

Credit Card Applicants Analysis

BY

Karthikeyan K

Data source: data.world

Dataset Description:

This data set talks about the details of the credit card applicants in Banks of America. Which have attributes of 13 and rows about 1600 which describes.

- ID – Applicant's ID
- CODE_GENDER – Gender of the Applicant
- FLAG_OWN_CAR – Does Applicant owns a car or not
- FLAG_OWN_REALTY – Does Applicant owns a Land or not
- AMT_INCOME_TOTAL – Income of an Applicant
- NAME_INCOME_TYPE – Type of Income
- NAME_EDUCATION_TYPE – Applicant's Education
- NAME_FAMILY_STATUS – Is Applicant is Married or not
- NAME_HOUSING_TYPE – Is Applicant live in Rent House or not
- DAYS_BIRTH – No of days were applicant born
- OCCUPATION_TYPE – Applicant's Occupation
- CNT_FAM_MEMBERS – Total family members of the Applicant

Assumption:

- From the given dataset we can issue credit card to the clients based on their income because they will pay the dues correctly
- And also, don't give the credit card to the old age people or people have age 50-60 because anytime they leave their job this leads to shortage of money to pay their dues.
- We can issue the credit card to the people who have own land and own car because banks easily mortgage those things when applicants don't pay their dues correctly
- And also, we can issue the credit card to the applicants who have married because both members may pay their dues correctly.
- From all the assumptions we subset the data to issue the credit card for the clients. From the assumptions these people may be pay their dues on time.

SOURCE CODE:

```
library(dplyr)

library(lattice)
library(plyr)

library(readr)
library(ggplot2)

credit_application <- read_csv("credit_application.csv")

## Rows: 2078 Columns: 14

## — Column specification —————
## Delimiter: ","
## chr (8): CODE_GENDER, FLAG_OWN_CAR, FLAG_OWN_REALTY, NAME_INCOME_TYPE, NAME_...
## dbl (6): ID, AMT_INCOME_TOTAL, DAYS_BIRTH, DAYS_EMPLOYED, FLAG_PHONE,

df=credit_application
head(df,5)

## # A tibble: 5 × 14
##       ID CODE_...1 FLAG_...2 FLAG_...3 AMT_I...4 NAME_...5 NAME_...6 NAME_...7 NAME_...8 DAYS_...9
##   <dbl> <chr>   <chr>   <chr>   <dbl> <chr>   <chr>   <chr>   <chr>   <dbl>
## 1 5.01e6 M      Y      Y      427500 Working Higher... Civil ... Rented... -12005
## 2 5.01e6 M      Y      Y      427500 Working Higher... Civil ... Rented... -12005
## 3 5.01e6 M      Y      Y      112500 Working Second... Married House ... -21474
## 4 5.01e6 F      N      Y      270000 Commer... Second... Single... House ... -19110
## 5 5.01e6 F      N      Y      270000 Commer... Second... Single... House ... -19110
## # ... with 4 more variables: DAYS_EMPLOYED <dbl>, FLAG_PHONE <dbl>,
## #   OCCUPATION_TYPE <chr>, CNT_FAM_MEMBERS <dbl>, and abbreviated variable
## #   names 1CODE_GENDER, 2FLAG_OWN_CAR, 3FLAG_OWN_REALTY, 4AMT_INCOME_TOTAL,
## #   5NAME_INCOME_TYPE, 6NAME_EDUCATION_TYPE, 7NAME_FAMILY_STATUS,
## #   8NAME_HOUSING_TYPE, 9DAYS_BIRTH

colnames(df) =
c("id","gender","owncar","ownland","income","incometype","edu","status","house","
age","exp","occupation","familycount")

#filling empty string
colSums(is.na(df))

##           id           gender           owncar           ownland           income           incometype
##           0             0             0             0             0             0
##           edu           status           house           age           exp
##           0             0             0             0             0
## occupation familycount
##           620             0
```

```
df['occupation'][is.na(df['occupation'])] = 'NIL'
summary(df)
```

