ACADGILD

SESSION 11: Linear Models

Assignment 2

Table of Contents

1. Introduction

This assignment will help you understand the concepts learnt in the session.

2. Objective

This assignment will test your skills on the basics of Regression Analysis and Modeling.

3. Prerequisites

Not applicable.

4. Associated Data Files

Not applicable.

5. Problem Statement

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

Perform the below operations:

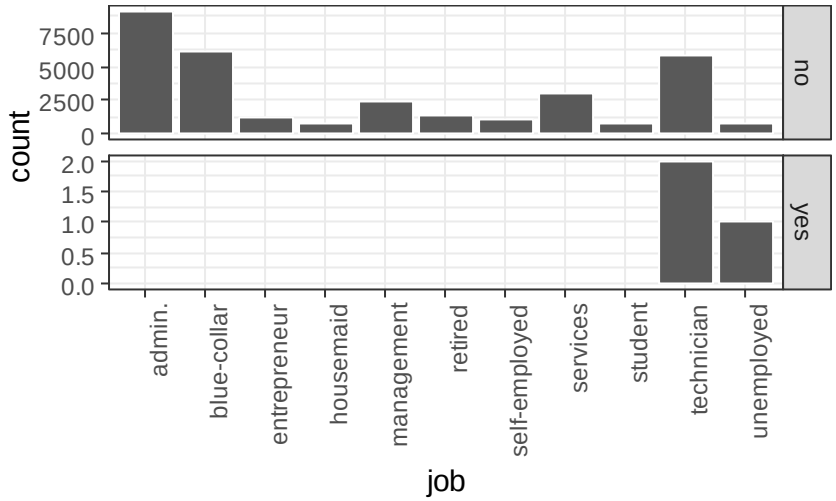
a. Is there any association between job and default?
b. Is there any significant difference in duration of last call between?
People having housing loan or not?
c. Is there any association between consumer price index and consumer?
d. Is the employment variation rate consistent across Job types?
e. Is the employment variation rate same across Education?
f. Which group is more confident?

Ans 1- Association between job and default.

