R Project Classification

4375 Machine Learning with Dr. Mazidi

Joshua Jan

3/27/2022

Data website link: https://www.kaggle.com/wenruliu/adult-income-dataset

Data cleaning comment: The data set consists of 48842 observations and 15 attributes. The columns are filled with different numbers, such as int and char. I used essential R functions to look at the variables and numbers while detecting if there are any NAs or empty data. I also discarded education, final weight, and relationship because they are irrelevant to what I want to know in the tables and graphs.

```
library(ggplot2)
library(plyr)
library(ROCR)
library(kernlab)
```

```
##
## Attaching package: 'kernlab'

## The following object is masked from 'package:ggplot2':
##
##     alpha
```

```
library(e1071)

adult <- read.csv("adult_income.csv")
str(adult)
```

```
## 'data.frame':    48842 obs. of  15 variables:
##  $ age            : int  25 38 28 44 18 34 29 63 24 55 ...
##  $ workclass      : chr  "Private" "Private" "Local-gov" "Private" ...
##  $ fnlwgt         : int  226802 89814 336951 160323 103497 198693 227026 104626 369667 104996 ...
##  $ education      : chr  "11th" "HS-grad" "Assoc-acdm" "Some-college" ...
##  $ educational.num: int  7 9 12 10 10 6 9 15 10 4 ...
##  $ marital.status : chr  "Never-married" "Married-civ-spouse" "Married-civ-spouse" "Married-civ-spous
##  $ occupation     : chr  "Machine-op-inspct" "Farming-fishing" "Protective-serv" "Machine-op-inspct"
##  $ relationship   : chr  "Own-child" "Husband" "Husband" "Husband" ...
##  $ race           : chr  "Black" "White" "White" "Black" ...
##  $ gender         : chr  "Male" "Male" "Male" "Male" ...
##  $ capital.gain   : int  0 0 0 7688 0 0 0 3103 0 0 ...
##  $ capital.loss   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ hours.per.week : int  40 50 40 40 30 30 40 32 40 10 ...
##  $ native.country : chr  "United-States" "United-States" "United-States" "United-States" ...
##  $ income         : chr  "<=50K" "<=50K" ">50K" ">50K" ...
```

```r
names(adult)
```

```
##  [1] "age"             "workclass"       "fnlwgt"          "education"      
##  [5] "educational.num" "marital.status"  "occupation"      "relationship"   
##  [9] "race"            "gender"          "capital.gain"    "capital.loss"   
## [13] "hours.per.week"  "native.country"  "income"
```

```r
summary(adult)
```

```
##       age         workclass             fnlwgt          education        
##  Min.   :17.00   Length:48842       Min.   :  12285   Length:48842      
##  1st Qu.:28.00   Class :character   1st Qu.: 117551   Class :character  
##  Median :37.00   Mode  :character   Median : 178145   Mode  :character  
##  Mean   :38.64                      Mean   : 189664                     
##  3rd Qu.:48.00                      3rd Qu.: 237642                     
##  Max.   :90.00                      Max.   :1490400                     
##  educational.num marital.status      occupation        relationship      
##  Min.   : 1.00   Length:48842       Length:48842       Length:48842      
##  1st Qu.: 9.00   Class :character   Class :character   Class :character  
##  Median :10.00   Mode  :character   Mode  :character   Mode  :character  
##  Mean   :10.08                                                           
##  3rd Qu.:12.00                                                           
##  Max.   :16.00                                                           
##      race              gender           capital.gain    capital.loss   
##  Length:48842       Length:48842       Min.   :    0   Min.   :   0.0  
##  Class :character   Class :character   1st Qu.:    0   1st Qu.:   0.0  
##  Mode  :character   Mode  :character   Median :    0   Median :   0.0  
##                                        Mean   : 1079   Mean   :  87.5  
##                                        3rd Qu.:    0   3rd Qu.:   0.0  
##                                        Max.   :99999   Max.   :4356.0  
##  hours.per.week  native.country        income         
##  Min.   : 1.00   Length:48842       Length:48842      
##  1st Qu.:40.00   Class :character   Class :character  
##  Median :40.00   Mode  :character   Mode  :character  
##  Mean   :40.42                                        
##  3rd Qu.:45.00                                        
##  Max.   :99.00
```

```r
head(adult)
```

```
##   age workclass fnlwgt    education educational.num     marital.status
## 1  25   Private 226802         11th               7      Never-married
## 2  38   Private  89814      HS-grad               9 Married-civ-spouse
## 3  28 Local-gov 336951   Assoc-acdm              12 Married-civ-spouse
## 4  44   Private 160323 Some-college              10 Married-civ-spouse
## 5  18         ? 103497 Some-college              10      Never-married
## 6  34   Private 198693         10th               6      Never-married
##          occupation  relationship  race gender capital.gain capital.loss
## 1 Machine-op-inspct     Own-child Black   Male            0            0
## 2   Farming-fishing       Husband White   Male            0            0
## 3   Protective-serv       Husband White   Male            0            0
## 4 Machine-op-inspct       Husband Black   Male         7688            0
```

```
## 5                 ?     Own-child White Female            0            0
## 6     Other-service Not-in-family White   Male            0            0
##   hours.per.week native.country income
## 1             40  United-States  <=50K
## 2             50  United-States  <=50K
## 3             40  United-States   >50K
## 4             40  United-States   >50K
## 5             30  United-States  <=50K
## 6             30  United-States  <=50K
```