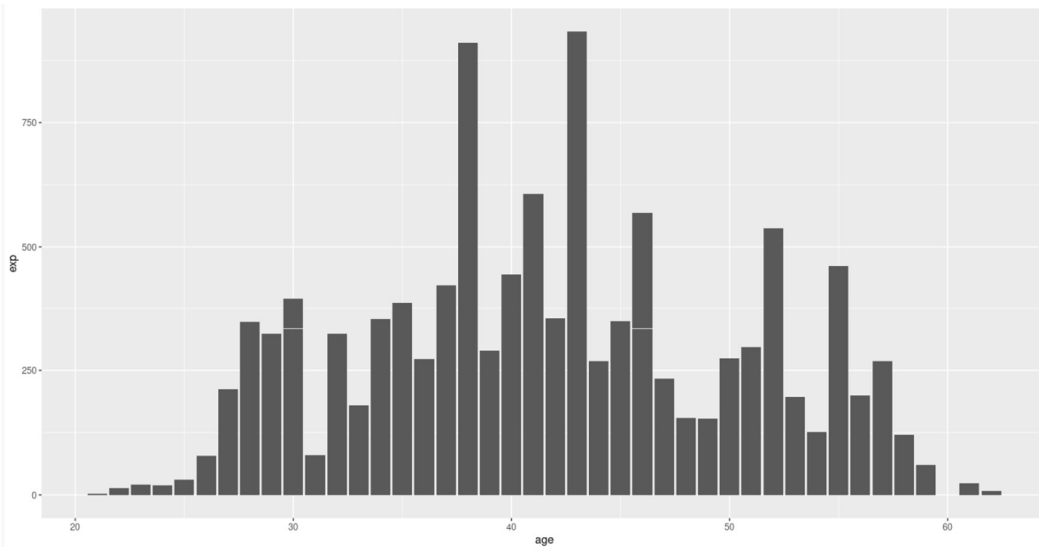
```
##      id          gender      own car      own land
## Min.   :1968310   Length:2078   Length:2078   Length:2078
## 1st Qu.:5009457   Class :character   Class :character   Class :character
## Median :5010142   Mode  :character   Mode  :character   Mode  :character
## Mean    :5116383
## 3rd Qu.:5010773
## Max.    :6736968
##      income      income type      edu      status
## Min.   : 33300   Length:2078   Length:2078   Length:2078
## 1st Qu.:135000   Class :character   Class :character   Class :character
## Median :166500   Mode  :character   Mode  :character   Mode  :character
## Mean    :205785
## 3rd Qu.:270000
## Max.    :1350000
##      house      age      exp
## Length:2078   Min.   : -23768   Min.   : -10936
## Class :character   1st Qu.: -19958   1st Qu.: -3327
## Mode  :character   Median : -15759   Median : -1773
##                      Mean    : -16249   Mean    : 62266
##                      3rd Qu.: -12927   3rd Qu.: -495
##                      Max.    : -7489    Max.    :365243
##      occupation      family count
## Length:2078   Min.   :1.000
## Class :character   1st Qu.:2.000
## Mode  :character   Median :2.000
##                      Mean    :2.165
##                      3rd Qu.:2.000
##                      Max.    :5.000
```

```
#converting age days to years #age wise experience
#converting exp days to years
```

