120 Years of Olympic games

Basic Information

Team Information

Sai Vamshi D	U1266122
Kiranmayee D	U1265299

Github repo:

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

Background and Motivation

The modern **Olympic Games** or **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.

Thousands of athletes compete against each other every 4 years. But more than that, the Games are not just between athletes, they are between nations. On an average 92 countries will compete against each other. But more than that, the Olympians are a symbol of honor for their country. They are representatives of their nation, showing the world that their country has what it takes to compete with the world.

Furthermore, it's also a uniting force that brings together hundreds of nations, participating side by side. There are no current efficient tools for visualizing on geographic distribution of medals in each country, as well as the statistics of categories of each sport.

We, therefore, propose an interactive visualization tool for the summer and winter Olympics over 120 years of history.

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.

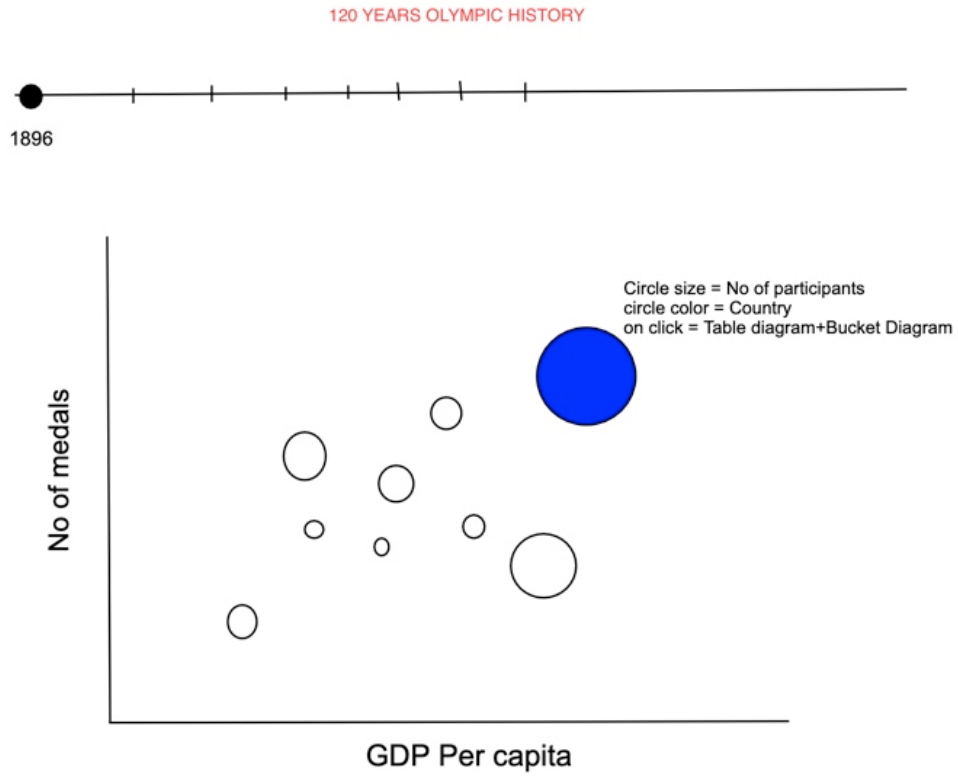
Designs:

Year slider:

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

Scatterplot

Scatterplot shows comparison between number of medals won by a country and its GDP per capita. How increase in GDP resulted in increased of participation from that country and also increase in 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 total number of participants in the Olympics from that particular country

Position of the circle encodes 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, 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.

Medals Table

This table showcases information of 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.

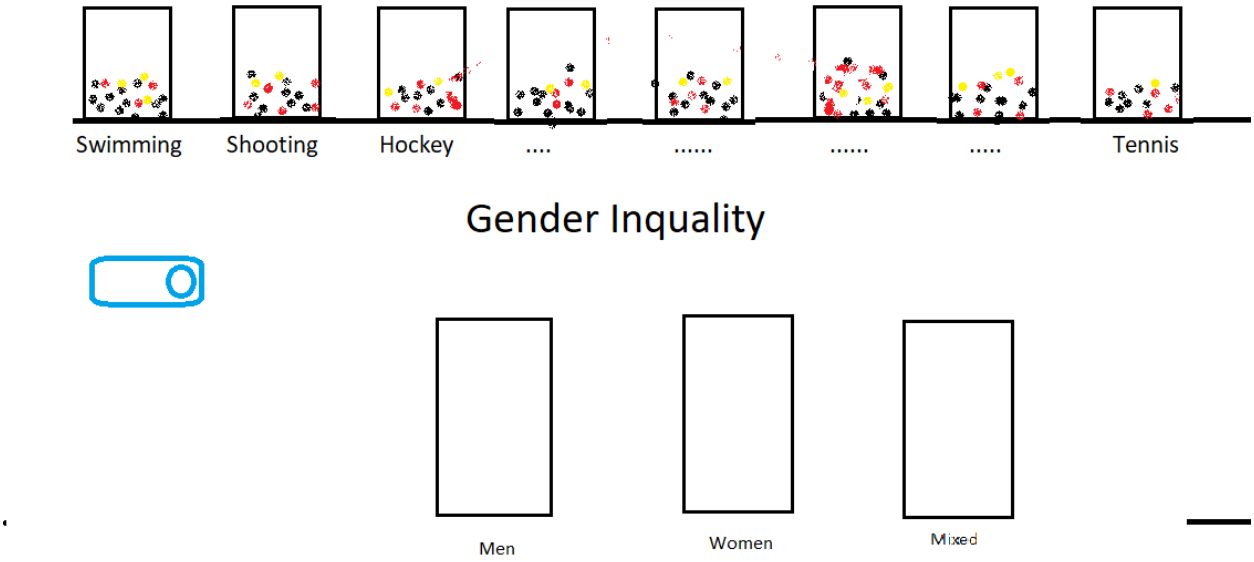
Gender Inequality

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.

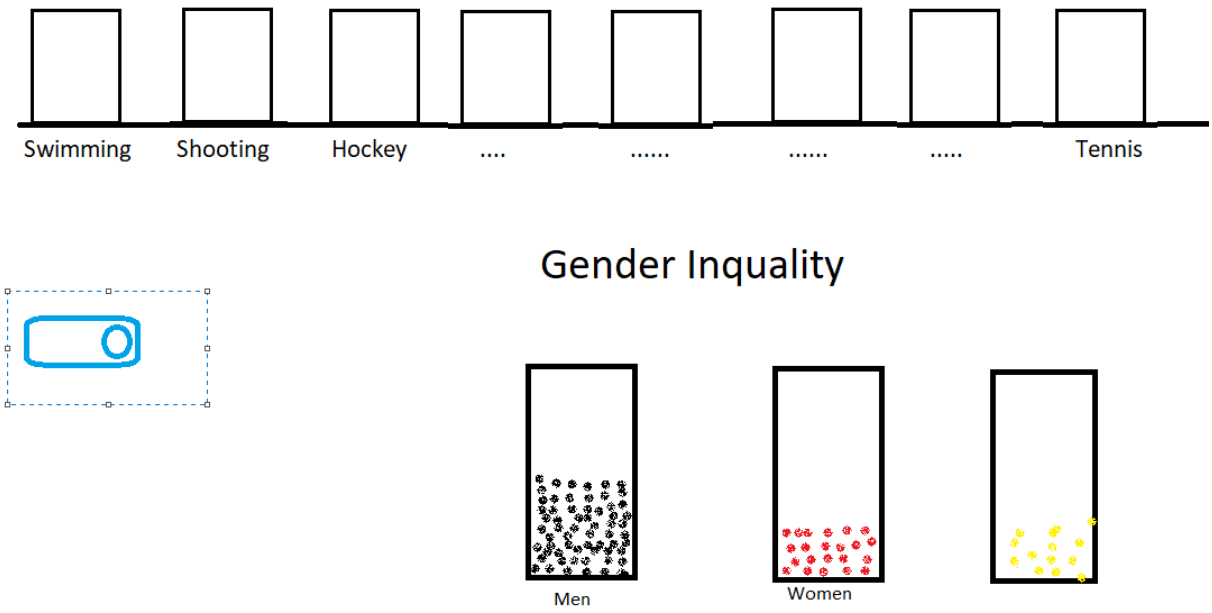
There is a toggle switch provided at the top, which when pressed all these athletes are sorted into three different bins based on their gender. This shows which shows the inequality between different genders.

These bins are synchronized with the year slider, showing data related to particular year.

Categorical Data:

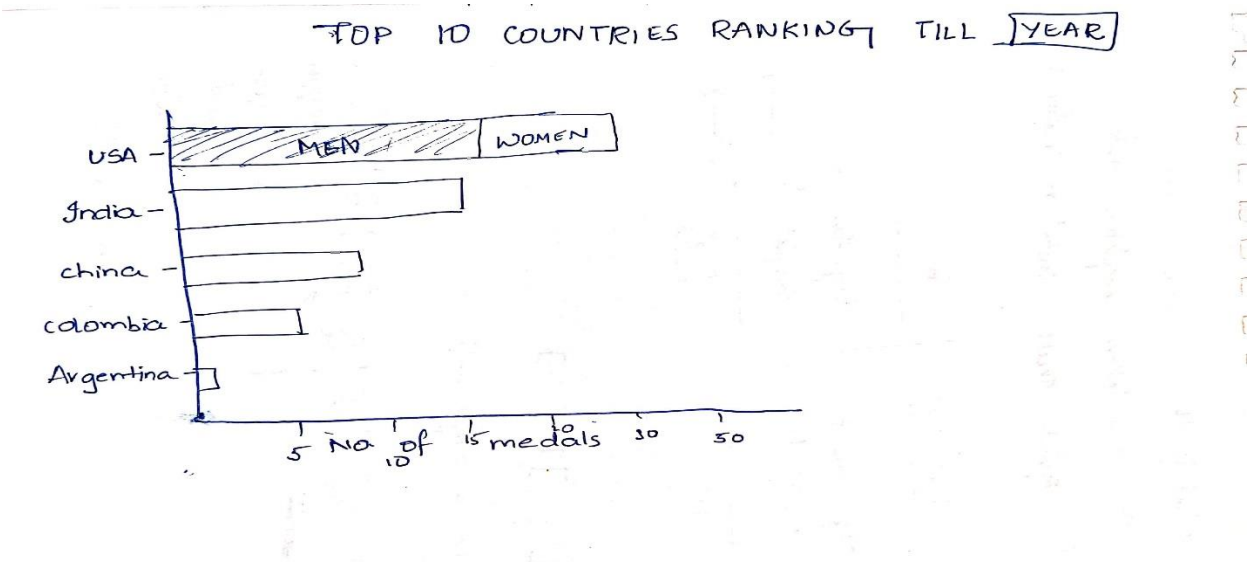


Aggregated Data



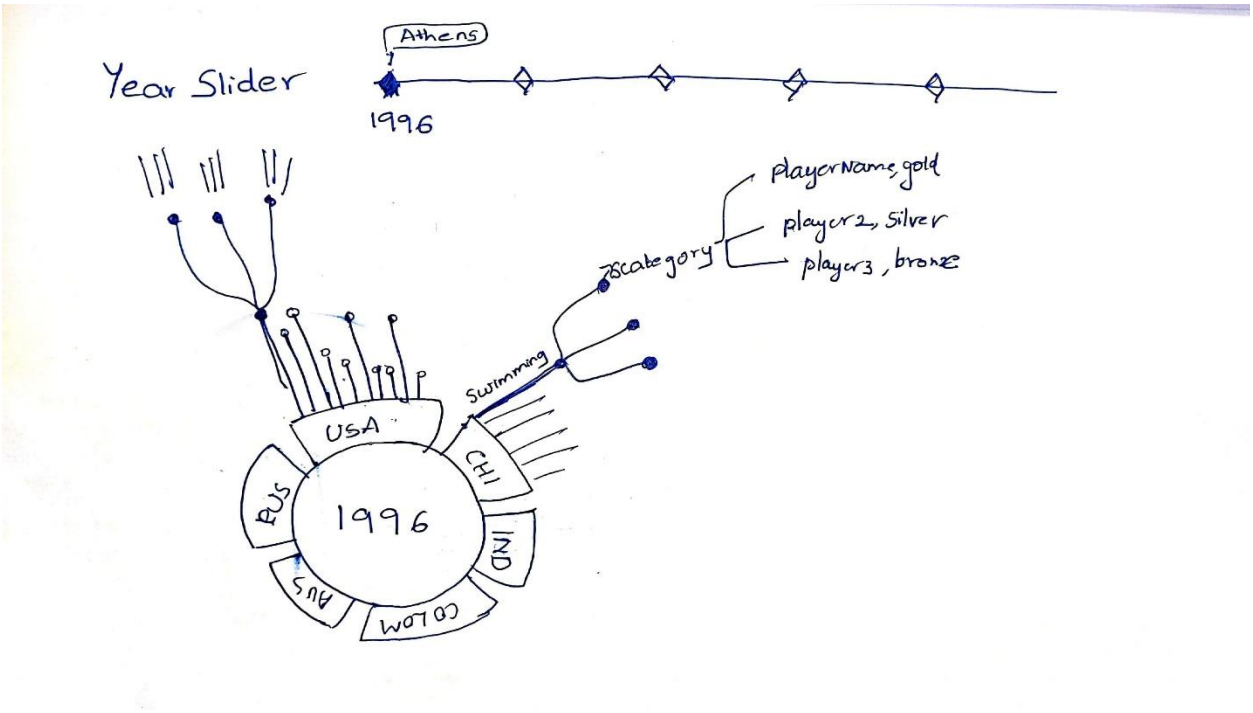
Bar Graph

This bar graph shows, ranking of different countries based on 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 country name and X – axis represent 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.



Optional visualisations:

Histogram



Above histogram is synchronized to year by the year slider, where in 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 top 20 players in the world over last 120 years.

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

Mandatory Features:

1. To show comparison of countries participating and number of medals they won.
2. To compare number of medals in each sport categories.
3. To show the gender inequality in different categories of sports.
4. To show changes in ranking of top performing countries over years.

Optional Features:

1. To show sports subcategory players information.
2. Word map for representing top players.

Project schedule:

Week 1:

1. Research the ideas for visualization project.
2. Gather the data set from online data sources.
3. Ideate on different visualizations for the data set.
4. Sketch the mandatory and optional designs.
5. Create Project Proposal document

Week 2:

1. Data preprocessing using python.
2. Develop code to read the preprocessed data.
3. Optimize on the designs and decide the final 3 or 4 designs after peer review.

Week 3:

1. Design visualization with selected design 1(Scatter plot).
2. Design Table Chart and include sorting feature

Week 4:

1. Design Bin chart for categorical data.
2. Design Ranking bar chart for the top 10 countries.

Week 5:

1. Work on the project website and other modifications required.
2. Add story to the visualizations.
3. Final Project Due