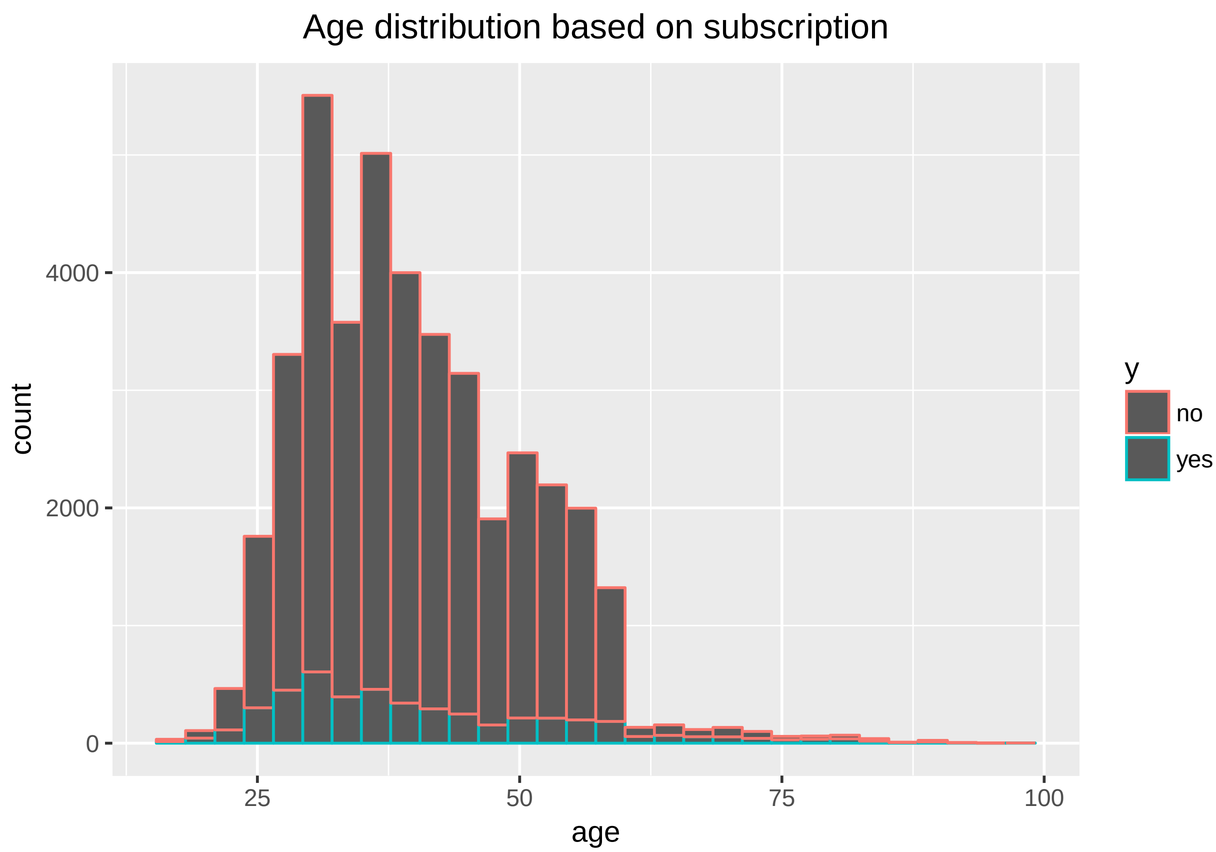
Bank Marketing Project

```
### Importing Data ###

setwd("~/Downloads/ML/bank-additional")

# Importing the csv file and keeping stringsAsFactors= T for automatically converting all
# the string variables into factors
bank <- read.table("bank-additional-full.csv",header=TRUE,sep=";")
str(bank)
```

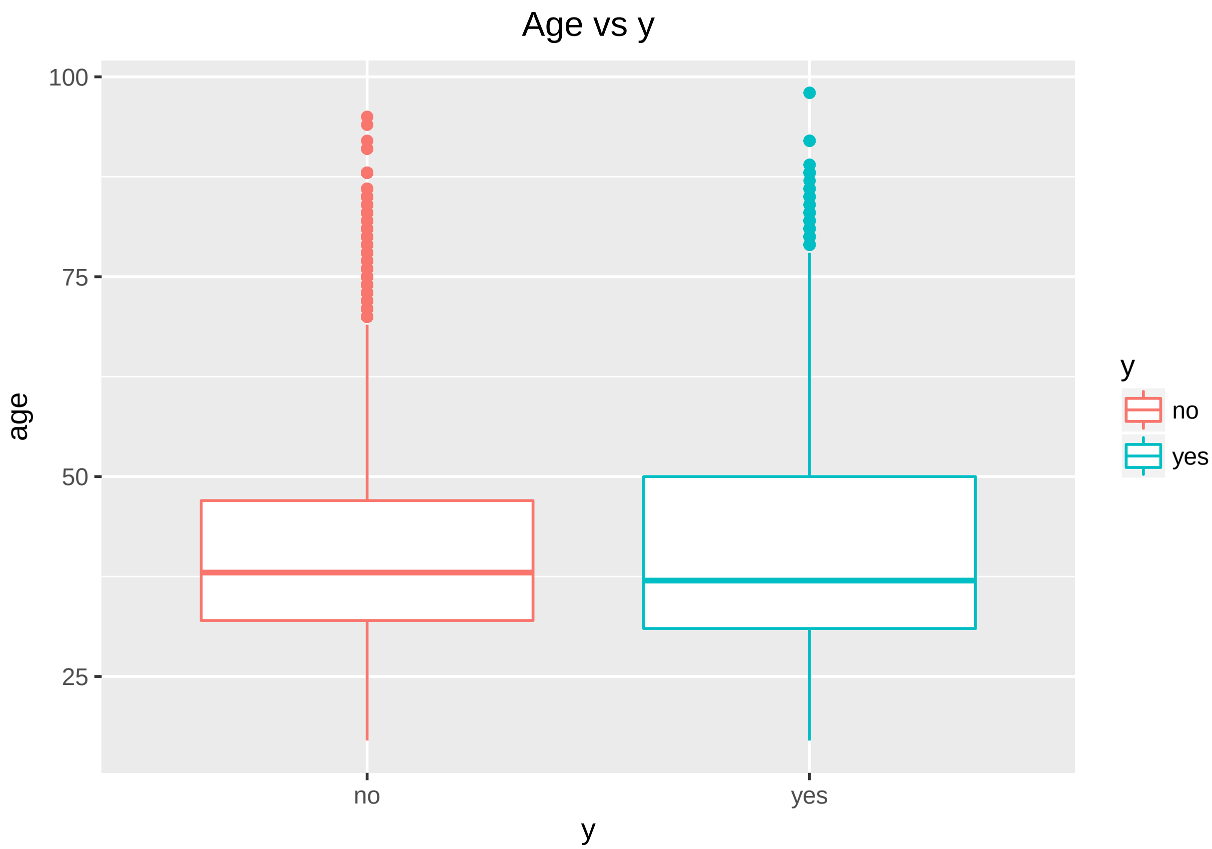
```
## 'data.frame':    41188 obs. of  21 variables:
##  $ age           : int  56 57 37 40 56 45 59 41 24 25 ...
##  $ job           : Factor w/ 12 levels "admin.","blue-collar",..: 4 8 8 1 8 8 1 2 10 8 ...
##  $ marital       : Factor w/ 4 levels "divorced","married",..: 2 2 2 2 2 2 2 2 3 3 ...
##  $ education     : Factor w/ 8 levels "basic.4y","basic.6y",..: 1 4 4 2 4 3 6 8 6 4 ...
##  $ default       : Factor w/ 3 levels "no","unknown",..: 1 2 1 1 1 2 1 2 1 1 ...
##  $ housing       : Factor w/ 3 levels "no","unknown",..: 1 1 3 1 1 1 1 1 3 3 ...
##  $ loan          : Factor w/ 3 levels "no","unknown",..: 1 1 1 1 3 1 1 1 1 1 ...
##  $ contact       : Factor w/ 2 levels "cellular","telephone": 2 2 2 2 2 2 2 2 2 2 ...
##  $ month         : Factor w/ 10 levels "apr","aug","dec",..: 7 7 7 7 7 7 7 7 7 7 ...
##  $ day_of_week   : Factor w/ 5 levels "fri","mon","thu",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ duration      : int  261 149 226 151 307 198 139 217 380 50 ...
##  $ campaign      : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ pdays         : int  999 999 999 999 999 999 999 999 999 999 ...
##  $ previous      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ poutcome      : Factor w/ 3 levels "failure","nonexistent",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ 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             : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
```

```
### Exploratory Data Analysis ###

# Visualizing different variables in the data set and their relations with the dependent
# variable y

library(ggplot2)

## The variable age shows that most of the people are between 25 and 60 as expected and it
# is slightly right skewed. And the distribution of people who subscribed is almost evenly
# distributed with repspect to the number of people in that age group. ##
ggplot(data=bank, aes(x=age, col=y))+
    geom_histogram()+
    ggtitle("Age distribution based on subscription")+
    theme(plot.title = element_text(hjust = 0.5))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
# A boxplot of "age vs y" also reflects the same information
ggplot(bank, aes(x=y, y=age, col=y))+
    geom_boxplot()+
    ggtitle("Age vs y")+
    theme(plot.title = element_text(hjust = 0.5))
```




```
## Creating a table to visualize the relationship between the job role of a person and
# variable y indicating whether a person subscribed for the plan or not
table(bank$job, bank$y)
```

```
##
##                  no  yes
##   admin.       9070 1352
##   blue-collar  8616  638
##   entrepreneur 1332  124
##   housemaid     954  106
##   management   2596  328
##   retired      1286  434
##   self-employed 1272  149
##   services     3646  323
##   student       600  275
##   technician   6013  730
##   unemployed    870  144
##   unknown       293   37
```

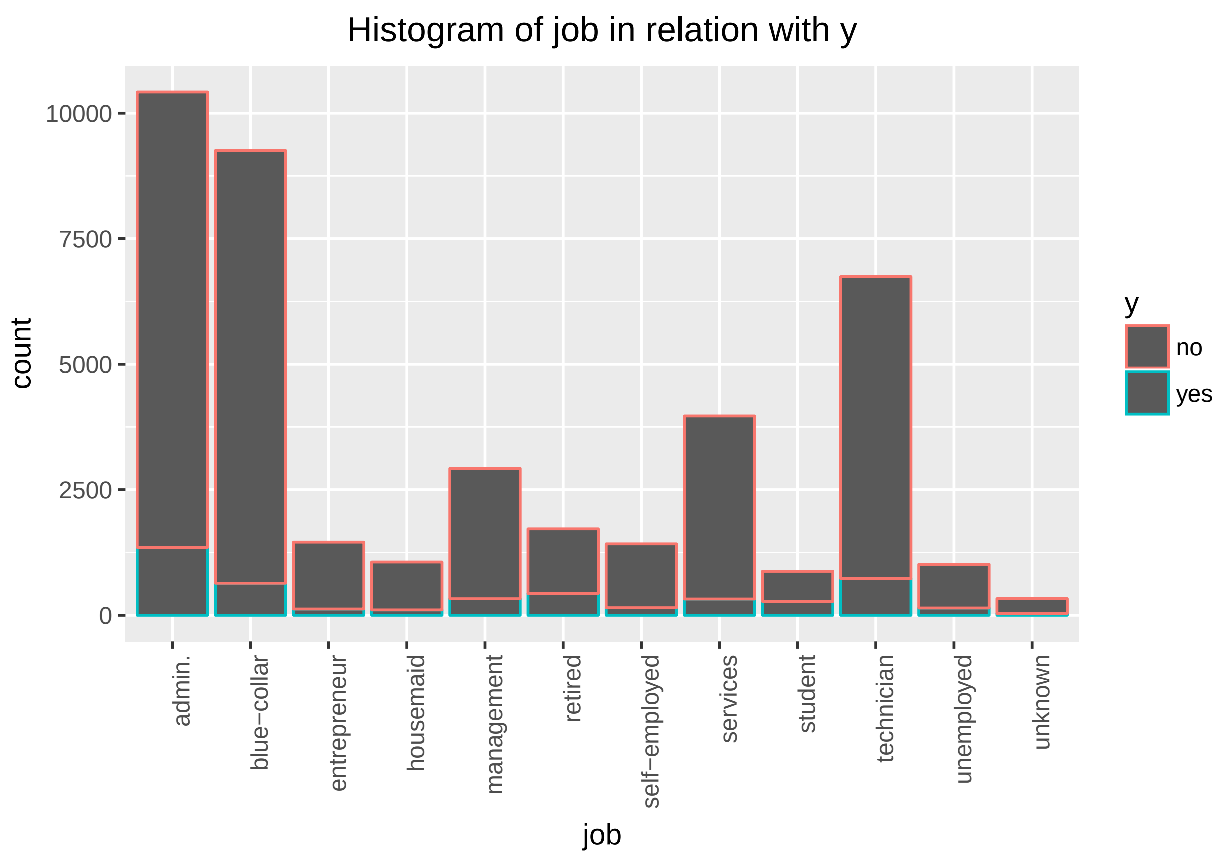
```
# Lets look at the proportion of people subscribing with repect to their job roles
prop.table(table(bank$job, bank$y), 1)
```

```
##
##                         no        yes
##   admin.        0.87027442 0.12972558
##   blue-collar   0.93105684 0.06894316
##   entrepreneur  0.91483516 0.08516484
##   housemaid     0.90000000 0.10000000
##   management    0.88782490 0.11217510
##   retired       0.74767442 0.25232558
##   self-employed 0.89514426 0.10485574
```

```
##    services      0.91861930 0.08138070
##    student       0.68571429 0.31428571
##    technician    0.89173958 0.10826042
##    unemployed    0.85798817 0.14201183
##    unknown       0.88787879 0.11212121
```

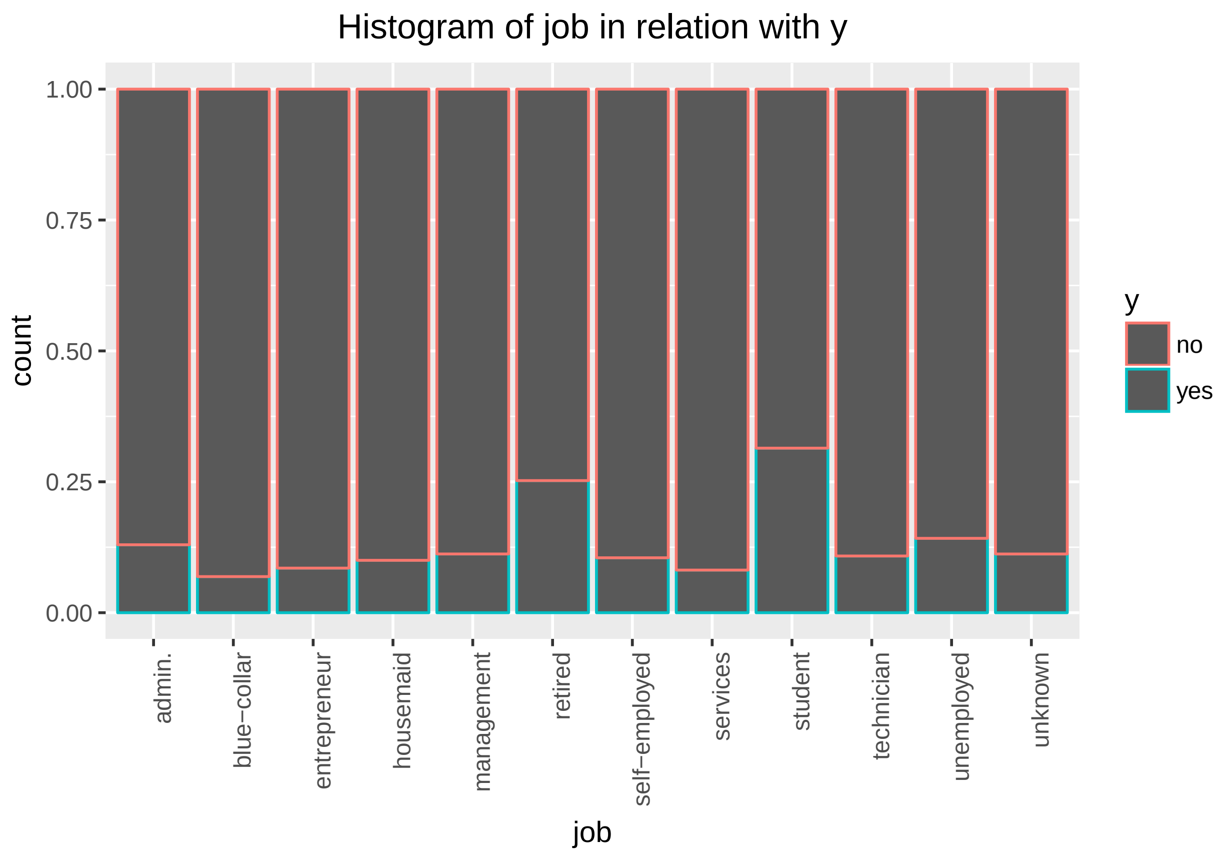
```r
# Plotting the proportions of each category shows that students and retired people have
# very high probability of saying "yes" to the subscription compared to all other
# categories. And blue-collar, entrepreneur andservices are the least probable categories
# for saying "yes". ##
ggplot(bank, aes(x=job, col=y))+
    geom_histogram(stat="count")+
    theme(axis.text.x = element_text(angle = 90, hjust = 1))+
    ggtitle("Histogram of job in relation with y")+
    theme(plot.title = element_text(hjust = 0.5))
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```




```r
ggplot(bank, aes(x=job, col=y))+
    geom_histogram(stat="count", position = "fill")+
    theme(axis.text.x = element_text(angle = 90, hjust = 1))+
    ggtitle("Histogram of job in relation with y")+
    theme(plot.title = element_text(hjust = 0.5))
```

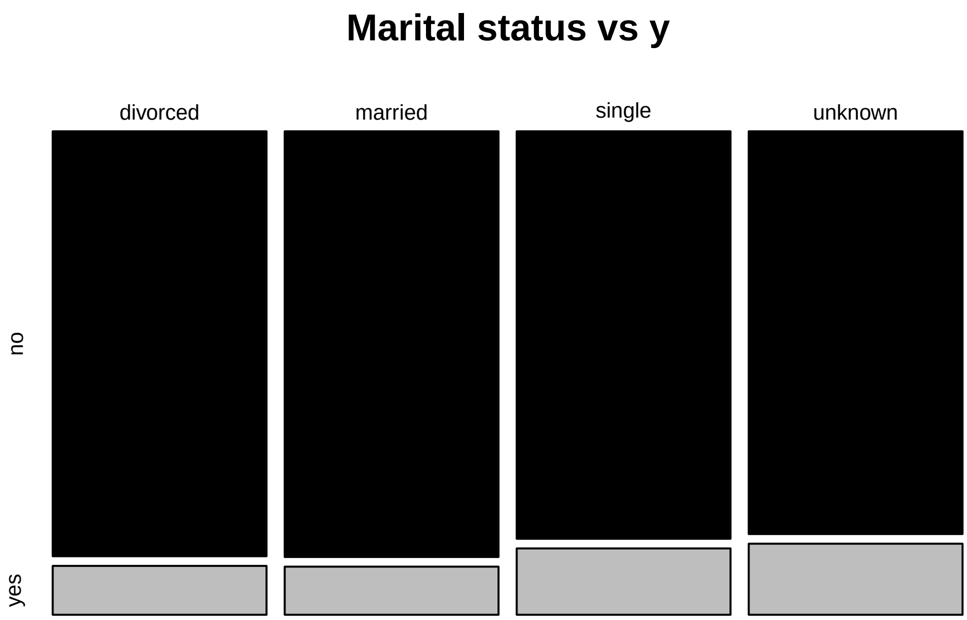
```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```




```
## A similar plot of proportions for marital status vs y shows the spread of y is almost
# evenly distributed and there is very little to no insight ##
prop.table(table(bank$marital, bank$y), 1)
```

```
##
##                    no       yes
##   divorced 0.8967910 0.1032090
##   married  0.8984275 0.1015725
##   single   0.8599585 0.1400415
##   unknown  0.8500000 0.1500000
```

```
plot((prop.table(table(bank$marital, bank$y), 1)),
     main="Marital status vs y", col=c("black","grey"))
```

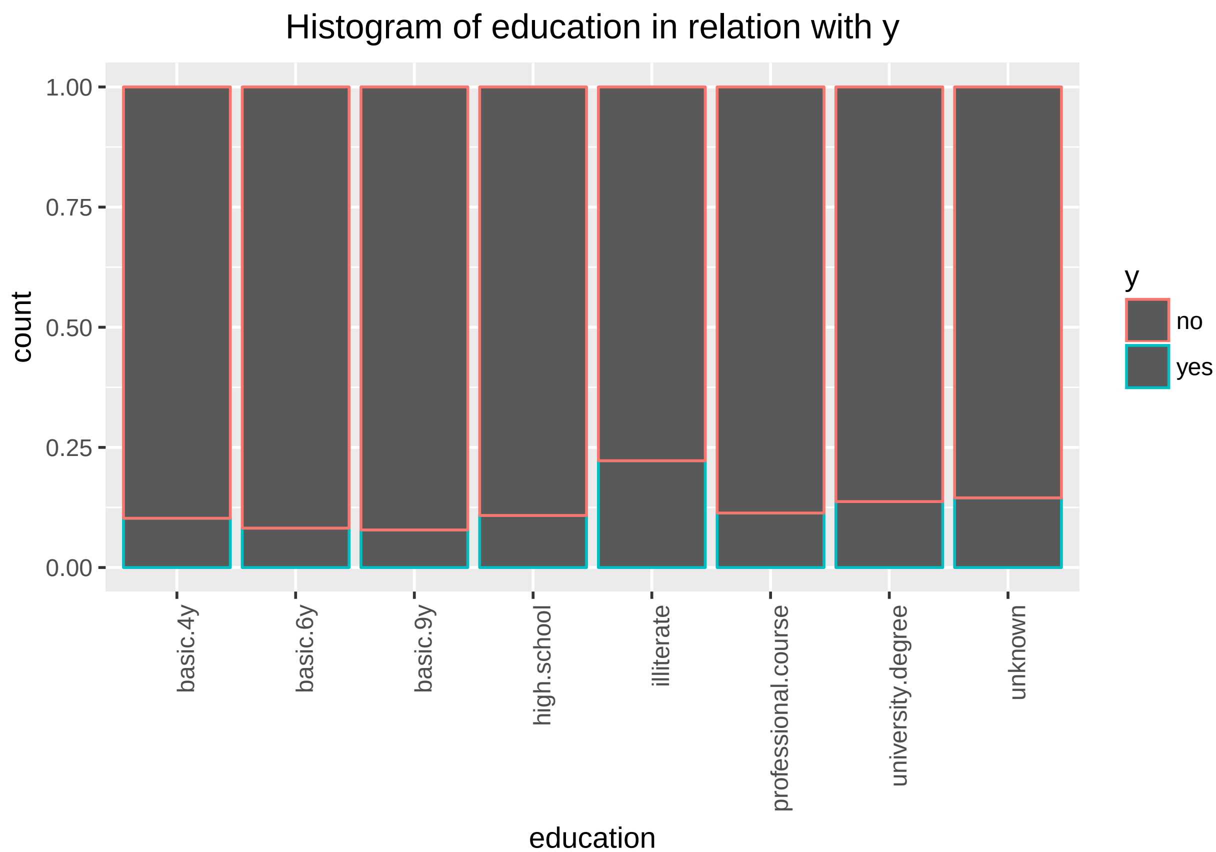



```
## The plot of proportions shows that people who are illiterate, people who has a
# university degree and the unknown category has more chance of taking the subscription
prop.table(table(bank$education, bank$y), 1)
```

```
##
##                              no        yes
##   basic.4y            0.89750958 0.10249042
##   basic.6y            0.91797557 0.08202443
##   basic.9y            0.92175352 0.07824648
##   high.school         0.89164477 0.10835523
##   illiterate          0.77777778 0.22222222
##   professional.course 0.88651535 0.11348465
##   university.degree   0.86275477 0.13724523
##   unknown             0.85499711 0.14500289
```

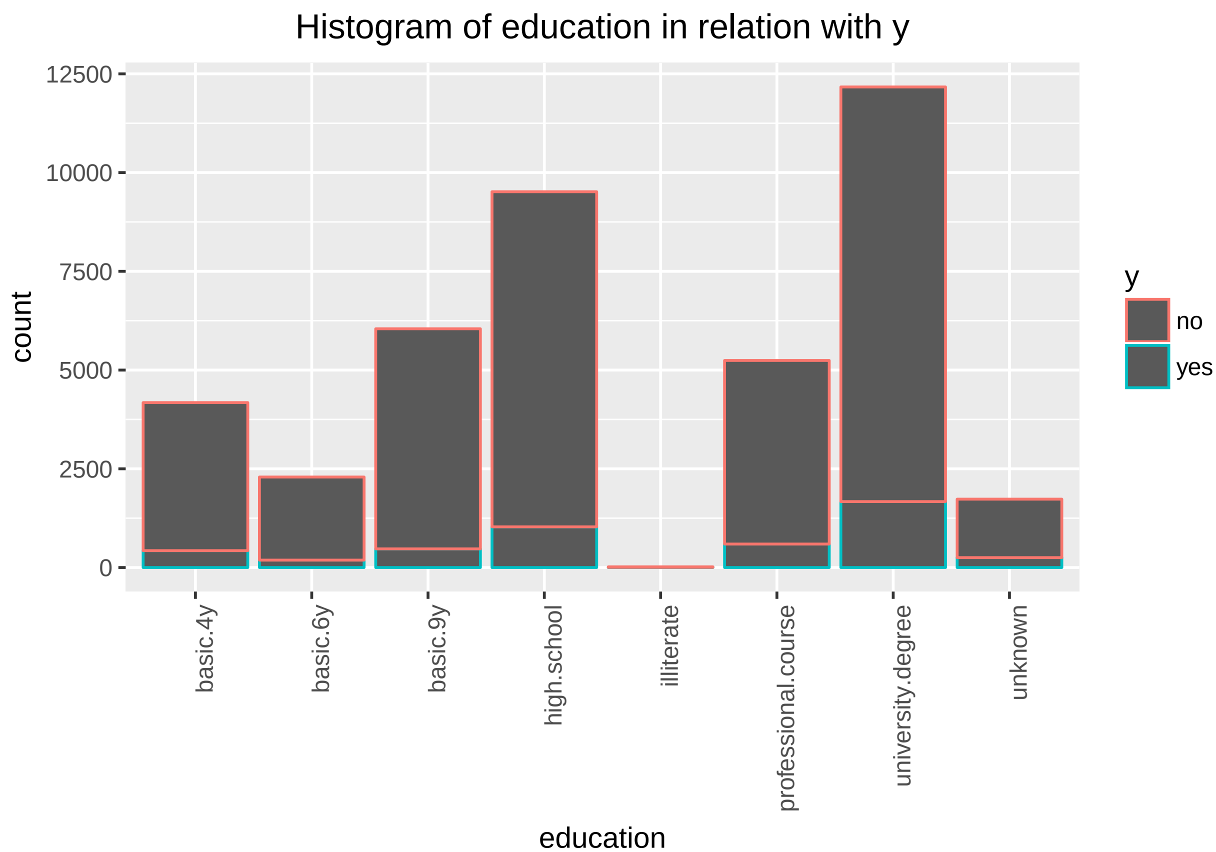
```
ggplot(bank, aes(x=education, col=y))+
    geom_histogram(stat="count", position = "fill")+
    theme(axis.text.x = element_text(angle = 90, hjust = 1))+
    ggtitle("Histogram of education in relation with y")+
    theme(plot.title = element_text(hjust = 0.5))
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```




```r
# But histogram of education shows that concentration of people who are illiterate is very
# small compared to other categories. So people having a university degree are more
# reasonable target.
ggplot(bank, aes(x=education, col=y))+
    geom_histogram(stat="count")+
    theme(axis.text.x = element_text(angle = 90, hjust = 1))+
    ggtitle("Histogram of education in relation with y")+
    theme(plot.title = element_text(hjust = 0.5))
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```

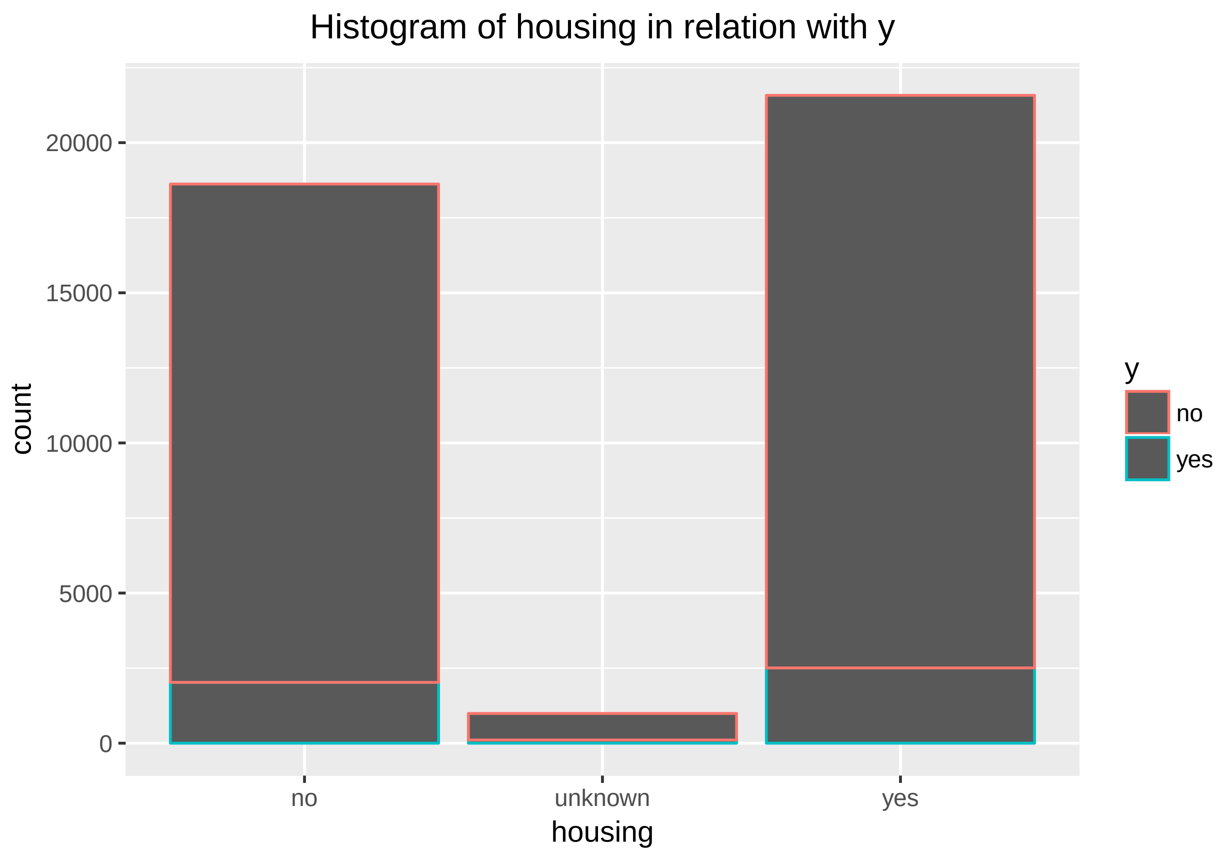



```
# The default variable is very less informative as it has only 3 values in the category of
# people who have defaulted and also large number of unkown values. So we can't get any
# understanding of its relation with y
table(bank$default)
```

```
##
##      no unknown     yes
##   32588    8597       3
```

