
120 YEARS OF OLYMPIC GAMES



Final Report

-ISM6419.792S22 Data Visualization

-KEERTHANA YELCHURI
U6516-2476

Contents

Introduction:	3
Methodology:	4
Data Processing:	4
Data Model:	4
Data Fields:	5
Analysis:	5
Conclusion:	11
Sources and References:	11

Introduction:

The Olympic Games are a well-known sporting platform that has been acknowledged all over the world since the late nineteenth century. Its origins, on the other hand, can be traced back to the Greek empire around 3,000 years ago, when it was limited to a single sprint race held in Greece's Olympia and only open to freeborn Greeks. It has evolved since then and is now a centre for all athletes from across the world to demonstrate their abilities in more than 28 different sporting competitions. It is currently hosted every two years in various countries under the labels Summer Olympics and Winter Olympics, each with its own set of games. It has evolved into a space that represents the individual candidates' power and serves as a source of pride for the countries they represent. It is one of my favorite seasons, and I eagerly anticipate it each year. Many athletes become generational heroes, waving their countries' flags and, more importantly, providing a good example for children and youngsters.

From 776 BC to 393 AD, the ancient Olympic Games were held in Olympia, Greece. After 1503 years, it resurfaced. In 1896, the first modern Olympics were held in Athens, Greece. The Games are referred to as the "Modern Olympics" from Athens 1896 to Rio 2016. Between 1912 and 1920 and 1936 and 1948, there are two long periods without any Games, both related to WWI and WWII.

The purpose of this report and visualizations is to answer some research questions that I've been considering. During the preparation of this report, I focused on the Tableau visualization tool because of its ease of use and different options for enhancing the appearance and feel of the graphs. It also speeds up the process of creating an interactive presentation, which is useful for delivering more data and comparing it all on the same page or dashboard. The most important advantage of visual analytics is that it makes complex data easier to analyze while presenting it in a clear, succinct, and suitable manner. This report gives a useful visual examination of the Olympic games from 1896 to 2016.

The following are the questions:

1. What is the average age of Olympic athletes, and how many medals have they earned overall, by gender and sport?

In this question, I'm trying to figure out how an athlete's age affects his or her ability to win or lose. I'm also curious to see if female athletes win more or fewer medals than male athletes.

2. Do GDP and population have any bearing on a country's ability to win? Find out how some of the world's best countries fared in the Olympics.

My goal is to see if a country's GDP and population have any bearing on how many medals it receives. Does having high GDP help the country to win more medals? I'd also like to know how many medals some of the top countries have won in total, as well as in which host cities they won them.

3. Identify a trend in the number of female athletes competing in the Olympics and the number of events in Olympics.

I'd like to know if the total number of female Olympians competing in the Games has increased or reduced over time, as well as whether the number of events has increased or decreased.

Methodology:

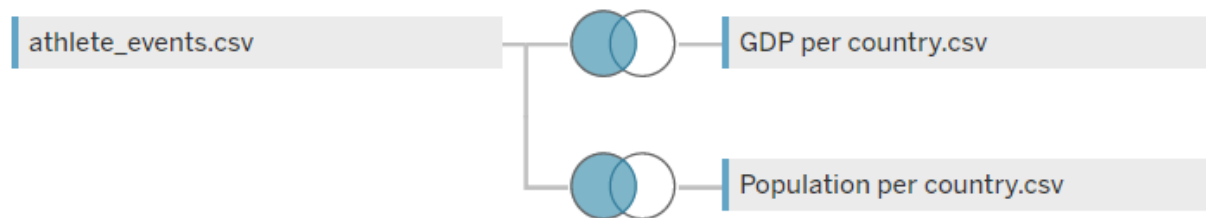
To analyze and answer to my research questions I got information from a variety of sites, including Kaggle and Gapminder.

