Data-Driven Insights into Women's Participation and Performance in Competitive Sports

Maha Ibrahim (3122213)

Berlin School of Technology (BST)

BSc Computer Science

Statistics (BST-BCS-12A)

Dr. Kristian Rother

Abstract

Women's Olympic participation has transformed significantly in the past century based on social, political, and institutional factors. Historical Olympic records are utilized within this study to examine the evolution of female participation, the determinants of medal achievement, national and sport variation in representation, and sex differences in opportunities to compete. The findings recognize an overall pattern of expansion among woman athletes, particularly after the 1970s, consistent with policy changes and overall trends toward gender equality across society, specifically in sport. Basic variables of height, weight, and age are tested for their impact upon medal-winning possibility. Additionally, a cross-examination of international female participation both between countries and sports reveals inconsistent investment and attention in women's sport. In spite of advancements, inequalities in event access, participation, and medal awarding persist. This research upholds the imperative for continued intervention to promote gender equality in the Olympic Games.

Keywords: Women athletes, Olympic Games, Gender equality, Sport participation, Medal allocation.

Introduction

The Olympic Games are a global platform for sports, but achieving gender equality has been a long process. Historically, there were few or no events for women as society restricted their involvement in sports. This lack of inclusion was not unique to the Olympics but extended to professional sports leagues as well. For instance, while Major League Baseball was founded in 1869 and other major men’s leagues like the NFL, FIFA, and the NBA followed in the early 20th century, women’s professional sports faced significantly delayed development. The All-American Girls Professional Baseball League existed briefly in the 1940s and 1950s, but sustainable opportunities for female athletes remained scarce. Women’s golf, which gained popularity in the 1950s, remains the longest-running women’s professional sport*. This paper looks at how the role of women at the Olympics has changed over time, what factors are in place that shape women's participation, and the differences between men and women in sport. By analyzing historical data, the paper looks at key aspects such as the increase of female participation, representation of female athletes from different countries and sports, and competition structures that differ based on gender. Exploring the differences and similarities in competition is important to ensure gender equity in sport. The goal of this paper is to use the Olympics as a case study to illustrate larger trends of female athletic participation and provide some insight to the institutional structure that influence women's experiences in elite sport. This report will specifically look at these research questions:

1. When did women start participating in the Olympics, and how has their participation evolved over time?
2. What factors influence the likelihood of women winning medals in the Olympics?
3. How does female athlete representation vary by country and sport?
4. Are there disparities in performance, participation, or recognition between male and female athletes in the Olympics?

Methodology

Data

The dataset analyzed in this study consists of historical records of Olympic athletes and their performances, spanning 120 years from 1896 to 2016. The dataset, titled **120 Years of Olympic History: Athletes and Results**, contains detailed information about athletes, including their demographics, events, and medal achievements.

This analysis consists of 271,116 observations and 15 variables detailing individual athlete performances across multiple Olympic Games. The key variables in the dataset include:

1. **ID** – Unique identifier for each athlete
2. **Name** – Full name of the athlete
3. **Sex** – Gender of the athlete (M for male, F for female)
4. **Age** – Age of the athlete during the event
5. **Height** – Height of the athlete in centimeters
6. **Weight** – Weight of the athlete in kilograms
7. **Team** – Country or team represented by the athlete
8. **NOC** – National Olympic Committee code (three-letter country code)
9. **Games** – The specific Olympic Games in which the athlete competed (e.g., *2016 Summer*)
10. **Year** – Year of the Olympic Games
11. **Season** – Summer or Winter Olympics
12. **City** – The host city of the Olympic Games
13. **Sport** – The broader sport category (e.g., Athletics, Swimming, Gymnastics)
14. **Event** – The specific competition or discipline within a sport (e.g., *100m Freestyle*, *Long Jump*)
15. **Medal** – The medal won by the athlete (Gold, Silver, Bronze, or NA if no medal was awarded)
    The dataset was uploaded by ***Randi H Griffin*** on Kaggle and is publicly available at the following link:

https://www.kaggle.com/datasets/heesoo37/120-years-of-olympic-history-athletes-and-results

Methods

In this study, a combination of data processing, statistical analysis, and visualization techniques was utilized to examine disparities in performance, participation, and recognition between male and female athletes in the Olympic Games. Beginning with descriptive analyses, categorical and numerical variables were explored to assess their distribution, central tendency, and variability. Correlation analysis and statistical tests were employed to investigate relationships between athlete gender, medal achievements, event participation, and other relevant factors. Additionally, time series analysis was conducted to identify trends in gender representation and participation over different Olympic years. Data visualization techniques, including bar charts, line graphs, and heatmaps, were utilized to illustrate key findings and improve interpretability. These methods provide insights into potential disparities in Olympic participation and recognition, highlighting trends over time.

