

CS 445 / 545 Final Project Report

Olympics

Megan Stanton
Computer Engineering
Email: mstanto4@vols.utk.edu

Jaidin Jackson
Computer Engineering
Email: jjack113@vols.utk.edu

Joshua Gray
Computer Engineering
Email: jgray51@vols.utk.edu

1. Objective

Every two years (with the exception of 2020), athletes around the world come together to represent their country and compete in the Olympic Games. Those who have watched the Olympics throughout the years have more than likely noticed that some countries have much better odds at earning a medal in certain events than others. Some of the most popular examples include Jamaica's track and field team, America's basketball team, and Canada's hockey team. And there are many others that have emerged over the past 20 years.

The objective of our project is to find out whether there are any external factors at play that might influence the outcomes for these events. There are two main questions that we aim to answer: do warmer countries have a tendency to perform better in the Summer Olympics as opposed to colder countries and does the event being conducted inside or outside affect the country's performance? In other words, we are looking for any correlation between a country's climate and its performance in the Summer Olympics, as well as any correlation between a country's performance and the type of event.

2. Data That Was Used

For this project, we collected data from several different data sets. The first set of data we analyzed contained the medals that were given out to each country for each Olympic event from the 1976 Olympics to the 2008 Olympics. This data was gathered from [kaggle.com](https://www.kaggle.com). In order to condense the size of the data that we were interested in, we decided to record only the medals for six events. The events that we chose were archery, diving 3M springboard, diving 10M springboard, gymnastics floor exercises, field hockey, and volleyball.

The second data set we utilized contained that average annual temperatures for various countries around the world. This data was gathered from en.climate-data.org. In order to make the information easier to analyze, we only recorded the average annual temperatures for countries which earned at least one medal any of the four events. We ended up recording the temperatures for 32 different countries.

3. Models / Algorithms

The original data set was read in using a python script to separate the sports that we decided to analyze into their own text files. The sports that were chosen were archery, diving three meter springboard, diving ten meter springboard, gymnastics floor exercises, hockey, and volleyball. The python script placed the data for each sport into a text file corresponding to the sports name. For example, the archery data was placed in the "archery.txt" file. After the data was sorted, the data was read into another python script to be analyzed. This file was named "AnalyzeData.ipynb" and contains all the charts seen in the "Results" section. For each sport, the countries were placed into three dictionaries, where each one contained either the gold, silver, or bronze medal count. Next, a graph was created for each sport that displayed the gold, silver, and bronze medal count on the same graph utilizing a stacked bar chart. In order to determine the overall best country at each sport, we calculated a total score for each country. This total score was calculated by counting gold medals as three points, silver medals as two points, and bronze medals as one point. The top three countries for each sport are located in Table 1 in the "Results" section.

Our next objective was to determine if warmer countries have a tendency to perform better in the summer Olympics as opposed to colder countries. In order to determine if this was true, the countries average temperature was graphed on the y axis. The x-axis represents the countries from left to right according to the amount of medals they won, with the left-hand side having the larger medal count. The bars were also color coordinated based on the temperature in order to make it simpler to determine correlations. The color codes chosen were "tomato", "gold", "yellow", "palegreen", and "mediumturquoise." The "Tomato" color had temperatures greater than twenty degrees Celsius. The "gold" color had temperatures ranging from fifteen degrees Celsius to twenty degrees Celsius. The "yellow" color had temperatures ranging from ten degrees Celsius to fifteen degrees Celsius. The "palegreen" color had temperatures ranging from five degrees Celsius to ten degrees Celsius. The "mediumturquoise" color had temperatures ranging from zero degrees Celsius to five degrees Celsius.

