inc            3.169223e+01
## Employment_State.gov                   1.635410e+01
## Employment_Without.pay                 1.920413e-01
## Education_11th                         4.895572e+00
## Education_12th                         2.465664e+00
## Education_1st.4th                      5.191785e-01
## Education_5th.6th                      1.419470e+00
## Education_7th.8th                      5.467591e+00
## Education_9th                          2.809688e+00
## Education_Assoc.acdm                   8.105621e+00
## Education_Assoc.voc                    9.961570e+00
## Education_Bachelors                    4.661084e+01
## Education_Doctorate                    1.088214e+01
## Education_HS.grad                      2.401448e+01
## Education_Masters                      2.656794e+01
## Education_Preschool                    2.523108e-01
## Education_Prof.school                  1.554456e+01
## Education_Some.college                 1.744635e+01
## Marital_status_Married.AF.spouse       2.044273e+00
## Marital_status_Married.civ.spouse      4.540250e+02
## Marital_status_Married.spouse.absent   3.168091e+00
## Marital_status_Never.married           1.327596e+02
## Marital_status_Separated               5.756307e+00
## Marital_status_Widowed                 5.052245e+00
## Occupation_Armed.Forces                6.818373e-04
## Occupation_Craft.repair                2.228452e+01
## Occupation_Exec.managerial             7.808279e+01
## Occupation_Farming.fishing             1.746278e+01
## Occupation_Handlers.cleaners           1.217100e+01
## Occupation_Machine.op.inspct           1.443226e+01
## Occupation_Other.service               2.438119e+01
## Occupation_Priv.house.serv             3.328196e-01
## Occupation_Prof.specialty              5.585062e+01
## Occupation_Protective.serv             1.294748e+01
## Occupation_Sales                       2.522474e+01
## Occupation_Tech.support                2.143473e+01
## Occupation_Transport.moving            1.696974e+01
## Relationship_Not.in.family             6.326625e+01
## Relationship_Other.relative            6.975224e+00
## Relationship_Own.child                 4.183853e+01
## Relationship_Unmarried                 2.859783e+01
## Relationship_Wife                      3.566675e+01
## Race_Asian.Pac.Islander                1.158793e+01
## Race_Black                             1.526600e+01
## Race_Other                             3.588875e+00
## Race_White                             2.266579e+01
```

```
## Sex_Male                                 7.155483e+01
## US                                       2.311997e+01
```

```
yhat.rndfor <- predict(fit.rndfor, test)
mse.tree <- mean((yhat.rndfor - y_test) ^ 2)
print(mse.tree)
```

```
## [1] 0.096711
```

```
#Boosting Model
fit.btree <- gbm(formula, data = train2,
                 distribution = "gaussian",
                 n.trees = 100,
                 interaction.depth = 2,
                 shrinkage = 0.001)
relative.influence(fit.btree)
```

```
## n.trees not given. Using 100 trees.
```

```
##                                Age                           Final_Weight
##                             0.0000                                 0.0000
##                 Years_of_education                           Capital_Gain
##                         15491.9513                               564.7292
##                       Capital_Loss                             Work_Hours
##                             0.0000                                 0.0000
##               Employment_Local.gov                     Employment_Private
##                             0.0000                                 0.0000
##             Employment_Self.emp.inc            Employment_Self.emp.not.inc
##                             0.0000                                 0.0000
##               Employment_State.gov                 Employment_Without.pay
##                             0.0000                                 0.0000
##                     Education_11th                         Education_12th
##                             0.0000                                 0.0000
##                  Education_1st.4th                      Education_5th.6th
##                             0.0000                                 0.0000
##                  Education_7th.8th                          Education_9th
##                             0.0000                                 0.0000
##               Education_Assoc.acdm                    Education_Assoc.voc
##                             0.0000                                 0.0000
##                Education_Bachelors                    Education_Doctorate
##                             0.0000                                 0.0000
##                  Education_HS.grad                      Education_Masters
##                             0.0000                                 0.0000
##                Education_Preschool                  Education_Prof.school
##                             0.0000                                 0.0000
##             Education_Some.college      Marital_status_Married.AF.spouse
##                             0.0000                                 0.0000
##    Marital_status_Married.civ.spouse Marital_status_Married.spouse.absent
##                         41124.3867                                 0.0000
##        Marital_status_Never.married               Marital_status_Separated
##                             0.0000                                 0.0000
##             Marital_status_Widowed                Occupation_Armed.Forces
##                             0.0000                                 0.0000
##             Occupation_Craft.repair             Occupation_Exec.managerial
##                             0.0000                                 0.0000
##          Occupation_Farming.fishing           Occupation_Handlers.cleaners
##                             0.0000                                 0.0000
##        Occupation_Machine.op.inspct               Occupation_Other.service
##                             0.0000                                 0.0000
##          Occupation_Priv.house.serv               Occupation_Prof.specialty
##                             0.0000                                 0.0000
##          Occupation_Protective.serv                       Occupation_Sales
##                             0.0000                                 0.0000
##             Occupation_Tech.support            Occupation_Transport.moving
##                             0.0000                                 0.0000
##          Relationship_Not.in.family            Relationship_Other.relative
##                             0.0000                                 0.0000
##             Relationship_Own.child                 Relationship_Unmarried
##                             0.0000                                 0.0000
##                  Relationship_Wife               Race_Asian.Pac.Islander
##                             0.0000                                 0.0000
##                         Race_Black                             Race_Other
##                             0.0000                                 0.0000
##                         Race_White                               Sex_Male
##                             0.0000                                 0.0000
```

```
##                                     US
##                                 0.0000
```

```
yhat.btree <- predict(fit.btree, train, n.trees = 100)
mse.btree <- mean((yhat.btree - y_train) ^ 2)
print(mse.btree)
```

```
## [1] 0.1775245
```