

# Olympics Data Analysis

Zihang WANG  
Duytan Le  
Liyang FANG



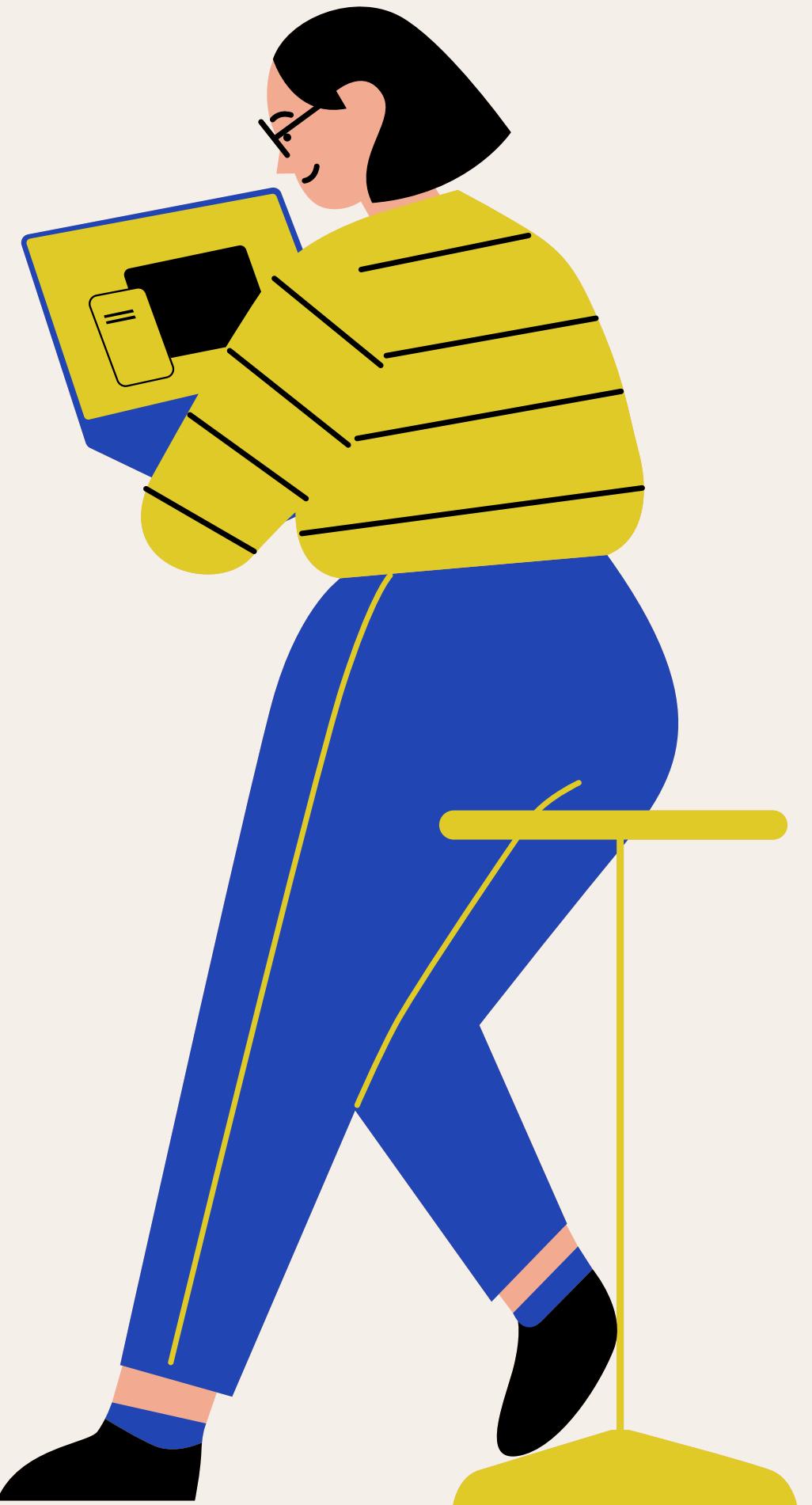
# O1 - Introduction of Dataset

# O2 - Pre-processing

# O3 - Data Visualization

# O4 - Analysis and Conclusion

Visualization



# 01 - Introduction of Dataset

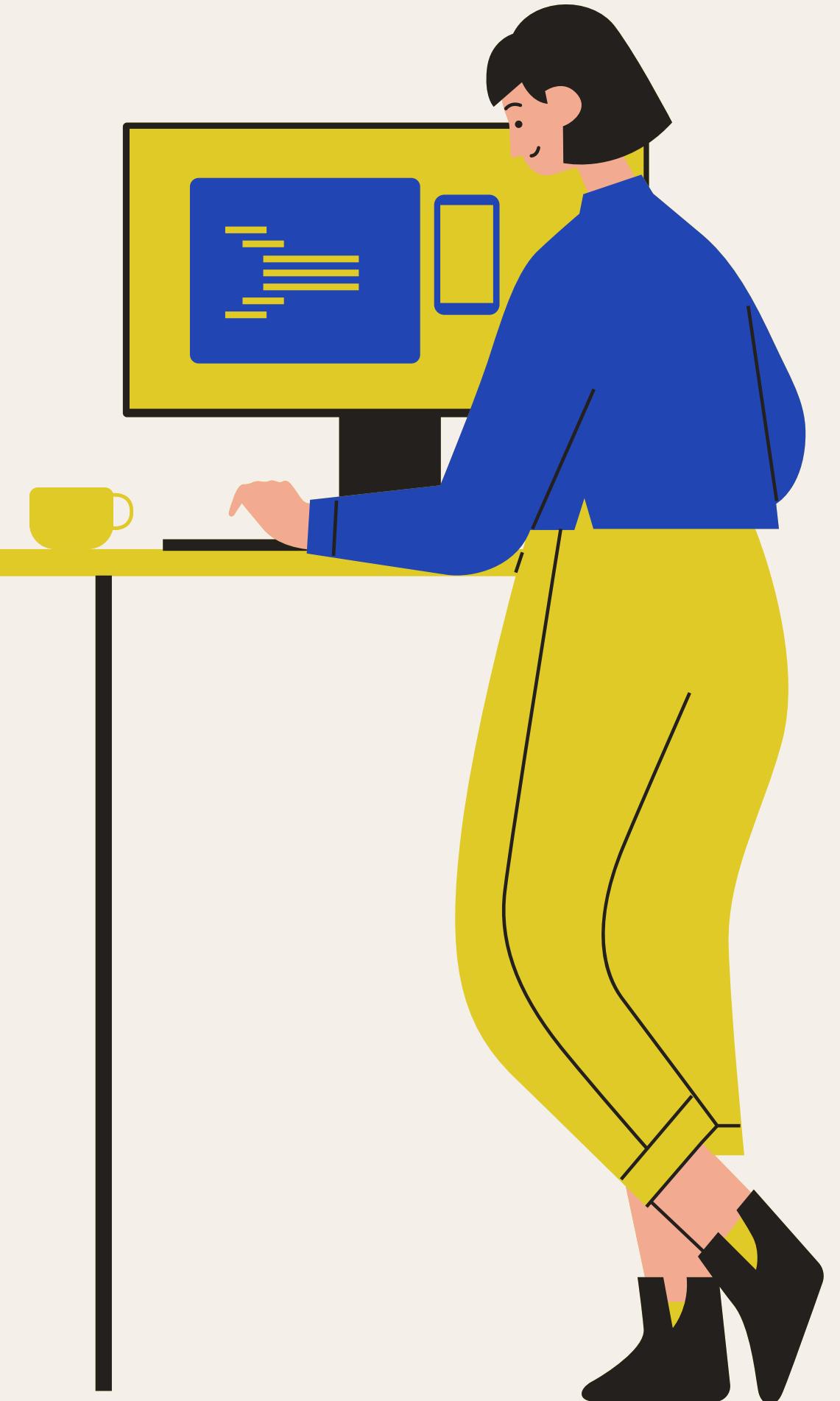
Athlete and Regions	Tokyo athlete	Medal
ID - feature	URL	Athlete Link
Name	Name	Medal Type
Sex	Height	Medal Code
Age	Birth	Medal Date
Height	Birth Place	Athlete Name
Weight	Country	Athlete Sex
Country	Residence	Country Code
Medal	Discipline	Discipline Code
		Event

