Dataset

120 years of Olympic History, retrieved from Kaggle

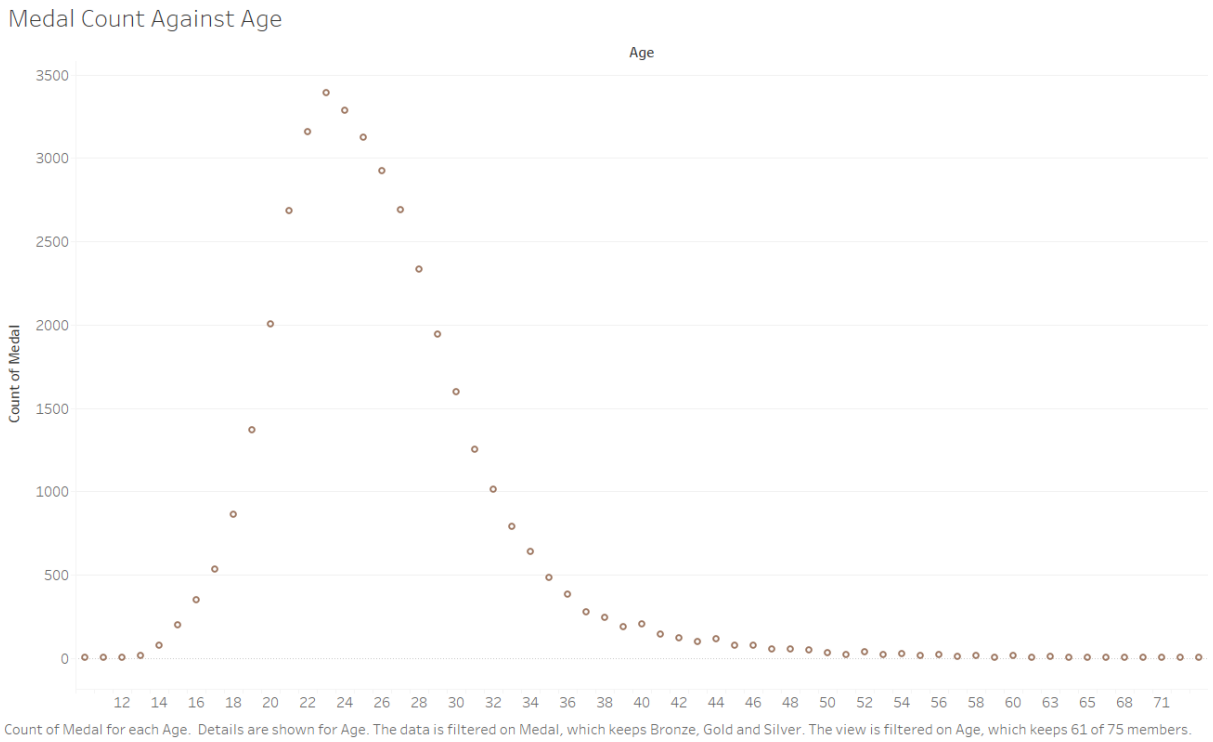
Initial Question

Does age affect athletic performance?

