Project 1

International Mathematical Olympiad (IMO) Data

Tri Phan, Cuong Pham

Agenda

DATA INTRODUCTION

QUESTION 1

QUESTION 2

Q&A

Data Introduction

International Mathematical Olympiad (IMO) Data is a publicly available dataset recorded the World Championship Mathematics Competition for High School students from 1954 to 2024.

There are 3 datasets

`country_results_df.csv`: Aggregated country-level performance data, including team size and medal counts.

`individual_results_df.csv`: Individual contestant performance and awards.

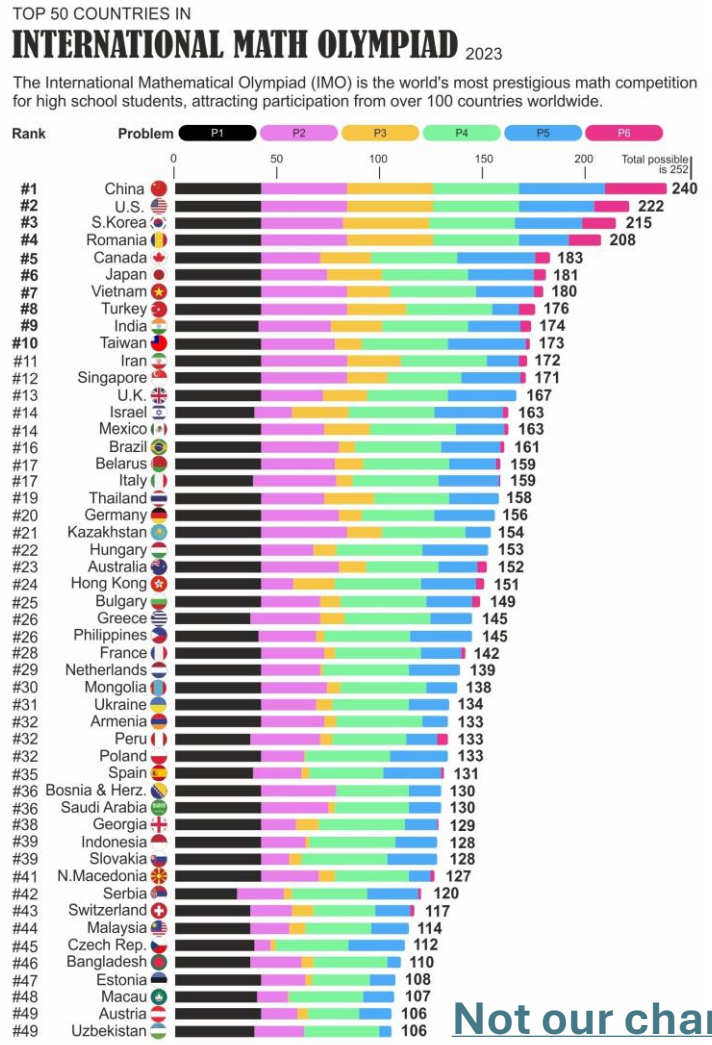
`timeline_df.csv`: Historical timeline of IMO events, including host countries and participant demographics.

Data Introduction

`country_results_df.csv`: Aggregated country-level performance data, including team size and medal counts.

- `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
- `awards_gold` (integer): Number of gold medals
- `awards_silver` (integer): Number of silver medals
- `awards_bronze` (integer): Number of bronze medals
- Others: p1 - p7, awards_honorable_mentions, leader, deputy_leader

Question 1. How do female participants' ratings influence countries' performance?



Not our chart

Data Introduction

`timeline_df.csv`: Historical timeline of IMO events, including host countries and participant demographics.

- `year` (integer): Year of IMO
- `country` (character): Host country
- `city` (character): Host city
- Others: edition, all_contestant , male_contestant, female_contestant, start_date, end_date

Question 2. How on-site temperature would affect the country-aggregated performance ?

Question 1. How do female participants' ratings influence countries' performance?

