Five Design Sheet – Design study of your data: Analysis of the ATP Match Statistics.

Hassan Younas
hsy25fzp@bangor.ac.uk
Bangor University

Introduction

The dataset used for this visualisation design study originates from Jeff Sackmann's *tennis_atp* repository, one of the most complete open-source collections of professional men's tennis data. It contains detailed historical information on player biographical attributes, match results, ranking histories, and match-level performance statistics. The dataset spans over five decades of the Open Era, with weekly ATP rankings mostly complete from 1985 onward and match statistics consistently available for tour-level events from 1991.

This assignment builds upon the data analysis performed in Assessment 1 and now focuses on designing appropriate visualisation concepts to communicate meaningful patterns within the dataset. Using the Five Design Sheet (FdS) methodology, several alternative visualisation ideas were developed to explore different aspects of the data, including global participation, player-level performance profiles, and the relationship between match statistics and outcomes. The aim of this design study is to select one suitable visualisation concept that effectively communicates an engaging and insightful tennis data story, and to document the design reasoning, comparisons, and evaluations that lead to the final choice.

Brief Outline of the data

The dataset comprises multiple interconnected CSV files providing structured information about ATP players, tournaments, rankings, and match statistics. Key components include:

Player Dataset

This file contains biographical information for each player, including first and last name, player ID, handedness (right/left), birth date, nationality (country code), and height in centimetres. These attributes support analysis related to demographics, player development, and cross-country comparisons.

Ranking Dataset

Weekly ATP ranking records include ranking date, ranking position, player ID, and ranking points (when available). Rankings are consistently recorded from 1985 onward, enabling longitudinal analysis of player progression, dominance, and career trajectories across multiple decades.

Match Results & Statistics

Match datasets (e.g., *atp_matches_2023.csv*) include tournament details, surface type, round, player IDs, scores, match winners, and match-level statistical variables such as aces, double faults, first serves in, break points saved, and total points won. These statistics are recorded most reliably for top-level ATP events from 1991 onward, with expanded coverage for Challengers and qualifiers in later years. Many match rows also store redundant biographical and ranking information for both players to simplify analysis.

Dataset Purpose

The dataset supports a wide range of analytical tasks, from evaluating player performance factors, studying match outcomes, and analysing ranking dynamics, to exploring country-level dominance and comparing statistical strengths between players. Because the data includes both numerical and categorical variables across many years, it is highly suitable for visual storytelling, statistical graphics, and interactive dashboards.

Alternative Designs:

Three alternative visualisation concepts were developed using the FdS methodology. Each design focuses on a different perspective of the dataset and offers unique storytelling possibilities.

1. World Map Choropleth – Country Dominance (Sheet 2)

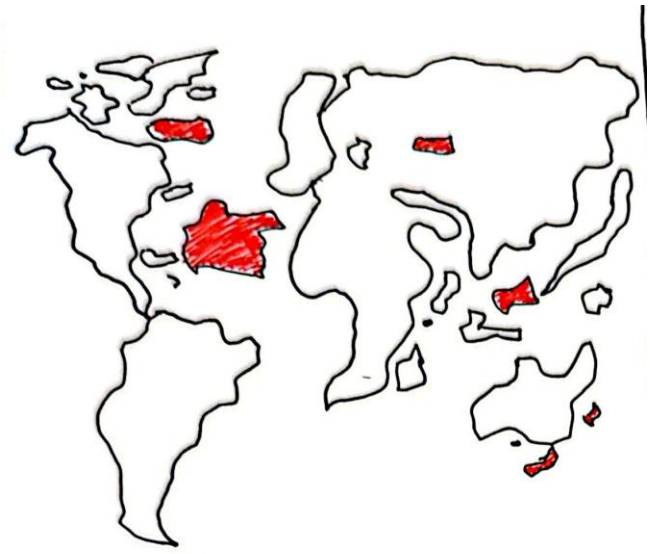


Figure 1: FDS World Map Choropleth

This concept visualises global tennis dominance by aggregating player-level data at the country level. The world map uses colour intensity to represent metrics such as number of top-100 players, total ranking points accumulated by each nation, or the proportion of match wins by players from that country. Countries with stronger professional tennis activity appear in darker shades, producing an intuitive and visually compelling representation of global participation and performance.

