GT Data Analytics Bootcamp
Group Project 1 Write-Up
October 2018

Project Name: Going for the Gold—Exploring Data on Olympic Medal Wins, 1896-2012
Group Members: Anum Ali Mohammed and Thach Tran

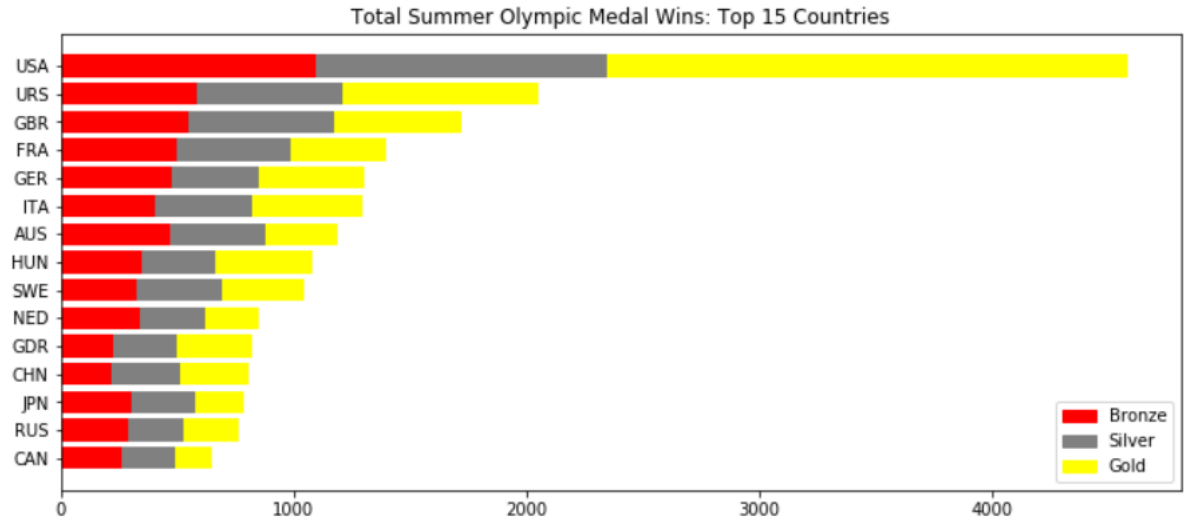
Project Overview:
The goal of our project was to understand trends in Olympic medal wins, in order to predict outcomes for future Olympic games. We considered historical data from the all Summer Olympic games from the 1896 (the first year of the modern Olympics) to the present. Data for all Olympic games except the most recent, Rio 2016, were included in our sample. We focused our analysis on exploring how the following 3 factors affect a country's number of Olympic medal wins: (1) Host Status, (2) Female Participation, and (3) Level of Economic Development.

Key Findings:

Part 1: Effect of Host Status on Medal Wins

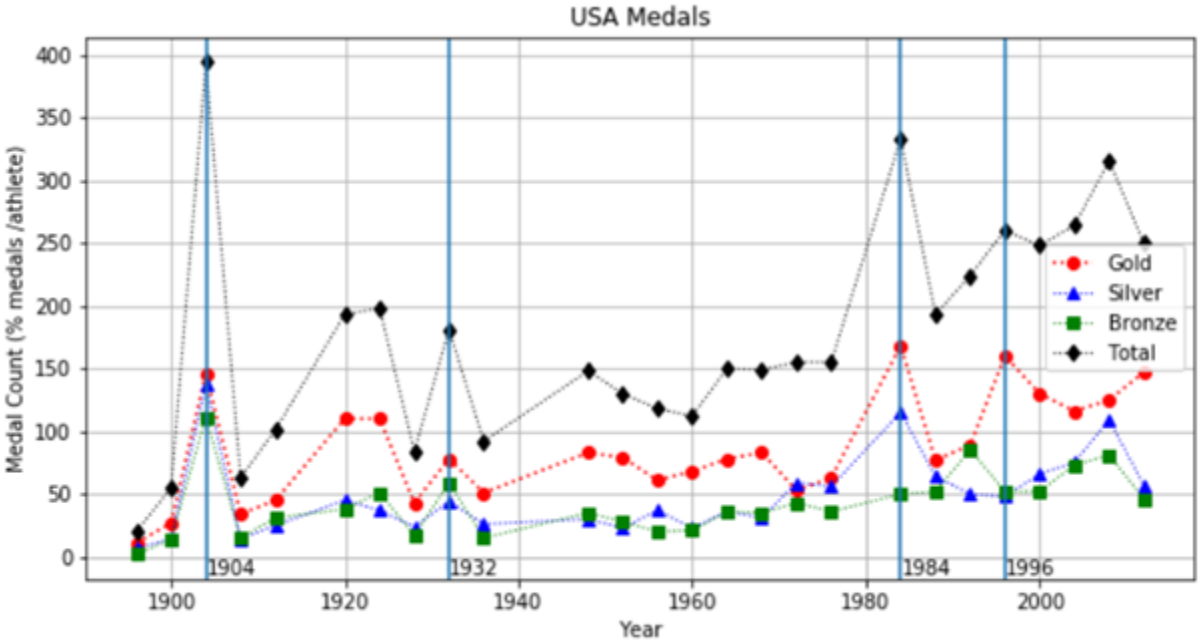
Key Question: Does hosting the Olympics have an effect on number of medals won?

