Bank Marketing

ELavarasan

2023-02-19

```
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(lattice)
library(ggplot2)
df<-read.csv("bank.csv")
df

##       age           job  marital education default balance housing loan
## contact
## 1     58    management  married   tertiary      no    2143     yes   no
## unknown
## 2     44    technician   single  secondary      no      29     yes   no
## unknown
## 3     33  entrepreneur  married  secondary      no       2     yes  yes
## unknown
## 4     47   blue-collar  married    unknown      no    1506     yes   no
## unknown
## 5     33       unknown   single    unknown      no       1      no   no
## unknown
## 6     35    management  married   tertiary      no     231     yes   no
## unknown
## 7     28    management   single   tertiary      no     447     yes  yes
## unknown
## 8     42  entrepreneur divorced   tertiary     yes       2     yes   no
## unknown
## 9     58       retired  married    primary      no     121     yes   no
## unknown
## 10    43    technician   single  secondary      no     593     yes   no
## unknown
## 11    41        admin. divorced  secondary      no     270     yes   no
## unknown
## 12    29        admin.   single  secondary      no     390     yes   no
```

```
unknown
## 13    53    technician  married secondary       no      6     yes   no
unknown
## 14    58    technician  married   unknown        no     71     yes   no
unknown
## 15    57      services  married secondary        no    162     yes   no
unknown
## 16    51       retired  married   primary        no    229     yes   no
unknown
## 17    45        admin.   single   unknown        no     13     yes   no
unknown
## 18    57   blue-collar  married   primary        no     52     yes   no
unknown
## 19    60       retired  married   primary        no     60     yes   no
unknown
## 20    33      services  married secondary        no      0     yes   no

##       day month duration campaign pdays previous poutcome    y
## 1       5   may      261        1    -1        0  unknown   no
## 2       5   may      151        1    -1        0  unknown   no
## 3       5   may       76        1    -1        0  unknown   no
## 4       5   may       92        1    -1        0  unknown   no
## 5       5   may      198        1    -1        0  unknown   no
## 6       5   may      139        1    -1        0  unknown   no
## 7       5   may      217        1    -1        0  unknown   no
## 8       5   may      380        1    -1        0  unknown   no
## 9       5   may       50        1    -1        0  unknown   no
## 10      5   may       55        1    -1        0  unknown   no
## 11      5   may      222        1    -1        0  unknown   no
## 12      5   may      137        1    -1        0  unknown   no
## 13      5   may      517        1    -1        0  unknown   no
## 14      5   may       71        1    -1        0  unknown   no
## 15      5   may      174        1    -1        0  unknown   no
## 16      5   may      353        1    -1        0  unknown   no
## 17      5   may       98        1    -1        0  unknown   no
## 18      5   may       38        1    -1        0  unknown   no
## 19      5   may      219        1    -1        0  unknown   no
## 20      5   may       54        1    -1        0  unknown   no
```

```
filter(df,age<=30)
```

```
##       age          job  marital education default balance housing loan
## 1      28   management   single  tertiary      no     447     yes  yes
## 2      29       admin.   single secondary      no     390     yes   no
## 3      28  blue-collar  married secondary      no     723     yes  yes
## 4      25     services  married secondary      no      50     yes   no
## 5      25  blue-collar  married secondary      no      -7     yes   no
## 6      29   management   single  tertiary      no       0     yes   no
## 7      24   technician   single secondary      no    -103     yes  yes
```

```
## 8     30    technician  married secondary        no      152      yes  yes
## 9     27    technician   single  tertiary        no        0      yes   no
## 10    29      services divorced secondary        no       31      yes   no
## 11    29        admin.   single secondary        no      818      yes  yes
## 12    28    unemployed   single  tertiary        no        0      yes   no
## 13    23   blue-collar  married secondary        no       94      yes   no
## 14    26       student   single secondary        no        0      yes   no
## 15    26        admin.   single secondary        no       82      yes   no
## 16    28   blue-collar  married   primary        no      324      yes   no
## 17    22   blue-collar   single secondary        no        0      yes   no
## 18    30    technician   single secondary        no     2573      yes   no
## 19    30       student   single  tertiary        no       25      yes   no
## 20    30   blue-collar   single secondary        no        0      yes   no


##          contact day month duration campaign pdays previous poutcome    y
## 1        unknown   5   may      217        1    -1        0  unknown   no
## 2        unknown   5   may      137        1    -1        0  unknown   no
## 3        unknown   5   may      262        1    -1        0  unknown   no
## 4        unknown   5   may      342        1    -1        0  unknown   no
## 5        unknown   5   may      365        1    -1        0  unknown   no
## 6        unknown   5   may      363        1    -1        0  unknown   no
## 7        unknown   5   may      145        1    -1        0  unknown   no
## 8        unknown   5   may      213        2    -1        0  unknown   no
## 9        unknown   5   may      400        1    -1        0  unknown   no
## 10       unknown   5   may      194        1    -1        0  unknown   no
## 11       unknown   5   may       61        1    -1        0  unknown   no
## 12       unknown   5   may      125        2    -1        0  unknown   no
## 13       unknown   5   may      193        1    -1        0  unknown   no
## 14       unknown   5   may      610        2    -1        0  unknown   no
## 15       unknown   5   may      228        1    -1        0  unknown   no
## 16       unknown   5   may      175        1    -1        0  unknown   no
## 17       unknown   5   may      179        2    -1        0  unknown   no
## 18       unknown   5   may       67        2    -1        0  unknown   no
## 19       unknown   5   may       89        2    -1        0  unknown   no
## 20       unknown   5   may        5        2    -1        0  unknown   no

##      day month duration campaign pdays previous poutcome    y
## 1      5   may      261        1    -1        0  unknown   no
## 2      5   may      139        1    -1        0  unknown   no
## 3      5   may      217        1    -1        0  unknown   no
## 4      5   may      164        1    -1        0  unknown   no
## 5      5   may      296        1    -1        0  unknown   no
## 6      5   may      255        2    -1        0  unknown   no
## 7      5   may      230        1    -1        0  unknown   no
## 8      5   may      336        1    -1        0  unknown   no
## 9      5   may      180        2    -1        0  unknown   no
## 10     5   may      363        1    -1        0  unknown   no
## 11     5   may      253        1    -1        0  unknown   no
## 12     5   may      179        1    -1        0  unknown   no
```

