Winter Olympic Analysis: Through the Ages

Introduction

In a world with so much turmoil, the Olympics serves as an event that momentarily slows us all down and connects the world for a common cause: competition. When it comes to the Olympics, the thing that we all care most about from each competition is who won. Which countries won the most medals? Which athletes broke records or upset the leading contenders? Has a newly participating country dethroned the previous champion? Or perhaps an unassuming athlete has cleared the house. The details about the winner and the number of medals awarded to a country are the details people remember. We used this to guide our analyses and wanted to work in the theme of how the Olympics have changed over time. The dataset begins in 1896 and since then, our world has experienced immense change. Specifically, our project used a choropleth map, a contingency table, a line graph, a parallel coordinate plot, and network graph on an Olympic dataset to explore characteristics of athletes and each country's participation statistics. The following is a list of topics and questions our team sough to answer.

Male vs. female ratio for every sport
Number of medals won by athlete
Average age of the athletes
Total medals earned by host cities
Performance of medaling athletes
Time series of countries' medal and participant counts per game.
How many athletes, sports, and nations are there?
Where do most athletes come from?
What is the characteristic of the athletes (e.g., gender and physical size)?

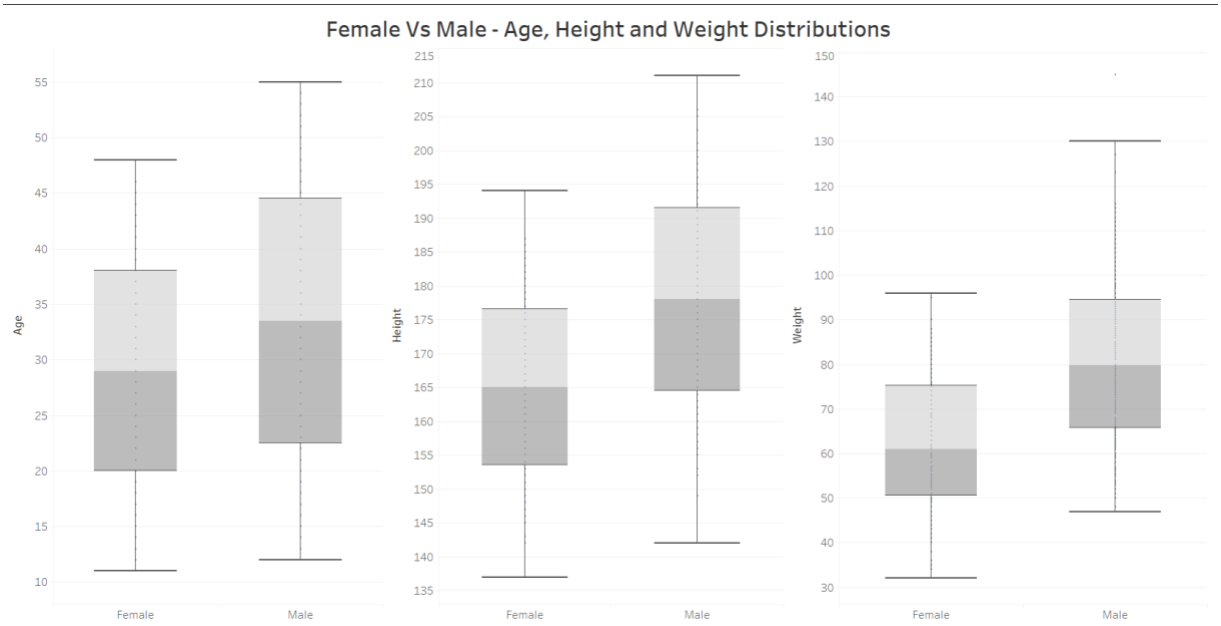
The data set was retrieved from Kaggle as the file, athlete_events.csv. It contains 271,116 data points of Olympic data for every summer (222,553) and winter (48,565) Olympic game held between 1896 and 2016. Each data point is an athlete participation record for a specific sporting event and contains 15 variables:

