



OLYMPICS STATS

PROCESS BOOK

CS 6630 - DATA VISUALIZATION

CONTENTS

- Basic Information
- Overview and Motivation
- Related work
- Questions
- Data
- Exploratory Data Analysis
- Design Evolution
- Implementation
- Evaluation and Conclusion

Basic Information

PROJECT TITLE: Olympics Stats

TEAM MEMBERS:

- Krunal Jain (u1140629) – krunal.bitsy@gmail.com
- Abhishek Lakshminarayana (u1145084) – abhishekl9426@gmail.com
- Vinod Reddy Gooty (u1142935) – gootyvinod12@gmail.com

REPOSITORY: https://github.com/krunaljain/dataviscourse-olympics_stats

WEBSITE: https://krunaljain.github.io/dataviscourse-olympics_stats/

Overview and Motivation

On April 6, 1896, the first modern Olympic Games was held in Athens, Greece, with athletes from 14 countries participating. From then the places where Olympics takes place started to change. Later, every year the number of countries participating increased. From year 1900 women started to participate in the Olympics.

The Olympics are leading international sporting events featuring summer and winter sports competitions in which thousands of athletes from around the world participate in a variety of competitions. The Olympic Games are considered the world's foremost sports competition with more than 200 nations participating. The Olympic Games are held every four years, with the Summer and Winter Games alternating by occurring every four years but two years apart.

All countries vouch for success at Olympic games, which are held once every 4 years. The performance of various countries over the years has shown some interesting trends. The aim of our project is to capture and visualize this trend and provide an insight into how different regions across the world fared in these games.

The motivation for us to consider visualizing Olympics stats is that how modern Olympics has changed over years along with several factors like men/women ration, world wars, GDP of participating countries. And coming from India, we always wonder why India doesn't perform well in Olympics. While we play with the data, we shall try to analyze why some countries have never made an impact in Olympics.

In this project we are going to make visualizations by which we will be able to see a trend in Olympics over the years. This will include medals won by countries, country ranking, men/women ratio, host nation performance. Overall, we can infer from the designs what the trend is and will be worth reckoning to know in detail. Our goal is to make visualizations synchronized with each other so that it's easy to deduce the learning from it.

Related Work

Inspiration for this project came from personal interest in Olympics sports. Throughout the course, we studied different visualization techniques and their applications in specific use cases. Through this project, we wish to incorporate these visualizations to highlight the Olympic performances of different countries over the years.

Questions

The visualization project tries to analyze certain statistics of Olympics and will help us giving details to the people who are very keen to know what's the trend. The main questions we want to answer with this project include the following:

- How European countries dominated at the beginning?
- How host nations perform better compared to previous years?
- Did world wars affect Olympics? How?
- Does medal count for sports vary over years?
- Men/women ratio for country over years?

Data

We are using one dataset. Here is the link to the dataset <https://www.kaggle.com/the-guardian/olympic-games> . This dataset has 3 CSV files, dictionary.csv, summer.csv, winter.csv. Since we are concentrating only on summer Olympics in this Project, we skip winter.csv file.

Dictionary file has list of country name and respective country code, population and GDP per Capita. Summer file has year column in ascending order of Olympics and respectively where it took place, different sport and its classification and 1st/2nd/3rd winner details in each classification.

We have two kind of views in our project:

- Basic Views: For these views, we can directly use the data available in the CSV file and do projections to filter out columns depending on our requirement.
- Aggregated views: For these types of views, the data is aggregated based on some parameter. So, we need to do data processing.

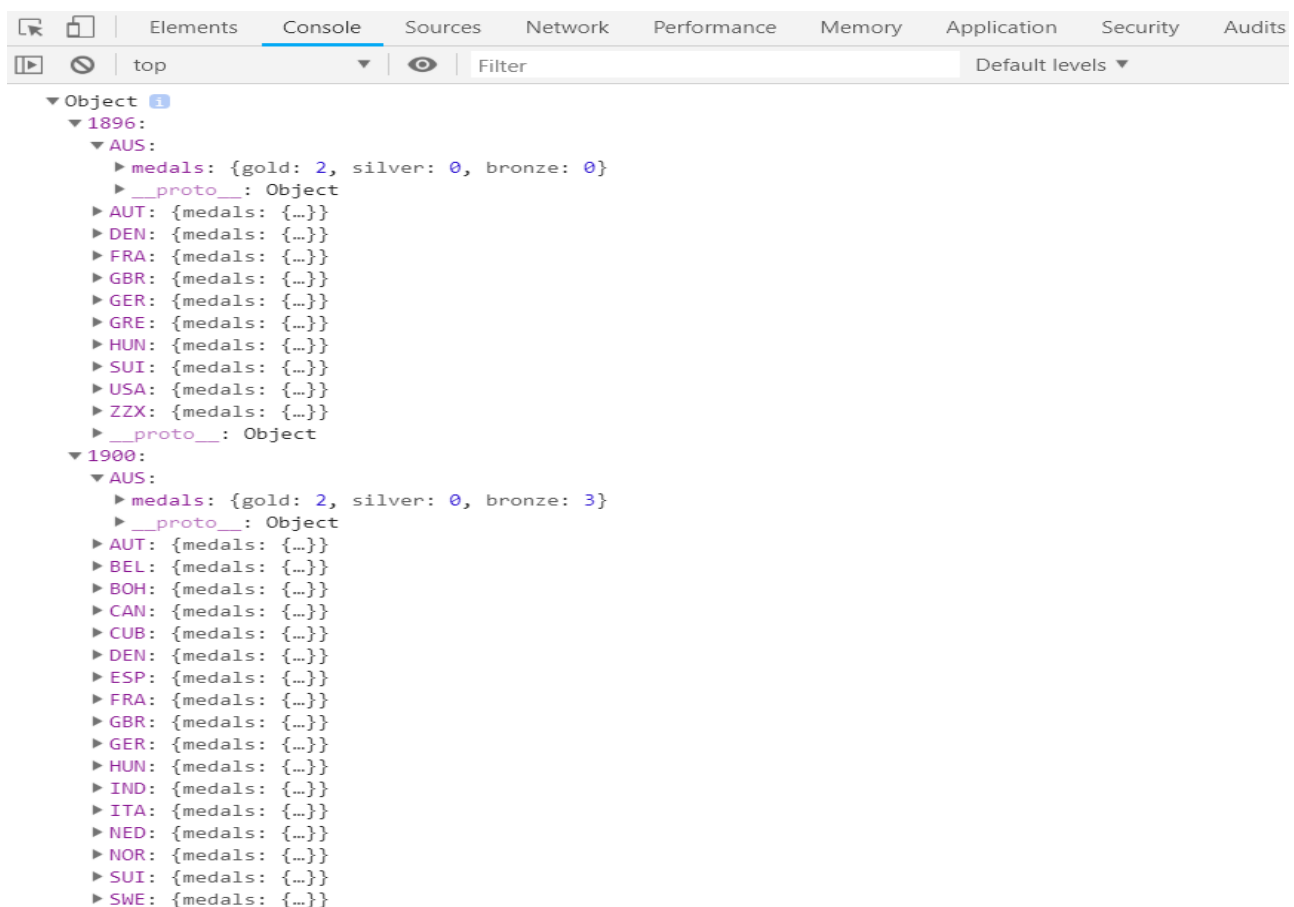
Exploratory Data Analysis

The schema of the dataset is as explained above. For each view, we had to aggregate and group data according to the desired parameters. The following are the different transformations we have done on the data depending on the requirements of the view. In

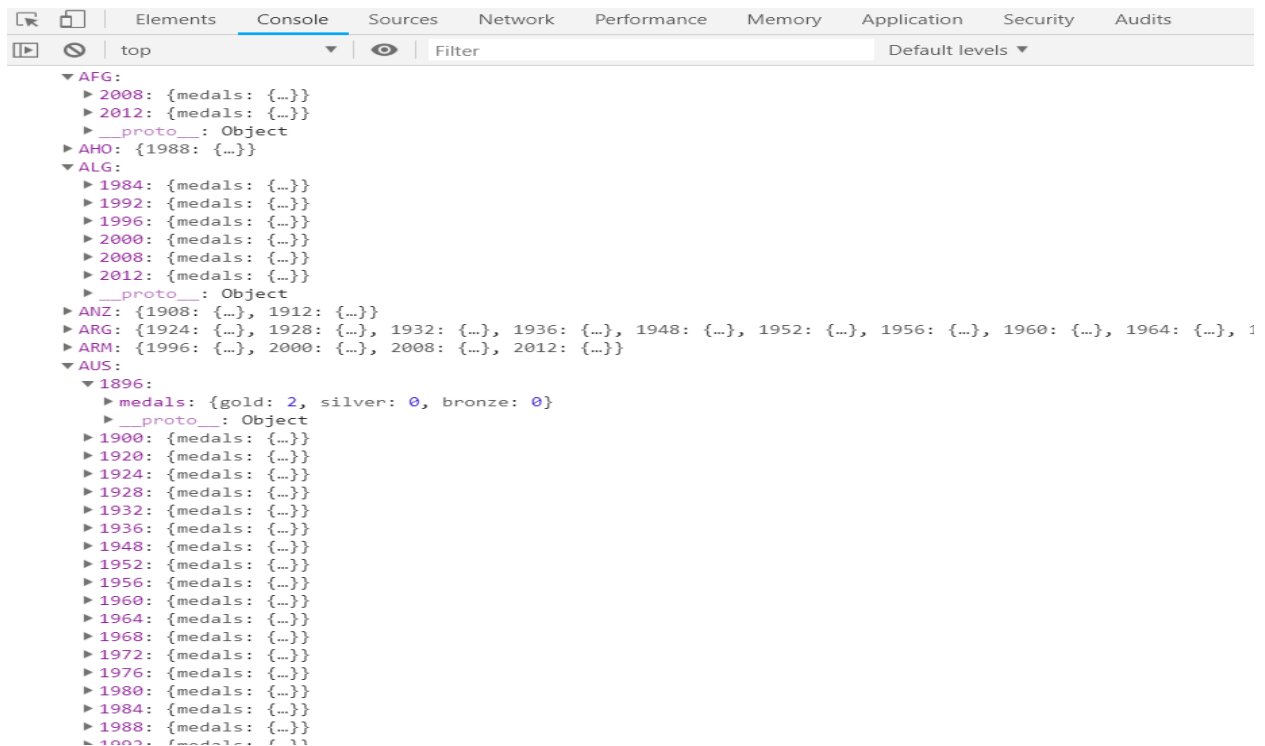
the code, we implemented a function named `aggregate`, which takes in two parameters, the `aggregate` parameter and the `grouping` parameter. Based on the values provided to this function, we can get the data required for all the views in our project.

