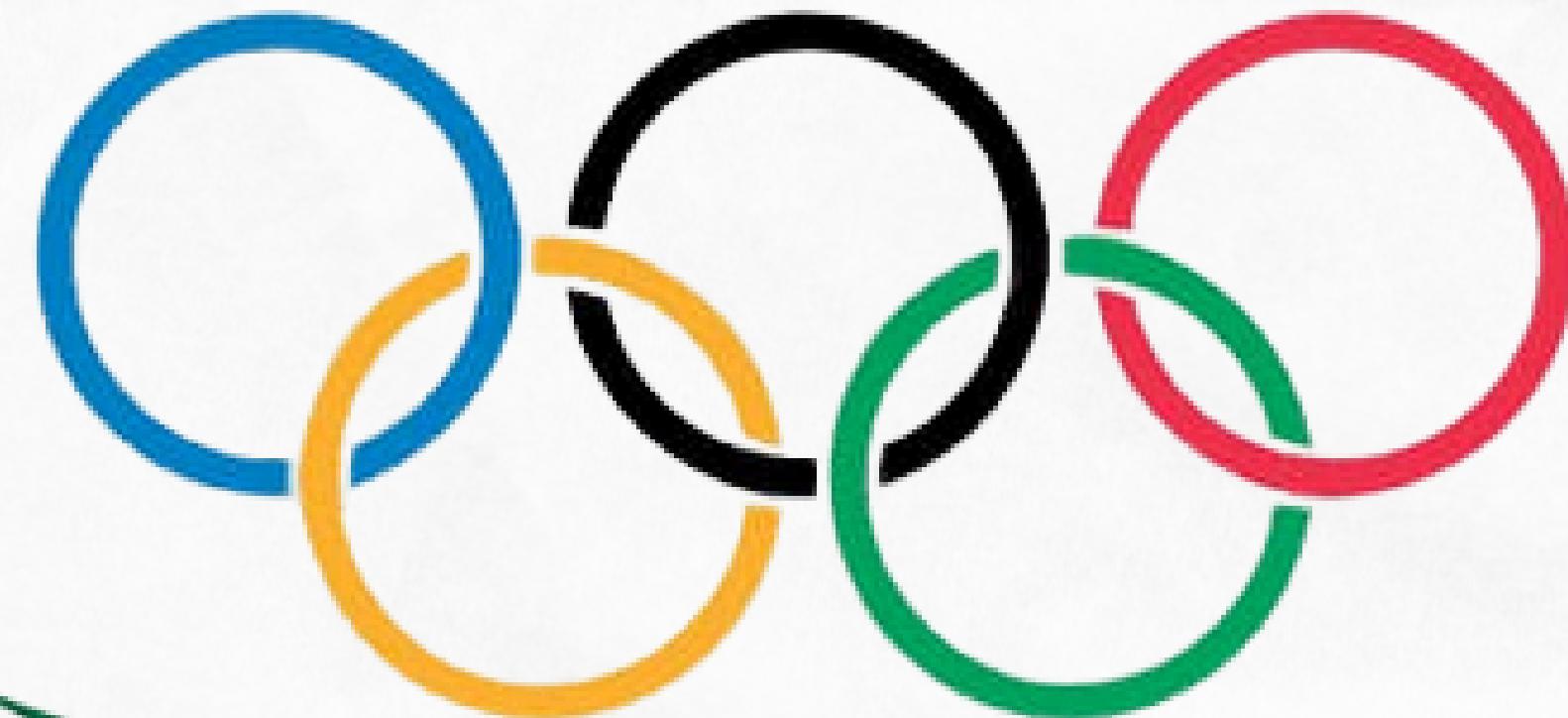


THE IMPACT OF HOSTING AND PHYSICAL ATTRIBUTES ON OLYMPIC PERFORMANCE AND PARTICIPATION

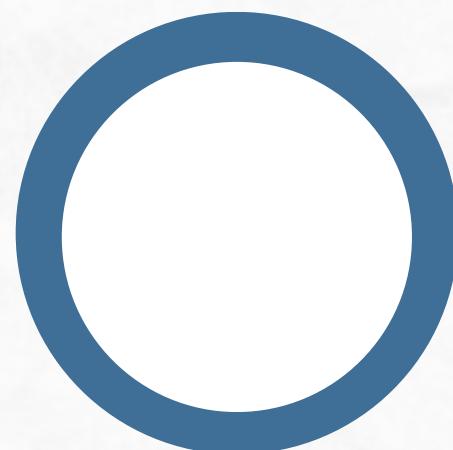
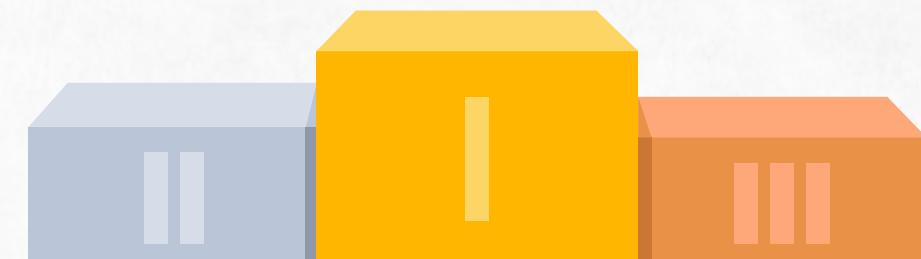
PROJECT ONE - GROUP 3



