

Olympics Analysis

About the data set:

This is a historical dataset on the modern Olympic Games, including all the Games from Athens 1896 to Rio 2016. The data is quite granular as it is aggregated per athlete and event per year.

The motivation behind the analysis was to understand how the Olympics have evolved over time, identify specific trends, and investigate the participation of different countries, genders, etc. I was also particularly interested in exploring the concept of the home field advantage to try to understand if we see statistically significant backing for this concept.

I analyzed this dataset by exploring through Tableau and creating a [dashboard](#) and I performed statistical testing using RStudio.

Bottom Lines:

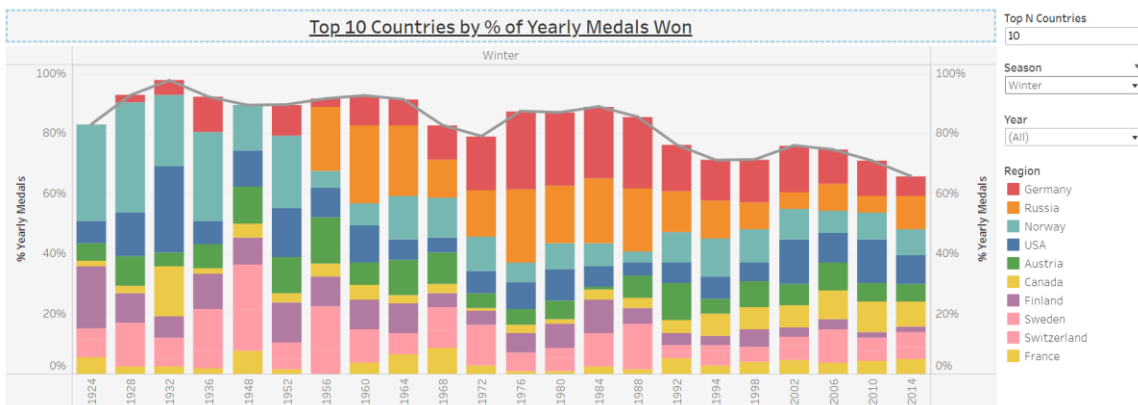
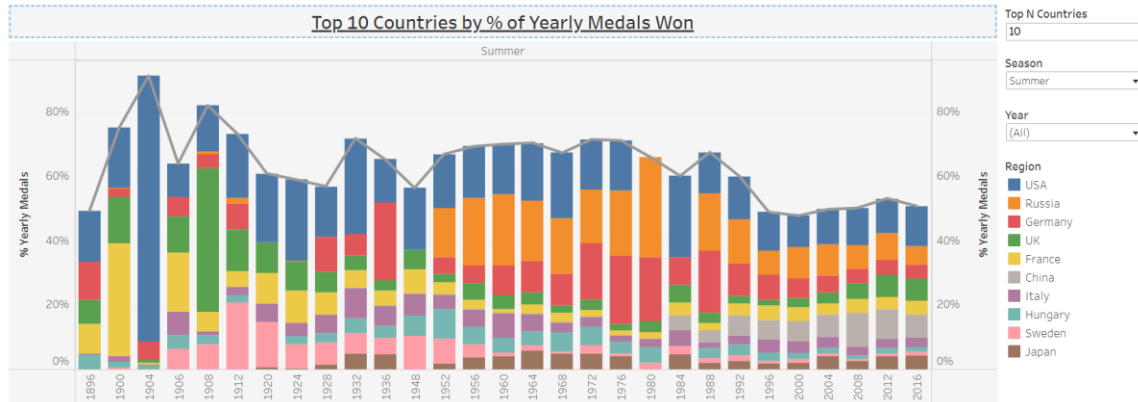
- The total % of yearly medals won by the top 10 countries in the Summer Olympics is close to ~50% while in the Winter Olympics it is closer to ~70%. This makes sense that we see fewer countries dominating the medals since Winter sports generally require specific climates and often require more infrastructure and are more costly to train for.
- While having a larger percentage of athletes helps increase yearly percentage of medals won, in order to be a top-ranking medal country, they must also have a higher % of athletes who received a medal (i.e. “medal conversion rate”).
- The % of medals won by the top 10 countries is decreasing, a strong indicator that more countries are improving their Olympic performance and the medal winning countries are becoming more diversified.
- While at a glance it appears that there is some sort of home field advantage, there is no statistically significant difference between the number of medals won during hosting Olympics and non-hosting Olympics. Perhaps with more data in the coming years this could change.
- Progressively more athletes are female between 25-29, so my Olympic dream is not dead yet!

