Comparing Average Rainfall in Two Different Cities

A PROJECT REPORT

Submitted by

OM PRAKASH M -2303811724321079

in partial fulfillment of requirements for the award of the course

AGB1252-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

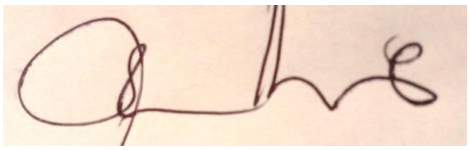
JUNE- 2025

K. RAMAKRISHNAN COLLEGE OF TECHNOLOGY (AUTONOMOUS)

SAMAYAPURAM – 621 112

BONAFIDE CERTIFICATE

Certified that this project report on **COMPARING AVERAGE RAINFALL OF TWO CITIES** is the bonafide work of **OM PRAKASH M-2303811724321079** 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

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 AVERAGE RAINFALL OF TWO CITIES** is the result of original work done by me and best of my 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 **AGB1252 – FUNDAMENTALS OF DATA SCIENCE USING R.**

.

Signature

OM PRAKASH M

Name of the Student

Place: Samayapuram

Date:30/05/2025

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 **S.Murugavalli, 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

This project, titled **comparing average rainfall of two cities**, presents an interactive web application built using the R programming language and the Shiny framework. The objective of the project is to facilitate a clear and intuitive comparison of monthly rainfall patterns between two different cities over a one-year period. Users input rainfall data as comma-separated values for each city, and the application processes and visualizes this information through dynamic bar plots and line charts. Additionally, a detailed textual report summarizes the average rainfall for both cities, indicating which city experiences greater precipitation. By combining data visualization with basic statistical analysis, this tool serves as an accessible and educational platform for understanding climatological differences between locations. The application is particularly useful for environmental studies, agriculture planning, and general weather analysis.

ABSTRACT WITH POs AND PSOs MAPPING

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

ABSTRACT	POs MAPPED	PSOs MAPPED
This project presents an interactive web application developed using the R programming language and the Shiny framework. The primary objective is to provide a clear and intuitive comparison of monthly rainfall patterns between two different cities over a one-year period.Users input rainfall data as comma-separated values (CSV) for each city. The application processes this data and generates dynamic visualizations, including bar plots and line charts, to depict the rainfall trends effectively. In addition to the visual outputs, a detailed textual summary is provided, highlighting the average monthly rainfall for each city and identifying which city receives greater overall precipitation.	**PO1 -3** **PO2 -3** **PO3 -3** **PO4 -3** **PO5 -3**	**PSO1 -3** **PSO2 -3** **PSO3 -3**

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

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 OBJECTIVE

The primary objective of this project is to develop an interactive Shiny web application that allows users to compare the monthly average rainfall between two cities. The application aims to help users visualize rainfall patterns, analyze seasonal variations, and determine which city receives more rainfall on average. By providing graphical representations and summary statistics, the tool enables effective comparison and supports data-driven insights into regional climate differences.

1.2 OVERVIEW

The project **"Comparing Average Rainfall of Two Cities"** is an interactive web application developed using the R programming language and the Shiny framework. It aims to facilitate the comparison of monthly rainfall data between two cities over a one-year period. Users input rainfall values as comma-separated entries for each city, which are then processed to generate dynamic visualizations such as bar plots and line charts. In addition to visual representations, the application provides a textual summary that calculates and compares the average rainfall of the two cities, helping identify which location receives more precipitation. The project effectively combines data analysis, visualization, and user interaction, making it a practical tool for climatology studies, environmental monitoring, and agricultural planning.

1.3 R CONCEPTS USED

1. **Data Frames**
2. **String Manipulation and Data conversion**
3. **Data Visualization with ggplot2**
4. **Shiny Web Application**
5. **Statistical Summarization**

CHAPTER 2
PROJECT METHODOLOGY

2.1 PROPOSED WORK

The methodology for the **Comparing Average Rainfall of Two Cities** project follows a systematic flow from data entry to statistical analysis and visual reporting, all integrated into a Shiny-based interactive web application.

