# COMPARING ANNUAL RAINFALL ACROSS STATES USING BOXPLOTS

**A PROJECT REPORT**

***Submitted by***

**MOHAMMED RAASITH H(2303811724321069)**

***in partial fulfillment of requirements for the award of the course***

**AGI1252-FUNDAMENTALS OF DATA SCIENCE USING R**

***in***

**ARTIFICIAL INTELLIGENCE AND DATA SCIENCE**

**K. RAMAKRISHNAN COLLEGE OF TECHNOLOGY**

(An Autonomous Institution, affiliated to Anna University Chennai and Approved by AICTE, New Delhi)

**SAMAYAPURAM – 621 112**

**JUNE- 2025**

# K. RAMAKRISHNAN COLLEGE OF TECHNOLOGY (AUTONOMOUS)

**SAMAYAPURAM – 621 112**

## BONAFIDE CERTIFICATE

Certified that this project report on **"COMPARING ANNUAL RAINFALL ACROSS STATES USING BOXPLOTS"** is the bonafide work of **MOHAMMED RAASITH H (2303811724321069)** who carried out the project work during the academic year 2024 - 2025 under my supervision.

**SIGNATURE**

Dr.T. AVUDAIAPPAN, M.E.,Ph.D.,

**HEAD OF THE DEPARTMENT**

ASSOCIATE PROFESSOR

Department of Artificial Intelligence

K.Ramakrishnan College of Technology (Autonomous)

Samayapuram–621112.

**SIGNATURE**

Ms.S. Murugavalli., M.E.,(Ph.D).,

**SUPERVISOR**

ASSISTANT PROFESSOR

Department of Artificial Intelligence

K.Ramakrishnan College of Technology (Autonomous)

Samayapuram–621112.

Submitted for the viva-voce examination held on ………….**02.06.2025**

**INTERNAL EXAMINER**

**EXTERNAL EXAMINER**

# DECLARATION

I declare that the project report on **“COMPARING ANNUAL RAINFALL ACROSS STATES USING BOXPLOTS”** is the result of original work done by us and best of our knowledge, similar work has not been submitted to **“ANNA UNIVERSITY CHENNAI”** for the requirement of Degree of **BACHELOR OF TECHNOLOGY**. This project report is submitted on the partial fulfilment of the requirement of the completion of the course **AGI1252 – FUNDAMENTALS OF DATA SCIENCE USING R.**

**Signature**

Name of the Student
MOHAMMED RAASITH H

Place: Samayapuram

Date:02.06.2025

**ACKNOWLEDGEMENT**

It is with great pride that I express our gratitude and in-debt to our institution "**K.Ramakrishnan College of Technology (Autonomous)**", for providing us with the opportunity to do this project.

I glad to credit honourable chairman **Dr. K. RAMAKRISHNAN**, **B.E.,** for having provided for the facilities during the course of our study in college.

I would like to express our sincere thanks to our beloved Executive Director **Dr. S. KUPPUSAMY, MBA, Ph.D.,** for forwarding to our project and offering adequate duration in completing our project.

I would like to thank **Dr. N. VASUDEVAN, M.Tech., Ph.D.,** Principal, who gave opportunity to frame the project the full satisfaction.

I whole heartily thanks to **Dr. T. AVUDAIAPPAN, M.E.,Ph.D.,** Head of the department, **ARTIFICIAL INTELLIGENCE** for providing his encourage pursuing this project.

I express our deep expression and sincere gratitude to our project supervisor **Mrs.S.GEETHA, M.E.,** Department of **ARTIFICIAL INTELLIGENCE,** for her incalculable suggestions, creativity, assistance and patience which motivated us to carry out this project.

I render our sincere thanks to Course Coordinator and other staff members for providing valuable information during the course.

I wish to express our special thanks to the officials and Lab Technicians of our departments who rendered their help during the period of the work progress.

**INSTITUTE**

**Vision:**

- To serve the society by offering top-notch technical education on par with global standards.