Strengths

- Provides an immediate high-level global overview
- Visually intuitive and appealing for poster presentation
- Highlights geographical patterns and national disparities
- Supports multiple dominance measurements (players, rankings, wins)

Limitations

- Geographic distortion may misrepresent small countries
- Countries with few but elite players (e.g., Serbia, Switzerland) may appear underrated
- Static choropleth cannot easily show temporal changes unless animated

2. Player Profile Performance Dashboard (Sheet 3)

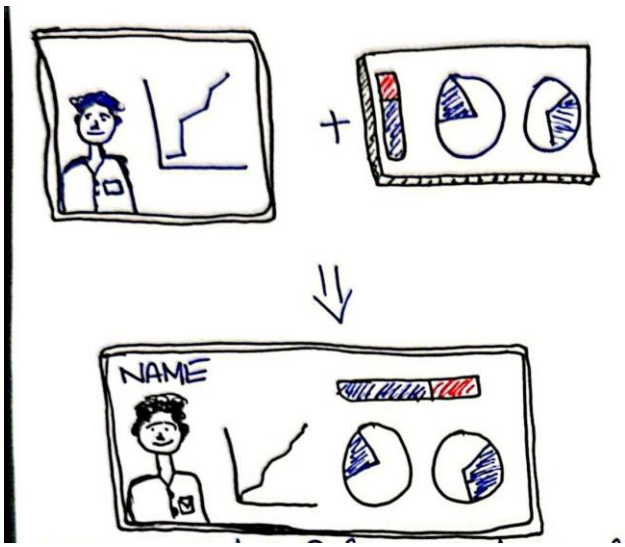


Figure 2: Player profile Performance Dashboard

This detailed dashboard focuses on a single player. It includes a player photograph, biographical information, ranking timeline, yearly performance bars, surface-wise pie charts, and mini-graphs of match statistics such as aces, double faults, and break-point performance. The aim is to create a magazine-style portrait of a player's entire career.

Strengths

- Very rich storytelling at the individual player level
- Combines multiple visualisation types for a complete profile
- Makes ranking evolution and performance trends easy to understand
- Suitable for comparing top players if extended to multiple dashboards

Limitations

- Too detailed for country-level or global storytelling
- Requires clean, complete player statistics

- May overwhelm the viewer if not carefully arranged
- Limited usefulness for analysing broader patterns across many players

3. Scatterplot of Fouls vs Match Winning (Sheet 4)

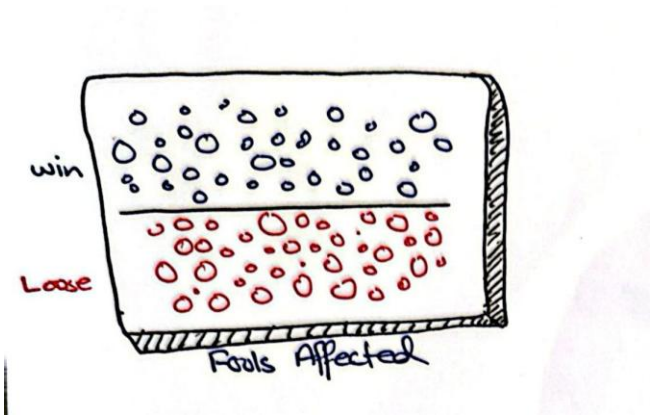


Figure 3: Scatterplot of Fouls vs match winning

This design examines the relationship between double faults and match outcomes. A scatterplot is used where a separating diagonal line divides the space: points above the line represent matches won, while points below represent matches lost. The plot can use colour to differentiate surfaces or tournament levels.

Strengths

- Directly tests a performance-related hypothesis
- Simple and analytically strong visualisation
- Highlights outliers (e.g., players winning despite many double faults)
- Easy to interpret in academic or analytical contexts

Limitations

- Focused on only one statistic
- Does not capture multidimensional factors influencing match outcomes
- Not visually rich enough for a poster-level story
- Lacks the global/narrative appeal of other designs

Critical Reflection Using CDS (Critical Design Strategy)

Using the CDS approach, each visualisation design was evaluated based on clarity, effectiveness, accessibility, novelty, data-mapping quality, and suitability for the final poster.

Clarity & Storytelling

- The **choropleth map** provides a very clear, instantly recognisable story about global tennis dominance.
- The **player dashboard** conveys deep insights but risks density and visual overload.
- The **scatterplot** is analytically strong but lacks narrative depth.

