Georgia State University

ScholarWorks @ Georgia State University

Computer Science Dissertations

Department of Computer Science

5-4-2021

Visual Analytics Of Sports Data

Shiraj Pokharel

Follow this and additional works at: https://scholarworks.gsu.edu/cs_diss

Recommended Citation

Pokharel, Shiraj, "Visual Analytics Of Sports Data." Dissertation, Georgia State University, 2021.
https://scholarworks.gsu.edu/cs_diss/173

This Dissertation is brought to you for free and open access by the Department of Computer Science at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Computer Science Dissertations by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

VISUAL ANALYTICS OF SPORTS DATA

by

SHIRAJ POKHAREL

Under the Direction of Ying Zhu, PhD

ABSTRACT

In this dissertation, we discuss analysis and visualization of performance anxiety in tennis matches along with confidence and momentum. We also discuss the micro-level analysis and visualization of tennis shot patterns with fractal tables and tactical rings, followed by discussion about mapping a tennis player's style of play with a visual analysis technique called tennis fingerprinting.

According to sports psychology, anxiety, confidence and momentum has a big impact on an athlete's performance in a sport event. Although much work has been done in sports data analysis and visualization, analysis of anxiety, confidence and momentum has rarely been included in recent literature. We propose a method to analyze a tennis player's anxiety level, confidence and momentum levels during a tennis match. This method is based on the psychological theories of anxiety and a database of over 4,000 professional tennis matches. Since sports data analysis and visualization can be a useful tool for gaining insights into the games, we present new techniques to analyze and visualize the shot patterns in tennis matches via our Fractal Tables and Tactical Rings. Tennis is a complicated game that involves a rich set of tactics and strategies. The current tennis analysis are usually conducted at a high level, which often fail to show the useful patterns and nuances embedded in low level data. However, based on a very detailed database of professional tennis matches, we have developed a system to analyze the serve and shot patterns so that an user can explore questions such as "What are the favorite patterns of this player? What are the most effective patterns for this player?" This can help tennis experts and fans gain a deeper insight and appreciation of the sport that are not usually obvious just by watching the match. Further, we present a new visual analytics technique called Tennis Fingerprinting to analyze tennis players' tactical patterns and styles of play. In tennis, style is a complicated and often abstract concept that cannot be easily described or analyzed. The proposed visualization method is an attempt to provide a concrete and visual representation of a tennis player's style.

INDEX WORDS: Visual Analytics, Data Visualization, Visual Knowledge Discovery, Visual Knowledge Representation, Sports Analytics.

VISUAL ANALYTICS OF SPORTS DATA

by

SHIRAJ POKHAREL

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2021

Copyright by
Shiraj Pokharel
2021

VISUAL ANALYTICS OF SPORTS DATA

by

SHIRAJ POKHAREL

Committee Chair:	Ying Zhu
Committee:	Anu Bourgeois
	David Cheshier
	Yanqing Zhang

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

May 2021

DEDICATION

This dissertation is dedicated to my always encouraging parents - Laximi Pokharel and Pitamber Pokharel, wife - Rachana Luitel and sister - Shivani Pokharel. Thank you for your unconditional love, guidance, and unflagging support.

ACKNOWLEDGEMENTS

This thesis is the result of research carried out over a period of many years. During this period, many people supported my work and helped me to bring it to a fruitful conclusion. The research would not have been possible without the support of most importantly - my advisor Dr. Ying Zhu for his patience and invaluable guidance throughout my graduate studies. I am filled with gratitude to the members of my dissertation committee - Dr. Anu Bourgeois, Dr. David Cheshier, and Dr. Yanqing Zhang for their advice and invaluable time.

TABLE OF CONTENTS

LIST OF FIGURES

PART 1

INTRODUCTION

Data visualization and visual analytics techniques are concerned about the algorithmic ways to communicate data or information via a visual medium. The need for simplification of the processes with regards to understanding complex and mostly obscure trends and patterns in the data is what drives this field forward. Structure, features, relationships in data need to be presented for insights and actionable information. Visual analytics thus demystifies large and complex data sets and guides the process towards qualitative understanding.

1.1 Problem Statement and Motivation

1.1.1 Performance Anxiety

Anxiety can have a great impact on human performance, particularly when the stake is huge, the competition is strong, or the time pressure is high. Perhaps the most visible cases of anxiety influencing performance is in sports events. Sports psychologists have long studied anxiety and considered it the most important factor that influences an athlete's performance. However, the relationship between anxiety and skilled performance in a competitive event is still not well understood. Such analysis is traditionally done by human with relatively small samples. In recent years, data-driven sports analysis have become a rapidly growing area because of the availability of large amount of sports performance data. However, anxiety has rarely been included in sports data analysis and visualizations. This work is an attempt to address this issue, but as with any other science, it is widely felt that such methodologies and tools themselves need continuous development and improvement. For detailed study of an individual player's im-match psychological state and hence modeling , visualizing and analytics of an individual athlete mandates a game which has a global popular appeal. Also the fact that Tennis is not a game that is timed helped us narrow down my choice. This dissertation describes a method to computationally build a tennis player's anxiety model based on detailed performance data. In this model, a player's anxiety is influenced by three main factors: uncertainty, anticipation, and threat. This model can estimate the anxiety level of a tennis player as the tennis match progresses. I have also developed data visualizations to help correlate anxiety with various performance measures, such as unforced errors, forced errors, winners, serve directions, first-serve faults, and double faults. This visual anxiety-performance analytics tool can help sports psychologists, tennis players, coaches, analysts, and fans better understand the relationship between anxiety and performance. For example, is there a correlation between high anxiety level and increased unforced errors? Does a player prefer certain types of serves or shots when the anxiety level is high? Although, I chose to study sports performance anxiety because of the abundance of sports performance data, the proposed anxiety-performance analysis methods and tools can be adapted to other fields if similar data sets are available.

1.1.2 Confidence and Momentum

Confidence and momentum have tremendous impact on sports performance especially in high stakes matches with pressure on performance. Sports psychologists and sports data journalists have long studied confidence and momentum and consider those to be important factors that influences an athlete's performance in a competitive sport event. However, because of the difficulty of measuring them during matches, the specific relationship between confidence and momentum and an athlete's performance in a competitive event is still not fully understood. Data visualization and analytics are being extensively used not just in the analysis of sports for commentators and the coaching staff but also in the form of news dissemination to augment the understanding of fans and general readers. Visual analytics techniques can be useful to understand the role confidence and momentum plays in an athlete's quest to win matches. In recent years, data-driven sports analysis and visual sports data analytics have become a rapidly growing area because of the availability of large amount of sports performance data. However, the practice of including confidence and momentum in visual analytics in sports is not common practice. The contribution in this work is an attempt to address this issue. As with any other science, it is widely felt that such methodologies and tools themselves need continuous development and improvement. Detailed study of an individual player's in-match psychological state with modeling visual analytics of an individual athlete mandates a game which has a global popular appeal. The fact that tennis is not a timed sport helped us narrow down the choice. This dissertation describes a method to build a tennis player's confidence and momentum model based on theories on sports psychology. Specially, in this model, an athlete confidence level is mainly based on an in-match performance with confidence starting out from an average baseline and rising or falling or maintaining the level accomplished, aided or impeded by the opponent's errors or brilliance respectively. I was able to build the confidence and momentum profile of a tennis player as the tennis match progresses. I have also developed data visualizations to correlate confidence and momentum with various performance measures such as unforced errors, forced errors, winners, serve directions, first-serve faults, double faults, aces and rally lengths. I chose to focus on tennis because there is an open source database of over 4,000 professional tennis matches, each of which includes detailed, shot-by-shot descriptions [8]. This analytics method which visually maps confidence and momentum performance can help fans, coaches and tennis players themselves appreciate the relationship between anxiety and performance and its influence in the player's game plan. It can also be used to find out if a player under low confidence and lacking momentum double faults more than normal or else used to ascertain a player's preferred serve direction under varied levels of confidence and momentum. Other example includes, a player and coach using this method to analyze their opponents or the player themselves by exploring a correlation between unforced errors in certain shots with high confidence level and momentum.

1.1.3 Patterns with Fractal Tables

Sports data analysis and visualization can help players, coaches and analysts gain insight into the game and therefore help them improve their performance. Such analysis can also help sports fans appreciate and enjoy the games on a deeper level. Therefore, most sports event broadcasters and web sites often present some statistics at the end of a match. For example, after each tennis match, a TV broadcaster or a news web site will list or sometimes visualize the total number of aces, unforced errors, winners, the serve percentages, etc. Such statistics are often displayed without explanation. It implies that a knowledgeable fan can gain useful insights from the statistics, such as "Player X lost because he made more unforced errors." Such high-level statistics can be useful but insufficient because they fail to reveal the complexity of tennis matches. As more detailed datasets become available, I was able to conduct micro-level data analysis that reveal deeper insight into the dynamics of the match.

This dissertation presents a new data analysis and visualization technique for analyzing tennis shot patterns. Shots are the basic building blocks of tennis matches. A tennis tactic is a combination of certain shots. In other words, each tennis tactic is a particular shot pattern. By analyzing and visualizing shot patterns, this system can show how different players use different tactics to try to gain advantage. Currently shot patterns are often visualized as heat maps or shot trajectory maps. Heat maps can show where the shots land on court, but they cannot differentiate various combinations of shots. They also do not differentiate long rallies and short rallies. Shot trajectory maps can potentially show shot combinations, but when many shots are displayed, such a map becomes very crowded and is hardly readable. I address these problems by introducing a new visualization technique – fractal tables. A fractal table is a table that can be subdivided recursively as needed. In this method, each fractal table contains tennis points of a certain rally length. Each cell in a fractal table represents an unique combination of shots – a shot pattern. A tennis match can be visualized by placing each point in a cell of a fractal table. Users conduct data analysis by placing different data points to the fractal tables and then look for patterns. Because the fractal tables can be recursively subdivided, they can accommodate tennis points of any rally length and any combination. The fractal tables provide a clear and efficient technique for analyzing tennis shot patterns on a micro-level. For example, long shots and short shots are separated. If a player favors certain shot patterns, they will be clearly visible in the fractal tables because more points will be placed in particular cells. The fractal tables can show which player has advantage (or disadvantage) in certain patterns. The visualization can clearly show why and how a player wins (or loss) from a tactical perspective. Because the fractal table can be easily subdivided and merged, we can conduct analysis at different levels of details. Multiple tactics can be merged into a tactical group. Overall, this new data analysis and visualization technique can help us better understand a player's tactical and strategical thinking during the match. I demonstrate the new technique by presenting a case analysis of a professional tennis match.

1.1.4 Tactical Patterns in Tennis with Tactical Rings

Many professional sports games are played at a hectic pace in a seemingly chaotic environment. As a result, non-expert viewers often miss the tactical patterns used by the players and therefore miss the opportunity to enjoy the games on a deeper level. It is not easy to describe such tactical patterns in words. As more detailed datasets become available, we can conduct data analysis to reveal the dynamics of the match. Sports data visualization and visual analytics can help reveal tactical patterns that are not otherwise obvious to the viewers. Current tennis analytics focuses primarily on high-level statistics (such as serve percentage and number of unforced errors). Although very useful, the high-level statistics often fail to reveal the complexity and the dynamic nature of tennis matches. The visual analysis of a tennis match mostly uses heatmaps or ball trajectory tracing charts. There are many shortcomings to these approaches. For example, a heatmap blends all the shots without showing the sequence of the shots, which is important for understanding tactical patterns. A ball trajectory tracing chart shows the shot sequence and the placement of the balls. However, by plotting many overlapping ball trajectories in one chart, it often creates clutter and is difficult to interpret. We address the above problems by introducing a new visualization technique – Tactical Rings. Tactical Rings are designed to visualize tactical patterns in tennis. Each point in tennis is a sequence of shots, and specific combinations of shots become tactical patterns due to their effectiveness. Most of the points in tennis contain one or more tactical patterns. Expert tennis players practice various tactical patterns and use them in their games to gain advantages. Some players are well known for their favorite patterns. Tennis players, coaches, and analysts often study tactical patterns of a player to analyze his/her game. However, tactical patterns are not always easy to spot, especially to non-expert viewers. Tactical rings are circles that can be recursively subdivided, leading to a set of concentric circles based on shot length. Each tactical ring contains tennis points of a certain rally length. Each cell in a tactical ring represents a unique shot pattern. An entire tennis match can be visualized by placing each point in a cell of tactical rings. Tactical rings have several benefits. It makes efficient use of space. It can accommodate tennis points of any rally length and any combination since they can be recursively divided. The sequence of shots is visualized without creating a clutter of overlapping lines. In this chapter, we demonstrate the application of tactical rings with case studies that focus on the first four shots, which is the primary space for tactical patterns.

1.1.5 Playing Style Analysis in Tennis

Sports data visualization and visual analytics is an effective medium to communicate the happenings, details, and otherwise obscure patterns in a match. Thus, data visualization and analytics are being extensively used not just in the analysis of sports but also in the form of news dissemination to augment the understanding of fans. For example, in 2020, the Australian Open Tennis Championship partnered with Infosys to provide viewers with a set of data visualizations (e.g., MatchBeats, Stats+, CourtVision, and Rally Analysis) for realtime match analytics.

Traditional tennis analytics generally focuses on high-level statistics, such as serve percentages, number of unforced errors, etc. As more detailed, shot-by-shot data sets become available, more micro-level data analysis techniques have been developed to reveal deeper insight into the dynamics of tennis matches. Commonly used low-level data visualizations include heatmap, ball trajectory chart, and ball contact locations. However, these visualization techniques do not sufficiently capture a tennis player's distinctive style of play.

Tennis experts and fans like to discuss tennis players' different styles. It is part of the joy of tennis. However, such analyses are often abstract and oversimplified. Players are labelled as "aggressive", "all-court player", "defensive", or "grinder". Such labels are often too simplistic and do not accurately capture the details and the dynamic nature of a player's style.

In this chapter, we present a new method to visualize a player's distinctive pattern of play. The goal is to visualize the characteristics of a tennis player's game, such as serve patterns and return patterns. Here we use the term "fingerprint" to refer to the characteristic of a player's game. Therefore, the proposed visualizations are used to show the "fingerprint" of a player's pattern of play.

The analysis presented in this chapter is based on micro-level performance data from professional tennis matches. Although we focus on tennis in this chapter, this idea can be extended to the analysis of other sports (including Esports), if micro-level performance data is available.

PART 2

BACKGROUND AND RELATED WORK

There has been a good number of research literature on sports data visualization and analytics in general and tennis visual analytics in particular. He and Zhu [1] proposed a data visualization that shows the progression of a tennis match. Users can highlight various performance data on the visualization, such as unforced errors, shot types, etc. [1] proposed a data visualization that shows the progression of scores of a tennis match. Users can highlight various performance data such as unforced errors on the chart. Some works like [2], [3], [4], [5], [6] have done some important work on statistical modeling of game analytics of tennis matches. However, none of the above works have included anxiety, confidence and momentum in their analysis. [7] analyzed the win-lose rate of different players at critical moments (e.g. break-points, set-points, etc.) and indentified mental characteristics of these players. [7]. Compared to the work on this dissertation, their analysis is largely based on a higher level, without low-level performance data such as unforced error, winners, and serve faults. They did not provide a data visualization interface for users to explore the data. [7] have a linear mixed probability model which they use to describe the score and the outcome of the score in terms of mentality with statistical models though it requires some effort on the part of the reader to figure as how the model variables correlate to the score and subsequently the mentality affected by the score of interest. It is worth reemphasizing that information visualization techniques especially in sports data visualization are chosen based out of data types. Like [11], [12], [13] the display of performance data related to on court performance are displayed as heatmaps and or markers. Further, like [14] game trajectory lines are displayed superimposed on a court image. As for display of both on-court and off-court statistics information techniques show wide variance as evidenced from [15] , [16] and [17] among others. From the perspective of analysis through sports psychology and tennis psychology in particular works by [18], [19], [20] are pertinent. [18] perform experiments to analyze the performance of expert and novice tennis players by making the experiment subjects identify the type of serve presented (flat, top-spin, sliced). [18] investigated visual search patterns.Those experiments showed that expert players were more focused on the shoulder, trunk areas whereas novice players tended to concentrate their search around the head of the server. The experiments were performed using the techniques of temporal visual occlusion. The conclusions emphasize the importance of combining sampling of eye movement and behavior parameters to enhance the understanding of the perceptual processes. [19] study to assess the degree to which different strategies are used by tennis coaches for influencing self-efficacy and their evaluation of the effectiveness of those strategies. The study results indicated that the most often used as well as effective strategies to enhance self-efficacy included encouraging positive self-talk, modeling confidence oneself, instruction-drilling, liberally rewarding statements, and verbal persuasion. [19] study also took us to looking at [20]. [20] studied tennis coaches in a way by surveying junior tennis coaches to ascertain their opinions on mental skills training including the specificity about the specific mental skills they teach and their effectiveness. [20] conclude that the mental skills most difficult to teach were related to handling of pressure, crisis and about dealing with self-confidence, and emotional control. Important also is research by [21] who states that people are contributors to their circumstances and not merely produced by circumstances. [21] states that among the mechanisms of agency the most important and significant is the belief of personal efficacy, which is stated to be the foundation of human motivation, well-being, and accomplishments.

Our method described in Parts 3 and 4 of this dissertation differs from the methods discussed earlier. We propose a method to analyze a tennis player's confidence level, momentum and serve confidence during a tennis match based on the psychological theories of confidence [36] and a database of over 4,000 professional tennis matches [8]. In our model, an athlete confidence level is mainly based on an in-match performance with confidence starting out from an average baseline and rising or falling or maintaining the level accomplished. We also enable the study of the correlation between a tennis player's confidence and momentum level and his/her performance measures, such as unforced errors, forced errors, winners, serve directions, first-serve faults, double faults, aces and rally lengths.

[24] have introduced techniques to visually encode the dynamics of a tennis match by using a hierarchical concept similar to layered icicle representations used for visualizing information hierarchies. [24] represent the time axis vertically as multiple aligned scales to indicate the duration of games and points and to support comparison tasks. Moreover, [24] have used color coding to indicate additional attributes attached to the data. [33] present a method which recommends the most likely serves of a player in a given context by utilizing a style prior. [33] contend that we can improve the prediction and or recommendation. Their approach also allows to quantify the similarity between players, which is useful for predictive analytics. [25] use probabilistic graphical models to model player behavior which can be used to find the factors such as location and speed of the incoming shot which are most conducive to a player hitting a winner or cause an error, and do what [25] call "live in-point" prediction - based on the shots being played during a rally they estimate the probability of the outcome (e.g., winner, continuation, or error) and the location of the next shot. [26] provide a visualization of an entire table-tennis match from time-oriented, statistical, and tactical analyses based perspective. The [26] proposed system has several well-coordinated views supports correlation identification through statistics and pattern detection of tactics with a score timeline and also allows cross analysis to gain insights. [27] worked with domain experts to present a visual analytics system for soccer analytics allowing users to track the spatio-temporal changes in formation and understand how and why such changes occur. The usefulness of the technique is due to the spatio-temporal nature of formations and other characteristics of soccer data, such as multivariate features, which makes analysis of formations in soccer a challenging problem. [28] present a system for analyzing high-frequency position-based soccer data at various levels of detail for analysis of movement features and game events.

Our method in Part 5 and Part 6 of the dissertation differs from the methods discussed above. With fractal tables and tactical rings, we propose a method to visually analyze a tennis match's inclusive four-shot space starting from the serve to include up to 3 return shots by using a database of around 4860 tennis matches [8]. Discussed in Part 5 of this dissertation - the fractal table, an inclusive four-shot space is clearly subdivided along the horizontal axis. Each shot space is further recursively subdivided into relevant categories, as 1 shot space is the serve which has a three cell space corresponding to the types of serve, namely wide , body and down the T. A two shot space would have cell space for each types of serve and the types of return namely forehand and backhand. Similarly the three-shot space would have cell space for all the types of serve , the types of return and a further subdivision for the shot by the server after receiving the return. The four shot space has cell space for all types of serves , the types of return, a further subdivision for the shot by the server after receiving the return and the final shot by the player who received the serve.

Similar to Fractal Tables described above, discussed in Part 6 of this dissertation - Tactical Rings are circles that can be recursively expanded by shot length, leading to the creation of a set of concentric circles. The innermost ring represents one-shot points (i.e., aces), the second ring represents two-shot points (i.e., a service followed by a return), the third ring represents three-shot points, and so on. Each ring is further divided into multiple segments, which are called cells. Each cell contains points with a particular shot combination, such as "wide-service - backhand return - forehand - backhand". A tactical pattern is a short combination of shots, usually 2 to 4 shots. Therefore, in our visualization, the points are visually sorted by tactical patterns.

As can be inferred from the discussion on existing literature in this part of the dissertation - traditional tennis analytics focuses on analyzing high-level statistics, such as serve percentages, serve winning percentages, etc. These statistics mostly deal with the outcome of points, games, and matches, not how the points are played. As already discussed before more recently, Wei, et al. used vision-based tennis ball tracking data to predict tennis players' serves [33] and shot directions [25].

The work presented in Part 7 of this dissertation is different from previous works in that we focus on visualizing a player's style, not the outcome of the points, games, or matches. This work was inspired in part by the previous work on literature analytics through visual fingerprinting [37]. We are interested in applying a similar technique to tennis analytics.