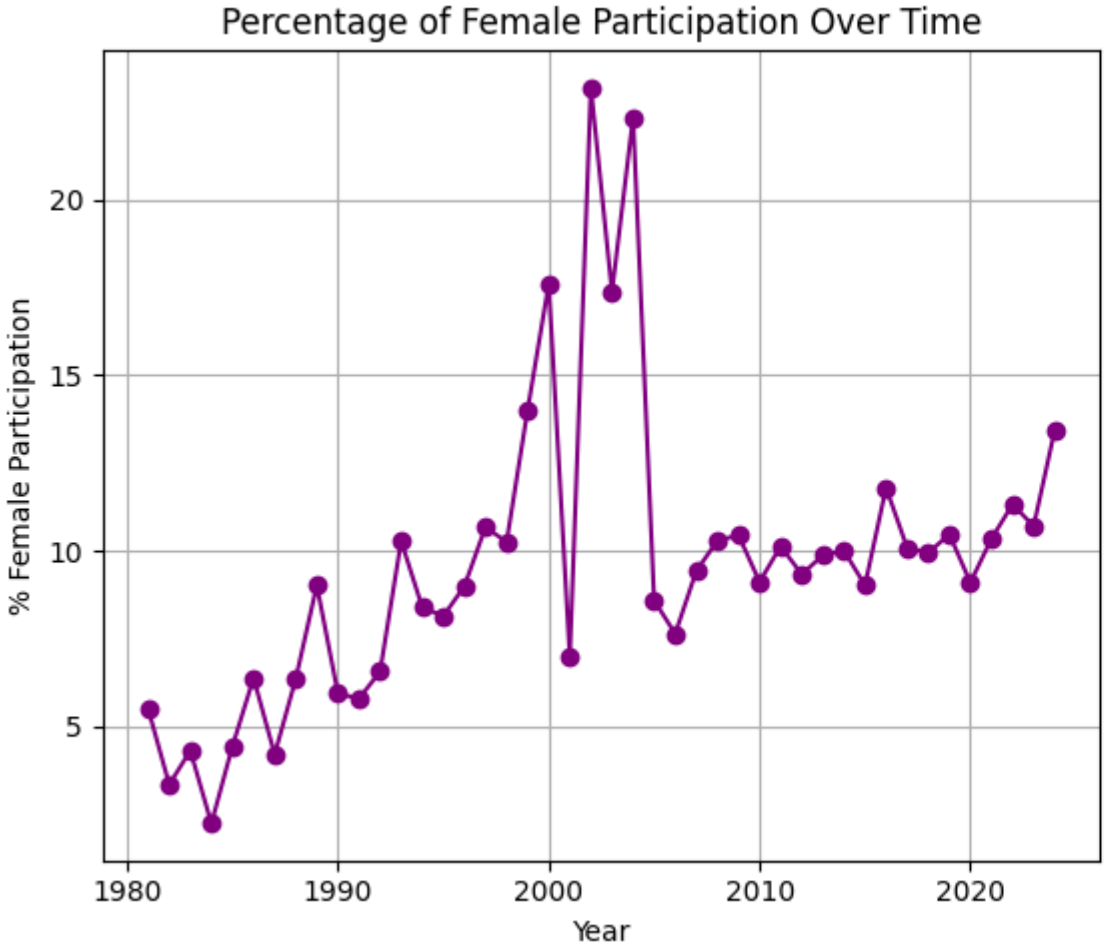
Preprocessing

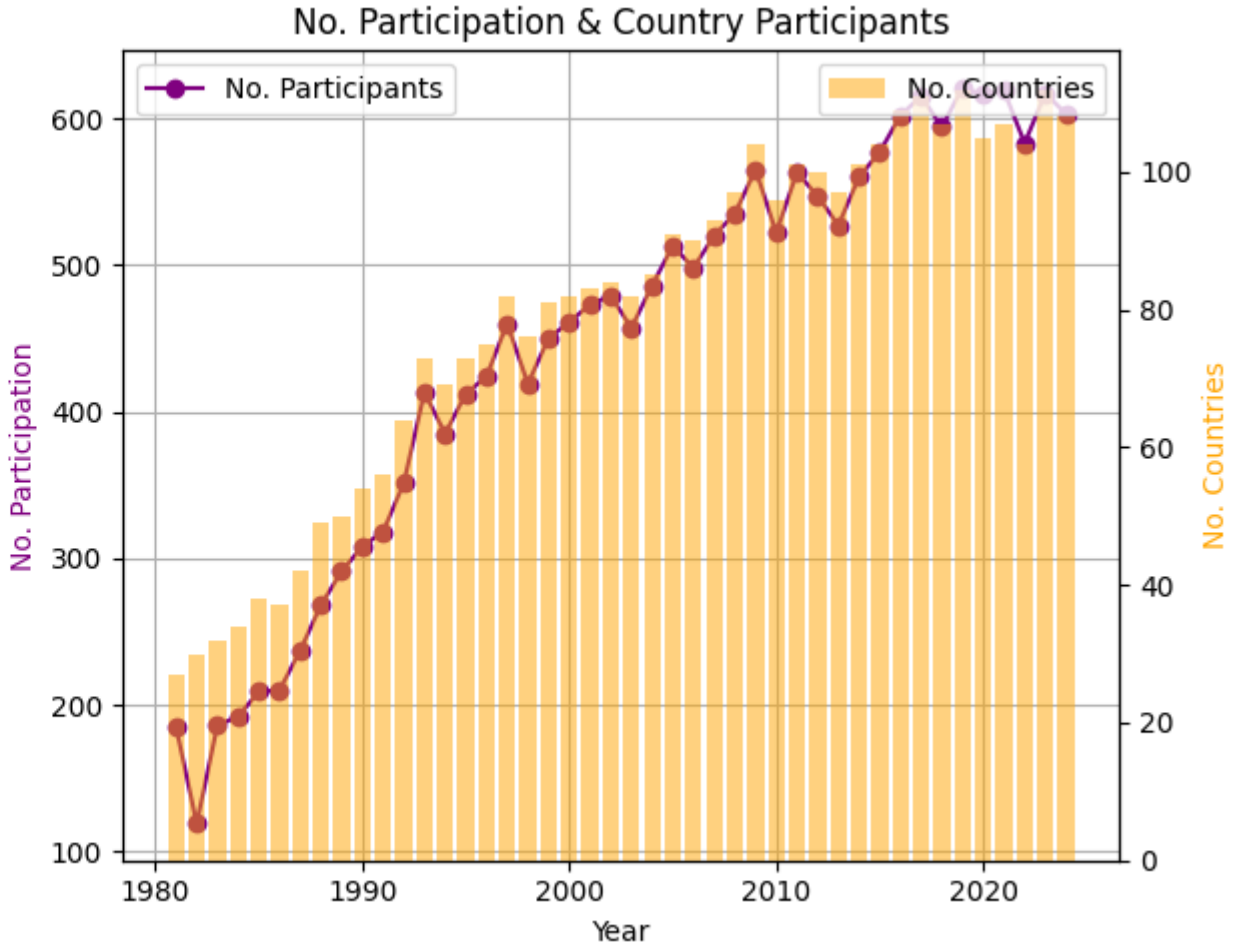
- Process data only from 1980.
- Handle missing data
  - If only `team_size_all` is provided, assume that all members are male.
  - If `team_size_female` is NA, fill with `team_size_all` - `team_size_male`

Aggregation

- Compute country performance by using a weighted sum of awards:
  - `country_performance = 3 * awards_gold + 2 * awards_silver + awards_bronze`
- Compute Female-Performance Ratio
  - `ratio = country_performance / team_size_female`
  - If `team_size_female = 0`, set `ratio = 0`.

