UEFA Competition Referee Analysis Tool
Group: 4

Fannar Freyr Jónuson   Leonard Hastrup   Petros Tsitsekidis   Szymon Zalewski

December 12, 2025

1 Introduction

In the ecosystem of elite football, UEFA competitions represent the pinnacle of the sport. Referees selected for these tournaments, the Champions League, Europa League, and Conference League, are tasked with managing games involving diverse playing styles, high cultural stakes, and immense pressure. A single decision in these continental matchups can determine not just the outcome of a match, but the financial and historical trajectory of a club. However, despite the elite status of these officials, the discourse surrounding their performance often remains subjective. Fans and pundits frequently debate whether referees adjust their strictness based on the prestige of the tournament: are they more lenient in the "showcase" Champions League compared to the lower-tier Conference League?

The motivation for this project stems from the difficulty in validating these perceptions. While raw disciplinary statistics exist, they are rarely presented in a way that allows for a fair comparison across the three distinct tiers of UEFA competition. A simple count of yellow cards is insufficient to judge a referee's performance without context; one must account for the number of appearances and the specific tournament level to understand the referee's true "personality." Furthermore, varying nationalities of referees add another layer of complexity, do officials from certain regions enforce the rules more strictly when performing on the European stage?

To address this, our project develops an interactive visualization system that explores refereeing behaviors across the three major UEFA competitions spanning the 2021-2025 seasons. By synthesizing raw data (cards, penalties, appearances) into a derived "Strictness Score," we aim to quantify officiating patterns that are usually left to anecdote.

The goal of this visualization is to allow users to investigate the consistency of referees across the UEFA hierarchy. It enables the exploration of key questions: Do specific referees become "stricter" as they move down the tiers from the Champions League to the Conference League? Are there visible correlations between a referee's age or nationality and their likelihood to issue cards? By visualizing these metrics, we provide an objective, analytical lens on the men in the middle of Europe's biggest matches.

2 The Visualization Task

To design an effective visualization, we utilized the nested model for visualization design [4]. First, we analyzed the problem domain by abstracting the data into types and defining the abstract tasks the user needs to perform.

2.1 Data Abstraction

The underlying dataset is a tabular collection of disciplinary statistics for football referees officiating in UEFA competitions (Champions League, Europa League, and Conference League) across four seasons (2021-2025). The data consists of Items (individual referee records per season/league) and Attributes (characteristics of those records).

Categorical (Nominal)

- Name (Identifier).

- Nationality (e.g., "English", "German").

- League (The context of the Performance: UCL, UEL, UECL).

Quantitative (Ratio)

- Age (Sequential data allowing for correlation analysis).

- Appearances, Yellow Cards, Double Yellow Cards, Red Cards, Penalties (Counts start from zero).

Derived Attribute: The Strictness Score

Raw counts of cards are misleading without context (e.g., 5 red cards in 20 games is different from 5 red cards in 5 games). To solve this, we derived a quantitative attribute called Strictness Score. This provides a normalized single metric to rank officials.

The formula is defined as a weighted sum of disciplinary actions divided by the number of appearances:

$$\text{Strictness} = \frac{(W_y \times \text{Yellows}) + (W_{dy} \times \text{DoubleYellow}) + (W_r \times \text{Reds}) + (W_p \times \text{Penalties})}{\text{Appearances}}$$

(Note: In our implementation, we assigned default weights based on the severity of the punishment, e.g., Yellow=1, DoubleYellow=3, Red=5, Penalty=3. However they are user adjustable.)

2.2 Task Abstraction

The visualization is designed to support three primary high-level goals: Discover (finding unknown patterns), Derive (generating the strictness metric), and Compare (evaluating differences).

We broke these down into specific domain tasks:

1. **Summarize & Overview (Global Analysis):**

   - Abstract Task: Summarize the distribution of strictness across the entire dataset.
   - User Goal: The user needs to see the "All 3 Leagues" view to understand the general baseline of officiating in Europe before drilling down. This answers the question: "What is the average strictness of a European referee regardless of the tournament?"

2. **Compare Distributions:**

   - Abstract Task: Compare attributes between derived clusters.
   - User Goal: Compare the strictness levels between the Champions League and the Conference League. Are referees significantly harsher in lower-tier competitions?

3. **Identify Extremes (Outliers):**

   - Abstract Task: Identify items with extreme values.
   - User Goal: Spot the single "strictest" referee in the dataset or find a referee who gave an unusually high number of penalties in a single season.

4. **Correlate Attributes:**

   - Abstract Task: Lookup and Correlate quantitative values.
   - User Goal: Determine if Age correlates with Strictness. (e.g., "Do referees mellow out as they get older?").