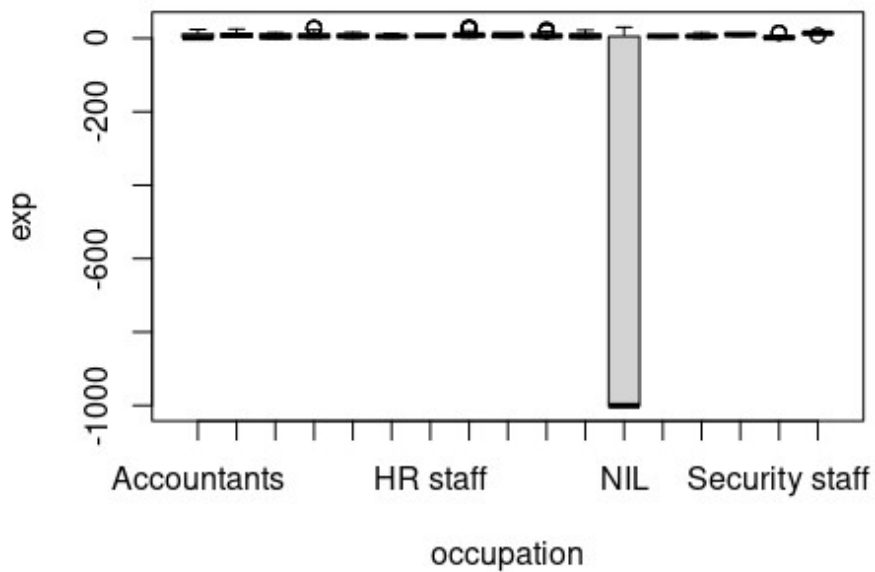
```
df$age=df$age/-365
df$exp=df$exp/-365
df$exp <- round(df$exp)
df$age <- round(df$age)
df=df %>% mutate(agerange = case_when(age>=20 & age<=30 ~ "20 to 30",
                                       age>=30 & age<=40 ~ "30 to 40",
                                       age>=40 & age<=50 ~ "40 to 50",
                                       age>=50 ~ "above 50"))
```

```
#age wise experience
```

```
ggplot(data=df, aes(x=age, y=exp))+ geom_bar(stat="identity")
```

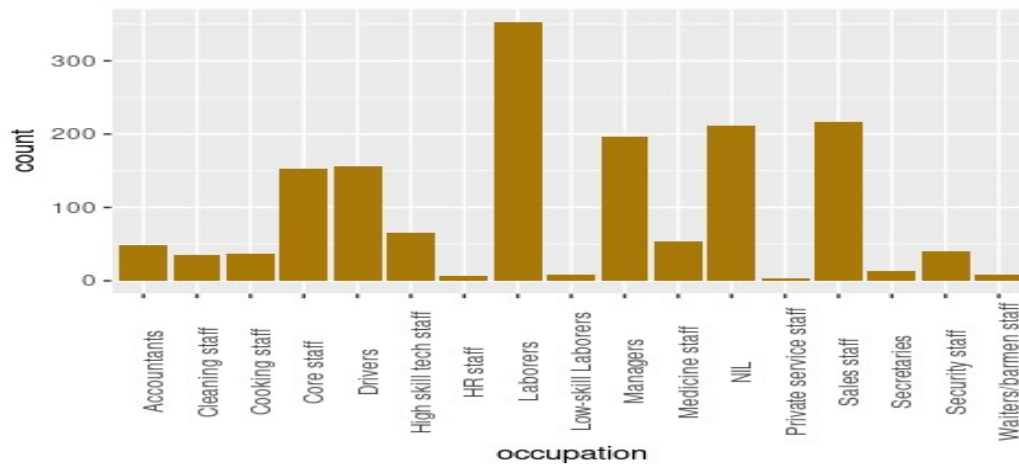


```
#removing outliers in exp
boxplot(exp~ occupation, data =df)
```

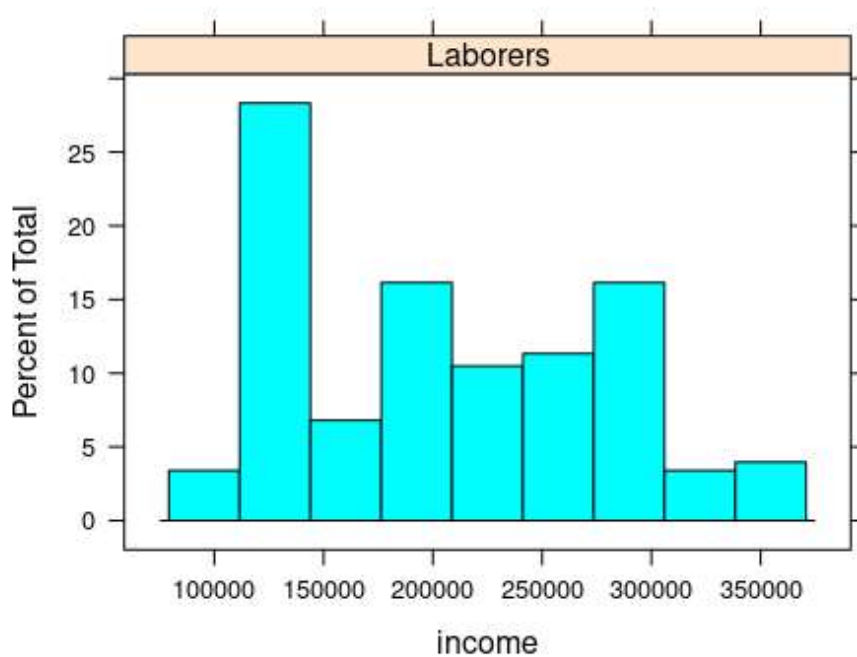


```
Q1 <- quantile(df$exp, .25)
Q3 <- quantile(df$exp, .75)
IQR <- IQR(df$exp)
df <- subset(df, df$exp > (Q1 - 1.5*IQR) & df$exp < (Q3 + 1.5*IQR))
```

