Introduction

The modern Olympic Games are quadrennial international sporting events featuring summer and winter sports competitions in which thousands of athletes from around the world participate in a variety of competitions. (Encyclopedia Britannica, 2018)

The goal of this report is to explore the participation and performance of Hungary in the Olympics. The main focus is on the total number, age and gender of athletes, the different Olympic sports in which Hungary has participated, and the number of medals won.

Data on Winter Games is not included in the analysis except where otherwise specified.

Research questions

- At which Olympic Games were Hungarian athletes the most successful?
- How many Hungarian athletes have participated in the Olympics?
- What are the most popular and most successful Olympic sports in Hungary?
- Who are the most decorated Olympians of Hungary?
- To what extent were the Hungarian Olympic Teams gender equal?

Key terms

IOC - International Olympic Committee
NOC - National Olympic Committee

Analysis

Data source

This report uses the "120 years of Olympic history: athletes and results" dataset from Kaggle (Kaggle.com, 2018), scraped by Randi H Griffin from www.sports-reference.com in May 2018 and released under CC0: Public Domain in CSV format, as its data source.

Exploratory data analysis and visualisation

Each row in the dataset corresponds to an individual athlete competing in an individual Olympic event. (Kaggle.com, 2018) The filtered dataset contains 6607 rows and 15 columns. The columns are:

Name	Description	Type
ID	Unique reference for each athlete	integer
Name	Athlete's name	string
Sex	Athlete's gender, M or F	string
Age	Athlete's age in years	float
Height	Athlete's height in centimeters	float
Weight	Athlete's weight in kilograms	float

Team	Name of the country or athletic club to which the athlete belonged	string
NOC	National Olympic Committee 3-letter code, HUN for Hungary	string
Games	Year and season	string
Year	The year the Game was held	integer
Season	Summer or Winter	string
City	Host city	string
Sport	Sport	string
Event	Event	string
Medal	Type of medal won, Gold, Silver, Bronze, or NaN	string

Pandas DataFrame.info() and DataFrame.describe() functions were used to generate descriptive statistics and a summary of the dataset:

```
print(hun_df.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6607 entries, 0 to 6606
Data columns (total 15 columns):
ID          6607 non-null int64
Name        6607 non-null object
Sex         6607 non-null object
Age         6461 non-null float64
Height      4726 non-null float64
Weight      4740 non-null float64
Team        6607 non-null object
NOC         6607 non-null object
Games       6607 non-null object
Year        6607 non-null int64
Season      6607 non-null object
City        6607 non-null object
Sport       6607 non-null object
Event       6607 non-null object
Medal       1135 non-null object
dtypes: float64(3), int64(2), object(10)
memory usage: 774.3+ KB
None
```

```
print(hun_df[['Age', 'Height', 'Weight']].describe())
```

```
              Age       Height       Weight
count  6461.000000  4726.000000  4740.000000
mean     25.359078   175.881507    71.660232
std       6.120854    10.531300    15.358222
min      12.000000   144.000000    35.000000
25%      21.000000   169.000000    61.000000
50%      25.000000   176.000000    70.000000
75%      28.000000   184.000000    80.000000
max      58.000000   209.000000   160.000000
```

The columns Age, Height, Weight, and Medal had missing values, as demonstrated here:

```
print(hun_df.isnull().sum())
```

```
ID            0
Name          0
Sex           0
Age         146
Height     1881
Weight     1867
Team          0
NOC           0
Games         0
Year          0
Season        0
City          0
Sport         0
Event         0
Medal      5472
```

The Medal column had 5472 missing values. Since only athletes who finish in the top three receive medals, the missing values in the Medal column were replaced with the value 'No medal':

```
hun_df['Medal'].fillna('No medal', inplace = True)
```

There was a significant amount of data missing from the Height, Weight and Age columns for the records from before 1960, as shown in the chart below:
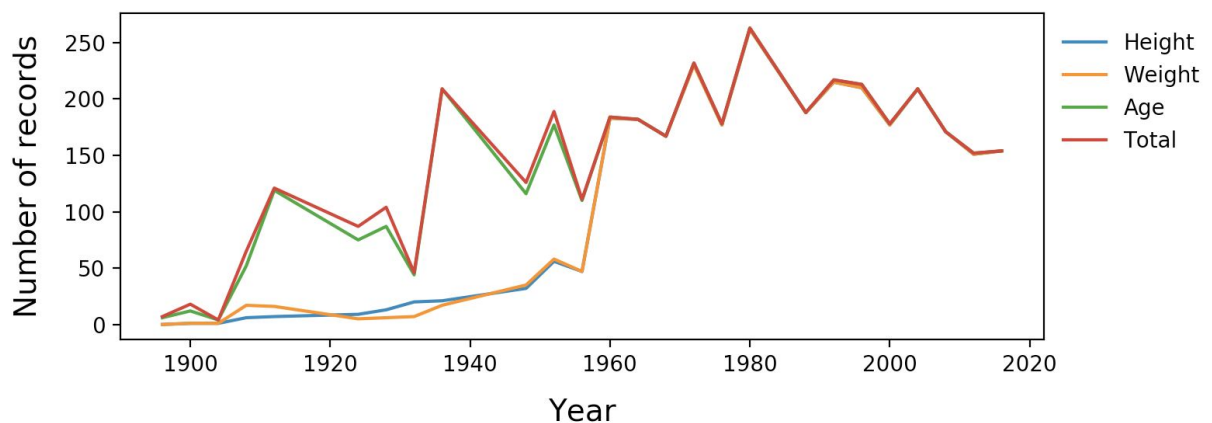
Figure 1: Number of height, weight and age data points per year vs total number of records per year

Hungary's athletes participated in every Olympic Games since 1896 with only two exceptions. Hungary was not invited to the 1920 Games after World War I, and the country was forced to participate in the Soviet-led boycott of the 1984 Summer Olympics (Wikipedia, 2018):

```
print(set(hun_df['Games']).symmetric_difference(events['Games']))
```

```
{'1920 Summer', '1984 Summer'}
```

Data preprocessing

Filtering

The dataset was reduced to only include results for Hungary by using boolean indexing to filter by the NOC code for Hungary: 'HUN'.

Transformation

The data types of the Medal, Sex and Season columns were converted from string to categorical, to enable statistical methods and plot types during analysis of those columns.

