# University Institute Of Computing

## **SUBJECT**-R PROGRAMING

## **PROJECT-** Weather Data Analysis


**SUBMITTED BY:-**

STUDENT NAME:- SACHIN SINGH

**BRANCH:-** BCA(DATA SCIENCE)

**SUB CODE:-** 24CAP-161

UID:- 24BCD10161

**SECTION:-**24BCD3(B)

**SEMESTER:-** SECOND

**SUBMITTED To:-**

TEACHER NAME:- MS. SHIVANI

u

# Mini Project Report: Weather Data Analysis in R

## 1. Introduction

Weather plays a significant role in our daily lives and has a broad impact on agriculture, transportation, and human comfort. This mini-project is focused on analyzing weather data from multiple cities using R, a powerful language for statistical computing and data visualization. The analysis includes exploring temperature, humidity, rainfall, and wind speed, along with generating insightful visualizations to better understand weather trends.

## 2. Aim

To develop an R program that performs exploratory data analysis on weather data, utilizing statistical measures and various graphical techniques to uncover patterns and insights.

## 3. Objectives

• Perform basic statistical analysis such as mean, median, and standard deviation on weather parameters.
• Visualize the data using bar plots, box plots, histograms, and scatter plots.
• Identify cities with extreme weather conditions (highest and lowest values).
• Interpret the results and derive meaningful conclusions from the data.

## 4. Tools and Technologies Used

• R Programming Language
• RStudio IDE (Optional)
• ggplot2 package for advanced visualizations
• Base R functions for statistical analysis and plotting

## 5. Input and Output

Input:
A data frame containing weather statistics for a list of cities with the following columns:
- City
- Temperature (°C)
- Humidity (%)
- Rainfall (mm)

- Wind Speed (km/h)

Output:
- Tabular display of the dataset
- Summary statistics for each weather parameter
- Graphical representations: bar plot, box plot, histogram, scatter plot
- Identification of cities with extreme weather conditions
- Observations and conclusions based on the analysis

## 6. Procedure

Step 1: Load the required libraries such as ggplot2 for plotting.
Step 2: Create a sample weather dataset with fictional or real data.
Step 3: Display the dataset in tabular form.
Step 4: Calculate statistical measures (mean, median, standard deviation) for each parameter.
Step 5: Generate visualizations to understand data distribution and relationships.
Step 6: Identify the cities with the highest and lowest values for each weather metric.
Step 7: Interpret and document insights gained from the analysis.

## 7. R Code Implementation

The following R code was written to perform the entire analysis and visualization:

```
# Load required libraries
library(ggplot2)

# Sample weather data
data <- data.frame(
  City = c("New York", "Los Angeles", "Chicago", "Houston", "Phoenix"),
  Temperature = c(22, 28, 16, 30, 35),
  Humidity = c(60, 50, 65, 70, 45),
  Rainfall = c(120, 50, 150, 75, 20),
  Wind_Speed = c(15, 10, 20, 12, 18)
)

print("Weather Data:")
print(data)

# Basic Statistics
weather_columns <- names(data)[-1]
for (column in weather_columns) {
  cat("\n", column, "\n")
```

```
  cat("Mean:", mean(data[[column]]), "\n")
  cat("Median:", median(data[[column]]), "\n")
  cat("Standard Deviation:", sd(data[[column]]), "\n")
}

# Visualizations
barplot(data$Temperature, names.arg = data$City, main="Avg Temperature",
col="steelblue")
boxplot(data$Humidity, data$Rainfall, data$Wind_Speed,
        names=c("Humidity", "Rainfall", "Wind Speed"),
        main="Distribution of Parameters", col=c("lightblue", "lightgreen", "lightcoral"))
hist(data$Temperature, main="Temperature Histogram", col="lightblue", breaks=5,
xlab="Temp (°C)")
plot(data$Temperature, data$Humidity, main="Temp vs Humidity",
     xlab="Temperature", ylab="Humidity", col="darkred", pch=19)
abline(lm(Humidity ~ Temperature, data = data), col="blue")

# Extreme Conditions
for (column in weather_columns) {
  max_index <- which.max(data[[column]])
  min_index <- which.min(data[[column]])
  cat("\n", column, "\n")
  cat("Max:", data$City[max_index], data[[column]][max_index], "\n")
  cat("Min:", data$City[min_index], data[[column]][min_index], "\n")
}
```

