



# The Winter Olympics



Sita Robinson

August 2018

**Abstract**—The Winter Olympics is a big international sporting event held in the winter. It includes a number of winter sports. The current winter sports include alpine skiing, biathlon, bobsleigh, cross country skiing, curling, figure skating, freestyle skiing, ice hockey, luge, nordic combined, speed skating, skeleton, ski jumping, and snowboarding. The purpose is to look at the Olympic data from the years 1924 to 2014 and learn more about the sports, events, the best and worst countries. The data was provided by the data.world website.

**Index Terms**—Information Visualization, Geographic Visualizations, Data Analysis

---

## 1 INTRODUCTION

THE Winter Olympics include a number of sports and events where a number of countries compete. The sports that occur during the Winter Olympics are either on snow or ice. These sports happen every four years and have been occurring from 1924. The dataset I chose was the Winter Olympic dataset from data.world and includes data from 1924 to 2014 which is every year not including the last Olympics in 2018. The data is provided in an Excel file format (.xlsx). The data includes 2865 rows and 9 columns. The names of the columns are "Year", "Sport", "Event", "Country", "Gender", "Medal Rank", "Medal", "Name of Athlete or Team", "Age of Athlete". The only missing values are in the "Age of Athlete" column and make up 692 out of the 2865 total rows. Other than that the data seems to generally be pretty clean so not much pre-processing was needed to produce the visualizations.

## 2 PURPOSE

The purpose of this project is to look at data on the Winter Olympics. The Winter Olympics have been occurring for a long time. Over those years, many different new sports and events have been added. The

first step is to look at the sports and events as well as look at who has overall won the most medals. A sport has a number of events that qualify for that sport which are specific to either men, women, or mixed. Then the next step is to look at who in the 2014 Olympics which is the last year this data covers won the most medals. The overall point of looking at this data is to find out more about the Olympics and which countries seem to be doing the best.

## 3 APPROACH

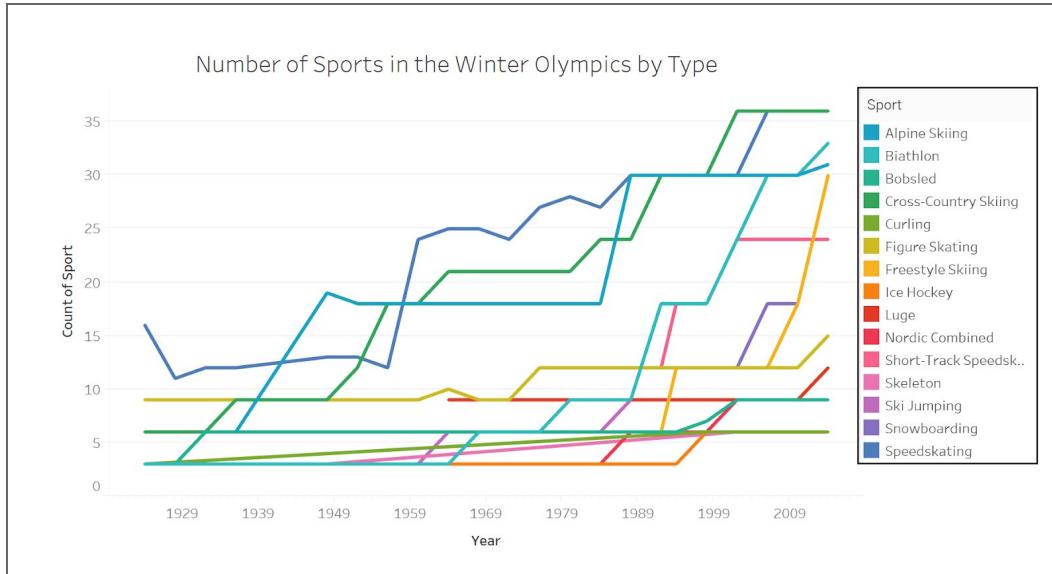
The pathway to approaching this was first looking at the Excel file after downloading it, looking at each column, and figuring out the purpose of the columns. The number of rows and columns were then taken into account followed by looking for null values. Then picking columns of interest, the proper visualizations were chosen. Some of the better visualizations are shown in the Visualizations section while the others are included in the Appendix of this paper.