**Mission:**

- Be a center of excellence for technical education in emerging technologies by exceeding the needs of industry and society.
- Be an institute with world class research facilities.
- Be an institute nurturing talent and enhancing competency of students to transform them as all – round personalities respecting moral and ethical values.

**DEPARTMENT**

**Vision:**

- To excel in education, innovation, and research in Artificial Intelligence and Data Science to fulfil industrial demands and societal expectations.

**Mission**

- To educate future engineers with solid fundamentals, continually improving teaching methods using modern tools.
- To collaborate with industry and offer top-notch facilities in a conducive learning environment.
- To foster skilled engineers and ethical innovation in AI and Data Science for global recognition and impactful research.
- To tackle the societal challenge of producing capable professionals by instilling employability skills and human values.

**PROGRAM EDUCATIONAL OBJECTIVES (PEO)**

- **PEO1:** Compete on a global scale for a professional career in Artificial Intelligence and Data Science.
- **PEO2:** Provide industry-specific solutions for the society with effective communication and ethics.
- **PEO3** Enhance their professional skills through research and lifelong learning initiatives.

**PROGRAM SPECIFIC OUTCOMES (PSOs)**

- **PSO1:** Capable of finding the important factors in large datasets, simplify the data, and improve predictive model accuracy.
- **PSO2:** Capable of analyzing and providing a solution to a given real-world problem by designing an effective program.

**PROGRAM OUTCOMES (POs)**

Engineering students will be able to:

1. **Engineering knowledge:** Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
2. **Problem analysis:** Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences
3. **Design/development of solutions:** Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations
4. **Conduct investigations of complex problems:** Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions
5. **Modern tool usage:** Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations
6. **The engineer and society:** Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice
7. **Environment and sustainability:** Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development
8. **Ethics:** Apply ethical principles and commit to professional ethics and

responsibilities and norms of the engineering practice.

9. **Individual and team work:** Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

10. **Communication:** Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

11. **Project management and finance:** Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

12. **Life-long learning:** Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

# ABSTRACT

The rapid advancement of data analytics has made visual representation an essential tool for understanding regional climate patterns. In this project, we developed a Rainfall Comparison System using the R programming language to visualize and compare annual rainfall across Indian states through boxplots. The primary aim of this application is to allow users to explore state-wise rainfall variability using preloaded or simulated datasets, and generate clean, comparative plots. This interactive platform is particularly valuable for geographers, environmental analysts, and learners seeking a deeper understanding of rainfall distribution and anomalies across regions. The user interface is designed for clarity, with minimal aesthetics that emphasize functional data presentation over unnecessary visual complexity. By leveraging libraries such as lattice, plotly, and base R graphics, the system enables the rendering of insightful boxplots that respond to categorical grouping and visual customization. This project not only demonstrates the utility of R for environmental data visualization but also showcases how multi-library integration can enhance interactivity and comparison in statistical graphics. Overall, this project serves as a practical example of how open-source tools can be used to democratize access to climate insights. It also lays the foundation for further enhancements, including seasonal trend comparisons, integration with real rainfall datasets, and expanded geographic coverage.

**ABSTRACT WITH POs AND PSOs MAPPING**

**CO 5 : BUILD DATABASES FOR SOLVING REAL-TIME PROBLEMS.**

| ABSTRACT | POs MAPPED | PSOs MAPPED |
|---|---|---|
| The rapid growth of data analytics has made visual tools essential for understanding climate patterns. This project presents a Rainfall Comparison System built in R, enabling users to visualize and compare annual rainfall across Indian states using boxplots. Designed with simplicity and clarity, the interactive platform supports both preloaded and simulated datasets, catering to geographers, environmental analysts, and learners.<br>Leveraging R libraries like lattice, plotly, and base graphics, the system produces clean, customizable plots that highlight rainfall variability and anomalies. Emphasizing functional design over visual complexity, it demonstrates the power of open-source tools for environmental data analysis. The project also sets the stage for future enhancements, such as seasonal trend comparisons and integration with real-time datasets. | **PO1 -3**<br>**PO2 -3**<br>**PO3 -3**<br>**PO5 -3**<br>**PO7 -2**<br>**PO11-2** | **PSO1 -3**<br>**PSO2 -3** |

