

Tennis Plots: Game, Set, and Match

Michael Burch and Daniel Weiskopf

Visualization Research Center, University of Stuttgart,
Allmandring 19, 70569 Stuttgart

{michael.burch,daniel.weiskopf}@visus.uni-stuttgart.de

Abstract. In this paper we introduce Tennis Plots as a novel diagram type to better understand the differently long time periods in tennis matches on different match structure granularities. We visually encode the dynamic tennis match by using a hierarchical concept similar to layered icicle representations used for visualizing information hierarchies. The time axis is represented vertically as multiple aligned scales to indicate the durations of games and points and to support comparison tasks. Color coding is used to indicate additional attributes attached to the data. The usefulness of Tennis Plots is illustrated in a case study investigating the tennis match of the women's Wimbledon final 1988 between Steffi Graf and Martina Navratilova lasting 1 hour, 19 minutes, and 31 seconds and being played over three sets (5:7, 6:2, 6:1). Interaction techniques are described in the case study in order to explore the data for insights.

Keywords: time-varying data, sports data, hierarchical data.

1 Introduction

Tennis matches are time-dependent sports events which can be subdivided into several playing phases. Finding insights in such time-varying data is difficult by just inspecting the data manually in a text file or by just watching the match, which can sometimes last a couple of hours. Statisticians or visualizers often analyze the data by aggregating the whole match or single sets into some statistical numbers which do not allow one to find insights in the evolution of the match. But exploring time-dependent sports data is interesting for sportsmen, managers, the mass media, as well as for the spectator in order to get informed by a simple all-showing diagram.

In this paper we introduce Tennis Plots: an interactive visualization technique that visually encodes the playing and pausing phases of a tennis match in a multiple aligned scales representation similar to layered icicle plots. Such a visualization first gives an overview of the complete match and simultaneously shows different phases and the match structure. A static diagram for dynamic data has many benefits due to mental map preservation, reduction of cognitive efforts, good performance of comparison tasks, and the application of interaction techniques [1].

In this visualization design we map the scored points as well as the time spent to score them and the pauses as differently thick and color coded rectangles depending on the durations. We visually map the time dimension to the vertical axis that is used to visualize the durations of the single events such as the time until a point is scored and the pauses in-between. The horizontal axis is also used as time axis showing the overall temporal structure of the tennis match in the form of sets. Interaction techniques are integrated targeting a better exploration of the time-varying data.

2 Related Work

Analyzing time-varying data can be challenging, depending on the type of data, the characteristics of the time axis, and the applied visual metaphor [2]. In particular, tennis matches are based on a time axis which can be split into a sequence of intervals which is also characteristic for calendar-based representations [3]. Klaasen and Magnus [4] propose a method for forecasting tennis match winners. They do not base their analysis on the beginning of the match but in particular during the running match. In the work of Terroba et al. [5] a tennis model based on Markov decision processes is presented for describing the dynamic interaction between tennis players. The goal of the work is to extract optimal strategies. Also a transformation of tennis videos into time-varying discrete data is proposed.

Visualization techniques are rarely applied strategies but are a good means to uncover visual patterns which can give insights into interesting phenomena hidden in time-varying data. For example, Burch and Weiskopf [6] illustrate and classify time-varying patterns by a visualization technique investigating dynamic water level data which are of quantitative nature. In particular, for tennis matches, the TennisViewer [7] by Jin and Banks has been proposed. Although the hierarchical structure of the tennis match is visible combined with a color coding on treemap [8] [9] boxes to indicate which of the two players scored which point, the comparison of multiple time axes is difficult. In our work we add explicit timelines to the hierarchical organization of the tennis match in order to derive time-varying insights on different levels of temporal granularity.

Jin and Banks [7] use competition trees to organize the tennis match. They also follow the idea of encoding the hierarchical structure of the match into layered representations similar to layered icicle plots [10] used for hierarchy visualization. However, in their work, the explicit temporal aspect in the match is not displayed, i.e. we see which point was scored before which other, but there is no absolute time axis integrated telling the viewer about the actual spent time when either playing or pausing in the match. Tu and Shen [11] also apply the concept of treemaps to illustrate time-varying data by adding explicit timelines to the treemap boxes, but the timeline is directly integrated into each treemap box making interval comparison tasks hard to solve visually.

