Olympic Dataset Analysis for SportsStats

You are a data scientist working for a data analytics firm. Your firm has explored a multitude of data sources and is tasked with providing key insights that your clients can make actionable. Your manager has asked you to provide some data analytics guidance for one of the firm's clients.

In a typical scenario, you would iteratively work with your client to understand the data wanting to be analyzed. Having a solid understanding of the data and any underlying assumptions present is crucial to the success of a data analysis project. However, in this case, you will need to do a little more of the "heavy lifting".

To begin, you will prepare a project proposal detailing:

- The questions we are wanting to answer,
- initial hypothesis about the data relationships, and
- the approach you will take to get your answers.

NOTE: The proposal is just a plan for how we will travel. It's there to help keep you on your path by keeping the end goal in mind. You will then will execute your plan and in the end present your findings in a month to your management.

Step 1: Preparing for Your Proposal

You will document your preparation in developing the project proposal. This includes:

1. Which client/dataset did you select and why?
2. Describe the steps you took to import and clean the data.
3. Perform initial exploration of data and provide some screenshots or display some stats of the data you are looking at.
4. Create an ERD or proposed ERD to show the relationships of the data you are exploring.

1. Selecting the Ideal Dataset: Key Criteria and Justifications

Selected Dataset

SportsStats (Olympics Dataset - 120 years of data)

Why This Dataset Is Ideal for Aspiring Data Scientists

As an aspiring data scientist, I am seeking opportunities to practice and enhance my skills through real-world, complex datasets. The Olympics dataset offers a rich, multifaceted source of data that perfectly aligns with my goal of mastering data science. Below are the key reasons why this dataset is an excellent choice for honing a wide range of data science skills:

A. Comprehensive Dataset:

- Includes diverse attributes such as demographics, sports, events, medals, and regions, making it excellent for practicing skills ranging from data preprocessing to advanced analysis.

B. Big Data Characteristics:

- **Volume:** Spanning 120 years of data with potentially thousands of athletes, events, and results, this dataset qualifies as big data in scale.
- **Variety:** Contains multiple data types (numerical, categorical, geographical, and temporal), offering diverse challenges and insights.
- **Velocity:** While primarily historical, real-time Olympics data (e.g., live stats or medal counts) could be integrated for streaming data practice.
- **Veracity:** The dataset is well-structured and reliable, ensuring factual accuracy and minimal bias.

C. Real-World Relevance:

- Simulates real-world challenges, such as predicting medal counts, clustering athletes by performance, or analyzing trends in sports evolution.
- Insights derived from the dataset can influence areas like sports science, policymaking, and media, demonstrating practical applications to future employers.

D. Machine Learning Potential:

- Enables supervised learning tasks, such as predicting medal outcomes based on athlete metrics.
- Supports unsupervised learning tasks, such as clustering athletes by sport, performance, or region.
- Allows for advanced techniques like time-series analysis to explore performance trends over decades.

E. Data Visualization Practice:

- Offers rich categories (e.g., sports, countries, medals) ideal for creating impactful visualizations such as heatmaps, time-series plots, and interactive dashboards.

F. Portfolio Value:

- Projects using globally recognized events like the Olympics enhance a portfolio, demonstrating the ability to handle large, meaningful datasets.

G. Complexity for Skill Development:

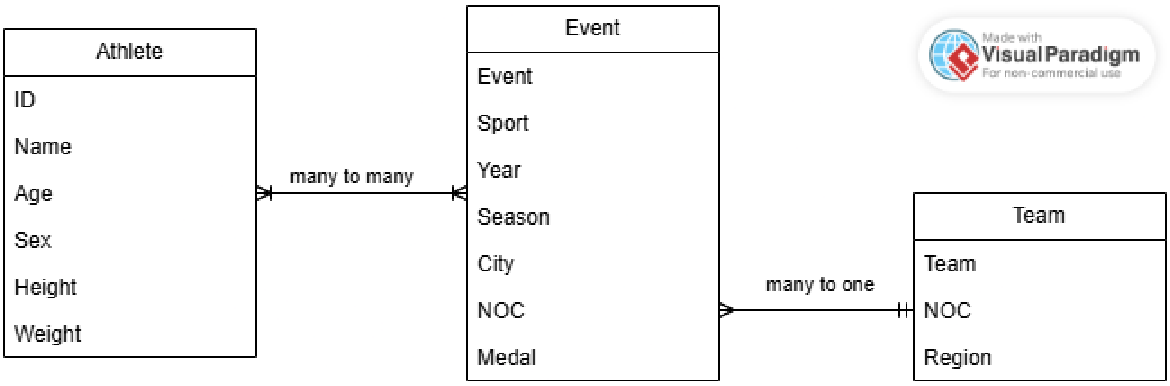
- Challenges include:
  - Handling missing values (e.g., incomplete athlete data for certain years or events).
  - Integrating external data (e.g., GDP, population, or sports investments) for deeper insights.
  - Scaling analysis to uncover patterns across regions, genders, or decades.

By leveraging this dataset, I aim to build a robust portfolio, enhance my technical and analytical skills, and gain hands-on experience with real-world data challenges. These projects will prepare me for a successful career in data science and demonstrate my ability to tackle complex datasets with meaningful results.

2. Proposed ERD

ERD Description:

1. Athlete (ID, Name, Age, Sex, Height, Weight)
2. Event (Event, Sport, Year, Season, City, NOC, Medal)
3. Team (Team, NOC, Region)



2. Data Import and Cleaning

Import Required Libraries

In [1]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import matplotlib.cm as cm
```

In [2]:
```
# Load datasets
athlete_data = pd.read_csv('athlete_events.csv')
noc_data = pd.read_csv('noc_regions.csv')
```

Display Dataset Information

This code displays a summary of the athlete_data and noc_data DataFrames, detailing the number of entries, columns, data types, and non-null counts. It aids in identifying missing values and understanding the datasets' structure for effective analysis.

In [3]:
```python
print("Athlete Data Info:\n")
athlete_data.info()
print("\nNOC Data Info:\n")
noc_data.info()
```

```
Athlete Data Info:

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 271116 entries, 0 to 271115
Data columns (total 15 columns):
 #   Column  Non-Null Count   Dtype  
---  ------  --------------   -----  
 0   ID      271116 non-null  int64  
 1   Name    271116 non-null  object 
 2   Sex     271116 non-null  object 
 3   Age     261642 non-null  float64
 4   Height  210945 non-null  float64
 5   Weight  208241 non-null  float64
 6   Team    271116 non-null  object 
 7   NOC     271116 non-null  object 
 8   Games   271116 non-null  object 
 9   Year    271116 non-null  int64  
 10  Season  271116 non-null  object 
 11  City    271116 non-null  object 
 12  Sport   271116 non-null  object 
 13  Event   271116 non-null  object 
 14  Medal   39783 non-null   object 
dtypes: float64(3), int64(2), object(10)
memory usage: 31.0+ MB

NOC Data Info:

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 230 entries, 0 to 229
Data columns (total 3 columns):
 #   Column  Non-Null Count  Dtype 
---  ------  --------------  ----- 
 0   NOC     230 non-null    object
 1   region  227 non-null    object
 2   notes   21 non-null     object
dtypes: object(3)
memory usage: 5.5+ KB
```

Preview the data

In [4]:
```
print("Athlete Data Preview:")
athlete_data.head(5)
```

```
Athlete Data Preview:
```

Out[4]:

	ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	
0	1	A Dijiang	M	24.0	180.0	80.0	China	CHN	1992 Summer	1992	Summer	Barcelona	Ba
1	2	A Lamusi	M	23.0	170.0	60.0	China	CHN	2012 Summer	2012	Summer	London	
2	3	Gunnar Nielsen Aaby	M	24.0	NaN	NaN	Denmark	DEN	1920 Summer	1920	Summer	Antwerpen	
3	4	Edgar Lindenau Aabye	M	34.0	NaN	NaN	Denmark/Sweden	DEN	1900 Summer	1900	Summer	Paris	
4	5	Christine Jacoba Aaftink	F	21.0	185.0	82.0	Netherlands	NED	1988 Winter	1988	Winter	Calgary	

In [5]:
```
print("NOC Data Preview:")
noc_data.head()
```

```
NOC Data Preview:
```

Out[5]:

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

Step 2: Data Cleaning

1. Visualizing Missing Data in the Athlete Dataset

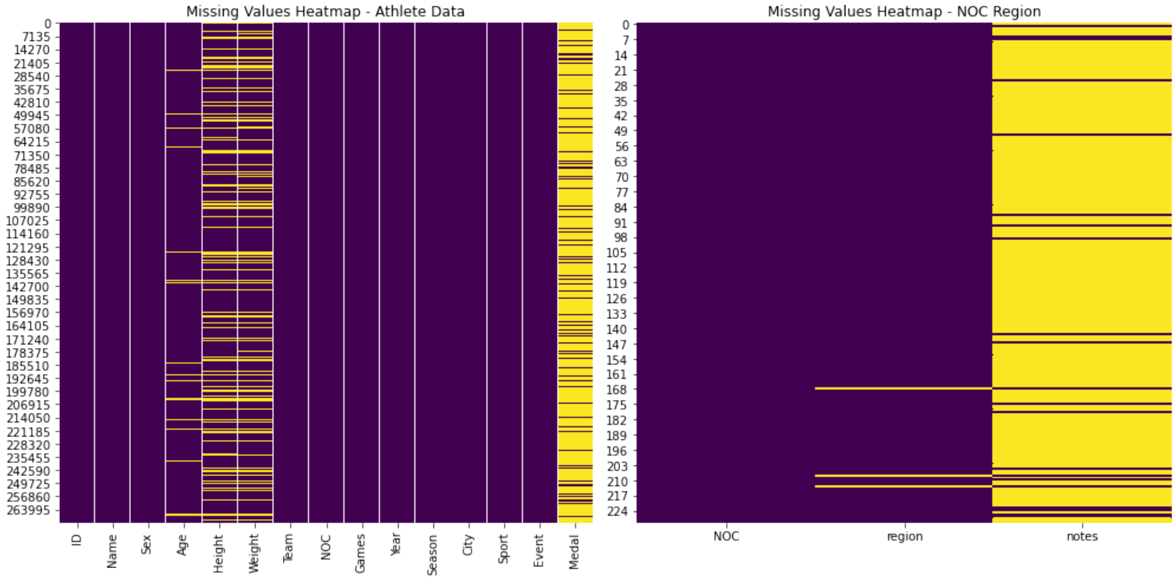
In [6]:
```
print("\nMissing values in Athlete Data:")
missing_athlete = athlete_data.isnull().sum()
missing_athlete
```

```
Missing values in Athlete Data:
```

Out[6]:
```
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
dtype: int64
```

Missing Data: Heatmap Visualization

In [7]:
```
# Missing values heatmap
# Create a subplot to show both heatmaps side by side
fig, axes = plt.subplots(1, 2, figsize=(14, 7))

# Missing values heatmap for athlete_data
sns.heatmap(athlete_data.isnull(), cbar=False, cmap='viridis', ax=axes[0])
axes[0].set_title("Missing Values Heatmap - Athlete Data")

# Missing values heatmap for noc_region
sns.heatmap(noc_data.isnull(), cbar=False, cmap='viridis', ax=axes[1])
axes[1].set_title("Missing Values Heatmap - NOC Region")

# Show the heatmaps
plt.tight_layout()
plt.show()
```



2. Imputing Missing Values for Height and Weight

Handling Missing Height and Weight Data: Approach and Rationale

To handle missing data for height and weight:

1. **Multiple Entries**: If an athlete has multiple records, the average value is calculated to ensure consistency.
2. **Single Entry**: If only one record exists, that data is used directly.
3. **No Records**: If no data is available for the athlete, the average for the athlete's sport event is used.
4. **Final Fallback**: If no sport-level data is available, the global average is used as a fallback.

This strategy ensures completeness and consistency while minimizing the impact of missing values.

Why Averaging Is Used for Height and Weight Imputation

Averaging is effective for height and weight imputation because these attributes typically remain consistent over time. In cases where data is missing, using the average helps maintain data integrity without introducing significant errors. By retaining height and weight, even if imputed, the athlete's record stays complete and allows for meaningful comparisons and analysis across athletes in the same sport.

```python
In [8]: # Impute missing height and weight based on 'Name' and group by sex
athlete_data['Height'] = athlete_data.groupby(['Name', 'Sex'])['Height'].transform(lambda x: x.fillna(round(x.mean() if len(x) > 1 else x.iloc[0], 1)))
athlete_data['Weight'] = athlete_data.groupby(['Name', 'Sex'])['Weight'].transform(lambda x: x.fillna(round(x.mean() if len(x) > 1 else x.iloc[0], 1)))

# Impute missing values using the average for the sport-event and group by sex (if possible)
athlete_data['Height'] = athlete_data.groupby(['Event', 'Sex'])['Height'].transform(lambda x: x.fillna(round(x.mean(), 1)))
athlete_data['Weight'] = athlete_data.groupby(['Event', 'Sex'])['Weight'].transform(lambda x: x.fillna(round(x.mean(), 1)))

# Final fallback: Impute any remaining missing values with the global average, still grouped by sex
athlete_data['Height'] = athlete_data.groupby('Sex')['Height'].transform(lambda x: x.fillna(round(x.mean(), 1)))
athlete_data['Weight'] = athlete_data.groupby('Sex')['Weight'].transform(lambda x: x.fillna(round(x.mean(), 1)))
```

3. Imputing Missing Values for Age

Rationale for Addressing Missing Age Data

Calculate the athlete's age if their name appears more than once in the data. Check if their age is listed in a different entry with a different year, and then calculate their age based on the year difference. If no age information is found, remove their data.

Why Age Data Is Treated Differently

Age is a more critical and fixed piece of information. If it's missing or inconsistent, it can lead to inaccuracies in analysis, such as skewing performance trends over age groups. Age can be calculated using other records for the same athlete, but if it remains missing after attempting to calculate it, dropping the row helps maintain the integrity of the dataset. Age is fundamental for categorizing athletes by age groups or assessing performance over time, so ensuring that this data is accurate or removing invalid entries is crucial.

```python
In [9]: # Impute missing Age using available records for the same athlete
# If age is missing for a record, calculate it using the event year difference and known age of the athlete

def calculate_age(row):
    if pd.isna(row['Age']):
        same_athlete = athlete_data[athlete_data['ID'] == row['ID']]
        known_age = same_athlete.loc[same_athlete['Age'].notna(), 'Age']
        known_year = same_athlete.loc[same_athlete['Age'].notna(), 'Year']
        if not known_age.empty:
            year_difference = row['Year'] - known_year.iloc[0]
            calculated_age = known_age.iloc[0] + year_difference
            return calculated_age
    return row['Age']

athlete_data['Age'] = athlete_data.apply(calculate_age, axis=1)

# Drop rows where Age is still missing as it is critical
athlete_data = athlete_data.dropna(subset=['Age'])

# Convert Age to integer to remove decimals (e.g., 25.0 -> 25)
athlete_data['Age'] = athlete_data['Age'].astype(int)
```

4. Merging the Athlete and NOC Datasets

Resolving Team/NOC Inconsistencies During Merge

To ensure accurate and meaningful analysis of the Olympics Dataset, resolving inconsistencies between the `Team` and `NOC` (National Olympic Committee) fields is critical. These inconsistencies can arise due to variations in naming conventions, geopolitical changes, or data entry errors. By merging the dataset with a reliable and standardized NOC reference dataset, we aim to:

1. **Improve Data Integrity**: Eliminate discrepancies, ensuring consistent representation of countries across records.
2. **Ensure Accuracy**: Prevent errors in calculations, such as medal counts or performance trends by country.
3. **Preserve Historical Continuity**: Standardize representations across time, e.g., linking the Soviet Union (URS) with Russia (RUS).
4. **Enable Advanced Analysis**: Ensure consistent data for clustering, predictions, or trend analysis.
5. **Enhance Visualization**: Avoid confusing duplicates or misrepresentations in graphs or maps.

Additionally, we handle missing `Region` data by replacing it with the corresponding `Team` value, ensuring that no records are left incomplete.

Implementation

In [10]:
```python
# Merge with NOC data to resolve Team/NOC inconsistencies
athlete_data = pd.merge(athlete_data, noc_data, how='left', on='NOC')

# Replace missing 'Region' values with 'Teams' data from athlete_data
athlete_data['region'] = athlete_data['region'].fillna(athlete_data['Team'])
```

Example of an Issue Resolved:

<u>Before merging:</u>

- **Team A**: Soviet Union, **NOC**: URS, **Region**: NaN
- **Team B**: Russia, **NOC**: RUS

<u>After merging:</u>

- **Team A & B**: Russia (including historical USSR data), **NOC**: RUS

If `Region` remains missing after the merge, it will default to the `Team` value, ensuring no gaps in the dataset.

Final Outcome:

By resolving inconsistencies and filling missing values, the dataset becomes more robust, enabling:

- Accurate analyses of medal counts and trends.
- Fair representation in visualizations.
- Enhanced insights for stakeholders.

5. Removing Duplicates

Introduction

The code processes athlete data, which often contains duplicate rows for the same athlete in the same year due to participation in multiple events. Removing duplicates is crucial to create a cleaner dataset and enable meaningful analysis.

Why Remove Duplicates?

- **Simplified Analysis:** Aggregating data by athlete and year provides a concise summary of performance.
- **Accuracy:** Prevents double-counting of medals or events, avoiding skewed results.
- **Efficiency:** Reduces dataset size, improving performance and clarity.

Key Steps in the Code

- **Add Medal Count Columns:** New columns (Gold, Silver, Bronze) categorize medal types for each row.
- **Group and Aggregate Data:** The dataset is grouped by ID and Year, summarizing key metrics (e.g., summing medals, averaging Height/Weight, taking the first occurrence of text fields).
- **Identify Most Frequent Event:** The most_frequent_event function finds the event with the most medals or participation for each athlete-year.

In [11]:

```python
def most_frequent_event(df):
    # Check if the athlete has any gold medals
    if df['Gold'].sum() > 0:
        # If they won gold, select the event with the most gold medals
        event = df[df['Gold'] == 1]['Event'].value_counts().idxmax()
    elif df['Silver'].sum() > 0:
        # If no gold, select the event with the most silver medals
        event = df[df['Silver'] == 1]['Event'].value_counts().idxmax()
    elif df['Bronze'].sum() > 0:
        # If no gold or silver, select the event with the most bronze medals
        event = df[df['Bronze'] == 1]['Event'].value_counts().idxmax()
    else:
        # If no medals, select the event they participated in most frequently
        event = df['Event'].value_counts().idxmax()
    return event

# Create new columns for Gold, Silver, and Bronze medal counts
athlete_data['Gold'] = athlete_data['Medal'].apply(lambda x: 1 if x == 'Gold' else 0)
athlete_data['Silver'] = athlete_data['Medal'].apply(lambda x: 1 if x == 'Silver' else 0)
athlete_data['Bronze'] = athlete_data['Medal'].apply(lambda x: 1 if x == 'Bronze' else 0)

# Apply the function to groupby
athlete_data_rem_dupli = athlete_data.groupby(['ID', 'Year'], as_index=False).agg({
    'Name': 'first',
    'Sex': 'first',
    'Age': 'first',
    'Height': 'mean',
    'Weight': 'mean',
    'Team': 'first',
    'NOC': 'first',
    'Games': 'first',
    'Season': 'first',
    'City': 'first',
    'Sport': 'first',
    'Medal': 'count',  # Count the total medals won (already split into Gold, Silver, Bronze)
    'Gold': 'sum',  # Sum the Gold medals
    'Silver': 'sum',  # Sum the Silver medals
    'Bronze': 'sum',  # Sum the Bronze medals
    'region': 'first',
    'notes': 'first',
})

# Apply the most_frequent_event function to find the most participated or most-medaled event
athlete_data_rem_dupli['Most_Participated_Event'] = athlete_data.groupby(['ID', 'Year']).apply(most_frequent_event).values

# Show the cleaned dataset
print(athlete_data_rem_dupli)
```

```
            ID  Year                      Name Sex  Age  Height  Weight  \
0            1  1992                 A Dijiang   M   24   180.0    80.0
1            2  2012                  A Lamusi   M   23   170.0    60.0
2            3  1920       Gunnar Nielsen Aaby   M   24   177.5    73.1
3            4  1900      Edgar Lindenau Aabye   M   34   182.5    95.6
4            5  1988  Christine Jacoba Aaftink   F   21   185.0    82.0
...        ...   ...                       ...  ..  ...     ...     ...
180680  135568  2016    Olga Igorevna Zyuzkova   F   33   171.0    69.0
180681  135569  1976                Andrzej ya   M   29   179.0    89.0
180682  135570  2014                  Piotr ya   M   27   176.0    59.0
180683  135571  1998        Tomasz Ireneusz ya   M   30   185.0    96.0
180684  135571  2002        Tomasz Ireneusz ya   M   34   185.0    96.0

                  Team  NOC        Games  Season            City  \
0                China  CHN  1992 Summer  Summer       Barcelona
1                China  CHN  2012 Summer  Summer          London
2              Denmark  DEN  1920 Summer  Summer       Antwerpen
3       Denmark/Sweden  DEN  1900 Summer  Summer           Paris
4          Netherlands  NED  1988 Winter  Winter         Calgary
...                ...  ...          ...     ...             ...
180680         Belarus  BLR  2016 Summer  Summer  Rio de Janeiro
180681        Poland-1  POL  1976 Winter  Winter       Innsbruck
180682          Poland  POL  2014 Winter  Winter           Sochi
180683          Poland  POL  1998 Winter  Winter          Nagano
180684          Poland  POL  2002 Winter  Winter  Salt Lake City

                 Sport  Medal  Gold  Silver  Bronze       region notes  \
0           Basketball      0     0       0       0        China   NaN
1                 Judo      0     0       0       0        China   NaN
2             Football      0     0       0       0      Denmark   NaN
3           Tug-Of-War      1     1       0       0      Denmark   NaN
4        Speed Skating      0     0       0       0  Netherlands   NaN
...                ...    ...   ...     ...     ...          ...   ...
180680      Basketball      0     0       0       0      Belarus   NaN
180681            Luge      0     0       0       0       Poland   NaN
180682     Ski Jumping      0     0       0       0       Poland   NaN
180683       Bobsleigh      0     0       0       0       Poland   NaN
180684       Bobsleigh      0     0       0       0       Poland   NaN

                   Most_Participated_Event
0              Basketball Men's Basketball
1             Judo Men's Extra-Lightweight
2                  Football Men's Football
3              Tug-Of-War Men's Tug-Of-War
4       Speed Skating Women's 1,000 metres
...                                    ...
180680       Basketball Women's Basketball
180681          Luge Mixed (Men)'s Doubles
180682  Ski Jumping Men's Large Hill, Team
180683                 Bobsleigh Men's Four
180684                 Bobsleigh Men's Four

[180685 rows x 20 columns]
```

6. Preview of Cleaned Data

In [12]:

```
# Preview cleaned data
print("\nCleaned Data Preview:")
athlete_data
```

```
Cleaned Data Preview:
```

Out[12]:

	ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	
0	1	A Dijiang	M	24	180.0	80.0	China	CHN	1992 Summer	1992	Summer	Bar
1	2	A Lamusi	M	23	170.0	60.0	China	CHN	2012 Summer	2012	Summer	L
2	3	Gunnar Nielsen Aaby	M	24	177.5	73.1	Denmark	DEN	1920 Summer	1920	Summer	Antw
3	4	Edgar Lindenau Aabye	M	34	182.5	95.6	Denmark/Sweden	DEN	1900 Summer	1900	Summer	
4	5	Christine Jacoba Aaftink	F	21	185.0	82.0	Netherlands	NED	1988 Winter	1988	Winter	C
...	...	...	...	...	...	...	...	...	...	...	...	
261637	135569	Andrzej ya	M	29	179.0	89.0	Poland-1	POL	1976 Winter	1976	Winter	Inn
261638	135570	Piotr ya	M	27	176.0	59.0	Poland	POL	2014 Winter	2014	Winter	
261639	135570	Piotr ya	M	27	176.0	59.0	Poland	POL	2014 Winter	2014	Winter	
261640	135571	Tomasz Ireneusz ya	M	30	185.0	96.0	Poland	POL	1998 Winter	1998	Winter	N
261641	135571	Tomasz Ireneusz ya	M	34	185.0	96.0	Poland	POL	2002 Winter	2002	Winter	Sa

261642 rows × 20 columns

In [13]:
```python
print("\nCleaned Data after Removing Duplicates Preview:")
athlete_data_rem_dupli
```

```
Cleaned Data after Removing Duplicates Preview:
```

Out[13]:

	ID	Year	Name	Sex	Age	Height	Weight	Team	NOC	Games	Season	
0	1	1992	A Dijiang	M	24	180.0	80.0	China	CHN	1992 Summer	Summer	Bai
1	2	2012	A Lamusi	M	23	170.0	60.0	China	CHN	2012 Summer	Summer	I
2	3	1920	Gunnar Nielsen Aaby	M	24	177.5	73.1	Denmark	DEN	1920 Summer	Summer	Ant
3	4	1900	Edgar Lindenau Aabye	M	34	182.5	95.6	Denmark/Sweden	DEN	1900 Summer	Summer	
4	5	1988	Christine Jacoba Aaftink	F	21	185.0	82.0	Netherlands	NED	1988 Winter	Winter	(
...	...	...	...	...	...	...	...	...	...	...	...	
180680	135568	2016	Olga Igorevna Zyuzkova	F	33	171.0	69.0	Belarus	BLR	2016 Summer	Summer	.
180681	135569	1976	Andrzej ya	M	29	179.0	89.0	Poland-1	POL	1976 Winter	Winter	Inr
180682	135570	2014	Piotr ya	M	27	176.0	59.0	Poland	POL	2014 Winter	Winter	
180683	135571	1998	Tomasz Ireneusz ya	M	30	185.0	96.0	Poland	POL	1998 Winter	Winter	N
180684	135571	2002	Tomasz Ireneusz ya	M	34	185.0	96.0	Poland	POL	2002 Winter	Winter	Sa

180685 rows × 20 columns

Missing Data After Cleaning: Heatmap Visualization

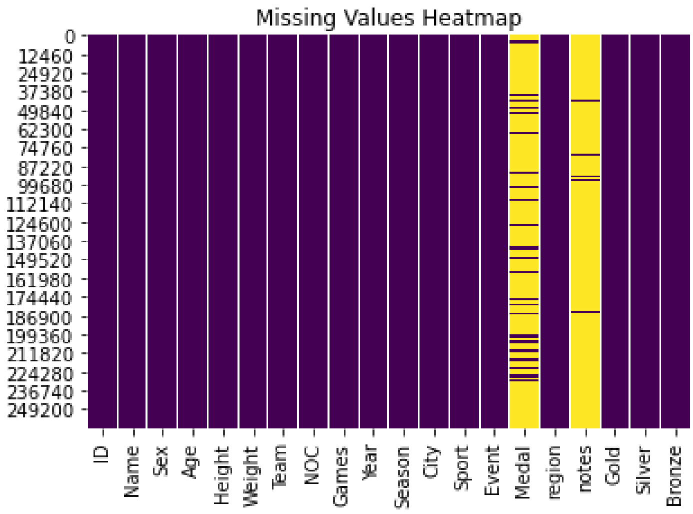
After cleaning the dataset, it's essential to check for any remaining important missing values to ensure data completeness and accuracy. A heatmap visualization provides a clear overview of where missing data still exists, highlighting potential gaps for further cleaning.

In [14]:
```python
# Check if any missing values are left
print("\nMissing values in Athlete Data:")
missing_athlete = athlete_data.isnull().sum()
missing_athlete
```

```
Missing values in Athlete Data:
```

Out[14]:
```
ID              0
Name            0
Sex             0
Age             0
Height          0
Weight          0
Team            0
NOC             0
Games           0
Year            0
Season          0
City            0
Sport           0
Event           0
Medal      222591
region          0
notes      256868
Gold            0
Silver          0
Bronze          0
dtype: int64
```

In [15]:
```python
# Missing values heatmap
sns.heatmap(athlete_data.isnull(), cbar=False, cmap='viridis')
plt.title("Missing Values Heatmap")
plt.show()
```



Step 3: Initial Exploration

Descriptive statistics

Descriptive Statistics Overview

In this section, we will calculate the descriptive statistics for the Age, Height, and Weight columns of the athlete data. Descriptive statistics are essential for summarizing the key features of the data, allowing us to understand the distribution, central tendencies (such as mean and median), and variability (such as standard deviation and range) of these attributes.

- **Age:** Represents the age of athletes.
- **Height:** Represents the height of athletes in centimeters.
- **Weight:** Represents the weight of athletes in kilograms.

The descriptive statistics include several important metrics:

- **Mean:** The average value of each attribute.
- **Standard Deviation (std):** Indicates how spread out the values are around the mean. A higher standard deviation suggests more variability, while a lower value indicates less.
- **Percentiles (25%, 50%, 75%):** These provide a breakdown of the data distribution:
  - **25% (Q1):** The value below which 25% of the data falls.
  - **50% (Median):** The middle value, separating the data into two equal halves.
  - **75% (Q3):** The value below which 75% of the data falls.

By examining these statistics, we can gain valuable insights into the general demographics and physical characteristics of athletes across the dataset. This information helps us identify trends and potential outliers, and it sets the stage for deeper analysis or modeling.

In [16]:
```
# Print the descriptive statistics for Age, Height, and Weight grouped by Sex
print("\nDescriptive Statistics:")
athlete_data_describe = athlete_data_rem_dupli.groupby('Sex')[['Age', 'Height', 'Weig
ht']].describe().T
print(athlete_data_describe)
```

```
Descriptive Statistics:
Sex                      F              M
Age    count  48535.000000  132150.000000
       mean      24.414114      26.282611
       std        5.599307       6.193119
       min       11.000000      10.000000
       25%       21.000000      22.000000
       50%       24.000000      25.000000
       75%       28.000000      29.000000
       max       74.000000      97.000000
Height count  48535.000000  132150.000000
       mean     169.018062     179.452655
       std        8.284259       8.880971
       min      127.000000     127.000000
       25%      164.000000     174.000000
       50%      168.700000     179.000000
       75%      174.000000     185.000000
       max      213.000000     226.000000
Weight count  48535.000000  132150.000000
       mean      61.451543      76.897041
       std       10.256846      12.975896
       min       25.000000      28.000000
       25%       55.000000      69.000000
       50%       60.000000      75.400000
       75%       67.000000      84.000000
       max      167.000000     214.000000
```

Conclusion: Descriptive Statistics Analysis

The descriptive statistics provide a detailed understanding of the athletes' physical characteristics. Below are the key findings from the dataset:

1. Age:

- Female athletes have a **mean age** of **24.41** years, while male athletes have a mean age of **26.28** years.
- The **age range** for female athletes spans from **10** to **74** years, and for male athletes, it spans from **10** to **97** years.
- The **standard deviation** for age is **5.60** for females and **6.19** for males, indicating moderate variability in age across both sexes.
- The **percentiles** show that 25% of female athletes are **21** years or younger, and 25% of male athletes are **22** years or younger. The median age for females is **24**, while it's **25** for males.

2. Height:

- The **average height** for female athletes is **169.00** cm, and for male athletes, it is **179.45** cm.
- The **height range** for female athletes extends from **127** cm to **213** cm, while for male athletes, it spans from **127** cm to **226** cm.
- The **standard deviation** of height is **8.27** cm for females and **8.72** cm for males, showing that there's significant height variability within both sexes.
- The **percentiles** indicate that 25% of female athletes are **164** cm or shorter, and 25% of male athletes are **174** cm or shorter. The median height for females is **168.7** cm, and for males, it's **179.0** cm.

3. Weight:

- The **average weight** for female athletes is **61.45** kg, while for male athletes, it is **76.90** kg.
- The **weight range** for female athletes goes from **25** kg to **167** kg, and for male athletes, it extends from **28** kg to **214** kg.
- The **standard deviation** for weight is **10.26** kg for females and **12.96** kg for males, suggesting that weight distribution is more spread out for male athletes.
- The **percentiles** show that 25% of female athletes weigh **55** kg or less, and 25% of male athletes weigh **69.0** kg or less. The median weight for females is **60.0** kg, while for males, it is **75.40** kg.

Summary:

These descriptive statistics provide a comprehensive view of the physical characteristics of Olympic athletes. Key insights include:

- **Age**: Male athletes are, on average, older than female athletes, with both sexes displaying a wide age range.
- **Height**: There is a clear height difference between male and female athletes, with males being taller on average.
- **Weight**: Male athletes have a higher average weight, with a broader weight distribution compared to females.

These findings set the stage for deeper analysis or modeling, as they give us a clear understanding of how physical traits vary across different groups of athletes. The variability in age, height, and weight will be useful for further examination of performance, trends, and other factors that may influence athletic success.

Step 4: Hypotheses

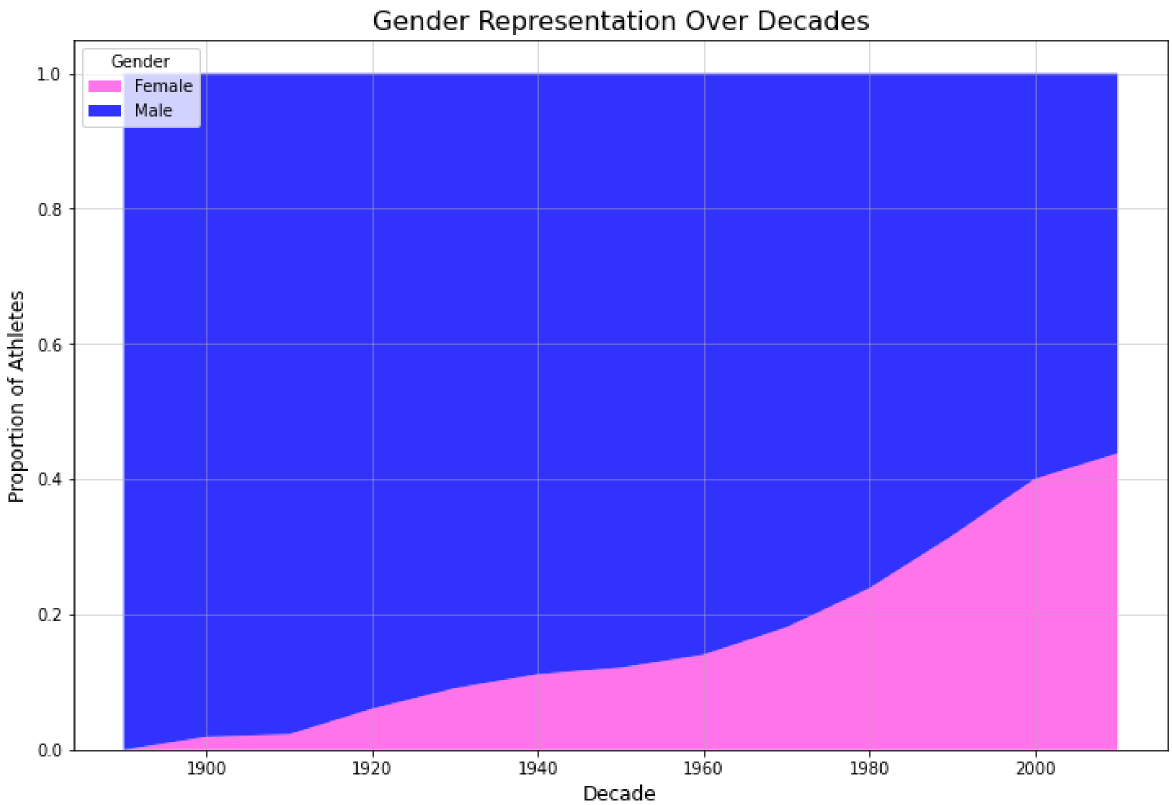
- **Hypothesis 1:** Gender representation has improved significantly in recent decades.
- **Hypothesis 2:** Athletes in their 20s are most likely to win Olympic medals, with distinct trends for males and females.
- **Hypothesis 3:** Certain sports or events are predominantly male or female, with a significant gender disparity in participation and achievement.
- **Hypothesis 4:** Athlete demographics (age, height, weight) have evolved over time due to advancements in sports science.
- **Hypothesis 5:** Countries with a better BMI (Body Mass Index) tend to win more medals, suggesting a possible correlation between physical size and athletic success.

Step 5: Data Analysis Approach

Hypothesis 1: Gender representation has improved significantly in recent decades.

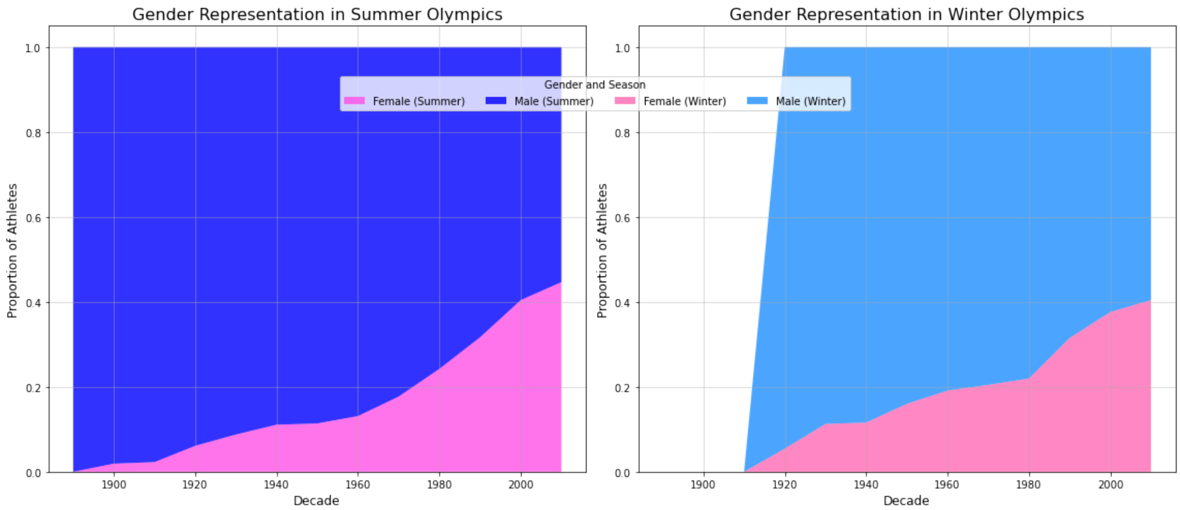
We hypothesize that in the earlier years, sports participation was predominantly male due to societal norms and limited opportunities for women. However, I believe that in recent decades, gender representation has become more balanced as efforts toward equality and inclusion in sports have gained momentum.

In [17]:
```
# Group athletes by decade and sex to calculate gender representation
def categorize_decade(year):
    return (year // 10) * 10

athlete_data_rem_dupli['Decade'] = athlete_data_rem_dupli['Year'].apply(categorize_de
cade)