Topics Covered:

1. [Summarizing Yearly Trends](#)
2. [Investigating the Home Field Advantage](#)
3. [Investigating Athlete’s Performance by Physical Characteristics](#)
 - a. [Age](#)
 - b. [Gender](#)
 - c. [Weight](#)

Summarizing Yearly Trends

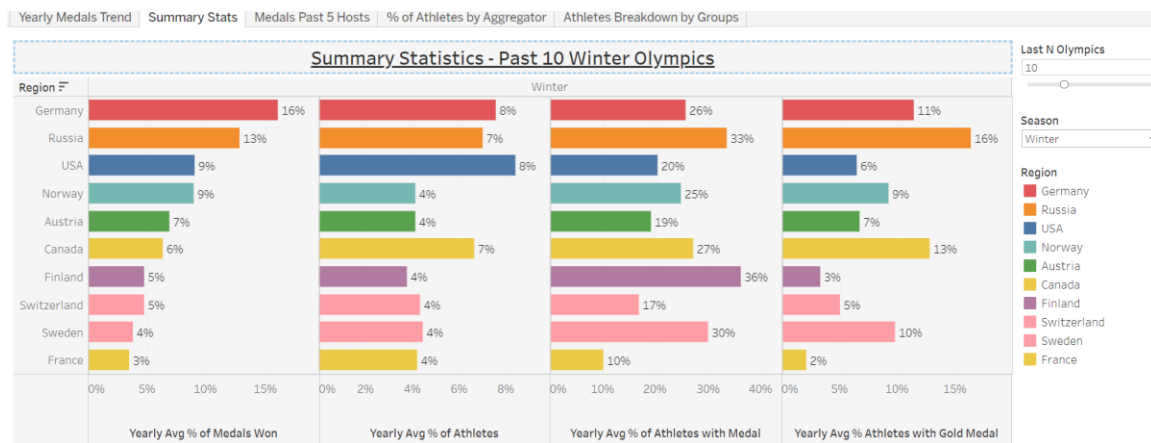
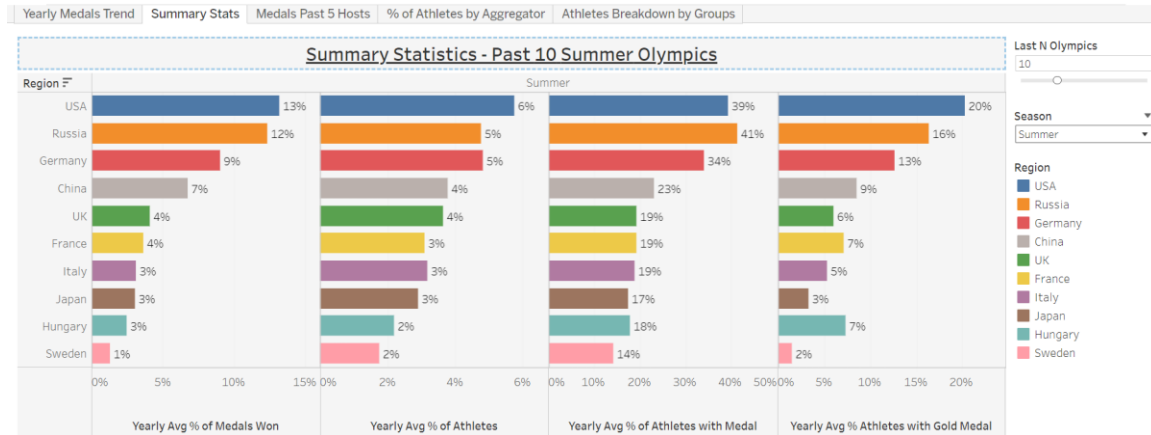
- Since the number of total medals won each year varies, I investigated the medal share % per country for the top 10 countries since 1896.
- We can see that the USA has always been a strong contender and Russia, which started to compete in the 1950s, started off right away with strong performance. Germany was a much stronger contender in the 70s and 80s and has since declined in medal share. China started having strong performances in the 1980s and has continued to grow ever since.
- The medal share % of the top 10 countries in the Summer Olympics is close to ~50% while in the Winter Olympics it is closer to ~70%, this isn't surprising as the Winter Olympics seem to be more heavily dominated by countries that have winter weather and are more affluent and can therefore train for winter sports more easily.
- We can also see that the medal share % for the Summer Olympics was much higher and dropped in the mid-90s where more countries seemingly started to have better performance. The medal share % for the Winter Olympics is also decreasing over time.



