

Dataset Choice:

The dataset used is based on the Olympics. The dataset was used in previous labs, and can be found on Kaggle [here](#). The data is a historical dataset on the Olympic Games, aging from the first Olympic Games in Athens 1896 to Rio 2016. There are 15 columns in the dataset, which are explained below:

1. **ID** (integer): The unique identifier that is used for each athlete.
2. **Name** (character): The name of the athlete.
3. **Sex** (character): The sex of the athlete.
4. **Age** (integer): The age of the athlete.
5. **Height** (integer): The height of the athlete in centimetres.
6. **Weight** (double): The weight of the athlete in kilogrammes.
7. **Team** (character): The name of the team the athlete represents.
8. **NOC** (character): The National Olympic Committee the athlete represents.
9. **Games** (character): The Olympic Games the athlete participated at.
10. **Year** (integer): The year of the games that the athlete competed at.
11. **Season** (character): The season in which the games were held.
12. **City** (character): The host city of the Olympic Games that year.
13. **Sport** (character): The sport of the competing athlete.
14. **Event** (character): The event that the athlete participated in.
15. **Medal** (character): The medal that the athlete won in that event, if any.

Data Exploration Findings:

The first discovery when exploring and analysing the data is that the USA is the most winning country/National Olympic Committee in the competition's history. USA has won 5,592 medals in total, which is more than double the amount of the second most winning NOC (Soviet Union: 2,503 medals). This shows that USA has been historically strong at the Olympics since the Games' inception.

The second discovery is that the USA is most likely to win a gold medal when they win. They have more gold medals wins than silver and bronze wins combined. Comparing their wins to the other 4 most winning NOCs, their medal distribution is much more even, with approximately equal distribution between the three medal types.

The third discovery is that the USA has certain sports that they excel in. Their two most successful sports are swimming and athletics, winning around 1,060~ medals in each sport. Apart from those two sports, the USA is not nearly as dominant in other sports, with their third most successful sport, Rowing, only winning 370~ medals in total. It should also be noted that in both swimming and athletics, their gold medal winnings outweigh both their silver and bronze winnings. This shows that the USA has produced and continues to produce some of the best swimming and athletics athletes in the world.

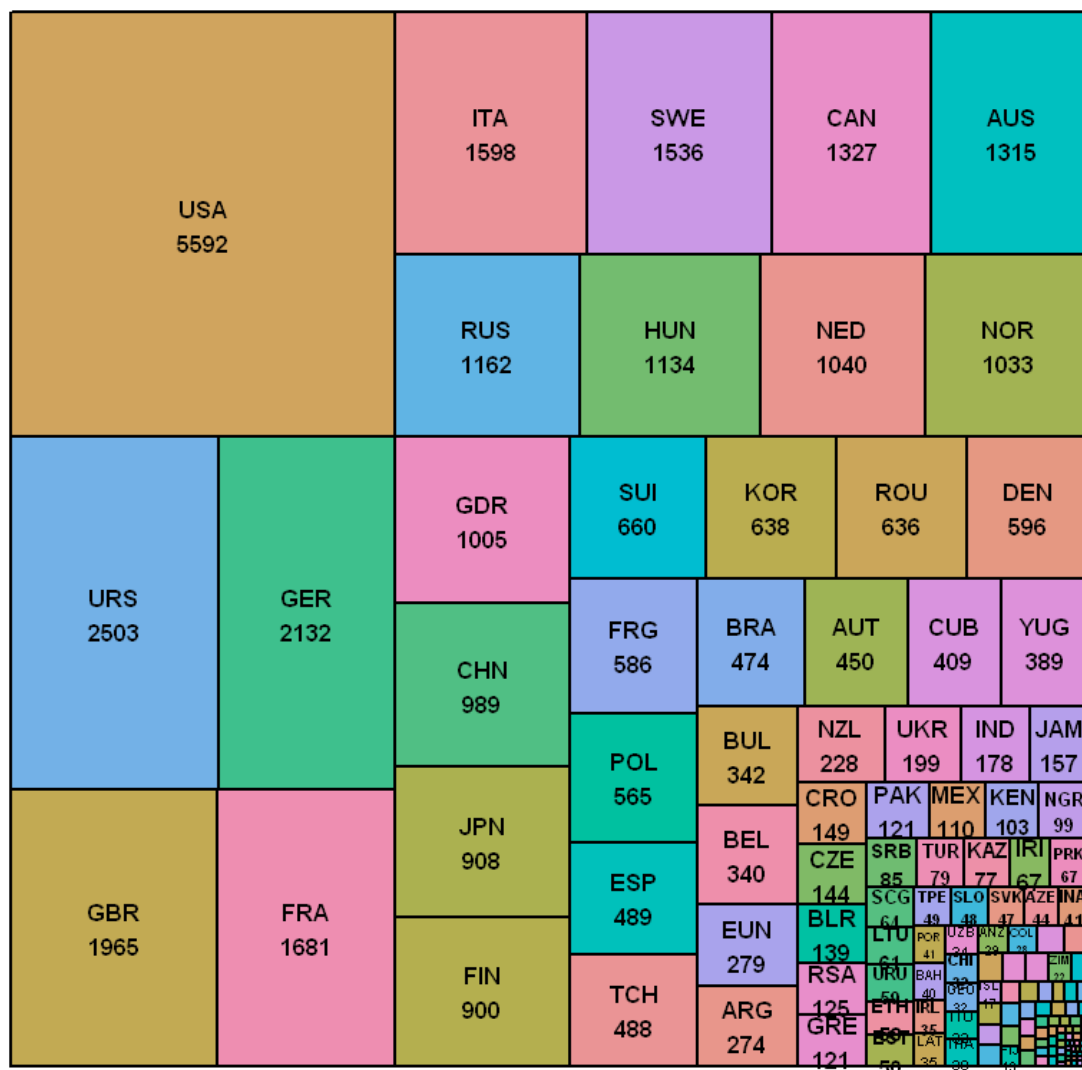
The last discovery is that the USA has more male medal winners than female medal winners. Looking at the distribution between male and female winners, male winners (68%) account for over double of female winners (32%). This shows that there is a gender imbalance between winners, and could suggest that the male athletes are trained better than female athletes.

