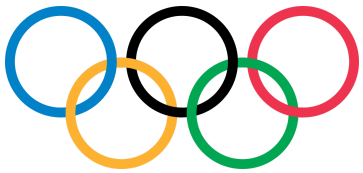
Melissa McGee, Austin Sanders, Beijing Wu
MIDS W200 Summer '21 Section 8
July 21, 2021



Project 2 Proposal: Economic Impact of Olympics on Host Cities

Team Members:

- Melissa McGee
- Austin Sanders
- Beijing Wu

Name of GitHub Repository:

- Project2-McGee-Sanders-Wu
- Link: https://github.com/UC-Berkeley-I-School/Project2_McGee_Sanders_Wu

Overview:

With the Tokyo 2020 Olympics starting soon, we thought it would be interesting to look at various countries' performance at the Olympics and how it may correlate with various economic indicators.

What are economic characteristics (socioeconomic scores, freedom index, GDP, birth rate, death rate, population size) of various countries that correlate with high performance (high medal count) for specific sports?

- How does this change over time?
- What are common characteristics of countries that consistently perform well in all sports categories?
- What are common characteristics of countries that consistently perform poorly in all sports categories?

Primary Dataset:

Our primary dataset is from Kaggle 120 years of Olympic history: athletes and results. It contains 271,116 rows and 15 columns. Each row corresponds to an individual athlete competing in an individual Olympic event. The columns for the athlete that we plan to use: medal, country, event, sex, year.

Supplemental Dataset(s):

We aimed to find datasets that contained data per country per year on one or more of the following economic characteristics that we believe will be interesting for this project:

- GDP: GDP per capita (current $US)
- Population: Population, total | Data (worldbank.org)
- Birth Rate: Death rate, crude (per 1,000 people)
- Death Rate: Mortality rate, under-5 (per 1,000 live births)
- Socioeconomic status scores: Country Socioeconomic Status Scores: 1880-2010 | Population-weighted measures of SES
- Freedom Index: Economic Data and Statistics on World Economy and Economic Freedom (heritage.org) and Human Freedom Index: 2020 | Cato Institute
- Gender equality: Data - OECD

We plan to join these datasets onto the primary dataset on both the Country and Year columns. Data for some of these economic indicators are only available for recent years. When doing analysis, we may focus on a window of recent years for some economic indicators, while other economic indicators with more historical data may have a longer view.

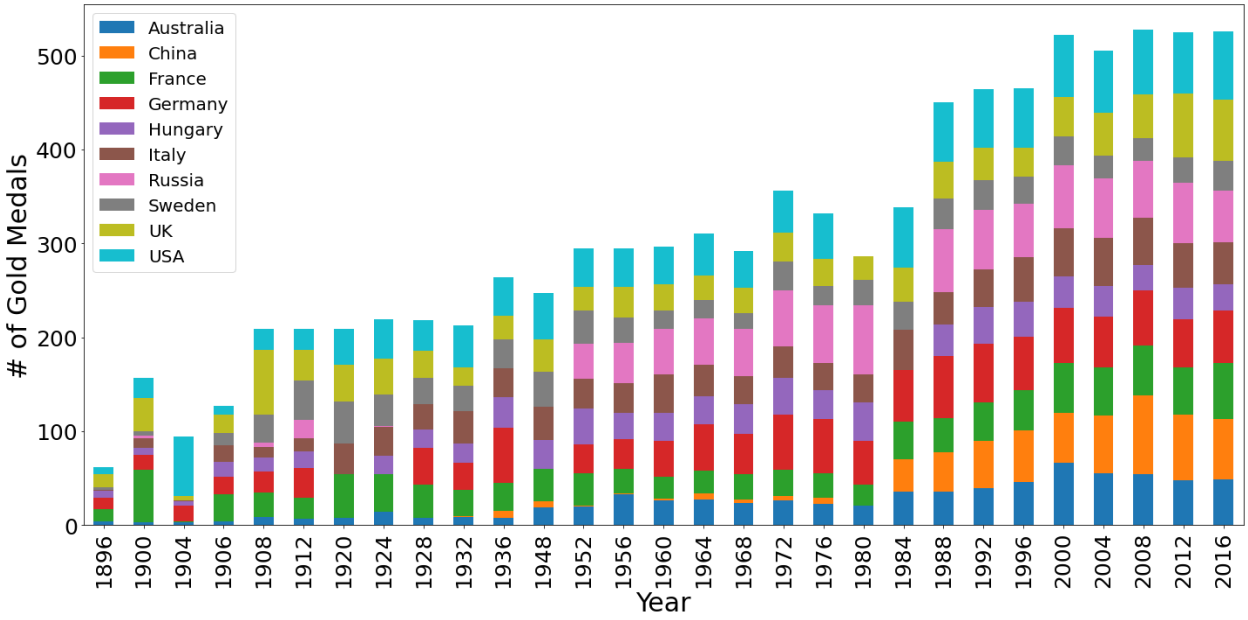
Initial Exploration of Datasets:

First, we wanted to see the countries with the most gold medals over all games. We merged "`athlete_events.csv`" and "`noc_regions.csv`" on column "`NOC`". We substituted `NaN` with "`No Medal`" for column "`Medal`". The top ten countries that had the most gold medals in summer Olympics are as follows:

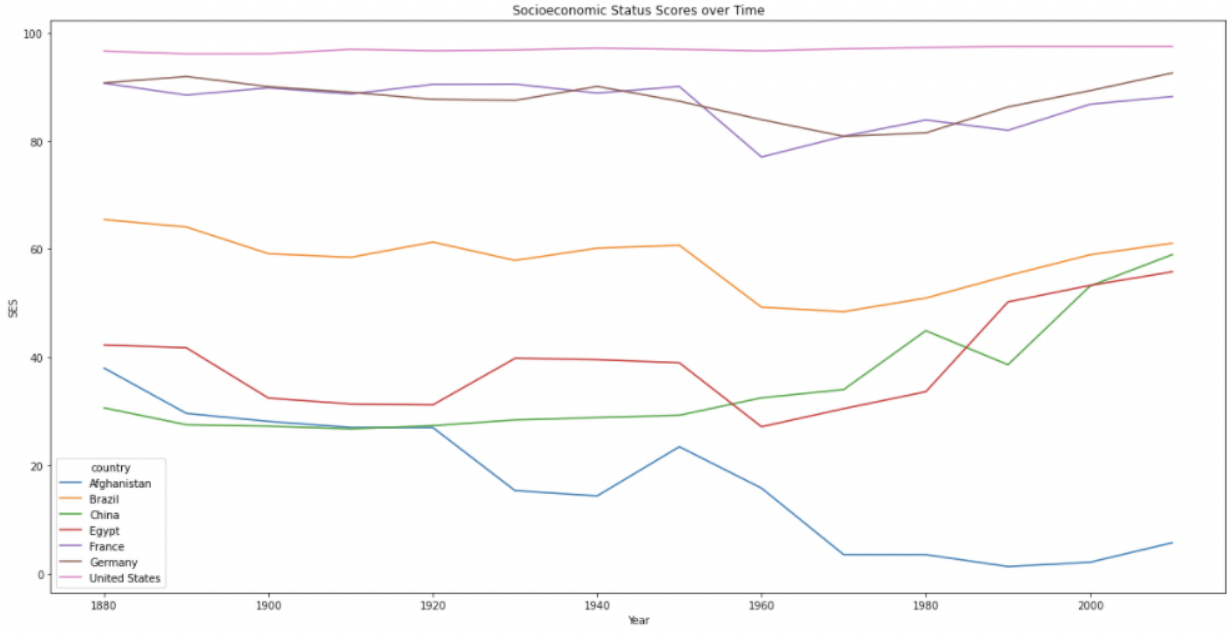
Country	Medal	size
USA	Gold	2472
Russia	Gold	1220
Germany	Gold	1075
UK	Gold	636
Italy	Gold	518
France	Gold	465
Hungary	Gold	432
Australia	Gold	362
Sweden	Gold	354
China	Gold	335

We then wanted to see how medal share among countries may change over time. Over the years, the total number of gold medals won by these top 10 countries is increasing. As shown in the plot below, the US team did not win any gold medal in the 1980 Olympics. In fact, the US team did not compete in the 1980 Olympics held in Moscow as a protest against the late 1979

Soviet invasion of Afghanistan. China has noticeably won a lot more gold medals since the 1980s, and all 10 countries have been consistently doing well since 1988. An interesting question that this plot brings to mind is, *"what are the common characteristics of these countries?"*



We additionally took a preliminary look at one of the economic datasets. The plot below shows the socioeconomic status scores over time for a select group of countries: Afghanistan, Brazil, China, Egypt, France, Germany, and the United States. We see that while the United States has enjoyed relatively high, and relatively stable SES scores over time, Afghanistan's scores have been decreasing and Brazil/Egypt's have been a little tumultuous. We look forward to seeing how this might correlate with performance in the Olympics during different years.



Final Report Overview:

We want to analyze the way economic characteristics impact a country's performance at the Olympics. We will create multiple graphs to display any significant relationships between a country's success and their socioeconomic dynamics.

First, we'll present a high-level overview of Olympics data - the average number of countries that participate, what the top five performing countries are, what the bottom five performing countries are, etc.

We'll then present a high-level overview of the economic indicators of interest, providing various charts comparing different countries' data over time for a select number of indicators.

Then, we'll merge the two, presenting findings on how different economic indicators may be correlated with performance at the Olympics. For example, we will look at the performance of female athletes compared to male athletes for countries with a high degree of gender equality versus a low degree of gender equality; additionally, we will look into how this might change over time as gender equality for a country (or countries) gets worse or better over time.

Of course, while we will not be able to definitely attribute causation with these investigations, we hope to find interesting correlations that we will then summarize in the final parts of the report.

References / Inspiration:

- Predicting Olympic Medal Counts: the Effects of Economic Development on Olympic Performance
- Visual Analysis of Olympics Data: How do different factors come into play into winning Olympics medals?