```
## 13       5   may         104           1    -1         0  unknown  no
## 14       5   may          91           1    -1         0  unknown  no
## 15       5   may         273           1    -1         0  unknown  no
## 16       5   may         154           1    -1         0  unknown  no
## 17       5   may         188           2    -1         0  unknown  no
## 18       5   may          48           1    -1         0  unknown  no
## 19       5   may         426           1    -1         0  unknown  no
## 20       5   may         197           1    -1         0  unknown  no
```

```r
filter(df,job=="technician" )
```

```
##       age         job  marital education default balance housing loan
contact
## 1     44 technician   single secondary      no      29     yes   no
unknown
## 2     43 technician   single secondary      no     593     yes   no
unknown
## 3     53 technician  married secondary      no       6     yes   no
unknown
## 4     58 technician  married   unknown      no      71     yes   no
unknown
## 5     36 technician   single secondary      no     265     yes  yes
unknown
## 6     57 technician  married secondary      no     839      no  yes
unknown
## 7     57 technician divorced secondary      no      63     yes   no
unknown
## 8     53 technician  married secondary      no      -3      no   no
unknown
## 9     44 technician  married secondary      no       0     yes   no
unknown
## 10    55 technician divorced secondary      no       0      no   no
unknown
## 11    24 technician   single secondary      no    -103     yes  yes
unknown
## 12    53 technician divorced secondary      no     989     yes   no
unknown
## 13    55 technician  married secondary      no    1205     yes   no
unknown
## 14    43 technician  married secondary      no    1937     yes   no
unknown
## 15    53 technician  married secondary      no     384     yes   no
unknown
## 16    42 technician   single secondary      no     690     yes   no
unknown
## 17    41 technician  married secondary      no    1270     yes   no
unknown
```

```
## 18      30 technician  married secondary          no      152    yes yes
unknown
## 19      39 technician  married secondary          no        0    yes   no
unknown
## 20      27 technician   single  tertiary          no        0    yes   no

##      day month duration campaign pdays previous poutcome   y
## 1      5   may      151        1    -1        0  unknown  no
## 2      5   may       55        1    -1        0  unknown  no
## 3      5   may      517        1    -1        0  unknown  no
## 4      5   may       71        1    -1        0  unknown  no
## 5      5   may      348        1    -1        0  unknown  no
## 6      5   may      225        1    -1        0  unknown  no
## 7      5   may      242        1    -1        0  unknown  no
## 8      5   may     1666        1    -1        0  unknown  no
## 9      5   may      225        2    -1        0  unknown  no
## 10     5   may      160        1    -1        0  unknown  no
## 11     5   may      145        1    -1        0  unknown  no
## 12     5   may      812        1    -1        0  unknown  no
## 13     5   may      158        2    -1        0  unknown  no
## 14     5   may      181        1    -1        0  unknown  no
## 15     5   may      176        1    -1        0  unknown  no
## 16     5   may       20        1    -1        0  unknown  no
## 17     5   may     1389        1    -1        0  unknown yes
## 18     5   may      213        2    -1        0  unknown  no
## 19     5   may      203        1    -1        0  unknown  no
## 20     5   may      400        1    -1        0  unknown  no
```

```
filter(df,job=="entrepreneur")
```

```
##      age          job  marital education default balance housing loan
contact
## 1      33 entrepreneur  married secondary      no       2     yes  yes
unknown
## 2      42 entrepreneur divorced  tertiary     yes       2     yes   no
unknown
## 3      52 entrepreneur  married secondary      no     113     yes  yes
unknown
## 4      38 entrepreneur   single  tertiary      no     243      no  yes
unknown
## 5      57 entrepreneur divorced secondary      no     -37      no   no
unknown
## 6      53 entrepreneur  married   unknown      no      19     yes   no
unknown
## 7      35 entrepreneur divorced secondary      no     -69     yes   no
unknown
## 8      57 entrepreneur  married secondary      no     294     yes   no
```

```
unknown
## 9     43 entrepreneur   married    unknown     no    100     yes   no
unknown
## 10    59 entrepreneur divorced secondary     no    901     yes   no
unknown
## 11    41 entrepreneur   married    primary     no    236     yes   no
unknown
## 12    56 entrepreneur divorced secondary     no    155      no   no
unknown
## 13    33 entrepreneur    single   tertiary     no   1068     yes   no
unknown
## 14    36 entrepreneur   married   tertiary     no   1169     yes   no
unknown
## 15    49 entrepreneur   married secondary     no    167     yes  yes
unknown
## 16    35 entrepreneur   married secondary    yes    204     yes   no
unknown
## 17    53 entrepreneur   married secondary     no    288      no   no
unknown
## 18    52 entrepreneur   married   tertiary     no   -184     yes  yes
unknown
## 19    38 entrepreneur   married   tertiary     no    898     yes   no
unknown
## 20    41 entrepreneur   married    unknown     no     89     yes   no

##       day month duration campaign pdays previous poutcome   y
## 1       5   may       76        1    -1        0  unknown  no
## 2       5   may      380        1    -1        0  unknown  no
## 3       5   may      127        1    -1        0  unknown  no
## 4       5   may      174        1    -1        0  unknown  no
## 5       5   may      173        1    -1        0  unknown  no
## 6       5   may      252        1    -1        0  unknown  no
## 7       5   may      300        1    -1        0  unknown  no
## 8       5   may      746        2    -1        0  unknown  no
## 9       5   may      702        1    -1        0  unknown  no
## 10      5   may       73        3    -1        0  unknown  no
## 11      5   may      151        1    -1        0  unknown  no
## 12      5   may      118        3    -1        0  unknown  no
## 13      5   may      396        2    -1        0  unknown  no
## 14      6   may      168        2    -1        0  unknown  no
## 15      6   may      198        3    -1        0  unknown  no
## 16      6   may      424        2    -1        0  unknown  no
## 17      6   may       69        1    -1        0  unknown  no
## 18      6   may      392        2    -1        0  unknown  no
## 19      6   may      132        2    -1        0  unknown  no
## 20      6   may      333        2    -1        0  unknown  no
```

```
filter(df,job=="blue-collar")
```

```
##      age          job  marital education default balance housing loan
contact
## 1    47 blue-collar  married   unknown      no    1506     yes   no
unknown
## 2    57 blue-collar  married   primary      no      52     yes   no
unknown
## 3    28 blue-collar  married secondary      no     723     yes  yes
unknown
## 4    32 blue-collar   single   primary      no      23     yes  yes
unknown