Data Mapping Quality

- The choropleth uses colour intensity effectively to show dominance, but requires careful colour-brewer palettes to avoid rainbow scales.
- The player dashboard uses multiple encodings (size, colour, shape), but could confuse inexperienced viewers.
- The scatterplot's binary win/loss division is clear but too simplistic for multi-factor tennis analysis.

Audience & Purpose

- The choropleth suits spectators, students, and analysts wanting broad comparisons.
- The player dashboard is ideal for detailed analyst-level investigations.
- The scatterplot is best for statistical audiences.

Novelty & Creativity

- The choropleth map offers creative potential, especially when enhanced with dominance metrics.
- The player dashboard is visually rich but less original.
- The scatterplot is traditional and expected.

Feasibility

- The choropleth is straightforward to implement using libraries such as GeoPandas, Plotly, or D3.js.
- The player dashboard requires more time and multiple coordinated charts.
- The scatterplot is the easiest but least impactful.

Overall Conclusion

The **choropleth world map** offers the strongest blend of clarity, storytelling power, visual appeal, feasibility, and relevance to the broad ATP dataset. It best represents the wide geographical spread of tennis talent and aligns most closely with the goals of Assessment 3.

Final Chosen Design - World Map Choropleth (Sheet 5)

The final chosen design builds on the choropleth concept from Sheet 2 and incorporates implementation-level details required for the realisation sheet.

Description of the Visualization

The final visualisation is a world map choropleth representing **country dominance in ATP tennis**, using aggregated metrics such as:

- Number of ATP players in the dataset
- Number of Top 100 or Top 10 players historically
- Total ranking points generated by players from each country
- Win percentages at ATP-level events

The map uses a colour-brewer-compliant sequential colour scale where darker colours indicate stronger dominance.

Algorithms & Processing

- Aggregation algorithm:
    - Group data by country_code
    - Compute dominance metrics (e.g., total ranking points, player counts)
    - Normalise values between 0–1 for consistent colouring
- Choropleth rendering:
    - Use geographical shapefiles and merge with aggregated data
    - Map normalised dominance values to colour scale
- Map projection:
    - Robinson projection for balanced global display

Dependencies & Libraries

Possible implementation tools:

- **Python:** Pandas, Matplotlib/Plotly, Numpy
- **Javascript:** D3.js,
- **Tableau (optional)**
- Colour palettes: ColorBrewer (per assignment rules)

Interaction (Optional for Poster)

- Hover labels with country stats
- Highlight of top countries
- Tooltip with player names and ranking summaries

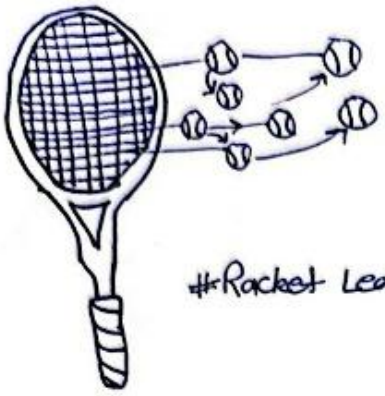
Technical Requirements

- Minimum display resolution:
    - 2500 px width for poster clarity

Why This Design Was Selected

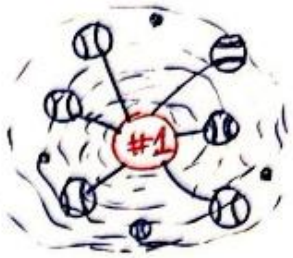
- Best balance of global storytelling and visual clarity
- Strong aesthetic impact suitable for an academic poster
- Easily understood by non-experts
- Represents the dataset at a macro level
- Allows identification of meaningful patterns such as dominance of Spain, USA, France, Serbia, and Switzerland
- Flexible enough for additional supporting charts in Assessment 3

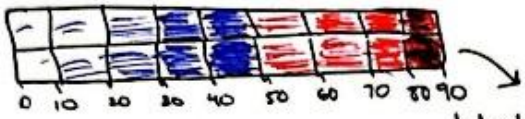
Sheet 1 Hassan Younas 01.12.2021



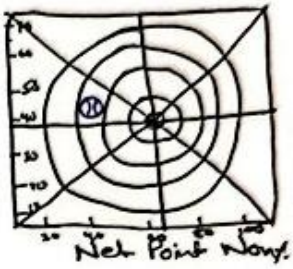
#1 Racket Leader

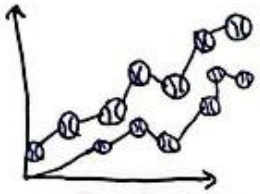
Ranking of Players





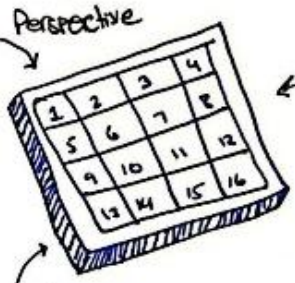
Highest Number of Streaks.