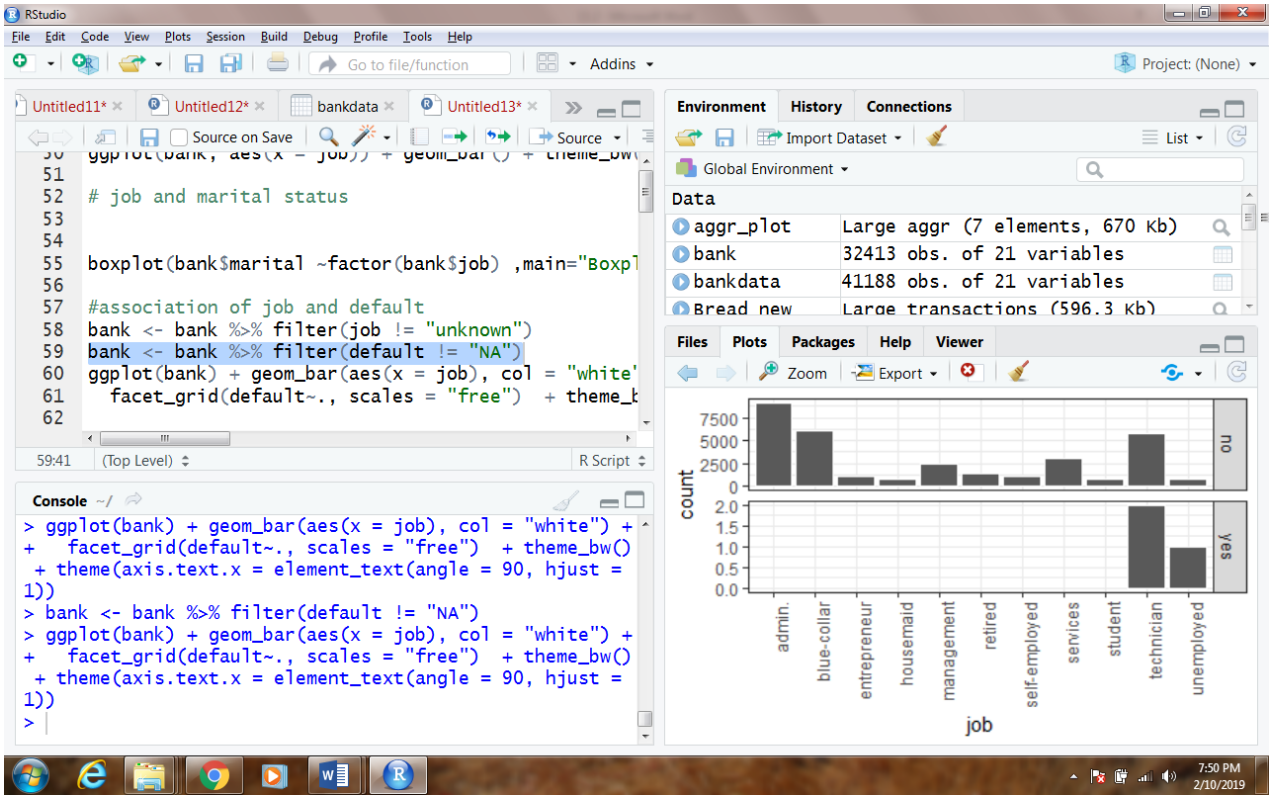
```
bank <- bank %>% filter(job != "unknown")

bank <- bank %>% filter(default != "NA")

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



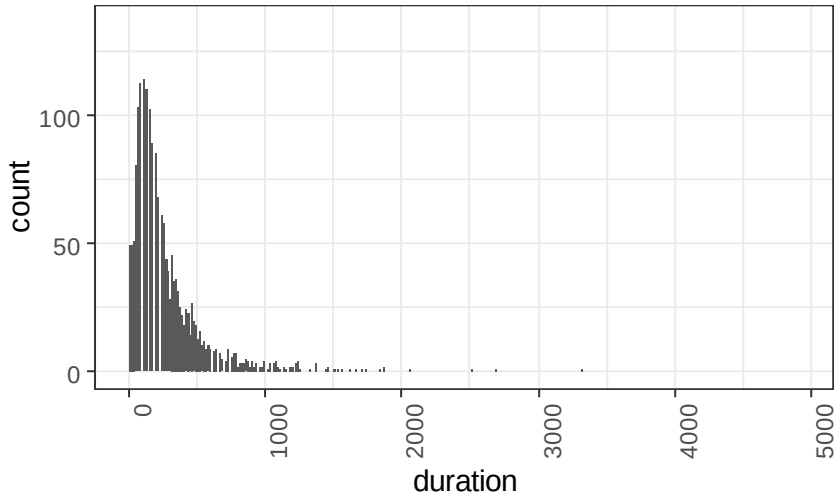
Technicians default maximum and admin defaults minimum. Only unemployed and technicians default.

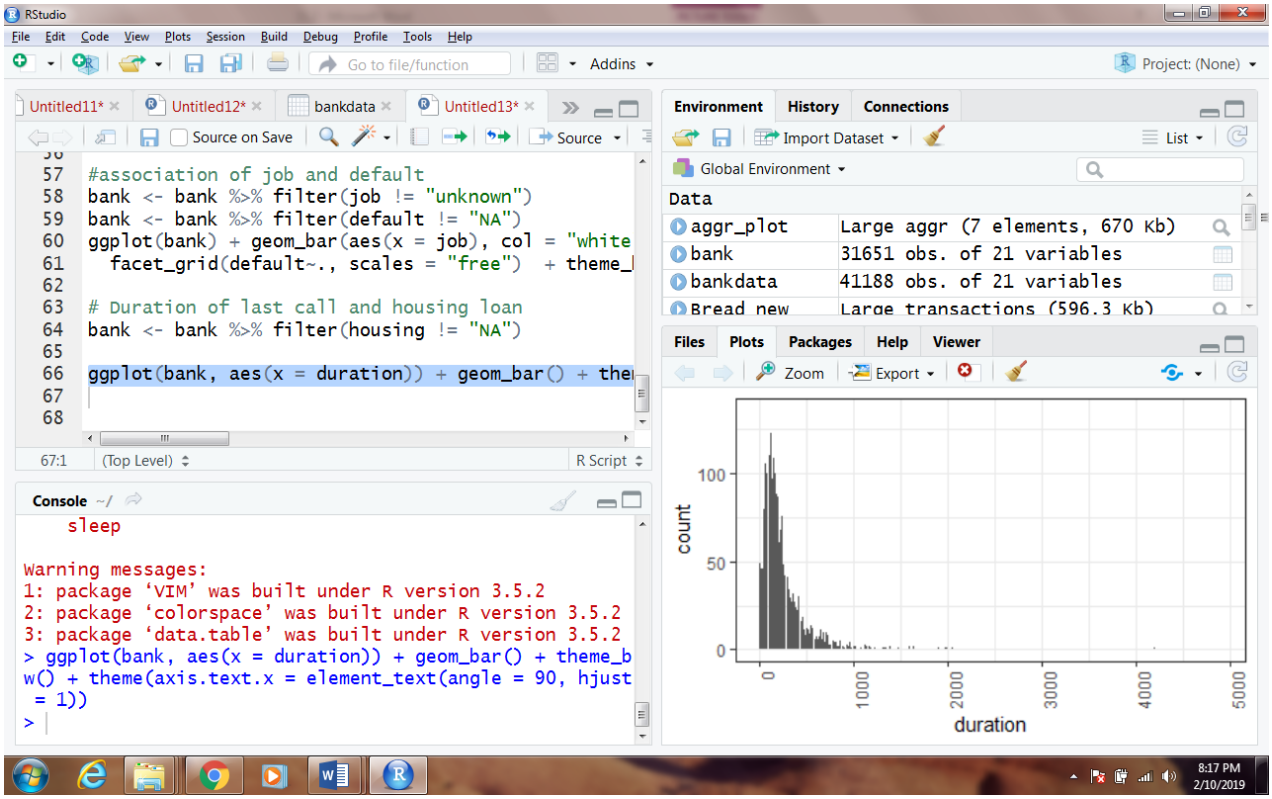


B - Is there any significant difference in duration of last call between?

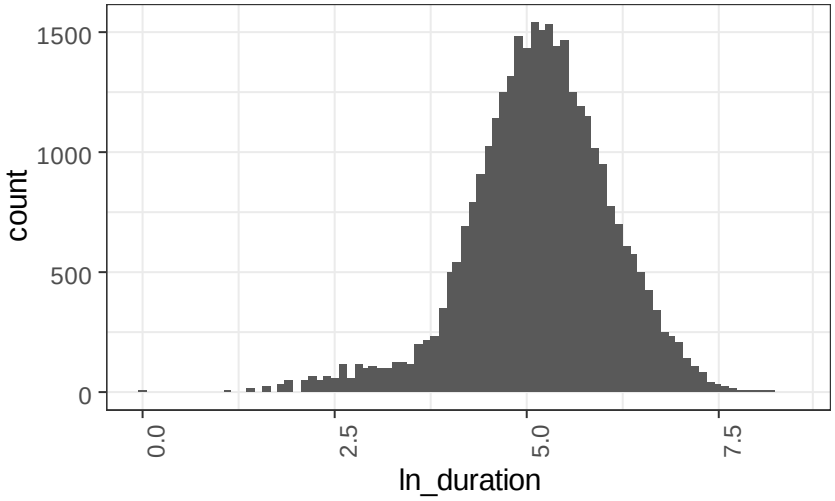
People having housing loan or not?

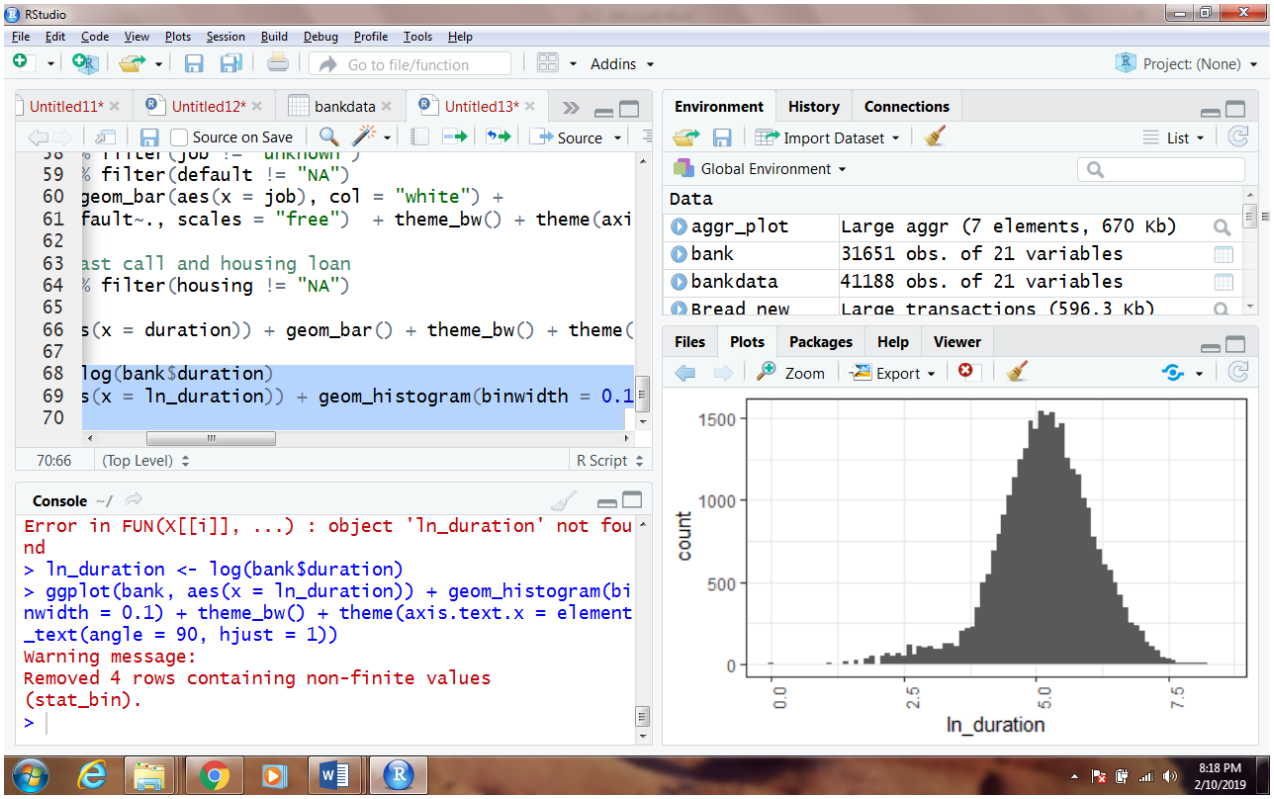
ggplot(bank, aes(x = duration)) + geom_bar() + theme_bw() + theme(axis.text.x = element_text(angle = 90, hjust = 1))


