

# ADIDAS SALES DATA

Exploratory Data analysis on Adidas  
sales data using R Progrraming



# Introduction - Adidas

Adidas was founded in 1949 by Adolf Dassler in Germany.

The company has grown to become one of the largest sportswear manufacturers in the world.

The brand's iconic logo, consisting of three parallel stripes, symbolizes performance, quality, and durability.

Beyond performance gear, Adidas has also made a significant impact on street fashion, partnering with celebrities and designers such as Kanye West for the Yeezy line.



# Objective of the project

The primary objective of this project is to conduct Exploratory Data Analysis (EDA) on a selected dataset to extract meaningful insights and understand underlying patterns. This involves a comprehensive examination of the data's structure, contents, and interrelationships through various analytical techniques. The EDA process includes:

## **Data Understanding and Cleaning**

## **Univariate Analysis**

## **Bivariate and Multivariate Analysis**



# Research Methodology

**Data Collection**

**Data Cleaning**

**Exploratory Data Analysis (EDA)**

**Statistical Analysis**



# Data and Variables

Retailer

Retailer\_name

Invoice.Date

Region

State

City

Product\_name

Price.per.unit

Units\_sold

Total\_sales

Operating.profit

Operating.margin

Sales.method

# Question 1 :- Identify the most common retailer who deals in adidas shoes

```
most_common_retailer_name ← Adidas_sales_data %>%  
  count(Retailer_name) %>%  
  arrange(desc(n)) %>% |  
  slice(5)  
  
print(most_common_retailer_name)  
View(most_common_retailer_name)
```

	Retailer_name	n
1	West Gear	72

## Question 2 :- Find the average of total sales by particular region

```
avg_total_sales_by_region <- Adidas_sales_data %>%
  group_by(Region) %>%
  summarize(avg_total_sales_by_region = mean(Total_Sales, na.rm = TRUE)) %>%
  arrange(desc(avg_total_sales_by_region))

View(head(avg_total_sales_by_region))
```

	Region	avg_total_sales_by_region
1	West	398099.0
2	Northeast	343338.9
3	South	306692.7
4	Midwest	132614.6

# Question 3 :- Identify the most common product sold by adidas

```
most_common_product_name ← Adidas_sales_data %>%  
  count(Product_name) %>%  
  arrange(desc(n)) %>%  
  slice(5)  
  
print(most_common_product_name)  
View(most_common_product_name)
```

	Product_name	n
1	Women's Athletic Footwear	82

# Question 4 :- Find the top 5 highest selling products along with their unit sold.

```
top_5_highest_selling_products <- Adidas_sales_data %>%
  select(Product_name, Total_Sales, Units.Sold) %>%
  filter(!is.na(Total_Sales), !is.na(Units.Sold)) %>%
  arrange(desc(Total_Sales)) %>%
  slice(1:5)

View(top_5_highest_selling_products)
```

	Product_name	Total_Sales	Units.Sold
1	Men's Street Footwear	780000	1200
2	Men's Street Footwear	765000	1275
3	Men's Street Footwear	750000	1250
4	Men's Street Footwear	750000	1250
5	Men's Street Footwear	747500	1150

# Question 5 :- Find the 5 lowest selling products along with their unit sold.

```
top_5_lowest_selling_products <- Adidas_sales_data %>%
  select(Product_name, Total_Sales, Units.Sold) %>%
  filter(!is.na(Total_Sales), !is.na(Units.Sold)) %>%
  arrange(Total_Sales) %>%
  slice(1:5)

View(top_5_lowest_selling_products)
```

	Product_name	Total_Sales	Units.Sold
1	Women's Athletic Footwear	11250	25
2	Women's Athletic Footwear	20000	50
3	Women's Athletic Footwear	30000	75
4	Women's Athletic Footwear	33750	75
5	Women's Athletic Footwear	33750	75

# Question 6 :- Analyze the impact of sales method on total sales of Adidas

```
Sales_Method_impact <- Adidas_sales_data %>%  
  group_by(Sales.Method) %>%  
  summarize(avg_total_sales = mean(Total_Sales, na.rm = TRUE))  
  
View(Sales_Method_impact)
```

	Sales.Method	avg_total_sales
1	In-store	506785.7
2	Outlet	315249.0

## Question 7 :- Find the median of no of units sold

```
median_Units_sold<- median  
(Adidas_sales_data$Units.Sold)  
print(paste("Median Units Sold:  
", median_Units_sold))
```

"Median Units Sold: 675"

## Question 8 :- Find the mean of Total sales

```
mean_total_sales <- mean  
(Adidas_sales_data$Total_Sales)  
print(paste("Mean of Total Sales:  
", mean_total_sales))
```

"Mean of Total Sales: 317957.575757576"

# Question 9 :- Calculate the range of operating profit

```
range_operating_profit ← range(Adidas_sales_data$Operating_Profit)  
print(paste("Range of Operating Profit:", diff(range_operating_profit)))
```

```
"Range of Operating Profit: 385500"
```

# Question 10 :- Calculate the variance of operating margin

```
var_operating_margin <- var(Adidas_sales_data$Operating.Margin)  
print(paste("Variance of Operating margin:", var_operating_margin))
```

```
"Variance of Operating margin: 0.00878746984010142"
```

## Question 11 :- Calculate correlation between units sold and total sales

```
correlation ← cor(Adidas_sales_data$Units.Sold,  
                  Adidas_sales_data$Total_Sales,  
                  method = "pearson")  
print(correlation)
```

