The Olympiad: Visualising Dominance and Bias in the Summer Olympics

Chew Yuxi, Tian Seet Yuen, Clifton Ngeow

Abstract—The Olympic Games prides itself on the values of competition and fairnes. However, several sporting events have been consistently dominated by a particular continent or even, country. Meanwhile, host countries are gifted the privilege of automatic qualification. These issues directly contradict the values of the Olympic Games, yet many are still unaware on the severity of these issues. To address this, our team have decided to adopt the style of journalism, with the ultimate goal to provide an engaging and educative reading experience for all. The Olympiad utilizes several visual elements, such as a sunburst diagram, time-series line graph and a scatter plot to assist readers in their understanding on the status of the Summer Olympics. Lastly, insights attained from these visualisations will also be shared in detail.

Index Terms—Data Visualisation, Summer Olympics, Olympic dynasties, home ground advantage, sunburst, scatter plot.

✦

1 INTRODUCTION

The Olympic Games can be traced back to 776BC and were dedicated to the gods of Olympia. Today, the Olympic Games has become the world's largest sports competition, with more than 200 nations participating every four years. [1]

The Olympic Games prides itself on the culture of excellence, competition and fairness. [2] However, during its history, the formation of Olympic Dynasties have become a common occurrence – whereby certain sport events are consistently dominated by a particular country or continent. Nevertheless, many are still unaware of the severity of this issue. The current state of the Summer Olympics largely represents a monopoly, with Olympic Dynasties in total control. All of which leads to a fundamental question – is it still a competition if the Summer Olympics are mostly a one-team race?

Meanwhile, every four years, different countries are given the opportunity to host this quadrennial event. In the sporting world, the phenomenon of Home Ground Advantage exists due to various reasons, eg. influence of home crowd or familiarity of playing environment. However, in the Summer Olympics, it is a tradition for host countries to be granted automatic qualification. [3] Once again, this contradicts the fundamental values of the Olympic Games – is it fair for the host country to be given this privilege?

2 MOTIVATION AND OBJECTIVES

On the surface, the Olympic Games exudes the spirit of competition and equality. However, the issue of Olympic Dynasties and Home Ground Advantage is very much prevalent in this prestigious event. Yet, many are still unaware about the severity of these issues. As such, the goal of this project is to provide drilled-down visual analyses that educate the public about the following:

1. Severity of Dominance by Continents in various sports
2. Severity of Dominance by Countries in various sports
3. Severity of Bias for Host Countries in medals won

Overall, this project seeks to assist the public to understand the Summer Olympics as it truly is – inclusive of all possible dominance and bias. In addition, this project seeks to adopt the style of journalism, and undertakes the challenge to present a wide variety of sports in a digestible and interactive manner.

3 DATA TRANSFORMATION

Two datasets were attained from Kaggle. [4] Data provided within these datasets included every athlete and results from the history of the Olympic Games. However, as our intended visualisations were very specific, it was essential to perform some form of data transformation and cleaning, to ultimately fit these data into our visualisations.

The data was transformed in the following ways:

1. Merge dataset by joining the National Olympic Committee (NOC) country code with relevant countries and continents
2. Compute number of participants based on each sport
3. Compute total medal count based on all medal types
4. If the sport has more than one participant, it is a team sport. Remove all but one observation of the team sport
5. Rename the name of the remaining observation to the country name. The winner of the team sport will be represented by the country rather than the name of the individual

6. Filter countries based on a list of host countries for the Summer Olympics

In the original dataset, each observation included the athlete's name, country, sport and medal won. As the data provided was too drilled-down, some form of aggregation was necessary to transform the data into the format which we require for our visualisations.

Next, with regard to continents, there was no information given. As such, another dataset was attained to match the country's NOC to the respective continent.

Lastly, to explore home ground advantage, we required only a list of host countries, along with their results and the corresponding years in which they hosted the Olympic Games. As such, we referenced a list of host countries [5] to extract only data which we required.

