Adult_Income_Logistic_Regression_Model

Nabil Momin

2024-06-10

```
library(corrgram)
library(corrplot)
```

```
## corrplot 0.92 loaded
```

```
library(caTools)
library(Amelia)
```

```
## Loading required package: Rcpp
```

```
## ##
## ## Amelia II: Multiple Imputation
## ## (Version 1.8.2, built: 2024-04-10)
## ## Copyright (C) 2005-2024 James Honaker, Gary King and Matthew Blackwell
## ## Refer to http://gking.harvard.edu/amelia/ for more information
## ##
```

```
library(dplyr)
```

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

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

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

```
library(ggplot2)

## Getting the data

adult <- read.csv('adult_sal.csv')

View(adult)

## Getting rid of extra column X

adult <- select(adult,- X)

View(adult)
```

```
summary(adult)
```

```
##       age        type_employer          fnlwgt          education        
##  Min.   :17.00   Length:32561       Min.   :  12285   Length:32561      
##  1st Qu.:28.00   Class :character   1st Qu.: 117827   Class :character  
##  Median :37.00   Mode  :character   Median : 178356   Mode  :character  
##  Mean   :38.58                      Mean   : 189778                     
##  3rd Qu.:48.00                      3rd Qu.: 237051                     
##  Max.   :90.00                      Max.   :1484705                     
##  education_num     marital           occupation        relationship      
##  Min.   : 1.00   Length:32561       Length:32561       Length:32561      
##  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               sex             capital_gain    capital_loss   
##  Length:32561       Length:32561       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   : 1078   Mean   :  87.3  
##                                        3rd Qu.:    0   3rd Qu.:   0.0  
##                                        Max.   :99999   Max.   :4356.0  
##   hr_per_week      country             income         
##  Min.   : 1.00   Length:32561       Length:32561      
##  1st Qu.:40.00   Class :character   Class :character  
##  Median :40.00   Mode  :character   Mode  :character  
##  Mean   :40.44                                        
##  3rd Qu.:45.00                                        
##  Max.   :99.00                                        
```

```
str(adult)
```

```
## 'data.frame':    32561 obs. of  15 variables:
##  $ age          : int  39 50 38 53 28 37 49 52 31 42 ...
##  $ type_employer: chr  "State-gov" "Self-emp-not-inc" "Private" "Private"
...
##  $ fnlwgt       : int  77516 83311 215646 234721 338409 284582 160187
209642 45781 159449 ...
##  $ education    : chr  "Bachelors" "Bachelors" "HS-grad" "11th" ...
##  $ education_num: int  13 13 9 7 13 14 5 9 14 13 ...
##  $ marital      : chr  "Never-married" "Married-civ-spouse" "Divorced"
"Married-civ-spouse" ...
##  $ occupation   : chr  "Adm-clerical" "Exec-managerial" "Handlers-
cleaners" "Handlers-cleaners" ...
##  $ relationship : chr  "Not-in-family" "Husband" "Not-in-family" "Husband"
...
##  $ race         : chr  "White" "White" "White" "Black" ...
##  $ sex          : chr  "Male" "Male" "Male" "Male" ...
##  $ capital_gain : int  2174 0 0 0 0 0 0 0 14084 5178 ...
##  $ capital_loss : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ hr_per_week  : int  40 13 40 40 40 40 16 45 50 40 ...
##  $ country      : chr  "United-States" "United-States" "United-States"
"United-States" ...
