Analysis of Winter Olympics Data

Arpita Jena

Context

This report is based on Winter Olympics data from 1924 to 2006. The dataset consists of medal history of participating countries and has information on sports category, discipline, year, even type etc. The analysis mainly focuses on various attributes of top countries like Norway, the United States, the Soviet Union, Germany.
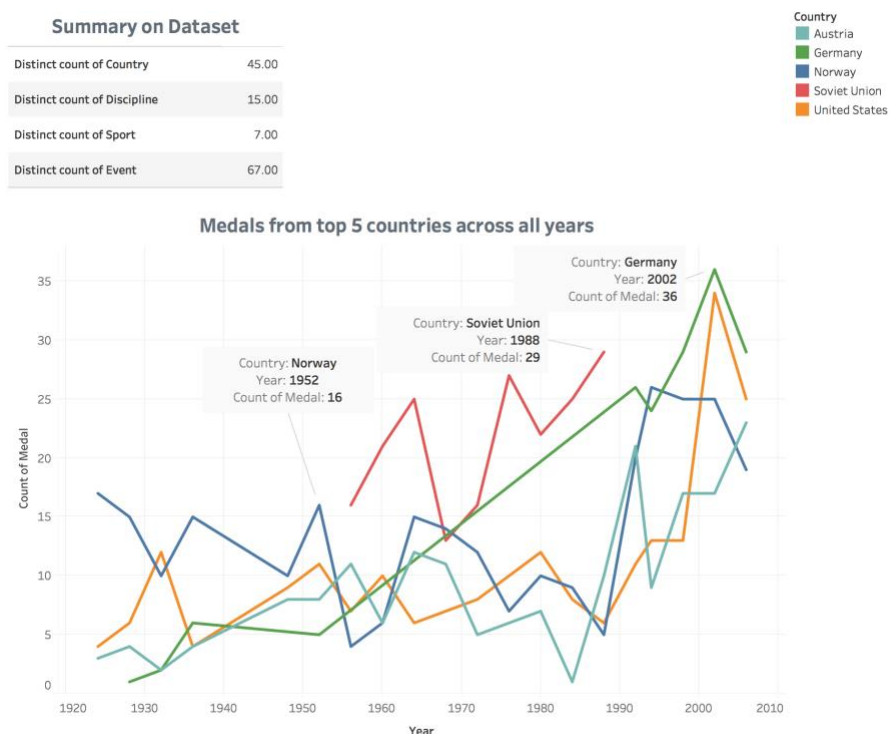
Motivation

As the data has many categorical features, visualizations are clustered around bar plots, line plots, and time series plots.
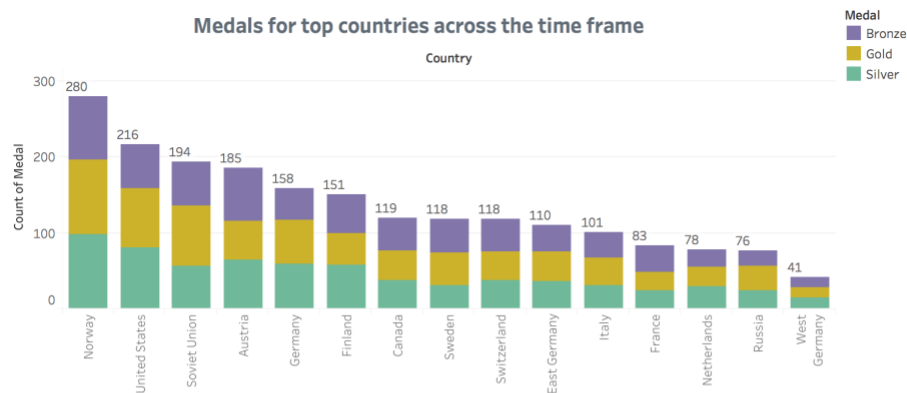
Approach

As the number of categories in each feature is comparatively high, the efficient way for visualization is to consider the top/bottom n features according to the requirements. Here, for each plot, a subset of categories is taken.

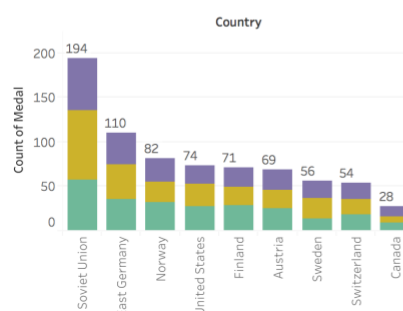Below is a brief summary and performance of top 5 countries across the time frame.

As seen above 45 countries participated during this time period. Till 1952 Norway was the top performing country after which the Soviet Union took its place and maintained its position till 1988. Germany seems to be focused on performance enhancement from the beginning and after the collapse of the Soviet Union, it is at the top of the leaderboard.
Had the dissolution of the Soviet Union not be done, it would have been one of the greatest competitors for other countries.
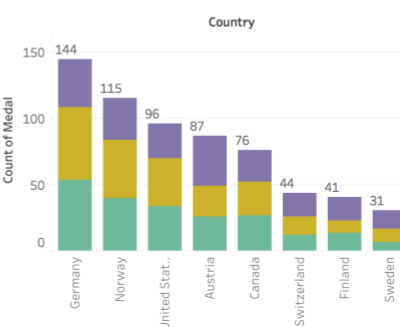
Now a look at the top-performing countries draws the following observation.
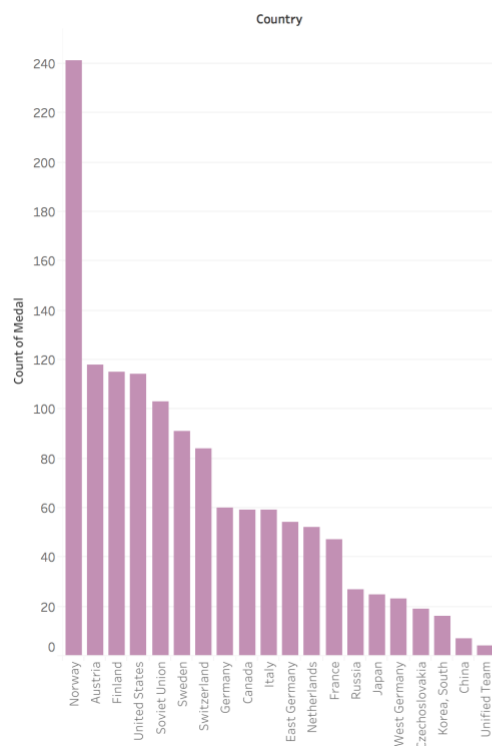
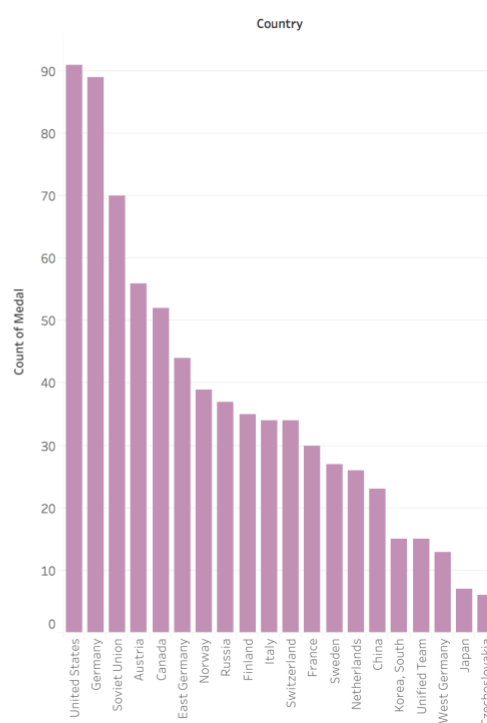## Medals for top countries across the time frame

Norway is the best-performing country in terms of total no of medals in the entire time period. The most interesting observation here is that - though being active for a limited time period the huge number of medals won by the Soviet Union puts it in the 3rd position in the overall time frame.

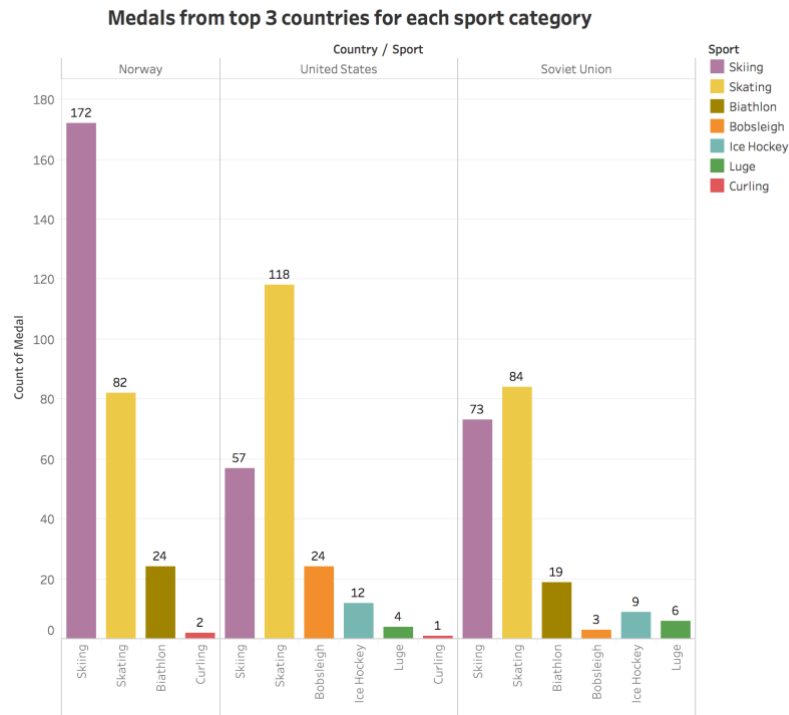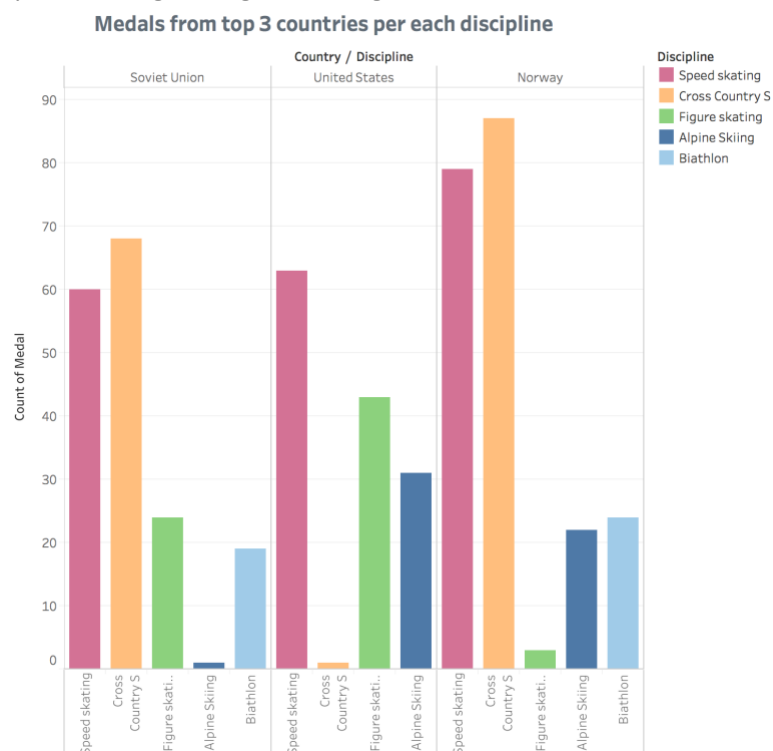In almost all countries medals won by males is relatively very high than that by females, with a few exceptions like China where females have won 3 times as many medals as males. South Korea also has a 50-50 distribution in terms of medal won by each category.

The bar plot below demonstrates that most of the victories of the top 3 countries come from skiing and skating.

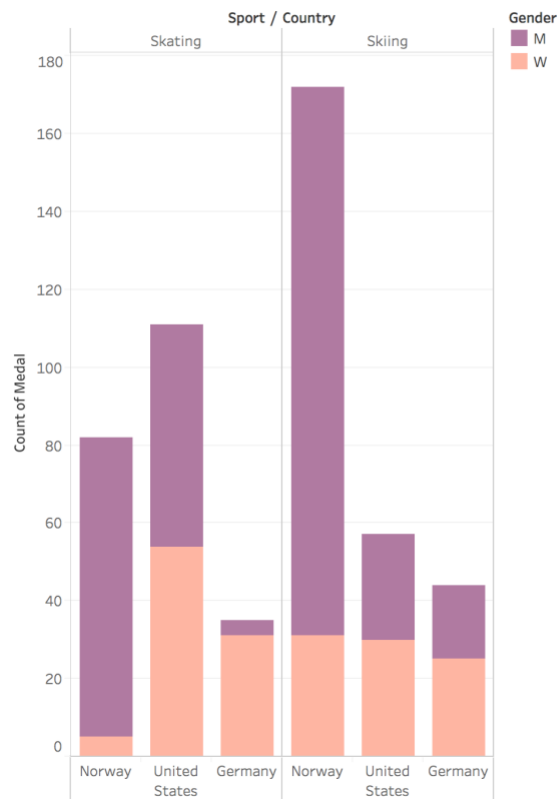**Medals from top 3 countries for each sport category**

The plot below takes into account the top 5 disciplines. 'Speed skating' is a popular sport after 'Cross country S' in Norway and Soviet Union. Whereas the United States has its maximum expertise in speed skating and figure skating.

Medals from top 3 countries per each discipline

Even though the uneven contribution of genders is evident across all countries, the following plot yields a slightly different result in the context of top 2 sports categories.

Surprisingly for Germany, there is more female contribution than male in both the sports. Even for the United States, the contribution of females in skiing exceeds by 3 than males. Whereas Norway has a very uneven distribution across gender.

Take Home Message
Except from finding some interesting outcomes here are some learnings from the dataset.

- It is recommended to perform selective plotting in case of categorical variables with high cardinality to avoid crowding of information.
- Attention should be paid to make right choice of color and plot type to effectively communicate the findings.