Project 3

25982672

How do trends in Olympic participation, performance, and medal distribution vary by country, sport, gender, and season over time, and what influence do hosting and seasonal factors have on these outcomes?

Introduction

The Olympics are one of the most prestigious and most celebrated sporting events in the world, providing an opportunity for countries and athletes to showcase their theathletic prowess across multiple disciplines. The olympics have existed for centuries , they serve as a lens to not just view the competition between different countries they also serve as a platform to observe various patterns of participation , performance disparities, and the influence of external factors such as hosting nations and seasonal variations. This project will explore those observations in order to find out how perfomance nad medal distribution vary by country,sport, gender and season overtime . This will be done by analyzing data across multiple Olympic games to investigate how the number of medals won by top countries has changed, which sports have historically contributed the most to medal counts, and whether there is a discernible "host country advantage. Also it will explore the relationship between the season in which a sport is played and the amount of medals awarded.

The following key questions will be explored::

1. How have the number of medals won by top-performing countries changed over time?
2. Which sports have historically contributed the most to the medal tallies of top-performing countries?
3. Is there a measurable "host country advantage" in Olympic medal counts?
4. How do Summer and Winter Olympics differ in terms of medal distribution and performance trends?

5. How has the gender distribution of Olympic athletes evolved over time, and how has this impacted overall participation?

Through a comprehensive analysis of these themes, we aim to shed light on the dynamics of the Olympic Games, enhancing our understanding of how historical, cultural, and environmental factors influence athletic outcomes. The insights gained from this project not only contribute to the discourse on sports and global competition but also highlight areas for potential improvement and equity in athletic representation.

Data Preparation

To conduct this analysis this project s data from multiple Olympic data sets which are `olympic_medals.csv`, `olympic_hosts.csv`, `olympic_athletes.csv`, and `olympic_results.csv`, which have information covering athlete information, medal outcomes, event details, and host country data.

Before the analysis of the data of the could begin key data cleaning and preparation steps were included to insure that the data analysis were as accurate as possible.

These steps include :

Duplicate Removal

Duplicate records were identified and removed across all datasets. This step ensures that each row represents a unique record, preventing inflation of any one country's or athlete's statistics.

Handling missing data

1. Rows where critical fields such as `country_name` and `medal_type` were missing were removed, as these are essential for the analysis of country-wise and medal-wise performance.
2. For non-essential fields such as `athlete_url`, missing data was left intact because this information was not necessary for the current analysis. However, the presence of missing data in certain columns (e.g., athlete birth years) could lead to minor biases, especially when analyzing trends like age distribution.
3. **Impact of Missing Data**: We acknowledge that missing data, especially for athletes' details, may skew the results in sections like athlete age analysis. However, given that the percentage of missing critical data was low, we determined that removing rows with missing key fields would not significantly impact the overall trends in medal distribution or gender analysis.

Standardization of Categorical Data:

- The project standardizes the `event_gender` and `medal_type` columns to ensure consistency in the analysis. For instance, `event_gender` values were capitalized (`Male`, `Female`), and all `medal_type` values were converted to uppercase (`GOLD`, `SILVER`, `BRONZE`).
- Additionally, the country names were standardized by merging variants like "Soviet Union" and "Russia" into a single entity, "Russia." This decision was made to maintain consistency in the analysis since splitting these entities would make it difficult to accurately trace performance trends over time.

