**Introduction**

This project aims to analyze and visualize company financial performance using statistical methods and data visualization techniques. By understanding the distribution of revenue, profit, and expenses, and identifying trends and outliers, businesses can make informed decisions to optimize operations and drive growth.

**Abstract**

The project leverages R programming for statistical analysis and graphical visualization of company performance metrics, including revenue, profit, and expenses. Through calculating key statistics (mean, median, standard deviation) and creating visualizations (bar plots, histograms, box plots), the analysis offers insights into financial trends and identifies top and bottom-performing companies. The outcome aids in strategic decision-making by identifying areas of improvement and optimization.

**Methodology**

1. **Data Preparation**:
   * Create a structured dataset containing company names, revenue, profit, and expense figures.
2. **Statistical Analysis**:
   * Apply statistical functions to compute mean, median, and standard deviation for each financial metric.
3. **Visualization**:
   * Develop bar plots to represent average metrics across companies, histograms for distribution analysis, and box plots for data dispersion.
4. **Performance Insights**:
   * Identify top and bottom performers using comparison techniques.

**Technology Used**

1. **Programming Language**: R - for statistical computations and graphical visualization.
2. **Libraries**:
   * base R functions for statistical analysis.
   * R plotting functions like barplot(), hist(), and boxplot() for visualizations.
3. **Tools**:
   * Integrated Development Environment (IDE) like RStudio for writing and executing code.

**Code:**

company\_data <- data.frame(

Company = c("Company A", "Company B", "Company C", "Company D", "Company E"),

Revenue = c(50000, 60000, 45000, 70000, 55000),

Profit = c(8000, 10000, 7000, 12000, 9000),

Expenses = c(42000, 50000, 38000, 58000, 46000)

)

# View the company data in a table format with borders (RStudio viewer)

View(company\_data) # <--- This opens a nice tab with the data in RStudio

# Function to Calculate Statistics

stats <- function(data) {

list(

Mean = mean(data),

Median = median(data),

SD = sd(data)

)

}

# Calculate Statistics for All Columns

revenue\_stats <- stats(company\_data$Revenue)

profit\_stats <- stats(company\_data$Profit)

expenses\_stats <- stats(company\_data$Expenses)

# Identify Top and Bottom Performers

highest\_company <- function(data) {

company\_data$Company[which.max(data)]

}

lowest\_company <- function(data) {

company\_data$Company[which.min(data)]

}

top\_revenue <- highest\_company(company\_data$Revenue)

least\_revenue <- lowest\_company(company\_data$Revenue)

# Visualizations

barplot(colMeans(company\_data[,-1]), main = "Average Metrics Across Companies", col = c("blue", "green", "red"))

hist(company\_data$Revenue, main = "Revenue Distribution", col = "blue")

boxplot(company\_data[,-1], main = "Financial Data Distribution", col = c("blue", "green", "red"))

# Output Results

cat("Revenue Statistics:\n")

cat("Mean:", revenue\_stats$Mean, "\n")

cat("Median:", revenue\_stats$Median, "\n")

cat("Standard Deviation:", revenue\_stats$SD, "\n")

cat("\nProfit Statistics:\n")

cat("Mean:", profit\_stats$Mean, "\n")

cat("Median:", profit\_stats$Median, "\n")

cat("Standard Deviation:", profit\_stats$SD, "\n")

cat("\nExpenses Statistics:\n")

cat("Mean:", expenses\_stats$Mean, "\n")

cat("Median:", expenses\_stats$Median, "\n")

cat("Standard Deviation:", expenses\_stats$SD, "\n")

cat("\nTop Revenue Company:", top\_revenue, "\n")

cat("Lowest Revenue Company:", least\_revenue, "\n")

# Summary Table of Statistics for All Metrics

summary\_table <- data.frame(

Metric = c("Revenue", "Profit", "Expenses"),

Mean = c(revenue\_stats$Mean, profit\_stats$Mean, expenses\_stats$Mean),

Median = c(revenue\_stats$Median, profit\_stats$Median, expenses\_stats$Median),

SD = c(revenue\_stats$SD, profit\_stats$SD, expenses\_stats$SD)

