Visualization Project

2022-11-02

Tokoyo Olympics 2020

Varun Pramod Agrawal

Roll No: MDS202251

DG1107: Visualization

Email: varuna@cmi.ac.in

Chennai Mathematical Institute

Abstract

In this project, We will perform EDA on Countries and their performances in Tokyo Olympics 2020. We will also see the gender trends across various sports disciplines. For this, we will use visualization techniques to describe the performance and other patterns in the data. Then we aim to deploy an analytics dashboard that will present the analysis in a user-friendly and attractive way.

Keywords : Country, Medal, Discipline, R, Ggplot2, Visualization.

Table of Contents

1. *Introduction:*

The Tokyo 2020 Games were an unprecedented demonstration of unity and solidarity as the world came together for the first time following the onset of the COVID-19 pandemic for an Olympic Games focused on the pure essentials: a celebration of athletes and sport. The event program for the 2020 Summer Olympics was approved by the IOC executive board on 9 June 2017. IOC president Thomas Bach stated that their goal was to give the Games "youthful" and "urban" appeal and to increase the number of female participants.

The Games featured 339 events in 33 different sports, encompassing a total of 50 disciplines with around 206 National Olympic Committees and over 11,000 athletes.

2. *DataSet Description:*

We have chosen following 3 datasets from Kaggle:

- Medals
  - Have details of the medal tally and the ranking of the 93 countries and the Country names having been translated to English while uploading
- Teams
  - Have details of teams of the countries taking part in which disciplines in total of 743 rows
- Gender Entries
  - The distribution of Males and Females across 47 different sports disciplines.

3. *Objectives:*

Using the above dataset, we will describe the following:

- How did a country perform compared to others
- How is the gender of athletes distributed across different sports categories?
- How many different sports the country has taken part in
- Which countries have the most number of teams participating.

The above objectives can be achieved by plotting bar plots, pie charts, and line plots.The above objectives can be achieved by plotting bar plots, pie chart and line plots.

4. *EDA Summary*

```
## Scale for 'fill' is already present. Adding another scale for 'fill', which
## will replace the existing scale.
```