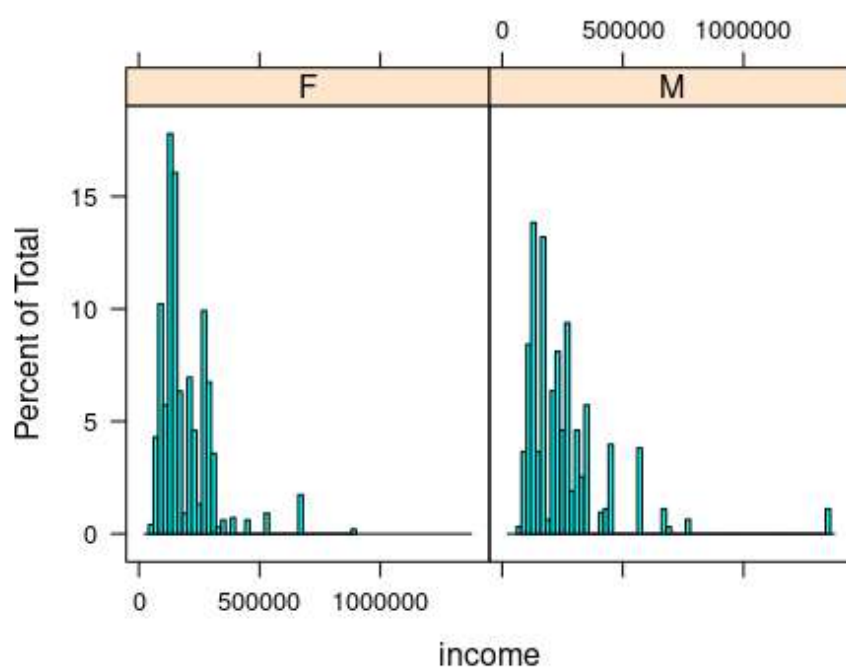
```
#count of each applicant's job
df %>% ggplot(aes(occupation))+
  geom_bar(stat="Count",fill="#A87809")+theme(axis.text.x = element_text(angle =
90))
```



```
hist=subset(df,occupation=="Laborers")
histogram(~income|occupation,data=hist)
```

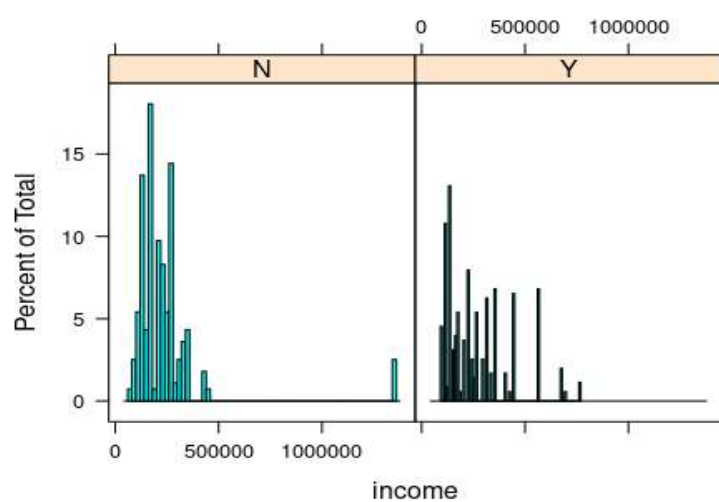


```
#distribution of income by gender
histogram(~income|gender,data=df,breaks=50)
```