)

# Display the Summary Statistics Table

cat("\nSummary Statistics Table:\n")

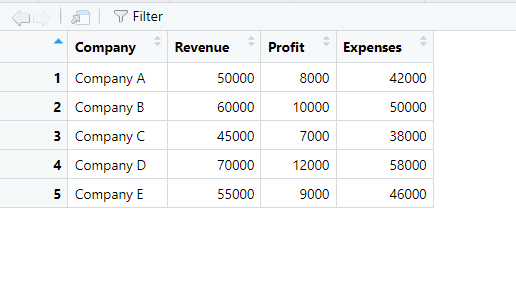
print(summary\_table, row.names = FALSE) # Outputs a structured summary table

# Company-wise Financials Table

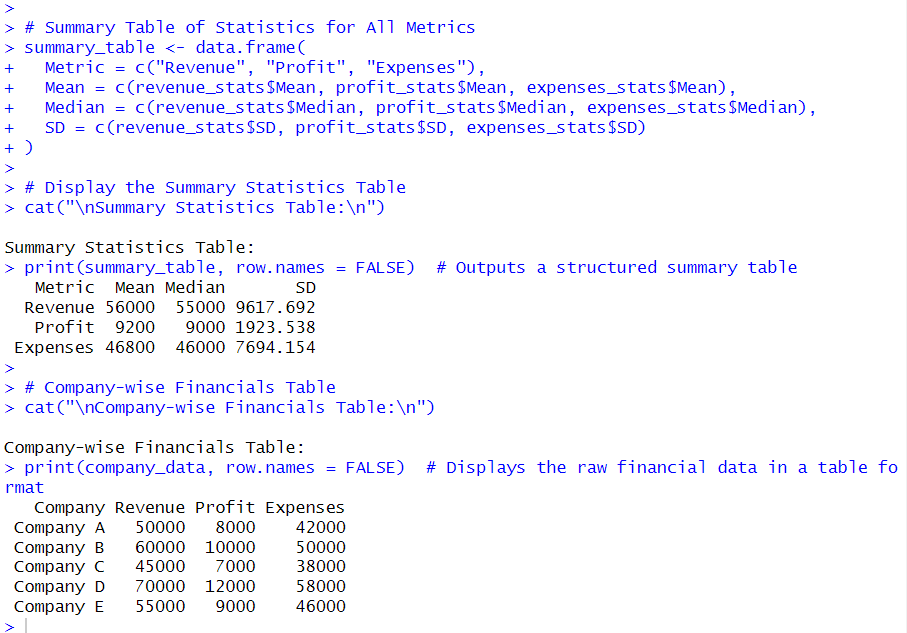
cat("\nCompany-wise Financials Table:\n")

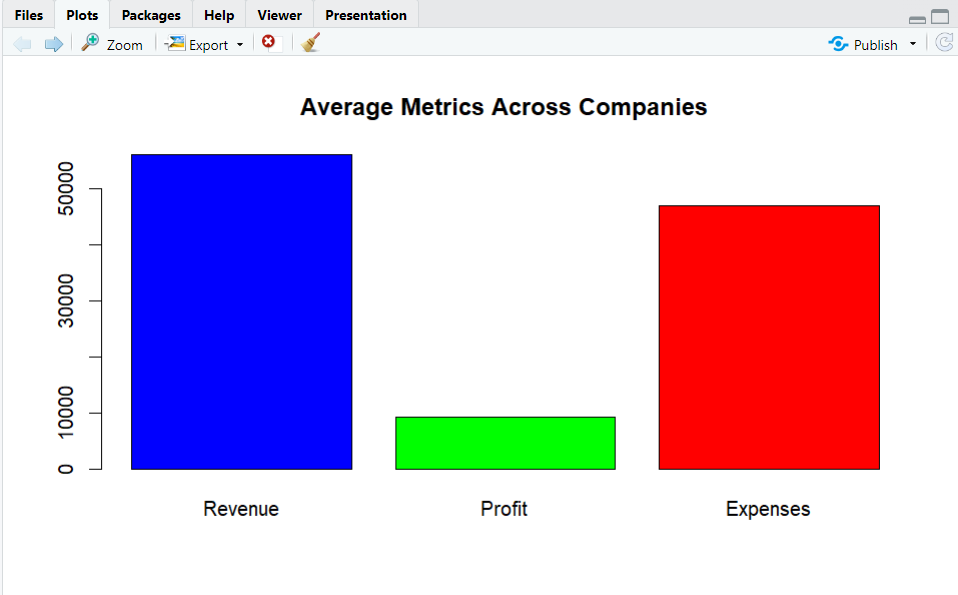
print(company\_data, row.names = FALSE) # Displays the raw financial data in a table format

