Ejemplo-Loan-Analysis.R

franciscodavila

2021-08-14

```
library(gmodels)
library(ggplot2)
```

```
## Registered S3 methods overwritten by 'tibble':
##   method     from
##   format.tbl pillar
##   print.tbl  pillar
```

```
library(tidyr)
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(pROC)
```

```
## Type 'citation("pROC")' for a citation.
```

```
##
## Attaching package: 'pROC'
```

```
## The following object is masked from 'package:gmodels':
##
##     ci
```

```
## The following objects are masked from 'package:stats':
##
##     cov, smooth, var
```

```r
library(knitr)
library(Sim.DiffProc)
```

```
## Package 'Sim.DiffProc', version 4.8
## browseVignettes('Sim.DiffProc') for more informations.
```

```r
library(bazar)
library(scatterplot3d)
library(MASS)
```

```
##
## Attaching package: 'MASS'
```

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

```r
loan_data_ch1 <- readRDS("~/Desktop/R:STATA/R/Admon. Riesgo Financiero/loan_data_ch1.rds")
str(loan_data_ch1)
```

```
## 'data.frame':    29092 obs. of  8 variables:
##  $ loan_status   : int  0 0 0 0 0 0 1 0 1 0 ...
##  $ loan_amnt     : int  5000 2400 10000 5000 3000 12000 9000 3000 10000 1000 ...
##  $ int_rate      : num  10.7 NA 13.5 NA NA ...
##  $ grade         : Factor w/ 7 levels "A","B","C","D",..: 2 3 3 1 5 2 3 2 2 4 ...
##  $ emp_length    : int  10 25 13 3 9 11 0 3 3 0 ...
##  $ home_ownership: Factor w/ 4 levels "MORTGAGE","OTHER",..: 4 4 4 4 4 3 4 4 4 4 ...
##  $ annual_inc    : num  24000 12252 49200 36000 48000 ...
##  $ age           : int  33 31 24 39 24 28 22 22 28 22 ...
```