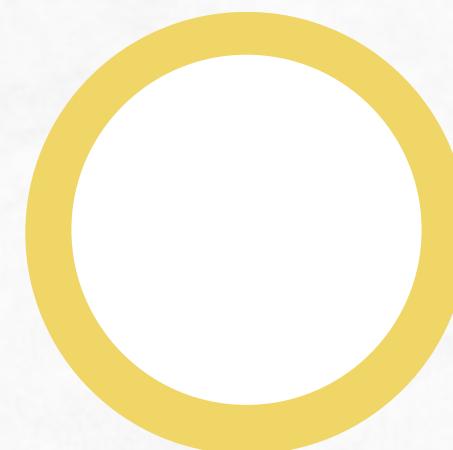
OUR DATA SET

120 YEARS OF OLYMPIC HISTORY

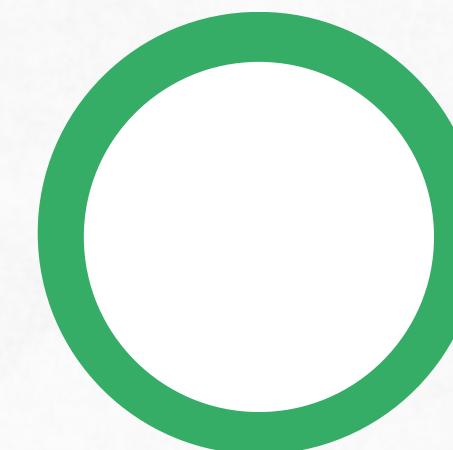
Athletes & Results 1896 to 2016



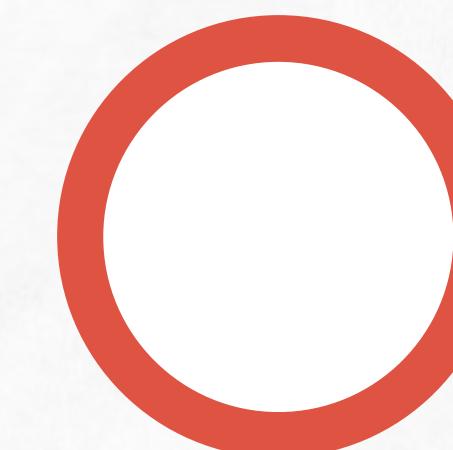
ID



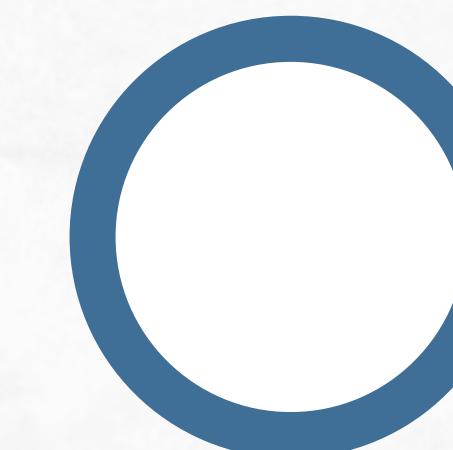
NAME



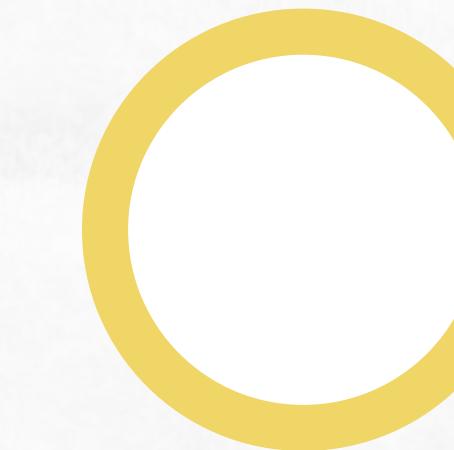
SEX



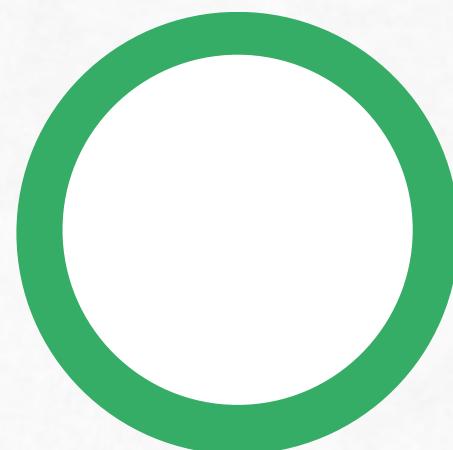
AGE



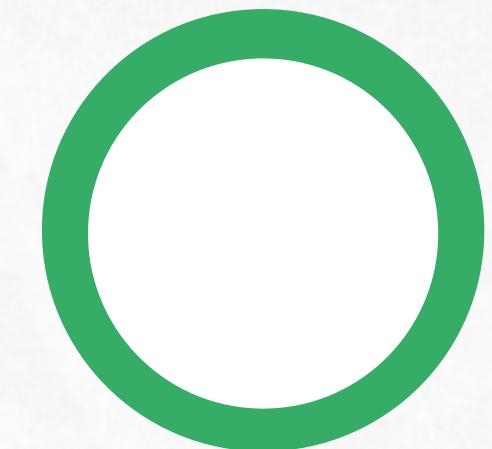
HEIGHT



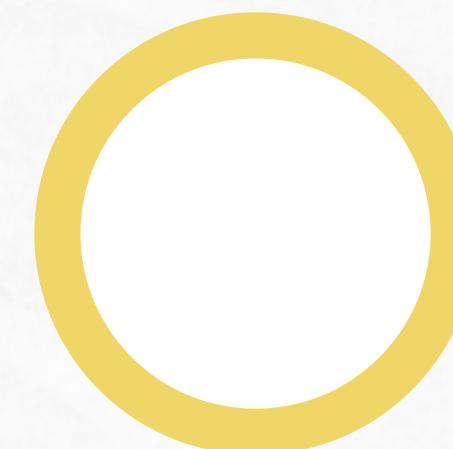
WEIGHT



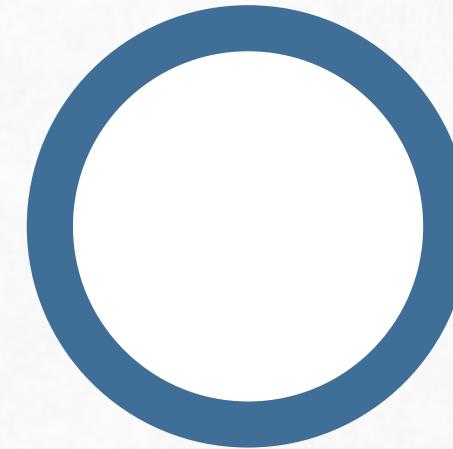
TEAM



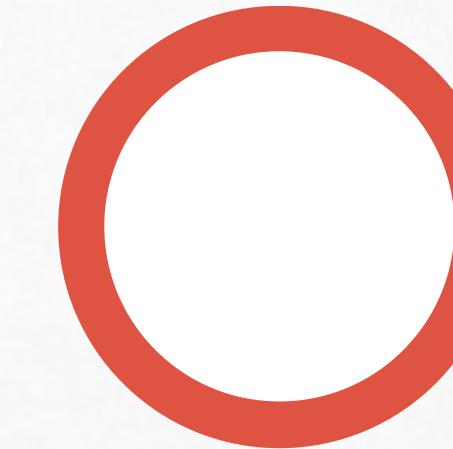
GAMES



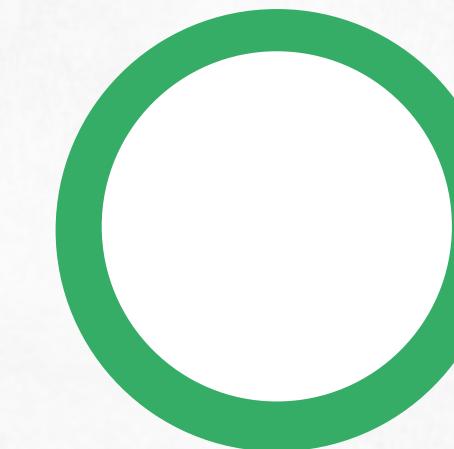
YEAR



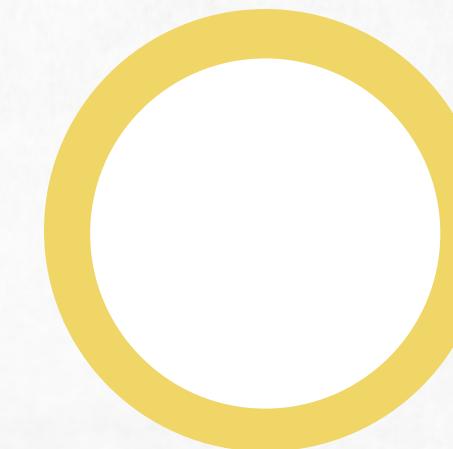
SEASON



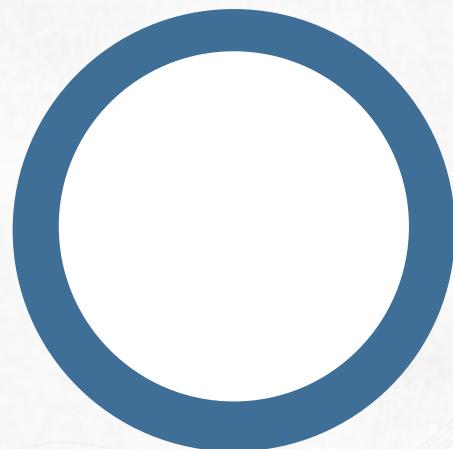
CITY



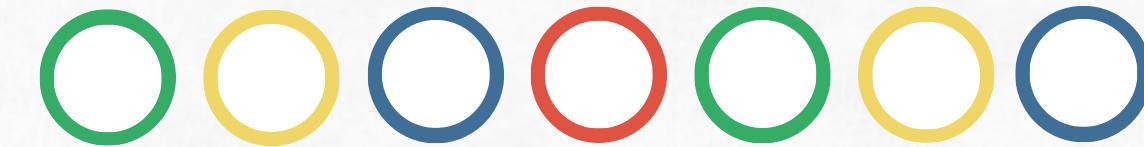
SPORT



EVENT



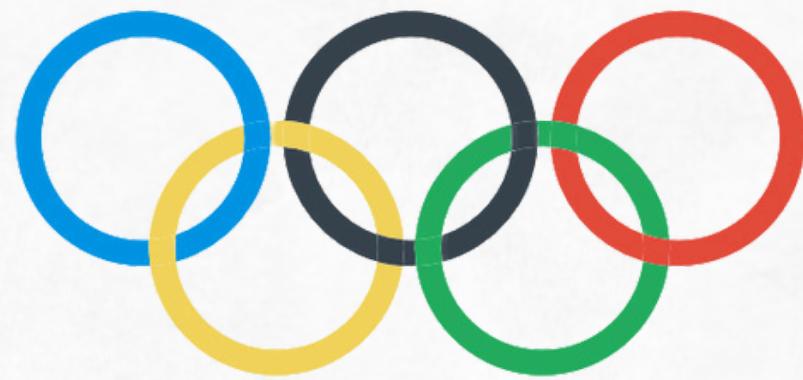
MEDAL



POTENTIAL LIMITATIONS OF OUR DATA SET

- **Missing Data:** There was a varying amount of missing height, weight, and age data, which could introduce biases if not handled appropriately.
- **Data Collection Methods:** Data collection methods for athletes might have changed over the years, affecting the consistency and accuracy of the data (the dataset begins in 1896).
- **Change in Olympic Scheduling:** Until 1992, the Summer and Winter Olympics were held in the same year every four years. Starting in 1994, the schedule changed so that the Winter and Summer Olympics occur in alternating even-numbered years.
- **Changes in Event Offerings:** Over the years, the types and number of events have changed, with some events being added and others removed. This affects the number of medals available and the distribution of athletes across sports.
- **Changes in Eligibility:** The introduction of professional athletes into the Olympics has impacted the composition and performance of athletes competing in the Games.





DATA CLEANING & FORMATTING

- New DataFrame for only the last 15 Olympics
 - 1960 to 2016
- DNM for Medal Nulls
- Dropped any row with nulls
 - missing height, weight, age data

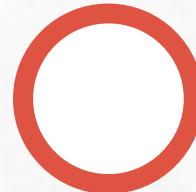
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 271116 entries, 0 to 271115
Data columns (total 15 columns):
 #   Column   Non-Null Count   Dtype  
--- 
 0   ID        271116 non-null    int64  
 1   Name      271116 non-null    object  
 2   Sex       271116 non-null    object  
 3   Age        261642 non-null    float64 
 4   Height     210945 non-null    float64 
 5   Weight     208241 non-null    float64 
 6   Team       271116 non-null    object  
 7   NOC        271116 non-null    object  
 8   Games      271116 non-null    object  
 9   Year       271116 non-null    int64  
 10  Season     271116 non-null    object  
 11  City       271116 non-null    object  
 12  Sport      271116 non-null    object  
 13  Event      271116 non-null    object  
 14  Medal      39783 non-null    object  
dtypes: float64(3), int64(2), object(10)
memory usage: 31.0+ MB
```



