Data 5310

Final Project Report

Divyasree Vammigari

Garrett Ringler

Saluwa Umuhoza

Ruqhaiya Fatima Syeda

Olympics Participation and Economic Trends

Introduction

The Olympic Games are a global stage where nations showcase their athletic and cultural identities. Yet, success at the Games often reflects more than just athletic prowess. It is influenced by a nation's economic resources, policies, and resilience. Understanding this connection provides valuable insights into how opportunity and investment shape global sports success.

This project explores these dynamics by merging two key datasets:

- **Olympic Data**: Detailing performances from 1960 to 2023, including medal counts, participation, and demographics.
- **GDP Data**: Providing annual GDP figures for participating nations over the same period.

By integrating these datasets, this analysis addresses key questions such as:

- Do wealthier nations consistently dominate the Olympics, or do smaller nations sometimes outperform expectations?
- How does economic investment in sports influence medal counts?
- What cultural and regional factors contribute to success in specific sports?
- How has Olympic performance evolved over time as national economies have grown or shifted?

While this analysis offers valuable insights, it also faces challenges, including gaps in historical GDP data and variations in the number of events across sports. Despite these complexities, the visualizations in this report aim to tell a clear and engaging story, designed to be accessible to a general audience.

Ultimately, this report seeks to go beyond the numbers, uncovering the resilience and resourcefulness of nations that excel against economic odds and revealing the deeper stories behind Olympic success.

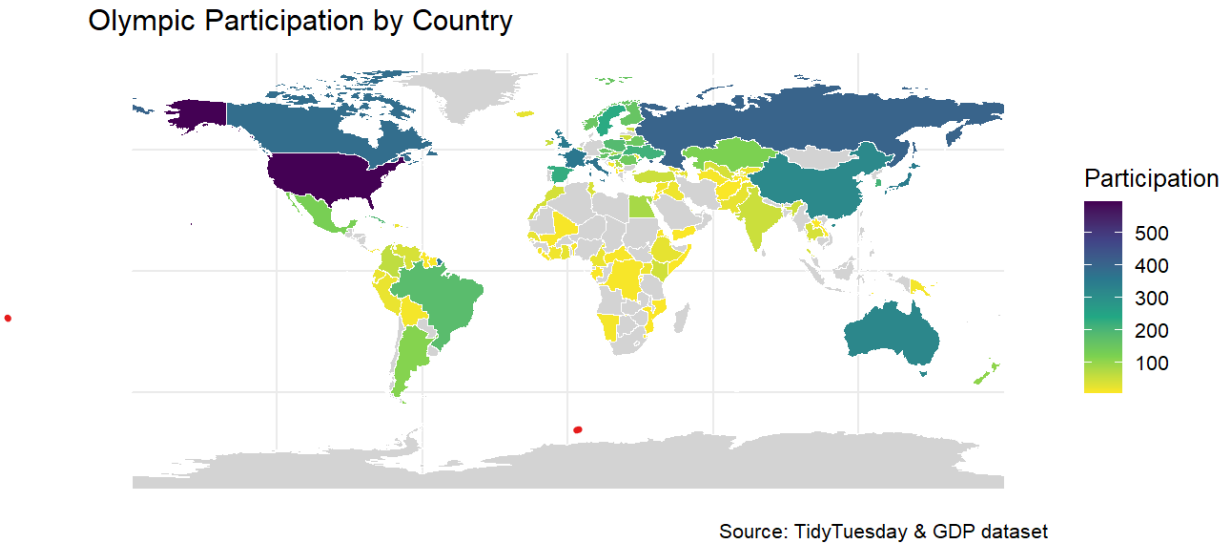
Data Methodology

The Olympic data used for this report was gathered from Kaggle and the GDP data was sourced from World Bank. Both of these data sets were relatively clean already so minimal work was required to prepare the data for this project. There were only a couple of key changes that needed to be made which could be achieved purely by filtering the data. The first item being clearing out incomplete early Olympic data. The earlier years of the Olympics had many missing values, so the team decided it was best to filter out the data. Since the plan was to merge with a GDP data set that included data from 1960 onwards, the team decided to just filter the Olympic data to 1960 and beyond as well. This filtering accomplished many of the things we had to do to prepare for this project. Other minor filtering and manipulation was done for certain graphs such as sorting by top countries and sports with the highest medal counts and taking the log values for GDP to further dive into the lower GDP country trends. The final step we took in this process was merging the

