Visualizing the International Mathematics Olympiad results

Members
Member 1: Chau Minh Khai 21khai.cm@vinuni.edu.vn
Member 2: Nguyen Tuan Anh 21anh.nt@vinuni.edu.vn
Member 3: Ta Viet Thang 21thang.tv@vinuni.edu.vn

1 Introduction

Our project analyzes historical data from the International Mathematical Olympiad (IMO)—the world's most prestigious mathematics competition for high school students. Established in 1959 in Romania, the IMO has grown into a global event, with over 100 countries participating annually. The competition challenges students with six advanced mathematical problems over two days, and medals are awarded based on individual and team performance.

The dataset used in this analysis spans the entire history of the competition from 1959 to 2024 and is organized into three components: **Country results**, **Individual results**, and **Timeline data**.

We addresses the following two research questions:

1. **How has the IMO evolved over time in terms of participation, problem difficulty, and national performance?**
   We analyze trends using: Number of countries and contestants (`country`, `team_size_all`), Average problem scores (`p1`–`p6/7`) across years, Medal counts (`awards_gold`, `awards_silver`, `awards_bronze`) by country.
2. **How do gender and socioeconomic factors relate to IMO participation and performance?**
   We focus on trends in female participation and performance (`team_size_female`, `female_contestant`, `award`), Correlations between GDP indicators (from support datasets) and team performance (`total` score, `awards_gold`).

Exploring these questions allows us to examine how countries have engaged with elite mathematical competition over time, and how factors such as gender representation and economic development influence outcomes. Despite some inconsistencies in early data and minor changes in scoring formats, the dataset provides a rich foundation for exploring the evolving landscape of mathematics competitions and their educational implications.

2 Question 1

2.1 Introduction

In this question, we explore how the IMO has changed over time in terms of country participation, problem difficulty, and top-performing nations. This question only uses the IMO datasets.

- **Country participation**: How have nations from different continents and regions participated in the IMO over time?
- **Problem difficulty**: Analyze the changes in the difficulty levels of the six IMO problems over time.
- **Top-performing nations**: Find nations which are most successful in the history of IMO.

We're interested in this topic because the IMO reflects broader trends in global education and talent development. By analyzing key data points—such as participant counts, participation distribution across continents, top-performing nations aim to uncover major shifts in the competition's structure and impact over the past 65 years.

2.2 Approach

2.2.1 Country participation

In this question, the global map is used to visualization. We used a interaction map which has a slide bar to select the year. In the map, nations, which is participated in the IMO in such year, are colored. I think this type of graph is the best option since users can easily observe participating nations by their geographical location overtime from 1959 to now. Moreover, This geographical visualization can show clearly the participating trends of continents overtime, which is the result that are needed to answer the question.

2.2.2 Problem difficulty

In this question, we used a line graph with 6 lines representing respectively average points scored on each of the six IMO problems from 1960 to 2024. A line graph is ideal for analyzing changes in difficulty levels of the six IMO problems over time because it clearly shows trends and patterns across years. By plotting the average score for each problem annually, we can visually compare how the difficulty of each problem has evolved. The continuous lines make it easy to identify which problems are consistently more challenging (lower average points) and spot any sudden shifts or long-term trends. Additionally, using different colors for each problem allows for easy distinction and comparison between them.

2.2.3 Top-performing nations

In this question, we use a bar chart to show to the top 20 countries with the highest number of gold medals in the history of the International Mathematical Olympiad. This graph is ideal for identifying the most successful nations in IMO history because it clearly ranks countries by their total gold medals. The bar lengths and color gradient make it easy to compare performance at a glance. It highlights which countries have consistently excelled over time.