1. Data Collection / Import

- Users can either manually enter monthly rainfall values or upload rainfall data in CSV or Excel formats.
- The application supports input for two cities, with monthly rainfall data from January to December.

2. Data Preprocessing

- The input data is cleaned using base R functions. Missing values are automatically filled with the average rainfall of the corresponding city to ensure consistency.
- Data is converted from text to numeric format and validated to ensure it contains 12 months of data per city.

3. User Interface Design

- Built using Shiny's layout functions such as fluidPage, sidebarLayout, and widgets like textInput, fileInput, and actionButton.
- Users are provided with tabs to view bar plots, trend lines, and rainfall summary reports.

4. Server Logic Implementation

- Reactive expressions handle user input and file upload dynamically.
- The server logic computes city-wise average rainfall and prepares the data for plotting and reporting.

5. Statistical Analysis

- Core statistical functions such as mean() are applied to compute the average rainfall for each city.
- Conditional logic is used to compare the two averages and generate a textual interpretation of the result.

6. Data Visualization

- Bar plots and line plots are generated using base R plotting functions to visually compare rainfall data month-by-month.
- Dynamic Y-axis scaling and legends are included to make the charts more informative and readable.

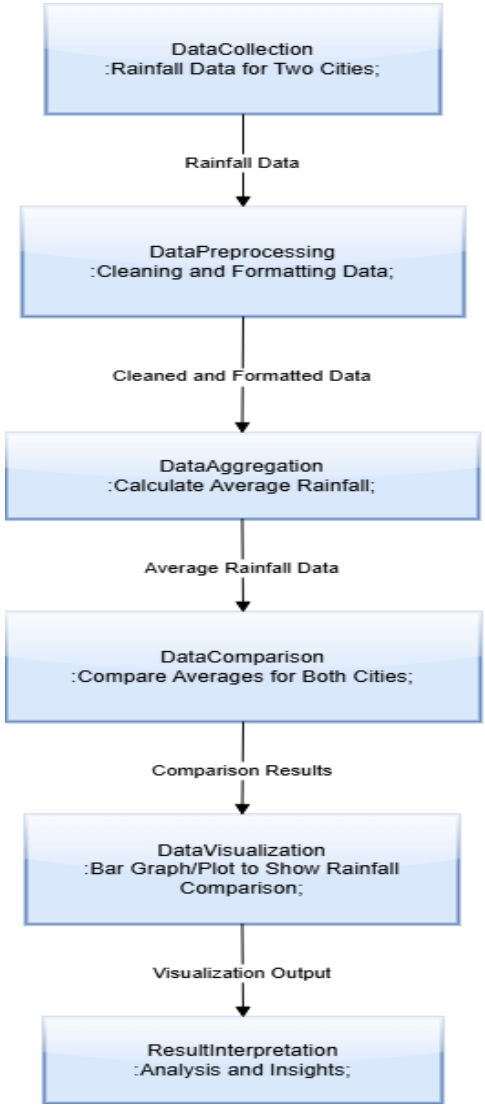
7. Interactive Outputs

- All graphs and text outputs are updated in real-time based on user input or uploaded files.
- Users immediately see how their data changes affect the visual and analytical results.

8. Deployment

- The Shiny application is designed to run locally or on a web server using shinyapps.io, allowing access through any modern browser.
- The deployment ensures interactive analysis without requiring local R installation.

2.2 BLOCK DIAGRAM



CHAPTER 3

MODULE DESCRIPTION

3.1 DATA COLLECTION MODULE

The Import Module is responsible for loading the commute time dataset into the R environment. Users can upload CSV files containing data on commute durations and transport modes. The module ensures that the uploaded data is correctly read using functions like read.csv() or readr::read_csv(). This module serves as the entry point for data analysis, supporting both custom user-uploaded files and built-in sample datasets to facilitate testing and demonstrations.

