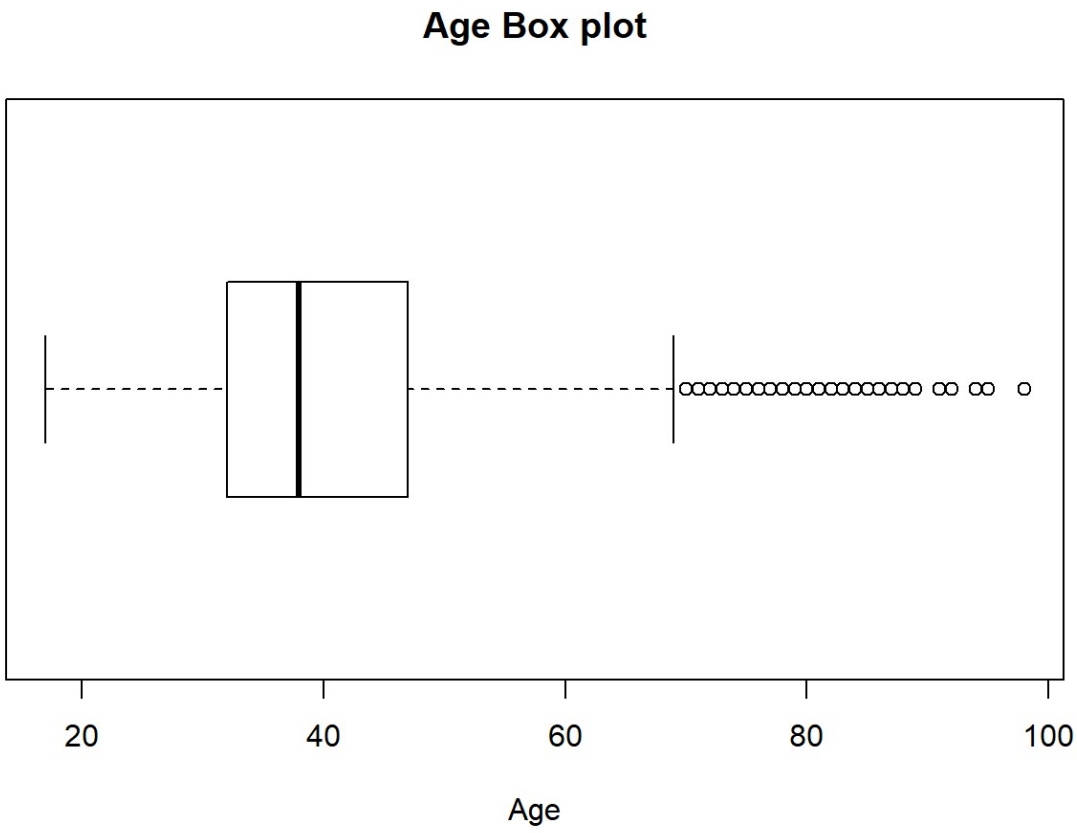
Bank_Data-_EDA.R

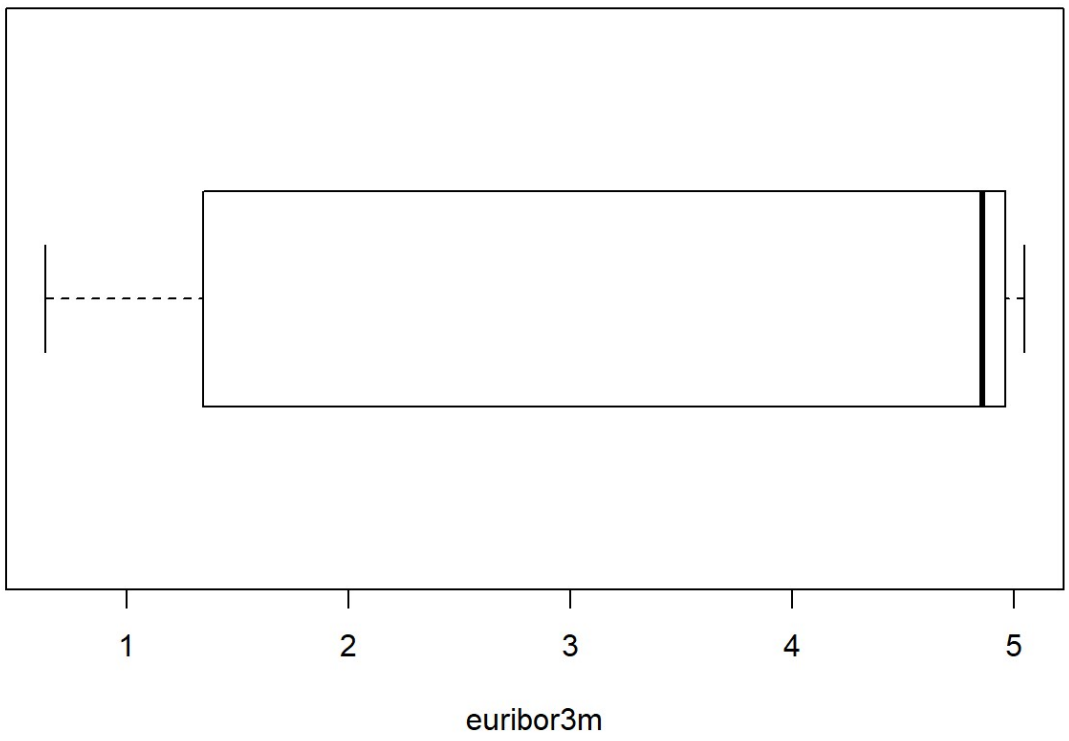
deept

Fri Feb 15 13:25:19 2019

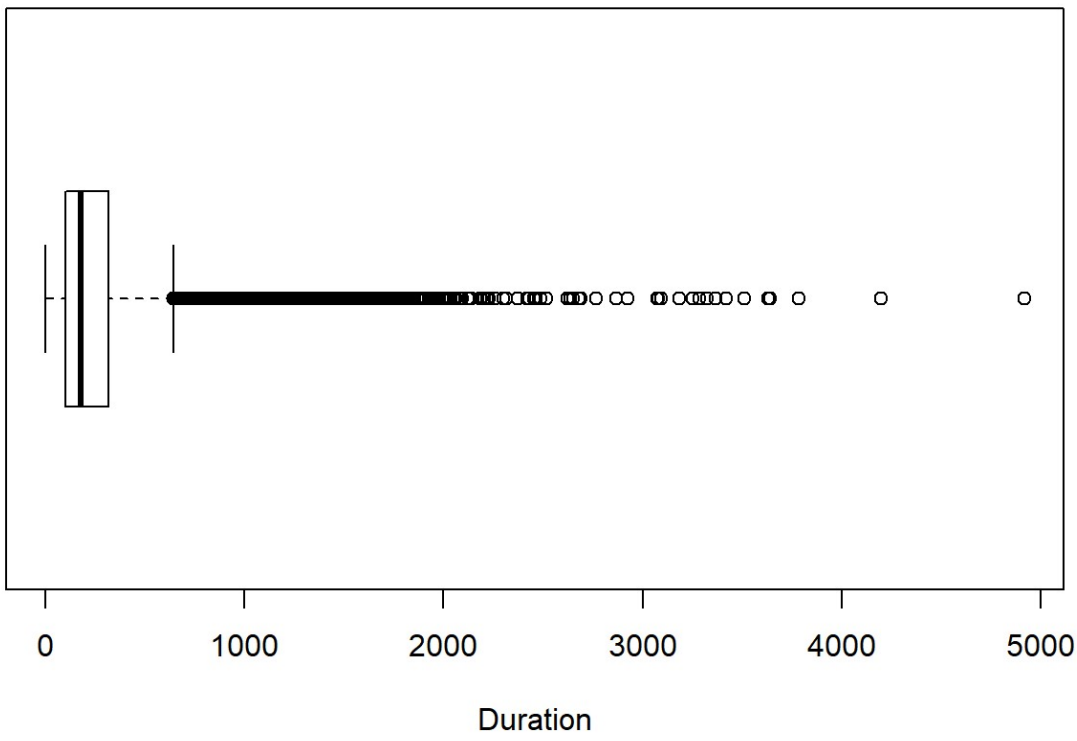
```
bank <- read.csv("~/Spring 19 Sem/Multi Analysis/bank-additional/bank-additional-full.
csv", sep=";")


boxplot(bank$age, main="Age Box plot",yaxt="n", xlab="Age", horizontal=TRUE)
```




```
boxplot(bank$euribor3m, main="Euribor3m Box plot",yaxt="n", xlab="euribor3m", horizont
al=TRUE)
```




```
boxplot(bank$duration, main="Duration Box plot",yaxt="n", xlab="Duration", horizontal=
TRUE)
```




```
library(MVA)
```

