COMP4010: Project 1

Visualizing the International Mathematics Olympiad

Nguyen Tuan Anh
Chau Minh Khai
Ta Viet Thang

Table of Contents

Why IMO?

- The International Mathematics Olympiad (IMO): The world's most prestigious math competition for high school students since 1959.

- Why visualize this data? It offers unique insights into:
  - Educational system performance globally
  - Cultural and socioeconomic influences on STEM.
  - Gender equity trends in mathematics.

- Our goal: To use data visualization to explore the IMO's rich 65-year history and uncover patterns in participation, performance, and demographics.

Dataset

Official IMO records covering 65 years (1959-2024)

- **Country Results:** Team size (overall, M/F), problem scores (P1-P6/P7), medal counts, team leaders.
- **Individual Results:** Contestant names, country, individual problem scores, total score, rank, award.
- **Timeline Data:** Edition, year, host country/city, participant counts (total, M/F), dates.

variable	class	description
year	integer	Year of IMO
country	character	Participating country
team_size_all	integer	Participating contestants
team_size_male	integer	Male contestants
team_size_female	integer	Female contestants
p1	integer	Score on problem 1
p2	integer	Score on problem 2
p3	integer	Score on problem 3
p4	integer	Score on problem 4
p5	integer	Score on problem 5
p6	integer	Score on problem 6
p7	integer	Score on problem 7
awards_gold	integer	Number of gold medals
awards_silver	integer	Number of silver medals
awards_bronze	integer	Number of bronze medals
awards_honorable_mentions	integer	Number of honorable mentions
leader	character	Leader of country team
deputy_leader	character	Deputy leader of country team

variable	class	description
year	integer	Year of IMO
contestant	character	Participant's name
country	character	Participant's country
p1	integer	Score on problem 1
p2	integer	Score on problem 2
p3	integer	Score on problem 3
p4	integer	Score on problem 4
p5	integer	Score on problem 5
p6	integer	Score on problem 6
total	integer	Total score on all problems
individual_rank	integer	Individual rank
award	character	Award won

variable	class	description
edition	integer	Edition of International Mathematical Olympiad (IMO)
year	integer	Year of IMO
country	character	Host country
city	character	Host city
countries	integer	Number of participating countries
all_contestant	integer	Number of participating contestants
male_contestant	integer	Number of participating male contestants
female_contestant	integer	Number of participating female contestants
start_date	Date	Start date of IMO
end_date	Date	End date of IMO

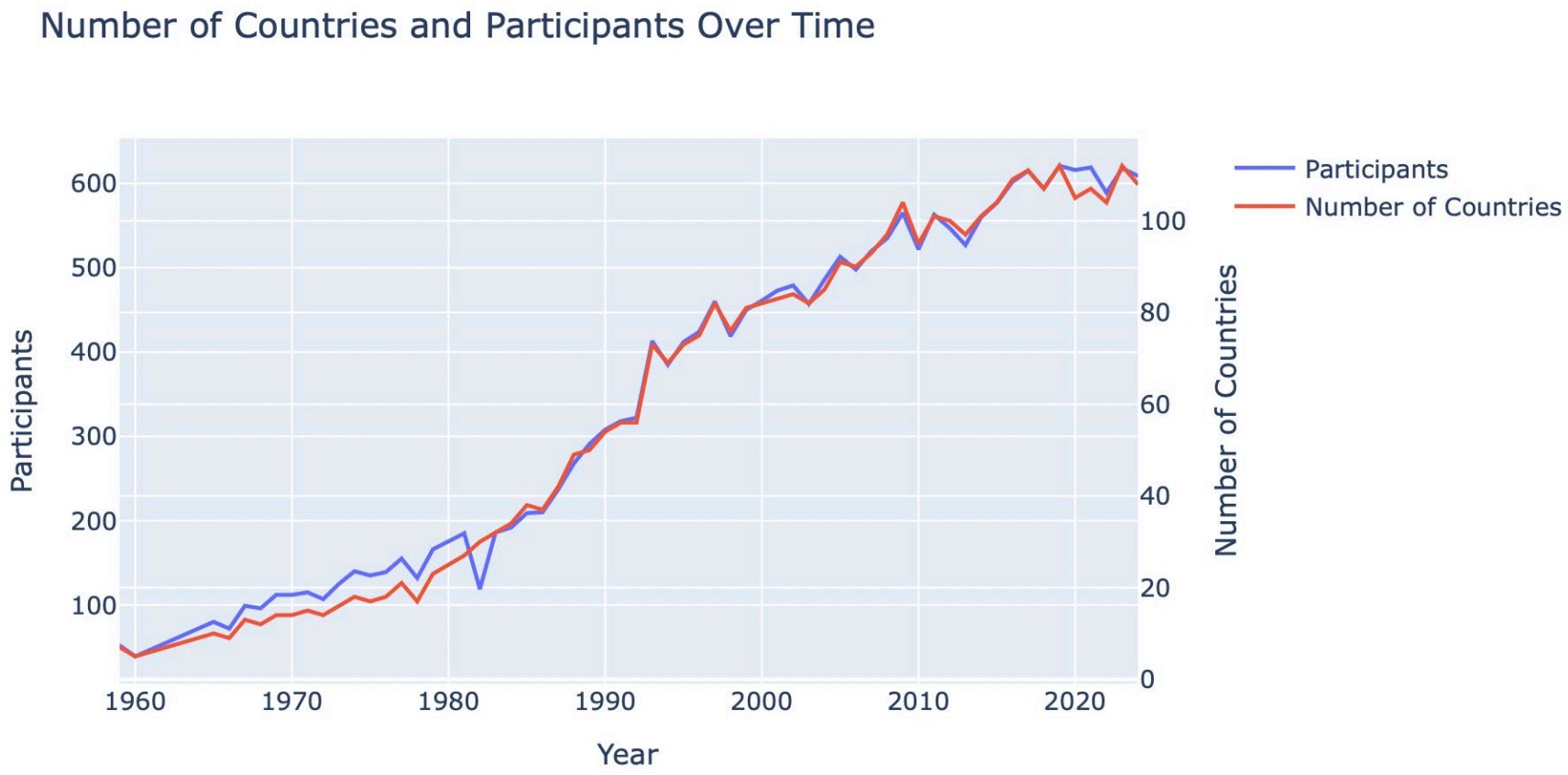
Research questions

- **RQ1: Competition evolution:** How has the IMO changed over 65 years regarding
  - Global participation?
  - Competition structure (hosting, difficulty)?
  - Performance metrics (scores, medals)?

