

DATA ANALYSIS ON OLYMPICS DATA: Report

Inspiration:

I was reading an article(<https://nuunlife.com/blogs/news/a-timeline-of-women-in-the-olympics>): Women in Olympics and there were so many fun facts that the author shared. I was very much intrigued by the study and wanted to do analysis on the Olympics data myself. Also, I have always been interested in sports, thus, it was so much fun to treat the data and find patterns and anomalies.



INTERESTED FINDINGS AND VISUALISATIONS:

1. **SOURCE OF DATA:** It is a dataset on the modern Olympic Games, including all the Games from Athens 1896 to Rio 2016. The data have been scraped from www.sports-reference.com in May 2018.
2. The dataset has **271116 rows and 15 columns**.
3. The youngest athlete to have participated in Olympics was at the age of **10** and oldest was at the age of **97**.
4. The minimum height of the athlete was **127 cm** and maximum height was **226 cm**
5. The minimum weight of the athlete was **25 kg** and Maximum weight of the athlete was **214kg**.
6. The name of an athlete that appears the most is **Robert Tait McKenzie**, about **58 times**. Verified this information on Wikipedia as McKenzie participated in every Olympic Summer Games from 1896 in Athens to the 1932 L.A. Olympics
7. Out of the two genders, **male** has dominated the Olympics
8. The athletes from **United States** have participated the most in the Olympics till now.
9. Out of **39783** medals awarded, **13372** were gold medals.
10. **Men's Football** as the game and **London** as city appeared the most in the data.
11. **Most Gold Medals**

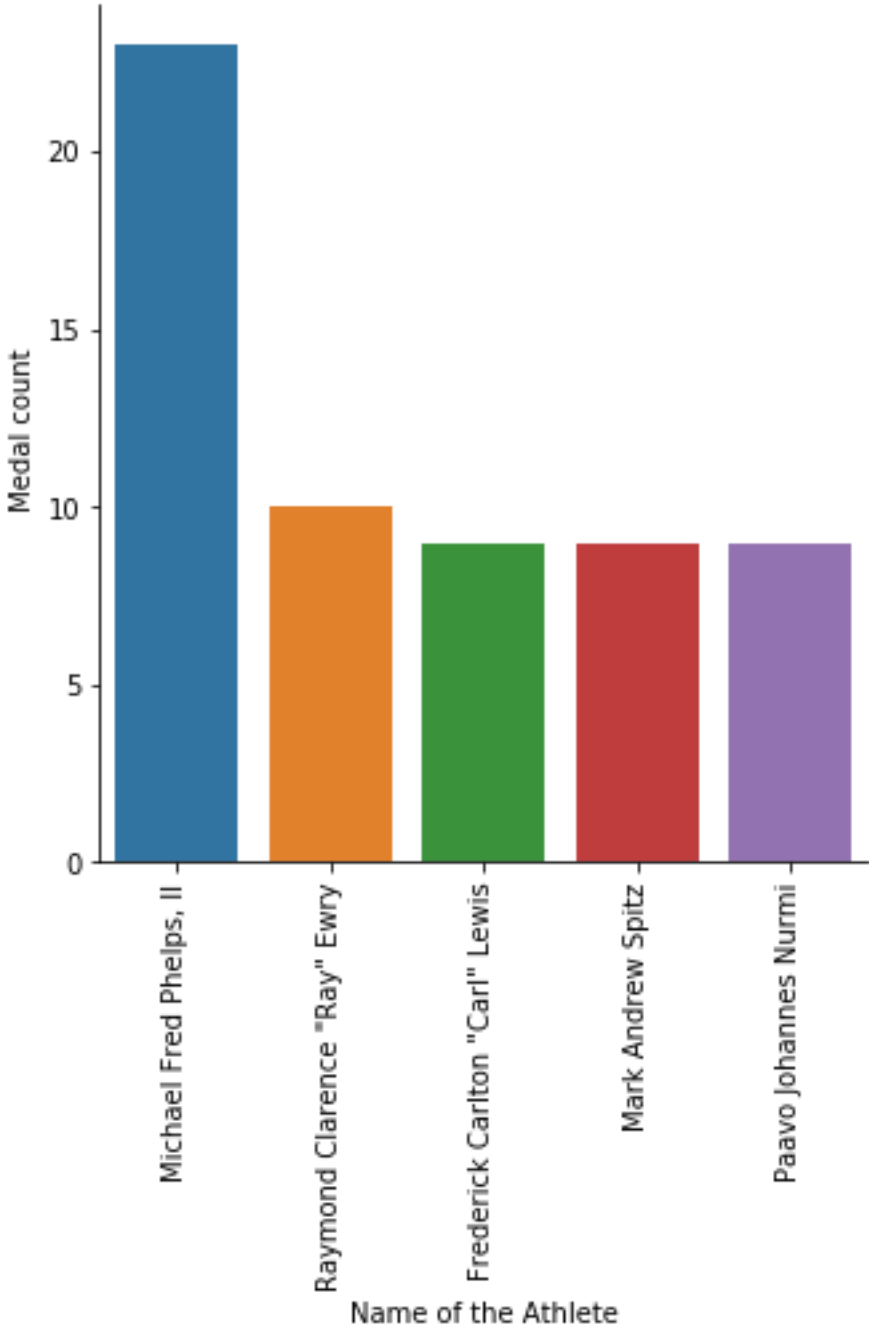
Question1 : Who won the most gold medals???

```
In [19]: gold_df.Name.value_counts().reset_index(name = "Medal count").head(5)
```