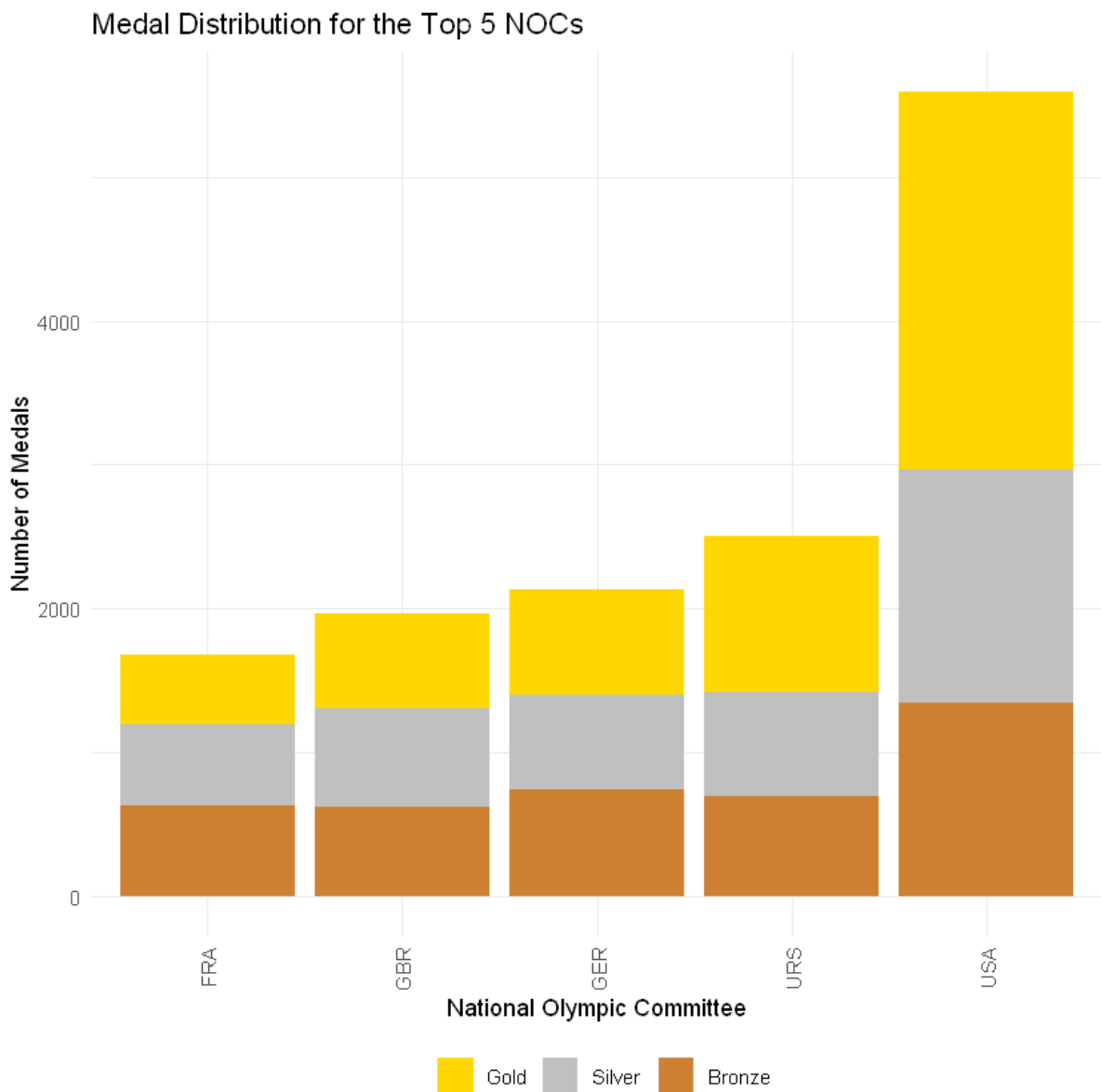
Big Idea:

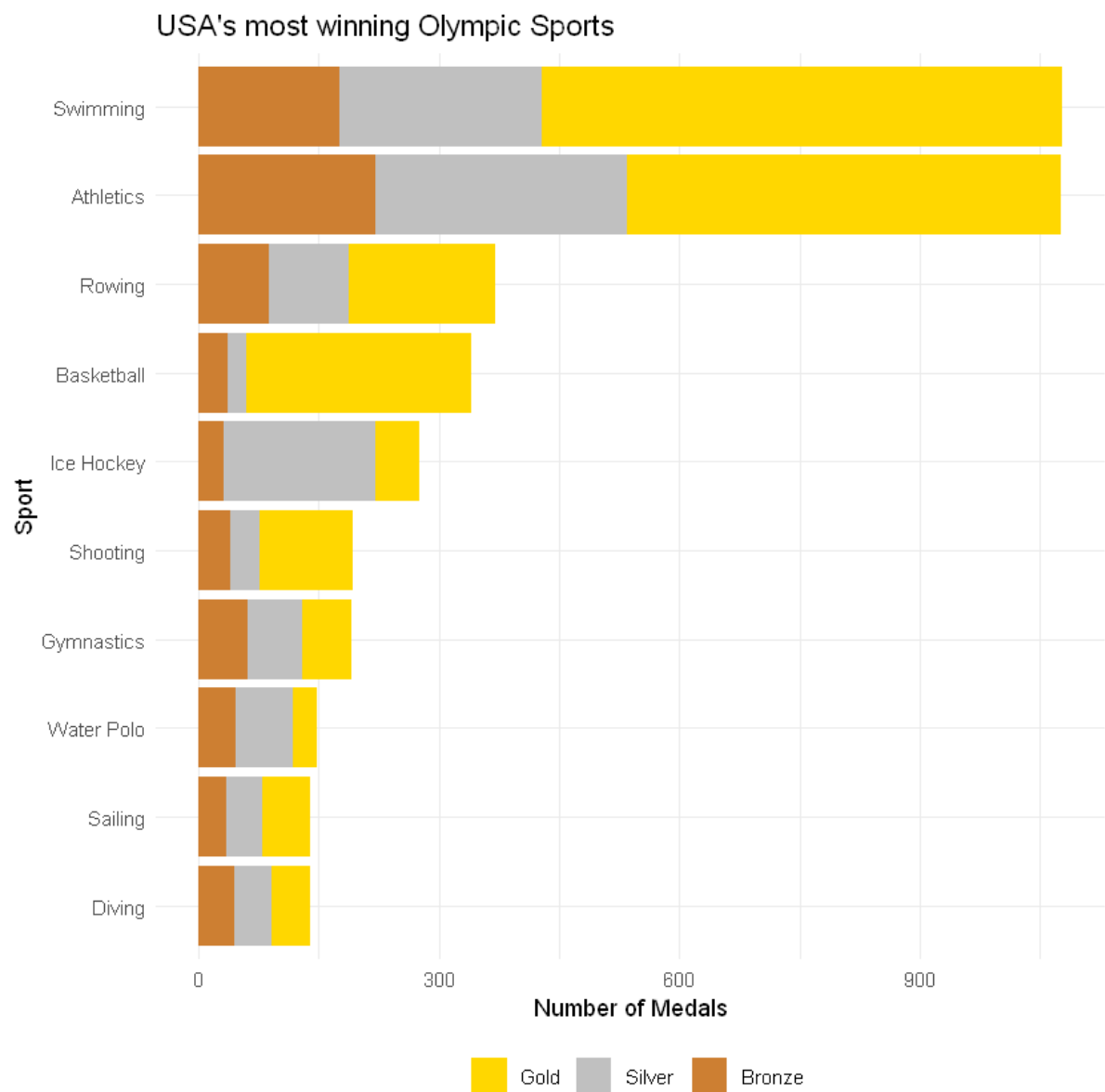
The USA is the most successful country in Olympic history, with a significant number of medals won than any other country. When winning a medal, a USA athlete is most likely to win a gold medal. The USA is dominant in certain sports, such as swimming and athletics. There is a gender imbalance amongst medal winners, with male winners accounting for over double the female winners.

Visualisations:

Total Medals won in Olympic History per NOC







Gender Distribution of USA's Medal Winners

