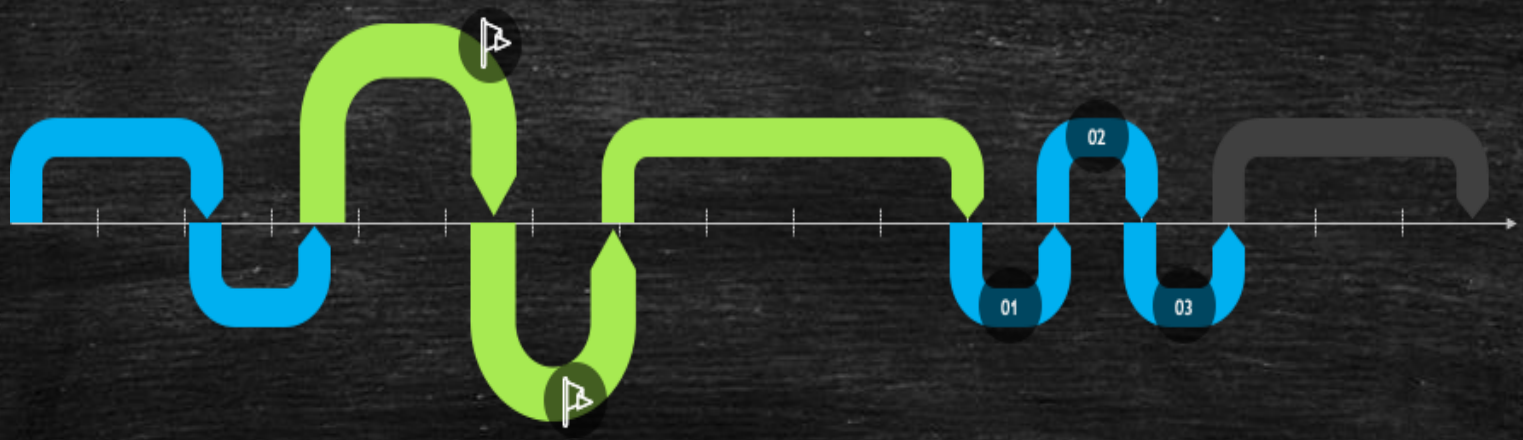
Sports Stats:

An Exploration of the World's Best Athletes

- Taylor Segell
12.27.21



Executive Summary

Who is Sports Stats?

Sports Stats is a sports analysis firm partnering with local news and elite personal trainers to provide "interesting" insights to help their partners.

They are looking for insights (i.e. Patterns/Trends) highlighting certain groups/events/countries, etc. for the purpose of developing a news story or discovering key health insights.

Problem Statement

- Sports Stats need to understand what news surrounding the Olympics is worth reporting on and needs to ensure that the assumptions made are based on factual statistics.
- What can historical Olympics data about health, sports, countries, regions, etc.
- What factors have the strongest correlation to winning a medal.

Questions To Answer

- What demographics such as age group gender, race have the biggest impacts on winning medals?
- Does height or weight play a pivotal role in medals won?
- Are there any countries that you would expect to not perform well based off of GDP, population, etc. that are successful?
- What are some of the metrics we can look to for a predictor variable?
- Is there a shift in the sports that are highly participated in over time?

Hypothesis

1. On a per capita basis smaller and poor countries will probably have a higher medal count and some of them will be outliers.

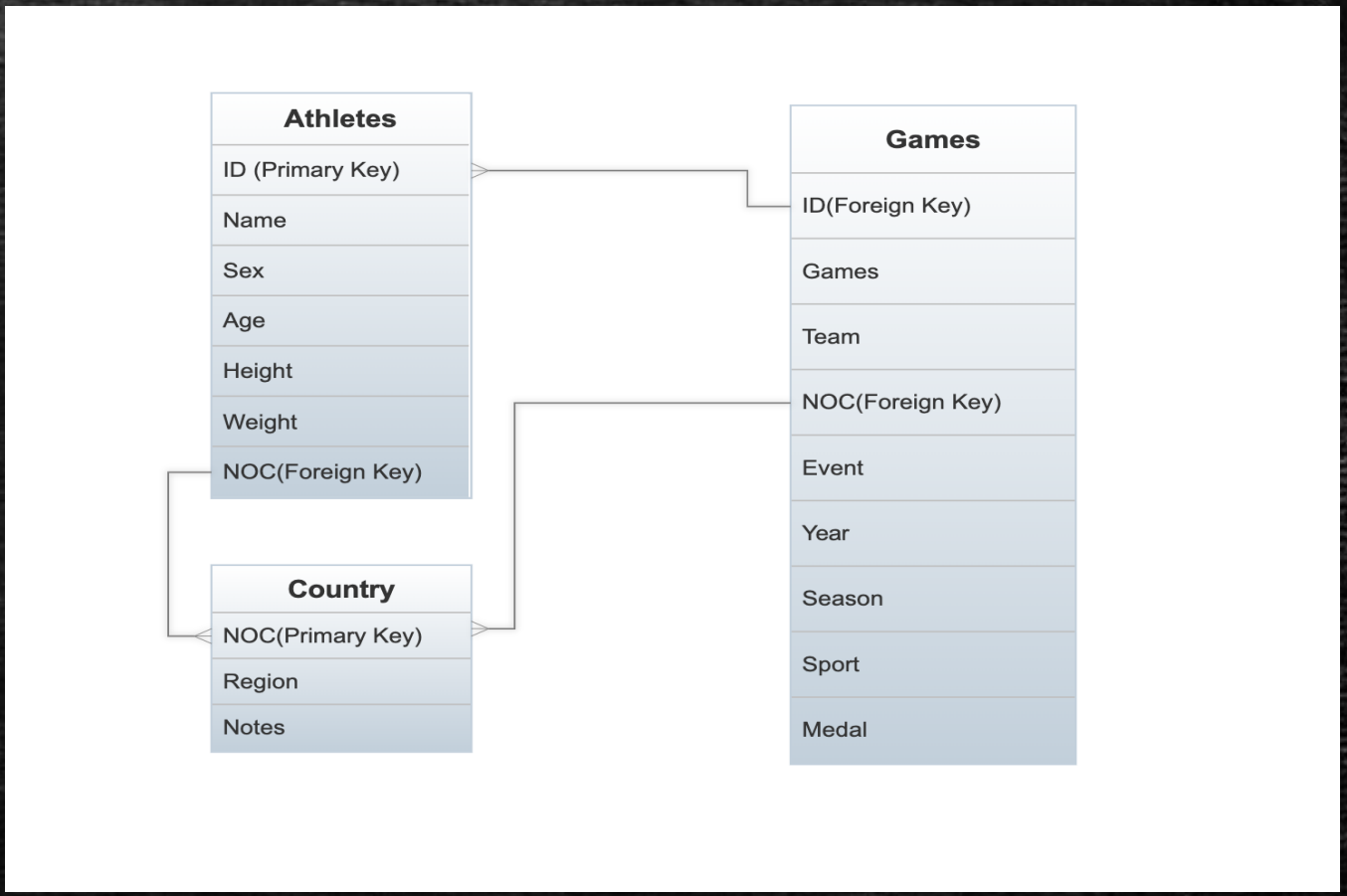
2. Countries farther North will excel in the winter and vice versa for the South in the summer.

