Chess Analytics: The Relationship Between Rating Discrepancy and Various Metrics

By Jim Yang, Ke Miao, and Michael Nguy

Motivation

This project was inspired by insights from top chess players who have highlighted underrated chess talents emerging from developing countries with solid chess cultures, such as China and India. These players often demonstrate superior performance in tournaments compared to their official chess ratings. Thus, we were driven to adopt a data-centric approach to explore the merits of these insights.

Historically, researchers have shown interest in factors influencing chess Olympiad performance and tournament participation. For instance, Forrest et al. (2020) found a correlation between a country's education capital and its ability to produce strong players, given the understanding that education capital is closely linked to economic strength. Visser (2016) highlighted the impact of monetary incentives on chess player performance and performance disparities in tournaments.

Additionally, Bodek (2019) observed that regional barriers could limit chess player Elo ratings. These previous findings prompted us to explore our research question from an economic perspective, focusing on regional aspects. Building upon Celik's work (2014), which demonstrated a positive relationship between GDP and performance in classic winter or summer Olympics, and recognizing chess as deeply rooted in chess culture, we decided to investigate chess tournament results to ascertain if they could be explained using economic data, specifically GDP, by country, and metrics related to chess culture.

Background

The International Chess Federation, referred to by its French acronym FIDE (Fédération Internationale des Échecs), acts as the governing body of international chess competitions. FIDE hands out chess titles, such as the Grandmaster title, as well as tracks player's ratings (chess Elo). Additionally, FIDE organizes the Chess Olympiad, a biennial chess tournament where teams representing nations of the world gather to compete. Each team consists of five players: Four starters and one substitute. The teams play against different opposing federations for 11 rounds, and the team that wins the most rounds is declared the winner.

Objectives

The purpose of the project was to 1.) combine different datasets with FIDE Olympiad chess data from 2012 through 2022 to extract a useful byproduct, 2.) create insightful visualizations that shed light on the global chess landscape, and 3.) answer our main question:

What factors influence a country's rating discrepancy in chess?

We will try to answer our main question by addressing several related questions, such as:

- Is there a correlation between a country's economic situation and being underrated in chess?
- Which countries have a strong chess culture?
- Can chess culture explain how teams perform in the Chess Olympiad?
- Which countries have consistently over-performed or underperformed at the Chess Olympiads?

We aim to offer a comprehensive analysis that sheds light on the current state of the chess landscape. Furthermore, our goal is to uncover valuable insights into how players are currently rated, specifically focusing on fairness. We also aim to investigate whether socioeconomic factors, such as poverty, may be a contributing factor to players being unjustly underrated.

Chess Elo is a rating system used to measure a player's skill level in chess. It assigns a numerical rating to each player, with higher ratings indicating stronger players. When two players compete, the change in their Elo ratings after the game depends on the outcome and the rating difference between them. We will primarily look at the changes in Elo of a country's Olympiad team after playing in the Olympiad and explore this change's relationship with several other factors.

Data Sources

Primary Dataset

FIDE Chess Olympiad Results

Description: This dataset comprises the outcomes of all games participated in by a player throughout the tournament. It includes essential columns like player name, federation, the number of games played, pre-tournament rating, and the calculated performance rating at the end of the tournament. The players are grouped by their respective federations.
Size: 1.99 MB in total, distributed across six separate XLSX files, each corresponding to an Olympiad that took place between 2010 and 2022 (with the exception of 2020, which was canceled due to the COVID-19 pandemic).
No. of attributes: 23
Time Coverage: 2010-2022, with the exception of 2020 due to the Covid-19 pandemic.
Format: XLXS
Access: Download (also available via script)
Location: chess-results.com

Secondary Dataset

World Development Indicators

Description: Contains various economic indicators for each country, such as income distribution, income per capita, GDP per capita, GNI per capita, and population statistics, among others.
Size: 1.19 MB
No. of attributes: 24
Time Coverage: 2007-2018
Format: csv
Access: Download
Location: databank.wordbank.org

Capital Coordinates Data

Description: It contains country name, capital city, latitude, longitude, population, capital and type for countries. Will be used to help determine chess culture.
Size: 12.42 KB
No. of attributes: 6
Time Coverage: Last updated 2020
Format: csv
Access: Download (via script)
Location: https://gist.github.com/ofou/df09a6834a8421b4f376c875194915c9

FIDE Rated Players Data

Description: Has information on all FIDE rated players. It includes player ID, name, federation, rating, and any chess titles
Size: 435 MB, distributed across 14 separate files.
No. of attributes: 7
Time Coverage: 2010-2022
Format: TXT
Access: Download (via script)
Location: ratings.fide.com

Methodology

Data Manipulation

Introduction to Data Manipulation

The primary objective of our data manipulation process was to prepare our datasets to ensure that they contained relevant and usable information. This preparation was crucial in facilitating the merging of datasets for subsequent analysis. One key aspect of this preparation involved ensuring that each dataset included a column storing the federation code for each row. This federation code served as a vital identifier that would enable the seamless merging of the datasets in the later stages of our project.