Data

Visualization

# O2 - Pre-processing

1. Merge datasets
2. Selection of data (eg. regions, countries, sex)
3. Drop out the repeat data
4. Replace null data
5. Insert or delete to correct the error data



# 02 - Pre-processing

```
ID          0  
Name        0  
Sex          0  
Age         9474  
Height      60171  
Weight      62875  
Team          0  
NOC          0  
Games         0  
Year          0  
Season        0  
City          0  
Sport         0  
Event         0  
Medal      231333  
region        0  
notes      266077  
dtype: int64
```

	NOC	Team
578	SGP	Singapore
6267	ROT	Refugee Olympic Athletes
44376	SGP	June Climene
61080	UNK	Unknown
64674	TUV	Tuvalu
80986	SGP	Rika II
108582	SGP	Singapore-2
235895	SGP	Singapore-1

	NOC	region	notes
0	AFG	Afghanistan	NaN
1	AHO	Curacao	Netherlands Antilles
2	ALB	Albania	NaN
3	ALG	Algeria	NaN
4	AND	Andorra	NaN

# 02 - Pre-processing

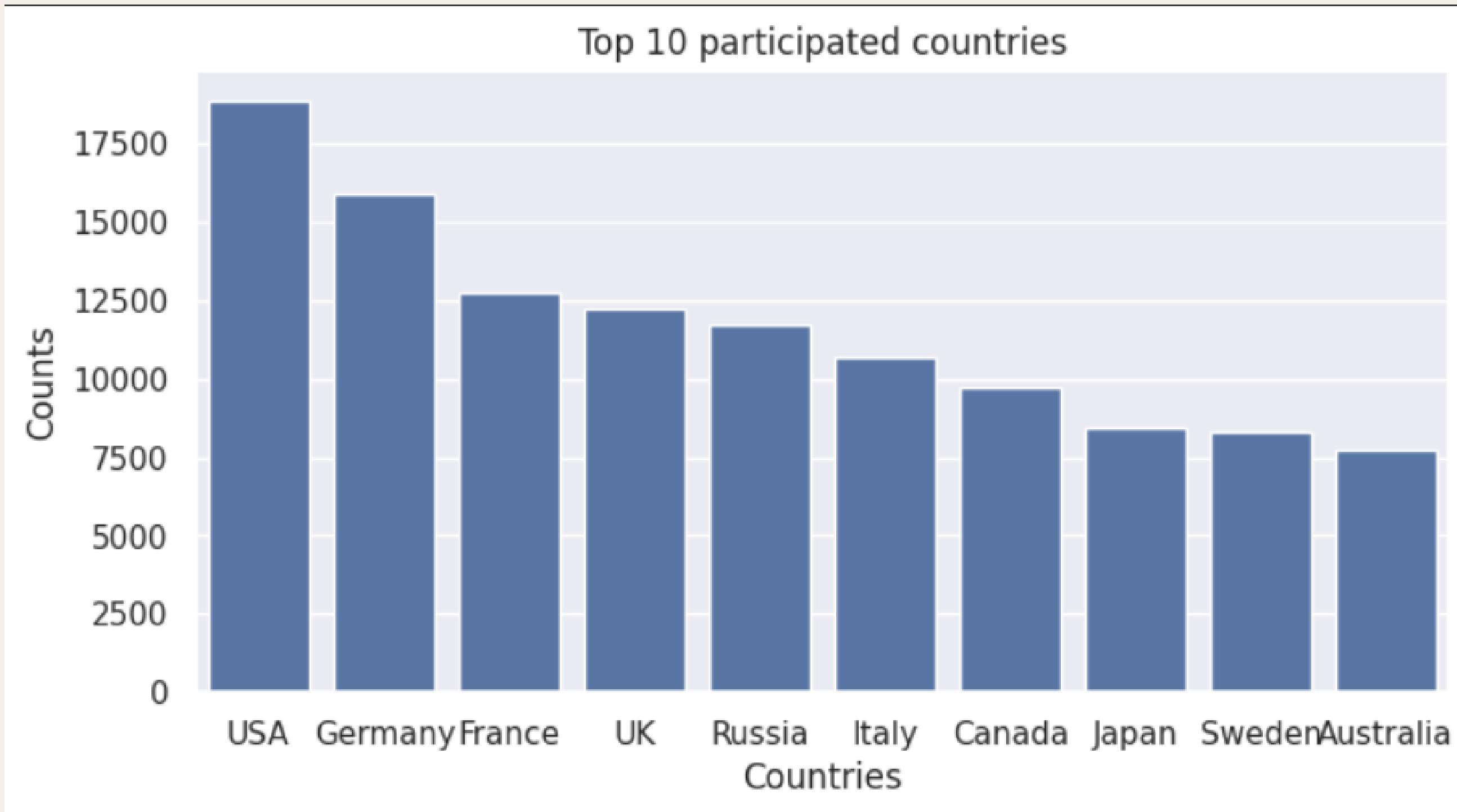
2]:

	Name	Sex	Year	Season	Sport	Medal	Country
0	KIM Je Deok	Male	2020	Summer	Archery	Gold Medal	South Korea
1	AN San	Female	2020	Summer	Archery	Gold Medal	South Korea
2	SCHLOESSER Gabriela	Female	2020	Summer	Archery	Silver Medal	Netherlands
3	WIJLER Steve	Male	2020	Summer	Archery	Silver Medal	Netherlands
4	ALVAREZ Luis	Male	2020	Summer	Archery	Bronze Medal	Mexico
...	...	...	...	...	...	...	...
2403	HARAI Balazs	Male	2020	Summer	Water Polo	Bronze Medal	Hungary
2404	HOSNYANSZKY Norbert	Male	2020	Summer	Water Polo	Bronze Medal	Hungary
2405	JANSIK Szilard	Male	2020	Summer	Water Polo	Bronze Medal	Hungary
2406	MANHERCZ Krisztian	Male	2020	Summer	Water Polo	Bronze Medal	Hungary
2407	MEZEI Tamas	Male	2020	Summer	Water Polo	Bronze Medal	Hungary

2408 rows × 7 columns

# 03 - Data Visualization

## Main Participates Countries

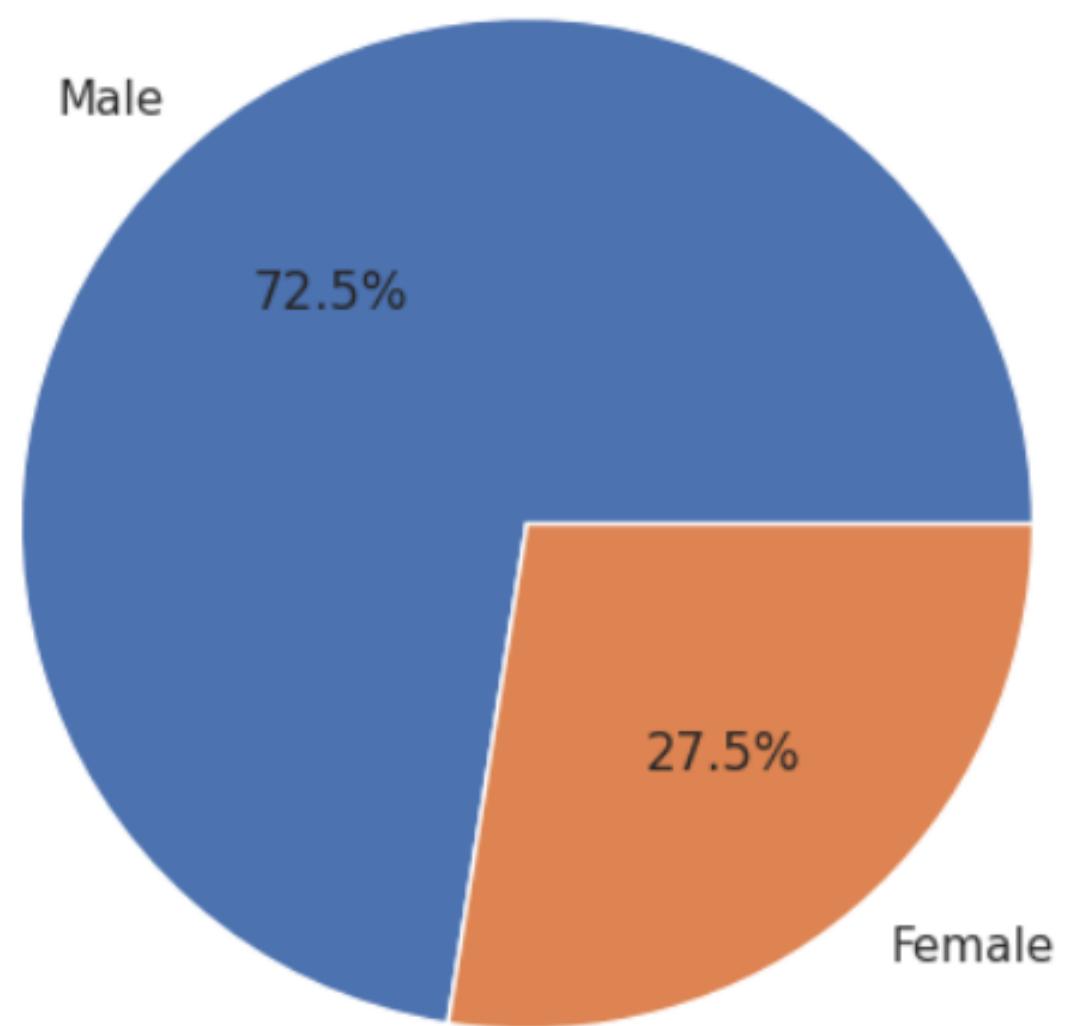


1. Mostly comes from developed countries
2. USA has a significant gap with others

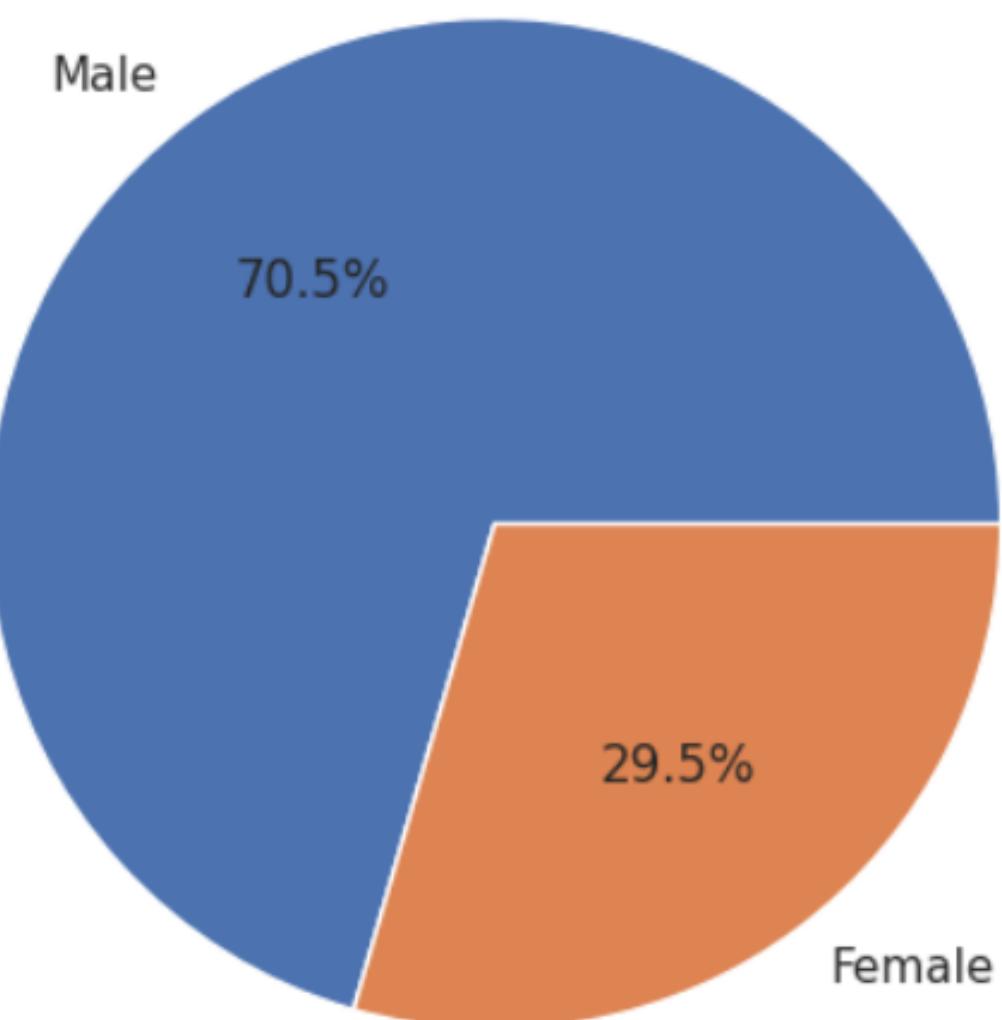
# 03 - Data Visualization

## Gender Analysis

Gender Distribution of participation in 120 years



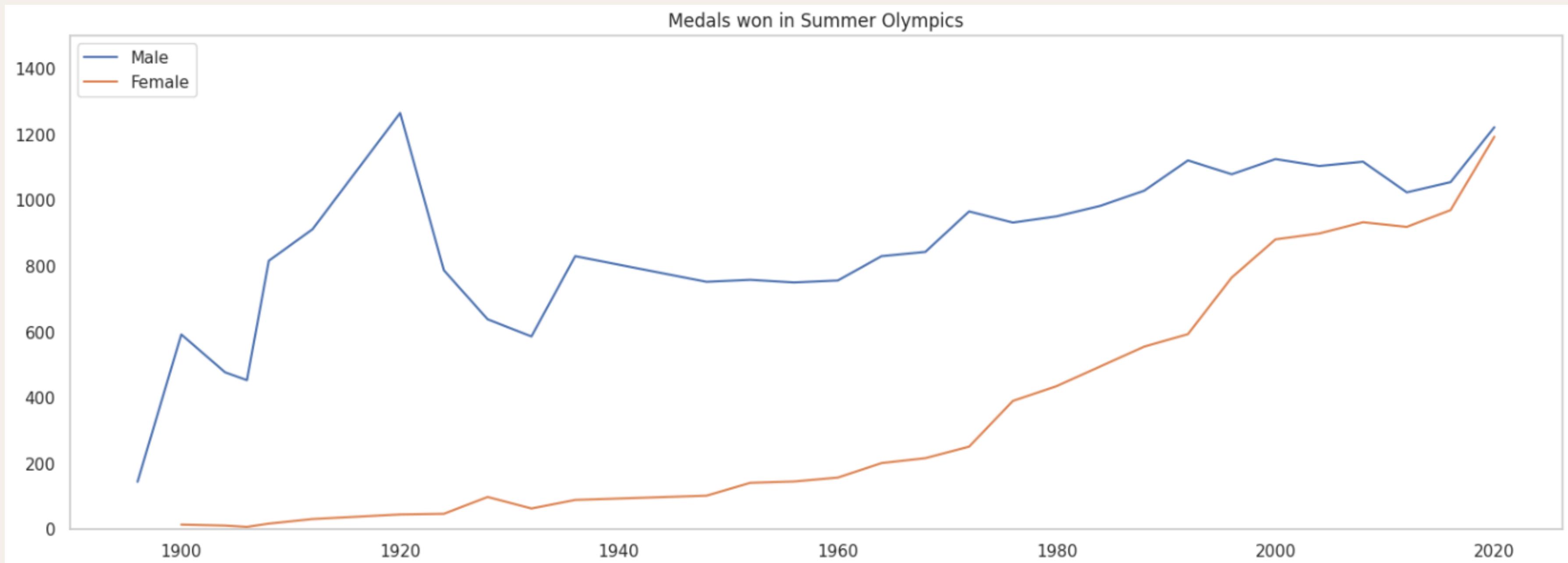
Gender Distribution of medals



1. Similar percentage between genders and medals' gender distribution
2. two percent higher than male

# 03 - Data Visualization

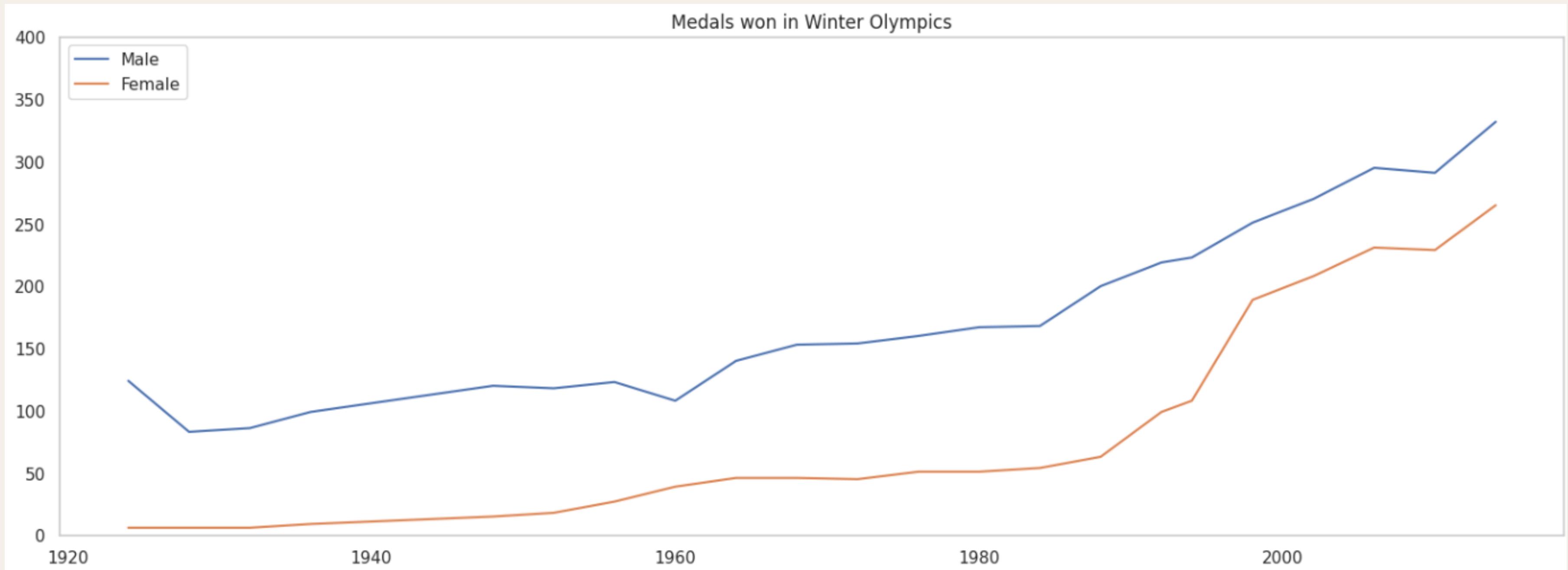
## Medal Statistics



Initially males were much higher than females, the number of medals won by females increased obviously since 1980