2.3 Analysis and visualization

2.3.1 Country participation

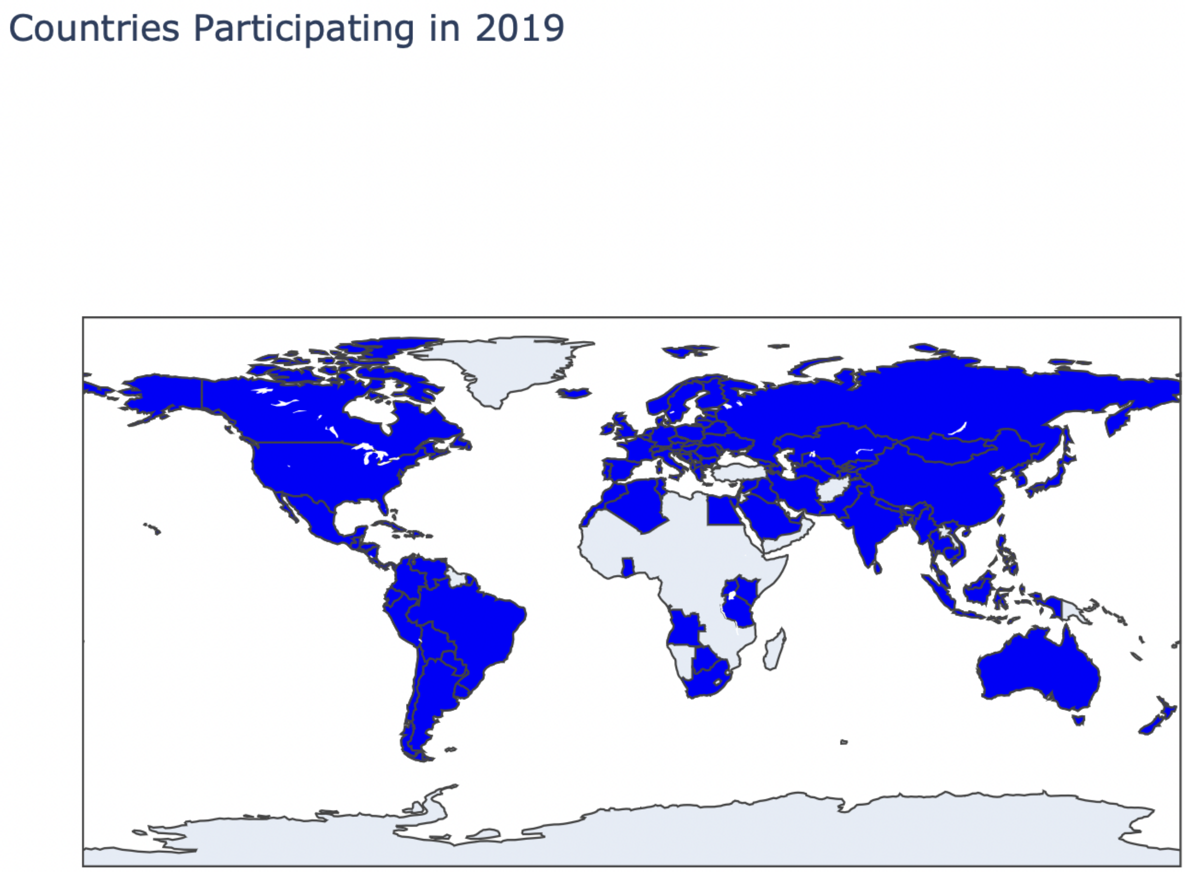

Figure 1: Map of participating nations in 2019

In 1959, the first International Mathematical Olympiad (IMO) was held as a regional event with only seven participating countries, all from Eastern Europe. This reflected the competition's limited geographic reach in its early years. Over time, especially during the period from 1980 to 2000, the IMO experienced significant growth in global participation. More countries from Latin America, Africa, and Asia began to join, gradually transforming the IMO into a truly international event. By 2024, the competition includes most nations from Europe, the Americas, and Asia, demonstrating its broad global appeal and recognition. However, participation from African countries remains relatively low, with only a few nations from the continent competing regularly, indicating that regional disparities in access to the competition still persist.