```r
head(loan_data_ch1)
```

```
##   loan_status loan_amnt int_rate grade emp_length home_ownership annual_inc age
## 1           0      5000    10.65     B         10           RENT      24000  33
## 2           0      2400       NA     C         25           RENT      12252  31
## 3           0     10000    13.49     C         13           RENT      49200  24
## 4           0      5000       NA     A          3           RENT      36000  39
## 5           0      3000       NA     E          9           RENT      48000  24
## 6           0     12000    12.69     B         11            OWN      75000  28
```

```r
CrossTable(loan_data_ch1$home_ownership)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |         N / Table Total |
## |-------------------------|
##
```

```
##
## Total Observations in Table:  29092
##
##
##            |  MORTGAGE |     OTHER |       OWN |      RENT |
##            |-----------|-----------|-----------|-----------|
##            |     12002 |        97 |      2301 |     14692 |
##            |     0.413 |     0.003 |     0.079 |     0.505 |
##            |-----------|-----------|-----------|-----------|
##
##
##
##
```

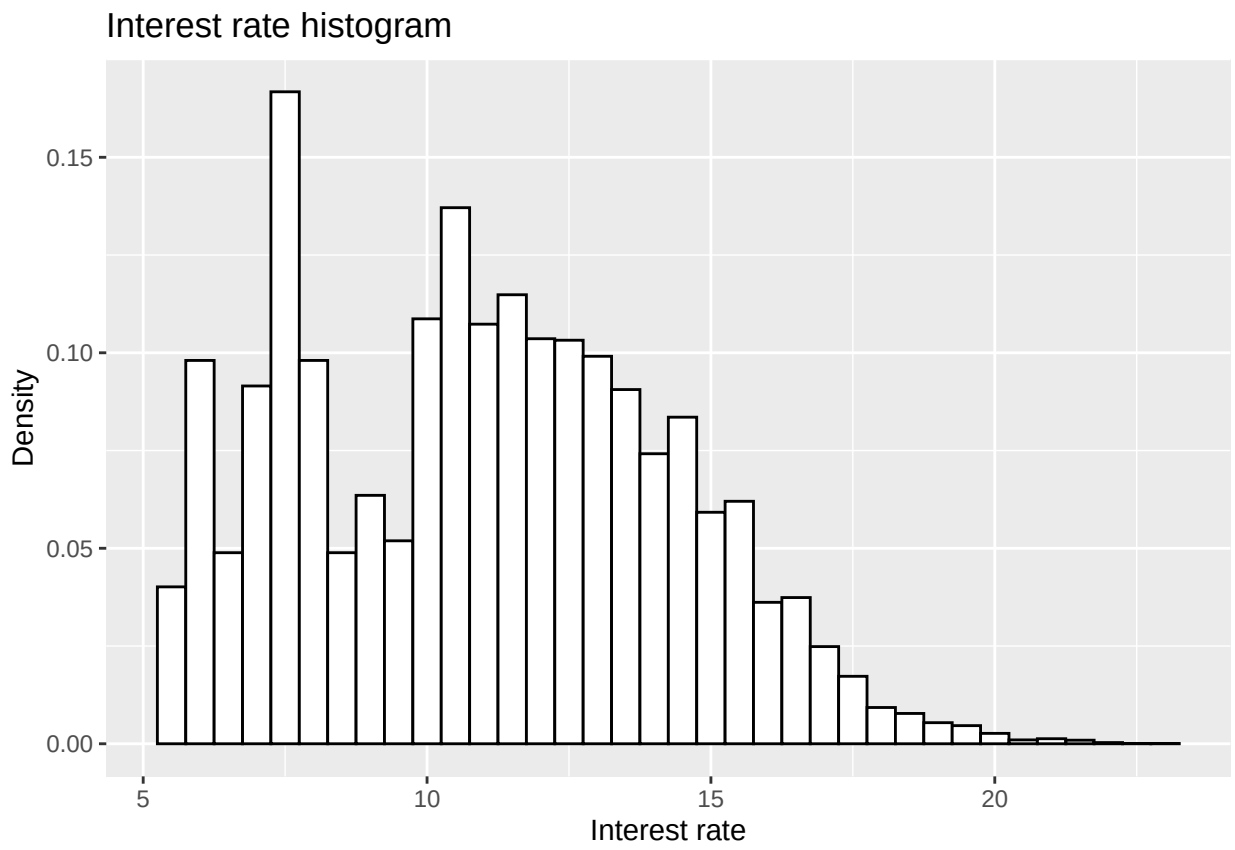
```
#41.3% posee hipoteca, 0.3% tiene otro, 7.9% posee casa y 50.5% renta#
CrossTable(loan_data_ch1$home_ownership, loan_data_ch1$loan_status, prop.r=TRUE, prop.c=FALSE, prop.t=FA
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |           N / Row Total |
