Walmart Data Analysis using R:

```
> library(readxl)

> walmart <- read_excel("C:/Users/Syed Abdul Sami/OneDrive/Desktop/walmart.xlsx")

> View(walmart)

> d<-walmart
> str(d)

Classes ‘tbl_df’, ‘tbl’ and 'data.frame':         1000 obs. of  17 variables:

 $ Invoice ID           : chr  "750-67-8428" "226-31-3081" "631-41-3108" "123-19-1176" ...

 $ Branch               : chr  "A" "C" "A" "A" ...

 $ City                 : chr  "Yangon" "Naypyitaw" "Yangon" "Yangon" ...

 $ Customer type        : chr  "Member" "Normal" "Normal" "Member" ...

 $ Gender               : chr  "Female" "Female" "Male" "Male" ...

 $ Product line         : chr  "Health and beauty" "Electronic accessories" "Home and lifestyle"
"Health and beauty" ...

 $ Unit price           : num  74.7 15.3 46.3 58.2 86.3 ...

 $ Quantity             : num  7 5 7 8 7 7 6 10 2 3 ...

 $ Tax 5%               : num  26.14 3.82 16.22 23.29 30.21 ...

 $ Total                : num  549 80.2 340.5 489 634.4 ...

 $ Date                 : POSIXct, format: "2019-01-05" "2019-03-08" ...

 $ Time                 : POSIXct, format: "1899-12-31 13:08:00" "1899-12-31 10:29:00" ...

 $ Payment              : chr  "Ewallet" "Cash" "Credit card" "Ewallet" ...

 $ cogs                 : num  522.8 76.4 324.3 465.8 604.2 ...

 $ gross margin percentage: num  4.76 4.76 4.76 4.76 4.76 ...

 $ gross income         : num  26.14 3.82 16.22 23.29 30.21 ...

 $ Rating               : num  9.1 9.6 7.4 8.4 5.3 4.1 5.8 8 7.2 5.9 ...


> length(d)
```

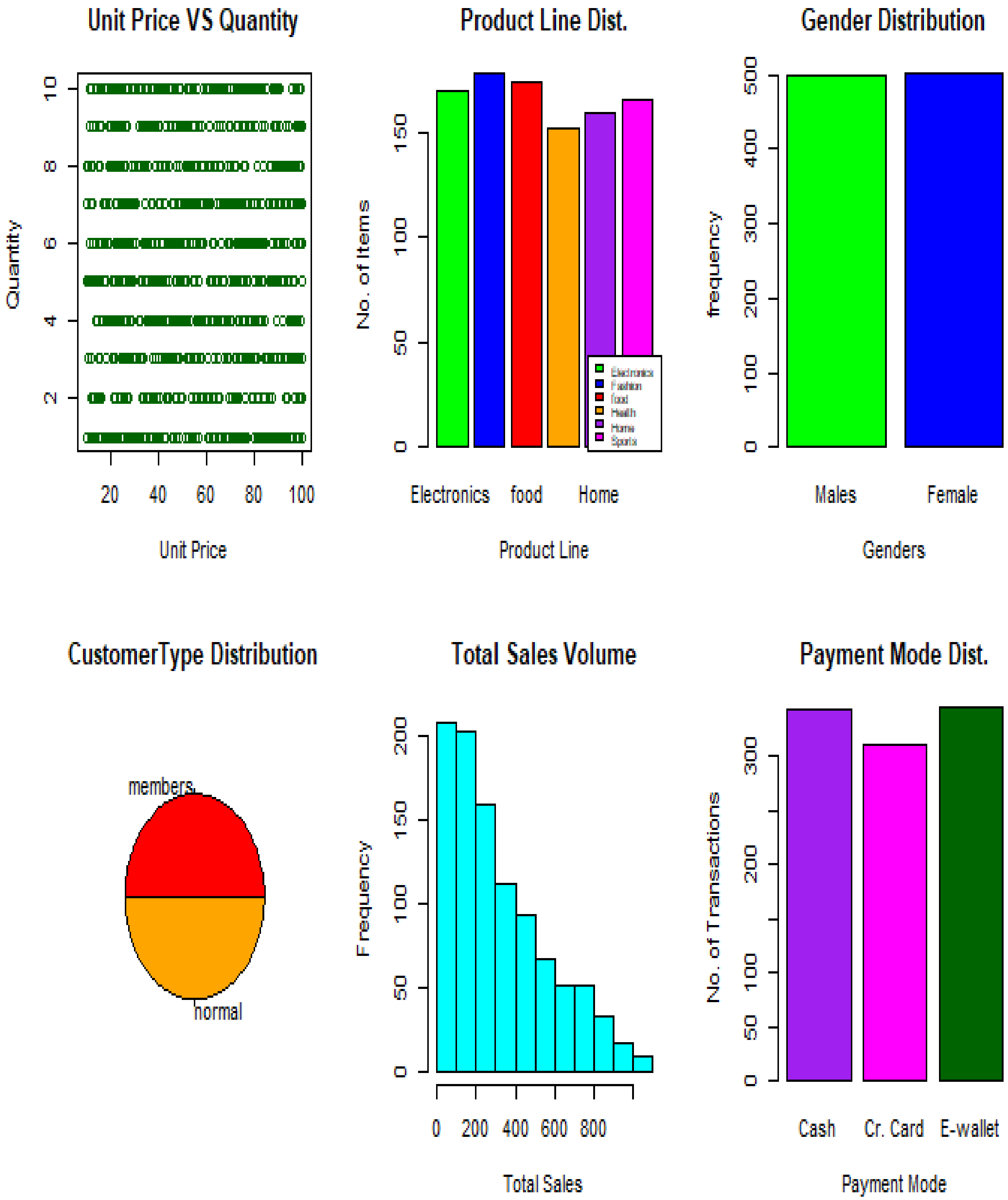
```
[1] 17

> remduplicate <- function(x){
+ uniValues <- unique(x)
+ return(uniValues)
+ }

> res <- remduplicate(d)
> print(res)

> attach(d)

> lm(Total ~ `Unit price` + Quantity)

Call:
lm(formula = Total ~ `Unit price` + Quantity)

Coefficients:
 (Intercept)  `Unit price`      Quantity
    -324.522         5.814        58.772

lm(Total ~ Quantity, data = d)

Call:
lm(formula = Total ~ Quantity, data = d)

Coefficients:
(Intercept)     Quantity
```

```
     -3.993       59.339

lm(Total ~ `Unit price`, data = d)

Call:
lm(formula = Total ~ `Unit price`, data = d)

Coefficients:
 (Intercept)  `Unit price`
      -4.582         5.884

totalsales <- sum(Total)
> print(totalsales)
[1] 322966.7

> highestsale <- d$Total[which.max(d$Total)]
> print(highestsale)
[1] 1042.65

> highestsellingproductline <- d$`Product line`[which.max(d$Total)]
> print(highestsellingproductline)
[1] "Fashion accessories"

> avgsale <- mean(Total)
> print(avgsale)
[1] 322.9667

> males <- sum(d$Gender=='Male')
> males
```

```
[1] 499

> females <- sum(d$Gender=='Female')

> females

[1] 501


> count(d, City)

# A tibble: 3 × 2

  City        n

  <chr>     <int>

1 Mandalay    332

2 Naypyitaw   328

3 Yangon      340


> count(d, `Customer type` )

# A tibble: 2 × 2

  `Customer type`     n

  <chr>           <int>

1 Member            501

2 Normal            499


> count(d, `Product line` )

# A tibble: 6 × 2

  `Product line`           n

  <chr>                <int>

1 Electronic accessories   170