ID	Unique number assigned to each athlete
Name	Athlete's name
Sex	Athlete's Sex (M/F)
Age	Athlete's Age
Height	Athlete's height(cm)
Weight	Athlete's weight(kg)
Team	Athletes country
NOC	National Olympic Committee 3-letter country code
Games	Year and season
Year	Year Olympic games were held (1896 - 2016)
Season	Season of event (Summer/Winter)
City	Host city
Sport	Olympic sport category
Event	Competitions held within a sport
Medal	Medal type received (Gold, Silver, Bronze, or NA)

Exploratory Analysis

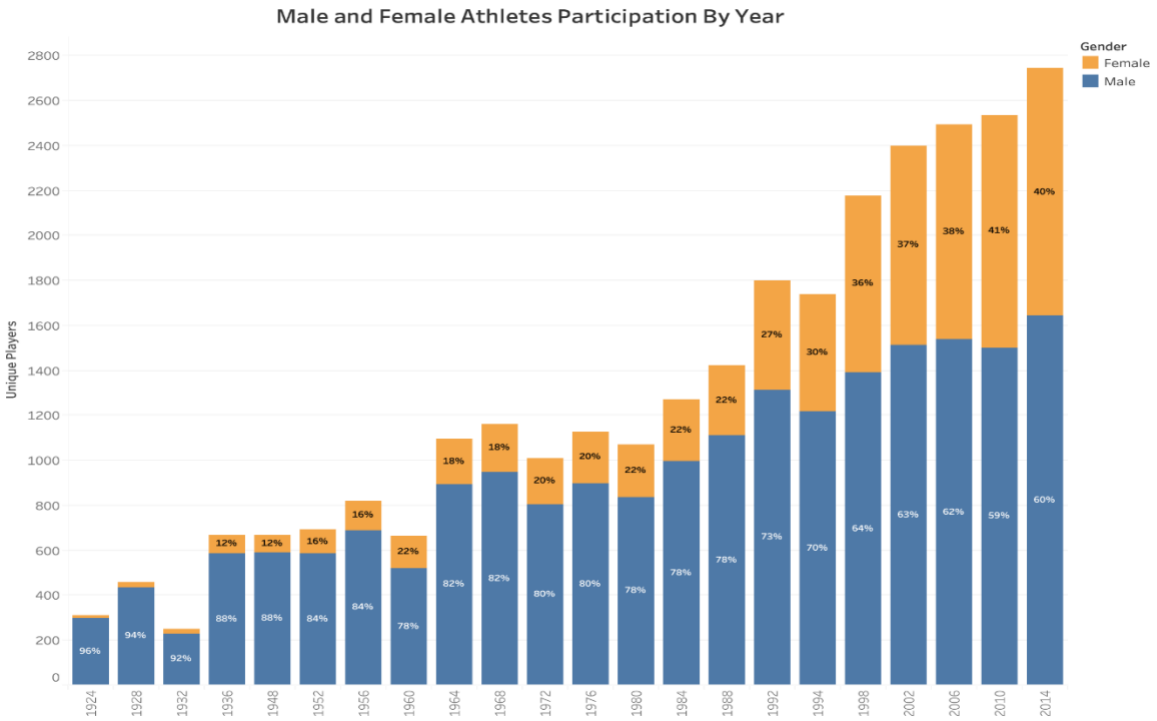
Before diving into different tracts of analysis, we needed to gain a basic understanding of the characteristics of our dataset. We explored the distribution of age, height, and weight of athletes that participated in the Winter Olympics. We also evaluated the proportion of male and female participation in the Winter Olympics – which proved to be a point of emphasis for one of the core visualizations. Another group member explored height and weight of athletes participating in the Summer Olympics and plotted the distribution of weight for athletes from 1896 to 2016. These three exploratory visualizations are merely a representation of the total number of visualizations that were created and the remainder of these can be found in the Appendix section of the report.

One of the key aspects in the Winter Olympics that the team analyzed and compared is the Male and Female Athletes participation and other criteria in the Winter Olympics. We firstly looked at the distribution of age, height and weight of Male and Female athletes that participated in the Winter Olympics. Below are the box plots that compares the distribution of age, height and weight for Male and Female athletes.