Looking at the past 10 Olympics for each season:

- USA and Russia were leaders in Summer and Germany and Russia were leaders in Winter in terms of average % of total medals.
- Usurpingly, having a larger percentage of athletes helps increase medals percentage, though the top-ranking countries also have higher % of athletes who received a medal (i.e. medal conversion rate).

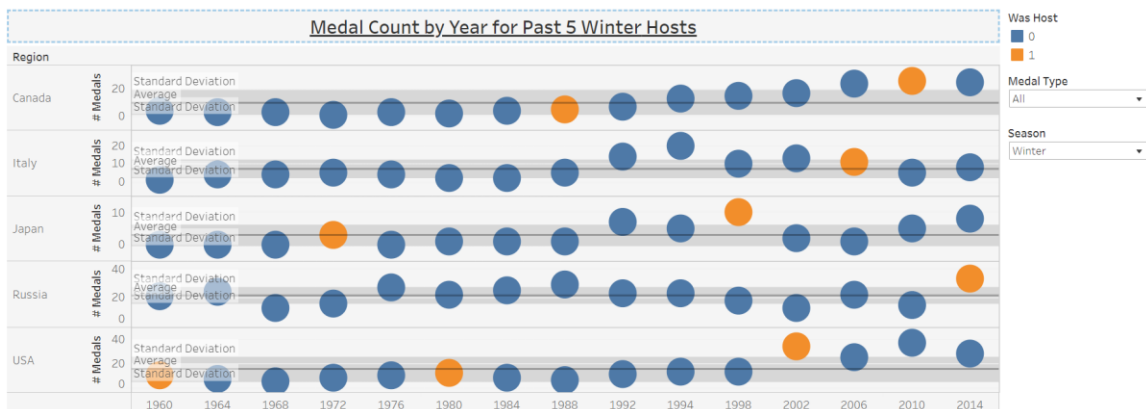
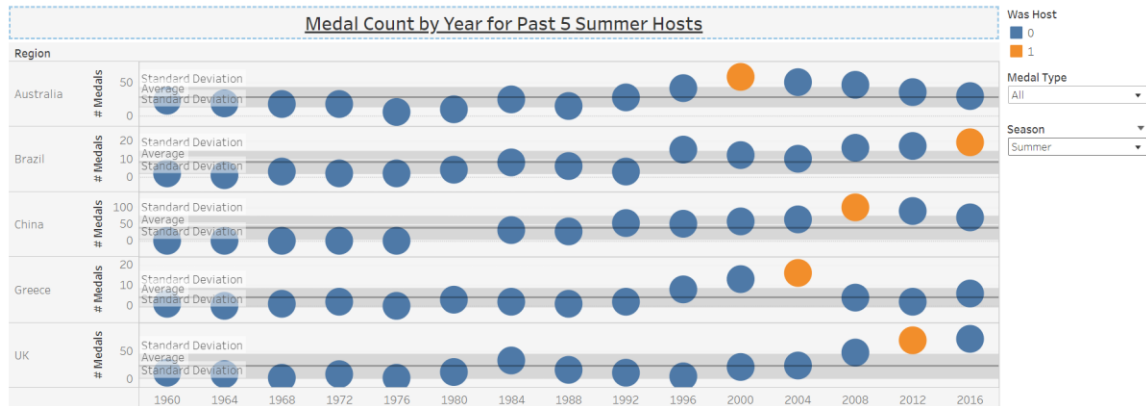
- Having a large percentage of athletes alone will not be enough to ensure high medal count, but rather countries also need to have highly competitive athletes with higher medal conversion rates. For example, Russia and Germany have ~the same amount of summer athletes but Russia outranks Germany by 3% of medals.



