

## ACADGILD ASSIGNMENT 11.2

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?

### ANSWER:

```
> library(readxl)
> bankdata <- read_excel("C:/Users/Veena/Desktop/bank-additional/bankdata.xlsx")
> View(bankdata)

";", escape_double = FALSE, trim_ws = TRUE)
#Lets look at dataset and generate initial understanding about the column types
str(bank_full)
#A deep check for NA in a particular column let say age
if(length(which(is.na(bank_full$age)==TRUE)>0)){
print("Missing Value found in the specified column")
} else
print("All okay: No Missing Value found in the specified column")
# Check another example say
if(length(which(is.na(bank_full$campaign)==TRUE)>0)){print("Missing Value found in the specified
column")} else
print("All okay: No Missing Value found in the specified column")
head(bank_full) ## Displays first 6 rows for each variable
str(bank_full) ## Describes each variables
summary(bank_full) ## Provides basic statistical information of each variable
## DATA EXPLORATION - Check for Missing Data
## Option 1
is.na(bank_full) ## Displays True for a missing value
## Since it is a large dataset, graphical display of missing values will prove to be easier
##Option 2
require(Amelia)
missmap(bank_full,main="Missing Data - Bank ", col=c("red","grey"),legend=FALSE)
## No red colour stripes are visible. hence no missing values.
summary(bank_full) ## displays missing values if any under every variable
```

#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.

#i. Pearson's chi-squared test of independence (significance test)

#### a. Is there any association between job and default?

```
with(bank_full, chisq.test( job, default))
with(bank_full, table( job, default ) )
# OR
with(bank_full, prop.table(table( job,default)))
#Pearson's Chi-squared test

data: job and default
X-squared = 60.343, df = 11, p-value = 8.008e-09
default
job no yes
admin. 5097 74
blue-collar 9531 201
entrepreneur 1432 55
housemaid 1218 22
management 9294 164
retired 2238 26
self-employed 1546 33
services 4079 75
student 935 3
technician 7467 130
unemployed 1273 30
unknown 286 2
default
job no yes
admin. 1.127381e-01 1.636770e-03
blue-collar 2.108115e-01 4.445821e-03
```

```
entrepreneur 3.167371e-02 1.216518e-03  
housemaid 2.694035e-02 4.866072e-04  
management 2.055694e-01 3.627436e-03  
retired 4.950123e-02 5.750813e-04  
self-employed 3.419522e-02 7.299109e-04  
services 9.022141e-02 1.658888e-03  
student 2.068081e-02 6.635553e-05  
technician 1.651589e-01 2.875406e-03  
p-value = 8.008e-09
```

#Pearson's Chi-squared test

#since the p-value is < 2.2e-16 is less than the cut-off value of 0.05, we can reject the null hypothesis in favor of alternative hypothesis and conclude, that the variables,( job & default- p-value = 8.008e-09) are dependent to each other.

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

```
with(bank_additional_full, chisq.test(duration,housing))  
with(bank_additional_full, table( duration,housing ) )  
# OR  
with(bank_additional_full, prop.table(table(duration, housing)))  
#data: duration and housing  
#X-squared = 3162.3, df = 3086, p-value = 0.1657  
#P value is above 0.05#
```

Chi-squared approximation may be incorrect

Pearson's Chi-squared test

data: duration and housing

X-squared = 3162.3, df = 3086, p-value = 0.1657

housing

duration no unknown yes

0 1 0 3

1 2 0 1

2 1 0 0

3 2 0 1

4 2 0 10

5 16 0 14

6 13 0 24

7 22 1 31

8 27 3 39

9 33 2 42

10 36 1 35

11 34 2 45

12 24 2 39

13 44 0 33

14 25 2 43

15 34 1 33  
16 35 1 44  
17 34 1 41  
18 43 1 40  
19 24 3 34  
20 29 0 32  
21 30 1 42  
22 35 2 39  
23 21 5 39  
24 30 2 32  
25 25 2 38  
26 23 1 37  
27 31 1 33  
28 25 0 24  
29 31 0 36  
30 17 2 35  
31 28 3 32  
32 20 1 21  
33 19 0 27  
34 31 3 34  
35 34 3 30  
36 42 1 39  
37 32 2 36  
38 26 2 32  
39 32 0 39  
40 26 2 38  
41 33 2 55  
42 35 1 45  
43 31 2 46  
44 33 3 49  
45 28 3 39  
46 37 3 37  
47 25 3 45  
48 43 3 46  
49 49 3 44  
50 41 3 52  
51 41 1 59  
52 48 4 50  
53 49 0 44  
54 46 2 58  
55 48 4 69  
56 51 2 57  
57 41 3 65  
58 44 0 66  
59 53 6 72  
60 47 1 57  
61 49 6 68

62 49 0 59  
63 55 1 71  
64 63 2 74  
65 57 2 64  
66 57 3 48  
67 65 6 76  
68 68 5 67  
69 63 3 64  
70 59 3 67  
71 76 9 67  
72 63 7 91  
73 74 8 85  
74 64 3 69  
75 65 0 73  
76 67 0 86  
77 56 3 87  
78 56 3 86  
79 71 0 67  
80 69 3 76  
81 69 6 72  
82 57 3 94  
83 59 6 84  
84 66 6 59  
85 85 1 84  
86 55 3 70  
87 70 3 89  
88 64 5 81  
89 68 3 82  
90 73 3 94  
91 75 2 70  
92 72 7 77  
93 76 6 69  
94 66 2 68  
95 63 5 80  
96 86 2 67  
97 79 3 76  
98 70 2 72  
99 55 3 71  
100 61 2 82  
101 62 3 80  
102 64 4 73  
103 68 1 78  
104 70 10 81  
105 62 4 79  
106 67 4 88  
107 60 6 84  
108 62 2 73

109 66 1 91  
110 71 3 68  
111 72 2 86  
112 61 3 80  
113 62 1 79  
114 74 6 76  
115 58 1 76  
116 62 5 59  
117 64 2 68  
118 49 1 81  
119 72 4 79  
120 52 2 68  
121 61 4 76  
122 69 4 84  
123 69 6 74  
124 65 5 94  
125 67 5 80  
126 62 2 88  
127 68 3 75  
128 70 4 73  
129 64 2 70  
130 62 4 76  
131 69 5 60  
132 56 3 53  
133 57 4 82  
134 74 1 58  
135 78 9 69  
136 77 8 83  
137 61 6 56  
138 51 4 64  
139 87 5 63  
140 64 2 79  
141 52 4 62  
142 57 1 58  
143 63 5 65  
144 64 2 71  
145 59 3 68  
146 54 3 45  
147 56 7 68  
148 51 6 66  
149 51 0 68  
150 49 2 82  
151 56 3 70  
152 49 2 64  
153 47 6 67  
154 62 8 66  
155 63 5 68

156 47 5 76  
157 71 4 70  
158 43 4 82  
159 65 4 71  
160 48 4 85  
161 57 3 69  
162 46 3 57  
163 46 3 64  
164 55 4 71  
165 66 3 67  
166 73 1 68  
167 44 4 57  
168 59 1 71  
169 50 3 53  
170 37 3 60  
171 59 3 53  
172 53 2 68  
173 51 3 71  
174 57 2 63  
175 48 2 65  
176 39 4 58  
177 47 3 56  
178 59 5 63  
179 46 7 55  
180 58 3 56  
181 49 1 63  
182 39 5 57  
183 39 1 50  
184 56 3 54  
185 56 3 55  
186 37 1 48  
187 54 2 52  
188 58 1 55  
189 63 2 49  
190 54 2 40  
191 54 2 53  
192 49 4 51  
193 45 3 58  
194 40 0 54  
195 31 3 54  
196 40 0 48  
197 50 0 57  
198 56 2 49  
199 49 2 44  
200 60 1 47  
201 41 2 67  
202 45 4 47

203 50 3 40  
204 47 2 42  
205 41 0 46  
206 50 1 41  
207 54 2 58  
208 46 3 41  
209 45 2 47  
210 42 2 56  
211 50 1 57  
212 44 3 46  
213 47 3 40  
214 39 2 50  
215 38 4 49  
216 33 3 44  
217 41 2 49  
218 36 1 39  
219 40 0 45  
220 37 1 48  
221 45 1 43  
222 52 2 39  
223 38 0 31  
224 30 1 42  
225 42 3 45  
226 29 3 61  
227 39 4 39  
228 28 1 44  
229 26 3 40  
230 36 6 40  
231 34 0 45  
232 38 1 36  
233 26 1 34  
234 37 4 37  
235 29 3 30  
236 27 1 44  
237 31 1 34  
238 32 1 43  
239 35 2 55  
240 31 4 52  
241 36 1 37  
242 23 3 32  
243 34 0 45  
244 34 0 41  
245 49 2 40  
246 39 3 37  
247 31 2 52  
248 24 3 43  
249 33 2 41

250 40 1 39  
251 34 1 48  
252 36 1 43  
253 31 1 33  
254 27 4 41  
255 38 1 33  
256 25 1 39  
257 36 3 32  
258 29 1 38  
259 33 2 36  
260 23 0 38  
261 29 1 32  
262 35 3 34  
263 37 1 36  
264 31 0 30  
265 23 0 43  
266 26 0 32  
267 33 1 38  
268 35 4 32  
269 23 0 31  
270 38 1 18  
271 27 5 25  
272 26 5 36  
273 23 0 37  
274 23 1 32  
275 21 1 34  
276 23 2 24  
277 33 1 29  
278 27 1 25  
279 19 2 22  
280 27 0 30  
281 30 0 31  
282 21 4 31  
283 31 1 30  
284 27 0 26  
285 29 1 30  
286 19 0 29  
287 25 0 21  
288 24 2 31  
289 23 2 28  
290 29 1 26  
291 29 2 28  
292 18 0 35  
293 28 0 34  
294 24 1 35  
295 35 4 28  
296 33 1 23

```
297 28 3 32
298 29 1 25
299 22 1 24
300 18 0 28
301 35 2 25
302 20 1 25
303 16 0 23
304 25 0 29
305 26 1 35
306 23 0 35
307 18 0 19
308 16 4 21
309 16 1 27
310 10 0 23
311 24 5 16
312 25 1 23
313 20 3 27
314 27 1 27
315 18 2 19
316 23 2 20
317 21 3 33
318 18 1 37
319 28 2 16
320 16 1 30
321 14 0 21
322 31 1 25
323 20 1 27
324 20 0 33
325 17 0 25
326 19 0 31
327 24 1 16
328 19 1 29
329 25 1 34
330 10 0 18
331 18 1 16
332 27 2 24
[ reached getOption("max.print") -- omitted 1211 rows ]
housing
duration no unknown yes
0 2.427892e-05 0.000000e+00 7.283675e-05
1 4.855783e-05 0.000000e+00 2.427892e-05
2 2.427892e-05 0.000000e+00 0.000000e+00
3 4.855783e-05 0.000000e+00 2.427892e-05
4 4.855783e-05 0.000000e+00 2.427892e-04
5 3.884627e-04 0.000000e+00 3.399048e-04
6 3.156259e-04 0.000000e+00 5.826940e-04
7 5.341362e-04 2.427892e-05 7.526464e-04
```