## 8. Observations and Insights

• Cities like Phoenix and Houston show higher temperatures compared to others.
• Rainfall is highest in Chicago, while Phoenix is the driest.
• The scatter plot and regression line suggest a possible inverse relationship between temperature and humidity.
• Box plots reveal variability in rainfall and wind speed across different cities.

## 9. Learning Outcomes

• Understood how to perform data manipulation and statistical analysis in R.
• Gained practical experience in using data visualization to explore datasets.
• Learned how to interpret data trends and summarize insights.
• Applied knowledge of R programming to a real-world scenario involving weather statistics.

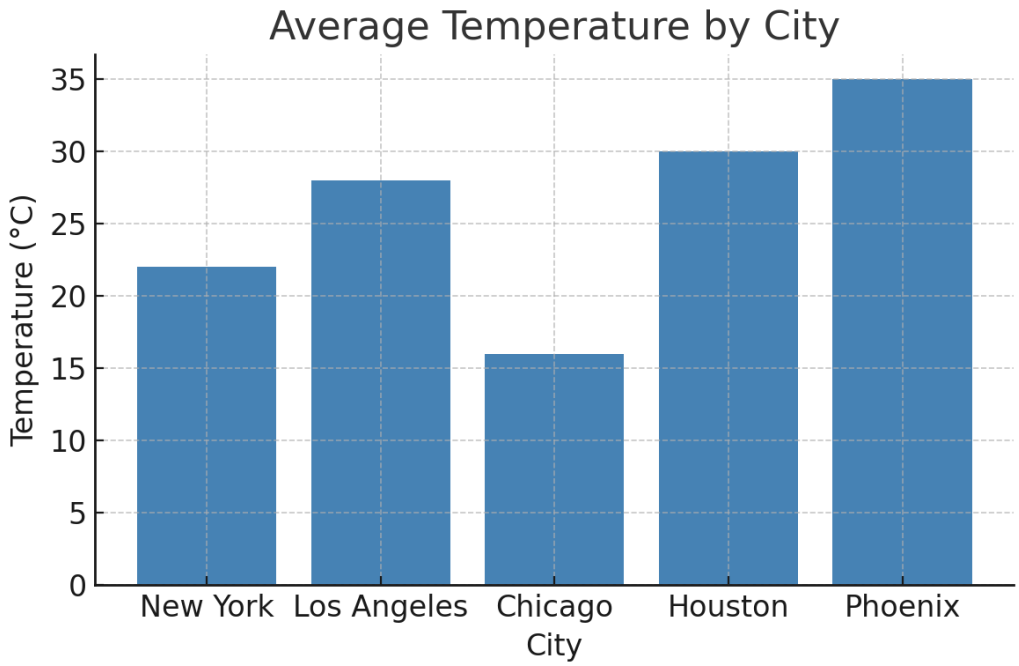
## 10. Conclusion

This mini project successfully demonstrated how R can be used for exploratory data analysis and visualization of weather data. It provided valuable insights into weather patterns, variability, and inter-parameter relationships across different cities. Such projects help develop analytical thinking and programming skills, useful for data science and environmental analytics.