Note: 1- Low, 2-Medium, 3- High

## TABLE OF CONTENTS

# CHAPTER 1

# INTRODUCTION

In the current digital age, data is one of the most valuable resources. The environmental and geographical sectors, in particular, generate large volumes of data related to climate patterns and regional statistics. Policymakers, climate scientists, and educators rely on this data to identify trends, evaluate environmental conditions, and support data-driven planning. However, raw environmental datasets in CSV or tabular formats can often be overwhelming and difficult to interpret without appropriate visualization tools. To address this need, our project titled "COMPARING ANNUAL RAINFALL ACROSS STATES USING BOXPLOTS" aims to create a user-friendly and interactive visualization tool that simplifies the process of analyzing and comparing annual rainfall data across different regions. Built using the R programming language, this application simulates and displays rainfall values across multiple Indian states such as Tamil Nadu, Kerala, Karnataka, and Andhra using dynamic boxplots.

What makes this project highly relevant and impactful is its clarity, simplicity, and flexibility. Users are not restricted to static charts or limited visual styles — instead, they can interact with rainfall data using three powerful visualization libraries: base R graphics, lattice, and plotly. This flexibility allows users to switch between static and interactive views depending on their needs, whether for academic presentation, research insight, or learning purposes. The interface and graphs are styled for readability using color-coded visuals and minimal design elements, helping to ensure data-driven insights are easy to interpret and communicate. The plots created reflect the variability, range, and median rainfall values in each state, helping users grasp climatic differences across regions in a single glance. This project stands as an example of how open-source tools like R and its visualization packages can be applied to environmental studies.

## 1.1 OBJECTIVE

The primary objective of this project is to design and implement an interactive data visualization tool that enables users to compare annual rainfall across different Indian states using boxplots. the project aims to:

Simplify the interpretation of regional rainfall data for users with little technical background.Enable dynamic visualization of climatic variations using various boxplot techniques from base R, lattice, and plotly.Offer a lightweight and efficient alternative to more complex GIS or statistical software used for climate analysis.

## 1.2 OVERVIEW

This R project visualizes annual rainfall distribution across Indian states using boxplots. It supports three visualization methods: base R, lattice, and Plotly for interactive output. The rainfall data is stored in a data frame with state names and annual rainfall values, with multiple entries per state to simulate yearly or district-level variation. Rainfall values are generated using the rnorm() function to mimic realistic data patterns. For visualization, base R uses the boxplot() function with color coding and labeled axes, while lattice employs bwplot() with advanced styling options. Plotly's plot_ly() creates interactive, hover-enabled boxplots. Each method effectively highlights data variation, central tendency, and outliers. Titles and axis labels are automatically generated to ensure clarity and context throughout the visualizations.

## 1.3 R PROGRAMMING CONCEPTS USED

This project uses core R programming concepts and advanced visualization libraries to efficiently analyze and compare rainfall data across Indian states. A structured dataset is created using data.frame(), with rep() and rnorm() functions to simulate realistic annual rainfall values for states like Tamil Nadu, Kerala, Karnataka, and Andhra. The data is organized to ensure consistency and repeatability, with the State column treated as a factor to support grouped visualizations.

For visual representation, base R's boxplot() function is used for simple comparisons, with customizable colors, labels, borders, and orientation. The lattice package provides more stylized boxplots through bwplot() and par.settings, allowing detailed control over visual elements for publication-ready graphics. Meanwhile, the plotly package brings interactivity to the analysis using plot_ly(), offering features like tooltips, hover effects, and dynamic grouping to enhance user engagement.

To ensure visual clarity and consistency across methods, color palettes are selected using the col and colors parameters, contributing to both aesthetics and accessibility. Overall, the project effectively combines basic R tools with modern libraries to create a comprehensive, interactive dashboard for comparing grouped rainfall data, offering deeper insights into regional climate patterns.