count
100
50
0
0
1000
2000
3000
4000
5000
duration

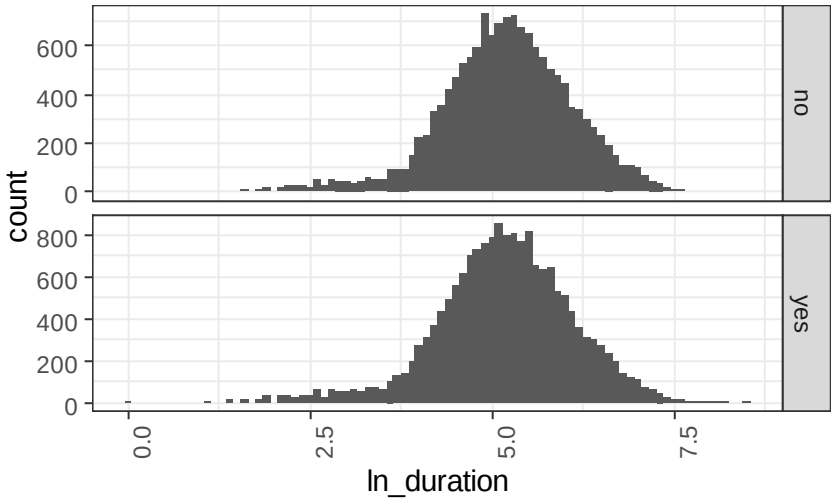

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Addins
Project: (None)
Untitled11*
Untitled12*
bankdata
Untitled13*
Source on Save
Source
#association of job and default
bank <- bank %>% filter(job != "unknown")
bank <- bank %>% filter(default != "NA")
ggplot(bank) + geom_bar(aes(x = job), col = "white
facet_grid(default~., scales = "free") + theme_
# Duration of last call and housing loan
bank <- bank %>% filter(housing != "NA")
ggplot(bank, aes(x = duration)) + geom_bar() + the