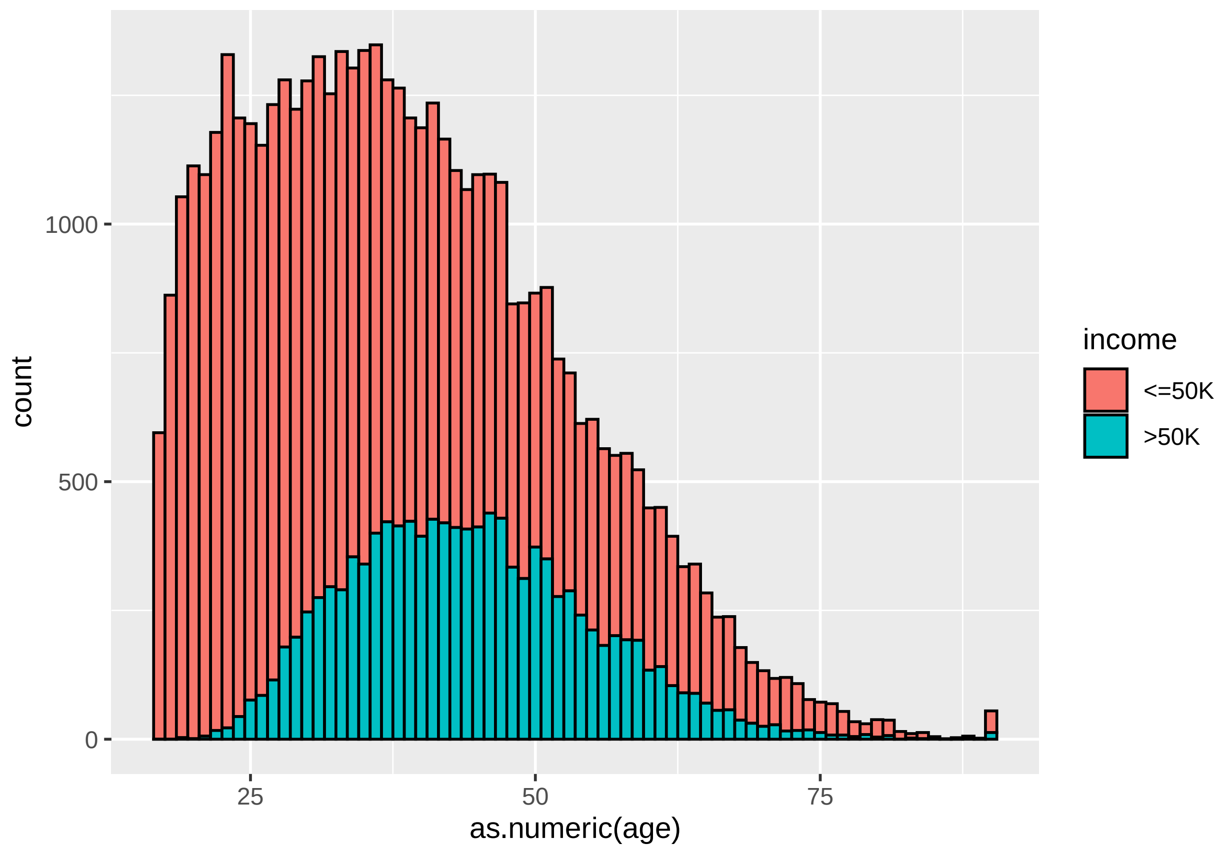
```
tail(adult)
```

```
##       age    workclass fnlwgt    education educational.num     marital.status
## 48837  22      Private 310152 Some-college              10      Never-married
## 48838  27      Private 257302   Assoc-acdm              12 Married-civ-spouse
## 48839  40      Private 154374      HS-grad               9 Married-civ-spouse
## 48840  58      Private 151910      HS-grad               9            Widowed
## 48841  22      Private 201490      HS-grad               9      Never-married
## 48842  52 Self-emp-inc 287927      HS-grad               9 Married-civ-spouse
##              occupation  relationship  race gender capital.gain capital.loss
## 48837   Protective-serv Not-in-family White   Male            0            0
## 48838      Tech-support          Wife White Female            0            0
## 48839 Machine-op-inspct       Husband White   Male            0            0
## 48840      Adm-clerical     Unmarried White Female            0            0
## 48841      Adm-clerical     Own-child White   Male            0            0
## 48842   Exec-managerial          Wife White Female        15024            0
##       hours.per.week native.country income
## 48837             40  United-States  <=50K
## 48838             38  United-States  <=50K
## 48839             40  United-States   >50K
## 48840             40  United-States  <=50K
## 48841             20  United-States  <=50K
## 48842             40  United-States   >50K
```