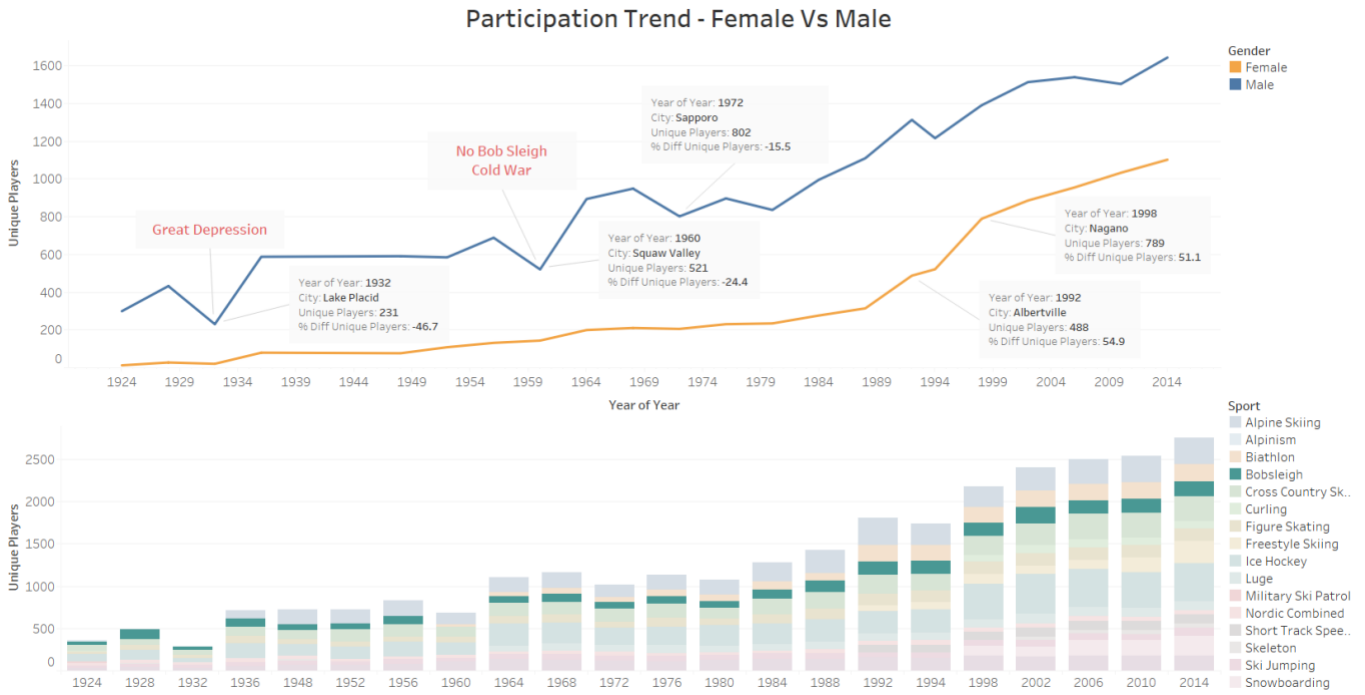
The above distributions suggest that the overall Female athletes are younger than the Male athletes, are shorter in height than the Male athletes and weight significantly less than the Male athletes.

Next, we look at the Male and Female Athletes participation by year. For this we created a bar plot of number of unique athletes participating by Year and divided by gender.



The above bar plot tells us that the participation in the Winter Olympics has increased yearly for both Male and Female athletes. During the initial Winter Olympics, the participation was almost entirely Male athletes. Since then that trend has changed and in some of the recent Winter Olympics, we see women athletes represent around 40% of the total participating athletes. The overall number of participating athletes has steadily increased with two major declines in participation in 1932 and 1960.

Visualizations



As we noticed in Exploratory analysis the overall participation dipped in the year 1932 and 1960. We further researched to find the reason behind and looked at the participation trend of Male and Female Athletes in the Winter Olympics. Now, we explore this a bit further. The above image contains two graphs. The first graph is a line plot that gives the number of unique Male and Female athletes that participated in the Winter Olympics by year with some important annotations for key events. The second graph gives the number of athletes that participated in the Winter Olympics by Year and Sport with a highlight on the Bob Sleigh event. The second graph help explains a finding in the first plot. In the first line graph we can see that there is a drastic dip in male participation in year 1932 and 1960.