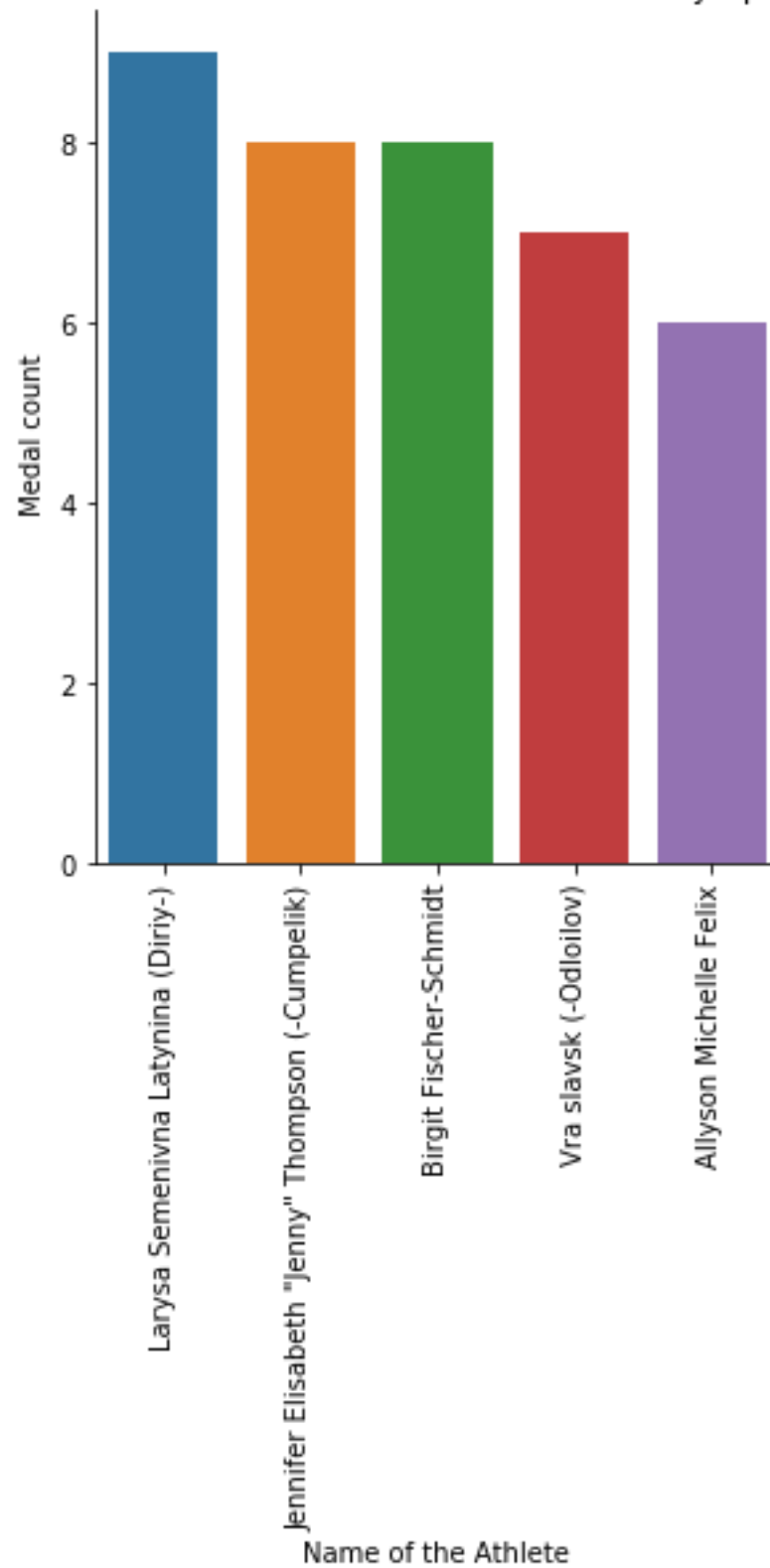
```
Out[19]:
```

	index	Medal count
0	Michael Fred Phelps, II	23
1	Raymond Clarence "Ray" Ewry	10
2	Mark Andrew Spitz	9
3	Larysa Semenivna Latynina (Diriy-)	9
4	Frederick Carlton "Carl" Lewis	9

Top 5 male athletes to win Gold Medals in Olympics History



Top 5 female athletes to win Gold Medals in Olympics History



13. Country that won maximum Gold Medals

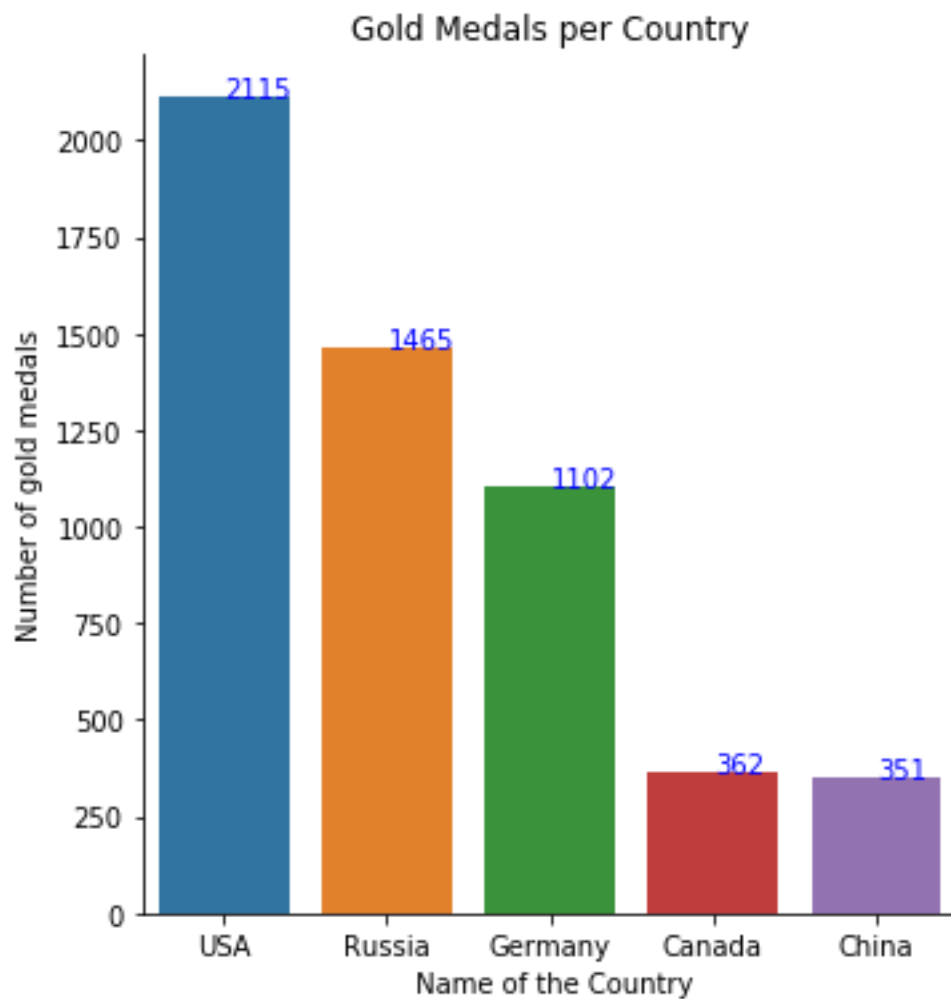
Question 3: Which country won the maximum gold medals?

```
In [24]: gold_df["region"].value_counts().reset_index(name = "Medal count").head(5)
```

