library(readxl)

## Warning: package 'readxl' was built under R version 4.4.1

library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.4.1

# Load the Excel file  
file\_path <- "D:\\akshay.xlsx"  
data <- read\_excel(file\_path)  
  
# View the first few rows of the data  
head(data)

## # A tibble: 6 × 17  
## `Invoice ID` Branch City `Customer type` Gender `Product line` `Unit price`  
## <chr> <chr> <chr> <chr> <chr> <chr> <dbl>  
## 1 750-67-8428 A Yangon Member Female Health and be… 74.7  
## 2 226-31-3081 C Naypyi… Normal Female Electronic ac… 15.3  
## 3 631-41-3108 A Yangon Normal Male Home and life… 46.3  
## 4 123-19-1176 A Yangon Member Male Health and be… 58.2  
## 5 373-73-7910 A Yangon Normal Male Sports and tr… 86.3  
## 6 699-14-3026 C Naypyi… Normal Male Electronic ac… 85.4  
## # ℹ 10 more variables: Quantity <dbl>, `Tax 5%` <dbl>, Total <dbl>,  
## # Date <dttm>, Time <dttm>, Payment <chr>, cogs <dbl>,  
## # `gross margin percentage` <dbl>, `Gross Income` <dbl>, Rating <dbl>

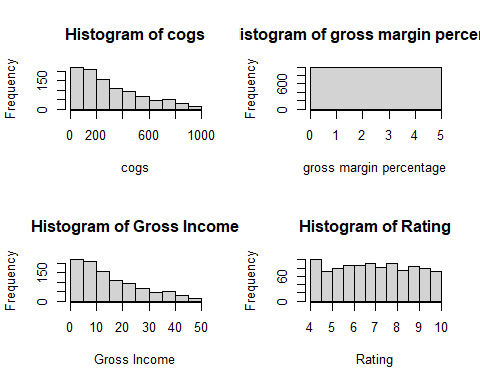
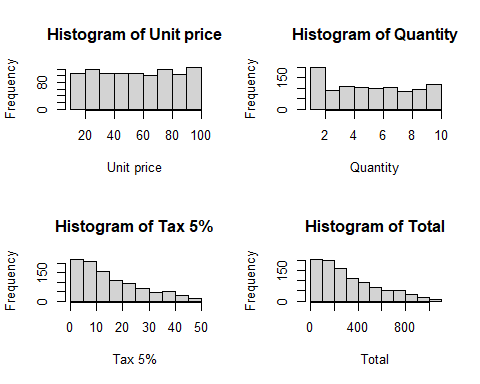
# Inspect the data  
str(data)

## tibble [1,000 × 17] (S3: tbl\_df/tbl/data.frame)  
## $ Invoice ID : chr [1:1000] "750-67-8428" "226-31-3081" "631-41-3108" "123-19-1176" ...  
## $ Branch : chr [1:1000] "A" "C" "A" "A" ...  
## $ City : chr [1:1000] "Yangon" "Naypyitaw" "Yangon" "Yangon" ...  
## $ Customer type : chr [1:1000] "Member" "Normal" "Normal" "Member" ...  
## $ Gender : chr [1:1000] "Female" "Female" "Male" "Male" ...  
## $ Product line : chr [1:1000] "Health and beauty" "Electronic accessories" "Home and lifestyle" "Health and beauty" ...  
## $ Unit price : num [1:1000] 74.7 15.3 46.3 58.2 86.3 ...  
## $ Quantity : num [1:1000] 7 5 7 8 7 7 6 10 2 3 ...  
## $ Tax 5% : num [1:1000] 26.14 3.82 16.22 23.29 30.21 ...  
## $ Total : num [1:1000] 549 80.2 340.5 489 634.4 ...  
## $ Date : POSIXct[1:1000], format: "2019-01-05" "2019-03-08" ...  
## $ Time : POSIXct[1:1000], format: "1899-12-31 13:08:00" "1899-12-31 10:29:00" ...  
## $ Payment : chr [1:1000] "Ewallet" "Cash" "Credit card" "Ewallet" ...  
## $ cogs : num [1:1000] 522.8 76.4 324.3 465.8 604.2 ...  
## $ gross margin percentage: num [1:1000] 4.76 4.76 4.76 4.76 4.76 ...  
## $ Gross Income : num [1:1000] 26.14 3.82 16.22 23.29 30.21 ...  
## $ Rating : num [1:1000] 9.1 9.6 7.4 8.4 5.3 4.1 5.8 8 7.2 5.9 ...