## 5    59 blue-collar  married secondary      no       0     yes   no
unknown
## 6    25 blue-collar  married secondary      no      -7     yes   no
unknown
## 7    60 blue-collar  married   unknown      no     104     yes   no
unknown
## 8    54 blue-collar  married secondary      no    1291     yes   no
unknown
## 9    47 blue-collar  married   unknown      no     306     yes   no
unknown
## 10   40 blue-collar   single   unknown      no      24     yes   no
unknown
## 11   57 blue-collar  married   primary      no     249     yes   no
unknown
## 12   49 blue-collar  married   unknown      no     154     yes   no
unknown
## 13   35 blue-collar   single secondary      no   12223     yes  yes
unknown
## 14   57 blue-collar  married secondary      no    5935     yes  yes
unknown
## 15   55 blue-collar  married   primary      no      23     yes   no
unknown
## 16   44 blue-collar  married secondary      no     582      no  yes
unknown
## 17   51 blue-collar  married   primary      no     173     yes   no
unknown
## 18   60 blue-collar  married   unknown      no      54     yes   no
unknown
## 19   55 blue-collar  married secondary      no     383      no   no
unknown
## 20   44 blue-collar  married secondary      no    -674     yes   no

##      day month duration campaign pdays previous poutcome   y
## 1      5   may       92        1    -1        0  unknown  no
## 2      5   may       38        1    -1        0  unknown  no
## 3      5   may      262        1    -1        0  unknown  no
## 4      5   may      160        1    -1        0  unknown  no
## 5      5   may      226        1    -1        0  unknown  no
## 6      5   may      365        1    -1        0  unknown  no
## 7      5   may       22        1    -1        0  unknown  no
```

```
## 8        5    may       266        1    -1         0    unknown   no
## 9        5    may        13        1    -1         0    unknown   no
## 10       5    may       185        1    -1         0    unknown   no
## 11       5    may       164        1    -1         0    unknown   no
## 12       5    may       357        1    -1         0    unknown   no
## 13       5    may       177        1    -1         0    unknown   no
## 14       5    may       258        1    -1         0    unknown   no
## 15       5    may       291        1    -1         0    unknown   no
## 16       5    may       211        1    -1         0    unknown   no
## 17       5    may       529        2    -1         0    unknown   no
## 18       5    may       221        1    -1         0    unknown   no
## 19       5    may       287        1    -1         0    unknown   no
## 20       5    may       257        1    -1         0    unknown   no
```

```
str(df)
```

```
## 'data.frame':    45211 obs. of  17 variables:
##  $ age      : int  58 44 33 47 33 35 28 42 58 43 ...
##  $ job      : chr  "management" "technician" "entrepreneur" "blue-collar"
...
##  $ marital  : chr  "married" "single" "married" "married" ...
##  $ education: chr  "tertiary" "secondary" "secondary" "unknown" ...
##  $ default  : chr  "no" "no" "no" "no" ...
##  $ balance  : int  2143 29 2 1506 1 231 447 2 121 593 ...
##  $ housing  : chr  "yes" "yes" "yes" "yes" ...
##  $ loan     : chr  "no" "no" "yes" "no" ...
##  $ contact  : chr  "unknown" "unknown" "unknown" "unknown" ...
##  $ day      : int  5 5 5 5 5 5 5 5 5 5 ...
##  $ month    : chr  "may" "may" "may" "may" ...
##  $ duration : int  261 151 76 92 198 139 217 380 50 55 ...
##  $ campaign : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ pdays    : int  -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 ...
##  $ previous : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ poutcome : chr  "unknown" "unknown" "unknown" "unknown" ...
##  $ y        : chr  "no" "no" "no" "no" ...
```

```
summary(df)
```

```
##       age             job              marital            education
##  Min.   :18.00   Length:45211       Length:45211       Length:45211
##  1st Qu.:33.00   Class :character   Class :character   Class :character
##  Median :39.00   Mode  :character   Mode  :character   Mode  :character
##  Mean   :40.94
##  3rd Qu.:48.00
##  Max.   :95.00
##    default             balance          housing              loan
##  Length:45211       Min.   : -8019   Length:45211       Length:45211
##  Class :character   1st Qu.:    72   Class :character   Class :character
##  Mode  :character   Median :   448   Mode  :character   Mode  :character
##                     Mean   :  1362
```

```
##                        3rd Qu.:  1428
##                        Max.   :102127
##     contact               day             month               duration
##  Length:45211        Min.   : 1.00    Length:45211        Min.   :   0.0
##  Class :character    1st Qu.: 8.00    Class :character    1st Qu.: 103.0
##  Mode  :character    Median :16.00    Mode  :character    Median : 180.0
##                      Mean   :15.81                        Mean   : 258.2
##                      3rd Qu.:21.00                        3rd Qu.: 319.0
##                      Max.   :31.00                        Max.   :4918.0
##     campaign            pdays             previous            poutcome
##  Min.   : 1.000    Min.   : -1.0    Min.   :  0.0000    Length:45211
##  1st Qu.: 1.000    1st Qu.: -1.0    1st Qu.:  0.0000    Class :character
##  Median : 2.000    Median : -1.0    Median :  0.0000    Mode  :character
##  Mean   : 2.764    Mean   : 40.2    Mean   :  0.5803
##  3rd Qu.: 3.000    3rd Qu.: -1.0    3rd Qu.:  0.0000
##  Max.   :63.000    Max.   :871.0    Max.   :275.0000
##       y
##  Length:45211
##  Class :character
##  Mode  :character
##
##
##
```

```r
sample_index=sample(1:nrow(df),20,replace = FALSE)
sample_index
```

```
##  [1] 11762 35161 11737 32192 23307 36291 41704 29132 14636 23647  6686
## 36438
## [13] 27152 11273  9101 26244  3456 10427 40865 42259
```

```r
str(df$age)
```

```
##  int [1:45211] 58 44 33 47 33 35 28 42 58 43 ...
```

```r
str(df$job)
```

```
##  chr [1:45211] "management" "technician" "entrepreneur" "blue-collar" ...
```

```r
str(df$marital)
```

```
##  chr [1:45211] "married" "single" "married" "married" "single" "married"
## ...
```

```r
str(df$education)
```

```
##  chr [1:45211] "tertiary" "secondary" "secondary" "unknown" "unknown" ...
```

```r
str(df$y)
```

```
##  chr [1:45211] "no" "no" "no" "no" "no" "no" "no" "no" "no" "no" "no" "no"
## ...
```

```
s1=select(df,2)
levels(s1)

## NULL

t=table(s1)
t

## job
##         admin.    blue-collar   entrepreneur        housemaid      management
##           5171           9732           1487            1240            9458
##        retired  self-employed       services         student       technician
##           2264           1579           4154             938            7597
##     unemployed        unknown
##           1303            288

plot(t)
```
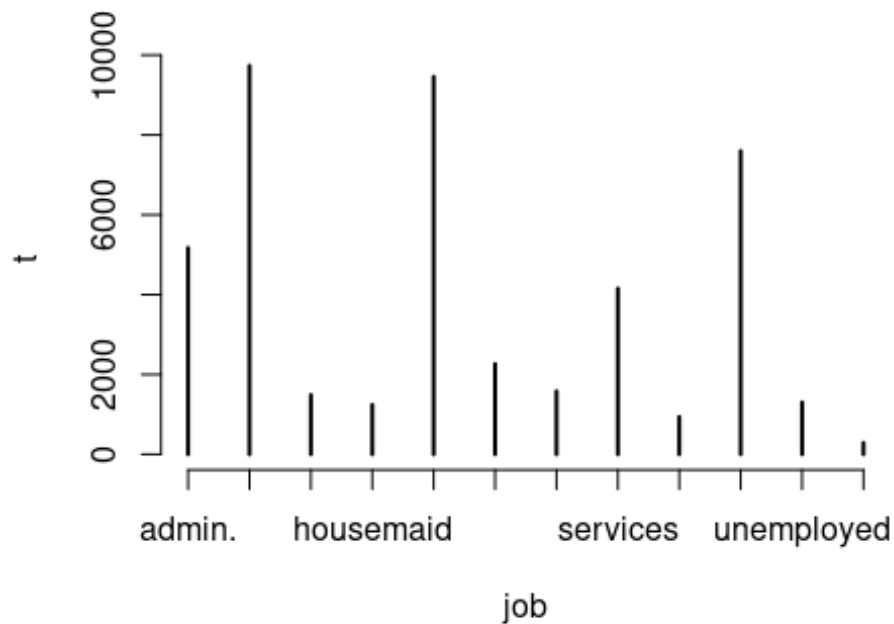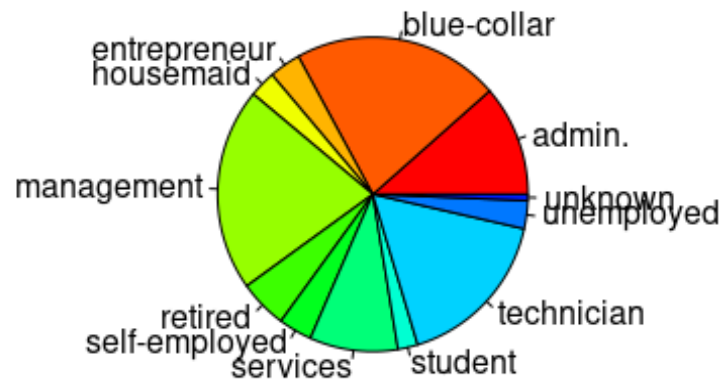


```
a=table(df$job)
b=table(df$education)
c=table(df$marital)
pie(a,main = "job status for pie chart",col = rainbow(length(df)))
```
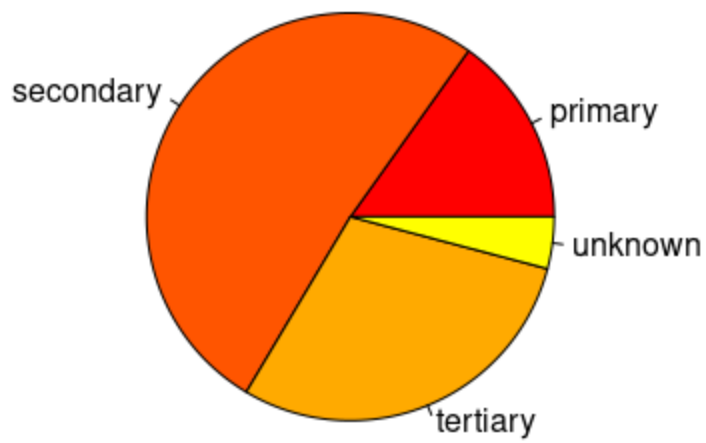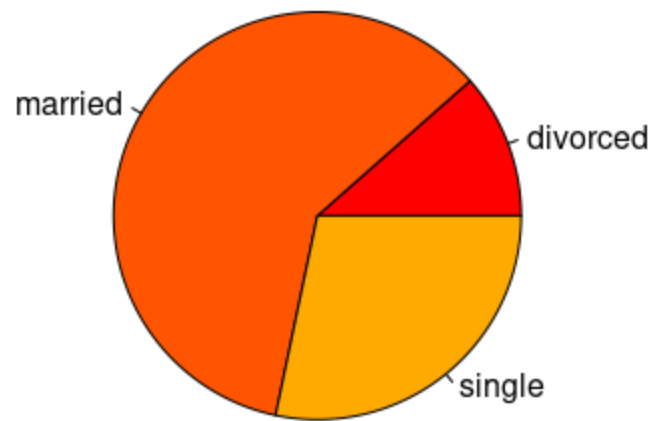
job status for pie chart


```r
pie(b,main = "education status for pie chart",col = rainbow(length(df)))
```