The IOC does not consider the 1906 Intercalated Games as Olympic Games (Olympics at Sports-Reference.com, 2018), so all 1906 events and competitors have been excluded from this analysis.

Number of athletes representing Hungary in the Olympics

Athletes can compete in one or more events across one or more Olympic Games. Every athlete has a unique ID number, so to determine the number of athletes representing Hungary in the Olympics, a count of the distinct values in the ID column can be used:

```
print('Total number of Hungarian athletes: ', hun_df['ID'].nunique())
```

```
Total number of Hungarian athletes:  2761
```
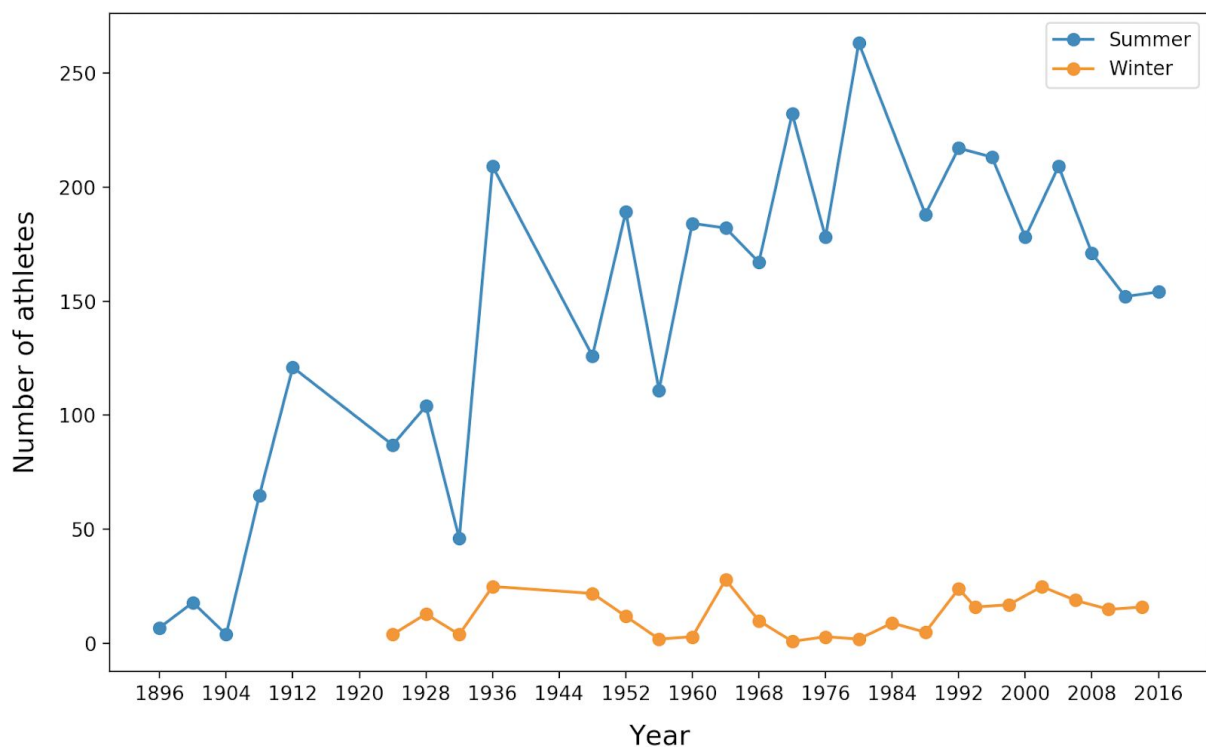

Figure 2: Line graph showing total number of Hungarian athletes per Olympic Games from 1896 to 2016 for each season.

The above graph reveals that the number of athletes competing in the Olympic Games has fluctuated significantly over the years. Participation in the Winter Games is notably consistently lower than that of the Summer Games. The total number of Hungarian participants reached a peak of 263 in 1980.
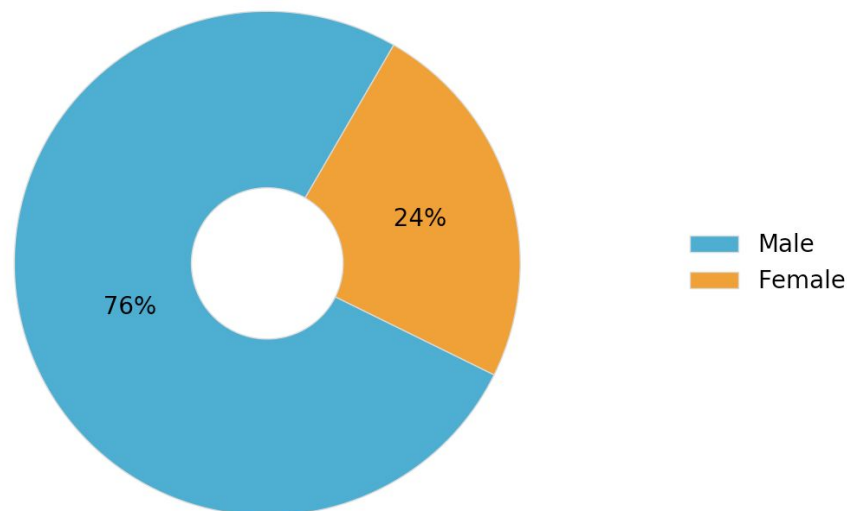
Gender