Consider the example of the primary view of the project, i.e the country wise distribution of medals for each Olympic. This view would require the aggregation of our primary dataset based on the year parameter and then grouping by countries to get medal count for each country for a specific year. Assume the dataset csv to be loaded in **Olympics data** object. We can get the data in the desired form by simply calling the `aggregate` function with the `aggregate` parameter as `year` and grouping parameter as `country`. We came up with a robust design and implementation of this function to get the data for all the views.

Example 1: Aggregated by years and grouped by countries



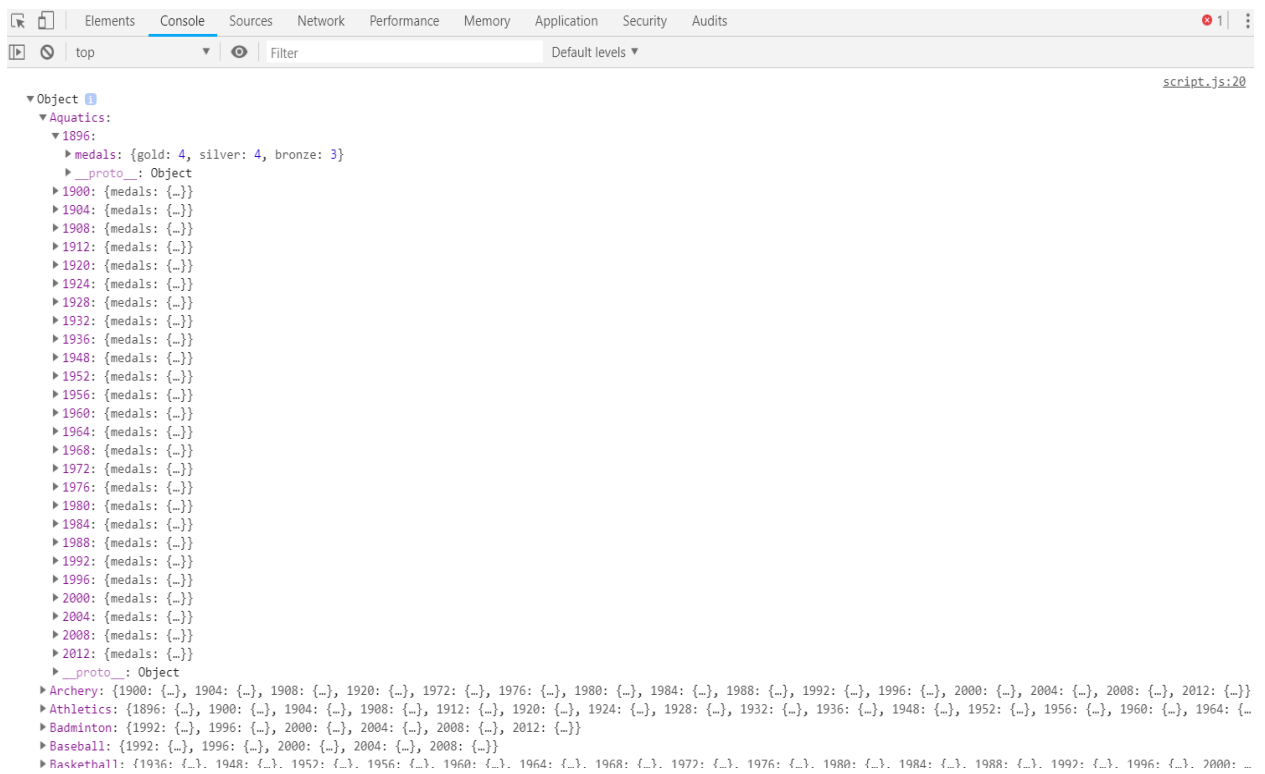
Example 2: Aggregated by countries and grouped by years



The screenshot shows the Chrome DevTools Console with the 'Console' tab selected. The top toolbar shows 'top' as the selected scope and 'Filter' as the search filter. The 'Default levels' dropdown is also visible. The console displays a large, nested object structure representing aggregated data by country and year. The structure is as follows:

- AFG:**
 - 2008: {medals: {...}}
 - 2012: {medals: {...}}
 - __proto__: Object
- AHO:** {1988: {...}}
- ALG:**
 - 1984: {medals: {...}}
 - 1992: {medals: {...}}
 - 1996: {medals: {...}}
 - 2000: {medals: {...}}
 - 2008: {medals: {...}}
 - 2012: {medals: {...}}
 - __proto__: Object
- ANZ:** {1908: {...}, 1912: {...}}
- ARG:** {1924: {...}, 1928: {...}, 1932: {...}, 1936: {...}, 1948: {...}, 1952: {...}, 1956: {...}, 1960: {...}, 1964: {...}, 1968: {...}, 1972: {...}, 1976: {...}, 1980: {...}, 1984: {...}, 1988: {...}, 1992: {...}, 1996: {...}, 2000: {...}, 2004: {...}, 2008: {...}, 2012: {...}}
- ARM:** {1996: {...}, 2000: {...}, 2008: {...}, 2012: {...}}
- AUS:**
 - 1896: {medals: {gold: 2, silver: 0, bronze: 0}, __proto__: Object}
 - 1900: {medals: {...}}
 - 1920: {medals: {...}}
 - 1924: {medals: {...}}
 - 1928: {medals: {...}}
 - 1932: {medals: {...}}
 - 1936: {medals: {...}}
 - 1948: {medals: {...}}
 - 1952: {medals: {...}}
 - 1956: {medals: {...}}
 - 1960: {medals: {...}}
 - 1964: {medals: {...}}
 - 1968: {medals: {...}}
 - 1972: {medals: {...}}
 - 1976: {medals: {...}}
 - 1980: {medals: {...}}
 - 1984: {medals: {...}}
 - 1988: {medals: {...}}
 - 1992: {medals: {...}}
 - 1996: {medals: {...}}
 - 2000: {medals: {...}}
 - 2004: {medals: {...}}
 - 2008: {medals: {...}}
 - 2012: {medals: {...}}
 - __proto__: Object

Example 3: Aggregated by sport and grouped by years



The screenshot shows the Chrome DevTools Console with the 'Console' tab selected. The top toolbar shows 'top' as the selected scope and 'Filter' as the search filter. The 'Default levels' dropdown is also visible. The console displays a large, nested object structure representing aggregated data by sport and year. The structure is as follows:

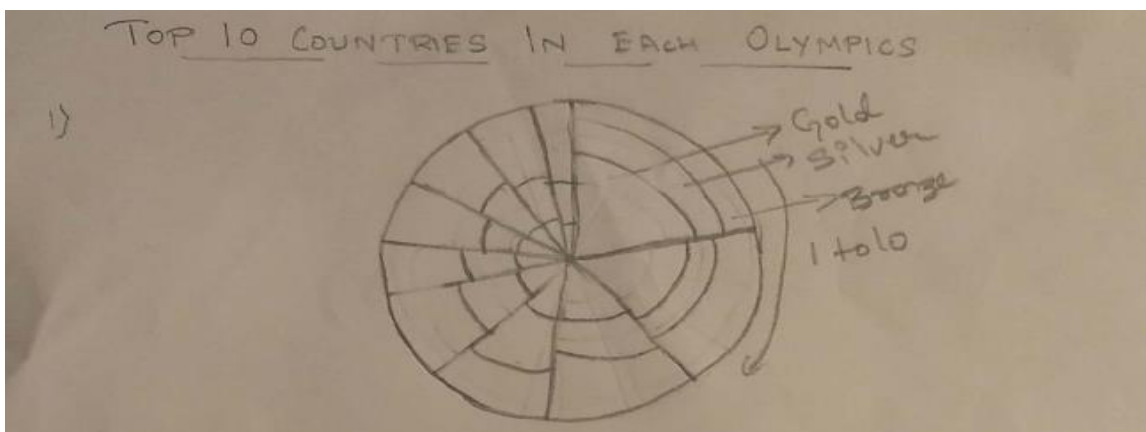
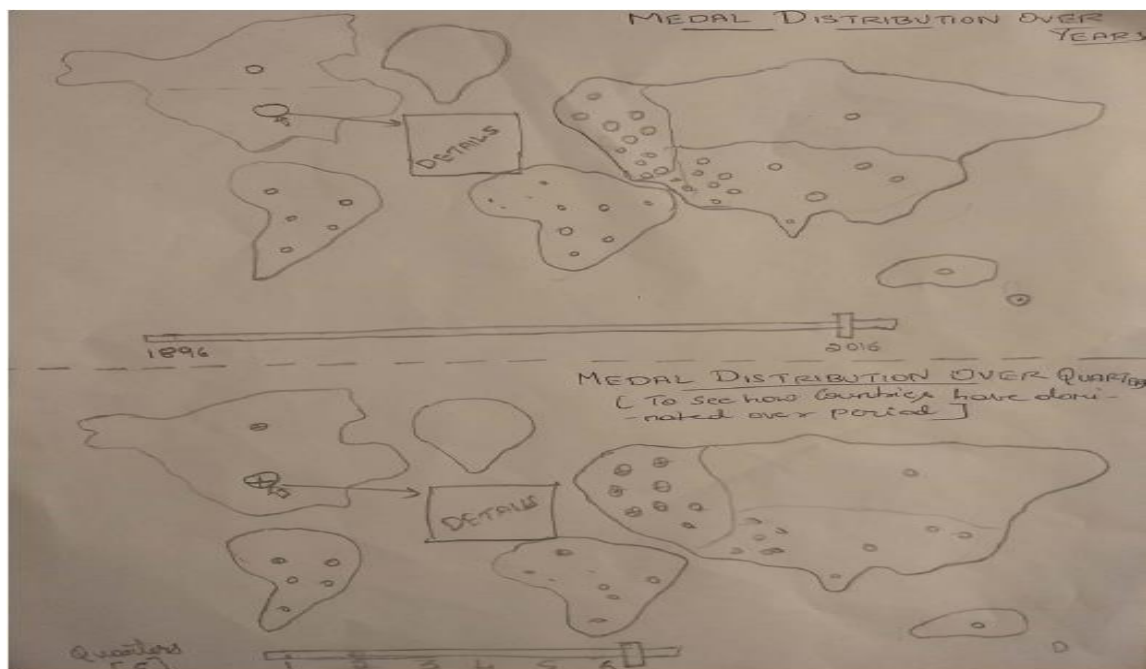
- Object**
 - Aquatics:**
 - 1896: {medals: {gold: 4, silver: 4, bronze: 3}, __proto__: Object}
 - 1900: {medals: {...}}
 - 1904: {medals: {...}}
 - 1908: {medals: {...}}
 - 1912: {medals: {...}}
 - 1920: {medals: {...}}
 - 1924: {medals: {...}}
 - 1928: {medals: {...}}
 - 1932: {medals: {...}}
 - 1936: {medals: {...}}
 - 1948: {medals: {...}}
 - 1952: {medals: {...}}
 - 1956: {medals: {...}}
 - 1960: {medals: {...}}
 - 1964: {medals: {...}}
 - 1968: {medals: {...}}
 - 1972: {medals: {...}}
 - 1976: {medals: {...}}
 - 1980: {medals: {...}}
 - 1984: {medals: {...}}
 - 1988: {medals: {...}}
 - 1992: {medals: {...}}
 - 1996: {medals: {...}}
 - 2000: {medals: {...}}
 - 2004: {medals: {...}}
 - 2008: {medals: {...}}
 - 2012: {medals: {...}}
 - __proto__: Object
 - Archery:** {1900: {...}, 1904: {...}, 1908: {...}, 1920: {...}, 1972: {...}, 1976: {...}, 1980: {...}, 1984: {...}, 1988: {...}, 1992: {...}, 1996: {...}, 2000: {...}, 2004: {...}, 2008: {...}, 2012: {...}}
 - Athletics:** {1896: {...}, 1900: {...}, 1904: {...}, 1908: {...}, 1912: {...}, 1920: {...}, 1924: {...}, 1928: {...}, 1932: {...}, 1936: {...}, 1948: {...}, 1952: {...}, 1956: {...}, 1960: {...}, 1964: {...}, 1968: {...}, 1972: {...}, 1976: {...}, 1980: {...}, 1984: {...}, 1988: {...}, 1992: {...}, 1996: {...}, 2000: {...}, 2004: {...}, 2008: {...}, 2012: {...}}
 - Badminton:** {1992: {...}, 1996: {...}, 2000: {...}, 2004: {...}, 2008: {...}, 2012: {...}}
 - Baseball:** {1992: {...}, 1996: {...}, 2000: {...}, 2004: {...}, 2008: {...}}
 - Basketball:** {1936: {...}, 1948: {...}, 1952: {...}, 1956: {...}, 1960: {...}, 1964: {...}, 1968: {...}, 1972: {...}, 1976: {...}, 1980: {...}, 1984: {...}, 1988: {...}, 1992: {...}, 1996: {...}, 2000: ...}

Design Evolution

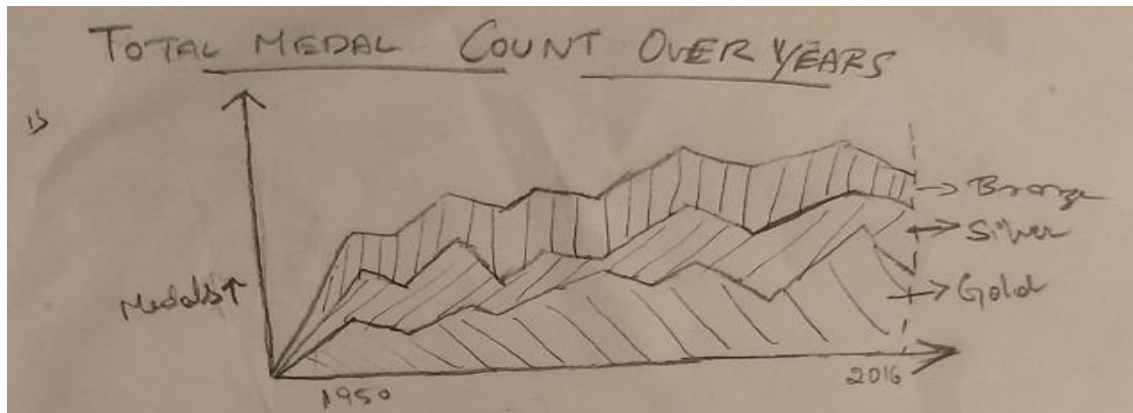
Proposal Design

In our proposal we had 5 views.

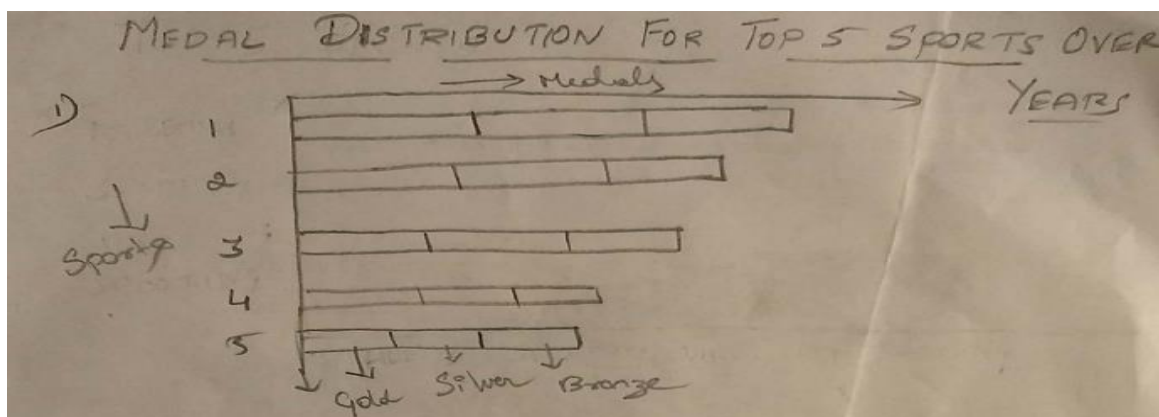
- ❖ **Map:** Once we launch the website our first visualization is a map which shows country with medal count. We use a circle to denote the countries that participated in that particular year and circle size represents the medals won.
- ❖ **Pie Chart:** We use this chart to view top 10 countries in selected year. Here each pie represents a country and each pie is divided into 3 parts representing different medal types.



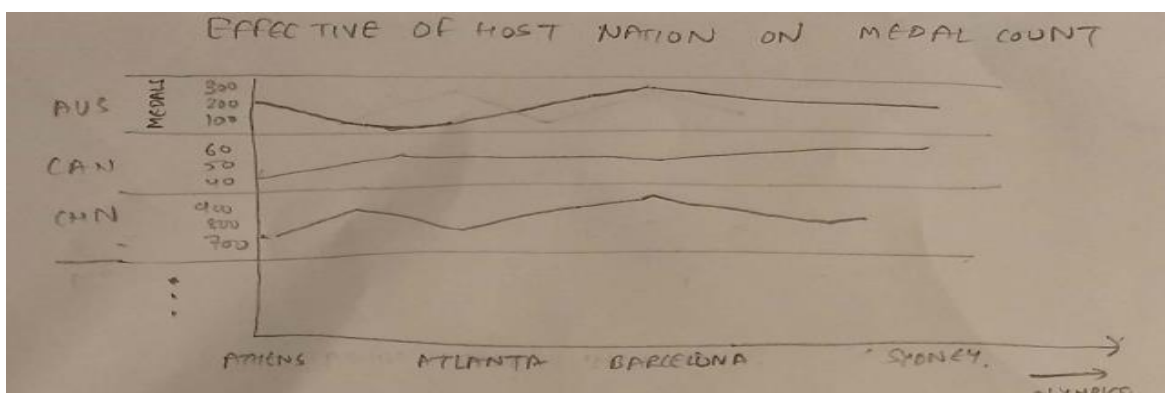
- ❖ **Area Chart:** This is used to visualize overall medal count in Olympics over the years. In this, x-axis represents year of Olympics and y-axis represents medal count. We have 3 areas in this chart with each representing different medal types.



- ❖ **Stacked Bar Chart:** Used to visualize top 5 sports (highest medal distribution) in each Olympics. Where x-axis represents Medal count and y-axis represents 5 sports. We have 3 stacks in this bar with each representing different medal types.



- ❖ **Line Chart 1:** Host nation performance can be visualized using this chart. Multiple line chart is stacked with each other to compare with other country performance.

