College of Professional Studies

Northeastern University San Jose

MPS Analytics

Course: ALY6070: Communication and Visualization for Data Analytics

Assignment:

Assignment 3 — Individual RShiny Application

Submitted on:

March 22, 2023

Submitted to:

Prof: VENKATA DUVVURI

Submitted by:

NIKSHITA RANGANATHAN

ABSTRACT

R Shiny is a web application framework for the R programming language that allows users to build interactive web applications and dashboards. It is an open-source package maintained by RStudio, and it is built on top of the shiny package in R.

R Shiny applications consist of two main components: the user interface (UI) and the server logic. The UI is defined using R code that specifies the layout of the application and the widgets that users can interact with. The server logic is also defined using R code, and it controls how the application responds to user input and generates output.

Features:

- Interactivity: R Shiny allows users to create interactive web applications where users can change parameters and immediately see the results. This makes it easy to explore data and communicate results to others.
- Flexibility: R Shiny can be used to create a wide range of applications, from basic dashboards and predictive models to intricate data visualizations.
- Reproducibility: R Shiny applications are built in R, which means that they are fully reproducible. This makes it easy to share code and results with others.
- Customization: R Shiny allows users to customize the look and feel of their applications with CSS and HTML, giving them full control over the user experience.
- Open source: R Shiny is an open-source project, which means that it is free to use and can be modified and extended by anyone.

Advantages:

- Easy to learn: R Shiny has a relatively low learning curve, especially for developers who are already familiar with R.
- Rapid application development: R Shiny allows developers to quickly create interactive applications without having to learn a new programming language or web framework.
- Scalability: R Shiny applications can be deployed to a variety of environments, from local machines to cloud servers, making it easy to scale applications to meet growing demands.
- Community support: There are many resources available for getting started and troubleshooting problems with R Shiny because it has a large and active community of users and developers.

Uses:

- Data visualization: R Shiny is often used to create interactive data visualizations that allow users to explore and understand data.
- Dashboards: R Shiny can be used to create interactive dashboards that provide real-time information and allow users to drill down into details.
- Predictive modeling: R Shiny can be used to create interactive predictive models that allow users to test different scenarios and see the impact on outcomes.
- Teaching and learning: R Shiny can be used in the classroom to create interactive exercises and demonstrations that help students understand statistical concepts and data analysis techniques.

INTRODUCTION

The Olympic Games, as a whole, is a global sporting event that brings together athletes from around the world to compete in various sports. The Summer Olympics and Winter Olympics are held every four years, respectively. The Olympics have a long history and have been an important cultural and political event since the first modern Olympic Games in 1896. It is also an important platform for promoting international cooperation and understanding, with athletes from all over the world coming together to compete in a spirit of sportsmanship and fair play. The Games have been used to promote social and political causes, such as the inclusion of women's sports and the fight against apartheid in South Africa.

The purpose of this project is to gain a better understanding of the factors that contribute to success in the Olympic Games. This information can be used to inform training and selection strategies for future Olympic teams and can also provide insights into the broader social and cultural contexts of the Games.

About this Dataset: The "120 Years of Olympic History: Athletes and Results" dataset on Kaggle is a rich and comprehensive dataset that contains information about every Olympic athlete and medalist since the first modern Olympic Games in Athens in 1896. The dataset includes a wide range of variables, such as athlete demographics, sport and event details, and medal outcomes.

The dataset contains over 270,000 records and 15 variables, including:

- ID - Unique identifier for each athlete
- Name - Athlete's name
- Sex - Gender of the athlete
- Age - Age of the athlete at the time of competition
- Height - Height of the athlete in centimeters
- Weight - Weight of the athlete in kilograms
- Team - Name of the athlete's team
- NOC - National Olympic Committee code
- Games - Year and season of the Olympic Games
- Year - Year of the Olympic Games
- Season - Summer or Winter Olympic Games
- City - Host city of the Olympic Games
- Sport - Sport in which the athlete competed
- Event - Event in which the athlete competed
- Medal - Medal won by the athlete (gold, silver, or bronze)

Research Questions :

1. What is the relationship between weight and height for Olympic athletes, and how has this relationship changed over time? Is there a correlation between height or weight and success in certain sports?