Figure 3: Pie chart showing overall ratio of participants in Summer Games from 1896 to 2016 according to gender.
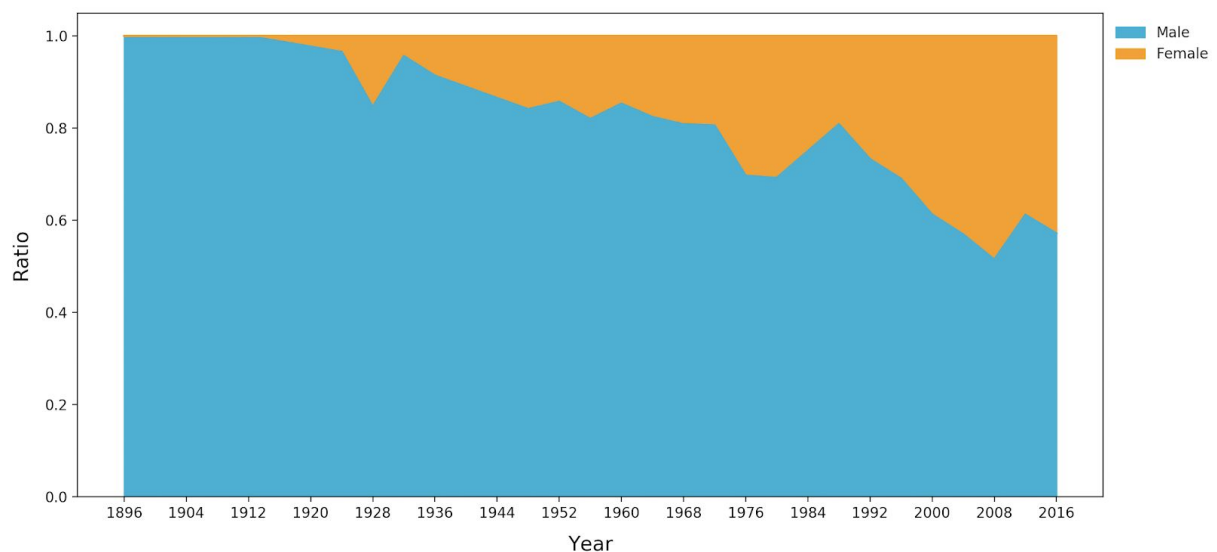
Overall, only one in four Hungarian competitors are women.

Figure 4. Area graph showing ratio of participants in the Summer Olympics according to gender

The above area plot reveals the shift of gender ratio over time. No Hungarian women participated in the Olympics until 1924, after which the proportion of women gradually increased over time. The sharpest uninterrupted rise was between 1988 and 2008, when the percentage of women increased from just 19% to 48.5%. Following a significant decrease (10%) in 2012, the proportion of women again increased in 2016.
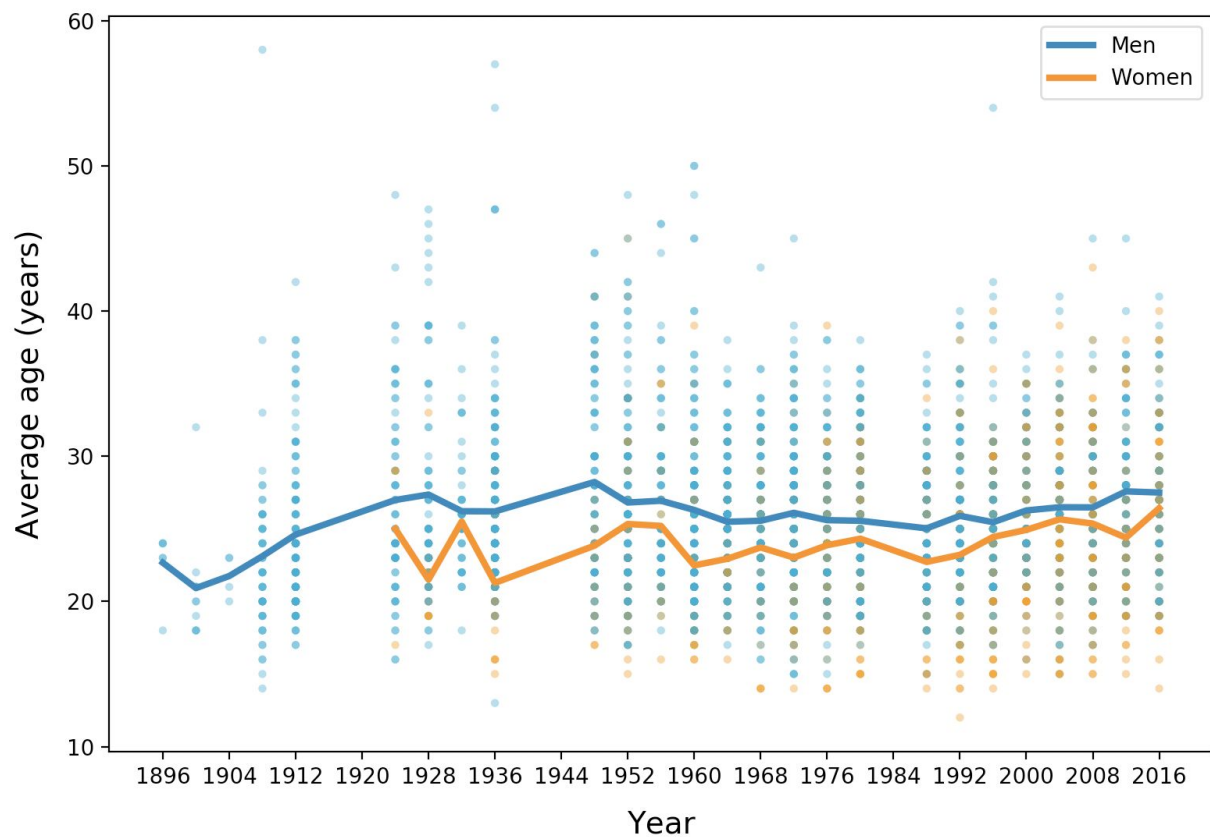
Age of athletes

Figure 5: Line graph showing the average age of athletes in Summer Games according to gender. Scatter plot showing the age of individual athletes, colour coded by gender

The above graph reveals that the average age of female athletes is consistently lower than their male counterparts.