Out[24]:

	index	Medal count
0	USA	2115
1	Russia	1465
2	Germany	1102
3	Canada	362
4	China	351

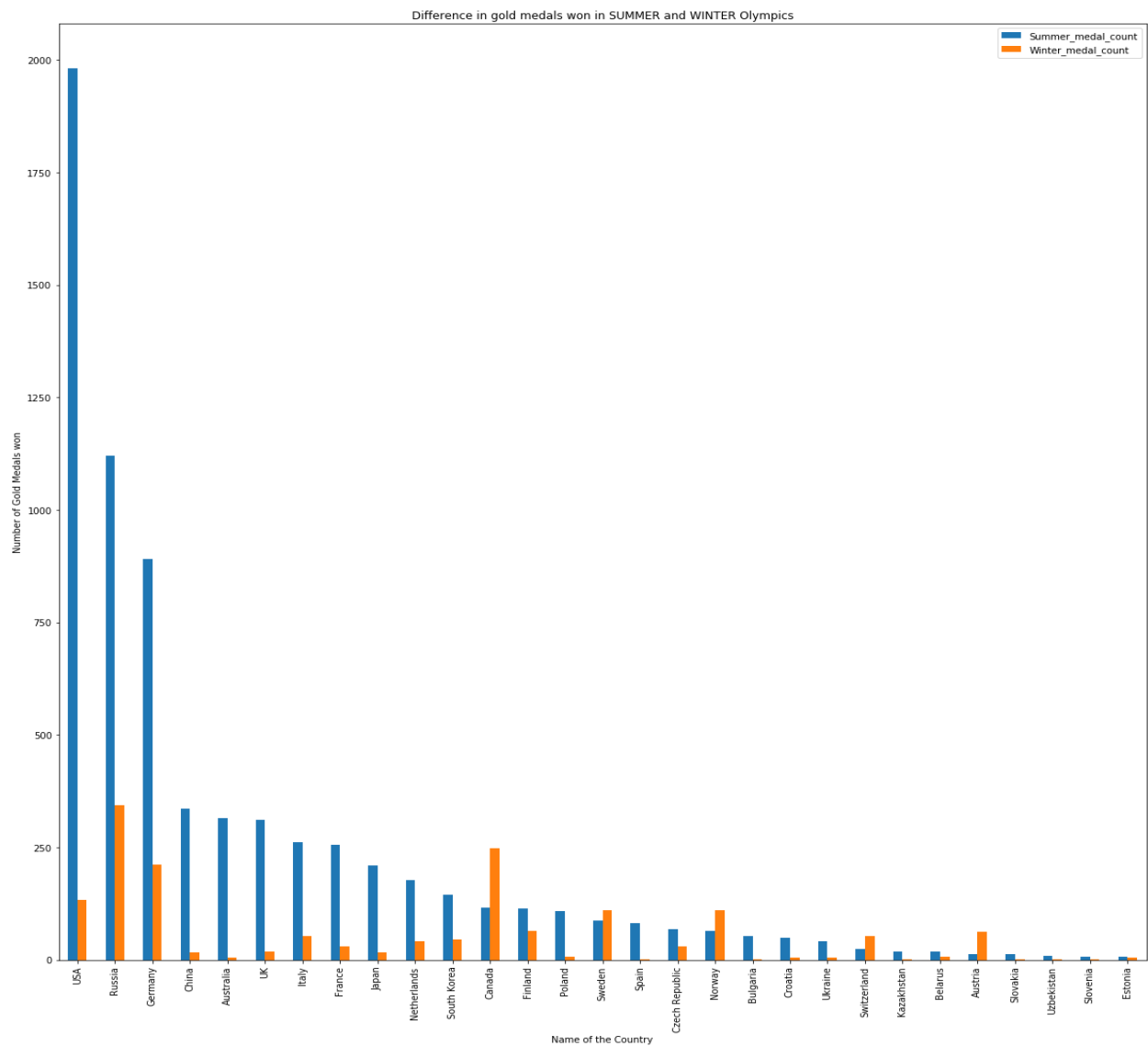
United States won the maximum gold medals followed by Soviet Union.



14. Which country has the biggest difference between their summer gold medal counts and winter gold medal counts relative to their total gold medal count?

Out[33]:

	Season_x	region	medal count_x	Season_y	medal count_y	Medal count difference	Relative medal count difference
0	Summer	USA	1981	Winter	134	1847	0.873286
1	Summer	Russia	1121	Winter	344	777	0.530375
2	Summer	Germany	890	Winter	212	678	0.615245
3	Summer	China	335	Winter	16	319	0.908832
4	Summer	Australia	314	Winter	5	309	0.968652



15. Who was the youngest and oldest player to won gold in Olympics?

Question 5: Who was the youngest and oldest player to have won a gold in olympics ? In which sport? and from which country?

```
In [38]: # sorting the dataframe by descending order of column Age
gold_df[["Name", "Age", "Height", "Weight", "Team", "Games", "Year", "Event"]].sort_values(by = "Age", ascending = False).head(5)
```

```
Out[38]:
```

	Name	Age	Height	Weight	Team	Games	Year	Event
173097	William Herbert "Bill" Northam	59.0	183.0	81.0	Barrenjoey	1964 Summer	1964	Sailing Mixed 5.5 metres
222156	Nicholas David "Nick" Skelton	58.0	175.0	76.0	Great Britain	2016 Summer	2016	Equestrianism Mixed Jumping, Individual
169083	Josef Carl Peter Neckermann	56.0	183.0	65.0	West Germany	1968 Summer	1968	Equestrianism Mixed Dressage, Team
13396	Nikolaus "Klaus" Balkenhol	56.0	178.0	83.0	Germany	1996 Summer	1996	Equestrianism Mixed Dressage, Team
222155	Nicholas David "Nick" Skelton	54.0	175.0	76.0	Great Britain	2012 Summer	2012	Equestrianism Mixed Jumping, Team