##  $ income       : chr  "<=50K" "<=50K" "<=50K" "<=50K" ...
```

```
## Data cleaning

table(adult$type_employer)
```

```
## 
##                 ?      Federal-gov        Local-gov     Never-worked 
##              1836              960             2093                7 
##           Private     Self-emp-inc Self-emp-not-inc        State-gov 
##             22696             1116             2541             1298 
##       Without-pay 
##                14
```

```
table(adult$marital)
```

```
## 
##               Divorced      Married-AF-spouse     Married-civ-spouse 
##                   4443                     23                  14976 
## Married-spouse-absent          Never-married              Separated 
##                    418                  10683                   1025 
##                Widowed 
##                    993
```

```
adult$type_employer <- gsub('Never-worked','Unemployed',adult$type_employer)
adult$type_employer <- gsub('Without-pay','Unemployed',adult$type_employer)

adult$type_employer <- gsub('Self-emp-inc','Self-emp',adult$type_employer)
adult$type_employer <- gsub('Self-emp-not-inc','Self-
emp',adult$type_employer)
View(adult$type_employer)
View(adult)
table(adult$type_employer)
```

```
## 
##           ? Federal-gov   Local-gov     Private    Self-emp   State-gov 
##        1836         960        2093       22696        3657        1298 
##  Unemployed 
##          21
```

```
## Getting rid of the ? and replacing it with NA

adult[adult=='?'] <- NA

## Cleaning column type employer
```

```
adult$type_employer <- gsub(' Local-gov','SL-gov',adult$type_employer)
adult$type_employer <- gsub('State-gov','SL-gov',adult$type_employer)
table(adult$type_employer)
```

```
## 
## Federal-gov   Local-gov     Private    Self-emp      SL-gov  Unemployed 
##         960        2093       22696        3657        1298          21
```

```
adult$type_employer <- gsub('Local-gov','SL-gov',adult$type_employer)
table(adult$type_employer)
```

```
## 
## Federal-gov     Private    Self-emp      SL-gov  Unemployed 
##         960       22696        3657        3391          21
```

```
## Cleaning the column education now

table(adult$education)
```

```
## 
##         10th         11th         12th      1st-4th      5th-6th      7th-
8th 
##          933         1175          433          168          333          
646 
##          9th   Assoc-acdm    Assoc-voc    Bachelors    Doctorate      HS-
grad 
##          514         1067         1382         5355          413        
10501 
##      Masters    Preschool  Prof-school Some-college 
##         1723           51          576         7291
```

```
edu <- function(ed){
  ed <- as.character(ed)
  if (ed =='10th' | ed =='11th' |ed=='12th' | ed=='1st-4th' | ed=='5th-6th' |
ed=='7th-8th' | ed=='9th' | ed=='Preschool'){
    return('School')
  }else if(ed == 'Assoc-acdm' | ed=='Assoc-voc'){
    return('Associate')
  }else{
    return(ed)
  }
}

adult$education <- sapply(adult$education,edu)

## Cleaning the column marital now

marital <- function(mar){
  mar <- as.character(mar)
  if (mar=='Seperated' | mar=='Divorced' | mar=='Widowed'){
    return('Not-Married')
  }else if(mar =='Never-married'){
    return(mar)
  }else{
    return('Married')
  }
}

adult$marital <- sapply(adult$marital,marital)

table(adult$marital)

##
##       Married Never-married   Not-Married
##         16442         10683          5436

## Cleaning relationship

relation <- function(relate){
  relate <- as.character(relate)
  if(relate == 'Not-in-family' | relate=='Other-relative' | relate=='Own-
child' | relate=='Unmarried'){
    return('Complicated')
  }else{
    return(relate)
  }
}

adult$relationship <- sapply(adult$relationship,relation)

table(adult$relationship)

##
## Complicated     Husband        Wife
##       17800       13193        1568

## Cleaning country

table(adult$country)

##
##                   Cambodia                     Canada
##                         19                        121
##                      China                   Columbia
##                         75                         59
##                       Cuba         Dominican-Republic
##                         95                         70
##                    Ecuador                El-Salvador
##                         28                        106
##                    England                     France
##                         90                         29
##                     Germany                      Greece
##                         137                          29
##                   Guatemala                       Haiti
##                          64                          44
##          Holand-Netherlands                    Honduras
##                           1                          13
##                        Hong                     Hungary
##                          20                          13
##                       India                        Iran
##                         100                          43
##                     Ireland                       Italy
##                          24                          73
##                     Jamaica                       Japan
##                          81                          62
##                        Laos                      Mexico
##                          18                         643
##                   Nicaragua Outlying-US(Guam-USVI-etc)
##                          34                          14
##                        Peru                 Philippines
##                          31                         198
##                      Poland                    Portugal
##                          60                          37
##                 Puerto-Rico                    Scotland
##                         114                          12
##                       South                      Taiwan
##                          80                          51
##                    Thailand             Trinadad&Tobago
##                          18                          19
##               United-States                     Vietnam
##                       29170                          67
##                  Yugoslavia
##                          16
```

```
Asia <- c('China','Hong','India','Iran','Cambodia','Japan', 'Laos' ,
          'Philippines' ,'Vietnam' ,'Taiwan', 'Thailand')

North.America <- c('Canada','United-States','Puerto-Rico' )

Europe <- c('England' ,'France', 'Germany' ,'Greece','Holand-
Netherlands','Hungary',
            'Ireland','Italy','Poland','Portugal','Scotland','Yugoslavia')

Latin.and.South.America <- c('Columbia','Cuba','Dominican-
Republic','Ecuador',
                             'El-Salvador','Guatemala','Haiti','Honduras',
                             'Mexico','Nicaragua','Outlying-US(Guam-USVI-
etc)','Peru',
                             'Jamaica','Trinadad&Tobago')
Other <- c('South')
```

```r
group_country <- function(ctry){
  if (ctry %in% Asia){
    return('Asia')
  }else if (ctry %in% North.America){
    return('North.America')
  }else if (ctry %in% Europe){
    return('Europe')
  }else if (ctry %in% Latin.and.South.America){
    return('Latin.and.South.America')
  }else{
    return('Other')
  }
}

adult$country <- sapply(adult$country,group_country)

## Factoring

str(adult)
```

