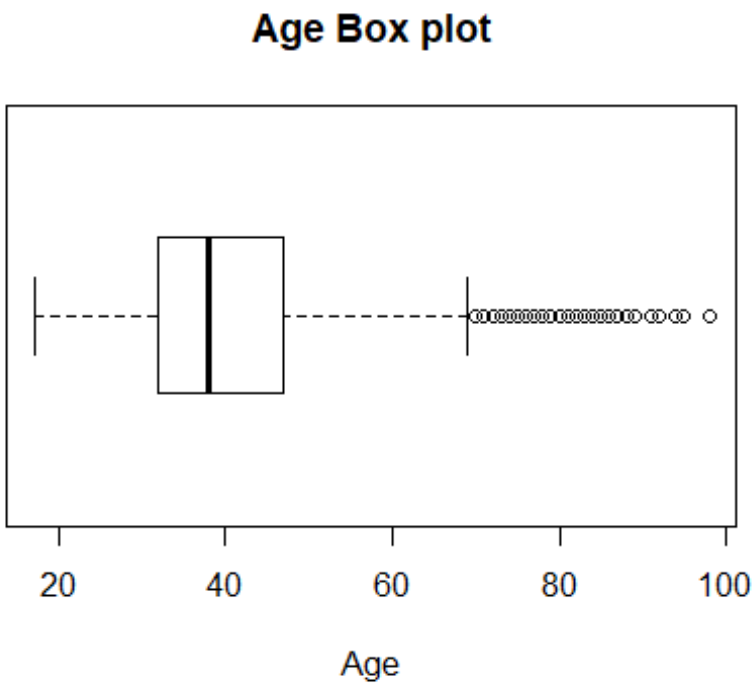
Bank_Data-_EDA.R

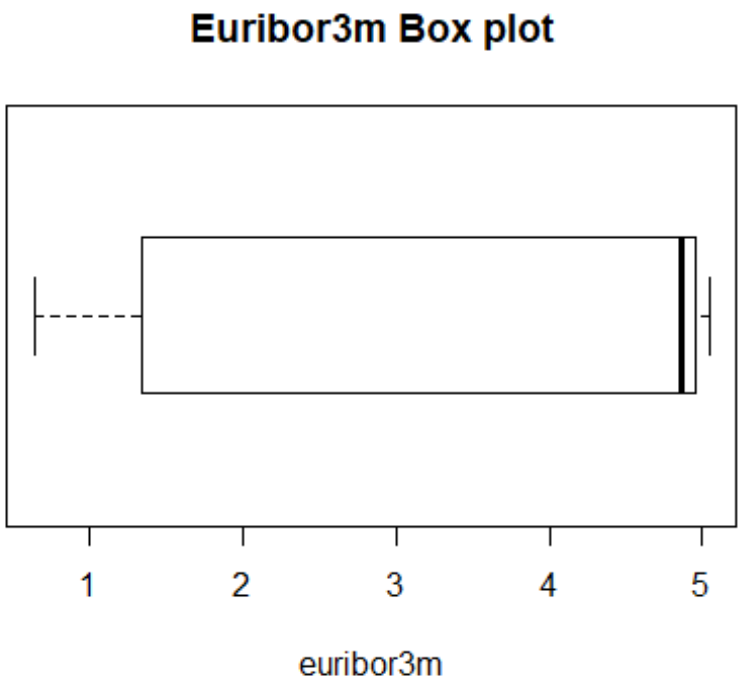
deept

Fri Feb 15 17:21:10 2019

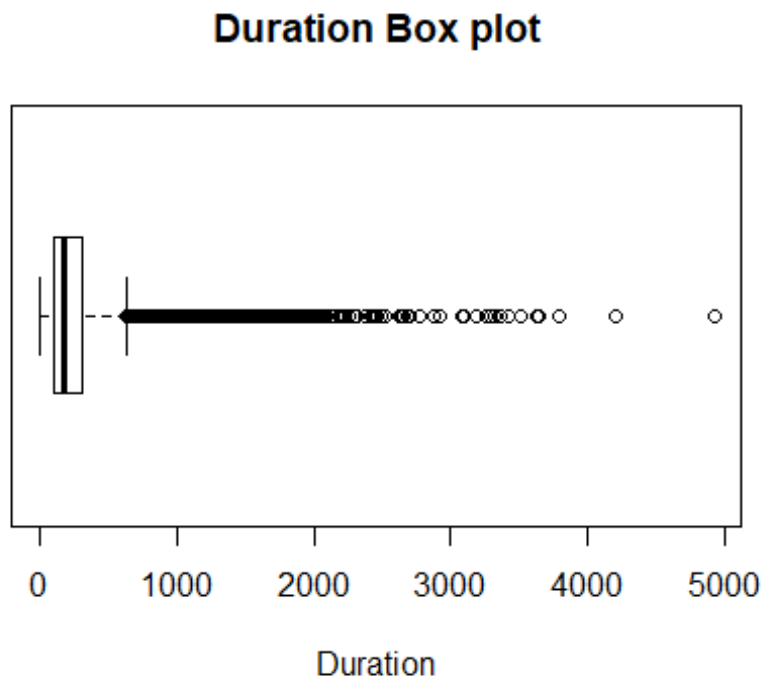
```
bank <- read.csv("~/Spring 19 Sem/Multi Analysis/bank-additional/bank-additional-full.csv", sep=";")


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



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



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



