Global Evolution of Elite Men's Tennis (1970s–2020s)

Geographical Analysis of ATP Top 100 Players

PRIFYSGOL BANGOR UNIVERSITY

Hassan Younas
Bangor University

Introduction

Men's professional tennis has undergone significant transformation over the past five decades, evolving from a sport dominated by a small number of nations into a truly global competition. This project explores the **geographical distribution and regional dominance of elite men's tennis players** using **ATP Top 100 ranking data** from the **1970s to the present day**. By combining player demographic information with historical ranking data, the study examines **where the world's top players come from and how this distribution has changed over time**. Through a **choropleth map as the main visualisation**, supported by regional trend analysis, the project reveals long-term patterns of dominance, decline, and emergence across different regions. The visualisation-driven approach allows complex temporal and geographical trends to be communicated clearly, supporting a data-driven narrative about the global evolution of elite men's tennis.

The Challenge

Several challenges were encountered during the development of this project, primarily related to **data integration, consistency, and visual representation**. One of the main difficulties was combining ATP player information with historical ranking datasets spanning multiple decades, as player names, country codes, and data formats were not always consistent across sources. This required extensive preprocessing to standardise country names, resolve duplicates, and ensure accurate mapping between players and rankings. Another challenge involved preparing the data for **geographical visualisation**, particularly ensuring that all countries aligned correctly with Tableau's geospatial recognition for the choropleth map. Additionally, representing fair comparisons across decades posed a challenge, as the number of countries represented and the global reach of tennis increased over time, potentially influencing visual interpretation. Finally, designing a poster that balanced **visual clarity with narrative depth** required careful consideration of layout, colour usage, and annotation, to ensure the main story remained clear without overwhelming the viewer with excessive detail.

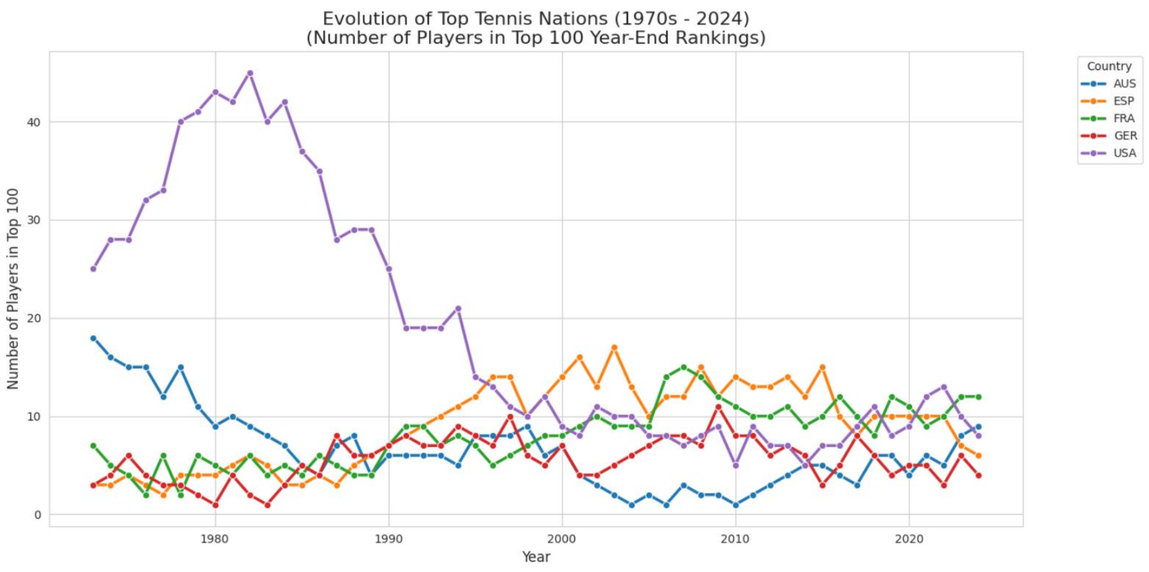

Figure 1. Evolution of Top Tennis Nations (1970s–2024). A time-series analysis of the Top 100 rankings, illustrating the long-term decline in US dominance contrasted with the rising consistency of European nations.