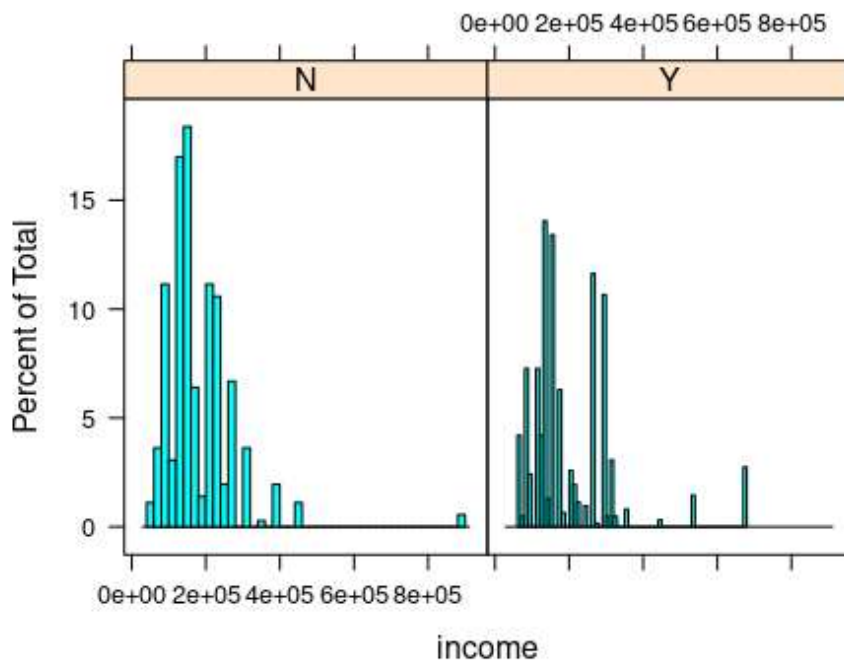


```
dm=subset(df,gender=='M')
dfe=subset(df,gender=='F')
```

#distribution of income by gender who having own land
`histogram(~income|ownland,data=dm,breaks=50)`



```
histogram(~income|ownland,data=dfe,breaks=50)
```



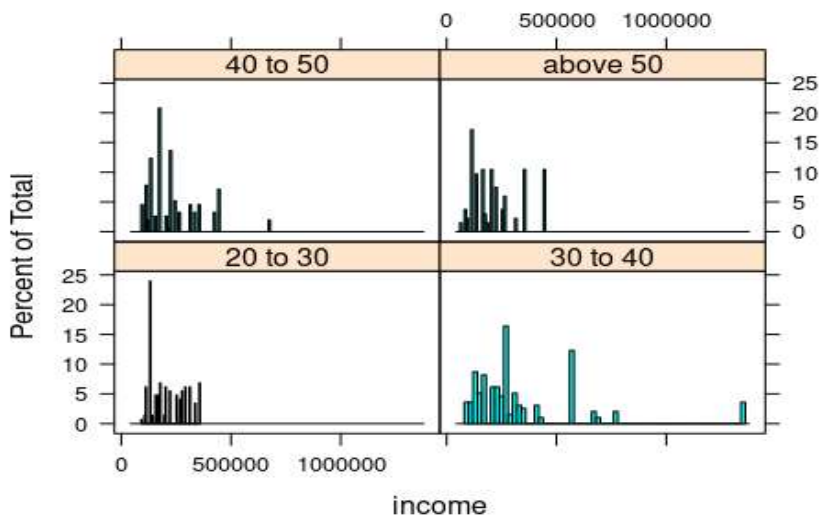
#subsetting ownland by applicant gender

```
d3=subset(dm,ownland=='Y')
```