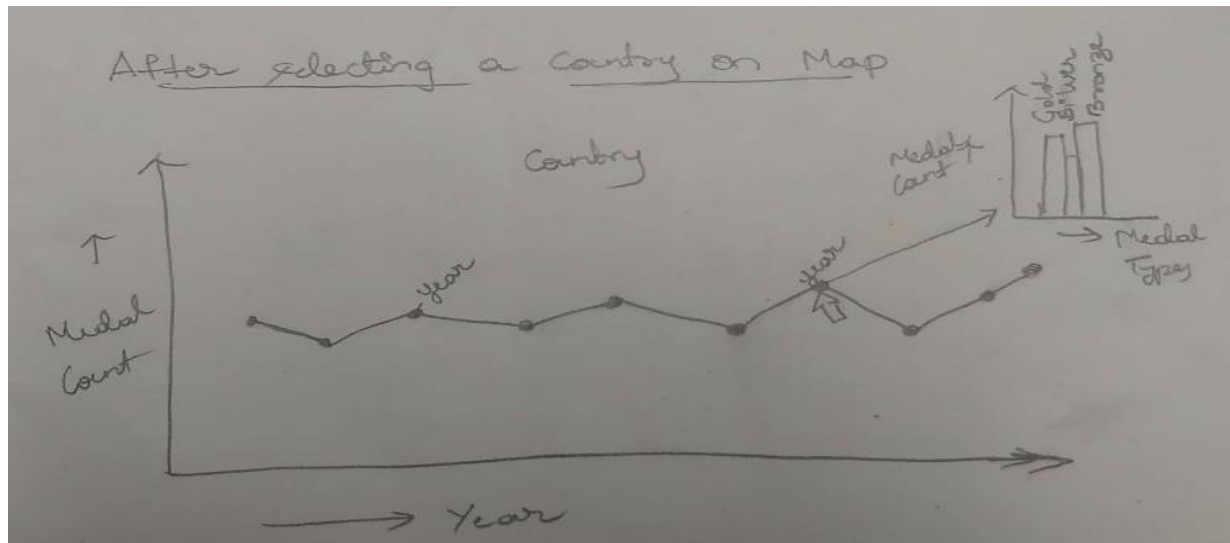


Design Changes

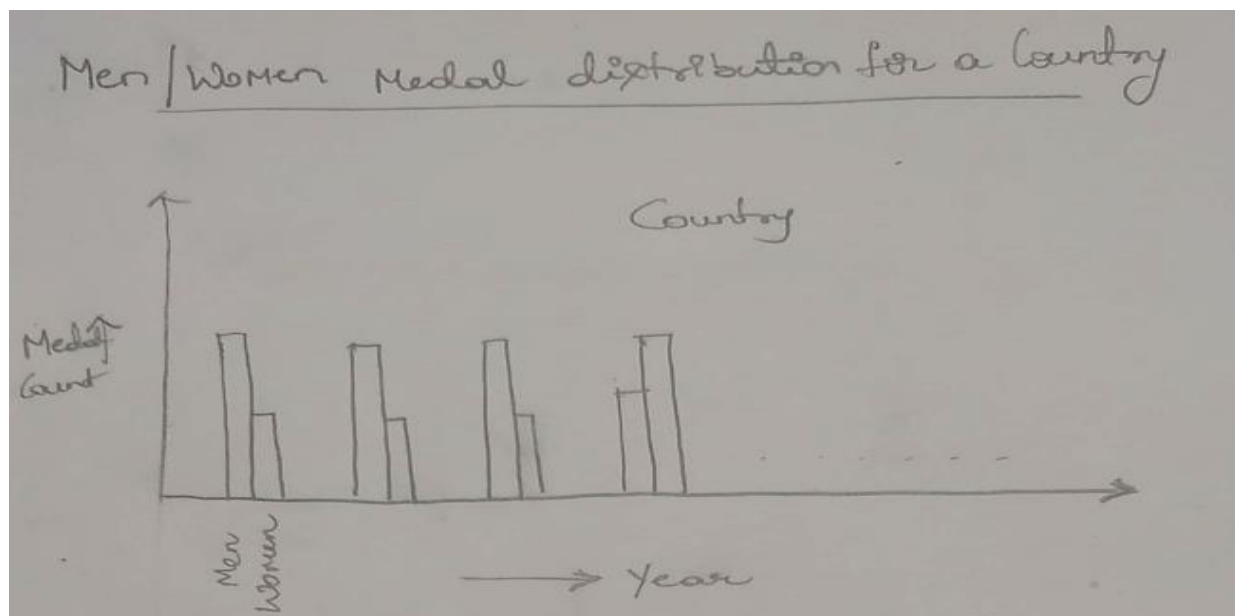
After getting feedback from team during feedback hours in class and discussing with our mentor **Mengjiao Han**, we decided to make some changes to the design and the interaction.

Additions

- **Line chart 2:** Used to visualize medal count over years for that country.

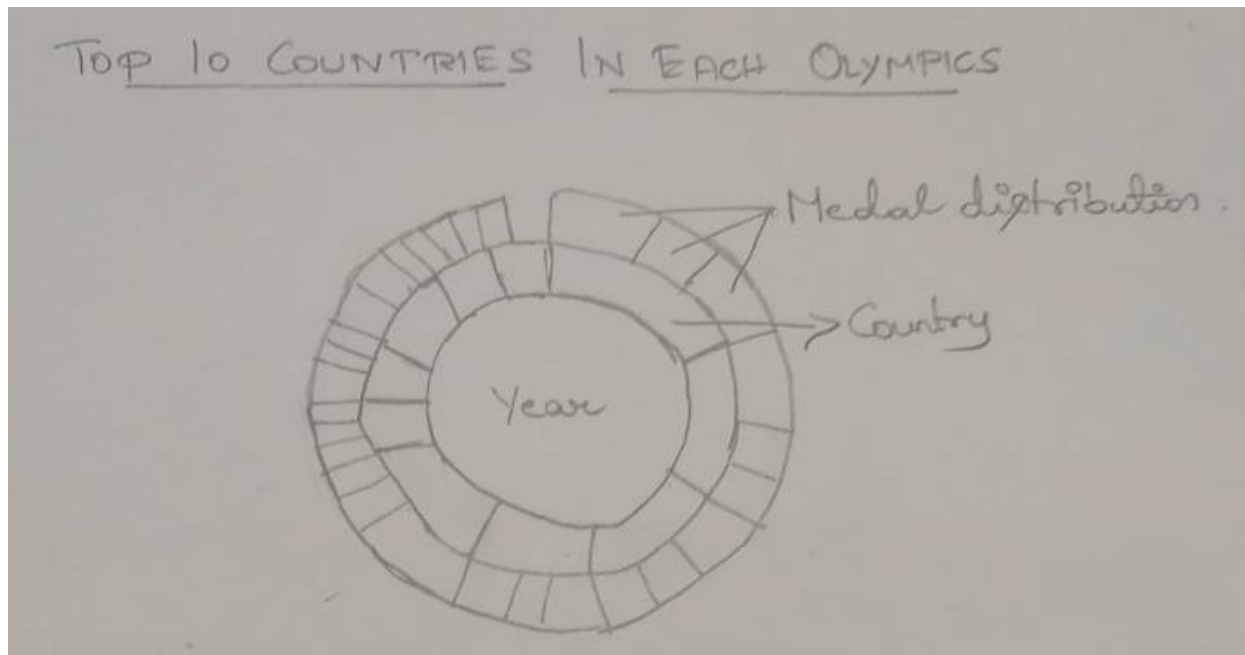


- **Bar chart:** Used to visualize men/women medal count for a country over years. X-axis represents years and Y-axis represents medal count.

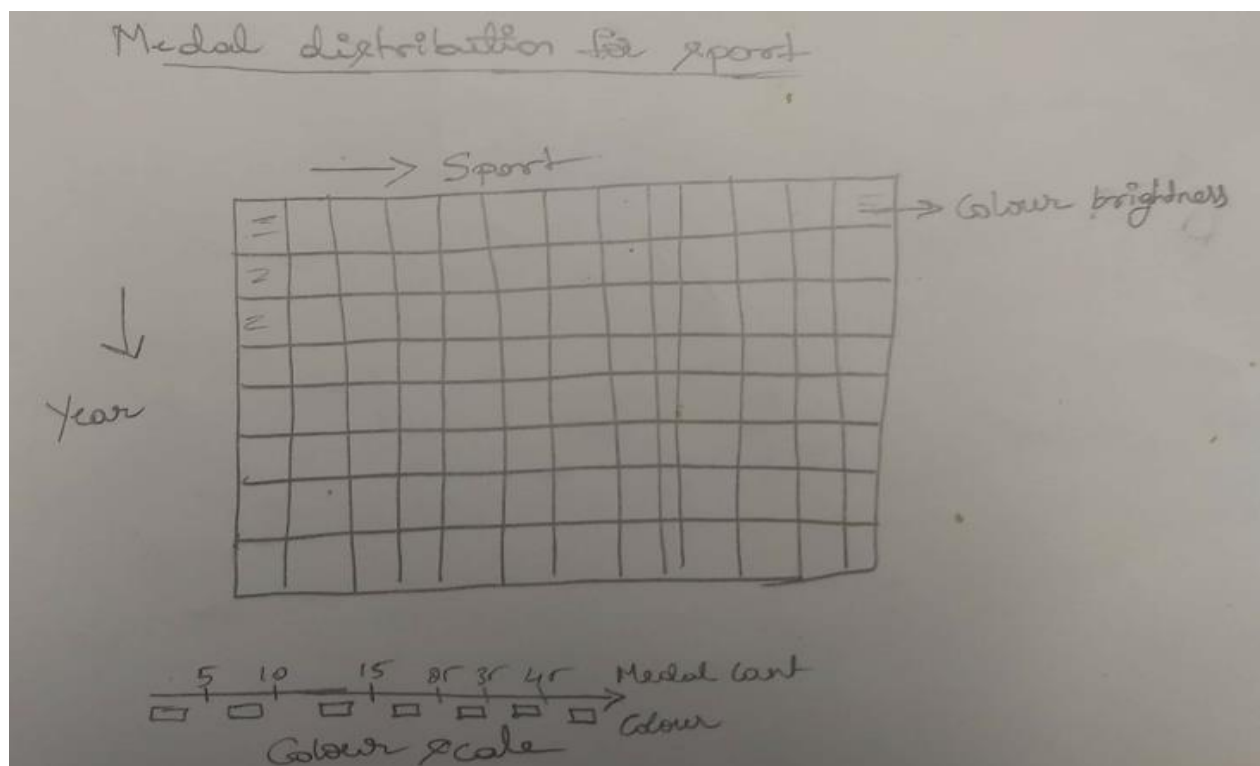


Changes

- **Pie Chart to Sunburst:** After getting suggestion from mentor we planned on visualizing top 10 countries in each Olympics using sunburst instead of pie chart.