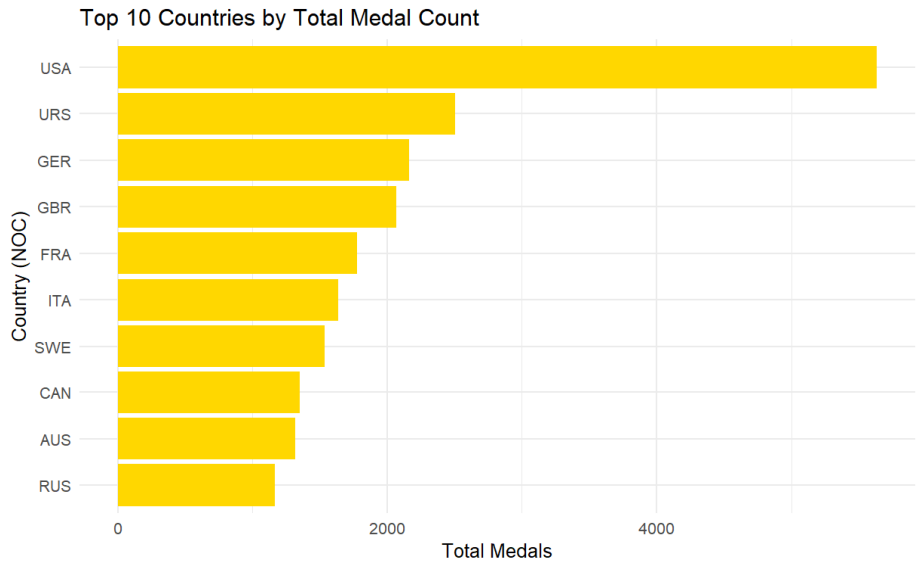
Olympic and GDP data which required making sure the country names lined up. After all these steps were taken, the data was in a great state for full analysis for this project.

Results

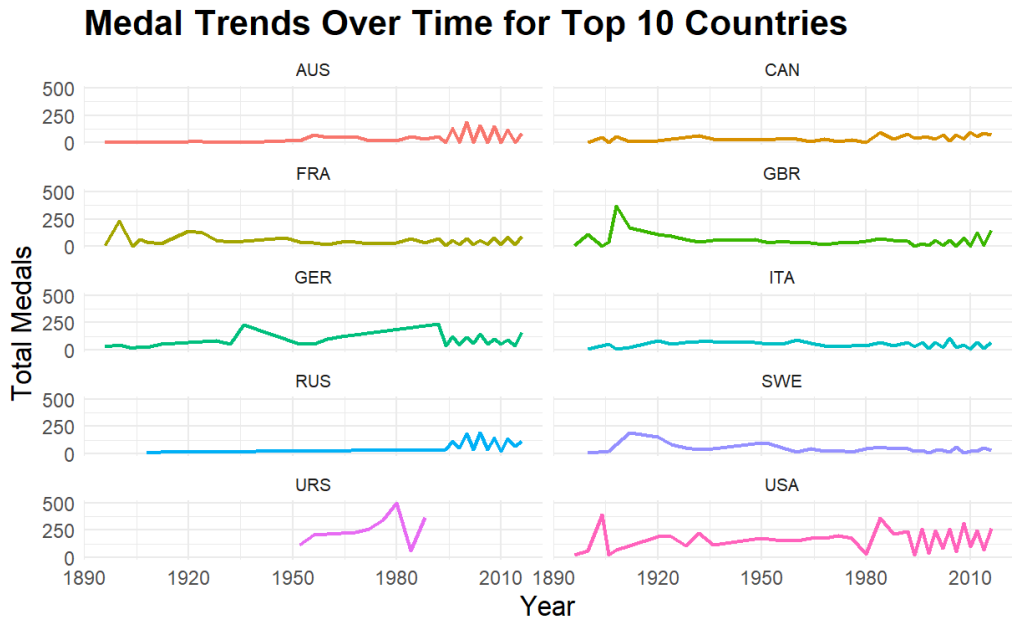
The following map displays the Olympic participation of countries around the world using a choropleth visualization. Countries are shaded based on their number of participants, with darker purple indicating higher participation and lighter yellow-green showing lower participation. Grey shading represents missing data for some countries. The color scale uses the "viridis" palette for clear distinction and accessibility. Overall, this showcases key countries that tend to have high participation.