8 6.555307e-04 7.283675e-05 9.468777e-04  
9 8.012042e-04 4.855783e-05 1.019714e-03  
10 8.740410e-04 2.427892e-05 8.497621e-04  
11 8.254832e-04 4.855783e-05 1.092551e-03  
12 5.826940e-04 4.855783e-05 9.468777e-04  
13 1.068272e-03 0.000000e+00 8.012042e-04  
14 6.069729e-04 4.855783e-05 1.043993e-03  
15 8.254832e-04 2.427892e-05 8.012042e-04  
16 8.497621e-04 2.427892e-05 1.068272e-03  
17 8.254832e-04 2.427892e-05 9.954356e-04  
18 1.043993e-03 2.427892e-05 9.711566e-04  
19 5.826940e-04 7.283675e-05 8.254832e-04  
20 7.040886e-04 0.000000e+00 7.769253e-04  
21 7.283675e-04 2.427892e-05 1.019714e-03  
22 8.497621e-04 4.855783e-05 9.468777e-04  
23 5.098572e-04 1.213946e-04 9.468777e-04  
24 7.283675e-04 4.855783e-05 7.769253e-04  
25 6.069729e-04 4.855783e-05 9.225988e-04  
26 5.584151e-04 2.427892e-05 8.983199e-04  
27 7.526464e-04 2.427892e-05 8.012042e-04  
28 6.069729e-04 0.000000e+00 5.826940e-04  
29 7.526464e-04 0.000000e+00 8.740410e-04  
30 4.127416e-04 4.855783e-05 8.497621e-04  
31 6.798097e-04 7.283675e-05 7.769253e-04  
32 4.855783e-04 2.427892e-05 5.098572e-04  
33 4.612994e-04 0.000000e+00 6.555307e-04  
34 7.526464e-04 7.283675e-05 8.254832e-04  
35 8.254832e-04 7.283675e-05 7.283675e-04  
36 1.019714e-03 2.427892e-05 9.468777e-04  
37 7.769253e-04 4.855783e-05 8.740410e-04  
38 6.312518e-04 4.855783e-05 7.769253e-04  
39 7.769253e-04 0.000000e+00 9.468777e-04  
40 6.312518e-04 4.855783e-05 9.225988e-04  
41 8.012042e-04 4.855783e-05 1.335340e-03  
42 8.497621e-04 2.427892e-05 1.092551e-03  
43 7.526464e-04 4.855783e-05 1.116830e-03  
44 8.012042e-04 7.283675e-05 1.189667e-03  
45 6.798097e-04 7.283675e-05 9.468777e-04  
46 8.983199e-04 7.283675e-05 8.983199e-04  
47 6.069729e-04 7.283675e-05 1.092551e-03  
48 1.043993e-03 7.283675e-05 1.116830e-03  
49 1.189667e-03 7.283675e-05 1.068272e-03  
50 9.954356e-04 7.283675e-05 1.262504e-03  
51 9.954356e-04 2.427892e-05 1.432456e-03  
52 1.165388e-03 9.711566e-05 1.213946e-03  
53 1.189667e-03 0.000000e+00 1.068272e-03  
54 1.116830e-03 4.855783e-05 1.408177e-03

55 1.165388e-03 9.711566e-05 1.675245e-03  
56 1.238225e-03 4.855783e-05 1.383898e-03  
57 9.954356e-04 7.283675e-05 1.578130e-03  
58 1.068272e-03 0.000000e+00 1.602408e-03  
59 1.286783e-03 1.456735e-04 1.748082e-03  
60 1.141109e-03 2.427892e-05 1.383898e-03  
61 1.189667e-03 1.456735e-04 1.650966e-03  
62 1.189667e-03 0.000000e+00 1.432456e-03  
63 1.335340e-03 2.427892e-05 1.723803e-03  
64 1.529572e-03 4.855783e-05 1.796640e-03  
65 1.383898e-03 4.855783e-05 1.553851e-03  
66 1.383898e-03 7.283675e-05 1.165388e-03  
67 1.578130e-03 1.456735e-04 1.845198e-03  
68 1.650966e-03 1.213946e-04 1.626687e-03  
69 1.529572e-03 7.283675e-05 1.553851e-03  
70 1.432456e-03 7.283675e-05 1.626687e-03  
71 1.845198e-03 2.185102e-04 1.626687e-03  
72 1.529572e-03 1.699524e-04 2.209381e-03  
73 1.796640e-03 1.942313e-04 2.063708e-03  
74 1.553851e-03 7.283675e-05 1.675245e-03  
75 1.578130e-03 0.000000e+00 1.772361e-03  
76 1.626687e-03 0.000000e+00 2.087987e-03  
77 1.359619e-03 7.283675e-05 2.112266e-03  
78 1.359619e-03 7.283675e-05 2.087987e-03  
79 1.723803e-03 0.000000e+00 1.626687e-03  
80 1.675245e-03 7.283675e-05 1.845198e-03  
81 1.675245e-03 1.456735e-04 1.748082e-03  
82 1.383898e-03 7.283675e-05 2.282218e-03  
83 1.432456e-03 1.456735e-04 2.039429e-03  
84 1.602408e-03 1.456735e-04 1.432456e-03  
85 2.063708e-03 2.427892e-05 2.039429e-03  
86 1.335340e-03 7.283675e-05 1.699524e-03  
87 1.699524e-03 7.283675e-05 2.160824e-03  
88 1.553851e-03 1.213946e-04 1.966592e-03  
89 1.650966e-03 7.283675e-05 1.990871e-03  
90 1.772361e-03 7.283675e-05 2.282218e-03  
91 1.820919e-03 4.855783e-05 1.699524e-03  
92 1.748082e-03 1.699524e-04 1.869477e-03  
93 1.845198e-03 1.456735e-04 1.675245e-03  
94 1.602408e-03 4.855783e-05 1.650966e-03  
95 1.529572e-03 1.213946e-04 1.942313e-03  
96 2.087987e-03 4.855783e-05 1.626687e-03  
97 1.918034e-03 7.283675e-05 1.845198e-03  
98 1.699524e-03 4.855783e-05 1.748082e-03  
99 1.335340e-03 7.283675e-05 1.723803e-03  
100 1.481014e-03 4.855783e-05 1.990871e-03  
101 1.505293e-03 7.283675e-05 1.942313e-03

102 1.553851e-03 9.711566e-05 1.772361e-03  
103 1.650966e-03 2.427892e-05 1.893755e-03  
104 1.699524e-03 2.427892e-04 1.966592e-03  
105 1.505293e-03 9.711566e-05 1.918034e-03  
106 1.626687e-03 9.711566e-05 2.136545e-03  
107 1.456735e-03 1.456735e-04 2.039429e-03  
108 1.505293e-03 4.855783e-05 1.772361e-03  
109 1.602408e-03 2.427892e-05 2.209381e-03  
110 1.723803e-03 7.283675e-05 1.650966e-03  
111 1.748082e-03 4.855783e-05 2.087987e-03  
112 1.481014e-03 7.283675e-05 1.942313e-03  
113 1.505293e-03 2.427892e-05 1.918034e-03  
114 1.796640e-03 1.456735e-04 1.845198e-03  
115 1.408177e-03 2.427892e-05 1.845198e-03  
116 1.505293e-03 1.213946e-04 1.432456e-03  
117 1.553851e-03 4.855783e-05 1.650966e-03  
118 1.189667e-03 2.427892e-05 1.966592e-03  
119 1.748082e-03 9.711566e-05 1.918034e-03  
120 1.262504e-03 4.855783e-05 1.650966e-03  
121 1.481014e-03 9.711566e-05 1.845198e-03  
122 1.675245e-03 9.711566e-05 2.039429e-03  
123 1.675245e-03 1.456735e-04 1.796640e-03  
124 1.578130e-03 1.213946e-04 2.282218e-03  
125 1.626687e-03 1.213946e-04 1.942313e-03  
126 1.505293e-03 4.855783e-05 2.136545e-03  
127 1.650966e-03 7.283675e-05 1.820919e-03  
128 1.699524e-03 9.711566e-05 1.772361e-03  
129 1.553851e-03 4.855783e-05 1.699524e-03  
130 1.505293e-03 9.711566e-05 1.845198e-03  
131 1.675245e-03 1.213946e-04 1.456735e-03  
132 1.359619e-03 7.283675e-05 1.286783e-03  
133 1.383898e-03 9.711566e-05 1.990871e-03  
134 1.796640e-03 2.427892e-05 1.408177e-03  
135 1.893755e-03 2.185102e-04 1.675245e-03  
136 1.869477e-03 1.942313e-04 2.015150e-03  
137 1.481014e-03 1.456735e-04 1.359619e-03  
138 1.238225e-03 9.711566e-05 1.553851e-03  
139 2.112266e-03 1.213946e-04 1.529572e-03  
140 1.553851e-03 4.855783e-05 1.918034e-03  
141 1.262504e-03 9.711566e-05 1.505293e-03  
142 1.383898e-03 2.427892e-05 1.408177e-03  
143 1.529572e-03 1.213946e-04 1.578130e-03  
144 1.553851e-03 4.855783e-05 1.723803e-03  
145 1.432456e-03 7.283675e-05 1.650966e-03  
146 1.311061e-03 7.283675e-05 1.092551e-03  
147 1.359619e-03 1.699524e-04 1.650966e-03  
148 1.238225e-03 1.456735e-04 1.602408e-03

149 1.238225e-03 0.000000e+00 1.650966e-03  
150 1.189667e-03 4.855783e-05 1.990871e-03  
151 1.359619e-03 7.283675e-05 1.699524e-03  
152 1.189667e-03 4.855783e-05 1.553851e-03  
153 1.141109e-03 1.456735e-04 1.626687e-03  
154 1.505293e-03 1.942313e-04 1.602408e-03  
155 1.529572e-03 1.213946e-04 1.650966e-03  
156 1.141109e-03 1.213946e-04 1.845198e-03  
157 1.723803e-03 9.711566e-05 1.699524e-03  
158 1.043993e-03 9.711566e-05 1.990871e-03  
159 1.578130e-03 9.711566e-05 1.723803e-03  
160 1.165388e-03 9.711566e-05 2.063708e-03  
161 1.383898e-03 7.283675e-05 1.675245e-03  
162 1.116830e-03 7.283675e-05 1.383898e-03  
163 1.116830e-03 7.283675e-05 1.553851e-03  
164 1.335340e-03 9.711566e-05 1.723803e-03  
165 1.602408e-03 7.283675e-05 1.626687e-03  
166 1.772361e-03 2.427892e-05 1.650966e-03  
167 1.068272e-03 9.711566e-05 1.383898e-03  
168 1.432456e-03 2.427892e-05 1.723803e-03  
169 1.213946e-03 7.283675e-05 1.286783e-03  
170 8.983199e-04 7.283675e-05 1.456735e-03  
171 1.432456e-03 7.283675e-05 1.286783e-03  
172 1.286783e-03 4.855783e-05 1.650966e-03  
173 1.238225e-03 7.283675e-05 1.723803e-03  
174 1.383898e-03 4.855783e-05 1.529572e-03  
175 1.165388e-03 4.855783e-05 1.578130e-03  
176 9.468777e-04 9.711566e-05 1.408177e-03  
177 1.141109e-03 7.283675e-05 1.359619e-03  
178 1.432456e-03 1.213946e-04 1.529572e-03  
179 1.116830e-03 1.699524e-04 1.335340e-03  
180 1.408177e-03 7.283675e-05 1.359619e-03  
181 1.189667e-03 2.427892e-05 1.529572e-03  
182 9.468777e-04 1.213946e-04 1.383898e-03  
183 9.468777e-04 2.427892e-05 1.213946e-03  
184 1.359619e-03 7.283675e-05 1.311061e-03  
185 1.359619e-03 7.283675e-05 1.335340e-03  
186 8.983199e-04 2.427892e-05 1.165388e-03  
187 1.311061e-03 4.855783e-05 1.262504e-03  
188 1.408177e-03 2.427892e-05 1.335340e-03  
189 1.529572e-03 4.855783e-05 1.189667e-03  
190 1.311061e-03 4.855783e-05 9.711566e-04  
191 1.311061e-03 4.855783e-05 1.286783e-03  
192 1.189667e-03 9.711566e-05 1.238225e-03  
193 1.092551e-03 7.283675e-05 1.408177e-03  
194 9.711566e-04 0.000000e+00 1.311061e-03  
195 7.526464e-04 7.283675e-05 1.311061e-03