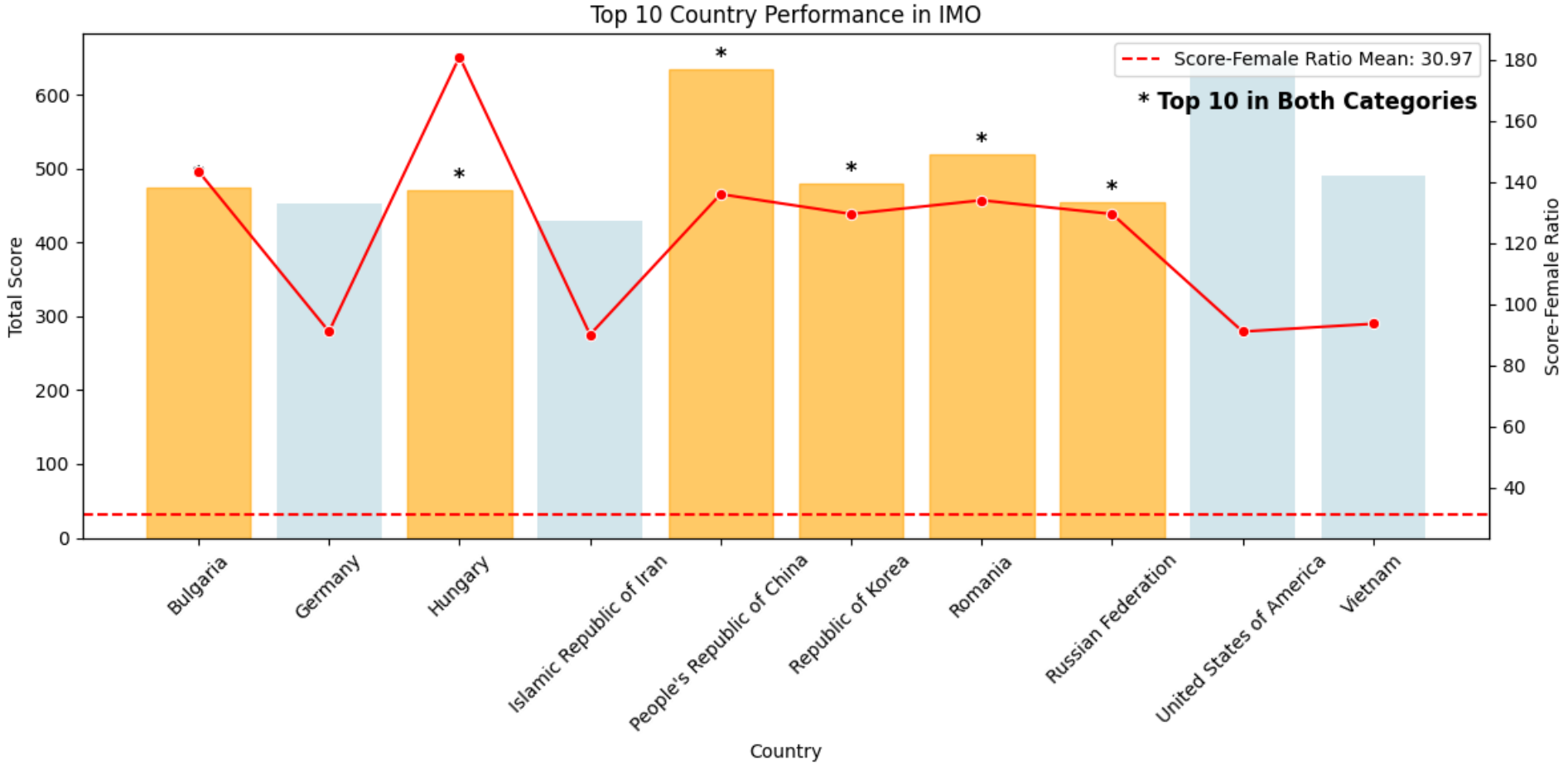
Question 1. How do female participants' ratings influence countries' performance?





