Assignment 1

Bhavika Dobanaboina

```
knitr::opts_chunk$set(message = FALSE)
knitr::opts_chunk$set(warning = FALSE)
```

```
library(readr)
library(dplyr)

#Reading the data:
grocery_data <-read_csv("Dataset.csv")

#Printing Number of Columns
ncol(grocery_data)
```

```
## [1] 16
```

```
#printing number of rows
nrow(grocery_data)
```

```
## [1] 34867
```

```
#printing datatypes of all variables
str(grocery_data)
```

```
## spc_tbl_ [34,867 x 16] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ index           : num [1:34867] 0 1 2 3 4 5 6 7 8 9 ...
##  $ Date            : chr [1:34867] "2/19/2016" "2/20/2016" "2/27/2016" "3/12/2016" ...
##  $ Year            : num [1:34867] 2016 2016 2016 2016 2016 ...
##  $ Month           : chr [1:34867] "February" "February" "February" "March" ...
##  $ Customer Age    : num [1:34867] 29 29 29 29 29 29 29 29 29 29 ...
##  $ Customer Gender : chr [1:34867] "F" "F" "F" "F" ...
##  $ Country         : chr [1:34867] "United States" "United States" "United States" "United States" .
##  $ State           : chr [1:34867] "Washington" "Washington" "Washington" "Washington" ...
##  $ Product Category: chr [1:34867] "Accessories" "Clothing" "Accessories" "Accessories" ...
##  $ Sub Category    : chr [1:34867] "Tires and Tubes" "Gloves" "Tires and Tubes" "Tires and Tubes" ..
##  $ Quantity        : num [1:34867] 1 2 3 2 3 1 2 1 2 2 ...
##  $ Unit Cost       : num [1:34867] 80 24.5 3.67 87.5 35 66 52 60 8 2.5 ...
##  $ Unit Price      : num [1:34867] 109 28.5 5 116.5 41.7 ...
##  $ Cost            : num [1:34867] 80 49 11 175 105 66 104 60 16 5 ...
##  $ Revenue         : num [1:34867] 109 57 15 233 125 78 120 68 20 6 ...
##  $ Column1         : num [1:34867] NA NA NA NA NA NA NA NA NA NA ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   index = col_double(),
```

```