196 9.711566e-04 0.000000e+00 1.165388e-03  
197 1.213946e-03 0.000000e+00 1.383898e-03  
198 1.359619e-03 4.855783e-05 1.189667e-03  
199 1.189667e-03 4.855783e-05 1.068272e-03  
200 1.456735e-03 2.427892e-05 1.141109e-03  
201 9.954356e-04 4.855783e-05 1.626687e-03  
202 1.092551e-03 9.711566e-05 1.141109e-03  
203 1.213946e-03 7.283675e-05 9.711566e-04  
204 1.141109e-03 4.855783e-05 1.019714e-03  
205 9.954356e-04 0.000000e+00 1.116830e-03  
206 1.213946e-03 2.427892e-05 9.954356e-04  
207 1.311061e-03 4.855783e-05 1.408177e-03  
208 1.116830e-03 7.283675e-05 9.954356e-04  
209 1.092551e-03 4.855783e-05 1.141109e-03  
210 1.019714e-03 4.855783e-05 1.359619e-03  
211 1.213946e-03 2.427892e-05 1.383898e-03  
212 1.068272e-03 7.283675e-05 1.116830e-03  
213 1.141109e-03 7.283675e-05 9.711566e-04  
214 9.468777e-04 4.855783e-05 1.213946e-03  
215 9.225988e-04 9.711566e-05 1.189667e-03  
216 8.012042e-04 7.283675e-05 1.068272e-03  
217 9.954356e-04 4.855783e-05 1.189667e-03  
218 8.740410e-04 2.427892e-05 9.468777e-04  
219 9.711566e-04 0.000000e+00 1.092551e-03  
220 8.983199e-04 2.427892e-05 1.165388e-03  
221 1.092551e-03 2.427892e-05 1.043993e-03  
222 1.262504e-03 4.855783e-05 9.468777e-04  
223 9.225988e-04 0.000000e+00 7.526464e-04  
224 7.283675e-04 2.427892e-05 1.019714e-03  
225 1.019714e-03 7.283675e-05 1.092551e-03  
226 7.040886e-04 7.283675e-05 1.481014e-03  
227 9.468777e-04 9.711566e-05 9.468777e-04  
228 6.798097e-04 2.427892e-05 1.068272e-03  
229 6.312518e-04 7.283675e-05 9.711566e-04  
230 8.740410e-04 1.456735e-04 9.711566e-04  
231 8.254832e-04 0.000000e+00 1.092551e-03  
232 9.225988e-04 2.427892e-05 8.740410e-04  
233 6.312518e-04 2.427892e-05 8.254832e-04  
234 8.983199e-04 9.711566e-05 8.983199e-04  
235 7.040886e-04 7.283675e-05 7.283675e-04  
236 6.555307e-04 2.427892e-05 1.068272e-03  
237 7.526464e-04 2.427892e-05 8.254832e-04  
238 7.769253e-04 2.427892e-05 1.043993e-03  
239 8.497621e-04 4.855783e-05 1.335340e-03  
240 7.526464e-04 9.711566e-05 1.262504e-03  
241 8.740410e-04 2.427892e-05 8.983199e-04  
242 5.584151e-04 7.283675e-05 7.769253e-04

243 8.254832e-04 0.000000e+00 1.092551e-03  
244 8.254832e-04 0.000000e+00 9.954356e-04  
245 1.189667e-03 4.855783e-05 9.711566e-04  
246 9.468777e-04 7.283675e-05 8.983199e-04  
247 7.526464e-04 4.855783e-05 1.262504e-03  
248 5.826940e-04 7.283675e-05 1.043993e-03  
249 8.012042e-04 4.855783e-05 9.954356e-04  
250 9.711566e-04 2.427892e-05 9.468777e-04  
251 8.254832e-04 2.427892e-05 1.165388e-03  
252 8.740410e-04 2.427892e-05 1.043993e-03  
253 7.526464e-04 2.427892e-05 8.012042e-04  
254 6.555307e-04 9.711566e-05 9.954356e-04  
255 9.225988e-04 2.427892e-05 8.012042e-04  
256 6.069729e-04 2.427892e-05 9.468777e-04  
257 8.740410e-04 7.283675e-05 7.769253e-04  
258 7.040886e-04 2.427892e-05 9.225988e-04  
259 8.012042e-04 4.855783e-05 8.740410e-04  
260 5.584151e-04 0.000000e+00 9.225988e-04  
261 7.040886e-04 2.427892e-05 7.769253e-04  
262 8.497621e-04 7.283675e-05 8.254832e-04  
263 8.983199e-04 2.427892e-05 8.740410e-04  
264 7.526464e-04 0.000000e+00 7.283675e-04  
265 5.584151e-04 0.000000e+00 1.043993e-03  
266 6.312518e-04 0.000000e+00 7.769253e-04  
267 8.012042e-04 2.427892e-05 9.225988e-04  
268 8.497621e-04 9.711566e-05 7.769253e-04  
269 5.584151e-04 0.000000e+00 7.526464e-04  
270 9.225988e-04 2.427892e-05 4.370205e-04  
271 6.555307e-04 1.213946e-04 6.069729e-04  
272 6.312518e-04 1.213946e-04 8.740410e-04  
273 5.584151e-04 0.000000e+00 8.983199e-04  
274 5.584151e-04 2.427892e-05 7.769253e-04  
275 5.098572e-04 2.427892e-05 8.254832e-04  
276 5.584151e-04 4.855783e-05 5.826940e-04  
277 8.012042e-04 2.427892e-05 7.040886e-04  
278 6.555307e-04 2.427892e-05 6.069729e-04  
279 4.612994e-04 4.855783e-05 5.341362e-04  
280 6.555307e-04 0.000000e+00 7.283675e-04  
281 7.283675e-04 0.000000e+00 7.526464e-04  
282 5.098572e-04 9.711566e-05 7.526464e-04  
283 7.526464e-04 2.427892e-05 7.283675e-04  
284 6.555307e-04 0.000000e+00 6.312518e-04  
285 7.040886e-04 2.427892e-05 7.283675e-04  
286 4.612994e-04 0.000000e+00 7.040886e-04  
287 6.069729e-04 0.000000e+00 5.098572e-04  
288 5.826940e-04 4.855783e-05 7.526464e-04  
289 5.584151e-04 4.855783e-05 6.798097e-04

290 7.040886e-04 2.427892e-05 6.312518e-04  
291 7.040886e-04 4.855783e-05 6.798097e-04  
292 4.370205e-04 0.000000e+00 8.497621e-04  
293 6.798097e-04 0.000000e+00 8.254832e-04  
294 5.826940e-04 2.427892e-05 8.497621e-04  
295 8.497621e-04 9.711566e-05 6.798097e-04  
296 8.012042e-04 2.427892e-05 5.584151e-04  
297 6.798097e-04 7.283675e-05 7.769253e-04  
298 7.040886e-04 2.427892e-05 6.069729e-04  
299 5.341362e-04 2.427892e-05 5.826940e-04  
300 4.370205e-04 0.000000e+00 6.798097e-04  
301 8.497621e-04 4.855783e-05 6.069729e-04  
302 4.855783e-04 2.427892e-05 6.069729e-04  
303 3.884627e-04 0.000000e+00 5.584151e-04  
304 6.069729e-04 0.000000e+00 7.040886e-04  
305 6.312518e-04 2.427892e-05 8.497621e-04  
306 5.584151e-04 0.000000e+00 8.497621e-04  
307 4.370205e-04 0.000000e+00 4.612994e-04  
308 3.884627e-04 9.711566e-05 5.098572e-04  
309 3.884627e-04 2.427892e-05 6.555307e-04  
310 2.427892e-04 0.000000e+00 5.584151e-04  
311 5.826940e-04 1.213946e-04 3.884627e-04  
312 6.069729e-04 2.427892e-05 5.584151e-04  
313 4.855783e-04 7.283675e-05 6.555307e-04  
314 6.555307e-04 2.427892e-05 6.555307e-04  
315 4.370205e-04 4.855783e-05 4.612994e-04  
316 5.584151e-04 4.855783e-05 4.855783e-04  
317 5.098572e-04 7.283675e-05 8.012042e-04  
318 4.370205e-04 2.427892e-05 8.983199e-04  
319 6.798097e-04 4.855783e-05 3.884627e-04  
320 3.884627e-04 2.427892e-05 7.283675e-04  
321 3.399048e-04 0.000000e+00 5.098572e-04  
322 7.526464e-04 2.427892e-05 6.069729e-04  
323 4.855783e-04 2.427892e-05 6.555307e-04  
324 4.855783e-04 0.000000e+00 8.012042e-04  
325 4.127416e-04 0.000000e+00 6.069729e-04  
326 4.612994e-04 0.000000e+00 7.526464e-04  
327 5.826940e-04 2.427892e-05 3.884627e-04  
328 4.612994e-04 2.427892e-05 7.040886e-04  
329 6.069729e-04 2.427892e-05 8.254832e-04  
330 2.427892e-04 0.000000e+00 4.370205e-04  
331 4.370205e-04 2.427892e-05 3.884627e-04  
332 6.555307e-04 4.855783e-05 5.826940e-04  
[ reached getOption("max.print") -- omitted 1211 rows ]

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

```

#Is there any association between consumer price index and consumer?
with(bank_additional_full, chisq.test(cons.price.idx,cons.conf.idx))
with(bank_additional_full, table(cons.price.idx,cons.conf.idx))
# OR
with(bank_additional_full, prop.table(table(cons.price.idx,cons.conf.idx)))
#p-value < 2.2e-16 and it is very much less than 0.05.we can reject the null hypothesis in favor of
alternative hypothesis and conclude, that the variables, (job & Marital-p-value < 2.2e-
16),(con.price.idx , consumer- are dependent to each other.
Chi-squared approximation may be incorrect
Pearson's Chi-squared test
data: cons.price.idx and cons.conf.idx
X-squared = 1029700, df = 625, p-value < 2.2e-16
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 0 0 0 0 0 0 0 0 0 0 0
92.379 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
92.469 0 0 0 0 0 0 0 0 0 0 0 0
92.649 0 0 0 0 0 0 0 0 0 0 0 0
92.713 0 0 0 0 0 0 0 0 0 0 0 0
92.756 0 0 0 0 1 0 0 0 0 0 0 0
92.843 0 282 0 0 0 0 0 0 0 0 0 0
92.893 0 0 0 0 5794 0 0 0 0 0 0
92.963 0 0 0 0 0 0 0 0 0 715 0
93.075 0 0 0 2458 0 0 0 0 0 0 0
93.2 0 0 0 0 0 0 3616 0 0 0
93.369 0 0 0 0 0 0 0 0 0 0 0 0
93.444 0 0 0 0 0 0 0 0 0 0 0 0
93.749 0 0 0 0 0 0 0 0 0 0 0 0
93.798 0 0 0 0 0 0 0 0 0 0 0 67
93.876 0 0 0 0 0 0 0 0 0 0 0 0
93.918 0 0 0 0 0 6685 0 0 0 0 0
93.994 0 0 0 0 0 0 0 0 0 0 0 0
94.027 0 0 0 0 0 0 0 0 0 0 0 0
94.055 0 0 0 0 0 0 0 0 0 0 0 0
94.199 0 0 0 0 0 0 0 0 0 0 0 0
94.215 0 0 0 0 0 0 0 0 0 0 0 0
94.465 0 0 0 0 0 0 0 4374 0 0
94.601 0 0 204 0 0 0 0 0 0 0 0
94.767 128 0 0 0 0 0 0 0 0 0 0
cons.conf.idx
cons.price.idx -40.3 -40 -39.8 -38.3 -37.5 -36.4 -36.1 -34.8 -34.6 -33.6 -33
92.201 0 0 0 0 0 0 0 0 0 0 0 0
92.379 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

