

120 Years of Olympic History Data Visualization Project

Bojian Zhang

11/16/2018

Abstract:

In this project, Olympic Games dataset from Kaggle is used to draw some insightful charts. Selected tools are Tableau and python plotly. This project portrays a panorama of Olympic Games history, including medal distribution among countries, trend of athletes' physical condition, popularity of different sports and so on.

Introduction:

This project can be mainly divided into three parts: how many countries joined in Olympics Games and how many medals they won; What are most popular sports in history and what are their popularity trends; Demographic information of athletes.

At first, some hypothesis is issued.

Hypothesis 1: European countries won most medals in history.

Hypothesis 2: Athletics and swimming are popular sports.

Hypothesis 3: Average Weight and Height of athletes increased continuously.

Hypothesis 4: Average Weight and Height of European athletes are larger than that of Asian athletes.

Background:

This dataset contains 15 columns and 271,116 rows.

A dataset description is shown in following table.

Variable Name	Data Type	Variable Meaning
ID	Numeric	Athlete ID
Name	Text	Athlete name
Sex	Categorical	Athlete gender
Age	Numeric	Athlete age
Height	Numeric	Athlete height
Weight	Numeric	Athlete weight
Team	Text	Athlete team
NOC	Text	Athlete team

Games	Text	Olympic games
Year	Numeric	Olympic games year
Season	Text	Summer or winter
City	Text	Host city
Sport	Text	Sport name
Event	Text	Event name
Medal	Numeric	Athlete medal

Data Visualization:

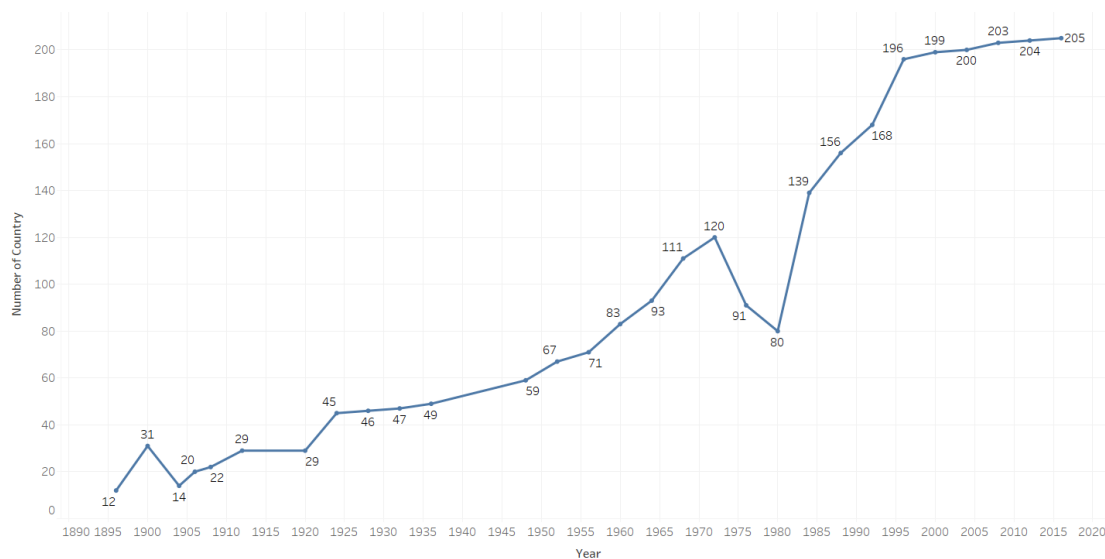


Figure 1. Number of Countries Attended Olympic Games

This line chart shows number of countries which joined in Olympic Games from 1896 to 2016. This line shows an upward trend, especially in last twenty years, almost all countries or regions in the world joined in Olympic Games. The 1976 Montreal Olympic Games was resisted by some African countries and the 1980 Moscow Olympic Games was also boycotted because Soviet Union invaded Afghanistan in 1979. So, number of countries declined in these two years.

