Use the link given below and locate the bank marketing dataset.
https://archive.ics.uci.edu/ml/machine-learning-databases/00222/

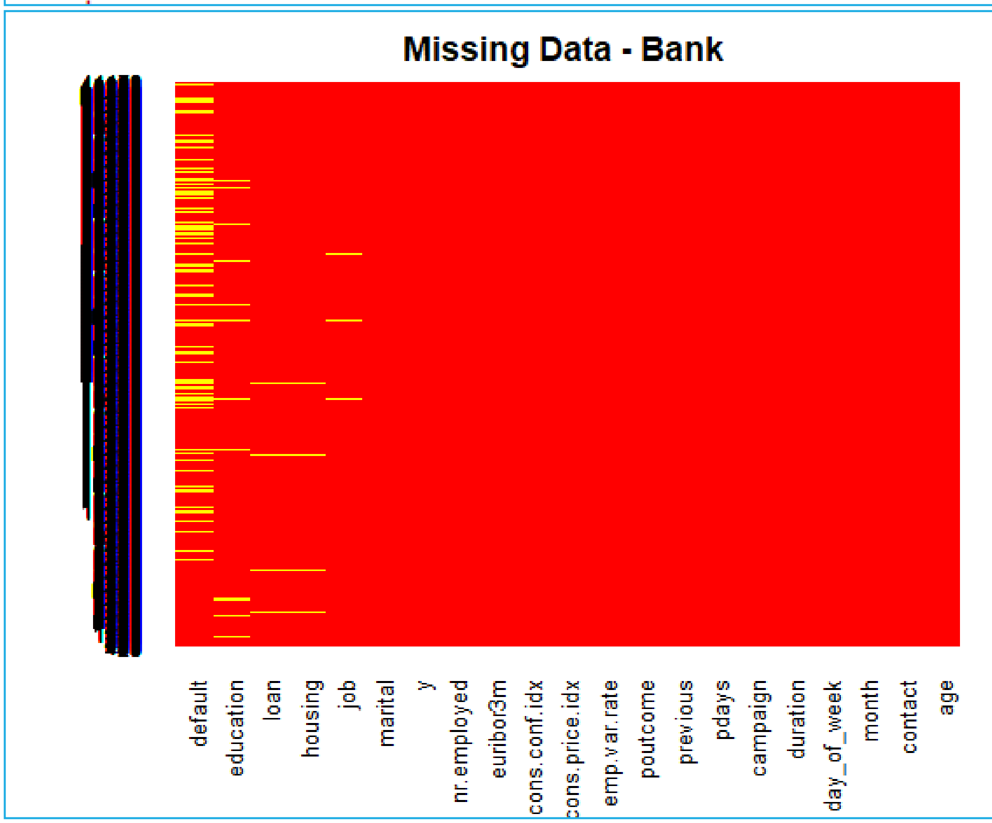
```
# Load the Data and check for the missing values

library(readr)
bank<- read_delim("G:/DATA ANALYTICS/DATA/bank-additional/bank-additional/bankdata.csv", ";", escape_double = FALSE, trim_ws = TRUE)

## Describes each variables using structure command
str(bank)

## Displays first 6 rows for each variable
head(bank)

## Summary Provides basic statistical information of each variable
summary(bank)

## DATA EXPLORATION - Check for Missing Data
## Since it is a large dataset, graphical display of missing values will prove to be easier
##Option 1

require(Amelia)
missmap(bank,main="Missing Data - Bank ", col=c("yellow","red"),legend=FALSE)

#cleaning the data of NA values for better analysis purpose

bank_full<-bankdata[complete.cases(bank), ]
View(bank_full)
missmap(bank_full,col=c("yellow","red"), legend = FALSE)
## No yellow colour stripes are visible. hence no missing values.

summary(bank_full)
```

```
> # Load the Data
> library(readr)
> bank<- read_delim("G:/DATA ANALYTICS/DATA/bank-additional/bank-additional/bankdata.csv", ";", escape_double = FALSE, trim_ws = TRUE)
Parsed with column specification:
cols(
  .default = col_character(),
  age = col_double(),
  duration = col_double(),
  campaign = col_double(),
  pdays = col_double(),
  previous = col_double(),
  emp.var.rate = col_double(),
  cons.price.idx = col_double(),
  cons.conf.idx = col_double(),
  euribor3m = col_double(),
  nr.employed = col_double()
)
See spec(...) for full column specifications.
```