```

92.469 0 0 0 0 0 0 0 0 178 0  
92.649 0 0 0 0 0 0 0 0 0 0 0 0  
92.713 0 0 0 0 0 0 0 0 0 0 0 172  
92.756 0 0 0 0 0 0 0 0 0 0 0 0 0  
92.843 0 0 0 0 0 0 0 0 0 0 0 0 0  
92.893 0 0 0 0 0 0 0 0 0 0 0 0 0  
92.963 0 0 0 0 0 0 0 0 0 0 0 0 0  
93.075 0 0 0 0 0 0 0 0 0 0 0 0 0  
93.2 0 0 0 0 0 0 0 0 0 0 0 0 0  
93.369 0 0 0 0 0 0 264 0 0 0  
93.444 0 0 0 0 0 5175 0 0 0 0  
93.749 0 0 0 0 0 0 0 174 0 0  
93.798 0 0 0 0 0 0 0 0 0 0 0 0  
93.876 0 212 0 0 0 0 0 0 0 0 0  
93.918 0 0 0 0 0 0 0 0 0 0 0 0  
93.994 0 0 0 0 0 7763 0 0 0 0 0  
94.027 0 0 0 233 0 0 0 0 0 0 0  
94.055 0 0 229 0 0 0 0 0 0 0  
94.199 0 0 0 0 303 0 0 0 0 0  
94.215 311 0 0 0 0 0 0 0 0 0  
94.465 0 0 0 0 0 0 0 0 0 0 0  
94.601 0 0 0 0 0 0 0 0 0 0 0  
94.767 0 0 0 0 0 0 0 0 0 0 0  
cons.conf.idx  
cons.price.idx -31.4 -30.1 -29.8 -26.9  
92.201 770 0 0 0  
92.379 0 0 267 0  
92.431 0 0 0 447  
92.469 0 0 0 0  
92.649 0 357 0 0  
92.713 0 0 0 0  
92.756 0 0 0 0  
92.843 0 0 0 0  
92.893 0 0 0 0  
92.963 0 0 0 0  
93.075 0 0 0 0  
93.2 0 0 0 0  
93.369 0 0 0 0  
93.444 0 0 0 0  
93.749 0 0 0 0  
93.798 0 0 0 0  
93.876 0 0 0 0  
93.918 0 0 0 0  
93.994 0 0 0 0  
94.027 0 0 0 0  
94.055 0 0 0 0  
94.199 0 0 0 0

94.215 0 0 0  
94.465 0 0 0  
94.601 0 0 0  
94.767 0 0 0  
cons.conf.idx  
cons.price.idx -50.8 -50 -49.5 -47.1 -46.2  
92.201 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000  
92.379 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000  
92.431 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000  
92.469 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000  
92.649 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000  
92.713 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000  
92.756 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000  
92.843 0.0000000000 0.0068466544 0.0000000000 0.0000000000 0.0000000000  
92.893 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.1406720404  
92.963 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000  
93.075 0.0000000000 0.0000000000 0.0000000000 0.0596775760 0.0000000000  
93.2 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000  
93.369 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000  
93.444 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000  
93.749 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000  
93.798 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000  
93.876 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000  
93.918 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000  
93.994 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000  
94.027 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000  
94.055 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000  
94.199 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000  
94.215 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000  
94.465 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000  
94.601 0.0000000000 0.0000000000 0.0049528989 0.0000000000 0.0000000000  
94.767 0.0031077013 0.0000000000 0.0000000000 0.0000000000 0.0000000000  
cons.conf.idx  
cons.price.idx -45.9 -42.7 -42 -41.8 -40.8  
92.201 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000  
92.379 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000  
92.431 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000  
92.469 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000  
92.649 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000  
92.713 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000  
92.756 0.0002427892 0.0000000000 0.0000000000 0.0000000000 0.0000000000  
92.843 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000  
92.893 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000  
92.963 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0173594251  
93.075 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000  
93.2 0.0000000000 0.0000000000 0.0877925609 0.0000000000 0.0000000000  
93.369 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000





```

94.465 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
94.601 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
94.767 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
cons.conf.idx
cons.price.idx -26.9
92.201 0.0000000000
92.379 0.0000000000
92.431 0.0108526755
92.469 0.0000000000
92.649 0.0000000000
92.713 0.0000000000
92.756 0.0000000000
92.843 0.0000000000
92.893 0.0000000000
92.963 0.0000000000
93.075 0.0000000000
93.2 0.0000000000
93.369 0.0000000000
93.444 0.0000000000
93.749 0.0000000000
93.798 0.0000000000
93.876 0.0000000000
93.918 0.0000000000
93.994 0.0000000000
94.027 0.0000000000
94.055 0.0000000000
94.199 0.0000000000
94.215 0.0000000000
94.465 0.0000000000
94.601 0.0000000000
94.767 0.0000000000

```

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

```

#
with(bank_additional_full, chisq.test( job,emp.var.rate))
with(bank_additional_full, table( job,emp.var.rate) )
# OR
with(bank_additional_full, prop.table(table( job,emp.var.rate)))
#p-value < 2.2e-16 is very much less than 0.05
Pearson's Chi-squared test
data: job and emp.var.rate
X-squared = 4676.8, df = 99, p-value < 2.2e-16
emp.var.rate
job -3.4 -3 -2.9 -1.8 -1.7 -1.1 -0.2 -0.1 1.1 1.4
admin. 321 47 562 2231 246 187 3 940 1601 4284
blue-collar 64 9 99 2519 58 33 3 575 2295 3599

```

entrepreneur 24 1 38 306 14 7 0 265 289 512  
 housemaid 32 9 41 120 18 16 1 70 229 524  
 management 98 12 121 593 47 38 0 522 553 940  
 retired 193 33 181 338 96 83 0 72 215 509  
 self-employed 40 6 60 287 24 12 0 187 253 552  
 services 32 2 88 1040 47 40 0 311 932 1477  
 student 62 20 144 311 72 73 0 21 66 106  
 technician 145 22 234 1243 110 115 2 575 1060 3237  
 unemployed 44 9 76 164 31 28 1 141 171 349  
 unknown 16 2 19 32 10 3 0 4 99 145  
 emp.var.rate  
 job -3.4 -3 -2.9 -1.8 -1.7  
 admin. 7.793532e-03 1.141109e-03 1.364475e-02 5.416626e-02 5.972613e-03  
 blue-collar 1.553851e-03 2.185102e-04 2.403613e-03 6.115859e-02 1.408177e-03  
 entrepreneur 5.826940e-04 2.427892e-05 9.225988e-04 7.429348e-03 3.399048e-04  
 housemaid 7.769253e-04 2.185102e-04 9.954356e-04 2.913470e-03 4.370205e-04  
 management 2.379334e-03 2.913470e-04 2.937749e-03 1.439740e-02 1.141109e-03  
 retired 4.685831e-03 8.012042e-04 4.394484e-03 8.206274e-03 2.330776e-03  
 self-employed 9.711566e-04 1.456735e-04 1.456735e-03 6.968049e-03 5.826940e-04  
 services 7.769253e-04 4.855783e-05 2.136545e-03 2.525007e-02 1.141109e-03  
 student 1.505293e-03 4.855783e-04 3.496164e-03 7.550743e-03 1.748082e-03  
 technician 3.520443e-03 5.341362e-04 5.681266e-03 3.017869e-02 2.670681e-03  
 unemployed 1.068272e-03 2.185102e-04 1.845198e-03 3.981742e-03 7.526464e-04  
 unknown 3.884627e-04 4.855783e-05 4.612994e-04 7.769253e-04 2.427892e-04  
 emp.var.rate  
 job -1.1 -0.2 -0.1 1.1 1.4  
 admin. 4.540157e-03 7.283675e-05 2.282218e-02 3.887054e-02 1.040109e-01  
 blue-collar 8.012042e-04 7.283675e-05 1.396038e-02 5.572011e-02 8.737982e-02  
 entrepreneur 1.699524e-04 0.000000e+00 6.433913e-03 7.016607e-03 1.243081e-02  
 housemaid 3.884627e-04 2.427892e-05 1.699524e-03 5.559872e-03 1.272215e-02  
 management 9.225988e-04 0.000000e+00 1.267359e-02 1.342624e-02 2.282218e-02  
 retired 2.015150e-03 0.000000e+00 1.748082e-03 5.219967e-03 1.235797e-02  
 self-employed 2.913470e-04 0.000000e+00 4.540157e-03 6.142566e-03 1.340196e-02  
 services 9.711566e-04 0.000000e+00 7.550743e-03 2.262795e-02 3.585996e-02  
 student 1.772361e-03 0.000000e+00 5.098572e-04 1.602408e-03 2.573565e-03  
 technician 2.792075e-03 4.855783e-05 1.396038e-02 2.573565e-02 7.859085e-02  
 unemployed 6.798097e-04 2.427892e-05 3.423327e-03 4.151695e-03 8.473342e-03  
 unknown 7.283675e-05 0.000000e+00 9.711566e-05 2.403613e-03 3.520443e-03

#### e. Is the employment variation rate same across Education?

```

with(bank_additional_full, chisq.test( education,emp.var.rate))
with(bank_additional_full, table( education, emp.var.rate) )
# OR
with(bank_additional_full, prop.table(table( education,emp.var.rate)))
  
```

Pearson's Chi-squared test

data: education and emp.var.rate

X-squared = 1451.6, df = 63, p-value < 2.2e-16

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	141	17	106	843	75	59	3	238	993	1701	
basic.6y	36	0	35	584	18	9	0	154	592	864	
basic.9y	69	16	110	1628	53	27	0	504	1428	2210	
high.school	216	36	358	2366	183	143	4	809	1857	3543	
illiterate	0	0	3	3	0	0	0	3	2	7	
professional.course	131	19	196	1041	93	113	3	470	887	2290	
university.degree	411	70	758	2403	301	242	0	1414	1627	4942	
unknown	67	14	97	316	50	42	0	91	377	677	

emp.var.rate

	education	-3.4	-3	-2.9	-1.8
basic.4y	3.423327e-03	4.127416e-04	2.573565e-03	2.046713e-02	
basic.6y	8.740410e-04	0.000000e+00	8.497621e-04	1.417889e-02	
basic.9y	1.675245e-03	3.884627e-04	2.670681e-03	3.952608e-02	
high.school	5.244246e-03	8.740410e-04	8.691852e-03	5.744392e-02	
illiterate	0.000000e+00	0.000000e+00	7.283675e-05	7.283675e-05	
professional.course	3.180538e-03	4.612994e-04	4.758668e-03	2.527435e-02	
university.degree	9.978635e-03	1.699524e-03	1.840342e-02	5.834224e-02	
unknown	1.626687e-03	3.399048e-04	2.355055e-03	7.672138e-03	

emp.var.rate

	education	-1.7	-1.1	-0.2	-0.1
basic.4y	1.820919e-03	1.432456e-03	7.283675e-05	5.778382e-03	
basic.6y	4.370205e-04	2.185102e-04	0.000000e+00	3.738953e-03	
basic.9y	1.286783e-03	6.555307e-04	0.000000e+00	1.223657e-02	
high.school	4.443042e-03	3.471885e-03	9.711566e-05	1.964164e-02	
illiterate	0.000000e+00	0.000000e+00	0.000000e+00	7.283675e-05	
professional.course	2.257939e-03	2.743518e-03	7.283675e-05	1.141109e-02	
university.degree	7.307954e-03	5.875498e-03	0.000000e+00	3.433039e-02	
unknown	1.213946e-03	1.019714e-03	0.000000e+00	2.209381e-03	

emp.var.rate

	education	1.1	1.4
basic.4y	2.410896e-02	4.129844e-02	
basic.6y	1.437312e-02	2.097698e-02	
basic.9y	3.467029e-02	5.365640e-02	
high.school	4.508595e-02	8.602020e-02	
illiterate	4.855783e-05	1.699524e-04	
professional.course	2.153540e-02	5.559872e-02	
university.degree	3.950180e-02	1.199864e-01	
unknown	9.153151e-03	1.643683e-02	