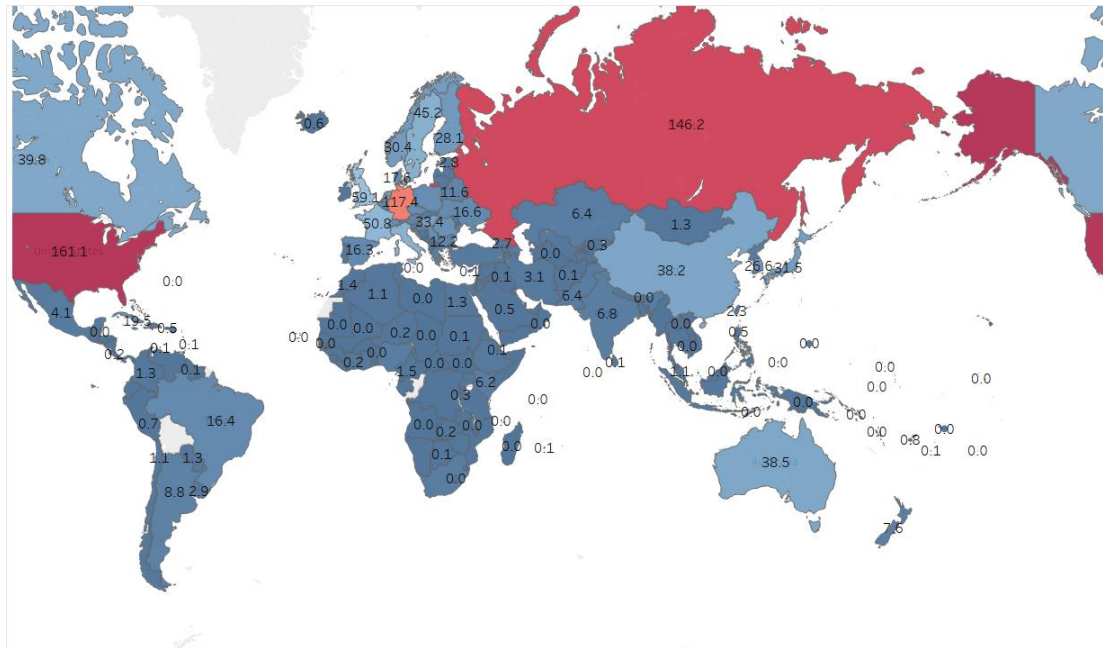


Figure 2. Average Number of Medals a Country Won in Olympic Games

Numbers in this map are calculated as number of medals a country won in history divided by number of Olympic Games this country attended. We can see from the map that the U.S., Russia and Germany are three sports powers, because they won many more medals than other countries. Another reason for their great performance is they are good at team event, they acquired many medals in one champion. North America, Europe and East Asia better performed in Olympic Games. African countries won very few medals in history. Hypothesis 1 is confirmed.