- **Stacked Bar chart to Heat Map:** Here instead of displaying top 5 sports we planned on displaying all sports over the year using heat map instead of stacked bar chart.



Feedback

Now, after the feedback which we had got from the other team and plus from Mentor proved to be worthy for us. We could come with the final design after incorporating their suggestions which were suiting our project. For example, we decided to include the Sunburst and Heatmap instead of pie chart and stacked bar chart and including the home button and Filters as well to make it look more clean and effective visualization.

There were few imperative points which are really important for us which we found really effective such as we need to make sure that our data doesn't look Biased and the design should look cleaner.

Our focus is to make Final design elegant and so we derive a learning/inferences from that design. We will be incorporating all the important feedbacks which we have got from our peers and especially the TA's as they have the expertise in that. At this stage feedback played an important role for us as we could figure out what's more important for us.

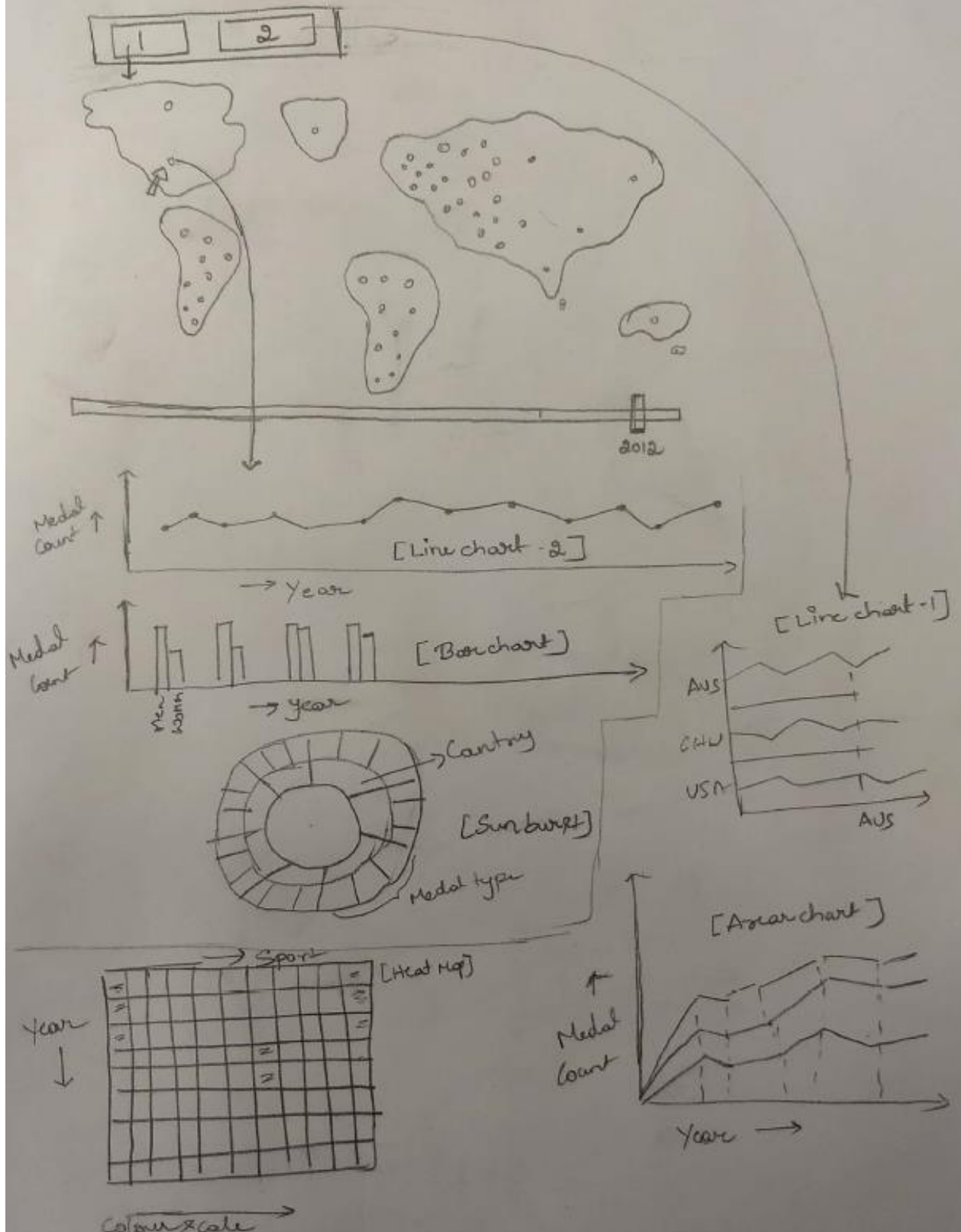
Final Design

Below is the final design of the Project. Which has two options:

- On launch, default option-1 will be selected, and a map will be displayed with country and their medal count for a default year. Once you click on a country, three visualization will be displayed.
 - Line Chart 2
 - Bar Chart
 - Sunburst
- On selecting option-2, three visualization will be displayed,
 - Line Chart 1
 - Area Chart
 - Heat Map

All the visualizations will be interactive and will plan on using brush after implementing the below design.

FINAL DESIGN



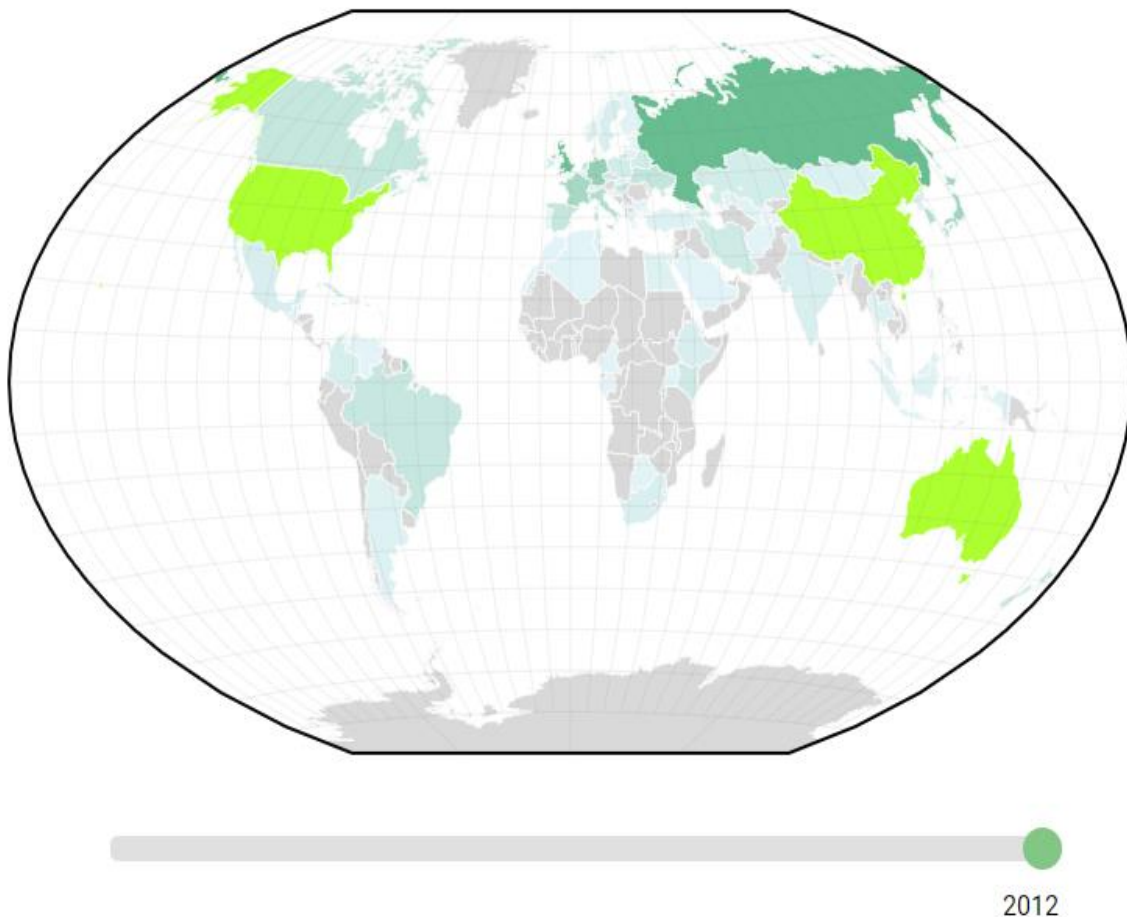
Implementation

This section discusses the intent and content of each functionality of the interactive visualization components with clear and well-referenced images showing the key design and interaction elements.

World Chart (Countries Participation)

Olympics stats visualization starts with a world map. This visualization is used to represent the countries which participated in Olympics in particular year and won the medals. The color concentration of country states the medal count. Higher the concentration more medals won and vice versa. The slider below world map can be used to view the visualization from year 1896(time when modern Olympics started) to 2012. Moving the slider from left to right shows the increase in countries participation over Olympics. This world map and slider is linked to three other charts. i.e. bar chart, sunburst and line chart.