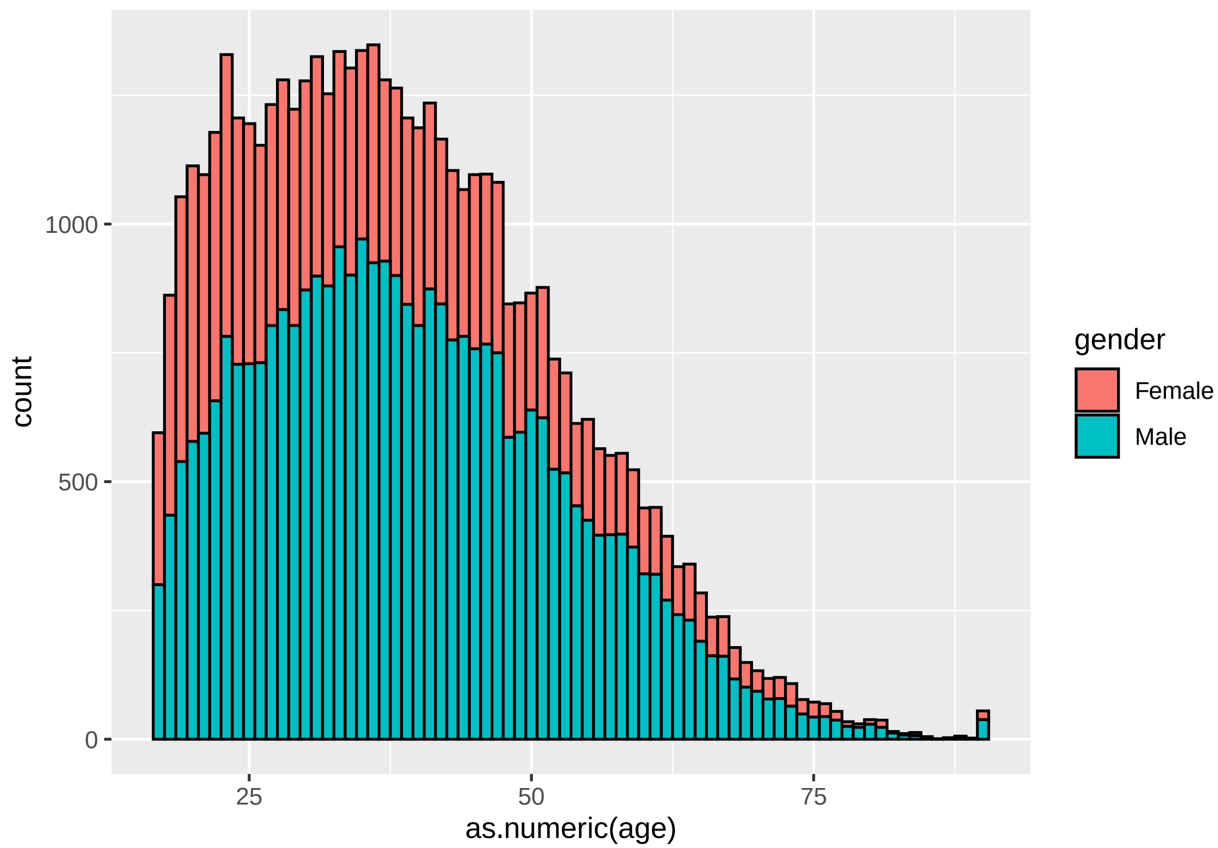
```
adult$educatoin <- NULL
adult$fnlwgt <- NULL
adult$relationship <- NULL
```