education status for pie chart


```
pie(c,main = "marital status for pie chart",col = rainbow(length(df)))
```
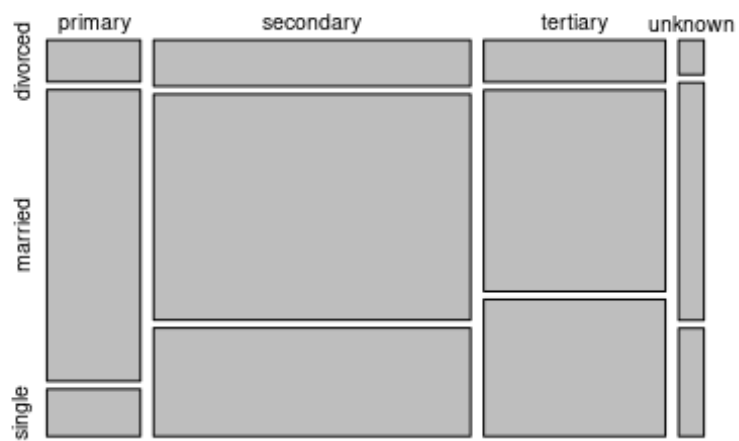
marital status for pie chart


```
tablem=table(df$education,df$marital)
tablem
```

```
##
##             divorced married single
##    primary       752    5246    853
##    secondary    2815   13770   6617
##    tertiary     1471    7038   4792
##    unknown       169    1160    528
```
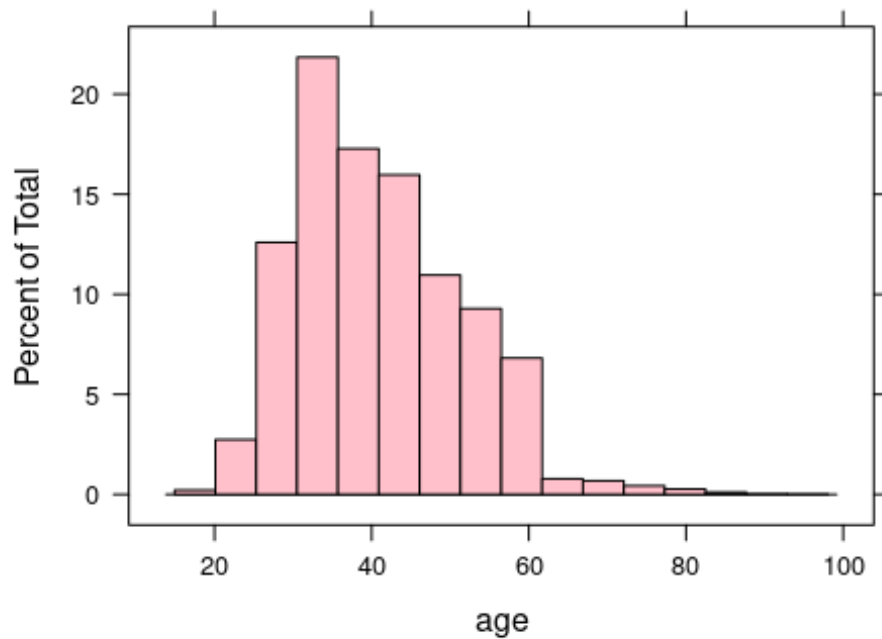
```
mosaicplot(tablem)
```

