# ONLINE SALES DATASET

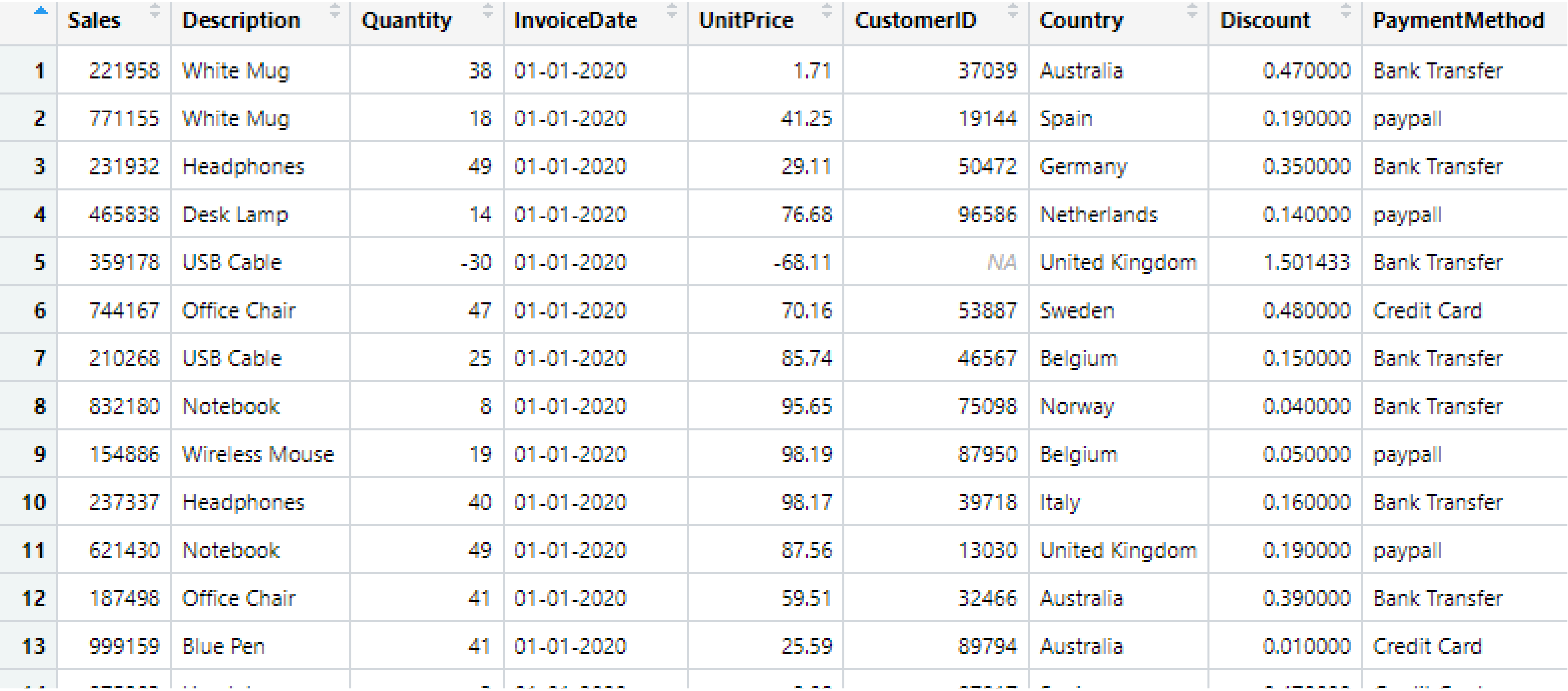
# INTRODUCTION

## Purpose and Scope

**To analyze online sales data for insights into product performance, regional trends, and discounts. Focus on categories, regions, sales, country, pricing, and discounts.**

# LOAD & VIEW DATASET

```
Sales_Data <- read.csv("C:/Users/Vanshika Gupta/Desktop/online
sale dataset.csv")
View(Sales_Data)
```