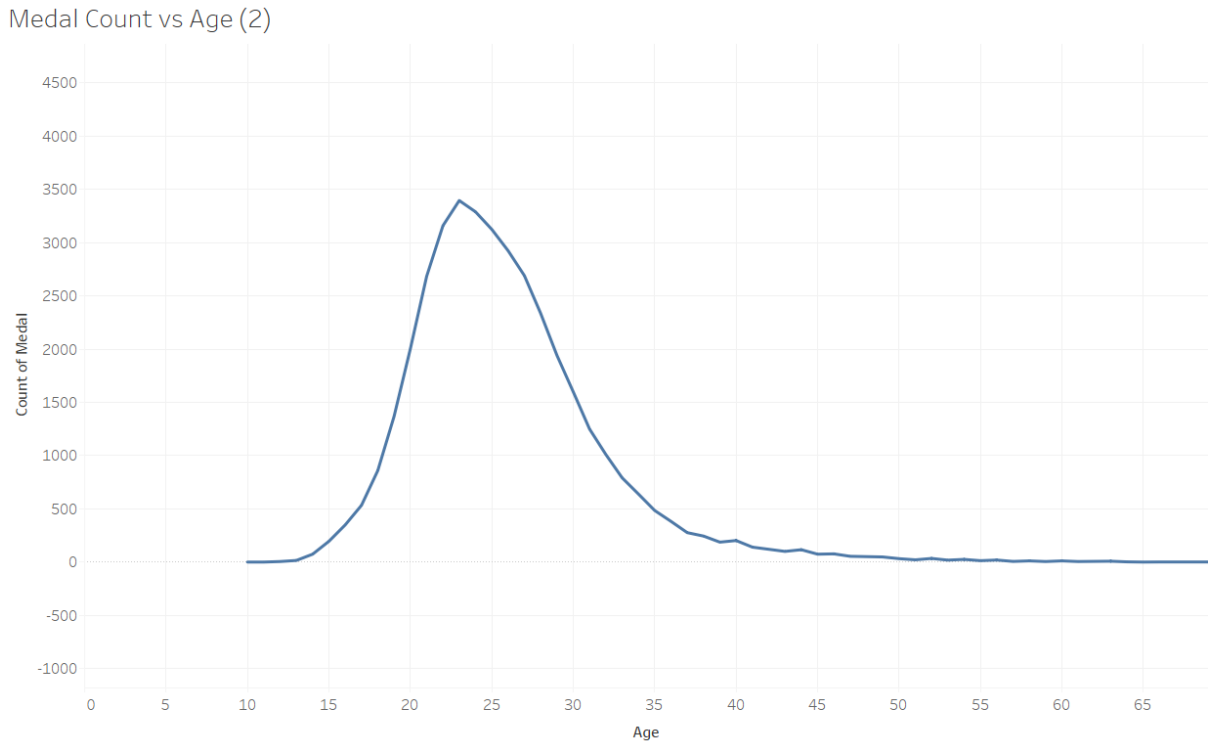
First, I decided to compare total medal count across all ages. I graph the total number of bronze, silver, and gold medals by age. Here, I had to make sure to filter out N/A values for age and medal type. From the graph, I see that there is a notable metal count peak for athletes in their 20s.

Columns: Age

Rows: CNT(Medal)



Count of Medal for each Age. Details are shown for Age. The data is filtered on Medal, which keeps Bronze, Gold and Silver. The view is filtered on Age, which keeps 61 of 75 members.



The trend of count of Medal for INT([Age]). The data is filtered on Medal, which keeps Bronze, Gold and Silver. The view is filtered on INT([Age]), which keeps non-Null values only.

Now, Olympic sports are separated by sex, so it would make since to separate the data by male and female as well.

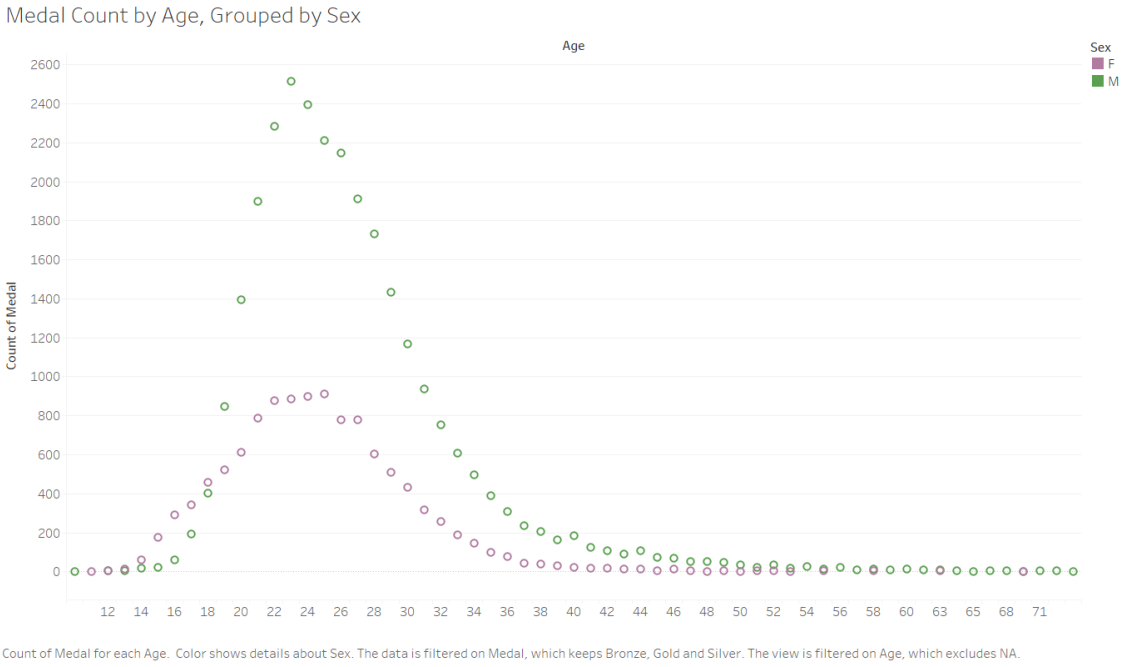
Q: How does Olympic performance vary by age and sex?

