Olympic Data Analysis

The basic idea for this analysis was to look at how countries performed throughout the Olympic summer games by looking at athlete participation and weighted medal score (Gold = 3 pts, Silver = 2pts, Bronze = 1pt). Furthermore, we also wanted to look to see whether a country’s military history has any negative correlation to the studied metrics.

The csv we used listed all the athletes that participated in both the summer and winter games. Initially, we tried looking at both to include more data points but the discrepancy between in terms of athlete number and number of events/medals made it difficult. For consistency, we chose to focus on just summer events and also, we eliminated data points for Olympic events that took place at or before 1904 due to the novelty of the games which created heavy outliers.

To filter the data, we mainly relied on the .loc function to draw from only the intended fields (season, team) and then grouped the data by year. East/West Germany and Russia/Soviet Union/Russia Empire/Unified Team were treated as one. We then tallied the number of unique athletes that participated in that year and plotted. For the medal count, a data frame was created which filtered by medal and uniquely associated events. A loop was then created to iterate through the list and assign a point value based on medal and then produce the total in a separate column. Both of these were then plotted on the same graph along with a linear regression line. Finally, we looked at all major military conflicts associated with each country and overlaid onto the graph.

🡪 The conclusions we found were mostly in line with our hypothesis but not necessarily conclusive. If we go by country, we see that the US had a drop off following WWII and participation remained low throughout the Vietnam war. After the war ended participation increased again with maybe a slight dip during and after the Iraq war.

🡪 Russia’s population was decimated by the Russian Revolution and as a result, there were only 3 Russian participants in 1924. Russia did not participate in the Olympic games again until 1948 where participation remained increasingly high until the advent of the soviet afghan war and ever since it has been on a decline in both medal count and participation.

🡪 Germany’s data is more scattered, but participation peaked in 1936 prior to WWII and then dropped by ~230 in 1948.

🡪 For France and Britain, it is difficult to derive any hard conclusions, but participation dropped for an extended period of time following WWII. Interestingly for both, despite steady rise in participation, medal count has been relatively stagnant.

🡪 Finally, another accidental finding by tracking medal count throughout the Olympics is correlation between hosting the Olympics and medal count. It isn’t shown in the graphs, however, each dramatic spike in both attendance and medal count is almost always associated with being the host country. This isn’t necessarily surprising, but noteworthy, nonetheless.