PART 3

ANALYSIS AND VISUALIZATION OF SPORTS PERFORMANCE ANXIETY IN TENNIS MATCHES

This chapter discusses our a method to analyze a tennis player's anxiety level during a tennis match with a method based on the psychological theories of anxiety and a database of over 4,000 professional tennis matches. In our model, an athlete's anxiety level is based on three factors: uncertainty, anticipation, and threat. We have also developed data visualizations to help users study the potential correlation between a tennis player's anxiety level and his/her skilled performance, such as unforced errors, forced errors, winners, serve directions, first-serve faults, and double faults.

3.1 Data

This analysis is based on the tennis match data from Tennis Abstract [8], an open source project that, as of the summer of 2015, provides point-by-point, shot-by-shot statistics of over 4000 professional tennis matches. The data includes the type of shot, direction of shot, types of serves, direction of serves, depth of returns, types of errors, etc. This data set is more detailed (and more useful) than the data retrieved from any tennis video analysis, which can identify the players' movements but cannot yet identify the type of shots and type of errors.

3.2 Basic Performance Anxiety Model

Anxiety is a type of fear reaction. Based on psychological and neuroscience studies [10], we have identified three factors that influence the level of anxiety: uncertainty, anticipation, and threat. These three factors not only apply to sports but also to other fields.

3.2.1 Uncertainty

We define uncertainty as a function of the gap in scores. The bigger the gap in scores, the lower the uncertainty. Uncertainty is at the highest when the score is tied. In this model, the uncertainty level (integer) ranges from 1 to 4, with 4 being the highest. (We do not include a 0 level uncertainty because in competitive sports there is always some uncertainty until the game is over.) For example, if a tennis game progresses as 0-0, 15-0, 15-15, 30-15, 40-15, the corresponding uncertainty levels would be 4, 3, 4, 3, 2 respectively.

3.2.2 Anticipation

We define anticipation as how close a player is to win a game. The closer to the end of the game, the higher the anticipation level. Specifically, the anticipation level is a function of the number of points a player needs to win in order to win the game. In this model, the anticipation index (integer) ranges from 0 to 3, with 3 being the highest. If the game goes to deuce, the anticipation index can increase to 4. For example, if a game runs 0-0, 15-0, 30-0, 30-15, and 40-15, the corresponding anticipation levels for the winning player would be 0, 1, 2, 2, and 3. At the start of the game, the anticipation is 0. At 30-0, the leading player is 2 points away from winning the game, the corresponding anticipation level is 2. At 40-15, the leading player is one point away from winning, therefore the anticipation level is 3.

3.2.3 Threat

We define threat as how close a player is to lose a game. If the opponent is closer to winning a game, a player perceives a higher level of threat and, as a result, a higher level of anxiety. In this model, threat index (integer) ranges from 0 to 3, with 3 being the highest. If the game goes to deuce, the threat index can increase to 4. The threat level is a function of how many points the opponent needs to win a game. For example, if a game goes 0-0, 15-0, 30-0, 30-15, 40-15, the corresponding threat levels for the winning player would be 0, 0, 0, 1, 1. One player's anticipation is the opposing player's threat.

3.3 Performance Anxiety Model for Tennis

In this model, the Anxiety score is a combination of Uncertainty, Anticipation, and Threat.

$$Anxiety = Uncertainty + Anticipation + Threat \tag{3.1}$$

Additionally a few minor factors influence a player's anxiety level in a complicated game like tennis.Statistics indicates that professional tennis players have a high probability to win their service games. For example, ATP players usually win 70% to 90% of their service games, while WTA players usually win 60% to 80% of their service games. This means a slightly higher expectation (and higher anxiety) for the server to win the current game because the server knows the opponent will likely win the next game (the opponent's service game). If the player has a weak serve, then the anxiety level should be even higher. For the same reason, a player is slightly less anxious when returning serves. If the player has a higher return game won percentage, the anxiety level should be even lower.

Therefore, in this model the anxiety score from Equation (1) is multiplied by a serve anxiety index. The serve anxiety index for the server is larger than 1 and is calculated based on the winning percentage of a player's service games. For example, if a player has a 75% service game won percentage, then the serve anxiety index is 100/75 = 1.33. (If a player's statistics is unknown, the average percentage is used.) On the other hand, When a player is returning serve, the serve anxiety index is smaller than 1 and is calculated based on the the winning percentage of the player's return game. For example, if a player has a 25% return game won rate, then the serve anxiety index is (1-0.25)=0.75.

3.4 Case Studies

In this section we demonstrate this method and visualization output (Figures 2.1 to 2.6) using the Tennis Abstract match data for the 2017 US Open semi-final between ATP players Kevin Anderson and Pablo Carreno Busta (henceforth referred to as KA and PCB respectively), where KA defeated PCB 4-6, 7-5, 6-3, 6-4. Each row in the visualization represents a set. The top row represents the first set, the second row the second set, and so on. Each individual chart represents a game. Based on the ATP tour stats [31], KA and PCB's service game won percentages are 86% and 76% respectively. Their return games won percentages are 17% and 25% respectively.

The programs are implemented with Python. The data visualizations are implemented with Python library Bokeh [9]. The game-by-game anxiety indexes are plotted as individual line charts, with performance data such as first serve fault, double fault, serve directions plotted as colored markers on the anxiety score lines. Compared with pure statistical analysis, these visualizations can help users correlate anxiety with various performance measures and conduct in-depth analysis on point-by-point, game-by-game basis.

3.4.1 Uncertainty

Figure 2.1 shows the correlation between uncertainty and unforced errors, forced errors, and winners for player PCB. In tennis, every point ends with one of the three outcomes: unforced error, forced error, or winner. An unforced error is an error made when a player is balanced and has enough time to make the shot, indicating poor performance. For top professional players, most of the unforced errors are attributed to nervousness, which is directly related to anxiety. A forced error is an error made by a player who is either out of balance (e.g. stretching to reach the ball) or has very little time to react. A forced error may not be directly related to anxiety but it may indicate that the player's previous shot is weak. A winner is a winning shot that the opponent is unable to reach, indicating excellent level of performance. This analysis and visualization shows a strong correlation between uncertainty and unforced errors and winners. Out of 83 unforced errors, 59% of them were made when uncertainty index is 3, and 30% of them were made when the uncertainty index is 4. Nine unforced errors were made when the uncertainty index is 2, and no unforced error was made when the uncertainty index is 1 (lowest). Out of 49 winners, 54% of them were made when the uncertainty index is 3, and 29% of them were made when the uncertainty index is 4. This shows both the negative and positive impact of uncertainty on the performance: when the

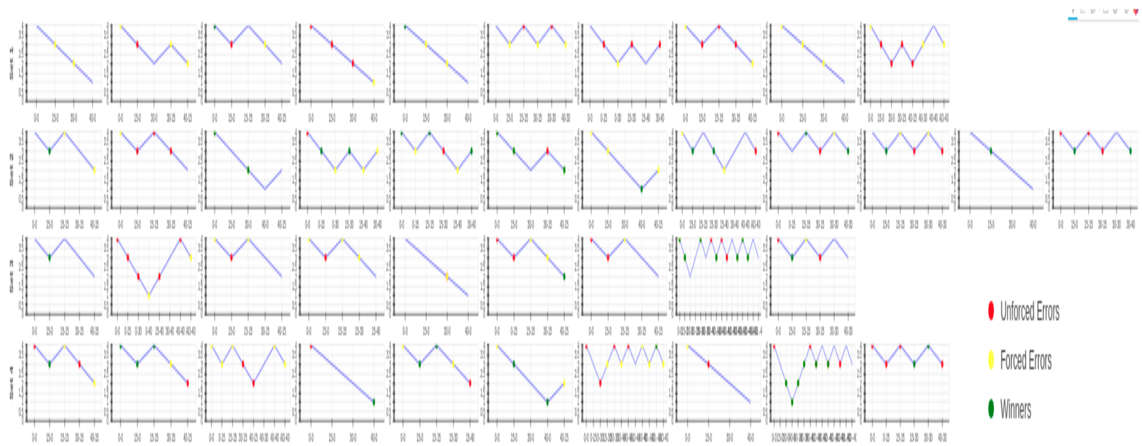

Figure (3.1) Correlating Uncertainty with Unforced Errors, Forced Errors and Winners for PCB.

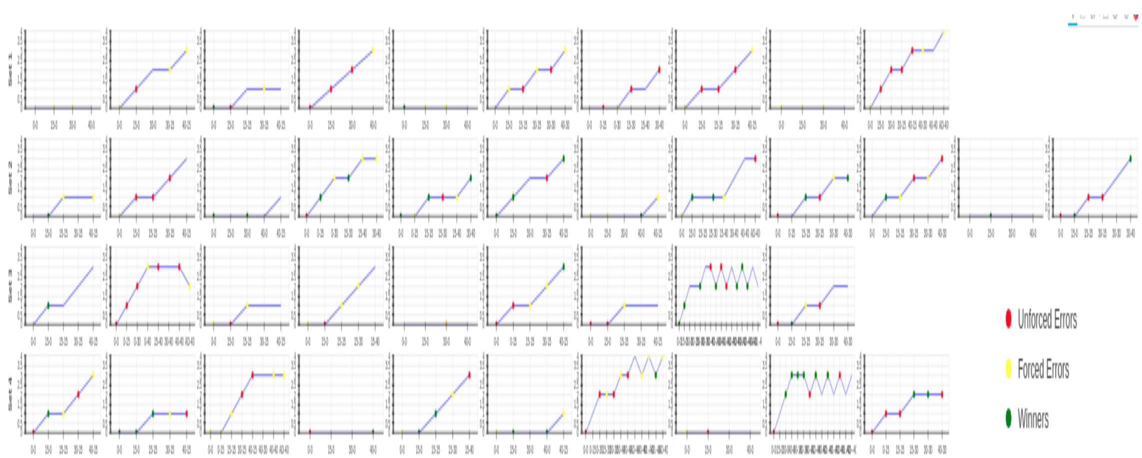

Figure (3.2) Correlating Anticipation with Unforced Errors, Forced Errors and Winners for PCB.

scores are close, the player made more unforced errors but also made more winning shots.

3.4.2 Anticipation

Figure 2.2 shows the correlation between anticipation and the performance for PCB. Overall, this analysis suggests a weak correlation between anticipation and performance. The unforced errors are more or less evenly distributed at different anticipation levels. However, there are more winners when the anticipation level is low. For all the winners, about 35% of them were made when the anticipation index is 0, compared to 24% for the anticipation index 1, 20% for the anticipation index 2, and 20% for the anticipation index 3. This suggests that PCB was more aggressive when the anticipation level is low (e.g. early in the game).

3.4.3 Threat

Figure 2.3 shows the correlation between threat and PCB's performance. There is a strong correlation between low threat level and unforced errors. Among the unforced errors, 36% of them were made at the threat index of 0, 35% made at the threat index of 1, 16% were made at the threat index of 2, and only 12% were made at the threat index of 3. This suggests that PCB was a little careless when the threat level is low but made a conscious effort to avoid unforced errors when the threat level is high (close to losing). There is also a correlation between low threat level and winners. Among the winners, 31% of them were at the threat index 0 and 1 respectively, 15% were made at the threat index 2, and 20% were made at the threat index 3. This suggests that PCB was slightly but noticeably more cautious when the threat level is high.

3.4.4 Anxiety Score

Figure 2.4 shows the correlation between the combined anxiety index and performance for PCB. This analysis shows that unforced errors exhibit a normal distribution with regards to the anxiety index. Among unforced errors, 64% of them occur in the anxiety index range 5 to 10 (medium), 25% of them occur in the anxiety index range 0 to 5 (low), and 11%

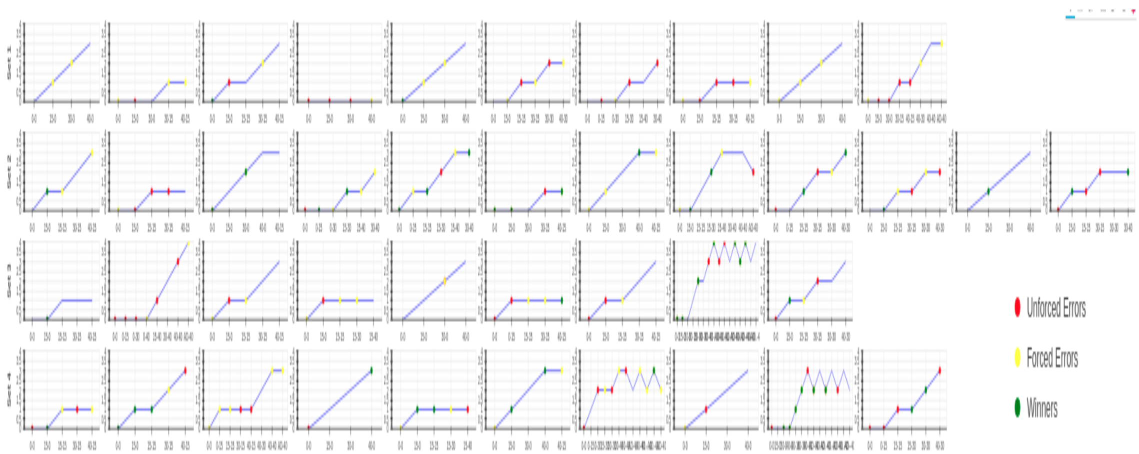

Figure (3.3) Correlating Threat with Unforced Errors, Forced Errors and Winners for PCB.

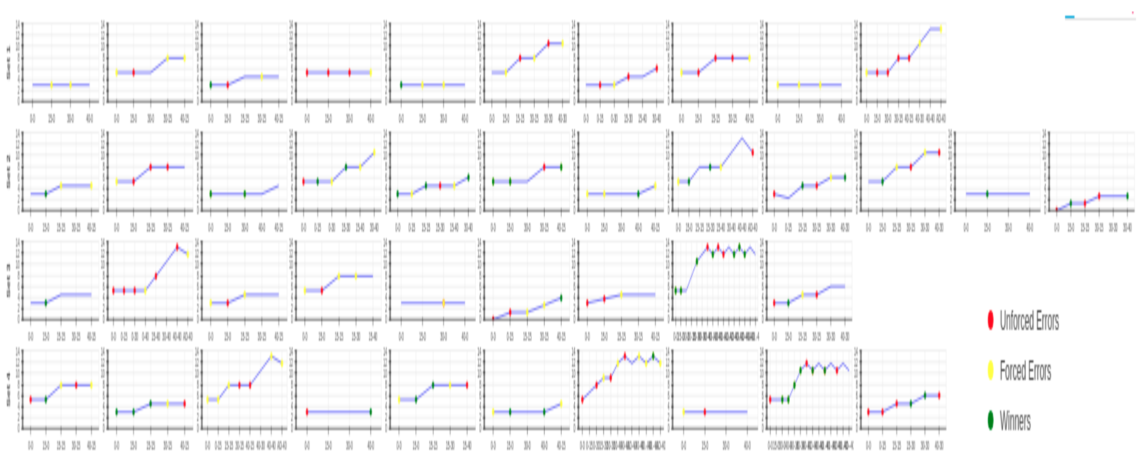

Figure (3.4) Correlating (Combined) Anxiety with Unforced Errors, Forced Errors and Winners for PCB.

occur in the anxiety index range 10 and above (high). This suggests that PCB managed to reduce his unforced errors at high pressure situations, perhaps by playing more cautiously. Most unforced errors occur in medium anxiety situations. This analysis shows that most of the winners were produced in low to medium anxiety situations. Among all winners, 38% of them were produced in the anxiety index range 0 to 5 (low), 44% of them produced in the anxiety index range 5 to 10 (medium), and only 15% were produced in the anxiety index range 10 or higher. This suggests that PCB was more aggressive in low anxiety situations but took less risk in high anxiety situations. However, our data visualizations show there are exceptions to this general pattern, thus allowing more nuanced analysis. In the eighth game of the third set and the ninth game of the fourth set, PCB produced 13 winners, with 8 of them in high anxiety situations (anxiety index of 10 or higher). This indicates that PCB decided to be more aggressive and took more risk when he was close to losing a set.

3.4.5 Serve Direction

Figure 2.5 shows the correlation between anxiety and serve directions. There are three directions in a tennis serve: Wide, Body, and Down-the-T. In high anxiety situations, a player tends to use his/her most effective serve. Analysis here shows that PCB's favorite serve direction in high anxiety situations (anxiety index of 10 or higher) is wide - 44% of his serves in high anxiety situations are wide. Data visualizations also show an interesting pattern. PCB did not use any Down-the-T serve in high anxiety situations until the 8th game in the 3rd set and the 7th and 9th games in the last set, in which he suddenly served seven times Down-the-T at critical moments. Similarly, he only served 7 times to the Body in the first 29 games in high anxiety situations, but served 9 times to the Body in the 8th game of the 3rd set and the 9th game of the 4th set in high anxiety situations. This may be a calculated decision to change his serve patterns to surprise his opponent at critical moments.

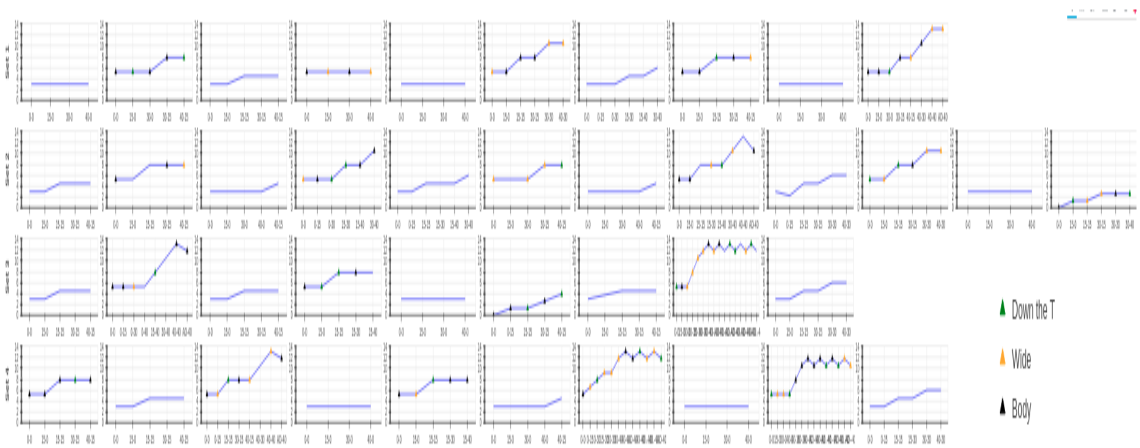

Figure (3.5) Correlating (Combined) Anxiety with Serve Direction for PCB.

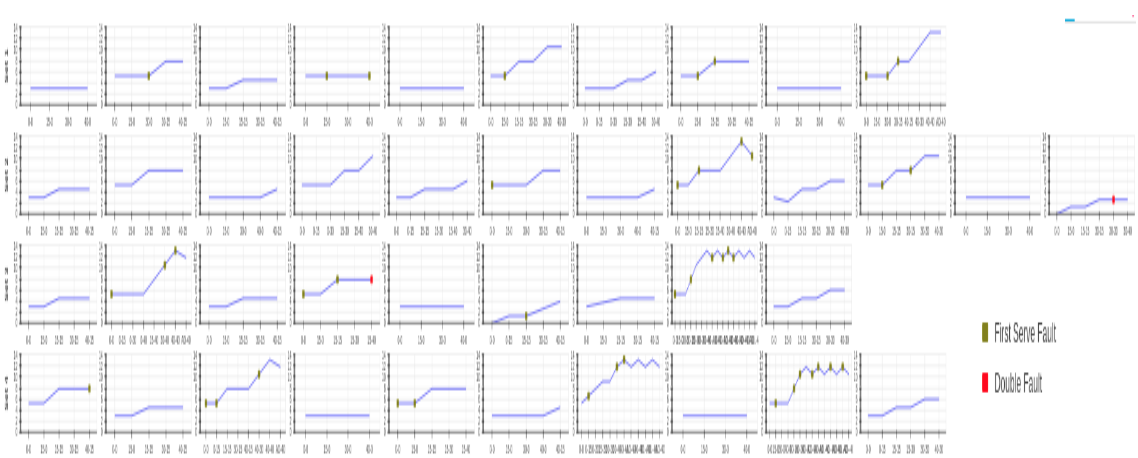

Figure (3.6) Correlating (Combined) Anxiety with First Serve Fault and Double Fault for PCB.

3.4.6 First Serve Fault and Double Fault

Figure 2.6 shows the correlation between first serve faults and double faults with anxiety index. This analysis shows that PCB made only 4 first serve faults in high anxiety situations (anxiety index of 10 or higher) in the first 14 service games. Then he made 12 first serve faults in the last 6 service games in high anxiety situations. This suggests that the heightened anxiety near the end of the match might cause these errors. Or perhaps PCB decided to serve more aggressively later in the match when he was losing. PCB made two double faults, both near the end of the games, perhaps indicating that high level of threats or anticipation led to the errors. Based on the above analysis, we can give a brief summary of current world number 21 PCB's pattern of play in this match. In high anxiety situations, he was cautious and able to reduce unforced errors and serve faults. In medium to low anxiety situations, he was more aggressive but also made many unforced errors. However, when close to losing the match, he could become very aggressive in high anxiety situations.

3.5 Discussion

This chapter described a method to computationally analyze and visualize the relationship between a player's anxiety and performance in a tennis match. My case study shows that we are able to discover useful patterns through these analysis and visualizations. My method has its limitations. It is an indirect estimate of an athlete's anxiety status, not a direct measurement of the athlete's physiological state (which is usually difficult to do in a professional match). My methods do not include factors outside of the match, such as physical condition, environment, etc. In the future, we plan to expand my data analysis and visualizations to other factors such as confidence, while also planning to expand these methods to other fields where anxiety-performance analysis is useful.

