DATA ANALYSIS OF CREDIT CARD CUSTOMERS

STATISTIC FINAL PROJECT

Report

submitted to fulfill the assignment in the Statistic course.

By :

Muhammad Farhan Akbar   (1301192246)

Khaidir Mauladan   (1301190000)

Rifky Aulia Putra Purwanto (1301200000)

Naufal Xeela Pandityatama (1301190000)

INFORMATICS MAJOR

SCHOOL OF COMPUTING

TELKOM UNIVERSITY

BANDUNG

2020

1. Credit Card Customer Characteristic

1.1. Reading Data from CSV File

Here is the code for reading a data from csv file using read.csv() method:

```
data <- read.csv("BankChurners.csv")
```

1.2. Attrition Flag Data

After we processing the data using R, we know that 1627 are Attrited Customer and 8500 existing customers. Below is the code to make bar chart for Attrition Flag data:

```
pie(table(data$Attrition_Flag), main = "Attrition Flag")
table(data$Attrition_Flag)
```

Attrition Flag
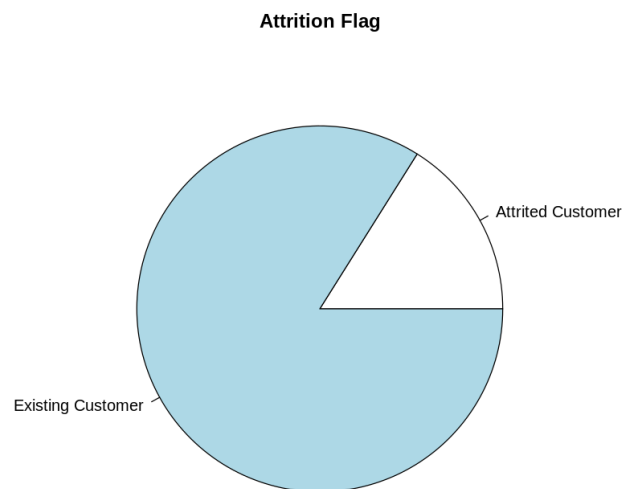
Attrited Customer: 1627

Existing Customer: 8500

1.3. Customer Age Data Distribution

For Customer Age data, we use histogram to visualize it and we use summary() method to know more information from Customer Age data. Here is the code and the histogram of Customer Age data distribution:

```
hist(data$Customer_Age, main = "Customer Age")
summary(data$Customer_Age)
#To Find mode
names(sort(-table(data$Customer_Age)))[1]
```
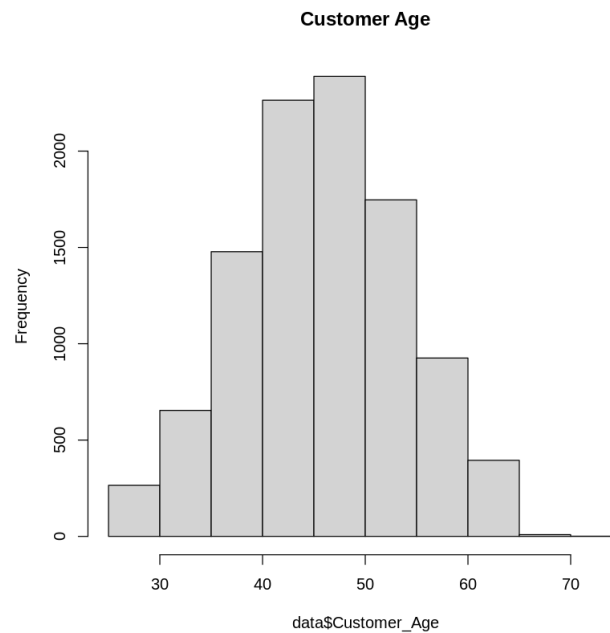
Customer Age

From the summary() method, we get:

Minimum = 26.00

1st Quartile = 41.00

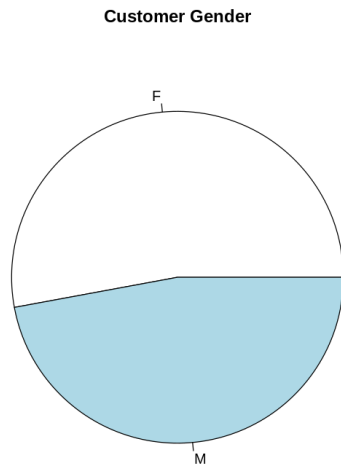Median = 46.00

Mean = 46.33

3rd Quartile = 52.00

Maximum = 73.00

Mode = 44.00

1.4. Customer Gender Data

Below are the code and the output of pie chart from Customer Gender data:

```
pie(table(data$Gender))
table(data$Gender)
```
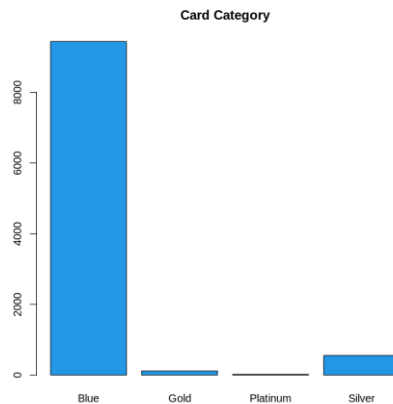
Customer Gender

We also know that there are 5358 females and 4769 male from 10127 customers.