Graph comment: The majority of the people make less than 50K per year. Around 38 to 50 years old is when other people make over 50K. Statistically, males make up the majority of the gender group than females.

```
ggplot(adult) + aes(x = as.numeric(age), group = income, fill = income) +
  geom_histogram(binwidth=1, color='black')
```




```
ggplot(adult) + aes(x = as.numeric(age), group = gender, fill = gender) +
  geom_histogram(binwidth=1, color='black')
```



Data exploration comment: I categorized government jobs into one level, self-employed to another, and NAs to the last level. After that, I factor in the working class to display the table. According to the table results, there are always fewer people making more than 50K in every work category.

```
adult$workclass <- gsub('^Federal-gov', 'Government', adult$workclass)
adult$workclass <- gsub('^Local-gov', 'Government', adult$workclass)
adult$workclass <- gsub('^State-gov', 'Government', adult$workclass)

adult$workclass <- gsub('^Self-emp-inc', 'Self-Employed', adult$workclass)
adult$workclass <- gsub('^Self-emp-not-inc', 'Self-Employed', adult$workclass)

adult$workclass <- gsub('^Never-worked', 'Other', adult$workclass)
adult$workclass <- gsub('^Without-pay', 'Other', adult$workclass)
adult$workclass <- gsub('^\\?', 'Other', adult$workclass)
adult$workclass <- gsub('^Other', 'Other/Unknown', adult$workclass)

adult$workclass <- as.factor(adult$workclass)

count <- table(adult[adult$workclass == 'Government',]$income)["<=50K"]
count <- c(count, table(adult[adult$workclass == 'Government',]$income)[">50K"])
count <- c(count, table(adult[adult$workclass == 'Other/Unknown',]$income)["<=50K"])
count <- c(count, table(adult[adult$workclass == 'Other/Unknown',]$income)[">50K"])
count <- c(count, table(adult[adult$workclass == 'Private',]$income)["<=50K"])
count <- c(count, table(adult[adult$workclass == 'Private',]$income)[">50K"])
count <- c(count, table(adult[adult$workclass == 'Self-Employed',]$income)["<=50K"])
count <- c(count, table(adult[adult$workclass == 'Self-Employed',]$income)[">50K"])
```

```
count <- as.numeric(count)

industry <- rep(levels(adult$workclass), each = 2)
income <- rep(c('<=50K', '>50K'), 4)
df1 <- data.frame(industry, income, count)
df1
```

```
##        industry income count
## 1    Government  <=50K  4531
## 2    Government   >50K  2018
## 3 Other/Unknown  <=50K  2563
## 4 Other/Unknown   >50K   267
## 5       Private  <=50K 26519
## 6       Private   >50K  7387
## 7 Self-Employed  <=50K  3542
## 8 Self-Employed   >50K  2015
```