```
## 'data.frame':    32561 obs. of  15 variables:
##  $ age           : int  39 50 38 53 28 37 49 52 31 42 ...
##  $ type_employer: chr  "SL-gov" "Self-emp" "Private" "Private" ...
##  $ fnlwgt        : int  77516 83311 215646 234721 338409 284582 160187
209642 45781 159449 ...
##  $ education     : chr  "Bachelors" "Bachelors" "HS-grad" "School" ...
##  $ education_num: int  13 13 9 7 13 14 5 9 14 13 ...
##  $ marital       : chr  "Never-married" "Married" "Not-Married" "Married"
...
##  $ occupation    : chr  "Adm-clerical" "Exec-managerial" "Handlers-
cleaners" "Handlers-cleaners" ...
##  $ relationship : chr  "Complicated" "Husband" "Complicated" "Husband" ...
##  $ race          : chr  "White" "White" "White" "Black" ...
##  $ sex           : chr  "Male" "Male" "Male" "Male" ...
##  $ capital_gain : int  2174 0 0 0 0 0 0 0 14084 5178 ...
##  $ capital_loss : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ hr_per_week  : int  40 13 40 40 40 40 16 45 50 40 ...
##  $ country       : chr  "North.America" "North.America" "North.America"
"North.America" ...
##  $ income        : chr  "<=50K" "<=50K" "<=50K" "<=50K" ...
```

```r
adult$education <- factor(adult$education)
adult$country <- factor(adult$country)
adult$marital <- factor(adult$marital)
adult$type_employer <- factor(adult$type_employer)
adult$relationship <- factor(adult$relationship)
adult$sex <- factor(adult$sex)
adult$income <- factor(adult$income)
str(adult)
```

```
## 'data.frame':    32561 obs. of  15 variables:
##  $ age          : int  39 50 38 53 28 37 49 52 31 42 ...
##  $ type_employer: Factor w/ 5 levels "Federal-gov",..: 4 3 2 2 2 2 2 3 2 2
...
##  $ fnlwgt       : int  77516 83311 215646 234721 338409 284582 160187
209642 45781 159449 ...
##  $ education    : Factor w/ 8 levels "Associate","Bachelors",..: 2 2 4 7 2
5 7 4 5 2 ...
##  $ education_num: int  13 13 9 7 13 14 5 9 14 13 ...
##  $ marital      : Factor w/ 3 levels "Married","Never-married",..: 2 1 3 1
1 1 1 1 2 1 ...
##  $ occupation   : chr  "Adm-clerical" "Exec-managerial" "Handlers-
cleaners" "Handlers-cleaners" ...
##  $ relationship : Factor w/ 3 levels "Complicated",..: 1 2 1 2 3 3 1 2 1 2
...
##  $ race         : chr  "White" "White" "White" "Black" ...
##  $ sex          : Factor w/ 2 levels "Female","Male": 2 2 2 2 1 1 1 2 1 2
...
##  $ capital_gain : int  2174 0 0 0 0 0 0 0 14084 5178 ...
##  $ capital_loss : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ hr_per_week  : int  40 13 40 40 40 40 16 45 50 40 ...
##  $ country      : Factor w/ 5 levels "Asia","Europe",..: 4 4 4 4 3 4 3 4 4
4 ...
##  $ income       : Factor w/ 2 levels "<=50K",">50K": 1 1 1 1 1 1 1 2 2 2
...