Data Processing:

On the three datasets I utilized for this project, namely “120 years of Olympic history: athletes and results”, “Gapminder GDP per capita, constant PPP dollars- v25” and “Gapminder Total Population v6”, I used Excel to perform some simple data transformations and cleaning. The following is a list of the datasets I utilized and the data processing I performed on them:

- a) Athlete_events: This dataset has the information about all the players who have played in Olympics starting from Athens 1896 to Rio 2016. This dataset was the main dataset that I have used and did not need any kind of pre-processing.
- b) GDP per country: From 1990 to 2019, this dataset contains year-by-year data on each country's total GDP. This dataset also includes a column containing the three-letter abbreviation of each country in lower case, in addition to the country name. In this dataset, the only change I made was to capitalize the column with the abbreviations.
- c) Population per country: From 1990 to 2019, this dataset contains year-by-year information about each country's total population. This dataset, like the "GDP per nation" dataset, has an extra column with each country's three-letter acronym in lower case. I simply capitalized the column with the abbreviations in this dataset.

Data Model:



- a) A left join was used to link athlete_events.csv to the GDP per nation dataset. The constraint that Year (from dataset athlete_events) = time (GDP per country) and NOC (from dataset athlete_events) = GEO (GDP per country) is used to combine these two tables.
- b) Similarly, Athlete_events.csv is linked to Population per country dataset using a left join. The join between these two tables is based on the condition that Year (from dataset athlete_events) = time (Population per country) and NOC (from dataset athlete_events) = GEO (Population per country).

Data Fields:

The following is a list of data attributes for each data source utilized in this project:

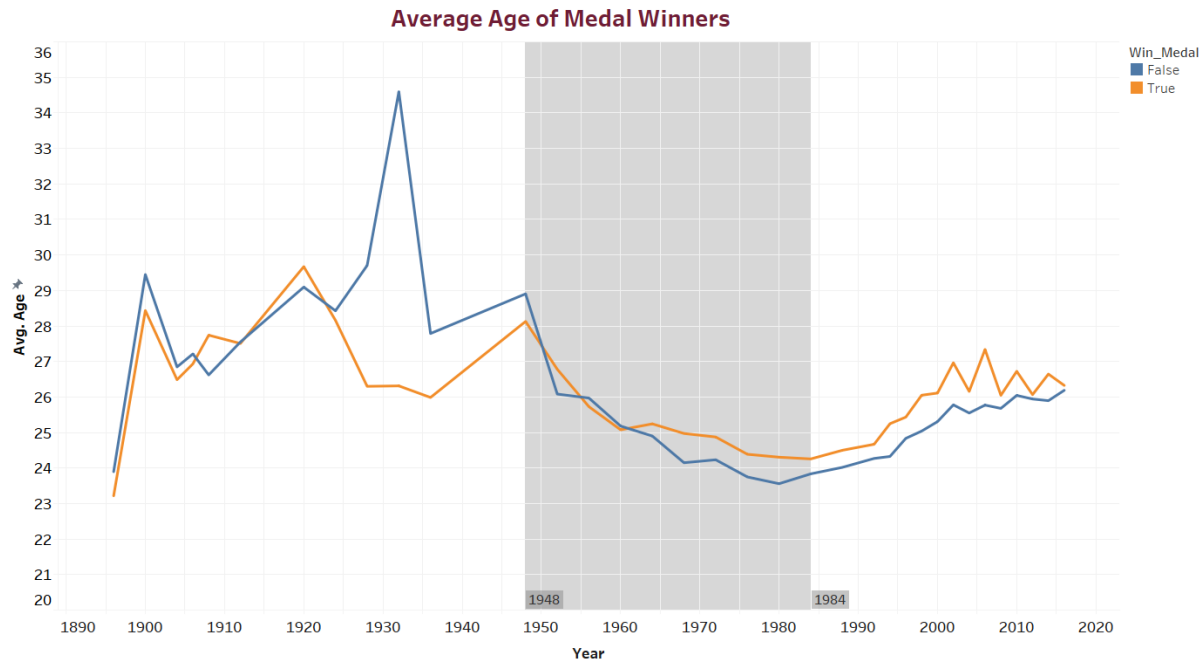
Data Source	Attributes	Details
athlete_events	ID	Unique number for each athlete
	Name	Athlete's name
	Sex	M or F
	Age	Integer
	Height	In centimeters
	Weight	In kilograms
	Team	Team name (Country)
	NOC	National Olympic Committee 3-letter code
	Games	Year and season
	Year	Integer
	Season	Summer or Winter
	City	Host city
	Sport	Sport
	Event	Event
	Medal	Gold, Silver, Bronze, or NA
GDP per country	GEO	3-letter abbreviation for each country
	name	Name of country
	time	year
	GDP Total	Total GDP of the country in the respective year
Population per country	GEO	3-letter abbreviation for each country
	name	Name of country
	time	year
	Population Total	Total population of the country in the respective year

Analysis:

To find answers to my research questions I've made a total of ten visualizations and two dashboards.

