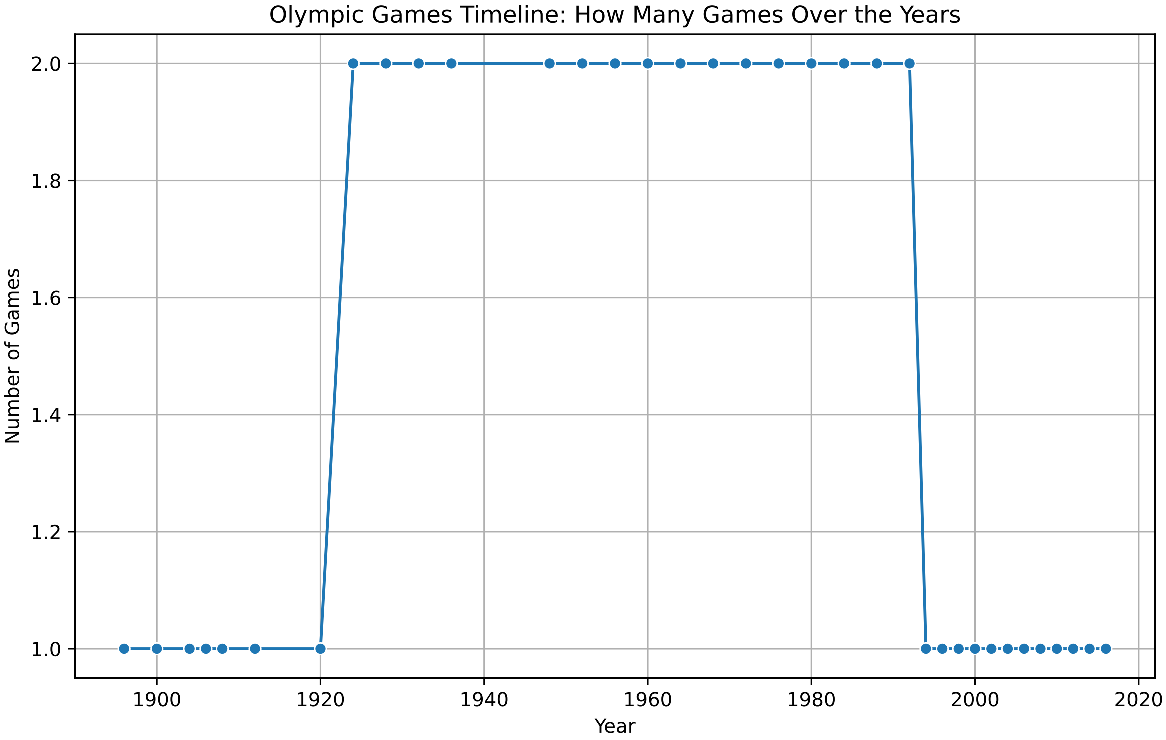
SUMMER OLYMPICS

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load the datasets
noc_regions = pd.read_csv('~/dataset/noc_regions.csv')
athlete_events = pd.read_csv('~/dataset/athlete_events.csv')


# 1.  How Many Olympic Games Have We Had?
total_games_played = athlete_events['Games'].nunique()
print(f'Total games played: {total_games_played}')

# Visualization 1: Olympic Games Timeline
games_per_year = athlete_events[['Year', 'Games']].drop_duplicates().groupby('Year').size()
plt.figure(figsize=(10, 6))
sns.lineplot(x=games_per_year.index, y=games_per_year.values, marker='o')
plt.title('Olympic Games Timeline: How Many Games Over the Years')
plt.xlabel('Year')
plt.ylabel('Number of Games')
plt.grid(True)
plt.show()

# 2.  When and Where? Olympic Games Timeline
all_games_held = athlete_events[['Games', 'Sport', 'City', 'Year']].drop_duplicates().sort_va
print(all_games_held)

# Visualization 2: Olympic Games Across Cities
plt.figure(figsize=(12, 8))
sns.scatterplot(data=all_games_held, x='Year', y='Games', hue='City', style='Sport', s=100)
plt.title('Olympic Games Across Cities: When and Where')
plt.xlabel('Year')
plt.ylabel('Games')
plt.legend(bbox_to_anchor=(1.05, 1), loc='upper left')
plt.grid(True)
plt.show()

# 3.  Sports Variety: What Have We Played?
sports_played = athlete_events[['Sport', 'Games']].drop_duplicates()
print(sports_played)

# Visualization 3: Variety of Sports Over Time
sports_per_game = athlete_events.groupby('Games')['Sport'].nunique().reset_index()
plt.figure(figsize=(12, 8))
sns.barplot(data=sports_per_game, x='Games', y='Sport', palette='viridis')
plt.xticks(rotation=45, ha='right')
plt.title('Variety of Sports Played in Each Olympic Game')
plt.xlabel('Games')
plt.ylabel('Number of Sports')
plt.show()

# 4.  Who Showed Up? Nation Participation
all_nations = athlete_events.merge(noc_regions, left_on='Team', right_on='region')[['Games',
total_no_games = all_nations.groupby('Games')['Team'].nunique().reset_index().rename(columns=
print(total_no_games)

# Visualization 4: Nations Joining the Party
plt.figure(figsize=(12, 8))
sns.lineplot(data=total_no_games, x='Games', y='Number_of_Nations', marker='o')
plt.xticks(rotation=45, ha='right')
plt.title('Nations Participating in Each Olympic Game')
plt.xlabel('Games')
plt.ylabel('Number of Nations')
plt.grid(True)
plt.show()

# 5.  Participation Extremes: Highest & Lowest
all_countries = athlete_events.merge(noc_regions, left_on='Team', right_on='region')[['Games
total_countries = all_countries.groupby('Games').size().reset_index(name='Total_Countries')
lowest_countries = total_countries.loc[total_countries['Total_Countries'].idxmin()]
highest_countries = total_countries.loc[total_countries['Total_Countries'].idxmax()]
print(f'Lowest Countries: {lowest_countries["Games"]} - {lowest_countries["Total_Countries"]
print(f'Highest Countries: {highest_countries["Games"]} - {highest_countries["Total_Countrie
# Visualization 5: Highs and Lows of Participation
plt.figure(figsize=(12, 8))
sns.barplot(data=total_countries, x='Games', y='Total_Countries', palette='coolwarm')
plt.xticks(rotation=45, ha='right')
plt.title('Participation Over Time: Highs and Lows')
plt.xlabel('Games')
plt.ylabel('Number of Countries')
plt.show()

# 6.   Ever-Present Nations: Who's Always Been There?
total_games = athlete_events['Games'].nunique()
the_countries = athlete_events.merge(noc_regions, left_on='Team', right_on='region')[['Games
all_countries_that_participated = the_countries.groupby('region').size().reset_index(name='Ev
nations_in_all_games = all_countries_that_participated[all_countries_that_participated['Every
print(nations_in_all_games)

