Beyond Descriptive Statistics

July 2025

Mia Troiano

Correlations

There are not many correlations in this dataset. One possible correlation that I examined was the relationship between age and medal status, thinking maybe older athletes would have more medals. However, the correlation was r = 0.022, showing a very poor correlation.

Another correlation examined was the size of the team and the medal count, assuming that a larger team will have a higher medal count. Broken down by year of the games, the correlation between size of the team and total medals is r = 0.8, which shows a high positive correlation.

Connections

Another relationship to be examined is which countries were most successful in which sports. I looked at the top 3 performers throughout the years in each sport. This led to visual analysis of each individual sport, with a portion of the table displayed below.

	SPORT	COUNTRY	TOTAL	GOLD_	SILVER_I	BRONZE_I
15	Athletics	USA	1080	542	317	221
16	Athletics	Russia	374	124	129	121
17	Athletics	Germany	365	90	132	143
18	Badminton	China	64	28	13	23
19	Badminton	South Korea	35	11	12	12
20	Badminton	Indonesia	28	11	9	8
21	Baseball	Cuba	112	64	48	0
22	Baseball	USA	68	24	0	44
23	Baseball	Japan	64	0	20	44
24	Basketball	USA	341	281	24	36
25	Basketball	Russia	194	60	50	84
26	Basketball	Serbia (Yugoslavia)	84	12	48	24

Figure 1. A snapshot of the table showing the top 3 countries in each event in order by total medals won over the 100+ years of the dataset.

Further analysis was performed to see how many times each country appeared in the top 3 performers for any sport, which would show the best overall performance at the Olympic games. Analysis shows that the United States appeared in the top 3 performers for 43 different sports, with Germany and the UK having the second and third most appearances.

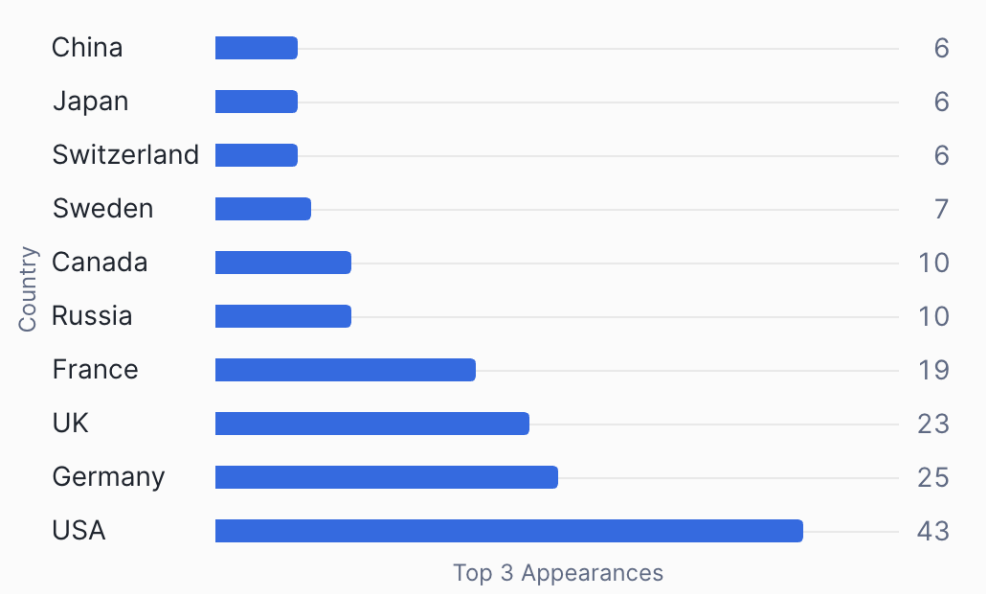

Figure 2. A bar chart showing how often each country appeared in the top 3 performers for each sport over the 100+ years of the Olympics.

Another analysis that could occur is the success rate of each country in the Olympics, which would be the quotient of the total medal count per athlete's representation. However, this analysis would need to consider that a single athlete can win multiple medals, which would need to be accounted for in the final calculation.