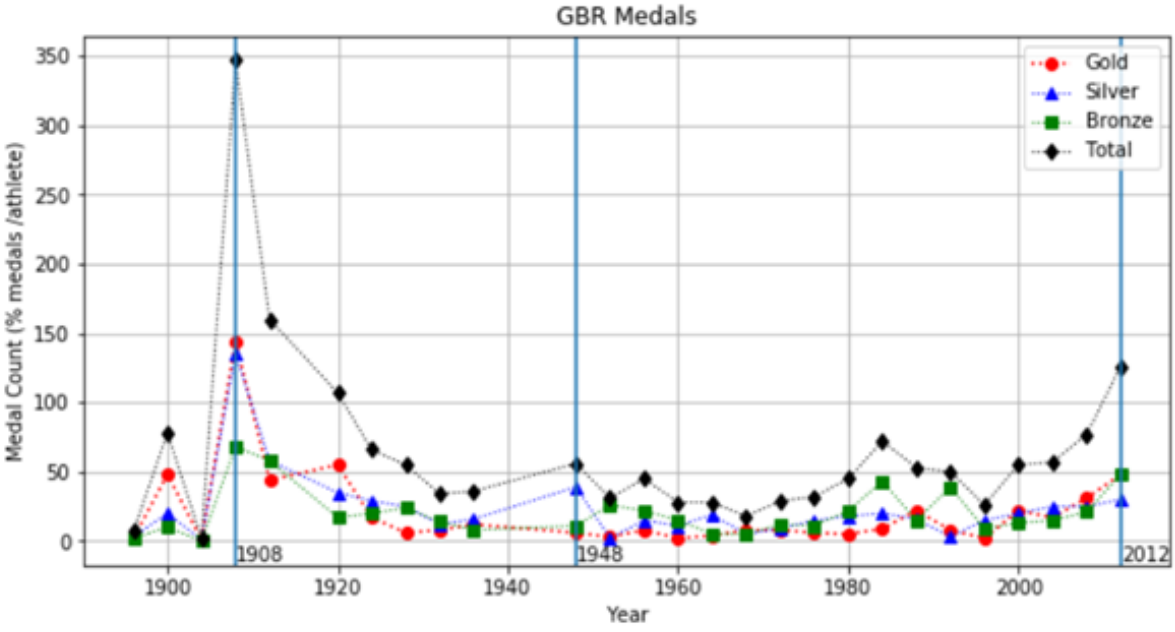
The first factor we considered was "Host Status". Every 4 years, the Olympics are held in a different city around the world. The city where the games take place is referred to as the "host." We decided to explore the effect of hosting the Olympics on the host country's number of Olympic medal wins, because we noticed that of the top 5 medal winning countries of all time (USA, USSR, Great Britain, France, and Germany), 4 had hosted the Olympics multiple times (USA, Great Britain, France). This trend made our group wonder if host countries are somehow more likely to win Olympic medals.