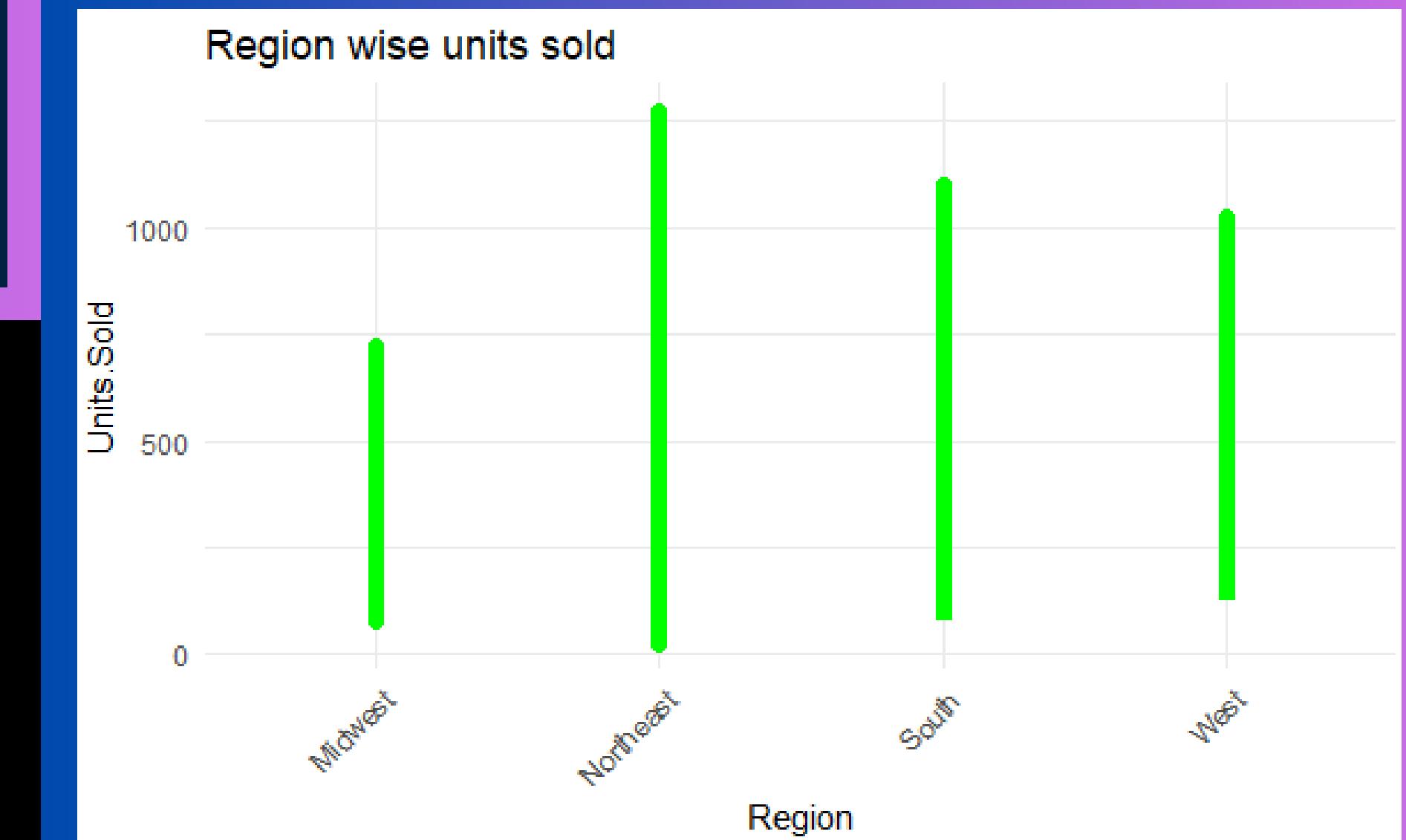
0.9024587

# Question 12 :- Create a line chart of Region wise Units sold

```
Adidas_sales_dataset_clean <- Adidas_sales_data %>%
  filter(!is.na(Region), !is.na(Units.Sold)) %>%
  mutate(Units.Sold)

# Create a line chart of Region wise Units sold

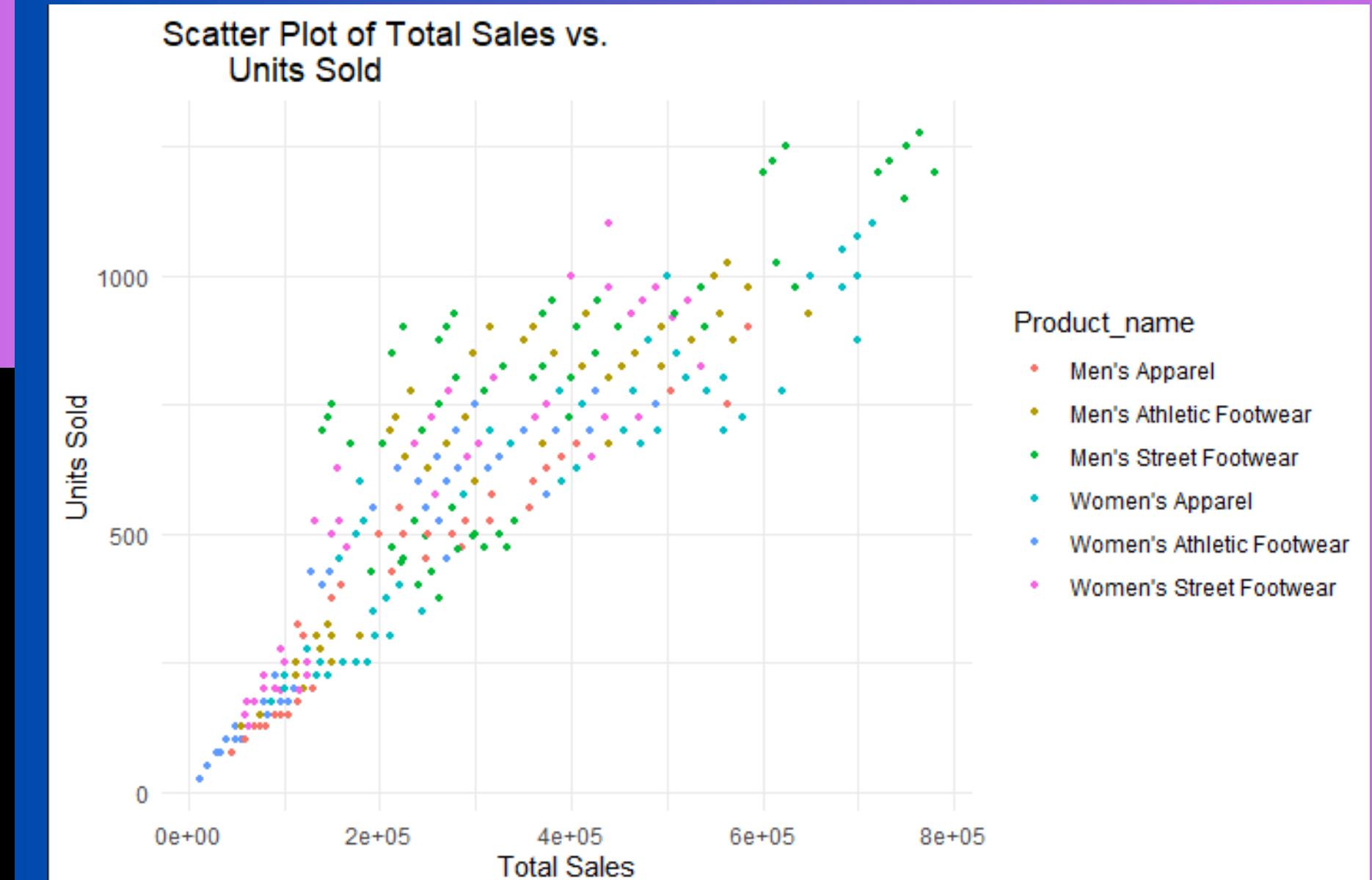
ggplot(Adidas_sales_data, aes(x = Region, y = Units.Sold)) +
  geom_line(color = "green", size = 2.5) +
  labs(title = "Region wise units sold",
       x = "Region",
       y = "Units.Sold") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



