



SportsStats

SportsStats is a sports analysis firm partnering with local news and elite personal trainers to provide “interesting” insights to help their partners. Insights could be patterns/trends highlighting certain groups/events/countries, etc. for the purpose of developing a news story or discovering key health insights.

Questions + Hypothesis

Questions

- 1) Are more women participating?
- 2) Are more countries participating?
- 3) Have some sports increased or decreased in popularity?
- 4) Are Summer sports more popular than Winter sports?

Hypothesis

- 1) More countries are participating over time.
- 2) More women participating continues to increase.
- 3) Countries that perform well in the Summer Olympics tend to also perform well in the Winter Olympics.
- 4) Taller athletes are more successful in the Olympics.
- 5) Northern countries are more successful at Winter sports.

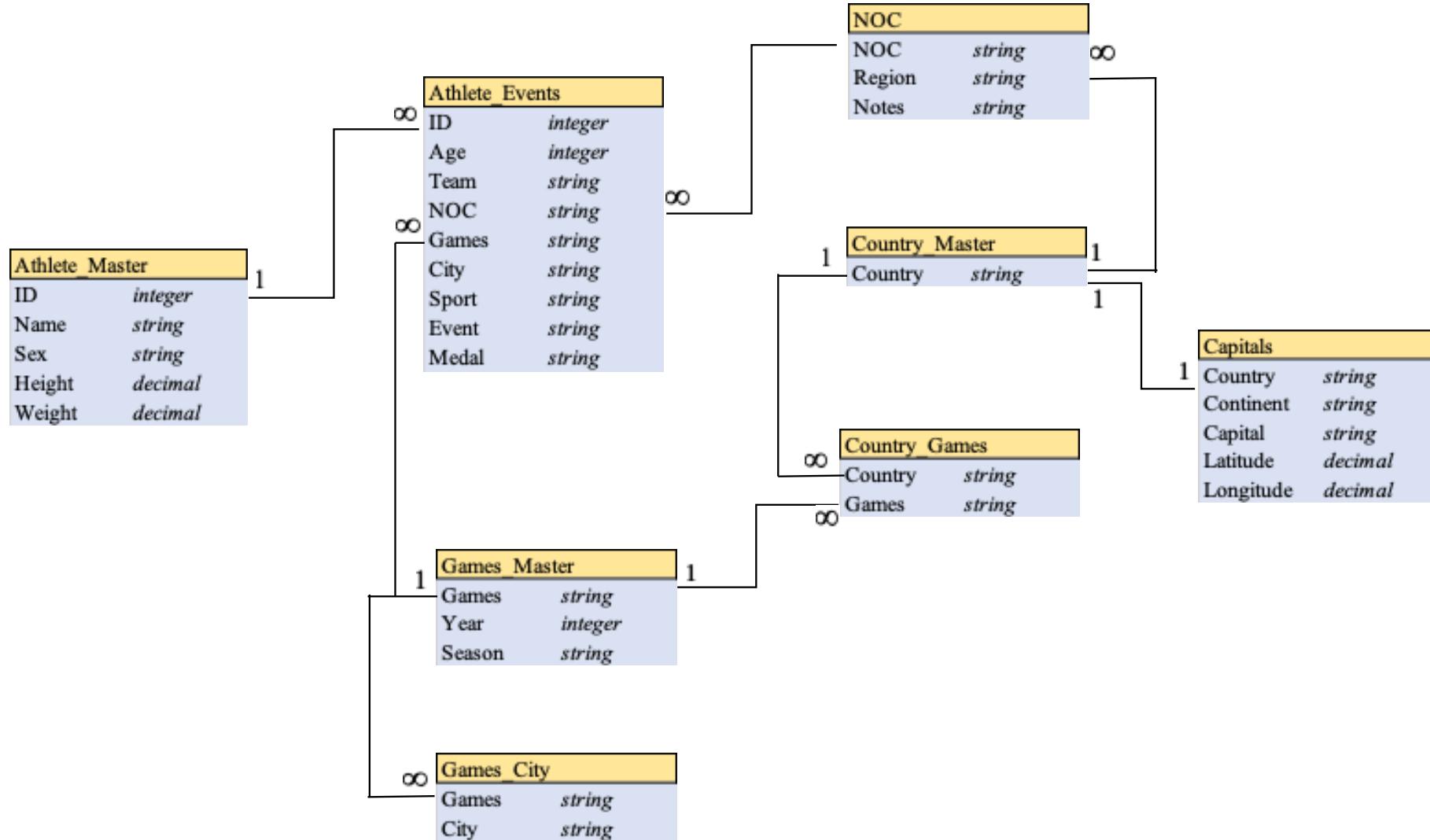
SportsStats dataset

- `athletes_events` (Athlete Events) – a table of every event for each athlete for all Olympic games between 1896 and 2016.
- `noc_regions` (NOC) – a table of National Olympic Committees codes for countries that exist today or did in the past. They include lists of territories such as Puerto Rico, Cayman Islands, and special administrative regions such as Hong Kong.
- `capitals` (not included with SportsStats) – a table with the capital of every country/territory and its Latitude/Longitude, used for mapping and comparisons of Latitude and other measures.

Country Column, Data Cleaning Challenges

- Used NOC table and team information to create a field “country” which groups related NOC regions into similar countries. Based on the noc_regions.region
- Countries with more than one NOC code grouped together. Ex. East and West Germany under one country called “Germany,” etc.
- The Soviet Union’s data is grouped together with Russia to create a country called “Russia (incl. USSR)”. Former Soviet States are separate starting in 1992.
- Former Yugoslavian states were listed in NOC under “Serbia.”
- USA, UK and the Netherlands territories are listed as separate countries, as are Taiwan, Hong Kong, and Palestine.
- Singapore changed their NOC code from “SIN” to “SGP.”
- Spelling of Bolivia was fixed in the NOC table.

Entity Relationship Diagram

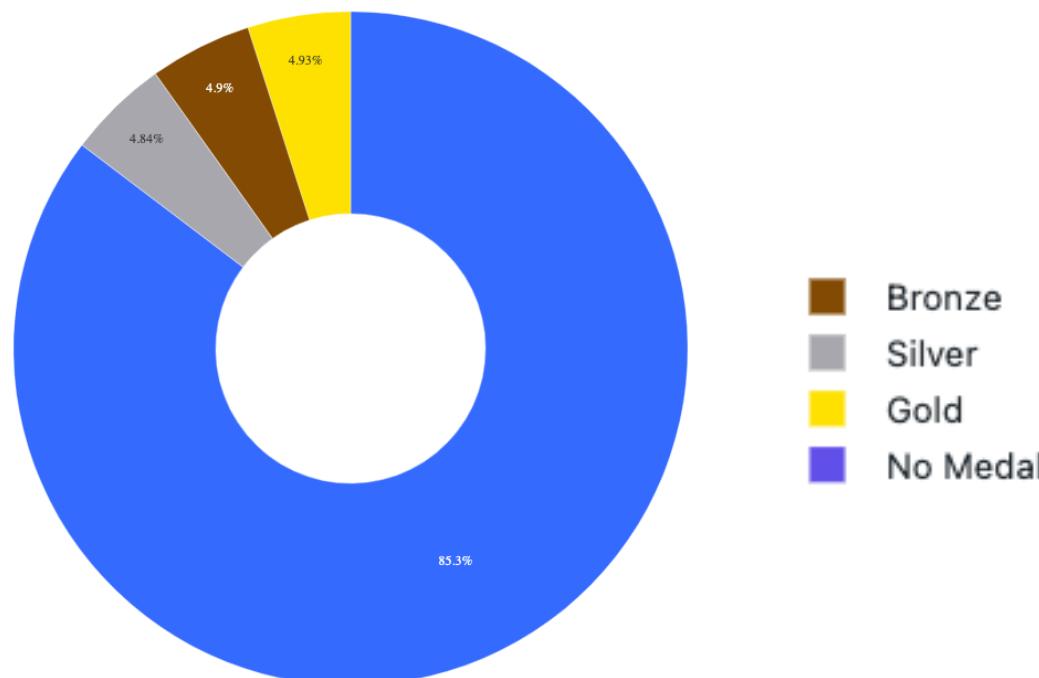
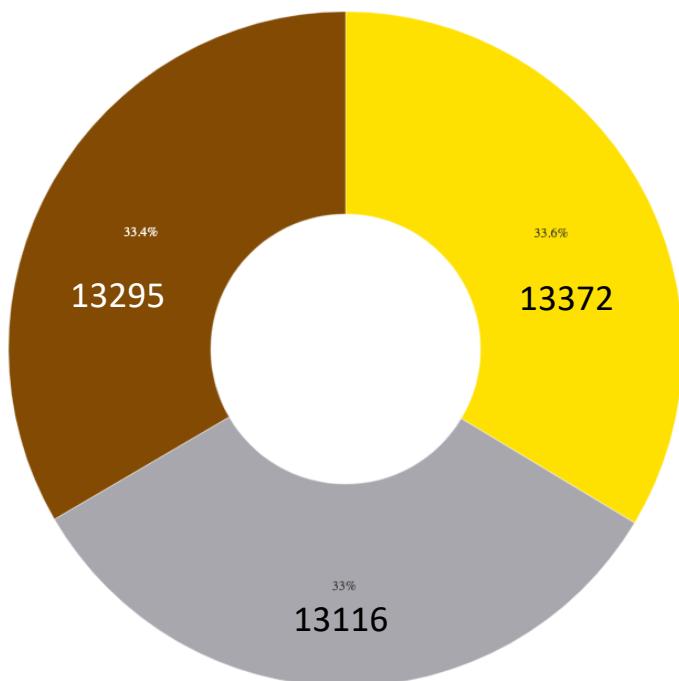


Dataset Totals

Total records :	271116	Total Countries:	209
Total athletes:	135571	Continents:	6
Total sports:	66	Cities hosted:	42
Total events:	765	Total games:	51
Athletes who played for 2 countries:	670		
Athletes who played for 3 countries :	13		
Athletes Winter + Summer games:	163		

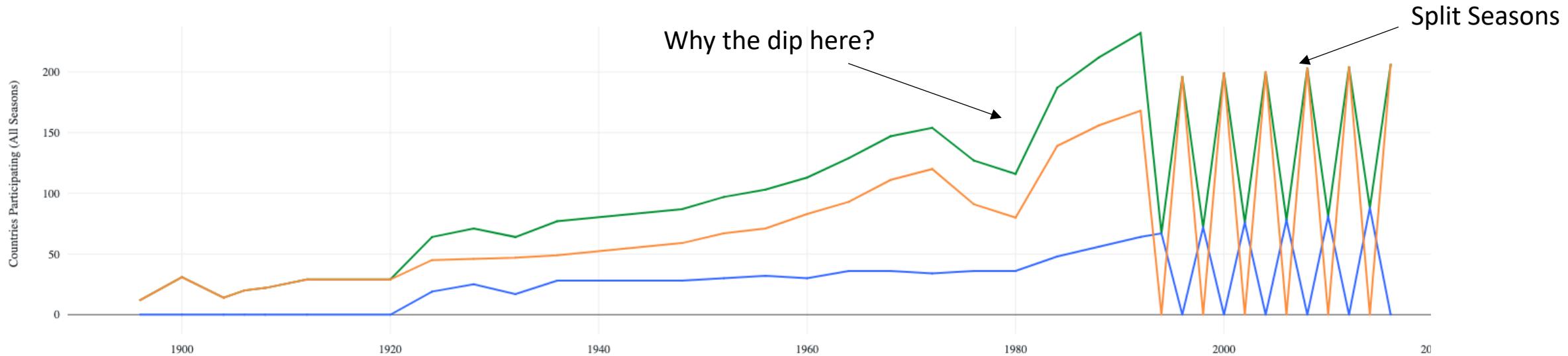
Equal number of Gold/Silver/Bronze?

- Presumably, each event would have a gold, silver, and bronze winner, unless there are less than three participants. How can there be more bronze medals than silver medals?
- According to Wikipedia, there are athletes who have their medals stripped away: https://en.wikipedia.org/wiki/List_of_stripped_Olympic_medals
- Also, team sports give medals to each member of the team. Not all teams for each event are equal in size.



Are more countries participating over time?

- Yes, for the most part. There were downticks in 1976 and 1980, due to boycotts. After 1992, when they split Summer/Winter into different years, you can see that the number of countries participating in Winter sports is much fewer. The green line is both seasons. Some countries participated in one season, but not the other.



Boycotts

1976 Boycott:

- African nations boycotted to protest the inclusion of New Zealand, which had previously played a Rugby game against apartheid South Africa.

1980 + 1984 Boycotts:

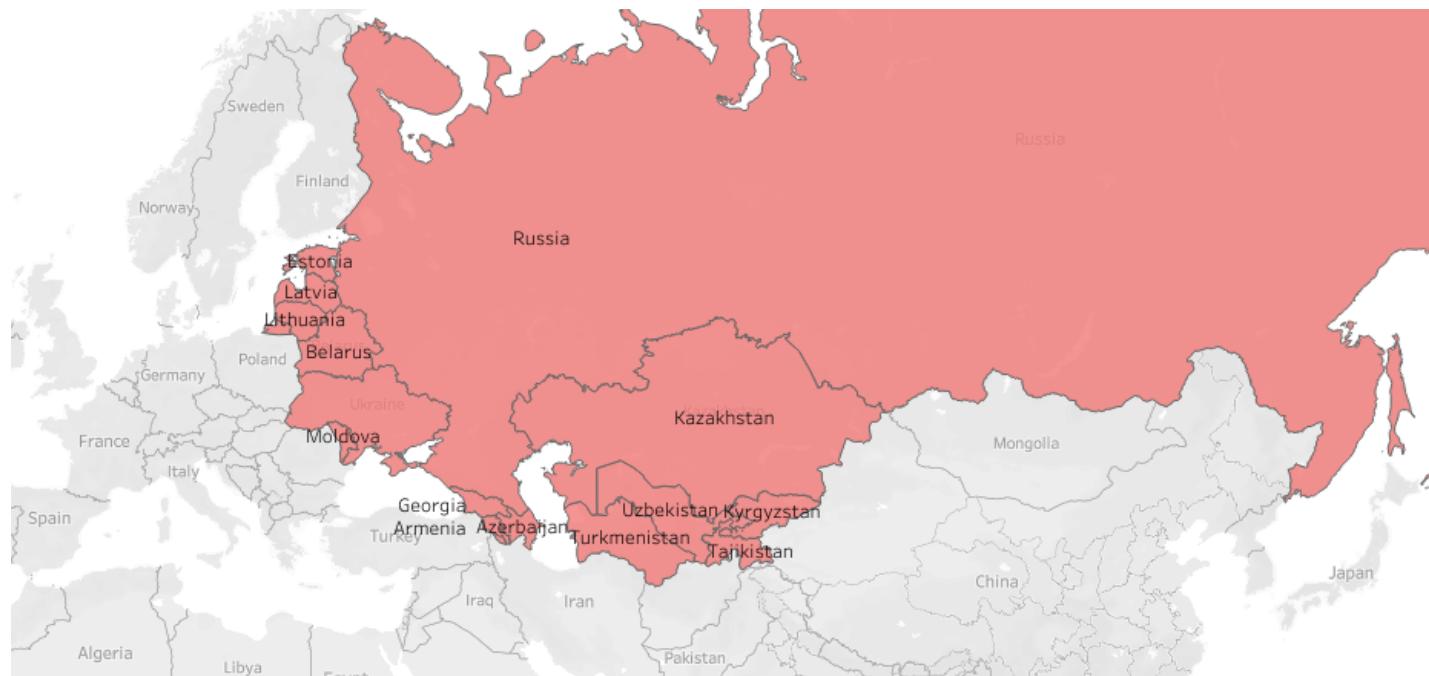
- The USA boycotted the Soviet games in 1980 to protest their invasion of Afghanistan, and the Soviet block boycotted in 1984 in response.

Sources: https://en.wikipedia.org/wiki/1980_Summer_Olympics_boycott

https://en.wikipedia.org/wiki/1976_Summer_Olympics

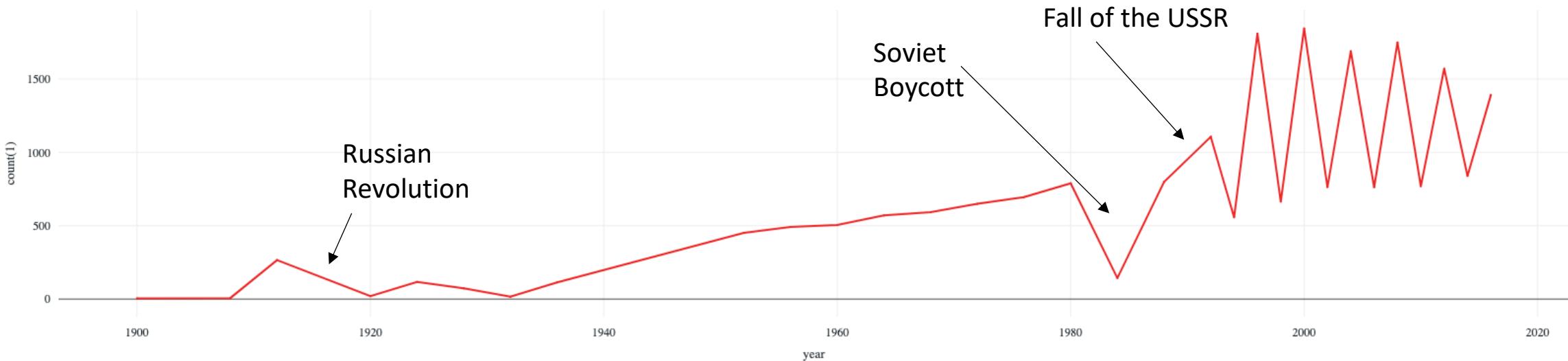
Soviet Union

- Former Soviet Socialist Republics: Armenia, Azerbaijan, Belarus, Estonia, Georgia, Kazakhstan, Kyrgyzstan, Latvia, Lithuania, Moldova, Russia, Tajikistan, Turkmenistan, Ukraine, Uzbekistan
- All events between 1920-1988 Olympics for the Soviet Union are included in "Russia (incl. USSR)" as we don't have the data to break out into separate states.



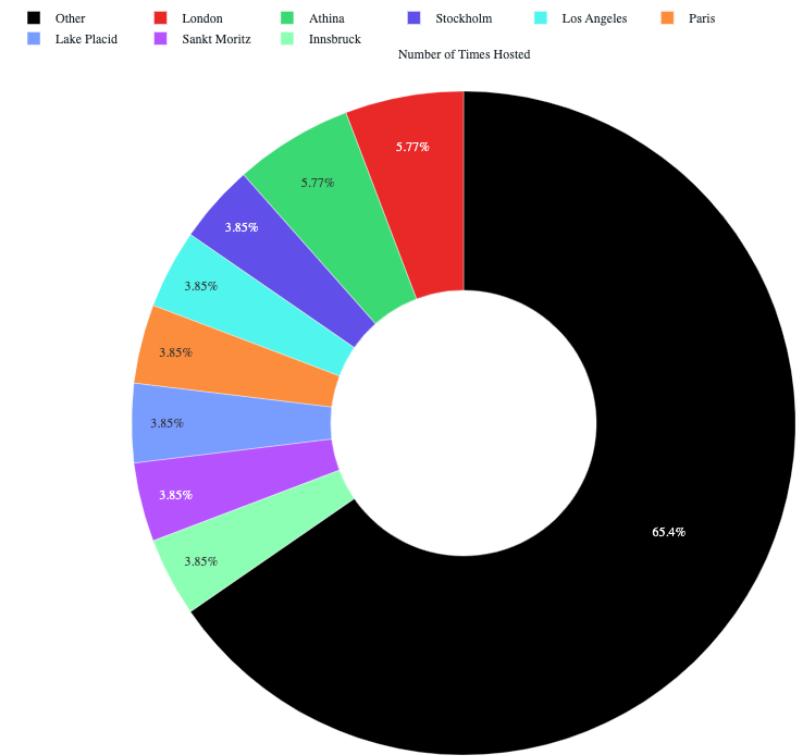
Russian Revolution/Fall of USSR

- Did the Russian Revolution or fall of the Soviet Union affect participation in Russia and countries once part of the Soviet Union?
- There was a major downturn in the year 1920, three years after the Russian Revolution.
- However, after the fall of the Soviet Union in 1991, the number of events went up, even accounting for the split of Summer/Winter after 1992.

