120 Years of Olympic History

Presented by Gianna Burgess

TABLE OF CONTENTS

INTRODUCTION

This project analyzes 120 years of Olympic history (1896-2016) using athlete and NOC region datasets. The goal is to analyze trends, demographics, medal distribution, the impact of physical attributes on Olympic success, and overall participation.

Data Cleaning

Merging Datasets

The two raw datasets were merged for better analysis.

Duplicate Handling

1385 duplicate entries in athlete participation in same events were dropped.

Type Conversion

Season and Year columns were converted into categorical data.

Missing Value Imputation

Missing values were imputed in the age, weight, height, and medal columns.

New Feature Creation

BMI and age group columns were created for more specific analysis of the dataset.

Exploratory Data Analysis

01 Scope

There are 135,571 unique athletes, 205 countries, and 66 unique sports represented in the data.

02 Gender participation

Male participation has historically dominated, but the gender gap is closing.

03 Popular Sports

Gymnastics and athletics are the most popular in terms of participation.

04 Distribution of Medals

Nearly equal distribution of medal types.

Exploratory Data Analysis

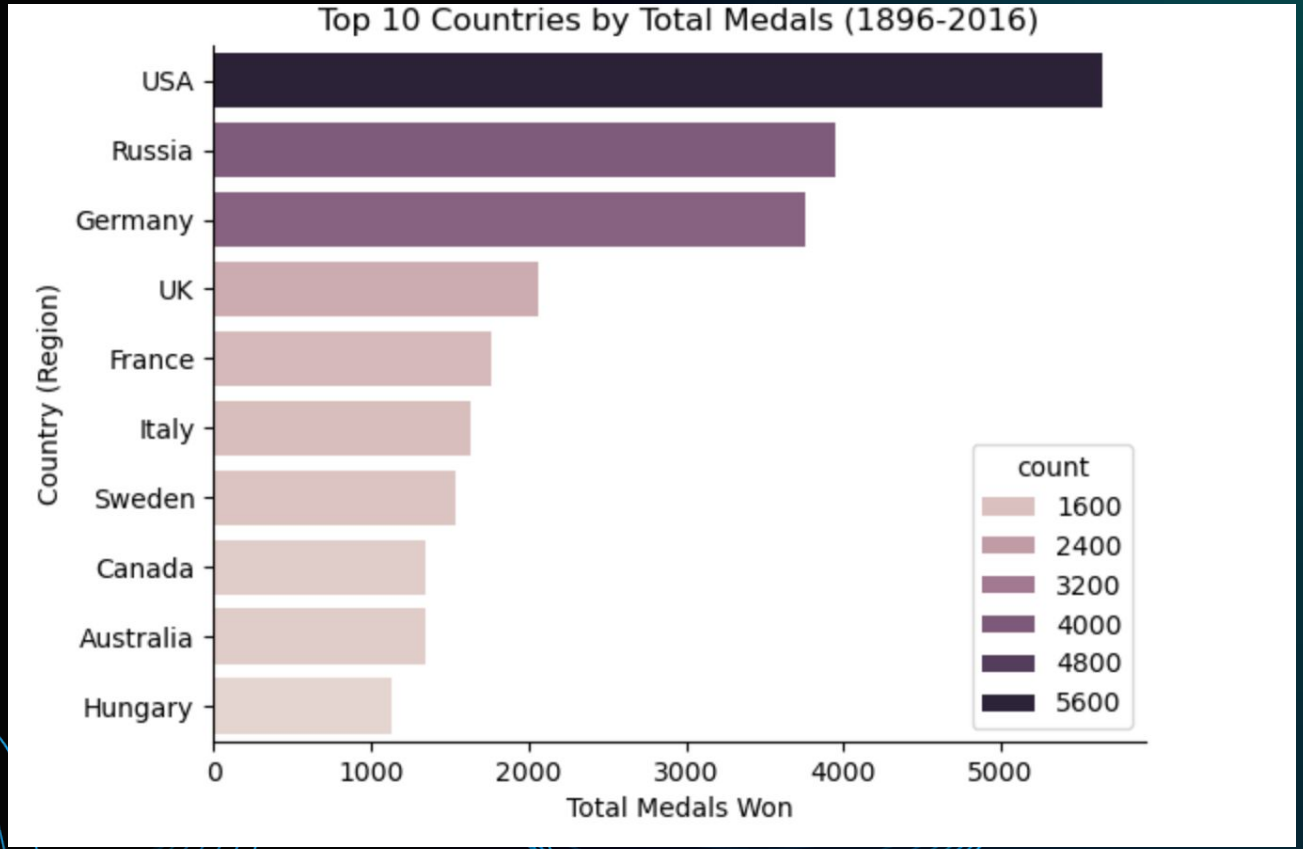
01 Overall Olympic Participation

Olympic participation has increased from the first games in 1896 to the last one in 2016.

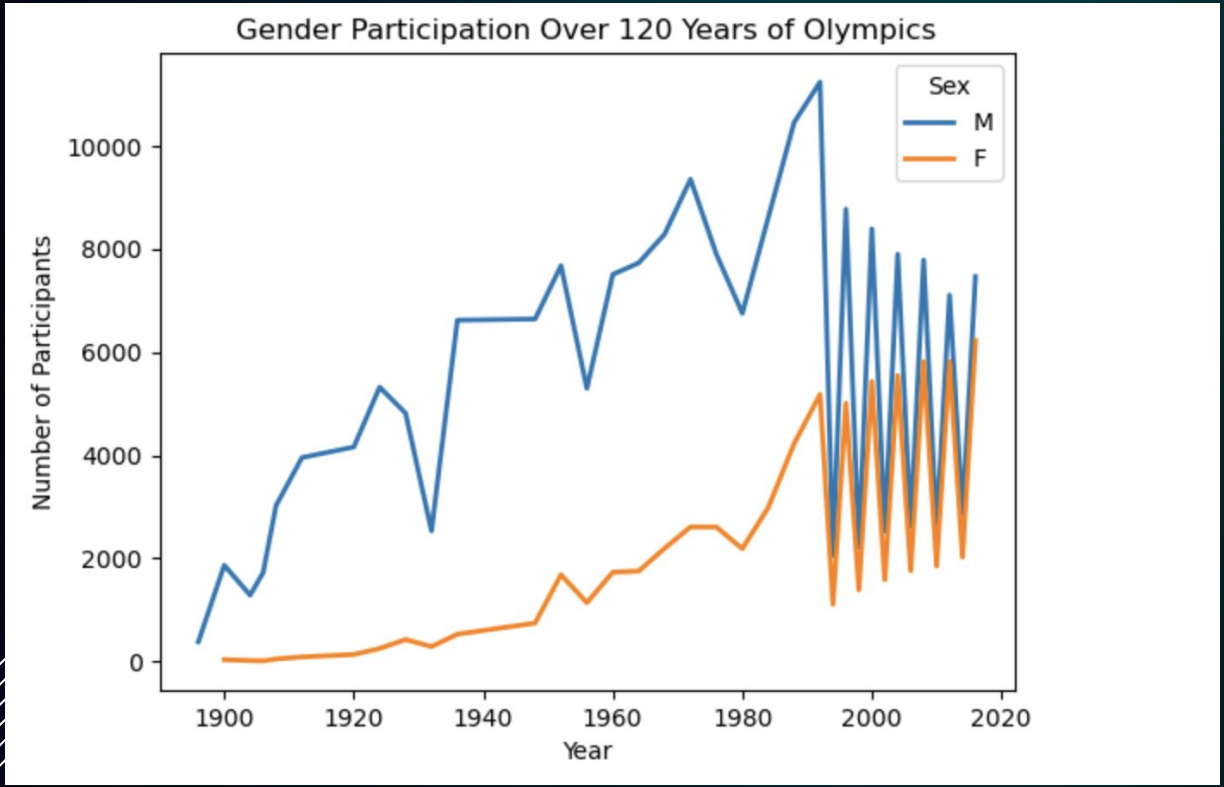
02 Physical Attributes of Athletes

The median athlete is 25 years old, 178 cm tall, and weighs 73.5 kg.

