120 years of Olympic Games history

Would it be possible to predict the future Olympic Games outcome with past data ?

Big Data Treatment

Lorine Pose, Théo Rup

Introduction

This report aims to present the work we did about the treatment of the database *120 years of Olympic history: athletes and results* found on www.kaggle.com. This dataset includes all the data from Athens 1896 games to Rio 2016 games, but we decided to keep only summer Olympic Games. The Olympic Games are a world-famous event gathering the best athletes from all around the globe. Each athlete represents his/her country during one or more events he/she is involved in. For each event, the 3 best athletes receive respectively a gold, silver and bronze medal. In this work, we tried to bring a statistical schema out in order to predict the next Olympic Games data in 2021.

I : Evolution of the competition format

In this part, we will study the competition format and its possible evolution over the editions. For this, we have displayed the number of events offered to male and female athletes. We obtained the line chart shown in Figure 1 from which we understand that the Olympic Games are not "fixed" but, on the contrary, in constant evolution. At first, we can see that men and women did not compete in the same number of events. However, this gap is reducing edition after edition. Secondly, we can note that, after a nearly continuous growth, the number of events for men tend to stabilize around 175 from now on.

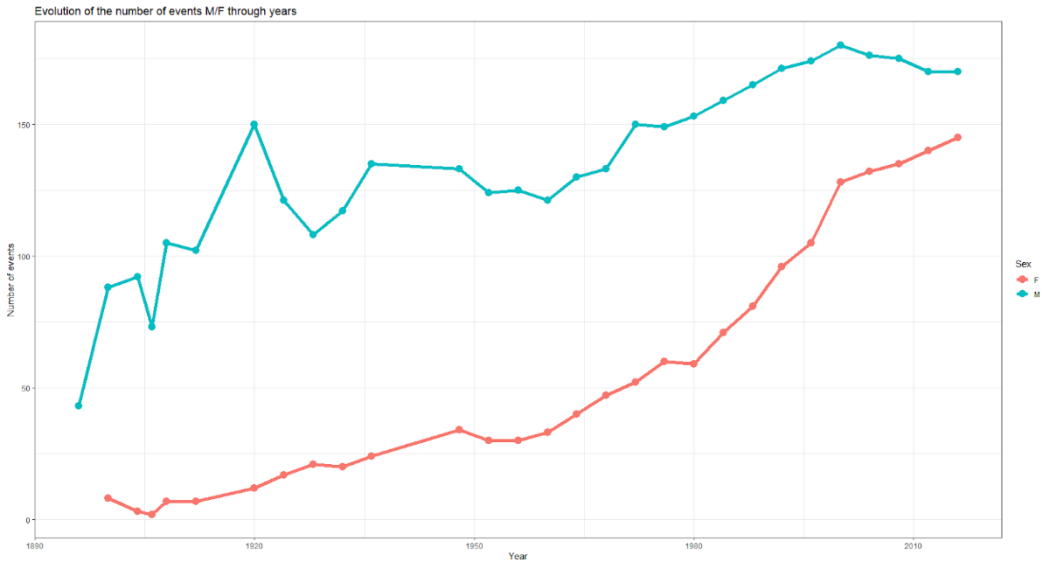

Figure 1 : Evolution of the number of events through editions depending on the sex of the athletes

The evolution in the number of events goes along with the number of registered athletes as shown by the bar chart in Figure 2. This emphasizes the Olympic Games will to gather the world of sport. We logically observe a tendency to have as many participants of each sex.

