Project Title: Olympic Athlete Analysis

Peter Sheldon, Thomas Byrne, Daniel Valdes, Jack Bitcon

Data Set: Olympic History

<https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results#athlete_events.csv>

**Introduction:**

Our team’s goal was to analyze a data set regarding the history of past Olympic games, both winter and summer, and examine trends to see if we could get a picture of what athletes and teams stood out from the rest of the pack. We examined things like body composition, gender, nationality, event etc. to determine what makes winning athletes and teams tick.

**Gender:**

Gender was used by our group to simulate predictive strategizing. By using the two graphs that are included and any iterations of the appropriate data sets that feed into these, any given country can first analyze how they stack up to any other given country in terms of total medal count. Then once their potential targets/rivals are determined, they can use the second graph of total gold medals available per sport to look for opportunity areas to dedicate budgetary funds/training facilities in an attempt to break into the upper echelon of medal count. (E.g. Japan could see that there is a lot of opportunity in swimming and dedicate a team to the 4x100 Relay training nonstop in an effort to carve out a niche on the global Olympic stage.

**Country:**

As one may expect, some of the larger and more developed countries (e.g. U.S., U.K, Germany) had a significantly larger medal count relative to some of the smaller counterparts both in terms of average and total medal count. One interesting standout is that the dominance of the former U.S.S.R continued after the breakup and bleed into the states that made up the union, now operating as independent sovereign countries. The original countries that made a mainstay in the first Olympics have historically had. We also included a graphic that shows the average number of athletes for the U.S and China over the years. As one can see, the graphs show spikes in similar time periods which would indicate the competitive nature and rivalry between the two countries, mainly PRC attempting to keep up in sheer numbers with the U.S. One could conclude that the absolute difference between any of the lines could be a result of budgetary allocation, cultural importance, etc.

**Biometrics:**

For fun, we decided to look at some of the biometric information for athletes in the dataset. Weight distribution was significantly more spread out than height distribution, which is likely a similar situation for the global population as a whole. We found that on average the weight is on the lower end of the spectrum for athletes as well which is to be expected.