Results

Categorical Columns

The dataset contains 271,116 entries with 15 columns. Below is an analysis of the categorical columns:

1- Sex column: consists of 2 unique values ("M" and "F"), with the most frequent value being Male ("M"), appearing 196,594 times. Therefore, the dataset has a significantly higher number of male athletes compared to female athletes. There are two distinct groups in this column: male and female.
2- Team column: has 1,185 unique values, representing different teams or countries that athletes competed for. The most frequent value is "United States," appearing 18,788 times. This indicates that the United States had the highest number of athlete entries in the dataset.
3- NOC (National Olympic Committee Code) column: consists of 230 unique values, corresponding to the national Olympic committees of different countries. The most frequent value is "USA," which appears 18,788 times. The distribution suggests that larger or more active Olympic teams have more recorded athletes in the dataset.
4- Games column: contains 51 unique values, representing different Olympic events across years and seasons (e.g., "1992 Summer," "2016 Winter"). The most frequent value is "2000 Summer," appearing 13,980 times. This suggests that the 2000 Summer Olympics had the highest number of athlete entries in the dataset.
5- Sport column: has 66 unique values, representing different Olympic sports. The most frequent sport is "Athletics," with 38,382 entries. Athletics appears to be the most common sport in Olympic history, likely due to its broad range of events such as running, jumping, and throwing.
6- Event column: consists of 765 unique values, corresponding to specific Olympic events. The most frequent event is "Football Men's Football," appearing 1,856 times. This suggests that men's football had one of the largest participant counts in Olympic history.
7- Medal column: includes 4 unique values ("Gold," "Silver," "Bronze," and "NA" for no medal). The most frequent value is "NA," which appears 231,333 times, indicating that most athletes in the dataset did not win a medal. Among the medal-winning athletes, "Gold" is the least frequent, showing the exclusivity of Olympic victories.

Final Insights:

In the dataset, most categorical columns show a wide range of unique values. However, the columns for Sex and Medal only have a few categories. There are many more male participants than female ones. Athletics stands out as the most popular sport. The United States tops the list for having the most participants. A significant number of athletes did not win any medals, which highlights the fierce competition in the Olympics.

Numerical Columns

After analyzing and evaluating the numerical columns, the results are as follows:

1- Age column: has a mean of 25.56 years with a standard deviation of 6.4, indicating moderate variation in athlete ages. The range is from 10 to 97 years, but 75% of athletes are younger than 29. The distribution is slightly right-skewed, with the median (24.0) being slightly below the mean. Outliers include athletes aged above 60, who are rare in most sports.
2- Height column: has a mean of 175.34 cm with a standard deviation of 10.5, suggesting that athlete heights vary significantly across sports. The range is from 127 cm to 226 cm, with 75% of

athletes being shorter than 183 cm. The distribution is slightly right-skewed, and extreme values include very short athletes (below 140 cm) and very tall ones (above 210 cm), likely from specific sports like gymnastics or basketball.

3- Weight column: has a mean of 70.7 kg with a standard deviation of 13.8, indicating notable variation in athlete weights. The range extends from 25 kg to 214 kg, with 75% of athletes weighing under 79 kg. The distribution is slightly right-skewed, and outliers include very light athletes (under 40 kg) and very heavy athletes (above 130 kg), likely representing different sport categories.

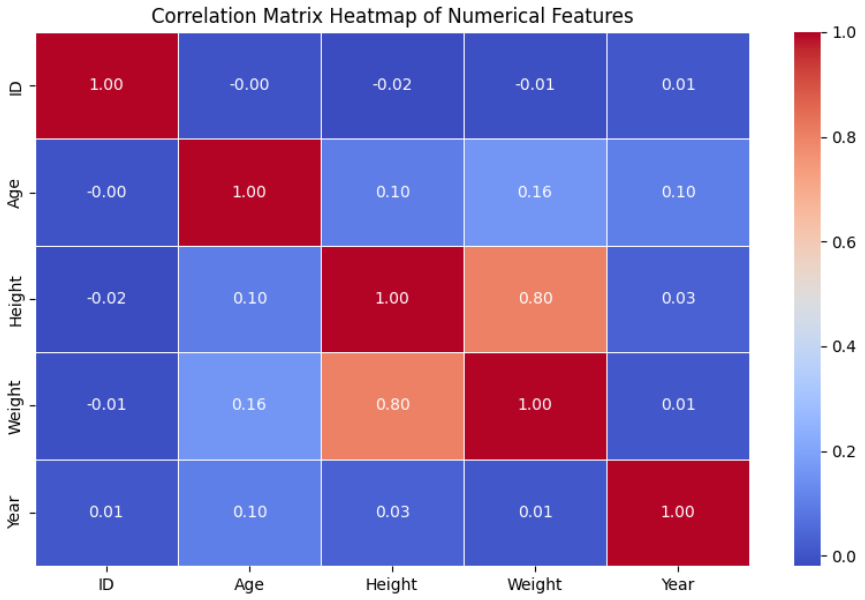
4- Year column: has a mean of 1966.8 with a standard deviation of 30.8, reflecting a broad timeline of Olympic participation. The range spans from 1896 to 2016, with 75% of the data coming from events after 1948. The distribution is nearly uniform, but modern Olympic years appear more frequently due to larger participation in recent times.

