Retail Case Study

Anshu

```
setwd("F:\\BA\\R case studies\\R case study 2 (Retail)")

Customer <- read.csv("F:\\BA\\R case studies\\R case study 2
(Retail)\\Customer.csv")
prod_cat_info <- read.csv("prod_cat_info.csv")
Transactions <- read.csv("transactions.csv")
```

Question no.1

```
Customer$Gender <- as.character(Customer$Gender)
Customer$DOB <- as.Date(Customer$DOB, format = "%d-%m-%Y")

Customer_Final <- merge(x=Transactions,y=Customer,by.x = "cust_id",by.y =
"customer_Id")
Customer_Final <- merge(x=Customer_Final,y= prod_cat_info,by
="prod_cat_code",all = T)
```

Using dplyr

```
Customer_Final1 <- full_join(Transactions,prod_cat_info,by = "prod_cat_code")
Customer <- dplyr::rename(Customer,"cust_id"="customer_Id")
Customer_Final1 <- full_join(Customer_Final1,Customer,by = "cust_id")
```

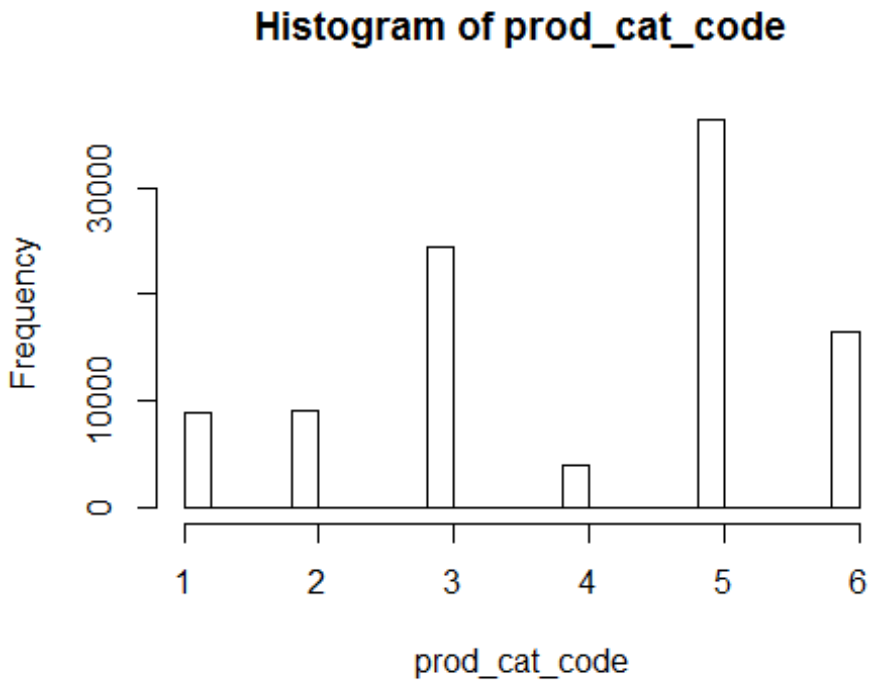
Question no.1

```
summary(Customer_Final)

##  prod_cat_code      cust_id       transaction_id           tran_date
##  Min.   :1.000   Min.   :266783   Min.   :3.269e+06   25-08-2012:  153
##  1st Qu.:3.000   1st Qu.:268956   1st Qu.:2.492e+10   13-07-2011:  144
##  Median :5.000   Median :270982   Median :5.011e+10   25-09-2011:  144
##  Mean   :4.003   Mean   :271030   Mean   :5.007e+10   3/2/2014  :  142
##  3rd Qu.:5.000   3rd Qu.:273120   3rd Qu.:7.528e+10   21-12-2013:  141
##  Max.   :6.000   Max.   :275265   Max.   :9.999e+10   1/1/2014  :  140
##  NA's   :4                                            (Other)   :98433
##  prod_subcat_code      Qty              Rate              Tax
##  Min.   : 1.000   Min.   :-5.000   Min.   :-1499.0   Min.   :  7.35
##  1st Qu.: 4.000   1st Qu.: 1.000   1st Qu.:  313.0   1st Qu.: 98.28
##  Median : 7.000   Median : 3.000   Median :  713.0   Median :199.92
##  Mean   : 6.797   Mean   : 2.438   Mean   :  637.9   Mean   :248.87
##  3rd Qu.:10.000   3rd Qu.: 4.000   3rd Qu.: 1109.0   3rd Qu.:366.98
##  Max.   :12.000   Max.   : 5.000   Max.   : 1500.0   Max.   :787.50
##  NA's   :4        NA's   :4        NA's   :4         NA's   :4
##    total_amt               Store_type         DOB
##  Min.   :-8270.9   e-shop        :19964   Min.   :1970-01-02
```