# Question 13 :- Create a scatter plot of Total sales VS. Units sold

```
Adidas_sales_dataset_clean <- Adidas_sales_data %>%
  filter(!is.na(Total_Sales), !is.na(Units.Sold))

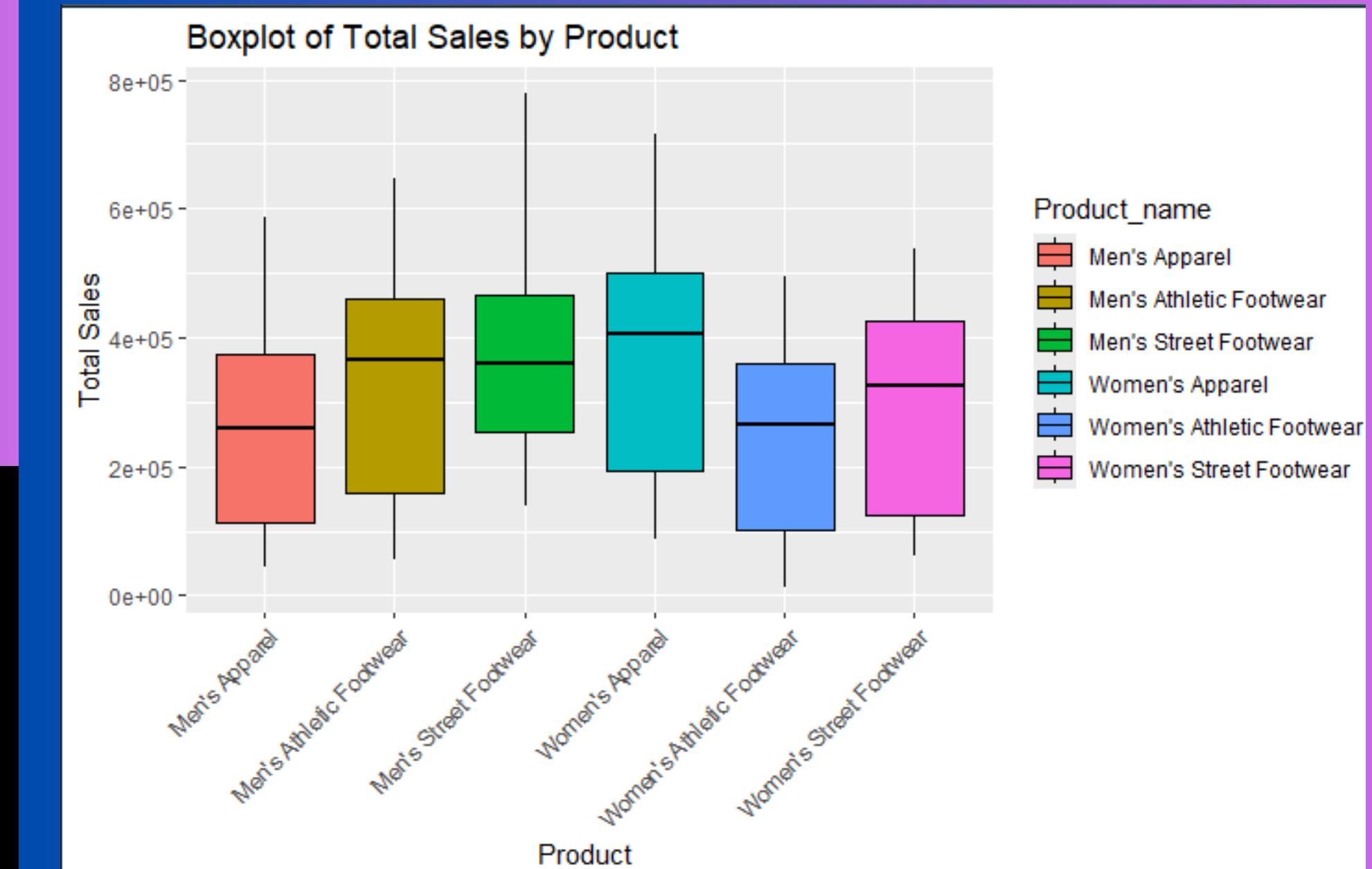
# Create a scatter plot of Total_Sales vs. Units_Sold
ggplot(Adidas_sales_dataset_clean, aes(x = Total_Sales,
                                         y = Units.Sold)) +
  geom_point(aes(color = Product_name), size = 1) +
  labs(title = "Scatter Plot of Total Sales vs.
        Units Sold",
       x = "Total Sales",
       y = "Units Sold") +
  theme_minimal()
```



# Question 14 :- Create a box plot of total sales by product