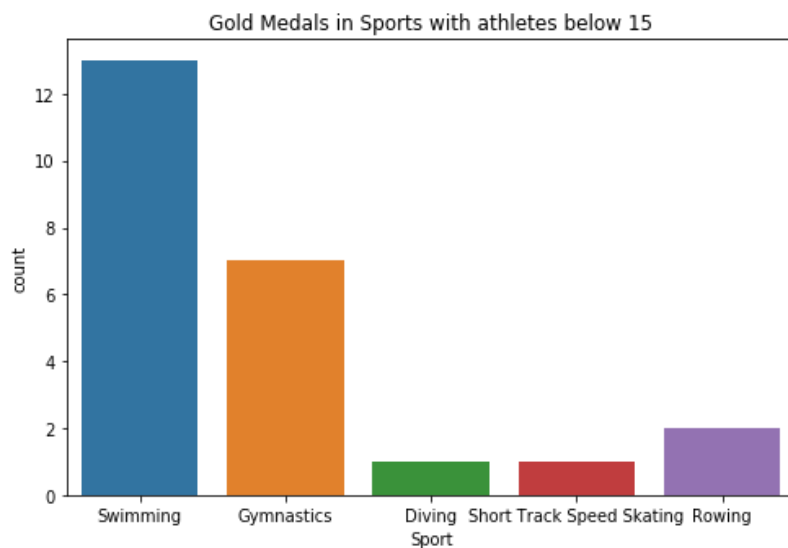
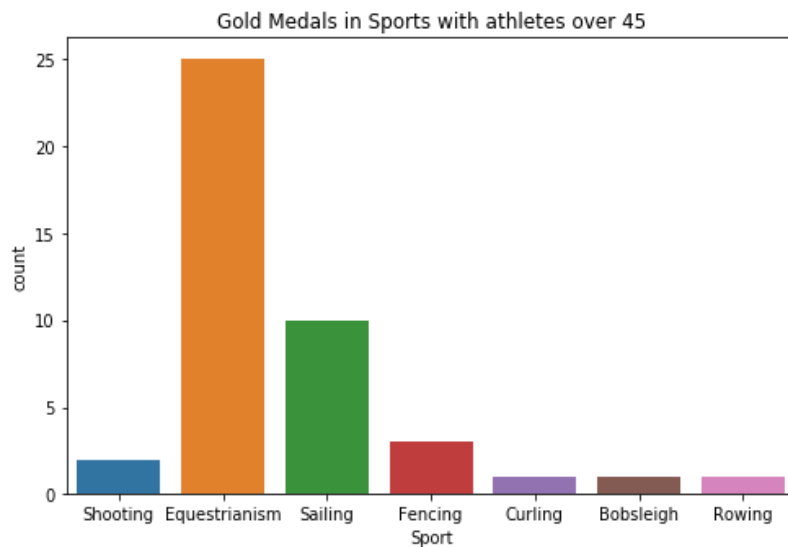
```
In [39]: # sorting the dataframe by ascending order of Age
gold_df[["Name", "Age", "Height", "Weight", "Team", "Games", "Year", "Event"]].sort_values(by = "Age").head(5)
```

```
Out[39]:
```

	Name	Age	Height	Weight	Team	Games	Year	Event
73581	Fu Mingxia	13.0	160.0	48.0	China	1992 Summer	1992	Diving Women's Platform
119277	Kim Yun-Mi	13.0	160.0	48.0	South Korea	1994 Winter	1994	Short Track Speed Skating Women's 3,000 metres...
52010	Donna Elizabeth de Varona (-Pinto)	13.0	168.0	61.0	United States	1960 Summer	1960	Swimming Women's 4 x 100 metres Freestyle Relay
268602	Klaus Zerta	13.0	165.0	48.0	Germany	1960 Summer	1960	Rowing Men's Coxed Pairs
43773	Nadia Elena Comneci (-Conner)	14.0	162.0	45.0	Romania	1976 Summer	1976	Gymnastics Women's Individual All-Around

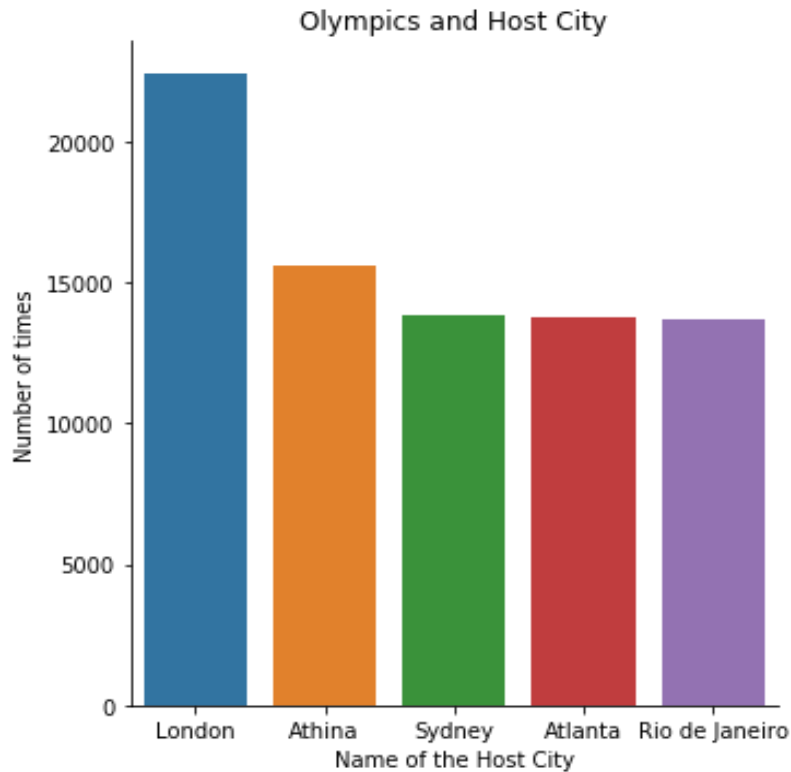
The oldest player to win Olympics gold medal is William Herbert "Bill" Northam in the event "Sailing Mixed 5.5 meters" at the age of 59 and the youngest player to win gold is Fu Mingxia in the event of "Diving Women's Platform".