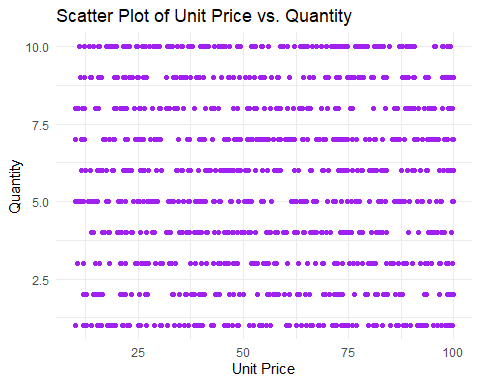
summary(data)

## Invoice ID Branch City Customer type   
## Length:1000 Length:1000 Length:1000 Length:1000   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## Gender Product line Unit price Quantity   
## Length:1000 Length:1000 Min. :10.08 Min. : 1.00   
## Class :character Class :character 1st Qu.:32.88 1st Qu.: 3.00   
## Mode :character Mode :character Median :55.23 Median : 5.00   
## Mean :55.67 Mean : 5.51   
## 3rd Qu.:77.94 3rd Qu.: 8.00   
## Max. :99.96 Max. :10.00   
## Tax 5% Total Date   
## Min. : 0.5085 Min. : 10.68 Min. :2019-01-01 00:00:00.0   
## 1st Qu.: 5.9249 1st Qu.: 124.42 1st Qu.:2019-01-24 00:00:00.0   
## Median :12.0880 Median : 253.85 Median :2019-02-13 00:00:00.0   
## Mean :15.3794 Mean : 322.97 Mean :2019-02-14 00:05:45.5   
## 3rd Qu.:22.4453 3rd Qu.: 471.35 3rd Qu.:2019-03-08 00:00:00.0   
## Max. :49.6500 Max. :1042.65 Max. :2019-03-30 00:00:00.0   
## Time Payment cogs   
## Min. :1899-12-31 10:00:00.00 Length:1000 Min. : 10.17   
## 1st Qu.:1899-12-31 12:43:00.00 Class :character 1st Qu.:118.50   
## Median :1899-12-31 15:19:00.00 Mode :character Median :241.76   
## Mean :1899-12-31 15:24:41.88 Mean :307.59   
## 3rd Qu.:1899-12-31 18:15:00.00 3rd Qu.:448.90   
## Max. :1899-12-31 20:59:00.00 Max. :993.00   
## gross margin percentage Gross Income Rating   
## Min. :4.762 Min. : 0.5085 Min. : 4.000   
## 1st Qu.:4.762 1st Qu.: 5.9249 1st Qu.: 5.500   
## Median :4.762 Median :12.0880 Median : 7.000   
## Mean :4.762 Mean :15.3794 Mean : 6.973   
## 3rd Qu.:4.762 3rd Qu.:22.4453 3rd Qu.: 8.500   
## Max. :4.762 Max. :49.6500 Max. :10.000

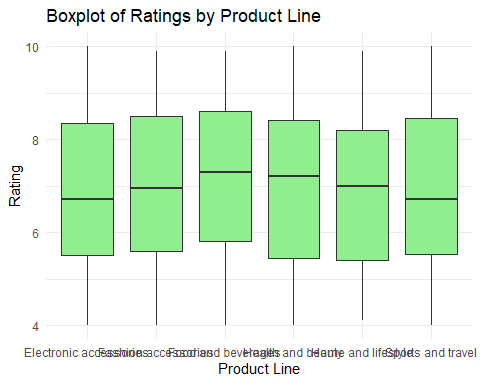
# Histograms  
hist\_data <- data %>% select\_if(is.numeric)  
par(mfrow = c(2, 2))  
for (col in colnames(hist\_data)) {  
 hist(hist\_data[[col]], main = paste("Histogram of", col), xlab = col)  
}



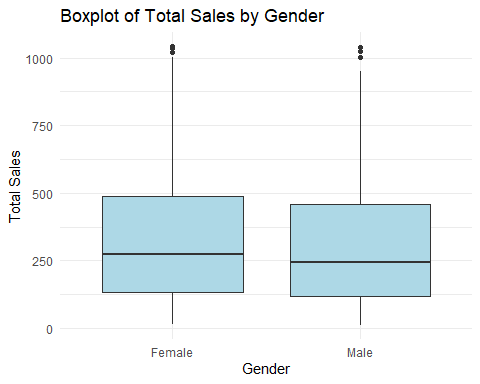
# Scatter plot of Unit Price vs. Quantity  
ggplot(data, aes(x = `Unit price`, y = Quantity)) +  
 geom\_point(color = "purple") +  
 labs(title = "Scatter Plot of Unit Price vs. Quantity", x = "Unit Price", y = "Quantity") +  
 theme\_minimal()



# Boxplot of Ratings by Product Line  
ggplot(data, aes(x = `Product line`, y = Rating)) +  
 geom\_boxplot(fill = "lightgreen") +  
 labs(title = "Boxplot of Ratings by Product Line", x = "Product Line", y = "Rating") +  
 theme\_minimal()



# Boxplot by categorical variable (e.g., Gender)  
ggplot(data, aes(x = Gender, y = Total)) +  
 geom\_boxplot(fill = "lightblue") +  
 labs(title = "Boxplot of Total Sales by Gender", x = "Gender", y = "Total Sales") +  
 theme\_minimal()



# Time series plot  
ggplot(data, aes(x = Date, y = Total)) +  
 geom\_line(color = "blue") +  
 labs(title = "Time Series of Total Sales", x = "Date", y = "Total Sales") +  
 theme\_minimal()