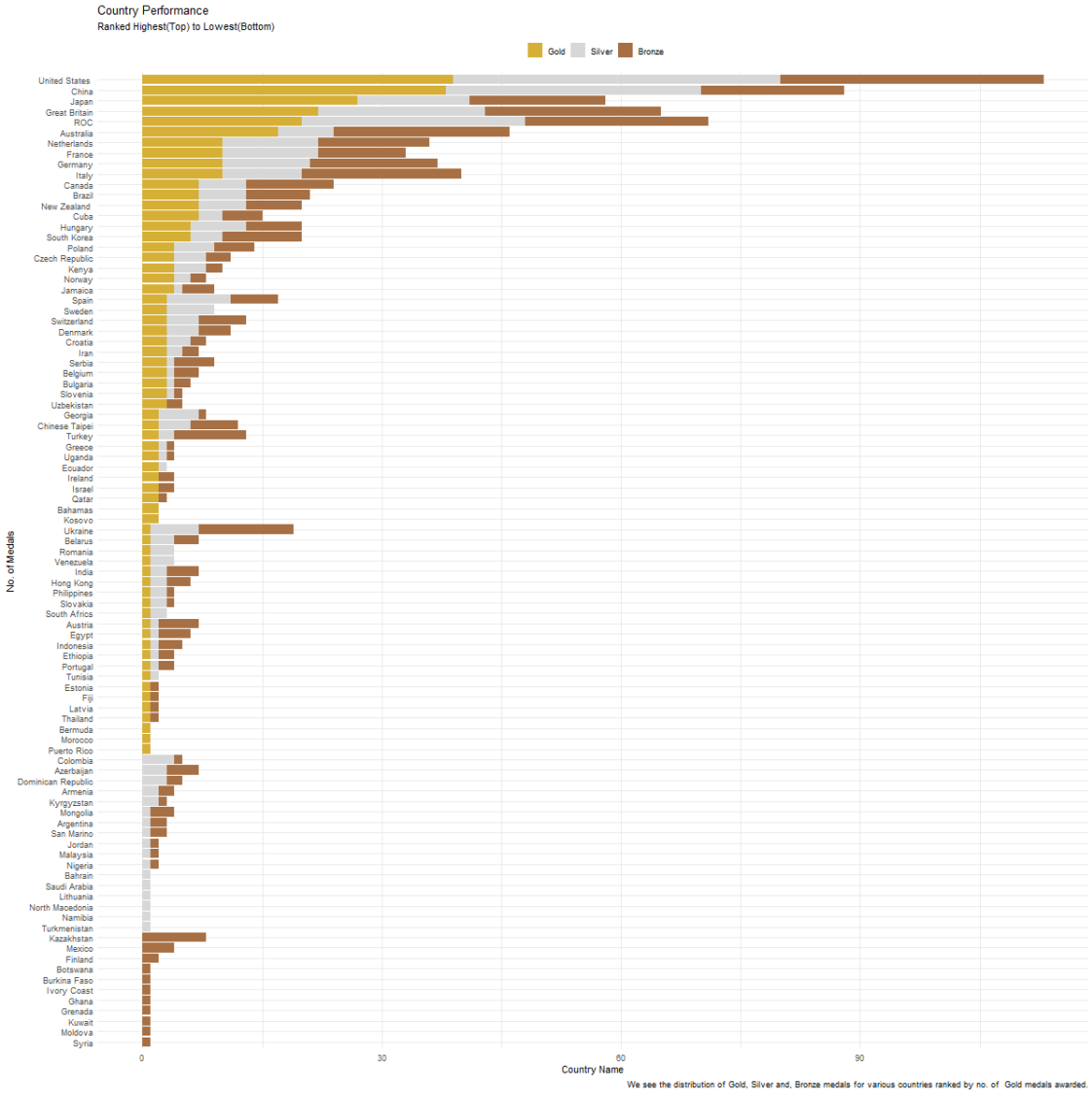

Figure 1: We see the distribution of Gold, Silver and, Bronze medals for various countries ranked by no. of Gold medals awarded.

```
## Scale for 'fill' is already present. Adding another scale for 'fill', which
## will replace the existing scale.
```

