**A.J. Lozano**

**4/15/2019**

**Readme:**

* **Which dataset you chose. If not part of Udacity's dataset options, document the source of your data.**

I chose to explore a data set from Kaggle that held a 120 year history of the Olympics. I’ve always enjoyed watching the games and felt like the data set would be fun to explore.

The data was gathered from the following link:

<https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results>

* **Main findings from the exploratory data analysis, and how you chose the results to put in your explanatory analysis.**

My purpose for exploring the Olympic Dataset was out of curiosity and a joy of watching the game. I also wanted to use the analysis as a reference that I could look back on from time to time.

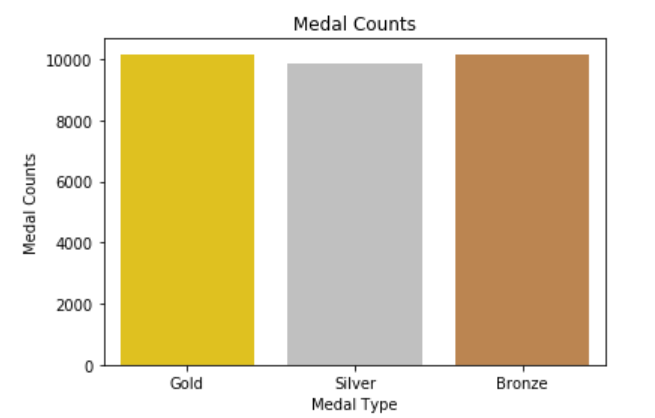
I wanted to learn about medal, country, and athlete characteristics, what’s changed over time, differences between Summer and Winter Games, differences between sports, etc.

The ‘athlete\_events\_csv’ data frame initially held 271,116 entries within 15 columns. I copied the original data frame to work on manipulation and univariate, bivariate and multivariate exploration.

I dropped all null rows for a tidy and complete dataset. I ended up with 30,181 rows.

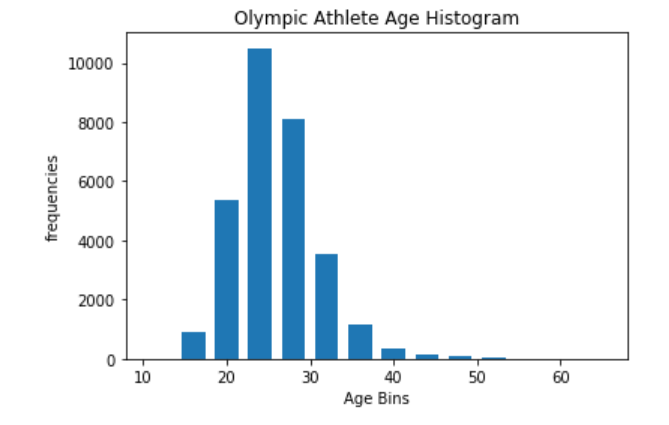
**Exploration:**

I started my exploration by understanding the count and type of medals within the data frame.

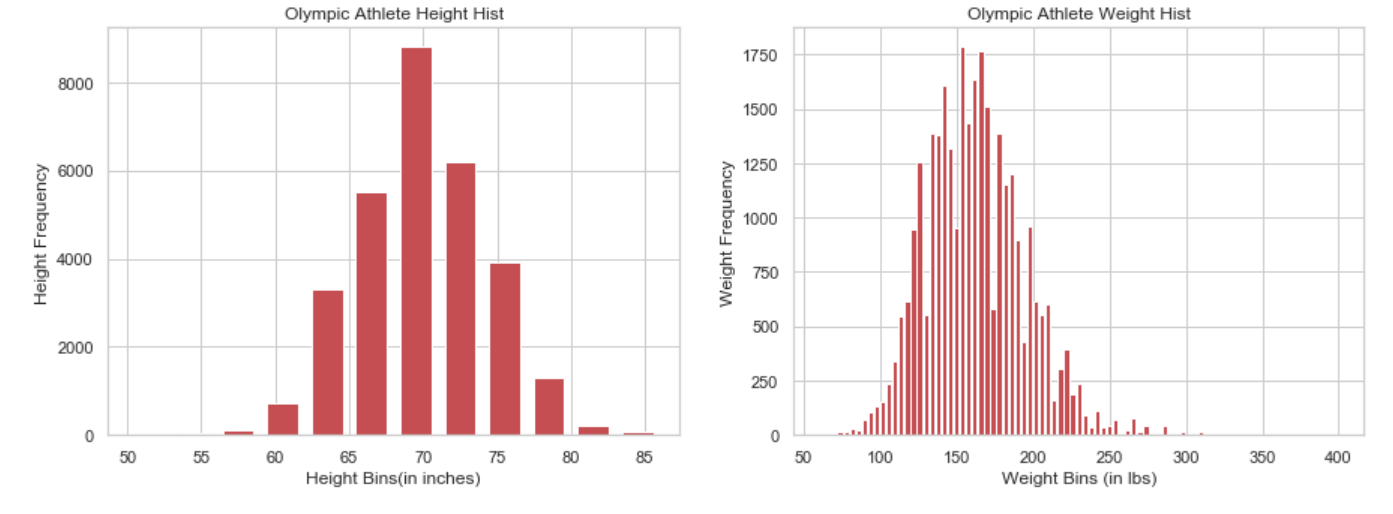


Gold accounted for 10,167, Silver totaled 10,148 and Bronze held 9,866.

From there I wanted to better understand age characteristics of the athletes. I felt like a histogram would easily show the frequency and age ranges of competitors.



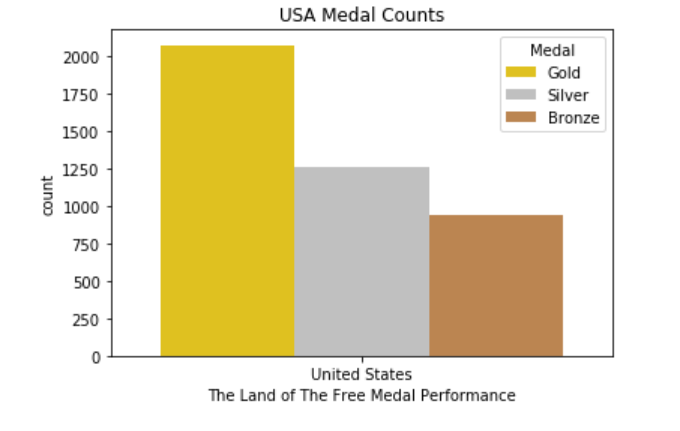
Most athletes are between 20 and 30 years of age. The average age is just over 25 years old.

Naturally, I wanted to get a better picture of height and weight characteristics. I again went to the histogram. 

The average height is just under 70 inches, while the average weight is about 163 pounds.

