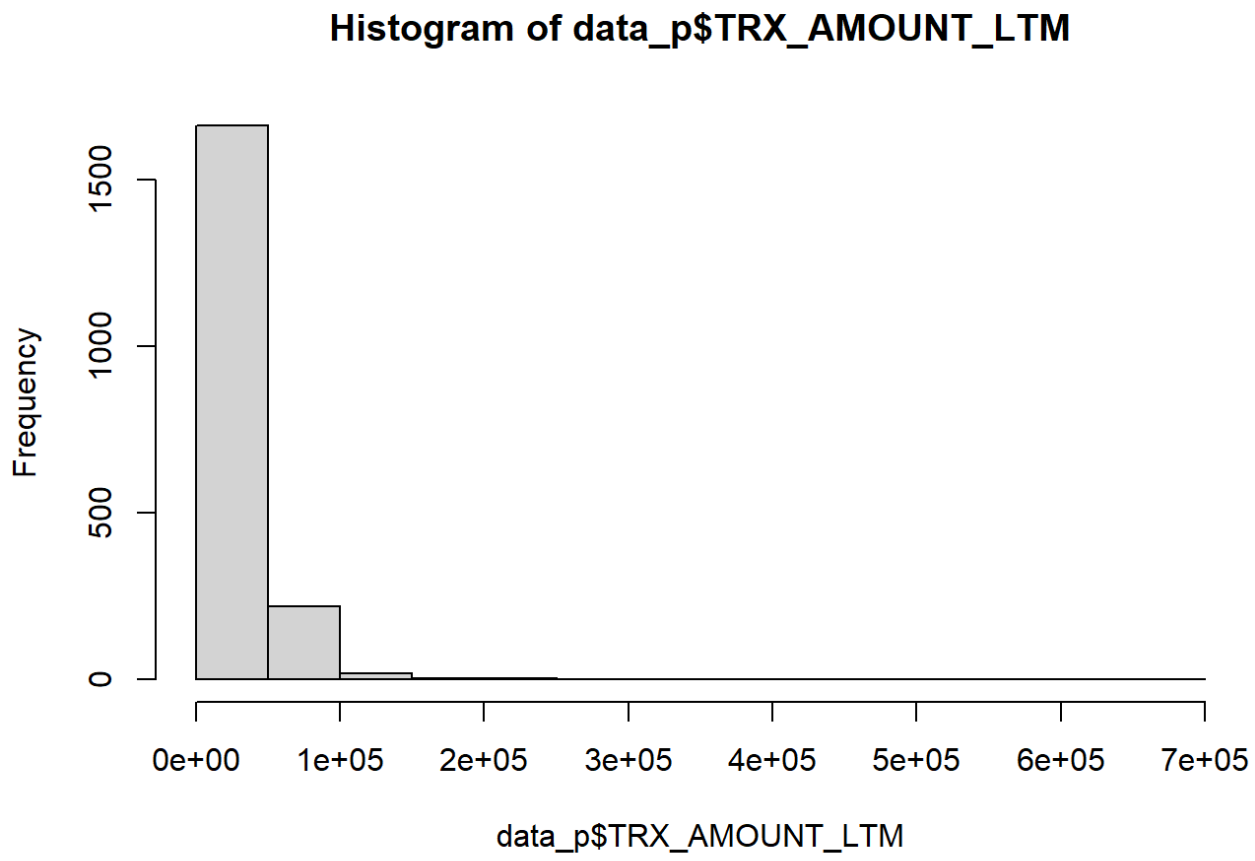
Analiz Portofolio

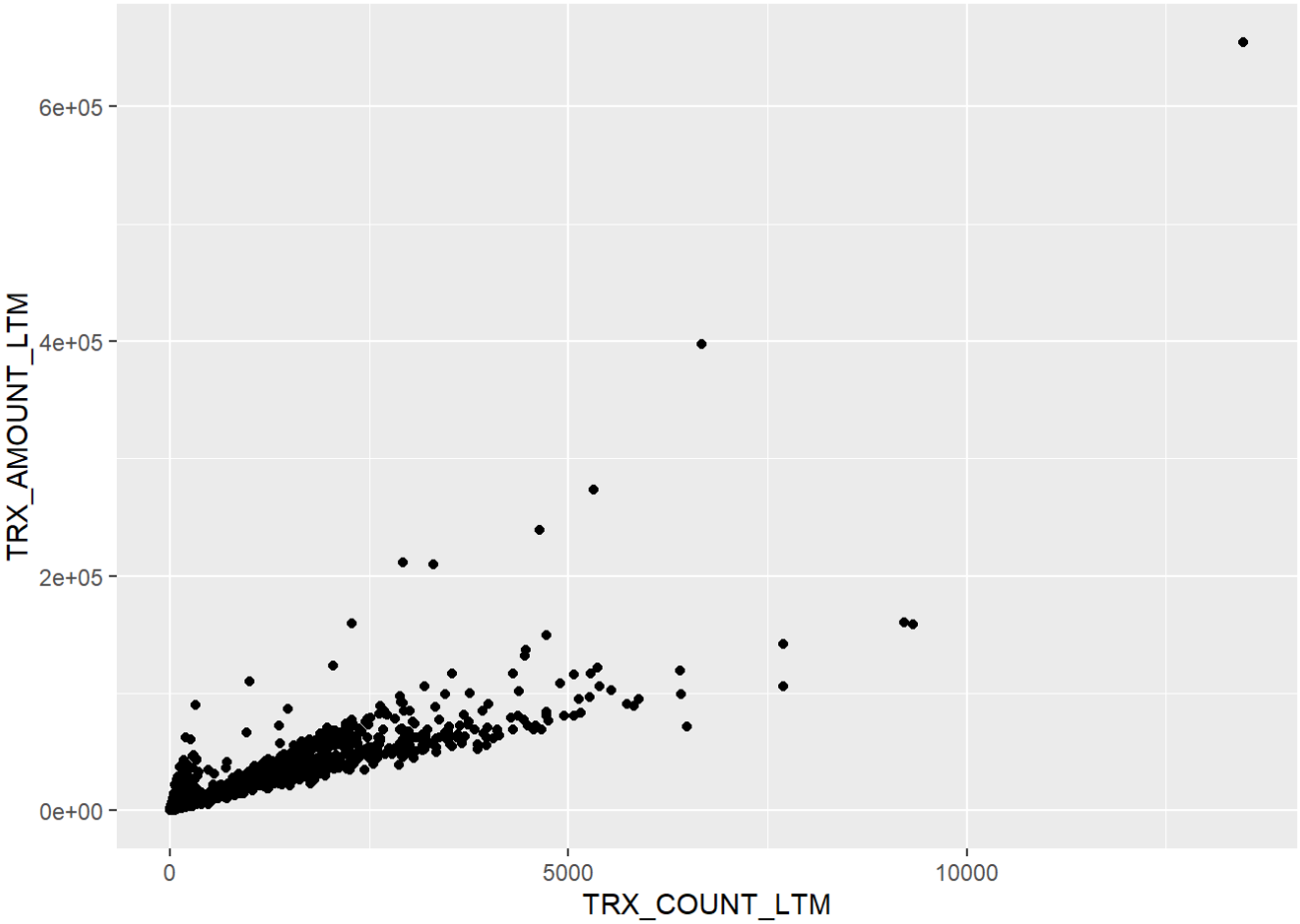
```
##Read data
data <- read.csv("Portfolio_for_test1.csv")

###Data validation
data_p <- subset(data,
                 select = -c(CB_COUNT,CB_AMOUNT,FRAUD_COUNT,FRAUD_AMOUNT))
head(data_p)
```

```
##         GROUPID       CLIENTID  GROUPNAME MERCHANT_NAME  CLIENTNAME  MCC
## 1  Group_ID3005  Client_ID9130  Group2867  Merchant2804  Client8159 5977
## 2 Group_ID16605 Client_ID26649 Group15836 Merchant15358 Client23056 5977
## 3  Group_ID2848  Client_ID8964  Group2724  Merchant2656  Client8013 5977
## 4  Group_ID3742 Client_ID10045  Group3595  Merchant3533  Client8975 5977
## 5  