Due to great depression the participation from male athletes was low in 1932.

And in the year 1960 the male participation went down because of two reason:

1. No Bob sleigh- In history of winter Olympics it was the first time that bob sleigh was not happening due to some budget constraints (second part of the graph shows histogram for same).

2. Cold war – Many of the countries did not sent their athletes due to the ongoing cold war situation.

And Overall participation from the female athletes seems improving.

Analysis and Discussion

For many athletes across the globe, reaching the Olympic games is the ultimate accomplishment. With such a robust dataset, our group wanted to approach the data from a wide scope. We knew that there were many avenues to pursue for an analysis, but we decided to focus on showing how the Olympic games have changed and subtleties that lie within the dataset that challenged our own notions of the Olympics.

One of the first realizations that stood out to group was how global the Olympics have become. Visualization C1 does a thorough job of displaying this, as it is easy to quickly observe how many countries have participated in the Olympics from 1896 through 2016. While each country doesn't necessarily win the same number of medals or send the same number of athletes to the games, we know that participation is now on a global scale.

We also saw the introduction of female athletes in the early half of the 1900s and wanted to further explore how the participation of women changed over time. Visualization C2 plots the participation trends of male and female athletes, in the Winter Olympics, from 1924 to 2016. Through this visualization, it's easy to see that women are beginning to make up a larger proportion of the athletes participating in the Winter Olympics. Additionally, the visualization shows dips in participation for specific years and calls out the reasons as to why this happened. In both cases, it was from major events whose impacts were felt on a global scale (Great Depression and Cold War). In many ways, this dataset mimicked many of the main events and accomplishments of the twentieth century.

As we got further into our analysis, we became interested in the height and weight of athletes and how they may vary depending on the sport or country participating. After exploring general trends in the data, we focused on showing the difference of median height across the ten most winning countries, that participated in the Summer Olympics, and the ten sports with the highest variance of height amongst its participants. We were able to learn that it really depends most on the sport and country at hand. While similar trends were observed across sports, there were countries that stood out as either being generally taller or shorter than the competition.

We also wanted to evaluate if we could identify trends across the top performing countries for the Winter Olympics, in hopes of illuminating a blueprint for future success. Unfortunately, we observed a variety of trends and the true takeaway is that it doesn't matter how many athletes a country sends, or how young the athletes are, or how many events they participate in; if the country focuses on sending the best athletes, its medal winning rate will prove to be strong. Lastly, the performance of athletes who competed in the second most popular winter sport, Alpine Skiing, was analyzed to uncover trends amongst medaling athletes. Through visualization C5 we were able to observe how winning in one event related to wining in other Alpine Skiing Events.

With the visualizations included in our report, we truly believe that we've only scratched the surface. For instance, it might be interesting to evaluate how the physical attributes of medaling athletes have changed over time or how the height and weight of medalists compare to each other depending on the sport. Other areas to pursue could include researching how countries with lower participation rates tended to perform, in terms of medals won, and which countries have made more of an effort to send equal numbers of male and female athletes to the Olympics. Regardless of the research topic pursued, there is a plethora of data at our disposal and there certainly are insights waiting to be discovered.

APPENDIX

APPENDIX 1 – Individual Reports

For the final project analysis, I was interested in identifying the features of male and female athletes participating in winter Olympics and for this I plotted box plot to see overall variation among age, height and weight. The distributions suggest that the overall Female athletes are younger than the Male athletes, have a lesser height than the Male athletes and weights significantly lesser than the Male athletes.

Then focused on one of the key aspects in the Winter Olympics that the overall Male and Female Athletes participation by year. For this I created a bar plot of number of unique athletes participating by Year and divided by whether they are Male or Female. I noticed during the initial Winter Olympics, the participation was from Male Athletes majorly but over the period that trend has changed and in some of the recent Winter Olympics we see Women Athletes constitute ~40% of the total athletes participating. also noticed that the overall participation dipped in the year 1932 and 1960.

