Data Visualization

Introduction:
The provided data set has statistics about Australian Open Tennis for 11 years i.e. from 2004 to 2014 (Men Only). Since the data provided has lot of information, it requires thorough analysis of the data. For analyzing the dataset, I have used Informatica's Rev tool and Rstudio. For the visualization, I have used JavaScript and D3 library. Text Editor being used is Sublime.

Story:
In my story, I am visualizing the data in the ascending of the year. A common man can easily find the winner of a year and his top 3 service/match variables which have contributed for the win. The more information about the matches can be found by hovering the mouse over the donut charts. Since, the pattern of the service/match variable of winners remain consistent throughout all the matches in a year, I am displaying the values for only quarters, semi and final for both the winner and the opponent and, I have not overwhelmed the visualization with lot of data.

As soon as, user opens the visualization, by default he can see the details of year 2004. An option to choose the required year has been provided. Let me start the story with 2004. The winner of year 2004 is Federer and his top service/match variables are ace, return and first point so 3 donut charts are displayed with these variable names. In the chart, green represents the statistics for the winner and red represents the statistics for the opponents. Winner will always have good number for the variables compared to opponents, so we could see that green occupies the major portion of the donut. There is an exception for error attribute, it must occupy low space to be a good service/match variable. Thus, a lot of information could be captured within few seconds. If user wants to discover why these service variables have been selected, then he can hover the mouse over the donut chart which gives information about value of the variables, the level and the player.

Design:
Green color and it shades have been used to represent the statistics of the winner in quarter, semi and final. Red color and it shades have been used to represent the statistics of the opponent in quarter, semi and final. This is because they are the generic notation and can be easily understood by the common audience. I have used donut charts for visualization of my story as it is efficient than others to display the information at the center space. Also, it is easy to grasp the comparison.

Justification:
The represented interactive visualization helps to get brief view about the winner and his top service/match variables for each year. Also, if someone is interested in more information then he/she can fetch more information by hovering the mouse over the donut chart. The information has been spread across the several pages without being overwhelmed i.e. it is discovered in layers (without loss of any meaning). The comparison has been made easier using donut charts. For any winner, considered top 3 service/match variables have been consistent through out all the matches so top 3 matches have been considered to represent the pattern in the visualization.