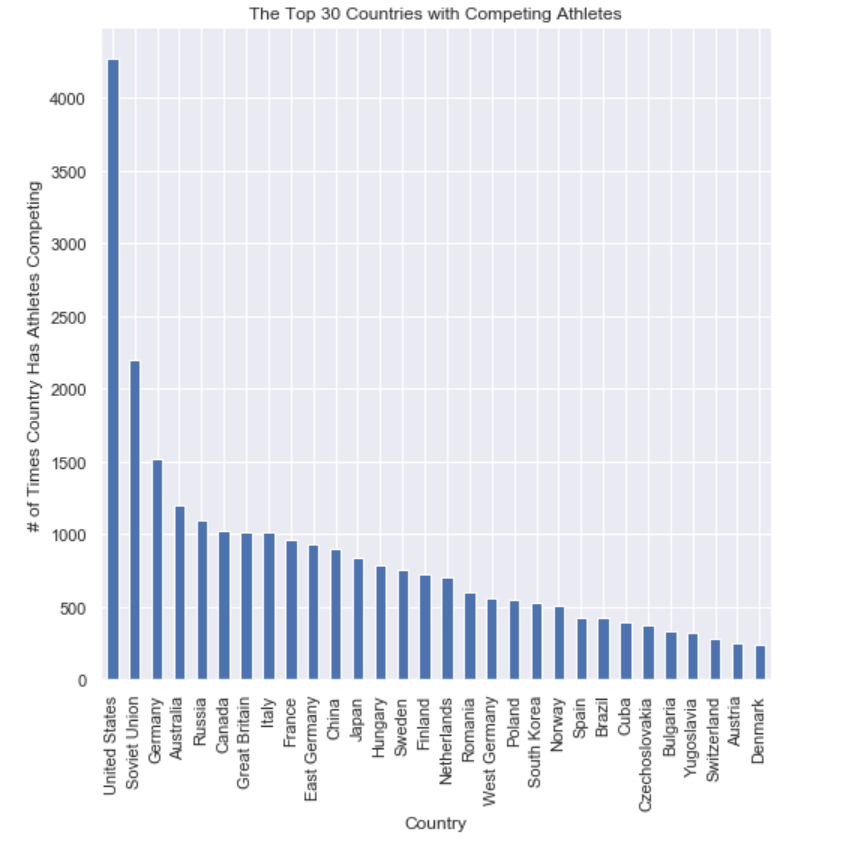
So who’s competing in the games…

The data frame contained 263 unique participating countries and 20,767 unique athletes. The U.S. has sent the most athletes with an amount of 4,273. Once I knew the U.S. sent the most participants I wanted to know what our hardware looks like.



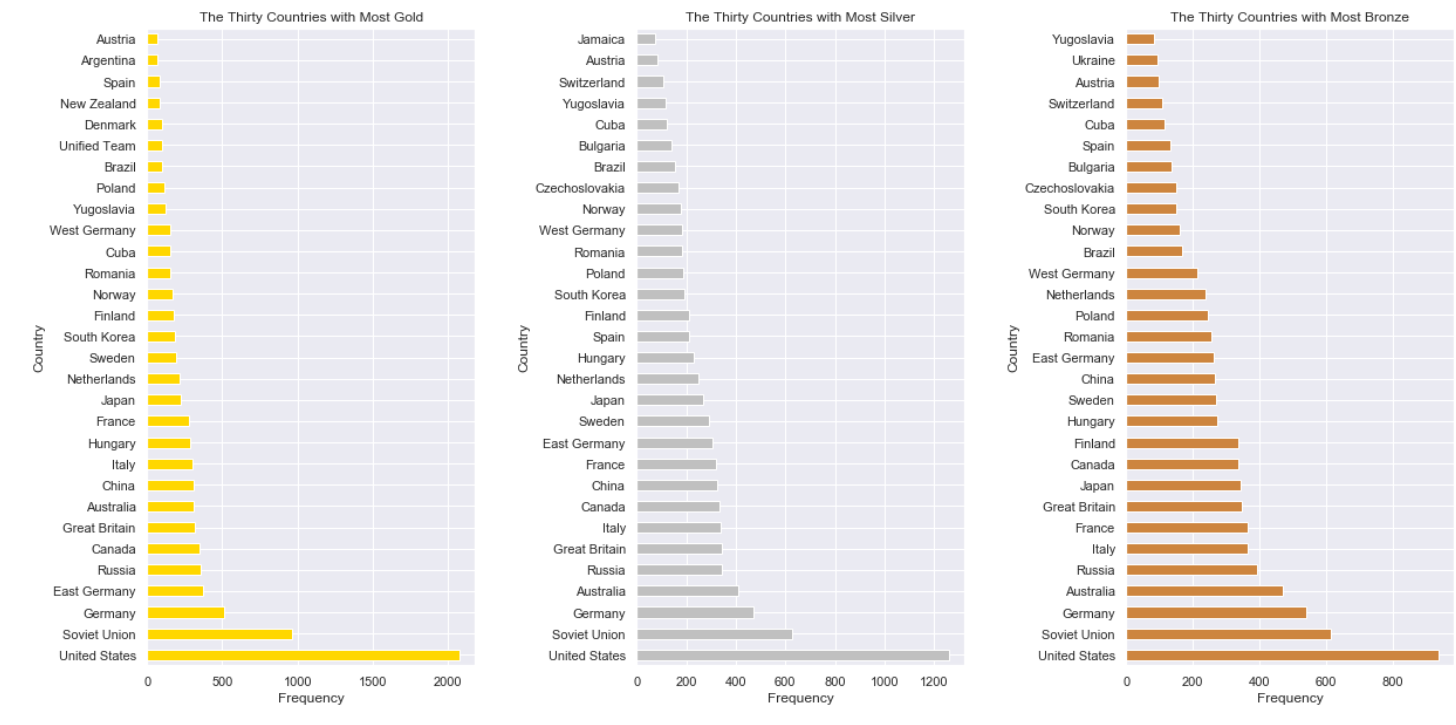
The U.S has over 2000 gold medals, 1250 silver medals and almost 1000 bronze medals.

I moved on to looking at the competition and how often the U.S. goes against different countries.



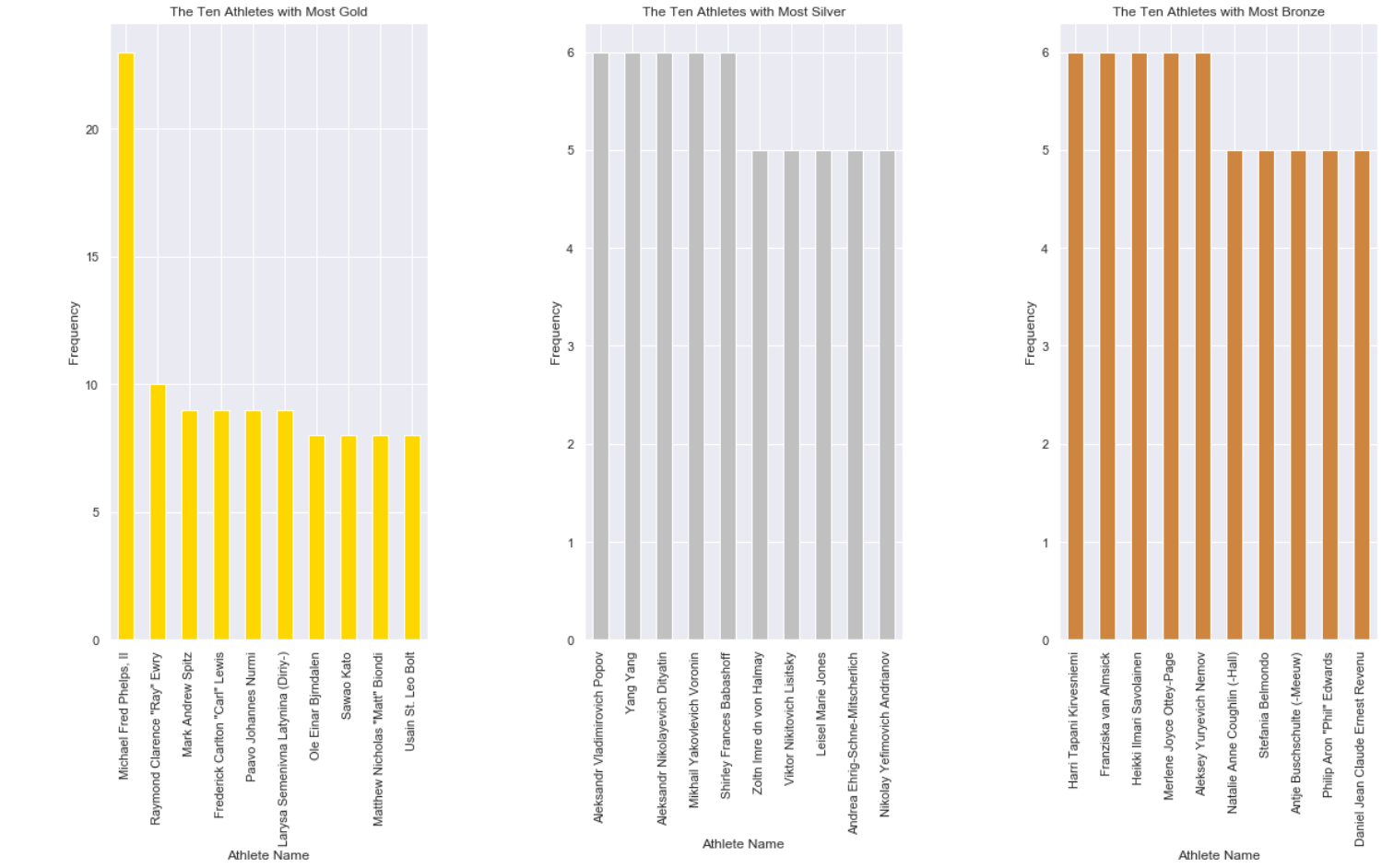
It looks like the Soviet Union and Germany are our biggest Olympic foes.