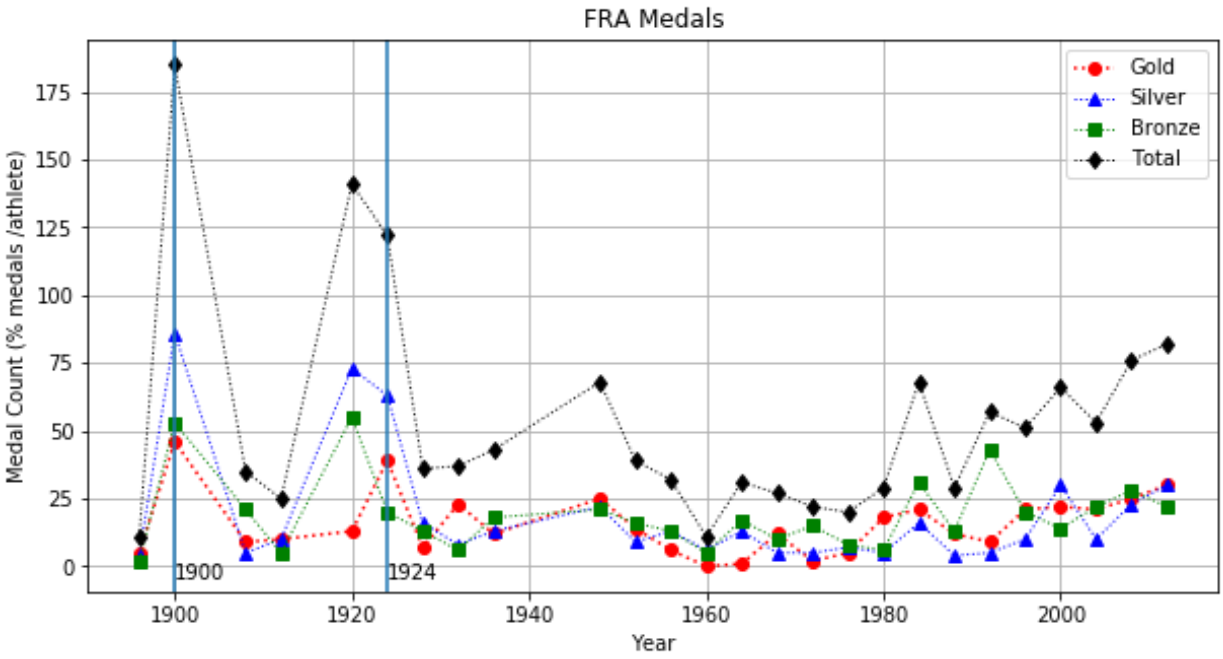


	Times Hosted
USA	4
UK	3
Germany	2
Greece	2
France	2
Mexico	1
Belgium	1
Sweden	1
Netherlands	1
Norway	1
Australia	1
South Korea	1
Australia/Sweden	1
Canada	1
Italy	1
China	1
Russia	1
Spain	1
Japan	1

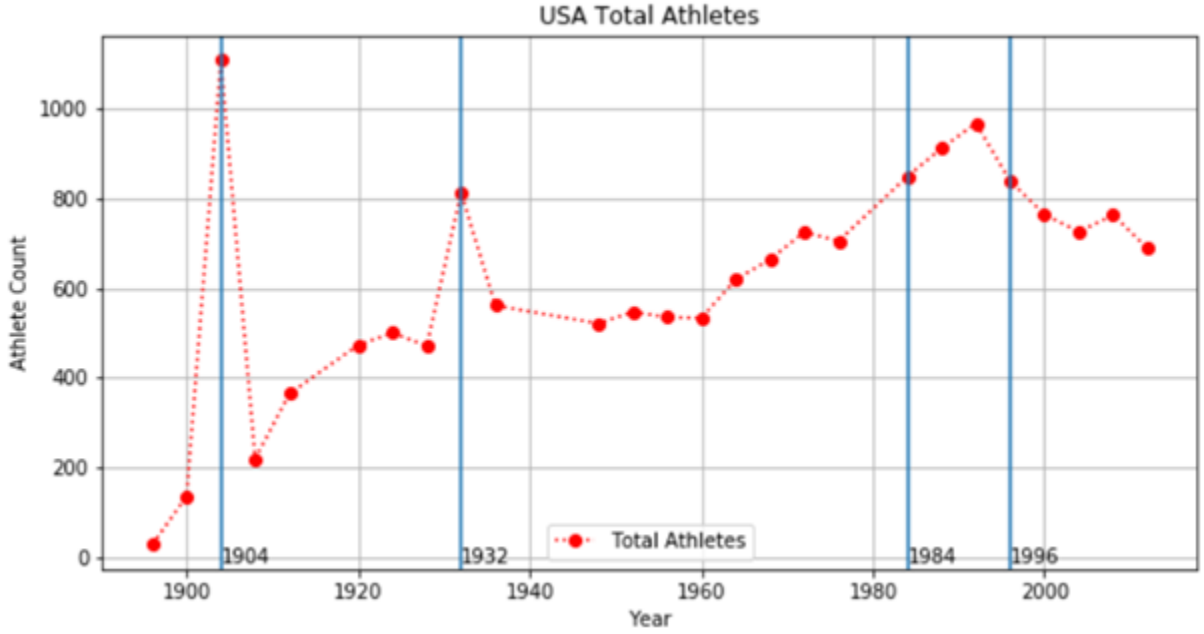
In order to explore the effect of host status on number of medal wins, we created timeseries graphs for number of medals won per year for a subset of 3 countries who had hosted the Olympics multiple times—the USA, which has hosted the Olympics 4 times, Great Britain, which has hosted it 3 times, and France, which has hosted it twice. On each timeseries graph, we drew vertical blue lines for the years the Olympics were hosted in that country. Looking at the graphs, we noticed dramatic spikes during host years, where the number of medal wins increased dramatically during the host year then dropped back down the following year.

