Visualizing the Olympics

Visualizing the Olympics

Code

INFO 526 - Fall 2024 - Final Project

This project leverages historical data from Summer Olympic Games to uncover insights into trends in athlete characteristics and country-level performance.

AUTHOR
MintMindsForDataViz

AFFILIATION
School of Information, University of Arizona

Abstract

This project analyzes data from past Olympics to explore patterns and trends in athlete performance and demographics, focusing on basketball and country-level performance. The first question investigates the evolution of basketball players' heights over time, utilizing a ridge plot to visualize changes and a logistic regression model to assess how height influences the probability of winning a medal. Results reveal significant trends in player height and its relationship with competitive success. The second question examines how countries' performance has shifted across Olympic Games, with particular attention to the potential influence of major world events. Together, these analyses provide insight into the interplay between individual attributes and global factors in shaping Olympic outcomes.

Introduction and Data

For our project, we have chosen to analyze Olympic data. Our motivation is to learn more about the evolution of athletes over time and which countries have dominated the games. The dataset used comes from Kaggle and describes characteristics of Olympic athletes such as name, gender, country, age, height, weight, event, and medal if applicable for every Olympic athlete from Athens 1896 to Rio 2016.

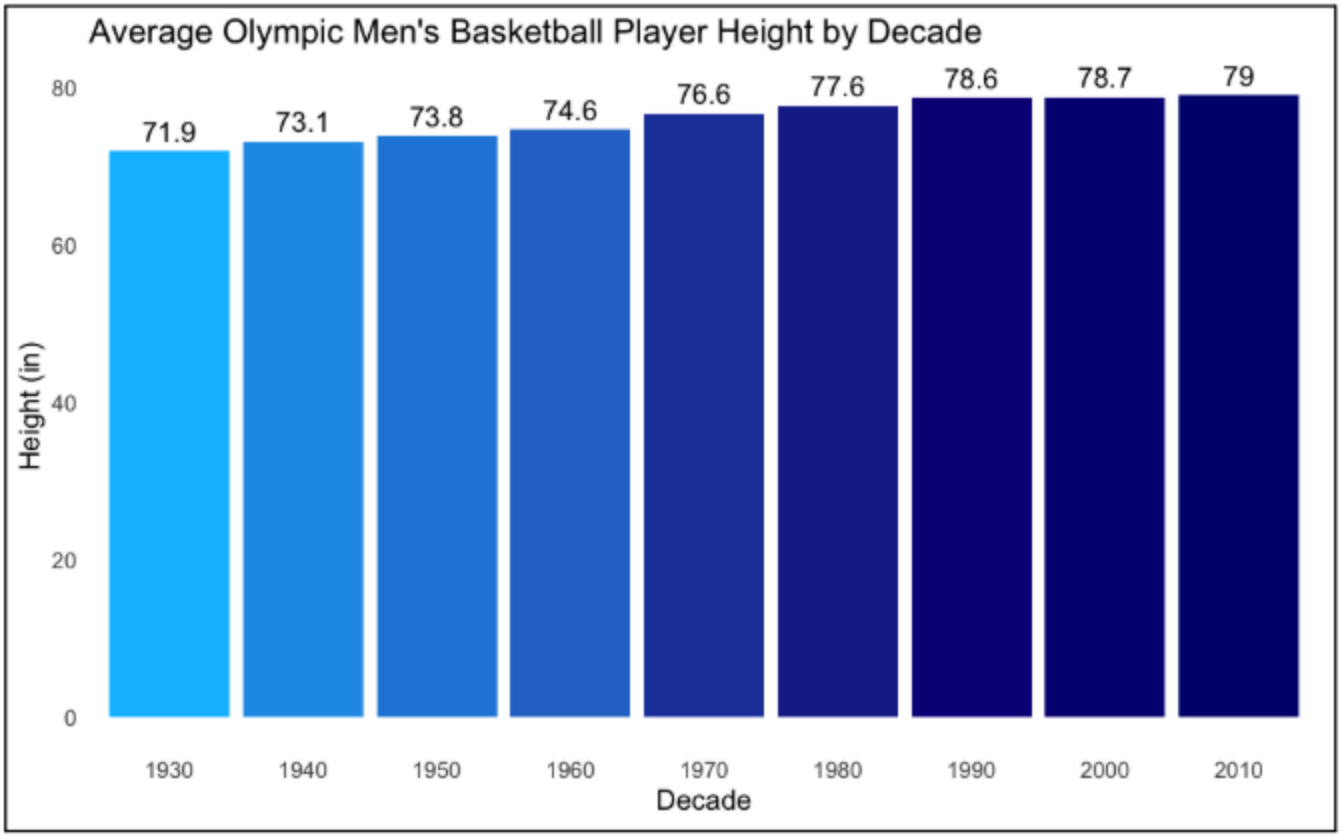
To learn more about the evolution of athletes our first research question states: Is there any ideal height for Olympics men's basketball and has it changed over time? To analyze the performance of different nations our second question states: How has the performance of different countries evolved and what major events influenced changes?

