CIS 568 – Data Visualization (2024 Spring)

End Sem Project Report (Group - 6)

Submitted By:

Ankitaben Mungalpara (ID: 02129569)

Email ID: amungalpara@umassd.edu

Siddharth Jain (ID: 02102211)

Email ID: sjain9@umassd.edu

Table of Contents:

Selected Story

Exploring 120 Years of Olympic History Through Athletes and Results

Introduction – Project Objective and Scope

We address two important questions in this section: **what are we studying** and **why have we selected this area of study?** We seek to collect insights into the Games' changing dynamics by examining Olympic athlete participation in gender representation, age distribution, and medal outcomes. We chose this study because we want to make a significant contribution to the conversations about sports diplomacy, gender equality, and the legacy of the Olympics. Our goal is to offer a more comprehensive understanding of Olympic history and its wider social impacts through data visualization.

For more than a century, the Olympic Games—regarded as the height of athletic achievement—have been a source of inspiration and integration. Since its founding in 1896, the Olympics have served as a platform for both the celebration of sporting achievements and the changing face of international society. From its modest beginnings with a few hundred male competitors to a global platform where thousands of athletes from diverse backgrounds showcase their talents, the Games have undergone significant transformations throughout its rich history.

Even though the Olympics have long been praised for their capacity to unite people from different countries through friendly competition, they have also come under fire, especially when it comes to concerns about inclusivity and gender representation. In the past, female athletes have encountered obstacles and differences in engagement with respect to their male counterparts, which is indicative of wider societal injustices. To promote gender equality within the Olympic movement and address these disparities, deliberate efforts have been made in the last few decades.

This project explores the complex dynamics surrounding the participation of Olympic athletes, concentrating on trends in medal distribution, age analysis, and gender representation. Utilizing information spanning more than 120 years of Olympic history, our goal is to disentangle the intricacies that underlie athlete demographics and performance results.

To evaluate the progress made toward gender parity, comprehend the connection between age and athletic performance, and discover prominence about sports or by different countries, this project focuses on gender representation, age distribution, and medal distribution in Olympic sports.

In summary, this study seeks to bring knowledge about the complex patterns of Olympic athlete involvement, providing insightful information on gender dynamics, trends in participation, and the larger sociocultural background of the Olympic Games. We will use various data visualization approaches to clarify historical Olympic trends and provide insights into different nations' participation.

Data Collection

The dataset utilized in this research project was sourced from Kaggle, particularly via a hyperlink provided here: Olympic history spanning 120 years: athletes and outcomes

The "athlete_events.csv" collection includes details on specific athletes competing in Olympic competitions from the 1896 Games in Athens to the 2016 Games in Rio. There are 15 columns and 271,116 rows in the dataset, each row representing a different participant participating in an Olympic event. Among the columns are:

1. **ID**: Unique identifier for each athlete.
2. **Name**: Name of the athlete.
3. **Sex**: Gender of the athlete (M or F).
4. **Age**: Age of the athlete (integer).
5. **Height**: Height of the athlete in centimeters.
6. **Weight**: Weight of the athlete in kilograms.
7. **Team**: Name of the team the athlete represents.
8. **NOC**: National Olympic Committee 3-letter code.
9. **Games**: Year and season of the Olympic Games.
10. **Year**: Year of the Olympic Games (integer).
11. **Season**: Summer or Winter.
12. **City**: Host city of the Olympic Games.
13. **Sport**: Sport in which the athlete competes.
14. **Event**: Specific event in which the athlete participates.
15. **Medal**: Medal won by the athlete (Gold, Silver, Bronze, or NA).

This dataset is an excellent resource for performing in-depth analyses of Olympic trends, gender dynamics, and medal distributions since it offers a comprehensive view of Olympic athlete participation and performance over the last century.

Methodology

We focus on one key question in this section: **how are we analyzing the Olympic games over time?** This section examines a wide range of data visualization methods for examining various aspects of Olympic performance and participation:

1) **Scatter Plot Analysis:** We investigate the relationship between medal counts, participant numbers, and GDP using scatter plots. We can investigate possible correlations between these variables thanks to this visual aid.
2) **Medal Counts Over Time:** We generate visualizations that offer insights into changes in medal distribution over time by extracting counts of bronze, gold, and silver medals for participating countries between 1980 and 2016.
3) **Age Distribution Analysis:** To examine the age distribution of athletes in sports and nations, histograms are used. The age distribution of Olympic athletes in various sports is revealed by this visualization technique.
4) **Trends in Gender Participation:** We examined gender participation over time by nation using line graphs. We can see historical trends and patterns in the representation of gender in the Olympics thanks to this visualization technique.
5) **Medal Distribution by Sport and Country:** Bar graphs are used to examine how many medals each sport has won for nations and periods. This kind of visualization offers a thorough analysis of medal accomplishments in various sports.
6) **Leading Participating Nations:** Pie charts are used to show which nations competed at the highest level in the Olympics between 1980 and 2016. This visualization technique provides a clear picture of how different countries' participation in the Olympics was distributed over the given time frame.

This section, in summary, discusses the methodology used to perform a thorough analysis of Olympic trends. A range of visualization techniques were employed to gain insights into the distribution of medals, athlete demographics, gender representation, and dynamics of country participation.

Results and Findings

1. Scatter Plot Analysis:

We observed a positive correlation in the scatter plot between medal counts and participant numbers for the year 2008. Countries that achieved higher medal success were linked to larger circle sizes, which indicate higher participant numbers. Upon closer inspection, it was discovered that these anomalies frequently matched countries that had made significant investments in sports facilities and training initiatives. This analysis emphasizes the importance of strategic investment in athletic growth and the significant effect of participant numbers on medal success.