3.2. DATA STORAGE AND MANAGEMENT MODULE

The **Data Storage and Management Module** in the project **"Comparing Average Rainfall of Two Cities"** is responsible for handling user input, validating data, and organizing it for analysis. Users enter monthly rainfall data for two cities as comma-separated values. The module ensures that exactly 12 numeric, non-negative values are provided for each city, corresponding to the months of the year. Once validated, the data is stored temporarily in R as a structured data frame

3.3 USER INTERFACE (UI) MODULE

The UI Module is designed using Shiny's layout functions and input widgets such as fileInput, selectInput, and checkboxGroupInput. It provides an intuitive and user-friendly interface where users can upload their data, choose transport modes, and select which types of graphs or statistics to display. This module focuses on enhancing user experience and making the data analysis process interactive and accessible even for non-programmers.

3.4 PREDICTION AND FORECASTING MODULE

The Prediction and Forecasting Module in the "Comparing Average Rainfall of Two Cities" project is designed to extend the functionality of the application by offering future insights based on historical rainfall data. This module enables users to anticipate upcoming rainfall patterns, supporting decision-making in areas such as agriculture, urban planning, and water resource management.

3.5 ALERT AND NOTIFICATION MODULE

The **Alert and Notification Module** is a valuable enhancement to the "Comparing Average Rainfall of Two Cities" project, aimed at increasing user engagement and awareness by proactively delivering important updates based on rainfall data analysis. This module is responsible for monitoring user input or uploaded data in real-time and generating automated alerts or notifications when specific conditions are met.

3.6 DATA VISUALIZATION MODULE

This module generates various graphical representations of the commute data using the ggplot2 package. It includes histograms, boxplots, bar charts, and density plots to show the distribution and comparison of commute times across different transport modes. These visualizations are updated interactively and allow users to explore patterns, identify outliers, and gain insights from their data visually.

CHAPTER 4
CONCLUSION

4.1 CONCLUSION

This project successfully demonstrates how R and Shiny can be used to create an interactive and user-friendly application for comparing average monthly rainfall between two cities. By integrating both manual input and file upload features, users can analyze their own data flexibly. The application processes the input, calculates statistical summaries, and generates insightful visualizations such as bar charts and trend lines. These tools help users easily identify differences in rainfall patterns over the year. The use of reactive programming in Shiny ensures real-time updates and a dynamic interface. Statistical functions like mean() and data validation techniques were effectively implemented. The visualization components make data interpretation intuitive, even for non-technical users. The project emphasizes the importance of clean data handling and user interaction in data-driven applications. Overall, it highlights the power of R in building interactive decision-support tools. Future improvements can include adding more statistical tests and forecasting capabilities.

4.2 FUTURE SCOPE

1.Inclusion of Multiple Cities

- Extend the application to compare rainfall data for more than two cities.
- Add dynamic inputs to handle any number of cities with appropriate visualization scaling.

2. Data Forecasting and Trend Analysis

- Integrate time series forecasting models (e.g., ARIMA, Holt-Winters) to

predict future rainfall.

- Display projected rainfall trends to support agricultural or city planning decisions.

3. Advanced Statistical Analysis

- Include additional metrics like standard deviation, variance, and correlation.
- Add hypothesis testing (e.g., t-test) to evaluate the significance of differences in rainfall.

4. Geographical Visualization

- Use packages like leaflet or ggmap to show rainfall distribution on interactive maps.
- Link rainfall data to geographical coordinates for better spatial analysis.

5. Data Source Integration

- Enable real-time data fetching from APIs like OpenWeather or Indian Meteorological Department.
- Automate the update process for rainfall data instead of manual entry.