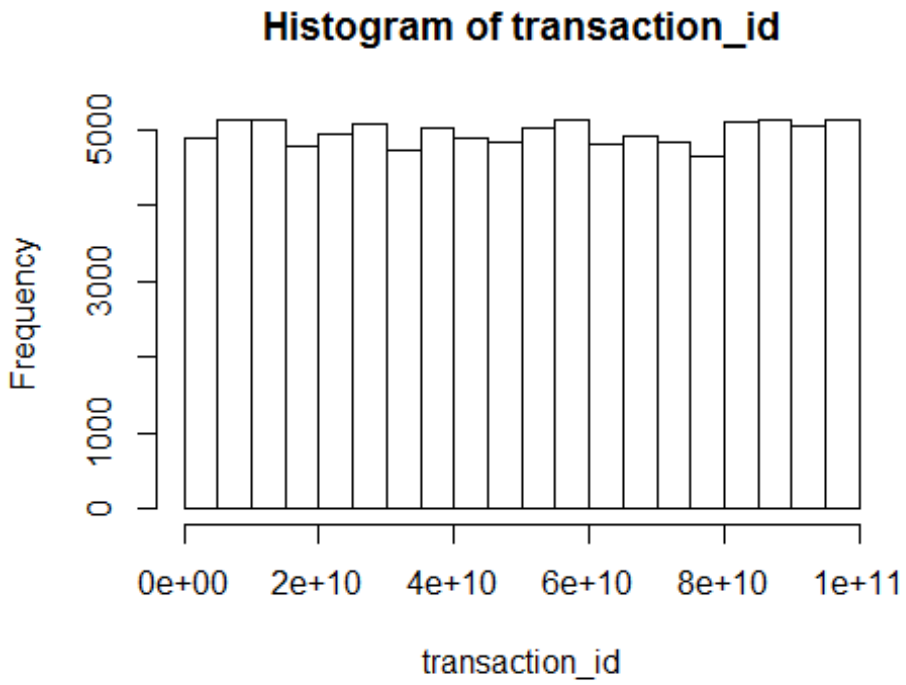
```
##  1st Qu.:  762.5   e-Shop        :20222   1st Qu.:1975-10-02
##  Median : 1761.4   Flagship store:19816   Median :1981-06-23
##  Mean   : 2114.6   MBR           :19974   Mean   :1981-07-16
##  3rd Qu.: 3585.7   TeleShop      :19321   3rd Qu.:1987-06-02
##  Max.   : 8287.5                          Max.   :1992-12-29
##  NA's   :3
##     Gender           city_code                  prod_cat
##  Length:99297       Min.   : 1.000   Bags            : 3996
##  Class :character   1st Qu.: 3.000   Books           :36414
##  Mode  :character   Median : 5.000   Clothing        : 8880
##                     Mean   : 5.467   Electronics     :24490
##                     3rd Qu.: 8.000   Footwear        : 8997
##                     Max.   :10.000   Home and kitchen:16516
##                     NA's   :36       NA's            :    4
##  prod_sub_cat_code   prod_subcat
##  Min.   : 1.000    Mens    : 7957
##  1st Qu.: 4.000    Women   : 7957
##  Median : 7.000    Academic: 6069
##  Mean   : 7.069    Children: 6069
##  3rd Qu.:10.000    Comics  : 6069
##  Max.   :12.000    (Other) :65172
##  NA's   :4         NA's    :    4
```

Histograms

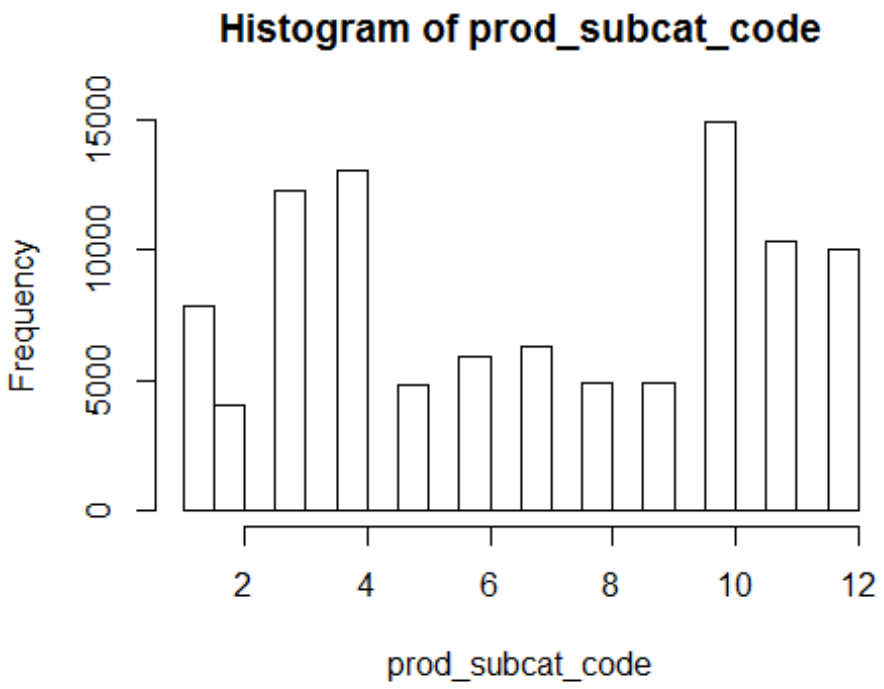
```
hist_prod_cat_code <- hist(Customer_Final$prod_cat_code,main = "Histogram of
prod_cat_code",xlab = "prod_cat_code")
```



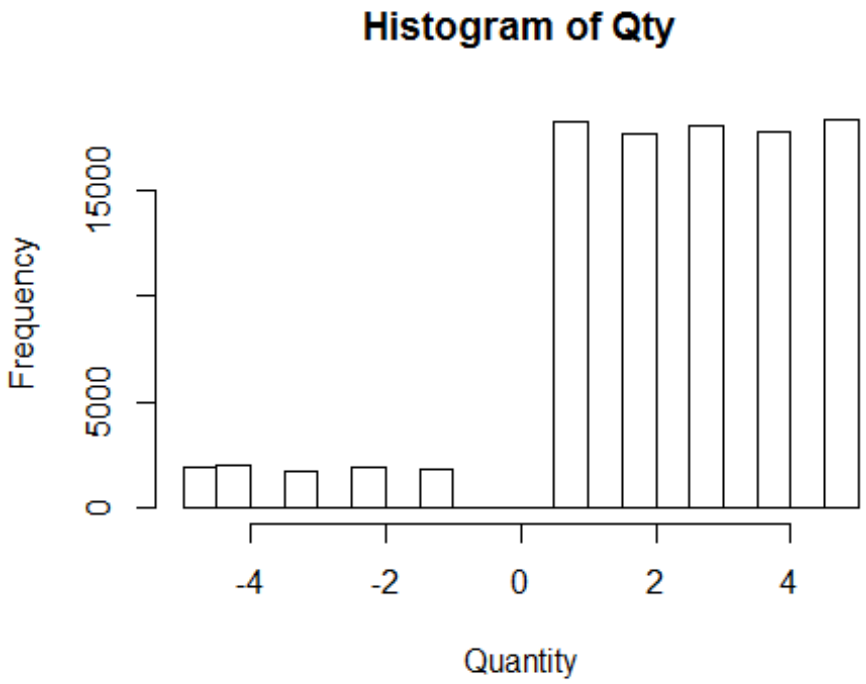

```
hist_transaction_id <- hist(Customer_Final$transaction_id,main = "Histogram
of transaction_id",xlab = "transaction_id")
```