# Count the number of athletes per decade and gender
gender_representation = athlete_data_rem_dupli.groupby(['Decade', 'Sex']).size().rese
t_index(name='Count')

# Calculate total athletes per decade
total_per_decade = gender_representation.groupby('Decade')['Count'].sum().reset_index
(name='Total')

# Merge total back to the gender data to calculate proportions
gender_representation = gender_representation.merge(total_per_decade, on='Decade')
gender_representation['Proportion'] = gender_representation['Count'] / gender_represe
ntation['Total']

# Pivot the data for a stacked area chart
pivot_data = gender_representation.pivot(index='Decade', columns='Sex', values='Propo
rtion').fillna(0)

# Plot a stacked area chart
plt.figure(figsize=(12, 8))
plt.stackplot(
    pivot_data.index,
    pivot_data['F'], pivot_data['M'],   # Male and Female proportions
    labels=['Female', 'Male'],
    colors=[ '#FF53E6', 'blue'],  # Custom colors for genders
    alpha=0.8
)

# Title and Labels
plt.title("Gender Representation Over Decades", fontsize=16)
plt.xlabel("Decade", fontsize=12)
plt.ylabel("Proportion of Athletes", fontsize=12)
plt.legend(title="Gender", loc='upper left', fontsize=10)
plt.grid(True, alpha=0.5)

# Show plot
plt.show()
```


Gender Representation Over Decades
Gender
Female
Male
Proportion of Athletes
0.0
0.2
0.4
0.6
0.8
1.0
1900
1920
1940
1960
1980
2000
Decade

In [18]:
```
# Group athletes by decade and sex to calculate gender representation
def categorize_decade(year):
    return (year // 10) * 10

athlete_data_rem_dupli['Decade'] = athlete_data_rem_dupli['Year'].apply(categorize_de
cade)

# Count the number of athletes per decade, gender, and season
gender_representation = athlete_data_rem_dupli.groupby(['Decade', 'Sex', 'Season']).s
ize().reset_index(name='Count')

# Calculate total athletes per decade and season
total_per_decade_season = gender_representation.groupby(['Decade', 'Season'])['Coun
t'].sum().reset_index(name='Total')

# Merge total back to the gender data to calculate proportions
gender_representation = gender_representation.merge(total_per_decade_season, on=['Dec
ade', 'Season'])
gender_representation['Proportion'] = gender_representation['Count'] / gender_represe
ntation['Total']

# Pivot the data for a stacked area chart (Separate for Winter and Summer)
pivot_data = gender_representation.pivot_table(index='Decade', columns=['Sex', 'Seaso
n'], values='Proportion', aggfunc='sum').fillna(0)

# Set up subplots for Winter and Summer
fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(16, 7))  # 1 row, 2 columns

# Colors for each combination of gender and season
colors = {
    ('F', 'Summer'): '#FF53E6',  # Female Winter
    ('M', 'Summer'): 'blue',     # Male Winter
    ('F', 'Winter'): '#FF69B4',  # Female Summer (lighter pink)
    ('M', 'Winter'): '#1E90FF'   # Male Summer (lighter blue)
}

# Stack plot for Winter (on the left)
ax1.stackplot(
    pivot_data.index,
    pivot_data[('F', 'Summer')], pivot_data[('M', 'Summer')],   # Female and Male pro
portions for Winter
    labels=['Female (Summer)', 'Male (Summer)'],
    colors=[colors[('F', 'Summer')], colors[('M', 'Summer')]],  # Use the defined col
ors
    alpha=0.8
)
ax1.set_title("Gender Representation in Summer Olympics", fontsize=16)
ax1.set_xlabel("Decade", fontsize=12)
ax1.set_ylabel("Proportion of Athletes", fontsize=12)
ax1.grid(True, alpha=0.5)

# Stack plot for Summer (on the right)
ax2.stackplot(
    pivot_data.index,
    pivot_data[('F', 'Winter')], pivot_data[('M', 'Winter')],   # Female and Male pro
portions for Summer
    labels=['Female (Winter)', 'Male (Winter)'],
    colors=[colors[('F', 'Winter')], colors[('M', 'Winter')]],  # Use the defined col
ors
    alpha=0.8
)
ax2.set_title("Gender Representation in Winter Olympics", fontsize=16)
ax2.set_xlabel("Decade", fontsize=12)
ax2.set_ylabel("Proportion of Athletes", fontsize=12)
ax2.grid(True, alpha=0.5)

# Add legends
```

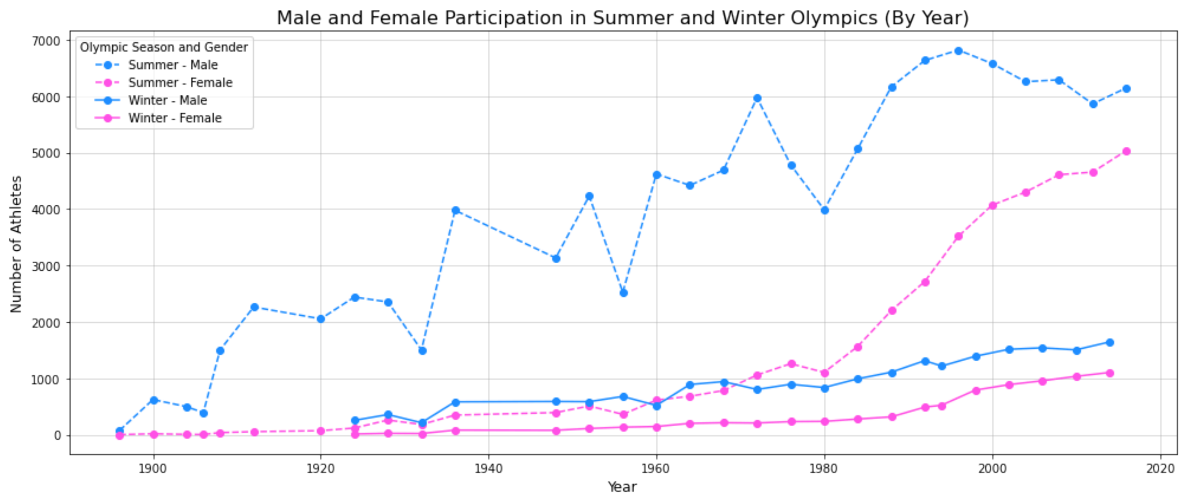
```
fig.legend(title="Gender and Season", bbox_to_anchor=(0.5, 0.85), loc='upper center',
ncol=4, fontsize=10)

# Show plot
plt.tight_layout()
plt.show()
```



Male and Female participation by year in Summer and Winter Olympics

In [19]:

```
# Filter data for Summer and Winter Olympics
summer_data = athlete_data[athlete_data['Season'] == 'Summer']
winter_data = athlete_data[athlete_data['Season'] == 'Winter']

# Group by Year and Sex to count unique athletes for Summer Olympics
summer_participation = summer_data.groupby(['Year', 'Sex'])['ID'].nunique().unstack(f
ill_value=0)

# Group by Year and Sex to count unique athletes for Winter Olympics
winter_participation = winter_data.groupby(['Year', 'Sex'])['ID'].nunique().unstack(f
ill_value=0)

# Plotting
plt.figure(figsize=(14, 6))

# Summer Olympics
plt.plot(summer_participation.index, summer_participation['M'], marker='o', label='Su
mmer - Male', color='#1E90FF', linestyle='--')
plt.plot(summer_participation.index, summer_participation['F'], marker='o', label='Su
mmer - Female', color='#FF53E6', linestyle='--')

# Winter Olympics
plt.plot(winter_participation.index, winter_participation['M'], marker='o', label='Wi
nter - Male', color='#1E90FF', linestyle='-')
plt.plot(winter_participation.index, winter_participation['F'], marker='o', label='Wi
nter - Female', color='#FF53E6', linestyle='-')

# Labels, title, and legend
plt.title('Male and Female Participation in Summer and Winter Olympics (By Year)', fo
ntsize=16)
plt.xlabel('Year', fontsize=12)
plt.ylabel('Number of Athletes', fontsize=12)
plt.legend(title='Olympic Season and Gender', fontsize=10)
plt.grid(alpha=0.5)
plt.tight_layout()

# Show plot
plt.show()
```



Conclusion: Gender Representation in the Olympics

The analysis of gender representation over the decades reveals significant progress in female participation in the Olympics. Below are the key findings:

1. **Early Exclusion of Women (1890s-1920s):**
    - In the 1890s, women accounted for **0%** of participants. By 1900, they represented just **2%** of the total (59 women compared to 3,009 men).
    - By the 1920s, female athletes made up only **6%** of the total participants, reflecting their limited involvement during the early years of the modern Olympics.

1. **Gradual Growth (1930s-1970s):**
    - From the 1930s to the 1970s, the percentage of female athletes steadily increased:
        - **9%** in the 1930s
        - **11%** in the 1940s
        - **12%** in the 1950s
        - **14%** in the 1960s
        - **18%** in the 1970s

1. **Rapid Increase (1980s-2010s):**
    - Female participation grew significantly during this period:
        - **24%** in the 1980s
        - **32%** in the 1990s
        - **40%** in the 2000s
        - **44%** in the 2010s

Final Insight:

The percentage of female athletes has increased from **0% in the 1890s to 44% by the 2010s**, showcasing a remarkable shift toward gender inclusivity in the Olympics. This growth reflects broader societal changes, advocacy for gender equality, and initiatives by international organizations to promote women's participation in sports. While substantial progress has been made, there is still work to be done to achieve full gender parity in all sports and events.

Hypothesis 2: Athletes in their 20s are most likely to win Olympic medals, with distinct trends for males and females.

Distribution of Medals by Age and Gender

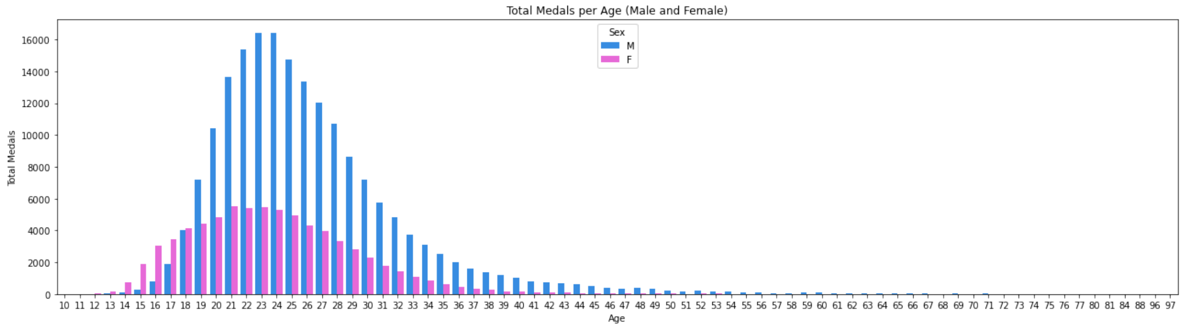
The Distribution of Medals by Age and Gender explores how athletes' achievements in various medals (Gold, Silver, Bronze) vary across different age groups and genders. This analysis provides insight into the patterns and trends of medal-winning performances, showing whether certain age groups or genders dominate particular medal categories.

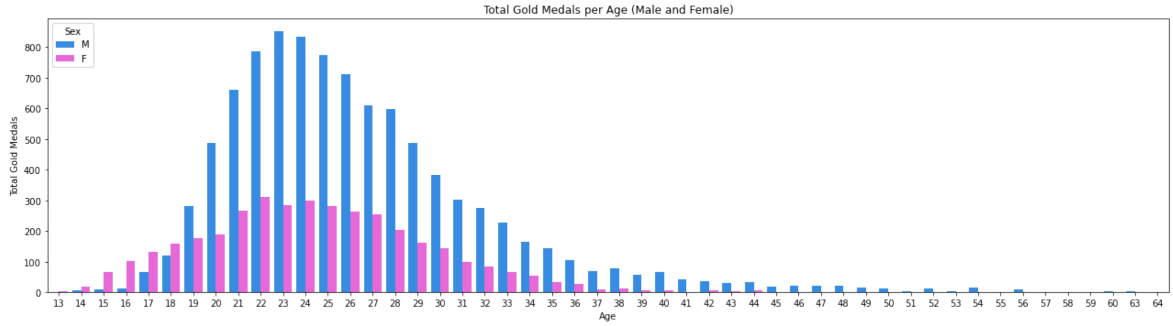
The visualization is broken down into four distinct charts:

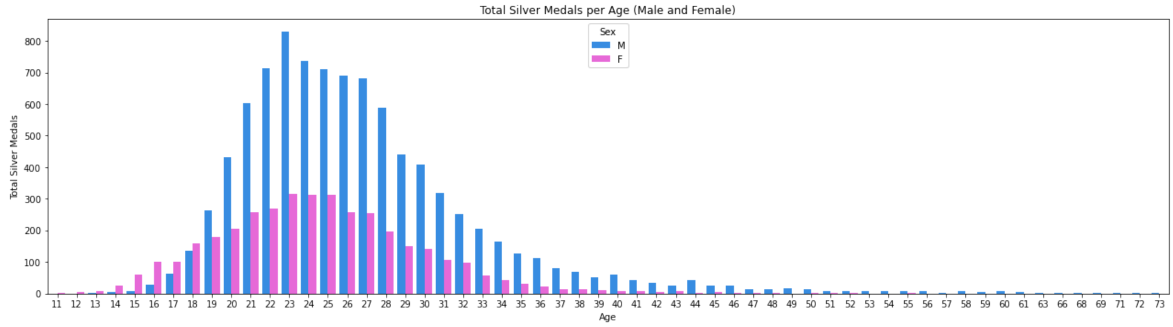
- **Total Medals per Age** – This chart highlights the overall distribution of medals won by athletes across various age groups, segmented by gender.
- **Total Gold Medals per Age** – This chart focuses specifically on the distribution of Gold medals by age, offering a glimpse into the ages at which athletes are most likely to win the top honors.
- **Total Silver Medals per Age** – Here, the analysis shifts to Silver medals, helping identify the age trends of athletes who just missed the Gold.
- **Total Bronze Medals per Age** – Finally, this chart examines the distribution of Bronze medals, showcasing the age groups where athletes often finish in third place.

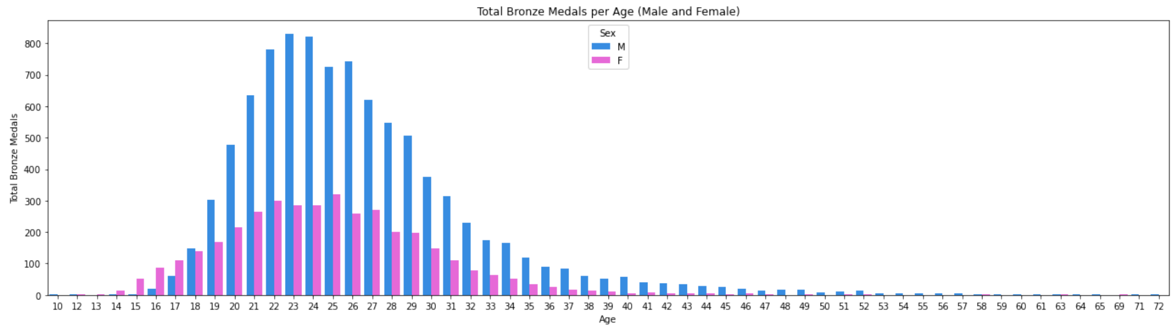
By comparing male and female athletes within each of these charts, we can uncover interesting trends and differences in medal distribution, helping us understand the role of age and gender in athletic performance.

In [20]:
```
# Color map for male and female
color_map = {'F': '#FF53E6', 'M': '#1E90FF'}

# Create a figure with 4 subplots (since you need 4 charts, not 5)
fig, axes = plt.subplots(4, 1, figsize=(18, 20))

# Total medals per age for both male and female
sns.countplot(x='Age', hue='Sex', data=athlete_data, ax=axes[0], palette=color_map)
axes[0].set_title('Total Medals per Age (Male and Female)')
axes[0].set_xlabel('Age')
axes[0].set_ylabel('Total Medals')

# Total Gold medals per age for both male and female
sns.countplot(x='Age', hue='Sex', data=athlete_data[athlete_data['Medal'] == 'Gold'],
ax=axes[1], palette=color_map)
axes[1].set_title('Total Gold Medals per Age (Male and Female)')
axes[1].set_xlabel('Age')
axes[1].set_ylabel('Total Gold Medals')

# Total Silver medals per age for both male and female
sns.countplot(x='Age', hue='Sex', data=athlete_data[athlete_data['Medal'] == 'Silve
r'], ax=axes[2], palette=color_map)
axes[2].set_title('Total Silver Medals per Age (Male and Female)')
axes[2].set_xlabel('Age')
axes[2].set_ylabel('Total Silver Medals')

# Total Bronze medals per age for both male and female
sns.countplot(x='Age', hue='Sex', data=athlete_data[athlete_data['Medal'] == 'Bronz
e'], ax=axes[3], palette=color_map)
axes[3].set_title('Total Bronze Medals per Age (Male and Female)')
axes[3].set_xlabel('Age')
axes[3].set_ylabel('Total Bronze Medals')

# Adjust layout for better spacing
plt.tight_layout()

# Show the plot
plt.show()
```









Conclusion:

The analysis supports Hypothesis 2, highlighting that athletes in their 20s are most likely to win Olympic medals, with gender-specific trends emerging. For male athletes, the peak in gold medal achievements occurs in their late 20s, particularly between ages 22 and 23, where medal counts are at their highest. Female athletes, on the other hand, show their strongest performance between ages 19 and 22, with a noticeable decrease in medal counts thereafter.

Additionally, the data indicates a consistent decline in the likelihood of winning medals as athletes age, particularly after their mid-30s. Male athletes generally experience a later peak in performance, reaching their highest medal counts between ages 22 and 25, while female athletes see a more pronounced decline starting at age 25.

In conclusion, Olympic medalists in their 20s—especially males—show a clear dominance in terms of medal counts. Female athletes, while also peaking in their early 20s, exhibit a steeper decline in medal success beyond this age range. These findings support the hypothesis that younger athletes, especially those in their 20s, have the highest probability of winning Olympic medals.

Hypothesis 3: Gender Disparity in Sports Participation and Achievement

In this analysis, we explore the hypothesis that certain sports or events are predominantly male or female, with a significant gender disparity in participation and achievement. This disparity is believed to be influenced by a combination of historical, cultural, and societal factors. Traditionally, some sports have been perceived as more suited to one gender over another, leading to an imbalance in the number of athletes participating in various events.

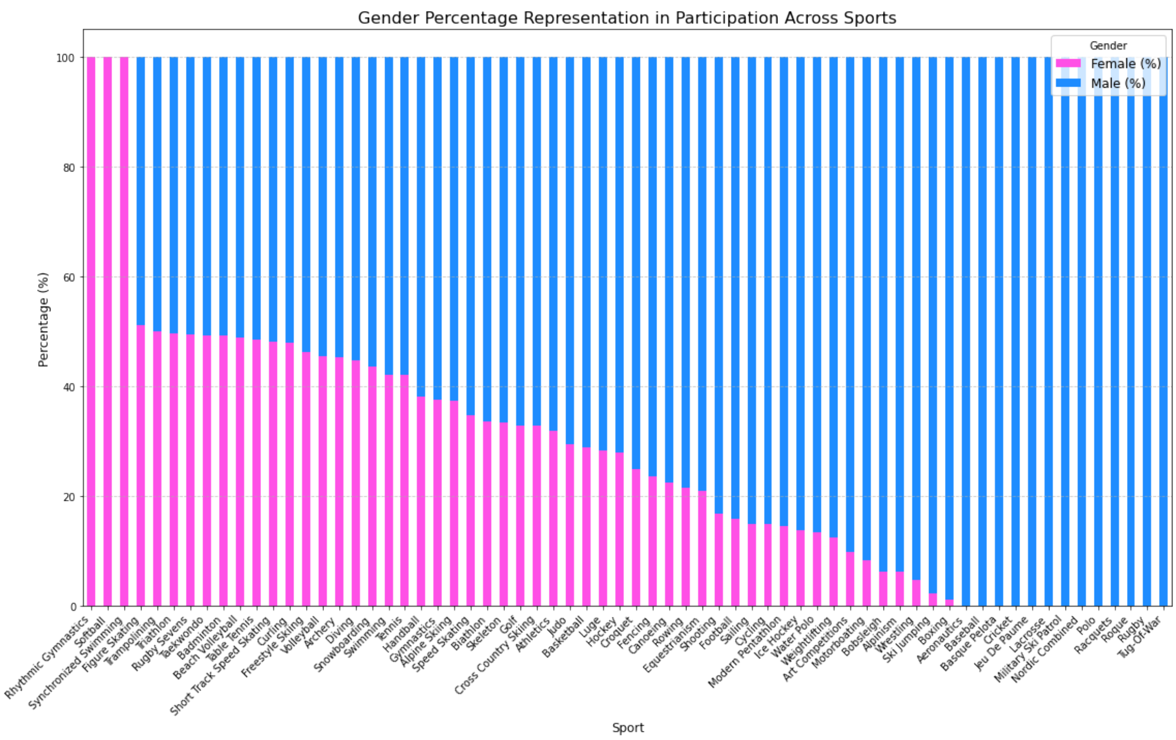
To test this hypothesis, we will examine the data of Olympic sports and their gender representation. We will investigate patterns where certain events have a skewed gender representation, while others may demonstrate more balanced participation. This analysis will also consider the historical and societal influences that have shaped these trends over time.

Gender Percentage Representation in Participation Across Sports

In [21]:

```
# Group by Sport and Sex, then count the number of participants
gender_disparity_by_sport = athlete_data_rem_dupli.groupby(['Sport', 'Sex']).size().r
eset_index(name='Count')

# Pivot the data to make 'Male' and 'Female' columns
pivot_disparity = gender_disparity_by_sport.pivot(index='Sport', columns='Sex', value
s='Count').fillna(0)

# Calculate percentage representation for each gender
pivot_disparity['Total'] = pivot_disparity.sum(axis=1)
pivot_disparity['Female (%)'] = (pivot_disparity['F'] / pivot_disparity['Total']) * 1
00
pivot_disparity['Male (%)'] = (pivot_disparity['M'] / pivot_disparity['Total']) * 100

# Keep only the percentage columns
percentage_disparity = pivot_disparity[['Female (%)', 'Male (%)']]

# Sort sports by total participation for better visualization
percentage_disparity = percentage_disparity.sort_values(by=['Female (%)', 'Male
(%)'], ascending=False)

# Create the plot
plt.figure(figsize=(16, 10))

# Plot the percentage as a stacked bar chart
percentage_disparity.plot(kind='bar', stacked=True, figsize=(16, 10), color=['#FF53E
6', '#1E90FF'])

# Add title and labels
plt.title("Gender Percentage Representation in Participation Across Sports", fontsize
=16)
plt.xlabel("Sport", fontsize=12)
plt.ylabel("Percentage (%)", fontsize=12)
plt.xticks(rotation=45, ha='right', fontsize=10)
plt.legend(title="Gender", labels=['Female (%)', 'Male (%)'], fontsize=12)

# Add grid for better readability
plt.grid(axis='y', linestyle='--', alpha=0.7)

# Show the plot
plt.tight_layout()
plt.show()
```

```
<Figure size 1152x720 with 0 Axes>
```



Conclusion: Gender Disparity in Sports Participation and Achievement

The data highlights clear patterns of gender disparity in Olympic sports, with certain sports being predominantly male or female. The gender participation breakdown shows that historical and societal factors heavily influence the distribution of athletes across different events.

- **Predominantly Female Sports:**
  Sports like **Rhythmic Gymnastics**, **Softball**, and **Synchronized Swimming** have **100% female** participation, indicating that these events are traditionally and culturally viewed as more suitable for women. This reflects societal expectations and the historical exclusion of women from other sports.

- **Predominantly Male Sports:**
  Conversely, sports like **Baseball**, **Wrestling**, and **Motorboating** have **100% male** participation, showcasing a deep-rooted gender bias and the cultural perception that these activities are suited primarily for men. This disparity also mirrors historical trends where women were often excluded or discouraged from participating in such sports.

- **Mixed Participation Sports:**
  In many sports, gender representation is much more balanced. For instance, **Trampolining**, **Triathlon**, and **Rugby Sevens** show near-even male and female participation, reflecting a more equitable distribution in these events.

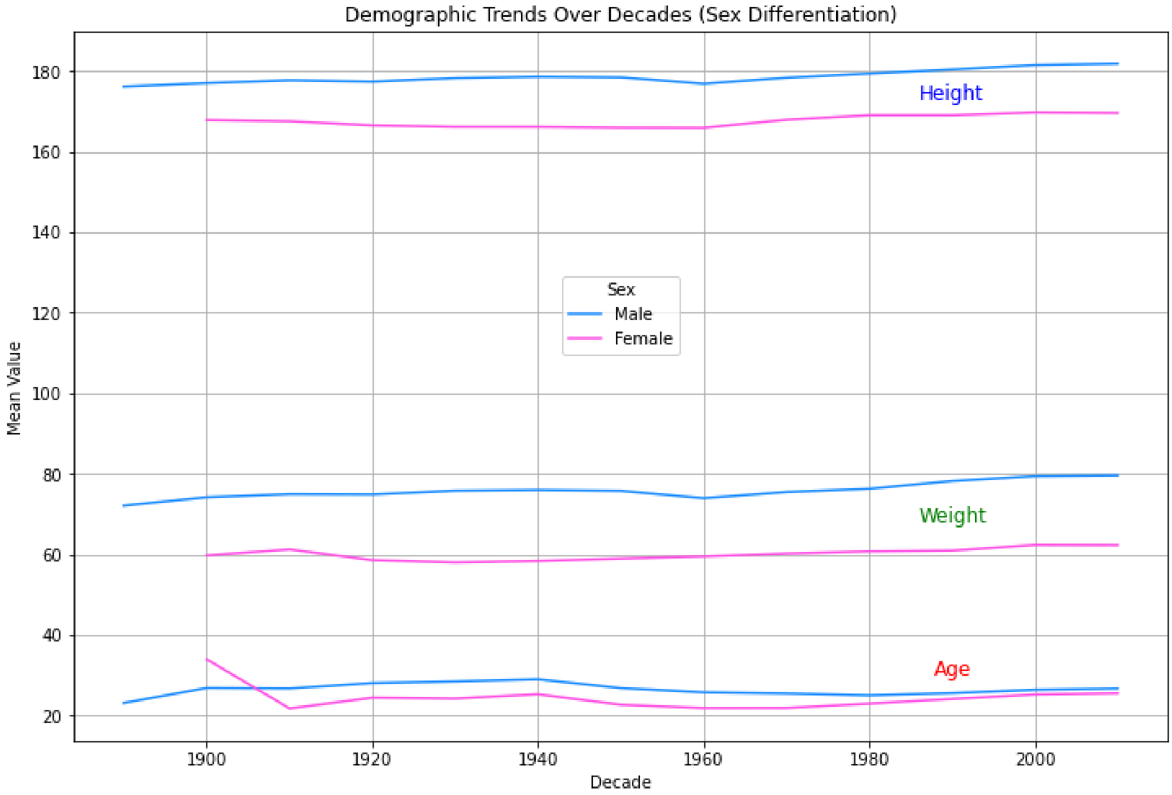
- **Sports with Significant Male Dominance:**
  Some sports, such as **Football**, **Weightlifting**, and **Boxing**, show a **dominant male presence**, with women representing a very small percentage. These disparities are influenced by cultural and historical factors, where male athletes have traditionally had more opportunities and recognition.

Overall, the gender disparities across Olympic sports are the result of long-standing societal beliefs, historical barriers to female participation, and the evolution of gender roles in sports. While significant progress has been made toward greater gender equality, particularly in mixed and female-dominated sports, there is still a considerable gap in many traditional male sports. This underscores the ongoing need for policies and initiatives that promote inclusivity and encourage participation from all genders in every sport.

Hypothesis 4: Athlete demographics (age, height, weight) have evolved over time due to advancements in sports science.

Evolution of Athlete Height, Age, and Weight Over the Decades

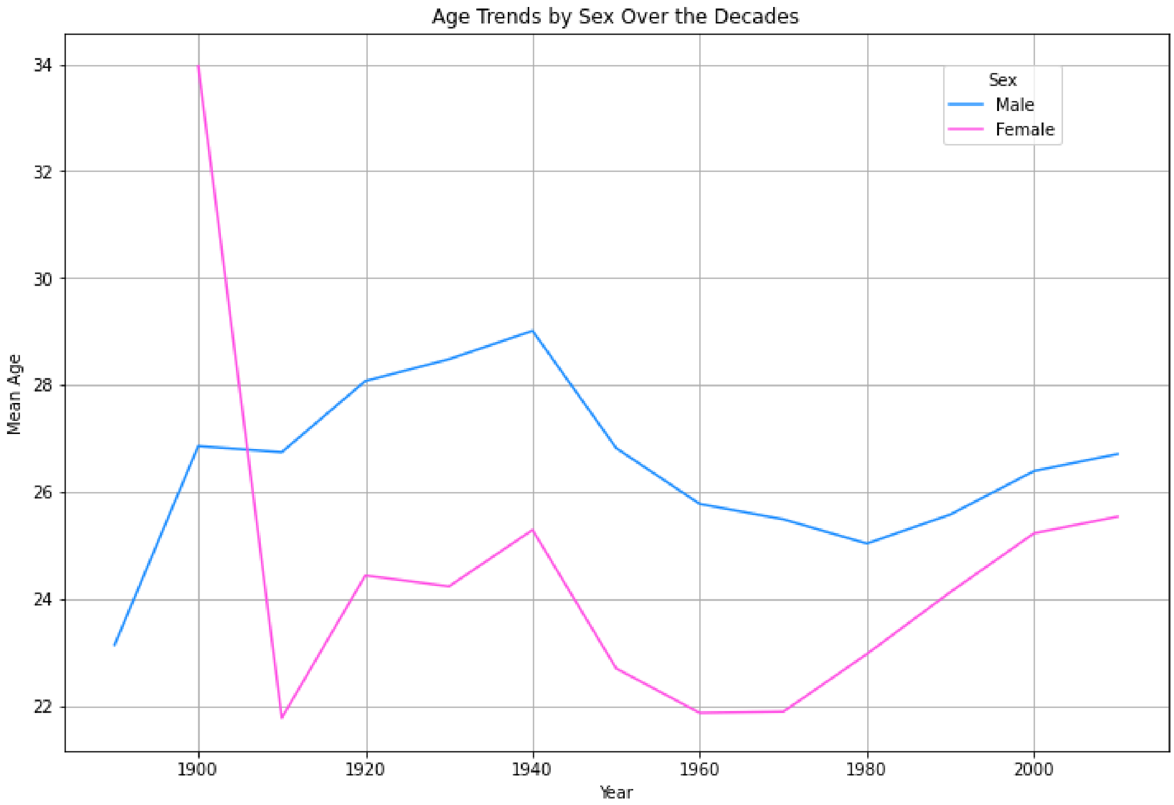
In [22]:
```
# Group athletes by decade to explore demographic trends
def categorize_decade(year):
    return (year // 10) * 10

athlete_data_rem_dupli['Decade'] = athlete_data_rem_dupli['Year'].apply(categorize_de
cade)

# Group by Decade, Sex, and Season (Summer/Winter) and calculate mean for Age, Heigh
t, and Weight
demographic_trends = athlete_data_rem_dupli.groupby(['Decade', 'Sex']).agg({
    'Age': 'mean',
    'Height': 'mean',
    'Weight': 'mean'
}).reset_index()

# Custom color palette
palette = {'F': '#FF53E6', 'M': '#1E90FF'}

# Create the plot for all three metrics (Height, Age, Weight) in one plot
plt.figure(figsize=(12, 8))

# Plot Height
sns.lineplot(data=demographic_trends, x='Decade', y='Height', hue='Sex', markers=Tru
e, palette=palette)
# Plot Age
sns.lineplot(data=demographic_trends, x='Decade', y='Age', hue='Sex', markers=True, p
alette=palette, legend=False)
# Plot Weight
sns.lineplot(data=demographic_trends, x='Decade', y='Weight', hue='Sex', markers=Tru
e, palette=palette, legend=False)

# Title and Labels
plt.title("Demographic Trends Over Decades (Sex Differentiation)")
plt.xlabel("Decade")
plt.ylabel("Mean Value")

# Add legend with bbox_to_anchor to move it up
plt.legend(title='Sex',labels=['Male', 'Female'], bbox_to_anchor=(0.5, 0.60), loc='ce
nter')

# Annotate the lines with labels for better clarity
plt.text(x=1990, y=173, s="Height", color='blue', fontsize=12, ha='center')
plt.text(x=1990, y=30, s="Age", color='red', fontsize=12, ha='center')
plt.text(x=1990, y=68, s="Weight", color='green', fontsize=12, ha='center')

# Add grid and show plot
plt.grid(True)
plt.show()
```



1. Evolution of Athlete Age Over the Decades

In [23]:

```python
# Group athletes by decade to explore demographic trends
def categorize_decade(year):
    return (year // 10) * 10

athlete_data_rem_dupli['Decade'] = athlete_data_rem_dupli['Year'].apply(categorize_de
cade)

# Group by Year, Sex and calculate mean for Age
demographic_trends_age = athlete_data_rem_dupli.groupby(['Decade', 'Sex']).agg({
    'Age': 'mean'
}).reset_index()

# Plot only Age
plt.figure(figsize=(12, 8))
sns.lineplot(data=demographic_trends_age, x='Decade', y='Age', hue='Sex', markers=Tru
e, palette=palette)

# Add legend with bbox_to_anchor to move it up and avoid the "Sex" title
plt.legend(title='Sex', labels=['Male', 'Female'], bbox_to_anchor=(0.85, 0.90), loc
='center')

plt.title("Age Trends by Sex Over the Decades")
plt.xlabel("Year")
plt.ylabel("Mean Age")
plt.grid(True)
plt.show()
```



Conclusion: Age Trends and Historical Influences Over the Years

The analysis of the average age of Olympic athletes over the years reveals clear patterns shaped by historical events, societal shifts, and advancements in sports science. Key findings from the data indicate the impact of global events, including the World Wars, the Great Depression, and the Cold War, on the evolution of athlete demographics.

1. Early 20th Century (1896–1936):

- **Female Athletes:** Female participation in the Olympics began in **1900**, with an initial average age of **29.4 years**. The high average age in **1904** (**48.8 years**) likely reflects the limited number of female athletes, many of whom were older, and possibly the effects of societal norms that restricted women's participation in sports. The **World War I** (1914–1918) and the **Great Depression** (1929–1939) further exacerbated these trends, contributing to sporadic participation and older athletes.

- **Male Athletes:** Male athletes' ages fluctuated between **23 and 29 years** during these years, with a noticeable increase in the **1932 Olympics** (**31.0 years**). This increase could have been influenced by disruptions in the athlete pool caused by the **Great Depression** and **World War I**, as fewer young athletes could participate in global competitions.

2. Mid-20th Century (1948–1964):

- **Female Athletes:** By the **1948 Olympics**, female athletes' average age had dropped to **25.3 years**, reflecting societal shifts post-WWII and growing acceptance of women in competitive sports. This period marked the beginning of greater professional support for women in sports, which continued into the **1960s** when the average age stabilized between **22 and 23 years**.

- **Male Athletes:** Male athletes' ages stabilized around **26–29 years**, likely due to the recovery from **World War II** and the influence of **Cold War** training programs. This era saw improved training facilities and youth development programs, particularly in newly emerging nations.

3. Late 20th Century (1976–2000):

- **Female Athletes:** The average age for female athletes remained consistent between **23 and 25 years**, reflecting the growing professionalism in women's sports and the influence of youth-focused training programs.
- **Male Athletes:** Similarly, the average age of male athletes remained around **24–26 years** during this period, a consistency influenced by the **Cold War** rivalry and the early identification of talent through state-sponsored training programs.

4. 21st Century (2004–2016):

- **Female Athletes:** The average age for female athletes gradually increased to **25–26 years**, reflecting the extended careers made possible by advancements in sports medicine, nutrition, and the increasing professionalization of women's sports.

- **Male Athletes:** The average age for male athletes also saw a slight increase, rising to **26–27 years** by **2016**, as improvements in recovery and longer athletic careers became more common due to the evolution of sports science.

Summary:

The trends in the data highlight how major global events such as **World War I**, **World War II**, the **Great Depression**, and the **Cold War** have influenced the average ages of Olympic athletes. These disruptions, alongside advancements in sports science, nutrition, and societal changes, have led to the evolving age demographics.

- In the early years, **female athletes** were significantly older due to limited participation, while **male athletes** maintained more consistent ages.

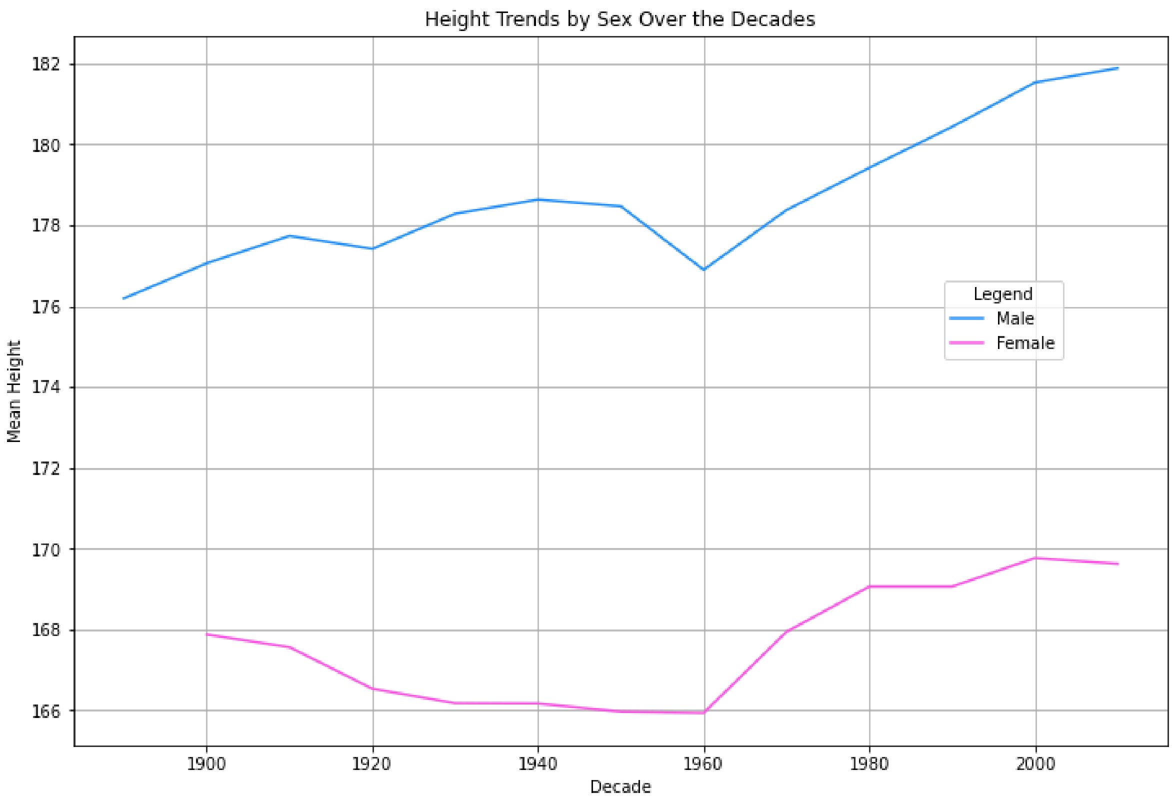
- In the post-WWII and Cold War eras, both genders saw a rise in younger athletes, influenced by professional training programs and growing global participation.

- By the 21st century, advancements in sports medicine and increased career longevity led to a slight increase in the average age for both male and female athletes, with more athletes competing at older ages due to better health management.

These findings underline the profound impact of both historical events and scientific advancements on the athletic careers of Olympians over the years.

2. Evolution of Athlete Height Over the Decades

In [24]:
```python
# Group athletes by decade to explore demographic trends
def categorize_decade(year):
    return (year // 10) * 10

athlete_data_rem_dupli['Decade'] = athlete_data_rem_dupli['Year'].apply(categorize_de
cade)