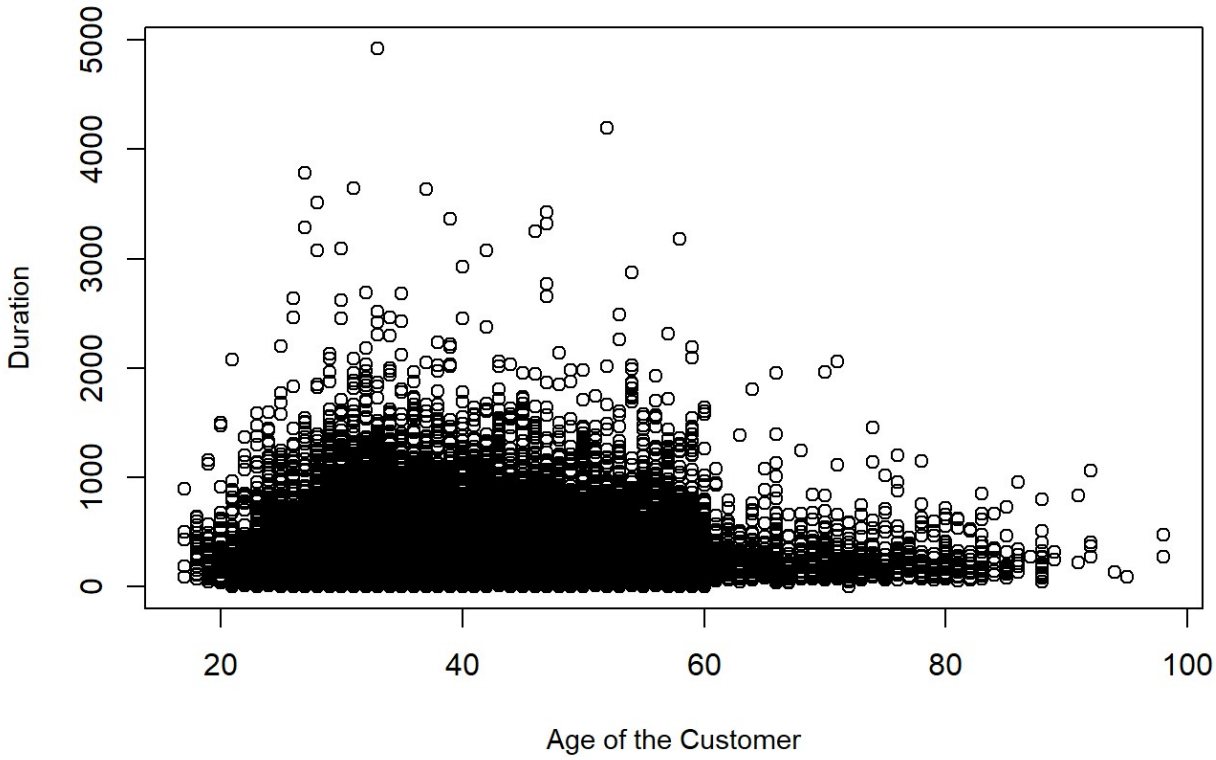
```
## Warning: package 'MVA' was built under R version 3.5.2
```

```
## Loading required package: HSAUR2
```

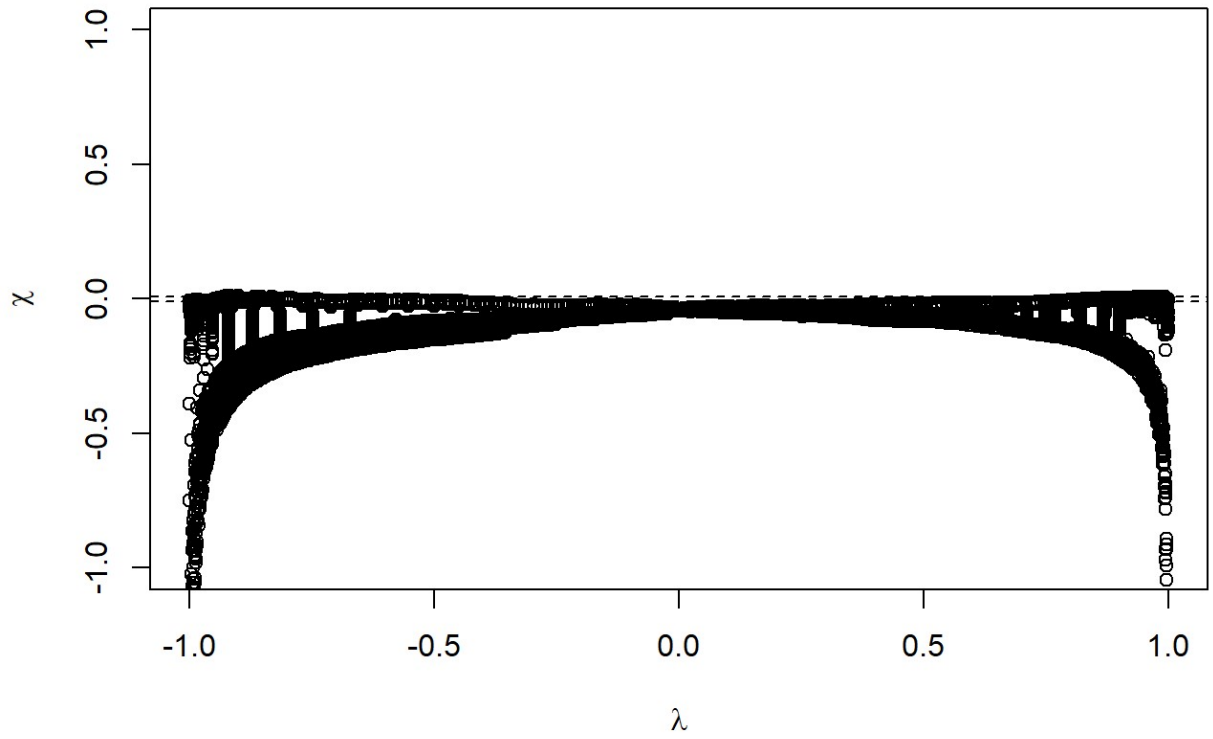
```
## Warning: package 'HSAUR2' was built under R version 3.5.2
```

```
## Loading required package: tools
```

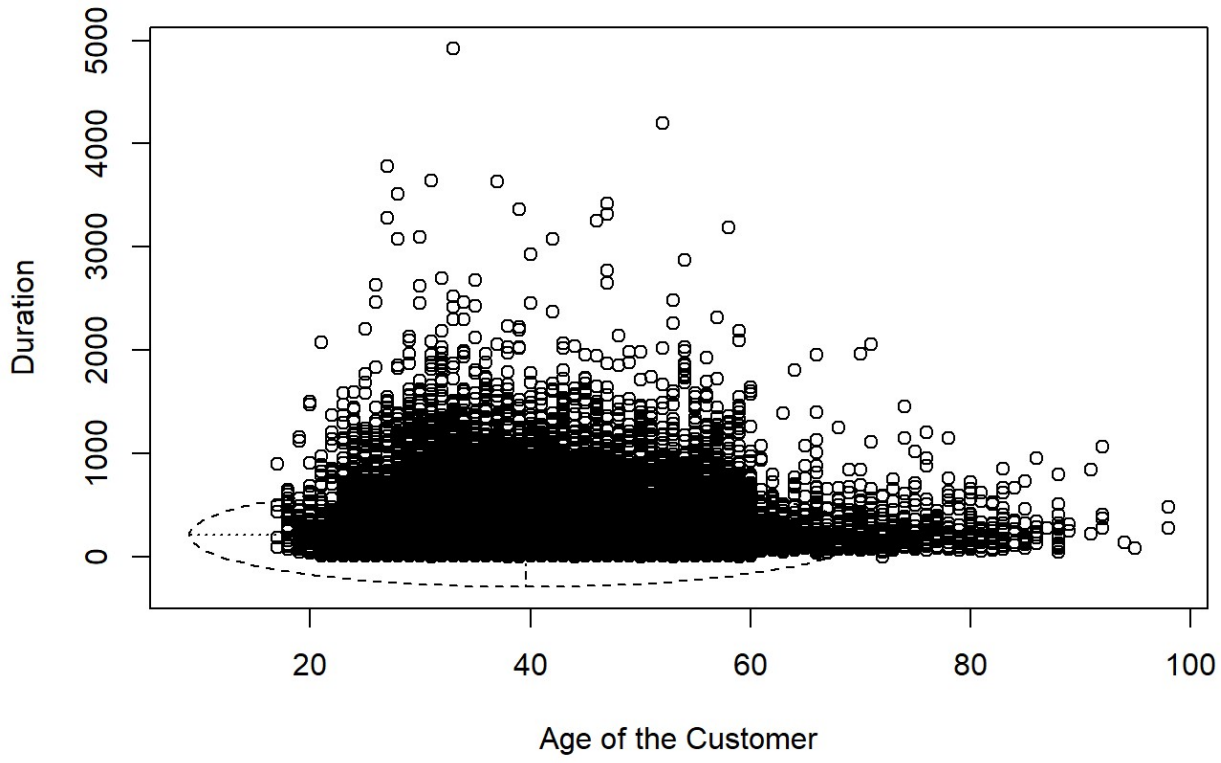
```
#Chiplot
mlab = "Age of the Customer"
plab = "Duration"
with(bank, plot(age, duration, xlab = mlab, ylab = plab, cex.lab = 0.9))
```



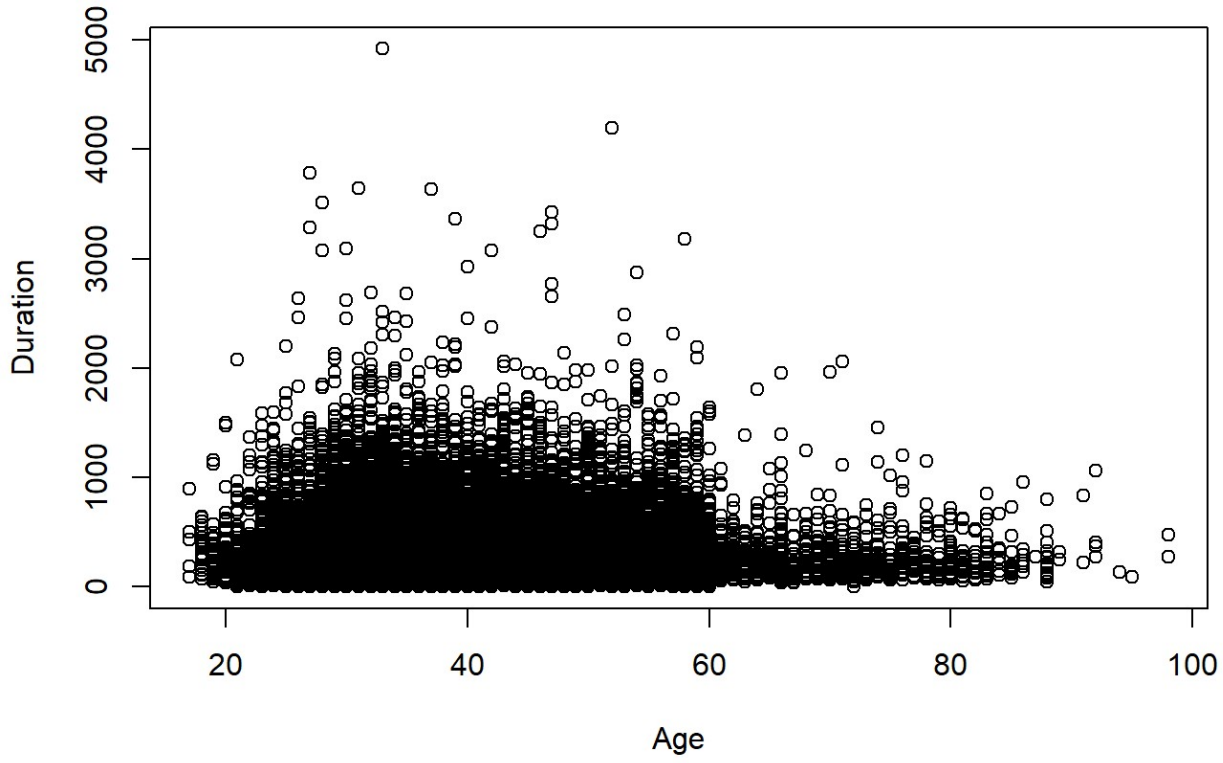

```
with(bank, chiplot(age, duration))
```



