Visualization Project Proposal

Basic Info

Project Name: Data Visualization of English Premier League

Group Member 1:

Name: CHEN YANG   email: chen.yang@utah.edu   uID: u0738066

Group Member 2:

Name: Hao Sha   email: u1078499@utah.edu   uID: u1078499

Link to Github repository: https://github.com/lucifer2012/dataviscourse-pr-premier-league

Background and Motivation

The Premier League is an English professional league for men's association football clubs. At the top of the English football league system, it is the country's primary football competition. Contested by 20 clubs, it operates on a system of promotion and relegation with the English Football League (EFL; known as "The Football League" before 2016–17). Welsh clubs that compete in the English football league system can also qualify. According to statistics, it's currently the most watched and welcomed soccer league in the world, which is broadcast in 212 territories to over 4.7 billion audience. In the past season, its average match attendance exceeds 36,000. With 47 clubs having competed in the league since 1992, there are only six teams are able to win the championship, including Manchester United, Chelsea, Arsenal, Manchester City, Blackburn Rovers and Leicester City[reference].

Having supported Manchester United for over 15 years, I am and still will be a fan for England Premier League. For each season, one of the biggest and most interesting question is which team is going to win the championship. Does the champion teams usually score the most goals or do they win the title because they have the best defence? For the teams downgraded to a lower rank, is it because they have the worst defence or because they could not find a way to score more goals? During each seasons, there are also many "big" games, like the derby between Manchester United and Arsenal. Are we able to predict the result more or less based on their performances against each other in the past years? These are all the keen questions that bother soccer fans. Thanks to visualization, now we are able to extract and display some key features of each team to explore why they can/can't win championships.

Project Objectives

There is an old saying that "good attack would help you win the audience, but good defense would help you win the champion". So our first two objectives are to testify if this is true.

i. The first goal is to explore the relationship between the defense and final ranking of the season. So the idea is to plot those two rankings of the season, and then to decide if they are related.

ii. The second goal is to decide if there exists a strong relationship between the goals that a team makes and its final ranking.

iii. Another great application of this visualization is to make a simple prediction of a specific game based on the rival history of the corresponding teams. In the world of football, there used to be a strong pattern that people could follow when it comes to prediction of a game. For example, Aston Villa could only defeated Manchester United once since 1999. Therefore, we are going to display the game results of a team pair for six seasons to help with predictions for the new games.

Data Source

Our data would be a relatively complete dataset of each team for 6 seasons. For each team, there would be pretty detailed information of its individual games, including half time results, full time results, corners and home/away team shots on target and so forth. The data of each season would be stored in a independent csv file, with the names of columns being the abbreviation of specific results statistics.

  i. Full explaination of abbreviations in column names

     http://www.football-data.co.uk/notes.txt

 ii. 2015-2016 Season

     http://www.football-data.co.uk/mmz4281/1516/E0.csv

iii. 2014-2015 Season

     http://www.football-data.co.uk/mmz4281/1415/E0.csv

 iv. 2013-2014 Season

     http://www.football-data.co.uk/mmz4281/1314/E0.csv

  v. 2012-2013 Season

     http://www.football-data.co.uk/mmz4281/1213/E0.csv

 vi. 2011-2012 Season

     http://www.football-data.co.uk/mmz4281/1112/E0.csv

vii. 2010-2011 Season

     http://www.football-data.co.uk/mmz4281/1011/E0.csv

Data Processing

Due to the original data was used for betting, it has some information that we do not need to include in our visualization, such as odds and rates. We expect to clean those unnecessary data away during our visualization. Besides, we probably need to extract some useful information of single teams to make a summary of each team during each season. The column names come in abbreviations in the original dataset, so we might need to change column names if necessary.

Visualization Design

The design is divided into two parts, I and II. They would be implemently in order according to the project schedule.

  i. Design I:

     1. The first element is the team list/chart. The point is to give the audience the chance to investigate the performance of a team for six season. The second point is to give the audience the chance to select and then compare the two teams. (part I)

     2. The second element would be two scatter plots. The two plots are intended to explore whether attack or defence is more relevant to final ranking. The X-axis would be the final ranking of each team, the Y-axis would be the ranking of defence/attack in that season. There would be a button to allow audience to select a specific season. The button would make changes to the team list/chart as well(part I). The team chosen in part I will be highlighted on the graph(Fig 1).