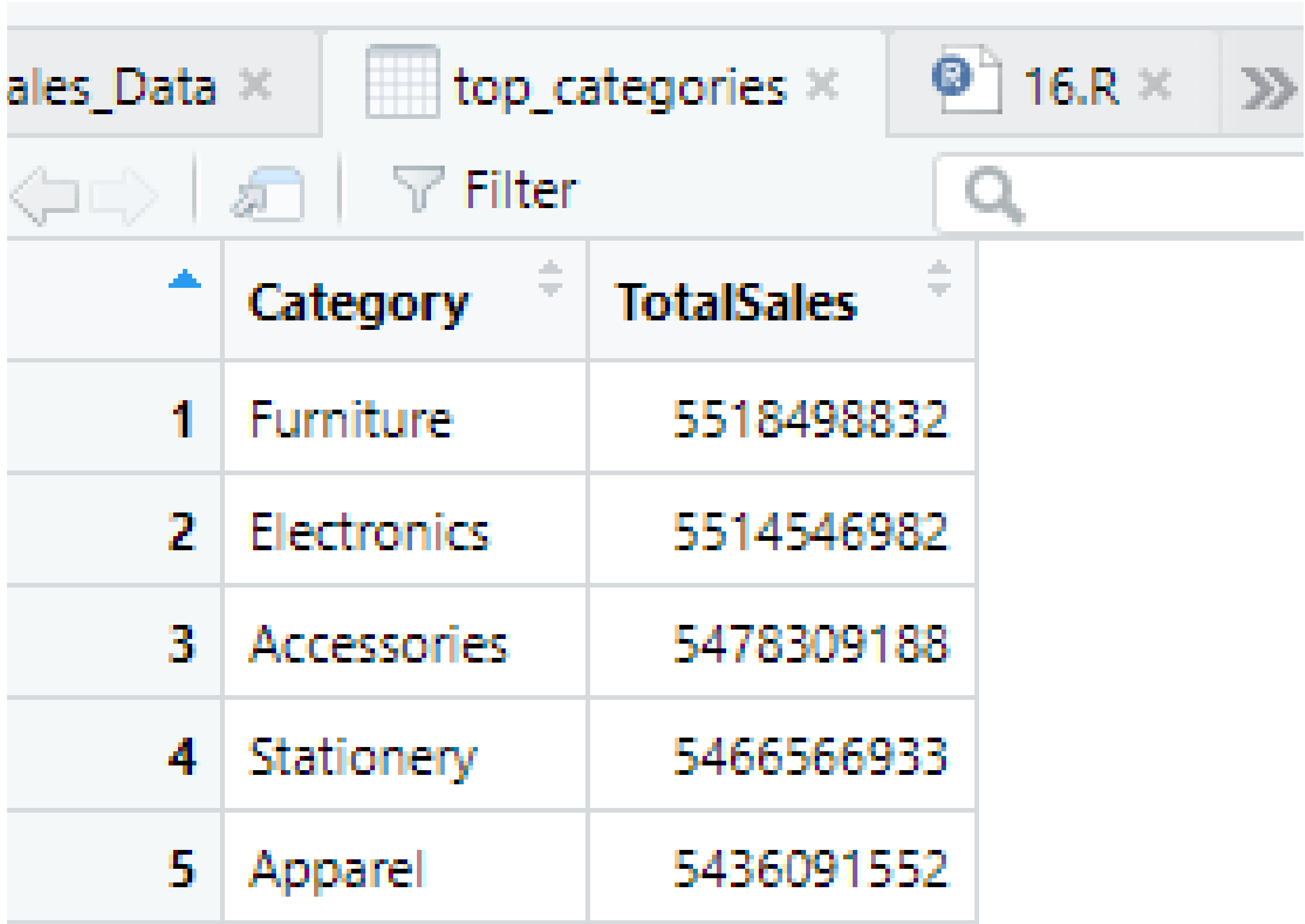


| | Sales | Description | Quantity | InvoiceDate | UnitPrice | CustomerID | Country | Discount | PaymentMethod |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 221958 | White Mug | 38 | 01-01-2020 | 1.71 | 37039 | Australia | 0.470000 | Bank Transfer |
| 2 | 771155 | White Mug | 18 | 01-01-2020 | 41.25 | 19144 | Spain | 0.190000 | paypall |
| 3 | 231932 | Headphones | 49 | 01-01-2020 | 29.11 | 50472 | Germany | 0.350000 | Bank Transfer |
| 4 | 465838 | Desk Lamp | 14 | 01-01-2020 | 76.68 | 96586 | Netherlands | 0.140000 | paypall |
| 5 | 359178 | USB Cable | -30 | 01-01-2020 | -68.11 | NA | United Kingdom | 1.501433 | Bank Transfer |
| 6 | 744167 | Office Chair | 47 | 01-01-2020 | 70.16 | 53887 | Sweden | 0.480000 | Credit Card |
| 7 | 210268 | USB Cable | 25 | 01-01-2020 | 85.74 | 46567 | Belgium | 0.150000 | Bank Transfer |
| 8 | 832180 | Notebook | 8 | 01-01-2020 | 95.65 | 75098 | Norway | 0.040000 | Bank Transfer |
| 9 | 154886 | Wireless Mouse | 19 | 01-01-2020 | 98.19 | 87950 | Belgium | 0.050000 | paypall |
| 10 | 237337 | Headphones | 40 | 01-01-2020 | 98.17 | 39718 | Italy | 0.160000 | Bank Transfer |
| 11 | 621430 | Notebook | 49 | 01-01-2020 | 87.56 | 13030 | United Kingdom | 0.190000 | paypall |
| 12 | 187498 | Office Chair | 41 | 01-01-2020 | 59.51 | 32466 | Australia | 0.390000 | Bank Transfer |
| 13 | 999159 | Blue Pen | 41 | 01-01-2020 | 25.59 | 89794 | Australia | 0.010000 | Credit Card |

# KEY FINDINGS AND VISUALIZATION

# TOP 5 CATEGORIES BY SALES

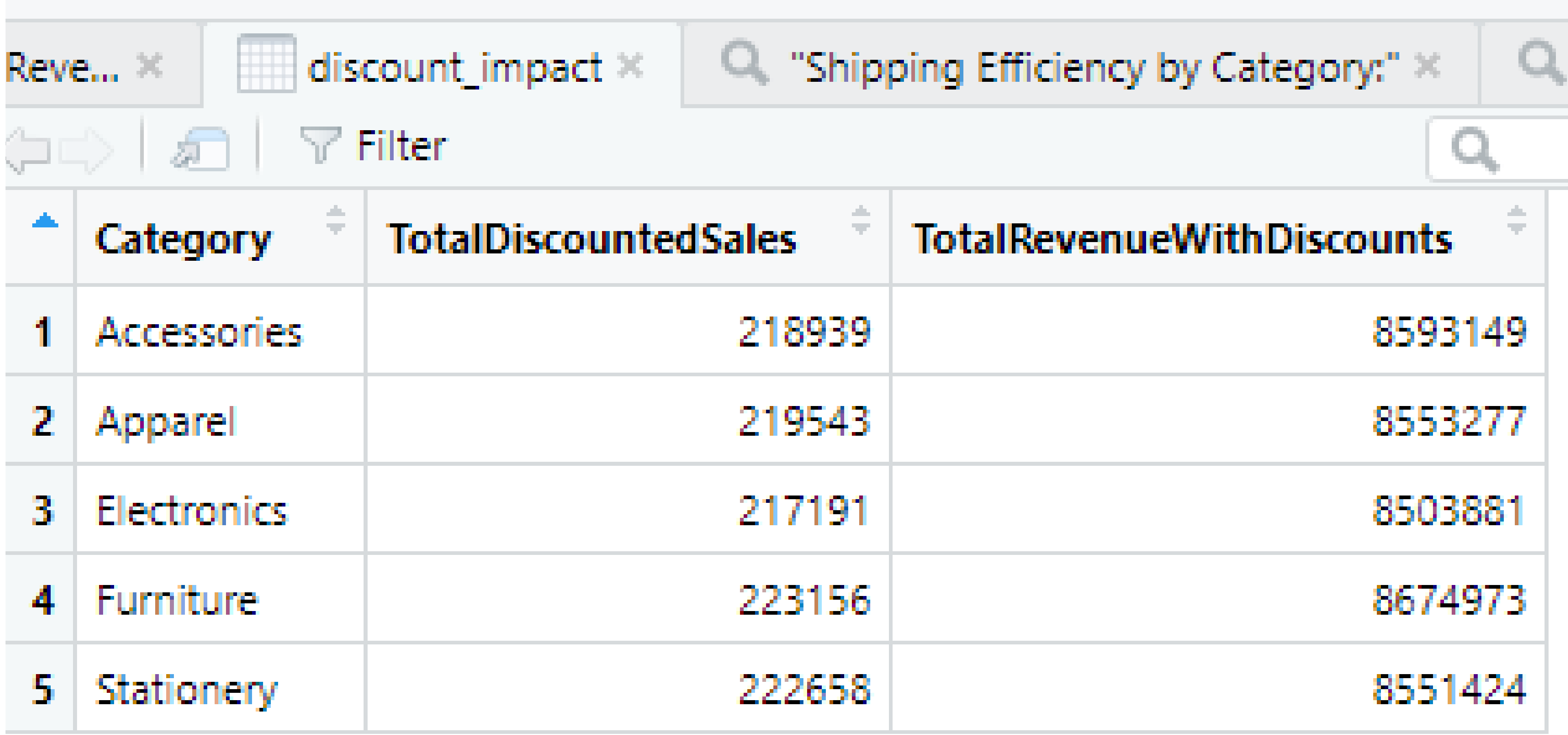
```
top_categories <- Sales_Data %>%
group_by(Category) %>%
summarise(TotalSales = sum(Sales)) %>%
arrange(desc(TotalSales)) %>%
head(5)
```



| | Category | TotalSales |
|---|---|---|
| 1 | Furniture | 5518498832 |
| 2 | Electronics | 5514546982 |
| 3 | Accessories | 5478309188 |
| 4 | Stationery | 5466566933 |
| 5 | Apparel | 5436091552 |

