Dynamic R Shiny Dashboard

INFO 526 - Project2

Data Decoders

INTRODUCTION:

This project is an interactive dashboard application built with R and Shiny with data pre-processing capability. The application enables users to carry out real-time exploratory data analysis by uploading a dataset and code for data transformations and plots they intend to show. The users can list their data transformations, code for the plots and the filters of their choice and can call our package to create a Shiny Dashboard. This reduces their time to create a Shiny Dashboard and gives an opportunity to explore how a Shiny Dashboard looks before being introduced in the class. The dashboard takes any structured .csv file, but for the demonstration we used the Paralympic Games dataset from the Data DNA Challenge Data Archive. The dataset has comprehensive records of athletes participating and winning medals from different countries, sports, and years.The dashboard was developed to provide insights from this specific dataset, but it is also versatile for limitless other uses and also gives exposure to the new users of a Shiny dashboard. The dashboard also gives the users an opportunity to conduct complete preprocessing upon upload, including removing duplicate records, managing null or missing values, and treating outliers based on number patterns.

The Paralympic Games dataset has several important variables:

- npc_new: the National Paralympic Committee (country code)
- year and games: the season and year of the event (e.g., "2000 Summer")
- sport: the sport that is being played (e.g. Athletics, Swimming)
- npc_gold, npc_silver, npc_bronze: kind of medal won (Gold, Silver, Bronze)
- season: whether it is a Summer or Winter game.

These are the variables that are significant in addressing key questions, for example, which countries have the most medals over time, or which regions do better in endurance sports like Athletics and Swimming.

JUSTIFICATION OF APPROACH:

We created a modular R package that accepts raw data, user-defined transformation code, code for plots and filters that might be applicable to produce meaningful dashboards without writing extensive Shiny code. It reduces the barrier to entry for students and streamlines dashboard development for course projects. Instead of hardcoding the UI and server for the Paralympic dataset, we created a code that functions dynamically and builds dashboard pages from user-supplied code, plots and filters. Our approach reflects flexibility and scalability which are key factors for reuse and adaptation to any dataset or project. Also our approach for pre-processing options help learners understand the importance of data quality before visualization and supports best practices in data science. It also supports reproducibility and encourages modular coding practices. .

VISUALIZATION:

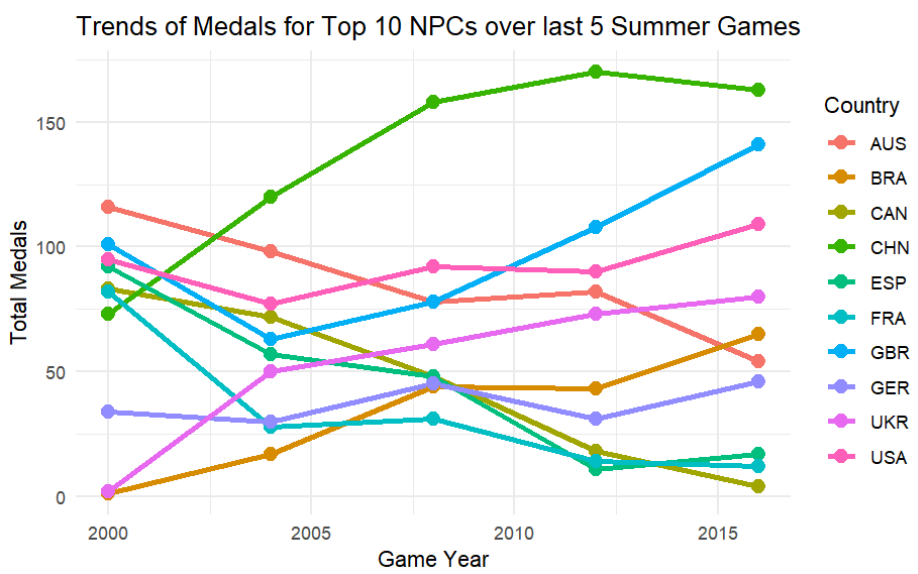
Using R Shiny development the Paralympic Dashboard delivers an interactive platform to analyze Paralympic Games information extensively. The system enables users to instantly see updated graphics through their interactive input operations which include filtering the data by year, country, and season. The application maintains powerful preprocessing functions which let users normalize and preprocess data through an intuitive interface. The application provides tools to process data which encompass null value removal and outlier scaling and duplicate removal and column name standardization. The prepared data makes sure that the dashboard delivers both high-quality insights along with statistical reliability. The system enables regular users and analysts to handle complicated datasets effectively while enabling them to explore their data collections through targeted investigations.

