The History of the Olympic Games

Spring 2020 - Final Project

EM 224: Informatics and Software Development

May 12, 2020

Regan Tarasewicz

Overview of Research Question

Due to the recent outbreak of the coronavirus, people all across the globe have experienced cancelations in all aspects of their lives. One major change has been the lack of sports, impacting both the athletes and their fans. Knowing that so many people are missing out on sports, including the Olympics which were scheduled for this summer, I wanted to focus this final project on the history of the Games. I wanted to explore the evolution of the Olympics from the beginning, 1896, up until the modern Games. In this paper will be graphs depicting how the Olympics have grown over the years through the number of events played for both the Winter and Summer Games. There will also be visuals for the five all-time winningest countries, or teams, with their medal counts and insights on their more successful years.

Dataset Description

The dataset used to visualize the trends in this paper is titled "120 years of Olympic history: athletes and results." The data itself was published on www.sports-reference.com and then collected into a csv file, which was uploaded to www.kaggle.com. This dataset is about the individual athletes that competed in events between 1896 and 2016, from both the Summer and Winter Games. Each row describes an athlete with the following: ID, Name, Sex, Age, Height, Team, NOC (National Olympic Committee), Games, Year, Season, City, Sport, Event, and Medal. There are 271,117 total rows in the dataset. The direct link to the dataset is the following: https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results/data

Dataset Provenance

The following steps were taken with the dataset for analysis:

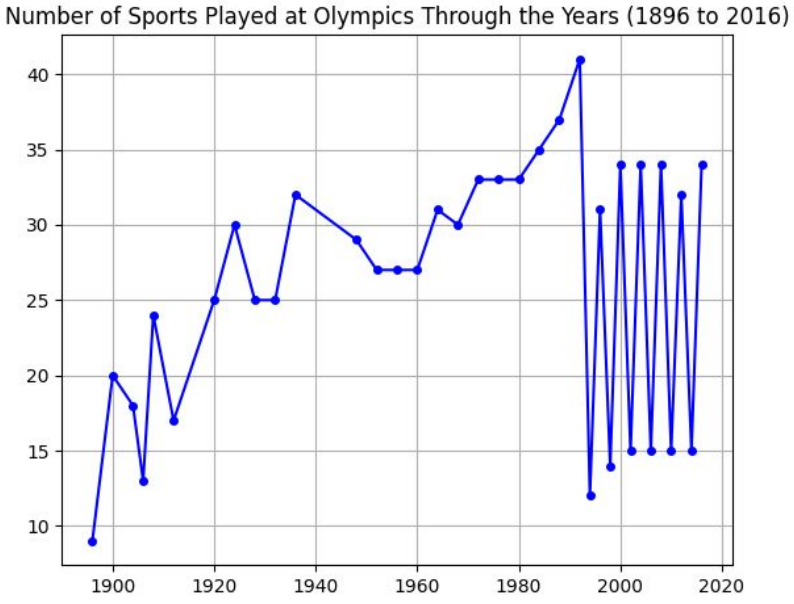
1. The dataset 'athlete_events.csv' was downloaded from the site.
2. Using the pandas library, the csv file was opened and the necessary attributes were collected.
    a. Teams, Year, Season, Sport, and Medal categories
3. The data frame was manipulated to access different values for each analysis.
4. Different plots were created to best describe the research questions.