Data exploration comment: White-collar refers to people avoiding physical labor, whereas blue-collar is the opposite. Service jobs belong to the service category, and NAs belong to Other/Unknown. Blue and white-collar make up the majority which makes sense because it is a broad term to categorize specific jobs. There is a clear difference (3x) between professional occupations that are above 50K and service jobs that are above 50K.

```
adult$occupation <- gsub('Adm-clerical', 'White-Collar', adult$occupation)
adult$occupation <- gsub('Craft-repair', 'Blue-Collar', adult$occupation)
adult$occupation <- gsub('Exec-managerial', 'White-Collar', adult$occupation)
adult$occupation <- gsub('Farming-fishing', 'Blue-Collar', adult$occupation)
adult$occupation <- gsub('Handlers-cleaners', 'Blue-Collar', adult$occupation)
adult$occupation <- gsub('Machine-op-inspct', 'Blue-Collar', adult$occupation)
adult$occupation <- gsub('Other-service', 'Service', adult$occupation)
adult$occupation <- gsub('Priv-house-serv', 'Service', adult$occupation)
adult$occupation <- gsub('Prof-specialty', 'Professional', adult$occupation)
adult$occupation <- gsub('Protective-serv', 'Service', adult$occupation)
adult$occupation <- gsub('Tech-support', 'Service', adult$occupation)
adult$occupation <- gsub('Transport-moving', 'Blue-Collar', adult$occupation)
adult$occupation <- gsub('^\\?', 'Other/Unknown', adult$occupation)
adult$occupation <- gsub('Armed-Forces', 'Other/Unknown', adult$occupation)
adult$occupation <- as.factor(adult$occupation)

df2 <- data.frame(table(adult$income, adult$occupation))
names(df2) <- c('income', 'occupation', 'count')
df2
```

```
##   income    occupation count
## 1  <=50K   Blue-Collar 12504
## 2   >50K   Blue-Collar  2547
## 3  <=50K Other/Unknown  2554
## 4   >50K Other/Unknown   270
## 5  <=50K  Professional  3388
## 6   >50K  Professional  2784
## 7  <=50K         Sales  4029
## 8   >50K         Sales  1475
## 9  <=50K       Service  6659
```

```
## 10   >50K       Service   935
## 11  <=50K  White-Collar  8021
## 12   >50K  White-Collar  3676
```

Logistic regression comment: Income is used as a response variable for all three algorithms to compare with other predictors. Income is an excellent result based on different factors since money is not equally distributed to everyone. Logistic regression provided data for people making more than 50K per year. Response leaning towards 1 indicates a higher chance to make over 50K, whereas an answer of 0 indicates otherwise. As a result, the threshold is kept to 0.5 as an indicator. The model has an accuracy of 85%, which is not bad with all things considered.

```
adult$education <- as.factor(adult$education)
adult$marital.status <- as.factor(adult$marital.status)
adult$race <- as.factor(adult$race)
adult$gender <- as.factor(adult$gender)
adult$native.country <- as.factor(adult$native.country)
adult$income <- as.factor(adult$income)

size <- round(.8 * dim(adult)[1])
train <- adult[1:size,]
test <- adult[-(1:size),]

beg <- proc.time()
lm <- glm(income ~ ., data = train, family = binomial('logit'))
```