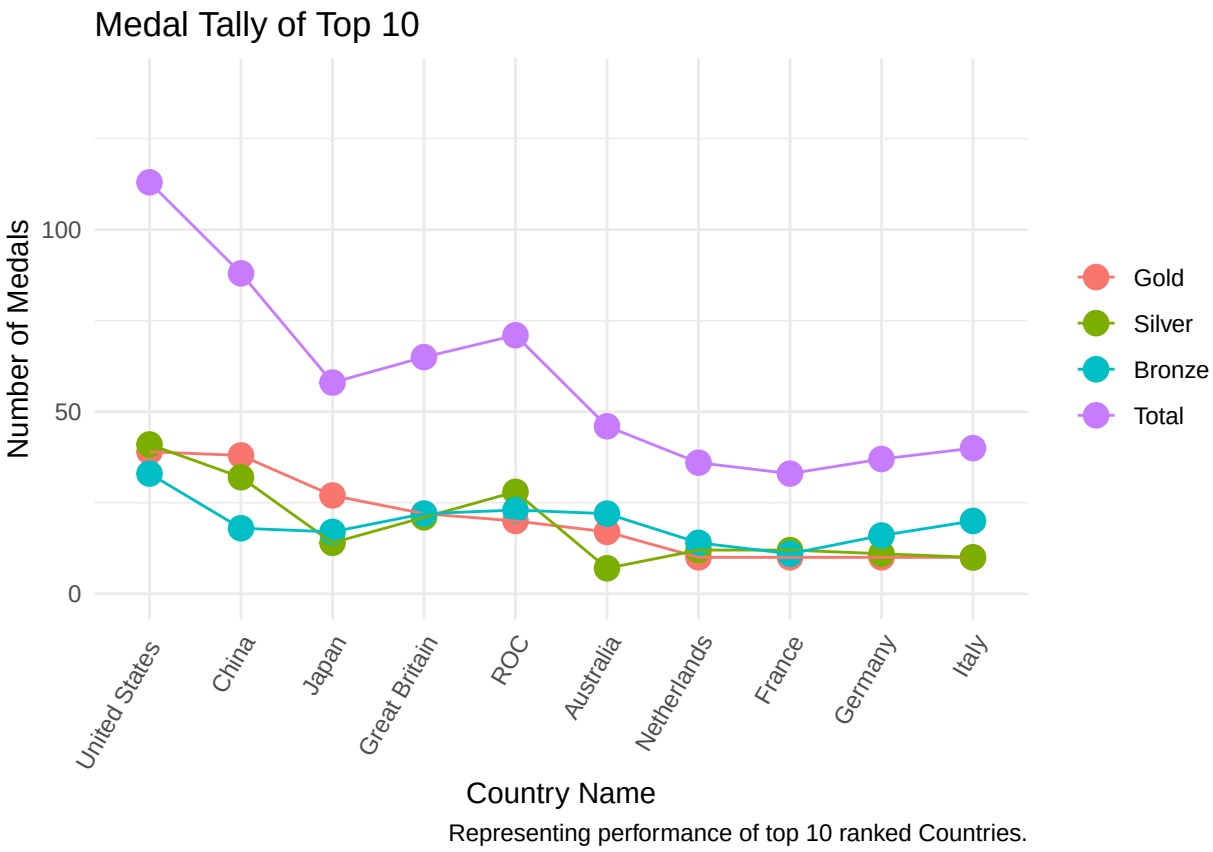

Figure 2: Representing performance of top 10 ranked Countries.

