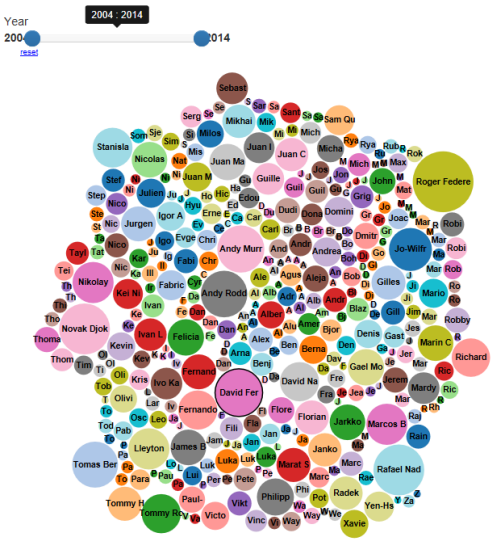
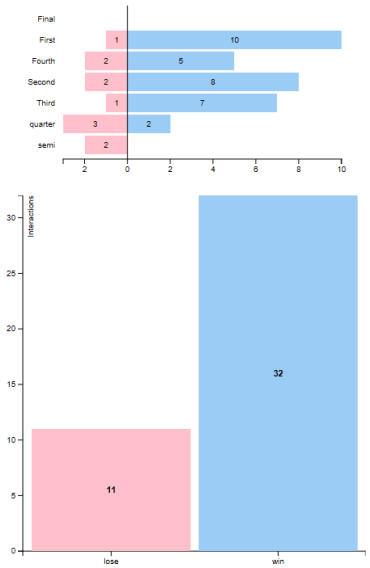
**1213217718**

**Sagar Patni**




**Main Idea/ Story:**

The idea is to explore the 10 year AUS Open Matches data using data visualization technique.

The plot shows an interactive exploratory data visualization which helps users to identify relationship between players and statistics of the player w.r.t. each round.

**Interactions:**

User can select the year range from the slider and the application will filter the data for respective years. User select a single player whose statistics to be explored by using the bubble chart.

After selecting the filter, the bar graphs shows the number of games won/ lose by user and the second bar graphs shows number of games won/lose by the rounds.

**Data Generation:**

The csv file data is converted to json data using python pandas code attached in file data_cleaning.py, The data schema is as shown below,


```
{
   "year":2004,
   "player":"Vadim Kutsenko",
   "round":"First",
   "result":"lose",
   "interactions":-1,
   "key":"Vadim Kutsenko",
   "ace":2.0,
   "double":5.0,
   "fastServe":217.0,
   "avgFirstServe":168.0,
   "avgSecServe":5.0,
   "error":38.0,
   "positiveInteractions":1
}
```


The above object shows the data of first round for user “Vadim Kut” for year 2004. The attributes ace, double, fastServe, avgFirstServe, avgSecServe, error are averaged for the round.

The file has multiple json objects for each player w.r.t. different rounds and each object contains averaged data for the respective rounds. In this application only the data about number of games won/ lose is shown but other json data in the objects can similarly shown using bar graphs.

For example avgFirstServe for games won/ lose in different rounds.

The data is filter on application using crossfilter.js to show statistics of the player.

**Design Rationale:**

The bubble charts are sized w.r.t. players who has played most games which makes easier for user to check statistics of popular player and on hover it shows player name on tooltip. The selections are highlighted which aids user to find relationship among data.

The colors of the bar graph are chosen such that it clearly shows the numbers of games won and lose.