# Group by Year, Sex and calculate mean for Height
demographic_trends_height = athlete_data_rem_dupli.groupby(['Decade', 'Sex']).agg({
    'Height': 'mean'
}).reset_index()

# Plot only Height
plt.figure(figsize=(12, 8))
sns.lineplot(data=demographic_trends_height, x='Decade', y='Height', hue='Sex', marke
rs=True, palette=palette)

# Add legend with bbox_to_anchor to move it up
plt.legend(title='Legend',labels=['Male', 'Female'], bbox_to_anchor=(0.85, 0.60), loc
='center')

plt.title("Height Trends by Sex Over the Decades")
plt.xlabel("Decade")
plt.ylabel("Mean Height")
plt.grid(True)
plt.show()
```



Conclusion: Athlete Height Trends by Sex Over the Decades

The analysis of athlete height trends by sex over the decades reveals several key insights into how the average heights of male and female athletes have evolved from the 1896 Olympics to the 2016 Olympics, with historical events such as **World War I**, **World War II**, the **Great Depression**, and the **Cold War** potentially influencing these changes.

Key Findings:

Male Athlete Heights:

- The average height of male athletes has generally shown a **consistent increase**, with minor fluctuations throughout the decades.

- In the early years, male athletes' heights were relatively stable, with small increases from **1896** (**176.19 cm**) to **1900** (**177.90 cm**), followed by continued growth. By **2016**, the average height of male athletes reached **182.08 cm**.

- This gradual increase in male athlete height suggests that physical attributes have evolved, likely due to improvements in **nutrition**, **training**, and **global physical standards**, with post-World War II recovery periods playing a role in enabling such changes.

Female Athlete Heights:

- Unlike male athletes, female athletes' heights exhibited more **variability** over the decades.

- In the early years, female athletes' average height was inconsistent, fluctuating from **169.85 cm** in **1900** to **167.56 cm** in **1920**, and continuing to show variation through the 1930s.

- From the **1960s** onward, however, the average height of female athletes began to steadily rise, reaching **170.32 cm** by **2016**. This trend aligns with increased **female participation** in sports and greater societal acceptance, likely spurred by **advances in nutrition** and **sports science**.

- Events such as the **Cold War** might also have contributed to increased access to sports training programs for both male and female athletes.

Gender Comparison:

- **Male athletes** have consistently been taller than **female athletes**, with the height difference between the sexes maintaining a steady gap over the years. In **1896**, the gap was about **10 cm**, and by **2016**, the difference had widened to approximately **12 cm**.

- While both male and female athlete heights have increased over time, the increase in **male athlete height** has been more **steady**, compared to the more **variable increase** in female athlete height.

- This steady increase in male height might reflect broader societal trends such as improved living standards, the impact of **World War II** and **post-WWII recovery**, as well as the greater inclusion of physical training as a norm.

Conclusions:

- The increasing heights of athletes, particularly male athletes, seem to be largely driven by better **training**, **diet**, and **genetics**, with improvements in **sports science** and **nutrition** since the mid-20th century facilitating these trends.

- The steady rise in female athletes' heights, especially from the **1960s onward**, reflects a change in **sports participation** and the growing **acceptance of women in sports**.

- The trends also show that the **impact of major global events**, like **WWI**, **WWII**, and the **Cold War**, likely disrupted sports participation during specific periods, potentially influencing height measurements.

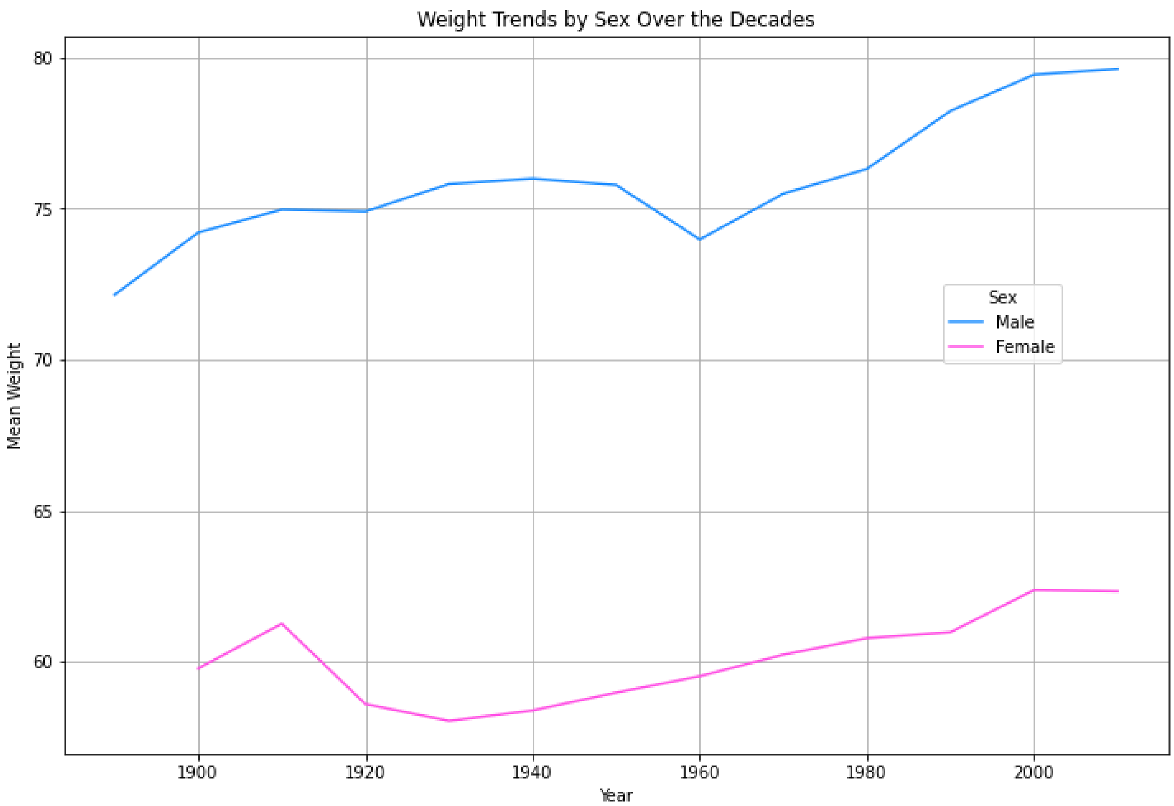
- While the overall increase in athlete height is notable, male athletes have experienced a larger and more consistent rise compared to female athletes.

In conclusion, the height trends of athletes across genders not only reflect improvements in **athletic training** and **nutrition**, but also broader **societal shifts**, such as greater inclusivity in sports for women and the effects of historical events. The male athletes' height increase appears more steady and consistent, while female athletes' growth, though noticeable, displays more variability, particularly influenced by **social changes** and **global conflicts**.

3. Evolution of Athlete Weight Over the Decades

In [25]:

```
# Group athletes by decade to explore demographic trends
def categorize_decade(year):
    return (year // 10) * 10

athlete_data_rem_dupli['Decade'] = athlete_data_rem_dupli['Year'].apply(categorize_de
cade)

# Group by Year, Sex and calculate mean for Weight
demographic_trends_height = athlete_data_rem_dupli.groupby(['Decade', 'Sex']).agg({
    'Weight': 'mean'
}).reset_index()

# Plot only Weight
plt.figure(figsize=(12, 8))
sns.lineplot(data=demographic_trends, x='Decade', y='Weight', hue='Sex', markers=Tru
e, palette=palette)

# Add legend with bbox_to_anchor to move it up
plt.legend(title='Sex',labels=['Male', 'Female'], bbox_to_anchor=(0.85, 0.60), loc='c
enter')

plt.title("Weight Trends by Sex Over the Decades")
plt.xlabel("Year")
plt.ylabel("Mean Weight")
plt.grid(True)
plt.show()
```



Conclusion: Trends in Female and Male Athlete Weight Over the Decades

The weight of athletes has evolved over time, with distinct trends observed for both male and female athletes. Analyzing the data for female and male athletes separately reveals the following key insights:

Female Athlete Weight Trends:

- The weight of female athletes has experienced fluctuations throughout the decades. In the early years, particularly in **1896**, the weight data for females is not available.

- Between **1900** and the **1950s**, female athletes' weight remained relatively stable, with small increases and decreases, primarily hovering around the **60 kg** mark.

- From the **1960s** to the **1980s**, there was a gradual increase in weight, with values consistently above **58 kg**.

- In the later years, from the **1990s** to **2010s**, the weight of female athletes has generally increased, reaching about **61-62 kg** by **2016**. In **2016**, the average weight for female athletes was **61.87 kg**, a noticeable increase from earlier decades.

Male Athlete Weight Trends:

- Male athletes have seen a steady increase in weight since the early years of the data. The weight started around **71.80 kg** in **1896**, and by **2016**, it had increased to about **78.48 kg**.

- The **1940s** and **1950s** marked a period of more substantial weight increases, with male athletes gaining more weight compared to earlier years.
- This trend continued through the **1970s** to **1990s**, where the male athlete's weight increased at a consistent pace.

- Post-2000, the weight of male athletes has remained relatively stable, with slight variations observed between **77 kg** and **78.5 kg**. In **2016**, the average weight for male athletes was **78.48 kg**, consistent with the highest observed values.

Observations:

- There is a clear upward trend in both male and female athlete weight over the decades, with male athletes experiencing a more consistent and larger increase.

- For female athletes, the weight increase is more gradual, especially noticeable in the later decades, with the largest increase observed between **2004** and **2016**, from **61.21 kg** to **61.87 kg**.

- The increase in athlete weight over time could be influenced by factors such as changes in **training methods**, **nutrition**, and the increasing **specialization of sports**.

Conclusion:

While both male and female athlete weights have risen over the decades, the weight growth for males has been more pronounced and consistent. Male athletes have experienced a steady increase in weight, from around **71.80 kg** in **1896** to **78.48 kg** in **2016**, reflecting changes in training and nutrition, as well as potential physiological shifts. In contrast, female athletes have seen a more gradual increase, reaching an average of **61.87 kg** in **2016**. This could be influenced by changes in female sports participation and growing acceptance of women in competitive athletics.

In conclusion, the weight trends for athletes over time suggest broader societal and physiological changes in the sports world, with male athletes showing a more consistent weight increase, while female athletes' weight growth has been more gradual but steady, particularly in the later decades.

General Conclusion for the Hypothesis

The evolution of athlete demographics—age, height, and weight—over time demonstrates a clear correlation with advancements in sports science, nutrition, and global societal changes. These demographic shifts highlight the dynamic nature of competitive sports and reflect the influence of historical events, technological progress, and societal attitudes toward athletes.

- **Age Trends:** Athletes' average age has fluctuated throughout history, with early years marked by older participants due to limited training programs and smaller pools of competitors. Post-World War II advancements in youth sports development brought younger athletes into the spotlight. More recently, improvements in sports medicine and recovery techniques have enabled longer athletic careers, slightly increasing the average age for both male and female athletes in the 21st century.

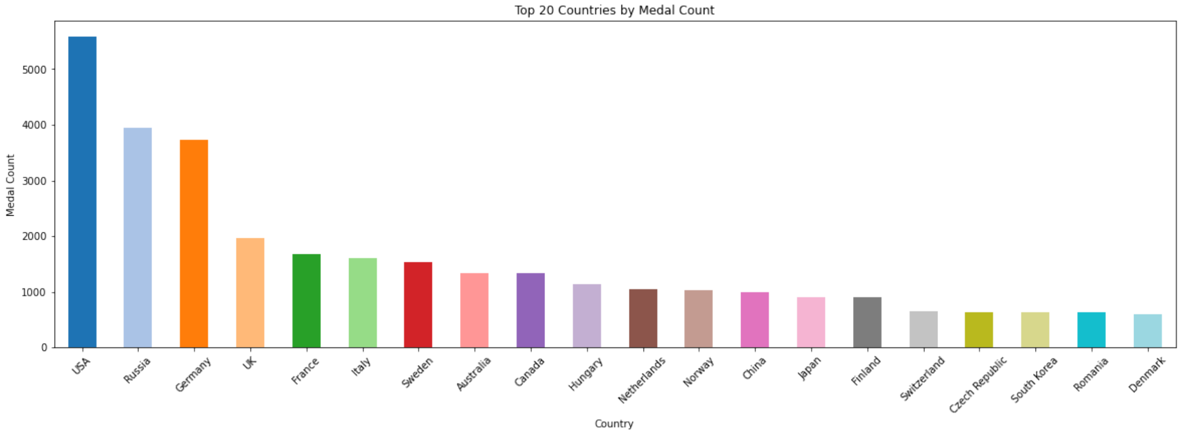
- **Height Trends:** Athlete height has shown an upward trajectory, with male athletes displaying a steady increase and female athletes experiencing more variability until the 1960s, after which heights stabilized and began to rise. This growth aligns with improved nutrition, genetics, and the increasing emphasis on physical attributes specific to various sports disciplines.

- **Weight Trends:** Both male and female athletes have seen a gradual increase in weight over the decades, reflecting enhanced training methods, nutrition, and the growing physical demands of modern sports. Male athletes have shown a more consistent and significant weight increase compared to females, highlighting differences in specialization and physiological adaptation in sports.

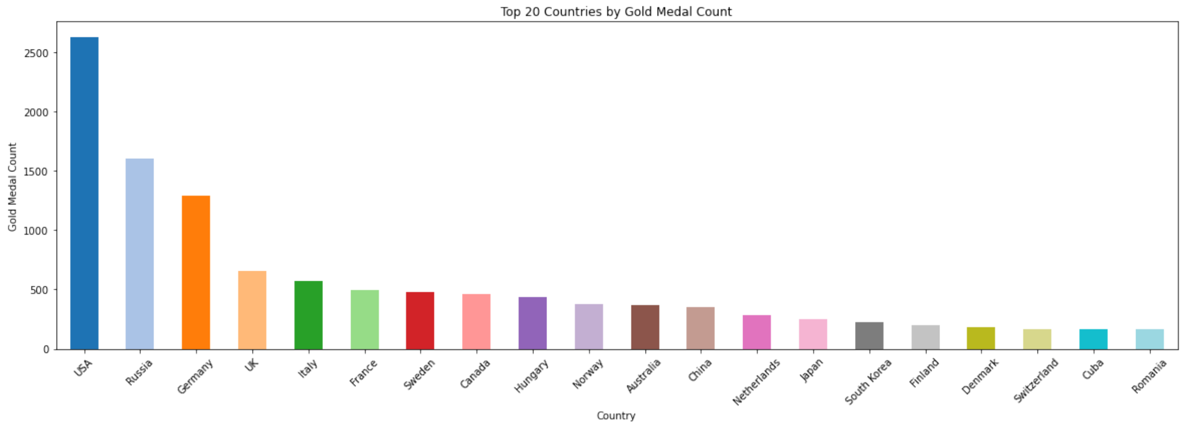
In summary, the changes in age, height, and weight among athletes over time underscore the transformative impact of sports science, societal shifts, and global participation. These factors collectively illustrate how athletics has evolved to optimize human performance and adapt to the changing landscape of competitive sports.

Hypothesis 5: Countries with a better BMI (Body Mass Index) tend to win more medals, suggesting a possible correlation between physical size and athletic success.

In [26]:
```
# Medal count by country
top_countries = athlete_data.groupby('region')['Medal'].count().sort_values(ascending
=False).head(20)

region = athlete_data['region']
colors = plt.cm.tab20.colors[:len(region)]

# Bar plot for top countries
plt.figure(figsize=(20, 6))
top_countries.plot(kind='bar', color=colors)
plt.title("Top 20 Countries by Medal Count")
plt.ylabel("Medal Count")
plt.xlabel("Country")
plt.xticks(rotation=45)
plt.show()
```



In [27]:
```python
# Filter for Gold medals only
gold_medals = athlete_data[athlete_data['Medal'] == 'Gold']

# Medal count by country (Gold medals only)
top_gold_countries = gold_medals.groupby('region')['Medal'].count().sort_values(ascen
ding=False).head(20)

# Generate colors for the bars
colors = plt.cm.tab20.colors[:len(top_gold_countries)]

# Bar plot for top countries with Gold medals
plt.figure(figsize=(20, 6))
top_gold_countries.plot(kind='bar', color=colors)
plt.title("Top 20 Countries by Gold Medal Count")
plt.ylabel("Gold Medal Count")
plt.xlabel("Country")
plt.xticks(rotation=45)
plt.show()
```



Visualizing the Top 10 Countries by Medal Count Across Different Sports

Introduction:

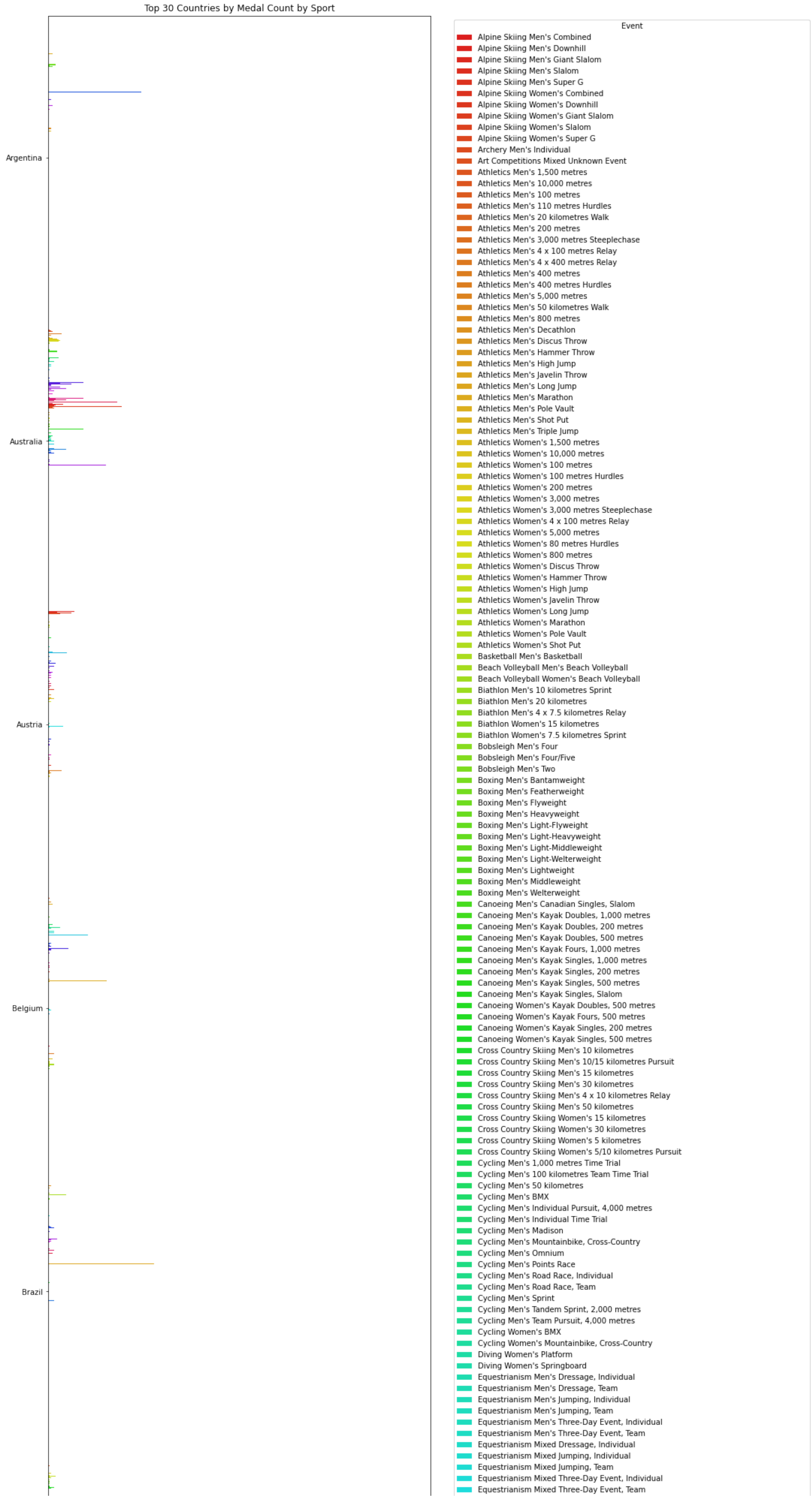
In this analysis, we will explore the distribution of Olympic medals among the top 10 countries. By using distinct colors to represent different sports, we will create a bar plot to visualize how each country has performed across various sports. The goal is to showcase which countries excelled in specific events and how their medal counts compare across different disciplines. The plot will provide insights into the relative strengths of each nation in the Olympic Games.

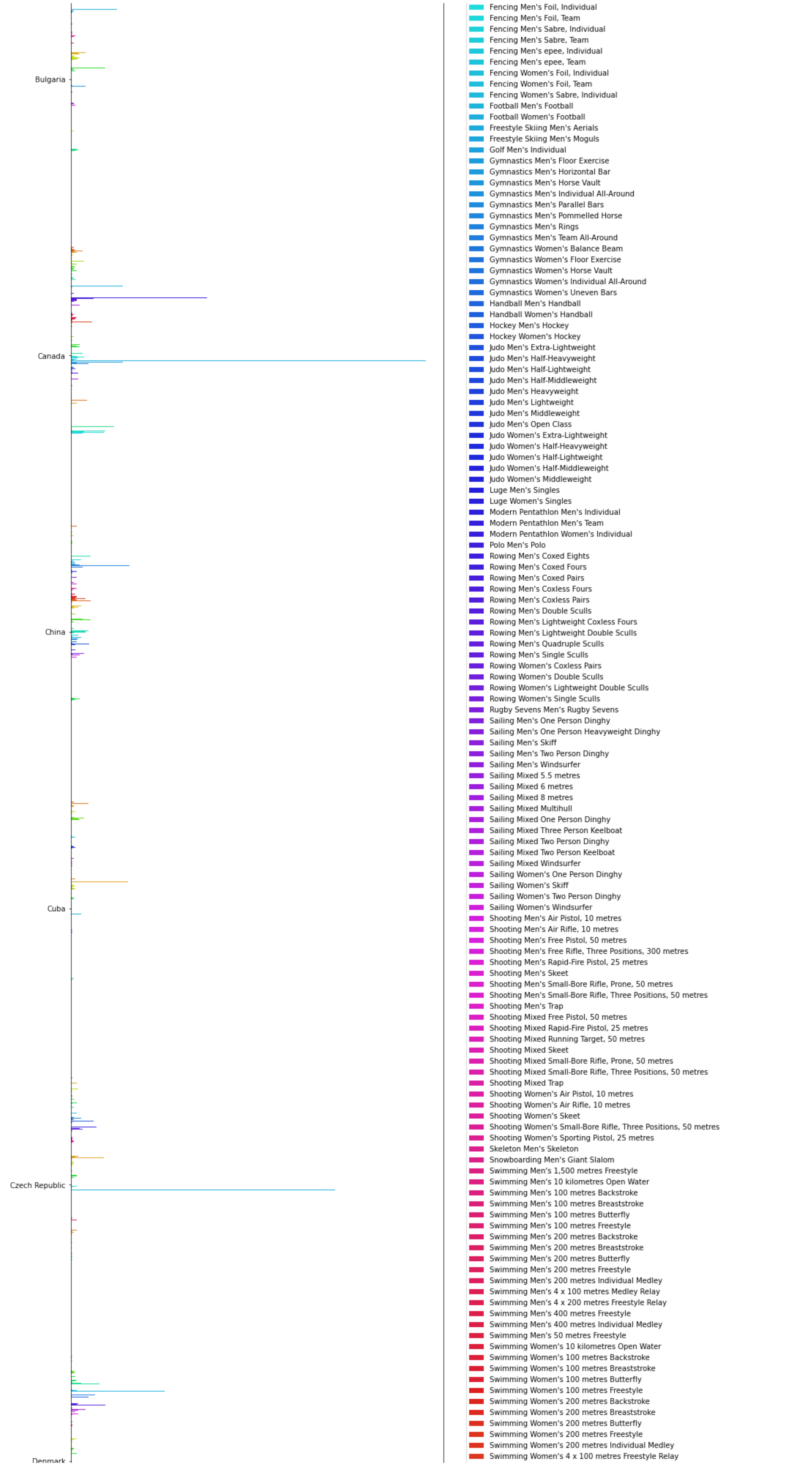
How Visualizing Medal Distribution Supports the BMI and Athletic Success Hypothesis

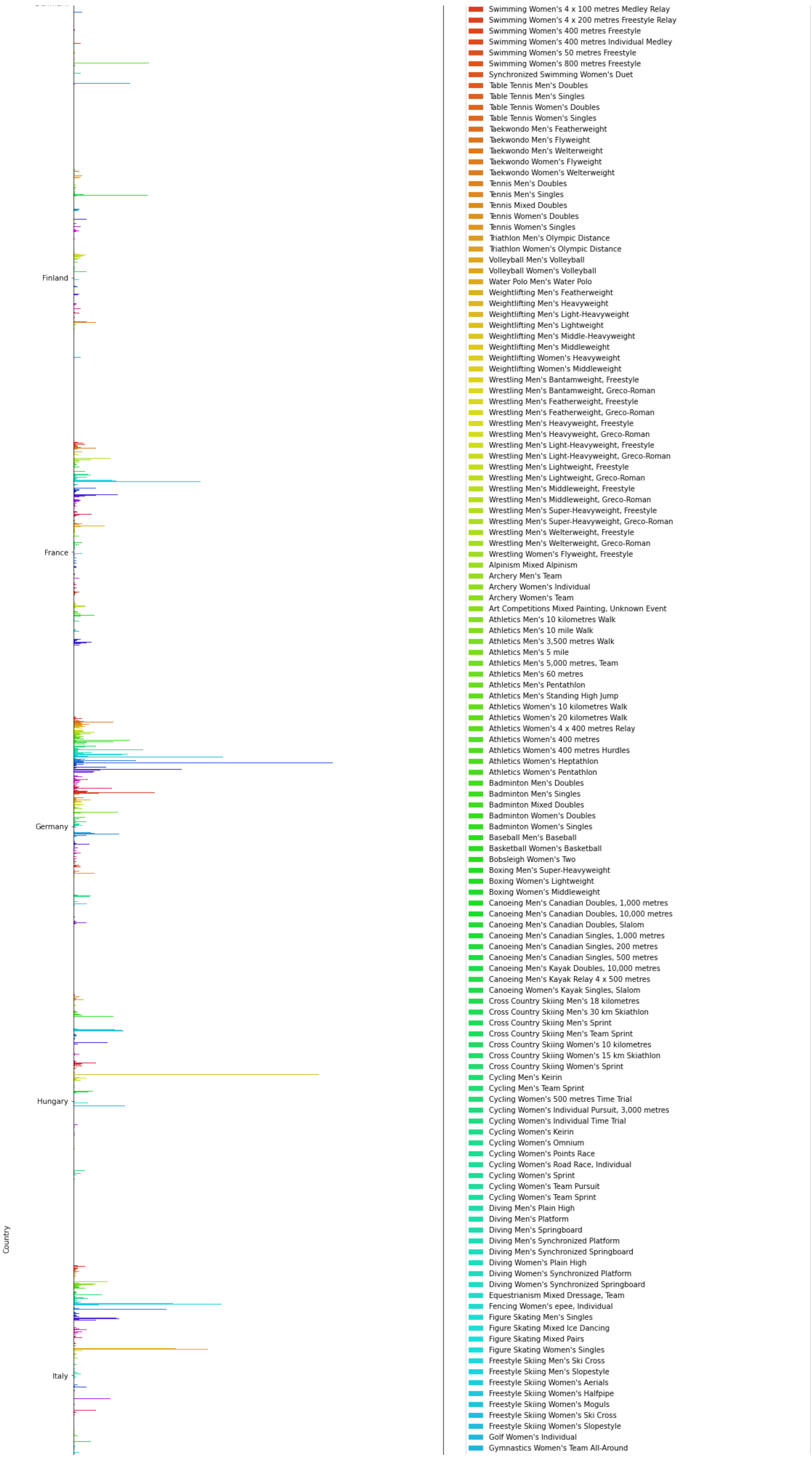
Visualizing the top 10 countries by medal count across different sports provides critical insights into the relationship between BMI and athletic success by highlighting patterns of dominance in specific events. Different sports require varying physical attributes, with some emphasizing endurance, strength, or agility, while others, like basketball or rowing, may rely heavily on height, weight, and overall physical size. By correlating these sports with the medal-winning countries and their average athlete BMIs, this visualization can reveal whether countries with specific physical profiles tend to excel in sports where size and body composition play a crucial role.

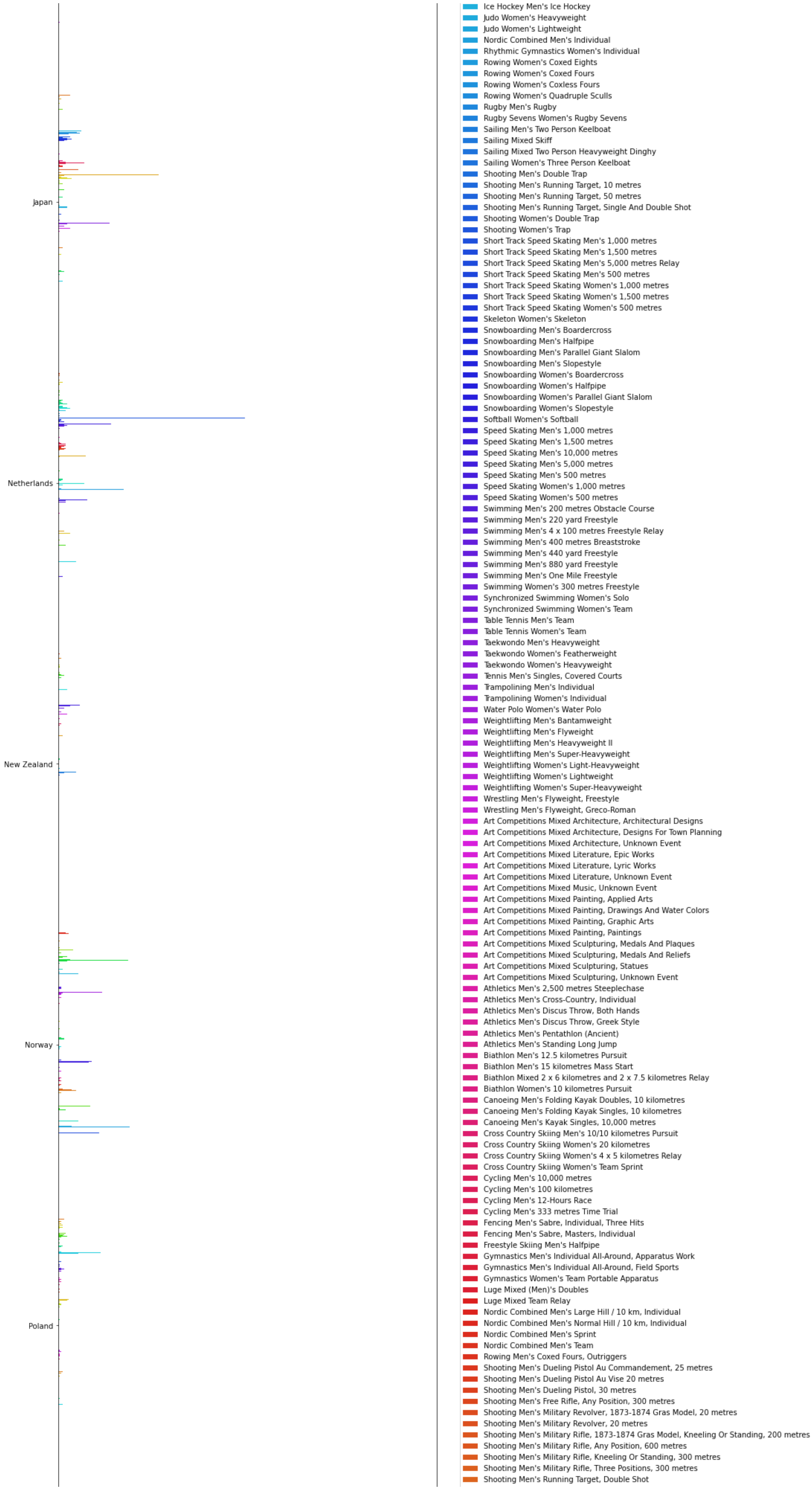
For example, if countries with higher average BMIs dominate sports requiring physical power, such as weightlifting or basketball, it supports the idea that physical size is a significant factor in athletic success. Additionally, this analysis can help identify exceptions, where countries excel in sports regardless of physical size, providing a more nuanced understanding of the interplay between BMI and medal performance across disciplines.

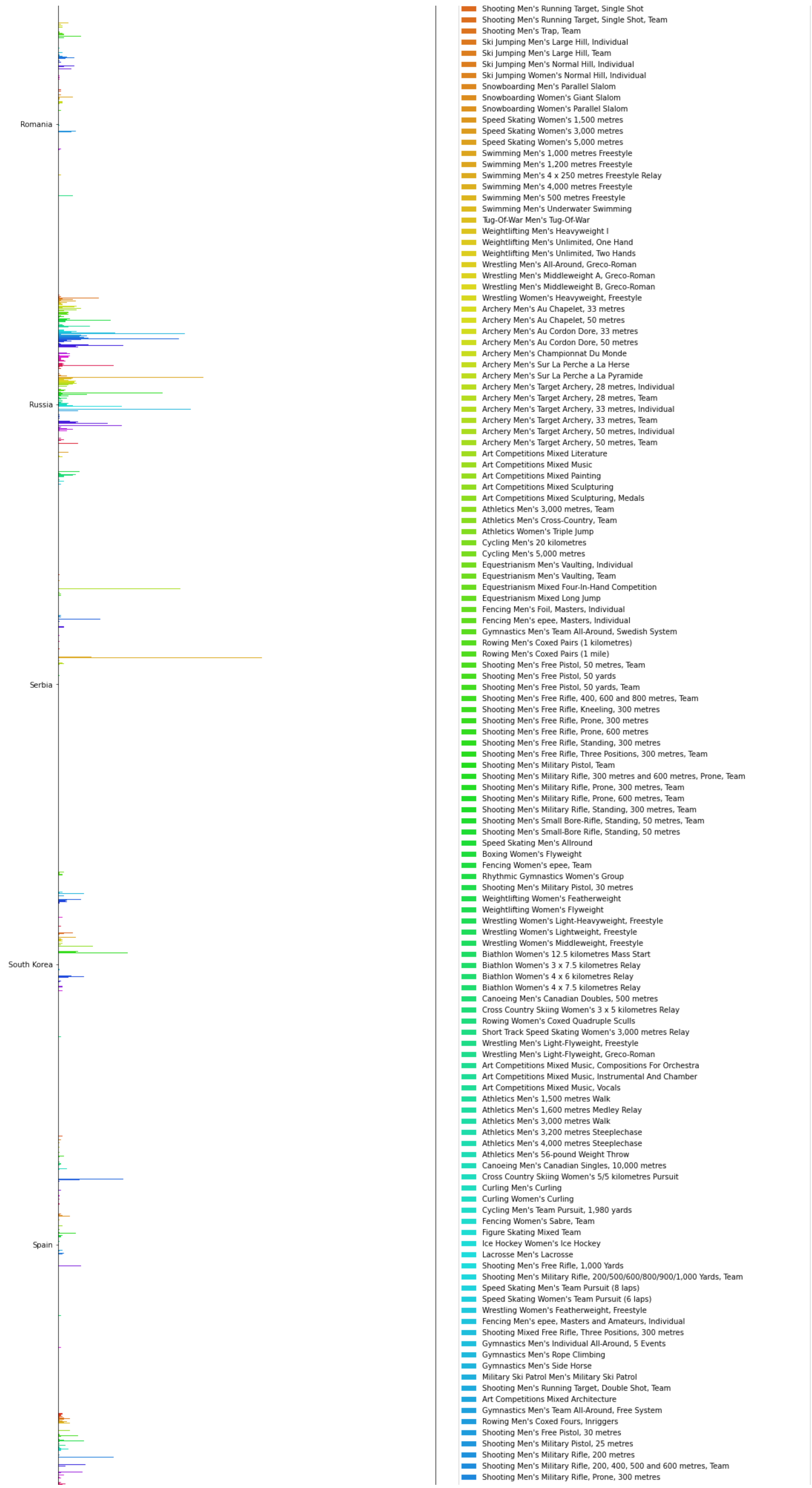
In [28]:
```
# Generate 200 distinct colors from multiple colormaps
colors = plt.cm.hsv(np.linspace(0, 1, 256))