# 03 - Data Visualization

## Medal Statistics

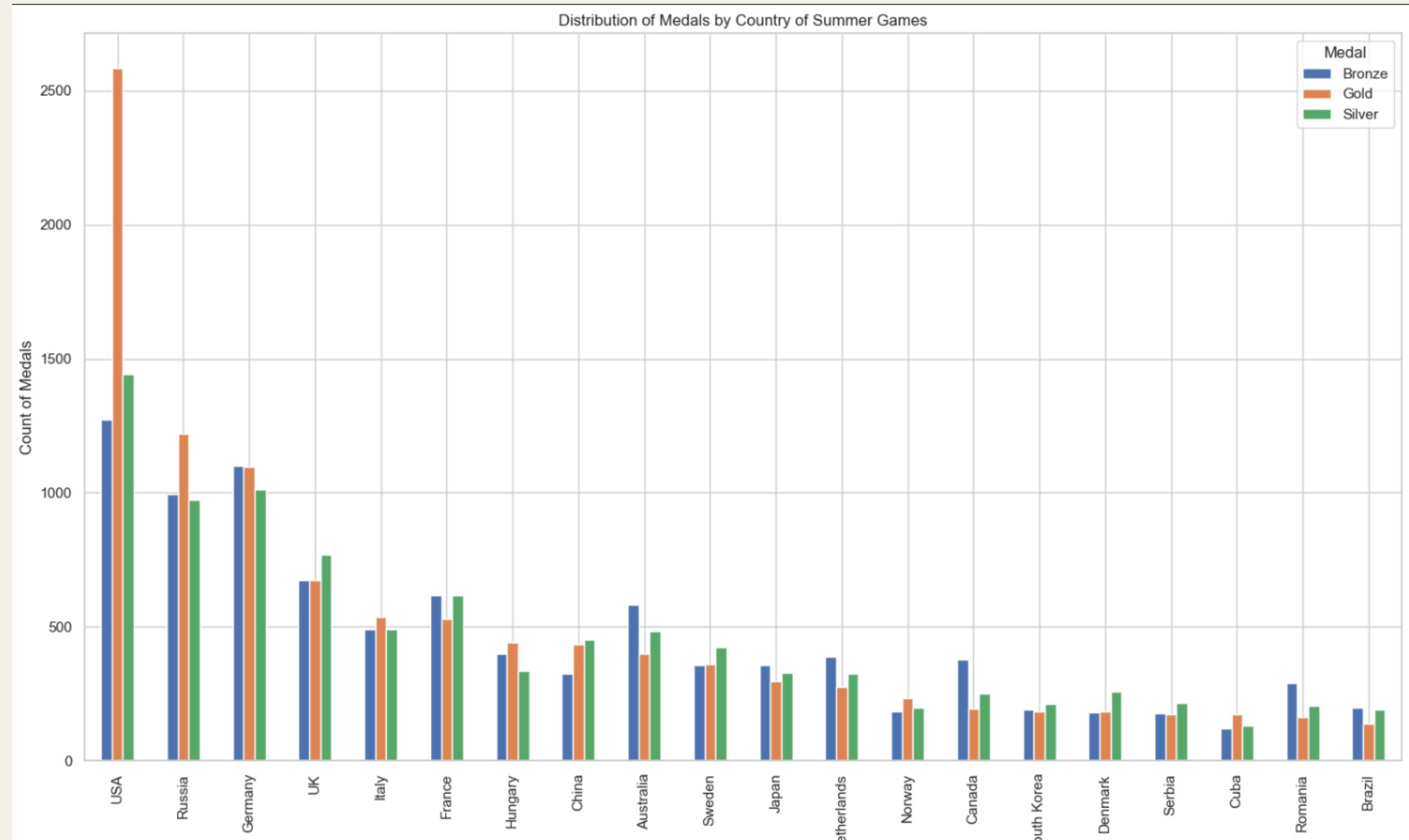


Different from summer, on winter there were kindly balance between male and female but still keep stable increasing since 2000 for female