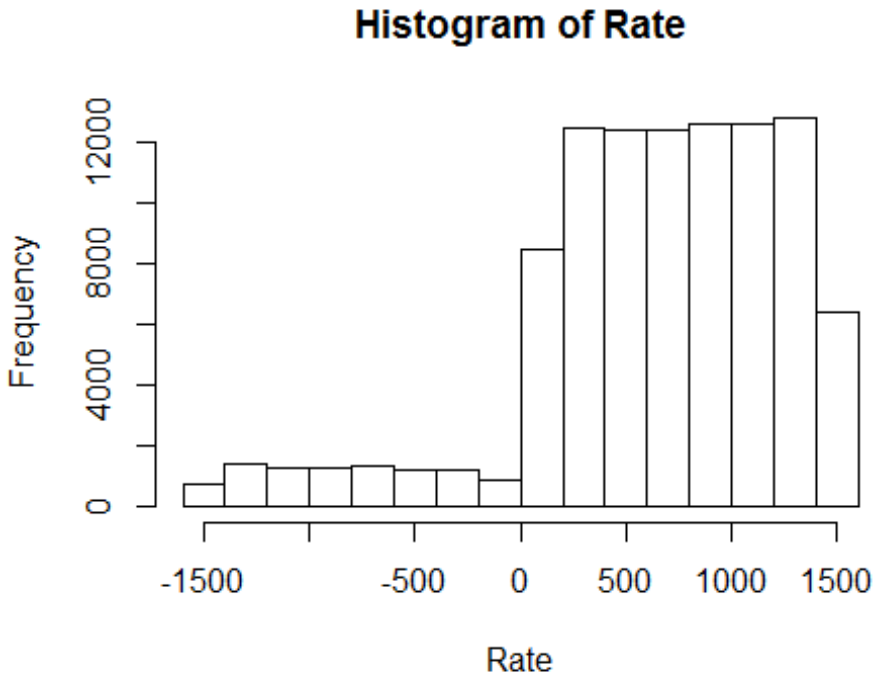
```r
hist_prod_subcat_code <- hist(Customer_Final$prod_subcat_code,main =
"Histogram of prod_subcat_code",xlab = "prod_subcat_code" )
```



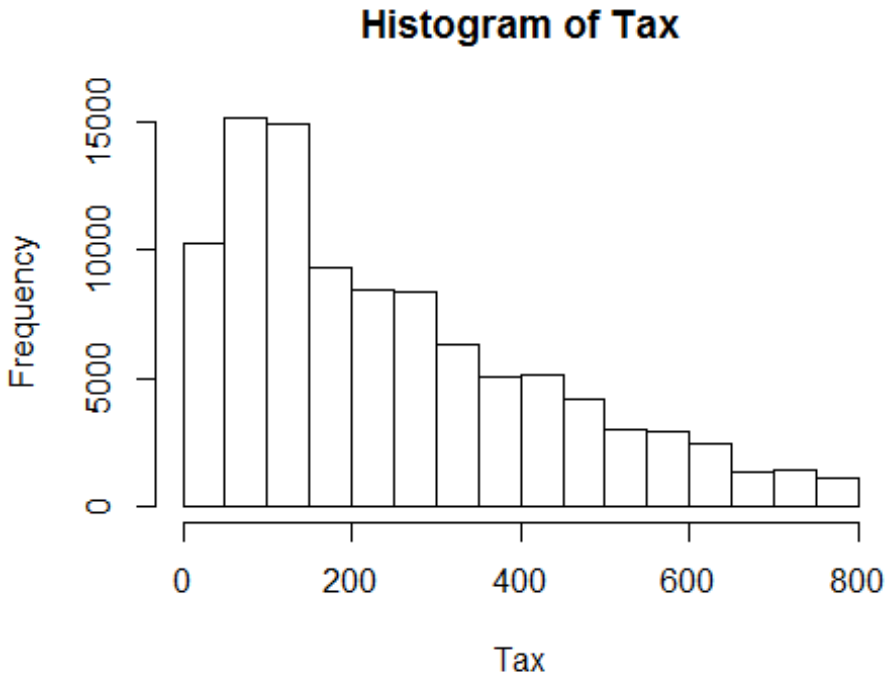

```r
hist_Qty <- hist(Customer_Final$Qty,main = "Histogram of Qty",xlab =
"Quantity")
```



