Big Data

Ting Chen, Tehema Teiti

December 21, 2018

1 Introduction

Nowadays, massive amount of data are available and are provided freely on specialized websites like kaggle, covering a lot of domains. With a large amount of data, it is possible to analyzed them in order to understand the real world, comparing variables between them and being able to deduce conclusions. It is important to define an interesting issue on which we will answer it by interrogating accurate data sets and producing relevant analyses.

For this project, we have chosen a data set from kaggle about the history of the Olympic Games since Athens 1896 [1]. This data set contains 271 116 observations of 15 variables, representing the results for each athletes, regarding many informations like the year, the country, the team...

What are the parameters that impact on the winning a medal at the Olympic Games ? In order to answer this issue, this document will present the analyzes that we made on this data set through four graphs that we will explain.

2 Analyzes

First, we have been interested by the ranking of the countries that have the most medals. The figure 1 represents the top ten countries that have the most medals over the years since 1896. We can see that the United States leads this ranking with 5637 medals, having more than the double of medals compared to the second country with 2503 medals (and even more than the following countries).

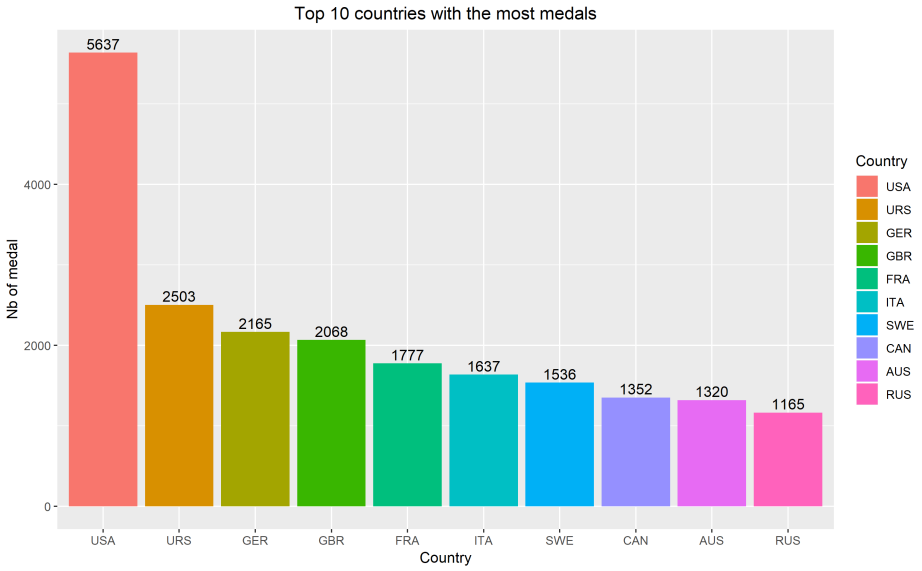

Figure 1: Top 10 countries with the most medals over the years

Then, after seeing that the United States leads this ranking, we focused on the number of athletes of each country of the top ten. The figure 2 represents the number of athletes of the top ten countries. On this graph, we can see that the United States has also the most athletes among the top ten countries. However, the second country, representing the Soviet Union, is the second country having the less athletes among the top ten countries. Regarding the trend of this graph, there is not a strong correlation between the number of athletes and the number of medals for a country.

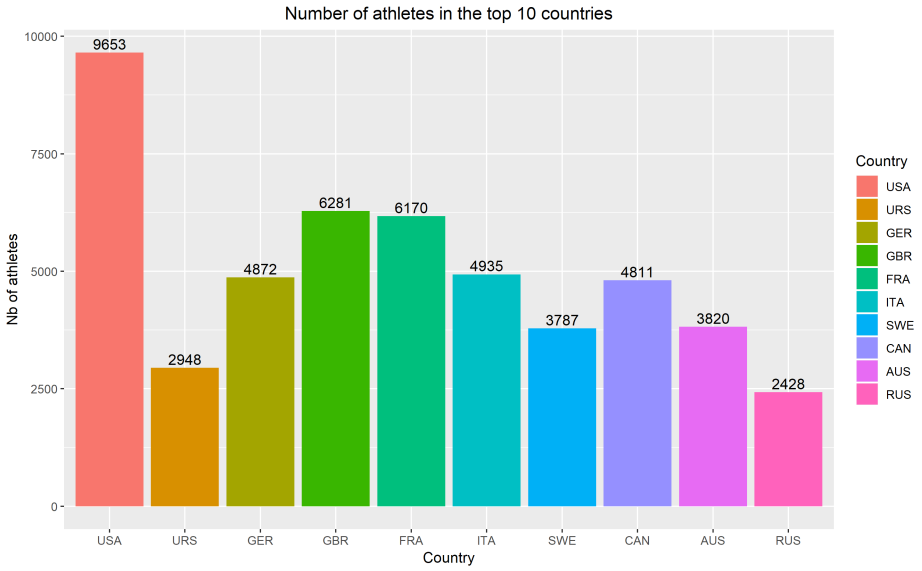

Figure 2: Number of athletes in the top 10 countries

After that, we have tried to see if there is a relation between the age of athletes who won medals during the Olympic Games over all countries and over all years. The figure 3 is an histogram of the number of medals won regarding age groups. The bin width is 4 years, and this histogram represents how many medals an age group obtained. We can see that this histogram represents a normal distribution recognizable with its bell shape. The most of medals are gathered between the groups 18 to 22 years old, 22 to 26 years old and 26 to 30 years old.