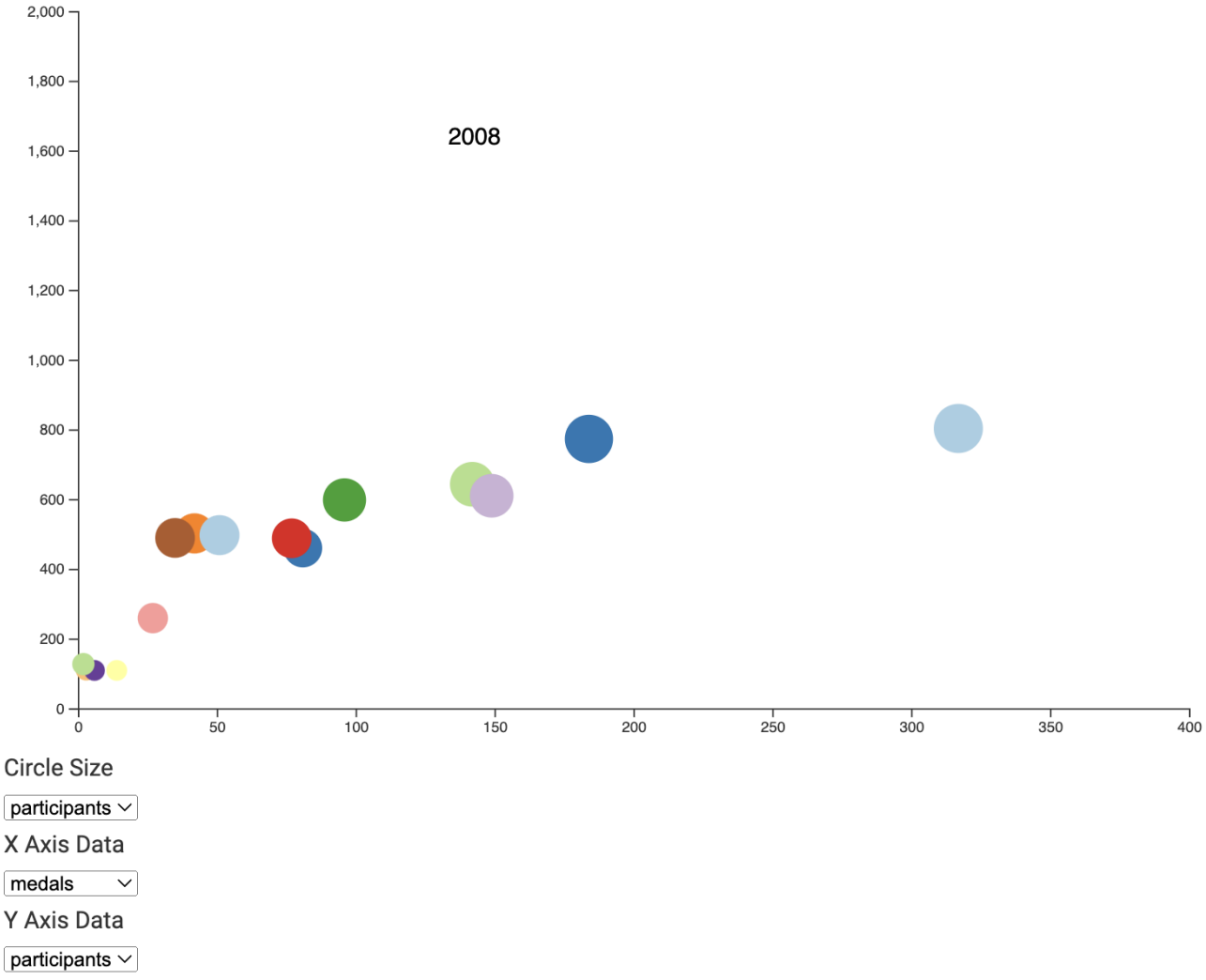


Fig. 1: Relationship Between Medal Counts and Participant Numbers in the 2008 Olympics

2. Scatter Plot Analysis:

Using a scatter plot, where the size of the circles represents the number of medals, we investigated the relationship between GDP and medal counts in our analysis for the year 2012. The scatter plot showed that the distribution of nations varied depending on the GDP and medal count range. For nations with higher GDPs, larger circle sizes were found to correspond with higher medal counts.

There were, nevertheless, some noteworthy outliers, signifying nations with noteworthy medal success in relation to GDP. This implies that while economic power plays a part in medal success, other elements like talent spotting and strategic investments in sports development are also important. Overall, the analysis shows how different factors affect a nation's performance on the international sporting arena, and the complicated relationship between GDP and Olympic medal counts is highlighted.