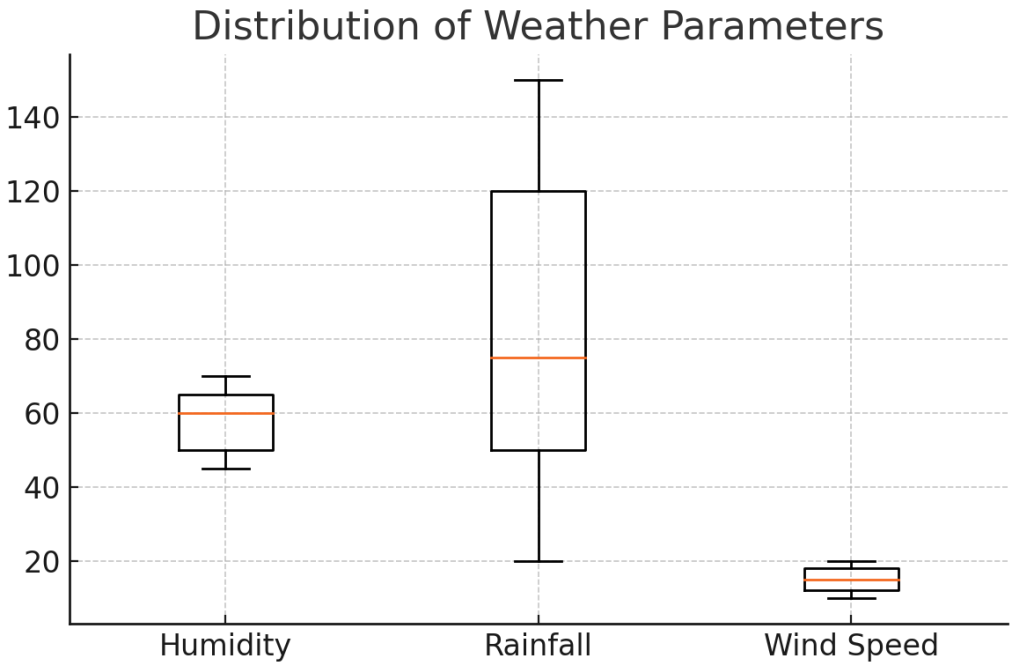
## 11. Output Visualizations

The following visualizations are generated from the analysis in R:

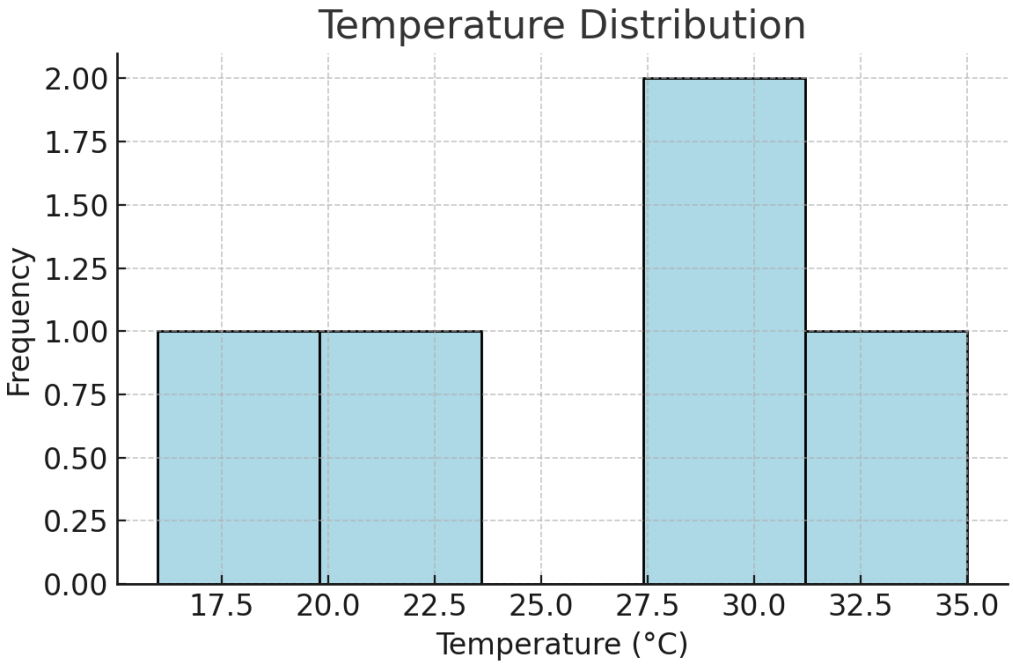
### Bar Plot: Average Temperature by City



### Box Plot: Distribution of Weather Parameters




### Histogram: Temperature Distribution



## Scatter Plot: Temperature vs. Humidity