Olympics and Social Indicators

Jatavata Balakoti, Sujal Yadav, Vinay Chavan, Utkarsh Kesharwani

Introduction

This report describes a basic RShiny application developed to explore data on the Olympics from 1896 to 2016, focusing on women's participation and achievements. The app features a range of visualizations, including bar graphs, scatter plots, a line chart, heat map, and pie chart, which allow users to examine trends in medals, participation by gender, and other related factors.

Through a drop-down menu, users can filter views by continent and medal type or explore data comparisons by GDP, life expectancy, and sex ratio. This application offers a simple, interactive overview of historical data, highlighting gender-based trends and regional differences in Olympic success.

Explaining the Data

What were the variables?

The following datasets were used in this analysis:

1. **Olympics Data (`olympics`):**
    - **Columns:** Athlete ID, Age, Sex, Height, Weight, Team, NOC (National Olympic Committee), Year, Season, Sport, and Medal Type.
    - This dataset contains detailed records of Olympic athletes, including demographic details (age, sex, height, weight) and participation in specific Olympic events. It includes information on which sport they competed in, their Olympic year, and whether they won a medal.
2. **GDP Data (`gdp_data`):**
    - **Columns:** Country, Year (2000–2023), Nominal GDP.

- This dataset tracks the GDP of various countries from 2000 to 2023, providing insight into the economic status of the countries during the period under study.

3. **Life Expectancy Data (`life_expectancy_data`):**
   - **Columns:** Rank, Country, Life Expectancy (for both sexes), Female Life Expectancy, Male Life Expectancy.
   - This dataset provides life expectancy statistics for countries, segmented by sex, giving insight into the general health and well-being of populations.
4. **Continent Data (`continent_data`):**
   - **Columns:** Country, Continent.
   - This dataset provides the continent each country belongs to, which is useful for analyzing trends and comparing performance across different regions.

Where was it scraped from?

The datasets for this analysis were obtained through a combination of web scraping and direct downloads, as outlined below:

1. GDP Data: The **GDP data** was scraped from the Wikipedia page **"List of countries by past and projected GDP (nominal)"** using the `rvest` library in R. Relevant tables (4, 5, and 6) were extracted, cleaned by removing commas, and converted to numeric format. The cleaned data was saved as `gdp.csv`.
2. Life Expectancy Data: The **Life Expectancy data** was scraped from the Worldometers website using `rvest`. The relevant table was extracted and saved as `life_expectancy.csv`.
3. Sex Ratio Data: The **Sex Ratio data** was scraped from the Wikipedia page **"List of sovereign states by sex ratio"** using `rvest`. The data was cleaned and saved as a CSV file.
4. Olympics Data: The **Olympics data** was directly downloaded from a GitHub repository and loaded into R for analysis. The dataset contains Olympic medals and athlete information.
5. Continent Data: The **Continent data** was manually extracted from the Statistics Times website and saved as a CSV file.

Potential Biases in the Data

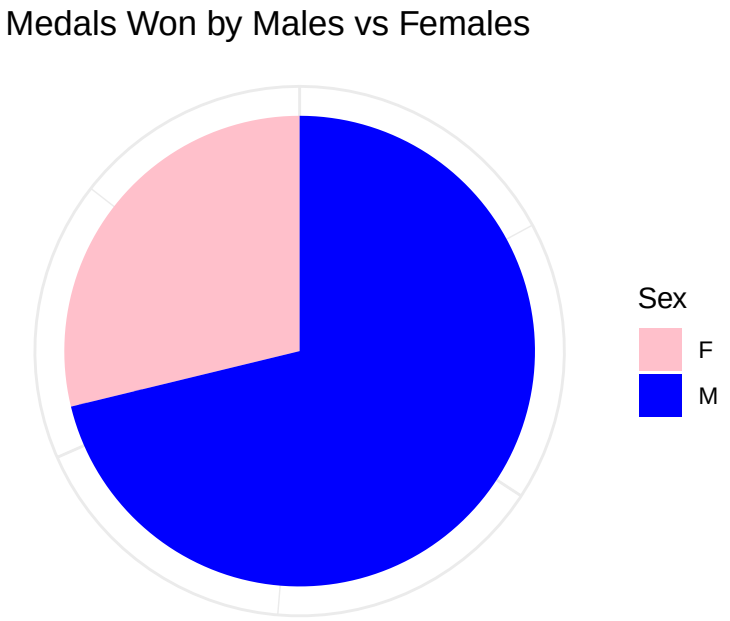
Several factors may influence the analysis and interpretation of the data in this application:

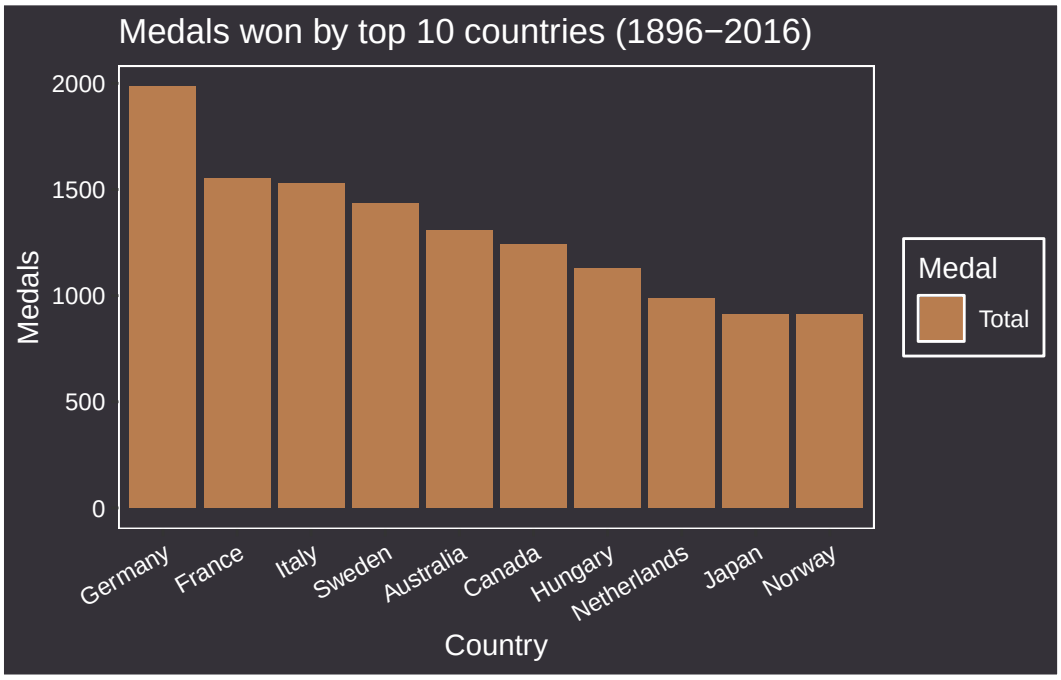
- **Historical Gender Inequality**: Early Olympic events had limited or no participation by women, skewing long-term analyses and underrepresenting female achievements, which may affect trends over time.
- **Economic and Geopolitical Disparities**: Wealthier countries tend to have more resources for athlete development, leading to an overrepresentation in medal counts, which may suggest a correlation between GDP and Olympic success that is influenced by funding availability rather than athletic performance.
- **Regional Representation**: Socioeconomic conditions in developing nations may limit athlete participation, leading to underrepresentation in global comparisons and potentially distorting continent-level analyses.

Interesting Questions to Ask

- How has female participation in the Olympics changed from 1896 to 2016, and which continents or countries have contributed most to this growth?
- Which countries and continents have won the most medals in women's events, and how has this changed over time?
- What are the top sports in which female athletes from different countries or continents have excelled, and are there regional patterns in sport specialization?
- Is there a relationship between a country's GDP and the number of medals won by female athletes? How does this vary by continent?
- How does female life expectancy correlate with the number of Olympic medals a country's female athletes have won?
- What trends exist in the gender ratio of Olympic athletes over time, and when did key shifts toward gender balance occur?
- In 2016, what was the distribution of male and female athletes by country, and how does this compare across continents?

