The URL for our Team GitHub repository is {https://github.com/Brandeis-BUS111-FinalProject/Prilgrim-Bank.git}.

Team: Yecheng Li, Doris Long, Vera Wang, Jikun Zhou

Class: BUS 111A

Project: Pilgrim case study

```
### Installing & Loading Packages
library(rmarkdown) ### This loads the package
```

Read the file and process the missing data

```
### Read the given dataset
consumerDB = read.csv("dataset.csv") ### read the given dataset

### Check who left the bank in 2000, named the variable as "LeftBank", which is "00"in X0Online and X0B
consumerDB$LeftBank = is.na(consumerDB$X0Online) & is.na(consumerDB$X0Billpay)

### Find the median for 1999 Age and Income
AgeMedian_1999 = median(consumerDB$X9Age,na.rm = TRUE)
IncomeMedian_1999 = median(consumerDB$X9Inc,na.rm = TRUE)

### Present the data for 1999 Age/Income median
AgeMedian_1999
```

```
## [1] 4
```

```
IncomeMedian_1999
```

```
## [1] 6
```

```
### Check who didn't left the bank in 2000, and the income or age in 1999 were missing, we named the va
consumerDB$fixAge = !consumerDB$LeftBank & is.na(consumerDB$X9Age)
consumerDB$fixIncome = !consumerDB$LeftBank & is.na(consumerDB$X9Inc)

### For "fixAge"== TRUE, we substitute "NA" to be "4", which is the median
### For "fixIncome"== TRUE, we substitute "NA" to be "6", which is the median
consumerDB[consumerDB$fixAge,]$X9Age = 4
consumerDB[consumerDB$fixIncome,]$X9Inc = 6
```

```
### Sort the concumerDB and get a Table that sepcifically contains data for 1999
### Name this table as "statsTable1999"
statsTable1999= consumerDB[,2:6]
X9Billpay = consumerDB[,10]
statsTable1999= cbind(statsTable1999,X9Billpay)
```

Statistics Summary for 1999 Data

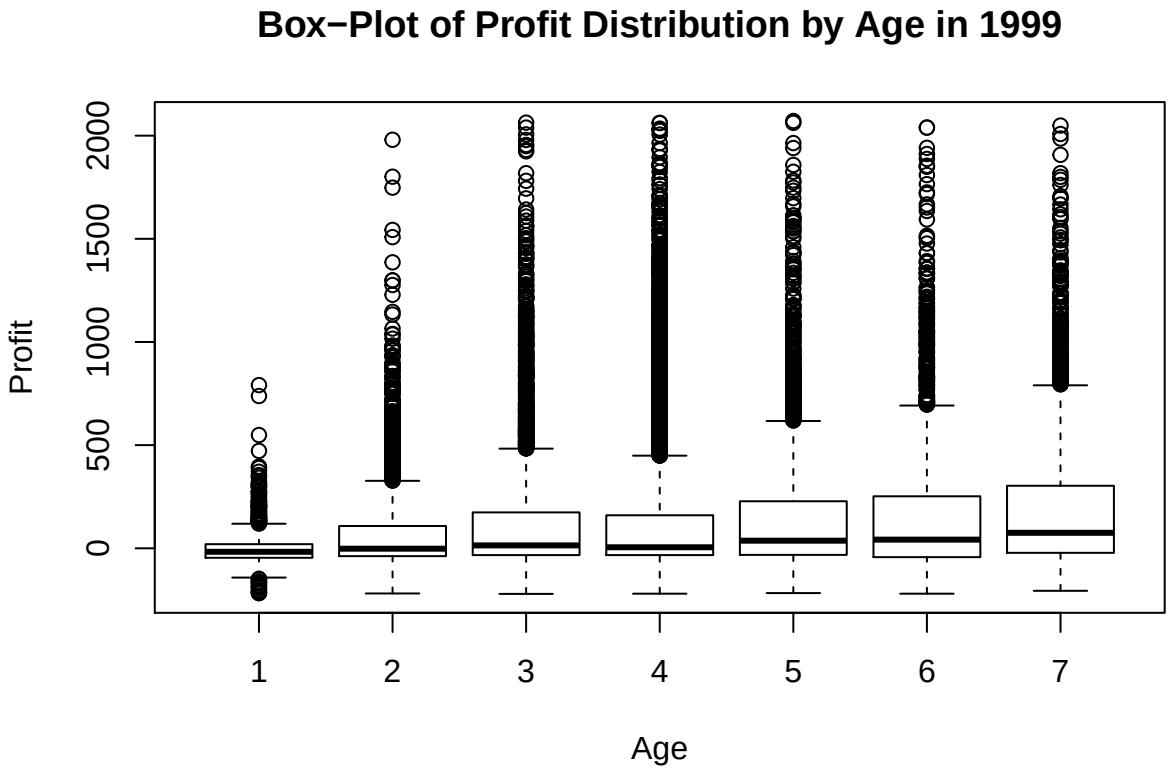
```
### Note: Please install.packages("psych")
library(psych)
```

```
### Describe() method in psych package will give a summary for statsTable
### The Summary_Table
Summary_Table=t(describe(statsTable1999))
Summary_Table = round(Summary_Table,2) ### round the decimal points to 2 digits

### This summary gives the mean, median, standard deviation, min, max and range for 1999 Profit, Age, In
Summary_Table_New = Summary_Table[c(3:5,8:10),c(1:6)]
Summary_Table_New
```