Conclusion

This analysis highlights a significant geopolitical shift in professional tennis, marked by the historical decline of US hegemony and the rising consistency of European nations. Simultaneously, the global distribution map reveals a sharp disparity in representation, showing that nations in Asia and Africa significantly trail behind these dominant regions in the elite leaderboard.

References

[1] J. Sackmann,"tennis_atp",GitHub.[Online].Available: https://github.com/JeffSackmann/tennis_atp/blob/master/README.md. [Accessed: Nov. 06, 2025].

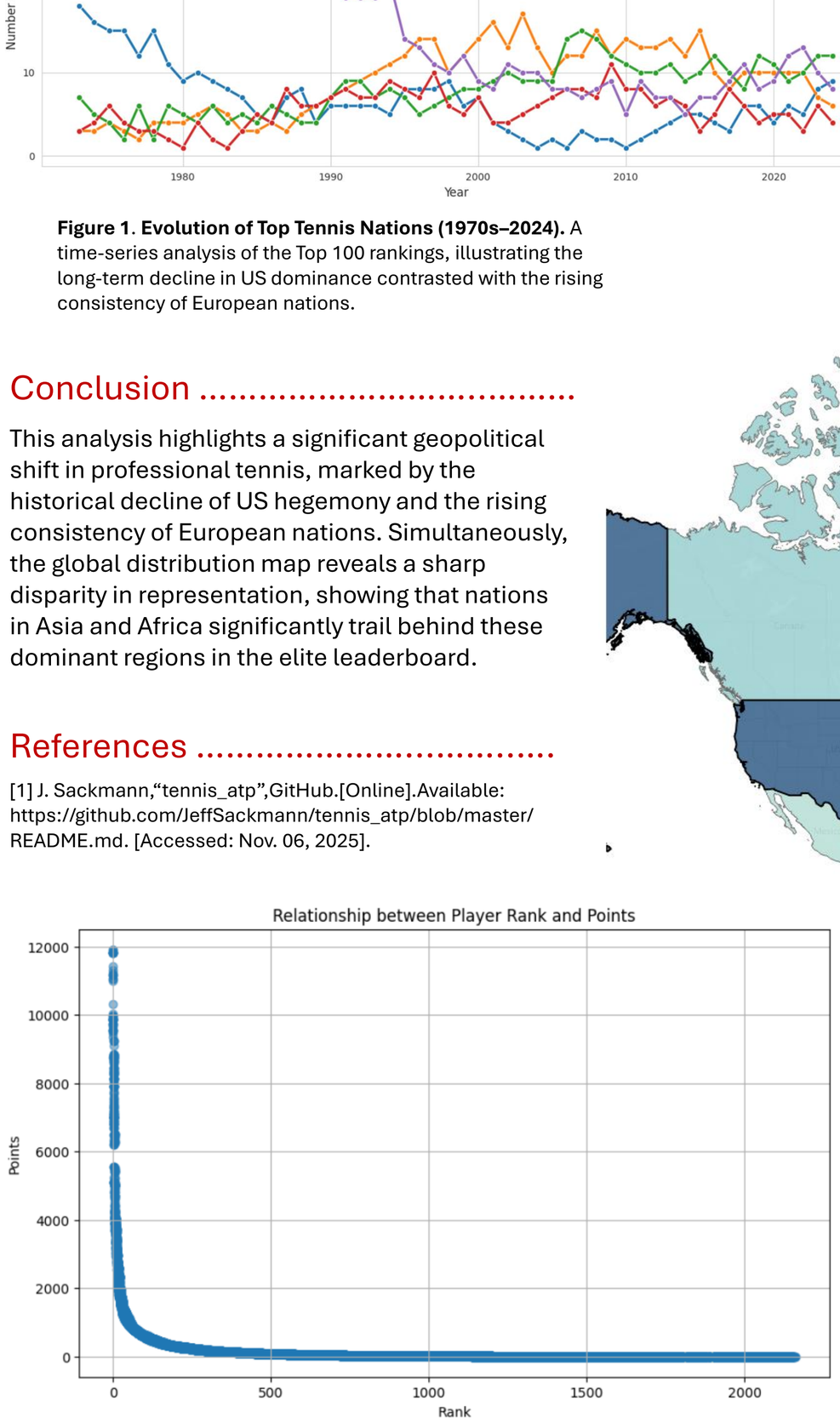

Figure 3: Inverse Correlation of Ranking vs. Points. This scatter plot illustrates that player ranking is inversely proportional to accumulated points.

Design Methodology

The design methodology for this project followed an **iterative, user-centred visualisation process**, informed by principles of information visualisation and the Critical Design Strategy (CDS). The process began with exploratory data analysis and preprocessing in Python to understand the structure, limitations, and inconsistencies within the ATP player and ranking datasets. Early design sketches and prototype visualisations were used to evaluate different visual encodings, leading to the selection of a **choropleth map** as the primary visualisation due to its effectiveness in communicating geographical distributions. Throughout the design process, several challenges influenced methodological decisions, including ensuring consistent country mappings across decades and avoiding misleading interpretations caused by varying country sizes and data density. Supporting visualisations, such as regional aggregation views and a temporal line graph, were introduced to address these challenges and provide complementary perspectives that strengthened the narrative. Design iterations focused on improving clarity, visual hierarchy, and interpretability through careful use of colour, layout, annotations, and metadata, ensuring the final poster effectively balances analytical depth with accessibility for a broad audience.