```

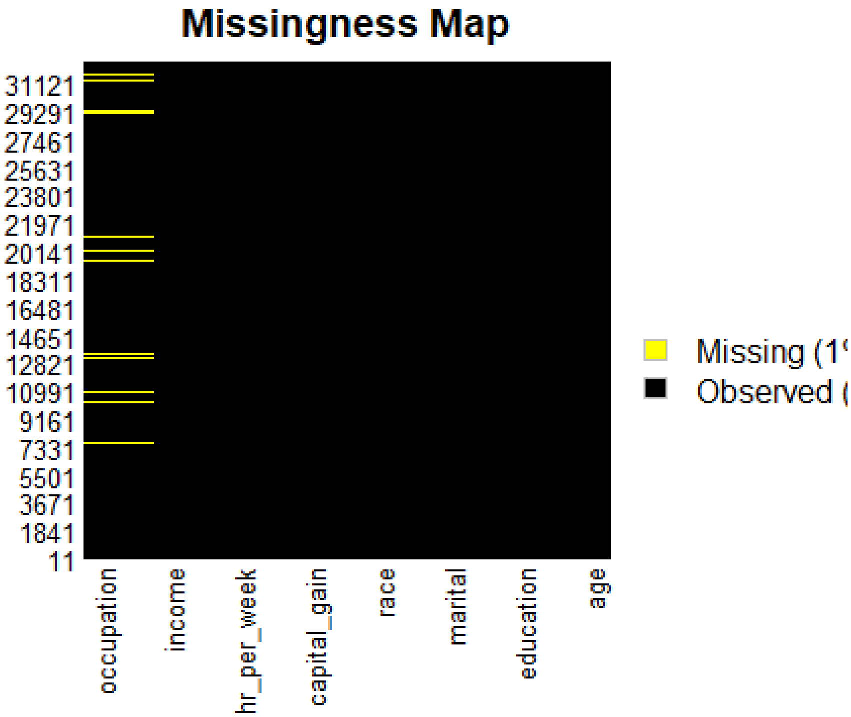
```
any(is.na(adult))
```

```
## [1] TRUE
```

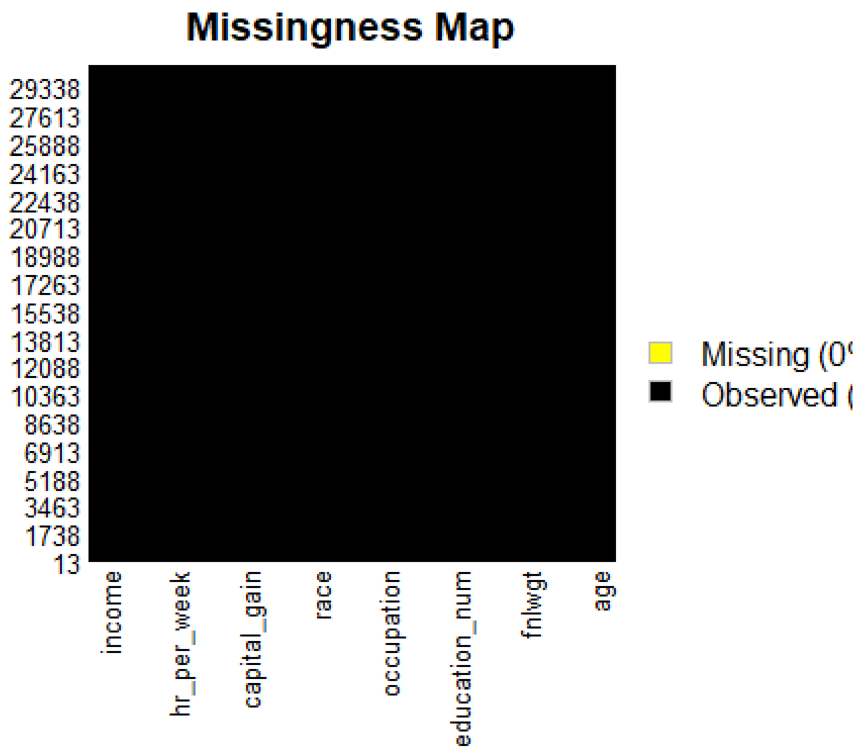
```
## we need to repeat the factor function so we don't see that ? in the str

str(adult)
```

```
## 'data.frame':    32561 obs. of  15 variables:
##  $ age          : int  39 50 38 53 28 37 49 52 31 42 ...
##  $ type_employer: Factor w/ 5 levels "Federal-gov",..: 4 3 2 2 2 2 2 3 2 2
...
##  $ fnlwgt       : int  77516 83311 215646 234721 338409 284582 160187
209642 45781 159449 ...
##  $ education    : Factor w/ 8 levels "Associate","Bachelors",..: 2 2 4 7 2
5 7 4 5 2 ...
##  $ education_num: int  13 13 9 7 13 14 5 9 14 13 ...
##  $ marital      : Factor w/ 3 levels "Married","Never-married",..: 2 1 3 1
1 1 1 1 2 1 ...