```
Adidas_sales_dataset_clean <- Adidas_sales_data %>%
  filter(!is.na(Total_Sales), !is.na(Units.Sold))

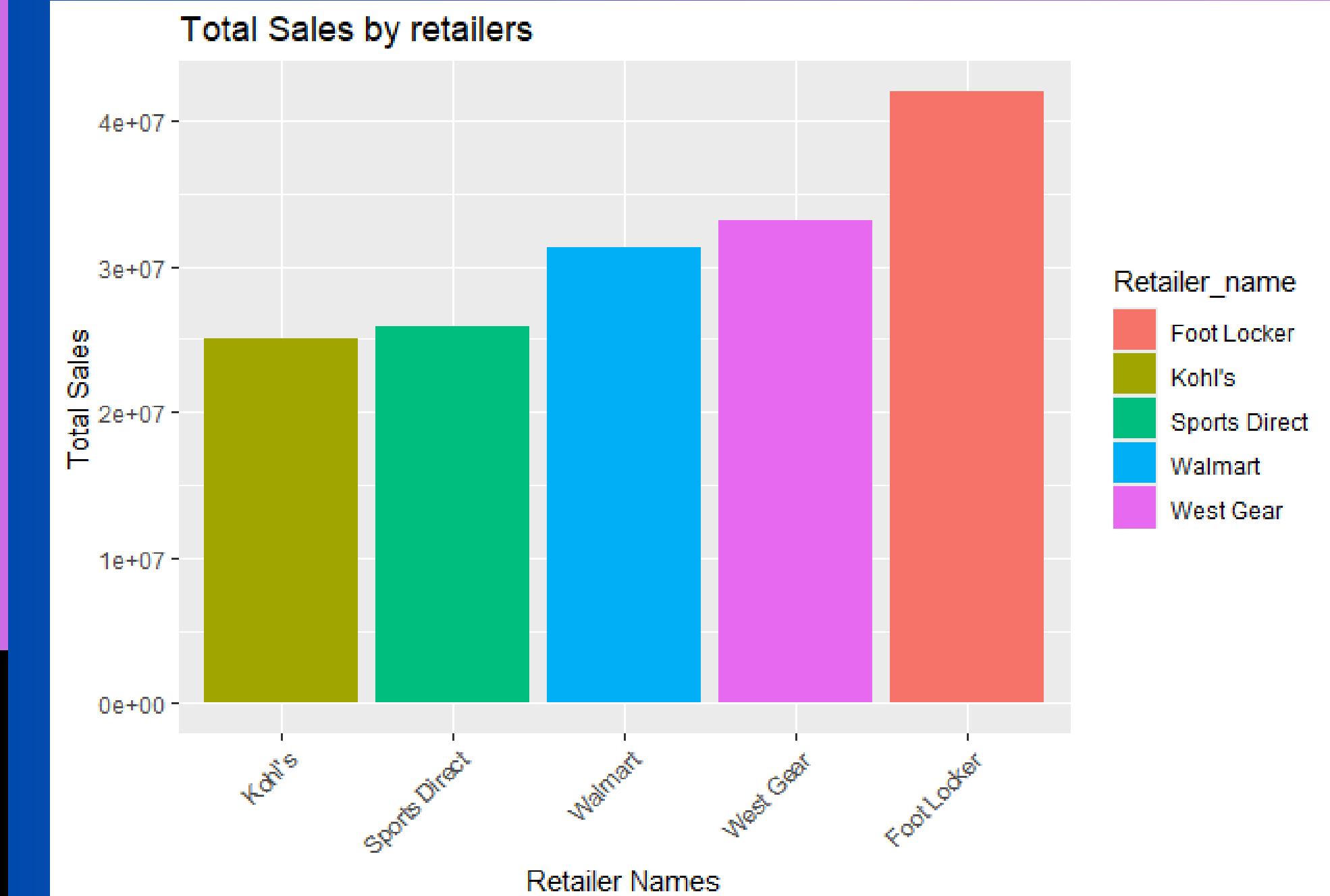
# Boxplot for Total Sales by Product
ggplot(Adidas_sales_dataset_clean, aes(x =
  Product_name, y = Total_Sales)) +
  geom_boxplot(aes(fill = Product_name,
                    , color = "black") +
  labs(title = "Boxplot of Total Sales by Product",
       x = "Product",