Implementation

The implementation of this project began with extensive **data preprocessing** to ensure consistency and reliability across multiple ATP datasets spanning several decades. Player demographic data was first cleaned and standardised before being integrated with individual ranking datasets for each decade. This integration required resolving inconsistencies in player names, country codes, and missing values. During this stage, an exploratory analysis was conducted to examine the relationship between **player ranking and ranking points**, revealing an **inverse relationship**, where higher-ranked players accumulate more points. This validation step helped confirm the correctness of the ranking data before visualisation. The processed data was then reshaped and aggregated to prepare dedicated datasets for **Tableau-based geographical visualisation**, including country- and region-level summaries. In parallel, a separate implementation was developed to generate a **line graph illustrating the evolution of top tennis nations from the 1970s to 2024**, which served as a supporting temporal view. The main challenge throughout the implementation was maintaining comparability across decades with differing levels of data completeness and global representation, which was addressed through careful filtering, aggregation, and validation prior to constructing the final **decade-wise choropleth maps** in Tableau.

Our approach

Our approach focused on transforming the idea of tracking the global evolution of elite men's tennis into clear, data-driven visualisations. After preprocessing and integrating historical ATP ranking data, we translated geographical patterns into a decade-wise choropleth map supported by temporal line graphs, addressing challenges such as inconsistent country mappings and cross-decade comparability through careful aggregation and validation.