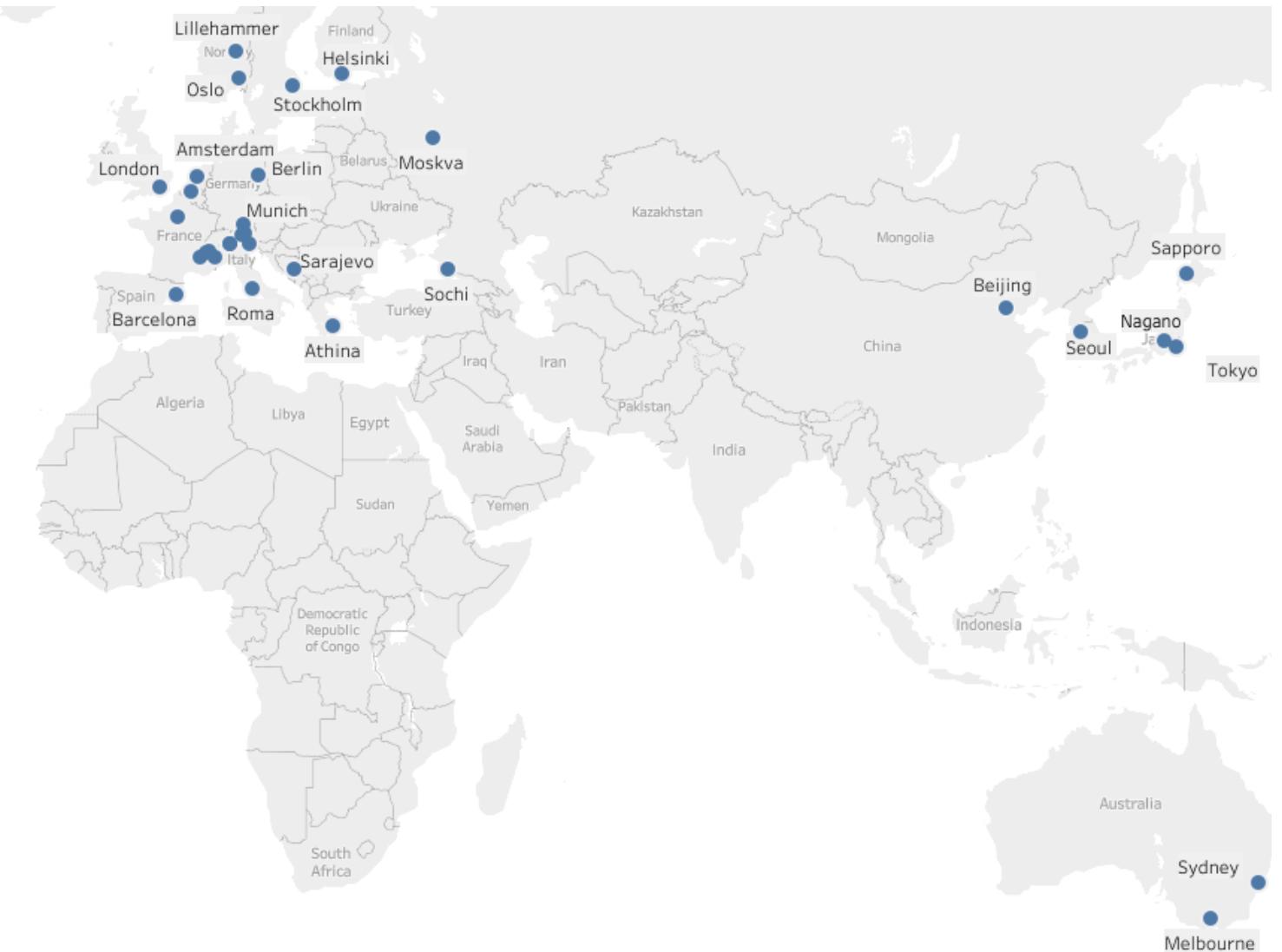


What City has Hosted the Olympics the Most?

- London + Athina (Athens) have hosted 3 times.
- Stockholm, Los Angeles, Paris, Lake Placid, Sankt Moritz, Innsbruck have hosted 2 times.
- Every other host city has hosted one time.
- Stockholm hosted in the games in 1912. The 1956 Summer Olympics were in Melbourne, but equestrian events were in Stockholm b/c of Australian quarantine laws, so it has 2 cities for the games.



Host Cities



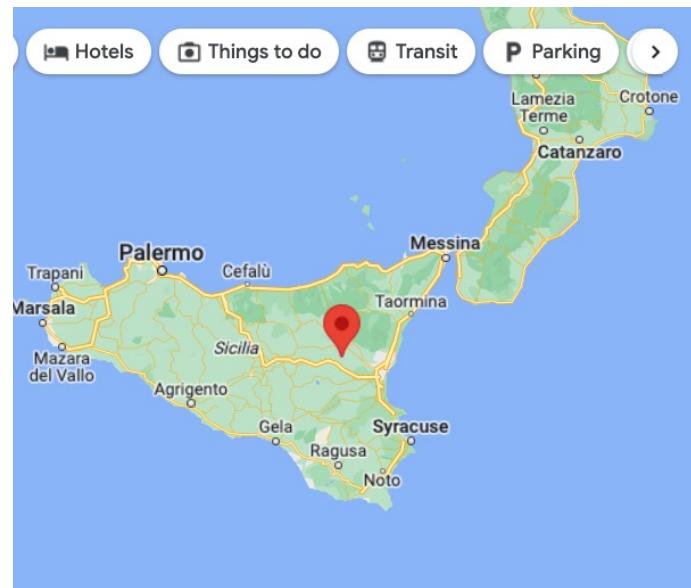
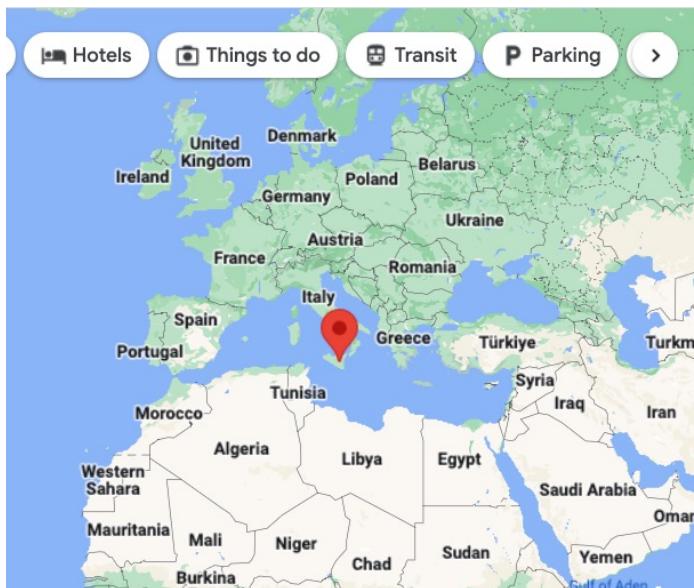
Close-up of European Host Cities

- Europe hosted so many times it needed a close-up.
- Another example of how the Olympics skew north.



Average Latitude/Longitude (athlete events)

- Just out of curiosity...what is the average latitude and longitude (using the country's capital). This is a average by athlete event, not individual athlete, or individual country.
- The average latitude and longitude is in Sicily, Italy.
- <https://www.google.com/maps/place/37%C2%B0036'36.0%22N+14%C2%B0043'48.0%22E/@38.4277752,14.229174,7.08z/>
- Is this measurement even relevant?



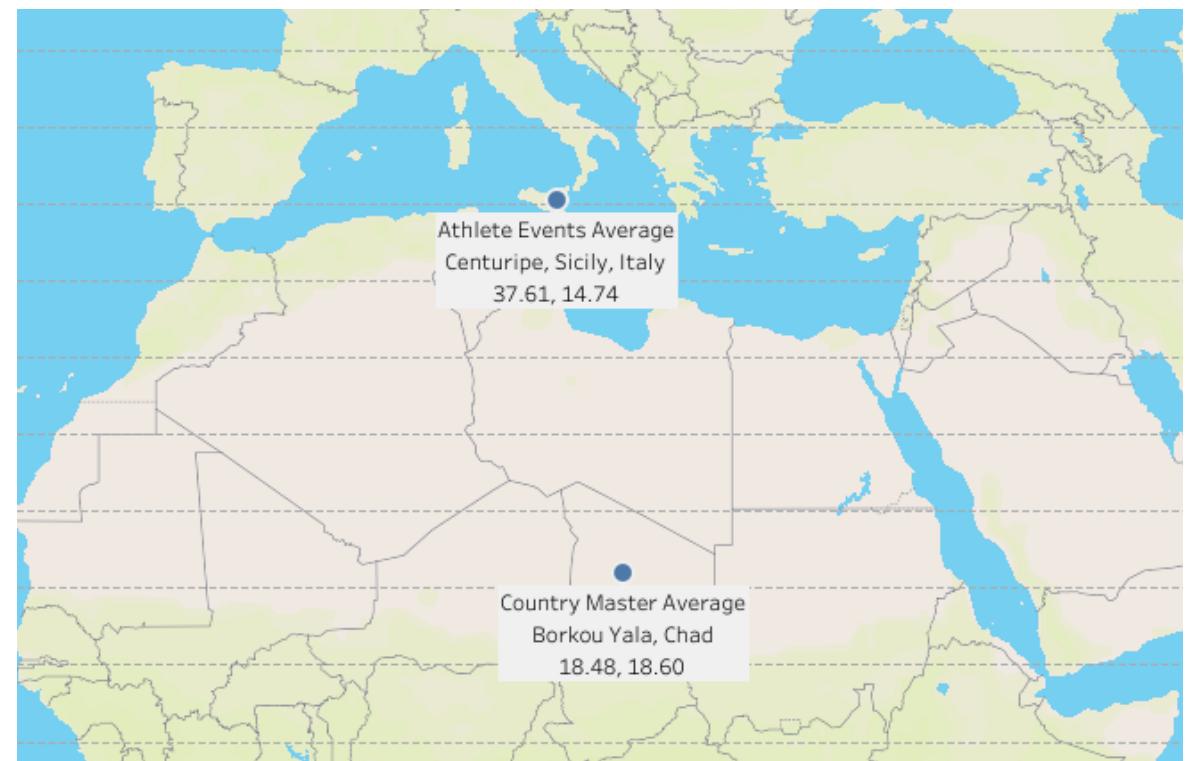
Average Latitude/Longitude (of countries)

- Using the Country Master table, the average latitude/longitude is in the country of Chad.
- <https://www.google.com/maps/place/18%C2%B0028'12.0%22N+18%C2%B0036'00.0%22E/@18.4704906,-17.260612,3z/data=!4m4!3m3!8m2!3d18.47!4d18.6>

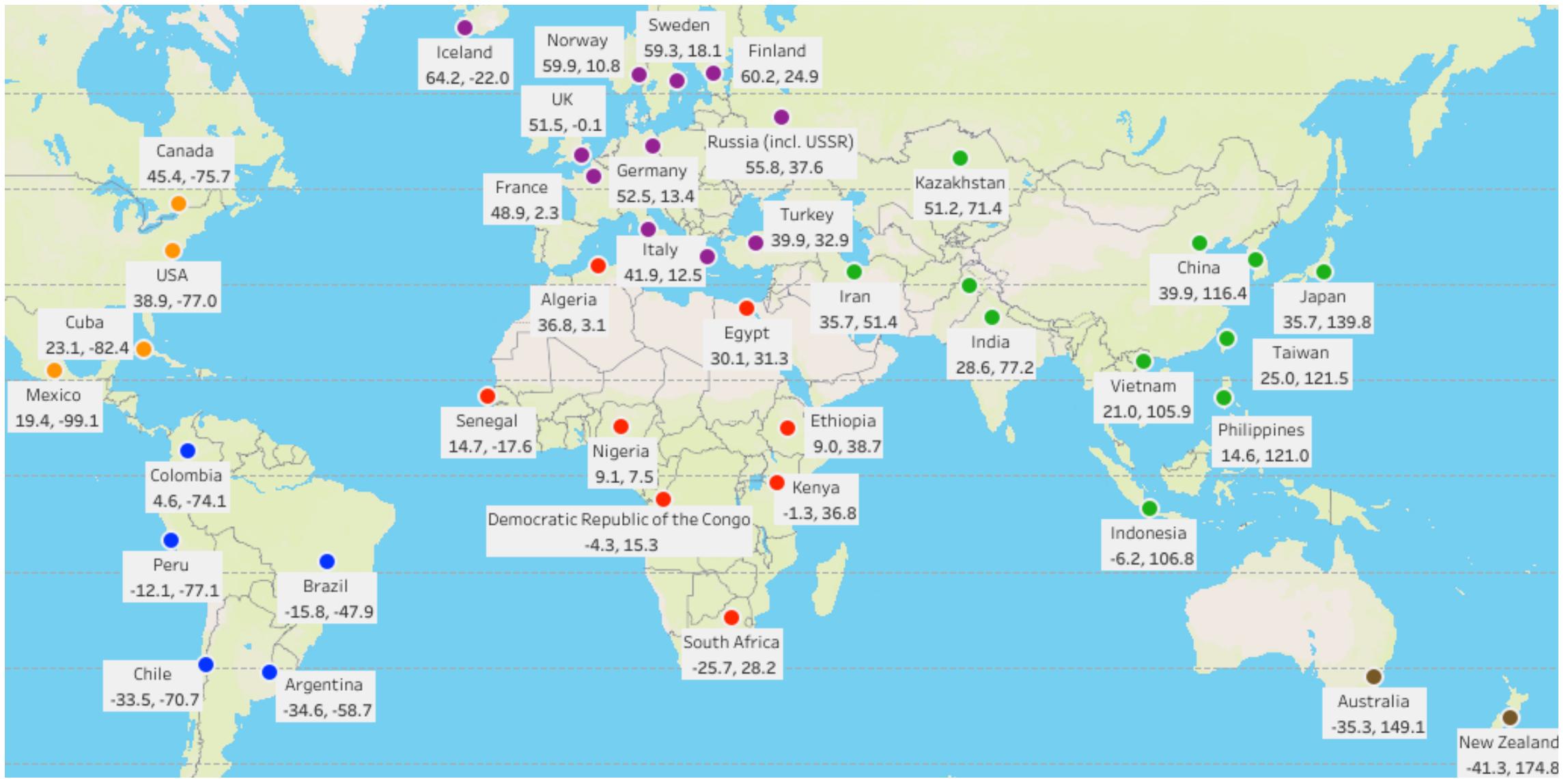


Compare Average Latitude/Longitude

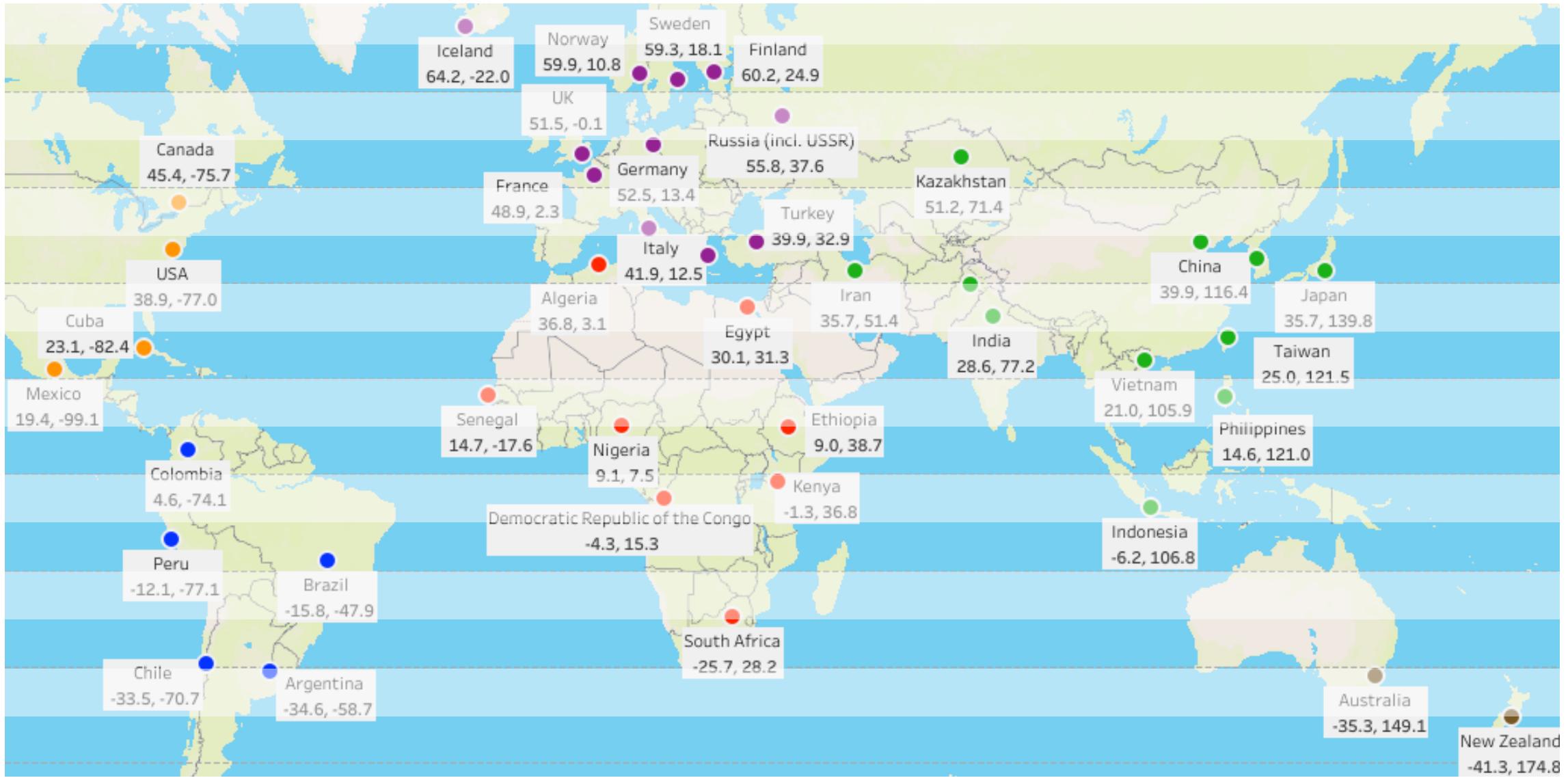
- Maybe the average latitude/longitude either by the Athlete Events or Country Master table is relevant.
- The "average" of this geographic location of countries is 1108 miles (2053 km) north of the equator, while the average geographic location of events (aka participation) is 2256 miles (4178 km) north of the equator.
- In other words, the Olympics are skewed north. Representation in the games are dominated by northern countries over the Global South.



Latitude Map Reference



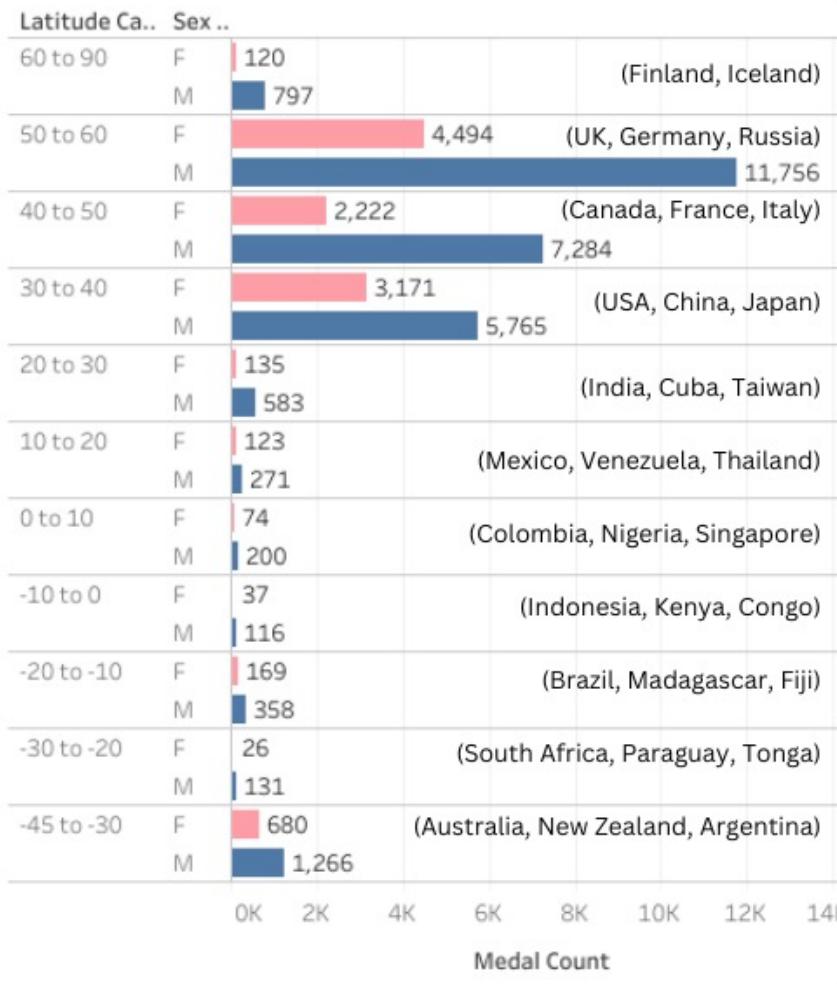
Latitude Category (10° Latitude bands)



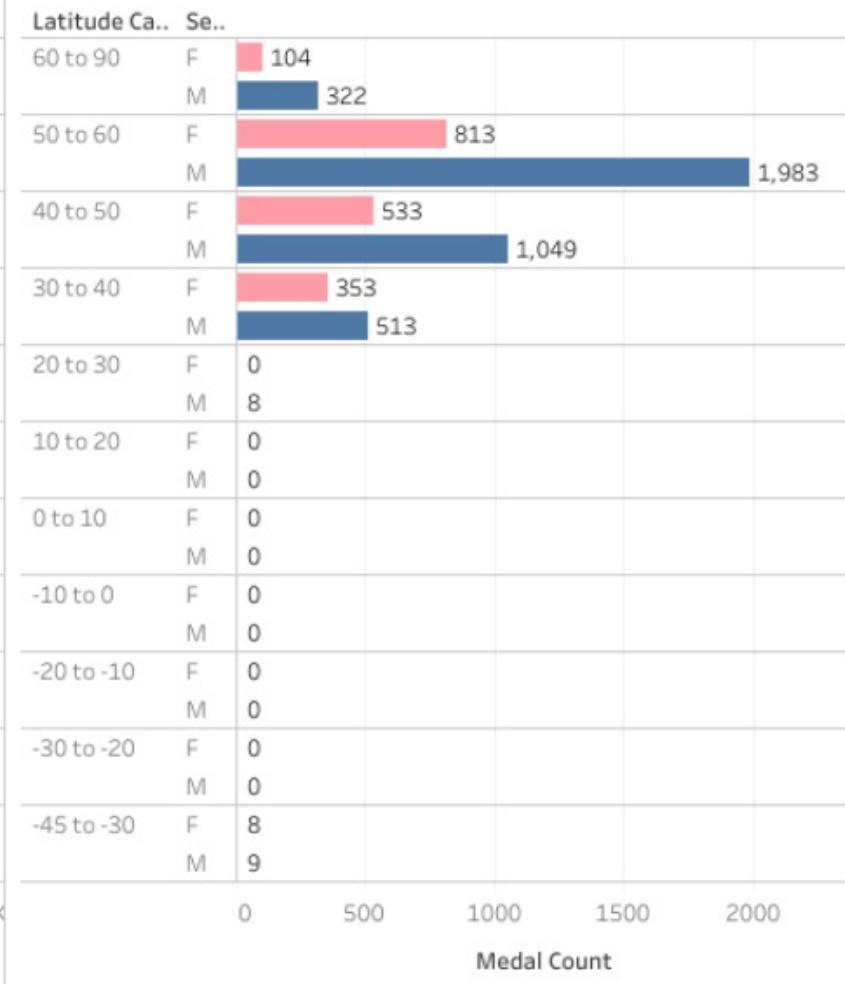
Are Northern countries more successful at Winter sports?