       y = "Total Sales") +
  theme(axis.text.x = element_text(angle
    = 45, hjust = 1))
```



# Question 15 :- Create a bar graph of total sales made by retailers

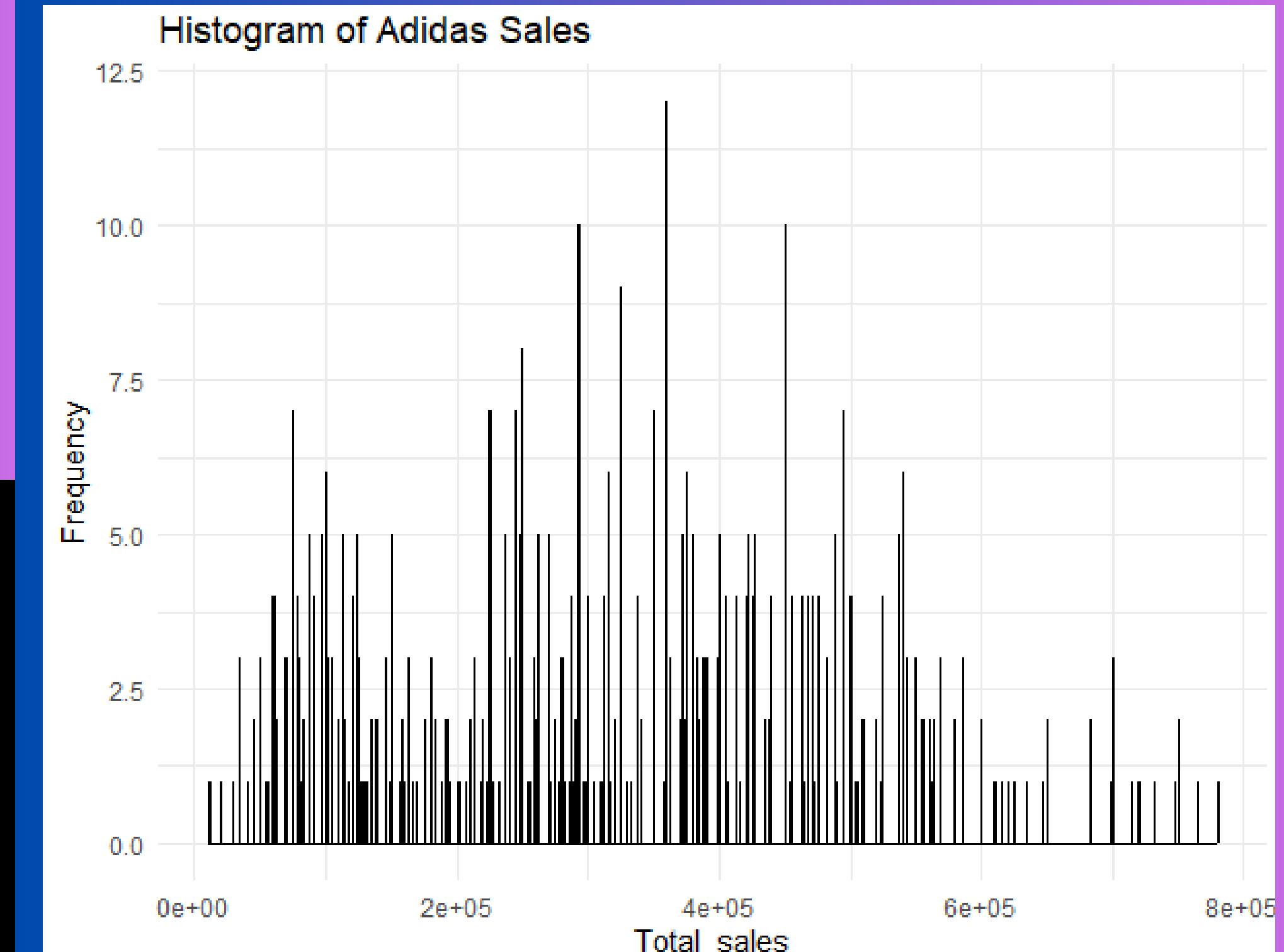
```
sales_summary <- Adidas_sales_data %>%
  group_by(Retailer_name) %>%
  summarise(Total_Sales = sum(Total_Sales,
                               na.rm = TRUE)) %>%
  arrange(desc(Total_Sales))