##  $ occupation   : chr  "Adm-clerical" "Exec-managerial" "Handlers-
cleaners" "Handlers-cleaners" ...
##  $ relationship : Factor w/ 3 levels "Complicated",..: 1 2 1 2 3 3 1 2 1 2
...
##  $ race         : chr  "White" "White" "White" "Black" ...
##  $ sex          : Factor w/ 2 levels "Female","Male": 2 2 2 2 1 1 1 2 1 2
```

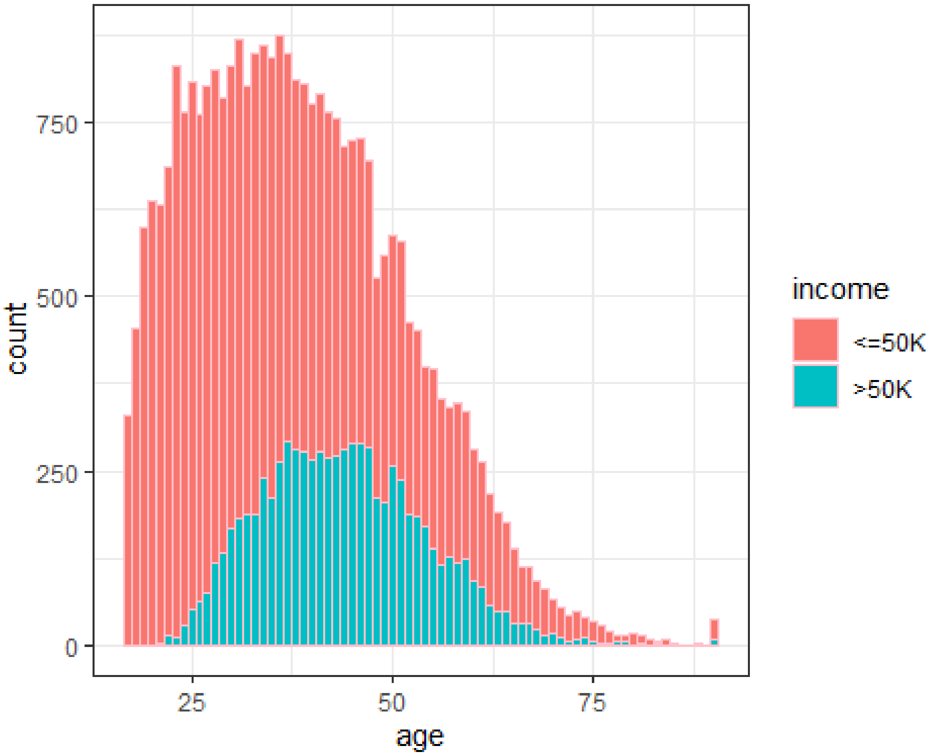
```
...
##  $ capital_gain : int  2174 0 0 0 0 0 0 0 14084 5178 ...
##  $ capital_loss : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ hr_per_week  : int  40 13 40 40 40 40 16 45 50 40 ...
##  $ country      : Factor w/ 5 levels "Asia","Europe",..: 4 4 4 4 3 4 3 4 4
4 ...
##  $ income       : Factor w/ 2 levels "<=50K",">50K": 1 1 1 1 1 1 1 2 2 2
...