Countries Participation



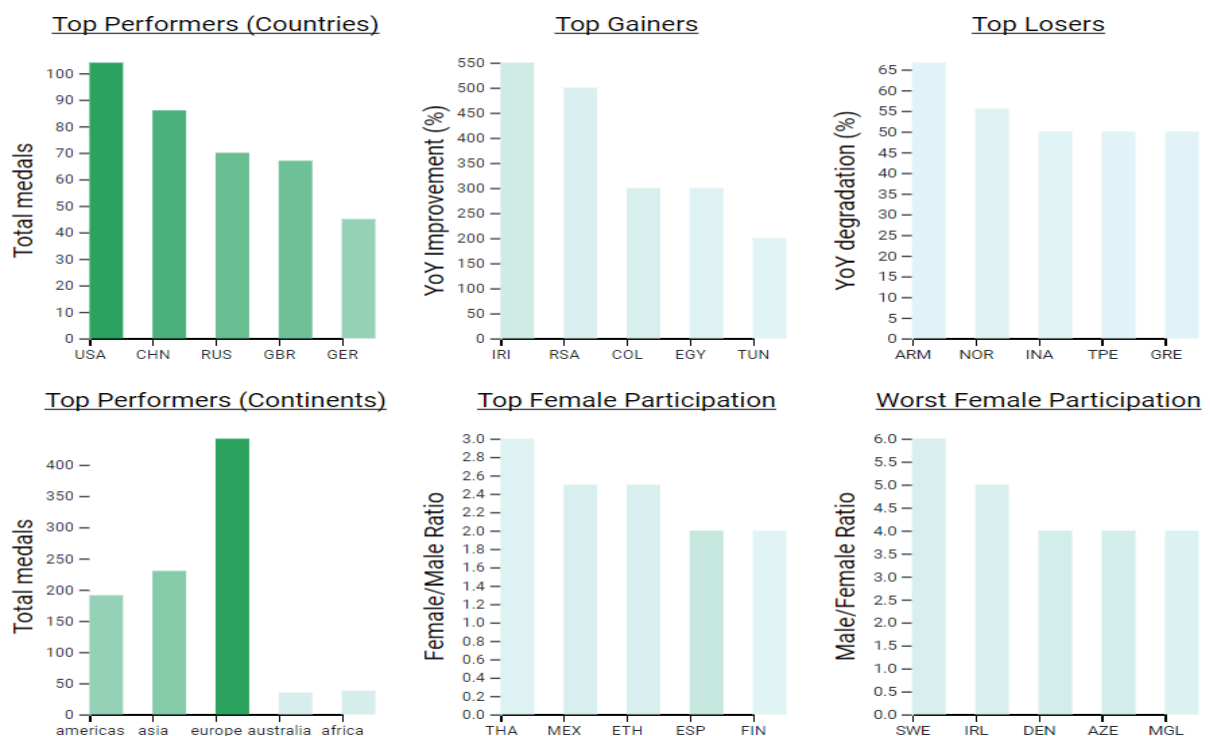
Bar Charts (Statistics)

Six bar charts are used to represent top performers (Country), Top Gainers, Top Losers, Top Performer (Continent) and Female Participation (Top and Worst) over the years of Olympic Games history. As we move the slider these bar chart changes showing different statistics.

The first bar chart titled Top Performers and the fourth bar chart titled Top Performers use color concentration and bar heights to visualize the top performers on the Country and Continent criteria respectively. Top performer visualization gives us a moving picture of two things- thorough picture of countries and continents that perform well constantly (Stay in the chart most of the time) and countries with varied performance (Appear in chart few times). This gives clear picture of countries which started to dominate in later part of Olympics. On mouseover, bars display the medal split.

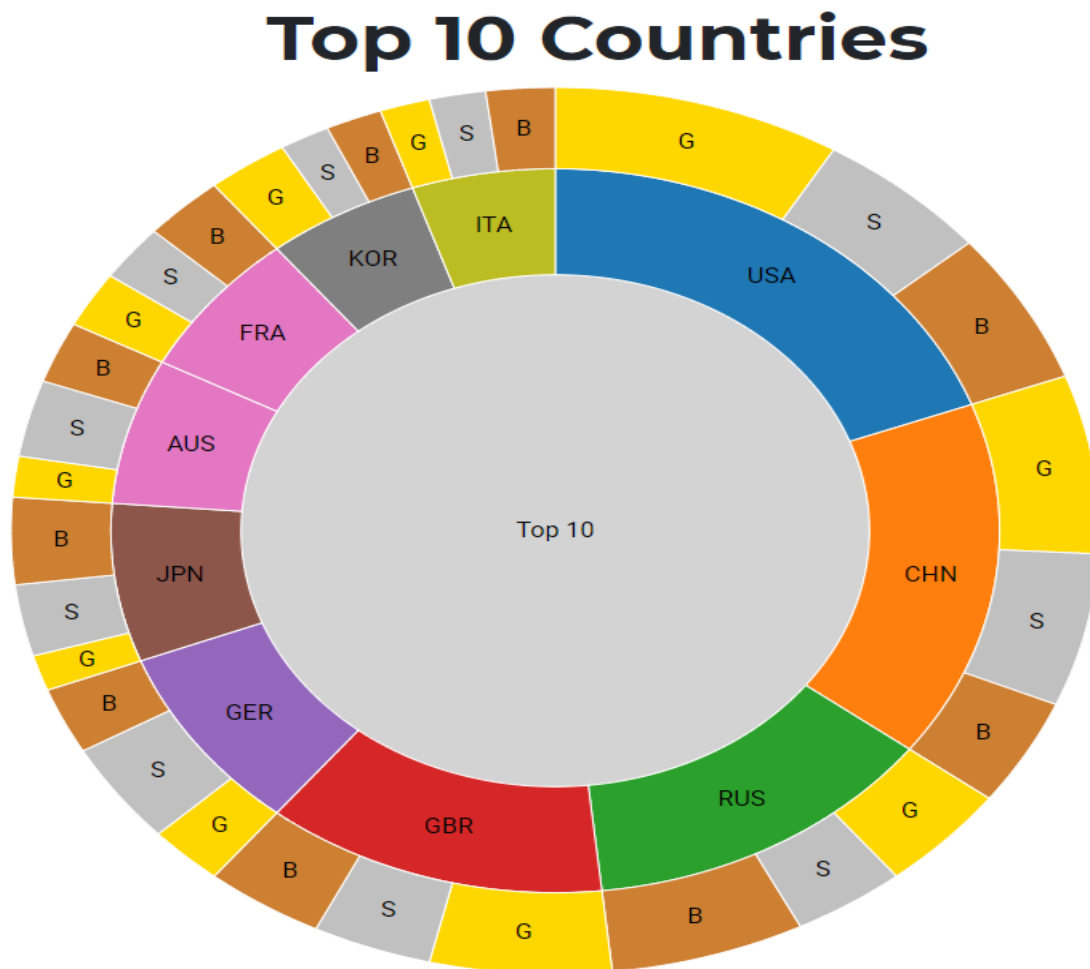
The second and third bar chart named top gainers and top losers visualizes about countries which gained their medal count and lost their medal count respectively in the next Olympics. Here y-Axis of both charts represents percentage of improvement and decrement respectively.

The Fifth and sixth chart named top and low female participation visualizes female to male and male to female ration. As we wanted to check the countries which send more female participants and less female participants respectively. We used this visualization to check the statistics. Here Y-axis represents the ratio female to male and male to female respectively and x-axis is used to rank the counties from top to low.



Sunburst (Top 10 Performers with medal distribution)

This Chart is used to visualize the top 10 performers over each Olympics. Upon moving the slider, the sunburst changes. The first circle shows the country list and the second circle shows the medal split for each country. The chart is interactive, if you select any country it generates a pie chart with medal distribution for that country. If you press at center It comes to actual chart.

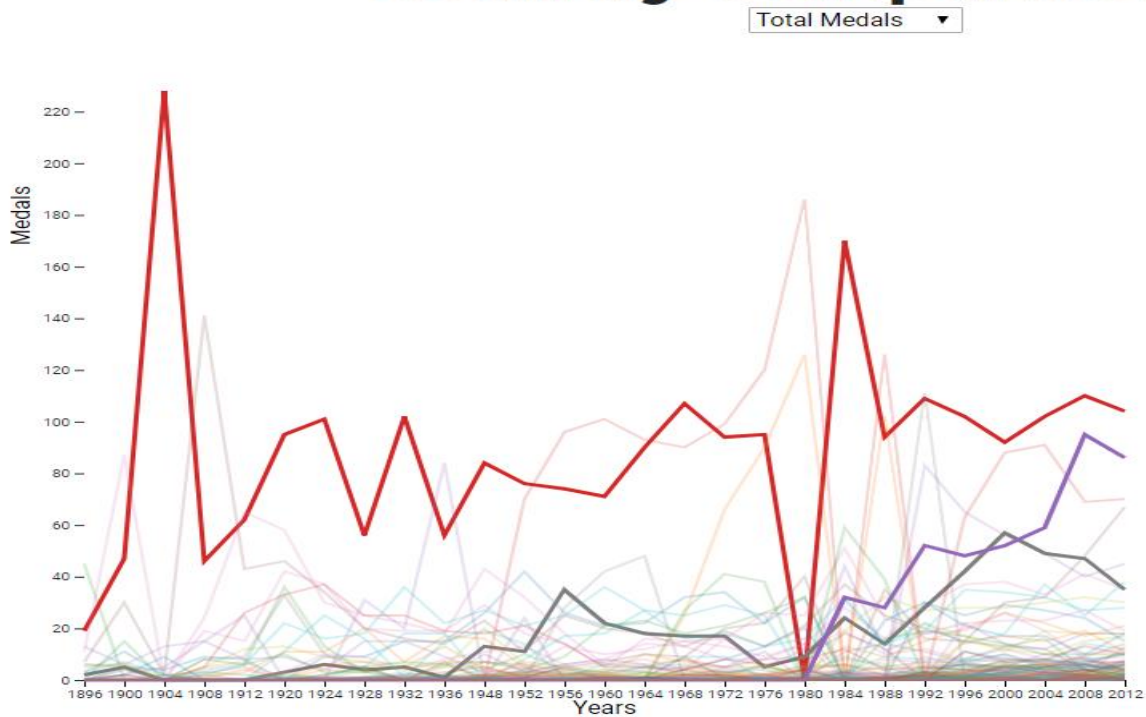


Line chart (Countries Comparison)

This Graph is used to visualize performance of all the countries in a single view. It gives a clear picture of constant and varied performing countries. Upon mouseover on line, it highlights the line and dulls other line so that we can see the performance of specific country. This graph is interactive with the world map. Upon selecting a country on world map, it highlights specific country line.