PART 4

ANALYSIS AND VISUALIZATION OF SPORTS CONFIDENCE AND MOMENTUM IN TENNIS

This chapter discusses our proposed method to analyze a tennis player's confidence level and serve confidence during a tennis match using known theories of psychology in confidence and techniques to identify a player's momentum in a match. In our model, an athlete confidence level is mainly based on an in-match performance with confidence starting out from an average baseline and rising or falling or maintaining the level accomplished. We have also developed data visualizations to help users study the correlation between a tennis player's confidence and momentum level and their performance measures, such as unforced errors, forced errors, winners, serve directions, first-serve faults, double faults, aces and rally lengths.

4.1 Data

The tennis match data is made available by Tennis Abstract [8], a project maintained and contributed to by tennis fans who have till the summer of 2015 contributed the Shot-by-Shot statistics of over 4000 matches which contain shot-by-shot data for every point of a match, including the type of shot, direction of shot, depth of returns, types of errors, etc. [8] contains both the raw point-by-point data from these matches and extensive match-level aggregate totals. Match scores are available for both men's and women's game. The metadata for the games and statistics include the player names, tournament, date, surface, etc [8]

4.2 Confidence Score

The confidence score has been modeled as consisting of the point confidence and game confidence. Point Confidence is the confidence gained by the player at the point level and subsequently the game confidence is the confidence gained or lost at the game level. Both these individual confidence scores are combined to form a confidence score which is then modulated by the way the points were won to give us a total Confidence Score. Confidence Score varies from 1 to 5 with 1 being the lowest and 5 the highest. A baseline confidence level is set at 3.

4.2.1 Point Confidence

A player winning 3 points in a row gains confidence with a level up by 1. Player wins 4th point then confidence goes by half-a point. This also would have been winning the game in love. And since confidence win and loss is symmetric as one player's loss in confidence is the opponent's gain. So, when a player loses 3 point in a row , confidence is lost by a level down by 1. Lose 4 points in a row then confidence drops further by half-point.

4.2.2 Game Confidence

Similarly,when a player wins 3 games in a row game then confidence goes up by 1 when starting the 4th game. From that point on every consecutive win in a game means a confidence level up by 1 point when starting the next game. As an example regarding a singles' match – In a set, when Player A has won the first 3 games , when starting the 4th game the confidence level does not start at average confidence level of 3 but at 4. If Player A also happens to win the 4th game then for the 5th game, Player A's starting confidence would be up again by half-point at 4.5. If Player A has won 5 games in a row then for the 6th game the starting confidence would again go up by half-point thus starting at 5.

The confidence score is not limited by the game. The confidence level does not reset after a game is over but carries over to the next game. For example, if a player wins the last two points in a game and then wins the first point in the next game, these three points are considered three points in a row.

4.2.3 Modulation

Point level confidence has to be modulated by how the point was won. If the point is won by a a forced error or a player hitting a winner, then the point confidence for the point winner is up by 1.0. If the point is won by an unforced error, the point confidence is up by 0.8. If a player wins a service game, the game confidence would not change. If a player wins a return game, the confidence should be up by 1.0. However, if a player loses a service game, the game confidence should be down by 1.0. If s/he then wins his next service game, her/his confidence should go up by 0.5. If s/he wins another service game, the game confidence should be up by another 0.5, which means going back to the player's baseline confidence level. Professional Players are expected to win their service games. So winning a service game is not a big confidence booster, but breaking a return game would be a boost in confidence for the next game where that player serves. So if a player breaks a return game then for the next game, the starting confidence level would be up by 1. Which means after breaking the return game the player starts the subsequent service game at confidence level 4. In any set if a player breaks twice then, that should be a significant boost in confidence where the confidence level going into the next game goes up again by 1.

The composite Confidence Index (CI) is correlated with :

- First Serve Faults and Double Faults.
- Aces.
- Rally Length.
- Serve directions.
- Unforced errors, Winners and the forced errors.

4.3 Serve Confidence

Serve Confidence also varies from 1 to 5 with 1 being the lowest and 5 the highest. A baseline serve confidence level is set at 3. If three first-serves go in then Serve Confidence goes up by a level. Though not common in professional matches, if the 4th serve goes in as well then the serve-confidence goes up by half-point. If a player hits an Ace then the serve-confidence goes up by 1.2. A second ace in a row would boost confidence by 1. Hitting a service winner also boosts the serve-confidence by a point. With regards to missing the first-serves , if two first-serves are missed in a row then the serve-confidence goes down by half-point. Missing 3rd consecutive first-serve, serve-confidence goes further down from the by half-point again. If a player double-faults then serve-confidence goes down by a point. Since, double faulting is a serious lapse in game performance a second double-fault makes serve-confidence go further down by a point. Serve Confidence also carries over from one game to another. The serve confidence is not limited by the game. The serve confidence level does not reset after a game is over but carries over to the next game.

The Serve Index (SI) is correlated with :

- First Serve Faults and Double Faults.
- Aces.
- Rally Length.
- Serve directions.
- Unforced errors, Winners and forced errors.

4.4 Momentum

Momentum is a prolonged period in which one player significantly outperforms the opponent. Confidence model looks at a player's performance. To find out the player's performance outperforming an opponent for a non-insignificant amount of time, we look at the score pattern on the game level where the games looked into would be contiguous without the abstraction of a "set" in a tennis match. In our model, the game level momentum is a 5-game sliding window. When a game is finished, the sliding window moves forward by one game. A player has the momentum if s/he has a least a 3 two-game lead in this 5-game window, for example a Win(W) - Loss(L) sequence of games as WWWLW, LWWWW, WLWWW, WWWWL, etc.

4.5 Model Implementation

We have implemented our anxiety model including all the constituent model parameters visualized based on score and score gaps using a Python library Bokeh [9]. The game by game score data was plotted on the x-axis and on the y-axis respective plots of confidence score of an athlete, serve confidence of both athletes and momentum identification of both players in a single's tennis match.

4.6 Case Studies : Confidence Index

In this section we discuss our analysis and visualization of the confidence index (CI), serve confidence(SC) and momentum of a tennis athlete. Each row in the chart represents a set. Each set consists of multiple games represented by respective plots. Top row represents the first set, rows added subsequently representing the match progression. For the case studies we have chosen a 2017 men's singles US Open semi-final match between Kevin Anderson and Pablo Carreno Busta (henceforth referred to as KA and PCB respectively) where KA defeated PCB by 4-6, 7-5, 6-3, 6-4. From the ATP tour stats [31] we see that service games win percentage for KA and PCB stands at around 86% and 76% respectively. Their respective return games win percentage stands at around 17% and 25% [31]. The charts can be read horizontally - when a game starts, is going on or nears the end. They can also be read vertically - study of pattern when Confidence Index and or Serve Confidence is up or down.

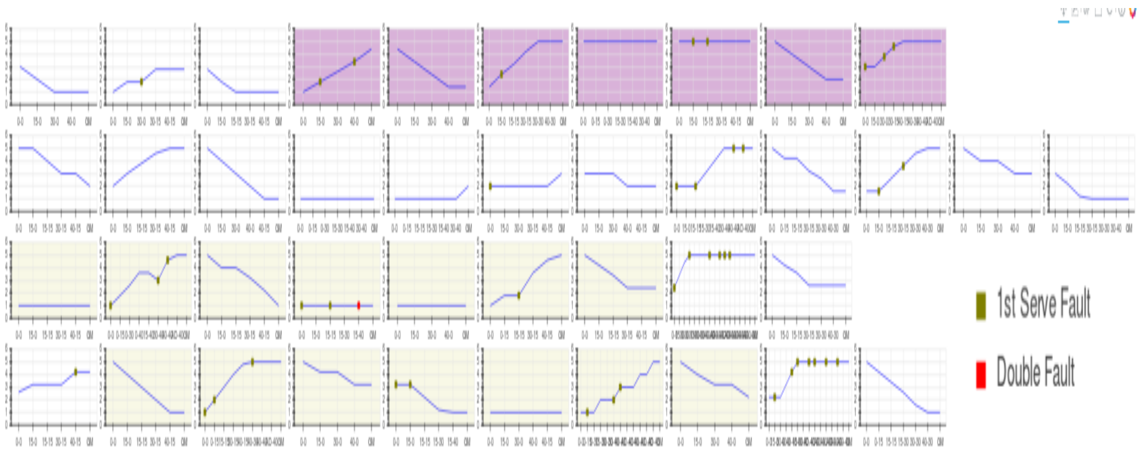

Figure (4.1) Busta Confidence Score correlated with his 1st serve fault and double fault

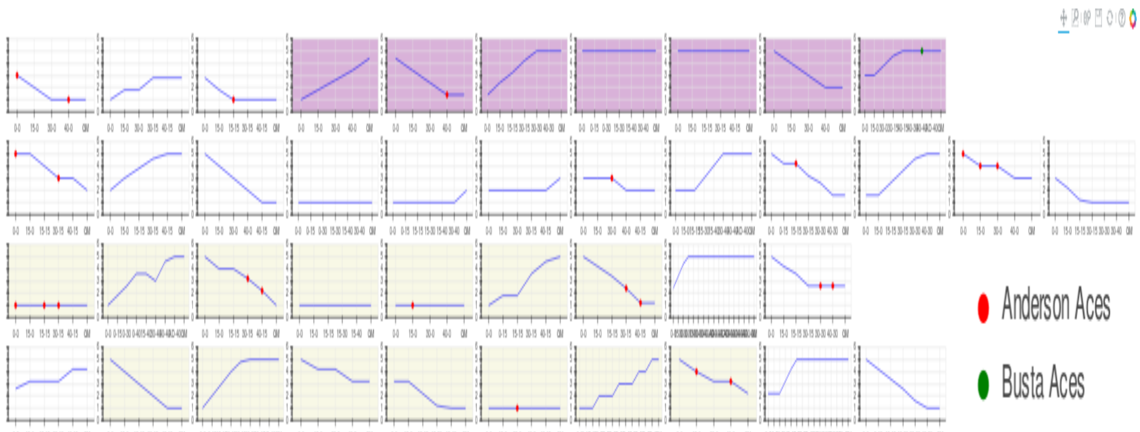

Figure (4.2) Busta Confidence Score correlated with his Aces hit in the match

4.6.1 Confidence Index vs First Serve Faults and Double Faults

We see from Figure 3.1 the correlation between PCB's Confidence Index and his first serve faults and double-faults. First serve faults generally are seen to occur at the beginning of the games and more specifically within the first half of a game, then the players tend to get cautious about not losing the point. It is observed that as a general pattern, we see more first serve faults when the PCB's CI is on the rise and or already high. In competitive games like Set 3, Game 8 and Set 4, Game 7 and Set 4, Game 9 have a higher proportion of first-serve faults per game. The reasons attributed to this is that in moments when confidence is rising and is high, PCB is aggresive and wants to impose his game on the opponent, so as to seek domination. PCB tends to have first-serve faults at the beginning of the game, whereas KA tends to fault towards the end of a game especially if those are the games which he had dominated. Also, as is expected from ATP professionals, we hardly see double-faults throughout the match committed by PCB. Infact double-fault has just one example Set 3 , Game 4. The reason is simple. Players don't want to lose the point on a double-fault. Once they commit a first-serve fault, extra care is taken to put the second serve in. Momentum is denoted with the color background where the purple color denotes PCB's momentum Set 1, Game 4 to Set 1, Game 10 and beige color denotes KA's momentum, Set 3 Game 1 to Set 3 Game 7 and Set 4 Game 2 to Set 4 Game 7. We see that the rate of first-serve fault for PCB continues largely independent of his or his opponent's momentum.

4.6.2 Confidence Index vs Aces

Figure 3.2 the correlation between PCB's Confidence Index and the Aces hit by both KA and PCB. The red circles on Figure 3.2 indicate aces hit by KA and green circles indicate aces hit by PCB. There is a clear pattern here. KA is a player who hits a lot of aces as compared to PCB. KA being an attack minded player as can be seen from Figure 3.2 has hit aces whenever PCB's confidence index(CI) is either decreasing or down already. There are some interesting patterns also from the perspective of momentum. Momentum is denoted with the color background where the purple color denotes PCB's momentum Set 1, Game 4

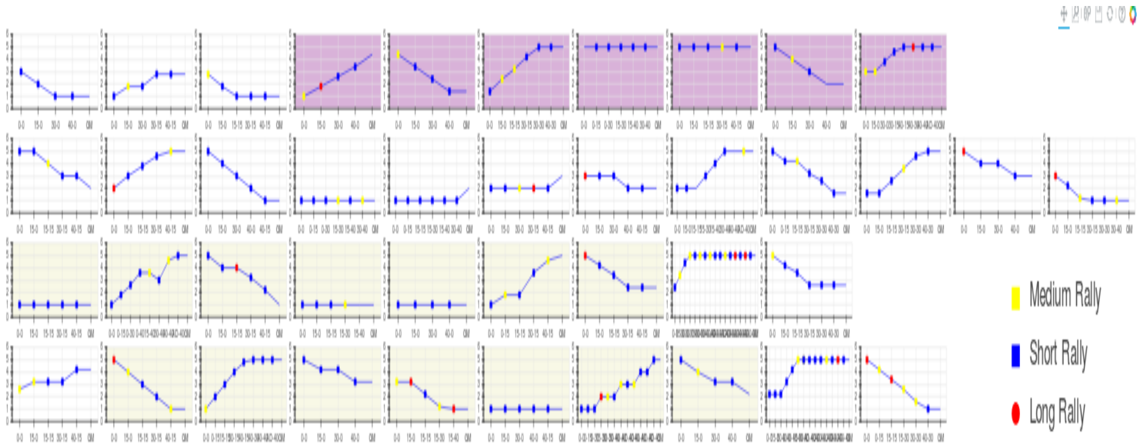

Figure (4.3) Busta Confidence Score correlated with rally length for all points

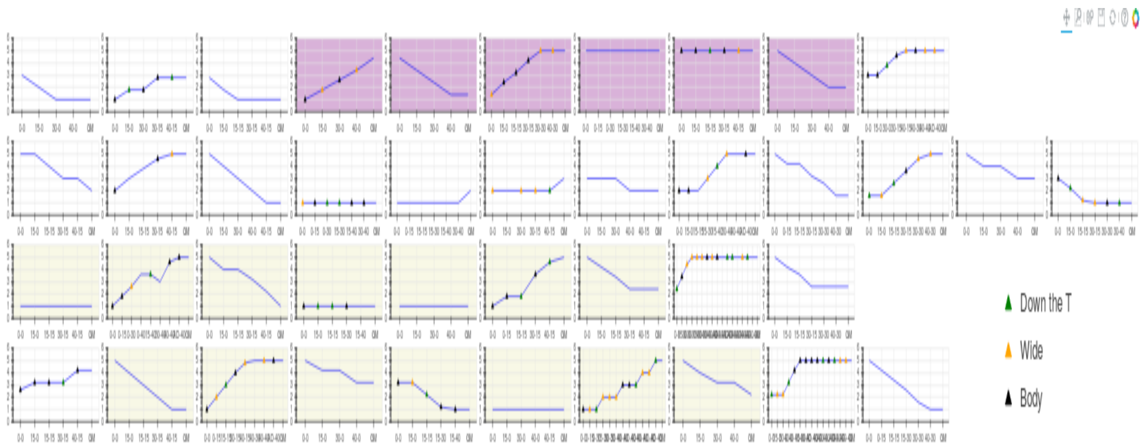

Figure (4.4) Busta Confidence Score correlated with his Serve Directions

to Set 1, Game 10 and beige color denotes KA's momentum, Set 3 Game 1 to Set 3 Game 7 and Set 4 Game 2 to Set 4 Game 7. The prolonged period on the first set when PCB has momentum, he was hit with just a single ace, as compared to other games throughout the match where KA has regularly hit aces. Infact, the total count of KA's aces is 24. Compared to his opponent, hitting a greater number of aces has put KA in a very comfortable position in the match.

4.6.3 Confidence Index vs Rally Length

Rally length of 0 to 4 is classified as short-rally , a length of 5 to 8 as medium and a really length equaling or exceeding 9 as long. In Figure 3.3 the pattern observed from each of the point level, game level and set level scenario is that , although there are instances of medium and long rallies this match is an overwhelmingly short-rally match. And this trend holds largely independent of the confidence level of the player PCB. The medium length rallies and the long rallies even do not have a conclusive pattern with regards to Serve Confidence in that both are seen at varied levels of confidence whether high, medium or low. Set 4 looks to have more medium rallies in comparison where as longer rallies are more evenly distributed. For the comparison we see from the figure we find that Set 1 has 10 medium rallies, Set 2 has 13 although it must be noted that Set 2 has been a long set owing to more number of games played. Set 3 has 11 medium rallies and finally Set 4 has 10 medium rallies. This points to the consistency in medium length rallies. As for the long rallies - Set 1 has just 2 long rallies in the entire set, Set 2 has 5 long rallies although again it must be noted that Set 2 has been a long set owing to more number of games played. Set 3 has 4 long rallies and Set 4 has 7 long rallies. The indications we get from this is that fatigue could be factor. We know that players are generally averse to playing rallies hit shorter rallies as they want to win the point earlier, which requires more effort. We have to recall that this match is a Semi-Final of the US Open, which by itself is the last grand slam of the season. So owing to pretty late time of the season and late stage of the tournament, fatigue could be a factor in sizable medium to longer rally length. Momentum is denoted with the color background where the purple color denotes PCB's momentum Set 1, Game 4 to Set 1, Game 10 and beige color denotes KA's momentum, Set 3 Game 1 to Set 3 Game 7 and Set 4 Game 2 to Set 4 Game 7. From the lens of momentum, we do not see a clear evidence of medium or long rallies when either player has had momentum. The rally length situation is in contrast to the situation with the number of aces discussed in the earlier sub-section where one particular player had a disproportionate share under his name.

4.6.4 Confidence Index vs Serve Directions

Figure 3.4 shows us the pattern between PCB's confidence score and serve directions and we can not really see much of a correlation between the confidence levels and directions of the serve namely wide , body or down the T. Serves have a mixed record of being served in directions that can be considered - namely - Body, Wide and Down The T. This patterns is something which is entirely reasonable as in any competitive match, one way to keep your opponent from reading your serve is to not to be consistent with any particular direction of serve. An important question to attempt an answer is that - Is there a serve pattern for any player during their times of high, medium or low confidence? Momentum is denoted with the color background where the purple color denotes PCB's momentum Set 1, Game 4 to Set 1, Game 10 and beige color denotes KA's momentum, Set 3 Game 1 to Set 3 Game 7 and Set 4 Game 2 to Set 4 Game 7.From the angle of momentum we see that in case of PCB momentum his serves hardly go down the T and most of his serves are wide and a few towards the body. In the cases where KA has the momentum we do see an increase in the serves that are down the T. There also are comparable numbers of PCB serves that go wide and towards the body when KA has the momentum.

4.6.5 Confidence Index vs Unforced Errors, Forced Errors and Winners

Figure 3.5 shows us the pattern between PCB's confidence score and his unforced errors, forced errors and winners. It can be figured out that the number of unforced errors is seen especially in Set 2 , Set 3 and Set 4 while remarkably low in Set 1. We know that PCB carried Set 1 and had momentum from Game 4 onwards till he won the set, while he lost Set 2, Set 3 and Set 4 on his way to losing the match. Also the pattern of unforced errors is that they are quite spread out in the games - beginning, middle or end of the games. Further, unforced errors have mostly associated with the decline in PCB's confidence trend line. We can also see unforced errors when PCB's confidence is really low and flat like games which he lost on love. Momentum is denoted with the color background where the purple color denotes PCB's momentum Set 1, Game 4 to Set 1, Game 10 and beige color denotes KA's

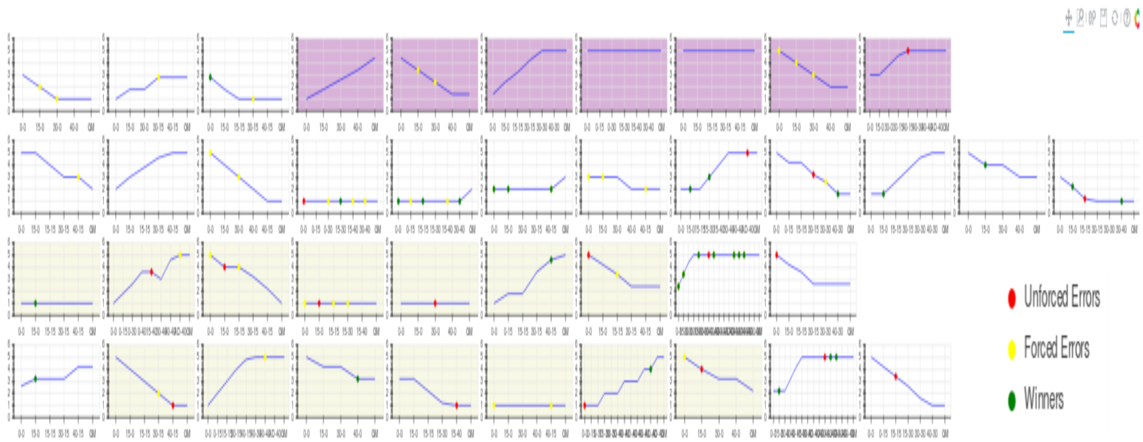

Figure (4.5) Busta Confidence Score correlated with his Unforced Errors , Forced Errors and Winners

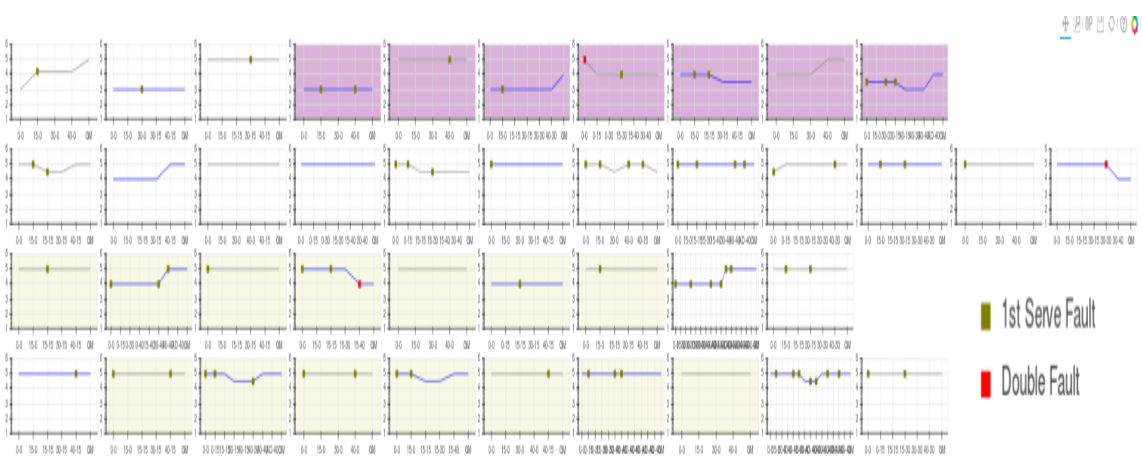

Figure (4.6) Serve Confidence correlated with First Serve Fault and Double Fault momentum, Set 3 Game 1 to Set 3 Game 7 and Set 4 Game 2 to Set 4 Game 7. During PCB's momentum he has committed just one unforced error, while there have been quite a few forced-errors. But, during the games where KA has had the momentum, PCB's rate of unforced-errors increases along with forced-errors which also seem to appear with regularity.