The below visualization displays data on the average age of competitors competing in the Olympics. We can see the average age of athletes who won medals (shown in orange) and the average age of athletes who did not win any medals (shown in blue). We can observe that the average age of athletes who won medals was high in the 1900s (about 27-28), and it changed slightly until 1948. From 1948 through nearly the end of the 1990s, the average age drops to 24-25, then rises to 26-27 until the Rio 2016 Olympics. On the other hand, we can see that in the different time frames, the average age of athletes who did not win

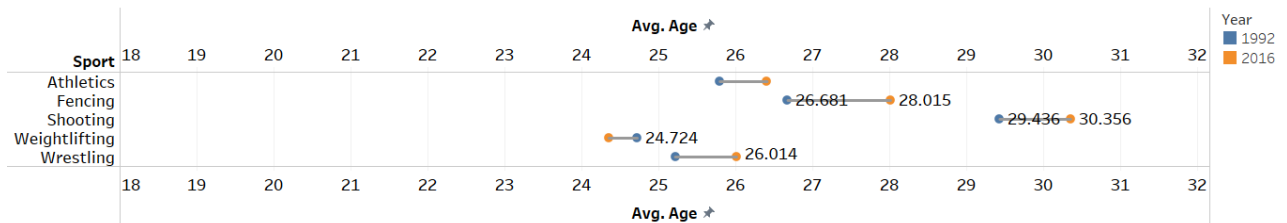
any medals was either higher or lower than that of those who did.



Let's take a look at the average age of the sports that have been ongoing since the modern Olympics began. I attempted to determine the age, particularly for the stable era, which lasted from 1992 to 2016. We can see that Shooting has the most age flexibility, with an average age of roughly 29-30 years, while Weightlifting has the least flexibility, with an average age of around 24 years.

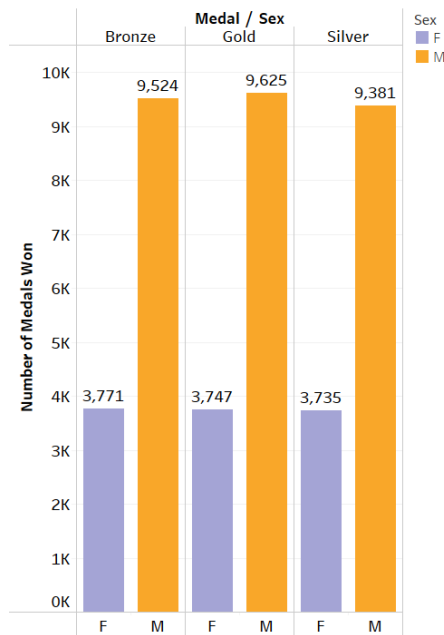
Avg Age by Sport - Stable Era

Stable Era was between 1988-2016



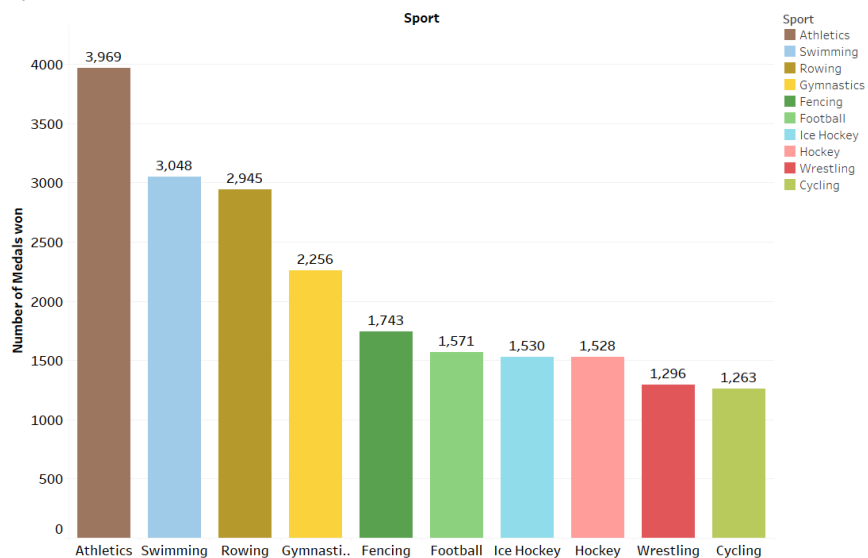
The below graph shows the number of medals won based on gender. We can see that the medals won by Men are almost three times compared to the medals won by female athletes. I was confused as to why this would happen and we might find our answer to this in the upcoming sections.