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

```
end <- proc.time()
summary(lm)
```

```
##
## Call:
## glm(formula = income ~ ., family = binomial("logit"), data = train)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -5.2374  -0.5168  -0.2208  -0.0615   3.3316
##
## Coefficients: (1 not defined because of singularities)
##                                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)                           -7.540e+00  2.921e-01 -25.811  < 2e-16
## age                                    2.416e-02  1.440e-03  16.771  < 2e-16
## workclassOther/Unknown                -1.150e+00  6.646e-01  -1.731  0.08345
## workclassPrivate                       3.365e-02  4.634e-02   0.726  0.46770
## workclassSelf-Employed                -2.943e-01  6.012e-02  -4.895 9.86e-07
## education11th                          4.094e-02  1.889e-01   0.217  0.82845
## education12th                          4.856e-01  2.393e-01   2.029  0.04242
## education1st-4th                      -6.219e-01  4.468e-01  -1.392  0.16394
## education5th-6th                      -2.494e-01  2.840e-01  -0.878  0.37976
## education7th-8th                      -4.176e-01  2.072e-01  -2.016  0.04379
## education9th                          -3.231e-01  2.394e-01  -1.349  0.17719
## educationAssoc-acdm                    1.570e+00  1.594e-01   9.849  < 2e-16
```

```
## educationAssoc-voc                     1.417e+00  1.539e-01   9.208  < 2e-16
## educationBachelors                     2.099e+00  1.430e-01  14.685  < 2e-16
## educationDoctorate                     2.922e+00  1.961e-01  14.902  < 2e-16
## educationHS-grad                       8.661e-01  1.396e-01   6.204 5.52e-10
## educationMasters                       2.437e+00  1.518e-01  16.055  < 2e-16
## educationPreschool                    -5.193e+00  4.873e+00  -1.066  0.28657
## educationProf-school                   2.875e+00  1.850e-01  15.542  < 2e-16
## educationSome-college                  1.253e+00  1.416e-01   8.854  < 2e-16
## educational.num                               NA         NA      NA       NA
## marital.statusMarried-AF-spouse        2.577e+00  4.526e-01   5.692 1.25e-08
## marital.statusMarried-civ-spouse       2.286e+00  6.188e-02  36.935  < 2e-16
## marital.statusMarried-spouse-absent    2.411e-01  1.892e-01   1.274  0.20255
## marital.statusNever-married           -4.314e-01  7.525e-02  -5.732 9.90e-09
## marital.statusSeparated                4.161e-02  1.437e-01   0.290  0.77209
## marital.statusWidowed                 -6.187e-02  1.399e-01  -0.442  0.65843
## occupationOther/Unknown                6.626e-01  6.639e-01   0.998  0.31822
## occupationProfessional                 7.167e-01  5.991e-02  11.962  < 2e-16
## occupationSales                        4.099e-01  5.542e-02   7.397 1.39e-13
## occupationService                      1.540e-01  5.798e-02   2.656  0.00791
## occupationWhite-Collar                 7.347e-01  4.591e-02  16.004  < 2e-16
## raceAsian-Pac-Islander                 6.245e-01  2.343e-01   2.666  0.00769
## raceBlack                              1.979e-01  2.006e-01   0.987  0.32371
## raceOther                              3.488e-01  2.930e-01   1.190  0.23387
## raceWhite                              4.348e-01  1.907e-01   2.280  0.02262
## genderMale                             2.578e-01  4.580e-02   5.630 1.80e-08
## capital.gain                           3.134e-04  9.210e-06  34.028  < 2e-16
## capital.loss                           6.731e-04  3.350e-05  20.092  < 2e-16
## hours.per.week                         2.903e-02  1.398e-03  20.762  < 2e-16
## native.countryCambodia                 2.669e-01  6.307e-01   0.423  0.67222
## native.countryCanada                   7.676e-01  2.521e-01   3.045  0.00233
## native.countryChina                   -8.489e-01  3.378e-01  -2.513  0.01198
## native.countryColumbia                -1.905e+00  6.740e-01  -2.826  0.00472
## native.countryCuba                     1.308e-01  3.124e-01   0.419  0.67534
## native.countryDominican-Republic      -7.774e-01  5.642e-01  -1.378  0.16827
## native.countryEcuador                 -8.825e-01  6.949e-01  -1.270  0.20413
## native.countryEl-Salvador             -6.935e-01  5.057e-01  -1.371  0.17032
## native.countryEngland                  6.812e-01  3.122e-01   2.182  0.02909
## native.countryFrance                   7.182e-01  5.374e-01   1.337  0.18138
## native.countryGermany                  2.749e-01  2.626e-01   1.047  0.29511
## native.countryGreece                   7.886e-02  4.034e-01   0.195  0.84500
## native.countryGuatemala               -1.682e+00  1.144e+00  -1.471  0.14137
## native.countryHaiti                    1.412e-01  4.993e-01   0.283  0.77737
## native.countryHoland-Netherlands      -9.735e+00  3.247e+02  -0.030  0.97608
## native.countryHonduras                 1.444e-01  1.175e+00   0.123  0.90220
## native.countryHong                    -1.736e-01  6.676e-01  -0.260  0.79488
## native.countryHungary                  5.253e-01  6.631e-01   0.792  0.42827
## native.countryIndia                   -1.544e-01  3.069e-01  -0.503  0.61477
## native.countryIran                    -3.084e-01  4.316e-01  -0.714  0.47492
## native.countryIreland                  1.418e+00  5.087e-01   2.788  0.00531
## native.countryItaly                    7.850e-01  3.147e-01   2.495  0.01261
## native.countryJamaica                  6.632e-01  3.898e-01   1.701  0.08892
## native.countryJapan                   -1.490e-02  3.708e-01  -0.040  0.96795
## native.countryLaos                    -1.153e+01  7.809e+01  -0.148  0.88260
## native.countryMexico                  -7.513e-01  2.350e-01  -3.197  0.00139
```