The mean age for female medalists is a bit lower.

Columns: Age

Rows: CNT(Medal)

Mark: Sex



Count of Medal for each Age. Color shows details about Sex. The data is filtered on Medal, which keeps Bronze, Gold and Silver. The view is filtered on Age, which excludes NA.

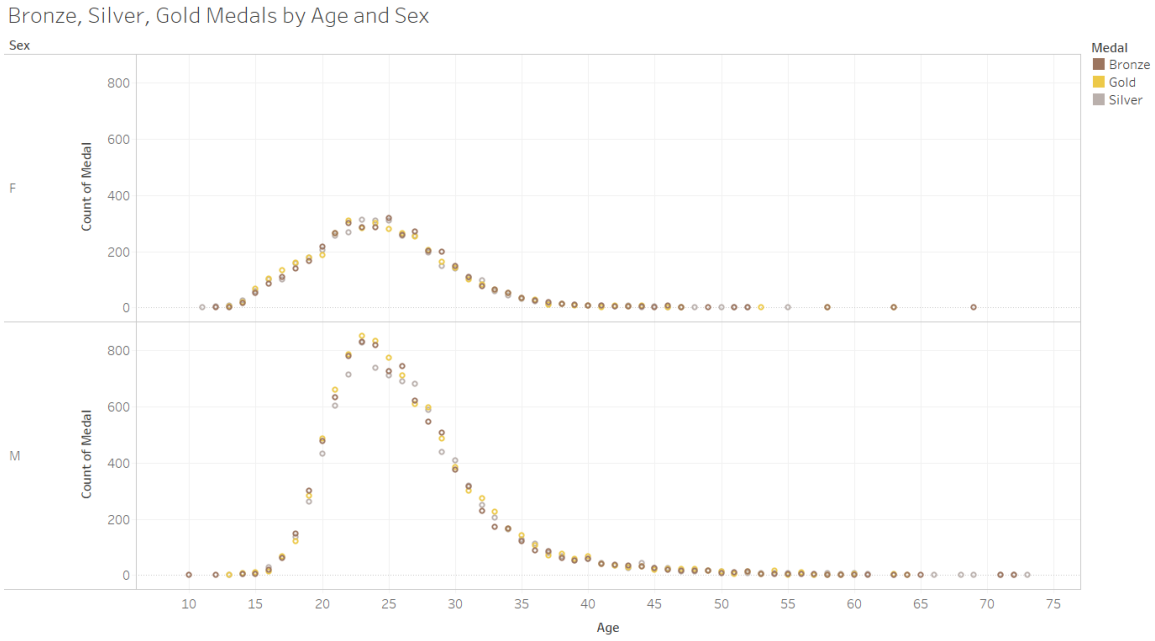
Next, I decided to break down the medal data furthermore, by using color to distinguish between different types of medals.

Q: How does Olympic performance as demonstrated by medal count vary by age and sex with breakdown by medal count?

Columns: Age

Rows: Sex, CNT(Medal)

Marks: Medal



The plot of count of Medal for Int([Age]) broken down by Sex. Color shows details about Medal. The data is filtered on Age, which has multiple members selected. The view is filtered on Medal, which keeps Bronze, Gold and Silver.