Comparison btw Players

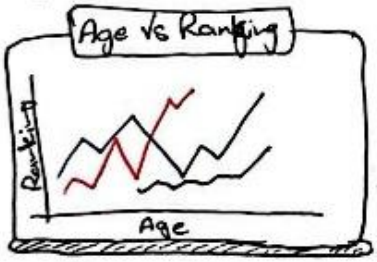
Player's Perspective



Yearly Calendar

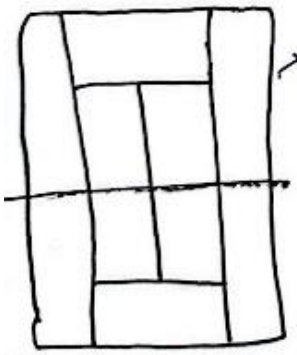
Colour density

Do Age affect Player's Ranking?

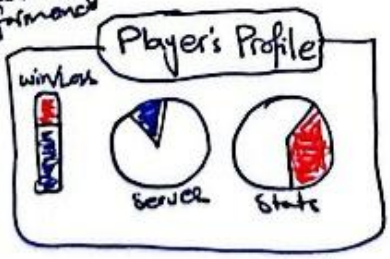


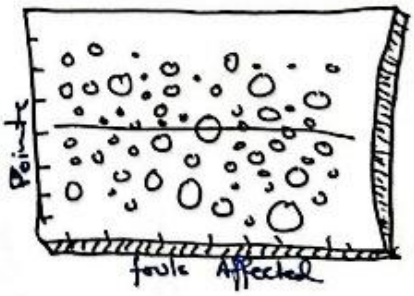
World map

Country dominance?



how surface affect Player performance



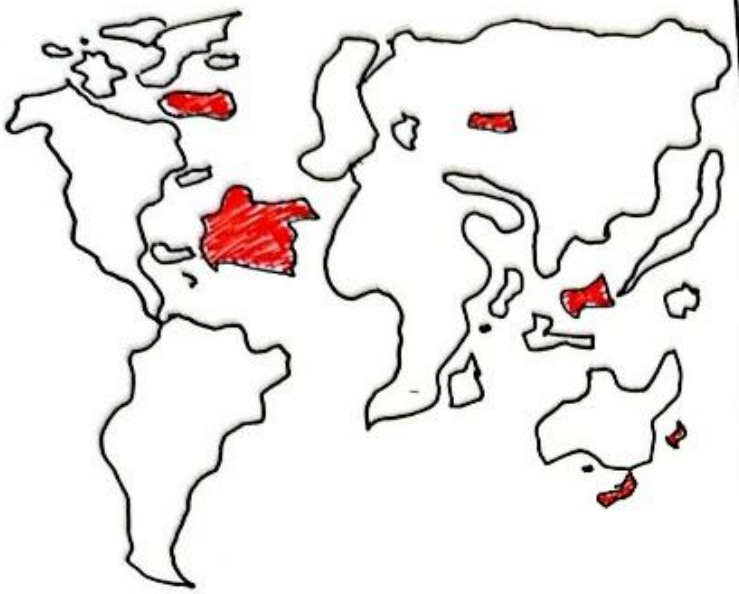


Surface Analysis

grass Vs hard

Which Player is best in which surface.

Big Picture/Layout



WORLD MAP CHOROPLETH
→ This design explores which country dominates ATP tennis ranking data.