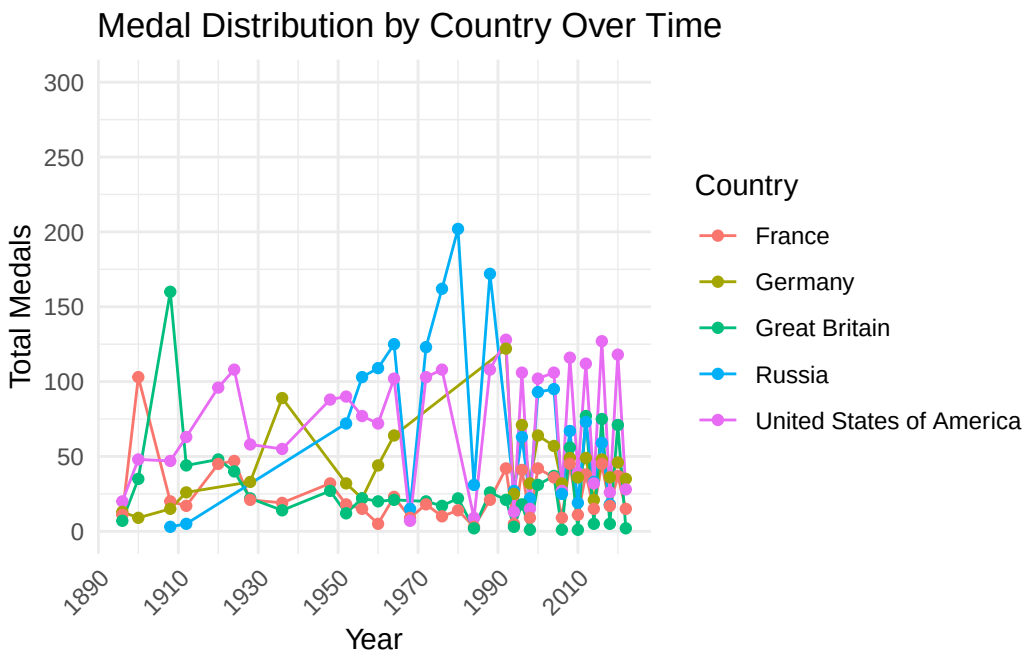
Merging Datasets:

The cleaned datasets were merged to create a comprehensive table that included medals, athletes, and host country details.

Exploratory Data Analysis and Reporting

How have the number of medals won by top countries changed over time?

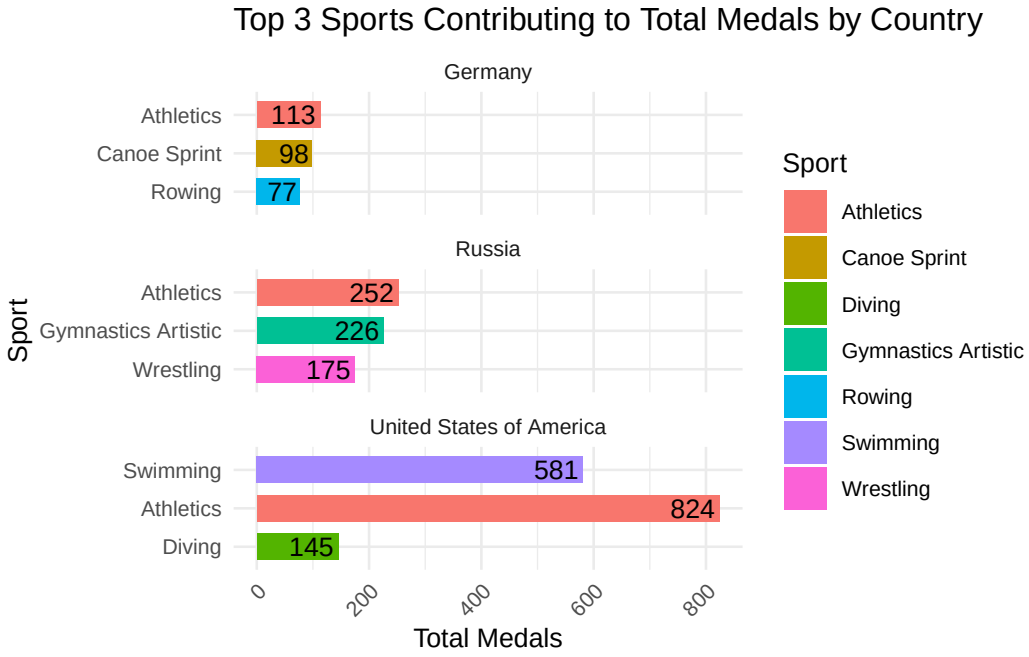
The first research question relates to the change in the number of medals won by top-performing countries over the years. Traditionally, few countries like the United States and Russia, have monopolized the Olympics, especially in the Summer Games. A graph visualizing medal count overtime for the top 5 medal-earning countries, notably USA, Russia, Germany and Great Britain, reflects different trends.