4. Results

Figures 1-6 represent the data that was collected for each sport. It displays the gold, bronze, and silver medal data in a stacked bar chart format. For archery, South Korea had the most gold, bronze, and silver medals. They had 6 gold, 5 silver, and 8 bronze medals. For diving 3M springboard, China had the most gold and bronze medals, with 10 and 7 medals respectively. While the United States had the most silver medals at 5 medals. For diving 10M springboard, China had the most gold medals at 8 medals, while the United States had the most silver and bronze medals, at 5 and 5 medals respectively. For gymnastics, Romania had the most gold medals at 6 medals, the United States had the most silver medals at 5 medals, and the Soviet Union had the most bronze medals at 5 medals. For field hockey, the Netherlands had the most gold and silver medals, at 64 and 64 medals respectively. While Spain had the most bronze medals at 48 medals. For volleyball, the United States had the most gold medals at 46 medals, Brazil had the most silver medals at 28 medals, and the Soviet Union had the most bronze medals at 36 medals. In archery, South Korea had the highest amount of total medals. In diving, China had the highest amount of total medals. In gymnastics, Romania had the highest amount of total medals followed by the Soviet Union. In field hockey, the Netherlands had the highest amount of total medals followed by Australia. In volleyball, Brazil had slightly more medals than the United States, but the United States more gold medals.