APPENDICES

APPENDIX A – SOURCE CODE

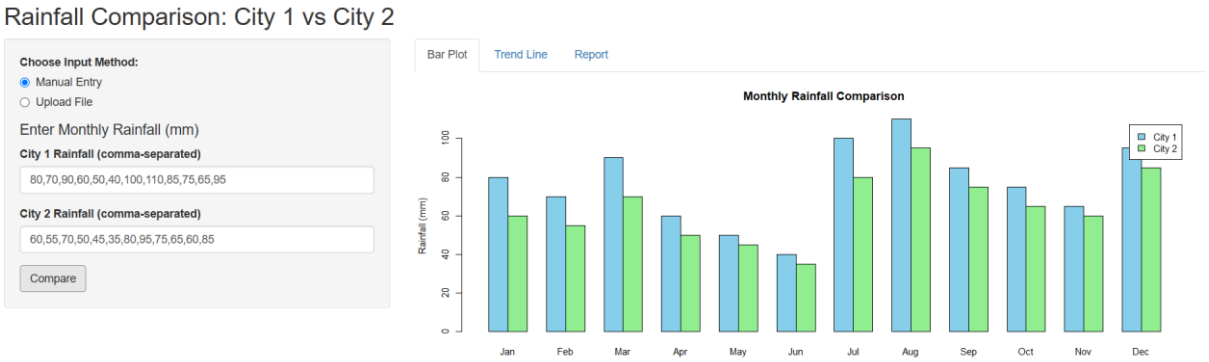
```
# install.packages("shiny")
# install.packages("readxl")

library(shiny)
library(readxl)

months <- month.abb  # Jan to Dec

# UI
ui <- fluidPage(
  titlePanel("Rainfall Comparison: City 1 vs City 2"),

  sidebarLayout(
    sidebarPanel(
      radioButtons("input_type", "Choose Input Method:",
                   choices = c("Manual Entry" = "manual", "Upload File" = "file")),

      conditionalPanel(
        condition = "input.input_type == 'manual'",
        h4("Enter Monthly Rainfall (mm)"),
        textInput("city1", "City 1 Rainfall (comma-separated)",
                  "80,70,90,60,50,40,100,110,85,75,65,95"),
        textInput("city2", "City 2 Rainfall (comma-separated)",
                  "60,55,70,50,45,35,80,95,75,65,60,85")
      ),

      conditionalPanel(
        condition = "input.input_type == 'file'",
        fileInput("datafile", "Choose CSV or Excel File",
                  accept = c(".csv", ".xlsx", ".xls")),
        helpText("File must have two columns for rainfall (City1 and City2), with
12 rows each.")
      ),

      actionButton("analyze", "Compare")
```