To start answering our first question we conducted exploratory data analysis and created visualizations to explore the heights of Olympic men's basketball players over time.

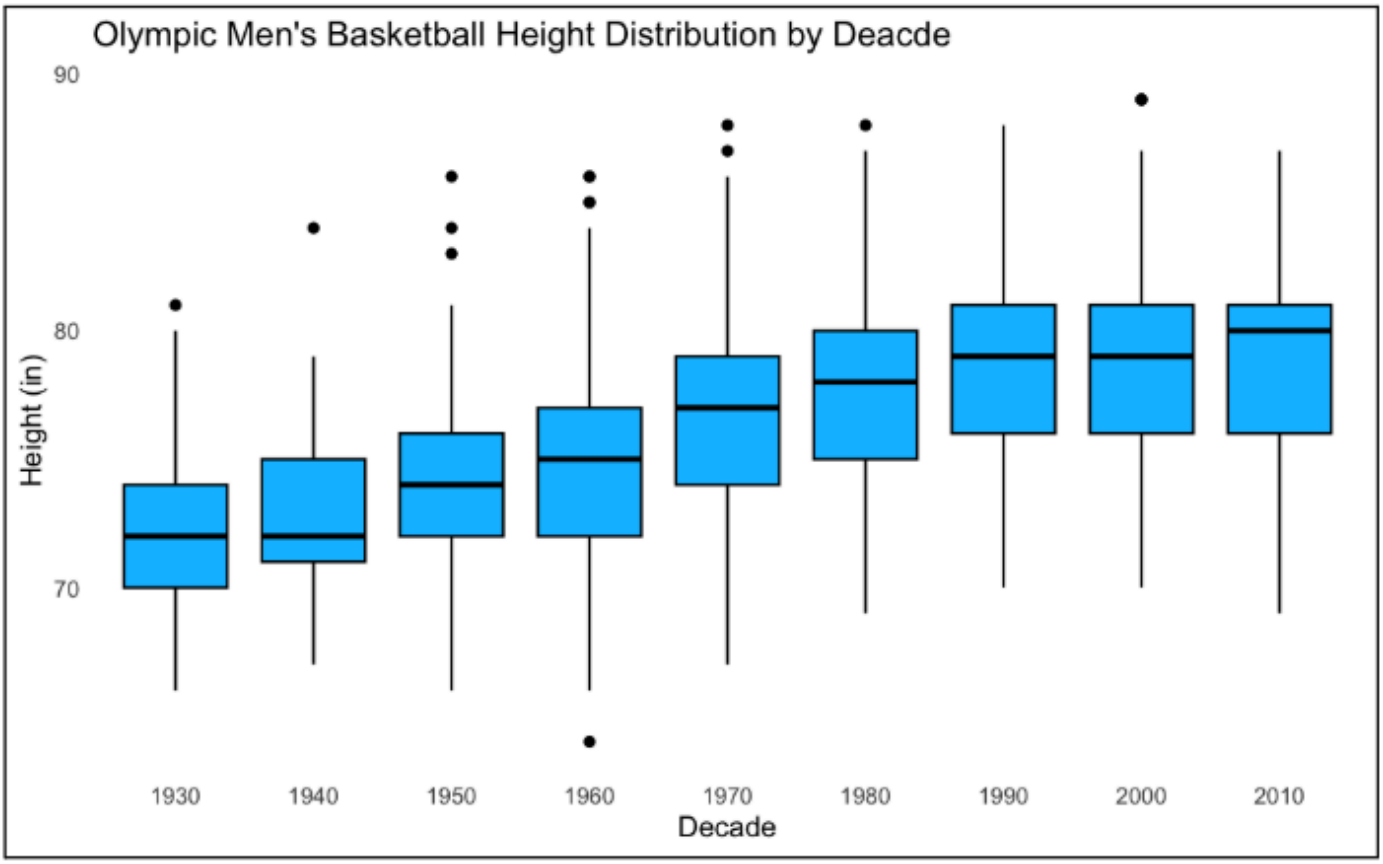
To create the plots shown below and the ridgeline plot used to help answer our first research question we created a column displaying the height in inches by dividing the height in centimeters column by 2.54. A

decade column was also created which grouped all Olympic games into their corresponding decade. A few basketball players' heights were missing, so all athletes missing height entries were removed.