# Group by Region, Sport, and Medal, and count the number of medals
medals_by_country_sport = athlete_data.groupby(['region', 'Event'])['Medal'].count().
reset_index()

# Sort by number of medals, descending order, and get the top 30 countries
top_countries = medals_by_country_sport.groupby('region')['Medal'].sum().nlargest(3
0).index
top_medals = medals_by_country_sport[medals_by_country_sport['region'].isin(top_count
ries)]

# Create a bar plot to display the results
plt.figure(figsize=(16, 180))

# Plot the data, assigning a distinct color to each sport
sns.barplot(data=top_medals, x='Medal', y='region', hue='Event', palette=colors)

# Title and labels
plt.title("Top 30 Countries by Medal Count by Sport")
plt.xlabel("Number of Medals")
plt.ylabel("Country")
plt.legend(title='Event', bbox_to_anchor=(1.05, 1), loc='upper left')

# Show the plot
plt.tight_layout()
plt.show()
```

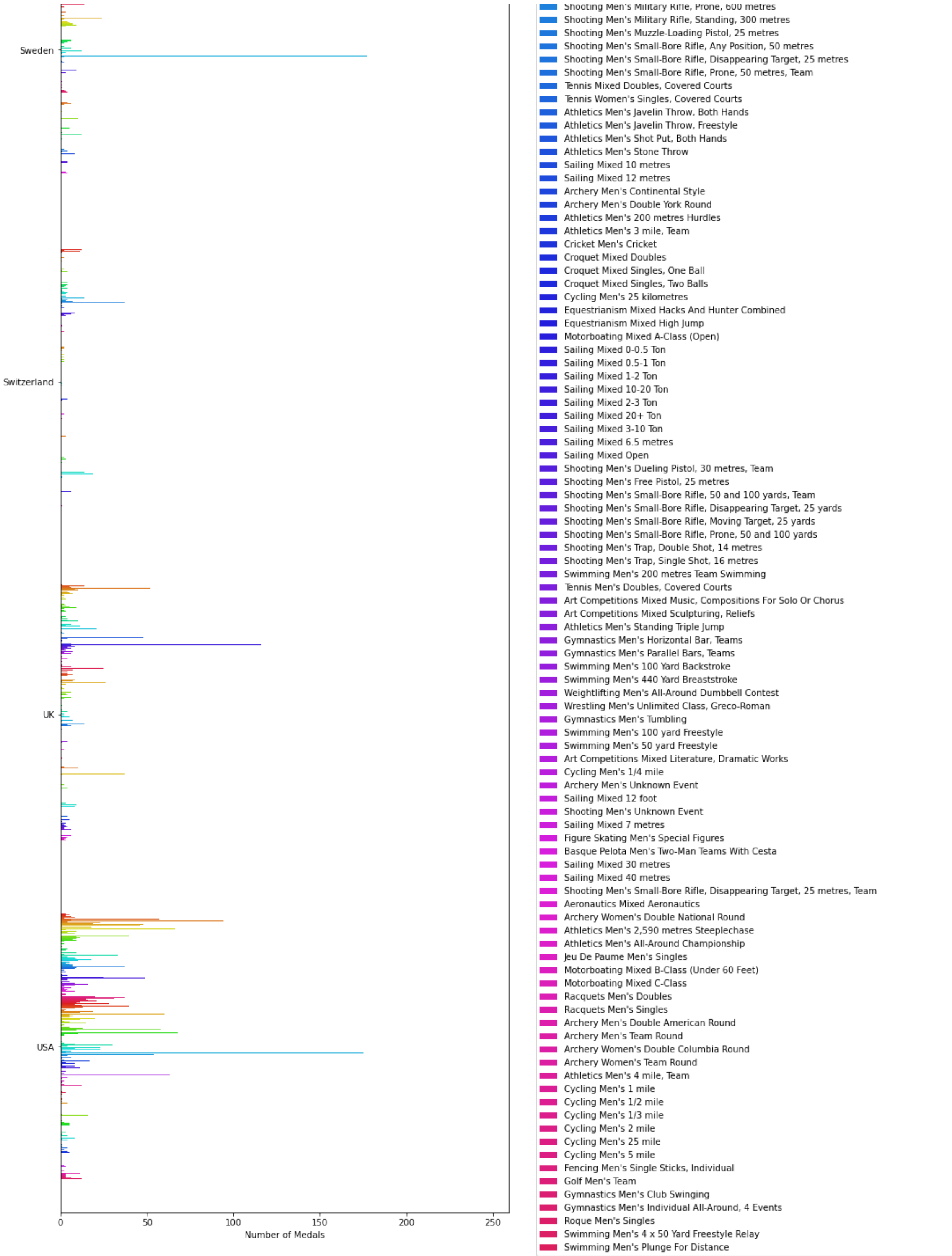

Top 30 Countries by Medal Count by Sport
Argentina
Australia
Austria
Belgium
Brazil
Event
Alpine Skiing Men's Combined
Alpine Skiing Men's Downhill
Alpine Skiing Men's Giant Slalom
Alpine Skiing Men's Slalom
Alpine Skiing Men's Super G
Alpine Skiing Women's Combined
Alpine Skiing Women's Downhill
Alpine Skiing Women's Giant Slalom
Alpine Skiing Women's Slalom
Alpine Skiing Women's Super G
Archery Men's Individual
Art Competitions Mixed Unknown Event
Athletics Men's 1,500 metres
Athletics Men's 10,000 metres
Athletics Men's 100 metres
Athletics Men's 110 metres Hurdles
Athletics Men's 20 kilometres Walk
Athletics Men's 200 metres
Athletics Men's 3,000 metres Steeplechase
Athletics Men's 4 x 100 metres Relay
Athletics Men's 4 x 400 metres Relay
Athletics Men's 400 metres
Athletics Men's 400 metres Hurdles
Athletics Men's 5,000 metres
Athletics Men's 50 kilometres Walk
Athletics Men's 800 metres
Athletics Men's Decathlon
Athletics Men's Discus Throw
Athletics Men's Hammer Throw
Athletics Men's High Jump
Athletics Men's Javelin Throw
Athletics Men's Long Jump
Athletics Men's Marathon
Athletics Men's Pole Vault
Athletics Men's Shot Put
Athletics Men's Triple Jump
Athletics Women's 1,500 metres
Athletics Women's 10,000 metres
Athletics Women's 100 metres
Athletics Women's 100 metres Hurdles
Athletics Women's 200 metres
Athletics Women's 3,000 metres
Athletics Women's 3,000 metres Steeplechase
Athletics Women's 4 x 100 metres Relay
Athletics Women's 5,000 metres
Athletics Women's 80 metres Hurdles
Athletics Women's 800 metres
Athletics Women's Discus Throw
Athletics Women's Hammer Throw
Athletics Women's High Jump
Athletics Women's Javelin Throw
Athletics Women's Long Jump
Athletics Women's Marathon
Athletics Women's Pole Vault
Athletics Women's Shot Put
Basketball Men's Basketball
Beach Volleyball Men's Beach Volleyball
Beach Volleyball Women's Beach Volleyball
Biathlon Men's 10 kilometres Sprint
Biathlon Men's 20 kilometres
Biathlon Men's 4 x 7.5 kilometres Relay
Biathlon Women's 15 kilometres
Biathlon Women's 7.5 kilometres Sprint
Bobsleigh Men's Four
Bobsleigh Men's Four/Five
Bobsleigh Men's Two
Boxing Men's Bantamweight
Boxing Men's Featherweight
Boxing Men's Flyweight
Boxing Men's Heavyweight
Boxing Men's Light-Flyweight
Boxing Men's Light-Heavyweight
Boxing Men's Light-Middleweight
Boxing Men's Light-Welterweight
Boxing Men's Lightweight
Boxing Men's Middleweight
Boxing Men's Welterweight
Canoeing Men's Canadian Singles, Slalom
Canoeing Men's Kayak Doubles, 1,000 metres
Canoeing Men's Kayak Doubles, 200 metres
Canoeing Men's Kayak Doubles, 500 metres
Canoeing Men's Kayak Fours, 1,000 metres
Canoeing Men's Kayak Singles, 1,000 metres
Canoeing Men's Kayak Singles, 200 metres
Canoeing Men's Kayak Singles, 500 metres
Canoeing Men's Kayak Singles, Slalom
Canoeing Women's Kayak Doubles, 500 metres
Canoeing Women's Kayak Fours, 500 metres
Canoeing Women's Kayak Singles, 200 metres
Canoeing Women's Kayak Singles, 500 metres
Cross Country Skiing Men's 10 kilometres
Cross Country Skiing Men's 10/15 kilometres Pursuit
Cross Country Skiing Men's 15 kilometres
Cross Country Skiing Men's 30 kilometres
Cross Country Skiing Men's 4 x 10 kilometres Relay
Cross Country Skiing Men's 50 kilometres
Cross Country Skiing Women's 15 kilometres
Cross Country Skiing Women's 30 kilometres
Cross Country Skiing Women's 5 kilometres
Cross Country Skiing Women's 5/10 kilometres Pursuit
Cycling Men's 1,000 metres Time Trial
Cycling Men's 100 kilometres Team Time Trial
Cycling Men's 50 kilometres
Cycling Men's BMX
Cycling Men's Individual Pursuit, 4,000 metres
Cycling Men's Individual Time Trial
Cycling Men's Madison
Cycling Men's Mountainbike, Cross-Country
Cycling Men's Omnium
Cycling Men's Points Race
Cycling Men's Road Race, Individual
Cycling Men's Road Race, Team
Cycling Men's Sprint
Cycling Men's Tandem Sprint, 2,000 metres
Cycling Men's Team Pursuit, 4,000 metres
Cycling Women's BMX
Cycling Women's Mountainbike, Cross-Country
Diving Women's Platform
Diving Women's Springboard
Equestrianism Men's Dressage, Individual
Equestrianism Men's Dressage, Team
Equestrianism Men's Jumping, Individual
Equestrianism Men's Jumping, Team
Equestrianism Men's Three-Day Event, Individual
Equestrianism Men's Three-Day Event, Team
Equestrianism Mixed Dressage, Individual
Equestrianism Mixed Jumping, Individual
Equestrianism Mixed Jumping, Team
Equestrianism Mixed Three-Day Event, Individual
Equestrianism Mixed Three-Day Event, Team


Bulgaria
Canada
China
Cuba
Czech Republic
Denmark
Fencing Men's Foil, Individual
Fencing Men's Foil, Team
Fencing Men's Sabre, Individual
Fencing Men's Sabre, Team
Fencing Men's epee, Individual
Fencing Men's epee, Team
Fencing Women's Foil, Individual
Fencing Women's Foil, Team
Fencing Women's Sabre, Individual
Football Men's Football
Football Women's Football
Freestyle Skiing Men's Aerials
Freestyle Skiing Men's Moguls
Golf Men's Individual
Gymnastics Men's Floor Exercise
Gymnastics Men's Horizontal Bar
Gymnastics Men's Horse Vault
Gymnastics Men's Individual All-Around
Gymnastics Men's Parallel Bars
Gymnastics Men's Pommelled Horse
Gymnastics Men's Rings
Gymnastics Men's Team All-Around
Gymnastics Women's Balance Beam
Gymnastics Women's Floor Exercise
Gymnastics Women's Horse Vault
Gymnastics Women's Individual All-Around
Gymnastics Women's Uneven Bars
Handball Men's Handball
Handball Women's Handball
Hockey Men's Hockey
Hockey Women's Hockey
Judo Men's Extra-Lightweight
Judo Men's Half-Heavyweight
Judo Men's Half-Lightweight
Judo Men's Half-Middleweight
Judo Men's Heavyweight
Judo Men's Lightweight
Judo Men's Middleweight
Judo Men's Open Class
Judo Women's Extra-Lightweight
Judo Women's Half-Heavyweight
Judo Women's Half-Lightweight
Judo Women's Half-Middleweight
Judo Women's Middleweight
Luge Men's Singles
Luge Women's Singles
Modern Pentathlon Men's Individual
Modern Pentathlon Men's Team
Modern Pentathlon Women's Individual
Polo Men's Polo
Rowing Men's Coxed Eights
Rowing Men's Coxed Fours
Rowing Men's Coxed Pairs
Rowing Men's Coxless Fours
Rowing Men's Coxless Pairs
Rowing Men's Double Sculls
Rowing Men's Lightweight Coxless Fours
Rowing Men's Lightweight Double Sculls
Rowing Men's Quadruple Sculls
Rowing Men's Single Sculls
Rowing Women's Coxless Pairs
Rowing Women's Double Sculls
Rowing Women's Lightweight Double Sculls
Rowing Women's Single Sculls
Rugby Sevens Men's Rugby Sevens
Sailing Men's One Person Dinghy
Sailing Men's One Person Heavyweight Dinghy
Sailing Men's Skiff
Sailing Men's Two Person Dinghy
Sailing Men's Windsurfer
Sailing Mixed 5.5 metres
Sailing Mixed 6 metres
Sailing Mixed 8 metres
Sailing Mixed Multihull
Sailing Mixed One Person Dinghy
Sailing Mixed Three Person Keelboat
Sailing Mixed Two Person Dinghy
Sailing Mixed Two Person Keelboat
Sailing Mixed Windsurfer
Sailing Women's One Person Dinghy
Sailing Women's Skiff
Sailing Women's Two Person Dinghy
Sailing Women's Windsurfer
Shooting Men's Air Pistol, 10 metres
Shooting Men's Air Rifle, 10 metres
Shooting Men's Free Pistol, 50 metres
Shooting Men's Free Rifle, Three Positions, 300 metres
Shooting Men's Rapid-Fire Pistol, 25 metres
Shooting Men's Skeet
Shooting Men's Small-Bore Rifle, Prone, 50 metres
Shooting Men's Small-Bore Rifle, Three Positions, 50 metres
Shooting Men's Trap
Shooting Mixed Free Pistol, 50 metres
Shooting Mixed Rapid-Fire Pistol, 25 metres
Shooting Mixed Running Target, 50 metres
Shooting Mixed Skeet
Shooting Mixed Small-Bore Rifle, Prone, 50 metres
Shooting Mixed Small-Bore Rifle, Three Positions, 50 metres
Shooting Mixed Trap
Shooting Women's Air Pistol, 10 metres
Shooting Women's Air Rifle, 10 metres
Shooting Women's Skeet
Shooting Women's Small-Bore Rifle, Three Positions, 50 metres
Shooting Women's Sporting Pistol, 25 metres
Skeleton Men's Skeleton
Snowboarding Men's Giant Slalom
Swimming Men's 1,500 metres Freestyle
Swimming Men's 10 kilometres Open Water
Swimming Men's 100 metres Backstroke
Swimming Men's 100 metres Breaststroke
Swimming Men's 100 metres Butterfly
Swimming Men's 100 metres Freestyle
Swimming Men's 200 metres Backstroke
Swimming Men's 200 metres Breaststroke
Swimming Men's 200 metres Butterfly
Swimming Men's 200 metres Freestyle
Swimming Men's 200 metres Individual Medley
Swimming Men's 4 x 100 metres Medley Relay
Swimming Men's 4 x 200 metres Freestyle Relay
Swimming Men's 400 metres Freestyle
Swimming Men's 400 metres Individual Medley
Swimming Men's 50 metres Freestyle
Swimming Women's 10 kilometres Open Water
Swimming Women's 100 metres Backstroke
Swimming Women's 100 metres Breaststroke
Swimming Women's 100 metres Butterfly
Swimming Women's 100 metres Freestyle
Swimming Women's 200 metres Backstroke
Swimming Women's 200 metres Breaststroke
Swimming Women's 200 metres Butterfly
Swimming Women's 200 metres Freestyle
Swimming Women's 200 metres Individual Medley
Swimming Women's 4 x 100 metres Freestyle Relay

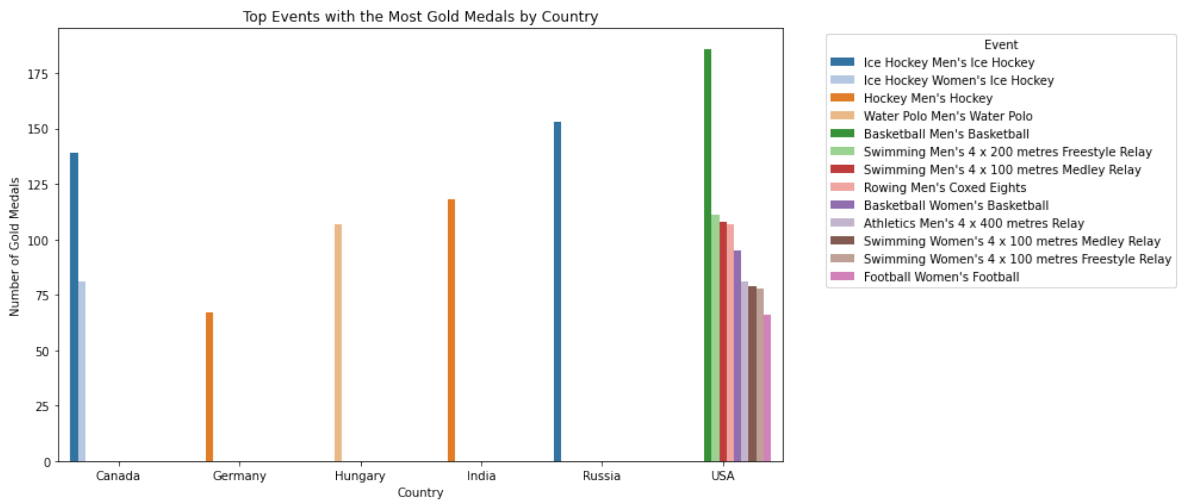

Finland
France
Germany
Hungary
Italy
Country
Swimming Women's 4 x 100 metres Medley Relay
Swimming Women's 4 x 200 metres Freestyle Relay
Swimming Women's 400 metres Freestyle
Swimming Women's 400 metres Individual Medley
Swimming Women's 50 metres Freestyle
Swimming Women's 800 metres Freestyle
Synchronized Swimming Women's Duet
Table Tennis Men's Doubles
Table Tennis Men's Singles
Table Tennis Women's Doubles
Table Tennis Women's Singles
Taekwondo Men's Featherweight
Taekwondo Men's Flyweight
Taekwondo Men's Welterweight
Taekwondo Women's Flyweight
Taekwondo Women's Welterweight
Tennis Men's Doubles
Tennis Men's Singles
Tennis Mixed Doubles
Tennis Women's Doubles
Tennis Women's Singles
Triathlon Men's Olympic Distance
Triathlon Women's Olympic Distance
Volleyball Men's Volleyball
Volleyball Women's Volleyball
Water Polo Men's Water Polo
Weightlifting Men's Featherweight
Weightlifting Men's Heavyweight
Weightlifting Men's Light-Heavyweight
Weightlifting Men's Lightweight
Weightlifting Men's Middle-Heavyweight
Weightlifting Men's Middleweight
Weightlifting Women's Heavyweight
Weightlifting Women's Middleweight
Wrestling Men's Bantamweight, Freestyle
Wrestling Men's Bantamweight, Greco-Roman
Wrestling Men's Featherweight, Freestyle
Wrestling Men's Featherweight, Greco-Roman
Wrestling Men's Heavyweight, Freestyle
Wrestling Men's Heavyweight, Greco-Roman
Wrestling Men's Light-Heavyweight, Freestyle
Wrestling Men's Light-Heavyweight, Greco-Roman
Wrestling Men's Lightweight, Freestyle
Wrestling Men's Lightweight, Greco-Roman
Wrestling Men's Middleweight, Freestyle
Wrestling Men's Middleweight, Greco-Roman
Wrestling Men's Super-Heavyweight, Freestyle
Wrestling Men's Super-Heavyweight, Greco-Roman
Wrestling Men's Welterweight, Freestyle
Wrestling Men's Welterweight, Greco-Roman
Wrestling Women's Flyweight, Freestyle
Alpinism Mixed Alpinism
Archery Men's Team
Archery Women's Individual
Archery Women's Team
Art Competitions Mixed Painting, Unknown Event
Athletics Men's 10 kilometres Walk
Athletics Men's 10 mile Walk
Athletics Men's 3,500 metres Walk
Athletics Men's 5 mile
Athletics Men's 5,000 metres, Team
Athletics Men's 60 metres
Athletics Men's Pentathlon
Athletics Men's Standing High Jump
Athletics Women's 10 kilometres Walk
Athletics Women's 20 kilometres Walk
Athletics Women's 4 x 400 metres Relay
Athletics Women's 400 metres
Athletics Women's 400 metres Hurdles
Athletics Women's Heptathlon
Athletics Women's Pentathlon
Badminton Men's Doubles
Badminton Men's Singles
Badminton Mixed Doubles
Badminton Women's Doubles
Badminton Women's Singles
Baseball Men's Baseball
Basketball Women's Basketball
Bobsleigh Women's Two
Boxing Men's Super-Heavyweight
Boxing Women's Lightweight
Boxing Women's Middleweight
Canoeing Men's Canadian Doubles, 1,000 metres
Canoeing Men's Canadian Doubles, 10,000 metres
Canoeing Men's Canadian Doubles, Slalom
Canoeing Men's Canadian Singles, 1,000 metres
Canoeing Men's Canadian Singles, 200 metres
Canoeing Men's Canadian Singles, 500 metres
Canoeing Men's Kayak Doubles, 10,000 metres
Canoeing Men's Kayak Relay 4 x 500 metres
Canoeing Women's Kayak Singles, Slalom
Cross Country Skiing Men's 18 kilometres
Cross Country Skiing Men's 30 km Skiathlon
Cross Country Skiing Men's Sprint
Cross Country Skiing Men's Team Sprint
Cross Country Skiing Women's 10 kilometres
Cross Country Skiing Women's 15 km Skiathlon
Cross Country Skiing Women's Sprint
Cycling Men's Keirin
Cycling Men's Team Sprint
Cycling Women's 500 metres Time Trial
Cycling Women's Individual Pursuit, 3,000 metres
Cycling Women's Individual Time Trial
Cycling Women's Keirin
Cycling Women's Omnium
Cycling Women's Points Race
Cycling Women's Road Race, Individual
Cycling Women's Sprint
Cycling Women's Team Pursuit
Cycling Women's Team Sprint
Diving Men's Plain High
Diving Men's Platform
Diving Men's Springboard
Diving Men's Synchronized Platform
Diving Men's Synchronized Springboard
Diving Women's Plain High
Diving Women's Synchronized Platform
Diving Women's Synchronized Springboard
Equestrianism Mixed Dressage, Team
Fencing Women's epee, Individual
Figure Skating Men's Singles
Figure Skating Mixed Ice Dancing
Figure Skating Mixed Pairs
Figure Skating Women's Singles
Freestyle Skiing Men's Ski Cross
Freestyle Skiing Men's Slopestyle
Freestyle Skiing Women's Aerials
Freestyle Skiing Women's Halfpipe
Freestyle Skiing Women's Moguls
Freestyle Skiing Women's Ski Cross
Freestyle Skiing Women's Slopestyle
Golf Women's Individual
Gymnastics Women's Team All-Around


Japan
Netherlands
New Zealand
Norway
Poland
Ice Hockey Men's Ice Hockey
Judo Women's Heavyweight
Judo Women's Lightweight
Nordic Combined Men's Individual
Rhythmic Gymnastics Women's Individual
Rowing Women's Coxed Eights
Rowing Women's Coxed Fours
Rowing Women's Coxless Fours
Rowing Women's Quadruple Sculls
Rugby Men's Rugby
Rugby Sevens Women's Rugby Sevens
Sailing Men's Two Person Keelboat
Sailing Mixed Skiff
Sailing Mixed Two Person Heavyweight Dinghy
Sailing Women's Three Person Keelboat
Shooting Men's Double Trap
Shooting Men's Running Target, 10 metres
Shooting Men's Running Target, 50 metres
Shooting Men's Running Target, Single And Double Shot
Shooting Women's Double Trap
Shooting Women's Trap
Short Track Speed Skating Men's 1,000 metres
Short Track Speed Skating Men's 1,500 metres
Short Track Speed Skating Men's 5,000 metres Relay
Short Track Speed Skating Men's 500 metres
Short Track Speed Skating Women's 1,000 metres
Short Track Speed Skating Women's 1,500 metres
Short Track Speed Skating Women's 500 metres
Skeleton Women's Skeleton
Snowboarding Men's Boardercross
Snowboarding Men's Halfpipe
Snowboarding Men's Parallel Giant Slalom
Snowboarding Men's Slopestyle
Snowboarding Women's Boardercross
Snowboarding Women's Halfpipe
Snowboarding Women's Parallel Giant Slalom
Snowboarding Women's Slopestyle
Softball Women's Softball
Speed Skating Men's 1,000 metres
Speed Skating Men's 1,500 metres
Speed Skating Men's 10,000 metres
Speed Skating Men's 5,000 metres
Speed Skating Men's 500 metres
Speed Skating Women's 1,000 metres
Speed Skating Women's 500 metres
Swimming Men's 200 metres Obstacle Course
Swimming Men's 220 yard Freestyle
Swimming Men's 4 x 100 metres Freestyle Relay
Swimming Men's 400 metres Breaststroke
Swimming Men's 440 yard Freestyle
Swimming Men's 880 yard Freestyle
Swimming Men's One Mile Freestyle
Swimming Women's 300 metres Freestyle
Synchronized Swimming Women's Solo
Synchronized Swimming Women's Team
Table Tennis Men's Team
Table Tennis Women's Team
Taekwondo Men's Heavyweight
Taekwondo Women's Featherweight
Taekwondo Women's Heavyweight
Tennis Men's Singles, Covered Courts
Trampolining Men's Individual
Trampolining Women's Individual
Water Polo Women's Water Polo
Weightlifting Men's Bantamweight
Weightlifting Men's Flyweight
Weightlifting Men's Heavyweight II
Weightlifting Men's Super-Heavyweight
Weightlifting Women's Light-Heavyweight
Weightlifting Women's Lightweight
Weightlifting Women's Super-Heavyweight
Wrestling Men's Flyweight, Freestyle
Wrestling Men's Flyweight, Greco-Roman
Art Competitions Mixed Architecture, Architectural Designs
Art Competitions Mixed Architecture, Designs For Town Planning
Art Competitions Mixed Architecture, Unknown Event
Art Competitions Mixed Literature, Epic Works
Art Competitions Mixed Literature, Lyric Works
Art Competitions Mixed Literature, Unknown Event
Art Competitions Mixed Music, Unknown Event
Art Competitions Mixed Painting, Applied Arts
Art Competitions Mixed Painting, Drawings And Water Colors
Art Competitions Mixed Painting, Graphic Arts
Art Competitions Mixed Painting, Paintings
Art Competitions Mixed Sculpturing, Medals And Plaques
Art Competitions Mixed Sculpturing, Medals And Reliefs
Art Competitions Mixed Sculpturing, Statues
Art Competitions Mixed Sculpturing, Unknown Event
Athletics Men's 2,500 metres Steeplechase
Athletics Men's Cross-Country, Individual
Athletics Men's Discus Throw, Both Hands
Athletics Men's Discus Throw, Greek Style
Athletics Men's Pentathlon (Ancient)
Athletics Men's Standing Long Jump
Biathlon Men's 12.5 kilometres Pursuit
Biathlon Men's 15 kilometres Mass Start
Biathlon Mixed 2 x 6 kilometres and 2 x 7.5 kilometres Relay
Biathlon Women's 10 kilometres Pursuit
Canoeing Men's Folding Kayak Doubles, 10 kilometres
Canoeing Men's Folding Kayak Singles, 10 kilometres
Canoeing Men's Kayak Singles, 10,000 metres
Cross Country Skiing Men's 10/10 kilometres Pursuit
Cross Country Skiing Women's 20 kilometres
Cross Country Skiing Women's 4 x 5 kilometres Relay
Cross Country Skiing Women's Team Sprint
Cycling Men's 10,000 metres
Cycling Men's 100 kilometres
Cycling Men's 12-Hours Race
Cycling Men's 333 metres Time Trial
Fencing Men's Sabre, Individual, Three Hits
Fencing Men's Sabre, Masters, Individual
Freestyle Skiing Men's Halfpipe
Gymnastics Men's Individual All-Around, Apparatus Work
Gymnastics Men's Individual All-Around, Field Sports
Gymnastics Women's Team Portable Apparatus
Luge Mixed (Men)'s Doubles
Luge Mixed Team Relay
Nordic Combined Men's Large Hill / 10 km, Individual
Nordic Combined Men's Normal Hill / 10 km, Individual
Nordic Combined Men's Sprint
Nordic Combined Men's Team
Rowing Men's Coxed Fours, Outriggers
Shooting Men's Dueling Pistol Au Commandement, 25 metres
Shooting Men's Dueling Pistol Au Vise 20 metres
Shooting Men's Dueling Pistol, 30 metres
Shooting Men's Free Rifle, Any Position, 300 metres
Shooting Men's Military Revolver, 1873-1874 Gras Model, 20 metres
Shooting Men's Military Revolver, 20 metres
Shooting Men's Military Rifle, 1873-1874 Gras Model, Kneeling Or Standing, 200 metres
Shooting Men's Military Rifle, Any Position, 600 metres
Shooting Men's Military Rifle, Kneeling Or Standing, 300 metres
Shooting Men's Military Rifle, Three Positions, 300 metres
Shooting Men's Running Target, Double Shot


Romania
Russia
Serbia
South Korea
Spain
Shooting Men's Running Target, Single Shot
Shooting Men's Running Target, Single Shot, Team
Shooting Men's Trap, Team
Ski Jumping Men's Large Hill, Individual
Ski Jumping Men's Large Hill, Team
Ski Jumping Men's Normal Hill, Individual
Ski Jumping Women's Normal Hill, Individual
Snowboarding Men's Parallel Slalom
Snowboarding Women's Giant Slalom
Snowboarding Women's Parallel Slalom
Speed Skating Women's 1,500 metres
Speed Skating Women's 3,000 metres
Speed Skating Women's 5,000 metres
Swimming Men's 1,000 metres Freestyle
Swimming Men's 1,200 metres Freestyle
Swimming Men's 4 x 250 metres Freestyle Relay
Swimming Men's 4,000 metres Freestyle
Swimming Men's 500 metres Freestyle
Swimming Men's Underwater Swimming
Tug-Of-War Men's Tug-Of-War
Weightlifting Men's Heavyweight I
Weightlifting Men's Unlimited, One Hand
Weightlifting Men's Unlimited, Two Hands
Wrestling Men's All-Around, Greco-Roman
Wrestling Men's Middleweight A, Greco-Roman
Wrestling Men's Middleweight B, Greco-Roman
Wrestling Women's Heavyweight, Freestyle
Archery Men's Au Chapelet, 33 metres
Archery Men's Au Chapelet, 50 metres
Archery Men's Au Cordon Dore, 33 metres
Archery Men's Au Cordon Dore, 50 metres
Archery Men's Championnat Du Monde
Archery Men's Sur La Perche a La Herse
Archery Men's Sur La Perche a La Pyramide
Archery Men's Target Archery, 28 metres, Individual
Archery Men's Target Archery, 28 metres, Team
Archery Men's Target Archery, 33 metres, Individual
Archery Men's Target Archery, 33 metres, Team
Archery Men's Target Archery, 50 metres, Individual
Archery Men's Target Archery, 50 metres, Team
Art Competitions Mixed Literature
Art Competitions Mixed Music
Art Competitions Mixed Painting
Art Competitions Mixed Sculpturing
Art Competitions Mixed Sculpturing, Medals
Athletics Men's 3,000 metres, Team
Athletics Men's Cross-Country, Team
Athletics Women's Triple Jump
Cycling Men's 20 kilometres
Cycling Men's 5,000 metres
Equestrianism Men's Vaulting, Individual
Equestrianism Men's Vaulting, Team
Equestrianism Mixed Four-In-Hand Competition
Equestrianism Mixed Long Jump
Fencing Men's Foil, Masters, Individual
Fencing Men's epee, Masters, Individual
Gymnastics Men's Team All-Around, Swedish System
Rowing Men's Coxed Pairs (1 kilometres)
Rowing Men's Coxed Pairs (1 mile)
Shooting Men's Free Pistol, 50 metres, Team
Shooting Men's Free Pistol, 50 yards
Shooting Men's Free Pistol, 50 yards, Team
Shooting Men's Free Rifle, 400, 600 and 800 metres, Team
Shooting Men's Free Rifle, Kneeling, 300 metres
Shooting Men's Free Rifle, Prone, 300 metres
Shooting Men's Free Rifle, Prone, 600 metres
Shooting Men's Free Rifle, Standing, 300 metres
Shooting Men's Free Rifle, Three Positions, 300 metres, Team
Shooting Men's Military Pistol, Team
Shooting Men's Military Rifle, 300 metres and 600 metres, Prone, Team
Shooting Men's Military Rifle, Prone, 300 metres, Team
Shooting Men's Military Rifle, Prone, 600 metres, Team
Shooting Men's Military Rifle, Standing, 300 metres, Team
Shooting Men's Small Bore-Rifle, Standing, 50 metres, Team
Shooting Men's Small-Bore Rifle, Standing, 50 metres
Speed Skating Men's Allround
Boxing Women's Flyweight
Fencing Women's epee, Team
Rhythmic Gymnastics Women's Group
Shooting Men's Military Pistol, 30 metres
Weightlifting Women's Featherweight
Weightlifting Women's Flyweight
Wrestling Women's Light-Heavyweight, Freestyle
Wrestling Women's Lightweight, Freestyle
Wrestling Women's Middleweight, Freestyle
Biathlon Women's 12.5 kilometres Mass Start
Biathlon Women's 3 x 7.5 kilometres Relay
Biathlon Women's 4 x 6 kilometres Relay
Biathlon Women's 4 x 7.5 kilometres Relay
Canoeing Men's Canadian Doubles, 500 metres
Cross Country Skiing Women's 3 x 5 kilometres Relay
Rowing Women's Coxed Quadruple Sculls
Short Track Speed Skating Women's 3,000 metres Relay
Wrestling Men's Light-Flyweight, Freestyle
Wrestling Men's Light-Flyweight, Greco-Roman
Art Competitions Mixed Music, Compositions For Orchestra
Art Competitions Mixed Music, Instrumental And Chamber
Art Competitions Mixed Music, Vocals
Athletics Men's 1,500 metres Walk
Athletics Men's 1,600 metres Medley Relay
Athletics Men's 3,000 metres Walk
Athletics Men's 3,200 metres Steeplechase
Athletics Men's 4,000 metres Steeplechase
Athletics Men's 56-pound Weight Throw
Canoeing Men's Canadian Singles, 10,000 metres
Cross Country Skiing Women's 5/5 kilometres Pursuit
Curling Men's Curling
Curling Women's Curling
Cycling Men's Team Pursuit, 1,980 yards
Fencing Women's Sabre, Team
Figure Skating Mixed Team
Ice Hockey Women's Ice Hockey
Lacrosse Men's Lacrosse
Shooting Men's Free Rifle, 1,000 Yards
Shooting Men's Military Rifle, 200/500/600/800/900/1,000 Yards, Team
Speed Skating Men's Team Pursuit (8 laps)
Speed Skating Women's Team Pursuit (6 laps)
Wrestling Women's Featherweight, Freestyle
Fencing Men's epee, Masters and Amateurs, Individual
Shooting Mixed Free Rifle, Three Positions, 300 metres
Gymnastics Men's Individual All-Around, 5 Events
Gymnastics Men's Rope Climbing
Gymnastics Men's Side Horse
Military Ski Patrol Men's Military Ski Patrol
Shooting Men's Running Target, Double Shot, Team
Art Competitions Mixed Architecture
Gymnastics Men's Team All-Around, Free System
Rowing Men's Coxed Fours, Inriggers
Shooting Men's Free Pistol, 30 metres
Shooting Men's Military Pistol, 25 metres
Shooting Men's Military Rifle, 200 metres
Shooting Men's Military Rifle, 200, 400, 500 and 600 metres, Team
Shooting Men's Military Rifle, Prone, 300 metres


Sweden
Switzerland
UK
USA
0
50
100
150
200
250
Number of Medals
Shooting Men's Military Rifle, Prone, 600 metres
Shooting Men's Military Rifle, Standing, 300 metres
Shooting Men's Muzzle-Loading Pistol, 25 metres
Shooting Men's Small-Bore Rifle, Any Position, 50 metres
Shooting Men's Small-Bore Rifle, Disappearing Target, 25 metres
Shooting Men's Small-Bore Rifle, Prone, 50 metres, Team
Tennis Mixed Doubles, Covered Courts
Tennis Women's Singles, Covered Courts
Athletics Men's Javelin Throw, Both Hands
Athletics Men's Javelin Throw, Freestyle
Athletics Men's Shot Put, Both Hands
Athletics Men's Stone Throw
Sailing Mixed 10 metres
Sailing Mixed 12 metres
Archery Men's Continental Style
Archery Men's Double York Round
Athletics Men's 200 metres Hurdles
Athletics Men's 3 mile, Team
Cricket Men's Cricket
Croquet Mixed Doubles
Croquet Mixed Singles, One Ball
Croquet Mixed Singles, Two Balls
Cycling Men's 25 kilometres
Equestrianism Mixed Hacks And Hunter Combined
Equestrianism Mixed High Jump
Motorboating Mixed A-Class (Open)
Sailing Mixed 0-0.5 Ton
Sailing Mixed 0.5-1 Ton
Sailing Mixed 1-2 Ton
Sailing Mixed 10-20 Ton
Sailing Mixed 2-3 Ton
Sailing Mixed 20+ Ton
Sailing Mixed 3-10 Ton
Sailing Mixed 6.5 metres
Sailing Mixed Open
Shooting Men's Dueling Pistol, 30 metres, Team
Shooting Men's Free Pistol, 25 metres
Shooting Men's Small-Bore Rifle, 50 and 100 yards, Team
Shooting Men's Small-Bore Rifle, Disappearing Target, 25 yards
Shooting Men's Small-Bore Rifle, Moving Target, 25 yards
Shooting Men's Small-Bore Rifle, Prone, 50 and 100 yards
Shooting Men's Trap, Double Shot, 14 metres
Shooting Men's Trap, Single Shot, 16 metres
Swimming Men's 200 metres Team Swimming
Tennis Men's Doubles, Covered Courts
Art Competitions Mixed Music, Compositions For Solo Or Chorus
Art Competitions Mixed Sculpturing, Reliefs
Athletics Men's Standing Triple Jump
Gymnastics Men's Horizontal Bar, Teams
Gymnastics Men's Parallel Bars, Teams
Swimming Men's 100 Yard Backstroke
Swimming Men's 440 Yard Breaststroke
Weightlifting Men's All-Around Dumbbell Contest
Wrestling Men's Unlimited Class, Greco-Roman
Gymnastics Men's Tumbling
Swimming Men's 100 yard Freestyle
Swimming Men's 50 yard Freestyle
Art Competitions Mixed Literature, Dramatic Works
Cycling Men's 1/4 mile
Archery Men's Unknown Event
Sailing Mixed 12 foot
Shooting Men's Unknown Event
Sailing Mixed 7 metres
Figure Skating Men's Special Figures
Basque Pelota Men's Two-Man Teams With Cesta
Sailing Mixed 30 metres
Sailing Mixed 40 metres
Shooting Men's Small-Bore Rifle, Disappearing Target, 25 metres, Team
Aeronautics Mixed Aeronautics
Archery Women's Double National Round
Athletics Men's 2,590 metres Steeplechase
Athletics Men's All-Around Championship
Jeu De Paume Men's Singles
Motorboating Mixed B-Class (Under 60 Feet)
Motorboating Mixed C-Class
Racquets Men's Doubles
Racquets Men's Singles
Archery Men's Double American Round
Archery Men's Team Round
Archery Women's Double Columbia Round
Archery Women's Team Round
Athletics Men's 4 mile, Team
Cycling Men's 1 mile
Cycling Men's 1/2 mile
Cycling Men's 1/3 mile
Cycling Men's 2 mile
Cycling Men's 25 mile
Cycling Men's 5 mile
Fencing Men's Single Sticks, Individual
Golf Men's Team
Gymnastics Men's Club Swinging
Gymnastics Men's Individual All-Around, 4 Events
Roque Men's Singles
Swimming Men's 4 x 50 Yard Freestyle Relay
Swimming Men's Plunge For Distance