3. The third element would incorporate a result bar and two pie charts. When the performance of a single team is examined, the result bar would incorporate the data of six seasons that this team attended. Slide on the result bar would give you a summary of the team's performance over this period, which should be displayed in the two pie charts. The first pie chart would show the win, draw and lose rate over that period, while the second pie chart would show the goals made and conceded during those matches. When mouse hovers over the result bar, it should display the match and result of that specific game with tooltips. When mouse hovers over the pie chart, it should display the detailed information, i.e, the exact numbers. (part II)

ii. Design II:

1. The first element is the team list/chart. The point is to give the audience the chance to investigate the performance of a team for six season. The second point is to give the audience the chance to select and then compare the two teams. (part I)

2. The second element would be a bar chart, which could display a 3 bars for a single year, representing attack performance, defend performance, and final rank, respectively. When choosing another team in the team list/chart, the two teams' data will be displayed side by side. However, this design cannot be used to compare all team's data at the same time, which cannot fulfill our intentions (Fig 2).(part I)

3. The third part is same as design I. (part II)

iii. Design III:

1. The first element is same as Design I. (part I)

2. The second element is same as Design I. (part I)

3. The third part is a radar graph, which will show wins, winning during half-time, attack, defence, and come-ups. In this radar graph, we can compare two teams's data during a chosen series of games. After some discussion, we found that radar graph is sometimes unclear when data are close. Also, those attributes are not parrallel so its not best choice to use radar graph here (Fig 3). (part II)

Conclusion

We choose design I. That's because in design I, it's easy to explore if there is liner relationship between the ranking and defense/attack. The problem with disign II is it might not give the audience a straghtforward answer when they are trying to dig up the importance of attack/defense to final rankings. And then, in design III, although the radar chart looks pretty cool and is used in many soccer games to access the capacities of players, it might not be a good application here. That's because the due to the scale problem, the changes in the radar chart might not be that obvious to the audience. So generally speaking, we think design I is better.

Must-have Features

i. There would be a drop-down menu to allow audience to selected different seasons. The data in team chart/list and scatter plots would change accordingly.

ii. In team chart/list, when a team or two teams are selected, the selected teams would be highlighted. Meanwhile, the data in result bar as well as the pie charts would change as different teams are selected.

iii. When the mouse hovers on the points in the scatter plot, it should dispay exact final/defense/attack rankings with tooltip.

iv. When mouse slides on the result bars and select matches over different periods, it should display the statitics of these games and make changes to the pie charts.

v. When the mouse hovers on different sections in the scatter plot, it should dispay exact values with tooltip.

Optimal Features

i. Replace the team list with the badges of each team in blocks. When the mouse clicks on a team, the specific badge/block would enlarge to highlight selection.

ii. Add the the map of England to our webpage so as to highlight the postions of the teams. Click on a team in the team list should locate its stadium in the map. In default, it should show the stadium of the champion in that season once the data of a season is loaded.

Project Schedule

i. 10/24/16—10/30/16: Proposal submitted. By the end of the week, data collection and cleaning should be mostly done.

ii. 10/30/16—11/06/16: Done with framework of the html and javascript files. If data cleaning is not done last week, then finish up.

iii. 11/7/16—11/13/16: Finish up with design part I.

iv. 11/13/16—11/19/16: Finish up with design part II.