1.5. Card Category Data

We also visualize and showing the information about Card Category data. Below are the code and the bar chart of Card Category Data:

```
barplot(table(data$Card_Category), col = 4, main = "Card Category")
table(data$Card_Category)
```
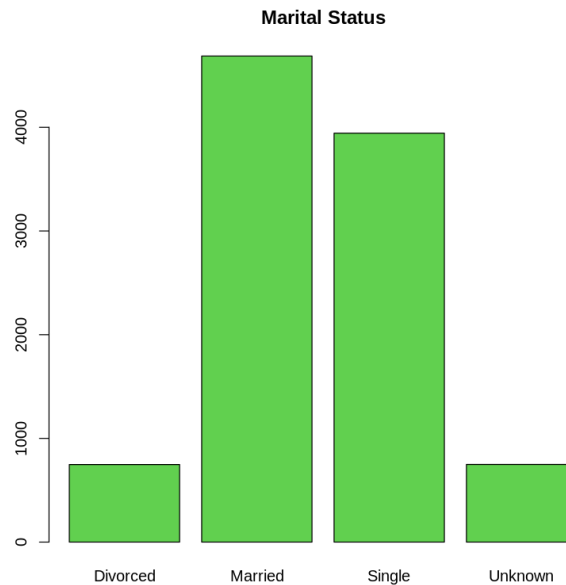
Card Category

Blue: 9436

Gold: 116

Platinum: 20

Silver: 555

1.6. Marital Status Data

Below are the code and the bar chart of Marital Status data:

```
barplot(table(data$Marital_Status), main = "Marital Status", col = 3)
table(data$Marital_Status)
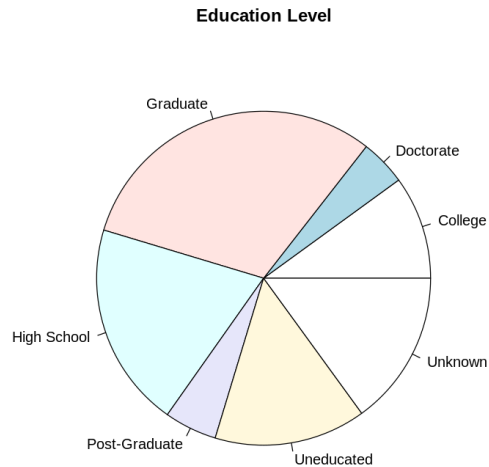```

**Marital Status**



Divorced: 748

Married: 4687

Single: 3943

Unknown: 749

## 1.7.  Education Level Data

We also know about the data distribution of customer Education Level. Below are the code and the pie chart of Education Level data:

```
pie(table(data$Education_Level), main = "Education Level")
table(data$Education_Level)
```

Education Level

College: 1013

Doctorate: 451
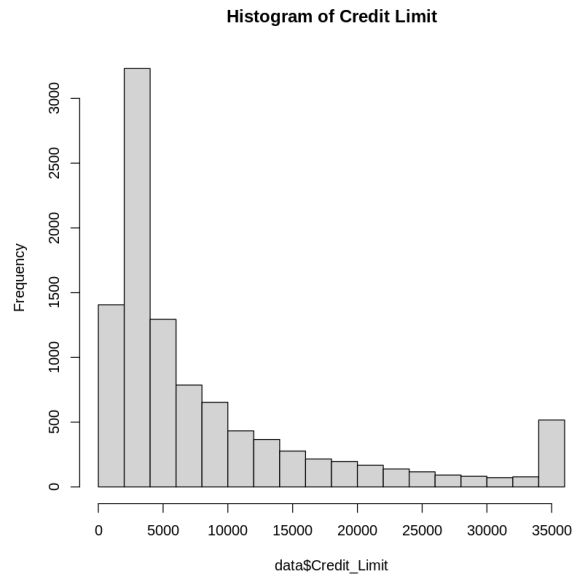
Graduate: 3128

High School: 2013

Post-graduate: 516

Uneducated: 1487

Unknown: 1519

1.8.   Credit Limit Data Distribution

For the Credit Limit data, we use histogram to visualize it and use summary() method to know more information about Credit limit. Here are the code and the output:

```
hist(data$Credit_Limit)
summary(data$Credit_Limit)
```

Histogram of Credit Limit

From the summary() method, we get:

Minimum = 1438

1st Quartile = 2555

Median = 4549

Mean = 8632

3rd Quartile = 11068

Maximum = 34516

1.9. Correlation Test and Scatter Plot of Credit Limit and Avg Open To Buy

We will show the correlation between Credit Limit and Avg Open To Buy using cor.test() method and visualize it using plot() method. Below are the code and the output for cor.test() method and plot() method:

```
cor.test(data$Credit_Limit, data$Avg_Open_To_Buy)
```

```
            Pearson's product-moment correlation

data:  data$Credit_Limit and data$Avg_Open_To_Buy
t = 1118.9, df = 10125, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.9958212 0.9961338
sample estimates:
       cor
0.9959805
```

```
plot(data$Credit_Limit, data$Avg_Open_To_Buy)
```
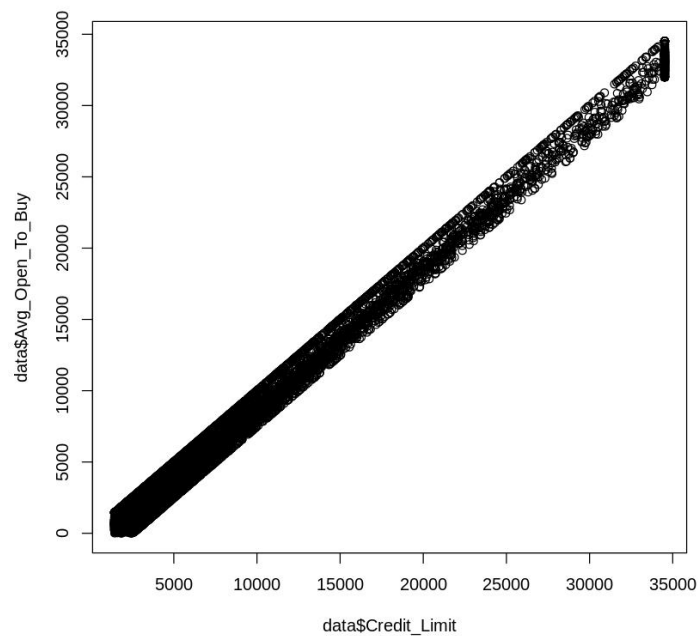

The correlation is 0.9959805. So, we can conclude that Credit Limit and Avg Open To Buy are very correlated.

2. Existing Customer Characteristic