Investigating the Home Field Advantage

I also analyzed the impact of the home field advantage. Below you can see the past 5 hosts of the Summer/Winter olympics and their medal count since 1960.

- For almost all the countries, their medal count when they host is way above their average, and for some it is the highest number of medals ever received.
- This can be explained by the fact that competitors from host countries have lower qualification standards than their foreign peers.



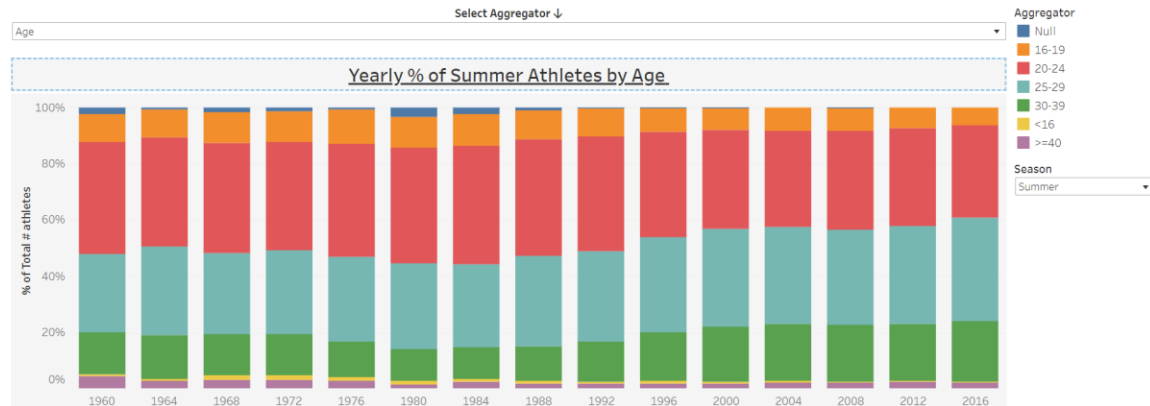
- To understand if this home field advantage truly holds any weight, I performed significance testing on the medal counts for hosting vs non hosting years for the 10 countries above.
- Since the distribution is not normal, I performed the Wilcox signed rank test in R Studio.
- Despite the suggestive evidence of the home field advantage, it appears that the advantage is not statistically significant ($p\text{-value} \leq .05$) in all cases except for Japan for Gold medals. It's worth noting, however, that many of the summer p -values are approaching significance and it may require a larger sample size (there is only one host year for many countries).