```r
library(MVA)
```

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

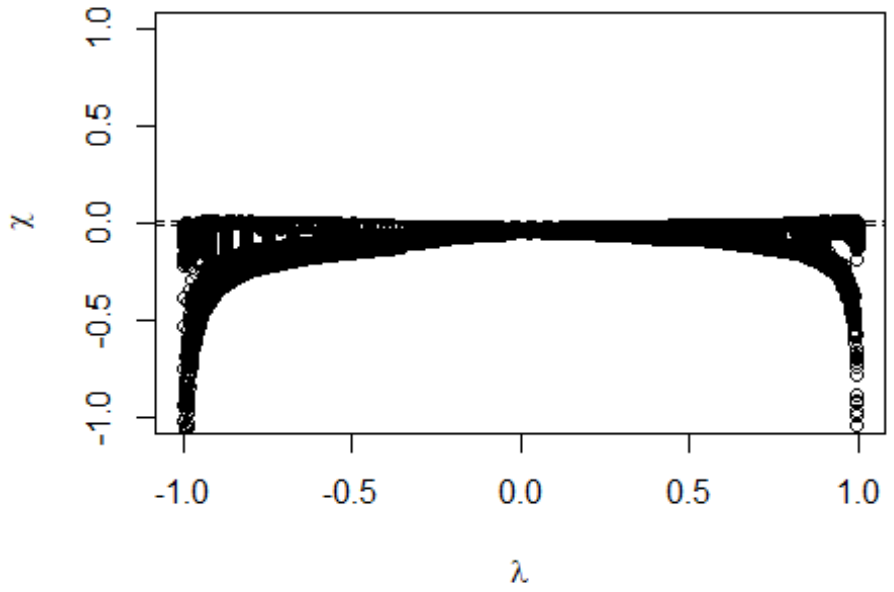
```
## Loading required package: HSAUR2
```

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

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

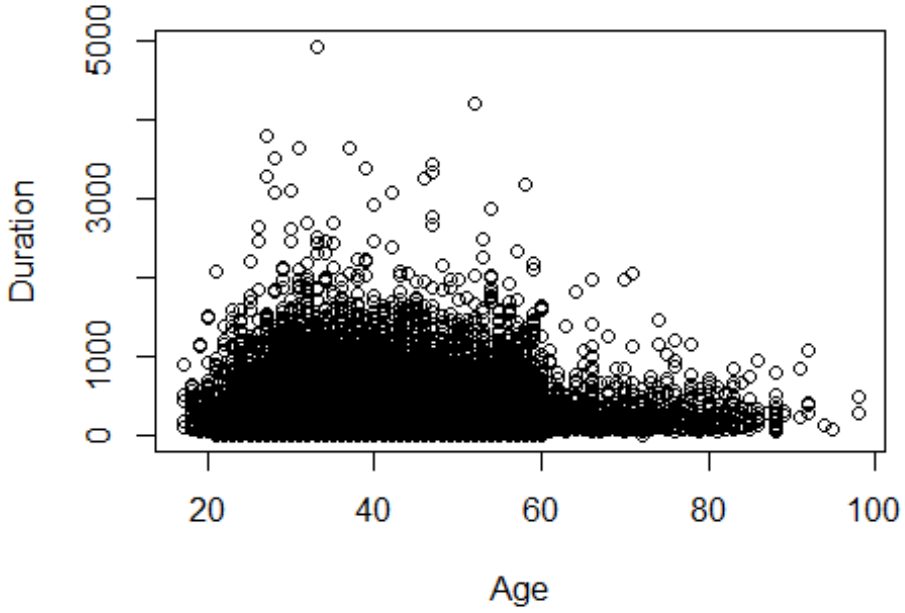
```r
#Chiplot
mlab = "Age of the Customer"
plab = "Duration"
with(bank, plot(age, duration, xlab = mlab, ylab = plab, cex.lab = 0.9))
```

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

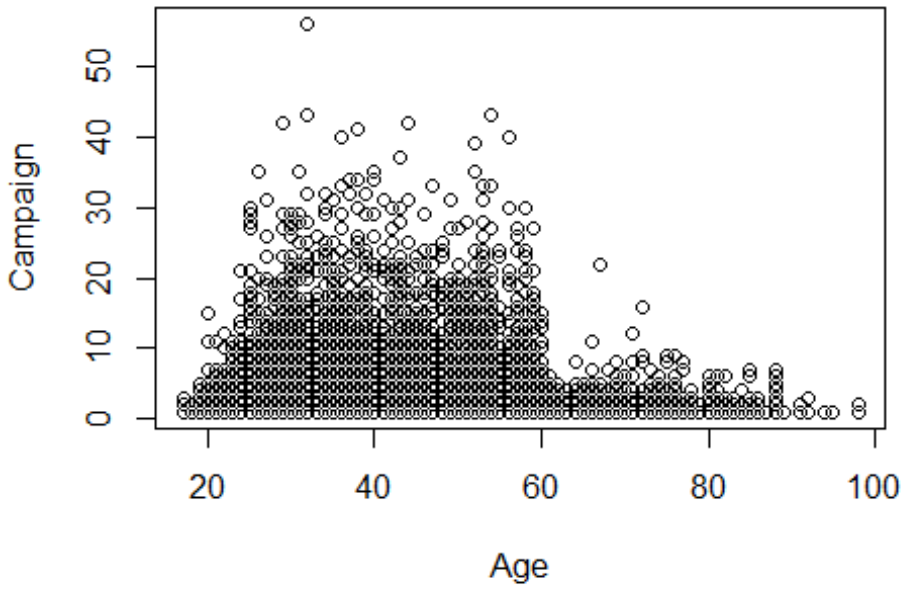