```
In [29]: # Filter for Gold medals only
gold_medals = athlete_data[athlete_data['Medal'] == 'Gold']

# Group by region and Event, and count the number of Gold medals
medals_by_country_sport = gold_medals.groupby(['region', 'Event'])['Medal'].count().n
largest(15).reset_index()

# Sort by number of Gold medals for each country, descending order, and take the top
5 events for each country
top_events_per_country = (
    medals_by_country_sport.groupby('region')
    .apply(lambda x: x.nlargest(20, 'Medal'))
    .reset_index(drop=True)
)

# Generate a distinct color palette for the bars
colors = plt.cm.tab20.colors[:len(top_events_per_country['Event'].unique())]

# Create a bar plot to display the results
plt.figure(figsize=(14, 6))  # Adjusted height for better visualization

# Plot the data, assigning a distinct color to each event
sns.barplot(data=top_events_per_country, x='region', y='Medal', hue='Event', palette=
colors)

# Title and labels
plt.title("Top Events with the Most Gold Medals by Country")
plt.xlabel("Country")
plt.ylabel("Number of Gold Medals")
plt.legend(title='Event', bbox_to_anchor=(1.05, 1))

# Show the plot
plt.tight_layout()
plt.show()
```



Observation

In this chart, Men's Basketball events have the highest number of awarded medals over the past 120 years, followed closely by Ice Hockey. However, since these are team sports, counting each medal individually inflates the totals. To ensure a more accurate representation of each region's success, we should adjust our calculations to count only one medal per sport.

Counting Gold Medals in Team Sports: A More Accurate Reflection of Regional Success

Introduction:

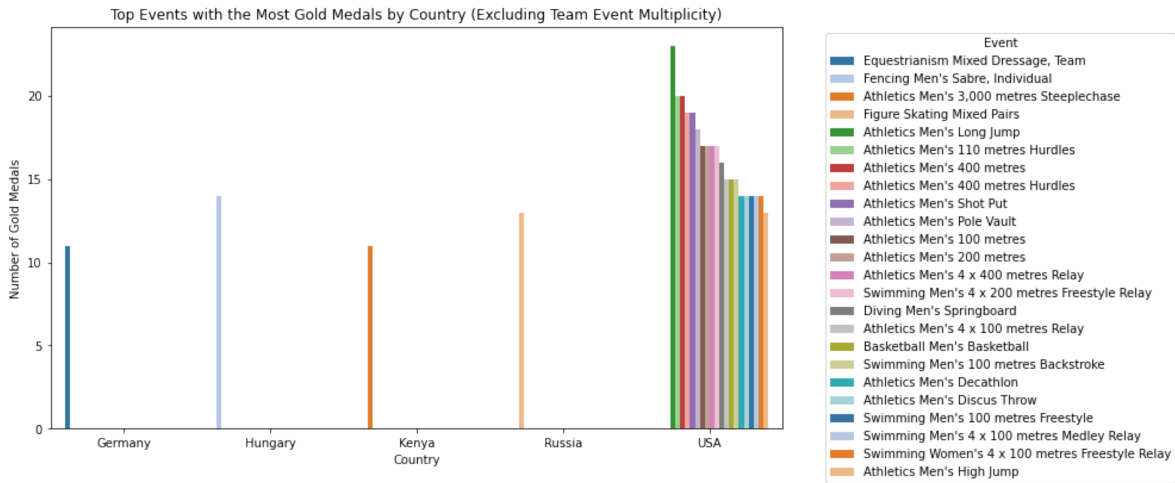
Gold medals represent the highest achievement in global competitions. However, in team sports, the way these medals are counted can sometimes distort the true performance of countries. Events like basketball, soccer, and volleyball typically involve multiple athletes, and a single team from a region may win the Gold medal. Counting individual medals for every athlete on the winning team can result in overestimating a region's success. This analysis aims to address this by counting only one medal per region-event pair, providing a more accurate depiction of success in team sports.

Why It's Relevant to Count Only One Medal for Team Sports:

In team sports, multiple athletes from the same country may win a Gold medal in the same event. Counting each athlete's medal can artificially inflate the total number of Gold medals for a region. For example, in Basketball or Soccer, each player on the winning team receives a medal. However, counting every individual medal for the same event may misrepresent a country's actual achievement. By counting only one Gold medal per region for each team event, we can present a fairer and more accurate portrayal of a country's success in these competitions.

This method helps provide a clearer comparison of regional dominance across both individual and team events, eliminating the overinflation caused by the number of athletes in team sports. It allows us to more accurately gauge where countries truly excel in global sporting competitions.

```
In [30]: # Filter for Gold medals only
gold_medals = athlete_data[athlete_data['Medal'] == 'Gold']

# Remove duplicates where multiple people from the same region won a gold medal in th
e same event
gold_medals_unique = gold_medals.drop_duplicates(subset=['region', 'Event', 'Year'])

# Group by region and event, and count the number of Gold medals (count only 1 medal
per region-event pair)
medals_by_country_sport = gold_medals_unique.groupby(['region', 'Event'])['Medal'].co
unt().nlargest(30).reset_index()

# Sort by the number of Gold medals for each country, descending order, and take the
top 5 events for each country
top_events_per_country = (
    medals_by_country_sport.groupby('region')
    .apply(lambda x: x.nlargest(20, 'Medal'))
    .reset_index(drop=True)
)

# Generate a distinct color palette for the bars
colors = plt.cm.tab20.colors[:len(top_events_per_country['Event'].unique())]

# Create a bar plot to display the results
plt.figure(figsize=(14, 6))  # Adjusted height for better visualization

# Plot the data, assigning a distinct color to each event
sns.barplot(data=top_events_per_country, x='region', y='Medal', hue='Event', palette=
colors)

# Title and labels
plt.title("Top Events with the Most Gold Medals by Country (Excluding Team Event Mult
iplicity)")
plt.xlabel("Country")
plt.ylabel("Number of Gold Medals")
plt.legend(title='Event', bbox_to_anchor=(1.05, 1))

# Show the plot
plt.tight_layout()
plt.show()
```

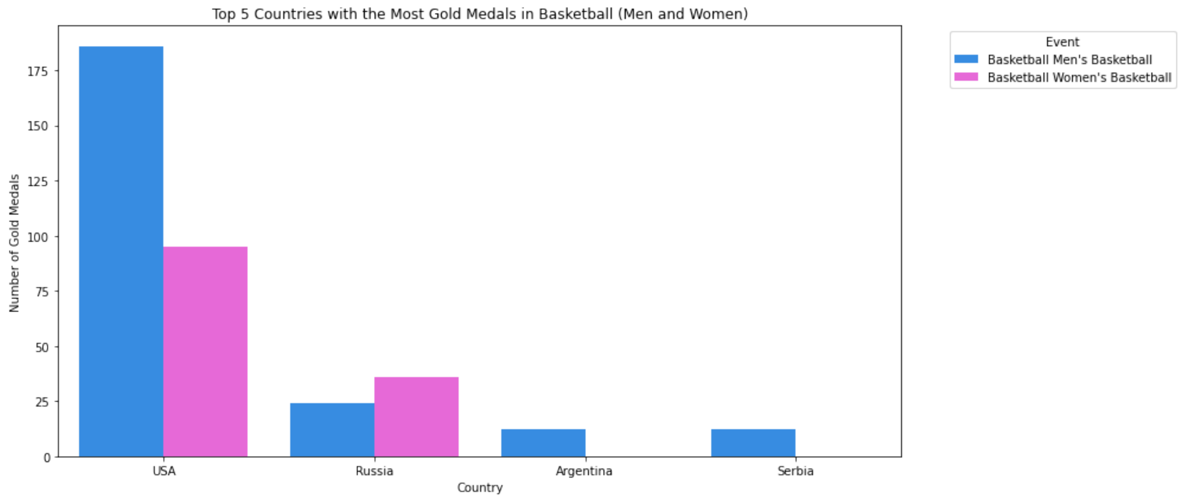


Observation

In this chart, by counting only one medal per event, we observe that Athletics Men's Long Jump holds the record for the most medals won over the past 120 years. This contrasts with the previous finding that basketball had the highest medal count, which is explained by the fact that long jump involves a single athlete per medal, whereas basketball medals are awarded to entire teams.

In [31]:

```
# Filter for Gold medals in the specified events
event_of_interest = ['Basketball Men\'s Basketball', 'Basketball Women\'s Basketbal
l']
gold_medals = athlete_data[
    (athlete_data['Medal'] == 'Gold') & (athlete_data['Event'].isin(event_of_interes
t))
]

# Group by region and Event, and count the number of Gold medals
medals_by_country_sport = (
    gold_medals.groupby(['region', 'Event'])['Medal']
    .count()
    .reset_index()
    .rename(columns={'Medal': 'Gold Count'})
)

# Get the top 5 countries for each event
top_events_per_country = (
    medals_by_country_sport.groupby('Event')
    .apply(lambda x: x.nlargest(200, 'Gold Count'))
    .reset_index(drop=True)
)

# Define colors for each event
sports_colors = {
    'Basketball Men\'s Basketball': '#1E90FF',
    'Basketball Women\'s Basketball': '#FF53E6'
}

# Create a bar plot to display the results
plt.figure(figsize=(14, 6))

# Plot the data, assigning a distinct color to each event
sns.barplot(
    data=top_events_per_country,
    x='region',
    y='Gold Count',
    hue='Event',
    palette=sports_colors
)

# Title and labels
plt.title("Top 5 Countries with the Most Gold Medals in Basketball (Men and Women)")
plt.xlabel("Country")
plt.ylabel("Number of Gold Medals")
plt.legend(title='Event', bbox_to_anchor=(1.05, 1), loc='upper left')

# Show the plot
plt.tight_layout()
plt.show()
```



Comparing Average BMI in Basketball Events (Men's and Women's) Across Top Countries

Introduction:

In this analysis, we will compare the average BMI (Body Mass Index) of athletes from the top 5 countries, excluding the USA, with the USA in the Men's and Women's Basketball events. The goal is to see how the USA compares to other leading countries in terms of athlete composition, considering both medal performance and physical attributes.

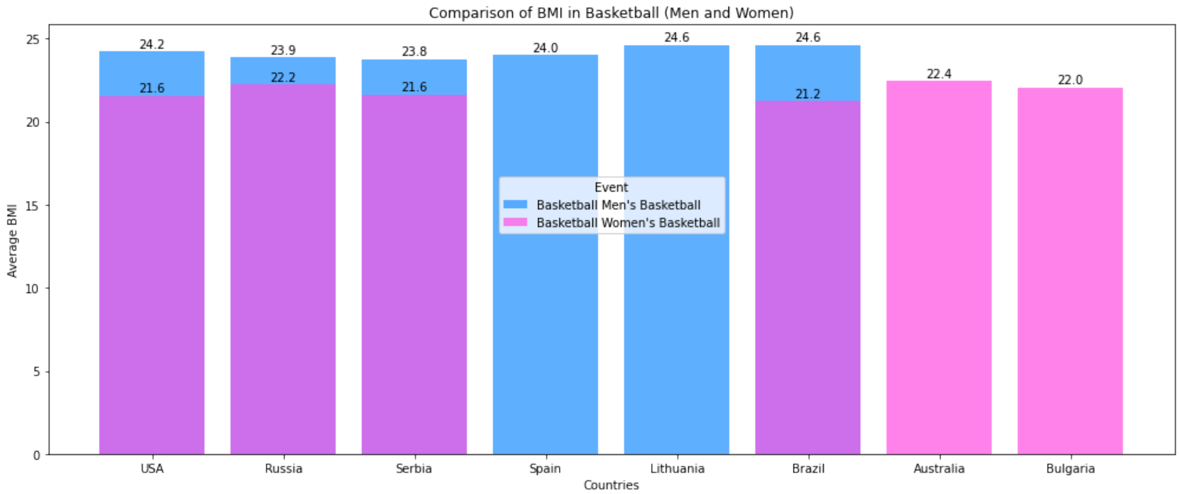
Focusing on all medals—gold, silver, and bronze—provides a more complete picture of a country's success. While gold medals are often seen as the ultimate measure of achievement, silver and bronze medals also reflect strong performances. By including all medal types, we avoid overemphasizing one aspect and instead capture the full scope of a country's athletic accomplishments. This broader perspective is particularly valuable when comparing the USA to other top-performing nations in these events.

Comparison of Average BMI in Men's and Women's Basketball by top 6 Country

In [32]:

```
# Define the events of interest
event_of_interest = ['Basketball Men\'s Basketball', 'Basketball Women\'s Basketbal
l']

# Filter for medal winners in the selected events
medal_winners = athlete_data[
    (athlete_data['Medal'].notnull()) & (athlete_data['Event'].isin(event_of_interes
t))
]

# Calculate BMI for each athlete
medal_winners['BMI'] = medal_winners['Weight'] / (medal_winners['Height'] / 100) ** 2

# Group by region and event to calculate average BMI and medal count
avg_bmi = (
    medal_winners.groupby(['region', 'Event'])
    .agg({'BMI': 'mean', 'Medal': 'count'})
    .rename(columns={'Medal': 'Medal Count'})
    .reset_index()
)

# Identify the top 5 countries by medal count for each sport, excluding the US
top_countries_by_sport = (
    avg_bmi[avg_bmi['region'] != 'USA']
    .groupby('Event')
    .apply(lambda x: x.nlargest(5, 'Medal Count'))
    .reset_index(drop=True)
)

# Add US data for comparison
us_data = avg_bmi[avg_bmi['region'] == 'USA']
comparison_data = pd.concat([us_data, top_countries_by_sport])

# Define colors for each event
sports_colors = {
    'Basketball Men\'s Basketball': '#1E90FF',
    'Basketball Women\'s Basketball': '#FF53E6'
}

# Plot comparison for the events of interest
plt.figure(figsize=(14, 6))

for event in event_of_interest:
    event_data = comparison_data[comparison_data['Event'] == event]
    countries = event_data['region']
    bmis = event_data['BMI']
    color = sports_colors[event]

    bars = plt.bar(
        countries, bmis, color=color, alpha=0.7, label=event
    )

    # Add value labels to bars
    for bar, bmi in zip(bars, bmis):
        plt.text(
            bar.get_x() + bar.get_width() / 2,
            bar.get_height() + 0.1,
            f'{bmi:.1f}',
            ha='center',
            va='bottom',
            fontsize=10
        )

# Customize the plot
plt.title("Comparison of BMI in Basketball (Men and Women)")
plt.ylabel("Average BMI")
plt.xlabel("Countries")
```

```python
plt.legend(title="Event", bbox_to_anchor=(0.5, 0.5))
plt.tight_layout()
plt.show()
```

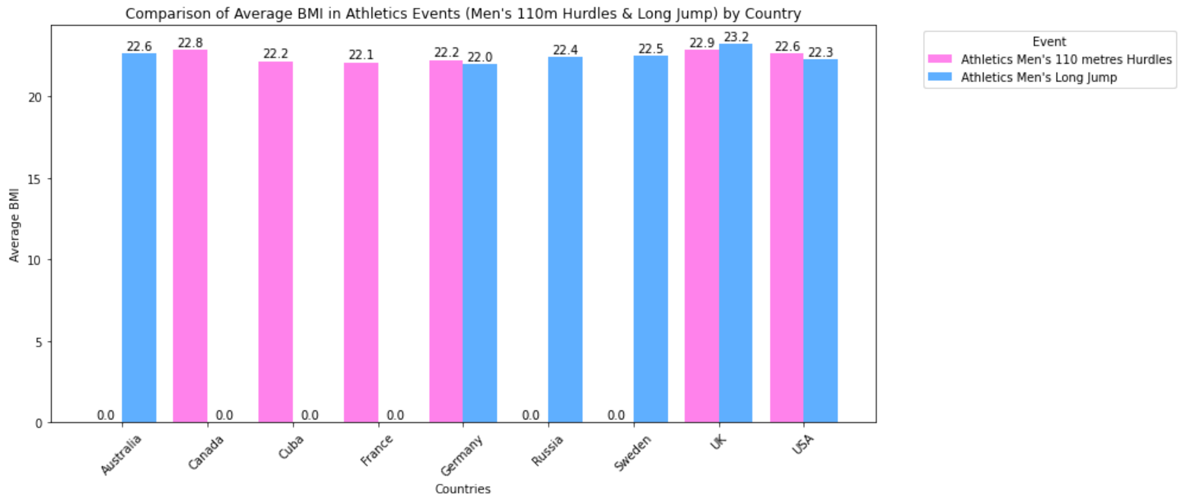


Comparison of Average BMI in Men's Athletics Events: 110m Hurdles & Long Jump Across Countries

In [33]:

```
# Define the events of interest
event_of_interest = ['Athletics Men\'s 110 metres Hurdles', 'Athletics Men\'s Long Ju
mp']

# Filter for medal winners in the selected events
medal_winners = athlete_data[
    (athlete_data['Medal'].notnull()) & (athlete_data['Event'].isin(event_of_interes
t))
]

# Calculate BMI for each athlete
medal_winners['BMI'] = medal_winners['Weight'] / (medal_winners['Height'] / 100) ** 2

# Group by region and event to calculate average BMI and medal count
avg_bmi = (
    medal_winners.groupby(['region', 'Event'])
    .agg({'BMI': 'mean', 'Medal': 'count'})
    .rename(columns={'Medal': 'Medal Count'})
    .reset_index()
)

# Identify the top 5 countries by medal count for each sport, excluding the US
top_countries_by_sport = (
    avg_bmi[avg_bmi['region'] != 'USA']
    .groupby('Event')
    .apply(lambda x: x.nlargest(5, 'Medal Count'))
    .reset_index(drop=True)
)

# Add US data for comparison
us_data = avg_bmi[avg_bmi['region'] == 'USA']
comparison_data = pd.concat([us_data, top_countries_by_sport])

# Pivot data for grouped bar plot
pivot_data = comparison_data.pivot(index='region', columns='Event', values='BMI').fil
lna(0)

# Define colors for each event
sports_colors = {
    'Athletics Men\'s Long Jump': '#1E90FF',  # Blue color for Long Jump
    'Athletics Men\'s 110 metres Hurdles': '#FF53E6'  # Pink color for 110 meters Hur
dles
}

# Define x positions for grouped bars
x_positions = np.arange(len(pivot_data.index))  # Number of countries
bar_width = 0.4

# Plot grouped bars
plt.figure(figsize=(14, 6))

for i, event in enumerate(event_of_interest):
    bars = plt.bar(
        x_positions + i * bar_width,  # Adjust positions for each event
        pivot_data[event],  # BMI values for the event
        width=bar_width,
        color=sports_colors[event],
        label=event,
        alpha=0.7
    )

    # Add the y-axis results (BMI) on top of the bars
    for bar in bars:
        yval = bar.get_height()  # Get the height (BMI value) of the bar
        plt.text(
            bar.get_x() + bar.get_width() / 2,  # Position the label at the center of
the bar
```

```
            yval + 0.1,  # Slightly above the top of the bar
            f'{yval:.1f}',  # Format the BMI value to one decimal place
            ha='center',
            va='bottom',  # Position the text at the top of the bar
            fontsize=10
        )

# Add labels and legend
plt.xticks(x_positions + bar_width / 2, pivot_data.index, rotation=45)  # Center labels
plt.title("Comparison of Average BMI in Athletics Events (Men's 110m Hurdles & Long Jump) by Country")
plt.ylabel("Average BMI")
plt.xlabel("Countries")
plt.legend(title="Event", bbox_to_anchor=(1.05, 1))
plt.tight_layout()
plt.show()
```



Conclusion: The Complex Role of BMI in Athletic Success Across Different Sports

The hypothesis that countries with a better BMI (Body Mass Index) tend to win more medals appears to hold some validity when analyzing the data from basketball and athletics events. In basketball, the **USA** stands out with a notably high BMI across both men's and women's events, along with a significant medal count—**222 medals** for men's basketball and **119 medals** for women's basketball. Other countries like **Russia**, **Serbia**, and **Spain** also show relatively higher BMIs, correlating with their respective success in basketball, though the relationship is less direct than in the USA.

In athletics, the data from the **110 meters hurdles** and **long jump** events reveals a similar trend. The **USA**, with moderate BMI values (22.64 for men's 110 meters hurdles and 22.29 for men's long jump), leads the medal count in both events. Other countries such as **Cuba**, **the UK**, and **Germany** show competitive performances, with BMIs generally hovering around or below the average for their respective events. This suggests that while BMI is a factor in athletic performance, it is not the sole predictor of success.

This data supports the idea that certain sports require different physiques. In **basketball**, a higher BMI can be indicative of greater muscle mass and size, which are advantageous for strength, jumping ability, and overall physical presence on the court. Larger BMI values are associated with the power and physicality needed in basketball, which is a contact-heavy sport. Conversely, in **sports like the 110 meters hurdles** or **long jump**, agility, speed, and explosive power are more important than sheer size, which could explain why athletes in these disciplines tend to have lower BMIs.

Overall, while there seems to be a loose correlation between BMI and medal count, especially in **basketball**, this relationship is not entirely conclusive across all events. Other factors, such as **training**, **strategy**, and **historical dominance** in a sport, likely play significant roles in determining a country's medal count. Therefore, while BMI may contribute to athletic success, it should not be viewed as a standalone determinant for predicting medals in international sports. The role of BMI likely varies between sports, with some disciplines favoring athletes with a higher BMI, while others benefit from leaner physiques more conducive to speed and agility.

Comparison of Average Height in Men's and Women's Basketball Across Countries

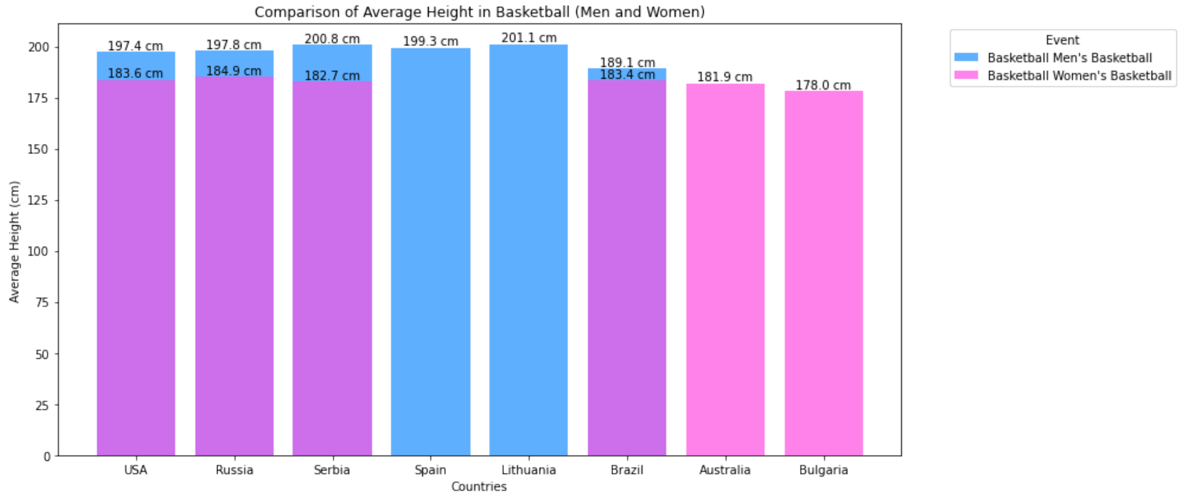
It’s widely acknowledged that body composition, as reflected in BMI (Body Mass Index), plays a significant role in the athletic performance of many sports. However, in basketball, the correlation between BMI and success, particularly in winning gold medals, may not be as strong as initially thought. This raises the question: Could a player’s height be a more significant factor in determining success on the court?

Basketball is a sport that inherently rewards players with greater height, as they can more effectively defend, block shots, and rebound. Tall players also have an advantage when it comes to scoring, with easier access to the basket during both offensive and defensive plays. While a high BMI might indicate strength or muscle mass, it doesn’t directly correlate with performance the way height does in basketball, where players like Michael Jordan, LeBron James, and Shaquille O’Neal have shown that being taller can offer a significant advantage.

In fact, height seems to contribute more directly to a player’s ability to dominate in the paint, alter shots, and secure rebounds—all essential skills for winning critical games, including gold medal matches. This hypothesis suggests that while BMI might offer some insight into a player’s conditioning and athleticism, height could be the true key to unlocking success at the highest levels of competition in basketball.

In [34]:

```python
# Define the events of interest
event_of_interest = ['Basketball Men\'s Basketball', 'Basketball Women\'s Basketbal
l']

# Filter for medal winners in the selected events
medal_winners = athlete_data[
    (athlete_data['Medal'].notnull()) & (athlete_data['Event'].isin(event_of_interes
t))
]

# Group by region and event to calculate average height and medal count
avg_height = (
    medal_winners.groupby(['region', 'Event'])
    .agg({'Height': 'mean', 'Medal': 'count'})
    .rename(columns={'Height': 'Average Height', 'Medal': 'Medal Count'})
    .reset_index()
)

# Identify the top 5 countries by medal count for each sport, excluding the US
top_countries_by_sport = (
    avg_height[avg_height['region'] != 'USA']
    .groupby('Event')
    .apply(lambda x: x.nlargest(5, 'Medal Count'))
    .reset_index(drop=True)
)

# Add US data for comparison
us_data = avg_height[avg_height['region'] == 'USA']
comparison_data = pd.concat([us_data, top_countries_by_sport])

# Define colors for each event
sports_colors = {
    'Basketball Men\'s Basketball': '#1E90FF',
    'Basketball Women\'s Basketball': '#FF53E6'
}

# Plot comparison for the events of interest
plt.figure(figsize=(14, 6))

for event in event_of_interest:
    event_data = comparison_data[comparison_data['Event'] == event]
    countries = event_data['region']
    heights = event_data['Average Height']
    color = sports_colors[event]

    bars = plt.bar(
        countries, heights, color=color, alpha=0.7, label=event
    )

    # Add value labels to bars
    for bar, height in zip(bars, heights):
        plt.text(
            bar.get_x() + bar.get_width() / 2,
            bar.get_height() + 0.1,
            f'{height:.1f} cm',
            ha='center',
            va='bottom',
            fontsize=10
        )

# Customize the plot
plt.title("Comparison of Average Height in Basketball (Men and Women)")
plt.ylabel("Average Height (cm)")
plt.xlabel("Countries")
plt.legend(title="Event", bbox_to_anchor=(1.05, 1), loc='upper left')
plt.tight_layout()
plt.show()
```



Conclusion: Does Height Matter in Winning at Basketball?

Based on the data provided for both men's and women's basketball events, it is evident that height plays a significant role in the success of basketball teams, particularly among medal-winning countries. For instance, the **USA** leads the medal count in both men's and women's basketball, with an average height of **197.45 cm** for men and **183.58 cm** for women. Countries like **Russia**, **Serbia**, and **Spain**, which have relatively taller athletes (average height above **197 cm** for men), also rank highly in terms of medal count.

In **men's basketball**, countries with taller average heights, such as **Serbia** (**200.76 cm**) and **Lithuania** (**201.06 cm**), show a strong correlation between height and success, suggesting that height may be advantageous in terms of winning medals. Similarly, in **women's basketball**, countries with higher average heights, like **Russia** (**184.92 cm**), perform better, although the height advantage is less pronounced than in the men's event.

However, while there is a correlation between height and medal success, it is not a definitive determinant. Other factors, including **skill level**, **strategy**, and **overall team dynamics**, also significantly influence a team's performance. Therefore, while height seems to contribute to success in basketball, it is not the sole factor responsible for winning medals.

General Conclusion: The Interplay of BMI, Height, and Athletic Success

Both **BMI** (Body Mass Index) and **height** appear to have an influence on success in certain sports, though their impact varies across different disciplines. In basketball, both factors seem to correlate with medal success, particularly for countries like the **USA**, **Russia**, and **Serbia**, which perform well in both men's and women's events. The **USA**, for instance, stands out with high average **BMI** values and significant medal counts, while its athletes' average **height**—197.45 cm for men and 183.58 cm for women—further supports their success on the court.

In **basketball**, a higher **BMI** is often associated with greater strength and muscle mass, beneficial for the physical demands of the sport, such as jumping ability and overall presence on the court. Similarly, taller athletes, as seen with **Serbia** and **Lithuania**, who have average heights of over 200 cm for men, also show a clear advantage in this sport, suggesting that height plays a role in winning medals. However, the relationship between BMI, height, and success in **basketball** is not absolute—other elements like **skill level**, **strategy**, and team dynamics are crucial to performance.

In contrast, **athletics** events like the **110 meters hurdles** and **long jump** show that while **BMI** may play a role, it is less predictive of success than in basketball. Here, athletes with moderate or lower **BMI** values tend to excel, suggesting that agility, speed, and explosive power—rather than size—are more critical for success. Thus, in certain disciplines like the hurdles or long jump, leaner physiques may be more advantageous, emphasizing the varying role of BMI across sports.

Overall, while both **BMI** and **height** can contribute to athletic success, they are not definitive factors. The complexity of sports performance means that a range of other aspects, including training, mental fortitude, and historical dominance, are likely more significant in determining a country's medal count. Therefore, **BMI** and **height** should be viewed as part of a larger puzzle, where they may help to explain success in some sports but cannot serve as the sole predictor of achievement.

Step 6: Machine Learning

In [35]:
```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import mean_squared_error, classification_report, confusion_matr
ix

# Filtering for relevant features for prediction tasks
selected_features = ['Age', 'Height', 'Weight', 'Gold', 'Silver', 'Bronze', 'Sport',
'Medal']
data = athlete_data[selected_features].dropna()

# Encoding categorical variables
data = pd.get_dummies(data, columns=['Sport'], drop_first=True)

# Defining Features and Target Variables
X_regression = data[['Age', 'Height', 'Weight']]  # Features for regression
y_regression = data[['Gold', 'Silver', 'Bronze']]  # Targets for regression

X_classification = data.drop(columns=['Medal'])
y_classification = data['Medal']

# Splitting the Data
X_train_reg, X_test_reg, y_train_reg, y_test_reg = train_test_split(X_regression, y_r
egression, test_size=0.2, random_state=42)
X_train_class, X_test_class, y_train_class, y_test_class = train_test_split(X_classif
ication, y_classification, test_size=0.2, random_state=42)

# --- Regression Model ---
print("\nTraining Linear Regression Model for Medal Count Prediction...")
reg_model = LinearRegression()
reg_model.fit(X_train_reg, y_train_reg)
```

```
Training Linear Regression Model for Medal Count Prediction...
```

Out[35]:
```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)
```

Evaluating Regression Model Performance with Mean Squared Error

Introduction:

Evaluating the performance of machine learning models is crucial for ensuring their accuracy and effectiveness. In this analysis, we use the Mean Squared Error (MSE) as a key metric to assess the predictive capability of a regression model. By comparing the predicted and actual values from the test dataset, the MSE quantifies the average squared difference, providing insight into the model's precision and overall fit to the data. A lower MSE value typically indicates better model performance.

In [36]:
```
# Prediction and Evaluation
y_pred_reg = reg_model.predict(X_test_reg)
mse = mean_squared_error(y_test_reg, y_pred_reg)
print("Mean Squared Error (Regression):", mse)
```

```
Mean Squared Error (Regression): 0.2222058072403971
```

Conclusion for Prediction and Evaluation

The regression model has a Mean Squared Error (MSE) of 0.2222. This indicates that, on average, the squared differences between the predicted and actual values are relatively small. The lower the MSE, the better the model fits the data. While this result suggests that the model performs reasonably well, further evaluation using additional metrics (such as R-squared or MAE) and validation with real-world scenarios can provide a more comprehensive assessment of its effectiveness.

Building and Training a Random Forest Classifier for Medal Prediction

Introduction:

Predictive classification models play a critical role in decision-making processes across various fields. In this analysis, a Random Forest Classifier is employed to predict medal outcomes. Random Forest, an ensemble learning method, leverages multiple decision trees to enhance predictive accuracy and reduce overfitting. This powerful algorithm is trained using labeled data to classify instances and make accurate medal predictions.

In [37]:
```
# --- Classification Model ---
print("\nTraining Random Forest Classifier for Medal Prediction...")
class_model = RandomForestClassifier()
class_model.fit(X_train_class, y_train_class)
```

```
Training Random Forest Classifier for Medal Prediction...
```

Out[37]:
```
RandomForestClassifier(bootstrap=True, ccp_alpha=0.0, class_weight=None,
                       criterion='gini', max_depth=None, max_features='auto',
                       max_leaf_nodes=None, max_samples=None,
                       min_impurity_decrease=0.0, min_impurity_split=None,
                       min_samples_leaf=1, min_samples_split=2,
                       min_weight_fraction_leaf=0.0, n_estimators=100,
                       n_jobs=None, oob_score=False, random_state=None,
                       verbose=0, warm_start=False)
```

Conclusion:

The training of the Random Forest Classifier has been successfully completed with default hyperparameters, including 100 decision trees (n_estimators=100) and the Gini impurity criterion. While the current model has been fitted, further evaluation through metrics such as accuracy, precision, recall, and F1 score is recommended. Additionally, hyperparameter tuning and cross-validation may help optimize the model for improved prediction performance.

Evaluation of Classification Model Performance

Introduction:

In this analysis, we evaluate the performance of a classification model by examining the Classification Report and Confusion Matrix. The classification report provides key metrics such as precision, recall, and F1-score for each class, while the confusion matrix gives a detailed breakdown of correctly and incorrectly predicted instances. These evaluation techniques help assess the model's accuracy and effectiveness in distinguishing between different categories: Bronze, Silver, and Gold.

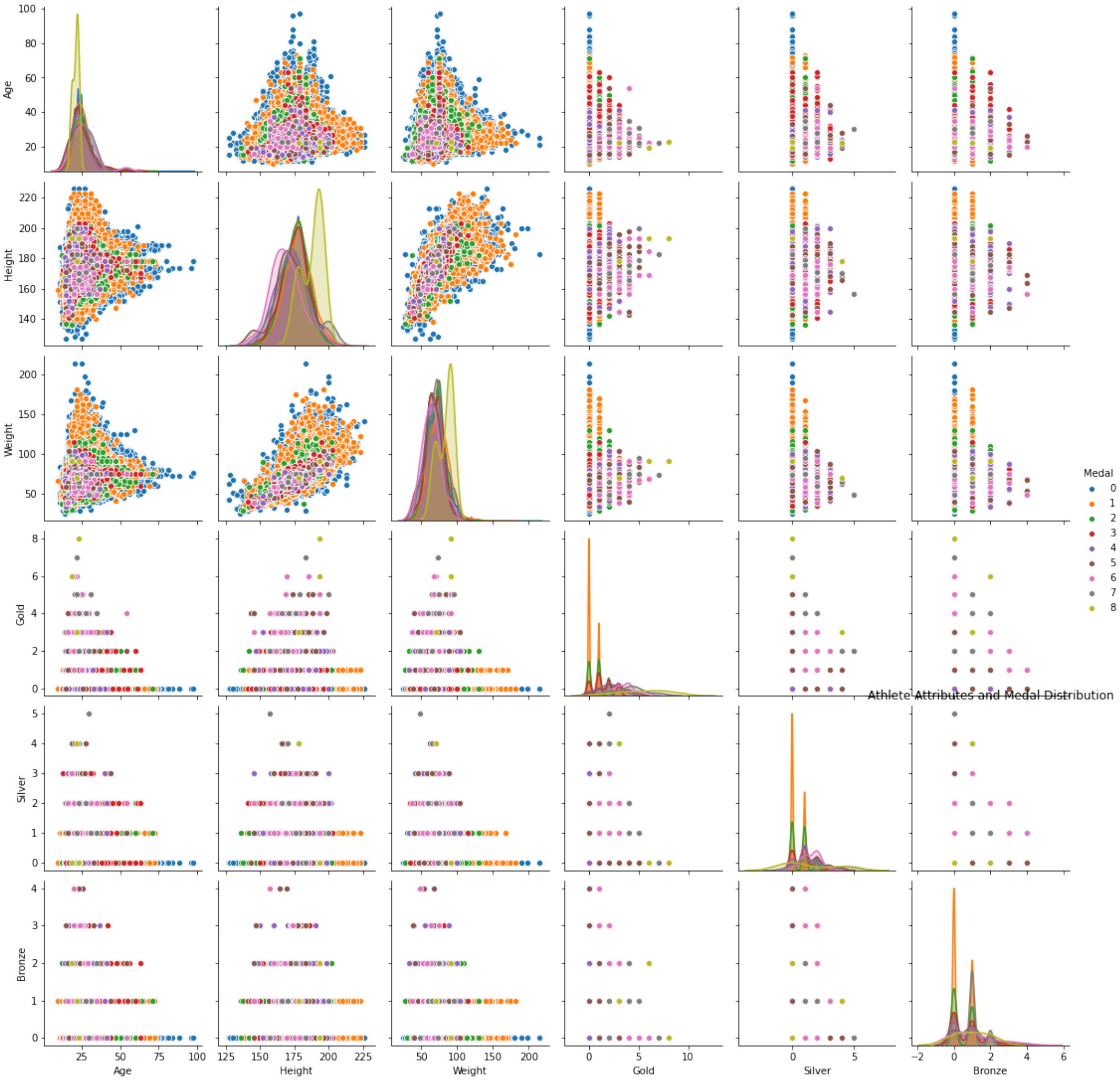
In [38]:
```python
# Prediction and Evaluation
y_pred_class = class_model.predict(X_test_class)
print("\nClassification Report:")
print(classification_report(y_test_class, y_pred_class))

print("Confusion Matrix:")
print(confusion_matrix(y_test_class, y_pred_class))
```