# DISCOUNT IMPACT- SALES VOLUME AND REVENUE WITH DISCOUNTS

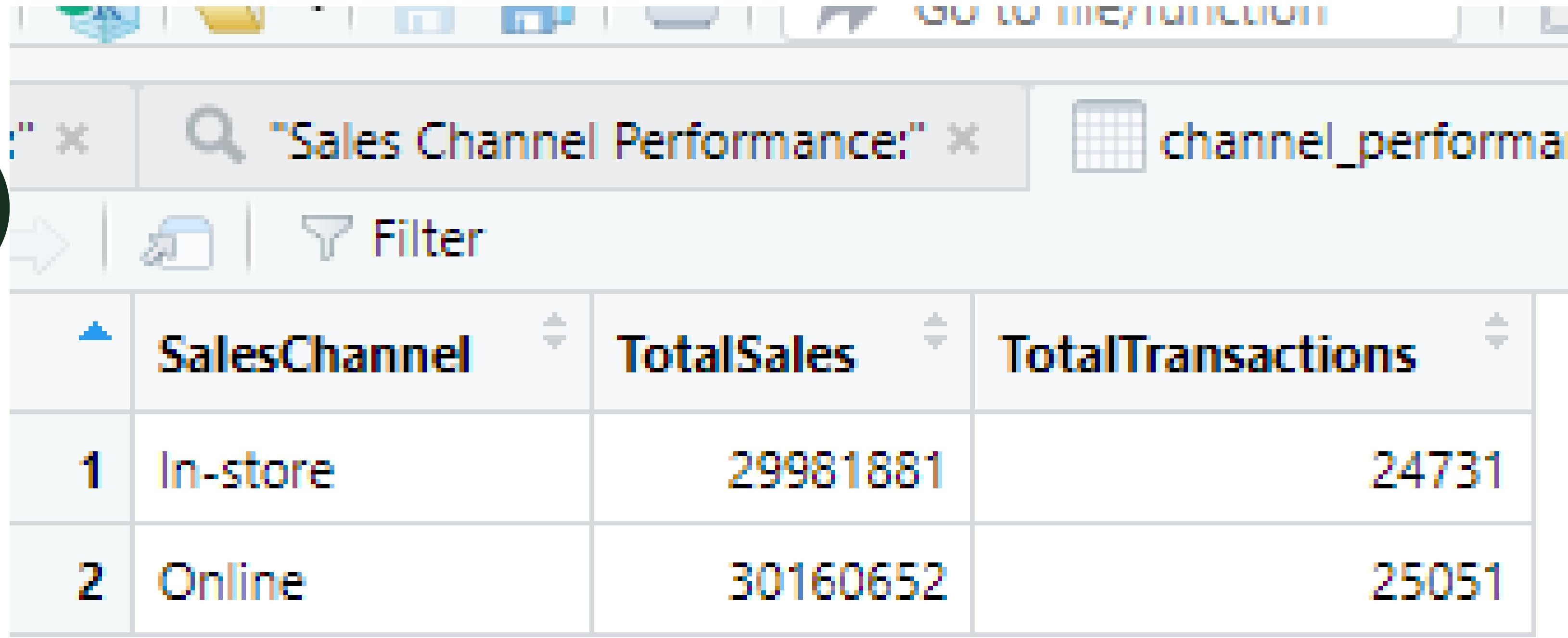
```
discount_impact <- Sales_Data %>%
filter(Discount > 0) %>%
group_by(Category) %>%
summarise(TotalDiscountedSales =
sum(Quantity),
TotalRevenueWithDiscounts = sum
((Quantity * UnitPrice)  *
(1 - Discount), na.rm = TRUE))

View(discount_impact)
```



|  | Category | TotalDiscountedSales | TotalRevenueWithDiscounts |
|---|---|---|---|
| 1 | Accessories | 218939 | 8593149 |
| 2 | Apparel | 219543 | 8553277 |
| 3 | Electronics | 217191 | 8503881 |
| 4 | Furniture | 223156 | 8674973 |
| 5 | Stationery | 222658 | 8551424 |

# SALES CHANNEL PERFORMANCE: COMPARE ONLINE VS IN-STORE SALES

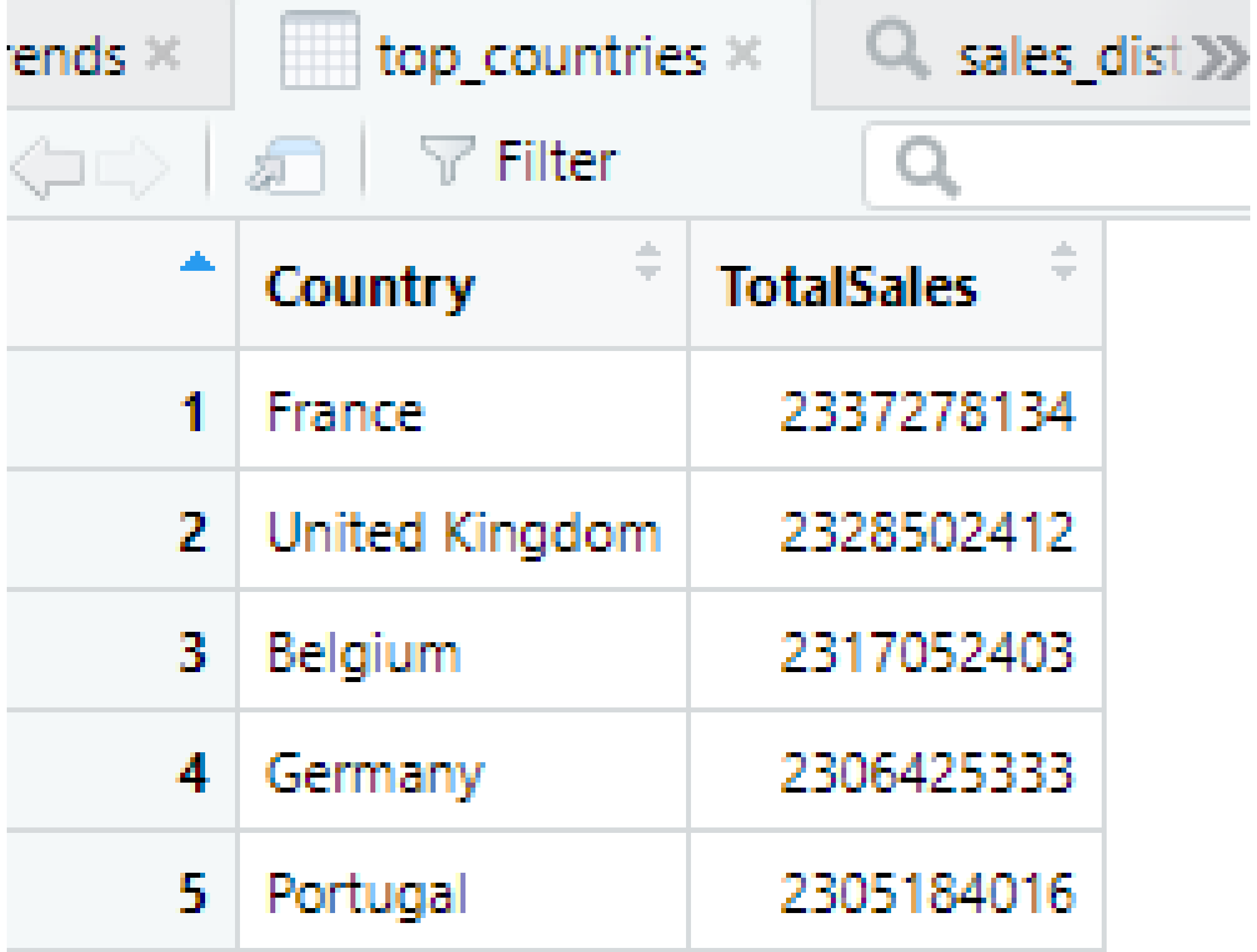
```
channel_performance <- Sales_Data %>%
group_by(SalesChannel) %>%
summarise(TotalSales = sum(Quantity *
UnitPrice, na.rm = TRUE),
TotalTransactions = n())

View(channel_performance)
```



|   | SalesChannel | TotalSales | TotalTransactions |
|---|---|---|---|
| 1 | In-store | 29981881 | 24731 |
| 2 | Online | 30160652 | 25051 |