67:1
(Top Level)
R Script
Console ~/
sleep
Warning messages:
1: package 'VIM' was built under R version 3.5.2
2: package 'colorspace' was built under R version 3.5.2
3: package 'data.table' was built under R version 3.5.2
> ggplot(bank, aes(x = duration)) + geom_bar() + theme_b
w() + theme(axis.text.x = element_text(angle = 90, hjust
= 1))
>
Environment
History
Connections
Import Dataset
List
Global Environment
Data
aggr_plot
Large aggr (7 elements, 670 Kb)
bank
31651 obs. of 21 variables
bankdata
41188 obs. of 21 variables
Bread new
Large transactions (596.3 Kb)
Files
Plots
Packages
Help
Viewer
Zoom
Export
count
100
50
0
0
1000
2000
3000
4000
5000
duration
8:17 PM
2/10/2019

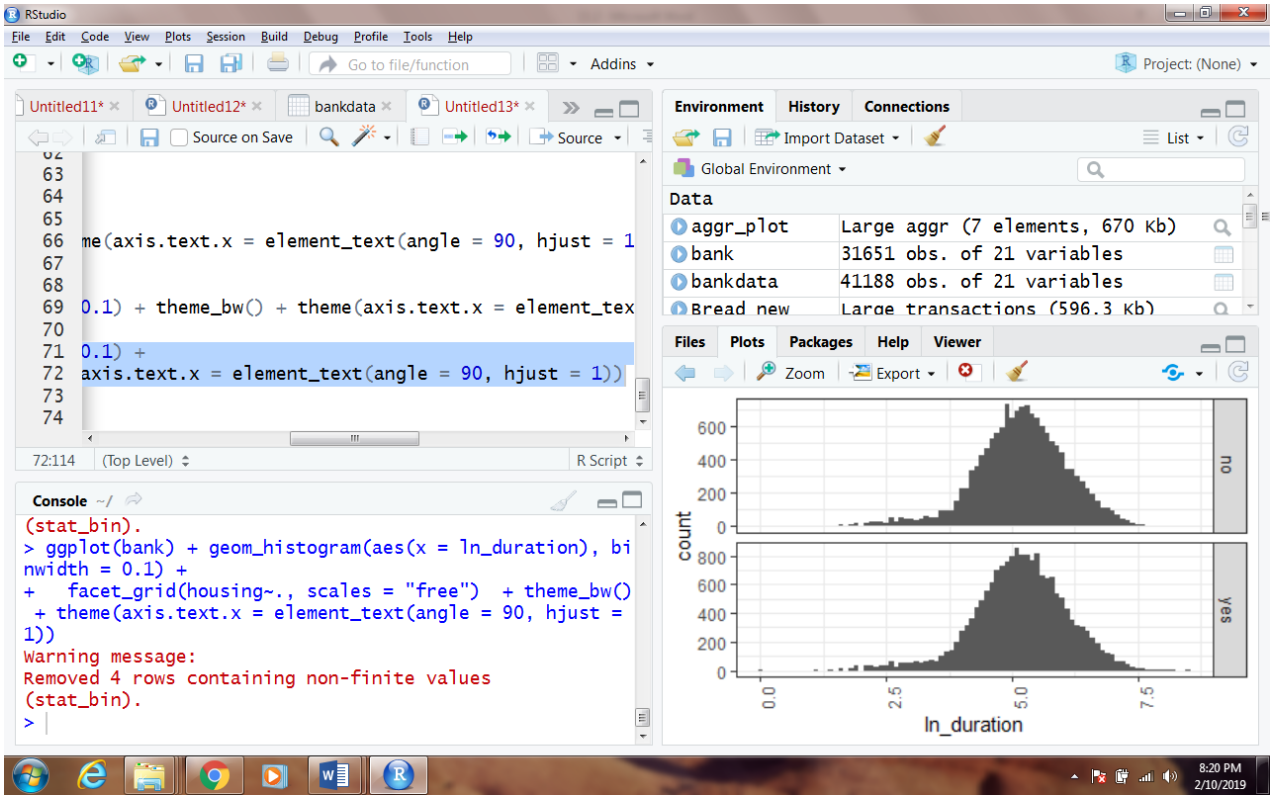
```
ln_duration <- log(bank$duration)
ggplot(bank, aes(x = ln_duration)) + geom_histogram(binwidth = 0.1) + theme_bw() +
theme(axis.text.x = element_text(angle = 90, hjust = 1))
```





ggplot(bank) + 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))

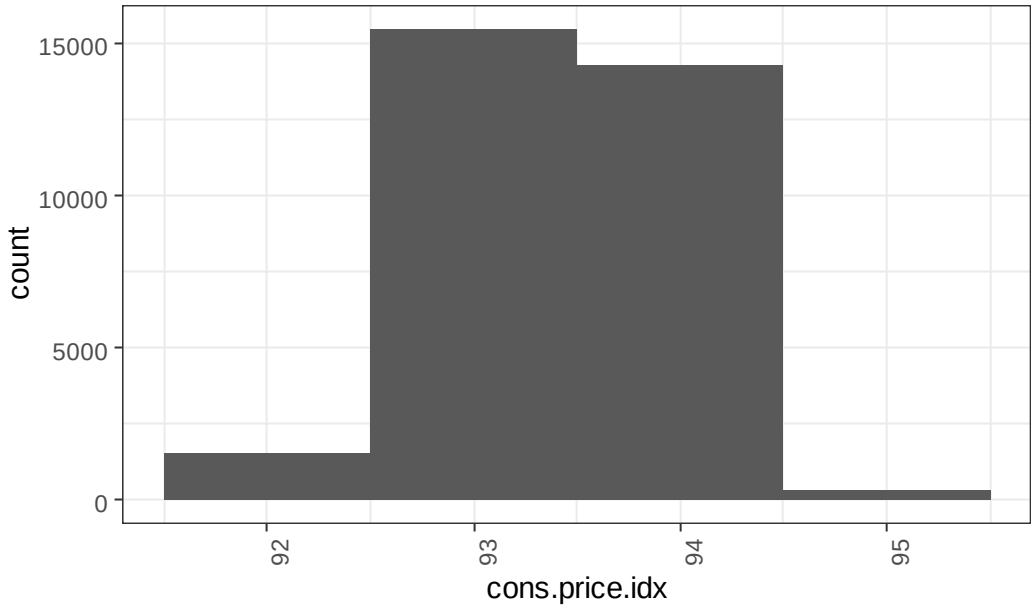


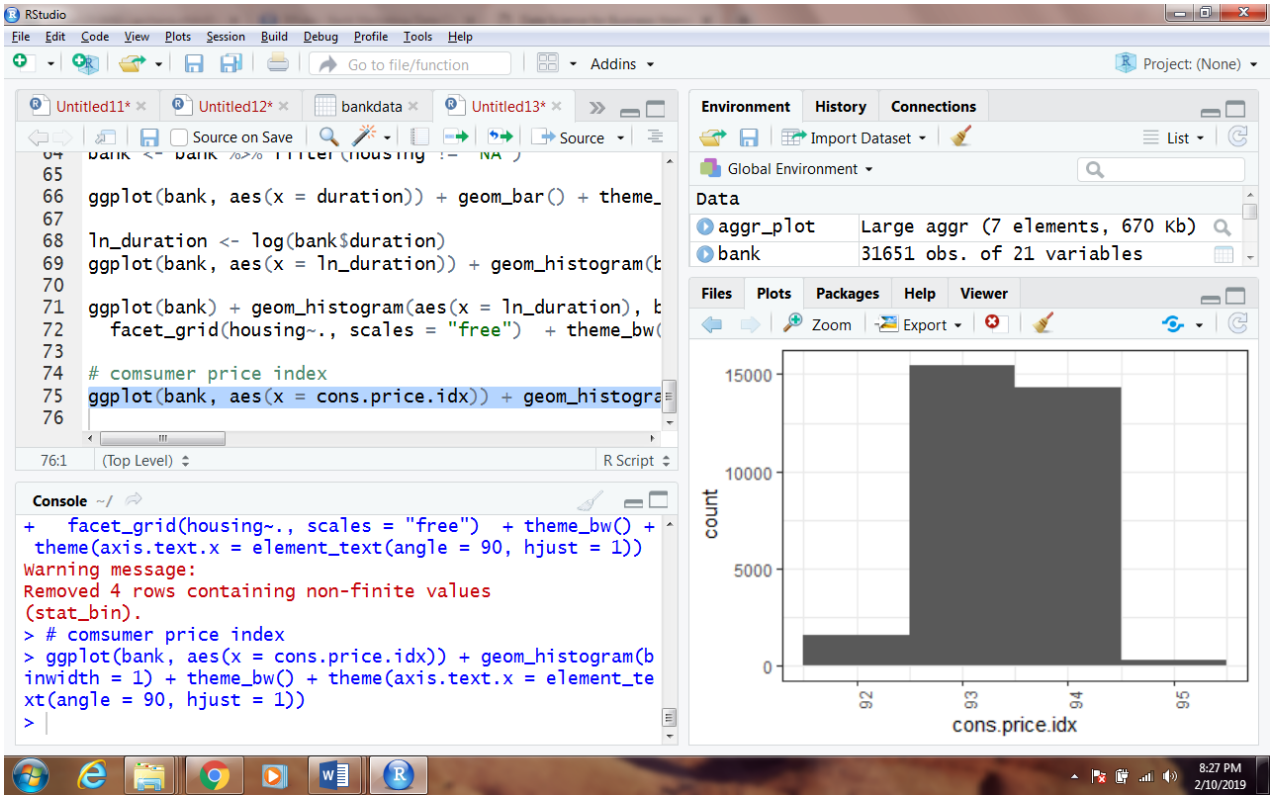


In call duration of 5 min 800 customers have taken home loan and 600 have not taken.

C- Is there any association between consumer price index and consumer