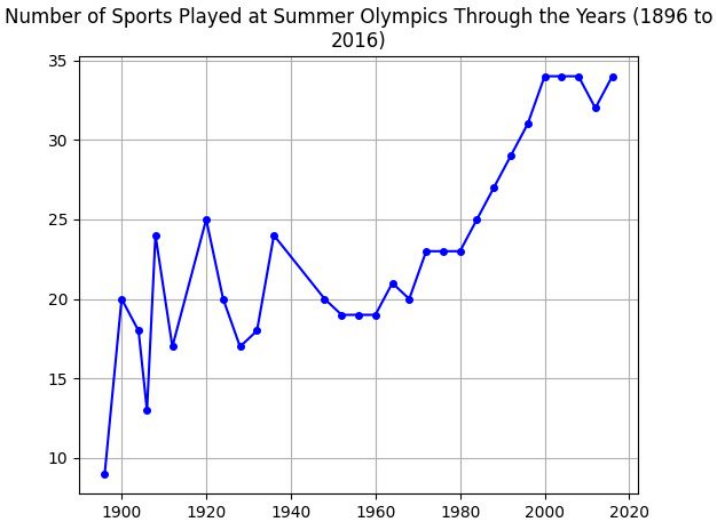
Methodology

Once I had the columns needed for this study, I went ahead with further manipulations for each set of data I wanted to collect. In order to see the different number of sports per year, I needed to drop any duplicated rows with the same year and sport. This allowed for each sport to be represented once per year, and then I could count the frequency of the years to get their total sport counts. I used a similar method for only the Summer and Winter Games, however there was an added manipulation of removing all rows that did not have the appropriate season. For the medal count analysis, I took the original data frame that was read in the code and needed to remove any NA's from the medal column so that only winning teams remained. From that I was able to get the most common team names and plot their total medal counts. With the team names saved, I was then able to manipulate the data frame further to only ever have one of those teams still listed and then took the count of each type of medal: gold, silver, or bronze for those distributions. Also with that, I used a similar process as the first analysis and took the frequency at which each year still existed for those countries to get the distribution of when those teams had the most success. Each of these plots were created using the matplotlib.pyplot library, showing bar graphs, line graphs, and pie charts.

Results

In the graph on the right can be seen the results from the number of sports played at each Olympics over the years. It's important to note that until 1992, both the Summer and Winter Games were played in the same year, heightening their total game count. Due to this seemingly skewed collection of data, I decided to split the games up by the season.