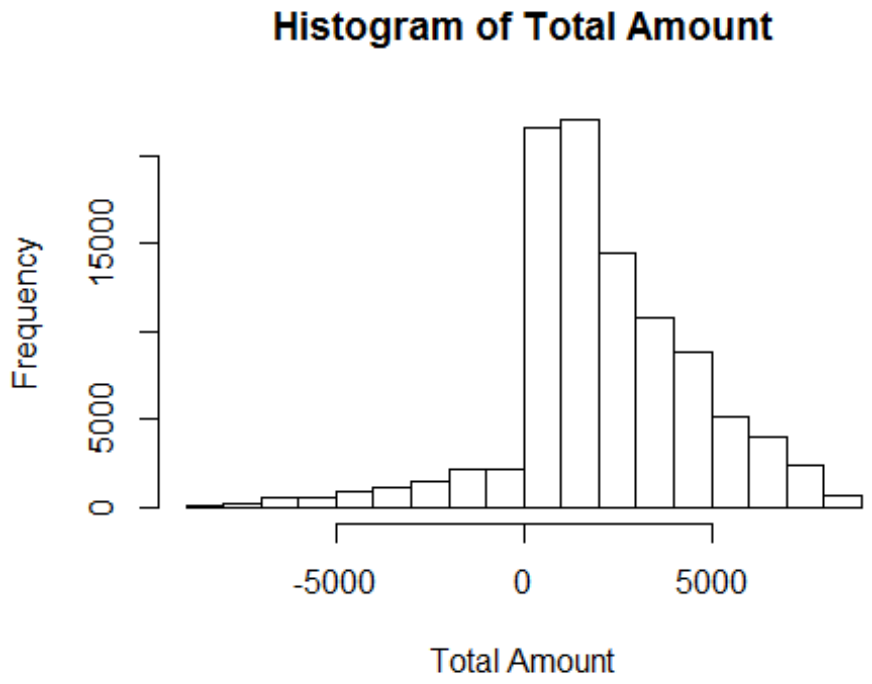

```
hist_Rate <- hist(Customer_Final$Rate,main = "Histogram of Rate",xlab =
"Rate")
```



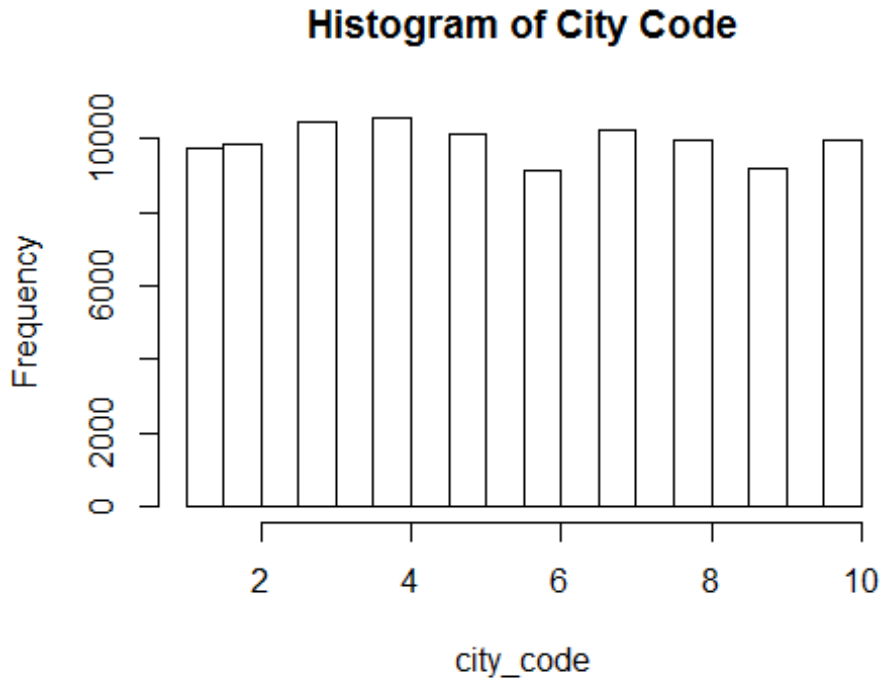
```
hist_Tax <- hist(Customer_Final$Tax,main = "Histogram of Tax",xlab = "Tax")
```



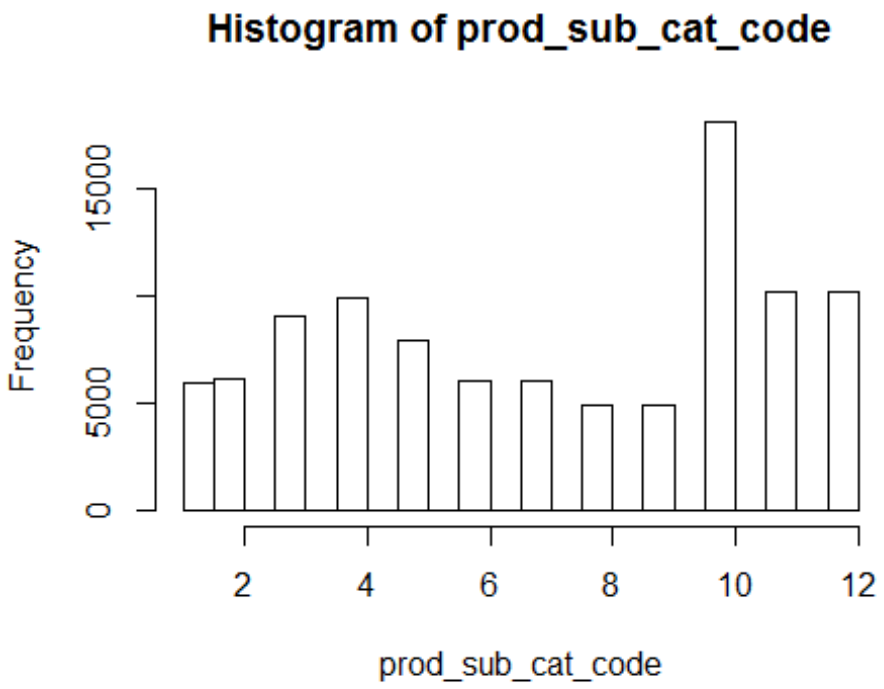

```
hist_Total_amt <- hist(Customer_Final$total_amt,main = "Histogram of Total
Amount",xlab = "Total Amount")
```




```
hist_city_code <- hist(Customer_Final$city_code,main = "Histogram of City
Code",xlab = "city_code")
```

