Interactive Visualization of Olympic games

Kiranmayee.D

Sai Vamshi.D

Table of Contents:

Team Information

Sai Vamshi D	U1266122
Kiranmayee D	U1265299

Github repo:

https://github.com/saivamshidobbali/120YearsOfOlympics

Background and Motivation

The Olympic games are the world's foremost sports competition that first took place in 1896 in Athens, Greece. They are held every four years, with Summer and Winter Games alternating by occurring every 2 years apart. Thousands of athletes from all over the world participate in the events. It is one of the greatest events where the whole world come together and cheer for their idols. The games are considered not just between the athletes, but between the nations. Athletes are representatives of their nation, showing the world that their country can compete in the world.

It is important for people to know the significance of the Olympic games and their countries position in the performance. Even though mainstream media, social media promotes and telecasts news regarding Olympic games, it only happens during the time of the events. Also, there are no tools that represent the history of the games over the years which people can access whenever needed. Therefore, we propose an interactive visualization tool for the Summer and Winter Olympics that presents, the number of medals won, the country which won, details of the athletes, categories of sports throughout the years.

Questions:

The aim of the project is to :

1. Effectively design a visualization system where users can interact and gain information on the Olympic games.
2. User should be able to select the year to see the results of the games pertaining to that year.
3. Is the country's GDP and the number of medals won correlated?
4. How is the participation of women athletes changing from history?
5. Comparison between Gold, Silver and Bronze medals won by each country?

Data

This is a historical dataset on the modern Olympic Games, including all the Games from Athens 1896 to Rio 2016. Note that the Winter and Summer Games were held in the same year up until 1992. After that, they staggered them such that Winter Games occur on a four-year cycle starting with 1994, then Summer in 1996, then Winter in 1998, and so on. A common mistake people make when analyzing this data is to assume that the Summer and Winter Games have always been staggered.

Dataset:
https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results

Data Preprocessing:

Need to drop unused columns and limit the number of countries represented to 20 using python. Need to filter out noisy data and remove the duplicates. Data regarding countries' GDP need to be gathered and combined with Olympics dataset.

Data files used are:

athelete_events_modified.csv

gdppc.csv

truncated_latest.csv

Used the csv data from the above files and converted into json using d3.nest() method.

Design Evolution:

Design 1:

Year slider:

It synchronizes all the views in this design with a particular year. (i.e 1896 - 2016). It also shows information related to the location of the Olympics in that particular year.

Scatterplot

Scatterplot shows a comparison between the number of medals won by a country and its GDP per capita. How is an increase in GDP resulted in an increase in participation from that country and also an increase in the number of medals won by that country over a period of 120 years.
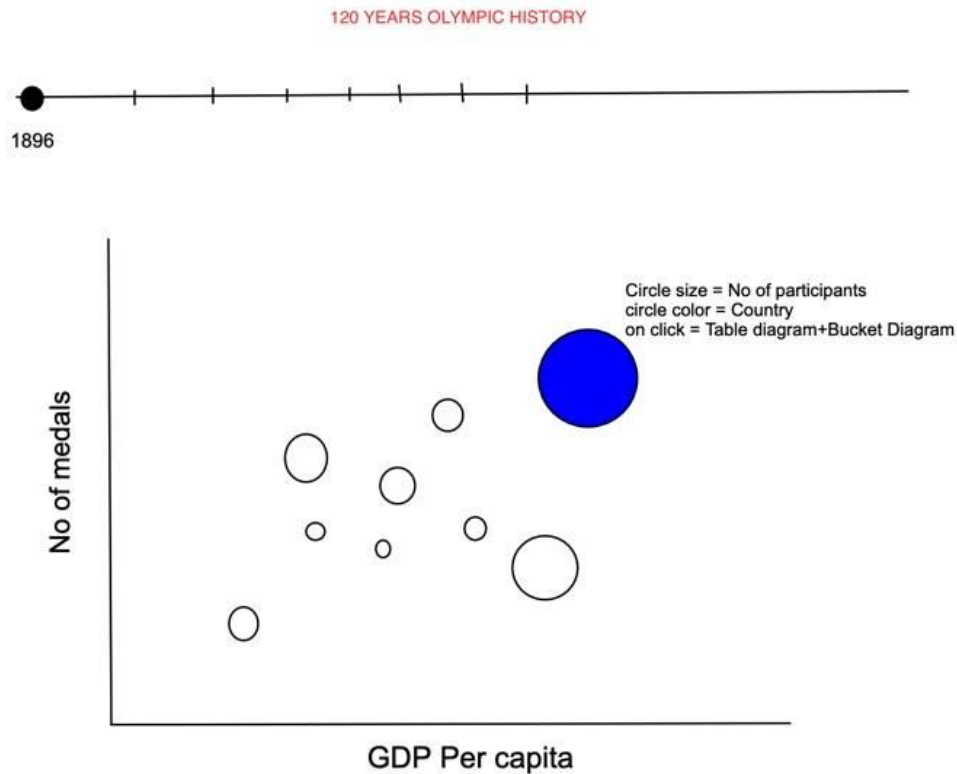
Marks and Channels
- Color represents the country name.
- Size of the circle represents the total number of participants in the Olympics from that particular country
- The position of the circle encodes the number of medals won by the country and its GDP per capita.
- Year slider fetches information for that particular year.
- By clicking a particular circle that circle gets highlighted, the rest of the circles are diluted and also a drop-down listing information related to each and every sport played by that country is seen in Medals Table.