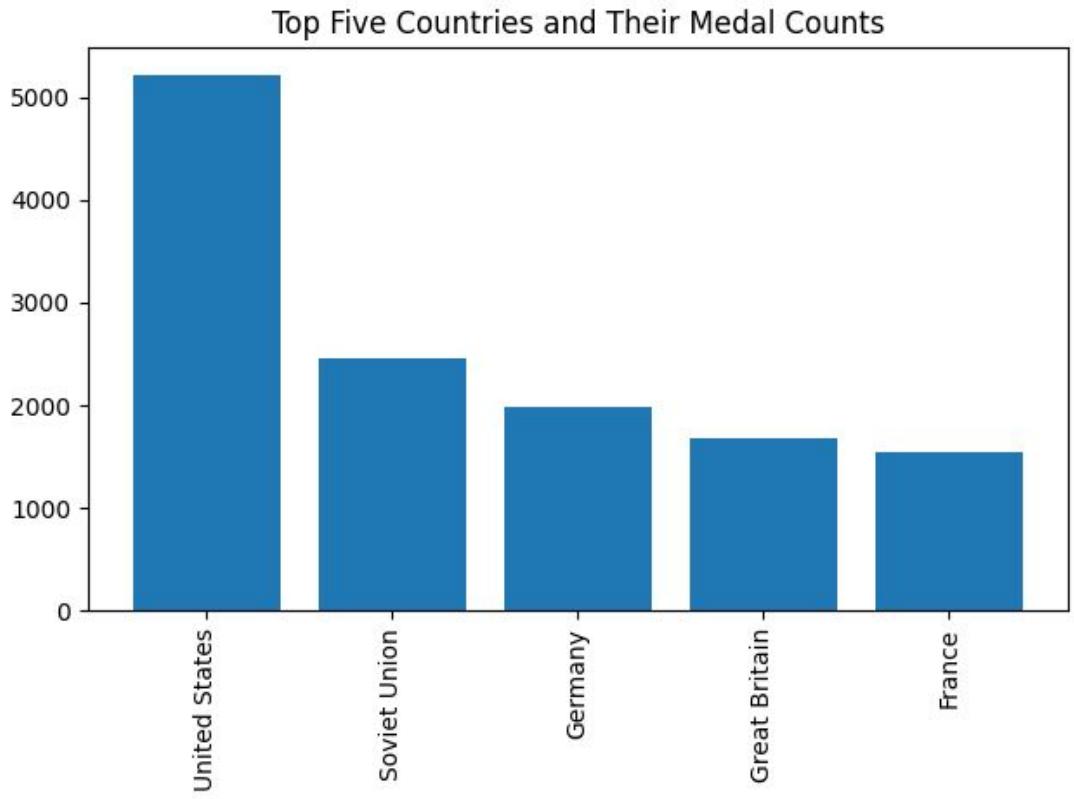




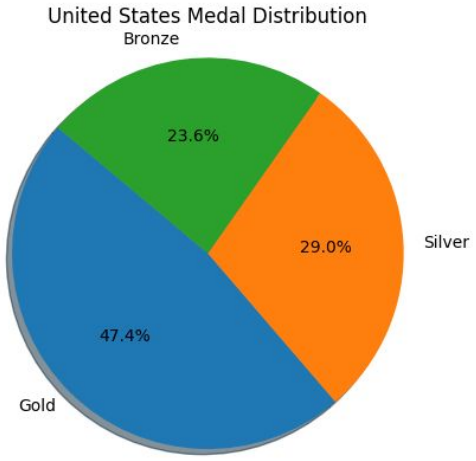
The first graph shows the Summer Olympics. It is plain to see that the overall trend of the amount of sports played has increased over the years, starting at only 9 back in 1896 growing to 34 in 2016. Most of the drops in the amount of sports played are due to outside influences, such as different wars going on at the time. The second graph depicts the Winter Olympics data. This information only goes back to 1924 with a start of 10 sports, as these Games were not introduced as early as the Summer Games. In both of the graphs, there are significant increases in the 1990's, due to the Games being split to different years, allowing each to expand on their own sport count. One last thing to note about this information is that these are the total sport counts, not event counts.

The next study focuses on the all-time most successful countries based on total medal counts. This dataset is historically accurate with stating each country's name as it was the year they played, for example Russia is not grouped with the Soviet Union as they represented different NOCs. Also, from 1972 until 1988, Germany was split into East and West Germany, and therefore the counts from those years are excluded. Below is the visualization of the top five

teams and their medal counts being, United States with 5,219, the Soviet Union with 2,451, Germany with 1,984, Great Britain with 1,673, and France with 1,550.

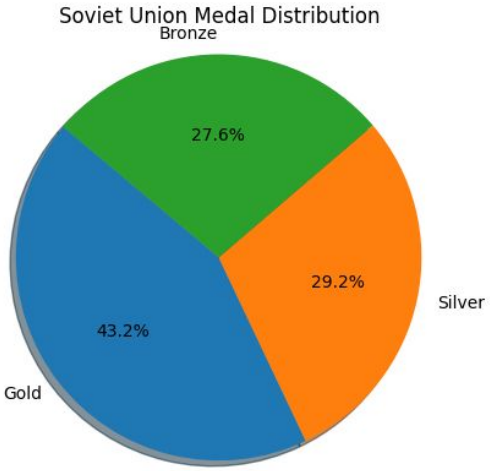


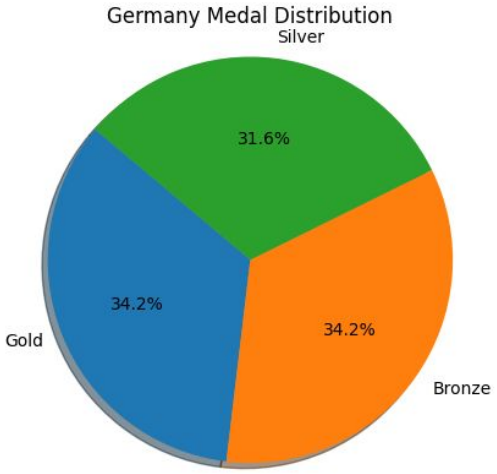
For a deeper look into the medals won by the top three of these countries, the following three pie charts show the distribution of the medal type since the success shown in the previous plot does not discriminate between the three medals. First are the statistics for the United States.



Gold medals contribute to 47.4% of the total medals, which is 2,474 of them. For silver, they won 1,512 medals and lastly were 1,233 bronze medals.

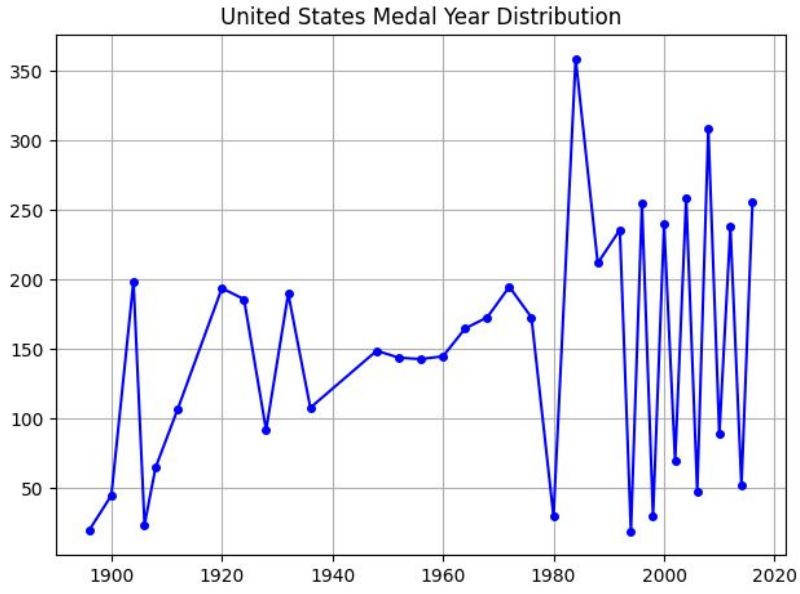
The Soviet Union showed a similar distribution with 1,058 gold medals accounting for 43.2% of their total. Silver medals were also the second