# CHAPTER 2

# PROJECT METHEDOLOGY

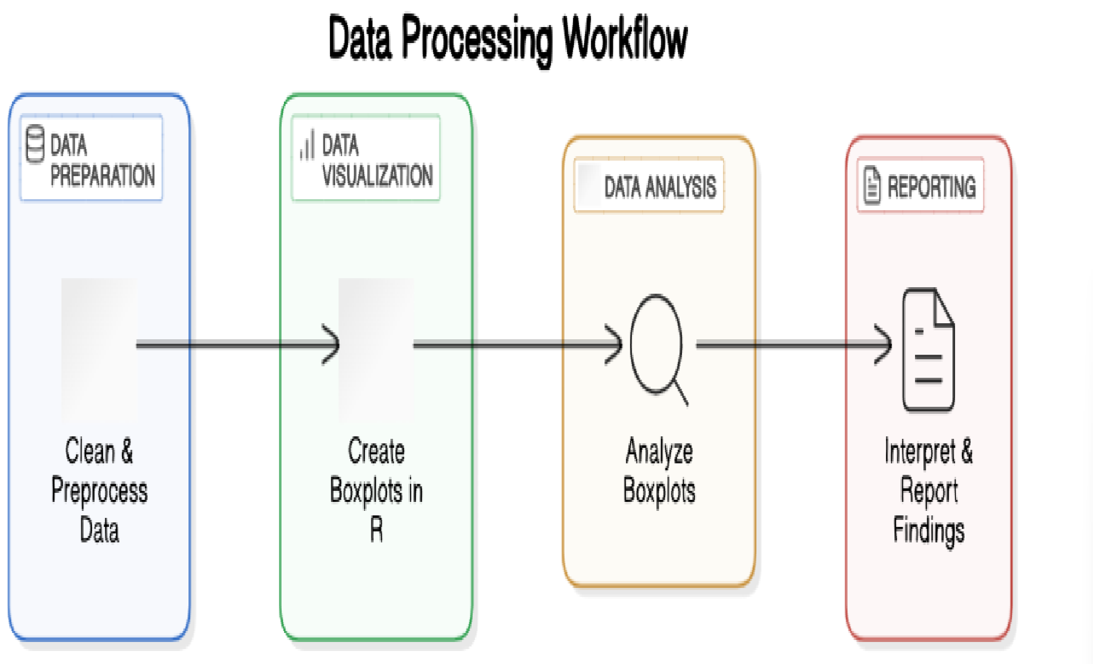
## 2.1 PROPOSED WORK

The proposed system is an interactive rainfall comparison application built in R, designed to help users analyze and interpret climatic variations using both static and dynamic boxplots. It provides a reusable and customizable framework that visualizes annual rainfall data across Indian states using a consistent dataset structure with columns like State and AnnualRainfall. Users can either generate synthetic data with rnorm() or upload CSV files containing real values.

For visualization, the system supports base R's boxplot() for quick comparisons showing medians, interquartile ranges, and outliers, while the lattice package's bwplot() delivers more polished graphics suitable for reporting. The plotly package introduces interactivity, allowing users to hover, zoom, and explore rainfall patterns more deeply.

This tool addresses the needs of educators, researchers, and students in climate science by offering an accessible way to interpret large datasets visually without relying on complex platforms. It is portable across R-compatible systems, flexible for use with rainfall data from any region, and scalable for future .

# 2.2 BLOCK DIAGRAM



# CHAPTER 3

# MODULE DESCRIPTION

## 3.1 FILE UPLOAD & DATA INGESTION MODULE

**Purpose:**

The purpose of this module is to allow users to either upload real rainfall datasets or generate synthetic data for visualization. It ensures that the application can work flexibly with user-supplied inputs or internally created datasets.