Continuing into country participation and earnings. The bar chart below ranks these top 10 countries by their total Olympic medal count, with the USA leading by a substantial margin. Following the USA are former USSR, Germany, and Great Britain, showing a strong history of athletic success. The bright yellow bars highlight the dominance of these nations in Olympic history, emphasizing their consistency in winning medals over time.

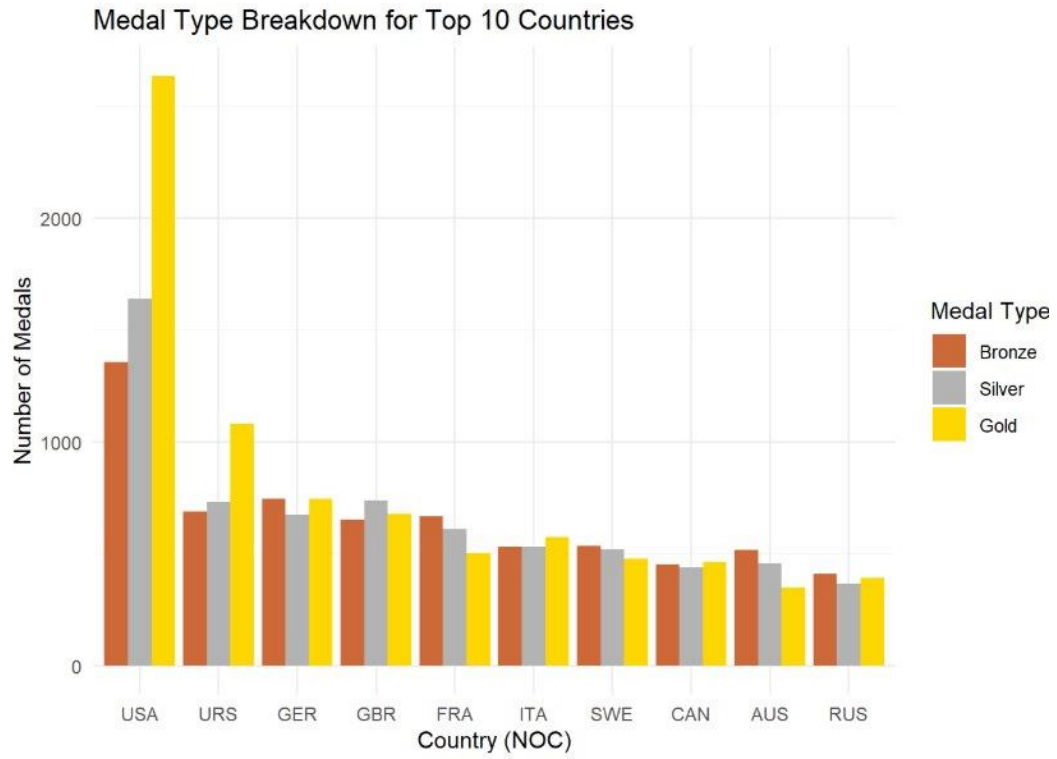


The next set of graphs showcases the constant upward and downward trends in medal gains even by the best countries. The graphs showcase the 10 countries with the most total medals. After adjusting the scale so that each graph has the same Y axis scale we can see how flat some of the lines are at a larger scale while the USA and Russia/USSR tend to consistently do well in medal counts. We can also see where Winter Olypmics split affects the timeseries here as an option to further explore in a different project.