## 4 VISUALIZATIONS

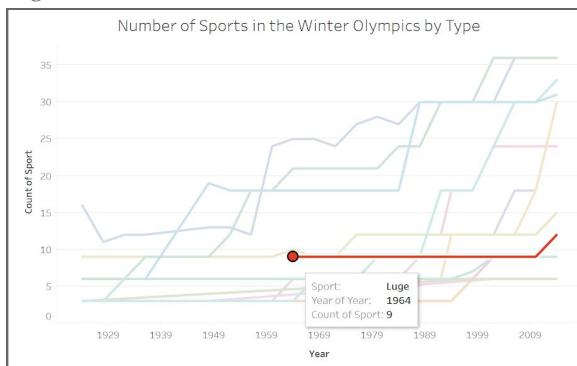
The visualization software I used was Tableau and RAWGraphs. Specifically from Tableau I used

geographic maps including the symbol map and the regular map options. I also used the horizontal bar and line option. For RAWGraphs, I used the circular dendrogram option.

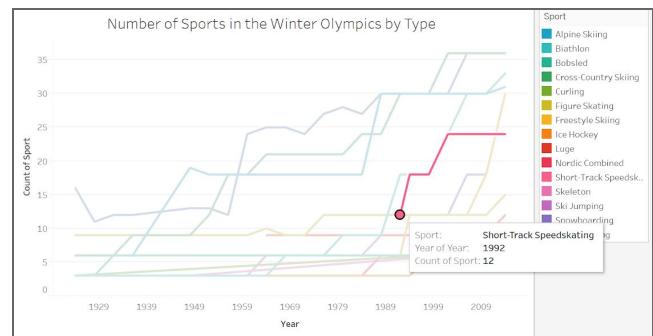
#### 4.1 FIGURE 1: NUMBER OF SPORTS OVER ALL YEARS



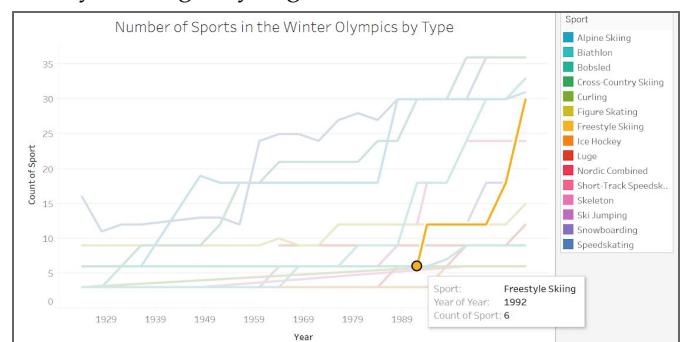
Shows the count in the dataset of the number of sports over the years by category. Luge seems to have only began in 1964.



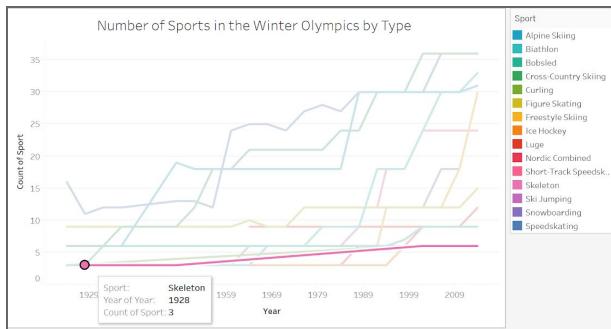
Short track speedskating only began in 1992.