By explicitly supporting the transition from a Summarized View (all leagues) to a Filtered View (specific league), the tool allows the user to verify if global trends hold true in local contexts

3   Related Work

The analysis of officiating behavior has historically focused on statistical regression models to detect bias rather than interactive exploration. Academic studies have extensively documented "home advantage" and "social pressure" on referees using static datasets [1, 3]. However, in the public domain, stakeholders typically rely on databases like Transfermarkt or FBref [SOMEONE FIND THIS REFERENCE PLEASE]. While comprehensive, these platforms present data in dense, siloed tables that obscure trends and make cross-league comparison difficult.

Our project bridges this gap by applying Information Visualization principles to this domain. We drew inspiration from the "Overview first, zoom and filter" paradigm [5] to structure our dashboard. For specific visual encodings, we adapted the "player radar" concept often used in industry analytics (e.g., StatsBomb) into a "Referee Passport". Notably, we utilized Waffle Charts instead of traditional Pie Charts for our "Card DNA" view. This decision is grounded in graphical perception research [2], which demonstrates that human observers estimate proportions more accurately in grid-based layouts than via angular slices, particularly when comparing small values like red cards.

4   The Visualization Solution

Our solution is implemented as a single-page React application using D3.js for rendering complex vector graphics and Material UI for the control interface. The design follows Shneiderman's Information Seeking Mantra: "Overview first, zoom and filter, then details-on-demand [5]". All of these components are linked together such that when the user chooses a subset of the data, then that same subset is displayed on the rest of the views.

4.1   Overview

The dashboard is divided into three logical zones:

1. **Global Controls (Top Bar):** A persistent control panel allows users to filter the dataset by **Season** (2021-2025), **Nationality**, **Referee Name**, **Age Range**, and **Experience (Appearances)**. A global settings dialog allows users to adjust the weights of the "Strictness Formula" dynamically, empowering them to define what "strict" means to them (e.g., weighting red cards higher than penalties).

2. **Macroscopic Views (The Overview):** For the general overview of how the data distributes between the different leagues, we have chosen to place an upset plot at the top. This gives the user direct access to which subsets it is possible to look into, when comparing the different leagues to eachother.

3. **Comparative View** The next section features two coordinated views:

   - **Geographic Strictness Map:** A choropleth map showing average strictness per country.
   - **Population Scatterplot:** A distribution plot placing every referee on an Age/Experience vs. Strictness plane.

4. **Microscopic (Details-on-Demand):** The lower section adapts based on the user's selection:

   - No Selection: Shows general trends.
   - Multi-Selection: Activates the **League Comparison** Scatter (League X vs. League Y).
   - Single Selection: Activates the **Referee Passport** (Heatmap + Waffle Chart) for deep profiling.

4.2   Design Decisions & Justification

We grounded our design choices in visualization theory, specifically focusing on effective mark and channel usage.

A. **Strictness Encoding (Geographic & Correlation Views)** For the quantitative "Strictness" attribute in our main overview displays, we utilized a **Sequential Single-Hue Color Scale** (Light Aqua `#8de4d3` → Medium Green `#a0d66f` → Dark Green `#1c5e39`).

- **Justification:** We transitioned to a colorblind-safe sequential scale to improve accessibility. Since strictness is a magnitude, we utilize **luminance** to encode value: lighter shades represent leniency, while darker shades represent high strictness. This ensures the pattern is visible to all users, avoiding common red-green deficiencies.

B. **The Geographic View**

- **Encoding:** We used a choropleth map where countries are colored by the average strictness of their referees (using the sequential green scale).
- **Interaction:** We implemented Linked Filtering. Clicking a country (e.g., Spain) acts as a global filter, instantly updating the scatterplots to show only Spanish referees.
- **Handling Sparsity:** To avoid misleading averages from countries with only 1 referee, we added a "Minimum Referees" slider and a "Region Mode" toggle (aggregating countries into regions like "Nordic" or "Eastern Europe").

C. **The Correlation View**

- Marks & Channels: Each referee is a point.
  - **X-Position:** Mapped to Experience (Appearances) or Age (user-toggled). We chose position because it is the strongest channel for quantitative comparison.
  - **Y-Position:** Mapped to Strictness Index.
  - **Color:** Mapped to Strictness (redundant encoding to reinforce the pattern).
- **Occlusion Handling:** Non-selected referees are rendered with low opacity (0.3), while selected referees use high opacity (1.0) and a stroke border. This "Focus+Context" technique ensures the user can see their selection within the global distribution.