We, instead, rely on a time-to-space mapping of the time-varying data, which has several benefits compared to animated diagrams. Comparisons, which are important tasks in visualizations, are done visually which would be problematic

with animation. The concept which we follow is denoted as mapping to identical aligned scales [12]. This data mapping strategy is ranked very high for comparing quantitative data values. In our work we follow the Visual Information Seeking Mantra [13]: Overview first, zoom and filter, then details-on-demand.

3 Visualization Technique

In this paper we describe a visualization technique for exploring a tennis match on different levels of structural granularity, i.e. either on match, set, game, or score levels.

3.1 Data Model

We model a tennis match as an ordered finite sequence S_I of $n \in \mathbb{N}$ intervals $S_I := \{I_1, \dots, I_n\}$. Each single interval $I_i \in S_I$ is attached by two timestamps t_{s_i} and t_{e_i} , expressing start and end time points for that interval I_i . The division into intervals depends on the events occurring in a tennis match, i.e. different playing and pausing phases, which again depend on the performances of the single players.

By this data model we subdivide the time axis into non-overlapping intervals which completely cover the time axis, i.e. $t_{e_i} = t_{s_{i+1}} \forall 1 \leq i \leq n - 1$ and $t_{s_i} \leq t_{e_i} \forall 1 \leq i \leq n$. It may be noted that if we allow $t_{s_i} = t_{e_i}$ also time points instead of time intervals can be modeled. The duration of an interval I_i is defined as $t_{d_i} := t_{e_i} - t_{s_i}$.

Moreover, each interval I_i has additional data attributes which can be modeled as an ordered list $L_i := \{a_1, \dots, a_m\}$. The a_i 's can for example be used to attach the current score in the match, additional events such as faults, double faults, net faults, aces, pauses, breaks, or the player who is to serve in this game.

The attached score data is taken to derive structural timeline subdivisions, i.e. a tennis match consists of sets, games, and the single scores. This allows us to generate a hierarchical organization of the match while still preserving the chronological order between the intervals S_I .

Moreover, this data model can be used to compute derived data values based on several filter criteria, for example the average duration of pauses taken by a specific player, the maximum length until a point is scored, or the sum of durations for single sets or games, i.e. derived data on different structural granularities.

3.2 Design Decisions

To design a suitable visualization technique for such time-varying data with attached data attributes having an inherent hierarchical structure we first illustrate how the sequence of intervals S_I is visually encoded.

Figure 1 (a) shows the vertical stacking of time intervals $S'_I \subseteq S_I$. Each interval $I_i \in S'_I$ is mapped to a rectangle where the width is fixed and the

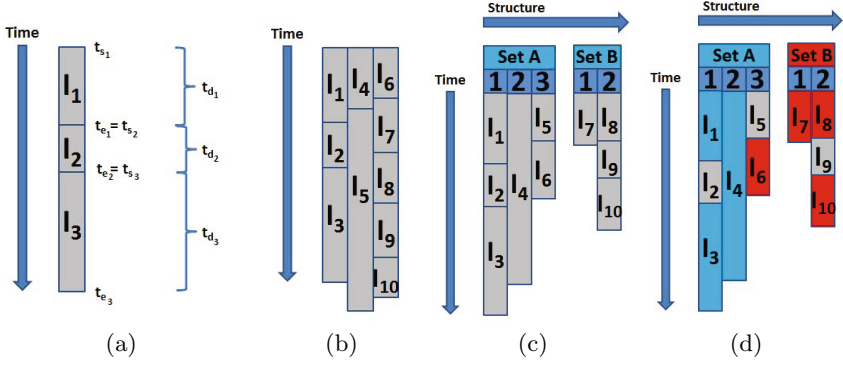


Fig. 1. Generating a Tennis Plot: (a) Stacking of time intervals. (b) Multiple sequences of time intervals. (c) Combination of time axes and the match structure. (d) Color coding to visualize categorical data.

height encodes the duration t_{d_i} of I_i . This visualization design allows the user to visually explore a number of intervals by inspecting the sum of heights of all intervals stacked on top of each other.

If we have to deal with m sets of intervals, i.e. $\mathbb{S} := \{S_{I_1}, \dots, S_{I_m}\}$ where each $S_{I_k} \subseteq S_I$ we need a representation that supports comparison tasks among all interval sets. Figure 1 (b) illustrates how we design such a visual representation. The single S_{i_k} are mapped in chronological order by their start time points from left to right in order to achieve a readable diagram. The sum of durations for each interval sequence can be inspected while additionally allowing comparisons among all of them.