- **RQ2: Participation demographics:** What patterns emerge when analyzing participation and performance by:
  - Gender distribution over time and region?
  - Correlation with Socioeconomic factors (GDP)?

RQ1 Findings

Number of Countries and Participants Over Time



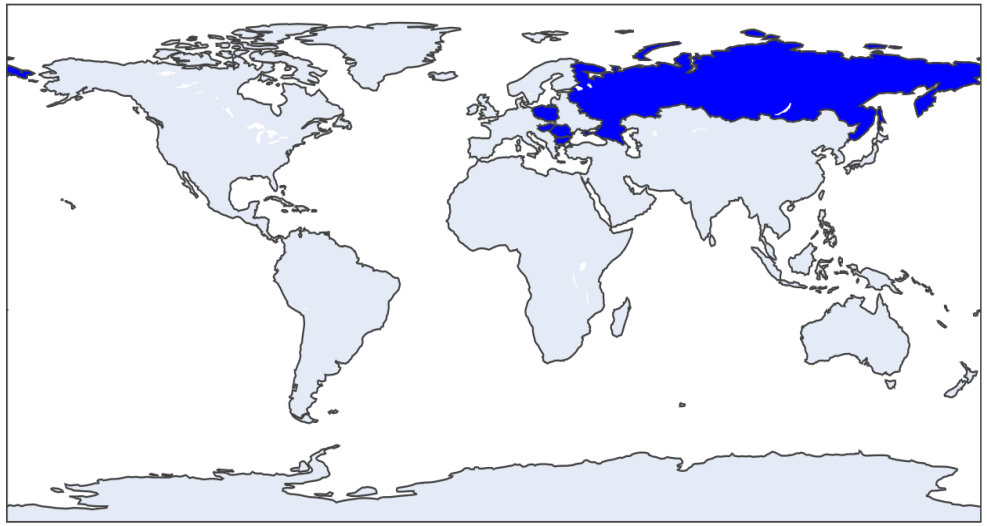
Correlation: 0,99774

RQ1 Findings

Countries Participating in the Competition

1959

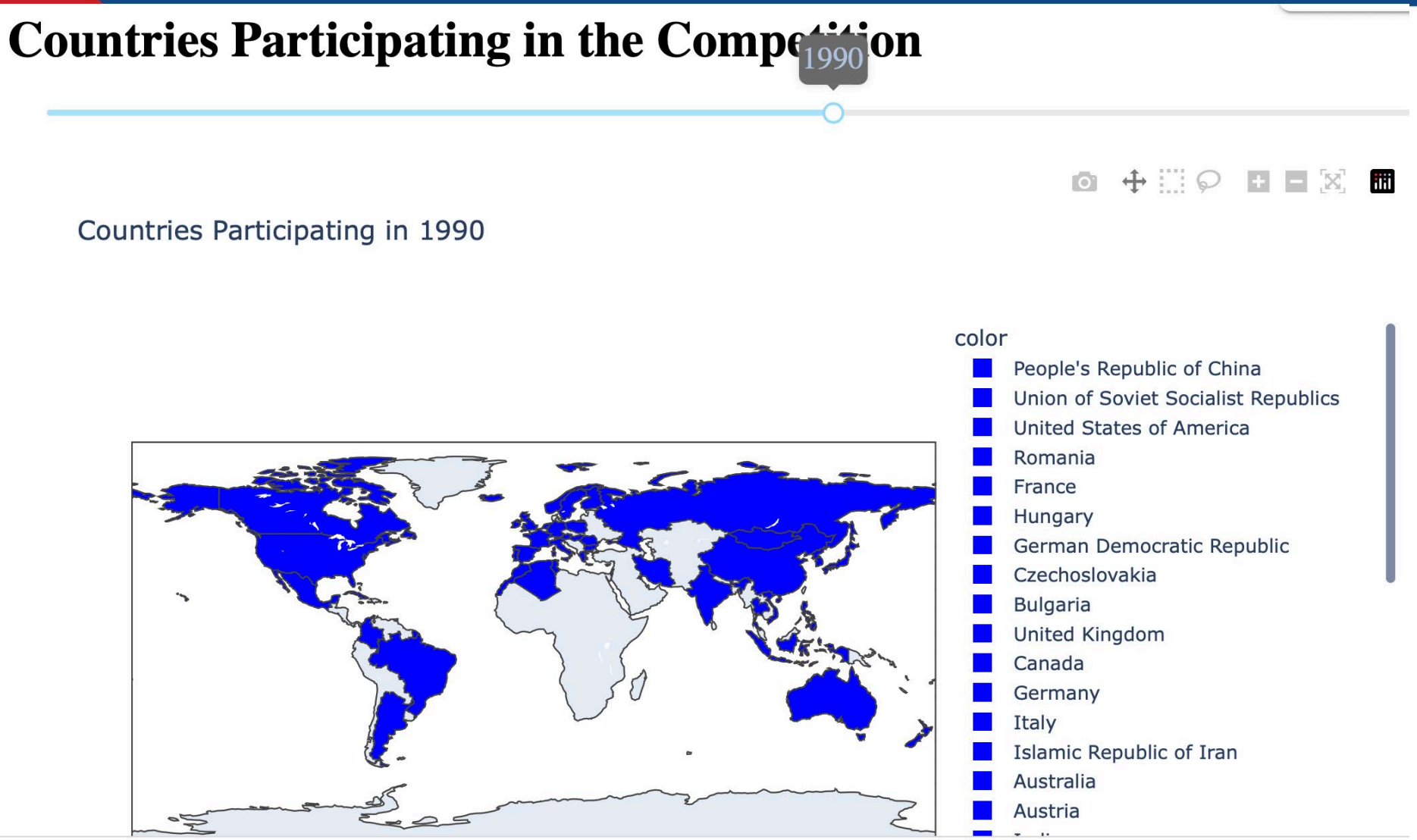
Countries Participating in 1959



color
- Romania
- Hungary
- Czechoslovakia
- Bulgaria
- Poland
- Union of Soviet Socialist Republics
- German Democratic Republic