##   ..   Date = col_character(),
##   ..   Year = col_double(),
##   ..   Month = col_character(),
##   ..   `Customer Age` = col_double(),
##   ..   `Customer Gender` = col_character(),
##   ..   Country = col_character(),
##   ..   State = col_character(),
##   ..   `Product Category` = col_character(),
##   ..   `Sub Category` = col_character(),
##   ..   Quantity = col_double(),
##   ..   `Unit Cost` = col_double(),
##   ..   `Unit Price` = col_double(),
##   ..   Cost = col_double(),
##   ..   Revenue = col_double(),
##   ..   Column1 = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```r
#quantitative variables
numerical_variables<-grocery_data[c(5,11,12,13,14,15)]

# categorical variables
category_variables<-grocery_data[c(7:10)]

# printing descriptive statistics for quantitative variable
summary(numerical_variables)
```

```
##   Customer Age      Quantity       Unit Cost         Unit Price
##  Min.   :17.00   Min.   :1.000   Min.   :   0.67   Min.   :   0.667
##  1st Qu.:28.00   1st Qu.:1.000   1st Qu.:  45.00   1st Qu.:  53.667
##  Median :35.00   Median :2.000   Median : 150.00   Median : 179.000
##  Mean   :36.38   Mean   :2.003   Mean   : 349.88   Mean   : 389.232
##  3rd Qu.:44.00   3rd Qu.:3.000   3rd Qu.: 455.00   3rd Qu.: 521.000
##  Max.   :87.00   Max.   :3.000   Max.   :3240.00   Max.   :5082.000
##  NA's   :1       NA's   :1       NA's   :1         NA's   :1
##       Cost         Revenue
##  Min.   :   2   Min.   :   2.0
##  1st Qu.:  85   1st Qu.: 102.0
##  Median : 261   Median : 319.0
##  Mean   : 576   Mean   : 640.9
##  3rd Qu.: 769   3rd Qu.: 902.0
##  Max.   :3600   Max.   :5082.0
##  NA's   :1
```

```r
#printing descriptive statistics for categorical variable

pivot_table<-table(category_variables$`Product Category`)
pivot_table
```

```
##
## Accessories       Bikes    Clothing
##       22534        7093        5239
```

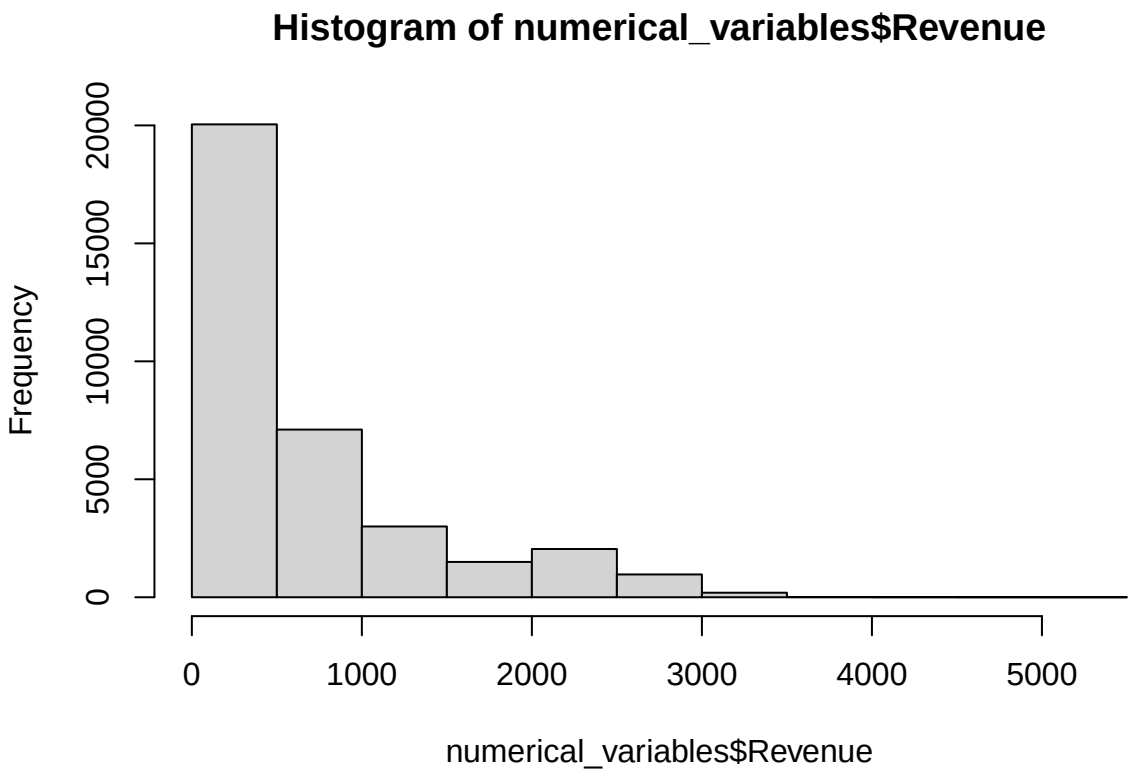
```r
pivot_table1<-table(category_variables$`Sub Category`)