Here, we summarize the most significant steps taken during our data manipulation process. For a more detailed look at the code and specific actions, please refer to our project's "clean_manipulate" Jupyter notebook.

General Data Cleansing and Handling Missing Data

- We initiated the data manipulation process with standard data cleansing techniques. Our primary goal was identifying and addressing uninterpretable or 'garbage' data values.
- When uninterpretable values were encountered, we replaced them with ‘nan,’ signifying missing or uninformative data.
- Subsequently, we removed rows or columns containing these ‘nan' values from our datasets.
- Given the diverse nature of our data across various countries and the sensitivity to the years in which data was collected, we refrained from using imputation methods, as these might not be appropriate for our dataset.
- Some aspects of missing data handling were deferred to the analysis stage. For instance, during analysis, we either ignored countries with ‘nan' values or aggregated values only for years where data was not ‘nan' for each respective country.

Chess Olympiad Dataset

- Dealt with six separate Olympiad datasets, each covering a different year between 2010 and 2022.
- Each dataset had its own structure, including variations in columns and header rows.
- Used 'pd.read_excel()' to read each Excel file into pandas DataFrames.
- Specified column indices to extract the desired information from each dataset.
- Employed a method to distinguish player data from header rows where we filtered rows based on whether the value in the first column was numeric (player board number) or non-numeric.
- Ensured that the same columns were selected for each of the six Olympiad DataFrames.
- Combined the cleaned datasets using 'pd.concat()' to create a single, comprehensive dataset.
- Added a 'year' column before merging to indicate the original year of each player's data row.

Methodology

Data Manipulation Continued

FIDE Player Ratings Dataset

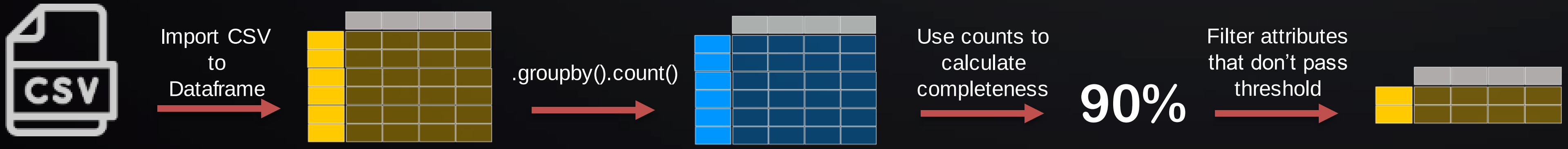
This dataset is presented in fixed width file(fwf) format. Standard fwf files should provide a width schema file indicating column widths. However in our case such schema file is not provided. Therefore for us, the key to decoding this fwf file was determining the columns and column width of each file.

To determine column names, we looked at the first line of each file. We ran into several constraints: we could not split on spaces because column name 'ID number' contains space; we could not split on uppercase characters because 'ID **N**umber' was a column name and 'jan13' was also a column name. Ultimately, for the first row in each file, we had to iterate through each character to find the indexes of capital letters and the letter 'j' (from 'jan13') as the beginning index for a column, eliminating extra indices between position zero and 10 (the first column is 'ID **n**umber' but sometimes the 'n' in 'ID **N**umber' is capitalized).

Next, we calculated the column widths. We did this based on the values of the differences between adjacent column names. Now that we had our widths, we used the pd.read_fwf(file_path), passing in our calculated column widths to import the data into a dataframe. Afterwards, we selected the relevant columns, as well as added a year column, then merged the players df for each year into one big FIDE rated players dataframe. Finally, we handled some edge cases for certain federations and standardized the 'title' column, as some chess titles were stored differently in some files.

World Bank Dataset

The dataset is loaded from a CSV file, and a list of specific years (from 2007 to 2018) was created, representing the years of economic data to be used. Next, missing data, represented as ".." in the dataset, was replaced with NaN values using the .replace() method. We calculated the percentage of countries with available data for each economic indicator using .groupby().count() to count how many countries have data for this column, then divided by the total number of countries. A threshold (set at 90%) was applied to filter out economic indicators that didn't meet the specified data completeness criteria of being included in 90% or more of the countries. The list of economic indicators that meet the threshold is stored in a list, and rows in the dataset are filtered using the .isin() method to keep only those rows corresponding to the selected economic indicators that meet our threshold.