In the 1959, the first competition is just a regional event with 7 nations from East Europe

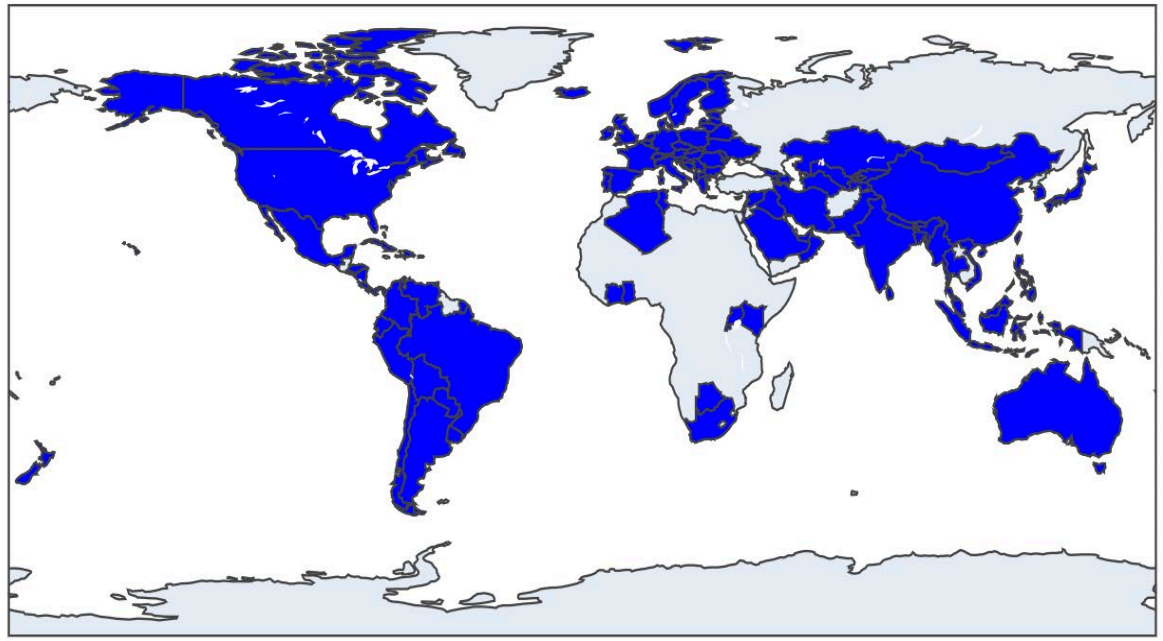
RQ1 Findings



Overtime, more Latin America, Africa and Asia nations join the competition.

RQ1 Findings

Countries Participating in 2024

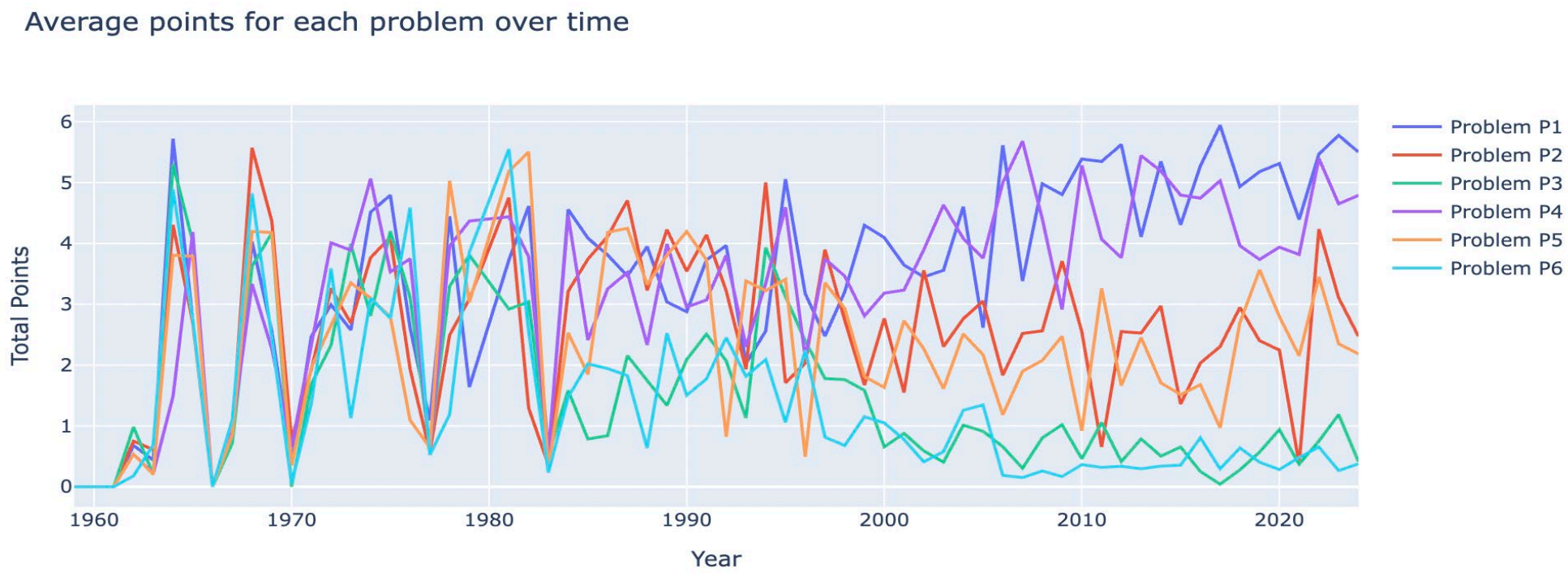


color
- United States of America
- People's Republic of China
- Republic of Korea
- India
- Belarus
- Singapore
- United Kingdom
- Hungary
- Poland
- Türkiye
- Taiwan
- Romania
- Bosnia and Herzegovina
- Italy
- Japan
- Israel
- Mongolia
- Hong Kong

In 2024, Most nations in Europe, America, Asia are participating in the IMO, but just a few nations in Africa are competing in the IMO.