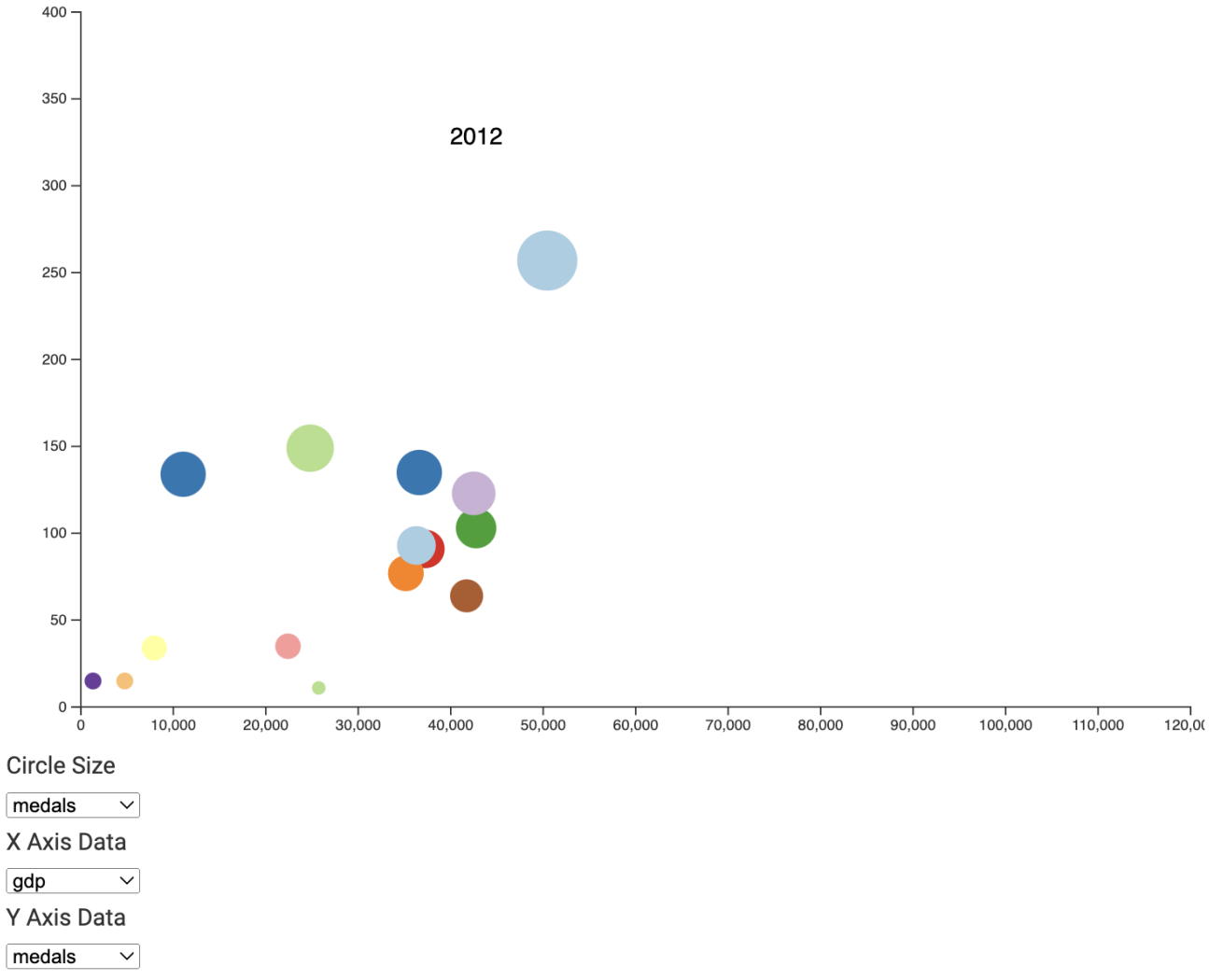


Fig. 2: Relationship Between GDP and Olympic Medal Counts in 2012

3. Medal Counts Over Time:

We examined the numbers of silver, bronze, and gold medals won by various nations at the 2004 Summer Olympics. The medal counts for a few chosen countries are shown in the table below:

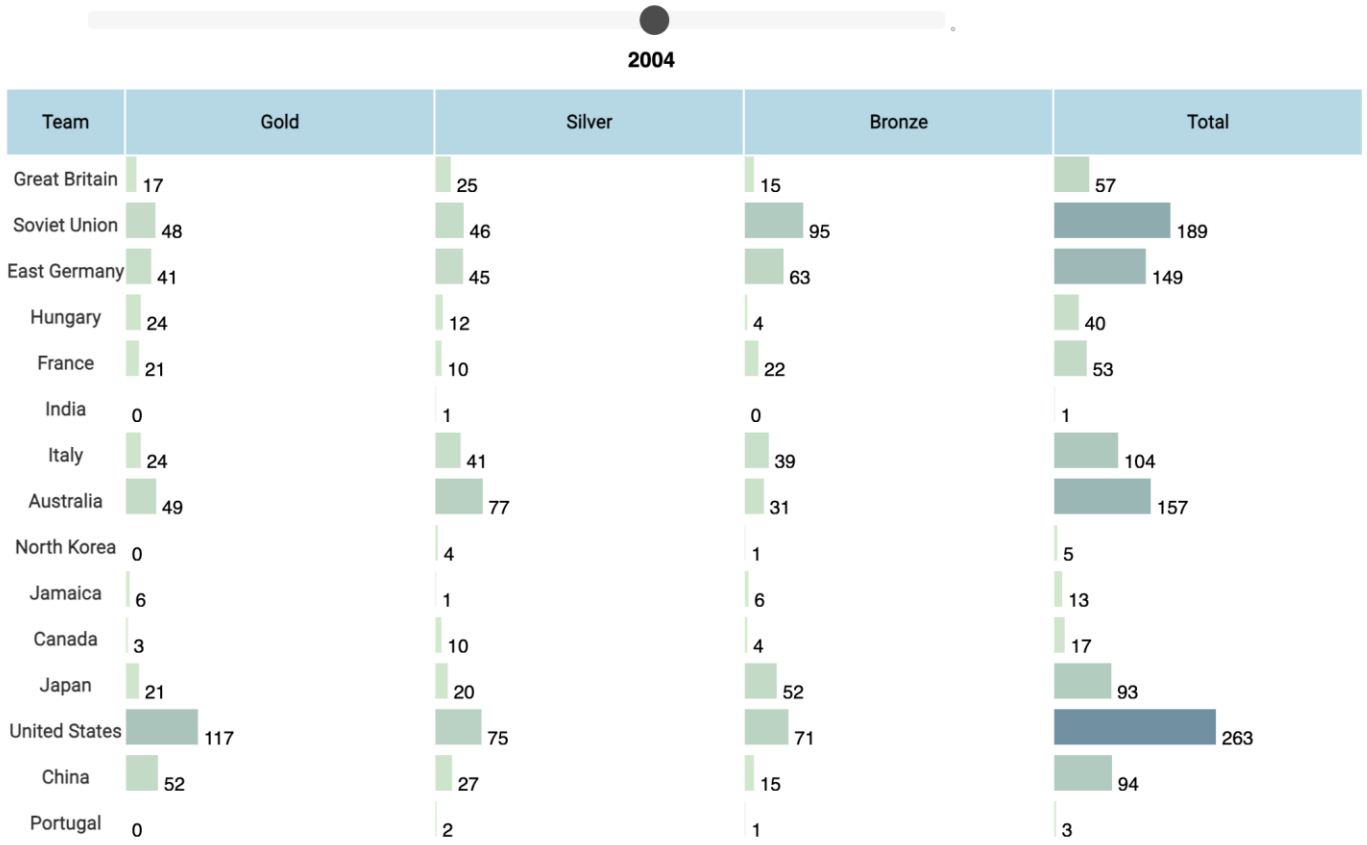


Fig. 3: Medal Counts for Select Countries in the 2004 Olympics

The table shows significant differences in the number of medals won by various nations. With 263 medals overall, the United States won the most, followed by the Soviet Union with 189 medals. This analysis sheds light on the medal hauls of the various participating nations in the 2004 Olympics, emphasizing the competition and disparities in performance between them.