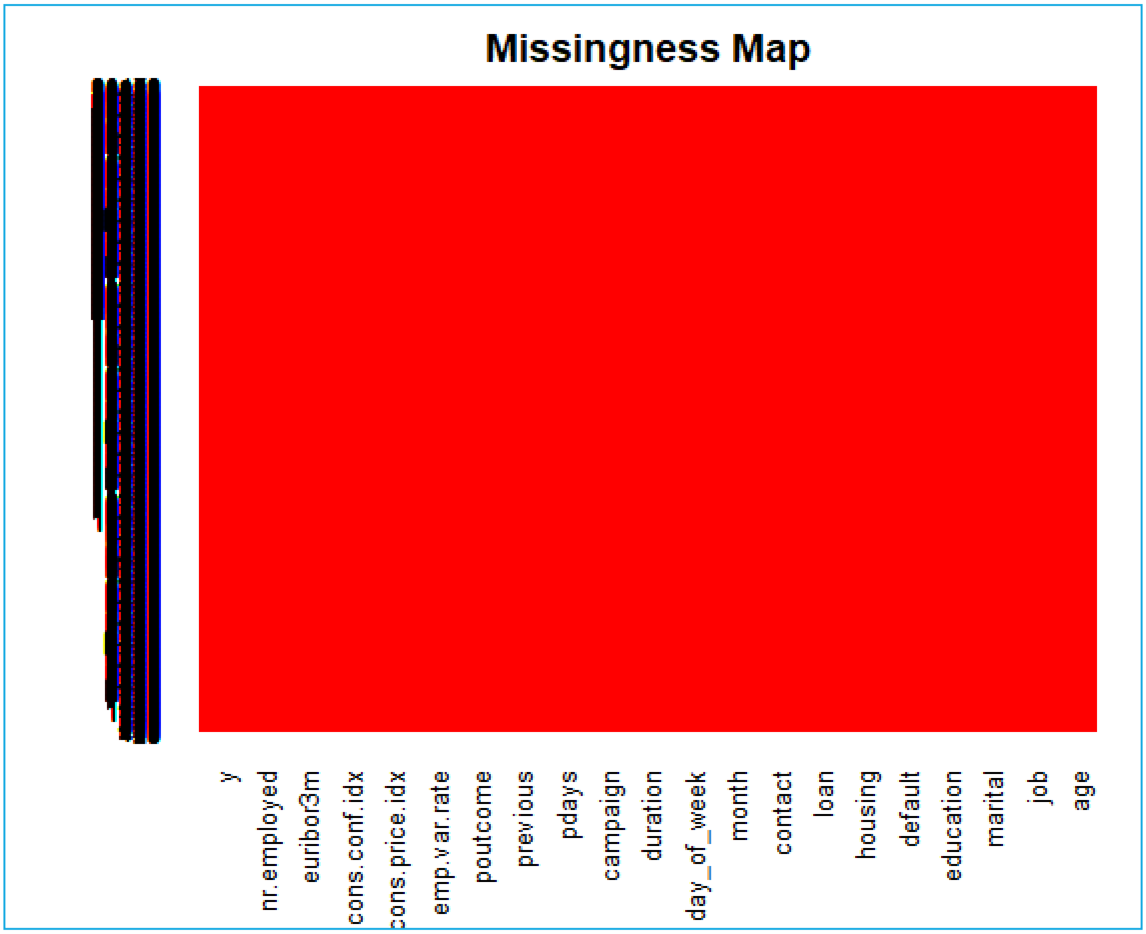
```
> str(bank) ## Describes each variables
Classes ‘spec_tbl_df’, ‘tbl_df’, ‘tbl’ and 'data.frame':	41188 obs. of  21 variables:
 $ age           : num  56 57 37 40 56 45 59 41 24 25 ...
 $ job           : chr  "housemaid" "services" "services" "admin." ...
 $ marital       : chr  "married" "married" "married" "married" ...
 $ education     : chr  "basic.4y" "high.school" "high.school" "basic.6y" ...
 $ default       : chr  "no" NA "no" "no" ...
 $ housing       : chr  "no" "no" "yes" "no" ...
 $ loan          : chr  "no" "no" "no" "no" ...
 $ contact       : chr  "telephone" "telephone" "telephone" "telephone" ...
 $ month         : chr  "may" "may" "may" "may" ...
 $ day_of_week   : chr  "mon" "mon" "mon" "mon" ...
 $ duration      : num  261 149 226 151 307 198 139 217 380 50 ...
 $ campaign      : num  1 1 1 1 1 1 1 1 1 1 ...
 $ pdays         : num  999 999 999 999 999 999 999 999 999 999 ...
 $ previous      : num  0 0 0 0 0 0 0 0 0 0 ...
 $ poutcome      : chr  "nonexistent" "nonexistent" "nonexistent" "nonexistent" ...
 $ emp.var.rate  : num  1.1 1.1 1.1 1.1 1.1 1.1 1.1 1.1 1.1 1.1 ...
 $ cons.price.idx: num  94 94 94 94 94 ...
 $ cons.conf.idx : num  -36.4 -36.4 -36.4 -36.4 -36.4 -36.4 -36.4 -36.4 -36.4 -36.4 ...
 $ euribor3m     : num  4.86 4.86 4.86 4.86 4.86 ...
 $ nr.employed   : num  5191 5191 5191 5191 5191 ...
 $ y             : chr  "no" "no" "no" "no" ...
 - attr(*, "spec")=
  .. cols(
  ..   age = col_double(),
  ..   job = col_character(),
  ..   marital = col_character(),
  ..   education = col_character(),
  ..   default = col_character(),
  ..   housing = col_character(),
  ..   loan = col_character(),
  ..   contact = col_character(),
  ..   month = col_character(),
  ..   day_of_week = col_character(),
  ..   duration = col_double(),
  ..   campaign = col_double(),
  ..   pdays = col_double(),
  ..   previous = col_double(),
  ..   poutcome = col_character(),
  ..   emp.var.rate = col_double(),
  ..   cons.price.idx = col_double(),
  ..   cons.conf.idx = col_double(),
  ..   euribor3m = col_double(),
  ..   nr.employed = col_double(),
  ..   y = col_character()
  .. )
```

```
> head(bank) ## Displays first 6 rows for each variable
# A tibble: 6 x 21
    age job    marital education default housing loan  contact month day_of_week duration campaign pdays previous poutcome emp.var.rate cons.price.idx
  <dbl> <chr>  <chr>   <chr>     <chr>   <chr>   <chr> <chr>   <chr> <chr>          <dbl>    <dbl> <dbl>    <dbl> <chr>           <dbl>          <dbl>
1    56 hous~  married basic.4y  no      no      no    teleph~ may   mon              261        1   999        0 nonexis~          1.1           94.0
2    57 serv~  married high.sch~ NA      no      no    teleph~ may   mon              149        1   999        0 nonexis~          1.1           94.0
3    37 serv~  married high.sch~ no      yes     no    teleph~ may   mon              226        1   999        0 nonexis~          1.1           94.0
4    40 admi~  married basic.6y  no      no      no    teleph~ may   mon              151        1   999        0 nonexis~          1.1           94.0
5    56 serv~  married high.sch~ no      no      yes   teleph~ may   mon              307        1   999        0 nonexis~          1.1           94.0
6    45 serv~  married basic.9y  NA      no      no    teleph~ may   mon              198        1   999        0 nonexis~          1.1           94.0
# ... with 4 more variables: cons.conf.idx <dbl>, euribor3m <dbl>, nr.employed <dbl>, y <chr>
> summary(bank) ## Provides basic statistical information of each variable
      age            job              marital           education           default            housing              loan             contact
 Min.   :17.00   Length:41188       Length:41188       Length:41188       Length:41188       Length:41188       Length:41188       Length:41188
 1st Qu.:32.00   Class :character   Class :character   Class :character   Class :character   Class :character   Class :character   Class :character
 Median :38.00   Mode  :character   Mode  :character   Mode  :character   Mode  :character   Mode  :character   Mode  :character   Mode  :character
 Mean   :40.02
 3rd Qu.:47.00
 Max.   :98.00
    month           day_of_week           duration         campaign          pdays          previous        poutcome          emp.var.rate
 Length:41188       Length:41188       Min.   :   0.0   Min.   : 1.000   Min.   :  0.0   Min.   :0.000   Length:41188       Min.   :-3.40000
 Class :character   Class :character   1st Qu.: 102.0   1st Qu.: 1.000   1st Qu.:999.0   1st Qu.:0.000   Class :character   1st Qu.:-1.80000
 Mode  :character   Mode  :character   Median : 180.0   Median : 2.000   Median :999.0   Median :0.000   Mode  :character   Median : 1.10000
                                       Mean   : 258.3   Mean   : 2.568   Mean   :962.5   Mean   :0.173                      Mean   : 0.08189
                                       3rd Qu.: 319.0   3rd Qu.: 3.000   3rd Qu.:999.0   3rd Qu.:0.000                      3rd Qu.: 1.40000
                                       Max.   :4918.0   Max.   :56.000   Max.   :999.0   Max.   :7.000                      Max.   : 1.40000
 cons.price.idx  cons.conf.idx     euribor3m      nr.employed        y
 Min.   :92.20   Min.   :-50.8   Min.   :0.634   Min.   :4964   Length:41188
 1st Qu.:93.08   1st Qu.:-42.7   1st Qu.:1.344   1st Qu.:5099   Class :character
 Median :93.75   Median :-41.8   Median :4.857   Median :5191   Mode  :character
 Mean   :93.58   Mean   :-40.5   Mean   :3.621   Mean   :5167
 3rd Qu.:93.99   3rd Qu.:-36.4   3rd Qu.:4.961   3rd Qu.:5228
 Max.   :94.77   Max.   :-26.9   Max.   :5.045   Max.   :5228
>
```

```
> require(Amelia)
Loading required package: Amelia
Loading required package: Rcpp
##
## Amelia II: Multiple Imputation
## (Version 1.7.5, built: 2018-05-07)
## Copyright (C) 2005-2019 James Honaker, Gary King and Matthew Blackwell
## Refer to http://gking.harvard.edu/amelia/ for more information
##
> missmap(bank,main="Missing Data - Bank ", col=c("yellow","red"),legend=FALSE)
```



```
> #cleaning the data of NA values for better analysis purpose
> bank_full<-bank[complete.cases(bank), ]
> missmap(bank_full,col=c("yellow","red"), legend = FALSE)
```