16. How many athletes have won the gold medals above the age of 50 and below the age of 15? In which sports?



From the above visualisations, we can clearly see that people above age 45 scored gold in sports like sailing, shooting and even horse riding, whereas for athletes below 15, swimming and gymnastics are very popular. The sports are very contrasting in nature as swimming is about dynamics and gymnastics is all about flexibility.

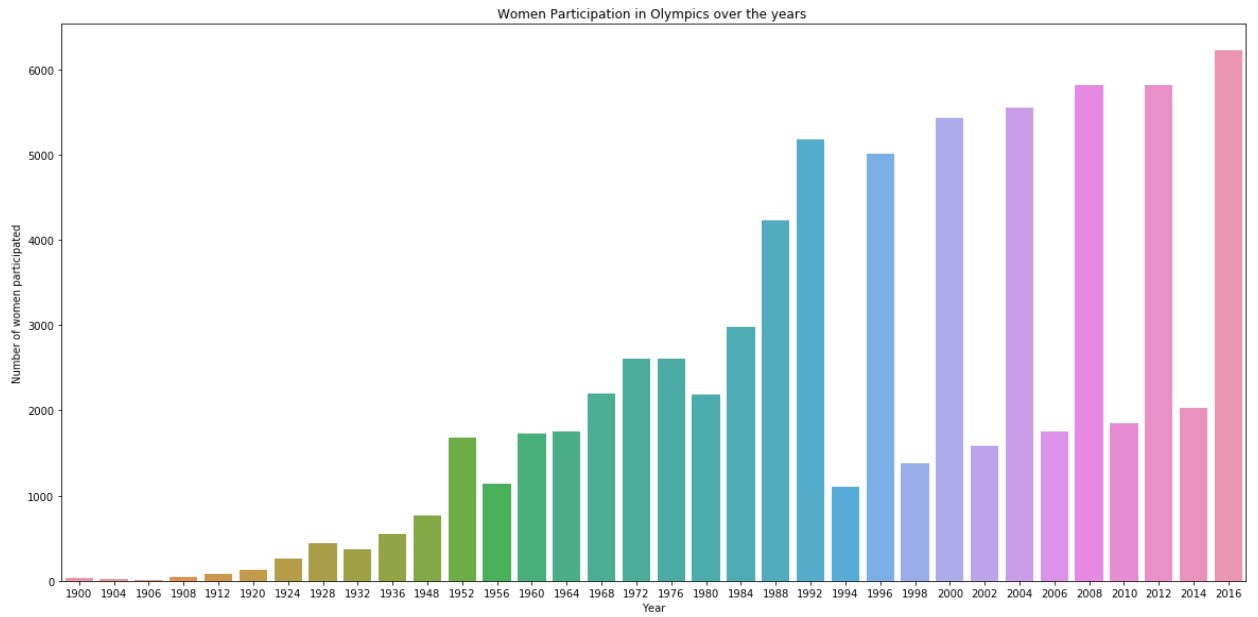
17. Which city hosted the Olympics the most?



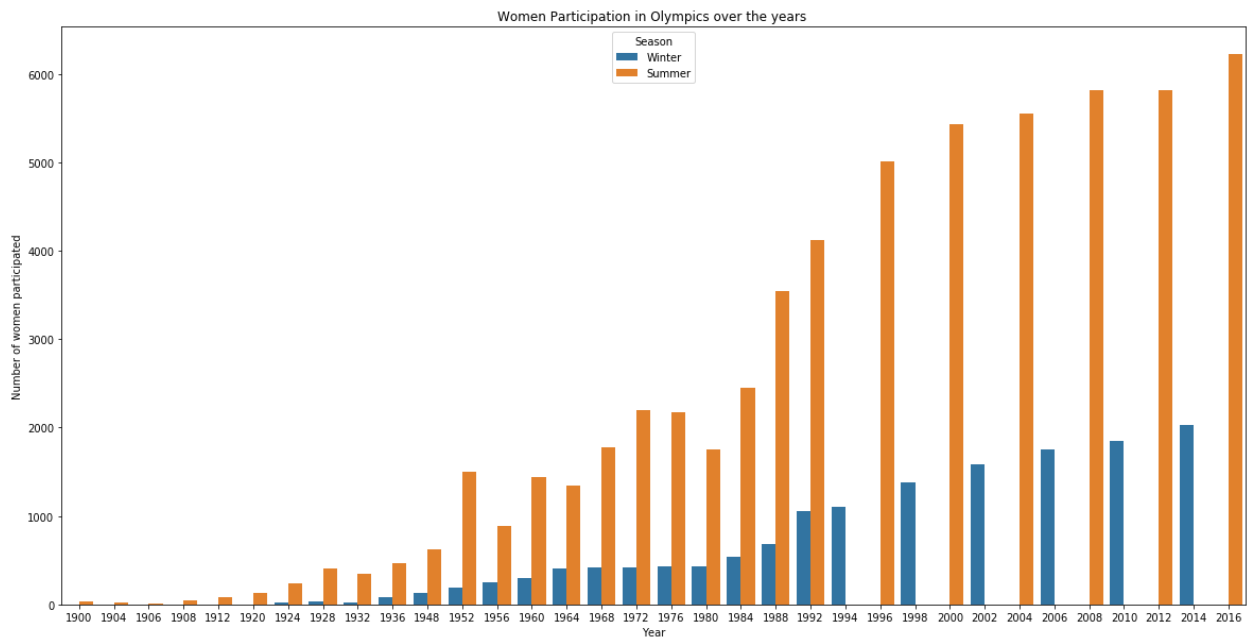
London is clearly a winner in terms of how many times Olympics were hosted with Athina on second, and, on the other hand, Sydney, Atlanta, and Rio de Janeiro were very close.