3. Competitors that are taller and weigh more will have an advantage.

4. The strongest correlation to medal's won will be height, weight, and countries wealth.

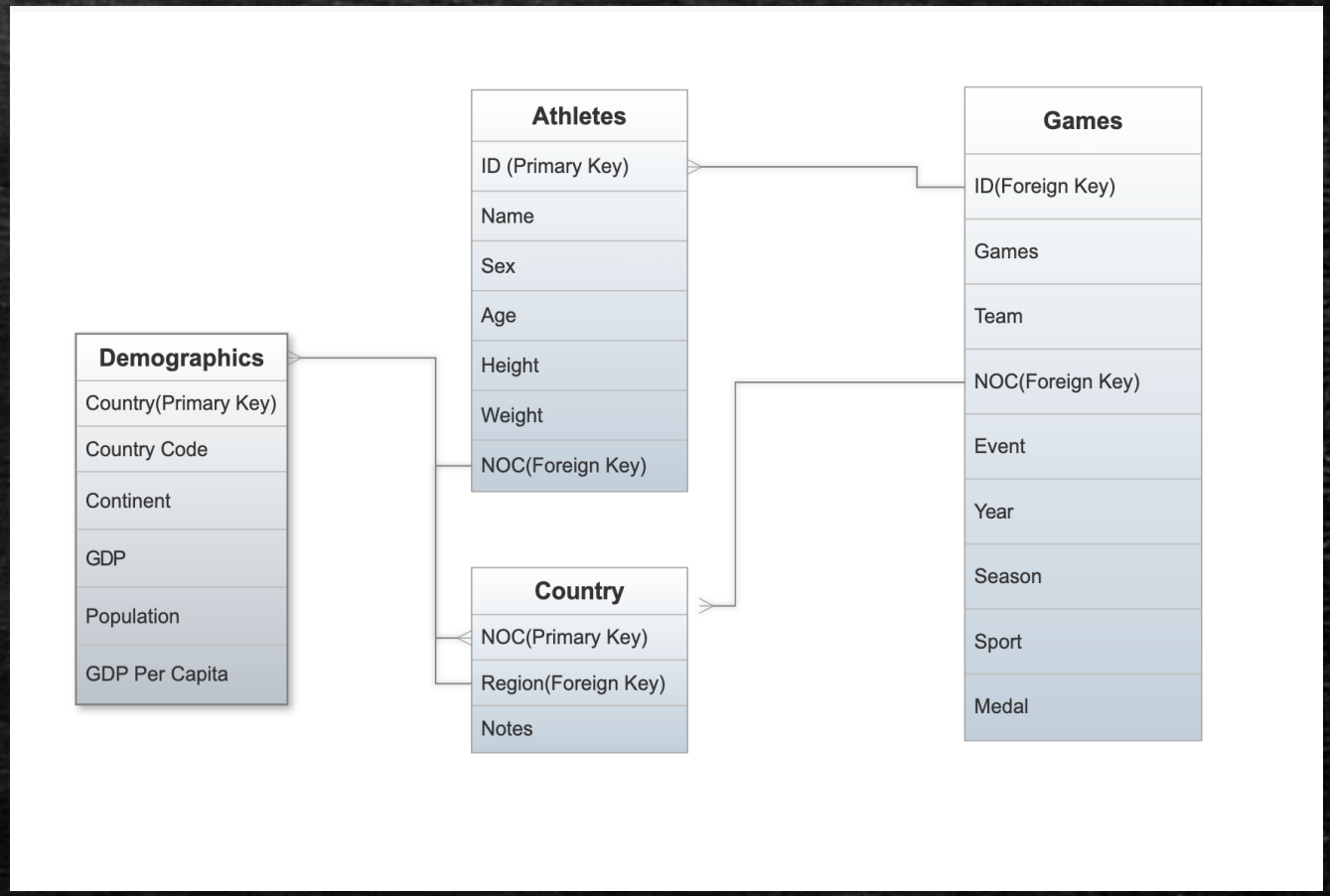
Original ER Diagrams

Original ER Diagrams showing relationships of different tables I split the data into to make analysis easier



Final ER Diagrams

Final ER Diagrams showing relationships of different tables after adding additional data to my analysis



Road Map

Approach

Sports Stats Inquiry

Plan to retrieve Observations/Trends from Historical Olympic Data to assist News leads and Health Insights

Data Preparation

Collecting relevant data that may be of value and fill in any gaps, given data may not address/cover.

Cleaning and Wrangling

Using SQL to make sure data is C.R.A.C.E.R.

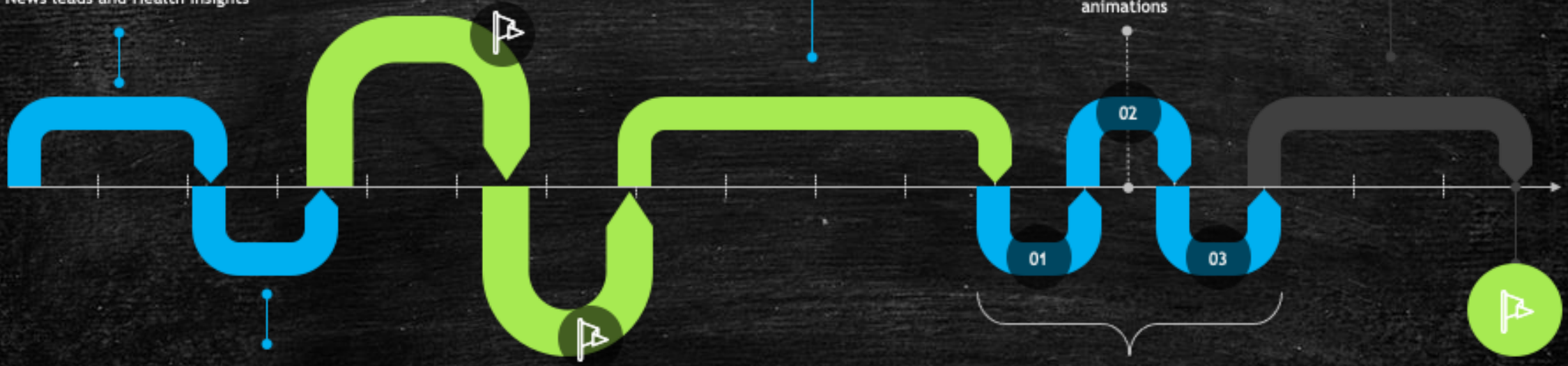
- Complete
- Reliable
- Accurate
- Consistent
- Error free
- 0 Redundancy

Visualizations

Using the data to tell us a story through graphs and animations

Recommendation

Present insights found to Sports Stats and what they can use for accurate reporting



Our Mission

Return in depth analysis of data provided and relevant external data to meet SportsStats needs.

Questions Hypothesis

Formulation of what to look for and basic assumptions from what the data will show us

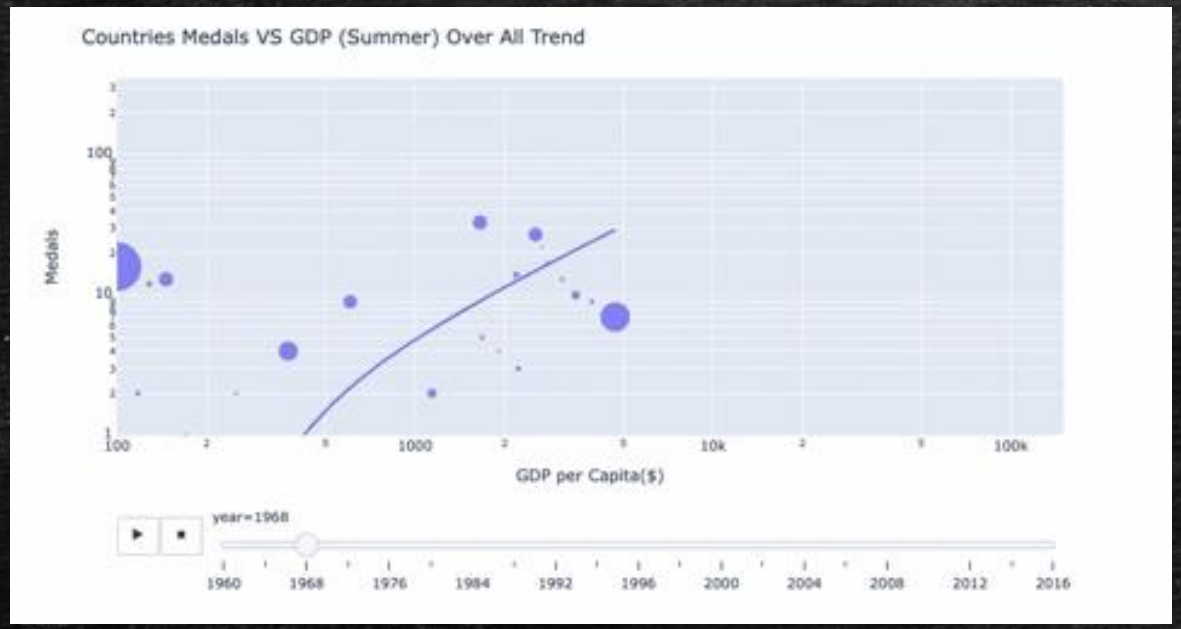
Analysis of Data

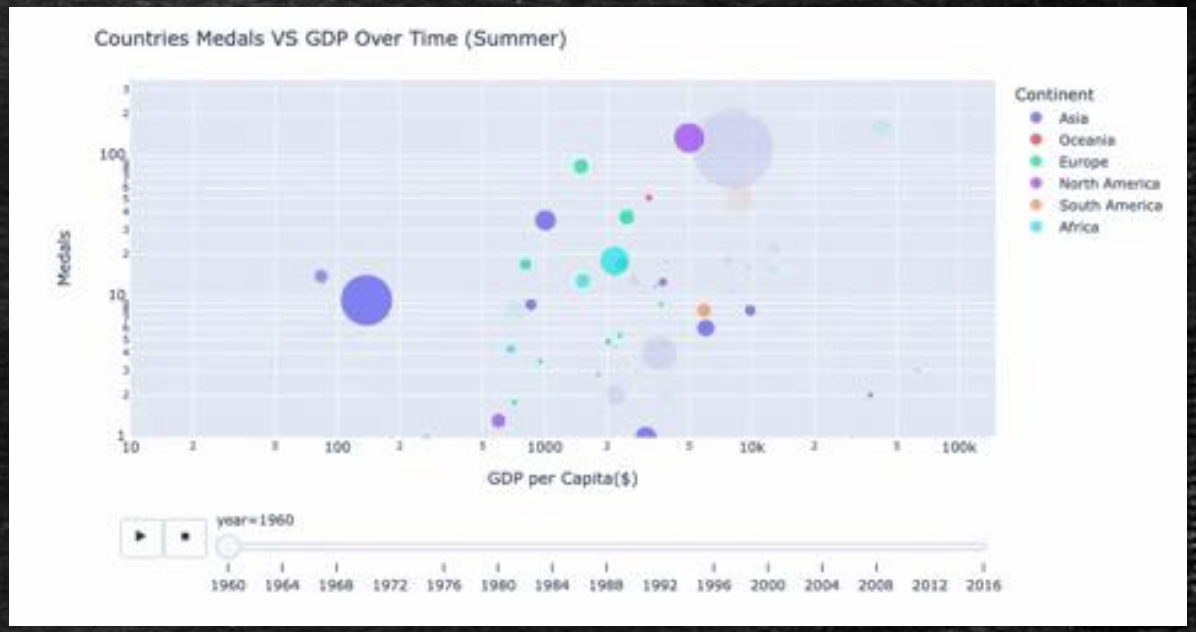
EDA of clean data and statistical/ correlation discovery

