CS3300 Project 1: Road to No. 100 Winter Olympics Gold Medals

Chen Li (cl953)
Yanjun Chen (yc2392)
Zhengnan Zhao (zz528)

A. A description of the data

In honor of Team USA' 100th Winter Olympics gold medal, this project serves as a historical review of American athletes' road to this milestone. What we are emphasizing here are all the winners of the 100 medals, so for team competitions, we decided not to count the winners of the same gold medal as 1, as we hope to show individual's effort. Therefore, the total gold medal count in our visualization exceeds 100, which might be a little bit confusing -- that's why we added a paragraph on our visualization to clarify this intention.

We first obtained the original dataset *winter.csv* from **Kaggle**[1], which was collected from **IOC Research and Reference Service.** The variables that particularly interest us are:

- Year (1924-2014): the year in which each Winter Olympics took place
- City: Winter Olympics host city
- Discipline: a branch of sports that comprising one or more events
- Country: a medal winner's country
- Gender: a medal winner's gender
- Medal: the class of a medal awarded (gold, silver, or bronze)

We further filtered the dataset by restricting Country = "US" & Medal= "Gold". Following that, we collected the latest information about Team USA's gold medals at PyeongChang 2018 Winter Olympics from **the Olympic Studies Center**[2]**,** and appended the new data to the filtered data. We then made a search of every host city's latitude and longitude coordinate, and merged this location dataset with the whole dataset. In addition, we utilized the world-50m.json data file from one of the d3-geo projections[3] to map data points onto a world map. Finally, we collected icons for all the Winter Olympic disciplines for later usage.

By sub-setting and merging data sources, we were able to focus on exploring the following aspects:

1) Team USA's gold medal winning trend over the years, breaking down by genders;
2) Gold medals count by disciplines;
3) A world map projection of host cities.

---

[1]Data was provided by the IOC Research and Reference Service and published by The Guardian's datablog. "Olympics Sports and Medals." https://www.kaggle.com/the-guardian/olympic-games/data

[2] Olympic Studies Center. "PyeongChang 2018."https://www.olympic.org/pyeongchang-2018/results/en/general/competition-schedule.htm

[3]“d3-geo projections” https://github.com/d3/d3-geo/blob/master/test/data/world-50m.json

It should be noted that some of the disciplines involve multiple participants, and therefore, award more than one gold medals for a winning team. This leads to the fact that the total gold medal count in our visualization exceeds 100.

B. A description of the mapping from data to visual elements

We divide the canvas into three sections, a bar chart showing the gold medals won by the U.S. athletes counted by each Winter Olympics, a world map showing the host city of each Winter Olympics, and a bubble chart showing the gold medals won by the U.S. athletes counted by each discipline. We connect the bar chart and the world map with host cities to give our audience a better sensemaking experience of the 100 gol]d medals mapped across time and geography.