- Yes, according to the data. It shows an almost stepwise decline in medals for Winter sports for each category of latitude as you go south
- The same is true for medal count in the Summer, but it is even more prevalent in the Winter Season, in which countries near the equator have zero medals.
- The medal count for both seasons is much larger than the Winter only, because Summer has many more events.

Medals by Latitude and Sex (Both Seasons)

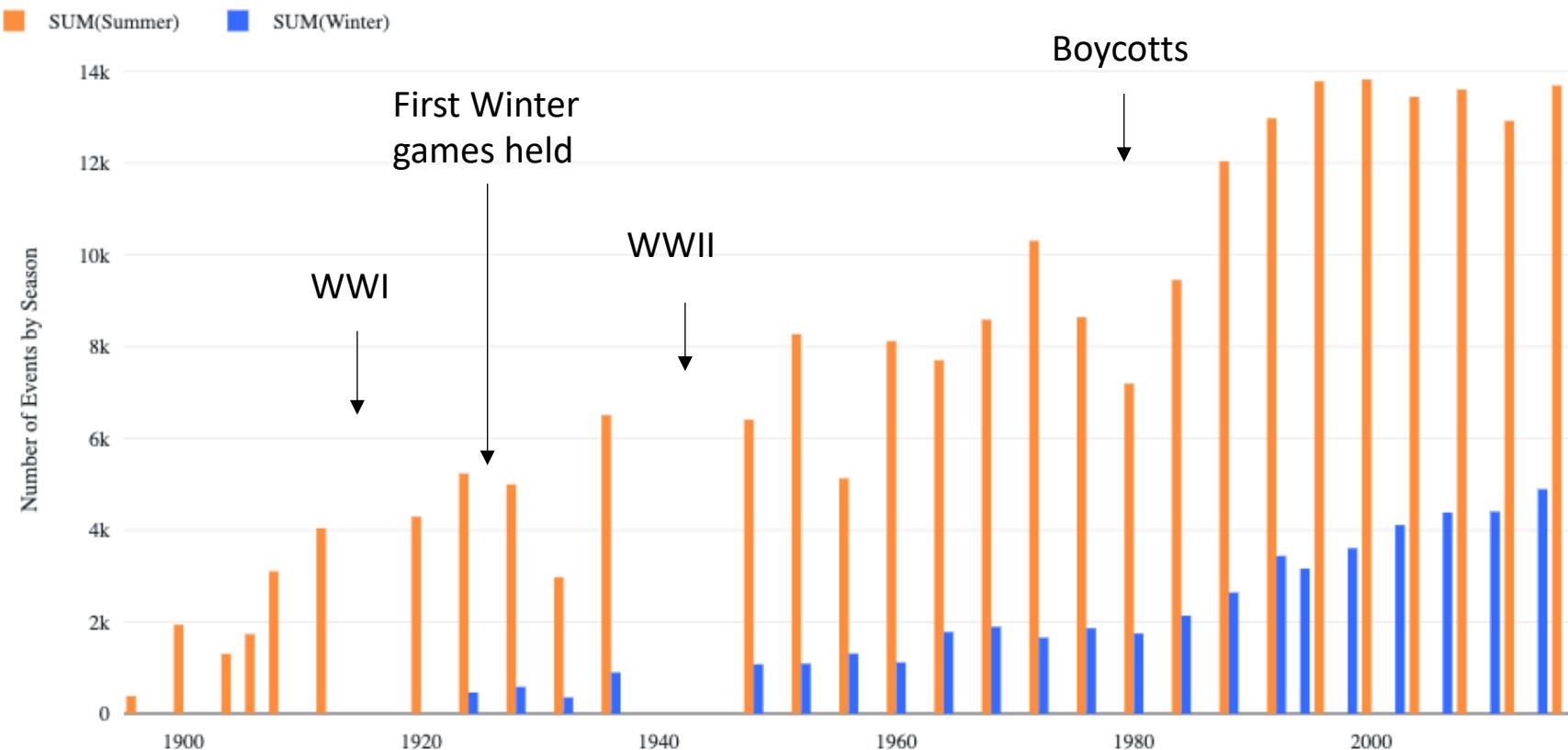


Medals by Latitude and Sex (Winter Only)



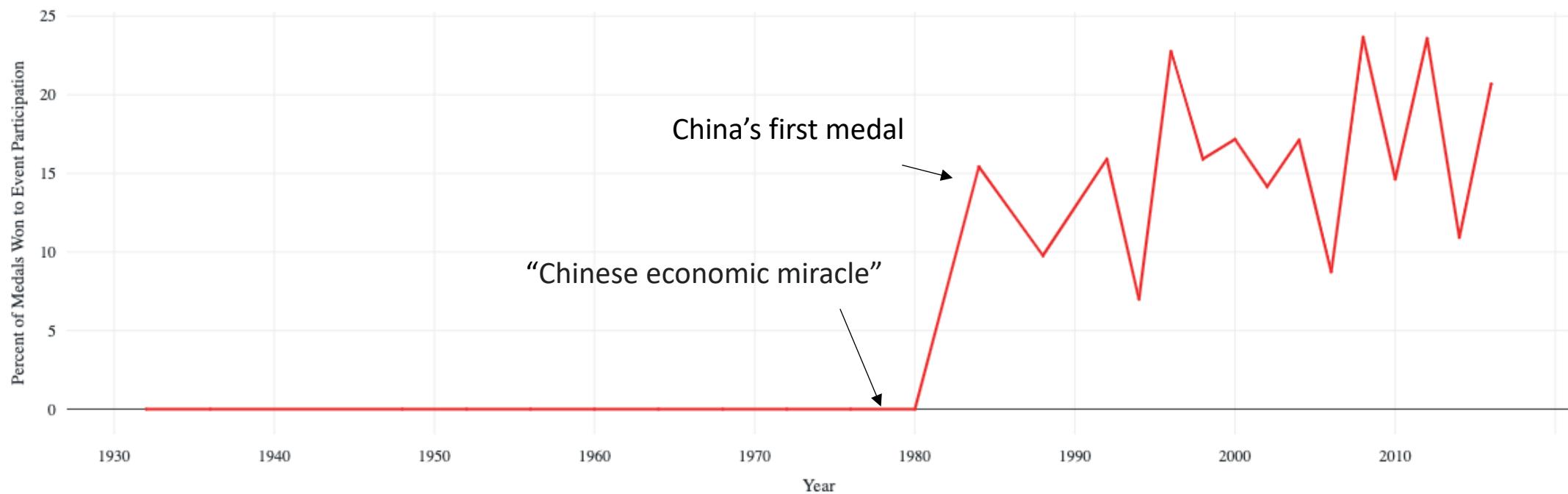
Are Summer sports more popular than Winter sports?

- Yes, and you can see by the numbers, as well as the percent values in years prior to 1992. After 1992, each year was 100% Summer or 100% Winter
- There is missing data during World Wars I + II, as the games were not held.



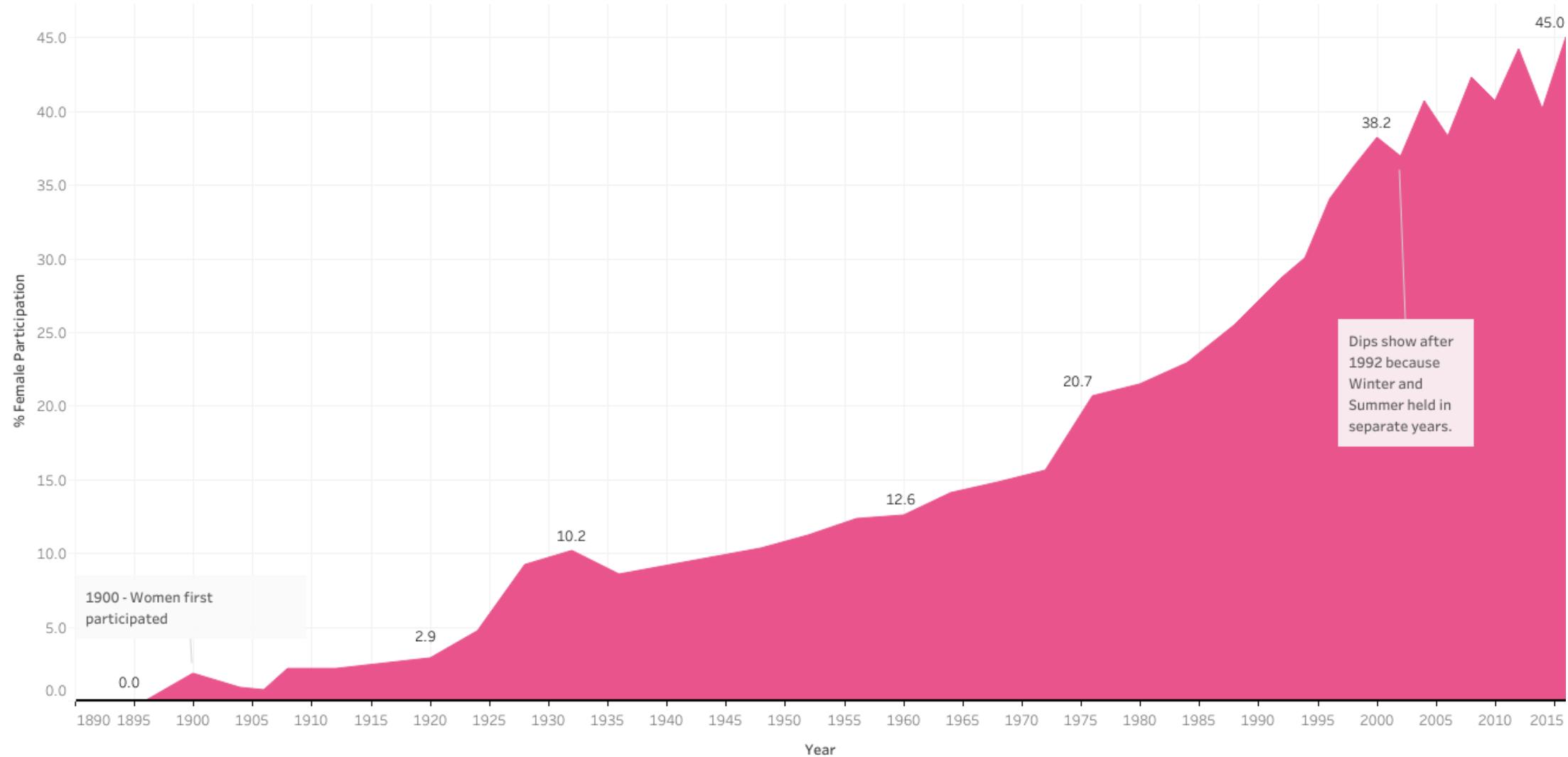
China

- This visualization looks at the rise of China, a country which has grown faster economically in the past 40 years than any other country.
- As you can see from the chart, China did not win a medal before 1984. However, by the 2008 Summer games, it peaked at almost 24%!



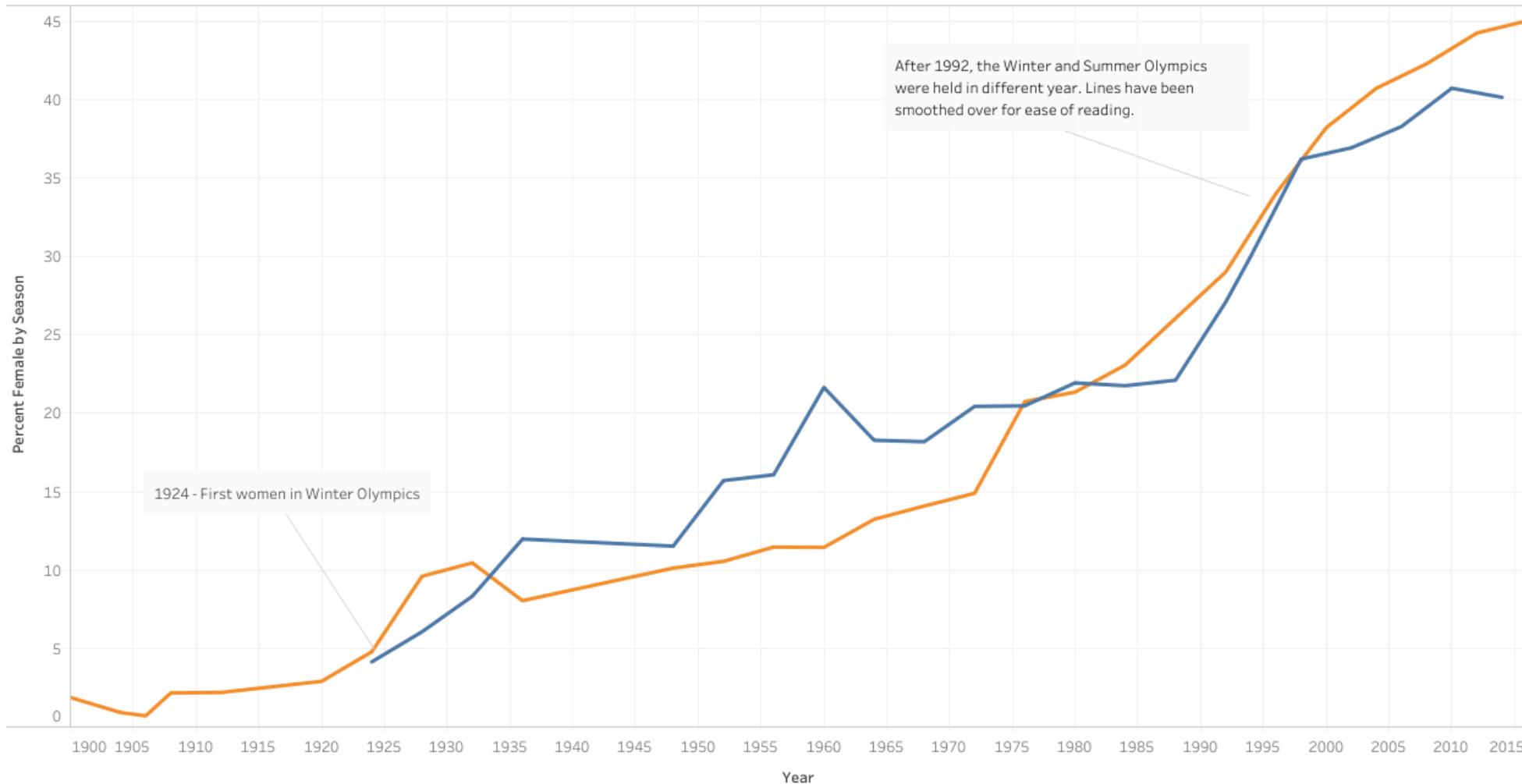
Female Participation by Year

Percent of Female Participation Per Year



Female Participation by Season

Percent of Female Participation Per Year by Season



Largest Percentage of Women Per Country

- What about percentage of women participating per country? The first three countries that show up as having the largest percentage of women are Kosovo, Timor-Leste, and Palau, very small countries with few athletes, but fourth place is China.

	country	# (F)	# (M)	# Total	% F
1	Kosovo	5	3	8	62.5
2	Timor-Leste	5	4	9	55.56
3	Palau	13	12	25	52
4	China	2992	2834	5826	51.36
5	Marshall Islands	7	7	14	50
6	Cape Verde	9	9	18	50

Smallest Percentage of Women Per Country

- What about percentage of women participating per country? The first three countries that show up as having the largest percentage of women are Kosovo, Timor-Leste, and Palau, very small countries with few athletes, but fourth place is China.

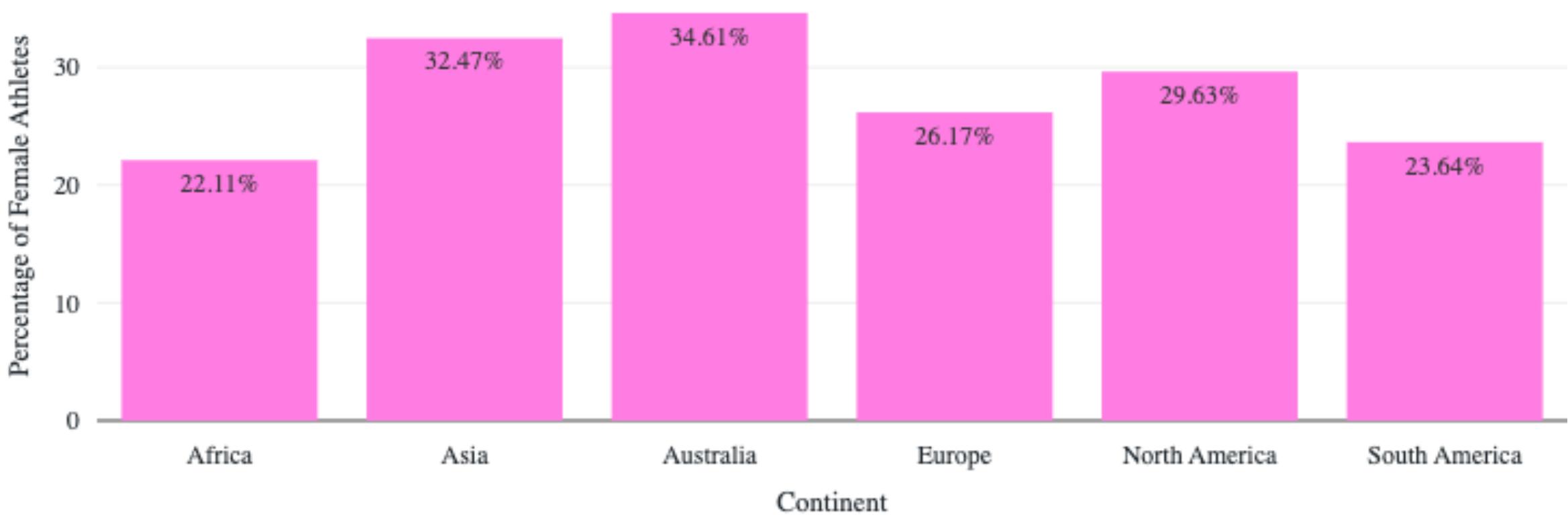
	country	# (F)	# (M)	# Total	% F
1	Kuwait	4	280	284	1.41
2	Pakistan	11	551	562	1.96
3	Saudi Arabia	5	225	230	2.17
4	Iraq	7	225	232	3.02
5	Qatar	7	185	192	3.65
6	Afghanistan	5	121	126	3.97

Afghanistan, Female

- Even though Afghanistan is in sixth place, I was curious about the years where women participated.
- Four women, five events. It is interesting that no women participated before the Taliban took over in the 1990s. Very inspiring to us all!

	id	name	games	sport	event
1	99303	Friba Razayee	2004 Summer	Judo	Judo Women's Middleweight
2	132125	Robina Muqim Yaar	2004 Summer	Athletics	Athletics Women's 100 metres
3	132125	Robina Muqim Yaar	2008 Summer	Athletics	Athletics Women's 100 metres
4	61961	Tahmina Kohistani	2012 Summer	Athletics	Athletics Women's 100 metres
5	133277	Kamia Yousufi	2016 Summer	Athletics	Athletics Women's 100 metres

Percent Female by Continent



Which Olympian won the most medals?