```
<class 'pandas.core.frame.DataFrame'>
Index: 196084 entries, 196158 to 156986
Data columns (total 15 columns):
 #   Column   Non-Null Count   Dtype  
--- 
 0   ID        196084 non-null    int64  
 1   Name      196084 non-null    object  
 2   Sex       196084 non-null    object  
 3   Age        196084 non-null    float64 
 4   Height     196084 non-null    float64 
 5   Weight     196084 non-null    float64 
 6   Team       196084 non-null    object  
 7   NOC        196084 non-null    object  
 8   Games      196084 non-null    object  
 9   Year       196084 non-null    int64  
 10  Season     196084 non-null    object  
 11  City       196084 non-null    object  
 12  Sport      196084 non-null    object  
 13  Event      196084 non-null    object  
 14  Medal      196084 non-null    object  
dtypes: float64(3), int64(2), object(10)
memory usage: 23.9+ MB
```

RESEARCH QUESTIONS

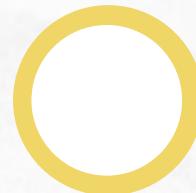
OUR DATA SET



One: Has one country hosted more than others in the last 30 Olympic Occurrences? Does hosting the olympics impact percentage of medals won by home country?



Two: What is the age distribution for the Summer & Winter Olympics? Which sport(s) have the highest average age?



Three: Does height or weight impact which sport someone is likely to compete in? Has the average height or weight changed over the last 30 Olympic occurrences?

Bonus T-Test: Is the average height or average weight significantly different for the Summer and Winter Olympics?



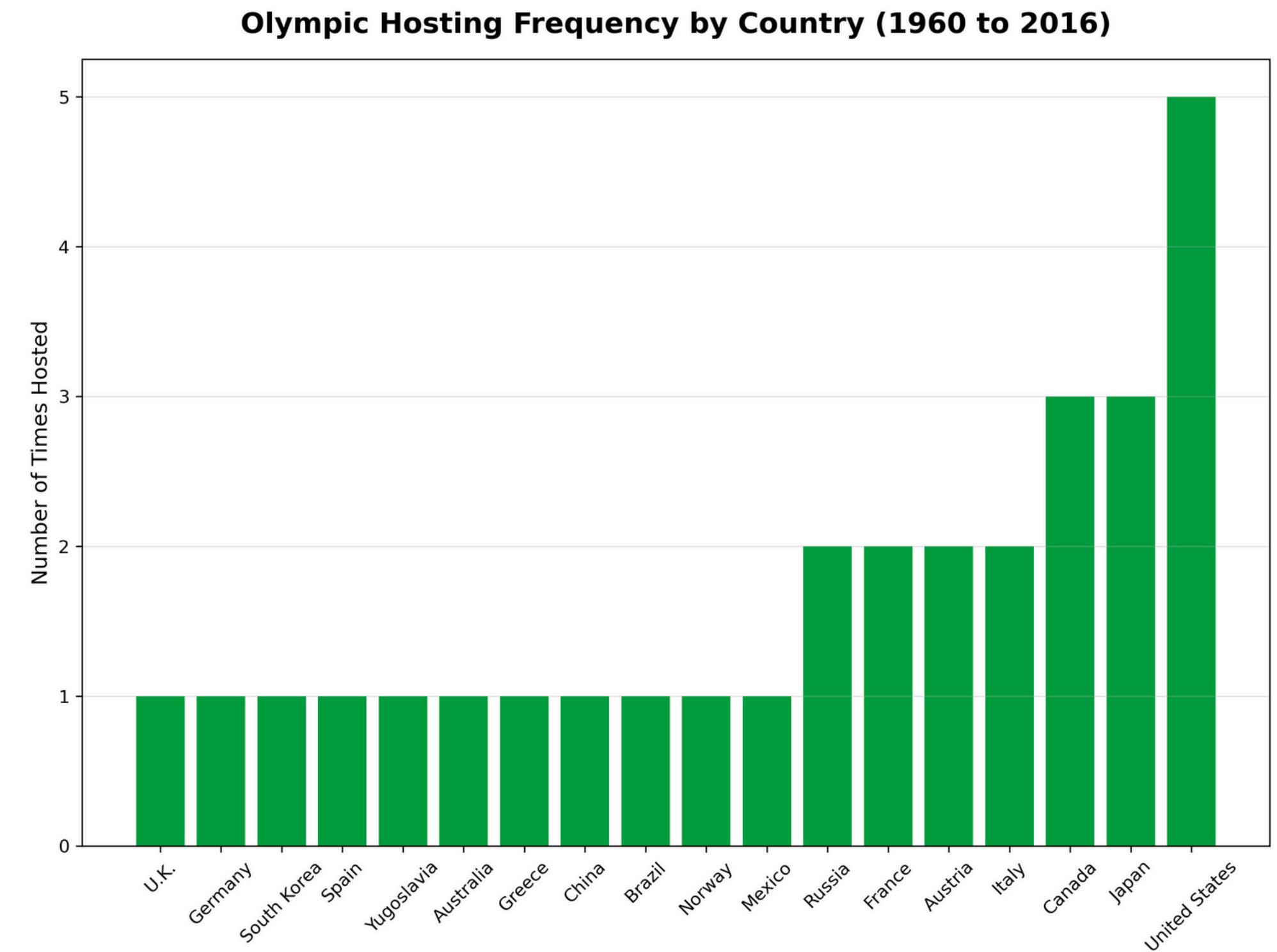
Regression: Does age impact whether or not an athlete is more likely to medal?

RESEARCH QUESTION ONE:

Has one country hosted more than others in the last 30 Olympic Occurrences?

Does hosting the olympics impact percentage of medals won by home country?