Group_ID2895  Client_ID9013  Group2767  Merchant2702  Client8055 5977
## 6  Group_ID2976  Client_ID9100  Group2841  Merchant2778  Client8133 5977
##   ONLINEPOS   SCHEME TRX_COUNT TRX_AMOUNT TRX_COUNT_LTM TRX_AMOUNT_LTM      MSC
## 1       POS     Visa         0       0.00           355      10096.920   0.0000
## 2       POS       MC       366   31477.25           310      26161.911 244.3526
## 3       POS Other_CC        27     448.33           171       2648.158   1.4130
## 4       POS     Visa         0       0.00           110       4038.731   0.0000
## 5       POS       MC         0       0.00           955      31179.243   0.0000
## 6       POS       MC         0       0.00          3166      55353.234   0.0000
##      MSC_LTM    NET_MSC NET_MSC_LTM      ATV      ID REFUND_COUNT REFUND_AMOUNT
## 1  34.345726   0.000000   -9.401826 28.40433 1980627            0             0
## 2 196.352905 131.764192  106.307628 84.41740 1971473            0             0
## 3   9.217931   0.202509    1.887268 15.45056 1980627            0             0
## 4  17.451744   0.000000    1.055076 36.65478 1980627            0             0
## 5 104.866034   0.000000   10.875268 32.66316 1980627            0             0
## 6 197.218159   0.000000   37.031138 17.48456 1980627            0             0
##   REFUND_COUNT_LTM REFUND_AMOUNT_LTM
## 1                0                 