The additional structure of the tennis match, i.e. the organization into sets, games, and scores is illustrated in Figure 1 (c). We use the vertical direction for representing time and the horizontal direction to represent the additional match structure, which has also a temporal nature similar to a calendar-based representation [3], i.e. the division into time intervals by preserving the inherent match structure.

Color coding is used as the final design step to map additional attributes to the plot. Typically, this additional data is of categorical nature, e.g. the single players, pausing and playing phases, or net faults and faults (see Figure 1 (d)).

Further information can easily be added to such a plot aligned to each timeline. Such data attachments are for example derived values such as average set durations, game duration, average pause durations, or average time for waiting until a player will serve for the next point (for each player separately) and the like.

We support several interaction techniques to browse, navigate, and explore the data. Some of the most important ones are event, time, structure, and text filters. Moreover, we support time normalization, logarithmic scaling, zooming, color codings, highlighting, and details-on-demand. Also additional derived values such as average, maximum, minimum and the like are displayable by statistical plots.

4 Case Study

We illustrate the usefulness of Tennis Plots by first transforming time-varying tennis match data into a specific data format combined with additional events. In the second step we visualize this data by the novel diagram and finally explore it by interaction techniques to find interesting insights which are hard to be found by either watching the tennis match or browsing through the textual time-varying data.

In our case study, we inspect the women’s Wimbledon final of 1988 between Steffi Graf and Martina Navratilova, which lasted 1 hour, 19 minutes, and 31 seconds. The match was played over three sets, resulting in 5:7, 6:2, and 6:1.

There are several visual patterns which can be detected when inspecting the plot. Those can be classified by either looking at single time axes, comparing a certain number of time axes, and by investigating the additional match structure.

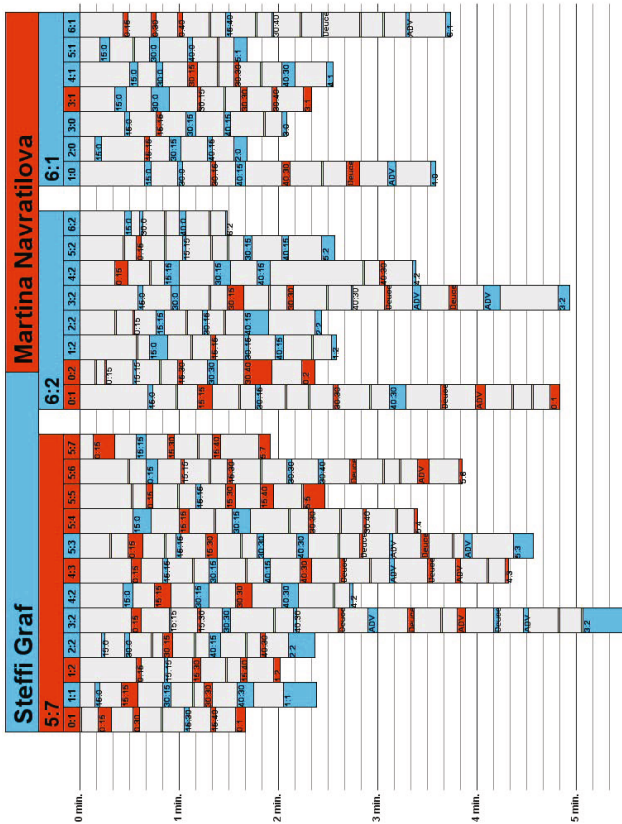


Fig. 2. Tennis Plot of the women’s Wimbledon final in 1988 between Steffi Graf and Martina Navratilova. Time axes are integrated and the single point durations are shown side-by-side.

When looking at single time axes in Figure 2, we directly see that the pausing phases are much longer than the playing phases during a tennis match. This can be done for each time axis separately. Apart from doing this inspection visually the tool can also show additional derived values for average durations of both categories of phases which is not shown in this figure. Looking at those additional statistics we can confirm our first impression obtained by reading the Tennis Plot. Moreover, if we have a look at single time axes we see the sequence of points whereas the color coding and the labeling information helps to rapidly understand which player scored at what time and also in which sequential order.

Comparing several time axes is beneficial to explore the different durations taken until one game is over. Here, we can for example see that the 3:2 game in the opening set took much longer than all other games (Figure 2). The 6:2 game in the second set took the least time in the match, which can be observed by looking at the timeline integrated into the plot. For this game we also directly see that Steffi Graf (blue color) won by scoring four points in a row and all of them in a very short time. To make these observations time-based comparison tasks have to be conducted first and then the single time axes have to be visually explored.