So what do the trophy cases of each country look like?



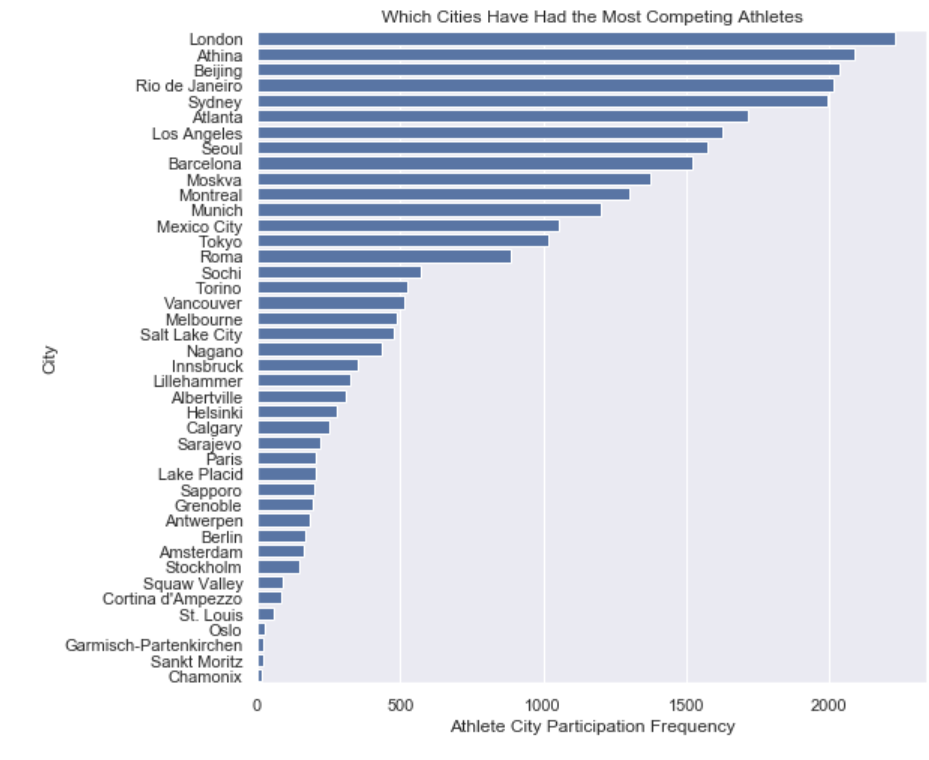
The U.S dominates in Gold, Silver and Bronze. The Soviet Union and Germany are second and third in all respective medal categories. Other notable countries include Australia and Russia.

Once I understood how countries performed, I wanted to see who the most decorated athletes are.



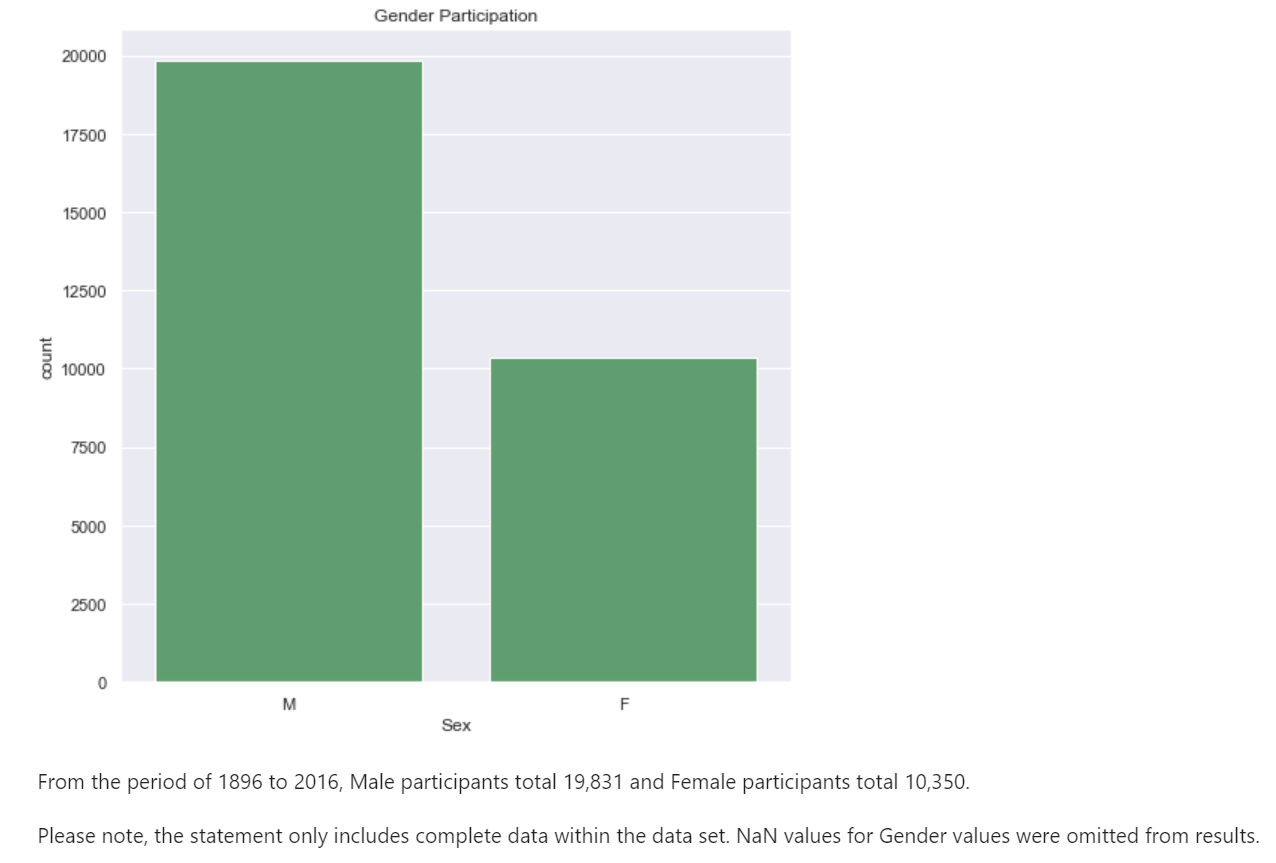
Michael Phelps has the most gold medals. Usain Bolt Is the 10th winningest gold medal athlete.

The Olympics are held in different countries and cities all throughout the world. I was curious to know what city has hosted the most athletes. There are 42 unique host cities.