2. How has the number of athletes in the Olympic Games varied over time for each gender, and what factors may have contributed to these trends?
3. Are there any trends or patterns in the types of sports that are included in the Summer and Winter Olympic Games?
4. How have the teams represented in the Olympic Games evolved over the years, and what factors have influenced these changes?
5. Is there a correlation between age and gender among Olympic athletes? Are there any sports where age or gender plays a significant role in performance?
6. Which countries have won the most medals in the Olympic Games?

ANALYSIS

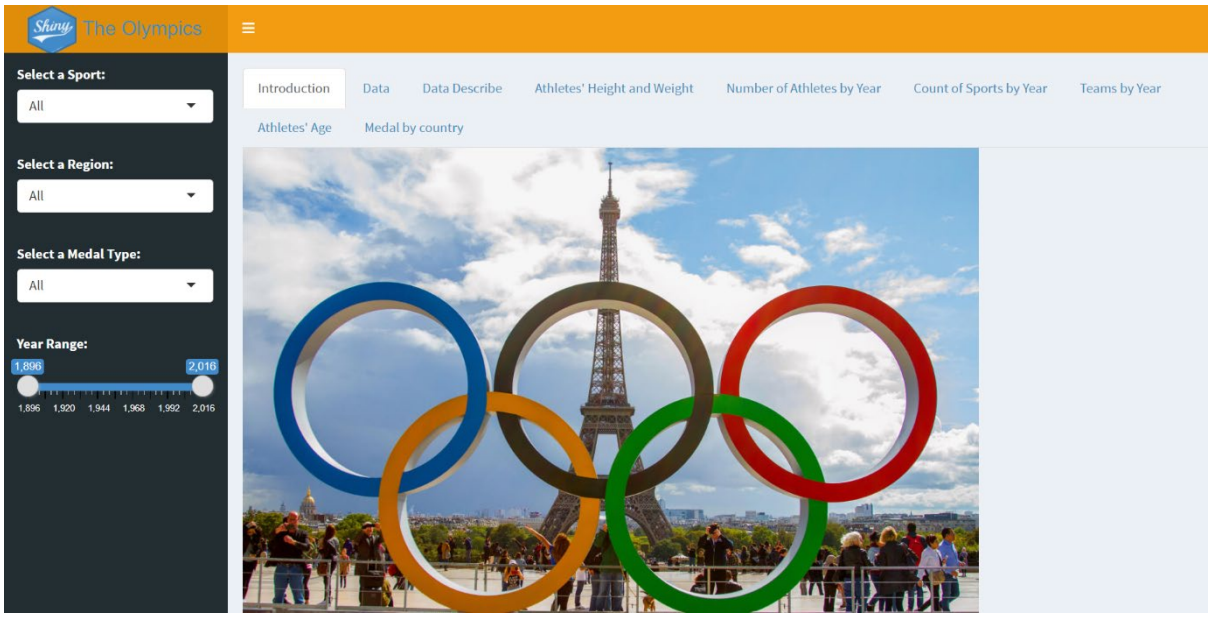

Figure 1 – Dashboard

This is a Shiny app that provides visualizations and data summaries for a dataset of Olympic athletes and events. The dashboard allows the user to filter the data based on sport, region, medal type, and year range. The dashboard contains several tabs that display different data visualizations. The app uses several libraries including shiny, shinydashboard, leaflet, dplyr, ggplot2, DT, summarytools, htmlTable, ochRe, htmlwidgets, and tidyr.

Graph 1:

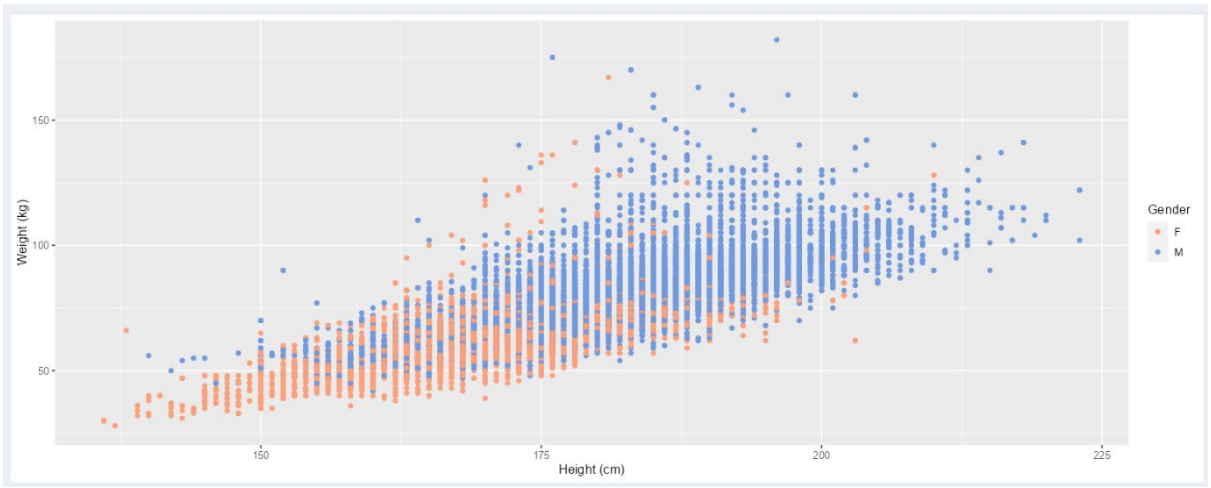

Figure 2 – Scatterplot

I wanted to analyze the connection between the weight and height of the athletes. So I used a scatterplot. A scatter plot is a type of graph that shows the relationship between two variables. The plot shows that as height increases, weight tends to increase as well. This relationship is

described as positive and linear. The plot also shows that, on average, men are taller and heavier than women. However, there are some exceptions to this general pattern, which may be related to differences in sports and other factors.

The positive relationship between height and weight can be explained by a number of factors, including genetics, nutrition, and physical activity. In addition, physical activity can play a role in the relationship between height and weight. Taller individuals may have a greater amount of lean body mass, which includes muscle and may therefore require more energy to maintain their body weight. This can lead to increased calorie intake and higher overall body weight.

Graph 2:

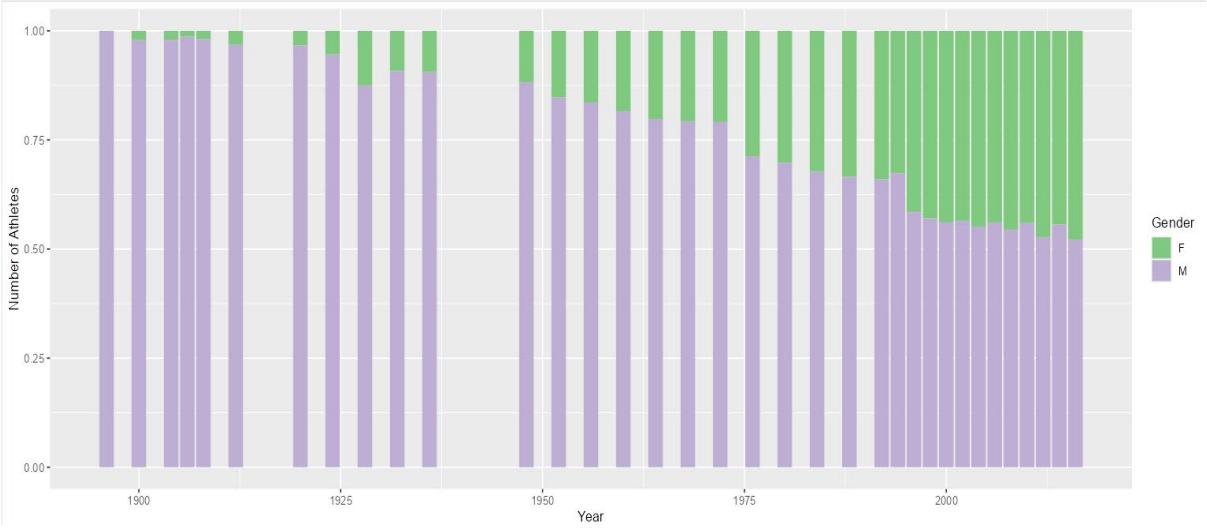

Figure 3 – Percent stacked barchart

Women's participation in the Summer Olympics has increased significantly over the years, but there still remains a gender gap in terms of overall representation.

Prior to the 1980s, female athletes were not well-represented in the Summer Olympics. This was partly due to the fact that many countries did not send female athletes to compete, and partly due to the fact that some sports did not have events for women.

However, starting in the 1980s, the committee made a conscious effort to promote gender equality in the Olympic Games, and more sports were added to the program that included events for women. As a result, the number of female athletes participating in the Summer Olympics has been steadily increasing, with the gap between male and female athletes closing over time.