Final Calculations, New Metrics and Hypothesis Testing

Insights and Trends Through Visualizations

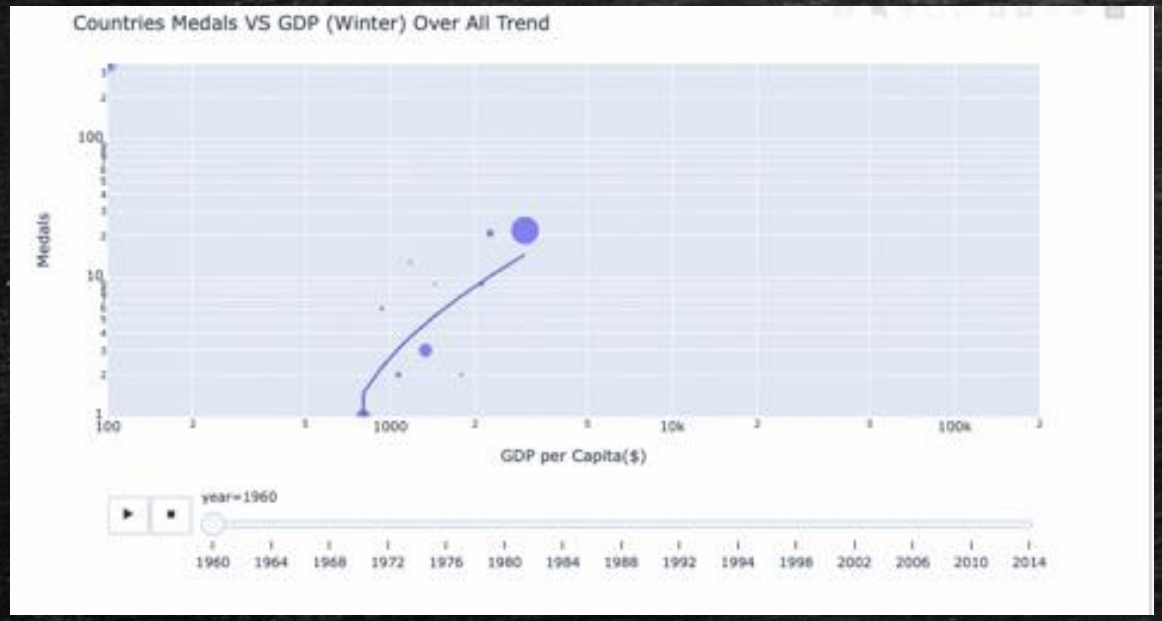
Hypothesis I:
Medal Count VS GDP Per Capita (Summer)

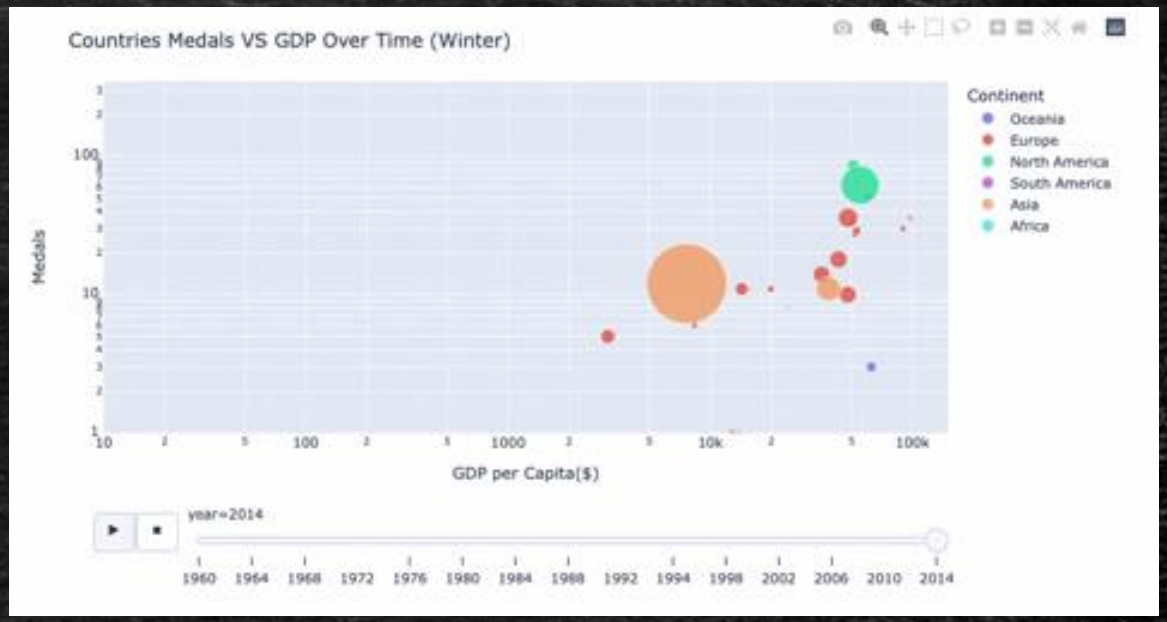




Overall countries with higher GDP per capita leads to higher medal counts in the Summer Olympics. However it does not seem to have the strongest correlation as it did in the past. Over time it has lost as strong as a correlation.

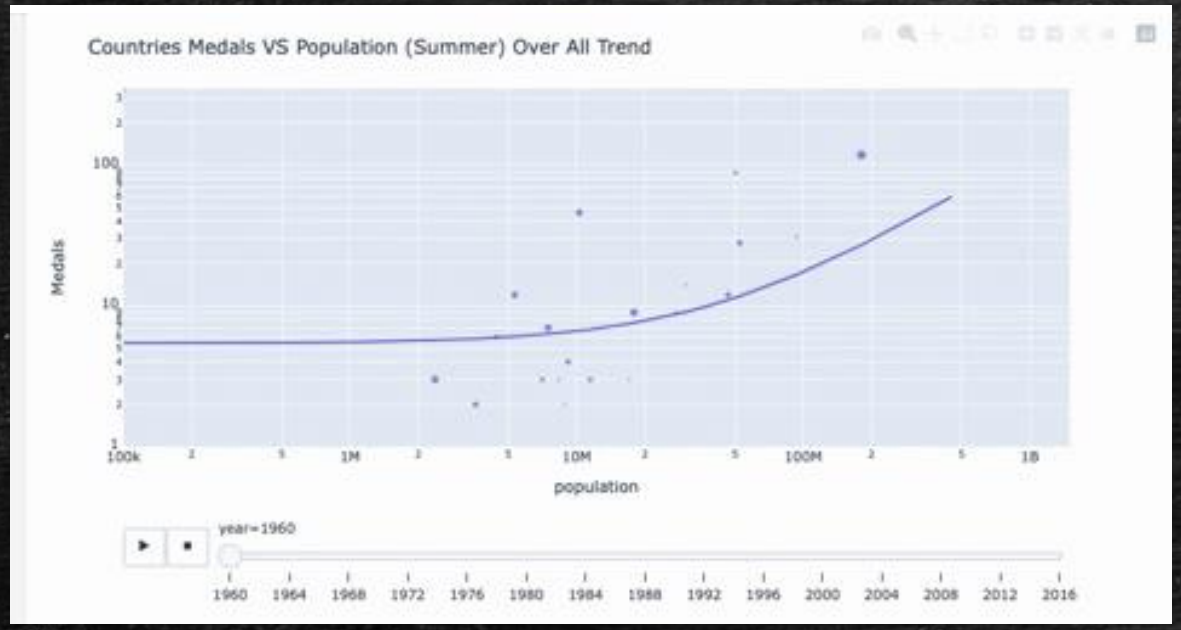
Medal Count VS GDP Per Capita (Winter)