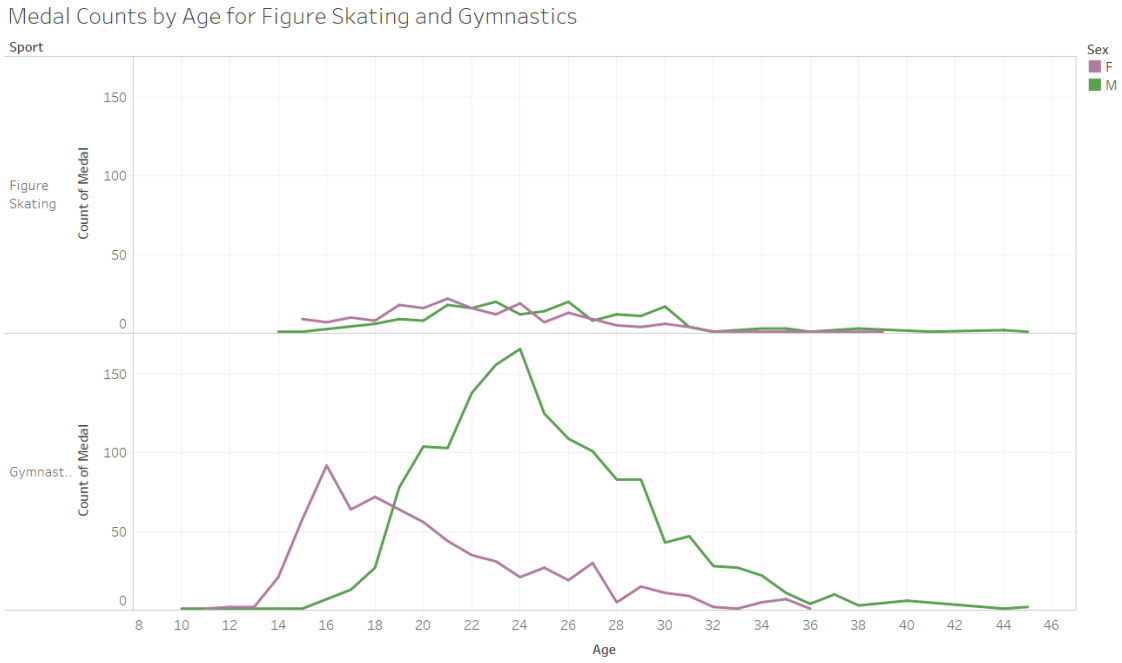
I used yellow for gold, gray for silver, and brown for bronze. These colors were meant to be reminiscent of the actual colors of the medals. The problem we run into here is that the counts and distributions for each medal type is too similar. Therefore, I decided to not distinguish between medal type.

So, we now have established the notable performance level differences across age for all sports, I decided to look at the breakdown of medal distribution by age grouped by sports. I wanted to see if there were any sports that showed a notable age range medal curve. I decided to start by investigating two sports known for early retirement: gymnastics and figure skating. We can see that males have a wider age range for medalists. Additionally, for females, the age is for the most number of medalists is 16, while for males it is 24.

Q: How does Olympic performance as demonstrated by medal count vary by age and for figure skaters and gymnasts?

Columns: Sport, Age

Rows: Sex, Medal



The trend of count of Medal for INT([Age]) broken down by Sport. Color shows details about Sex. The data is filtered on Medal and Age. The Medal filter keeps Bronze, Gold and Silver. The Age filter keeps 34 members. The view is filtered on Sport, which keeps Figure Skating and Gymnastics.