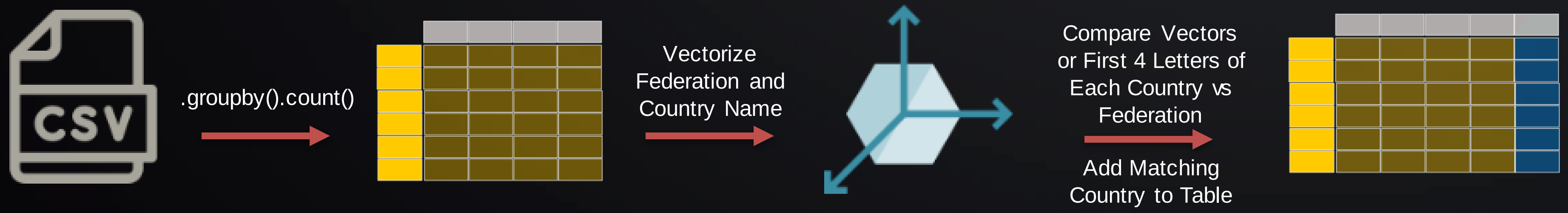


Methodology

Data Manipulation Continued

Capital Coordinates

This dataset had no primary key in the form of country code, so we had to match it with a country codes table found on Wikipedia (ISO 3166-1 alpha-3). The country codes table had four columns, two of which were ISO and IOC codes (which FIDE also uses), the remaining columns were FIFA code and federation name. Our goal was to match the country name in our capitals dataframe with the federation name in the country codes table. in short, the function was designed to match a list of query strings (in this case, federation names) with a list of potential matching answers (country names). The function employs the TF-IDF (Term Frequency-Inverse Document Frequency) vectorization technique to calculate the similarity between queries and potential answers. In the first stage, we use cosine similarity to return a list of best matches between queries and answers, along with a list of queries that didn't find a match. In the second stage, for countries/federations that are not similar enough per the judgement of TF-IDF, we try to match the first four letters of each unpaired federation with country names until a match is found. We probably could have hard coded since there weren't that many countries. Finally, we hardcoded three countries that wasn't able to find a match through the previous steps.



Analysis

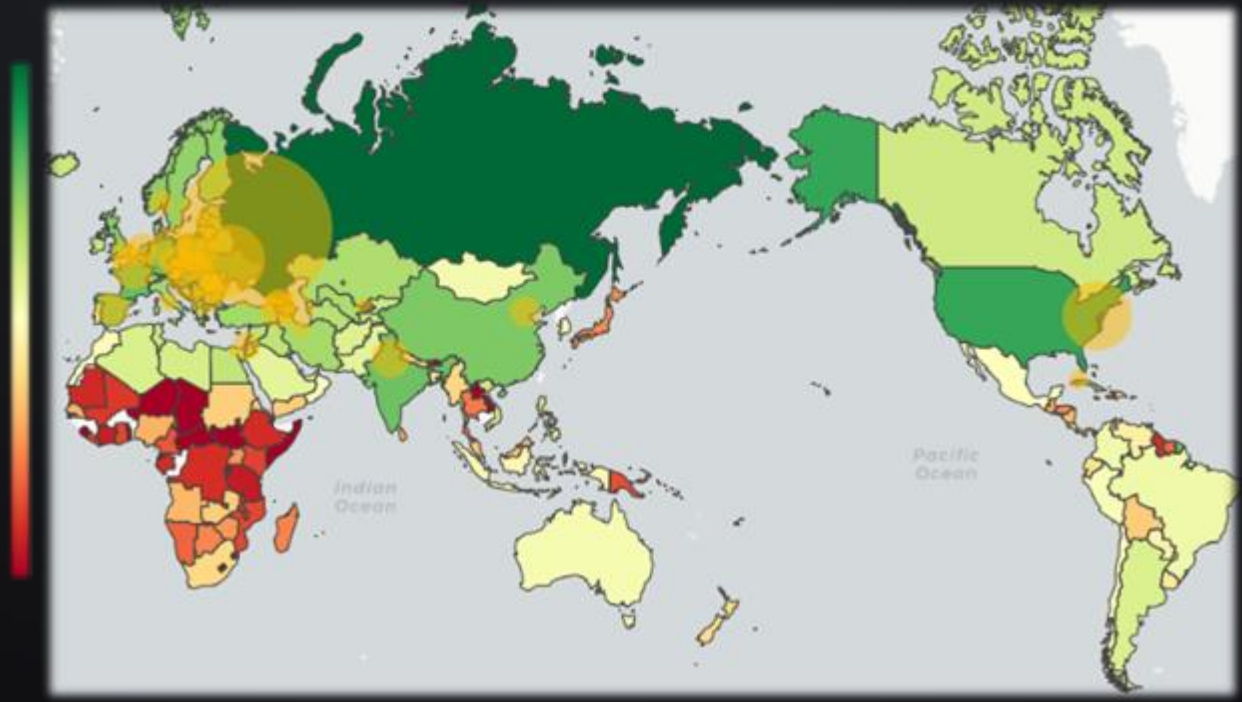
Analysis – Chess Culture Index

We further manipulated the FIDE player ratings dataset to calculate what we termed a country's "chess culture." We considered these five metrics to calculate chess culture:

1. **Chess Masters Density:** Offers insights into the popularity but may be influenced by population.
   - We get the count of those holding chess titles of FM, IM, GM for each country using .groupby(federation, title). We applied weights to each title, then divided by population to get a 'master density' metric.
2. **Chess Masters Count:** Provides a quantitative measure of dedicated players.
   - We get the overall count of those who hold chess titles for each country. We multiplied this with the result of step 1 for each country to get a country's "*base culture*," balancing quantity and density.
3. **Total Number of Top Players:** Indicates the country's ability to produce strong players.
   - We get the top players by filtering by rating > 2700. Then we use .groupby().first() on federation and player_id.
4. **Total Years of Having Top Players:** Reflects dominance and consistency.
   - We again get the top players by filtering by rating > 2700. Then we use .groupby().count(); each year as a top player boosts this metric.
5. **Proximity to Chess Culture Hubs:** Countries with prominent chess figures impact neighboring country's chess scene.
   - Using haversine values, we see if a country is within reach of any chess hub. For chess hubs within reach, we do matrix multiplication of the hub's base culture by a weighted distance value relative to the reach of the hub, then add the additional influence to the *base culture* to get a final chess culture value. If a federation is not within reach of any hubs, the *base culture* is the country's chess culture.

We visualized chess culture using a choropleth map (figure 1). The color range shows a country's chess culture, while a circle over a country's capital is that country's range of influence on other countries. Using our culture metrics, we found that 1.) Russia had by far the richest culture (due to having a long history of top chess players, 2.) Countries that had world chess champions (India, China, USA) are rich in chess culture. Norway also had a strong culture (due to former world champion Magnus Carlsen's dominance as the world's best player).

Figure 1. Choropleth Map, colored by chess culture



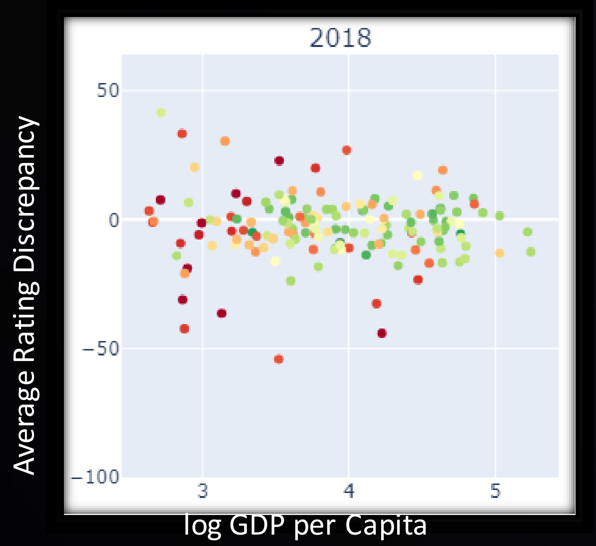
Analysis

Chess Rating Change vs Normalized GDP

Using the FIDE Olympiad dataset, we calculated each country's average rating change (*or rating discrepancy*) after playing in the tournaments (Figure 2). With this discrepancy data, the chess culture data calculated previously, and the World Bank dataset, we created several scatterplots to explore any relationships or trends in our datasets.

- We found that more federations lost rating than gained rating during Olympiads. This held true for all 6 Olympiads we looked at.
- Visually, there is **no apparent correlation** between log GDP per capita and rating discrepancy (Figure 3).
- In **2022**, countries generally lost more rating than usual. A possible explanation is that the COVID-19 pandemic worsened rating misalignment, obscured talent progress, and led to countries '**underperforming**' in an underrated field (Figure 4).

Figure 2. Choropleth Map, colored by average rating discrepancy across all Olympiads



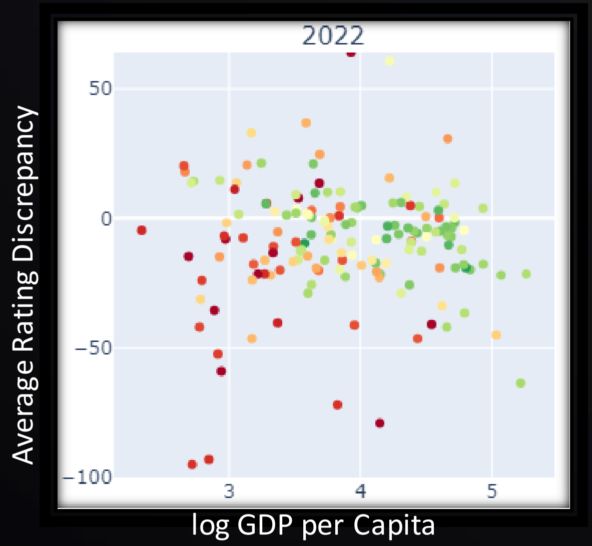


Figure 3 and 4. Scatterplot of rating discrepancy during Chess Olympiad tournament play vs. normalized GDP per capita, with chess culture encoded as the color range. (2018, 2022)