## |-------------------------|
##
##
## Total Observations in Table:  29092
##
##
##                              | loan_data_ch1$loan_status
## loan_data_ch1$home_ownership |         0 |         1 | Row Total |
## -----------------------------|-----------|-----------|-----------|
##                     MORTGAGE |     10821 |      1181 |     12002 |
##                              |     0.902 |     0.098 |     0.413 |
## -----------------------------|-----------|-----------|-----------|
##                        OTHER |        80 |        17 |        97 |
##                              |     0.825 |     0.175 |     0.003 |
## -----------------------------|-----------|-----------|-----------|
##                          OWN |      2049 |       252 |      2301 |
##                              |     0.890 |     0.110 |     0.079 |
## -----------------------------|-----------|-----------|-----------|
##                         RENT |     12915 |      1777 |     14692 |
##                              |     0.879 |     0.121 |     0.505 |
## -----------------------------|-----------|-----------|-----------|
##                 Column Total |     25865 |      3227 |     29092 |
## -----------------------------|-----------|-----------|-----------|
##
##
```

```
ggplot(loan_data_ch1,aes(x=int_rate))+
  geom_histogram(aes(y=..density..),binwidth=0.5,colour="black",
                 fill="white")+
  labs(y="Density",
```

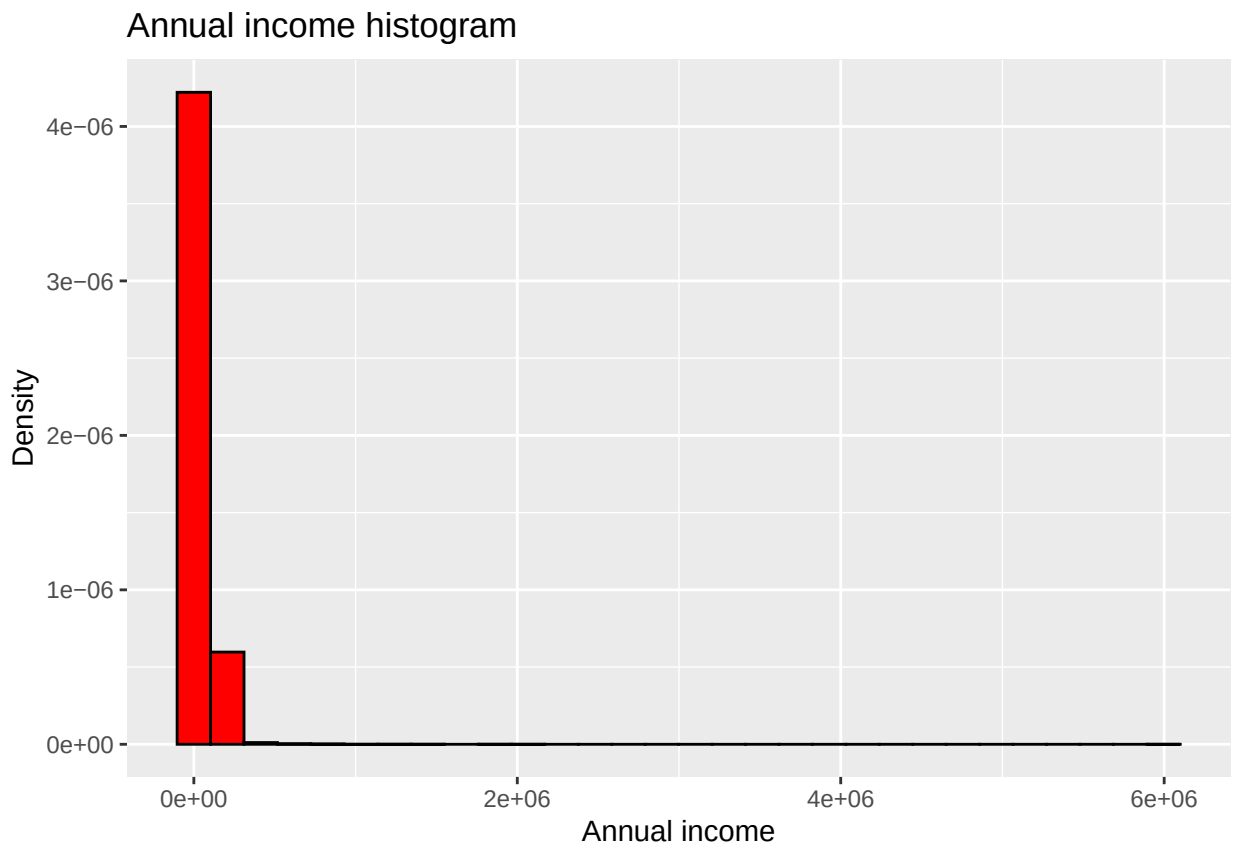
```
       x="Interest rate",
       title="Interest rate histogram",
       subtitle=NULL)+
  theme(legend.position="bottom",legend.title=element_blank())
```