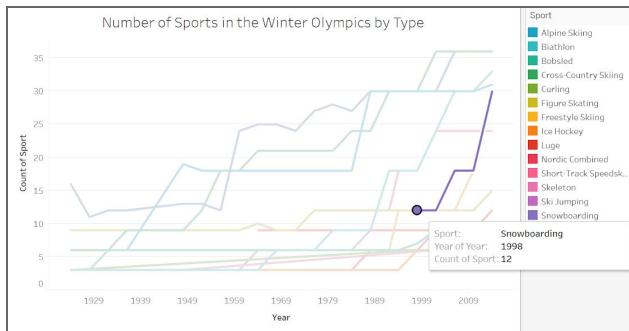
Freestyle skiing only began in 1992.



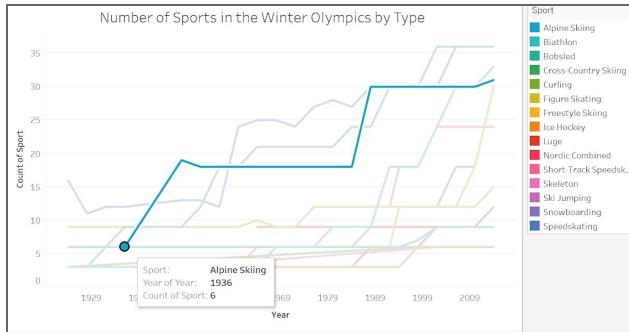
Skeleton seems to have started in 1928.



Snowboarding in this dataset started in 1998.

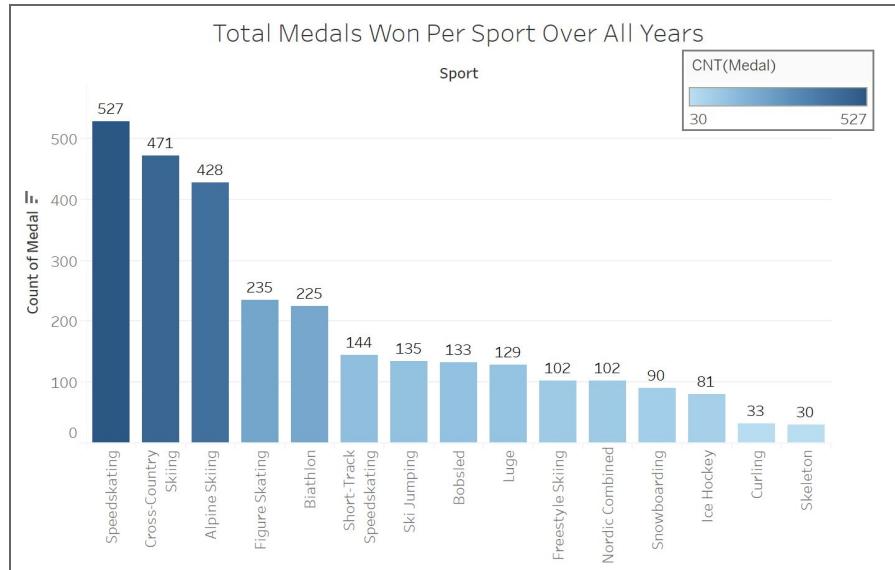


Alpine Skiing seems to have started in 1936.

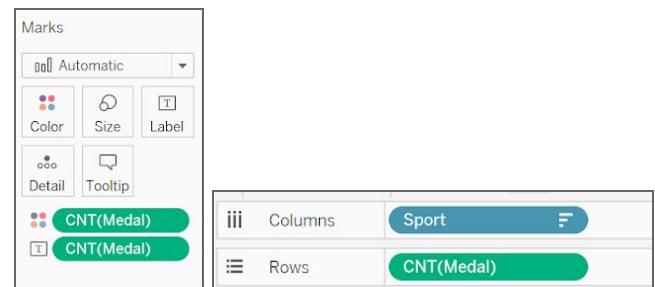


This visualization is a little too crowded for my taste but is fine with the use of the hover and click tool in Tableau. All the other sports seem to have started in 1924 which was the first Winter Olympics. This is just from the data that is included in the dataset so the sports may have started occurring before then. It is not clear if this is just a sample or not. Along with the sports being added there are also the events. The options chosen to get this result are setting the "Sport" to color, the column to year, and the row to count of "Sport". The options chosen to get this result are shown in the next column.