4. Age Distribution Analysis:

We observed a wide range of age demographics represented in our analysis of the athlete age distribution for the United States across all sport games. The histogram showed a bell-shaped curve, which suggested that the ages of the athletes were distributed normally.

Most athletes were found to be between the ages of 20 and 30, which is the peak athletic performance period for many sports. There were, nevertheless, some noteworthy exceptions, with athletes ranging in age from teenagers to people in their 40s and beyond.

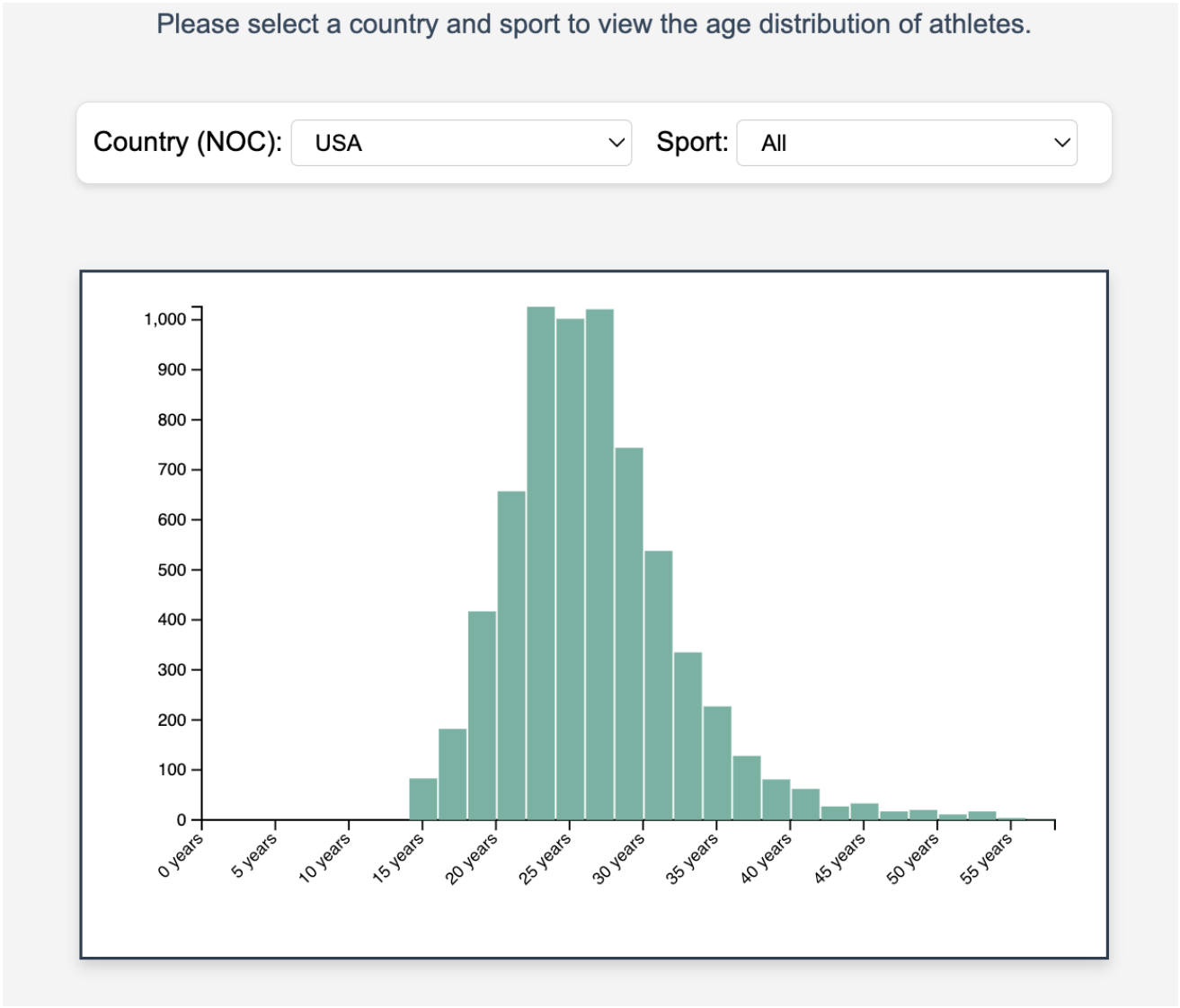


Fig. 4: Athlete Age Distribution for USA Across All Olympic Sports

The distribution indicates that American athletes span a wide range of ages, which is consistent with the variety of sports and the length of athletic careers. The analysis emphasizes the range of age groups that support American participation in Olympic sports.

5. Age Distribution Analysis:

Athlete age distribution analysis for swimming events in Australia showed that most participants were between the ages of 15 and 30. According to this age distribution, Australian swimmers who compete in the Olympics typically represent their country in their peak sporting years. This age group of athletes is concentrated in competitive swimming, where swimmers usually peak between their late teens and early twenties. The significance of early talent development and youth development programs in Australia's swimming community is highlighted by this finding. All things considered; this analysis offers insightful information about the age distribution of Australian swimmers competing at the Olympics.

