plots.R

Saloni

2019-11-14

```
#Name: Saloni Mishra
#Purpose: Class project
#install.packages("tidyr")
#install.packages("lmtest", repos = "http://cran.us.r-project.org")
#install.packages("devtools")
#install.packages("tidyverse")
#install.packages("caret")
#install.packages("car")
#install.packages("hrbrthemes")
#install.packages("olsrr")
library(olsrr)

##
## Attaching package: 'olsrr'

## The following object is masked from 'package:datasets':
##
##     rivers

library(hrbrthemes)

## NOTE: Either Arial Narrow or Roboto Condensed fonts are required to use
these themes.

##       Please use hrbrthemes::import_roboto_condensed() to install Roboto
Condensed and

##       if Arial Narrow is not on your system, please see
http://bit.ly/arialnarrow

library(tidyverse)

## -- Attaching packages ---------------------------------------------
tidyverse 1.2.1 --

## v ggplot2 3.2.1     v purrr   0.3.3
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   1.0.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0

## -- Conflicts ------------------------------------------------
tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(caret)
```

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

```
##
## Attaching package: 'caret'
```

```
## The following object is masked from 'package:purrr':
##
##     lift
```

```r
library(car)
```

```
## Loading required package: carData
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

```
## The following object is masked from 'package:purrr':
##
##     some
```

```r
library(VIF)
```

```
##
## Attaching package: 'VIF'
```

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

```r
library(MASS)
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

```
## The following object is masked from 'package:olsrr':
##
##     cement
```

```r
library(leaps)
library(readxl)
library(dplyr)
library(RColorBrewer)
library(reshape2)
```

```
##
## Attaching package: 'reshape2'

## The following object is masked from 'package:tidyr':
##
##     smiths

library(ISLR)
library(ggplot2)
library(digest)
library(tidyr)
library(ggplot2)
library(dplyr)

getPalette = colorRampPalette(brewer.pal(12, "Set1"))

## Warning in brewer.pal(12, "Set1"): n too large, allowed maximum for
palette Set1 is 9
## Returning the palette you asked for with that many colors

my_data<-read_excel("C:\\Users\\Arvind\\Downloads\\default of credit card
clients.xls")

## New names:
## * `` -> ...1

View(my_data)
#str(my_data)
colnames(my_data)<-my_data[1,]
mydata<-data.frame(apply(my_data[-1,], 2, as.numeric))
View(mydata)
#colnames(mydata)<-my_data[1,]
mydata[,c(3:5,7:12,25)]<-data.frame(apply(my_data[-1,c(3:5,7:12,25)], 2,
as.factor))
# mydata[,3:5]<-data.frame(apply(my_data[-1,3:5], 2, as.factor))
# mydata[,7:12]<-data.frame(apply(my_data[-1,7:12], 2, as.factor))
# mydata[,25]<-data.frame(apply(my_data[-1,25], 2, as.factor))
View(mydata)
str(mydata)

## 'data.frame':    30000 obs. of  25 variables:
##  $ ID                        : num  1 2 3 4 5 6 7 8 9 10 ...
##  $ LIMIT_BAL                 : num  20000 120000 90000 50000 50000 50000
500000 100000 140000 20000 ...
##  $ SEX                       : Factor w/ 2 levels "1","2": 2 2 2 2 1 1 1 2
2 1 ...
##  $ EDUCATION                 : Factor w/ 7 levels "0","1","2","3",..: 3 3
3 3 3 2 2 3 4 4 ...
##  $ MARRIAGE                  : Factor w/ 4 levels "0","1","2","3": 2 3 3 2
2 3 3 3 2 3 ...
##  $ AGE                       : num  24 26 34 37 57 37 29 23 28 35 ...
```

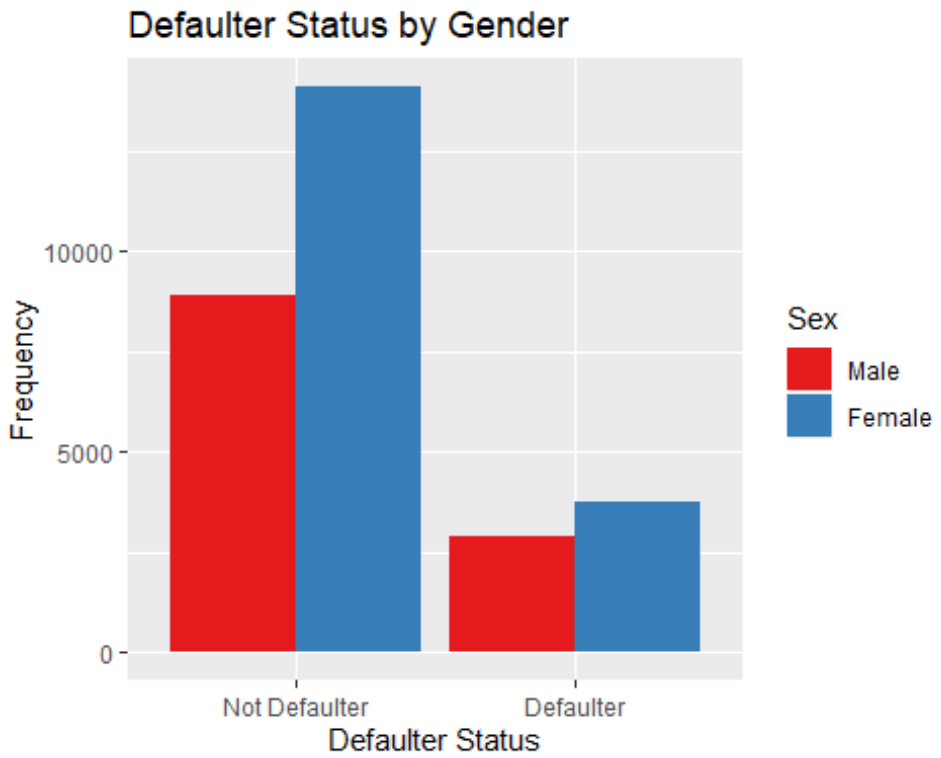
```
##  $ PAY_0                     : Factor w/ 11 levels "-1","-2","0",..: 5 1 3
3 1 3 3 3 3 2 ...
##  $ PAY_2                     : Factor w/ 11 levels "-1","-2","0",..: 5 5 3
3 3 3 3 1 3 2 ...
##  $ PAY_3                     : Factor w/ 11 levels "-1","-2","0",..: 1 3 3
3 1 3 3 1 5 2 ...
##  $ PAY_4                     : Factor w/ 11 levels "-1","-2","0",..: 1 3 3
3 3 3 3 3 3 2 ...
##  $ PAY_5                     : Factor w/ 10 levels "-1","-2","0",..: 2 3 3
3 3 3 3 3 3 1 ...
##  $ PAY_6                     : Factor w/ 10 levels "-1","-2","0",..: 2 4 3
3 3 3 3 1 3 1 ...
##  $ BILL_AMT1                 : num  3913 2682 29239 46990 8617 ...
##  $ BILL_AMT2                 : num  3102 1725 14027 48233 5670 ...
##  $ BILL_AMT3                 : num  689 2682 13559 49291 35835 ...
##  $ BILL_AMT4                 : num  0 3272 14331 28314 20940 ...
##  $ BILL_AMT5                 : num  0 3455 14948 28959 19146 ...
##  $ BILL_AMT6                 : num  0 3261 15549 29547 19131 ...
##  $ PAY_AMT1                  : num  0 0 1518 2000 2000 ...
##  $ PAY_AMT2                  : num  689 1000 1500 2019 36681 ...
##  $ PAY_AMT3                  : num  0 1000 1000 1200 10000 657 38000 0 432
0 ...
##  $ PAY_AMT4                  : num  0 1000 1000 1100 9000 ...
##  $ PAY_AMT5                  : num  0 0 1000 1069 689 ...
##  $ PAY_AMT6                  : num  0 2000 5000 1000 679 ...
##  $ default.payment.next.month: Factor w/ 2 levels "0","1": 2 2 1 1 1 1 1 1
1 1 ...
```

```
mydata$SEX<-ordered(mydata$SEX,levels = c(1,2), labels = c("Male", "Female"))
mydata$EDUCATION<-ordered(mydata$EDUCATION,levels = c(1,2,3,4), labels =
c("graduate school", "university", "high school", "others"))
mydata$MARRIAGE<-ordered(mydata$MARRIAGE,levels = c("1","2","3"), labels =
c("married","single","others"))
mydata$`default payment next month`<-
ordered(mydata$default.payment.next.month,levels = c("0","1"), labels =
c("Not Defaulter", "Defaulter"))
View(mydata)
#mydata$PAY_0<-ordered(mydata$PAY_0,levels = c(-1,-2,0,1,2,3,4,5,6,7,8),
labels = c("one month early","two months early","pay duly", "payment delay
for one month", "payment delay for two months","payment delay for three
months","payment delay for four months","payment delay for five
months","payment delay for six months", "payment delay for seven months",
"payment delay for eight months", "payment delay for nine months and above"))
summary(mydata)
```