P-VALUES – ONE-SIDED WILCOX TEST										
	SUMMER					WINTER				
Parameter	AUS	BRA	UK	CHN	GRC	JPN	RUS	ITA	USA	CAN
Gold Medals	10%	6%	6%	6%	5%	2%	8%	10%	21%	47%
Total Medals	7%	7%	10%	6%	6%	7%	7%	21%	23%	13%

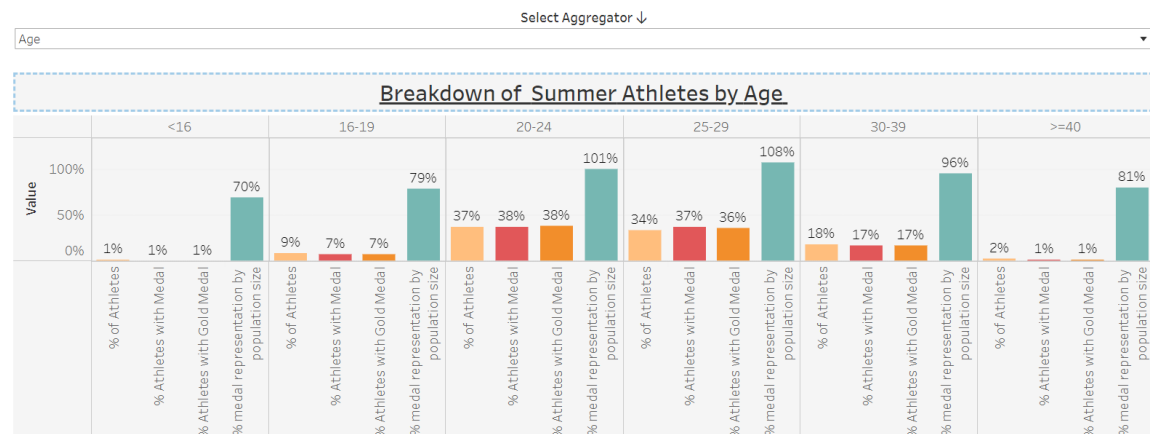
Investigating Athlete's Performance by Physical Characteristics

Age

- Progressively more athletes are between 25-29 and 30-39 starting in the mid-90s. Good news, there's still time!!

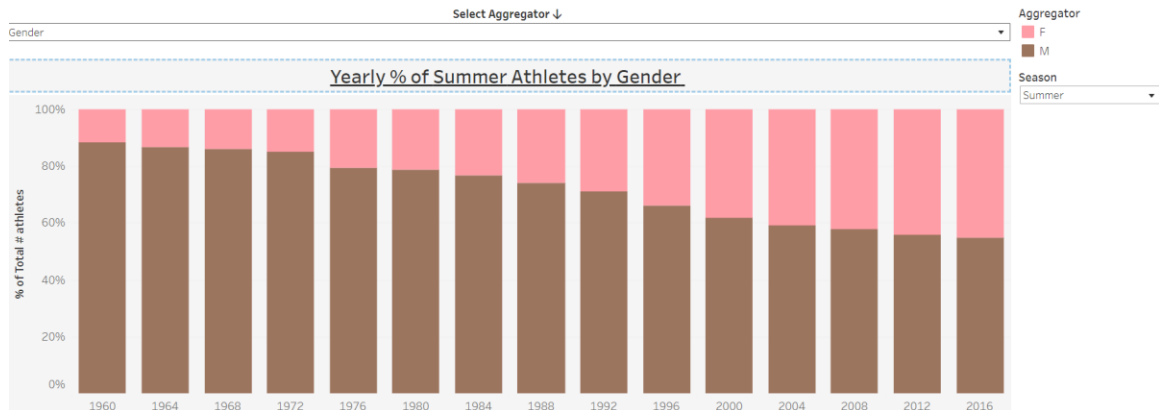


- For the past 10 Summer Olympics I calculated the % of athletes and % of athletes with medal/gold medal by age group. For example, 20–24-year-old athletes made up 37% of athletes and 38% of athletes with medal.
- I created a KPI to understand what the value of each age group is in terms of medal by creating a ratio of % of athletes with medal divided by the total % of athletes. In this case, it seems that the **25–29-year-old athletes have the highest % of athletes with medal given their share of the population**. This is somewhat surprising since I would expect the 20–24-year-old athletes to be more in their physical prime.

