Process Book

Project Title: Data behind the champions

Team name: **[C3]** VISBnB

Team members:

Chang, Chi-Hsuan - chc155@mail.harvard.edu

Xiao, Zhuyun - xiz309@mail.harvard.edu

Hua, Ji - jih723@mail.harvard.edu

Gui, Yufei - yug689@mail.harvard.edu

Background and Motivation

Ji is a “big” Golden State Warriors fan in our group.The warriors are the Champions of NBA in 2014-15 season. Some professional commenters think the warriors are leading the League to a new stage, for their innovative strategy of making good use of three-point shots. They pay less attention to mid-range shots (8-24 feet) and more focus on three-point shots and under-basket shots. The core player of Golden State Warriors, Stephen Curry, is thought to be the talented representative of three-point shooter. To see the difference of attack strategy between the warriors and other teams, we visualized the playoff matches the warriors played in last season. Also, we have made player comparison to see how good is Stephen Curry at three point shooting as well as his teammates.

Related Work

1. http://marquelaries.sportsblog.com/posts/6557294/nba-hot-topic--can-stephen-curry-keep-up- this-mvp-pace-.html This article discusses whether Stephen Curry can keep the MVP title, which we are also interested about.
2. http://fivethirtyeight.com/features/stephen-curry-is-the-revolution/ This article analyses why Stephen Curry is outstanding. It brings some ideas to us about how to evaluate a player’s performance.
3. http://www.nba.com/warriors/news/warriors-become-second-team-nba-history-win-65-games-consecutive-seasons Warriors become second team in NBA history to win 65 games in consecutive seasons, which inspires us to analyze the reasons for its success.
4. http://www.landofbasketball.com/player_comparison.html This two players comparison website is kind of similar to our 1v1 player comparison part, but we hope to add more fancy interactive visualizations.

Questions

We aim to explore:

1. **How did the Golden State Warriors win the 2014 - 2015 season?**
   The winning path of the Golden State Warriors in the 2014 - 2015 season. We will have a look at all the playoff matches that GSW played and visualize the shooting distribution as well as other stat items.

2. **How the performances of players affected GSW’s chance of winning the 2014 - 2015 NBA championship?**
   As we all know, a team is based on its players. GSW is the same. To make a more aggressive attacking strategy, the Warriors need to be more good at three points and “under-basket” shots.

This project can provide audience (we're targeting on NBA fans) information on how the Golden State Warriors won the last seasonal champion (Goal 1) and we may discuss whether it will become the next legend in NBA history. We will also discuss on how players' ability may take part in such success (Goal 2).

Data

Data source:

We use data from the NBA website: http://stats.nba.com/?ls=iref:nba:gnav.
NBA website includes intensive performance data (play type, pace/ possessions per game, team work (lineup), offensive efficiency, and shooting zone) for players and teams, which allows us for cross-sectional (performances between teams and Golden State Warriors) analysis in the 2014 - 2015 season.

Data collection method:

We use R as our main data collection and wrangling tool. In R, we use NBA official website's API to fetch the json file through url using R package ("jsonlite" and "httr"). The url format is like this: http://stats.nba.com/stats/{endpoint}/?{params}. This is the usual way for data collection from the NBA websites, and detail descriptions of Endpoint and Parameters can be found here:
https://github.com/seemethere/nba_py/wiki/stats.nba.com-Endpoint-Documentation

Also, some of our R scripts are inspired by established R package (ballr), which can be found here: https://github.com/toddwschneider/ballr

Data Processing:

For our project, we will analyze the play style of different teams (i.e. scoring strategy, pace, score area) as well as different core players in the 2014 - 2015 season. However, these data are stored in different webpages so we need to merge these data and produce tidy CSV file for later web page visualization. These steps are implemented by R packages dplyr and tidyr.