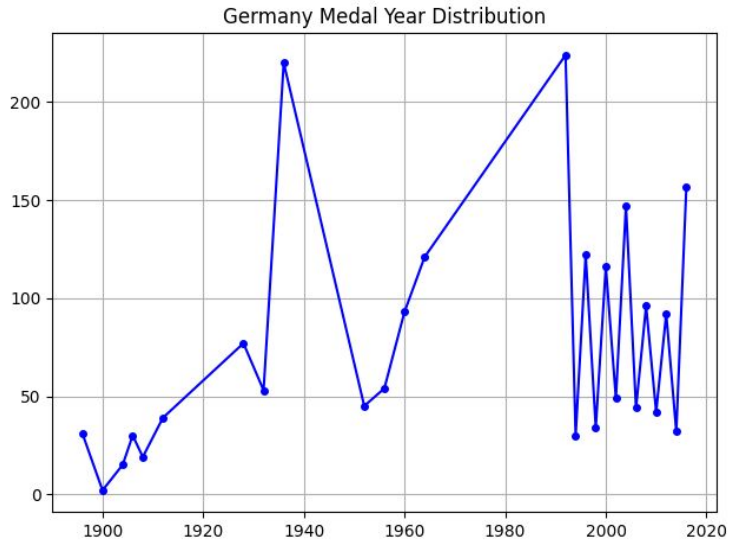


most common medal for them at 716. Finally, they had 677 silver medals. As for Germany, their gold and bronze medals both account for 34.2% of their medals, contributing 679 and 678, respectively. Close behind in count were their silver medals at 627 for the final 31.6%. These statistics show that the success of a team is due to all placing athletes, not simply the gold winning performances.

The next analysis of the medal counts answers when these teams were most successful, particularly the top three. Each point on the lines in the next three graphs represent a year the team participated in the games. First on the right, the distribution of the winnings for the United States. In the 35 years accounted on the graph, the United States had the most success in 1984 with 359 total medals won and their least successful year came 10 year later in 1994 when they only won 19 medals. Their success in 1984 has been attributed to not only a hometown



advantage in Los Angeles, but also the boycott that the Soviet Union led that year in response to the 1980 boycott led by the Americans. The spike for the Soviet Union winnings in 1980 can be seen in the next chart. That year they won 488 medals, over 100 more than any other year.



Although the Soviets had less overall success than the United States, their success came from only nine games. Given the total medals by each of these two teams and the amount of games they won them in, the Soviet Union won an average of 272 medals per game while the United States only an average of 149 medals per game. Lastly, Germany experienced two high winning years in 1936 and 1992, with 220 and 224 medals won. The 26 games that Germany participated in as a complete country, they won an average of 76 medals per year, significantly less than the Soviet Union or the United States. Overall, the down spikes seen in both plots for Germany and the United States are due to the inclusion of both the Summer and Winter Games. Since the Games had been played in the same year prior to 1992, medal counts were more steady between years. The first charts in this paper showed that the number of sports played in the Winter Games is less than in the Summer Games, accounting for the lower counts for the winter years.

Conclusion

Through this analysis it can be seen that the Olympic Games grew from the early years until it mostly stabled out around 2000. With the increase in the number of sports there was also a general upward trend in the number of medals won for the top countries, since the opportunity to win at more sports increased. While plainly the United States have received the most medals all time, it is not so easy to say they have been the most successful. Comparing the average medal counts per games played, the Soviet Union did better by 123 medals. This leads into thinking about what defines success and what makes a country the best at the Olympics. While current situations have left us without the Olympics this year, this study has been able to connect me to sports and I hope it was able to do the same for all readers.