On further investigation found that in 1932: Due to great depression the participation from male athletes was low. And in the year 1960 the male participation went down because of two reason:

1. No Bob sleigh- In history of winter Olympics it was the first time that bob sleigh was not happening due to some budget constraints (second part of the graph shows histogram for same).

2. Cold war – Many of the countries did not sent their athletes due to the ongoing cold war situation.

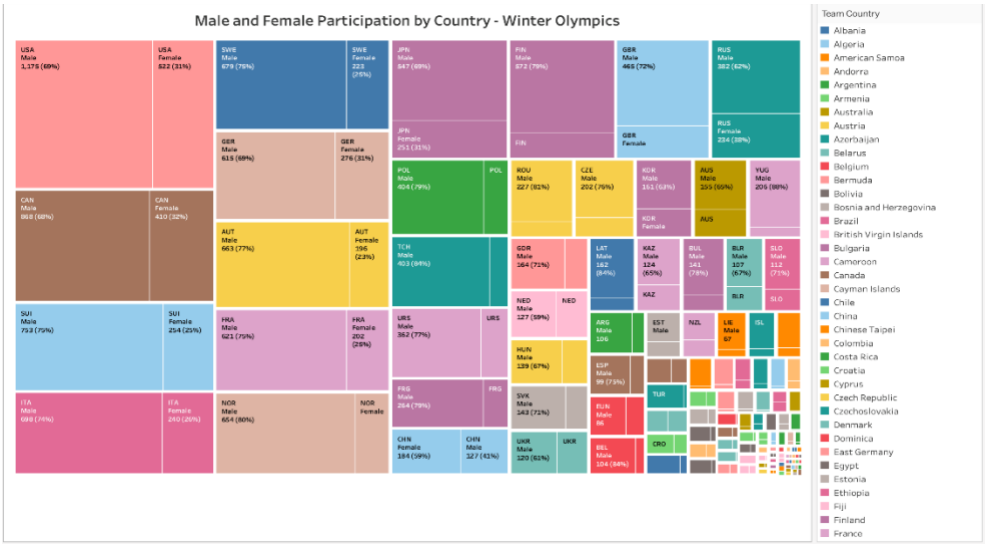
And Overall participation from the female athletes seems improving.

I was also interested in understanding the medals won by male and female athletes for their respective countries. Hence Created a Mosaic plot that depicts the different type of Medals Won by Country and Gender (appendix (C2 figure 6)). This a great visualization to understand how countries Male and Female Athletes performed in the Winter Olympics (in terms of getting Gold, Silver and Bronze Medals). In the (appendix (C2 figure 6)) graph we can see that the United States and Canada has the highest number of medals in Winter Olympics.

APPENDIX 2 – Results of formative analysis

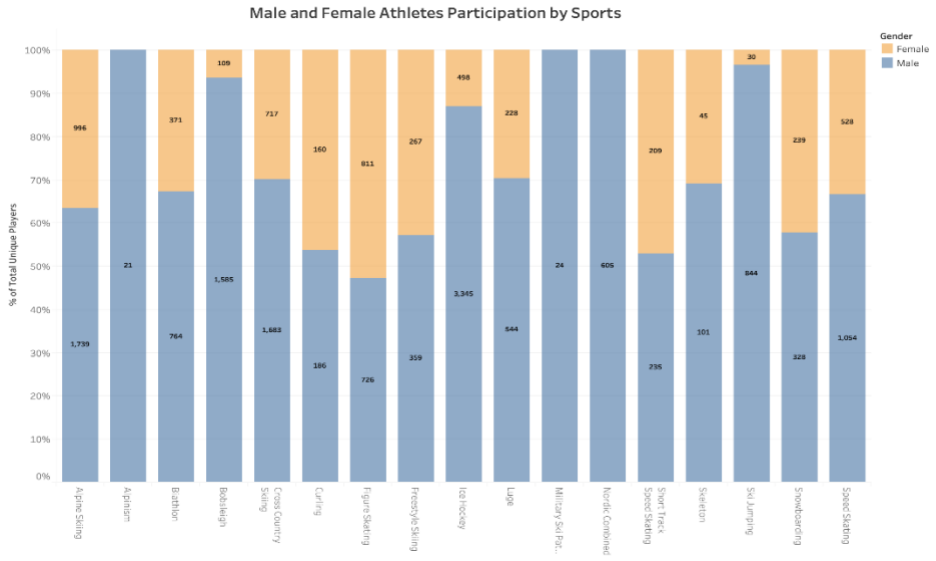
The following graph is a heat map showing the Male and Female Athletes participation by Country for all Winter Olympics held until 2015.