18. Women Participation in Olympics

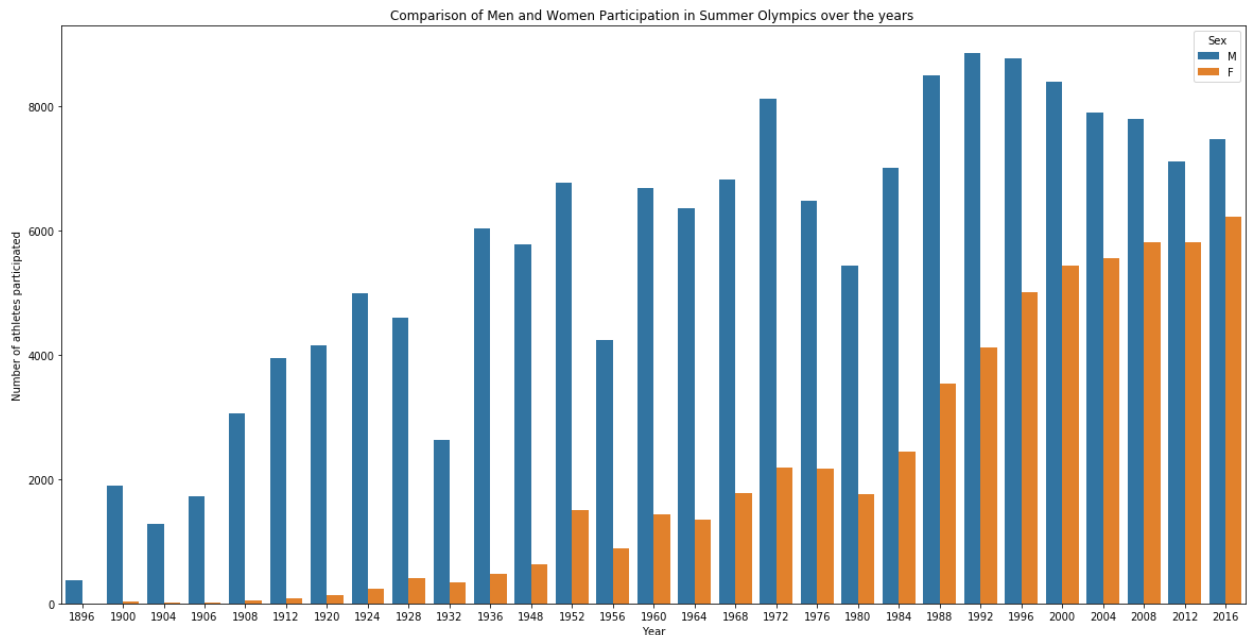




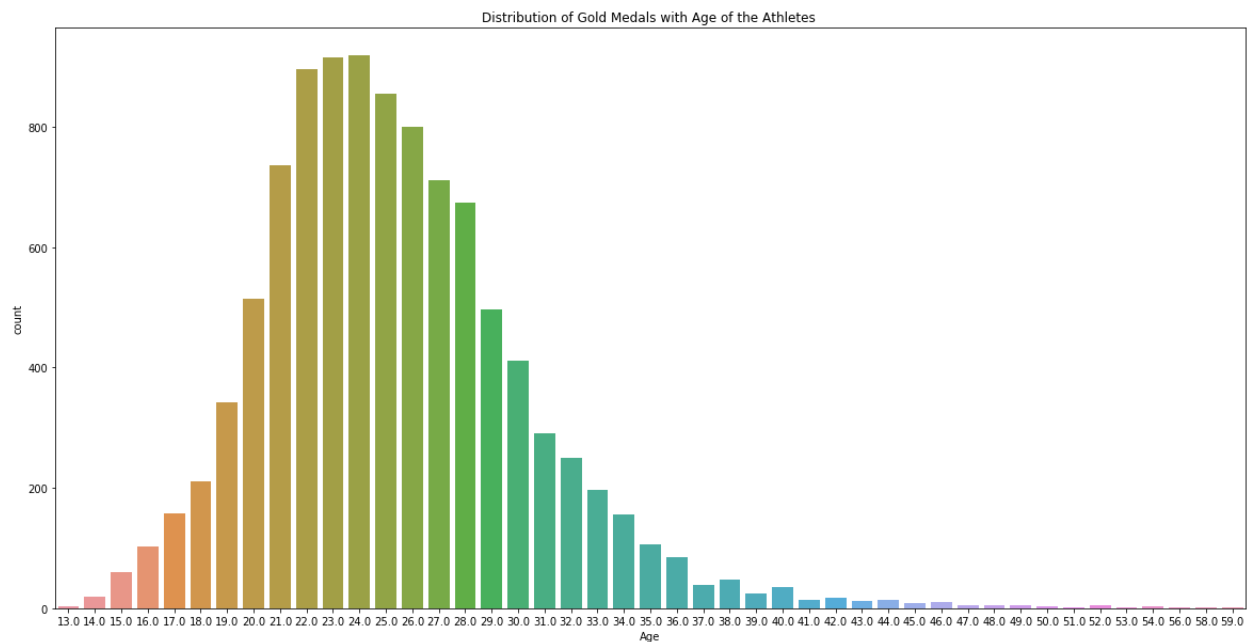
19. Comparison of Women Participation in Summer and Winter Olympics



