

Summer Olympic Games (1896-2024): Data Visualization

King's College London

Akshita Singh

7CCSMSDV – Introduction to Data Visualization

ACADEMIC HONESTY INTEGRITY

Students at King's are part of an academic community that values trust, fairness and respect and actively encourages students to act with honesty and integrity. It is a college policy that students take responsibility for their work and comply with the university's standards and requirements. Online proctoring / invigilation will not be used for our online assessments. By submitting their answers students will be confirming that the work submitted is completely their own. Misconduct regulations remain in place during this period and students can familiarize themselves with the procedures on the College website. *I agree to abide by the expectations as to my conduct, as described in the academic honesty and integrity statement.*

1 PART 1: ANALYTICS

1.1 Exploratory Research Questions Proposed

Q1: *Analyse the development of teams' performances over time. Are there detectable trends in the performance of teams over the last decade?*

This question looks at the changes in medal performances during the past ten years, which includes the last three Summer Olympics in 2012, 2016, and 2020. The number of medals won in each Olympic year is used to determine each team's (country's) performance using the official Olympic medal dataset from Kaggle. To facilitate time-series comparison, a pivot table in Excel was made to compile the total medals earned by each nation during Olympic years. Line graphs are used to demonstrate discernible trends, such as consistency, improvement, or decline.

Q2: *Which country has won the most medals in each Olympic year? (with a focus on six selected countries)*

The focus of this question is Olympic domination. Excel pivot tables were used to organise medals by country and filter them by year from the entire dataset. A ranking for the chosen six nations was produced for every Olympic year from 1896 to 2024. Each year, the nation with the most medals overall was determined. Peak years for each nation and changes in competition over time are highlighted in the

analysis. Bar charts were made to easily display the top-performing nations by year.

Q3: *How have the top 10 countries' medal counts changed over the last 10 Olympics?*

The focus of this question is Olympic domination. Medals from the entire dataset were sorted by year and Long-term medal trends for the most prosperous countries are examined in this question. First, the Grand Totals pivot table in Excel was used to determine the total number of medals for each nation. Together, the cumulative totals from 1988 to 2024 were used to determine the top 10 medal-winning nations. A secondary pivot table then displayed the medal totals for each of these nations over all ten Olympics. Excel was used to build a multi-line chart that showed variations over time, emphasizing trends, advancements, or decreases. This provides a historical snapshot of elite performance.

With a focus on identifying dominant nations, consistency, and changes in competitiveness, these questions were developed to examine Olympic team performance during the past ten years as well as over the longer term (since 1896).

1.2 Data Used & Appropriateness

The Summer Olympics Medals (1896–2024) dataset from Kaggle [1] is the dataset that was used. It includes medal information broken down by year, event, and athlete, enabling national aggregate. It is suitable due to:

It provides a solid foundation for long-term trend analysis because it covers more than a century.

It contains enough information to address all three questions: country, year, and medal kind.

It is neat and organized, making it appropriate for pivot analysis using Excel.

1.2.1 Data Cleaning and Preparation Process

Data preparation was crucial because of the original dataset's size and scope. The main actions performed in Excel were:

Step 1: Sort by Relevant Countries and Years Only the years 2016, 2020, and 2024 were pertinent for Q1. Only these three were extracted by applying a filter to the "Year" column. Argentina, China, India, Japan, Russia, and the United States were the six countries chosen for Q1 using an additional filter.

For Q2, the dataset was not restricted by year- instead full historical dataset was used on- Switzerland, Sweden, Spain, South Korea, West Germany, Russia, Japan.

For Q3, no country filter was initially applied — instead, top 10 countries were identified by their cumulative medal count across the last 10 Olympics (1988–2024).

Step 2: Use pivot tables to generate medal counts to total the number of medals won annually by nation:

Pivot Tables were built with “Team” as Columns, “Year” as Rows, and “Medal” set to “Count”.

This made it possible to compile a clear summary of the number of medals each nation earned during each Olympic year.

Step 3: We utilized Excel formula =SUM() to get total medals for each country across years.

=MAX() determined which country received the most medals in each year (Q2).

=LARGE() and =INDEX() assisted in determining the top ten nations in terms of total medals during the third quarter.