RQ1 Findings

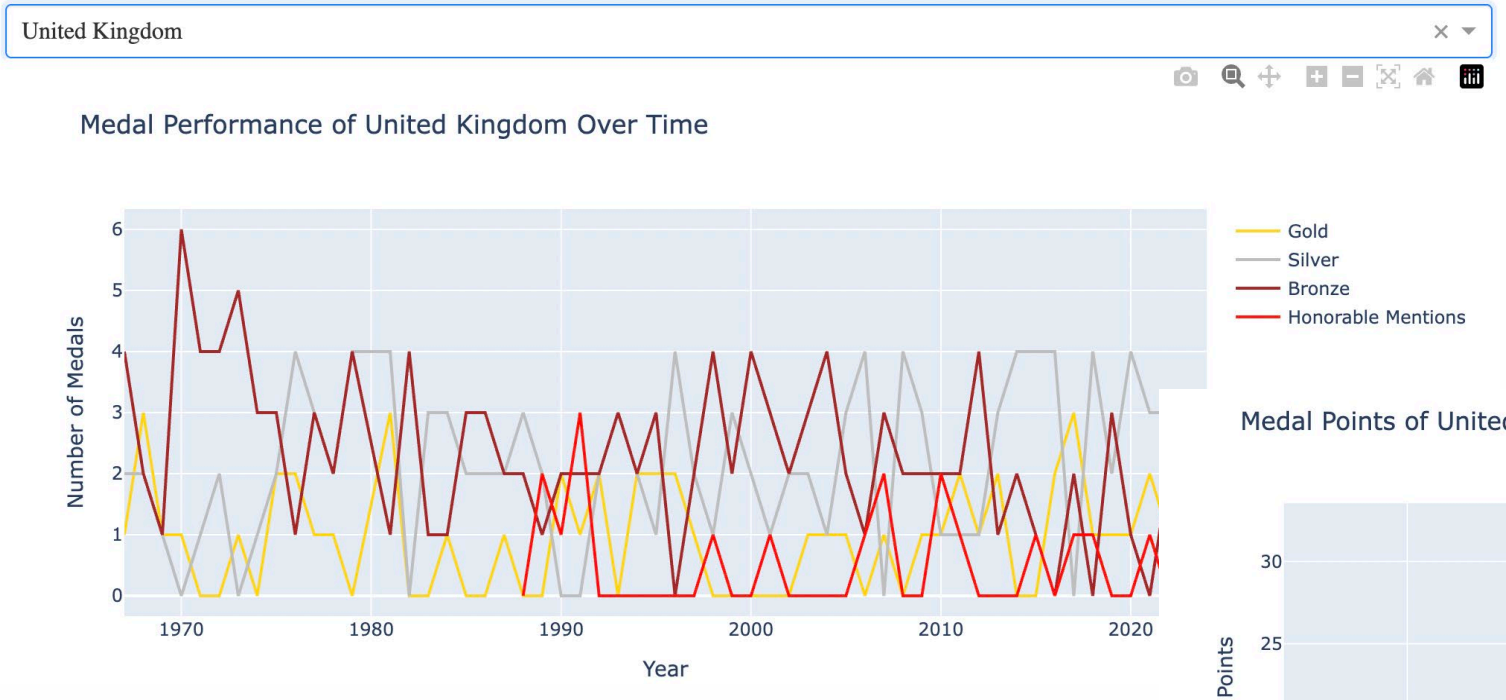
Average points for each problem over time

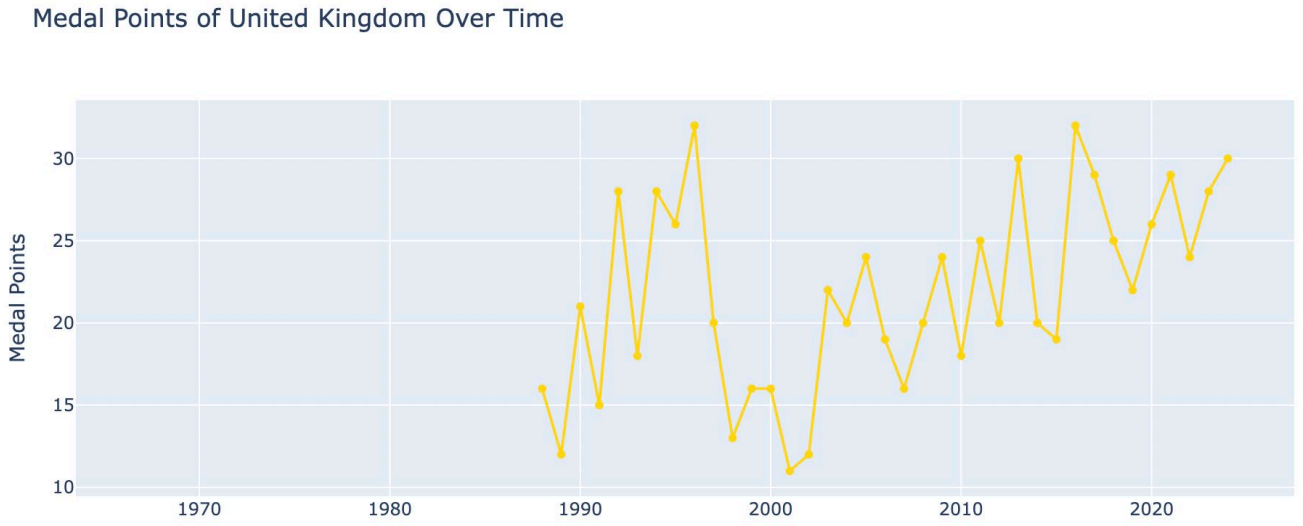


Before 2000, the average points for problems are strongly fluctuate
From 2000 to now, Problem 1 and Problem 4 are easiest, then Problem 2 and Problem 5, hardest problems are Problem 3 and Problem 6