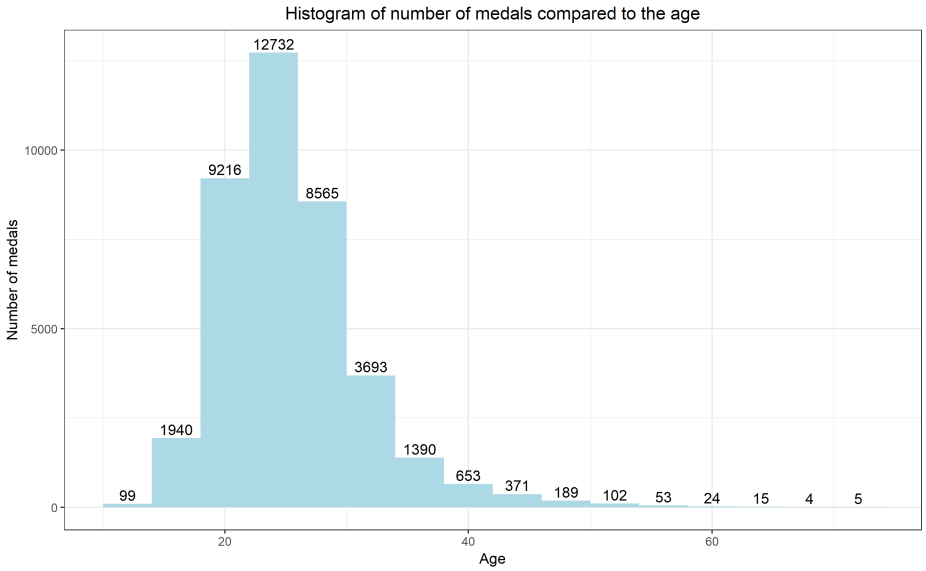

Figure 3: Histogram of number of medals compared to the age

Finally, we wanted to see how the BMI (Body Mass Index) impacts the performance of ahteletes. The BMI is calculated by the following formula :

$$BMI = \frac{Weight}{Height^2} \tag{1}$$

According to WHO (World Health Organization), the usual BMI is between 18.5 and 24.9. The figure 4 represents the number of medals distributed by BMI of athletes and by gender (men in blue and women in red). We can see that, globally, every athletes have the usual BMI. Futhermore, we can observe two groups of BMI : BMI of women, which is between 20 and 22, and BMI of men, between 23 and 25 approximately. In this way, for each gender, every athletes who won a medal have globally the same BMI.

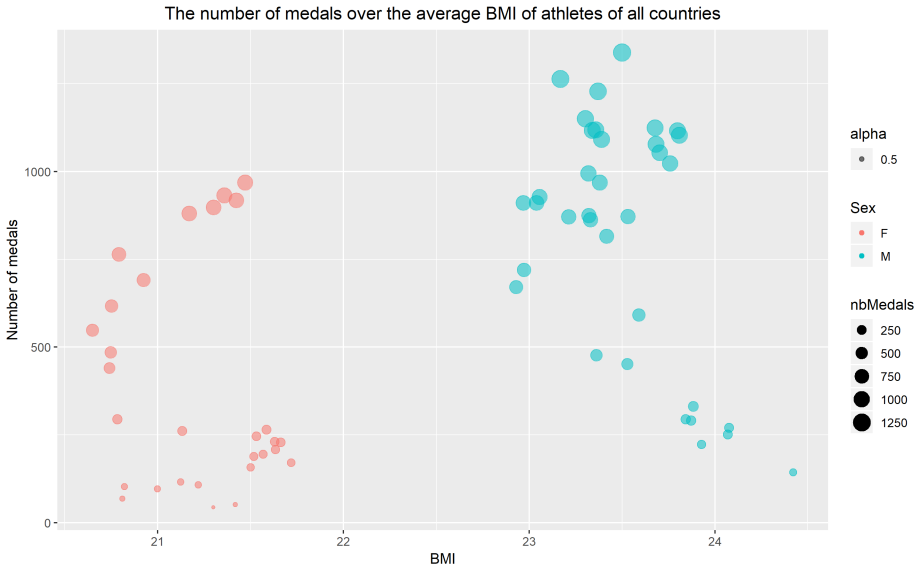

Figure 4: Number of medals over the average BMI of athletes of all countries

3 Conclusion

To conclude, we cannot say that the number of medals obtained by a country is correlated by the number of athletes regarding the top ten countries. Even if the United States has the most athletes and the most medals among these countries, the second country, the Soviet Union, has less athletes implying a more concentration of medals per athlete. However, most of medals are won by athletes aged around 18 to 30 years old, so we can say that this age group has more chance to win a medal. Finally, we distinguish that athletes who won medals have approximately the same BMI, intrinsically with the gender.

References

[1] 120 years of Olympic history: athletes and results. `https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results`.