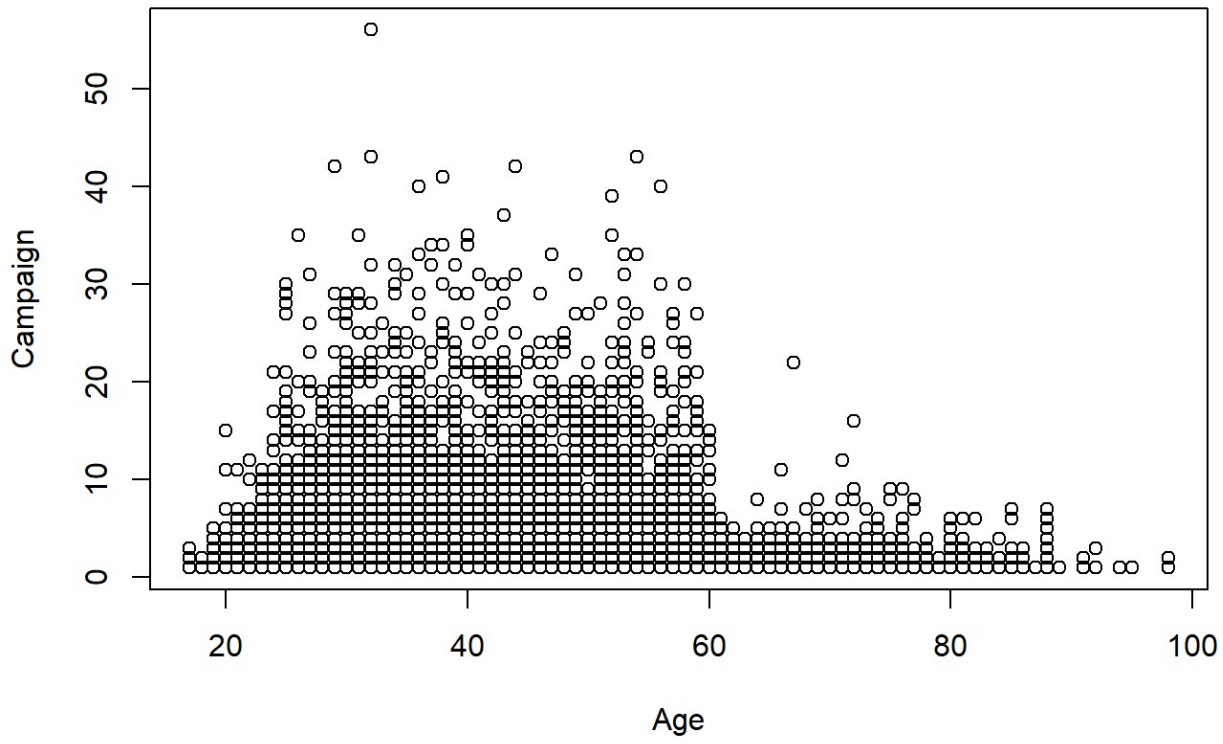

```
#bvplot
bank_age_dur=data.frame(bank$age, bank$duration)
bvbox(bank_age_dur, mtitle = "", xlab = mlab, ylab = plab)
```




```
y_int=ifelse(bank$y=='no', 0, 1)

plot(bank$age, bank$duration, pch=c(1,16)[y_int],xlab="Age",ylab="Duration")
```

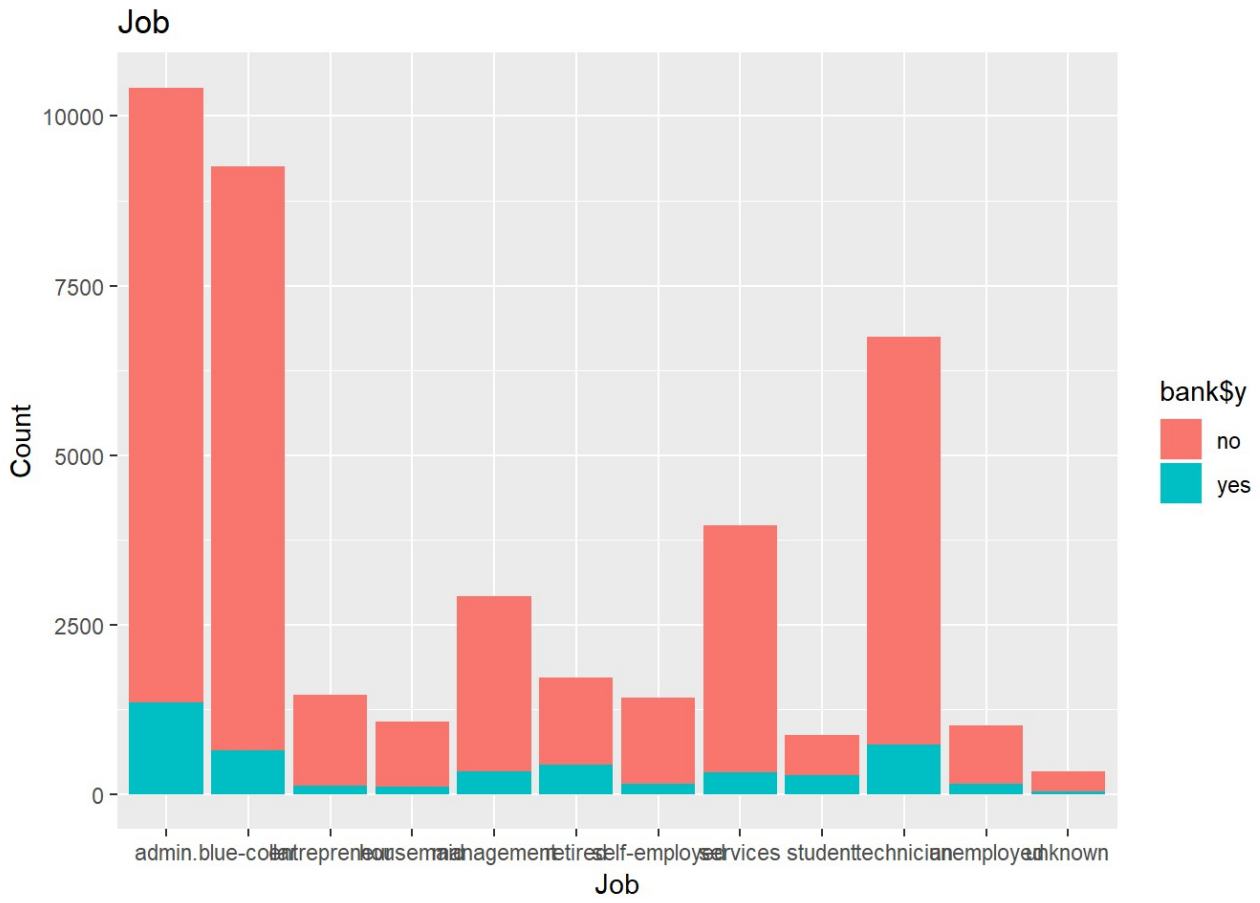



```
plot(bank$age, bank$campaign, pch=c(1,16)[y_int],xlab="Age", ylab="Campaign")

library(ggplot2)
```