0
## 2                0                 0
## 3                0                 0
## 4                0                 0
## 5                0                 0
## 6                0                 0
```

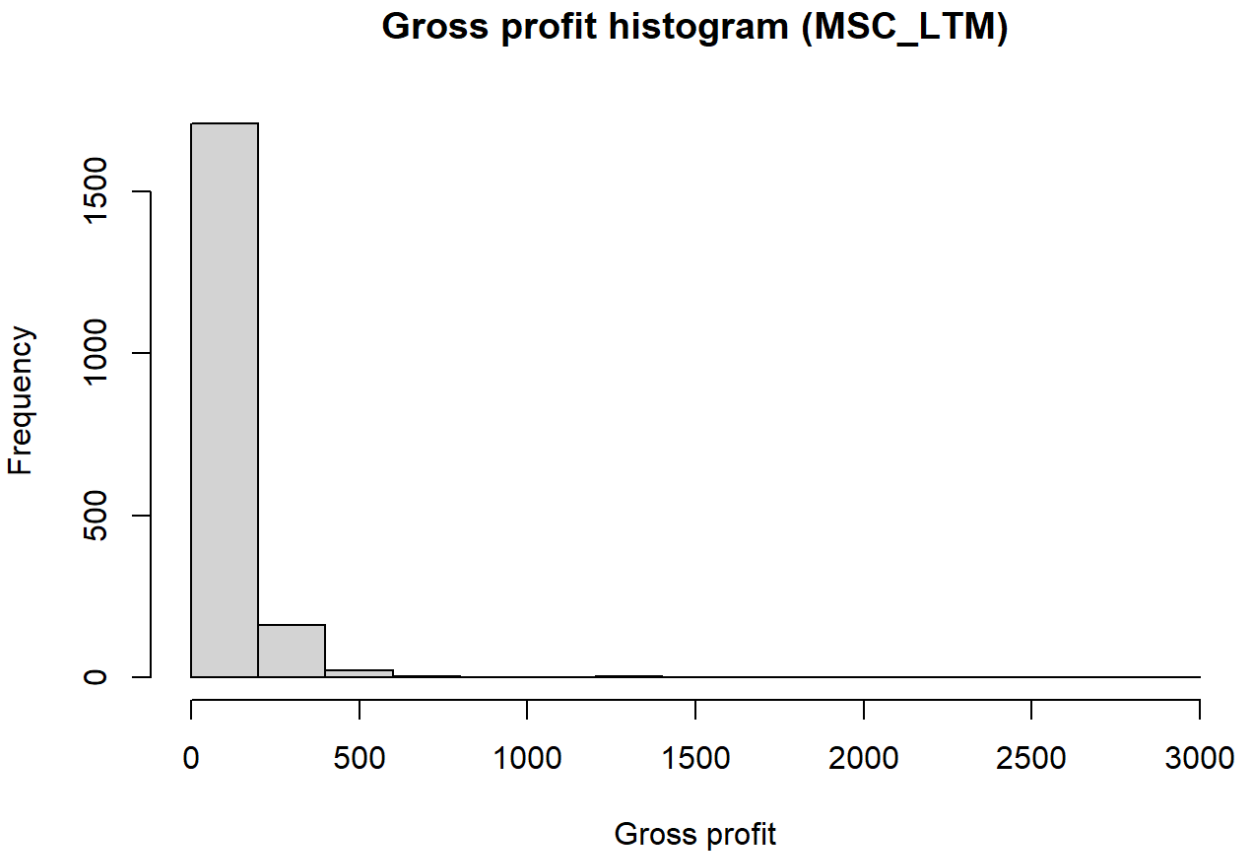
```r
###Analyzing customer activity

###Transaction volume analysis
mean_amount <- mean(data_p$TRX_AMOUNT_LTM)
print(mean_amount)
## [1] 19798.87
median_amount <- median(data_p$TRX_AMOUNT_LTM)
print(median_amount)
## [1] 5075.129
hist(data_p$TRX_AMOUNT_LTM)

###Activity segmentation
online_transactions <- subset(data_p, ONLINEPOS == "ONLINE")
physical_transactions <- subset(data_p, ONLINEPOS == "POS")

##Data visualization
library(ggplot2)