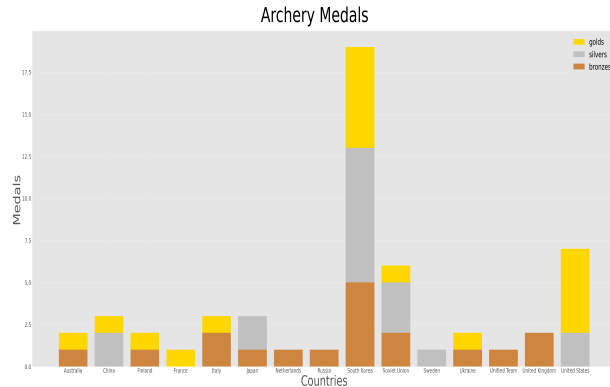


Figure 1: Medal Data for Archery

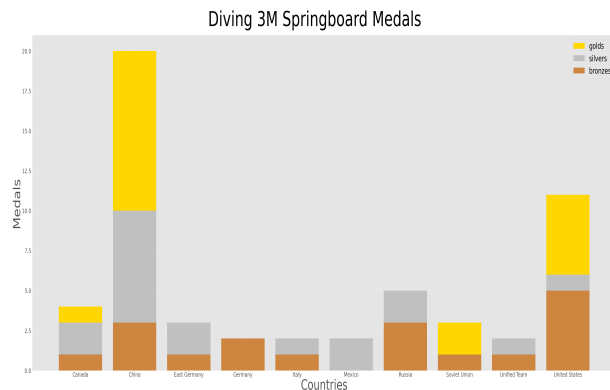


Figure 2: Medal Data for Diving 3M Springboard

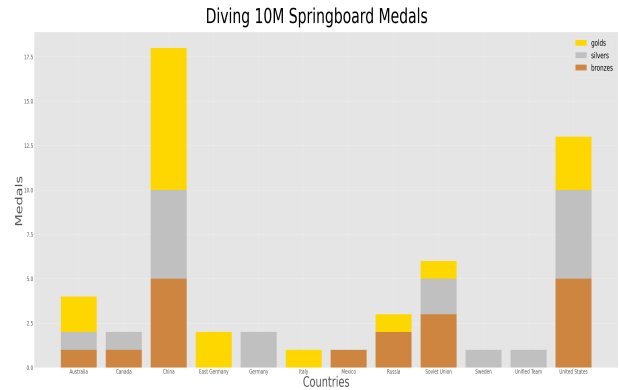


Figure 3: Medal Data for Diving 10M Springboard

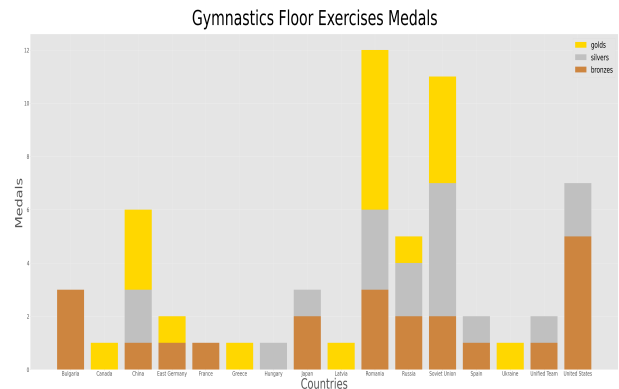


Figure 4: Medal Data for Gymnastics

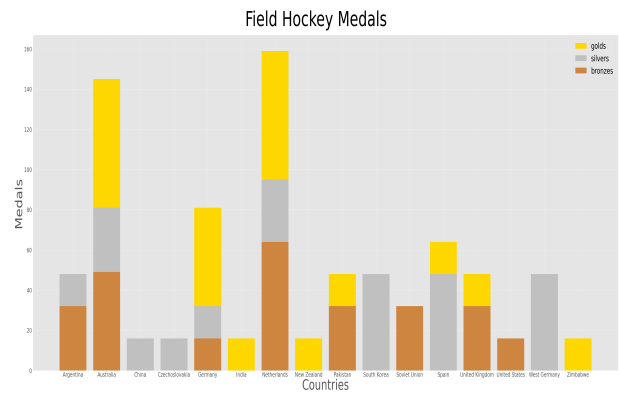


Figure 5: Medal Data for Field Hockey

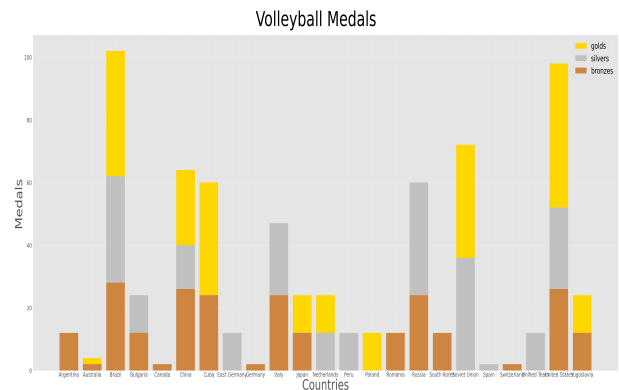


Figure 6: Medal Data for Volleyball

Table 1 demonstrates the top three medal earning countries for each corresponding sport, and what score that country was given. The score was calculated by counting gold medals as three points, silver medals as two points, and bronze medals as one point. The United States placed in the top three in four sports. The Soviet Union also placed in the top three in four sports. China came in first place in both diving events, and came in third in gymnastics. The rest of the countries in the top three, Romania, Australia, Germany, Netherlands, and Brazil, only placed once.

Table 1: Total Scores for All Sports
SU = Soviet Union USA = United States

| Sport | First | Second | Third |
|-------------------|--------------------|------------------|----------------|
| Archery Scores | South Korea 39 | USA 19 | SU 11 |
| Diving 3M Scores | China 47 | USA 22 | Canada 8 |
| Diving 10M Scores | China 39 | USA 24 | SU 10 |
| Gymnastics Scores | Romania 27 | SU 24 | China 14 |
| Hockey Scores | Netherlands 318 | Australia 305 | Germany 195 |
| Volleyball Scores | USA 216 | Brazil 216 | SU 180 |

Our first objective in looking at the climate of the winning countries was to determine if warmer countries have a tendency to perform better in the summer Olympics as opposed to colder countries. We also wanted to examine whether the event being conducted inside or outside affected the country's performance. The following figures depict the average temperature for the given country. The x-axis represents the countries from left to right according to the amount of medals they won, with the left-hand side having the larger medal count.