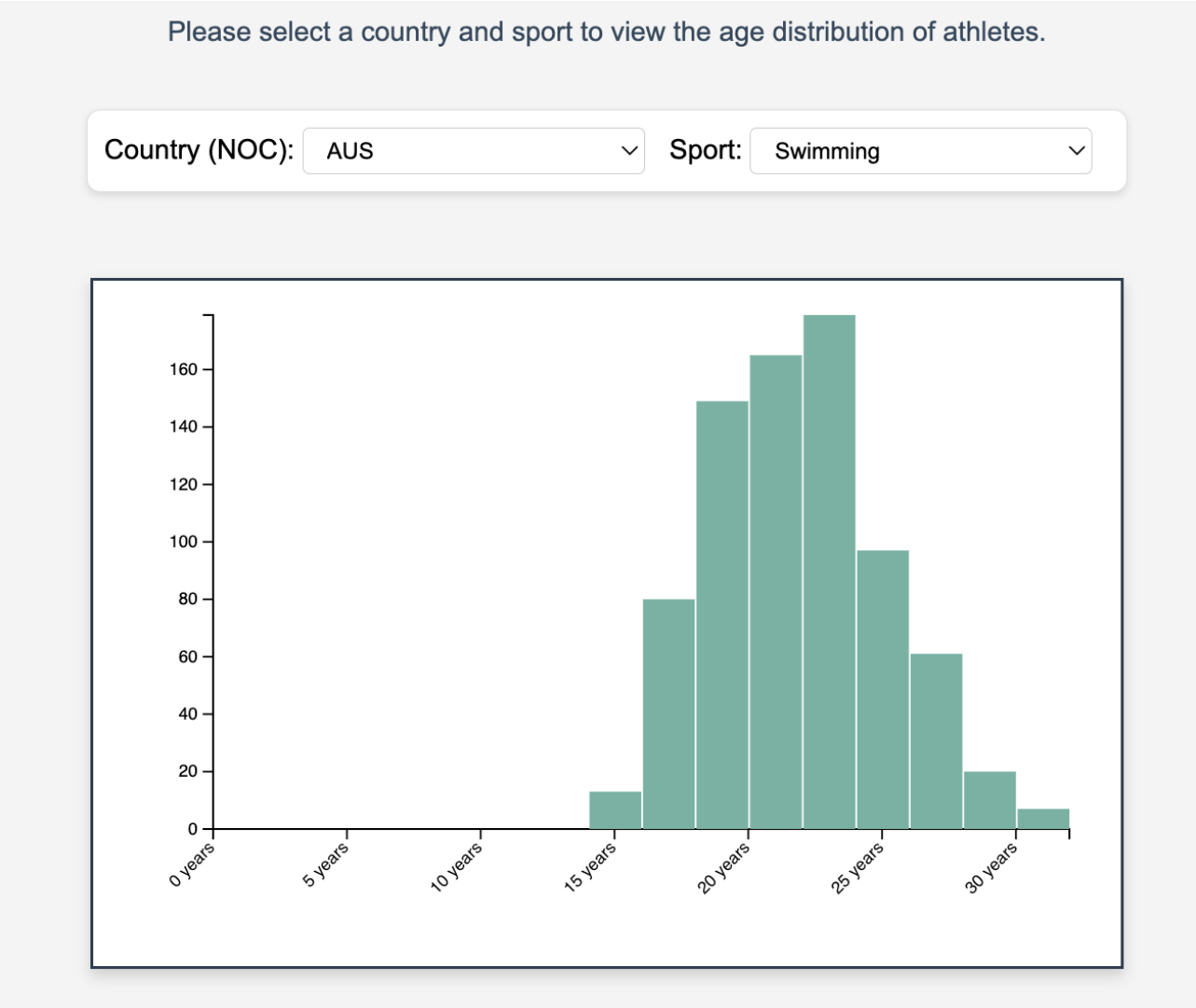
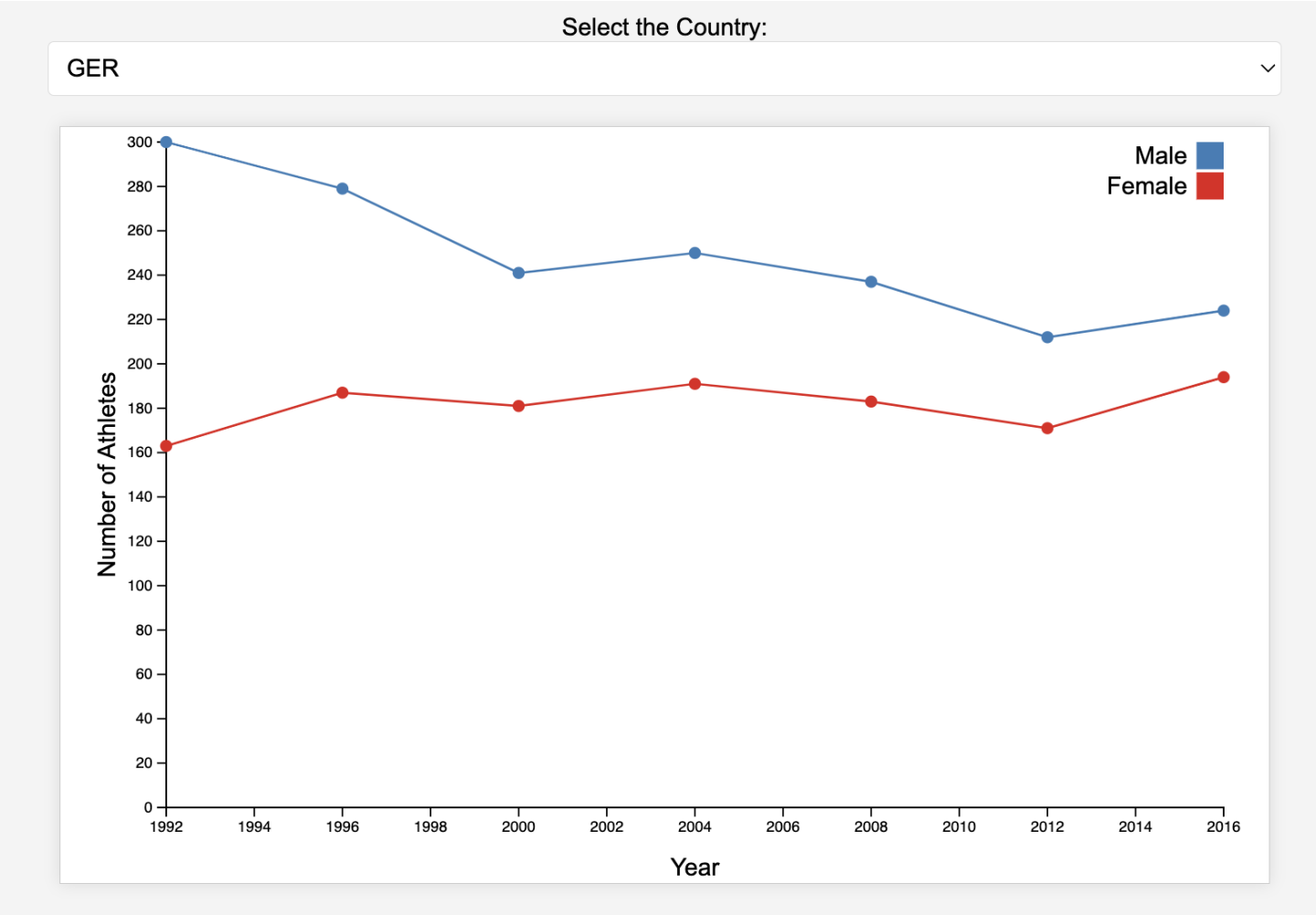