Step 4: Create a multi-line chart for 6 teams (2016-2024) using Q1.

Q2: Coloured bar charts indicating which country won each Olympic year among the listed countries.

Q3: A multi-line chart of the top ten nations, illustrating long-term trends from 1988 to 2024.

1.3 Relationship Between the Dataset and the Research Questions

All three research problems are addressed utilising the same source dataset, Kaggle's [1] Summer Olympics Medals (1896-2024). This dataset serves as a unified source from which specialised subsets were extracted to meet certain analytical objectives. This approach's strength is its consistency: while the raw data stays unchanged, it allows for a variety of analysis depending on how it is filtered, sorted, or aggregated.

Each study topic uses a distinct selection of nations and time periods, but the dataset's main structure — with 'Team', 'Year', and 'Medal' variables — allows for smooth comparison and integration between tasks.

Q1: Performance Trends Over the Last Decade.

Focusses on six particular countries and three Olympic years (2016-2024).

The dataset was filtered to show short-term trends in medal counts.

This subgroup enables us to assess how teams have developed recently, identify growing or decreasing nations, and find consistent performances at recent Olympics.

Q2: Top Country Each Year (Historical Comparison)

Each Olympic year involves a different collection of nations (for example, Switzerland, Sweden, Spain, South Korea, West Germany, Russia, and Japan).

In contrast to Q1, this larger historical viewpoint emphasizes supremacy per Olympic edition rather than consistency.

The filtered data helps discern eras of dominance, country peaks, or fading trends, supplementing Q1's short-term view.

Q3: Long-Term Trends for the Top Ten Countries

Uses the entire dataset at first, then selects the top ten nations based on cumulative medals from the previous ten Olympics.

The top ten are constantly chosen using formulae, guaranteeing that the most competitive nations are analysed.

This question offers a longitudinal perspective and ties to Q1 by extending trend analysis over a longer time period and across more nations.

1.3.1 Interconnections Among Questions

Temporal linkage: All questions use the 'Year' variable to ensure alignment across different Olympic periods.

Dimensional consistency: The 'Team' field is the category anchor for all visualizations and comparisons.

Comparative value: Q1 and Q3 both concentrate on trends (short-term vs. long-term), but Q2 offers individual peak-performance insights that supplement the wider trend perspectives.

Scalability: Thanks to the versatility of Excel's pivot tables and formulae, these several analytical routes may be explored from a single dataset, assuring consistency and decreasing error.

2 PART 2: DESIGN AND DISCUSSION

2.1 Question 1: – Analyse the development of teams' performances over time.

Are there any noticeable trends in team performance over the previous decade?

Design: Line graphic comparing medal counts each country from 2016 to 2024.

Design Rationale: A multi-line chart depicted the performance of chosen teams (USA, China, Japan, Russia, India, and Argentina) throughout three Olympic Games. The X-axis indicates the year (temporal dimension), while the Y-axis displays the

total number of medals (quantitative measure). Each line represents a nation, allowing for trend identification through slope and fluctuation comparisons.

This pattern was selected because it clearly displays upward or decreasing performance trends. The use of unique colour-coded lines within accessible palettes provides readability and comparability even for users with colour vision problems.

2.2 Question 2: Analyse the trend in which countries have won most medals

Design: Coloured bar chart with different nations displayed over all Olympic years from 1896 to 2024.

Design rationale: This visualisation employs a grouped bar chart format to display the year-by-year medal performance of seven selected countries.

Countries include Switzerland, Sweden, Japan, West Germany, Spain, South Korea, and Russia.

The X-axis depicts all Summer Olympic years from 1896 to 2024, while the Y-axis displays the total number of medals. Each nation is encoded with a separate colour, allowing for simple comparison within and between years. Countries are stacked or grouped based on medal contributions, letting viewers to determine who dominated each year.

This visualisation was used because it enables:

The highest bar indicates the leading country of the year.

Historical context—how global events (e.g., wars, boycotts) influenced medal patterns.

A comparative examination of each country's medal increase or reduction over time.

To guarantee great contrast and distinguishability, colour was used in accordance with accessibility guidelines. Bars are separated with enough cushioning to minimise overlap and improve legibility, especially during closely packed time periods.