No yellow colour stripes are visible. hence no missing values.

```
> summary(bank_full)
      age            job               marital            education           default            housing              loan              contact         
 Min.   :17.00   Length:30488       Length:30488       Length:30488       Length:30488       Length:30488       Length:30488       Length:30488      
 1st Qu.:31.00   Class :character   Class :character   Class :character   Class :character   Class :character   Class :character   Class :character  
 Median :37.00   Mode  :character   Mode  :character   Mode  :character   Mode  :character   Mode  :character   Mode  :character   Mode  :character  
 Mean   :39.03                                                                                                                                        
 3rd Qu.:45.00                                                                                                                                        
 Max.   :95.00                                                                                                                                        
    month           day_of_week           duration         campaign          pdays          previous          poutcome          emp.var.rate     
 Length:30488       Length:30488       Min.   :   0.0   Min.   : 1.000   Min.   :  0.0   Min.   :0.0000   Length:30488       Min.   :-3.40000  
 Class :character   Class :character   1st Qu.: 103.0   1st Qu.: 1.000   1st Qu.:999.0   1st Qu.:0.0000   Class :character   1st Qu.:-1.80000  
 Mode  :character   Mode  :character   Median : 181.0   Median : 2.000   Median :999.0   Median :0.0000   Mode  :character   Median : 1.10000  
                                       Mean   : 259.5   Mean   : 2.521   Mean   :956.3   Mean   :0.1943                      Mean   :-0.07151  
                                       3rd Qu.: 321.0   3rd Qu.: 3.000   3rd Qu.:999.0   3rd Qu.:0.0000                      3rd Qu.: 1.40000  
                                       Max.   :4918.0   Max.   :43.000   Max.   :999.0   Max.   :7.0000                      Max.   : 1.40000  
 cons.price.idx  cons.conf.idx      euribor3m      nr.employed        y            
 Min.   :92.20   Min.   :-50.8   Min.   :0.634   Min.   :4964   Length:30488      
 1st Qu.:93.08   1st Qu.:-42.7   1st Qu.:1.313   1st Qu.:5099   Class :character  
 Median :93.44   Median :-41.8   Median :4.856   Median :5191   Mode  :character  
 Mean   :93.52   Mean   :-40.6   Mean   :3.460   Mean   :5161                     
 3rd Qu.:93.99   3rd Qu.:-36.4   3rd Qu.:4.961   3rd Qu.:5228                     
 Max.   :94.77   Max.   :-26.9   Max.   :5.045   Max.   :5228                     
```

The Pearson's chi-squared test of independence is one of the most basic and common hypothesis tests in the statistical analysis of categorical data. It is a significance test. Given two categorical random variables, X and Y, the chi-squared test of independence determines whether or not there exists a statistical dependence between them. Formally, it is a hypothesis test. The chi-squared test assumes a null hypothesis and an alternate hypothesis. The general practice is, if the p-value that comes out in the #result is less than a pre-determined significance level, which is 0.05 usually, then we reject the null hypothesis.

H0: The two variables are independent
H1: The two variables are dependent

The null hypothesis of the chi-squared test is that the two variables are independent and the alternate hypothesis is that they are related.

To establish that two categorical variables (or predictors) are dependent, the chi-squared statistic must have a certain cutoff. This cutoff increases as the number of classes within the variable (or predictor) increases. Pearson's chi-squared test of independence (significance test)

Perform the below operations:

a. Is there any association between job and default?

```
chisq.test(bank_full$job ,bank_full$default)
with(bank_full, chisq.test( job, default))
with(bank_full, table( job, default) )
# OR
with(bank_full, prop.table(table( job,default)))

#X-squared = 18.2, df = 10, p-value = 0.05168
```

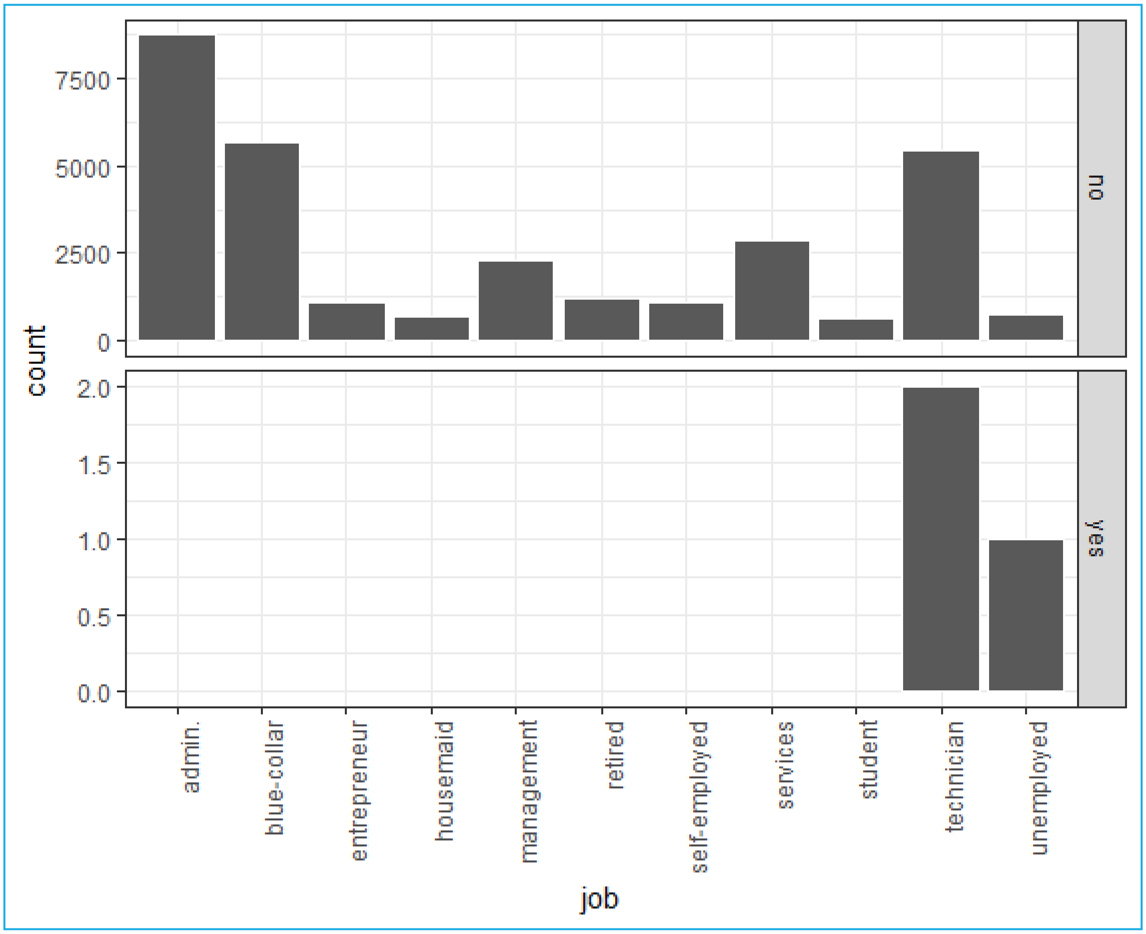
Console ~/

```
> chisq.test(bank_full$job ,bank_full$default)

	Pearson's Chi-squared test

data:  bank_full$job and bank_full$default
X-squared = 18.2, df = 10, p-value = 0.05168

Warning message:
In chisq.test(bank_full$job, bank_full$default) :
  Chi-squared approximation may be incorrect
> # OR
> with(bank_full, chisq.test( job, default))

	Pearson's Chi-squared test

data:  job and default
X-squared = 18.2, df = 10, p-value = 0.05168

Warning message:
In chisq.test(job, default) : Chi-squared approximation may be incorrect
> with(bank_full, table( job, default) )
               default
job               no  yes
  admin.        8737    0
  blue-collar   5675    0
  entrepreneur  1089    0
  housemaid      690    0
  management    2311    0
  retired       1216    0
  self-employed 1092    0
  services      2857    0
  student        610    0
  technician    5471    2
  unemployed     737    1
> # OR
> with(bank_full, prop.table(table( job,default)))
               default
job                       no          yes
  admin.        2.865718e-01 0.000000e+00
  blue-collar   1.861388e-01 0.000000e+00
  entrepreneur  3.571897e-02 0.000000e+00
  housemaid     2.263186e-02 0.000000e+00
  management    7.580031e-02 0.000000e+00
  retired       3.988454e-02 0.000000e+00
  self-employed 3.581737e-02 0.000000e+00
  services      9.370900e-02 0.000000e+00
  student       2.000787e-02 0.000000e+00
  technician    1.794477e-01 6.559958e-05
  unemployed    2.417345e-02 3.279979e-05
>
```