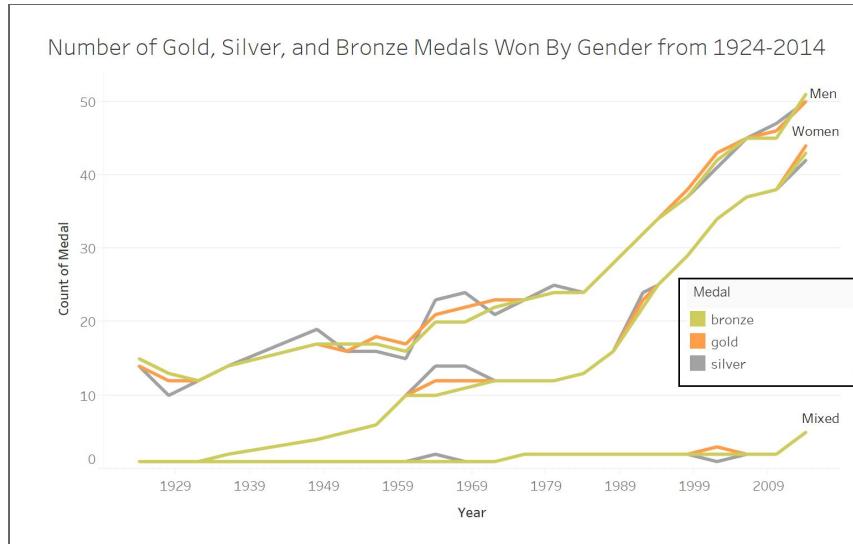
## 4.2 FIGURE 2: TOTAL MEDALS WON PER SPORT OVER ALL YEARS



Speedskating has had the most medals won with 527 overall followed by 471 for Cross-Country Skiing and 428 for Alpine Skiing. The sport with the least number of medals over all years is Skeleton with 30 metals. The color is representative of the number of the number of medals per sport for all years. The dark blue represents more medals won and the color becomes a lighter blue as the medal count goes down. Something to keep in mind also is that some of these sports were added later along with more and more events that are being added each year which might explain why some event categories have much more medals. Take a look at *Figure 1* in *Section 4.1* for more information on the sports and when they were added. The options chosen to produce the results obtained from Tableau was a count of "Medal" for the color and the label. The column filter was "Sport" and the Row filter was set to a count of "Medal". The options chosen to get this visualization are shown to the right.