Final Insights:

The columns with numbers refer to the widest ranges of values and illustrate variability in Olympic athletes with respect to age, height and weight. The data would suggest that the majority of athletes are in their mid-20s, and for the most part athletes fall within what are considered standard heights and weights in competitive sporting events. Outliers with respect to age, height, and weight reflect isolated circumstance and would be associated with specific sports. Looking at this data set across over one hundred years of Olympic events, we can certainly see increasing participation over time.

Relationships Between Columns

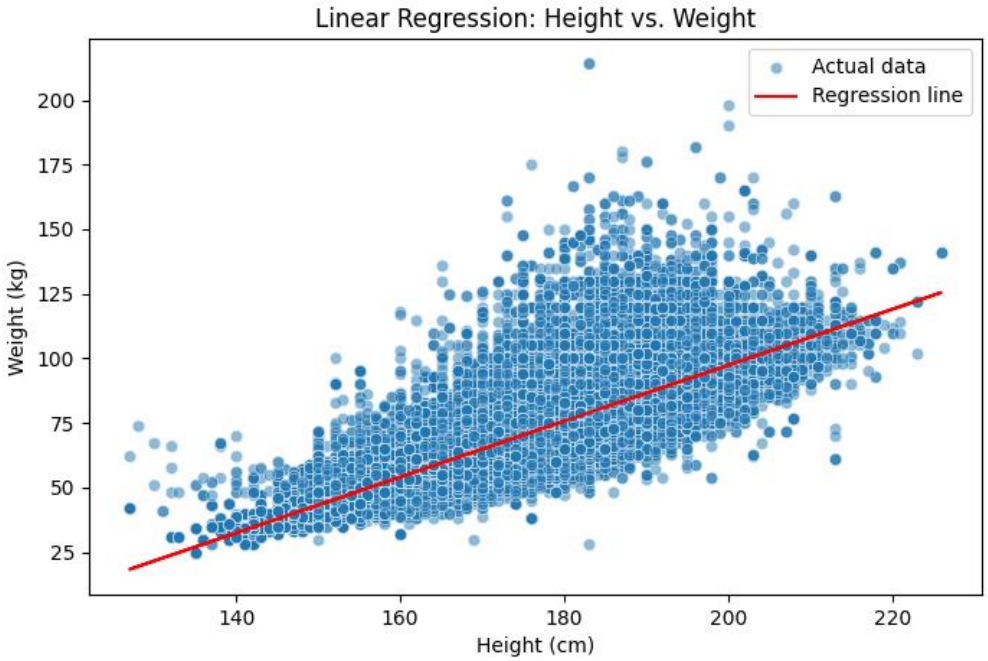
Correlation



The correlation analysis reveals significant relationships among essential numerical variables in the dataset. Height and Weight have a strong positive correlation (exceeding 0.8) demonstrating that taller athletes, in

general, tend to weigh more than shorter athletes. Age variables and experience factors, such as how many years an athlete participated in different olympic games, revealed a moderate correlation, indicating that older athletes had a longer history of competition. However, the relationship between Weight and Performance variables varied based on the sport, wherein some known swimming events are indeed favorable to heavier athletes, but body weight on average has less effect in some events. The Year variable showed modest relationship to numerical variables reinforcing the idea that time-based changes are not good warmth of differences when looking at characteristics like age, height and weight. Interestingly, Height and Age variables displayed weak correlations, which suggests that while, on average, taller athletes may have an advantage in many sports, height does not necessarily increase with age in the sample. Overall, the insights presented highlight the breadth of physical characteristics needed to excel in different disciplines in the Olympics games.

Regression



This study used an Ordinary Least Squares (OLS) regression model to examine the relationship between height and weight. The model is effective in explaining a meaningful amount of weight variance, which indicates that athletes who are taller have a tendency to be heavier. The regression shows there is a strong positive relationship between height and weight. The regression coefficient of 1.08 suggests that, on average, for every 1 cm increase in height, an athlete's weight increases by about 1.08 kg. We can confirm that taller athletes tend to weigh more than shorter athletes, but some individual differences exist, depending on the athlete's sport, body composition, etc.