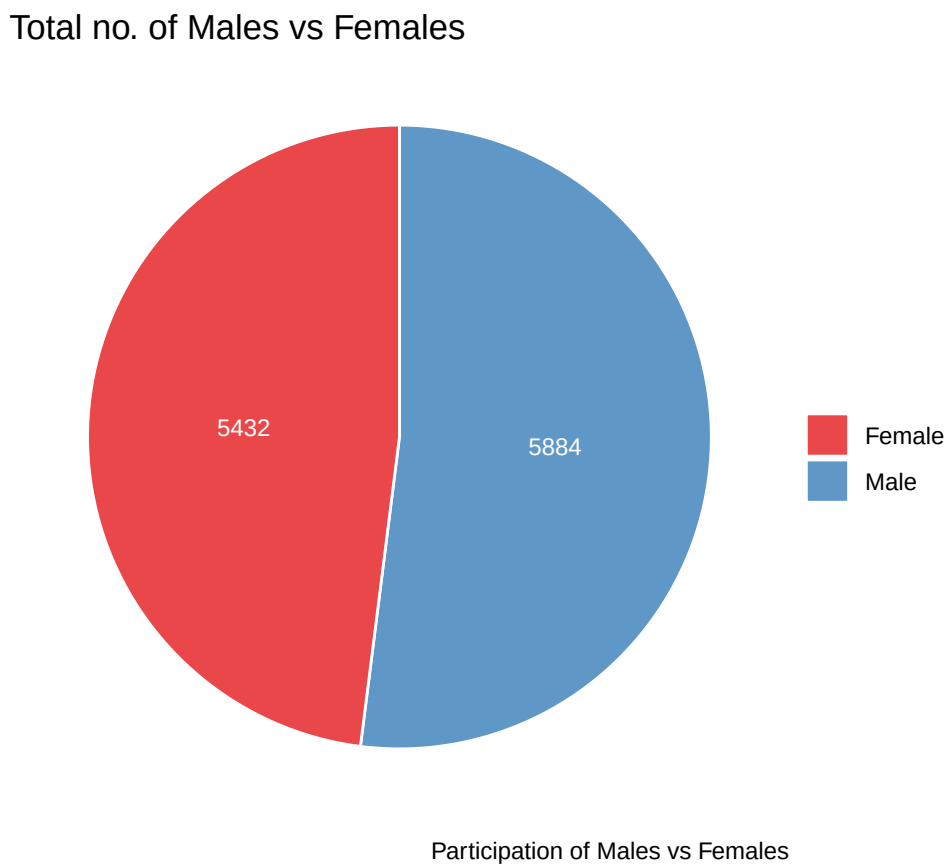

Figure 3: **Pie Chart** Participation of Males vs Females

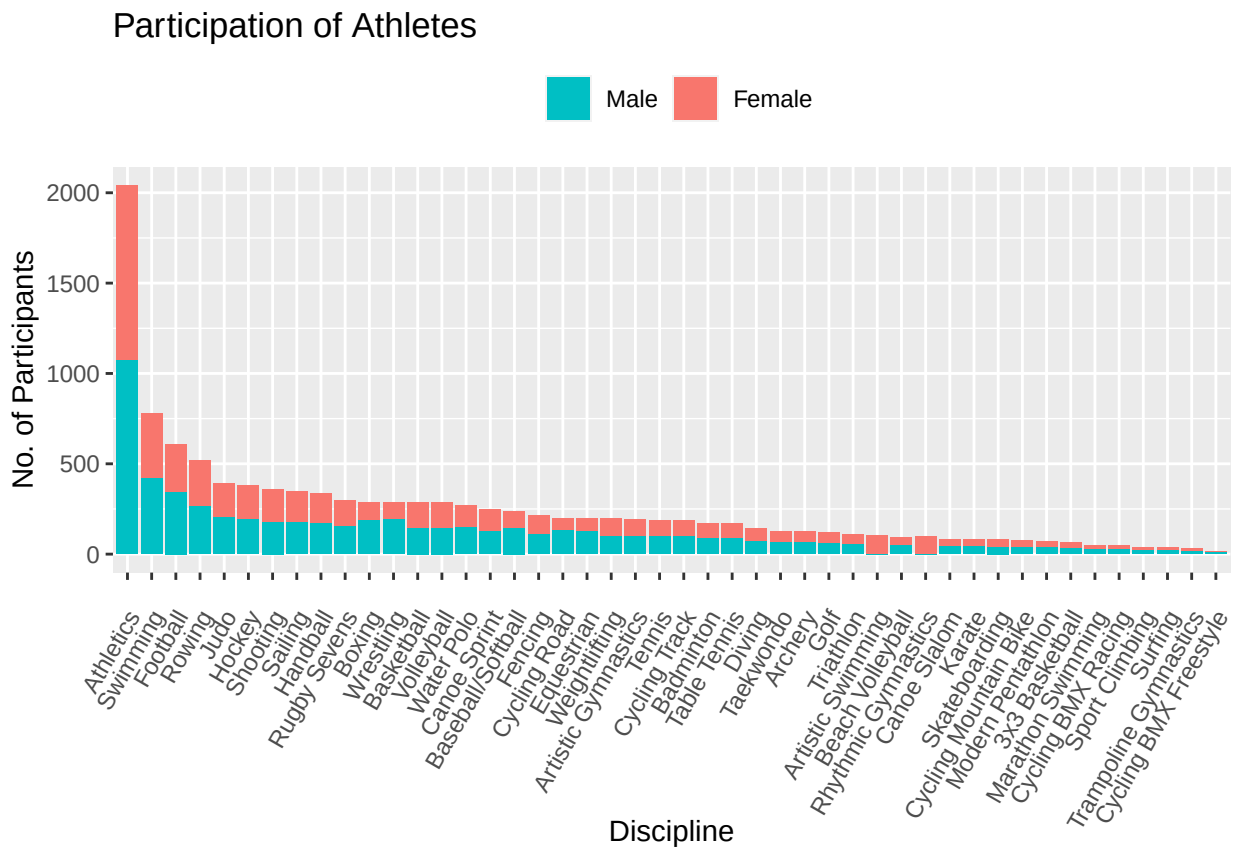

Figure 4: Participation of athletes in each discipline.