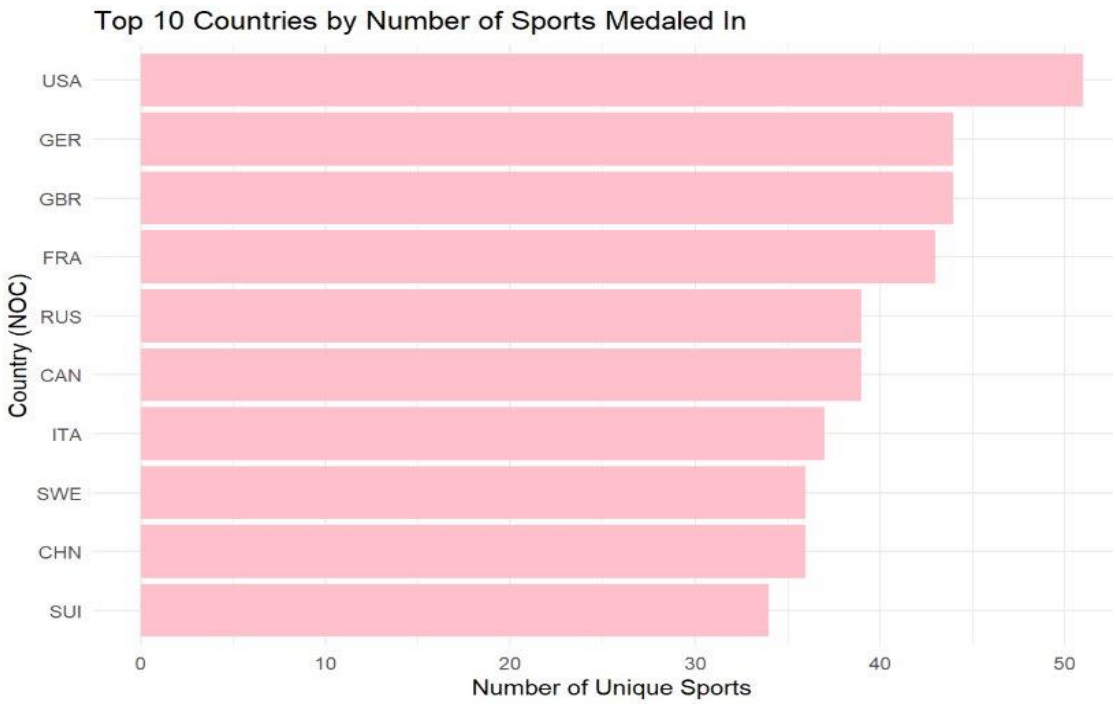


The Winter vs Summer Olympic split leads to the next graph which illustrates the sports that tend to be more prevalent in medal counts. Here we see that Swimming, Rowing, and Athletics are the top three sports yielding medals each Olympic year while sports like Triathlon and Skeleton yield the lowest count each Olympic year. There are a couple of factors we can consider for why this may happen. The first factor is likely the number of events within that sport that happens every Olympic year, sports like Swimming and Athletics have more events while Triathlon is a longer event that has less total events. This combined graph also helps illustrate a potential problem that the years for Winter and Summer Olympics went from being the same year prior to 1994 to being split up in 1994 and having them alternate every 2 years versus doing both every 4 years. Which of the Winter sports only Ice Hocky has high medal counts. These factors lead into the next graph which illustrates the medal types by those top 10 countries. The Winter vs Summer split could lead to lower counts for countries like Australia.