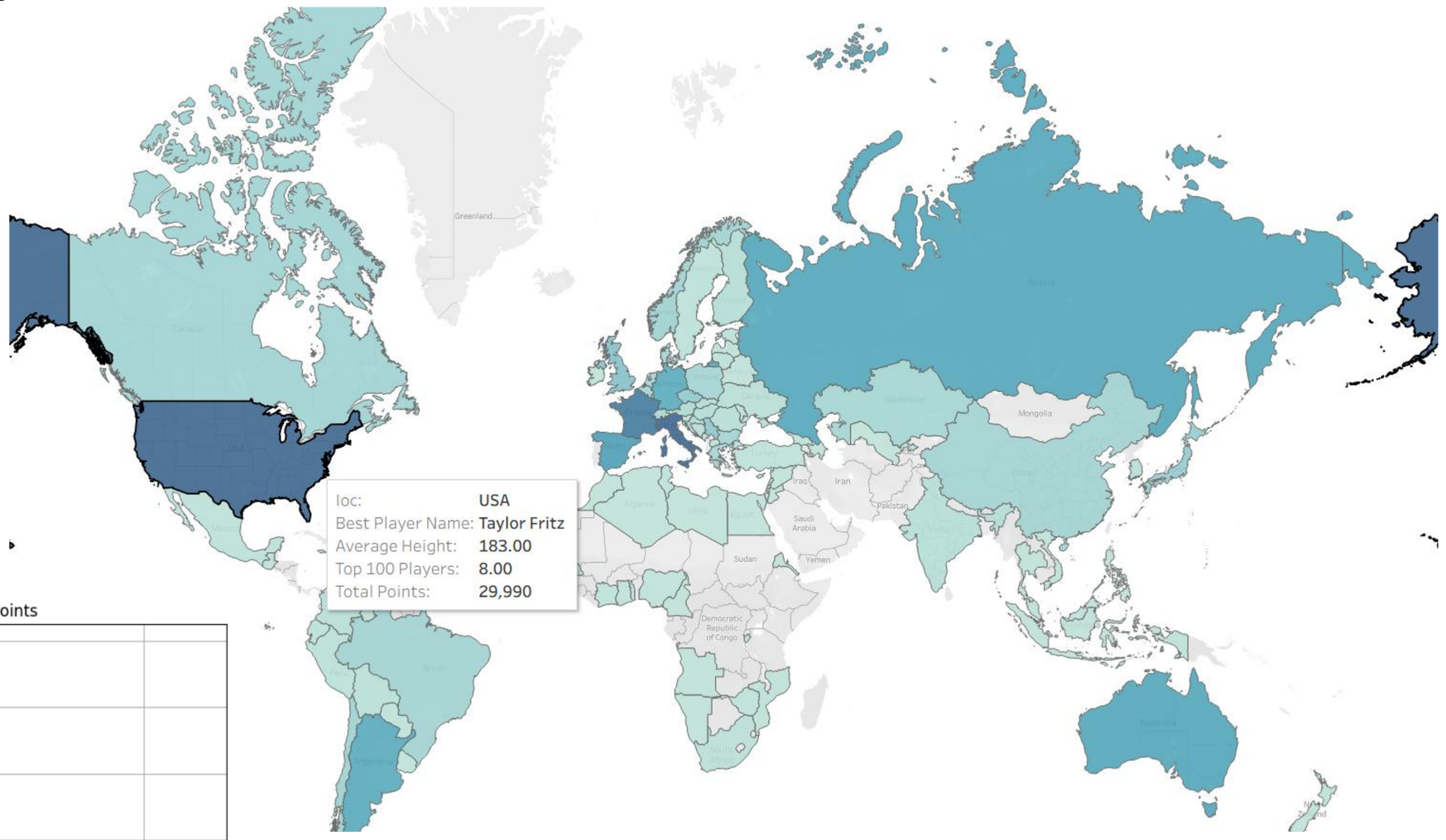

Figure 2. Choropleth Analysis of World Rankings. A global map illustrating the distribution of the Top 100 players, highlighting national dominance through a synthesis of total accumulated points and the number of ranked athletes per country.

Global Evolution of Elite Men's Tennis: A Visual Analytics Study of ATP Rankings (1970s–2020s)

Hassan Younas
Bangor University
Bangor, United Kingdom
hsy25fzp@bangor.ac.uk

Abstract

Men's professional tennis has evolved significantly over the past five decades, transitioning from dominance by a small number of nations to a more globally distributed competitive landscape. This project presents a visual analytics study of ATP Top 100 ranking data from the 1970s to the present day, integrating player demographics with historical performance metrics. Using a combination of analytical visualisations and an interactive choropleth map, the study explores relationships between ranking and points, temporal shifts in national dominance, and the geographic distribution of elite players. The work demonstrates how structured data preprocessing and thoughtful visual design can support clear storytelling about long-term trends in global sport.