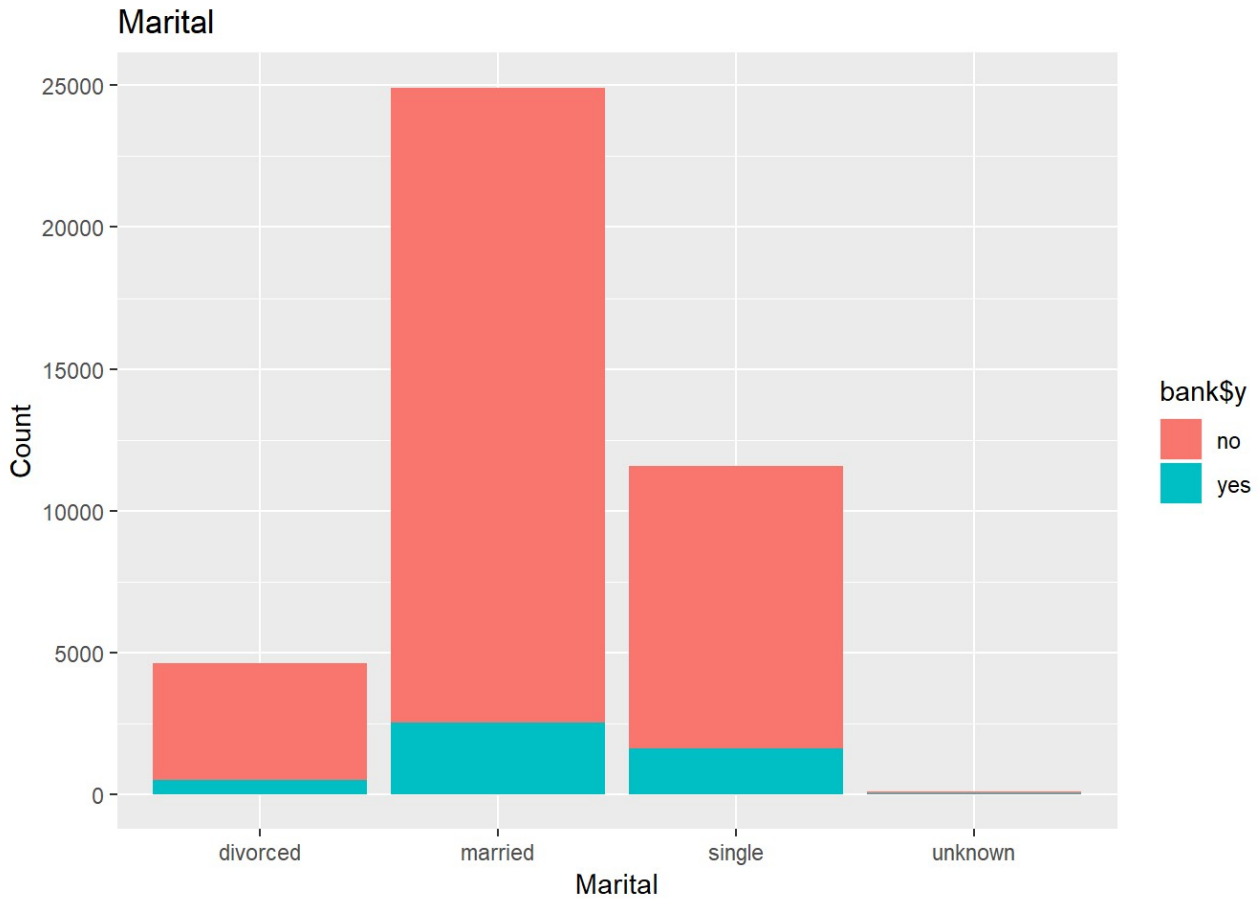



```
ggplot(bank,aes(x=bank$age,fill=bank$y)) + geom_histogram(binwidth=1) +
  labs(y= "Count", x="Age", title = "Age")
```

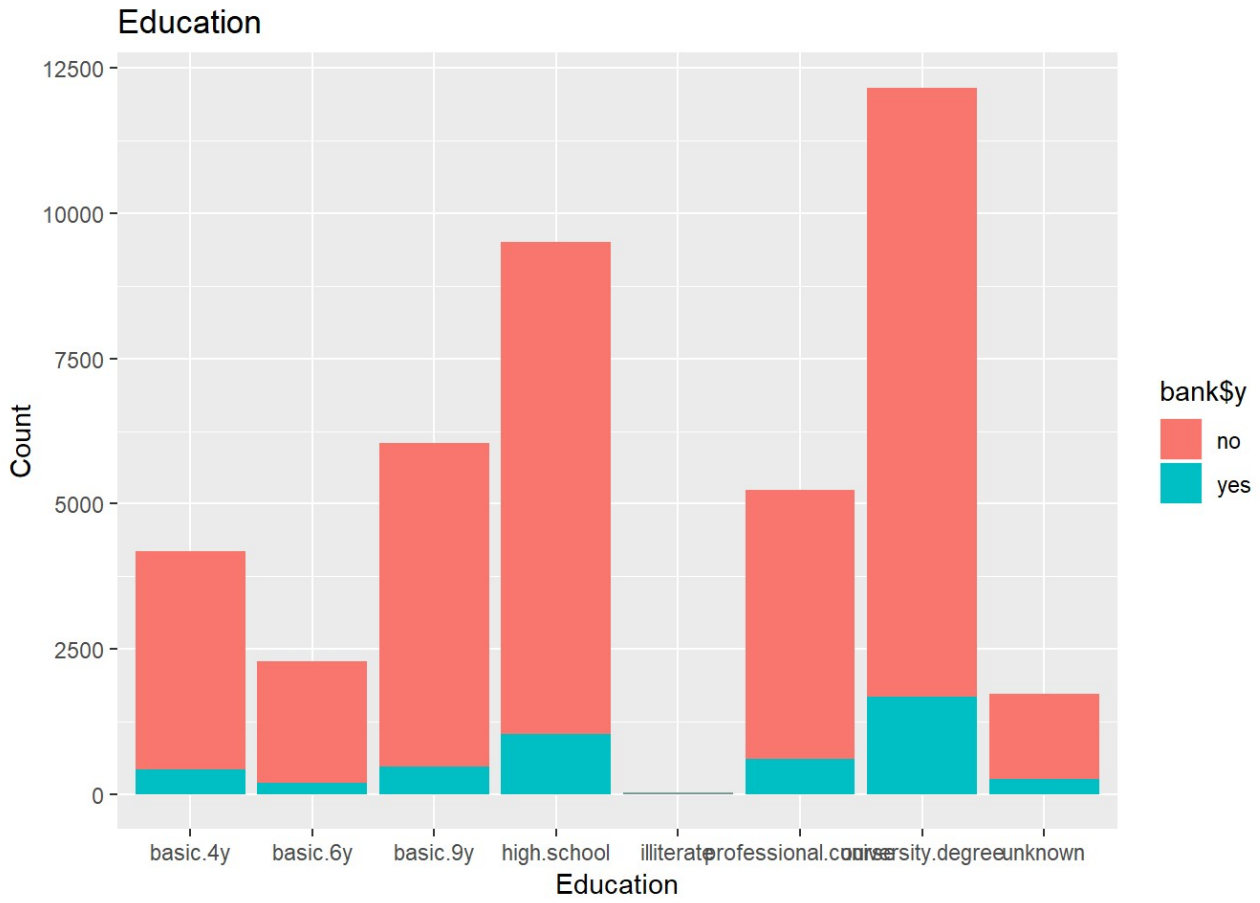

```
ggplot(bank, aes(x=bank$job,fill=bank$y)) + geom_bar()+
  labs(y= "Count", x="Job", title = "Job")
```



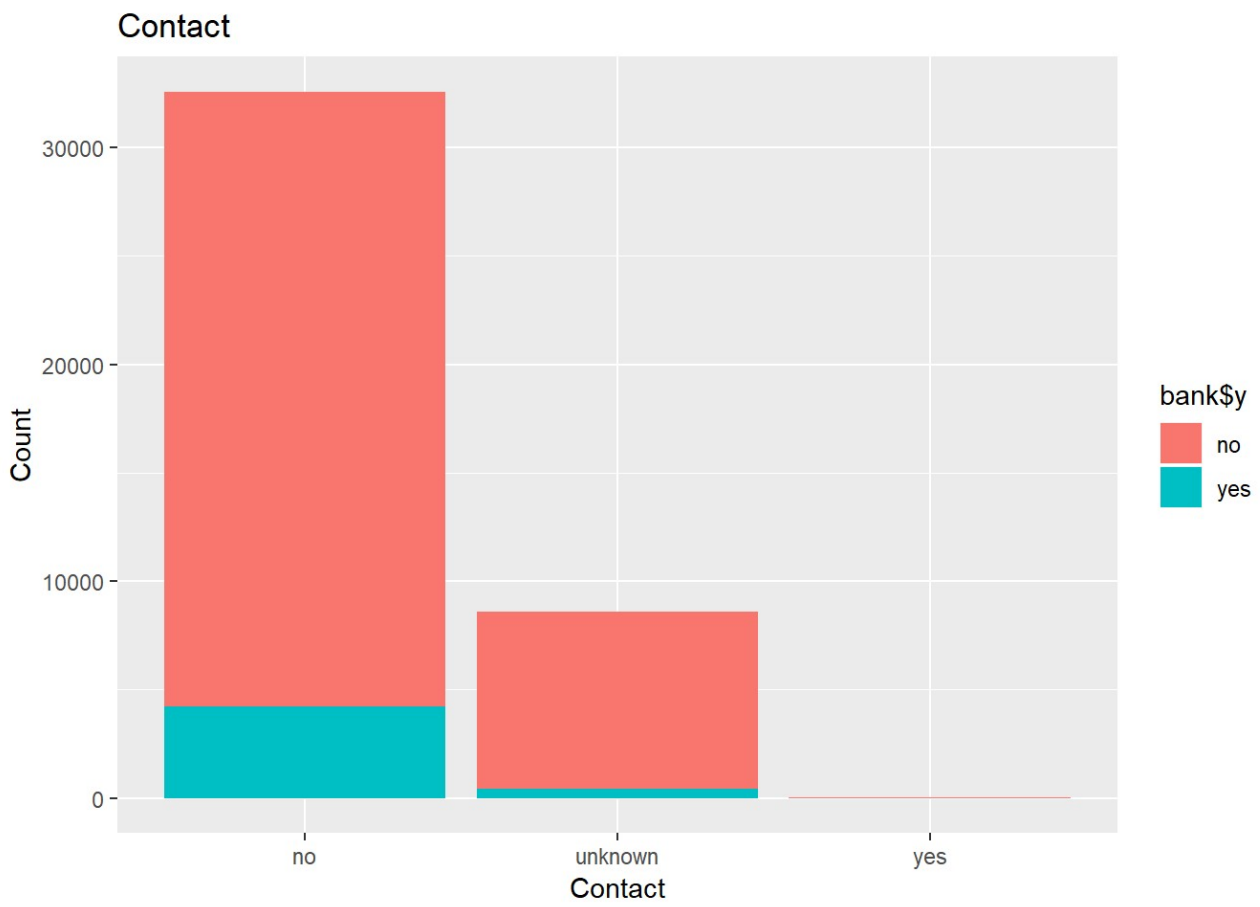

```
ggplot(bank, aes(x=bank$marital,fill=bank$y)) + geom_bar() +
  labs(y= "Count", x="Marital", title = "Marital")
```