```
## native.countryNicaragua                    -3.634e-01  6.521e-01  -0.557  0.57732
## native.countryOutlying-US(Guam-USVI-etc) -4.319e-01  1.122e+00  -0.385  0.70018
## native.countryPeru                         -9.204e-01  7.096e-01  -1.297  0.19462
## native.countryPhilippines                  -5.160e-02  2.551e-01  -0.202  0.83970
## native.countryPoland                        1.561e-01  3.627e-01   0.430  0.66687
## native.countryPortugal                      9.106e-01  4.048e-01   2.249  0.02449
## native.countryPuerto-Rico                  -1.102e-01  3.303e-01  -0.334  0.73863
## native.countryScotland                     -1.902e+00  1.125e+00  -1.692  0.09072
## native.countrySouth                        -9.210e-01  3.956e-01  -2.328  0.01989
## native.countryTaiwan                        7.732e-02  4.336e-01   0.178  0.85847
## native.countryThailand                     -8.630e-01  6.545e-01  -1.318  0.18734
## native.countryTrinadad&Tobago              -1.775e+00  1.114e+00  -1.594  0.11102
## native.countryUnited-States                 2.150e-01  1.244e-01   1.729  0.08385
## native.countryVietnam                      -1.663e+00  6.019e-01  -2.763  0.00572
## native.countryYugoslavia                    8.159e-01  6.714e-01   1.215  0.22427
##
## (Intercept)                                ***
## age                                        ***
## workclassOther/Unknown                     .
## workclassPrivate
## workclassSelf-Employed                     ***
## education11th
## education12th                              *
## education1st-4th
## education5th-6th
## education7th-8th                           *
## education9th
## educationAssoc-acdm                        ***
## educationAssoc-voc                         ***
## educationBachelors                         ***
## educationDoctorate                         ***
## educationHS-grad                           ***
## educationMasters                           ***
## educationPreschool
## educationProf-school                       ***
## educationSome-college                      ***
## educational.num
## marital.statusMarried-AF-spouse            ***
## marital.statusMarried-civ-spouse           ***
## marital.statusMarried-spouse-absent
## marital.statusNever-married                ***
## marital.statusSeparated
## marital.statusWidowed
## occupationOther/Unknown
## occupationProfessional                     ***
## occupationSales                            ***
## occupationService                          **
## occupationWhite-Collar                     ***
## raceAsian-Pac-Islander                     **
## raceBlack
## raceOther
## raceWhite                                  *
## genderMale                                 ***
## capital.gain                               ***
```

```
## capital.loss                               ***
## hours.per.week                             ***
## native.countryCambodia
## native.countryCanada                       **
## native.countryChina                        *
## native.countryColumbia                     **
## native.countryCuba
## native.countryDominican-Republic
## native.countryEcuador
## native.countryEl-Salvador
## native.countryEngland                      *
## native.countryFrance
## native.countryGermany
## native.countryGreece
## native.countryGuatemala
## native.countryHaiti
## native.countryHoland-Netherlands
## native.countryHonduras
## native.countryHong
## native.countryHungary
## native.countryIndia
## native.countryIran
## native.countryIreland                      **
## native.countryItaly                        *
## native.countryJamaica                      .
## native.countryJapan
## native.countryLaos
## native.countryMexico                       **
## native.countryNicaragua
## native.countryOutlying-US(Guam-USVI-etc)
## native.countryPeru
## native.countryPhilippines
## native.countryPoland
## native.countryPortugal                     *
## native.countryPuerto-Rico
## native.countryScotland                     .
## native.countrySouth                        *
## native.countryTaiwan
## native.countryThailand
## native.countryTrinadad&Tobago
## native.countryUnited-States                .
## native.countryVietnam                      **
## native.countryYugoslavia
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 42846  on 39073  degrees of freedom
## Residual deviance: 25645  on 38994  degrees of freedom
## AIC: 25805
##
## Number of Fisher Scoring iterations: 11
```

```r
prob <- predict(lm, test, type = 'response')
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from a rank-deficient fit may be misleading
```

```r
pred1 <- rep('<=50K', length(prob))
pred1[prob>=.5] <- '>50K'
tb1 <- table(pred1, test$income)
tb1
```

```
##
## pred1   <=50K >50K
##   <=50K  6872  973
##   >50K    492 1431
```

```r
acc1 <- mean(pred1 == test$income)
print(paste("Accuary =", acc1))
```

```
## [1] "Accuary = 0.850020475020475"
```

```r
time <- end - beg
time
```

```
##    user  system elapsed
##    3.28    0.02    3.32
```