Question 1: During the past five Summer Paralympic Games what modifications occurred in the medal achievements of top 10 NPCs (National Paralympic Committee)?

Visualization Steps:

Users choose the Summer season option in the filter then view medal data for the five most recent years.The system determines the top 10 NPCs through total medal accumulation.This visual displays medal patterns for each NPC during five Olympics through line charts.

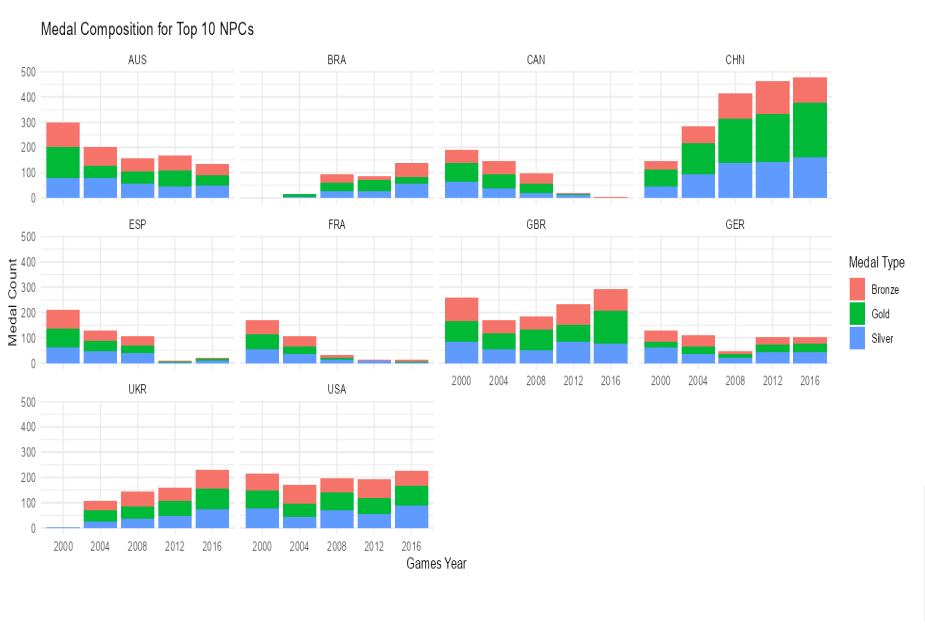
Trends of Medals for Top 10 NPCs over Last 5 Summer Games:



China (CHN) together with the United States (USA) maintain consistent dominance in Paralympic medals by either maintaining stable counts or increasing their totals. The trending pattern of different NPCs appears unstable because their investments and athlete development strategies change along with their strategic event targets.

Medal Composition for Top 10 NPCs:

A set of faceted bar plots displays medal distributions of Gold, Silver and Bronze for the top 10 NPCs throughout five Summer Games.