New Metric

BMI

Body Mass Index (BMI) was calculated for each athlete, calculated as following formula: weight/(height/100)^2. BMI has been used to categorize population level body composition, with the following categories: underweight <18.5, normal 18.5-24.9, overweight 25.0-29.9, obese 30.0-34.9, extremely obese >35.0. The average BMI for males and females for each year of the games was then calculated to examine the average body composition of the athletes for that year. An assumption would be that the BMI for both

genders decreases over the years as an extremely fit physique becomes increasingly popular in most sports.

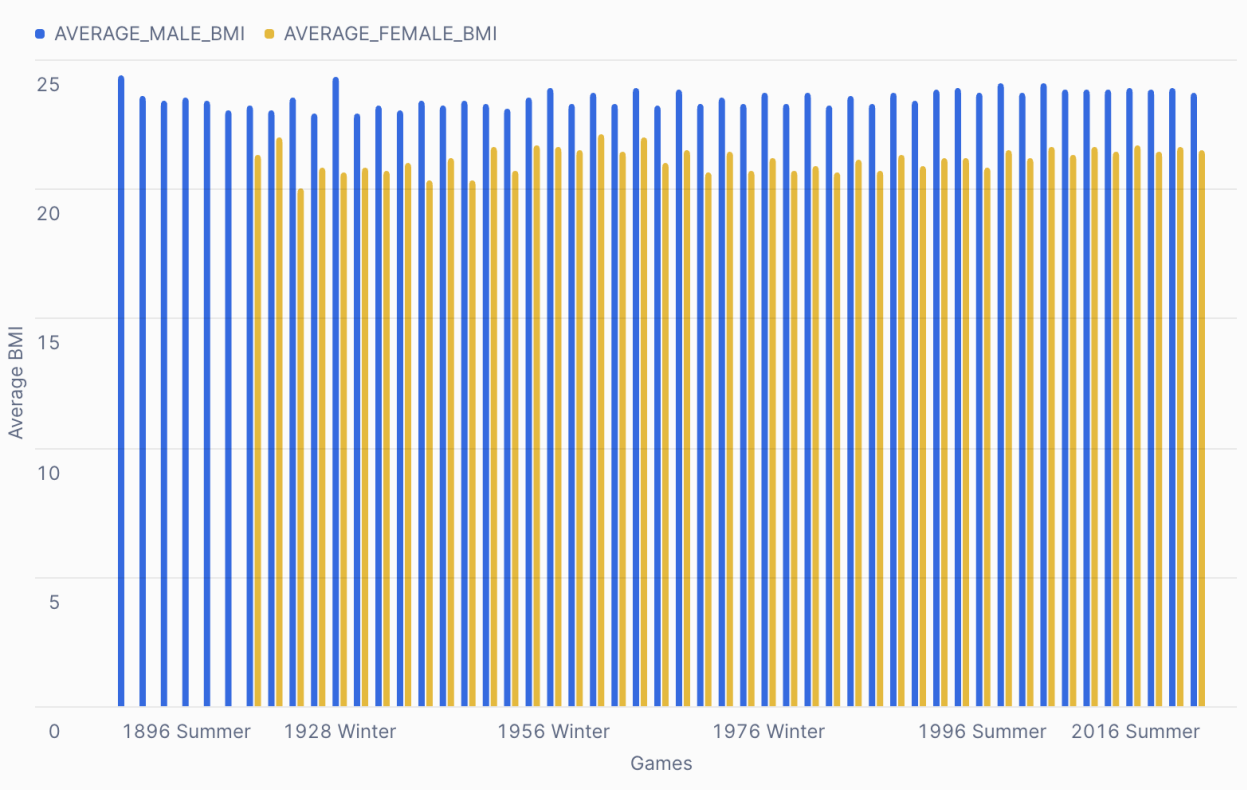

Figure 3. A bar chart showing the average BMI for females and males, broken up by year of the games in ascending order.

Correlation: The correlation between male BMI and year of the games is r = 0.067 and the correlation between female BMI and year of the games is r = 0.059. Both correlations are very minimal, which shows that there is very little correlation between these two variables. This shows that there is little to no relationship between the year of the games and the average BMI of the athletes, which shows that there was no real change in the overall body composition of these elite level athletes over the years. It should be noted that the BMI was an average for all the males or females in the Olympics that year, the BMI could vary greatly by country or by sporting event.