The difference between male and female athlete age is explored in more detail in the following charts:
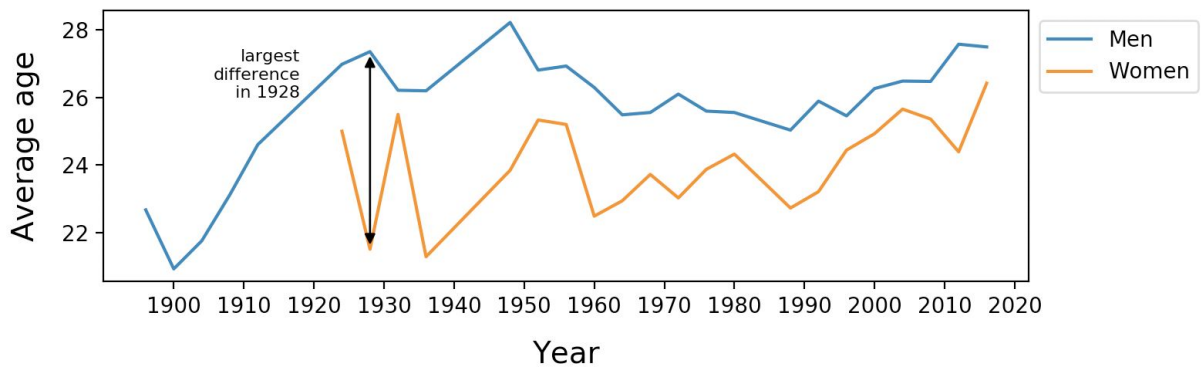

Figure 6: Line graph showing the average age of athletes in Summer Games according to gender.
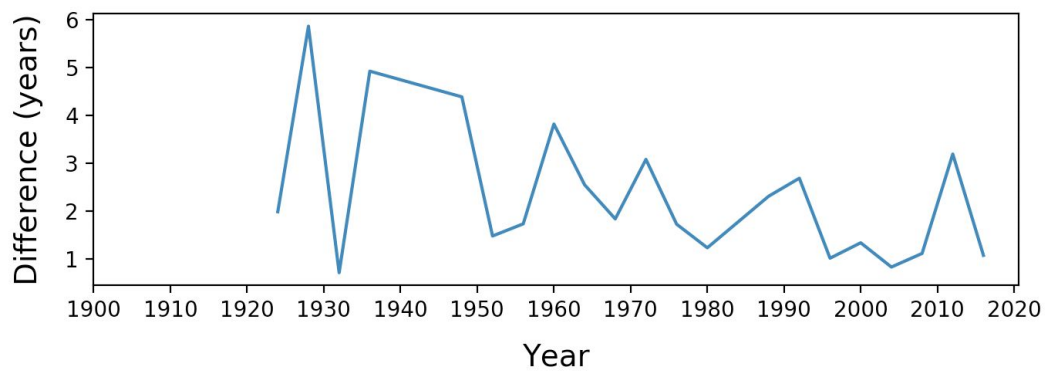
Figure 7: Line graph showing the difference between the average age of male and female athletes in Summer Games

The above charts reveal that the difference between the average age of male and female athletes has fluctuated over the years with no noticeable pattern.

Sports and events

Hungary has participated in 41 Olympic sports and 414 different events over the years.
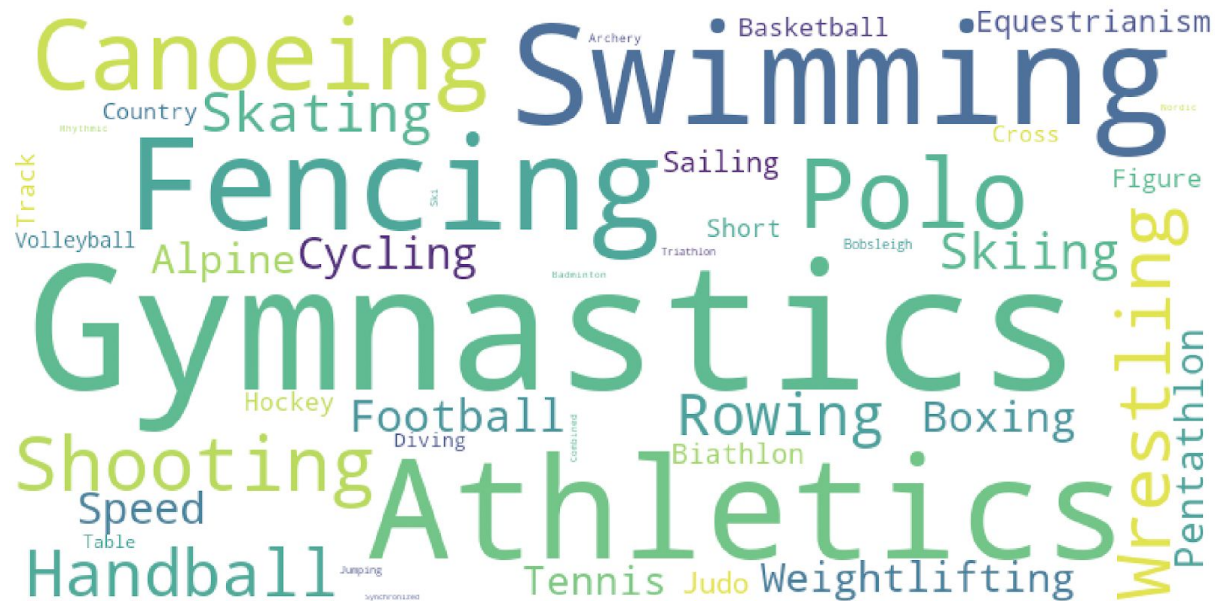
Figure 8. Word cloud showing the sports by number of athletes