tablem

#histogram
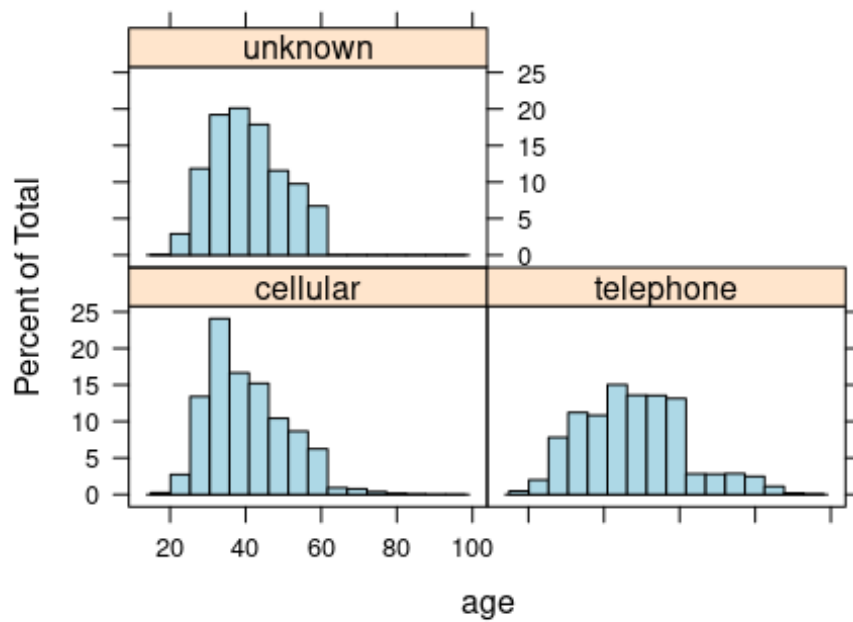
```
histogram(~age,data =df,main = "age distribution level",col = "pink")
```

age distribution level

```
histogram(~age|contact,data =df,main = "communication with people",col =
"lightblue")
```


communication with people

```
histogram(~age|marital,data = df,main = "age by marital status")
```

age by marital status
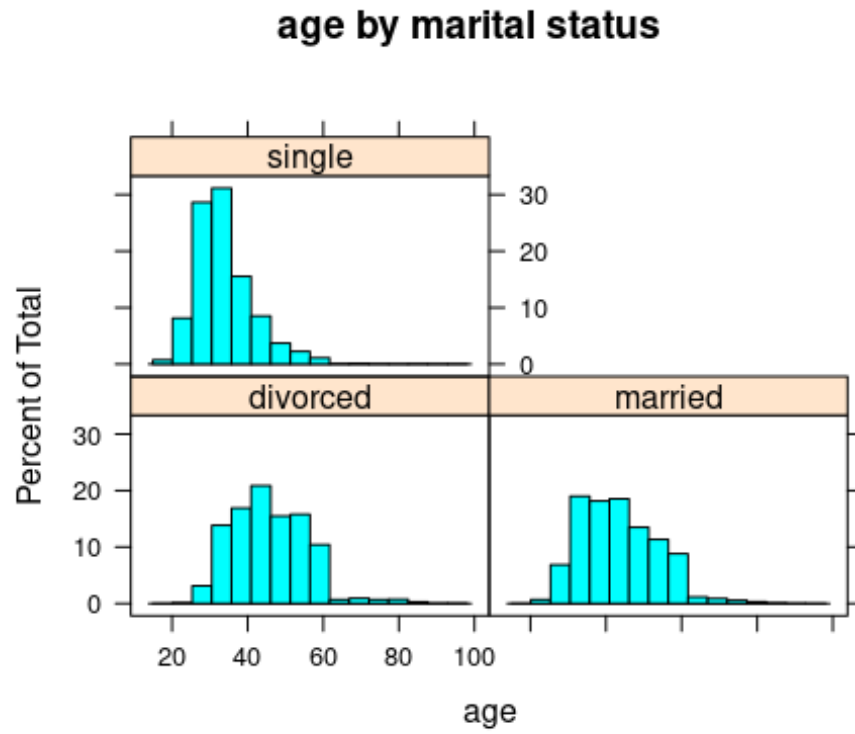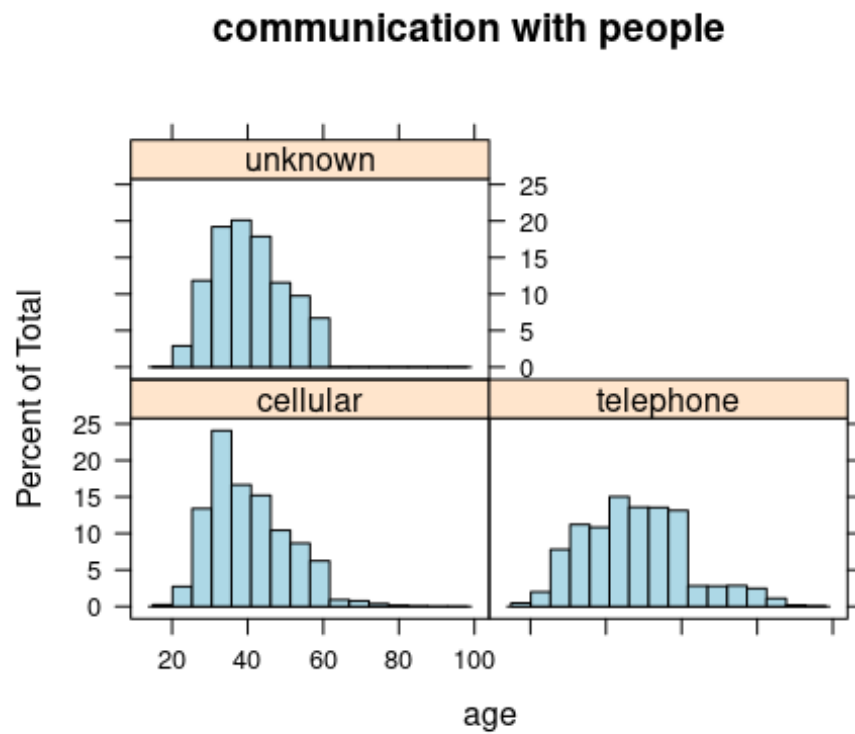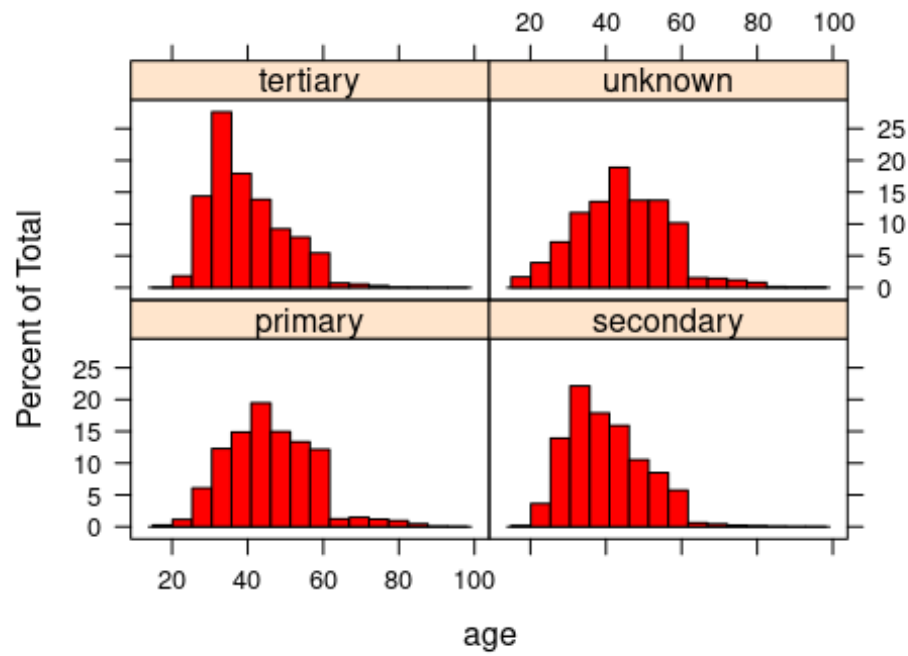
```
histogram(~age|contact,data =df,main = "communication with people",col =
"lightblue")
```
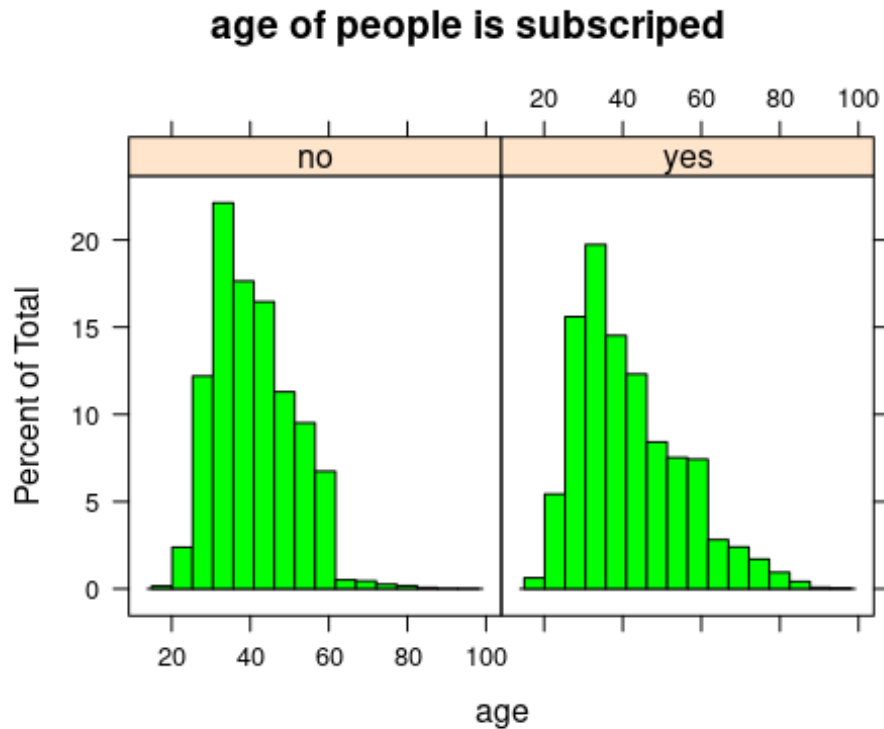
communication with people


```
histogram(~age|education,data =df,main = "age by education level",col =
"red")
```