4.7 Case Studies : Serve Confidence

Similar to the earlier section, in this section we discuss our analysis and visualization of the Serve Confidence(SC) and momentum of a tennis athlete. As earlier, each row in the chart represents a set. Each set consists of multiple games represented by respective plots. Top row represents the first set, rows added subsequently representing the match progression. We have chosen the same match for the case studies - 2017 men's singles US Open semi-final match between Kevin Anderson(KA) and Pablo Carreno Busta(PCB) where KA defeated PCB by 4-6, 7-5, 6-3, 6-4.

A general picture we see from Figures 3.6 to 3.10 is that PCB's Serve Confidence(SC) is usually low in the first set and improves in the later sets, which tends to suggest that a serve is not his strong point coming into a match. This explains some patterns in his SC data in that, as a strategy he does not rely heavily on his service-games for results, as in later sets especially Set 2 , 3 and 4 - although his Serve Confidence is higher he has still lost games and conceded the game momentum to KA.

4.7.1 Serve Confidence vs First Serve Faults and Double Faults

In Figure 3.6 we map each games service confidence of the player's with PCB's first-serve faults and double faults. First serve faults generally are seen to occur at the beginning of the games and more specifically within the first half of a game, then the players tend to get cautious about not losing the point. As a trend we see that in the first set, PCB's serve confidence is generally lower compared to KA. What we also see is that PCB carried Set 1 despite his SC not being that higher with respect to his opponent. PCB has double-faulted rarely and whenever he has that has led to his loss of SC. In general, we see more first serve

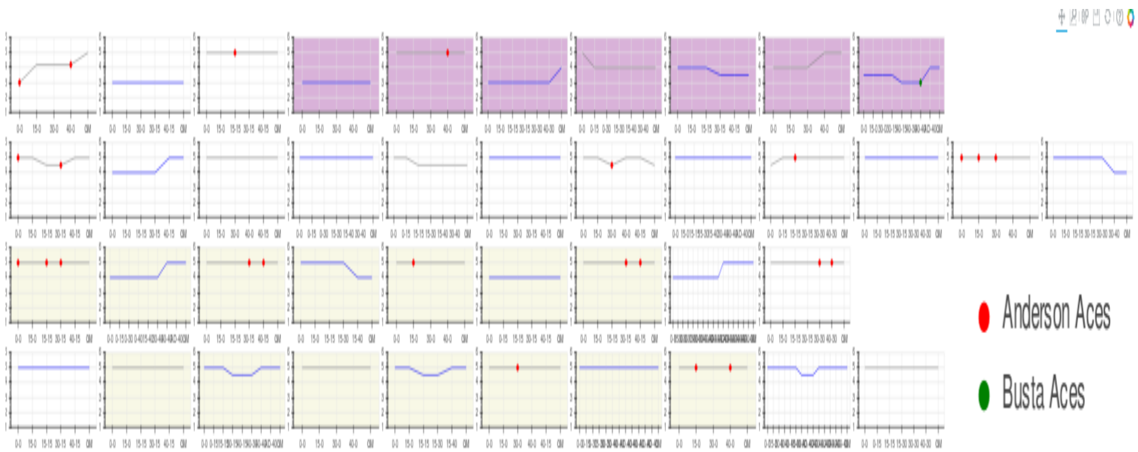

Figure (4.7) Serve Confidence correlated with Aces

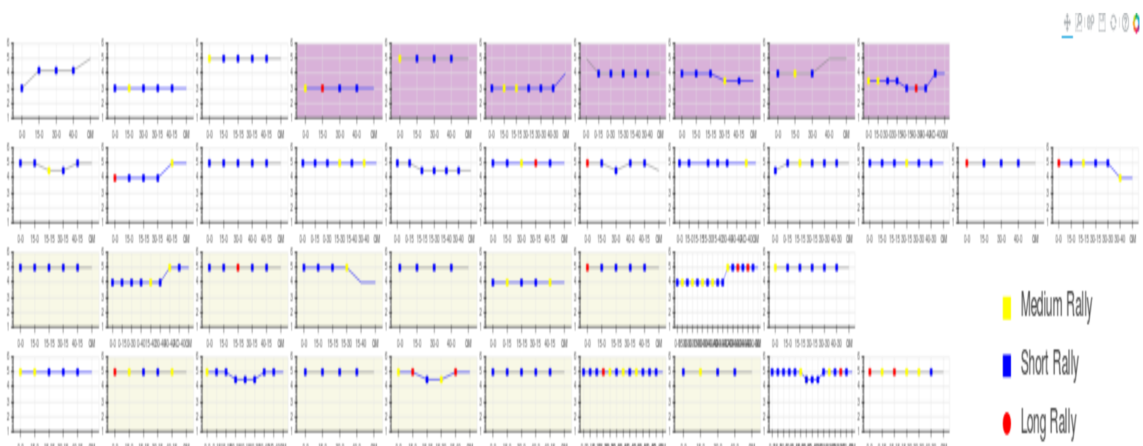

Figure (4.8) Serve Confidence correlated with rally length for all points faults when the PCB's SC is average and or already high. In competitive games like Set 3, Game 8 and Set 4, Game 7 and Set 4, Game 9 have a higher proportion of first-serve faults per game. The reasons attributed to this is that largely irrespective of the state of his SC, PCB is aggressive wants to impose his game on the opponent, so as to dominate the game. Also, PCB tends to have first-serve faults at the beginning of the game, whereas KA tends to fault towards the end of a game especially if those are the games which he had dominated. Momentum is denoted with the color background where the purple color denotes PCB's momentum Set 1, Game 4 to Set 1, Game 10 and beige color denotes KA's momentum, Set 3 Game 1 to Set 3 Game 7 and Set 4 Game 2 to Set 4 Game 7. During games of PCB's momentum his rate of first-serve faults are high which do not substantially decrease during games of KA's momentum. This suggests that the desire to be aggressive to impose his game on his opponent is largely independent of PCB's game level momentum or the lack of it.

4.7.2 Serve Confidence vs Aces

Figure 3.7 tracks the correlation between the SC and the Aces hit by both KA and PCB. The red circles on Figure 3.7 indicate aces hit by KA and green circles indicate aces hit by PCB. There is a clear pattern here. KA is a player who hits a lot of aces largely independent of his SC level, as compared to PCB. KA being an attack minded player as can be seen from Figure 3.7 has hit aces mostly at the beginning and or in the middle of the games, which points to his strategy of aggressive play. There are some interesting patterns also from the perspective of momentum. Momentum is denoted with the color background where the purple color denotes PCB's momentum Set 1, Game 4 to Set 1, Game 10 and beige color denotes KA's momentum, Set 3 Game 1 to Set 3 Game 7 and Set 4 Game 2 to Set 4 Game 7. The prolonged period on the first set when PCB has momentum, he was hit with just a single ace, as compared to other games throughout the match where KA has regularly hit aces. Infact ,the total count of KA's aces is 24. Compared to his opponent, hitting a greater number of aces has put KA in a very comfortable position in the match.

4.7.3 Serve Confidence vs Rally Length

Rally length of 0 to 4 is classified as short-rally , a length of 5 to 8 as medium and a really length equaling or exceeding 9 as long. In Figure 3.8 clear pattern observed from each of the point level, game level and set level scenario is that , although there are instances of medium and long rallies this match is an overwhelmingly short-rally match. And this trend holds largely independent of the SC level of both players. The medium length rallies and the long rallies even do not have a conclusive pattern with regards to Serve Confidence in that both are seen at varied levels of confidence whether high, medium or low. Set 4 looks to have more medium rallies in comparison where as longer rallies are more evenly distributed. For the comparison we see from the figure we find that Set 1 has 10 medium rallies, Set 2 has 13 although it must be noted that Set 2 has been a long set owing to more number of games played. Set 3 has 11 medium rallies and finally Set 4 has 10 medium rallies. This points to the consistency in medium length rallies. As for the long rallies - Set 1 has just 2 long rallies in the entire set, Set 2 has 5 long rallies although again it must be noted that Set 2 has been a long set owing to more number of games played. Set 3 has 4 long rallies and Set 4 has 7 long rallies. The indications we get from this is that fatigue could be factor. We know that players are generally averse to playing rallies hit shorter rallies as they want to win the point earlier, which requires more effort. We have to recall that this match is a Semi-Final of the US Open, which by itself is the last grand slam of the season. So owing to pretty late time of the season and late stage of the tournament, fatigue could be a factor in sizable medium to longer rally length. Momentum is denoted with the color background where the purple color denotes PCB's momentum Set 1, Game 4 to Set 1, Game 10 and beige color denotes KA's momentum, Set 3 Game 1 to Set 3 Game 7 and Set 4 Game 2 to Set 4 Game 7. Also, from the lens of momentum, we do not see a clear evidence of medium or long rallies when either player has had momentum. The rally length situation is in contrast to the situation with the number of aces discussed in the earlier sub-section where KA had a disproportionate share under his name.

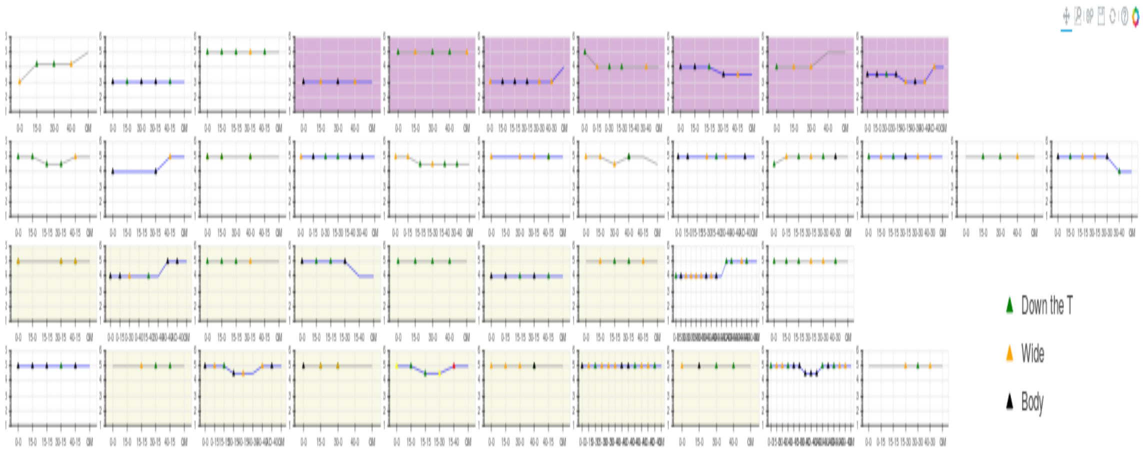

Figure (4.9) Serve Confidence correlated with Serve Direction

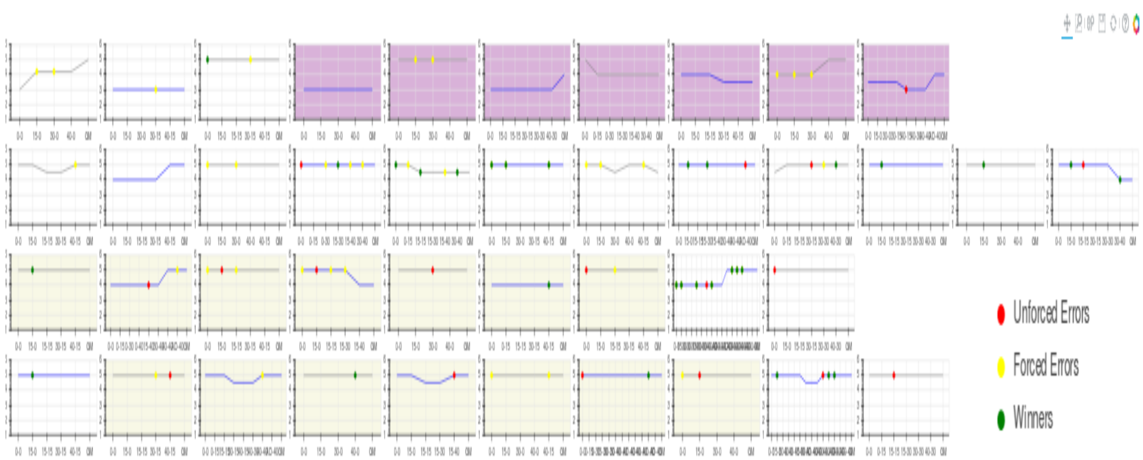

Figure (4.10) Serve Confidence correlated with Busta's Unforced Errors, Forced Errors and Winners

4.7.4 Serve Confidence vs Serve Directions

Figure 3.9 shows us the pattern between both players' SC and serve directions and we can not see much of a correlation pattern between the SC and directions of the serve - namely wide , body or down the T. This is entirely reasonable as in any competitive match, one way to keep your opponent from reading your serve is to not to be consistent with any particular direction of serve. An important question to attempt an answer is that - Is there a serve pattern for any player during their times of high, medium or low confidence? Momentum is denoted with the color background where the purple color denotes PCB's momentum Set 1, Game 4 to Set 1, Game 10 and beige color denotes KA's momentum, Set 3 Game 1 to Set 3 Game 7 and Set 4 Game 2 to Set 4 Game 7. From the angle of momentum we see that in case of PCB momentum his serves hardly go down the T and most of his serves are wide and a few towards the body. In the cases where KA has the momentum we do see an increase in the serves that are down the T. There also are comparable numbers of PCB serves that go wide and towards the body when KA has the momentum. In Set 1 which PCB carried, KA has many serves which are down the T and a few wide. In subsequent sets that ratio appears to improve and KA has some more serves towards the body.

4.7.5 Serve Confidence vs Unforced Errors, Forced Errors and Winners

Figure 3.10 shows us the pattern between both players' SC and PCB's unforced errors, forced errors and winners. It can be figured out that the number of unforced errors is seen especially in Set 2 , Set 3 and Set 4 while remarkably low in Set 1. We know that PCB carried Set 1 and had momentum from Game 4 onwards till he won the set, while he lost Set 2, Set 3 and Set 4 on his way to losing the match. Also the pattern of unforced errors is that they are quite spread out in the games - beginning, middle or end of the games. Further, unforced errors have mostly associated with the decline in PCB's confidence trend line as seen in Figure 3.5 but here the association is holding despite SC holding high and steady from Set 2 onwards. PCB's shot-winners are not seen to have any particular pattern in the plots in general but in the Set 3 Game 8 and Set 4 Game 9 have more winners being hit during high SC as those games have entered deuce which suggests extremely high competition. Momentum is denoted with the color background where the purple color denotes PCB's momentum Set 1, Game 4 to Set 1, Game 10 and beige color denotes KA's momentum, Set 3 Game 1 to Set 3 Game 7 and Set 4 Game 2 to Set 4 Game 7. During PCB's momentum he has committed just one unforced error, while there have been quite a few forced-errors. But, during the games where KA has had the momentum, PCB's rate of unforced-errors increases along with forced-errors which also seem to appear with regularity.

4.8 Discussion

In this paper we described a method to build a confidence index, serve confidence and momentum profile for a tennis player during a tennis match and developed data visualizations to correlate those with the performance measures. Our confidence index, serve confidence and momentum profile method is entirely based on the point scores, game scores in a match and how these scores influence the player's confidence and momentum. Some of the limitations of our approach is that although we have developed good techniques to relate the scores with confidence and momentum, a player's confidence and momentum levels may be influenced by other factors as well which are not dependent on the players themselves. Injury profile, previous performances against the same opponent, media generated expectation, fan pressure, etc tend to have some influence and are difficult to model and measure. However, once the match starts, the scores are the primary factors that influence confidence and momentum. In future, we plan to enhance our data analysis and visualization methodologies while also planning to expand our methods to other fields where confidence and momentum based performance analysis is useful.

PART 5

MICRO-LEVEL ANALYSIS AND VISUALIZATION OF TENNIS SHOT PATTERNS WITH FRACTAL TABLES

This chapter discusses a new technique to analyze and visualize the shot patterns in tennis matches. Tennis is a complicated game that involves a rich set of tactics and strategies. The current tennis analysis are usually conducted at a high level, which often fail to show the useful patterns and nuances embedded in low level data. However, based on a very detailed database of professional tennis matches, we have developed a system to analyze the serve and shot patterns so that an user can explore questions such as "What are the favorite patterns of this player? What are the most effective patterns for this player?" This can help tennis experts and fans gain a deeper insight and appreciation of the sport that are not usually obvious just by watching the match.

5.1 Data

The tennis match data is made available by Tennis Abstract [8], a project maintained and contributed to by tennis fans who have contributed the shot-by-shot statistics of around 4860 matches which contain shot-by-shot data for every point of a match, including the type of shot, direction of shot, depth of returns, types of errors, etc. [8] contains both the raw point-by-point data from these matches and extensive match-level aggregate totals. Match scores are available for both men's and women's game. The metadata for the games and statistics include the player names, tournament, date, surface, etc [8]

5.2 Method

Our visualization technique is designed to analyze an inclusive four-shot space in tennis matches. We know from [31] that majority of the rally length in tennis , consistently close to 70% falls within a four shot space which includes the serve. Thus an inclusive four shot space from the serve(one shot space) up to a serve with a three shot rally (four shot space) is representative of the whole match for analysis purposes. Towards that end, our method consists of the following.

5.2.1 Fractal Table

A fractal table is a table that can be subdivided recursively as needed. Fractal table for the four shot space is created by recursively dividing a shot space. The one shot space is equally divided into the cells for each type of serve and consequently each cell thus represents a point played in lieu of the type of serve. Tennis has three types of serves known by their placements, namely, Wide(W) , Body(B) and down-the-T(T). So three equal area cells denote the serve placements. The same method is then recursively applied to the two shot space which accounts for the serve and the return so that the area for each cell remains the same for a total of six cells of equal area. In the three shot space, this recursive approach is used to account for the serve, a return and the shot after the return so that the area for each cell remains the same for a total of twelve cells of equal area. The four shot space sees the serve , its return followed by a shot and the subsequent final shot by the returner so that the area for each cell remains the same for a total of twenty four cells of equal area. Because the fractal table can be recursively subdivided in this way they can accommodate tennis points of any rally length. We can also easily group and compare shots with any combination of our choice.

5.2.2 Data Visualization

Figure 4.1 depicts the fractal table which consists of an inclusive four shot space. Each shot space is further subdivided to its relevant categories which for the serve is either Wide(W) , Body(B) or Down The T (T) along and parallel to the horizontal axis. The return-shot is categorized to either a Forehand (F) or a Backhand (B) parallel to the vertical axis. Shot by the server after receiving the return is categorized parallel to the horizontal axis and finally the shot by the player who returned the serve is categorized parallel to the vertical axis and the serve-return axis. Thus each sub-space denotes a permutation category which in the one shot space (the serve) is W, B, T and in the two shot space include F and B returns to each serve type. The returns on the two shot space are demarcated vertically as stated above. The three shot space has divisions denoting permutations for the serve type , the return and a shot by the server. This means a serve type separation along and parallel to the horizontal axis, a return demarcated parallel to the vertical axis and the server's shot demarcated parallel to the horizontal axis. Finally, the four shot space has cell space divisions denoting permutations for the serve type, the return, a shot by the server and a shot by a the returner. The play sequence would be a serve type separation along and parallel to the horizontal axis, a return demarcated parallel to the vertical axis , the server's shot upon receiving the return demarcated parallel to the horizontal axis and finally the returner's shot demarcated parallel to the vertical axis and the serve-return axis.