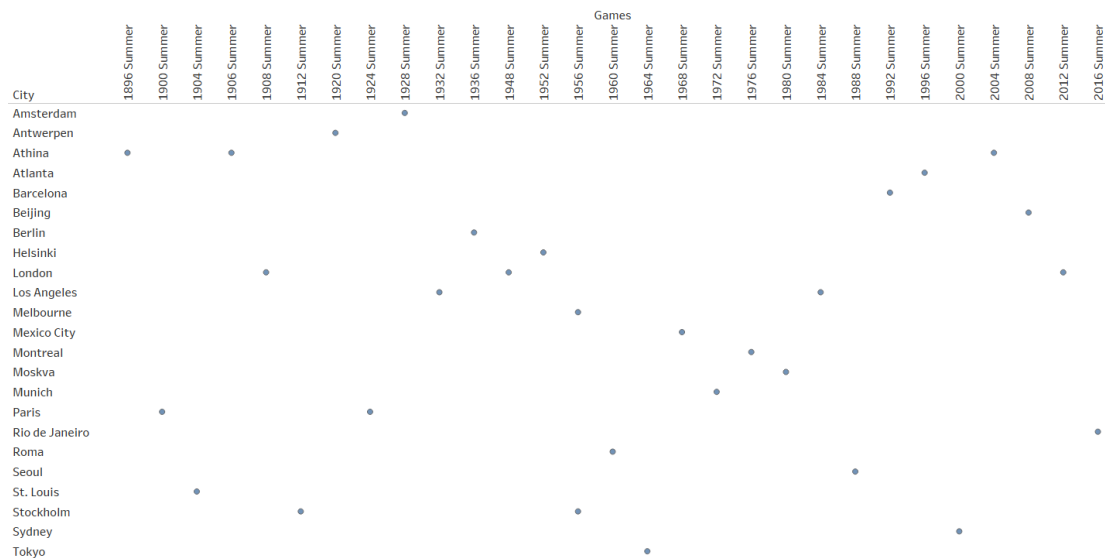


Figure 3. Host Cities Distribution

This chart shows host cities distribution. Some cities host more than once summer Olympic Games in history. For example, Athens and London host three summer Olympic Games in history.

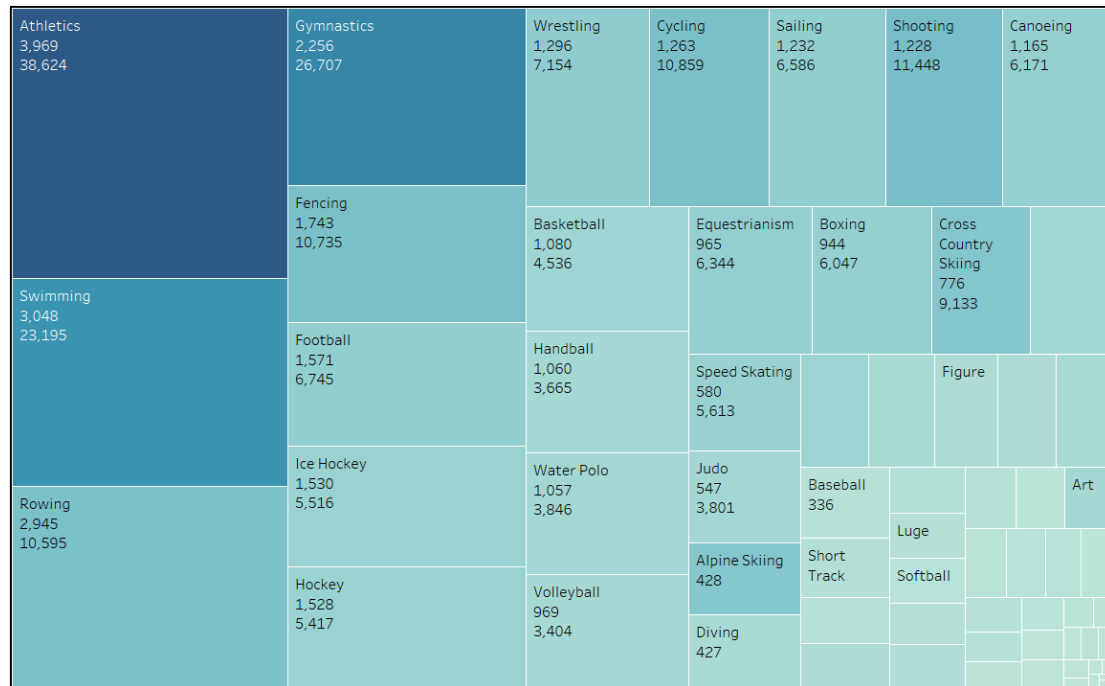


Figure 4. Medals and Athletes by Sport

Area of squares in above tree map indicates count of medals of a sport. Athletics has largest area and it produced 3,969 medals in history. Swimming and rowing are two popular sports as well, they produced 3,048 and 2,945 medals respectively. Color of squares in above tree map reflects count of athletes who attended a sport. Athletics, swimming and gymnastics are three popular sports. Large number of athletes participated in these sports. It is worth noticing that small squares with deep color, such as alpine skiing, are competitive sports. Because a large number of athletes fight for a small number of medals. Hypothesis 2 is confirmed.