v. 11/19/16—11/25/16: Debug and wrap up. If there is more time, we would try the optional features.
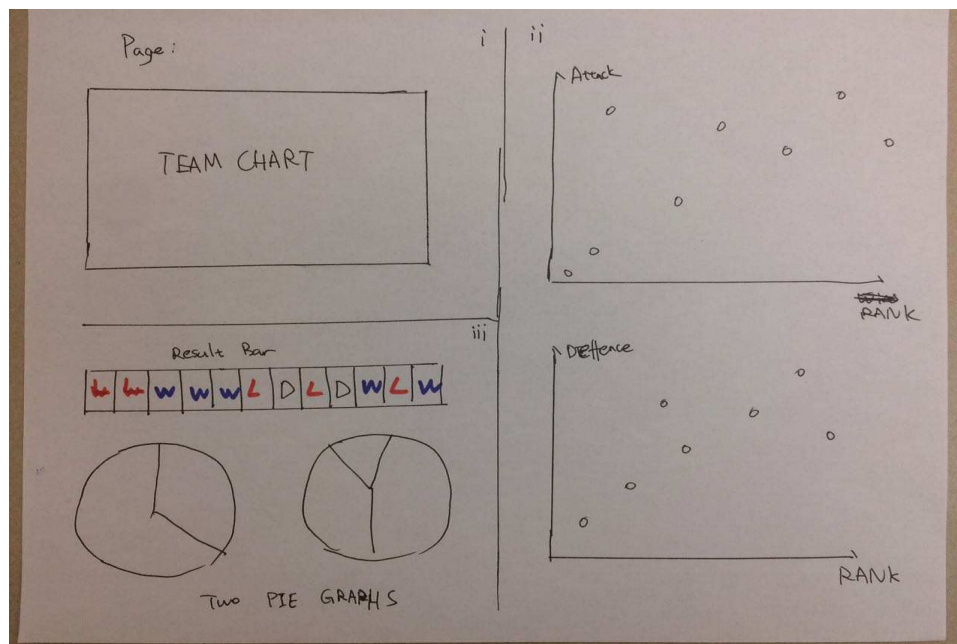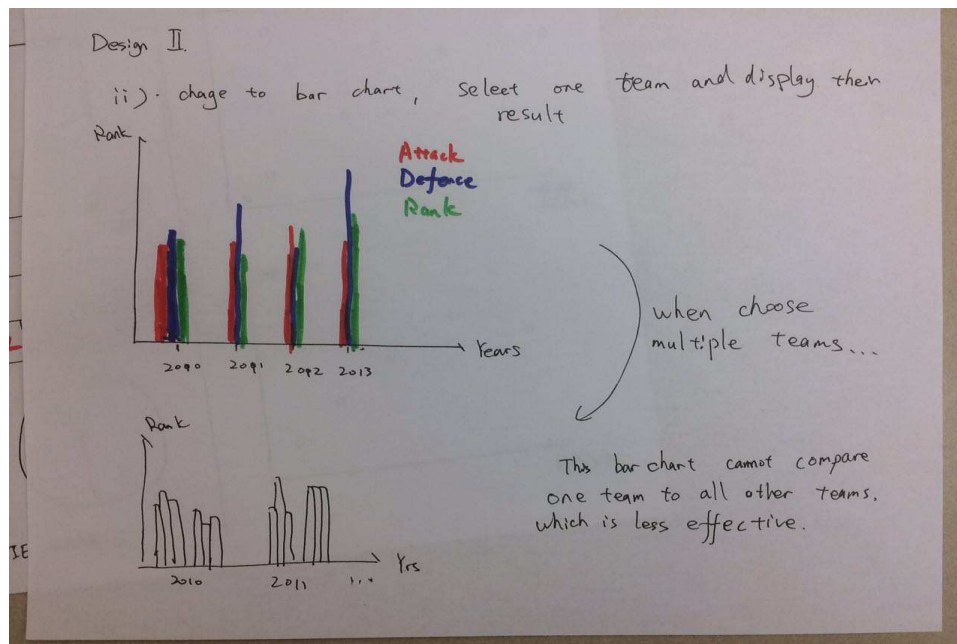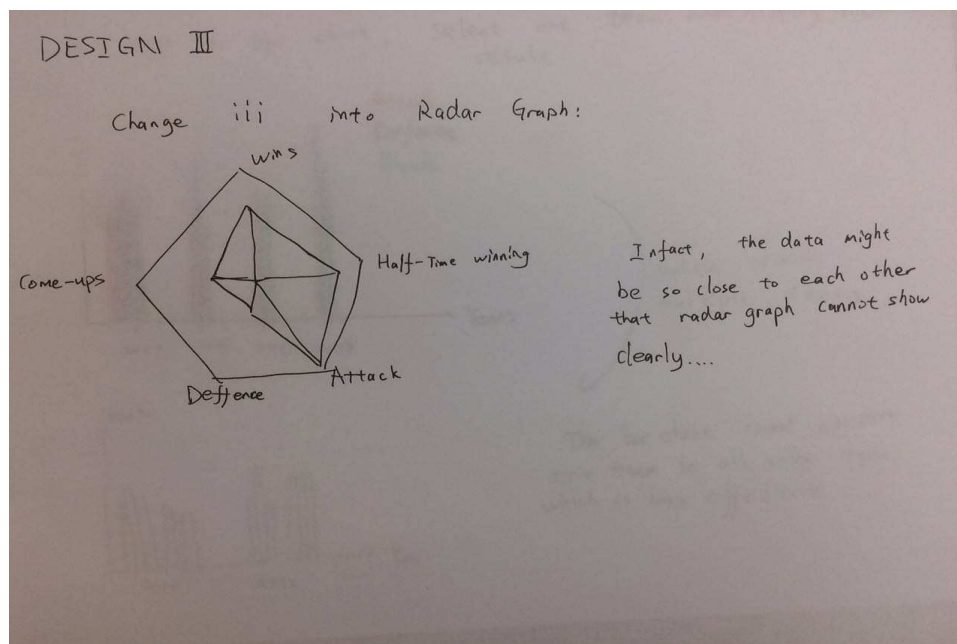
Appendix

Figure 1.

Design II.

ii). chage to bar chart, select one team and display their
result

Attack
Defence
Rank

when choose
multiple teams...

This bar chart cannot compare
one team to all other teams.
which is less effective.

Figure 2.

DESIGN III

Change iii into Radar Graph:

Infact, the data might
be so close to each other
that radar graph cannot show
clearly....

Figure 3.