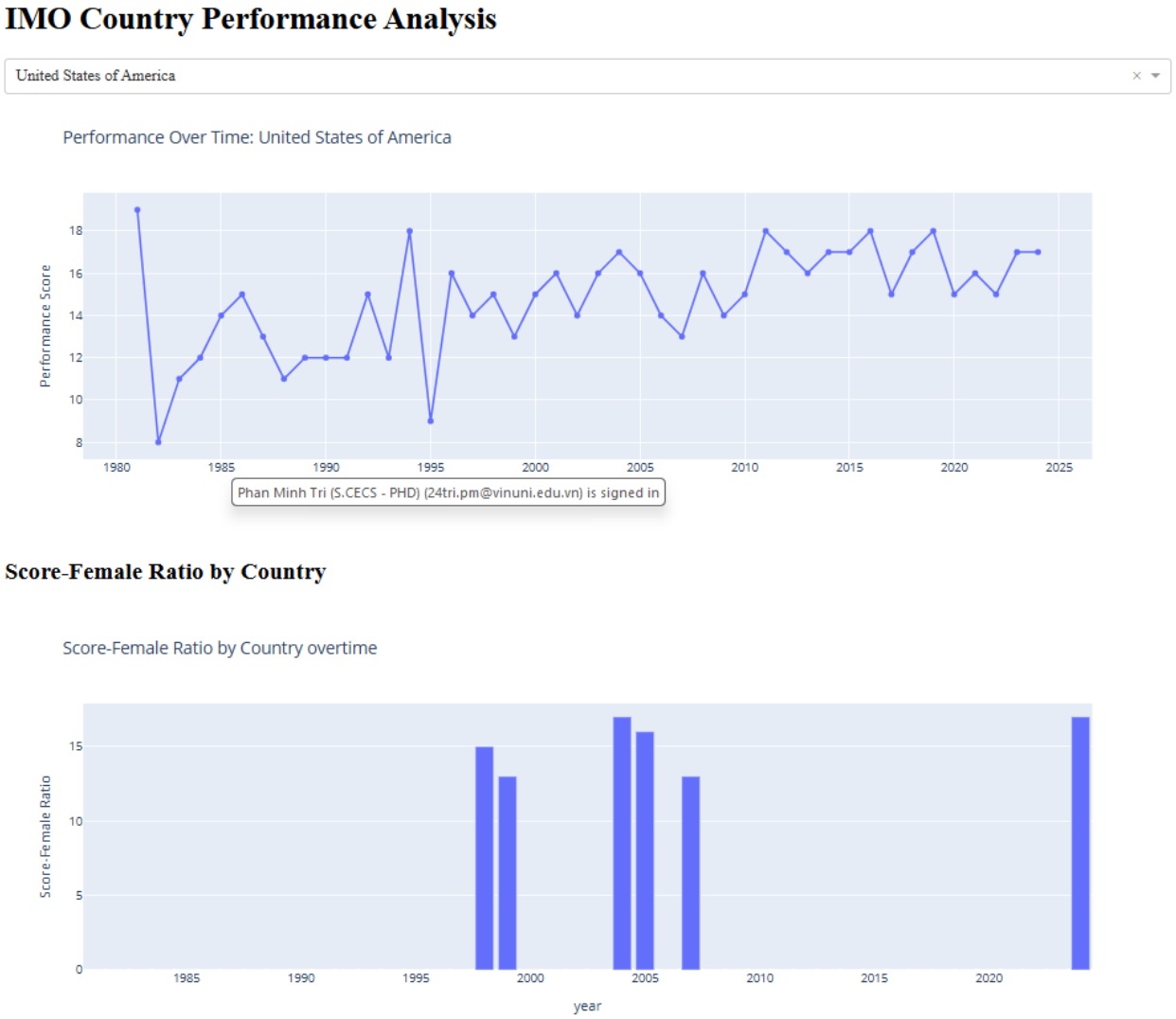
➔ Although the number of participants increased over time, the female participation rate has remained around 10% since 2005.

Question 1. How do female participants' ratings influence countries' performance?



➔ The participation of females does NOT affect a country's performance; high-performing countries maintain their performance.

Question 1. How do female participants' ratings influence countries' performance?



Interactive chart to view individual country performance and its Performance-Female Ratio over time.