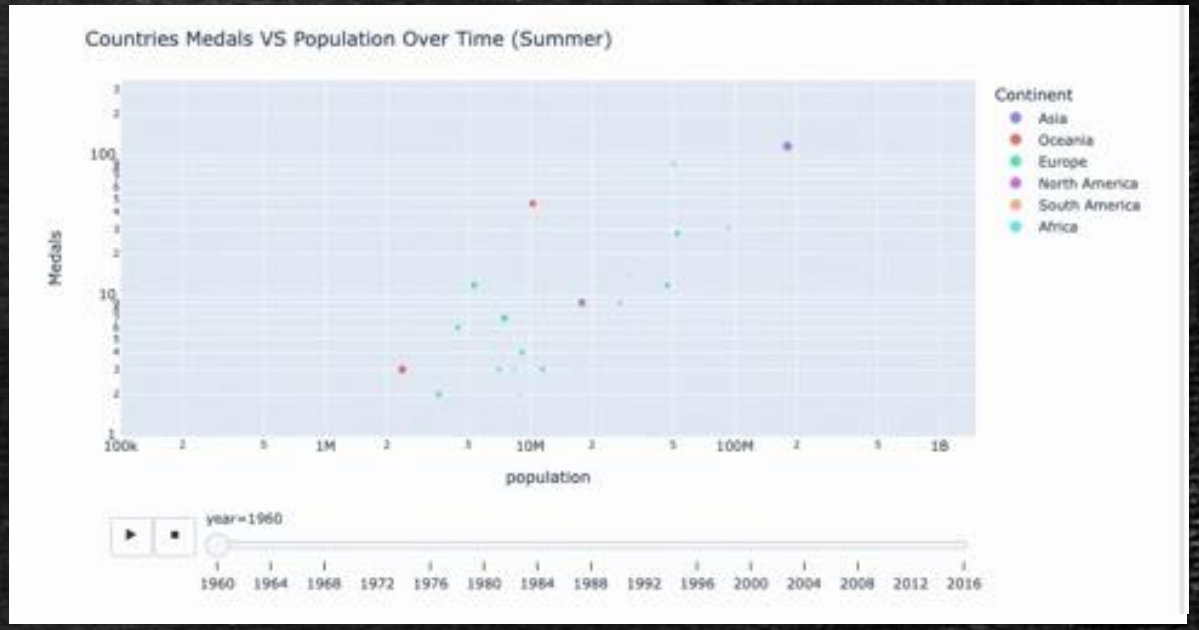




The Winter Olympics tends to have a stronger correlation with higher GDP per capita countries taking home medals. This probably has to do with the fact the majority of developing countries are near the equator while the more developed countries are up North.

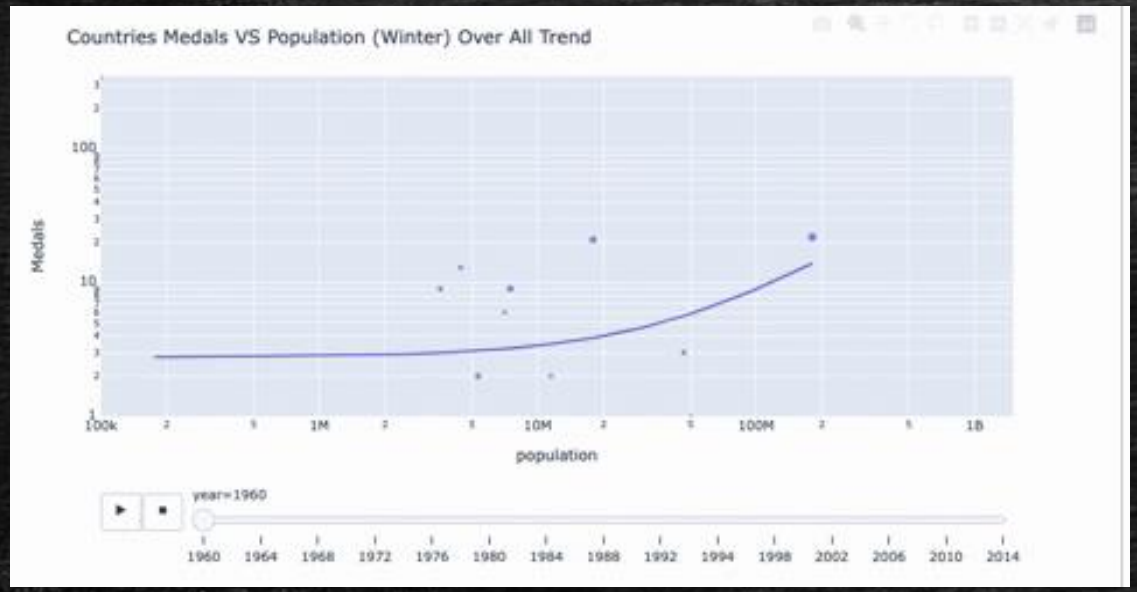
Medal Count VS Population (Summer)

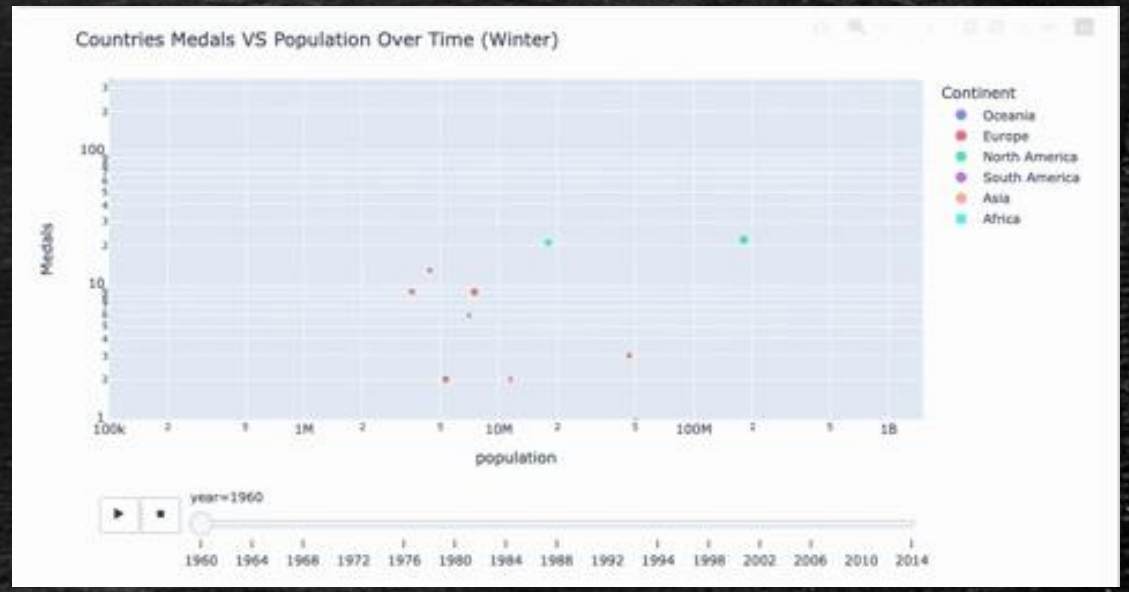




Overall countries with higher populations leads to higher medal counts in the Summer Olympics. This is a pretty obvious statistic when you look at it from an overall stand point and not Athletes competed per medal.

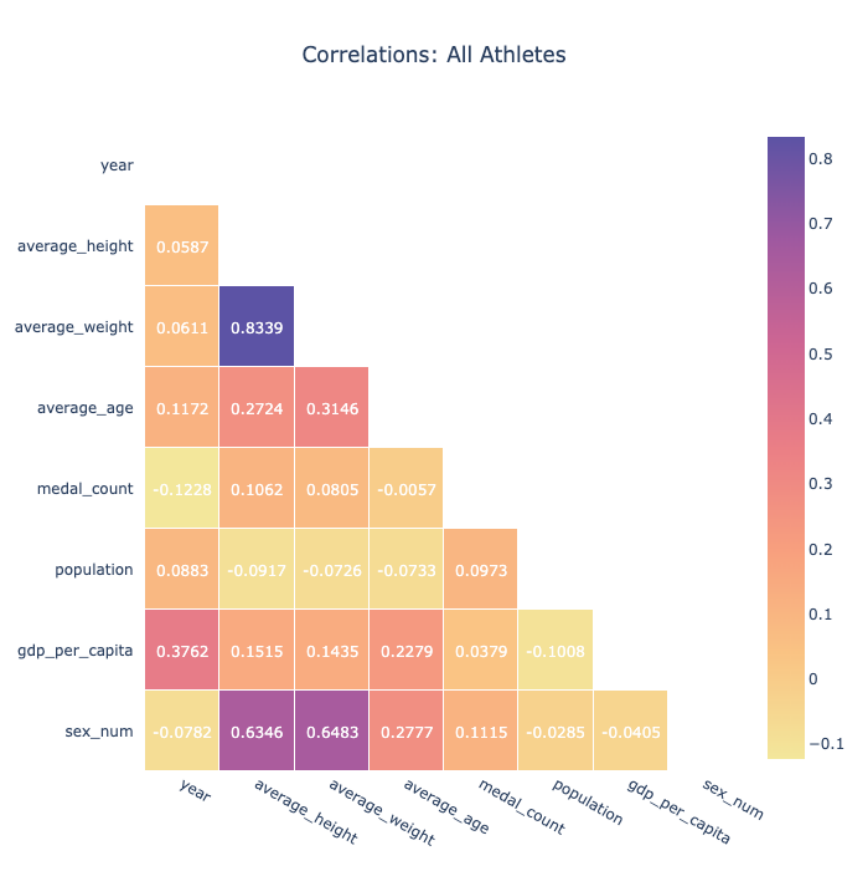
Medal Count VS Population (Winter)