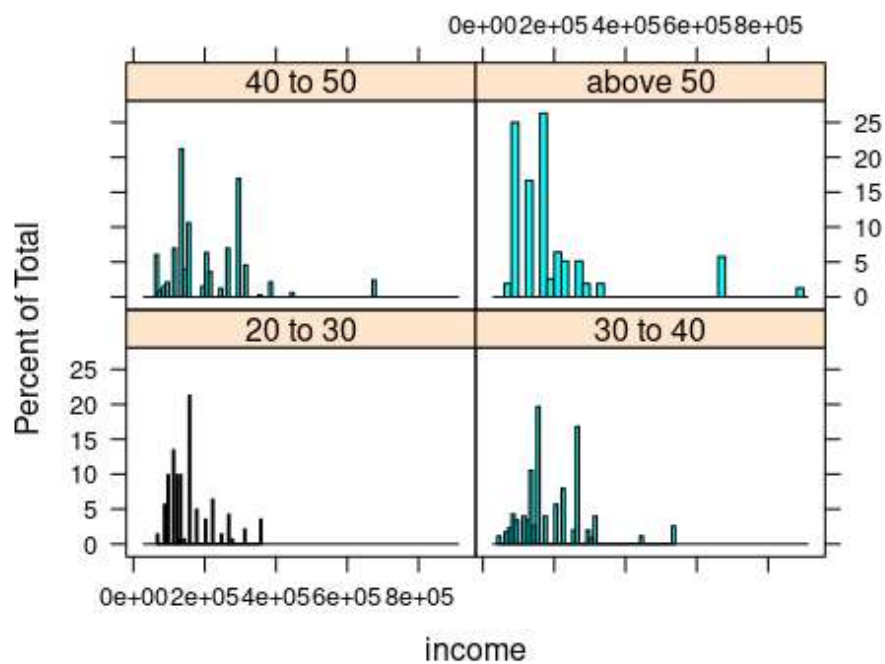
```
d4=subset(dfe,ownland=='Y')
```

#distribbution of income by age range

```
histogram(~income|agerange,data=dm,breaks=50)
```



```
histogram(~income|agerange,data=dfe,breaks=50)
```

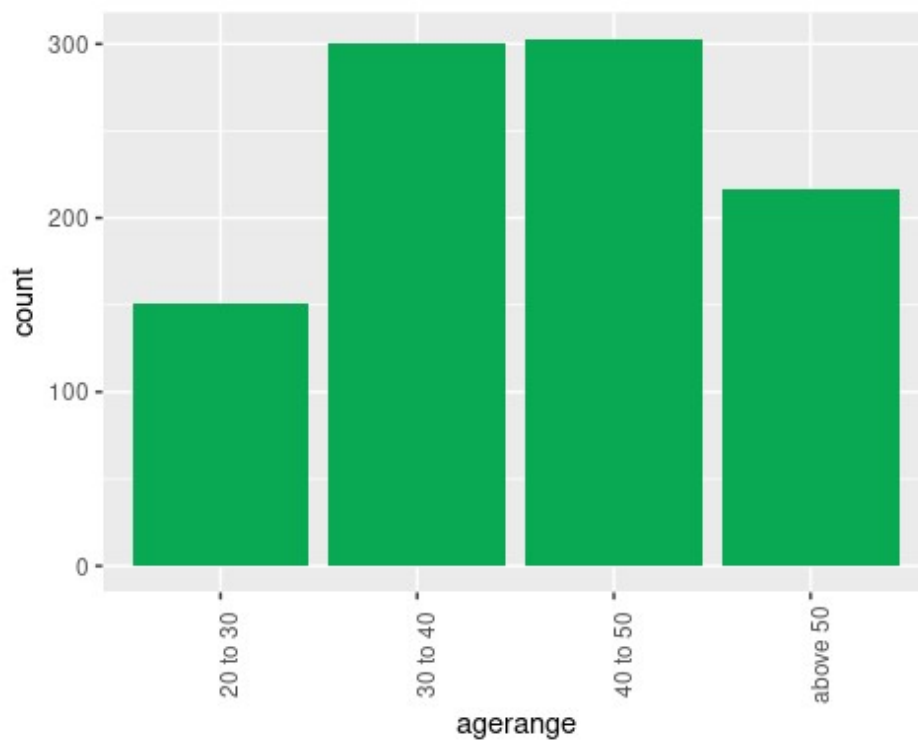


```
df5 = rbind(d3,d4)
dim(df5)
```

```
## [1] 971 15
```