Fig. 5: Athlete Age Distribution for Australian Swimmers in Olympic Events

6. Trends in Gender Participation:

Over time, Germany's gender participation trends in the Olympics indicate a slight decline in male participation and a slight increase in female participation.

- There were 181 female and 241 male participants from Germany in 2000.
- The number of male participants dropped to 224 by 2016, while the number of female participants rose to 194.



Fig, 6: Gender Participation Trends in German Olympic Teams (1992-2016)

Overall, there was a slight decrease in the number of men competing, but a rise in the number of women, suggesting that German Olympic teams were becoming more inclusive of women during this time.

7. Medal Distribution by Sport and Country:

Japan had some interesting achievements in several sports at the 2008 Olympics. In particular:

- **Softball**: Japan proved their supremacy in this sport by winning an amazing 15 gold medals.
- **Gymnastics**: The Japanese squad demonstrated their skill in gymnastics events by winning seven silver medals.
- **Wrestling**: Japan proved their mettle in this age-old Olympic sport by winning two gold medals in the sport.
- **Swimming**: Japanese athletes demonstrated their competitiveness in the pool by winning six bronze medals in swimming events.

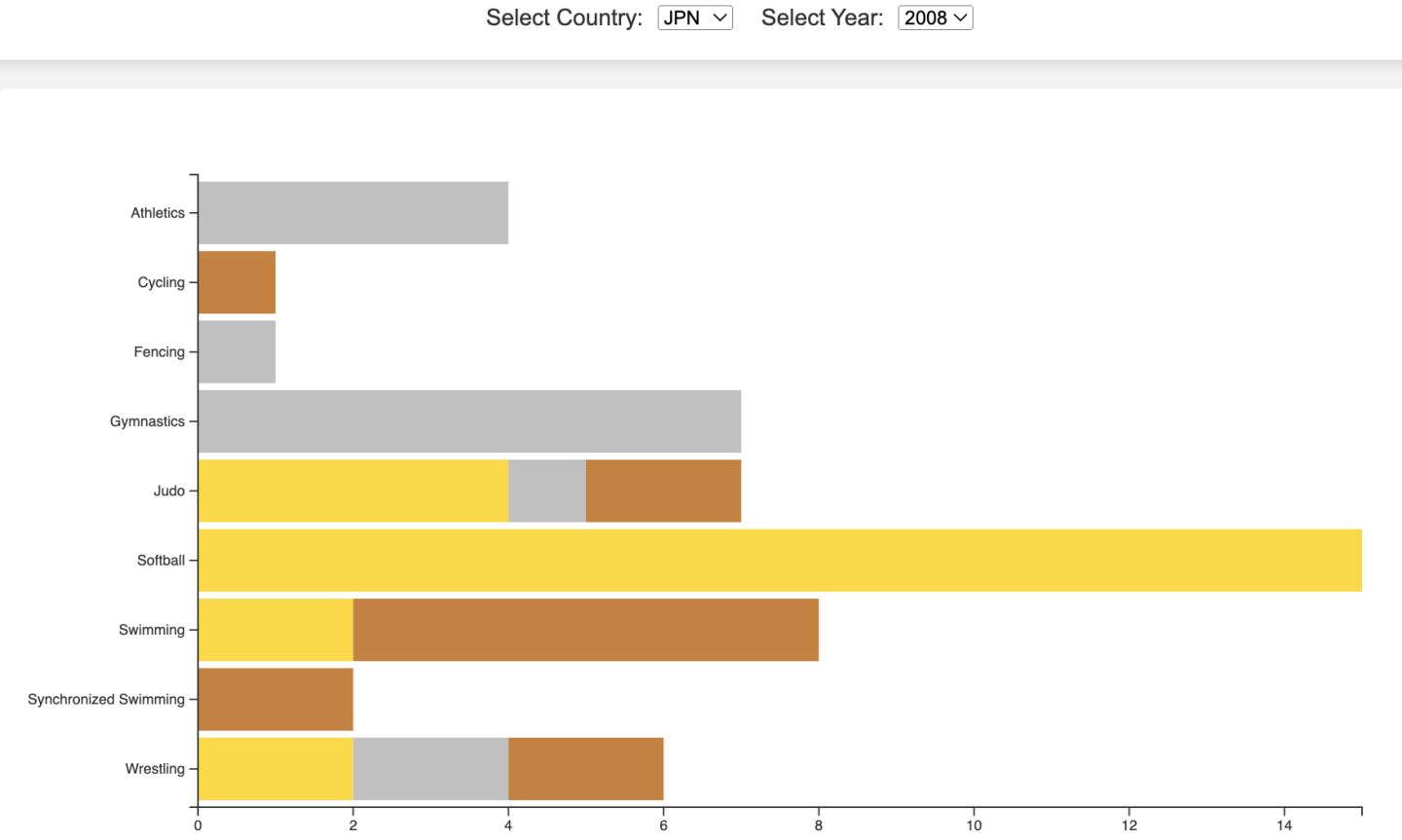


Fig. 7: Medal Counts for Japan Across Different Sports in the 2008 Olympics

Japan's overall medal totals in these sports highlighted their excellent Olympic performance in 2008 and enhanced their reputation in the world of sports.