Each permutation is considered a state space, and thus each state space in the inclusive four shot space is denoted by equal area square cells. Color is used to differentiate the four shot spaces from each other.It can be seen from Figure 4.1 that the one shot (serve) space is colored dark-orange, the two-shot space (serve - return) is colored khaki, the three shot space(serve- return - server's shot) is colored green and the four shot space(serve - return - server's shot - returner's shot) is colored blue.

As an example, in Figure 4.1, the three dark-orange colored boxes only record the aces and service winners. The vertical bisection of the khaki rectangle leads to the left half being forehand return and right half representing the backhand return. The khaki colored box labelled BB denotes serve to the body (B) and returned backhand (B). The box labelled TF denotes serve down the T and return forehand (F). A green colored box labelled BFB denotes a serve to the body (B), followed by a forehand return (F) and the server's backhand shot (B). Likewise box TBF denotes a serve down the T, a backhanded return (B) followed by a forehand shot by the server (F). Here horizontal trisection leads to space to mark the serve type category which means two squares each for a serve type. The vertical bisection of the green rectangle leads to the left half being forehand return and right half representing the backhand return. A blue colored box labelled BFFB denotes a serve to the body (B) followed by a forehand return (F), the server's forehand shot (F) and the returner's backhand shot (B). Likewise a box labelled WBFF denotes a serve wide (W), followed by a backhand return (B), followed by the server's forehand shot (F) and the returner's forehand shot (F). Even here horizontal trisection leads to space to mark the serve type category which means two squares each for a serve type. The vertical bisection of the blue rectangle leads to the left half being forehand return and right half representing the backhand return. We see that each half is further bisected to forehand and backhand halves to accommodate the relevant permutations for the last shot. So, within each of the shot space we denote a point played which follows a particular shot pattern discussed above to the relevant box. A data point is plotted on a relevant square cell within the fractal table. For example a point played in the match which was served wide(W) to the backhand was returned with a backhand(B) and was shot with a forehand(F) and the final shot was a forehand(F) is plotted in the relevant square cell WBFF. The visuaization was implemented in Javascript library D3 [32].

5.2.3 Visual Analytics Techniques

Different visual analytics techniques can be used to understand the data and thus draw inference and insight from our visualization. Starting with the one shot space as in Figure 4.1, we could just look at the aces that have been served and analyze the serve pattern of the players. Apart from the aces, by looking at the two, three and the four shot spaces we can also analyze which player tends to serve either to the backhand or the forehand. This shot placement pattern can be very useful to the players themselves in analyzing how their serve and shot patterns affect their game plan and whether it hurts or helps their chance of winning. Data trends also tell us if the players are attacking or defensive in their approach. For example, a player whose points are mostly won on the one shot space and not many on the four shot space would most likely be an attack minded player with a tendency to finish points earlier. Such players tend to not prefer longer rallies and hence not run much in the

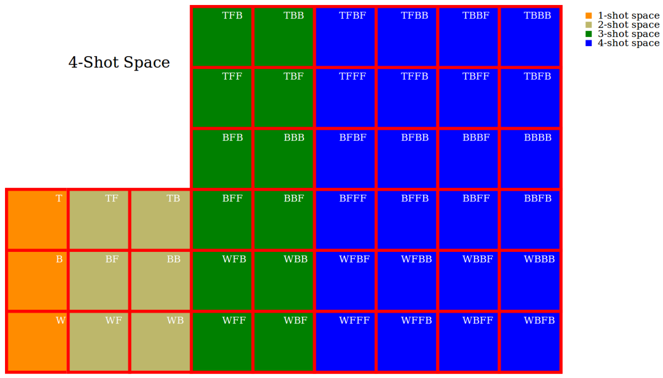

Figure (5.1) A four shot space

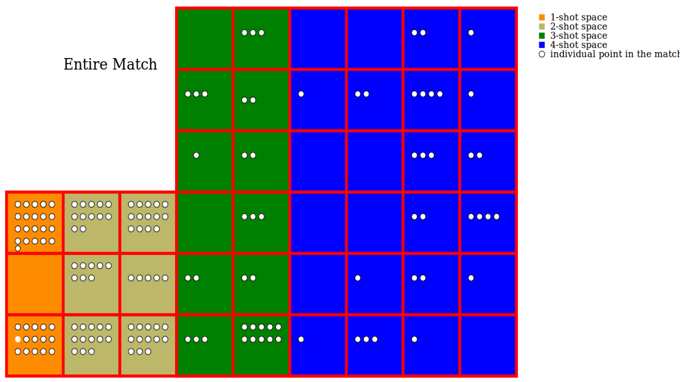

Figure (5.2) Points Plotted for the entire match.

court. But, players who tend to concede more aces with few service winners and who tend to have won more points with longer rallies are defensive in approach. Defensive players tend to run more doing rallies to outlast their opponent by capitalizing the opponent's error.

5.3 Case Studies

In this section we demonstrate our method and visualization output using the match data for the 2017 US Open semi-final between ATP players Kevin Anderson and Pablo Carreno Busta (henceforth referred to as KA and PCB respectively), where KA defeated PCB 4-6, 7-5, 6-3, 6-4. Our analysis and visualization are based on the match data from Tennis Abstract.

5.3.1 Full Match Analysis

The two players played 247 points, including 36 one-shot points, 65 two-shot points, 31 three-shot points and 31 four-shot points. out of which 163 points were within an inclusive four-shot space. Therefore, 66% of the shots are within the one to four-shot patterns. In our analysis, we focus on the first four shots in tennis for several reasons. Above four shots, there are relatively few points for each shot pattern for statistical analysis. Some leading tennis analysts, such as Craig O'Shannessy (tennis analysts for the New York Times and world 1 Novak Djokovic), believe that the first four shots are the most important in tennis. We also want to keep our fractal tables relatively small so they can fit the space of this paper. Figure 4.2 shows all the one to four-shot points grouped by shot patterns. Each white circle represents a shot.

The data visualization reveals some patterns in the player's shot selections. In the one-shot space (i.e. aces or serve faults), we see that more points were served down the T than wide, and there was no serve to the body. This indicates that serving down the T is more likely to produce an ace. In comparison, the two-shot space contains almost equal number of forehand and backhand return of serves. The two-shot space has the most points than any other shot pattern, perhaps indicating the effectiveness of the players' serves and returns. In the three and four-shot spaces, the majority of the points were played with a backhand return of serve. For example, in the four-shot space, 74% of the points were played with a backhand return of serve. This indicates that serving to the backhand tends to result in short rallies.

5.3.2 Set By Set Analysis

As seen in Figure 4.3 fractal table, the first set which PCB carries, sees a fair number of points on the one shot space which are aces or one shot winners on serve. The two shot space sees more points on the forehand return than backhand return. Infact the two shot space also sees the serves towards the body which had not been observed in the one shot space. The four points played out being served to the body have all had forehand returns

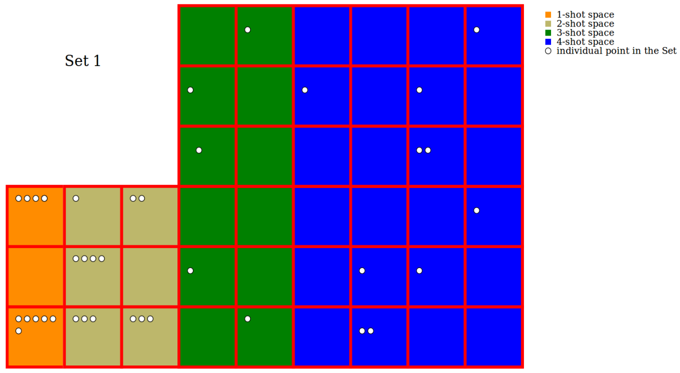

Figure (5.3) Points Plotted for the First set.

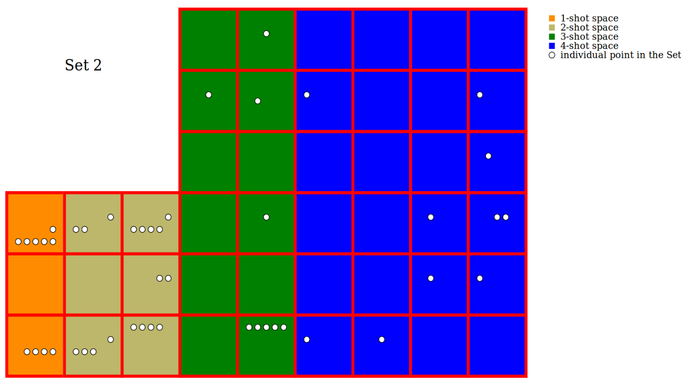

Figure (5.4) Points Plotted for the Second set.

and no points have been played that was served to the body and had a backhand return. Though not many points have been played out in the three shot space the forehand backhand return distribution seems to be fairly even. The four shot space again sees the return skewed towards a backhand return to the serve. As for the serves themselves in the four shot space the serve towards the body are associated with backhand returns. A serve down the T with a forehand return only has a single data point. In the first set the two and three shot spaces sees more forehand returns whereas the four shot space sees more backhand returns.

The second set which KA carries, sees slightly less number of points on the one shot space which are aces or one shot winners on serve, as seen in the fractal table in Figure 4.4. In contrast to the first set the two shot space sees more points on the backhand return than forehand return. The two shot space also sees the serves towards the body which had not been observed in the one shot space. Only two points were served to the body in the two

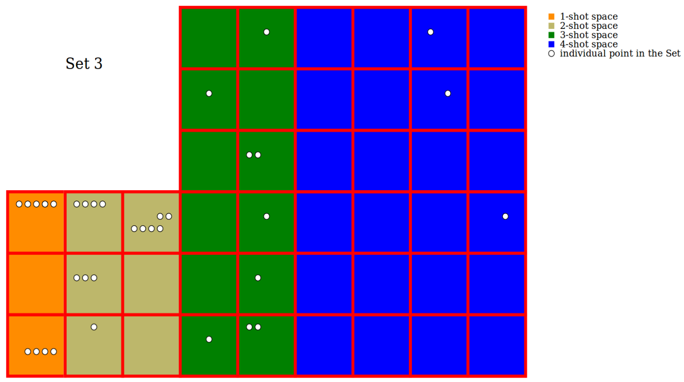

Figure (5.5) Points Plotted for the Third set.

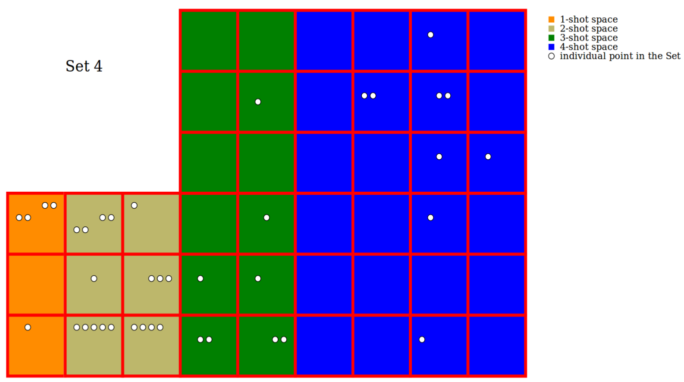

Figure (5.6) Points Plotted for the Fourth set.

shot space and each had backhand return. From the two shot space up to four shot space sees serve return distribution skewed towards backhand return. This trend is in contrast to the first set. The three shot space has a single data point on the forehand return. As for the serves themselves in the four shot space the serve towards the body are a little more associated with backhand returns. A serve down the T with forehand and backhand return only have a single data point each.

As seen in the fractal table in Figure 4.5, the third set won by KA, sees a total of nine points on the one shot space which are aces or one shot winners on serve, with near even distributions on wide and down the T serves. In contrast to the first two sets, in the third set the two shot space sees almost even points on the forehand and backhand returns. In fact the two shot space also sees three points with serves towards the body and had a forehand return. Serve wise more serves were down the T in the two shot space. As in the second set and in contrast to the first set, the three shot space in the third set sees the distribution of the serve return skewed towards backhand return. The three shot space has two data points on the forehand return and seven points on the backhand return. The four shot space just has three data points and sees the return skewed towards a backhand return to the serve. As for the serves themselves in the four shot space the serve wide has no data points, serve towards the body has one data point with backhand return. A serve down the T with backhand return has a two data points each with the final shot being forehand and backhand each.

As seen in the fractal table in Figure 4.6, the fourth set which KA carries, sees a total of five points on the one shot space which are aces or one shot winners on serve, with just a single point served wide and rest all served down the T. In contrast to the first two sets, and similar to the third set, the fourth set sees the two shot space nearly even on points on the forehand and backhand returns. In fact the two shot space also sees three points with serves towards the body and had a backhand return and one with a forehand return.Serve wise more serves were Wide and serves to the body and down the T almost equal in number as compared to the third set where serve wise more serves were down the T in the two shot space. In contrast to the second and third set, the three shot space in the third set sees forehand backhand return distribution slightly skewed towards backhand return. In fact the three shot space has three data points on the forehand return and five points on the backhand return. The four shot space looks less sparse as compared to the third set. Here in the fourth set, the four shot space just has two data points on the forehand return with serve down the T. No serves were seen wide and towards the body. The return skewed towards a backhand return to the all three types of serve.

5.3.3 Service Games Analysis

The fractal tables in figures 4.7 and 4.8 respectively show us the first serve and second serve service games for both players for the entire match. For the first serve service games in Figure 4.7 we see that in the one shot space, KA has hit plenty of aces to give himself an edge

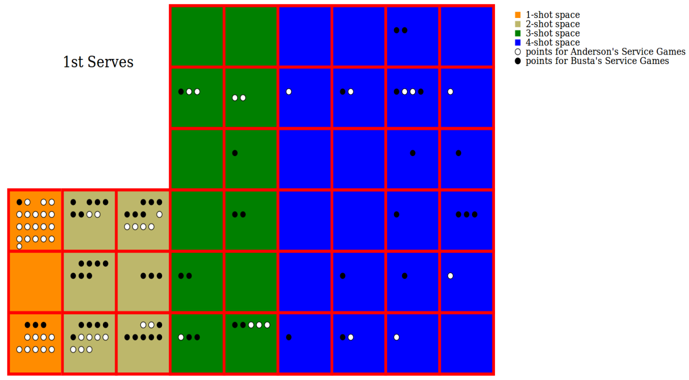

Figure (5.7) First Serve Service Games for both players.

over his opponent. The trend is that his aces and or service-winners are more towards down the T than wide. As for PCB he has few aces and or one shot winners and they have mostly come serving wide than down the T. The two shot space in Figure 4.7, sees that PCB has much more first serve service games in comparison to KA. PCB's first serves service games are also more evenly distributed serving wide , body and down the T for both forehand and backhand returns. As for KA many of his first serves service games have been served wide and have had a forehand return from the opponent. KA received more backhand returns when he served down the T. In the three shot space PCB has near even distribution of serves with a forehand and backhand returns. KA has slightly more first serves service games with a backhand return than forehand. A three shot space is also where the server's shot finishes the point. From that dimension, when PCB has served wide he has ended the point slightly more with a forehand than backhand. When serving towards the body, PCB has received backhand returns and he has tended to end the point with a little more forehands than backhands. In the four shot space, we see that KA has fewer first serves service games which have gone to the three shot rally as compared to his opponent PCB. KA also has served more down the T as opposed to serving wide. He has no first serves service games served to the body in the four shot space. PCB has more evenly distributed first serves service games serve direction wise. Overall even here the skew tends to be more towards the backhand return for both players serving.

Figure 4.8 gives us the second serves service games for both players. In direct contrast

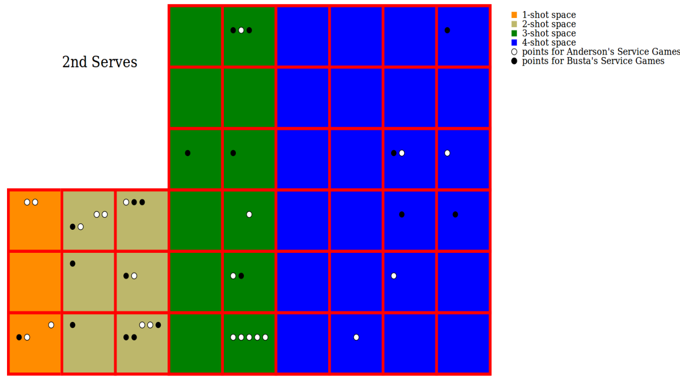

Figure (5.8) Second Serve Service Games for both players.

to Figure 4.7 we see few data points on the one shot space for both players in Figure 4.8. One strong rationale for this trend is that a fault in the second serve would be a double fault leading to the loss of point for the server. So, the server needs to be careful with the serve and hence they tend to play safe so as not to double fault. Figure 4.8 also shows that the skew toward a backhanded return is much less pronounced in the two shot space for both players. In the three shot and four shot space the service return is heavily skewed to backhand. In the three shot space we see that KA has tended to serve more wide as compared to body and T while in the four-shot space his serves have tended more towards wide and body and none down the T. The four shot space also sees PCB serves towards body and T and none served wide. These trends suggest their most comfortable areas of serve when faced with a prospect of a double fault. From the service games we observe that both players are more likely to serve to the opponent's backhand. We can also see looking at the shot spaces that a serve to the forehand is more likely to result in the point ending in the two shot space while serve to the backhand has lead to longer rallies. As for the players themselves, KA has similar serve directions in both the fractal tables in figures 4.7 and 4.8, which suggests he is confident in his serve. The serve trends show that PCB is more likely to serve towards the body. Serving towards the body is serving through the middle where the height of the net is a bit lower and considered safe. This serve trend combined with his lack of aces in the game tells us that PCB is more cautious with his serves. The other reason could be that since KA is taller compared to PCB, serving KA to the body is a tactic used against a tall opponent to slow his return speed by slowing down his body maneuvers.

5.3.4 Point Outcome Analysis

This section details the analysis of the outcomes of points played in the match with regards to unforced errors , forced errors and winners.

Figure 4.9 fractal table shows us the unforced errors committed by the players within the inclusive four-shot space for the duration of the match. A compelling trend observed is that for both players up to the two shot space there is a clear absence of unforced errors. Unforced errors have mostly been committed in the three and four shot space meaning two and three shot rally space respectively. It can be see that in the three shot space a backhand return to an opponent's serve most likely has led to the server's subsequent shot being an unforced error. This trend holds true for both the players. However, in the three shot space, KA's unforced errors have mostly come when he has served wide and down the T whereas PCB unforced errors are more distributed in the serve cell space. For forehand returns KA's unforced errors have come serving down the T and PCB has it more distributed. This suggests that KA has some vulnerabilities when serving wide and down the T. That aspect is something KA could work on to improve his game and for his opponent's to exploit to pressure KA during competitive matches. The four shot space sees more of KA unforced errors than PCB unforced errors. KA has committed unforced errors in all serve directions. Also, PCB has tended to commit unforced errors in backhand returns in all serve directions. A serve toward body and down the T coupled with a backhand return has tended to end in an unforced error in a three shot rally where KA looks more vulnerable compared to his opponent PCB.

Overall when we count the number of unforced errors we see more KA's unforced errors in both three shot and four shot spaces. This means he has unforced errors when both serving and returning. For KA the trend seen thus far is that while he has scored more aces as seen in Figure 4.2, we also see more unforced errors from him. Also seen in Figure 4.9 is a trend that longer rallies tend to favor PCB owing to a distinct lack of unforced errors in

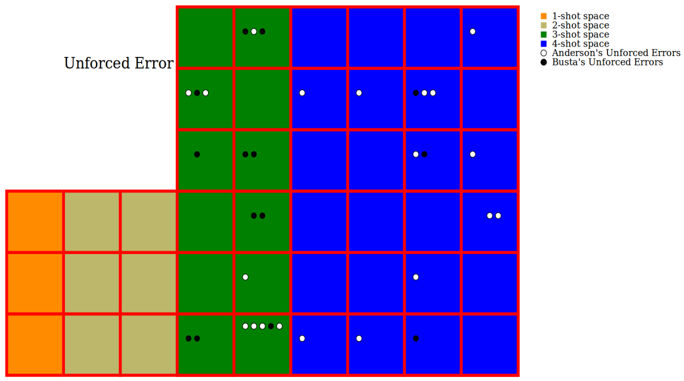

Figure (5.9) Unforced Errors for both players.

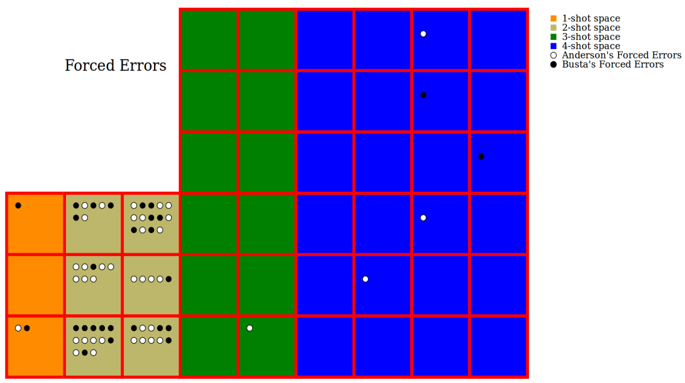

Figure (5.10) Forced Errors for both players.

comparison to his opponent.