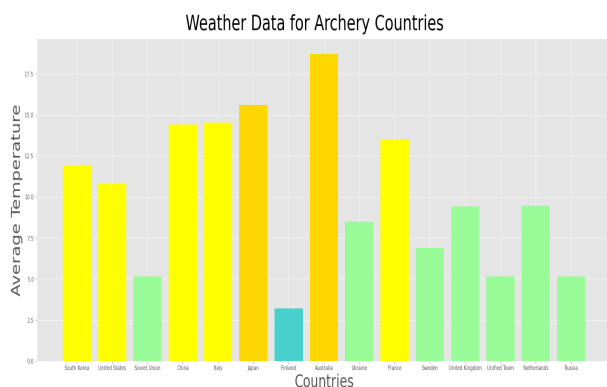


Figure 7: Weather Data for Archery

The moderate countries (yellow) performed the best in archery, with France being an outlier. The warm countries (gold) performed average. However, Finland is an extremely cold country and is also found in this range. The cold countries (pale green) performed the worst in archery, with the Soviet Union being an outlier. The colder countries are

found more so on the right side of the graph, whereas the moderate and warm countries are found more so on the left side or the middle. This even being conducted outside could be a reason for colder countries performing slightly better. However, there are still outliers that dispute this claim. Therefore, there is no obvious correlation between the country's temperature and its performance in archery.

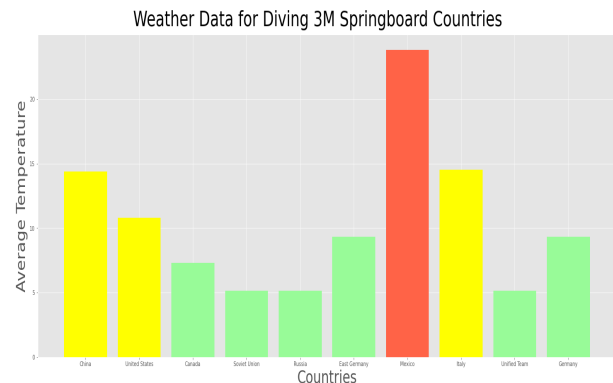


Figure 8: Weather Data for Diving 3M Springboard

The moderate countries (yellow) performed the best in 3M diving, with Italy being an outlier. The cold countries (pale green) performed about average, with the Unified Team and Germany both being outliers. Mexico, an extremely warm country (tomato), also performed average. Therefore, there is no obvious correlation between the country's temperature and its performance in the 3M diving springboard event.

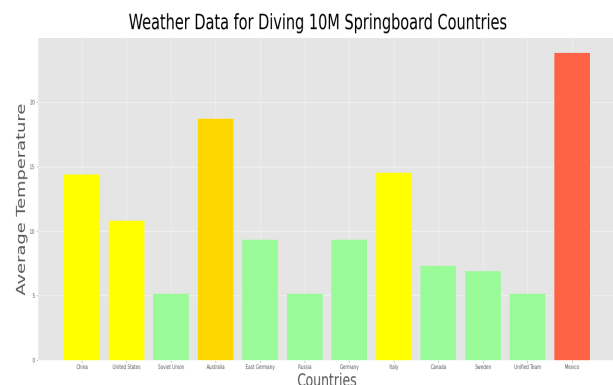


Figure 9: Weather Data for Diving 10M Springboard

The moderate countries (yellow) performed the best in the 10M diving event, with Italy being an outlier again. Australia, a warm country (gold), performed slightly above average. The cold countries (pale green) performed both average and below average, with the Soviet Union being an outlier. Mexico, an extremely warm country (tomato), also performed very below average. Therefore, there is no obvious correlation between the country's temperature and its performance in the 10M diving springboard event.