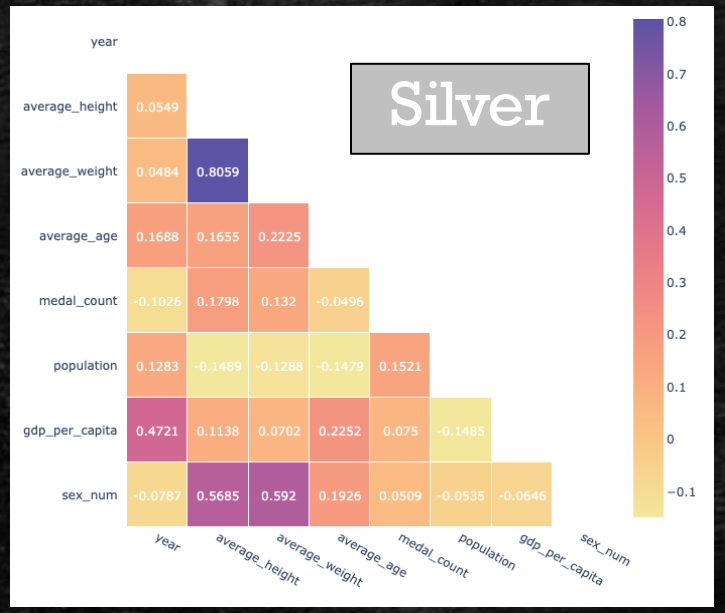
The Winter Olympics almost had a negative correlation with higher populated countries taking home medals. While this has changed over time to be more balanced it is the only instance found with a negative trend of some sort.

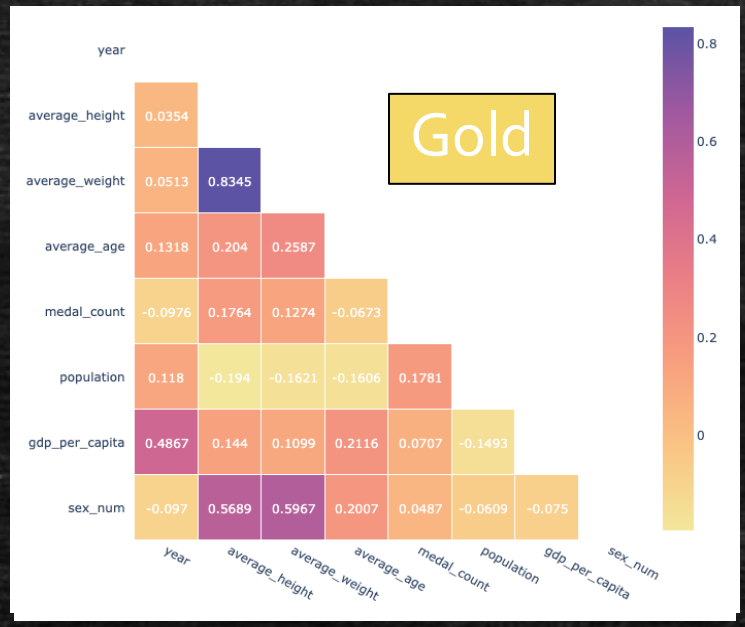
Overall Correlations of Athletes

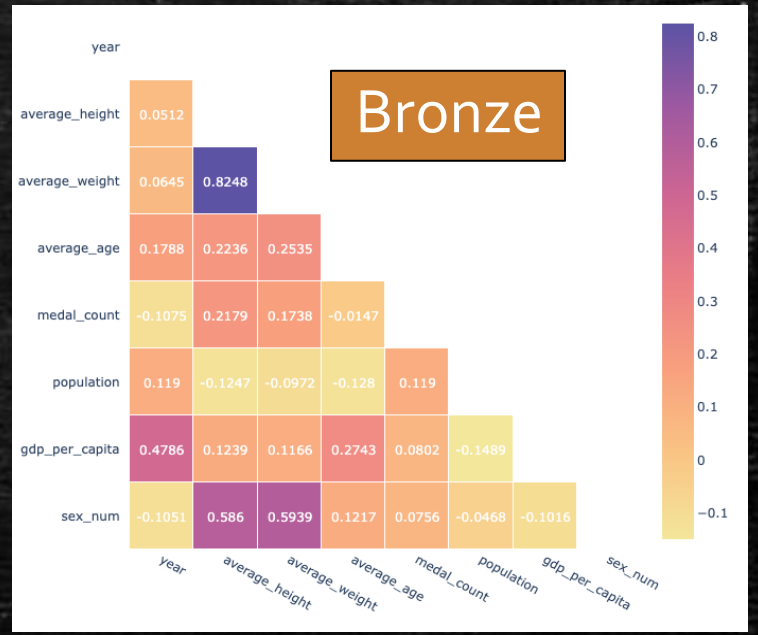


Whil it is obvious height and weight will have the strongest correlation, it is not to surprising that they would also have the strongest correlation for medals one. Olympians are typically physical specimens. The only other major correlation with medal count is population. GDP per capita seems to play little role.