```r
#bvplot
bank_age_dur=data.frame(bank$age, bank$duration)
bvbox(bank_age_dur, mtitle = "", xlab = mlab, ylab = plab)
```

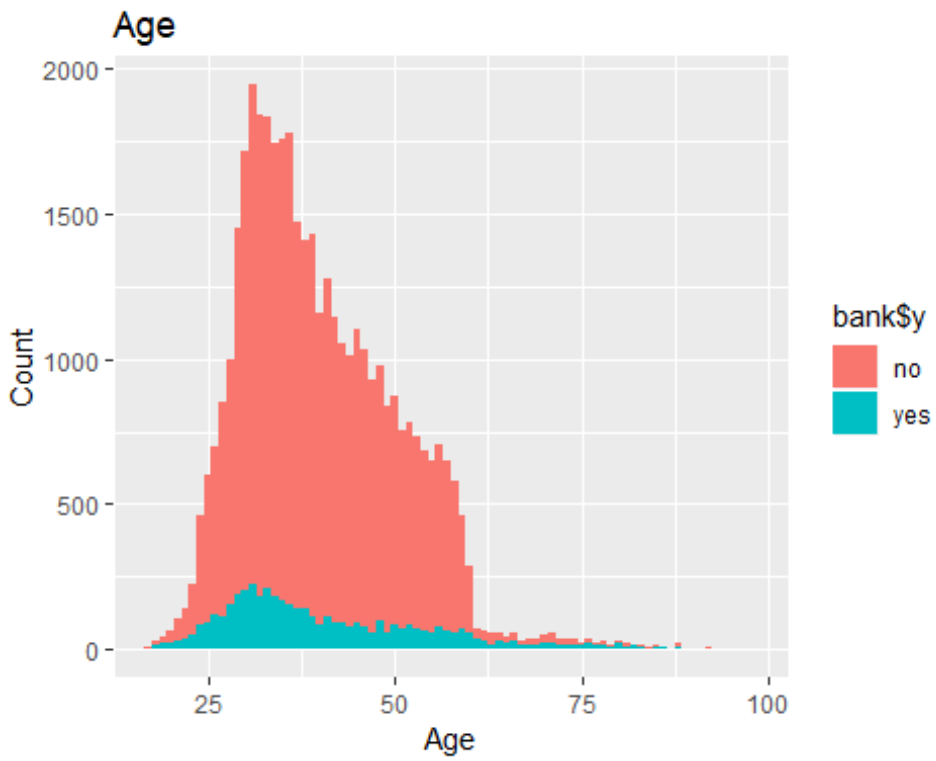
```r
y_int=ifelse(bank$y=='no', 0, 1)

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



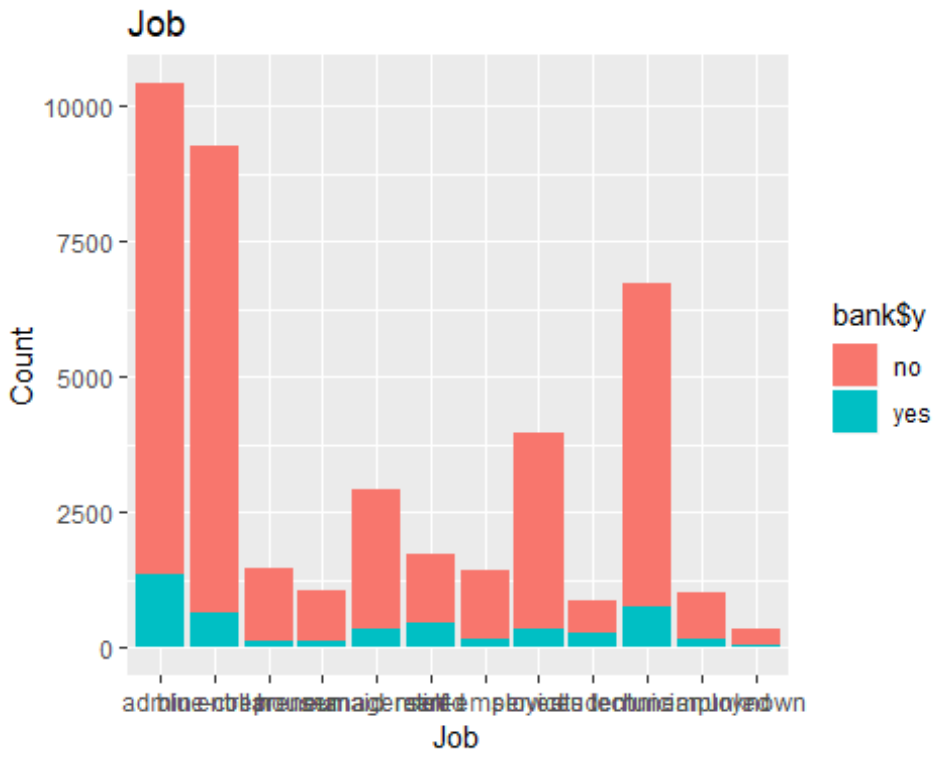
```r
plot(bank$age, bank$campaign, pch=c(1,16)[y_int],xlab="Age", ylab="Campaign")

library(ggplot2)
```



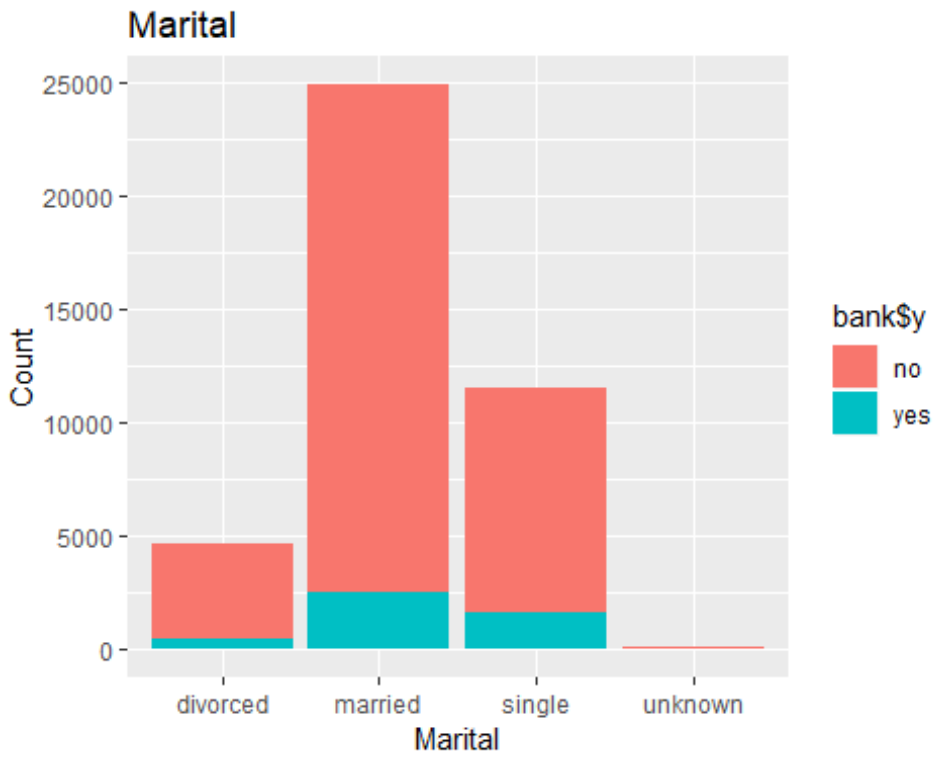
```
ggplot(bank,aes(x=bank$age,fill=bank$y)) + geom_histogram(binwidth=1) +
  labs(y= "Count", x="Age", title = "Age")
```