# TOP 5 COUNTRIES BY TOTAL SALES

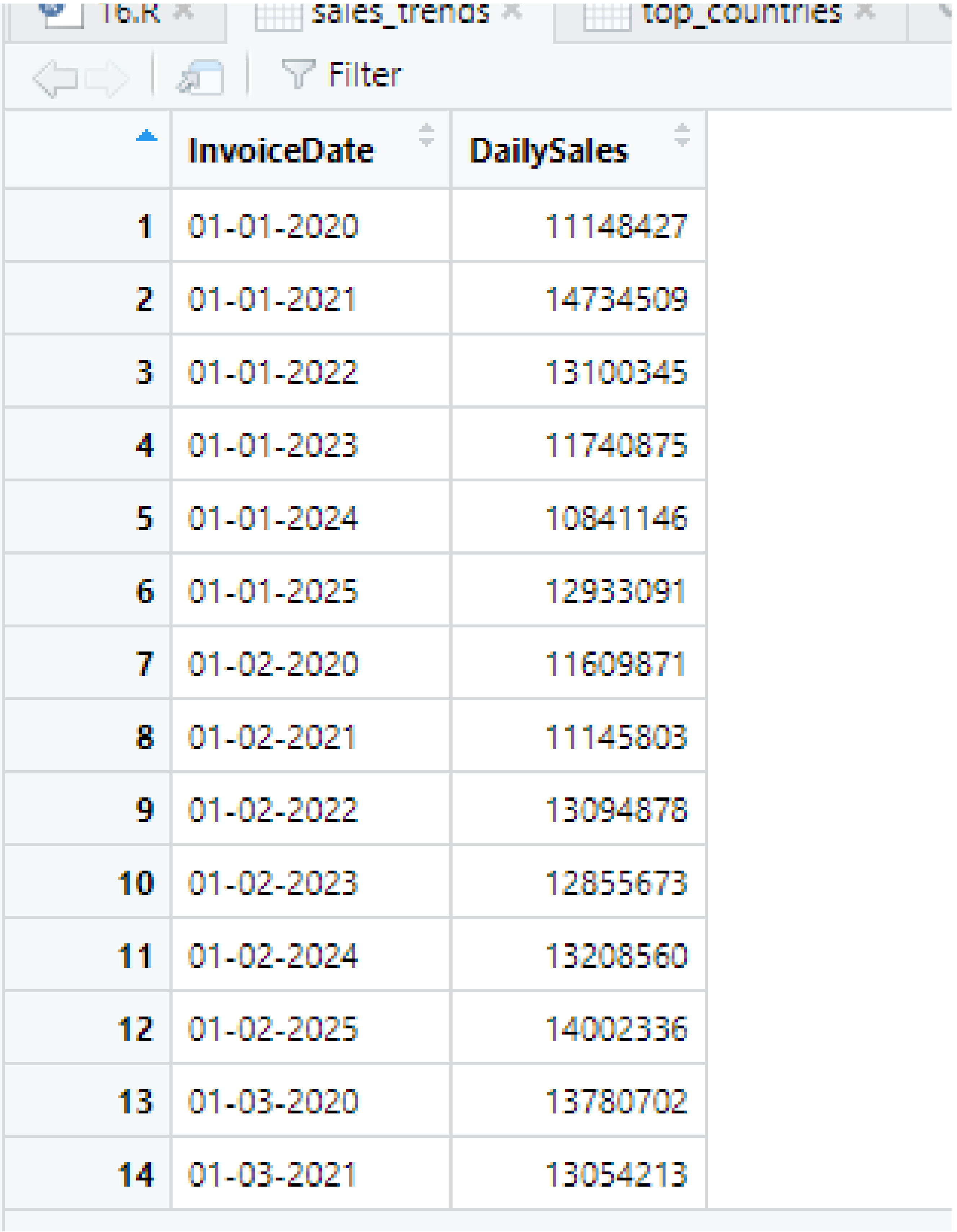
```
top_countries <- Sales_Data %>%
group_by(Country) %>%
summarise(TotalSales = sum(Sales)) %>%
arrange(desc(TotalSales)) %>%
head(5)

View(top_countries)
```



| | Country | TotalSales |
|---|---|---|
| 1 | France | 2337278134 |
| 2 | United Kingdom | 2328502412 |
| 3 | Belgium | 2317052403 |
| 4 | Germany | 2306425333 |
| 5 | Portugal | 2305184016 |

# SALES TRENDS OVER TIME

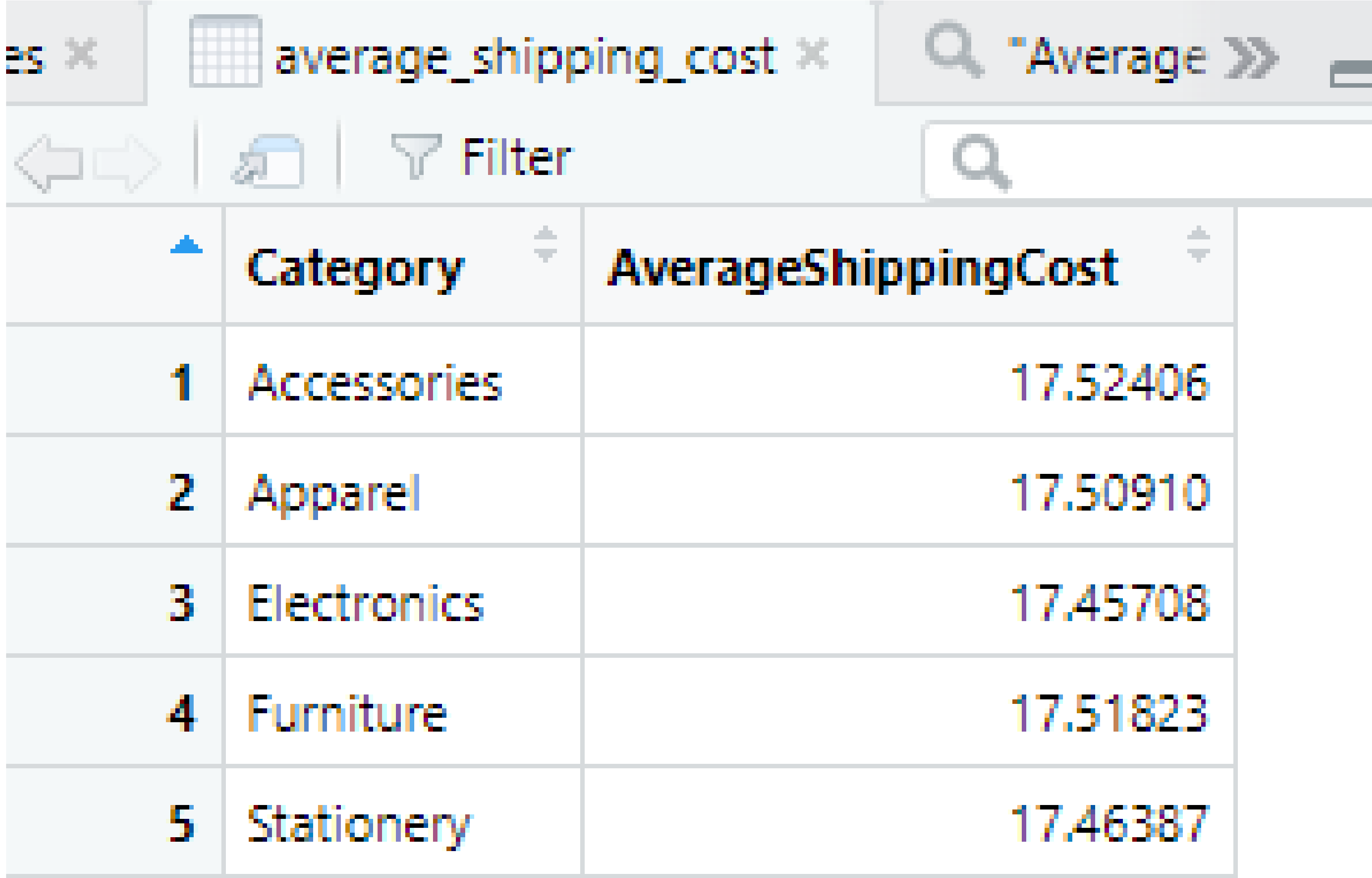
```
sales_trends <- Sales_Data %>%
group_by(InvoiceDate) %>%
summarise(DailySales = sum(Sales))
View(sales_trends)
```



| | InvoiceDate | DailySales |
|---|---|---|
| 1 | 01-01-2020 | 11148427 |
| 2 | 01-01-2021 | 14734509 |
| 3 | 01-01-2022 | 13100345 |
| 4 | 01-01-2023 | 11740875 |
| 5 | 01-01-2024 | 10841146 |
| 6 | 01-01-2025 | 12933091 |
| 7 | 01-02-2020 | 11609871 |
| 8 | 01-02-2021 | 11145803 |
| 9 | 01-02-2022 | 13094878 |
| 10 | 01-02-2023 | 12855673 |
| 11 | 01-02-2024 | 13208560 |
| 12 | 01-02-2025 | 14002336 |
| 13 | 01-03-2020 | 13780702 |
| 14 | 01-03-2021 | 13054213 |

