WINTER OLYMPICS

Process Book

Group: Bugger
Liu Haowen, Sun Haoxin, Wu Yujie

Spring Semester 2022

1 INTRODUCTION

The XXIV Olympic Winter Games in Beijing closed early this year, in which lots of brilliant athletes contributed their wonderful performances to the Game. However, not many people know details about them, and elite athletes in past Games easily fade from people's memory. There are few websites to sort out those pieces of information and display them intuitively. Although there are statistics of medals from different countries online, it lacks visualization of detailed discipline information of each country. Therefore, we design a website to record the information and wonderful moments of elite athletes in all Winter Olympic Games visually. Besides, we visualize each country's entries based on the medals awarded. In this way, people can have a better understanding of winter sports and this global event.

To do this, we design a homepage and two detailed information pages to visually display our data in two different aspects: athletes and countries. In general, the homepage is designed to show basic information about Winter Olympic Games and winter sports, so that users can get to know the Game before diving into details. Then the first information page implements an interactive world map and represents every athlete by a dot lying in his/her homeland, which visualizes athletes' data. By applying different filters like years they won medals and events, users can freely explore athletes' information that they are interested in. The last page is used to provide some intuitive insights by showing statistical graphs, which summarize different countries' competition information.

In this process book, we present the dataset we use in our design, then demonstrate how we develop our data visualization websites, challenges we meet during this process and the final version of our design.

2 DATASET

We explore the dataset *Olympic Games, 1986-2021* [1] combined with the dataset *Beijing 2022 Olympic Winter Games* [2] to visually present the data of excellent athletes in all Winter Olympics since 1924. Both of them are derived from *Kaggle*.

2.1 Description

The first dataset contains the data of Olympics from Athens 1896 Olympics to the latest Tokyo 2020 Summer Olympics. Containing more than $150,000$ pieces of data, it consists of four tables, including information about athletes, host time and city, medals, and detailed results of every Game.

To keep data up to date, we complement the above dataset with the second one, which contains the data about medals and athletes in Beijing 2022. It consists of ten tables, but only three tables — Athletes, Events and Medals — matter in our project.

2.2 Data preprocessing

Before applying these two datasets in our website design, we first do data preprocessing via Python, to clean the data and extract useful information.

As this project is dedicated to displaying information about excellent athletes in Winter

Olympics, we first figure out and merge useful parts of two datasets. We take medalists as the criterion to decide whether athletes are elite or not, so we focus on the data of medals, remove the data of athletes who got no medal, and use other tables to complement the information of medalists. Besides, we also do data cleaning and processing in the following aspects.

- **Winter and Summer.** The first dataset contains the data of both Summer and Winter Games, so we only reserve data for the Winter Olympic Games.

- **Name Format.** The name format of athletes is not unified, i.e., the first and last names are in different orders in these datasets, which may be confusing to users. Thus, we first figure out their first and last names separately, then unify the format.

- **Country Name Change or Reconstruction.** Some countries experience name changes or reconstruction from 1924 to 2022. For example, one of the participating countries in 1956 is the Soviet Union. In 1968, German Democratic Republic joined in Winter Olympic Games. Some countries appear in the dataset with different names such as China which sometimes is also called People's Republic of China. So, we unify country names before proceeding with the statistics.

2.3 Data analysis

After doing preprocessing and getting clean data, we do some analysis by Python to figure out some statistical features so that we can have a better understanding of our data and design our websites based on these features.

The number of medals each committee won determines the density of dots representing athletes on the world map, so we carefully look at this feature. We find that most of these athletes are from Europe, especially northern and central Europe countries like Norway, Germany, and Finland. Thus, users should be able to zoom in/out on the interactive map in our design, otherwise they cannot figure out specific athletes in these regions.

In the temporal dimension, it is clearly shown that the number of medals has increased over time, especially from the 1920s to the 2000s. Categorical characteristics can be noticed by counting the total number of medals in each discipline and event, and we find that these numbers vary a lot among different events. For example, Speed Skating has 545 medals while Short Track only has 18 medals. These features are then used as our guidelines when designing the statistical page and the filters in the interactive world map.

It can also be observed that the number of male athletes outnumbers that of female athletes. After analysis of the processed dataset, the majority of countries have more male participants than female participants.

3 VISUALIZATION DESIGN

To better display the data to the audience, multiple visualization methods are considered and implemented in our projects. In this section, we demonstrate the path we take to reach the final goal of our project.