Sheet 2

HASSAN YOUNAS

World Map Visualization

02.12.2025

COMPONENTS

A. Country Representation
- A world map, with intensity = dominance score

B. Data Input
- Total Player in dataset from each Co
- Number of Players in TOP K50, TOP 50 or To
- Total Ranking Points accumulated by Cou
- Grand Slams win
- Avg ranking per country

C. Visual Encoding
- Colour intensity = Strength/Dominance
- Highlight size by winning

D. Time slider year by year

PROS & CONS

Pros:-
- Global Perspective
- Easy to understand
- Make Ranking Pattern obvious
- Strong storytelling
- Visually attractive

Cons:-
- Country Population bias
- Map Clutter
- Combining metric will be tricky
- Geographic Maps can mislead as area ≠ tennis strength
- A time-based component may be needed, but maps don't show time naturally.

FOCUS

Key Question:-

① which countries Produced the highest-ranked Players over time?

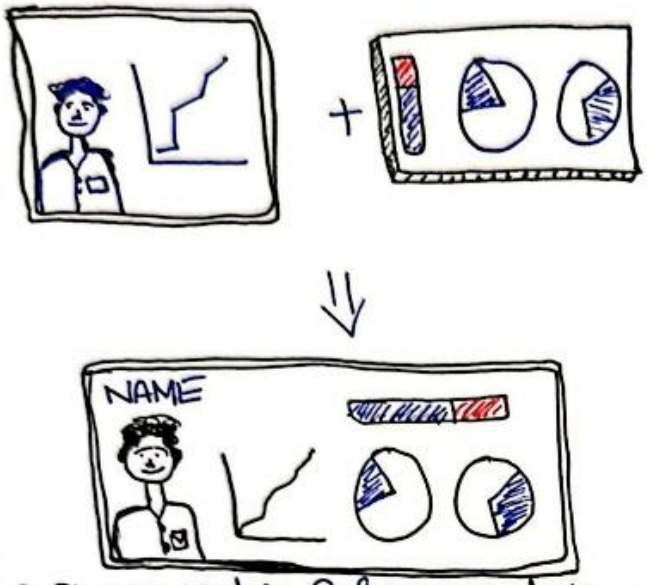
② which countries consistently appear in the top 10/top 50/Top 100?

③ Are certain countries improving or declining in tennis performance?

④ Do smaller countries produce elite Players?

⑤ How do big nations (USA, Spain, France, Germany, UK) compare?

Big Picture



A Player-centric Performance dashboard that shows an individual tennis Player's Picture, ranking, match stats & career summary.

Focus

↳ To showcase how a single ATP Player has performed throughout their career.

↳ To show metrics such as ranking over time, matches won/lost, aces, double faults, surface Performence, and yearly summary.

↳ To Present an engaging, magazine-style Profile combining Player Photograph + visual Analytics

Sheet 3

HASSAN YOUNAS

Player Profile Performance Dashboard

02.12.2025

COMPONENTS

A. Player Profile Section:-
- ↳ Player Picture
- ↳ Name
- ↳ Country flag
- ↳ Career high ranking

B. Ranking Timeline (Main chart)

C. Yearly Performance Bars (Barchart)

D. Surface Performance (Pie chart)

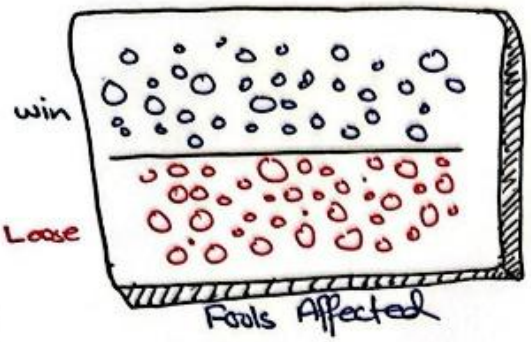
E. Visual Encoding
- ↳ Colour brewer safe Pelette
- ↳ Photo-humanized the data
- ↳ Icons

Pros & Cons:-

Pros:-
- Very engaging & visually Appealing
- Combines Identity + Stats
- Storytelling is strong
- Shows full career Progression
- Easy to compare Players
- Uses multiple visualization types

Cons:-
- Too much information may feel dense
- Not suitable for comparing many Players at once.
- requires clean, complete stats per Player.
- May overwhelm viewers if not designed carefully.

Big Picture layout



This design examines whether there is a relationship blw foul-related statistics and match outcomes. A scatterplot is used to visually compare foul counts to match results. A central reference line divides the Plot Points above the line indicate match won & vice versa

Sheet 4

HASSAN YOUNAS

Relationship b/w Fouls & Match Winning

03.12.2025

Components

A. Scatterplot Layout
   - X-axis: Foul counts
   - Y-axis: Match outcome

B. Central Decision Line

C. Colour Encoding
   - Blue Point → wins
     Red Point → Loose

D. Size of Points, based on no. of fouls per match

Focus

↳ To understand whether high foul counts reduce the chance of winning.