# Visualization 6: Nations That Never Missed a Game
plt.figure(figsize=(12, 8))
sns.barplot(data=nations_in_all_games, y='region', x='Every_Participants', palette='Set3')
plt.title('Nations That Participated in Every Olympic Game')
plt.xlabel('Number of Games Participated')
plt.ylabel('Country')
plt.show()

# 7.  Summer Classics: Sports in Every Summer Olympics
total_summer_games = athlete_events[athlete_events['Season'] == 'Summer']['Games'].nunique()
sports_in_summer = athlete_events[athlete_events['Season'] == 'Summer'].groupby('Sport')['Gan
sports_in_all_summer_games = sports_in_summer[sports_in_summer['Games'] == total_summer_games
print(sports_in_all_summer_games)

# Visualization 7: Sports in Every Summer Olympics
plt.figure(figsize=(12, 8))
sns.barplot(data=sports_in_all_summer_games, y='Sport', x='Games', palette='Spectral')
plt.title('Sports That Played in Every Summer Olympics')
plt.xlabel('Number of Games')
plt.ylabel('Sport')
plt.show()

# 8.  One-Hit Wonders: Sports Played Only Once
sports_played_once = athlete_events.groupby('Sport')['Games'].nunique().reset_index()
sports_played_once = sports_played_once[sports_played_once['Games'] == 1]
print(sports_played_once)
```

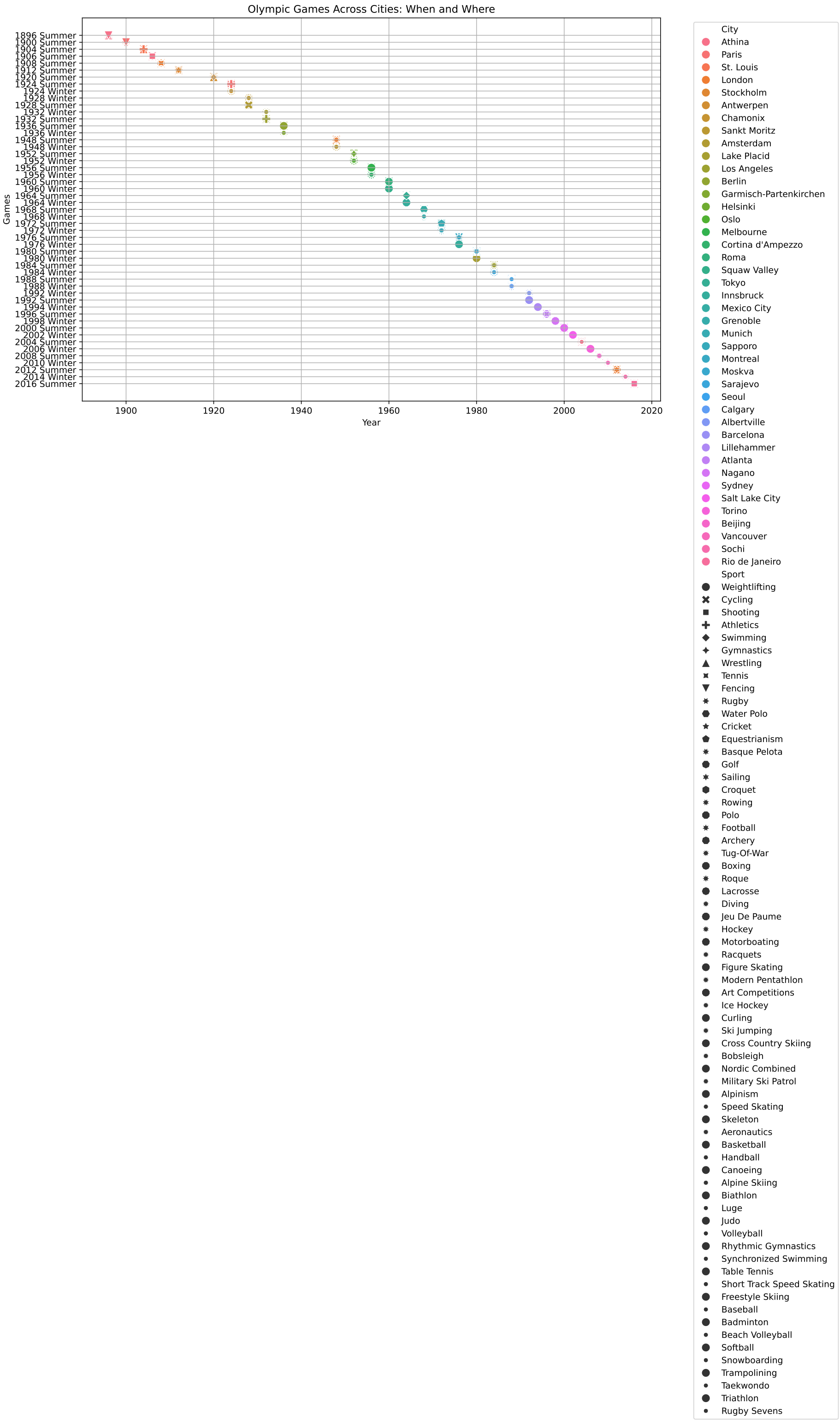
```
# Visualization 8: Sports Played Only Once in the Olympics
plt.figure(figsize=(12, 8))
sns.barplot(data=sports_played_once, y='Sport', x='Games', palette='Pastel1')
plt.title('One-Hit Wonders: Sports Played Only Once in the Olympics')
plt.xlabel('Number of Games')
plt.ylabel('Sport')
plt.show()

# 9.  Golden Oldies: The Oldest Gold Medalists
oldest_gold_medalist = athlete_events[athlete_events['Medal'] == 'Gold'].sort_values(by='Age
print(oldest_gold_medalist)

# Visualization 9: Age of Gold Medalists
plt.figure(figsize=(12, 8))
sns.histplot(athlete_events[athlete_events['Medal'] == 'Gold']['Age'], bins=30, kde=True, co
plt.title('Age Distribution of Gold Medalists')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.show()

# 10.  Gender Showdown: Ratio of Male to Female Athletes
gender_count = athlete_events.groupby('Sex').size()
ratio = gender_count.max() / gender_count.min()
print(f'Ratio: 1:{ratio:.2f}')

# Visualization 10: Gender Ratio in the Olympics
plt.figure(figsize=(8, 8))
gender_count.plot(kind='pie', autopct='%1.1f%%', startangle=140, colors=['skyblue', 'pink'],
plt.title('Gender Ratio in the Olympics')
plt.ylabel('')
plt.show()