```
Classification Report:
              precision    recall  f1-score   support

      Bronze       1.00      1.00      1.00      2568
        Gold       1.00      1.00      1.00      2672
      Silver       1.00      1.00      1.00      2571

    accuracy                           1.00      7811
   macro avg       1.00      1.00      1.00      7811
weighted avg       1.00      1.00      1.00      7811

Confusion Matrix:
[[2568    0    0]
 [   0 2672    0]
 [   0    0 2571]]
```

Conclusion:

The classification model achieved perfect performance across all metrics, with a precision, recall, and F1-score of 1.00 for every class and an overall accuracy of 100% on the test dataset. The confusion matrix further supports this result, showing no misclassifications for any class. Such results indicate an exceptionally well-trained model, though it may warrant further validation to ensure the absence of data leakage or overfitting.

In [39]:
```python
# --- Data Visualization ---
sns.pairplot(athlete_data_rem_dupli, vars=['Age', 'Height', 'Weight', 'Gold', 'Silve
r', 'Bronze'], hue='Medal')
plt.title("Athlete Attributes and Medal Distribution")
plt.show()
```



Optimal Athlete Profiles: Correlations Between Age, Height, Weight, and Medal Performance

From the data visualization, we can observe several key trends:

- **Age and Medal Performance:** Athletes in their mid-20s, particularly around 25 years of age, tend to win more medals. There is a noticeable decline in medal performance as athletes age, especially in the Gold category. This suggests that younger athletes are more likely to achieve higher medals, with the best age for optimal performance appearing to be around 25 years old.

- **Height and Medal Distribution:** Athletes with a height between 160 cm and 180 cm appear to have the highest concentration of medals, especially in the Silver and Bronze categories. However, athletes who aim for Gold tend to be slightly taller, with an optimal height around 180 cm, indicating that height plays a role in reaching the top tier of medals.

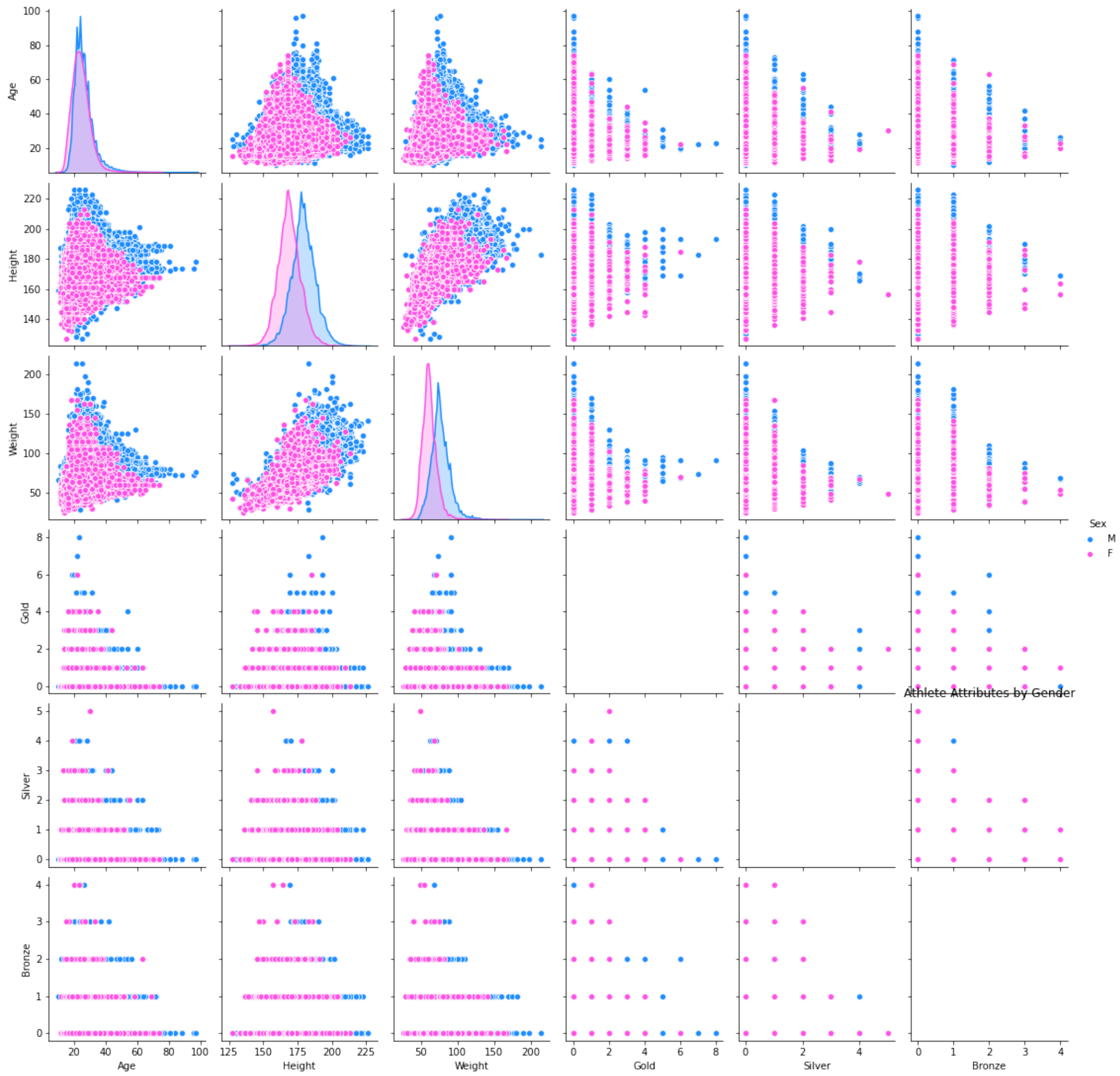
- **Weight and Medal Concentration:** For weight, the highest concentration of medals is seen in athletes weighing between 50 kg and 100 kg, with the ideal weight for medal success hovering around 75 kg. This trend suggests that athletes within this weight range tend to perform better overall.

- **Gold vs. Other Medals:** As Gold medals increase, the number of Silver and Bronze medals tends to decrease, and vice versa. This indicates a competitive balance between medal types, where excelling in one category may detract from performance in others.

These insights provide valuable correlations between physical attributes like age, height, and weight and an athlete's likelihood of winning medals, particularly Gold.

In [40]:

```
# Define the color map for genders
color_map = {'F': '#FF53E6', 'M': '#1E90FF'}

# --- Data Visualization ---
sns.pairplot(athlete_data_rem_dupli, vars=['Age', 'Height', 'Weight', 'Gold', 'Silve
r', 'Bronze'], hue='Sex', palette=color_map)
plt.title("Athlete Attributes by Gender")
plt.show()
```



Optimal Physical Attributes for Medal-Winning Female and Male Athletes

Analyzing Ideal Age, Height, and Weight for Gold, Silver, and Bronze

Female Athletes:

- **Ideal Athlete Profile:**
  - **Age:** Around 24 to 25 years seems to be common among top performers.
  - **Height:** Heights around 170 cm are typical for those with medals.
  - **Weight:** Weights ranging from 60 to 64 kg are ideal.

- **Medal Candidate:**
  - **Gold Medal Candidate:** Female athletes around 24.34 years old, 168.66 cm, and 60.98 kg have the closest attributes to the top performers.
  - **Silver Medal Candidate:** Female athletes around 24.97 years old, 170.39 cm, and 63.39 kg could be a silver medal candidate, showing competitive figures.
  - **Bronze Medal Candidate:** Female athletes around 23.45 years old, 168.97 cm, and 60.25 kg would likely be a strong contender for a bronze medal.

Male Athletes:

- **Ideal Athlete Profile**
  - **Age:** Around 26 to 27 years is ideal.
  - **Height:** Heights around 179 cm (5'10") are ideal.
  - **Weight:** Weights between 75 and 79 kg are suitable.

- **Medal Candidate:**
  - **Gold Medal Candidate:** Male athletes around 26.24 years old, 179.20 cm, and 76.49 kg are closely aligned with top athletes for gold.
  - **Silver Medal Candidate:** Male athletes around 26.55 years old, 180.60 cm, and 78.96 kg could be in the running for silver.
  - **Bronze Medal Candidate:** Male athletes around 26.63 years old, 179.35 cm, and 75.70 kg are also strong contenders for a bronze medal.

Conclusion:

- **For Gold:**
  - **Female:** 24.34 years old, 168.66 cm, 60.98 kg
  - **Male:** 26.24 years old, 179.20 cm, 76.49 kg

- **For Silver:**
  - **Female:** 24.97 years old, 170.39 cm, 63.39 kg
  - **Male:** 26.55 years old, 180.60 cm, 78.96 kg

- **For Bronze:**
  - **Female:** 23.45 years old, 168.97 cm, 60.25 kg
  - **Male:** 26.63 years old, 179.35 cm, 75.70 kg

This analysis suggests that the ideal athletes for each medal category are those whose physical attributes (age, height, and weight) are close to the mean for their respective genders.

In [41]:
```python
# Group by 'Sex' and 'Medal' and calculate the average for Age, Height, and Weight
average_stats = athlete_data_rem_dupli.groupby(['Sex', 'Gold', 'Silver', 'Bronze'])
[['Age', 'Height', 'Weight']].mean()

# Display the results
print(average_stats)
```

```
                              Age      Height     Weight
Sex Gold Silver Bronze                                  
F   0    0      0       24.340682  168.655913  60.979976
                1       24.974034  170.391469  63.388249
                2       23.452055  168.973607  60.252648
                3       25.000000  186.000000  75.000000
         1      0       24.636078  170.709894  63.656241
...                           ...         ...        ...
M   5    1      1       26.500000  189.500000  85.136667
    6    0      0       20.000000  169.000000  68.000000
                2       19.000000  193.000000  91.000000
    7    0      0       22.000000  183.000000  73.000000
    8    0      0       23.000000  193.000000  91.000000

[118 rows x 3 columns]
```

In [42]:
```
# Ensure all rows will be printed
pd.set_option('display.max_rows', None)

# Group by 'Sex', 'Sport', 'Gold', 'Silver', 'Bronze' and calculate the mean of 'Ag
e', 'Height', and 'Weight'
average_stats = athlete_data_rem_dupli.groupby(['Sex', 'Sport', 'Gold', 'Silver', 'Br
onze'])[['Age', 'Height', 'Weight']].mean()