2.3.2 Problem difficulty

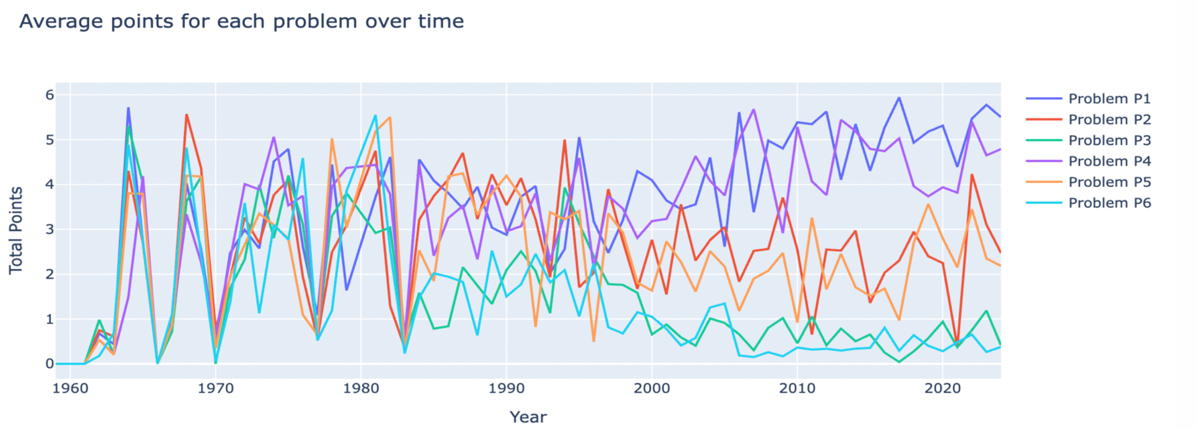

Figure 2: Average points for each problem over time

Before 2000, average problem scores fluctuated significantly year to year. Since 2000, a clear pattern has emerged: Problems 1 and 4 tend to be the easiest, followed by Problems 2 and 5. Problems 3 and 6 consistently rank as the most difficult.

2.3.3 Top-performing nations

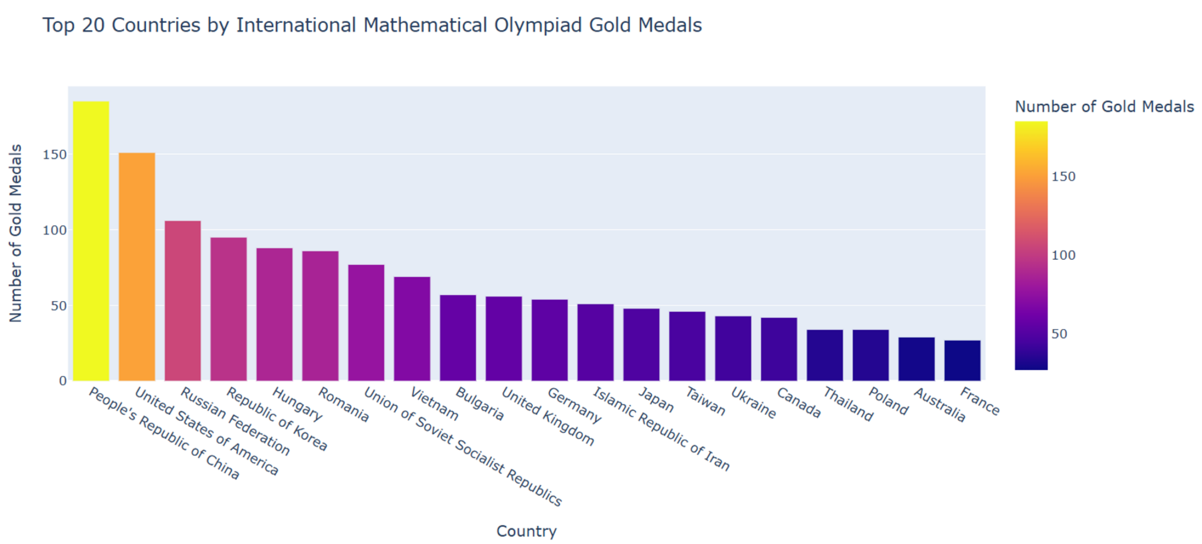

Figure 3: Top 20 Countries by International Mathematical Olympiad Gold Medals

China, USA, and Russia are top 3 countries dominating the leaderboard, each earning over 100 gold medals. China leads significantly, reflecting its consistent excellence since entering in 1985. Korea, Hungary, and Romania follow, showing strong long-term performance. Emerging countries like Vietnam and Iran also appear in the top rankings.

3 Question 2

3.1 Introduction

In this question, we explored on the relationship between the historical International Mathematics Olympiads results, together with some demographical and socioeconomic backgrounds of participating countries. More specifically, we will explore on these perspectives:

- **Gender distribution**: How has the female participants performed overtime?
- **Socioeconomic Factors**: How do external factors such as GDP and GDP per capita affect the performance of teams? Are there any evidences on the relationship between these socioeconomic factors vs. team performance across the time?