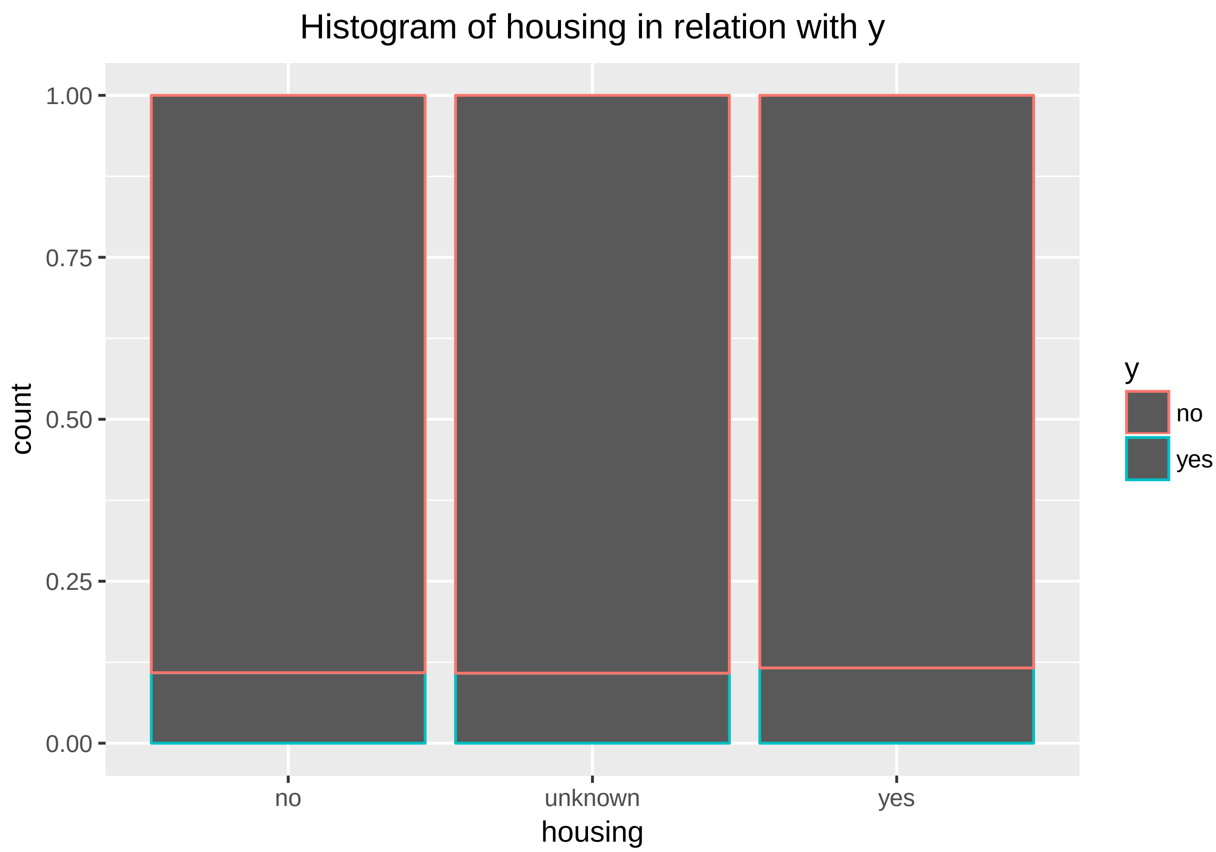
```
# The proportion plots show that the variable y is almost uniformly distributed in all the
# 3 categories of both housing and loan variables which indicates that these variables has
# very less correlation with variable y
ggplot(bank, aes(x=housing, col=y))+
    geom_histogram(stat="count")+
    ggtitle("Histogram of housing in relation with y")+
    theme(plot.title = element_text(hjust = 0.5))
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```



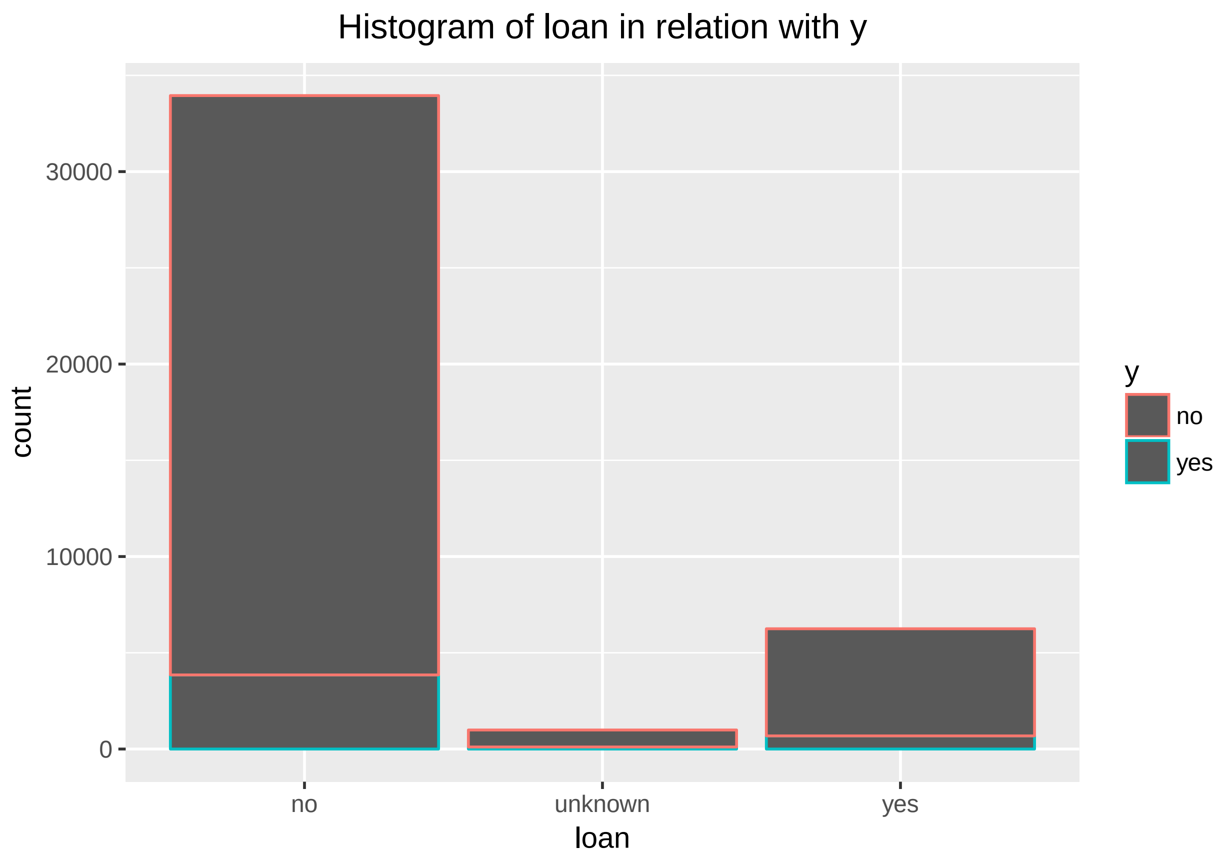

```r
ggplot(bank, aes(x=housing, col=y))+
    geom_histogram(stat="count", position = "fill")+
    ggtitle("Histogram of housing in relation with y")+
    theme(plot.title = element_text(hjust = 0.5))
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```



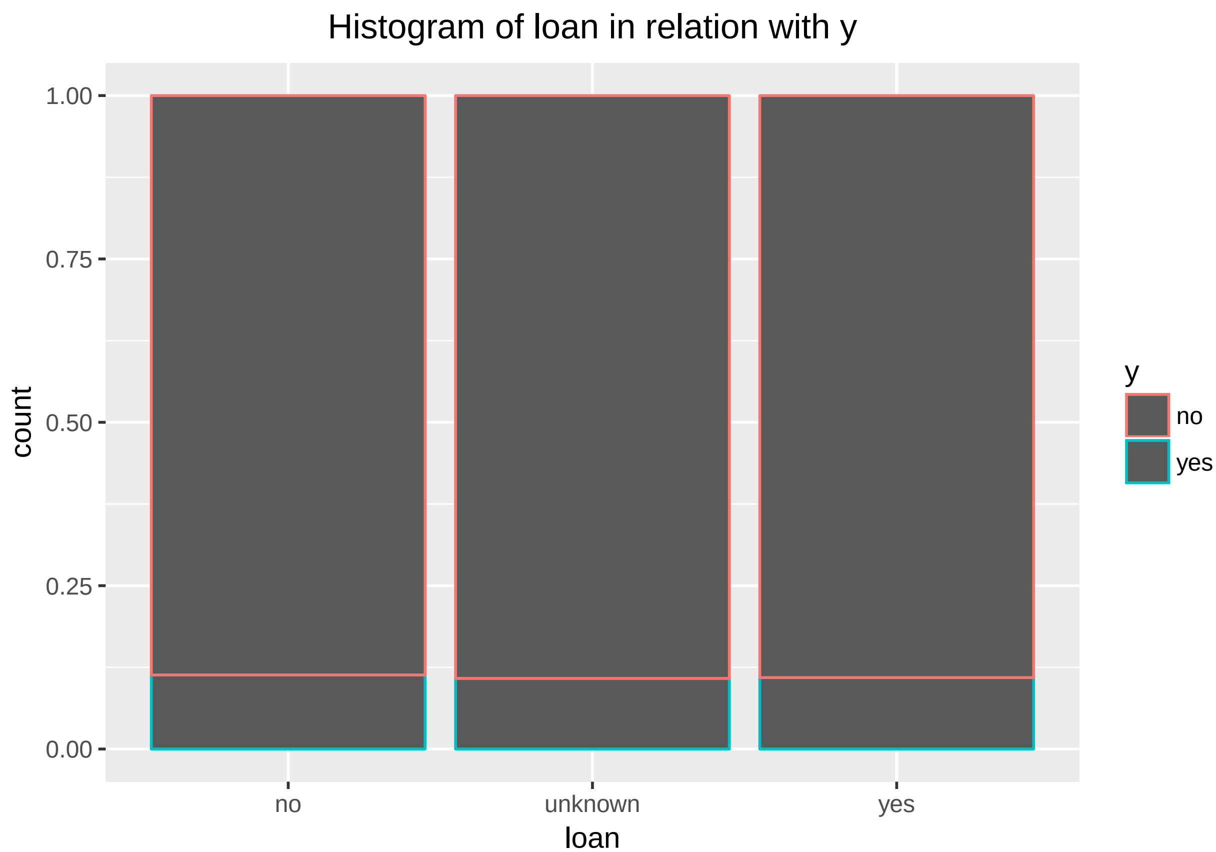

```
ggplot(bank, aes(x=loan, col=y))+
    geom_histogram(stat="count")+
    ggtitle("Histogram of loan in relation with y")+
    theme(plot.title = element_text(hjust = 0.5))
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```

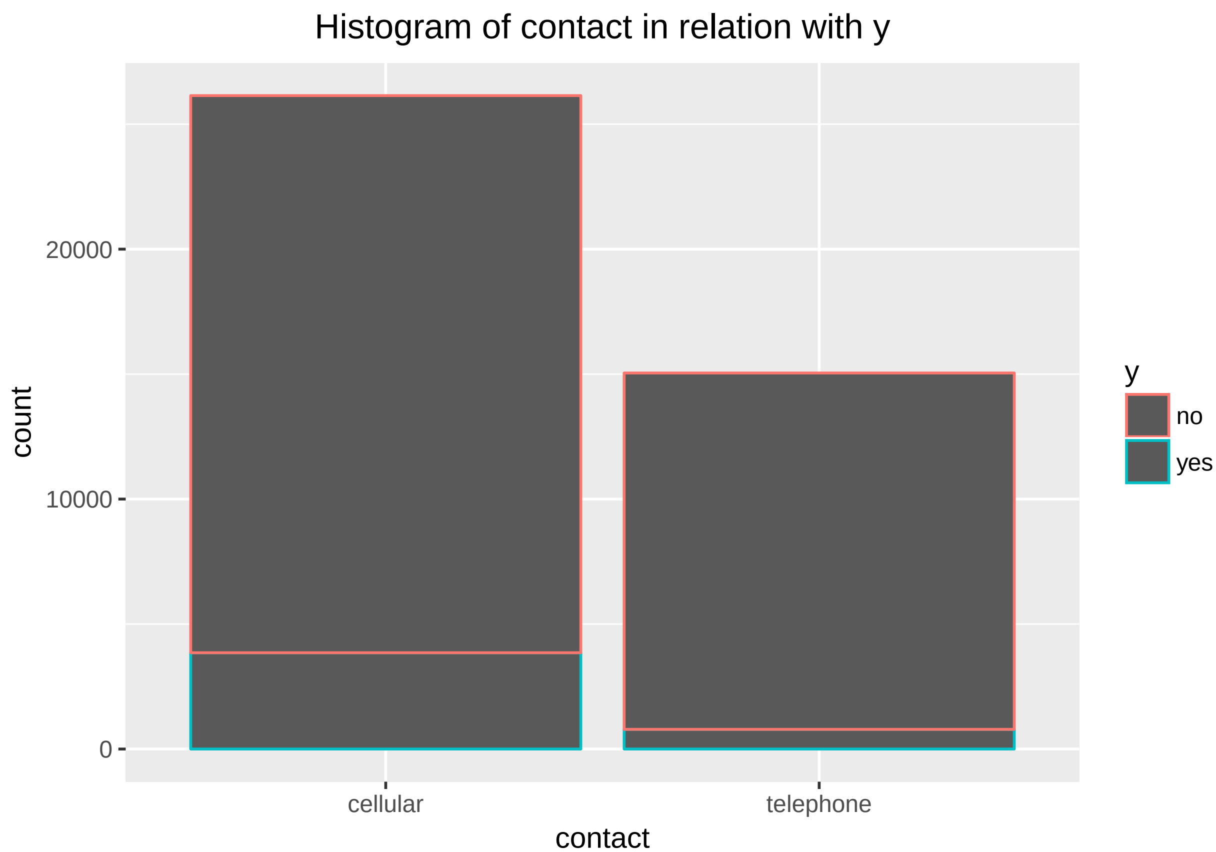



```r
ggplot(bank, aes(x=loan, col=y))+
    geom_histogram(stat="count", position = "fill")+
    ggtitle("Histogram of loan in relation with y")+
    theme(plot.title = element_text(hjust = 0.5))
```

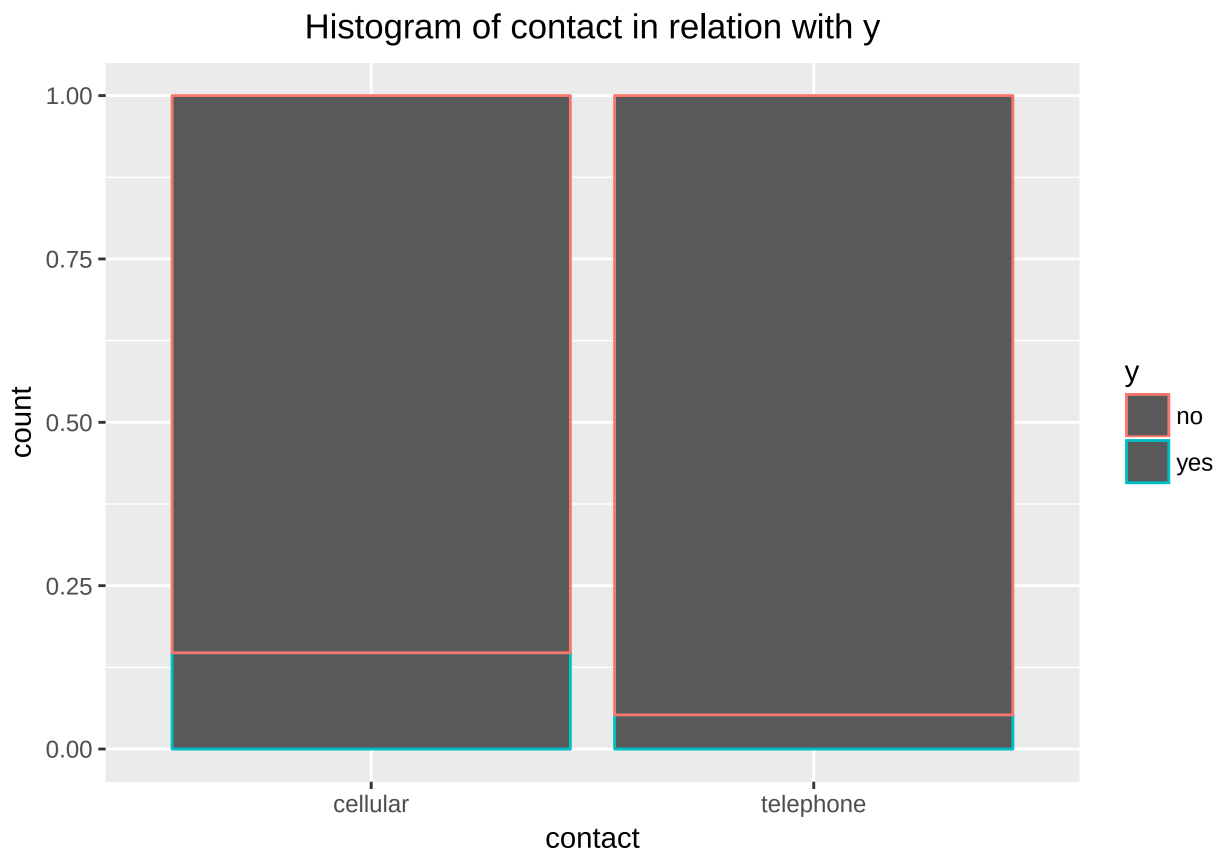
```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```




```
# The people who were contacted through cellular phone have slighlty high probability of
# taking the plan compared to the people who were contacted through a telephone
ggplot(bank, aes(x=contact, col=y))+
    geom_bar()+
    ggtitle("Histogram of contact in relation with y")+
    theme(plot.title = element_text(hjust = 0.5))
```

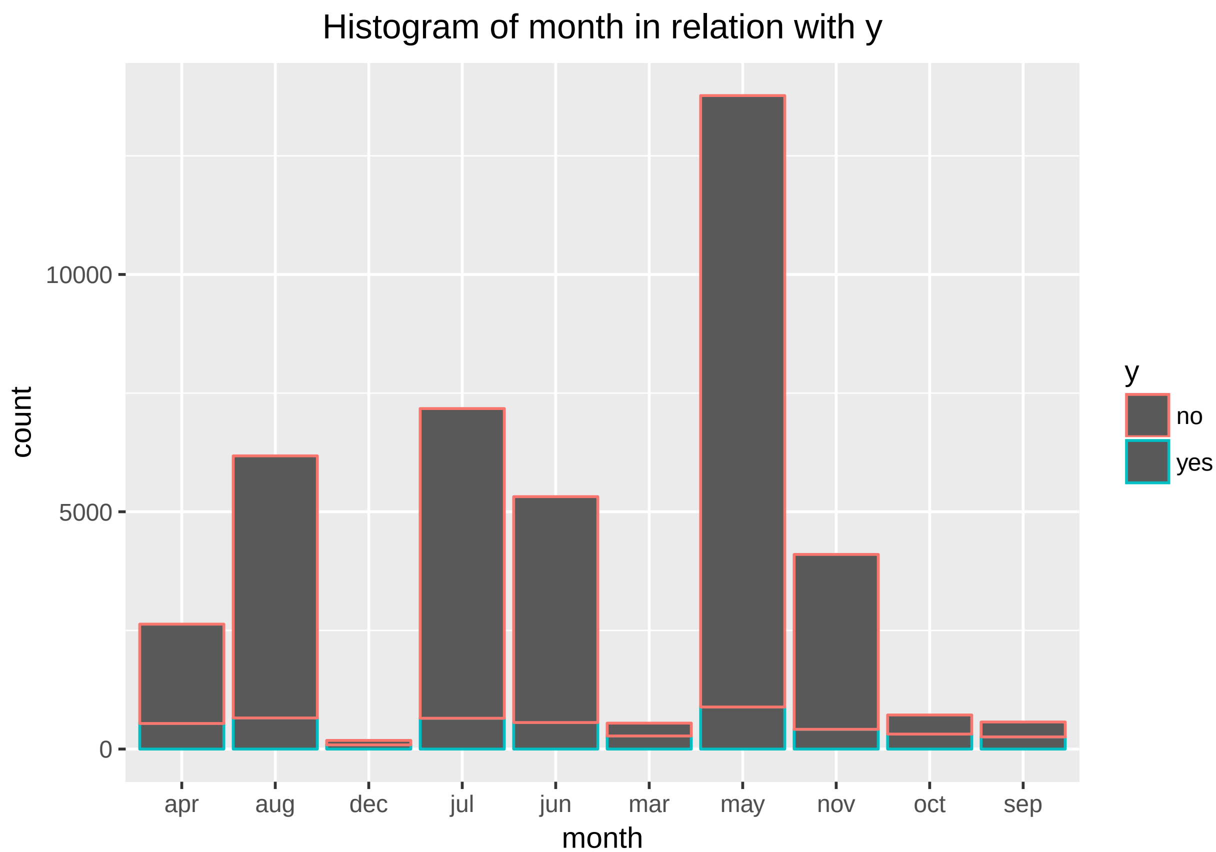



```r
ggplot(bank, aes(x=contact, col=y))+
    geom_bar(position = "fill")+
    ggtitle("Histogram of contact in relation with y")+
    theme(plot.title = element_text(hjust = 0.5))
```




```r
# The plot of proportion table and scatterplot of month and day of the week on which
# people were contacted shows that the months december, march, october and september
# have a very high probability of people taking the plan compared to other months.
# But the histogram of month in accordance with y shows that very less number of people
# were actually contacted in those months.
ggplot(bank, aes(x=month, col=y))+
    geom_histogram(stat="count")+
    ggtitle("Histogram of month in relation with y")+
    theme(plot.title = element_text(hjust = 0.5))
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```

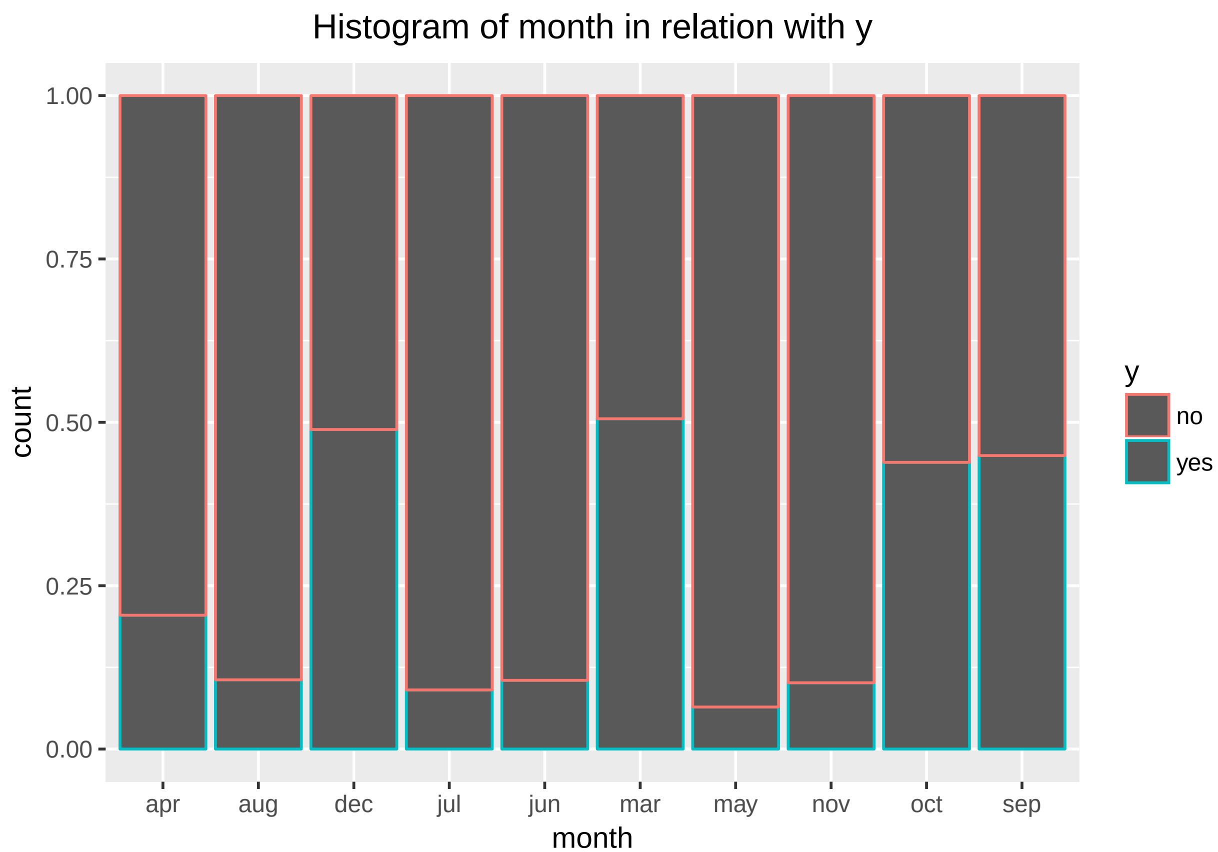



```
prop.table(table(bank$month, bank$y),1)
```

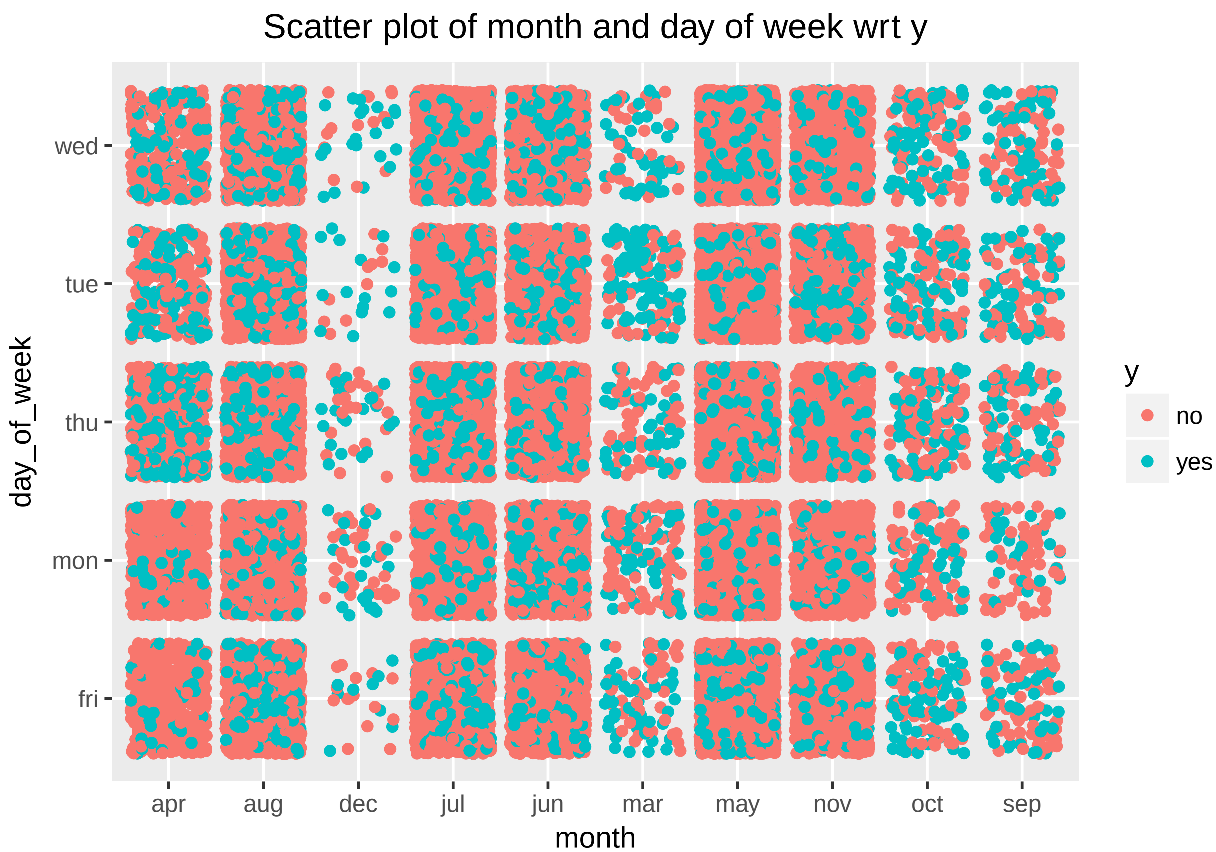
```
##
##              no        yes
##   apr 0.79521277 0.20478723
##   aug 0.89397863 0.10602137
##   dec 0.51098901 0.48901099
##   jul 0.90953443 0.09046557
##   jun 0.89488530 0.10511470
##   mar 0.49450549 0.50549451
##   may 0.93565255 0.06434745
##   nov 0.89856133 0.10143867
##   oct 0.56128134 0.43871866
##   sep 0.55087719 0.44912281
```

```
ggplot(bank, aes(x=month, col=y))+
    geom_histogram(stat="count", position = "fill")+
    ggtitle("Histogram of month in relation with y")+
    theme(plot.title = element_text(hjust = 0.5))
```

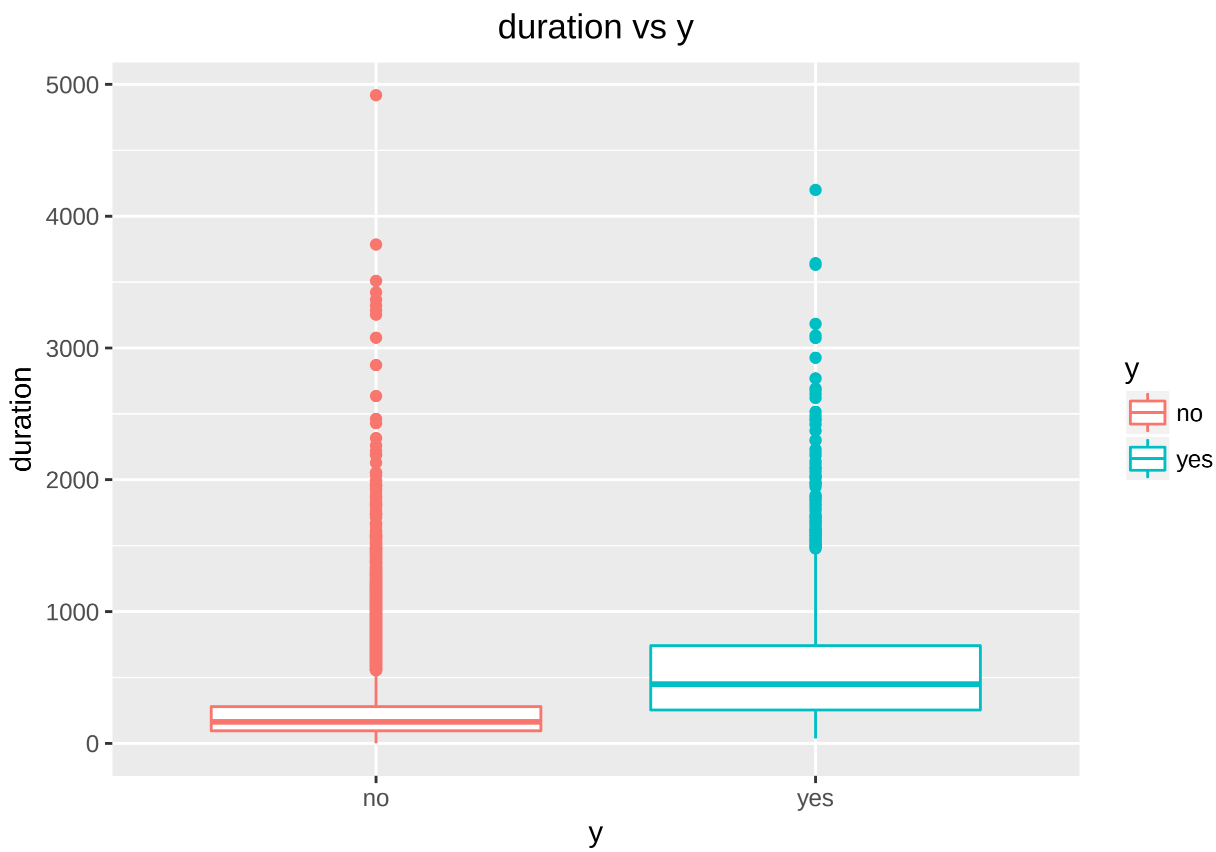
```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```



