CSC111 Project 2 Proposal:
The Analysis of Summer Olympics Through External Effects (1940 - 2020)

Anh Dang Phuong, Dimitrios Gkiokmema, Dora Gombar, Maya Edri

March 5, 2024

1., Problem Description and Research Question

How did external - geopolitical, and societal - factors influence the outcomes and dynamics of the Olympic Games?

Over the last 100 years, many historical occurrences have created tension while trying to perform in the Olympic Games, leaving an indelible mark on its history, fundamentally reshaping the competition landscape, and altering the trajectory of athletic achievement.

Take, for instance, the impact of World Wars on the Olympics. Since the inception of the first modern Olympic Games in 1896, the sports game has faced cancellation on only three occasions: once during World War I (1916) and twice during World War II (1940, 1944). Yet, even in the post-war era, the scars of conflict lingered, leading to the decision to ban German and Japanese athletes from participating in 1948.
Afterward, during the Cold War, the Olympics became a battleground for ideological supremacy between the United States and the Soviet Union. The intense rivalry between the superpowers spilled onto the athletic stage, with each nation leveraging sporting success to bolster their respective political agendas. Following the Soviet invasion of Afghanistan, tensions between the United States and the Soviet Union escalated, leading to President Jimmy Carter's announcement of a boycott of the 1980 Moscow Summer Games by the United States. In response, the Soviet Union boycotted the 1984 Summer Olympics in Los Angeles.
Finally, the fall of the Soviet Union in 1991 was a watershed moment in modern history, and its impact rippled across various facets of global affairs, including the Olympic competition. The Soviet Union formally dissolved and broke into fifteen separate nations, which altered the Olympic community's balance of power and presented logistical challenges as new nations tried to establish their sporting infrastructure.

The abovementioned examples - which are only fragments of the whole picture - perfectly illustrate the intricate background of the Olympic Games, indicating that sports results come not only from human capital but also from geopolitical and societal factors - often foreseen. **Our project aims to emphasize, represent, and visualize such (international) historical events through the lens of our statistical computation approach: we plan to use graphical tree representations, pie and bar charts, and graphs.**

2., Computational Plan: sample functions

Our project will incorporate datasets on world population, global regionalization, and Olympic Games outcomes (1940 - 2020) to show the interconnectedness of these factors. Our datasets will be downloaded as .csv files, and we plan to extract (read) the data similarly to the .txt files used in Project 1 or the .csv files in Exercise 3. After processing the data, we will create two Node-based Trees, which will be the central part of our main.
One Tree will have a hierarchy of world, continents, regions (northern, eastern, central, etc.), countries, and an '*AnnualData*' class. *(see Figure 1)* The leaves, '*AnnualData*' class instances, will represent the year of the Game and contain information about the type and number of medals won, players who participated in the games, and the Game's host country. We might make object classes for the other hierarchy levels later in the project, but currently, we are thinking of simply writing the hierarchy's name in the Tree's roots. For example, since continents don't have

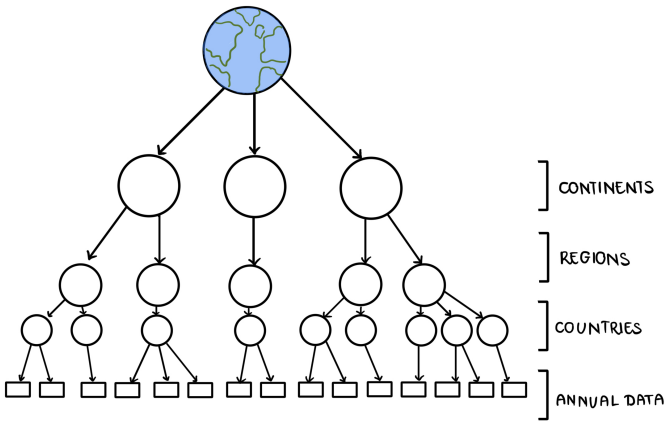

Figure 1: *Our representation of the Node-based Tree structure*

a class, the continent's name will be a tree's root.

The **other Tree** will store the host countries similarly but with a single string as leaves. However, instead of this Tree, we have been considering to use a global variable, a map, to store years mapped to host countries.

The methods and functions we create will use the data stored in the trees to perform calculations, some of our ideas to provide some idea of our approach:

I. 'Easier to complete' computations (probably used during more compound analysis):

- Compute the medals in a given year
- Compute the medals for a given country, continent, and regions
- Compare the number of Gold, Silver, and Bronze
- Ranking (which country, continent, or region ranked the i^{th} place for the number of (gold/silver/bronze/total) medals in the given year?)
- Access to the annual data for a given (country, year). That is, the method prints information stored as an '*AnnualData*' object, but in a better format other than a default one.

II. Parts of our compound analysis:

- Historical events (wars, stock market crash, etc.): measuring their impact by calculating the standard deviation of the medals given at the Olympic games each year and seeing if the number of medals lies far from the mean. An outlier is a data point that is "three standard deviations from the mean may be considered an outlier" (Peter K Dunn, Scientific Research and Methodology)

  If such outliers coincide with historical events, our hypothesis is proven correct. Lastly, our results will be displayed on graphs drawn in PyGame. The graphs will clearly show the interaction of population and historical events in the Olympic games.