#as p-value is > 0.05 there is no association between job and default

```
ggplot(bank_full) + geom_bar(aes(x = job), col = "white") +
  facet_grid(default~., scales = "free") + theme_bw() + theme(axis.text.x = element_text(angle
= 90, hjust = 1))
```

Technicians default maximum and admin defaults minimum. Only unemployed and technicians default.



b. Is there any significant difference in duration of last call between people having housing loan or not?

```
chisq.test(bank_full$duration ,bank_full$housing)

# OR
with(bank_full, chisq.test(duration,housing))
with(bank_full, table( duration,housing) )

# OR
with(bank_full, prop.table(table(duration, housing)))

#data: duration and housing
#X-squared = 1440.8, df = 1440, p-value = 0.4893
#P value is above 0.05 hence there is no association between people having housing loan or not
```

```
> chisq.test(bank_full$duration ,bank_full$housing)

        Pearson's Chi-squared test

data:  bank_full$duration and bank_full$housing
X-squared = 1440.8, df = 1440, p-value = 0.4893
```

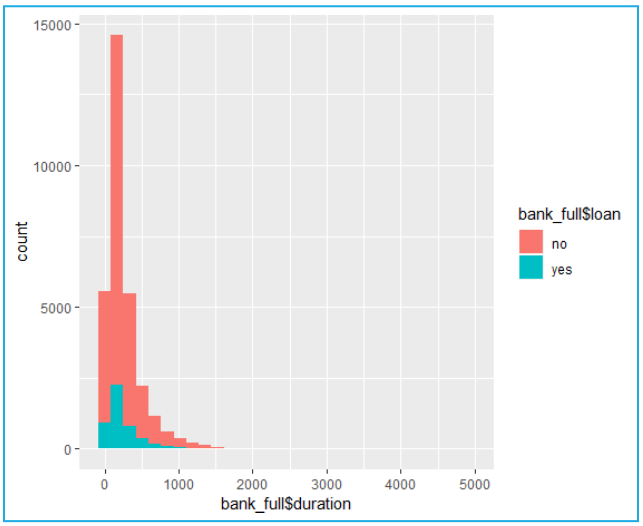
```
> # OR
> with(bank_full, chisq.test(duration,housing))

        Pearson's Chi-squared test

data:  duration and housing
X-squared = 1440.8, df = 1440, p-value = 0.4893
```

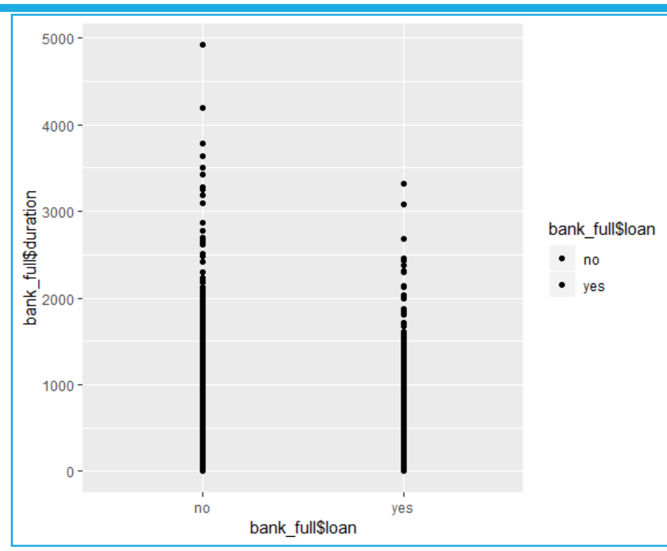
```
> with(bank_full, table( duration,housing) )
        housing
duration no yes
       0  1   3
       1  2   1
       2  1   0
       3  2   1
       4  2  10
       5 13  11
       6 12  19
       7 17  28
       8 19  32
       9 25  36
      10 26  23
      11 25  37
      12 19  34
      13 34  24
      14 19  36
      15 28  26
      16 27  27
      17 25  27
      18 31  28
      19 16  27
      20 23  25
      21 22  31
```

```
> # OR
> with(bank_full, prop.table(table(duration, housing)))
        housing
duration           no          yes
       0 3.279979e-05 9.839937e-05
       1 6.559958e-05 3.279979e-05
       2 3.279979e-05 0.000000e+00
       3 6.559958e-05 3.279979e-05
       4 6.559958e-05 3.279979e-04
       5 4.263973e-04 3.607977e-04
       6 3.935975e-04 6.231960e-04
       7 5.575964e-04 9.183941e-04
       8 6.231960e-04 1.049593e-03
       9 8.199948e-04 1.180792e-03
      10 8.527945e-04 7.543952e-04
      11 8.199948e-04 1.213592e-03
      12 6.231960e-04 1.115193e-03
      13 1.115193e-03 7.871950e-04
      14 6.231960e-04 1.180792e-03
      15 9.183941e-04 8.527945e-04
      16 8.855943e-04 8.855943e-04
      17 8.199948e-04 8.855943e-04
      18 1.016793e-03 9.183941e-04
      19 5.247966e-04 8.855943e-04
```