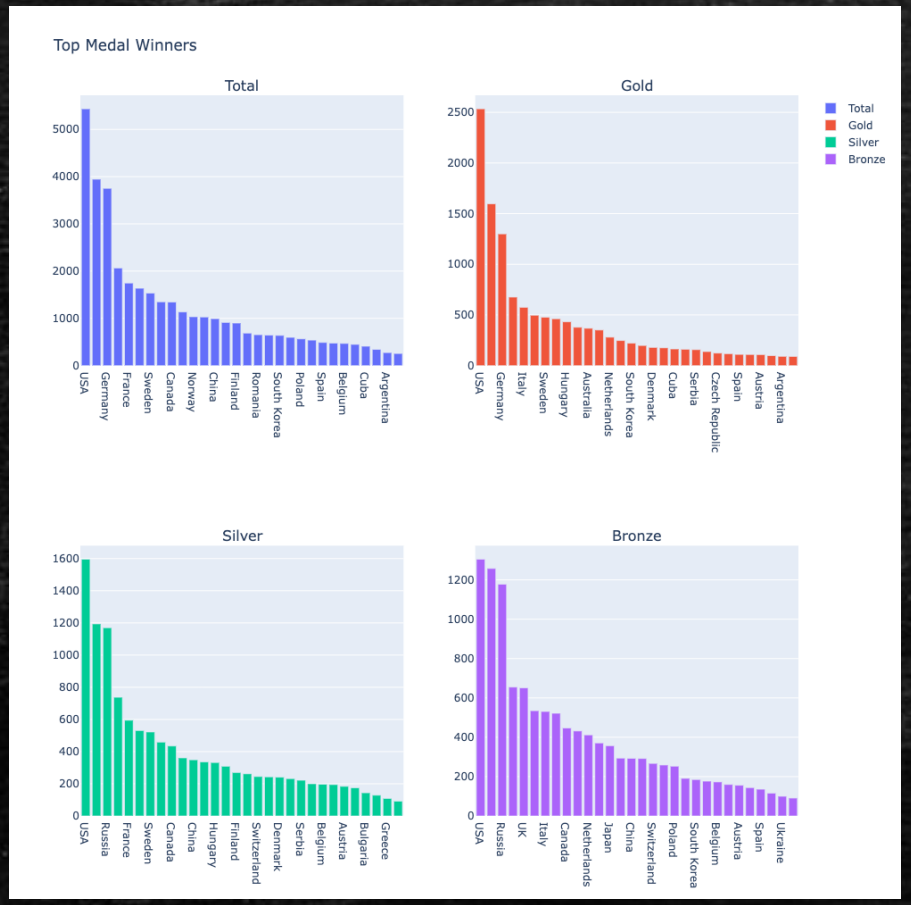
Correlations By Medal Recipients

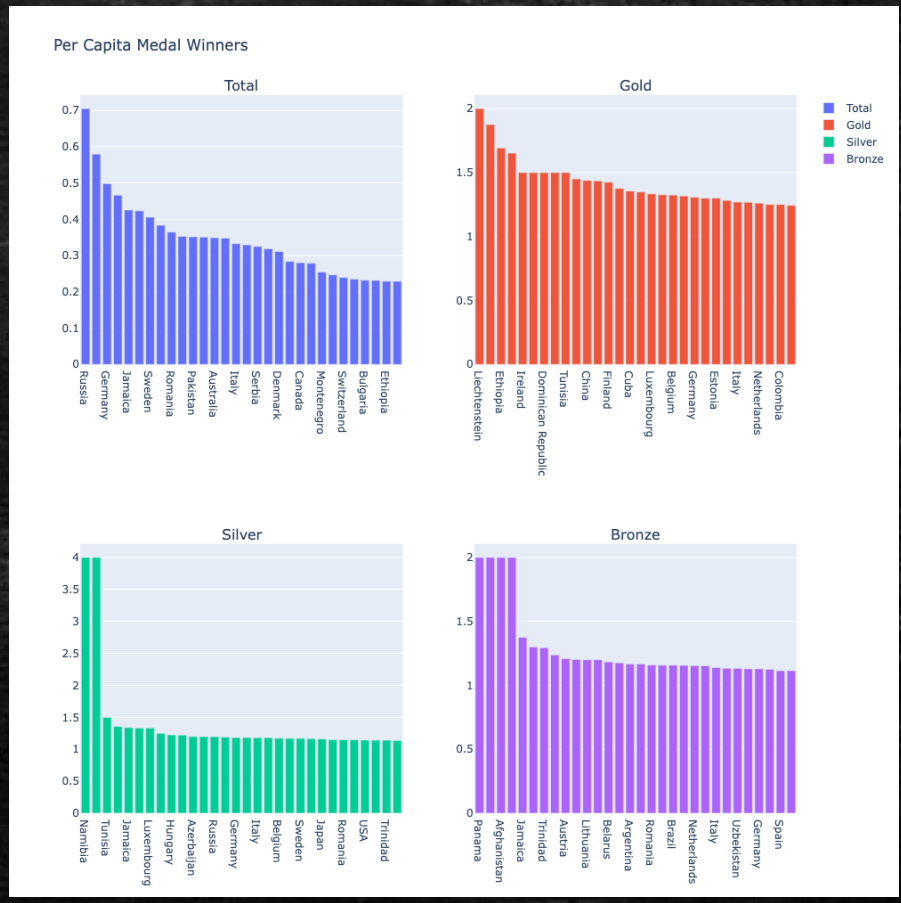






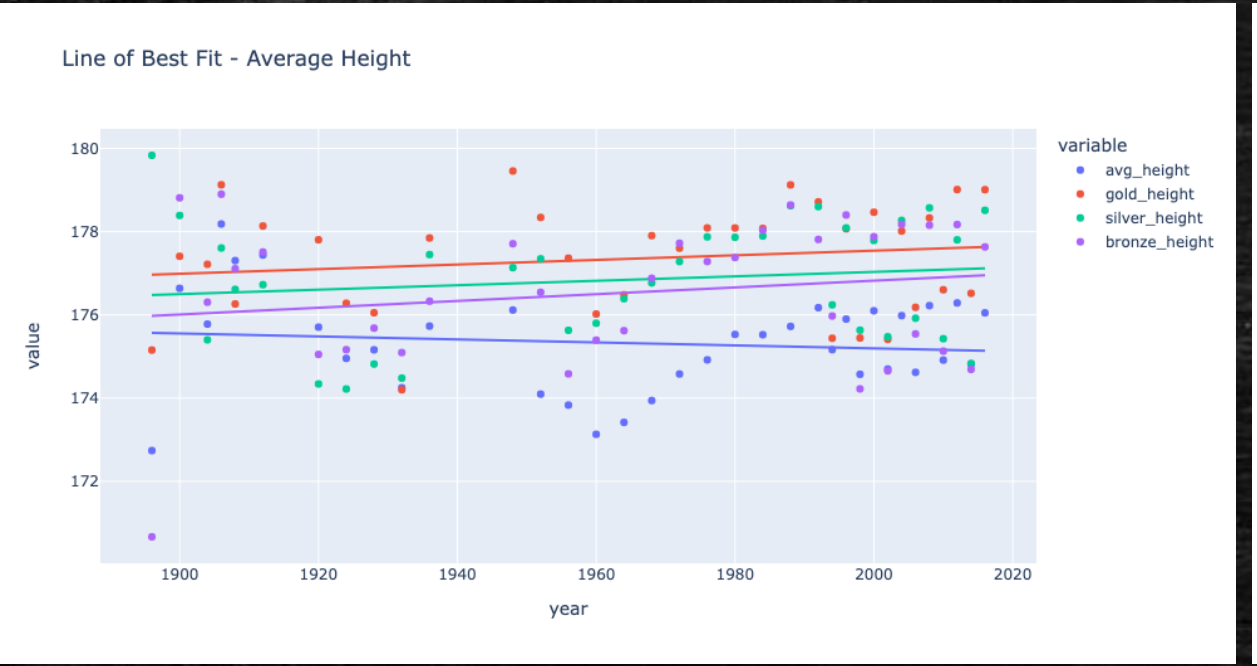
Top Versus Per Capita Winners

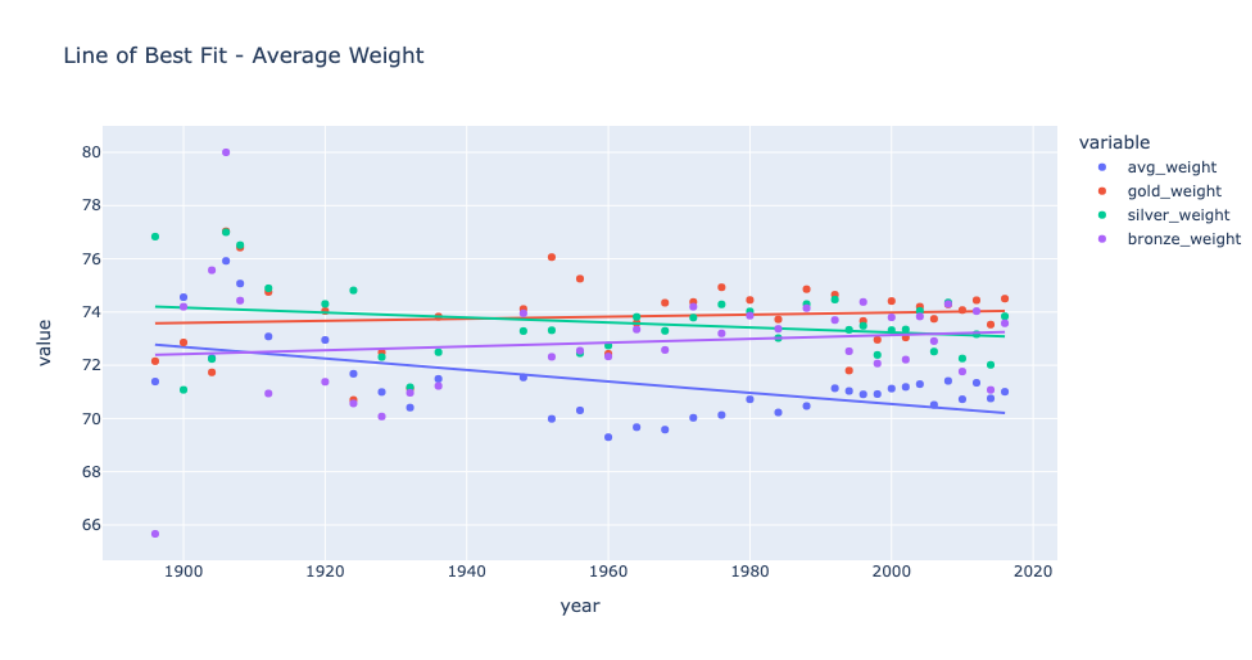




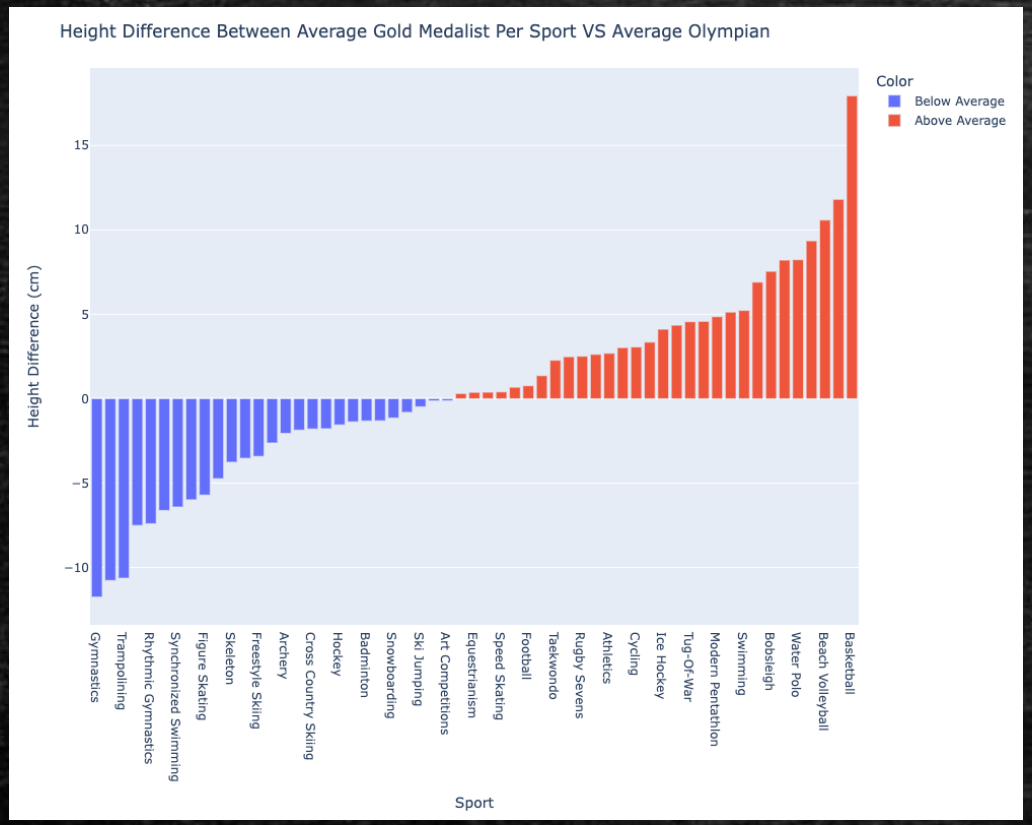
Average Height vs Weight by Medal

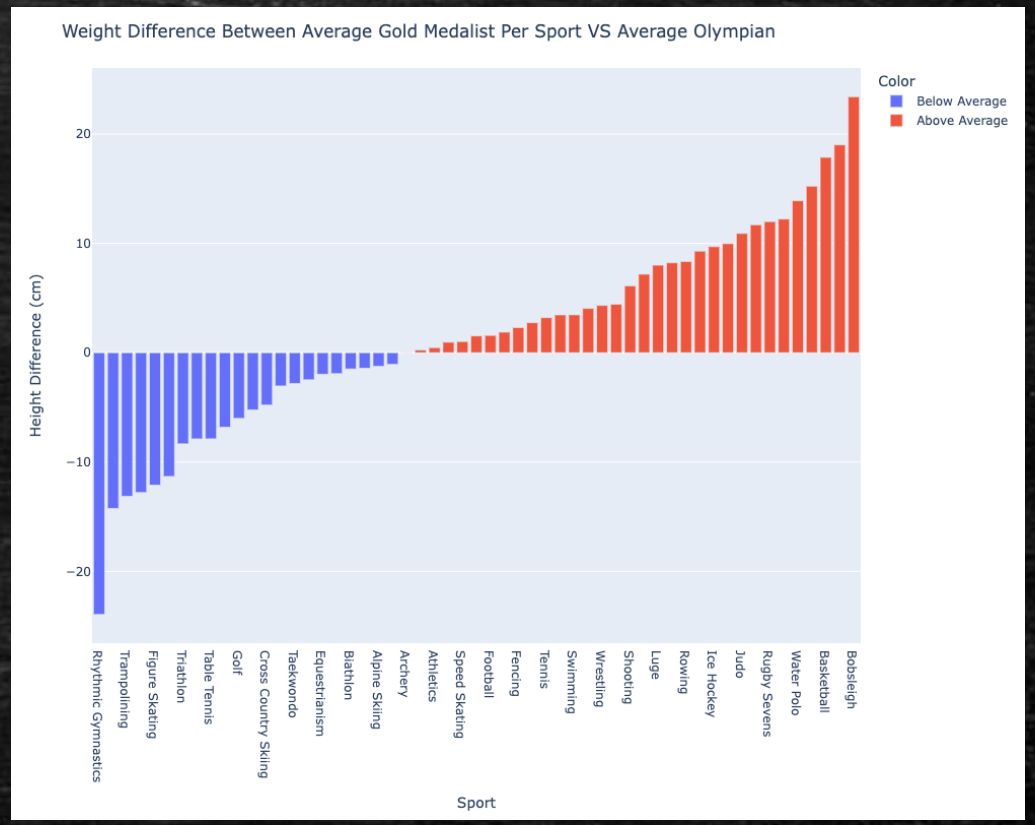
As we found earlier height and weight both play pivotal roles in the efficacy of an