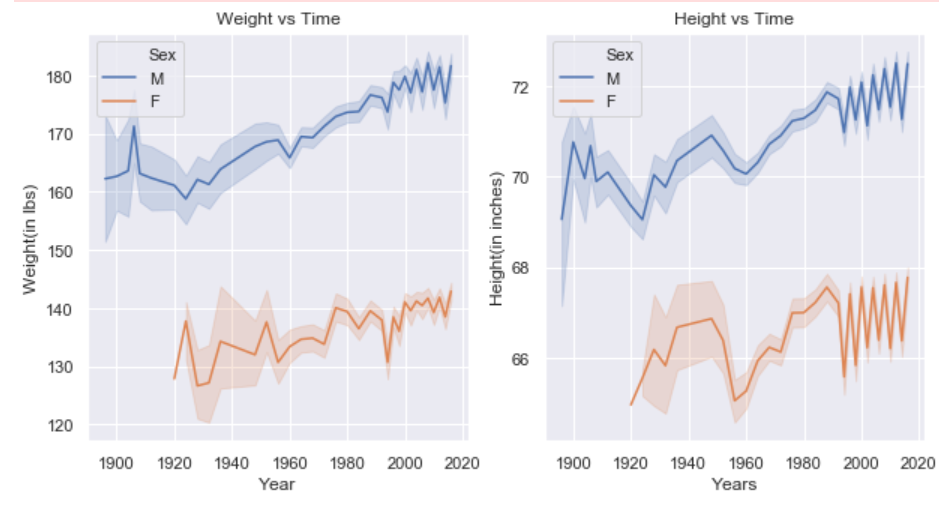


London has hosted the most athletes and Chamonix the least.

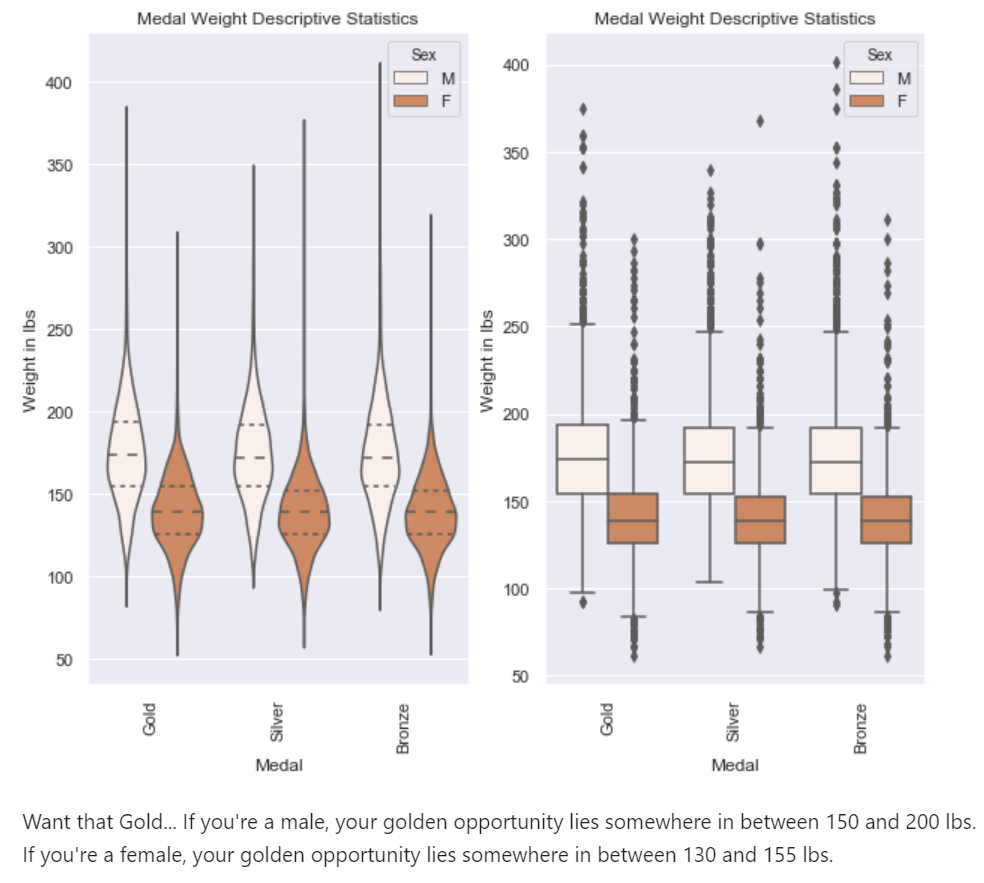
Gender participation and characteristics were a focus of the exploration. I thought I would start looking into that by understanding gender participation frequencies with the data set.

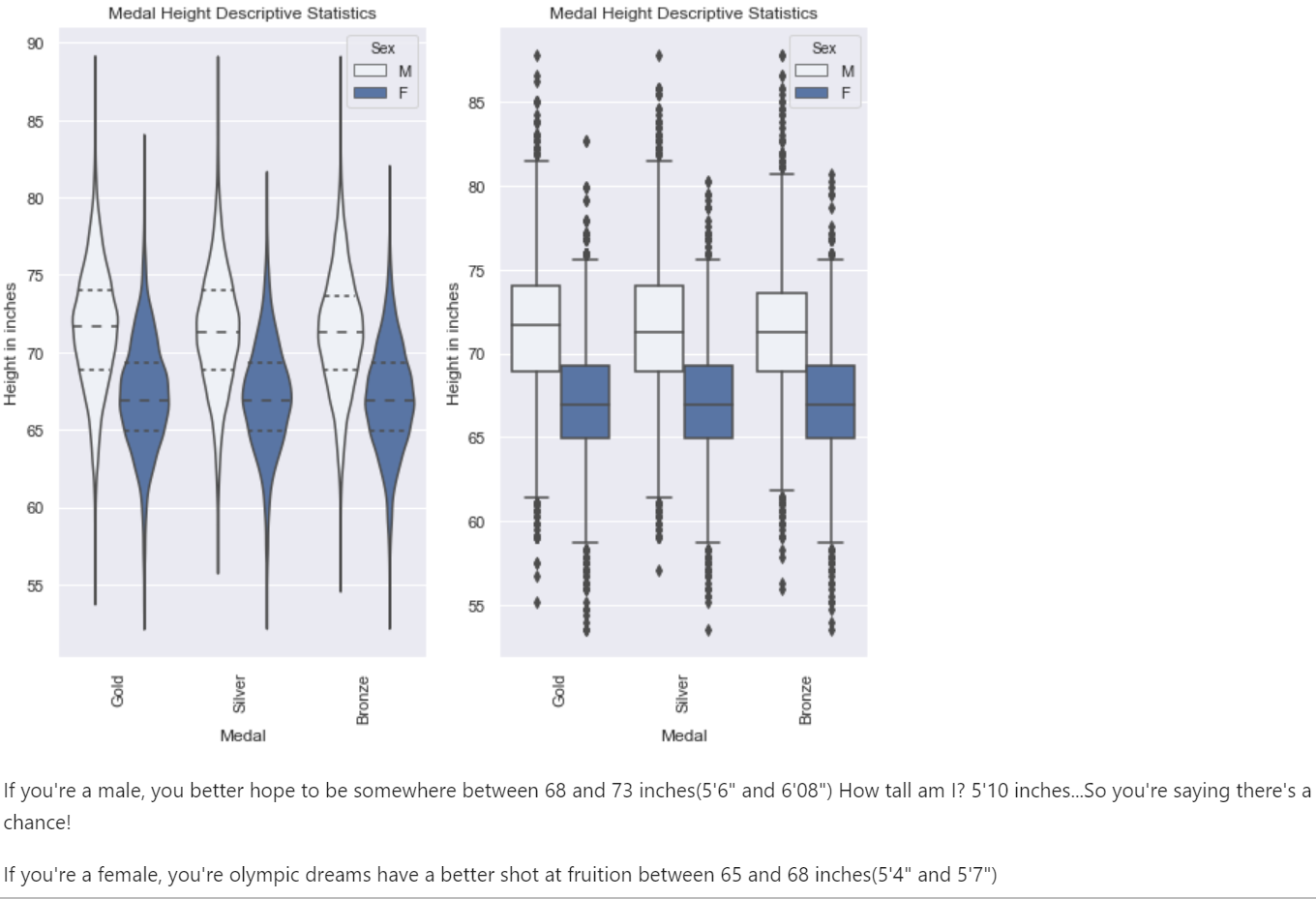


I started the bivariate analysis by getting a picture of changes over time. How had athletes bodies, height, and weight changed over time?