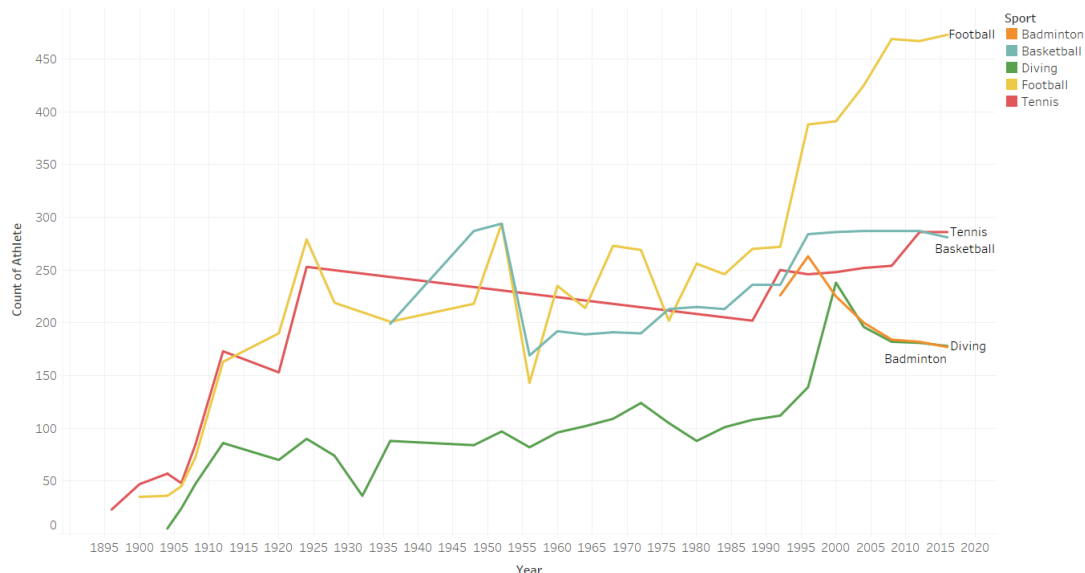


Figure 5. Popularity of five sports

This line chart compares trends of popularity of five sports. Tennis, diving and football became Olympic Games sports early. But they experienced different trends. Count of athletes of tennis and football rise sharply before 1925, then they remained stable until 1990. After that count of football athletes increased again, up to more than 450 in 2016 Olympic Games. While count of

tennis athletes remained the same as before. Diving is the least popular one, although count of diving athletes kept increasing, it declined after 2000 Olympic Games. Basketball became an Olympic Games sport in 1936. Count of basketball athletes is near to tennis. Badminton is a new Olympic Games sport, it has a history of less than 20 years. Count of badminton athletes is near to diving.

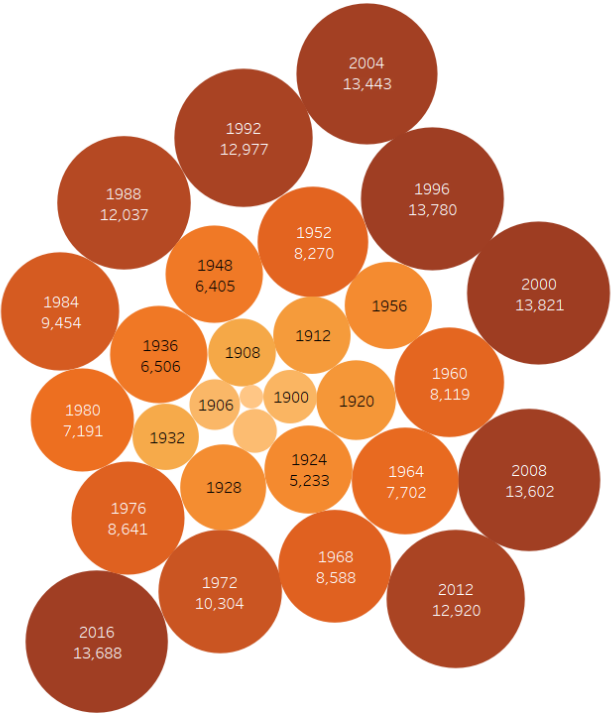


Figure 6. Number of Athletes in Summer Olympic Games

This bubble chart shows number of athletes in summer Olympic Games. Size of bubble represents number of athletes. From this chart, we can see that number of athletes increased in history.

