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?**
   It helps understand data structure, detect errors, missing values, patterns, and relationships before choosing or training a model.
2. **How do distributions influence algorithm selection in ML?**
   Some algorithms assume specific distributions (e.g., normality), while skewed or non-linear distributions may require robust or non-parametric models.
3. **What insights can outliers provide in business data?**
   Outliers can indicate anomalies, fraud, rare events, data errors, or high-value opportunities needing special attention.
4. **Why are visual summaries preferred over raw tables?**
   Visuals quickly reveal trends, patterns, and anomalies that are hard to identify from large numerical tables.
5. **How does visualization improve business intelligence?**
   It enables faster decision-making by clearly communicating insights, trends, and performance metrics to stakeholders.

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

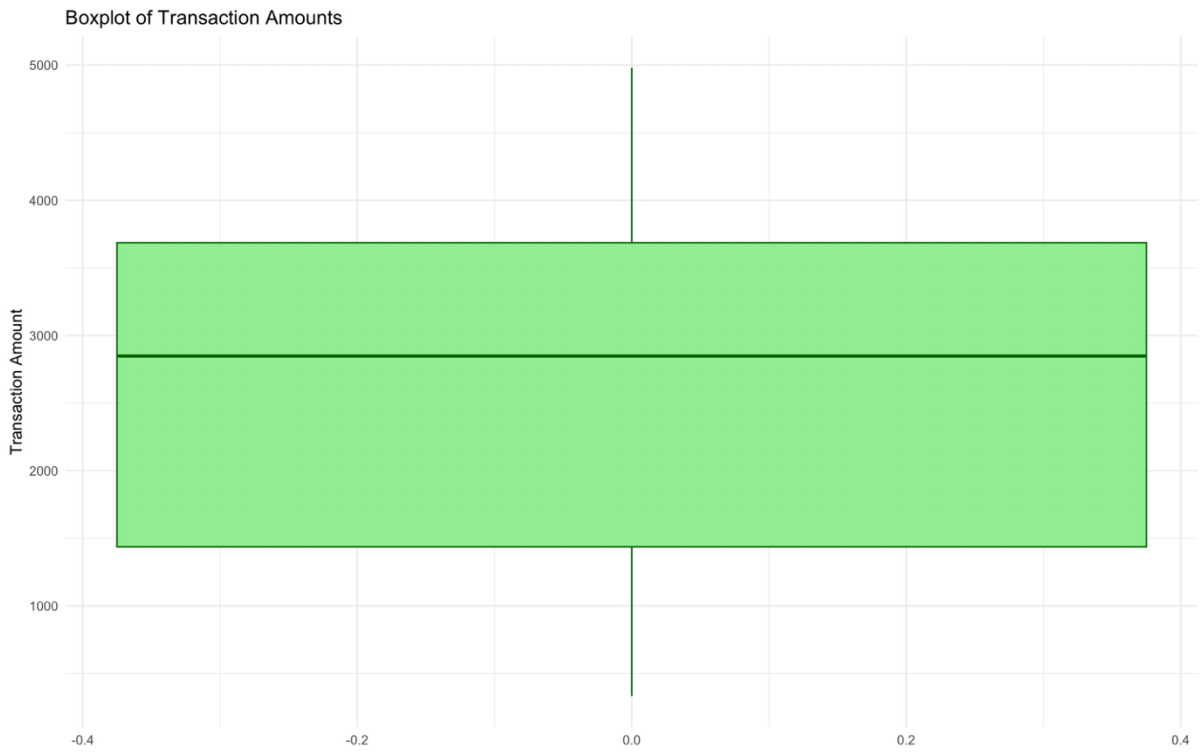
Code:

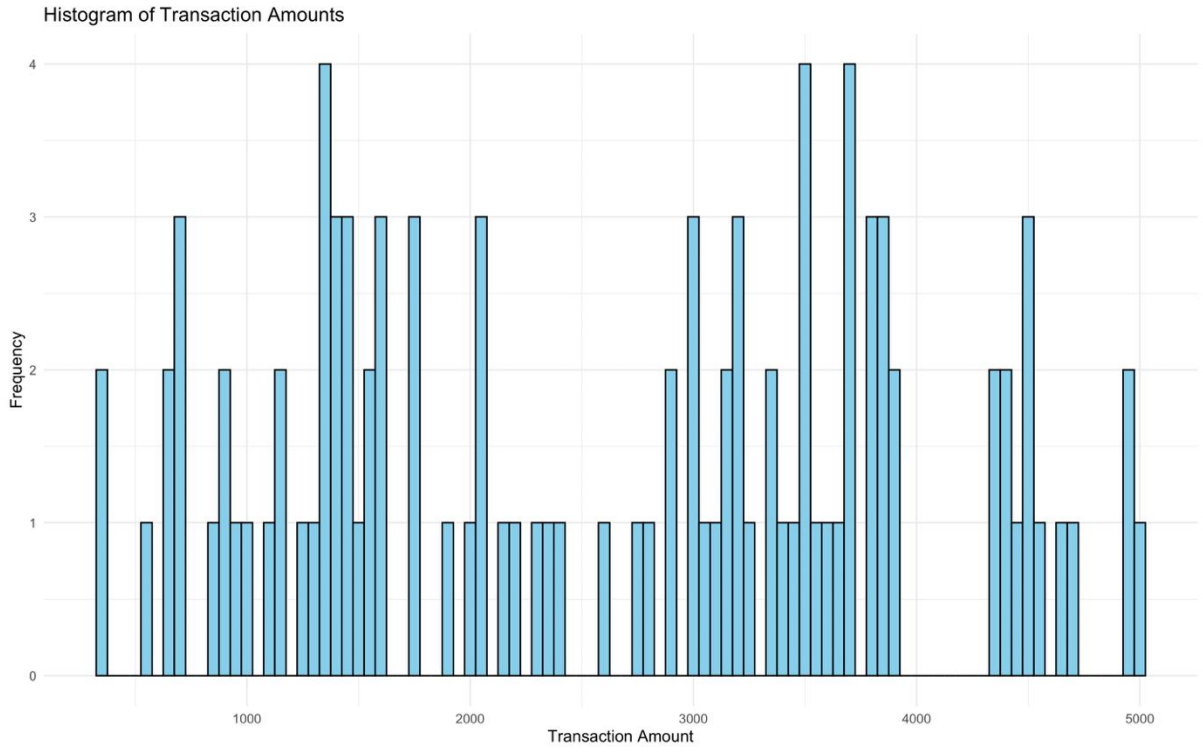
Ex-2 Implementation of data visualization techniques