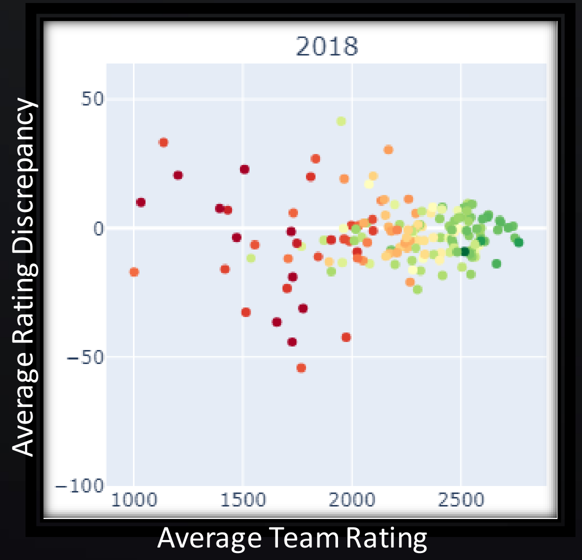


Figure 5. Average Rating Discrepancy vs. Average Team Rating with culture as color.

- Plotting a scatterplot of average rating discrepancy of a country vs. their team rating, and encoding chess culture as a color range, we saw that high culture countries have a higher average team rating and lower average rating deviation.

Analysis

Correlation Heatmap, Line plots of Performance Discrepancy by GDP per capita group

We explored the correlation between different factors and performance_discrepancy

- Log GDP per capita calculated to cancel out the massive population in the denominator when calculating GDP per capita.
- **Observation:** Among the three different culture factors: master_rate (log of the master score of the country scaled by country population), master_metric(master score multiplied with the log of the master score of the country scaled by country population), chess_culture_index (distance to chess hub based standardized chess culture index), chess_culture_index shows the relatively strongest correlation with performance_discrepancy
- **Observation:** GDP per capita shows a relatively stronger negative correlation with performance_discrepancy but not too strong
- **Observation:** Interaction between chess_culture_index and log_gdp_per_capita seems to have some correlation with performance_discrepancy

We explored individual, mean, and standard deviation of performance discrepancies over the 6 Olympiads for different gdp_per_capita groups.

- **Case 1:** top 25 and bottom 25 gdp_per_capita groups: The top group mean is higher than the bottom group mean, and the standard deviation is similar among the two groups.
- **Case 2:** top 20 and bottom 80 gdp_per_capita groups: mean similar among two groups, standard deviation similar among two groups.
- **Case 3:** top 80 and bottom 80 gdp_per_capita groups: mean similar among two groups, standard deviation similar among two groups.
- **Observation:** As group size increases, the top and bottom group means become similar.
- **Observation:** 2012 and 2014 seems to have more discrepancies in both top and bottom groups of different sizes