```
ggplot(bank, aes(x = cons.price.idx)) + geom_histogram(binwidth = 1) + theme_bw() + theme(axis.text.x =
element_text(angle = 90, hjust = 1))
```

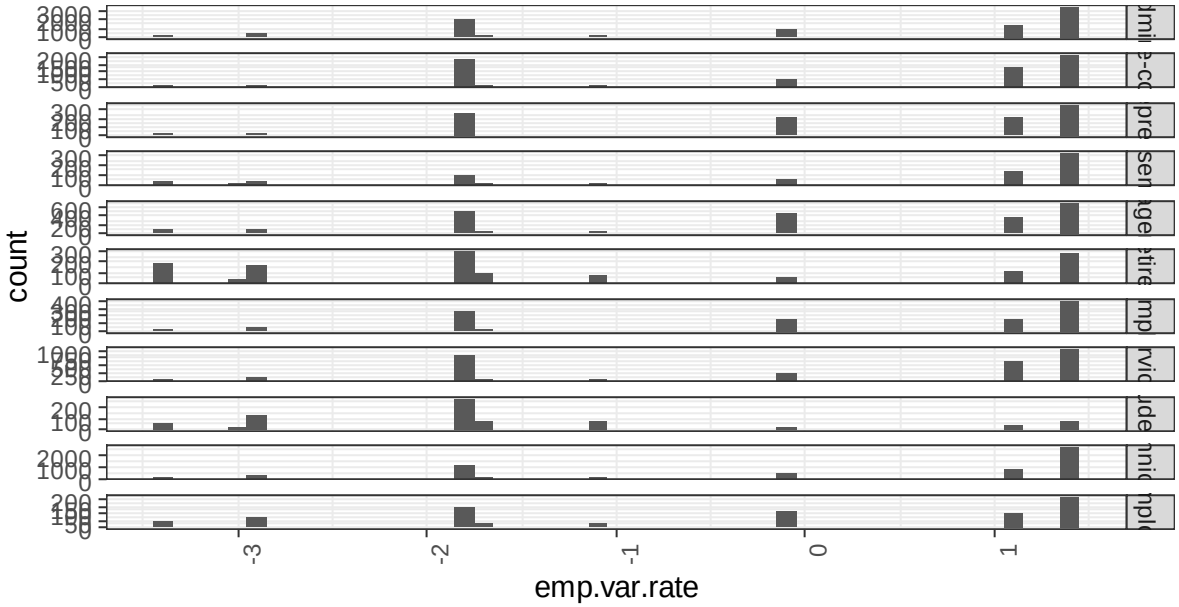




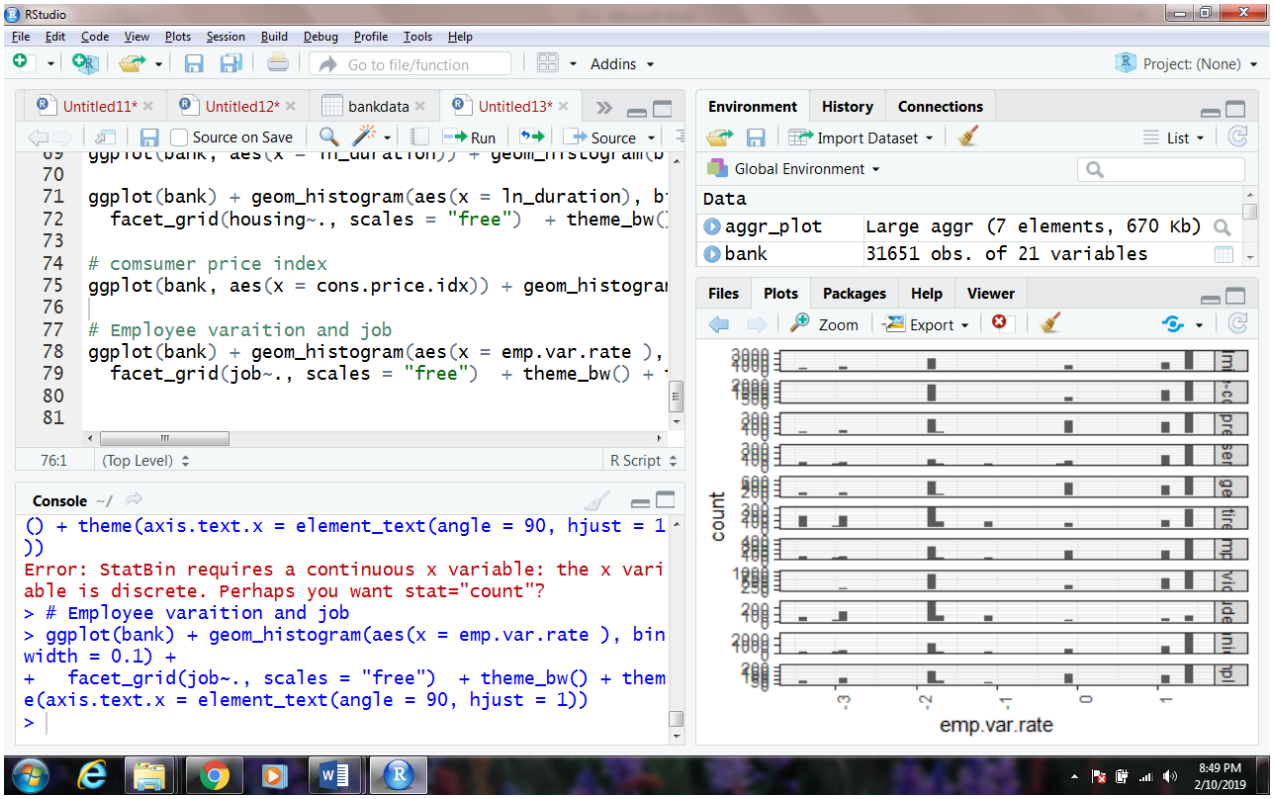
Maximum consumers have consumer price index between 92.5 and 94.5

D - Is the employment variation rate consistent across Job types?