# Display the results
average_stats

# Reset to default row limit if needed
#pd.reset_option('display.max_rows')
```

Out[42]:

Sex	Sport	Gold	Silver	Bronze	Age	Height	Weight
F	Alpine Skiing	0	0	0	22.304985	167.096233	62.333871
				1	23.119048	167.358730	62.716667
				2	22.800000	167.746667	63.273333
			1	0	22.918919	167.389640	62.949099
				1	24.222222	167.888889	64.111111
				2	27.000000	166.966667	62.066667
			2	0	23.166667	169.161111	62.177778
		1	0	0	22.611111	166.604630	62.401852
				1	23.500000	175.000000	72.500000
				2	24.000000	170.000000	81.000000
			1	0	23.666667	168.888889	65.011111
				1	29.000000	170.000000	68.000000
		2	0	0	24.000000	172.000000	67.285714
				1	25.000000	171.000000	64.000000
			1	0	24.000000	162.000000	56.000000
		3	1	0	20.000000	175.000000	76.000000
	Alpinism	1	0	0	43.000000	167.600000	59.800000
	Archery	0	0	0	27.095486	167.017101	61.613281
				1	25.870968	166.019355	61.212903
				2	22.000000	168.000000	64.000000
			1	0	25.451613	168.793548	63.574194
			2	0	55.000000	167.600000	59.800000
		1	0	0	28.857143	168.192857	62.635714
				1	24.250000	166.000000	61.500000
				2	63.000000	167.600000	59.800000
			1	0	21.666667	167.333333	64.500000
		2	0	0	22.285714	165.928571	59.471429
		3	0	0	44.000000	167.600000	59.800000
	Art Competitions	0	0	0	43.673077	167.551282	59.800000
				1	48.000000	167.600000	59.800000
			1	0	37.750000	167.600000	59.800000
		1	0	0	34.000000	167.600000	59.800000

					Age	Height	Weight
Sex	**Sport**	**Gold**	**Silver**	**Bronze**			
	Athletics	**0**	**0**	**0**	25.184777	169.352567	60.399591
				1	25.164306	170.522049	62.932011
				2	21.750000	170.500000	62.000000
			1	**0**	25.135693	170.508456	62.671485
				1	25.200000	169.500000	59.050000
				2	23.000000	176.500000	62.500000
			2	**0**	24.181818	168.818182	58.000000
				1	32.000000	181.500000	67.500000
		1	**0**	**0**	25.178451	170.979461	63.688384
				1	25.363636	170.136364	57.590909
			1	**0**	24.791667	168.833333	58.958333
				1	26.750000	170.250000	62.250000
			2	**0**	22.000000	169.000000	57.250000
		2	**0**	**0**	25.950000	170.750000	60.750000
				1	23.500000	173.500000	65.500000
			1	**0**	24.857143	169.000000	60.714286
		3	**0**	**0**	22.000000	171.750000	58.500000
			1	**0**	28.000000	170.000000	57.000000
		4	**0**	**0**	30.000000	175.000000	63.000000
	Badminton	**0**	**0**	**0**	25.026515	168.180871	61.411711
				1	24.428571	169.271429	61.503571
			1	**0**	24.291667	169.479167	62.416667
		1	**0**	**0**	24.750000	168.890000	61.800000
				1	21.500000	167.500000	61.500000
			1	**0**	26.000000	172.933333	65.683333
		2	**0**	**0**	25.000000	173.000000	60.000000
	Basketball	**0**	**0**	**0**	25.491309	182.202781	73.675782
				1	25.136364	183.143939	72.696970
			1	**0**	25.053846	181.846154	73.238462
		1	**0**	**0**	26.534351	184.026718	75.191603
	Beach Volleyball	**0**	**0**	**0**	28.129167	178.693750	68.254167
				1	29.500000	180.833333	69.333333
			1	**0**	29.916667	178.241667	66.866667
		1	**0**	**0**	29.250000	181.000000	70.833333

					Age	Height	Weight
Sex	Sport	Gold	Silver	Bronze			
	Biathlon	0	0	0	25.024299	166.487165	57.289704
				1	26.862069	166.689655	56.168966
				2	30.000000	168.666667	57.000000
			1	0	27.041667	168.041667	58.916667
				1	23.000000	167.750000	54.000000
			2	0	25.333333	169.666667	61.666667
			3	0	26.000000	158.000000	48.000000
		1	0	0	26.000000	167.624603	58.853968
				1	27.200000	165.200000	58.000000
				2	26.000000	161.500000	57.000000
			1	0	28.800000	169.200000	62.200000
				1	29.500000	164.500000	56.500000
			2	0	26.500000	173.000000	63.000000
		2	0	0	27.000000	161.666667	53.333333
			1	0	24.000000	169.500000	59.000000
		3	0	0	27.000000	168.000000	58.000000
	Bobsleigh	0	0	0	27.579832	173.092437	72.537815
				1	29.250000	174.875000	73.750000
			1	0	29.375000	173.375000	74.625000
		1	0	0	28.625000	172.625000	74.000000
	Boxing	0	0	0	26.875000	169.262500	62.047917
				1	25.833333	167.333333	61.925000
			1	0	28.166667	169.500000	62.000000
		1	0	0	24.833333	168.500000	62.000000
	Canoeing	0	0	0	24.847172	169.003700	64.186462
				1	26.166667	170.171795	65.483333
			1	0	25.732143	170.551786	65.989286
				1	26.000000	170.412500	64.218750
			2	0	26.000000	170.333333	65.833333
		1	0	0	26.882353	170.411765	65.178431
				1	28.166667	171.833333	67.666667
			1	0	28.071429	172.928571	66.928571
				1	27.500000	170.500000	67.000000
		2	0	0	26.857143	171.571429	66.857143
			1	0	24.500000	172.000000	68.500000
		3	0	0	29.000000	168.000000	63.000000
	Croquet	0	0	0	39.500000	167.600000	59.800000

Sex	Sport	Gold	Silver	Bronze	Age	Height	Weight
	Cross Country Skiing	0	0	0	24.866388	166.540894	57.492759
				1	27.261538	167.658462	57.970000
				2	31.200000	165.420000	57.090000
			1	0	27.456522	167.847826	58.434783
				1	27.636364	165.100000	57.018182
			2	0	28.000000	165.344444	57.494444
				1	34.000000	169.000000	59.500000
		1	0	0	26.589744	166.105128	57.822222
				1	27.090909	166.545455	58.363636
				2	23.000000	174.000000	63.000000
				4	23.000000	164.000000	54.000000
			1	0	27.125000	164.000000	56.680208
				1	27.400000	165.200000	53.200000
			2	0	27.000000	159.500000	57.750000
		2	0	0	26.000000	162.500000	53.500000
			1	0	25.750000	163.500000	54.750000
			2	1	31.000000	164.000000	55.000000
		3	0	0	28.666667	162.333333	60.333333
				1	28.000000	176.000000	68.500000
			1	0	27.000000	167.000000	58.000000
				1	30.500000	167.500000	60.500000
			2	0	25.000000	167.000000	58.000000
	Curling	0	0	0	28.792208	167.399351	62.185714
				1	31.434783	166.260870	63.000000
			1	0	32.434783	169.043478	64.773913
		1	0	0	34.136364	168.090909	65.727273
	Cycling	0	0	0	26.530769	167.619286	58.771648
				1	27.191489	169.659574	62.744681
				2	27.500000	175.000000	66.500000
			1	0	27.577778	168.400000	61.022222
				1	32.000000	175.000000	70.000000
			2	0	28.000000	171.000000	65.333333
				1	26.000000	170.000000	70.000000
		1	0	0	26.800000	171.325000	63.500000
				1	28.000000	165.500000	65.166667
			1	0	31.666667	162.666667	52.333333
		2	0	0	24.333333	165.333333	60.666667
		3	1	0	30.000000	168.000000	69.000000

Sex	Sport	Gold	Silver	Bronze	Age	Height	Weight
	Diving	**0**	**0**	**0**	21.458886	161.286649	54.102056
				1	23.245283	161.400000	54.745283
				2	22.500000	157.500000	49.500000
			1	**0**	20.652174	160.435870	53.095652
				1	21.857143	159.028571	53.621429
			2	**0**	24.333333	160.000000	49.000000
		1	**0**	**0**	19.419355	160.067742	52.117742
				1	21.250000	154.562500	46.675000
			1	**0**	19.111111	160.916667	51.994444
		2	**0**	**0**	21.545455	161.386364	51.790909
	Equestrianism	**0**	**0**	**0**	34.188540	167.292514	58.182810
				1	35.693878	169.312245	60.603061
				2	37.000000	175.000000	71.000000
			1	**0**	33.076923	169.071538	59.710769
				1	33.375000	167.750000	58.500000
			2	**0**	28.000000	172.000000	56.000000
		1	**0**	**0**	33.344828	168.905172	57.224138
				1	35.500000	171.166667	59.666667
			1	**0**	35.500000	169.166667	60.333333
		2	**0**	**0**	24.750000	171.250000	60.750000
	Fencing	**0**	**0**	**0**	26.481447	168.344852	60.320501
				1	27.244444	169.276667	60.992778
				2	23.000000	166.666667	60.000000
			1	**0**	25.552083	170.658333	61.418750
				1	26.875000	169.250000	61.250000
			2	**0**	28.000000	167.500000	56.500000
		1	**0**	**0**	26.156627	169.563855	60.355422
				1	27.700000	169.900000	62.700000
			1	**0**	26.000000	170.250000	60.125000
		2	**0**	**0**	23.555556	170.666667	59.777778
	Figure Skating	**0**	**0**	**0**	20.322126	160.537310	49.976952
				1	22.783333	160.475000	50.014167
			1	**0**	21.883333	159.886667	49.546667
			2	**0**	24.000000	165.000000	53.000000
		1	**0**	**0**	22.542373	160.128814	48.779661
				1	24.000000	161.033333	49.433333
			1	**0**	21.000000	158.000000	48.200000
		2	**0**	**0**	27.000000	159.000000	44.000000
	Football	**0**	**0**	**0**	24.542254	167.236901	60.514507
				1	25.656863	170.438235	63.616667
			1	**0**	25.323232	166.929293	59.929293
		1	**0**	**0**	26.326733	168.712871	62.069307

					Age	Height	Weight
Sex	**Sport**	**Gold**	**Silver**	**Bronze**			
	Freestyle Skiing	**0**	**0**	**0**	24.217848	164.729134	58.332546
				1	25.588235	164.882353	58.235294
			1	**0**	25.176471	164.588235	58.205882
		1	**0**	**0**	24.823529	165.529412	59.500000
	Golf	**0**	**0**	**0**	27.403226	168.782258	63.058065
				1	33.500000	169.350000	74.200000
			1	**0**	21.000000	166.850000	63.400000
		1	**0**	**0**	25.500000	168.350000	64.200000

					Age	Height	Weight
Sex	**Sport**	**Gold**	**Silver**	**Bronze**			
	Gymnastics	**0**	**0**	**0**	19.276786	156.292128	48.097453
				1	20.250000	155.498387	47.372003
				2	18.666667	156.700000	49.583333
			1	**0**	18.867347	155.362136	46.863484
				1	21.555556	157.164286	47.668651
				2	17.500000	156.000000	42.500000
				3	17.000000	150.000000	40.000000
			2	**0**	19.400000	153.600000	44.400000
				1	15.666667	145.000000	39.666667
				3	15.000000	147.000000	39.000000
			3	**0**	25.000000	158.000000	48.000000
		1	**0**	**0**	20.113924	153.855289	45.072333
				1	19.727273	155.967532	47.246753
				2	18.666667	152.666667	42.666667
				3	27.000000	160.000000	55.000000
			1	**0**	21.952381	155.102041	46.414966
				1	23.666667	159.547619	51.238095
				2	19.500000	158.440476	46.285714
				4	20.000000	156.642857	48.714286
			2	**0**	19.250000	155.500000	45.750000
				1	25.333333	160.666667	54.333333
				2	16.000000	145.000000	42.000000
			3	**0**	16.000000	145.000000	41.000000
				1	18.000000	160.000000	45.000000
		2	**0**	**0**	19.000000	151.666667	44.666667
				1	17.600000	156.820000	47.550000
				2	19.000000	158.000000	47.000000
			1	**0**	20.250000	151.910714	44.678571
				1	19.000000	158.314286	49.214286
			2	**0**	22.500000	159.321429	46.857143
				1	20.000000	156.000000	46.000000
				2	29.000000	161.000000	52.000000
			5	**0**	30.000000	156.642857	48.714286
		3	**0**	**0**	16.000000	160.000000	47.000000
			1	**0**	19.000000	154.666667	48.000000
				1	14.000000	162.000000	45.000000
			2	**1**	20.500000	153.000000	45.000000
		4	**0**	**1**	19.000000	143.000000	47.000000
			1	**0**	16.000000	145.000000	40.000000
				1	21.000000	161.000000	52.000000
			2	**0**	30.500000	158.321429	53.357143

					Age	Height	Weight
Sex	**Sport**	**Gold**	**Silver**	**Bronze**			
	Handball	**0**	**0**	**0**	26.054645	174.976393	69.080546
				1	26.867089	174.686076	68.798734
			1	**0**	24.490566	173.836478	67.723270
		1	**0**	**0**	25.303226	175.049032	68.883871
	Hockey	**0**	**0**	**0**	25.173293	165.774822	60.466871
				1	26.149068	166.947826	61.024845
			1	**0**	24.880503	166.608805	59.698113
		1	**0**	**0**	25.943038	166.984810	61.012658
	Ice Hockey	**0**	**0**	**0**	23.814978	167.281938	64.079295
				1	23.193878	167.653061	65.760204
			1	**0**	24.237624	170.603960	69.267327
		1	**0**	**0**	25.792079	170.524752	69.455446
	Judo	**0**	**0**	**0**	25.192865	166.305463	66.901003
				1	24.897959	166.324490	68.006122
			1	**0**	25.040816	166.273469	68.879592
		1	**0**	**0**	25.285714	165.110204	67.785714
	Luge	**0**	**0**	**0**	23.616822	168.454829	66.553894
				1	23.533333	173.733333	69.266667
			1	**0**	24.333333	173.400000	69.800000
		1	**0**	**0**	24.692308	170.615385	66.538462
		2	**0**	**0**	26.000000	183.000000	76.000000
	Modern Pentathlon	**0**	**0**	**0**	25.496644	169.731544	58.114094
				1	26.200000	175.000000	61.800000
			1	**0**	25.400000	175.200000	62.800000
		1	**0**	**0**	25.800000	170.200000	56.200000
	Motorboating	**0**	**0**	**0**	26.000000	167.600000	59.800000
	Rhythmic Gymnastics	**0**	**0**	**0**	18.688213	167.147148	48.609316
				1	19.250000	170.754545	49.422727
			1	**0**	19.113636	169.845455	49.061364
		1	**0**	**0**	18.431818	171.113636	49.340909
	Rowing	**0**	**0**	**0**	25.252303	176.274415	69.204252
				1	25.536697	177.511927	71.034404
			1	**0**	26.009434	177.891509	72.052594
				1	24.421053	174.421053	69.000000
			2	**0**	28.333333	179.666667	75.333333
		1	**0**	**0**	25.528571	178.056190	72.507143
				1	30.500000	180.250000	74.500000
			1	**0**	26.250000	172.750000	67.750000
		2	**0**	**0**	27.200000	178.300000	74.200000

					Age	Height	Weight
Sex	**Sport**	**Gold**	**Silver**	**Bronze**			
	Rugby Sevens	**0**	**0**	**0**	26.535714	167.055357	65.696429
				1	25.500000	168.000000	68.833333
			1	**0**	24.000000	170.250000	71.333333
		1	**0**	**0**	24.916667	170.083333	68.416667
	Sailing	**0**	**0**	**0**	26.596713	169.299874	62.708976
				1	27.162791	170.688372	62.506977
			1	**0**	26.976744	171.339535	63.504651
		1	**0**	**0**	27.422222	169.528889	61.964444
			1	**0**	32.000000	167.600000	59.800000
	Shooting	**0**	**0**	**0**	28.965909	164.919508	60.989489
				1	30.351351	165.405405	63.270270
				2	27.000000	160.000000	52.000000
			1	**0**	28.097561	166.682927	63.463415
				1	30.333333	164.000000	59.000000
		1	**0**	**0**	27.194444	166.138889	66.513889
				1	24.833333	164.166667	59.166667
			1	**0**	24.600000	163.000000	54.400000
		2	**0**	**0**	30.000000	154.000000	63.000000
	Short Track Speed Skating	**0**	**0**	**0**	22.340517	164.062931	56.407471
				1	22.750000	165.785714	58.000000
				2	19.500000	165.500000	60.500000
			1	**0**	23.307692	165.884615	58.988462
				1	22.600000	168.000000	61.600000
				2	23.000000	164.000000	63.000000
			2	**0**	24.500000	166.500000	63.500000
			3	**0**	20.000000	165.000000	60.000000
		1	**0**	**0**	20.380952	163.312698	55.626984
				1	23.000000	162.362500	56.262500
			1	**0**	20.500000	163.750000	58.750000
				1	18.500000	170.000000	58.000000
		2	**0**	**0**	20.000000	170.000000	63.000000
				1	21.500000	165.500000	56.000000
			1	**0**	25.000000	166.000000	58.000000
		3	**0**	**0**	20.500000	165.500000	58.000000
	Skeleton	**0**	**0**	**0**	27.648148	167.775926	60.962963
				1	25.000000	167.500000	56.500000
			1	**0**	30.000000	172.000000	66.000000
		1	**0**	**0**	26.500000	166.250000	61.000000
	Ski Jumping	**0**	**0**	**0**	21.037037	164.370370	52.125926
				1	18.000000	165.000000	57.000000
			1	**0**	30.000000	164.000000	52.000000
		1	**0**	**0**	22.000000	171.000000	62.000000

Sex	Sport	Gold	Silver	Bronze	Age	Height	Weight
	Snowboarding	**0**	**0**	**0**	24.571865	165.996942	60.231346
				1	26.266667	168.066667	60.266667
			1	**0**	24.600000	167.866667	59.800000
		1	**0**	**0**	24.000000	167.733333	63.000000
	Softball	**0**	**0**	**0**	26.023411	168.511706	66.110368
				1	27.508475	171.169492	68.406780
			1	**0**	26.716667	170.133333	68.416667
		1	**0**	**0**	26.066667	171.316667	72.408333
	Speed Skating	**0**	**0**	**0**	23.204703	167.133645	61.754968
				1	25.731707	167.634146	62.614634
				2	27.666667	167.833333	62.666667
			1	**0**	25.090909	168.012121	62.045455
				1	23.333333	169.666667	60.833333
				2	33.000000	168.000000	60.000000
			2	**0**	26.400000	170.600000	64.600000
				1	27.000000	169.000000	62.666667
		1	**0**	**0**	25.888889	169.012037	64.145062
				1	22.625000	168.750000	64.375000
			1	**0**	26.363636	169.454545	62.727273
				1	25.000000	163.000000	59.000000
			2	**0**	25.750000	167.750000	64.000000
				2	26.000000	173.000000	71.000000
		2	**0**	**0**	25.600000	166.600000	60.200000
				1	22.000000	171.000000	53.000000
				2	25.000000	165.000000	61.000000
			1	**0**	25.000000	170.000000	65.000000
			2	**0**	22.000000	180.000000	72.000000
			3	**0**	27.000000	168.000000	65.000000
		3	**0**	**0**	23.000000	166.000000	66.000000
		4	**0**	**0**	24.000000	163.000000	59.000000

					Age	Height	Weight
Sex	**Sport**	**Gold**	**Silver**	**Bronze**			
	Swimming	**0**	**0**	**0**	19.509146	170.655250	60.820024
				1	19.552301	172.763668	63.070642
				2	20.000000	173.577826	63.083768
				3	25.000000	186.000000	75.000000
			1	**0**	19.275701	172.820016	63.055452
				1	20.466667	173.103333	64.106667
				2	20.111111	174.444444	63.007407
			2	**0**	20.000000	175.479365	65.479365
				1	21.600000	174.800000	66.800000
				2	14.000000	181.000000	66.000000
			3	**0**	27.000000	177.500000	63.500000
				1	26.000000	183.000000	62.000000
		1	**0**	**0**	19.463087	173.546421	63.505593
				1	19.928571	170.992262	61.476190
				2	21.000000	170.800000	64.280000
			1	**0**	20.285714	173.638095	64.040476
				1	20.000000	175.714286	64.642857
				2	23.000000	179.500000	61.500000
			2	**0**	20.647059	175.466176	65.349020
				1	22.000000	180.000000	60.000000
				3	25.000000	173.000000	63.000000
			3	**0**	25.500000	173.500000	63.500000
				1	24.000000	176.000000	69.000000
			4	**0**	19.000000	178.000000	67.000000
		2	**0**	**0**	18.513514	172.868919	62.275676
				1	17.800000	175.200000	67.000000
				2	23.000000	167.000000	63.500000
				3	33.000000	183.000000	68.000000
			1	**0**	20.000000	174.969697	64.393939
				1	21.666667	172.333333	60.666667
			2	**0**	16.000000	174.250000	67.750000
				1	18.500000	174.000000	63.500000
		3	**0**	**0**	18.500000	173.790741	62.281481
				1	26.500000	168.500000	70.000000
			1	**0**	21.000000	175.090625	63.962500
				1	19.333333	175.333333	66.000000
		4	**0**	**0**	23.000000	183.000000	74.000000
				1	17.000000	188.000000	77.000000
			1	**0**	18.000000	177.500000	66.000000
		6	**0**	**0**	22.000000	185.000000	70.000000

					Age	Height	Weight
Sex	**Sport**	**Gold**	**Silver**	**Bronze**			
	Synchronized Swimming	**0**	**0**	**0**	22.012915	168.290683	55.582103
				1	22.673077	167.230769	55.176923
				2	23.428571	167.571429	55.857143
			1	**0**	23.803922	168.372549	57.431373
				1	24.500000	171.500000	55.750000
			2	**0**	26.625000	169.625000	56.125000
		1	**0**	**0**	24.080000	169.740000	56.920000
		2	**0**	**0**	23.833333	171.083333	56.333333
	Table Tennis	**0**	**0**	**0**	25.907018	165.298684	58.363421
				1	24.142857	163.480952	55.395238
				2	23.333333	164.683333	55.050000
			1	**0**	25.200000	164.150000	58.100000
				1	24.000000	168.000000	60.000000
		1	**0**	**0**	23.200000	164.600000	55.800000
				1	23.500000	163.500000	58.000000
			1	**0**	24.166667	167.833333	60.500000
		2	**0**	**0**	23.000000	163.857143	57.428571
	Taekwondo	**0**	**0**	**0**	23.577093	170.514978	61.176652
				1	22.531250	172.750000	62.421875
			1	**0**	24.100000	170.250000	60.075000
		1	**0**	**0**	22.200000	171.750000	61.350000
	Tennis	**0**	**0**	**0**	24.243548	171.851801	61.835242
				1	26.742857	172.341429	61.682381
				2	27.500000	172.900000	62.900000
			1	**0**	26.483871	174.051613	63.170968
				1	24.600000	171.190000	59.796667
			2	**0**	36.000000	172.700000	62.633333
		1	**0**	**0**	25.954545	175.654545	66.242424
				1	20.800000	173.780000	60.580000
			1	**0**	31.000000	172.900000	62.900000
				1	23.000000	172.700000	62.633333
		2	**0**	**0**	27.833333	174.283333	65.450000
				1	20.000000	172.700000	62.633333
	Trampolining	**0**	**0**	**0**	25.836066	162.032787	53.049180
				1	22.600000	159.200000	50.200000
			1	**0**	25.400000	162.540000	54.180000
		1	**0**	**0**	25.000000	159.800000	52.400000
	Triathlon	**0**	**0**	**0**	27.834677	166.887097	54.699194
				1	28.800000	167.000000	55.200000
			1	**0**	28.600000	171.800000	56.400000
		1	**0**	**0**	30.800000	167.600000	53.800000

					Age	Height	Weight
Sex	**Sport**	**Gold**	**Silver**	**Bronze**			
	Volleyball	**0**	**0**	**0**	24.300281	179.474275	68.768288
				1	25.074074	178.657407	69.329012
			1	**0**	24.705128	180.467949	71.532051
		1	**0**	**0**	24.391026	179.474359	70.988462
	Water Polo	**0**	**0**	**0**	25.158249	175.377104	69.303030
				1	24.828125	175.093750	71.281250
			1	**0**	24.921875	175.015625	70.656250
		1	**0**	**0**	25.761905	177.476190	72.714286
	Weightlifting	**0**	**0**	**0**	24.134078	160.583799	67.386872
				1	24.428571	159.885714	68.242857
			1	**0**	22.771429	160.400000	69.071429
		1	**0**	**0**	23.800000	159.942857	69.314286
	Wrestling	**0**	**0**	**0**	25.127119	163.877119	60.569915
				1	26.781250	164.062500	61.000000
			1	**0**	26.277778	164.388889	60.333333
		1	**0**	**0**	24.055556	162.833333	59.944444

					Age	Height	Weight
Sex	**Sport**	**Gold**	**Silver**	**Bronze**			
M	**Aeronautics**	**1**	**0**	**0**	26.000000	178.400000	75.400000
	Alpine Skiing	**0**	**0**	**0**	23.832071	177.773345	78.260119
				1	24.500000	178.236333	80.214000
				2	25.500000	179.000000	85.500000
			1	**0**	25.413043	177.946739	79.071739
				1	27.714286	178.157143	81.376190
			2	**0**	27.250000	182.000000	87.250000
				1	22.000000	176.000000	85.000000
		1	**0**	**0**	25.555556	179.169444	81.098148
				1	24.833333	180.016667	79.000000
			1	**0**	23.333333	177.722222	79.716667
				1	30.333333	185.666667	92.000000
		2	**0**	**0**	26.000000	180.166667	83.000000
				1	22.000000	178.433333	79.833333
		3	**0**	**0**	22.000000	177.666667	76.200000
	Alpinism	**1**	**0**	**0**	38.533333	178.420000	75.453333
	Archery	**0**	**0**	**0**	28.462349	177.993725	76.324096
				1	27.567568	179.181982	78.372973
				2	30.500000	179.200000	87.700000
			1	**0**	29.166667	175.812037	76.866667
				1	25.000000	182.000000	77.666667
			2	**0**	38.000000	178.400000	75.400000
				1	45.666667	178.400000	75.400000
			3	**1**	40.000000	178.400000	75.400000
		1	**0**	**0**	30.116279	178.270543	77.495349
				1	30.000000	182.000000	100.000000
				2	56.000000	170.033333	75.400000
			1	**0**	25.333333	182.333333	71.666667
			2	**0**	63.000000	170.033333	75.400000
			3	**1**	44.000000	178.400000	75.400000
		2	**0**	**0**	22.666667	178.800000	79.466667
				1	26.000000	170.033333	75.400000
			1	**0**	44.000000	178.400000	75.400000
			2	**0**	34.000000	178.400000	75.400000
		4	**2**	**0**	54.000000	178.400000	75.400000
	Art Competitions	**0**	**0**	**0**	45.052784	177.138894	76.702377
				1	42.209302	176.233272	77.774771
			1	**0**	43.733333	175.357284	78.151481
				1	41.000000	170.136364	79.918182
		1	**0**	**0**	41.068182	176.098004	78.385765
			1	**0**	39.500000	178.580000	84.000000

					Age	Height	Weight
Sex	**Sport**	**Gold**	**Silver**	**Bronze**			
	Athletics	**0**	**0**	**0**	25.587272	179.903756	74.032892
				1	25.050420	181.186006	76.024585
				2	24.040000	178.037000	72.066000
				3	24.000000	176.000000	64.000000
			1	**0**	25.117898	181.325691	76.611210
				1	24.046512	179.287984	74.059690
				2	21.333333	177.322222	69.255556
				3	23.000000	172.000000	64.000000
			2	**0**	25.100000	179.695667	71.829333
		1	**0**	**0**	24.917508	181.618827	77.361588
				1	24.735294	177.858824	71.313235
				2	21.000000	160.000000	51.000000
				3	23.000000	178.000000	71.000000
			1	**0**	25.576271	180.255367	73.843785
				1	23.500000	181.344444	81.150000
			2	**0**	27.600000	180.800000	73.800000
		2	**0**	**0**	24.841270	180.889418	73.873545
				1	25.000000	187.500000	82.000000
			1	**0**	25.600000	178.806667	75.413333
			2	**0**	20.000000	188.000000	81.000000
				1	24.000000	183.000000	64.000000
			3	**0**	24.500000	186.850000	81.990000
		3	**0**	**0**	25.444444	183.777778	76.888889
			1	**0**	22.000000	171.666667	61.166667
		4	**0**	**0**	22.666667	183.000000	75.333333
			2	**0**	28.000000	175.000000	66.000000
		5	**0**	**0**	26.000000	174.000000	65.000000
	Badminton	**0**	**0**	**0**	26.433333	179.493056	74.231019
				1	26.000000	180.000000	74.206897
			1	**0**	27.259259	177.407407	71.666667
		1	**0**	**0**	25.692308	178.884615	72.961538
				1	26.000000	183.000000	75.000000
	Baseball	**0**	**0**	**0**	26.233393	182.687253	85.633573
				1	25.160714	182.505357	85.416964
			1	**0**	27.080357	181.800000	85.497321
		1	**0**	**0**	26.517857	182.987500	86.416964
	Basketball	**0**	**0**	**0**	25.256953	194.273059	90.840847
				1	25.315315	196.777477	94.053153
			1	**0**	25.244541	197.444541	94.434934
		1	**0**	**0**	25.072650	197.782051	95.344444
	Basque Pelota	**1**	**0**	**0**	26.000000	178.400000	75.400000

Sex	Sport	Gold	Silver	Bronze	Age	Height	Weight
	Beach Volleyball	0	0	0	29.888889	193.198413	89.333333
				1	29.166667	193.666667	90.166667
			1	0	30.166667	194.550000	90.416667
		1	0	0	30.500000	193.608333	91.708333
	Biathlon	0	0	0	26.624883	178.431977	72.418117
				1	26.866667	178.158333	72.449722
				2	24.000000	180.500000	80.500000
			1	0	27.319149	178.639362	71.862411
				1	29.000000	180.214286	71.921429
			2	0	30.200000	179.600000	75.200000
				1	30.500000	179.500000	68.500000
		1	0	0	27.589744	180.164103	74.602564
				1	23.571429	178.357143	71.064286
			1	0	25.900000	178.770000	71.253333
				1	25.333333	177.666667	70.666667
			2	0	22.500000	179.000000	71.500000
		2	0	0	29.250000	177.250000	71.112500
				1	31.000000	182.350000	77.266667
			1	0	23.333333	182.000000	72.666667
		3	0	0	29.000000	176.000000	72.000000
		4	0	0	28.000000	178.000000	65.000000
	Bobsleigh	0	0	0	28.938572	181.867165	89.716418
				1	28.902174	181.577174	91.102174
				2	32.625000	180.500000	94.350000
			1	0	29.674157	181.766292	89.301124
				1	29.750000	179.250000	81.000000
			2	0	29.444444	180.888889	86.577778
		1	0	0	30.113924	182.654430	91.903797
				1	30.750000	184.833333	90.583333
			1	0	31.333333	185.444444	92.822222
		2	0	0	31.083333	184.750000	96.950000
	Boxing	0	0	0	23.013272	172.702423	65.029303
				1	22.877358	174.480425	67.500708
			1	0	23.054852	173.824473	67.312658
		1	0	0	23.012500	174.480000	67.651667
			1	0	31.000000	172.500000	61.000000
		2	0	0	20.000000	167.450000	52.000000

					Age	Height	Weight
Sex	**Sport**	**Gold**	**Silver**	**Bronze**			
	Canoeing	**0**	**0**	**0**	25.454080	180.625236	78.877290
				1	25.315556	181.701556	81.261630
				2	26.454545	180.177273	78.150000
			1	**0**	26.598131	181.653037	81.101636
				1	25.190476	181.066667	81.161905
			2	**0**	23.909091	184.818182	85.727273
				1	22.000000	175.000000	85.000000
			3	**0**	26.000000	185.000000	88.000000
		1	**0**	**0**	25.625698	182.613966	81.583240
				1	25.909091	183.565909	84.136364
				2	19.000000	188.000000	87.000000
			1	**0**	26.612903	182.424194	81.258065
				1	28.000000	179.000000	85.000000
		2	**0**	**0**	25.461538	181.582692	81.926923
				1	23.000000	188.000000	87.000000
		3	**0**	**0**	26.500000	185.500000	87.500000
	Cricket	**0**	**1**	**0**	35.571429	178.400000	75.400000
		1	**0**	**0**	25.363636	178.400000	75.400000
	Croquet	**0**	**0**	**0**	30.000000	175.520000	74.600000
				1	39.000000	178.400000	75.400000
			1	**0**	20.000000	178.400000	75.400000
		1	**0**	**1**	58.000000	178.400000	75.400000
			1	**0**	22.000000	178.400000	75.400000
		2	**0**	**0**	15.000000	178.400000	75.400000

					Age	Height	Weight
Sex	**Sport**	**Gold**	**Silver**	**Bronze**			
	Cross Country Skiing	**0**	**0**	**0**	26.002846	177.104945	71.364274
				1	27.651685	176.842697	70.444195
				2	28.000000	178.554167	73.875000
			1	**0**	28.297297	177.563063	71.068243
				1	26.000000	182.479487	75.674359
				2	27.200000	177.800000	69.693333
			2	**0**	29.000000	182.706667	76.513333
				1	26.500000	191.000000	92.500000
			3	**0**	31.000000	174.000000	68.000000
		1	**0**	**0**	27.483333	178.093333	72.233889
				1	28.400000	177.383333	71.737778
				2	27.666667	175.333333	68.000000
			1	**0**	27.285714	179.711905	71.854762
				1	35.000000	173.000000	66.000000
			2	**0**	29.250000	178.383333	74.141667
				1	26.000000	177.000000	72.000000
		2	**0**	**0**	28.166667	184.037500	76.741667
				1	32.000000	170.500000	68.000000
			1	**0**	27.500000	179.000000	73.500000
				1	23.000000	186.500000	81.500000
		3	**0**	**0**	24.000000	183.000000	68.000000
			1	**0**	28.000000	180.000000	73.000000
	Curling	**0**	**0**	**0**	32.152866	180.751592	81.089172
				1	32.384615	180.592308	80.246154
			1	**0**	35.600000	179.973333	83.116667
		1	**0**	**0**	33.285714	181.314286	83.000000

					Age	Height	Weight
Sex	**Sport**	**Gold**	**Silver**	**Bronze**			
	Cycling	**0**	**0**	**0**	24.567714	177.450157	71.660430
				1	24.181818	178.884017	74.441131
				2	24.800000	178.423333	74.290000
			1	**0**	24.405109	178.710375	74.299325
				1	24.904762	178.582540	75.551190
				3	21.000000	176.014286	75.400000
			2	**0**	26.300000	176.093333	75.658333
				1	22.000000	176.900000	71.366667
		1	**0**	**0**	24.011811	178.958333	74.571063
				1	23.000000	177.928571	74.988810
				2	21.000000	178.270000	71.730000
			1	**0**	24.269231	179.942308	75.089744
				1	21.000000	183.900000	77.687500
		2	**0**	**0**	23.833333	179.197222	77.485926
				1	24.000000	171.000000	75.000000
			1	**0**	23.666667	179.200000	76.415556
			3	**1**	19.000000	176.014286	75.400000
		3	**0**	**0**	23.800000	179.100000	78.100000
		4	**0**	**1**	20.000000	181.000000	75.400000
	Diving	**0**	**0**	**0**	23.262712	171.416269	67.378443
				1	23.125000	170.987500	67.608929
				2	24.666667	173.200000	69.333333
			1	**0**	22.964286	171.785714	67.632738
				1	22.600000	170.200000	65.800000
			2	**0**	17.000000	170.000000	58.000000
		1	**0**	**0**	23.044444	169.474444	65.235556
				1	24.000000	169.416667	62.333333
			1	**0**	22.363636	173.227273	67.454545
				2	26.000000	172.000000	65.000000
		2	**0**	**0**	23.714286	170.907143	66.542857

					Age	Height	Weight
Sex	**Sport**	**Gold**	**Silver**	**Bronze**			
	Equestrianism	**0**	**0**	**0**	33.980507	175.690566	69.916139
				1	35.587065	176.317579	69.791725
				2	34.750000	174.956250	70.768750
			1	**0**	36.012739	176.464597	70.941794
				1	37.117647	175.778431	69.513725
				2	33.000000	174.233333	68.433333
			2	**0**	31.000000	175.421429	71.185714
		1	**0**	**0**	35.903030	176.804848	70.625051
				1	35.785714	177.557143	71.646429
			1	**0**	37.194444	176.070833	69.409259
				1	41.000000	173.000000	66.920000
			2	**0**	38.000000	173.025000	65.500000
		2	**0**	**0**	34.125000	176.911458	69.631250
				1	35.000000	173.850000	68.150000
	Fencing	**0**	**0**	**0**	28.774348	180.062902	75.318243
				1	29.138889	180.895365	75.949187
				2	29.400000	181.692333	77.298333
			1	**0**	29.624242	179.622247	76.169289
				1	30.111111	180.516358	76.573148
			2	**0**	30.437500	177.453125	75.757292
				1	41.000000	180.244444	75.216667
			3	**1**	27.000000	200.000000	74.550000
		1	**0**	**0**	29.000000	180.275089	76.279264
				1	28.720000	182.993000	78.526000
				2	40.000000	178.866667	74.866667
			1	**0**	27.650000	180.366875	76.339167
				1	34.333333	180.288889	76.955556
			2	**0**	25.000000	167.000000	74.483333
		2	**0**	**0**	32.459459	178.703378	75.343694
				1	37.000000	180.275000	75.150000
			1	**0**	35.000000	186.741667	75.005556
				2	37.000000	183.000000	75.060000
			2	**0**	32.000000	179.067500	74.725000
		3	**0**	**0**	20.000000	179.700000	74.666667
			1	**0**	20.000000	183.000000	69.000000
			2	**0**	36.000000	180.080000	75.300000
		5	**0**	**0**	25.000000	188.000000	74.960000

					Age	Height	Weight
Sex	**Sport**	**Gold**	**Silver**	**Bronze**			
	Figure Skating	**0**	**0**	**0**	23.217090	175.739607	69.140531
				1	25.532258	175.056452	69.570968
			1	**0**	25.034483	176.727586	70.717241
			2	**0**	24.500000	168.000000	71.000000
		1	**0**	**0**	25.916667	177.068333	71.105000
				1	24.000000	177.500000	72.575000
			1	**0**	26.500000	178.300000	71.700000
		2	**0**	**0**	30.000000	187.000000	77.000000
	Football	**0**	**0**	**0**	23.374606	177.500170	73.023698
				1	23.669856	177.626316	73.563158
			1	**0**	23.437811	178.631095	74.533209
				1	26.000000	174.166667	60.000000
		1	**0**	**0**	24.168704	177.301956	73.233741
	Freestyle Skiing	**0**	**0**	**0**	24.511111	176.453778	74.618444
				1	24.470588	175.294118	73.470588
			1	**0**	24.823529	176.117647	76.647059
		1	**0**	**0**	24.705882	177.941176	78.117647
	Golf	**0**	**0**	**0**	32.156863	179.010294	79.090196
				1	31.000000	178.308333	78.183333
			1	**0**	31.100000	177.571667	78.305833
				1	27.500000	176.750000	77.300000
		1	**0**	**0**	23.916667	178.500000	77.670833
			1	**0**	19.000000	176.750000	77.300000

Sex	Sport	Gold	Silver	Bronze	Age	Height	Weight
	Gymnastics	0	0	0	24.549598	168.165227	64.031332
				1	24.740876	169.964736	65.915319
				2	23.153846	166.265345	61.552097
				4	24.000000	169.000000	68.000000
			1	0	24.434483	171.528592	68.071926
				1	25.000000	166.810685	62.760728
				2	23.600000	163.258214	61.892500
				3	30.500000	166.000000	60.000000
				4	26.000000	169.037500	67.900000
			2	0	24.625000	167.636781	63.823884
				1	25.800000	166.884563	64.886786
				2	33.000000	167.662500	63.362500
			3	0	27.333333	167.583796	62.354167
			4	0	24.000000	169.000000	64.000000
				1	28.000000	167.662500	63.362500
		1	0	0	24.498316	171.145074	67.653752
				1	24.391304	165.571860	60.878768
				2	29.000000	167.771329	63.949107
			1	0	28.076923	165.793269	61.364038
				1	23.285714	164.617857	60.675000
				2	23.200000	170.000000	63.800000
			2	0	26.300000	164.114782	59.483393
				1	32.000000	171.000000	71.000000
			3	0	26.000000	158.000000	52.000000
				1	25.000000	160.000000	58.000000
			4	0	21.000000	166.000000	62.000000
		2	0	0	24.888889	165.331805	62.173810
				1	22.750000	164.917857	60.092857
			1	0	26.000000	165.000000	60.280000
				1	31.000000	165.335714	61.185714
				2	27.500000	164.500000	62.000000
				3	22.000000	174.000000	75.000000
			2	0	24.666667	166.172685	62.020833
				1	27.000000	166.000000	60.000000
			4	1	23.000000	170.000000	65.000000

					Age	Height	Weight
Sex	**Sport**	**Gold**	**Silver**	**Bronze**			
		3	**0**	**0**	24.500000	162.500000	59.150000
				1	23.666667	169.000000	64.333333
				2	24.000000	167.662500	63.362500
			1	**0**	26.000000	165.383482	62.219196
				1	32.000000	171.831250	63.362500
				2	28.000000	163.831250	60.681250
			2	**0**	25.000000	163.000000	59.000000
				1	28.000000	166.400000	63.344286
			4	**1**	22.000000	178.000000	70.000000
		4	**0**	**0**	26.000000	159.000000	70.000000
			1	**0**	23.000000	167.000000	63.000000
				1	25.000000	163.000000	59.000000
			2	**0**	30.000000	167.662500	63.362500
				1	25.500000	168.500000	65.500000
		5	**1**	**0**	25.000000	168.922222	67.400000
		6	**0**	**0**	20.000000	169.000000	68.000000
	Handball	**0**	**0**	**0**	26.421592	188.437595	88.760796
				1	26.774359	190.413333	91.689231
			1	**0**	26.939698	189.506030	89.926131
		1	**0**	**0**	27.252577	190.435052	92.039175
	Hockey	**0**	**0**	**0**	25.877514	176.594747	73.007154
				1	25.488166	177.622189	74.317825
			1	**0**	25.817365	177.849401	74.293114
		1	**0**	**0**	26.193084	177.404611	74.144380
	Ice Hockey	**0**	**0**	**0**	26.113181	180.848496	83.551855
				1	26.411192	181.331265	84.192336
			1	**0**	26.004831	181.161836	84.277295
		1	**0**	**0**	25.972973	181.599509	84.720885
	Jeu De Paume	**0**	**0**	**0**	35.000000	178.916667	75.400000
				1	29.000000	178.500000	75.400000
			1	**0**	39.000000	178.500000	75.400000
		1	**0**	**0**	25.000000	177.250000	75.400000
				1	29.000000	176.833333	75.400000
	Judo	**0**	**0**	**0**	25.399121	177.060967	82.552989
				1	25.491429	178.230857	86.464000
			1	**0**	25.290698	178.383721	86.047674
		1	**0**	**0**	24.595238	177.072619	85.716667
				1	30.000000	180.000000	93.000000
			1	**0**	27.000000	185.000000	110.000000
		2	**0**	**0**	31.000000	187.750000	94.000000
	Lacrosse	**0**	**1**	**0**	26.230769	174.000000	76.700000
		1	**0**	**0**	27.217391	174.000000	76.704348

					Age	Height	Weight
Sex	**Sport**	**Gold**	**Silver**	**Bronze**			
	Luge	**0**	**0**	**0**	24.766625	179.344542	80.652447
				1	27.837838	180.105405	82.251351
				2	26.000000	177.000000	84.666667
			1	**0**	26.282051	179.641026	82.435897
			2	**0**	42.000000	185.000000	95.000000
		1	**0**	**0**	24.921053	179.650000	81.957895
				1	24.000000	185.500000	86.500000
			1	**0**	25.000000	175.500000	78.500000
		2	**0**	**0**	25.333333	184.000000	85.000000
	Military Ski Patrol	**0**	**0**	**0**	22.857143	178.064286	74.300000
				1	23.500000	177.750000	74.266667
			1	**0**	24.333333	178.400000	75.400000
		1	**0**	**0**	25.000000	178.237500	74.875000
	Modern Pentathlon	**0**	**0**	**0**	26.445545	179.118564	72.236035
				1	26.954545	179.144697	71.809848
				2	29.250000	178.400000	74.535417
			1	**0**	28.454545	180.645455	73.703409
				1	24.800000	180.100000	72.875000
			2	**0**	29.000000	173.500000	67.500000
		1	**0**	**0**	27.028571	181.395238	73.973333
				1	29.166667	179.500000	74.166667
			1	**0**	29.857143	179.428571	72.285714
		2	**0**	**0**	24.000000	178.800000	70.800000
			2	**0**	32.000000	175.000000	77.110000
	Motorboating	**0**	**0**	**0**	36.000000	180.257143	76.542857
		1	**0**	**0**	46.000000	181.000000	77.000000
		2	**0**	**0**	29.333333	178.400000	75.400000
	Nordic Combined	**0**	**0**	**0**	23.860972	175.855276	67.726382
				1	24.384615	177.579487	67.941026
				2	27.000000	186.000000	67.000000
				3	26.000000	179.000000	68.000000
			1	**0**	24.206897	178.186207	67.979310
				1	23.400000	179.530000	66.960000
			2	**0**	23.000000	179.133333	67.883333
			3	**0**	29.000000	183.000000	67.000000
		1	**0**	**0**	26.296296	177.174074	67.040741
				1	26.500000	186.000000	68.300000
			1	**0**	24.857143	177.428571	64.714286
				1	27.000000	179.000000	64.000000
		2	**0**	**0**	24.000000	179.500000	63.925000
			1	**0**	30.000000	179.000000	68.000000
		3	**0**	**0**	22.000000	177.000000	65.000000

Sex	Sport	Gold	Silver	Bronze	Age	Height	Weight
	Polo	**0**	**0**	**0**	37.416667	174.600000	66.000000
				1	36.263158	174.431579	66.000000
			1	**0**	31.916667	174.741667	66.000000
		1	**0**	**0**	36.045455	174.590909	66.000000
	Racquets	**0**	**0**	**1**	19.000000	176.000000	75.400000
			1	**0**	24.000000	176.000000	75.400000
				1	45.000000	176.000000	75.400000
		1	**0**	**1**	21.000000	176.000000	75.400000
	Roque	**0**	**0**	**1**	37.000000	178.400000	75.400000
			1	**0**	59.000000	178.400000	75.400000
		1	**0**	**0**	64.000000	178.400000	75.400000
	Rowing	**0**	**0**	**0**	25.142909	186.604536	83.394763
				1	25.777943	187.230179	84.076080
				2	23.888889	184.194444	74.616667
			1	**0**	25.672107	186.999011	84.155143
				1	23.400000	181.253333	79.112500
		1	**0**	**0**	25.267751	187.828107	85.202959
				1	26.875000	186.675000	85.606250
			1	**0**	22.285714	187.235714	79.592857
				1	26.000000	182.216667	69.350000
		2	**0**	**0**	24.333333	186.183333	84.166667
		3	**0**	**0**	21.000000	180.033333	79.433333
	Rugby	**0**	**1**	**0**	24.560000	174.706000	77.800000
		1	**0**	**0**	24.266667	177.226667	77.275000
		2	**1**	**0**	24.000000	178.000000	77.000000
	Rugby Sevens	**0**	**0**	**0**	26.026549	182.654867	90.548673
				1	25.615385	180.384615	88.692308
			1	**0**	26.250000	183.333333	91.916667
		1	**0**	**0**	25.692308	186.384615	96.461538
	Sailing	**0**	**0**	**0**	30.447393	179.837882	78.589126
				1	32.474048	180.273356	79.809343
			1	**0**	31.948171	179.508079	78.146189
				1	38.571429	178.400000	75.400000
				2	30.000000	178.400000	75.400000
			2	**0**	39.400000	178.400000	75.400000
		1	**0**	**0**	32.116711	180.014854	77.784748
				1	50.000000	178.400000	75.400000
			1	**0**	42.000000	178.400000	75.400000
		2	**0**	**0**	30.600000	178.400000	75.400000

					Age	Height	Weight
Sex	**Sport**	**Gold**	**Silver**	**Bronze**			
	Shooting	**0**	**0**	**0**	34.025387	175.905709	78.133957
				1	33.479263	176.080968	79.025568
				2	35.000000	174.936190	78.790357
				3	36.000000	174.430952	78.430952
			1	**0**	33.991489	175.828562	79.362730
				1	38.421053	176.958440	82.469900
				2	38.333333	175.796429	76.941806
			2	**0**	38.000000	174.705556	79.090556
				1	38.000000	176.720556	77.749722
				2	37.000000	172.620000	75.091250
			3	**0**	32.000000	173.382500	75.210000
		1	**0**	**0**	32.600000	176.292783	78.850100
				1	34.772727	175.477314	80.104678
				2	38.500000	175.951875	76.670714
			1	**0**	35.153846	174.321965	77.539607
				1	35.500000	175.940000	76.501250
			2	**0**	32.000000	178.400000	80.516667
				1	30.000000	175.000000	78.500000
		2	**0**	**0**	34.785714	175.193265	81.700383
				1	45.250000	177.460417	77.993750
			1	**0**	33.333333	175.753333	79.025556
				1	30.000000	173.140000	75.400000
				2	39.000000	174.347500	75.545000
		3	**0**	**0**	33.500000	177.920833	77.694444
				1	32.000000	176.500000	76.431429
			1	**0**	41.000000	172.400000	75.762500
			2	**0**	33.500000	174.618750	76.481944
		4	**1**	**1**	35.000000	180.000000	75.162500
				2	35.000000	175.823077	75.253846
		5	**1**	**1**	31.000000	179.000000	75.273333

Sex	Sport	Gold	Silver	Bronze	Age	Height	Weight
	Short Track Speed Skating	**0**	**0**	**0**	23.000000	176.121271	71.839494
				1	22.310345	174.689655	70.586207
				2	19.000000	173.000000	63.000000
			1	**0**	24.115385	173.983333	70.865385
				1	20.600000	175.200000	72.000000
				2	27.000000	173.000000	68.000000
			2	**0**	22.250000	173.500000	67.250000
		1	**0**	**0**	23.590909	176.240909	71.875000
				1	25.250000	170.500000	70.750000
				2	23.000000	173.000000	68.000000
			1	**0**	21.500000	173.641667	71.958333
			2	**0**	19.000000	168.000000	60.000000
		2	**0**	**0**	24.500000	173.500000	67.000000
				1	26.000000	184.000000	75.000000
			1	**0**	20.000000	173.000000	60.000000
		3	**0**	**1**	24.000000	170.000000	65.000000
	Skeleton	**0**	**0**	**0**	31.115044	180.892920	82.162389
				1	31.500000	183.366667	80.333333
			1	**0**	29.166667	182.033333	79.500000
		1	**0**	**0**	31.000000	181.540000	85.720000
			1	**0**	23.000000	179.500000	82.150000
	Ski Jumping	**0**	**0**	**0**	23.521087	175.915737	66.221198
				1	24.159091	178.242045	65.538636
				2	23.250000	176.312500	59.937500
			1	**0**	23.822222	178.222222	64.864444
				1	24.428571	177.542857	64.271429
			2	**0**	28.000000	173.500000	59.500000
			3	**0**	17.000000	182.000000	65.000000
		1	**0**	**0**	23.421053	177.075000	64.705263
				1	24.666667	175.750000	62.250000
				2	23.500000	180.500000	62.000000
			1	**0**	24.300000	174.025000	62.125000
		2	**0**	**0**	24.400000	173.800000	56.800000
				1	16.000000	179.000000	60.000000
			1	**0**	22.000000	175.000000	62.000000
		3	**0**	**0**	26.500000	176.262500	61.287500
	Snowboarding	**0**	**0**	**0**	24.764302	178.353089	76.477346
				1	24.071429	181.500000	79.928571
			1	**0**	23.071429	175.214286	73.635714
				1	29.000000	184.000000	77.000000
		1	**0**	**0**	25.000000	180.153846	80.769231
		2	**0**	**0**	27.000000	179.000000	82.000000

					Age	Height	Weight
Sex	**Sport**	**Gold**	**Silver**	**Bronze**			
	Speed Skating	**0**	**0**	**0**	24.564664	178.840667	75.890589
				1	24.800000	179.832056	77.632556
				2	25.000000	181.706250	77.656250
				3	27.000000	180.000000	75.000000
			1	**0**	24.610169	179.240113	77.409605
				1	25.250000	182.886111	79.141667
				2	27.000000	190.000000	92.000000
			2	**0**	25.200000	181.986667	81.280000
				1	24.000000	178.800000	75.600000
				3	26.000000	177.340000	74.020000
		1	**0**	**0**	25.513514	181.277928	78.629279
				1	26.666667	181.888889	79.283333
			1	**0**	24.476190	182.076190	79.200000
				1	28.000000	177.446667	75.206667
		2	**0**	**0**	26.285714	176.800000	75.600000
				1	24.500000	180.000000	76.000000
			1	**0**	26.000000	179.500000	75.000000
		3	**0**	**0**	26.666667	185.977778	80.800000
			1	**0**	31.000000	180.000000	75.600000
				1	30.000000	167.000000	67.000000
		5	**0**	**0**	21.000000	185.000000	84.000000

Sex	Sport	Gold	Silver	Bronze	Age	Height	Weight
	Swimming	0	0	0	21.452670	183.570914	77.288278
				1	21.913621	185.121788	78.816944
				2	22.136364	185.537500	80.267045
			1	0	21.896552	185.185166	79.336188
				1	21.735849	185.091038	79.908176
				2	23.000000	182.123810	80.166667
			2	0	21.625000	185.145833	80.072222
				1	21.100000	184.504167	76.225000
			3	0	21.500000	184.000000	76.500000
				1	22.000000	180.375000	76.175000
		1	0	0	21.578125	186.670009	80.897396
				1	22.000000	185.960078	80.875969
				2	20.000000	187.193750	82.862500
			1	0	21.574468	187.914539	81.423404
				1	21.400000	185.700000	81.380000
			2	0	22.090909	191.478788	85.072727
		2	0	0	21.215686	185.606863	78.948319
				1	22.875000	184.004167	79.352083
				2	24.333333	193.000000	89.333333
			1	0	22.066667	190.417778	85.986667
				1	20.714286	190.142857	84.000000
			2	0	21.600000	198.000000	88.400000
				1	27.000000	188.000000	88.000000
		3	0	0	21.363636	183.909091	78.909091
				1	21.000000	191.000000	88.000000
			1	0	21.000000	184.666667	78.666667
				1	19.000000	183.000000	70.000000
			2	0	17.000000	196.000000	104.000000
		4	0	0	18.000000	180.000000	79.000000
			1	0	20.000000	198.000000	88.000000
			2	0	27.000000	193.000000	91.000000
		5	1	0	31.000000	193.000000	91.000000
				1	22.000000	200.000000	95.000000
		6	0	2	19.000000	193.000000	91.000000
		7	0	0	22.000000	183.000000	73.000000
		8	0	0	23.000000	193.000000	91.000000

Sex	Sport	Gold	Silver	Bronze	Age	Height	Weight
	Table Tennis	**0**	**0**	**0**	27.490066	177.252318	71.967881
				1	27.000000	177.354545	69.009091
				2	22.500000	180.000000	73.500000
			1	**0**	25.263158	176.736842	69.157895
				1	25.500000	170.000000	61.500000
			2	**0**	28.000000	164.000000	70.000000
		1	**0**	**0**	23.384615	174.384615	68.000000
				1	25.000000	179.000000	71.500000
			1	**0**	26.000000	175.750000	73.000000
		2	**0**	**0**	24.750000	174.250000	69.000000
	Taekwondo	**0**	**0**	**0**	24.889362	181.511064	74.221702
				1	23.937500	186.156250	76.250000
			1	**0**	24.300000	184.300000	76.000000
		1	**0**	**0**	23.900000	185.050000	75.500000
	Tennis	**0**	**0**	**0**	26.330986	184.592158	78.672316
				1	26.745098	185.462418	80.058824
				2	25.333333	188.522222	78.355556
			1	**0**	27.702703	184.164414	79.334685
				1	27.000000	185.158333	79.408333
			2	**0**	26.400000	179.816667	74.896667
				1	31.500000	186.910000	79.613333
		1	**0**	**0**	28.310345	185.659770	80.399138
				1	27.500000	187.386111	82.136111
			1	**0**	28.000000	186.266667	80.141667
				1	38.000000	182.058333	78.745833
		2	**0**	**0**	26.500000	183.358333	79.275000
				1	24.333333	181.068889	76.744444
			1	**0**	21.000000	178.000000	80.066667
		3	**0**	**0**	23.000000	185.566667	80.066667
	Trampolining	**0**	**0**	**0**	24.983607	171.262295	65.895082
				1	22.000000	172.600000	66.000000
			1	**0**	25.600000	171.600000	67.000000
		1	**0**	**0**	23.600000	171.200000	63.800000
	Triathlon	**0**	**0**	**0**	27.868526	180.143426	68.744223
				1	25.200000	178.800000	66.800000
			1	**0**	28.600000	180.000000	70.600000
		1	**0**	**0**	27.200000	184.400000	72.000000

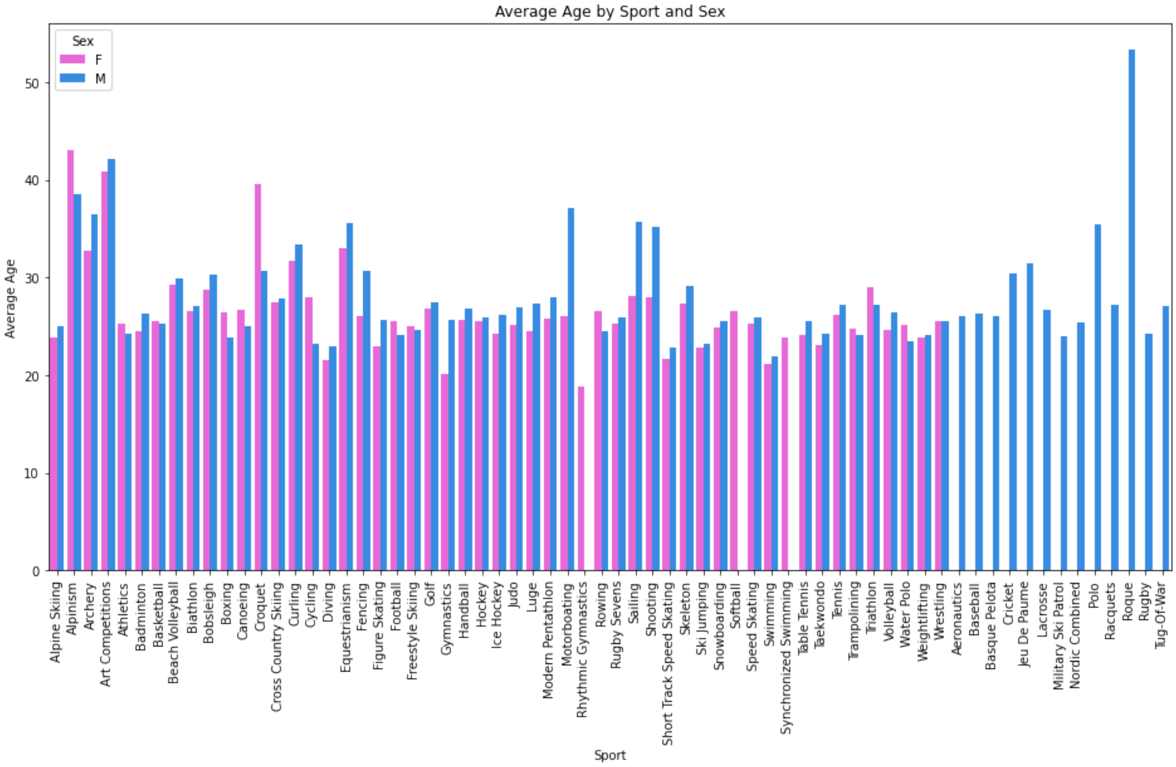
Sex	Sport	Gold	Silver	Bronze	Age	Height	Weight
	Tug-Of-War	0	0	0	28.400000	182.585000	96.613000
				1	28.173913	182.326087	94.315217
			1	0	30.766667	182.429167	96.356667
				1	27.000000	183.000000	93.000000
				3	25.000000	170.000000	75.000000
		1	0	0	30.421053	183.124123	94.984474
				1	29.000000	181.962500	97.075000
				2	21.000000	170.000000	82.000000
				3	26.000000	190.000000	88.000000
			1	0	24.750000	178.725000	93.412500
	Volleyball	0	0	0	25.457627	193.010169	86.458806
				1	26.411043	193.480982	87.403067
			1	0	26.945783	193.885542	87.692771
		1	0	0	26.969880	194.291566	89.330120
	Water Polo	0	0	0	25.655098	186.453883	86.963966
				1	25.664311	187.709435	88.866519
				2	21.333333	183.016667	80.147222
			1	0	26.071161	187.621317	89.205899
				1	18.000000	185.450000	80.725000
			2	0	22.333333	182.950000	78.966667
		1	0	0	26.398577	187.731791	89.463108
				1	20.000000	186.000000	83.250000
			1	0	26.000000	184.400000	82.150000
		2	0	0	25.800000	182.573333	78.870000
		3	0	0	25.500000	186.300000	84.033333
				1	19.000000	185.025000	80.650000
	Weightlifting	0	0	0	25.737537	168.942009	80.061131
				1	26.076471	169.281765	81.963725
			1	0	25.450292	169.150292	82.685965
		1	0	0	25.463277	168.494068	83.153390
			1	0	21.000000	172.850000	89.450000
				1	21.000000	178.828571	89.828571
	Wrestling	0	0	0	25.745354	172.513333	75.367252
				1	26.197150	173.335590	77.943587
				2	24.000000	183.000000	100.250000
			1	0	26.065041	172.950542	76.621274
				1	27.000000	181.825000	83.000000
			2	0	25.000000	179.325000	81.725000
		1	0	0	26.007937	173.091931	76.409127
				1	24.000000	173.850000	77.450000
			1	0	23.000000	178.666667	90.166667
		2	0	0	27.666667	180.316667	93.733333

In [43]:

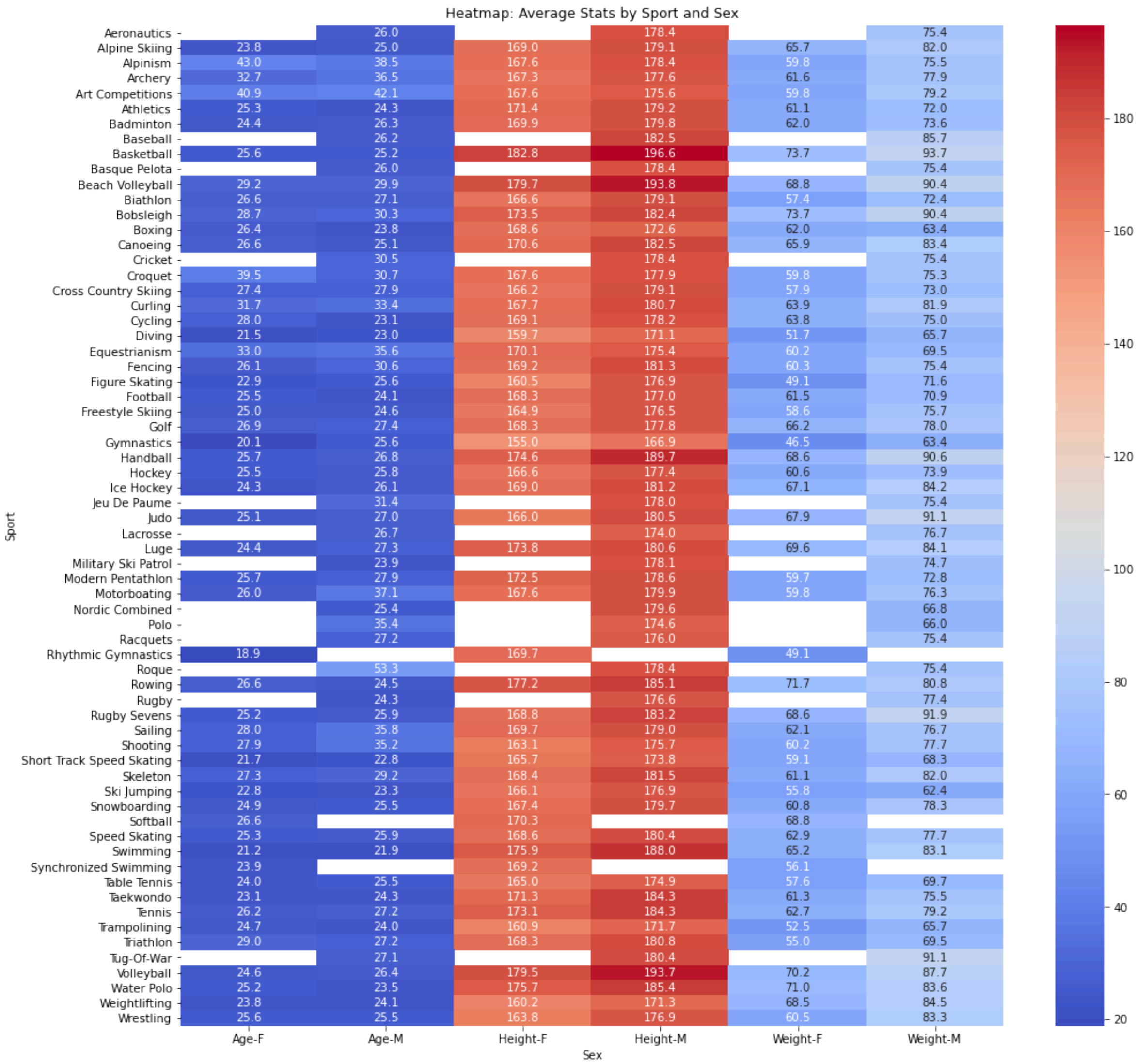
```
# Reset the index of `average_stats` to get the grouping columns for plotting
average_stats_reset = average_stats.reset_index()
color_map = {'F': '#FF53E6', 'M': '#1E90FF'}
# Grouped Bar Chart for Age, Height, Weight across Sex, Sport, and Medal counts
def grouped_bar_chart():
    fig, ax = plt.subplots(figsize=(16, 8))
    sns.barplot(data=average_stats_reset, x='Sport', y='Age', hue='Sex', ci=None, ax=
ax, palette=color_map)
    ax.set_title('Average Age by Sport and Sex')
    ax.set_ylabel('Average Age')
    ax.set_xlabel('Sport')
    plt.xticks(rotation=90)
    plt.show()