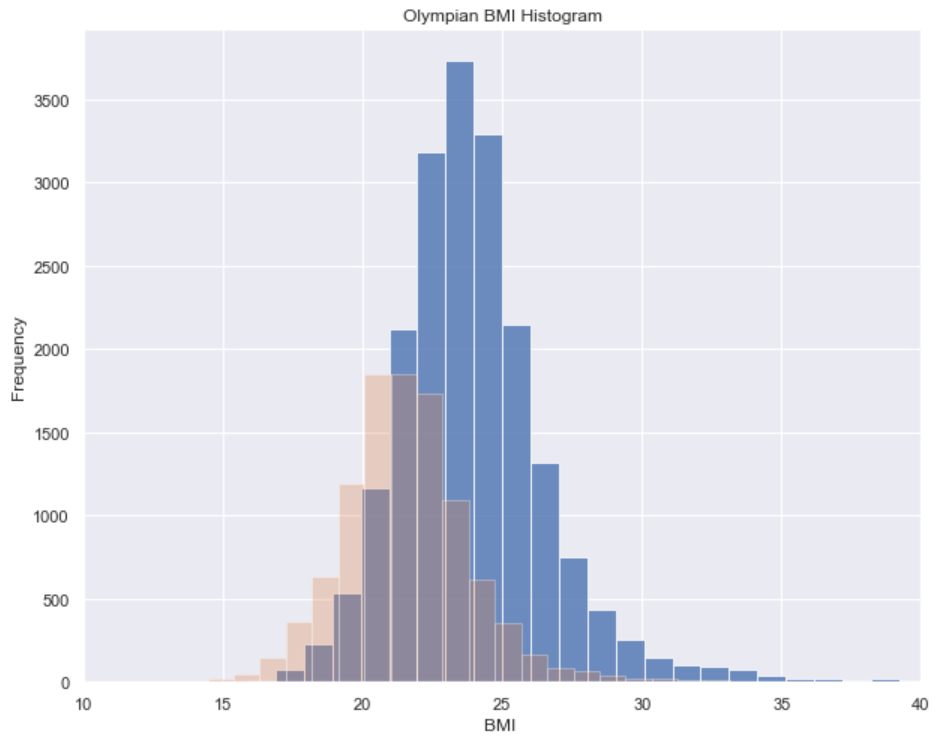


Athlete height and weight, regardless of gender, has trended upward since the beginning of the games. I also looked at height and weight characteristics for awarded athletes.



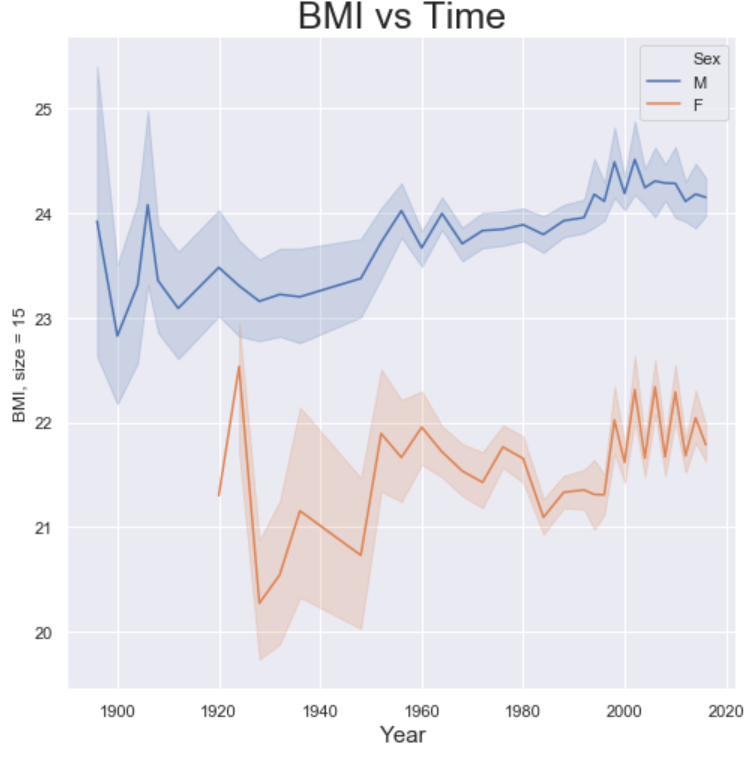


BMI is used to broadly define weight groups. Underweight BMI is less than 18.5, normal BMI is 18.5 to 24.9 and your considered to have an overweight BMI at 25 plus. Obese is considered 30 plus.



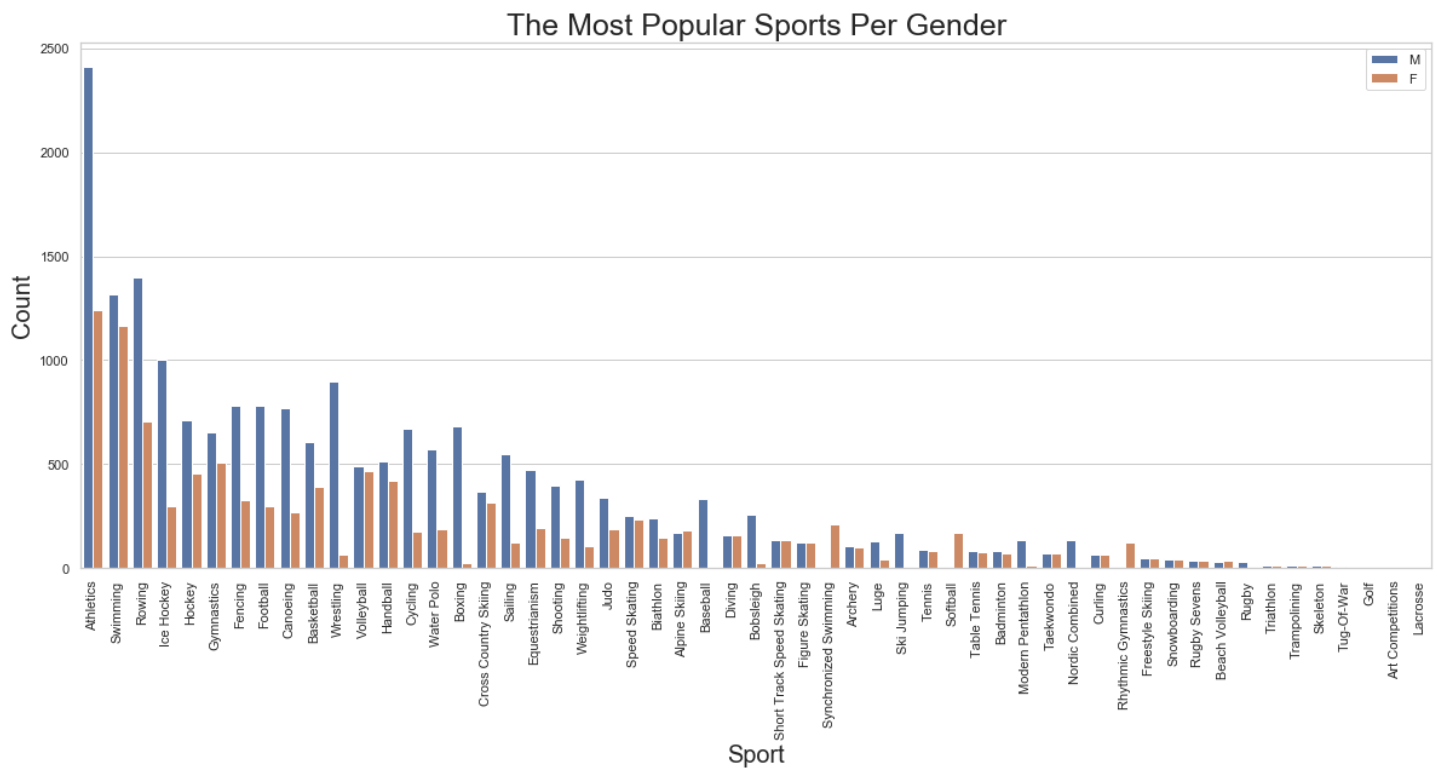
The male median BMI is 24 and the average is 24. Female’s have a median BMI of 21 and an average of 22. BMI descriptive statistics fall into the normal category regardless of gender.