```
ggplot(bank) + 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 =
```

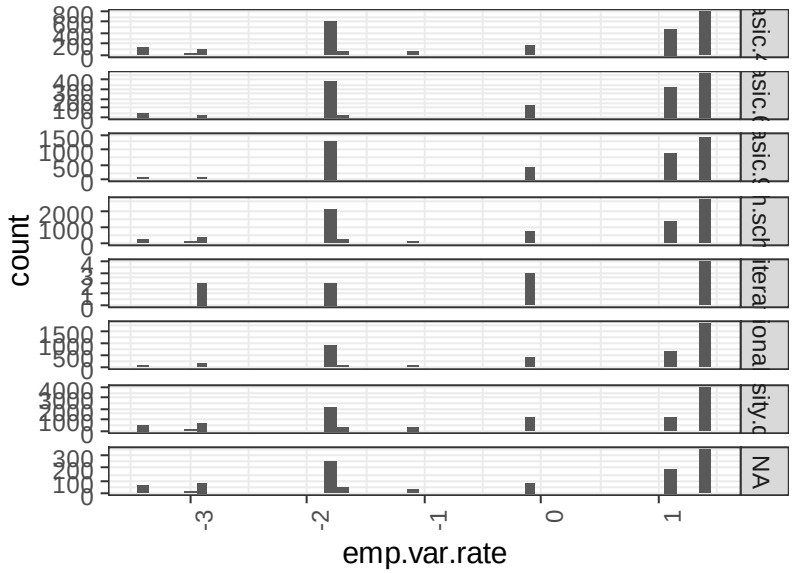


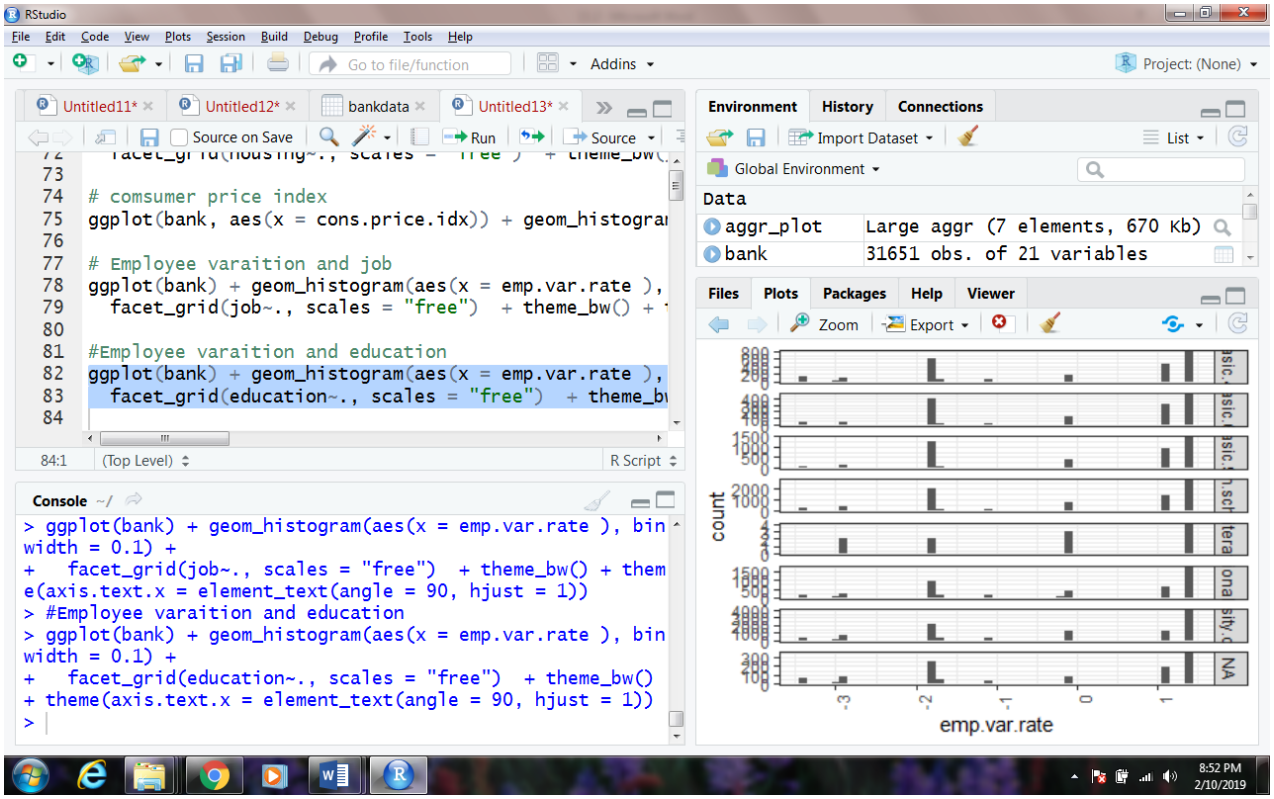
Yes employment variation rate consistent across Job types



Employee variation rate and education

ggplot(bank) + geom_histogram(aes(x = emp.var.rate), binwidth = 0.1) +
facet_grid(education~., scales = "free") + theme_bw() + theme(axis.text.x = element_text(angle = 90, hjust = 1))

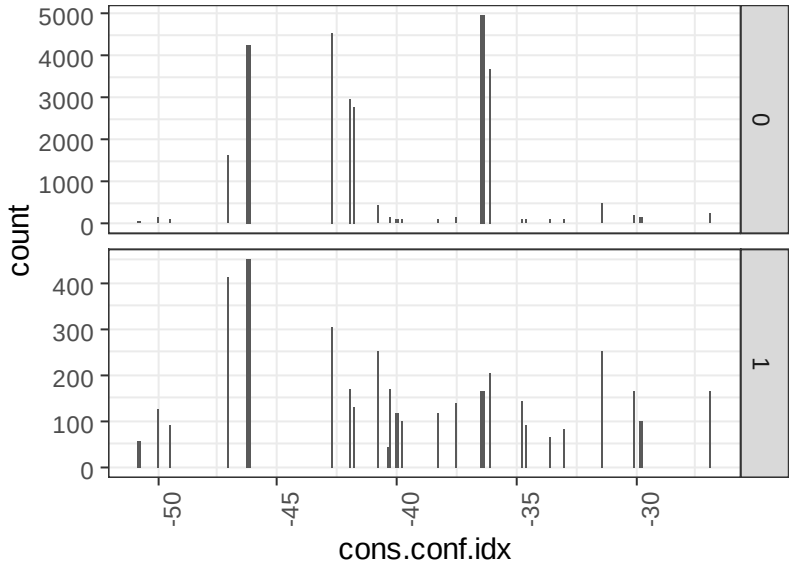

count
emp.var.rate
-3
-2
-1
0
1
NA


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Untitled11* Untitled12* bankdata Untitled13*
Source on Save Run Source
comsumer price index
ggplot(bank, aes(x = cons.price.idx)) + geom_histogra
Employee varaition and job
ggplot(bank) + geom_histogram(aes(x = emp.var.rate),
facet_grid(job~., scales = "free") + theme_bw() +
#Employee varaition and education
ggplot(bank) + geom_histogram(aes(x = emp.var.rate),
facet_grid(education~., scales = "free") + theme_bu
84:1 (Top Level) R Script
Console ~/
> ggplot(bank) + geom_histogram(aes(x = emp.var.rate), bin
width = 0.1) +
+ facet_grid(job~., scales = "free") + theme_bw() + them
e(axis.text.x = element_text(angle = 90, hjust = 1))
> #Employee varaition and education
> ggplot(bank) + geom_histogram(aes(x = emp.var.rate), bin
width = 0.1) +
+ facet_grid(education~., scales = "free") + theme_bw()
+ theme(axis.text.x = element_text(angle = 90, hjust = 1))
>
Environment History Connections
Import Dataset List
Global Environment
Data
aggr_plot Large aggr (7 elements, 670 Kb)
bank 31651 obs. of 21 variables
Files Plots Packages Help Viewer
Zoom Export
count
emp.var.rate
8:52 PM
2/10/2019

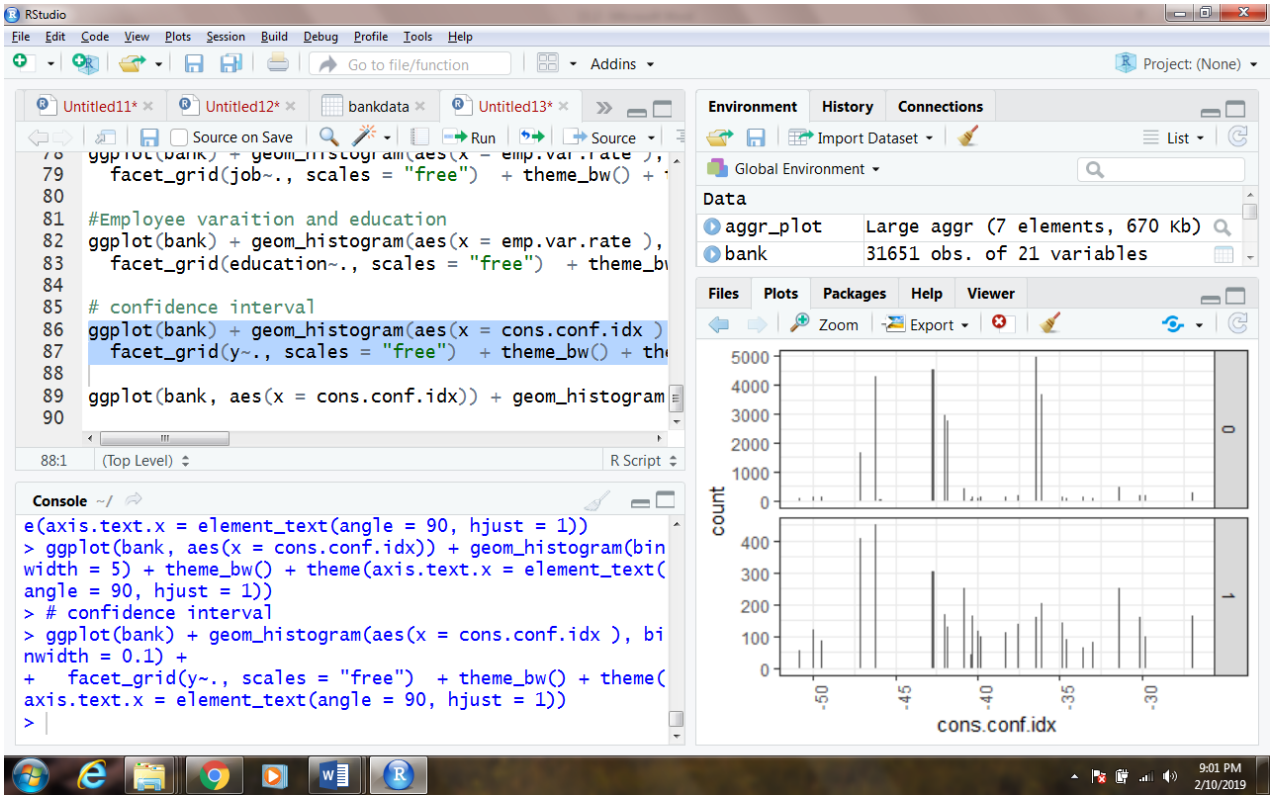
As per graph the employee variation rate is same for all education.

F - Which group is more confident?