```
## Warning: Removed 2776 rows containing non-finite values (stat_bin).
```



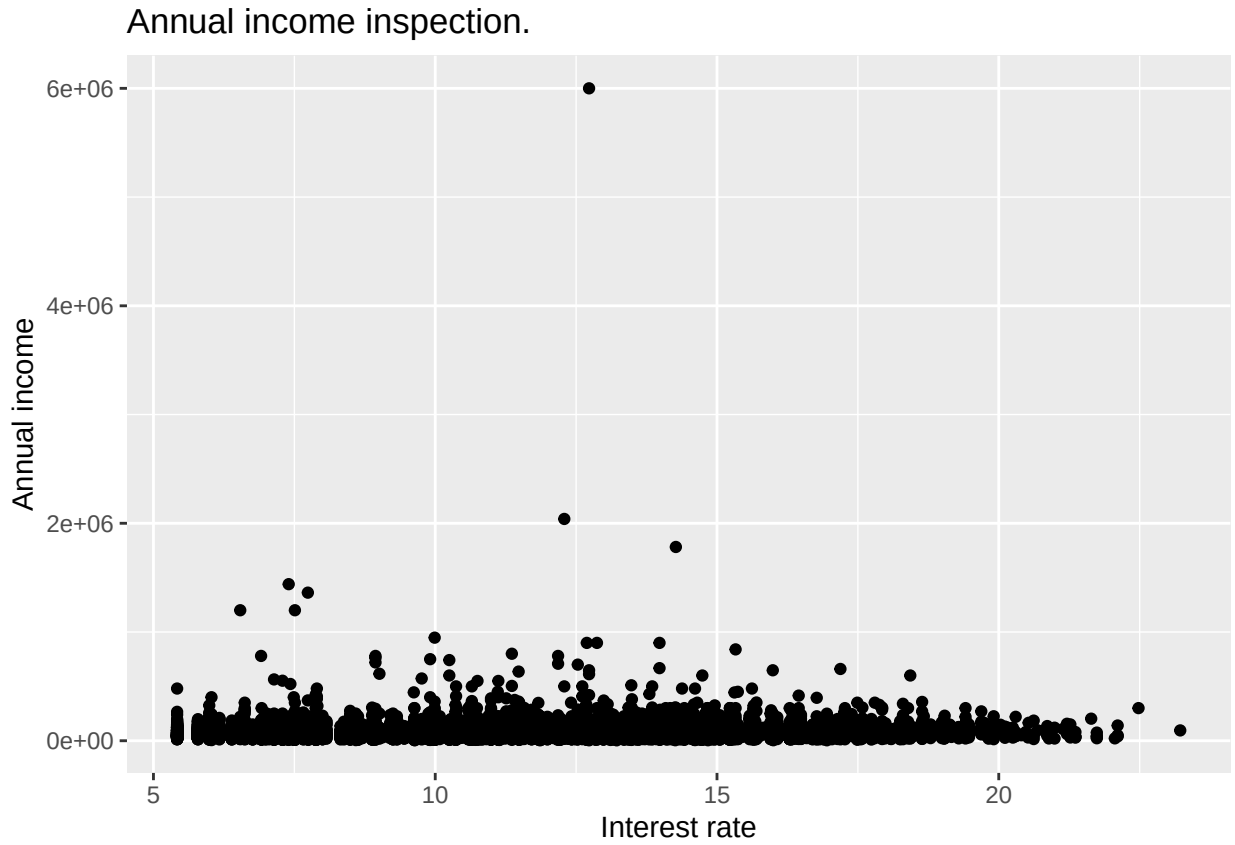

```
ggplot(loan_data_ch1,aes(x=annual_inc))+
  geom_histogram(aes(y=..density..),colour="black",fill="red")+
  labs(y="Density",
       x="Annual income",
       title="Annual income histogram",
       subtitle=NULL)+
  theme(legend.position="bottom",legend.title=element_blank())
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
ggplot(loan_data_ch1, aes(int_rate, annual_inc)) +
  geom_point() +
  labs(y="Annual income",
       x="Interest rate",
       title="Annual income inspection.",
       subtitle=NULL)+
  theme(legend.position="bottom",legend.title=element_blank())
```