```



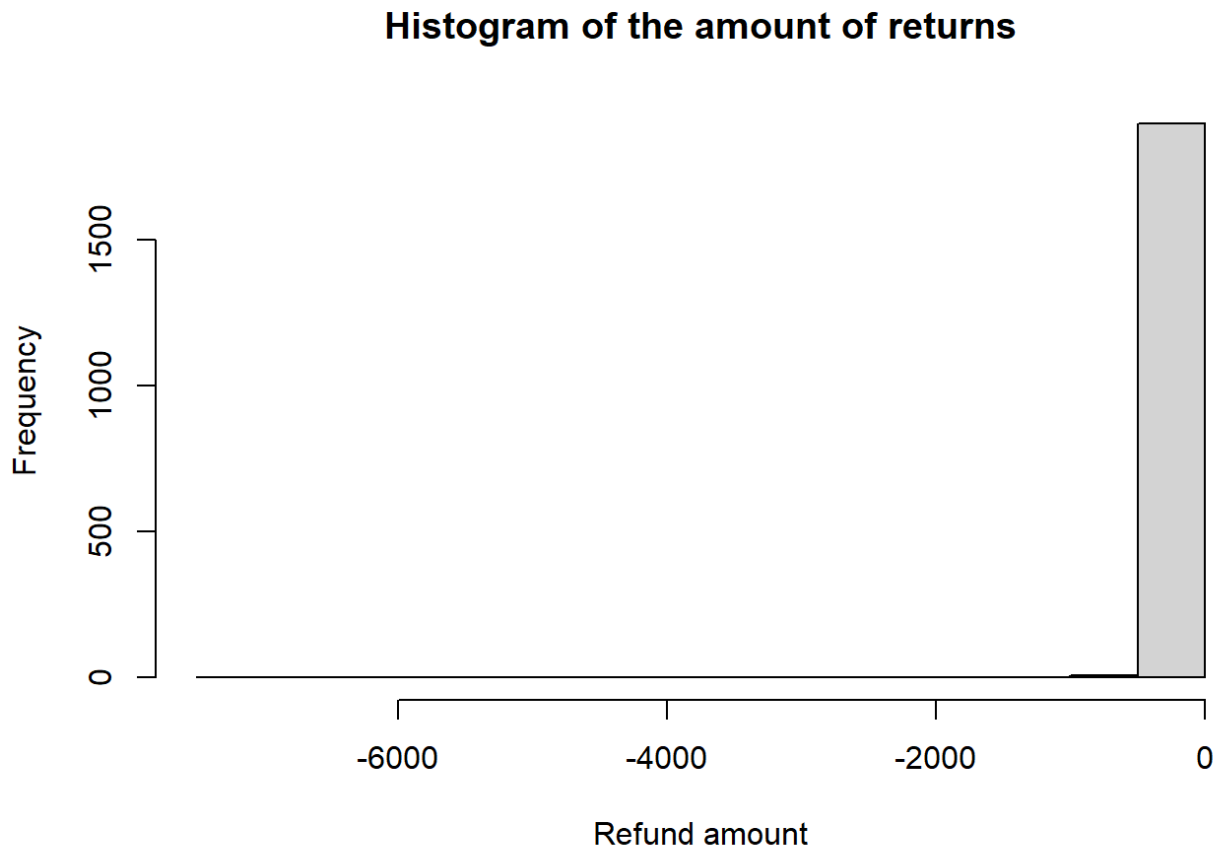

```r
ggplot(data_p, aes(x = TRX_COUNT_LTM, y = TRX_AMOUNT_LTM)) + geom_point()
```




```
###Analysis of financial performance


##Calculation of statistics
mean_msc <- mean(data_p$MSC_LTM)
print(mean_msc)
## [1] 79.70745
median_msc <- median(data_p$MSC_LTM)
print(median_msc)
## [1] 20.76443
total_net_msc <- sum(data_p$NET_MSC_LTM)
print(total_net_msc)
## [1] 23330.01
standard_deviation_msc <- sd(data_p$MSC_LTM)
print(standard_deviation_msc)
## [1] 145.4234
hist(data_p$MSC_LTM, main = "Gross profit histogram (MSC_LTM)", xlab = "Gross prof
it")
```




```
####Returns analysis


###Calculation of statistics