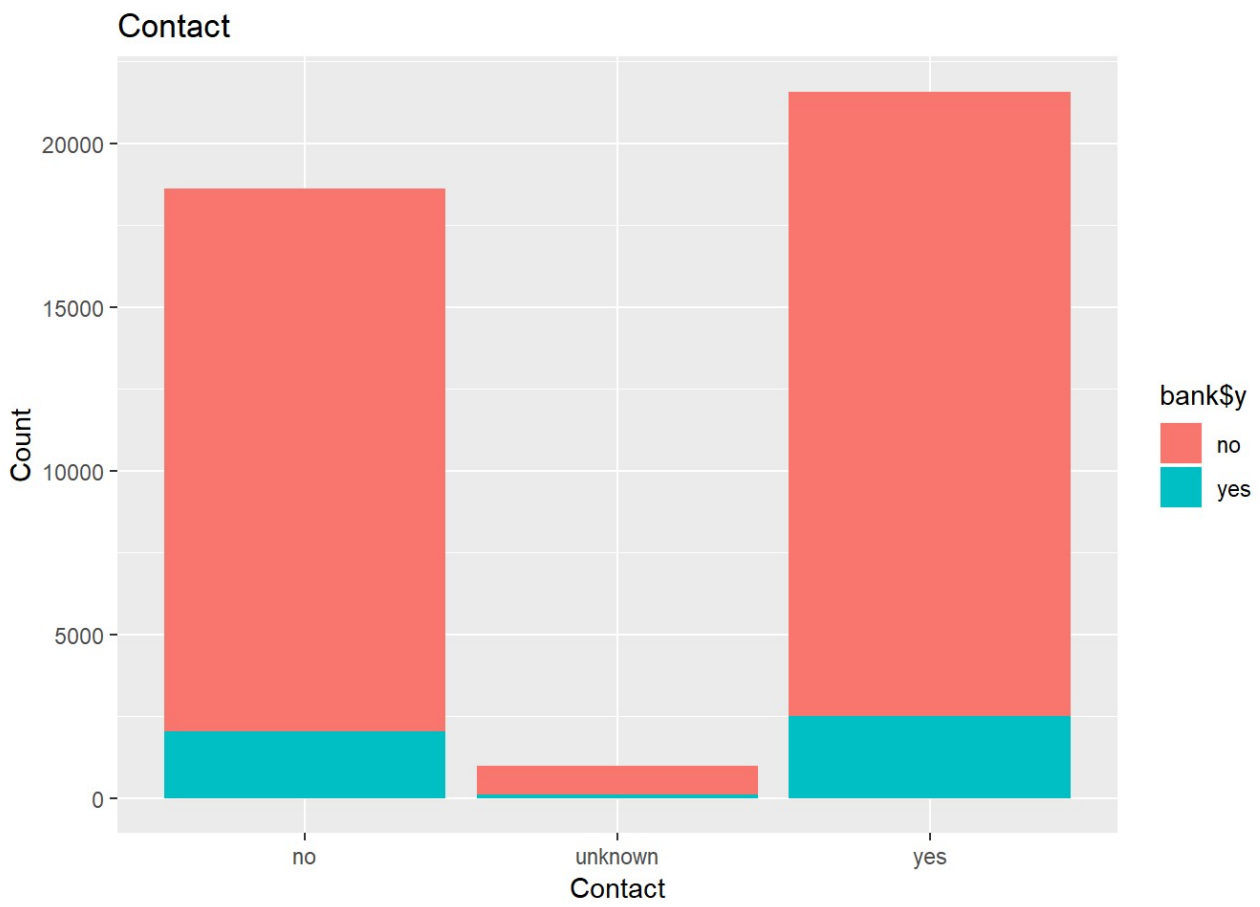

```
ggplot(bank, aes(x=bank$education,fill=bank$y)) + geom_bar()+
  labs(y= "Count", x="Education", title = "Education")
```



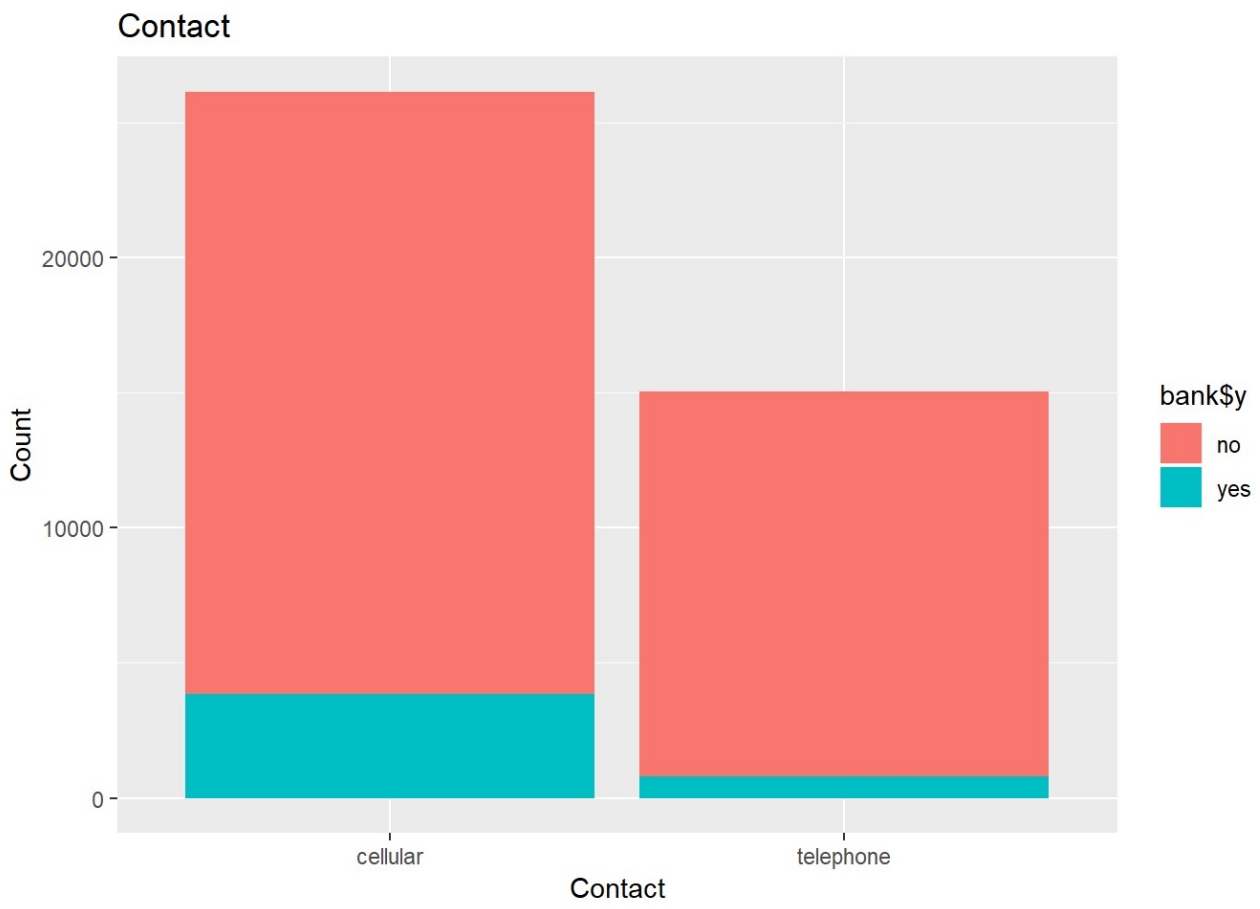

```
ggplot(bank, aes(x=bank$default,fill=bank$y)) + geom_bar()+
  labs(y= "Count", x="Contact", title = "Contact")
```



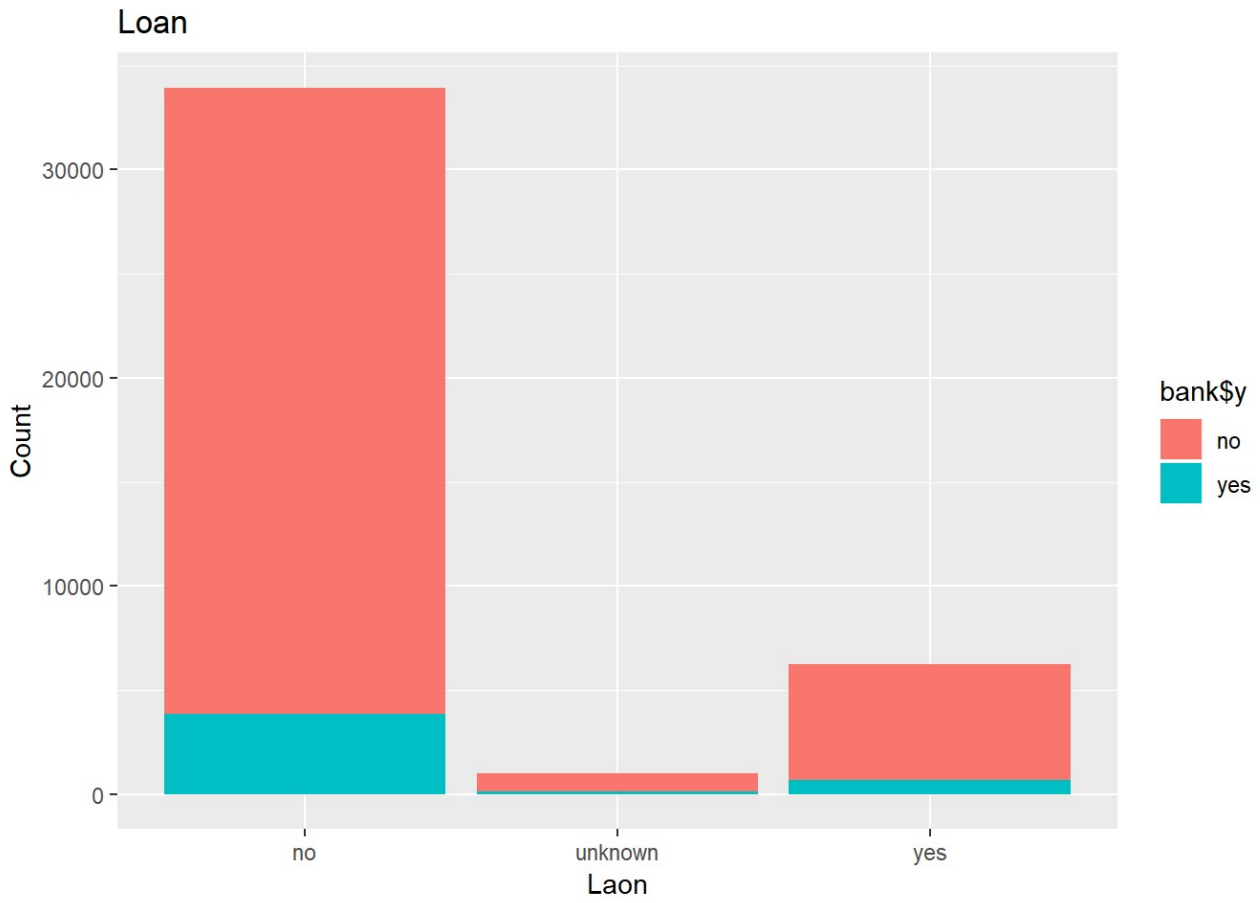

```
ggplot(bank, aes(x=bank$housing,fill=bank$y)) + geom_bar()+
  labs(y= "Count", x="Contact", title = "Contact")
```



