

Srinivas Vallabhaneni

CSE 578

The data provided by Dr. Sharon helped me analyze the statistical part of the game of tennis and understand how each feature is significant and affects the outcome of games. In my analysis of this dataset I tried to prove that if a player does not loose points in Australian open games, he would be statistically in a better position to win the game. I try to prove this seemingly subjective assumption by representing this using a derived feature called *dynamic play score* which considers First point won, second point won, return points and Ace points. The score is calculated as follows:

Dynamic play Score= First points won+ Second points won+ return points+ ace points

Notes: 1) All the features used above are the key attributes, calculated using PCA for better data classification.

2) Uses linear addition for better representation of the pie chart.

Data Analysis and Story:

To get meaningful insights from the data provided, I segregated the data into winner data and loser data and the separately checked for missing values. I then supplemented the missing values with the average value of the feature. Then both the winners and loser data is appended vertically. So now we normalize the data to a standard scale and then subjected to Principle Component Analysis (PCA) which resulted in obtaining key attributes for data classification. These key attributes were used to formulate the Dynamic Score. The data was then ordered based on the Dynamic score to get the top 10 players of Australian opens dataset.

Then using this dataset, a bar chart is plotted to represent the top 10 players against their dynamic score over all the matches. Additionally, a pie chart is created beside the bar chart which represents the distribution of the dynamic score (scores of the key attributes obtained using PCA).

Some additional statistics of the Australian open series are represented on the page to represent some information about the series.

Interaction:

When cursor is hovered over the bars then the values of the pie chart changes according to the scores of that player. And, the image of the player with some statistics are represented below the bar chart to make the visualization more informative. These images and the player statistics are removed when the cursor is moved away.

In the same way, when cursor is hovered over the pie chart, then the bar chart represents the player and the score of the feature that is hovered over the pie chart.

Rationales of my visualizations:

My visualizations involve a bar chart and a pie chart. I have chosen Bar chart to compare the statistics of the top 10 players, because bar chart represents a linear comparison between different classes in a way that makes it easy to differentiate between the better and the best. I have pie chart to represent the breakup of the dynamic score of each player because dynamic score is computed using linear addition of 4 different players of the Australian open statistics. Hence pie chart would be the idea visualization candidate to represent something of this sort. I have used tennis ball with Australian open on it because it is the best way to describe about the tennis data and it's about the Australian opens. In the heading I have used Australian open logo as this visualization represents about the Australian open data.

Moving on to the choice of my color usage, I have chosen steel blue to display the heading and the bars as the color of the Australian open logo is steel blue and its good for representation purposes. I have used 4 completely diverse and attractive colors for the pie chart to make it more eye catching. I have used cream color as the background for better representation of the charts.

Note: File can be viewed in any browser, but for better visualization Chrome is preferred.