Additional data cleaning included:

1. Filter seasons based on Summer Edition
2. Remove irrelevant columns such as ID, Height and Weight
3. Renaming the data headers to fit our visualisations

4 VISUALISATION DESIGN

4.1 Data Journalism

To appeal to the average reader, the team adopted the role of a journalist, and presented the visualisation in the form of a news article. A sample of the data journalism article is shown below. Users can read the insights of author and view the appropriate visualisation simultaneously.
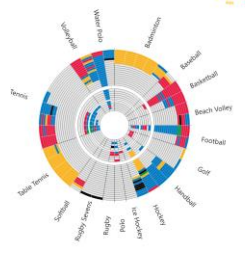

Figure 1: Sample of article

To meet the project goal, our team decided on 3 main elements – sunburst diagram, time-series line graph and scatter plot. Each visualisation is intended to provide the user a better understanding of the Olympic Games.

4.2 Sunburst Diagram

The goal of the sunburst diagram is to visualise dominance by continents and countries. We decided to use a sunburst to tackle the issue of showing a large number of sports at once. This was crucial, because other alternatives would require the user to toggle between sports often. The action of toggling is extremely disruptive to the reading process.

A traditional sunburst diagram is a multilevel pie chart used to display the proportion of various categories at every hierarchical level. In this case, each level of the sunburst diagram represents an Olympic Event. Thus, each individual arc is separated from the next arc by a time period of 4 years, due to the quadrennial nature of the Summer Olympic Games.
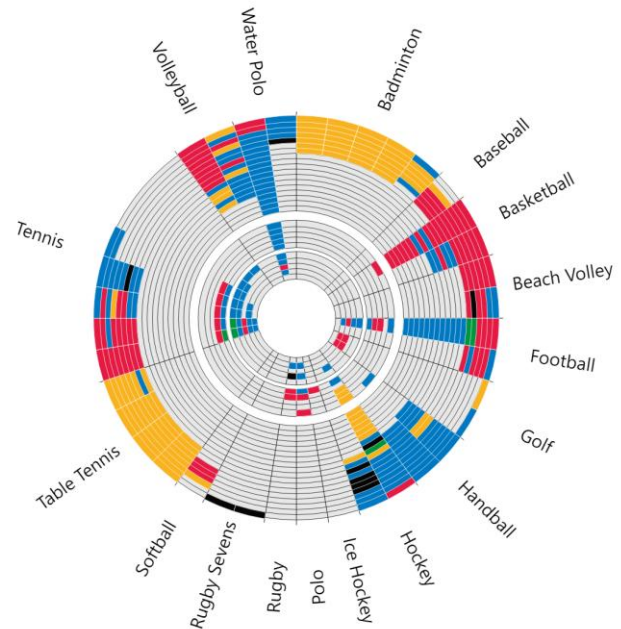

Figure 2: Sunburst Diagram of Ball Sports

The sunburst diagram is then segmented by various sports. Due to the large number of sporting events, five sunburst charts had to be generated, with each diagram adhering to a specific theme. The theme of the figure above is **ball sports**.

Figure 3: Continent Legend of Sunburst Diagram
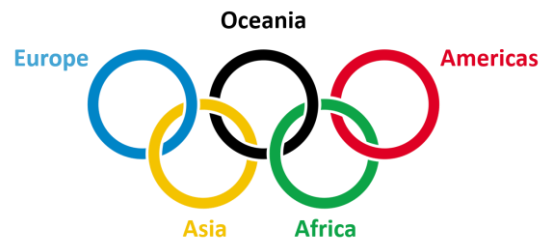
Each sub-unit in the sport category represents a gold medal that is won by a participant. The sub-units are coloured according to the continent that the participant belongs to. The legend above details the colours that were used to represent each continent. Thus, by observing the number of sub-units of a particular colour across multiple iterations of the Olympic Games, the

reader is able to assess whether the continent has been relatively dominant in the sport.

Figure 4: Tooltip shown upon hovering across a sub-unit

Tooltips were also included for each sub-unit, displaying relevant information such as the name of the participant, the country that won the gold medal and the sporting event, the sport discipline.

To allow the user to focus on particular countries as he or she reads the web article and visualisation, we incorporated the use of interactive text. An excerpt of text from the visual guide included in the article is shown below.

Hover over the text to highlight the relevant units in the sunburst diagrams. In this case, medals won by the United States are highlighted.