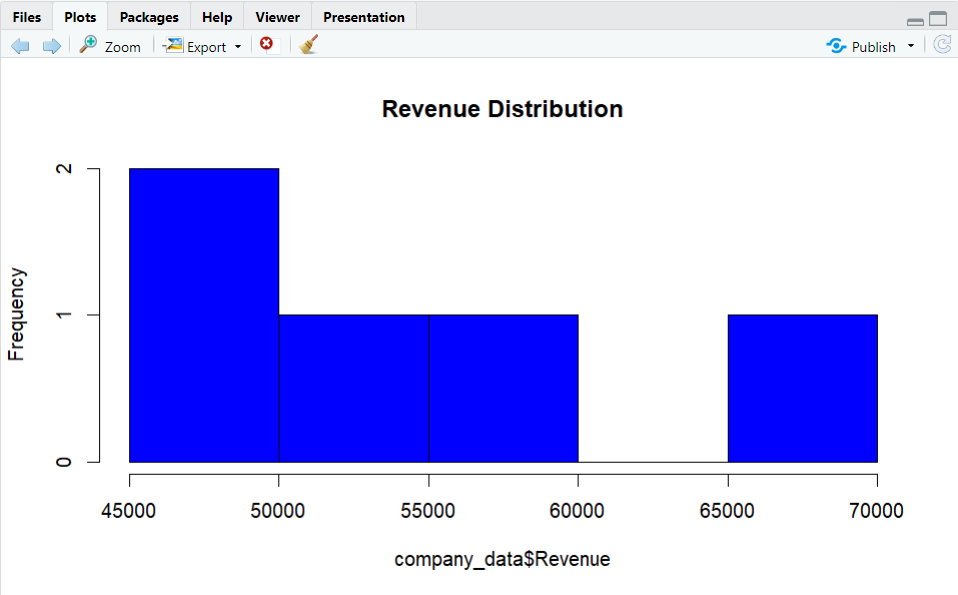
OUTPUT:- # Company Performance Dataset

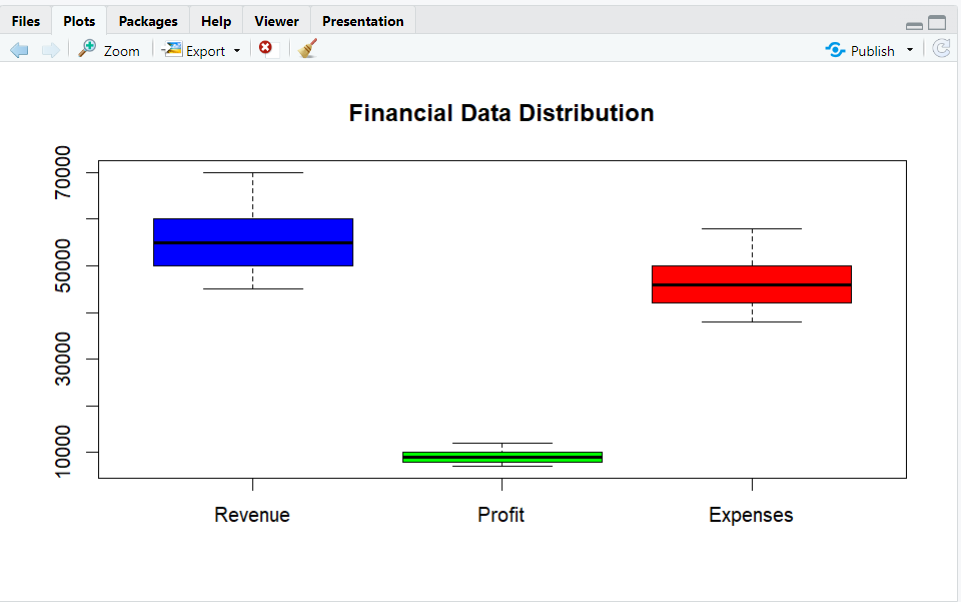












**Explanation of the Code:**

**Data Creation**

R

company\_data <- data.frame(

Company = c("Company A", "Company B", "Company C", "Company D", "Company E"),

Revenue = c(50000, 60000, 45000, 70000, 55000),

Profit = c(8000, 10000, 7000, 12000, 9000),

Expenses = c(42000, 50000, 38000, 58000, 46000)

)

* **What it does**: Creates a data frame called company\_data that holds the performance metrics (Revenue, Profit, and Expenses) for five companies.
* **Purpose**: The dataset serves as the input for analysis and visualization.

**View the Data**

R

View(company\_data) # <--- This opens a nice tab with the data in RStudio

* **What it does**: Opens the company\_data in a spreadsheet-like viewer for easier exploration.
* **Purpose**: Useful for verifying the dataset visually (RStudio feature).

**Function to Calculate Statistics**

R

stats <- function(data) {

list(

Mean = mean(data),

Median = median(data),

SD = sd(data)

)

}

* **What it does**: Defines a function stats() that takes a numeric vector and calculates:
  + Mean: The average of the values.
  + Median: The middle value when sorted.
  + SD: The standard deviation, showing data spread.
* **Purpose**: Streamlines statistical calculations for different columns.

**Apply Statistics Function**

R

revenue\_stats <- stats(company\_data$Revenue)

profit\_stats <- stats(company\_data$Profit)

expenses\_stats <- stats(company\_data$Expenses)

* **What it does**:
  + Applies the stats() function to the Revenue, Profit, and Expenses columns in company\_data.
  + Saves the results in separate variables (revenue\_stats, profit\_stats, expenses\_stats).
* **Purpose**: Stores the calculated metrics for later output and reference.

**Identify Top and Bottom Performers**

R

highest\_company <- function(data) {

company\_data$Company[which.max(data)]

}

lowest\_company <- function(data) {

company\_data$Company[which.min(data)]

}

top\_revenue <- highest\_company(company\_data$Revenue)

least\_revenue <- lowest\_company(company\_data$Revenue)

* **What it does**:
  + Defines two functions:
    - highest\_company(): Finds the company with the highest value in the input column.
    - lowest\_company(): Finds the company with the lowest value in the input column.
  + Uses these functions to determine:
    - top\_revenue: The company with the highest revenue.
    - least\_revenue: The company with the lowest revenue.
* **Purpose**: Identifies companies that are performing the best and worst in terms of revenue.

**Data Visualizations**

1. **Bar Plot**

R

barplot(colMeans(company\_data[,-1]), main = "Average Metrics Across Companies", col = c("blue", "green", "red"))

* + **What it does**: Creates a bar plot showing average Revenue, Profit, and Expenses.
  + **Details**:
    - company\_data[,-1]: Excludes the first column (Company names).
    - colMeans(): Calculates the mean for each metric.
    - col: Specifies bar colors for clarity.
  + **Purpose**: Summarizes company performance in a single chart.