pivot_table1
```

```
##
##        Bike Racks       Bike Stands Bottles and Cages              Caps
##               103               145              5295              1517
##          Cleaners           Fenders             Gloves           Helmets
##               545               762               480              4176
##   Hydration Packs           Jerseys    Mountain Bikes        Road Bikes
##               396              2000              2737              3022
##            Shorts             Socks   Tires and Tubes     Touring Bikes
##               566               364             11112              1334
##             Vests
##               312
```

```r
#transforming variables

filter(category_variables,category_variables$`Product Category` =="Clothing")
```

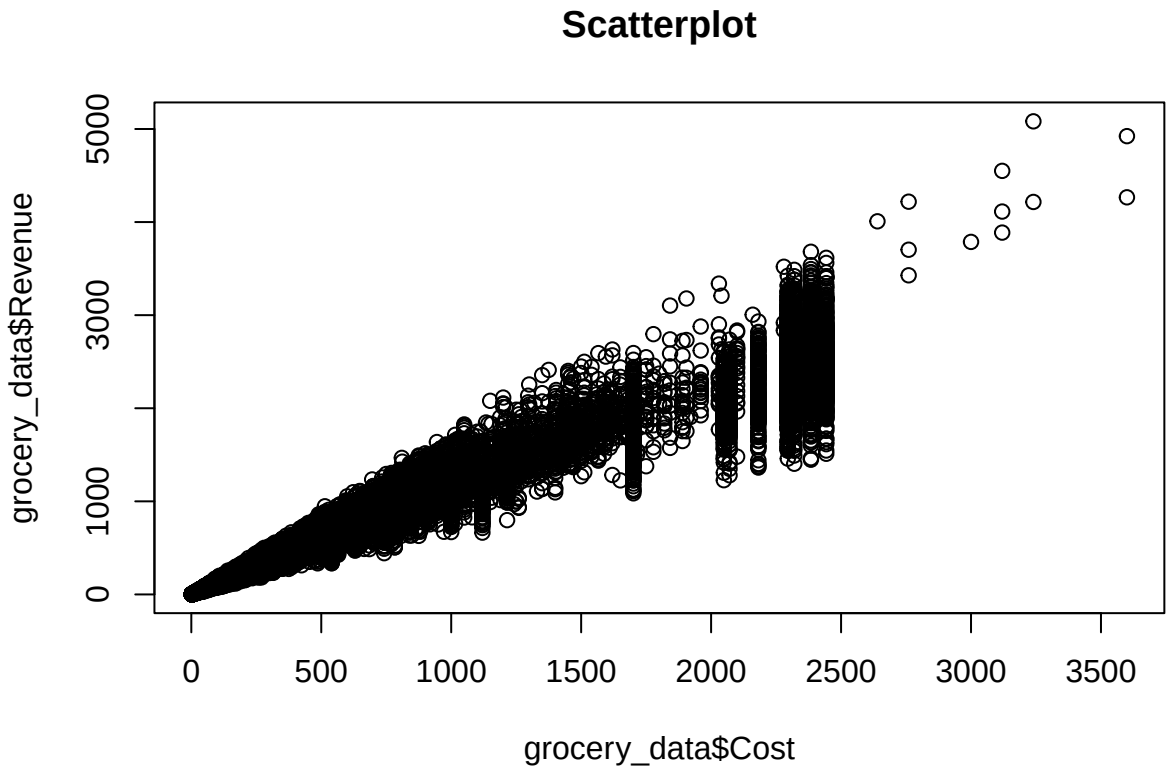
```
## # A tibble: 5,239 x 4
##    Country       State      `Product Category` `Sub Category`
##    <chr>         <chr>      <chr>              <chr>
##  1 United States Washington Clothing           Gloves
##  2 United States California Clothing           Jerseys
##  3 United States California Clothing           Jerseys
##  4 United States California Clothing           Jerseys
##  5 United States California Clothing           Jerseys
##  6 United States California Clothing           Jerseys
##  7 United States California Clothing           Jerseys
##  8 United States California Clothing           Jerseys
##  9 United States Washington Clothing           Socks
## 10 United States Washington Clothing           Socks
## # i 5,229 more rows
```

```r
#plotting one quantitative variable

hist(numerical_variables$Revenue)
```




```
barplot(pivot_table1)
```


```
# plotting scatter plot

plot(x=grocery_data$Cost,y=grocery_data$Revenue,main ="Scatterplot")
```


Scatterplot
grocery_data$Revenue
0
1000
3000
5000
0
500
1000
1500
2000
2500
3000
3500
grocery_data$Cost