#data: duration and housing
#X-squared = 1440.8, df = 1440, p-value = 0.4893
#P value is above 0.05 hence there is no association between people having housing loan or not

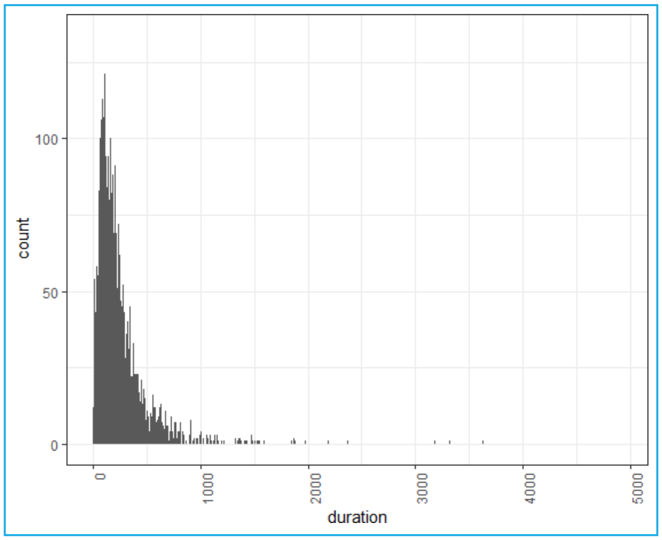
```
library(ggplot2)
bank_full$duration<-as.numeric(bank_full$duration)
ggplot(bank_full, aes(x=bank_full$duration, fill=bank_full$loan))+geom_histogram()
ggplot(bank_full, aes(x=bank_full$loan,y=bank_full$duration, fill=bank_full$loan))+geom_point()
```

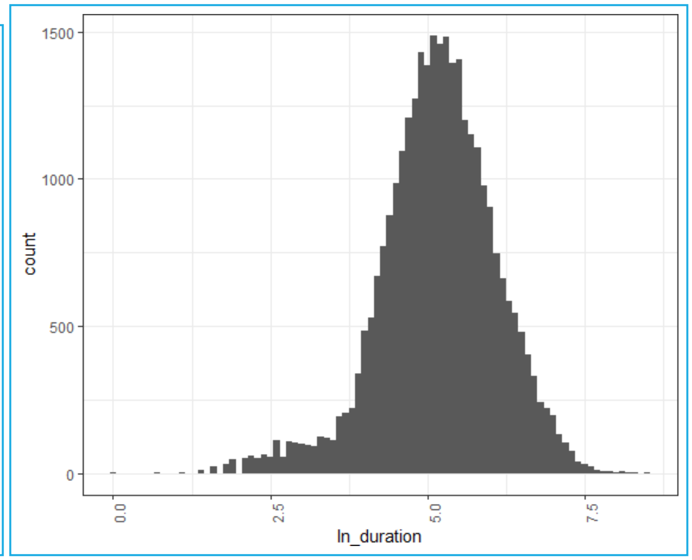


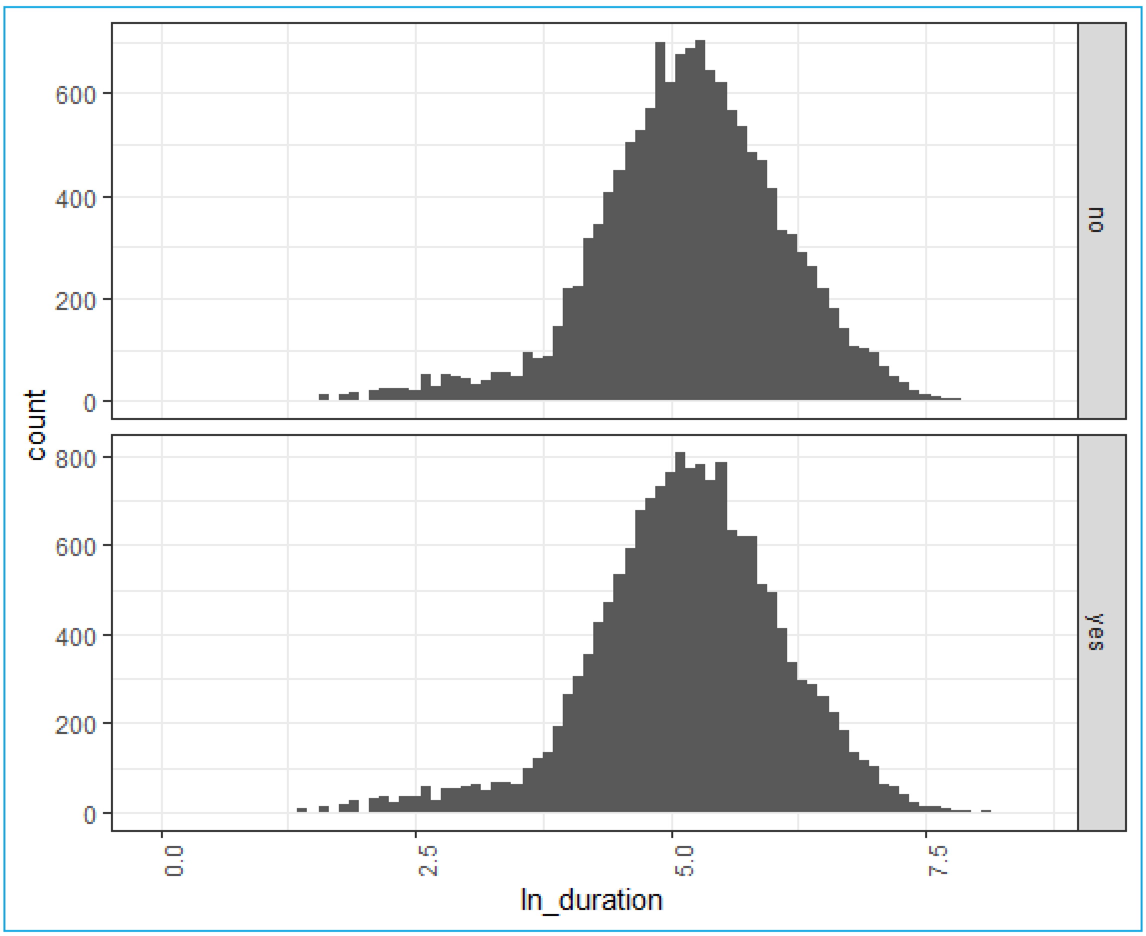


#As per the plots it is visible that duration taken is more for customers without loan.

```
ggplot(bank_full, aes(x = duration)) + geom_bar() + theme_bw() + theme(axis.text.x =
element_text(angle = 90, hjust = 1))
ln_duration <- log(bank_full$duration)
ggplot(bank_full, aes(x = ln_duration)) + geom_histogram(binwidth = 0.1) + theme_bw() +
theme(axis.text.x = element_text(angle = 90, hjust = 1))
ggplot(bank_full) + geom_histogram(aes(x = ln_duration), binwidth = 0.1) +
  facet_grid(housing~., scales = "free") + theme_bw() + theme(axis.text.x = element_text(angle = 90,
hjust = 1))
```