Interpretation: The USA although not consistent has led the medal tally across the Olympic Games, especially during the era between 1990 to 2020. Its dominance in Summer sports, particularly swimming and athletics, has allowed it to maintain a clear lead. Russia (and its Soviet-era predecessor) showed similar dominance during the Cold War, though its medal count declined in the post-Soviet era, reflecting political and economic challenges. This analysis ties into the broader geopolitical context of the 20th century, where Olympic success often served as a proxy for global influence. Political factors, including World War II, and the Cold War, caused visible fluctuations in these trends, particularly for countries like Russia and Germany.

Which sports have historically contributed the most medals to each of the top medal earning countries?

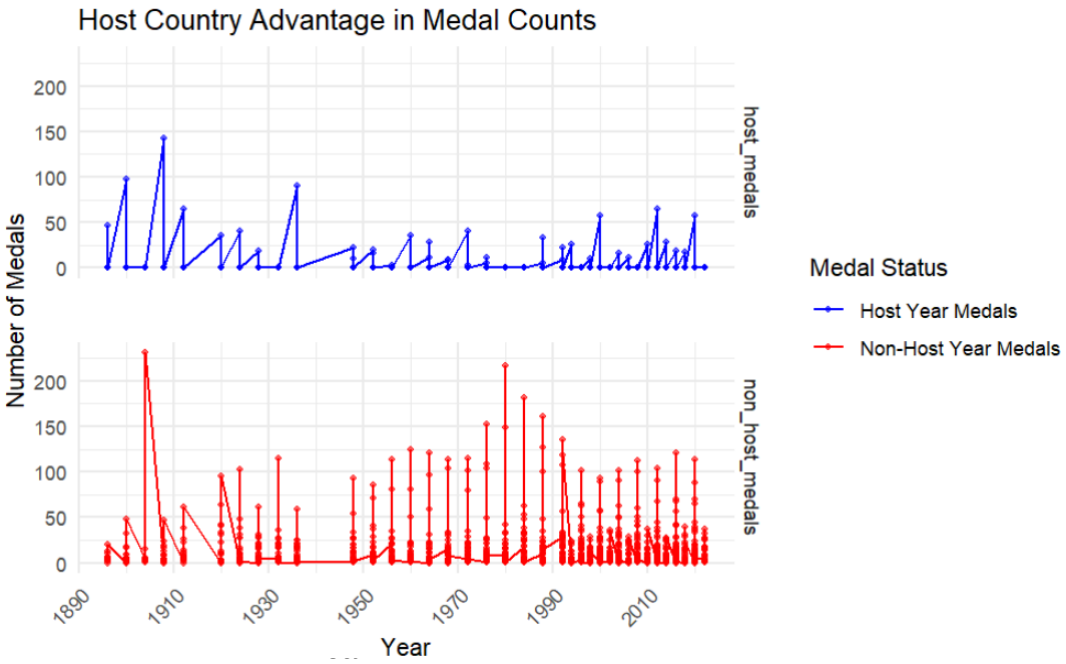
The following analysis delves into which sports have historically contributed the most medals to top-performing countries. Countries tend to dominate specific disciplines, often influenced by cultural focus, institutional investment, and environmental factors.



Interpretation: For the USA, athletics (track and field), swimming and diving have contributed the giants share of medals, affirming the country's historical strengths in these sports. Russia, meanwhile, has consistently excelled in gymnastics, wrestling, and Athletics this reflects the country's focus on both Summer on these sports. This analysis highlights the strategic decisions nations make in focusing on particular sports. By investing in disciplines where they have historical success or competitive advantages, countries can maximize their overall medal counts.