# Plotting the bar graph using ggplot2
ggplot(sales_summary, aes(x = reorder
                           (Retailer_name, Total_Sales), y = Total_Sales,
                           fill = Retailer_name)) +
  geom_bar(stat = "identity") +
  labs(title = "Total product sold by retailers",
       x = "Retailer Names", y
       = "Total Sales") +
  theme(axis.text.x = element_text(angle = 45,
                                   hjust = 1))
```



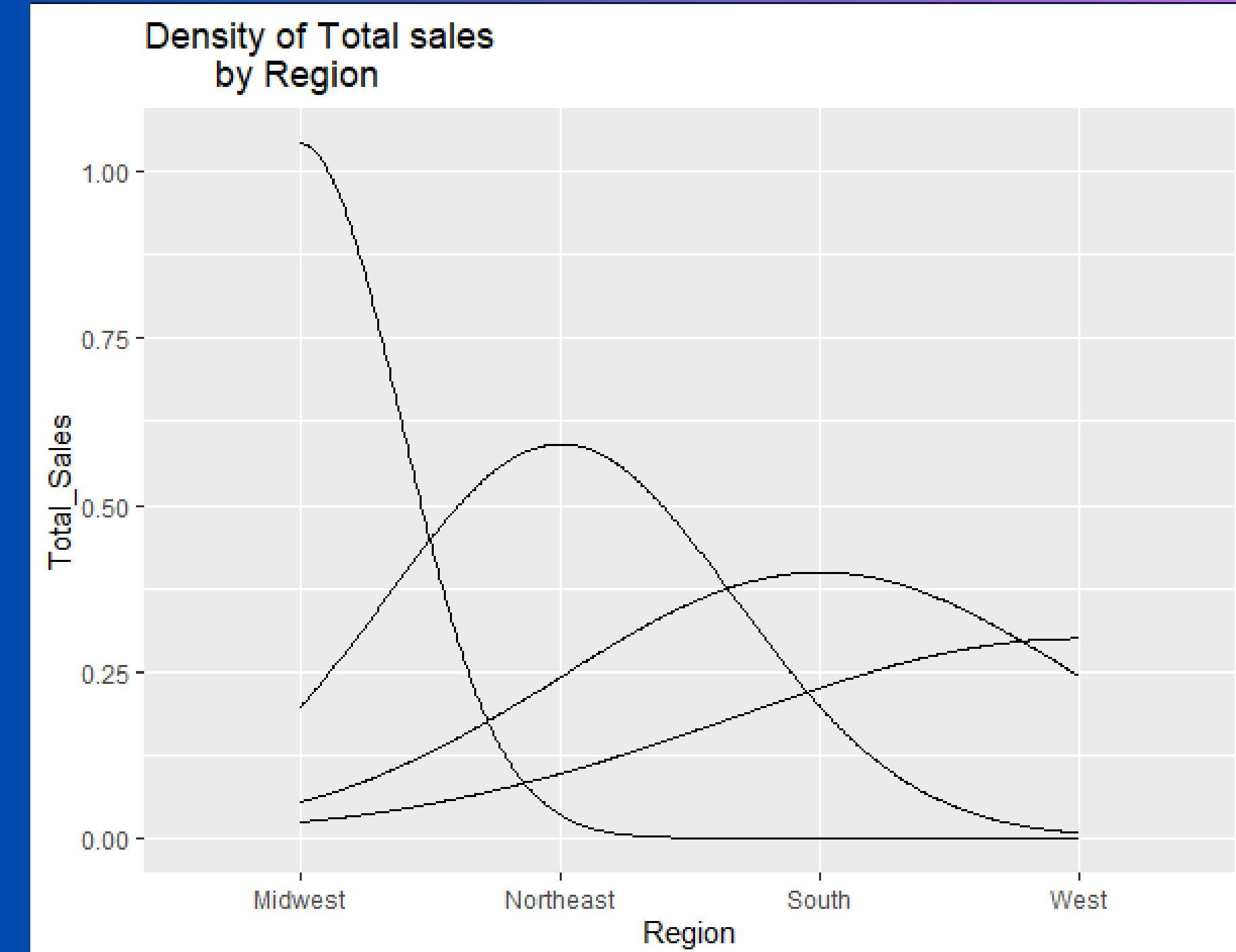
# Question 16 :- Create a histogram of total sales

```
head(Adidas_sales_data)
ggplot(Adidas_sales_data, aes
      (x = Total_Sales)) +
  geom_histogram(binwidth = 100, fill =
    "green", color = "black", alpha = 0.7) +
  labs(title = "Histogram of Adidas Sales",
       x = "Total_sales",
       y = "Frequency") +
  theme_minimal()
```



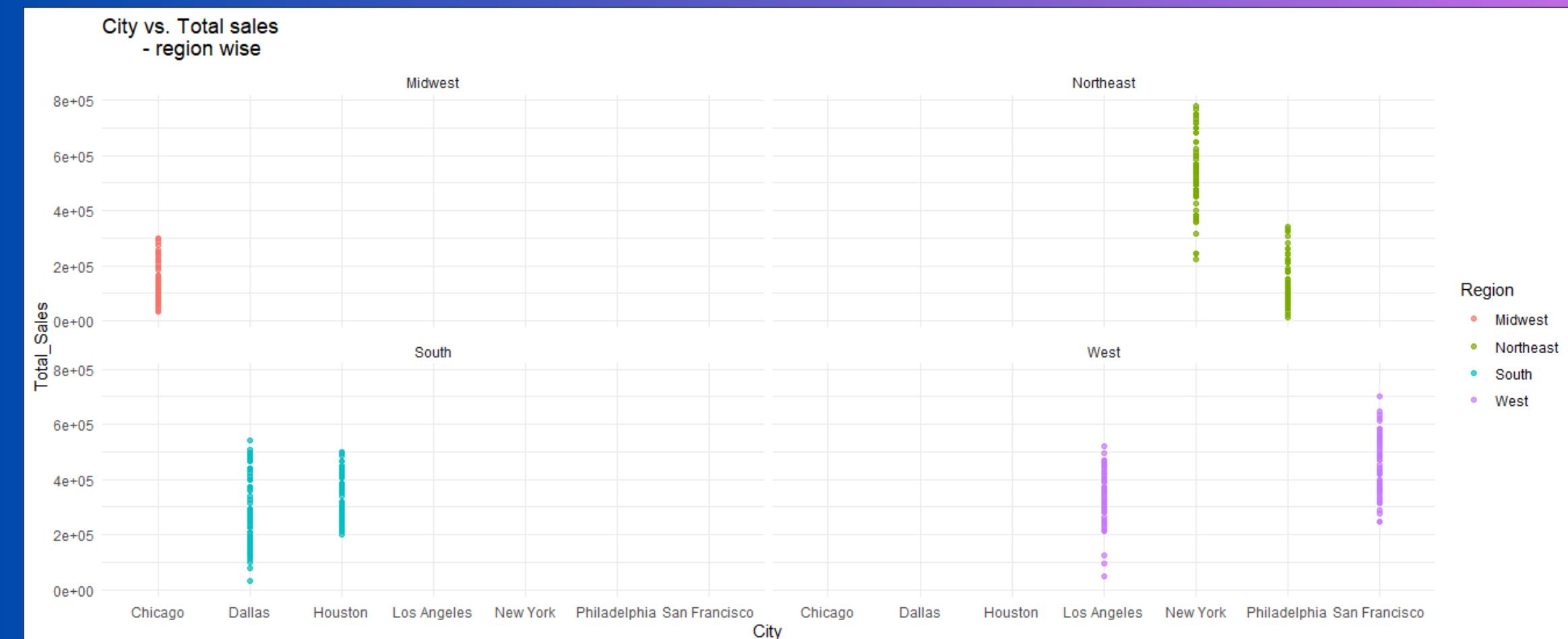
# Question 17 :- Create a density chart of total sales - region wise