Table Chart:

This table showcases information on the number of medals (gold, silver, bronze) won by each country in each sport. All the columns can be sorted.

This table is linked to year slider, it fetches information of a particular year.

Initially only aggregated values of medals won by a country are shown as bars, on clicking a specific country, we can see a drop-down listing information related to each and every sport played by that country and also respective country is highlighted in scatterplot.

COUNTRY	GOLD	SILVER	BRONZE	TOTAL
USA	30			
Swimming				
Gymnastics				
Shooting				
Weight lifting				
CHINA				
Swimming				
Gymnastics				
Shooting				
Weight lifting				

Marks and channels:

Marks = Horizontal Bars

Channels = Colors, Length of the bar.

Gender Inequality

Each of the circles represent a particular athlete, hover on the circle to get athlete information. These athletes are colored based on their gender and segregated into different sports categories they represent.
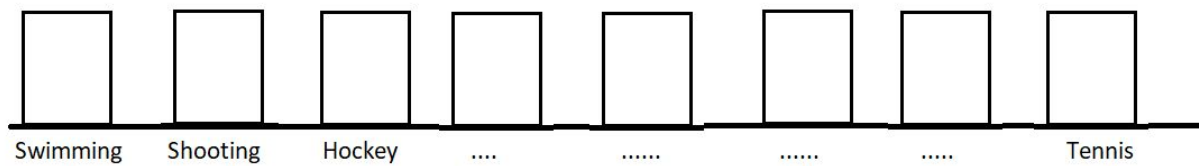
There is a toggle switch provided at the top, which when pressed all these athletes are sorted into three different bins based on their gender. This shows which shows the inequality between different genders.

These bins are synchronized with the year slider, showing data related to a particular year.
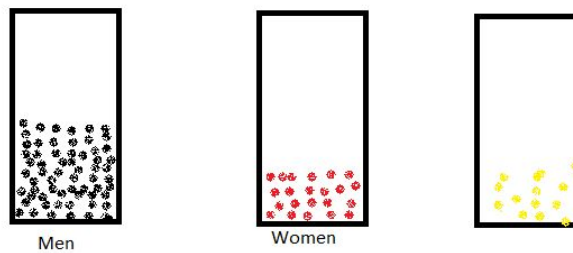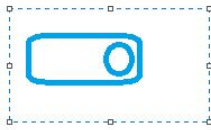
Categorical Data:

Swimming   Shooting   Hockey   ....   Tennis

Gender Inquality

Men   Women   Mixed

Aggregated Data

Swimming   Shooting   Hockey   ....   Tennis
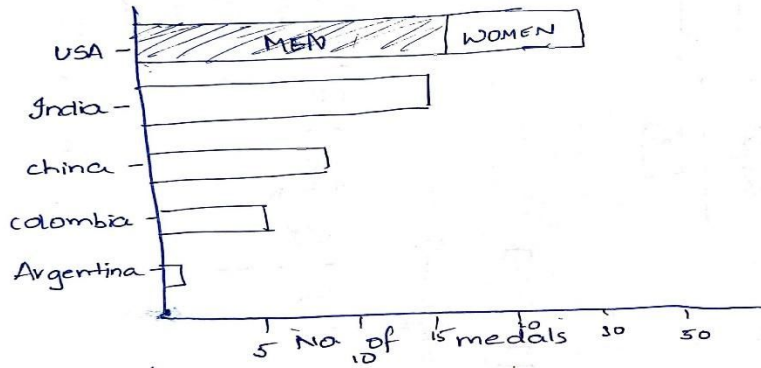
Gender Inquality

Men   Women

Bar Graph

This bar graph shows, ranking of different countries based on the number of medals won by them until that particular year. It is also synchronized by year slider with other views showing data until a particular year. Y-axis represents the country name and X-axis represents the number of medals won. It also encodes male and female medal winner in the bar graph using color, basically it is a stacked bar graph.

TOP 10 COUNTRIES RANKING TILL YEAR

Horizontal bar chart titled "TOP 10 COUNTRIES RANKING TILL YEAR" showing number of medals by country (USA, India, China, Colombia, Argentina), with bars divided into MEN and WOMEN segments. X-axis labeled "No. of medals" with values 5, 10, 15, 10, 30, 50.

Design 2:

Tree map

Design

Treemap

Area of each rectangle encodes number of medals won by a country over a period of time.

The aggregated Map over 120 years shows how these medals are distributed among countries.

Year slider

USA   China   Canada

As Above Treemap shows cummulative value of total Medals won by that country untill that year shown by year slider.

This dynamic change in the size of the country shows how the medals are getting distribute and how some countries have heed start, others start late.

Year slider

Category (dropdown)

| All |
| Swimming |
| Shooting |

USA | China

Along with Aggregated tree map that shows all the medals won by a country over a period, they will be a dropdown menu to choose one particular sports category, which when selected shows statistics for that one sport alone.