The United States has hosted the most Olympics since 1960, hosting 5 times. The second two closest other Countries that have hosted are Canada and Japan, each hosting 3 times since 1960. It is much more likely for a large potential hosting country to only host the Olympics once, maybe twice over 30 Olympic occurrences.



RESEARCH QUESTION ONE:

Has one country hosted more than others in the last 30 Olympic Occurrences?

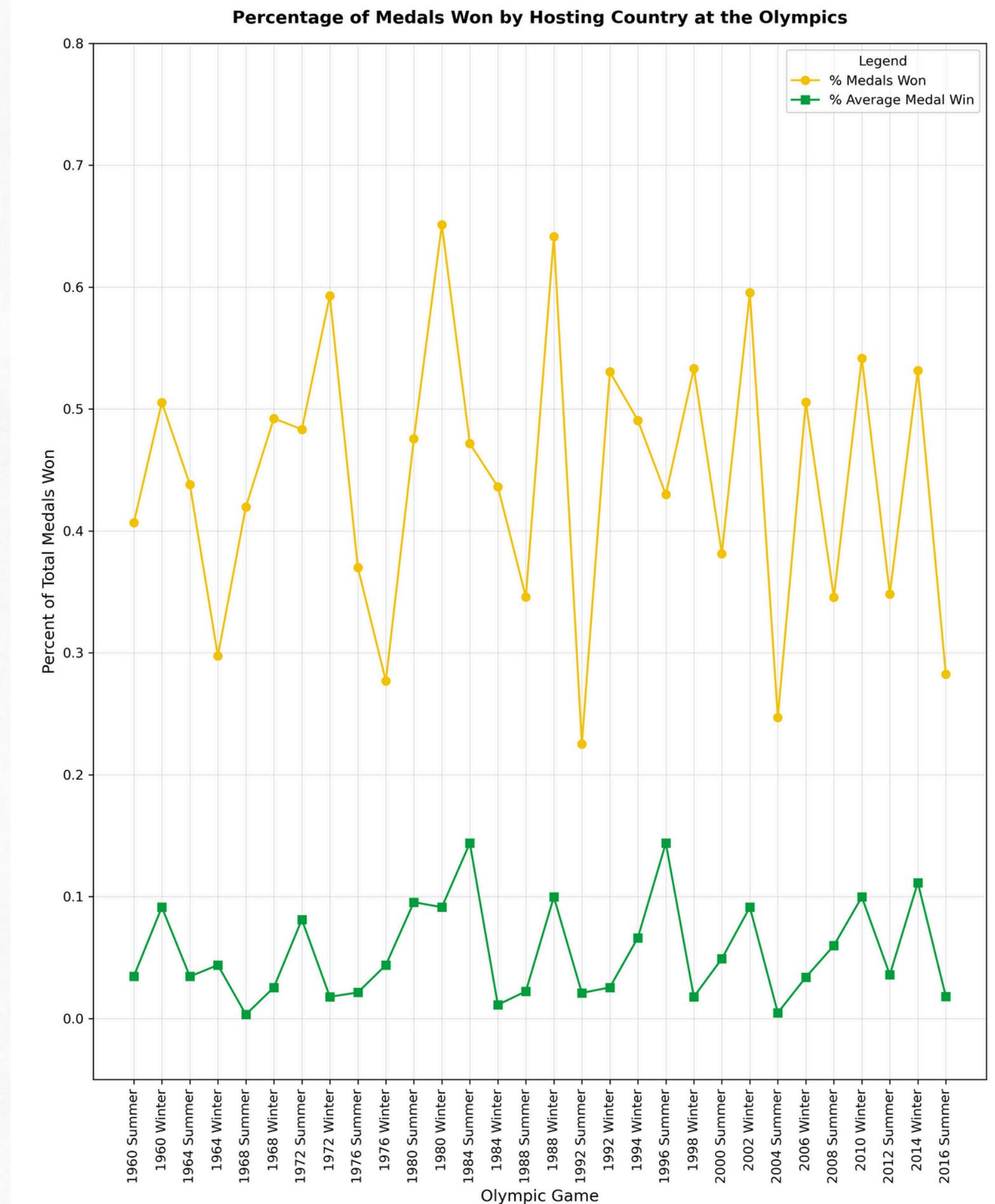
Does hosting the olympics impact percentage of medals won by home country?

MEDALS WON BY HOSTNG COUNTRY ANALYSIS

Hosting the Olympics appears to impact the number of medals won. All countries that have hosted the Games since 1960 have seen an increase in the percentage of medals won during their hosting year, as shown in the graph. The green line displays the average number of medals won by the host country in a typical Olympic year, correlating to either the Summer or Winter Games, depending on which they hosted. The yellow line displays the percentage of medals won in the year they hosted. While some countries saw larger boosts, all 18 countries experienced an increase in their percentage of medals won.

Some potential contributing factors here could be:

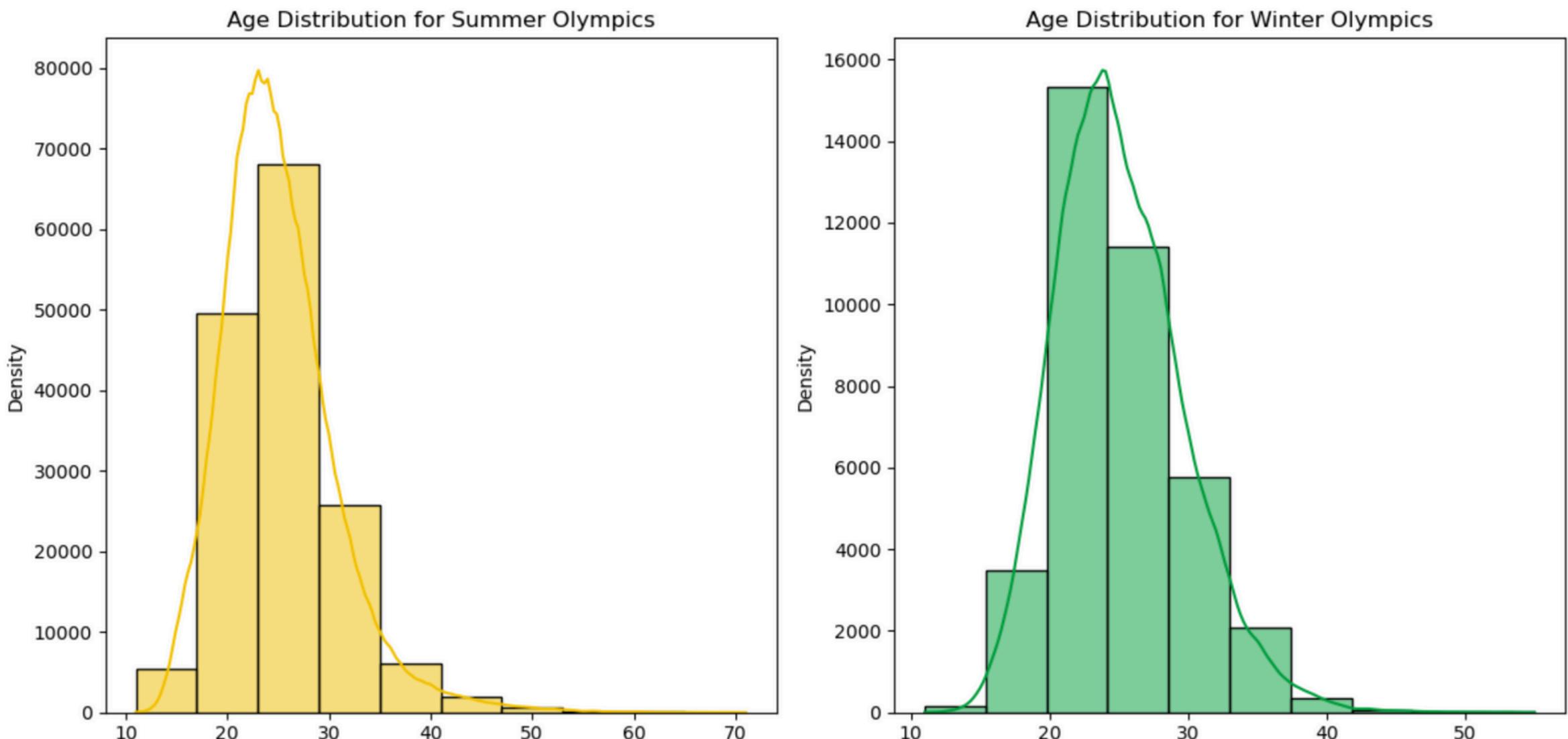
- Athletes don't have to travel and adjust to time zone changes
- Local support
- Automatic qualification for extra events as the hosting country



RESEARCH QUESTION TWO: What is the age distribution for the Summer & Winter Olympics?

AGE DISTRIBUTION ANALYSIS

The data presented shows that during the Summer Olympics, the average age is closer to the early 20s while, the average age for Winter Olympics is in the mid 20s.

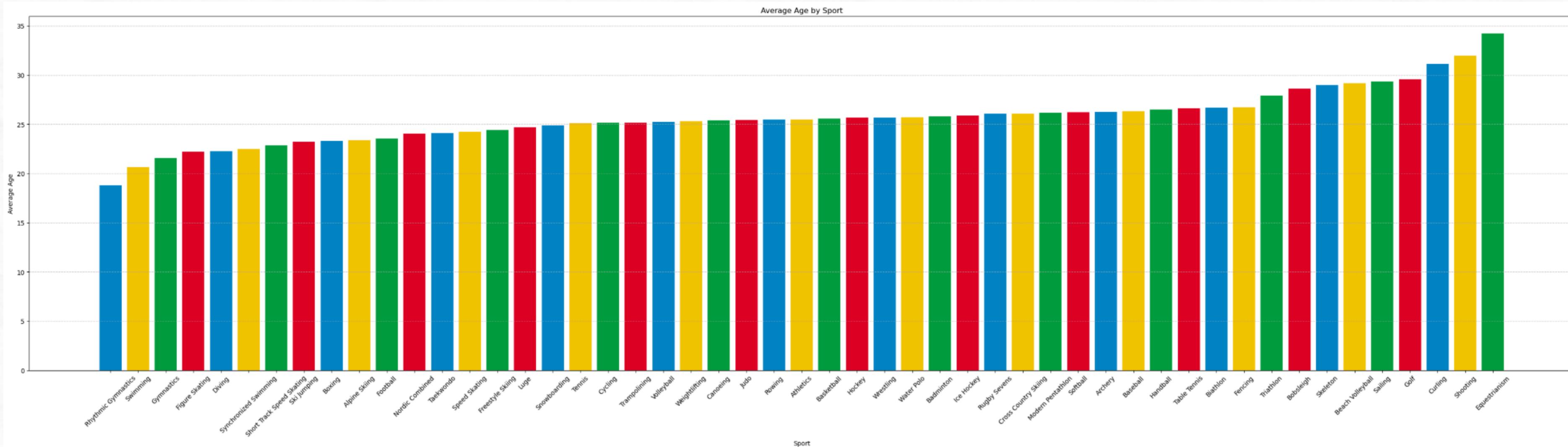


RESEARCH QUESTION TWO:

Which sport(s) have the highest average age?

HIGHEST AVERAGE ANALYSIS

Data shows that the sport with highest average age is Equestrianism. Equestrian sports don't put as much physical labor on the athlete in comparison to sports like gymnastics, boxing, and figure skating.



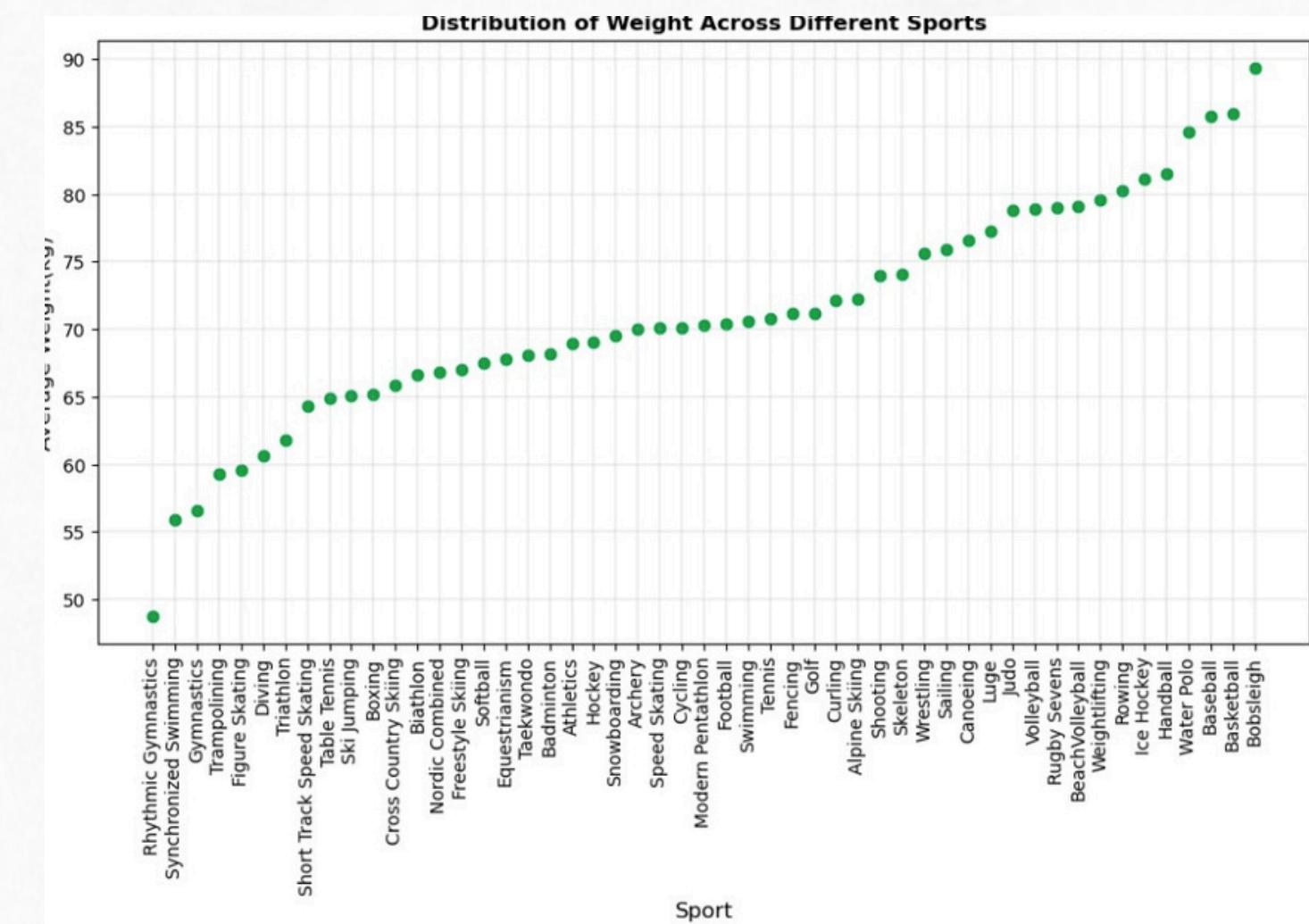