mean_refunds <- mean(data_p$REFUND_AMOUNT_LTM)
print(mean_refunds)
## [1] -13.30586
total_refunds <- sum(data_p$REFUND_AMOUNT_LTM)
print(total_refunds)
## [1] -25387.59
####Data visualization
hist(data_p$REFUND_AMOUNT_LTM, main = "Histogram of the amount of returns", xlab =
"Refund amount")
```



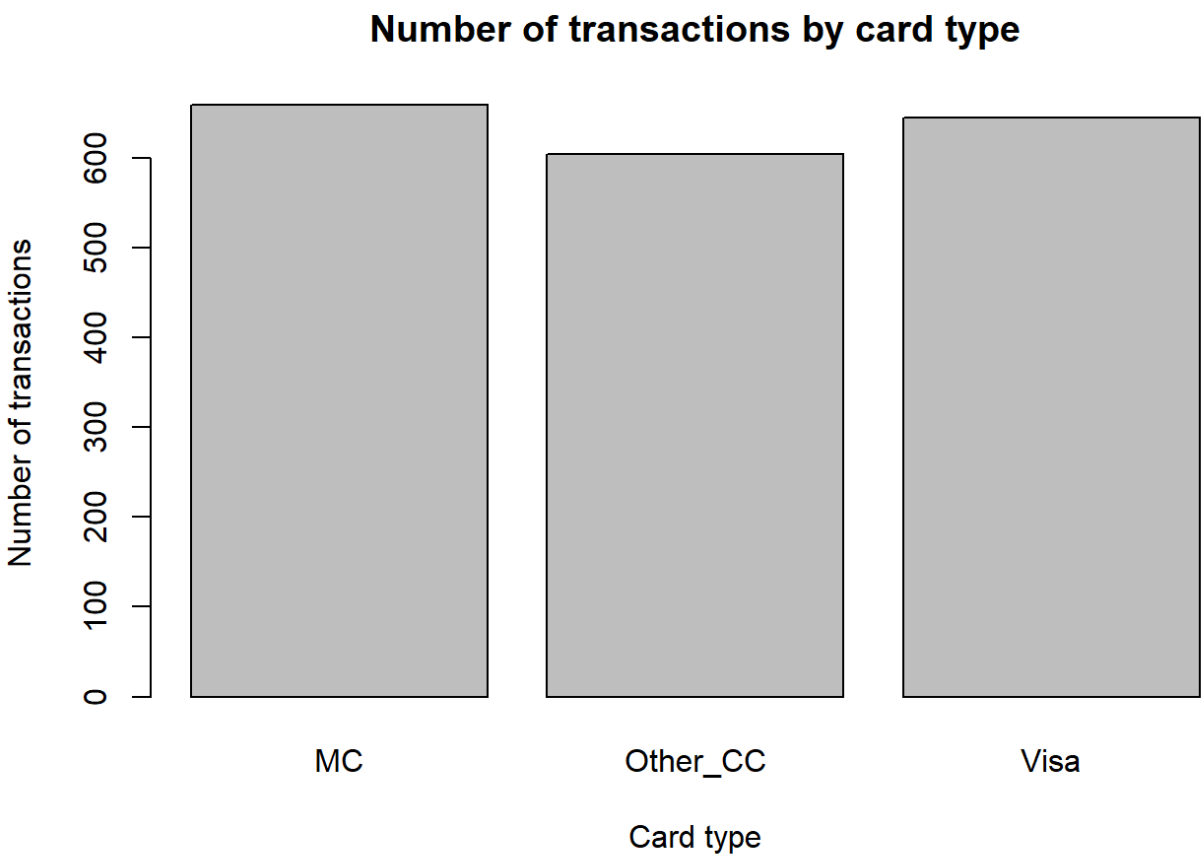

```
###Analyzing map types


# Counting the number of each card type
card_types <- table(data_p$SCHEME)
print(card_types)
```

```
##
##       MC Other_CC     Visa
##      659      604      645
```

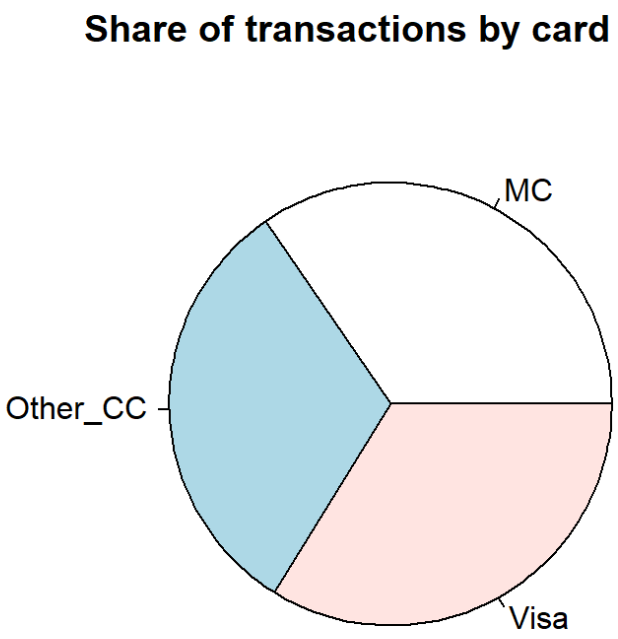
```
# Calculation of the share of each card type
card_types_percentage <- prop.table(card_types) * 100
print(card_types_percentage)