This style improves on standard line charts by providing accurate, absolute medal counts, rather than merely trends. While a line chart might indicate change over time, a bar chart displays scale and dominance each year, which is more appropriate for this question's intent: "which country won the most each year".

2.3 Question 3: - How have the top 10 countries' medal counts changed over the last 10 Olympics?

Design: Long-term multi-line chart with 10 countries across 10 Olympic Games (1988–2024)

The chart has a multi-series line arrangement, with the X-axis representing Olympic years and the Y-axis displaying total medals. Each country is represented by a distinct colour line, making it easier to follow performance over time.

This design was chosen for its capacity to identify long-term adjustments, steady improvements, stable leadership, and

abrupt shifts. For example, the United States' solid leadership and China's development since 2000 are readily obvious.

Compared to stacked bar charts or pie charts, this style excels in emphasising relative location and detecting intersecting patterns. Colour consistency and clear spacing were prioritised for clarity. The design also incorporates best practices covered in class, such as line encoding for temporal data and using slope as a comparison channel.

REFERENCES

- [1] S. de Oliveira, *Summer Olympics Medals (1896–2024)*, Kaggle, 2023. [Online]. Available: <https://www.kaggle.com/datasets/stefanydeoliveira/summer-olympics-medals-1896-2024>
- [2] C. V. Gonzalez, Simulation and Data Visualization Coursework: VGC-2020-2021, King's College London, Department of Informatics, 2021.
- [3] M. Bostock, V. Ogievetsky, and J. Heer, "D³ Data-Driven Documents," *IEEE Trans. Visualization & Computer Graphics*, vol. 17, no. 12, pp. 2301–2309, Dec. 2011.
- [4] A. Cairo, *The Functional Art: An Introduction to Information Graphics and Visualization*, 1st ed. Berkeley, CA: New Riders, 2012.
- [5] J. Heer, M. Bostock, and V. Ogievetsky, "A Tour Through the Visualization Zoo," *Communications of the ACM*, vol. 53, no. 6, pp. 59–67, June 2010.
- [6] KCL Module Materials – 7CCSMSDV, "Introduction to Data Visualization," Lecture Slides and Labs, Department of Informatics, King's College London, 2025.
- [7] I. Kant, *Critica della ragion pura*, p. 1. Translated by G. Gentile & G. L. Radice, 6th ed., Bari: Laterza, 1949. (Cited in accordance with course policy on AI tools)
- [8] A. Singh, "Interactive Drill-Down Dashboard of Olympic Medal Counts (2016–2024)," created using D3.js as part of coursework for 7CCSMSDV – Introduction to Data Visualization, King's College London, Apr. 2025.