age by education level

```
histogram(~age|y,data =df,main = "age of people is subscriped",col = "green")
```


age of people is subscriped

```
#box plot

ggplot(df,aes(x=education,y=age))+geom_boxplot()+
facet_wrap(~contact,scales = "free_y")+
labs(x="Education",y="age",title = "age by education level")+
theme(axis.text.x = element_text(angle = 90))
```

age by education level
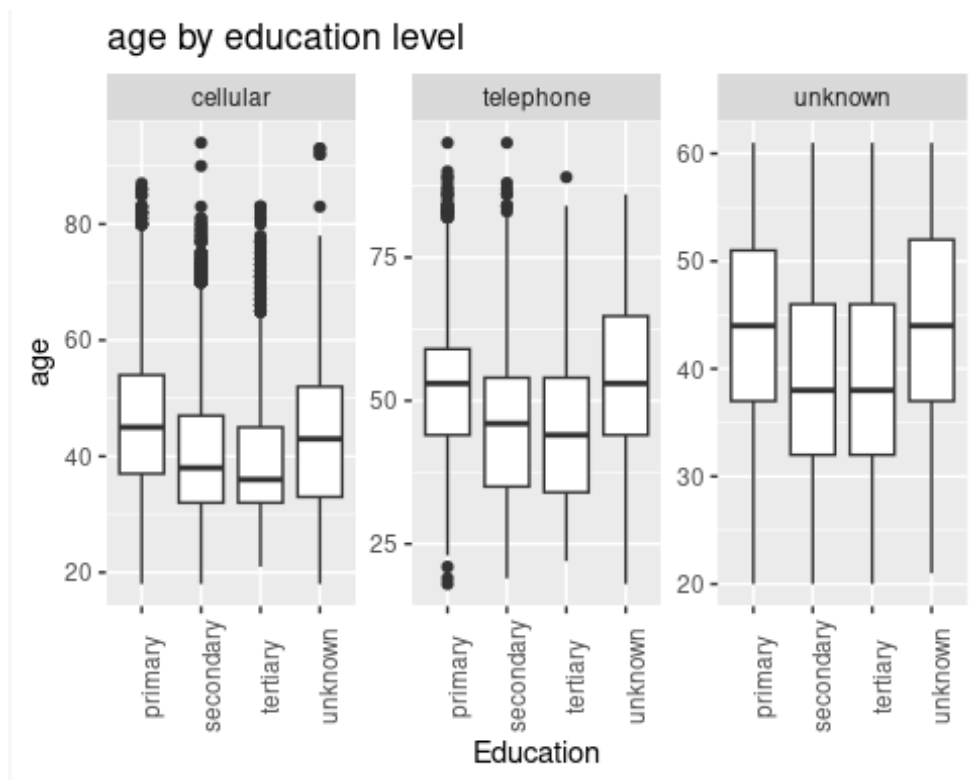
```
ggplot(df,aes(x=y,y=age))+geom_boxplot()+
facet_wrap(~marital,scales = "free_y")+
labs(x="subscriped",y="age",title = "age of people subscribed")+
theme(axis.text.x = element_text(angle = 90))
```