```

```
##
##       MC Other_CC     Visa
## 34.53878 31.65618 33.80503
```

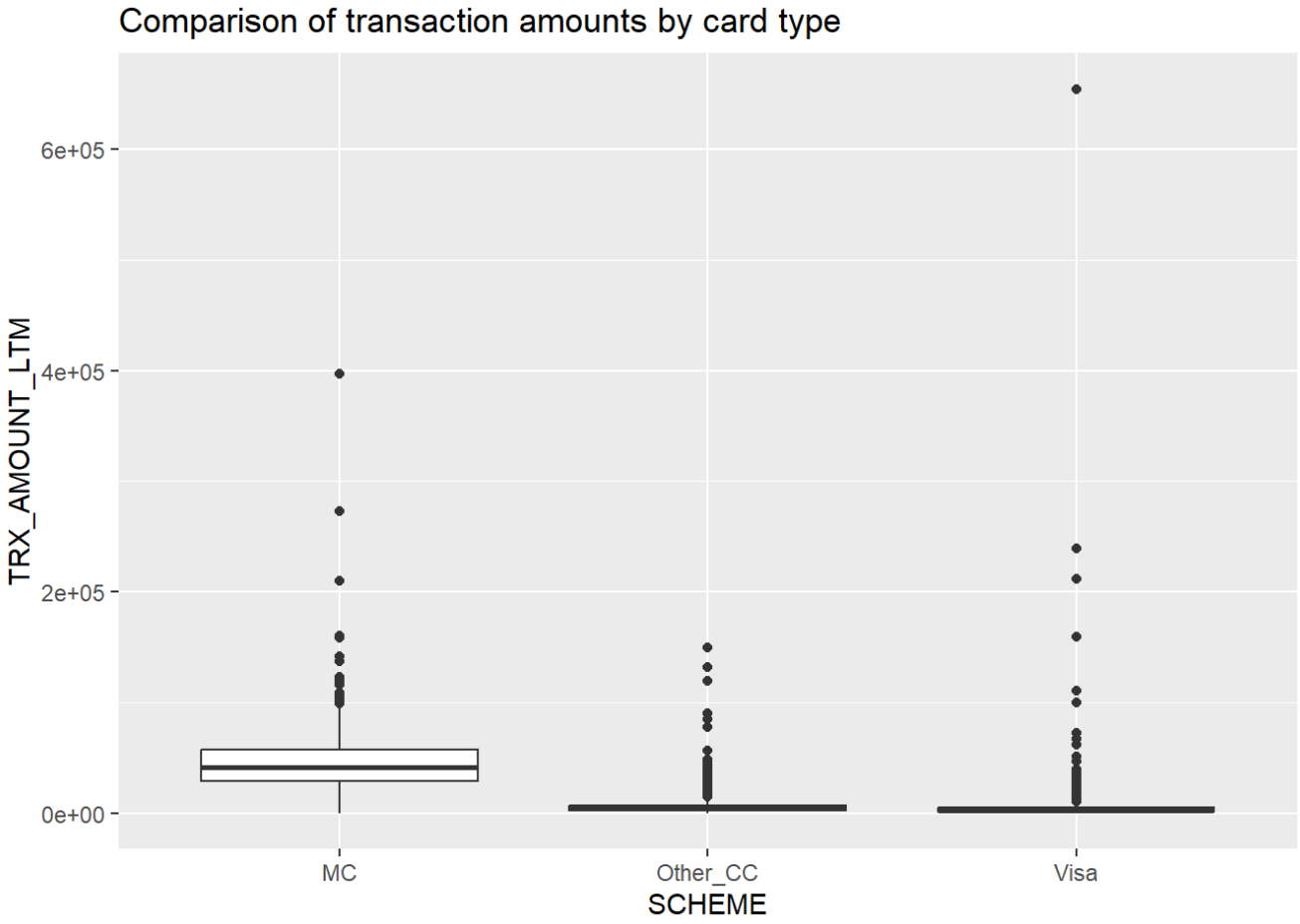
```
barplot(card_types, main = "Number of transactions by card type", xlab = "Card typ
e", ylab = "Number of transactions")
```




```
pie(card_types_percentage, main = "Share of transactions by card type")
```




```
#### Comparison of card types by other parameters (e.g. transaction amount)

library(ggplot2)

ggplot(data_p, aes(x = SCHEME, y = TRX_AMOUNT_LTM)) + geom_boxplot() + labs(title
= "Comparison of transaction amounts by card type")
```


Comparison of transaction amounts by card type
TRX_AMOUNT_LTM
6e+05
4e+05
2e+05
0e+00
MC
Other_CC
Visa
SCHEME