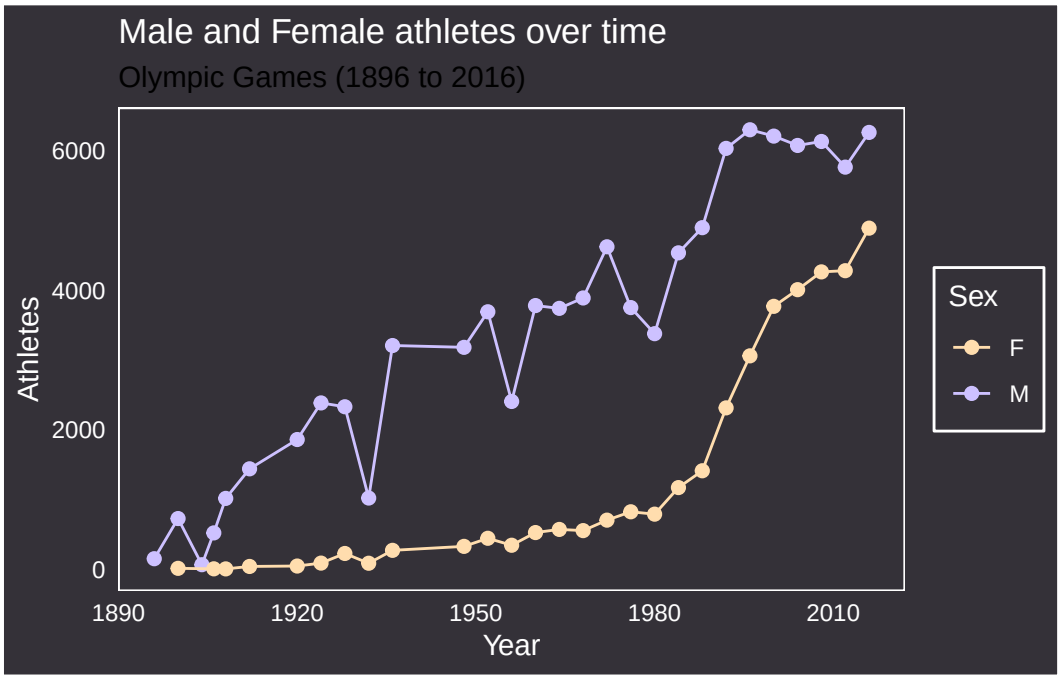
Important Visualisation



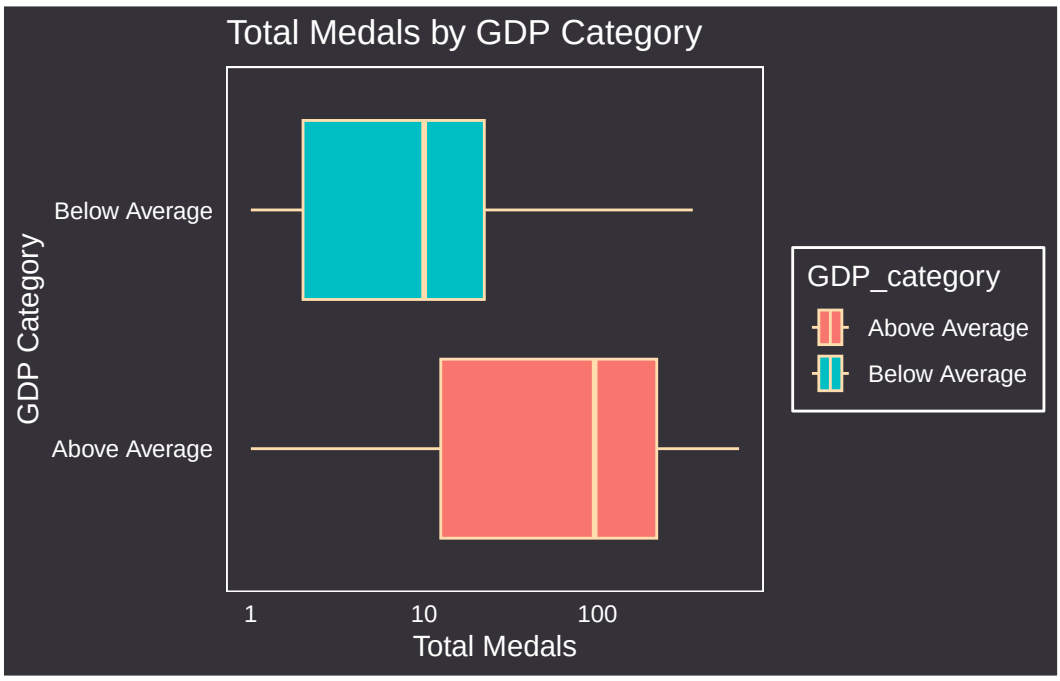


The top ten countries with the most Olympic medals from 1896 to 2016 are predominantly from developed, wealthier nations, particularly Europe. This reflects the impact of better