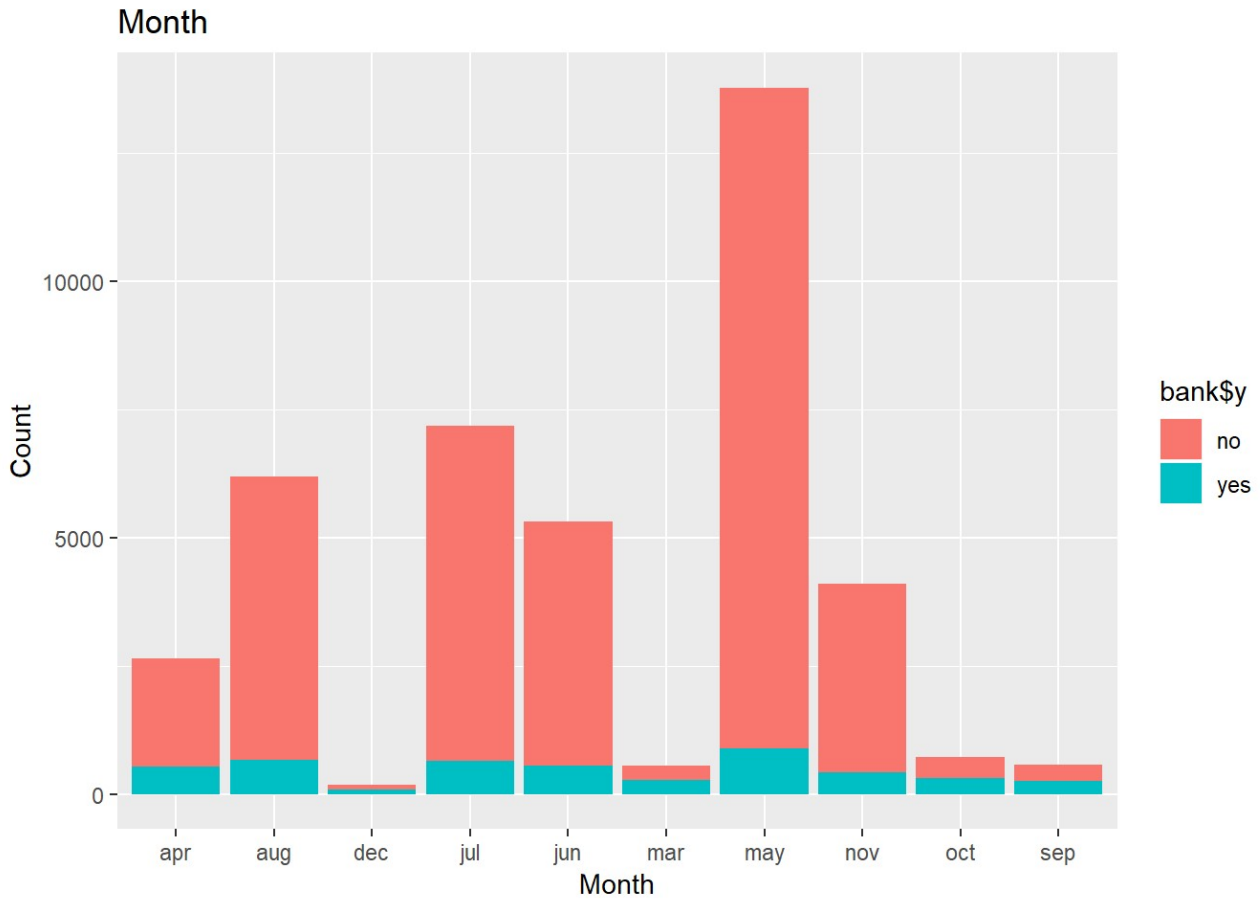

```
ggplot(bank, aes(x=bank$contact,fill=bank$y)) + geom_bar()+
  labs(y= "Count", x="Contact", title = "Contact")
```



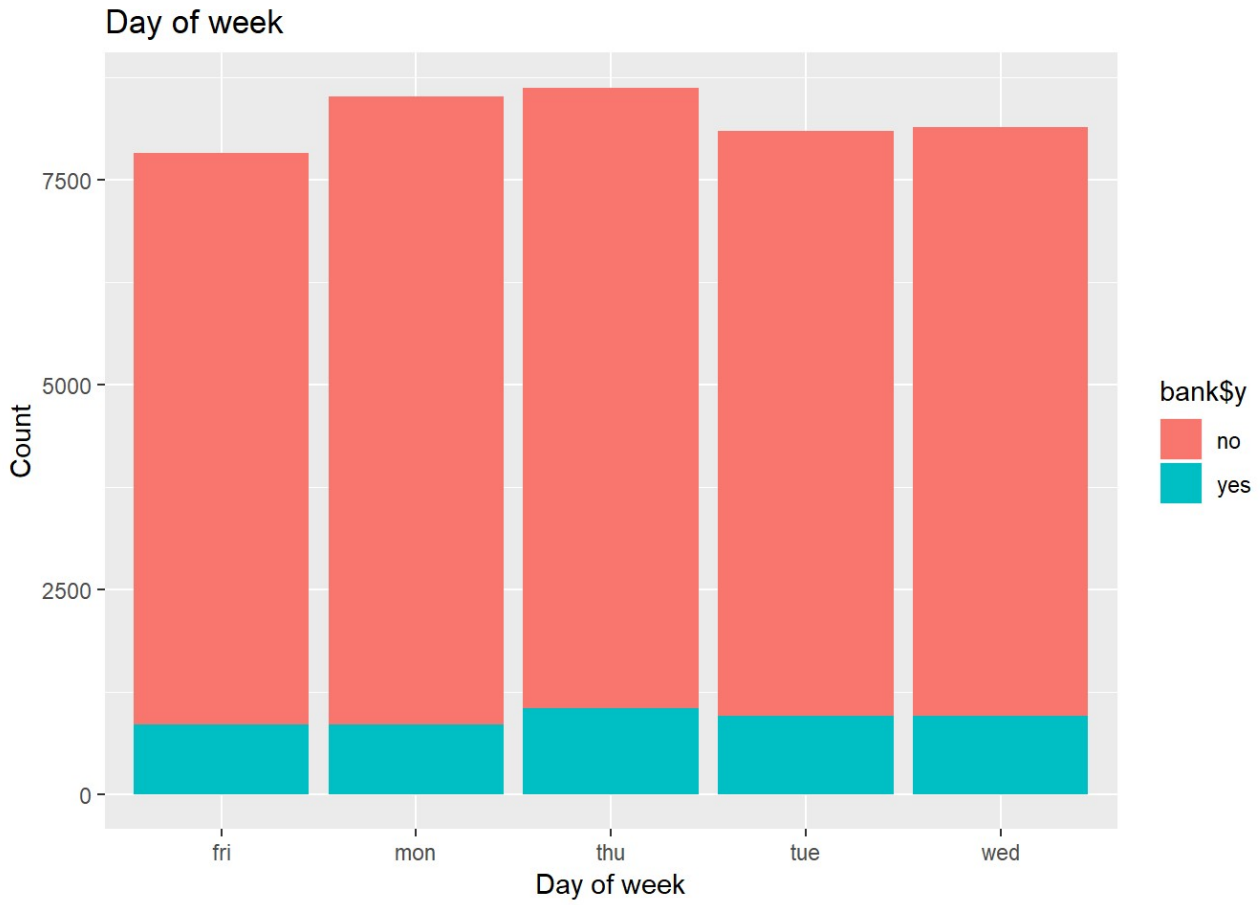

```
ggplot(bank, aes(x=bank$loan,fill=bank$y)) + geom_bar()+
  labs(y= "Count", x="Laon", title = "Loan")
```