Appendix A

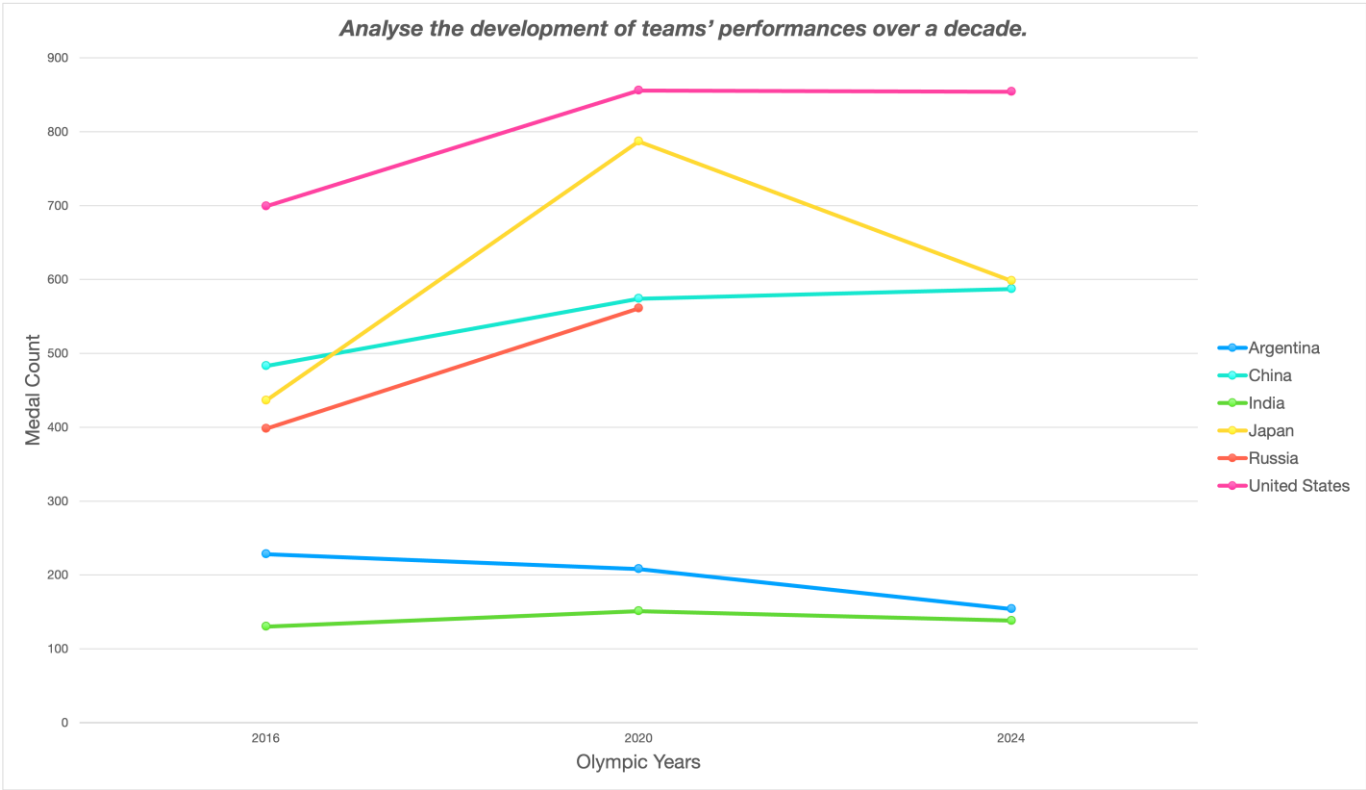


Figure B.1: Line chart representing short-term medal trends of Argentina, China, India, Japan, Russia, and the United States from 2016 to 2024. Each line illustrates how a country's performance evolved over three Olympic events, highlighting consistency, spikes, or declines.