In fact, at the 2016 Summer Olympics in Rio de Janeiro, Brazil, there were nearly as many female athletes as male athletes, with women making up 45% of all participants. This can be observed in the 100% stacked bar graph (Figure 3). However, there is still work to be done to achieve full gender equality in the Olympics, as some sports still do not offer equal

opportunities for women, and some countries continue to send fewer female athletes to compete.

Graph 3:

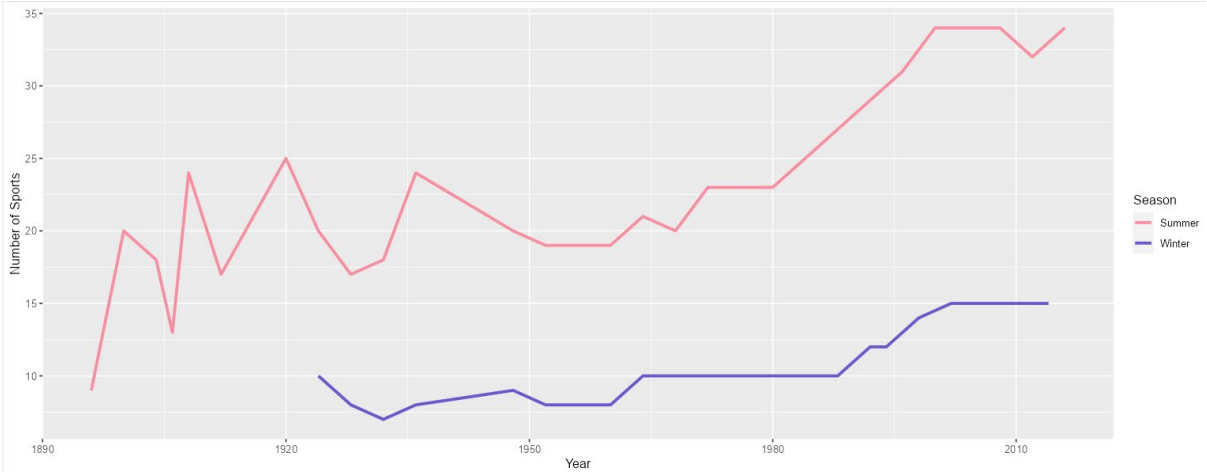

Figure 4 – Line chart

The line chart shows that the number of sports in the Summer Olympics has consistently been higher than the number of sports in the Winter Olympics. This is likely due to the fact that the Summer Olympics have been around for much longer than the Winter Olympics and have had more time to add new sports over the years. The Summer Olympics offer more than 40 different sports and hundreds of events, while the Winter Olympics only cover around 15 sports and fewer events.

Furthermore, the line chart shows that the number of sports in both the Summer and Winter Olympics has increased over time. This is due to the addition of new sports and events that reflect changing interests and global trends. For example, skateboarding and surfing are set to make their debut at the 2021 Tokyo Olympics, reflecting the growing popularity of these sports around the world.

Graph 4:

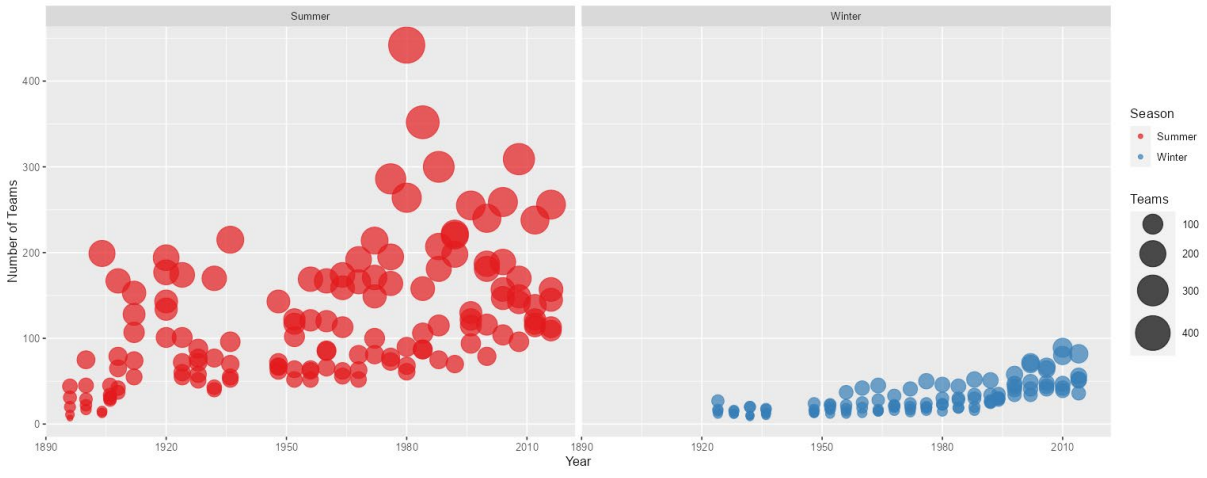

Figure 5 – Bubble chart

On the other hand, the bubble chart shows that many more teams participate in the Summer Olympics than in the Winter Olympics. This is likely due to the wider range of sports and events offered in the Summer Olympics, as well as the popularity of those sports around the world. Sports like athletics, gymnastics, and basketball have large followings and are played in many countries, while winter sports like skiing, snowboarding, and ice hockey are often limited to countries with colder climates.

Additionally, the cost of competing in the Winter Olympics can be prohibitively high for many countries, as winter sports often require expensive equipment and facilities. This may limit the number of countries that are able to participate in the Winter Olympics, especially those from warmer climates.

Graph 5 :

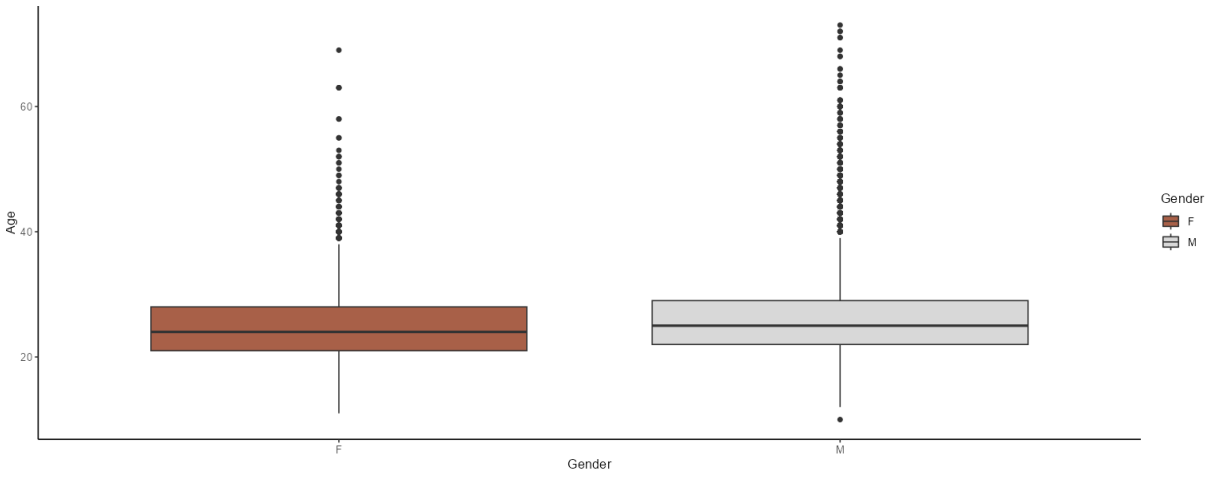

Figure 6 – Boxplot

According to the dataset, it appears that the age of male athletes participating in the Olympics tends to be slightly higher than the age of female athletes.

It's worth noting, however, that the age difference between male and female Olympians is not always significant, and there are many examples of male and female athletes who have competed at similar ages and achieved great success in their respective sports.

Graph 6 :