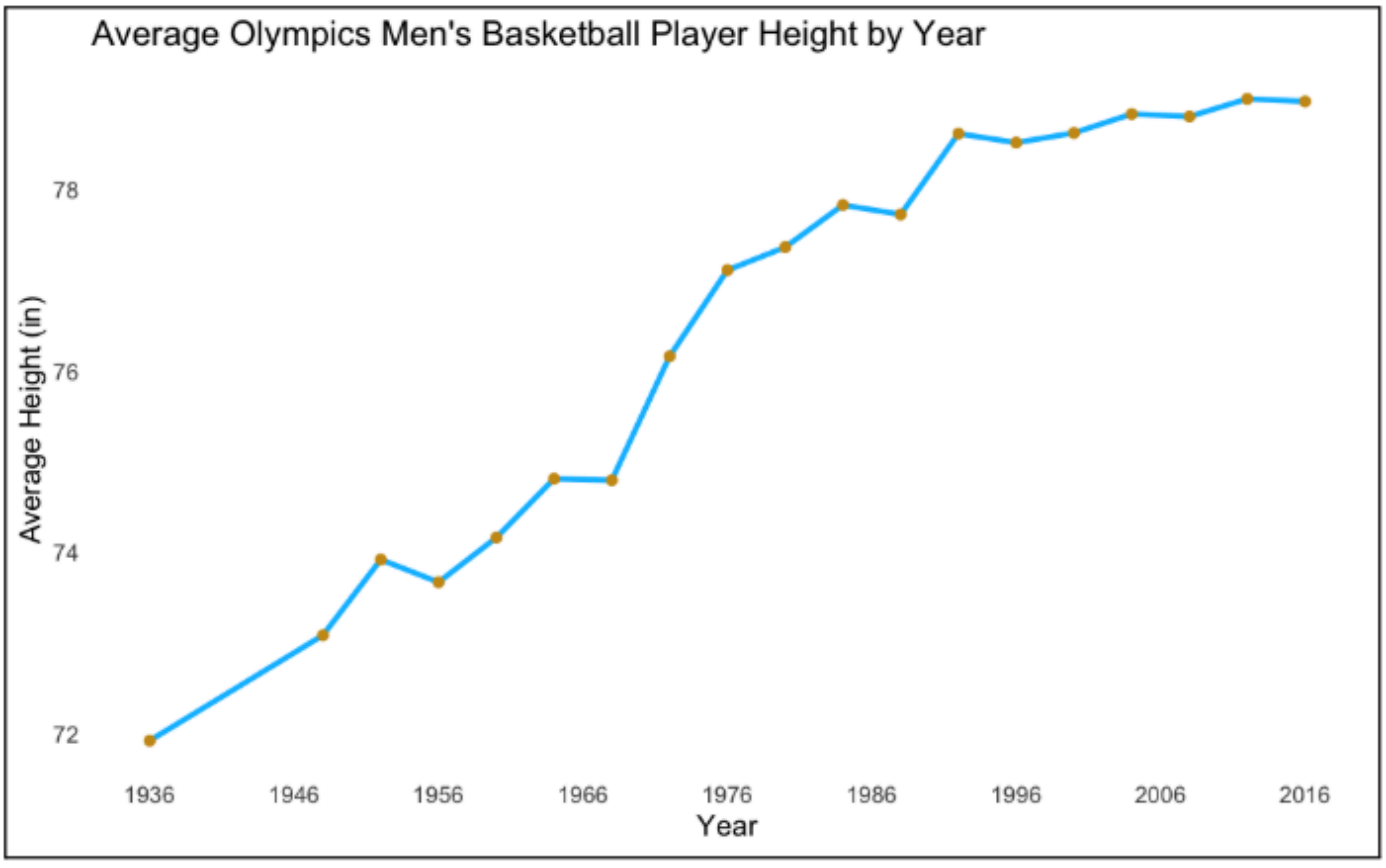
Knowing that the average height throughout all of Olympics men's basketball is 77 inches, we created the following visualizations:



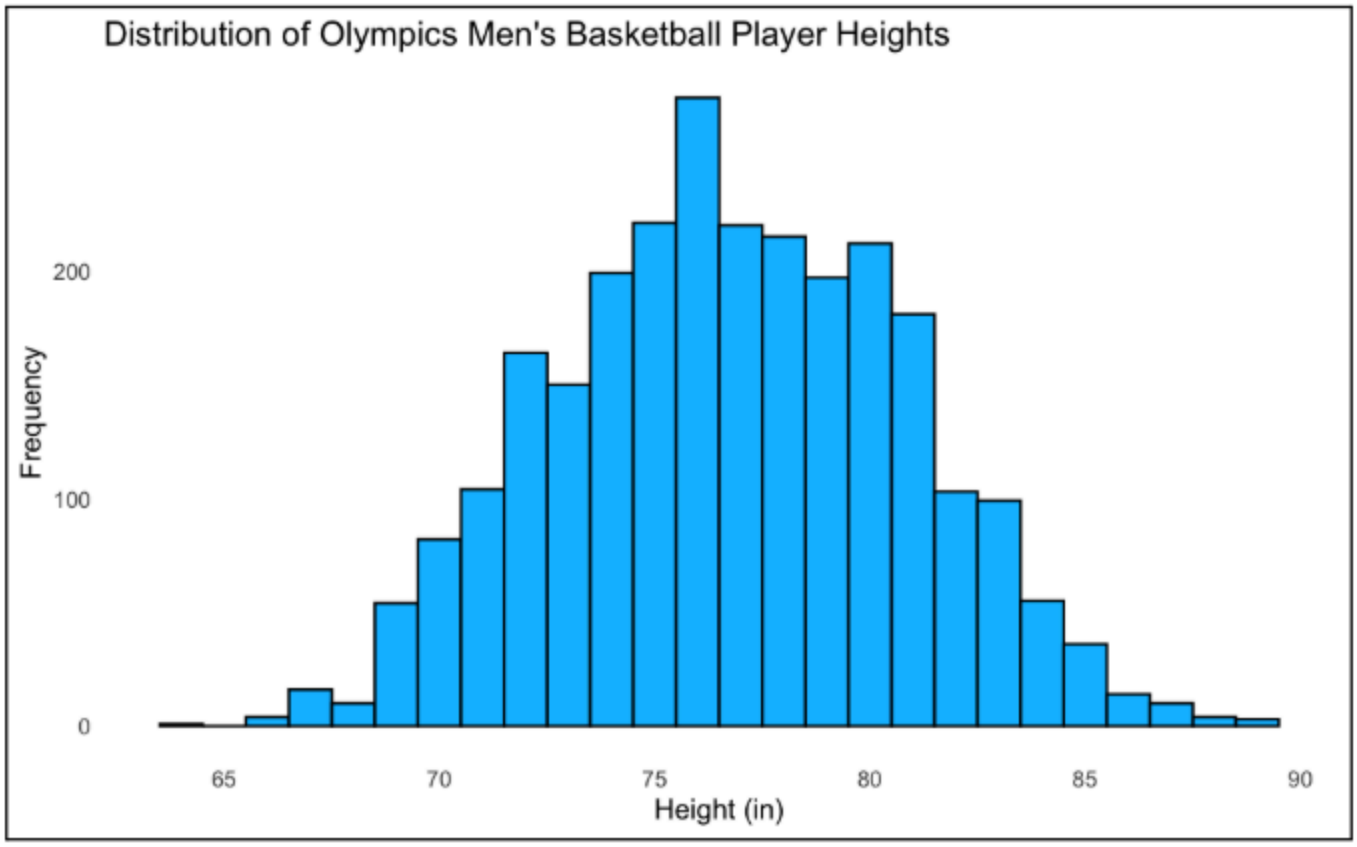
One thing to note is that basketball did not start in the Olympics until 1936. Based on the visualization above, it is clear that the height of Olympic men's basketball players has increased by just over 7 inches in these 80 years.



These box plots show a similar story to the bar plots above. While heights increased from 1930 through 1980, by 1990 the growth had slowed.



The line plot above displays the average height of Olympic Men's basketball athletes since 1936. Now we can see in more detail that player heights reached a new peak in 1992 and have stayed fairly consistent since then.



The final exploratory visualization for Olympic men's basketball player heights confirms that the average height between 1936 and 2016 is around 77 inches, or 6 feet 5 inches. It also shows the small frequency of players below 70 inches (5 feet 10 inches) and above 85 inches (7 feet 1 inch).

Methodology

The visualization used for displaying Olympic men's basketball players' heights over each decade is a ridge-line plot. A ridge-line plot makes it easy to see how the most common Olympic basketball player heights have increased by showing the changes in height distributions for each decade. To build the ridge-line plot we used the created height in inches and decade columns.

The methodology used to answer our first research question is a logistic regression model to see the probabilities of winning a medal based on height. The model will not take into account the type of medal won, but simply if a medal was won or not since a logistic regression model is used for binary classification. To create the model, we created a binary column showcasing a 1 if a medal was won and a 0 if not. We then created a data frame consisting of only heights and whether a medal was won or not. All NAs from the height column were dropped and this data frame was then split into training and testing data for the logistic regression model.

To answer our second research question which focuses on the distribution of medals by country, a few transformations were done. First, a data frame (medal data) showing the count of each type of medal for each country from each Olympics is shown. For example, starting in 1896 we can see the distribution of bronze,

silver, gold, and total medals for each country up until 2016. Next, we had to identify the top 10 countries leading in total medals and filter so that only those countries appeared in our medal data. Finally, we created a data frame consisting of events that affected Olympic participation and the years that they occurred, so these years could be plotted.

To showcase medal trends over time, we created a line plot displaying the total medals for the top 10 countries. The plot also showcases the year's major world events occurred, which can explain some fluctuations in medal counts. We chose a line plot because it effectively displays countries' changes in medal counts over time and allows for a simple display of the world event years.

We also created an area chart illustrating the distribution of gold medals won by top countries. An area chart was chosen because it highlights the changes in gold medal dominance across countries and the overall growth of the Olympic games in terms of total gold medals.

A dashboard was created to showcase all of these visualizations in one place. Although not showcased on the dashboard, in our presentation we have included an animation of medal trends over time for the top 10 countries in total medal count.

