Olympics Data

Brett Sutow

11/16/2020

# Problem Statement and Question

Summer Olympics: We are going to look at this data to determine if individual country or person has a larger impact on Olympic victories. It will go into detail to look at winners of the Summer Olympics who have won multiple Olympic medals, from their various countries. It will also look at the detailed information about where the events took place. It will also go into detail about the specific sports. People will be interested in this because it allows people to understand when and where medals were won, and if country has an impact.

How has the Olympic games changed over the years? Are there countries that have a better chance at winning games? Do we see specific sports that are dominated by countries or areas? Are we more likely to see repeat winners if they participate multiple times? Does winning one medal type lead to winning a different one?

## Setup and Data Overview

A picture containing timeline

Description automatically generated

## 

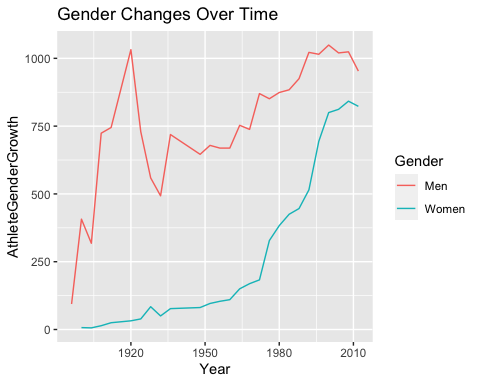
## How the Olympics have changed over the years? The below shows the changes in sports and athlete winners over time

The graphs below show how the Olympic games have changed over time. The graphs highlight the number of increase we have seen in athletes over the years, as well as highlighting the number of sports increases. The final graph shows the increase male & females over time as Olympic participants.

Chart, line chart

Description automatically generatedChart, line chart

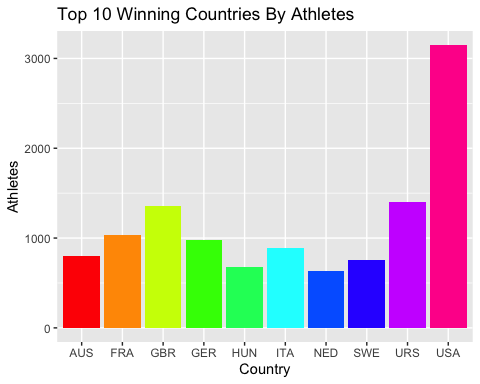
Description automatically generated

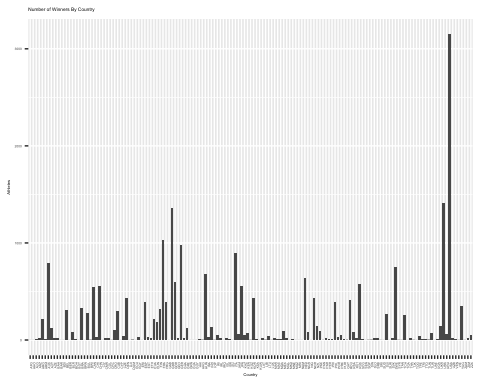


## 

## Shows the countries that have had the most winners: Gold, Silver, and Bronze totaled

The graph below shows the top countries that have won in the Olympic games over time. This is a total of Gold, Silver, and Bronze medal winners. It highlights the overall number of winners.

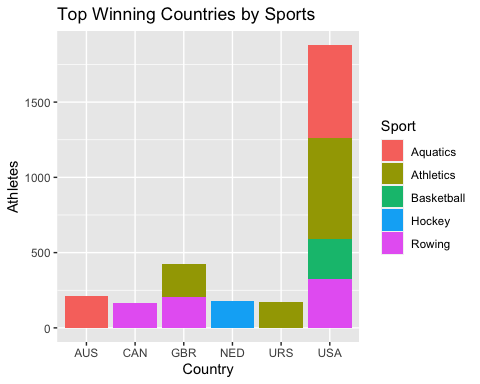




## 

## Country winners by specific sport type

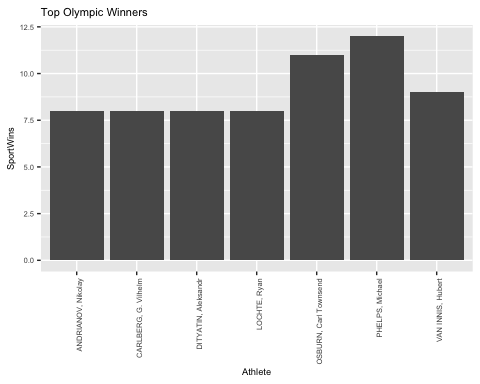
The graph below shows the top countries that have won in the Olympic games over time. This is a total of Gold, Silver, and Bronze medal winners. It highlights what countries dominated specific sports through showing how many athletes have won that sport. In this graph it shows the dominance by the USA in aquatics.