So have BMI’s changed over time?

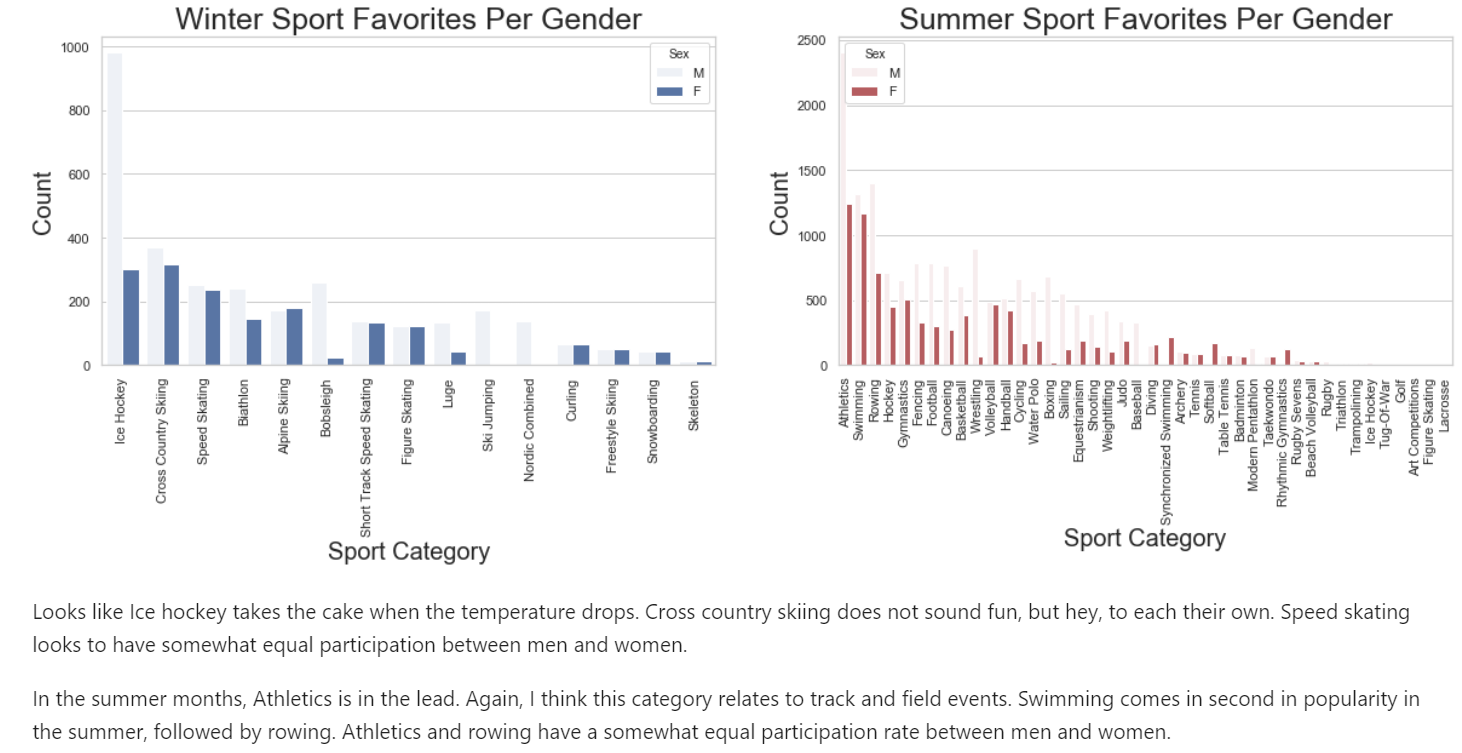


The simple answer is Yes. Both Male and Female BMI’s have trended upward since the beginning.

So with so many sports to compete in…which ones are the preferences per gender?

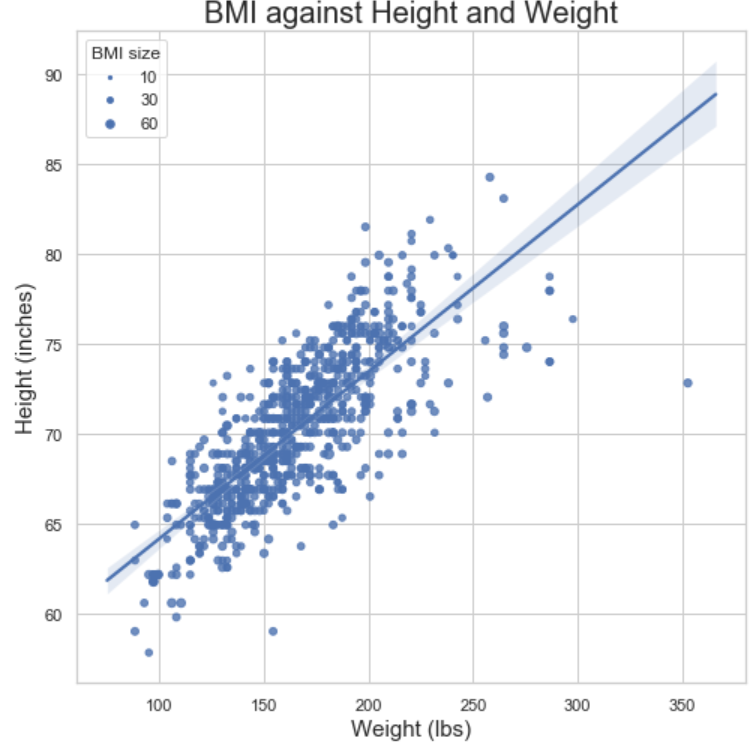


Athletics is the clear favorite for men and holds a slight edge over swimming for women. Rowing comes in third for both genders. I also broke the preferences per the summer and winter seasons.

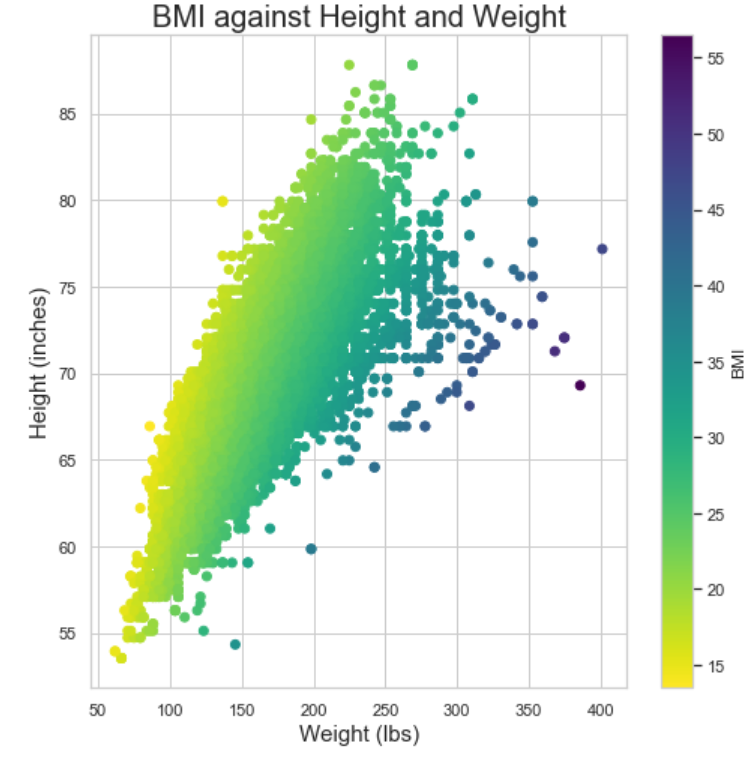


Multivariate exploration looks at many different angles of exploration within the variables. Some of the graphs are repeated, but in different formats.