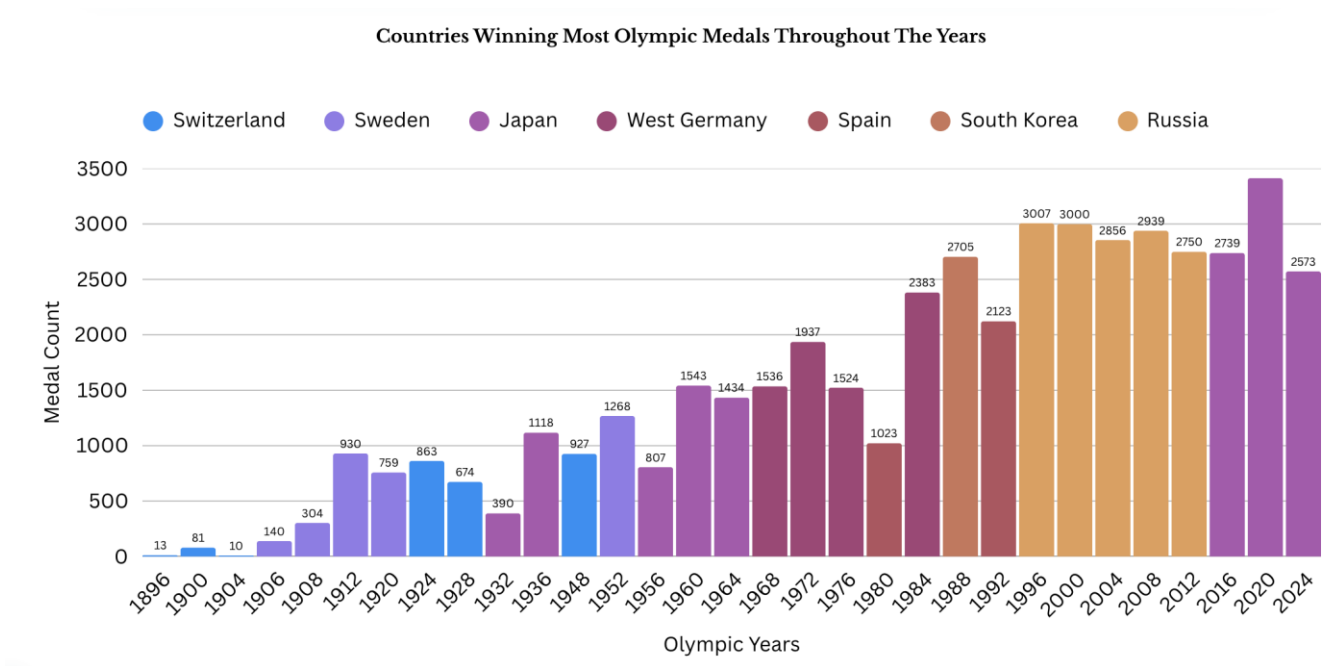


Figure B.2: Grouped bar chart displaying year-wise medal distribution for Switzerland, Sweden, Spain, South Korea, West Germany, Russia, and Japan from 1896 to 2024. The chart shows comparative dominance in each Olympic year, helping identify which country led during each edition.