**Description:**

This module supports uploading CSV files or generating rainfall data using rnorm(). It reads and loads the data into a structured data frame with "State" and "AnnualRainfall" columns. It checks for the presence of necessary fields and converts state names to factors for grouped plotting. The module handles missing or inconsistent entries and provides feedback to users. It ensures the data is cleaned and formatted for visualization modules

**Key Functions:**

1. Accept CSV file input through a user interface component.
2. Generate synthetic rainfall data using rnorm() when no file is uploaded.
3. Validate the structure and contents of the dataset (check for required columns)

## 3.2 USER INPUT SELECTION MODULE

**Purpose:**

The purpose of this module is to let users customize and control the visualization process by selecting plot types, data sources, and display preferences. It enhances user engagement and ensures flexible, tailored analysis based on selected options.

**Description:**

This module provides an interactive UI for selecting visualization types such as Base R, Lattice, or Plotly. Users can also choose to upload data or use simulated datasets. This module offers an interactive interface for users to choose visualization types like Base R, Lattice, or Plotly. It supports both uploaded and simulated datasets. Users can filter states, select color themes, and adjust plot orientation. Inputs are captured dynamically, and plots update instantly without reloading. It ensures smooth, valid, and customizable user interactions for flexible analysis.It allows filtering by state, selecting color themes, and adjusting axis orientation. The module captures inputs dynamically and updates the output instantly without needing a page reload. It ensures valid selections and prevents unsupported configurations. This makes the application responsive, user-friendly, and adaptable to individual analysis preferences.

**Key Functions:**

1. Allow users to select the type of boxplot (Base R, Lattice, Plotly).
2. Enable filtering of specific states or data ranges.
3. Provide customization options like color and axis orientation.

## 3.3 DATA PROCESSING AND REACTIVE LOGIC MODULE

**Purpose:**

The purpose of this module is to handle data processing and manage the reactive logic that updates visualizations based on user inputs. It ensures that the data is filtered, transformed, and prepared correctly for plotting, providing accurate and timely results. This module supports dynamic interaction and responsiveness throughout the application.

**Description:**

This module processes the uploaded or generated rainfall data according to user selections and filters. It applies necessary transformations like grouping, summarizing, or cleaning to prepare data for visualization. The reactive framework ensures that any change in input triggers automatic updates in the processed data and corresponding plots. It manages dependencies between inputs and outputs, maintaining smooth and efficient performance. Error handling and data validation are incorporated to avoid incorrect visualizations. The module ensures seamless integration between data changes and visualization updates, making the application interactive and reliable.

**Key Functions:**

1. Process and filter data based on user input selections.
2. Apply necessary data transformations and cleaning steps.
3. Implement reactive logic to update visualizations instantly.

# 3.4 VISUALIZATION MODULE

**Purpose:**

The purpose of this module is to generate and display rainfall data visualizations based on processed data and user preferences. It supports multiple plotting methods to provide both static and interactive graphical outputs. This module helps users easily interpret rainfall patterns through clear and customizable boxplots.

**Description:**

This module creates visual representations of rainfall distribution using Base R's boxplot(), Lattice's bwplot(), and Plotly's interactive plotting functions. It adapts the visual output according to user-selected options such as color schemes, plot orientation, and state filtering. The module integrates seamlessly with reactive data inputs to update plots instantly when data or settings change. It ensures that each plot is clearly labeled with titles and axis descriptions for better readability. The interactive Plotly plots allow users to hover and zoom for deeper data exploration. This module balances simplicity and interactivity to cater to varied user needs in rainfall analysis.

**Key Functions:**

1. Generate static boxplots using Base R's boxplot().
2. Create stylized boxplots with Lattice's bwplot().
3. Render interactive, hover-enabled plots using Plotly's plot_ly().

# CHAPTER 4

# CONCLUSION AND FUTURE SCOPE

## CONCLUSION

In this project, we developed an interactive data visualization tool in R that compares annual rainfall across Indian states using boxplots. The aim was to create a clear, visually rich interface that enables users to understand rainfall distribution patterns across multiple regions over time. By generating boxplots through multiple visualization libraries — including **base R**, **lattice**, and **plotly** — the tool offers both static and interactive representations for deeper analysis.

The application achieves this goal through a structured approach that involves:

- Generation or uploading of rainfall datasets,
- Flexible user input for selecting the preferred plotting library,
- Reactive or script-based backend logic for dynamic plot generation,
- And clean, multi-library visualizations showing rainfall trends by state.