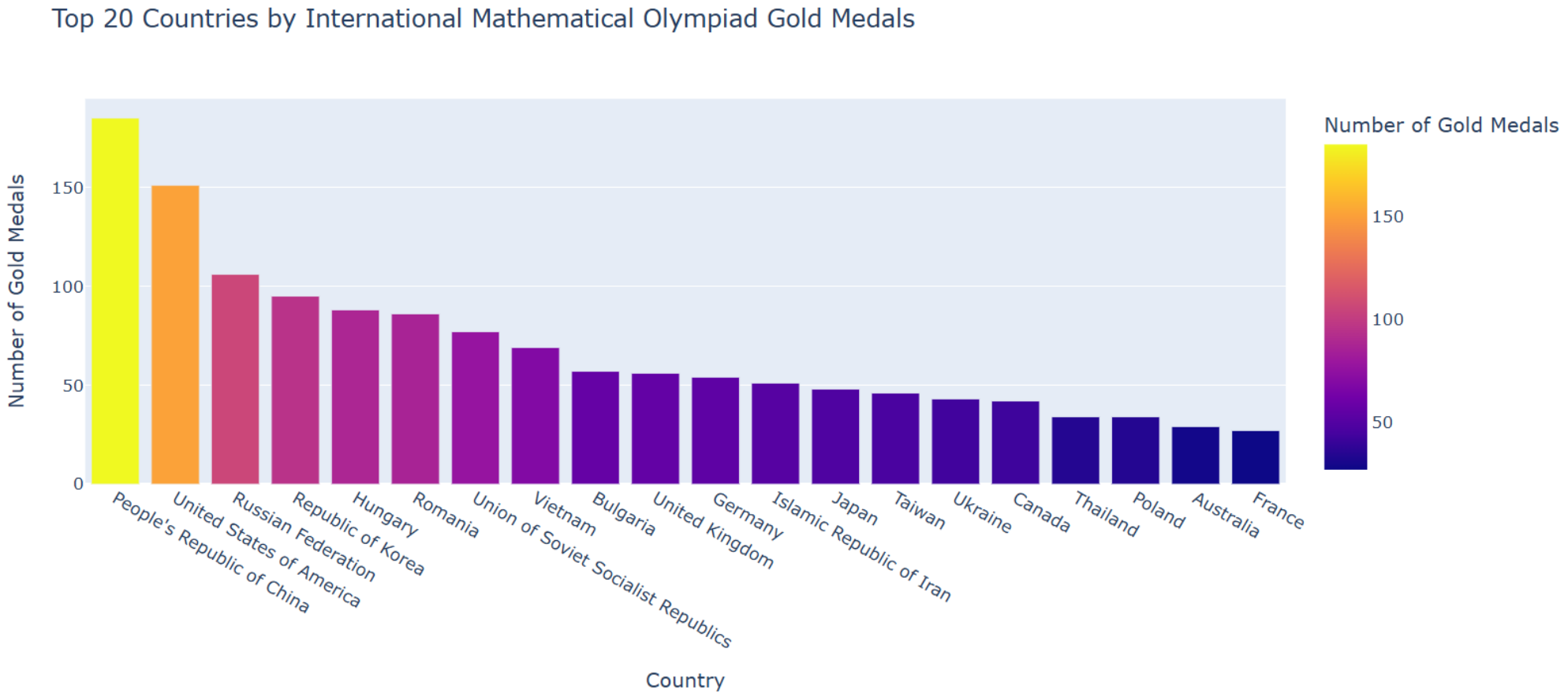
RQ1 Findings

Country Medal Performance Over Time





RQ1 Findings



China, USA and Russia are top 3 in gold medals during the whole history of IMO.

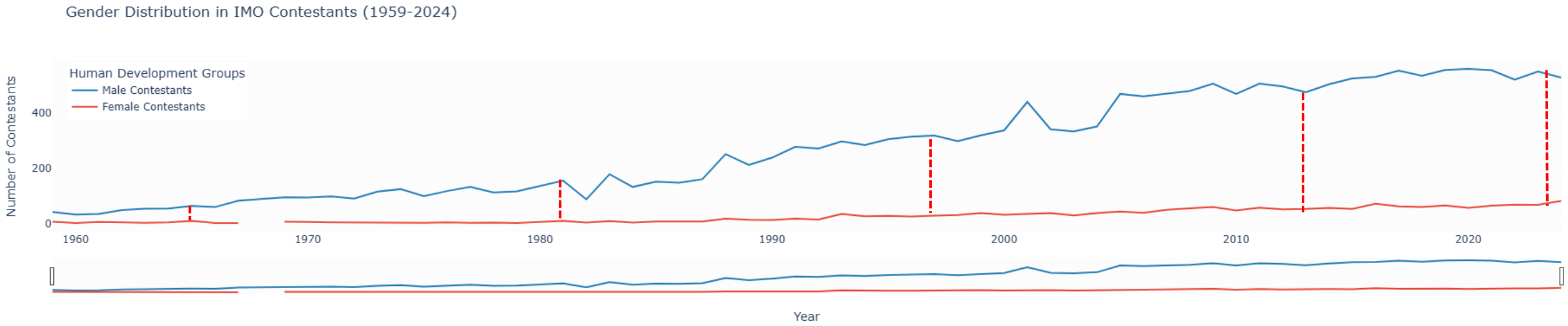
RQ2 Findings

1. **Gender distribution over time and region?**

- We have added one more datasets: Gender Inequality Index, which providing more insights about Human Development Groups, as well as the Gender Inequality Index from 1990 to 2021.
- Knowing that the IMO teams are often 06 students, we will calculate `female_ratio`, the proportional of female members in a team.
- We will try to get to know more about the overall trend, relationship of this ratio between high – low human development group & geographical regions.