```
[1]exst <- subset(data, Attrition_Flag == "Existing Customer")
```
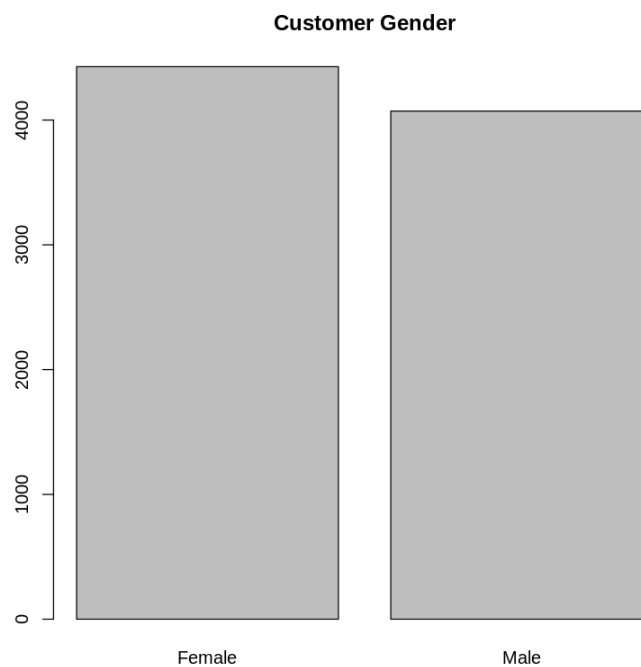
The code above is to read and subset "Existing Customer"

2.1. Customer Gender

The processed data resulted in 4428 female existing customers and 4072 male existing customers. The number of female customers is 8,75% higher than male customers.

```
[1]barplot(table(exst$Gender), main = "Customer Gender", names.ar
   g = c("Female", "Male"))
[2]table(exst$Gender)
```
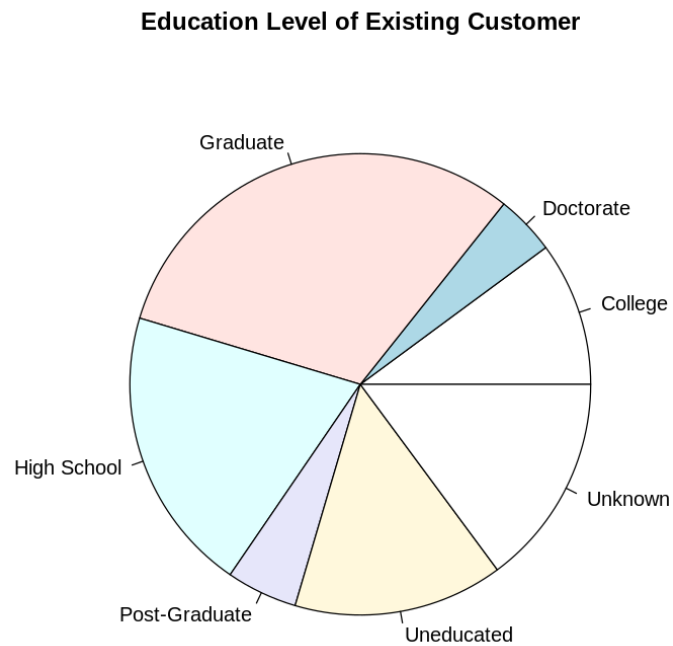
Customer Gender

2.2. Education Level

We compile recorded existing customer's education level which resulted in 2641 Graduates, 1707 High School students, 1263 unknown education, 1250 uneducated, 859 College students, 424 Post-graduates, and 356 Doctorates.
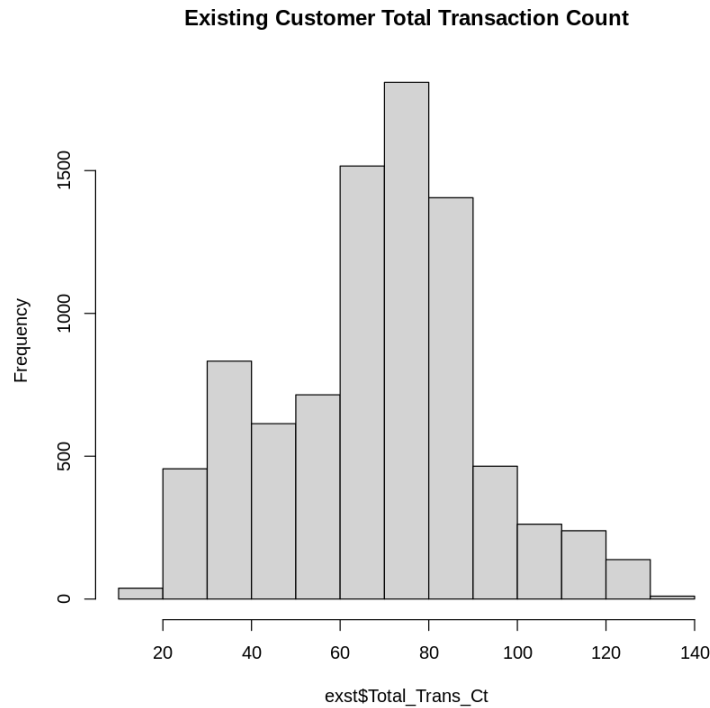
Education Level of Existing Customer

2.3. Total Transaction Count

The histogram below is the total transaction count of each existing customers. With the highest transaction count number of 139 transactions of a customer, lowest number of 11 transactions, with 1st quartile of 54 transactions, 3rd quartile of 82 transactions, and an average of 68,67 transactions, also with the median of 71 recorded transactions.