The rainfall data is simulated using the rnorm() function, covering four states: *Tamil Nadu, Kerala, Karnataka,* and *Andhra*. For each state, random annual rainfall values are generated to represent regional

This solution is both flexible and scalable, designed to help students, researchers, meteorologists, and policy planners analyze climatic variation easily. The use of interactive plots ensures a more engaging experience, especially for non-technical users or learners new to data visualization.

## FUTURE SCOPE

While the current version successfully meets its core objective, the following enhancements can significantly increase its value and applicability:

### 1.Real Dataset Integration

Enable the application to ingest CSV files with actual meteorological rainfall data from IMD or other open datasets for real-world analysis.

### 2.Seasonal Comparison Support

Extend functionality to handle *monthly* or *seasonal* rainfall averages (e.g., monsoon vs. dry season), allowing users to toggle between time periods.

### 3. Advanced Statistical Insights

Integrate features such as:

- Mean and median overlays on boxplots,
- Standard deviation visualization,
- Outlier detection indicators.

### 4. Interactive UI with Shiny

Convert the script into a Shiny web application that includes:

- File upload interface,
- Plot type selection,
- Real-time rendering.

## 5. Region-wise Rainfall Mapping

Incorporate geospatial visualization using leaflet or ggmap to represent rainfall data geographically across India.

### 6. Export & Reporting Tools

Add functionality for:

- Downloading boxplots as images (PNG/PDF),
- Exporting summary statistics (mean, median, IQR) in tabular form.

### 7. Mobile and Web Responsiveness

If implemented as a Shiny app, optimize layout and responsiveness for mobile and tablet usa