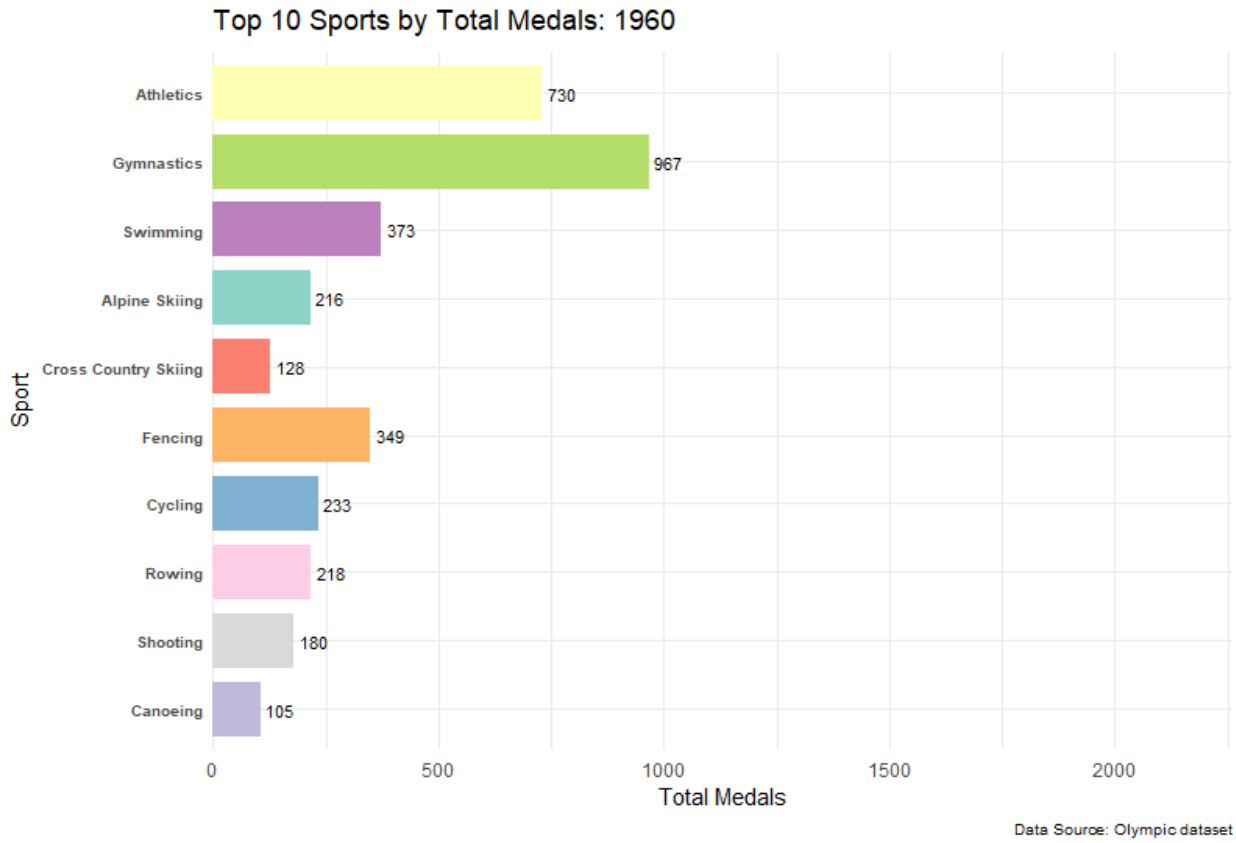
This next graph illustrates the medal breakdowns of the top 10 most medaled countries. With this graph we can see the distribution of medals from the previous set of graphs to help continue the story of which countries tend to win by showcasing the level of winning. For example, Australia is one of the countries with the highest medal count but most of those medals seem to be Silver or lower while the USA wins more golds by far. Which again looking at the previous graph many of these medals are from the Summer Olympics which with more Winter events could swing some of these bars.