```
[1]hist(exst$Total_Trans_Ct, main = "Existing Customer Total Tran
   saction Count")
[2]summary(exst$Total_Trans_Ct)
```

Existing Customer Total Transaction Count

2.4. Customer Age

The histogram below is the recorded customer ages. With the 73 years of age for the oldest customer, 26 years of age for the youngest, with 1st quartile of 41 years of age, 3rd quartile of 52 years of age, and an average of 46,26 years of age, also with a median of 46 years old.

```
[1]hist(exst$Customer_Age, main = "Customer Age of Existing Custo
   mer")
[2]names(sort(-table(exst$Customer_Age)))[1]
[3]summary(exst$Customer_Age)
```

Customer Age of Existing Customer
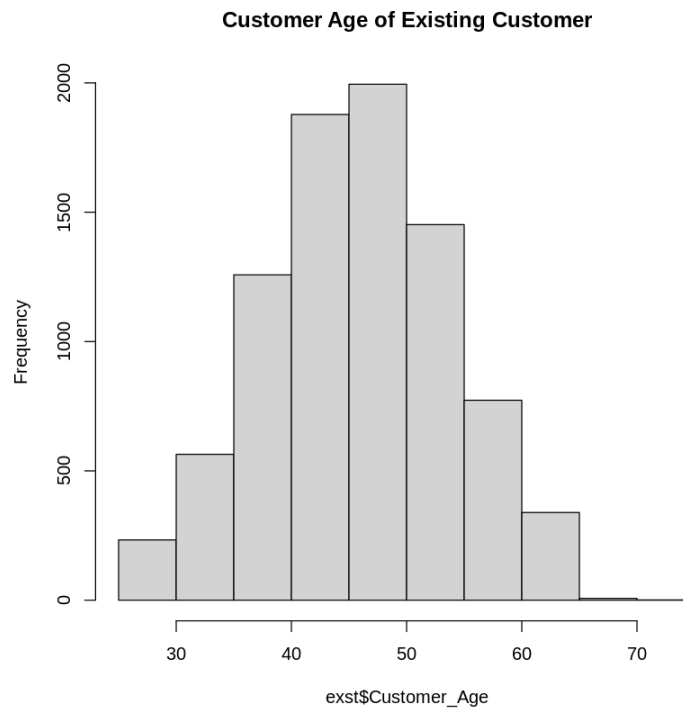
2.5. Total Transaction Amount

The boxplot below is the recorded amount of total transaction from existing customers. With the highest transaction of $18,484, $816 for the lowest transaction, with 1st quartile of $2,385, 3rd quartile of $4,781, and an average of $4,655, also a median of $4,100.

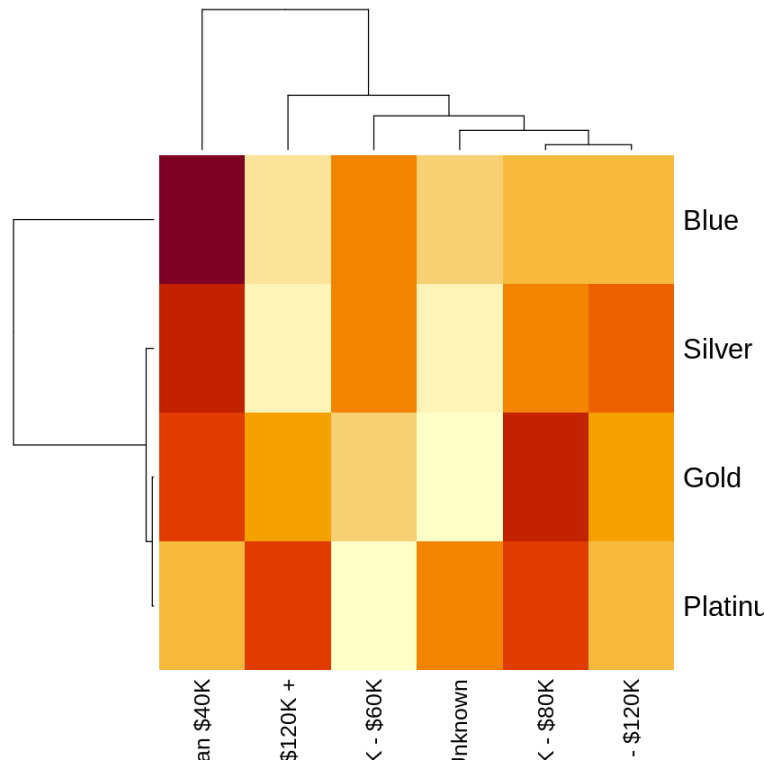```
[1]boxplot(exst$Total_Trans_Amt)
[2]summary(exst$Total_Trans_Amt)
```
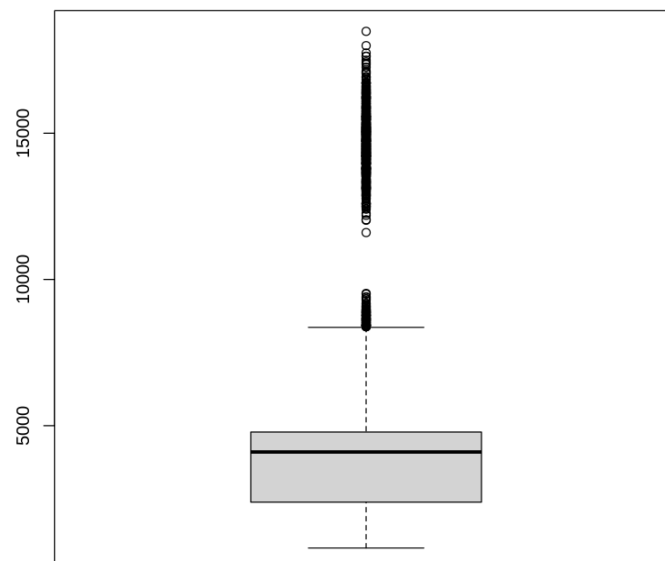
2.6. Card Type with Income

The heatmap below is the heat comparison for Card type on y-axis and income category on x-axis. With highest frequency of blue card customer with less than $40K income.

