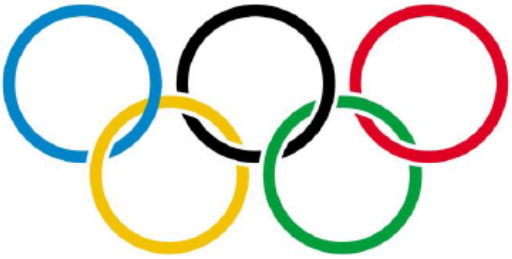
Olympics Games Analysis Dashboard



DATA*6200 F22 - Data Manipulation and Visualization

Created by Mahesh Balasaheb Raut
University of Guelph Id: 1227509

Table of Contents

I Abstract

The dashboard is to perform analysis on the Olympic dataset provided by Kaggle. The Kaggle dataset for Olympic events from 1896 to 2016 shows an incredible amount of research for recording decades of study into an accessible format. The dashboard provides visualizations of data on the Olympic Games, the international sporting event held every four years between countries. The winter and summer Olympic games were played in the same year until 1992. They have divided two years apart after that. 1912–20 and 1936–1948, corresponding to World Wars I and II, were two lengthy periods without any Games. Various features from the dataset were looked at while designing the dashboard. The growth of the games presents ideas on how the games have grown in numbers for athletes, events, and nations over time since its inception. The sports leaderboard by medals shows the total count of medals by Sport and Gender for the top 10 countries by Gold, Silver, and Bronze category. The medal tally contributes to the medal results table for a selected year won by the countries. The map view of the dashboard reveals total medal winnings on a map by the country for all years from 1896 to 2016.

II Introduction

Greece's Athens hosted the first modern Olympics in 1896. The Games from Athens 1896 to Rio 2016 are collectively referred to as the "modern Olympics." Baron Pierre de Coubertin presented the idea in 1894.

This dashboard offers an easy-to-use visualization analysis of the Olympic games from 1896 to 2016 that takes into account different system levels. The Olympic analysis dashboard is to convey complicated data accurately, and appropriately. The Olympic analysis dashboard was made considering various system levels that include defining the tasks and data in terms of the issue. Then the outline of data into data kinds and operations is done.

The dashboard's main objective is to describe how the user could profit from the produced system having visual representations. User interfaces created for this dashboard are straightforward and easy to use while educating people about the Olympics.

Questions targeted in the dashboard

- How has the participation of countries changed over time?
- How many athletes participated over time?
- How many events are contested over time?
- Which nations dominate each event category?
- What are the total medals won by each country at every Olympic event?
- What nations won the most medals since the beginning of the Olympics?

Data Characteristics

- Year-wise data
- Season(Winter and Summer Olympics data)
- Sports level data
- Athletes level data

III Dataset

A dashboard was made using two distinct data frames. After investigating the dataset, it was determined that no data should be removed. Some features such as athlete's age, weight, and height were not executed in the shiny dashboard but can be looked at in future work. Also, the idea of a host city for the Olympic Games was decided to show on the map but was dropped because the incorrect host city for the chosen year was given in the feature "City".

The NOC represents National Olympic Committee which is mapped to the region i.e. country as shown in the noc_regions dataset below. NOC is used while developing a medal tally as it correctly represents the country of an athlete. Features "NOC" and "region" in the noc_regions dataset are mapped to their respective NOC in the athlete events dataset.

1. Athlete Events data

```
## Observations: 271,116
## Variables: 15
## $ ID     <char> "1", "2", "3", "4", "5", "5", "5", "5", "5", ...
## $ Name   <char> "A Dijiang", "A Lamusi", "Gunnar Nielsen Aaby", ...
## $ Sex    <fact> M, M, M, M, F, F, F, F, F, F, M, M, M, M, M, M, ...
## $ Age    <inte> 24, 23, 24, 34, 21, 21, 25, 25, 27, 27, 31, 31, ...
## $ Height <dobl> 180, 170, NA, NA, 185, 185, 185, 185, 185, ...
## $ Weight <dobl> 80, 60, NA, NA, 82, 82, 82, 82, 82, 82, 75, ...
## $ Team   <char> "China", "China", "Denmark", "Denmark/Sweden", ...
## $ NOC    <char> "CHN", "CHN", "DEN", "DEN", "NED", "NED", "NED", ...
## $ Games  <char> "1992 Summer", "2012 Summer", "1920 Summer", ...
## $ Year   <inte> 1992, 2012, 1920, 1900, 1988, 1988, 1992, 1992, ...
## $ Season <fact> Summer, Summer, Summer, Summer, Winter, Winter, ...
## $ City   <char> "Barcelona", "London", "Antwerpen", "Paris", ...
## $ Sport  <char> "Basketball", "Judo", "Football", "Tug-Of-War", ...
## $ Event  <char> "Basketball Men's Basketball", ...
## $ Medal  <fact> NA, NA, NA, Gold, NA, NA, NA, NA, NA, NA, NA, ...
```

2. NOC Regions data

```
## Observations: 230
## Variables: 3
## $ NOC    <char> "AFG", "AHO", "ALB", "ALG", "AND", "ANG", "ANT", ...
## $ region <char> "Afghanistan", "Curacao", "Albania", "Algeria", ...
## $ notes  <char> NA, "Netherlands Antilles", NA, NA, NA, NA, ...
```

IV Design Elements

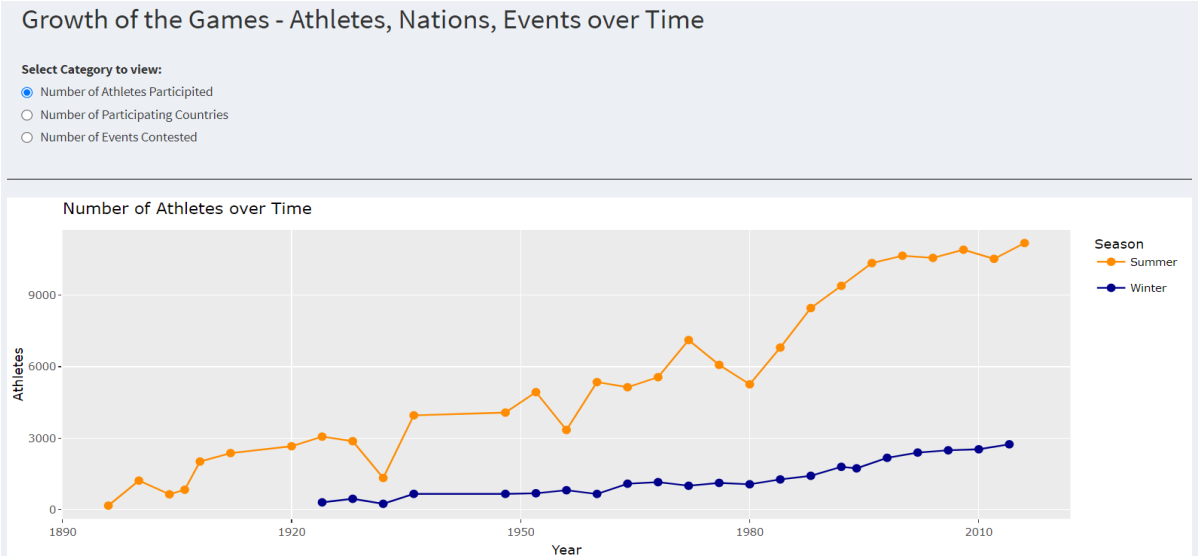
Olympic games are globally played and thus the year revealed from it is the interest of medal by countries and the total medal counts. Another interesting analysis is the participation of athletes and countries, which are represented on a graph. The dashboard sidebar consists of five menu items:

- About
- Growth of the Games
- Sports Leaderboard by Medals
- Medal Tally
- Map View

With more detailed questions, the user-friendly Olympic Shiny dashboard body becomes a trustworthy tool for interactively displaying visual representations. The following four major answers are revealed in the dashboard, and their uses from the dashboard are shown below:

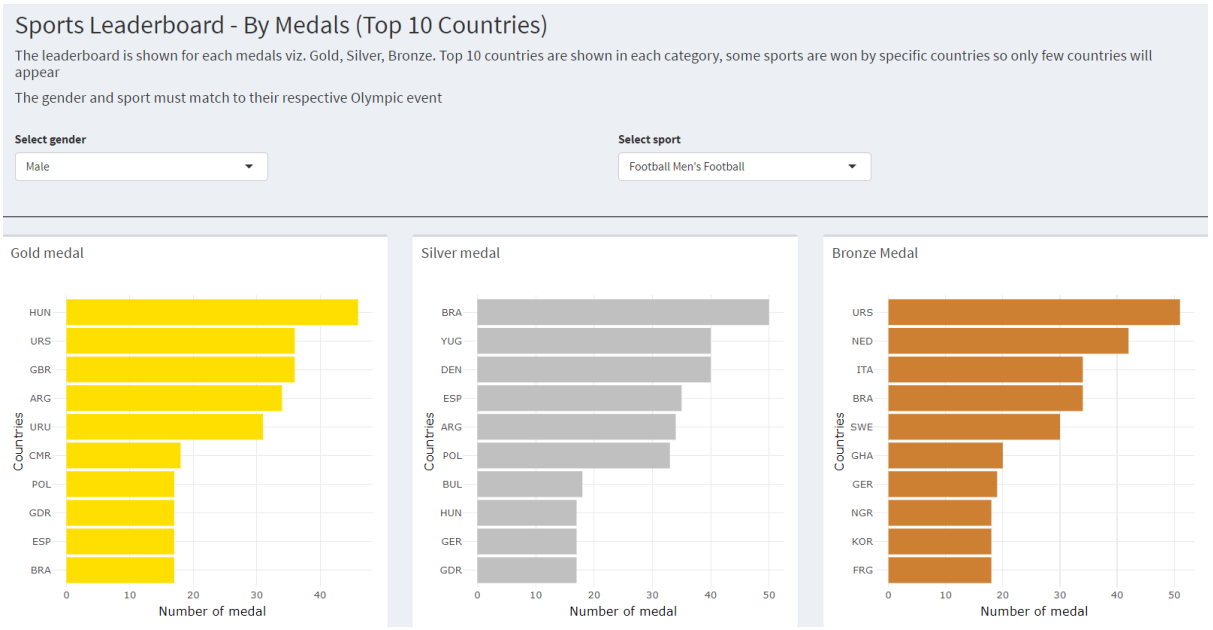
1. Plot showing Growth of the Games over time

The different categories for choosing athletes, nations, and events can be selected by a radio button, and the plots are categorized by Winter and Summer seasons.



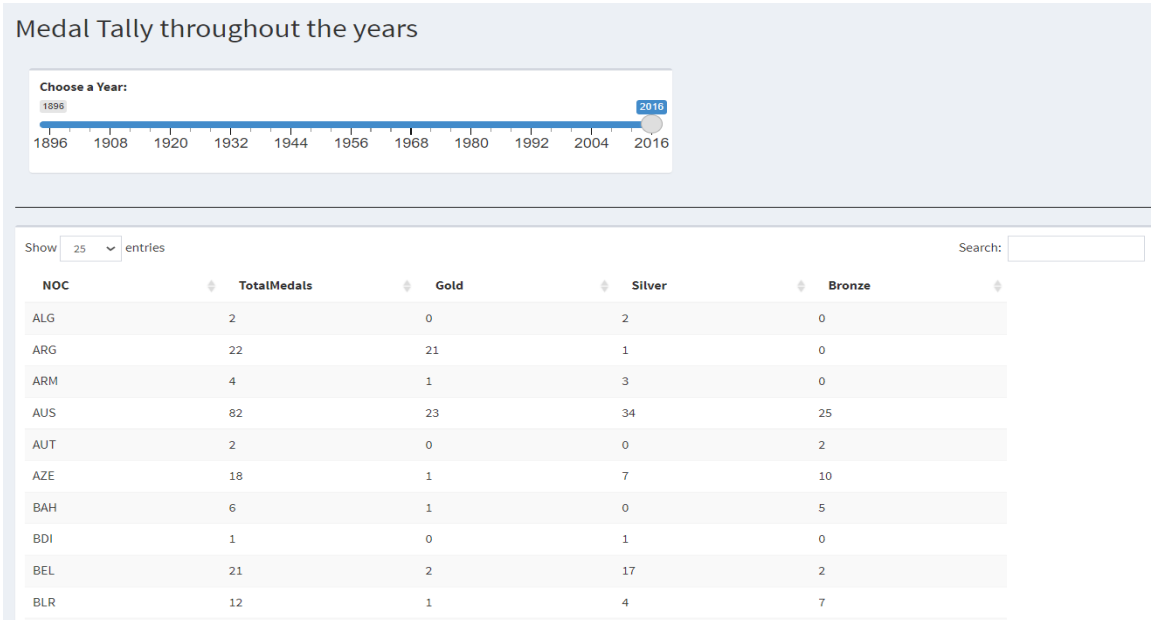
2. Sports Leaderboard by Medals

Two dropdown menus are provided for selecting gender and sport played. The gender and sport must match their respective Olympic event. Three boxes show gold, silver, and bronze medals won by each country in their rankings in that particular event. Only the top 10 countries are shown. Also, if the number of countries is less than 10, then only that many countries will be present in that box.



3. Table for medal tally by years

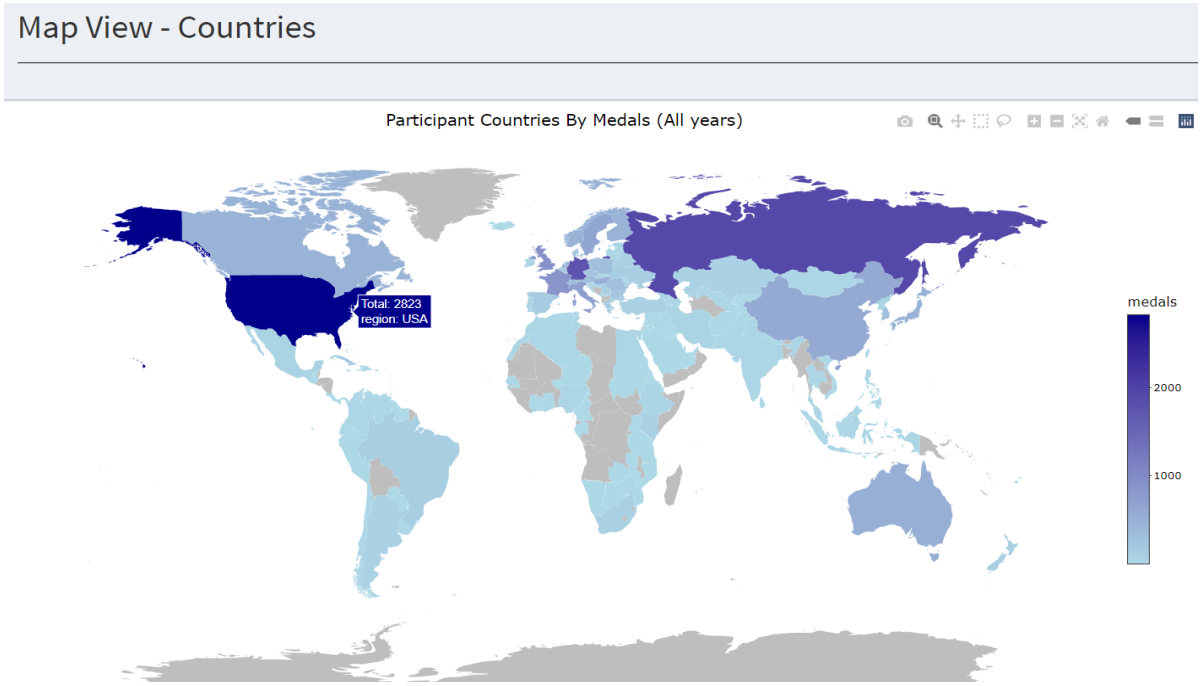
The slider input is used to select the year of the Olympics games. The table shows the medal tally for the selected year by country, gold, silver, total medals, and bronze. By default, the slider input will show the 2016 year and the table will show 25 entries. Please note, if the year selected from the slider input has no record of the Olympic Games played, then a blank table will be displayed.



NOC	TotalMedals	Gold	Silver	Bronze
ALG	2	0	2	0
ARG	22	21	1	0
ARM	4	1	3	0
AUS	82	23	34	25
AUT	2	0	0	2
AZE	18	1	7	10
BAH	6	1	0	5
BDI	1	0	1	0
BEL	21	2	17	2
BLR	12	1	4	7

4. Total medals won by countries - The map view

Total medal count of countries from 1896 to 2016 is shown on the map. The global map displays geospatial visual representation. The color range selected is light blue to dark blue where dark blue shows the most successful countries and light blue show the less successful countries in Olympics. The Grey color shows that data is not available for that country. Please be aware that the large dataset makes it difficult for the map to load quickly on the dashboard.



V Implementation

R programming was used to put the final concept into practice. The information is derived from the relevant sources, which were before indicated. The datasets were combined and organized before being aggregated in accordance with each visualization, after which the visualizations were created.

Libraries

The libraries used for building this application are mentioned below:

- Shiny: It facilitates the development of interactive web apps using R. Writing this is simple. There is no need for knowledge of web development.
- Shinydashboard: It is useful when creating a dashboard in R. It consists of a header, a sidebar, and a body.
- Plotly: An interactive web-based visualization is created using it. Zooming and tooltips are the interactions.

VI Key Findings

- Over time, both the number of nations participating and the total number of games played has increased.
- The USA is the top nation in gold, silver, and bronze medals.
- The USA, Soviet Union, Germany, Great Britain, France, Italy, Sweden, Australia, Canada, and Hungary were deemed to be the top participating nations in the Olympic games.
- Only a few nations (Canada, France, Norway, USA, Austria, and Australia) excelled in the snowboarding competitions at the Olympics. Despite having a tropical environment, Australia surprised everyone by capturing gold and silver medals.
- Russian athletes have dominated the rhythmic gymnastics division, winning nearly as many medals as the rest of the world combined.
- Only Canadian, Great Britain, and American Olympians win medals in the Lacrosse event.
- The top basketball-playing nations are the USA, Serbia, Russia, Brazil, and Australia.
- Only two regions viz. Asia and Europe got medals in badminton. China is the most successful of them, followed by Indonesia, Malaysia, Denmark, the UK, and South Korea.
- Athletics is found to be a very popular sport, where almost all countries win at least a medal over the year.

VII References

- https://www.kaggle.com/datasets/heesoo37/120-years-of-olympic-history-athletes-and-results
- https://www.digitalshadows.com/uploads/2018/02/Threats-to-the-Winter-Olympics.png
- https://olympics.com/en/
- https://www.freeiconspng.com/
- https://en.wikipedia.org/wiki/Olympic_Games
- https://shiny.rstudio.com/
- https://rstudio.github.io/shinydashboard/