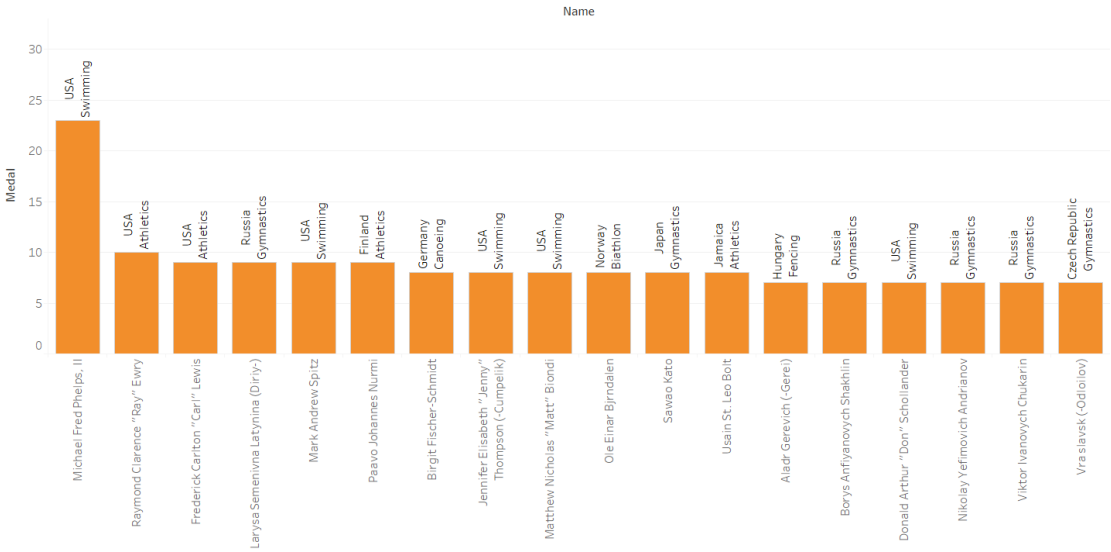


Figure 7. Star Athletes

This bar chart ranks athletes who won most gold medals in history. This chart shows name, nationality, sport of athletes and number of gold medals they won. Phelps, an American

swimmer won 23 gold medals in his career. These athletes concentrate on swimming, athletics and gymnastics sports. And most of them come from USA and Russia.