Top 10 Countries by Total Medals

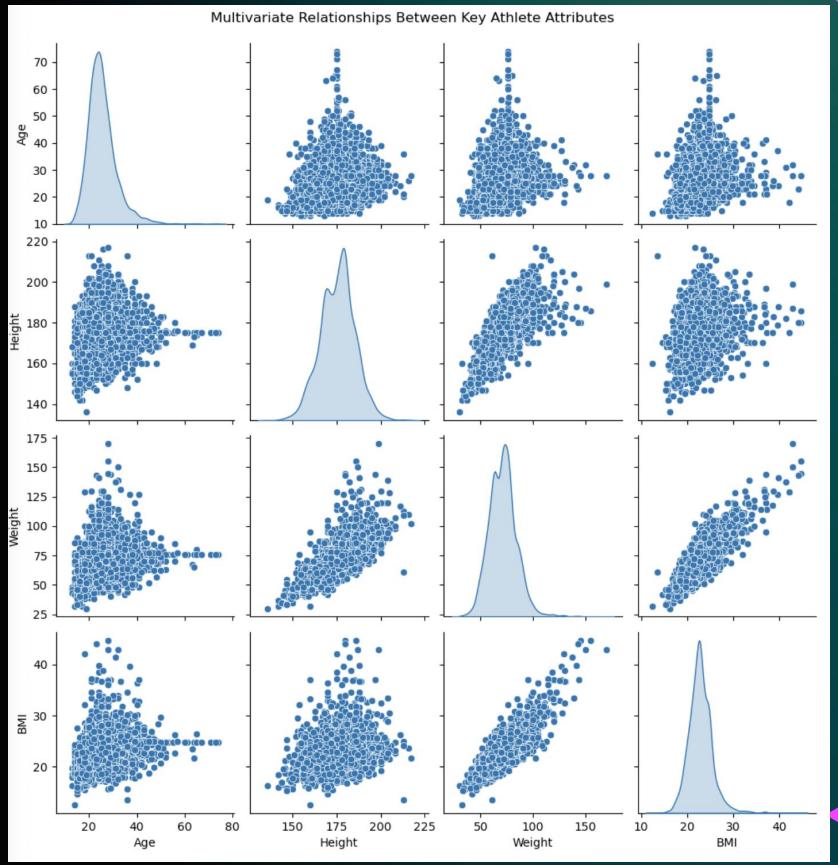


Gender Participation Over Time

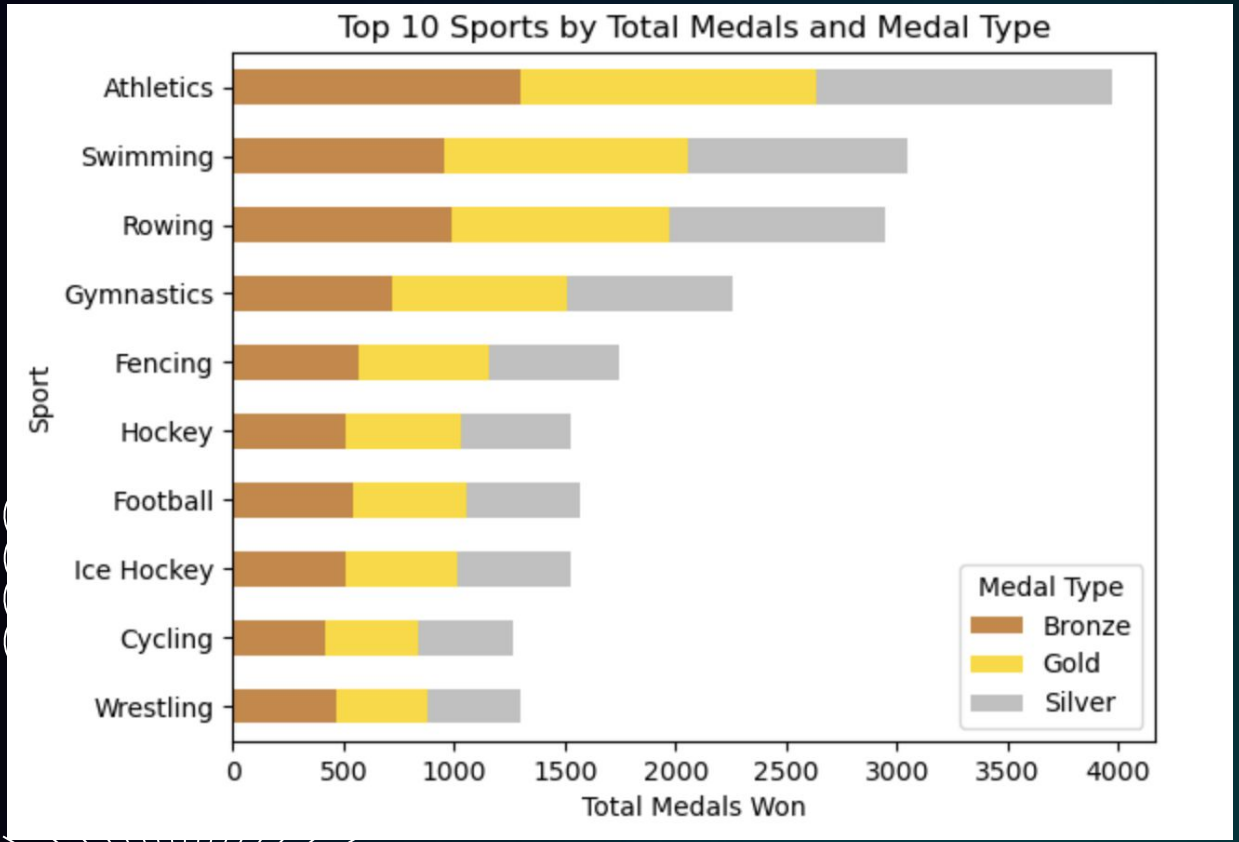


Age Distribution by Medal Type