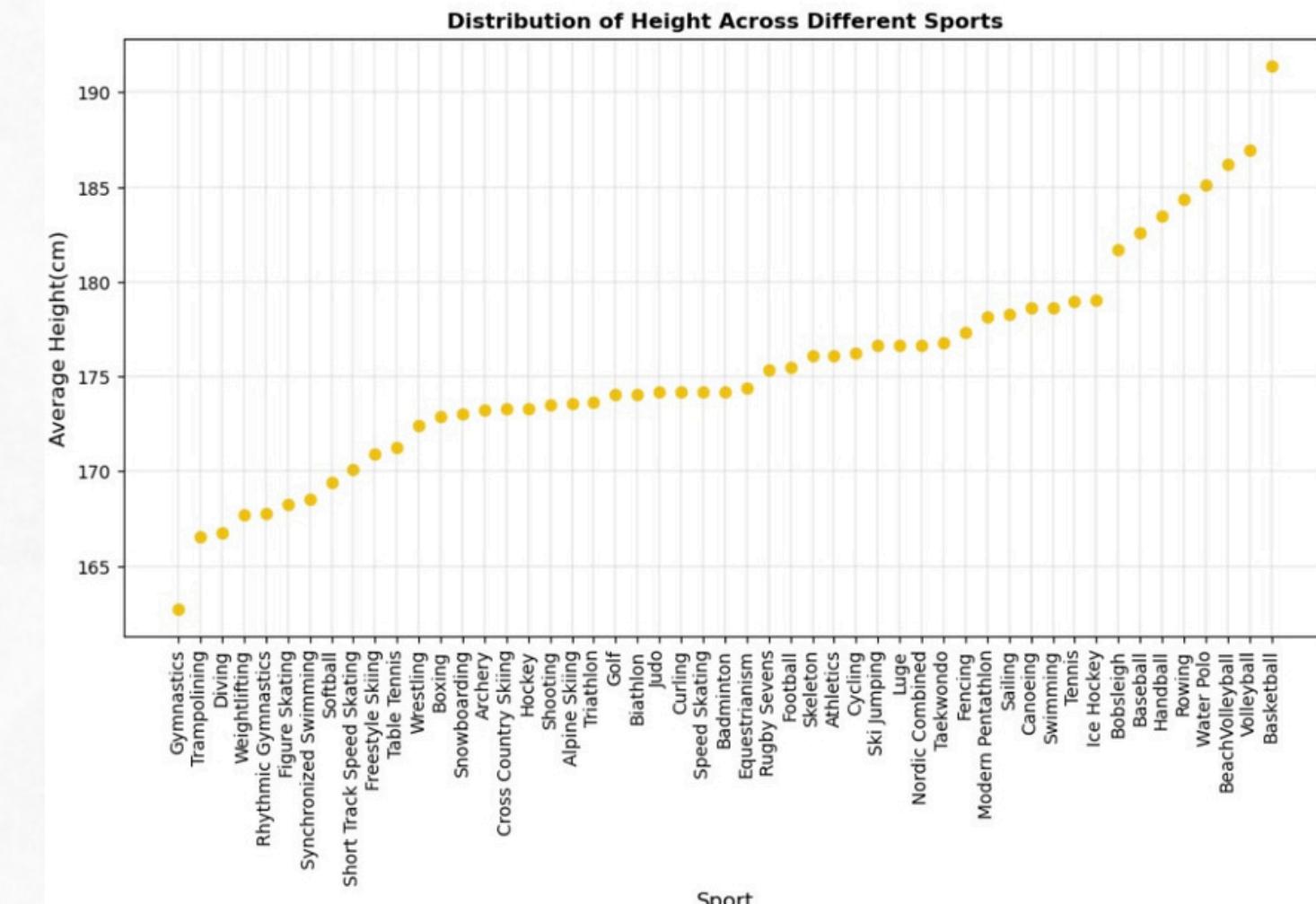
RESEARCH QUESTION THREE:

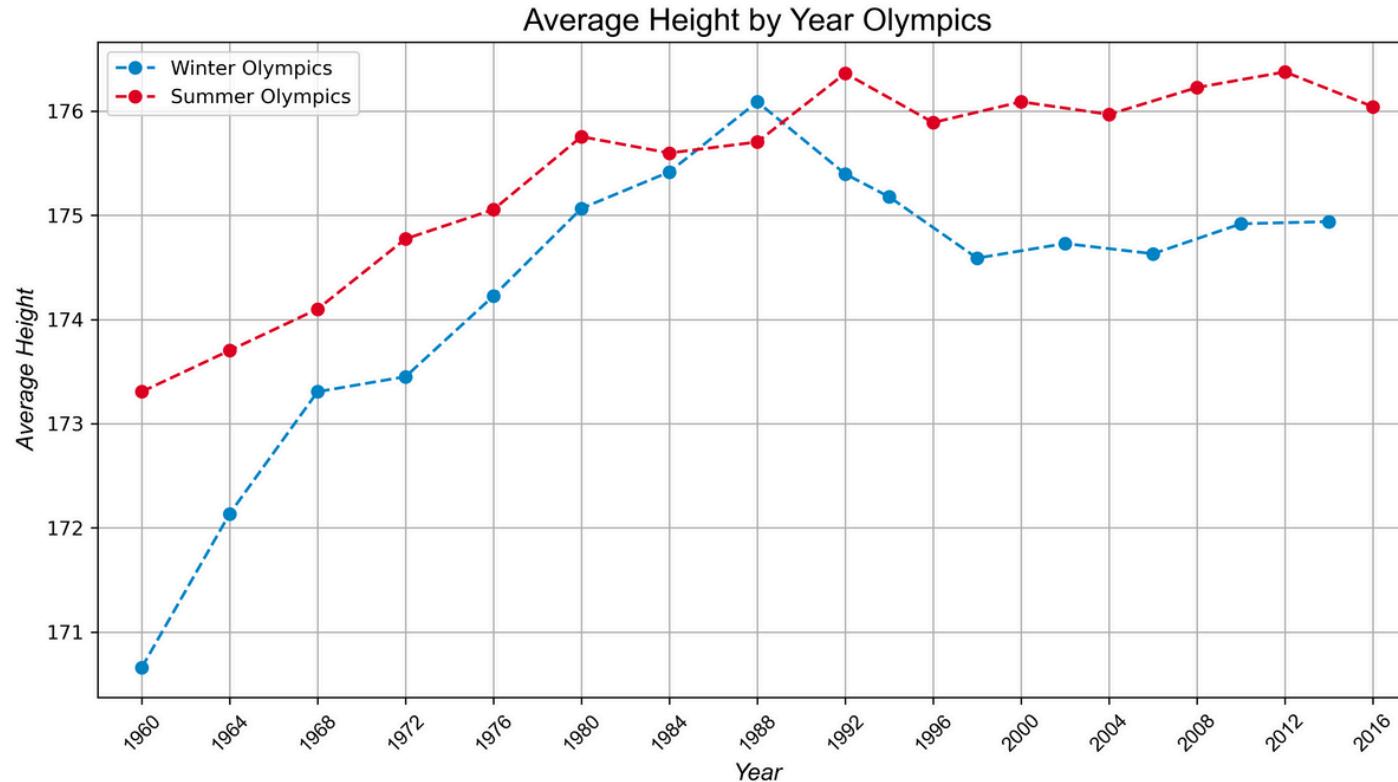
Does height or weight impact which sport someone is likely to compete in?

Has the average height or weight changed over the last 15 Olympics?

HEIGHT & WEIGHT ACROSS DIFFERENT EVENTS

It can be determined that both height and weight are impactful factors when competing in Olympic sports. Both factors can be advantages or disadvantages depending on the sport. For example, taller, heavier players may have advantages in sports like basketball, volleyball, and water polo because they have longer reach and strength. While in others, shorter, lighter statures offer greater agility and maneuverability as in gymnastics.



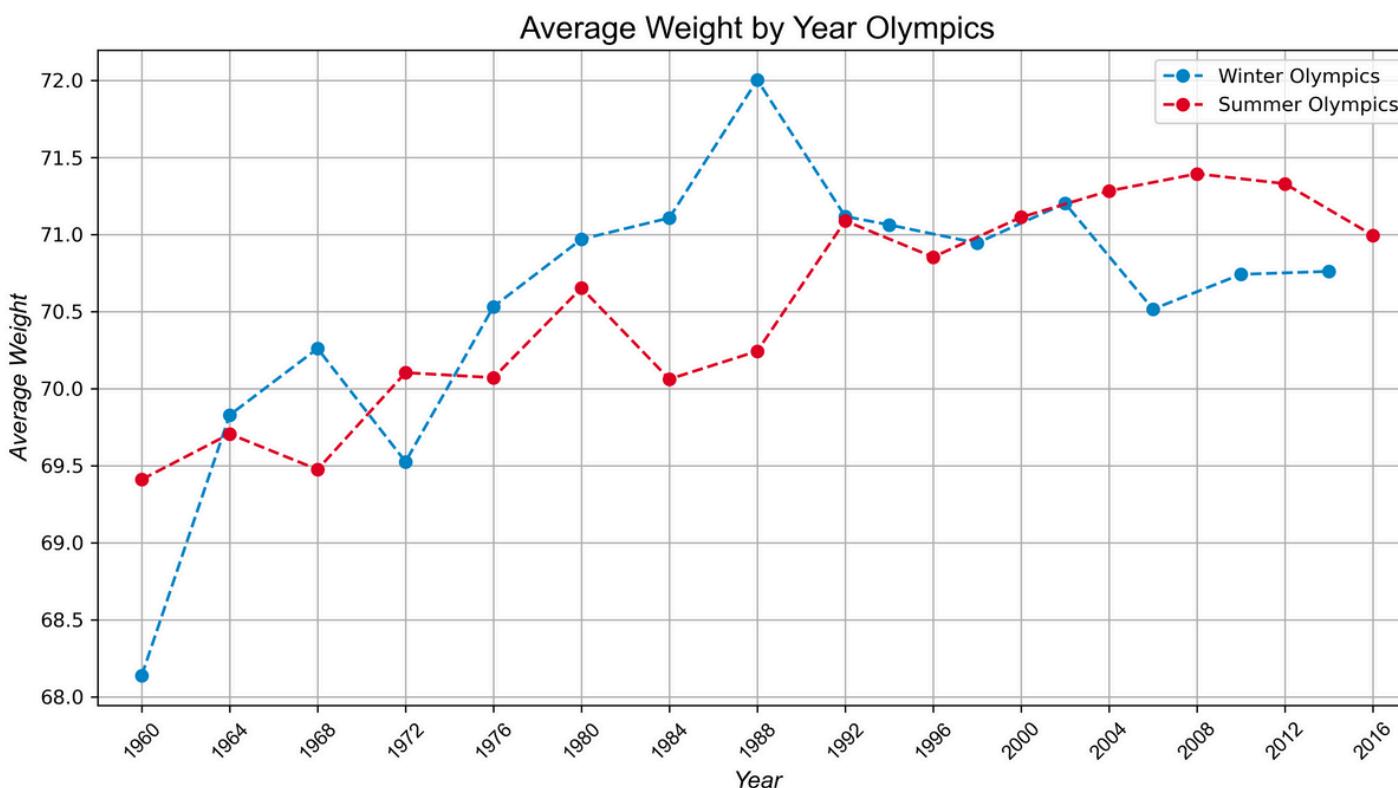


AVERAGE HEIGHT 1960 TO 2016

For both Summer and Winter Olympics the average height has mostly increased year over year.

There was a few downward years for the Winter Olympics which can be explained by a large volume of new events added, who's height averages brought down the overall mean.

While there is an upward trend, there is not a huge change in average height over the last 15 Olympics, with the greatest change being from ~170 cm to ~176 cm which is the difference from 5'6 to 5'7 1/2



AVERAGE WEIGHT 1960 TO 2016

For both Summer and Winter Olympics the average weight has a trend of increasing year over year. This year over year trend isn't quite as clean as the height averages but has a clear upward trend with some peaks and valleys.

The biggest change here is from 68 kg to 72 kg in the Winter Olympics. This is about a total 10 pound average increase, meaning this data has varied more overall vs. average height data.

T-TEST:

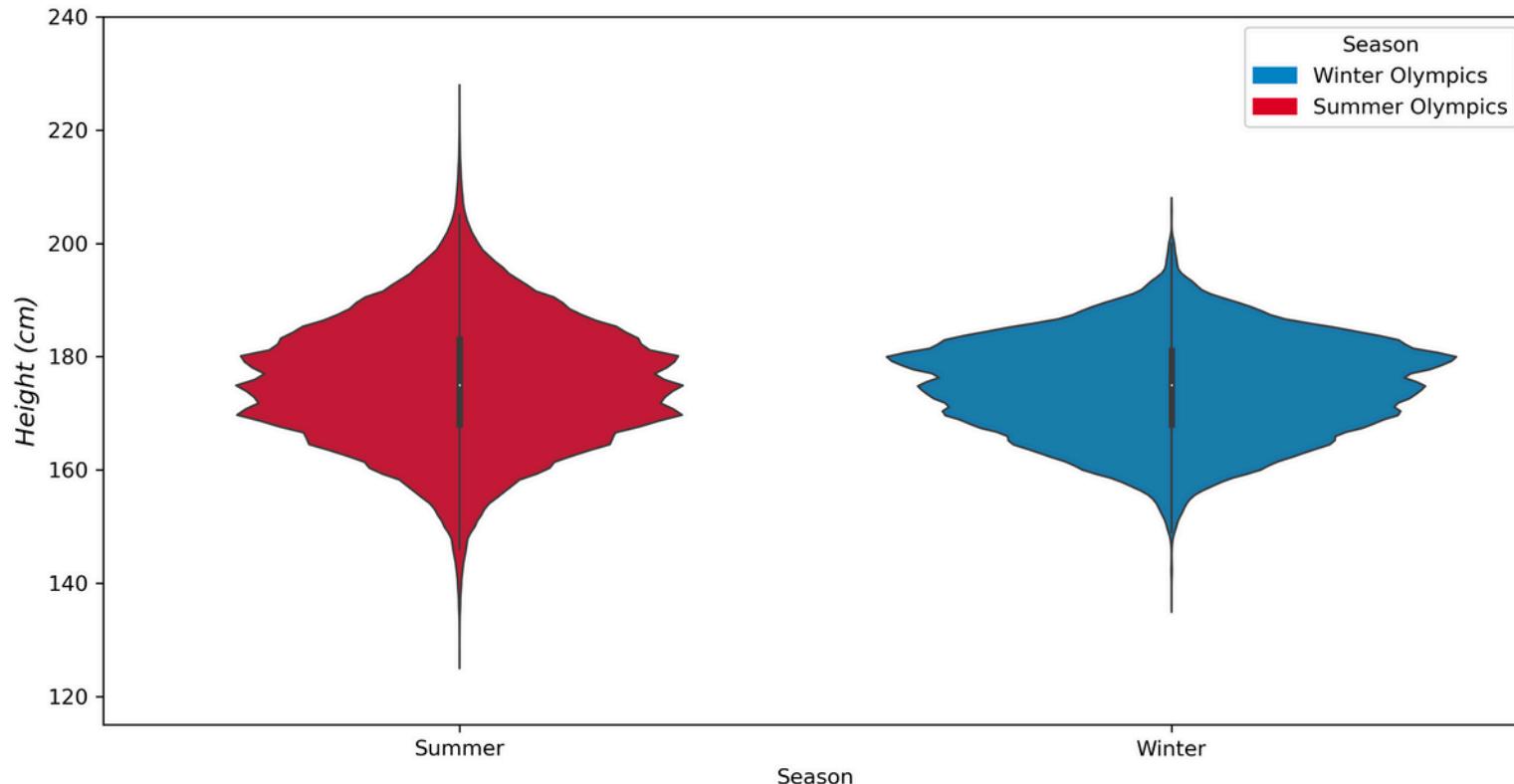
Is the average height or average weight significantly different from Summer to Winter Olympics?

NULL & ALTERNATIVE HYPOTHESES:

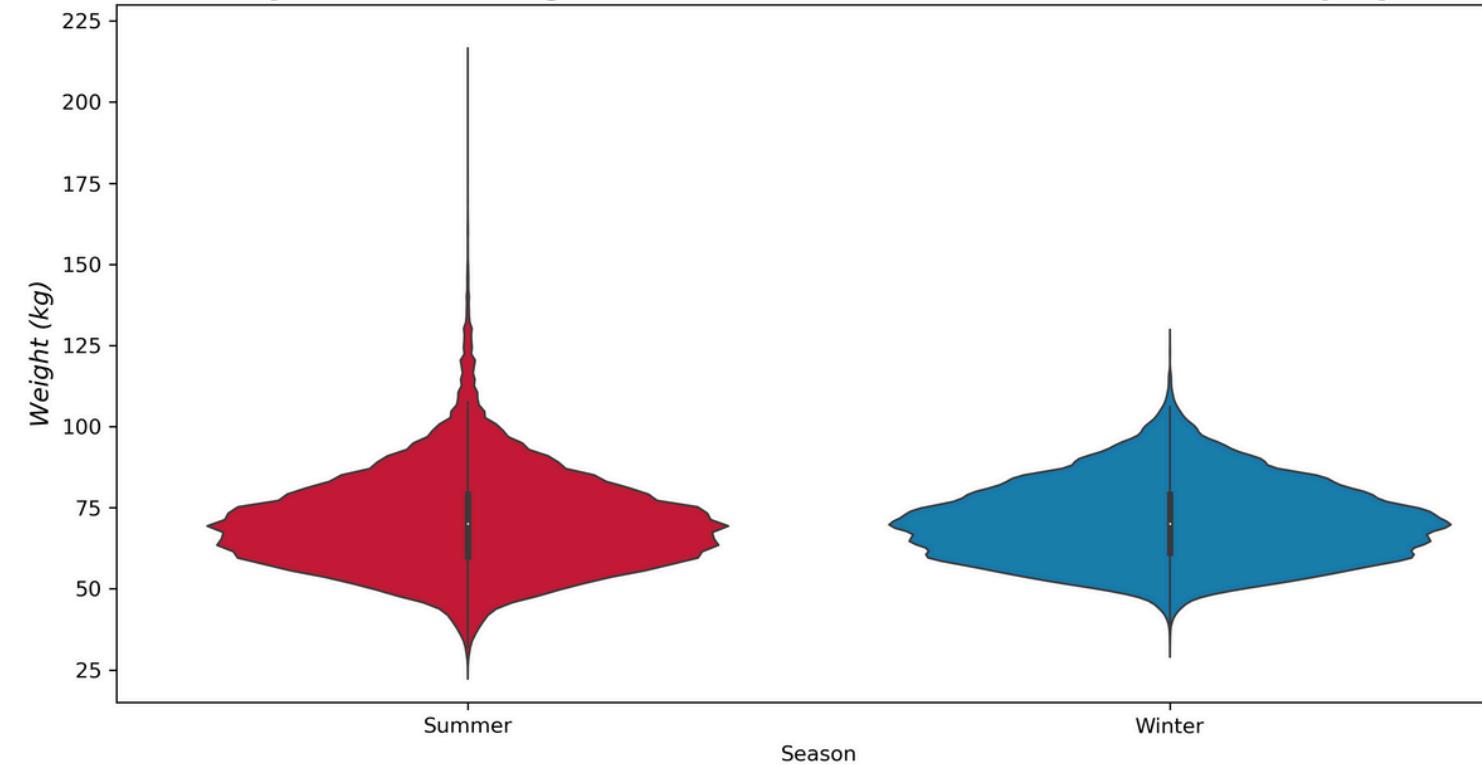