Tree map represents Medals won By top 15 Countries only

Marks: Rectangular shape
Channel: Area of Rectangle

treemap further gives us information about subcategory of each sport, like Mixed doubl Men's singles, women's singles in tennis. We can get information about No. of Medals won by each country in these sub-categories.
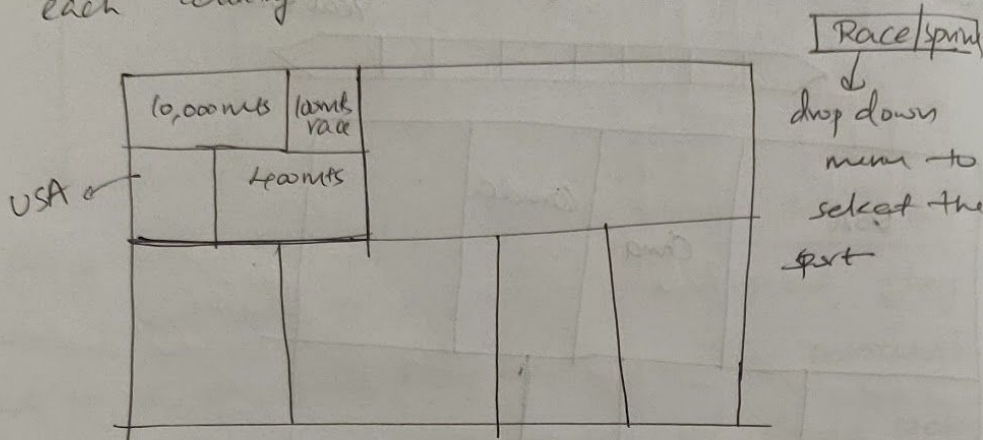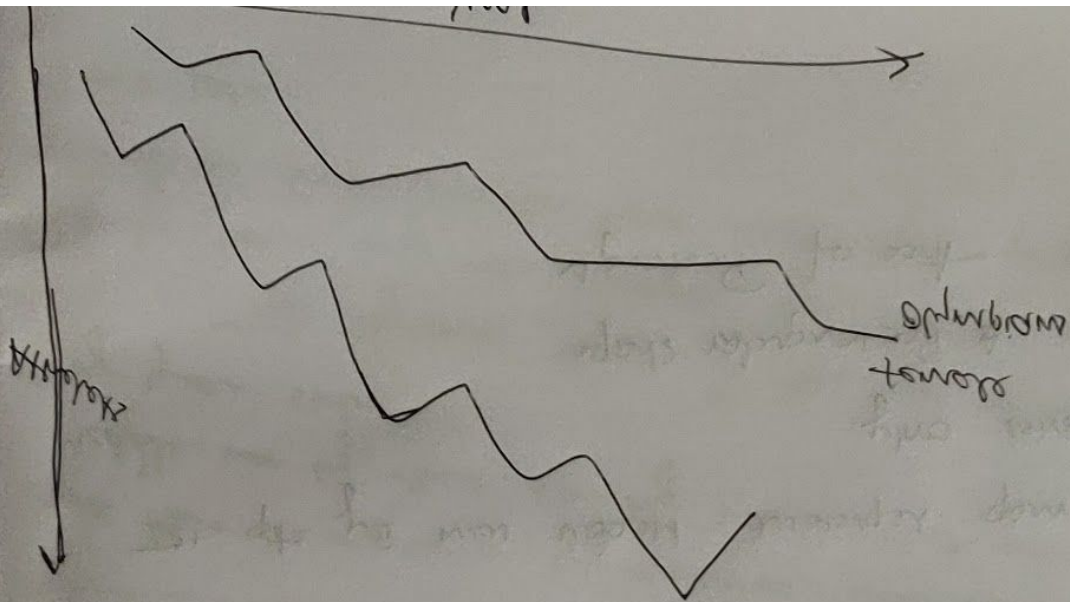

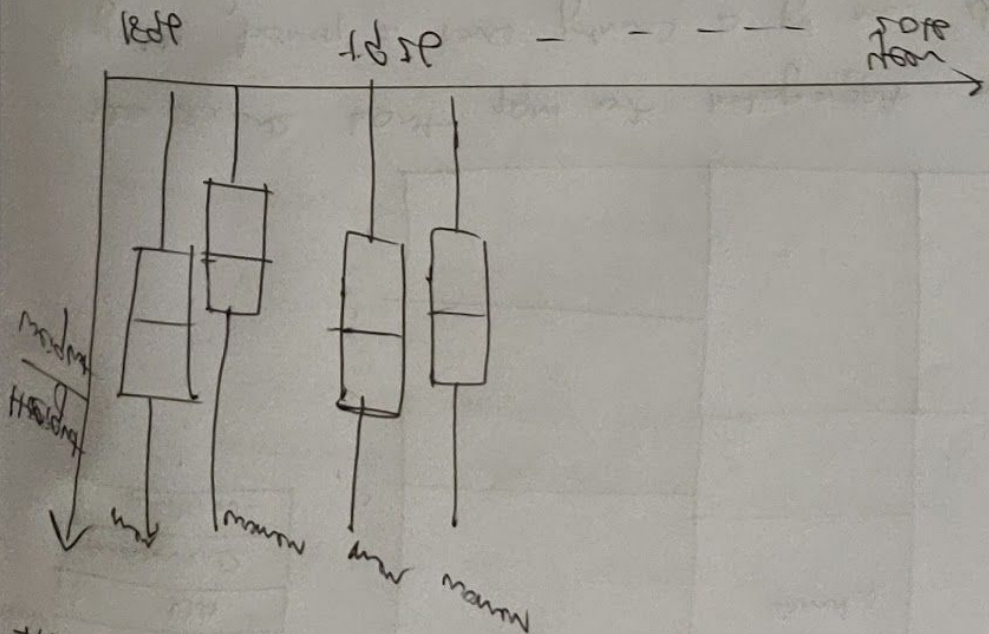Color — encodes country
Area — Medals won by each country.