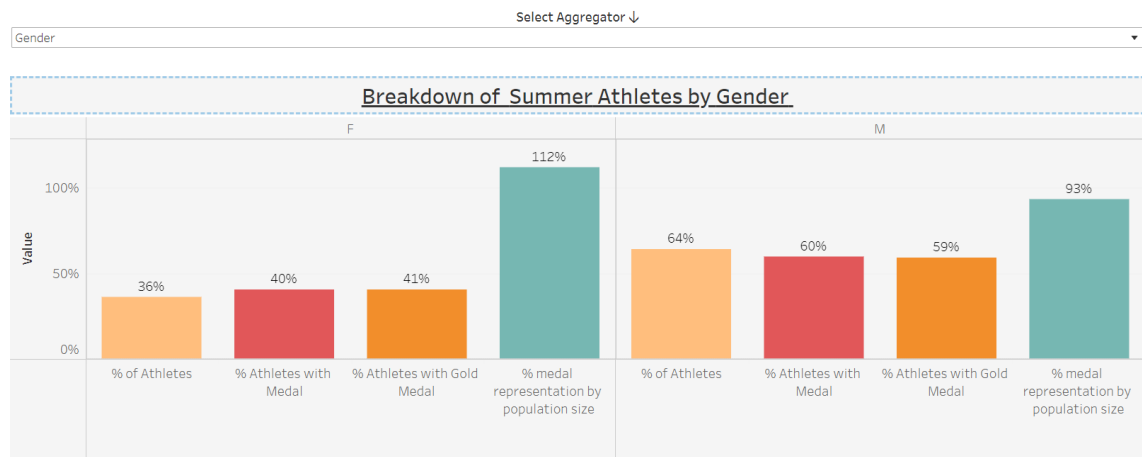


Gender

- In the 2016 Summer Olympics, women accounted to 45% of athletes! Compared to 1960 at only 11%, I would say we are making progress.

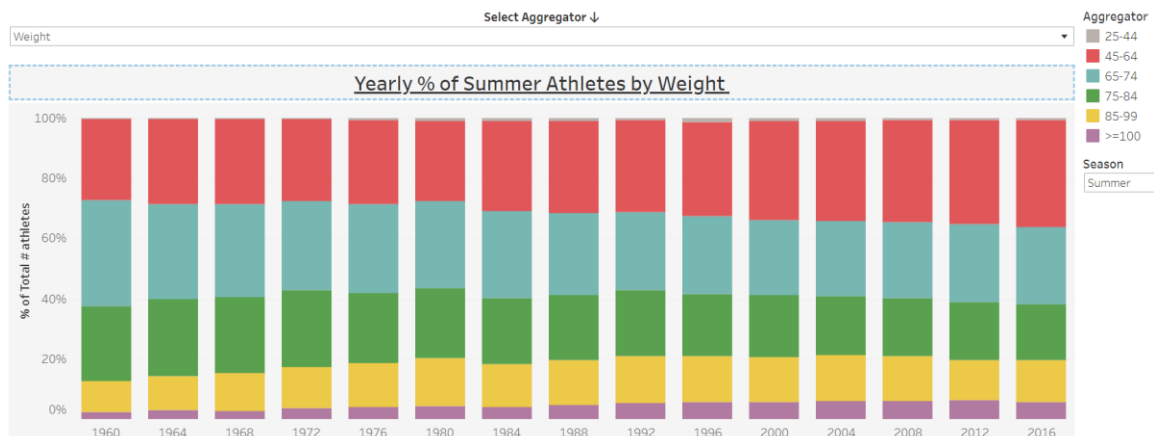


- Using the same KPI as above, we can see that females have higher medal representation given their share of the population. In the past 10 Olympics they made up 36% of athletes, 40% of athletes with medal, and 41% of athletes with gold medal!



Weight

- We can see that over time athletes with weight between 45-65 KG are increasing, which may have to do with the increase in women competing.



- Using the KPI from above and looking at groups accounting for at least 5% of athletes, it seems that heavier athletes have a higher medal representation, with athletes between 85-99 KG accounting for 15% of athletes and 19% of athletes with medals.