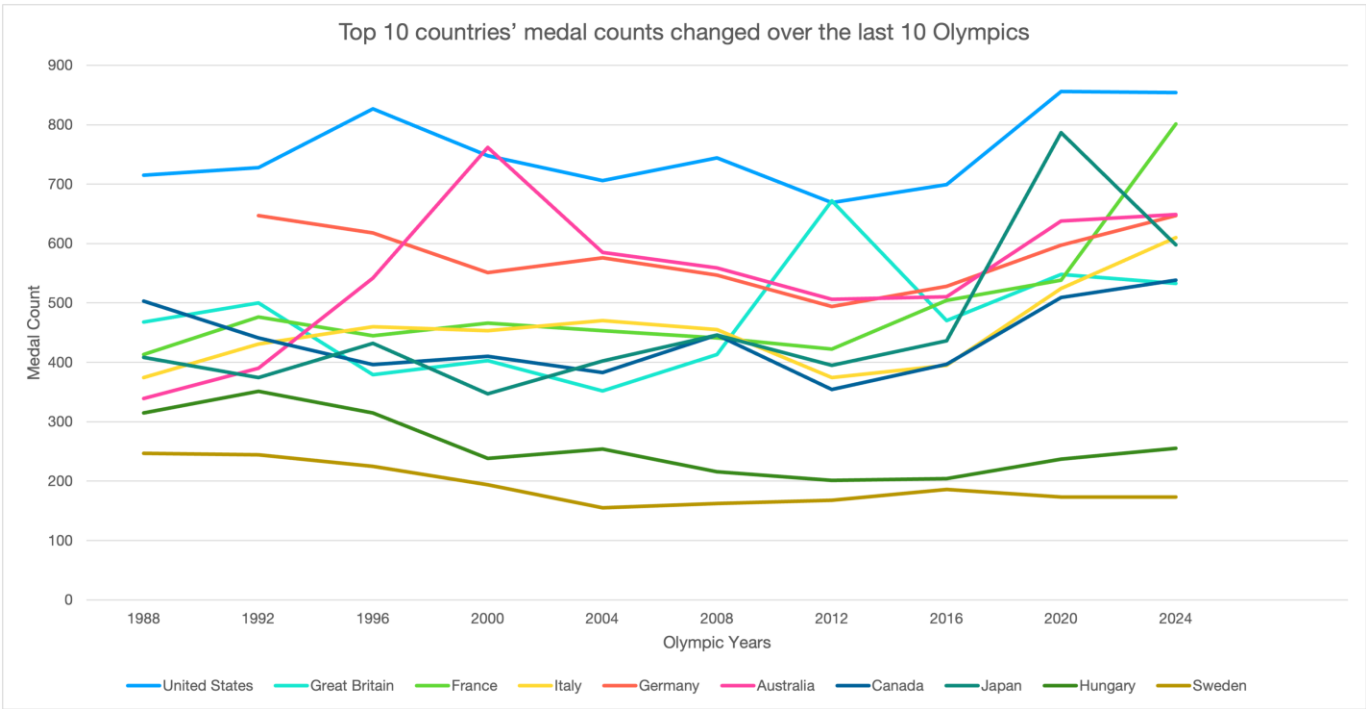


Figure B.3: Line graph tracing medal counts of the top 10 countries from 1988 to 2024. This figure visualizes long-term trends in medal dominance, revealing rising or declining patterns and country-wise consistency over the last 10 Olympic Games.

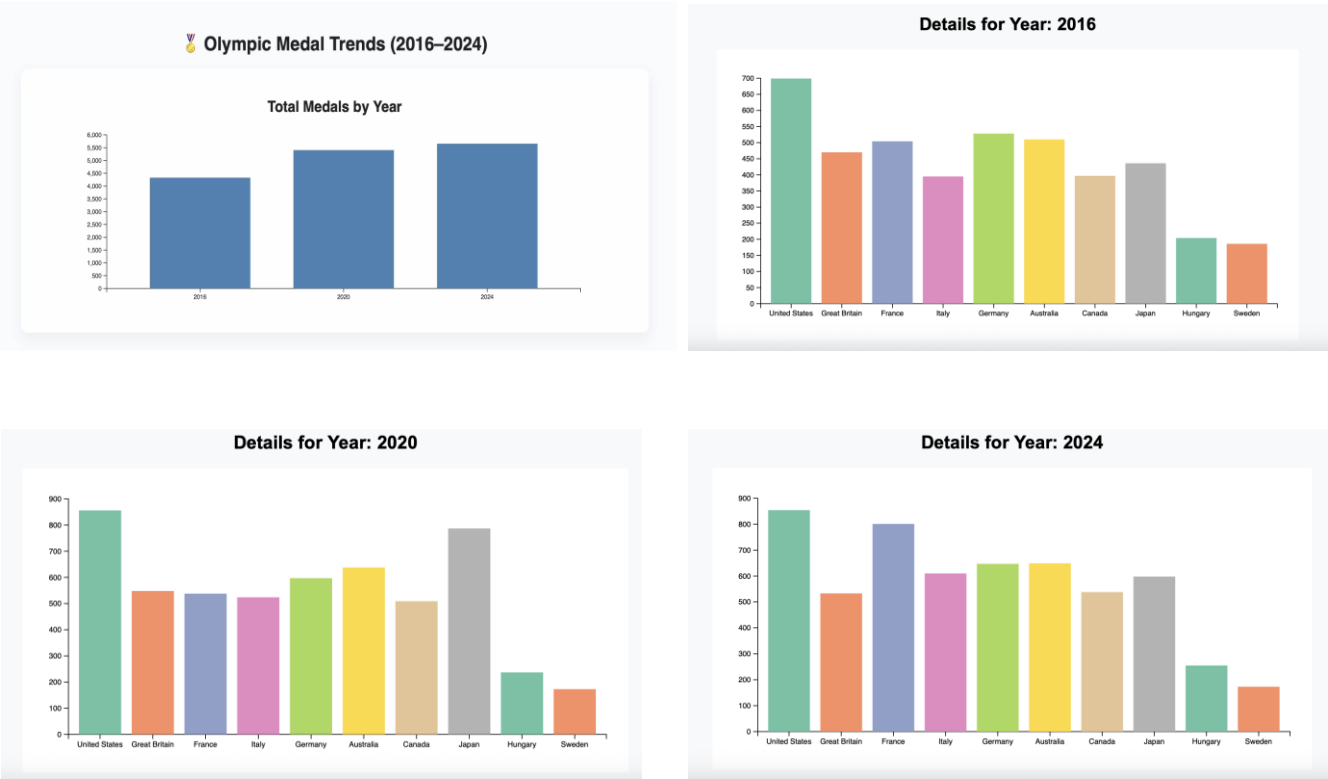


Figure C: Interactive Drill-Down Dashboard of Olympic Medal Counts (2016–2024)