# CALCULATE AVERAGE SHIPPING COST PER CATEGORY

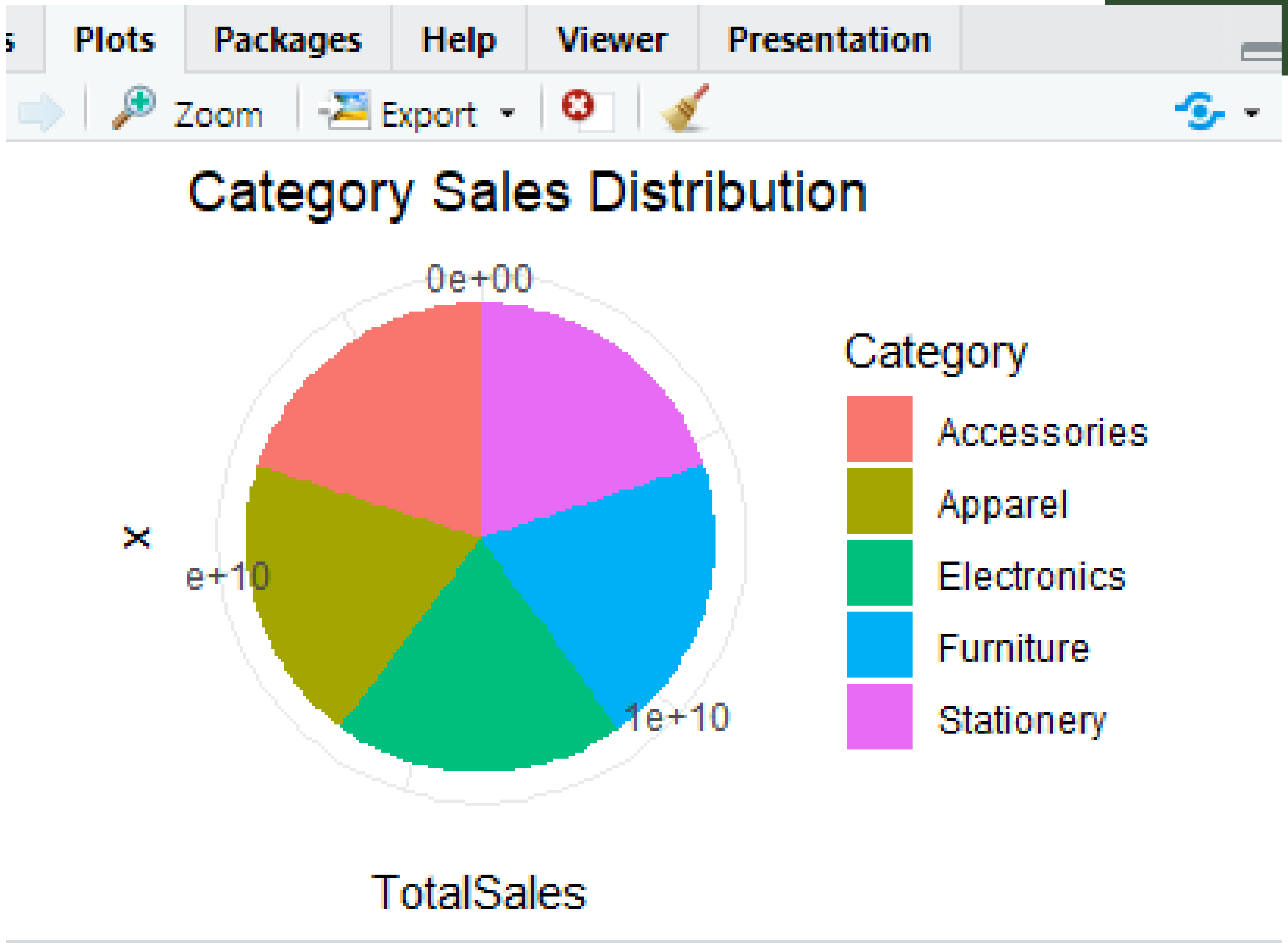
```
average_shipping_cost <- Sales_Data %>%
group_by(Category) %>%
summarise(AverageShippingCost =
mean(ShippingCost, na.rm = TRUE))

View(average_shipping_cost)
```



|  | Category | AverageShippingCost |
|---|---|---|
| 1 | Accessories | 17.52406 |
| 2 | Apparel | 17.50910 |
| 3 | Electronics | 17.45708 |
| 4 | Furniture | 17.51823 |
| 5 | Stationery | 17.46387 |

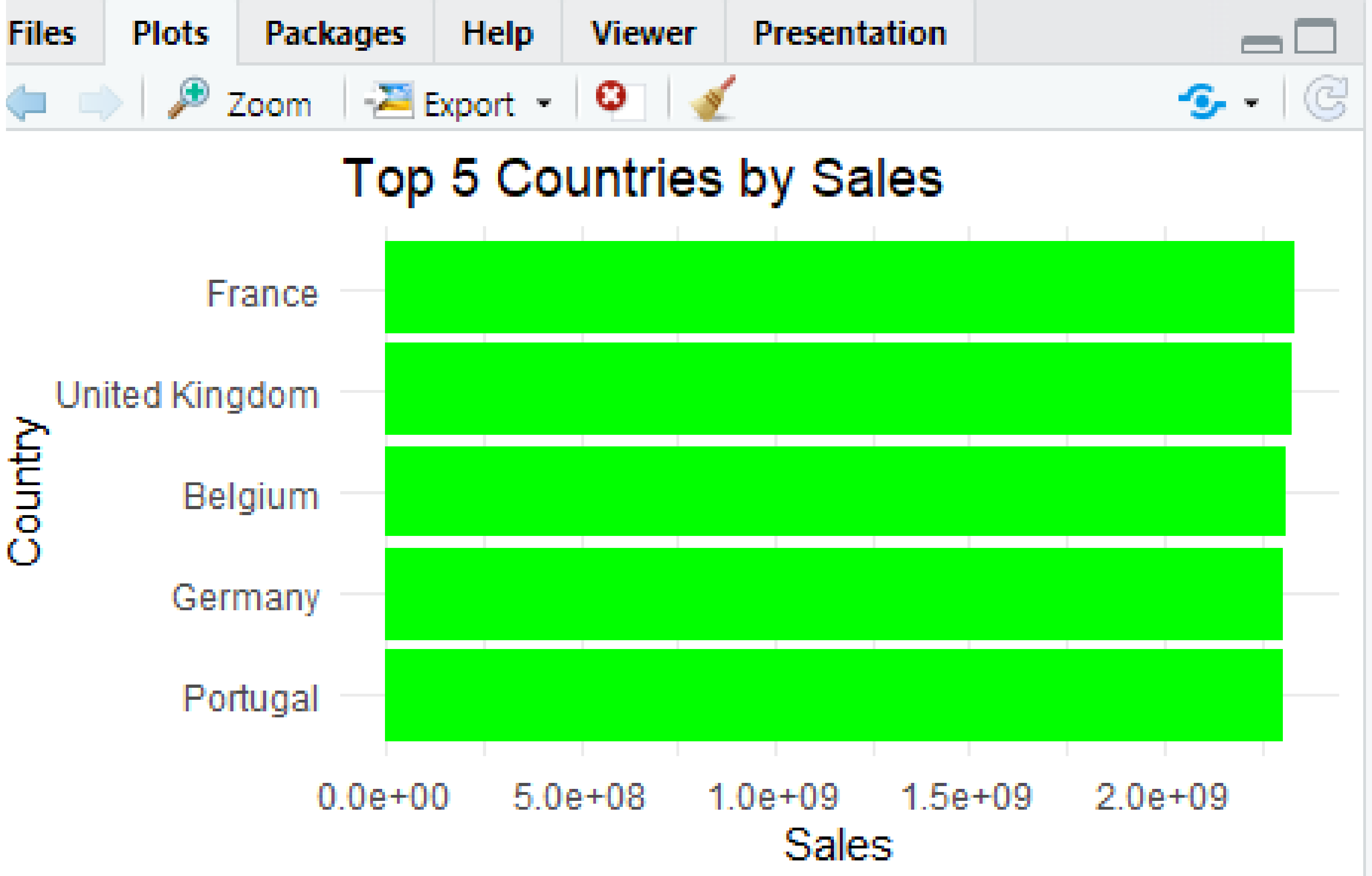
# PIE CHART - CATEGORY SALES CONTRIBUTION

```
category_sales <- Sales_Data %>%
group_by(Category) %>%
summarise(TotalSales = sum(Sales))

ggplot(category_sales, aes(x = "", y = TotalSales,
fill = Category)) +
geom_bar(stat = "identity", width = 1) +
coord_polar("y") +
labs(title = "Category Sales Distribution") +
theme_minimal()
```