`[1]heatmap(table(exst$Card_Category, exst$Income_Category))`

3. Attrited Customer Characteristic

## 3.1 Separating Attrition Flag data to Attrited Customer only

Subsetting data from the Attrition Flag dataset so that it only shows the Attrited Customer dataset by using the subset() function.

```
att <- subset(data, Attrition_Flag == "Attrited Customer")
```

3.2 Customer Gender from Attrited Customer

By creating a bar chart from the Attrited Customer dataset, we can display the amount of customers by gender with 930 of them are female and 697 are male and conclude that the majority of the Attrited Customers are female. Below are the functions that were used to create a bar chart using the barplot() function and table() function:

```
barplot(table(att$Gender), col = 6)
table(att$Gender)
```
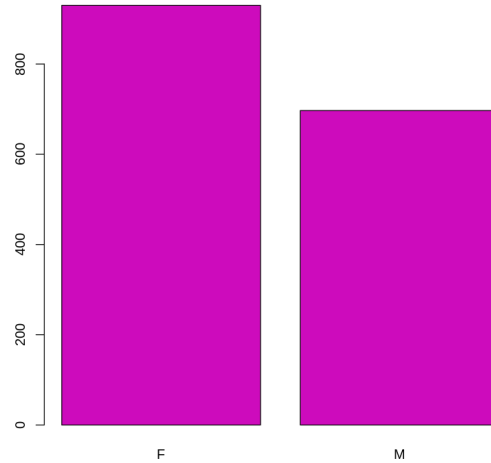

3.3 Customer Age from Attrited Customer

The histogram below displays the customer's age from the Attrited Customer dataset. By using certain functions, we can summarize that the minimum age value is 26 years old and the maximum age value is 68 years old. With the 1st quartile of 41 years old, 3rd quartile is 52 years old, a median of 47 years old, and with an average of 46.66 years old.

```
hist(att$Customer_Age, col = 7, main = "Customer Age from Attrited Customer")
#To find mode
names(sort(-table(att$Customer_Age)))[1]
summary(att$Customer_Age)
```
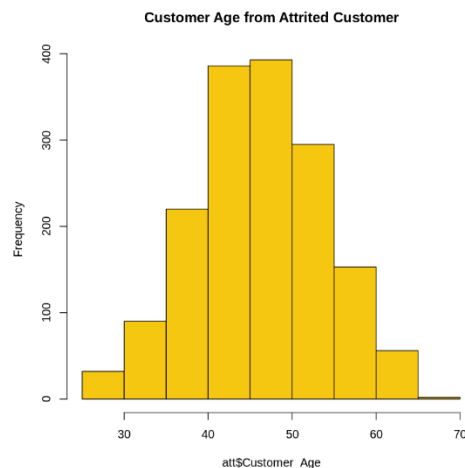
3.4 Education Level from Attrited Customer

The data that was extracted shows that there are various educational level from the Attrited customers. By using certan functions, we can show that there are 154 College students, 95 are Doctorate, 487 Graduates, 306 Highschool students, 92 are Post-Graduates, 237 are Uneducated, with 256 people have an Unknown background.

```
pie(table(att$Education_Level))
table(att$Education_Level)
```
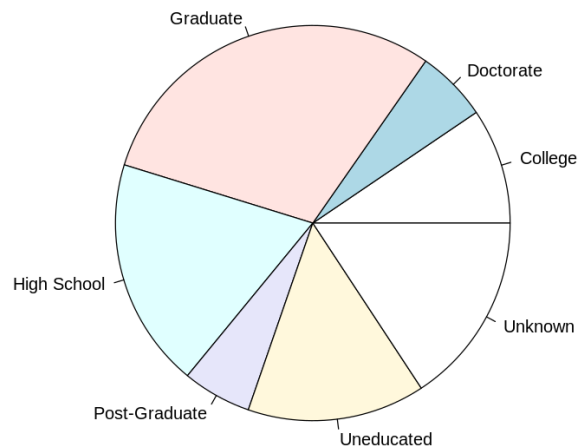

3.5 Total Transaction Count from Attrited Customer

Below is a histogram that displays the total transaction count of each attrited customers. With the lowest count of 10 transactions, and the highest is 94 transactions. With a 1st quartile of 37 transaction count, 3rd quartile of 51 transactions, a median of 43 transactions, and an average of 44.93 recorded transactions.

```
hist(att$Total_Trans_Ct, col = 10)
summary(att$Total_Trans_Ct)
```