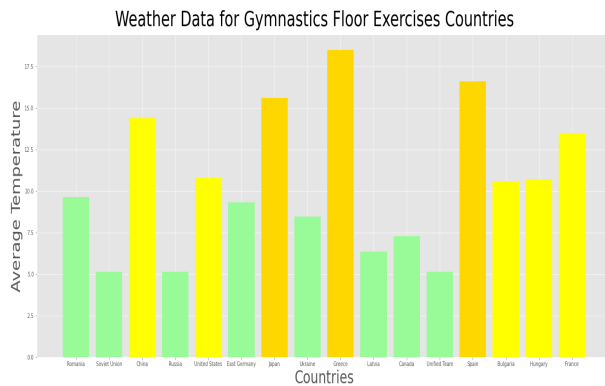


Figure 10: Weather Data for Gymnastics

The cold countries (pale green) performed the best in the gymnastics floor exercise event. However, there are cold countries that also performed average and below average. The warm countries (gold) performed either average or below average. The moderate countries (yellow) performed both above average and extremely below average. The colder countries are found more so on the left side of the graph, whereas the moderate and warm countries are found more so on the right side. This even being conducted inside could be a reason for colder countries performing slightly better. However, there are still outliers that dispute this claim. Therefore, there is no obvious correlation between the country's temperature and its performance in the gymnastics floor exercise event. There is a slight non-obvious correlation in that there are no hot countries in this event that obtained medals. All of the medal winning countries were below 15 degrees Celsius.

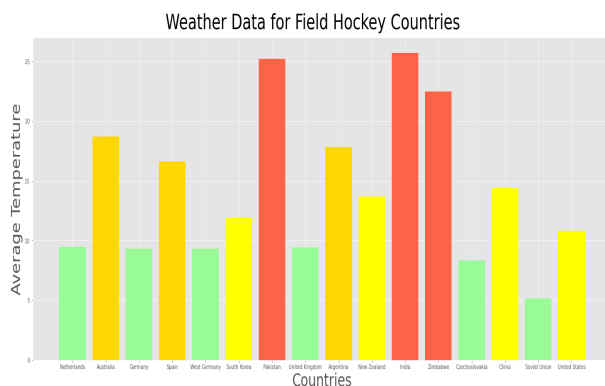


Figure 11: Weather Data for Field Hockey

The only two countries of the same temperature range that received a similar number of medals are India and Zimbabwe, both extremely warm countries. Countries that fall within the same temperature ranges (warm, moderate, cold, and extremely cold) are evenly spread out across the graph. Therefore, there is absolutely no correlation between a country's temperature and its performance in the field hockey event.

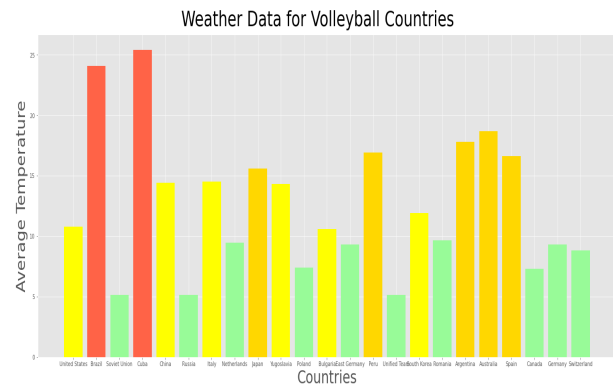


Figure 12: Weather Data for Volleyball

Countries that fall within the same temperature ranges (extremely warm, warm, moderate, cold, and extremely cold) are all spread out across the graph. There are a few countries of the same temperature range that are lumped together, but not enough that it would indicate any sort of correlation. Therefore, there is absolutely no correlation between a country's temperature and its performance in the volleyball event.

Each graph listed the countries from left to right depending on the number of medals the country received with the left being the greatest number of medals and the right being the least number of medals. We wanted to know if warmer countries performed better in the summer Olympics overall. If that were the case, the warmer countries would be primarily on the left side of each graph and the colder countries would be on the right side of each graph. After examining all the graphs, warmer countries did not perform better in the summer Olympics overall. Countries of all temperature ranges were found anywhere along the x-axis of the graphs. The only two events that indicated a slight correlation between a country's average temperature and performance in an event was archery and the gymnastics floor exercise. Colder countries seemed to perform better in gymnastics (an inside event), whereas warmer countries seemed to perform better in archery (an outside event). Nevertheless, there was no obvious correlation and were several outliers within the graphs for each event. The other events we observed did not indicate a correlation between a country's temperature and whether or not the event took place outside or inside.