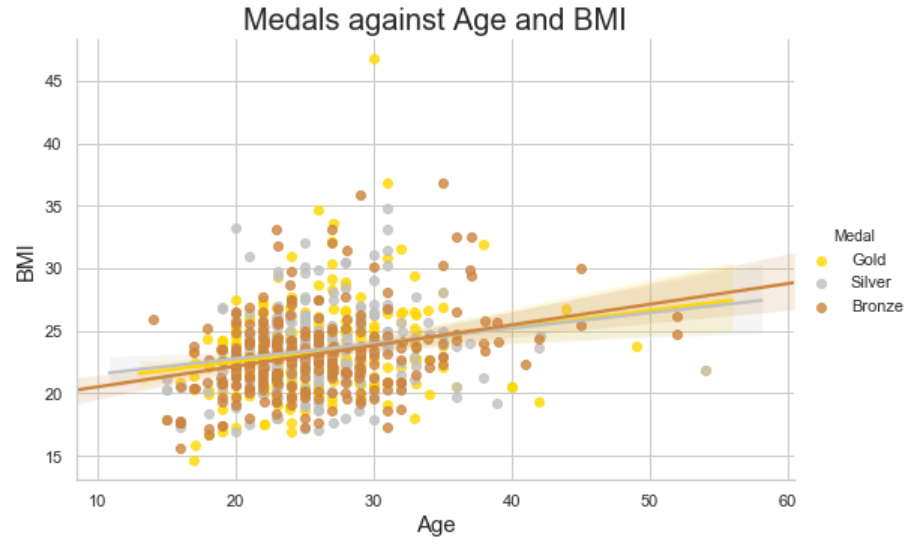
I first wanted to look at BMI’s relationship with height and weight.



Obviously the larger and taller you are, the higher the BMI. There is a clear correlation. However, using size as the BMI marker makes it much more difficult to interpret. Using color instead of size really helps visually.

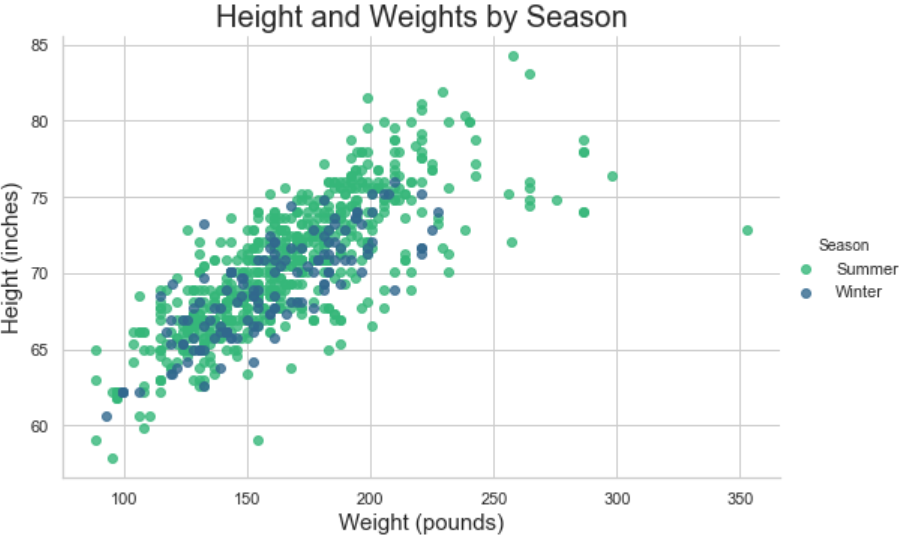


I continued to BMI exploration against age and the possibility of earning a medal.



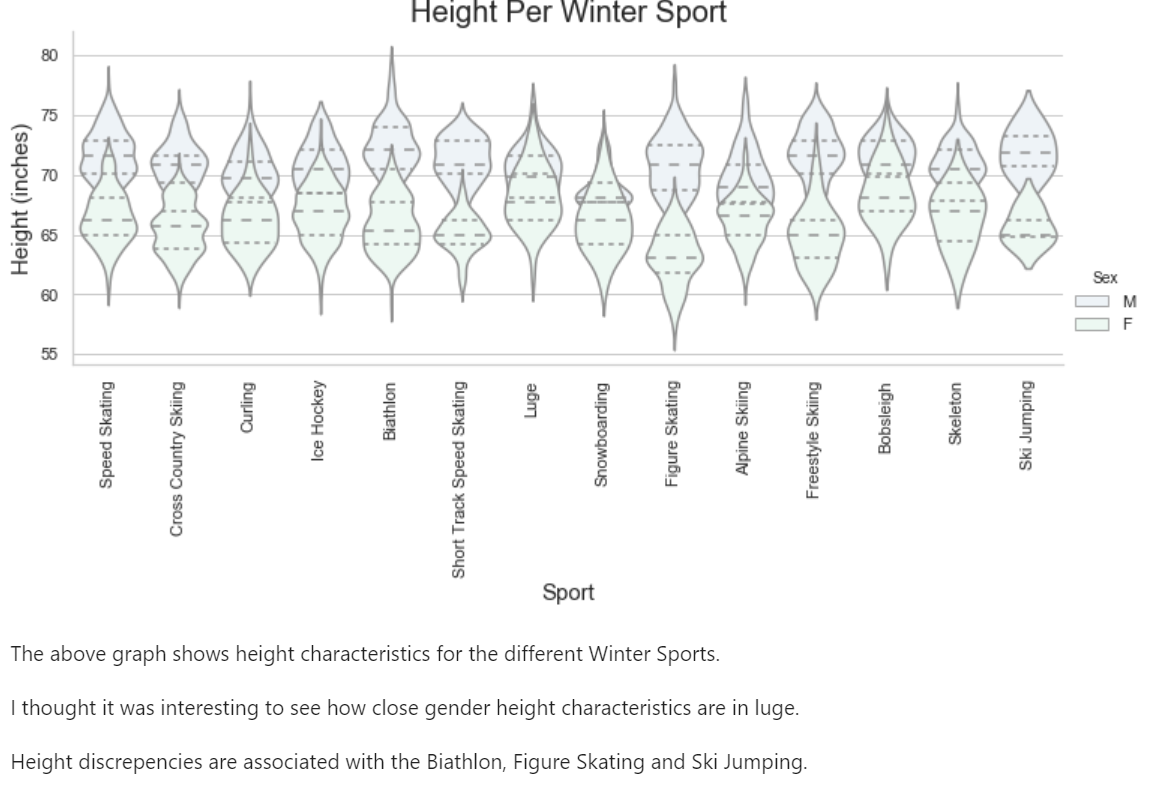
The graph shows that chances of winning a medal as you age is affected by BMI. If you are looking to compete as you get older, you need to keep a lower BMI.