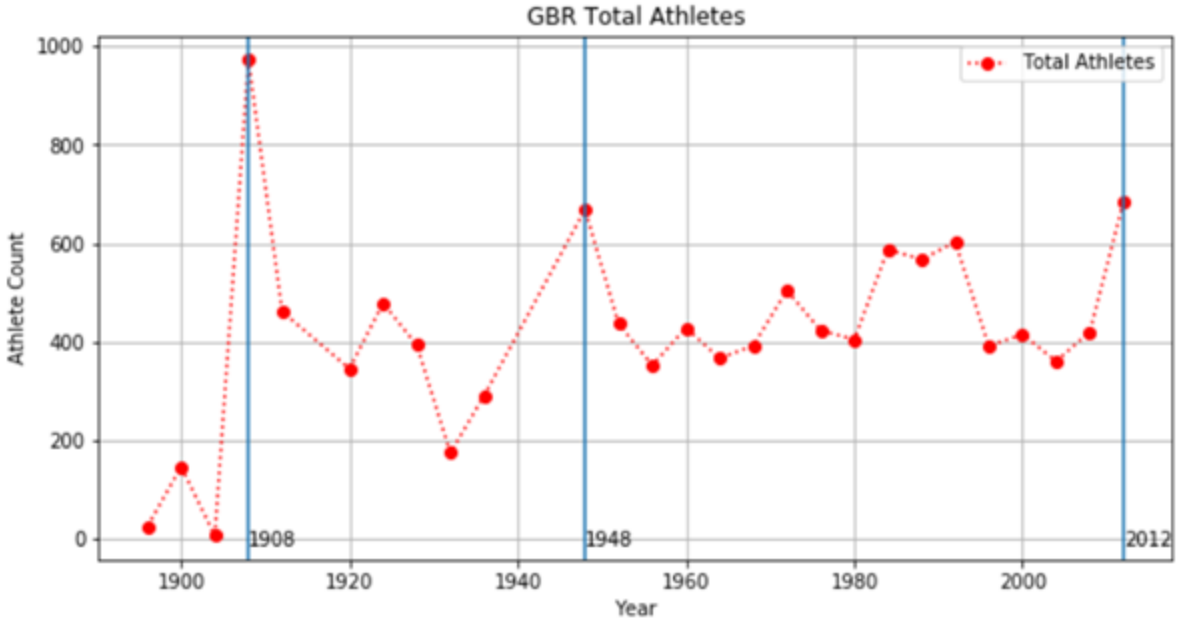

USA Medals
Medal Count (% medals /athlete)
Year
Gold
Silver
Bronze
Total
1904
1932
1984
1996