Histogram of att$Total_Trans_Ct
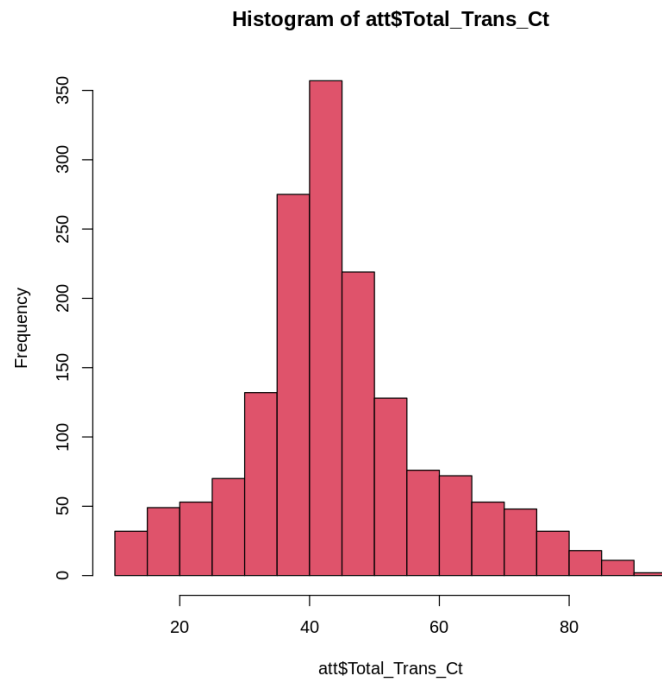
3.6 Total Transaction Amount from Attrited Customer

By using the boxplot chart, we can display and summarize the total transaction amount of attrited customers. With the lowest amount is $510, and the highest is $10,583. The dataset shows that the 1st quartile is valued at $1,904, 3rd quartile is $2,772, with a median of $2,329, and an average of $3,095.

```
boxplot(att$Total_Trans_Amt)
summary(att$Total_Trans_Amt)
```
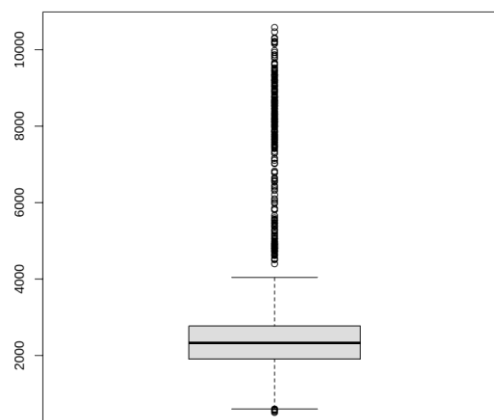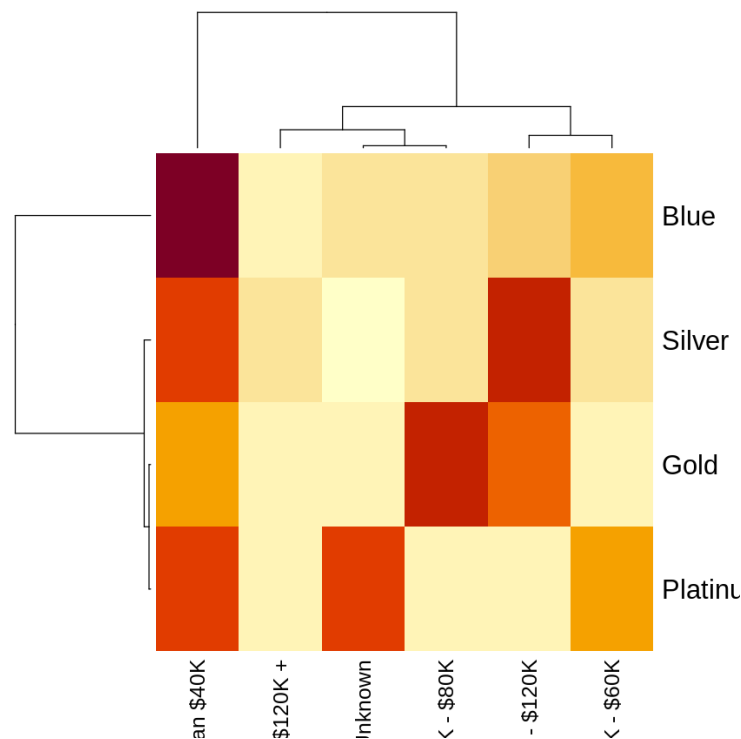
3.7 Card Type with Income of Attrited Customer

Below is a heatmap that displays the correlation between card type and customer's income. We can conclude that the warmest or the highest frequency are blue card customers with an income of <$40K or below than $40,000.

```
heatmap(table(att$Card_Category, att$Income_Category))
```
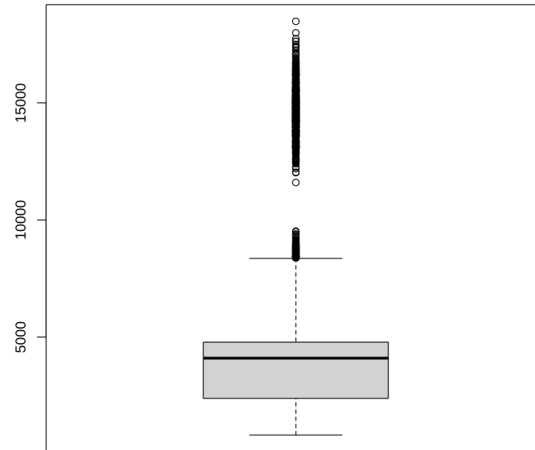

4. Comparing Between Existing and Attrited Customer Characteristic

4.1   Total Transaction

- Existing
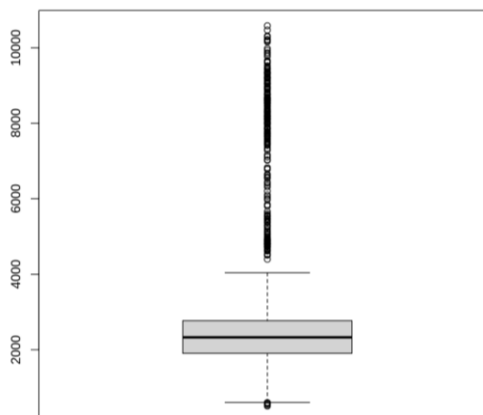
```
boxplot(exst$Total_Trans_Amt)
summary(exst$Total_Trans_Amt)
```