1 Introduction

Professional men's tennis provides a rich domain for longitudinal data analysis due to the availability of detailed ranking and performance records across multiple decades. Historically, elite tennis has been dominated by a small number of countries, particularly the United States and Australia. However, the modern era reflects a shift toward European dominance and increasing global participation.

The aim of this project is to examine how elite men's tennis has evolved geographically and temporally by analysing ATP Top 100 ranking data from the 1970s to 2024. The study focuses on how individual-level performance metrics, particularly ranking points, aggregate to reveal broader national and regional trends. The final output is a visual poster centred around a choropleth map, supported by analytical and temporal visualisations that guide the viewer through a coherent narrative of dominance, decline, and emergence.

2 Data and Implementation

2.1 Data Sources

Two primary datasets were used in this study. The first consisted of multiple decade-based ATP player rankings datasets, covering the 1970s, 1980s, 1990s, 2000s, 2010s, 2020s, and current rankings. Each dataset included the attributes *PlayerID*, *Ranking_Date*, *Rank*, and *Points*. The second dataset, *atp_players*, provided player-level demographic information including name, nationality, handedness, height, and place of birth.

2.2 Preprocessing and Data Integration

Significant preprocessing was required due to inconsistencies across historical datasets. Fields not relevant to the analysis, such as date of birth and Wikidata identifiers, were removed to reduce complexity. Rows containing missing or invalid values were excluded to ensure analytical reliability.

Each decade-based rankings dataset was then merged with the player dataset using *PlayerID* as the primary key. This process resulted in seven structurally consistent merged datasets, enabling direct comparison across decades. Special care was taken to standardise country codes and names to ensure compatibility with Tableau's geographic recognition system.

2.3 Validation and Exploratory Analysis

Before constructing the final visualisations, exploratory analysis was conducted in Python to validate the correctness of the data. A scatter plot was used to examine the relationship between ranking and points, revealing a strong inverse and non-linear relationship. This finding confirmed that points increase disproportionately at higher ranks and justified the use of ranking points as the primary quantitative measure for aggregation and visual encoding.

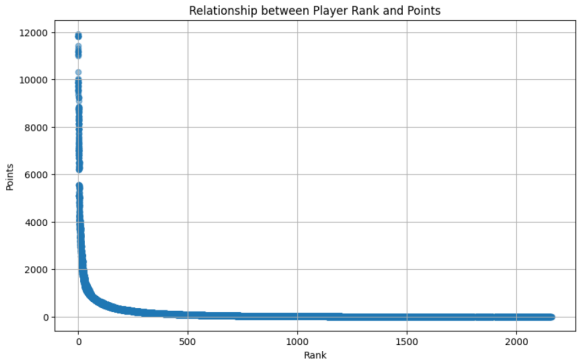

Figure 1: Inverse relationship between ATP ranking and accumulated points. Higher-ranked players hold disproportionately more points, validating the use of points as a continuous performance measure.

Additional exploratory analysis included examining player handedness using box plots. While the dataset was imbalanced, the analysis showed that right-handed players dominate the upper range of point distributions, providing contextual insight rather than a causal claim.

2.4 Aggregation and Tableau Preparation

For the choropleth visualisation, the merged datasets were further processed to aggregate player points by nationality. The analysis focused on the Top 100 players, as this group best represents elite

performance while maintaining visual clarity. Country-level summaries were created, including total accumulated points, number of Top 100 players, and average player height.