- Michael Phelps with 29 total medals, and 23 of them are gold.

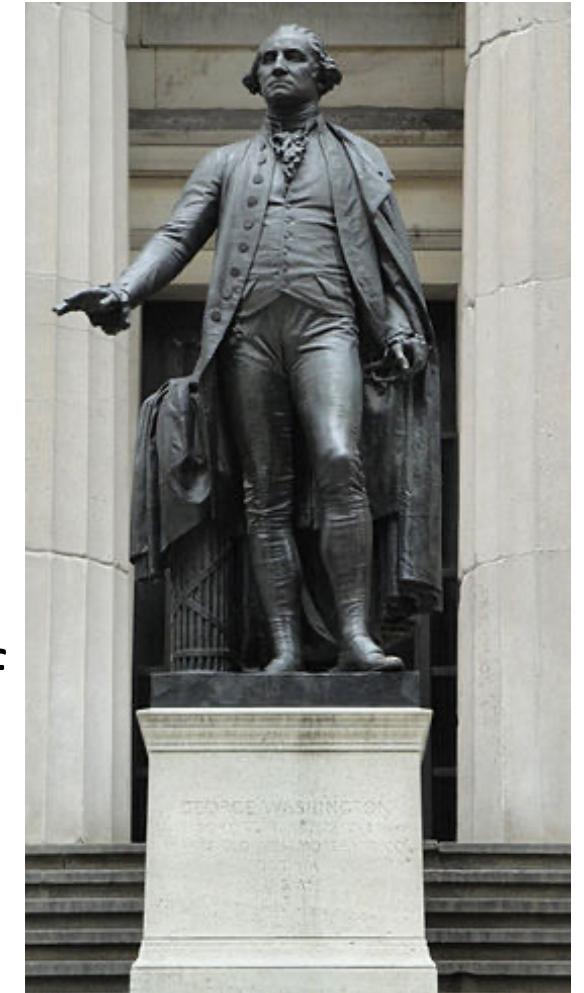
	id	name	country	medals	gold	silver	bronze
1	94406	Michael Fred Phelps, II	USA	28	23	3	2
2	67046	Larysa Semenivna Latynina (Dir	Russia (incl. USSR)	18	9	5	4
3	4198	Nikolay Yefimovich Andrianov	Russia (incl. USSR)	15	7	5	3
4	74420	Edoardo Mangiarotti	Italy	13	6	5	2
5	89187	Takashi Ono	Japan	13	5	4	4
6	11951	Ole Einar Bjørndalen	Norway	13	8	4	1

Youngest Olympian?

- In 1896, Dimitrios Loundras was only 10 years old when he won a bronze competing in gymnastics, "Gymnastics Men's Parallel Bars, Teams"
- I wondered if this is a data entry error. I have a hard time believing that a 10-year-old was competing against adults and coming 3rd place.
- However, it turns out I was wrong. He was actually 10 years old, the youngest ever in the Olympics.

Oldest Olympian?

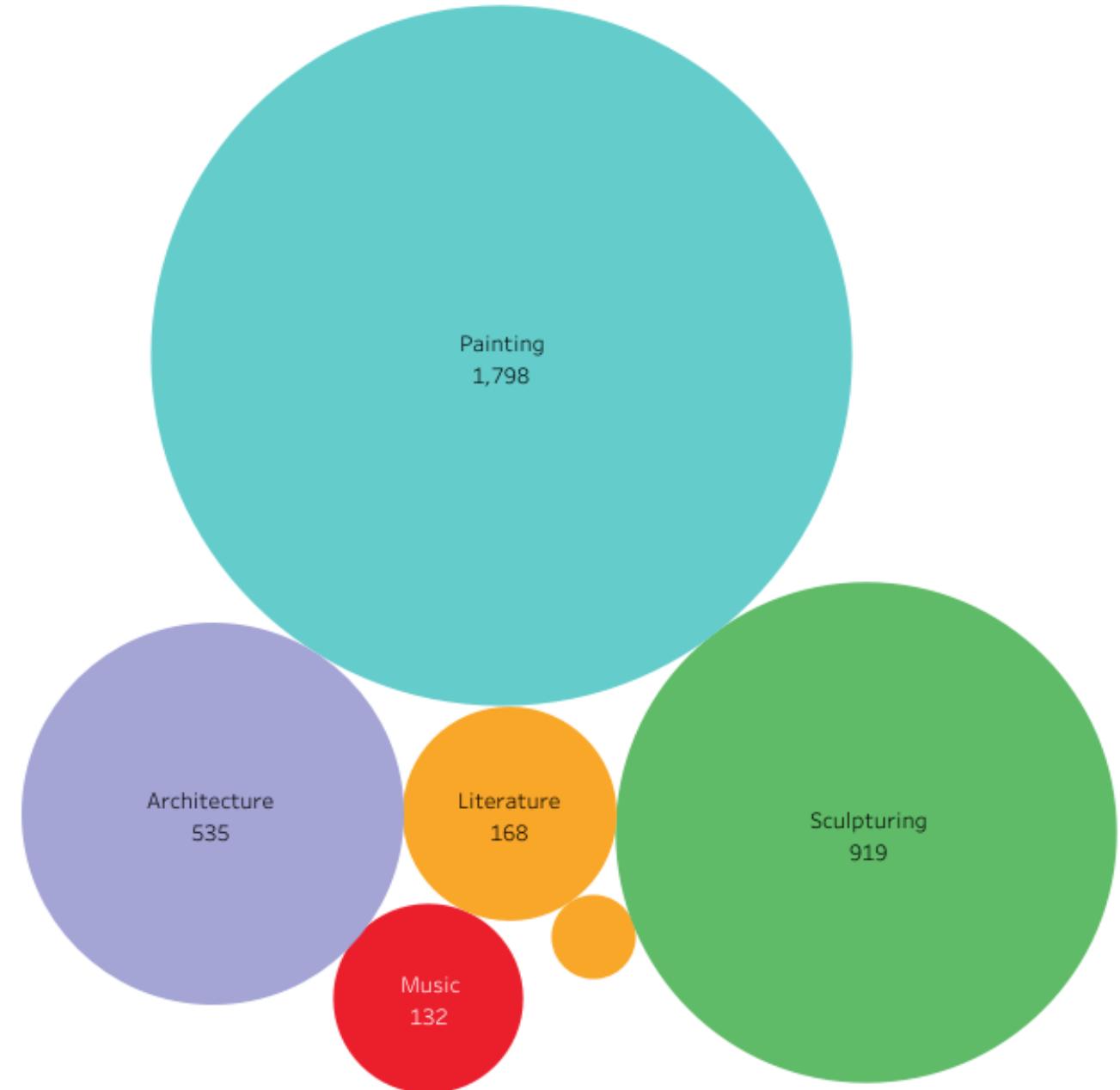
- John Quincy Adams Ward - 97 years (in the dataset)
- However, he only lived to be 79 years old. (d. 1910)
- At first, I thought someone typed “97” instead of “79.”
- His work was included in the 1928 event “Art Competitions Mixed Sculpturing, Statues.”
- He was a famous sculptor most known for the statue of George Washington on Wall Street, NYC. (right)
- Since data on age, height, and weight may be used to compare performance, it’s *probably* not a good idea to include dead people in the calculations.



Art Competitions

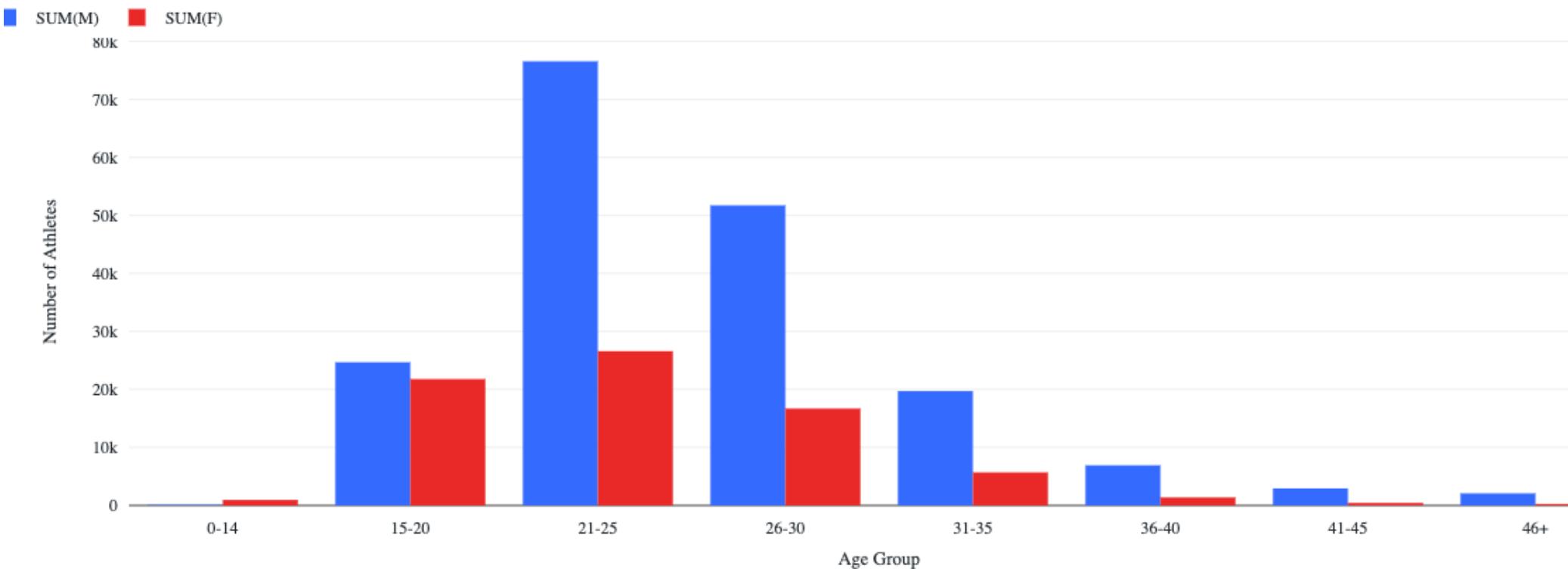
- Art Competitions was a Summer sport from 1912 to 1948. It peaked in 1932 when it was the most popular sport that year.
- Events in this category include Painting, Architecture, Literature, Music, and Sculpturing.
- Average age of "athletes" in this sport is 44.66 years old.

Art Competitions, Number of Events

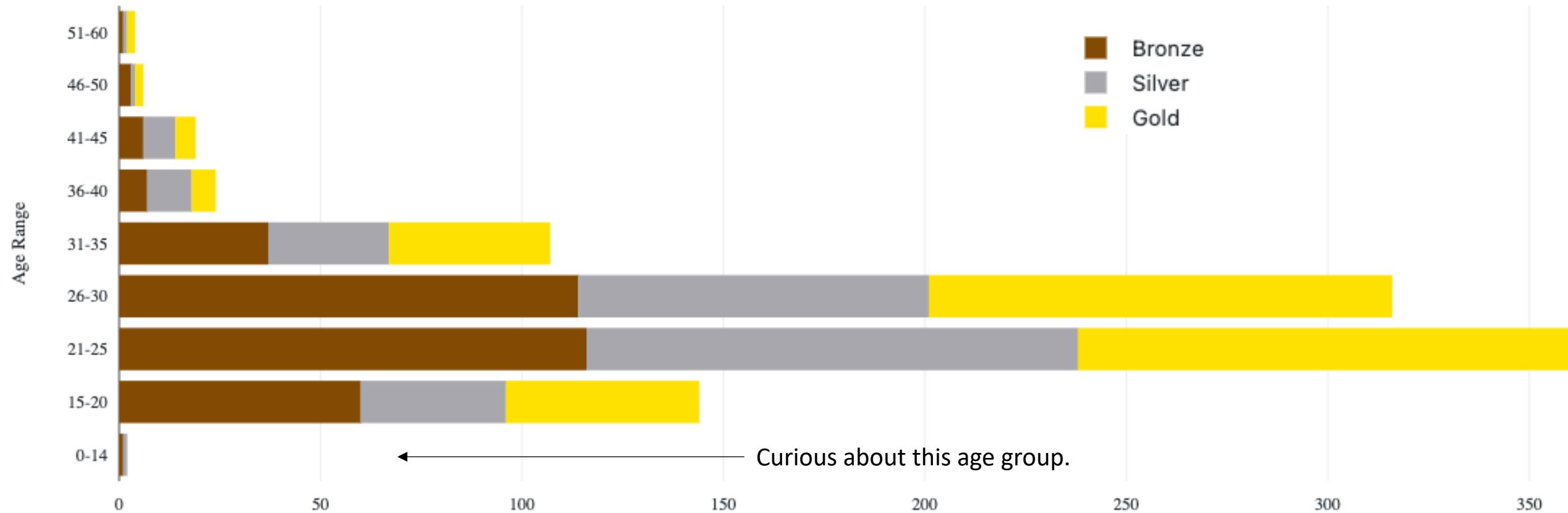


Athletes by Age Group

- Athletes counted for each age group they have a record for.
- Males outnumber females in every category except '0 to 14'

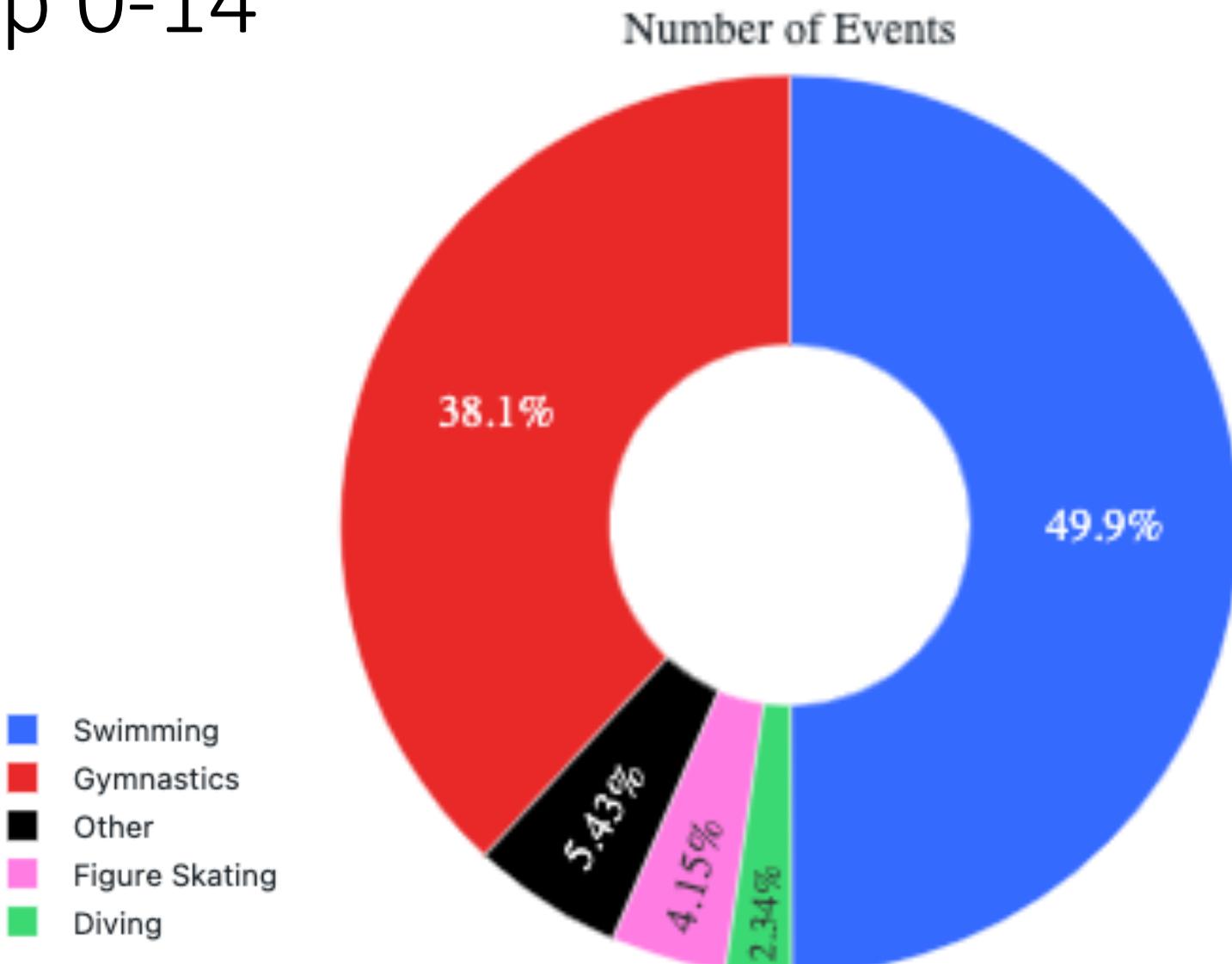


Medal by Age Group



Girls, Age Group 0-14

- Girls aged 0-14 participated in 16 different sports, but the bulk of them are in Swimming and Gymnastics.



Boys, Age Group 0-14

- Meanwhile, boys aged 0 to 14 participate in Swimming and Rowing events.

Number of Events

Event Category	Percentage
Swimming	37.3%
Rowing	29.1%
Other	13.4%
Diving	8.96%
Figure Skating	7.46%
Gymnastics	3.73%

Height, Weight, Age

- One of my hypothesis is that Height, Weight, and Age play a major role in your chances of success in the Olympics.
- Because of this, I have filtered out “Art Competitions” when calculating correlations between these measurements.
- Those participants are more “artists” than “athletes.”

Lack of Height or Weight Change

- Every single athlete (that has competed in more than one event) has the exact same weight and height in each of their events, even if they participated in multiple games over multiple years.
- I find this hard to believe, especially if the athlete starts participating at a young age. Nobody's weight stays *exactly* the same, even athletes who make exercise and nutrition a top priority.

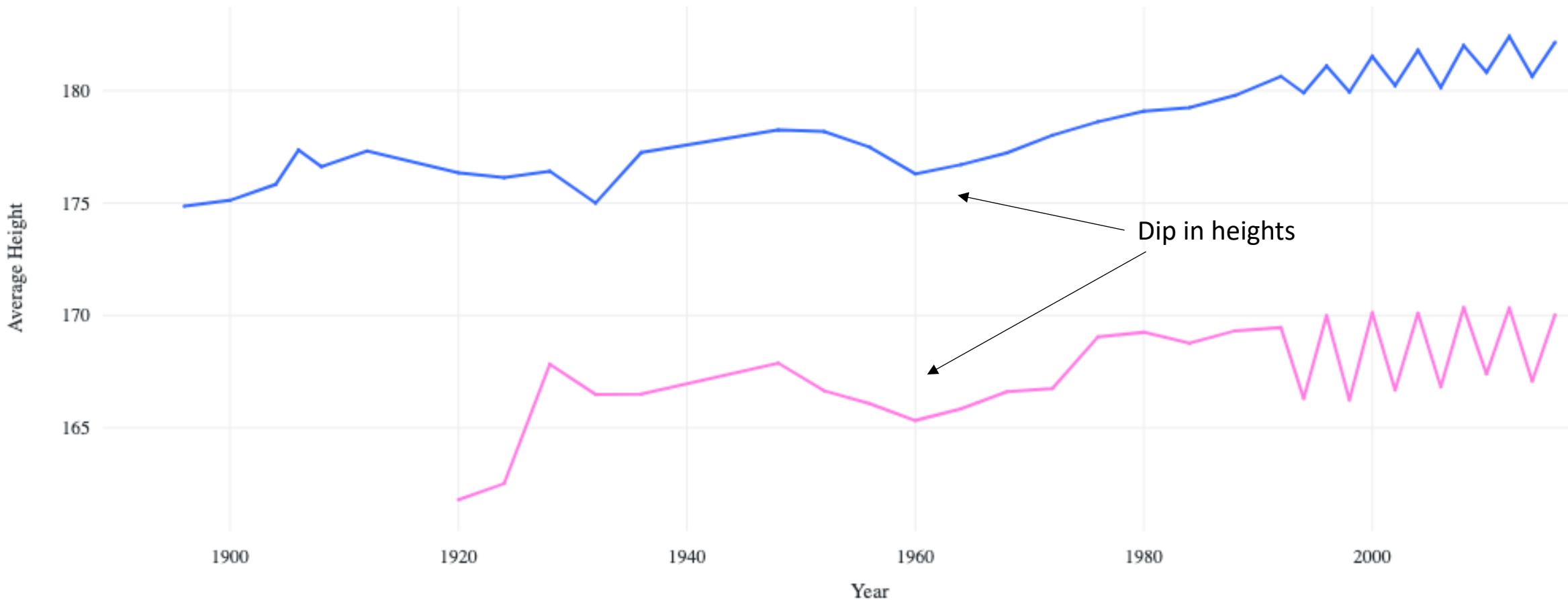
Increased Height



Has the average height of an athlete increased over the years?