Minimal	: 816
1st Quartile	: 2385
Median	: 4100
Mean	: 4655
3rd Quartile	: 4781
Max	: 18484

- Attrited

```
boxplot(att$Total_Trans_Amt)
summary(att$Total_Trans_Amt)
```


Minimal	: 510
1st Quartile	: 1904
Median	: 2329
Mean	: 3095
3rd Quartile	: 272
Maximum	: 10583

In the total transaction amount category, it can be concluded that existing customer transfer more money than the attrited customer. This can be shown from the above diagram that existing customer transfer an average 4655 $ in the last 12 months, whereas attrited customer only transfer an average 3095 $ in the last 12 months.
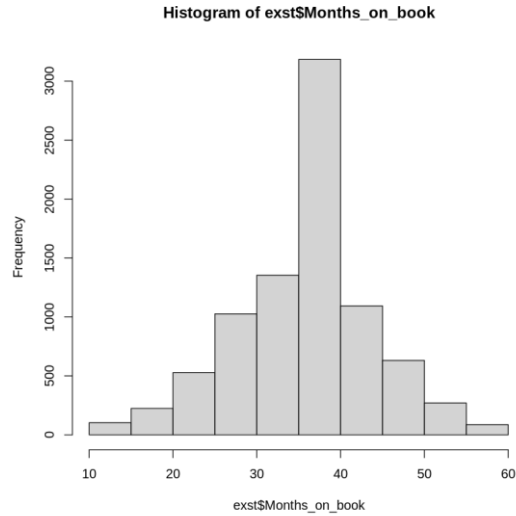
4.2.   Months on Book

- Attrited

```
hist(att$Months_on_book)
summary(att$Months_on_book)
```

Histogram of att$Months_on_book
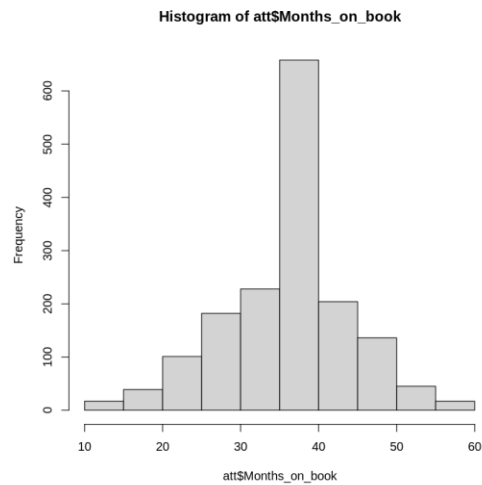
Minimal	: 13.00
1st Quartile	: 32.00
Median	: 36.00
Mean	: 36.18
3rd Quartile	: 40.00
Max	: 56.00

- Existing

```
hist(exst$Months_on_book)
summary(exst$Months_on_book)
```

Minimal   : 13.00
1st Quartile   : 31.00
Median   : 36.00
Mean   : 35.88
3rd Quartile   : 40.00
Maximum   : 56.00

In the months on book category, the data shows the longer the costumer used a certain bank the more likely that costumer will likely to attrited from that bank. This is shown where the mean for membership of an attirited costumer is 36.18 months, whereas an existing costumer is slightly lower at 35.88 months.
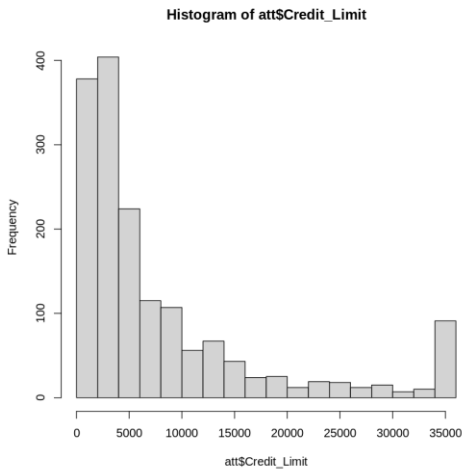
4.3.   Credit Limit

- Attrited

```
hist(att$Credit_Limit)
summary(att$Credit_Limit)
```

Histogram of att$Credit_Limit
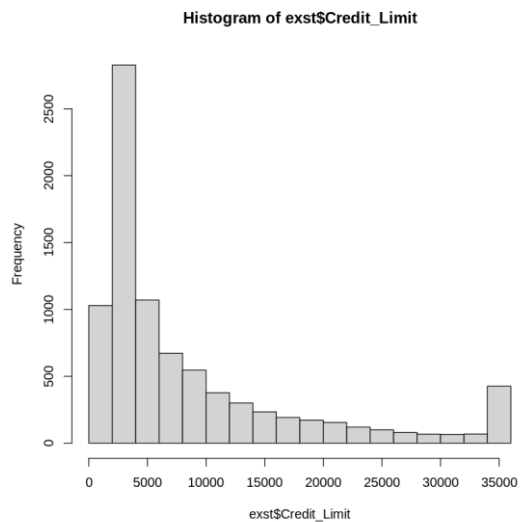

Minimal   : 1438
1st Quartile   : 2114
Median   : 4178
Mean   : 8136
3rd Quartile   : 9934
Maximum   : 34516

- Existing

```
hist(exst$Credit_Limit)
summary(exst$Credit_Limit)
```

Histogram of exst$Credit_Limit

Minimal   : 1438
1st Quartile   : 2602
Median   : 4644
Mean   : 8727
3rd Quartile   : 11253
Maximum   : 34516

In the Credit Limit category existing costumer have a higher credit limit than attrited costumer. This can be shown where the average credit limit for an existing costumer is 8727$ whereas credit limit for an attrited costumer is 8136$. Existing costumer have a higher credit limit is also because existing costumer have a longer month on book than attrited costumer, therefore they have more time to increase their credit limit.