#In call duration of 5 min 800 customers have taken home loan and 600 have not taken.

c. Is there any association between consumer price index and consumer?

chisq.test(bank_full$cons.price.idx ,bank_full$cons.conf.idx)
 # OR
with(bank_full, chisq.test(cons.price.idx,cons.conf.idx))
with(bank_full, table(cons.price.idx,cons.conf.idx))
 # OR
with(bank_full, prop.table(table(cons.price.idx,cons.conf.idx)))

#X-squared = 762200, df = 625, p-value < 2.2e-16

```
> chisq.test(bank_full$cons.price.idx ,bank_full$cons.conf.idx)

        Pearson's Chi-squared test

data:  bank_full$cons.price.idx and bank_full$cons.conf.idx
X-squared = 762200, df = 625, p-value < 2.2e-16
```

```
> # OR
> with(bank_full, chisq.test(cons.price.idx,cons.conf.idx))

        Pearson's Chi-squared test

data:  cons.price.idx and cons.conf.idx
X-squared = 762200, df = 625, p-value < 2.2e-16
```

```
> with(bank_full, table(cons.price.idx,cons.conf.idx))
```

cons.price.idx \ cons.conf.idx	-50.8	-50	-49.5	-47.1	-46.2	-45.9	-42.7	-42	-41.8	-40.8	-40.4	-40.3	-40	-39.8	-38.3	-37.5	-36.4	-36.1	-34.8	-34.6	-33.6	-33	-31.4	-30.1
92.201	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	676	0
92.379	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
92.431	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
92.469	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	157	0	0	0
92.649	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	326
92.713	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	147	0	0
92.756	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
92.843	0	261	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
92.893	0	0	0	0	4616	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
92.963	0	0	0	0	0	0	0	0	0	628	0	0	0	0	0	0	0	0	0	0	0	0	0	0
93.075	0	0	0	1970	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
93.2	0	0	0	0	0	0	0	3054	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
93.369	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	221	0	0	0	0	0
93.444	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3798	0	0	0	0	0	0
93.749	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	145	0	0	0	0
93.798	0	0	0	0	0	0	0	0	0	0	63	0	0	0	0	0	0	0	0	0	0	0	0	0
93.876	0	0	0	0	0	0	0	0	0	0	0	0	179	0	0	0	0	0	0	0	0	0	0	0
93.918	0	0	0	0	0	0	4646	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
93.994	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4938	0	0	0	0	0	0	0
94.027	0	0	0	0	0	0	0	0	0	0	0	0	0	0	199	0	0	0	0	0	0	0	0	0
94.055	0	0	0	0	0	0	0	0	0	0	0	0	0	210	0	0	0	0	0	0	0	0	0	0
94.199	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	266	0	0	0	0	0	0	0	0
94.215	0	0	0	0	0	0	0	0	0	0	0	278	0	0	0	0	0	0	0	0	0	0	0	0
94.465	0	0	0	0	0	0	0	0	2776	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
94.601	0	0	183	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
94.767	116	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

```
> # OR
> with(bank_full, prop.table(table(cons.price.idx,cons.conf.idx)))
              cons.conf.idx
cons.price.idx        -50.8           -50         -49.5         -47.1         -46.2         -45.9         -42.7           -42         -41.8         -40.8         -40.4
        92.201 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
        92.379 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
        92.431 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
        92.469 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
        92.649 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
        92.713 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
        92.756 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0003279979 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
        92.843 0.0000000000 0.0085607452 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
        92.893 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.1514038310 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
        92.963 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0205982682 0.0000000000
        93.075 0.0000000000 0.0000000000 0.0000000000 0.0646155865 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
        93.2   0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.1001705589 0.0000000000 0.0000000000 0.0000000000
        93.369 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
        93.444 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
        93.749 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
        93.798 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0020663868
        93.876 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
        93.918 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.1523878247 0.0000000000 0.0000000000 0.0000000000 0.0000000000
        93.994 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
        94.027 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
        94.055 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
        94.199 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
        94.215 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
        94.465 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0910522173 0.0000000000 0.0000000000
        94.601 0.0000000000 0.0000000000 0.0060023616 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
        94.767 0.0038047756 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
```

 # P value is less than 0.05 hence we can conclude, that the variables, con.price.idx , cons.conf.idx are highly dependent to each other.

d. Is the employment variation rate consistent across Job types?