```
## Warning: Removed 2776 rows containing missing values (geom_point).
```

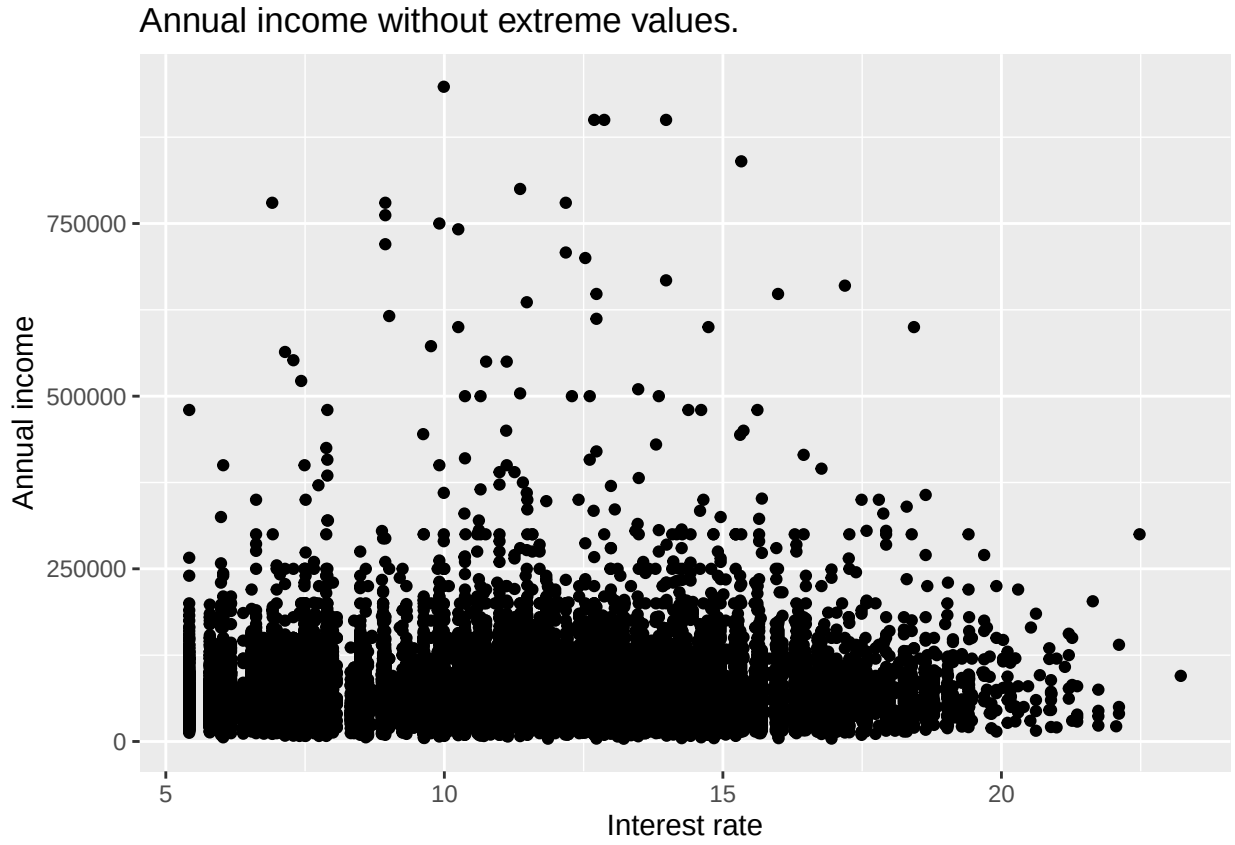



```
high_income <- loan_data_ch1[(loan_data_ch1$annual_inc>1000000),]
high_income
```

```
##       loan_status loan_amnt int_rate grade emp_length home_ownership annual_inc
## 4861            0     12025    14.27     C         13           RENT    1782000
## 13931           0     10000     6.54     A         16            OWN    1200000
## 15386           0      1500       NA     A          5       MORTGAGE    1900000
## 16713           0     12000     7.51     A          1       MORTGAGE    1200000
## 19486           0      5000    12.73     C         12       MORTGAGE    6000000
## 22811           0     10000       NA     A          1       MORTGAGE    1200000
## 23361           0      6400     7.40     A          7       MORTGAGE    1440000
## 23683           0      6600     7.74     A          9       MORTGAGE    1362000
## 28468           0      8450    12.29     C          0           RENT    2039784
##       age
## 4861   63
## 13931  36
## 15386  60
## 16713  32
## 19486 144
## 22811  40
## 23361  44
## 23683  47
## 28468  42
```