Figure 5: Excerpt from the article with interactive text in **Pink**

Upon hovering across the interactive text, the relevant sub-units in the appropriate sunburst chart will be highlighted as shown below. This allows the reader to easily follow the flow of the article, while interacting with the chart. By including such interactivity, we are able to communicate insights about dominance at the country level.
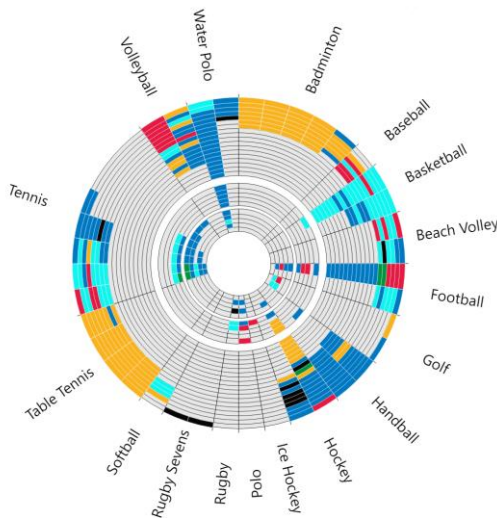

Figure 6: Sunburst diagram with the United States sub-units highlighted via interactive text

4.3 Time-Series Line Graph

A time series is a series of data points indexed in a timely order. A time-series line graph displays these data points via connecting them together through a line. As the history of Summer Olympics include results across every 4 years, a time-series line graph would be ideal to provide an overview of how results have changed over time.
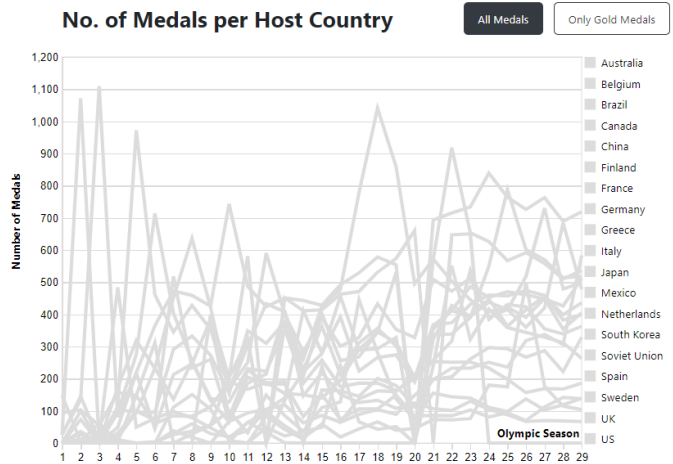

Figure 7: Line Chart displaying all Number of Medals for all host countries, by Olympic Season

In this case, we will utilize this element to visualise the magnitude of Home Ground Advantage. Since countries host the Summer Olympics on different occasions, any pattern or trend with respect to the year in which the country hosted in would be clearly displayed.

As the Olympic Games were cancelled during World War I and II, the Olympic Season is displayed as the x-axis instead to ensure the generation of continuous lines.
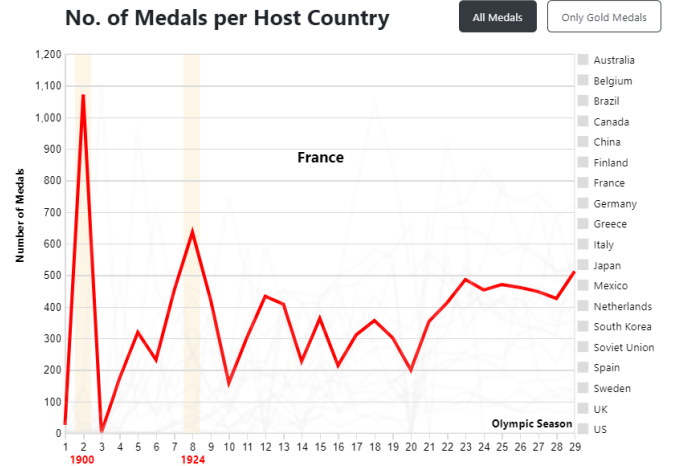
Figure 8: Highlighted Line Chart upon hovering "France"

Next, upon hovering over the country, the line corresponding to the hover will be highlighted, along with the cities and years in which the country hosted the Olympic Games in. This allows the user to immediately identify any trends with respect to host country bias.