20. Comparison of Men and Women Participation

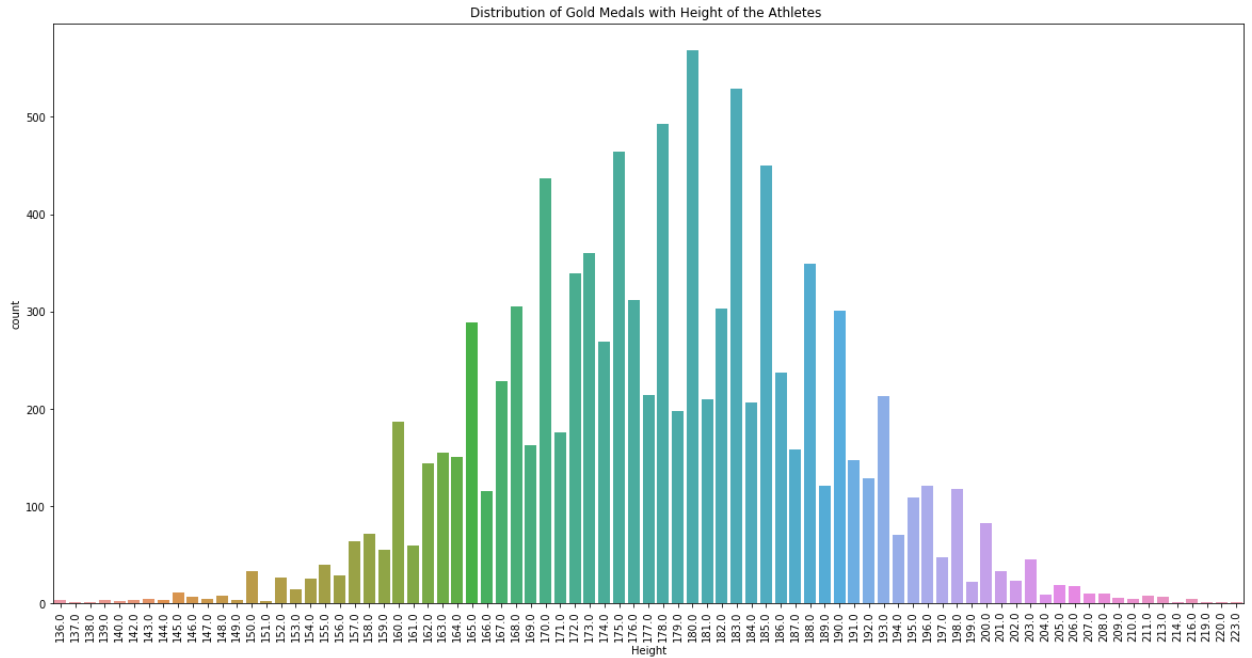


21. Distribution of Gold Medals among the Age of Athletes



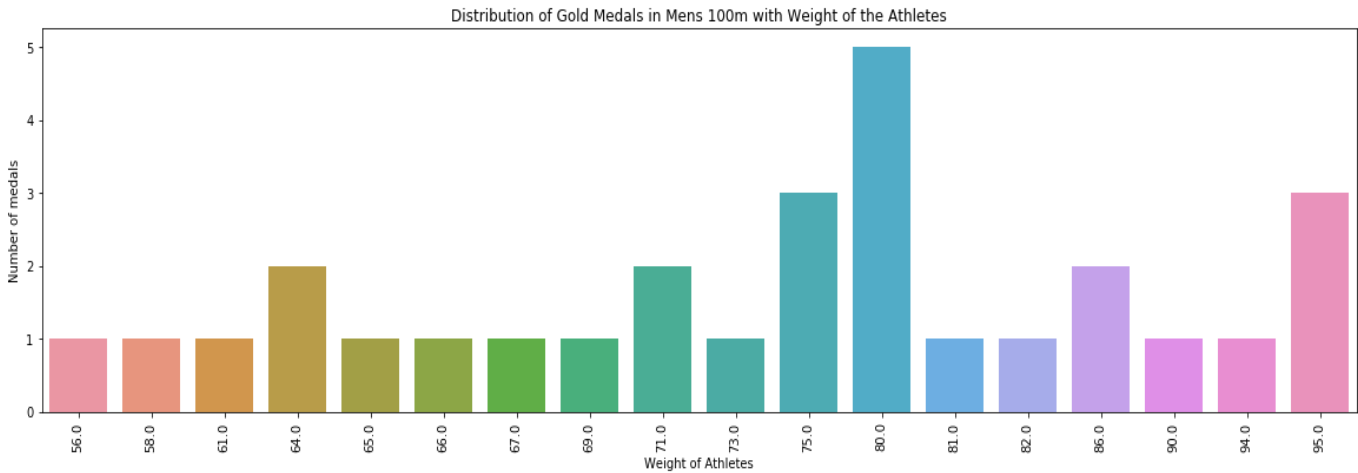
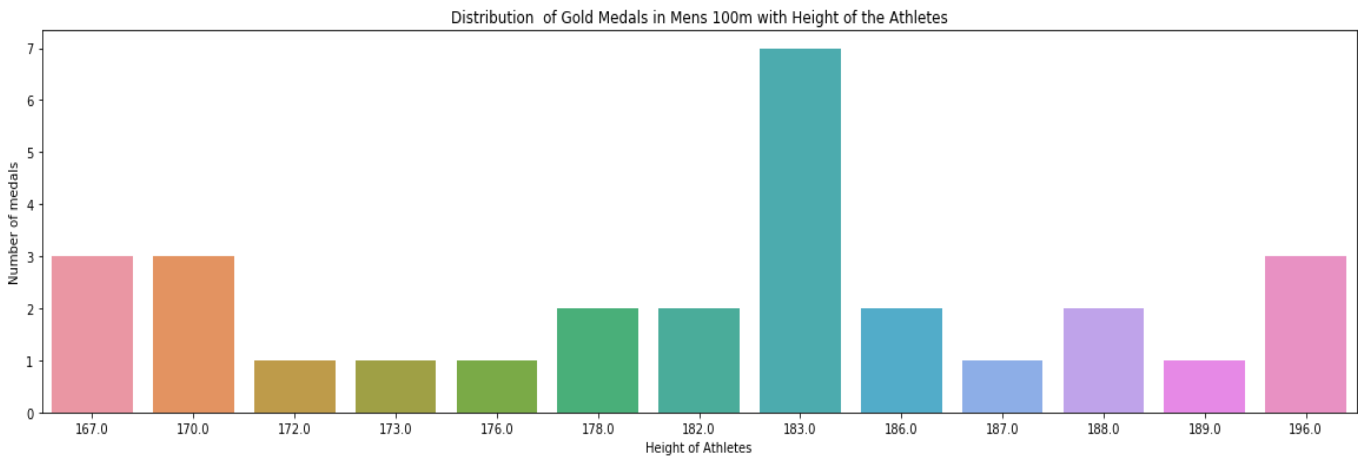
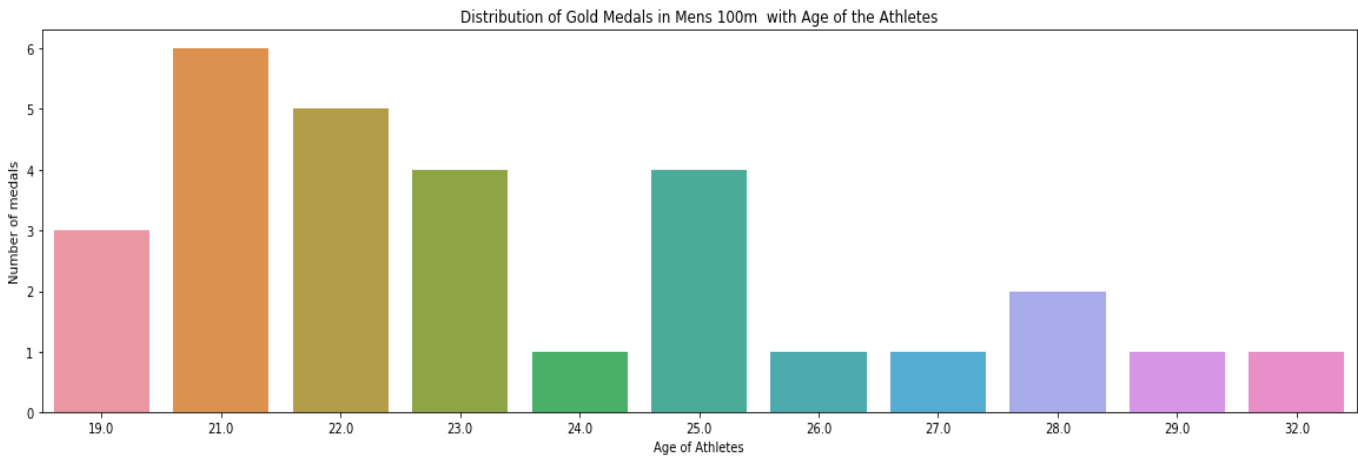
From the plot we can see that most of the athletes that won the gold medals have the age window of 20-30 and 24 is the peak age at which most of the athletes have won the gold medals.