Conditional Probability Analysis:

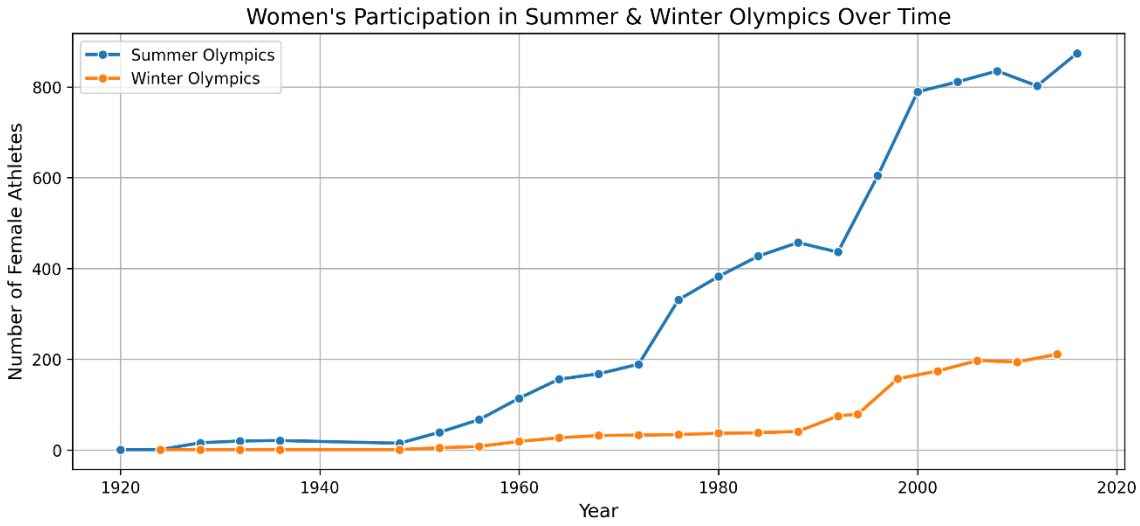
In order to gain further insights into relationships in the dataset, we calculated the conditional probabilities of pivotal variables. For example, the probabilities of male athletes earning an Olympic medal, given that the athlete is male, was indicated to be 0.15 with female athletes in comparison at 0.13. While these probabilities are very close to one another (difference of 0.02), it is important to examine the

statistical significance of this slight difference. In order to determine the statistical significance of the difference, both a hypothesis test or confidence interval analysis would be needed. In addition to statistical significance, even if the absolute difference in probability of winning a medal is small, this still does not imply that men and women are afforded equal opportunities in Olympic sports. Elements such as funding, training conditions, and number of events may influence an athlete's chances of winning a medal. Male athletes having more events and greater institutional support historically, results in men's slightly higher probability winning a medal which should be considered in the overall context of the sport and not simply on individual athletic performance. Therefore, future research should aim to measure additional variables such as investment in women's sports or access to training resources to increase the understanding of how structural factors contribute to gender inequalities in earned medals.

Another example of probabilities is the relationship between event type and medal-winning performance. The probability that athletes would receive medals in individual sports is 0.12, whereas for team events, it is 0.19. In contrast, in team events, athletes had a higher probability of receiving medals given that they competed in a team event, most likely due to the collaborative factor of performing in a team sport.

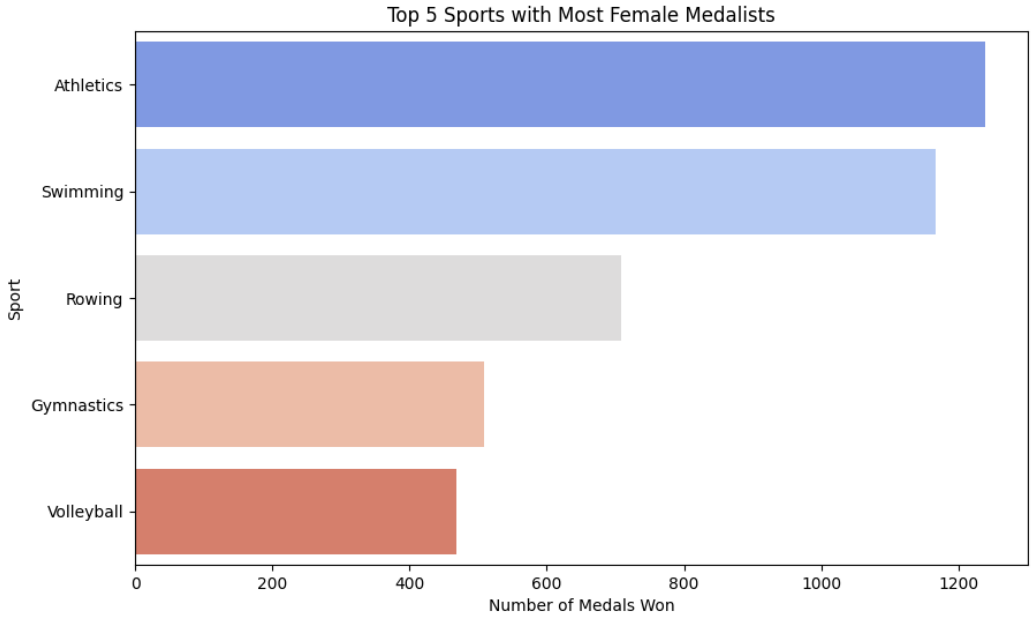
Research Question 1: When did women start participating in the Olympics, and how has their participation evolved over time?