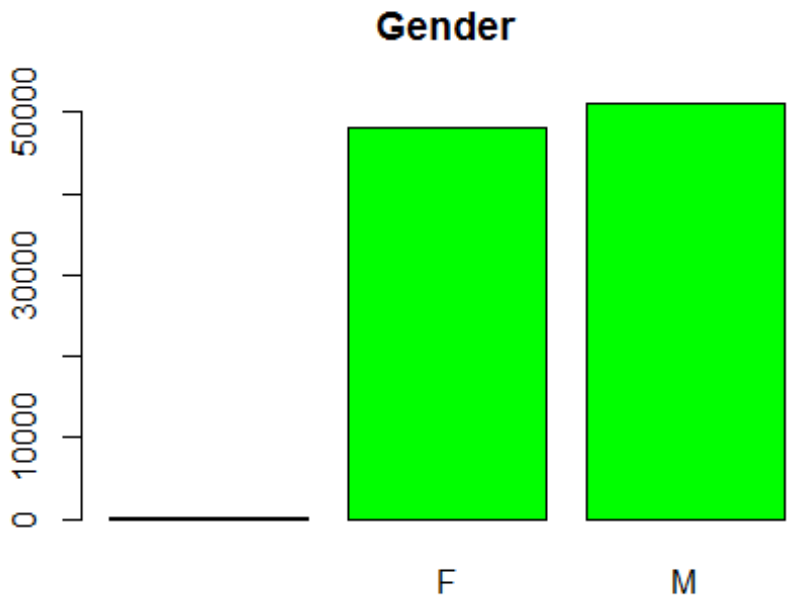


```
hist_prod_sub_cat_code <- hist(Customer_Final$prod_sub_cat_code,main =
"Histogram of prod_sub_cat_code",xlab = "prod_sub_cat_code")
```

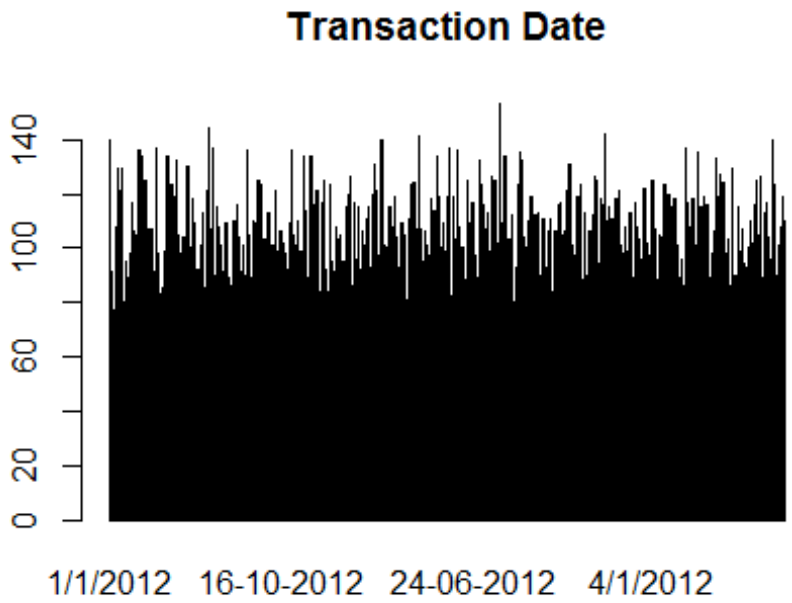


Frequency Bars

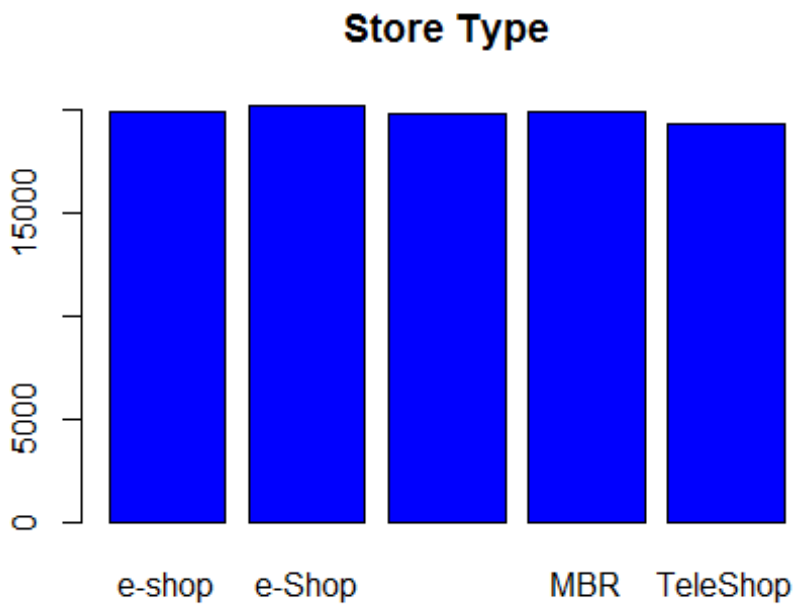
```
Gender <- table(Customer_Final$Gender)
Bar_Gender <- barplot(Gender,main = "Gender",col="Green")
```