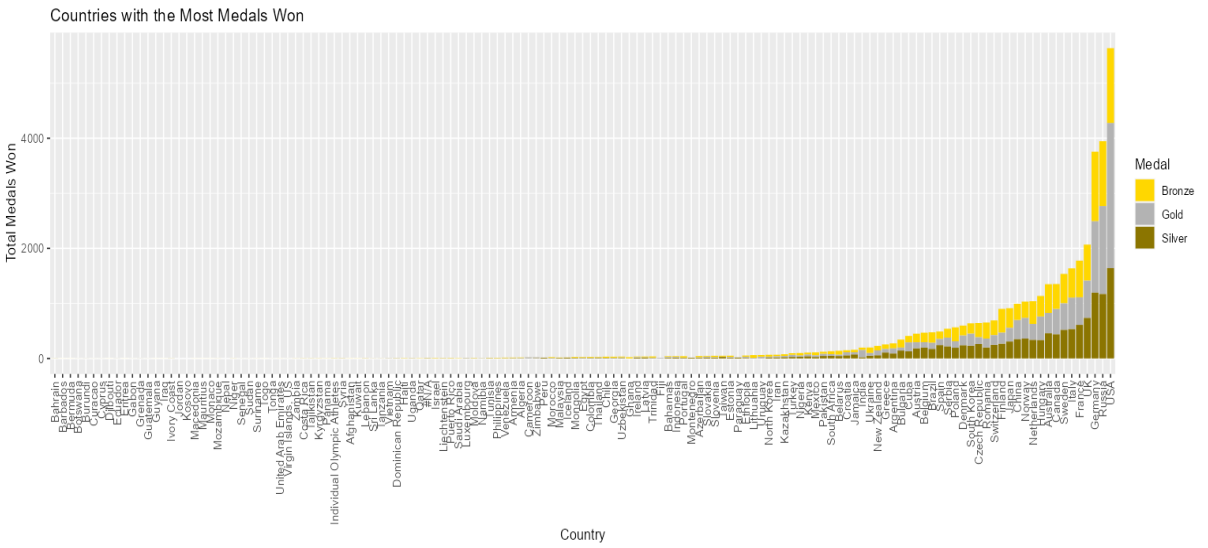

Figure 6 – Stacked bar graph

According to the above graph, the country with the most medals is the United States, with a total of 5637 medals. The second country with the most number of medals is Russia, with 2470 medals, and the third is Germany with 2159 medals.

There could be several reasons why the United States has won the most medals. One reason could be the size and diversity of the country, which provides a larger pool of potential athletes to choose from. Additionally, the United States has a strong tradition of supporting and investing in sports programs at the high school and college levels, which helps to identify and develop talented athletes. The country also has a well-developed system of professional sports leagues, which can serve as a pathway for athletes to reach the highest levels of international competition.

Other factors that may contribute to the United States' success in the Olympics could include cultural attitudes towards sports and physical activity, access to training facilities and resources, and the availability of quality coaching and support staff.

REFERENCES

P. (n.d.). *basic-dataset/netflix_titles.csv at master · prasertcbs/basic-dataset*. GitHub. https://github.com/prasertcbs/basic-dataset/blob/master/netflix_titles.csv