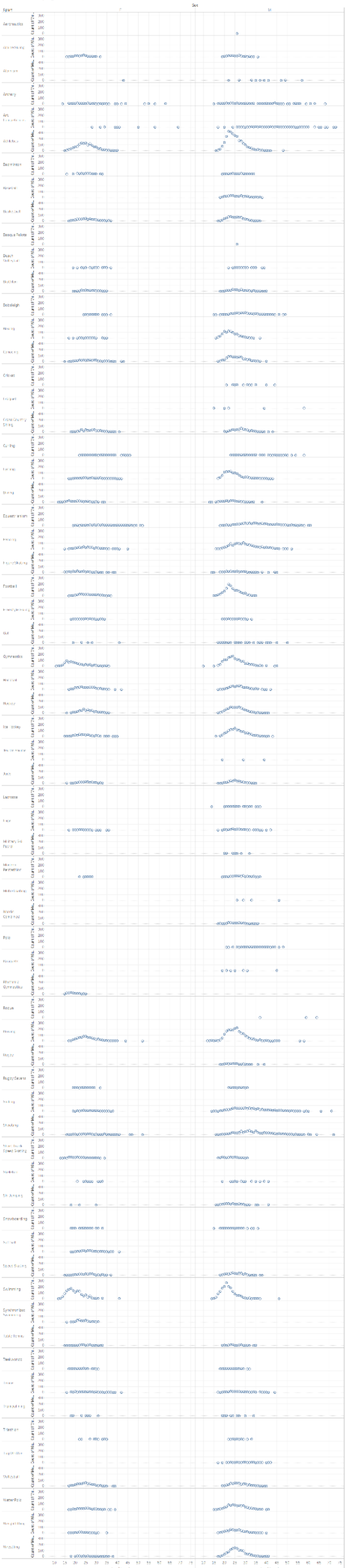
Then, I decided to backtrack and do some exploratory analysis and look across all sports. In terms of exploration this graph was helpful because I could see all the sports. This allowed me to pick out a couple of sports with notable distributions. In terms of visualizations, this was not helpful because there were too many sports, so it was hard to compare across sports. The graph turned out to be exceedingly long (see below).

Q: How does Olympic performance as demonstrated by medal count vary by age and sport?

Columns: Sport, Age

Rows: Sex, CNT(Medal)

Medal Count by Age Across Sports



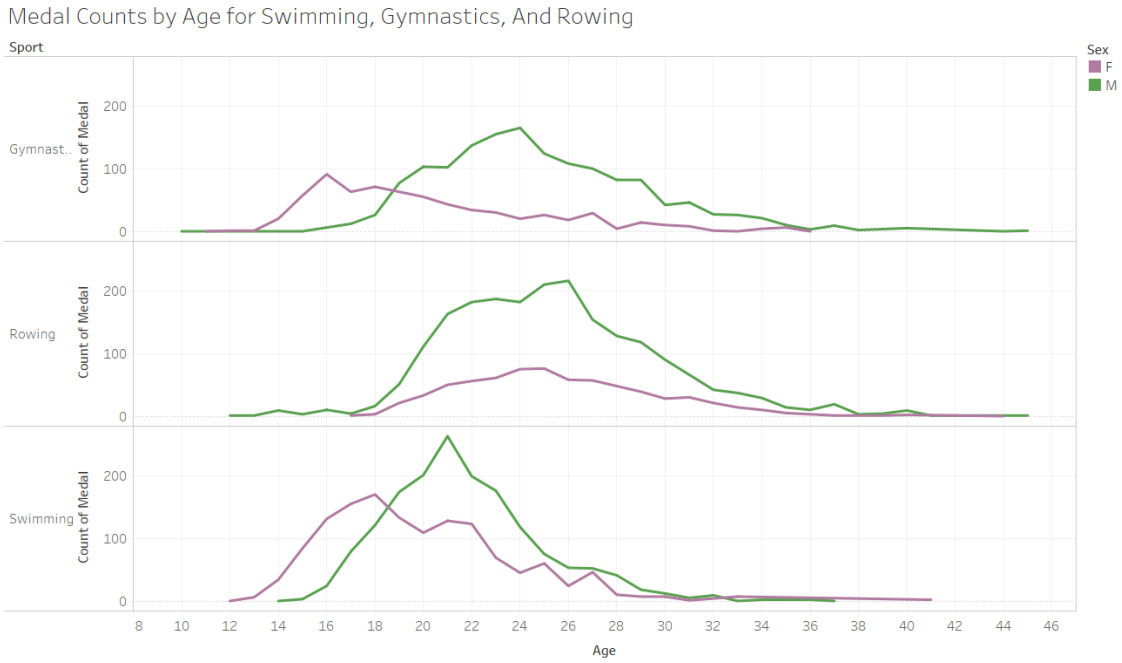
The plot of count of Medal for each Age (Years) broken down by Sex vs. Sport. The data is filtered on Age and Medal. The Medal filter keeps Bronze, Gold and Silver.