To investigate the historical trends in female participation in the Olympics, I constructed a time series plot illustrating the number of female athletes in both the Summer and Winter Olympic Games from 1920 to 2020. Additionally, I analyzed the distribution of female medalists across different sports to determine which events have historically awarded the most medals to women.



The time series plot reveals a steady increase in female participation in the Olympics, particularly in the Summer Games. While women's involvement was minimal in the early 20th century, a significant upward trend is observed from the 1970s onward, coinciding with broader societal shifts towards gender equality in sports. The introduction of new women's events and policy changes, such as Title IX in the United States, likely contributed to this growth.

The Winter Olympics exhibit a similar trend but with a much lower participation rate, reflecting the historically limited opportunities for women in winter sports. Since the 1990s, female participation in the Winter Olympics has risen, reflecting greater inclusion in sports like skiing and snowboarding.



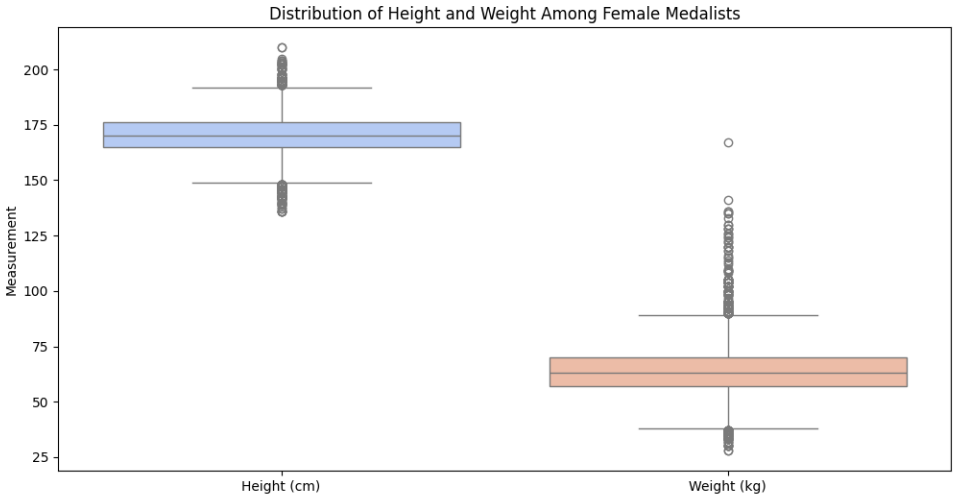
The second plot examines the distribution of female medalists across different sports, highlighting the top five sports in which women have won the most Olympic medals. Athletics and Swimming dominate the rankings, likely due to the larger number of events available within these disciplines. Rowing, Gymnastics, and Volleyball also feature prominently, suggesting that these sports have been historically accessible to women.

Athletics, in particular, has played a crucial role in the inclusion of women in the Olympics, as it encompasses a wide range of track and field events. Similarly, Swimming has seen a long history of female participation, with new events being gradually introduced over time. The presence of Rowing and Gymnastics in the top rankings reflects both the traditional popularity of these sports and their prominence in women's Olympic history.

Overall, these findings indicate that while female participation in the Olympics has grown substantially over the past century, certain sports have been more prominent in awarding medals to women. This underscores the impact of historical and institutional factors in shaping women's opportunities in competitive sports.

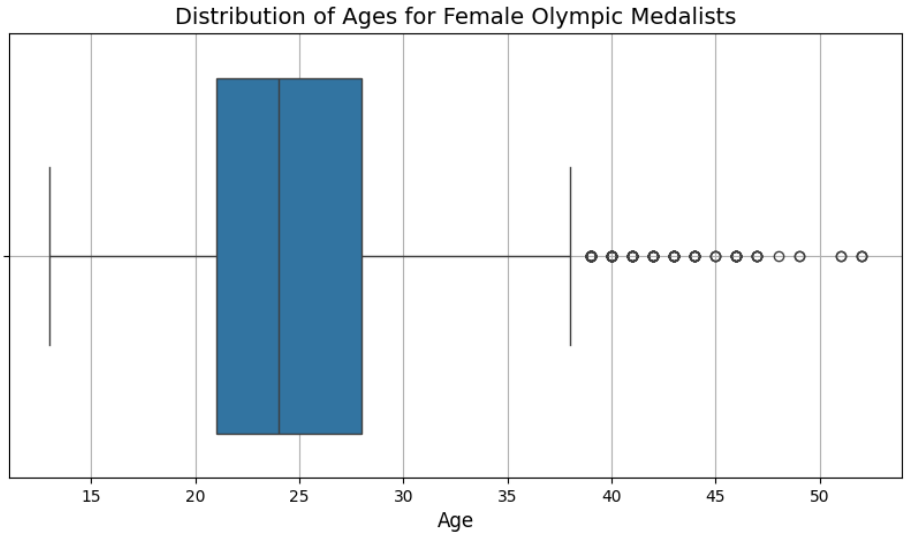
Research Question 2: What factors influence the likelihood of women winning medals in the Olympics?