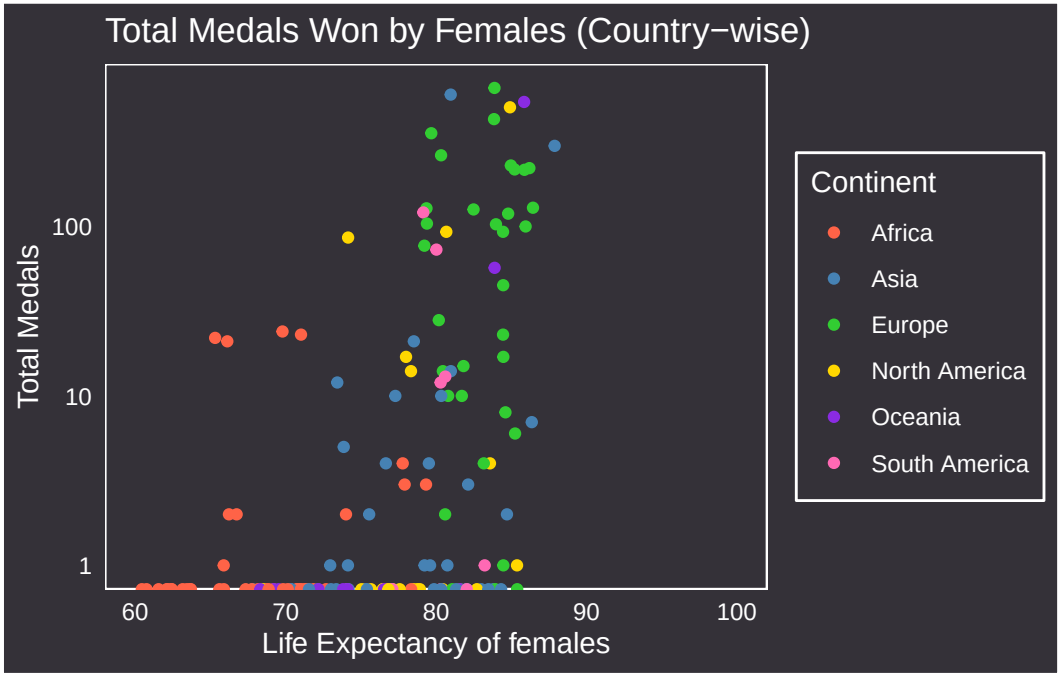
funding, infrastructure, and access to training on Olympic success.



Male athlete participation has historically been higher than female participation, with declines during global wars and political conflicts. However, since the 1980s, female participation has surged due to expanded events, increased gender equality awareness, and societal shifts. These changes were supported by women's sports organizations and legal frameworks promoting gender equality in sports.



The box plot shows that countries with below-average GDP tend to win fewer medals, with their median medal count lower than the first quartile of above-average GDP countries. This highlights the advantage of wealthier nations in Olympic performance.



The graph shows a correlation between female life expectancy and Olympic success. European countries, with higher life expectancy, win more medals, while African countries, with lower life expectancy, contribute fewer medals, suggesting links between healthcare, economic development, and Olympic performance.

Conclusion

This report presents an analysis of Olympic data from 1896 to 2016, focusing on female participation and medal achievements. The findings highlight significant trends, such as the steady increase in female athlete participation since the 1980s and the dominance of wealthier, developed countries in winning Olympic medals. The correlation between GDP and Olympic success, as well as the influence of female life expectancy on medal outcomes, underscores the importance of economic and social factors in shaping athletic performance. Despite the progress in gender equality, challenges such as historical gender disparities and regional economic inequalities continue to impact the representation and success of female athletes. Overall, this application provides valuable insights into the evolution of women's participation in the Olympics and the factors influencing their success on the global stage.

References

Life Expectancy Data : https://www.worldometers.info/demographics/life-expectancy/

Sex Ratio Data : https://en.wikipedia.org/wiki/List_of_sovereign_states_by_sex_ratio

Olympics Data : https://raw.githubusercontent.com/VinayChavan2006/R-project/main/Data/datasets/olym

Continent Data : https://statisticstimes.com/geography/countries-by-continents.php

GDP Data : https://en.wikipedia.org/wiki/List_of_countries_by_past_and_projected_GDP_(nominal)