Number of Medals based on Gender and Season Between 1896 to 2016

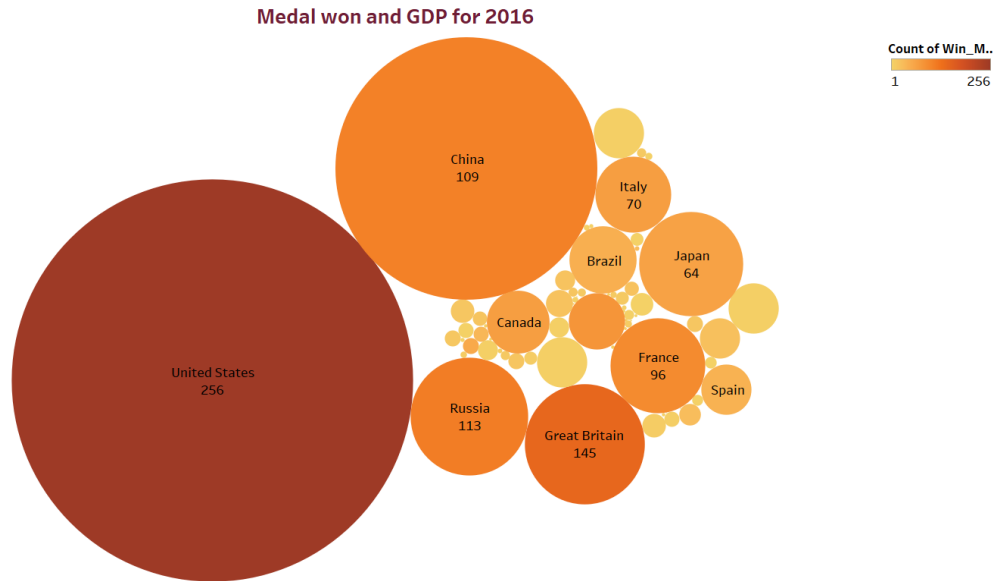


Let's look at the number of medals won in the Top 10 sports per season from 1896 to 2016. In the top ten sports at both the summer and winter Olympics combined, we can see that Athletics has the most medals won, while cycling has the least.

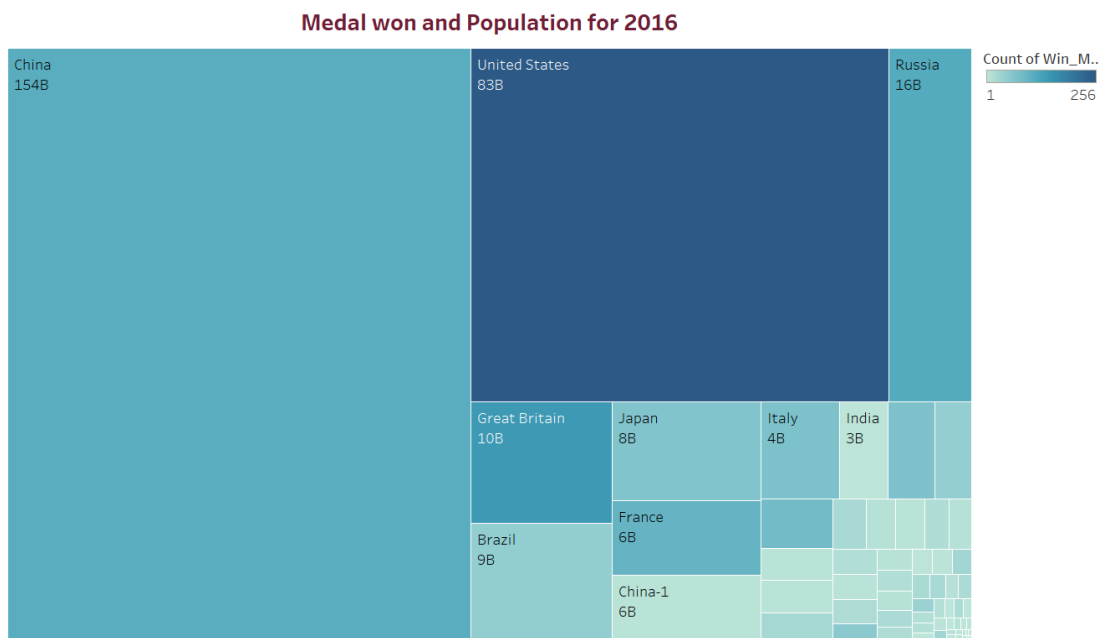
Medals won by Sports and Season Top 10 - Between 1896 to 2016