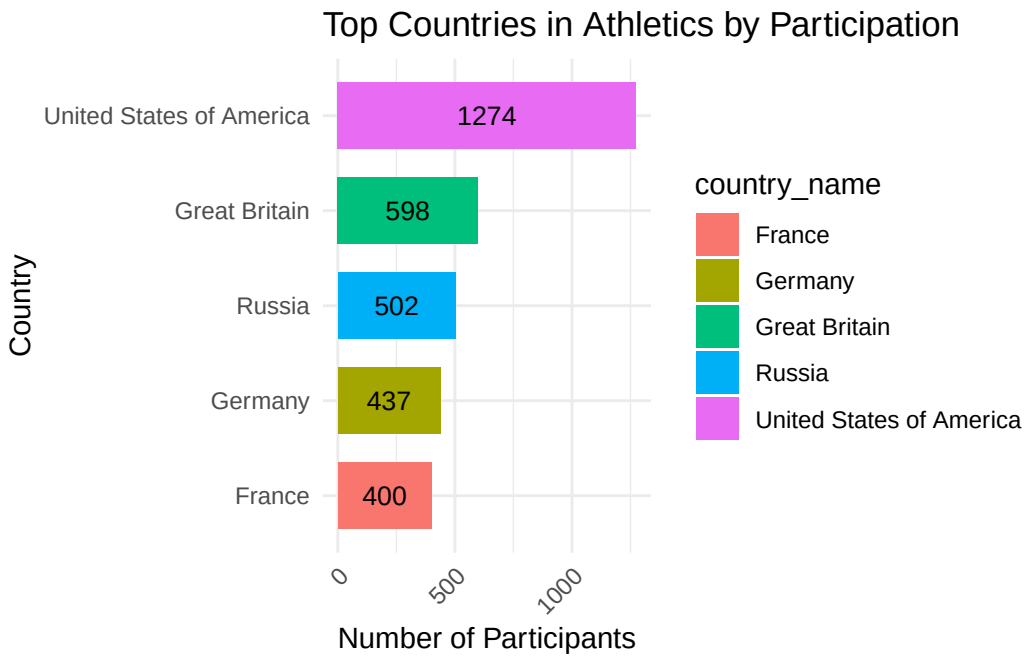
Is there a "host country advantage" where host countries win more medals?

The concept of a "host country advantage" posits that nations hosting the Olympics tend to perform better, often winning more medals than they would in non-hosting years. This could be attributed to various factors, including familiarity with local conditions, increased government investment, and national morale.



Interpretation: The top graph shows the number of medals won by host countries (in blue) over the years. It appears that host countries tend to win a higher number of medals during their Olympic Games. However, the fluctuations suggest that the advantage varies significantly from one event to another.The bottom graph illustrates the number of medals won by non-host countries (in red). There seems to be a consistent pattern of medal counts over the years, but host countries generally have a spike during their hosting year, indicating a distinct advantage.Hosting the Olympics provides a notable boost to the medal tally of the host nation, likely due to home support, familiarity with venues, and resources. The sharp spikes in the host medals during specific years indicate periods of significant success, while the fluctuations in non-host medals show a more stable but lower count.

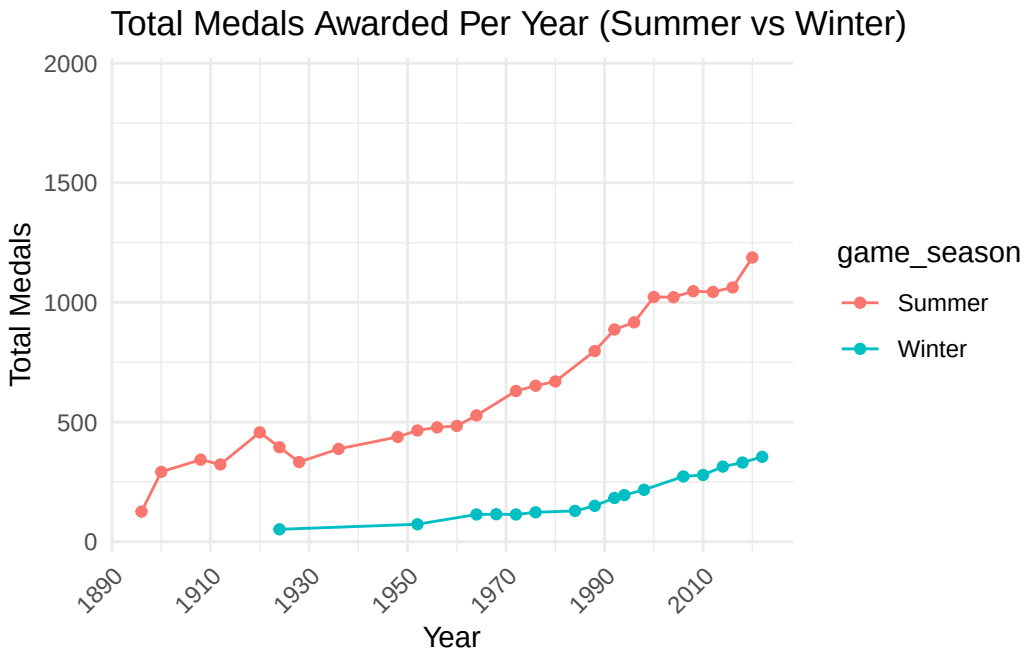
Which countries participate the most in athletics?