Separate Python scripts were used to generate a temporal line graph illustrating the evolution of leading tennis nations from the 1970s to 2024. Maintaining comparability across decades was a challenge due to changes in global participation; this was addressed through consistent filtering and aggregation strategies.

3 Visualisation and Storytelling

The poster follows a top-down narrative structure. Initial analytical visuals establish trust in the data by validating ranking metrics. This is followed by a time-series line graph showing the rise and decline of dominant nations, with the United States showing a clear long-term decline and European nations such as Spain and France demonstrating increased consistency.

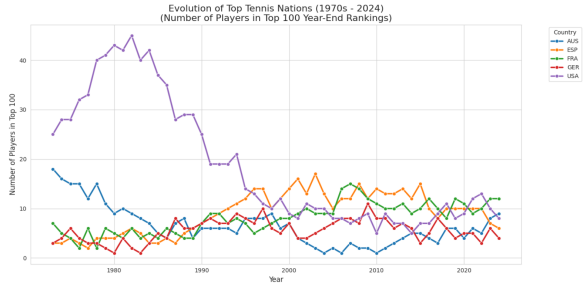

Figure 2: Evolution of Top Tennis Nations (1970s–2024). The decline of U.S. dominance contrasts with the rising consistency of European nations such as Spain and France.

The central visualisation is an interactive choropleth map encoding national dominance through aggregated ATP points. Colour intensity communicates relative performance, while interactive tooltips provide contextual metadata, including the top player per country and average height. Supporting text and annotations guide interpretation and mitigate known limitations of geographic visualisations, such as area bias.

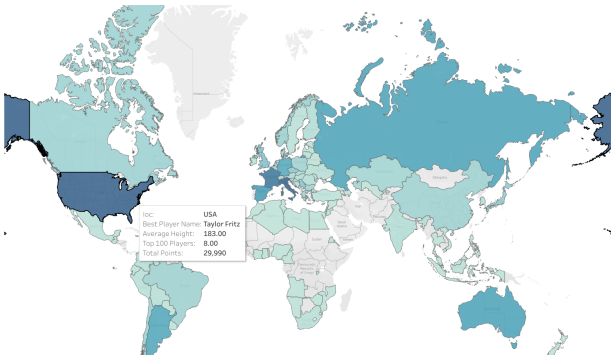

Figure 3: Global choropleth map showing the geographic distribution of ATP Top 100 players. Colour intensity represents aggregated ranking points by country, with tooltips providing player-level metadata.

4 Critical Reflection Using CDS

4.1 Concept

The conceptual framing of the project is coherent and well-defined, linking individual performance metrics to global patterns of dominance. The long-term temporal scope strengthens the analytical contribution, though the exclusive focus on Top 100 players introduces an elite bias that limits insights into broader participation trends.

4.2 Design

From a design perspective, the poster demonstrates effective use of visual hierarchy, consistent colour schemes, and clear annotations. The choropleth map is an appropriate choice for geographic storytelling, supported by complementary analytical and temporal views that reduce misinterpretation.

4.3 Structure

Structurally, the project benefits from a robust preprocessing pipeline and modular implementation. The separation of analytical validation, aggregation, and visualisation improved clarity and reproducibility. Future work could incorporate normalised metrics, such as population-adjusted player counts, to enhance cross-country comparability.

5 Conclusion

This project demonstrates how historical ATP ranking data can be transformed into meaningful visual narratives that reveal the shifting geography of elite men's tennis. The results highlight declining traditional dominance, rising European consistency, and persistent global inequalities in elite representation. The integration of structured data processing, analytical validation, and thoughtful visual design illustrates the value of visual analytics in sports data analysis.