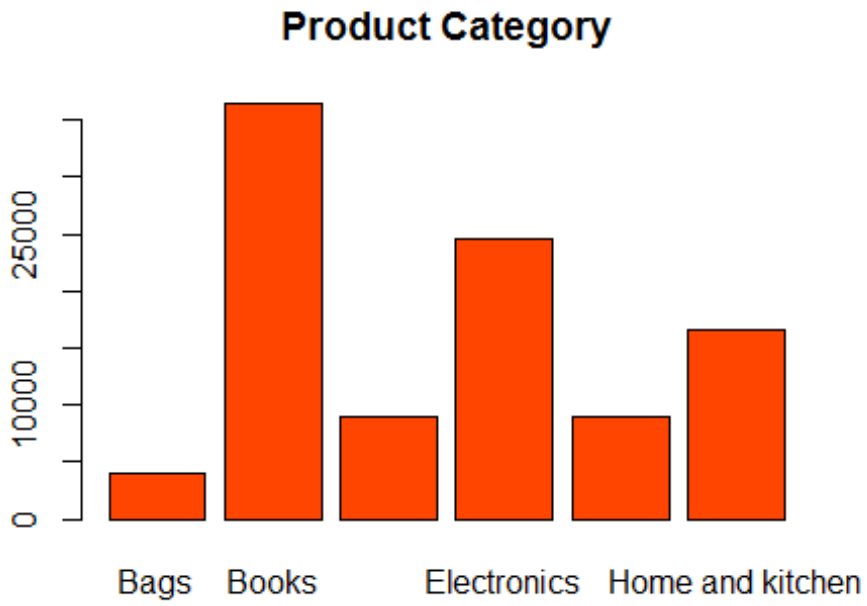

```
Tran_date <- table(Customer_Final$tran_date)
Bar_Tran_date <- barplot(Tran_date,main = "Transaction Date")
```



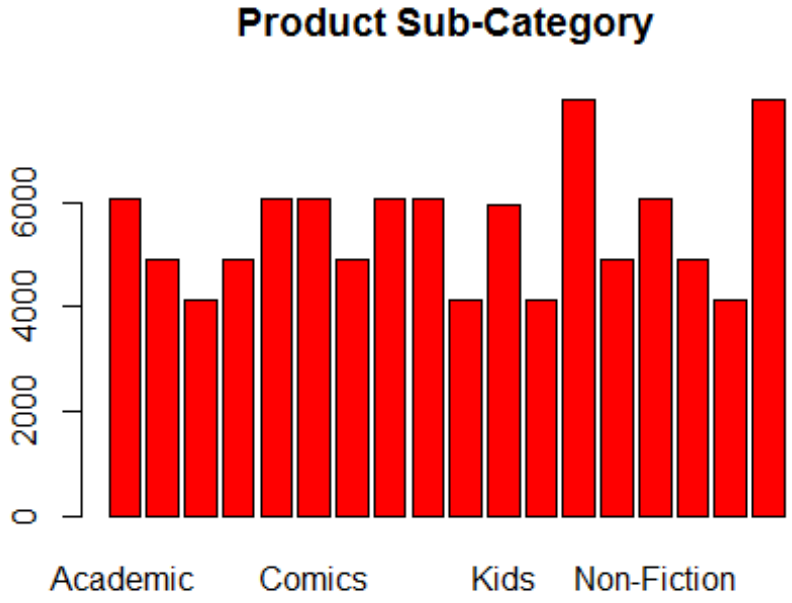
```r
Store_Type <- table(Customer_Final$Store_type)
Bar_Store_Type <- barplot(Store_Type,main = "Store Type",col="Blue")
```




```r
Prod_cat <- table(Customer_Final$prod_cat)
Bar_Prod_cat <- barplot(Prod_cat,main = "Product Category",col="orange red")
```




```
prod_subcat <- table(Customer_Final$prod_subcat)
Bar_prod_subcat <- barplot(prod_subcat,main = "Product Sub-
Category",col="red")
```



Question no.3(a)

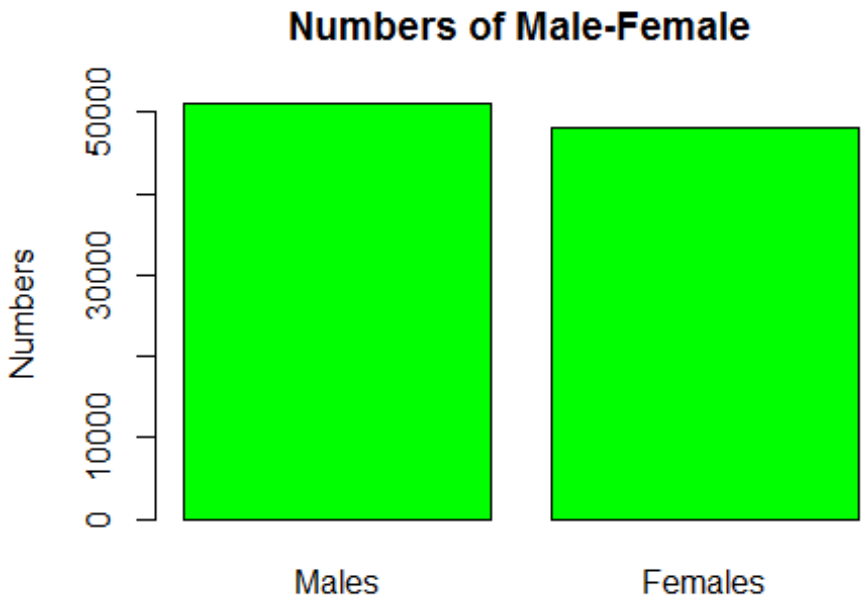
```
Transaction_Date <- parse_date_time(Customer_Final$tran_date, c("mdy",
"dmy"))
Transaction_Date <- as.Date(Transaction_Date,format = "%d-%m-Y")
TimePeriod <- print(max(Transaction_Date,na.rm = T)-
min(Transaction_Date,na.rm = T))

## Time difference of 1430 days
```