The size of each word in the above visualisation is representative of its frequency in the dataset: most Hungarian athletes compete in Gymnastics, Athletics, Swimming or Fencing. The bar plot below shows the top 10 sports with the highest number of participants:
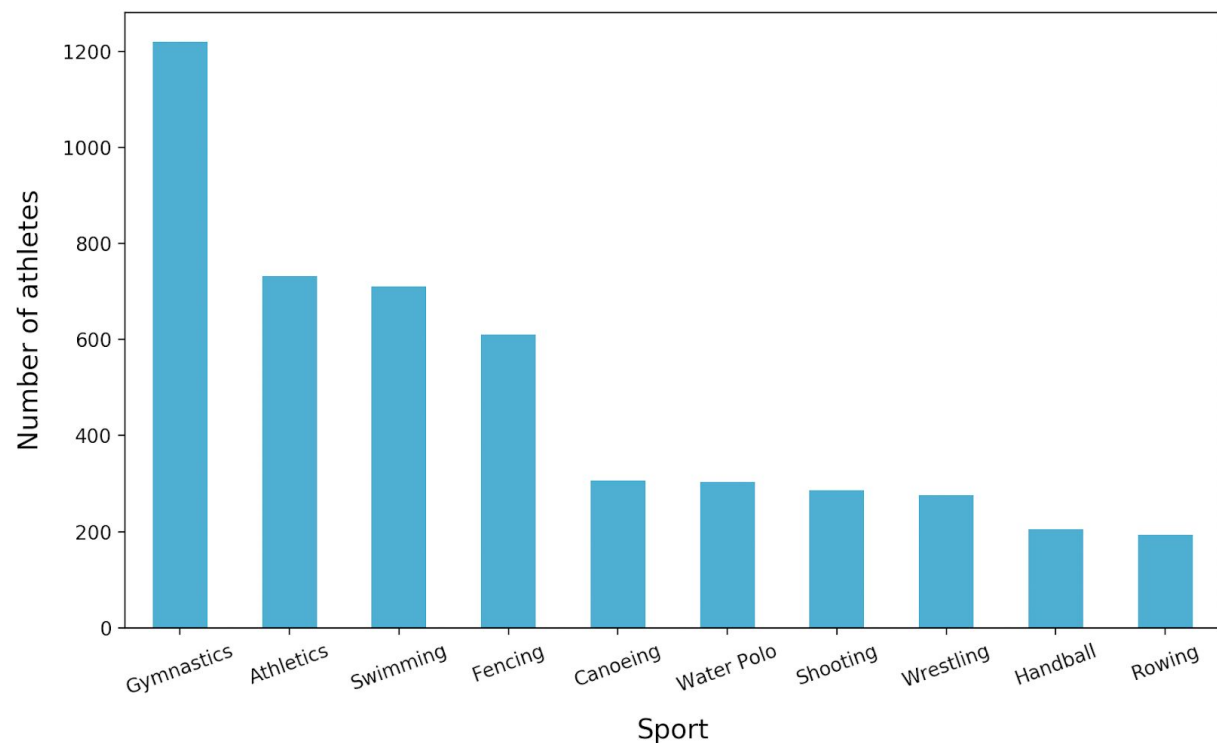

Figure 9. Bar graph showing number of Hungarian athletes by sport in the Summer Olympics

Medals

One way to measure a country's performance in the Olympics is to count the number of medals won by its athletes. Hungary has won 175 Gold, 147 Silver, and 168 Bronze medals across 18 different sports over the years. Hungarian athletes were the most successful at the 1952 Olympic Games in Helsinki with a total of 42 medals and the highest number of gold: 16.
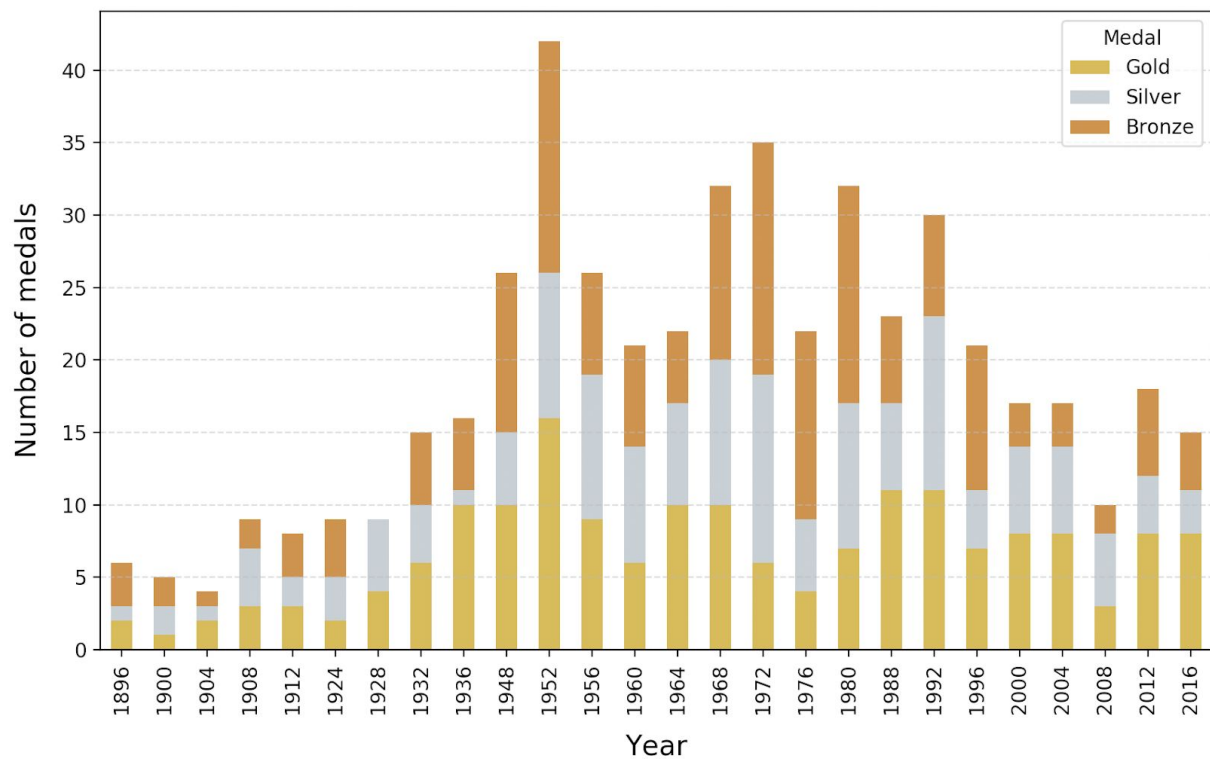

Figure 10. Stacked bar graph showing the total number of medals won by year and by type in the Summer Olympics