With insights from these questions, we could have some more understanding about impacts of education to countries' development in either social equality, or economic growth.

3.2 Approach

3.2.1 Support datasets

To provide information about gender equalities and economic growth, we are using 03 more datasets, for answering each of above questions

- For question about gender equality, we introduce Gender Inequality Index [Pet23] dataset, which providing more insights about Human Development Groups, as well as the Gender Inequality Index from 1990 to 2021. The dataset does not only provide us information about Gender Inequality Index, but also Development Group, ranking based on Human Development Index (HDI), grouped into 4 ranks: *Very High, High, Medium, and Low*
- For question about correlation between GDP and GDP per capita versus IMO performance, we extracted the GDP and GDP per capita in GDP per capita (1970 - 2022) [Hus22] and GDP from 1960 to 2020 [Chr22].

3.2.2 Data cleaning

When introducing new datasets, we are required to tackle several problems for the consistency between supporting datasets and main datasets on IMO results, including:

- **Country name inconsistency**. Some countries have variance in name, for example, "Viet Nam" and "Vietnam" refers to the same country. To tackle this, we manually find those inconsistency, with some supporting scripts to correct the name to be matched for merging.
- **Missing countries**. Due to political changes, or the countries are not participating the competition, as shown in Question 1, several country names are not appeared in the the supporting datasets. As the complexity of merging/unmerging historical names is significantly high, required both expert knowledges, we accept this and ignore the country name missing. Given any year, we will only merge countries appear in both datasets.

3.3 Analysis and visualization

3.3.1 Gender Distribution Trend

First, we have the gender distribution overall trend from the start of the competition. For this plot, we chose trendline as the data dimension is low, but we need to keep track of changes across competitions.

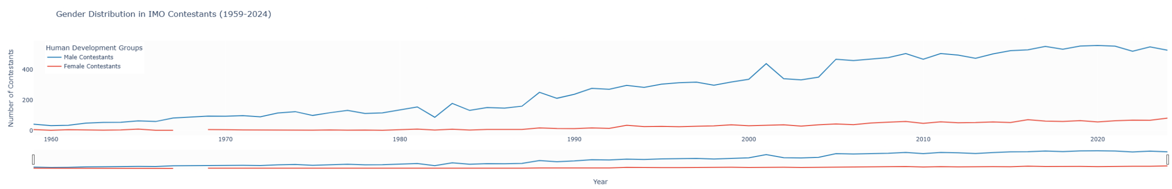

Figure 4: Overall Gender Distribution across IMOs

From the visualization, the participation of female contestants are not significantly increasing. Moreover, the gap between numbers of male versus female contestants are increasing. According to [Con24], this is mainly due to the attraction of the competition is on more male students comparing to female ones. To resolve this, they are hosting Girls' Mathematical Olympiad, to attract more female students into the competitions.

3.3.2 Gender Distribution versus Development Group

From the supporting dataset of gender equality, we created another trendline graphs to show the change across development group.

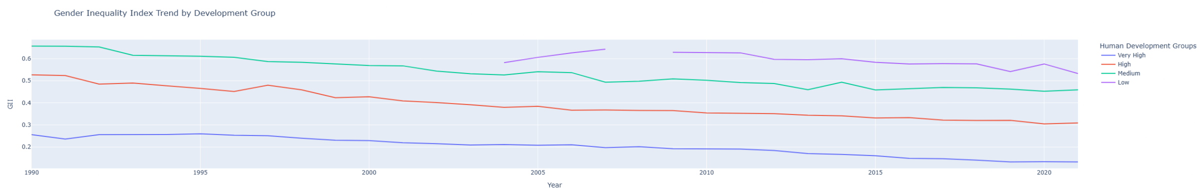

Figure 5: Overall Gender Distribution across IMOs

From the trendlines, although low developed countries are participating the competition relatively later than more developed countries, they are leading in the female participants ratio across the IMO competitions. The more developed the countries, the lower the female participants ratio in the team is. It is generally hard to answer this behavior, without more information. To the best of our knowledge, this could implies the overall trend of female contestants in 5. There are other interesting competitions rather than academic ones like IMO for developed countries, so the showing trend in this figure is unreliable to conclude any information about gender equality in education.