8. Medal Distribution by Sport and Country:

France made significant achievements in different sports at the 2000 Olympics. Particularly, France performed well in the sport of basketball, winning 12 silver medals. In addition, France demonstrated their skill in gymnastics by winning two silver medals.

France demonstrated its excellence in cycling at the 2000 Olympics, winning an astounding seven gold medals in the sport. This incredible performance highlights France's strength in the cycling competitions at the Games.

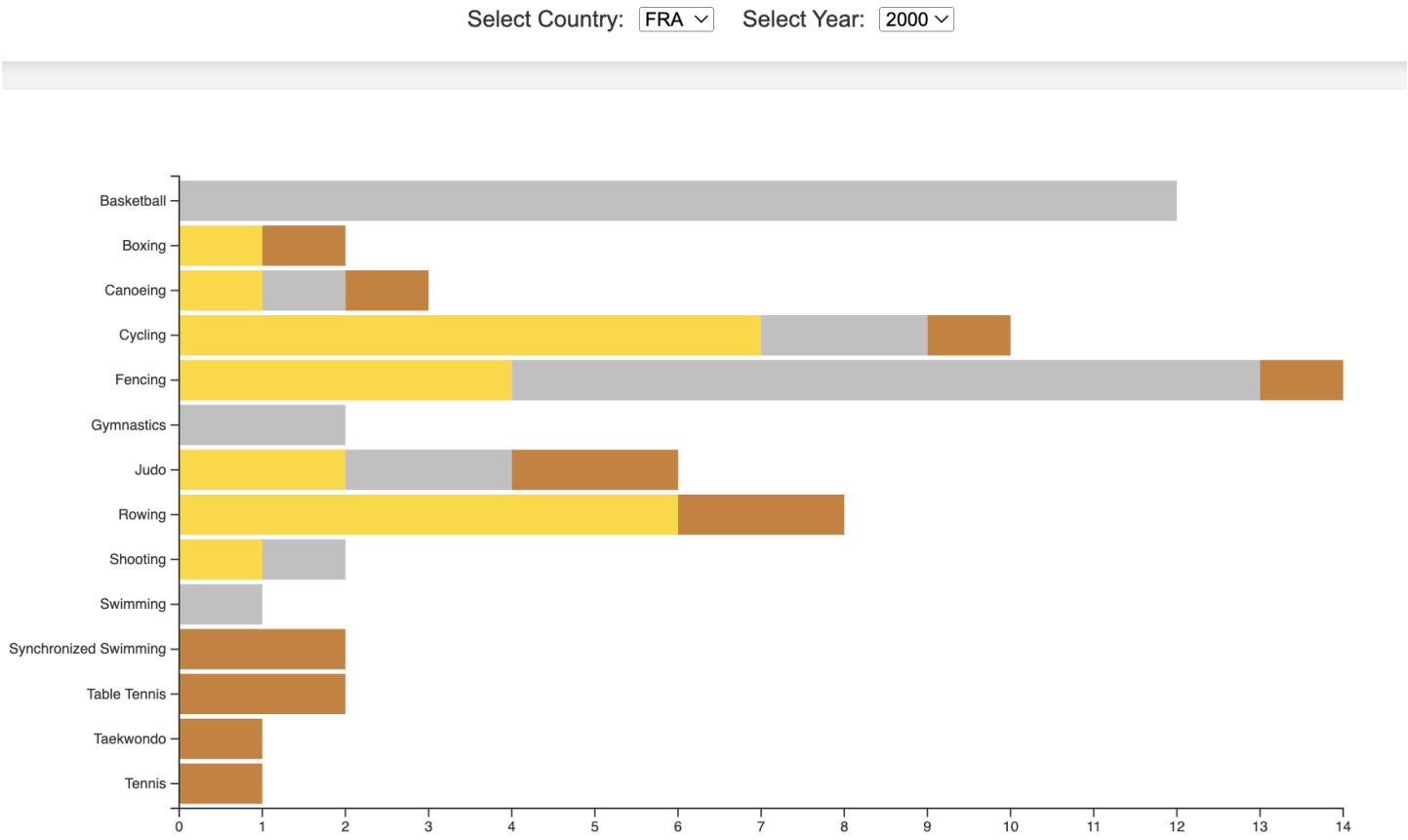


Fig. 8: Medal Distribution for France in the 2000 Olympics

In addition, France achieved success in the sport of table tennis, winning two bronze medals. This accomplishment demonstrates France's wide range of athletic talents and competitive strength in several Olympic sports during the 2000 Games.

9. Leading Participating Nations:

About 13.3% of all medals awarded at the 2012 Olympics went to Qatar (QAT), Portugal (POR), and China (CHN), according to the pie chart that shows the top 10 countries by medal counts. The United States of America (USA) came in second place with 11.9% of all the medals. North Korea (PRK) and Australia (AUS) received 8.0% and 7.8% of the total medals, respectively, while Japan (JPN), Canada (CAN), and Jamaica (JAM) each received between 8.1% and 9.2%. Rounding out the top 10 with 6.7% of the total medals was Italy (ITA). This distribution emphasizes the variety of nations participating in the 2012 Olympics and the international scope of the Games' representation of sporting excellence.