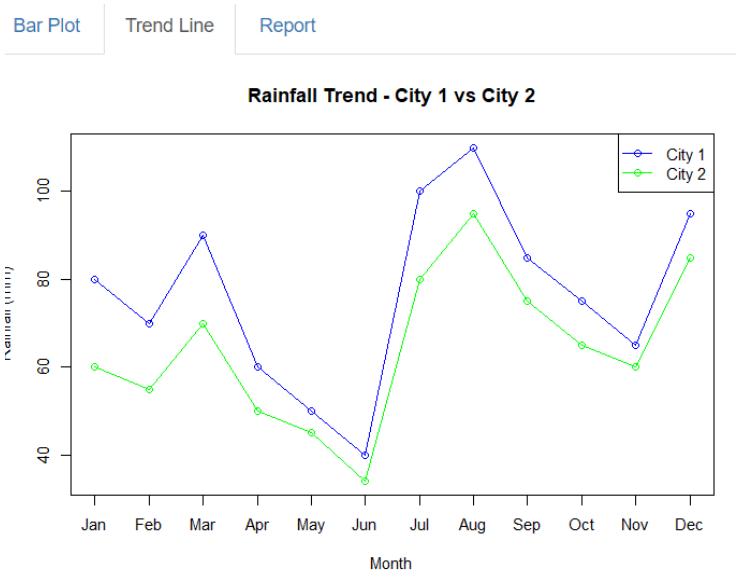
```
    ),

    mainPanel(
      tabsetPanel(
        tabPanel("Bar Plot", plotOutput("barPlot")),
        tabPanel("Trend Line", plotOutput("linePlot")),
        tabPanel("Report", verbatimTextOutput("report"))
      )
    )
  )
)

# Server
server <- function(input, output) {

  dataInput <- reactive({
    if (input$input_type == "manual") {
      vals1 <- as.numeric(unlist(strsplit(input$city1, ",")))
      vals2 <- as.numeric(unlist(strsplit(input$city2, ",")))

      if (length(vals1) != 12 || length(vals2) != 12) {
        stop("Please enter exactly 12 values for each city.")
      }

      vals1[is.na(vals1)] <- mean(vals1, na.rm = TRUE)
      vals2[is.na(vals2)] <- mean(vals2, na.rm = TRUE)

      df <- data.frame(City1 = vals1, City2 = vals2)

    } else if (input$input_type == "file") {
      req(input$datafile)
      ext <- tools::file_ext(input$datafile$name)

      if (ext == "csv") {
        df <- read.csv(input$datafile$datapath)
      } else if (ext %in% c("xlsx", "xls")) {
        df <- read_excel(input$datafile$datapath)
      } else {
        stop("Unsupported file type")
```

```
    }

    if (ncol(df) < 2 || nrow(df) < 12) {
      stop("File must contain at least 12 rows and 2 columns.")
    }

    df <- df[1:12, 1:2]
    df <- as.data.frame(lapply(df, function(x) {
      vals <- as.numeric(x)
      vals[is.na(vals)] <- mean(vals, na.rm = TRUE)
      return(vals)
    }))

    names(df) <- c("City1", "City2")
  }

  return(df)
})

observeEvent(input$analyze, {
  output$barPlot <- renderPlot({
    df <- dataInput()
    barplot(
      rbind(df$City1, df$City2),
      beside = TRUE,
      col = c("skyblue", "lightgreen"),
      names.arg = months,
      legend.text = c("City 1", "City 2"),
      main = "Monthly Rainfall Comparison",
      ylab = "Rainfall (mm)"
    )
  })

  output$linePlot <- renderPlot({
    df <- dataInput()
    plot(df$City1, type = "o", col = "blue", ylim = range(df),
         xaxt = "n", xlab = "Month", ylab = "Rainfall (mm)",
         main = "Rainfall Trend - City 1 vs City 2")
    lines(df$City2, type = "o", col = "green")
```

```
    axis(1, at = 1:12, labels = months)
    legend("topright", legend = c("City 1", "City 2"), col = c("blue", "green"),
lty = 1, pch = 1)
  })

  output$report <- renderText({
    df <- dataInput()
    avg1 <- mean(df$City1)
    avg2 <- mean(df$City2)
    result <- if (avg1 > avg2) {
      "City 1 has higher average rainfall."
    } else if (avg2 > avg1) {
      "City 2 has higher average rainfall."
    } else {
      "Both cities have equal average rainfall."
    }

    paste0(
      "--- Average Rainfall Report ---\n",
      "City 1 Average: ", round(avg1, 2), " mm\n",
      "City 2 Average: ", round(avg2, 2), " mm\n\n",
      result
    )
  })
 })
}

# Run the app
shinyApp(ui = ui, server = server)
```

APPENDIX B – SCREENSHOTS





REFERENCES

- Wickham, H., & Grolemund, G. (2016). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. O'Reilly Media, Inc.

- Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., ... & Yamaguchi, M. (2021). *Shiny: Web Application Framework for R*. R package version 1.7.1. https://shiny.rstudio.com

- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

- Firke, S. (2021). *janitor: Simple Tools for Examining and Cleaning Dirty Data*. R package version 2.1.0. https://cran.r-project.org/package=janitor

- OpenAI. (2025). *ChatGPT* [Large language model]. Accessed May 2025 from https://chat.openai.com