# 03 - Data Visualization

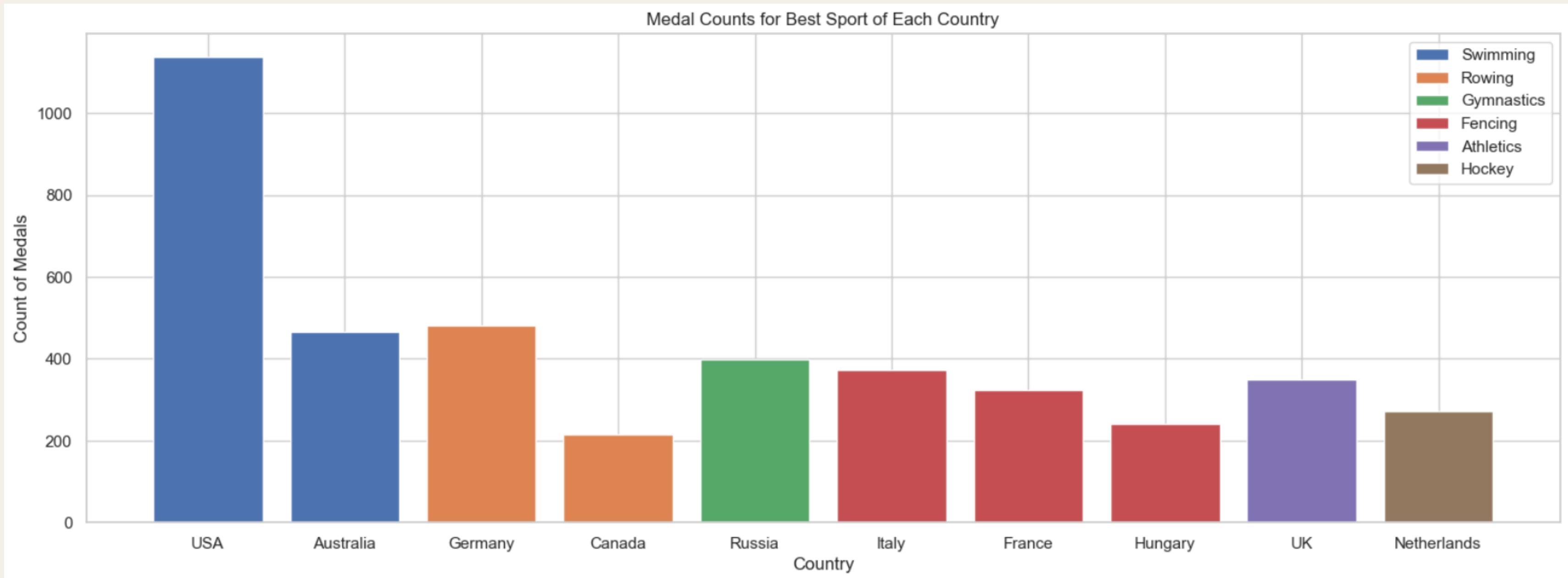
## Medal Statistics

1. USA dominates the medals
2. No much differences of medals between the last countries



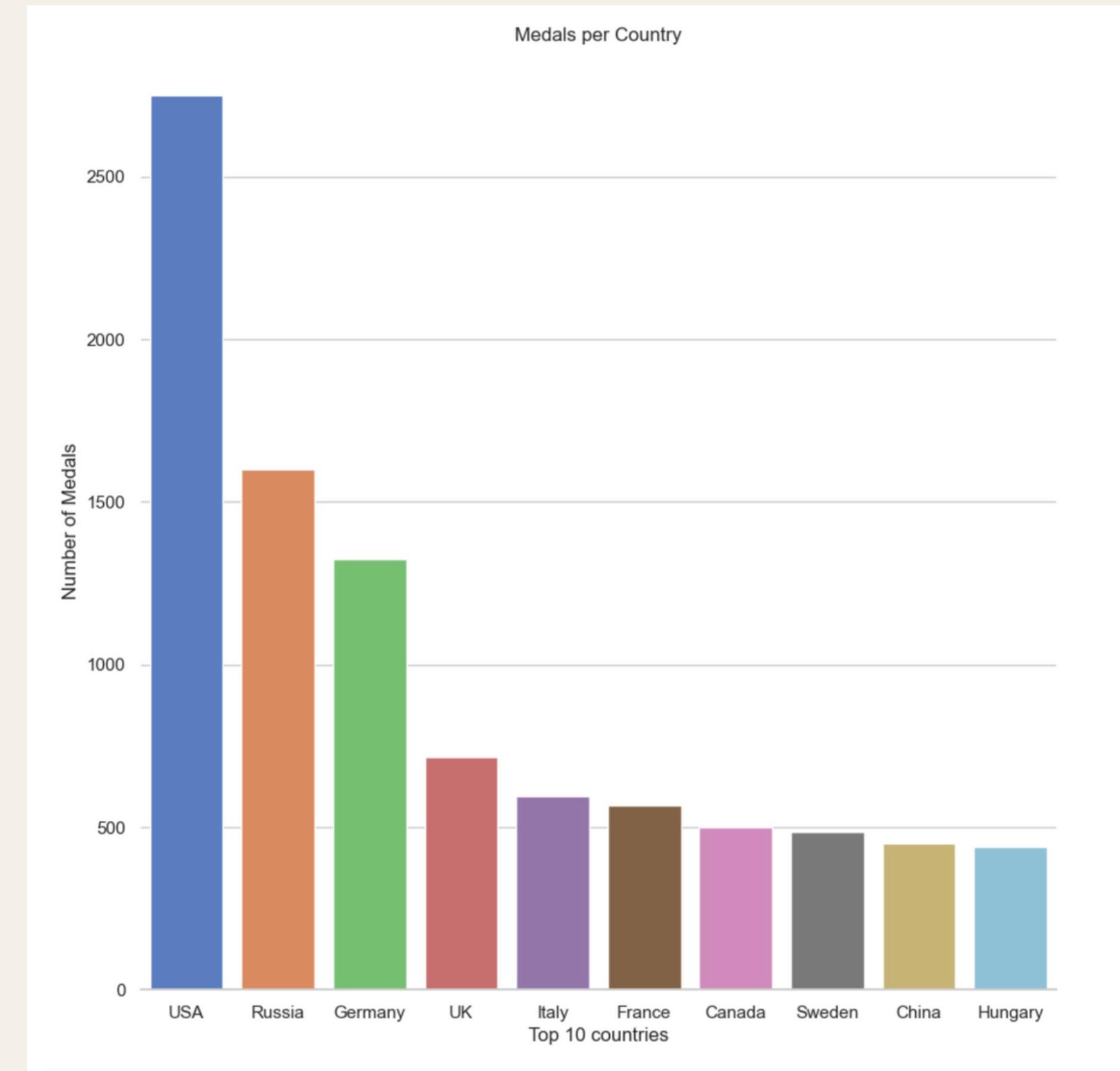
# 03 - Data Visualization

## Medal Statistics



# 03 - Data Visualization

## Medals by Country

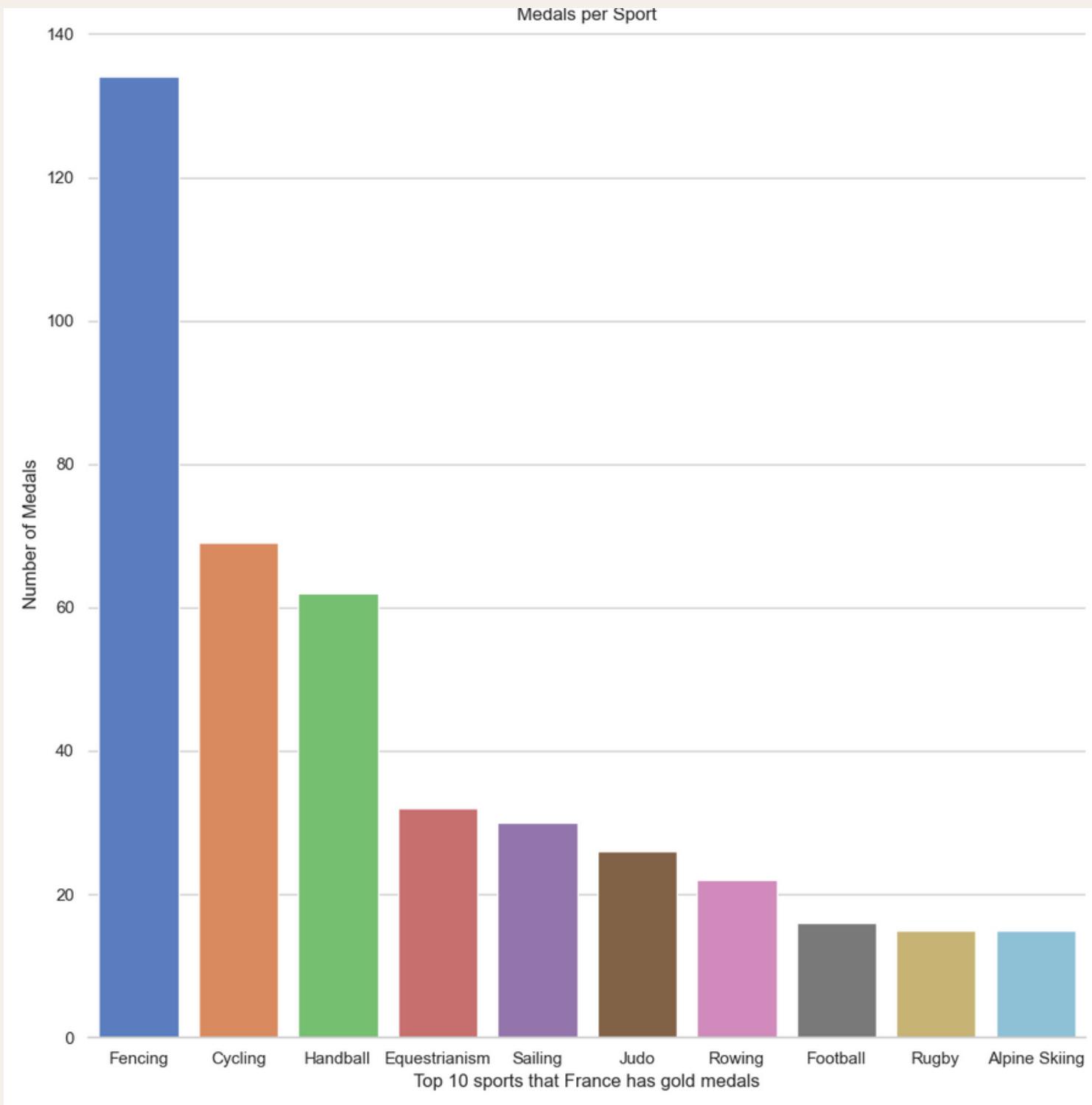


# 04 - Analysis & Conclusions

- U.S. **Dominance** 🇺🇸: With 2750 medals, the USA demonstrates a significant lead, indicating robust sports infrastructure, investment, and a large pool of athletes.
- Russia's **Strong Performance** 🇷🇺: Russia's 1599 medals reflect its historical strength in the Olympics, influenced by state-sponsored sports programs, particularly during the Soviet era.
- Germany's **Consistency** 🇩🇪: Germany's tally of 1322 medals highlights consistent Olympic performance, including the contributions of both East and West Germany before reunification.