22. Distribution of Gold Medals among the Height of Athletes

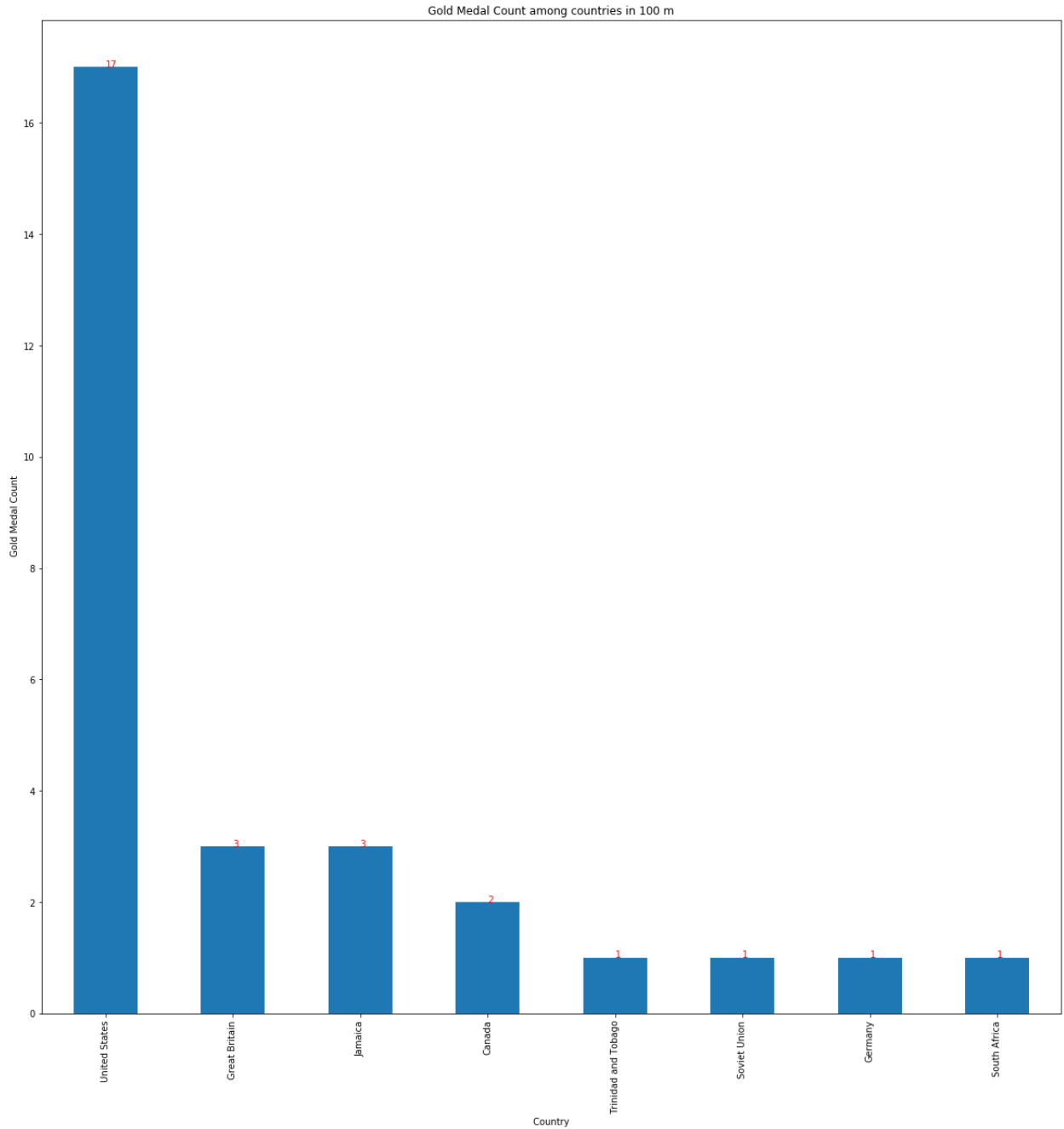


By looking at the distribution plot, There seems to be a pattern of a wave, looks like optimal height to win a gold medal is 170 to 190 cm, however there are few spikes at certain heights.

23. **Distribution of Gold Medals with the Age, Weight and Height of athletes in 100m men**



24. Gold medal count among countries in 100m



There is so much more analytics and visualisation techniques that can be used to find more interesting facts and patterns. TO BE CONTINUED.....