In the below visualization, for 2016, we can observe that the countries with the lowest GDP only have one medal, whilst the countries with the highest GDP have more than 50 medals, with the United States having the most with 265 medals. All of this shows that overall medal successes in the Olympics are linked to a country's total GDP.



In the next graph, we can notice some erratic behavior. It appears that a country's population has no bearing on its medal victories. Australia, with a population of over 2 billion people, has won roughly 82 medals, compared to Japan, which has an 8 billion people but just 64 medals.

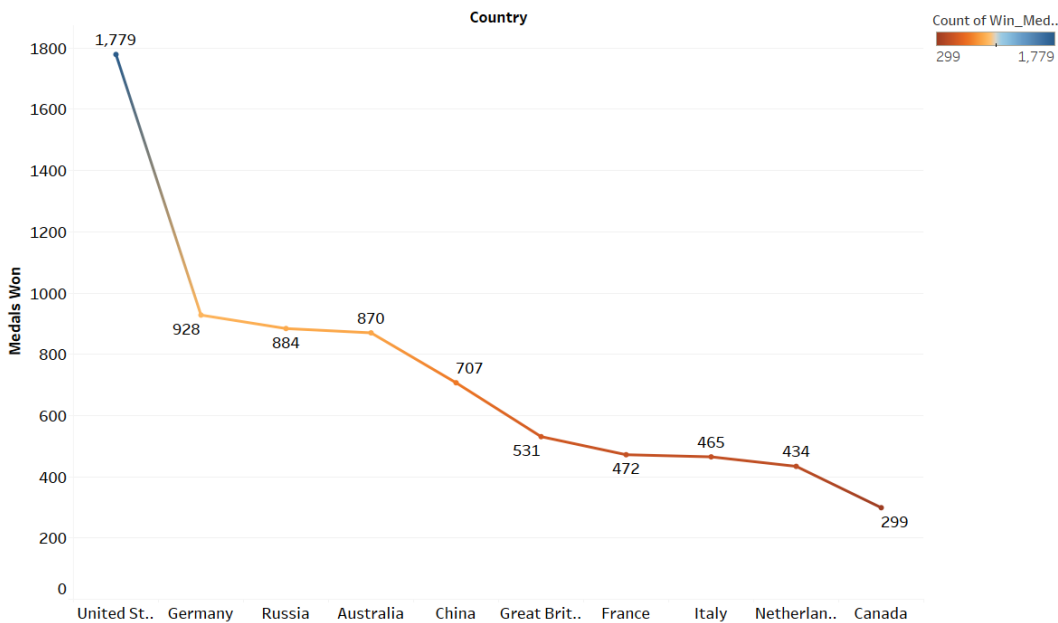


The below two graphs help us find the total medals won by some of the top countries in the world and also helps to understand the top host cities where these countries won these medals. This graph is for

Summer Olympics where we see that United Sates has the highest number of winnings. It won them in seven host cities including Atlanta, Beijing and Rio de Janeiro.