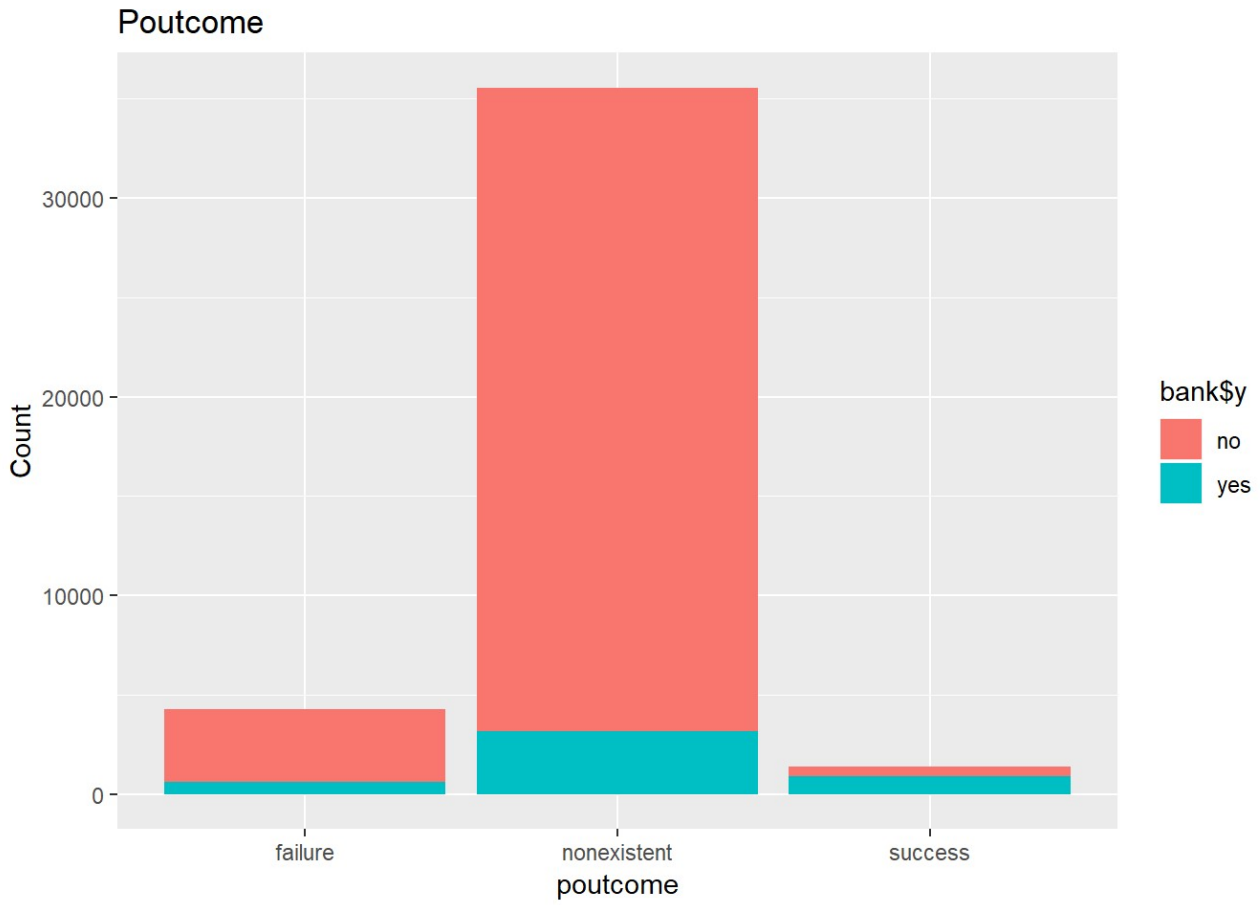

```
ggplot(bank, aes(x=bank$month,fill=bank$y)) + geom_bar()+
  labs(y= "Count", x="Month", title = "Month")
```



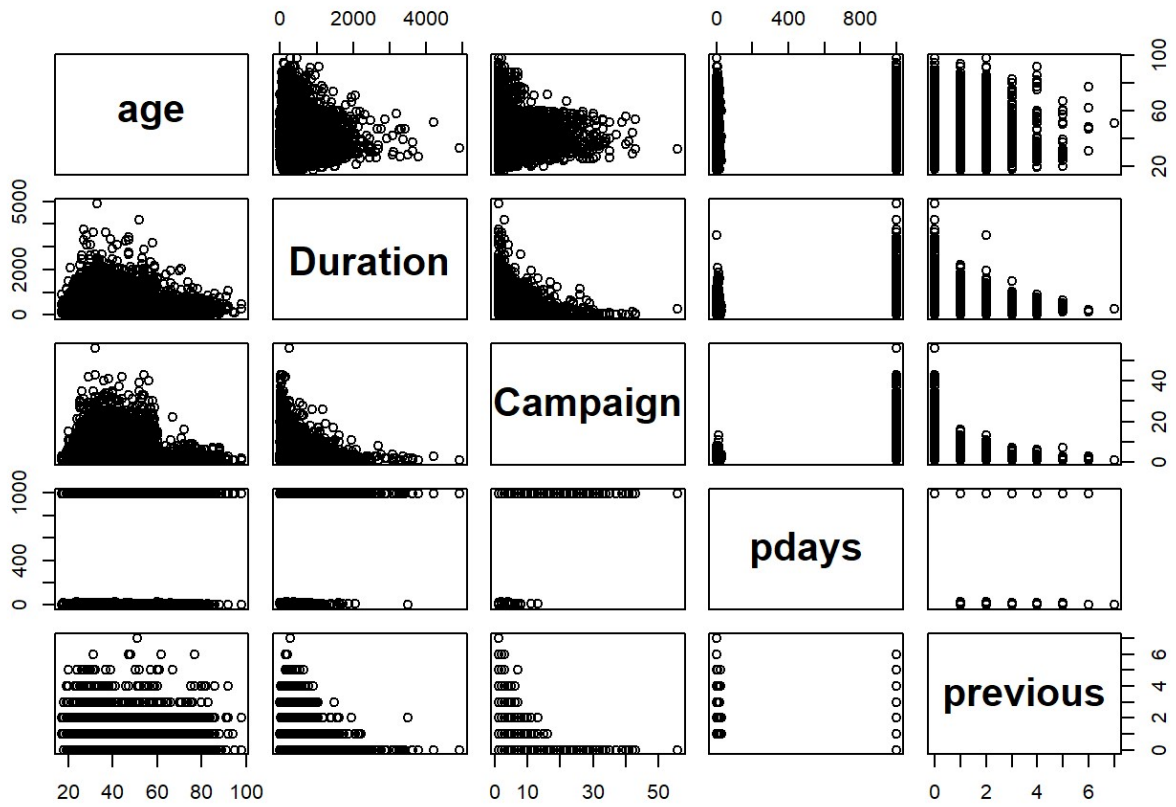

```
ggplot(bank, aes(x=bank$day_of_week,fill=bank$y)) + geom_bar()+
  labs(y= "Count", x="Day of week", title = "Day of week")
```




```
ggplot(bank, aes(x=bank$poutcome,fill=bank$y)) + geom_bar() +
  labs(y= "Count", x="poutcome", title = "Poutcome")
```



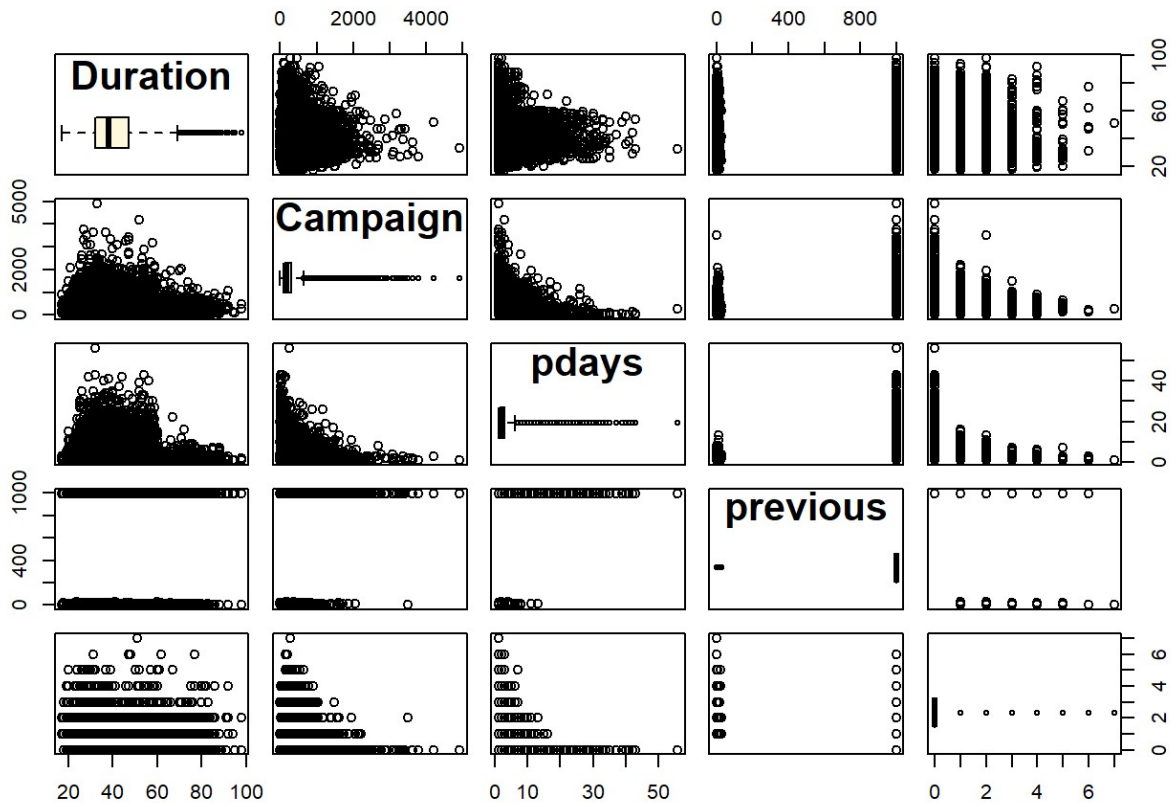

```
#correlation analysis
bank_int=bank[c(1,11:14)]
pairs(bank_int, labels=c("age","Duration","Campaign","pdays","previous"),pch=c(1,16)[y
_int],font.labels=2)
```




```
#Diagonal boxplot
library(SciViews)
```