Hungary's medals have come predominantly from fencing, canoeing, swimming and wrestling events. Fencing is the leading sport with 87 medals, of which 37 are gold.
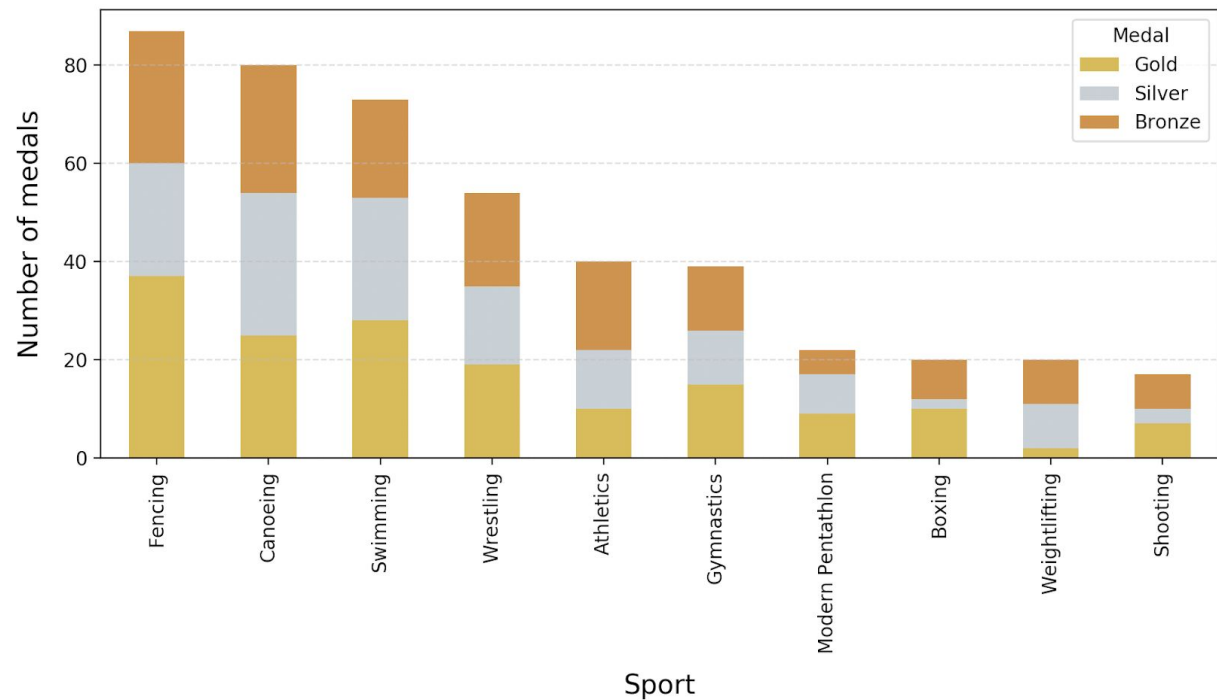
Figure 11. Stacked bar graph showing the number of medals won by sport and by type in the Summer Olympics. Sorted by total medal count, top 10 sports are shown
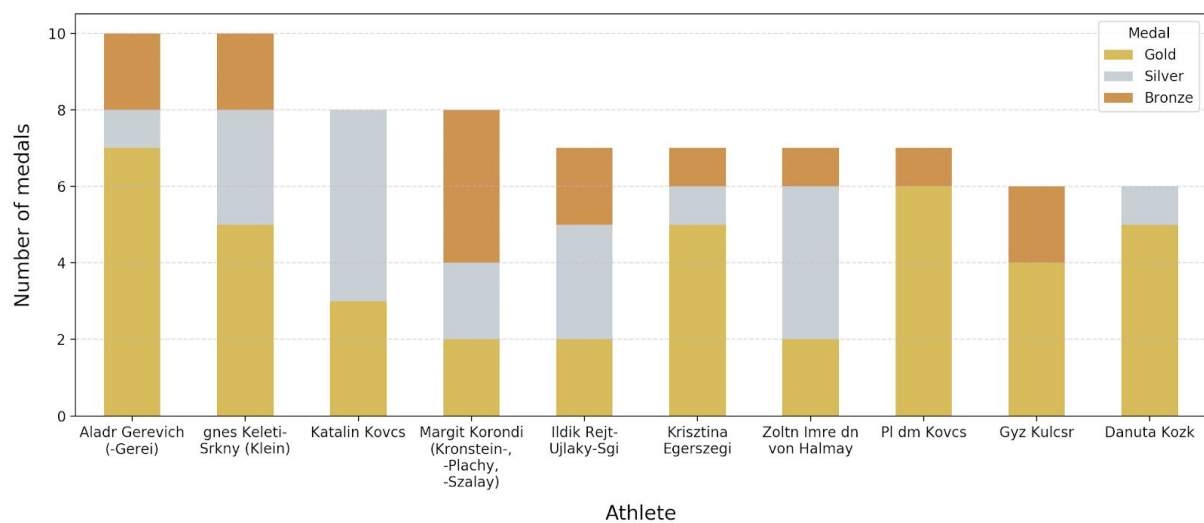
Figure 12. Stacked bar graph showing the number of medals won by athletes and by type in the Summer Olympics. Sorted by total medal count, top 10 athletes are shown

Several Hungarian athletes have won multiple Olympic medals. Aladar Gerevich, Hungary's most successful Olympian, was seven-times an Olympic gold medal winner.

Summary

In this report, different visualization techniques were used to gain insight into the participation and performance of Hungarian athletes in the Summer Olympics: a pie chart to show gender ratio, line charts to illustrate trend of participation over time, a bar chart to compare the number of athletes by sports and stacked bar charts to visualise the total number and type of medals won by year and by sport.

Key findings

- Hungary participated in every Olympics in the dataset except for the 1920 and 1984 Games
- Hungary has won a total of 490 medals in Summer Olympics
- Fencing is the top medal-producing sport for Hungary, for both total number of medals (87) and number of gold medals (37)
- Aladar Gerevich (fencing) and Agnes Keleti (gymnastics) are Hungary's most decorated Olympians, having each won a total of 10 medals
- The gender participation gap has narrowed over time, but still reflects a bias toward males

Evaluation and future work

I learned a lot about different visualisation techniques and choosing the right graph for my data. The report could have been improved by using a greater variety of graph types to show relationships between variables. Future work includes deeper analysis and comparison of Hungary's achievements with the achievements of other countries in the dataset.

Appendix A: Bibliography

Kaggle.com. (2018). 120 years of Olympic history: athletes and results. [online] Available at: https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results/home [Accessed 25 Nov. 2018].

Wikipedia. (2018). Hungary at the Olympics. [online] Available at: https://en.wikipedia.org/wiki/Hungary_at_the_Olympics [Accessed 13 Dec. 2018].