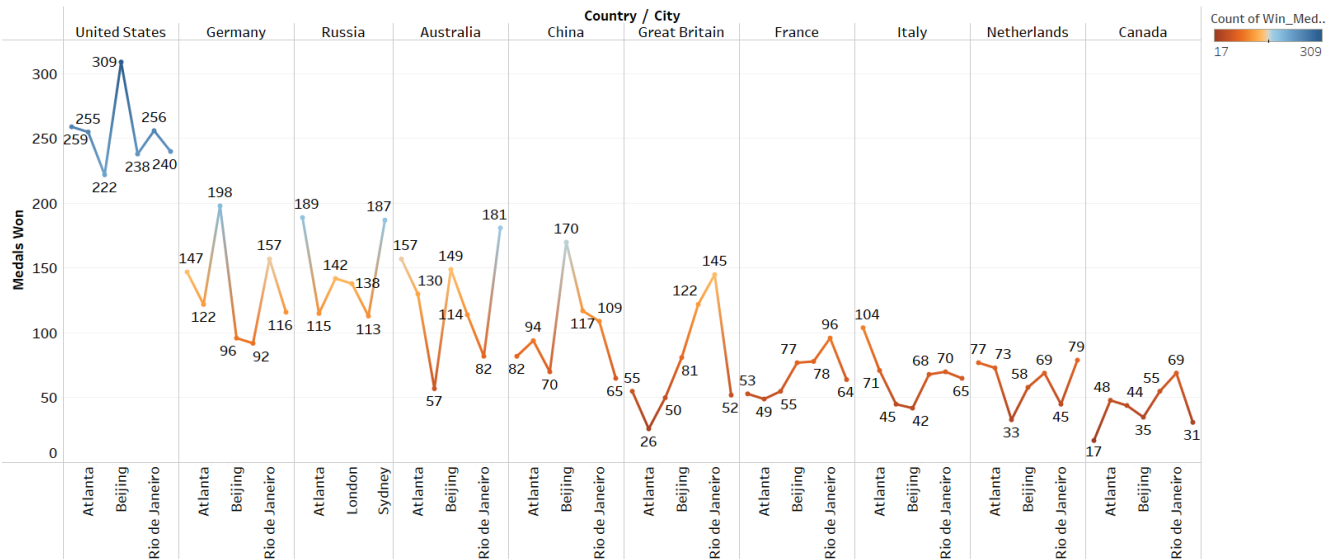
Medals Won By Top Countries and The Top Host Cities Where They Won These Medals -Stable Era

Stable Era was between 1992 to 2016



Medals Won By Top Countries and The Top Host Cities Where They Won These Medals -Stable Era

Stable Era was between 1992 to 2016

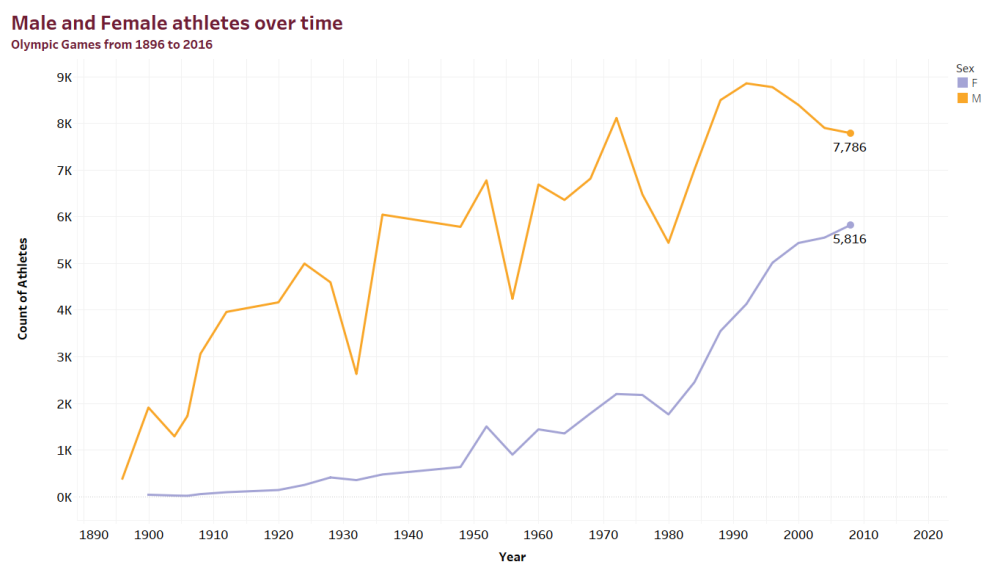


For 2016, we can find the countries who won the most medals during summer Olympics in the below map.