5. Primary Issues Encountered

There were several issues that we ran into while working on our project. The names of the athletes in our data sets gave us a couple of issues. The first issue we ran into involved how some of the athletes' names were spelled in our data set. There were a few athletes that had special characters in their name which made it hard for us to efficiently read the data using a computer program. We did learn how to search for a specific byte in a file which was very helpful in determining where the special character was located so we knew which special characters to remove.

Another issue involving the names of the countries was that some of the countries changed their names during the time frame we collected our data. For example, there were instances of East and West Germany winning medals even though Germany was reunified in 1990. Another example was the Soviet Union. Because Russia was named the Soviet Union from 1922 to 1991, we had several instances of medals won by both the Soviet Union and Russia. These issues caused problems with how we decided to count the medals for each country.

There was also an issue with Yugoslavia. Yugoslavia was a country from 1945 to 1992. We have instances of this country in our data set, however, since it is no longer a country, we could not find any information about its climate. Instead, we decided to use Croatia's climate information.

6. Future Work

If we were to make improvements to our current project, we would first take into account the fact that some countries' names changed over the years. Specifically, the countries Germany, Russia, and Yugoslavia—the details are outlined in the 'Primary Issues Encountered' section.

Our main goal was to identify a specific reason as to why certain countries perform better than others in a competitive event. Prior to this project, we suspected that countries that reside in warmer regions would perform better in the Summer Olympics. However, we did not consider that many of these events are conducted inside (sometimes outside), making the weather an unlikely influence on a country's performance. If we had analyzed the Winter Olympics, then we may have seen stronger trend because more events in the Winter Olympics are conducted outside as opposed to inside. Events in the Winter Olympics also require particular outdoor environments, such as snowy mountains for skiing and snowboarding. Given this fact, countries that experience these conditions year-round likely have a high interest rate in such sports, increasing the chances of producing an athlete who excels in them, whereas countries who do not experience these conditions year-round, if at all, likely do not have a high interest rate, decreasing the chances of producing an athlete who excels in them.

After completing this project, we also realized that we could have analyzed a different type of contest. The Olympics is a physical performance contest, but we could have chosen an academic contest, such as the World Chess Championship, or we could have chosen something as simple as analyzing countries' performances in academic subjects, such math and science. Given that countries have varying intensities and methods in how they educate their people, we believe that we would be able to identify a much stronger trend in the data. Ultimately, this project supplied us the basic knowledge to collect and analyze data and provided us a solid foundation to build off of as we continue to expand on our project and investigate our suspicions even further.

7. Organization Chart

Responsibilities of Team Members

Megan Stanton (Coordinator)

- Collect Medal Data
- Create Graphs for Medal Data

Jaidin Jackson (Recorder)

- Format Graphs
- Analyze Data/Create Presentation
- Record Progress

Joshua Gray (Checker)

- Collect Weather Data
- Create Graphs for Weather Data

Everyone

- Parse Data Sets

Timeline

October 15st:

- Find data sets for the summer Olympics.

November 15th:

- Create a program to read in the large data set and portion it out into smaller files, collecting only information for the sports that we have decided to examine.

November 20th:

- Compile the data into a program so we can create graphs and analyze them.
- Determine which countries have the highest average medal count for each sport.

November 21st:

- Gather data for the average climate of the countries that have the highest average medal count.

November 22nd:

- Create graphs for the average climate compared to the medal count of the countries.
- Prepare presentation.

November 26th:

- Complete objective, data, and algorithms sections of the final project report.

November 30th:

- Complete results, primary issues, and future work sections of the final project report.

December 1st:

- Revise draft of final project report.

December 6th:

- Submit final project report.