# CHAPTER 5

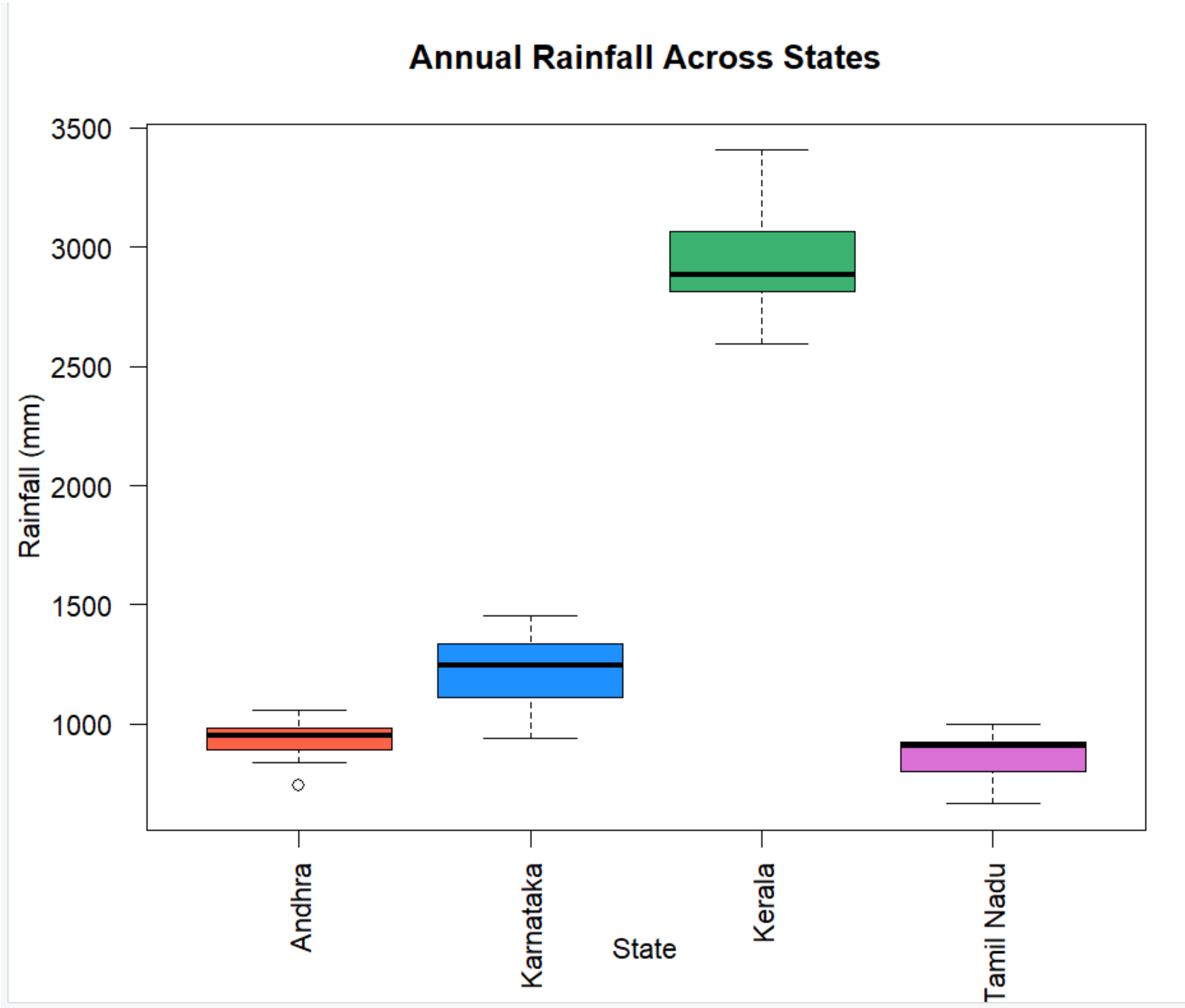
## APPENDIX A SOURCE CODE

```
boxplot(AnnualRainfall ~ State,
        data = rainfall_data,
        col = c("tomato", "dodgerblue", "mediumseagreen", "orchid"),
        main = "Annual Rainfall Across States",
        xlab = "State",
        ylab = "Rainfall (mm)",
        las = 2,
        border = "black")
install.packages("lattice")  # only once
library(lattice)


bwplot(AnnualRainfall ~ State,
       data = rainfall_data,
       main = "Annual Rainfall Across States",
       xlab = "State",
       ylab = "Rainfall (mm)",
       par.settings = list(box.rectangle = list(col = "darkblue"),
                           plot.symbol = list(col = "red", pch = 16),
                           box.umbrella = list(col = "green")))
install.packages("plotly")  # only once
library(plotly)


plot_ly(rainfall_data,
        x = ~State,
        y = ~AnnualRainfall,
        type = "box",
        color = ~State,
        colors = c("tomato", "steelblue", "mediumseagreen", "orchid")) %>%
  layout(title = "Annual Rainfall Across States",
         xaxis = list(title = "State"),
         yaxis = list(title = "Rainfall (mm)"))
```

# APPENDIX B SCREENSHOTS



## REFERENCES


1. Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
   https://ggplot2.tidyverse.org
   (Used for visualizing rainfall data in an aesthetically appealing and layered format.)
2. Sarkar, D. (2008). *Lattice: Multivariate Data Visualization with R*. Springer.
   https://lattice.r-forge.r-project.org/
   (Applied for generating advanced multi-panel boxplots using the bwplot() function.)
3. Sievert, C. (2020). *Interactive Web-Based Data Visualization with R, plotly, and shiny*. CRC Press.
   https://plotly-r.com/
   (Used to implement interactive rainfall boxplots for dynamic exploration.)
4. R Core Team. (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
   https://www.r-project.org
   (Primary environment for developing the data analysis and visualization components.)
5. Grolemund, G., & Wickham, H. (2011). *Dates and Times Made Easy with lubridate*. Journal of Statistical Software, 40(3), 1–25.
   https://www.jstatsoft.org/article/view/v040i03
   (Referenced for potential extension to seasonal or time-based rainfall comparisons.)
6. Indian Meteorological Department (IMD). (2024). *Rainfall Statistics of India*.
   https://mausam.imd.gov.in/
   (To be integrated for future real-data ingestion and rainfall trend analysis.)