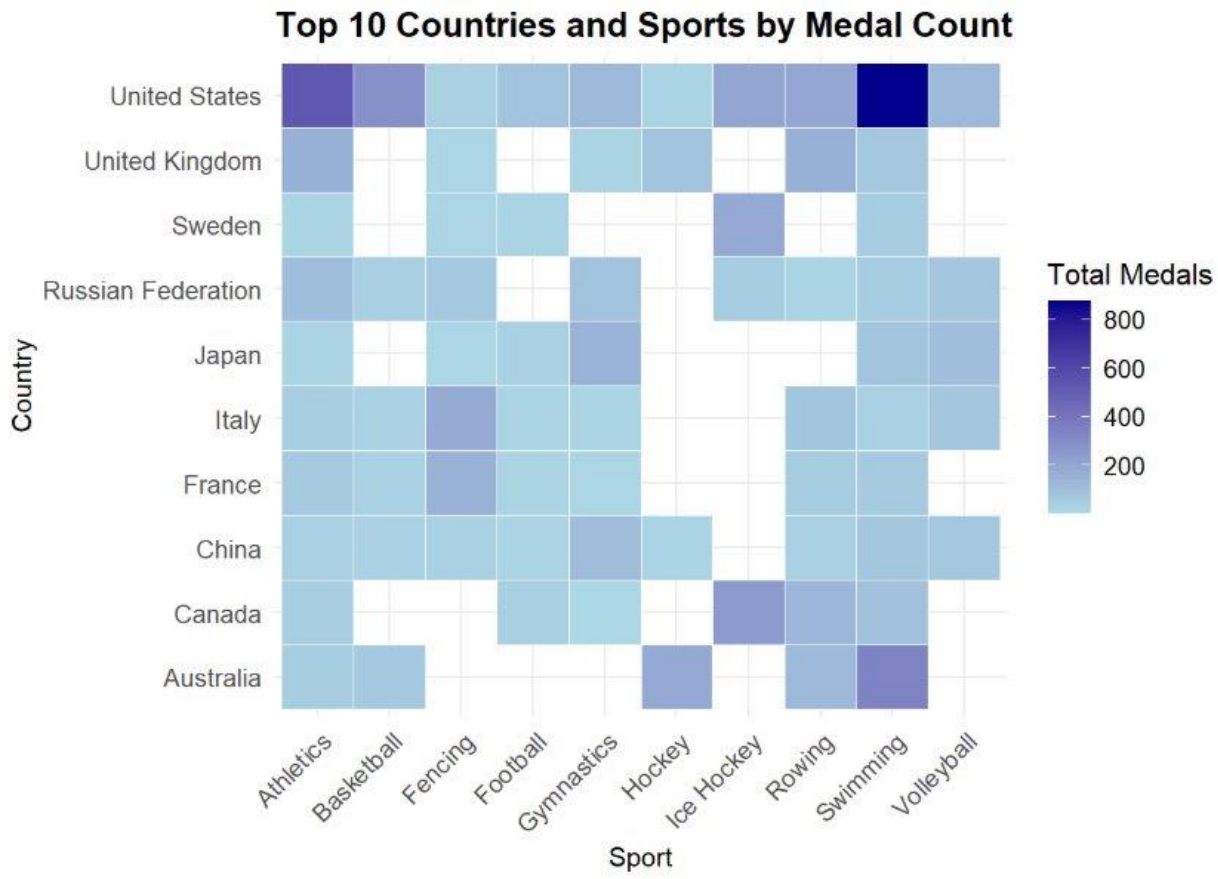
Moving more into sports distributions, the bar chart below ranks countries by the number of unique sports in which they have medaled. The USA leads, showcasing its dominance across a diverse range of sports. Germany and Great Britain follow, emphasizing their strong presence as well. This chart highlights the versatility and depth of certain nations in the Olympics.



The next graph is an animated bar graph which shows the top 10 sports by total medal counts from 1960 - 2016. The dominance of specific sports can be clearly seen, like Gymnastics, Athletics, and Swimming always top the lists. In winter sports, Alpine Skiing and Cross Country Skiing emerge quite at the top, signaling their importance during the Winter Olympics. The animated nature of the graph provides insights into how the Olympics have grown and diversified across the decades.