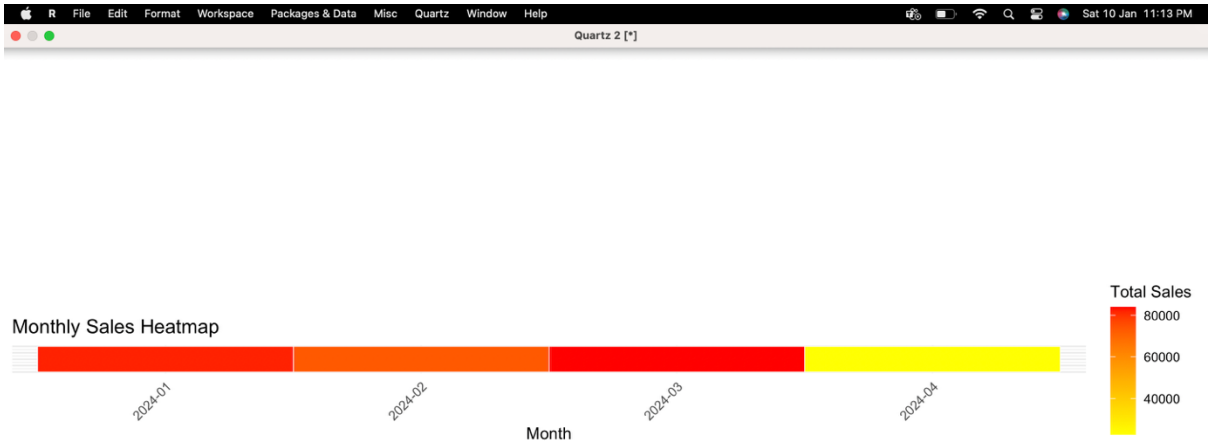
<functions> Help search

```r
# Load required libraries
library(ggplot2)
library(dplyr)
# Load your CSV (correct filename)
transactions_df <- read.csv("/Users/sugitha/Downloads/2.ecommerce_transactions.csv")

head(transactions_df)

# 1️⃣ Histogram of Transaction Amounts
ggplot(transactions_df, aes(x = Transaction_Amount)) +
  geom_histogram(binwidth = 50, fill = "skyblue", color = "black") +
  labs(title = "Histogram of Transaction Amounts",
       x = "Transaction Amount",
       y = "Frequency") +
  theme_minimal()

# 2️⃣ Boxplot to detect outliers
ggplot(transactions_df, aes(y = Transaction_Amount)) +
  geom_boxplot(fill = "lightgreen", color = "darkgreen") +
  labs(title = "Boxplot of Transaction Amounts",
       y = "Transaction Amount") +
  theme_minimal()

# Extract and print outliers
outliers <- boxplot.stats(transactions_df$Transaction_Amount)$out
cat("Outliers in Transaction Amounts:\n")
print(outliers)

# 3️⃣ Monthly Sales Heatmap
# Convert Transaction_Date to Date type
transactions_df$Transaction_Date <- as.Date(transactions_df$Transaction_Date, format="%Y-%m-%d")

# Extract month-year
transactions_df$Month <- format(transactions_df$Transaction_Date, "%Y-%m")

# Aggregate total sales by month
monthly_sales <- transactions_df %>%
  group_by(Month) %>%
  summarise(Total_Sales = sum(Transaction_Amount), .groups = "drop")

# Heatmap
ggplot(monthly_sales, aes(x = Month, y = 1, fill = Total_Sales)) +
  geom_tile(color = "white") +
  scale_fill_gradient(low = "yellow", high = "red") +
  labs(title = "Monthly Sales Heatmap",
       x = "Month",
       y = "",
       fill = "Total Sales") +
  theme_minimal() +
  theme(axis.text.y = element_blank(),
        axis.ticks.y = element_blank(),
        axis.title.y = element_blank(),
        axis.text.x = element_text(angle = 45, hjust = 1)) +
  coord_fixed(ratio = 0.1)

```

Output:


Boxplot of Transaction Amounts
Transaction Amount
5000
4000
3000
2000
1000
-0.4
-0.2
0.0
0.2
0.4


Histogram of Transaction Amounts
Frequency
4
3
2
1
0
1000
2000
3000
4000
5000
Transaction Amount



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

1. What does a right-skewed distribution indicate about customer behavior?

A right-skewed (positively skewed) distribution of transaction amounts means most customers spend smaller amounts, but a few customers make very large purchases. This suggests that high spenders are rare but can significantly impact revenue.

2. How can detected outliers impact business decisions?

Outliers can indicate fraudulent transactions, data entry errors, or unusual high-value purchases. Identifying them helps businesses:

- Investigate anomalies or potential fraud
- Adjust marketing strategies for high-value customers
- Avoid skewed analytics that could misguide decisions

3. Which visualization best supports anomaly detection?

Boxplots are particularly effective for anomaly detection because they clearly highlight data points outside the normal range (outliers). Histograms can show general distribution, but boxplots pinpoint unusual values.

4. How does EDA improve AI model accuracy?

EDA helps identify:

- Outliers, missing values, and errors

- Feature distributions and correlations
- Skewed or imbalanced data

By cleaning and understanding the data first, AI models learn from high-quality inputs, reducing bias and improving predictive accuracy.

5. How can visualization guide feature engineering?

Visualizations reveal patterns, trends, and relationships that may not be obvious from raw tables. For example:

- Scatterplots can suggest interaction features
- Histograms or density plots can guide transformations (e.g., log transform for skewed features)
- Heatmaps reveal correlations, suggesting features to combine or drop

ASSESSMENT

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