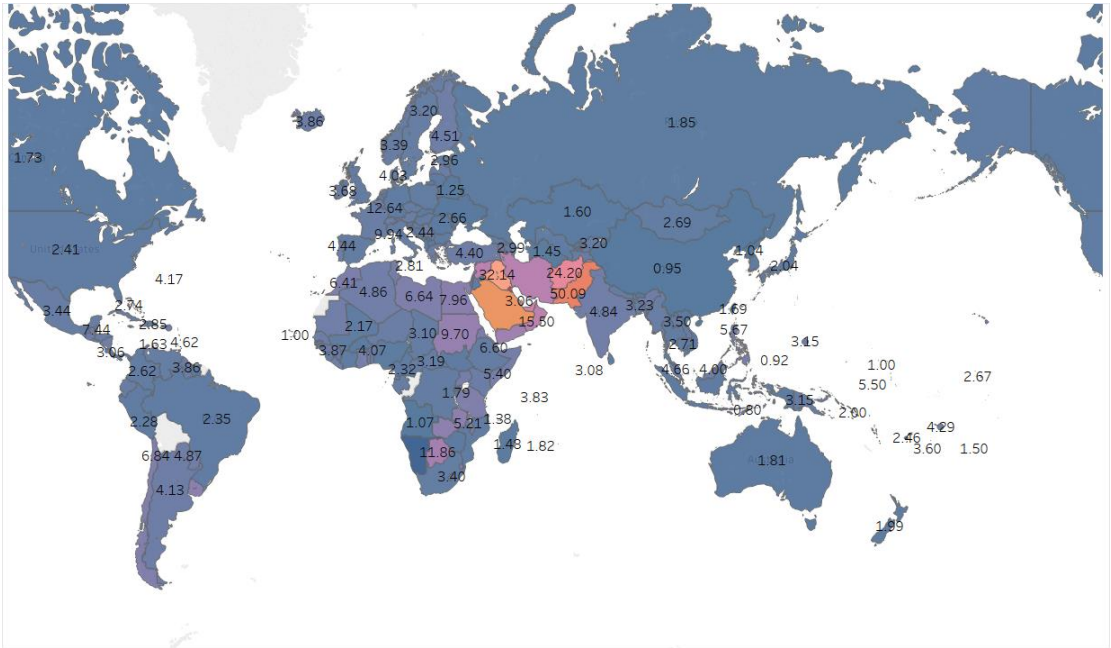


Figure 8. Athletes Gender Difference

Ratios in above map equals number of male athletes in a country divided by number of female athletes in this country. From the map we see that for most countries, male athletes are more than female athletes. Especially for some Islamic states, gender difference of athletes is very large. Chinese female athletes are more than Chinese male athletes.

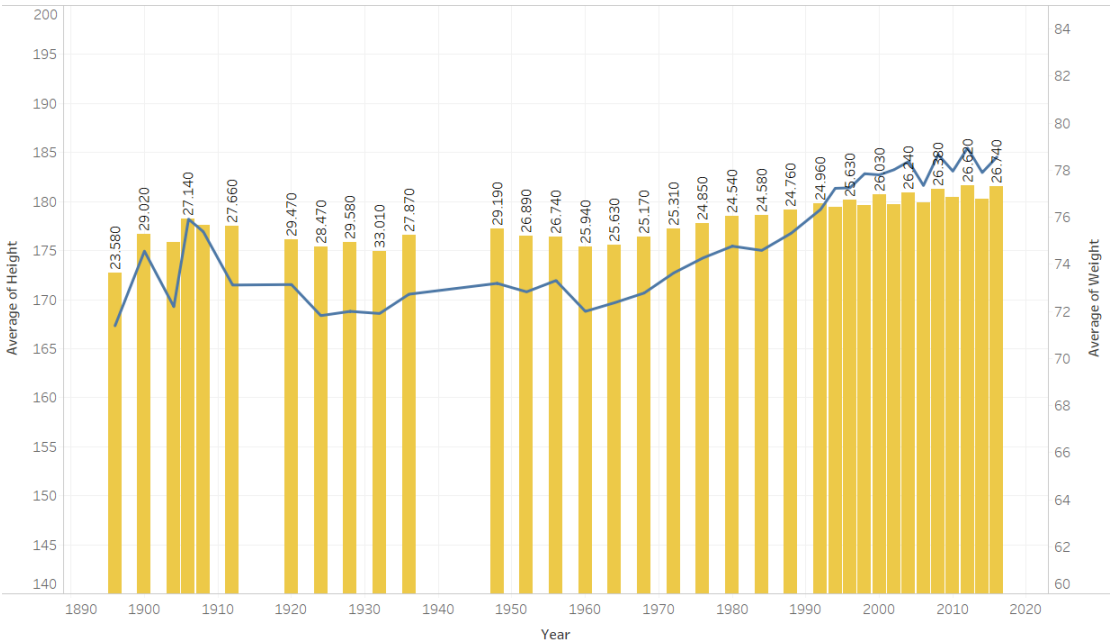


Figure 9. Average of Weight and Height of Male Athletes

Blue line in this combined chart displays average weight of male athletes in history and yellow bar indicates average height of male athletes in history. We can see from the chart that both average weight and average height of male athletes show upward trend, especially, after 1990,

two figures reached historic high. Number in the chart is average age.