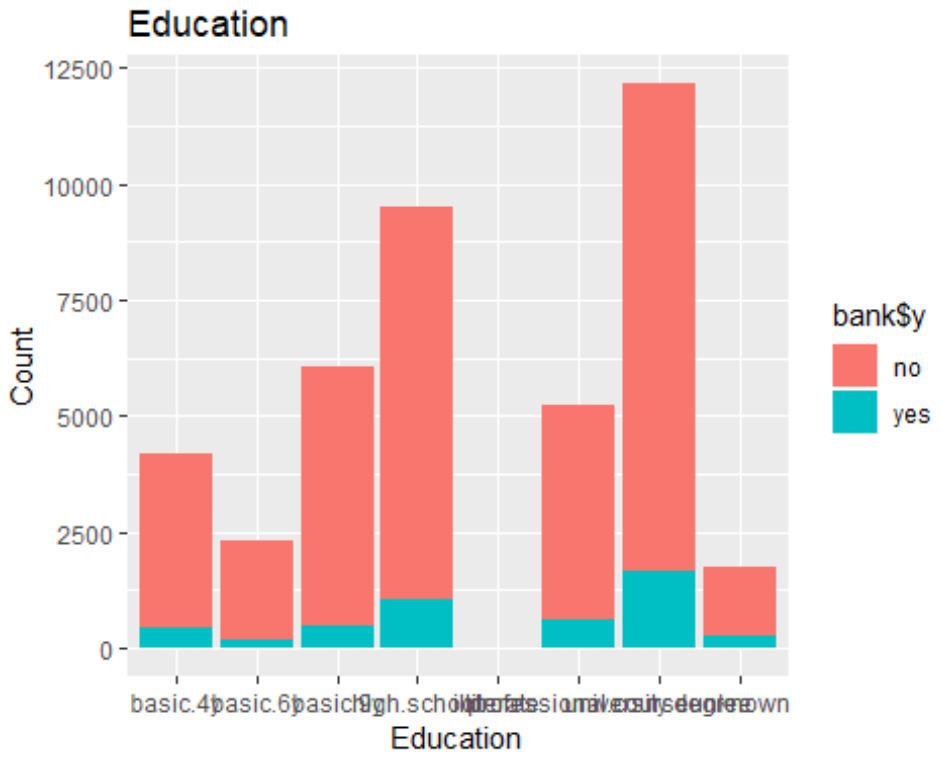

```
ggplot(bank, aes(x=bank$job,fill=bank$y)) + geom_bar()+
  labs(y= "Count", x="Job", title = "Job")
```



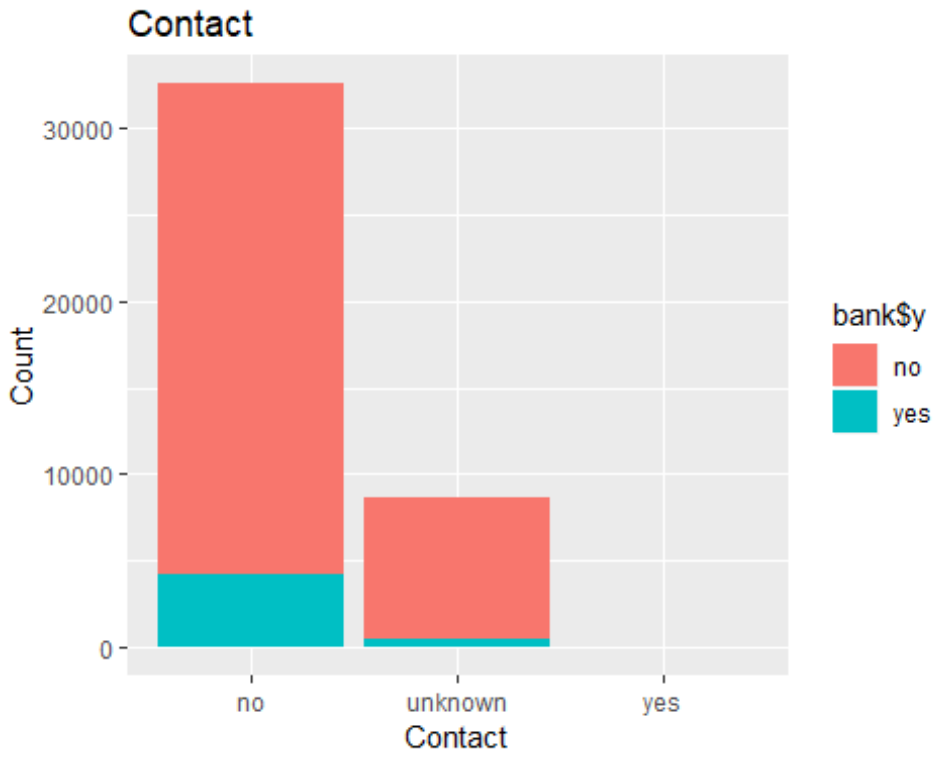

```
ggplot(bank, aes(x=bank$marital,fill=bank$y)) + geom_bar() +
  labs(y= "Count", x="Marital", title = "Marital")
```