Figure 4.10 fractal table shows us the forced errors committed by both players within the inclusive four shot space. In contrast to the unforced errors seen in Figure 4.9, forced errors are overwhelmingly concentrated in the two shot space, in a way that the forced errors seen on the one shot space, three shot space and the four shot space look like isolated data points. Forced errors on the two shot space mean forced errors on returning the serve. In Figure 4.10 we see that KA has more forced errors as compared to his opponent PCB. KA's forced errors are distributed well across both forehand and backhand returns. In comparison to his opponent most of KA's forced errors have come when served to the body while PCB has few forced errors committed when being served to the body. PCB has more forced errors when being served wide and down the T. Notwithstanding the fact that KA has hit more aces, PCB's serves has led to more of KA forced errors in the two shot space. So what this

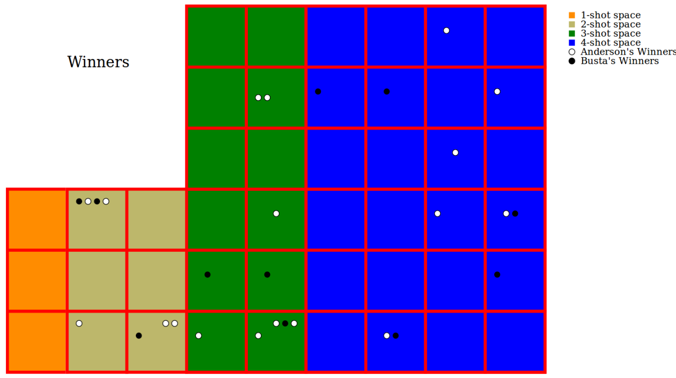

Figure (5.11) Winners scored by both players.

tells us is that although PCB does not have a powerful and speedy serve his serves are in fact good.

The fractal table in figure 4.11 shows us the winners scored by both the players within the inclusive four shot space. The two shot space has a few data points on winners. For both the players in the two shot space most winners are concentrated on two cell spaces, namely, serve wide - backhand return and serve down the T - forehand return. No winners have been scored by either players in the two shot space when they have been served to the body. The three shot space sees most winners concentrated around the backhand return space. Three shot space being a two shot rally indicates winners scored after the server received the return. KA has benefited more from this trend as compared to PCB. All of KA winners in the three shot space have been forehands and the serve area has been all three wide , body and down the T. The four shot space sees a subtle trend where PCB has more winners scored with a forehand return to the opponent's serve. KA has just one data point on the forehand return to his opponent's serve and that has come when having been served wide, while PCB has hit winners when having been served wide and down the T. None has a winner when having been served towards the body in the four shot space. The backhand return to serve seems to have been a good indicator for KA to have scored winners when the serve has been placed towards the body or down the T. KA's winners have come with both forehand and backhand in this space while PCB looks to have benefited from being served wide and towards the body.

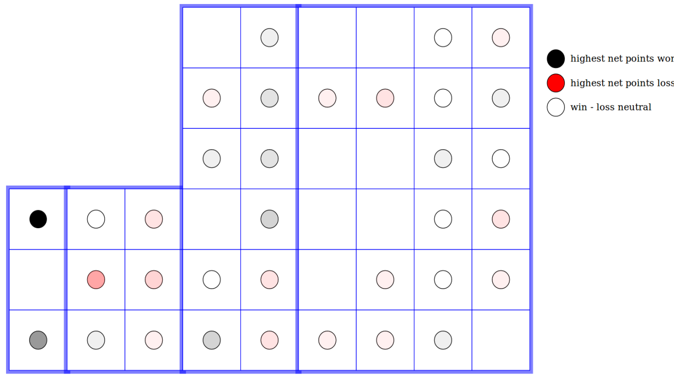

Figure (5.12) Anderson's Net Profit or Loss.

5.3.5 Player's Net Profit or Loss Analysis

The fractal tables in Figures 4.12 and 4.13 show the net points won making a profit or the net points lost by KA and PCB respectively in each cell in the fractal table within the inclusive four shot space. Apart from the background color the fractal table space is exactly the same as stated in Figure 4.1. A circle has been used to represent the win or loss with colors to represent the variations. A win is denoted by black, and loss is denoted by red and win-loss neutral states denoted by white. Color opacity has been used to incorporate the intensity of win and loss. To ensure the proper perspective when looking at data points with varied level of color intensity the fractal table has been rendered to a white background with blue border lines used to separate the shot spaces. Thin blue lines have been used to separate out the cell space with regards to serve, return and shot categories as earlier. In absolute scale the highest net win and loss is 17 points and lowest is 0, the sequence of win and loss point margin has been starting from highest 17 down to 7, 6, 3, 2, 1, and 0 , with 0 being the lowest. In Figure 4.12 which shows KA's win loss chart, we can see that he dominated his opponent with aces serving to the T and he made impressive gains with aces serving wide. In the 2-shot space we can see he has net loss of points in four boxes which were serve to the body with both forehand and backhand returns and serve down the T with a backhand return. There is one neutral box (white), a serve down the T to the forehand and one slightly black circle in the serve wide-forehand return box where his gains were feeble.

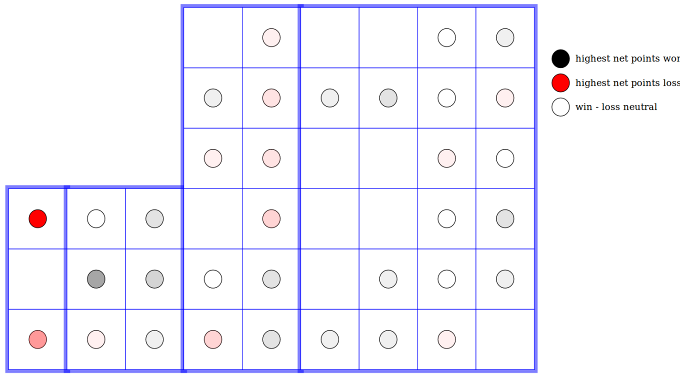

Figure (5.13) Busta's Net Profit or Loss.

The three-shot space indicates KA has lost points while serving or being served wide and down the T. His net loss in the wide serve space has come with a forehand return to the serve. The shot after the serve return where he lost net points is with both forehand and backhand. KA's net loss in the serve space which is down the T has come with a forehand return to the serve. The shot after the serve return where he lost net points is forehand. Although the gains were feeble his net profit in points have mostly come in the serve area to the body. The three shot space sees that serves to the body by both KA and PCB has had KA emerge as a net victor of points when served or serving to the body with either return type and either shot type to the return, apart from the BFF cell (Ref. Figure 4.1), which is serve to the body with a forehand return and a forehand shot space where no points were played out during the entirety of the match. For KA, the four-shot space sees that most of the points have been played out while being served to the backhand. Serve to the forehand has points played out only in five cell spaces as stated in Figure 4.1, namely, WFFF, WFFB, WFBB, TFFF and TFFB which are either serve wide with a forehand return or serve down the T with a forehand return. It is interesting to note that all these five cell spaces sees KA have a net loss of points. Although the net loss is feeble in others the net loss of points seems to be a little higher in TFFB which is serving down the T to the forehand. The four shot space sees no points being played out with serve to the body and a forehand return. The serve to the body with a backhand return looks to be pretty even for both players in the sense that BBFF and BBBB, both serving to the body towards the backhand spaces are neutral spaces for KA as well as PCB. KA has a net loss of points in the space BBFB serving to the body towards the backhand with server hitting the return with a forehand and point ending with a backhand shot. Further, KA has a slight net advantage in BBBF serving to the backhand with a backhand return. Overall, serve to the backhand is a mixed picture for KA in both serve areas wide and down the T.

Similar is the observations with PCB's net profit-loss chart seen in Figure 4.13, since we know that KA's net win is PCB's net loss and vice versa. PCB got dominated by KA with KA serving aces to the T and PCB also made significant loss with KA serving wide aces. In the two-shot space we can see PCB has a net profit of points in four boxes with one neutral box (white) and one slightly red circle in the serve wide-forehand return box where his net losses were with fine margins compared to neutral. In the three-shot space we see that PCB has won net points and profited while serving or being served wide and down the T. His net loss in the wide serve space has come with a backhand return to the serve. Elsewhere in the wide serve space PCB has had a net win. The shot after the serve return where he lost net points is both forehand and backhand. PCB's net win in the serve space which is down the T has come with a forehand return to the serve. The shot after the serve return where he won net points is forehand. Although the losses were slight his net loss has come in the serve area to the body where the verdict is again mixed as there have been spaces where he made a net profit and where it was neutral. The three shot space sees that serves to the body by both KA and PCB has had PCB emerge as a net loser of points when served or serving to the body with either return type and either shot type to the return. That is apart from the BFF space discussed in Figure 4.1 is a serve to the body, return forehand with the point ending forehand shot, where no points were played out during the entirety of the match. In the four-shot space these five cell spaces as discussed in Figure 4.1, namely WFFF, WFFB, WFBB, TFFF and TFFB all of whom are either serve to the body return forehand or serve down the T return forehand, indicate PCB has a net profit of points having won more than losing to his opponent KA. Although the net win has a small margin in others the net profit of points seems to be a little higher in TFFB which serves down the T with a forehand return. The four shot space sees no points being played out with serve to the body and a forehand return. As discussed above, the serve to the body with a backhand return looks to be pretty even for both players in the sense that both BBFF and BBBB spaces are neutral spaces for either player. PCB has a net profit of points in the space BBFB and a slight net loss in BBBF both of which are serve to the body with a backhand return but differ in point ending shots with backhand and forehand respectively. When it comes to serve wide or down the T serving or being served to the backhand is a mixed picture for PCB carving out slight losses , weak wins or outright neutral spaces.

These data trends suggest that this has been an interesting asymmetrical match. KA has advantages with aces but also makes more unforced and forced errors as the rally goes longer. PCB has advantage in making fewer unforced and forced errors in longer rallies. PCB also has disadvantages with aces though his serves are good as seen from many forced errors by KA. KA has more aces but also has more unforced and forced errors and PCB has good serves as well. The insight we can draw for this match from our fractal table based visualization technique is that once the serve is out of the picture, the open play is pretty even handed and no player has dominated the other, a fact obscured by the final scoreline (KA beat PCB 4-6, 7-5, 6-3, 6-4) where KA loses the first set and makes a three set comeback to victory. In fact a closer look at Figure 4.12 fractal table indicates that in the two to four shot spaces, of the 25 cell spaces wherever the ball has been played KA's profit stands at 10 cells and had incurred losses in 15. But the reason KA seems to have taken the match is due to his superior serve in terms of speed and placement leading to many aces which decisively took the match away from PCB. Overall we see a clear difference and contrast in style of play and tactics preferred by the players.

5.4 Discussion

In this paper we described a method to analyze and visualize the tactical patterns of a tennis match. We have developed a novel fractal table to visualize the tennis patterns and their match statistics. Compared with the conventional high-level tennis match statistics, our method provides a micro-level analysis that reveals the complicated dynamics of a tennis match. This tool can be used by tennis experts, players, coaches and serious fans to analyze and compare a player's performance with regards to various serves, returns and shot patterns. In the future, we plan to expand this micro-level analysis to other aspects of tennis matches and apply this approach to other sports.

PART 6

TACTICAL RINGS : A VISUALIZATION TECHNIQUE FOR ANALYZING TACTICAL PATTERNS IN TENNIS

In this chapter, we present a new technique to visualize and analyze shot-by-shot tactical patterns in tennis. Reserach literature in tennis data analysis and visualization focus more on high-level statistics and overviews such as heat-maps but did not handle data analytics at micro-level. Our visualization technique can reveal patterns and imbalances in a tennis match that lead to a deeper understanding of the game. We demonstrate the application and benefits of our visualization technique with case studies.

6.1 Data

Our work is based on the tennis match data from Tennis Abstract [8], an open source project that provides both high-level and micro-level data of more than 5000 professional tennis matches. The micro-level data includes shot-by-shot descriptions, such as the shot types, the shot directions, the return depth, error types, and many other technical features.

6.2 Method

6.2.1 Tactical Rings

Tactical Rings are circles that can be recursively expanded by shot length, leading to the creation of a set of concentric circles. The innermost ring represents one-shot points (i.e., aces), the second ring represents two-shot points (i.e., a service followed by a return), the third ring represents three-shot points, and so on. Each ring is further divided into multiple segments, which are called cells. Each cell contains points with a particular shot combination, such as ”wide-service - backhand return - forehand - backhand”. A tactical pattern is a short combination of shots, usually 2 to 4 shots. Therefore, in our visualization,

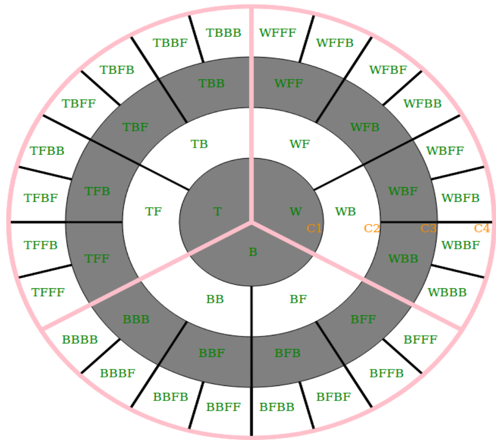

Figure (6.1) A Tactical Ring with four-shot space

the points are visually sorted by tactical patterns.

Due to limited space and also because the majority of the points in professional tennis matches fall within the four-shot space (about 70% according to ATP [31]), the visualizations in this paper only include the first four shots. However, our visualization technique can accommodate longer rallies by adding more rings.

Figure 5.1 shows how Tactical Rings are organized. The innermost ring C1 is the one-shot space (service aces) and is evenly divided by service directions: Wide(W), Body(B), and down the T(T).

The next ring C2 is the two-shot space (service and return). Like C1, the C2 ring is also divided into three service directions. Each service direction segment is further divided into two segments: forehand return and backhand return. Therefore, the C2 ring is divided into six cells, each representing a unique two-shot sequence.

The C2 ring is followed by the C3 ring (the three-shot space). The C3 ring inherits the six segments from the C2 ring, but each segment is further divided into two smaller segments based on the type of the third shot: forehand and backhand. Therefore, the C3 ring has 12 cells, each representing a unique three-shot sequence. The process continues in the C4 ring (four-shot space). Therefore, C4 ring has 24 cells, each representing a unique four-shot sequence. In Figure 5.1, the shot sequence for each cell is labeled in green letters: W(serving wide), B(serving to the body), T(serving down the T), F(forehand), and B(Backhand). For example, the cell marked as WBFF represents the shot pattern: wide-serve-backhand return-forehand-forehand. If letter B is the first letter on the left, it means serving to the body. If B appears in the subsequent letter, it means backhand shot. Therefore, there is no ambiguity.

Points in a tennis match are displayed as dots. Each dot is placed in a particular cell based on the shot sequence of that point. Aces are placed in the C1 ring, two-shot points are placed in the C2 ring, and so on. For example, for a point in which player A serves down the T (T), player B returns with forehand (F), player A hits a forehand shot (F), and player B hits a forehand shot (F), this point will be placed in the cell marked as TFFF. To avoid crowdedness, if too many dots are placed in one cell, we aggregate multiple dots into bigger dots. Additional information, such as the outcome of the point and the type of error, can be encoded in the color and shape of the dots.

We choose to use a ring layout because it is the most natural and spatially efficient layout for this data visualization. It can expand and shrink naturally and evenly in all directions. Longer points are placed on the outer rings because they have more patterns. For example, to visualize the five-shot points, we simply add a bigger C5 ring outside of the C4 ring. The bigger circumference of the C5 ring would naturally accommodate the larger number of cells (48) for the five-shot space. In comparison, a table-based data visualization would need to be extended either vertically or horizontally.

The visualization is implemented in Javascript library D3 [32].

6.2.2 Visual Analytics

Using Tactical Rings, we can conduct three types of analysis: static analysis, dynamic analysis, and comparative analysis. In static analyses, we can study a player's shot pattern distribution and ask questions such as the following.

- Does this player have an advantage when he/she serves in a certain direction?

- Does this player have favorite tactical patterns?

- Does this player have an advantage in short points or long points?
- Does this player make more errors in certain shot patterns (e.g., backhand rallies, lateral movement)?

In dynamic analyses, we can use Tactical Rings to study how a player uses different tactical patterns at different stages of the match. Here are some examples.

- Does this player change tactics as the match progresses (e.g., as the player is getting tired)?
- Does this player have favorite tactical patterns at critical moments (e.g., game points, breakpoints, must-win points)?

In comparative analyses, we can use Tactical Rings to visually compare the strengths and weaknesses as well as the styles of multiple players. Tactical Rings can visualize the tactical imbalances between two players. For example, we can study the strengths and weaknesses of two players by using color to differentiate the points won by each player or the errors made by each player in Tactical Rings. We can highlight the cells with long forehand rallies or backhand rallies to see who has a better forehand or backhand. Which player tends to win long points?

6.3 Case Studies

In this section, we use several case studies to demonstrate the application and usefulness of Tactical Rings. We use the three matches played by Rafael Nadal and Roger Federer at Wimbledon (grass court) from 2006 to 2008 [8] and the five matches between them in the French Open (clay court) from 2005 to 2011 [8].

6.3.1 Cumulative Match Analysis for Grass and Clay

Figure 5.2 shows the match data (first four shots) from Wimbledon. Figure 5.3 shows the match data (first four shots) from Roland Garros (the French Open). We can see that

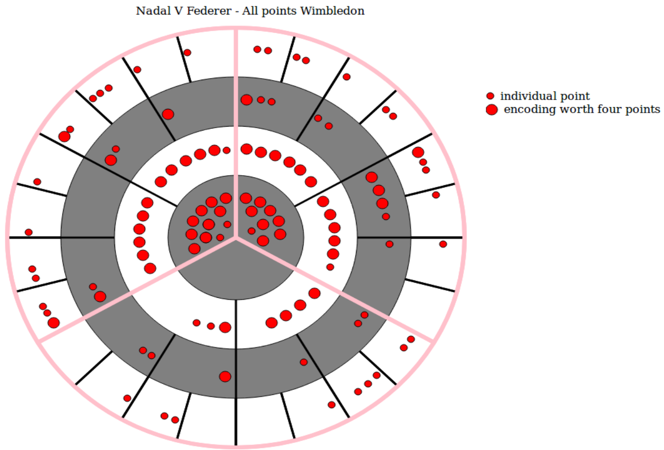

Figure (6.2) All points Wimbledon (grass court)

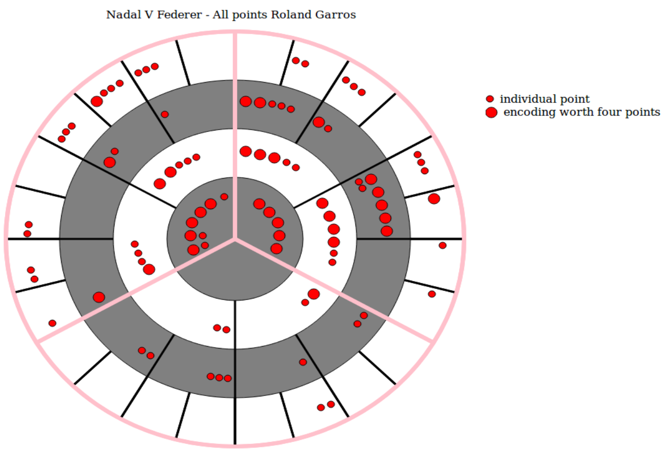

Figure (6.3) All points Roland Garros (clay court)

at Wimbledon, more points are played in the one or two-shot spaces. However, at Roland Garros, the points are more evenly distributed. There are far more aces at Wimbledon (grass) than at Roland Garros (clay). It also shows that points tend to end quicker on grass courts than on clay courts. In the one-shot space (i.e., aces or serve faults), we see serving to the body produces no ace.At Wimbledon, there are more aces serving to the T than serving wide, while at Roland Garros, there are aces serving wide than serving to the T.

The two-shot space has the most points than any other shot pattern, indicating the effectiveness of the players' serves and returns (i.e., either a return error or a return winner). In the two-shot space, the forehand and backhand returns are almost balanced. However, in the three and four-shot spaces, we see an imbalance – the majority of the points have a backhand return of serve, and this pattern exists for both grass and clay surfaces.

6.3.2 Point Outcome Analysis

Figures 5.4 and 5.5 show the unforced errors in the Tactical Rings. Unforced error is one of the most important metrics of a player's quality of play. From Figure 5.4, we see

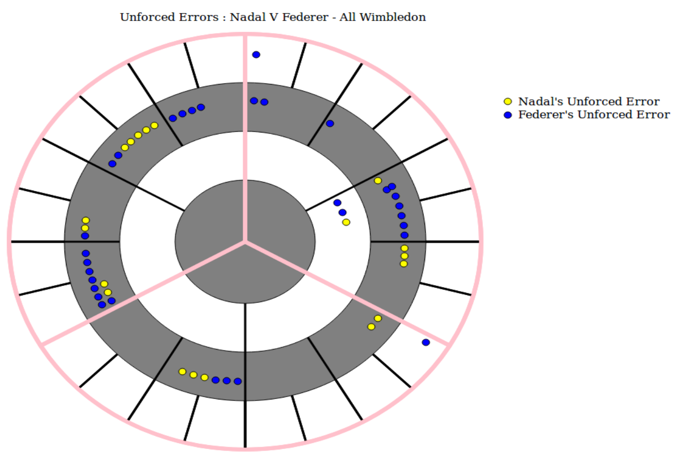

Figure (6.4) Federer v Nadal : Unforced Errors at Wimbledon

that the vast majority of the unforced errors happen in the three-shot space, and somewhat surprisingly, very few two-shot and four-shot points end in unforced error. Besides, Federer made more unforced errors than Nadal in the three-shot points, and he made more unforced errors in the WBF and TFF patterns. Note that in both patterns, the points end with a forehand shot. A plausible explanation is that both players played aggressively in the third shots (thus leading to unforced errors), and Federer was more aggressive than Nadal on the grass court, particularly using his forehand. It seems that Federer had two favorite tactical patterns on grass: WBF and TFF. In the WBF pattern, Federer served wide to Nadal's backhand and then played aggressively with his forehand shot. In the TFF pattern, Federer served to the T (Nadal's forehand) and then played aggressively with his forehand.