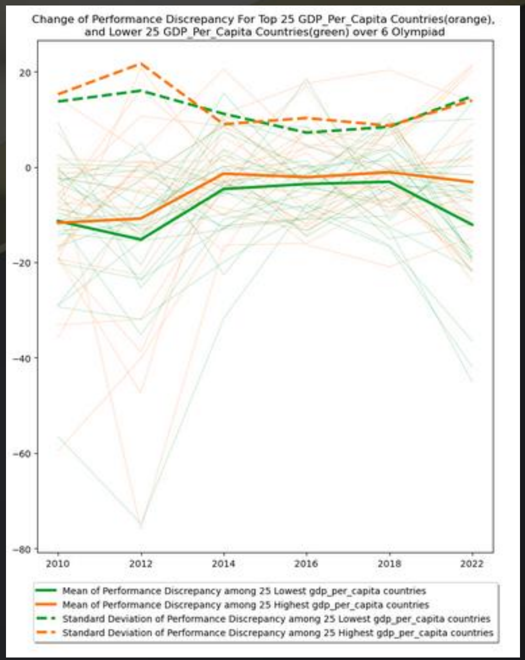
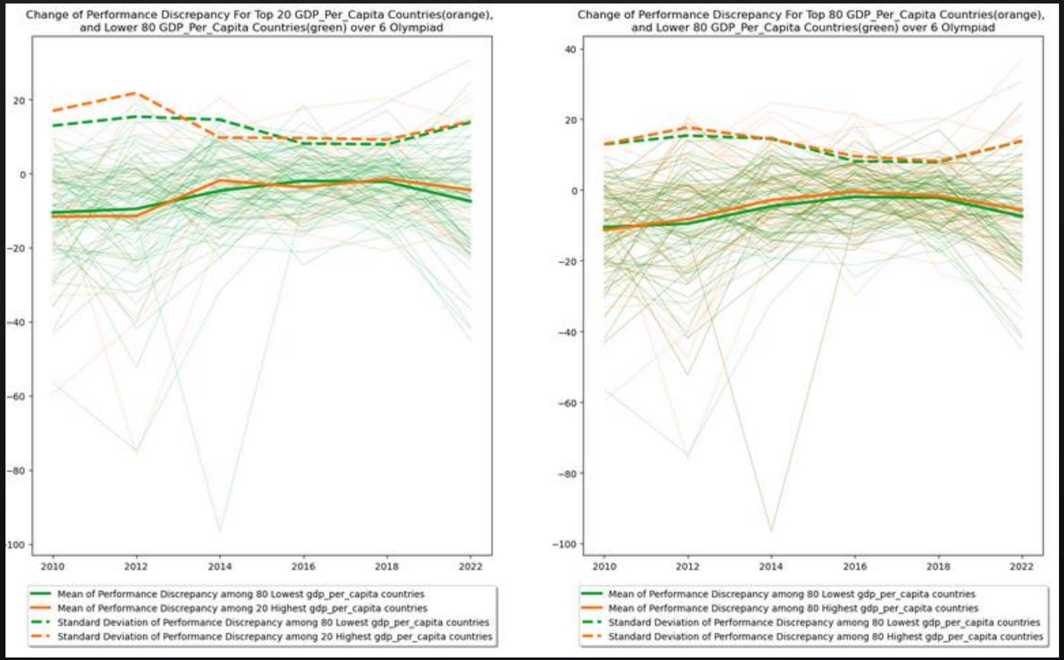


Figure 7. Performance discrepancy change over 6 Olympiads for top gdp_per_capita group(orange) and bottom gdp_per_capita group(green)

Figure 6. Heat map of correlation matrices for performance_discrepancy, culture factors, gdp_per_capita, log_gdp_per_capita and (chess_culture_index X log_gdp_per_capita)



Analysis

Line Plots of Consistently Overperform or Underperform Countries

We explored the countries whose **absolute** performance discrepancies are positive (overperforming) for at least 5 games out of the 6 Olympiads and countries whose absolute performance discrepancies are negative (underperforming) for at least 5 games out of the 6 Olympiads.

- **Observation**: Many more countries are underperforming, indicating that Olympiads may be more competitive than average FIDE tournaments.
- **Observation**: The countries that are consistently overperforming are LKA (Sri Lanka), NPL(Nepal), HKG(Hong Kong), KGZ(Kyrgyzstan), IDN(Indonesia), THA(Thailand), GEO(Georgia), PRI(Puerto Rico) These are all small countries mainly in Asia. This may be because these small countries participate less in FIDE-rated tournaments than other countries, so their ratings have not caught up to their true strength.

We explored the countries whose **normalized** performance discrepancies are positive (overperforming) for at least 5 out of the 6 Olympiads and countries whose absolute performance discrepancies are negative (underperforming) for at least 5 out of the 6 Olympiads.

- **How the normalization is done**: The average mean discrepancy over the years is around -5.77. We take this -5.77 as the benchmark. For an Olympiad tournament, if a country's absolute performance discrepancy is lower than the benchmark by more than a threshold value, we consider the country an underperformer in that year's Olympiad. Alternatively, If the absolute performance discrepancy value is higher than the benchmark by a threshold value, we consider the country an overperformer that year.
- **Observation**: The consistently overperforming countries in this context seem to be mainly small Asian countries, and consistent underperforming countries seem to be primarily small island (Oceania, Caribbean) countries.