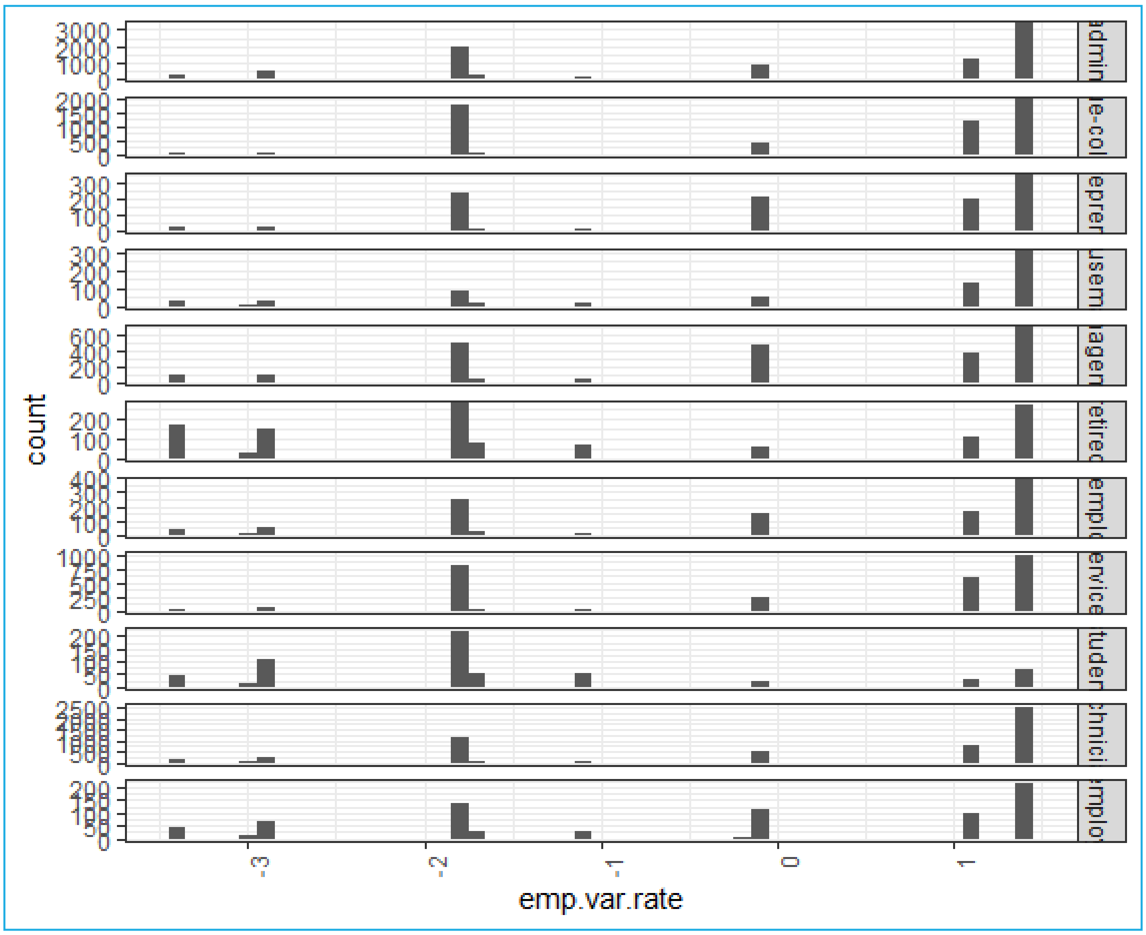
with(bank_full, chisq.test(job,emp.var.rate))
with(bank_full, table(job,emp.var.rate))
 # OR
with(bank_full, prop.table(table(job,emp.var.rate)))
 # X-squared = 3481.7, df = 90, p-value < 2.2e-16
#P value is less than 0.05 hence we can conclude, that the variables, employment variation rate consistent across job types

```
> with(bank_full, chisq.test( job,emp.var.rate))

        Pearson's Chi-squared test

data:  job and emp.var.rate
X-squared = 3481.7, df = 90, p-value < 2.2e-16
```

```
> with(bank_full, table( job,emp.var.rate) )
               emp.var.rate
job             -3.4   -3 -2.9 -1.8 -1.7 -1.1 -0.2 -0.1  1.1  1.4
  admin.         297   41  528 1985  235  177    3  840 1234 3397
  blue-collar     61    8   86 1760   55   31    3  456 1233 1982
  entrepreneur    22    1   31  243   14    7    0  217  200  354
  housemaid       30    9   33   84   17   16    1   54  137  309
  management      86    8  107  494   45   35    0  469  370  697
  retired        169   28  150  285   79   71    0   56  106  272
  self-employed   37    5   56  249   21   12    0  155  165  392
  services        32    2   75  828   39   39    0  241  600 1001
  student         43   17  106  220   55   52    0   18   31   68
  technician     131   19  221 1105   98  101    2  499  766 2531
  unemployed      43    9   68  139   29   24    1  112   96  217
>
```

```
> # OR
> with(bank_full, prop.table(table( job,emp.var.rate)))
               emp.var.rate
job                    -3.4           -3         -2.9         -1.8         -1.7         -1.1         -0.2         -0.1          1.1          1.4
  admin.       9.741538e-03 1.344791e-03 1.731829e-02 6.510758e-02 7.707951e-03 5.805563e-03 9.839937e-05 2.755182e-02 4.047494e-02 1.114209e-01
  blue-collar  2.000787e-03 2.623983e-04 2.820782e-03 5.772763e-02 1.803988e-03 1.016793e-03 9.839937e-05 1.495670e-02 4.044214e-02 6.500918e-02
  entrepreneur 7.215954e-04 3.279979e-05 1.016793e-03 7.970349e-03 4.591971e-04 2.295985e-04 0.000000e+00 7.117554e-03 6.559958e-03 1.161113e-02
  housemaid    9.839937e-04 2.951981e-04 1.082393e-03 2.755182e-03 5.575964e-04 5.247966e-04 3.279979e-05 1.771189e-03 4.493571e-03 1.013514e-02
  management   2.820782e-03 2.623983e-04 3.509578e-03 1.620310e-02 1.475991e-03 1.147993e-03 0.000000e+00 1.538310e-02 1.213592e-02 2.286145e-02
  retired      5.543165e-03 9.183941e-04 4.919969e-03 9.347940e-03 2.591183e-03 2.328785e-03 0.000000e+00 1.836788e-03 3.476778e-03 8.921543e-03
  self-employed 1.213592e-03 1.639990e-04 1.836788e-03 8.167148e-03 6.887956e-04 3.935975e-04 0.000000e+00 5.083967e-03 5.411965e-03 1.285752e-02
  services     1.049593e-03 6.559958e-05 2.459984e-03 2.715823e-02 1.279192e-03 1.279192e-03 0.000000e+00 7.904749e-03 1.967987e-02 3.283259e-02
  student      1.410391e-03 5.575964e-04 3.476778e-03 7.215954e-03 1.803988e-03 1.705589e-03 0.000000e+00 5.903962e-04 1.016793e-03 2.230386e-03
  technician   4.296773e-03 6.231960e-04 7.248754e-03 3.624377e-02 3.214379e-03 3.312779e-03 6.559958e-05 1.636710e-02 2.512464e-02 8.301627e-02
  unemployed   1.410391e-03 2.951981e-04 2.230386e-03 4.559171e-03 9.511939e-04 7.871950e-04 3.279979e-05 3.673576e-03 3.148780e-03 7.117554e-03
>
```

ggplot(bank_full) + geom_histogram(aes(x = emp.var.rate), binwidth = 0.1) +
facet_grid(job~., scales = "free") + theme_bw() + theme(axis.text.x = element_text(angle = 90, hjust = 1))
#Yes employment variation rate consistent across Job types



e. Is the employment variation rate same across Education?

with(bank_full, chisq.test(job,emp.var.rate))
with(bank_full, table(job,emp.var.rate))
 # OR
with(bank_full, prop.table(table(job,emp.var.rate)))
 # X-squared = 3481.7, df = 90, p-value < 2.2e-16
#P value is less than 0.05 hence we can conclude, that the variables, employment variation rate consistent across job types

```
> with(bank_full, chisq.test( education,emp.var.rate))

        Pearson's Chi-squared test

data:  education and emp.var.rate
X-squared = 915.91, df = 54, p-value < 2.2e-16
```