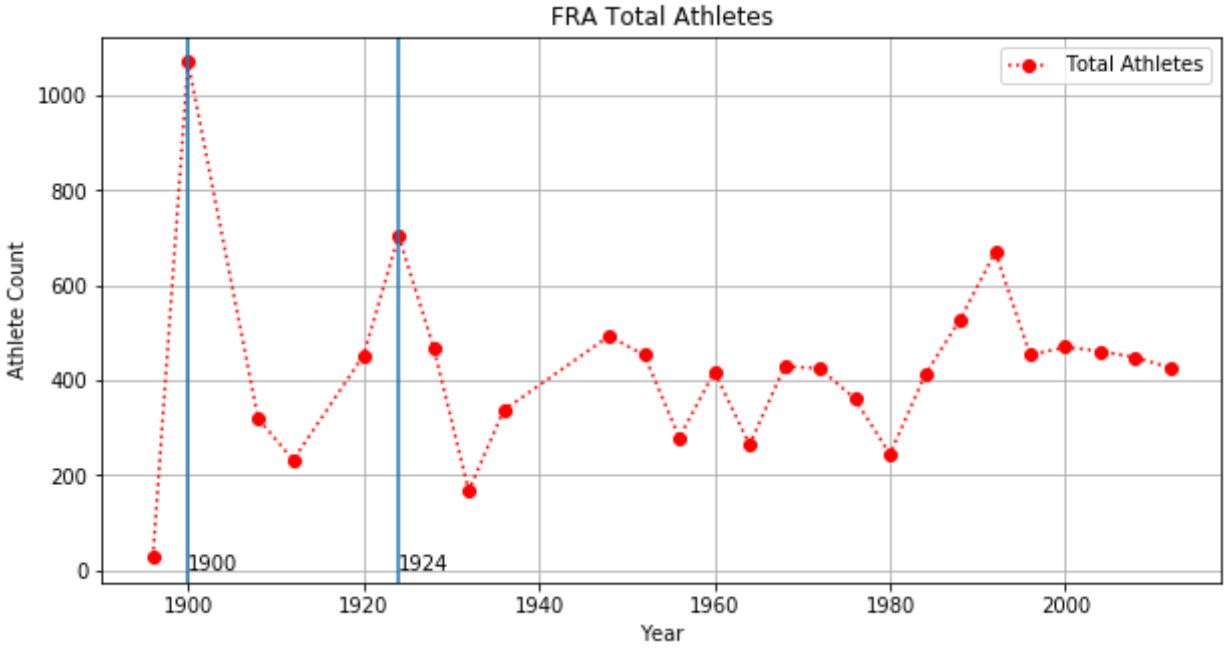

GBR Medals
Medal Count (% medals /athlete)
Year
Gold
Silver
Bronze
Total
1908
1948
2012



Our timeseries graphs made it clear that medal wins did increase significantly during host years. However, we wondered if the host effect was due to random chance, or if some other factor was affecting it that we needed to control for. We recognized that during host years, hosting countries may be more likely to send more athletes to the Olympics due to low barrier to entry (minimal travel costs, etc.), and we decided to try to control for the number of athletes and see if a host effect still existed. Unfortunately, we were unable to complete this portion of our analysis in time for the group presentations on Saturday October 13.