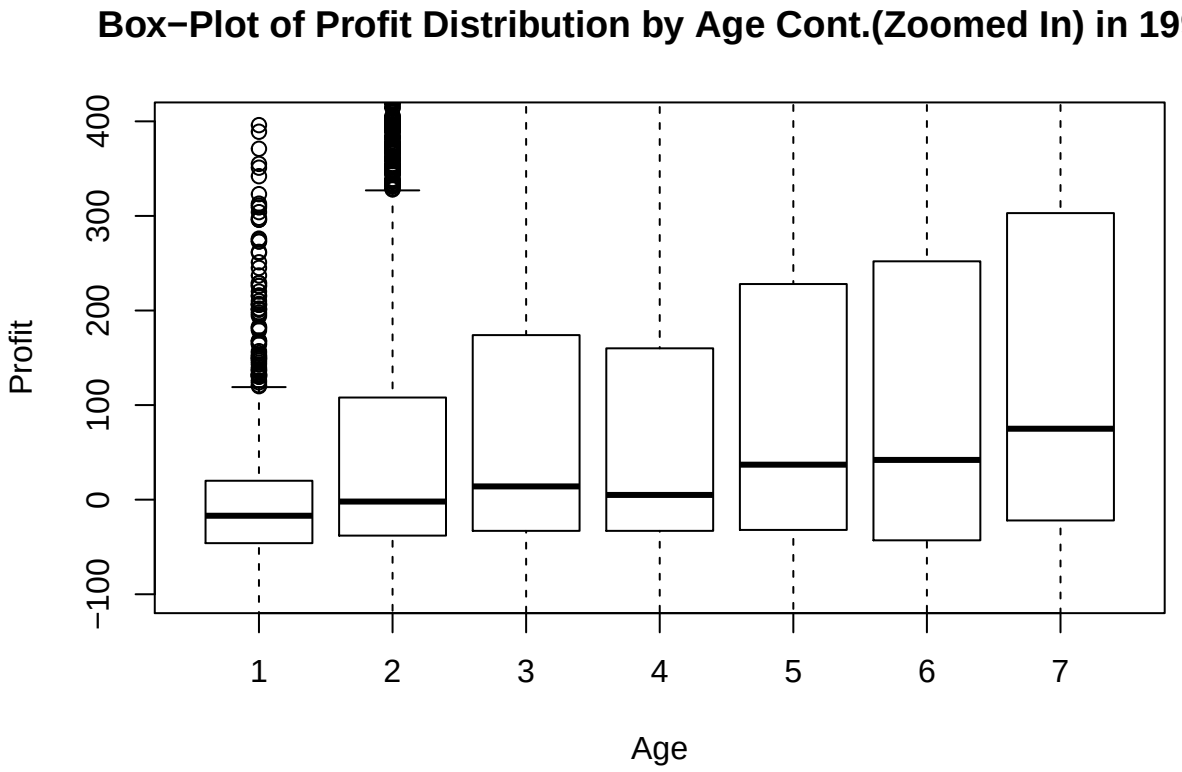
```
##        X9Profit X9Online X9Age X9Inc X9Tenure X9Billpay
## mean     111.50     0.12  4.04  5.55    10.16      0.02
## sd       272.84     0.33  1.49  2.15     8.45      0.13
## median     9.00     0.00  4.00  6.00     7.41      0.00
## min     -221.00     0.00  1.00  1.00     0.16      0.00
## max     2071.00     1.00  7.00  9.00    41.16      1.00
## range   2292.00     1.00  6.00  8.00    41.00      1.00
```

Graphic Summary

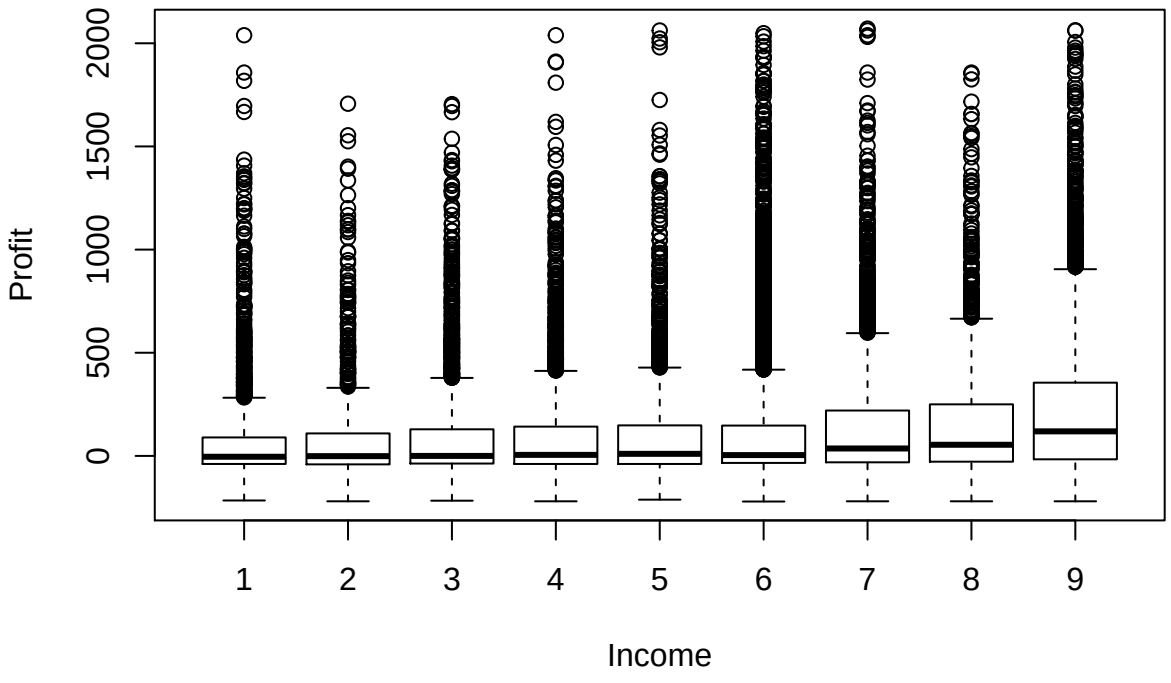
```
### This is a boxplot graph for Profit& Age
boxplot(X9Profit~X9Age, data = consumerDB,
        main = "Box-Plot of Profit Distribution by Age in 1999",  ## Sets Title to Plot
        xlab = "Age", ylab = "Profit") ### Sets X and Y Axes
```