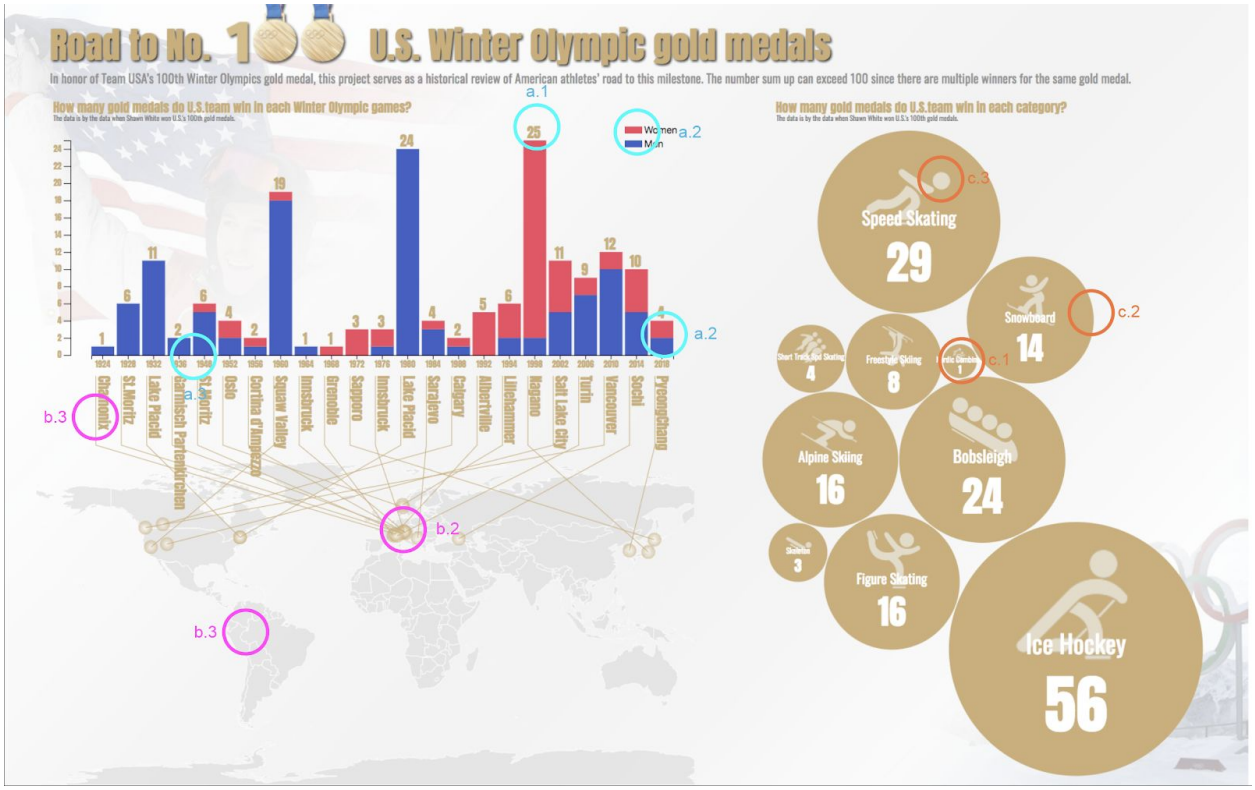

Fig.1: the final layout

Bar chart

We choose to use a bar chart to show Team USA's gold medals winning trend over the years, breaking down by genders. The design highlights are as below:

1. We set up SVG coordinates and map the amount of gold medals Team USA won at each Winter Olympic Games to the SVG coordinates. We use the height of the bar to represent the amount of the gold medals. We then show the count on the top of each bar to give a more direct sense of amount. (see fig.1 a.1)
2. We divide the bar by genders and use different colors(red v.s. blue) to show the difference. Also, we add legends on the top-right corner to clarify this difference. The reason of doing this is to show that U.S. women athletes are becoming more and more involved and achieved. (see fig.1 a.2)
3. We find that there exist a few special historical occurrences in the timeline of the Winter Olympic Games. For example, the Games were interrupted in 1940 and 1944 by World War II, and the next Winter Olympics after 1992 was in 1994, instead of 1996. If we were to use the natural ordering of the years to place the bars, the chart would look weird by having multiple gaps. Therefore, we write a function to create an additional reference: the ordering of the the Games, we then use this reference to place bars along the x axis. Thus, bars can be placed evenly and beautifully in this way.(see fig.1 a.3)

<u>World map</u>

Below the bar chart is a world map where we show the location of host cities of the Games. The design highlights are as below:

1. Firstly, we use Eisenlohr conformal projection method (d3.geoEquirectangular) to map the coordinates in JSON geographic file into SVG to set up the world map. Each city's coordinates (longitude and latitude) are then retrieved from our csv file to be mapped on this world map. Also, by employing this projection method, we use circles to represent these cities' positions.
2. We design the map in light grey fill color and white stroke color to make it fade into the background. All the other elements, which should be more discoverable, are designed with the theme color: gold.(see fig.1 b.1)
3. One of the biggest challenge for us is to map the circles on the map and try to leverage each circles' discoverability. As many events were held at European cities, the circles representing those events are overlapped together. Here we design the circle with a radial gradient: transparent in the center and hard border with high opacity. The layout improves greatly.(see fig.1 b.2)
4. We draw paths to connect the circles representing cities to the bar chart to combine the world map part with the bar chart part as an overall review of Team USA's gold medals in time and space dimension. (see fig.1 b.3)

Bubble chart

We further plot a bubble chart on the right hand side of the visualization. We choose a bubble chart because it is rather self-explanatory and thus balances well with the above two relatively complex graphs. The design highlights are as below:

1. We set the value of maximum dimension of the bubble areas so that text elements will be in harmony with different sizes of bubbles. However, there are some bubbles with very small text elements. To reconcile this, we add a Math.max() to set up a minimum font size.(see fig.1 c.1)
2. We convert the numeric value of medal count to bubble-chart specific format.(see fig.1 c.2)
3. We create circles to represent bubbles based on the medal count for each discipline. We then add icons of the sports over the circles to make it easier for audience to recognize the disciplines. In addition, we match the text for each bubble with disciplines.(see fig.1 c.3)

C. The story

Our visualization illustrates the gold medals based on various dimensions and angles, yet tries to present a coherent data story. Through winners' gender comparison, medal winning trend (for over 94 years) presentation, and medal count by disciplines exploration, this visualization provides several insights.

First, women athletes are closing their gender gap when it comes to the gold medal winning – no women athlete had won a gold medal for over a decade from 1924 to 1936. However, we can observe a fairly steady trend towards more women athletes' achievement over time, with females evidently outperforming their male teammates in Nagano, Japan in 1998. This might be due to increased funding and institutional opportunities (e.g. increased number of women's events) for female athletes. Our visualization aligns well with the fact that female athletes are approaching equal participation rate as male athletes.[4] Second, Team USA has been excelling at disciplines including ice hockey, speed skating, bobsleigh, alpine skiing, figure skating and snowboarding. Cities such as Lake Placid, Squaw Valley, and Salt Lake City fortunately witnessed some of the best moments when Team USA added gold medals to its collection – potentially resulted from the so called "home advantage" effect.

In short, our visualization is an exploration of the overall pattern of Team USA's gold medal winning path, breaking down by athletes' genders and different disciplines. Future visualizations would benefit from presenting total (gold, silver and bronze) medal count in both summer and winter games, and focusing more on the trend of female and male athletes' participation rather than achievement over time.

[4] Olympic Studies Center. “Women at the Olympic Games” https://www.olympic.org/women-in-sport/background/statistics