2 Fashion accessories      178

3 Food and beverages       174

4 Health and beauty        152

5 Home and lifestyle       160
```

```
6 Sports and travel       166
> count(d, Branch )
# A tibble: 3 × 2
  Branch     n
  <chr>  <int>
1 A        340
2 B        332
3 C        328

> TotalCustomers <- males + females
> TotalCustomers
[1] 1000
> percentageMales <- males/TotalCustomers*100
> percentageMales
[1] 49.9
> percentageFemales <- females/TotalCustomers*100
> percentageFemales
[1] 50.1

> TotalBranches <- count(d, Branch )
> TotalProductLine <- count(d, `Product line` )
> TotalCustomerType <- count(d, `Customer type` )
> TotalCities <- count(d, City)
> TotalGenders <- count(d, Gender)
> topCitywithHighestsales <- d$City[which.max(d$Total)]
> topCitywithHighestsales
[1] "Naypyitaw"

> x <- c(percentageFemales, percentageMales)
> x
```

```
[1] 50.1 49.9

> Lables <- c(Females, Males)

Error: object 'Females' not found

> Lables <- c('Females', 'Males')

> Lables

[1] "Females" "Males"
> Members <- sum(d$`Customer type` == 'Member')

> Normal <- sum(d$`Customer type` == 'Normal')

> percentMember <- Members/TotalCustomers*100

> percentNormal <- Normal/TotalCustomers*100

> percentMember

[1] 50.1

> percentNormal

[1] 49.9

paymentMode <- count(d, Payment)

> paymentMode

# A tibble: 3 × 2

  Payment       n

  <chr>     <int>

1 Cash        344

2 Credit card   311

3 Ewallet     345

> y <- c(percentMember,percentNormal)

> Lables1 <- c('members','normal')


> maxRating <- d$Rating[which.max(d$Rating)]

> maxRating

[1] 10

> minRating <- d$Rating[which.min(d$Rating)]
> minRating
[1] 4

> avgRating <- mean(Rating)
> avgRating
[1] 6.9727

> totalCGS <- sum(cogs)
> totalCGS
[1] 307587.4
#plots:
> colors <- c('green','blue','red','orange','purple','magenta','darkgreen','violet','cyan')

> plot(`Unit price`, Quantity, xlab = 'Unit Price', main = 'Unit Price VS Quantity', col = 'darkgreen')
> barplot(i,names.arg = pl ,xlab = 'Product Line', ylab = 'No. of Items', main = 'Product Line Dist.', col = colors )
> legend("bottomright", pl, cex = 0.6, fill = colors)
> barplot(x,names.arg = genders,xlab = 'Genders', ylab = 'frequency', main = 'Gender Distribution', col = colors )
> pie(y, Lables1, main = "CustomerType Distribution", col = colors[3:4])
> hist(Total, main = "Total Sales Volume", xlab = 'Total Sales', col = 'cyan')

> pm <- c(344,311,345)
> pmn <- c('Cash', 'Cr. Card', 'E-wallet')
> barplot(pm,names.arg = pmn, xlab = 'Payment Mode', ylab = 'No. of Transactions', main = 'Payment Mode Dist.', col = colors[5:8] )
```


Unit Price VS Quantity
Quantity
10
8
6
4
2
20
40
60
80
100
Unit Price
Product Line Dist.
No. of Items
150
100
50
0
Electronics
Fashion
food
Health
Home
Sports
Electronics
food
Home
Product Line
Gender Distribution
frequency
500
400
300
200
100
0
Males
Female
Genders
CustomerType Distribution
members
normal
Total Sales Volume
Frequency
200
150
100
50
0
0
200
400
600
800
Total Sales
Payment Mode Dist.
No. of Transactions
300
200
100
50
0
Cash
Cr. Card
E-wallet
Payment Mode