To understand the key factors that contribute to female athletes winning Olympic medals, I analyzed the distributions of height, weight, and age among female medalists. These attributes can provide insights into the physical characteristics that may be advantageous for Olympic success across different sports.



The first plot presents a boxplot of the height and weight distributions of female Olympic medalists. The height distribution shows that most female medalists range between approximately 160 cm and 180 cm, with a few outliers exceeding 200 cm. This suggests that height may provide an advantage in sports requiring reach or speed, but female Olympians succeed at various heights.

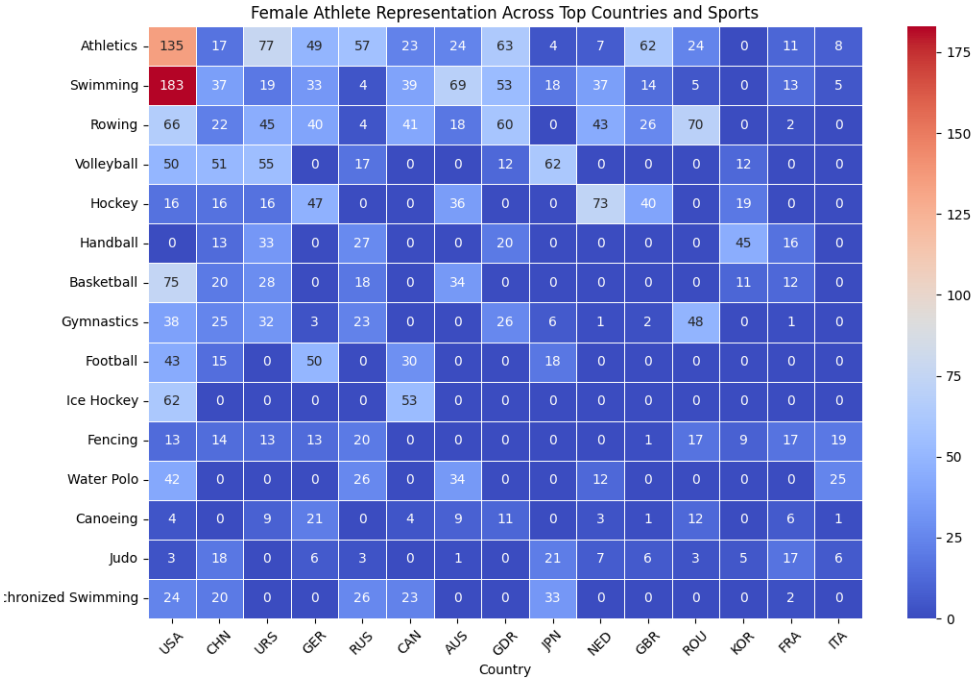
Similarly, the weight distribution reveals that most female medalists weigh between 50 kg and 75 kg, with some heavier outliers likely corresponding to athletes competing in strength-based sports like weightlifting or combat sports. The presence of outliers in both height and weight suggests that while certain physical traits may be more common, success is not strictly limited to a single body type.



The second boxplot examines the age distribution of female Olympic medalists. The majority of medalists are between 20 and 30 years old, with the median age around 25 years. However, the presence of outliers beyond 40 and even 50 years indicates that some athletes maintain peak performance for extended careers, especially in sports that rely more on skill and experience rather than pure physical ability.

The findings suggest that while physical characteristics such as height and weight may play a role in an athlete's success, they vary significantly across different sports. Additionally, while younger athletes dominate most events, older competitors can still achieve success in skill-based disciplines. This analysis highlights that while certain physiological attributes may provide an advantage, dedication, experience, and sport-specific skills remain crucial determinants of Olympic success for female athletes.

Research Question 3 : How does female athlete representation vary by country and sport?