age of people subscribed
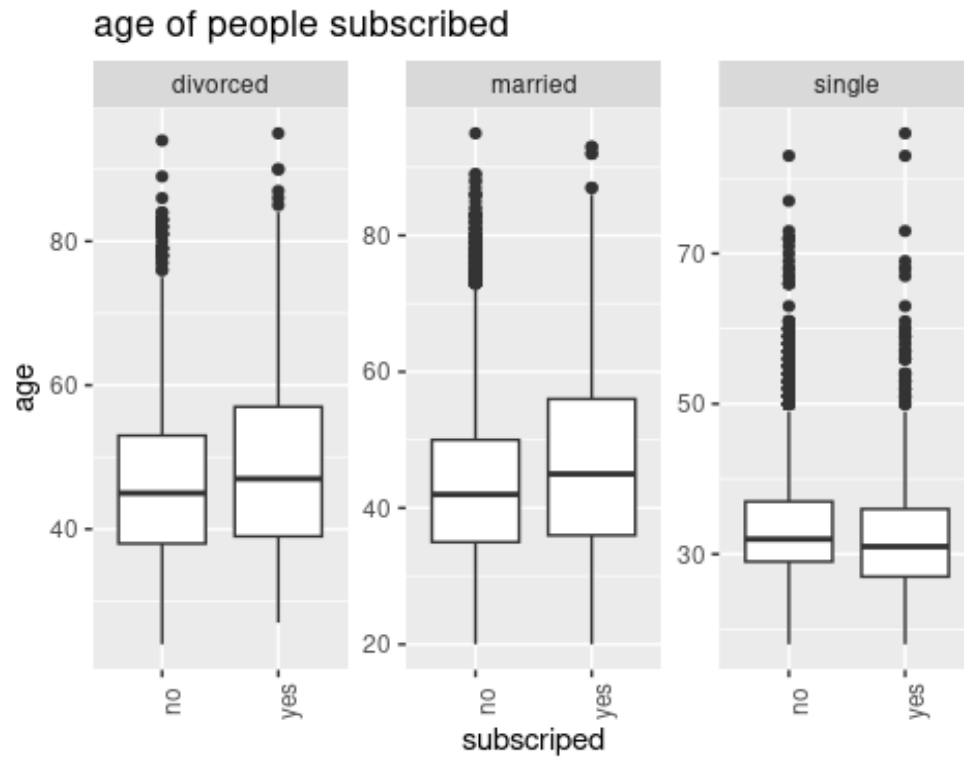

```
#scatterplot
ggplot(df,aes(x=job,y=age,color =y))+geom_point()+
facet_wrap(~education,scales = "freely")+
labs(x="job",y="age",title = "frequency distribution")+
theme(axis.text.x = element_text(angle = 90))
```

frequency distrubution
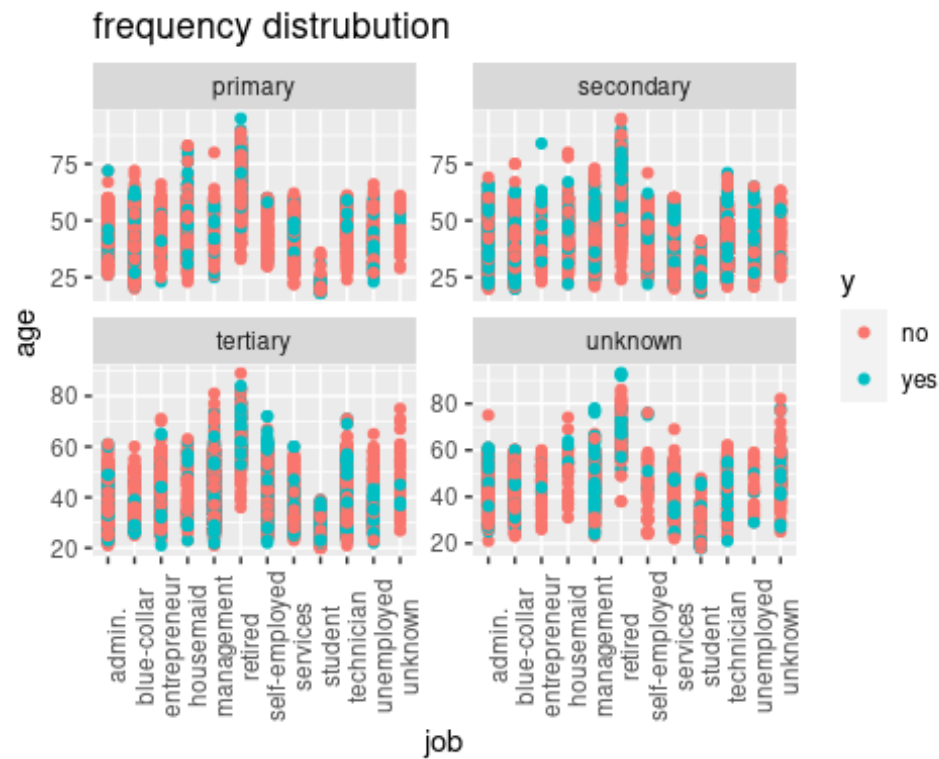
```
#barchat
ggplot(df,aes(x=job,fill=y)) + geom_bar(position="fill")
+scale_fill_brewer(palette="set2")+
facet_wrap(~education,scales = "free_y")+
labs(x="job",title = "frequency distribution")+
theme(axis.text.x = element_text(angle = 90))
```

frequency distribution
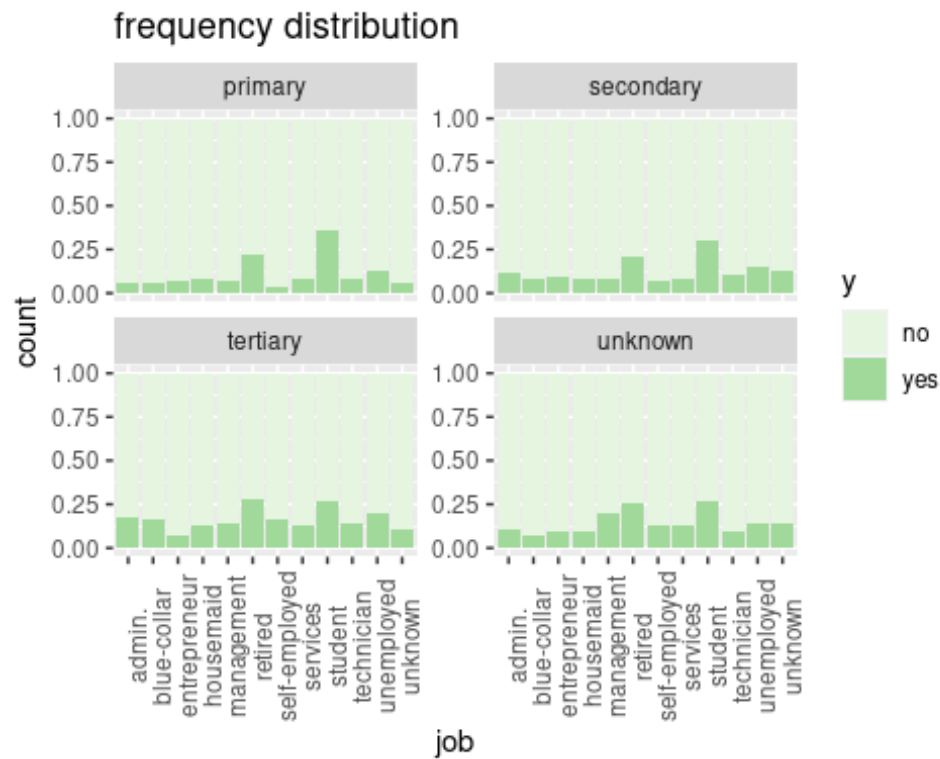
```
ggplot(df,aes(x=education,fill=y)) + geom_bar(position="fill")
+scale_fill_brewer(palette="set2")+
facet_wrap(~marital,scales = "free_y")+
labs(x="job",title = "frequency distribution")+
theme(axis.text.x = element_text(angle = 90))
```