It seems that both players were not very aggressive in their returns (hence the few unforced errors in returns), perhaps because the serves are fast on the grass court.

From Figure 5.5, we see that most of the unforced errors are in the two-shot and three-shot points, and very few four-shot points end in unforced error. Both players played aggressively in their third shots, particularly with forehand shots. Again, Federer made more unforced errors than Nadal did, indicating that he was the more aggressive player. Compared with Wimbledon, both players were more aggressive in their service returns on the clay court, hence the higher number of unforced errors in their second shots. Federer still favored the WBF pattern on the clay court, but did not use the TFF pattern very often.

Nadal made many unforced errors in the WB and WBF patterns at Roland Garros but not at Wimbledon. This seems to indicate that Nadal returned Federer's wide serves more aggressively with his backhand on the clay court. Nadal also used WBF pattern more aggressively on the clay court, serving to Federer's backhand and then playing aggressively with his forehand.

From the data visualization, the imbalance of unforced errors between the two players is evident. The difference between Wimbledon and Roland Garros in the two-shot space is intriguing. The many unforced errors in the three-shot WBF pattern also raise interesting questions.

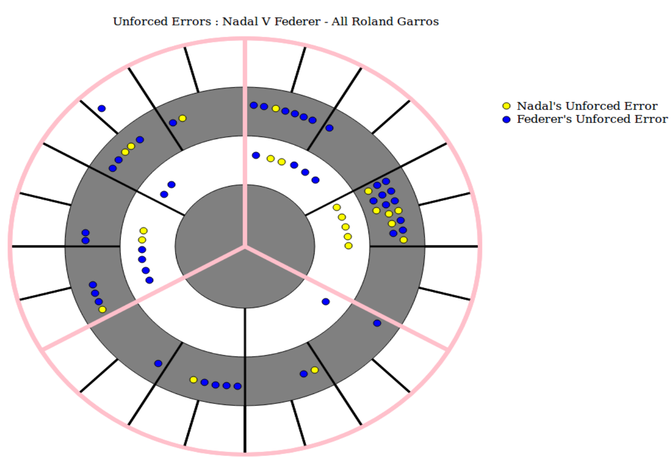

Figure (6.5) Federer v Nadal : Unforced Errors at Roland Garros

Here we assume that most unforced errors are caused by aggressive play. Other factors may also cause unforced errors, such as fatigue, timing, or wind. However, for top-level

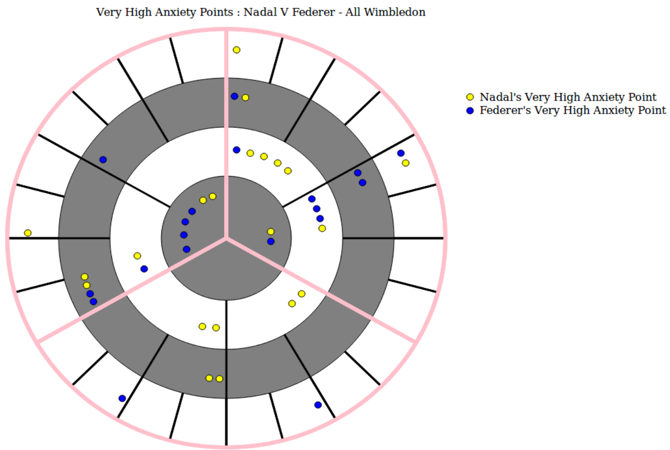

Figure (6.6) Federer v Nadal : Very High Anxiety Points at Wimbledon

players like Federer and Nadal playing in grand slam finals, we believe overly aggressive play is still the primary factor for unforced errors, especially within the first four shots. Our data visualizations are meant to reveal interesting patterns and raise questions for further investigation, not necessarily to give definitive answers. The true causes of these unforced errors may need to be investigated via detailed video analysis.

6.3.3 Very High Anxiety Points Outcome Analysis

Figures 5.6 and 5.7 show the very high anxiety points won by Federer and Nadal in Tactical Rings. Very high anxiety points [29] [30] are critical points (e.g., deuce points) whose outcome may lead the game to either direction.

In Figure 5.6, we see that Federer won four aces by serving to the T, and only one ace by serving wide. In Figure 5.7, he won three aces by serving to the T and two aces by serving wide. It seems that serving to the T is advantages for Federer at high anxiety points. On the other hand, there is no clear pattern for Nadal on both grass courts and clay courts. However, Nadal won far more high anxiety points in service return (two-shot space) than Federer did. The data seems to show that Nadal was a better returner at critical moments, while Federer was a better server at critical moments. Nadal returned Federer's wide serves particularly well at high anxiety moments.

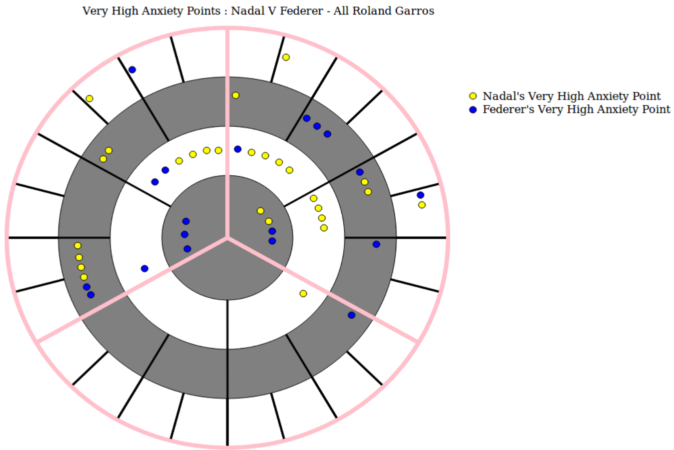

Figure (6.7) Federer v Nadal : Very High Anxiety Points at Roland Garros

From Figure 5.6, we see that, on grass, Federer won more points serving to the T but had no advantage serving wide or to the body. On grass, Federer and Nadal are almost evenly matched in the three and four-shot spaces at critical moments.

From Figure 5.7, we see that, on clay, Federer did better in the three-shot space serving wide, while Nadal did better in the three-shot space serving down the T. There are very few points in the four-shot space. No one has an advantage there.

Overall, our data visualizations suggest that Federer should serve more to the T on the grass court at critical moments. On grass, at critical moments, Federer should serve either wide or to the T, but not to the body. On the other hand, perhaps Nadal should serve more to the T on the clay court at critical moments.

The above analysis shows that Tactical Rings can be used to identify tactical imbalances in tennis games, which may help inform players' strategical decisionmaking.

6.4 Discussion

In this paper, we described Tactical Rings, a novel method to visually analyze the shot patterns of tennis matches. Compared with the conventional tennis match analysis and visualization techniques, our method provides a new way to analyze tennis tactics on a shot-by-shot basis. Through the case studies, we have shown that our visualizations can reveal interesting patterns and dynamic imbalances that are not otherwise obvious. Tactical Rings is an analytics tool not only useful to fans and tennis enthusiasts but also to players, coaches, and analysts who are always looking for ways to improve and gain small advantages in a highly competitive sport. In the future, we plan to expand this technique to other aspects of tennis matches and other sports when such detailed datasets are available.

PART 7

DATA VISUALIZATION AND ANALYSIS OF PLAYING STYLES IN TENNIS

In this chapter, we present a new visual analytics technique called Tennis Fingerprinting to analyze tennis players' tactical patterns and styles of play. Tennis is a game with a variety of styles, tactics, and strategies. Tennis experts and fans are often interested in discussing and analyzing tennis players' different styles. In tennis, style is a complicated and often abstract concept that cannot be easily described or analyzed. The proposed visualization method is an attempt to provide a concrete and visual representation of a tennis player's style. We demonstrate the usefulness of our method by analyzing matches played by Roger Federer and Rafael Nadal at Wimbledon, Roland Garros, and Australian Open. Although we focus on tennis data analysis and visualization in this chapter, this idea can be extended to the analysis of other competitive sports, including E-sports.

7.1 Data

Our analysis is based on the crowd-sourced Tennis Match Charting Project [8] that provides shot-by-shot data of more than 5000 professional tennis matches, including types of shot, directions of shots, depth of returns, types of errors, etc. The shot-by-shot data was created by human charters who watched the match videos and manually entered the data.

7.2 Method

7.2.1 Tennis Fingerprinting:

Tennis fingerprinting is an attempt to visualize a tennis player's distinctive playing style. For example, it displays a player's pattern of serves and returns of an entire match on a point-by-point basis. In this visualization, each block represents a point. Each point

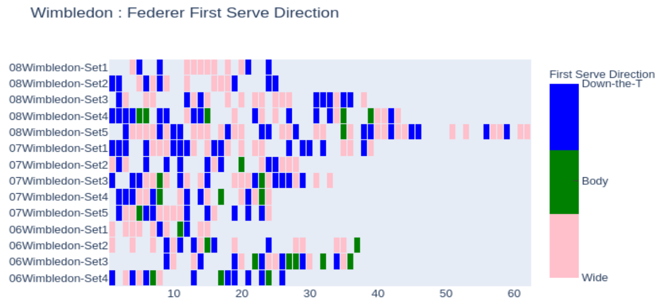

Figure (7.1) Federer First Serve at Wimbledon

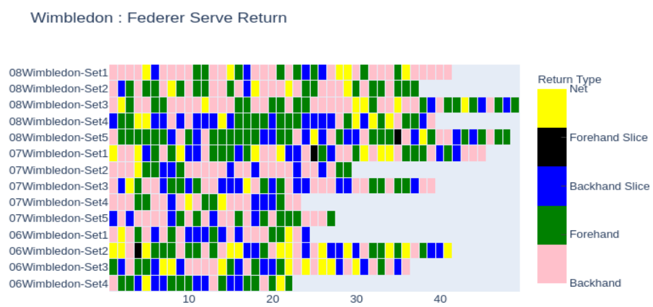

Figure (7.2) Federer Serve Return at Wimbledon

is color-coded for a particular variable to be analyzed (such as serve, return, and rally). Each horizontal line corresponds to a set. In this paper, we focus on two tactical patterns: serve pattern and return pattern. More tactical patterns can be explored using this type of visualization but due to space limit we will discuss them in the future.

Serve Directions: Serve is often considered the most important shot in professional tennis. In a tennis match, a player serves alternatively from two sides of the court: deuce side and ad side. On each side, the player has two chances to serve: first serve and second serve. If both serves are missed, it is called a double fault. For each serve, a player can decide to serve in three directions: wide, body, and down-the-T. The serve direction is also called serve placement.

In professional matches, serve speed and placement are both very important. Serve placement is often part of a tactical plan in which a player tries to win a point by a combination of shots, starting with the serve. Because of this, choosing the direction for each serve is often a complicated decision that may involve the following variables: serve side (deuce or ad), first or second serve, server's strength and weaknesses, the opponent's strengths and weaknesses, current scores, court surface condition, and wind condition. In the mean time, a player also tries to make the serve directions unpredictable to the returner.

Serve Return: Returning serve is also a very important shot in professional tennis. Good returns can create many opportunities to win points and break the opponent's service games. A player's return patterns also reveals his/her tactical decision making. For example, a player can choose to return with backhand or run around to return with a forehand, which is aggressive but risky. However, the types of return shots are often dictated by the serve directions. Therefore, the return pattern is also partly a reflection of the server's serve patterns.

Visualization Design: Figures 6.1 and 6.2 shows two examples of the proposed data visualization. Figure 6.1 visualizes the serve patterns. Figure 6.2 visualizes the service return patterns. The vertical axis shows different sets in selected tennis matches. The horizontal axis is the timeline. Therefore, each row represents the points of a set in a tennis match. Each color-coded block represents a serve or return. For serve patterns, pink represents a wide serve; green represents a serve to the body; blue represents a serve down-the-T. The color coding for the return of serves is as follows. Pink represents a backhand return; green represents a forehand return; blue represents a backhand slice; black represents a forehand slice; yellow represents a return to the net.

7.2.2 Visual Analytics

The proposed data visualizations can help users visually analyze a player's playing style. For example, it can be used to answer the following questions.

1. Does a player have a preferred service direction? Is a player's service pattern consistent

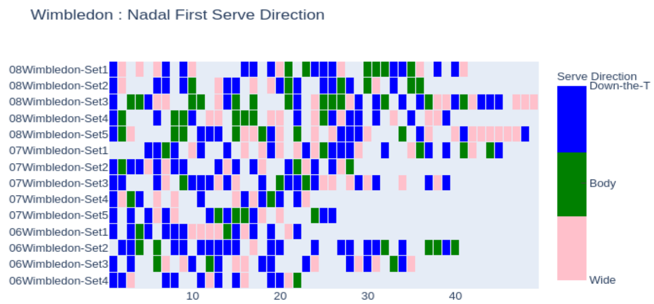

Figure (7.3) Nadal First Serve at Wimbledon

or change from match to match? Which serve pattern works well against a particular opponent?

2. Does a player have a preferred service-return pattern? What kind of return pattern works well against certain opponent?

3. Can we visually see the difference between the styles of different players?

4. Can we visually compare a player's serve or return patterns between matches to see if a player adjusted his/her serve, return style for different opponents?

7.3 Case Studies

In this section, we use several case studies to demonstrate the application of our proposed visualizations. We use the data [8] from three matches played by Rafael Nadal and Roger Federer at Wimbledon (grass court) from 2006 to 2008, the five matches between them in the French Open at the Roland Garros (clay court) from 2005 to 2011 and the two matches in the Australian Open (hard court) in 2009 and 2012.

7.3.1 First Serve Pattern Analysis

In this section, we analyze the first serve patterns for both players at Wimbledon, Roland Garros (French Open), and the Australian Open, respectively.

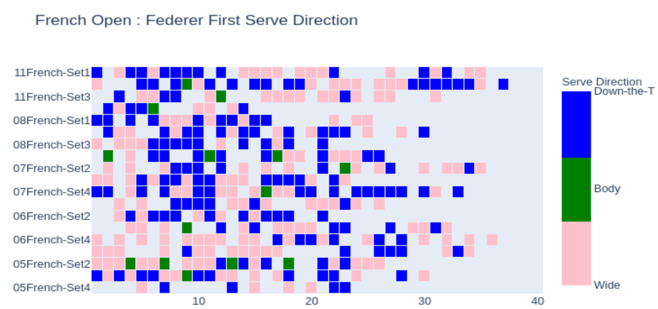

Figure (7.4) Federer First Serve at Roland Garros

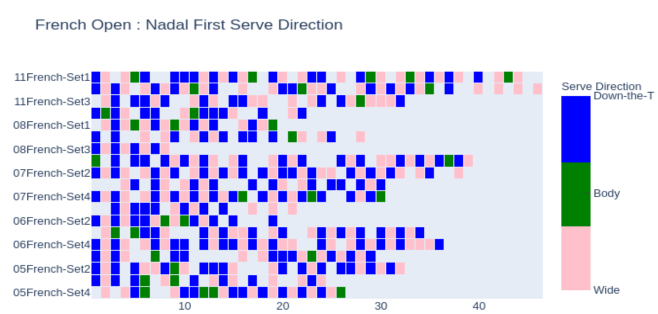

Figure (7.5) Nadal First Serve at Roland Garros

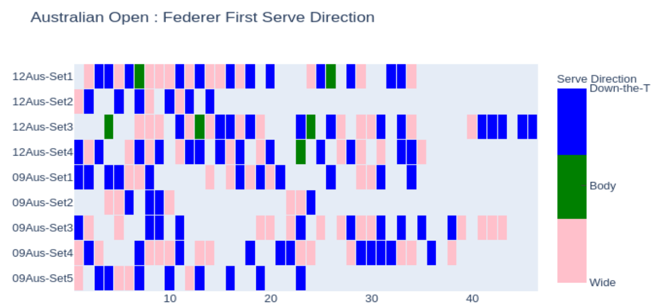

Figure (7.6) Federer First Serve at Australian Open

Figures 6.1 and 6.3 shows Federer and Nadal's first serve patterns at Wimbledon (grass court). From the visualizations, we can clearly see that in Figure 6.1 Federer tended to serve more wide (pink) than down-the-T (blue). Looking at the bars horizontally, we can see that Federer sometimes served wide in consecutive serves. He made only a small number of serves to the body (green). But interestingly, he made a cluster of the body serves in the third and fourth set of the 2006 match.

In contrast, Nadal's first serves at Wimbledon was more balanced, with similar numbers of wide, body, and down-the-T serves. Nadal made more body serves than Federer. Looking vertically at Nadal's figure, we can see that he tended to start each set serving down-the-T, as indicated by the vertical blue bars on the left side of figure 6.3.

Figures 6.4 and 6.5 show Federer and Nadal's first serve patterns at Roland Garros (clay court). The contrast between the two players is quite obvious. In figure 6.4 Federer preferred to serve wide or down-the-T, and made only a small number of the body serves. It seems that Federer liked to serve in clusters: consecutive wide or down-the-T serves, as visualized by the long horizontal pink and blue bars. In contrast, Nadal's serves are a mixed picture, seemly more random, with no obvious clusters. Again, we see Nadal had a tendency to start each set serving down-the-T, as indicated by the vertical blue bars on the left side of figure 6.5.

Figures 6.6 and 6.7 show Federer and Nadal's first serve patterns at the Australian Open (hard court). Again, we can see that Federer did not like to serve to the body. In the 2009

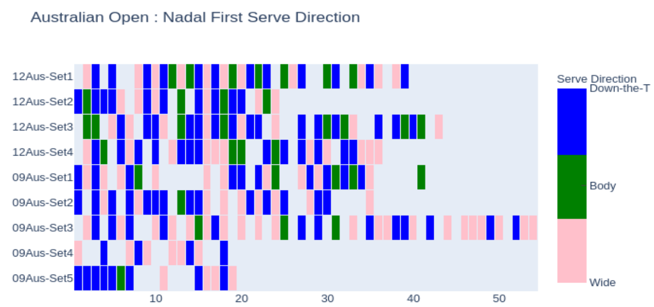

Figure (7.7) Nadal First Serve at Australian Open

match (bottom half of figure 6.6), there seems to be no body-serve. Again, Nadal's serves are more mixed. We still see strong examples of Nadal's tendency to start a set serving down-the-T. In figure 6.7, we can see two clusters of down-the-T serves (blue bars) at the beginning of two sets.

By analyzing the visualization of serve patterns for the three major tournaments, we can see some consistent patterns for each player. Federer took more risks in his first serves, serving wide and down-the-T and only a small number to the body. He tended to serve wide in consecutive serves, showing high confidence in this serve placement. Nadal's serves seem more mixed and random, but he had the tendency of serving down-the-T at the beginning of a set, perhaps a sign of confidence in this serve placement. From the visualization, we see no obvious change in each players first serve patterns on different court surfaces, indicating a largely consistent serving style.

7.3.2 Second Serve Pattern Analysis

Players are generally less aggressive in their second serves. As a result, second serves are generally slower than the first serves, and the placements of second serves also have more margin for errors. Therefore, there is generally more second serves to the body than the first serves. Figures 6.8 and 6.9 show the second serve patterns for both players at Wimbledon (grass court). The difference between their second serve patterns is not as distinctive as their first serve patterns, although Nadal still served more to the body (green) than Federer.

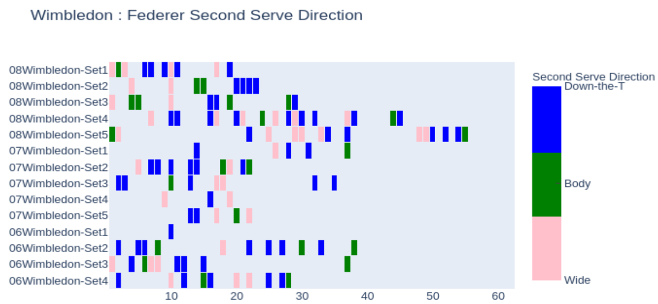

Figure (7.8) Federer Second Serve at Wimbledon

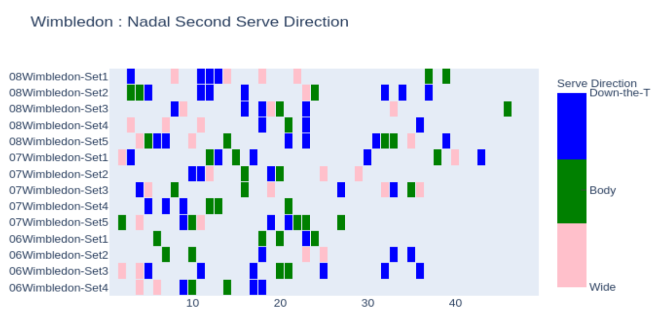

Figure (7.9) Nadal Second Serve at Wimbledon

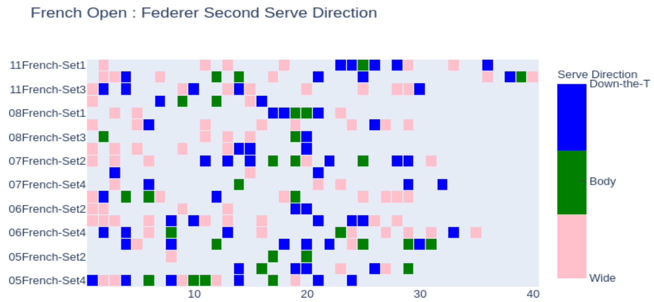

Figure (7.10) Federer Second Serve at Roland Garros

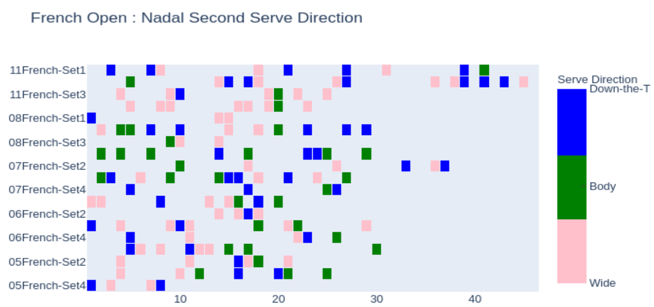

Figure (7.11) Nadal Second Serve at Roland Garros

Both players served more down-the-T (blue) than wide (pink).