I decided to focus in on narrow down my scope of sports and pick out three sports that had notable parabolic medal distributions. These sports were: figure skating, gymnastics, and rowing.

Q: How does Olympic performance as demonstrated by medal count vary by age for swimmers, gymnasts, and rowing?

Columns: Sport, Age

Rows: Sex, CNT(Medal)

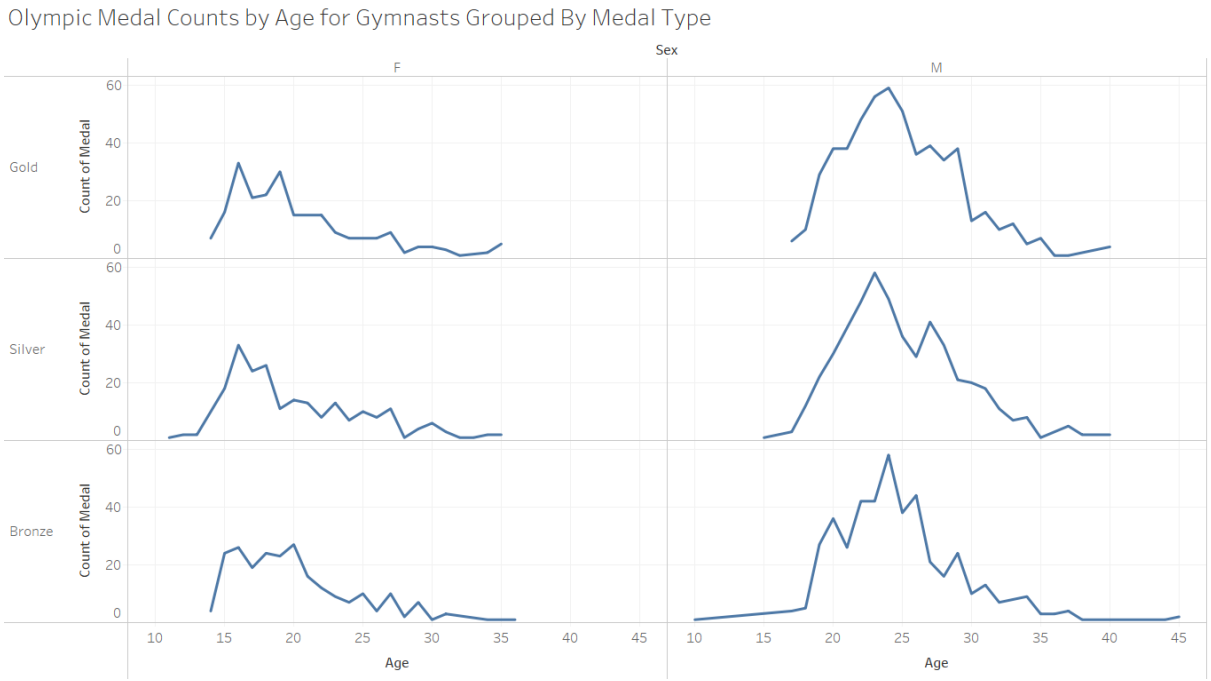


The trend of count of Medal for INT([Age]) broken down by Sport. Color shows details about Sex. The data is filtered on Medal and Age. The Medal filter keeps Bronze, Gold and Silver. The Age filter keeps 34 members. The view is filtered on Sport, which keeps Gymnastics, Rowing and Swimming.

I decided to just focus in on the distribution of medals for one sport: gymnastics. This was because I noticed that it had the most visible medal-age curve and had the most notable difference among the sexes.

I had to transform the data to filter out records with N/A for age or N/A for medal. I also had to cast age into an integer and selected only medal records for gymnastics.

Now I wanted to investigate if gymnastics was skewed by a certain medal counts. That is, is the age distribution for medal counts different types of medals? We already saw that there was not a significant difference for the pooled medal counts, but would limiting the sport to just gymnastics make a difference?