3.1 Design journey

Firstly, before developing the source code for our project, we design drafts for these three pages by hand as shown in Fig. 1 and Fig. 2, which act as guidelines for later software development.

The homepage is designed to introduce some basic information about Winter Olympics and winter sports. To be more specific, we aim to display past Winter Olympic Games and sports events in these Games on this page. The overall idea of the design is to achieve ease of use, and we facilitate users' experience by visualizing the data concisely and intuitively. A navigation bar on the top would allow users to access other pages conveniently. Since these contents are text-heavy, displaying them all at once would make the page too difficult to read. Based on the principle of "overview first and details on demand", our solution is to apply a vertical timeline to briefly show each past Games, and if users are interested in any specific Game, they can explore details (which may contain some texts) by clicking on the timeline. The same solution goes for the winter sports part. A table containing icons and names would be displayed, while corresponding and detailed information about events is available when clicking on icons.

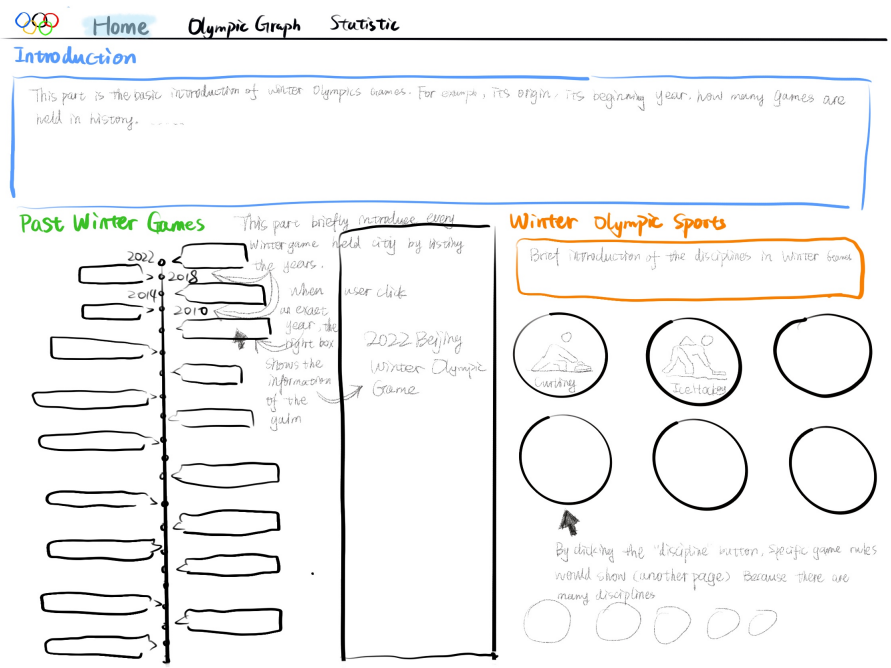


Figure 1: Design draft for the homepage.

According to [3], Geographic Information System is the most effective for problem identification and brainstorming. Therefore, to visualize the athletes' data, we decide to utilize an interactive map to facilitate users' interpretation. Our preliminary idea is intuitive: athletes are represented by dots located at their homelands on the map. Because there are too many data points in the dataset and exploratory analysis tells us that athletes lie denser in Europe and North America, it is impossible to display all these data points on the map all at once. Applying "overview first and details on demand" again, we plan to limit the number of data points shown on the map at one time and allow users to zoom in/out to see the entire dataset. Then the detailed information about each athlete, like his/her nationality, when he/she won a medal in what sports event and so on, would pop up when users click on these points on the map.

Besides, some users may only want to see athletes with specific features. For instance, some skiing enthusiasts just hope to get more knowledge about athletes in skiing and don't want to be bothered by information about athletes in other fields. So we plan to have filters for athletes' information on the top of this page in two dimensions: years they won medals and events they participate in. By doing this, we have different ways for visitors to use this website. Users without interest in certain athletes can just roam around the map and click randomly on the icons. In this case, they have the indications of regions of athletes. Users who want to know the information about a specific event or Game can apply the filters on the map, so that only desired athletes' information is shown on the screen.