- UK and Italy's **Contribution** 🇬🇧 🇮🇹: The UK (717 medals) and Italy (594 medals) show substantial Olympic achievements, potentially linked to strong traditions in specific sports.

# 03 - Data Visualization

Sports that France dominate 

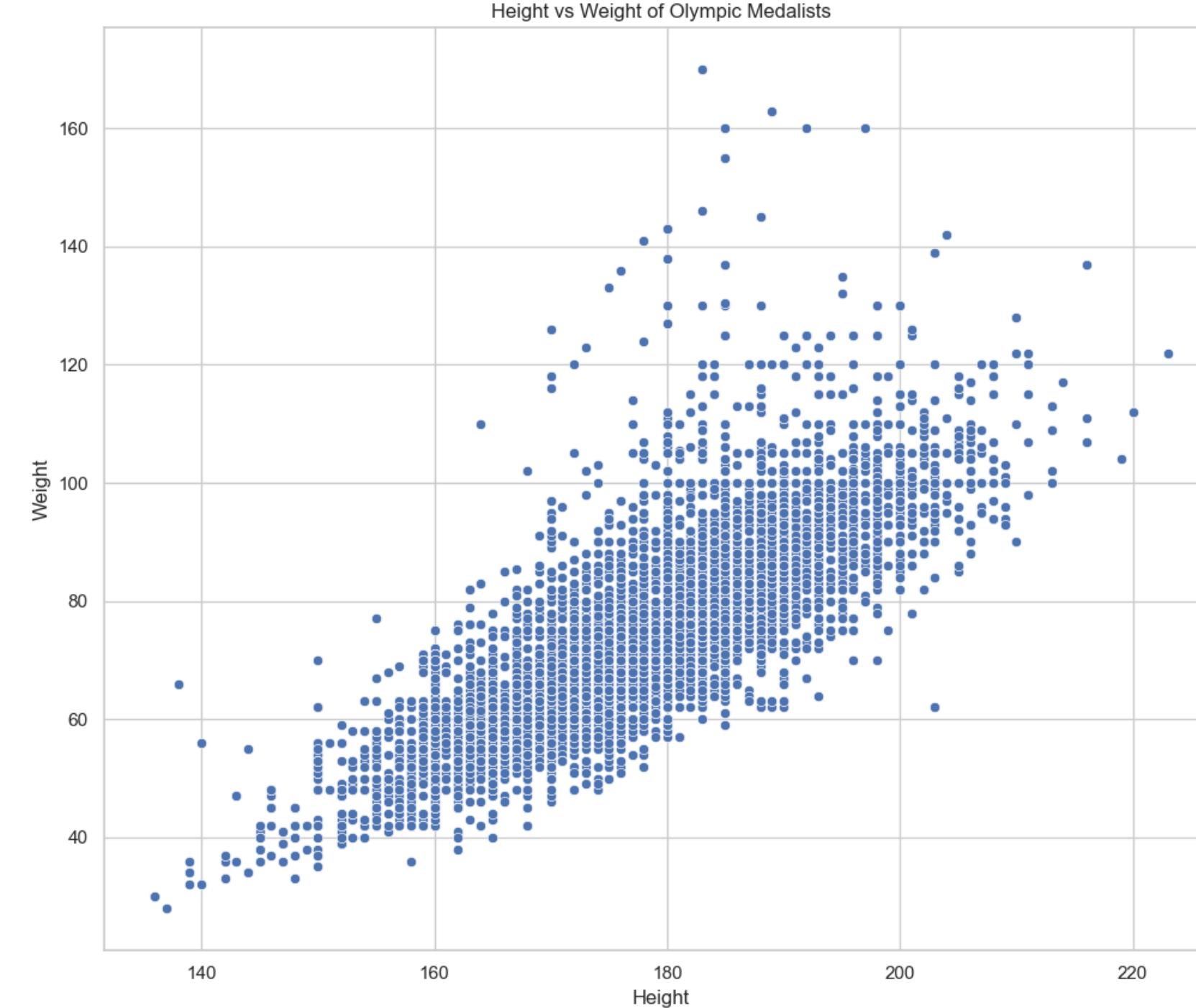


# 04 - Analysis & Conclusions

1. Fencing (134 Golds) : Dominant performance, reflecting a strong historical tradition and expertise in fencing.
2. Cycling (69 Golds) : Significant success, likely influenced by France's rich cycling culture and events like the Tour de France.
3. Handball (62 Golds) : High achievement, indicating effective national training programs and the popularity of the sport.
4. Equestrianism (32 Golds) : Strong performance, possibly due to excellent breeding, training facilities, and investment in the sport.
5. Sailing (30 Golds) : Notable success, benefiting from France's geographic advantages and sailing tradition.

