

Statistical analysis and Relation between Data

The process of developing the interactive visualization started with cleaning the Tennis US Men's 11 years' dataset. The dataset contained rows with missing values and were removed completely. Building up on the previous analysis for Static visualization and coming up with a novel metric of Aggressive Margin and Relative error as follows:

Points Won by Forcing error = player's total points - (player's Winner shots + opponent's error)

Aggressive points = player's total points + opponent's forced error - player's unforced error

Aggressive Margin = (Aggressive points of player - Aggressive points of Opponent)

I looked at the Error data and also **calculating a player's Aggressive points based on their Total Points, Winners, Unforced errors and Forced errors**. Along with my analysis of how aggressive play relates with the errors made by the opponent. Checking other attributes which also relate with a player's aggressive play I also looked at the player's Fast Serve for each game. I found a **continuing trend** that holds true for **80%** of the players and hence, I select the **top five** players for the chart.

Interestingly, on sorting the calculated aggressive margin, I found that the opponents were making relatively more errors. This means "*Aggressive margin is **directly proportional** to the errors made by the opponent and also most of the time the Fast Serve Speed had increased too for the player*".

Visualization Type decisions

The above 3 attributes would let us see the trend or pattern that I found in the data. To show the direct proportionality I made use of a **Scatter-Bubble Chart**. On the X-axis is the Aggressive margin and on the Y-axis is the Error which is "Relatively more errors made by the opponent" and radius of the bubble/dot is the Fastest Serve speed in a game.

To supplement this and to understand more factors that contributed to the game, I make use of a simple **Radar or Spider chart** that also displays the player's and his opponent's Points won by forced errors, Aces, First Point won, First Serves and Nets in a particular game.

Interactions in the Charts

The most important interaction is the one where you hover over the bubble and the bubbles for a particular get highlighted while for the other players it fades to an extent. This allows to see a clearer trend for each individual player. This hover action also displays a Tool-tip which gives information about that point and simultaneously also changes the views in the radar chart which makes it possible to easily compare the stats for the player and his opponent. Another interaction is the ability to click on the legend to enable/disable a particular player's plot. This can be used to focus only on some combination of players' performance.

Choice of Colors

The colors used for the scatter plot are quite vivid. This helps to easily represent each players' plot with a very **different and unique color** in the chart. The **legend** of the chart easily helps to identify a player corresponding to a particular color. The color for the radar chart use a shade of Red and Green which goes by convention of Red being a loser and blue being a winner.

Hosted Online (Use Mozilla Firefox): <https://kashyapbhansali.github.io/TennisViz/>