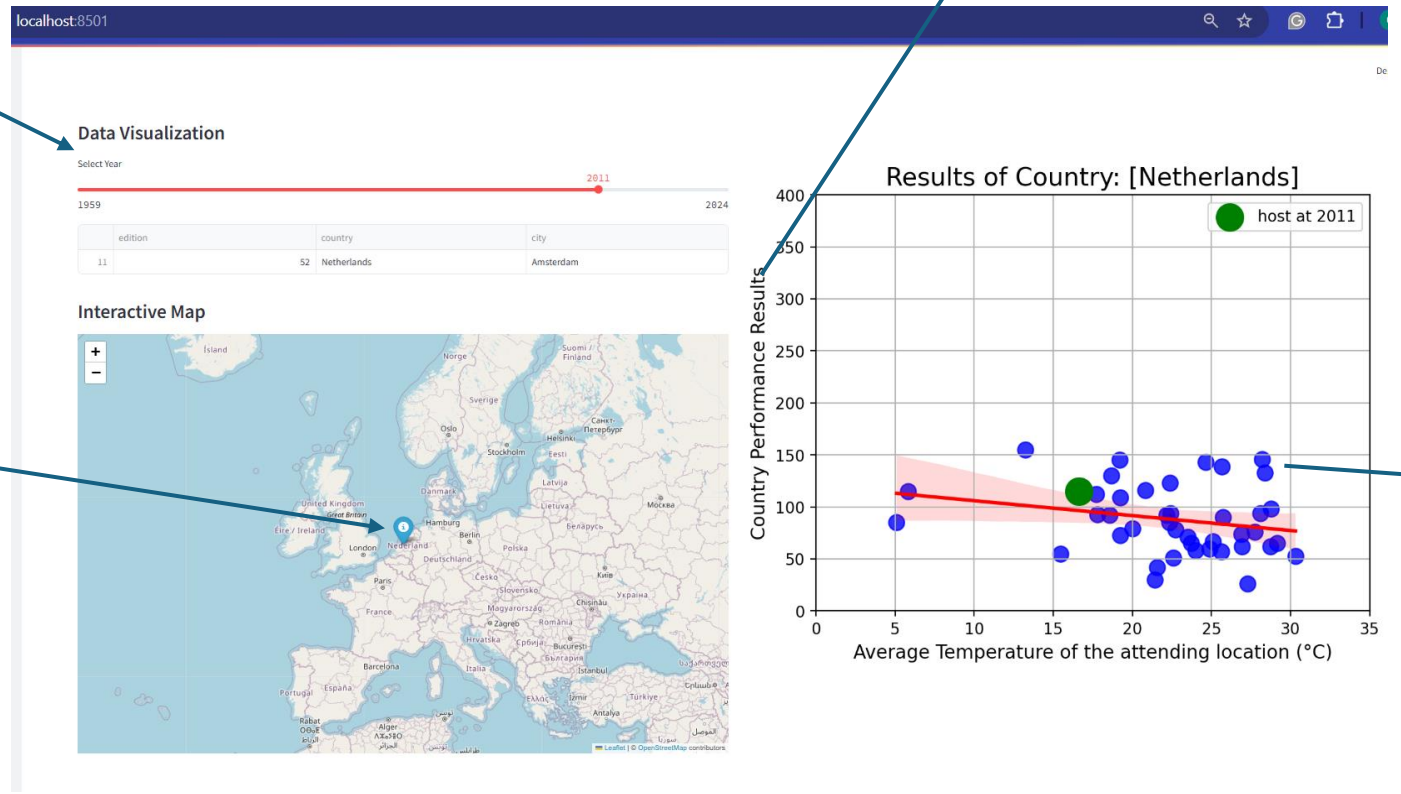
Question 2. How on-site temperature affect host country performance ?

Select the IMO "Year" (slider)

Country-level "scoring result" (total scores from P1 to P6)

The host city (map hover)

Data of all years that the country has participated



An **interactive visualization** exploring the relationship between host country performance and average temperature at IMO locations.

Question 2. How on-site temperature affect host country performance ?

Using external data:

(1) OpenMetéo Weather API to get the *day-specific temperature* given the city location;
(2) simplemaps.com to get the *location* (latitude and longitude) of the given city.