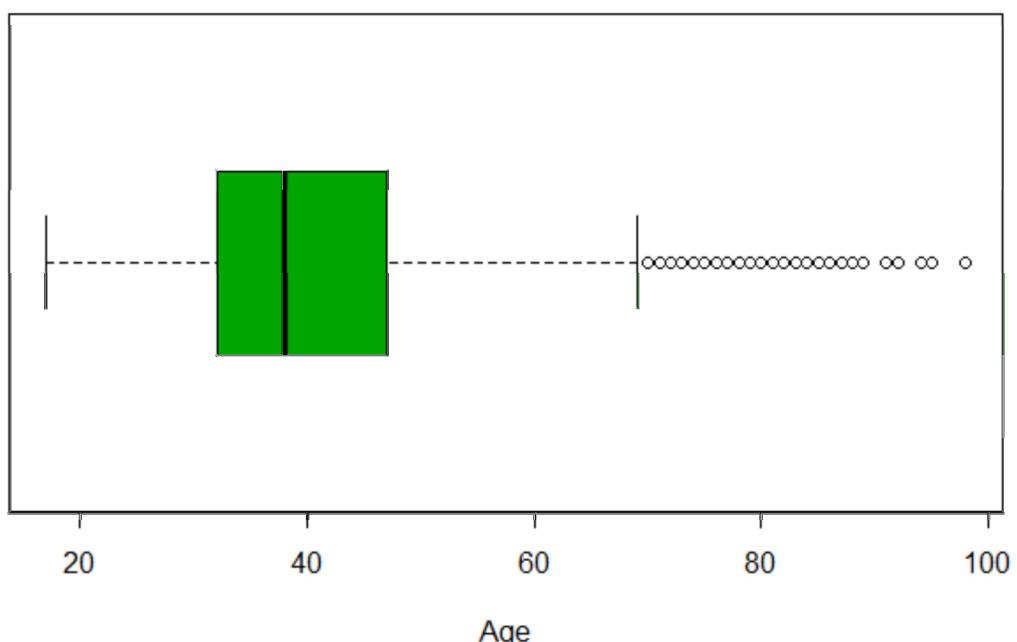
```
> library(readxl)
> bankdata <- read_excel("C:/Users/Veena/Desktop/bank-additional/bankdata.xlsx")
```

```

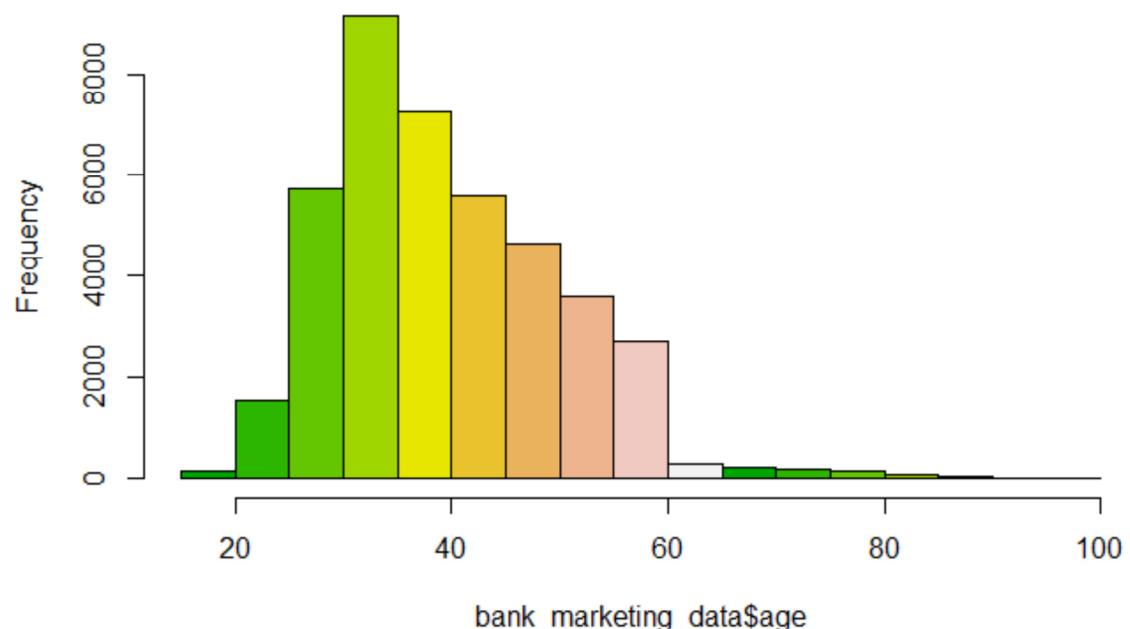
";", escape_double = FALSE, trim_ws = TRUE)
head(bank_marketing_data)
# We look at difference between mean and median in summary if it's more there might be outliers
boxplot(bank_marketing_data$age, main="Age Box plot",yaxt="n", xlab="Age", horizontal=TRUE,
col=terrain.colors(2))
# By plotting histogram we can ensure if there are outliers or not
## DATA VISUALISATION
## Use Box plots (Only for continuous variables)- To Check Ouliers
boxplot(bank_marketing_data$age~bank_marketing_data$contact, main=" AGE",ylab="age of
customers",xlab="contact")
boxplot(bank_marketing_data$age~bank_marketing_data$job, main=" AGE",ylab="age of
customers",xlab="job")
boxplot(bank_marketing_data$age~bank_marketing_data$education, main=" AGE",ylab="age of
customers",xlab="education")
boxplot(bank_marketing_data$age~bank_marketing_data$marital, main=" AGE",ylab="age of
customers",xlab="marital")
## Barplots for Categorical Variables
barplot(table(bank_marketing_data$job),col="red",main="JOB")
barplot(table(bank_marketing_data$marital),col="green",main="Marital")
barplot(table(bank_marketing_data$education),col="red",main="Education")
barplot(table(bank_marketing_data$emp.var.rate ),col="red",main="emp.var.rate")
hist(bank_marketing_data$age,col=terrain.colors(10))
#Correlation Analysis What we saw in the box plot can be emphasized by correlation plot, It can tell
if predictor is a good predictor or not a good predictor. This analysis can help us decide if we can
drop some columns/predictors depending upon its correlation with the outcome variable.
library(psych)
pairs.panels(bank_marketing_data[, c(1:8,17)])
pairs.panels(bank_marketing_data[, c(9:17)])
pairs.panels(bank_marketing_data[, c(1:8,19)])

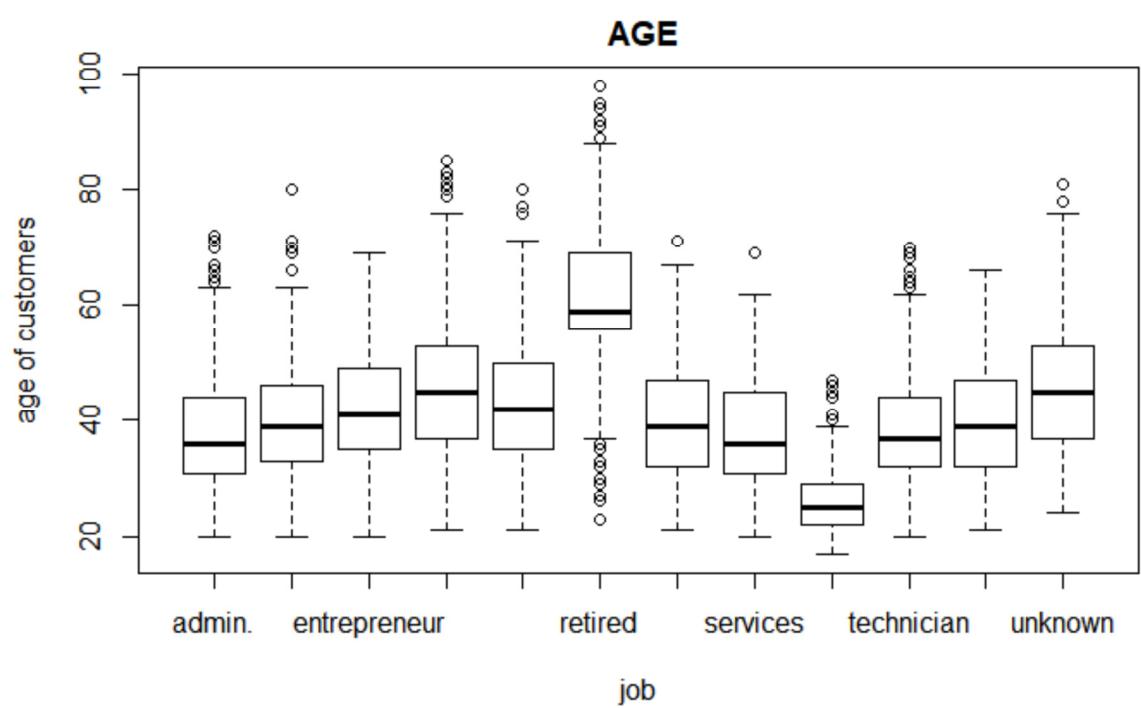
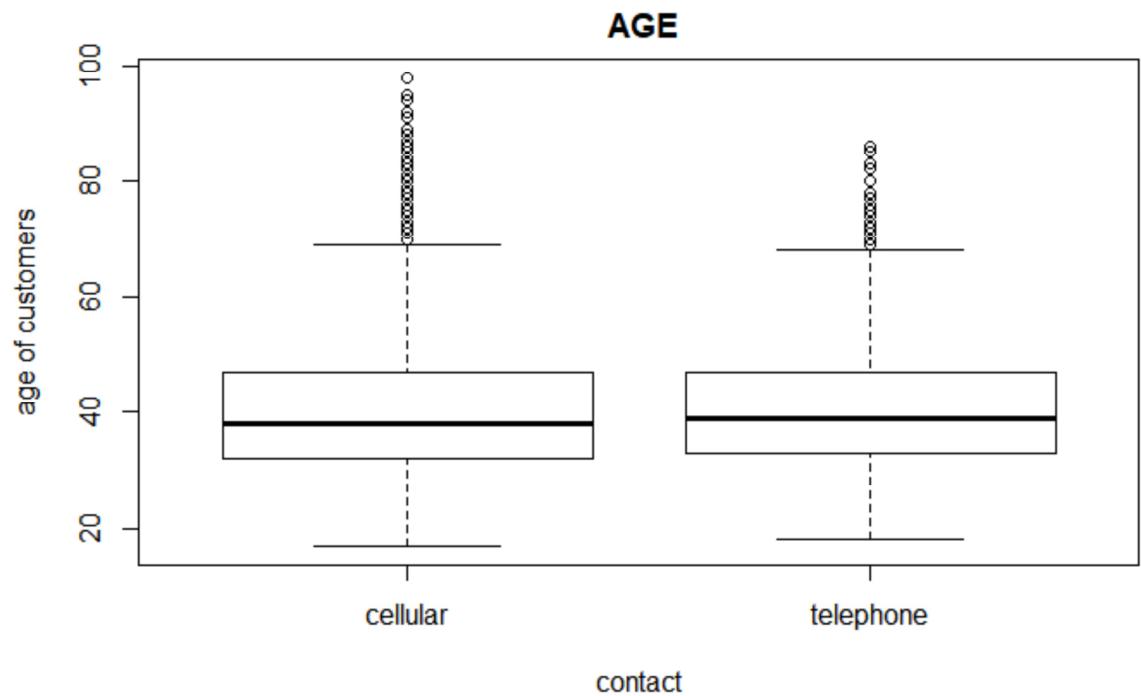
```