Figures 6.10 and 6.11 shows the second serve patterns for both players at Roland Garros (clay court). Both players served more to the body (green), particularly Federer. Compared with Wimbledon, Federer served more wide (pink) than down-the-T (blue). Again, Nadal's second serves are more evenly mixed. We also notice that Nadal chose to serve down-the-T (blue) at the start of multiple sets.

Figures 6.12 and 6.13 shows the second serve patterns for both players at the Australian Open (hard court). Again, we can see Federer made only a few body serves, and Nadal started four sets with serving down-the-T (see the blue bars on the left of Figure 6.13.

From the visual analysis, we see that Federer was still taking more risks in his second serves, opting to serve more to wide and down-the-T. However, on his less favorite clay courts, he seemed to be more cautious and served more to the body. Nadal's second serves

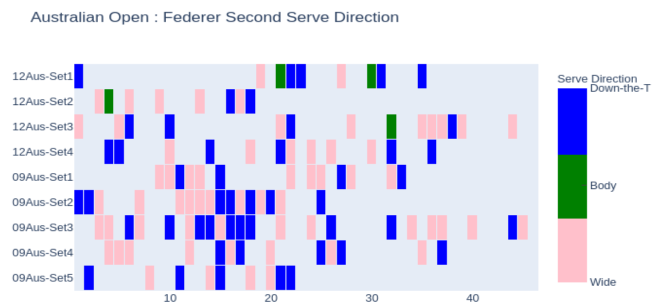

Figure (7.12) Federer Second Serve at Australian Open

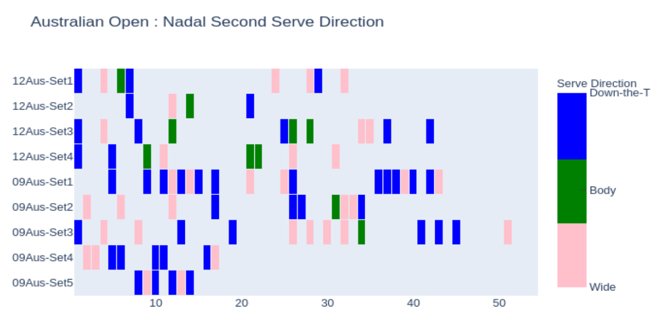

Figure (7.13) Nadal Second Serve at Australian Open

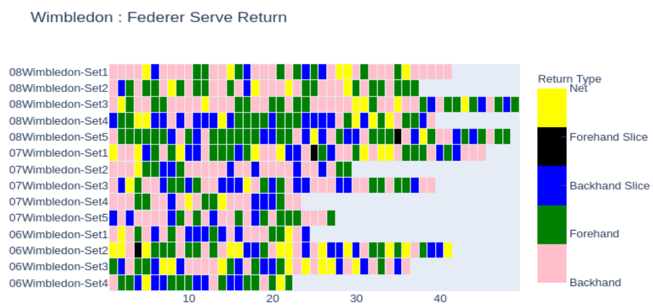

Figure (7.14) Federer Serve Return at Wimbledon

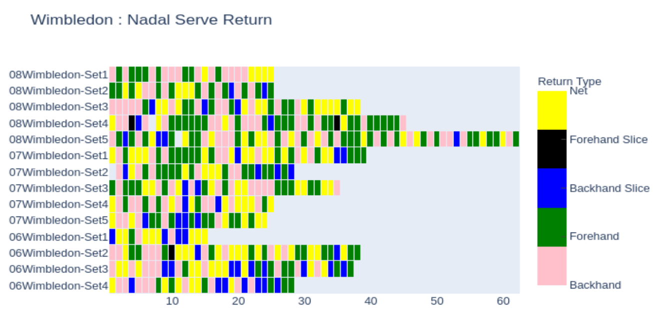

Figure (7.15) Nadal Serve Return at Wimbledon

are more or less evenly mixed, like his first serves. However, we also get a glimpse of his tendency to serve down-the-T at the beginning of many sets. Even in second serves, the visual patterns seem to suggest that Federer slightly favored wide serves while Nadal slightly favored down-the-T serves.

7.3.3 Return Pattern Analysis

In this section we analyze the service returns for both players on grass, clay, and hard courts.

Figures 6.14 and 6.15 show the service return patterns for both players at Wimbledon (grass court). From the visualization (Fig. 14), we can see that Federer made many backhand returns (pink) in the first three sets of their famous 2008 match. However, in the fourth and fifth sets, Federer made notably fewer backhand returns but more forehand returns (green)

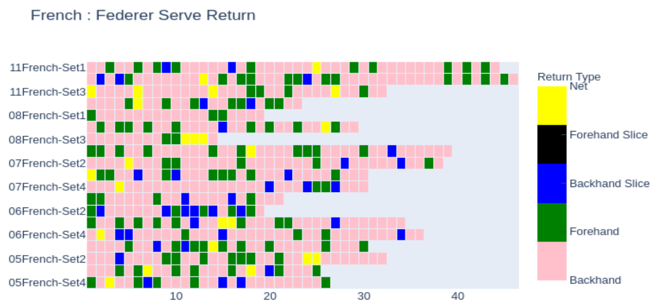

Figure (7.16) Federer Serve Return at Roland Garros

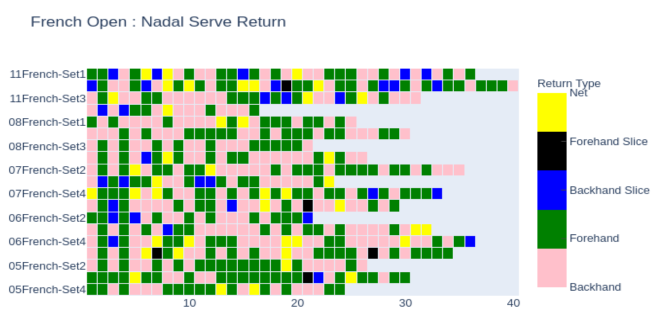

Figure (7.17) Nadal Serve Return at Roland Garros

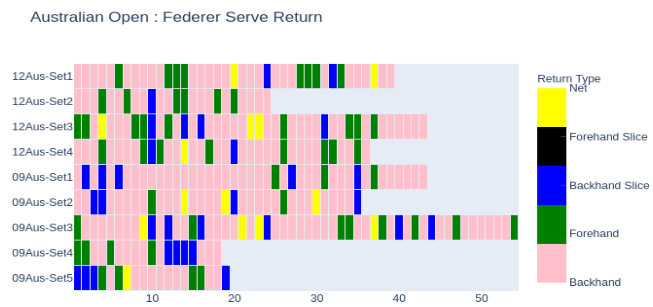

Figure (7.18) Federer Serve Return at Australian Open

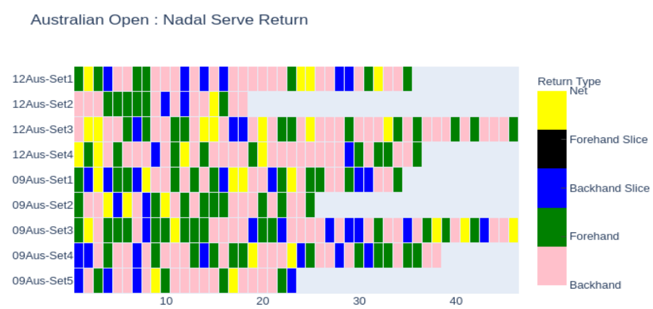

Figure (7.19) Nadal Serve Return at Australian Open

and backhand slice returns (blue). Again, in the final set of their 2006 match, we also see very few backhand returns (pink). Does this indicate a change in Federer's return strategy or Nadal's serves strategy? The visual analysis raises an interesting question.

On the other hand, in Fig. 15 Nadal made fewer backhand returns (pink) and very few backhand slices (blue), perhaps because he has a two-handed backhand. However, Nadal made more errors than Federer returning to the net (yellow), an indicator of the effectiveness of Federer's serves on grass courts.

Figures 6.16 and 6.17 show the service return patterns for both players at Roland Garros (clay court). From the visualization (Figure 6.17), we can see that Federer made far more backhand returns (pink) than other types of returns. This clearly shows Nadal's strategy of relentlessly serving to Federer's back on clay court. In the meantime, Nadal made many returns with his forehand (green). By comparing Fig. 14 and 16, we can see that Federer was forced to play a very different type of return game on clay. This is one of the reasons why he never performed well against Nadal on clay courts.

Figures 6.18 and 6.19 show the service return patterns for both players at the Australian Open (hard court). In Fig. 18, we can see that Federer was forced to make a large number of backhand returns (pink) like he did at Roland Garros. Comparing Figure 6.14, 6.16, and 6.18, we clearly see Nadal's strategy against Federer was serving to Federer's weaker backhand. But for some reason, Nadal was unable to do this to Federer on the grass court, or perhaps Federer was able to handle it effectively on the grass. It is an interesting question for further exploration. Comparing Fig. 15, 17, and 19, we do not see significant differences in Nadal's return patterns on different surfaces.

7.4 Discussion

In this chapter, we described a visualization technique to analyze tennis players' playing style. Tennis experts and fans are often interested in discussing and analyzing tennis players' different styles. Watching the clash of different styles, such as Federer and Nadal, is one of the main reasons many people enjoy tennis. However, style is also a complicated and often abstract concept that cannot be easily described or analyzed. The proposed visualization method is an attempt to provide a concrete and visual representation of a tennis player's style. This visualization technique, combined with expert knowledge, can lead to the discovery of deeper insight into this sport.

We demonstrated the usefulness of our methods with case studies of Federer and Nadal's matches at Wimbledon, Roland Garros, and Australian Open. Using our visualizations, we are able to discover some interesting patterns in both Federer and Nadal's service and return games. The contrast between the two players is clearly visible in our visualizations. In the current work, we only focused on serves and returns, the starting point of each point. In the future, we plan to expand our work to visualize tennis players' different styles of play during rallies.

Our idea can be extended other competitive sports to analyze player behavior and styles as long as micro-level performance data is available. This can be particularly useful for video games and Esports, where detailed performance data can be collected automatically in real-time.

PART 8

CONCLUSION AND FUTURE DIRECTIONS

In this dissertation, I have showcased multiple aspects of visual analytics in Sports with case studies that are real tennis matches.

I described a method to computationally analyze and visualize the relationship between a player's anxiety and performance in a tennis match. The case study shows that we are able to discover useful patterns through these analysis and visualizations. I also described a method to build a confidence index, serve confidence and momentum profile for a tennis player during a tennis match and developed data visualizations to correlate those with the performance measures. The methods have limitations though. It is an indirect estimate of an athlete's anxiety, confidence and momentum status, not a direct measurement of the athlete's physiological state (which is usually difficult to do in a professional match). The methods do not include factors outside of the match, such as physical condition, environment, etc.

This dissertation further described two methods to analyze and visualize the tactical patterns of a tennis match with the development of novel fractal table and tactical rings to visualize the tennis patterns and their match statistics. Compared with the conventional high-level tennis match statistics, fractal tables and tactical rings enable micro-level analysis that reveals the complicated dynamics of a tennis match.

Lastly, the dissertation described a visualization technique to analyze tennis players' playing style. Since style is also a complicated and often abstract concept that cannot be easily described or analyzed. The proposed visualization method is an attempt to provide a concrete visual representation of a tennis player's style. This visualization technique, combined with expert knowledge, can lead to the discovery of deeper insight into this sport.

There are also exciting new paradigms on which this research can expand. One of them would be to visually analyze more psychological metrics and build a comprehensive psychological analysis tool which can track players' mental condition more holistically.

The visual representation and analysis methods developed and showcased in this dissertation can be used to expand visual data analytics to other fields where relevant datasets are available.

In visual analytics, answers to much of "what more can be explored" is almost always a function of availability of relevant data. But some of the additional research questions mentioned above could be well explored with the dataset primarily used for this research.

REFERENCES

[1] X. He and Y. Zhu, "TennisMatchViz: a tennis match visualization," in Proceedings of International Conference on Visualization and Data Analysis (VDA), 2016.

[2] M. Madurska, "A set-by-set analysis method for predicting the outcome of professional singles tennis matches", Master's Thesis , Imperial College London, 2012.

[3] P. Newton and J. Keller, "Probability of winning at tennis I", Studies in Applied Mathematics, Theory and data Studies in Applied Mathematics, 114 (3), pp. 241-269, 2005.

[4] L. Riddle, "Probability models for tennis scoring systems", Journal of the Royal Statistical Society. Series C (Applied Statistics), 37(1), pp. 63-75, 1988.

[5] D. Jackson and K. Mosurski, "Heavy defeats in tennis: Psychological momentum or random effect?", CHANCE, 10(2), pp. 27-34, 1997.

[6] I. MacPhee, J. Rougier, and G. Pollard, "Server advantage in tennis matches", Journal of Applied Probability, 41(4), pp. 1182-1186, 2004.

[7] S. Kovalchik, M. Ingram, "Hot heads, cool heads, and tacticians: Measuring the mental game in tennis", MIT Sloan Sports Analytics Conference, 2016.

[8] Tennis Match Charting Project, `https://github.com/JeffSackmann/tennis_MatchChartingProject`.

[9] Bokeh, `http://www.bokeh.pydata.org`.

[10] D. W. Grupe and J. B. Nitschke. "Uncertainty and Anticipation in Anxiety: An integral neurobiological and psychological perspective", Nature Review Neuroscience, 14(7), pp. 488-501, 2013.

[11] H. Pileggi, C. D. Stolper, J. M. Boyle, and J. T. Stasko, “Snapshot: Visualization to propel ice hockey analytics,” Visualization and Computer Graphics, IEEE Transactions on, vol. 18, no. 12, pp.2819-2828, 2012.

[12] F. Beck, M. Burch, and D. Weiskopf, “Visual comparison of timevarying athletes’ performance,” The 1st Workshop on Sports Data Visualization. IEEE, 2013.

[13] B. Moon and R. Brath, “Bloomberg sports visualization for pitch analysis,” in The 1st Workshop on Sports Data Visualization. IEEE,2013.

[14] D. Demaj, “Geovisualizing spatio-temporal patterns in tennis: An alternative approach to post-match analysis,” 2013. [Online]. Available: `http://gamesetmap.com/?pageid=2`.

[15] R. Cava and C. Dal Sasso Freitas, “Glyphs in matrix representation of graphs for displaying soccer games results,” The 1st Workshop on Sports Data Visualization. IEEE, 2013.

[16] R. Sisneros and M. Van Moer, “Expanding plus-minus for visual and statistical analysis of nba box-score data,” in The 1st Workshop on Sports Data Visualization. IEEE, 2013.

[17] S. G. Owens and T. Jankun-Kelly, “Visualizations for exploration of american football season and play data,” The 1st Workshop on Sports Data Visualization. IEEE, 2013.

[18] C.Goulet, Chantal Bard, and Michelle Fleury,”Expertise Differences in Preparing to Return a Tennis Serve: A Visual Information Processing Approach”, Journal of Sport and Exercise Psychology,11, 382-398,1989.

[19] R.Weinberg, A.Jackson,”Building self-efficacy in tennis players: A coach’s perspective”, Journal of Applied Sport Psychology, 2:2, 164-174,1990.

[20] D.Gould , Russell Medbery , Nicole Damarjian, Larry Lauer,”A Survey of mental skills training knowledge, opinions, and practices of junior tennis coaches”, Journal of Applied Sport Psychology, 11:1, 28-50,1999.

[21] A.Bandura,"An agentic perspective on positive psychology",S. J. Lopez (Ed.), Praeger perspectives. Positive psychology: Exploring the best in people, Vol. 1. Discovering human strengths (pp. 167-196), 2008.

[22] Polk, T. and Yang, J. and Hu, Y. and Zhao, Y., "TenniVis: Visualization for Tennis Match Analysis", IEEE Transactions on Visual and Computer Graphics, 2014.

[23] J Liqun , DG Banks "TennisViewer: a browser for competition trees", IEEE Computer Graphics and Applications,Volume: 17,Issue:4 ,1997.

[24] M. Burch, D. Weiskopf,"Tennis Plots: Game, Set, and Match",Diagrammatic Representation and Inference. Diagrams 2014. Lecture Notes in Computer Science, vol 8578, 2014.

[25] X. Wei, P. Lucey, S. Morgan and S. Sridharan, "Forecasting the Next Shot Location in Tennis Using Fine-Grained Spatiotemporal Tracking Data," IEEE Transactions on Knowledge and Data Engineering, vol. 28, no. 11, pp. 2988-2997, 2016.

[26] Y. Wu et al., "iTTVis: Interactive Visualization of Table Tennis Data," in IEEE Transactions on Visualization and Computer Graphics, vol. 24, no. 1, pp. 709-718, 2018.

[27] Y. Wu et al., "ForVizor: Visualizing Spatio-Temporal Team Formations in Soccer," in IEEE Transactions on Visualization and Computer Graphics, vol. 25, no. 1, pp. 65-75, 2019.

[28] H. Janetzko, D. Sacha, M. Stein, T. Schreck, D. A. Keim and O. Deussen, "Feature-Driven Visual Analytics of Soccer Data",IEEE Symposium on Visual Analytics Science and Technology)VAST),pp. 13-22, 2014.

[29] S. Pokharel and Y. Zhu, “Analysis and Visualization of Sports Performance Anxiety in Tennis Matches”, in the Proceedings of the 13th International Symposium on Visual Computing, Lecture Notes in Computer Science, Vol. 11241, Springer, 2018. pp. 407-419.

[30] S. Pokharel and Y. Zhu, "Micro-level Data Analysis and Visualization of the Interrelation Between Confidence and Athletic Performance",in Proceedings of the 2nd International Conference on Computer Science and Software Engineering(CSSE), 2019.

[31] ATP Stats, [Online]. Available: `https://www.atpworldtour.com/en/stats` [Accessed: July 15, 2019].

[32] Design Driven Documents, [Online]. Available: `https://d3js.org/`. [Accessed: July 15, 2019].

[33] X. Wei, P. Lucey, S. Morgan, P. Carr, M. Reid, Machar, S. Sridharan, "Predicting Serves in Tennis Using Style Priors", in Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp 2207-2215, 2015.

[34] S. Pokharel and Y. Zhu, "Tactical Rings : A Visualization Technique for Analyzing Tactical Patterns in Tennis", in the Proceedings of the 14th International Symposium on Visual Computing, Lecture Notes in Computer Science, vol. 11845, 2019. pp. 481–491.

[35] S. Pokharel, Y. Zhu and S. Puri, "Micro-Level Analysis and Visualization of Tennis Shot Patterns with Fractal Tables," in the Proceedings of the 2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC), IEEE, 2019, pp. 97-102.

[36] J. Yacub, "Play Winning Tennis with Perfect Strategy", Wheatmark, ISBN: 978-1-60494-051-0, 2008.

[37] Daniel A. Keim and Daniela Oelke, "Literature Fingerprinting: A New Method for Visual Literary Analysis", In Proceedings of the 2007 IEEE Symposium on Visual Analytics Science and Technology(VAST). pp 115–122, 2007.

[38] Görg, Carsten, et al. "Combining computational analyses and interactive visualization for document exploration and sensemaking in jigsaw." IEEE Transactions on Visualization and Computer Graphics,pp 1646-1663, 2012.

[39] Eric Alexander, et al. ”Serendip: Topic model-driven visual exploration of text corpora.” Conference on Visual Analytics Science and Technology (VAST), 2014.

[40] Steffen Koch, et al. ”VarifocalReader—in-depth visual analysis of large text documents” IEEE transactions on visualization and computer graphics pp 1723-1732, 2014.

[41] L. García-González, D. Iglesias, A. Moreno, M. P. Moreno, and F. Del Villar, ”Tactical knowledge in tennis: A comparison of two groups with different levels of expertise”, Perceptual and Motor Skills, vol. 115, pp. 567–580, 2012.

[42] S. L. McPherson and M. Kernodle, ”Mapping two new points on the tennis expertise continuum: Tactical skills of adult advanced beginners and entry-level professionals during competition”, Journal of Sports Sciences, vol. 25, pp. 945-959, 2007.

[43] N. Kolman, T. Kramer, M. T. Elferink-Gemser, B. C. H. Huijgen, and C. Visscher, ”Technical and tactical skills related to performance levels in tennis: A systematic review”, Journal of Sports Sciences, vol. 37, pp. 108-121, 2019.