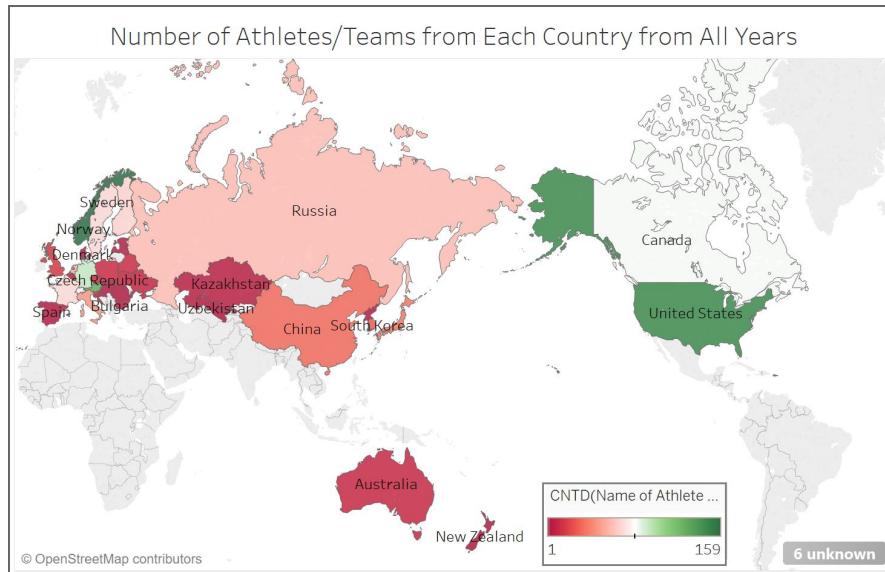
**4.3 FIGURE 3: MEDAL TYPES WON BY GENDER  
OVER ALL YEARS FROM 1924 TO 2014**



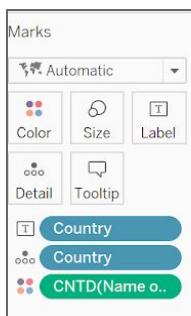
This visualization shows the data in a year by year basis for the different metal types and is separated by gender. The lines are colored by the medal types. The medal counts are getting higher each year. Each major grouping of lines are labeled by the sex of that athlete or team. Men seem to have the most number of metals in 2014 followed closely by women and mixed had the least. Men in 2014 seem to have the most bronze with around 51 medals followed by 50 gold and 50 silver. Women in 2014 seem to have the most gold medals with around 44, 43 for bronze, and 42 silver. Mixed gender had in 2014 5 medals each for gold, silver, and bronze. 1964-1972 seems to be an interesting years in that there is more than typical variation in the medal counts. Options chosen to get this result are shown in the column to the right:

The screenshot shows the configuration of a data visualization tool. The 'Marks' panel on the left includes settings for 'Color', 'Size', 'Label', 'Detail', 'Tooltip', and 'Path'. Under 'Color', two series are defined: 'Medal' (represented by a blue square with a circle) and 'Gender' (represented by a grey square with a circle). The 'Columns' panel on the right shows 'YEAR(Year)' assigned to the 'Columns' role. The 'Rows' panel shows 'CNT(Medal)' assigned to the 'Rows' role.

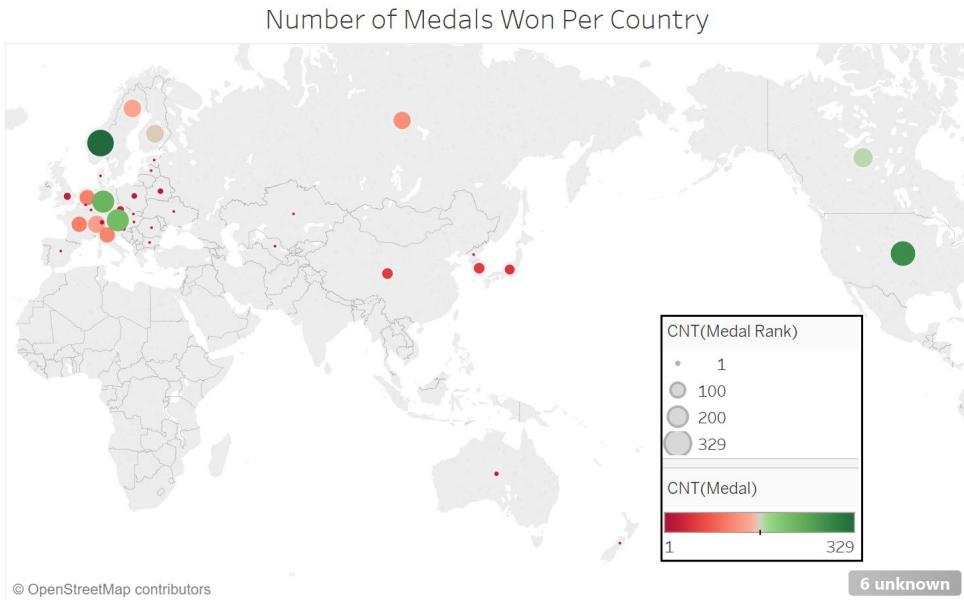
**4.5 FIGURE 4: NUMBER OF ATHLETES AND TEAMS FOR EACH COUNTRY OVER ALL YEARS**



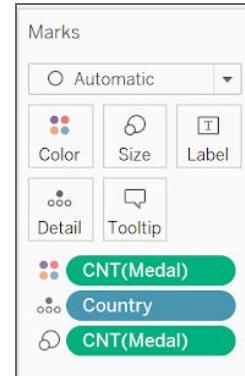
Norway seems to have the most overall athletes and teams over all time in this dataset followed by the United States, Austria, Germany and Canada. There are a number of countries with only one athlete or team. The dark red represents the least and dark green represents the most number of athletes and teams. The column and row filter has only “Longitude” and “Latitude” respectively. The label shows certain countries as well to make it easier to figure out without the hover and select tool. The options chosen to get this result is shown in the image below:



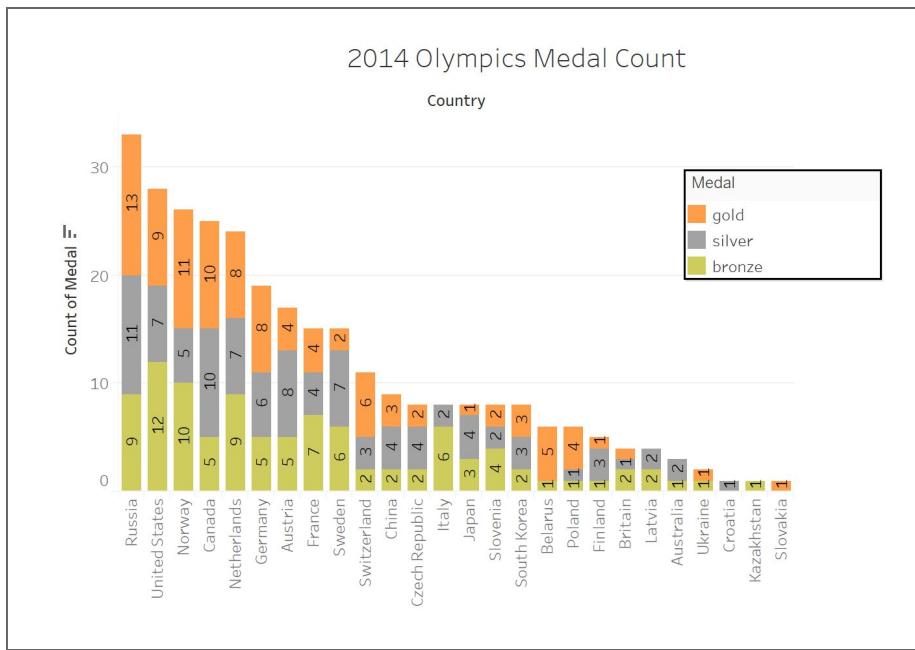
#### 4.6 FIGURE 5: NUMBER OF OVERALL MEDALS WON BY COUNTRY



Norway has the most number of medals (329 total medals) this is probably related to the fact that the most number of athletes and teams in this dataset overall were from Norway as shown in *Figure 4* in *Section 4.5*. The next country with the most number of medals is the United States with around 282 total medals. Next is Canada with a total of 170 metals. The size of the circles in this map are based on the count of the medals. Countries with the least amount of medals are the smallest circles and countries with the most number of medals is the bigger circles. The color is also used along with the size of the circles. The countries with the most medals are in green and the least number of medals are in red. Some countries with the only one medal is Romania, Uzbekistan, and Denmark. The options chosen to produce this visualization were setting the color filter to the count of "Medal", detail to "Country", and the size to the count of "Medal". The options chosen are shown to the right in the image.



**4.7 FIGURE 6: 2014 OLYMPICS MEDAL COUNT**



Russia had the most overall medals for the Winter Olympics in 2014 followed by the United States, Norway, and Canada. The breakdown of the medals are shown by the color for each medal type and labeled with the number of metals. You can see even though the United States in 2014 had the second most number of overall metals, Norway and Canada beats the U.S. in the count of gold medals. Russia has the most silver medals followed by Canada, and Austria. The United States has the most bronze medals followed by Norway and the Netherlands. The options chosen to get this result were setting "Medal" to the color filter, columns to "Country", and Rows to a count of "Medal"

The screenshot shows the Tableau interface with the following configurations:

- Marks:** Set to "Automatic".
- Color:** Set to "Medal".
- Size:** Set to "Automatic".
- Label:** Set to "Automatic".
- Detail:** Set to "Medal".
- Tooltip:** Set to "Automatic".
- Legend:** Titled "Medal" with categories: gold, silver, bronze.
- Columns:** Set to "Country".
- Rows:** Set to "CNT(Medal)".