RQ2 Findings

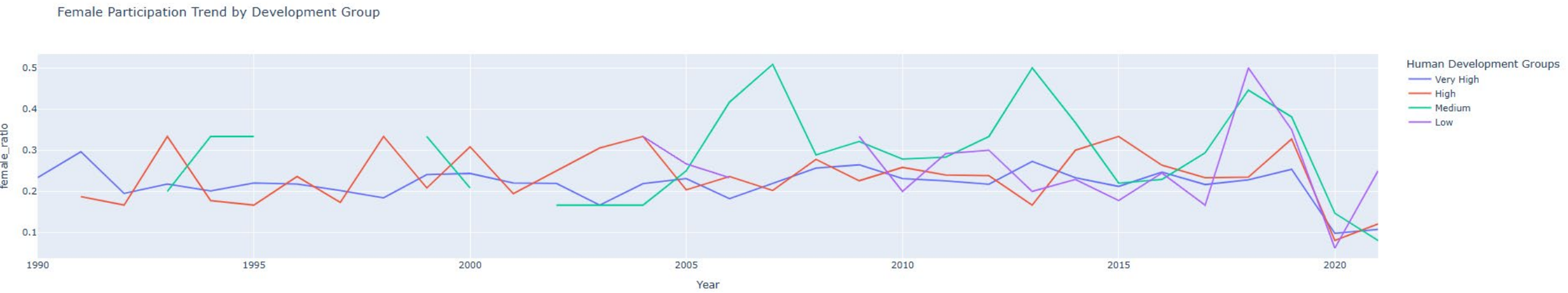
1. **Overall trend**



Overall, the gender gap in numbers of female contestants vs. male contestants is increasing.

RQ2 Findings

2. **By Human Development Group**



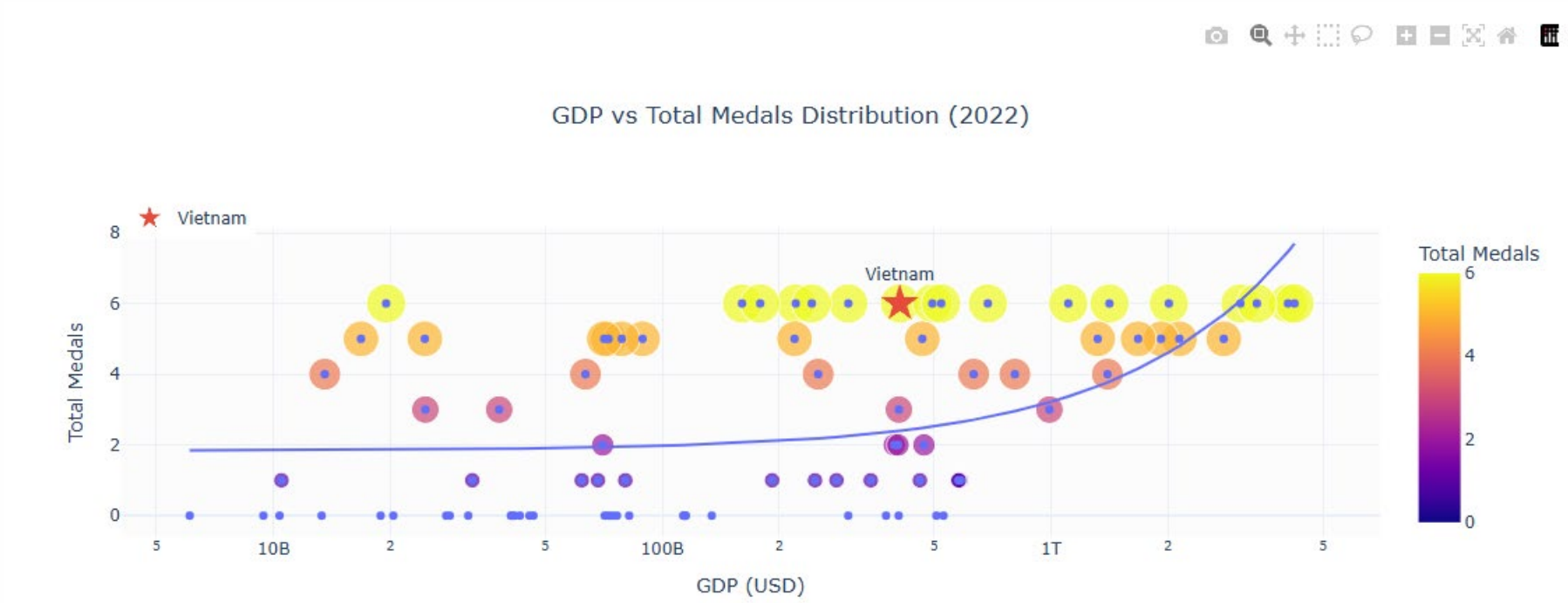
- Regardless of development group, the female ratio is going around 20-30% (equivalent to 1-2 students)
- From 2017, a downtrend in this ratio occurred in all human development group countries, but recently there is some increasing signal.