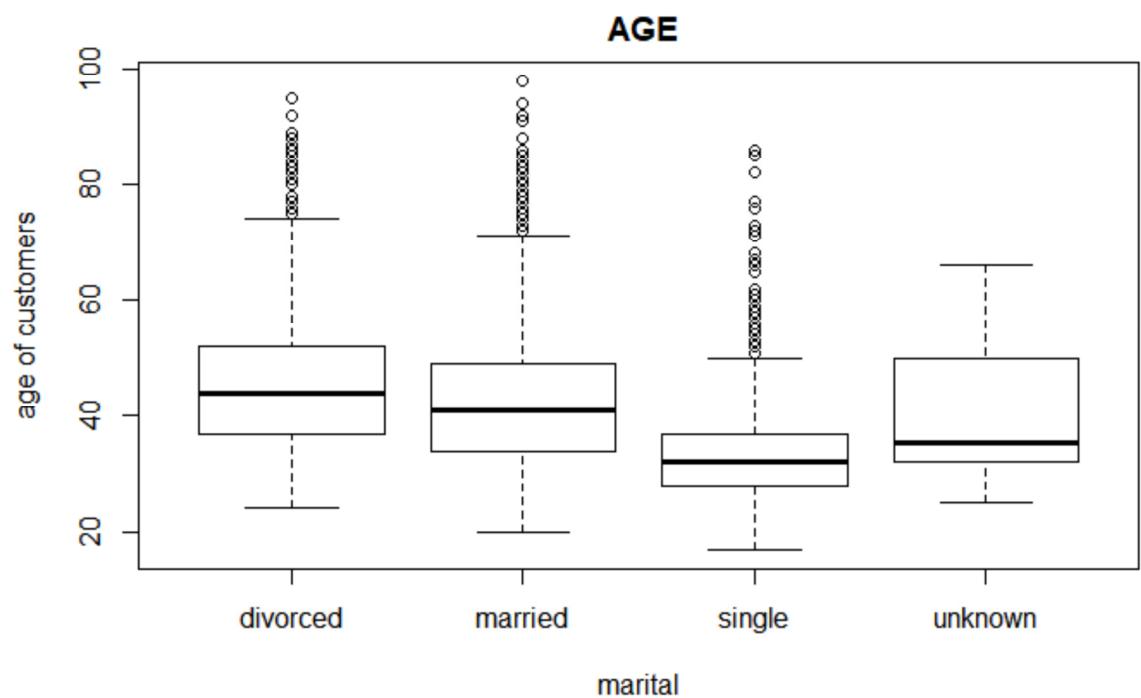
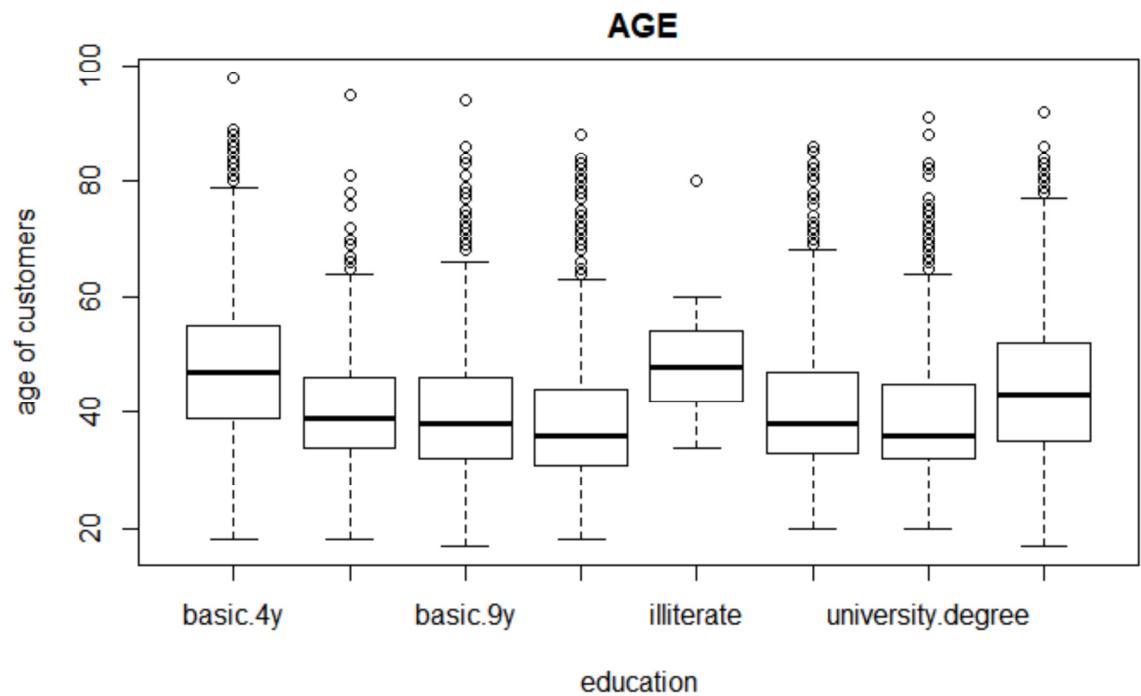
**Age Box plot**

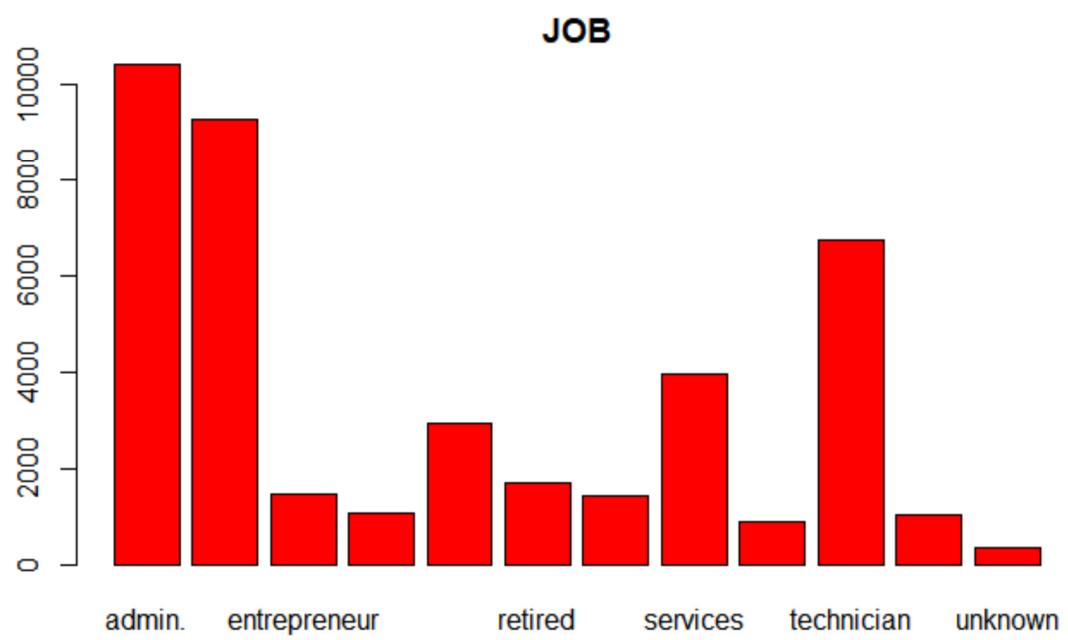
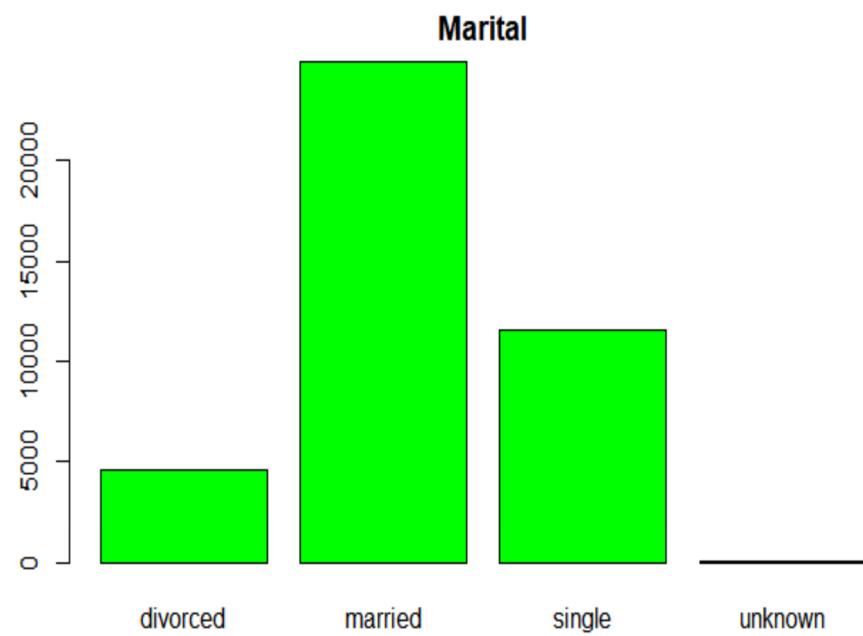


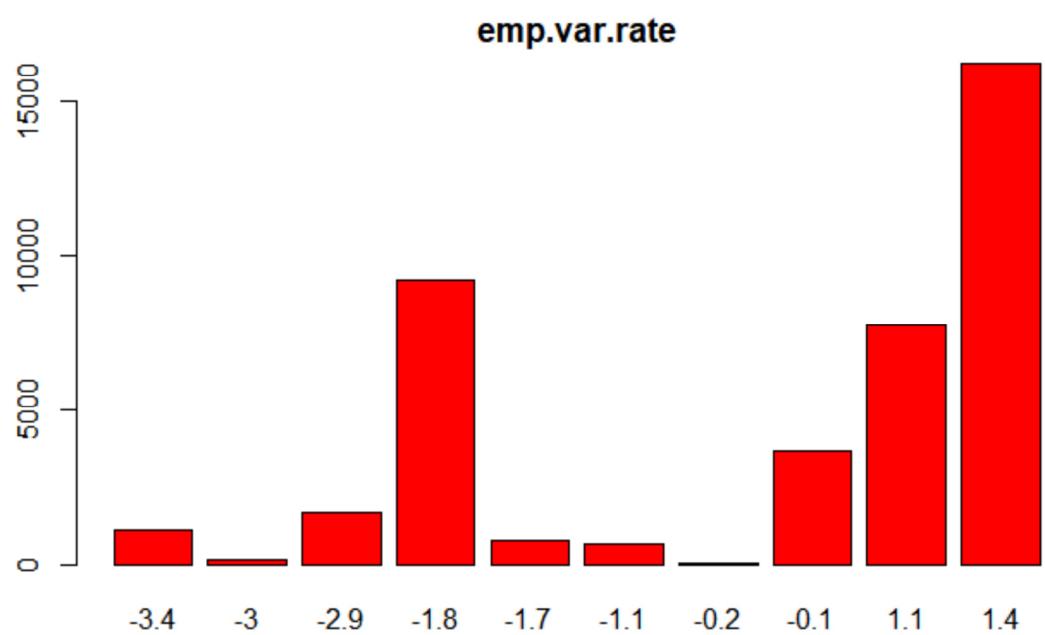
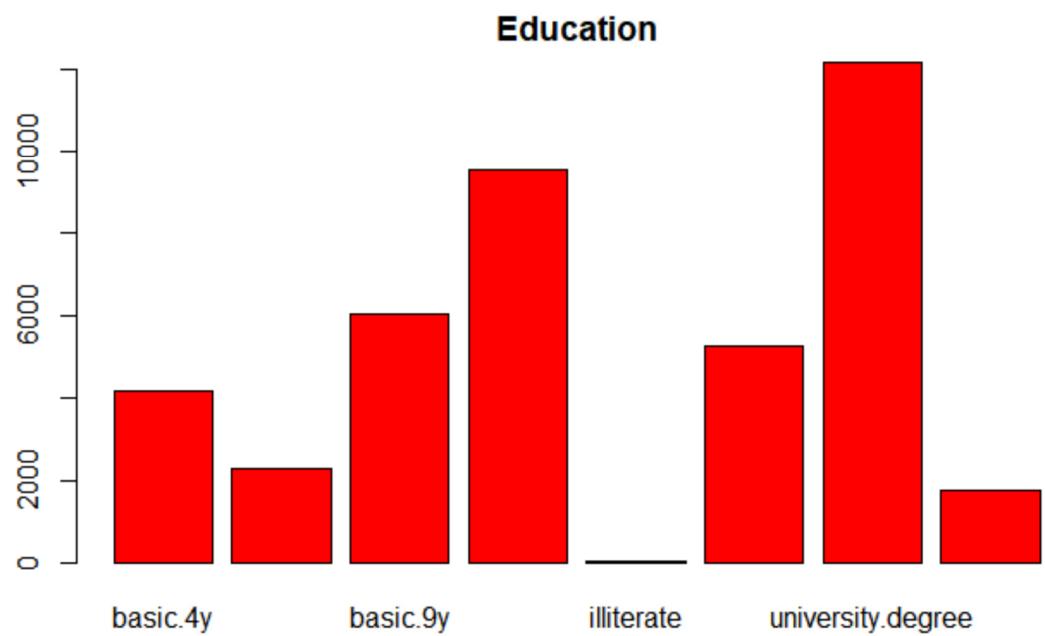
**Histogram of bank\_marketing\_data\$age**