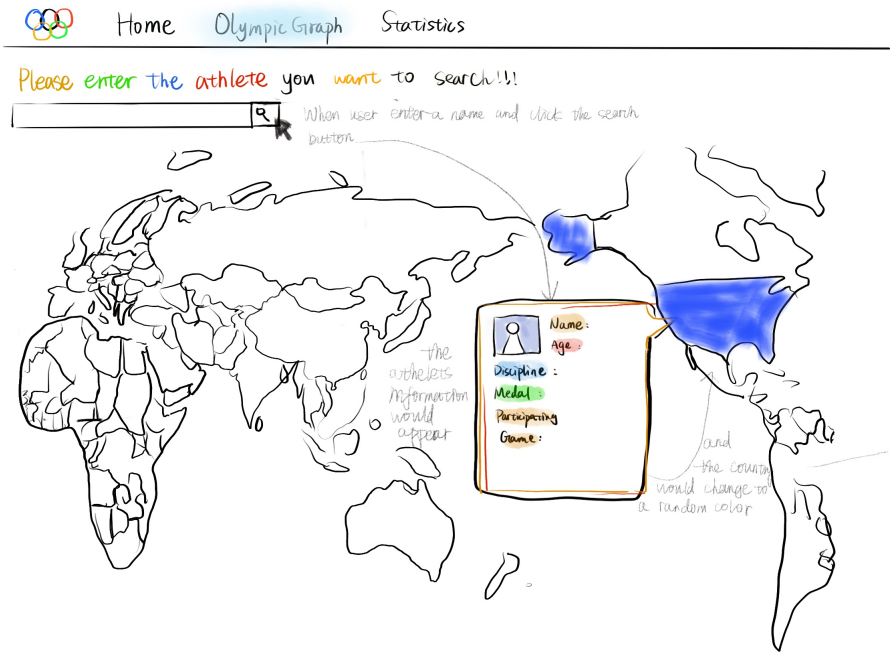


Figure 2: Design draft for the interactive map.

Apart from individual athletes' information which is visualized by an interactive map on the second page, the country is another popular topic and essential element in the Winter Olympic Games. We design the third page to intuitively summarize the statistical characteristics of countries in the dataset, displaying some figures based on the preprocessing results of this dataset. We reckon that some users may also want to compare countries' strengths with specific features. Thus, it is meaningful to visualize statistical characteristics of medals won by different countries and awarded by different disciplines. In our design, bar charts and radar figures are implemented to visualize the information. These figures are also interactive so that by hovering or clicking, users can get corresponding detailed information.

This page would facilitate people with orientations of a particular discipline and/or a country. For instance, people who want to develop their career in the winter games would be delighted to search on this page. They can explore the country's name to see if there are good resources for their target sports in their own countries.

Secondly, after having the above design ideas and drafts, we start to develop the source code for our project. In general, we use HTML and CSS to design the framework of these websites,

while the interactive parts are developed via JavaScript. Specifically, apart from basic JavaScript functions, we use the *D3* and *Leaflet* library to develop our interactive map and related filters. Bar charts reflecting statistical features on the third page are drawn via the *Chart* library, and some user interactions on our pages are based on the *jQuery* library.

Finally, after the development and verification on local hosts, we push and host our project on *github.io*, so that users can directly access our website through the URL rather than having to download all source code and supporting materials.

3.2 Challenges

It is not surprising that even though we have a quite comprehensive design before development, we still meet some problems and challenges during our development. This sector shows how we tackle these challenges and make modifications to our original ideas.

3.2.1 Text-heavy contents

In our previous plan, we design a sector named "introduction" to display some information about the Winter Olympics, like values of Olympism, unity and diversity spirits. However, this requires large text paragraphs, which are not suitable for visualization. Thus, we change our original design and use scrolling images with keywords to display this information, instead of using pure text introductions.

3.2.2 Athletes' homeland

As shown in the previous sector, we display athletes' data points on the world map based on their homeland. However, athletes from the same countries locate in the same places and thus cannot be distinguished on the map in this scheme. For example, dots representing athletes from France would all lie in the center point of France, making our visualization unclear and not user-friendly. To solve this problem, we add a random number to each dot's coordinates, so that they can separate from each other, resulting in better visualization.

3.2.3 Athletes' images

At first we only plan to use dots to represent athletes on the map, but we find that this is not vivid. Thus, we change our idea to use athletes' images on the map. However, the dataset doesn't contain these contents, and we have to collect them by ourselves. We use the library *Beautiful Soup* in Python to pull the image URL of each athlete from *Wikipedia*, complementing our original dataset. For athletes whose *Wiki* don't contain images, we use their Olympic icon as the profile images on the map.