When taking the match structure into account, i.e. the subdivision into points and sets, we can furthermore see that the first set took the most time of all sets. This insight is obtained by interactively stacking all intervals for each set separately on top of each other. Moreover, Martina Navratilova won the opening set but she lost the second and the third one. Looking at the time axes in Figure 2 we can see that Steffi Graf wins more and more points the more the match is coming to an end. Asking for details-on-demand again, we get the information that the number of won points changes from 38:46 (first set) to 31:18 (second set), and finally to 28:14 (third set), i.e. in the last set Steffi Graf won twice as many points as Martina Navratilova. Here we can detect in a more efficient way that the number of Steffi Graf's scored points is increasing towards the end of the match, i.e. in the second and third set.

5 Conclusion and Future Work

In this paper we introduced Tennis Plots as an interactive diagram supporting a viewer in inspecting time-varying data acquired from tennis matches. Instead of showing the data as statistical numbers the diagram focuses more on depicting the data on an interval basis. Playing and pausing phases as well as additional events and attributes can be visually inspected together with the match hierarchy, i.e. the subdivision into points, games, and sets. Interaction techniques are integrated to browse, navigate, and explore the data on different levels of time and structural granularities.

Although this approach is able to solve some exploration tasks, it might be extended by additional features and visual components. For example, a direct combination to the actual video of the match might be of interest to help users to directly browse to a specific point in time in the match in which they can get

better impressions of the real data and what actually happened there. Such detail information cannot be provided by the Tennis Plots. Moreover, other sports data might be of interest to be visually analyzed such as volleyball, basketball, and the like. Apart from visualizing sports data on an interval basis also more general time-varying data with an additional hierarchical structure might be interesting to be represented. For more general scenarios we are aware of the fact that additional problems will occur such as finding a suitable subdivision into intervals. Finally, a user study should be conducted investigating the performances of participants when analyzing data by these plots.

Open Access. This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Tversky, B., Bauer Morrison, J., Bétrancourt, M.: Animation: Can it Facilitate? *International Journal of Human-Computer Studies* 57(4), 247–262 (2002)
2. Aigner, W., Miksch, S., Schumann, H., Tominski, C.: *Visualization of Time-Oriented Data*. Springer (2011)
3. van Wijk, J.J., van Selow, E.R.: Cluster and Calendar Based Visualization of Time Series Data. In: *Proceedings of Infovis*, pp. 4–9 (1999)
4. Klaassen, F.J.G.M., Magnus, J.R.: Forecasting the Winner of a Tennis Match. *European Journal of Operational Research* 148(2), 257–267 (2003)
5. Terroba Acha, A., Kosters, W.A., Varona, J., Manresa-Yee, C.: Finding Optimal Strategies in Tennis from Video Sequences. *IJPRAI* 27(6) (2013)
6. Burch, M., Weiskopf, D.: Visualizing Dynamic Quantitative Data in Hierarchies – TimeEdgeTrees: Attaching Dynamic Weights to Tree Edges. In: *Proceedings of International Conference on Information Visualization Theory and Applications*, pp. 177–186 (2011)
7. Jin, L., Banks, D.C.: TennisViewer: A Browser for Competition Trees. *IEEE Computer Graphics and Applications* 17(4), 63–65 (1997)
8. Johnson, B., Shneiderman, B.: Tree Maps: A Space-Filling Approach to the Visualization of Hierarchical Information Structures. In: *Proceedings of IEEE Visualization*, pp. 284–291 (1991)
9. Shneiderman, B.: Tree Visualization with Tree-Maps: 2-D Space-Filling Approach. *ACM Transactions on Graphics* 11(1), 92–99 (1992)
10. Kruskal, J., Landwehr, J.: Icicle Plots: Better Displays for Hierarchical Clustering. *The American Statistician* 37(2), 162–168 (1983)
11. Tu, Y., Shen, H.-W.: Visualizing Changes of Hierarchical Data Using Treemaps. *IEEE Transactions on Visualization and Computer Graphics* 13(6), 1286–1293 (2007)
12. Cleveland, W.S., McGill, R.: Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *Journal of the American Statistical Association* 79(387), 531–554 (1984)
13. Shneiderman, B.: The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In: *Proceedings of the IEEE Symposium on Visual Languages*, pp. 336–343 (1996)