Figure 1



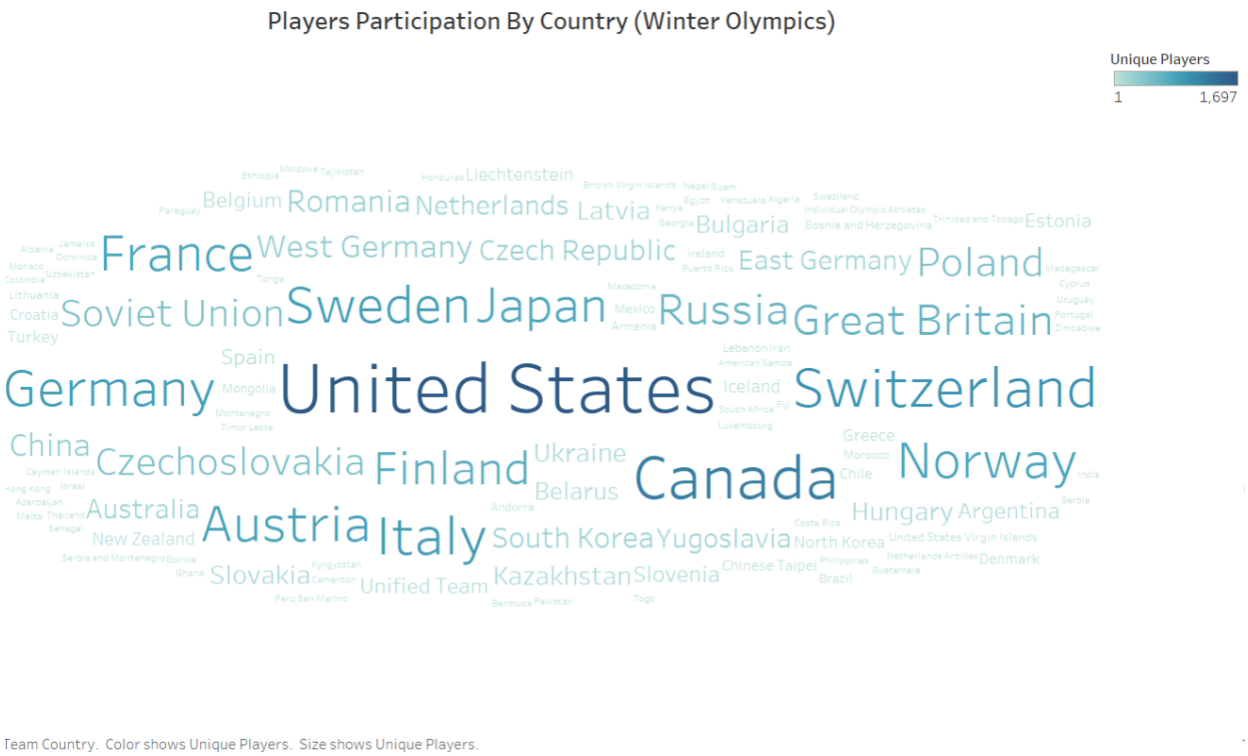
The above graph helped us visualize the participation by various countries in the Winter Olympics as well as the participation of Male and Female Athletes by Country in the Winter Olympics. We see that most of countries that participated in the Winter Olympics have more Male athletes compared to Female athletes.

Figure 2



The above graph shows the % of Male Vs Female Athletes Participation by various Sports in the Winter Olympics. Most of the games have more Male Athletes participating compared to Female Athletes. However, few sports like Figure Skating and Short Track Speed Skating have more participation from Female Athletes. Like the previous analysis it would interesting to see the trend in participation of these sports over time.