Line chart and Box plot

Above Line Chart shows difference between no. of male and female Athletes over the years. This actually shows gender gap between male vs female over the years.

Design 3:

Histogram
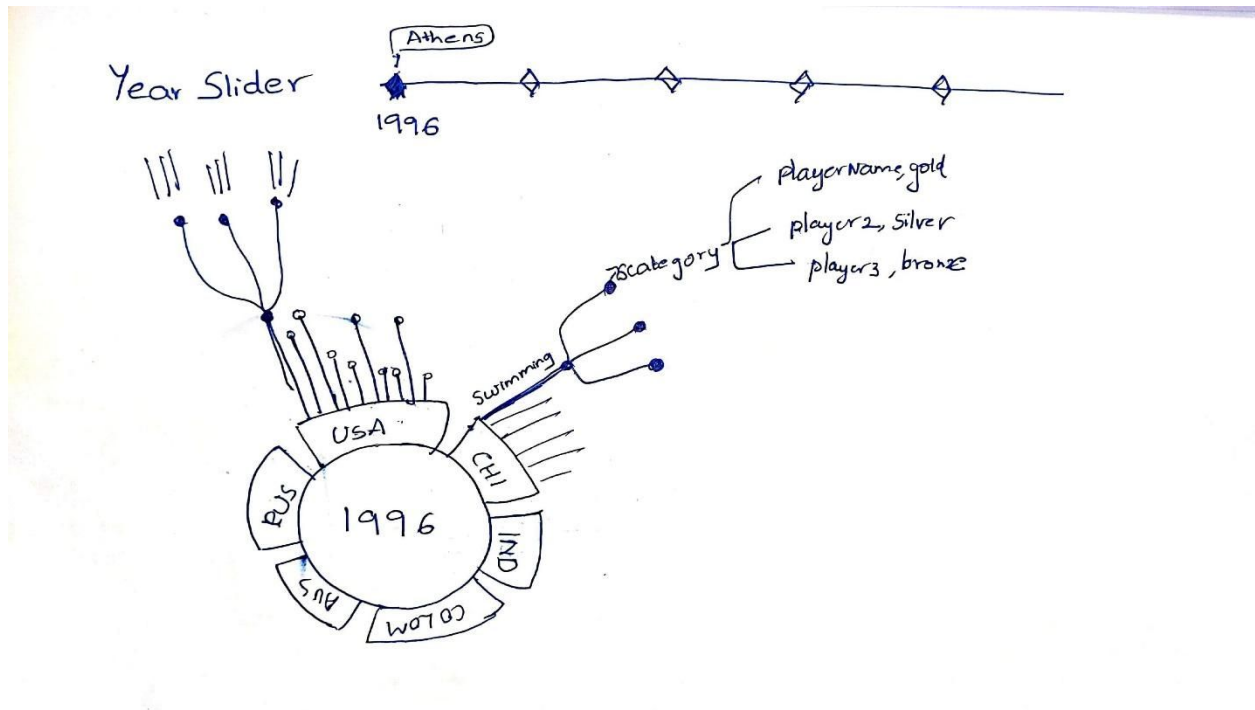
The above histogram is synchronized to year by year slider, wherein each segment represents a country and bars on top of the country represents medals won in each category, for example, swimming, shooting, tennis, etc.

When clicked on these bars it expands to represents sub-categories like 100mts Race, 400 mts race in running which further expands to list players who won the medals.

Word Map:

It shows aggregated values of medals won by the top 20 players in the world over the last 120 years.

Player with the highest number of medals is represented by the largest font, player with the lowest number of medals with the smallest font.

Final Design Implementation:

1) Scatter plot and Table chart

Choosing scatter plot from 1st design, as it can easily incorporate GDP data and number of participants. Also choosing table chart from 1st design as it acts as a support for the scatter plot as it gives more information like medals won by a country in specific sports category.