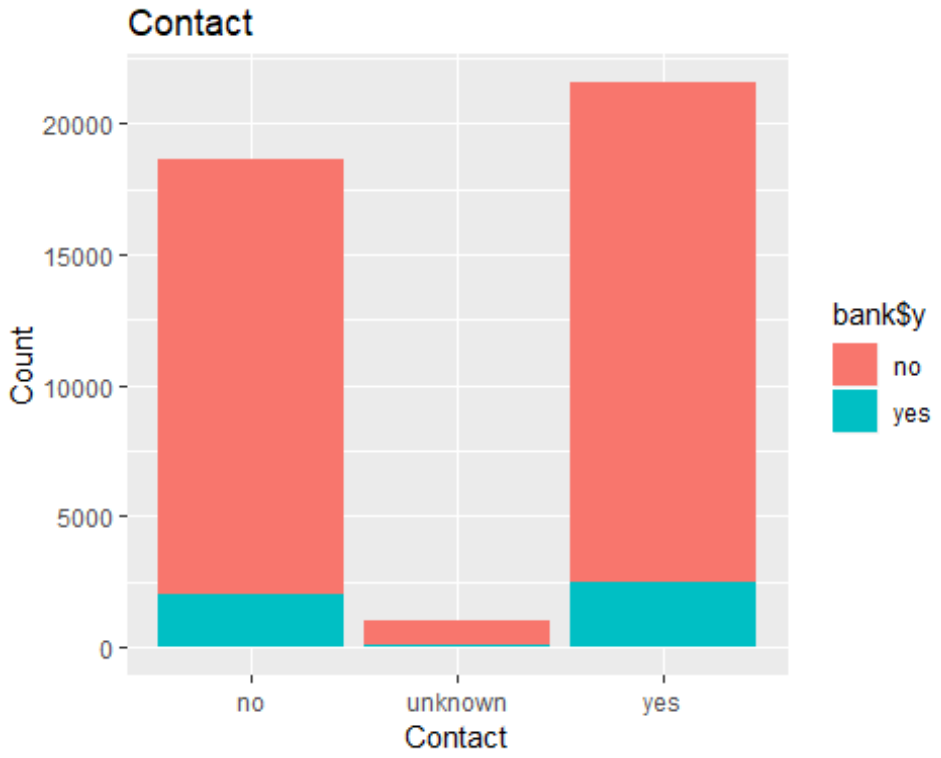
```
ggplot(bank, aes(x=bank$education,fill=bank$y)) + geom_bar()+
  labs(y= "Count", x="Education", title = "Education")
```



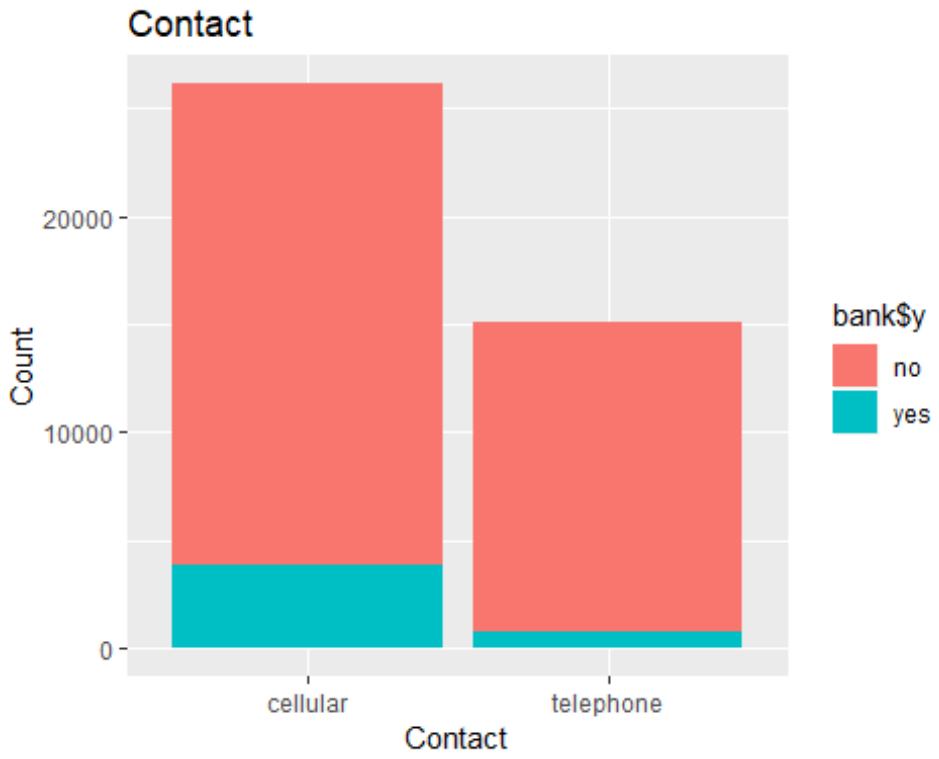

```
ggplot(bank, aes(x=bank$default,fill=bank$y)) + geom_bar()+
  labs(y= "Count", x="Contact", title = "Contact")
```