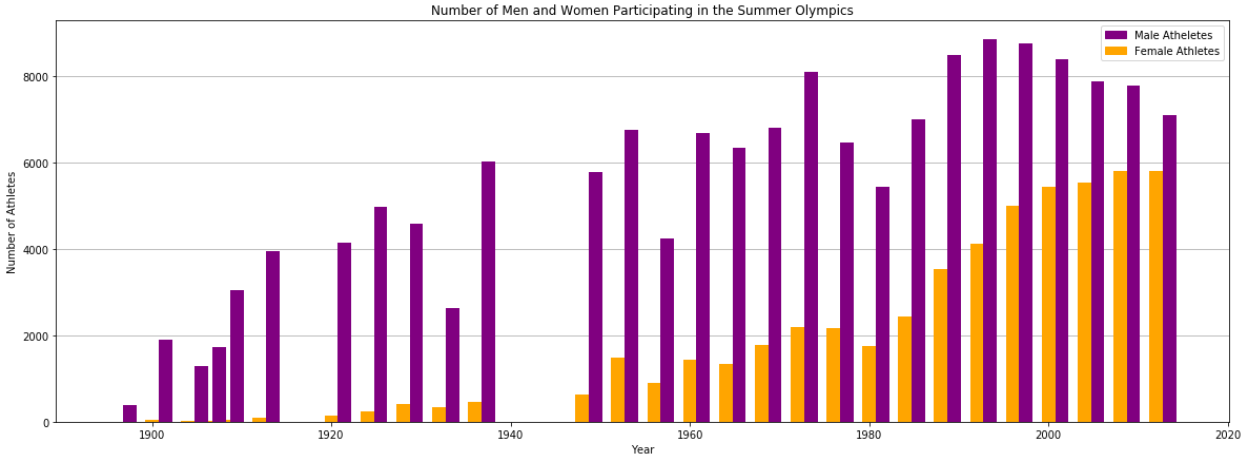

GBR Total Athletes
Total Athletes
Athlete Count
Year
1908
1948
2012
0
200
400
600
800
1000
1900
1920
1940
1960
1980
2000


FRA Total Athletes
Total Athletes
Athlete Count
Year
1900
1924
0
200
400
600
800
1000
1900
1920
1940
1960
1980
2000

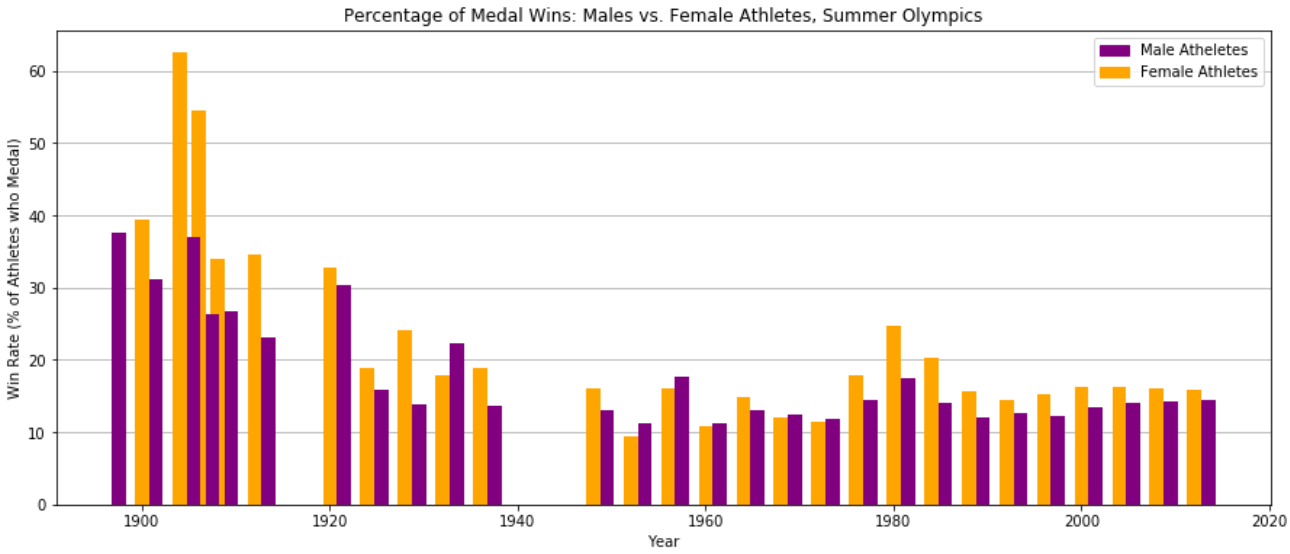
Part 2: Effect of Female Participation on Medal Wins

Key Question: Does the level of female participation in the Olympics have an effect on number of medals won? Does more women = more wins for country?