1. **Histogram**

R

hist(company\_data$Revenue, main = "Revenue Distribution", col = "blue")

* + **What it does**: Creates a histogram of the Revenue column.
  + **Purpose**: Shows the distribution of revenue across companies.

1. **Box Plot**

R

boxplot(company\_data[,-1], main = "Financial Data Distribution", col = c("blue", "green", "red"))

* + **What it does**: Creates a box plot for Revenue, Profit, and Expenses.
  + **Details**:
    - Highlights data spread, medians, and outliers for each metric.
  + **Purpose**: Visualizes variations within each financial metric.

**Summary statistics .**

**Line 1:**

r

summary\_table <- data.frame(

Metric = c("Revenue", "Profit", "Expenses"),

Mean = c(revenue\_stats$Mean, profit\_stats$Mean, expenses\_stats$Mean),

Median = c(revenue\_stats$Median, profit\_stats$Median, expenses\_stats$Median),

SD = c(revenue\_stats$SD, profit\_stats$SD, expenses\_stats$SD)

)

* **Purpose**: This line creates a **summary table** for the financial metrics (Revenue, Profit, Expenses) and includes statistics like **Mean**, **Median**, and **Standard Deviation (SD)**.
* **Explanation**:
  + data.frame(): This function creates a data frame, which is a structured table in R.
  + Metric: This column contains the names of the metrics — "Revenue", "Profit", and "Expenses".
  + Mean: This column will contain the mean (average) values for Revenue, Profit, and Expenses, which are pulled from the respective statistics (e.g., revenue\_stats$Mean for Revenue).
  + Median: Similarly, this column will contain the median values for each metric (revenue\_stats$Median for Revenue).
  + SD: This column will contain the standard deviation values for each metric (e.g., revenue\_stats$SD for Revenue).

The result is a table that summarizes the key statistics (mean, median, SD) for each financial metric.

**Line 2:**

r

cat("\nSummary Statistics Table:\n")

* **Purpose**: This line prints a heading or label before displaying the summary table.
* **Explanation**:
  + cat(): This function is used to concatenate and print output to the console.
  + \n: This represents a newline character, ensuring the text "Summary Statistics Table" is printed on a new line.

This prepares the user to see the summary table by printing a label above it.

**Line 3:**

r

print(summary\_table, row.names = FALSE)

* **Purpose**: This line prints the **summary table** that was created in Line 1.
* **Explanation**:
  + print(): This function displays the contents of the object passed to it — in this case, summary\_table.
  + row.names = FALSE: By default, data.frame objects print row names, but setting row.names = FALSE ensures that the row numbers (such as 1, 2, 3) are not displayed in the printed table. Only the data itself will be shown.

**Line 4:**

r

cat("\nCompany-wise Financials Table:\n")

* **Purpose**: Similar to Line 2, this prints a label or heading for the company-wise financials table.
* **Explanation**:
  + cat() with \n ensures that "Company-wise Financials Table" is printed on a new line, preparing for the display of company-specific financial data.

**Line 5:**

r

print(company\_data, row.names = FALSE)

* **Purpose**: This line prints the **company-wise financial data** stored in company\_data.
* **Explanation**:
  + print(): It displays the contents of company\_data (which could be a data frame containing detailed financials of multiple companies).
  + row.names = FALSE: As with the previous print statement, this ensures that the row names of the company\_data table are not displayed.

**Conclusion**

This project successfully analyzed and visualized company financial data using R. By calculating statistical metrics like mean, median, and standard deviation for revenue, profit, and expenses, valuable insights into performance trends were gained. Additionally, visualizations such as bar plots, histograms, and box plots provided clear representations of data distribution and helped identify top and bottom-performing companies.

The results demonstrate how R can serve as an effective tool for financial analysis, enabling businesses to make data-driven decisions and optimize their operations. With minimal data and computational steps, comprehensive insights were achieved, showcasing the importance of leveraging analytics in today's business environment.