- According to this query, no. The 1990s and the earliest part of the 20th Century had some of the tallest average athletes? Why?
- Women participated in small numbers until the 2nd half of the 20th Century.
- Even though there are records for female Olympians before 1920, they had null values for height, and weren't included in the average.
- What if we break it down by sex?

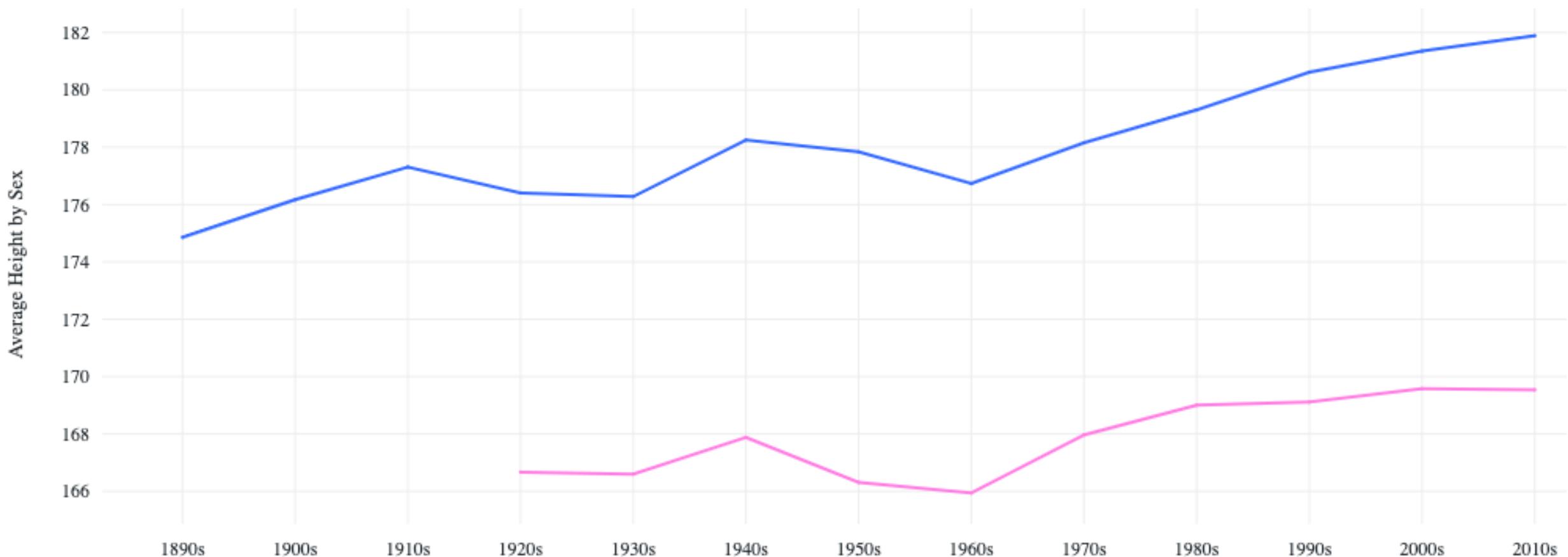
Height Increase By Sex



Average Height of Male/Female per Year

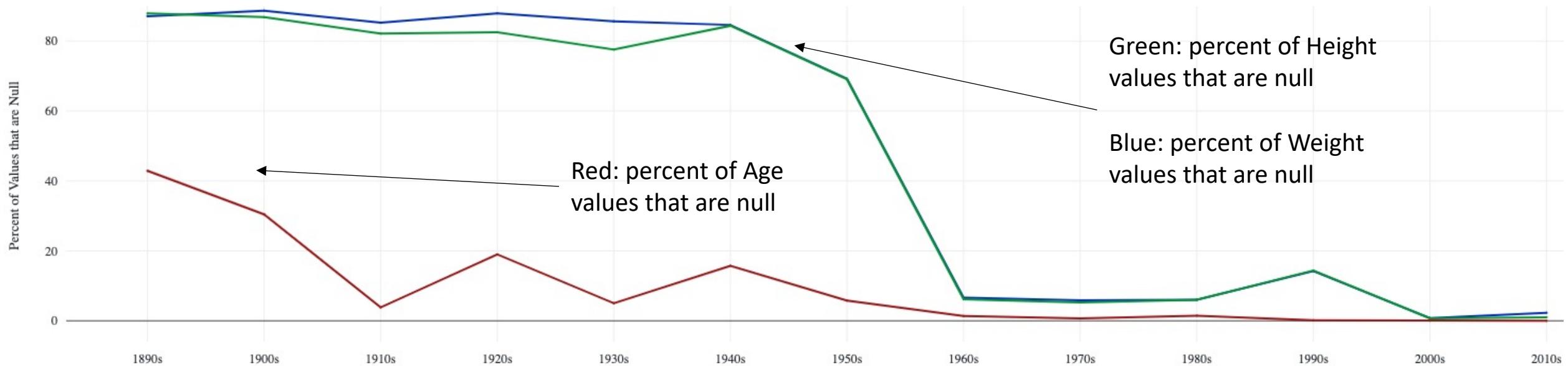
- Have they both increased when measured separately?
- The data says yes for both sexes, even if only slightly.
- It's interesting to see the season split after 1992
- I would have never guessed that the average Winter athlete is shorter than the average Summer athlete!

Height by Decade (smoothed out)



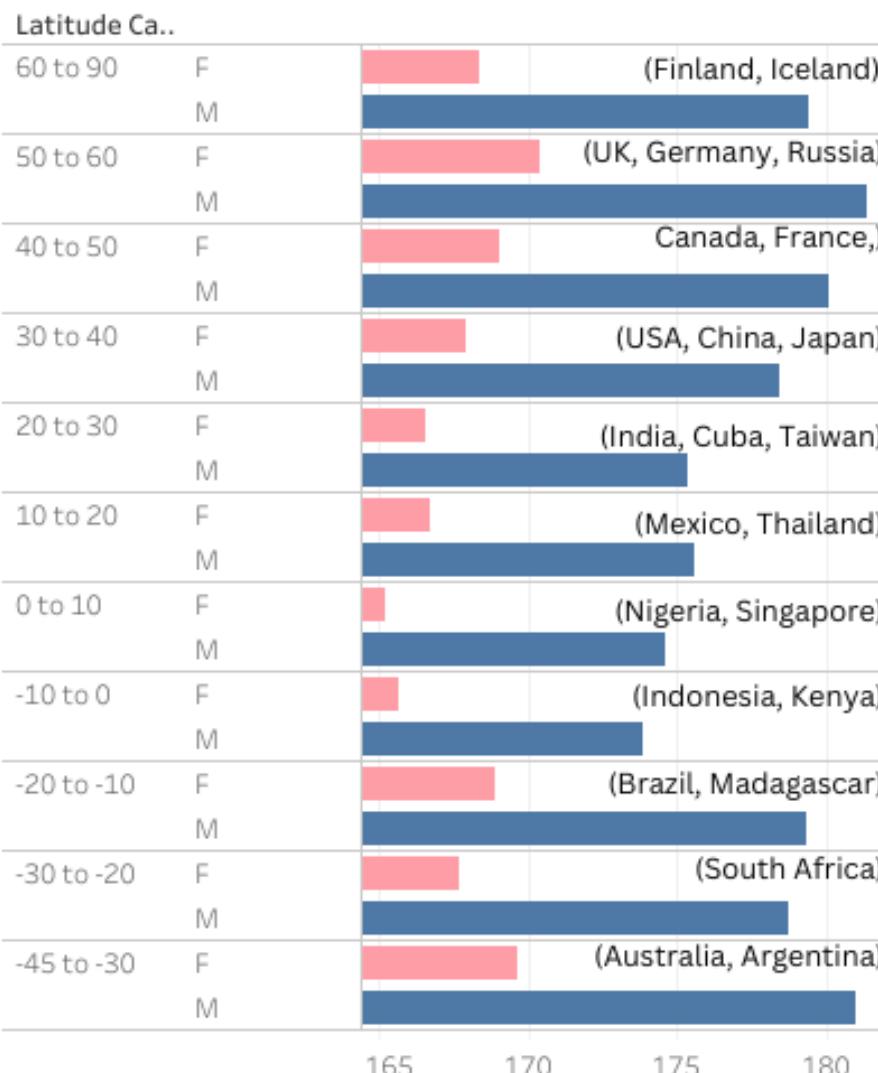
Visualization of height by decade and sex.

- What accounts for the dip in height between 1940 and 1970?
- Starting in the 1950s, the Olympics improved record-keeping for height/weight values as the percent of null values went down drastically.



Does Latitude affect height?

Latitude, Sex, and Height



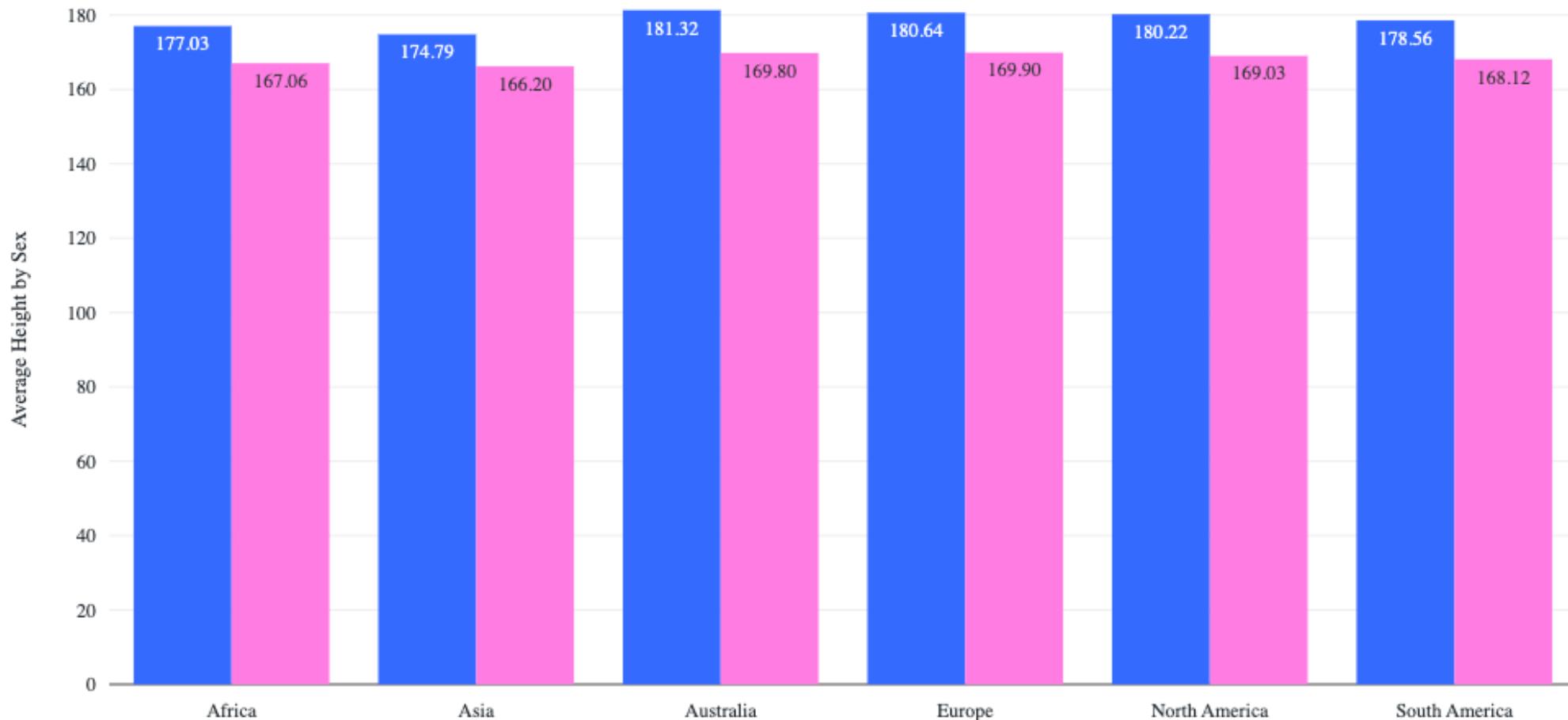
Height by Sex and Latitude

Sex

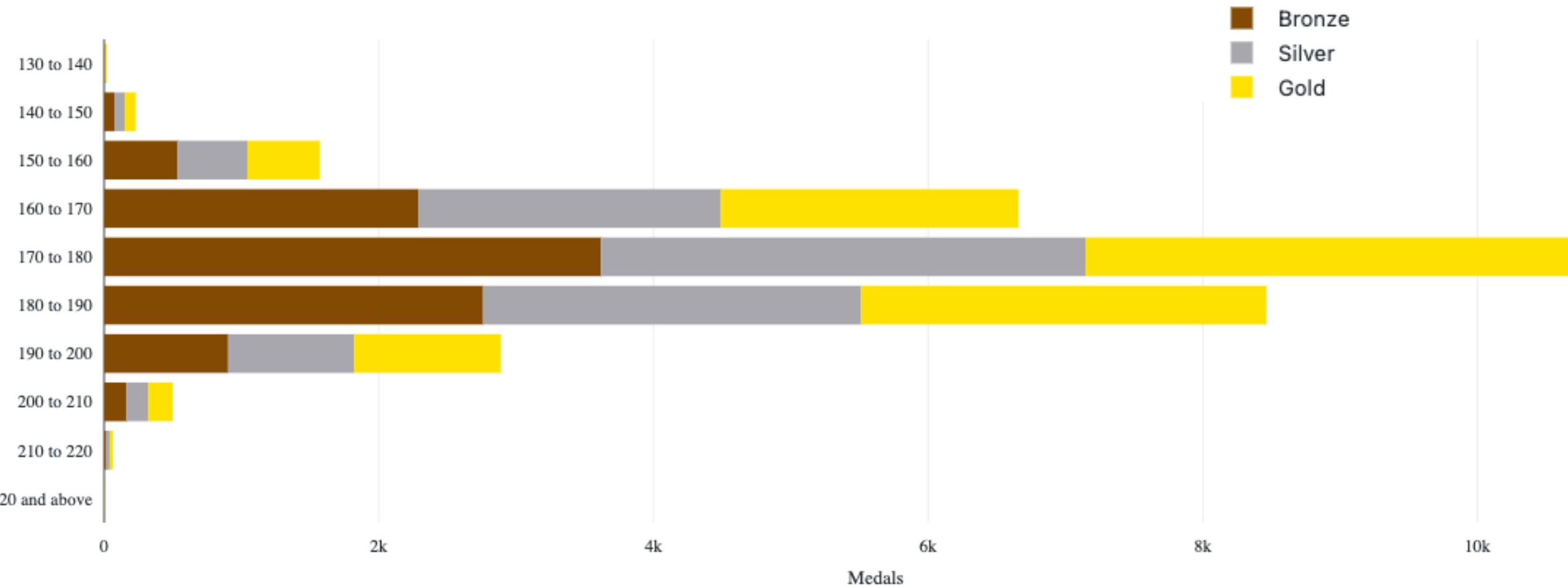
Latitude Ca..	F	M
60 to 90	168.36	179.42
50 to 60	170.38	181.36
40 to 50	169.09	180.09
30 to 40	167.94	178.41
20 to 30	166.56	175.33
10 to 20	166.77	175.61
0 to 10	165.22	174.60
-10 to 0	165.65	173.84
-20 to -10	168.90	179.33
-30 to -20	167.74	178.71
-45 to -30	169.64	180.94

Height by Continent by Sex

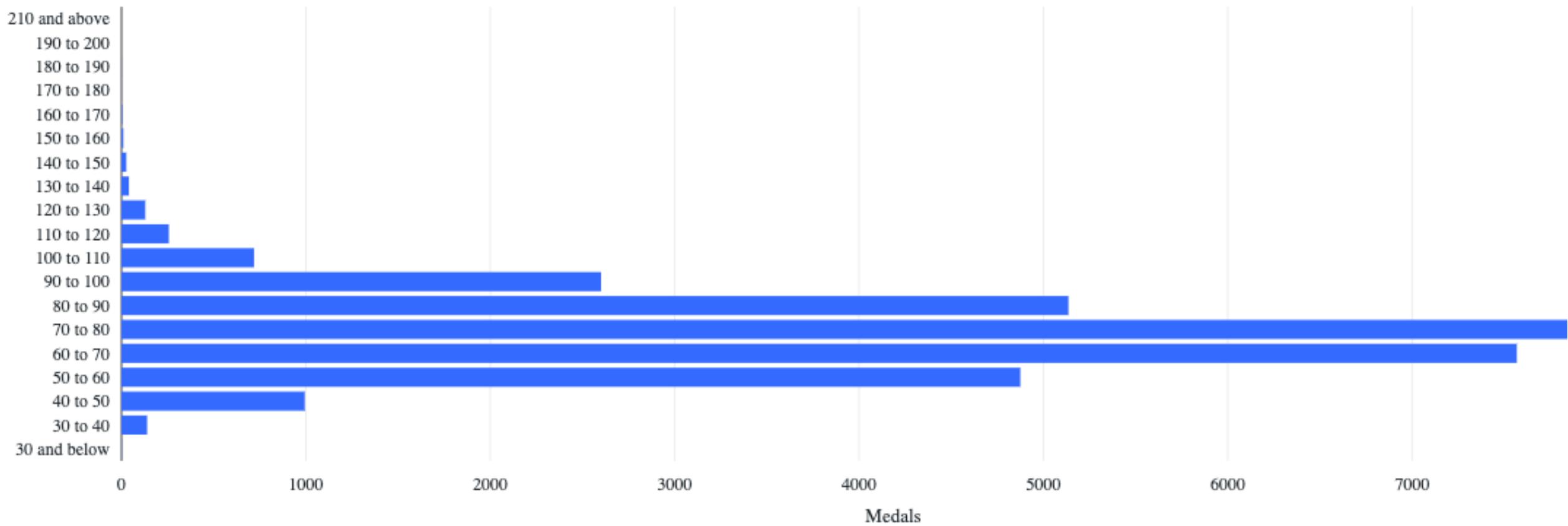
■ AVG(M) ■ AVG(F)