RQ2 Findings

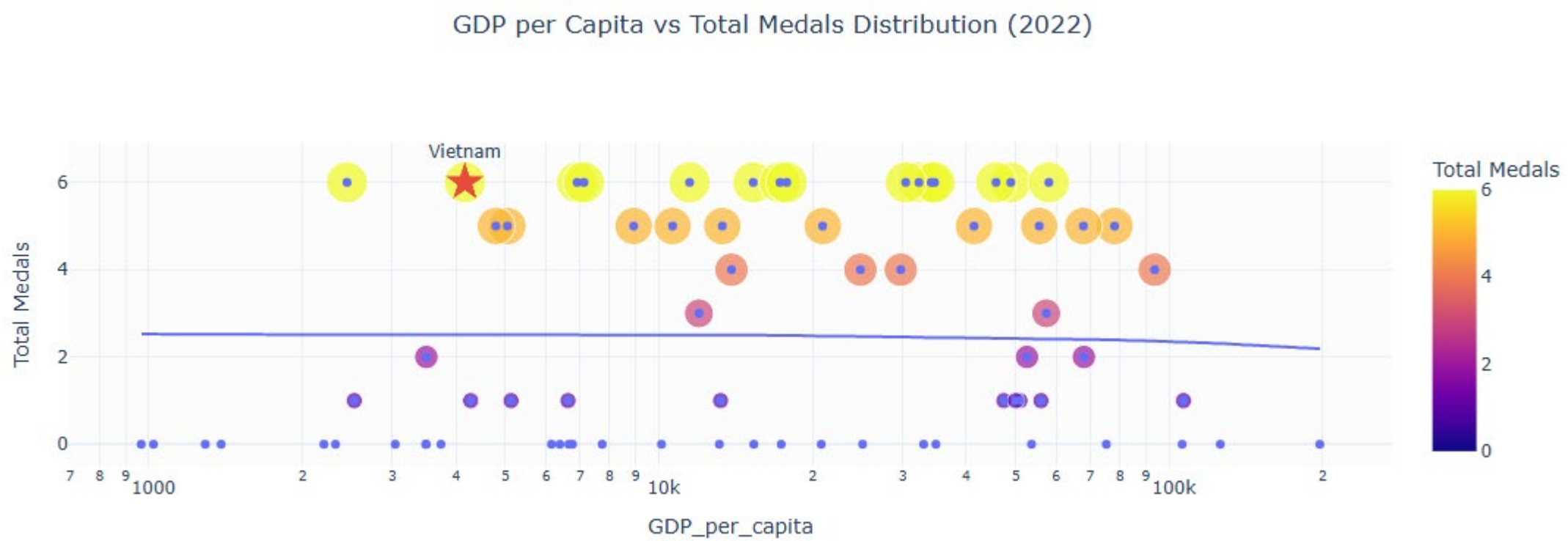
2. **Correlation with Socioeconomic factors?**

- We have added one more datasets about GDP of countries from 1965 to 2022.
- With this additional dataset, we are trying to find out whether a bigger economy correlated with more medals in the competitions.

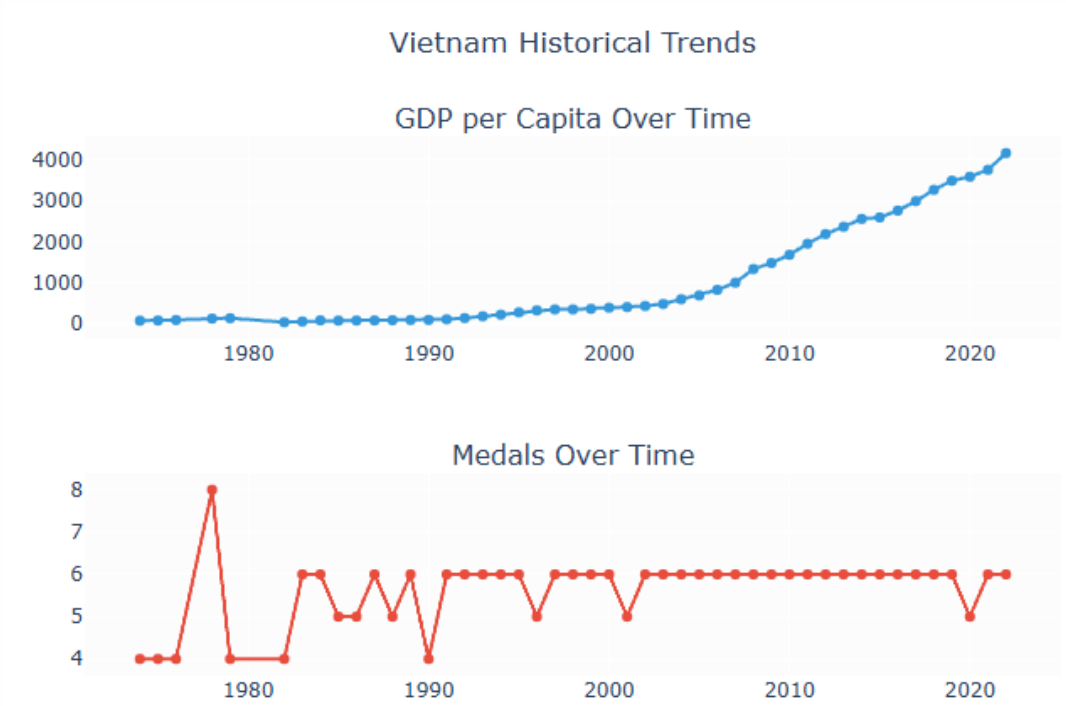
Different between GDP vs. GDP per capita