athlete. Gold winners are typically the tallest of the bunch and weigh more than the average.





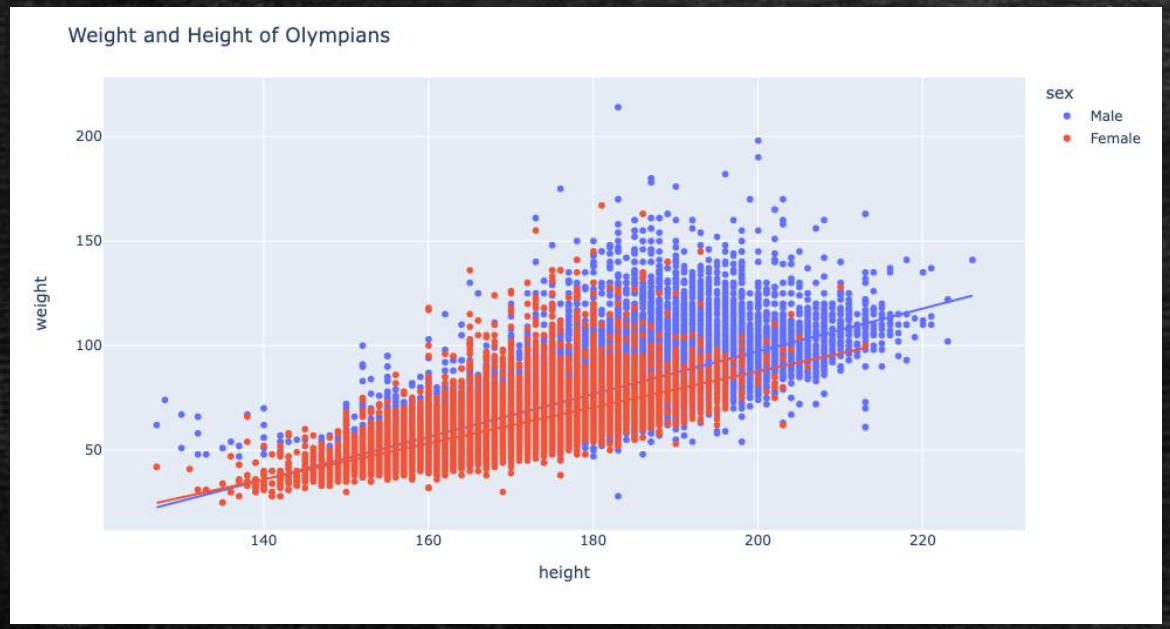
Olympian Verse Gold Medalists

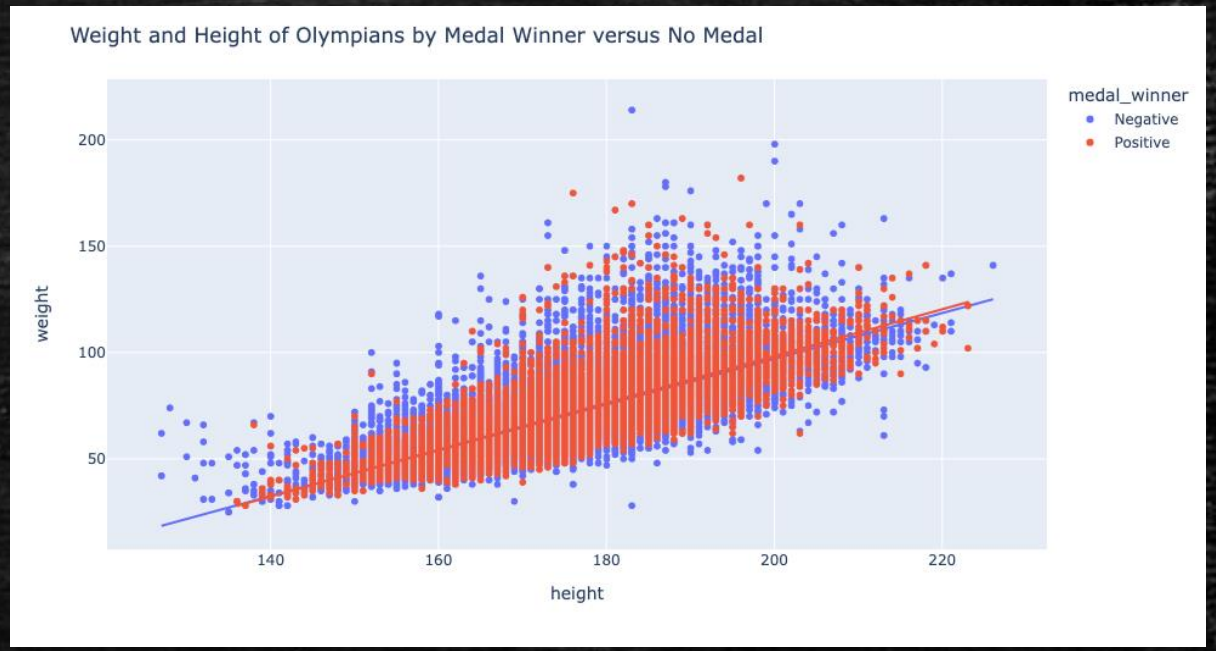




While it is true the gold medalist is typically larger than the average Olympian, when we break it down by sport we find that certain disciplines contradict that narrative such as Gymnastics, Figure Skating, Trampolining, and Skeleton