table(adult$type_employer)

##
## Federal-gov     Private    Self-emp      SL-gov  Unemployed
##         960       22696        3657        3391          21

## Using Amelia

missmap(adult,legend = TRUE,col = c('yellow','black'))
```



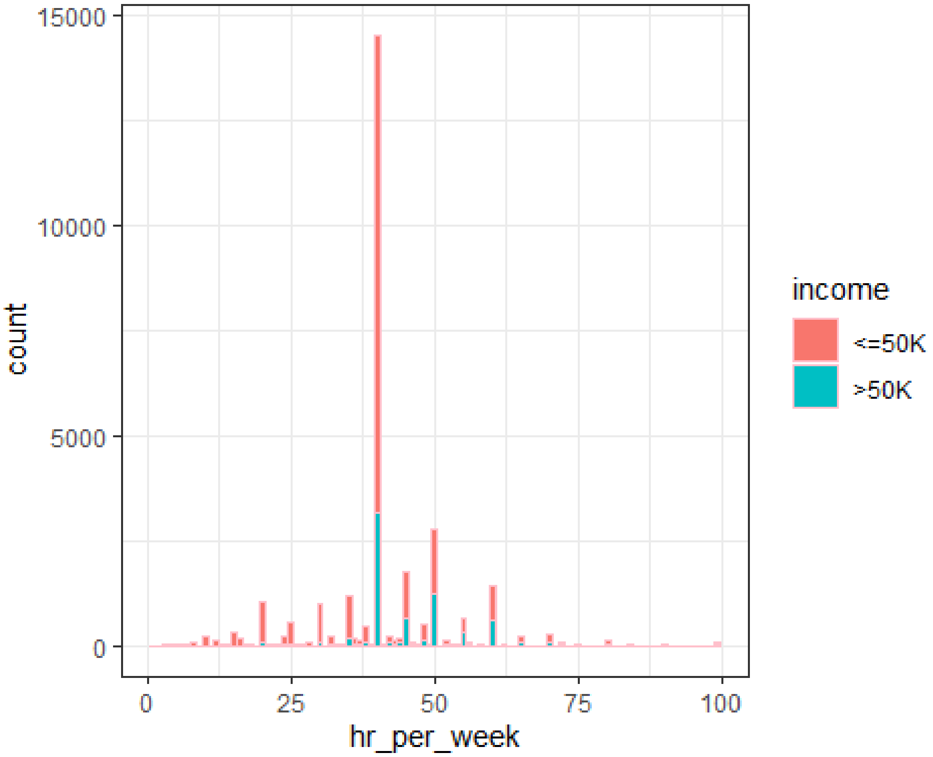
```
## Removing the NA value from the data set

adult <- na.omit(adult)

missmap(adult,legend = TRUE,col = c('yellow','black'))
```



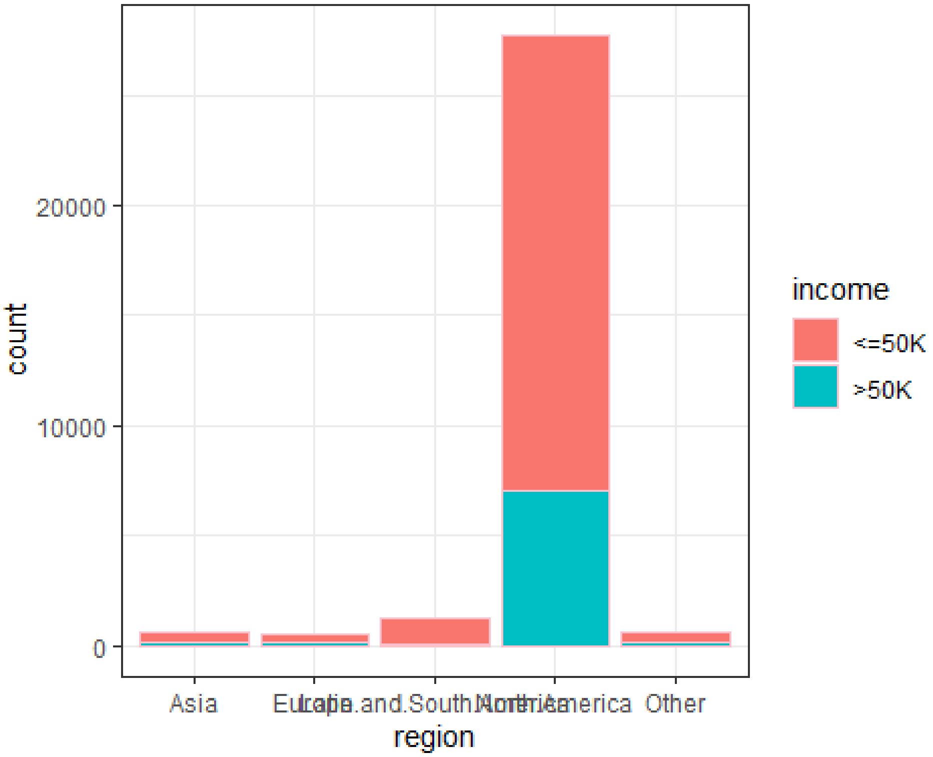
```
## EDA time

ggplot(adult,aes(age)) +
geom_histogram(aes(fill=income),color='pink',binwidth = 1) + theme_bw()
```



```r
ggplot(adult,aes(hr_per_week)) +
geom_histogram(aes(fill=income),color='pink',binwidth = 1) + theme_bw()
```



```
# rename the column name of country to region because it now does not make
sense

adult <- rename(adult,region=country)

table(adult$region)

## 
##                    Asia                  Europe Latin.and.South.America 
##                     634                     493                    1244 
##           North.America                   Other 
##                   27720                     627

ggplot(adult,aes(region)) + geom_bar(aes(fill=income),color='pink',binwidth =
1) + theme_bw()

## Warning in geom_bar(aes(fill = income), color = "pink", binwidth = 1):
Ignoring
## unknown parameters: `binwidth`
```



```
## Modeling

sample <- sample.split(adult$income,SplitRatio = 0.7)
train <- subset(adult,sample == TRUE)
test <- subset(adult,sample == FALSE)

str(train)
```