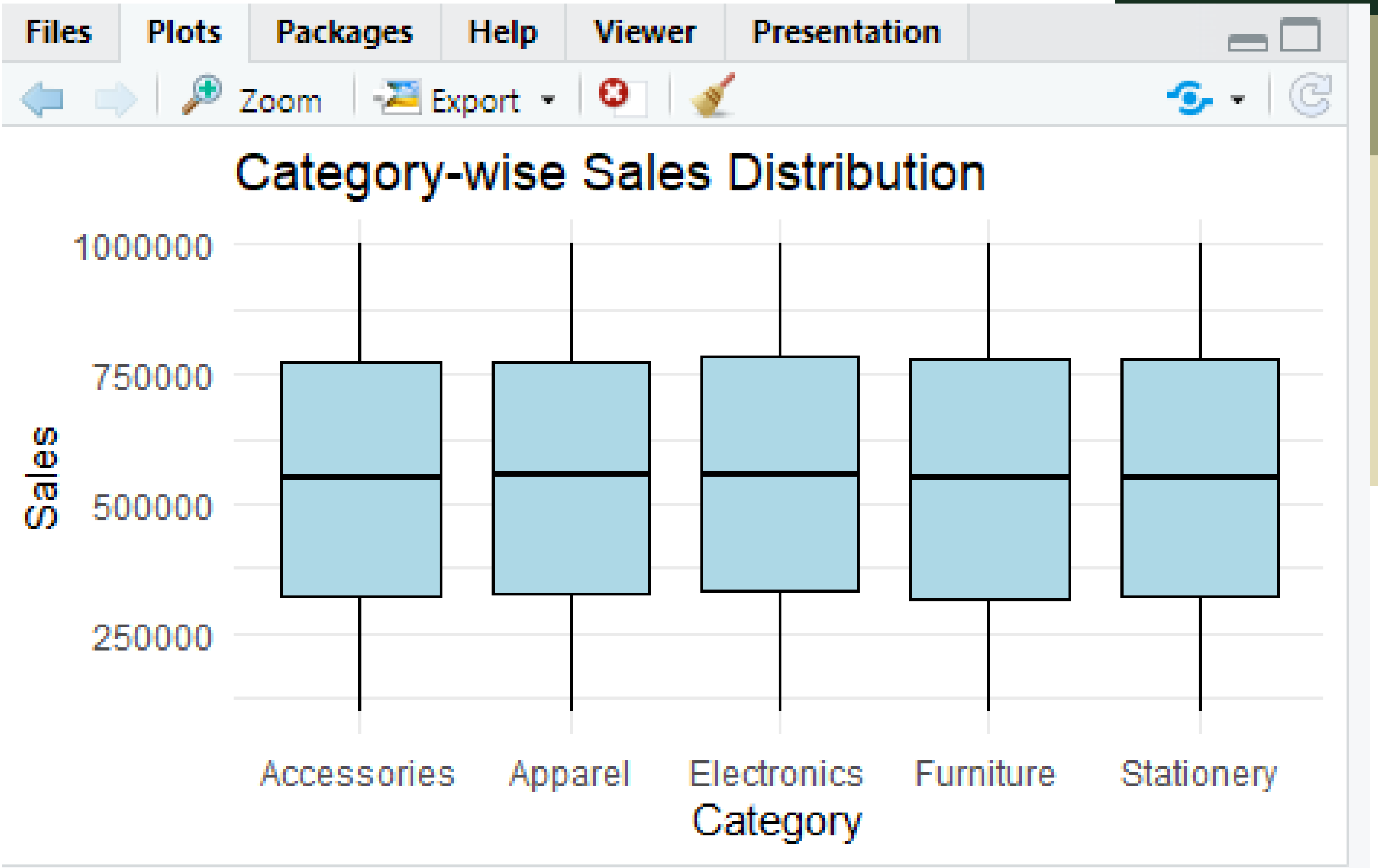
# BAR PLOT FOR TOP COUNTRIES

```
ggplot(top_countries, aes(x = reorder
(Country, TotalSales), y = TotalSales)) +
geom_bar(stat = "identity", fill = "green") +
coord_flip() +
labs(title = "Top 5 Countries by Sales",
x = "Country", y = "Sales") +
theme_minimal()
```



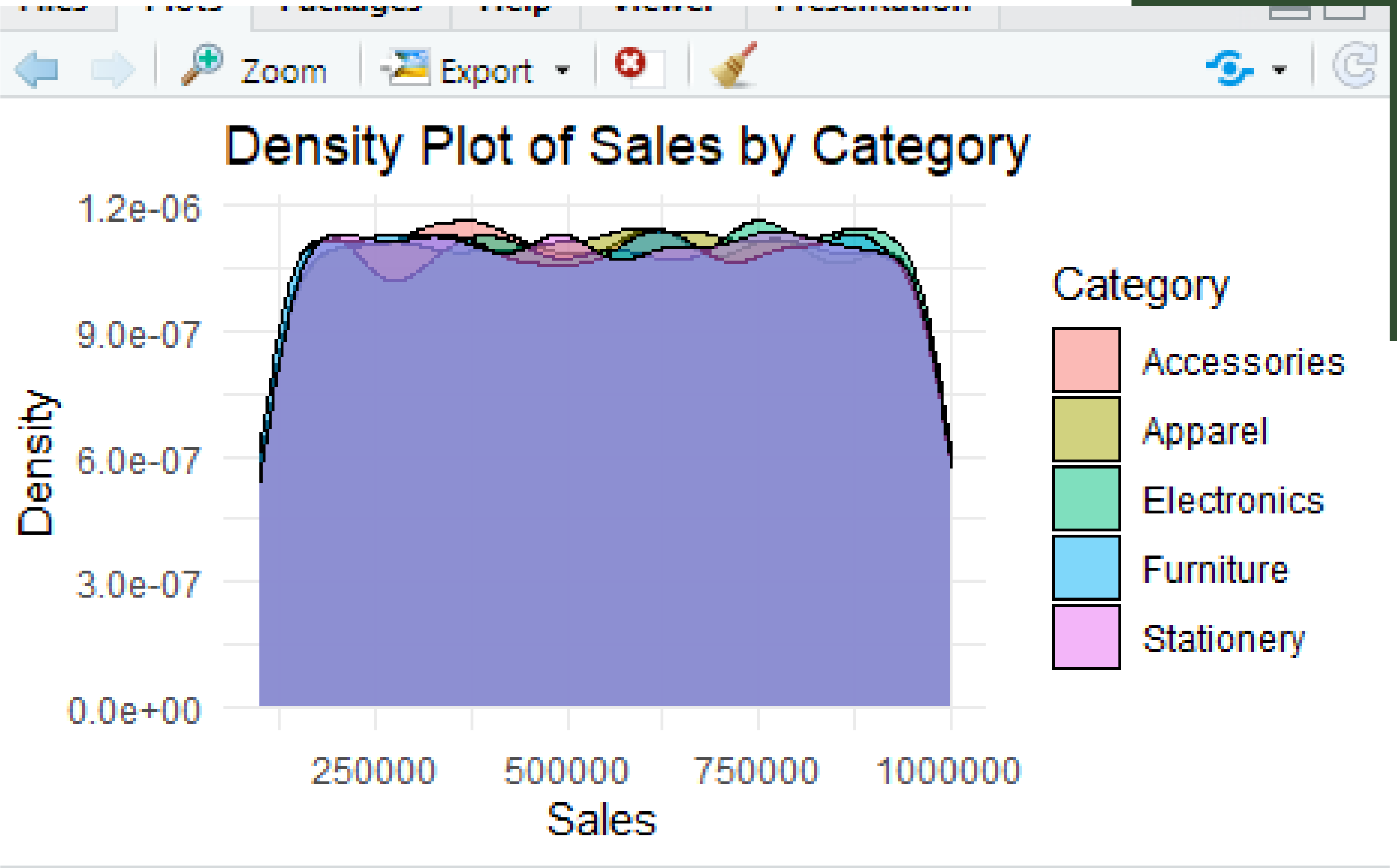
# BOX PLOT CATEGORY-WISE SALES DISTRIBUTION

```
ggplot(Sales_Data, aes(x = Category, y = Sales)) +
geom_boxplot(fill = "lightblue", color = "black") +
labs(title = "Category-wise Sales Distribution",
x = "Category", y = "Sales") +
theme_minimal()
```