3.3.3 GDP and GDP per capita versus IMO performance

For the next two visualizations, we are also using trendlines, but there are more figure is introduced. As we are comparing both IMO performance vs GDP/GDP per capita, the data complexity is increasing. To simplize the implementation, we made interactive visualization, where we can select two variables:

- Year: Selecting the year for analysis.
- Country: Selecting the country for their in-depth performance review.

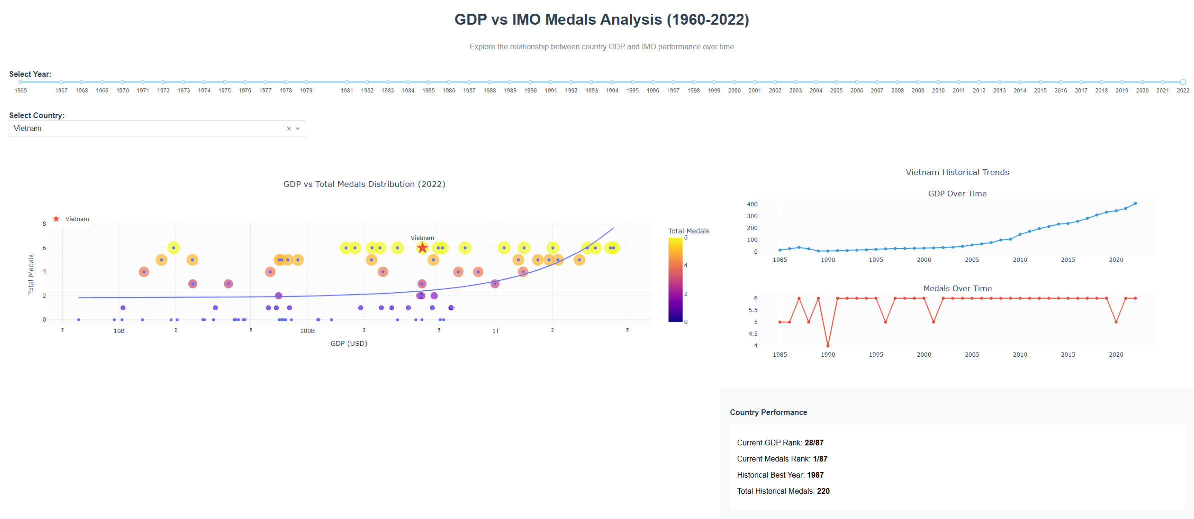

Figure 6: User interface for the economic comparision

There are three subplots inside this user interface:

1. A overall distribution of GDP vs. IMO medal counts across countries, as a scatter plot, where the x-axis is GDP, and y-axis is the medal count. The country's medal counts are color-noted by a heatmap, showing how many medals they have received in the selected year. Also, a Ordinary Least Square prediction line is drawn from the data of that year, indicating possible impacts of GDP to medal counts.
2. Two trendlines in the left, showing the distrbution of the GDP/GDP per capita and the medal counts of countries across time. This visualization aimed to show the relationship between economic factors to the performance of the team.
3. A small quick report showing the performance of the countries in both economic and IMO will summary the indicators of selected country.

We will analyze the trend in 2022.