D. **The Cross-League Comparison**

- **Problem:** Users want to know if a referee changes their behavior between leagues (e.g., "Are they stricter in the Champions League than the Europa League?").
- **Solution:** A scatterplot where X = League A Strictness and Y = League B Strictness.
- **Reference Line:** We added a $y = x$ diagonal line. Points above the line indicate referees who are stricter in League B; points below are stricter in League A. This geometric aid allows for instant assessment of bias.

E. **Single Referee Profiling** When a single referee is selected, we provide a "Visual Identity Card":

(a) **The "Passport" (Heatmap):** A matrix view (Rows: Competitions, Columns: Seasons) colored by strictness. This reveals temporal trends (e.g., "They were lenient in 2021 but strict in 2023").

(b) **The "Card DNA" (Waffle Chart):** A 10 x 10 grid representing the percentage split of disciplinary actions.

- **Justification:** Unlike the strictness views, we retained the **semantic categorical colors** (Yellow for cautions, Red for dismissals) here. We chose a Waffle Chart over a Pie Chart because square grids allow for more accurate area comparison, while the semantic colors ensure immediate recognition of specific card types without needing to reference the abstract strictness scale.

5 Use Case

The user starts at the upset plot then sees that there is 60 referees, who have data from all three different leagues. This is something they find quite interesting, so the click on it and thus filters that data for the rest of the dashboard. They then go down to the choropleth map, and finds it interesting there are large countries like Russia that is not in this subset, prompting them to analyze Russian referees further. They thus click on Russia, and back in at the upset plot concludes that they represented in the champions league. Next observation is that Russia's strictness is comparable to other countries, so maybe this is a general thing for Eastern Europe. They thus switch to the region view, and wants to compare eastern and western europe to eachother. At first glance western europe seems more lenient, but looking at the scatterplot, this is likely caused by the low number of apperances for the very strict referees. They thus alter the minimum appearances and sees that they are in fact quite similiar to eachother. Then going further with the minimum appearnces they can thus see that there generally is a higher amount of reappeances for western referees. Which they then can continue to drill down on, also altering the strictness level to get a better idea of the tendency of the regions.

Figure 1: Visualization showing the strictness discrepancy for English referees.

6 Limitations

While our visualization successfully enables the comparative analysis of referee strictness, it is subject to specific limitations regarding the underlying data and visual scalability.

Data & Metric Limitations

First, our "Strictness Score" algorithm is a linear weighted sum that does not capture the qualitative context of a foul. The data treats a "tactical" yellow card for time-wasting identically to a yellow card for a reckless tackle, meaning we cannot distinguish between strict rule enforcement and a loss of match control. Second, the model does not account for confounding external variables, such as the playing style of the teams involved; a referee may appear stricter simply because they officiated more aggressive teams. Finally, the geographic view suffers from sample size bias. Smaller nations often have only one international referee, meaning a single card-heavy performance can skew an entire country's color coding.

Visual & Technical Limitations

Visually, the "Global Overview" scatterplot suffers from occlusion when all seasons are active, as hundreds of data points cluster in the center, potentially hiding individual records. Additionally, while the user-configurable weights allow for flexibility, the default setting ($Yellow = 1, Red = 3$) imposes an initial analytical bias that may not reflect every user's perception of severity. Lastly, our temporal heatmaps show strictness per season but lack granular match sequencing, preventing users from seeing if a referee's behavior changes immediately following a controversial decision.

7 Division of Labor

- **Fannar** took part in the visualisation design, and designed and implemented the scatterplot and the grouped bar chart and their interactions.

- **Leonard** took part in visualisation design and implemented/designed the choropleth map and upset plot together with their interactions.

- **Petros** took part in the data collection, data preprocessing, exploratory data analysis, and worked on general issues that popped up in the integration of different visualisation components.

- **Szymon** created the initial project structure including layout, took part in visualization designed and implemented referee specific charts (Referee Passport and Card DNA). Also was involved in additional improvements and fixes that were needed during the project work and after feedback.

8   Bibliography

References

[1] Ryan H. Boyko, Adam R. Boyko, and Mark G. Boyko. Referee bias contributes to home advantage in english premiership football. *Journal of Sports Sciences*, 25(11):1185–1194, 2007. PMID: 17654230.

[2] William S. Cleveland and Robert McGill. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387):531–554, 1984.

[3] Peter Dawson and Stephen Dobson. The influence of social pressure and nationality on individual decisions: Evidence from the behaviour of referees. *Journal of Economic Psychology*, 31(2):181–191, 2010. The Economics and Psychology of Football.

[4] Tamara Munzner. A nested model for visualization design and validation. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):921–928, 2009.

[5] Ben Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *IEEE Symposium on Visual Languages, 1996*, pages 336–343, 1996.