Null: The average height or the average weight are not significantly different in Summer vs. Winter Olympics

Alternative: There is a significant difference between height or weight in Summer vs. Winter Olympics

Comparison of Height in the Last 15 Winter and Summer Olympics



Comparison of Weight in the Last 15 Winter and Summer Olympics



T-TEST OUTCOME:

Summer vs. Winter Height

Our P-Value for Summer vs. Winter Height differences is 1.32e-48. This number is smaller than our significance level of .05 which means we reject the null hypothesis and there is a significant difference between average athlete height for Summer and Winter Olympics.

Summer vs. Winter Weight

Our P-Value for Summer vs. Winter Weight differences is .084. This number is bigger than our significance level of .05 which means we fail to reject the null hypothesis and there is not a significant difference between average athlete weight for Summer and Winter Olympics.

T-Test Results for Summer/Winter Height:

T-Statistic: 14.655164939136906

P-Value: 1.3248279871475216e-48

Summer Median Height: 175.0

Summer Mean Height: 175.55377382465056

Summer Max Height: 226.0

Summer Min Height: 127.0

Winter Median Height: 175.0

Winter Mean Height: 174.67063902388585

Winter Max Height: 206.0

Winter Min Height: 137.0

T-Test Results for Summer/Winter Weight:

T-Statistic: -1.7235299044115997

P-Value: 0.08479431852185805

Summer Median Weight: 70.0

Summer Mean Weight: 70.64032401524777

Summer Max Weight: 214.0

Summer Min Weight: 25.0

Winter Median Weight: 70.0

Winter Mean Weight: 70.78158928756075

Winter Max Weight: 127.0

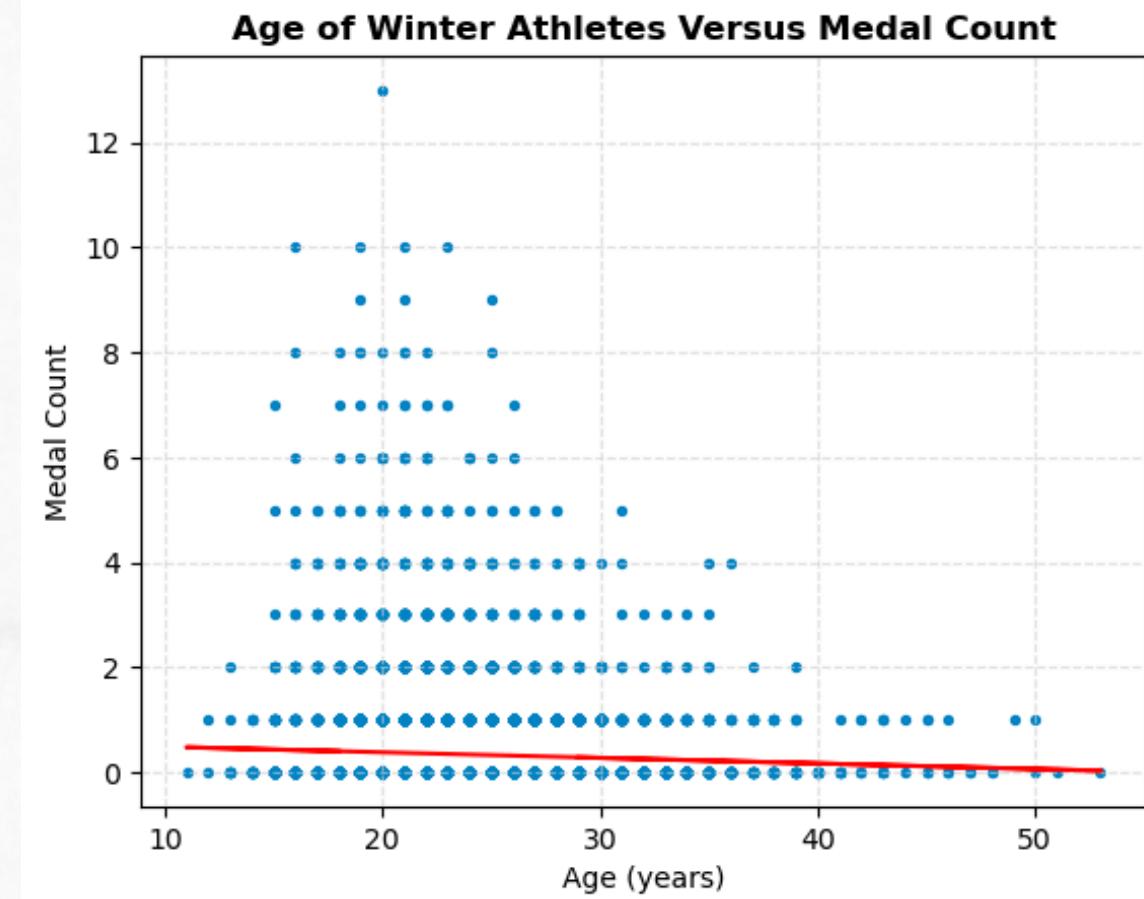
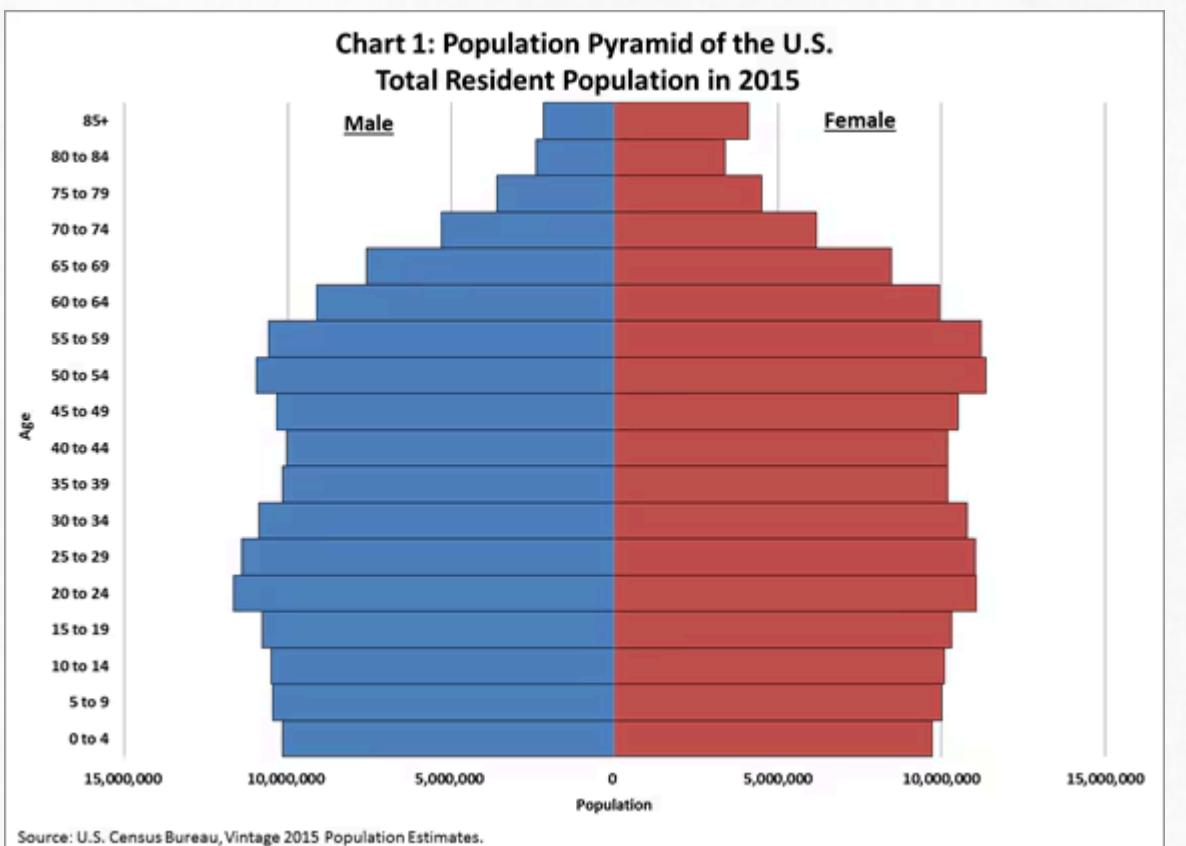
Winter Min Weight: 32.0

REGRESSION:

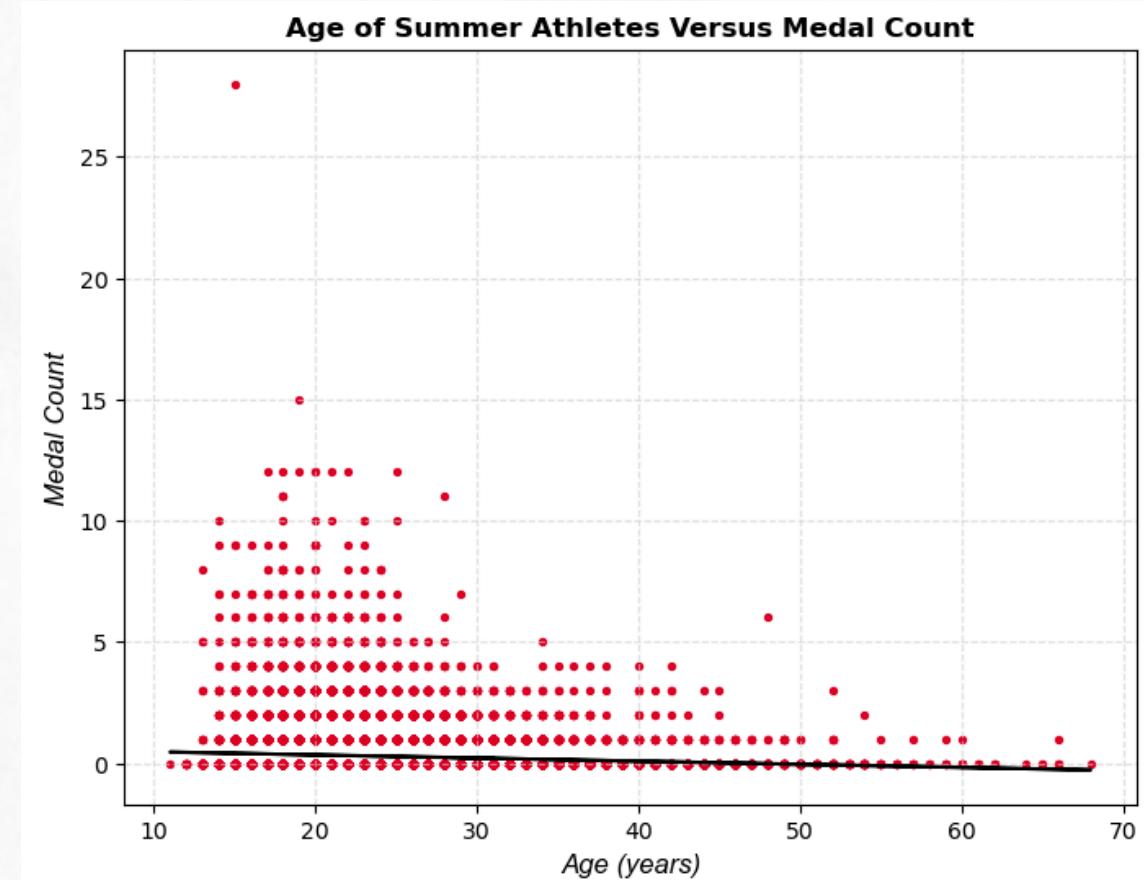
Does age, height, or weight impact whether or not an athlete is more likely to earn a high medal count?

While it certainly affects the chances it is not enough to be able to make accurate predictions based solely on age