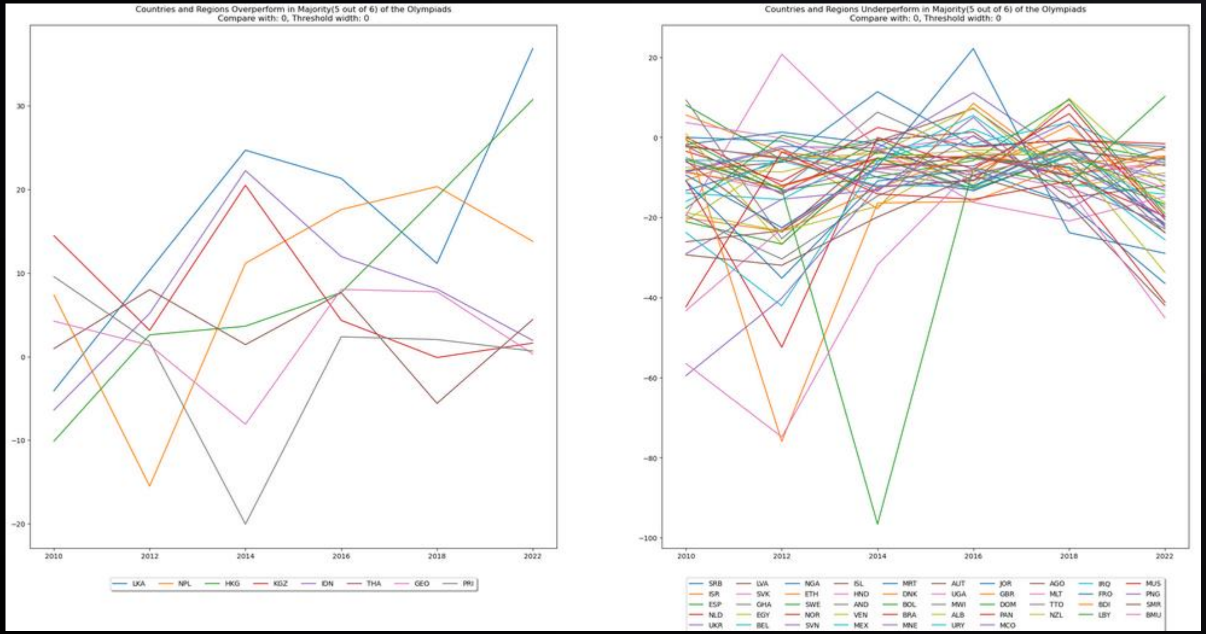


Figure 8. Countries having absolute performance discrepancy overperform or underperform at least 5 games out of 6 Olympiads

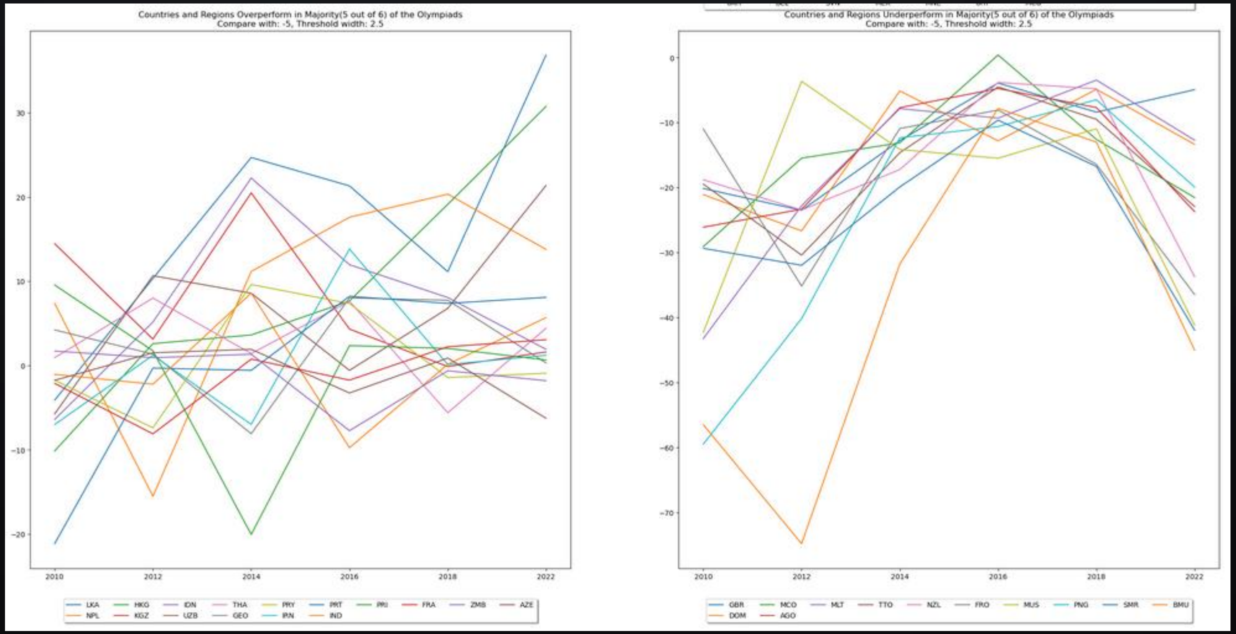


Figure 9. Countries having normalized performance discrepancy overperform or underperform at least 5 games out of 6 Olympiads

Discussion

Discussion, Limitations, and Next Steps

Summary

In conclusion, our study on rating discrepancies in chess Olympiads highlights the impact of limited international exposure on underrated players.