Naive Bayes comment: Naive Bayes took less time to compute than logistic regression, but the accuracy did diminish by around 0.034% due to simplicity of the model. Naive bayes works well with small data sets, that's why the results are not that great in this scenario. In addition, it doesn't model feature interactions so we can't really narrow down the data that way. Nevertheless, it is still interpretable to the user.

```r
beg <- proc.time()
nb <- naiveBayes(income ~., data = train)
end <- proc.time()
pred2 <- predict(nb, newdata = test, type = "class")
tb2 <- table(pred2, test$income)
tb2
```

```
##
## pred2   <=50K >50K
##   <=50K  6990 1424
##   >50K    374  980
```

```r
acc2 <- mean(pred2 == test$income)
print(paste("Accuary =", acc2))
```

```
## [1] "Accuary = 0.815929565929566"
```

```
time <- end - beg
time
```

```
##    user  system elapsed
##    0.13    0.00    0.12
```

SVM classification comment: SVM got the best accuracy compared to the other two algorithms used for classification. One reason is that SVM is robust to outliers since only support vectors define margins. However, computation time took very long due to large data.

```
beg <- proc.time()
svm <- ksvm(income ~ ., data = train)
end <- proc.time()
pred3 <- predict(svm, newdata = test, type = 'response')
tb3 <- table(pred3, test$income)
tb3
```

```
##
## pred3   <=50K >50K
##   <=50K  6978 1045
##   >50K    386 1359
```

```
acc3 <- mean(pred3 == test$income)
print(paste("Accuary =", acc3))
```

```
## [1] "Accuary = 0.853501228501228"
```

```
time <- end - beg
time
```

```
##    user  system elapsed
##  133.03    1.45  135.13
```

Result analysis: The majority of the observations in the dataset have income less than 50K. There are different factors to determine if people make more or less than 50K, such as gender, job types, and ethnicity. According to the first graph, making more money takes time and doesn't happen instantly. Most likely, promotions and job changes are happening behind the scenes. Nonetheless, no matter what factors determine a person's worth, it shouldn't undermine the possibility to achieve more than 50K within one's lifetime.