#subsetting age range because of regular income

```
df5 %>% ggplot(aes(agerange))+
  geom_bar(stat="Count",fill="#09a853")+theme(axis.text.x = element_text(angle =
90))
```

```
d6=subset(df5,agerange=='30 to 40')
d7=subset(df5,agerange=='40 to 50')
d8 = rbind(d6,d7)
summary(d8)
```

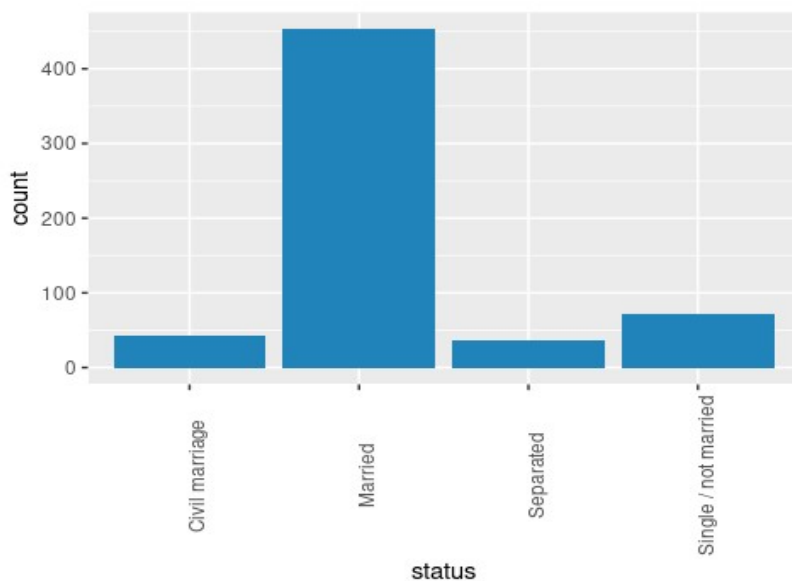
```
##      id                gender          owncar          ownland
## Min.   :1968310      Length:603      Length:603      Length:603
## 1st Qu.:5009104      Class :character    Class :character    Class :character
## Median :5010426      Mode  :character    Mode  :character    Mode  :character
## Mean   :5090131
## 3rd Qu.:5010828
## Max.   :6736968
##      income          incometype          edu          status
## Min.   : 67500      Length:603      Length:603      Length:603
## 1st Qu.:135000      Class :character    Class :character    Class :character
## Median :216000      Mode  :character    Mode  :character    Mode  :character
## Mean   :241381
## 3rd Qu.:297000
## Max.   :765000
##      house          age          exp          costlyphone
## Length:603      Min.   :31.0      Min.   : 0.000      Min.   :0.0000
## Class :character    1st Qu.:36.0      1st Qu.: 3.000      1st Qu.:0.0000
## Mode  :character    Median :41.0      Median : 7.000      Median :0.0000
##                      Mean   :40.5      Mean   : 7.254      Mean   :0.2521
##                      3rd Qu.:44.0      3rd Qu.:10.000     3rd Qu.:1.0000
##                      Max.   :50.0      Max.   :20.000     Max.   :1.0000
```

```
## occupation      familycount      agerange
## Length:603      Min.      :1.000      Length:603
## Class :character 1st Qu.:2.000      Class :character
## Mode :character Median :2.000      Mode :character
##                  Mean  :2.569
##                  3rd Qu.:3.000
##                  Max.  :5.000
```

#subsetting married people coz of both members income
count(d8, 'status')

```
##              status freq
## 1 Civil marriage    42
## 2 Married          453
## 3 Separated         36
## 4 Single / not married 72
```

```
d8 %>% ggplot(aes(status))+
  geom_bar(stat="Count", fill="#2084ba")+theme(axis.text.x = element_text(angle =
90))
```