Question no.3(b)

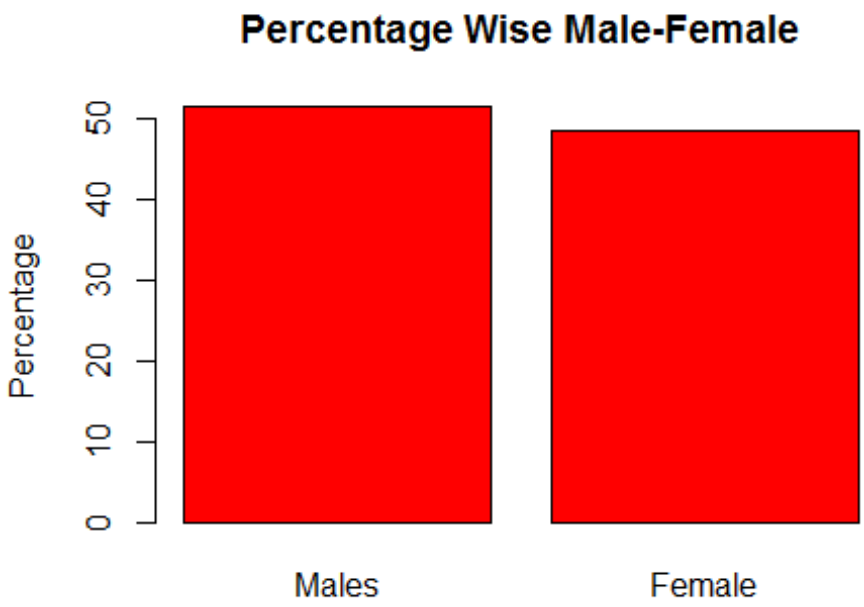
```
Total_Neg_Num <- (length(which(Customer_Final$total_amt<0)))
print(Total_Neg_Num)

## [1] 9294
```

Question no.4

```
No.ofMales <-
print(length(Customer_Final$Gender[Customer_Final$Gender=="M"]))

## [1] 51051

No.ofFemales <-
print(length(Customer_Final$Gender[Customer_Final$Gender=="F"]))

## [1] 48206

Sum <- sum(No.ofMales+No.ofFemales)
Total <- c(Males=No.ofMales,Females=No.ofFemales)
Bar_M_F <- barplot(Total,main = "Numbers of Male-Female",ylab =
"Numbers",col="Green")
```




```
PercentMales <- print((No.ofMales/Sum)*100)

## [1] 51.43315

PercentFemales <- print((No.ofFemales/Sum)*100)

## [1] 48.56685

Percentage <- c(Males=PercentMales,Female=PercentFemales)
Bar_M_F_Per <- barplot(Percentage, main = "Percentage Wise Male-Female",ylab
= "Percentage",col="red")
```

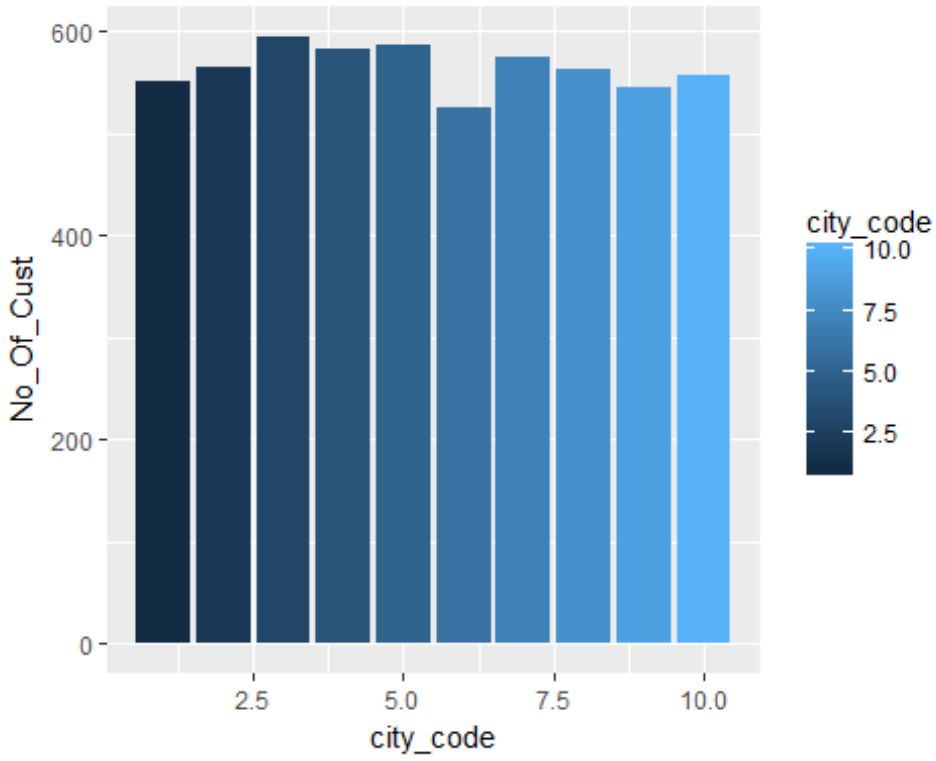


Question no.5(a)

```
CityWiseCust <- Customer %>% group_by(city_code) %>%
summarise(No_Of_Cust=length(city_code))
CityWiseCust$Percentage <-
(CityWiseCust$No_Of_Cust/sum(CityWiseCust$No_Of_Cust)*100)
CityWiseCust$Percentage <- round(CityWiseCust$Percentage,digits = 3)

MaxCityCustomer <-subset(CityWiseCust,Percentage=max(Percentage))
MaxCity <- print(MaxCityCustomer[which.max(CityWiseCust$Percentage),])

## # A tibble: 1 x 3
##   city_code No_Of_Cust Percentage
##       <int>      <int>      <dbl>
## 1         3        595       10.5
```