We've identified complex factors influencing rating disparities by analyzing data from the Chess Olympiad while factoring in geographical variables, chess culture, population size, and GDP per capita. Most federations tend to lose rating points on average in the Olympiad, possibly attributable to the presence of strong unrated players from newly admitted federations. However, there are consistently overperforming and underperforming countries, the former often being small countries' national teams that consist of non-top-tier players who rarely receive invitations to tournaments but have proved themselves worthy in local (possibly unrated) tournaments.

Overperforming countries do not follow a specific wealth or chess culture profile. In contrast, underperforming countries generally fall into two categories: economically disadvantaged with limited chess culture and economically well-off with rich chess traditions.

Interestingly, chess culture positively affects rating discrepancies, but the interaction between GDP per capita and chess culture shows a counterintuitive negative correlation. Many overperforming countries, predominantly in central/southern Asia, may face difficulties participating in FIDE-rated tournaments, which are more prevalent in Europe. This highlights the potential importance of the number of tournaments a federation hosts as a critical predictor of rating disparities.

In essence, these findings underscore the complexity of real-world problems and emphasize the need to consider a broad range of factors to avoid bias by omitting critical predictors in analytical assessments.

Limitations

The ratings and performance of chess players are complex. We made the decision to use chess Olympiad data instead of other tournament data because all countries were represented, not just those that were wealthy. Although this may have reduced sampling bias, this may have limited our sample counts—teams are only made of 4 to 5 players, and each team only plays 11 rounds.

Confounding variables such as travel or team selection (best players from a country aren't representative of the average player from that country) may play a role in team performance. We were not able to look at this during our project.

We had to assume that chess culture primarily resulted from having top chess players. Many other aspects of chess culture, such as chess clubs, chess park culture, tournaments, and forms of media, were not examined at this time.

Next Steps

We found that a group of countries did particularly poorly in 2022, losing a lot more points than any country lost in the Olympiads prior. It may be worthwhile to examine exactly what caused this.

Exploring data on confounding variables such as distance to the host site and team selection may bring more insights into chess rating discrepancy.

Running a regression model may provide us with more precise data on exactly which factors have the most influence on rating discrepancy. It may even be used to predict a team's future performance.

Statement of Work

Work on this project was highly collaborative. We had weekly audio/video meetings among the team members as well as frequent consultation audio/video meeting with our mentor. Each member of the team contributed substantially in brainstorming, retrieving, cleaning, analyzing, visualizing, and discussing data, as well as writing up code annotations and the report. We also communicatde about confusions and concerns timely through slack. The following is a more specific breakdown of responsibilities:

Jim Yang	Ke Miao	Michael Nguy
• Problem formulation and project proposal. • Sourcing, cleaning, and manipulation of datasets. • Coding analysis, creation of visualizations, and code reformatting. • Revision and expansion of other members' code. • Writing markdowns in notebooks. • Design and writing of technical report.	• Problem formulation and revision of project proposal. • Creation and management of version control repository. • Furthering of data manipulation and dataset merging. • Plotting and revisions of key visualizations, leading to several insights. • Writing a portion of markdowns in analysis.ipynb. • Writing, revision, and styling of technical report.	• Problem formulation and revision of project proposal. • Data manipulation of certain datasets. • Editing markdowns of notebooks. • Rough draft and outline of parts of the technical report. • General revisions of technical report.

Possible Future Improvement in Collaboration

A possible improvement is to better utilize the code repository. In our project, we used GitLab for basic version control. 1) In the future, we can use GitLab's review and approval mechanism to track peer feedback. 2) Also, one member encountered temporary difficulties in checking in code and had to email work to another member to upload. In the future, it may be a good idea to have a backup version control plan for unexpected conditions.

References

- Forrest, D. et al. Who Wins at the Chess Olympics? The Role of Resources and Education Capital. Working Paper in Economics of University of Liverpool, April 2020, https://www.liverpool.ac.uk/media/livacuk/schoolofmanagement/research/economics/who,wins,at,the,chess,olympics.pdf
- Bodek, M. Assessing Regional Miscalibration of Elo Ratings in Chess, March 2019, https://qss.dartmouth.edu/sites/program_quantitative_social_science/files/bodek_elo_paper_19w_vfinal_one_quarter.pdf
- Visser, M. Self-selection in tournaments: The case of chess players, Journal of Economic Behavior & Organization, June 2016, https://www.sciencedirect.com/science/article/pii/S0167268116300208
- Celik, O. B. Estimating the Determinants of Summer Olympic Game Performance, March 2014, https://www.researchgate.net/publication/267391137_Estimating_the_Determinants_of_Summer_Olympic_Game_Performance
- ISO 3166-1 alpha-3, 3 letter country codes, Wikipedia: https://en.wikipedia.org/wiki/ISO_3166-1_alpha-3