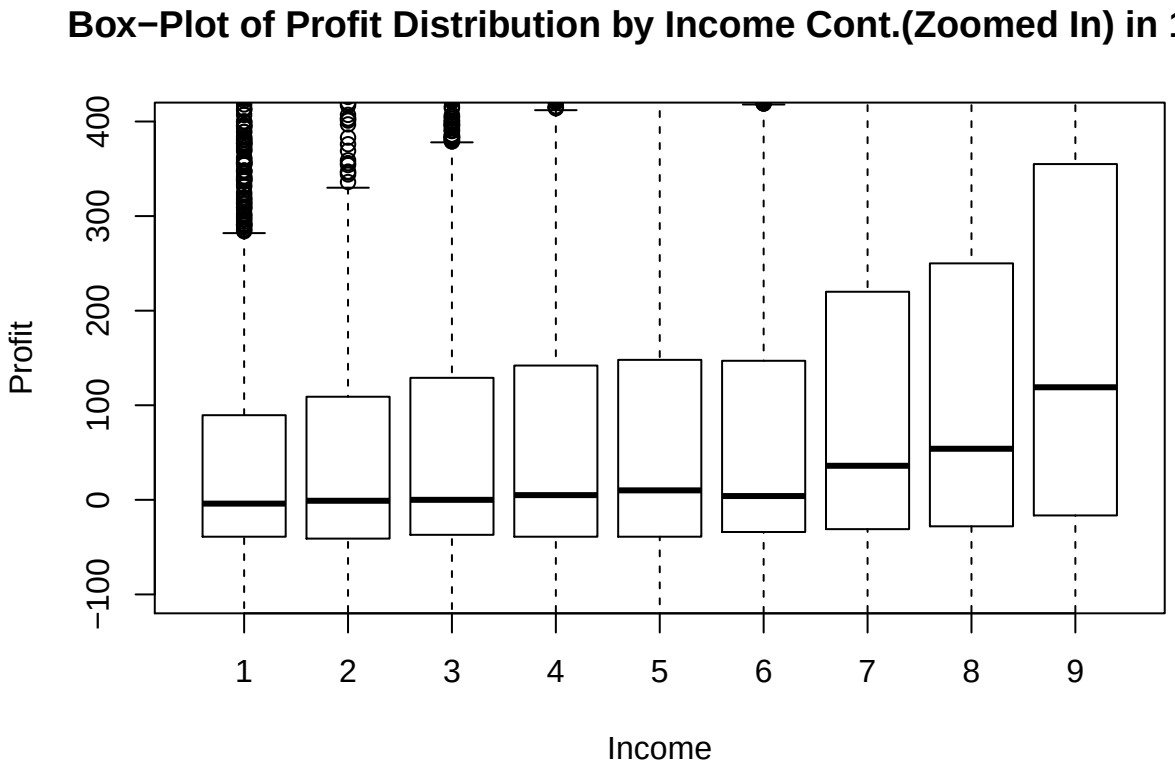

```
### Profit& Age(Zoomed in)
boxplot(X9Profit~X9Age, data = consumerDB,
        main = "Box-Plot of Profit Distribution by Age Cont.(Zoomed In) in 1999",  ## Sets Title to Plo
        xlab = "Age", ylab = "Profit",ylim=c(-100,400)) ### Sets X and Y Axes
```



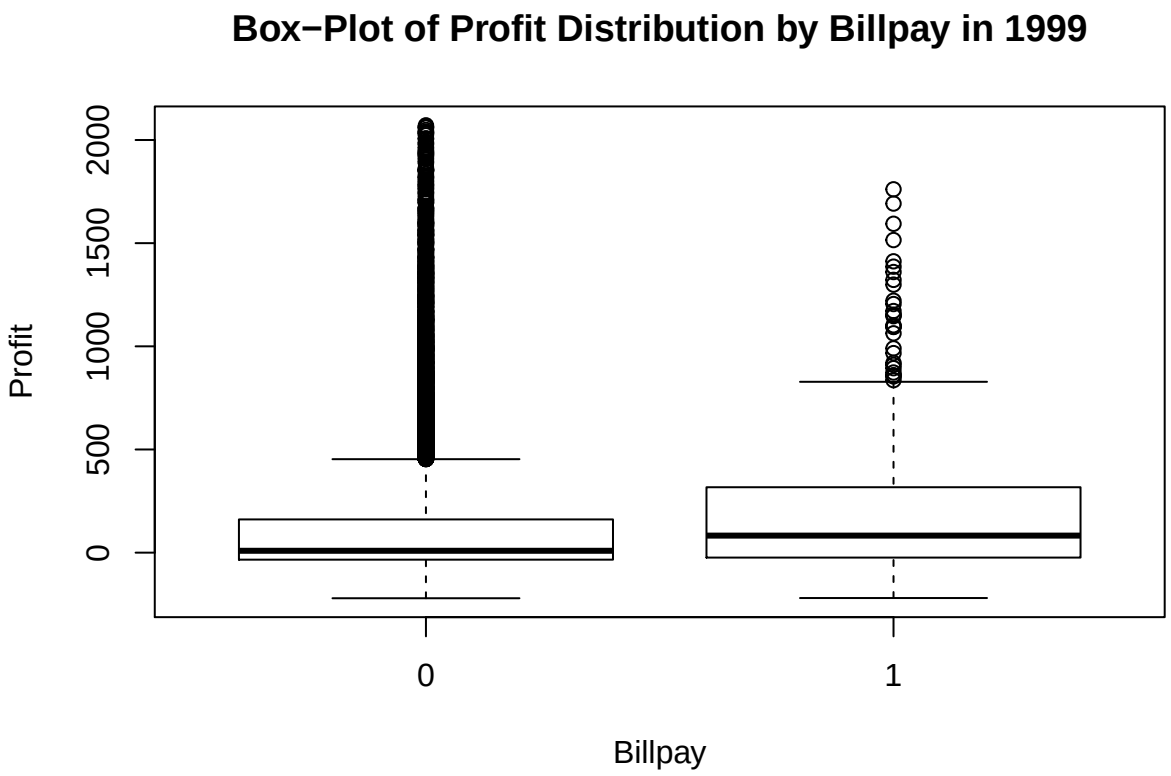

```
### This is a boxplot graph for Profit& Income
boxplot(X9Profit~X9Inc, data = consumerDB,
        main = "Box-Plot of Profit Distribution by Income in 1999",  ## Sets Title to Plot
        xlab = "Income", ylab = "Profit") ### Sets X and Y Axes
```