# 11.  Top 10 Gold Medalists
top_10_gold_medalists = athlete_events[athlete_events['Medal'] == 'Gold'].groupby(['Name', '
print(top_10_gold_medalists)

# Visualization 11: The Gold Medal Elites
plt.figure(figsize=(14, 8))
sns.barplot(data=top_10_gold_medalists, y='Name', x='Total_Gold_Medals', hue='Sex', palette=
plt.title('Top 10 Athletes with the Most Gold Medals')
plt.xlabel('Total Gold Medals')
plt.ylabel('Athlete')
plt.show()

# 12.   Top 10 Countries by Medals
top_10_countries = athlete_events[athlete_events['Medal'].notna()].groupby('Team').size().res
print(top_10_countries)

# Visualization 12: The Top 10 Most Successful Countries
plt.figure(figsize=(12, 8))
sns.barplot(data=top_10_countries, y='Team', x='Total_Medals', palette='Blues_d')
plt.title('Top 10 Most Successful Countries in Olympics')
plt.xlabel('Total Medals')
plt.ylabel('Country')
plt.show()
```

```
Total games played: 51
```



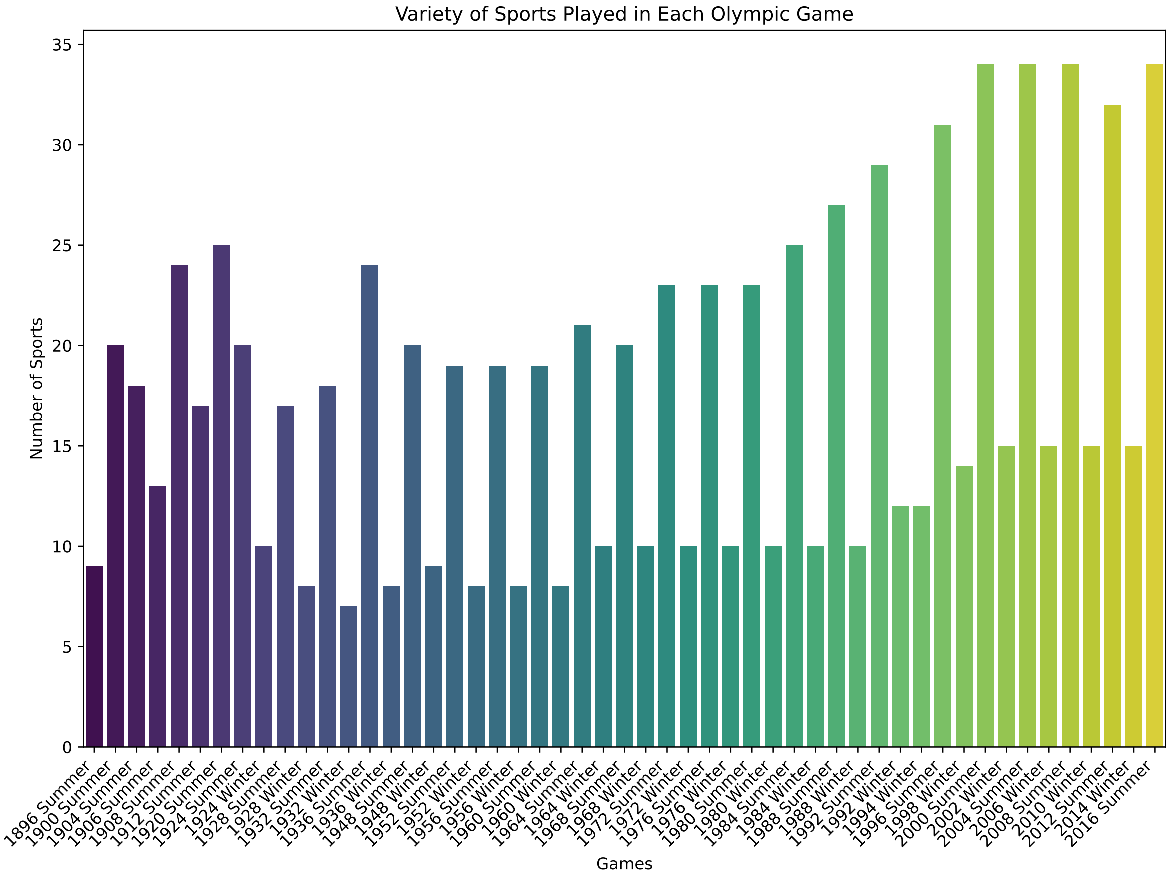
```
              Games             Sport           City  Year
63069  1896 Summer    Weightlifting         Athina  1896
10332  1896 Summer          Cycling         Athina  1896
```

```
8144   1896 Summer          Shooting          Athina  1896
7348   1896 Summer         Athletics          Athina  1896
7353   1896 Summer          Swimming          Athina  1896
...            ...               ...             ...   ...
282    2016 Summer           Cycling  Rio de Janeiro  2016
2175   2016 Summer  Beach Volleyball  Rio de Janeiro  2016
412    2016 Summer              Judo  Rio de Janeiro  2016
769    2016 Summer      Rugby Sevens  Rio de Janeiro  2016
386    2016 Summer          Shooting  Rio de Janeiro  2016