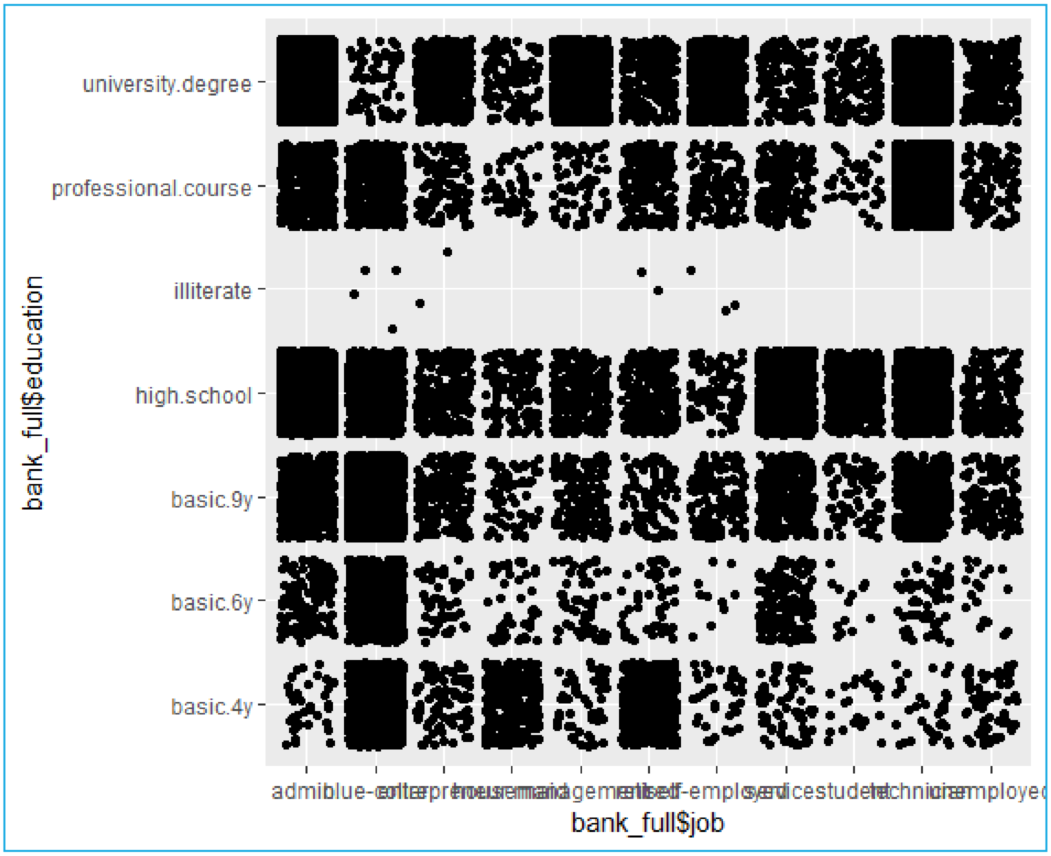
```
> with(bank_full, table( education, emp.var.rate) )
                     emp.var.rate
education             -3.4   -3 -2.9 -1.8 -1.7 -1.1 -0.2 -0.1  1.1  1.4
  basic.4y             130   13   95  596   68   54    3  179  446  796
  basic.6y              35    0   30  391   17    9    0  122  320  465
  basic.9y              64   15  100 1262   50   26    0  414  899 1446
  high.school          210   34  342 2020  177  140    4  708 1363 2701
  illiterate             0    0    2    2    0    0    0    3    0    4
  professional.course  127   19  183  930   88  105    3  409  654 1803
  university.degree    385   66  709 2191  287  231    0 1282 1256 4005
> 
```

```
> # OR
> with(bank_full, prop.table(table( education,emp.var.rate)))
                     emp.var.rate
education                     -3.4           -3         -2.9         -1.8         -1.7         -1.1         -0.2         -0.1          1.1          1.4
  basic.4y            4.263973e-03 4.263973e-04 3.115980e-03 1.954867e-02 2.230386e-03 1.771189e-03 9.839937e-05 5.871162e-03 1.462871e-02 2.610863e-02
  basic.6y            1.147993e-03 0.000000e+00 9.839937e-04 1.282472e-02 5.575964e-04 2.951981e-04 0.000000e+00 4.001574e-03 1.049593e-02 1.525190e-02
  basic.9y            2.099187e-03 4.919969e-04 3.279979e-03 4.139334e-02 1.639990e-03 8.527945e-04 0.000000e+00 1.357911e-02 2.948701e-02 4.742850e-02
  high.school         6.887956e-03 1.115193e-03 1.121753e-02 6.625558e-02 5.805563e-03 4.591971e-03 1.311992e-04 2.322225e-02 4.470611e-02 8.859223e-02
  illiterate          0.000000e+00 0.000000e+00 6.559958e-05 6.559958e-05 0.000000e+00 0.000000e+00 0.000000e+00 9.839937e-05 0.000000e+00 1.311992e-04
  professional.course 4.165573e-03 6.231960e-04 6.002362e-03 3.050380e-02 2.886382e-03 3.443978e-03 9.839937e-05 1.341511e-02 2.145106e-02 5.913802e-02
  university.degree   1.262792e-02 2.164786e-03 2.325505e-02 7.186434e-02 9.413540e-03 7.576752e-03 0.000000e+00 4.204933e-02 4.119654e-02 1.313632e-01
> 
```

#P value is less than 0.05 hence we can conclude, that the variables, employment variation rate and education are dependent

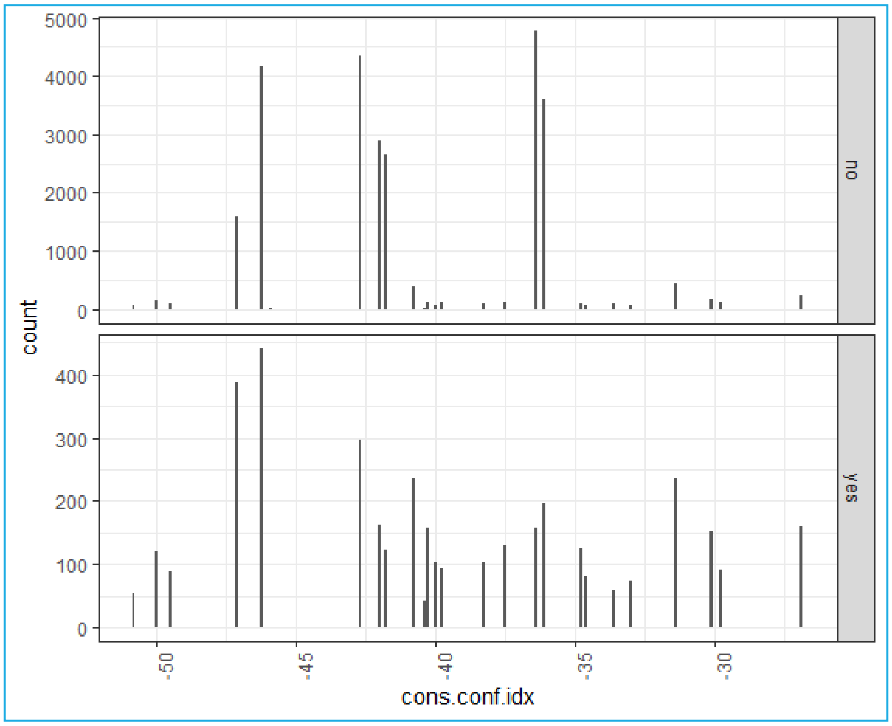
ggplot(bank_full, aes(x=bank_full$job, y=bank_full$education))+geom_jitter()

#Employment variation rate is not same as per the above plot Higher the education, Higher job profile



f. Which group is more confident?

 # on the basis of JOB
ggplot(bank_full) + geom_histogram(aes(x = cons.conf.idx), binwidth = 0.1) +
facet_grid(y~., scales = "free") + theme_bw() + theme(axis.text.x = element_text(angle = 90, hjust = 1))
#People who have not taken loan are more confident



on the basis of default

ggplot(bank_full) + geom_histogram(aes(x = cons.conf.idx), binwidth = 0.1) +
facet_grid(default~., scales = "free") + theme_bw() + theme(axis.text.x = element_text(angle = 90, hjust = 1))

#Non defaulters are more confident.