Encyclopedia Britannica. (2018). Olympic Games | History, Locations, & Winners. [online] Available at: https://www.britannica.com/sports/Olympic-Games [Accessed 25 Nov. 2018].
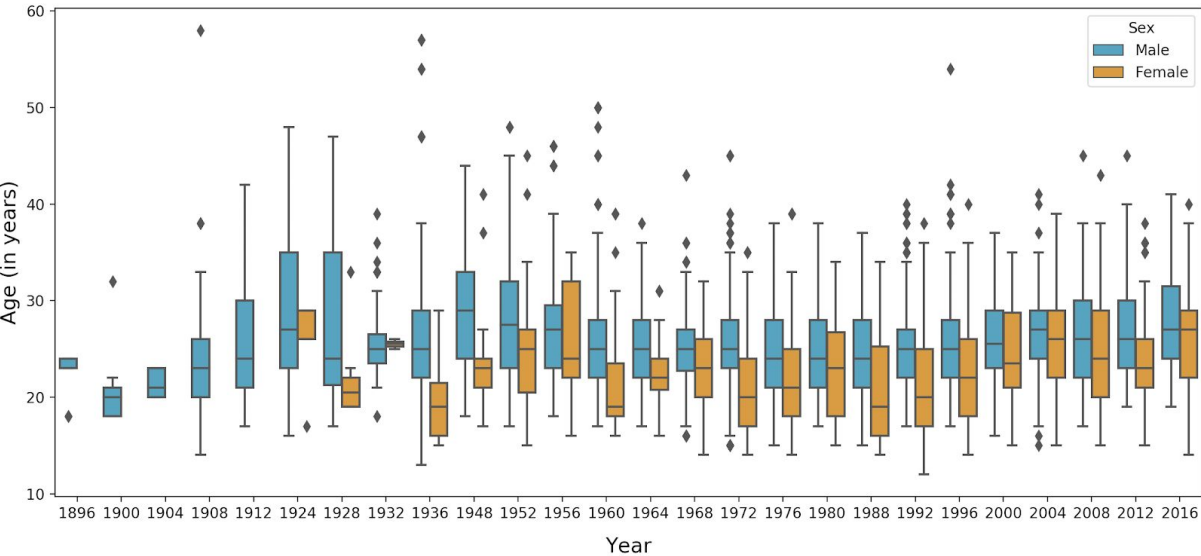
Olympics at Sports-Reference.com. (2018). What Events are Olympic? | Olympics at Sports-Reference.com. [online] Available at: https://www.sports-reference.com/olympics/about/events.html [Accessed 11 Dec. 2018].
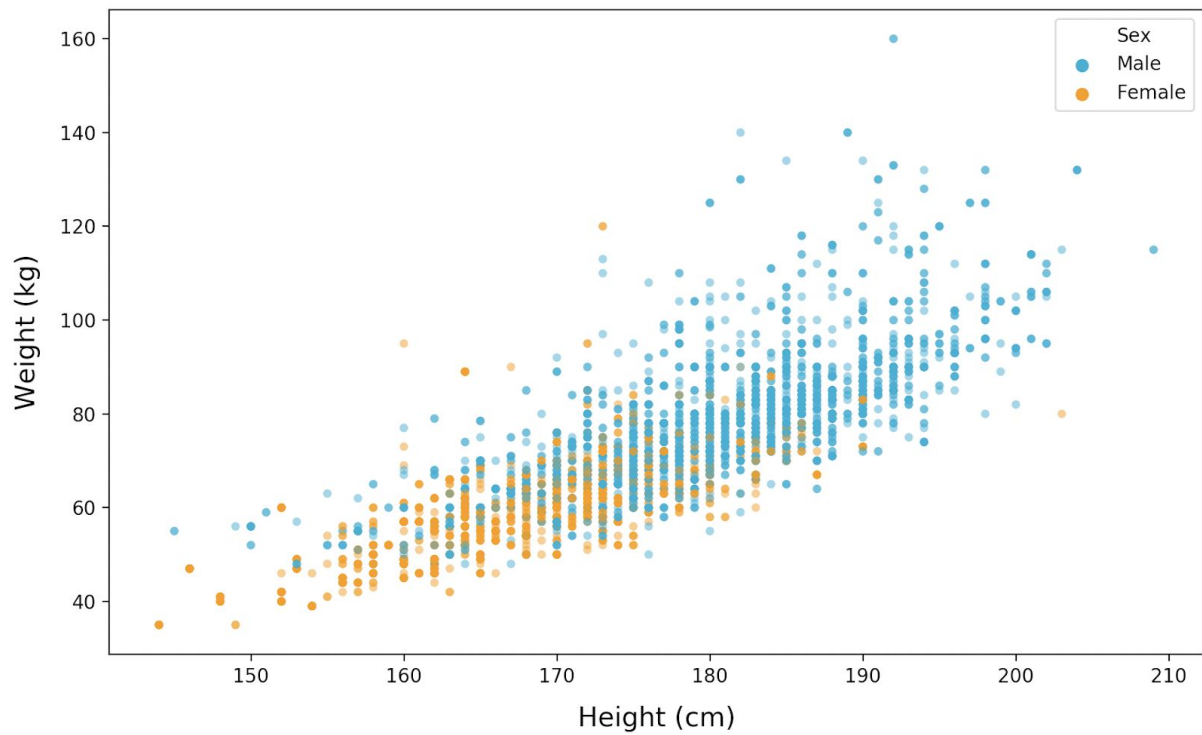
Appendix B: Additional graphs