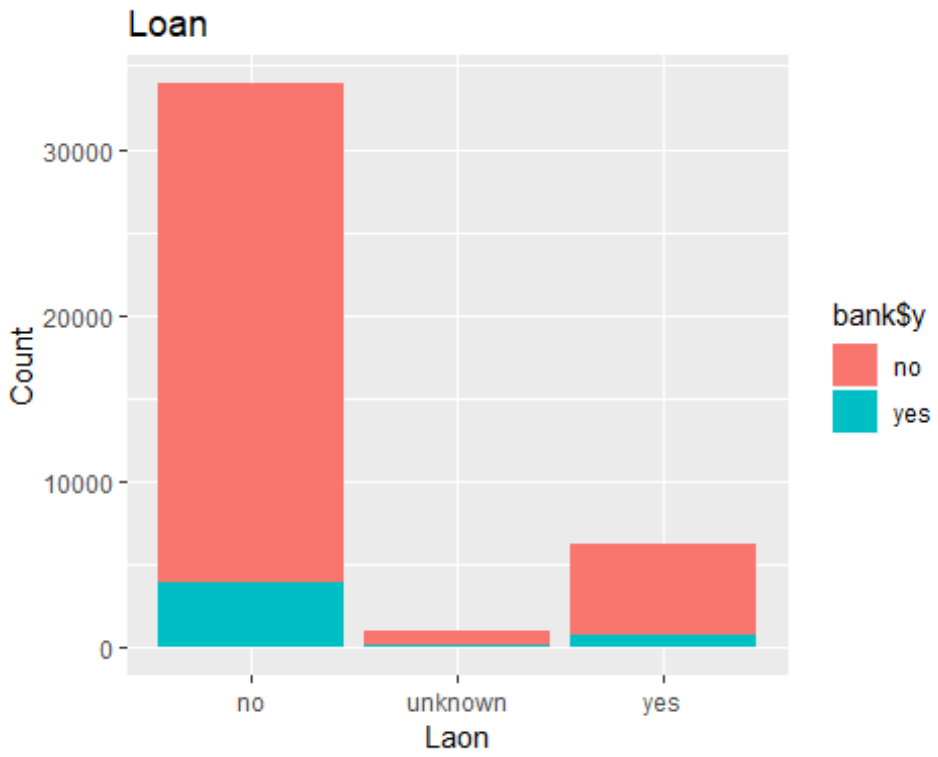

```
ggplot(bank, aes(x=bank$housing,fill=bank$y)) + geom_bar()+
  labs(y= "Count", x="Contact", title = "Contact")
```



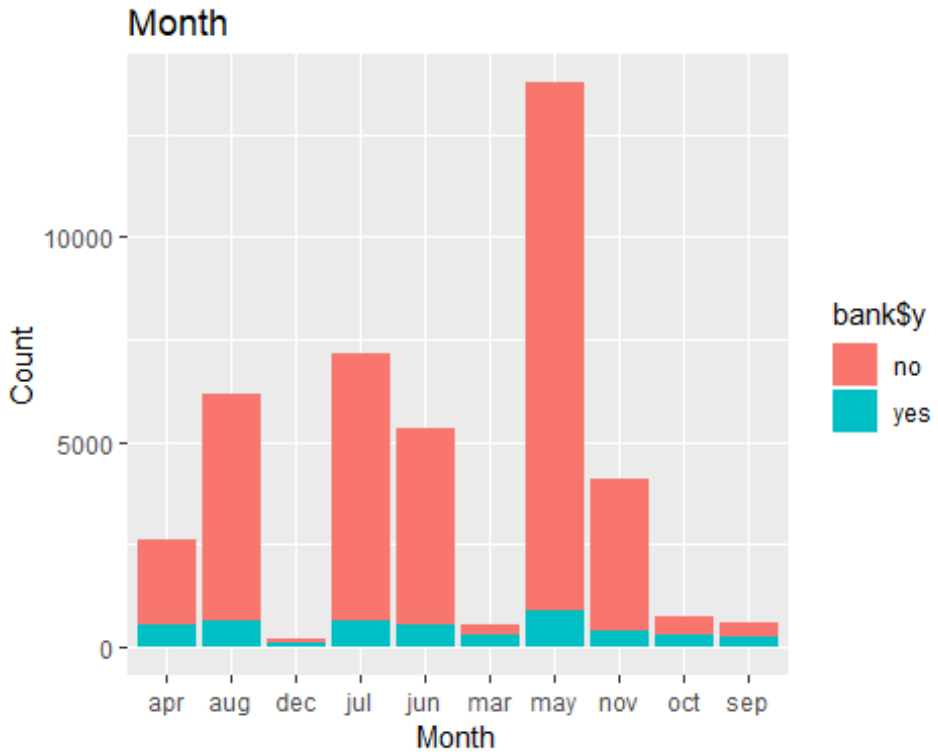
```
ggplot(bank, aes(x=bank$contact,fill=bank$y)) + geom_bar()+
  labs(y= "Count", x="Contact", title = "Contact")
```