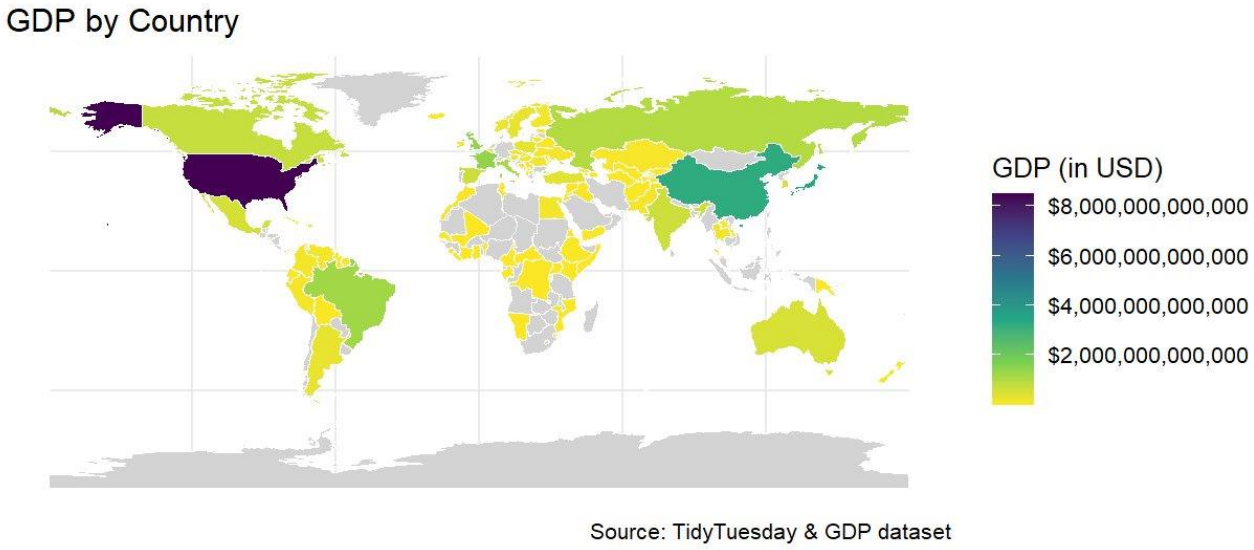


Moving into a heatmap shows the total Olympic medals won by the top 10 countries across the top 10 sports. The y-axis lists countries, while the x-axis shows sports, with darker blue tiles representing higher medal counts. The United States dominates several sports like swimming and athletics, evident from the dark blue tiles. Some sports, such as gymnastics and hockey, show more balanced distributions among countries. Blank or lighter tiles indicate fewer or no medals for a specific country sport combination.