Lastly, users are also able to select between all medals or only gold medals. As the degree of difficulty to attain

each medal type differs, i.e. Bronze, Silver and Gold, the number of medals won could differ significantly as well. As such, this option allows for further analysis and comparison between each medal tier.

4.4   Scatter Plot

A scatter plot is a two-dimensional data visualisation that uses dots to represent the values obtained for two different variables - one plotted along the x-axis and the other plotted along the y-axis.

In our case, our scatter plot shows the number of medals won and the number of participants of all Olympic history.
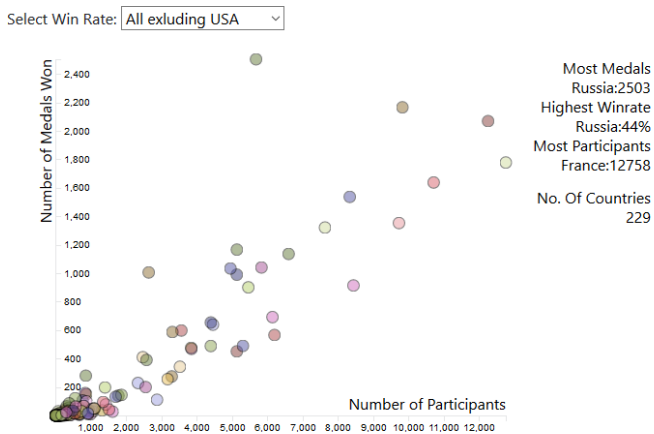
Figure 9: Scatter Plot to show relationship between Number of Participants and Medals Won

Via this scatter plot, we can observe the relationship between number of participants sent by a country and the number of medals won.

Secondly, the scatter plot allows viewer to identify the truly dominating countries, i.e. dots will appear as an outlier above the trend line. For instance, via Figure 10, it can be observed that Russia is a clear outlier and correspondingly, it has the highest win rate amongst all countries.

Included in the win rate drop down selection are bins of win rates for further exploration. Additionally, a graph without the USA is included to show a better distribution of win rates in the array of countries.

Lastly, hovering over the legend will also reveal the respective country on the plot in that category.
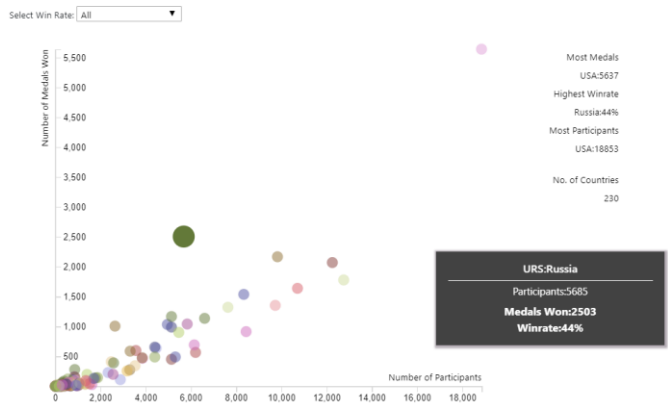
Figure 10: Scatter Plot with Russia hovered upon

5   WEB BASED APPLICATION

5.1   System Architecture

Our project uses a 2-tier client/server architecture. Our client-side was designed and developed using JavaScript. JavaScript is one of the fastest-growing language and is the most commonly used language due to a shift toward the use of mobile devices. Moreover, the flexibility and scale of JavaScript libraries allow users to access web applications on devices that has a browser. [6] We used Amazon Web Services as our server to host our application as it is quick to deploy and reliable with free resources for deployment.

5.2   Data Visualisation

Our application was developed with a strong emphasis on readability and storytelling. Using an article style design, our application wishes to show insights of the Olympics through the aid of interactive visuals. Despite looking more like an article, our visualisations still follow best practices for design, letting users explore and find their own insights with the interactive graphs.