Interpretation: In the analysis above the United States leads significantly with 1,274 participants in athletics, showcasing its long-standing dominance in the sport. Great Britain follows with 598 participants, while the Soviet Union (502), Germany (437), and France (400) show considerable participation but are significantly behind the USA. The high number of participants from the USA suggests a strong investment in athletics, which aligns with its overall success in Olympic sports. The Soviet Union's participation highlights its historical strength in athletics, reflecting its emphasis on sports during its existence. The data may imply that nations with more athletes in a particular sport tend to have better overall performance, as evidenced by the USA's dominance in athletics.

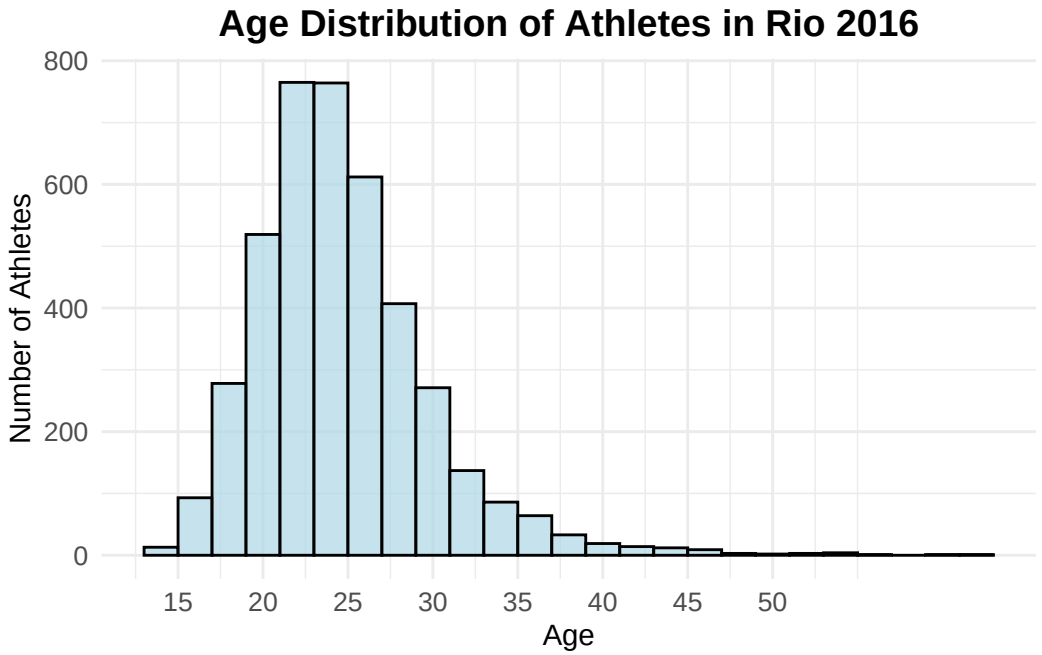
Is there a correlation between the season in which a sport is played and the amount of medals awarded?

The Olympic Games are held in two distinct seasons—Summer and Winter—which feature different sports and therefore offer different opportunities for nations to excel. This section examines the differences in medal distribution between the two seasons and whether participation in one season offers a competitive edge.



Interpretation: Summer Olympics consistently award more medals than Winter Olympics due to the larger number of sports and participating athletes. Countries like the USA, , and Russia have reaped the benefits of this, dominating Summer sports that offer higher medal tallies. In contrast, countries such as Norway, which specializes in Winter sports like cross-country skiing and biathlon, excel in the Winter Olympics but are at a disadvantage overall due to the fewer medals available.This seasonal imbalance highlights the inherent challenges of comparing overall medal counts across countries that focus on different Olympic seasons. Nations that dominate Winter sports may struggle to keep pace with Summer-heavy countries in total medal counts, despite excelling in their respective sports.