```
ggplot(Adidas_sales_data, aes  
      (x = Region, fill = Total_Sales)) +  
  geom_density(alpha = 1.5) +  
  labs(title = "Density of Total sales  
        by Region", x = "Region",  
       y = "Total_Sales")
```



# Question 18 :- Create a faceted chart of city vs. total sales region wise

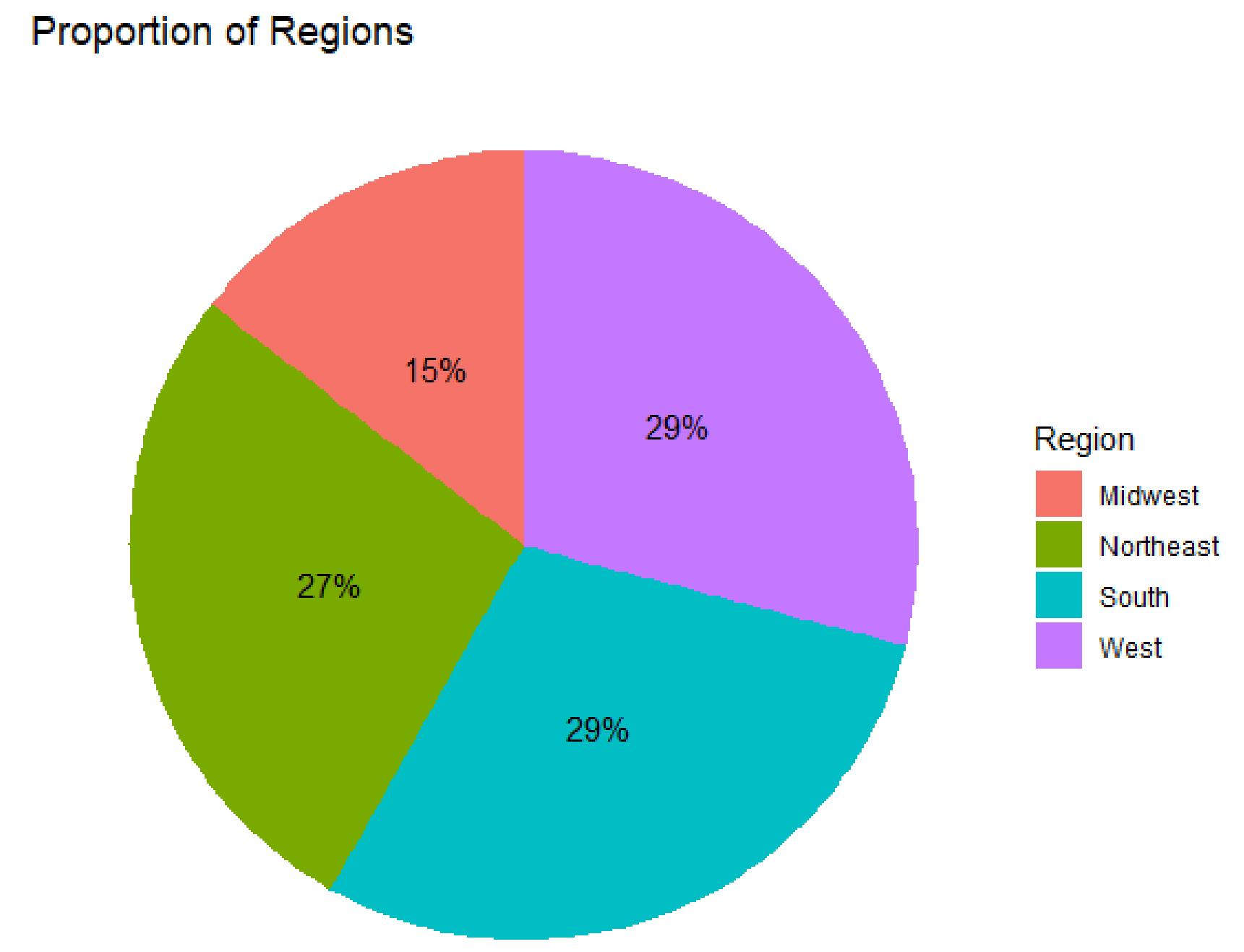
```
ggplot(Adidas_sales_data, aes  
      (x = City, y = Total_Sales)) +  
  geom_point(aes(color = Region),  
             alpha = 0.7) +  
  facet_wrap(~ Region) +  
  labs(title = "City vs. Total sales  
        - region wise", x = "City",  
       y = "Total_Sales") +  
  theme_minimal()
```



# Question 19 :- Create a pie chart showing proportion of regions

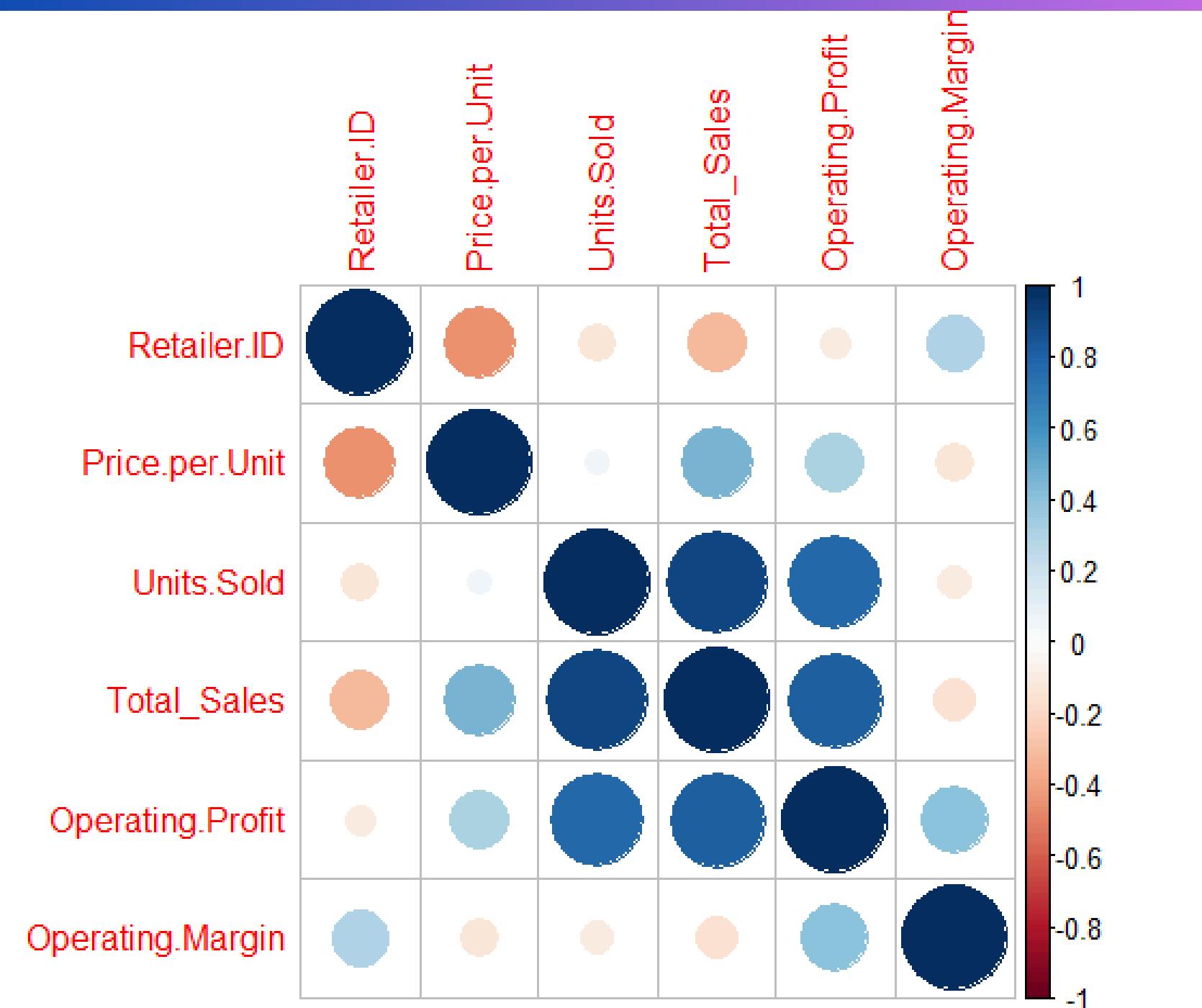
```
product_sold_in_regions <- Adidas_sales_data %>%
  count(Region) %>%
  mutate(percentage = n / sum(n) * 100)
print(product_sold_in_regions)

