Exploring 120 Years of Olympic History: A Visual Data Analysis by Franck Di Sanza

Introduction

The dataset under examination encompasses 120 years of Olympic history, featuring a myriad of variables such as athlete details, event information, games information, medals, biometrics, and nationality. This research seeks to address some specific issues related to the Olympic Games. It tries to study changes in distribution of medals of the various nations, explore the trends in age, weight as well as height of athletes in particular sporting activities and if possible, discover links if any that exist between performance of these athletes and their age, weight, or height. The study also examines whether there are certain trends in popularity and participation in various sports between Olympic editions with deviations. The study also explores the profiles of the most successful Olympic sport people with a focus on their success in particular sports. The second focus dwells on the evolving sex ratio in sport and the role of the host cities on participation and success during Olympic Games.

Visualization & Graphing Methods

My approach involves a series of diverse visualizations to effectively communicate insights. Each point needs to be shown in a specific way to help the viewer understand the data as efficiently as possible.

The Medal distribution graphs are unique in that way, both for the packed bubble or the tree map. For this purpose, packed bubbles and tree maps represent convenient visualizations of medal classification by countries against the backdrop of Olympic history. Packed bubbles are considered as one of the intuitive representations, where each bubble represents a country, and its size is the total number of medals acquired. It makes it possible to show immediately how various countries compare in their overall productivity. Proximity of bubbles makes it possible to determine clusters of countries with similar medal scores, allowing the discovery of trends and exceptions. However, tree maps provide a hierarchical perspective dividing medals by sport categories or type. The color-coding helps provide a layered visualization, indicating the total medal and its distribution among the countries. It provides a full picture of the medal distribution intricacy and gives an insightful visualization view on the competition in the Olympic Games. This is what prompted me to put both these visualizations side by side.

A diagram of a diagram of different colors

Description automatically generated with medium confidence

To analyze the evolution of the participant’s age and BMI, I add to get creative to show the data for a single game and the evolution over every game at the same time. Both graphs are identical so I will explain my thought process for the age one. The display the age spread for every game, a shape graph was used with the years as the columns and age as rows. On top of this is an average line that goes through all the years. I picked this way of graphing because it allows the viewer to analyze the age parameter for one game and its evolution through the years. The shapes are also color-coded to display the performance of the athletes of that age. Because one of the goals of this analysis is to understand the correlation between biometrics and performance, I had to find a way to show a indicator of performance, in that case, a medal count.

A screen shot of a graph

Description automatically generated

Furthermore, line graphs and bar graphs are among the essential tools in this context for presenting useful information about trends, patterns, and comparisons concerning our study of 120 years of Olympic history. Line graphs are perfect tools to show how a sport’s popularity changes through the years. Smooth lines clearly indicate the trajectory and make it possible to detect patterns and abrupt changes that attract attention. On the contrary, bar graphs are extremely useful in illustrating the medal count or participation per country that has hosted the Olympics. Therefore, this project employs line graphs and bar graphs strategically to ensure that the most relevant findings are not only analytically derived but also visually impactful, thus aiding us to comprehensively understand the complex dynamics of the Olympics’ landscape.

A graph of a graph showing the growth of a sports game

Description automatically generated with medium confidence

A screenshot of a graph

Description automatically generated

Usability Analysis

While the analysis is usable in the right context and the representations contained significant information, they appeared more consolidative compared to breakthroughs. I believe in the necessity of organizing known information into a compact whole to make it available and useful for those eager about the Olympic Games. Nevertheless, the addition of more interactive components would make the analysis more responsive and allow the viewer to be more interactive with the data provided.

Discussion Section

This study of the entire Olympic visualization gave essential information on how complicated and varied these Games are. A major thing which was noted during the study was the fact that it would be near impossible to carry out a complete analysis for the whole Olympic Games since the activities involved are very many and the bodies involved also show diversities. The data set presented many patterns, correlations and outliers and the truth is that the common trends usually conceal their specificities for their sports. For example, trying to find out universal links between age, weight, height, and sporting performances showed to be a difficult task, as each sport has its own peculiarities and specifics. It led to an awareness that a specific method was required to come up with scientific findings based on an individual’s sport.

Future Work

While Tableau proved effective for my visualizations by allowing me to customize, stack different types of graphs and being able to interact with them, certain limitations were identified. Tableau is not great at making complicated calculations that would have been interesting to make in this analysis. For example, calculating the difference between the performance of a host country while considering the evolution of the Olympic Games is complicated. Using a language such as Python on the side would have been a good addition to this analysis and I believe it could have made it more pertinent. It could have been done by outputting a different database through Python calculations and then using it in Tableau. An integration inside Tableau would be a good feature.