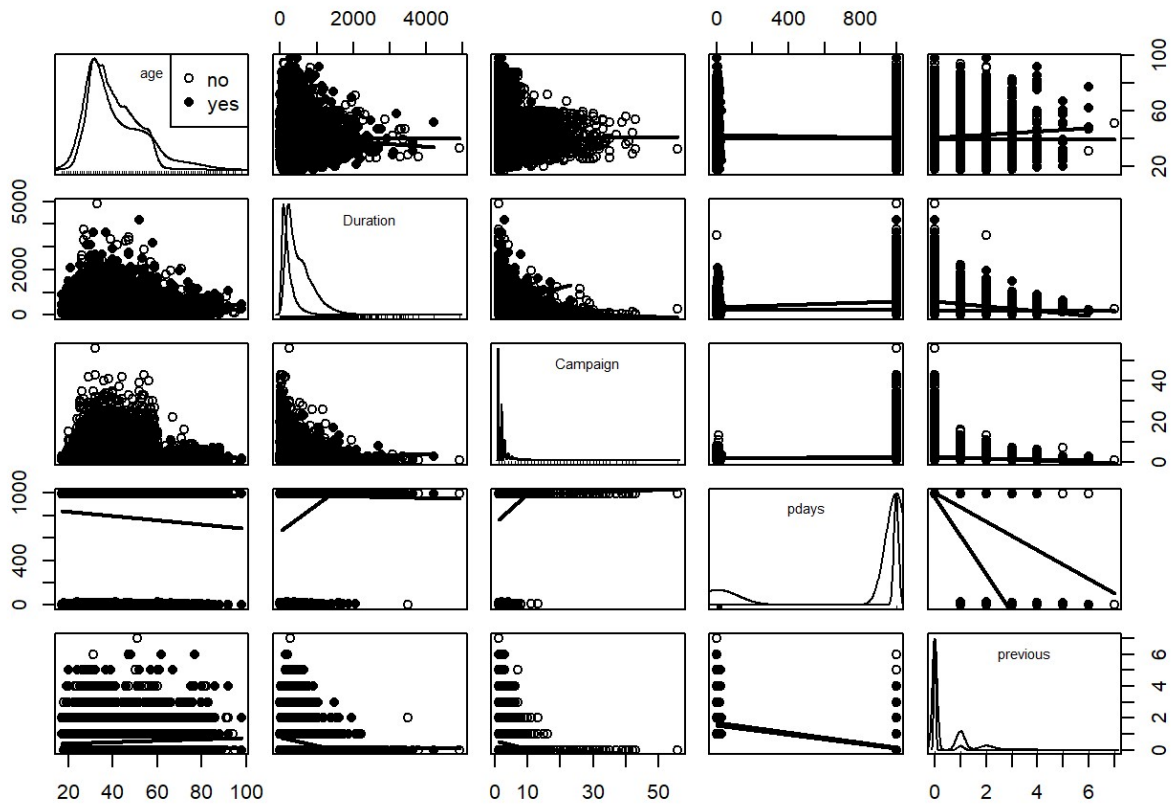
```
## Warning: package 'SciViews' was built under R version 3.5.2
```

```
pairs(bank_int, diag.panel = panel.boxplot, labels=c("Duration","Campaign","pdays","pr
evious"),pch=c(1,16)[y_int], font.labels=2)
```




```
library(car)
```

```
scatterplotMatrix(~age+duration+campaign+pdays+previous | bank$y, data=bank_int, var.l
abels=c("age","Duration","Campaign","pdays","previous"),cex.labels=0.7, diagonal="boxp
lot",smooth=FALSE,reg.line=FALSE,pch=c(1,16),col=rep("black",2), legend.plot=FALSE)
```




```
#Instead of using splom using psych library it includes splom , and give better correl
ation for factor features
library(psych)
```

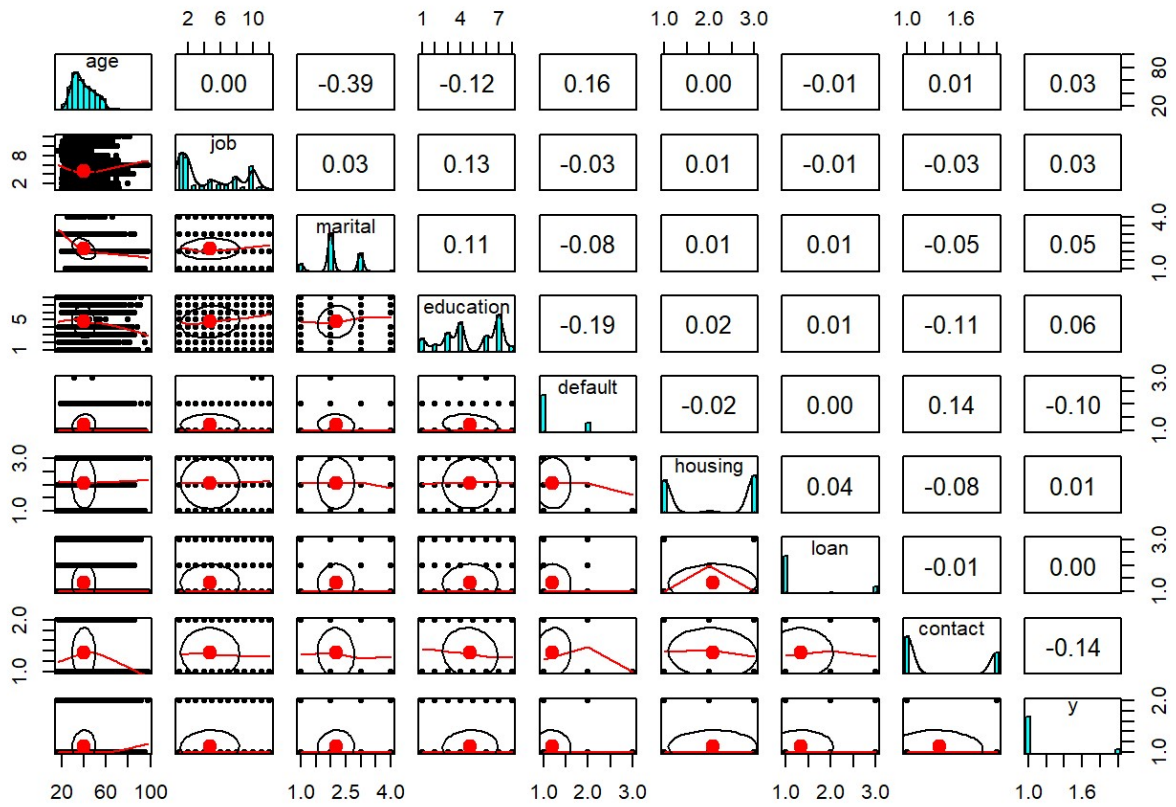
```
## Warning: package 'psych' was built under R version 3.5.2
```

```
##
## Attaching package: 'psych'
```

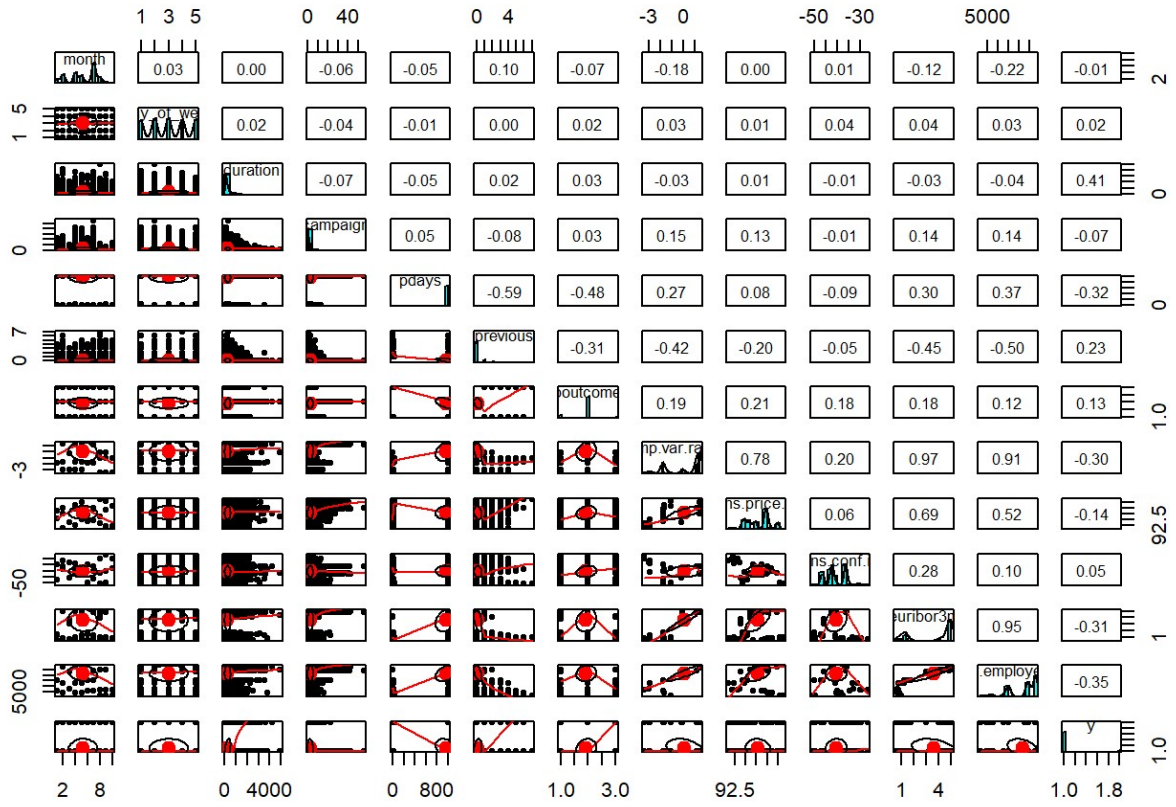
```
## The following object is masked from 'package:car':
##
##     logit
```

```
## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
```

```
pairs.panels(bank[,c(1:8,21)])
```




```
pairs.panels(bank[,c(9:21)])
```


month
0.03
0.00
-0.06
-0.05
0.10
-0.07
-0.18
0.00
0.01
-0.12
-0.22
-0.01
y_of_we
0.02
-0.04
-0.01
0.00
0.02
0.03
0.01
0.04
0.04
0.03
0.02
duration
-0.07
-0.05
0.02
0.03
-0.03
0.01
-0.01
-0.03
-0.04
0.41
ampaign
0.05
-0.08
0.03
0.15
0.13
-0.01
0.14
0.14
-0.07
pdays
-0.59
-0.48
0.27
0.08
-0.09
0.30
0.37
-0.32
previous
-0.31
-0.42
-0.20
-0.05
-0.45
-0.50
0.23
outcome
0.19
0.21
0.18
0.18
0.12
0.13
0.78
0.20
0.97
0.91
-0.30
0.06
0.69
0.52
-0.14
0.28
0.10
0.05
0.95
-0.31
-0.35
y
1 3 5
0 40
0 4
-3 0
-50 -30
5000
2 8
0 4000
0 800
1.0 3.0
92.5
1 4
1.0 1.8