Question no.5(b)

```
CatVsCust_Count <-ggplot(CityWiseCust) + aes(x = city_code , y = No_Of_Cust ,
fill = city_code) +
                    geom_bar(stat = "identity")
print(CatVsCust_Count)

## Warning: Removed 1 rows containing missing values (position_stack).
```



Question no.6

```
Store_Type1 <- Customer_Final %>% group_by(Store_type) %>%
summarise(TOtalAmount= sum(total_amt,na.rm = T))
Store_Type2 <- Customer_Final %>% group_by(Store_type,cust_id) %>%
summarise()
Store_Type3 <- Store_Type2 %>% group_by(Store_type) %>%
summarise(TotalQuantity = length(Store_type))

Store_Type1 <- merge(Store_Type1,Store_Type3,all = T)
StoreName_byValue <- Store_Type1[which.max(Store_Type1$TOtalAmount),]
print(StoreName_byValue)

##   Store_type TOtalAmount TotalQuantity
## 2     e-Shop    42993171          2943

StoreName_byQuantity <- Store_Type1[which.max(Store_Type1$TotalQuantity),]
print(StoreName_byQuantity)

##   Store_type TOtalAmount TotalQuantity
## 4        MBR    41700330          3000
```

Question no.7

```
Flagship_Stores <- Customer_Final %>% group_by(Store_type,prod_cat) %>%
summarise(TotalRevenue=sum(total_amt,na.rm=T))

Flagship_Stores_Clothing <-
```

```
Flagship_Stores[Flagship_Stores$Store_type=="Flagship store" &
Flagship_Stores$prod_cat=="Clothing",]
print(Flagship_Stores_Clothing)
```

```
## # A tibble: 2 x 3
## # Groups:   Store_type [2]
##   Store_type     prod_cat TotalRevenue
##   <fct>          <fct>           <dbl>
## 1 Flagship store Clothing     3583270.
## 2 <NA>           <NA>              NA
```

```
Flagship_Stores_Electronics <-
Flagship_Stores[Flagship_Stores$Store_type=="Flagship store" &
Flagship_Stores$prod_cat=="Electronics",]
print(Flagship_Stores_Electronics)
```

```
## # A tibble: 2 x 3
## # Groups:   Store_type [2]
##   Store_type     prod_cat    TotalRevenue
##   <fct>          <fct>              <dbl>
## 1 Flagship store Electronics    11075680.
## 2 <NA>           <NA>                 NA
```

Question no.8

```
Transact <- Customer_Final %>% group_by(cust_id,total_amt) %>% summarise()
Transact1 <- Transact %>% group_by(cust_id) %>% summarise(No.Of_Transaction
=length(which(total_amt>0)))
Customer_morethan_5_Transct <- print(length(which(Transact1$No.Of_Transaction
>= 5)))
```

```
## [1] 1775
```

Question no.9

```
Male_customer <- Customer_Final %>% group_by(Gender,prod_cat) %>%
summarise(Total_Revenue = sum(total_amt,na.rm = T))
Male_cust <- Male_customer[Male_customer$Gender=="M"& Male_customer$prod_cat
== "Electronics",]
print(Male_cust)
```

```
## # A tibble: 1 x 3
## # Groups:   Gender [1]
##   Gender prod_cat    Total_Revenue
##   <chr>  <fct>               <dbl>
## 1 M      Electronics     28515547.
```

Question no.10(a)

```
Customer_Final$tran_date <- parse_date_time(Customer_Final$tran_date,
c("mdy", "dmy"))  #imp
Customer_Final$tran_date <- as.Date(Customer_Final$tran_date,format = "%d-%m-
```

```
Y")

Customer_Final$Age <- (Customer_Final$tran_date - Customer_Final$DOB)/365.24
Customer_Final$Age <- round(Customer_Final$Age,digits = 2)

  #Elder is 18-25

Customer_Final$Age_Group <-
ifelse(Customer_Final$Age<=25,"Young",ifelse(Customer_Final$Age<=35,"Elder","
Mature"))
Aged <- Customer_Final %>% group_by(Age_Group,prod_cat) %>%
summarise(TotalRevenue=sum(total_amt,na.rm = T))

Age_Grp_Elec <- Aged[Aged$Age_Group=="Young" & Aged$prod_cat=="Electronics",]
print(Age_Grp_Elec)

## # A tibble: 1 x 3
## # Groups:   Age_Group [1]
##   Age_Group prod_cat    TotalRevenue
##   <chr>     <fct>              <dbl>
## 1 Young     Electronics    12819630.

Age_Grp_Book <- Aged[Aged$Age_Group=="Young" & Aged$prod_cat=="Books",]
print(Age_Grp_Book)

## # A tibble: 1 x 3
## # Groups:   Age_Group [1]
##   Age_Group prod_cat TotalRevenue
##   <chr>     <fct>           <dbl>
## 1 Young     Books       17362571.
```

Question no.10(b)

```
date_revenue <- Customer_Final %>% group_by(tran_date) %>%
summarise(TotalRevenue=sum(total_amt, na.rm = T))
Date <- date_revenue[date_revenue$tran_date>="2014-01-01" &
date_revenue$tran_date<="2014-03-01",]
TotalRevenueGen <- print(sum(date_revenue$TotalRevenue))

## [1] 209966944
```