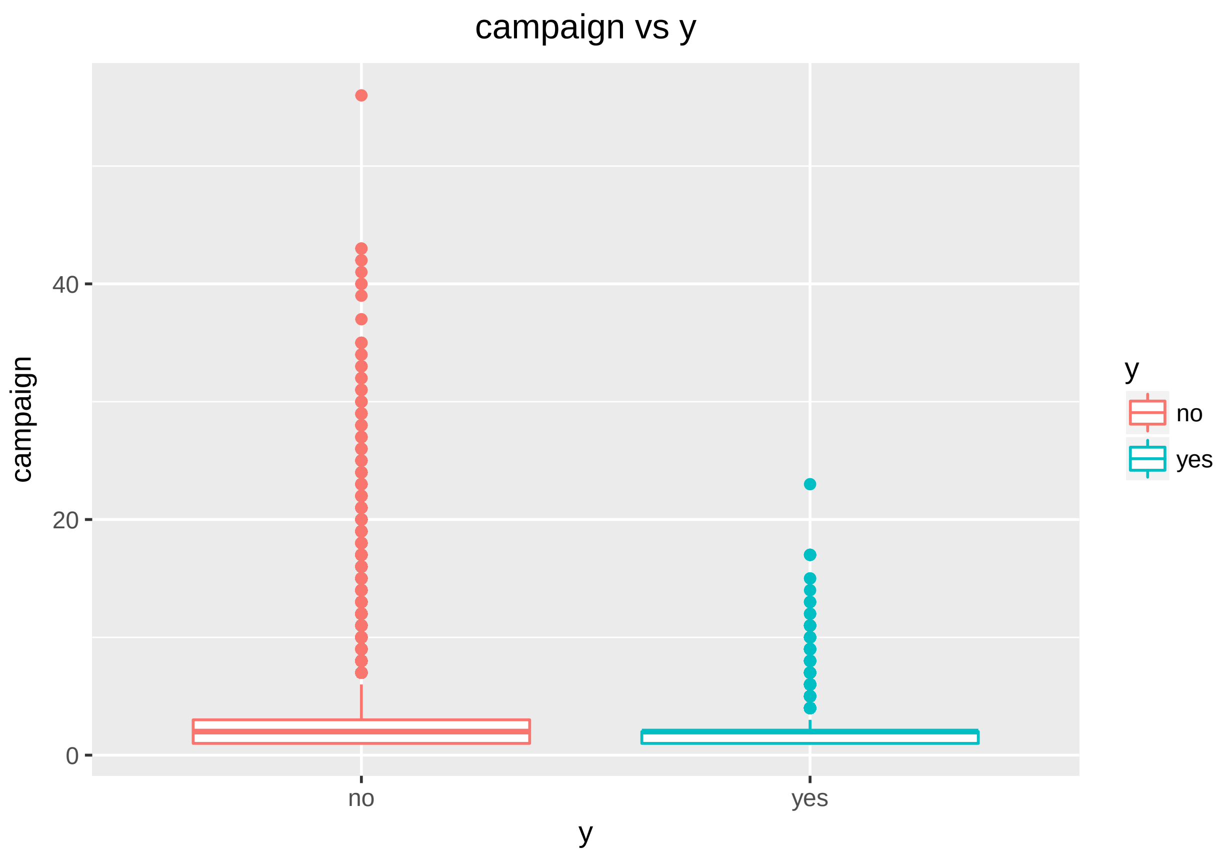

```r
ggplot(bank, aes(x=month, y=day_of_week, col=y)) +
    geom_jitter()+
    ggtitle("Scatter plot of month and day of week wrt y")+
    theme(plot.title = element_text(hjust = 0.5))
```




```
# The boxplot of call duration and whether the customer took the subscription clearly
# indicates that people who took subscription spend significanlty more amount of time
# speaking to the representative. The variable duration can be highly useful while
# predicting y. We can acutally predict that y="no" whenever call duration is 0 which
# means that if a customer didnt attend the call then he/she didnt subscribe. And the
# variable duration should be left out while building a realistic predictive model, as we
# only get to know call duration after speaking to the customer, but we would anyway
# know if a person subscribed or not (y="yes" or "no") by the end of the call.So this
# variable while building a true predictive model which will be used by the organization
# in future
ggplot(bank, aes(x=y, y=duration, col=y))+
    geom_boxplot()+
    ggtitle("duration vs y")+
    theme(plot.title = element_text(hjust = 0.5))
```




```r
# The proportion table of campaign(number of times client was contacted during this
# campaign) shows that it is very unlikely that client will say "yes" after contacting
# the client more than 15 times with a probability of 1.4%
ggplot(bank, aes(x=y, y=campaign, col=y))+
    geom_boxplot()+
    ggtitle("campaign vs y")+
    theme(plot.title = element_text(hjust = 0.5))
```




```r
table(bank$campaign, bank$y)
```

```
##
##          no   yes
##    1  15342  2300
##    2   9359  1211
##    3   4767   574
##    4   2402   249
##    5   1479   120
##    6    904    75
##    7    591    38
##    8    383    17
##    9    266    17
##    10   213    12
##    11   165    12
##    12   122     3
##    13    88     4
##    14    68     1
##    15    49     2
##    16    51     0
##    17    54     4
##    18    33     0
##    19    26     0
##    20    30     0
##    21    24     0
##    22    17     0
##    23    15     1
##    24    15     0
##    25     8     0
##    26     8     0
```

```
##   27    11     0
##   28     8     0
##   29    10     0
##   30     7     0
##   31     7     0
##   32     4     0
##   33     4     0
##   34     3     0
##   35     5     0
##   37     1     0
##   39     1     0
##   40     2     0
##   41     1     0
##   42     2     0
##   43     2     0
##   56     1     0
```

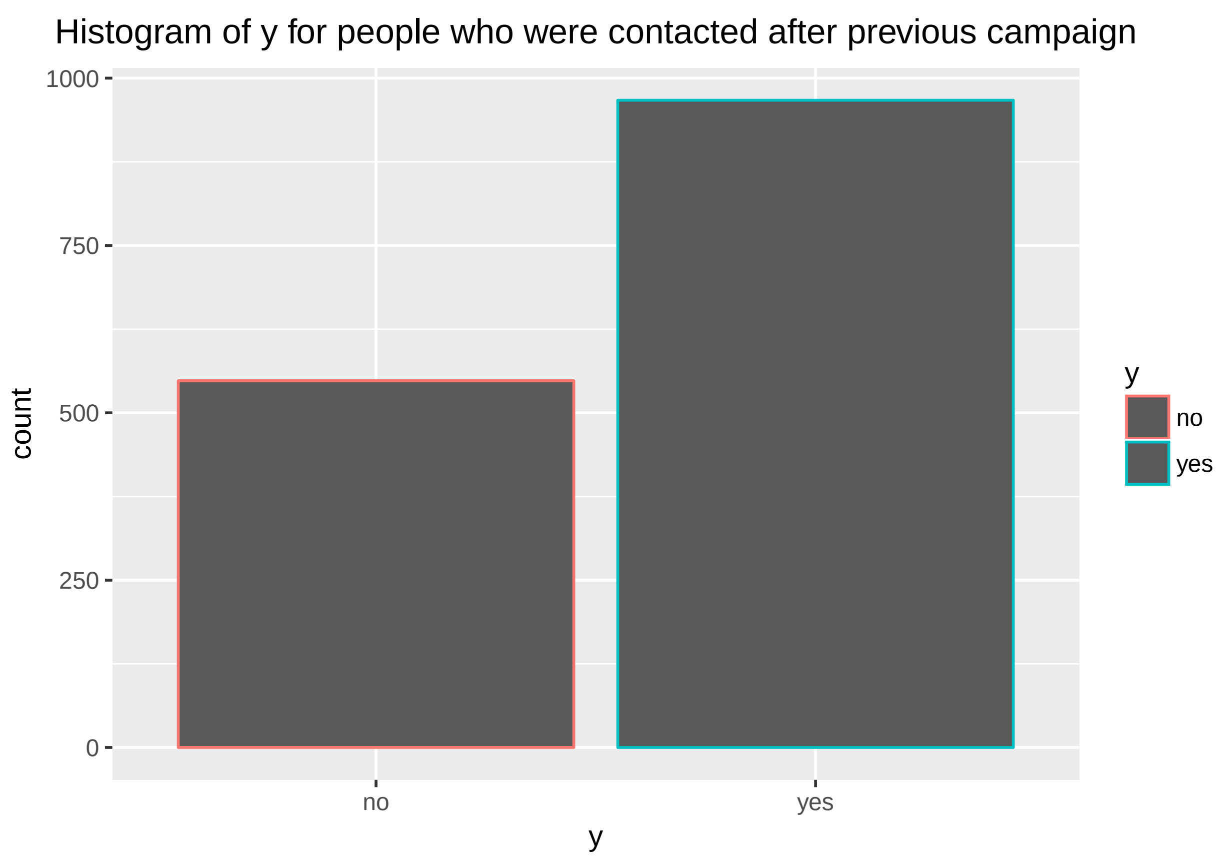
```r
prop.table(table(bank[bank$campaign>15, ]$y))
```

```
##
##         no        yes
## 0.98591549 0.01408451
```

```r
prop.table(table(bank[bank$campaign<15, ]$y))
```

```
##
##       no      yes
## 0.886396 0.113604
```

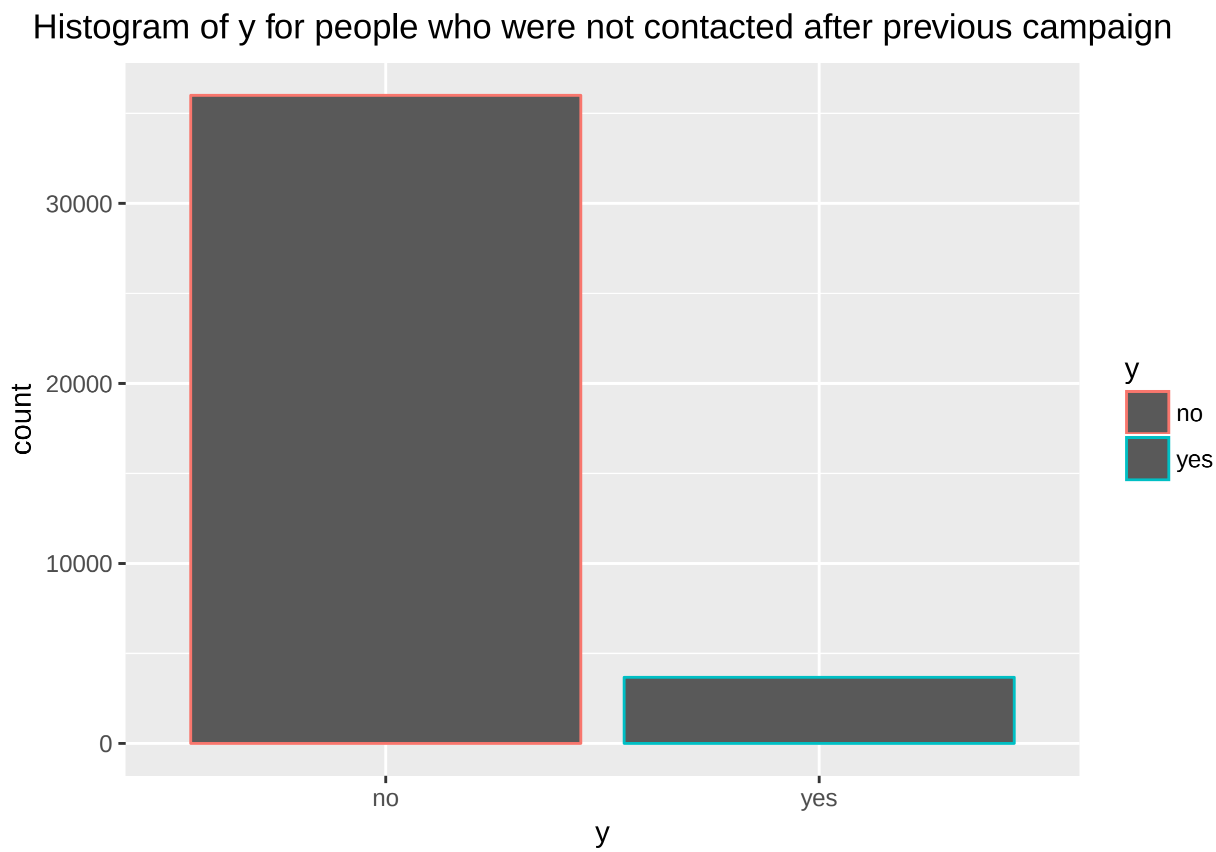
```r
# The proportion table of pdays (number of days that passed by after the client was last
# contacted from a previous campaign. If a client was not contacted previously pdays will
# be 999) of people who were contacted previously and people who weren't shows that
# previously contacted clients have 63.8% chance of taking the subscription while people
# who weren't contacted previously only have 9% chance of taking the plan.
ggplot(bank[bank$pdays!=999,], aes(x=y, col=y))+
    geom_bar()+
    ggtitle("Histogram of y for people who were contacted after previous campaign")+
    theme(plot.title = element_text(hjust = 0.5))
```




```
prop.table(table(bank[bank$pdays!=999,]$y))
```

```
##
##        no       yes
## 0.3617162 0.6382838
```

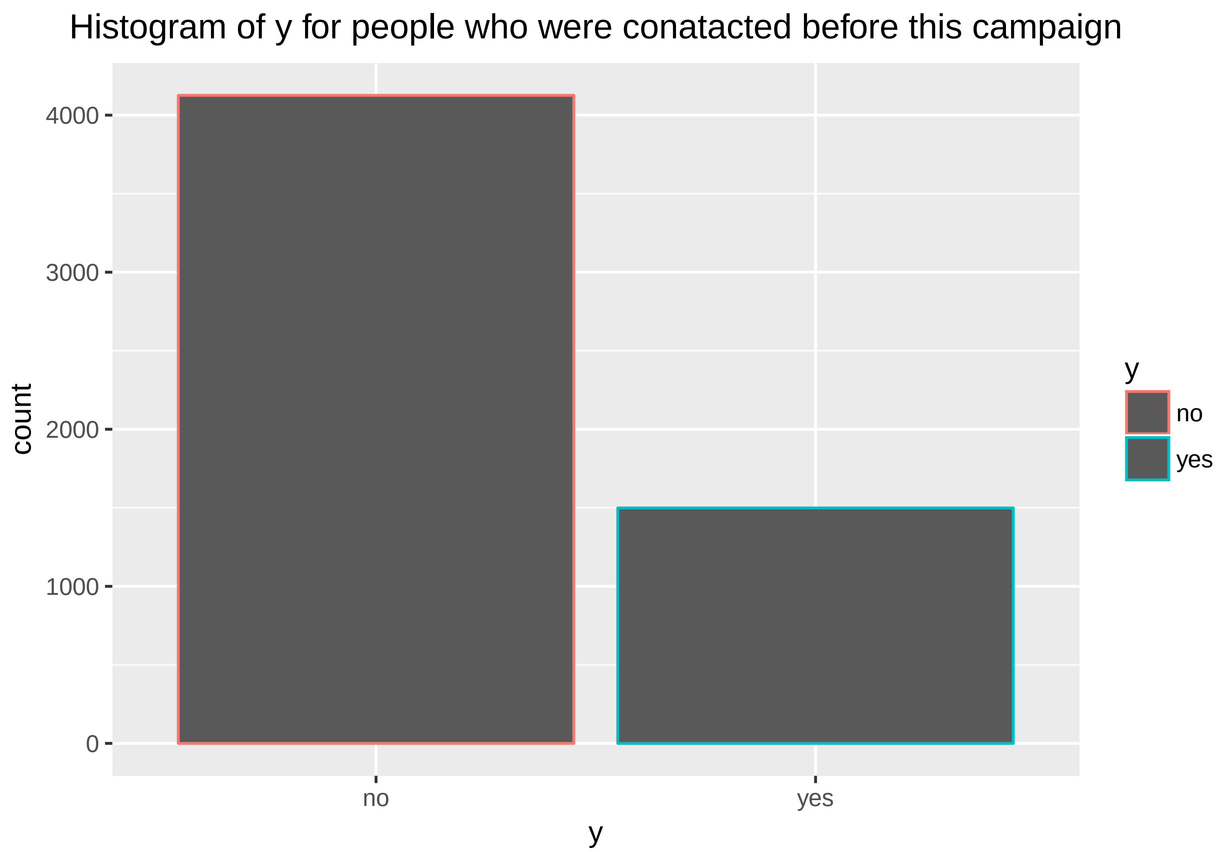
```
ggplot(bank[bank$pdays==999,], aes(x=y, col=y))+
    geom_bar()+
    ggtitle("Histogram of y for people who were not contacted after previous campaign")+
    theme(plot.title = element_text(hjust = 0.5))
```




```
prop.table(table(bank[bank$pdays==999,]$y))
```

```
##
##         no        yes
## 0.90741814 0.09258186
```

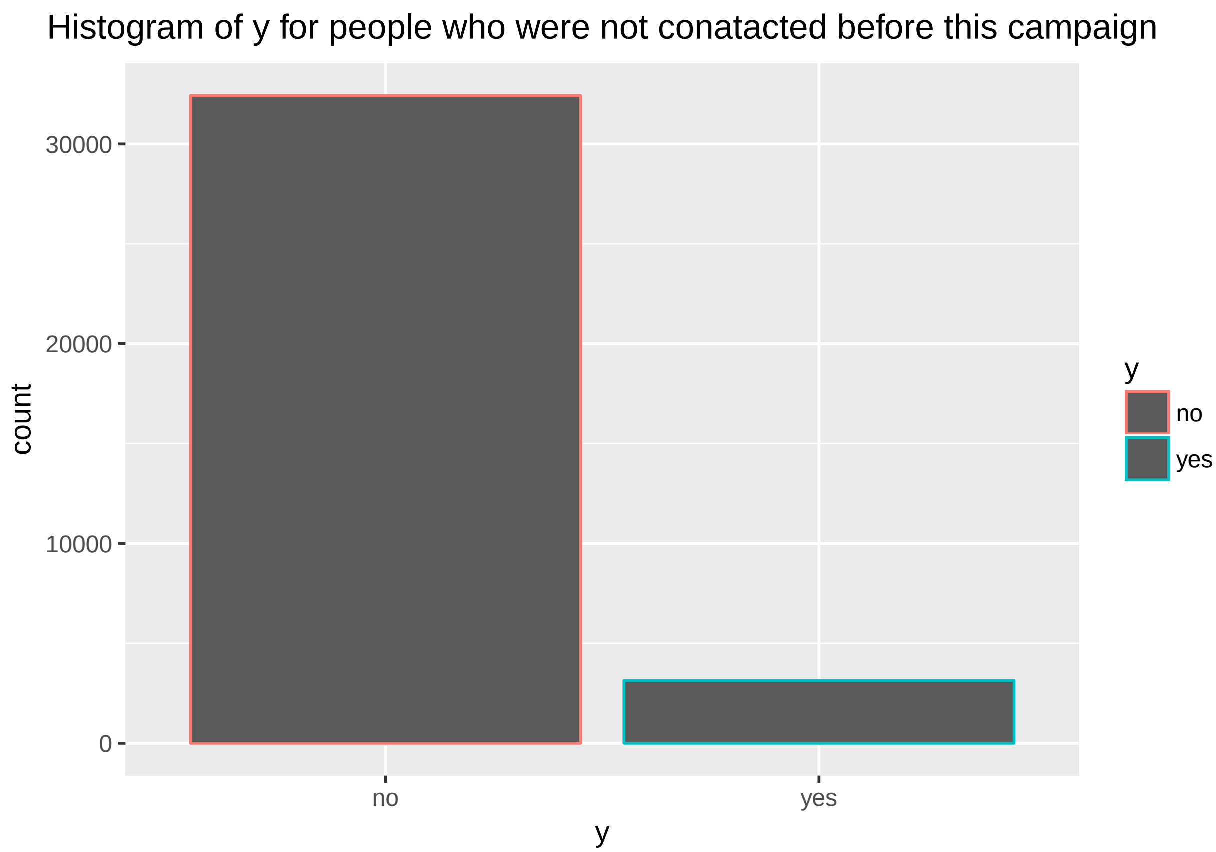
```
# The previous variable (number of contacts performed before this campaign and for this
# client) is very similar to campaign variable, and indicates that people who were
# contacted previously have higher chance of taking the subscription compared to the
# people who were not contacted before this campaign. By examining the relationship
# between pdays and previous using subsets of population, it is evident that, of those
# people who were contacted at least once during this campaign were all contacted at least
# once before this campaign. And complimentarily, the people who were not contacted
# even once before this campaign are also not contacted even once in this campaign. This
# indicates that the variable previous is highly dependent on the variable pdays and
# doesn't play significant role in giving extra information
ggplot(bank[bank$previous!=0,], aes(x=y, col=y))+
    geom_bar()+
    ggtitle("Histogram of y for people who were conatacted before this campaign")+
    theme(plot.title = element_text(hjust = 0.5))
```




```
prop.table(table(bank[bank$previous!=0,]$y))
```

```
##
##        no       yes
## 0.7335111 0.2664889
```

```
ggplot(bank[bank$previous==0,], aes(x=y, col=y))+
    geom_bar()+
    ggtitle("Histogram of y for people who were not conatacted before this campaign")+
    theme(plot.title = element_text(hjust = 0.5))
```




```r
prop.table(table(bank[bank$previous==0,]$y))
```

```
##
##         no        yes
## 0.91167787 0.08832213
```

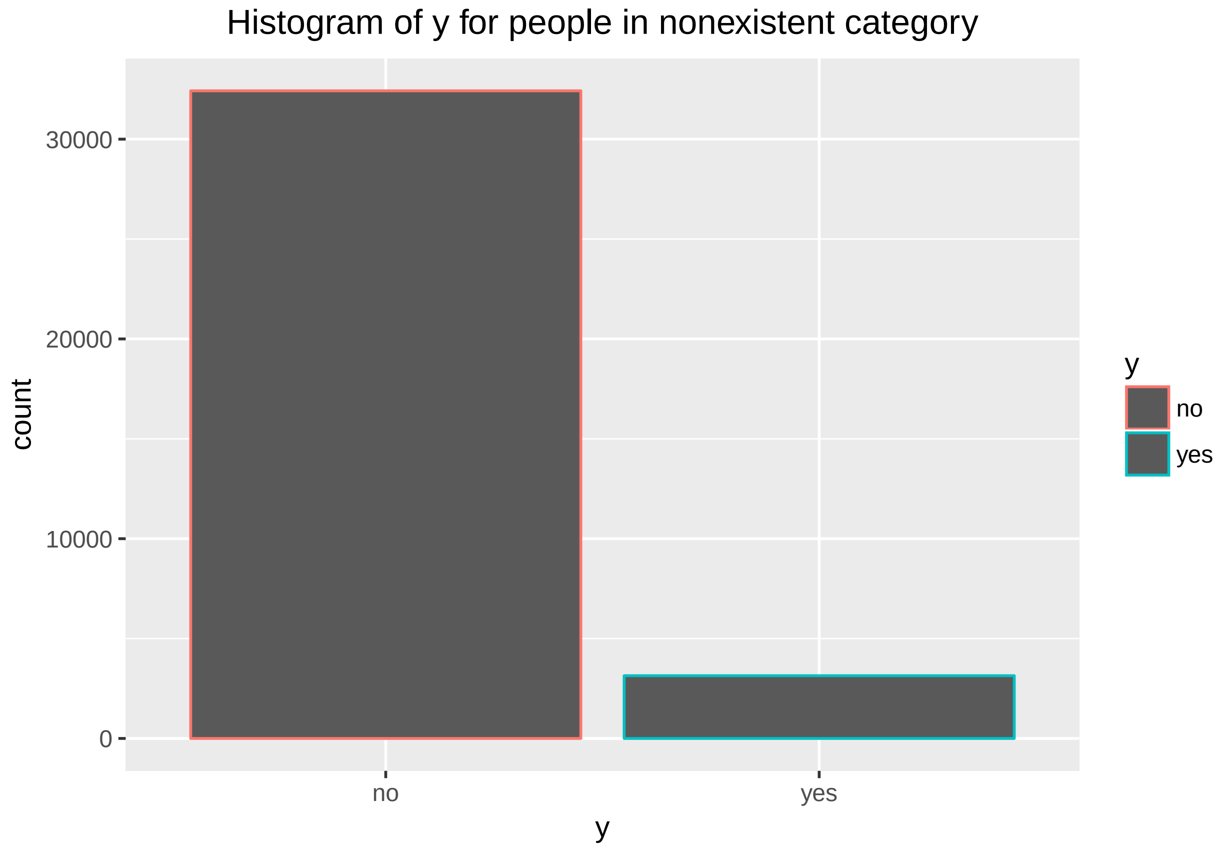
```r
table(bank[bank$pdays!=999,]$previous)
```

```
##
##   1   2   3   4   5   6   7
## 865 405 166  58  16   4   1
```

```r
table(bank[bank$previous==0,]$pdays)
```

```
##
##   999
## 35563
```

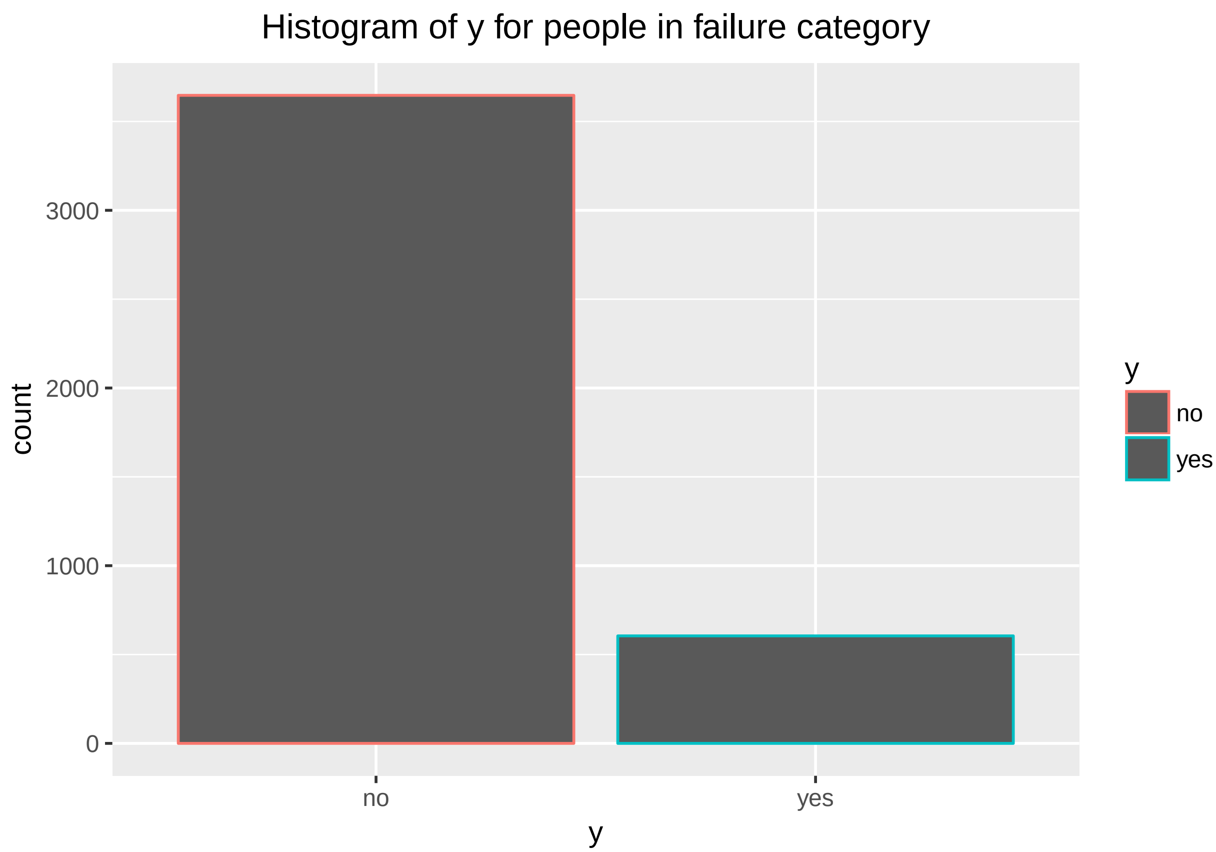
```r
# Bar plots of each category of the variable poutcome(outcome of the previous marketing
# campaign) indicates that people whose outcome of the previous campaign is a success
# have very high probability of taking the plan followed by people who rejected the offer
# in previous campaign. The people who were not contacted previously at all have very low
# chance of taking the plan.
ggplot(bank[bank$poutcome=="nonexistent",], aes(x=y, col=y))+
    geom_bar()+
    ggtitle("Histogram of y for people in nonexistent category")+
    theme(plot.title = element_text(hjust = 0.5))
```




```
prop.table(table(bank[bank$poutcome=="nonexistent",]$y))
```

```
##
##         no        yes
## 0.91167787 0.08832213
```

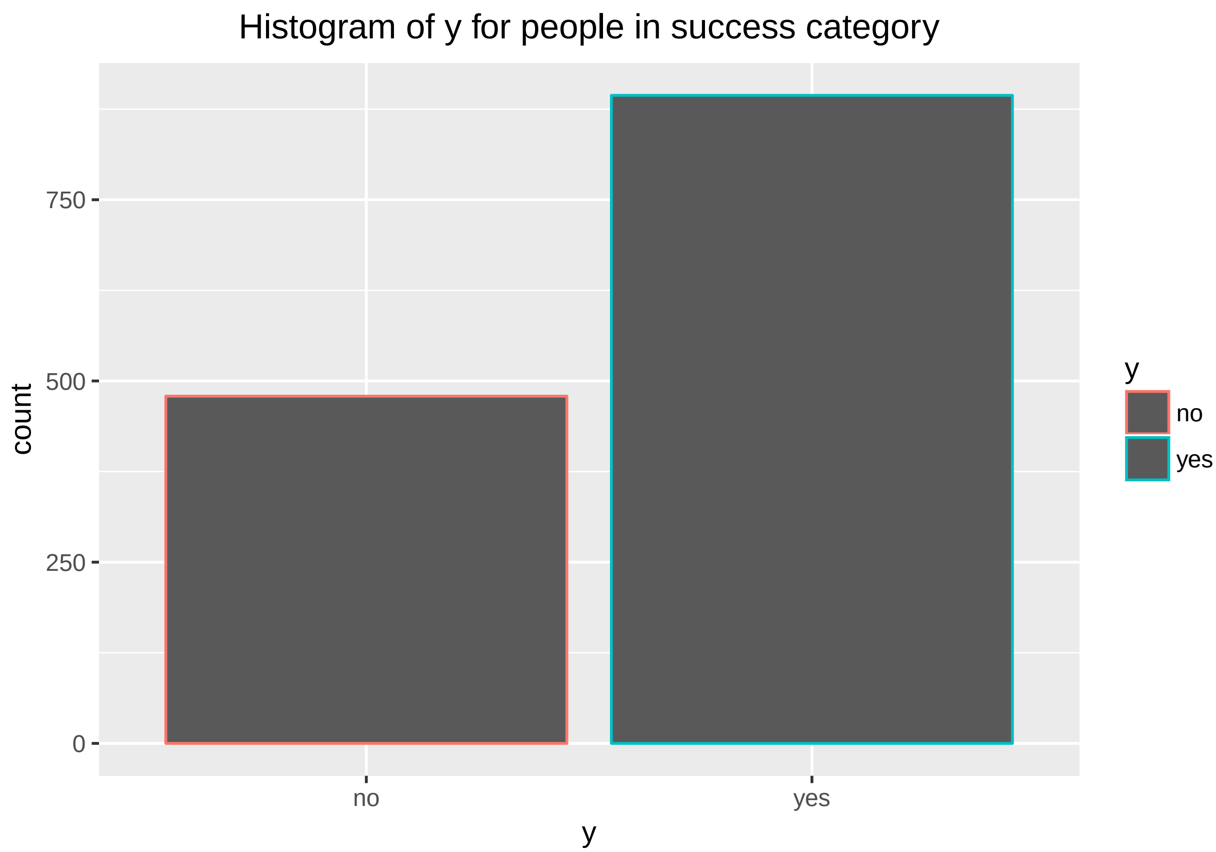
```
ggplot(bank[bank$poutcome=="failure",], aes(x=y, col=y))+
    geom_bar()+
    ggtitle("Histogram of y for people in failure category")+
    theme(plot.title = element_text(hjust = 0.5))
```




```
prop.table(table(bank[bank$poutcome=="failure",]$y))
```

```
##
##       no      yes
## 0.857714 0.142286
```

```
ggplot(bank[bank$poutcome=="success",], aes(x=y, col=y))+
    geom_bar()+
    ggtitle("Histogram of y for people in success category")+
    theme(plot.title = element_text(hjust = 0.5))
```




```
prop.table(table(bank[bank$poutcome=="success",]$y))
```

```
##
##        no       yes
## 0.3488711 0.6511289
```

```
# It is interesting to note that people who were non-existent in poutcome are the
# same people with pdays=999
table(bank[bank$poutcome=="nonexistent",]$y)
```

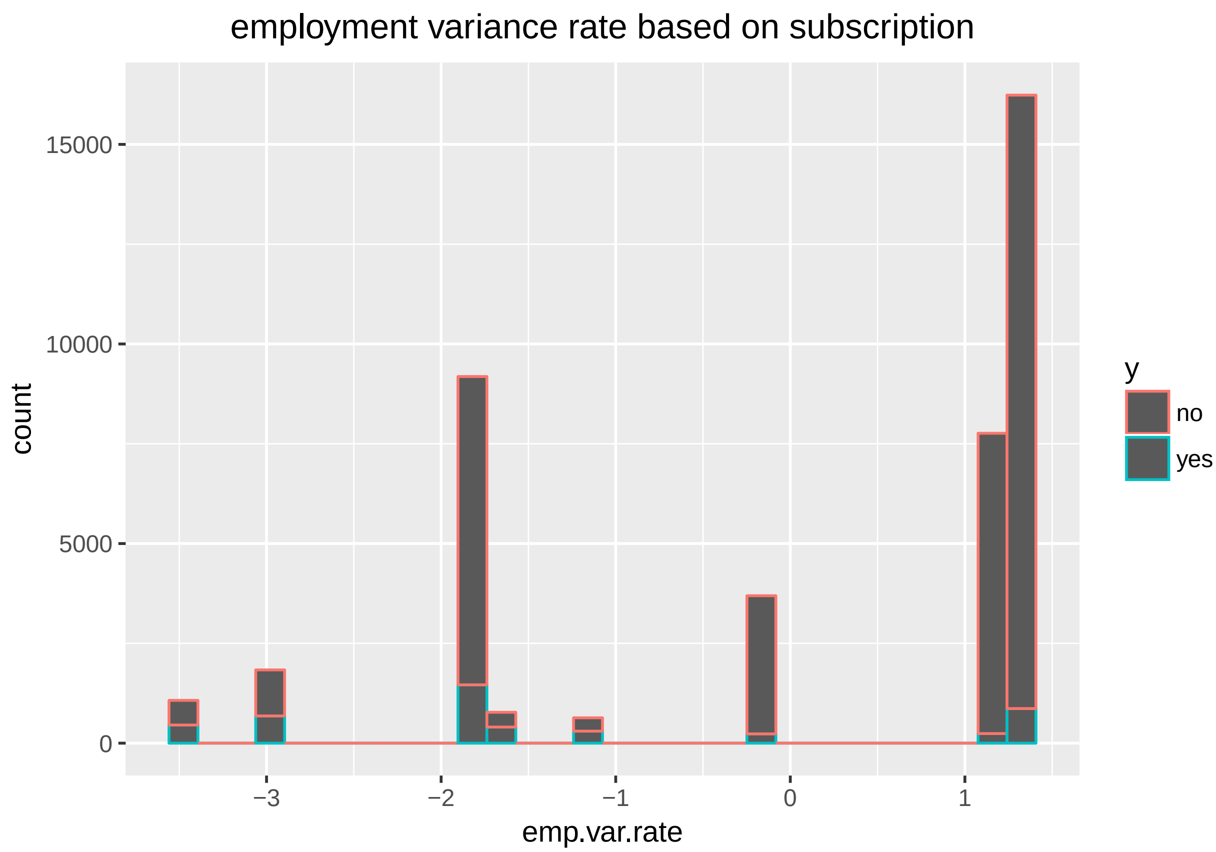
```
##
##    no   yes
## 32422  3141
```

```
table(bank[bank$poutcome=="nonexistent"&bank$pdays==999,]$y)
```

```
##
##    no   yes
## 32422  3141
```

```
# The probability taking the subscription significantly increases when employment
# variance rate is less than -1.
ggplot(data=bank, aes(x=emp.var.rate, col=y))+
    geom_histogram()+
    ggtitle("employment variance rate based on subscription")+
    theme(plot.title = element_text(hjust = 0.5))
```

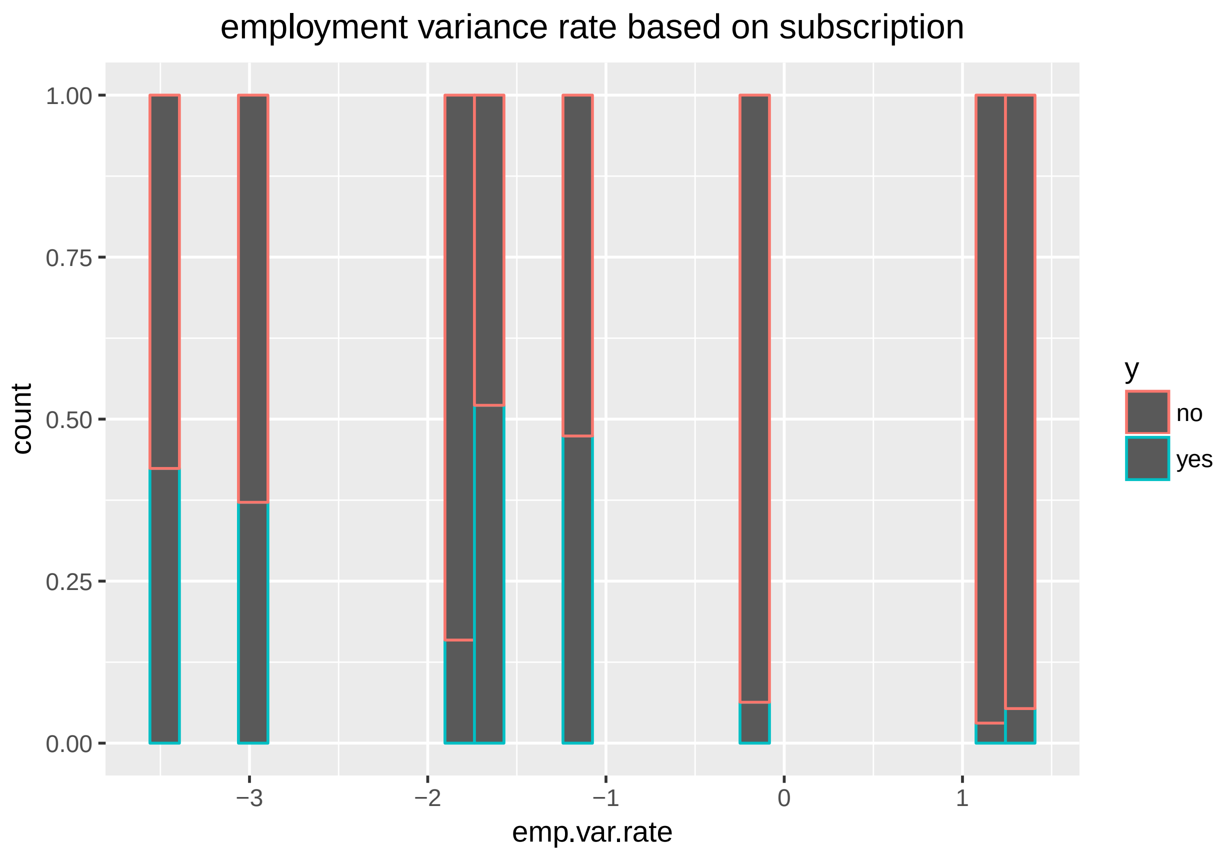
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
ggplot(data=bank, aes(x=emp.var.rate, col=y))+
    geom_histogram(position = "fill")+
    ggtitle("employment variance rate based on subscription")+
    theme(plot.title = element_text(hjust = 0.5))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

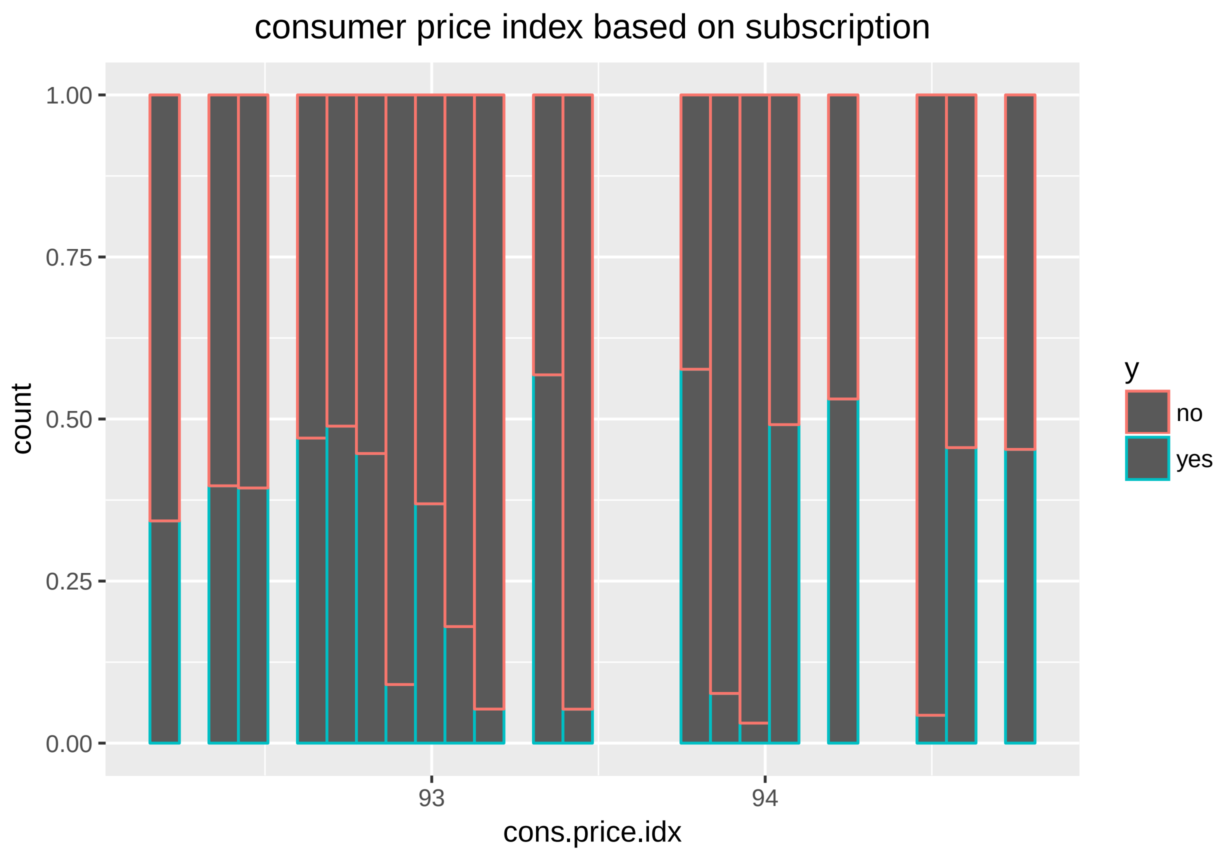
```
## Warning: Removed 44 rows containing missing values (geom_bar).
```



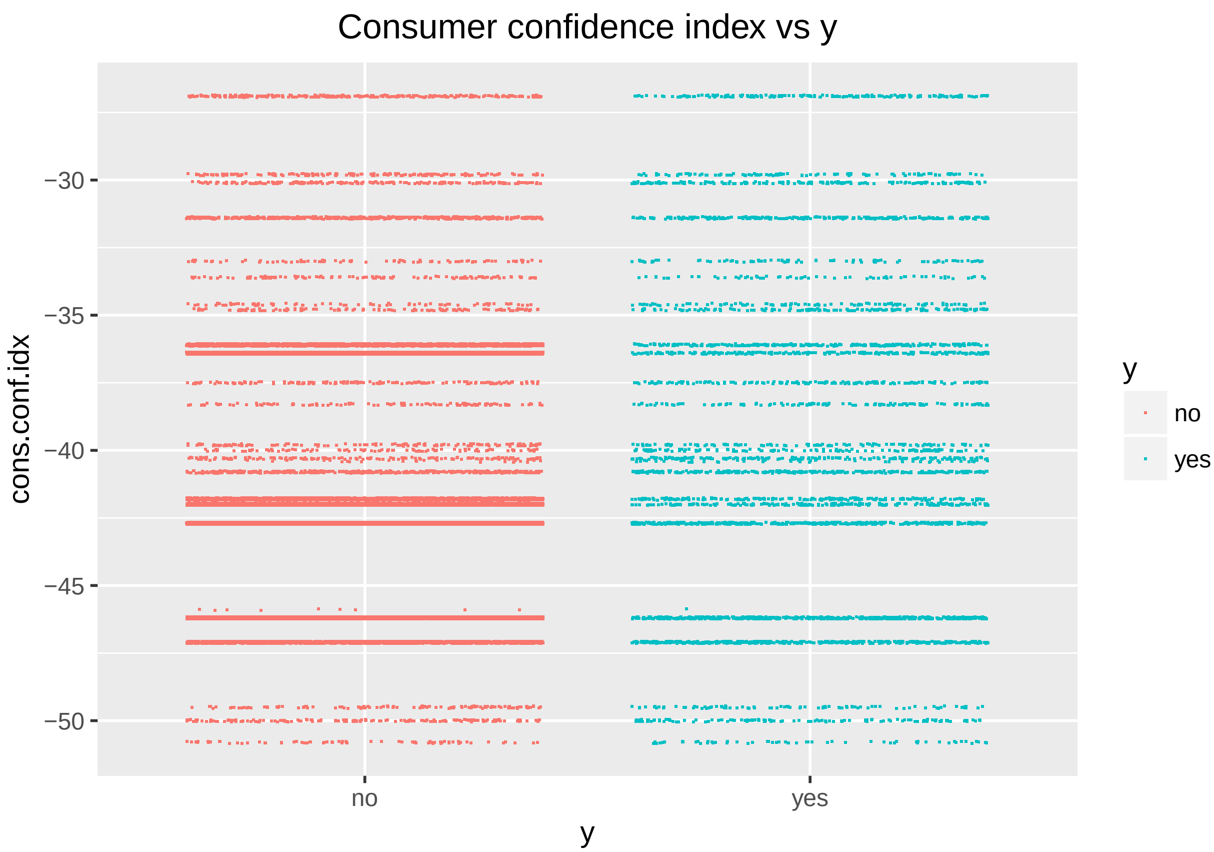

```
# A histogram of proportions of the variable consumer price index shows that the
# concentration of people taking the plan is evenly spread on both higher and lower
# sides of cpi, so we can't really make any generalizations about a particular group
# being more favourable for saying "yes"
ggplot(data=bank, aes(x=cons.price.idx, col=y))+
    geom_histogram(position = "fill")+
    ggtitle("consumer price index based on subscription")+
    theme(plot.title = element_text(hjust = 0.5))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 20 rows containing missing values (geom_bar).
```



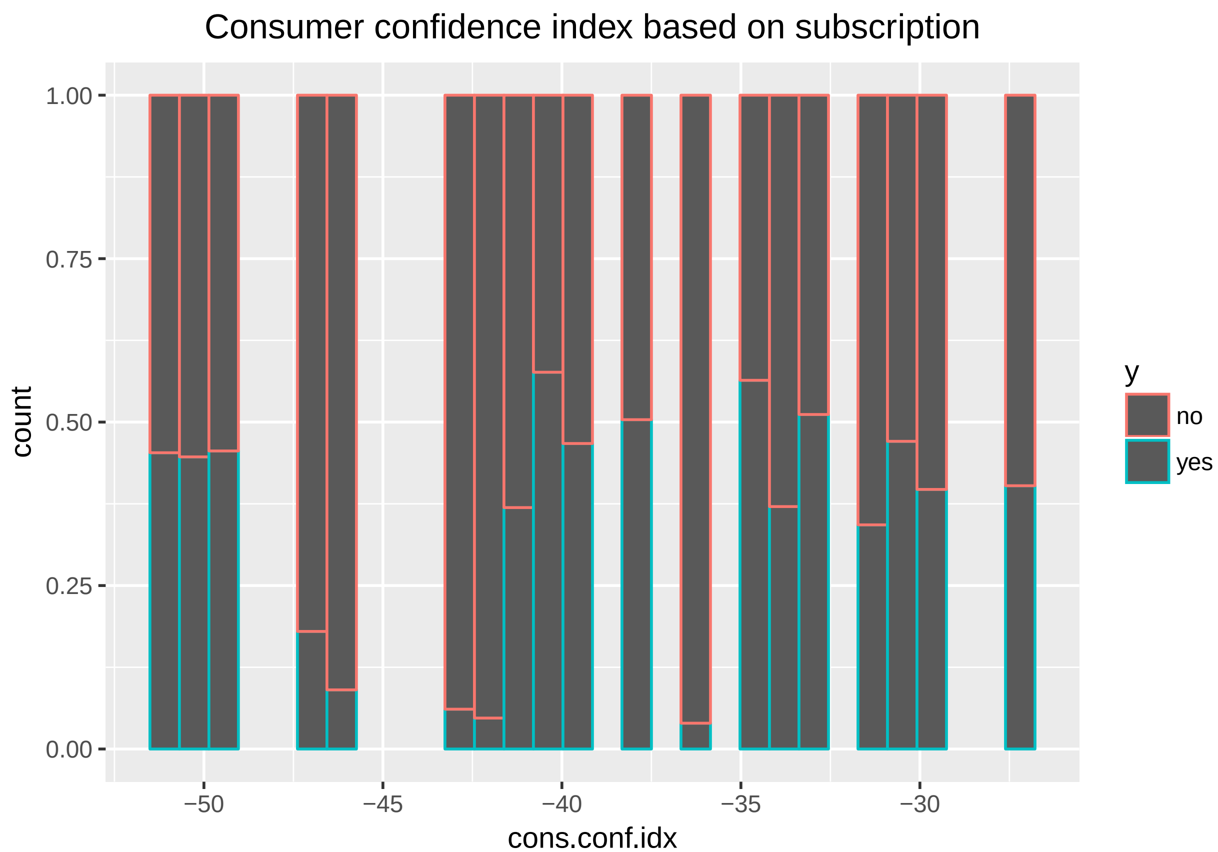

```
# The scatterplot and histogram of consumer confidence index indicates that the amount of
# people who are taking the subscription mostly depends on the amount of people present in
# that particular range of values and not actually on the consumer confidence index itself
ggplot(data=bank, aes(x=y, y=cons.conf.idx, col=y))+
    geom_jitter(shape=46)+
    ggtitle("Consumer confidence index vs y")+
    theme(plot.title = element_text(hjust = 0.5))
```




```
ggplot(data=bank, aes(x=cons.conf.idx, col=y))+
    geom_histogram(position = "fill")+
    ggtitle("Consumer confidence index based on subscription")+
    theme(plot.title = element_text(hjust = 0.5))
```

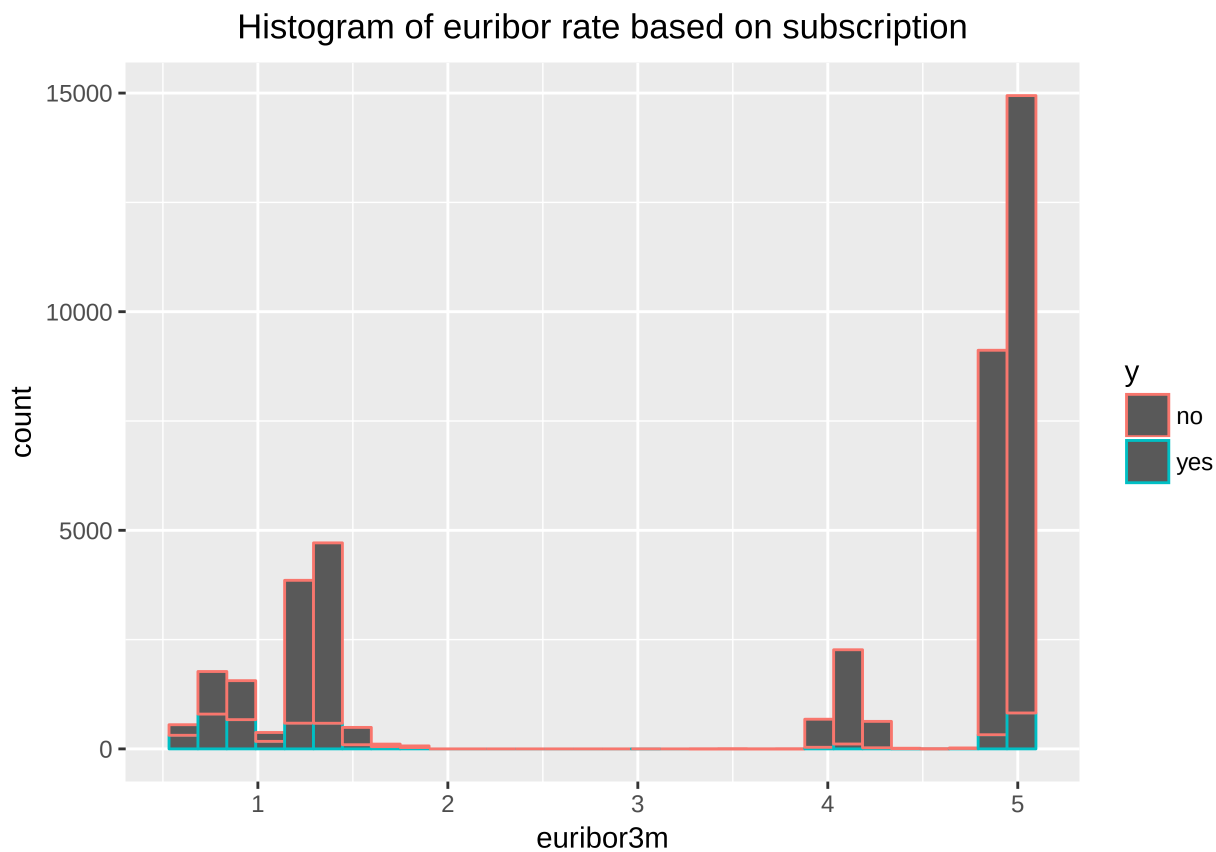
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 22 rows containing missing values (geom_bar).
```



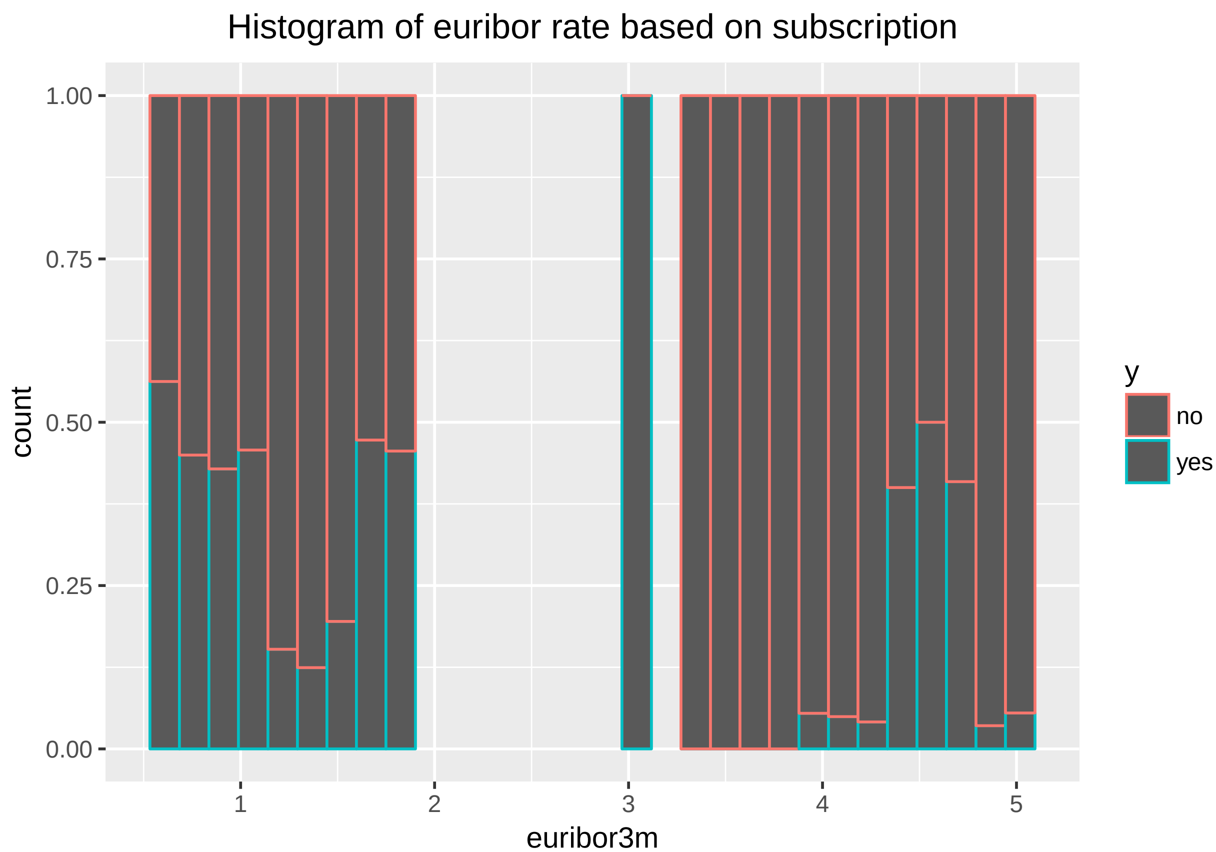

```r
# The histograms and proportion tables of euribor(Euro Interbank Offered Rate) shows that,
# as euribor decreases the probability of people taking the plan increases significantly.
# And when euribor drops below 1, there is very high chance of people taking the
# subscription.
ggplot(data=bank, aes(x=euribor3m, col=y))+
    geom_histogram()+
    ggtitle("Histogram of euribor rate based on subscription")+
    theme(plot.title = element_text(hjust = 0.5))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

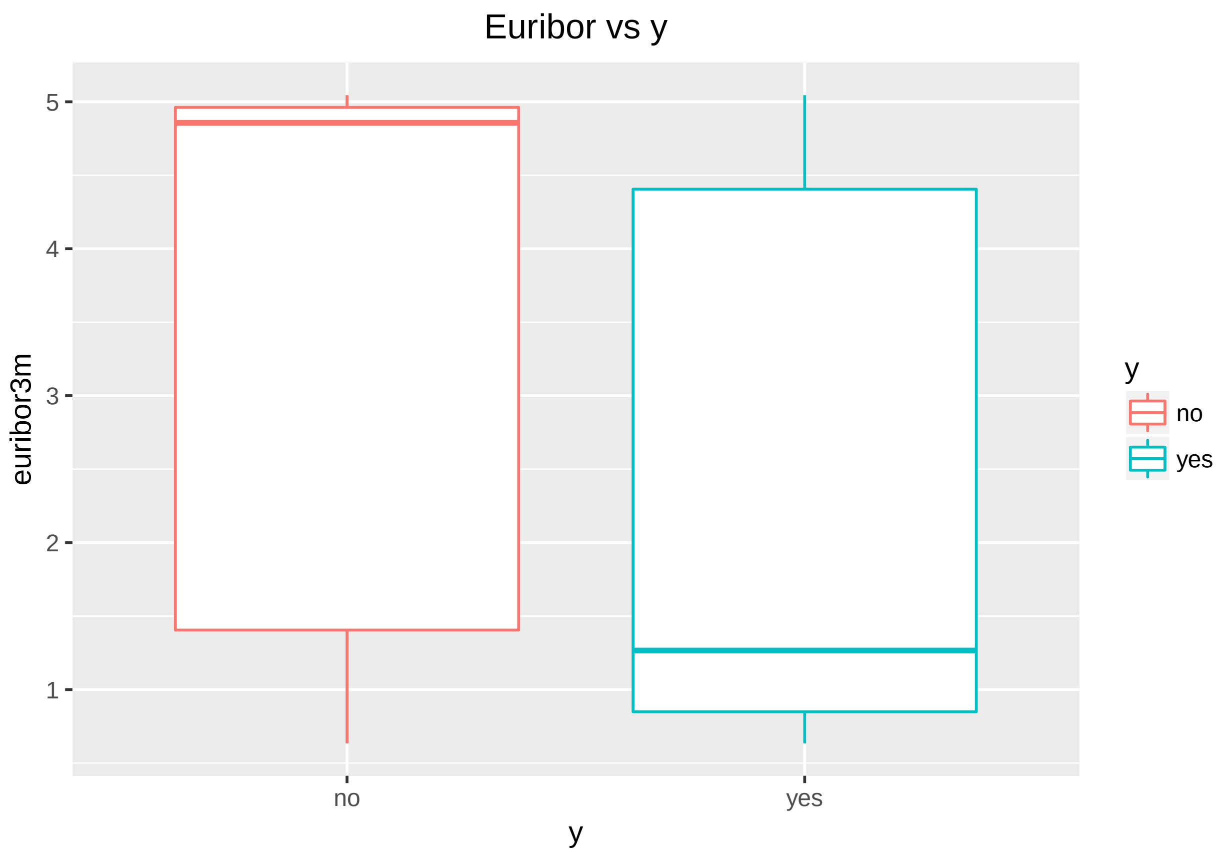



```
ggplot(data=bank, aes(x=euribor3m, col=y))+
    geom_histogram(position = "fill")+
    ggtitle("Histogram of euribor rate based on subscription")+
    theme(plot.title = element_text(hjust = 0.5))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Warning: Removed 16 rows containing missing values (geom_bar).
```




```r
ggplot(bank, aes(x=y, y=euribor3m, col=y))+
    geom_boxplot()+
    ggtitle("Euribor vs y")+
    theme(plot.title = element_text(hjust = 0.5))
```




```
prop.table(table(bank[bank$euribor3m<5,]$y))
```

```
##
##        no       yes
## 0.8874964 0.1125036
```

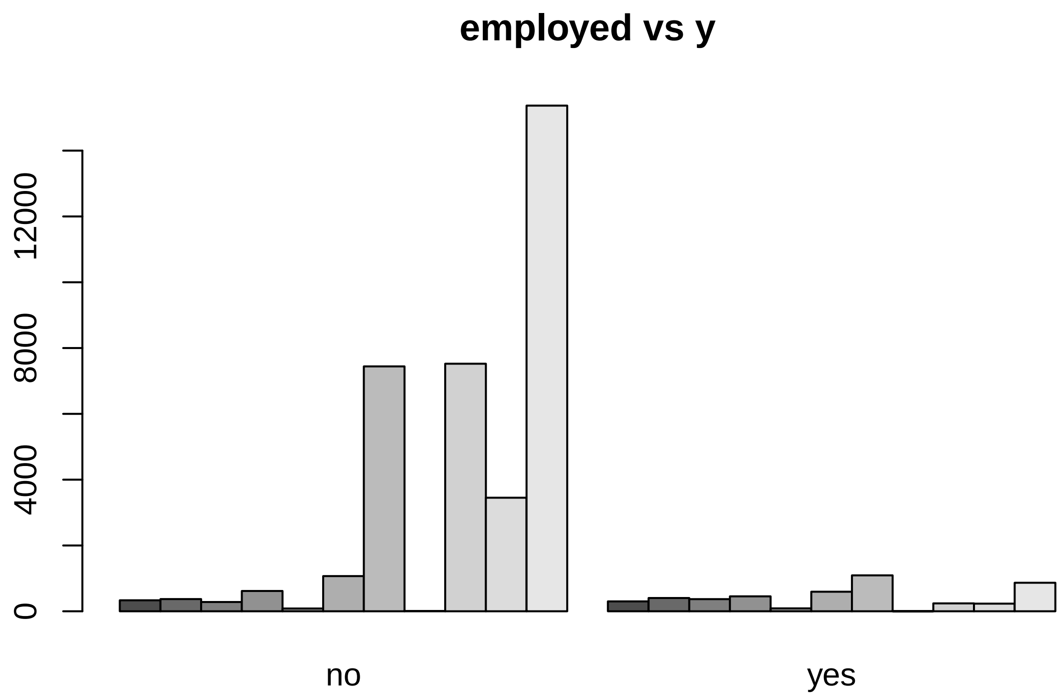
```
prop.table(table(bank[bank$euribor3m<3,]$y))
```

```
##
##        no       yes
## 0.7554453 0.2445547
```

```
prop.table(table(bank[bank$euribor3m<1,]$y))
```

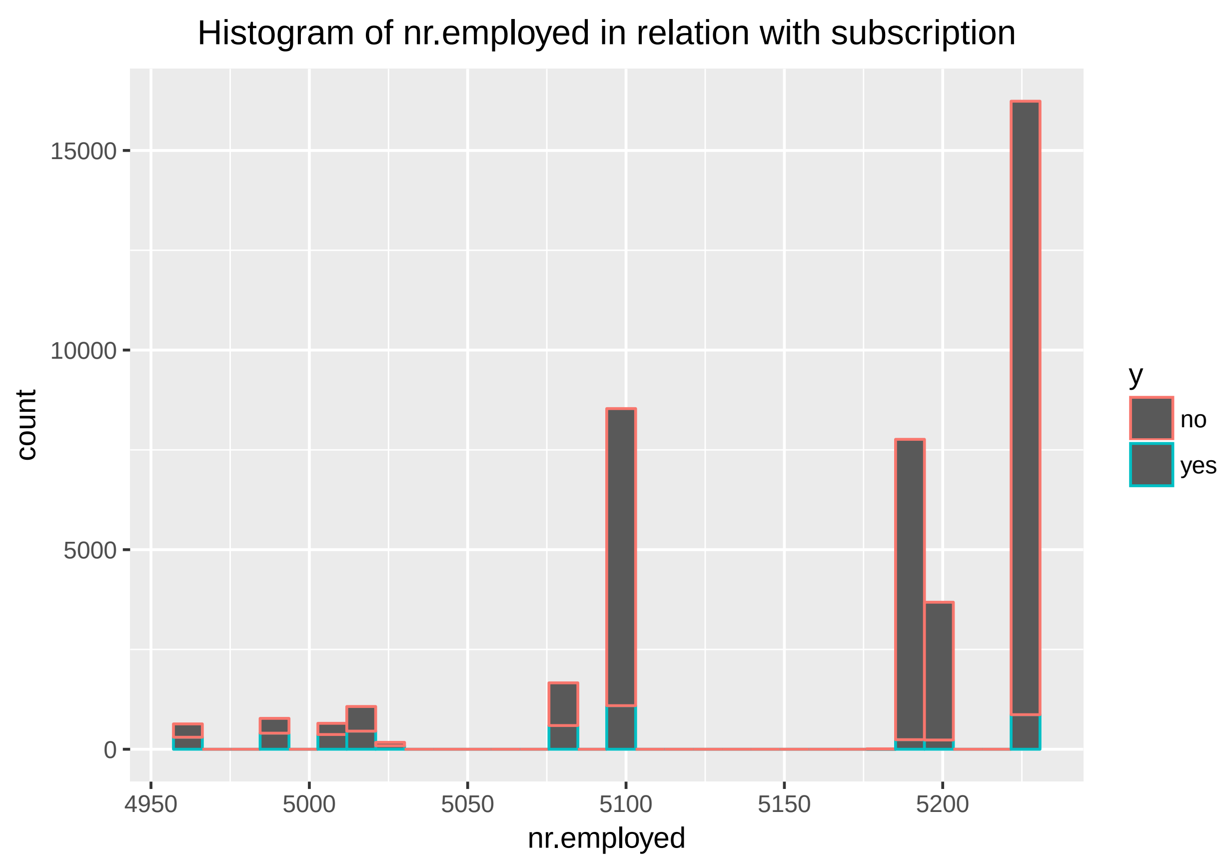
```
##
##        no       yes
## 0.5429306 0.4570694
```

```
# The barplot, histogram and proportion tables of the nr.employed shows that, as the
# number of employees increases the efficiency of the campaign decreases. When the number
# of employees are less than 5000 the probability of a client accepting the offer
# increases significantly.
bartable <- table(bank$nr.employed, bank$y)
barplot(bartable, beside = TRUE, legend = levels(unique(bank$nr.employed)),
        main="employed vs y")
```




```r
ggplot(data=bank, aes(x=nr.employed, col=y))+
    geom_histogram()+
    ggtitle("Histogram of nr.employed in relation with subscription")+
    theme(plot.title = element_text(hjust = 0.5))
```

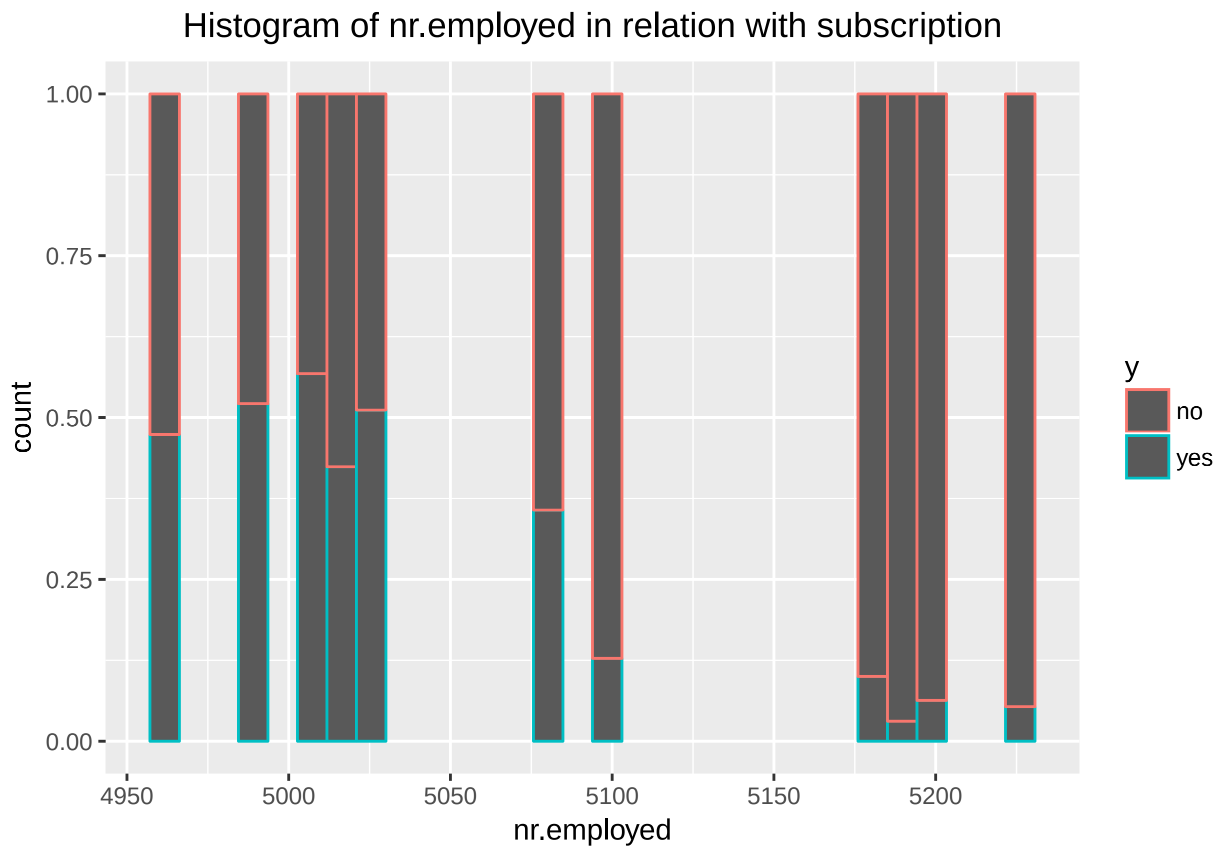
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```r
ggplot(data=bank, aes(x=nr.employed, col=y))+
    geom_histogram(position="fill")+
    ggtitle("Histogram of nr.employed in relation with subscription")+
    theme(plot.title = element_text(hjust = 0.5))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 38 rows containing missing values (geom_bar).
```




```r
prop.table(table(bank[bank$nr.employed<5228,]$y))
```

```
##
##        no       yes
## 0.8487617 0.1512383
```

```r
prop.table(table(bank[bank$nr.employed<5100,]$y))
```

```
##
##        no       yes
## 0.7554453 0.2445547
```

```r
prop.table(table(bank[bank$nr.employed<5000,]$y))
```

```
##
##  no yes
## 0.5 0.5
```

```r
# Getting the indexes of factor columns from bank data set, to convert them into
# numeric for creating a correlation plot
bank_dup <- bank
factors_index <- which(sapply(bank_dup, is.factor))
factors_index
```

```
##          job     marital   education     default     housing        loan 
##            2           3           4           5           6           7 
##      contact       month day_of_week    poutcome           y 
##            8           9          10          15          21
```

```r
# Converting factor columns to numeric
bank_dup[,factors_index] <- lapply(factors_index, function(fac)
    {as.numeric(bank_dup[,fac])})
str(bank_dup)
```

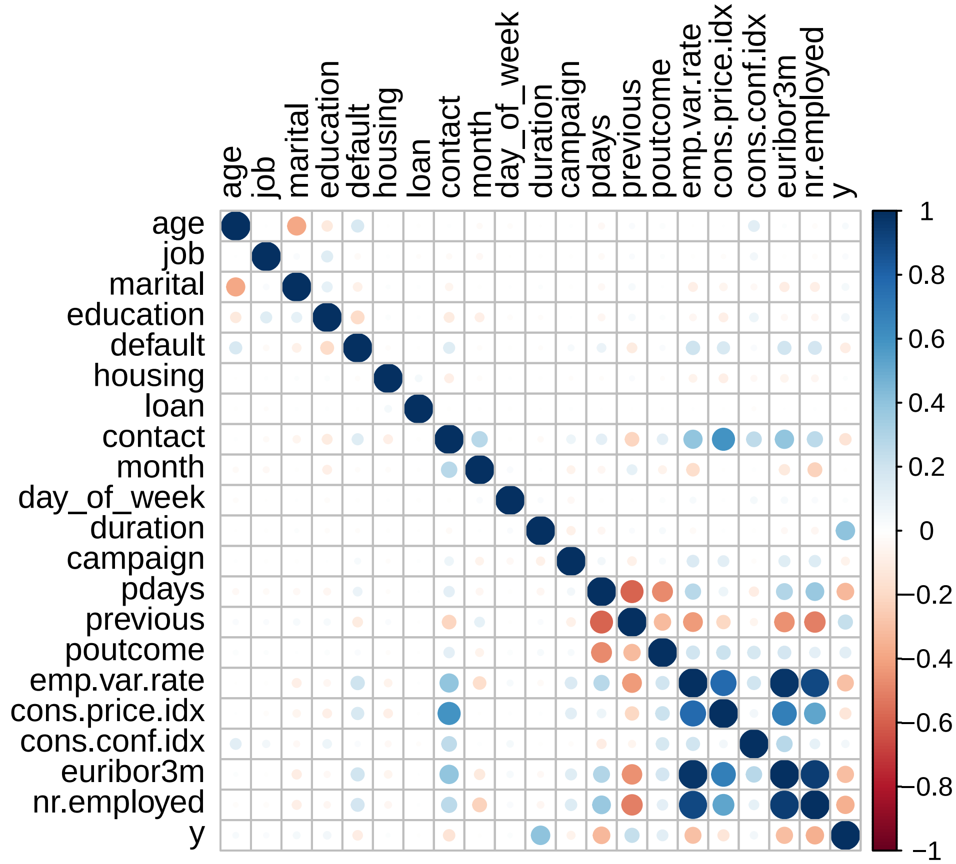
```
## 'data.frame':    41188 obs. of  21 variables:
##  $ age           : int  56 57 37 40 56 45 59 41 24 25 ...
##  $ job           : num  4 8 8 1 8 8 1 2 10 8 ...
##  $ marital       : num  2 2 2 2 2 2 2 2 3 3 ...
##  $ education     : num  1 4 4 2 4 3 6 8 6 4 ...
##  $ default       : num  1 2 1 1 1 2 1 2 1 1 ...
##  $ housing       : num  1 1 3 1 1 1 1 1 3 3 ...
##  $ loan          : num  1 1 1 1 3 1 1 1 1 1 ...
##  $ contact       : num  2 2 2 2 2 2 2 2 2 2 ...
##  $ month         : num  7 7 7 7 7 7 7 7 7 7 ...
##  $ day_of_week   : num  2 2 2 2 2 2 2 2 2 2 ...
##  $ duration      : int  261 149 226 151 307 198 139 217 380 50 ...
##  $ campaign      : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ pdays         : int  999 999 999 999 999 999 999 999 999 999 ...
##  $ previous      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ poutcome      : num  2 2 2 2 2 2 2 2 2 2 ...
##  $ 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             : num  1 1 1 1 1 1 1 1 1 1 ...
```

```r
# Correlation plot of bank explains the correlation between different columns of bank
library(corrplot)
```

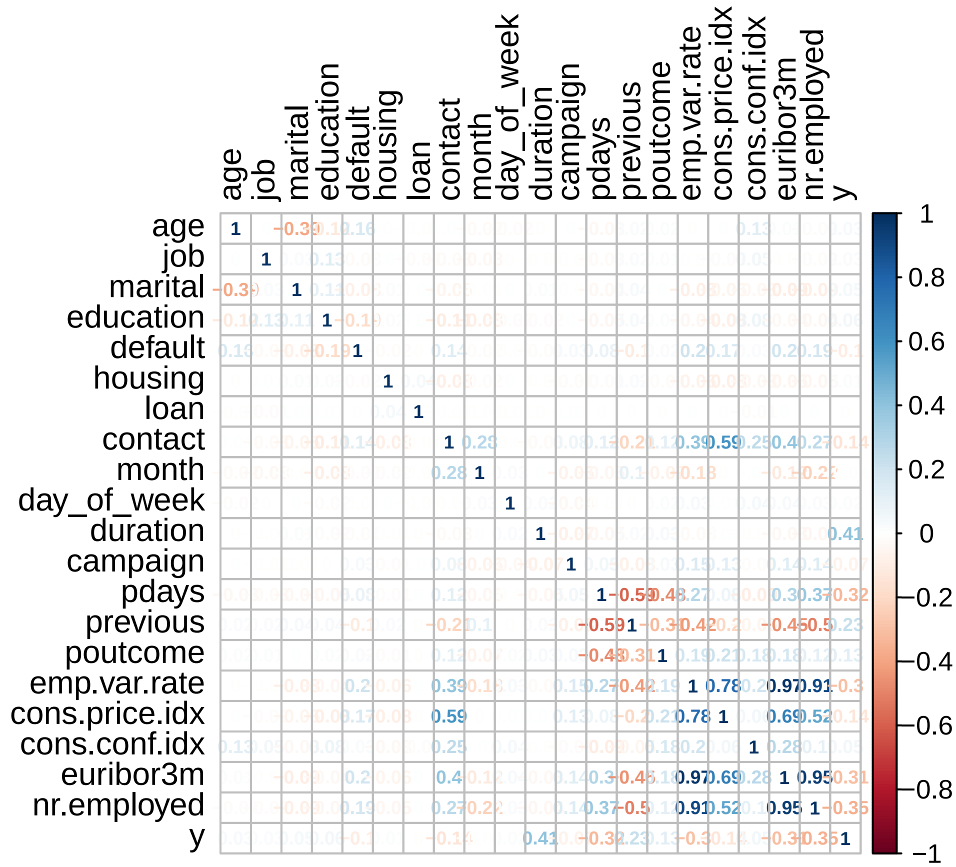
```
## Warning: package 'corrplot' was built under R version 3.4.2
```

```
## corrplot 0.84 loaded
```

```r
C <- cor(bank_dup)
corrplot(C, tl.col="black")
```



```
corrplot(C, method = "number",number.cex=0.6, tl.col="black")
```



```
### Model building and evaluation ###


# Normalizing the numeric features in bank to reduce the bias towards features with
# comparitively high numeric values
normalize <- function(x) {
  return((x - min(x)) / (max(x) - min(x)))
}
factors_index <- which(sapply(bank, is.factor))
factors_index
```

```
##         job     marital   education     default     housing        loan
##           2           3           4           5           6           7
##     contact       month day_of_week    poutcome           y
##           8           9          10          15          21
```

```
bank_n <- as.data.frame(lapply(bank[ ,-factors_index], normalize))
names(bank_n)
```

```
##  [1] "age"            "duration"       "campaign"       "pdays"
##  [5] "previous"       "emp.var.rate"   "cons.price.idx" "cons.conf.idx"
##  [9] "euribor3m"      "nr.employed"
```

```
bank[names(bank_n)] <- bank_n[names(bank_n)]
str(bank)
```

```
## 'data.frame':    41188 obs. of  21 variables:
##  $ age           : num  0.481 0.494 0.247 0.284 0.481 ...
##  $ job           : Factor w/ 12 levels "admin.","blue-collar",..: 4 8 8 1 8 8 1 2 10 8 ...
##  $ marital       : Factor w/ 4 levels "divorced","married",..: 2 2 2 2 2 2 2 2 3 3 ...
##  $ education     : Factor w/ 8 levels "basic.4y","basic.6y",..: 1 4 4 2 4 3 6 8 6 4 ...
##  $ default       : Factor w/ 3 levels "no","unknown",..: 1 2 1 1 1 2 1 2 1 1 ...
##  $ housing       : Factor w/ 3 levels "no","unknown",..: 1 1 3 1 1 1 1 1 3 3 ...
##  $ loan          : Factor w/ 3 levels "no","unknown",..: 1 1 1 1 3 1 1 1 1 1 ...
##  $ contact       : Factor w/ 2 levels "cellular","telephone": 2 2 2 2 2 2 2 2 2 2 ...
##  $ month         : Factor w/ 10 levels "apr","aug","dec",..: 7 7 7 7 7 7 7 7 7 7 ...
##  $ day_of_week   : Factor w/ 5 levels "fri","mon","thu",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ duration      : num  0.0531 0.0303 0.046 0.0307 0.0624 ...
##  $ campaign      : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ pdays         : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ previous      : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ poutcome      : Factor w/ 3 levels "failure","nonexistent",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ emp.var.rate  : num  0.938 0.938 0.938 0.938 0.938 ...
##  $ cons.price.idx: num  0.699 0.699 0.699 0.699 0.699 ...
##  $ cons.conf.idx : num  0.603 0.603 0.603 0.603 0.603 ...
##  $ euribor3m     : num  0.957 0.957 0.957 0.957 0.957 ...
##  $ nr.employed   : num  0.86 0.86 0.86 0.86 0.86 ...
##  $ y             : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
```

```
# Splitting the bank data set into training and testing sets
library(irr)
```

```
## Loading required package: lpSolve
```

```
library(ROCR)
```

```
## Loading required package: gplots
```

```
##
## Attaching package: 'gplots'
```

```
## The following object is masked from 'package:stats':
##
##     lowess
```

```
library(caret)
```

```
## Loading required package: lattice
```

```
## Warning in as.POSIXlt.POSIXct(Sys.time()): unknown timezone 'zone/tz/2017c.
## 1.0/zoneinfo/America/New_York'
```

```
library(gmodels)
set.seed(141)
train_ind <- createDataPartition(bank$y, p=0.75, list=FALSE)
train_data <- bank[train_ind, ]
test_data <- bank[-train_ind, ]

# The target variable y is uniformly distributed among both train and test sets
prop.table(table(train_data$y))
```

```
##
##        no       yes
## 0.8873458 0.1126542
```

```
prop.table(table(test_data$y))
```

```
##
##        no       yes
## 0.8873458 0.1126542
```

```
### Naive Bayes ###

library(e1071)

set.seed(141)
# Building naive bayes model using train data
bayes_model <- naiveBayes(train_data[-21], train_data$y, laplace = 3)
bayes_model
```

```
##
## Naive Bayes Classifier for Discrete Predictors
##
## Call:
## naiveBayes.default(x = train_data[-21], y = train_data$y, laplace = 3)
##
## A-priori probabilities:
## train_data$y
##        no       yes
## 0.8873458 0.1126542
##
## Conditional probabilities:
##             age
## train_data$y      [,1]      [,2]
##          no  0.2828580 0.1220641
##          yes 0.2955974 0.1711786
##
##             job
## train_data$y      admin. blue-collar entrepreneur   housemaid  management
##          no  0.247458739 0.236528582  0.035996648 0.026159507 0.071483222
##          yes 0.282992036 0.130546075  0.027303754 0.024744027 0.074800910
```

```
##             job
## train_data$y     retired self-employed    services     student  technician
##          no  0.034903632   0.034757897 0.098917915 0.017196779 0.163915911
##          yes 0.095847554   0.034129693 0.069681456 0.064846416 0.155858931
##             job
## train_data$y  unemployed     unknown
##          no  0.024301381 0.008379786
##          yes 0.030716724 0.008532423
##
##             marital
## train_data$y    divorced     married      single     unknown
##          no  0.112423878 0.613572549 0.271852095 0.002151479
##          yes 0.106529210 0.542096220 0.348224513 0.003150057
##
##             education
## train_data$y     basic.4y     basic.6y     basic.9y  high.school
##          no  0.1015855659 0.0575177693 0.1539639147 0.2306542737
##          yes 0.0924657534 0.0428082192 0.0970319635 0.2248858447
##             education
## train_data$y   illiterate professional.course university.degree
##          no  0.0005831966        0.1270275196      0.2877346455
##          yes 0.0019977169        0.1269977169      0.3567351598
##             education
## train_data$y      unknown
##          no  0.0409331146
##          yes 0.0570776256
##
##             default
## train_data$y           no      unknown          yes
##          no  0.7749452954 0.2248723559 0.0001823487
##          yes 0.9051304099 0.0940097449 0.0008598452
##
##             housing
## train_data$y        no   unknown       yes
##          no  0.4530999 0.0245806 0.5223195
##          yes 0.4396675 0.0246489 0.5356836
##
##             loan
## train_data$y        no   unknown       yes
##          no  0.8238512 0.0245806 0.1515682
##          yes 0.8314703 0.0246489 0.1438808
##
##             contact
## train_data$y  cellular telephone
##          no  0.6074333 0.3925667
##          yes 0.8293173 0.1706827
##
##             month
## train_data$y         apr         aug         dec         jul         jun
##          no  0.057213658 0.149375023 0.002623811 0.176341970 0.131044787
##          yes 0.113105413 0.138746439 0.021652422 0.138461538 0.120797721
##             month
## train_data$y         mar         may         nov         oct         sep
##          no  0.007834991 0.354651798 0.100761634 0.011406290 0.008746037
##          yes 0.060968661 0.190313390 0.091737892 0.069515670 0.054700855
##
```

```
##             day_of_week
## train_data$y       fri       mon       thu       tue       wed
##          no  0.1911325 0.2076861 0.2077591 0.1970393 0.1963830
##          yes 0.1825465 0.1814020 0.2257511 0.2068670 0.2034335
##
##             duration
## train_data$y       [,1]       [,2]
##          no  0.04490545 0.04230937
##          yes 0.11229918 0.08267516
##
##             campaign
## train_data$y       [,1]       [,2]
##          no  0.02983606 0.05225394
##          yes 0.01907524 0.02972799
##
##             pdays
## train_data$y      [,1]      [,2]
##          no  0.9849543 0.1213524
##          yes 0.7940512 0.4029340
##
##             previous
## train_data$y       [,1]       [,2]
##          no  0.01886625 0.05824195
##          yes 0.07089491 0.12302136
##
##             poutcome
## train_data$y    failure nonexistent    success
##          no  0.09956236  0.88719912 0.01323851
##          yes 0.13470909  0.67526512 0.19002580
##
##             emp.var.rate
## train_data$y      [,1]      [,2]
##          no  0.7593705 0.3091820
##          yes 0.4522450 0.3401953
##
##             cons.price.idx
## train_data$y      [,1]      [,2]
##          no  0.5465981 0.2181408
##          yes 0.4503395 0.2640743
##
##             cons.conf.idx
## train_data$y      [,1]      [,2]
##          no  0.4269046 0.1840212
##          yes 0.4642536 0.2579954
##
##             euribor3m
## train_data$y      [,1]      [,2]
##          no  0.7193968 0.3717477
##          yes 0.3399032 0.3971505
##
##             nr.employed
## train_data$y      [,1]      [,2]
##          no  0.8029169 0.2441870
##          yes 0.4973375 0.3322022
```

```r
# Applying the model on the test data to predict the dependent variable
bayes_pred <- predict(bayes_model, test_data[-21])
```

```r
str(bayes_pred)
```

```
##  Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
```

```r
# The model has accuracy of around 86% which is good. The kappa value of  0.42 suggests
# a moderate agreement between the true and predicted values
CrossTable(bayes_pred, test_data$y)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## | Chi-square contribution |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  10297
##
##
##              | test_data$y
##   bayes_pred |        no |       yes | Row Total |
## -------------|-----------|-----------|-----------|
##           no |      8177 |       444 |      8621 |
##              |    36.332 |   286.175 |           |
##              |     0.948 |     0.052 |     0.837 |
##              |     0.895 |     0.383 |           |
##              |     0.794 |     0.043 |           |
## -------------|-----------|-----------|-----------|
##          yes |       960 |       716 |      1676 |
##              |   186.883 |  1472.027 |           |
##              |     0.573 |     0.427 |     0.163 |
##              |     0.105 |     0.617 |           |
##              |     0.093 |     0.070 |           |
## -------------|-----------|-----------|-----------|
## Column Total |      9137 |      1160 |     10297 |
##              |     0.887 |     0.113 |           |
## -------------|-----------|-----------|-----------|
##
##
```

```r
confusionMatrix(bayes_pred,test_data$y, positive = "yes")
```

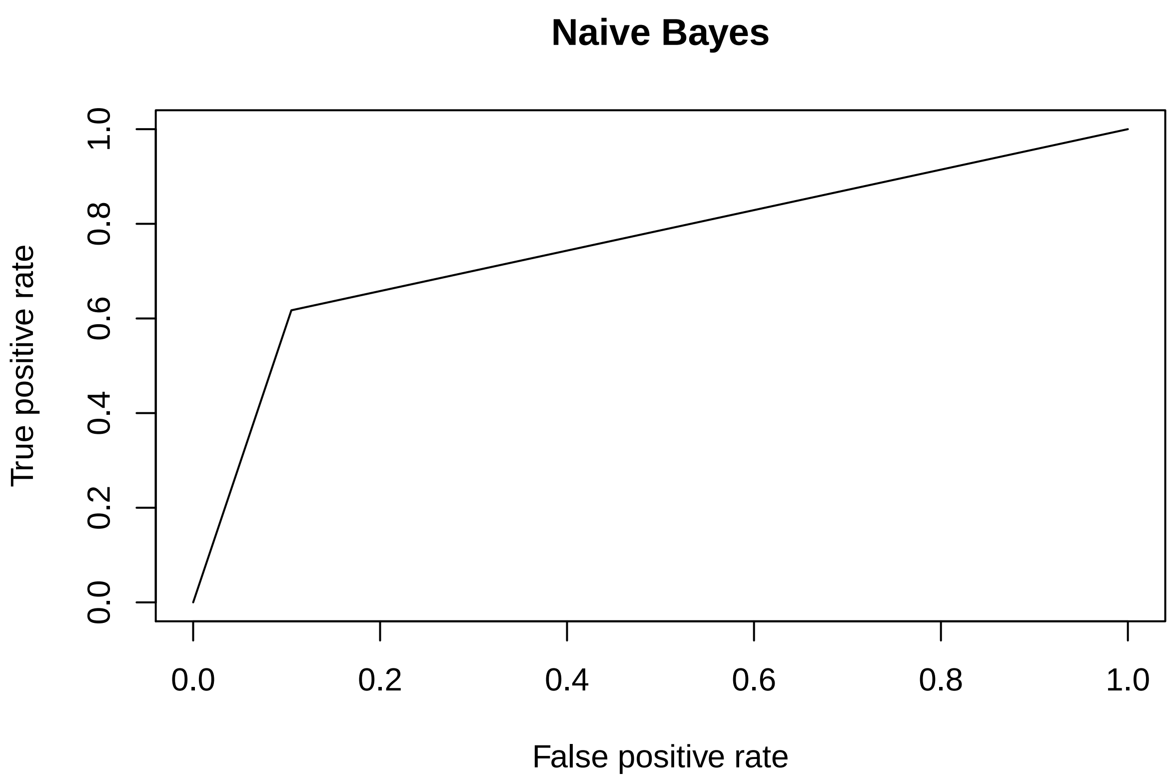
```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   no  yes
##        no  8177  444
##        yes  960  716
##
##                Accuracy : 0.8636
##                  95% CI : (0.8569, 0.8702)
##     No Information Rate : 0.8873
##     P-Value [Acc > NIR] : 1
##
```

```
##                  Kappa : 0.4289
##  Mcnemar's Test P-Value : <2e-16
##
##             Sensitivity : 0.61724
##             Specificity : 0.89493
##          Pos Pred Value : 0.42721
##          Neg Pred Value : 0.94850
##              Prevalence : 0.11265
##          Detection Rate : 0.06953
##    Detection Prevalence : 0.16277
##       Balanced Accuracy : 0.75609
##
##        'Positive' Class : yes
##
```

```
nb_accuracy <- confusionMatrix(bayes_pred,test_data$y, positive = "yes")$overall[[1]]
nb_accuracy <- nb_accuracy*100
nb_accuracy
```

```
## [1] 86.36496
```

```
# Evaluation of the model using ROC curve and AUC(area under the curve) shows the model
# is performing fine in predicting the false positives and false negatives with auc value
# of 0.75, which has value of 1 for ideal case.
pred_nb <- prediction(predictions = as.numeric(bayes_pred), labels = as.numeric(test_data$y))
perf_nb <- performance(pred_nb,measure = "tpr", x.measure = "fpr")
plot(perf_nb, main="Naive Bayes")
```




```
perf.auc_nb <- performance(pred_nb, measure = "auc")
nb_auc <- unlist(perf.auc_nb@y.values)
nb_auc
```

```
## [1] 0.756087
```

```r
### Decision tree ###

library(C50)

set.seed(141)
# Building a decision tree classification model using training set data
decision_tree <- C5.0(train_data[-21], train_data$y, trails=20)
# The summary of decision tree model shows that the variables poutcome, duration,
# nr.employed, month, age and emp.var.rate are the most important variables in predicting
# the target variable y
summary(decision_tree)
```

```
##
## Call:
## C5.0.default(x = train_data[-21], y = train_data$y, trails = 20)
##
##
## C5.0 [Release 2.07 GPL Edition]      Sun Dec 10 17:34:15 2017
## -------------------------------
##
## Class specified by attribute `outcome'
##
## Read 30891 cases (21 attributes) from undefined.data
##
## Decision tree:
##
## poutcome = success:
## :...duration <= 0.03273689:
## :   :...cons.conf.idx <= 0.376569: no (69/3)
## :   :   cons.conf.idx > 0.376569:
## :   :   :...campaign > 0.01818182:
## :   :       :...marital = divorced: yes (3/1)
## :   :       :   marital in {married,single,unknown}: no (25/1)
## :   :       campaign <= 0.01818182:
## :   :       :...month in {jun,oct}: no (39/13)
## :   :           month = apr:
## :   :           :...housing in {no,unknown}: no (3)
## :   :           :   housing = yes: yes (1)
## :   :           month = dec:
## :   :           :...duration <= 0.02806019: no (3)
## :   :           :   duration > 0.02806019: yes (2)
## :   :           month = jul:
## :   :           :...duration <= 0.02704351: no (6)
## :   :           :   duration > 0.02704351: yes (2)
## :   :           month = mar:
## :   :           :...housing in {no,unknown}: yes (5)
## :   :           :   housing = yes: no (5/1)
## :   :           month = may:
## :   :           :...duration <= 0.02704351: no (3)
## :   :           :   duration > 0.02704351: yes (3)
## :   :           month = nov:
## :   :           :...euribor3m <= 0.01836318: yes (14/1)
## :   :           :   euribor3m > 0.01836318: no (2)
## :   :           month = sep:
## :   :           :...housing = no: yes (4)
## :   :           :   housing in {unknown,yes}: no (11/4)
```

```
## :   :           month = aug:
## :   :           :...age <= 0.3580247:
## :   :               :...euribor3m <= 0.04919519: yes (7/2)
## :   :               :   euribor3m > 0.04919519: no (18/2)
## :   :               age > 0.3580247:
## :   :               :...housing in {no,unknown}: no (3)
## :   :                   housing = yes: yes (14/6)
## :   duration > 0.03273689:
## :   :...nr.employed <= 0.4257089: yes (649/116)
## :       nr.employed > 0.4257089:
## :       :...month = mar: yes (2)
## :           month in {aug,dec,jul,jun,nov,oct,sep}: no (15/2)
## :           month = apr:
## :           :...pdays <= 0.007007007: yes (30/7)
## :           :   pdays > 0.007007007: no (6/1)
## :           month = may:
## :           :...job in {entrepreneur,housemaid,management,retired,
## :               :       self-employed,student,unknown}: no (10)
## :               job in {technician,unemployed}: yes (15/4)
## :               job = admin.:
## :               :...day_of_week in {fri,wed}: yes (6/1)
## :               :   day_of_week in {mon,tue}: no (7)
## :               :   day_of_week = thu:
## :               :   :...housing = no: yes (1)
## :               :       housing in {unknown,yes}: no (3)
## :               job = services:
## :               :...education in {basic.4y,basic.6y,basic.9y,illiterate,
## :               :   :             professional.course,unknown}: yes (3)
## :               :   education = university.degree: no (1)
## :               :   education = high.school:
## :               :   :...housing = no: yes (2)
## :               :       housing in {unknown,yes}: no (2)
## :               job = blue-collar:
## :               :...duration <= 0.09109394: no (15/1)
## :                   duration > 0.09109394:
## :                   :...education in {basic.4y,basic.6y,high.school,illiterate,
## :                       :             professional.course,university.degree,
## :                       :             unknown}: yes (5)
## :                       education = basic.9y:
## :                       :...day_of_week = thu: yes (2)
## :                           day_of_week in {fri,mon,tue,wed}: no (4)
## poutcome in {failure,nonexistent}:
## :...duration <= 0.07991053:
##     :...nr.employed <= 0.4257089:
##     :   :...duration <= 0.03497357: no (1213/163)
##     :   :   duration > 0.03497357:
##     :   :   :...contact = telephone: no (150/48)
##     :   :       contact = cellular:
##     :   :       :...duration > 0.050427:
##     :   :           :...cons.price.idx <= 0.06936867:
##     :   :           :   :...day_of_week in {fri,thu,tue}: no (59/18)
##     :   :           :   :   day_of_week in {mon,wed}: yes (32/14)
##     :   :           :   cons.price.idx > 0.06936867:
##     :   :           :   :...education in {basic.6y,basic.9y,high.school,
##     :   :           :       :             illiterate,
##     :   :           :       :             university.degree}: yes (288/101)
```

```
##     :   :           :       education = professional.course:
##     :   :           :       :...age <= 0.2098765: yes (21/5)
##     :   :           :       :   age > 0.2098765: no (27/9)
##     :   :           :       education = basic.4y:
##     :   :           :       :...poutcome = nonexistent: yes (23/6)
##     :   :           :       :   poutcome = failure:
##     :   :           :       :   :...month in {apr,aug,dec,jul,mar,may,oct,
##     :   :           :       :       :         sep}: no (8)
##     :   :           :       :       month in {jun,nov}: yes (3)
##     :   :           :       education = unknown:
##     :   :           :       :...emp.var.rate > 0.3541667: no (4)
##     :   :           :           emp.var.rate <= 0.3541667:
##     :   :           :           :...job in {admin.,blue-collar,entrepreneur,
##     :   :           :               :       housemaid,management,self-employed,
##     :   :           :               :       student,unemployed,
##     :   :           :               :       unknown}: yes (12)
##     :   :           :               job in {retired,services,
##     :   :           :                       technician}: no (12/5)
##     :   :           duration <= 0.050427:
##     :   :           :...education in {basic.4y,high.school,illiterate,
##     :   :               :             unknown}: no (179/62)
##     :   :               education = basic.6y:
##     :   :               :...marital in {divorced,single,unknown}: yes (5)
##     :   :               :   marital = married:
##     :   :               :   :...age <= 0.3333333: no (4)
##     :   :               :       age > 0.3333333: yes (2)
##     :   :               education = basic.9y:
##     :   :               :...day_of_week in {mon,wed}: no (16/4)
##     :   :               :   day_of_week in {thu,tue}: yes (7/1)
##     :   :               :   day_of_week = fri:
##     :   :               :   :...job in {admin.,blue-collar,entrepreneur,
##     :   :               :       :       housemaid,management,self-employed,
##     :   :               :       :       student,technician,unemployed,
##     :   :               :       :       unknown}: no (4)
##     :   :               :       job in {retired,services}: yes (4)
##     :   :               education = professional.course:
##     :   :               :...job in {admin.,entrepreneur,housemaid,management,
##     :   :               :   :       services,student,unknown}: no (7/2)
##     :   :               :   job in {blue-collar,self-employed,
##     :   :               :   :       unemployed}: yes (9/1)
##     :   :               :   job = retired:
##     :   :               :   :...campaign <= 0: yes (4)
##     :   :               :   :   campaign > 0: no (3)
##     :   :               :   job = technician:
##     :   :               :   :...loan = no: no (26/8)
##     :   :               :       loan in {unknown,yes}: yes (5/1)
##     :   :               education = university.degree:
##     :   :               :...poutcome = failure:
##     :   :                   :...pdays <= 0.006006006: yes (4)
##     :   :                   :   pdays > 0.006006006: no (58/16)
##     :   :                   poutcome = nonexistent:
##     :   :                   :...job in {blue-collar,entrepreneur,management,
##     :   :                       :       technician,unknown}: no (40/14)
##     :   :                       job in {housemaid,services,
##     :   :                       :       unemployed}: yes (4)
##     :   :                       job = retired:
```

```
##     :   :                       :...emp.var.rate <= 0: no (2)
##     :   :                       :   emp.var.rate > 0: yes (2)
##     :   :                       job = self-employed:
##     :   :                       :...marital in {divorced,married}: yes (4)
##     :   :                       :   marital in {single,unknown}: no (6/1)
##     :   :                       job = student:
##     :   :                       :...duration <= 0.04209028: yes (3)
##     :   :                       :   duration > 0.04209028: no (2)
##     :   :                       job = admin.:
##     :   :                       :...day_of_week = fri: yes (8/3)
##     :   :                           day_of_week = wed: no (12/5)
##     :   :                           day_of_week = mon:
##     :   :                           :...campaign <= 0: no (7)
##     :   :                           :   campaign > 0: yes (7/3)
##     :   :                           day_of_week = thu:
##     :   :                           :...campaign <= 0: yes (10/3)
##     :   :                           :   campaign > 0: no (4)
##     :   :                           day_of_week = tue:
##     :   :                           :...nr.employed <= 0.2037807: yes (12)
##     :   :                               nr.employed > 0.2037807:
##     :   :                               :...duration <= 0.03924359: yes (2)
##     :   :                                   duration > 0.03924359: no (6)
##     :   nr.employed > 0.4257089:
##     :   :...age > 0.5308642:
##     :       :...duration <= 0.02765352: no (52/5)
##     :       :   duration > 0.02765352:
##     :       :   :...euribor3m > 0.1736568: yes (42/14)
##     :       :       euribor3m <= 0.1736568:
##     :       :       :...marital in {married,single,unknown}: no (9)
##     :       :           marital = divorced:
##     :       :           :...housing in {no,unknown}: no (2)
##     :       :               housing = yes: yes (2)
##     :       age <= 0.5308642:
##     :       :...month in {aug,jul,jun,may,nov,sep}: no (20581/110)
##     :           month in {apr,dec}:
##     :           :...euribor3m > 0.1736568: no (1167/99)
##     :           :   euribor3m <= 0.1736568:
##     :           :   :...duration <= 0.0351769: no (86/4)
##     :           :       duration > 0.0351769:
##     :           :       :...education in {basic.6y,illiterate,
##     :           :           :              professional.course,
##     :           :           :              university.degree,
##     :           :           :              unknown}: yes (46/18)
##     :           :           education in {basic.9y,high.school}: no (24/8)
##     :           :           education = basic.4y:
##     :           :           :...age <= 0.1111111: no (2)
##     :           :               age > 0.1111111: yes (3)
##     :           month in {mar,oct}:
##     :           :...duration <= 0.01911346: no (59/5)
##     :               duration > 0.01911346:
##     :               :...campaign > 0.05454545:
##     :                   :...duration <= 0.05002033: no (11)
##     :                   :   duration > 0.05002033: yes (2)
##     :                   campaign <= 0.05454545:
##     :                   :...month = oct:
##     :                       :...loan in {no,unknown}: yes (27/2)
```

```
##     :                        :   loan = yes: no (1)
##     :                        month = mar:
##     :                        :...default = yes: yes (0)
##     :                            default = unknown:
##     :                            :...job in {admin.,blue-collar,entrepreneur,
##     :                            :   :       housemaid,management,retired,
##     :                            :   :       self-employed,services,student,
##     :                            :   :       technician,unknown}: no (3)
##     :                            :   job = unemployed: yes (1)
##     :                            default = no:
##     :                            :...duration > 0.03761692: yes (49/10)
##     :                                duration <= 0.03761692:
##     :                                :...marital = divorced: yes (1)
##     :                                    marital in {single,
##     :                                    :           unknown}: no (35/11)
##     :                                    marital = married:
##     :                                    :...housing in {no,
##     :                                        :           unknown}: no (9/3)
##     :                                        housing = yes: yes (13/4)
##     duration > 0.07991053:
##     :...nr.employed <= 0.4257089:
##         :...emp.var.rate > 0.1041667:
##         :   :...loan in {unknown,yes}: yes (42/6)
##         :   :   loan = no:
##         :   :   :...poutcome = nonexistent: yes (125/26)
##         :   :       poutcome = failure:
##         :   :       :...emp.var.rate <= 0.3541667:
##         :   :           :...campaign <= 0.05454545: yes (40/8)
##         :   :           :   campaign > 0.05454545: no (3)
##         :   :           emp.var.rate > 0.3541667:
##         :   :           :...marital = divorced: yes (4)
##         :   :               marital in {married,single,unknown}: no (17/5)
##         :   emp.var.rate <= 0.1041667:
##         :   :...loan = unknown: yes (12/4)
##         :       loan = yes:
##         :       :...marital in {divorced,single,unknown}: no (11/1)
##         :       :   marital = married:
##         :       :   :...month in {aug,nov}: yes (8/1)
##         :       :       month in {apr,jul,mar,may,oct}: no (4)
##         :       :       month = dec:
##         :       :       :...duration <= 0.101464: yes (2)
##         :       :       :   duration > 0.101464: no (2)
##         :       :       month = jun:
##         :       :       :...day_of_week in {mon,thu}: yes (3)
##         :       :       :   day_of_week in {fri,tue,wed}: no (5/1)
##         :       :       month = sep:
##         :       :       :...duration <= 0.1122407: yes (2)
##         :       :           duration > 0.1122407: no (4)
##         :       loan = no:
##         :       :...cons.price.idx > 0.1745908: yes (87/23)
##         :           cons.price.idx <= 0.1745908:
##         :           :...contact = telephone:
##         :               :...marital in {divorced,single}: yes (11/4)
##         :               :   marital in {married,unknown}: no (15/2)
##         :               contact = cellular:
##         :               :...duration <= 0.09577064: yes (64/14)
```

```
##         :                   duration > 0.09577064:
##         :                   :...campaign > 0.05454545: yes (6)
##         :                       campaign <= 0.05454545:
##         :                       :...job in {housemaid,management,
##         :                           :       student}: yes (21/6)
##         :                           job in {self-employed,services,technician,
##         :                           :       unemployed,unknown}: no (27/9)
##         :                           job = blue-collar:
##         :                           :...euribor3m <= 0.02221718: yes (2)
##         :                           :   euribor3m > 0.02221718: no (3)
##         :                           job = entrepreneur:
##         :                           :...duration <= 0.1268808: yes (2)
##         :                           :   duration > 0.1268808: no (2)
##         :                           job = retired:
##         :                           :...cons.price.idx <= 0.1044427: no (17/5)
##         :                           :   cons.price.idx > 0.1044427: yes (3)
##         :                           job = admin.:
##         :                           :...month in {apr,dec,jul,jun,mar,may,nov,
##         :                               :         oct}: no (16/3)
##         :                               month = sep: yes (4)
##         :                               month = aug:
##         :                               :...day_of_week in {fri,
##         :                                   :               wed}: yes (5/1)
##         :                                   day_of_week in {mon,thu,
##         :                                                   tue}: no (5/1)
##         nr.employed > 0.4257089:
##         :...duration <= 0.1309475:
##             :...age > 0.5185185:
##             :   :...month in {apr,jul}: yes (12/1)
##             :   :   month in {aug,dec,jun,mar,may,nov,oct,sep}: no (19/3)
##             :   age <= 0.5185185:
##             :   :...month in {apr,aug,dec,jul,jun,may,nov,
##             :       :         sep}: no (2757/406)
##             :       month in {mar,oct}:
##             :       :...job in {admin.,entrepreneur,housemaid,management,
##             :           :       retired,self-employed,services,student,
##             :           :       technician,unknown}: yes (12)
##             :           job in {blue-collar,unemployed}: no (2)
##             duration > 0.1309475:
##             :...duration > 0.1699878:
##                 :...contact = cellular:
##                 :   :...job in {admin.,blue-collar,retired,self-employed,
##                 :   :   :       student,unknown}: yes (386/145)
##                 :   :   job = housemaid:
##                 :   :   :...education in {basic.4y,high.school,illiterate,
##                 :   :   :   :             professional.course,
##                 :   :   :   :             unknown}: no (9/2)
##                 :   :   :   education in {basic.6y,basic.9y,
##                 :   :   :                 university.degree}: yes (9/2)
##                 :   :   job = services:
##                 :   :   :...day_of_week = fri: no (13/3)
##                 :   :   :   day_of_week in {mon,thu,tue,wed}: yes (45/12)
##                 :   :   job = entrepreneur:
##                 :   :   :...campaign <= 0: yes (8/1)
##                 :   :   :   campaign > 0:
##                 :   :   :   :...default in {no,yes}: no (15/4)
```

```
##                  :   :   :       default = unknown: yes (5/1)
##                  :   :   job = management:
##                  :   :   :...marital in {married,unknown}: yes (38/12)
##                  :   :   :   marital = single: no (9/3)
##                  :   :   :   marital = divorced:
##                  :   :   :   :...education = high.school: no (1)
##                  :   :   :       education in {basic.4y,basic.6y,basic.9y,
##                  :   :   :                     illiterate,professional.course,
##                  :   :   :                     university.degree,
##                  :   :   :                     unknown}: yes (4)
##                  :   :   job = technician:
##                  :   :   :...month in {dec,jun,mar,may,oct,sep}: yes (13/3)
##                  :   :   :   month = apr:
##                  :   :   :   :...campaign <= 0.01818182: yes (5)
##                  :   :   :   :   campaign > 0.01818182: no (2)
##                  :   :   :   month = aug:
##                  :   :   :   :...loan in {no,unknown}: yes (30/11)
##                  :   :   :   :   loan = yes: no (5)
##                  :   :   :   month = jul:
##                  :   :   :   :...day_of_week in {fri,thu,tue,wed}: no (24/8)
##                  :   :   :   :   day_of_week = mon: yes (5)
##                  :   :   :   month = nov:
##                  :   :   :   :...day_of_week in {fri,mon,tue}: yes (6/1)
##                  :   :   :       day_of_week in {thu,wed}: no (9/1)
##                  :   :   job = unemployed:
##                  :   :   :...day_of_week in {mon,tue,wed}: yes (6)
##                  :   :       day_of_week in {fri,thu}:
##                  :   :       :...loan in {no,unknown}: no (4)
##                  :   :           loan = yes: yes (1)
##                  :   contact = telephone:
##                  :   :...month in {apr,aug,dec,mar,oct,sep}: yes (7/1)
##                  :       month = jul:
##                  :       :...campaign <= 0.09090909: no (16/5)
##                  :       :   campaign > 0.09090909: yes (4)
##                  :       month = nov:
##                  :       :...duration <= 0.2350549: yes (4)
##                  :       :   duration > 0.2350549: no (4)
##                  :       month = jun:
##                  :       :...job = housemaid: no (3/1)
##                  :       :   job in {student,technician,unemployed,
##                  :       :   :       unknown}: yes (20/7)
##                  :       :   job = admin.:
##                  :       :   :...day_of_week in {fri,wed}: yes (8)
##                  :       :   :   day_of_week in {mon,thu,tue}: no (17/6)
##                  :       :   job = entrepreneur:
##                  :       :   :...duration <= 0.2248882: no (7/1)
##                  :       :   :   duration > 0.2248882: yes (4)
##                  :       :   job = management:
##                  :       :   :...age <= 0.3950617: yes (5)
##                  :       :   :   age > 0.3950617: no (3)
##                  :       :   job = retired:
##                  :       :   :...marital in {divorced,married,unknown}: yes (4)
##                  :       :   :   marital = single: no (1)
##                  :       :   job = self-employed:
##                  :       :   :...age <= 0.2469136: yes (2)
##                  :       :   :   age > 0.2469136: no (2)
```

```
##                   :       :   job = services:
##                   :       :   :...campaign <= 0.01818182: no (2)
##                   :       :   :   campaign > 0.01818182: yes (3)
##                   :       :   job = blue-collar:
##                   :       :   :...marital = unknown: yes (0)
##                   :       :       marital = divorced: no (3/1)
##                   :       :       marital = single:
##                   :       :       :...housing in {no,unknown}: no (5)
##                   :       :       :   housing = yes: yes (2)
##                   :       :       marital = married:
##                   :       :       :...euribor3m > 0.9698481: yes (21/5)
##                   :       :           euribor3m <= 0.9698481: [S1]
##                   :       month = may:
##                   :       :...education = illiterate: no (1)
##                   :           education = professional.course: yes (17/5)
##                   :           education = basic.6y:
##                   :           :...default in {no,yes}: yes (9/2)
##                   :           :   default = unknown: no (9/3)
##                   :           education = unknown:
##                   :           :...default in {unknown,yes}: yes (4)
##                   :           :   default = no:
##                   :           :   :...age <= 0.2839506: no (5)
##                   :           :       age > 0.2839506: yes (3)
##                   :           education = high.school:
##                   :           :...job in {entrepreneur,technician}: yes (2)
##                   :           :   job in {housemaid,management,retired,
##                   :           :   :       self-employed,student,unemployed,
##                   :           :   :       unknown}: no (6/3)
##                   :           :   job = blue-collar:
##                   :           :   :...marital in {divorced,married,
##                   :           :   :   :           unknown}: yes (4)
##                   :           :   :   marital = single: no (2)
##                   :           :   job = services:
##                   :           :   :...campaign <= 0.01818182: no (17/4)
##                   :           :   :   campaign > 0.01818182: yes (6/1)
##                   :           :   job = admin.:
##                   :           :   :...marital in {divorced,unknown}: no (0)
##                   :           :       marital = married:
##                   :           :       :...duration <= 0.249085: no (6)
##                   :           :       :   duration > 0.249085: yes (3/1)
##                   :           :       marital = single:
##                   :           :       :...default in {no,yes}: yes (7/1)
##                   :           :           default = unknown: no (1)
##                   :           education = basic.4y:
##                   :           :...euribor3m <= 0.9571525:
##                   :           :   :...job in {admin.,blue-collar,housemaid,
##                   :           :   :   :       management,retired,self-employed,
##                   :           :   :   :       student,technician,unemployed,
##                   :           :   :   :       unknown}: no (8)
##                   :           :   :   job in {entrepreneur,services}: yes (2)
##                   :           :   euribor3m > 0.9571525:
##                   :           :   :...day_of_week = tue: no (1)
##                   :           :       day_of_week = wed: yes (4)
##                   :           :       day_of_week = mon:
##                   :           :       :...campaign <= 0: yes (2)
##                   :           :       :   campaign > 0: no (2)
```

```
##                  :           :       day_of_week = thu:
##                  :           :       :...marital = divorced: yes (1)
##                  :           :       :   marital in {married,single,
##                  :           :       :               unknown}: no (4)
##                  :           :       day_of_week = fri:
##                  :           :       :...campaign > 0: yes (2)
##                  :           :           campaign <= 0:
##                  :           :           :...loan in {no,unknown}: no (3)
##                  :           :               loan = yes: yes (1)
##                  :           education = basic.9y:
##                  :           :...age > 0.3209876: no (12)
##                  :           :   age <= 0.3209876:
##                  :           :   :...duration > 0.2767385: yes (5)
##                  :           :       duration <= 0.2767385:
##                  :           :       :...housing in {unknown,yes}: no (10/3)
##                  :           :           housing = no:
##                  :           :           :...duration <= 0.2185848: yes (6/1)
##                  :           :               duration > 0.2185848: no (9/2)
##                  :           education = university.degree:
##                  :           :...duration > 0.3186255: yes (6)
##                  :               duration <= 0.3186255:
##                  :               :...euribor3m > 0.9578327: yes (9/3)
##                  :                   euribor3m <= 0.9578327:
##                  :                   :...age <= 0.3333333: no (28/6)
##                  :                       age > 0.3333333:
##                  :                       :...euribor3m <= 0.9571525: no (5/1)
##                  :                           euribor3m > 0.9571525: yes (5/1)
##                  duration <= 0.1699878:
##                  :...contact = telephone: no (337/91)
##                      contact = cellular:
##                      :...default in {unknown,yes}: no (111/36)
##                          default = no:
##                          :...euribor3m <= 0.1736568: yes (120/42)
##                              euribor3m > 0.1736568:
##                              :...euribor3m <= 0.1838585: no (38/5)
##                                  euribor3m > 0.1838585:
##                                  :...marital = unknown: no (1)
##                                      marital = divorced:
##                                      :...day_of_week = thu: no (8/4)
##                                      :   day_of_week = wed: yes (9)
##                                      :   day_of_week = fri:
##                                      :   :...cons.price.idx <= 0.4844115: yes (2)
##                                      :   :   cons.price.idx > 0.4844115: no (4)
##                                      :   day_of_week = mon: [S2]
##                                      :   day_of_week = tue:
##                                      :   :...age <= 0.3333333: yes (7)
##                                      :       age > 0.3333333: no (6/2)
##                                      marital = married:
##                                      :...job in {blue-collar,entrepreneur,
##                                      :   :       housemaid,retired,student,
##                                      :   :       unknown}: no (53/18)
##                                      :   job = self-employed:
##                                      :   :...loan in {no,unknown}: no (7/1)
##                                      :   :   loan = yes: yes (2)
##                                      :   job = services:
##                                      :   :...emp.var.rate <= 0.6875: no (5/1)
```

```
##                                      :   :   emp.var.rate > 0.6875: yes (13/2)
##                                      :   job = technician:
##                                      :   :...age <= 0.2962963: yes (23/10)
##                                      :   :   age > 0.2962963: no (8)
##                                      :   job = unemployed:
##                                      :   :...cons.price.idx <= 0.4844115: yes (2)
##                                      :   :   cons.price.idx > 0.4844115: no (3)
##                                      :   job = management:
##                                      :   :...day_of_week in {fri,
##                                      :   :   :               mon}: no (8/3)
##                                      :   :   day_of_week = wed: yes (3)
##                                      :   :   day_of_week = thu:
##                                      :   :   :...emp.var.rate <= 0.6875: yes (2)
##                                      :   :   :   emp.var.rate > 0.6875: no (3)
##                                      :   :   day_of_week = tue:
##                                      :   :   :...emp.var.rate <= 0.6875: no (3)
##                                      :   :       emp.var.rate > 0.6875: yes (2)
##                                      :   job = admin.:
##                                      :   :...cons.price.idx > 0.4844115: yes (16/3)
##                                      :       cons.price.idx <= 0.4844115:
##                                      :       :...duration <= 0.1612444: no (25/5)
##                                      :           duration > 0.1612444: [S3]
##                                      marital = single:
##                                      :...campaign > 0.05454545: no (8)
##                                          campaign <= 0.05454545:
##                                          :...campaign > 0.03636364: yes (5)
##                                              campaign <= 0.03636364: [S4]
##
## SubTree [S1]
##
## education in {basic.4y,basic.6y,high.school,illiterate,professional.course,
## :              university.degree,unknown}: no (6/1)
## education = basic.9y:
## :...housing in {no,unknown}: yes (3)
##     housing = yes: no (1)
##
## SubTree [S2]
##
## education in {basic.4y,basic.6y,basic.9y,high.school,illiterate,
## :              professional.course,unknown}: no (3)
## education = university.degree: yes (1)
##
## SubTree [S3]
##
## loan in {no,unknown}: yes (3)
## loan = yes: no (1)
##
## SubTree [S4]
##
## job in {retired,self-employed,unemployed,unknown}: no (0)
## job in {entrepreneur,housemaid,student}: yes (5/1)
## job = admin.:
## :...campaign <= 0: no (13/3)
## :   campaign > 0: yes (9/2)
## job = blue-collar:
## :...duration <= 0.1667344: no (8)
```

```
## :   duration > 0.1667344: yes (2)
## job = management:
## :...day_of_week = thu: no (3/1)
## :   day_of_week in {fri,mon,tue,wed}: yes (2)
## job = services:
## :...day_of_week in {fri,mon}: yes (3)
## :   day_of_week in {thu,tue,wed}: no (6/1)
## job = technician:
## :...loan in {unknown,yes}: yes (4)
##     loan = no:
##     :...education in {basic.4y,basic.6y,basic.9y,high.school,illiterate,
##         :              professional.course,unknown}: no (10/3)
##         education = university.degree:
##         :...housing = no: no (2)
##             housing in {unknown,yes}: yes (3)
##
##
## Evaluation on training data (30891 cases):
##
##      Decision Tree
##    ----------------
##    Size      Errors
##
##     284 1990( 6.4%)   <<
##
##
##     (a)   (b)    <-classified as
##    ----  ----
##   26717   694    (a): class no
##    1296  2184    (b): class yes
##
##
##  Attribute usage:
##
##  100.00% duration
##  100.00% poutcome
##   99.22% nr.employed
##   83.36% month
##   81.80% age
##   10.55% contact
##    6.48% euribor3m
##    4.88% job
##    4.01% education
##    2.70% cons.price.idx
##    2.41% default
##    2.30% loan
##    2.24% campaign
##    2.07% marital
##    2.05% emp.var.rate
##    1.50% day_of_week
##    0.78% cons.conf.idx
##    0.39% housing
##    0.32% pdays
##
##
## Time: 0.2 secs
```

```r
# Applying model to the training set which classifies all the observations in training set
# as either "yes" or "no"
tree_pred <- predict(decision_tree, test_data[-21])
str(tree_pred)
```

```
##  Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
```

```r
# Evaluating model performance by comparing the predicted variable with true labels

# The model has an accuracy of 91.32% and p-value is < 2.2e-16 which implies that the model
# is performing well. The kappa value is 0.533 which indicates a moderate agreement between
# true and predicted values. The sensitivity and specificity of the model are 0.53 and 0.96,
# which are the false negative and false positive rates respectively
CrossTable(tree_pred, test_data$y)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## | Chi-square contribution |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  10297
##
##
##              | test_data$y
##    tree_pred |        no |       yes | Row Total |
## -------------|-----------|-----------|-----------|
##           no |      8781 |       538 |      9319 |
##              |    31.680 |   249.531 |           |
##              |     0.942 |     0.058 |     0.905 |
##              |     0.961 |     0.464 |           |
##              |     0.853 |     0.052 |           |
## -------------|-----------|-----------|-----------|
##          yes |       356 |       622 |       978 |
##              |   301.863 |  2377.692 |           |
##              |     0.364 |     0.636 |     0.095 |
##              |     0.039 |     0.536 |           |
##              |     0.035 |     0.060 |           |
## -------------|-----------|-----------|-----------|
## Column Total |      9137 |      1160 |     10297 |
##              |     0.887 |     0.113 |           |
## -------------|-----------|-----------|-----------|
##
##
```

```r
confusionMatrix(tree_pred,test_data$y, positive = "yes")
```

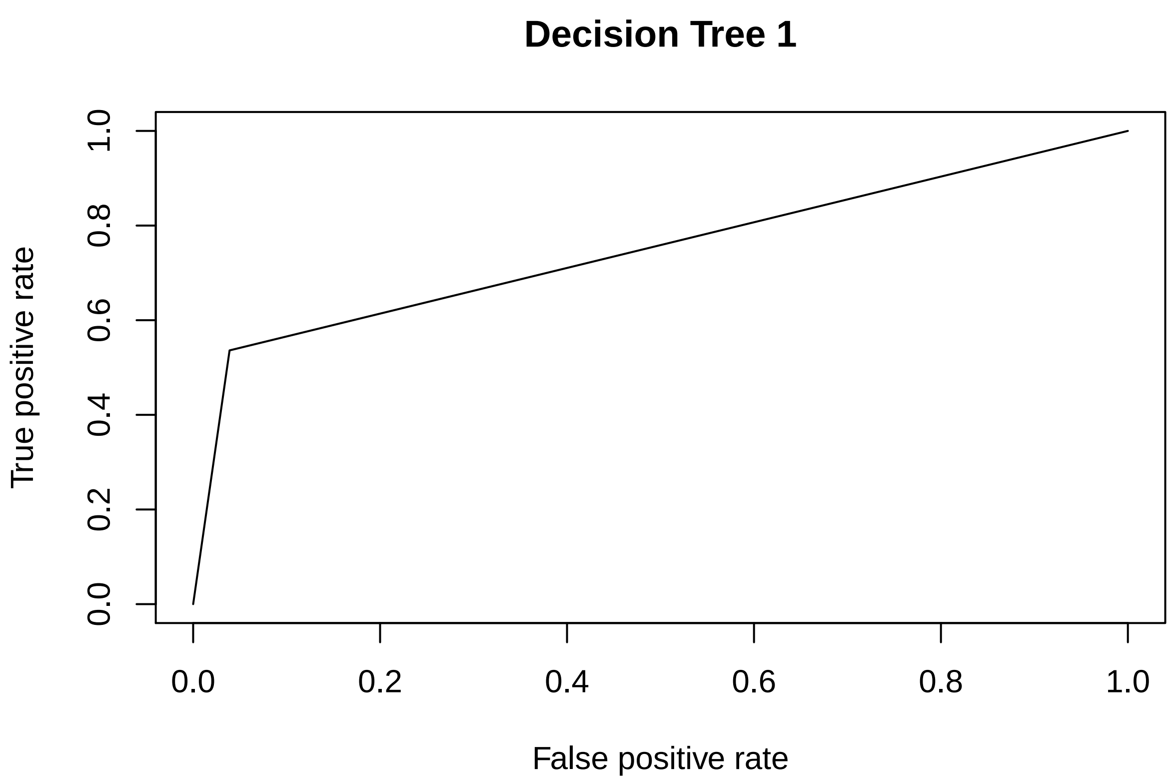
```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   no  yes
##         no  8781  538
```

```
##        yes  356  622
##
##                Accuracy : 0.9132
##                  95% CI : (0.9076, 0.9185)
##     No Information Rate : 0.8873
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.5338
##  Mcnemar's Test P-Value : 1.417e-09
##
##             Sensitivity : 0.53621
##             Specificity : 0.96104
##          Pos Pred Value : 0.63599
##          Neg Pred Value : 0.94227
##              Prevalence : 0.11265
##          Detection Rate : 0.06041
##    Detection Prevalence : 0.09498
##       Balanced Accuracy : 0.74862
##
##        'Positive' Class : yes
##
```

```
dt_accuracy <- confusionMatrix(tree_pred,test_data$y, positive = "yes")$overall[[1]]
dt_accuracy <- dt_accuracy*100
dt_accuracy
```

```
## [1] 91.31786
```

```
# The ROC curve suggests that the model is doing a fine job in predicting the true negatives
# and false positives with a area under the curve value of 0.74
pred_dt <- prediction(predictions = as.numeric(tree_pred), labels = as.numeric(test_data$y))
perf_dt <- performance(pred_dt,measure = "tpr", x.measure = "fpr")
plot(perf_dt, main="Decision Tree 1")
```



```r
perf.auc <- performance(pred_dt, measure = "auc")
dt_auc <- unlist(perf.auc@y.values)
dt_auc
```

```
## [1] 0.7486222
```

```r
# We can improve performance of the model in classifying FP and FN and thereby increasing
# the value of auc by assigning cost or penalty for the making a FP or FN mistake. Here we
# assignmed more penalty for FN than FP because it is better to make few extra calls for
# the people who won't actually take the plan, than classifying the person who would
# actually take the subscription as "no" and avoid contacting him altogether

# Creating cost error matrix
matrix_dimensions <- list(c("no", "yes"), c("no", "yes"))
names(matrix_dimensions) <- c("predicted", "actual")
matrix_dimensions
```

```
## $predicted
## [1] "no"  "yes"
##
## $actual
## [1] "no"  "yes"
```

```r
error_cost <- matrix(c(0, 1, 5, 0), nrow = 2, dimnames = matrix_dimensions)
error_cost
```

```
##          actual
## predicted no yes
##       no   0   5
##       yes  1   0
```

```r
# Training a decision tree model using the error_cost matrix
set.seed(141)
decision_tree_2 <- C5.0(train_data[-21], train_data$y, trails=20, costs=error_cost)
summary(decision_tree)
```

```
##
## Call:
## C5.0.default(x = train_data[-21], y = train_data$y, trails = 20)
##
##
## C5.0 [Release 2.07 GPL Edition]      Sun Dec 10 17:34:15 2017
## -------------------------------
##
## Class specified by attribute `outcome'
##
## Read 30891 cases (21 attributes) from undefined.data
##
## Decision tree:
##
## poutcome = success:
## :...duration <= 0.03273689:
## :   :...cons.conf.idx <= 0.376569: no (69/3)
## :   :   cons.conf.idx > 0.376569:
## :   :   :...campaign > 0.01818182:
## :   :       :...marital = divorced: yes (3/1)
## :   :       :   marital in {married,single,unknown}: no (25/1)
## :   :       campaign <= 0.01818182:
## :   :       :...month in {jun,oct}: no (39/13)
```

```
## :   :           month = apr:
## :   :           :...housing in {no,unknown}: no (3)
## :   :           :   housing = yes: yes (1)
## :   :           month = dec:
## :   :           :...duration <= 0.02806019: no (3)
## :   :           :   duration > 0.02806019: yes (2)
## :   :           month = jul:
## :   :           :...duration <= 0.02704351: no (6)
## :   :           :   duration > 0.02704351: yes (2)
## :   :           month = mar:
## :   :           :...housing in {no,unknown}: yes (5)
## :   :           :   housing = yes: no (5/1)
## :   :           month = may:
## :   :           :...duration <= 0.02704351: no (3)
## :   :           :   duration > 0.02704351: yes (3)
## :   :           month = nov:
## :   :           :...euribor3m <= 0.01836318: yes (14/1)
## :   :           :   euribor3m > 0.01836318: no (2)
## :   :           month = sep:
## :   :           :...housing = no: yes (4)
## :   :           :   housing in {unknown,yes}: no (11/4)
## :   :           month = aug:
## :   :           :...age <= 0.3580247:
## :   :               :...euribor3m <= 0.04919519: yes (7/2)
## :   :               :   euribor3m > 0.04919519: no (18/2)
## :   :               age > 0.3580247:
## :   :               :...housing in {no,unknown}: no (3)
## :   :                   housing = yes: yes (14/6)
## :   duration > 0.03273689:
## :   :...nr.employed <= 0.4257089: yes (649/116)
## :       nr.employed > 0.4257089:
## :       :...month = mar: yes (2)
## :           month in {aug,dec,jul,jun,nov,oct,sep}: no (15/2)
## :           month = apr:
## :           :...pdays <= 0.007007007: yes (30/7)
## :           :   pdays > 0.007007007: no (6/1)
## :           month = may:
## :           :...job in {entrepreneur,housemaid,management,retired,
## :               :       self-employed,student,unknown}: no (10)
## :               job in {technician,unemployed}: yes (15/4)
## :               job = admin.:
## :               :...day_of_week in {fri,wed}: yes (6/1)
## :               :   day_of_week in {mon,tue}: no (7)
## :               :   day_of_week = thu:
## :               :   :...housing = no: yes (1)
## :               :       housing in {unknown,yes}: no (3)
## :               job = services:
## :               :...education in {basic.4y,basic.6y,basic.9y,illiterate,
## :               :   :             professional.course,unknown}: yes (3)
## :               :   education = university.degree: no (1)
## :               :   education = high.school:
## :               :   :...housing = no: yes (2)
## :               :       housing in {unknown,yes}: no (2)
## :               job = blue-collar:
## :               :...duration <= 0.09109394: no (15/1)
## :                   duration > 0.09109394:
```

```
## :                    :...education in {basic.4y,basic.6y,high.school,illiterate,
## :                        :             professional.course,university.degree,
## :                        :             unknown}: yes (5)
## :                        education = basic.9y:
## :                        :...day_of_week = thu: yes (2)
## :                            day_of_week in {fri,mon,tue,wed}: no (4)
## poutcome in {failure,nonexistent}:
## :...duration <= 0.07991053:
##     :...nr.employed <= 0.4257089:
##     :   :...duration <= 0.03497357: no (1213/163)
##     :   :   duration > 0.03497357:
##     :   :   :...contact = telephone: no (150/48)
##     :   :       contact = cellular:
##     :   :       :...duration > 0.050427:
##     :   :           :...cons.price.idx <= 0.06936867:
##     :   :           :   :...day_of_week in {fri,thu,tue}: no (59/18)
##     :   :           :   :   day_of_week in {mon,wed}: yes (32/14)
##     :   :           :   cons.price.idx > 0.06936867:
##     :   :           :   :...education in {basic.6y,basic.9y,high.school,
##     :   :           :       :             illiterate,
##     :   :           :       :             university.degree}: yes (288/101)
##     :   :           :       education = professional.course:
##     :   :           :       :...age <= 0.2098765: yes (21/5)
##     :   :           :       :   age > 0.2098765: no (27/9)
##     :   :           :       education = basic.4y:
##     :   :           :       :...poutcome = nonexistent: yes (23/6)
##     :   :           :       :   poutcome = failure:
##     :   :           :       :   :...month in {apr,aug,dec,jul,mar,may,oct,
##     :   :           :       :       :         sep}: no (8)
##     :   :           :       :       month in {jun,nov}: yes (3)
##     :   :           :       education = unknown:
##     :   :           :       :...emp.var.rate > 0.3541667: no (4)
##     :   :           :           emp.var.rate <= 0.3541667:
##     :   :           :           :...job in {admin.,blue-collar,entrepreneur,
##     :   :           :               :       housemaid,management,self-employed,
##     :   :           :               :       student,unemployed,
##     :   :           :               :       unknown}: yes (12)
##     :   :           :               job in {retired,services,
##     :   :           :                       technician}: no (12/5)
##     :   :           duration <= 0.050427:
##     :   :           :...education in {basic.4y,high.school,illiterate,
##     :   :               :             unknown}: no (179/62)
##     :   :               education = basic.6y:
##     :   :               :...marital in {divorced,single,unknown}: yes (5)
##     :   :               :   marital = married:
##     :   :               :   :...age <= 0.3333333: no (4)
##     :   :               :       age > 0.3333333: yes (2)
##     :   :               education = basic.9y:
##     :   :               :...day_of_week in {mon,wed}: no (16/4)
##     :   :               :   day_of_week in {thu,tue}: yes (7/1)
##     :   :               :   day_of_week = fri:
##     :   :               :   :...job in {admin.,blue-collar,entrepreneur,
##     :   :               :       :       housemaid,management,self-employed,
##     :   :               :       :       student,technician,unemployed,
##     :   :               :       :       unknown}: no (4)
##     :   :               :       job in {retired,services}: yes (4)
```

```
##      :   :               education = professional.course:
##      :   :               :...job in {admin.,entrepreneur,housemaid,management,
##      :   :               :   :       services,student,unknown}: no (7/2)
##      :   :               :   job in {blue-collar,self-employed,
##      :   :               :   :       unemployed}: yes (9/1)
##      :   :               :   job = retired:
##      :   :               :   :...campaign <= 0: yes (4)
##      :   :               :   :   campaign > 0: no (3)
##      :   :               :   job = technician:
##      :   :               :   :...loan = no: no (26/8)
##      :   :               :       loan in {unknown,yes}: yes (5/1)
##      :   :               education = university.degree:
##      :   :               :...poutcome = failure:
##      :   :                   :...pdays <= 0.006006006: yes (4)
##      :   :                   :   pdays > 0.006006006: no (58/16)
##      :   :                   poutcome = nonexistent:
##      :   :                   :...job in {blue-collar,entrepreneur,management,
##      :   :                       :       technician,unknown}: no (40/14)
##      :   :                       job in {housemaid,services,
##      :   :                       :       unemployed}: yes (4)
##      :   :                       job = retired:
##      :   :                       :...emp.var.rate <= 0: no (2)
##      :   :                       :   emp.var.rate > 0: yes (2)
##      :   :                       job = self-employed:
##      :   :                       :...marital in {divorced,married}: yes (4)
##      :   :                       :   marital in {single,unknown}: no (6/1)
##      :   :                       job = student:
##      :   :                       :...duration <= 0.04209028: yes (3)
##      :   :                       :   duration > 0.04209028: no (2)
##      :   :                       job = admin.:
##      :   :                       :...day_of_week = fri: yes (8/3)
##      :   :                           day_of_week = wed: no (12/5)
##      :   :                           day_of_week = mon:
##      :   :                           :...campaign <= 0: no (7)
##      :   :                           :   campaign > 0: yes (7/3)
##      :   :                           day_of_week = thu:
##      :   :                           :...campaign <= 0: yes (10/3)
##      :   :                           :   campaign > 0: no (4)
##      :   :                           day_of_week = tue:
##      :   :                           :...nr.employed <= 0.2037807: yes (12)
##      :   :                               nr.employed > 0.2037807:
##      :   :                               :...duration <= 0.03924359: yes (2)
##      :   :                                   duration > 0.03924359: no (6)
##      :   nr.employed > 0.4257089:
##      :   :...age > 0.5308642:
##      :       :...duration <= 0.02765352: no (52/5)
##      :       :   duration > 0.02765352:
##      :       :   :...euribor3m > 0.1736568: yes (42/14)
##      :       :       euribor3m <= 0.1736568:
##      :       :       :...marital in {married,single,unknown}: no (9)
##      :       :           marital = divorced:
##      :       :           :...housing in {no,unknown}: no (2)
##      :       :               housing = yes: yes (2)
##      :       age <= 0.5308642:
##      :       :...month in {aug,jul,jun,may,nov,sep}: no (20581/110)
##      :           month in {apr,dec}:
```

```
##     :           :...euribor3m > 0.1736568: no (1167/99)
##     :           :   euribor3m <= 0.1736568:
##     :           :   :...duration <= 0.0351769: no (86/4)
##     :           :       duration > 0.0351769:
##     :           :       :...education in {basic.6y,illiterate,
##     :           :           :             professional.course,
##     :           :           :             university.degree,
##     :           :           :             unknown}: yes (46/18)
##     :           :           education in {basic.9y,high.school}: no (24/8)
##     :           :           education = basic.4y:
##     :           :           :...age <= 0.1111111: no (2)
##     :           :               age > 0.1111111: yes (3)
##     :           month in {mar,oct}:
##     :           :...duration <= 0.01911346: no (59/5)
##     :               duration > 0.01911346:
##     :               :...campaign > 0.05454545:
##     :                   :...duration <= 0.05002033: no (11)
##     :                   :   duration > 0.05002033: yes (2)
##     :                   campaign <= 0.05454545:
##     :                   :...month = oct:
##     :                       :...loan in {no,unknown}: yes (27/2)
##     :                       :   loan = yes: no (1)
##     :                       month = mar:
##     :                       :...default = yes: yes (0)
##     :                           default = unknown:
##     :                           :...job in {admin.,blue-collar,entrepreneur,
##     :                           :   :       housemaid,management,retired,
##     :                           :   :       self-employed,services,student,
##     :                           :   :       technician,unknown}: no (3)
##     :                           :   job = unemployed: yes (1)
##     :                           default = no:
##     :                           :...duration > 0.03761692: yes (49/10)
##     :                               duration <= 0.03761692:
##     :                               :...marital = divorced: yes (1)
##     :                                   marital in {single,
##     :                                   :           unknown}: no (35/11)
##     :                                   marital = married:
##     :                                   :...housing in {no,
##     :                                       :           unknown}: no (9/3)
##     :                                       housing = yes: yes (13/4)
##     duration > 0.07991053:
##     :...nr.employed <= 0.4257089:
##         :...emp.var.rate > 0.1041667:
##         :   :...loan in {unknown,yes}: yes (42/6)
##         :   :   loan = no:
##         :   :   :...poutcome = nonexistent: yes (125/26)
##         :   :       poutcome = failure:
##         :   :       :...emp.var.rate <= 0.3541667:
##         :   :           :...campaign <= 0.05454545: yes (40/8)
##         :   :           :   campaign > 0.05454545: no (3)
##         :   :           emp.var.rate > 0.3541667:
##         :   :           :...marital = divorced: yes (4)
##         :   :               marital in {married,single,unknown}: no (17/5)
##         :   emp.var.rate <= 0.1041667:
##         :   :...loan = unknown: yes (12/4)
##         :       loan = yes:
```

```
##          :       :...marital in {divorced,single,unknown}: no (11/1)
##          :       :   marital = married:
##          :       :   :...month in {aug,nov}: yes (8/1)
##          :       :       month in {apr,jul,mar,may,oct}: no (4)
##          :       :       month = dec:
##          :       :       :...duration <= 0.101464: yes (2)
##          :       :       :   duration > 0.101464: no (2)
##          :       :       month = jun:
##          :       :       :...day_of_week in {mon,thu}: yes (3)
##          :       :       :   day_of_week in {fri,tue,wed}: no (5/1)
##          :       :       month = sep:
##          :       :       :...duration <= 0.1122407: yes (2)
##          :       :           duration > 0.1122407: no (4)
##          :       loan = no:
##          :       :...cons.price.idx > 0.1745908: yes (87/23)
##          :           cons.price.idx <= 0.1745908:
##          :           :...contact = telephone:
##          :               :...marital in {divorced,single}: yes (11/4)
##          :               :   marital in {married,unknown}: no (15/2)
##          :               contact = cellular:
##          :               :...duration <= 0.09577064: yes (64/14)
##          :                   duration > 0.09577064:
##          :                   :...campaign > 0.05454545: yes (6)
##          :                       campaign <= 0.05454545:
##          :                       :...job in {housemaid,management,
##          :                           :       student}: yes (21/6)
##          :                           job in {self-employed,services,technician,
##          :                           :       unemployed,unknown}: no (27/9)
##          :                           job = blue-collar:
##          :                           :...euribor3m <= 0.02221718: yes (2)
##          :                           :   euribor3m > 0.02221718: no (3)
##          :                           job = entrepreneur:
##          :                           :...duration <= 0.1268808: yes (2)
##          :                           :   duration > 0.1268808: no (2)
##          :                           job = retired:
##          :                           :...cons.price.idx <= 0.1044427: no (17/5)
##          :                           :   cons.price.idx > 0.1044427: yes (3)
##          :                           job = admin.:
##          :                           :...month in {apr,dec,jul,jun,mar,may,nov,
##          :                               :         oct}: no (16/3)
##          :                               month = sep: yes (4)
##          :                               month = aug:
##          :                               :...day_of_week in {fri,
##          :                                   :               wed}: yes (5/1)
##          :                                   day_of_week in {mon,thu,
##          :                                                   tue}: no (5/1)
##          nr.employed > 0.4257089:
##          :...duration <= 0.1309475:
##              :...age > 0.5185185:
##              :   :...month in {apr,jul}: yes (12/1)
##              :   :   month in {aug,dec,jun,mar,may,nov,oct,sep}: no (19/3)
##              :   age <= 0.5185185:
##              :   :...month in {apr,aug,dec,jul,jun,may,nov,
##              :       :         sep}: no (2757/406)
##              :       month in {mar,oct}:
##              :       :...job in {admin.,entrepreneur,housemaid,management,
```

```
##             :           :       retired,self-employed,services,student,
##             :           :       technician,unknown}: yes (12)
##             :           job in {blue-collar,unemployed}: no (2)
##             duration > 0.1309475:
##             :...duration > 0.1699878:
##                 :...contact = cellular:
##                 :   :...job in {admin.,blue-collar,retired,self-employed,
##                 :   :   :       student,unknown}: yes (386/145)
##                 :   :   job = housemaid:
##                 :   :   :...education in {basic.4y,high.school,illiterate,
##                 :   :   :   :             professional.course,
##                 :   :   :   :             unknown}: no (9/2)
##                 :   :   :   education in {basic.6y,basic.9y,
##                 :   :   :                 university.degree}: yes (9/2)
##                 :   :   job = services:
##                 :   :   :...day_of_week = fri: no (13/3)
##                 :   :   :   day_of_week in {mon,thu,tue,wed}: yes (45/12)
##                 :   :   job = entrepreneur:
##                 :   :   :...campaign <= 0: yes (8/1)
##                 :   :   :   campaign > 0:
##                 :   :   :   :...default in {no,yes}: no (15/4)
##                 :   :   :       default = unknown: yes (5/1)
##                 :   :   job = management:
##                 :   :   :...marital in {married,unknown}: yes (38/12)
##                 :   :   :   marital = single: no (9/3)
##                 :   :   :   marital = divorced:
##                 :   :   :   :...education = high.school: no (1)
##                 :   :   :       education in {basic.4y,basic.6y,basic.9y,
##                 :   :   :                     illiterate,professional.course,
##                 :   :   :                     university.degree,
##                 :   :   :                     unknown}: yes (4)
##                 :   :   job = technician:
##                 :   :   :...month in {dec,jun,mar,may,oct,sep}: yes (13/3)
##                 :   :   :   month = apr:
##                 :   :   :   :...campaign <= 0.01818182: yes (5)
##                 :   :   :   :   campaign > 0.01818182: no (2)
##                 :   :   :   month = aug:
##                 :   :   :   :...loan in {no,unknown}: yes (30/11)
##                 :   :   :   :   loan = yes: no (5)
##                 :   :   :   month = jul:
##                 :   :   :   :...day_of_week in {fri,thu,tue,wed}: no (24/8)
##                 :   :   :   :   day_of_week = mon: yes (5)
##                 :   :   :   month = nov:
##                 :   :   :   :...day_of_week in {fri,mon,tue}: yes (6/1)
##                 :   :   :       day_of_week in {thu,wed}: no (9/1)
##                 :   :   job = unemployed:
##                 :   :   :...day_of_week in {mon,tue,wed}: yes (6)
##                 :   :       day_of_week in {fri,thu}:
##                 :   :       :...loan in {no,unknown}: no (4)
##                 :   :           loan = yes: yes (1)
##                 :   contact = telephone:
##                 :   :...month in {apr,aug,dec,mar,oct,sep}: yes (7/1)
##                 :       month = jul:
##                 :       :...campaign <= 0.09090909: no (16/5)
##                 :       :   campaign > 0.09090909: yes (4)
##                 :       month = nov:
```

```
##                   :       :...duration <= 0.2350549: yes (4)
##                   :       :   duration > 0.2350549: no (4)
##                   :       month = jun:
##                   :       :...job = housemaid: no (3/1)
##                   :       :   job in {student,technician,unemployed,
##                   :       :   :       unknown}: yes (20/7)
##                   :       :   job = admin.:
##                   :       :   :...day_of_week in {fri,wed}: yes (8)
##                   :       :   :   day_of_week in {mon,thu,tue}: no (17/6)
##                   :       :   job = entrepreneur:
##                   :       :   :...duration <= 0.2248882: no (7/1)
##                   :       :   :   duration > 0.2248882: yes (4)
##                   :       :   job = management:
##                   :       :   :...age <= 0.3950617: yes (5)
##                   :       :   :   age > 0.3950617: no (3)
##                   :       :   job = retired:
##                   :       :   :...marital in {divorced,married,unknown}: yes (4)
##                   :       :   :   marital = single: no (1)
##                   :       :   job = self-employed:
##                   :       :   :...age <= 0.2469136: yes (2)
##                   :       :   :   age > 0.2469136: no (2)
##                   :       :   job = services:
##                   :       :   :...campaign <= 0.01818182: no (2)
##                   :       :   :   campaign > 0.01818182: yes (3)
##                   :       :   job = blue-collar:
##                   :       :   :...marital = unknown: yes (0)
##                   :       :       marital = divorced: no (3/1)
##                   :       :       marital = single:
##                   :       :       :...housing in {no,unknown}: no (5)
##                   :       :       :   housing = yes: yes (2)
##                   :       :       marital = married:
##                   :       :       :...euribor3m > 0.9698481: yes (21/5)
##                   :       :           euribor3m <= 0.9698481: [S1]
##                   :       month = may:
##                   :       :...education = illiterate: no (1)
##                   :           education = professional.course: yes (17/5)
##                   :           education = basic.6y:
##                   :           :...default in {no,yes}: yes (9/2)
##                   :           :   default = unknown: no (9/3)
##                   :           education = unknown:
##                   :           :...default in {unknown,yes}: yes (4)
##                   :           :   default = no:
##                   :           :   :...age <= 0.2839506: no (5)
##                   :           :       age > 0.2839506: yes (3)
##                   :           education = high.school:
##                   :           :...job in {entrepreneur,technician}: yes (2)
##                   :           :   job in {housemaid,management,retired,
##                   :           :   :       self-employed,student,unemployed,
##                   :           :   :       unknown}: no (6/3)
##                   :           :   job = blue-collar:
##                   :           :   :...marital in {divorced,married,
##                   :           :   :   :           unknown}: yes (4)
##                   :           :   :   marital = single: no (2)
##                   :           :   job = services:
##                   :           :   :...campaign <= 0.01818182: no (17/4)
##                   :           :   :   campaign > 0.01818182: yes (6/1)
```

```
##                 :           :   job = admin.:
##                 :           :   :...marital in {divorced,unknown}: no (0)
##                 :           :       marital = married:
##                 :           :       :...duration <= 0.249085: no (6)
##                 :           :       :   duration > 0.249085: yes (3/1)
##                 :           :       marital = single:
##                 :           :       :...default in {no,yes}: yes (7/1)
##                 :           :           default = unknown: no (1)
##                 :           education = basic.4y:
##                 :           :...euribor3m <= 0.9571525:
##                 :           :   :...job in {admin.,blue-collar,housemaid,
##                 :           :   :   :       management,retired,self-employed,
##                 :           :   :   :       student,technician,unemployed,
##                 :           :   :   :       unknown}: no (8)
##                 :           :   :   job in {entrepreneur,services}: yes (2)
##                 :           :   euribor3m > 0.9571525:
##                 :           :   :...day_of_week = tue: no (1)
##                 :           :       day_of_week = wed: yes (4)
##                 :           :       day_of_week = mon:
##                 :           :       :...campaign <= 0: yes (2)
##                 :           :       :   campaign > 0: no (2)
##                 :           :       day_of_week = thu:
##                 :           :       :...marital = divorced: yes (1)
##                 :           :       :   marital in {married,single,
##                 :           :       :               unknown}: no (4)
##                 :           :       day_of_week = fri:
##                 :           :       :...campaign > 0: yes (2)
##                 :           :           campaign <= 0:
##                 :           :           :...loan in {no,unknown}: no (3)
##                 :           :               loan = yes: yes (1)
##                 :           education = basic.9y:
##                 :           :...age > 0.3209876: no (12)
##                 :           :   age <= 0.3209876:
##                 :           :   :...duration > 0.2767385: yes (5)
##                 :           :       duration <= 0.2767385:
##                 :           :       :...housing in {unknown,yes}: no (10/3)
##                 :           :           housing = no:
##                 :           :           :...duration <= 0.2185848: yes (6/1)
##                 :           :               duration > 0.2185848: no (9/2)
##                 :           education = university.degree:
##                 :           :...duration > 0.3186255: yes (6)
##                 :               duration <= 0.3186255:
##                 :               :...euribor3m > 0.9578327: yes (9/3)
##                 :                   euribor3m <= 0.9578327:
##                 :                   :...age <= 0.3333333: no (28/6)
##                 :                       age > 0.3333333:
##                 :                       :...euribor3m <= 0.9571525: no (5/1)
##                 :                           euribor3m > 0.9571525: yes (5/1)
##                 duration <= 0.1699878:
##                 :...contact = telephone: no (337/91)
##                     contact = cellular:
##                     :...default in {unknown,yes}: no (111/36)
##                         default = no:
##                         :...euribor3m <= 0.1736568: yes (120/42)
##                             euribor3m > 0.1736568:
##                             :...euribor3m <= 0.1838585: no (38/5)
```

```
##                                  euribor3m > 0.1838585:
##                                  :...marital = unknown: no (1)
##                                      marital = divorced:
##                                      :...day_of_week = thu: no (8/4)
##                                      :   day_of_week = wed: yes (9)
##                                      :   day_of_week = fri:
##                                      :   :...cons.price.idx <= 0.4844115: yes (2)
##                                      :   :   cons.price.idx > 0.4844115: no (4)
##                                      :   day_of_week = mon: [S2]
##                                      :   day_of_week = tue:
##                                      :   :...age <= 0.3333333: yes (7)
##                                      :       age > 0.3333333: no (6/2)
##                                      marital = married:
##                                      :...job in {blue-collar,entrepreneur,
##                                      :   :       housemaid,retired,student,
##                                      :   :       unknown}: no (53/18)
##                                      :   job = self-employed:
##                                      :   :...loan in {no,unknown}: no (7/1)
##                                      :   :   loan = yes: yes (2)
##                                      :   job = services:
##                                      :   :...emp.var.rate <= 0.6875: no (5/1)
##                                      :   :   emp.var.rate > 0.6875: yes (13/2)
##                                      :   job = technician:
##                                      :   :...age <= 0.2962963: yes (23/10)
##                                      :   :   age > 0.2962963: no (8)
##                                      :   job = unemployed:
##                                      :   :...cons.price.idx <= 0.4844115: yes (2)
##                                      :   :   cons.price.idx > 0.4844115: no (3)
##                                      :   job = management:
##                                      :   :...day_of_week in {fri,
##                                      :   :   :               mon}: no (8/3)
##                                      :   :   day_of_week = wed: yes (3)
##                                      :   :   day_of_week = thu:
##                                      :   :   :...emp.var.rate <= 0.6875: yes (2)
##                                      :   :   :   emp.var.rate > 0.6875: no (3)
##                                      :   :   day_of_week = tue:
##                                      :   :   :...emp.var.rate <= 0.6875: no (3)
##                                      :   :       emp.var.rate > 0.6875: yes (2)
##                                      :   job = admin.:
##                                      :   :...cons.price.idx > 0.4844115: yes (16/3)
##                                      :       cons.price.idx <= 0.4844115:
##                                      :       :...duration <= 0.1612444: no (25/5)
##                                      :           duration > 0.1612444: [S3]
##                                      marital = single:
##                                      :...campaign > 0.05454545: no (8)
##                                          campaign <= 0.05454545:
##                                          :...campaign > 0.03636364: yes (5)
##                                              campaign <= 0.03636364: [S4]
##
## SubTree [S1]
##
## education in {basic.4y,basic.6y,high.school,illiterate,professional.course,
## :              university.degree,unknown}: no (6/1)
## education = basic.9y:
## :...housing in {no,unknown}: yes (3)
##     housing = yes: no (1)
```

```
##
## SubTree [S2]
##
## education in {basic.4y,basic.6y,basic.9y,high.school,illiterate,
## :              professional.course,unknown}: no (3)
## education = university.degree: yes (1)
##
## SubTree [S3]
##
## loan in {no,unknown}: yes (3)
## loan = yes: no (1)
##
## SubTree [S4]
##
## job in {retired,self-employed,unemployed,unknown}: no (0)
## job in {entrepreneur,housemaid,student}: yes (5/1)
## job = admin.:
## :...campaign <= 0: no (13/3)
## :   campaign > 0: yes (9/2)
## job = blue-collar:
## :...duration <= 0.1667344: no (8)
## :   duration > 0.1667344: yes (2)
## job = management:
## :...day_of_week = thu: no (3/1)
## :   day_of_week in {fri,mon,tue,wed}: yes (2)
## job = services:
## :...day_of_week in {fri,mon}: yes (3)
## :   day_of_week in {thu,tue,wed}: no (6/1)
## job = technician:
## :...loan in {unknown,yes}: yes (4)
##     loan = no:
##     :...education in {basic.4y,basic.6y,basic.9y,high.school,illiterate,
##         :              professional.course,unknown}: no (10/3)
##         education = university.degree:
##         :...housing = no: no (2)
##             housing in {unknown,yes}: yes (3)
##
##
## Evaluation on training data (30891 cases):
##
##      Decision Tree
##    ----------------
##    Size      Errors
##
##     284 1990( 6.4%)   <<
##
##
##     (a)   (b)    <-classified as
##    ----  ----
##   26717   694    (a): class no
##    1296  2184    (b): class yes
##
##
##  Attribute usage:
##
##  100.00% duration
```

```
##  100.00% poutcome
##   99.22% nr.employed
##   83.36% month
##   81.80% age
##   10.55% contact
##    6.48% euribor3m
##    4.88% job
##    4.01% education
##    2.70% cons.price.idx
##    2.41% default
##    2.30% loan
##    2.24% campaign
##    2.07% marital
##    2.05% emp.var.rate
##    1.50% day_of_week
##    0.78% cons.conf.idx
##    0.39% housing
##    0.32% pdays
##
##
## Time: 0.2 secs
```

```r
# Applying the new model to test set
tree_pred_2 <- predict(decision_tree_2, test_data[-21])
str(tree_pred_2)
```

```
##  Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
```

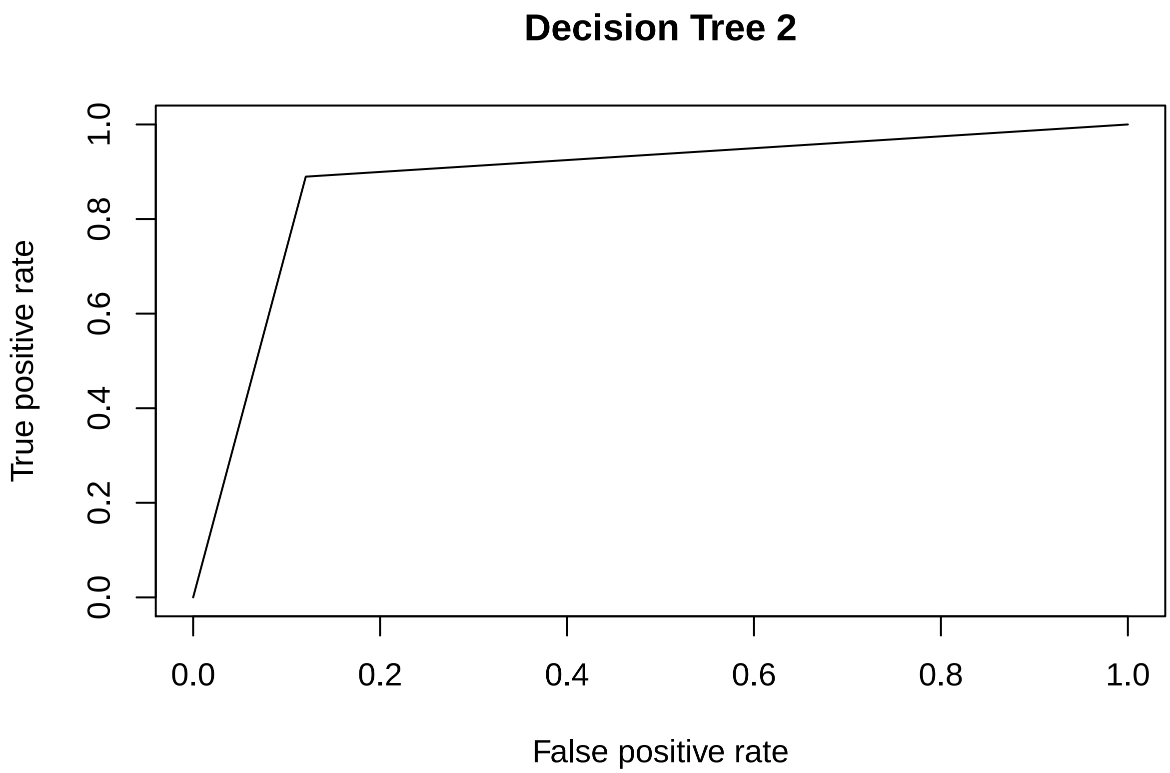
```r
# The accuracy of the model was slightly reduced to 88.06%. But the value of sensitivity
# is very significantly improved
confusionMatrix(tree_pred_2,test_data$y, positive = "yes")
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   no  yes
##        no  8036  128
##        yes 1101 1032
##
##                Accuracy : 0.8806
##                  95% CI : (0.8742, 0.8868)
##     No Information Rate : 0.8873
##     P-Value [Acc > NIR] : 0.9843
##
##                   Kappa : 0.563
##  Mcnemar's Test P-Value : <2e-16
##
##             Sensitivity : 0.8897
##             Specificity : 0.8795
##          Pos Pred Value : 0.4838
##          Neg Pred Value : 0.9843
##              Prevalence : 0.1127
##          Detection Rate : 0.1002
##    Detection Prevalence : 0.2071
##       Balanced Accuracy : 0.8846
##
##        'Positive' Class : yes
##
```

```r
dt2_accuracy <- confusionMatrix(tree_pred_2,test_data$y, positive = "yes")$overall[[1]]
dt2_accuracy <- dt2_accuracy*100
dt2_accuracy
```

```
## [1] 88.06448
```

```r
# The ROC curve also reflects improvement in the prediction of true positives with auc of 0.8845
pred_dt_2 <- prediction(predictions = as.numeric(tree_pred_2), labels = as.numeric(test_data$y))
perf_dt_2 <- performance(pred_dt_2,measure = "tpr", x.measure = "fpr")
plot(perf_dt_2, main="Decision Tree 2")
```




```r
perf.auc_dt_2 <- performance(pred_dt_2, measure = "auc")
dt2_auc <- unlist(perf.auc_dt_2@y.values)
dt2_auc
```

```
## [1] 0.8845781
```

```r
### Neural Networks ###

library(nnet)

set.seed(141)
# creating a neural network model using training set
nnet_model <- nnet(y~age + job + marital + education +
                       default + housing + loan + contact +
                       month + day_of_week + duration + campaign +
                       pdays + previous + poutcome + emp.var.rate +
                       cons.price.idx + cons.conf.idx + euribor3m +
                       nr.employed, data=train_data, size=9, decay=0.1)
```

```
## # weights:  496
## initial  value 31333.957381
## iter  10 value 9011.110826
## iter  20 value 7669.544914
```

```
## iter  30 value 6741.491465
## iter  40 value 6413.159454
## iter  50 value 6191.163433
## iter  60 value 5943.655987
## iter  70 value 5764.836959
## iter  80 value 5671.171255
## iter  90 value 5595.733635
## iter 100 value 5531.377260
## final  value 5531.377260
## stopped after 100 iterations
```

```
# The model has 53 input nodes, 9 hidden nodes and 1 output node
nnet_model$n
```

```
## [1] 53  9  1
```

```
# applying the neural network model to test set
nnet_pred <- predict(nnet_model, test_data[-21], type="class")
str(nnet_pred)
```

```
##  chr [1:10297] "no" "no" "no" "no" "no" "no" "no" "no" "no" "no" "no" ...
```

```
# The accuracy of the model is around 91% which is very good. And the kappa value is
# indicating a moderate agreement between predicted and true values.
CrossTable(nnet_pred, test_data$y)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## | Chi-square contribution |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  10297
##
##
##              | test_data$y
##    nnet_pred |        no |       yes | Row Total |
## -------------|-----------|-----------|-----------|
##           no |      8830 |       571 |      9401 |
##              |    28.555 |   224.920 |           |
##              |     0.939 |     0.061 |     0.913 |
##              |     0.966 |     0.492 |           |
##              |     0.858 |     0.055 |           |
## -------------|-----------|-----------|-----------|
##          yes |       307 |       589 |       896 |
##              |   299.605 |  2359.905 |           |
##              |     0.343 |     0.657 |     0.087 |
##              |     0.034 |     0.508 |           |
##              |     0.030 |     0.057 |           |
## -------------|-----------|-----------|-----------|
## Column Total |      9137 |      1160 |     10297 |
##              |     0.887 |     0.113 |           |
## -------------|-----------|-----------|-----------|
```

```
##
##
```

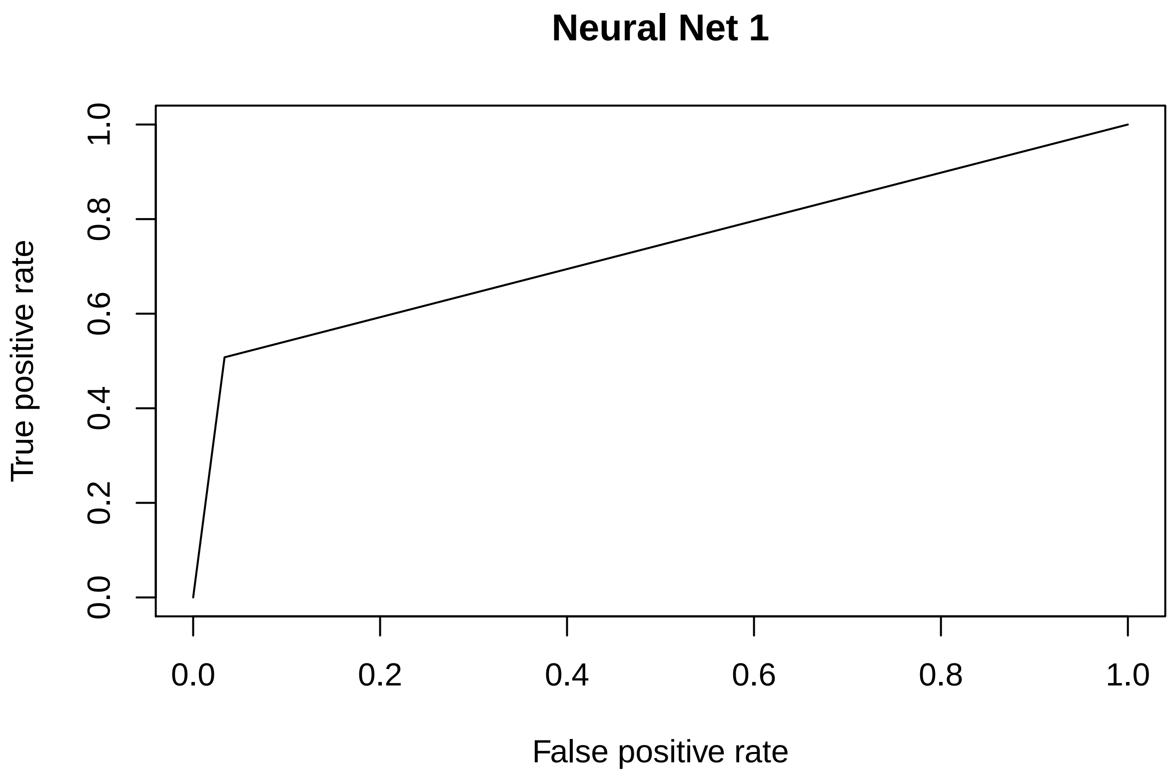
```r
confusionMatrix(nnet_pred,test_data$y, positive = "yes")
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   no  yes
##        no  8830  571
##        yes  307  589
##
##                Accuracy : 0.9147
##                  95% CI : (0.9092, 0.9201)
##     No Information Rate : 0.8873
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.5265
##  Mcnemar's Test P-Value : < 2.2e-16
##
##             Sensitivity : 0.50776
##             Specificity : 0.96640
##          Pos Pred Value : 0.65737
##          Neg Pred Value : 0.93926
##              Prevalence : 0.11265
##          Detection Rate : 0.05720
##    Detection Prevalence : 0.08702
##       Balanced Accuracy : 0.73708
##
##        'Positive' Class : yes
##
```

```r
nnet_accuracy <- confusionMatrix(nnet_pred,test_data$y, positive = "yes")$overall[[1]]
nn_accuracy <- nnet_accuracy*100
nn_accuracy
```

```
## [1] 91.47324
```

```r
# Building ROC curve and calculating AUC of the predicted and true values indicating the
# relationship between true positive rate and false positive rate.
nnet_pred_fac <- as.factor(nnet_pred)
pred_nn <- prediction(predictions = as.numeric(nnet_pred_fac), labels = as.numeric(test_data$y))
perf_nn <- performance(pred_nn,measure = "tpr", x.measure = "fpr")
plot(perf_nn, main="Neural Net 1")
```




```
perf.auc_nn <- performance(pred_nn, measure = "auc")
nn_auc <- unlist(perf.auc_nn@y.values)
nn_auc
```

```
## [1] 0.7370795
```

```
# Lets try to improve the model peformance by using the function pcaNNet which applies
# principal component analysis to the variables before building a neural network model.
set.seed(141)
nnet_model_2 <- pcaNNet(y~age + job + marital + education +
            default + housing + loan + contact +
            month + day_of_week + duration + campaign +
            pdays + previous + poutcome + emp.var.rate +
            cons.price.idx + cons.conf.idx + euribor3m +
            nr.employed, data=train_data, size=8, decay=0.1)
```

```
## # weights:  386
## initial  value 21417.939395
## iter  10 value 7230.132920
## iter  20 value 6255.473857
## iter  30 value 5607.495900
## iter  40 value 4768.984514
## iter  50 value 3981.698700
## iter  60 value 3767.468181
## iter  70 value 3662.719354
## iter  80 value 3578.469377
## iter  90 value 3525.404344
## iter 100 value 3480.951714
## final  value 3480.951714
## stopped after 100 iterations
```

```
# predicting the target variable of the training set using the model
nnet_pred_2 <- predict(nnet_model_2, test_data[,-21], type="class")
str(nnet_pred_2)
```

```
##  chr [1:10297] "no" "no" "no" "no" "no" "no" "no" "no" "no" "no" "no" ...
```

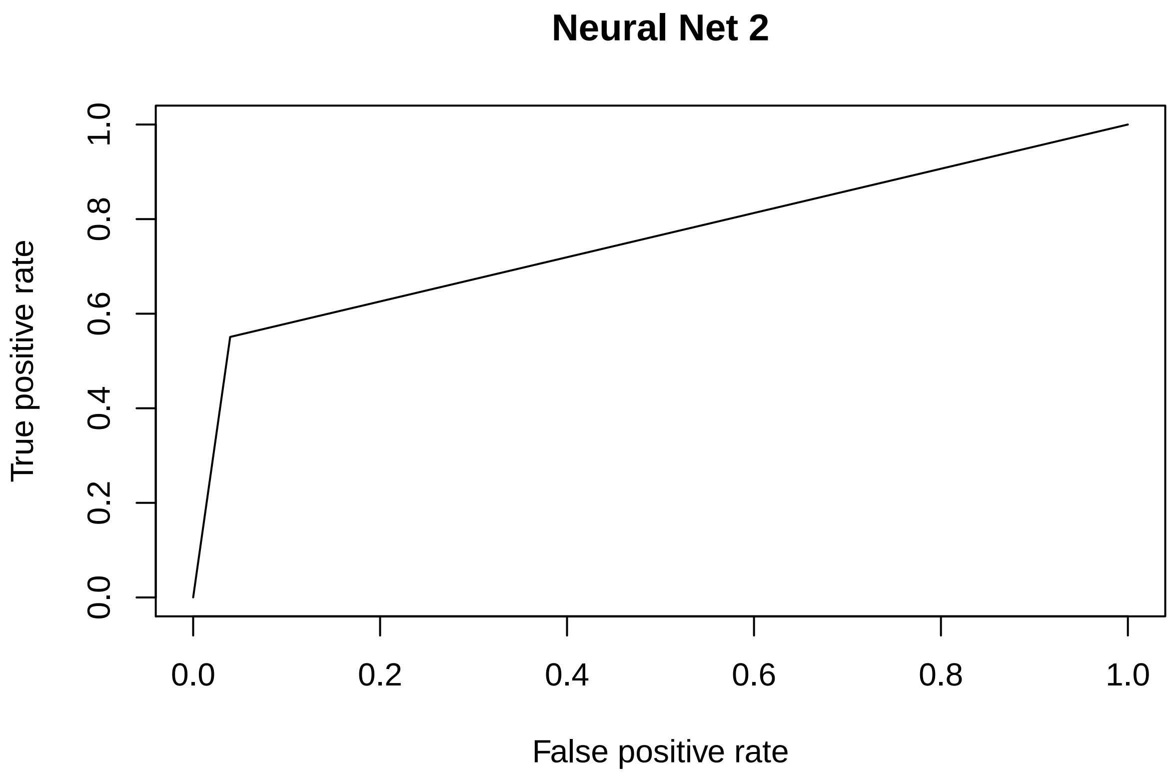
```r
# The sensitivity of the model fairly increased but it is still less efficient compared to
# the decision tree model
confusionMatrix(nnet_pred_2,test_data$y, positive = "yes")
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   no  yes
##        no  8775  521
##        yes  362  639
##
##                Accuracy : 0.9142
##                  95% CI : (0.9087, 0.9196)
##     No Information Rate : 0.8873
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.5438
##  Mcnemar's Test P-Value : 1.054e-07
##
##             Sensitivity : 0.55086
##             Specificity : 0.96038
##          Pos Pred Value : 0.63836
##          Neg Pred Value : 0.94395
##              Prevalence : 0.11265
##          Detection Rate : 0.06206
##    Detection Prevalence : 0.09721
##       Balanced Accuracy : 0.75562
##
##        'Positive' Class : yes
##
```

```r
nn2_accuracy <- confusionMatrix(nnet_pred_2,test_data$y, positive = "yes")$overall[[1]]
nn2_accuracy <- nn2_accuracy*100
nn2_accuracy
```

```
## [1] 91.42469
```

```r
# Plotting the ROC curve using the true and predicted values of target variable and
# computing area under the ROC curve
nnet_pred_fac_2 <- as.factor(nnet_pred_2)
pred_nn_2 <- prediction(predictions = as.numeric(nnet_pred_fac_2), labels = as.numeric(test_data$y))
perf_nn_2 <- performance(pred_nn_2,measure = "tpr", x.measure = "fpr")
plot(perf_nn_2, main="Neural Net 2")
```




```
perf.auc_nn_2 <- performance(pred_nn_2, measure = "auc")
nn2_auc <- unlist(perf.auc_nn_2@y.values)
nn2_auc
```

```
## [1] 0.7556215
```

```
### Support Vector Machine ###

library(kernlab)
```

```
##
## Attaching package: 'kernlab'
```

```
## The following object is masked from 'package:ggplot2':
##
##     alpha
```

```
# Building a SVM model using training set
set.seed(141)
svm_model <- ksvm(y~., data=train_data, kernel = "rbfdot", C=9)
svm_model
```

```
## Support Vector Machine object of class "ksvm"
##
## SV type: C-svc  (classification)
##  parameter : cost C = 9
##
## Gaussian Radial Basis kernel function.
##  Hyperparameter : sigma =  0.055949851456535
##
## Number of Support Vectors : 6414
##
## Objective Function Value : -38609.8
## Training error : 0.051471
```

```r
# Predicting the target variable by supplying test data for the model
svm_pred <- predict(svm_model, test_data[-21])
str(svm_pred)
```

```
##  Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
```

```r
# The accuracy of the SVM model is around 91% and a kappa value of 0.46
CrossTable(svm_pred, test_data$y)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## | Chi-square contribution |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  10297
##
##
##              | test_data$y
##     svm_pred |        no |       yes | Row Total |
## -------------|-----------|-----------|-----------|
##           no |      8845 |       620 |      9465 |
##              |    23.713 |   186.780 |           |
##              |     0.934 |     0.066 |     0.919 |
##              |     0.968 |     0.534 |           |
##              |     0.859 |     0.060 |           |
## -------------|-----------|-----------|-----------|
##          yes |       292 |       540 |       832 |
##              |   269.763 |  2124.849 |           |
##              |     0.351 |     0.649 |     0.081 |
##              |     0.032 |     0.466 |           |
##              |     0.028 |     0.052 |           |
## -------------|-----------|-----------|-----------|
## Column Total |      9137 |      1160 |     10297 |
##              |     0.887 |     0.113 |           |
## -------------|-----------|-----------|-----------|
##
##
```

```r
confusionMatrix(svm_pred,test_data$y, positive = "yes")
```

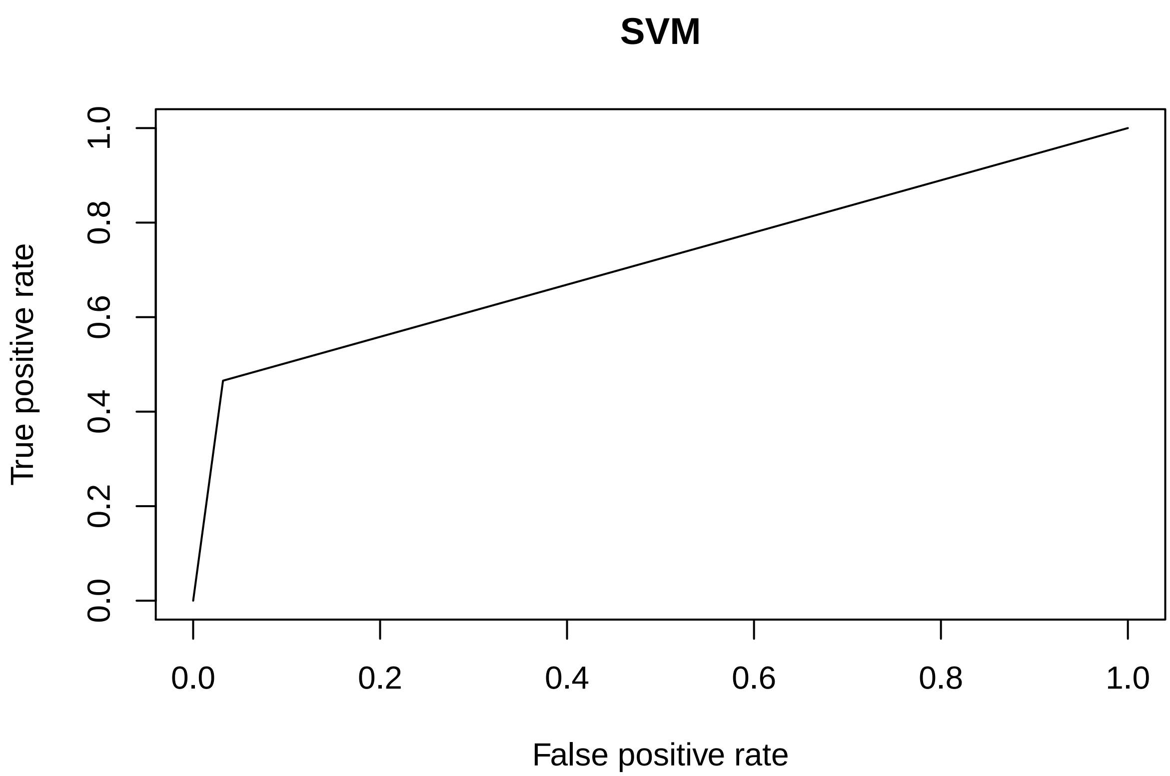
```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   no  yes
##        no  8845  620
##        yes  292  540
##
##                Accuracy : 0.9114
##                  95% CI : (0.9058, 0.9168)
##     No Information Rate : 0.8873
##     P-Value [Acc > NIR] : 7.62e-16
```

```
##
##                  Kappa : 0.4946
##  Mcnemar's Test P-Value : < 2.2e-16
##
##            Sensitivity : 0.46552
##            Specificity : 0.96804
##         Pos Pred Value : 0.64904
##         Neg Pred Value : 0.93450
##             Prevalence : 0.11265
##         Detection Rate : 0.05244
##   Detection Prevalence : 0.08080
##      Balanced Accuracy : 0.71678
##
##       'Positive' Class : yes
##
```

```r
svm_accuracy <- confusionMatrix(svm_pred,test_data$y, positive = "yes")$overall[[1]]
svm_accuracy <- svm_accuracy*100
svm_accuracy
```

```
## [1] 91.14305
```

```r
# ROC curve and AUC
pred_svm <- prediction(predictions = as.numeric(svm_pred), labels = as.numeric(test_data$y))
perf_svm <- performance(pred_svm,measure = "tpr", x.measure = "fpr")
plot(perf_svm, main="SVM")
```




```r
perf.auc_svm <- performance(pred_svm, measure = "auc")
svm_auc <- unlist(perf.auc_svm@y.values)
svm_auc
```

```
## [1] 0.7167796
```

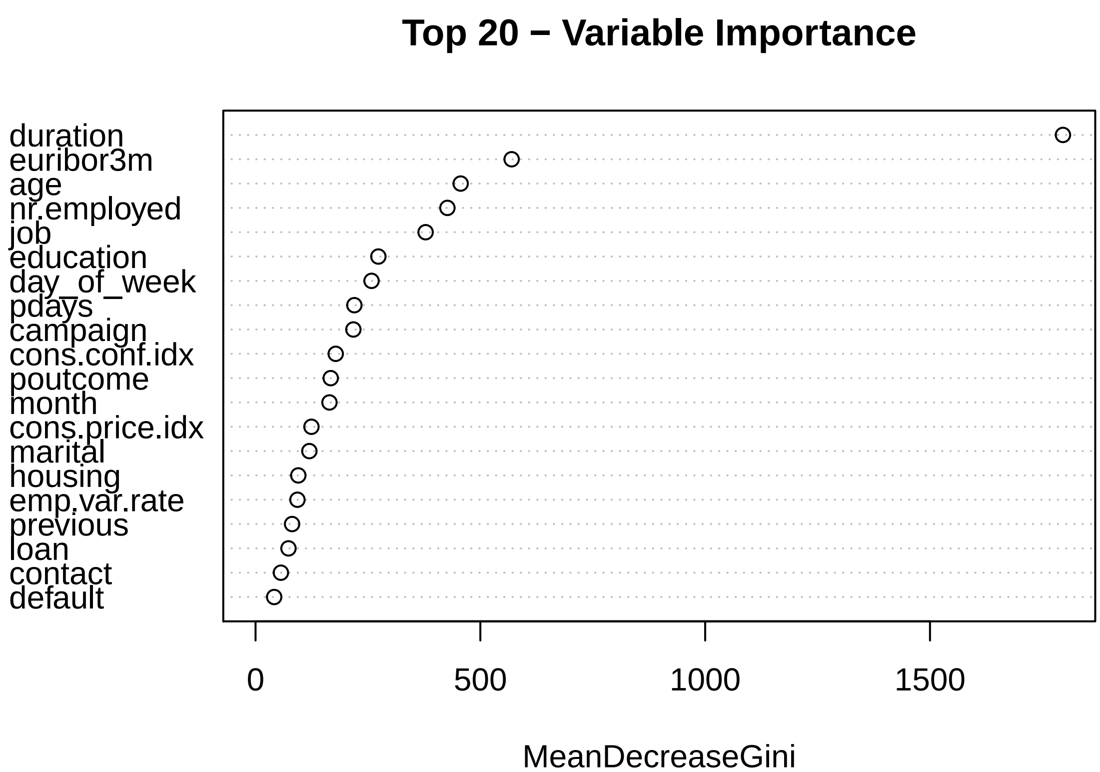
```r
### Random Forest ###

library(randomForest)
```

```
## randomForest 4.6-12

## Type rfNews() to see new features/changes/bug fixes.

##
## Attaching package: 'randomForest'

## The following object is masked from 'package:ggplot2':
##
##     margin
```

```r
# Building a random forest model using training data
set.seed(141)
rf_model <- randomForest(y~., ntree=80, data=train_data)
rf_model
```

```
##
## Call:
##  randomForest(formula = y ~ ., data = train_data, ntree = 80)
##                Type of random forest: classification
##                      Number of trees: 80
## No. of variables tried at each split: 4
##
##         OOB estimate of  error rate: 8.7%
## Confusion matrix:
##        no  yes class.error
## no  26371 1040  0.03794097
## yes  1648 1832  0.47356322
```

```r
# Generating a variable importance graph using the random forest model we built. The
# variables with high Gini index are the most important variables. So, the variables
# at the top of the y-axis are more important in building a model than variables at
# the bottom
varImpPlot(rf_model,
           sort = T,
           n.var=20,
           main="Top 20 - Variable Importance")
```




```
# Applying our model to test data for predicting the target variable y
rf_pred <- predict(rf_model, test_data[-21])
str(rf_pred)
```

```
##  Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
##  - attr(*, "names")= chr [1:10297] "1" "6" "10" "12" ...
```

```
# The accuracy of the model is around 91%
CrossTable(rf_pred, test_data$y)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## | Chi-square contribution |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  10297
##
##
##              | test_data$y
##      rf_pred |        no |       yes | Row Total |
## -------------|-----------|-----------|-----------|
##           no |      8810 |       547 |      9357 |
##              |    30.972 |   243.956 |           |
##              |     0.942 |     0.058 |     0.909 |
##              |     0.964 |     0.472 |           |
##              |     0.856 |     0.053 |           |
```

```
## -------------|-----------|-----------|-----------|
##          yes |       327 |       613 |       940 |
##              |   308.301 |  2428.403 |           |
##              |     0.348 |     0.652 |     0.091 |
##              |     0.036 |     0.528 |           |
##              |     0.032 |     0.060 |           |
## -------------|-----------|-----------|-----------|
## Column Total |      9137 |      1160 |     10297 |
##              |     0.887 |     0.113 |           |
## -------------|-----------|-----------|-----------|
##
##
```

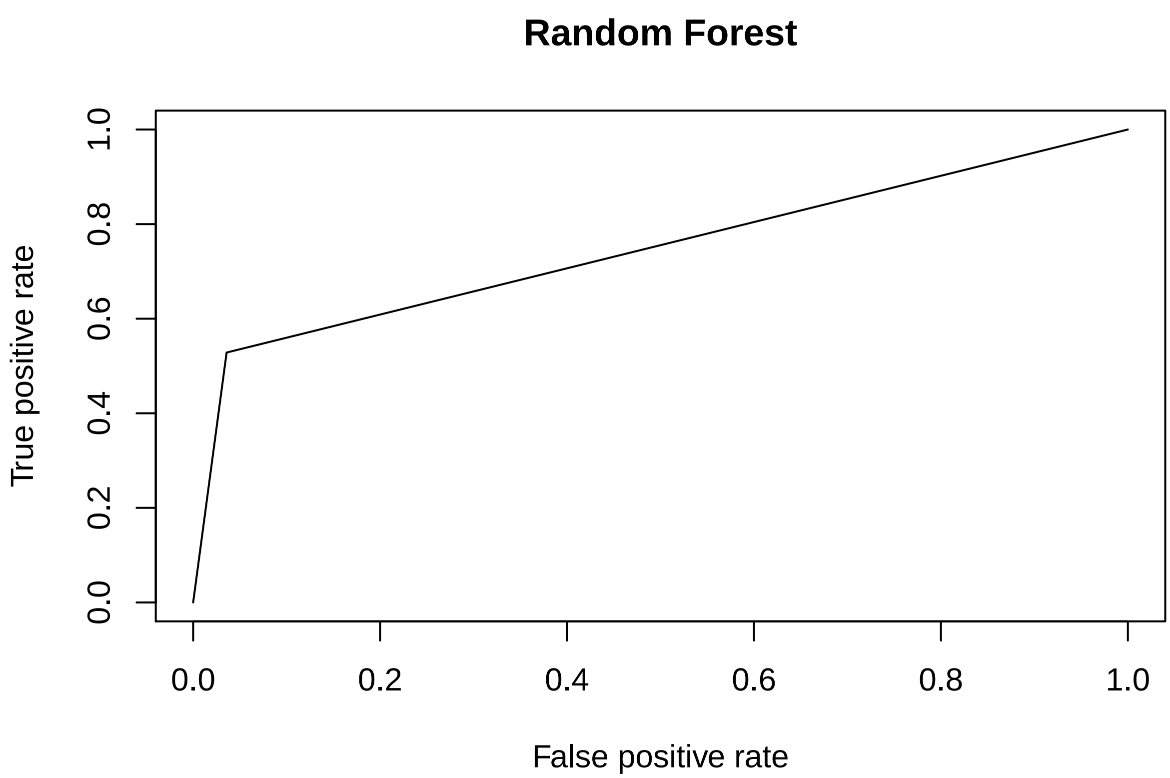
```r
confusionMatrix(rf_pred,test_data$y, positive = "yes")
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   no  yes
##        no  8810  547
##        yes  327  613
##
##                Accuracy : 0.9151
##                  95% CI : (0.9096, 0.9204)
##     No Information Rate : 0.8873
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.5371
##  Mcnemar's Test P-Value : 1.284e-13
##
##             Sensitivity : 0.52845
##             Specificity : 0.96421
##          Pos Pred Value : 0.65213
##          Neg Pred Value : 0.94154
##              Prevalence : 0.11265
##          Detection Rate : 0.05953
##    Detection Prevalence : 0.09129
##       Balanced Accuracy : 0.74633
##
##        'Positive' Class : yes
##
```

```r
rf_accuracy <- confusionMatrix(rf_pred,test_data$y, positive = "yes")$overall[[1]]
rf_accuracy <- rf_accuracy*100
rf_accuracy
```

```
## [1] 91.51209
```

```r
# ROC curve and AUC for the Random Forest model
pred_rf <- prediction(predictions = as.numeric(rf_pred), labels = as.numeric(test_data$y))
perf_rf <- performance(pred_rf,measure = "tpr", x.measure = "fpr")
plot(perf_rf, main="Random Forest")
```




```
perf.auc_rf <- performance(pred_rf, measure = "auc")
rf_auc <- unlist(perf.auc_rf@y.values)
rf_auc
```

```
## [1] 0.7463299
```

```
### Comparision of models ###

# Creating a matrix containing accuracy and AUC values of all the models
compare <- matrix(c(nb_accuracy, nb_auc, dt_accuracy, dt_auc, dt2_accuracy, dt2_auc,
                    nn_accuracy, nn_auc, nn2_accuracy, nn2_auc,
                    svm_accuracy, svm_auc, rf_accuracy, rf_auc),ncol=2, byrow = T)

compare <- as.data.frame(compare)

rownames(compare) <- c("NaiveBayes", "DecisionTree1", "DecisionTree2", "NeuralNetwork1",
                       "NeuralNetwork2", "SVM", "RandomForest")

names(compare) <- c("Accuracy", "AUC")
compare
```

```
##                Accuracy       AUC
## NaiveBayes     86.36496 0.7560870
## DecisionTree1  91.31786 0.7486222
## DecisionTree2  88.06448 0.8845781
## NeuralNetwork1 91.47324 0.7370795
## NeuralNetwork2 91.42469 0.7556215
## SVM            91.14305 0.7167796
## RandomForest   91.51209 0.7463299
```

```
# Adding a third column for the comparision matrix which serves as final evaluation metric
compare$evaluation <- (compare$Accuracy^2)*(compare$AUC)
compare
```

```
##                Accuracy       AUC evaluation
## NaiveBayes     86.36496 0.7560870   5639.582
## DecisionTree1  91.31786 0.7486222   6242.724
## DecisionTree2  88.06448 0.8845781   6860.215
## NeuralNetwork1 91.47324 0.7370795   6167.405
## NeuralNetwork2 91.42469 0.7556215   6315.842
## SVM            91.14305 0.7167796   5954.328
## RandomForest   91.51209 0.7463299   6250.112
```

```r
### If FN and FP rates are considered significant in addition to the accuracy, we need to
# select the model with highest evaluation metric. The 2nd Decision Tree model is a clear
# standout in this case ###
compare[order(compare$evaluation, decreasing = T), ]
```

```
##                Accuracy       AUC evaluation
## DecisionTree2  88.06448 0.8845781   6860.215
## NeuralNetwork2 91.42469 0.7556215   6315.842
## RandomForest   91.51209 0.7463299   6250.112
## DecisionTree1  91.31786 0.7486222   6242.724
## NeuralNetwork1 91.47324 0.7370795   6167.405
## SVM            91.14305 0.7167796   5954.328
## NaiveBayes     86.36496 0.7560870   5639.582
```

```r
### If model accuracy is the only metric to be considered, then we can select any of
# Random Forest, Neural network 1, Decision tree 1 or Support Vector Machines, as all
# these models have almost the same accuracy ###
compare[order(compare$Accuracy, decreasing = T), ]
```

```
##                Accuracy       AUC evaluation
## RandomForest   91.51209 0.7463299   6250.112
## NeuralNetwork1 91.47324 0.7370795   6167.405
## NeuralNetwork2 91.42469 0.7556215   6315.842
## DecisionTree1  91.31786 0.7486222   6242.724
## SVM            91.14305 0.7167796   5954.328
## DecisionTree2  88.06448 0.8845781   6860.215
## NaiveBayes     86.36496 0.7560870   5639.582
```

```r
### K-fold cross validation ###


# Performing K-fold cross validation on the selected models to get a better estimation
# of its future performance

# Random Forest #
# 10-fold cross validation of the random forest model
set.seed(141)
folds <- createFolds(bank$y, k=10)
cv_results <- lapply(folds, function (x) {
    bank_train <- bank[-x, ]
    bank_test <- bank[x, ]
    bank_model <- randomForest(y~., ntree=80, data=bank_train)
    bank_predict <- predict(bank_model, bank_test[-21])
    accuracy <- confusionMatrix(bank_predict, bank_test$y, positive = "yes")$overall[[1]]
    accuracy <- accuracy*100
    return(accuracy)
})

# The average accuracy of our random forest models is 91.5% which is impressive
cv_rf <- mean(unlist(cv_results))
```

```r
cv_rf
```

```
## [1] 91.55336
```

```r
# Neural network #
# 10-fold cross validation of the Neural network 1 model #
set.seed(141)
folds <- createFolds(bank$y, k=10)
cv_results <- lapply(folds, function (x) {
    bank_train <- bank[-x, ]
    bank_test <- bank[x, ]
    bank_model <- nnet(y~age + job + marital + education +
                           default + housing + loan + contact +
                           month + day_of_week + duration + campaign +
                           pdays + previous + poutcome + emp.var.rate +
                           cons.price.idx + cons.conf.idx + euribor3m +
                           nr.employed,data=bank_train, size=9, decay=0.1)
    bank_predict <- predict(bank_model, bank_test[-21], type="class")
    accuracy <- confusionMatrix(bank_predict, bank_test$y, positive = "yes")$overall[[1]]
    accuracy <- accuracy*100
    return(accuracy)
})
```

```
## # weights:  496
## initial  value 22089.048723
## iter  10 value 10022.227098
## iter  20 value 8867.879442
## iter  30 value 7838.038323
## iter  40 value 7412.422302
## iter  50 value 7053.834068
## iter  60 value 6950.034387
## iter  70 value 6852.040777
## iter  80 value 6779.366270
## iter  90 value 6734.139892
## iter 100 value 6691.217632
## final  value 6691.217632
## stopped after 100 iterations
## # weights:  496
## initial  value 14851.744068
## iter  10 value 9530.122508
## iter  20 value 7591.611898
## iter  30 value 7108.830103
## iter  40 value 6868.484648
## iter  50 value 6758.667989
## iter  60 value 6666.797042
## iter  70 value 6619.283937
## iter  80 value 6574.173241
## iter  90 value 6516.668431
## iter 100 value 6469.026437
## final  value 6469.026437
## stopped after 100 iterations
## # weights:  496
## initial  value 21008.214412
## iter  10 value 9991.519945
## iter  20 value 8218.147097
## iter  30 value 7358.687938
## iter  40 value 7153.273406
```

```
## iter  50 value 6999.118583
## iter  60 value 6871.425526
## iter  70 value 6775.439317
## iter  80 value 6711.209502
## iter  90 value 6667.747996
## iter 100 value 6629.596193
## final  value 6629.596193
## stopped after 100 iterations
## # weights:  496
## initial  value 20710.977832
## iter  10 value 9913.273557
## iter  20 value 8084.479475
## iter  30 value 7515.789067
## iter  40 value 7184.729051
## iter  50 value 6996.784704
## iter  60 value 6860.558637
## iter  70 value 6775.431046
## iter  80 value 6707.920893
## iter  90 value 6657.175065
## iter 100 value 6603.203579
## final  value 6603.203579
## stopped after 100 iterations
## # weights:  496
## initial  value 21505.172399
## iter  10 value 9871.647541
## iter  20 value 9040.838701
## iter  30 value 8223.511222
## iter  40 value 7649.727016
## iter  50 value 7312.263282
## iter  60 value 7170.727621
## iter  70 value 6983.606220
## iter  80 value 6876.192534
## iter  90 value 6807.789574
## iter 100 value 6756.335308
## final  value 6756.335308
## stopped after 100 iterations
## # weights:  496
## initial  value 23453.238456
## iter  10 value 10682.643341
## iter  20 value 9807.973937
## iter  30 value 8632.047387
## iter  40 value 8145.555106
## iter  50 value 7854.271070
## iter  60 value 7594.317945
## iter  70 value 7391.997000
## iter  80 value 7210.124825
## iter  90 value 7091.610037
## iter 100 value 7022.800105
## final  value 7022.800105
## stopped after 100 iterations
## # weights:  496
## initial  value 32134.744969
## iter  10 value 10701.027253
## iter  20 value 8258.479608
## iter  30 value 7486.135774
## iter  40 value 7339.871893
```

```
## iter  50 value 7245.080537
## iter  60 value 7061.461826
## iter  70 value 6922.190655
## iter  80 value 6810.266255
## iter  90 value 6724.143673
## iter 100 value 6665.325276
## final  value 6665.325276
## stopped after 100 iterations
## # weights:  496
## initial  value 38251.920488
## iter  10 value 10791.902372
## iter  20 value 8741.588341
## iter  30 value 7981.429351
## iter  40 value 7620.417152
## iter  50 value 7281.177400
## iter  60 value 7086.904004
## iter  70 value 6967.582628
## iter  80 value 6882.093157
## iter  90 value 6752.317339
## iter 100 value 6656.112322
## final  value 6656.112322
## stopped after 100 iterations
## # weights:  496
## initial  value 20609.319365
## iter  10 value 10144.478051
## iter  20 value 7840.238306
## iter  30 value 7261.513898
## iter  40 value 7114.246439
## iter  50 value 6969.709706
## iter  60 value 6855.065447
## iter  70 value 6804.080225
## iter  80 value 6744.642738
## iter  90 value 6667.198924
## iter 100 value 6604.121202
## final  value 6604.121202
## stopped after 100 iterations
## # weights:  496
## initial  value 74174.998010
## iter  10 value 10115.236609
## iter  20 value 7694.136562
## iter  30 value 7404.824948
## iter  40 value 7200.070132
## iter  50 value 7058.566793
## iter  60 value 6887.861754
## iter  70 value 6781.302972
## iter  80 value 6702.055970
## iter  90 value 6652.224646
## iter 100 value 6616.154756
## final  value 6616.154756
## stopped after 100 iterations
```

```r
# The average accuracy of the neural network model is 91.3%
cv_nn <- mean(unlist(cv_results))
cv_nn
```

```
## [1] 91.34698
```

```r
# Decision tree #
# 10-fold cross validation of the Decision tree 1 model
set.seed(141)
folds <- createFolds(bank$y, k=10)
cv_results <- lapply(folds, function (x) {
    bank_train <- bank[-x, ]
    bank_test <- bank[x, ]
    bank_model <- C5.0(bank_train[-21],bank_train$y, trails=20)
    bank_predict <- predict(bank_model, bank_test[-21], type="class")
    accuracy <- confusionMatrix(bank_predict, bank_test$y, positive = "yes")$overall[[1]]
    accuracy <- accuracy*100
    return(accuracy)
})

# The decision tree model has an average accuracy of 91.3%
cv_dt <- mean(unlist(cv_results))
cv_dt
```

```
## [1] 91.32757
```

```r
#
set.seed(141)
folds <- createFolds(bank$y, k=10)
cv_results <- lapply(folds, function (x) {
    bank_train <- bank[-x, ]
    bank_test <- bank[x, ]
    bank_model <- ksvm(y~., data=bank_train, kernel = "rbfdot", C=9)
    bank_predict <- predict(bank_model, bank_test[-21])
    accuracy <- confusionMatrix(bank_predict, bank_test$y, positive = "yes")$overall[[1]]
    accuracy <- accuracy*100
    return(accuracy)
})

# The SVM model has an average accuracy of 90.9%
cv_svm <- mean(unlist(cv_results))
cv_svm
```

```
## [1] 90.96581
```

```r
### Random forest model is more robust and stable in predicting future outcomes and it is
# the best model to use if accuracy is the only criterion ###
```