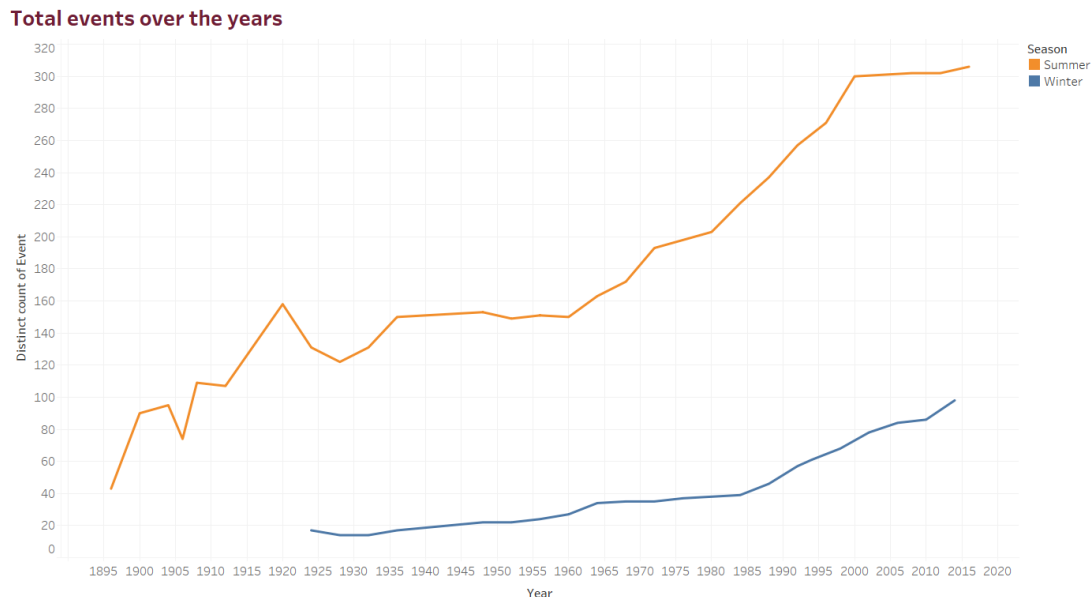


We know that humans have been generally sexist in many ways for a long time, unfortunately throughout history, and the inclusion of women in Olympic athletes is no exception. We can see if inclusion has increased or reduced at this worldwide event from 1896 to 2016 in the visualization. This chart is for the summer Olympics.

As you can see, there has been a noticeable increase in the involvement of the female sex in competitions; yet, the male sex continues to dominate, though the gap is not as large as it once was. This, I believe, is one of the reasons why on an overall scenario from 1896 to 2016, men's total winnings are nearly three times those of women.



The number of events in both the summer and winter Olympics has increased throughout time, as seen in the line graph below. We can also notice that the first winter Olympics were held in 1924, as well.



Conclusion:

- Looking at the economic effect, it looked like the total GDP of a country is more significant to find the winnings in the recent years compared to population.
- The age range of players winning the medals has decreased over the years and an optimal age for some of the sports can be identified in the recent years.
- Over the years, inclusion of female athletes has increased such that the gap between the number of females and male athletes taking part is minimal from 1896 to 2016.
- The number of events has increased over the years for both summer and winter Olympics.

Some of the **additional research questions** are:

- Does host countries have better chance of winning medals in Olympics?
- Has the number of athletes and countries taking part in Olympics increase over time or decrease?

Sources and References:

- I got the data about each athlete who took part in Olympics from 1896 to 2016 (athlete_events.csv) from the below Kaggle website.
<https://www.kaggle.com/datasets/heesoo37/120-years-of-olympic-history-athletes-and-results>
- The GDP data about each country (1990 to 2019) was from the below source:
 “Gapminder GDP per capita, constant PPP dollars- v25”
<https://www.gapminder.org/data/documentation/gd001/>
- The data about the total population of each country from 1800 to 2100 was from:
 “Gapminder Total Population v6”
<https://www.gapminder.org/data/documentation/gd003/>