# 03 - Data Visualization

Next, among these medals, we curious about the height and weight of medalists 🏅



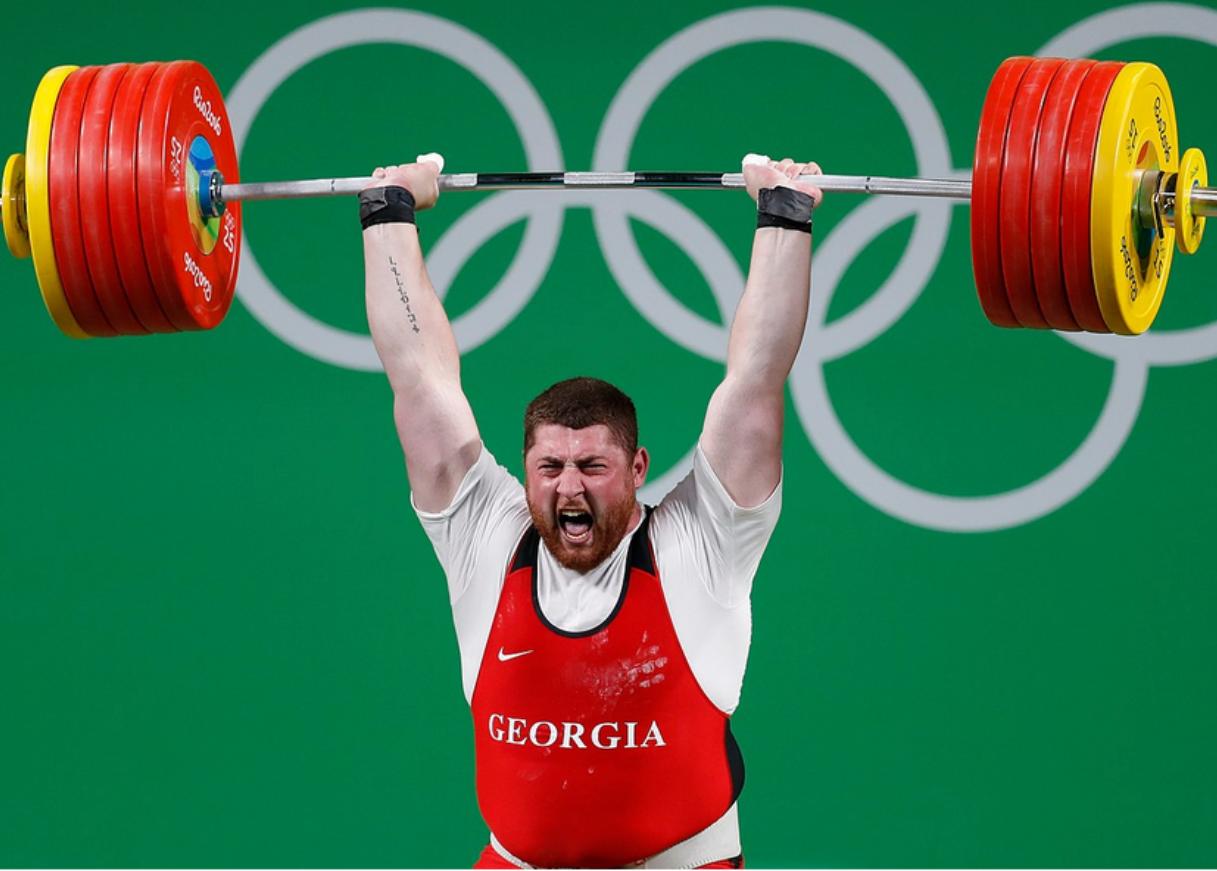
# 03 - Data Visualization

Heavier than 160 kg 🥇

```
1 notNullMedals.loc[notNullMedals['Weight'] > 160]
Executed at 2024.02.29 10:47:29 in 54ms
```

	Name	Sex	Age	Height	Weight	Year	Season	Sport	Medal	Country
3884	Andrey Ivanovich Chemerkin	Male	24.0	183.0	170.0	1996	Summer	Weightlifting	Gold	Russia
29824	Leonid Ivanovich Zhabotynskiy	Male	26.0	189.0	163.0	1964	Summer	Weightlifting	Gold	Russia
29825	Leonid Ivanovich Zhabotynskiy	Male	30.0	189.0	163.0	1968	Summer	Weightlifting	Gold	Russia

Voila! That's Weight-lifting



So. How about medalists taller than 210cm?

# 03 - Data Visualization

Taller than 210 cm 🥇

```
1 notNullMedals.loc[notNullMedals['Height'] > 210]
Executed at 2024.02.29 10:49:46 in 33ms
```

24 rows × 10 columns pd.DataFrame

	Name	Sex	Age	Height	Weight	Year	Season	Sport	Medal	Country
1843	Walter Jones "Walt" Bellamy, Jr.	Male	23.0	211.0	98.0	1960	Summer	Basketball	Gold	USA
2178	Oleksandr Mykhailovych Bielostieniy	Male	29.0	214.0	117.0	1988	Summer	Basketball	Gold	Russia
3810	Tyson Cleotis Chandler	Male	29.0	216.0	107.0	2012	Summer	Basketball	Gold	USA
4536	Melvin Grant "Mel" Counts	Male	22.0	213.0	102.0	1964	Summer	Basketball	Gold	USA
6638	Patrick Aloysius Ewing	Male	21.0	213.0	109.0	1984	Summer	Basketball	Gold	USA
6639	Patrick Aloysius Ewing	Male	29.0	213.0	109.0	1992	Summer	Basketball	Gold	USA
7422	Marcus Ross "Marc" Freiberger	Male	23.0	211.0	98.0	1952	Summer	Basketball	Gold	USA
10111	Valeriy Hryhorovych Hoborov	Male	23.0	211.0	115.0	1988	Summer	Basketball	Gold	Russia
10433	Dwight David Howard	Male	22.0	211.0	120.0	2008	Summer	Basketball	Gold	USA
10801	Darrall Tucker Imhoff	Male	21.0	211.0	98.0	1960	Summer	Basketball	Gold	USA

Most of them come from Basketball. That's make sense!





The finding continues...  
Thank You