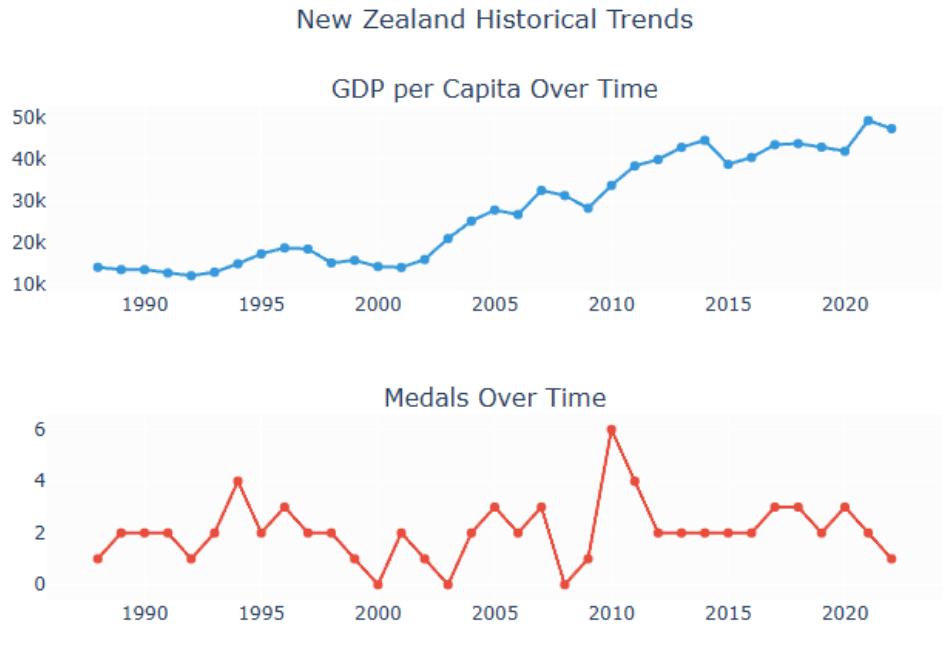


Different between GDP vs. GDP per capita

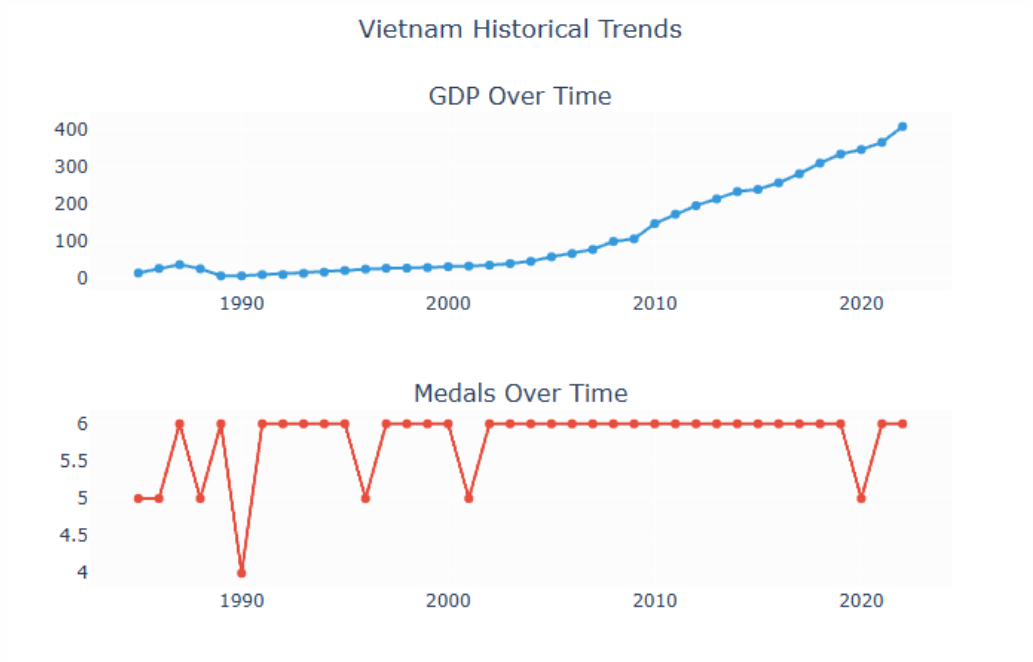


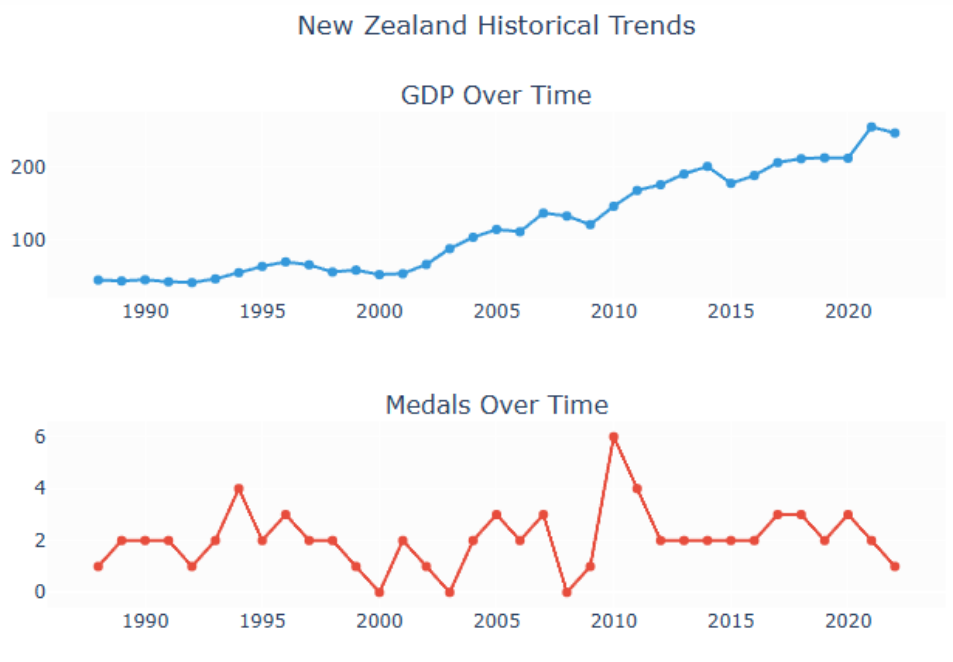
Does economic growth go along?





Does economic growth go along?





Conclusions

- Although IMO (as many other Olympiads) is very prestigious, as well as a good way for nurturing future talents
  - It does not show the stories of the whole socioeconomics pictures.
- IMO medalist number is **not a good indicator** for evaluating how good is a country either in education, or economics, etc.

Future work

- Tracking of IMO participants' educational and career paths after the competition. => Needs follow-up surveys or accessing anonymized career/educational datasets linked to participant cohorts.

- Link between specific national education policies and IMO performance. => Needs country-specific policy data (e.g., math curriculum changes, gifted program funding) and aligning it temporally with IMO results.

VINUNIVERSITY

Thank You!