Figure 3



The above Word Cloud gives the information related to the Players Participation by Country in the Winter Olympics. The size and the color are used as an indicator for number of unique players participating from a given country in the Winter Olympics.

Steps: -

1. Create Unique Players calculated field as: COUNTD([ID])
2. Drag Team Country field to Text
3. Set the Mark to Text
4. Drag the Unique Players field to Color and Size

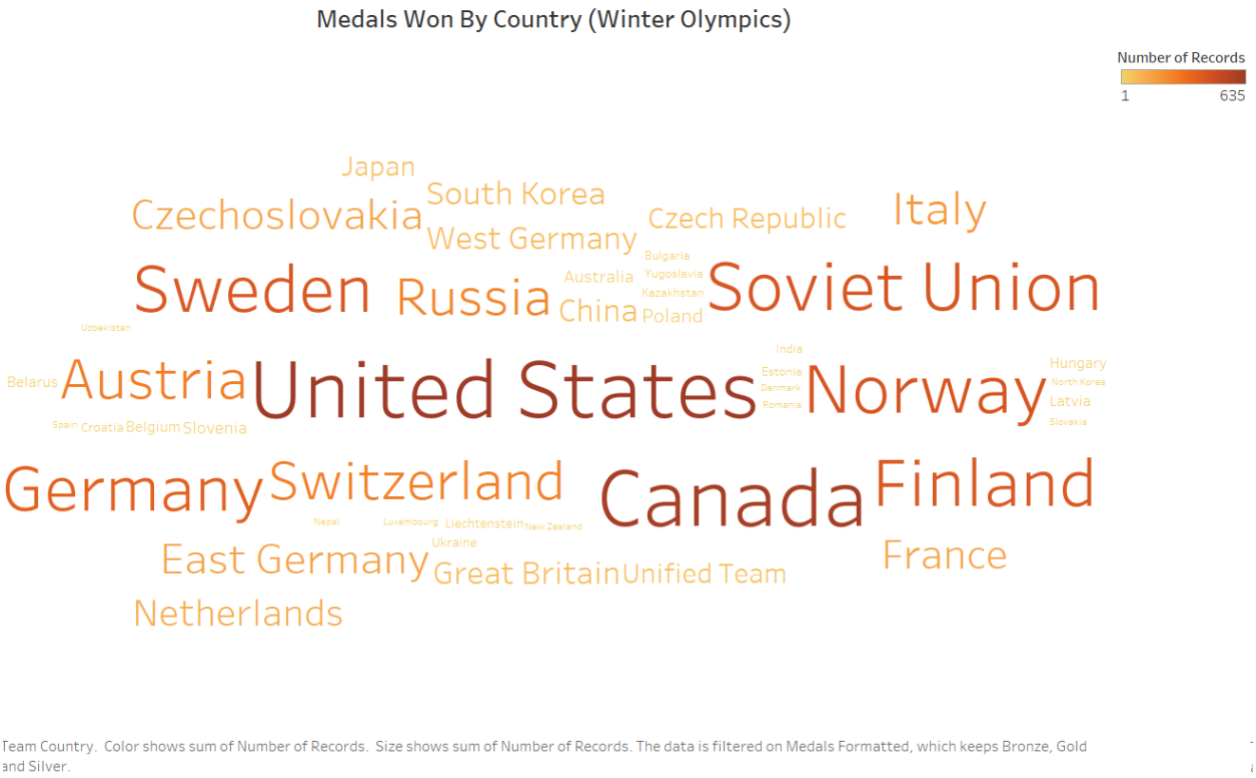
Figure 4

The above Word Cloud gives the information related to the Unique Games by Country in the Winter Olympics. The size and the color are used as an indicator for number of unique games played by a country.