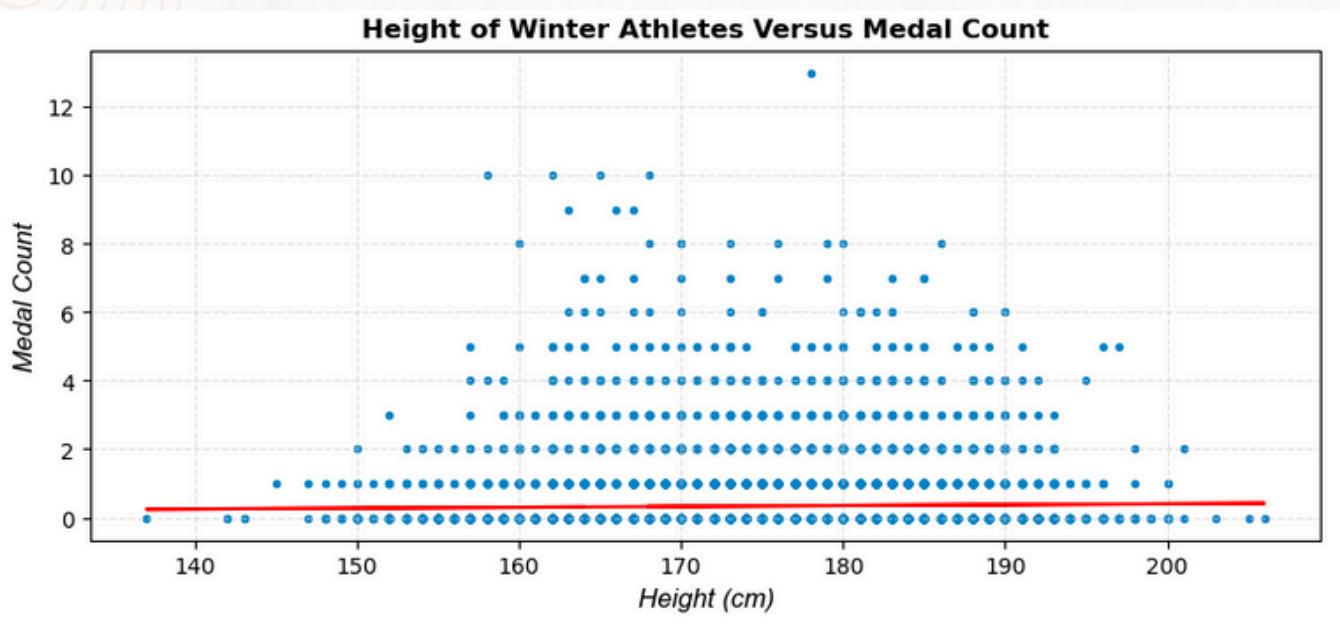
Comparing the age of the athletes to an example age distribution using US age demographics, we can see that there are clearly less athletes of a higher age than would be expected if it followed a traditional age distribution, and even in the regressions the line of best fit does trend slightly downward but the r squared value shows a very unclear relationship, even though there are less athletes of a higher age, the athletes that remain have similar chances to medal.



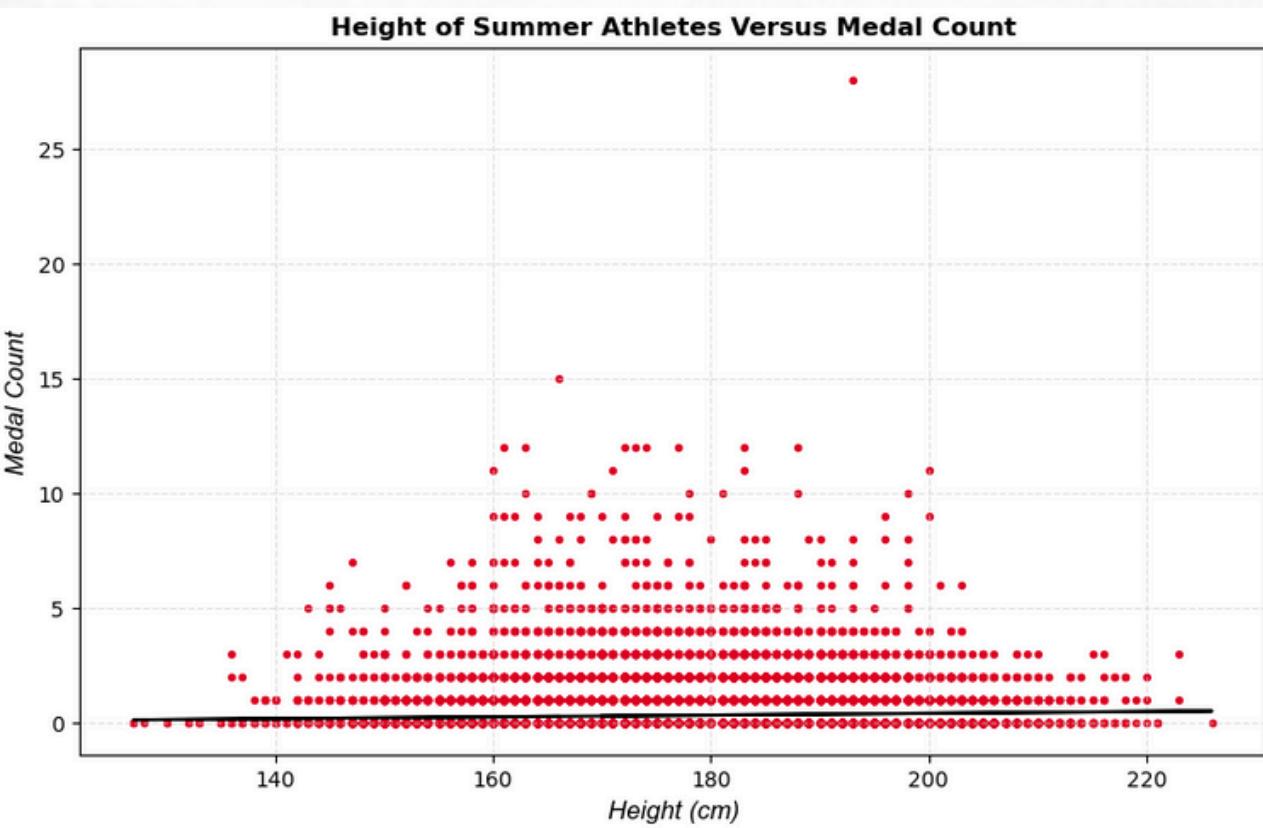
r - squared value of 0.0027719



HEIGHT AND WEIGHT REGRESSIONS



r squared value: 0.0006749

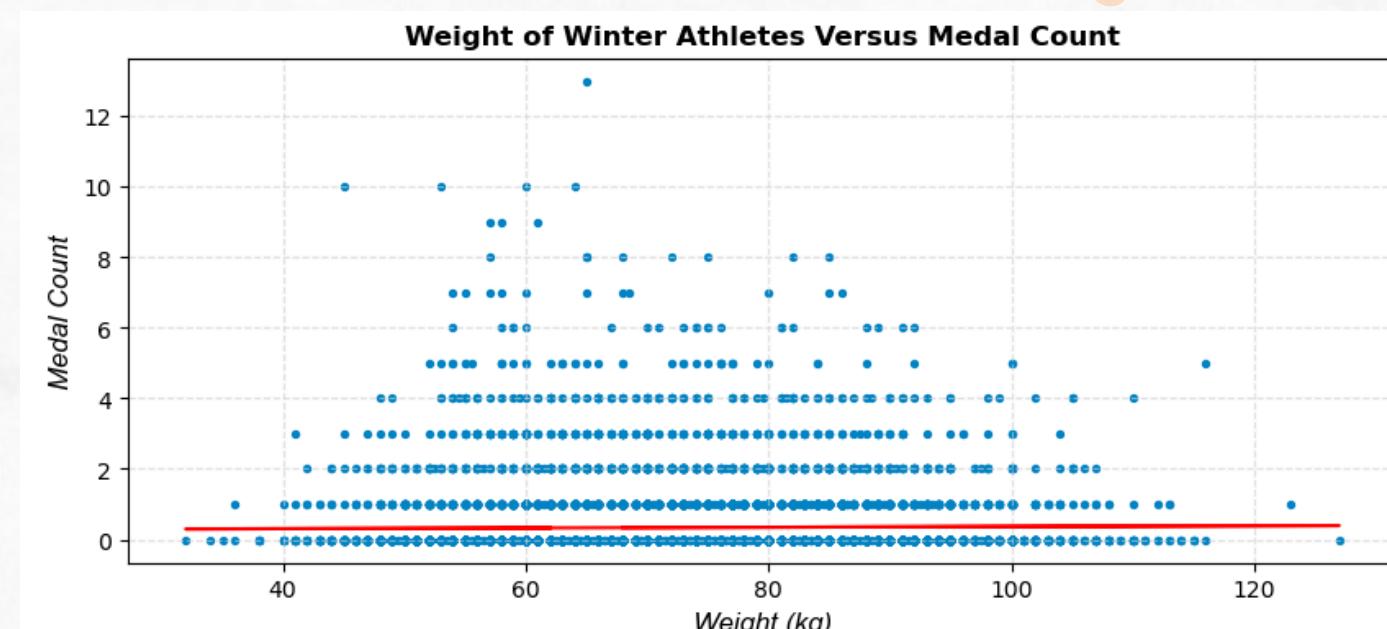


r squared value: 0.0034277

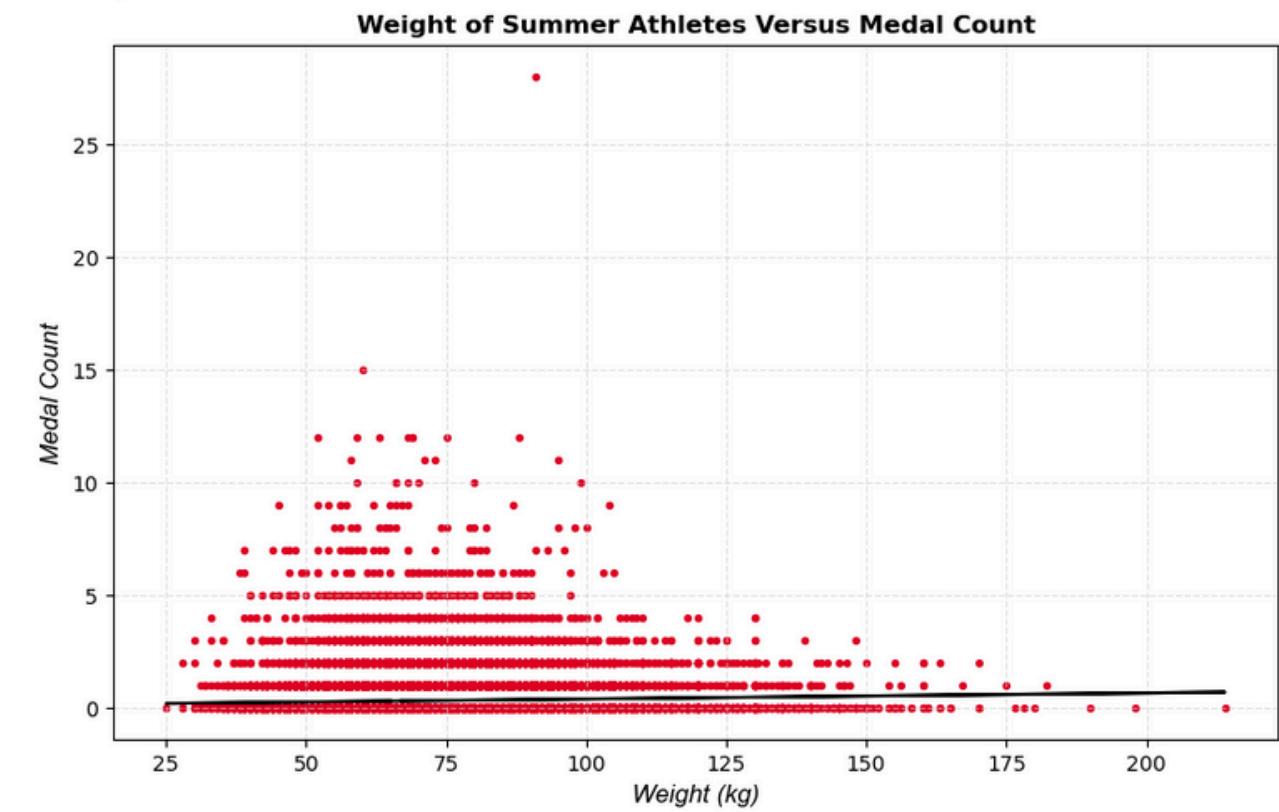
The height and weight showed similar results when regressed against medal count. Although there are slight trends, the relationships are largely insignificant, while having relatively normal distributions.

An interesting takeaway from these graphs is that while there might be no linear relationship, high earning medal count athletes are not outliers in height or weight.

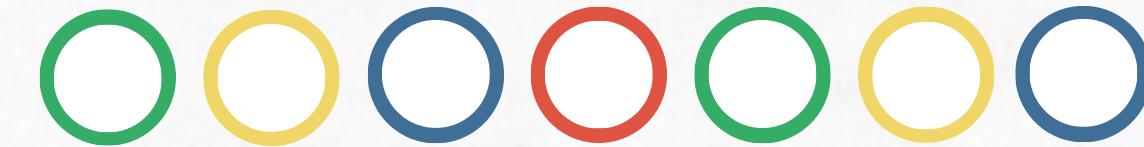
Building off that, despite a person's natural height and weight, there are opportunities to compete and find success.



r squared value: 0.0002371



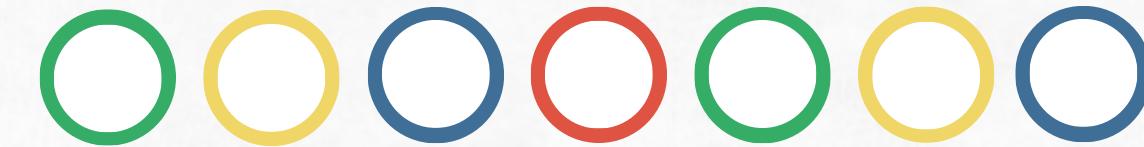
r squared value: 0.0031352



THE IMPACT OF HOSTING AND PHYSICAL ATTRIBUTES ON OLYMPIC PERFORMANCE AND PARTICIPATION

- Hosting appears to impact percentage of medals won by the hosting team. While we can't draw any relationship conclusions, it does appear that hosting increases the percentage of medals won.
- Age plays a role in the sport an athlete plays. The data shows signs that athletes who experience more longevity are those that do not play sports that do not require as much physical stress on the body or require more experience.