```
ggplot(bank) + 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))
```

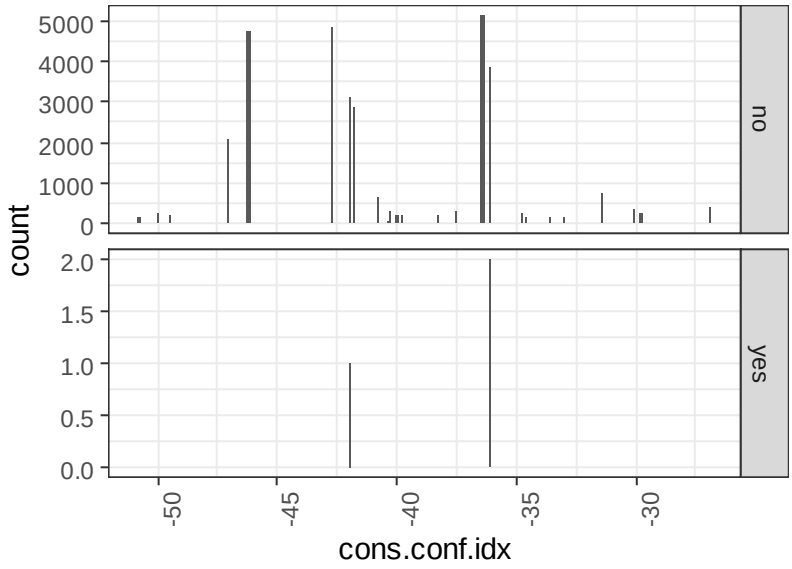


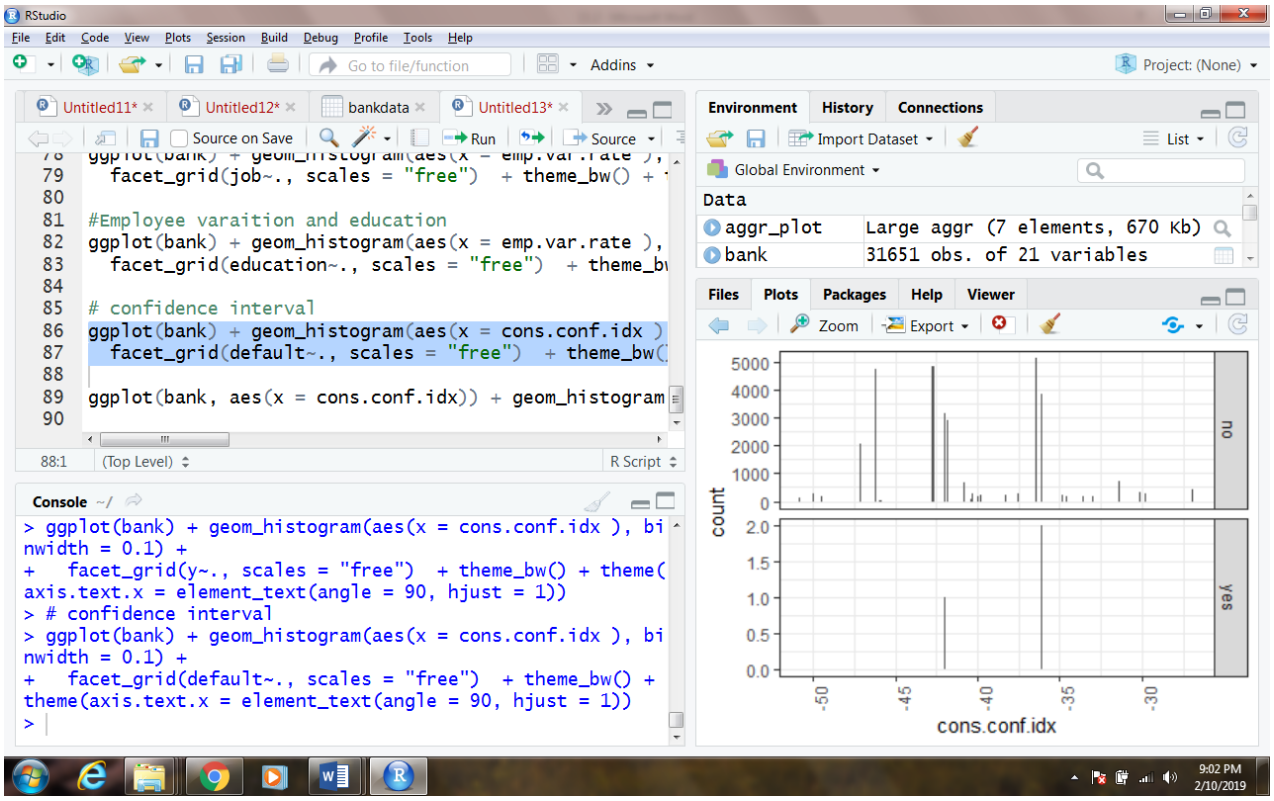
People who have not taken loan are more confident



6. –

ggplot(bank) + 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))


count
5000
4000
3000
2000
1000
0
no
2.0
1.5
1.0
0.5
0.0
yes
-50
-45
-40
-35
-30
cons.conf.idx


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Untitled11* Untitled12* bankdata Untitled13*
Source on Save Run Source
79 facet_grid(job~., scales = "free") + theme_bw() +
80
81 #Employee varaition and education
82 ggplot(bank) + geom_histogram(aes(x = emp.var.rate),
83 facet_grid(education~., scales = "free") + theme_bw
84
85 # confidence interval
86 ggplot(bank) + geom_histogram(aes(x = cons.conf.idx)
87 facet_grid(default~., scales = "free") + theme_bw()
88
89 ggplot(bank, aes(x = cons.conf.idx)) + geom_histogram
90
88:1 (Top Level) R Script
Console ~/
> ggplot(bank) + geom_histogram(aes(x = cons.conf.idx), bi
nwidth = 0.1) +
+ facet_grid(y~., scales = "free") + theme_bw() + theme(
axis.text.x = element_text(angle = 90, hjust = 1))
> # confidence interval
> ggplot(bank) + geom_histogram(aes(x = cons.conf.idx), bi
nwidth = 0.1) +
+ facet_grid(default~., scales = "free") + theme_bw() +
theme(axis.text.x = element_text(angle = 90, hjust = 1))
>
Environment History Connections
Import Dataset List
Global Environment
Data
aggr_plot Large aggr (7 elements, 670 Kb)
bank 31651 obs. of 21 variables
Files Plots Packages Help Viewer
Zoom Export
count
no
yes
cons.conf.idx
9:02 PM
2/10/2019

Non defaulters are more confident.

1. R file should be submitted where applicable.
2. R file should be in PDF or in .r format
3. Proper screenshots of the outputs should be submitted as well
4. The r codes, if submitted in any other format, will be subjected to deduction in marks

Note: Your solution will not be entertained if it is any other format, e.g., .zip, .doc, .rtf etc.

7. Approximate Time to Complete Task

20 mins.