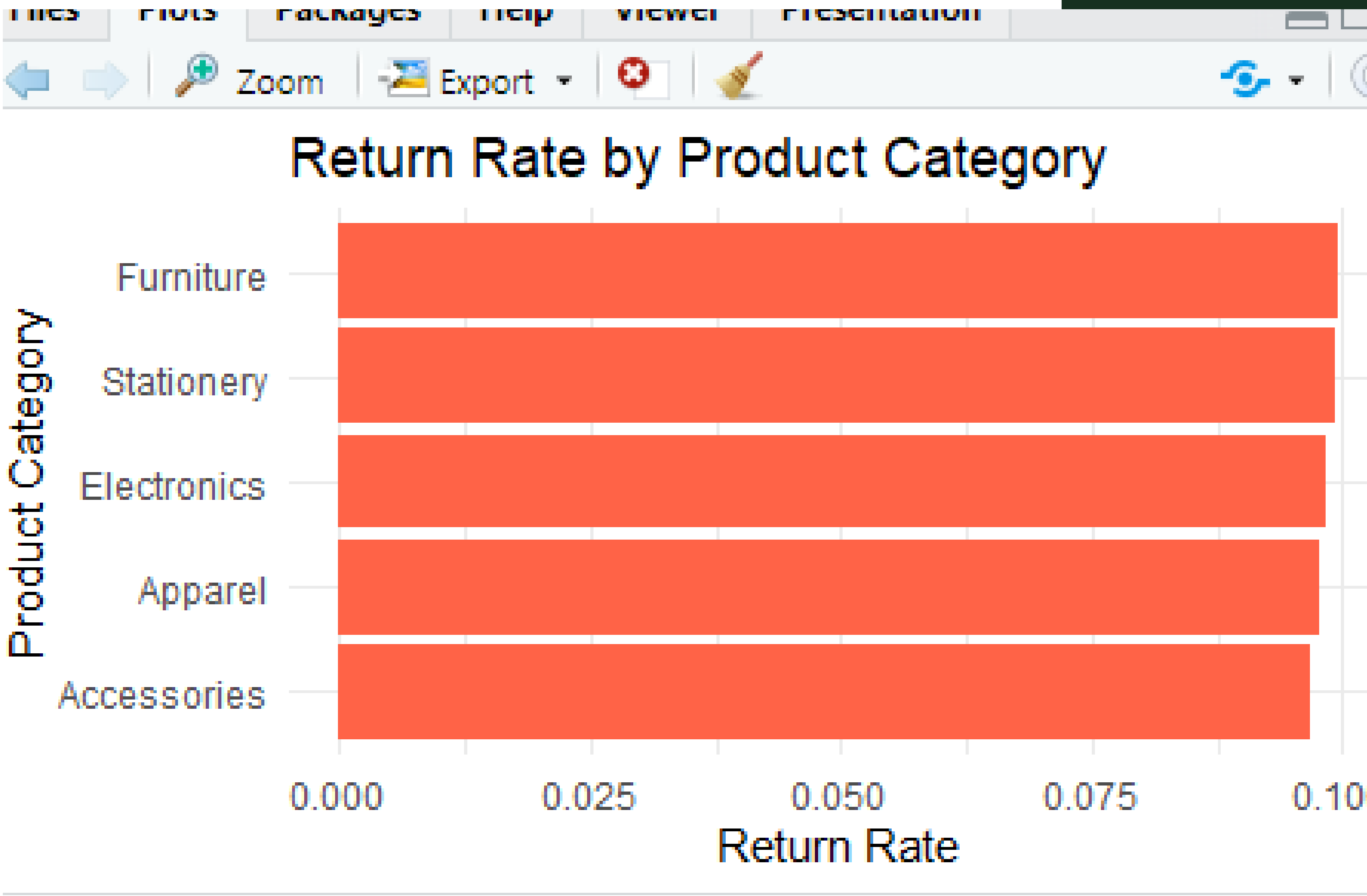
# DENSITY PLOT: SALES DISTRIBUTION

```
ggplot(Sales_Data, aes(x = Sales, fill = Category)) +
geom_density(alpha = 0.5) +
labs(title = "Density Plot of Sales by Category",
x = "Sales", y = "Density") +
theme_minimal()
```



