Part 1: Proposal

1. Research Question: How can we best predict an Olympic track athlete’s finishing time (events including 100M to the marathon)? What is the relationship between an athlete’s finishing time at the Olympics and (at the athlete-level) their sex, medal won, distance of competition, year of competition, and (at the country-level) the athlete’s nationality, their country GDP, and population?

Why this project is interesting? Rationale for each variable included??

1. *Article 1:* Universality, Limits and Predictability of Gold-Medal Performances at the Olympic Games

Need references/links

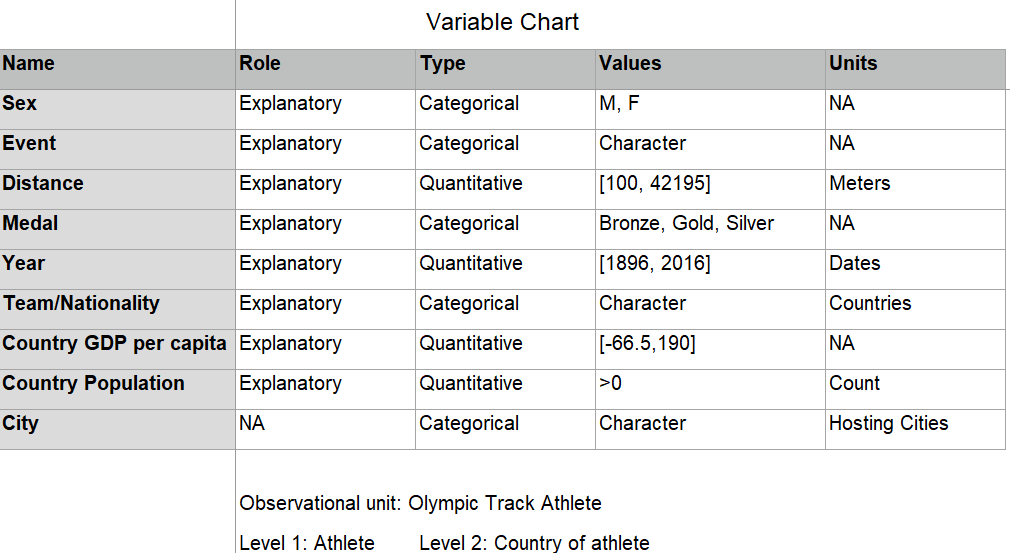
*Article 2:* Predicting Olympic Medal Counts: The Effects of Economic Development on Olympic Performance

Bian, X. 2005. Predicting Olympic Medal Counts: The Effects of Economic Development on Olympic Performance. The Park Place Economist, XIII: 37–44.

Link: <https://pdfs.semanticscholar.org/7293/1ab692bcab9e724b0e5ed4adb53b7ff8097f.pdf>

Article 1 provided a useful primer into all the thought and research that has gone into modeling Olympic performance and medaling. Article focused on all gold medals for all events and attempted to predict the “relative difference between improvements in two different editions of the game.” The authors showed that this statistic, if calculated with respect to an asymptotic performance value (i.e. peak physiological performance time, weight, length, etc. in the competition event, follows a normal distribution. It was an interesting idea to create a meaningful variable to predict Olympic performances. On a side note, the authors also indicated several times that there are some differences in analysis of swimming and all other non-swimming events.

Article 2 was an analysis on country-level characteristics of the Olympics. The author was interested in how socio-economic factors influenced a country’s medal count. He found GDP per capita and population statistically significant in predicting the number of medals won by a country, only including countries in the analysis that won at least one medal. The author said choosing only those countries to include most likely inflated the impact that GDP and population has on medal count. He also discussed addressing the problem of multicollinearity with the variables GDP per capita and population. Also included was a discussion on how GDP per capita and population affect each other claiming “marginal contribution of population growth to the Olympic medal winning process tends to decline as the population size gets bigger”, which gave us the idea to test an interaction with these two variables.



1. How we plan to address question
2. How we found our data

The link for the data of a country’s population and GDP per capita over time:

<https://www.gapminder.org/data/>

The link for Olympic track results from 1896-2016:

<https://www.kaggle.com/jayrav13/olympic-track-field-results>