3.3 Final product

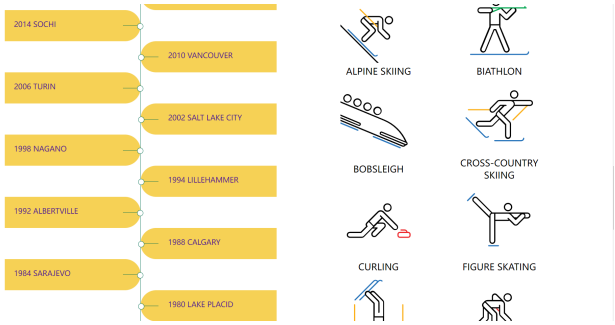
After solving the above-mentioned challenges, we derive the final edition of our project. In general, we have three pages to display the introduction, athletes' data and statistical characteristics of the data respectively.

When users open our website, we will first introduce the Winter Olympics to them, as shown in Fig. 3. On the top of this homepage, a navigation bar helps the audience to redirect to the two pages presenting the dataset. Then below the cover is a list of scrolling images about the Winter

Olympics, which provides information about the Game itself, like values and spirits of the Game. Next, a vertical timeline on the left side of this page summarizes the host city and time of all the past Games in chronological order, and on the right side, a table containing icons and names of winter sports disciplines is displayed. If users are interested in some specific Games, then they can browse more information about these Games by clicking on the corresponding leaves in the timeline, and they will see images and texts like Fig. 3(d), which introduce the Games in detail. The same design also goes for the table of winter sports events, where users can get to know more about specific disciplines by clicking on these icons.

(a) Design for the cover.

(b) Design for the Introduction sector.



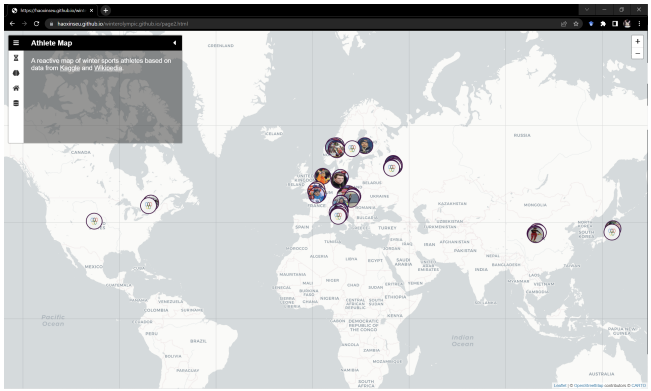
(c) Design for the Past Games and Events sector.

(d) Design for the detailed information.

Figure 3: Final design of the homepage.

By clicking "Interactive Map" on the navigation bar, users can jump to the "Athlete Map" page and see a reactive map dotted with markers of brilliant athletes, as shown in Fig 4(a). Users can zoom on the map with the mouse wheel. The density of markers on the map reveals the strength of winter sports in each area. By clicking a marker, users can see the brief introduction of the corresponding athlete, including the name, discipline, homeland, and which medal he/she won in which game, as shown in Fig 4(b).

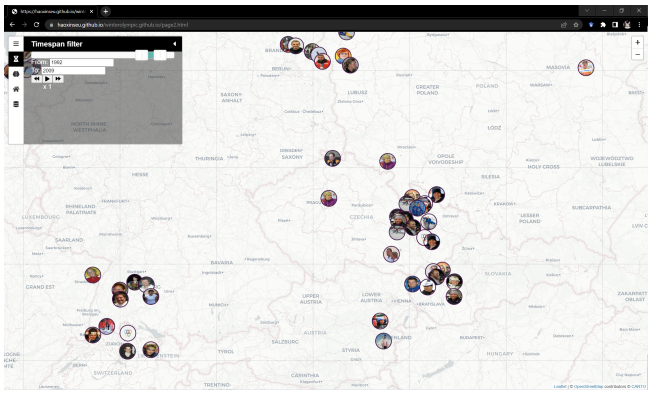
When turning sight to the sidebar and clicking the second icon, users can see a timespan filter, which can filter athletes by a specific time interval. Fig 5(a) shows that by inputting or sliding the timespan, users can choose to only show athletes in a certain period. In addition, after choosing an interval, users can click on the forward icon downside to auto-slide the time interval to see the trends of winter sports strength in each area. When clicking on the third icon of the sidebar, users can see a discipline filter, in which we can filter athletes by disciplines of winter sports by choosing certain disciplines, as shown in Fig 5(b). By clicking the fourth and fifth icons users can jump to "homepage" and "Analysis" page respectively.