Results

1. Olympic Men's Basketball Player Heights

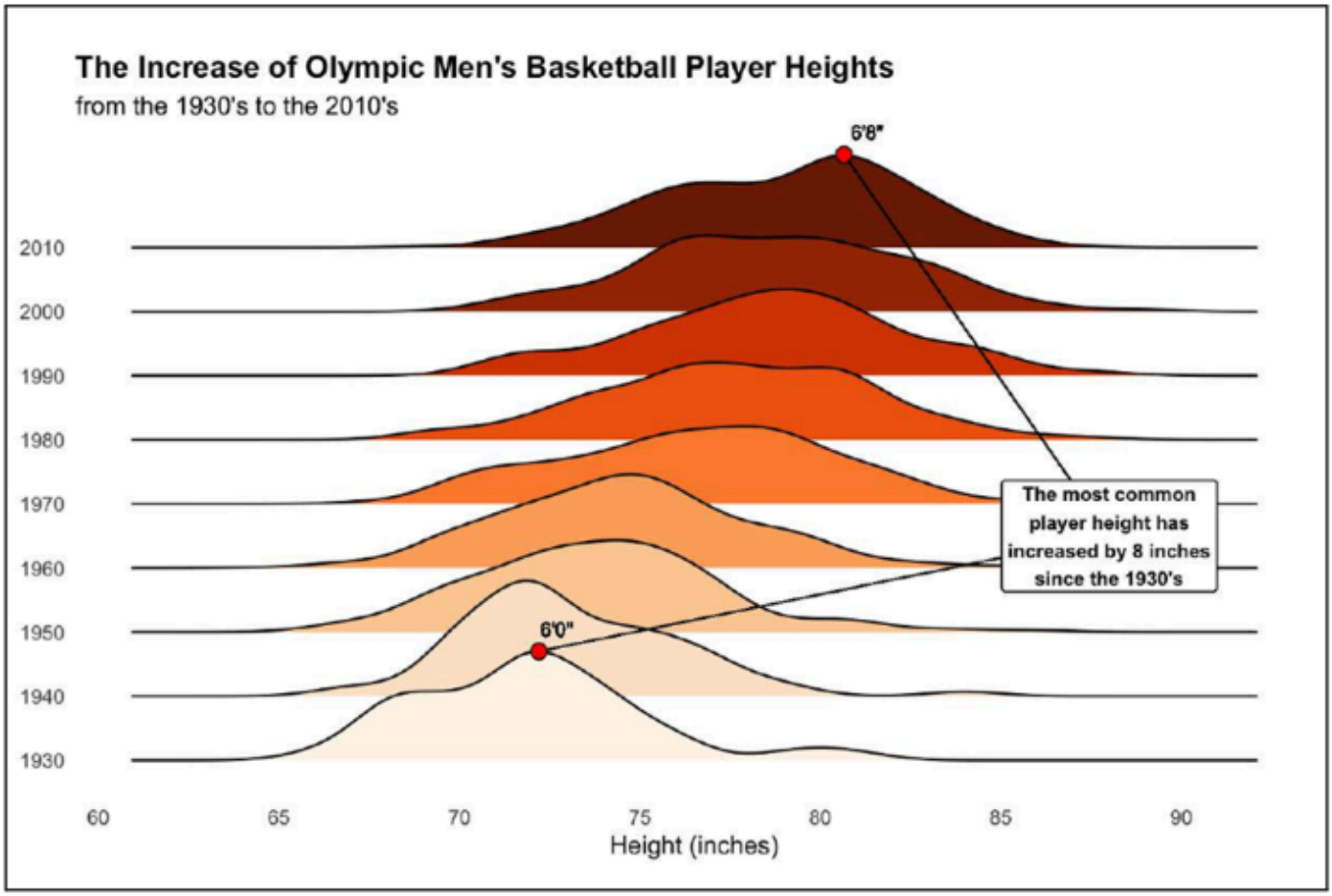
Insight: The most common or average height for Olympic men's basketball players has considerably increased over the years

Key Findings:

- In the 1930's, the common height was 6'0". By the 2010s, it had increased to 6 '8".
- This increase highlights the evolution of player selection criteria and the growing emphasis on height as an advantage in basketball.

Visualization:

The distribution of heights of men's basketball players from the 1930s to 2016 is displayed in the ridge plots showing the upward shift in height distribution across decades, with annotations emphasizing the shift.



2. Height and Medal-Winning probability

Insights:

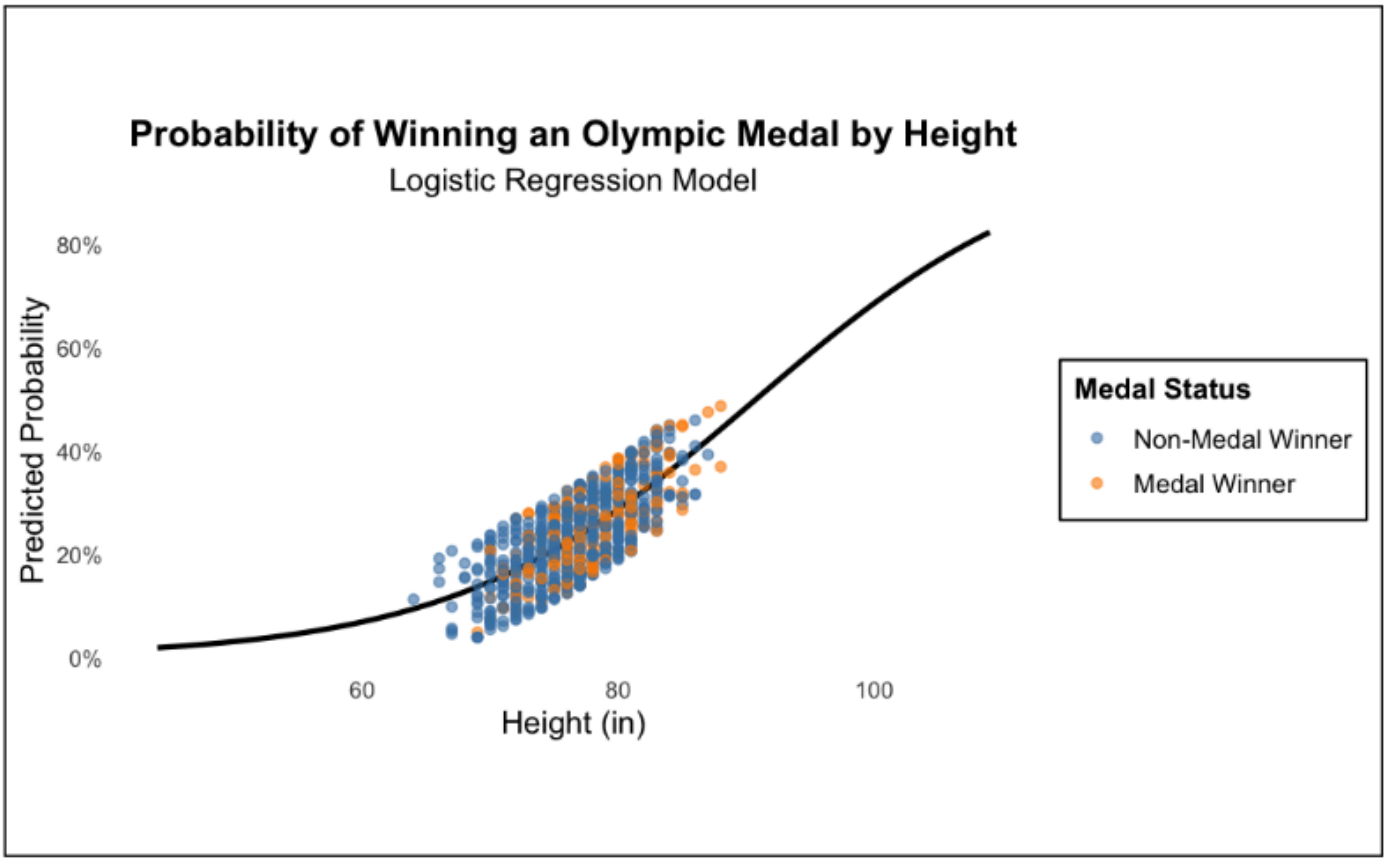
- Height is a significant factor in predicting winning probability in men's basketball, with a P-value of 6.05e-10.
- The odds of winning increase by 8.85% for every one-inch increase in height.
- Despite the statistical Significance, the AUC of 0.5973 indicates limited predictive power when height is used as the sole or major predictor

Visualization Highlights:

- The logistic regression curve shows the relationship between height and the probability of winning a medal.
- The jittered data points introduce a little nuance to the visualization, demonstrating the model's predictions alongside the observed data.

Supporting insights:

The moderate AUC score illustrates the need for a multifactorial approach to athlete evaluation beyond physical phenotypes.



3. Medal Trends over time (top 10 countries):

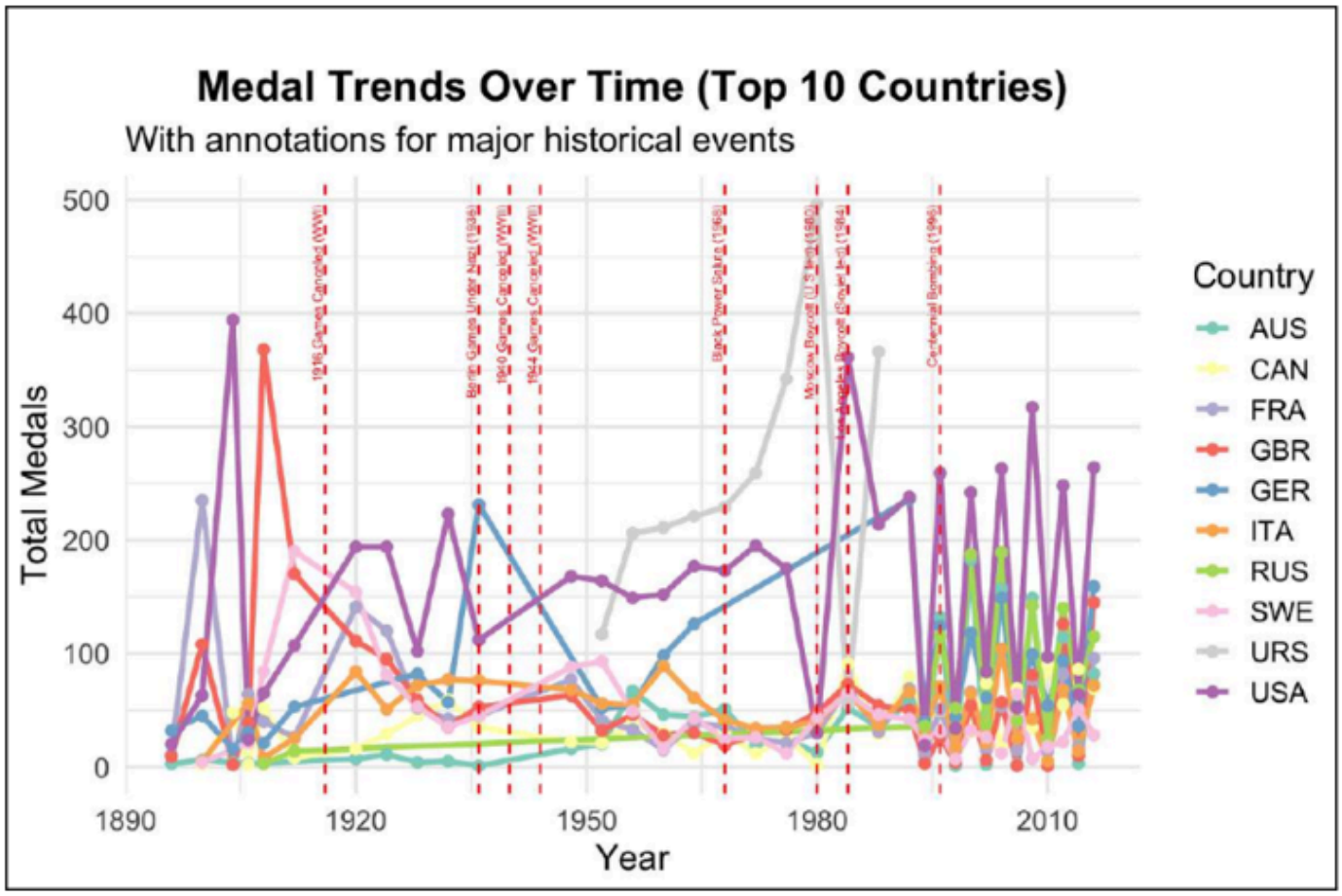
Insight: Medal trends of the top 10 countries reflect a longitudinal dynamic of consistency in the performance of these nations, fluctuations in competitive investment, and changes in focus in coverage of various sports.

Key Findings:

- The line plot shows how the total medals won by the 10 countries were distributed along the timeline.
- Some countries are growing steadily and may remain at the transcendence of investing in athletic performance.
- Changing levels of competition or participation over decades are suggested by differences in medal counts.

Visualization:

Line Plot: The plot represents the trajectory of total medals obtained by the top countries, in an unmistakable way to identify the periods in which they grew their performance and when there were some dips. The visualization shows how dominant countries have managed to keep their edge while others have surged and fallen.