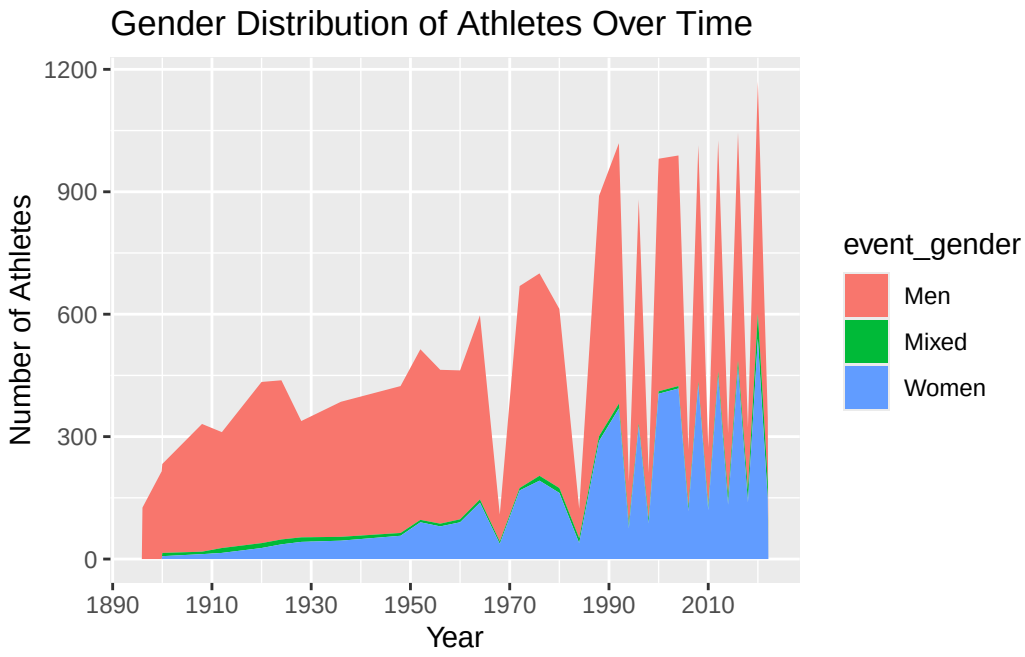
What is the distribution of athletes' ages in the 2016 Olympic Games? Plot a histogram to show the age distribution.



Interpretation: The age distribution of athletes in the 2016 Olympics shows that most participants were between 20 and 30 years old. Sports like gymnastics tend to attract younger athletes, with many peaking in their teens or early twenties. In contrast, endurance sports like marathon running and equestrian events see higher participation from older athletes, some of whom continue competing into their 40s. This age distribution underscores the varying physical demands of different sports and highlights the diverse pathways athletes take to reach Olympic competition.

How has the gender distribution of Olympic athletes changed over time?

The evolution of gender representation in the Olympics is a significant area of study, reflecting broader societal changes toward gender equality. Historically, men dominated Olympic participation, but recent decades have seen a concerted effort to include more women's events and encourage female participation.



Interpretation: The data shows a steady increase in female participation, particularly after the 1980s when more women's events were introduced into the Olympics. While male participation still outpaces female participation, the gap is narrowing. Events such as gymnastics, swimming, and athletics have seen substantial increases in female representation, helping to drive this trend. This trend toward gender parity reflects broader global efforts to promote equality and inclusion in sports. However, it also raises questions about whether true parity has been achieved or if certain barriers still exist, particularly in traditionally male-dominated sports.

Conclusion

This analysis has provided several key insights into the dynamics of Olympic participation and performance. The USA and Russia's dominance, particularly in the Summer Games, underscores the importance of strategic investments in sports that offer numerous medal opportunities, such as athletics and swimming. Russia's decline in the post-Cold War era, along with China's rise as a sporting powerhouse, reflects how political and economic conditions influence national performance in the Olympics. The concept of a host country advantage is well-supported by the data, with host nations consistently outperforming their usual medal counts during their hosting years. This is likely due to the home field advantage, greater preparation, and increased national pride. Gender analysis shows significant strides toward parity, though some barriers still exist, particularly in male-dominated sports. The growing

inclusion of women in the Olympics reflects broader societal changes and global efforts toward gender equality. Finally, the analysis of seasonal differences demonstrates the inherent challenges in comparing total medal counts across nations, particularly for countries that specialize in Winter sports. The Summer Games, with their broader range of events, offer more opportunities for medals, leading to an imbalance in total medal counts between countries that excel in Winter versus Summer sports. These findings contribute to a deeper understanding of the complexities behind Olympic success, highlighting the influence of historical, political, and societal factors on global athletic competition.