frequency distribution
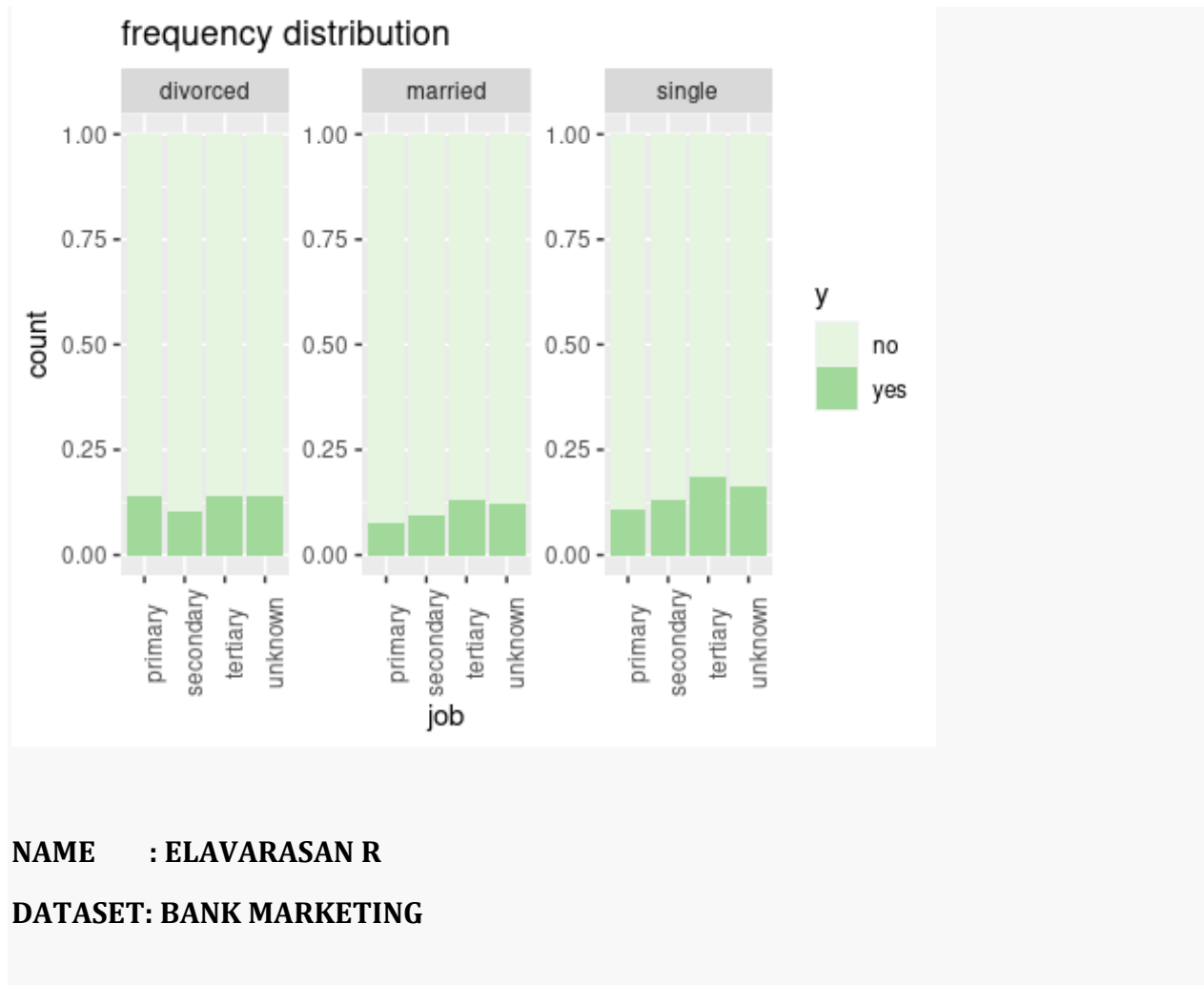
NAME   : ELAVARASAN R

DATASET: BANK MARKETING

DESCRIPTION:

The data is related with direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be ('yes') or not ('no') subscribed.

Age - (numeric)

Job : type of job (categorical: admin,blue-collar,entrepreneur,housemaid,management,retired,self-employed,services,student,technician,unemployed,unknown

Marital: marital status (categorical: divorced, married, single, unknown, note: divorced means divorced or widowed

Education: primary, secondary, tertiary, unknown

The classification goal is to predict if the client will subscribe (yes/no) a term deposit (variable y).

ASSUMPTION:

- Which type of age group people are subscribed?

- Which type of education people are subscribed?

- How many number of people are subscribed?

- How many number of people are not subscribed?

- Which type of job people are subscribed?

- Finally which type or people are maximum subscribed the term?

PROBLEM:

How am I going to increase the number of subscriber?

Inference:

- Averagely Portuguese banking institution collect the data is **45212.** Which people are subscribed or not subscribed.

- **39922** people are not subscribed.
- **5289** people are subscribed.
- Averagely the no of primary education level people are **6851** and secondary level people **23202** and tertiary people **13301** the unknown people are **1857**.
- **Distribution the age of people-**is right skewed distribution.
- **Distribution of the communicate with the people-**is normally the all contact people are distributed right skewed distribution.
- **The maximum of the 20-35 age group people are subscribed the term of deposited.**

Insight:

- Maximum of the married people are subscribed the term of deposited.

- Then tertiary education level people also subscribed the term of deposited.

- The primary education level students also subscribed the term of deposited.

- As well as secondary level student also subscribed the term of deposited but most of the people are not subscribed the term of deposited.

- The working people and the retired people minimum of the people are also subscribed the term of deposited

- Here maximum of the subscribed people is students, married, retired &management persons.