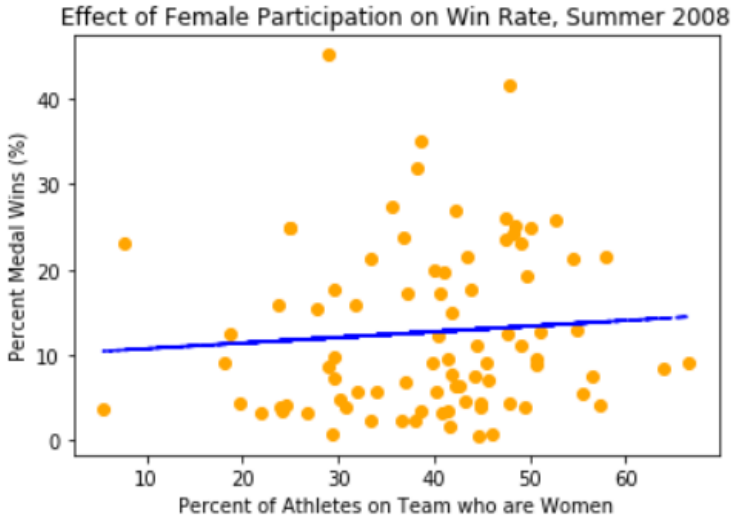
Following host status, the next factor we considered to see if it had an effect on the number of Olympic medal wins was female participation. We define female participation as the number of athletes on a country's Olympic team who are women. We chose to look into how a country's level of female participation is related to medal wins because we recognized that equal sex representation has historically been a major issue for the Olympics. At the first Olympic games in 1896, there were 0 female athletes. Fortunately, this has improved over time. However, even by the summer 2012 games, the Olympics had not achieved true gender parity, and more men than women were competing in the games.



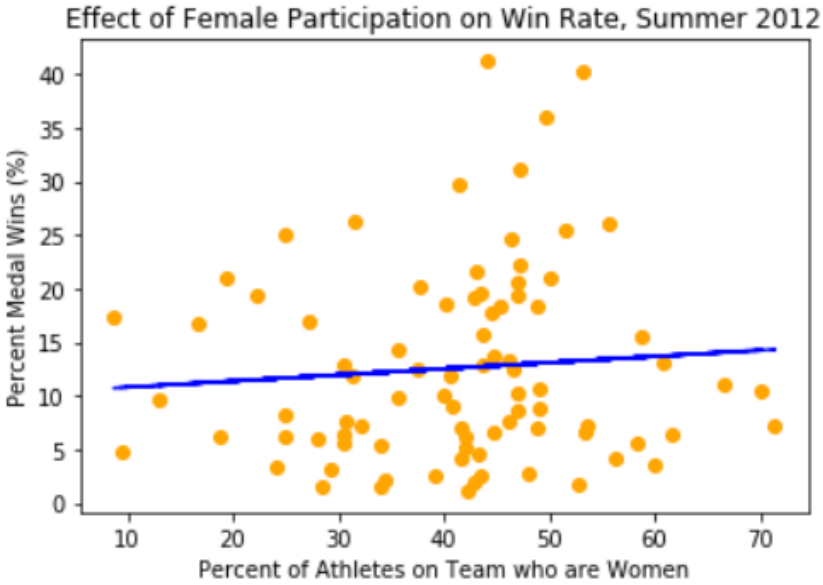
We noticed that although more men compete in the Olympics, women who enter are more likely to win. In other words, women have higher win rates than men and are more likely to earn medals. We define "win rate" as the percentage of events a country enters that are won (any color medal). The higher win rate among female athletes may be attributable to the fact that less women compete in the games, so there is less competition and higher likelihood of winning. Still, the higher win rates among female versus male athletes made us wonder if having more women on the team could bolster a country's overall performance in the Olympics. If women are more likely to women, shouldn't teams that have more women have higher win rates?



In order to analyze the effect of female participation on overall country win rates, we created a series of scatter plots which graphed the win rate for the country overall (percent medal wins for all athletes—male and female) versus the level of female participation (percent of athletes on the team who are women). If female participation has a significant effect on win rate, we would expect to see a strong positive correlation on these graphs. We graphed win rate vs. female participation for the 2008 and 2012 Olympics, the two most recent games for which we had data. Our analysis revealed a slight correlation between female participation and wins, as demonstrated by positive slope values. However, these correlations were not statistically significant, and both graphs had very high p-values (much higher than the level required for statistical significance, $p < 0.05$) Overall, it does not appear that the level of female participation has a significant effect on how many Olympic medals a country wins.