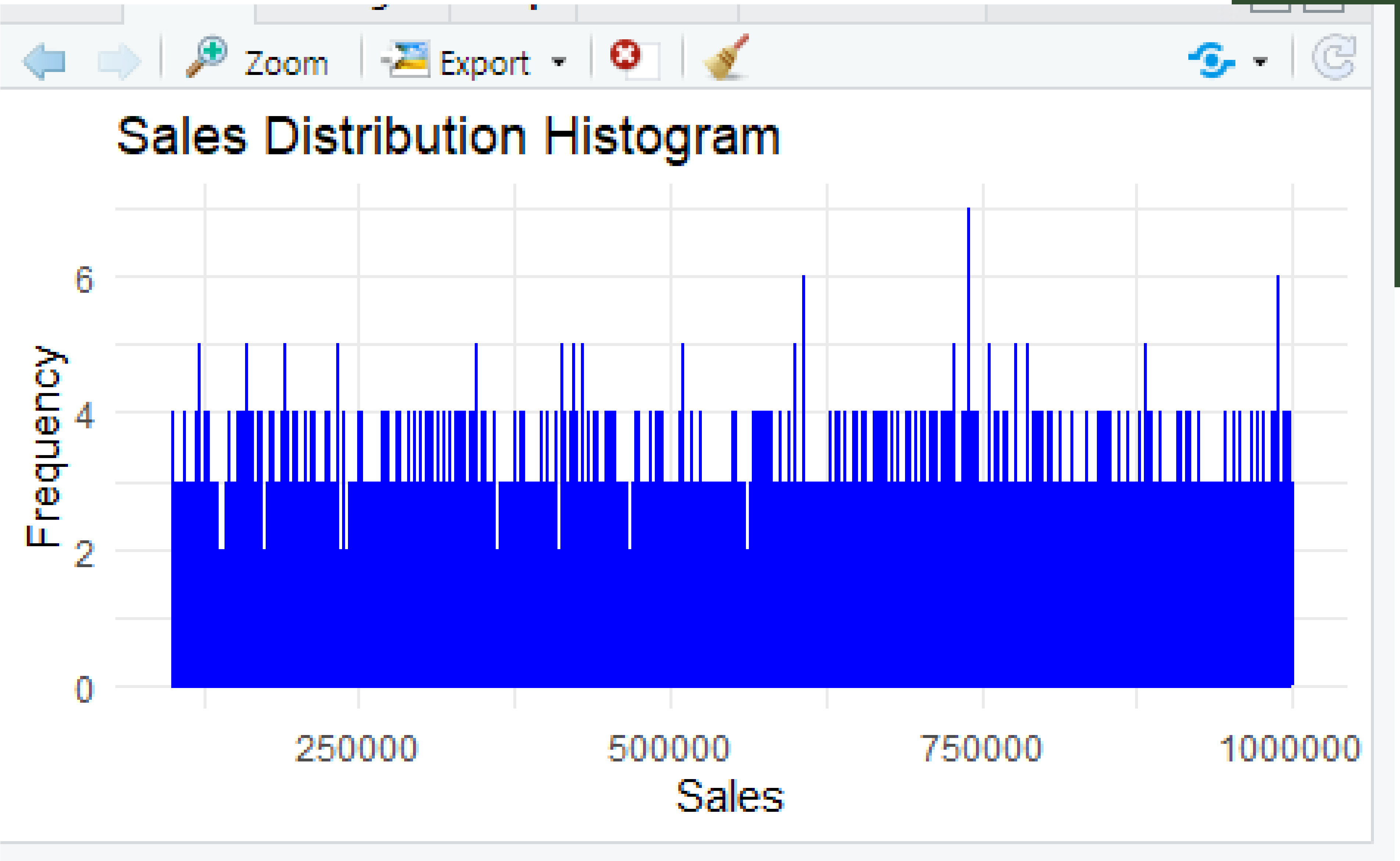
# RETURN RATE BY PRODUCT CATEGORY

```
return_rate_by_category <- Sales_Data %>%
group_by(Category) %>%
summarise(ReturnCount = sum(ReturnStatus ==
"Returned"),
TotalCount = n(),  ReturnRate = ReturnCount / TotalCount)
%>%   arrange(desc(ReturnRate))

ggplot(return_rate_by_category, aes(x = reorder(Category,
ReturnRate), y = ReturnRate)) +
geom_bar(stat = "identity", fill = "tomato") +
coord_flip() +
labs(title = "Return Rate by Product Category", x = "Product
Category", y = "Return Rate") +
theme_minimal()
```



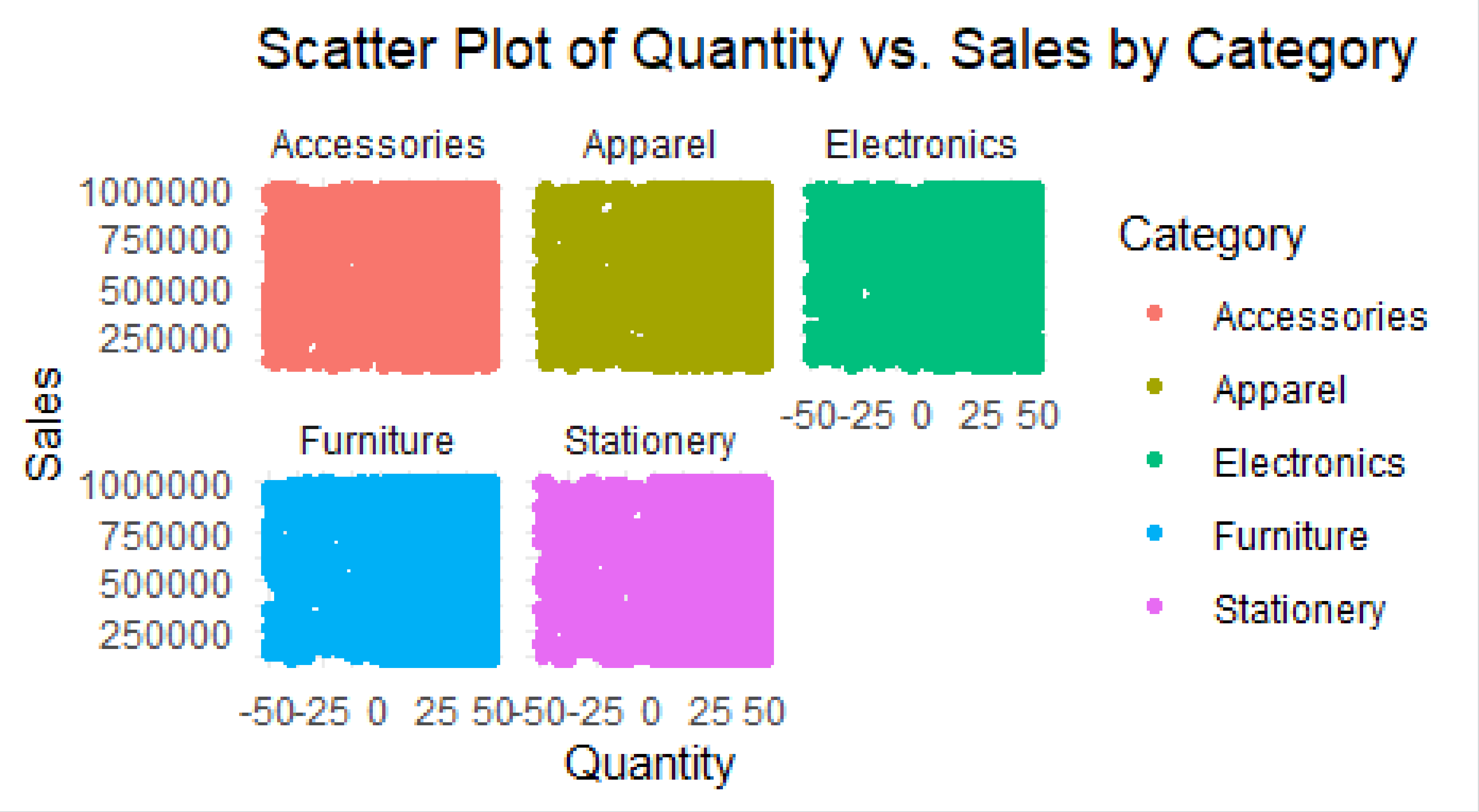
# HISTOGRAM OF SALES DISTRIBUTION

```
ggplot(Sales_Data, aes(x = Sales)) +
geom_histogram(binwidth = 10,
fill = "orange", color = "blue") +
labs(title = "Sales Distribution Histogram",
x = "Sales", y = "Frequency") +
theme_minimal()
```



# FACETED SCATTER PLOT- QUANTITY VS. SALES BY CATEGORY

```
ggplot(Sales_Data, aes(x = Quantity,
y = Sales, color = Category)) +
geom_point() +
facet_wrap(~ Category) +
labs(title = "Scatter Plot of Quantity vs.
Sales by Category",
x = "Quantity", y = "Sales") +
theme_minimal()
```



# CONCLUSION

THE ANALYSIS OF THE ONLINE SALES DATASET HAS PROVIDED VALUABLE INSIGHTS INTO VARIOUS ASPECTS OF SALES PERFORMANCE.

THE STUDY UNDERSCORES THE IMPORTANCE OF DATA-DRIVEN STRATEGIES IN OPTIMIZING SALES PERFORMANCE AND ENHANCING CUSTOMER EXPERIENCES. FUTURE RECOMMENDATIONS INCLUDE EXPLORING PREDICTIVE ANALYTICS FOR FORECASTING TRENDS AND REFINING DISCOUNT STRATEGIES FOR MAXIMIZED PROFITABILITY.

# THANK YOU