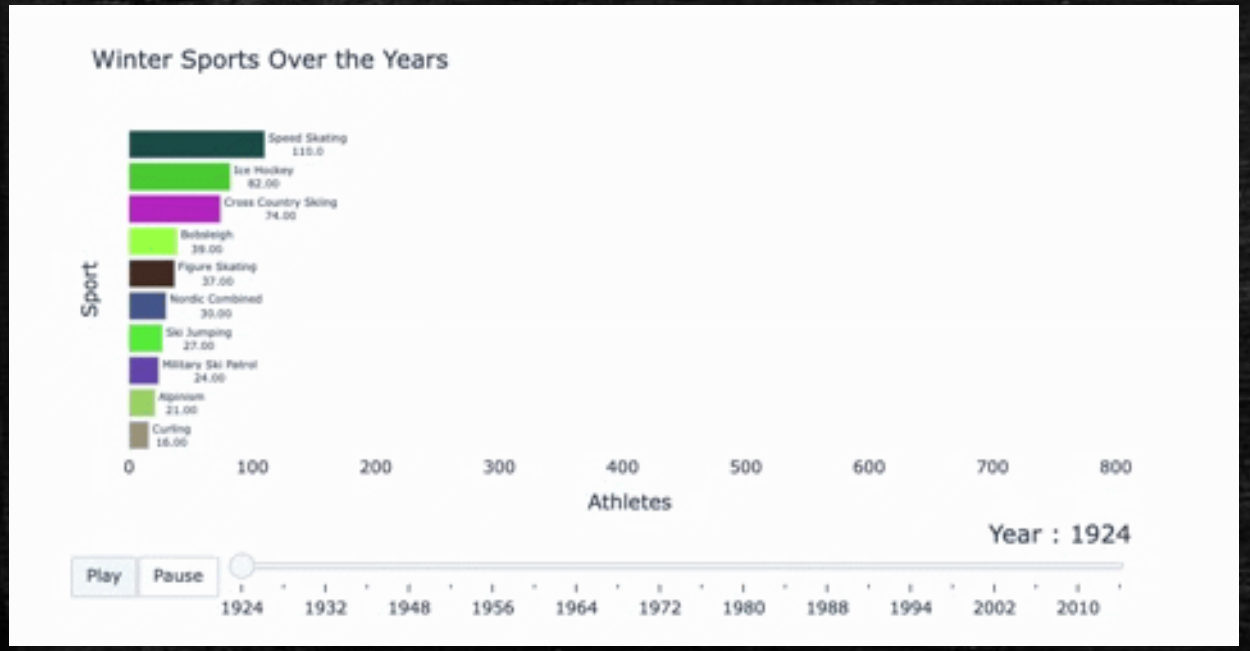
More on Height and Weight

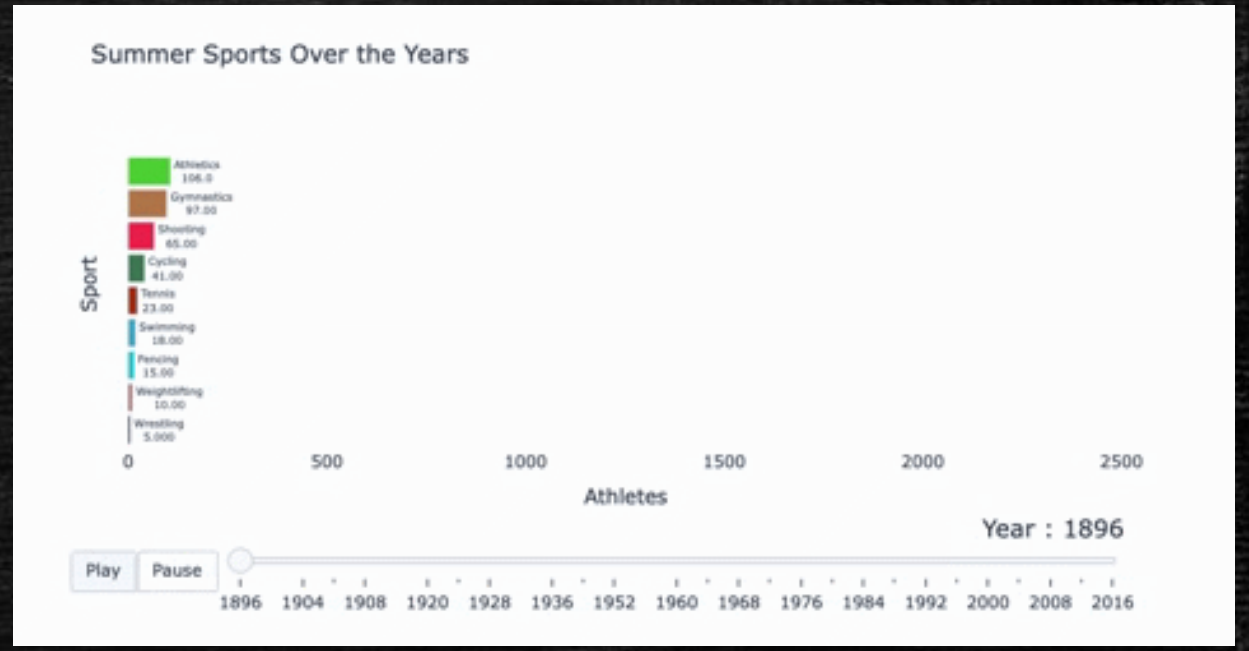




Most Popular Sports at the Games

From a participation standpoint the top sports in each of the summer and winter games have stayed relatively consistent over the years. Athletics(Track and Field) and Gymnastics have both stayed at the top while Swimming has probably the highest growth for summer games. As for winter sports , skiing has remained dominant while Snowboarding has had the most growth since the beginning of the games in 1924.





Revisiting our Hypotheses

1. Smaller/less developed countries will have a lower medal count.
   - RATING = PARTIALLY TRUE
   - For the most part this is true but smaller countries tend to do just as well when judged per capita athletes

2. Countries farther North will excel in the winter and vice versa for the South in the summer.
   - RATING = TRUE
   - Since many equatorial countries do not compete in winter Olympics this is a given.

Competitors that are taller and weigh more will have an advantage.
   - RATING = MIXED
   - While this is true broadly speaking, it is highly dependent on the event an athlete is competing in.

The strongest correlation to medal's won will be height, weight, and countries wealth.
   - RATING = PARTIALLY TRUE
   - Height and weight were the top two factors however BMI was a better predictor and population had a higher correlation than wealth for the summer Olympics and vice versa for the winter Olympics

Conclusion

After going through this analysis I was able to prove my hypothesis true to a degree. The last two about weight and height we're quite accurate overall and even the one regarding the season and which part of the world would win primarily the North in the Winter. Also per capita smaller countries do tend to win more medals than larger ones like China, USA, Etc. Especially at the winter games

The new Metrics I created regarding height and weight difference between average Olympian and medaling Olympian by sport was definitely need for a more rounded view. It was definitely a game changer

.

Recommendations

To those of you at SportStats. My recommendation would be to focus your stories and research on more Underdog based countries. While they do sound like they are unlikely to win being underdogs. The data shows they actually have just as good at the chance and sometimes even greater coming out on top. As an American I do believe America is the best of all but that is my bias, like you are likely to believe your homeland is the best. When you truly look at the historical Olympic data, every country and athlete has a shot to win gold and everyone wants to root for the underdog. There is nothing better.

Further Exploration

Honestly, I could probably go through a million other questions I have at the moment but I will save that for later like breaking it down amongst events in who excels in which events weather certain countries excel at certain events we're certain demographics to. There is an endless amounts of analysis to be found in this data.

The main thing needed though for a deeper analysis would be more data in relation to individual athletes and more regarding their country of origin. I would be curious to see the data in regard to amateur versus professional status in the athletes respected discipline.

-