slope = 0.06592339894132858
p-value = 0.47585518875601895

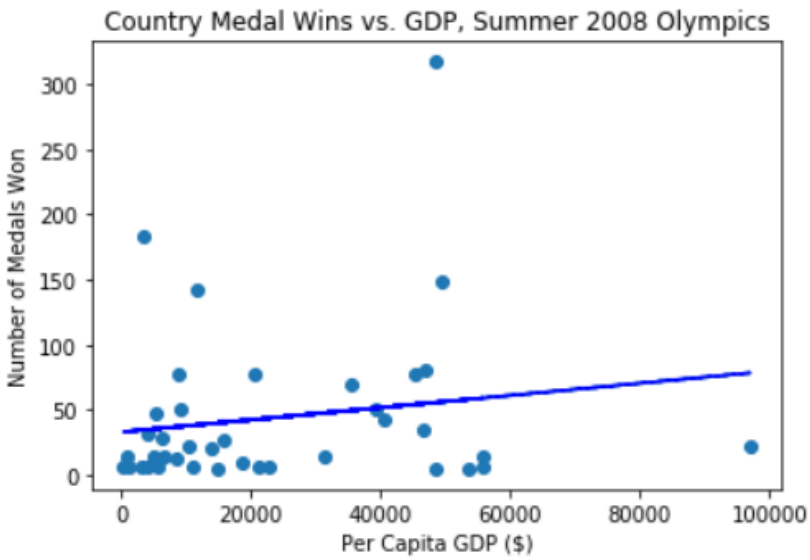


slope = 0.057491945122412895
p-value = 0.45792808675459784

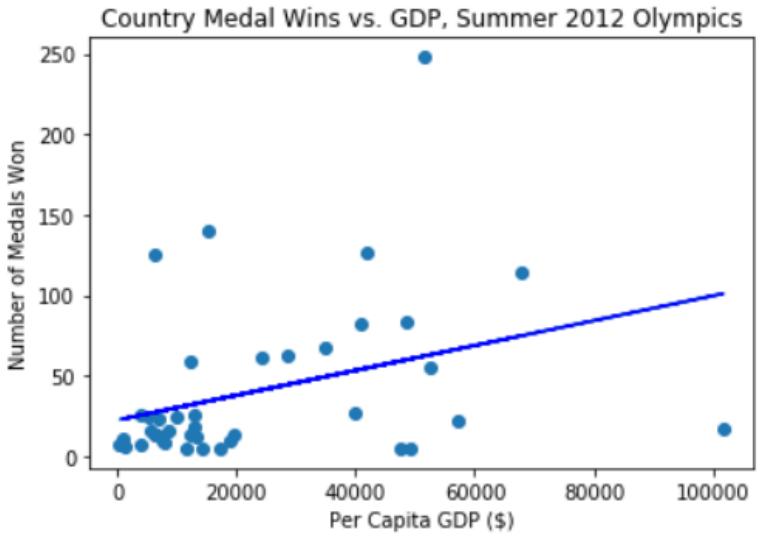
Part 3: Effect of a Country's Level of Economic Development on Medal Wins

Key Question: How does level of economic development affect a country's Olympic medal count? Do wealthier/more developed countries do better at the Olympics? Are low-income countries at a disadvantage?

The final factor our group considered as a possible factor that could affect Olympic medal wins was economic development. We recognized that majority of top medal-winning countries throughout Olympic history have been highly-developed, high income countries such as the US and nations in Western Europe. This made us wonder whether there is a significant correlation between a country's level of economic development and its performance at the Olympics. In order to explore this question, we conducted an analysis of per capita GDP versus medal wins for nations competing int the 2008 and 2012 summer Olympic games. If there was a positive correlation between level of development and Olympic wins, we would expect to see medal counts to increase at as GDP increases. Unfortunately, our results for this section of our data analysis project were inconclusive. In 2012 there was a statistically significant correlation between GDP and number of medals won ($p = 0.03$). However, although the 2008 graph showed a slight positive correlation, it was not statistically significant ($p = 0.31$). Thus, although there seems to be a correlation between increasing GDP and medal counts, it is unclear whether this relationship is statistically significant.



slope = 0.00046700489078327904
p-value = 0.3098971294408652



slope = 0.0007729784004581387
p-value = 0.030207419487391557

Conclusion:

We considered the effect of 3 country-level factors on the number of Olympic medal wins—host status, female participation, and level of economic development. All 3 of these factors are correlated, to some extent, with increased medal wins. However, not all correlations are statistically significant. Further research and analysis is needed to determine which factors are most important in predicting Olympic performance.