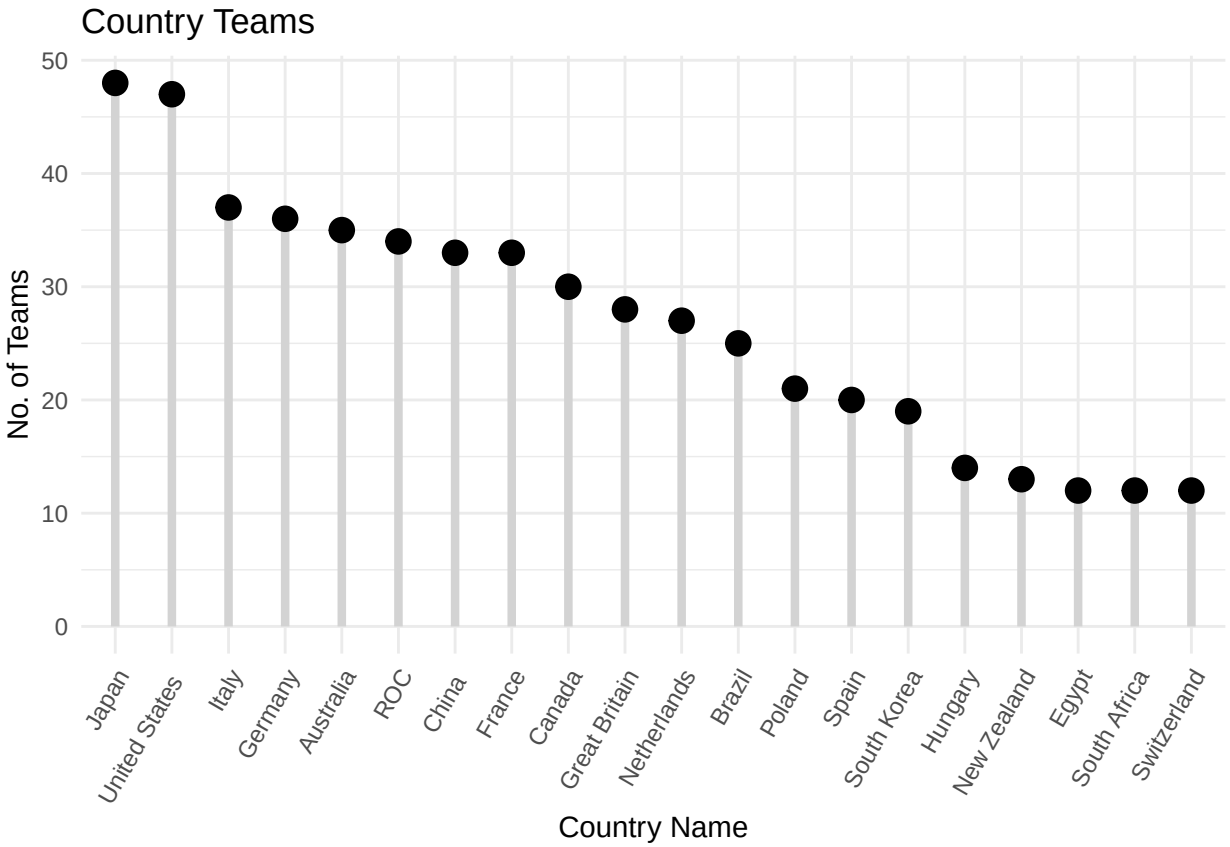

Figure 5: Top 20 countries with highest number of teams participating.

5. *Summary*

- the United States ranked no. 1 Country in terms of total no. of medals as well as no. of gold medals
- There is almost 50%-50% participation of Male and Female athletes in the Olympics
- Athletics got the most participation
- Japan had the most number of different teams taking part followed by the United States

5. *Acknowledgments*

I would like to thank Prof. Sourish Das for his guidance in making this project.

References:

- Kaggle
  - https://www.kaggle.com/datasets/jeronimojr/tokyo-2021-medal-table?select=tokyo_2021.csv
  - https://www.kaggle.com/datasets/arjunprasadsarkhel/2021-olympics-in-tokyo?select=Teams.xlsx
  - https://www.kaggle.com/datasets/arjunprasadsarkhel/2021-olympics-in-tokyo?select=EntriesGender.xlsx
- Wikipedia
- Tokyo Olympics Official Website