PRODIGY\_DS\_02

Anushaa\_2341490

2024-07-28

library(dplyr)

## Warning: package 'dplyr' was built under R version 4.3.3

##   
## 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 4.3.3

library(readxl)

## Warning: package 'readxl' was built under R version 4.3.3

library(corrplot)

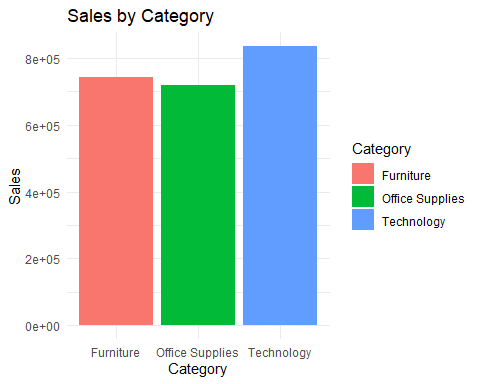
## Warning: package 'corrplot' was built under R version 4.3.3

## corrplot 0.92 loaded

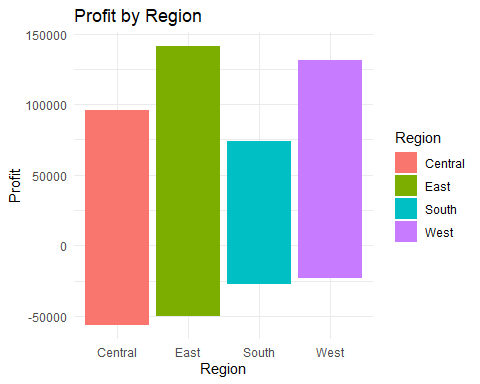
# Load the cleaned Sample Superstore dataset  
library(readxl)  
superstore <- read\_excel("C:/Users/anush/Downloads/Sample - Superstore\_PRODIGY.xls")  
View(superstore)

# Visualizations

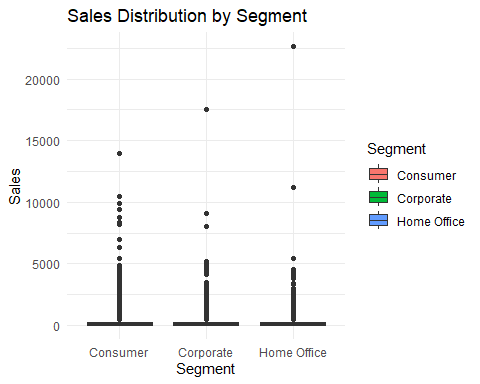
# Sales by Category  
ggplot(superstore, aes(x = Category, y = Sales, fill = Category)) +  
 geom\_bar(stat = "identity") +  
 labs(title = "Sales by Category", x = "Category", y = "Sales") +  
 theme\_minimal()



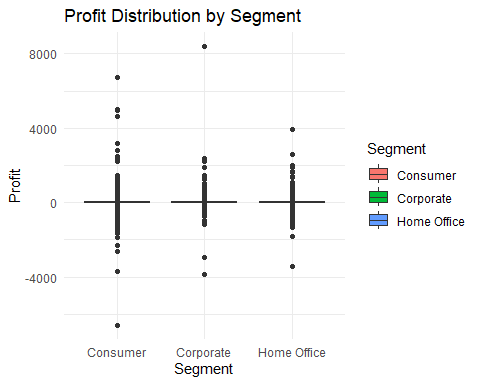
# Profit by Region  
ggplot(superstore, aes(x = Region, y = Profit, fill = Region)) +  
 geom\_bar(stat = "identity") +  
 labs(title = "Profit by Region", x = "Region", y = "Profit") +  
 theme\_minimal()



# Sales and Profit Distribution by Segment  
ggplot(superstore, aes(x = Segment, y = Sales, fill = Segment)) +  
 geom\_boxplot() +  
 labs(title = "Sales Distribution by Segment", x = "Segment", y = "Sales") +  
 theme\_minimal()



ggplot(superstore, aes(x = Segment, y = Profit, fill = Segment)) +  
 geom\_boxplot() +  
 labs(title = "Profit Distribution by Segment", x = "Segment", y = "Profit") +  
 theme\_minimal()



# Correlation Matrix  
numeric\_cols <- superstore %>% select(Sales, Profit, Quantity, Discount)  
corr\_matrix <- cor(numeric\_cols, use = "complete.obs")  
corrplot(corr\_matrix, method = "circle", type = "upper", tl.col = "black", tl.srt = 45, col = colorRampPalette(c("red", "white", "blue"))(200))