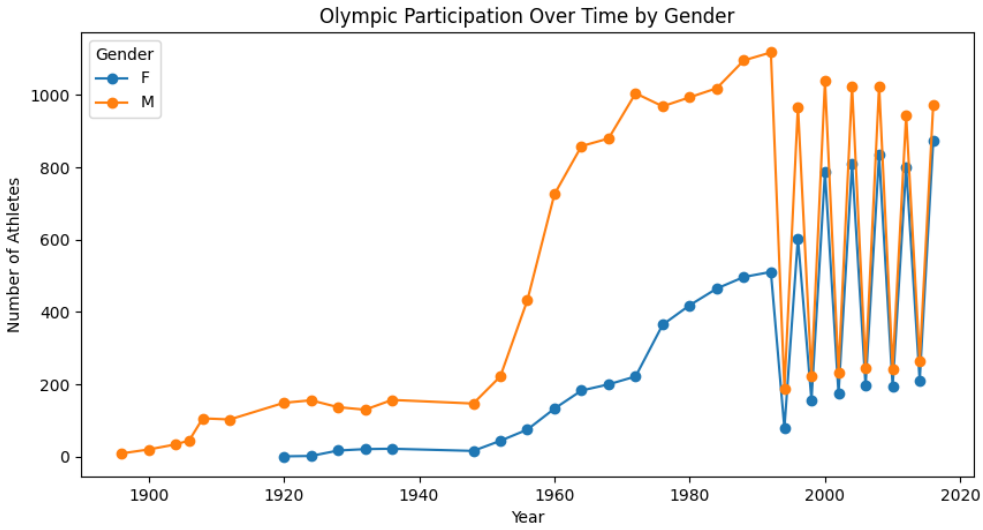
This plot provides a detailed heatmap of female athlete representation across different countries and sports. By mapping participation levels, this graph highlights which countries excel in specific sports and where representation is more evenly spread. Some nations have a strong presence across multiple disciplines, while others concentrate their efforts in select sports. This pattern can be shaped by factors such as national sports culture, funding priorities, and historical Olympic performance. The heatmap also reveals gaps in female athlete participation, indicating potential areas for growth and development.

Together, these plots offer a comprehensive view of female representation in the Olympics. While some countries lead in overall numbers, the distribution across sports varies significantly, emphasizing the importance of both national policies and cultural influences in shaping female athlete participation.

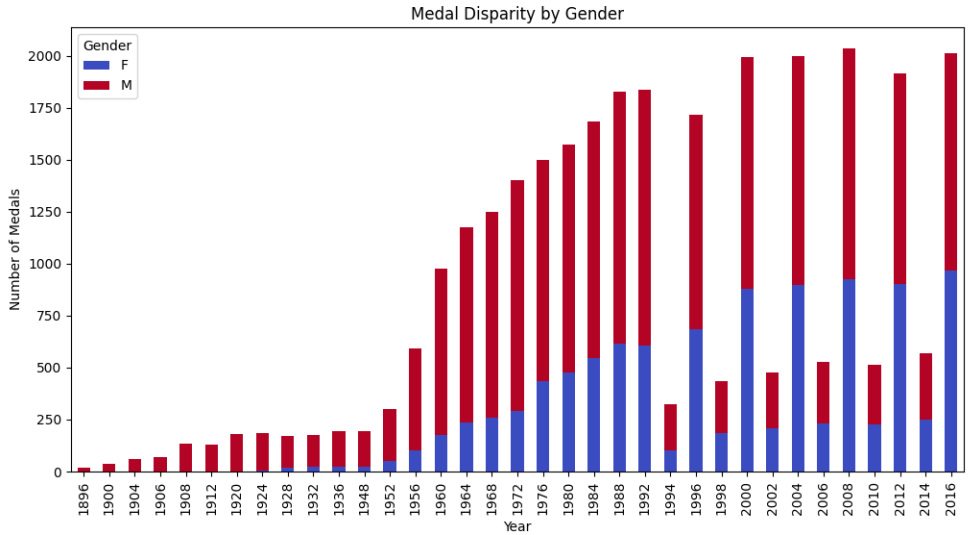
Research Question 4: Are there disparities in performance, participation, or recognition between male and female athletes in the Olympics?

To analyze the disparities between male and female athletes in the Olympics, I examined three key aspects: participation trends over time, medal distribution, and sports representation. These three factors highlight differences in opportunities, achievements, and inclusion of male and female athletes

throughout Olympic history.