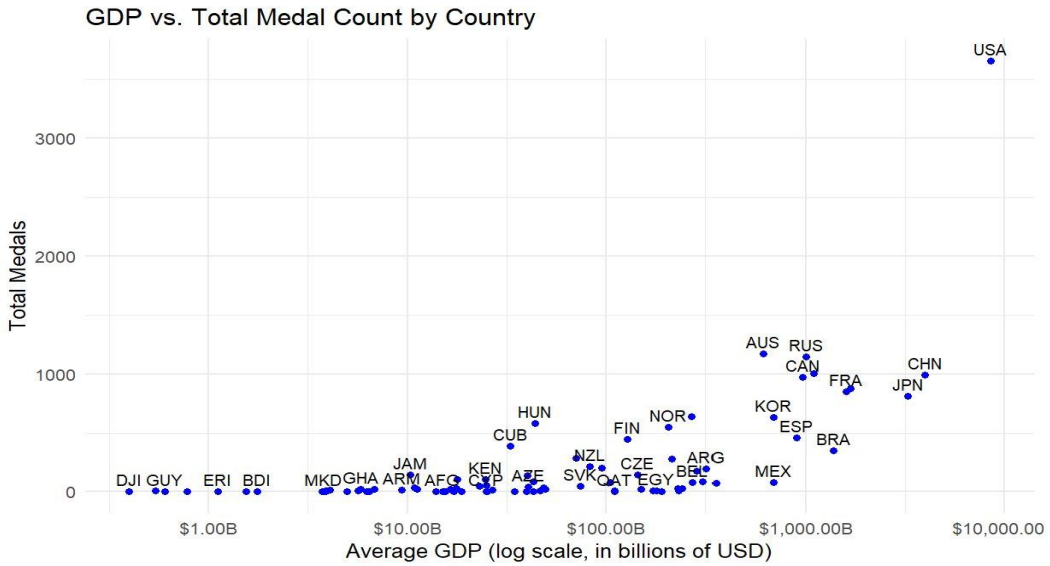


Wrapping things up, the next map displays the GDP of countries around the world using a choropleth visualization. Countries are shaded based on their GDP in USD, with darker purple indicating higher GDP and lighter yellow-green showing lower GDP. Grey shading represents missing data for some countries. The color scale uses the "viridis" palette for clear distinction and accessibility. This graph showcases similar trends to the first map which could mean GDP is a potential factor in participation/performance. With higher GDP comes higher potential money spent on training athletes. The next graph will dive deeper

into this.



The scatterplot examines the relationship between GDP and total medal count by country. It reveals a positive correlation and indicates that wealthier nations, such as the USA and China, tend to dominate in terms of medal counts. The USA is an outlier with both an exceptionally high GDP and medal count, which highlights the role of economic resources in fostering athletic success.



Discussion

Through this study, we have highlighted a compelling relationship between a nation's economic resources and its Olympic performance. The positive correlation between GDP and medal counts was evident and it underscores the significant role that wealth plays in bringing athletic success to a country. Wealthier nations such as the USA and China, consistently dominate the Olympics whether it's Summer or Winter sports. We strongly believe that their success is driven by their ability to invest in training, infrastructure, and sports development. Then we have also seen that smaller nations with limited GDP, like Djibouti or Guyana, face substantial challenges in achieving similar levels of success.

That being said, there are exceptions to this belief like Australia that has relatively smaller GDP compared to China or Japan, but it still leads in the Olympics with a higher medal count. This is where we think that other factors like cultural priorities, government policies, and global events also play a huge role in shaping a country's Olympic performance. Similarly, Finland, Hungary and Norway also represent another such example where they have a higher medal count but Mexico whose GDP is higher, has seen less success in the Olympics.

The visualizations also reveal some intriguing dynamics that are beyond economic resources. For instance, the dominance of sports like swimming and athletics in medal counts reflects the number of events in these sports along with cultural and historical factors that shape nations' focus on specific disciplines. Likewise, the change to alternating Winter and Summer Olympics after 1994 has impacted the medal distributions, especially for Norway and Switzerland that excel in Winter sports.

However, this analysis has a few limitations. There are gaps in historical GDP data and missing Olympic records for earlier years (before 1960) which limited the scope of the dataset. A potential bias is introduced when comparing countries or disciplines because of the variations in the number of events across sports. Addressing these issues would require a combination of different datasets that include details about the population size, government's investment on sports, and cultural attitudes toward specific sports.

Further research could unlock more subtle insights by exploring:

- The role of population size and per capita GDP in Olympic performance.
- Representation of gender in Olympic participation and medal tally.
- The influence of political or international events, like boycotts, on participation and success.
- Temporal trends of emergent sports and their impact on the general distribution of medals.

This study lays the groundwork for understanding the multifaceted factors that influence Olympic success. By exploring these avenues, we can gain deeper insights into how nations overcome economic and structural barriers to excel on the world's largest sporting stage.

References

- World Bank. (n.d.). GDP (current US$). Retrieved from https://data.worldbank.org/indicator/NY.GDP.MKTP.CD
- Dattabiplab. (n.d.). Olympics analysis dataset. Retrieved from https://github.com/dattabiplab/olympics-analysis