- Host Country Effect: analyzing the impact of hosting the Olympics on a country's performance by calculating the difference in medal counts for host countries in the year they hosted compared to non-host years.

  More profoundly, we will extract the years when the given country was a host—if it was—from the Tree or the global variable and then recursively traverse through the 'bigger Tree' to extract the information about non-hosted and hosted events. If the country has never been a host, we will not do the comparison, but we will return that it has not been a host yet.

- Team vs. Individual Sports Impact: analyzing the impact of team sports versus individual sports on a country's overall medal count and identifying countries that excel in one category over the other.

  More profoundly, we manually or, using another dataset, create two groups of sports—individual and team sports—and then develop statistics about which is dominant in each country over time, using recursive methods.

- Performance: through the investigated period, we will analyze the results of a give (input) country, continent, or region. We will measure the rate of change by taking the difference between consecutive years, and then take the average of these differences. This will give us the average of change for a country over the years.

  More profoundly, we plan to weigh each medal—e.g., gold = 3, silver = 2, bronze = 1—and use this measure to examine a country, region, or continent's overtime performance in the participating years.

- Change in Participating Countries: based on user input, we will investigate the rate of change in the amount of medals received by participating countries over time.

- Change in the number of Sports: by measuring medals and keeping track of the number of sports. We are also considering calculating the rate of change for the number of sports in each game, but this is still undecided.

For our tree representation, we are still exploring our options, but we are planning to use *PyGame*, and, as other methods, we have been considering the visualization method in Exercise 3 Part 2, the *dash* library, and another approach (explained below).

Lastly, from the tree, the results for our research questions will be displayed on graphs drawn in *matplotlib*. *Matplotlib* is a powerful and versatile Python library for creating static, animated, and interactive visualizations. It is particularly well-suited for our project due to its user-friendly interface and a rich set of functionalities. The library excels in generating a diverse array of plots, charts, and graphs. The graphs will clearly show the interaction of population and historical events in the Olympic games. Here is the break-down of displaying graphs through *matplotlib* (it takes as little as four lines of code):

1. import matplotlib.pyplot as plt
2. Make two lists of equal length (for the x and y axis)
3. plt.plot(x, y)
4. plt.show()

Graphs in matplotlib are very customizable; you can display multiple graphs at a time, multiple lines (or substitute the lines for points), add a title, x/y labels, and much more! We originally thought of using PyGame to display our graphs, but we would have needed to draw them manually; matplotlib does all the hard work for us by drawing the graphs by itself.

Our code will utilize numerous datasets to visualize how connected they are (population, Olympic games, and maybe some other data):

1. Information on the Olympic Games from 1896-2020:
IOC Research and Reference Service. 2017. "Olympic Sports and Medals, 1896-2014." Kaggle. 2017. https://www.kaggle.com/datasets/the-guardian/olympic-games?select=summer.csv.
(an example of its data is: 1896, Greece, Athens, Great Britain, GBR, 2, 3, 2)

2. Global regionalization:
International Organization on Standardization. 1999. "UNSD — Methodology." United Nations: Statistics Division. 1999. https://unstats.un.org/unsd/methodology/m49/.

3. Population data:
Ritchie, Hannah, Lucas Rodés-Guirao, Edouard Mathieu, Marcel Gerber, Esteban Ortiz-Ospina, Joe Hasell, and Max Roser. 2023. "Data Page: Population." Our World in Data. 2023. https://ourworldindata.org/grapher/population.

3., References

1. Carter, Jimmy. 1995. *Keeping Faith: Memoirs of a President.* University of Arkansas pbk. ed. Fayetteville, AR: University of Arkansas Press.

2. Dunn, Peter K. 2021. *Scientific Research and Methodology: An introduction to quantitative research in science and health.* https://bookdown.org/pkaldunn/Book/TypesOfResearch.html.

3. Grannan, Cydney. 2016. "7 Significant Political Events at the Olympic Games — Britannica." Encyclopedia Britannica. July 29, 2016. https://www.britannica.com/list/7-significant-political-events-at-the-olympic-games.

4. International Organization on Standardization. 1999. "UNSD — Methodology." United Nations: Statistics Division. 1999. https://unstats.un.org/unsd/methodology/m49/.

5. IOC Research and Reference Service. 2017. "Olympic Sports and Medals, 1896-2014." Kaggle. 2017. https://www.kaggle.com/datasets/the-guardian/olympic-games?select=summer.csv.

6. Johnston, Mindy. 2024. "Moscow 1980 Olympic Games — Boycott, Cold War, USSR, & Summer Games — Britannica." Encyclopedia Britannica. February 28, 2024. https://www.britannica.com/event/Moscow-1980-Olympic-Games.

7. Miller, David. 2012. *The Official History of the Olympic Games and the IOC: Athens to London*, 1894-2012. Edinburgh: Mainstream Pub.

8. Ritchie, Hannah, Lucas Rodés-Guirao, Edouard Mathieu, Marcel Gerber, Esteban Ortiz-Ospina, Joe Hasell, and Max Roser. 2023. "Data Page: Population." Our World in Data. 2023. https://ourworldindata.org/grapher/population.