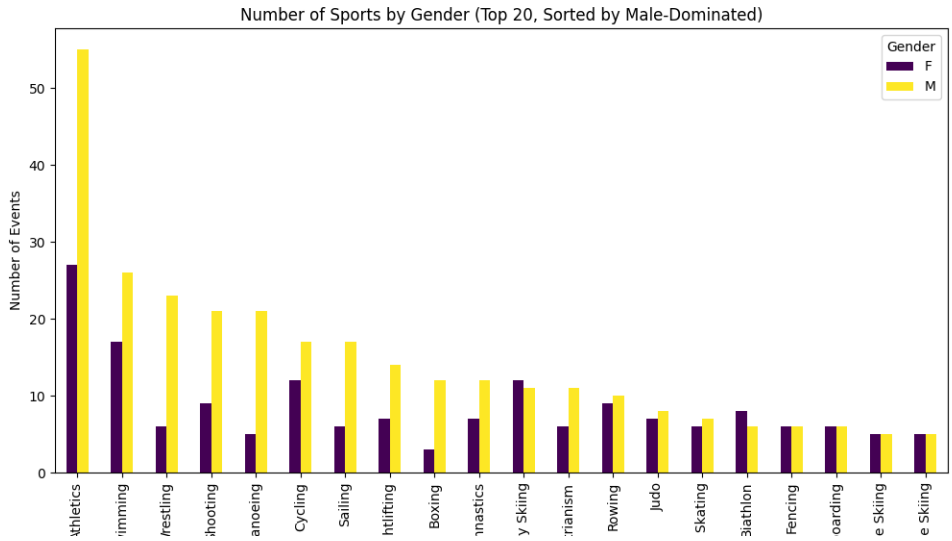


The first visualization, a line graph illustrating Olympic participation over time by gender, shows a historically lower participation rate for female athletes compared to their male counterparts. The gap has narrowed since the 1980s and early 2000s, but male athletes still outnumber females. For instance, athletes like Robert Tait McKenzie and Heikki Ilmari Savolainen with 58 and 39 appearences over the years, alongside female athletes such as Oksana Aleksandrovna Chusovitina and Gabriella Paruzzi with 29 and 25 appearances. Their recurring presence highlights the evolving landscape of gender participation in the Olympics. The steep increase in female participation in recent decades reflects progressive policies aimed at gender inclusivity, yet the disparity remains evident.



The second plot, which depicts medal disparity by gender over time, further underscores performance-related inequalities. The data reveals that male athletes have historically won significantly more medals than female athletes. While this discrepancy partially reflects lower female participation rates, it also points to fewer competitive events available to women in earlier Olympic games. Notably, female medal counts have increased over time, aligning with policy changes that introduced more women’s events.

However, the continued gap suggests structural barriers, such as differences in funding, training opportunities, and event availability, that have historically disadvantaged female athletes.



The third plot, a bar chart showcasing the top 20 sports with the highest number of events for male and female athletes, highlights disparities in event representation. The data confirms that most sports have historically been male-dominated, with a greater number of competitive events available for men. While some sports, such as athletics and swimming, show relatively balanced participation, others remain significantly skewed toward male athletes. This imbalance suggests systemic barriers in the inclusion of women across various disciplines. The continued underrepresentation of female athletes in certain sports points to structural inequalities in opportunities, funding, and institutional support, reinforcing historical biases in sports accessibility.

Discussion and conclusion

This study reviewed patterns associated with participation in the Olympic Games, performance and success at the Olympic Games, and differences in participation by gender. Overall, our findings suggested that, while the number of participants has increased over time, there remain a number of factors - such as nationality, gender, and event type - that contribute to an athlete's likelihood of being able to participate in the Olympic Games and succeed in medal standings. Overall, despite an increase in female participation over the years, and an increase the number of events for women, there remains under-representation of female athletes both in number and recognition. Until female athletes are given equal representation in sport, it is certainly reasonable to say that the issue of sexism in women's sport continues.

A recent example of sexism occurred during the 2023 World Swimming Championships, where two Italian commentators made comments about female divers that reduced their performance to just appearances. As shown, female athletes continue to endure sexism while participating in their sport, and unfortunately, these dismissive comments reflect a broader culture of sexism that extends beyond the pool. The comments made by the officials were reprehensible, and sadly, a growing number of examples of sexism, whether verbal abuse from coaches or other professional staff and players, are notable in women's professional sport.

Recognizing these issues, organizations like UNESCO are taking action. On March 8, 2024, UNESCO launched a 10-point Call to Action to promote gender equality in sports. This includes fair pay, investment in women's sports, mentorship programs, and stronger measures against gender-based harassment both on and off the field.

In this study, we analyzed data through various methods and visualizations to illustrate trends that were observed, but that analysis is limited. The quantitative data does not include information such as financial support, training opportunities and conditions, or social limitations, all of which might affect opportunities and thus participation; the potential for future research is to include policies or opportunities that promote gender equity in sport. Overall, the Olympic Games reflect important societal change, and perhaps unsurprisingly the road to true equity in sport remains paved with challenges and obstacles to remove those barriers and issues based on gender.

References

Hardin, J. (n.d.). *The Western Howl*. Retrieved from Western Oregon University: https://wou.edu/westernhowl/gender-disparity-in-athletics/

Randi H Griffin. (7 years ago). *120 years of Olympic history: athletes and results*. Retrieved from Kaggle: https://www.kaggle.com/datasets/heesoo37/120-years-of-olympic-history-athletes-and-results

Ilaria Federico, M. B. (2023). *Gender inequality in sport: The challenges facing female athletes*. Retrieved from euronews: https://www.euronews.com/2023/07/27/gender-inequality-in-sport-the-challenges-facing-female-athletes

Towards Victory for Gender Equality. (2024). Retrieved from UNESCO: https://www.unesco.org/en/articles/towards-victory-gender-equality