# Create a pie chart with percentages
ggplot(product_sold_in_regions , aes(x = "", y = percentage,
                                       fill = Region)) +
  geom_bar(width = 1, stat = "identity") +
  coord_polar(theta = "y") +
  geom_text(aes(label = paste0(round(percentage,0), "%")),
            position = position_stack(vjust = 0.5)) +
  labs(title = "Proportion of Regions") +
  theme_void()
```



# Question 20 :- Create a correlation matrix graph

```
install.packages("corrplot")
library(corrplot)
cor_matrix <- cor(select_if(Adidas_sales_data
    , is.numeric), use = "complete.obs")
corrplot(cor_matrix, method = "circle")
```



# Key Insights

- 1. Analysis Focus:** The project analyzed Adidas sales data across multiple dimensions.
- 2. Key Variables:** The analysis tracked critical metrics.
- 3. Analysis Depth:** The study incorporated the Statistical measures.
- 4. Visualization Approach:** Multiple visualization techniques were used to understand patterns.

# CONCLUSION

This project conducted a comprehensive Exploratory Data Analysis (EDA) on Adidas sales data.

The methodology included data cleaning using `na.omit()` and `distinct()` functions followed by thorough univariate, bivariate, and multivariate analyses.

The study explored multiple variables like retailer, units sold.

Various visualization techniques were used, like line charts, scatter plots.

The analysis aimed to uncover sales patterns, identify top-performing products and retailers understand the relationship between different sales metrics through statistical measures

Thank You!