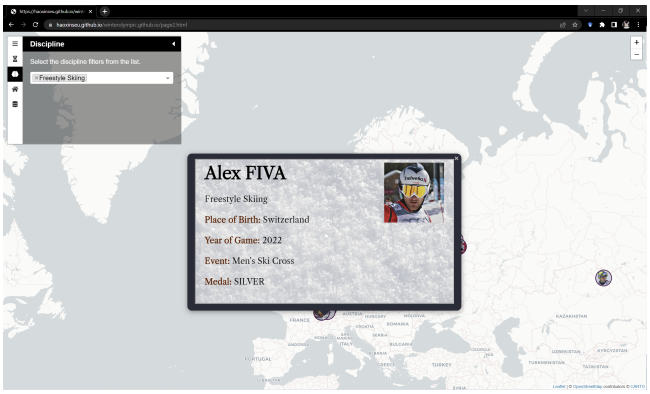
(a) Design for athlete map.

(b) Design for the athlete profile.

Figure 4: Final design of the Athlete Map page (1).

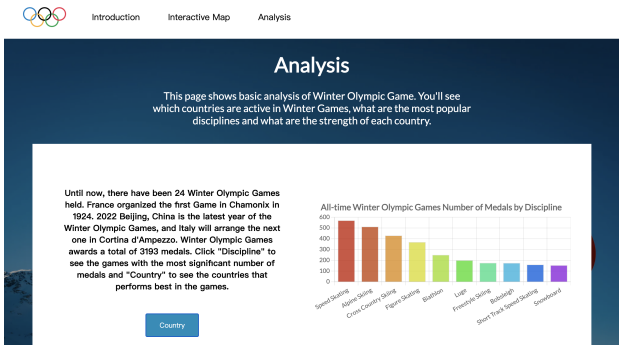


(a) Design for the timespan filter.

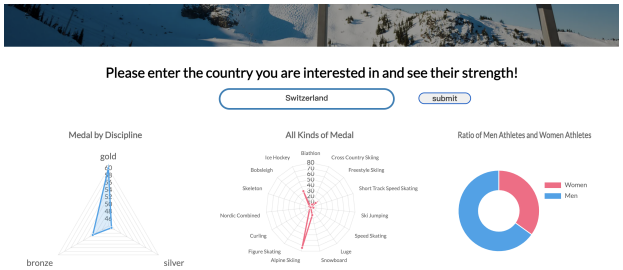


(b) Design for the discipline filter.

Figure 5: Final design of the Athlete Map page (2).



(a) Design for Analysis Page.



(b) Design for Country Search.

Figure 6: Final design of the Analysis page.

When users come to the "Analysis" page, they first see a brief introduction to this page. Scrolling down the website, there is an interactive button for users to change the displaying bar chart. If they click "Discipline", the number of medals awarded according to disciplines is shown while "Country" leads to a bar chart considering different countries. We set the button on this page because we want to reduce the cognitive load of users by showing only one graph on one page. In this case, they would be less stressed about to intake of the information on the dataset. These are followed by a search bar where users can input their interested countries. Once they click on the submit button, three statistical graphs would show if the country won a prize in the previous games. The first two are radar graphs showing the country's number of medals and the number of medals in a discipline. The final doughnut graph illustrates the ratio of female and male players of the country. The three graphs show the information in a visual and vivid way. The screenshot of the last page is shown in Fig. 6.

4 PEER ASSESSMENT

In general, all members in our team have roughly the same amount of work and are indispensable in our teamwork. We reached consent for the working loads for every milestone in this project. We explored how to implement these websites, and figured out solutions to each part all together before everyone began to work on software development, i.e., the most difficult parts of this project are finished via our discussion. Besides the discussion, the main contributions of each member are shown as follows.

- *Sun Haoxin:* Develop Python notebooks to do pre-processing and analysis on the original dataset. Pull athletes' image URLs from *Wikipedia*. Construct the homepage. Write the major part of this process book.
- *Liu Haowen:* Further process and analyze the dataset to discard invalid data and allocate athletes on the map. Construct the "Athlete Map" page with D3.js data viz based on processed dataset.
- *Wu Yujie:* Analyze the processed dataset and obtain the detailed information based on country. Design overall style of the website. Construct the "Analysis" page with the processed "country" information.

References

[1] Petro. Olympic games, 1986-2021 — results, medals, athletes and hosts from athens to tokyo, 2022. [Online; accessed 24-March-2022].

[2] Petro. Beijing 2022 olympic winter games — full data about medals and athletes. ice hockey and curling results, 2022. [Online; accessed 24-March-2022].

[3] Kheir Al-Kodmany. Using visualization techniques for enhancing public participation in planning and design: process, implementation, and evaluation. *Landscape and urban planning*, 45(1):37–45, 1999.