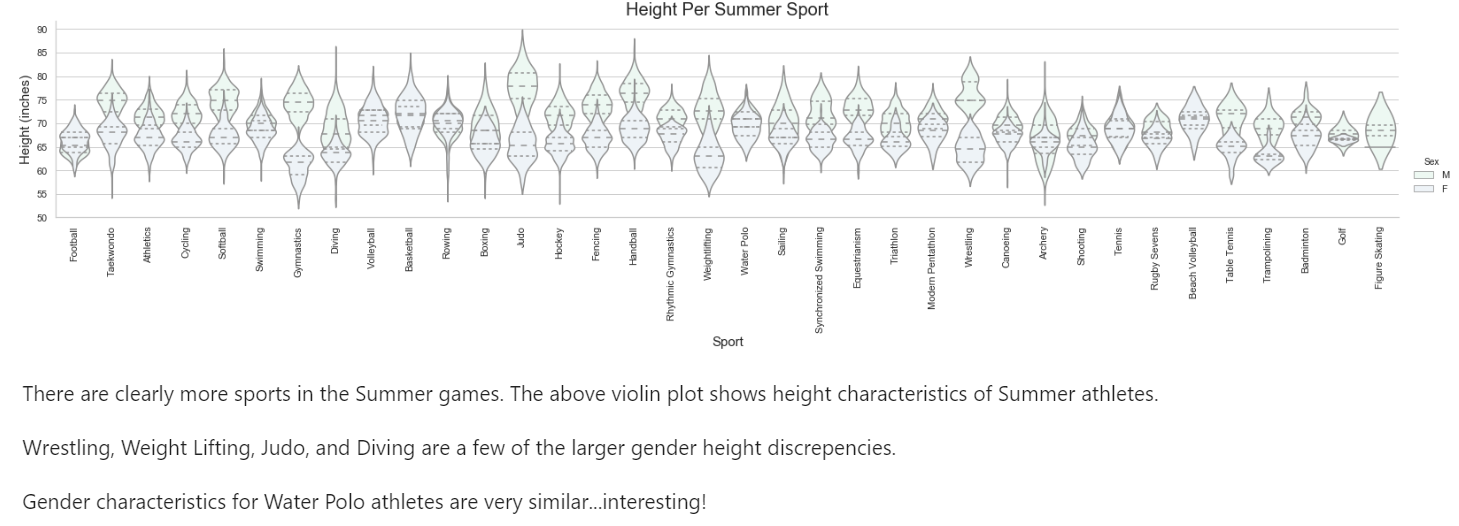
I transitioned to looking at a multivariate analysis of height and weight characteristics per season, per sport and per gender.

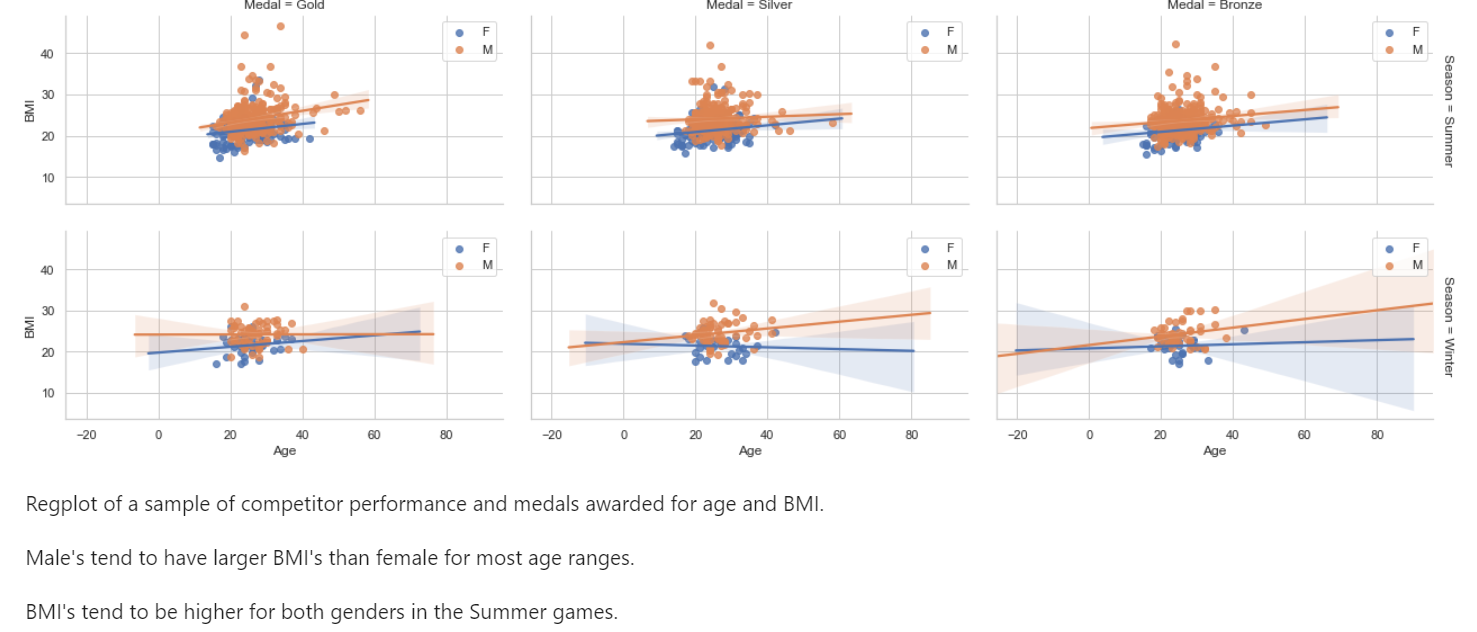


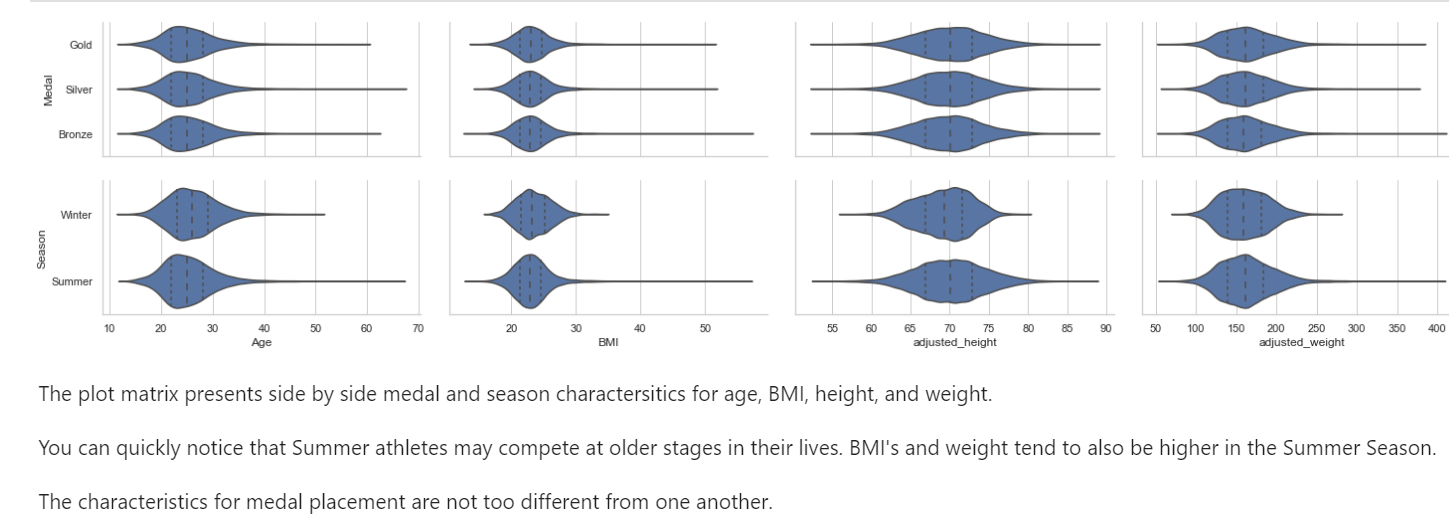
The graph above looks at a sample of athlete height and weight per season. Once averages come into play, there is not too much difference in the body composition as a whole between summer and winter athletes.

The funny thing is I would have guessed winter athletes as being heavier. However, winter athletes ended up being lighter...if ever so slightly.









**If you obtained feedback from others for your explanatory designs, document them here.**

N/A