Height to Medals Won

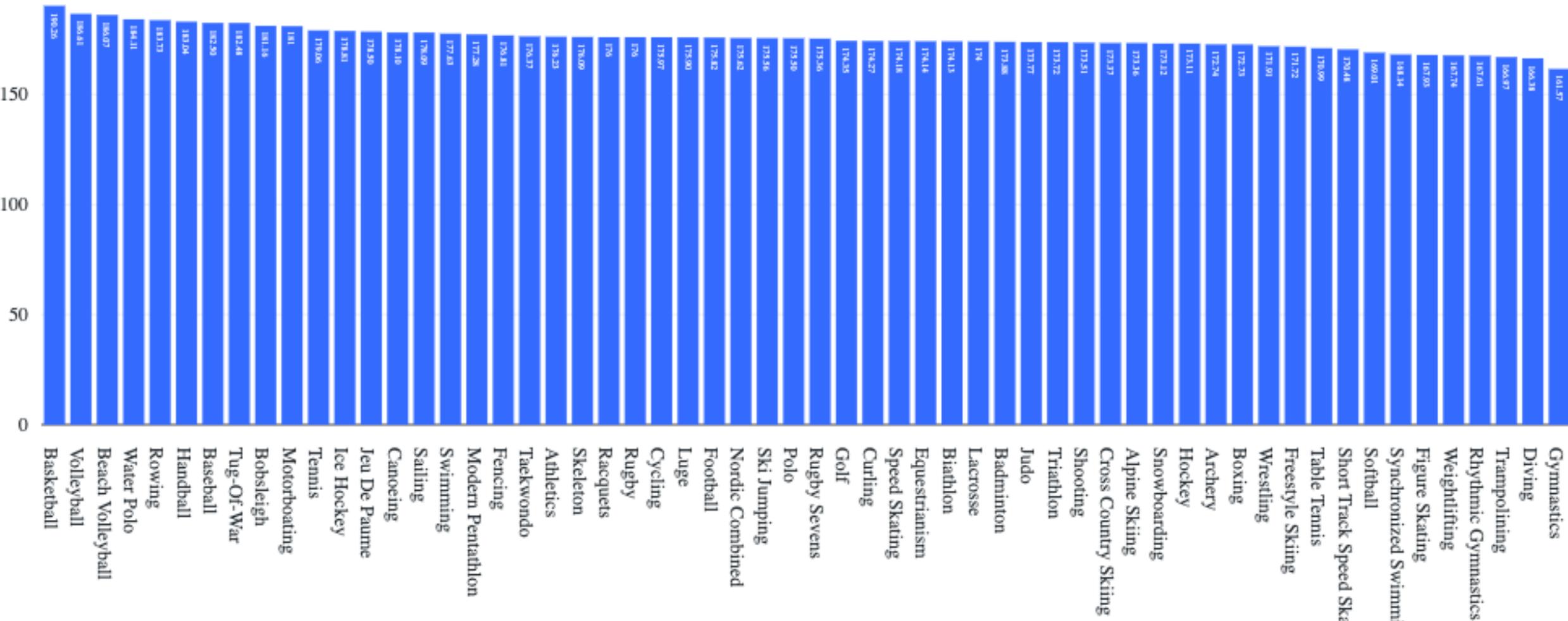


Medals Won by Weight Group



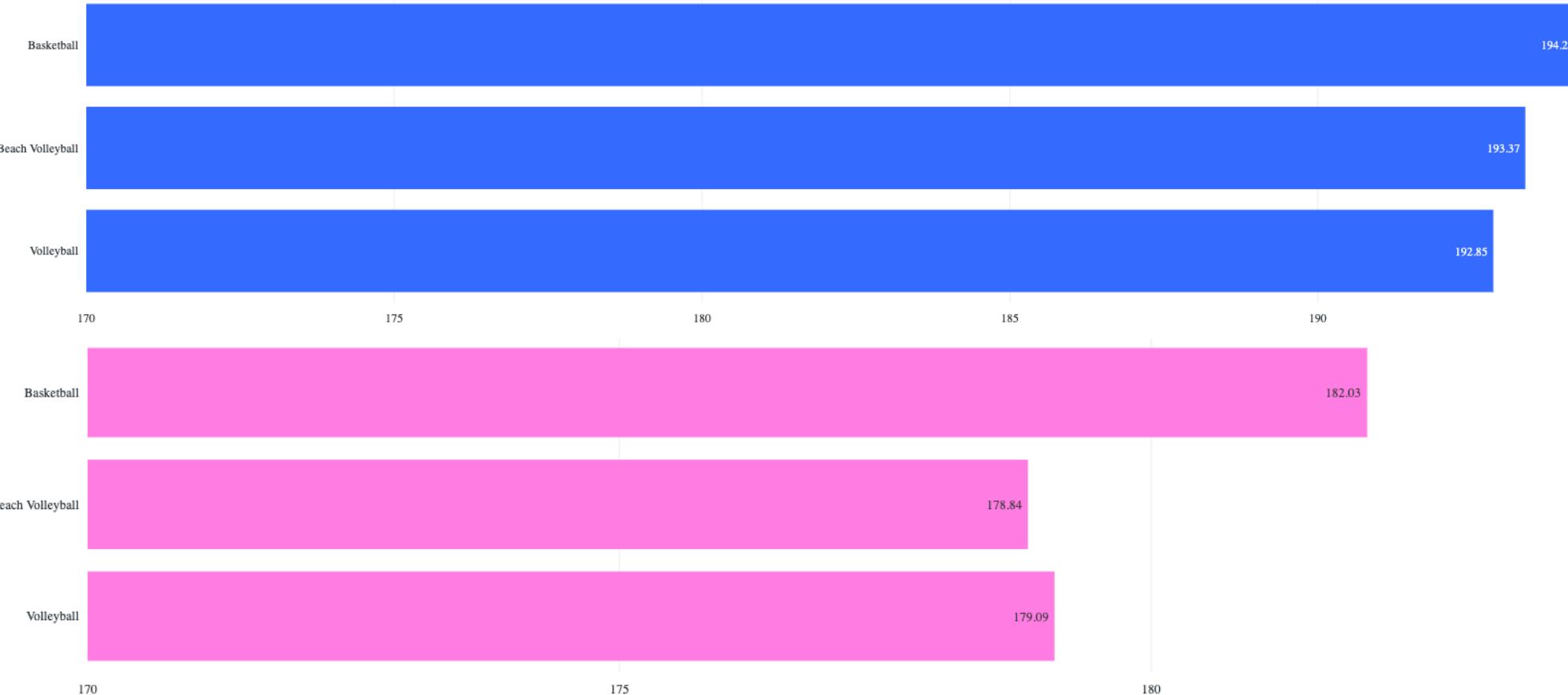
Height By Sport, by Sex

- Basketball has the tallest athletes, while Gymnastics has the shortest.

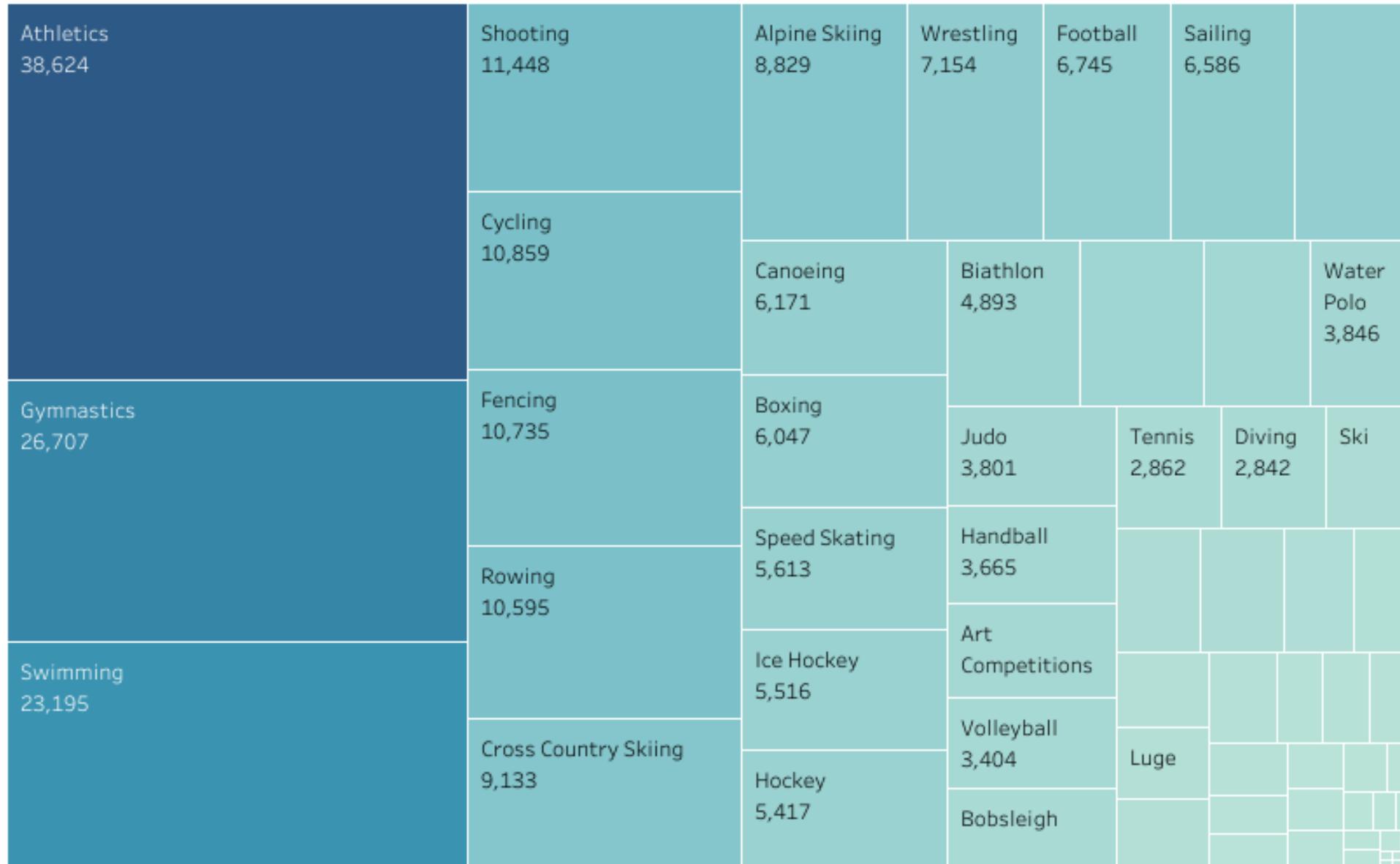


Top 3 Sports by Height, by Sex

- Both sexes top 3 are Basketball, Beach Volleyball, and Volleyball as having the tallest athletes.

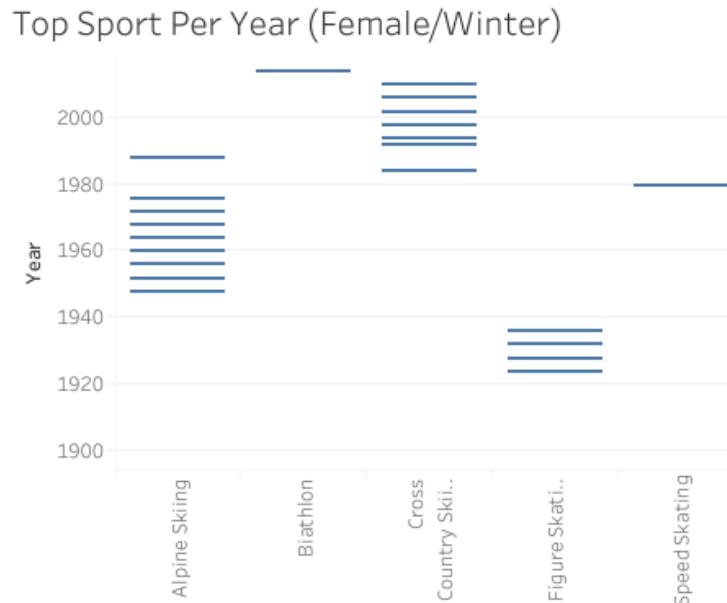
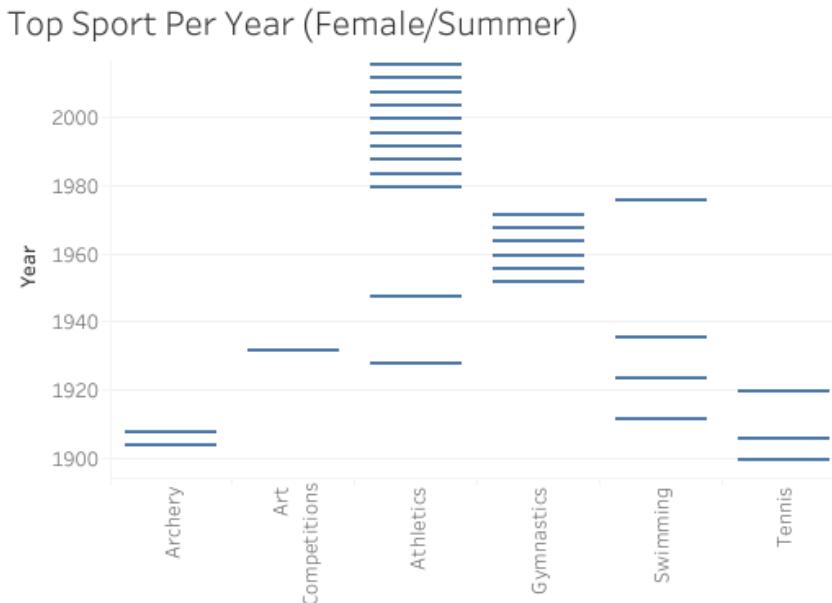
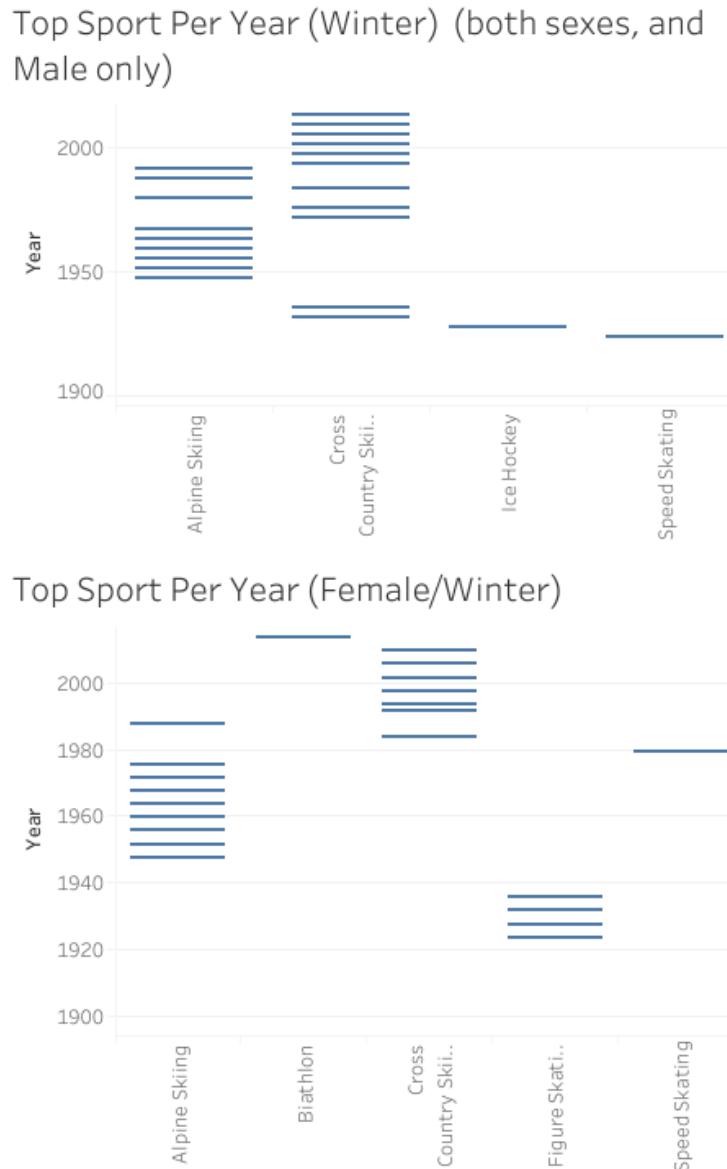
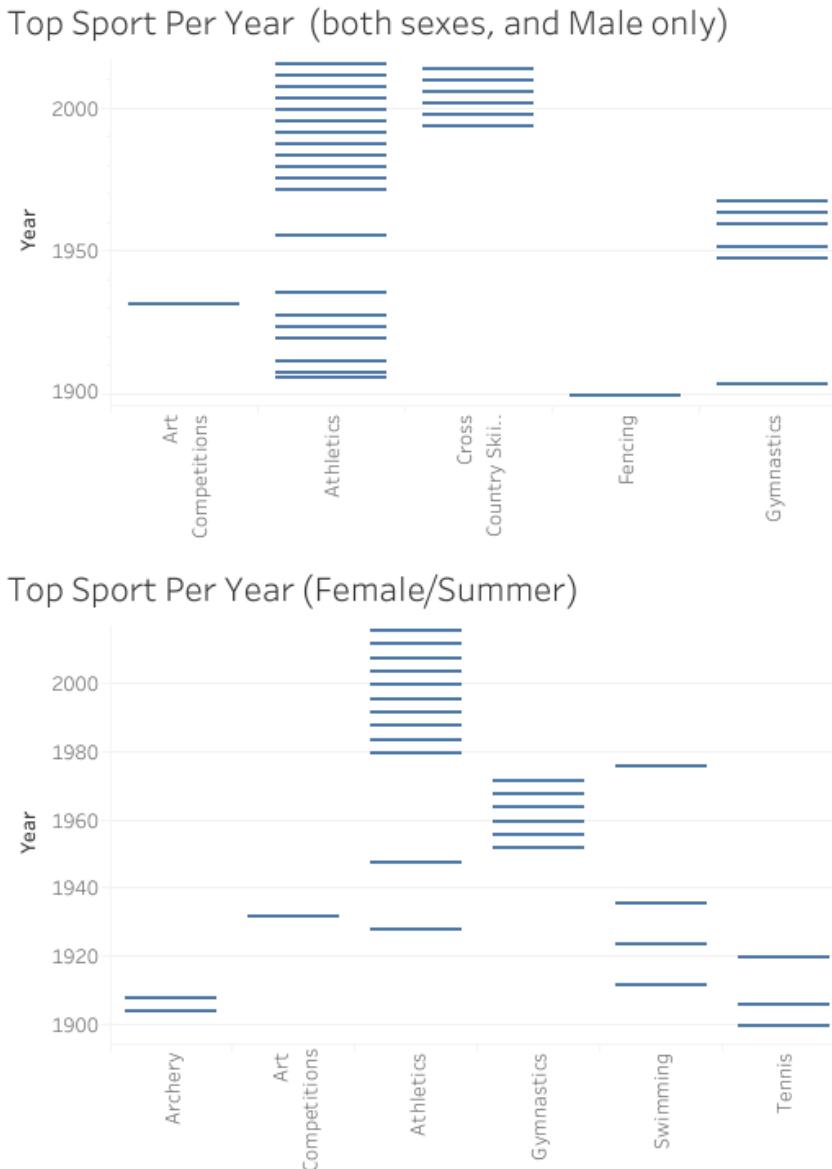


Number of Events Per sport



Top Sport of Each Year

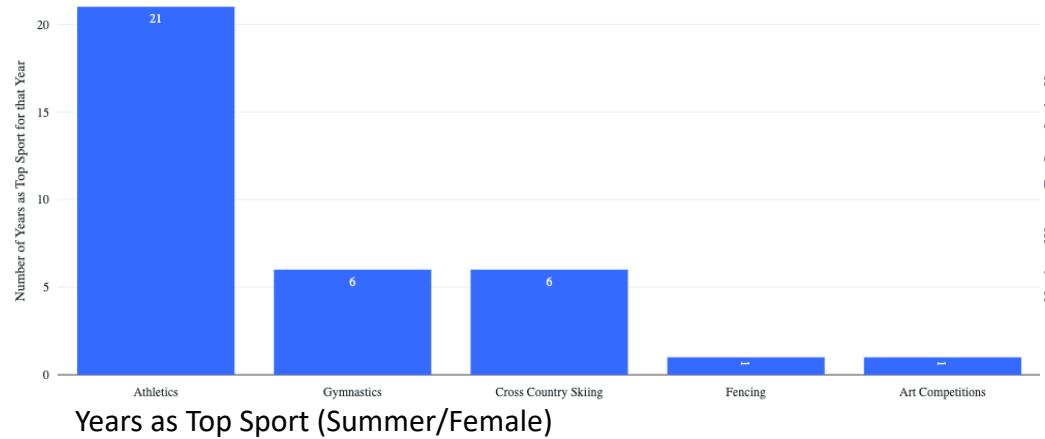
- One dash for each year that a sport was the top sport for that year.
- Shows the prevalence of certain sports over time.



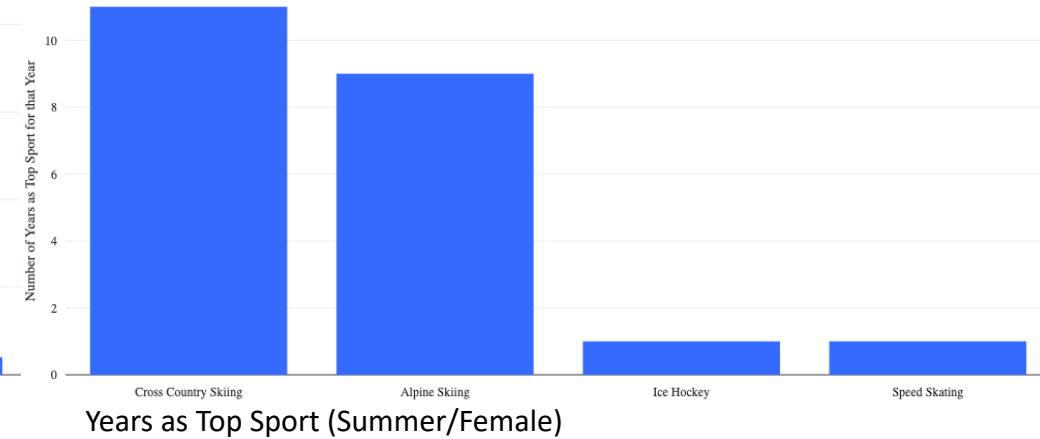
Number of Years as Top Sport

Same data as the previous slide, but instead of a sports popularity over time, it sums up the number of years that each sport was the top sport for that particular year, or season, or sex.