[906 rows x 4 columns]
```


Olympic Games Across Cities: When and Where
Games
1896 Summer
1900 Summer
1904 Summer
1906 Summer
1908 Summer
1912 Summer
1920 Summer
1924 Summer
1924 Winter
1928 Winter
1928 Summer
1932 Winter
1932 Summer
1936 Summer
1936 Winter
1948 Summer
1948 Winter
1952 Summer
1952 Winter
1956 Summer
1956 Winter
1960 Summer
1960 Winter
1964 Summer
1964 Winter
1968 Summer
1968 Winter
1972 Summer
1972 Winter
1976 Summer
1976 Winter
1980 Summer
1980 Winter
1984 Summer
1984 Winter
1988 Summer
1988 Winter
1992 Winter
1992 Summer
1994 Winter
1996 Summer
1998 Winter
2000 Summer
2002 Winter
2004 Summer
2006 Winter
2008 Summer
2010 Winter
2012 Summer
2014 Winter
2016 Summer
1900
1920
1940
1960
1980
2000
2020
Year
City
Athina
Paris
St. Louis
London
Stockholm
Antwerpen
Chamonix
Sankt Moritz
Amsterdam
Lake Placid
Los Angeles
Berlin
Garmisch-Partenkirchen
Helsinki
Oslo
Melbourne
Cortina d'Ampezzo
Roma
Squaw Valley
Tokyo
Innsbruck
Mexico City
Grenoble
Munich
Sapporo
Montreal
Moskva
Sarajevo
Seoul
Calgary
Albertville
Barcelona
Lillehammer
Atlanta
Nagano
Sydney
Salt Lake City
Torino
Beijing
Vancouver
Sochi
Rio de Janeiro
Sport
Weightlifting
Cycling
Shooting
Athletics
Swimming
Gymnastics
Wrestling
Tennis
Fencing
Rugby
Water Polo
Cricket
Equestrianism
Basque Pelota
Golf
Sailing
Croquet
Rowing
Polo
Football
Archery
Tug-Of-War
Boxing
Roque
Lacrosse
Diving
Jeu De Paume
Hockey
Motorboating
Racquets
Figure Skating
Modern Pentathlon
Art Competitions
Ice Hockey
Curling
Ski Jumping
Cross Country Skiing
Bobsleigh
Nordic Combined
Military Ski Patrol
Alpinism
Speed Skating
Skeleton
Aeronautics
Basketball
Handball
Canoeing
Alpine Skiing
Biathlon
Luge
Judo
Volleyball
Rhythmic Gymnastics
Synchronized Swimming
Table Tennis
Short Track Speed Skating
Freestyle Skiing
Baseball
Badminton
Beach Volleyball
Softball
Snowboarding
Trampolining
Taekwondo
Triathlon
Rugby Sevens

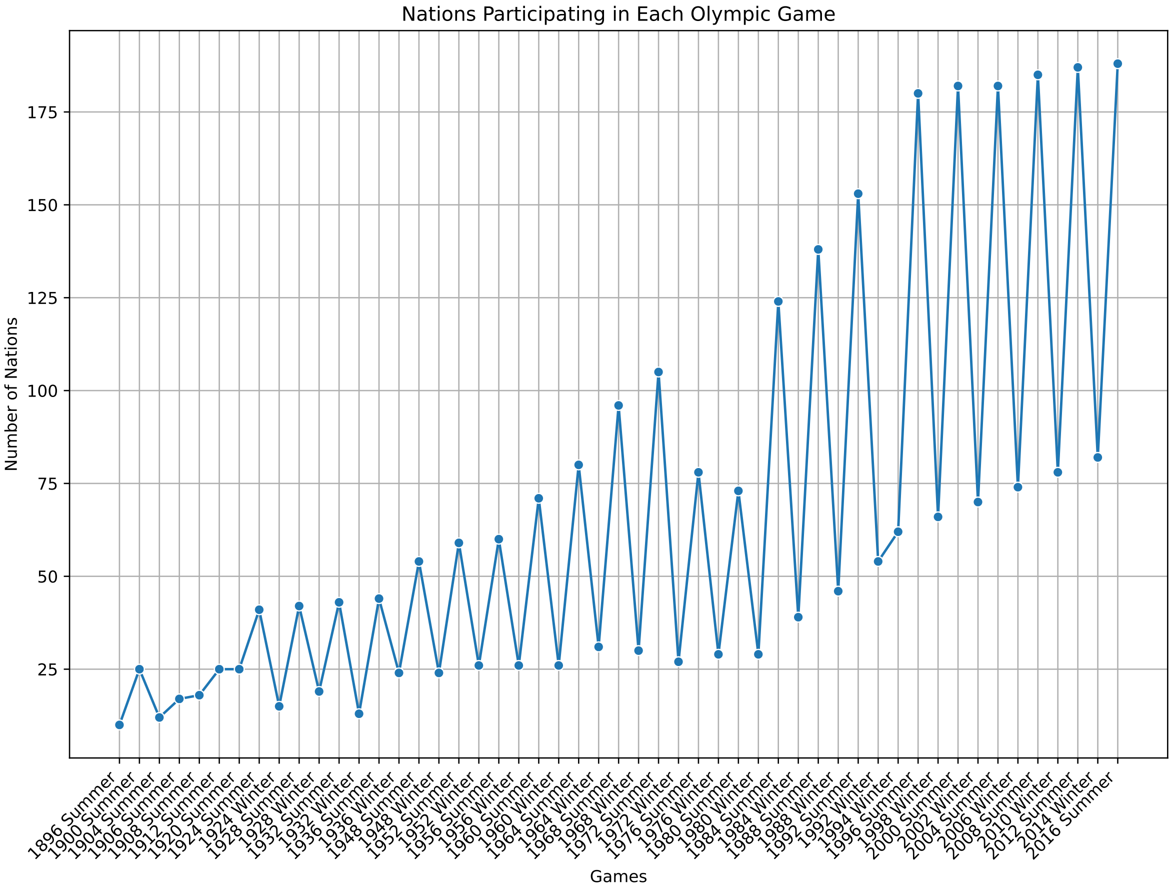
```
                 Sport        Games
0           Basketball  1992 Summer
1                 Judo  2012 Summer
2             Football  1920 Summer
3           Tug-Of-War  1900 Summer
4        Speed Skating  1988 Winter
...                ...          ...
63069    Weightlifting  1896 Summer
63071        Wrestling  1896 Summer
112109   Weightlifting  1904 Summer
213142        Alpinism  1932 Summer
214105     Aeronautics  1936 Summer

[906 rows x 2 columns]
```

```
/tmp/ipykernel_44969/1611472489.py:45: FutureWarning:



Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assi
```



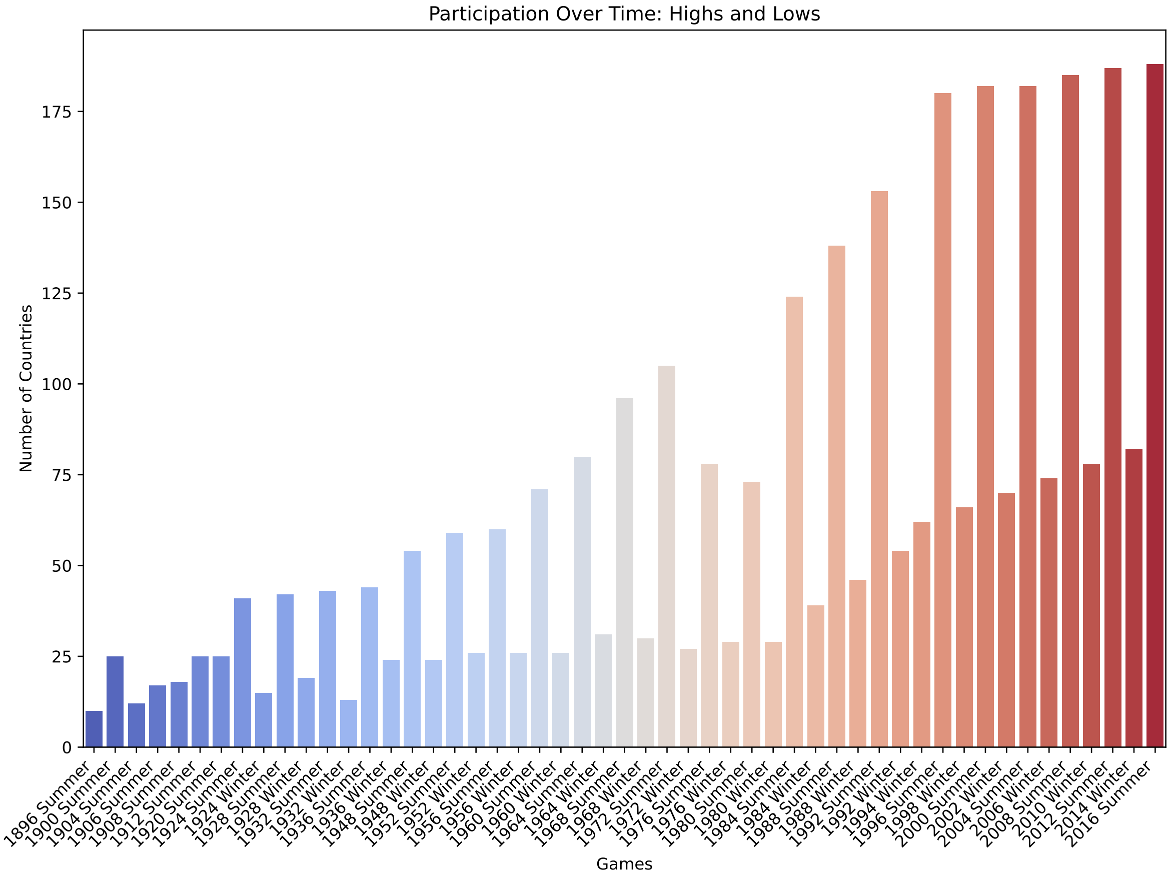
```
          Games  Number_of_Nations
0   1896 Summer                 10
1   1900 Summer                 25
2   1904 Summer                 12
3   1906 Summer                 17
4   1908 Summer                 18
5   1912 Summer                 25
6   1920 Summer                 25
7   1924 Summer                 41
8   1924 Winter                 15
9   1928 Summer                 42
10  1928 Winter                 19
11  1932 Summer                 43
12  1932 Winter                 13
13  1936 Summer                 44
14  1936 Winter                 24
15  1948 Summer                 54
```

```
16  1948 Winter                24
17  1952 Summer                59
18  1952 Winter                26
19  1956 Summer                60
20  1956 Winter                26
21  1960 Summer                71
22  1960 Winter                26
23  1964 Summer                80
24  1964 Winter                31
25  1968 Summer                96
26  1968 Winter                30
27  1972 Summer               105
28  1972 Winter                27
29  1976 Summer                78
30  1976 Winter                29
31  1980 Summer                73
32  1980 Winter                29
33  1984 Summer               124
34  1984 Winter                39
35  1988 Summer               138
36  1988 Winter                46
37  1992 Summer               153
38  1992 Winter                54
39  1994 Winter                62
40  1996 Summer               180
41  1998 Winter                66
42  2000 Summer               182
43  2002 Winter                70
44  2004 Summer               182
45  2006 Winter                74
46  2008 Summer               185
47  2010 Winter                78
48  2012 Summer               187
49  2014 Winter                82
50  2016 Summer               188
```



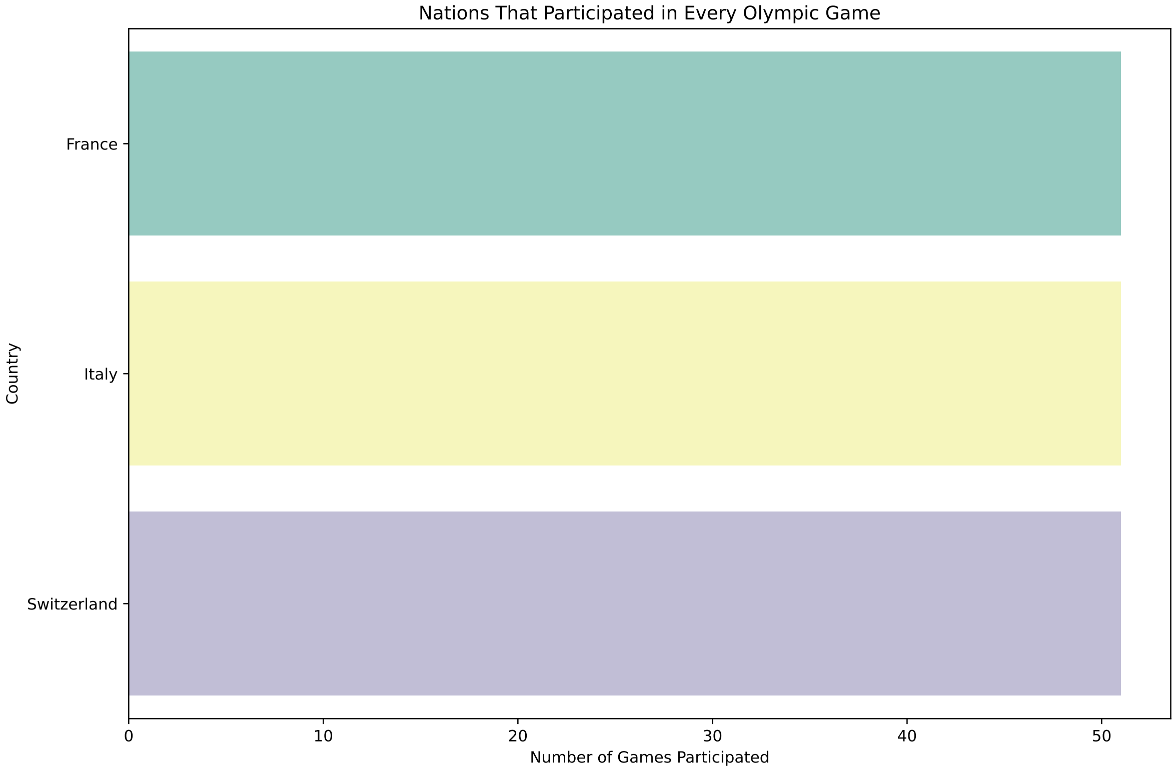
```
Lowest Countries: 1896 Summer - 10
Highest Countries: 2016 Summer - 188
```

```
/tmp/ipykernel_44969/1611472489.py:77: FutureWarning:



Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assig
```

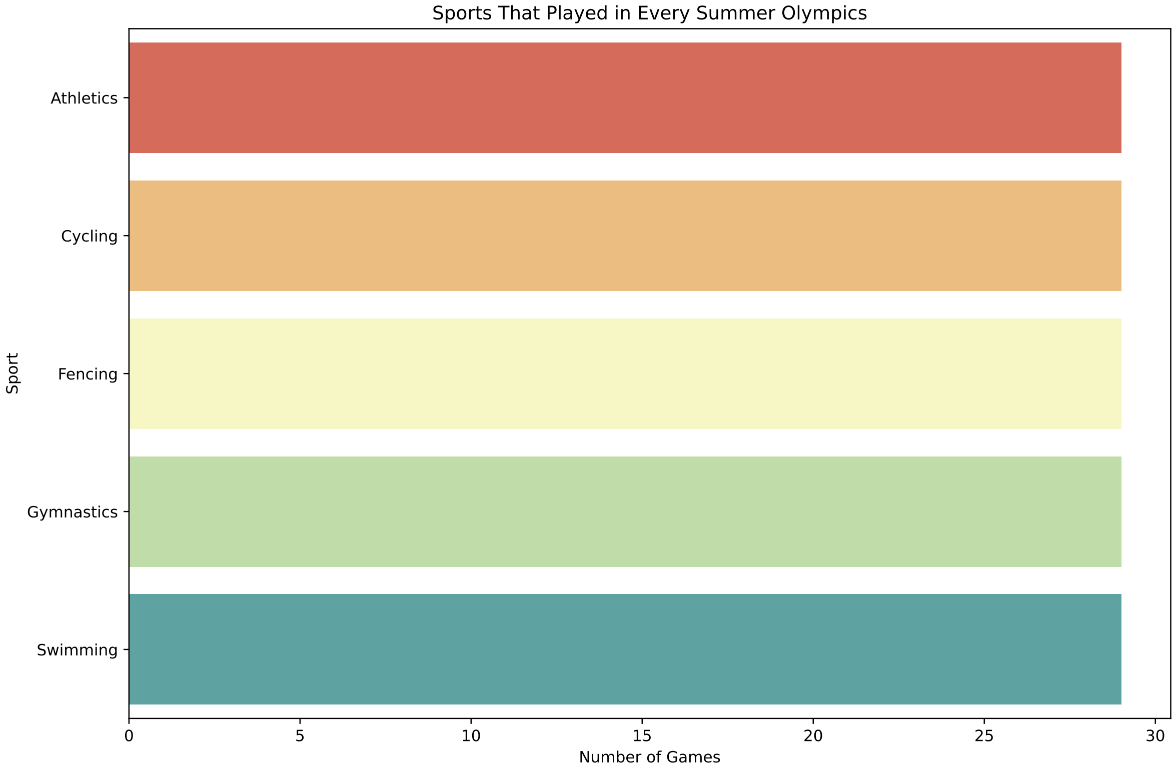


```
          region  Every_Participants
59        France                  51
82         Italy                  51
168  Switzerland                  51
```

```
/tmp/ipykernel_44969/1611472489.py:93: FutureWarning:



Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assi
```

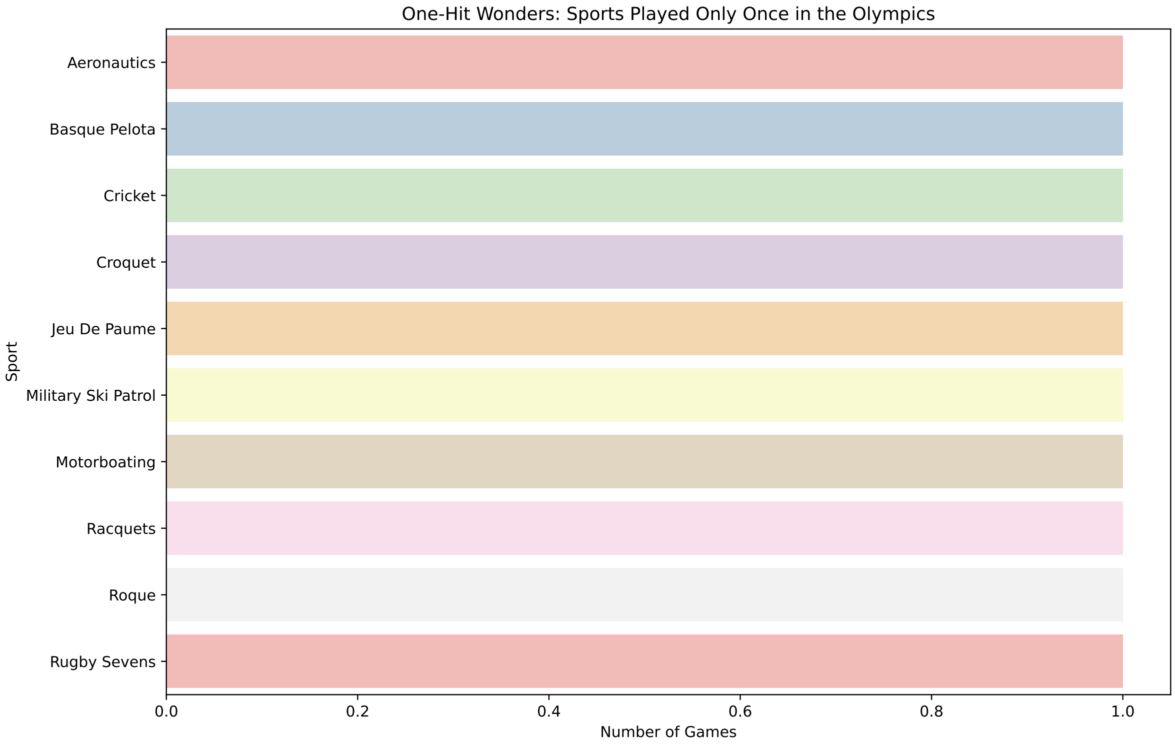


```
         Sport  Games
4    Athletics     29
14     Cycling     29
17     Fencing     29
21  Gymnastics     29
40    Swimming     29

/tmp/ipykernel_44969/1611472489.py:107: FutureWarning:



Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assig
```

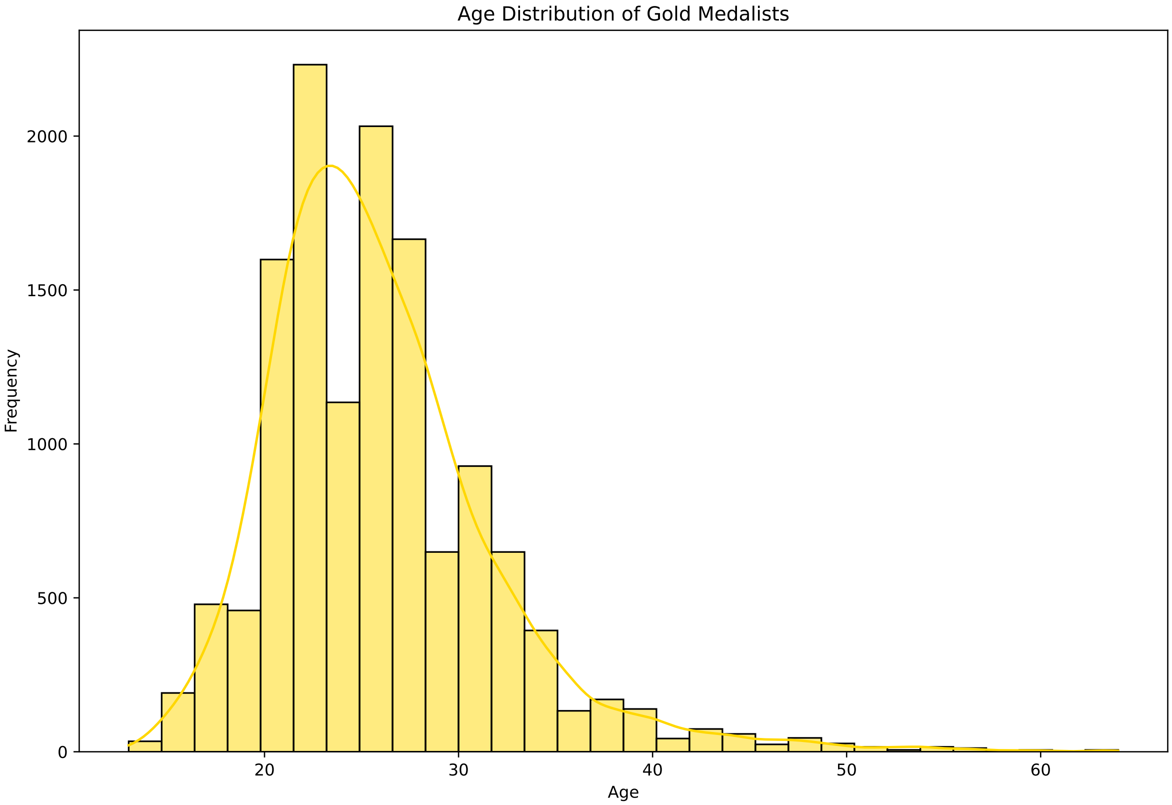


```
                Sport  Games
0         Aeronautics      1
9       Basque Pelota      1
15            Cricket      1
16            Croquet      1
31        Jeu De Paume      1
35  Military Ski Patrol      1
37        Motorboating      1
40            Racquets      1
42               Roque      1
45        Rugby Sevens      1
```

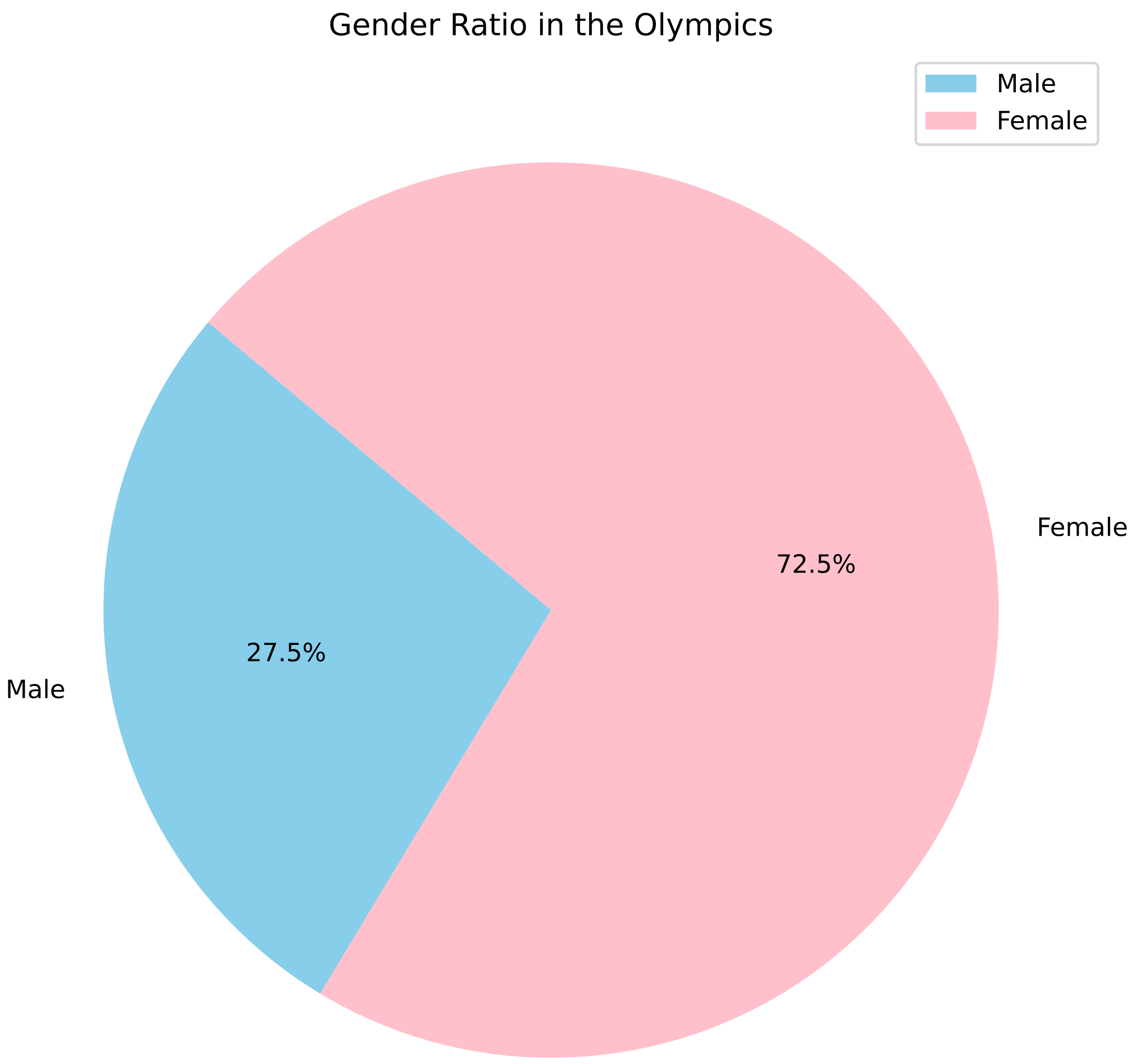
```
/tmp/ipykernel_44969/1611472489.py:120: FutureWarning:


Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assig
```



```
            ID               Name Sex   Age  Height  Weight    Team  NOC  \
233390  117046  Oscar Gomer Swahn   M  64.0     NaN     NaN  Sweden  SWE

              Games  Year  Season       City     Sport  \
233390  1912 Summer  1912  Summer  Stockholm  Shooting