## 

## Athlete winners and outcomes

The graph below shows the top athletes that have won in the Olympic games over time. This is a total of Gold, Silver, and Bronze medal winners. The overall leader is Michael Phelps with Osburn being a close second behind him.



## Overall, write a coherent narrative that tells a story with the data as you complete this section.

The data provides several insights on the various different aspects of the Olympic games from 1896 to 2012. The data provides valuable information that can create correlation and relationships between the data points. Due to the fact that most of the points are categorical, the data set was looked at in mainly graph formats. Through this clear graphical observations were able to be made. The data tells a story about how over the years the Olympic games have grown significantly. It also presents clear information that ties with historical events why there are dips in various data points. The Olympics have grown in participation, sports, and countries over the past 100 years. Through this output the information could be clearly analyzed.

## Summarize the problem statement you addressed.

The problem statement was to see visually if there is a relationship between athletes, sports and countries over the past 100 years of the Olympics. The goal was to see if there is a better chance for a specific country or athlete to win the games. It was also to look at which countries have dominated the games, and if the athletes have something to do with that. Rather than looking at gold medal wins, we looked at all medals wins, because these are considered to be winners as well. Gender was also a key data point that was looked at to gain a better understanding of correlations. Looking at the various medal totals and athlete wins, we were able to properly conclude various relationships between the data points.

## Summarize how you addressed this problem statement (the data used and the methodology employed).

The problem statement was addressed through the process of combining various data points in tables and graphs, through this process it was more easily analyzed. The first step in the process was to group the Olympic games by year. From there I analyzed the changes in athletes over that time period. Which showed a clear increase over the past 100 years in athlete participation, it also showed the various dips that can be attributed to historical events i.e. WWI and WWII. Still utilizing the year group I then analyzed the increase in sport discipline over time. Which showed the increase in sports as time has gone on. Then year was utilized to see the various changes in gender over time as well. Showing the clear increase in women since the beginning of the Olympic games.

The next point that was analyzed was the winners over time via both countries and athletes. Through looking at the data points it showed the top winning countries per athlete wins. As well as showing just overall the various winning countries. When analyzing the breakdown of top winners it shows clearly that the USA has historically been the top winner. The top winning countries were then broken down by sports to see if there have been countries that have historically dominated sports, which showed that the USA has dominated several sports. Proving that the USA is the top performing country in the Olympics.

The final subject that was analyzed was the top performing athletes. This was looked at through the particular athlete and sporting events they participated with. The important thing to note about this data outlook is that it provides you with an idea of who are repeat winners in various different events. The data shows the connection of top winning athletes in particular sports, and the table gives an outlook of their relationship with various countries.

## Summarize the interesting insights that your analysis provided.

Through the vast history of the Olympics, of over 100 years.

The United States one of the still relatively young countries around the world has dominated.

Michael Phelps is the top performing Olympic athlete of all-time o Large name 1st world countries are typically the top performers o Gender has increased significantly over the years. There is still a gap with women, but they have caught up.

Through the plots you can see the significance of historical events impact on Olympic events. This can be seen with years like: Great Depression, WWI, WWII. This seemed to send women on an increased trajectory.

## Summarize the implications to the consumer (target audience) of your analysis.

This is shows various sports data points for Olympic fans. It allows fans to gain a better understanding of how their country has performed, as well as changes over time in the Olympic games. It also provides a great understanding of the various athletes and their performances in specific sports. Overall it shows fans that the top performing country has been the USA, and that their swimmer Michael Phelps has been the top athlete. This information is based off medal wins. Many fans or data scientists would be able to utilize a data set like this to predict various future winners.

# Discuss the limitations of your analysis and how you, or someone else, could improve or build on it.

The data utilized mainly categorical information. Which made analysis much more difficult. If someone was more experienced, they would’ve had an easier time analyzing the information. Going off of that information having more data would’ve made the analysis much easier as well. Not specifically number of athletes but, other information like weight, height, or athletic numbers/ performance. These data points would’ve made it easier to predict potential winners and outcomes. I would’ve been able to go further into depth about the specifics of various athletes and teams, and why they were successful.

# References:

*120 years of Olympic history: athletes and results*. (2018, June 15). Kaggle. <https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results>

Sports Reference | Sports Stats, fast, easy, and up-to-date. (n.d.). Sports-Reference.Com. <https://www.sports-reference.com/>