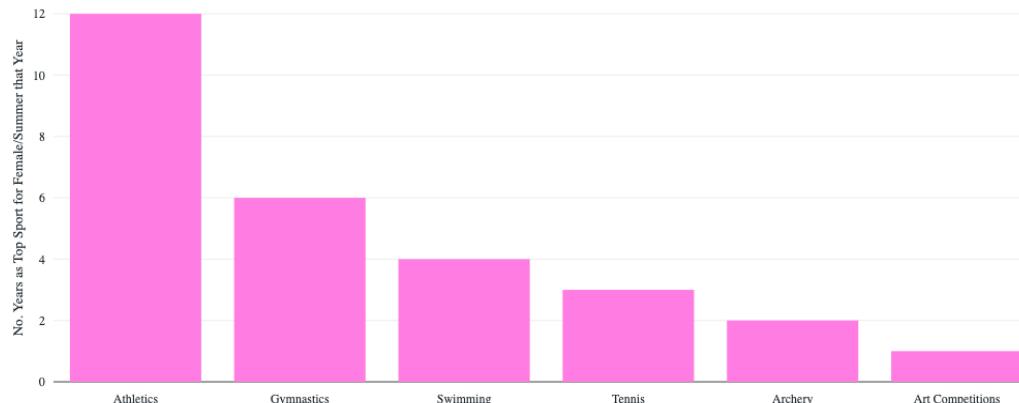
Years as Top Sport



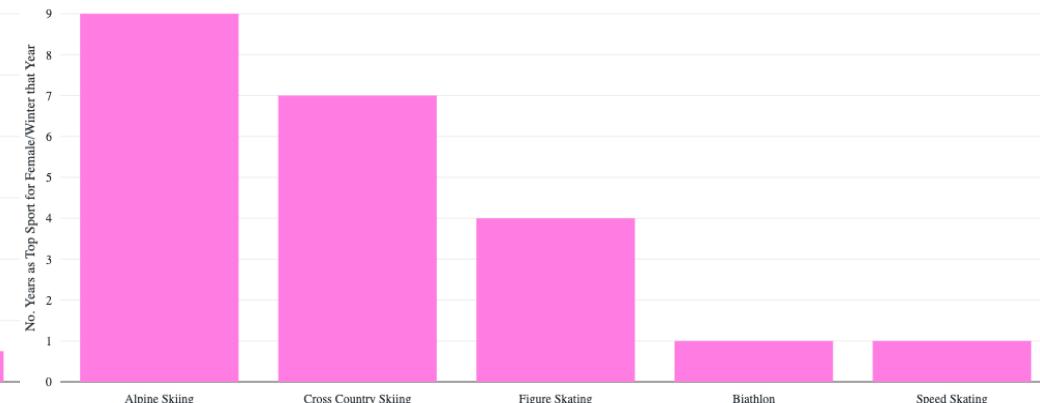
Years as Top Sport (Winter)



Years as Top Sport (Summer/Female)



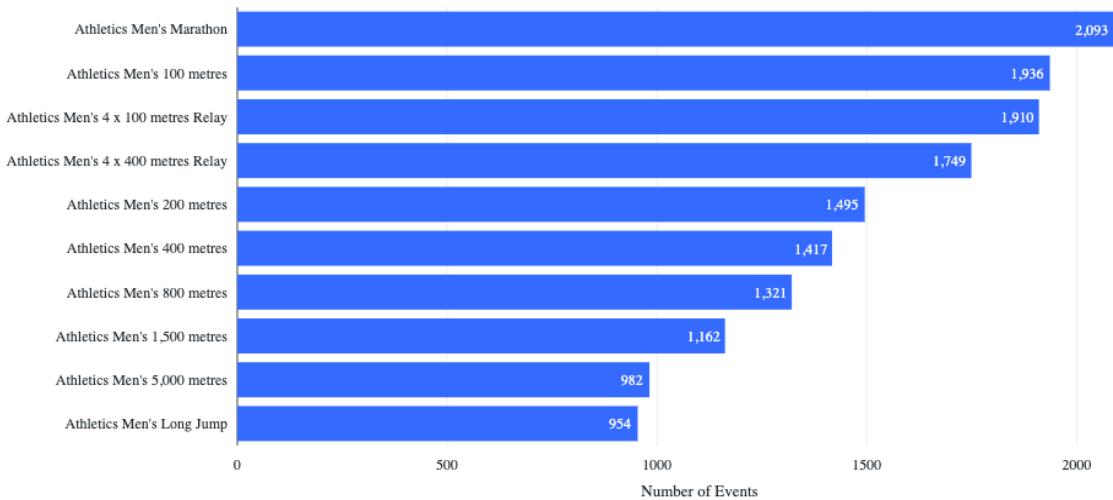
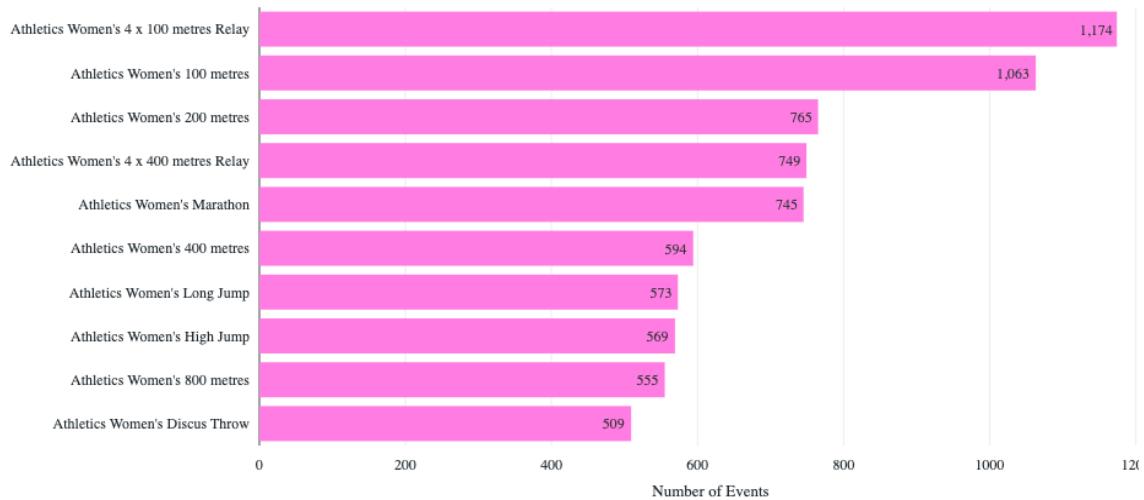
Years as Top Sport (Summer/Female)



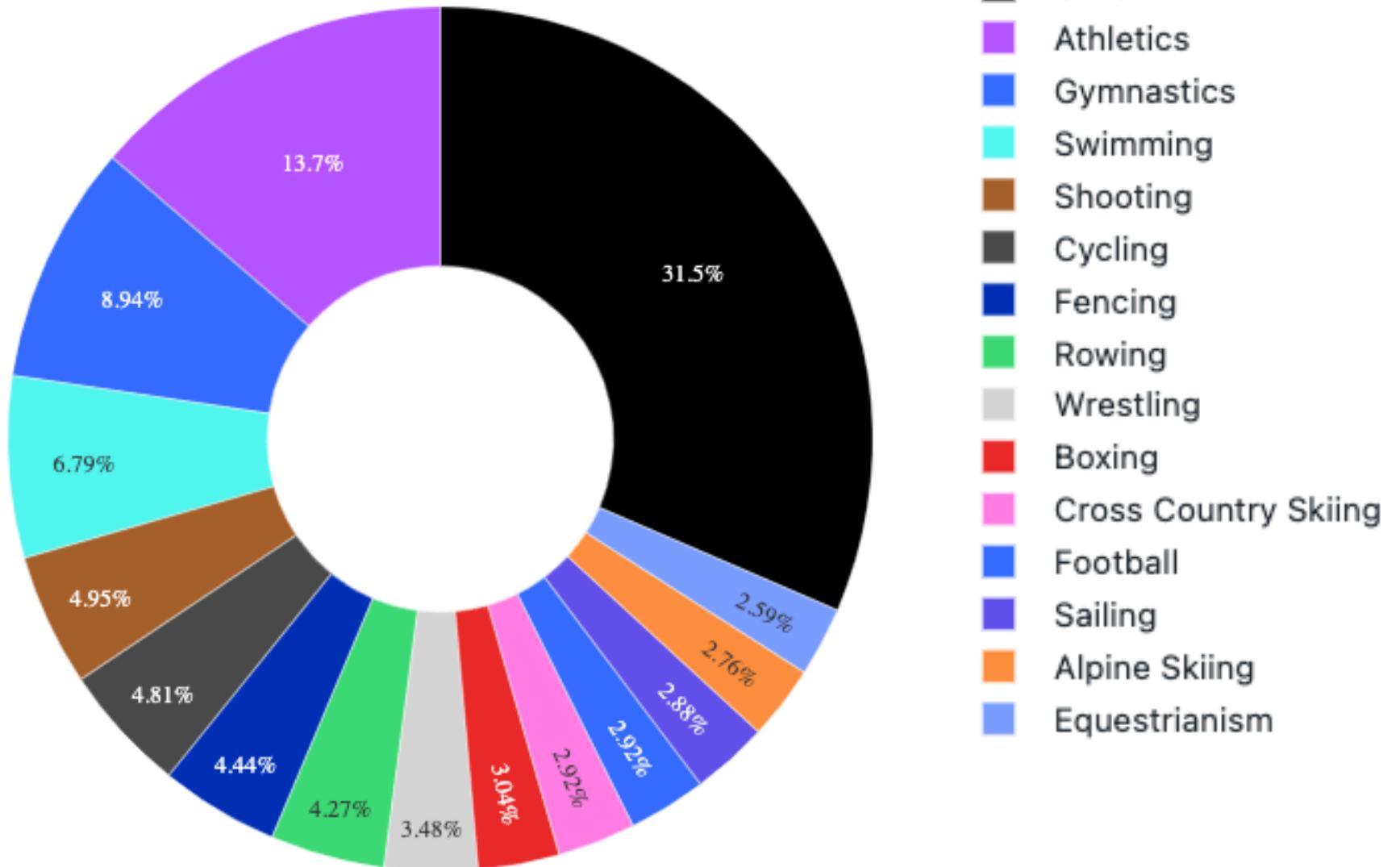
What are Athletics?

- Athletics has shown up as the top sport in the previous slides, so it warrants an explanation.
- Some events in this sport are decathlon, marathon, 100 metres hurdles, high jump, pole vault, shot put, discus throw, javelin throw, and relay.