```
## 'data.frame':    21503 obs. of  15 variables:
##  $ age          : int  39 50 38 28 37 52 31 42 37 30 ...
##  $ type_employer: Factor w/ 5 levels "Federal-gov",..: 4 3 2 2 2 3 2 2 2 4
...
##  $ fnlwgt       : int  77516 83311 215646 338409 284582 209642 45781
159449 280464 141297 ...
##  $ education    : Factor w/ 8 levels "Associate","Bachelors",..: 2 2 4 2 5
4 5 2 8 2 ...
##  $ education_num: int  13 13 9 13 14 9 14 13 10 13 ...
##  $ marital      : Factor w/ 3 levels "Married","Never-married",..: 2 1 3 1
1 1 2 1 1 1 ...
##  $ occupation   : chr  "Adm-clerical" "Exec-managerial" "Handlers-
cleaners" "Prof-specialty" ...
##  $ relationship : Factor w/ 3 levels "Complicated",..: 1 2 1 3 3 2 1 2 2 2
...
##  $ race         : chr  "White" "White" "White" "Black" ...
##  $ sex          : Factor w/ 2 levels "Female","Male": 2 2 2 1 1 2 1 2 2 2
...
##  $ capital_gain : int  2174 0 0 0 0 0 14084 5178 0 0 ...
##  $ capital_loss : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ hr_per_week  : int  40 13 40 40 40 45 50 40 80 40 ...
##  $ region       : Factor w/ 5 levels "Asia","Europe",..: 4 4 4 3 4 4 4 4 4
1 ...
##  $ income       : Factor w/ 2 levels "<=50K",">50K": 1 1 1 1 1 2 2 2 2 2
...
```

```
str(test)
```

```
## 'data.frame':    9215 obs. of  15 variables:
##  $ age          : int  53 49 32 54 43 56 19 23 20 22 ...
##  $ type_employer: Factor w/ 5 levels "Federal-gov",..: 2 2 2 2 2 4 2 4 2 4
...
##  $ fnlwgt       : int  234721 160187 186824 302146 117037 216851 168294
190709 266015 311512 ...
##  $ education    : Factor w/ 8 levels "Associate","Bachelors",..: 7 7 4 4 7
2 4 1 8 8 ...
##  $ education_num: int  7 5 9 9 7 13 9 12 10 10 ...
##  $ marital      : Factor w/ 3 levels "Married","Never-married",..: 1 1 2 1
1 1 2 2 2 1 ...
##  $ occupation   : chr  "Handlers-cleaners" "Other-service" "Machine-op-
inspct" "Other-service" ...
##  $ relationship : Factor w/ 3 levels "Complicated",..: 2 1 1 1 2 2 1 1 1 2
...
##  $ race         : chr  "Black" "Black" "White" "Black" ...
##  $ sex          : Factor w/ 2 levels "Female","Male": 2 1 2 1 2 2 2 2 2 2
...
##  $ capital_gain : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ capital_loss : int  0 0 0 0 2042 0 0 0 0 0 ...
##  $ hr_per_week  : int  40 16 40 20 40 40 40 52 44 15 ...
##  $ region       : Factor w/ 5 levels "Asia","Europe",..: 4 3 4 4 4 4 4 4 4
```

```
4 ...
##  $ income       : Factor w/ 2 levels "<=50K",">50K": 1 1 1 1 1 2 1 1 1 1
...