The data visualisation elements in our application is created with Data-Driven Documents (D3.js) which is a JavaScript library producing dynamic, interactive data visualisations in web browsers which users can manipulate [7]. D3.js makes use of the widely implemented Scalable Vector Array (SVG), HTML5, and CSS standards to produce elements which can be added, removed or edited according to the contents of the dataset [8]. Additionally, D3.js accepts a wide variety of data types such as CSV, JSON and HTML document fragments. [9]

5.3   Web Design

Our application was developed based on the cornerstone technologies of the World Wide Web. Namely, CSS, HTML and JavaScript. Using a combination of Bootstrap,

a framework to help design websites faster and easier, and jQuery, our team were able to develop the web design rapidly. As most of the frameworks we used are to develop web applications, our application is supported by most modern browsers.

In addition, we used Bootstrap's library to implement a mobile-responsive web page where it will scale properly even on a mobile device and jQuery to create a dynamic page.

6 INSIGHTS

Upon completion of the visualisations, the insights observed are as follows:

6.1 Dominance in Movement Sports

Athletics is a category of sports that consists of competitive running, jumping, throwing and walking. The Americas and Europe have been dominating a large proportion of the events up till the 1980s. From the 1984 Olympics onwards, Africa has been gaining steady traction, especially in Long Distance Running events.

Furthermore, the most dominant country is the United States. Historically, no other country has excelled in athletics as much the United States have, winning gold medals for athletics in every single Summer Olympic Games, except for the Moscow Games in 1980, where the US boycotted the event.

Competitive cycling is heavily dominated by Europe, which has been performing consistently well throughout the years. However, in the Beijing 2008, London 2012 and Rio De Janeiro 2016 games, the United Kingdom has played a strong hand, clinching many victories in the sport.

6.2 Dominance in Ball Sports

Asia, or more accurately, China, has dominated **Table Tennis and Badminton** in most Olympic games. This speaks tons about the country's cultural influence on the popularity of the sport, which is then reflected in its dominance. Table Tennis and Badminton are also sports that do not favor physical size advantages, which explains why the Asian powerhouse is able to dominate these sports and not others such as Basketball.

The continent of Europe has been consistently tyrannizing **Handball**. The Soviet Union (subsequently Russia) after its dissolution, seems to exert slight dominance in the sport. However, other countries such as Denmark, France and Norway have provided strong competition in the recent games.

Next, the United States are regularly dominating men's and women's **Basketball**, probably due to the influence of the NBA. In fact, the US never had to send all-star NBA teams to the Summer Olympics to clinch the gold, until the Beijing 2008 Olympic Games. This is because Argentina broke the American dominance in 2004. Basketball is one of the most severely dominated sports along with Table Tennis.

Lastly, Europe has held the golden mantle for **Water Polo** in many editions of the Summer Olympics. Hungary has been an especially powerful force in the sport, and historically, has had a slight dominance over the sport.

6.3 Dominance in Water Sports

The gold medal trend for **Diving** has taken a rather interesting turn. For the early part of Olympic history, the United States monopolized the gold medals. During the 1960s to 1980s, European countries started to nab the gold medals as well. In more recent times, the Sport is completely dominated by China.

After the introduction of **Synchronised Swimming** to the Summer Olympics in the Los Angeles 1984 Games, the Americas, or rather, the United States and Canada, have scooped up all the gold medals. But since the Sydney Games in 2000, Russia has ceremoniously seized the top spots to this day.

Lastly, **Canoeing, Rowing and Sailing** are heavily dominated by Europe. Within the continent, there are several powerhouses, such as Germany, Hungary and the United Kingdom. The level of dominance is not as severe as other sports.

6.4 Dominance in Combat Sports

Fencing is one of the oldest sports in existence, its origins can be traced to medieval duelling. Europe has eclipsed the sport, winning a large proportion of the gold medals for fencing events. Notable countries that excel in the sport are Italy and the Soviet Union and subsequently, Russia.

Next, Asia is by far the most dominant continent for **Taekwondo**. This is not surprising, as the sport has South Korean origins.

Lastly, although there does not seem to be a dominant continent, Japan has been extremely successful in **Judo**. Similar to Taekwondo, Judo roots are firmly entrenched in Japanese culture. Other countries have far less consistent performances.