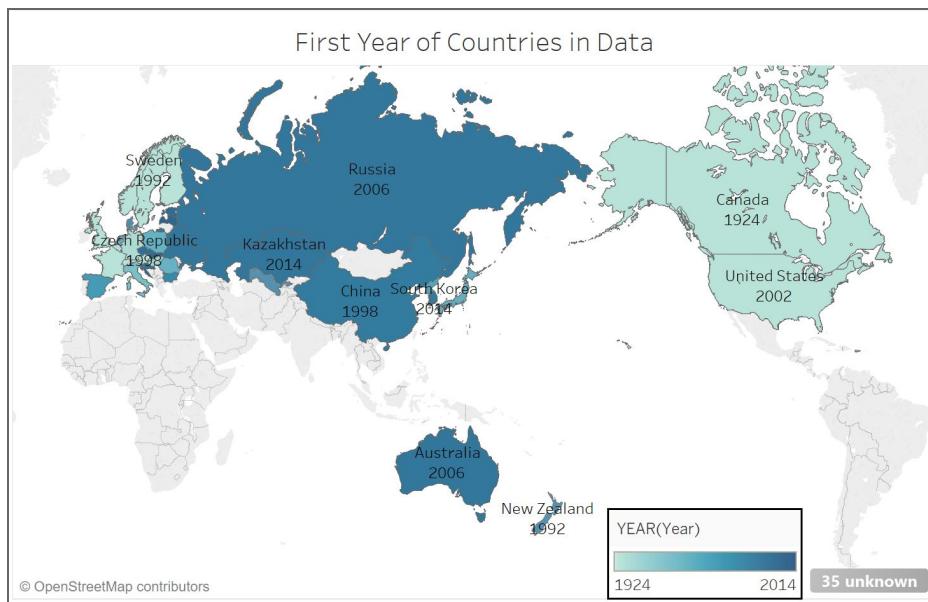
#### 4 CONCLUSION

The Winter Olympics occur every four years and include sports on snow and ice or sports typical to winter climates. Overall, men seem to have the most number of medals in recent years followed by women and then mixed. Speedskating has had the most number of medals followed by cross country skiing and alpine skiing. Norway seems to have had the most number of medals over all years followed by the United States and Canada. In 2014, Russia has the most number of medals overall followed by the United States, and Norway. It is interesting what insights can be gained from a dataset. The other points can be found within the specific visualization section but those are the most important.

## 5 APPENDIX

There are some visualizations that were attempted but were not the best. Some of them are included here because they are interesting and provide more insights into the data.