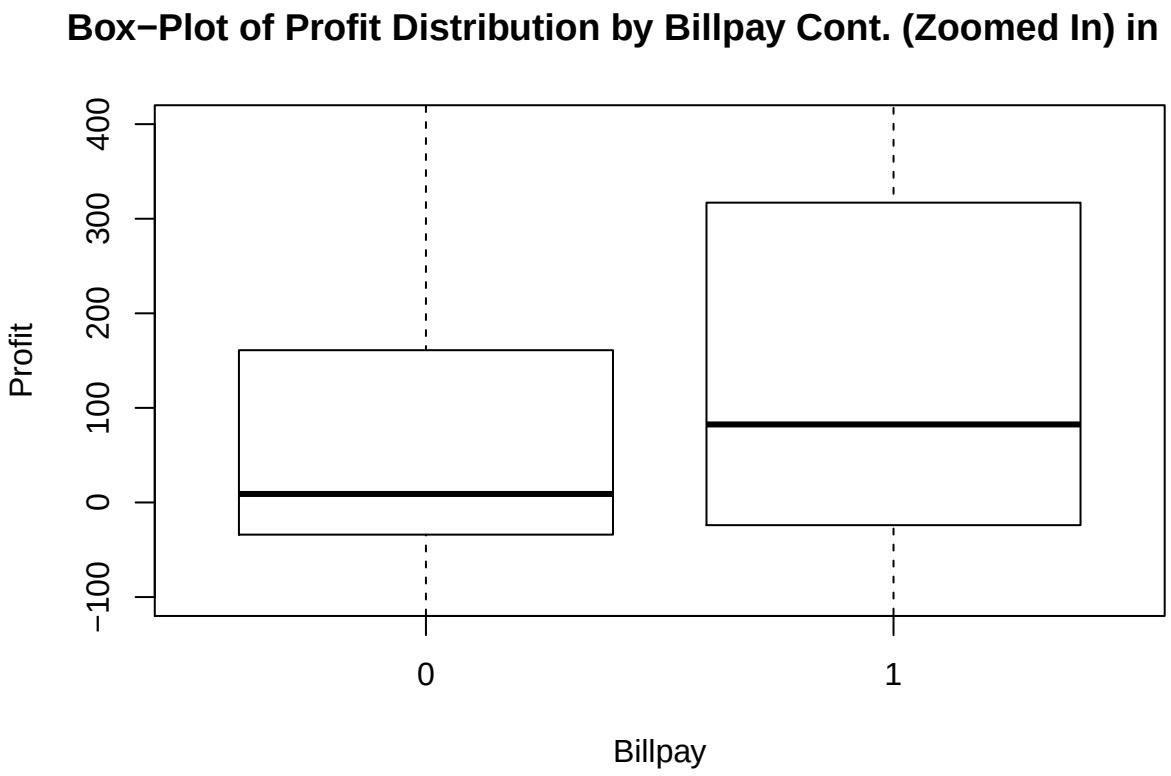
```
### Profit& Income(Zoomed in)
boxplot(X9Profit~X9Inc, data = consumerDB,
        main = "Box-Plot of Profit Distribution by Income Cont.(Zoomed In) in 1999",  ## Sets Title to
        xlab = "Income", ylab = "Profit",ylim=c(-100,400)) ### Sets X and Y Axes
```



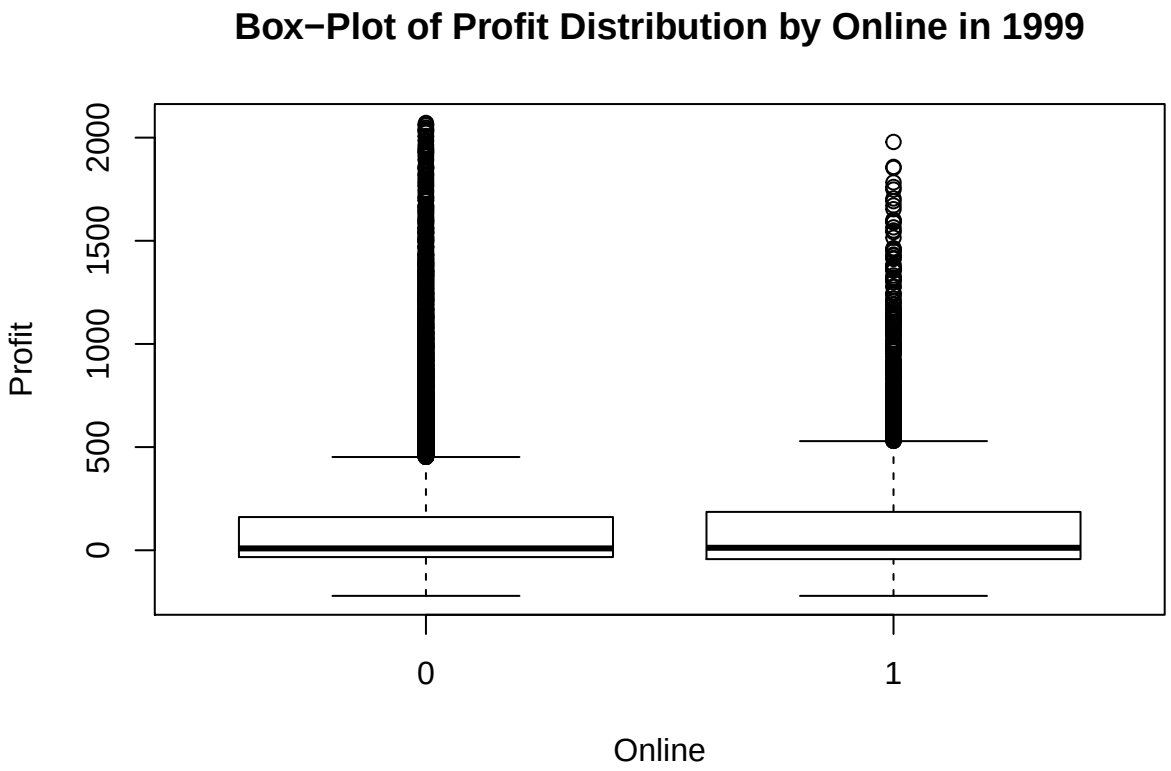

```
### This is a boxplot graph for Profit& Billpay
boxplot(X9Profit~X9Billpay, data = consumerDB,
        main = "Box-Plot of Profit Distribution by Billpay in 1999",  ## Sets Title to Plot
        xlab = "Billpay", ylab = "Profit") ### Sets X and Y Axes
```



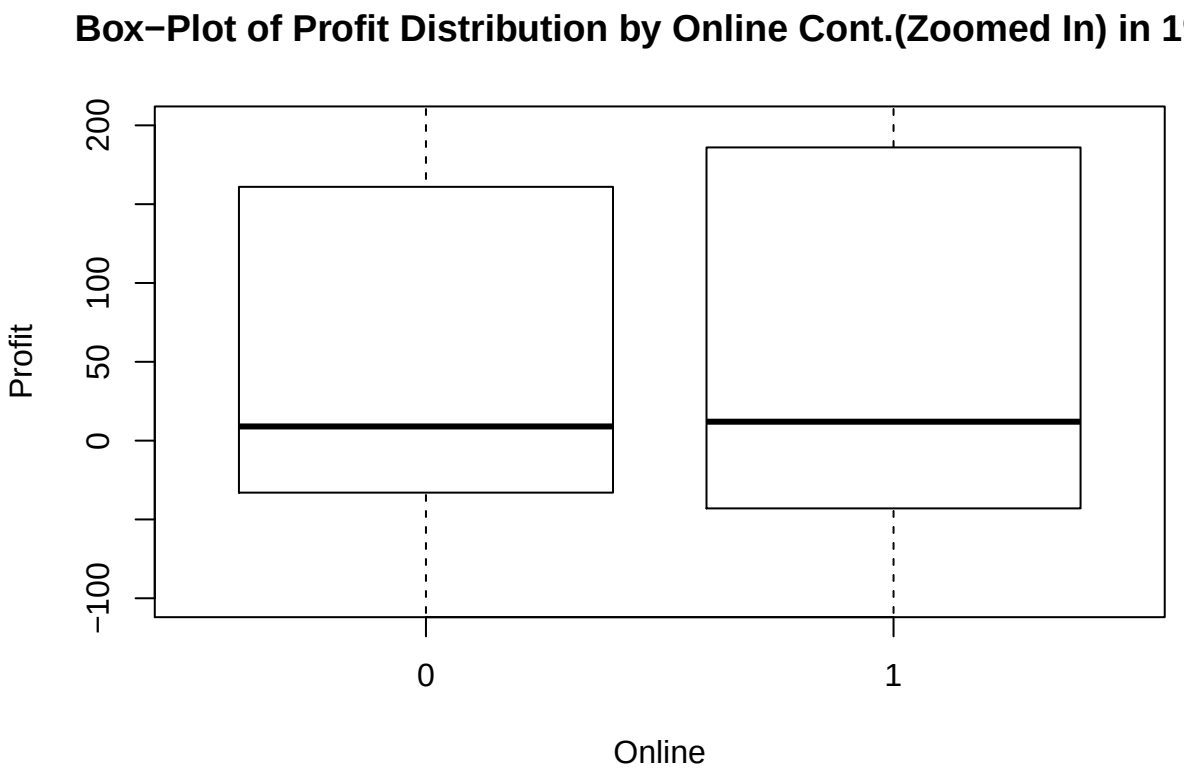

```
### Profit& Billpay(Zoomed in)
boxplot(X9Profit~X9Billpay, data = consumerDB,
        main = "Box-Plot of Profit Distribution by Billpay Cont. (Zoomed In) in 1999",  ## Sets Title t
        xlab = "Billpay", ylab = "Profit",ylim=c(-100,400)) ### Sets X and Y Axes
```



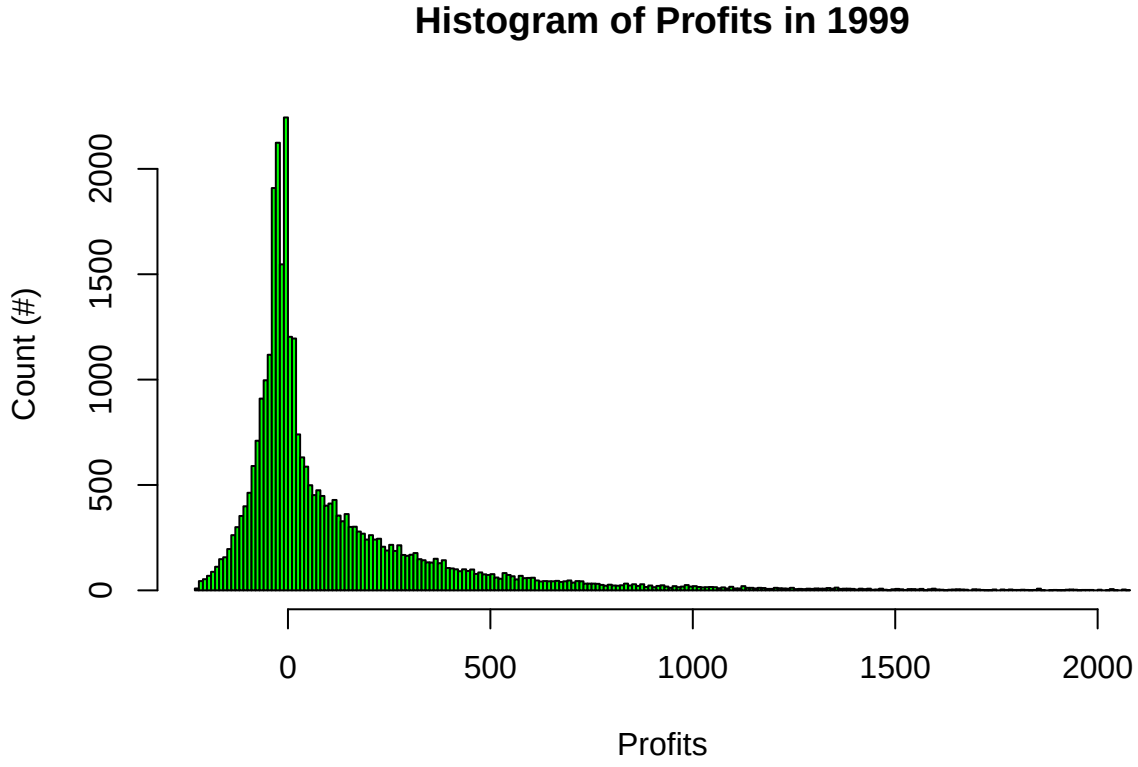
```
### This is a boxplot graph for Profit& Online
boxplot(X9Profit~X9Online, data = consumerDB,
        main = "Box-Plot of Profit Distribution by Online in 1999",  ## Sets Title to Plot
        xlab = "Online", ylab = "Profit") ### Sets X and Y Axes
```




```
### Profit& online(Zoomed in)
boxplot(X9Profit~X9Online, data = consumerDB,
        main = "Box-Plot of Profit Distribution by Online Cont.(Zoomed In) in 1999",  ## Sets Title to
        xlab = "Online", ylab = "Profit",ylim=c(-100,200)) ### Sets X and Y Axes
```

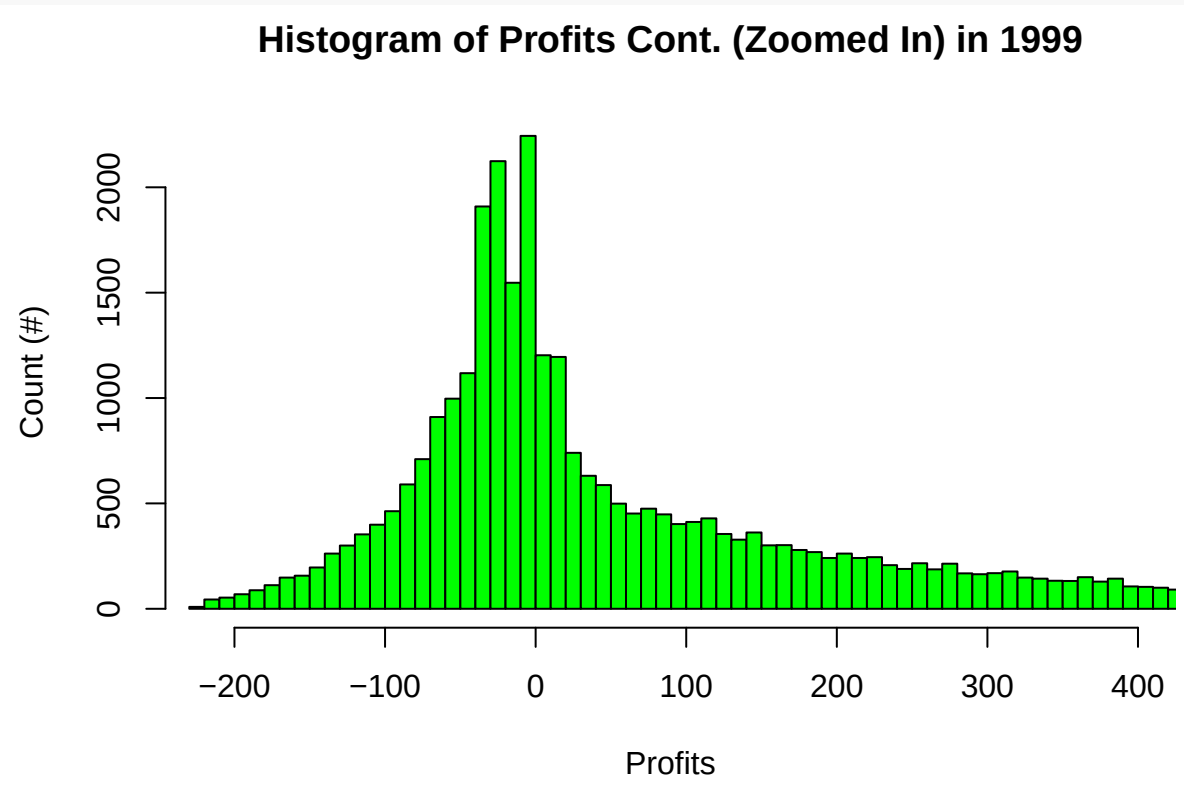



```
### Histogram of Profits
hist(consumerDB$X9Profit, main = "Histogram of Profits in 1999",
     xlab = "Profits", ylab = "Count (#)", col = "green", n = 200)
```

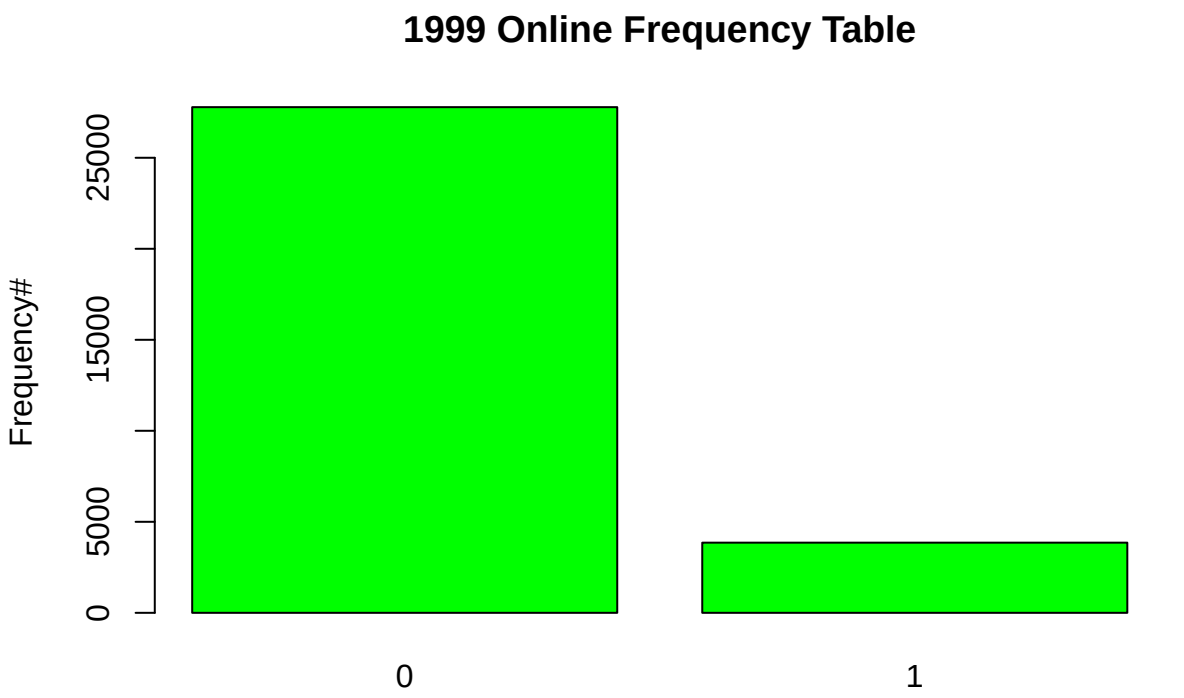



```
### Histogram of Profits (Zoomed in)
hist(consumerDB$X9Profit, main = "Histogram of Profits Cont. (Zoomed In) in 1999",
```

```
     xlab = "Profits", ylab = "Count (#)", col = "green", xlim = c(-221,400), n = 200)
```

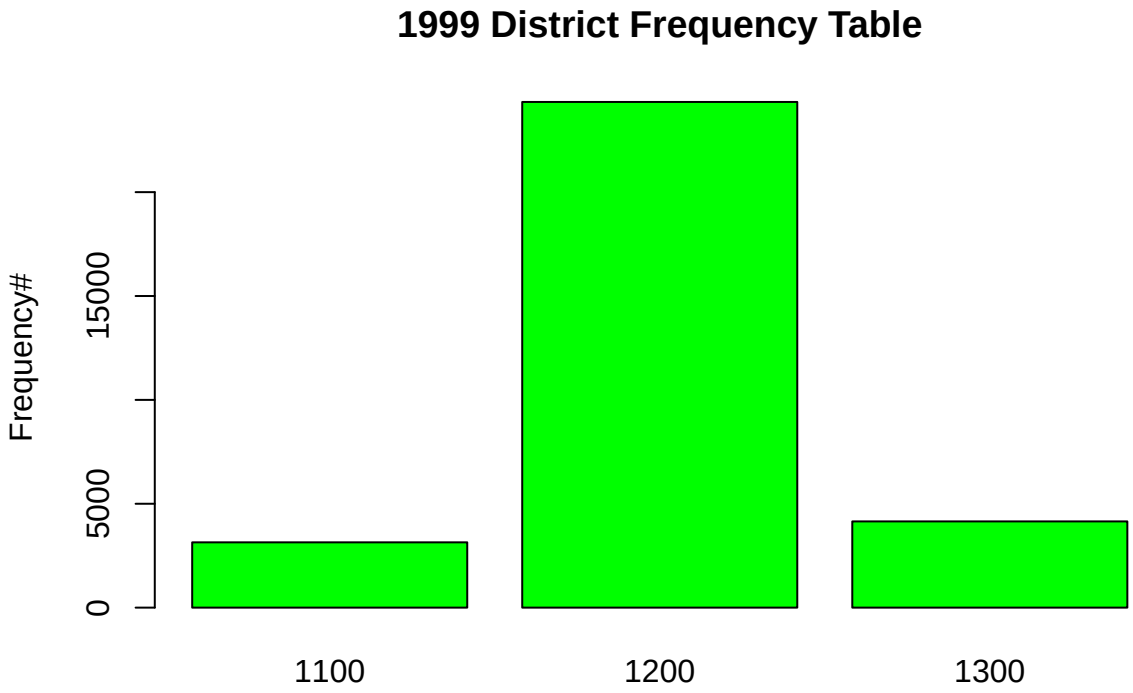



```
### This boxplot describes the frequency of Online consumer in 1999
barplot(table(consumerDB$X9Online),main="1999 Online Frequency Table",ylab = "Frequency#",col= "green")
```

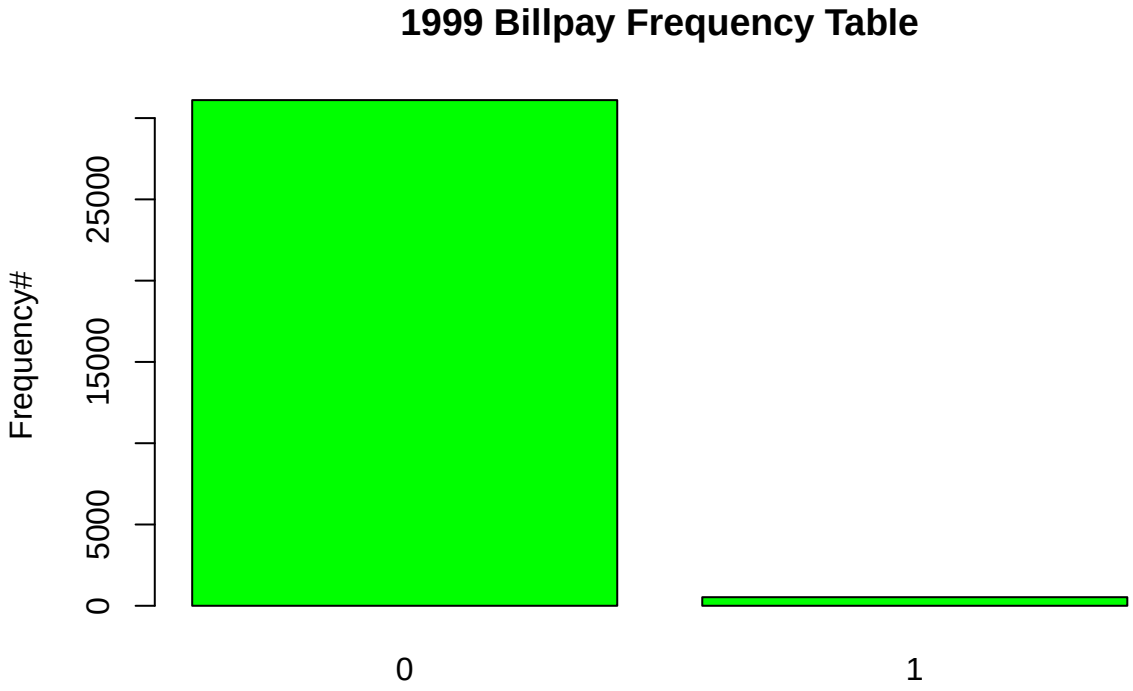



```
### table(consumerDB$X9Online)

### This boxplot describes the frequency of District in 1999
```

```r
barplot(table(consumerDB$X9District),main="1999 District Frequency Table",ylab = "Frequency#",col= "gre
```




```r
### table(consumerDB$X9District)

### This boxplot describes the frequency of Billpay consumer in 1999
barplot(table(consumerDB$X9Billpay),main="1999 Billpay Frequency Table",ylab = "Frequency#",col= "green
```




```r
### table(consumerDB$X9Billpay)

library(plyr)
### 1999 Income with online and billpay customer
```

```
### Note: NA represents those consumers that left the bank in 2000 and didn't have information in 1999 i
summary_Income_1999 = ddply(consumerDB, .(X9Inc),
                            summarize, '0' = sum(X9Online,X9Online==0), '1' = sum(X9Online,X9Online==1))
summary_Income_1999
```

```
##    X9Inc     0    1
## 1      1  2044  320
## 2      2   810  132
## 3      3  2571  568
## 4      4  2312  494
## 5      5  2369  606
## 6      6 10249 2458
## 7      7  3152  934
## 8      8  1742  542
## 9      9  2960 1042
## 10    NA  3425  612
```

```
### 1999 Age with online and billpay customer
### Note: NA represents those consumers that left the bank in 2000 and didn't have information in 1999 a
summary_Age_1999 = ddply(consumerDB, .(X9Age),
                            summarize, '0' = sum(X9Online,X9Online==0), '1' = sum(X9Online,X9Online==1))
summary_Age_1999
```

```
##   X9Age     0    1
## 1     1   710  274
## 2     2  3650 1572
## 3     3  5390 1668
## 4     4 10229 2540
## 5     5  3236  612
## 6     6  2290  230
## 7     7  2693  196
## 8    NA  3436  616
```

```
### Creating Tenure levels to different 14 levels
### The range of Tenure is 0 to 42
consumerDB$Tenure.Level = "0"
consumerDB[consumerDB$X9Tenure <= 3.00,]$Tenure.Level = "Level 01"
consumerDB[consumerDB$X9Tenure > 3.00 & consumerDB$X9Tenure <= 6.00,]$Tenure.Level = "Level 02"
consumerDB[consumerDB$X9Tenure > 6.00 & consumerDB$X9Tenure <= 9.00,]$Tenure.Level = "Level 03"
consumerDB[consumerDB$X9Tenure > 9.00 & consumerDB$X9Tenure <= 12.00,]$Tenure.Level = "Level 04"
consumerDB[consumerDB$X9Tenure > 12.00 & consumerDB$X9Tenure <= 15.00,]$Tenure.Level = "Level 05"
consumerDB[consumerDB$X9Tenure > 15.00 & consumerDB$X9Tenure <= 18.00,]$Tenure.Level = "Level 06"
consumerDB[consumerDB$X9Tenure > 18.00 & consumerDB$X9Tenure <= 21.00,]$Tenure.Level = "Level 07"
consumerDB[consumerDB$X9Tenure > 21.00 & consumerDB$X9Tenure <= 24.00,]$Tenure.Level = "Level 08"
consumerDB[consumerDB$X9Tenure > 24.00 & consumerDB$X9Tenure <= 27.00,]$Tenure.Level = "Level 09"
consumerDB[consumerDB$X9Tenure > 27.00 & consumerDB$X9Tenure <= 30.00,]$Tenure.Level = "Level 10"
consumerDB[consumerDB$X9Tenure > 30.00 & consumerDB$X9Tenure <= 33.00,]$Tenure.Level = "Level 11"
consumerDB[consumerDB$X9Tenure > 33.00 & consumerDB$X9Tenure <= 36.00,]$Tenure.Level = "Level 12"
consumerDB[consumerDB$X9Tenure > 36.00 & consumerDB$X9Tenure <= 39.00,]$Tenure.Level = "Level 13"
consumerDB[consumerDB$X9Tenure > 39.00 & consumerDB$X9Tenure <= 42.00,]$Tenure.Level = "Level 14"

### Format the table and present a piovt table with online consumers in different tenure levels
table(consumerDB$Tenure.Level,consumerDB$X9Online)
```

```
##
```

```
##               0    1
##   Level 01 5517  799
##   Level 02 5777  864
##   Level 03 4476  849
##   Level 04 2931  426
##   Level 05 2026  276
##   Level 06 1760  204
##   Level 07 1379  150
##   Level 08 1287  121
##   Level 09  872   65
##   Level 10  745   59
##   Level 11  532   26
##   Level 12  259    9
##   Level 13  131    4
##   Level 14   88    2
```

```
### Format the table and present a piovt table with billpay consumers in different tenure levels
table(consumerDB$Tenure.Level,consumerDB$X9Billpay)
```

```
##
##               0    1
##   Level 01 6222   94
##   Level 02 6511  130
##   Level 03 5220  105
##   Level 04 3292   65
##   Level 05 2262   40
##   Level 06 1939   25
##   Level 07 1507   22
##   Level 08 1391   17
##   Level 09  922   15
##   Level 10  795    9
##   Level 11  553    5
##   Level 12  268    0
##   Level 13  134    1
##   Level 14   90    0
```

Correlation Statistics

```
### Sort the table and ignore those data points that miss 1999 income or age value
### Name a new consumer Database "consumerDB2"
consumerDB2=consumerDB[!is.na(consumerDB$X9Inc),]
consumerDB2=consumerDB2[!is.na(consumerDB2$X9Age),]

### Calculate the profit's corrlation with other factors
cor(x=consumerDB2$X9Profit, y=consumerDB2$X9Inc)
```

```
## [1] 0.1299013
```

```
cor(x=consumerDB2$X9Profit, y=consumerDB2$X9Age)
```

```
## [1] 0.1350308
```

```
cor(x=consumerDB$X9Profit, y=consumerDB$X9Online)
```

```
## [1] 0.007049992
```

```r
cor(x=consumerDB$X9Profit, y=consumerDB$X9Billpay)
```

```
## [1] 0.04006637
```

```r
cor(x=consumerDB$X9Profit, y=consumerDB$X9Tenure)
```

```
## [1] 0.191133
```

```r
cor(x=consumerDB$X9Profit, y=consumerDB$X9District)
```

```
## [1] 0.00309511
```

```r
### Creating Profit levels to different 13 levels
### The range of Profit is -221 to 2071
consumerDB$Profit.Level = "0" ### default value

### Divide the profits by 10 and create a smaller range of Profit
consumerDB$ProfitProcess = consumerDB$X9Profit/10
consumerDB[consumerDB$ProfitProcess <= 0,]$Profit.Level = "Level 01"
consumerDB[consumerDB$ProfitProcess > 0 & consumerDB$ProfitProcess <= 10,]$Profit.Level = "Level 02"
consumerDB[consumerDB$ProfitProcess > 10 & consumerDB$ProfitProcess <= 20,]$Profit.Level = "Level 03"
consumerDB[consumerDB$ProfitProcess > 20 & consumerDB$ProfitProcess <= 40,]$Profit.Level = "Level 04"
consumerDB[consumerDB$ProfitProcess > 40 & consumerDB$ProfitProcess <= 60,]$Profit.Level = "Level 05"
consumerDB[consumerDB$ProfitProcess > 60 & consumerDB$ProfitProcess <= 80,]$Profit.Level = "Level 06"
consumerDB[consumerDB$ProfitProcess > 80 & consumerDB$ProfitProcess <= 100,]$Profit.Level = "Level 07"
consumerDB[consumerDB$ProfitProcess > 100 & consumerDB$ProfitProcess <= 120,]$Profit.Level = "Level 08"
consumerDB[consumerDB$ProfitProcess > 120 & consumerDB$ProfitProcess <= 140,]$Profit.Level = "Level 09"
consumerDB[consumerDB$ProfitProcess > 140 & consumerDB$ProfitProcess <= 160,]$Profit.Level = "Level 10"
consumerDB[consumerDB$ProfitProcess > 160 & consumerDB$ProfitProcess <= 180,]$Profit.Level = "Level 11"
consumerDB[consumerDB$ProfitProcess > 180 & consumerDB$ProfitProcess <= 200,]$Profit.Level = "Level 12"
consumerDB[consumerDB$ProfitProcess > 200,]$Profit.Level = "Level 13"

### Format the table and present a piovt table with online consumers in different profit levels
table(consumerDB$Profit.Level,consumerDB$X9Online)
```

```
##
##                0     1
##   Level 01 13009  1793
##   Level 02  5890   742
##   Level 03  2886   392
##   Level 04  3069   454
##   Level 05  1331   220
##   Level 06   667   113
##   Level 07   376    52
##   Level 08   213    41
##   Level 09   151    21
##   Level 10    86    15
##   Level 11    48     8
##   Level 12    38     3
##   Level 13    16     0
```

```r
### Format the table and present a piovt table with billpay consumers in different profit levels
table(consumerDB$Profit.Level,consumerDB$X9Billpay)
```

```
##
##                0     1
##   Level 01 14617   185
```

```
##   Level 02  6538    94
##   Level 03  3218    60
##   Level 04  3438    85
##   Level 05  1503    48
##   Level 06   757    23
##   Level 07   415    13
##   Level 08   245     9
##   Level 09   166     6
##   Level 10    98     3
##   Level 11    54     2
##   Level 12    41     0
##   Level 13    16     0
```

render("FirstProject_v10.Rmd",pdf_document())