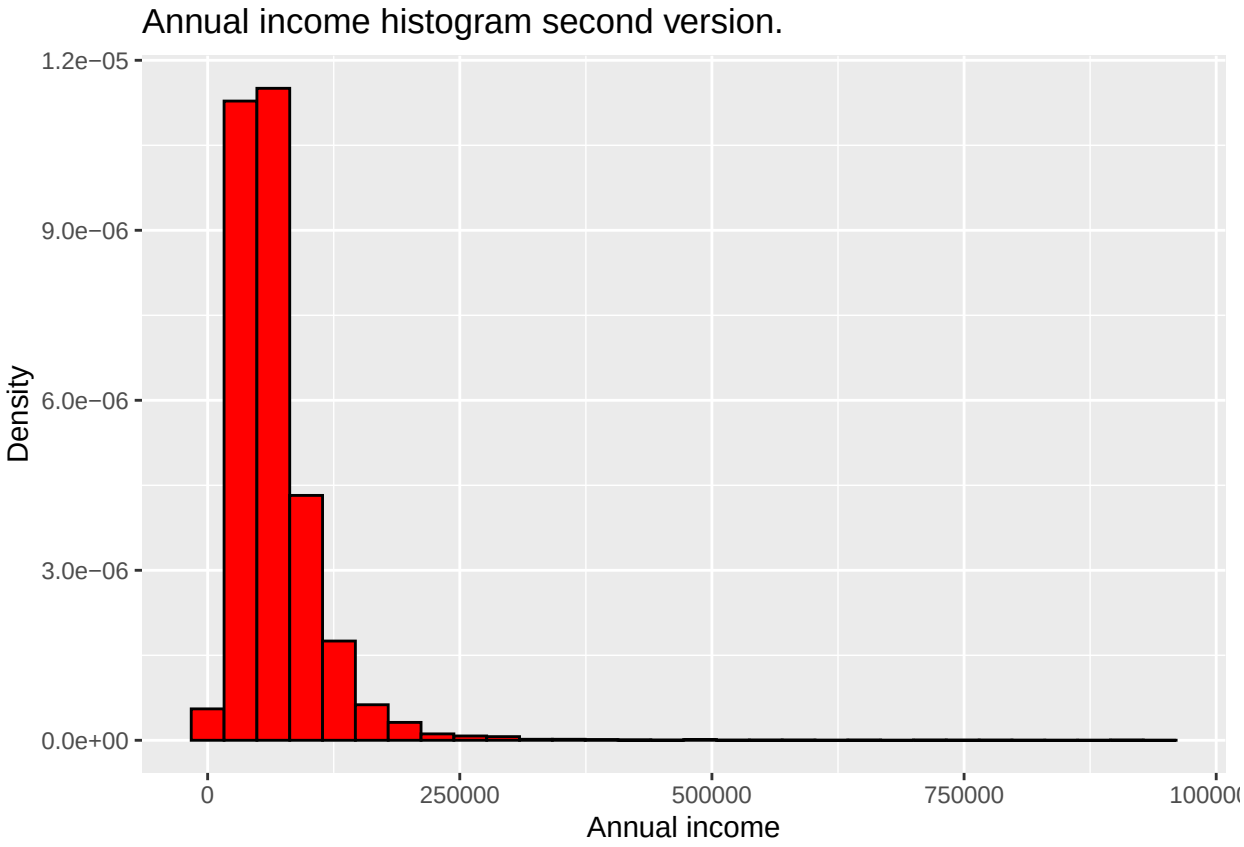
```
high_income_index<-data.frame(value=as.integer(rownames(high_income)))
loan_data_ch1<-loan_data_ch1[-high_income_index$value,]
ggplot(loan_data_ch1,aes(int_rate,annual_inc))+
  geom_point()+
  labs(y="Annual income",
       x="Interest rate",
       title="Annual income without extreme values.",
       subtitle=NULL)+
  theme(legend.position="bottom",legend.title=element_blank())
```

```
## Warning: Removed 2774 rows containing missing values (geom_point).
```




```
ggplot(loan_data_ch1,aes(x=annual_inc))+
  geom_histogram(aes(y=..density..),colour="black",fill="red")+
  labs(y="Density",
       x="Annual income",
       title="Annual income histogram second version.",
       subtitle=NULL)+
  theme(legend.position="bottom",legend.title=element_blank())
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```


Annual income histogram second version.
1.2e−05
9.0e−06
6.0e−06
3.0e−06
0.0e+00
Density
0
250000
500000
750000
Annual income