4. Gold Medal Distribution:

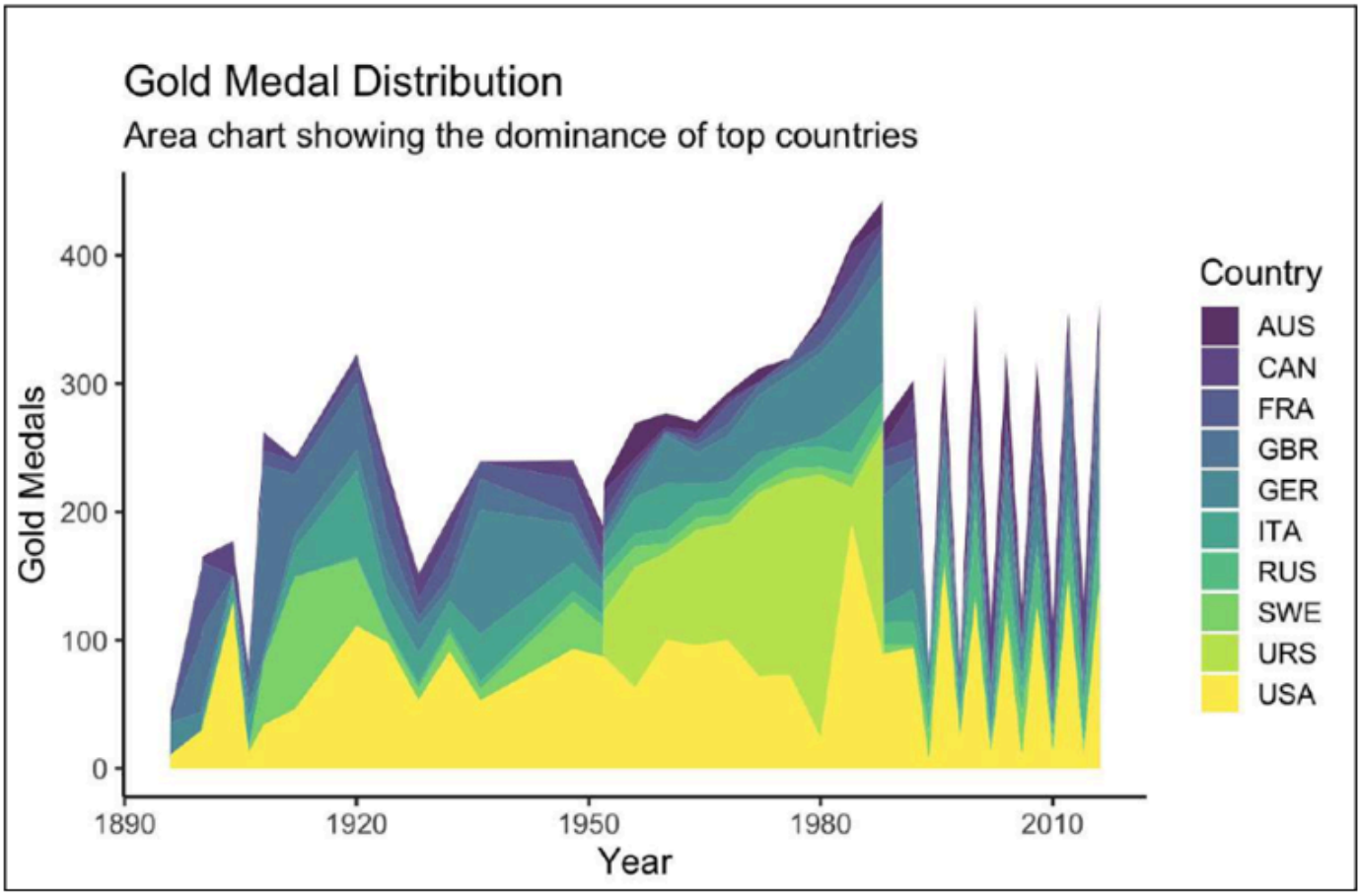
Insight: Gold medal dominance differs markedly from country to country with some nations displaying eras of sustained supremacy in terms of total gold medal wins.

Key Findings:

- The area chart shows the distribution of gold medals among the top-performing nations over time.
- Countries with a strategic focus on Olympic performance, like the USA and USSR, dominated certain decades.
- There has been an increased distribution of gold medals in the last few decades, indicating increased competition, with more countries contributing to the overall medal count.

Visualization:

Area Chart: This style of chart highlights the changes in dominant gold medal-winning countries over the decades, with transitions, signaling shifts in athlete power and ever-increasing competition as the Games moved into a new area.



Discussions:

Many articles drew their data from Olympic Games data, yielding deep insights into long-term trends in athletic performance and country-level dominance. The rise in the average height of men's basketball players (from 1932 to today) is indicative of evolving athletic demands probably influenced by progress in sports science and a competitive drive for physiological optimization. Although height was found to correlate strongly with the probability of being a medal winner, the relatively low predictive power of this model (AUC = 0.5973) demonstrates that success in sports is multifactorial, involving technical skills, teamwork, and psychological resilience.

The patterns of medals won by the top 10 countries over the years showed an unflinching dominance of a few nations, including the United States and the former Soviet Union, but also a growing global competitiveness in the past 15 years. These trends highlight the importance of consistent investment in sports infrastructure and talent development. Long-term trends also mirror the effects of historical and geopolitical developments that have disrupted fair flows of opportunities and resources.

The gold medal distribution over time did show how dominance is concentrated among a select few nations, though trends in recent games suggest a move towards broadening participation and success around the world. This would mean advancement in evening out the playing field such that emerging countries can compete against long-time leading countries.

So while this analysis revealed several important trends and factors driving success, there still seemed to be crucial limitations. The link between height and medal success is nuanced and needs to be explored in light of a variety of other factors including economic conditions, training regimes, and the impact of particular sports

disciplines. The focus on medal production by dominant countries, informative as it is, hides the achievements of athletes from less established nations who are making significant strides on the world stage.

Future work in this area might include exploratory analysis of non-medal performances, further analysis of the intersection between politics, economy, and sports, or dynamic tools to visualize trends in more effective ways. Such initiatives could deliver a more comprehensive perspective on the changing landscape of international sports and be transformative for countries seeking to improve their Olympic performance.

References

Griffin, R. (2018, June 15). 120 years of olympic history: Athletes and results. Kaggle. https://www.kaggle.com/datasets/heesoo37/120-years-of-olympic-history-athletes-and-results

Roos, D. (2020, March 24). When world events disrupted the Olympics. History.com. https://www.history.com/news/olympics-postponed-cancelled

Kennedy, L. (2021, July 26). 6 times the Olympics were boycotted. History.com. https://www.history.com/news/olympic-boycotts