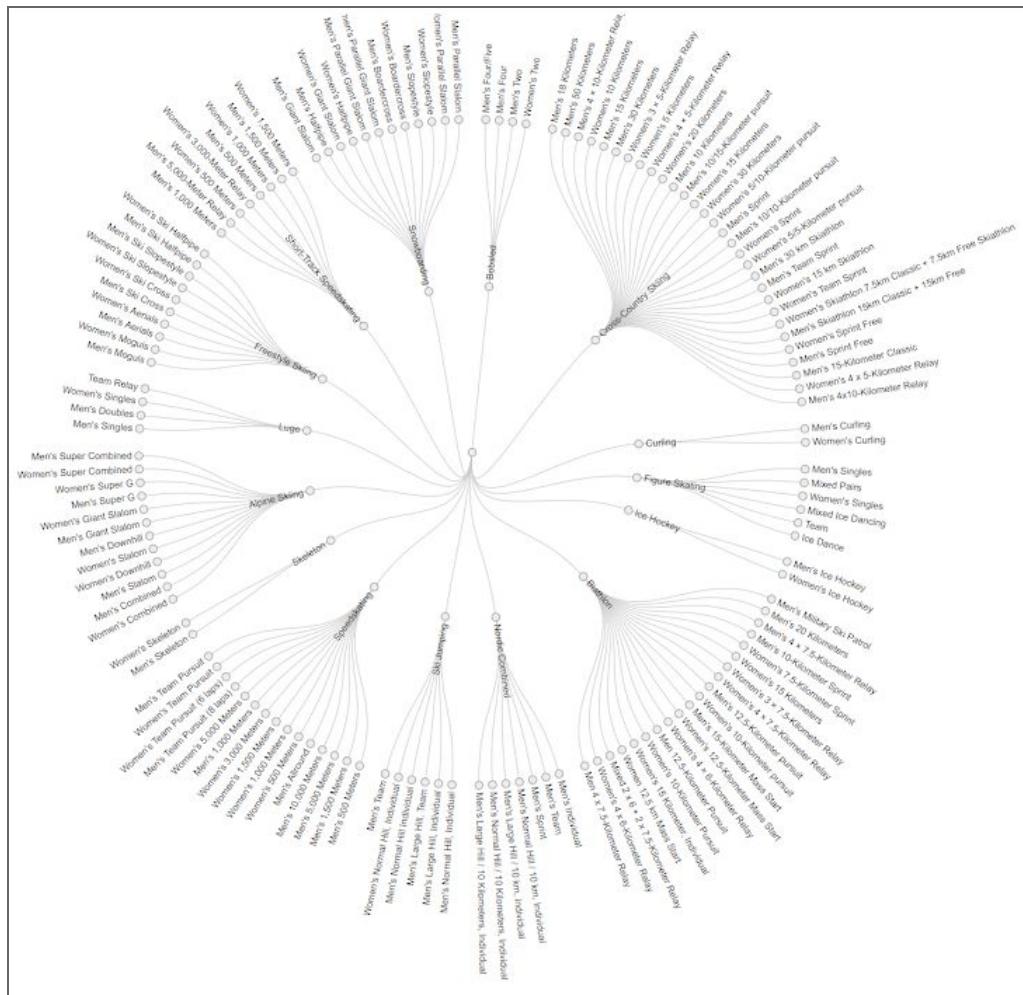
### 5.1 FIGURE 7: FIRST YEAR OF COUNTRIES IN OLYMPICS IN DATASET



This geographic visualization is interesting because you can see the first year a specific country seemingly entered the Olympics or what is just covered in this data. The problem with this is certain countries split or do not exist anymore and are now called something else. An example is Czechoslovakia which is now known as Slovakia and the Czech Republic. It seems that most countries entered the Winter Olympics recently.



## **5.2 FIGURE 8: SPORTS AND EVENTS IN THE WINTER OLYMPICS**



This is an interesting visualization but way too much data is shown and it takes a lot of space. I used RAWGraphs to produce this circular dendrogram. Some changes had to be made in the developer tools for the whole image to show without cutting out. From this you can see that cross country skiing has the most number of different events in this dataset. You can also see that a lot of the events are either specific for men, women, or mixed. The sport skeleton has the least number of event types. The options chosen to get this circular dendrogram was by dragging the "Sport" and "Event" dimension into the Hierarchy filter. An image of the filter and the options are shown to the right.

Map your Dimensions

- Year** number ✖
- Sport** string ✖
- Event** string ✖
- Country** string ✖
- Gender** string ✖
- Medal Rank** number ✖
- Medal** string ✖
- Name of Athlete or Team** string ✖
- Age of Athlete** number ✖

✖ \*

Hierarchy

Drag numbers, strings, dates here

**Sport** string ✖

**Event** string ✖

## REFERENCES

- [1] <https://www.olympic.org/sports>
- [2] <https://data.world/makeovermonday/2018w7-the-winter-olympics>