4.4. Average Open to Buy

- Attrited
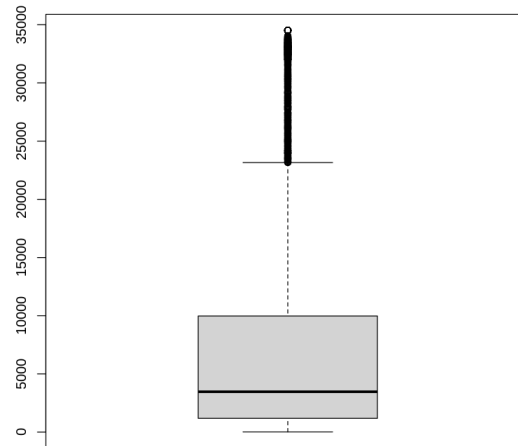
```
summary(att$Avg_Open_To_Buy)
boxplot(att$Avg_Open_To_Buy)
```
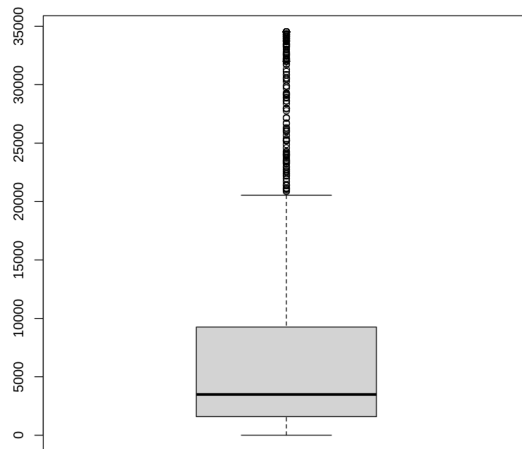

Minimal	: 3
1st Quartile	: 1587
Median	: 3488
Mean	: 7463
3rd Quartile	: 9258
Max	: 34516

- Existing

```
summary(exst$Avg_Open_To_Buy)
boxplot(exst$Avg_Open_To_Buy)
```

Minimal	: 15
1st Quartile	: 1184
Median	: 3470
Mean	: 7470
3rd Quartile	: 9978
Maximum	: 34516

In the average open to buy category attrited costumer have a lower average open to buy than an existing costumer. Attrited costumer have an average 7463$ of average open to buy, whereas existing costumer have an average 7470$. This is partly due to the low credit limit that an attrited costumer have, therefore lowering the average open to buy for an attrited costumer
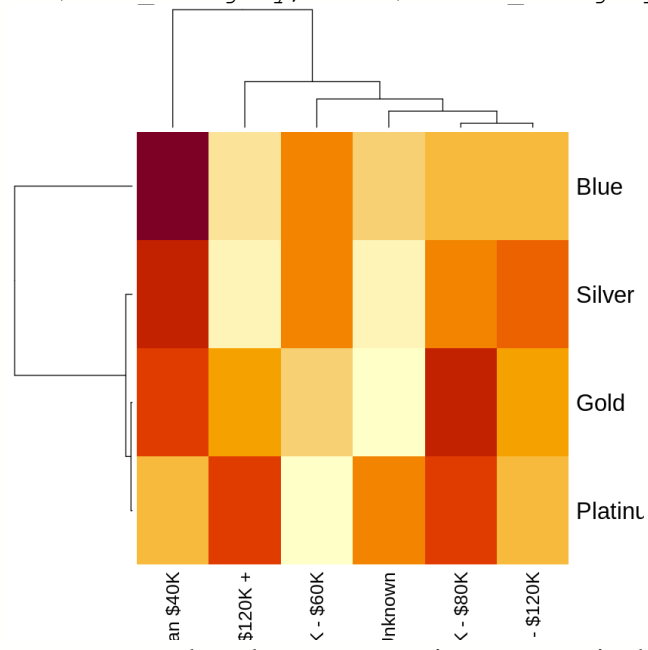
4.5. Income and Card Category

- Attrited

```
heatmap(table(att$Card_Category, att$Income_Category))
```
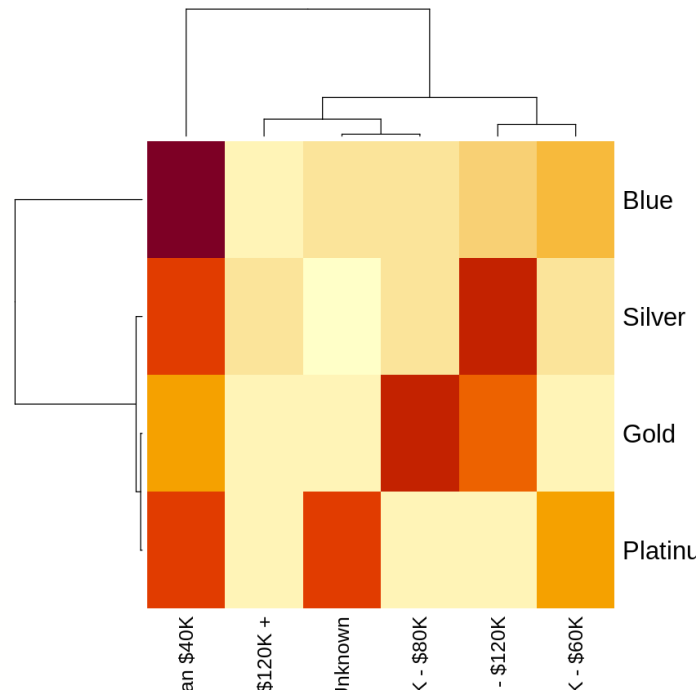
- Existing

```
heatmap(table(exst$Card_Category, exst$Income_Category))
```


In the income category and card category section most attrited have a blue category card and an income below 40k$, whereas existing costumer are more spread out evenly in terms of card category and income category