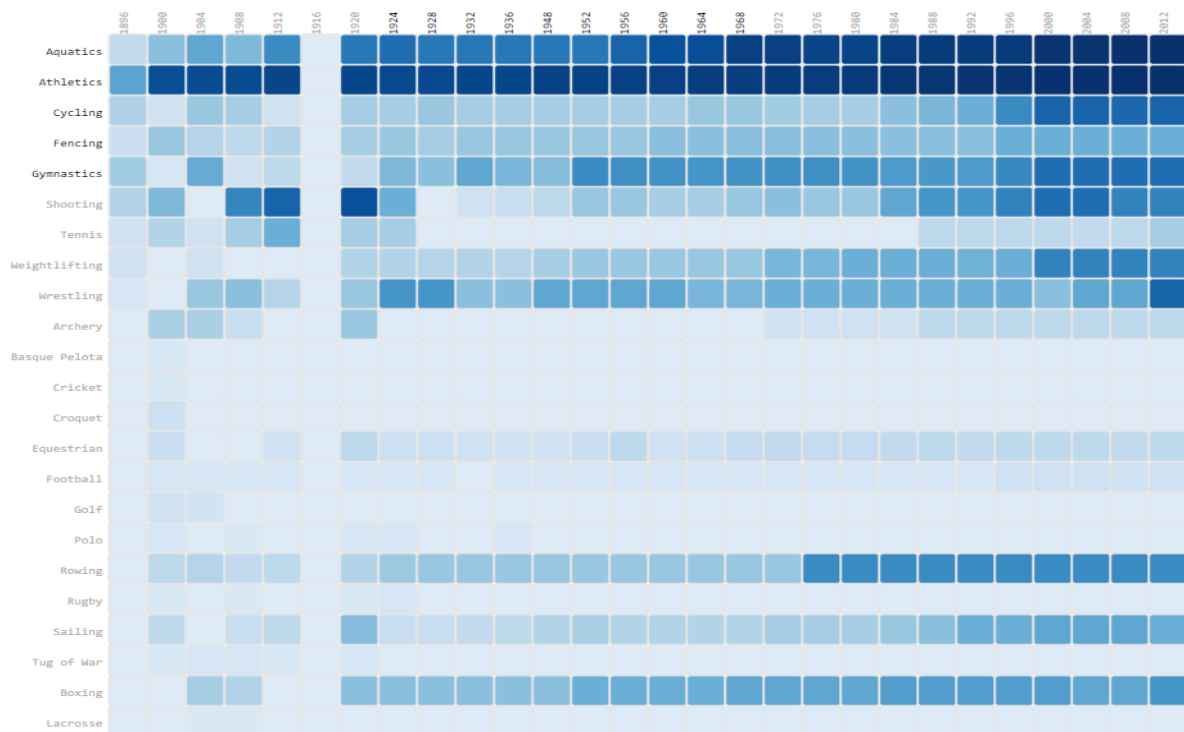
We have used a drop down in the chart which is used to visualize the country performance on different medal category. This gives clear picture of countries dominance over gold, silver and bronze medals.

Country comparison



Heat Map (Medal distribution for sport)

Sport-Wise Medal Distribution Over The Years

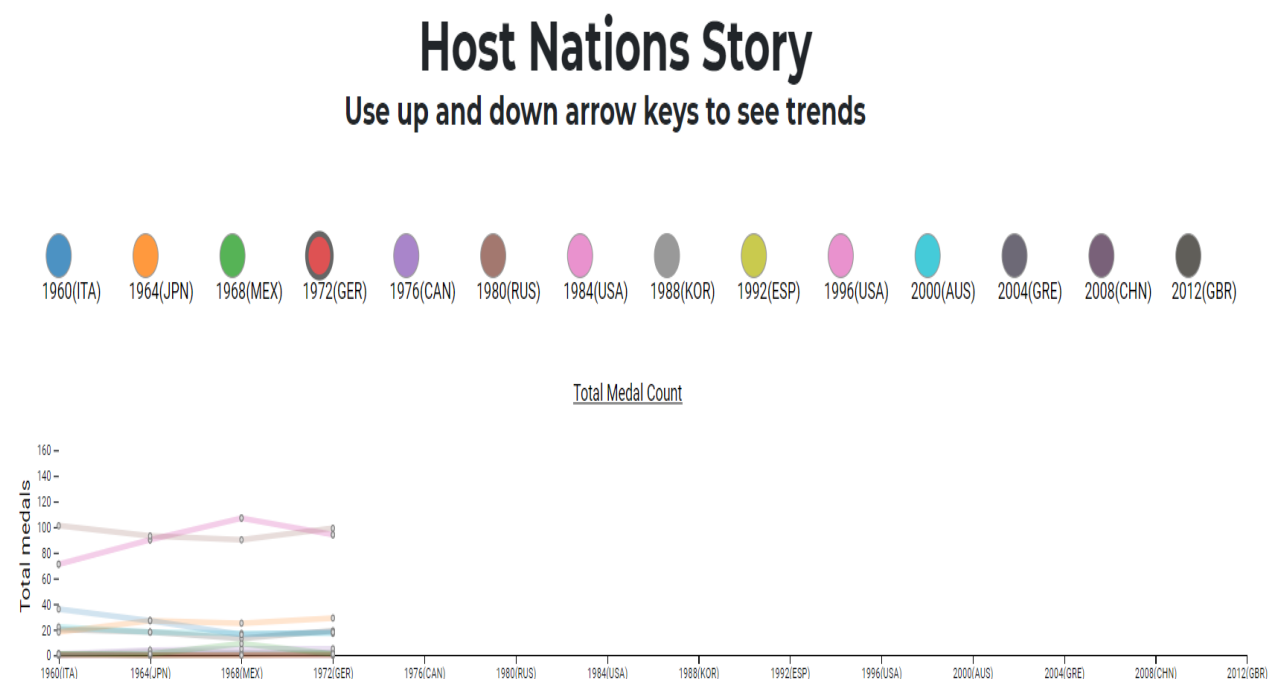


This Map is used to visualize the medal distribution for each sport over each Olympics. The y-axis represents the sports and x-axis represents Olympic years. The color concentration of block states the medal count. Higher the concentration more medal distribution and vice versa. Upon mouseover on block it gives the medal count.

Line Chart (Host Nation Performance – Story Telling)

This visualization says the story of host nation performance from 1960. It can be used to view the performance of countries which hosted the Olympics from 1960. In this chart we have a year wise button from 1960 to 2012 with host nation name and we have a line graph which is linked to year wise button. The y-axis represents medal count and x-axis represents years and host country names. upon clicking on a year wise button, it gives the line graph till that particular year from 1960. The lines include only the countries that have hosted from year 1960 to 2012. Upon mouse over on line it highlights the line and gives the name of the country referred to the line.

We called this chart story telling as it has this extra feature. That is using upward and downward keys on keyboard we can visualize the host nation performance on the line chart. For example, starting with an empty line chart if you go on pressing upward key, the line starts to move from year 1960 towards year 2012 telling the story of the host nations.

