Assignment #2: Interactive Visualization Design

"Presenting interactive statistical visualization of Australian Open from 2004 to 2014"

Data Preparation – For my visualization, I had to convert the csv data into a hierarchical data. I used CSV to JSON converter available online and then ran a python script to build a hierarchy in my JSON data files. I used 5 variables namely, firstPoint%, secPoint%, Net%, Break%, return%. Continuing further, I calculated Ace probability, Winners Ratio and Error Rate by dividing aces with winners, winners with total points and errors with total points respectively. Finally, since few match details of initial rounds were missing, and I couldn't have uneven distribution of rounds in my visualization, I have used fourth round and above match details only!

Story – There are total 165 matches available. Upon exploring the data, I found out that winners usually have higher firstPoint%, secPoint% and return% than their opponents. Also, players with higher ace probability didn't necessarily win. On the contrary, they generally ended up having higher error rate (that's interesting)! I have designed an interactive visualization to understand how different matches across 10 years of Australian Open panned out and how these trends are visible by producing radar chart and bar chart comparison of above-mentioned variables.

Design – I wanted to implement a different way of selecting a match, hence I have used a sunburst chart for selection of matches. The inner most circle denotes final round, the second inner most circle denotes semi-finals and so on so forth. The center circle displays details regarding the arc(match) over which the mouse hovers. I have used radio buttons to change the year. On selecting a match, the match gets highlighted on the sunburst, with bar chart and radar chart popping up below displaying the statistical comparison for the match. The radar chart comparing 5 variables is interactive with changing opacity and tooltip display. The bar chart comparing derived variables also has tooltip enabled. The legend, chart title and axis names are displayed for clarity. The page clears its content on year change while as default we display 2004 Aus Open stats.

Color – I have used 64 colors to represent different players. I tried using categorical20b, but it has limited usage for a dataset like this. I have represented the sunburst arcs with winner's colors. And the same color combination is used in bar chart and radar chart enabling consistency and clarity for the viewer. The color combination dims on selection and mouseover and enables aesthetically pleasing visualization. The background color is kept as white and text color is black for contrast.

Design Rationale – The story can be clearly understood by looking at the radar chart. The radar chart is widely used for comparison purposes and in my story, I'm comparing the stats of the two players involved in a match. *We can easily see a trend developing in the radar chart – "Winners cover larger polygon area compared to their opponents".* Thus, our hypothesis that players with higher serve % and

return % stands validated. The bar chart is another comparison that helps us validate our initial analysis – <u>We can spot that *"Players with higher ace percentage have higher error rate".*</u> I was able to fit all the rounds starting from fourth, quarter finals, semi finals and concluding with finals. The sunburst chart is efficient in selection and interaction, helping to display 15 match stats for each year by just selecting different arcs.