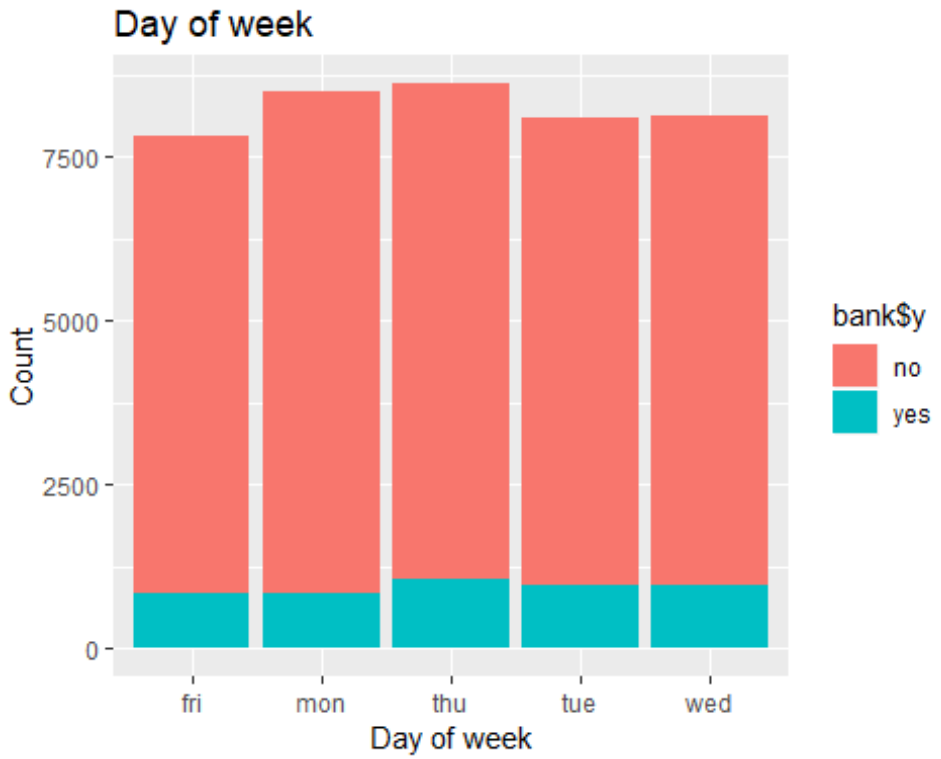

```
ggplot(bank, aes(x=bank$loan,fill=bank$y)) + geom_bar()+
  labs(y= "Count", x="Laon", title = "Loan")
```




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



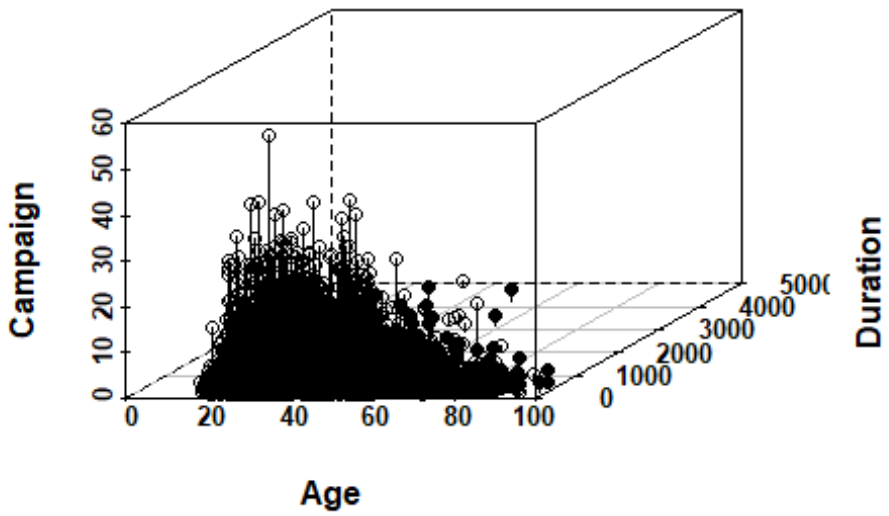
```r
ggplot(bank, aes(x=bank$day_of_week,fill=bank$y)) + geom_bar()+
  labs(y= "Count", x="Day of week", title = "Day of week")
```



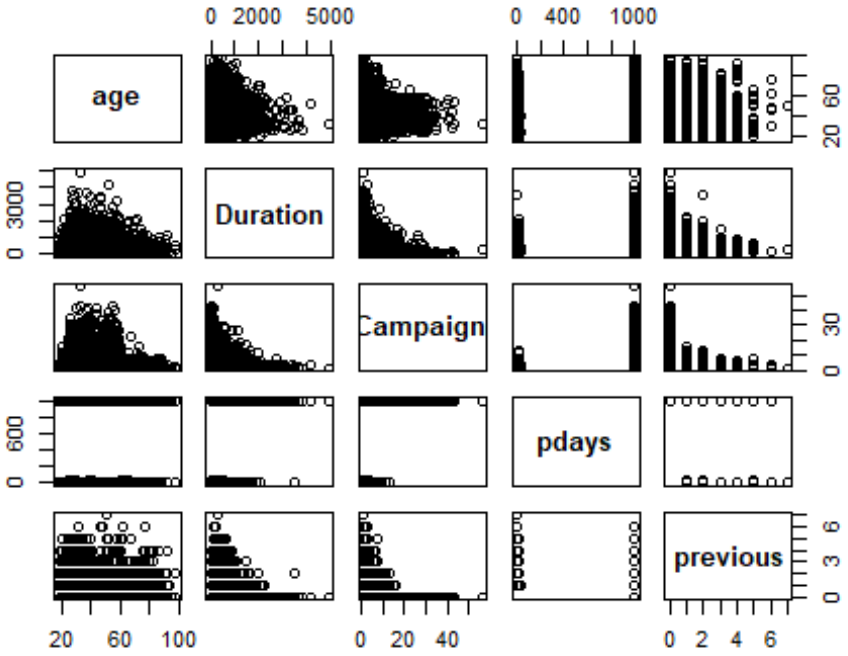

```r
ggplot(bank, aes(x=bank$poutcome,fill=bank$y)) + geom_bar() +
  labs(y= "Count", x="poutcome", title = "Poutcome")
```


Poutcome
30000
20000
10000
0
Count
failure
nonexistent
success
poutcome
bank$y
no
yes