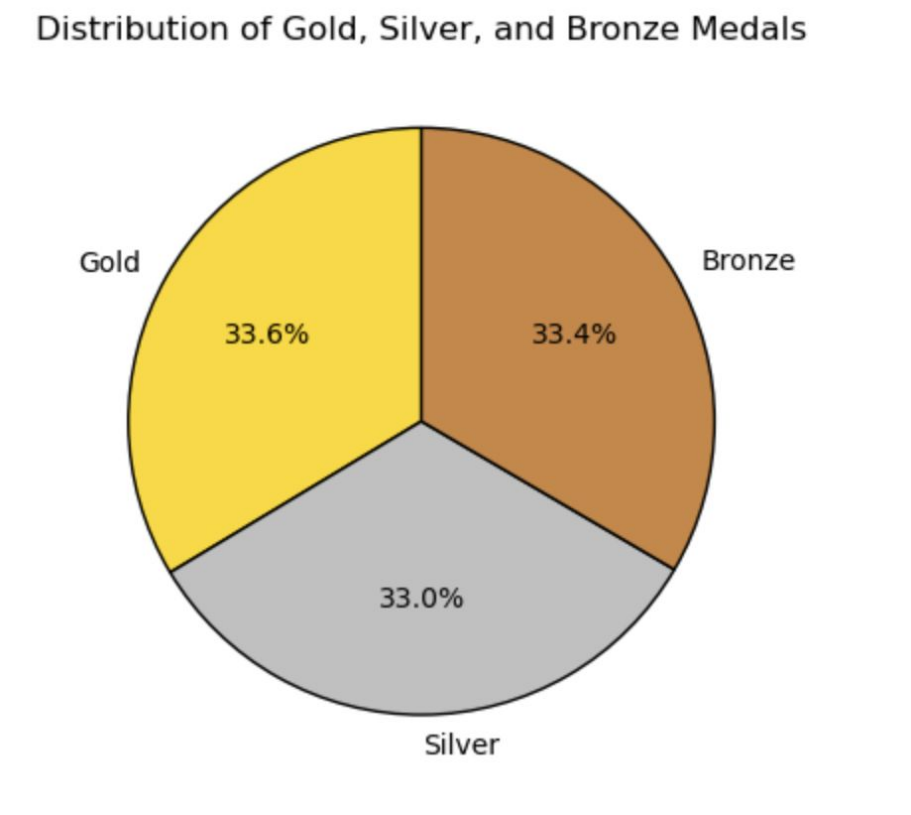
Pair Plot of Physical Attributes



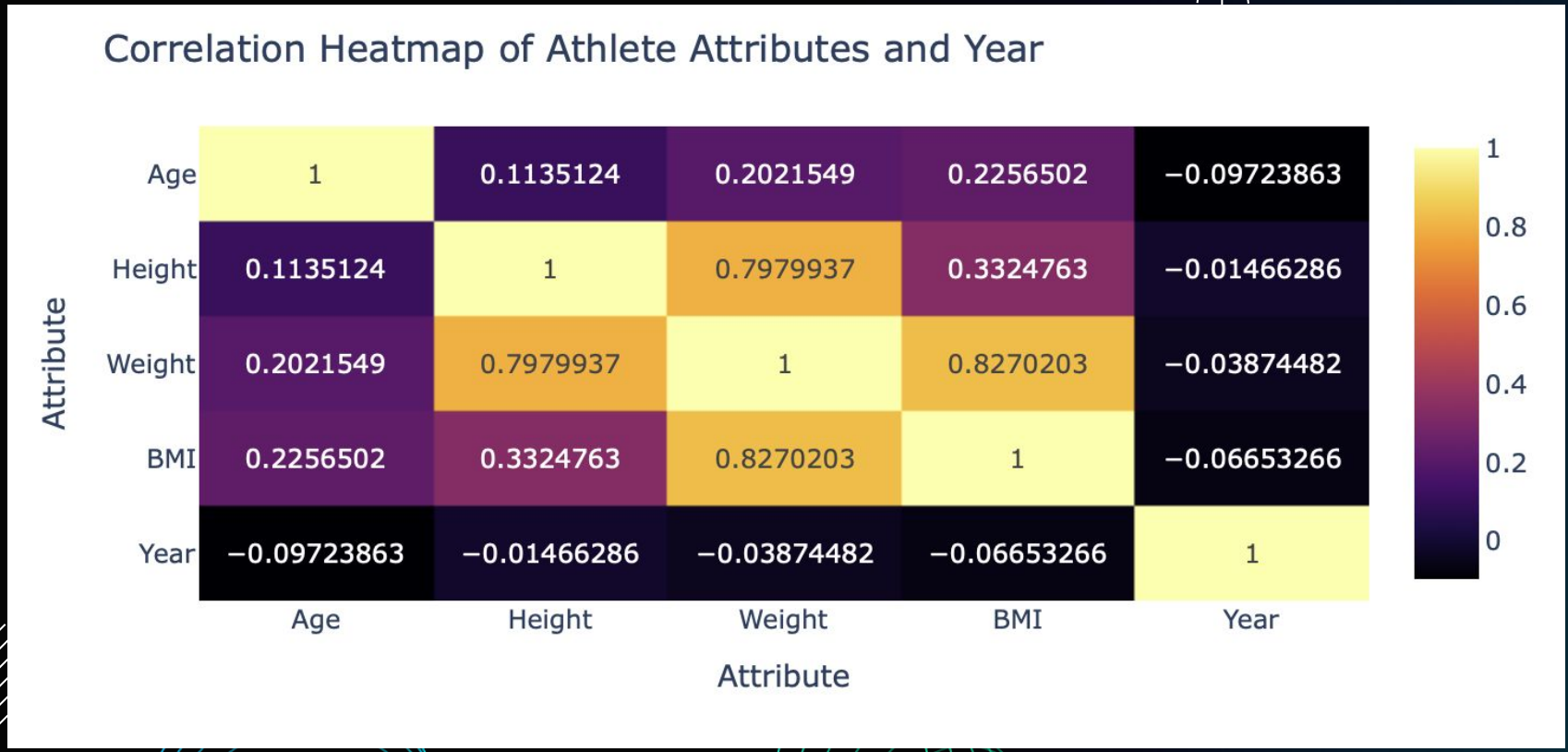
Medal Count by Sport



Medal Distribution



Heatmap of Correlation



Insights

Medal Dominance

- USA dominates
- Sports with more events dominate

Participation Trends

- Significant increase in participation
- The average athlete is not bigger than in past athletes

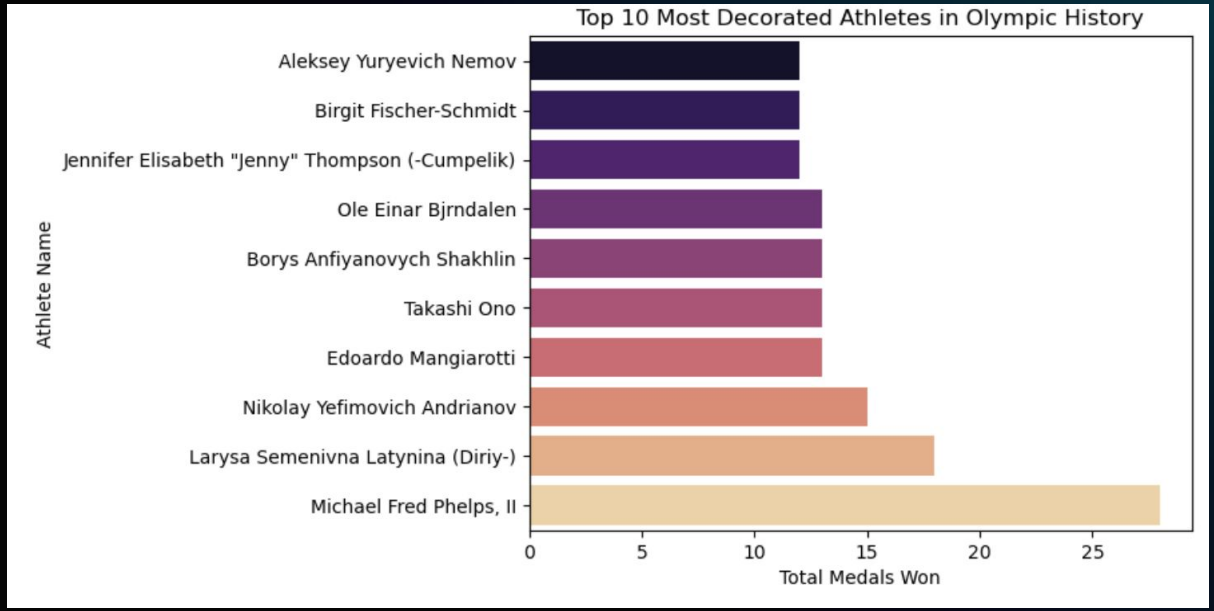
Gender, Age, and Physical Attributes

- Medal success peaks at 25
- The gender gap is closing
- Medalists are bigger than non-medalists