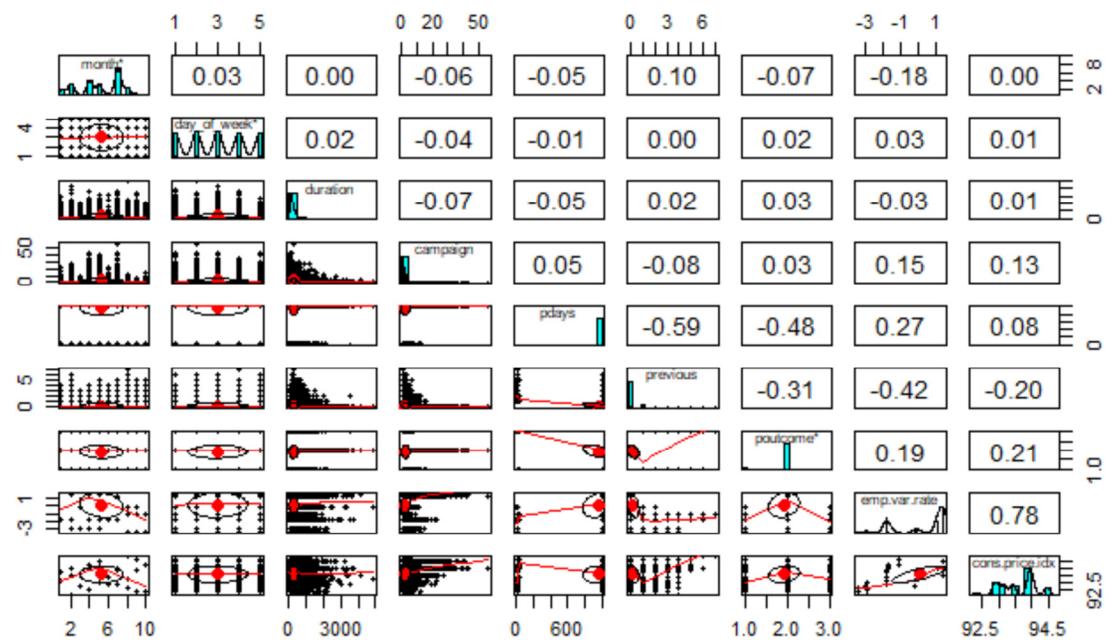
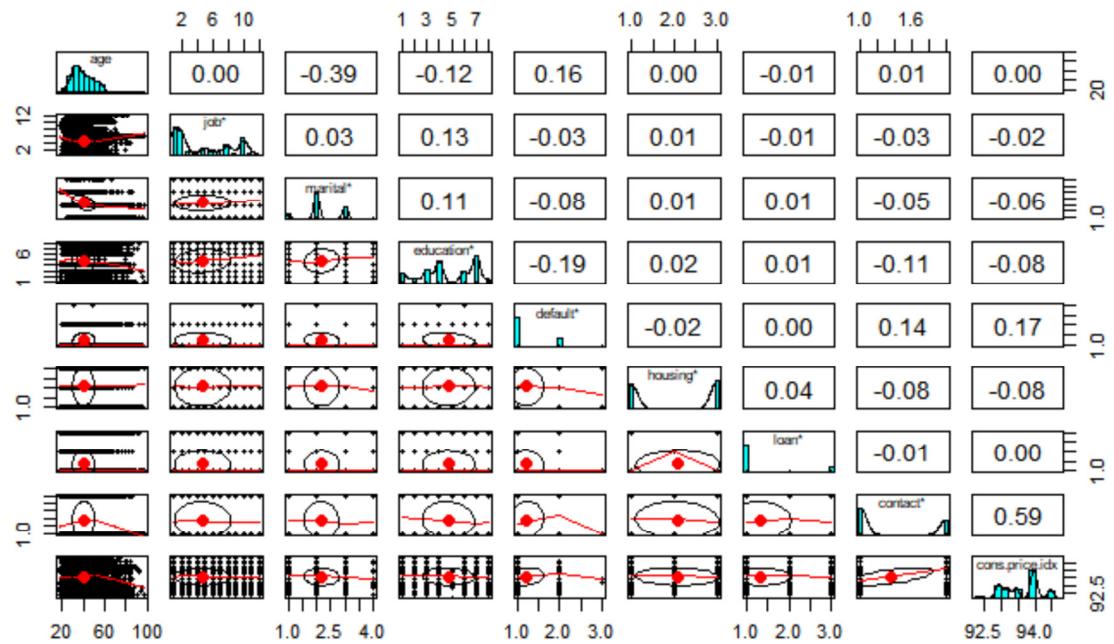


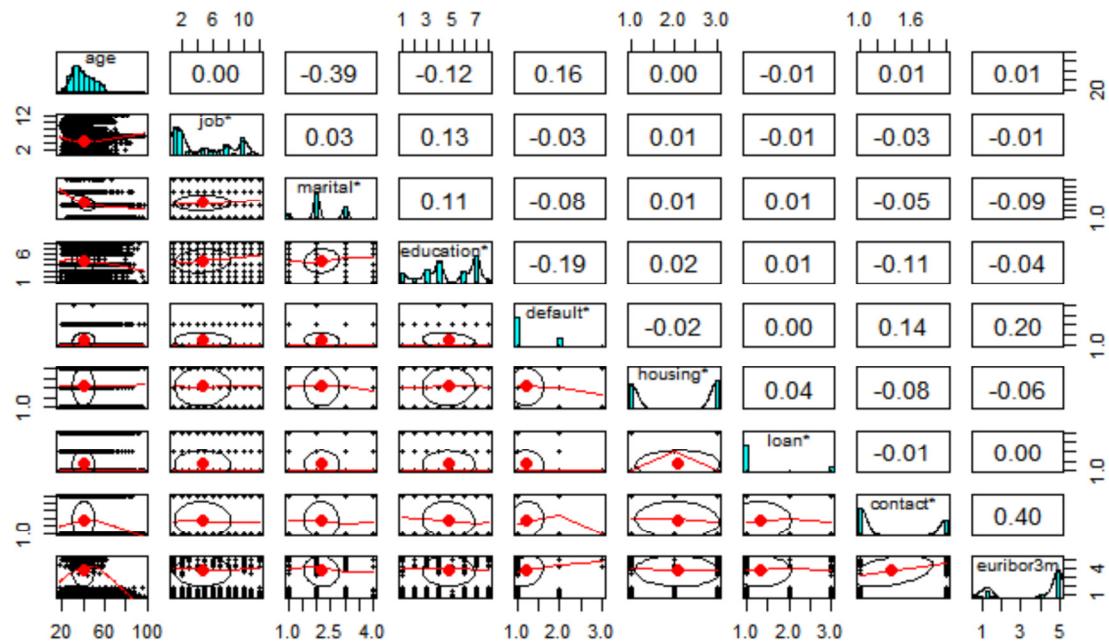












Subset Selection/ Feature-space reduction: Features-space can be reduced by selecting subsets based upon correlation values obtained

```
#####Subset Selection#####
lib
```

```
bank_marketing_data_sub<-bank_marketing_data[, c(1:4,7:9,12,14,15,17)]
```

```
str(bank_marketing_data_sub)
```

```
pairs.panels(bank_marketing_data_sub)
```

#3.4. Data transformation and Binning We do data transformation and binning for better modeling. We convert categorical variable into numerical using binning.

```
#####Binning and Data Transformation#####

```

```
#bank_marketing_data_sub$age <- cut(bank_marketing_data_sub$age, c(1,20,40,60,100))
```

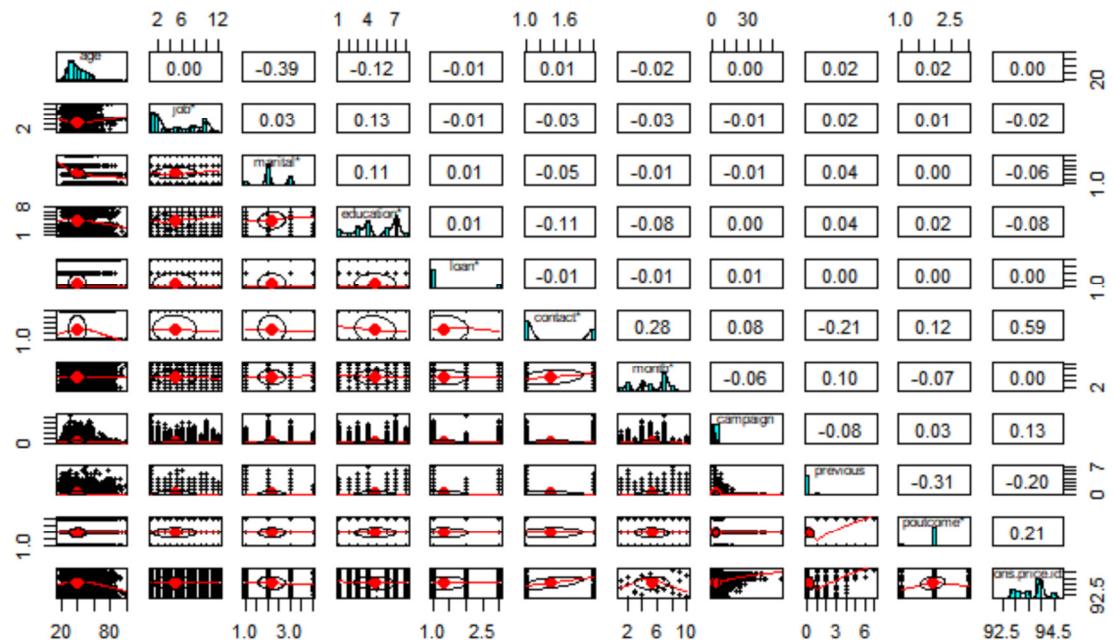
```
#bank_marketing_data_sub$is_divorced <- ifelse( bank_marketing_data_sub$marital == "divorced", 1, 0)
```

```
bank_marketing_data_sub$is_nr.employed <- ifelse( bank_marketing_data_sub$education == "employed", 1, 0)
```

```
#bank_marketing_data_sub$is_single <- ifelse( bank_marketing_data_sub$marital == "single", 1, 0)
```

```
bank_marketing_data_sub$nr.employed <- NULL
```

```
str(bank_marketing_data_sub)
```