6.5   Dominance in Other Sports

The dominance over **Weightlifting** has switched hands multiple times. Europe controlled the top spots during the earlier editions of the Olympics, as well as 1960s to 1990s. The Americas, or more specifically, the United States, laid claim to the gold medals for the 1948, 1952 and 1956 games. From the Sydney 2000 Games onwards, Asia has reigned over the sport, gradually kicking out all other continents. The most dominant country in recent times is China.

Archery is consistently dominated by South Korea for both men and women events. This is one of the most severely dominated sports.

The only powerhouse for **Rhythmic Gymnastics** has been the Soviet Union (subsequently Russia). Russia has successfully defended their dominant streak for 5 Olympic Games, since the Sydney 2000 Games, for both individual and team events.

6.6   Dominance by Medal Win Rate and Country

Firstly, the most dominant country based on medal win rate is the former **Soviet Union** (44%). This number is insane, as it suggests that almost 1 in 2 participants actually do achieve at least a Bronze medal in their respective sport event. However, this could be due to their hosting of the 1980 Summer Olympics in Moscow, which 67 countries did not participate in. Meanwhile, the **United States** boasts the highest number of participants in the history of Olympics. Yet, its win rate is only at 30%, significantly lower than that of Soviet Union.

Notably, countries with the highest win rates are mainly from the western region of the European Continent. Only 11 countries sit above the 20%-win rate, while 35 countries have a win rate of between 10% to 20%. Interestingly, the majority of these countries have hosted the Summer Olympics at least once in its history.

Rather unfortunately, 81 countries have never won a single medal in the Olympics.

6.7   Severity of Host Country Bias

Pertaining to the total number of medals won regardless of tier, countries consistently peak while hosting the Olympic Games. This could be due to the fact that Olympic hosts are guaranteed a spot in each team event. As such, the number of medals won is likely to increase because there are more athletes competing for the country.

However, as we focus solely on the **Gold Medal Tier**, the trend is not as consistent. We see that for countries such as **Mexico** and **Canada**, gold medals do not appear to peak during the seasons which they hosted in. It seems that even with the home ground advantage, it is difficult for athletes from host countries to dethrone the historically dominant countries. In a nutshell, champions end up grabbing the gold regardless of locality.

7   CONCLUSION

Ideally, readers of The Olympiad would walk away with a clearer picture on the severity of dominance and host country bias in the Summer Olympics.

Now that we have uncovered the severity of dominance and host country bias, The Olympiad could also explore the possible root causes of these issues. Some sportsmen have extreme physical advantages in terms of height, weight or other biological aspects. As such, some continents or countries will naturally perform better in certain sports due to the genetic advantages of their citizens. Understanding these differences can provide insights on why dominance occurs.

Meanwhile, The Olympiad could also expand its coverage into the financial downside of hosting the Olympics. While it is true that host countries do receive tons of benefits, eg. guaranteed qualification, increased tourism or media attention, however, it has been found that especially in recent times, the financial costs of hosting the Olympic Games far outweigh the gains.

Lastly, in discovering these controversial issues, we believe that The Olympiad has only just scratched the surface of combining journalism and visual analytics. With reference to the work we have done in exploring the Summer Olympics, The Olympiad serves as a stepping stone to the discovery of deeper insights, as well as coverage over other major sectors – eg. global crisis, politics or even fraud.

REFERENCES

[1] *The Olympics in Ancient Greece,* available at https://www.history.com/topics/sports/olympic-games

[2] *Promote Olympism in Society,* available at https://www.olympic.org/the-ioc/promote-olympism

[3] *Summer Olympic Games - Qualification,* available at http://olympics.wikia.com/wiki/Summer_Olympic_Games

[4] *120 years of Olympic history: athletes and results, available at https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results*

[5] *List of Olympic Games host cities,* available at https://en.wikipedia.org/wiki/List_of_Olympic_Games_host_cities#References

[6] *JavaScript Trends in 2018, available at https://codeburst.io/javascript-trends-in-2018-3fb0077259*

[7] *Bostock, M. (2012) Data-Driven Document. Available at d3js.org. Accessed 31 Oct. 2012.*

[8] *Dewar, M (2012) Getting Started with D3. O'Reilly Media Inc. USA.*

[9] *Data Structures D3.js Accepts, available at https://www.dashingd3js.com/data-structures-d3js-accepts*