Steps: -

1. Create Unique Games calculated the field as: COUNTD([Games])
2. Drag Team Country field to Text
3. Set the Mark to Text
4. Drag the Unique Games field to Color and Size

Figure 5

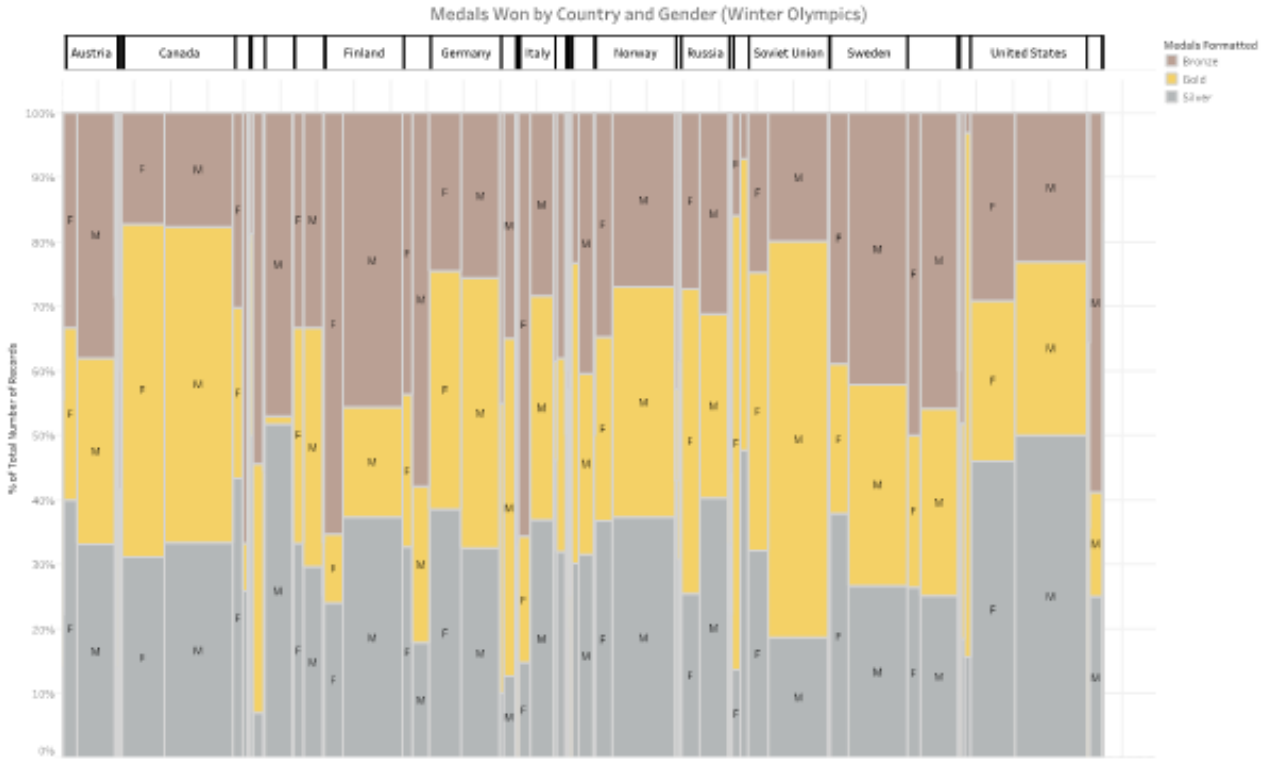


The above Word Cloud gives the information related to the Medals Won by Country in the Winter Olympics. The size and the color are used as an indicator for number of medals won by a country.

Steps: -

1. Drag Team Country field to Text
2. Set the Mark to Text
3. Drag the Number of Records fields to Color and Size
4. Filter the worksheet based on Medal Formatted; include only Bronze, Silver and Gold.

Figure 6



The figure above is a Mosaic plot that depicts the different type of Medals Won by Country and Gender. This a great visualization to understand how countries Male and Female Athletes performed in the Winter Olympics (in terms of getting Gold, Silver and Bronze Medals). In the above graph we can see that the United States and Canada has highest number of medals in Winter Olympics.

Steps: -

1. Create/Add Following Metrics to the Tableau Visualization
    a. Number of Records: Auto-generated
    b. % of Number of Records by Medals Won by each country and gender
    c. Records per Column: {EXCLUDE [Medals Formatted]: SUM([Number of Records])}
    d. # of Medals.