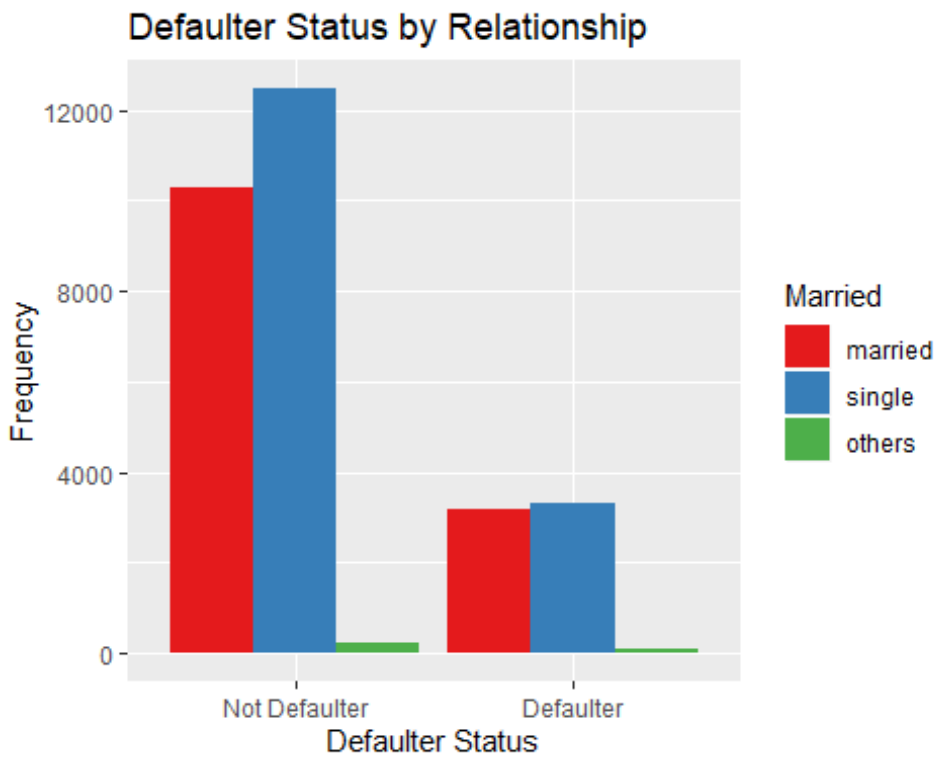
```
##        ID          LIMIT_BAL          SEX                   EDUCATION
##  Min.   :    1   Min.   :  10000   Male  :11888   graduate school:10585
##  1st Qu.: 7501   1st Qu.:  50000   Female:18112   university     :14030
##  Median :15000   Median : 140000                  high school    : 4917
##  Mean   :15000   Mean   : 167484                  others         :  123
```

```
##  3rd Qu.:22500   3rd Qu.: 240000                  NA's           :  345
##  Max.   :30000   Max.   :1000000
##
##     MARRIAGE          AGE             PAY_0           PAY_2
##  married:13659   Min.   :21.00   0      :14737   0      :15730
##  single :15964   1st Qu.:28.00   -1     : 5686   -1     : 6050
##  others :  323   Median :34.00   1      : 3688   2      : 3927
##  NA's   :   54   Mean   :35.49   -2     : 2759   -2     : 3782
##                  3rd Qu.:41.00   2      : 2667   3      :  326
##                  Max.   :79.00   3      :  322   4      :   99
##                                  (Other):  141   (Other):   86
##      PAY_3           PAY_4           PAY_5           PAY_6
##  0      :15764   0      :16455   0      :16947   0      :16286
##  -1     : 5938   -1     : 5687   -1     : 5539   -1     : 5740
##  -2     : 4085   -2     : 4348   -2     : 4546   -2     : 4895
##  2      : 3819   2      : 3159   2      : 2626   2      : 2766
##  3      :  240   3      :  180   3      :  178   3      :  184
##  4      :   76   4      :   69   4      :   84   4      :   49
##  (Other):   78   (Other):  102   (Other):   80   (Other):   80
##    BILL_AMT1         BILL_AMT2        BILL_AMT3         BILL_AMT4
##  Min.   :-165580   Min.   :-69777   Min.   :-157264   Min.   :-170000
##  1st Qu.:   3559   1st Qu.:  2985   1st Qu.:   2666   1st Qu.:   2327
##  Median :  22382   Median : 21200   Median :  20089   Median :  19052
##  Mean   :  51223   Mean   : 49179   Mean   :  47013   Mean   :  43263
##  3rd Qu.:  67091   3rd Qu.: 64006   3rd Qu.:  60165   3rd Qu.:  54506
##  Max.   : 964511   Max.   :983931   Max.   :1664089   Max.   : 891586
##
##    BILL_AMT5        BILL_AMT6          PAY_AMT1         PAY_AMT2
##  Min.   :-81334   Min.   :-339603   Min.   :     0   Min.   :      0
##  1st Qu.:  1763   1st Qu.:   1256   1st Qu.:  1000   1st Qu.:    833
##  Median : 18105   Median :  17071   Median :  2100   Median :   2009
##  Mean   : 40311   Mean   :  38872   Mean   :  5664   Mean   :   5921
##  3rd Qu.: 50191   3rd Qu.:  49198   3rd Qu.:  5006   3rd Qu.:   5000
##  Max.   :927171   Max.   : 961664   Max.   :873552   Max.   :1684259
##
##     PAY_AMT3         PAY_AMT4         PAY_AMT5           PAY_AMT6
##  Min.   :     0   Min.   :     0   Min.   :     0.0   Min.   :     0.0
##  1st Qu.:   390   1st Qu.:   296   1st Qu.:   252.5   1st Qu.:   117.8
##  Median :  1800   Median :  1500   Median :  1500.0   Median :  1500.0
##  Mean   :  5226   Mean   :  4826   Mean   :  4799.4   Mean   :  5215.5
##  3rd Qu.:  4505   3rd Qu.:  4013   3rd Qu.:  4031.5   3rd Qu.:  4000.0
##  Max.   :896040   Max.   :621000   Max.   :426529.0   Max.   :528666.0
##
##  default.payment.next.month default payment next month
##  0:23364                    Not Defaulter:23364
##  1: 6636                    Defaulter    : 6636
##
##
##
```

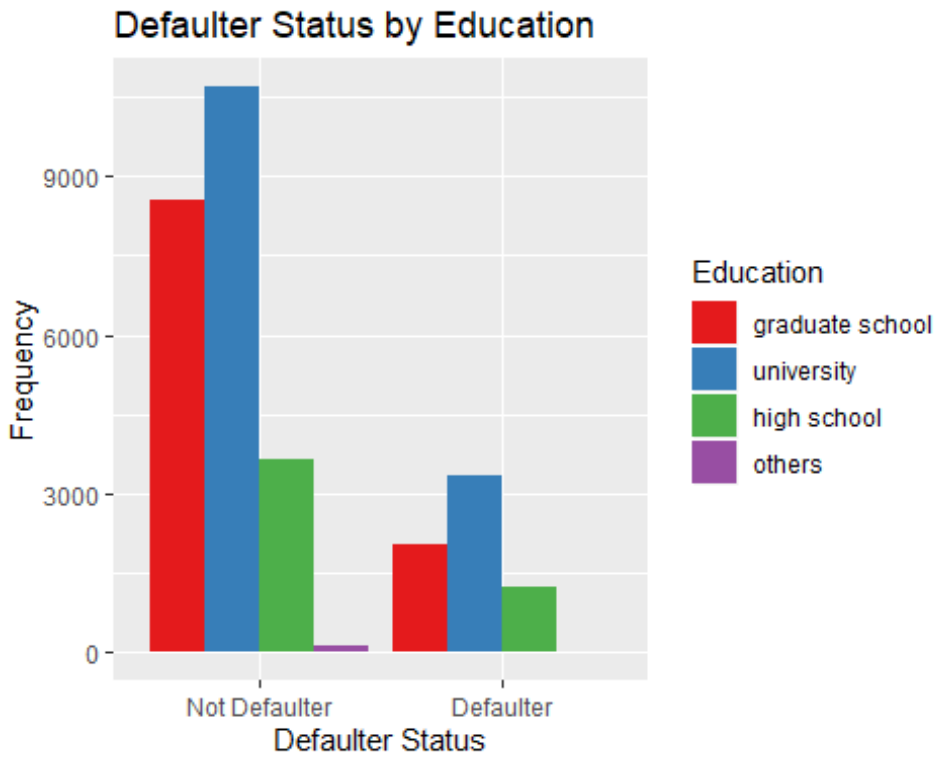
```
##
##

table(mydata$PAY_0)

##
##    -1    -2     0     1     2     3     4     5     6     7     8
##  5686  2759 14737  3688  2667   322    76    26    11     9    19

# Looking for incomplete cases
#mydata[!complete.cases(mydata),]
#Deleting NAs from data
mydata <- na.omit(mydata)
View(mydata)
#Putting all the rows in sequence after deleting
rownames(mydata)<- seq(length=nrow(mydata))

# EDA plots
#####################################################################
plot1<-data.frame(table(mydata$SEX,mydata$`default payment next month`))
names(plot1)<-c("Sex","Default","Frequency")
ggplot(plot1, aes(factor(Default), Frequency, fill =Sex )) +
  geom_bar(stat="identity", position = "dodge") +
  xlab("Defaulter Status")+
  ggtitle("Defaulter Status by Gender")+
  scale_fill_brewer(palette = "Set1")
```



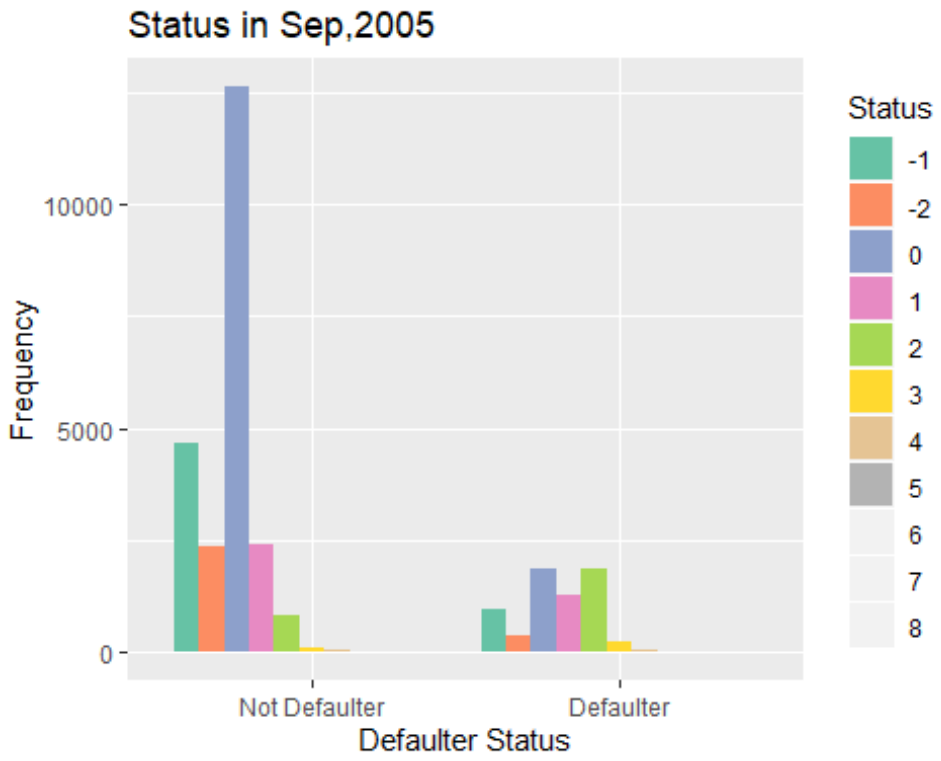
```
#######################################
plot2<-data.frame(table(mydata$MARRIAGE,mydata$`default payment next month`))
names(plot2)<-c("Married","Default","Frequency")
ggplot(plot2, aes(factor(Default), Frequency, fill =Married )) +
  geom_bar(stat="identity", position = "dodge") +
  xlab("Defaulter Status")+
  ggtitle("Defaulter Status by Relationship")+
  scale_fill_brewer(palette = "Set1")
```



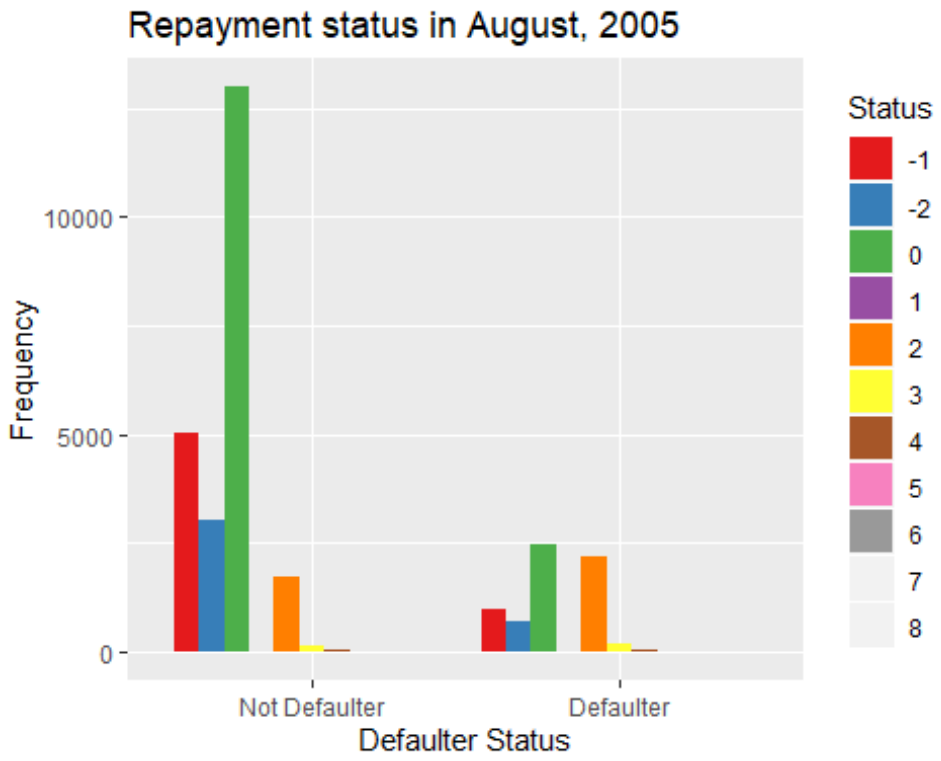

```
#############################################
plot3<-data.frame(table(mydata$EDUCATION,mydata$`default payment next
month`))
names(plot3)<-c("Education","Default","Frequency")
ggplot(plot3, aes(factor(Default), Frequency, fill =Education )) +
  geom_bar(stat="identity", position = "dodge") +
  xlab("Defaulter Status")+
  ggtitle("Defaulter Status by Education")+
  scale_fill_brewer(palette = "Set1")
```



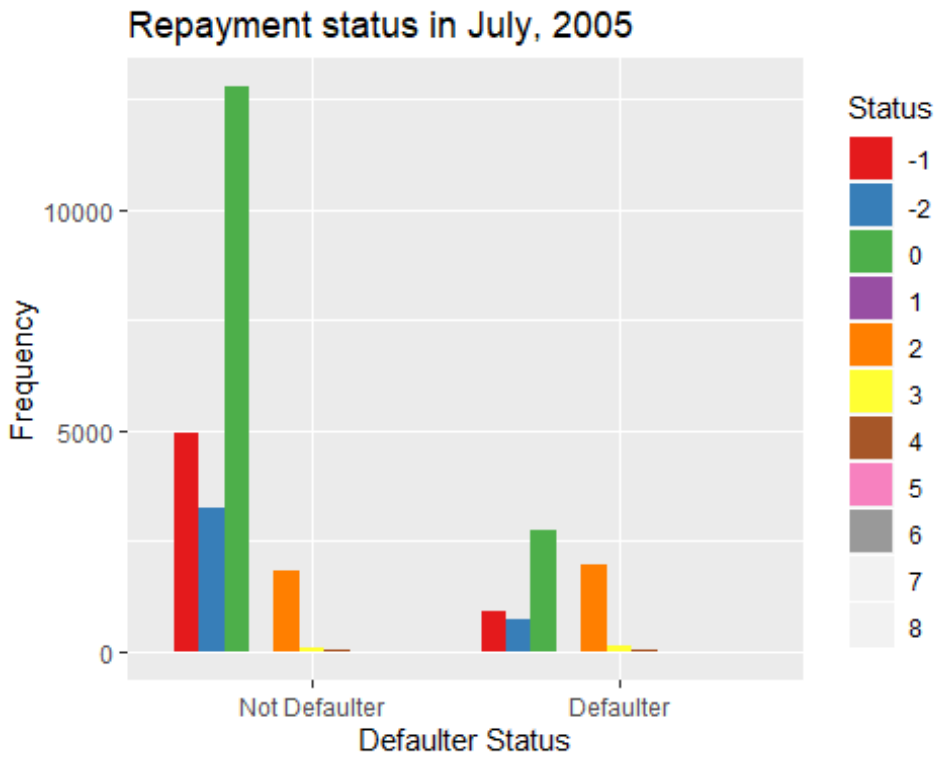

```
###############################################
plot4<-data.frame(table(mydata$PAY_0,mydata$`default payment next month`))
names(plot4)<-c("Status","Defaulter","Frequency")
ggplot(plot4, aes(factor(Defaulter), Frequency, fill = Status)) +
  geom_bar(stat="identity", position = "dodge") +
  xlab("Defaulter Status")+
  ggtitle("Status in Sep,2005")+
  scale_fill_brewer(palette = "Set2")

## Warning in RColorBrewer::brewer.pal(n, pal): n too large, allowed maximum
for palette Set2 is 8
## Returning the palette you asked for with that many colors
```



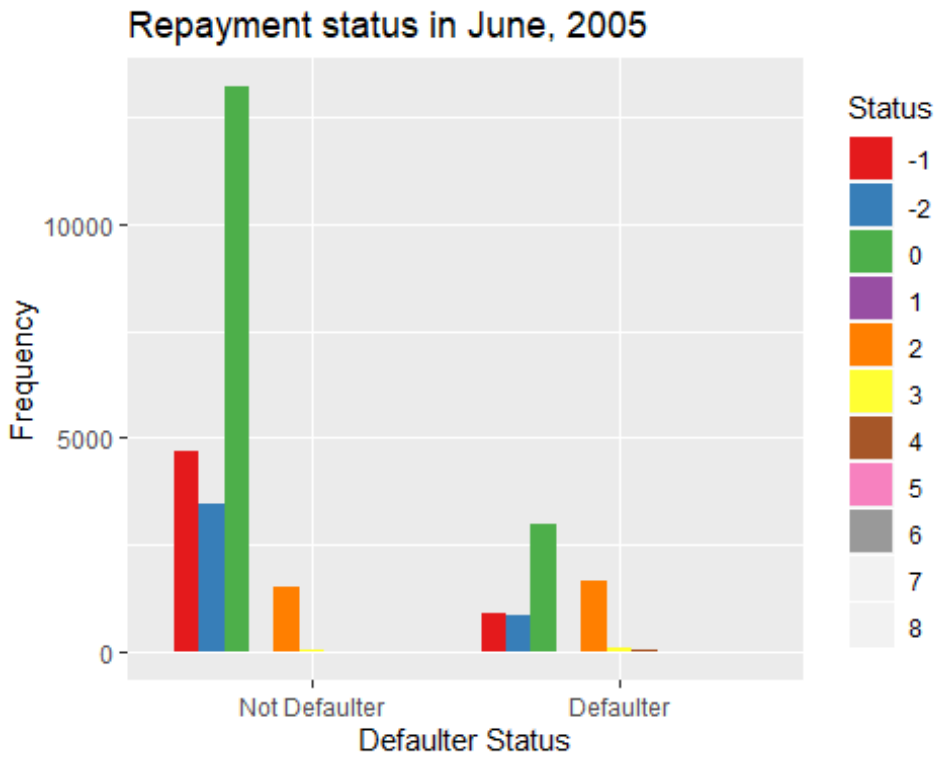

```
###############################################
plot5<-data.frame(table(mydata$PAY_2,mydata$`default payment next month`))
names(plot5)<-c("Status","Defaulter","Frequency")
ggplot(plot5, aes(factor(Defaulter), Frequency, fill =Status )) +
  geom_bar(stat="identity", position = "dodge") +
  xlab("Defaulter Status")+
  ggtitle("Repayment status in August, 2005")+
  scale_fill_brewer(palette = "Set1")

## Warning in RColorBrewer::brewer.pal(n, pal): n too large, allowed maximum
for palette Set1 is 9
## Returning the palette you asked for with that many colors
```



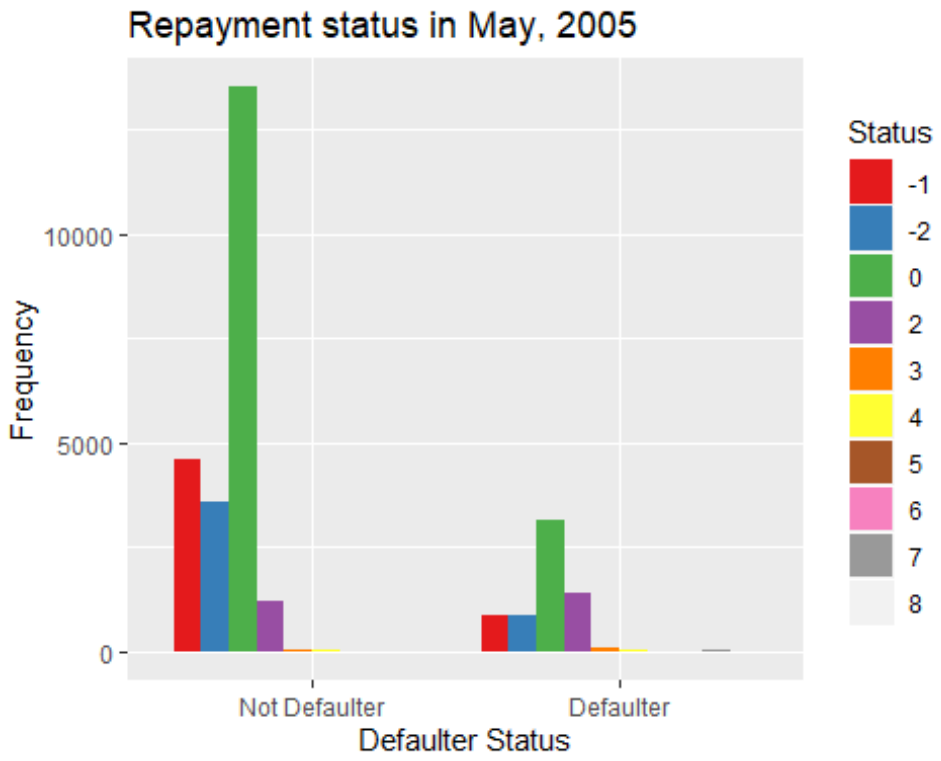

```
###############################################
plot6<-data.frame(table(mydata$PAY_3,mydata$`default payment next month`))
names(plot6)<-c("Status","Defaulter","Frequency")
ggplot(plot6, aes(factor(Defaulter), Frequency, fill =Status)) +
  geom_bar(stat="identity", position = "dodge") +
  xlab("Defaulter Status")+
  ggtitle("Repayment status in July, 2005")+
  scale_fill_brewer(palette = "Set1")

## Warning in RColorBrewer::brewer.pal(n, pal): n too large, allowed maximum
for palette Set1 is 9
## Returning the palette you asked for with that many colors
```



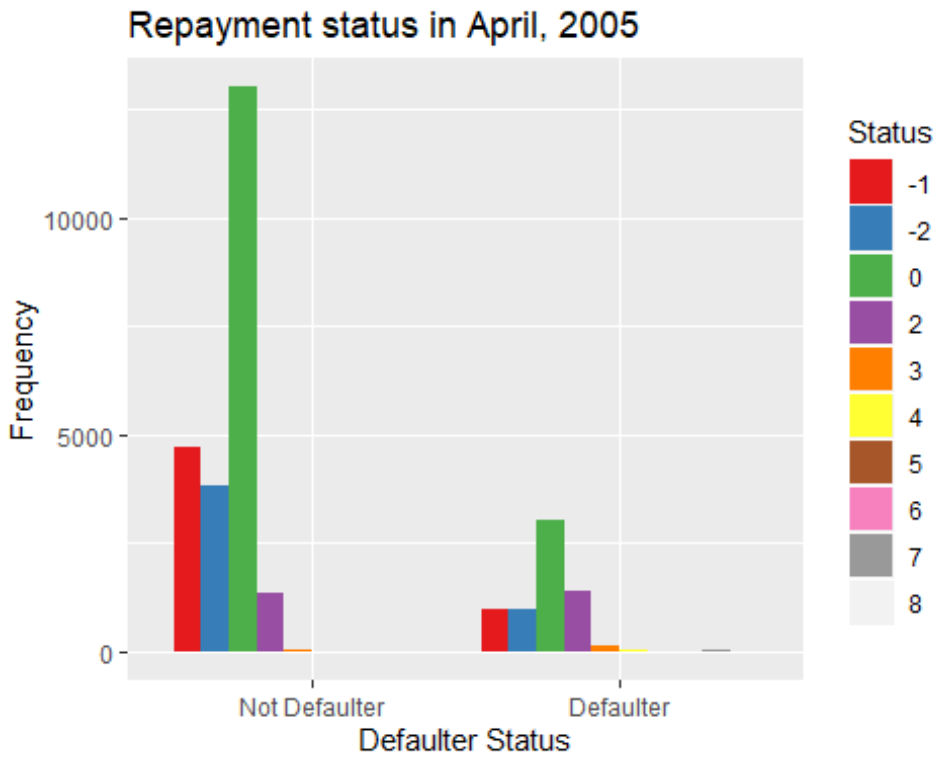

```
################################################
getPalette = colorRampPalette(brewer.pal(10, "Set1"))

## Warning in brewer.pal(10, "Set1"): n too large, allowed maximum for
palette Set1 is 9
## Returning the palette you asked for with that many colors

plot7<-data.frame(table(mydata$PAY_4,mydata$`default payment next month`))
names(plot7)<-c("Status","Defaulter","Frequency")
ggplot(plot7, aes(factor(Defaulter), Frequency, fill =Status )) +
  geom_bar(stat="identity", position = "dodge") +
  xlab("Defaulter Status")+
  ggtitle("Repayment status in June, 2005")+
  scale_fill_brewer(palette = "Set1")

## Warning in RColorBrewer::brewer.pal(n, pal): n too large, allowed maximum
for palette Set1 is 9
## Returning the palette you asked for with that many colors
```




```
###############################################
plot8<-data.frame(table(mydata$PAY_5,mydata$`default payment next month`))
names(plot8)<-c("Status","Defaulter","Frequency")
ggplot(plot8, aes(factor(Defaulter), Frequency, fill =Status)) +
  geom_bar(stat="identity", position = "dodge") +
  xlab("Defaulter Status")+
  ggtitle("Repayment status in May, 2005")+
  scale_fill_brewer(palette = "Set1")

## Warning in RColorBrewer::brewer.pal(n, pal): n too large, allowed maximum
for palette Set1 is 9
## Returning the palette you asked for with that many colors
```




```
################################################
plot9<-data.frame(table(mydata$PAY_6,mydata$`default payment next month`))
names(plot9)<-c("Status","Defaulter","Frequency")
ggplot(plot9, aes(factor(Defaulter), Frequency, fill = Status)) +
  geom_bar(stat="identity", position = "dodge")+
  xlab("Defaulter Status")+
  ggtitle("Repayment status in April, 2005")+
  scale_fill_brewer(palette = "Set1")

## Warning in RColorBrewer::brewer.pal(n, pal): n too large, allowed maximum
for palette Set1 is 9
## Returning the palette you asked for with that many colors
```

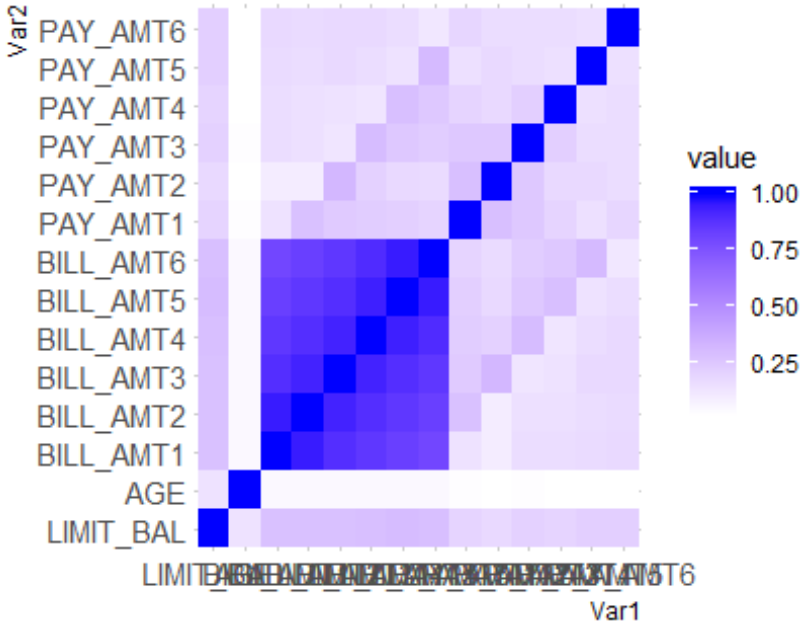



```
################################################
# plot for Collinearity (heat map)
################################################
data<-mydata[,c(2,6,13:24)]
data1 <- round(cor(data),2)
head(data1)
```

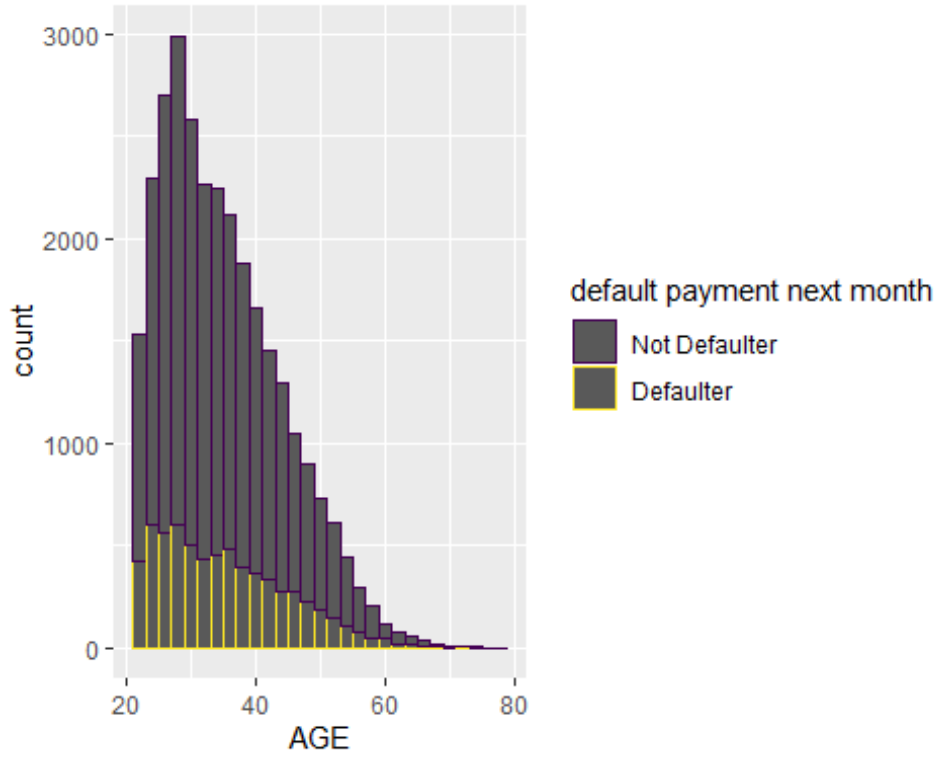
```
##           LIMIT_BAL  AGE BILL_AMT1 BILL_AMT2 BILL_AMT3 BILL_AMT4 BILL_AMT5
## LIMIT_BAL      1.00 0.14      0.28      0.28      0.28      0.29      0.30
## AGE            0.14 1.00      0.05      0.05      0.05      0.05      0.05
## BILL_AMT1      0.28 0.05      1.00      0.95      0.89      0.86      0.83
## BILL_AMT2      0.28 0.05      0.95      1.00      0.93      0.89      0.86
## BILL_AMT3      0.28 0.05      0.89      0.93      1.00      0.93      0.89
## BILL_AMT4      0.29 0.05      0.86      0.89      0.93      1.00      0.94
##           BILL_AMT6 PAY_AMT1 PAY_AMT2 PAY_AMT3 PAY_AMT4 PAY_AMT5 PAY_AMT6
## LIMIT_BAL      0.29     0.20     0.18     0.21     0.20     0.22     0.22
## AGE            0.05     0.03     0.02     0.03     0.02     0.02     0.02
## BILL_AMT1      0.80     0.14     0.10     0.16     0.16     0.17     0.18
## BILL_AMT2      0.83     0.28     0.10     0.15     0.15     0.16     0.17
## BILL_AMT3      0.86     0.24     0.32     0.13     0.14     0.18     0.18
## BILL_AMT4      0.90     0.23     0.21     0.30     0.13     0.16     0.18
```

```
data2<-melt(data1)
plot10 <- ggplot(data = data2, aes(x=Var1, y=Var2, fill=value,
                                   label= value))
plot11 <- plot10 + geom_tile()+
  scale_fill_gradient(low="white", high="blue") +
```

```
  theme_ipsum()
plot11

## Warning in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)): font
## family not found in Windows font database

## Warning in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)): font
## family not found in Windows font database

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## font family not found in Windows font database

## Warning in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)): font
## family not found in Windows font database

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## font family not found in Windows font database

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## font family not found in Windows font database

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## font family not found in Windows font database

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## font family not found in Windows font database

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## font family not found in Windows font database

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## font family not found in Windows font database

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x,
## x$y, : font family not found in Windows font database

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## font family not found in Windows font database

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## font family not found in Windows font database
```




```
#########################################################
#plot for Age (I will think about using it)
plot12<-ggplot(mydata, aes(x=AGE,color=`default payment next month`)) +
  geom_histogram(binwidth=2)
plot12
```



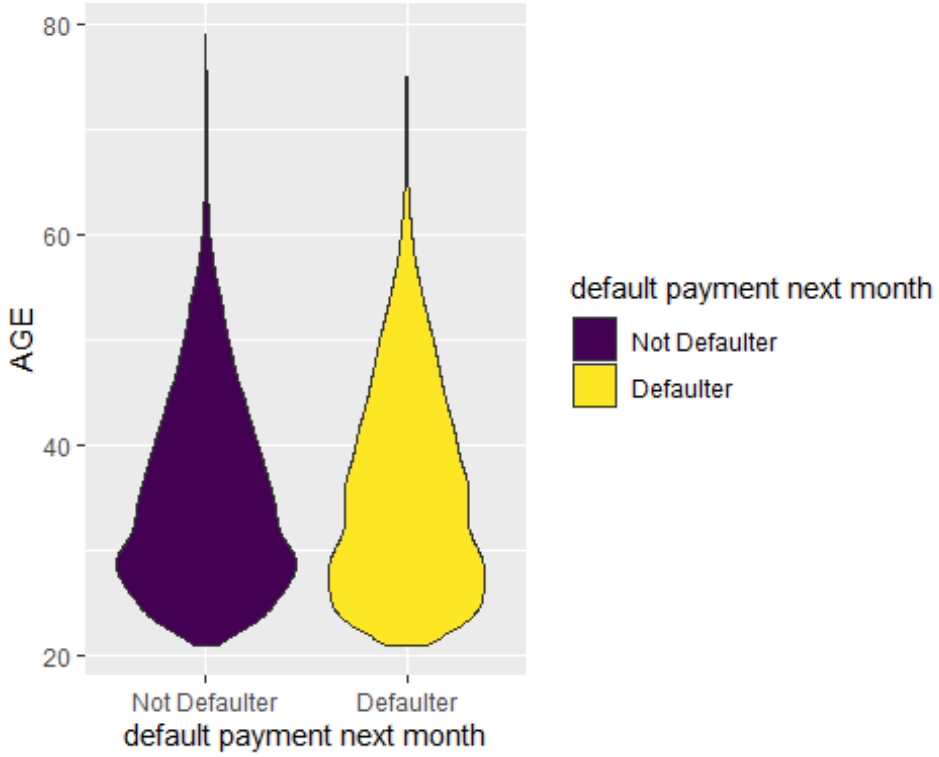

```
#pairs(mydata[,c(2,6,25,13:24)])
#########################################################
# Voilin Plot for Age
#########################################################
z<-data.frame(table(mydata$AGE))
names(z)<-c("AGE","Frequency")
head(z)
```

```
##   AGE Frequency
## 1  21        64
## 2  22       553
## 3  23       917
## 4  24      1117
## 5  25      1176
## 6  26      1245
```

```
plot13<-data.frame(table(mydata$AGE,mydata$`default payment next month`))
names(plot13)<-c("AGE","Defaulter","Frequency")
myda<-data.frame(apply(plot13[,-2], 2, as.numeric))
#ggplot(mydata, aes(x=`default payment next month`, y=AGE, fill=`default
payment next month`)) +
geom_violin()
```

```
## geom_violin: draw_quantiles = NULL, na.rm = FALSE
## stat_ydensity: trim = TRUE, scale = area, na.rm = FALSE
## position_dodge
```

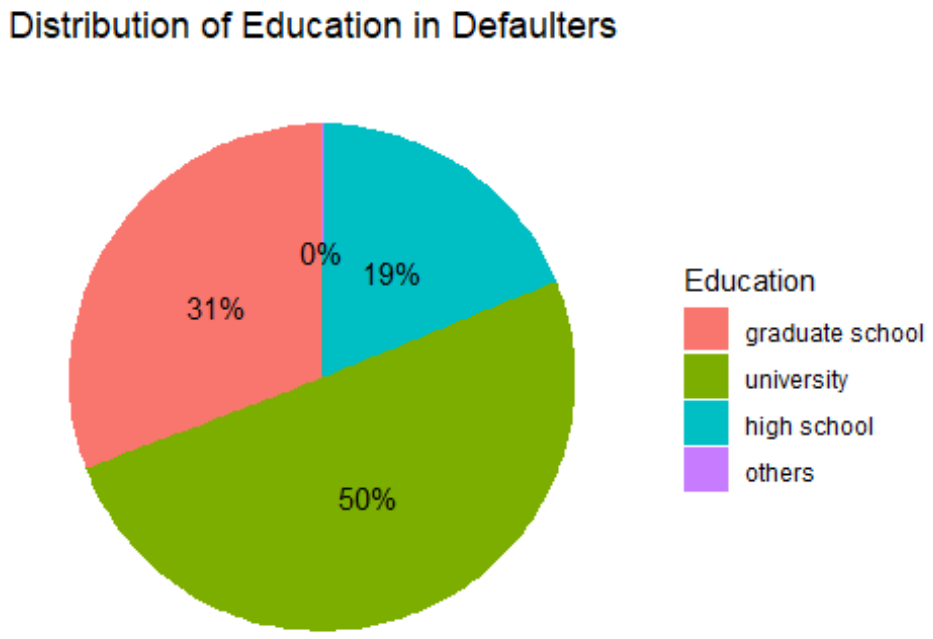
```
Voilin_plot<-ggplot(mydata, aes(x=`default payment next month`, y=AGE,
fill=`default payment next month`)) +
  geom_violin()
Voilin_plot
```




```
############################################################

#####################################################################
#plot13<-data.frame(table(mydata$EDUCATION))
#names(plot13)<-c("Education","Frequency")
#plot13$prop<-plot13$Frequency/sum(plot13$Frequency)
#ggplot(plot13, aes("", y=prop, fill = Education )) +
#geom_bar(stat="identity") +
#coord_polar("y", start=0)
#ggsave("plot13.png", width=4)
#############################################################

# Pie Chart
############################################################
mydata1<-mydata[mydata$`default payment next month`=="Defaulter",]
View(mydata1)
plot14<-data.frame(table(mydata1$EDUCATION))
names(plot14)<-c("Education", "Frequency")
View(plot14)
plot14$prop<-plot14$Frequency/sum(plot14$Frequency)
head(plot14)
```

```
##         Education Frequency        prop
## 1 graduate school      2036 0.308251325
## 2      university      3329 0.504012112
## 3     high school      1233 0.186676760
## 4          others         7 0.001059803
```

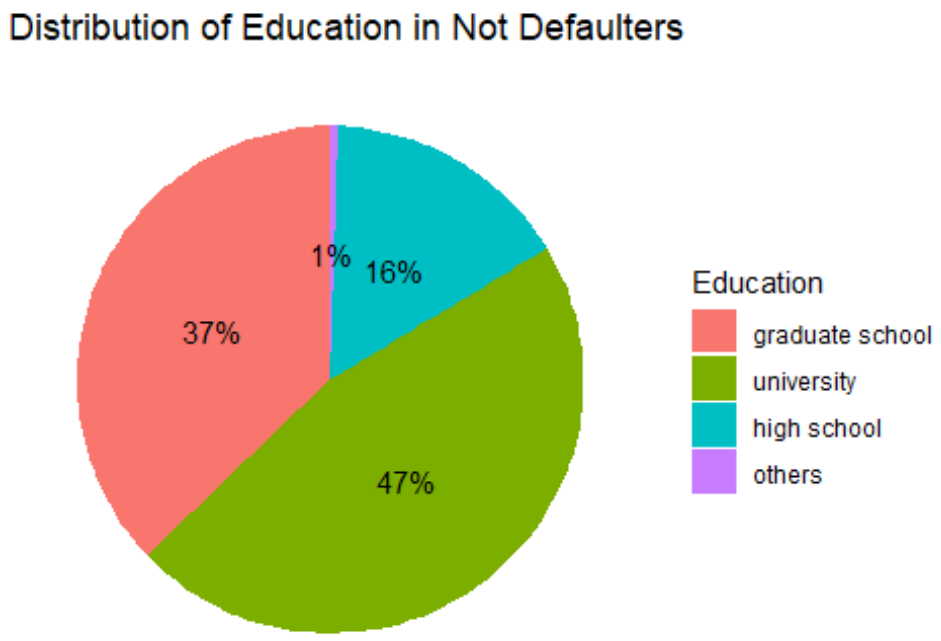
```
ggplot(plot14, aes("", y=prop, fill = Education )) +
  geom_bar(stat="identity") +
  coord_polar("y", start=0)+
  geom_text(aes(label = paste0(round(prop*100), "%")), position =
position_stack(vjust = 0.5))+
  labs(x = NULL, y = NULL, title = "Distribution of Education in
Defaulters")+
  theme_classic() + theme(axis.line = element_blank(),
                          axis.text = element_blank(),
                          axis.ticks = element_blank(),
                          plot.title = element_text(hjust = 0.1))
```




```
#################################################################
mydata2<-mydata[mydata$`default payment next month`=="Not Defaulter",]
View(mydata2)
plot15<-data.frame(table(mydata2$EDUCATION))
names(plot15)<-c("Education", "Frequency")
View(plot15)
plot15$prop<-plot15$Frequency/sum(plot15$Frequency)
head(plot15)
```

```
##         Education Frequency        prop
## 1 graduate school      8545 0.371586363
## 2      university     10695 0.465080884
## 3     high school      3640 0.158288398
## 4          others       116 0.005044356
```

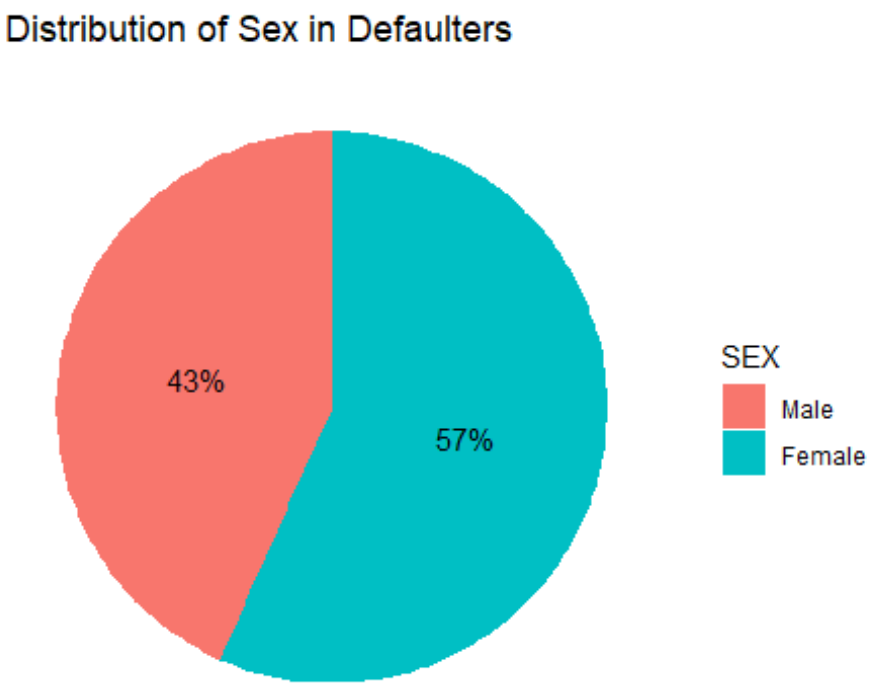
```
ggplot(plot15, aes("", y=prop, fill = Education )) +
  geom_bar(stat="identity") +
  coord_polar("y", start=0)+
  geom_text(aes(label = paste0(round(prop*100), "%")), position =
position_stack(vjust = 0.5))+
  labs(x = NULL, y = NULL, title = "Distribution of Education in Not
Defaulters")+
  theme_classic() + theme(axis.line = element_blank(),
                          axis.text = element_blank(),
                          axis.ticks = element_blank(),
                          plot.title = element_text(hjust = 0.1))
```




```
##############################################################################
##############################################################################
mydata1<-mydata[mydata$`default payment next month`=="Defaulter",]
View(mydata1)
plot16<-data.frame(table(mydata1$SEX))
names(plot16)<-c("SEX", "Frequency")
View(plot16)
plot16$prop<-plot16$Frequency/sum(plot16$Frequency)
head(plot16)
```

```
##      SEX Frequency      prop
## 1   Male      2861 0.4331567
## 2 Female      3744 0.5668433
```

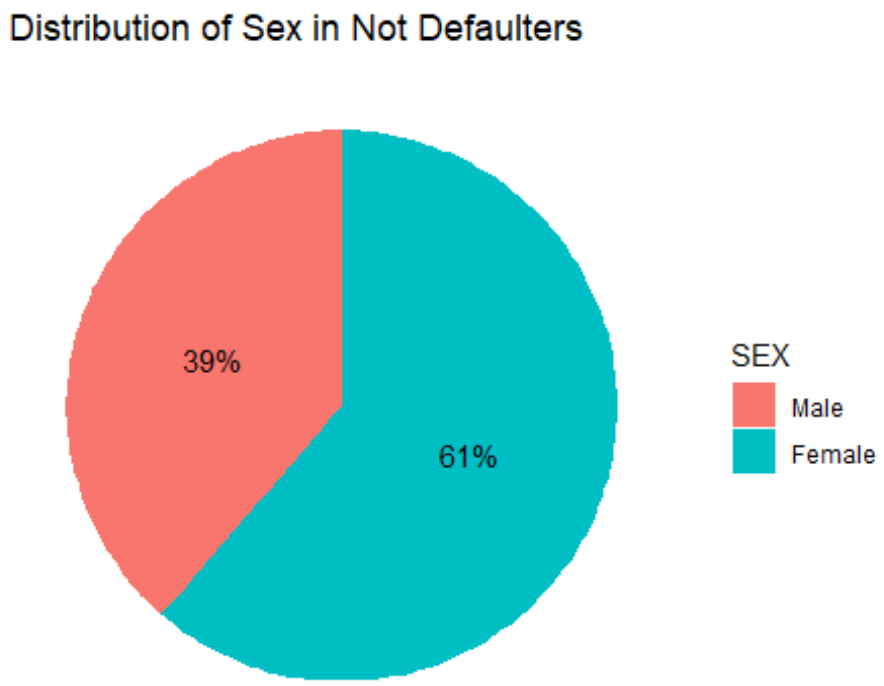
```
ggplot(plot16, aes("", y=prop, fill = SEX )) +
  geom_bar(stat="identity") +
  coord_polar("y", start=0)+
  geom_text(aes(label = paste0(round(prop*100), "%")), position =
position_stack(vjust = 0.5))+
  labs(x = NULL, y = NULL, title = "Distribution of Sex in Defaulters")+
  theme_classic() + theme(axis.line = element_blank(),
                          axis.text = element_blank(),
                          axis.ticks = element_blank(),
                          plot.title = element_text(hjust = 0.1))
```




```
#########################################################################
mydata1<-mydata[mydata$`default payment next month`=="Not Defaulter",]
View(mydata1)
plot17<-data.frame(table(mydata1$SEX))
names(plot17)<-c("SEX", "Frequency")
View(plot17)
plot17$prop<-plot17$Frequency/sum(plot17$Frequency)
head(plot17)
```

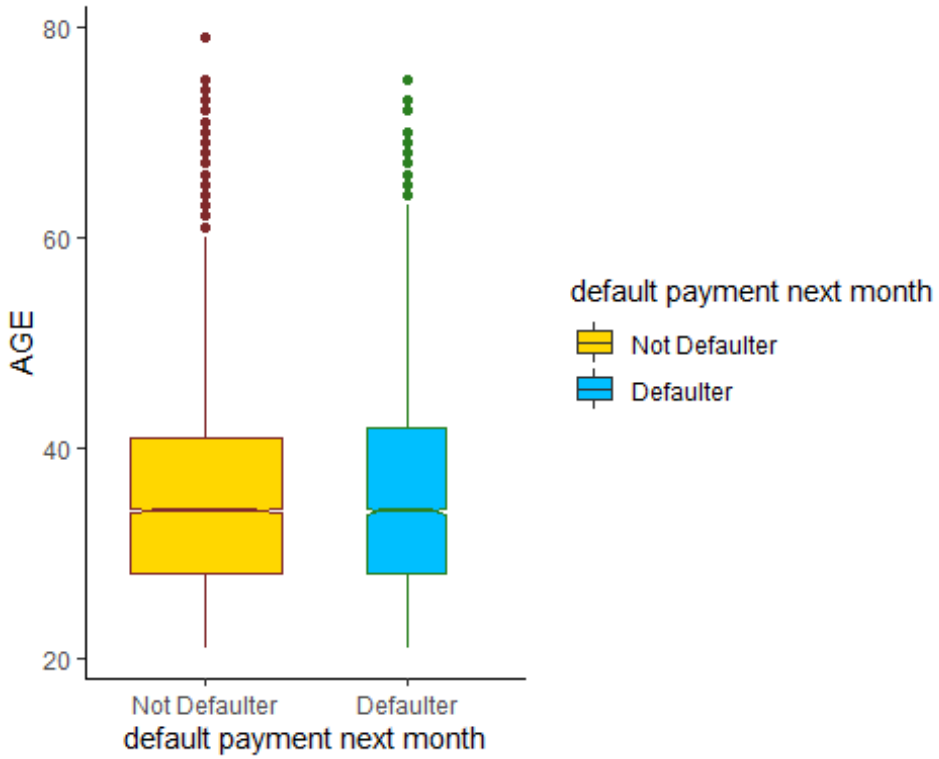
```
##      SEX Frequency      prop
## 1   Male      8885 0.3863715
## 2 Female     14111 0.6136285
```

```
ggplot(plot17, aes("", y=prop, fill = SEX )) +
  geom_bar(stat="identity") +
  coord_polar("y", start=0)+
  geom_text(aes(label = paste0(round(prop*100), "%")), position =
position_stack(vjust = 0.5))+
  labs(x = NULL, y = NULL, title = "Distribution of Sex in Not Defaulters")+
  theme_classic() + theme(axis.line = element_blank(),
                          axis.text = element_blank(),
                          axis.ticks = element_blank(),
                          plot.title = element_text(hjust = 0.1))
```

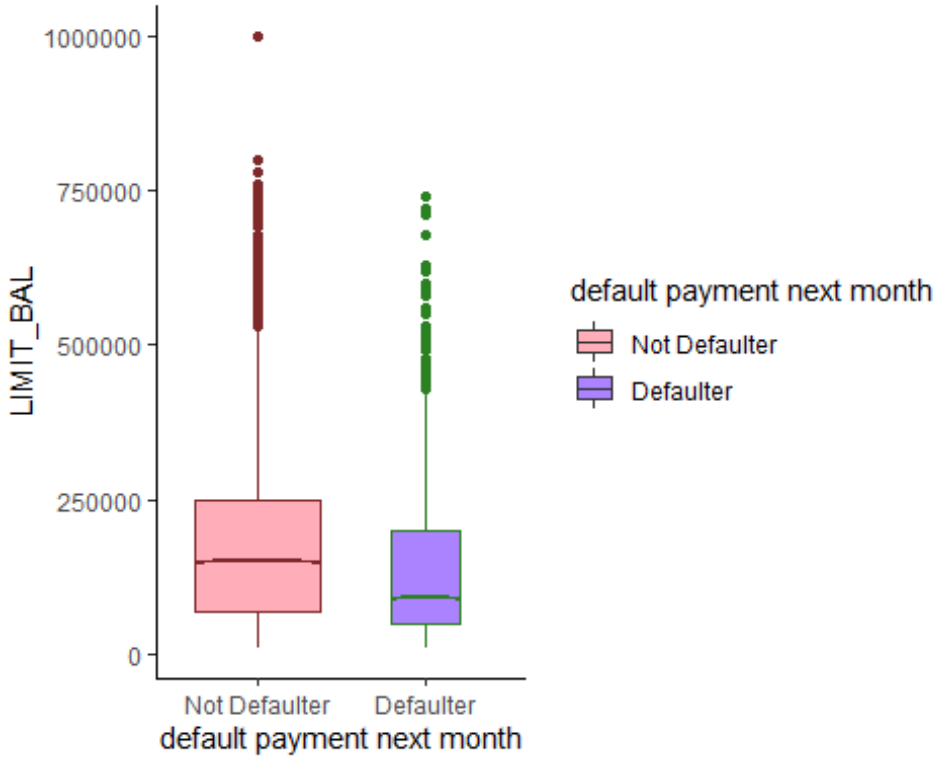



```
#################################################################################
#

# Boxplot
#################################################################################
#
myColors <- ifelse(levels(mydata$`default payment next month`)=="Defaulter" ,
rgb(0.168, 0.50, 0.129) ,
                   ifelse(levels(mydata$`default payment next month`)=="Not
Defaulter", rgb(0.50, 0.162, 0.168),
                          "grey90" ) )
plot18<-ggplot(mydata, aes(x = `default payment next month`, y = AGE, fill=
`default payment next month`, color="myColors")) +
  geom_boxplot(notch=TRUE, notchwidth = 0.7, varwidth = TRUE, col=myColors)+
  scale_fill_manual(values=c("gold", "deepskyblue"))+
```

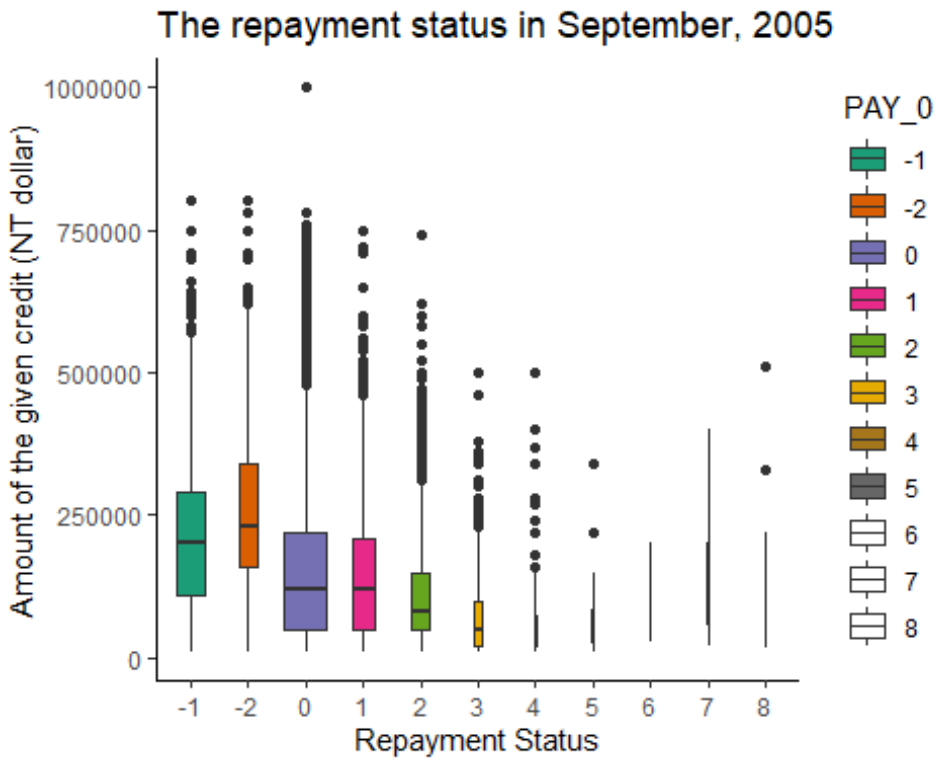
```
  theme_classic()
plot18
```




```
############################################################
plot19<-ggplot(mydata, aes(x = `default payment next month`, y = LIMIT_BAL,
fill= `default payment next month`, color="myColors")) +
  geom_boxplot(notch=TRUE, notchwidth = 0.7, varwidth = TRUE, col=myColors)+
  scale_fill_manual(values=c("lightpink1", "mediumpurple1"))+
  theme_classic()
plot19
```




```
############################################################
mydata2<-mydata[mydata$PAY_0 %in% c("-1","0","1", "2"),]
table(mydata$PAY_0)

##
##    -1    -2     0     1     2     3     4     5     6     7     8
##  5633  2708 14499  3662  2640   320    76    24    11     9    19

plot20<-ggplot(mydata, aes(x = PAY_0, y = LIMIT_BAL, fill=PAY_0)) +
  geom_boxplot(varwidth = TRUE)+
  scale_fill_brewer(palette="Dark2")+
  theme_classic()+
  ggtitle("The repayment status in September, 2005")+
  xlab("Repayment Status")+
  ylab("Amount of the given credit (NT dollar)")
plot20

## Warning in RColorBrewer::brewer.pal(n, pal): n too large, allowed maximum
for palette Dark2 is 8
## Returning the palette you asked for with that many colors
```




```
#####################################################################
#Scatter Plot
#####################################################################
```