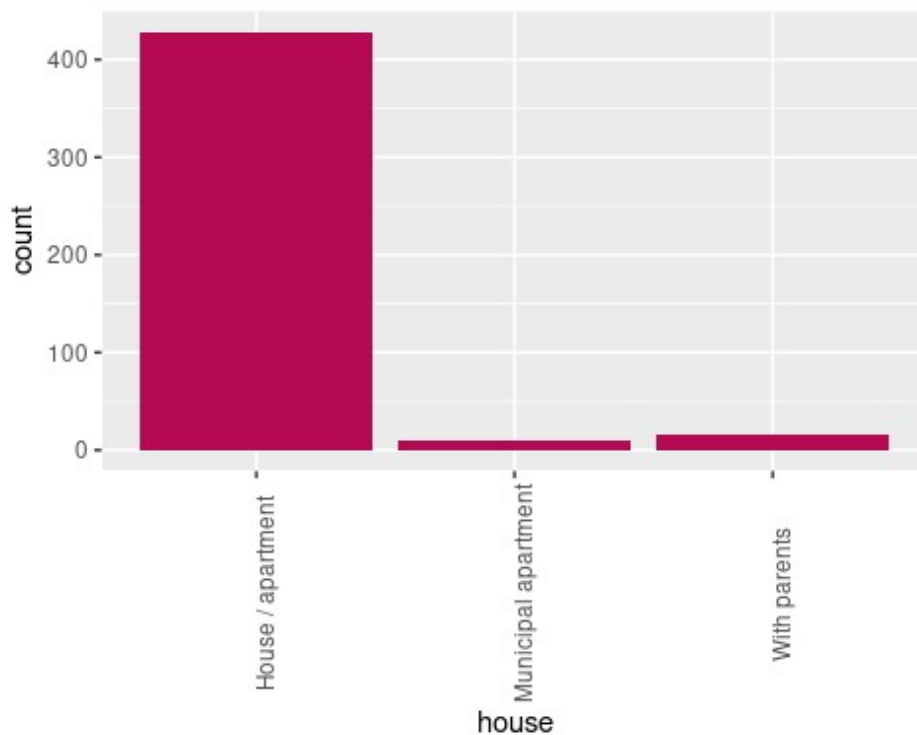


```
d9=subset(d8,status=='Married')
```

#subsetting married people who have own house
count(d9, 'house')

```
##              house freq
## 1 House / apartment 428
## 2 Municipal apartment 9
## 3 With parents      16
```

```
d9 %>% ggplot(aes(house))+
  geom_bar(stat="Count",fill="#b50954")+theme(axis.text.x = element_text(angle =
90))
```



```
d10=subset(d9,house=="House / apartment")
head(d10,5)
```

```
## # A tibble: 5 × 15
##       id gender owncar ownland income incometype edu  status house  age  exp
##   <dbl> <chr>  <chr>  <chr>   <dbl> <chr>    <chr> <chr>  <chr> <dbl> <dbl>
## 1 5008836 M      Y      Y      270000 Working  Seco... Marri... Hous... 35    3
## 2 5008837 M      Y      Y      270000 Working  Seco... Marri... Hous... 35    3
## 3 5008838 M      N      Y      405000 Commercia... High... Marri... Hous... 32    6
## 4 5008839 M      N      Y      405000 Commercia... High... Marri... Hous... 32    6
## 5 5008840 M      N      Y      405000 Commercia... High... Marri... Hous... 32    6
## # ... with 4 more variables: costlyphone <dbl>, occupation <chr>,
## #   familycount <dbl>, agerange <chr>
```

Inference:

- Credit card to be issued to the people who have married and having steady income with own land, own house. From these criteria applicants were chosen.
- Majority of the selected applicants who are laborer which indicates these people are the Blue collar workers who are Middle class peoples
- So, the Middle-class peoples need the Credit card for their future uses so bank's make profit out of them.

Insight:

- From the histogram of salary of the people are mostly right skewed.
- From that salary of the applicants most of those are having own house and own land with approximate median salary of \$1,50,000. From this inference we can issue the card fearlessly.
- The applicants of age 30 to 50 are the most Experienced in their job.
- From the median of family count, we assume approximately most of applicant's family members are 3 persons only.
- From the median of experience of the applicants are approximately around 7 years this indicates they are well settled in their job so the risk of firing the people is low.