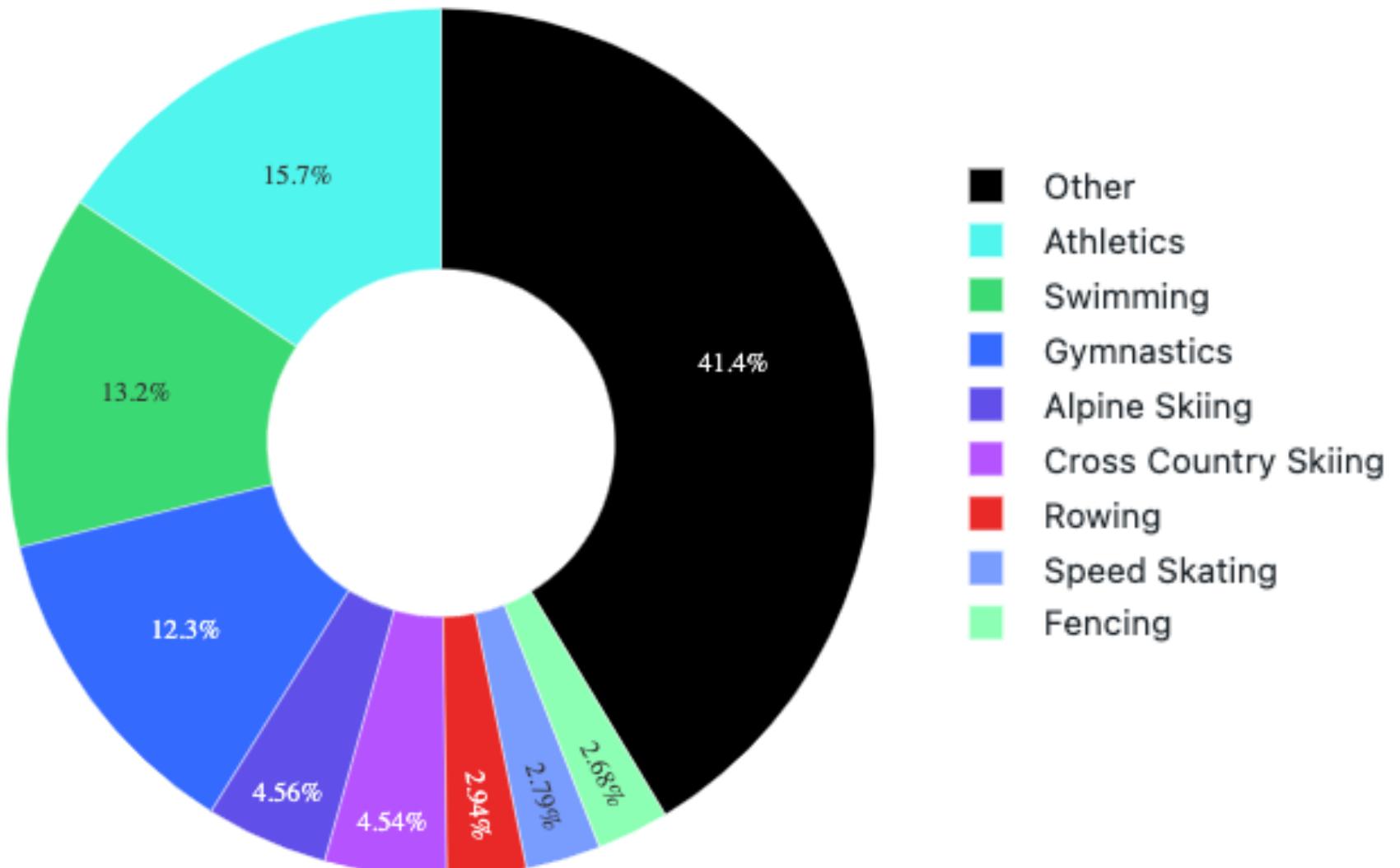
Top “Athletics” Events by Sex



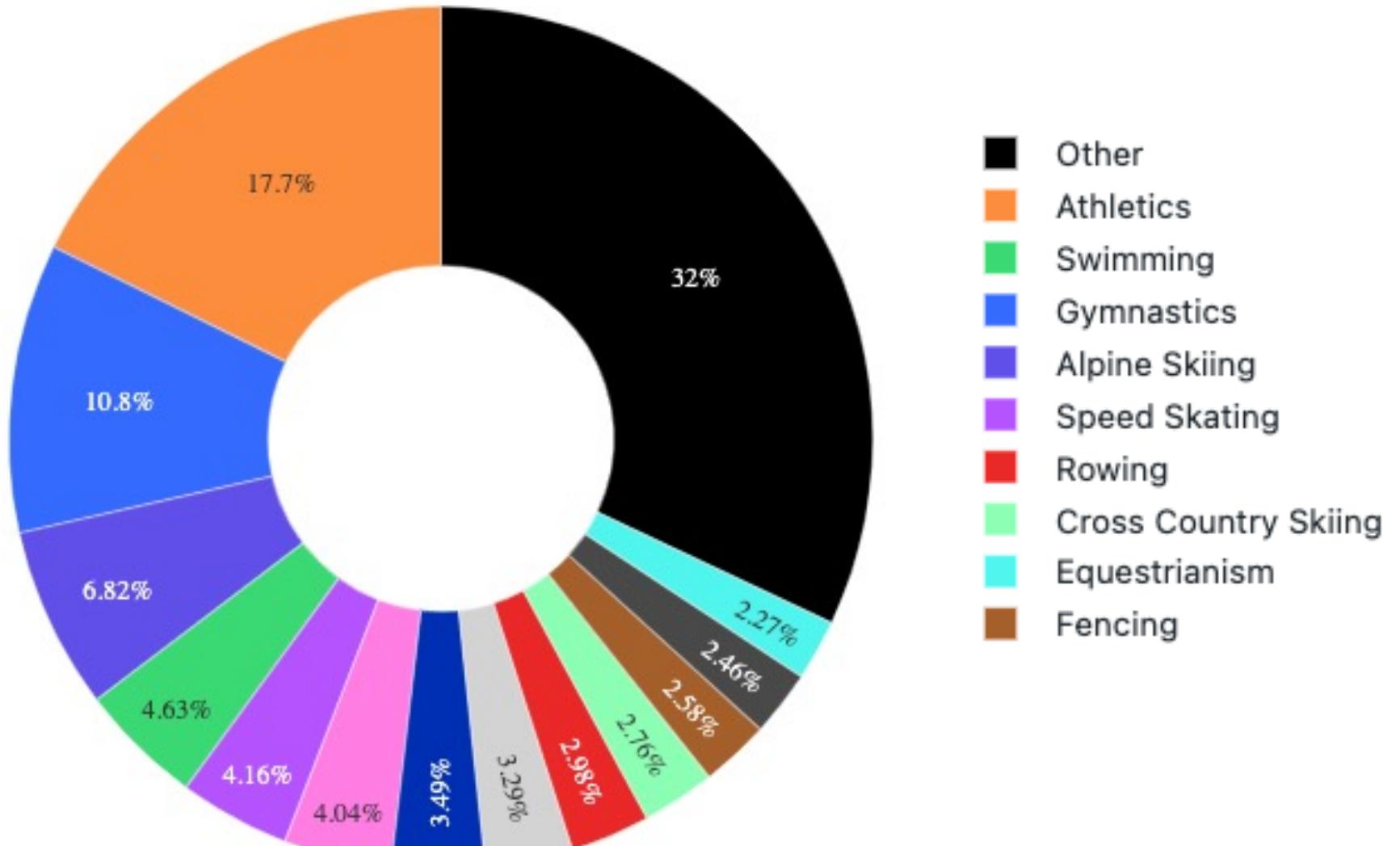
Top Sports for Males



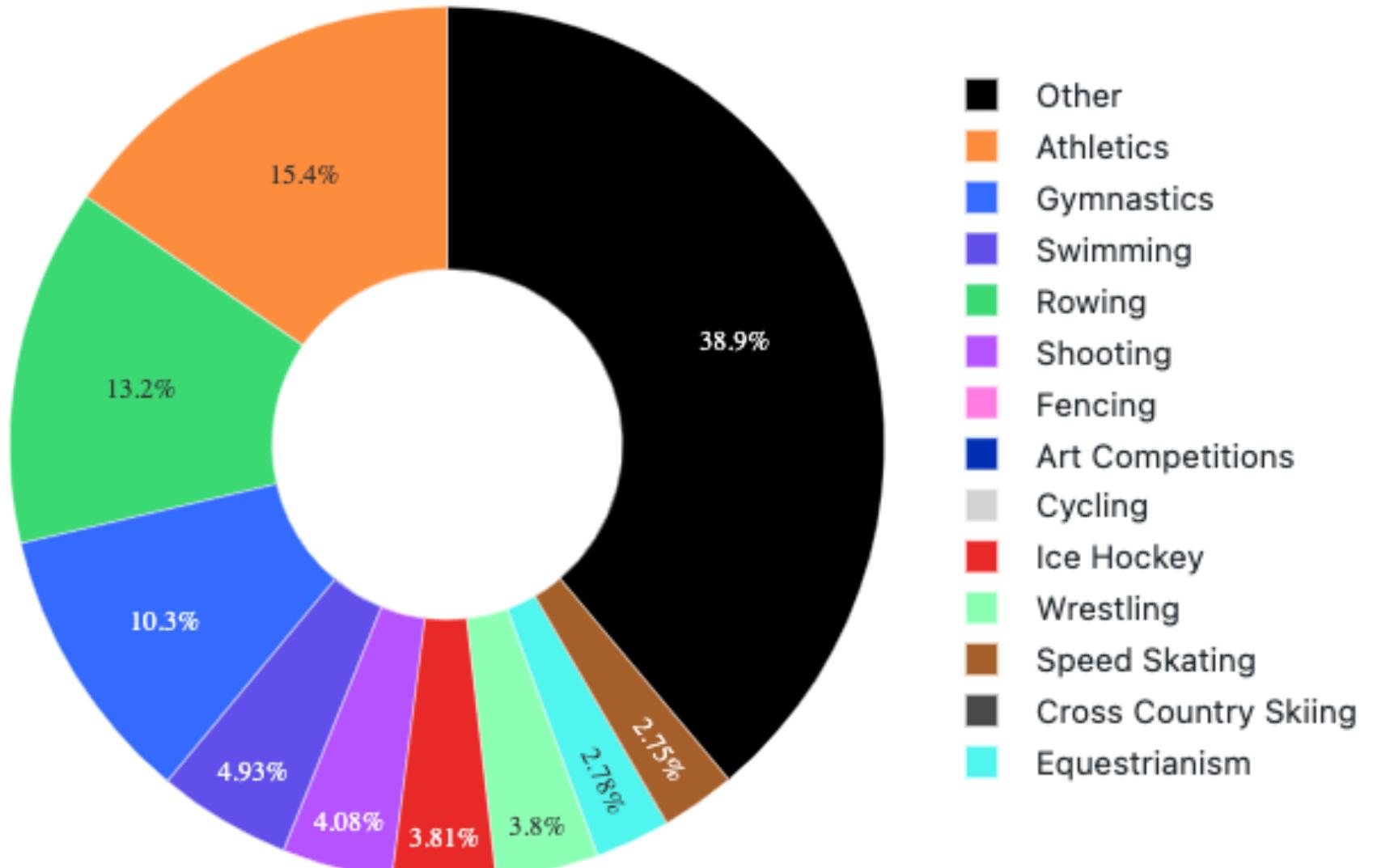
Top Sports for Females



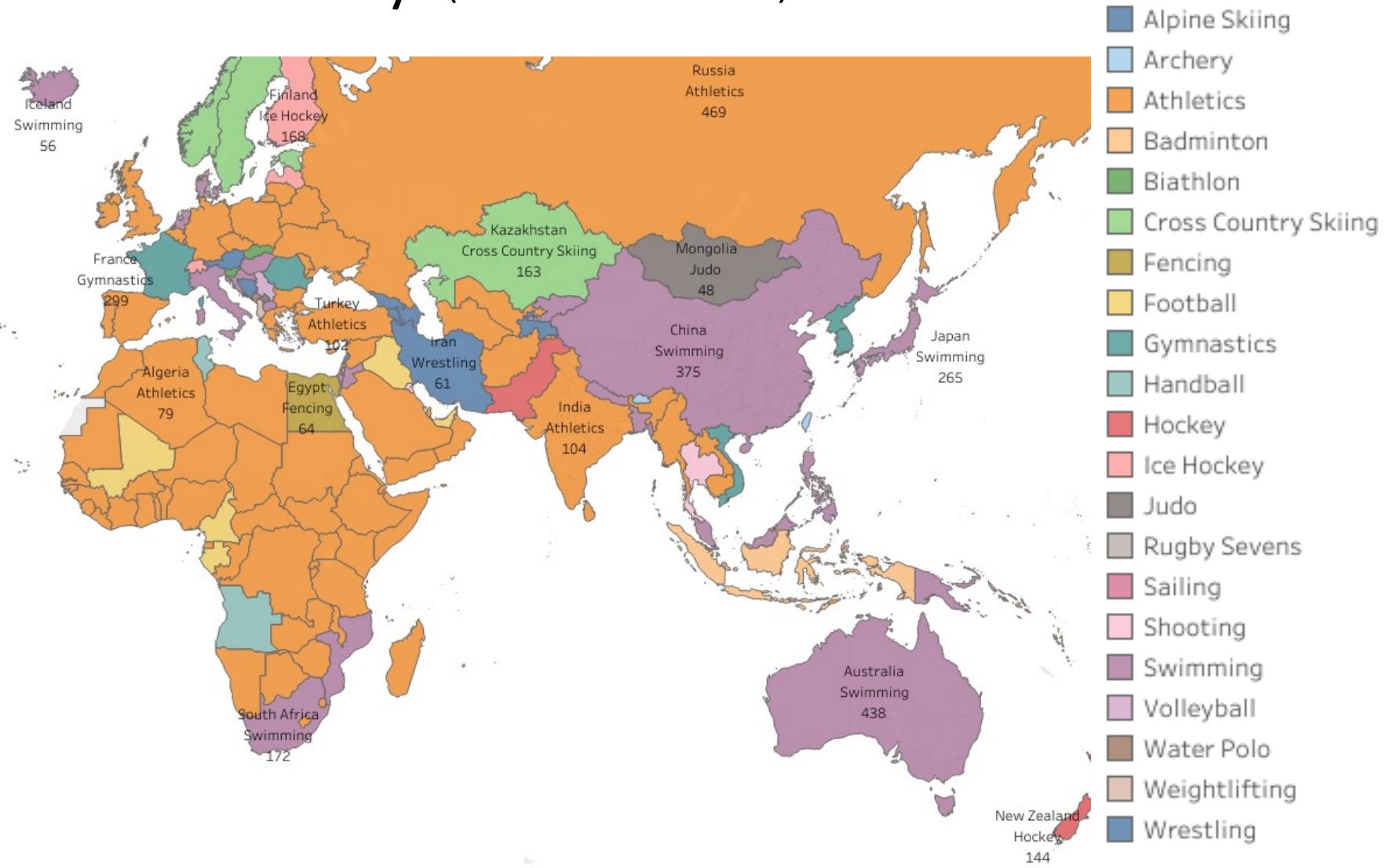
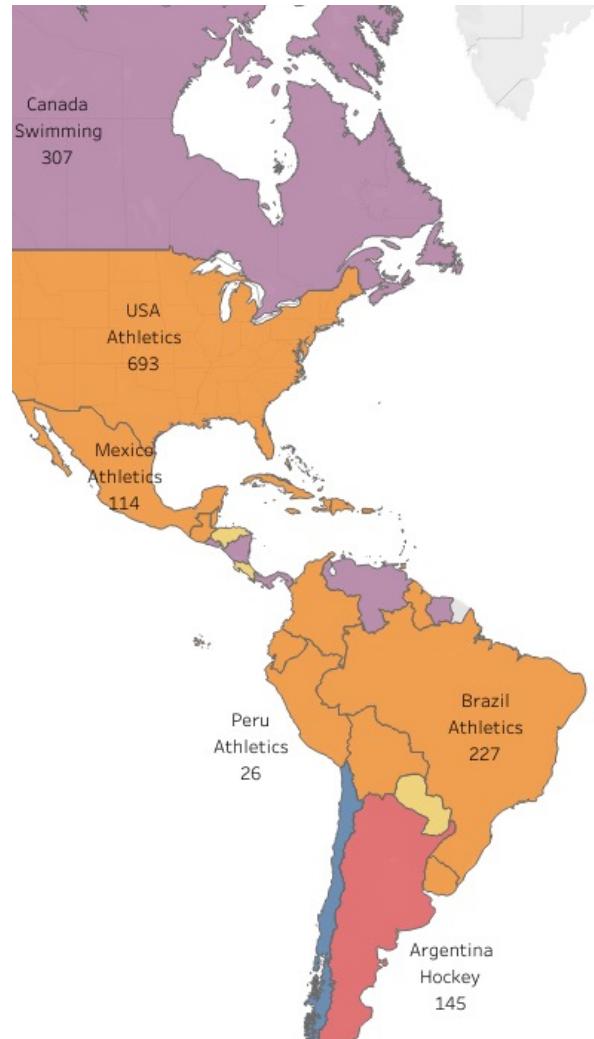
Top Sports for Males (USA only)



Top Sports for Females (USA only)



Top Sports per Country (2000-Present)

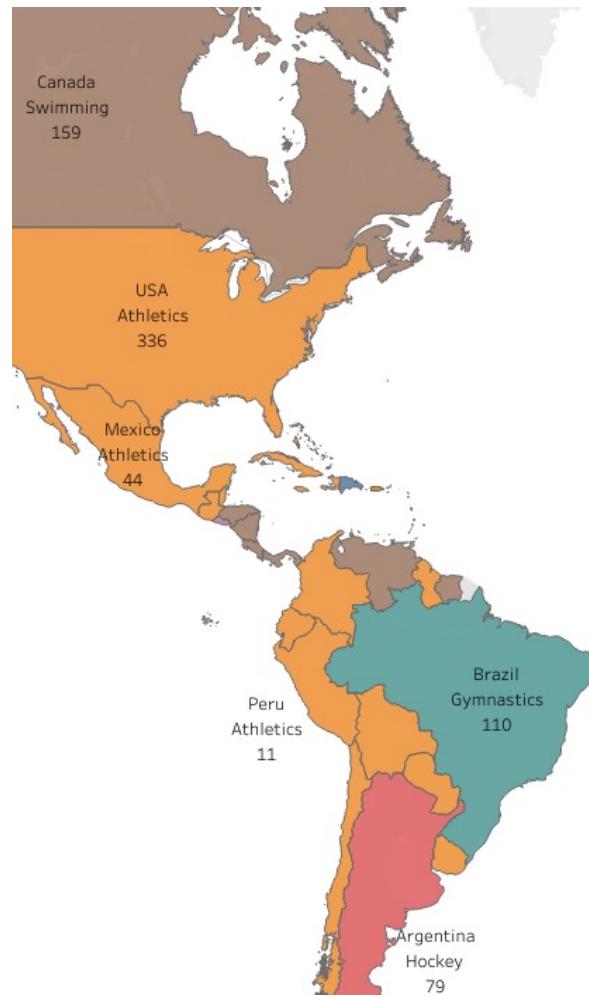


Top Sports Europe (2000-Present)



- Alpine Skiing
- Archery
- Athletics
- Badminton
- Biathlon
- Cross Country Skiing
- Fencing
- Football
- Gymnastics
- Handball
- Hockey
- Ice Hockey
- Judo
- Rugby Sevens
- Sailing
- Shooting
- Swimming
- Volleyball
- Water Polo
- Weightlifting
- Wrestling

Top Sports per Country, Female (2000-Present)

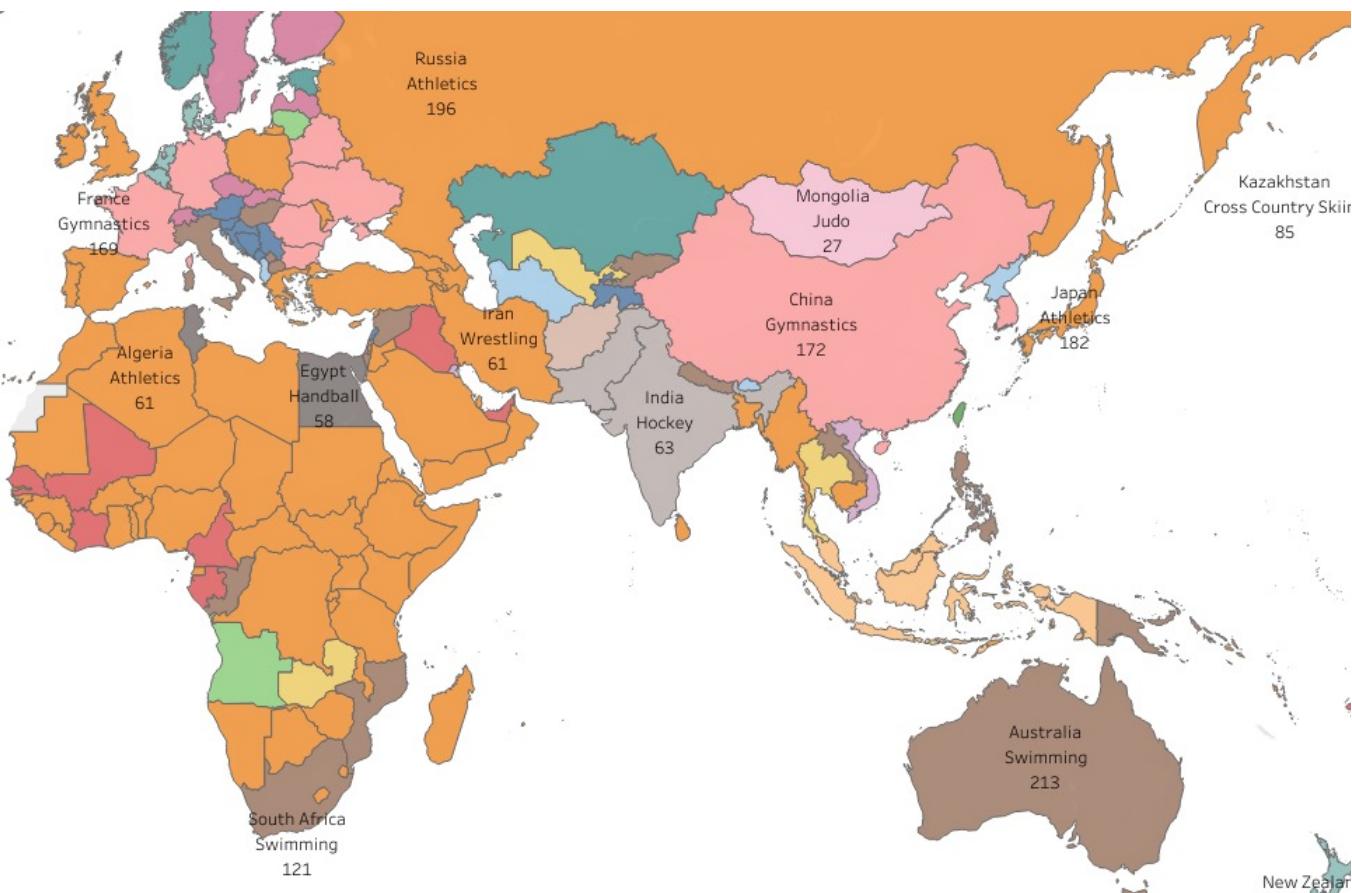


- Alpine Skiing
- Archery
- Athletics
- Badminton
- Basketball
- Biathlon
- Cross Country Skiing
- Diving
- Gymnastics
- Handball
- Hockey
- Ice Hockey
- Judo
- Rhythmic Gymnastics
- Rugby Sevens
- Sailing
- Shooting
- Softball
- Swimming
- Synchronized Swimming
- Volleyball
- Weightlifting

Top Sports Europe, Female (2000-Present)



Top Sports per Country, Male (2000-Present)



- Alpine Skiing
- Archery
- Athletics
- Badminton
- Baseball
- Basketball
- Bobsleigh
- Boxing
- Cross Country Skiing
- Cycling
- Football
- Gymnastics
- Handball
- Hockey
- Ice Hockey
- Judo
- Sailing
- Shooting
- Swimming
- Taekwondo
- Water Polo
- Weightlifting
- Wrestling

Top Sports Europe, Male (2000-Present)

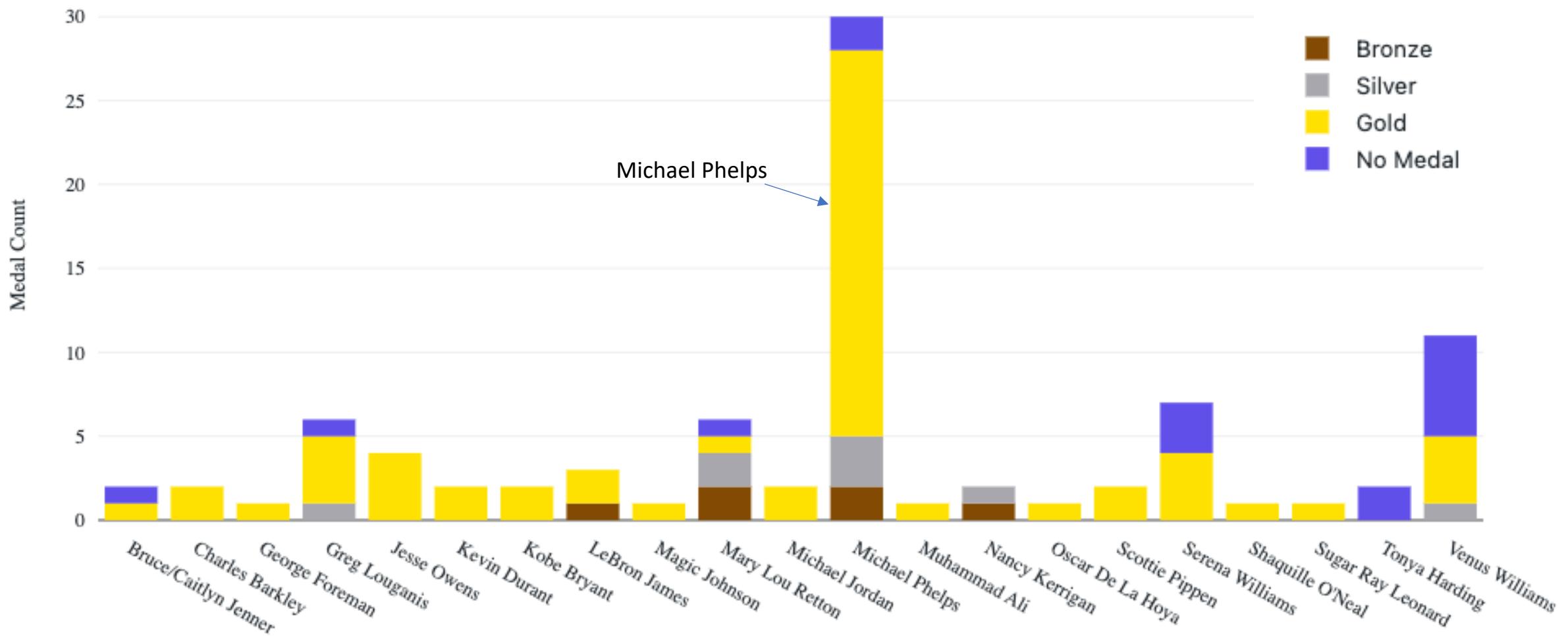


Famous Athletes from the USA

- What is the definition of “famous”? My very *un*-scientific definition is if I, personally, had ever heard of the athlete...perhaps not a useful measure at all, but I suppose many others have heard of many of these athletes.
- In addition to the famous athletes of today, there is Jesse Owens, who won 4 gold medals in the 1936 Berlin games held under Hitler, who, “as a black American man, was credited with "single-handedly crushing Hitler’s myth of Aryan supremacy.”

Famous US athletes

- You can thank Michael Phelps for making this chart hard to read. He won so many more medals than everyone else that it skews the chart.



Shout Out!

- In the USA, Olympic athletes are called “heroes” for winning medals. They often sponsor brands of athletic gear, star in commercials, or end up on the cover of a Wheaties cereal box.
- However, I’d like to give a shout out to the 4 women from Afghanistan who participated in the Olympics from 2004 to 2016. No matter what happens in that country from here on out, history will always know that these 4 women represented their country in defiance of long-standing attitudes against women.
- While none of these women won any medals, I have to say that they are the biggest heroes in this whole dataset!

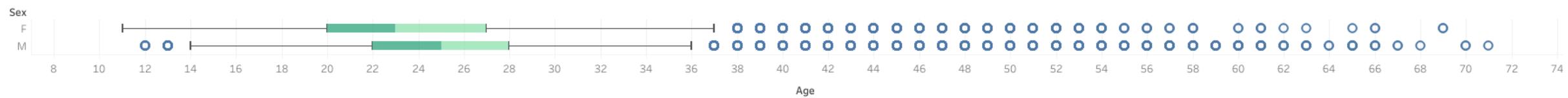
Basic Age, Height and Weight Statistics

- 271116 total records in the Athlete Events Table/Dataset

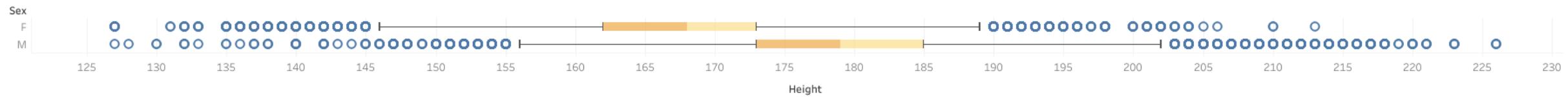
	MIN (1 st Quar)	2 nd Quartile	AVERAGE	MEDIAN (3 rd Quar)	4 th Quartile	MAX	Std Dev	NULL VALUES
Height	127	168	175.34	175	183	226	5.53	60,171
Weight	25	60	70.70	70	79	214	5.49	62,875
Age	10	21	25.56	24	28	97	6.39	9,474

Age, Height, Weight by Sex Box + Whisker Plot

Age



Height



Weight



Pearson Correlation for Athlete

- There doesn't seem to be a high correlation between an athlete's physical attributes (age, height, weight) and the amount or percent of medals won. The same goes for latitude.
- Age to Medals won: 0.071
- Height to Medals won: 0.034
- Weight to Medals won: 0.036
- Latitude to Medals won: 0.091
- Age to percent of Medals won: 0.041
- Height to percent of Medals won: -0.126
- Weight to percent of Medals won: 0.005
- Latitude to percent of Medals won: 0.096

Pearson Correlation for Country

- What is the correlation between percent of wins in the Summer Olympics to those of the Winter Olympics.
- Percent wins for Summer to Percent wins for Winter: 0.593
- Number of athletes in Summer to that in Winter: 0.630

I was surprised that the correlation is greater for Winter/Summer comparisons than the athlete's physical attributes, since there are many countries that do not participate in Winter games, and there are many fewer athletes and events for Winter.

Standard Deviation for Age, by Sex

- When separating the Standard Deviation by sex, I found:
- Male = 6.47
- Female = 5.79
- I wondered why there was such a great difference between sexes. This is due to the Art Competitions category, which had events for literature, art, and music in the early 20th Century, with some ages up to 97 years old, and threw off the values. When filtering out Art Competitions, the values are more reasonable at:
 - Male = 5.57
 - Female = 5.86

Hypothesis Results

Hypothesis		Proven?	Explanation
1	<i>More countries are participating.</i>	TRUE	12 Countries in 1986, 232 in 1991, 200-205 since.
2	<i>More women participating.</i>	TRUE	From 0% to 45% in Summer, 40% in Winter
3	<i>Summer performance correlates to Winter.</i>	TRUE	Pearson Correlation 0.593, higher than measures of height, weight, age
4	<i>Taller athletes win more.</i>	FALSE	More medals were won, and percentage of metals over events, by athletes that were toward the median of the dataset. The correlation to medals won (and %) was very low for height, weight, and age.
5	<i>Northern countries are more successful in Winter sports.</i>	TRUE	Very true! Scandinavian countries have higher percent wins (medals to events) than Russia and the USA, even though they are more populous and have more Winter medals overall. This is also true of Summer sports, less pronounced.

Conclusion

- This analysis has shown us that geography and country are the greatest predictors of whether an athlete will win a medal, even over physical characteristics such as Age, Height, and Weight.
- Height is increasing across the board, regardless of country or sex.
- Northern countries lead the Olympics, in medals, events participated in, and hosting the games.
- Over time, more countries are participating, and the percentage of woman to men is moving toward 50%.

Recommendations and Actions

The next step for my client is to use this data to focus on changing trends:

- 1) Increase in female participation
- 2) Changing popularity of some sports, while others fall out of favor.
- 3) Concentrate all Winter sports marketing toward Northern European countries, along with Russia and the United States.
- 4) Focus on which sports are popular per country, to specialize marketing in that country.
- 5) Focus on famous athletes in countries that have a lot of participation, such as the United States.