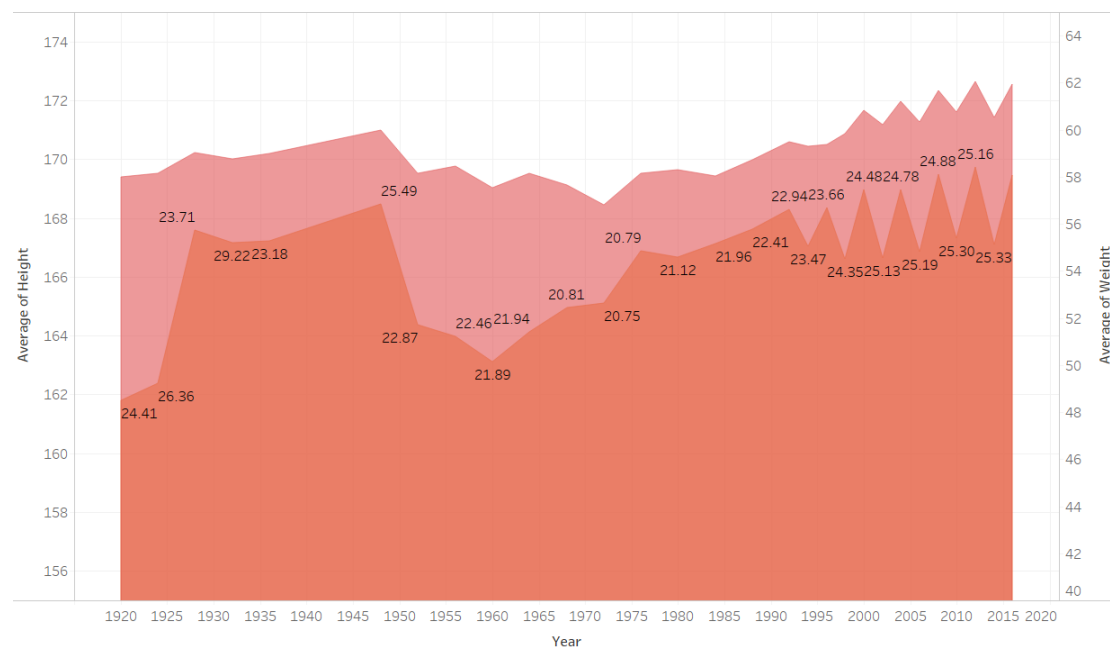


Figure 10. Average of Weight and Height of Female Athletes

As a comparison, this area chart shows average of weight and average height of female athletes. Pink area indicates average of weight and red area indicates average of height. These two figures also show upward trend. However, average weight and height reached high level before and after Great Depression and World War 2. Hypothesis 3 is rejected.

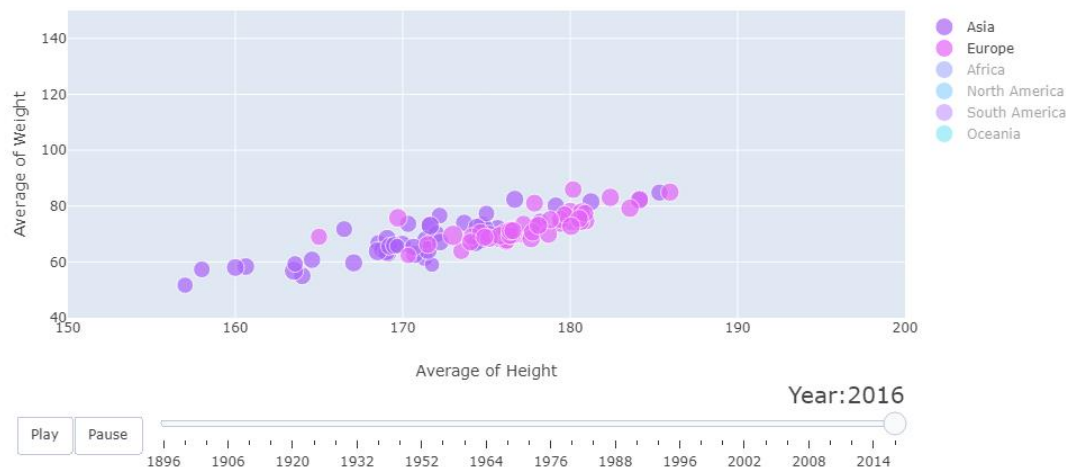


Figure 11. Interactive and Dynamic Scatterplot

This is an interactive and dynamic scatterplot drawn by python plotly. Area of circles means average age of athletes. From this screenshot we can figure out that average of height and weight of European athletes are larger than that of Asian athletes. Hypothesis 4 is confirmed.

Conclusion:

Olympic Games develops well in past 120 years, although it was influenced by international political situation in history, nearly all countries or regions in the world attend to this international game now. Countries in North America, Europe and North east Asia are sport powers because they won most medals and host most Olympic Games. Some sports are very popular in the world, while some sports, such as rugby, softball and some winter Olympic Games need more attention. We should encourage more female athletes to join in Olympic Games and developing countries to develop their sport industry.