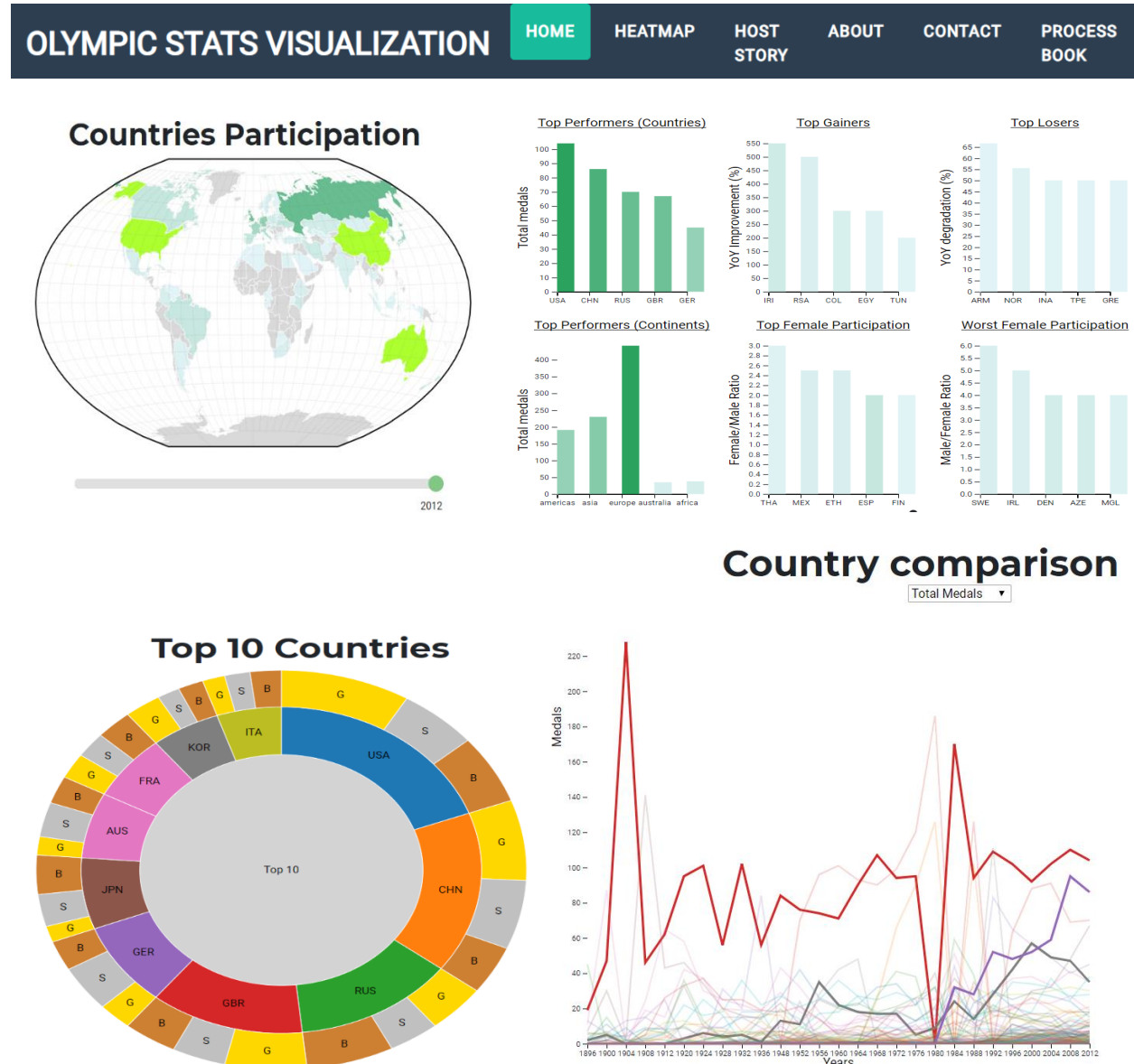


Changes Made During Implementation

We made one major change to our final design while implementation. After discussing we decided on visualizing **bar chart** with different statistics instead of **Area Chart** with overall medal count. The reason behind that was we could answer many questions using this statistics chart.

Evaluation and Conclusion

Visualization 1



Once we launch the visualization, first it directs to the home page which has 4 visualization linked to the slider as shown above. Once the slider moves the world map, bar chart and sunburst vary depending on particular year.

From the world map and line chart visualization, we can answer the question **How European countries dominated at the initial stage of Olympics?** As we can see that most of the European countries started to participate from the beginning of Olympics compared to countries like China and Russia (Which are dominating in recent times). Also, we can see how some countries like United States and Great Britain have dominated from beginning to recent times. Also,

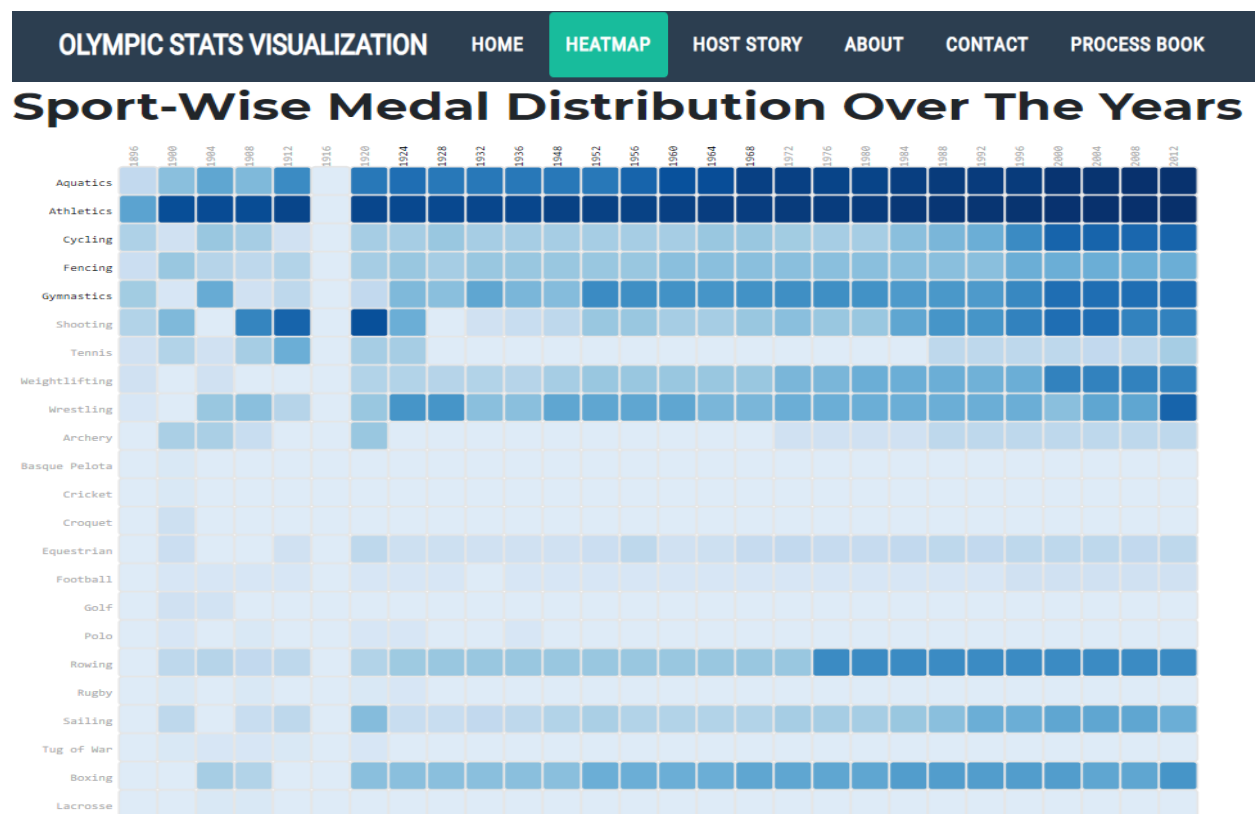
we can answer the question, **did world wars affect Olympics? How?** We can see that there was no Olympics in the year 1940 and 1944 which shows world cancelled Olympics for 2 consecutive year, and It can be seen USA boycotted 1980 Olympics held in Russia and Russia boycotted 1984 Olympics held in USA.

Also, we can use world map and line chart combo to compare the performance between multiple countries. As seen in the image, United states, China and Australia are highlighted in world map and the respective lines are highlighted in line chart.

From bar chart statistics we can evaluate the top 5 improvement and decrement slope of the countries in consecutive Olympics. Also gives top 5 men to women and women to men ratio. This helps to answer question **Which countries will send more women participants?**

From Sunburst we can visualize top 10 performed countries, the advantage here is distribution of medal category. This helps to visualize that countries with less overall medal count can be in higher position than a country with more overall medal count. This is because of the advantage of having more higher category medal (i.e. gold > silver or silver > bronze).

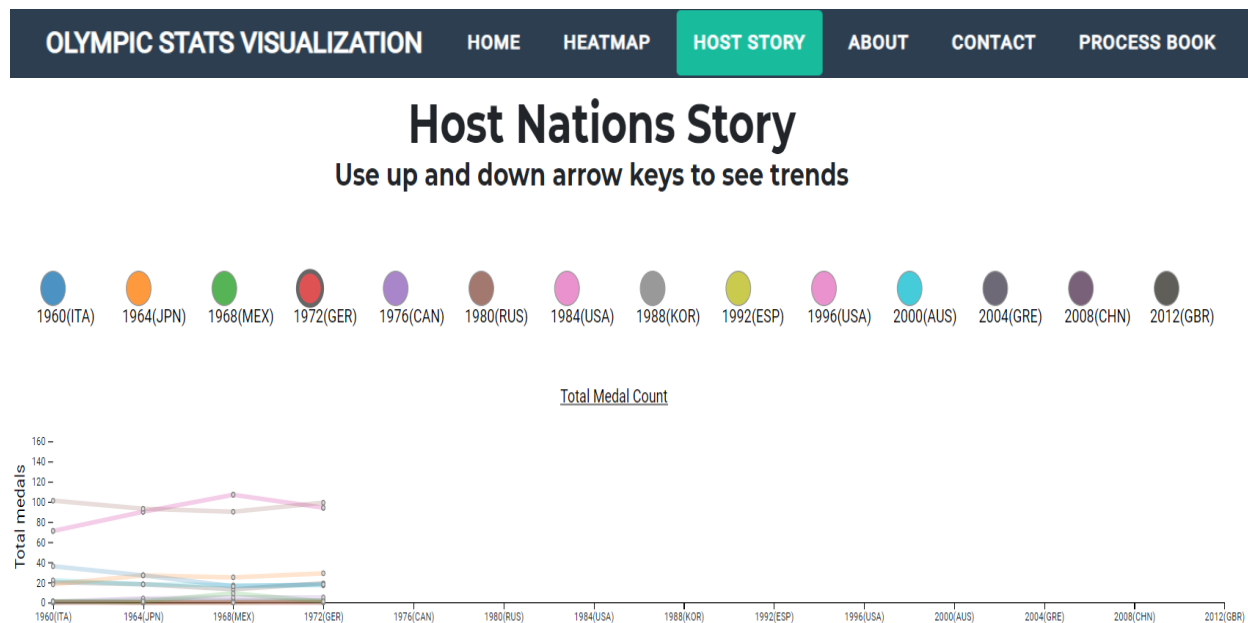
Visualization 2



This Heat map helps to visualize the change in trend of sports over years. As we can see, the medal count for Aquatics and Athletics has gradually increased over every Olympics. Also, we can

see that for some sports its has increased and then again dropped because of various reasons like less participants, or less support etc. For Example, Wrestling and shooting sport has up and down in medal count distribution. What also matters is once a host nation is decided for the Olympics, the type of games that more priority should be given in that country is planned.

Visualization 3



This visualization tries to answer host nation performance. There is trend that host nations win more medal compared to its previous Olympic medal count. We wanted to visualize this in a story. So, we used story telling concept to visualize the trend. It shows how most of the host nation countries perform well when they host the Olympics and win more medals compared to its previous Olympic. For example, Russia and United states performance in 1980 and 1984 Olympics respectively.

Conclusion

Through our visualizations, we were able to learn more about the datasets in our project and were able to answer the questions that we had at the beginning of the project. Overall, our project was to visualize the outcomes of the Olympic games from 1896-2012. We used varied visualizations to project the data and linked all the visualizations to make it more interactive, comprehensible and user friendly.