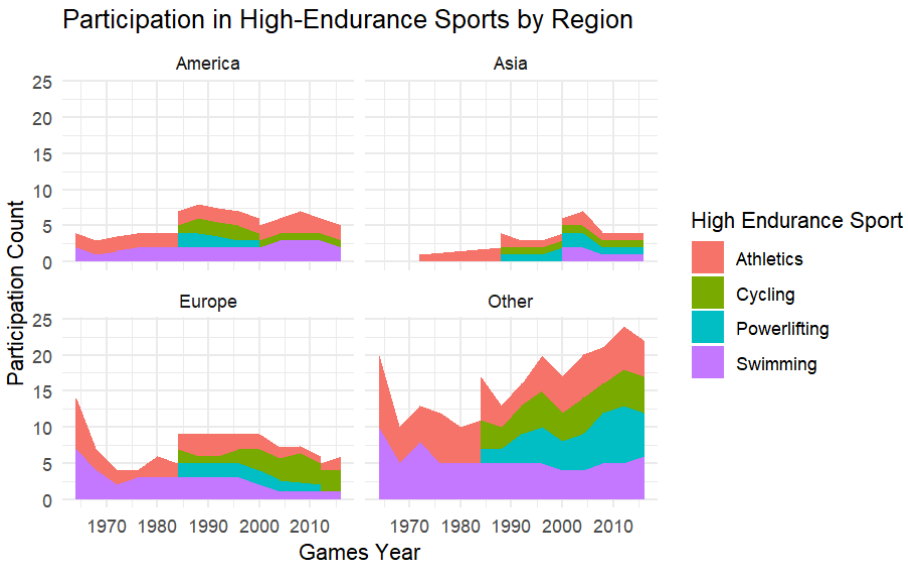


Medals exist in varying proportions which demonstrates performance variations between different teams. A nation receiving more Gold medals than other nations demonstrates superior elite-level performance since it shows evidence of better success at the highest level of competition.

Question 2: How have top level Masochism-based sport groups from specific regions controlled which sports they participate in and how has this domination evolved?

Participation in High-Endurance Sports by RegionSport Domination by Region (Bubble Chart):

Area plots tracking regional participation over time in Athletics, Swimming, Powerlifting, and Cycling.

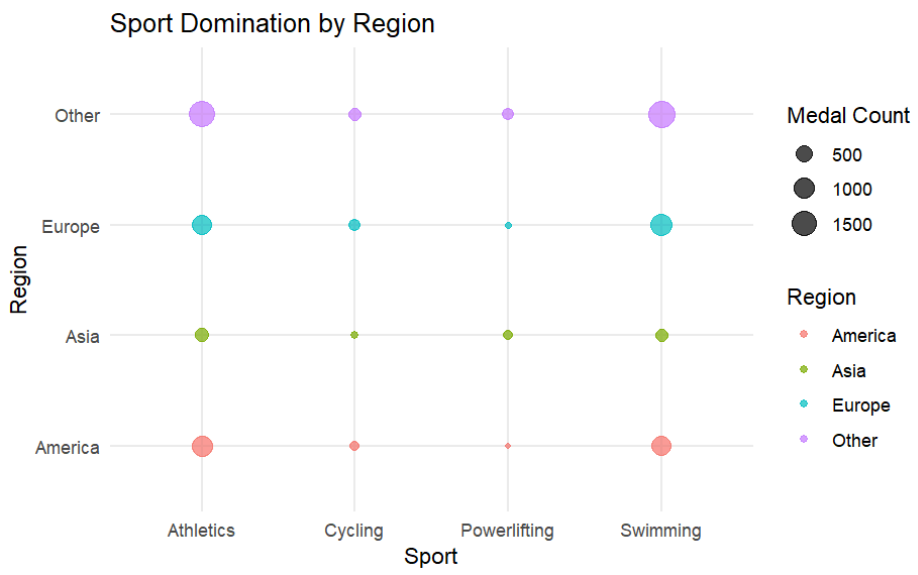


The athletic and swimming participation rates in Europe and Other

remain high and continue to grow. The observed trends match regional policies and sports creation infrastructure which along with athlete development systems focus on endurance events.

Sport Domination by Region (Bubble Plot):

The bubble chart displays sports and regions while using bubble size to represent the total number of medals achieved.



Different geographic areas control certain sports competitions with Track & Field leading in North America and Swimming leading in Asia because each area assigns particular attention to specific athletic disciplines.

DISCUSSION:

Through R Shiny technology the Paralympic Dashboard creates an interactive solution which enables easy analysis of Paralympic Games data. The tool's most valuable feature automatically prepares incoming datasets through processing null values and outliers and duplicate removal and column format normalization which produces high-quality inputs for visualization. Users can filter data through three options (year, country, and season) that generate real-time graphics which enable everyone from casual viewers to analysts to analyze complex questions effortlessly. The system enables users to present trends alongside medal analysis through its automatic design while allowing smooth user interaction regardless of programming knowledge or data preparation expertise.

The dashboard enables users to understand historical medal patterns and regional sporting superiority within its two essential analytical inquiries. The total medal counts during five Paralympic Games demonstrated consistent leadership by China and the USA but their medal compositions showcased distinct qualitative strengths between these nations. The endurance sports events at Athletics and Swimming competitions are dominated by Europe and Asia according to data from both bubble chart visuals and regional participation patterns. The measurements give details about sportswoman training strategies and show base structural adjustments that result from policies and development plans.This dashboard takes sports data into understandable stories which serve as flexible information tools for analytics-based decisions within sports science.