The above graph shows us the age distribution across medal types isn't notably different.

Q: How does Olympic performance as demonstrated by medal count vary by age and across sexes for gymnasts?

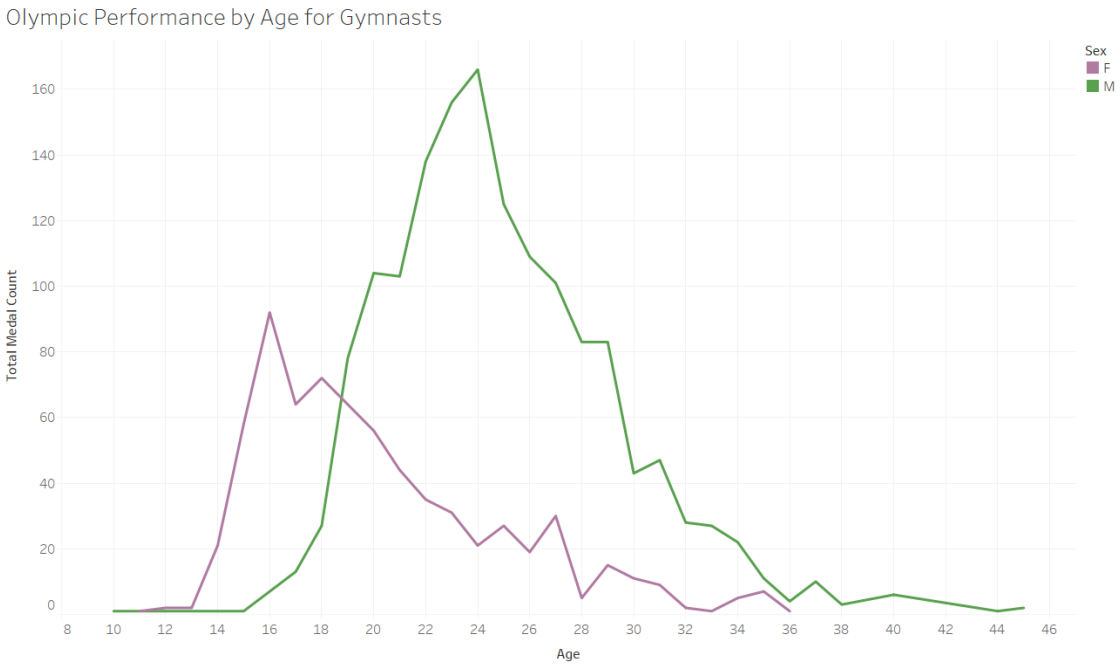


Figure 1. This graph shows the athletic performance, as quantified by medal count, at the Olympics as a function of age, and grouped by sex. The female age-medal distribution is much more right skewed than the male age-medal distribution, which follows a relatively normal distribution.

Men's gymnastics debuted at the Olympic Games in 1896. Women's gymnastics were added in 1936, starting with an all-around competition, and in 1952 individual events were added. For females, the top medal scoring age was 16, the average age was 20.2, and the median age was 24.9. For males, the top medal scoring age was 24, the average age 24.9, and the median age was 24. The female age-medal distribution is right skewed while the male age-medal distribution follows a normal distribution. One reason for the overall average of Olympic medalists, high risk for injury in the sport. There is also a notable difference among sexes. A possible explanation for this is the differences in competition. For the Olympics, each national team contains five members; men's gymnastics features six events, while women's gymnastics has four. For the women's team, this means they can recruit members that excel only in one event, while for the men's team athletes must be able to contribute high scores across multiple events. Therefore, athletes on the men's team may be older to be all-arounders.

Bibliography

A History of Gymnastics: From Ancient Greece to Modern Times. (n.d.). Retrieved from https://www.scholastic.com/teachers/articles/teaching-content/history-gymnastics-ancient-greece-modern-times/

Meyers, D. (2016, August 16). Comparing men's and women's gymnastics is futile: They're barely the same sport. Retrieved from https://www.theguardian.com/sport/blog/2016/aug/16/mens-womens-gymnastics-comparison-olympics