# Heatmap of Age, Height, Weight across Sport and Sex
def heatmap_chart():
    stats_for_heatmap = average_stats_reset.pivot_table(index='Sport', columns='Sex',
values=['Age', 'Height', 'Weight'])
    plt.figure(figsize=(16, 16))
    sns.heatmap(stats_for_heatmap, annot=True, cmap='coolwarm', fmt='.1f')
    plt.title('Heatmap: Average Stats by Sport and Sex')
    plt.xlabel('Sex')
    plt.ylabel('Sport')
    plt.show()

def radar_chart():

    # Create a side-by-side layout for all four radar charts
    fig, (ax1, ax2, ax3, ax4) = plt.subplots(1, 4, figsize=(24, 6), subplot_kw=dict(p
olar=True))

# Radar Chart 1: For Basketball
    data_basketball = average_stats_reset[average_stats_reset['Sport'] == 'Basketbal
l'].iloc[0][['Age', 'Height', 'Weight']]
    categories_basketball = ['Age', 'Height', 'Weight']
    data_basketball['Age'] = data_basketball['Age'] / 0.3  # Example scaling factor f
or Age
    data_basketball['Weight'] = data_basketball['Weight'] / 0.5  # Example scaling fa
ctor for Weight
    values_basketball = data_basketball.values
    angles_basketball = np.linspace(0, 2 * np.pi, len(categories_basketball), endpoin
t=False).tolist()

    values_basketball = np.concatenate((values_basketball, [values_basketball[0]]))
# Close the radar chart loop
    angles_basketball += angles_basketball[:1]

    ax1.fill(angles_basketball, values_basketball, color='blue', alpha=0.25)
    ax1.plot(angles_basketball, values_basketball, color='blue', linewidth=2)
    ax1.set_yticklabels([])  # Remove radial ticks
    ax1.set_xticks(angles_basketball[:-1])
    ax1.set_xticklabels(categories_basketball)
    ax1.set_title('Radar Chart: Average Stats for Basketball')

    # Radar Chart 2: For Art Competitions (Correcting Sport name)
    data_art_competitions = average_stats_reset[average_stats_reset['Sport'] == 'Art
Competitions'].iloc[0][['Age', 'Height', 'Weight']]
    categories_art_competitions = ['Age', 'Height', 'Weight']
    data_art_competitions['Age'] = data_art_competitions['Age'] / 0.3  # Example scal
ing factor for Age
    data_art_competitions['Weight'] = data_art_competitions['Weight'] / 0.5  # Exampl
e scaling factor for Weight
    values_art_competitions = data_art_competitions.values
    angles_art_competitions = np.linspace(0, 2 * np.pi, len(categories_art_competitio
ns), endpoint=False).tolist()

    values_art_competitions = np.concatenate((values_art_competitions, [values_art_co
```

```
mpetitions[0]]))  # Close the radar chart loop
    angles_art_competitions += angles_art_competitions[:1]

    ax2.fill(angles_art_competitions, values_art_competitions, color='blue', alpha=0.
25)
    ax2.plot(angles_art_competitions, values_art_competitions, color='blue', linewidt
h=2)
    ax2.set_yticklabels([])  # Remove radial ticks
    ax2.set_xticks(angles_art_competitions[:-1])
    ax2.set_xticklabels(categories_art_competitions)
    ax2.set_title('Radar Chart: Average Stats for Art Competitions')

    # Radar Chart 3: For Athletics (Correcting Sport name)
    data_athletics = average_stats_reset[average_stats_reset['Sport'] == 'Athletic
s'].iloc[0][['Age', 'Height', 'Weight']]
    categories_athletics = ['Age', 'Height', 'Weight']
    data_athletics['Age'] = data_athletics['Age'] / 0.3  # Example scaling factor for
Age
    data_athletics['Weight'] = data_athletics['Weight'] / 0.5  # Example scaling fact
or for Weight
    values_athletics = data_athletics.values
    angles_athletics = np.linspace(0, 2 * np.pi, len(categories_athletics), endpoint=
False).tolist()

    values_athletics = np.concatenate((values_athletics, [values_athletics[0]]))  # C
lose the radar chart loop
    angles_athletics += angles_athletics[:1]

    ax3.fill(angles_athletics, values_athletics, color='blue', alpha=0.25)
    ax3.plot(angles_athletics, values_athletics, color='blue', linewidth=2)
    ax3.set_yticklabels([])  # Remove radial ticks
    ax3.set_xticks(angles_athletics[:-1])
    ax3.set_xticklabels(categories_athletics)
    ax3.set_title('Radar Chart: Average Stats for Athletics')

    # Radar Chart 4: For Boxing
    data_boxing = average_stats_reset[average_stats_reset['Sport'] == 'Weightliftin
g'].iloc[0][['Age', 'Height', 'Weight']]
    categories_boxing = ['Age', 'Height', 'Weight']
    data_boxing['Age'] = data_boxing['Age'] / 0.1  # Example scaling factor for Age
    data_boxing['Weight'] = data_boxing['Weight'] / 0.1  # Example scaling factor for
Weight
    values_boxing = data_boxing.values
    angles_boxing = np.linspace(0, 2 * np.pi, len(categories_boxing), endpoint=Fals
e).tolist()

    values_boxing = np.concatenate((values_boxing, [values_boxing[0]]))  # Close the
radar chart loop
    angles_boxing += angles_boxing[:1]

    ax4.fill(angles_boxing, values_boxing, color='blue', alpha=0.25)
    ax4.plot(angles_boxing, values_boxing, color='blue', linewidth=2)
    ax4.set_yticklabels([])  # Remove radial ticks
    ax4.set_xticks(angles_boxing[:-1])
    ax4.set_xticklabels(categories_boxing)
    ax4.set_title('Radar Chart: Average Stats for Weightlifting')

    plt.tight_layout()
    plt.show()


# Bubble Chart for Age vs Height vs Weight (Size of bubbles)
def bubble_chart():
    plt.figure(figsize=(10, 6))
    sns.scatterplot(data=average_stats_reset, x='Age', y='Height', size='Weight', hue
='Sport', sizes=(10, 200), legend=False, alpha=0.6)
    plt.title('Bubble Chart: Age vs Height vs Weight')
    plt.xlabel('Age')
```

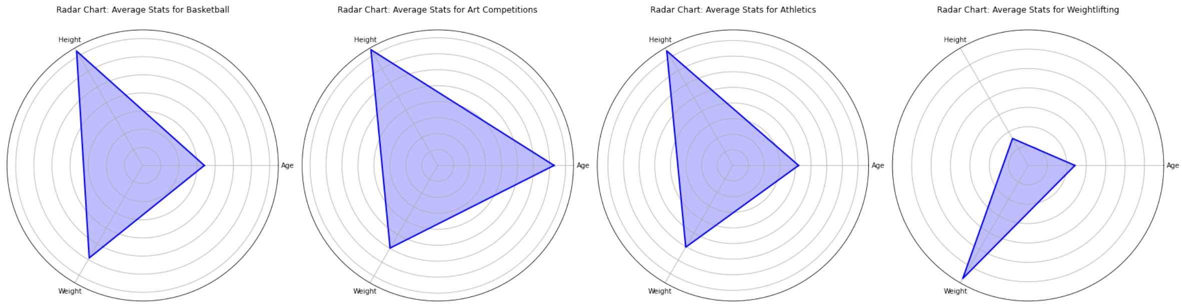
```
    plt.ylabel('Height')
    plt.show()

# Call all chart functions
grouped_bar_chart()
heatmap_chart()
radar_chart()
bubble_chart()
```

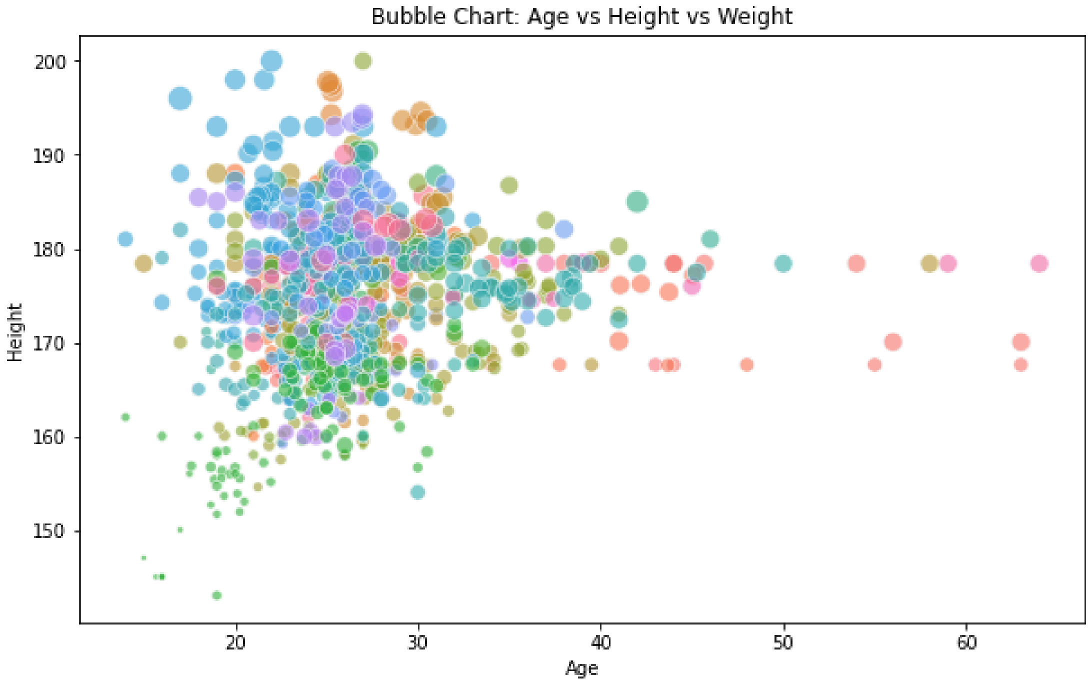


Average Age by Sport and Sex
Sex
F
M
Average Age
Sport



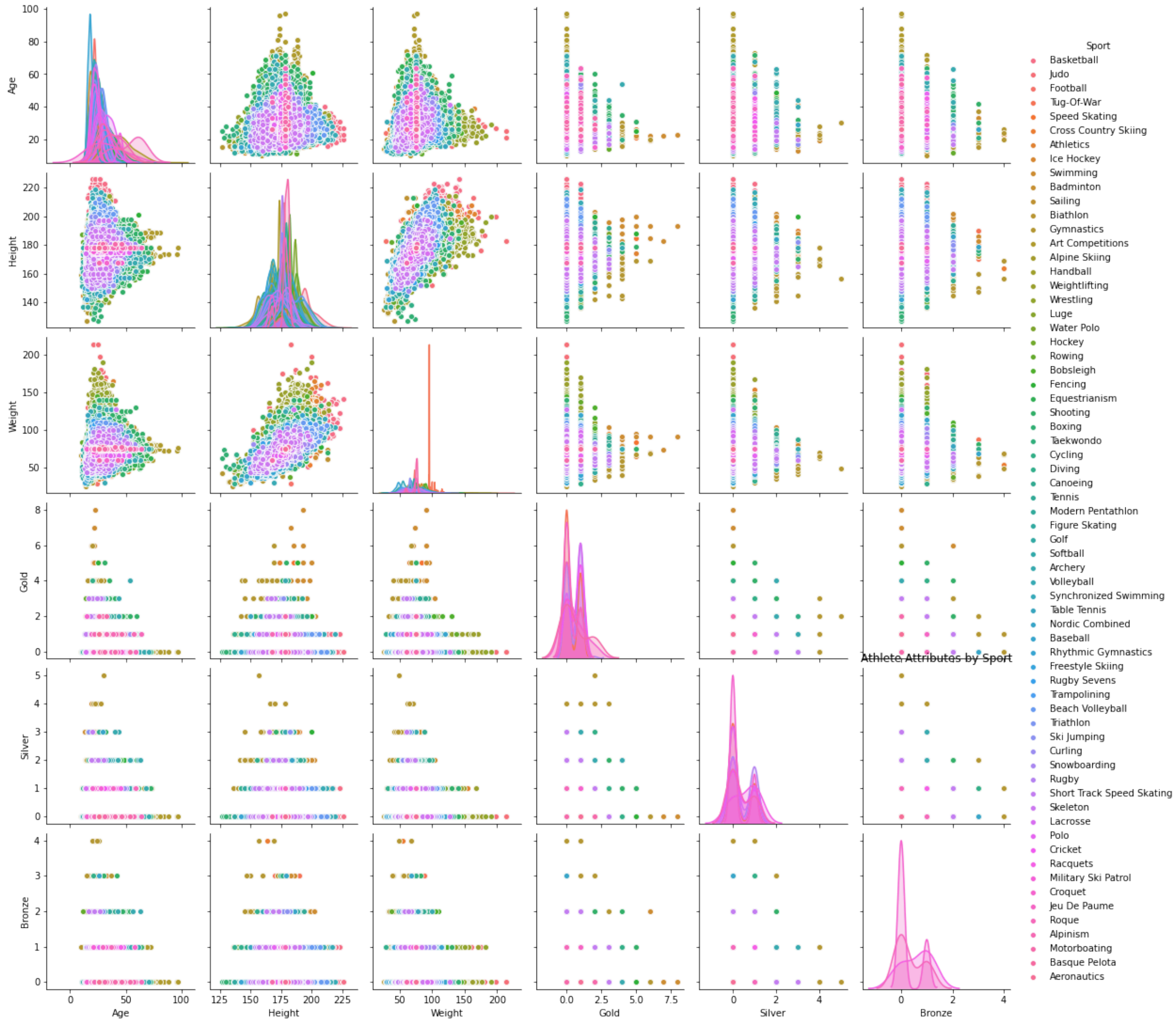
Heatmap: Average Stats by Sport and Sex
Sport
Sex
Age-F
Age-M
Height-F
Height-M
Weight-F
Weight-M


Radar Chart: Average Stats for Basketball
Radar Chart: Average Stats for Art Competitions
Radar Chart: Average Stats for Athletics
Radar Chart: Average Stats for Weightlifting
Height
Age
Weight


Bubble Chart: Age vs Height vs Weight
Height
Age
200
190
180
170
160
150
20
30
40
50
60

In [44]:

```
# --- Data Visualization ---
sns.pairplot(athlete_data_rem_dupli, vars=['Age', 'Height', 'Weight', 'Gold', 'Silve
r', 'Bronze'], hue='Sport')
plt.title("Athlete Attributes by Sport")
plt.show()
```



In [45]:
```
# Print the descriptive statistics for Age, Height, and Weight grouped by Sex and Eve
nt
print("\nDescriptive Statistics:")
athlete_data_describe = athlete_data_rem_dupli.groupby(['Sex', 'Sport'])[['Age', 'Hei
ght', 'Weight']].describe().T
print(athlete_data_describe)
```

```
Descriptive Statistics:
Sex                      F                                                         \
Sport        Alpine Skiing Alpinism     Archery Art Competitions    Athletics
Age    count   1524.000000      1.0  673.000000     1.660000e+02  9473.000000
       mean      22.394357     43.0   26.997028     4.360241e+01    25.179669
       std        3.854033      NaN    7.739432     1.316694e+01     4.754912
       min       14.000000     43.0   14.000000     1.700000e+01    12.000000
       25%       20.000000     43.0   21.000000     3.325000e+01    22.000000
       50%       22.000000     43.0   25.000000     4.400000e+01    25.000000
       75%       25.000000     43.0   31.000000     5.300000e+01    28.000000
       max       45.000000     43.0   63.000000     7.400000e+01    48.000000
Height count   1524.000000      1.0  673.000000     1.660000e+02  9473.000000
       mean     167.162191    167.6  167.066939     1.675542e+02   169.498744
       std        4.676852      NaN    5.422376     5.898744e-01     7.437509
       min      152.000000    167.6  152.000000     1.600000e+02   142.000000
       25%      165.000000    167.6  164.000000     1.676000e+02   165.000000
       50%      167.000000    167.6  167.250000     1.676000e+02   169.000000
       75%      170.000000    167.6  170.000000     1.676000e+02   175.000000
       max      187.000000    167.6  185.000000     1.676000e+02   194.000000
Weight count   1524.000000      1.0  673.000000     1.660000e+02  9473.000000
       mean      62.443362     59.8   61.704681     5.980000e+01    60.665794
       std        5.522667      NaN    7.720630     2.121556e-14    12.197158
       min       45.000000     59.8   42.000000     5.980000e+01    35.000000
       25%       60.000000     59.8   57.000000     5.980000e+01    53.000000
       50%       62.066667     59.8   61.800000     5.980000e+01    58.000000
       75%       65.000000     59.8   64.000000     5.980000e+01    64.000000
       max       90.000000     59.8   95.000000     5.980000e+01   136.000000

Sex                                                                  \
Sport          Badminton   Basketball Beach Volleyball    Biathlon
Age    count  606.000000  1256.000000       276.000000  642.000000
       mean    24.953795    25.517516        28.315217   25.305296
       std      3.838196     4.062619         4.133049    4.096374
       min     16.000000    16.000000        18.000000   17.000000
       25%     22.000000    22.000000        25.000000   22.000000
       50%     25.000000    25.000000        28.000000   25.000000
       75%     27.000000    28.000000        31.000000   28.000000
       max     44.000000    37.000000        39.000000   38.000000
Height count  606.000000  1256.000000       276.000000  642.000000
       mean   168.335314   182.455016       178.867391  166.610202
       std      5.891149     8.860684         5.931490    5.683041
       min    150.000000   161.000000       163.000000  150.000000
       25%    164.000000   176.000000       175.000000  163.000000
       50%    168.200000   183.000000       180.000000  166.500000
       75%    172.000000   190.000000       183.000000  170.750000
       max    184.000000   213.000000       196.000000  182.000000
Weight count  606.000000  1256.000000       276.000000  642.000000
       mean    61.487679    73.685748        68.352899   57.381347
       std      5.892785     9.666320         5.286135    4.920085
       min     43.000000    50.000000        55.000000   45.000000
       25%     58.000000    67.000000        65.000000   54.000000
       50%     61.000000    73.350000        68.000000   57.000000
       75%     65.000000    79.000000        72.000000   60.000000
       max     90.000000   128.000000        81.000000   72.000000

Sex                       ...           M                                \
Sport          Bobsleigh  ... Table Tennis   Taekwondo       Tennis
Age    count  143.000000  ...   673.000000  307.000000  1004.000000
       mean    27.832168  ...    27.279346   24.687296    26.492032
       std      4.087056  ...     5.809665    3.953149     4.969592
       min     19.000000  ...    15.000000   16.000000    15.000000
       25%     25.000000  ...    23.000000   22.000000    23.000000
       50%     28.000000  ...    26.000000   24.000000    26.000000
       75%     30.500000  ...    31.000000   27.000000    29.000000
       max     39.000000  ...    54.000000   37.000000    44.000000
Height count  143.000000  ...   673.000000  307.000000  1004.000000
       mean   173.181818  ...   177.131055  182.407492   184.639968
       std      6.399424  ...     7.102657    8.999171     5.822019
```

```
       min    158.000000  ...   150.000000  160.000000   162.000000
       25%    168.500000  ...   173.000000  176.000000   181.950000
       50%    173.000000  ...   177.000000  183.000000   184.750000
       75%    178.000000  ...   182.000000  189.000000   188.000000
       max    194.000000  ...   198.000000  207.000000   208.000000
Weight count  143.000000  ...   673.000000  307.000000  1004.000000
       mean    72.804196  ...    71.672808   74.632248    78.827945
       std      6.730484  ...     7.950965   13.416297     6.095224
       min     55.000000  ...    50.000000   54.000000    59.000000
       25%     68.500000  ...    66.000000   63.500000    75.300000
       50%     73.000000  ...    71.200000   74.000000    78.750000
       75%     77.500000  ...    76.000000   83.000000    81.000000
       max     95.000000  ...    99.000000  110.000000   111.000000

Sex                                                                     \
Sport         Trampolining   Triathlon  Tug-Of-War   Volleyball   Water Polo
Age    count     76.000000  266.000000  126.000000  1852.000000  3156.000000
       mean      24.736842   27.819549   29.309524    25.810475    25.739861
       std        3.906809    3.866322    6.141614     3.917882     4.449449
       min       18.000000   18.000000   17.000000    17.000000    14.000000
       25%       22.000000   25.000000   24.000000    23.000000    22.000000
       50%       24.000000   28.000000   29.000000    25.000000    25.000000
       75%       27.000000   30.000000   33.000000    28.000000    29.000000
       max       34.000000   42.000000   45.000000    41.000000    45.000000
Height count     76.000000  266.000000  126.000000  1852.000000  3156.000000
       mean     171.368421  180.195489  182.393188   193.244924   186.761339
       std        5.358712    6.046310    2.893182     7.809953     6.533559
       min      162.000000  164.000000  170.000000   159.000000   154.000000
       25%      168.000000  176.000000  182.500000   188.000000   184.000000
       50%      170.000000  180.000000  182.500000   194.000000   186.800000
       75%      175.000000  185.000000  182.500000   198.000000   190.000000
       max      185.000000  196.000000  195.000000   219.000000   206.000000
Weight count     76.000000  266.000000  126.000000  1852.000000  3156.000000
       mean      65.836842   68.803759   95.162579    86.909881    87.504832
       std        6.198986    5.323425    7.346365     8.189663     8.811717
       min       57.000000   54.000000   75.000000    56.000000    61.000000
       25%       62.000000   65.000000   95.000000    82.000000    82.000000
       50%       64.000000   69.000000   95.600000    86.900000    87.700000
       75%       69.000000   72.000000   95.600000    92.000000    91.000000
       max       84.000000   82.000000  118.000000   120.000000   125.000000

Sex
Sport         Weightlifting    Wrestling
Age    count    3248.000000  6077.000000
       mean       25.722291    25.811585
       std         4.332696     4.113980
       min        15.000000    15.000000
       25%        23.000000    23.000000
       50%        25.000000    25.000000
       75%        28.000000    28.000000
       max        45.000000    50.000000
Height count    3248.000000  6077.000000
       mean      168.950594   172.646976
       std         9.231694     9.496303
       min       140.000000   137.000000
       25%       162.000000   166.000000
       50%       169.100000   172.000000
       75%       175.000000   180.000000
       max       205.000000   214.000000
Weight count    3248.000000  6077.000000
       mean       80.473315    75.712100
       std        22.594234    19.086083
       min        50.000000    47.000000
       25%        62.000000    61.900000
       50%        75.400000    72.600000
       75%        90.000000    88.000000
       max       176.500000   190.000000
```

```
[24 rows x 116 columns]
```

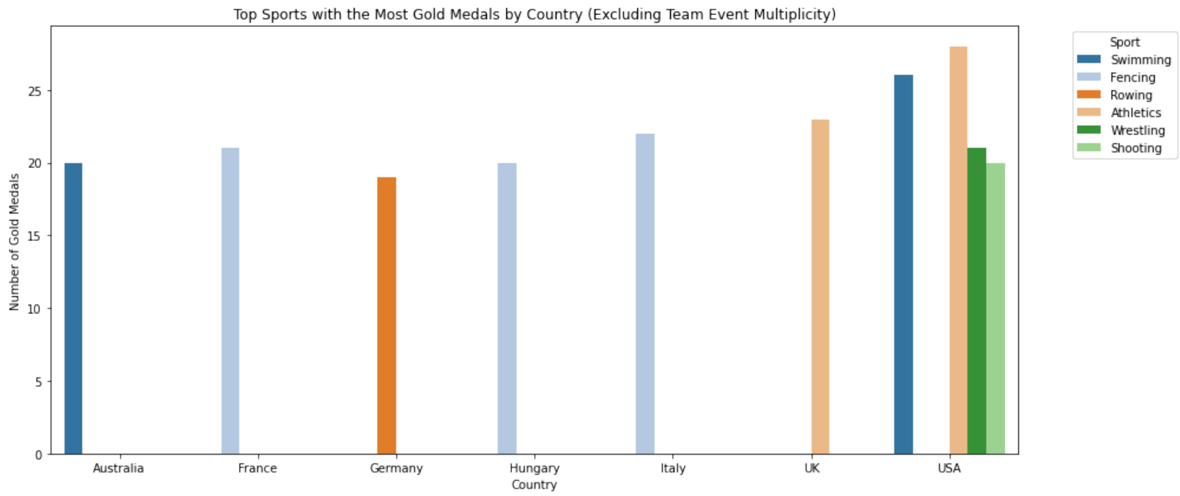
Conclusion

The data visualization, using a pairplot to explore the relationships between key athlete attributes (Age, Height, Weight, and Medal counts in Gold, Silver, and Bronze) across different sports, provides valuable insights into the diversity of athletic profiles based on sport and gender. The hue differentiation by sport allows for a clear comparison of how various sports require different physical characteristics, and the added transparency in the visualizations enhances clarity, particularly when multiple data points overlap.

From the results, we observe notable variations in physical attributes between genders and sports. For instance, sports like basketball, volleyball, and rowing, which generally involve higher body mass and height, show a significant distinction between male and female athletes. Conversely, sports such as gymnastics, figure skating, and diving feature relatively smaller body frames. Additionally, sports with high endurance demands, such as cycling and swimming, tend to have more balanced distributions in terms of age and physical size.

These insights could help identify trends in physical attributes across genders and sports, offering guidance for sports training, athlete selection, and performance analysis. This visualization approach, combined with detailed data, serves as a powerful tool for further in-depth analysis of athlete characteristics in relation to their respective disciplines.

In [46]:
```
# Filter for Gold medals only
gold_medals = athlete_data[athlete_data['Medal'] == 'Gold']

# Remove duplicates where multiple people from the same region won a gold medal in th
e same event
gold_medals_unique = gold_medals.drop_duplicates(subset=['region', 'Sport', 'Year'])

# Group by region and event, and count the number of Gold medals (count only 1 medal
per region-event pair)
medals_by_country_sport = gold_medals_unique.groupby(['region', 'Sport'])['Medal'].co
unt().nlargest(10).reset_index()

# Sort by the number of Gold medals for each country, descending order, and take the
top 5 events for each country
top_events_per_country = (
    medals_by_country_sport.groupby('region')
    .apply(lambda x: x.nlargest(10, 'Medal'))
    .reset_index(drop=True)
)

# Generate a distinct color palette for the bars
colors = plt.cm.tab20.colors[:len(top_events_per_country['Sport'].unique())]

# Create a bar plot to display the results
plt.figure(figsize=(14, 6))  # Adjusted height for better visualization

# Plot the data, assigning a distinct color to each event
sns.barplot(data=top_events_per_country, x='region', y='Medal', hue='Sport', palette=
colors)

# Title and labels
plt.title("Top Sports with the Most Gold Medals by Country (Excluding Team Event Mult
iplicity)")
plt.xlabel("Country")
plt.ylabel("Number of Gold Medals")
plt.legend(title='Sport', bbox_to_anchor=(1.05, 1))

# Show the plot
plt.tight_layout()
plt.show()
```



In [47]:

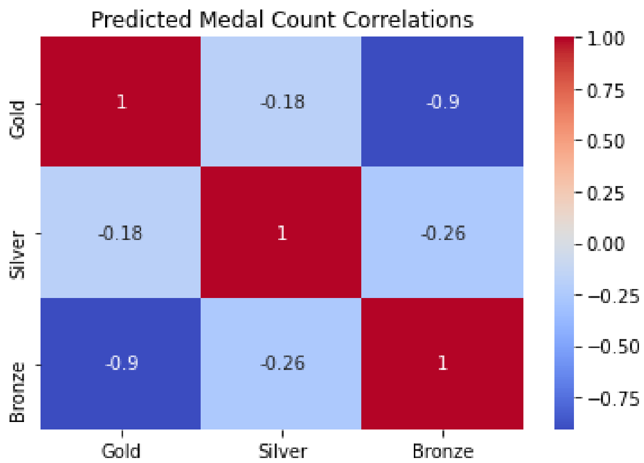
```python
from mpl_toolkits.mplot3d import Axes3D

# Group by 'Sex' and 'Sport', then calculate the average of 'Age', 'Height', and 'Wei
ght'
average_stats = athlete_data_rem_dupli.groupby(['Sex', 'Sport'])[['Age', 'Height', 'W
eight']].mean()

# Reset index to access 'Sex' and 'Sport' for plotting
average_stats = average_stats.reset_index()

# Extract the relevant columns for plotting
ages = average_stats['Age']
heights = average_stats['Height']
weights = average_stats['Weight']
sex = average_stats['Sex']
sports = average_stats['Sport']

# Create 3D plot
fig = plt.figure(figsize=(16, 10))
ax = fig.add_subplot(111, projection='3d')

# Plot the data, color by 'Sex' (could be customized to use 'Sport' too)
color_map = {'F': '#FF53E6', 'M': '#1E90FF'}
colors = [color_map[sex_] for sex_ in sex]

# Plot the points (with a linear style of markers)
scatter = ax.scatter(ages, heights, weights, c=colors, s=50)

# Label every 3rd point and ensure labels for 'Basketball', 'Athletics', 'Swimming',
'Weightlifting' are visible
for i in range(len(ages)):
    # Offset labels to left for females, right for males
    label_offset = 20  # Base label offset
    if sex[i] == 'F':  # Female labels go left
        label_offset = -label_offset  # Move left
    elif sex[i] == 'M':  # Male labels go right
        label_offset = label_offset  # Move right

    # Ensure the specific sports are always labeled
    if sports[i] in ['Basketball', 'Athletics', 'Swimming', 'Fencing','Baseball', 'Gy
mnastics', 'Wrestling']:
        ax.text(ages[i] + label_offset, heights[i], weights[i], sports[i], size=9, co
lor='black')
        ax.plot([ages[i], ages[i] + label_offset], [heights[i], heights[i]], [weights
[i], weights[i]], color='gray', linestyle='--')
    # Otherwise, label every 15th point
    #elif i % 70 == 0:
       # ax.text(ages[i] + label_offset, heights[i], weights[i], sports[i], size=12,
color='black')
       # ax.plot([ages[i], ages[i] + label_offset], [heights[i], heights[i]], [weight
s[i], weights[i]], color='gray', linestyle='--')

# Add labels
ax.set_xlabel('Age')
ax.set_ylabel('Height')
ax.set_zlabel('Weight')

# Add title
ax.set_title('3D Plot of Average Age, Height, and Weight by Sex and Sport')

# Show plot
plt.show()
```

Correlation Analysis of Predicted Medal Counts

Introduction:

In this analysis, we examine the correlation between the predicted medal counts for Gold, Silver, and Bronze categories. Using a heatmap generated by Seaborn, we visualize the pairwise correlations between the medal counts. This analysis helps to understand the relationships between the different medal categories in the predictions, shedding light on how the predicted values for each type of medal are related to one another. The heatmap provides an intuitive representation of the correlation matrix, making it easier to identify any strong or weak associations.

In [48]:
```
sns.heatmap(pd.DataFrame(y_pred_reg, columns=['Gold', 'Silver', 'Bronze']).corr(), an
not=True, cmap='coolwarm')
plt.title("Predicted Medal Count Correlations")
plt.show()
```



Conclusion:

The correlation analysis of the predicted medal counts reveals an interesting pattern. We see a strong negative relationship of -0.9 between Gold and Bronze medals. This means that when the predicted number of Gold medals goes up, the predicted number of Bronze medals tends to go down, and vice versa. There's also a weak negative relationship of -0.18 between Gold and Silver, and a moderate negative relationship of -0.26 between Silver and Bronze. In simpler terms, these results suggest that as one type of medal (like Gold) increases, it could lead to fewer predictions for other medal types (like Bronze or Silver). This pattern may reflect how the model is distributing the total number of medals across different categories, possibly influenced by competition dynamics.

Athlete Conclusion

The presented data set offers a comprehensive analysis of the relationship between age, height, and weight across male and female athletes participating in various sports. The following summarizes the key findings:

Athletic Profiles:

- **Basketball, Beach Volleyball, Water Polo (Male):** These sports feature taller and heavier athletes, with Water Polo athletes averaging over 186 cm in height and 86-89 kg in weight.
- **Endurance Sports (Biathlon, Aeronautics, Female):** These sports generally favor shorter, leaner athletes, typically under 180 cm and around 75 kg for males, while female endurance athletes also show lower weights and moderate heights.
- **Weightlifting and Wrestling (Male):** Athletes in these sports tend to be shorter but heavier, with weightlifters averaging 169 cm and 80+ kg. Wrestlers tend to maintain heights around 172 cm, with weights varying by weight class (75 kg to 90+ kg).
- **Power and Agility Sports (Female):** Sports like Basketball and Bobsleigh demand higher average heights and weights for strength and leverage, with female athletes exhibiting similar builds.
- **Precision and Longevity Sports (Archery, Art Competitions, Male and Female):** These sports show older athletes, with some continuing well into their 40s, as precision and skill can offset physical decline.

Age Trends:

- **Intensive Physical Demands (Bobsleigh, Alpine Skiing, Wrestling, Weightlifting, Male):** Athletes peak in their late 20s to early 30s, reflecting the intense physical nature of these activities.
- **Water Polo (Male):** Athletes tend to be in their mid-20s, balancing both physical conditioning and experience.
- **Precision and Longevity Sports (Male and Female):** These sports show older athletes, who continue to perform successfully based on skill and experience rather than just physical condition.

Performance Insights:

- **Strength and Stamina Sports (Basketball, Baseball, Beach Volleyball, Wrestling, Male):** Athletes exhibit consistent height and weight patterns, indicating rigorous training and physical maintenance.
- **Water Polo (Male):** The physical demands of water polo require robust athletic profiles, with athletes maintaining strength, stamina, and agility.
- **Weightlifting (Male):** Weightlifters show high weight averages due to muscle mass, even with shorter statures.
- **Endurance Sports (Female):** Female athletes in endurance sports like Biathlon and Cross Country Skiing show optimized physical attributes with lighter builds and moderate heights for stamina.

Notable Observations:

- **Statistical Outliers (Male):** Some older athletes maintain competitive builds in precision sports, where skill and experience play a larger role.
- **Dynamic Profiles (Water Polo and Wrestling, Male):** These sports show athletes with varying body types based on weight class and performance demands.
- **Physical Demands (Female):** The diverse range of physical demands across sports highlights how specific physical attributes correlate with athletic success, such as agility, strength, precision, and endurance.

This combined analysis underscores the significant role of anthropometric factors in shaping athletic profiles across both male and female athletes. Understanding how height, weight, and age intersect with the demands of each sport offers valuable insights for training, talent identification, and performance forecasting. Further research can explore deeper correlations between specific body measurements and athletic success to optimize athlete development strategies.

In [ ]: