

Olympic Data Visualization

Name: Danish Bhatkar

Email ID: dbhatka@g.clemson.edu, Clemson ID: C15563212

Name: Gaurav Patel

Email ID: gauravp@g.clemson.edu, Clemson ID: C19407183

Name: Shubham Narale

Email ID: snarale@g.clemson.edu, Clemson ID: C33915920

[Repository Link](#)

| [Website Link](#)

| [Video Demo](#)

Table of Contents

Sr. No.	Contents
1	Overview Motivation
2	Related Works
3	Questions
4	Data
5	Exploratory Data Analysis
6	Design Evolution
7	Implementation
8	Evaluation
9	Future Work

Overview & Motivation

The goal of the Olympic data visualization project is to create detailed, interactive representations that allow users to explore and communicate insights from past Olympic data. The project is introduced, along with its goals, objectives, and data sources, in this overview. The project's primary focus is the Olympic Games, a choice made because of the event's lengthy history and widespread influence. It provides a singular chance to explore a broad dataset and find patterns and trends across years, nations, sports, and events. The team combines a variety of viewpoints and talents because its members have varying interests in sports analytics and data visualization. All resources, including code, data, and documentation, are centrally located in the project repository, which is referenced throughout the page.

The main goal is to provide answers to important concerns regarding the Olympics, such as how medal distribution trends have changed over time, how various sports have changed over time, and how national performances have changed. There will be at least three main questions answered, each with its own unique graphics. The information comes from two CSV files: one lists the names and codes of the participating nations and the other lists the events, sports, and medal totals from the Summer Olympics. These files offer a thorough history record that is necessary for the study. The project is organized, with team members assigned specific tasks and weekly deadlines. This method prevents last-minute rushing and guarantees a balanced workload and timely completion. The paper also includes a list of characteristics that are not necessary but are nevertheless important for the project's success.

In conclusion, the goal of this Olympic data visualization project is to provide perceptive, captivating, and interactive representations that present fresh angles on the historical information from the Olympic Games. The initiative aims to create an engaging account of Olympic history by combining creative design with thorough data analysis.

Related Works

A number of creative methods have been used in the field of Olympic data visualization in an effort to clarify the complex nature of the Games. Prominent research frequently focuses on the regional distribution of medal accomplishments, shifts in athletic disciplines, and trends in athlete performance. These projects make use of a variety of statistics, including past performance, athlete demographics, and economic indices of the participating nations.

Utilizing interactive dashboards to let consumers examine data from various Olympic editions and sports is a recurring subject. To illustrate the global distribution of competitors and medalists, for example, some studies use geographical visualizations, while many others use temporal analysis to trace the history of the Olympics. Another well-liked technique for examining the links between nations based on their success and participation in the Olympics is network analysis. These connected publications offer insightful perspectives and useful tools that help guide and motivate our endeavor, especially with regard to data management, storytelling, and visualization strategies.

Questions

The Olympic dataset is a huge data collection that includes a variety of athlete qualities, the types of events that are held, the medals that different competitors have won, the countries of the competitors, and the years that these events were held. There is a lot of opportunity for data extraction and analysis with this dataset. A crucial first step in developing a dashboard based on this dataset was determining who our main target audience would be. It became evident that a broad range of customers, including those who might not be very knowledgeable about the Olympics, would be served by the dashboard. As such, we deliberately avoided doing an analysis that was unduly intricate or sophisticated.

Our aim was to develop a dashboard that could be accessed and comprehended by those with varying levels of Olympic experience. Without providing too much information to the consumers, we tried to make the data visually appealing and educational. This strategy was founded on the knowledge that, although the dataset provided a wealth of information, simplicity and clarity were essential for clear and concise communication. We wanted to make sure that our dashboard would be a useful tool for a wide range of people, from infrequent watchers to die-hard Olympic supporters, offering them clear and fascinating insights and information.

We focused on broader features of the Olympic dataset that would be interesting to a wide range of users while creating our dashboard. This involved looking at the patterns of performance of different nations, figuring out which countries dominate particular sports, and comparing the results of various nations. Our intention was to make this material easily understandable and entertaining for people with different degrees of Olympic experience.

To do this, we included four essential visual components—each intended to answer a distinct question—in the dashboard:

- 1) Performance Trends of a Country: This graphic aids in monitoring a nation's progress throughout time. It lets viewers examine how competitors from a country performed at previous Olympic Games, emphasizing any gains, losses, or consistency in their performance.
- 2) A Comparative Evaluation of Performances of Two or more Nationalities: This function allows us to compare the performances of several nations side by side. It's an effective technique for comparing national performance at the Olympics and for learning about each country's advantages and disadvantages.
- 3) Most Popular Sport in a Country: The goal of this visualization is to show which sports are the most popular or successful in a certain nation. It provides an insight into other countries' sports cultures by highlighting the areas in which their athletic prowess and passions are most concentrated.
- 4) Distribution of Medals Earned in Various Event/Sport Subcategories: This dashboard element explores the details of medal distribution in different sports and their sub-disciplines. It offers a thorough understanding of a nation's strong points and highlights their hegemonic domains within the Olympic system.

Regardless of the user's past experience or level of competence in Olympic sports, each of these graphics is essential to make the huge and complicated data of the Olympics comprehensible and entertaining for everybody. The dashboard is designed to be both user-friendly and educational, making it suitable for both casual viewers looking for a brief synopsis and more committed aficionados who want to delve further into the patterns and trends in Olympic history.

Data

Comprehensive details about every athlete that competed in the Olympic Games are included in the primary dataset file. It contains information on the athlete's name, gender, age, height and weight, the team they competed for, their National Olympic Committee (NOC) code, the year, season, and host city of the particular Olympic Games they attended, the sport and event they competed in, and any medals they may have won. This dossier provides a thorough

understanding of the athletes' backgrounds and accomplishments across a range of Olympic sports. The file also focuses on a more aggregate level, describing how different nations performed in different Olympic Games sporting disciplines and events. It gives an annual analysis of how many gold, silver, and bronze medals each nation has earned across all sports and disciplines.

This information is essential for examining a nation's historical Olympic results. The main file is connected by the dictionary.csv file. It provides a mapping of national names to Olympic country codes (NOC codes). Correlating the data concerning specific Olympic athletes requires this mapping.csv containing the data from the combined medal tally. A thorough analysis, including identifying which nations perform best in which sports, following national performance trends over time, and comprehending the role played by specific athletes in their nation's overall Olympic success, can be carried out by connecting the NOC code of athletes with the country names and codes in the dictionary. When combined, these data provide a thorough picture of the Olympics, including details on the accomplishments of specific athletes, national statistics, and historical trends in Olympic sports. This extensive data collection is a priceless resource for sports analysts, historians, and enthusiasts interested in exploring the many dimensions of Olympic history and performance.

Data Cleaning: We started by cleaning the Olympic dataset because we were having some issues getting it ready for visualization. This required removing several properties that we decided weren't needed for our visualization goals. After that, we started preprocessing the dataset. In order to accomplish the objectives of our project, we had to remove entries with null values at this phase in order to solve the problem of missing values in our dataset. At first, we thought about manipulating data with d3, but we discovered it was difficult. Therefore, in order to better meet our analysis objectives, we decided to edit the data in Excel. There, we included information on the number of casualties on particular days of the week and the number of cars for each day.

Exploratory Data Analysis

When we first started our Olympic data project, we had trouble deciding which dataset would be best for our study. Our investigation comprised a thorough examination of many Olympic datasets spanning a broad time period from 1896 to 2012. We had to go through a lot of datasets for this procedure, each of which offered a unique viewpoint and set of facts about the Olympic

Games, such as medal totals, participation, athlete profiles, and event-specific information. We examined and selected the dataset that most closely matched the goals of our project after giving it some thought. The selected dataset included an extensive collection of Olympic data, including specifics on the medal counts for each nation, athletes' accomplishments, and the range of sports and disciplines featured at the Olympic Games between 1896 and 2012. We were especially drawn to this dataset because of its thorough coverage and depth, which gave us a solid basis for our research and visualization activities.

Initially, we intended to present the entire medal count separated into three sections—Gold, Silver, and Bronze—using a stacked bar chart in our visualizations for the Olympic Games dataset. But as our investigation developed, we changed course and decided to use a line chart in its place. This choice was made after it became clear that a line chart would better illustrate the patterns and trends in medal distribution over time, particularly when combined with a time slider that included all of the years included in the dataset. In our case, the line chart's benefit was its capacity to clearly and continuously illustrate the dynamics of Olympic medal victories over time.

We improved the visualization's interactivity by adding a time slider, which lets visitors zoom in on particular years and see how the number of medals changed over time. This method worked better for our objective of examining and showcasing the development of Olympic accomplishments across time. The first step in interacting with the data was to use the global map. On the globe, users may pick individual nations or a set of up to four nations. The two visualizations, the line chart and the bubble chart, would then dynamically change to reflect the data relating to the selected nations as a result of this selection. Users were able to customize how they explored the dataset due to the interactive aspect of the globe map, which produced a more in-depth and individualized examination of Olympic results.

We also created a bubble chart that displayed every event in which the chosen nations took part. Each bubble's size corresponded to the overall number of medals won in that particular event, providing a clear visual depiction of the nations' prowess and accomplishments across a range of sports. This bubble chart offered a distinctive interpretation of the data, visually strikingly illustrating the range of occurrences and the magnitude of accomplishments. These interactive components—the bubble chart, line chart, and global map—combined to produce a thorough and interesting visualization tool. This tool revealed the rich history and achievements of the participating nations in the Olympic Games in addition to successfully presenting the Olympic data and enabling users to engage with and study the dataset in a more meaningful way.

Design and Evolution

To be assured, our original intention was to examine athlete participation by gender using a line chart. Our goal was to learn more about whether athletic involvement rates for men and women were similar. Since a line chart is a good tool for displaying trends and comparisons over time, we expected it to provide us a clear picture of the patterns of gender involvement over time. However, inconsistent data presented us with some difficulties. These problems may arise from a number of sources, such as incomplete or inaccurate data collection, missing numbers, or anomalies in data reporting guidelines. The goal of producing an accurately represented and readily interpreted graphic of the data was defeated by these discrepancies. We decided to change course responsibly after realizing this and giving up on our original plan.

The significance of data integrity in our visualization efforts was highlighted by our experience. Even the most well-designed visualizations can become unreliable when data quality is degraded. In order to guarantee data accuracy and consistency, we came to understand the necessity of more stringent data cleaning and validation procedures. To avoid such disappointments, we will concentrate on improving these areas in subsequent initiatives. In light of the difficulties we encountered, we may investigate different strategies in next initiatives. These might entail updating our data gathering procedure or putting in place more reliable data verification techniques. We may also think about utilizing data imputation techniques to approximate missing or inaccurate values if the discrepancies in our data are not too great. This, however, will rely on the kind and severity of the data problems.

Another goal of our project was to make a stacked bar chart that would show the number of medals won by different teams or nations in a competition. The X-axis would be used to arrange the bars according to the names of the teams or countries, while the Y-axis would show the overall number of medals won. The three sections of each bar in the graphic were intended to symbolize gold, silver, and bronze medals, respectively. The goal of this design was to present a clear and comparable picture of the medal distribution among the competing teams or nations. We successfully and flawlessly produced the stacked bar chart after putting this visualization into practice. But we found a serious problem with the portrayal. The silver section took up the majority of the bars, suggesting that there were more silver medals than gold and bronze. A lopsided visual depiction resulted from this uneven medal distribution.

It was difficult to visually evaluate and contrast the numbers of gold and bronze medals won by various teams or nations due to the dominance of the silver category. The overabundance of silver medals on the graphic obscured the other parts, which might cause confusion or a lack

of knowledge of the overall distribution of medals. This problem brought to light a weakness in our visualization strategy. When one category in a stacked bar chart has a large advantage over the others, it can adversely affect the visual balance of the chart and make it harder to understand the smaller categories. We may need to reevaluate our visualization approach in order to resolve this problem. Using a different kind of graphic, like a clustered bar chart or a normalized stacked bar chart, which can better handle notable data variances, is one strategy. As an alternative, we may look at new methods to partition the data, including concentrating on a smaller group of nations or teams or classifying the medals differently in order to provide a more visually balanced depiction.

We changed our strategy and chose to use a line chart to show the overall number of medals in our final dashboard considering the difficulties we encountered with the stacked bar chart. The total number of medals won was plotted on the Y-axis and the years on the X-axis in this line graph's design. Every nation we chose from our dashboard's global map was symbolized by a distinct colored line. This design decision made it possible to visualize the data clearly and effectively, which was essential for achieving our goals. The line chart offered our analysis a number of significant benefits.

First of all, it made the depiction of trends over time easy to understand. We may readily see how a nation's performance changed over time in terms of medal acquisition by charting the overall number of medals against the years. This temporal viewpoint proved especially useful for examining long-term patterns and spotting notable changes in performance. Furthermore, the chart's comparing skills were significantly improved by the use of distinct colored lines for each nation. Users may quickly assess how different nations' performance patterns compare to one another over time by looking at how many medals each nation has won. This comparison element allowed us to find leaders, laggards, and new patterns in the data, which was essential for our study.

We discussed the medal distribution problem we were having with the stacked bar chart and came to the conclusion that a bubble chart would be a better visual aid to use. We were able to depict the medal distribution in a way that was both participatory and visually appealing thanks to this creative method. Every bubble in our bubble chart represents a distinct sporting event. The overall number of medals won in each event was closely correlated with the size of these bubbles. It was simple to determine which sporting events were more prevalent in terms of medal winnings because of this design decision, which gave consumers an instant visual clue about the relative importance of each sport in terms of medal tallies. We added an interactive element to the bubble chart as well. An extensive breakdown of the medal distribution for that

particular sporting event would be shown in a tooltip that would appear when a user hovered over a bubble.

The number of gold, silver, and bronze medals won was included in this breakdown. Instantaneous, comprehensive information was provided by this feature without overcrowding the chart, improving the user experience. We improved the functionality of the bubble chart further in recognition of the intricate and multifaceted nature of sporting events, which frequently have several divisions and subcategories. For the bubble chart, we added a 'on-click' function to our JavaScript code. With the use of this feature, visitors could dig down into a primary event's several categories and subcategories by clicking on the bubble that represented it. For users who wished to go further into the data, this interactive feature was essential as it offered insights into the individual performance of nations or players in different facets of each sport. Our bubble chart was a static visual representation before this 'on-click' capability was included, but now it is a dynamic, exploring tool. Users may now explore the subtleties and specifics of each event in addition to viewing the overall medal distribution throughout the several sports. Finding trends and patterns in the athletic events at both the macro and micro levels was made possible by this depth of study.

Implementation

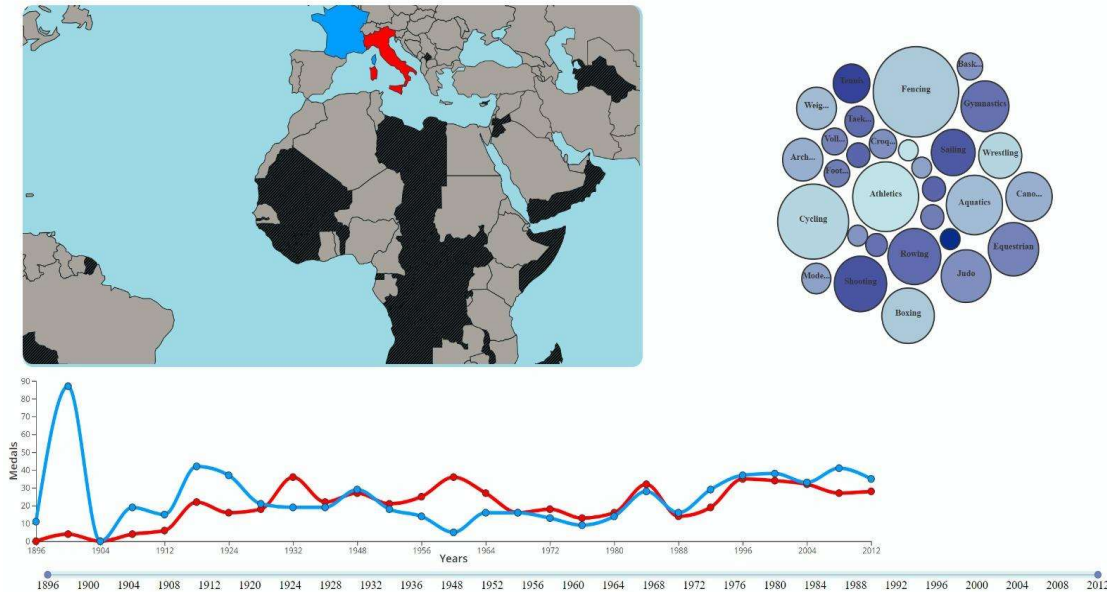
The dashboard's description gives a detailed overview of its features and objectives. Let's go into more detail about each element and how they work together to provide insights:

The global map that is interactive acts as a geographic filter. Users are able to examine the performance of teams by selecting nations. This approach of spatial selection is visually appealing and straightforward, making it simple to compare the results of other nations. One essential element for temporal analysis is the time slider. Users are able to choose a certain time frame, making it possible to analyze patterns over particular years. This may be especially helpful in determining how a team's performance has changed over time or in reaction to certain events (such modifications to policies or enhancements to training).

A traditional method for trend analysis is the line chart, which has the years on the X-axis and the medal count on the Y-axis. It shows the evolution of a team's performance throughout time. An upward trend in the line, for example, would point to better performance, but a downward trend might point to a decrease. Another level of analysis is provided by the bubble chart, which

may include additional variables such as the team size, the kind of medals won, or other pertinent elements. This kind of graphic is helpful for quickly and easily understanding complicated datasets using visual means.

Countries: **Italy** and **France** Event: **All** Year: **1896 to 2012**



The bubble chart on the dashboard makes it easier to analyze sports data at a more granular level by providing an intricate, dynamic representation of how particular nations performed in different competitions. Let's examine its attributes and the information it may offer in more detail: The bubble chart shows the individual competitions in which the chosen nations have taken home medals. This degree of specificity is essential for comprehending both overall performance and area-specific excellence. Each bubble's size corresponds to the quantity of medals won in that particular competition.

It is simple to determine which events a nation excels in thanks to this graphic depiction. Bigger bubbles quickly highlight instances in which a nation has excelled, highlighting its strong points. Using the global map and time slider, the bubble chart dynamically modified according to the chosen nation and chronology. This feature guarantees that, regardless of whether the user is comparing numerous nations or examining a single country's performance over time, the data provided is always pertinent to their current focus.

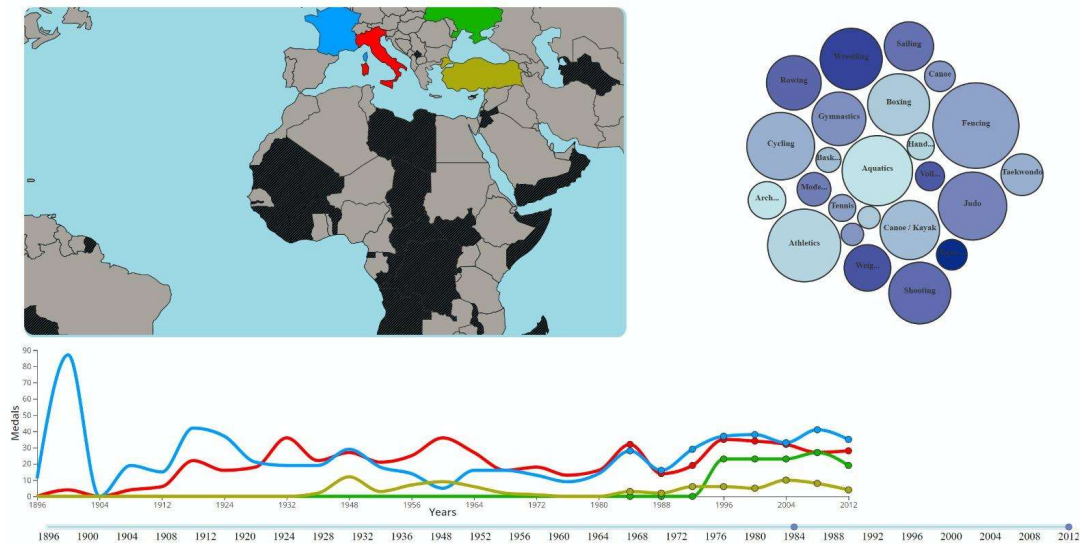
A sophisticated analysis is made possible by the capacity to drill down into certain categories within larger events (such as separate races in athletics). This can reveal patterns and trends within certain disciplines as well as at the sport level. A nation may excel in short-distance track competitions but not in long-distance ones, for instance.

The image is a composite of three visualizations related to the 2012 London Olympics:

- World Map:** A map of the world with Italy highlighted in red and the United States highlighted in blue.
- Bubble Chart:** A chart showing the size of the Olympic rings and the number of medals won by each country. The bubbles are labeled with the number of medals won: 5000 Walk, 50KM Walk, 20KM Walk, Shot..., Long..., High Jump, 4X100M Relay, 5000M, 800M, 400M, 200M, Discus Throw, Trip..., 110M, 300M, 100M, Pole Vault, 4X100M, and 5000M. The size of the bubble represents the number of medals won.
- Line Graph:** A line graph showing the number of medals won by the United States (blue line) and Italy (red line) from 1896 to 2012. The x-axis is labeled 'Years' and the y-axis is labeled 'Medals'.

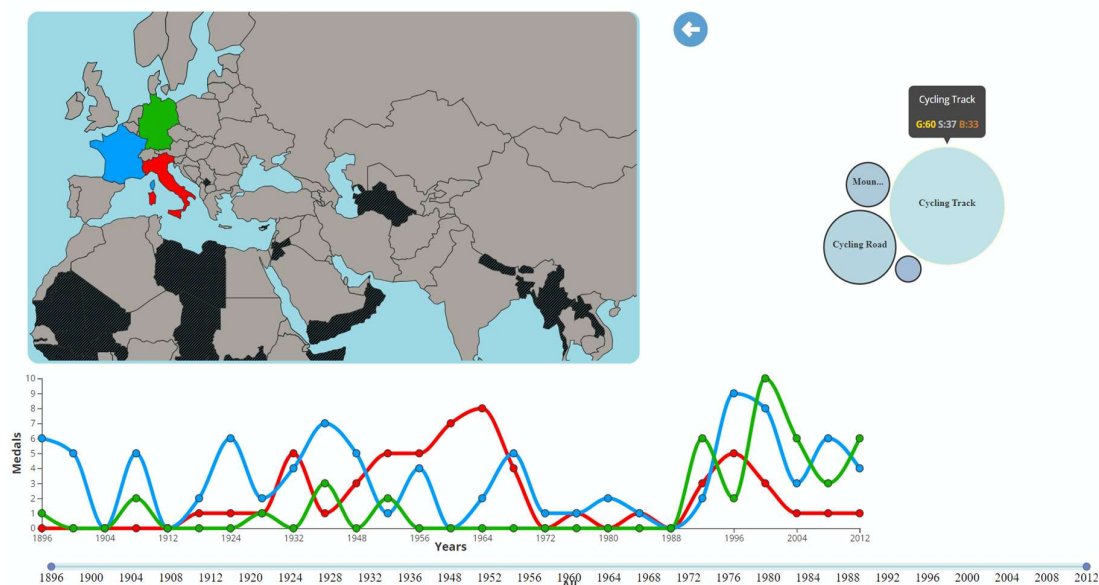
The data shown on the bubble chart is relevant to the nations and time period selected because of the synchronization between the global map, time slider, and bubble chart. A consistent and flawless user experience is offered by this integration, where modifications made to one component instantly affect other components. The dashboard facilitates targeted analysis by just showcasing data for the designated countries and time frame. Trends, patterns, and anomalies unique to the selected parameters may be observed by users without being distracted by extraneous data.

Countries: **Italy, France, Ukraine and Turkey** Event: **All** Year: **1984 to 2012**



Additionally we have provided a few features in the bubble chart to dig deeper and analyze the various categories and subcategories of a sport event. We can click on any of the bubbles/events that display the further categorization of that specific event. In the image given below, after clicking on the Cycling event, we can see a further breakdown of the events such as “Cycling Road” and “Cycling Track”. These bubbles provide more detailed information about the medals won by the selected countries.

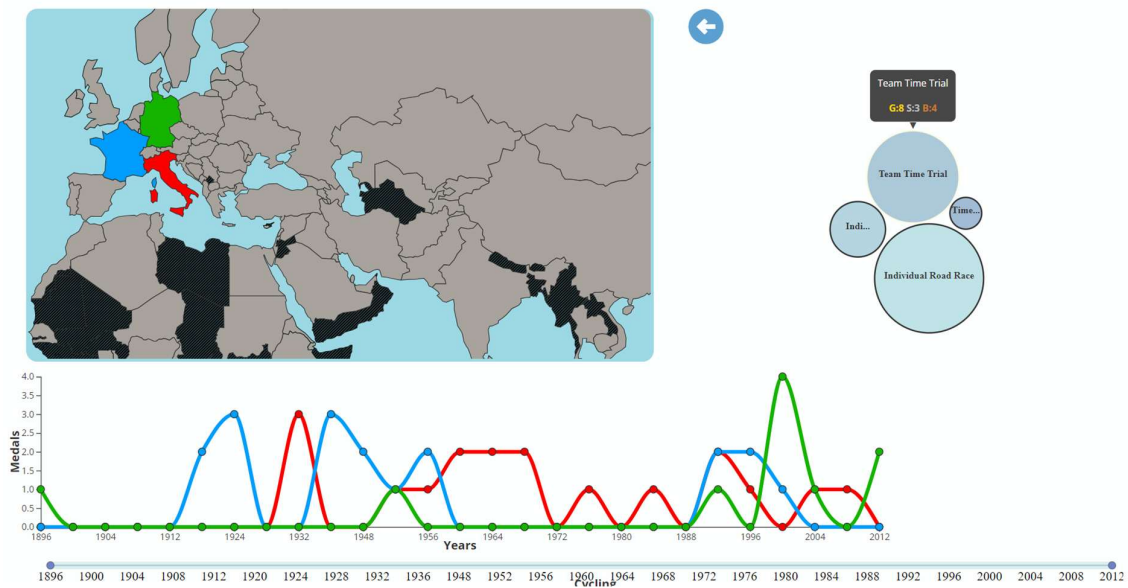
Countries: **Italy, France and Germany** Event: **Cycling** Years: **1896 to 2012**



Countries: **Italy, France and Germany**

Event: **Cycling Road**

Years: **1896 to 2012**

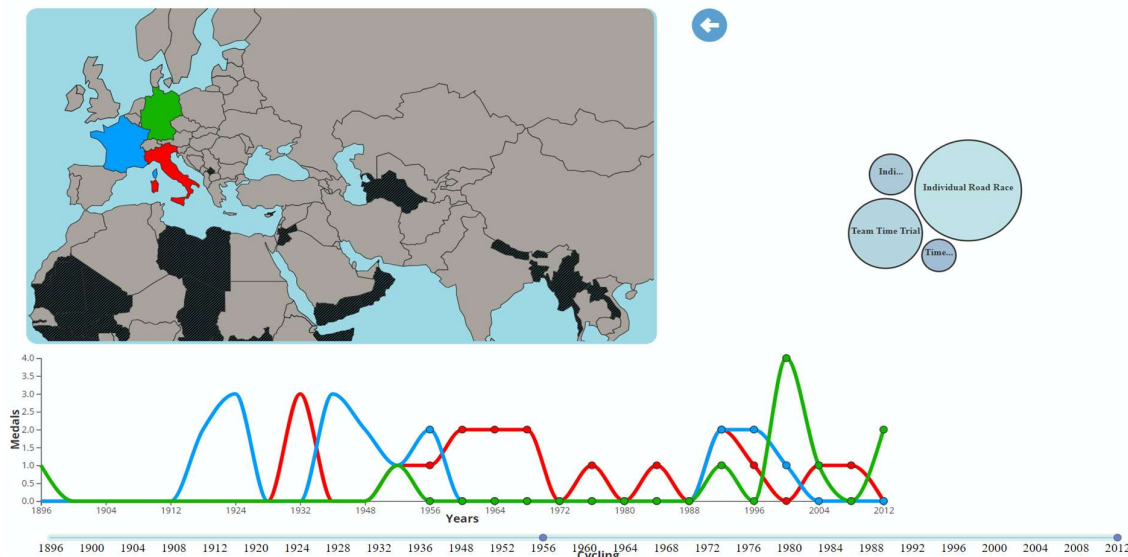


Apart from the total medal count, the bubble chart also provides information on the distribution of medals won, i.e., the number of gold, silver, and bronze medals won. This can be done by simply hovering over a bubble.

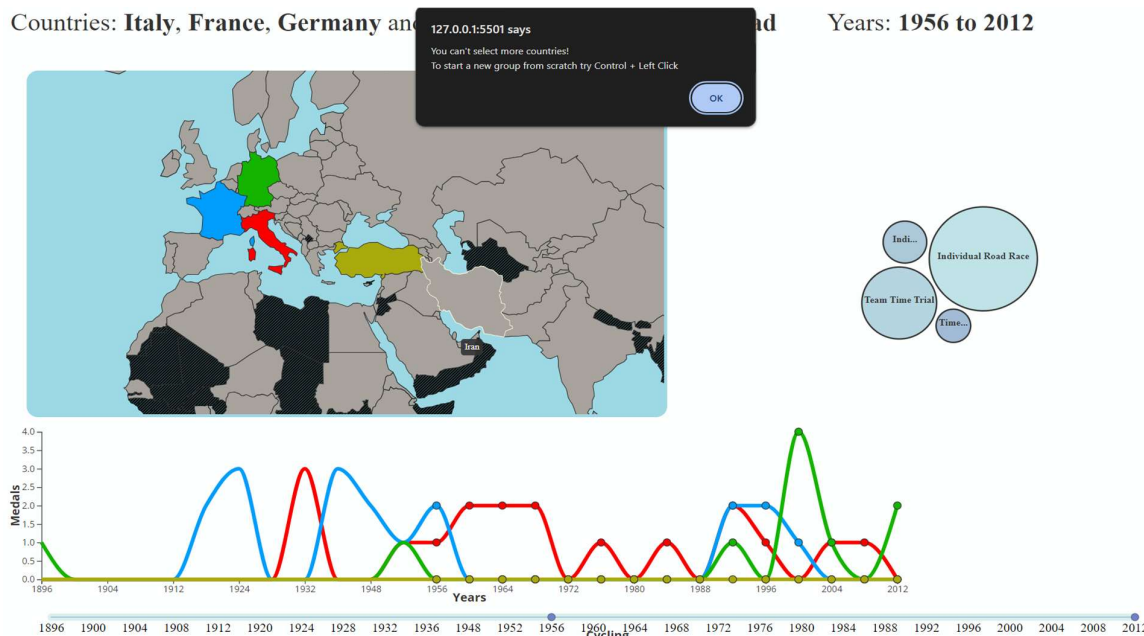
Countries: **Italy, France and Germany**

Event: **Cycling Road**

Years: **1956 to 2012**



As mentioned earlier, the world map allows us to select a maximum of 4 nations only. This feature is added to avoid clustering of multiple lines at the bottom and to keep the bubble chart clean by avoiding too many bubbles. On attempting to select a 5th nation, we get a pop-up saying that we can't select more countries as given in the below image.



Evaluation

The visualization's primary objective is to highlight nations that excel at the Olympics. This would be accomplished by examining a depiction of the globe map in which the nations are emphasized according to the total number of medals (Gold, Silver, and Bronze) they have won between 1896 and 2012. The graphic possesses a feature that allows users to compare the different medal categories (Gold, Silver, and Bronze) obtained in other nations or eras to provide even more depth. This might highlight intriguing patterns, such as a nation's predominance in particular sports or eras. Users could compare Olympic results for up to four different nations.

This function made it possible to compare the chosen countries directly and intelligently, showcasing their advantages and advancements in the Olympic Games across time. The tool totaled the medals (Gold, Silver, and Bronze) that the chosen nations had earned overall. This summary made it easy to grasp these nations' relative sports proficiency and provided a clear perspective on their total Olympic accomplishments.

The line graph does a good job of showing how each team's Olympic success has changed over time. A graphic narrative of each team's journey was created by charting medal counts or other performance measures against time, showing their highs and lows in terms of Olympic accomplishment. This tool was especially helpful for side-by-side performance comparisons between several teams. It was simple for users to see how different Olympic teams performed against one another. This comparison provided insights on performance decline, improvement, and consistency over time in addition to being quantitative (in terms of medal counts). The US team's supremacy was effectively displayed by the line chart. The United States' persistent success and dominance in the Olympics were clearly shown by the graph line that showed their tendency to stay above those of other nations. It was simple to recognize the United States as one of the most successful Olympic teams in history because of this feature of the depiction.

The visualization tool's bubble chart provided a comprehensive overview of the popularity of different sports in various countries, with an emphasis on high-performing nations like the US, Russia, China, and Germany. An explanation of its attributes and capabilities is provided below: The bubble chart first showed all the sports in which the selected country had medaled after one or more selections. Users were able to rapidly determine which sports were the most successful or well-liked in the chosen countries by using this summary. Sports like athletics and aquatics were shown to frequently top the charts for top-performing nations, demonstrating their broad appeal and achievement in these fields.

The chart was made to facilitate more investigation. After selecting a specific sport, such as athletics, the chart separated to display several subdisciplines within that activity. This function gave users the ability to go further into particular topics of interest and gave a more detailed picture of a nation's abilities in various sports. Taking the user experience a notch higher, the bubble chart showed the various events that fell within the chosen sport discipline once the user made their selection. For instance, if we choose Athletics for Italy, we may see events such as the 800 m, long jump, and 50 km walk.

This degree of specificity highlighted not just the sports a nation excels in but also the individual events within those sports, providing an even more detailed picture of that nation's Olympic accomplishments. The quantity of medals won in each sport or event, as well as the kinds of medals (Gold, Silver, Bronze), might have been represented by the size and color of the bubbles in the chart. An instant understanding of the relative importance and success of each sport or event for the chosen countries was given by this visual hint.

Future Work

These recommendations are intended to improve the interactive dashboard's capacity to communicate information, especially when it comes to monitoring team performance in sports or other comparable activities over an extended period. Let's dissect each recommendation:

Including a Visual Display for a Team's Win Total: This graphic attempts to highlight the top teams from various eras. It would enhance current data visualizations by emphasizing wins, which are a crucial sign of a team's performance. Depending on the type of data and the intended level of user participation, the visual might be either a bar chart or a line graph. It should be able to adapt to a time slider so that users may see how the performance of the team varies over certain periods of time. This update would improve the dashboard's usefulness for historical analysis and comparison by giving users a clear, concise picture of which teams have performed well during particular times.

Changing the Bubble Chart's Circles to Pie Charts: In situations where more than one country is chosen, the existing bubble chart may not be able to accurately show how medals are distributed among various teams in a particular competition. The dashboard may more clearly show how each country's performance compares within a single event by converting the circles that represent each team into pie charts. A country would be represented by each segment of the pie chart, with the size of the section corresponding to the quantity of medals or victories. This would make it easier for people to comprehend how competitive an event is, particularly when comparing teams or nations.

X-axis transitioning on the line chart: The goal is to improve the line chart's relevance and interaction, particularly in reaction to modifications made to the chosen time frame. In response to the time slider, the line chart's X-axis should dynamically modify its scale and domain. When a user chooses a time range between 2000 and 2020, for instance, the line chart needs to be updated to only show data from this time frame. With this feature, the line chart would respond better and be more specific to the user's current topic of interest. It increases the dashboard's usefulness and accuracy in data representation by guaranteeing that the data presented is always pertinent to the time period being evaluated.

All in all, these improvements aim to make the dashboard more accurate, dynamic, and responsive when presenting data, making it a more complete and approachable instrument for evaluating team performances over time.