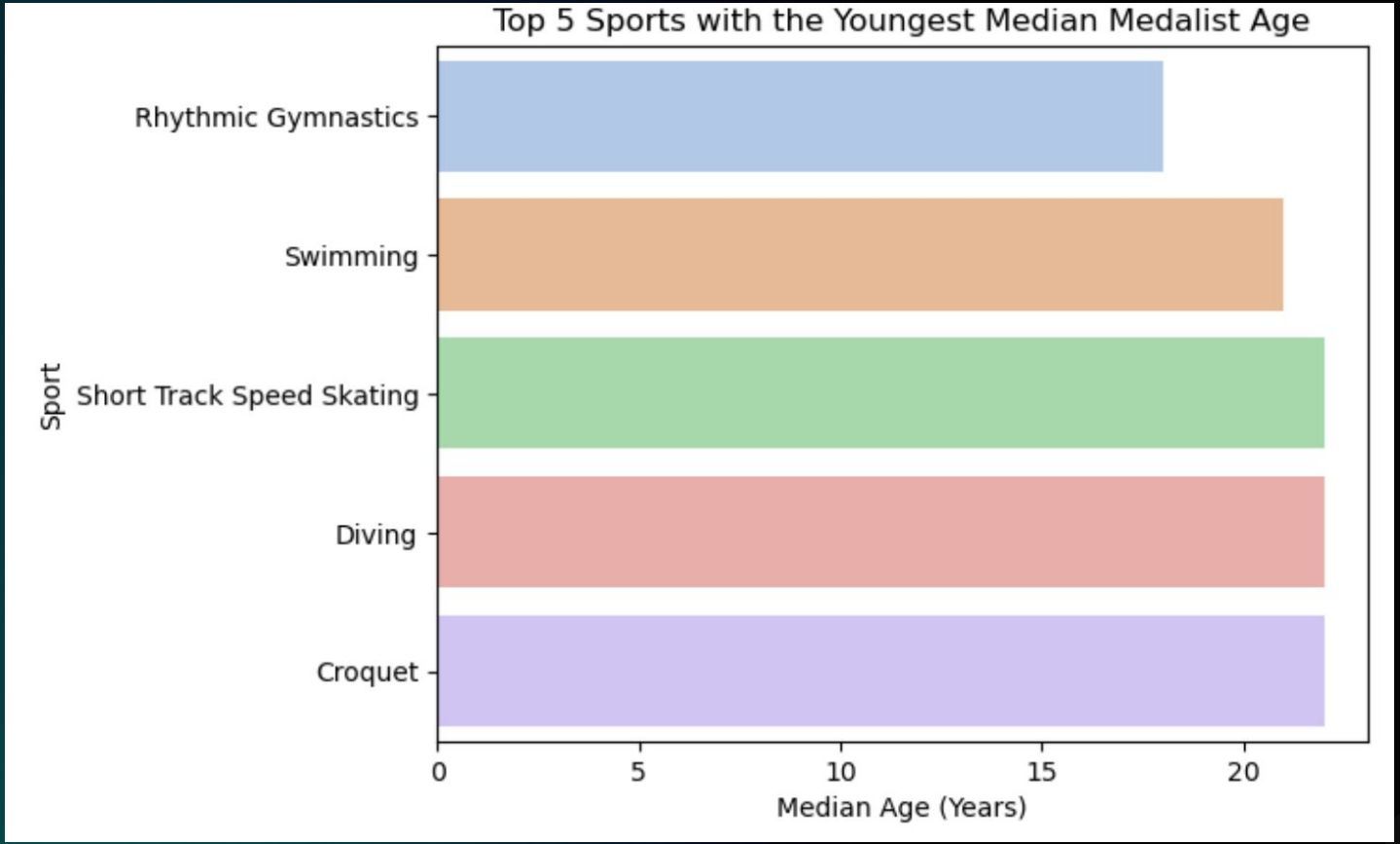
Limitations

- NaN values
- Structural bias
- Total medal count favors certain countries and sports

Who are the most decorated athletes?



Which sports do young people dominate?



Do host countries win more medals?