- In Scatter plot, X-axis, Y-axis , circle size parameters can be chosen.
- Color is organized for countries according to their. continent
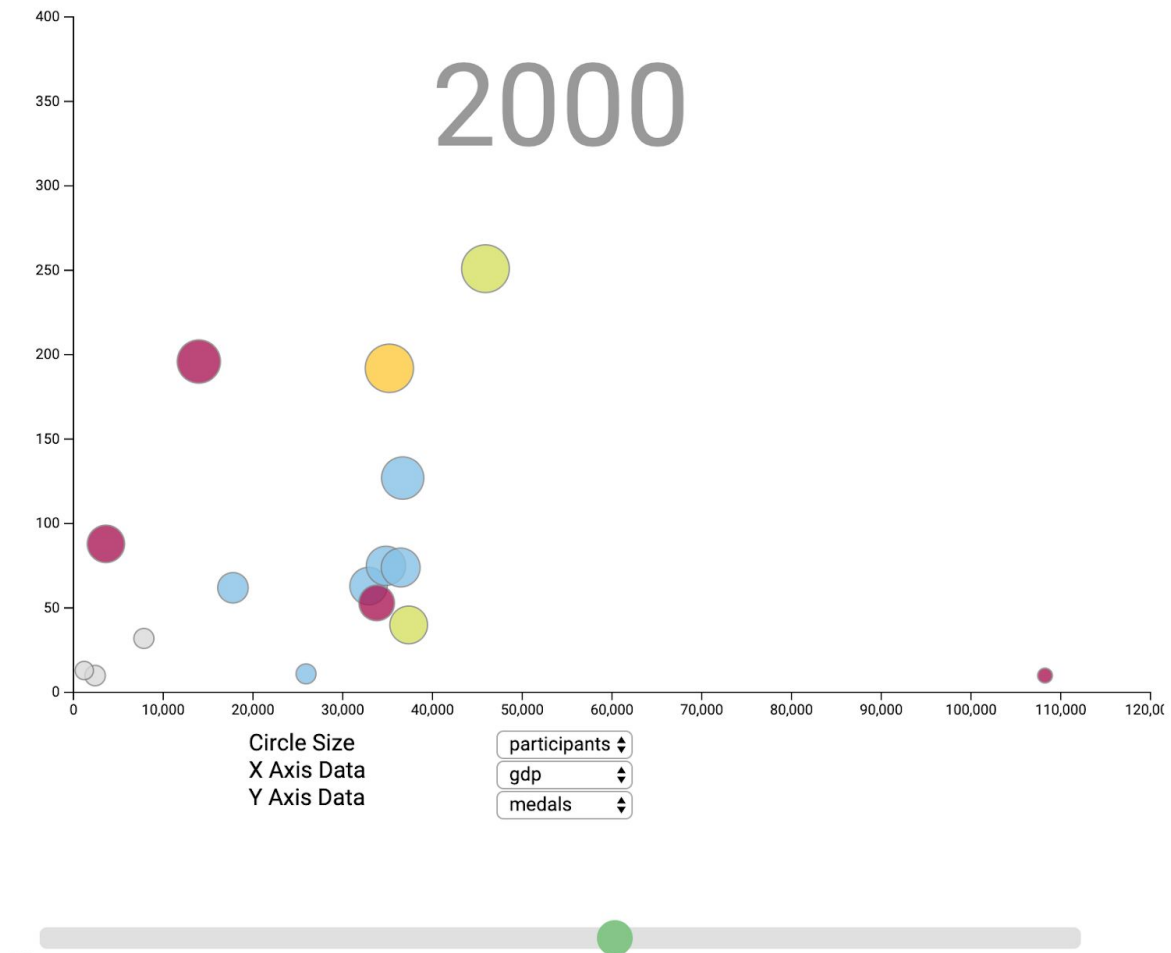
Team	Gold	Silver	Bronze	Total
Great Britain	22	12	20	54
Soviet Union	66	54	67	187
East Germany	31	64	23	118
Hungary	25	4	24	53
France	22	14	30	66
India	0	1	0	1
Italy	22	29	14	65
Australia	60	54	69	183
North Korea	0	3	1	4
Jamaica	0	3	20	23
Canada	4	23	4	31
Pakistan				
Japan	5	8	31	44
United States	130	51	61	242
China	39	17	23	79
Portugal	0	2	0	2
Qatar	0	1	0	1