                                                   Event Medal
233390  Shooting Men's Running Target, Single Shot, Team  Gold
```

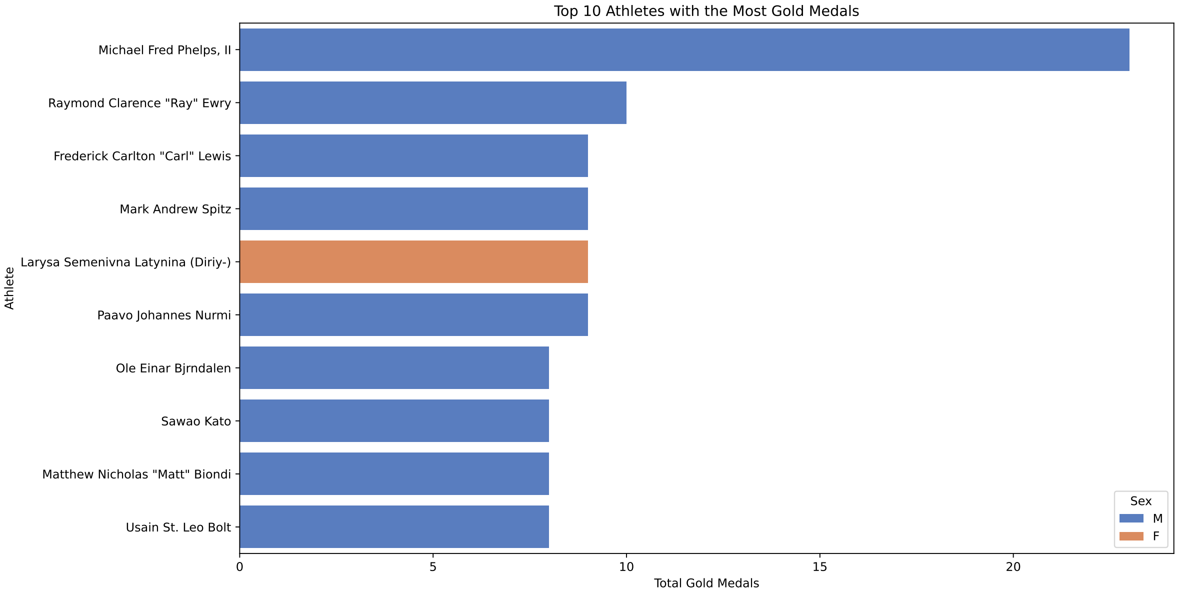

Age Distribution of Gold Medalists
Frequency
0
500
1000
1500
2000
Age
20
30
40
50
60

```
Ratio: 1:2.64
```



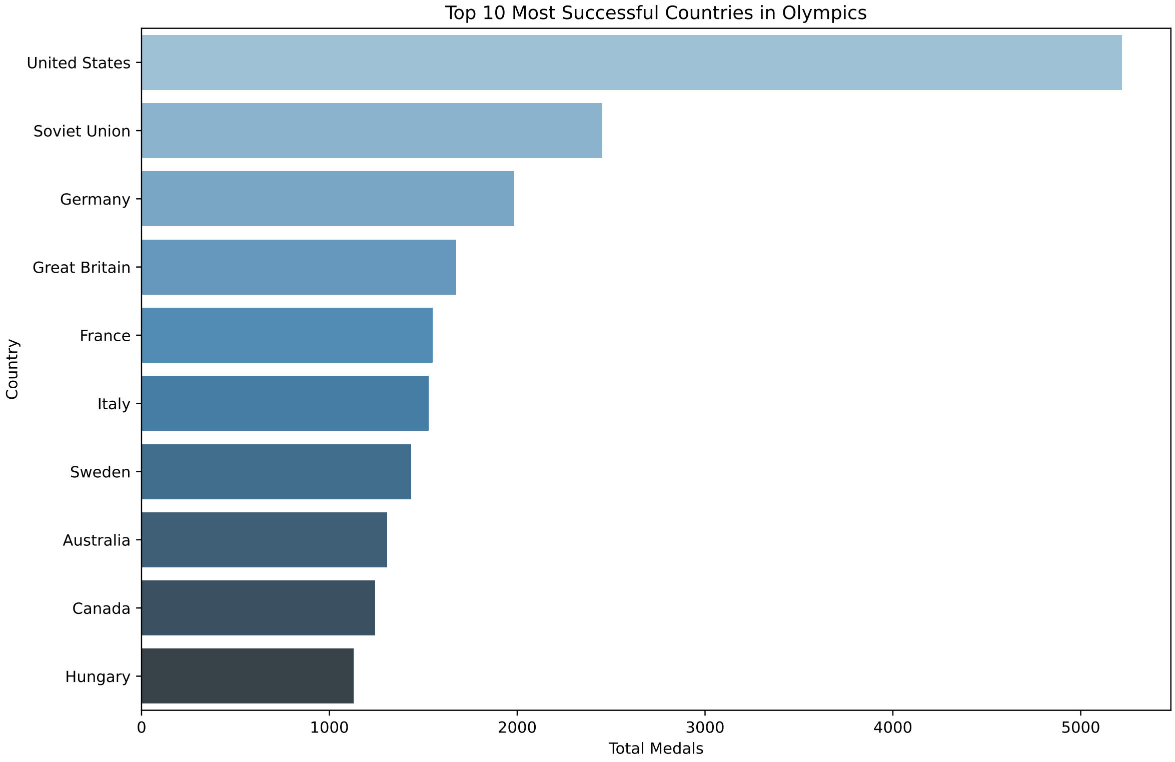
```
                                    Name           Team Sex  Total_Gold_Medals
6732             Michael Fred Phelps, II  United States   M                 23
7947         Raymond Clarence "Ray" Ewry  United States   M                 10
2977      Frederick Carlton "Carl" Lewis  United States   M                  9
6431                   Mark Andrew Spitz  United States   M                  9
5674  Larysa Semenivna Latynina (Diriy-)   Soviet Union   F                  9
7491                Paavo Johannes Nurmi        Finland   M                  9
7323                 Ole Einar Bjrndalen         Norway   M                  8
```

```
8497                    Sawao Kato          Japan   M              8
6585     Matthew Nicholas "Matt" Biondi  United States   M              8
9465                Usain St. Leo Bolt        Jamaica   M              8
```



```
               Team  Total_Medals
462   United States          5219
403    Soviet Union          2451
165         Germany          1984
171   Great Britain          1673
149          France          1550
215           Italy          1527
420          Sweden          1434
18        Australia          1306
67           Canada          1243
198         Hungary          1127
```

```
/tmp/ipykernel_44969/1611472489.py:168: FutureWarning:



Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assi
```


Top 10 Most Successful Countries in Olympics
Country
United States
Soviet Union
Germany
Great Britain
France
Italy
Sweden
Australia
Canada
Hungary
0
1000
2000
3000
4000
5000
Total Medals