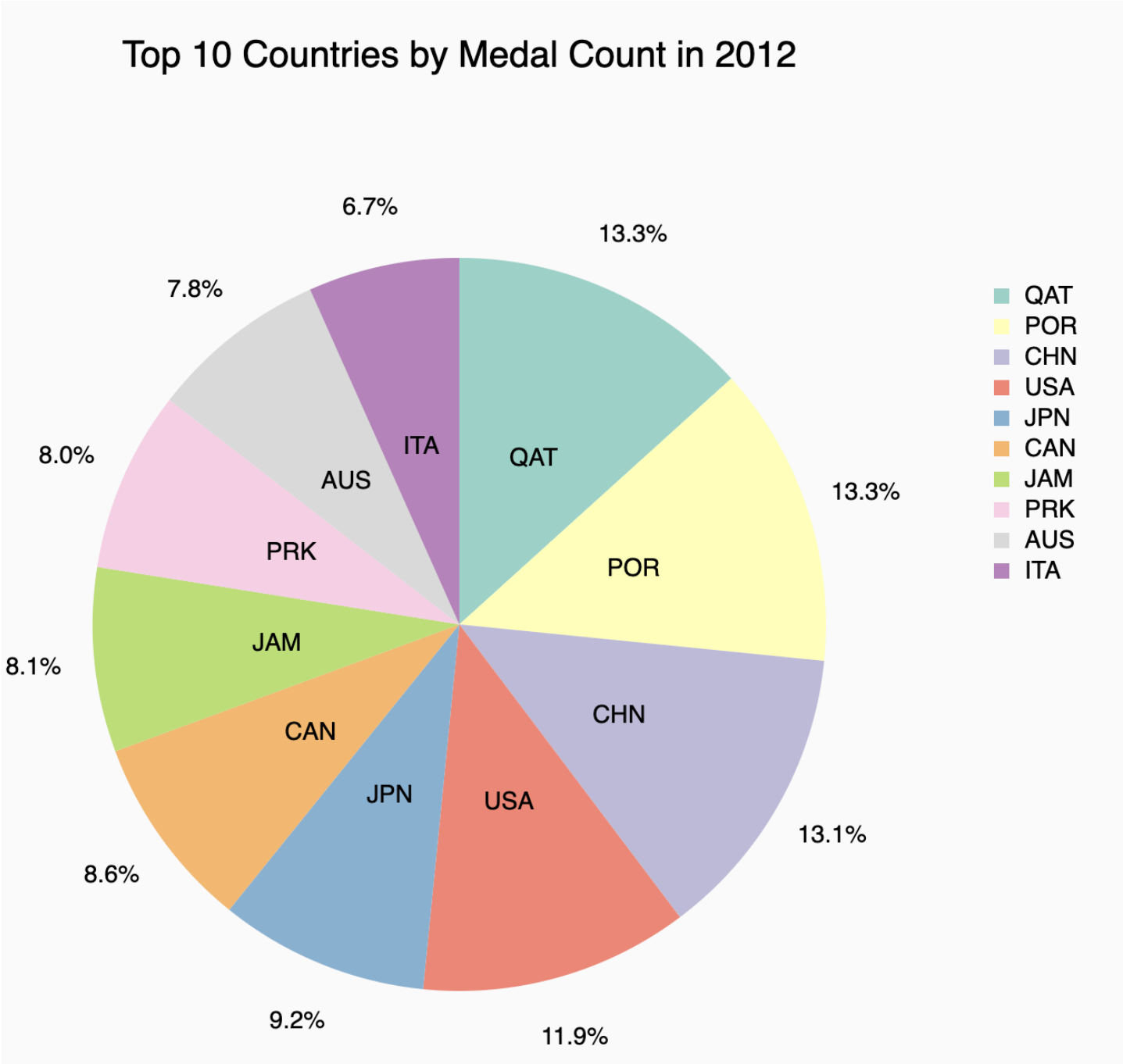


Fig. 9. Top 10 Countries by Medal Counts in the 2012 Olympics

Conclusion

This project's analysis of Olympic trends offers insightful information about how international sports competition is changing. We have gained a better understanding of the intricate dynamics that shape the Games by looking at gender representation, age distribution, and medal results over a century of Olympic history.

Although differences still exist, our analysis has shown encouraging progress toward gender equality in Olympic participation. The analysis of age distribution has identified the different career paths taken by athletes and the impact of sporting culture on performance. The examination of medal distribution additionally highlighted trends of dominance between nations and sports.

In general, this project adds to the current discussion about gender equality, and the wider social effects of the Olympic Games. Through the utilization of data-driven analysis and visualization methodologies, we have yielded significant insights that may guide forthcoming discourse and endeavors dedicated to advancing inclusivity and excellence in sports.

Project Structure and GitHub Repository

Folder Structure:

To ensure readability and clarity, the project is organized into folders in a structured manner. The following is the folder structure:

- **css**: The stylesheets (CSS files) used to style the web visualizations are contained in this directory.
- **js**: The JavaScript code files in this directory enable functionality and interactivity in the web visualizations.
- **data**: Datasets and data files used for analysis and visualization are kept in this directory.
- **HTML files:** Every HTML file in the project is associated with a particular web page or visualization.

GitHub repository:

Version control and teamwork are facilitated by the project code and associated files being hosted on GitHub. The Olympic Analysis repository on GitHub can be accessed at https://github.com/yourusername.

You can access the project code, data, and documentation by exploring the repository.

References

1) Kaggle Dataset: Heesoo37. (n.d.). 120 years of Olympic history: athletes and results. Kaggle. Retrieved from https://www.kaggle.com/datasets/heesoo37/120-years-of-olympic-history-athletes-and-results
2) GitHub Repository: [Author's Name]. (Year). Repository title. GitHub. Retrieved from [URL]
3) JavaScript: Mozilla Developer Network. (n.d.). JavaScript. MDN Web Docs. Retrieved from https://developer.mozilla.org/en-US/docs/Web/JavaScript
4) D3.js: D3.js. (n.d.). D3.js Documentation. GitHub. Retrieved from https://github.com/d3/d3/wiki