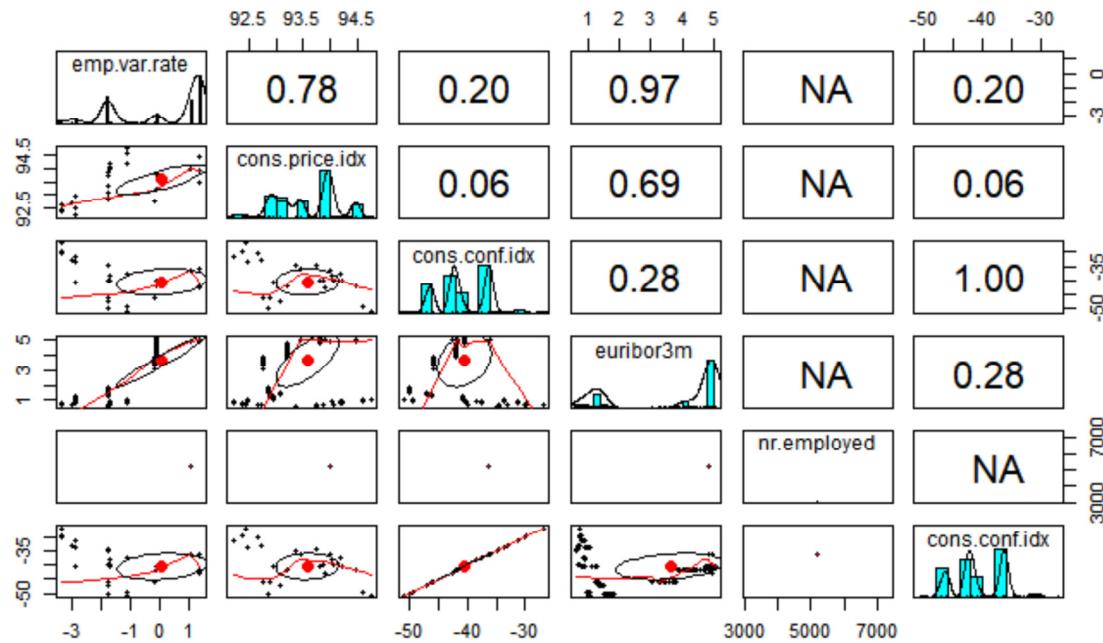
```
#scatter.smooth(x=bank_marketing_data$job, y=bank_marketing_data$emp.var.rate,
main="emp.var.rate ~ job") # scatterplot
# load library
library(corrplot)
# load the data
data<-bank_marketing_data
data(bank_marketing_data_sub)
# calculate correlations
correlations <- cor(bank_marketing_data[,16:19])
#create correlation plot
corrplot(correlations, method="circle")
```

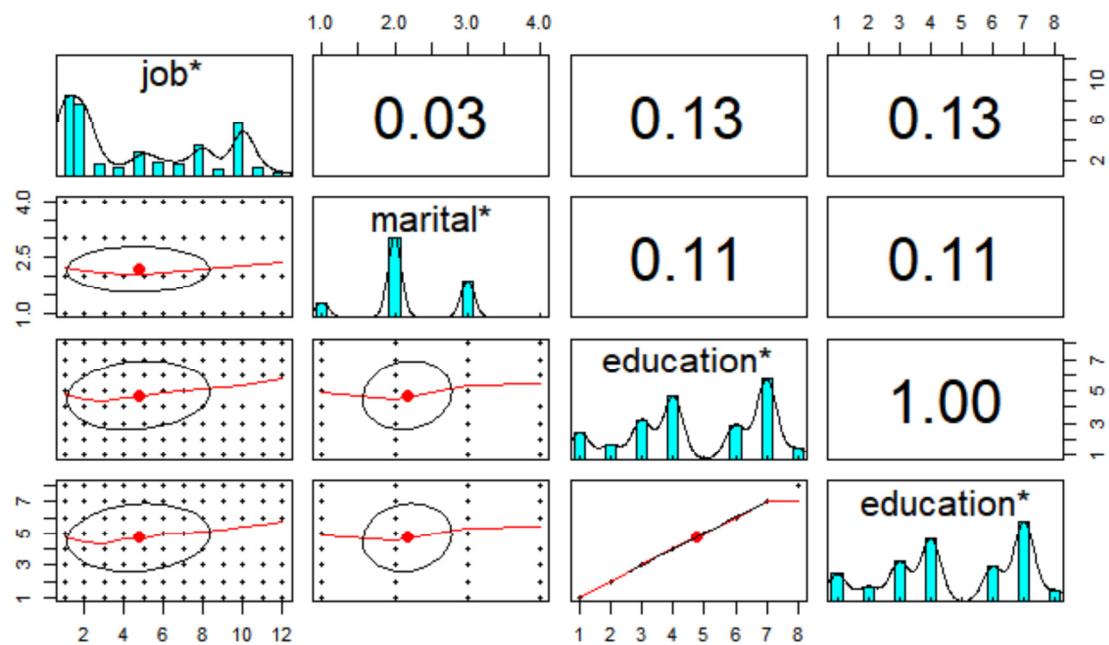
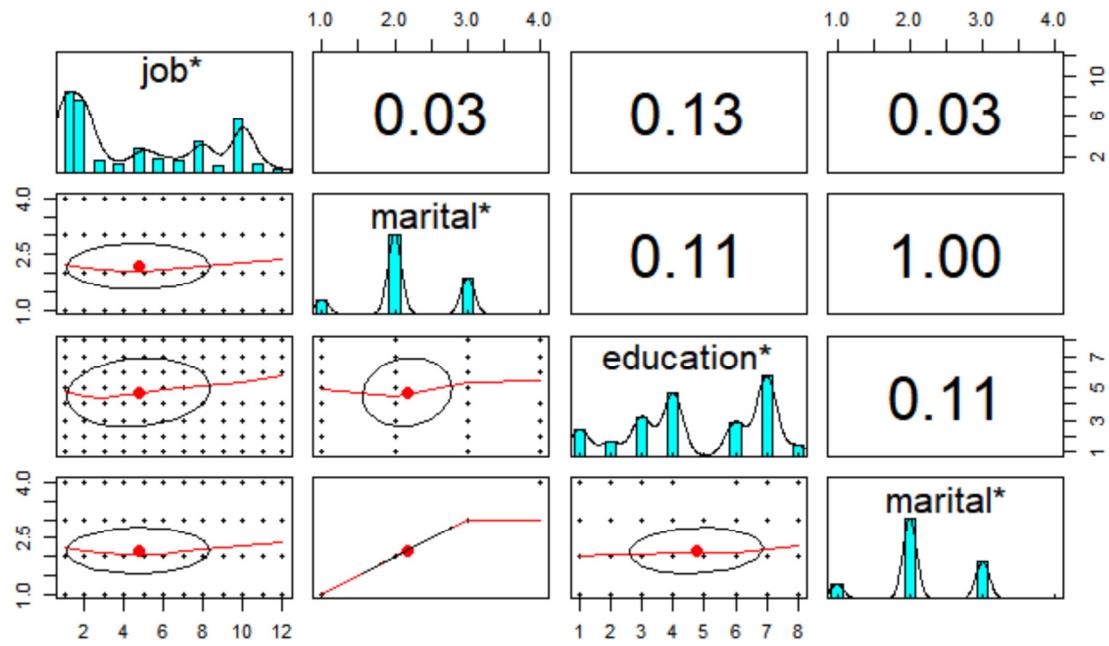


```
airs.panels(bank_marketing_data[, c(16:20,18)])
```

```
pairs.panels(bank_marketing_data[, c(2:4,3)])
```

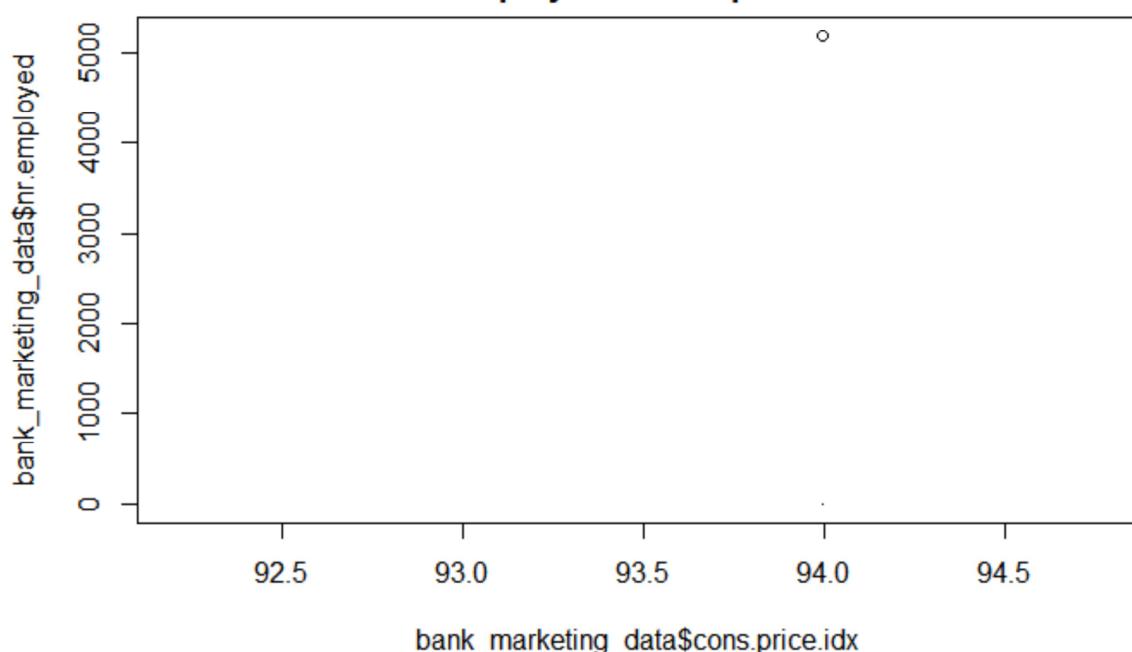
```
pairs.panels(bank_marketing_data[, c(2:4,4)])
```



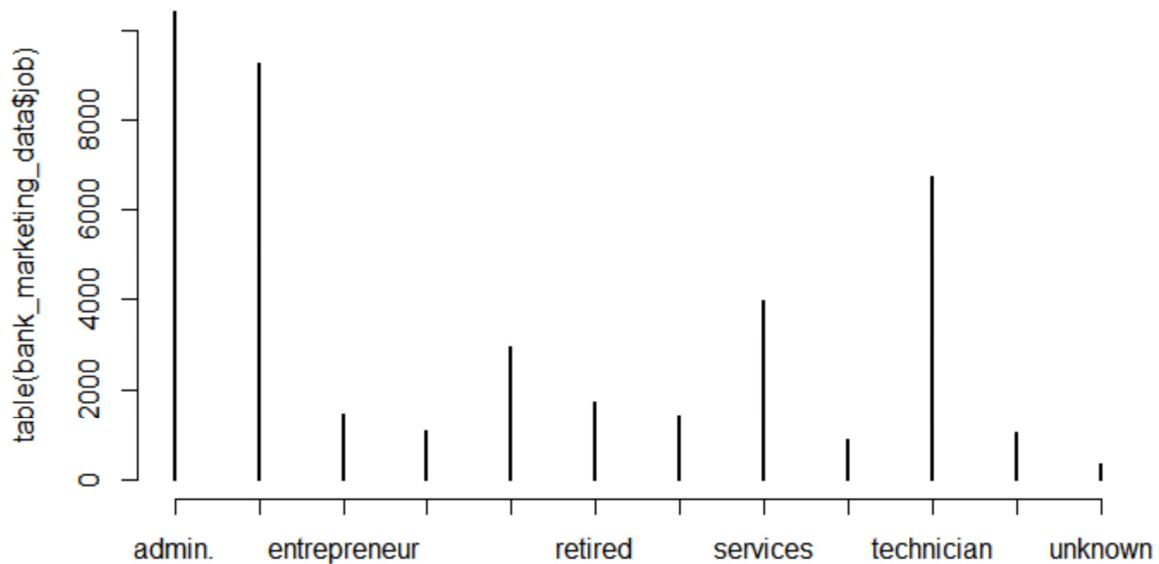


```
head(bank_marketing_data)
scatter.smooth(x=bank_marketing_data$cons.price.idx, y=bank_marketing_data$nr.employed,
main="nr.employed ~ cons.price.idx")
```

**nr.employed ~ cons.price.idx**



```
#cor(bank_marketing_data$age, bank_marketing_data$emp.var.rate)
head(bank_marketing_data)
table(bank_marketing_data$job)
table(bank_marketing_data$marital)
plot(table(bank_marketing_data$job))
```



```
library(psych)
pairs.panels(bank_marketing_data[,1:6])
```