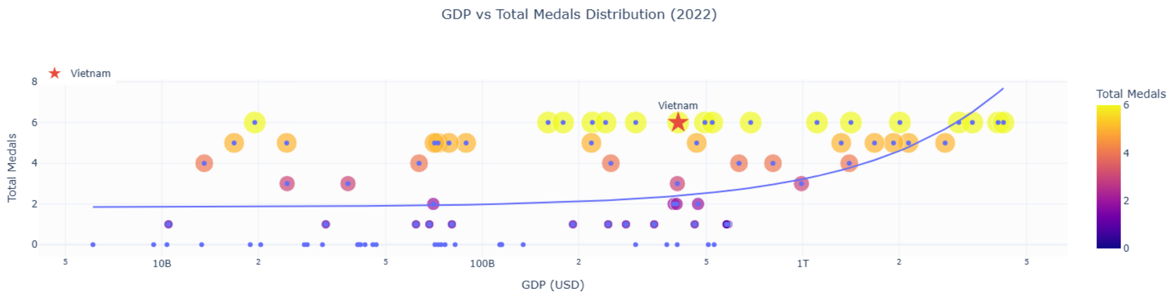

Figure 7: Distribution of medals vs. GDP in 2022

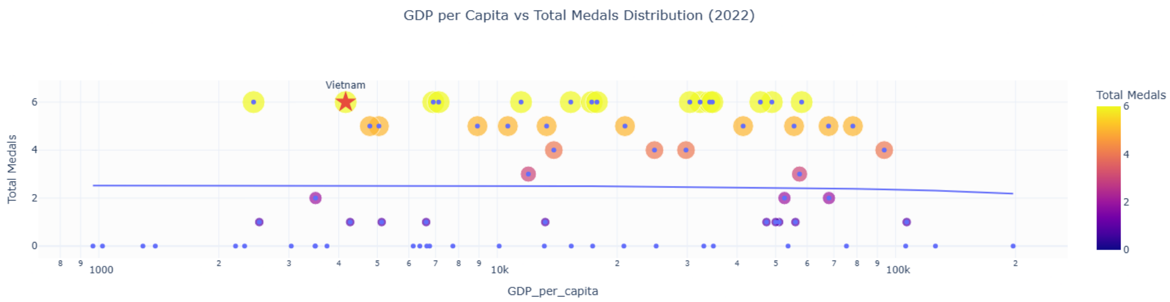

Figure 8: Distribution of medals vs. GDP in 2022

From the GDP figure in 2022, we can see that there is no significant trend in relationship between GDP and IMO medal counts. There are high GDP countries with not many medals in the competition, as well as countries has lower GDP reached 6/6 medals. This behavior is also happening in GDP per capita figure. However, as the supporting OLS trendline suggest, a country with higher GDP is likely to have more medals, as the line increases toward the increasement of GDP. However, there are more in-depth analysis required for verifying this fact, as the OLS trendline is not strictly follow the prediction, especially in GDP per capita figure.

Furthermore, IMO is a competition of advanced mathematics, comparing with the level of mathematics required by most of the countries. Moreover, the sample size for evaluating is just 06 students per countries per year, which is relatively much smaller comparing the whole number of students. There are others metric which could provides more general results, such as Programme for International Student

Assessments (PISA) [Con19]. Therefore, we suggest that IMO results and the GDP/GDP per capita are not reliable metrics for predicting or regression for each other, as well as the IMO results are not enough to conclude any insights about any countries' human development, as well as education level.

4 Conclusions

For Research Question 1, we observed clear trends in country participation, shifting problem difficulty, and long-term dominance by nations like China, the USA, and Russia. These trends reflect educational focus and national investment in math training, but they do not fully represent the broader educational landscape.

For Research Question 2, we found that while gender gaps are slowly closing, female participation remains low. Moreover, there is no strong or consistent correlation between socioeconomic indicators like GDP or GDP per capita and IMO success. Wealthier nations do not always outperform lower-income ones, highlighting that national performance in the IMO is influenced by many factors beyond economics.

Ultimately, while the IMO is valuable for recognizing individual excellence, its outcomes should not be overinterpreted as a measure of a country's overall educational or economic strength.

References

[Con19] Wikipedia Contributors. *Programme for International Student Assessment.* Feb. 2019. URL: https://en.wikipedia.org/wiki/Programme_for_International_Student_Assessment.

[Chr22] Rini Christy. *Countries GDP 1960-2020.* 2022. URL: https://www.kaggle.com/datasets/rinichristy/countries-gdp-19602020.

[Hus22] Ibrar Hussain. *Per Capita GDP of All Countries 1970 to 2022.* 2022. URL: https://www.kaggle.com/datasets/dataanalyst001/per-capita-gdp-of-all-countries-1970-to-2022.

[Pet23] Gianina-Maria Petrașcu. *Gender Inequality Index.* 2023. URL: https://www.kaggle.com/datasets/gianinamariapetrascu/gender-inequality-index.

[Con24] Wikipedia Contributors. *International Mathematical Olympiad.* Feb. 2024. URL: https://en.wikipedia.org/wiki/International_Mathematical_Olympiad.