EXPT NO:2	Implementation of data visualization techniques
DATE: 06.01.2026	

PRE-LAB QUESTIONS (PROVIDE BRIEF ANSWERS TO THE FOLLOWING QUESTIONS)

1. **Why is exploratory data analysis critical before model building?**

   EDA is essential because it allows you to understand the data's underlying patterns, identify anomalies or errors, and test assumptions before applying machine learning models. This ensures that the model is built on high-quality, relevant data.

2. **How do distributions influence algorithm selection in ML?**

   Different machine learning algorithms make different assumptions about data distribution. For instance, many linear models assume a normal (Gaussian) distribution; if data is highly skewed , you may need to apply data transformations or choose non-parametric algorithms that do not rely on specific distributions.

3. **What insights can outliers provide in business data?**

   In business, outliers often represent significant events such as high-value "VIP" transactions, fraudulent activities, or system errors. Identifying them through visualizations like boxplots helps a business decide whether to target these high-value segments or investigate potential security risks.

4. **Why are visual summaries preferred over raw tables?**

   The human brain processes images much faster than text or numbers. Visual summaries like histograms and heatmaps reveal trends, correlations, and clusters that are nearly impossible to detect just by looking at thousands of rows in a raw spreadsheet.

5. **How does visualization improve business intelligence?**

   Visualization transforms complex data into intuitive, actionable insights. It allows decision-makers to quickly identify peak sales periods, regional performance gaps, and customer spending habits, leading to faster and more accurate evidence-based business strategies.

IN-LAB EXERCISE:
OBJECTIVE:

To explore data distribution and variability using advanced visualization techniques.

SCENARIO:
A startup analyzes e-commerce transaction data to understand customer spending behavior and detect abnormal purchase patterns.

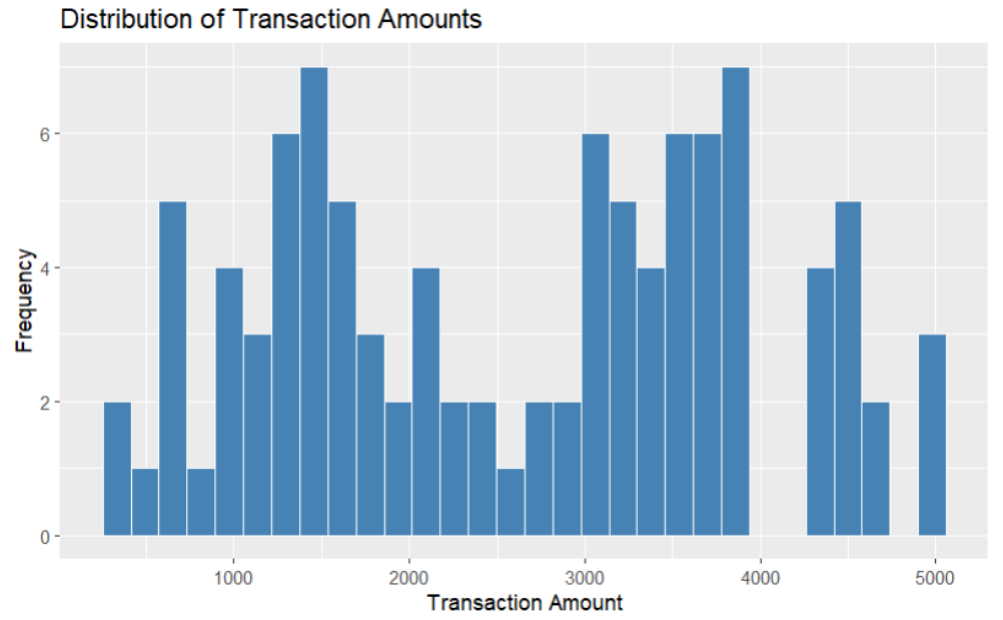
IN-LAB TASKS (Using R Language)