↳ To compare match outcomes visually against a foul metric

↳ To compare Patterns across multiple matches and Potentially across many Players.

↳ To show correlation blw fouls and the outcome clearly.

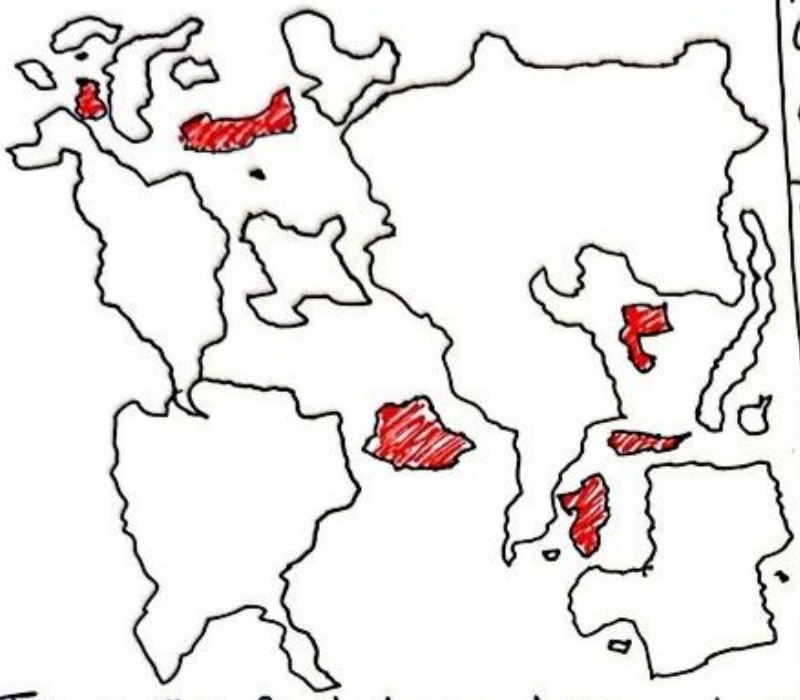
↳ To analyse Player discipline and its effect on Performance

Pros & Cons

Pros:
- Very clear statistical relation visualised.
- Scatterplot is ideal for spotting correlation
- Easy to identify outliers
- shows actionable insight.
- Clean, simple, and interpretable visualisation.

Cons:
- Not Good for storytelling
- Only shows two stats
- Doesn't capture long trends (only per match)
- May clutter.

Big Picture



This is the final chosen design: a choropleth world map showing tennis dominance by country.

SHEET 5

HASSAN YOUNAS

ATP Country Dominance Choropleth

03.12.2025

Components

A. Choropleth Map (center)
- Countries shaded based on dominance score
- Dark, Light, Grey colour

B. Tooltip Box
- When hovering a country it will show country name number of Players, Avg ranking, #Top 10, highest ranking.

C. Visual Encoding
- Color Intensity = Strength
- Highlight size by winning

D Time slider year by year

Details

① Algorithims:
↳ Aggregation Algorithm and Normalization will be done.

② Dominance Score:-

$$DS = \frac{\text{Total Ranking Point} + \#\text{TOP}}{\text{Number of Players}}$$

③ Dependencies:-
- Python
- Pandas
- MatPlotlib
- Numpy
- Tableau (optional)

④ Estimated Time:-
- Data Cleaning 3-4 hrs
- Map generation 2-3 hrs

Focus

Key Questions:-

① Which countries Produced the highest-ranked Players over time?

② Which countries consistently appear in the top 10/top 50/ top 100?

③ are certain countries improving or declining in tennis performance?

④ Do smaller countries produce elite Player?

⑤ How do big nations (USA, UK, Spain, France, Germany) compare

⑥ Geographic Pattern of tennis strength

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

Step 3

Sum values 64

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

ACKNOWLEDGMENTS

I wish to express my deepest gratitude to Jeff Sackmann for providing the ATP tennis dataset, which served as the cornerstone of this analysis. I am also profoundly grateful to my supervisor, Jonathan C. Roberts, for his expert guidance, constructive feedback, and unwavering support throughout the development of this project.

REFERENCES

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