- Height and weight appear to influence performance in Olympic sports, with taller, heavier athletes excelling in sports like basketball and volleyball, while shorter, lighter athletes tend to excel in sports like gymnastics.
- Height and weight appear to be increasing since 1960. While neither is a straight upward trend, both height and weight are higher than they were in the 1960 Olympics.
- An Olympians age, height, and weight are not good linear indicators of athletes who earn a high number of medals





HOW TO IMPROVE FOR A FUTURE ANALYSIS

- Enhance Data Completeness:
 - Focus on improving the collection and recording of height, weight, and age data to minimize biases and ensure more comprehensive analysis. This can be done through finding other data sets to match up with missing athlete data.
- Standardize Data Collection Methods:
 - Advocate for consistent data collection methods across different years and events to enhance the accuracy and comparability of the data.
- Analyze Changes in Event Offerings:
 - Examine how the addition or removal of events over time impacts the distribution of medals and athlete demographics.
- Investigate the Effects of Eligibility Changes:
 - Study how changes in eligibility criteria, such as the inclusion of professional athletes, have influenced competition outcomes and athlete participation.





WORKS CITED

- <https://www.census.gov/newsroom/blogs/random-samplings/2016/06/americas-age-profile-told-through-population-pyramids.html>
- <https://www.nature.com/articles/s41598-022-27259-8#:~:text=Another%20argument%20often%20made%20in,the%20Olympic%20Games%2C7.>
- <https://olympics.com/en/olympic-games/calgary-1988>
- <https://www.kaggle.com/datasets/heesoo37/120-years-of-olympic-history-athletes-and-results>
- <https://www.npr.org/sections/tokyo-olympics-live-updates/2021/08/07/1025452727/how-home-field-advantage-gives-olympic-host-countries-an-edge-and-more-gold-medals>



THANK YOU!

QUESTIONS?