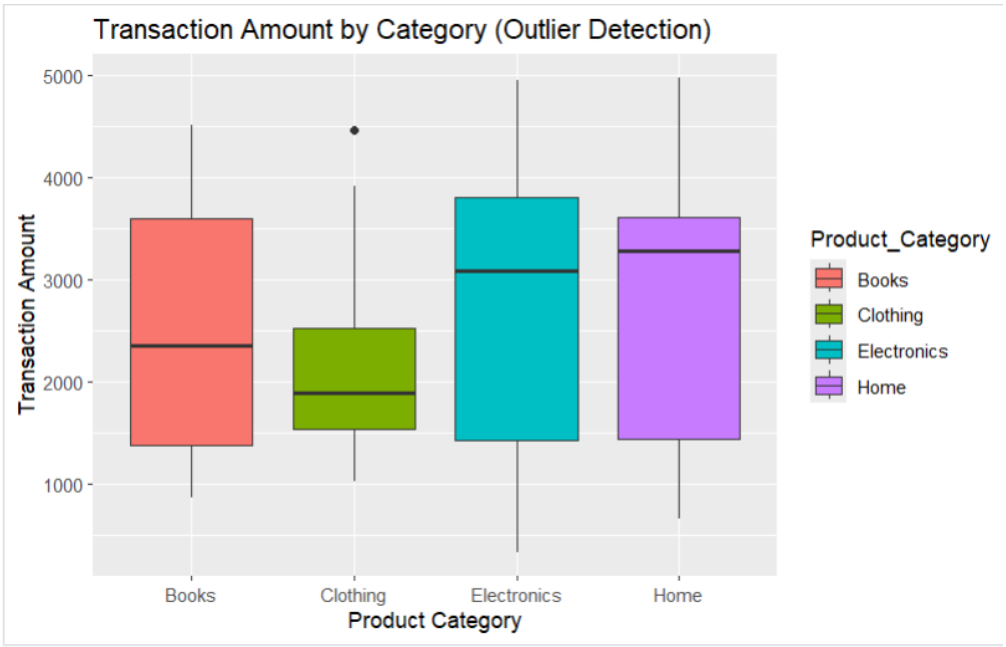
- Plot histogram of transaction amounts
- Use boxplot to detect outliers
- Create heatmap of monthly sales intensity