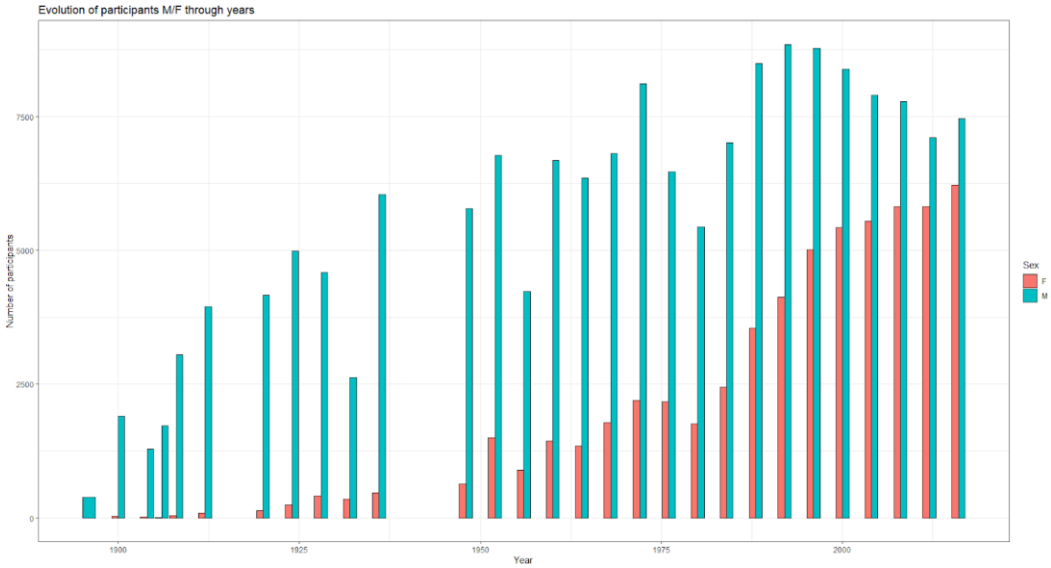

Figure 2 : Evolution of the number of athletes through editions depending on their sex

II : Influence of the origin of the athletes

We also studied if one or more countries were more represented among the winner athletes. For that, we displayed, for each edition, the 5 countries hosting the greatest number of Olympic champions (athletes who won the Gold medal). Figure 3 shows that a small number of countries is always taking the lead on the world ranking. Among these countries, the USA, Russia, China and the European countries are majority. It is important to note that the USA and Russia were respectively not present during the 1980 and 1984 Olympic Games for political reasons. To sum up, we can see that there is a link between the power of a country and its Olympic ranking. It might be because developed countries are making more investments in sport policies (infrastructures, trainers' formation…).

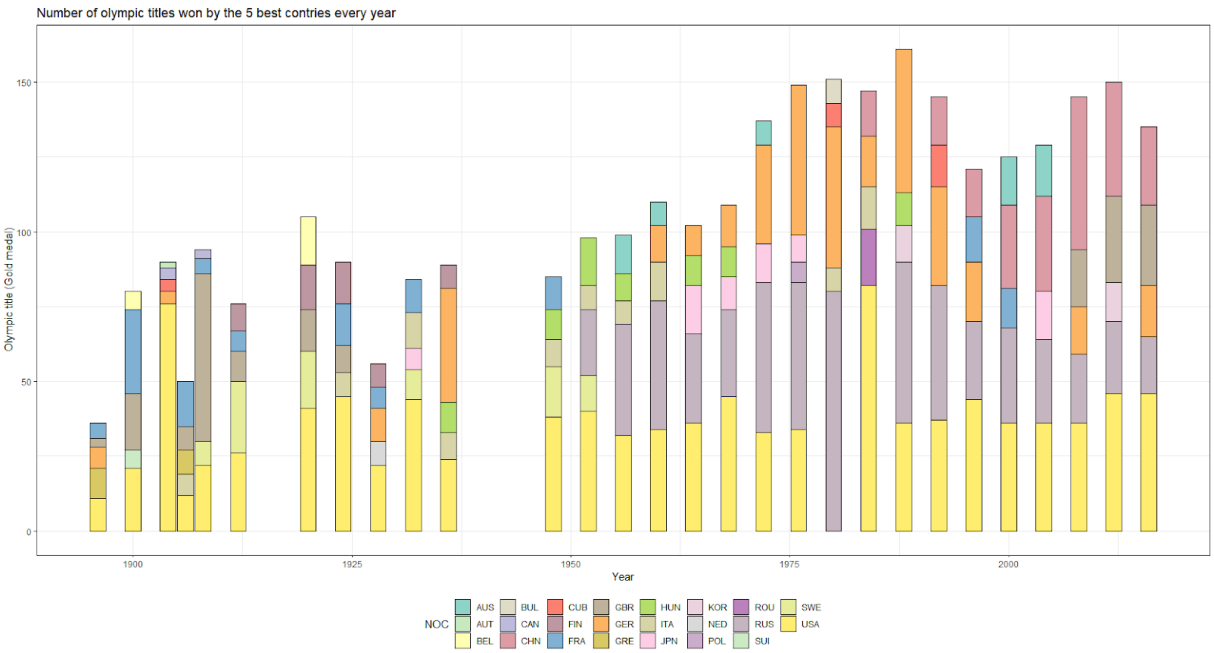

Figure 3 : Top 5 countries in the Olympic Games ranking since 1896

III. Influence of the physical characteristics of the athletes per discipline

We decided to analyse the height and weight characteristics to see how much it influences and differentiates the winning athletes from their adversaries. For the displayed results to be clearer, we only chose to study 5 disciplines that have different physical requirements and are among the most popular:

- Swimming Men's 100 metres Freestyle
- Gymnastics Men's Parallel Bars
- Basketball Men's Basketball
- Athletics Men's Shot Put
- Athletics Men's 100 metres

We studied the evolution of height and weight of the participants of each discipline (Figure 4 and 5). The linear regressions show that, depending on the discipline, the evolution of the weight and the height of the participants can be major. Except for the athletes practicing parallel bars gymnastics , the athletes in the selected disciplines tend to become taller and therefore to gain weight. Proportionally, the basketball players will be the tallest and the shot put athletes will have more imposing stature. Following those results, we can expect that athletes' physical characteristics in each sport are following a tendency that can be compared to natural selection to reach better performance in their fields.

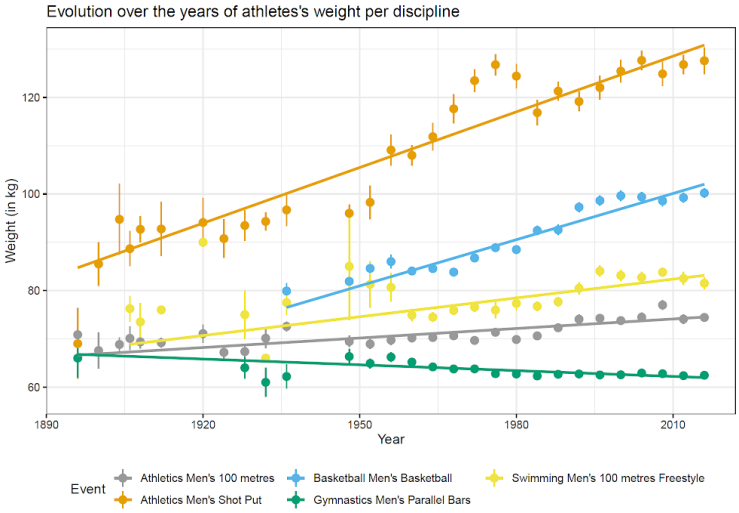

Figure 4 : Evolution over the years of athletes' weight in the selected disciplines

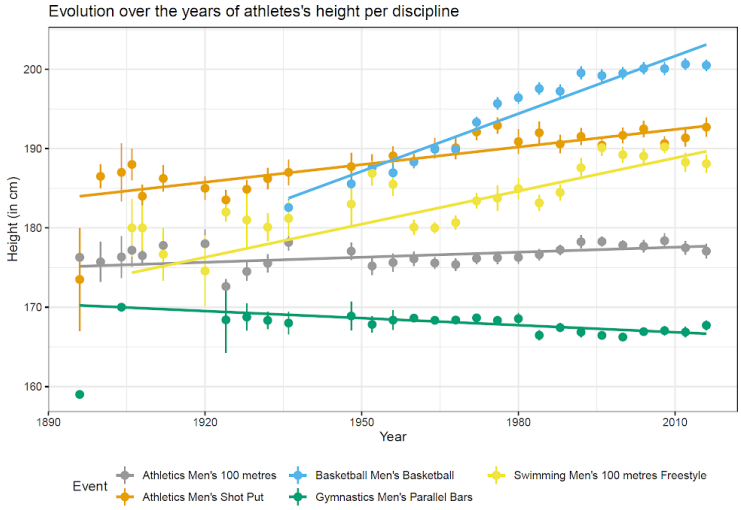

Figure 5 : Evolution over the years of athletes' height in the selected disciplines

Now that we know that the physics of the athletes of each discipline changes through the years, we take interest in how much the weight and height of an athlete will impact his chances to win. In order to do that, we compared the mean height and weight of all the participants of a discipline with the mean height and weight of the winners of those same disciplines. This analysis is based on the Olympics Games that took place since 2013, given that as we saw before that the stature of the athletes changed drastically and considering the previous data could skew the results.

As we can observe in Figure 6 and 7, even though the difference is not phenomenal, it seems to corroborate some of the previous tendencies. Indeed, for athletes practicing parallel bars, it seems preferable to be smaller and lighter as the evolution curves tends to show. Contrarily, the swimmers 100m athletes (and even basketball players) taller and bigger seemed to perform better. However, the logic is not respected for the shot put athletes.