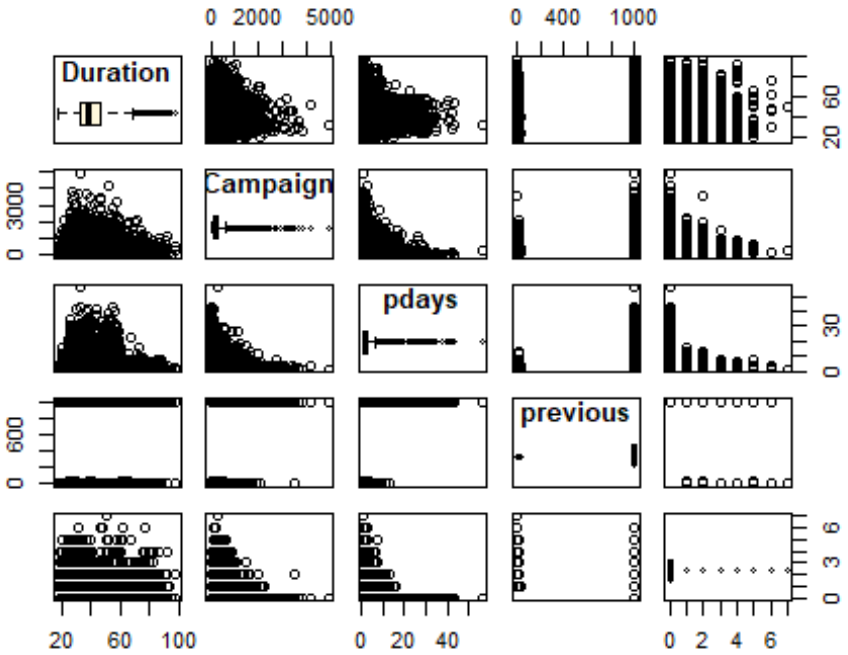
```r
#3d scatterplot
library(scatterplot3d)

## Warning: package 'scatterplot3d' was built under R version 3.5.2

s3d <-
scatterplot3d(bank$age,bank$duration,bank$campaign,pch=c(1,16)[as.numeric(ban
k$y)],xlab="Age", ylab="Duration", angle=45,zlab="Campaign",
lty.hide=2,type="h",y.margin.add=0.1,font.axis=2,font.lab=2)
```



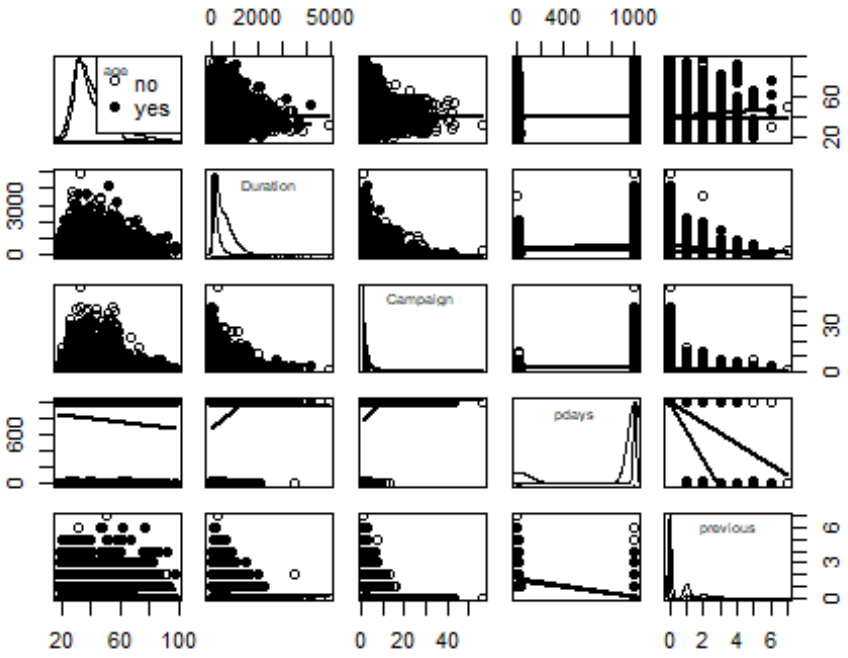
```r
#correlation analysis
bank_int=bank[c(1,11:14)]
pairs(bank_int,
labels=c("age","Duration","Campaign","pdays","previous"),pch=c(1,16)[y_int],f
ont.labels=2)
```



```r
#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","previous"),pch=c(1,16)[y_int],
font.labels=2)
```



```
library(car)

## Warning: package 'car' was built under R version 3.5.2

## Loading required package: carData

## Warning: package 'carData' was built under R version 3.5.2

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



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

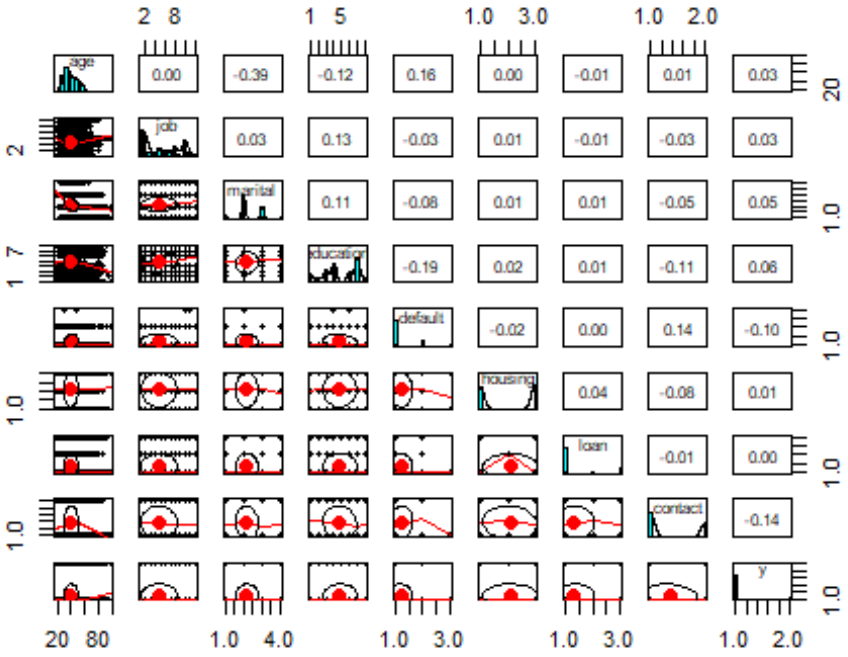
```
## Warning: package 'psych' was built under R version 3.5.2
```

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

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

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

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



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