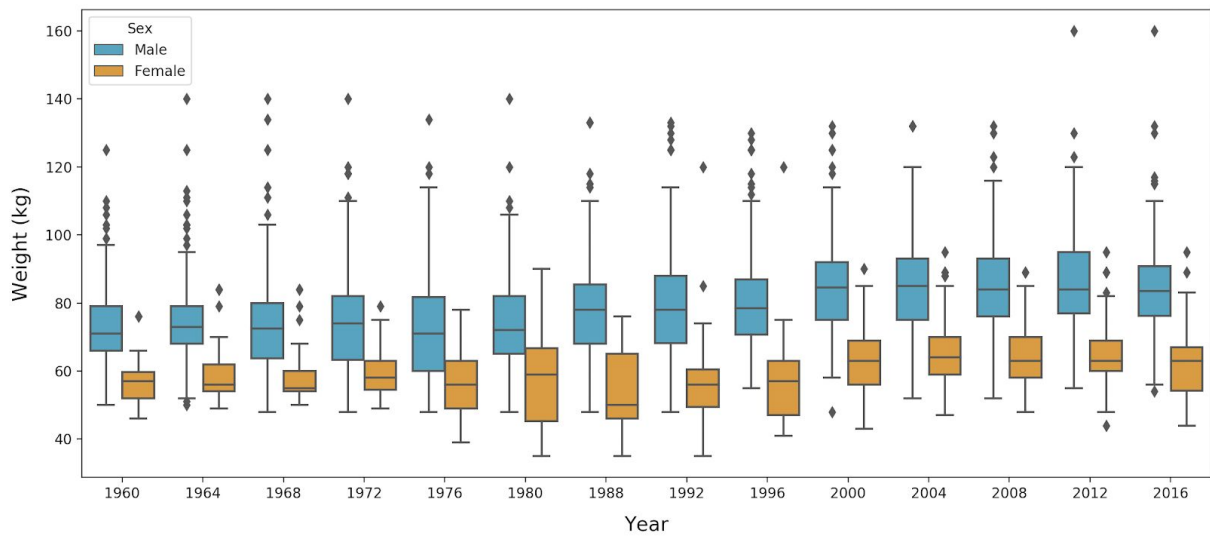**Popularity of sports by number of participants**
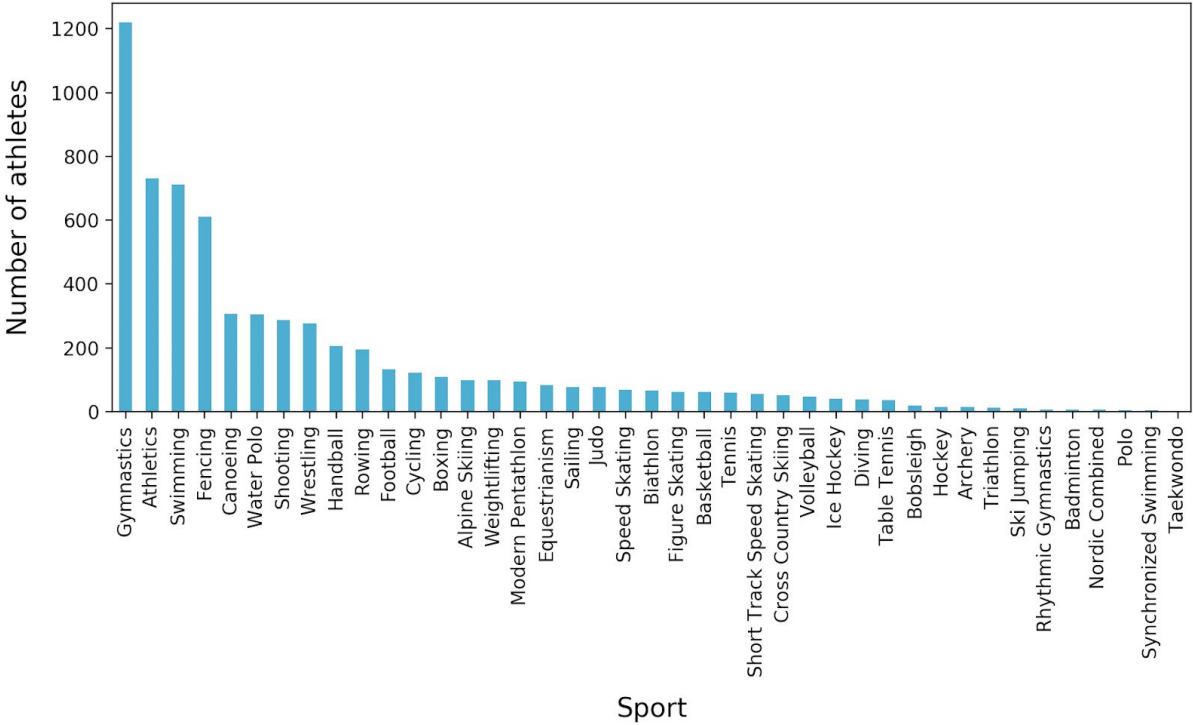
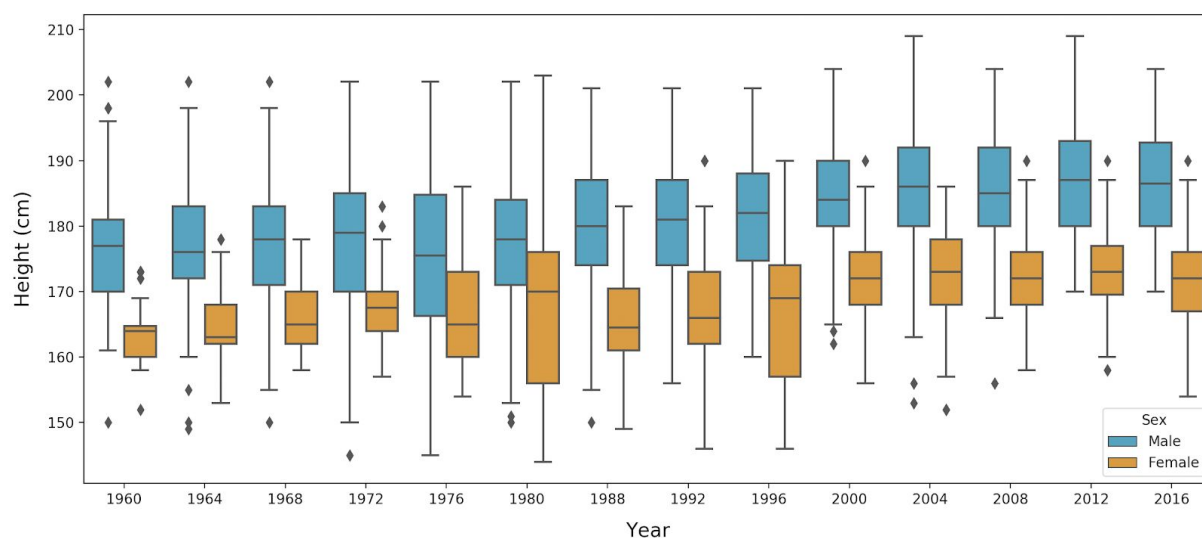**Age distribution of Hungarian athletes in Summer Olympics**

Height and weight of Hungarian athletes

Weight distribution of Hungarian atheletes in Summer Olympic games 1960-2016

Height distribution of Hungarian atheletes in Summer Olympic games 1960-2016

Medals won by Hungary according to gender