Critical Design Survey (CDS)

Step 1

Assign a name to the design:

Summarise essence:

Circle 5 (first impression) words:

clear confusing sensible indifferent clever reliable pointless
indistinctive complex organised moderate spectacular useless
average bad fulfilling useful fair vague beautiful

Step 2

			-2	-1	0	1	2	
1	Is suitable for the user and task	*Perception*	○	○	○	○	●	Unsuitable ⇢ Suitable
2	Is understandable for user and task to hand		○	○	○	○	●	Incomprehensible ⇢ Understandable
3	It doesn't require guesswork		○	○	○	○	●	Requires guesswork ⇢ Clear assumptions
4	Is trustworthy		○	○	○	○	●	Distrustful ⇢ Trustful
5	Would be useful		○	○	○	○	●	Useless ⇢ Useful
6	It would fit in with other technologies	*Environment*	○	○	○	●	○	Wrong setting ⇢ Right setting
7	Uses suitable technology		○	○	○	○	●	Unsuitable technology ⇢ Right technology
8	Has appropriate interaction		○	○	○	○	●	Unsuitable interaction ⇢ Appropriate interaction
9	Its sizing is correct		○	○	○	○	●	Unsuitable size ⇢ Suitable physical size
10	Gives a positive ambience		○	○	○	○	●	Poor vibe/ambience ⇢ Positive ambience
11	Suitable user interface	*Interface*	○	○	○	○	●	Unsuitable GUI ⇢ Suitable GUI
12	Ergonomic interface		○	○	○	○	●	Uncomfortable ⇢ Ergonomic
13	Facets are sized suitably		○	○	○	○	●	Poorly proportioned ⇢ Suitable sized facets
14	Interface suitably spaced		○	○	○	●	○	Poor facet spacing ⇢ Relevant spacing
15	Suitable quantity of interface parts		○	○	○	●	○	Unsuitable facet quantity ⇢ Suitable facet quantity
16	Has all necessary components	*Components*	○	○	○	○	●	Missing components ⇢ All necessary components
17	Has all suitable output/view types		○	○	○	○	●	Unsuitable types ⇢ Suitable view types
18	Clear relationships between parts		○	○	○	●	○	Unclear correspondences ⇢ Clear view relationships
19	Task can be easily performed		○	○	○	○	●	Task unfulfilled ⇢ Task easily performed
20	Suitable organisation of components		○	○	○	○	●	Poor component layout ⇢ Good component layout
21	Inspiring design	*Design*	○	○	○	●	○	Uninspiring ⇢ Inspiring
22	Aesthetic and visually attractive		○	○	○	○	●	Unattractive ⇢ Visually attractive (aesthetic)
23	Good composition and space utilisation		○	○	○	○	●	Poor layout ⇢ Good composition
24	Suitable coverage of data/underpinning facets		○	○	○	○	●	Unsuitable coverage ⇢ Suitable coverage
25	Clear instructions, labels, legends to give context		○	○	○	○	●	Poor labels/legends ⇢ Suitable legends/labels
26	Right choice of channels to communicate things clearly	*Visual marks*	○	○	○	○	●	Poor choice of channels ⇢ Good channel choices
27	Communicates appropriate relationships/morphisms		○	○	○	●	○	Inappropriate mappings ⇢ Appropriate mappings
28	The types of marks used, communicate things well		○	○	○	○	●	Inappropriate mark types ⇢ Suitable mark types
29	Components are shown at the right level of abstraction		○	○	○	○	●	Poor scale/zoom ⇢ Good scale/zoom
30	Nothing is hidden that shouldn't be hidden		○	○	○	○	●	Overplotting ⇢ Clear display, easy read

Step 3

Sum values: 6 | 48

Total: 54

Reflect

Improvements

Total: -60 —55 -50 -45 -40 -35 -30 -25 -20 -15 -10 -5 0 5 10 15 20 25 30 35 40 45 50 55 60

Average: -2 -1.83 -1.66 -1.5 -1.33 -1.16 -1 -0.83 -0.66 -0.5 -0.33 -0.16 0 0.16 0.33 0.5 0.66 0.83 1 1.16 1.33 1.5 1.66 1.83 2

Poor design — Good design