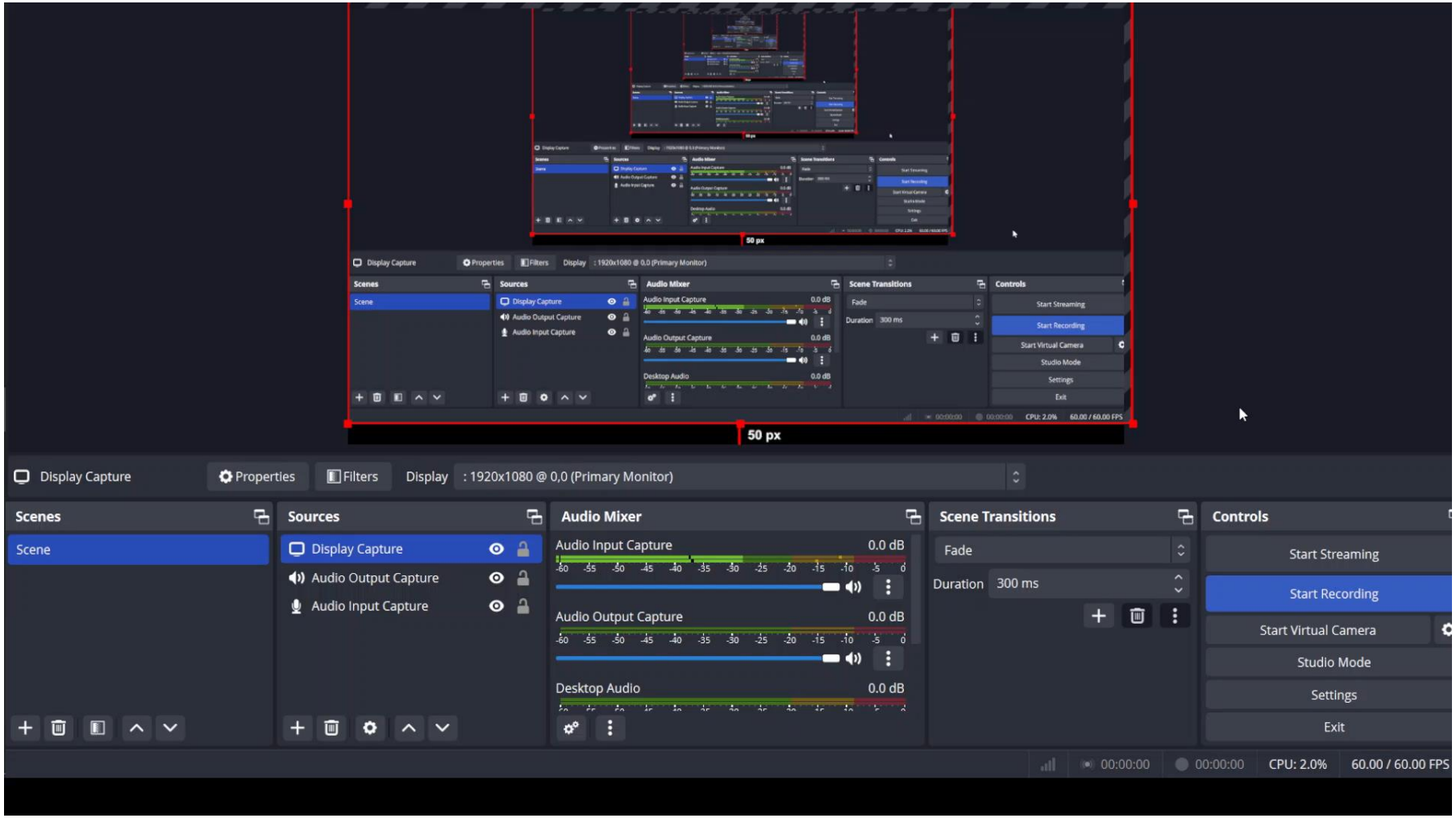
Analysis:

- Fetching: for each year, retrieve the *location* and *day-specific temperature* of the host city.
  - Align city & country name across all datasets.
  - Calculate average temperature across IMO event days.

- Cleaning:
  - remove all entries with none/missing values

- Scatter Plot:
  - Plot the *host-country's scoring result* & the *onsite-average temperature*,
  - (using all the year that the country participated)

Question 2. How on-site temperature affect host country performance ?



Video link

Tech stack *(streamlit, pydeck, folium, pandas, seaborn, matplotlib)*

Thank you!