It is important to notice that these analyses are only valid for the disciplines we chose. The results could be different for any other discipline depending on the physical requirements needed to excel in it.

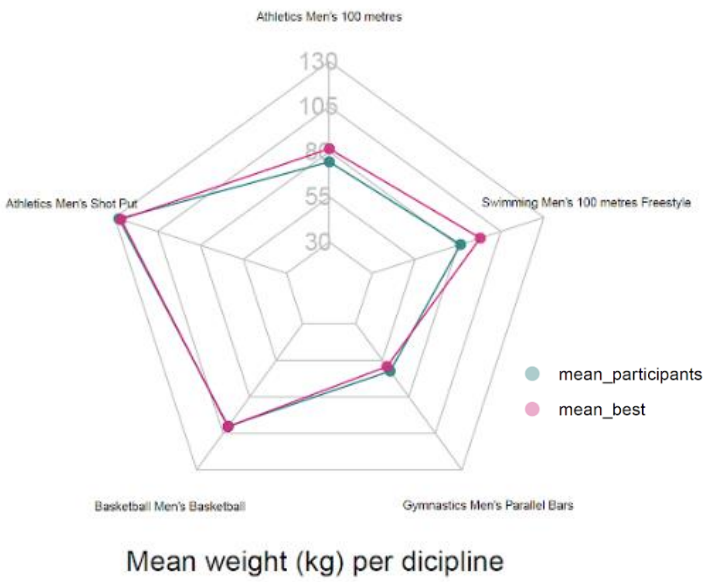

Figure 6 : Radar chart comparing the mean weight of all the athletes and of only the winners for the selected disciplines

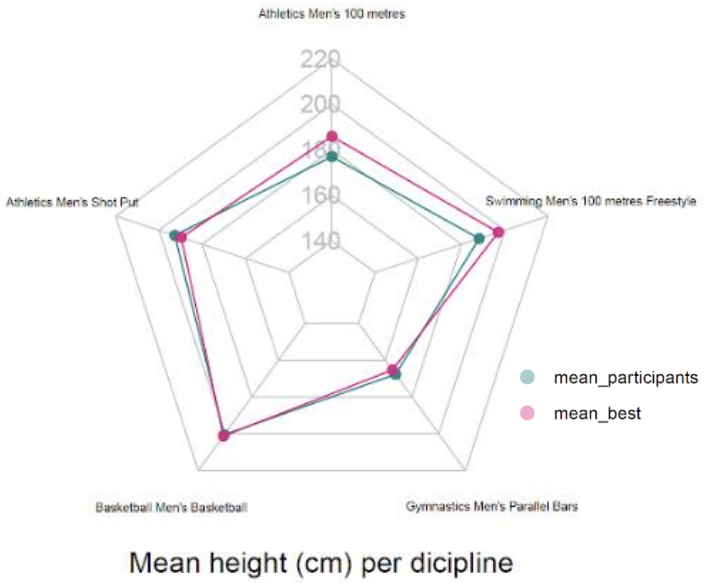

Figure 7 : Radar chart comparing the mean height of all the athletes and of only the winners for the selected disciplines

Nonetheless, the difference in these cases does not seem important enough to justify the ability to win of the athlete in each discipline, at least not only based on its height and weight.

Conclusion

The Olympic Games have developed significantly during the last decades. While the number of male disciplines tend to stabilize, the female ones are catching up rather fast. It could be strategically interesting for countries to invest on the female sport development to be a more competitive nation in the future. It is sure that the one who will adopt this policy will be more represented among the winners.

We could also easily notice that the countries more represented among the winners are the more powerful ones. This could be due to the possibility of those countries to invest in the development of sports, for instance by providing better sport infrastructures and encouraging such practices, giving access to the necessary resources. Based on this reasoning, we can assume that a majority of the winners of the next Olympic Games will be from those powerful countries, mostly from the USA but also from China, Russia and the main European powers.

Among a set of Olympic disciplines, we have analysed a linear evolution of the height and weight of the athletes, tending towards perfect proportions making them “adapted” to their sport. Nonetheless, these factors do not allow us to assume the future winners, since a lot of other factors and data are still missing from the dataset and couldn't be analysed to give a more complete evaluation of the subject.

To sum up, we can only point a trend for the upcoming years but it remains impossible to give the magic recipe transforming an athlete to a winner; which makes good sense with sport !

References

[1] 120 years of Olympic history: athletes and results (May 2018)
Available: https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results