```

```
model <- glm(income ~ . ,family = binomial(link='logit'),train)
```

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

```
View(train)
View(test)
View(adult)

summary(model)
```

```
## 
## Call:
## glm(formula = income ~ ., family = binomial(link = "logit"), 
##     data = train)
## 
## Coefficients:
##                                  Estimate Std. Error z value Pr(>|z|)    
## (Intercept)                    -7.630e+00  7.533e-01 -10.129  < 2e-16 ***
## age                             2.731e-02  1.980e-03  13.795  < 2e-16 ***
## type_employerPrivate           -4.784e-01  1.109e-01  -4.315 1.60e-05 ***
## type_employerSelf-emp          -8.066e-01  1.234e-01  -6.538 6.22e-11 ***
## type_employerSL-gov            -7.021e-01  1.245e-01  -5.639 1.71e-08 ***
## type_employerUnemployed        -1.152e+01  1.026e+02  -0.112 0.910671    
## fnlwgt                          6.291e-07  2.096e-07   3.002 0.002686 ** 
## educationBachelors              3.184e-01  1.195e-01   2.666 0.007680 ** 
## educationDoctorate              9.632e-01  3.153e-01   3.055 0.002254 ** 
## educationHS-grad               -2.237e-01  1.547e-01  -1.446 0.148262    
## educationMasters                5.412e-01  1.775e-01   3.050 0.002291 ** 
## educationProf-school            9.150e-01  2.588e-01   3.536 0.000406 ***
## educationSchool                -7.047e-01  3.317e-01  -2.125 0.033597 *  
## educationSome-college          -6.641e-02  1.149e-01  -0.578 0.563250    
## education_num                   1.312e-01  5.408e-02   2.426 0.015284 *  
## maritalNever-married           -8.265e-01  1.441e-01  -5.736 9.71e-09 ***
## maritalNot-Married             -1.784e-01  1.431e-01  -1.247 0.212556    
## occupationArmed-Forces         -8.427e-01  1.705e+00  -0.494 0.621067    
## occupationCraft-repair          2.787e-02  9.486e-02   0.294 0.768877    
## occupationExec-managerial       8.025e-01  9.146e-02   8.774  < 2e-16 ***
## occupationFarming-fishing      -1.141e+00  1.690e-01  -6.751 1.47e-11 ***
## occupationHandlers-cleaners    -7.640e-01  1.701e-01  -4.492 7.05e-06 ***
## occupationMachine-op-inspct    -3.280e-01  1.197e-01  -2.740 0.006143 ** 
## occupationOther-service        -7.929e-01  1.368e-01  -5.796 6.78e-09 ***
## occupationPriv-house-serv      -3.906e+00  2.009e+00  -1.945 0.051808 .  
## occupationProf-specialty        5.573e-01  9.701e-02   5.745 9.21e-09 ***
## occupationProtective-serv       5.969e-01  1.452e-01   4.111 3.94e-05 ***
## occupationSales                 2.910e-01  9.752e-02   2.984 0.002842 ** 
## occupationTech-support          5.650e-01  1.310e-01   4.314 1.60e-05 ***
## occupationTransport-moving     -1.669e-01  1.177e-01  -1.418 0.156313    
## relationshipHusband            1.524e+00  1.335e-01  11.415  < 2e-16 ***
## relationshipWife               2.884e+00  1.585e-01  18.198  < 2e-16 ***
## raceAsian-Pac-Islander         1.097e+00  3.371e-01   3.254 0.001140 **
## raceBlack                      8.001e-01  3.050e-01   2.623 0.008712 **
## raceOther                      4.484e-01  4.227e-01   1.061 0.288832
## raceWhite                      9.010e-01  2.929e-01   3.076 0.002100 **
## sexMale                        8.753e-01  9.247e-02   9.466  < 2e-16 ***
## capital_gain                   3.353e-04  1.269e-05  26.419  < 2e-16 ***
## capital_loss                   6.610e-04  4.588e-05  14.406  < 2e-16 ***
## hr_per_week                    3.079e-02  2.008e-03  15.328  < 2e-16 ***
## regionEurope                   3.470e-01  2.579e-01   1.345 0.178497
## regionLatin.and.South.America -3.223e-01  2.573e-01  -1.252 0.210426
## regionNorth.America            2.379e-01  2.063e-01   1.153 0.248801
## regionOther                   -2.608e-01  2.319e-01  -1.124 0.260813
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 24138  on 21502  degrees of freedom
## Residual deviance: 14090  on 21459  degrees of freedom
## AIC: 14178
##
## Number of Fisher Scoring iterations: 11
```

```
## Using the predict

test$predict.income <- predict(model,test,type='response')

table(test$income,test$predict.income >0.5)
```

```
##
##          FALSE TRUE
##   <=50K   6433  487
##   >50K     901 1394
```

```
## calculating how accurate how model is

acc <- (6409+1366)/(6409+511+929+1366)

## Our accuracy is 0.84

print(acc)
```

```
## [1] 0.843733
```

```
## Recall is 0.92

recall <- 6409/(6409+511)
```

```r
print(recall)

## [1] 0.9261561

## Precision is 0.87

precision <- 6409/(6409+929)

print(precision)

## [1] 0.8733987
```