Resources. (n.d.). Tableau. https://www.tableau.com/resources

APPENDIX

```
library(shiny)
library(shinydashboard)
library(leaflet)
library(dplyr)
library(ggplot2)
library(DT)
library(summarytools)
library(htmlTable)
library(ochRe)
library(htmlwidgets)
library(tidyr)

# Load the dataset
df <- read.csv("athlete_events.csv")

# Adding image
title<-tags$a(tags$img(src="r.png",height='50',width='50'),"The Olympics")
# Define the UI
ui <- dashboardPage(skin="yellow",
  dashboardHeader(title = title),
                  dashboardSidebar(
                  selectInput("sport", "Select a Sport:",
                              choices = c("All", unique(df$Sport)),
                              selected = "All"),
                  selectInput("region", "Select a Region:",
                              choices = c("All", unique(df$region)),
                              selected = "All"),
                  selectInput("medal", "Select a Medal Type:",
                              choices = c("All", "Gold", "Silver", "Bronze"),
                              selected = "All"),
                  sliderInput(
                    "Year", label = "Year Range:",
                    min = 1896, value = c(1896,2016) , max = 2016
                  )
                ),
                dashboardBody (

                  tabsetPanel(

                    # Introduction page
```

```
            tabPanel(
              "Introduction",
        tags$img(src="img.jpg",height="500px",width="auto"),h1("The Olympic Games R
        Shiny Dashboard"),
              p("The 120 Years of Olympic History dataset is a comprehensive collection of
        athlete and event information from the modern Olympic Games dating back to
        1896."),
              p("It includes data on over 270,000 athletes who have participated in more
        than 40,000 events across more than 150 countries. This rich and diverse dataset
        provides a wealth of opportunities for data analysis, visualization, and exploration."),
              p("Use the sidebar to filter the data  and navigate through the tabs to view
        different data visualizations.")
            ),
            tabPanel("Data",DT::dataTableOutput("table")),
            tabPanel("Data Describe",htmlOutput("summary")),
            tabPanel("Athletes' Height and Weight", plotOutput("height_weight")),
            tabPanel("Number of Athletes by Year", plotOutput("athlete_barplot")),
            tabPanel("Count of Sports by Year", plotOutput("sports_count")),
            tabPanel("Teams by Year", plotOutput("teams_count")),
            tabPanel("Athletes' Age", plotOutput("age_boxplot")),
            tabPanel("Medal by country", plotOutput("medals_plot"))
          )
        )
      )

# Define the server
server <- function(input, output) {

  # Filter the dataset based on user input
  filtered_df <- reactive({
    df %>%
      filter(if (input$sport == "All") TRUE else Sport == input$sport) %>%
      filter(if (input$region == "All") TRUE else region == input$region) %>%
      filter(if (input$medal == "All") TRUE else Medal == input$medal)%>%
      filter(Year >= input$Year[1] & Year <= input$Year[2])%>%
      filter(!is.na(Medal))
  })
  output$table<-DT::renderDataTable({
    datatable(subset(df,select = -c(Height,Weight,Team,NOC,City,Games)),
              options = list(pageLength= 30))
  })

  # creating the summary for selected variables
  output$summary<-renderUI({
    print(dfSummary(subset(df,select =c(Sex,Age,Height,Weight,Year,Season,Sport,Medal))
    ),method = 'render',headings = FALSE, bootstracp.css = FALSE)
  })
  # Render the athletes' height and weight plot
  palette <- c("lightsalmon", "#709AE1FF")
  output$height_weight <- renderPlot({
    filtered_df() %>%
      select(Name, Sex, Age, Height, Weight) %>%
      ggplot(aes(x = Height, y = Weight, color = Sex)) +
      geom_point() +
      scale_color_manual(values = palette) +
      labs(x = "Height (cm)", y = "Weight (kg)", color = "Gender")
  })
```

```
# Render the bar graph of number of athletes by year and gender
output$athlete_barplot <- renderPlot({
  filtered_df() %>%
    group_by(Year, Sex) %>%
    summarise(n = n()) %>%
    ggplot(aes(x = Year, y = n, fill = Sex)) +
    geom_bar(stat = "identity", position = "fill") +
    scale_fill_brewer(palette = "Accent") +
    labs(x = "Year", y = "Number of Athletes", fill = "Gender")
})


# Render the count of sports by year plot
output$sports_count <- renderPlot({
  filtered_df() %>%
    group_by(Year, Season) %>%
    summarize(count = n_distinct(Sport)) %>%
    ggplot(aes(x = Year, y = count, color = Season)) +
    geom_line(size = 1.2) +
    labs(x = "Year", y = "Number of Sports", color = "Season") +
    scale_color_manual(values = c("#F88fa2","#6A5ACD"))
})

# Render the teams by year plot
output$teams_count <- renderPlot({
  filtered_df() %>%
    group_by(Year, Season, Team) %>%
    summarize(count = n()) %>%
    filter(rank(desc(count)) <= 5) %>%
    ggplot(aes(x = Year, y = count, size = count, color = Season)) +
    geom_point(alpha = 0.7) + scale_size(range = c(2, 14))+
    labs(x = "Year", y = "Number of Teams", color = "Season", size = "Teams") +
    scale_color_brewer(palette = "Set1") +
    facet_wrap(~Season)
})
# Render the age boxplot
output$age_boxplot <- renderPlot({
  filtered_df() %>%
    select(Sex, Age) %>%
    ggplot(aes(x = Sex, y = Age, fill = Sex)) +
    geom_boxplot() +
    scale_fill_ochre(palette="winmar") +
    labs(x = "Gender", y = "Age", fill = "Gender") +
    theme_classic()
})
# Calculate total medals won by each country
medals_by_country <- df %>%
  filter(!is.na(Medal)) %>%
  filter(!is.na(region)) %>%
  group_by(region,Medal) %>%
  summarize(Total_Medals = n()) %>%
  arrange(desc(Total_Medals)) %>%
  top_n(10)

# Create the plot
output$medals_plot <- renderPlot({
  ggplot(medals_by_country, aes(x = reorder(region, Total_Medals), y =
         Total_Medals,fill=Medal)) +
```

```
    geom_bar(stat = "identity") +
    xlab("Country") +
    ylab("Total Medals Won") +
    ggtitle("Countries with the Most Medals Won")+
    theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))+
    scale_fill_manual(values = c("gold","gray70","gold4"))
  })

}

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