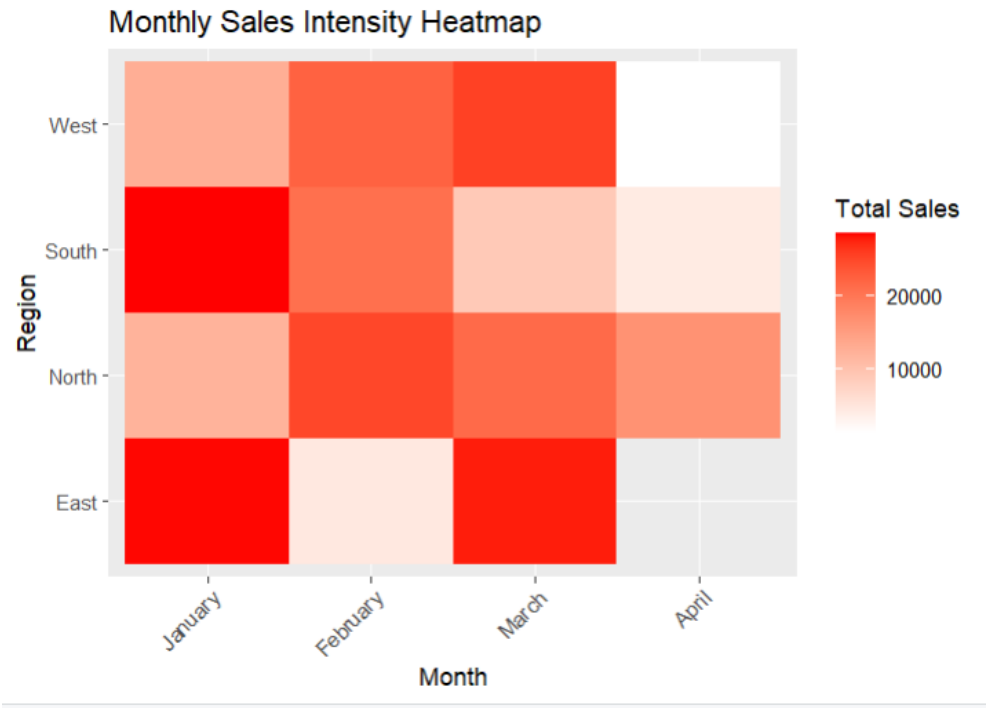
CODE:

```r
#.................
#Roll No : 23BAD076
#.................

library(ggplot2)
library(dplyr)
ecommerce_data <- read.csv("2.ecommerce_transactions.csv")
ecommerce_data$Transaction_Date <- as.Date(ecommerce_data$Transaction_Date)
ecommerce_data$Month <- format(ecommerce_data$Transaction_Date, "%B")

# --- histogram of transaction amounts ---
ggplot(ecommerce_data, aes(x = Transaction_Amount)) +
  geom_histogram(fill = "steelblue", color = "white", bins = 30) +
  labs(title = "Distribution of Transaction Amounts",
       x = "Transaction Amount",
       y = "Frequency")

# --- boxplot to detect outliers ---
ggplot(ecommerce_data, aes(x = Product_Category, y = Transaction_Amount, fill = Product_Category)) +
  geom_boxplot() +
  labs(title = "Transaction Amount by Category (Outlier Detection)",
       x = "Product Category",
       y = "Transaction Amount")

# --- heatmap of monthly sales intensity ---
heatmap_summary <- ecommerce_data %>%
  group_by(Month, Region) %>%
  summarise(Total_Sales = sum(Transaction_Amount, na.rm = TRUE))
ggplot(heatmap_summary, aes(x = Month, y = Region, fill = Total_Sales)) +
  geom_tile() +
  scale_fill_gradient(low = "white", high = "red") +
  labs(title = "Monthly Sales Intensity Heatmap",
       x = "Month",
       y = "Region",
       fill = "Total Sales") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

OUTPUT:


Distribution of Transaction Amounts
Frequency
0
2
4
6
1000
2000
3000
4000
5000
Transaction Amount


Transaction Amount by Category (Outlier Detection)
Transaction Amount
1000
2000
3000
4000
5000
Books
Clothing
Electronics
Home
Product Category
Product_Category
Books
Clothing
Electronics
Home



POST-LAB QUESTIONS (PROVIDE BRIEF ANSWERS TO THE FOLLOWING QUESTIONS)

1. **What does right-skewed distribution indicate about customer behavior?**
   In an e-commerce context, a right-skewed distribution indicates that the majority of customers make frequent, low-value purchases. Only a small number of customers are responsible for very high-value transactions.

2. **How can detected outliers impact business decisions?**
   Outliers can identify high-spending "VIP" customers for targeted loyalty programs or reveal fraudulent transactions that require better security measures.

3. **Which visualization best supports anomaly detection?**
   The Boxplot is best because it clearly marks "dots" that are far away from normal spending.

4. **How does EDA improve AI model accuracy?**
   EDA improves accuracy by allowing you to clean the data—removing noise, handling missing values, and correcting errors—before the model "learns" from it.

5. **How can visualization guide feature engineering?**
   It shows which factors (like region or month) actually matter, helping you pick the right data for the model.

ASSESSMENT

Description	Max Marks	Marks Awarded
Pre Lab Exercise	**5**	
In Lab Exercise	**10**	
Post Lab Exercise	**5**	
Viva	**10**	
Total	**30**	
	Faculty Signature	