R scripts and output csv files:

R script	Descriptions	Output csv file
players_url.R	Fetch NBA players (479 players in total) that played in the 2014-2015 season.	**player_info.csv (Screenshot 1)** Variables: "person_id", "display_last_comma_first", "display_first_last", "Rosterstatus", "From_year", "To_year", "Playercode", "Team_id", "Team_city", "Team_name",

		"Team_abbreviation", "Team_code", "Games_played_flag", "name"
teamdashboard_data.R	Fetch 2014-15 seasonal performances of teams (overall and shot distance).	**overallteamdashboard_2014_15.csv (Screenshot 2)** Variables: "Group_set", "Group_value", "Fgm", **Field Goals Made** "Fga", **Field Goals Attempted** "Fg_pct", **Field Goals Percentage (fgm/fga)** "Fg3m", **3 points Field Goals made** "Fg3a", **3 points Field Goals Attempted** "Fg3_pct", **3 points Field Goals Percentage** "Efg_pct", **effective field goals** "Blka", **Blockshot average** "Pct_ast_2pm", **asissted 2 points made percentage** "pct_uast_2pm", "pct_ast_3pm", "pct_uast_3pm", **unassisted 2/3 points made percentage** "Pct_ast_fgm", **assisted field goals made percentage** "Pct_uast_fgm", **unassisted field goals made percentage** "Cfid", **court field id** "cfparams", **court field** parameters "team_id" **Shot5ftteamdashboard_2014_15.csv (Screenshot3)** Variables: "Group_set", "Group_value", "Fgm", "Fga", "Fg_pct", "Fg3m", "Fg3a", "Fg3_pct", "Efg_pct", "Blka", "Pct_ast_2pm", "pct_uast_2pm", "pct_ast_3pm", "pct_uast_3pm",

		"pct_ast_fgm", "Pct_uast_fgm", "cfid", "cfparams", "team_id"
shots.R	Fetch 2014-15 seasonal performances and also the shooting zone data of players.	**final_shot_data.csv (Screenshot 4)** "grid_type", "game_id", "game_event_id", "player_id", "player_name", "team_id", "team_name", "period", "minutes_remaining", "seconds_remaining", "event_type", "action_type", "shot_type", "shot_zone_basic", "shot_zone_area", "shot_zone_range" "shot_distance", "Loc_x", "loc_y", "shot_attempted_flag", "shot_made_flag", "shot_made_numeric", "shot_value" **league_average_2014_15.csv (Screenshot 5)** Variables: "grid_type", "shot_zone_basic", "shot_zone_area", "shot_zone_range", "fga", "fgm", "fg_pct", "shot_value"
lineup.R	Fetch Players' lineup data in 2014-2015 season.	**lineup_2014_15.csv (Screenshot 6)** Variables: "group_set", "player1", "player2", "player3", "player4", "player5", "group_name", "team_id", "team_abbreviation",

		"gp", "w", "l", "w_pct", "min", "off_rating", "def_rating", "net_rating", "ast_pct", "ast_to", "ast_ratio", "oreb_pct", "dreb_pct", "reb_pct", "tm_tov_pct", "efg_pct", "ts_pct", "pace", "pie", "group_quantity"
shottime_def.R	Fetch close defense distance and shot time of players in 2014 - 2015 season.	**shottime_2014_15.csv (Screenshot 7)** Variables: "player_id", "player_name", "player_last_team_id", "player_last_team_abbreviation", "age", "gp", "g", "fga_frequency", "fgm", "fga", "fg_pct", "efg_pct", "fg2a_frequency", "fg2m", "fg2a", "fg2_pct", "fg3a_frequency", "fg3m", "fg3a", "fg3_pct", "shotclockrange", "closedefdistrange"
playtype_datawrangling.R	Data wrangling of playtype data extracted directly from the websites for players and teams in 2014 - 2015 season.	**player_playtype.csv (Screenshot 8)** Variables: "Player" "team_id" "team_abbreviation" "person_id"

		"Team" "GP" "Poss" "Freq" "PPP" "PTS" "FGM" "FGA" "FG." "eFG." "FT_Freq" "TO_Freq" "SF_Freq" "And_One_Freq" "Score_Freq" "Percentile" "type" **team_playtype.csv** **(Screenshot 9)** Variables: "Team", "GP", "Poss", "Freq", "PPP", "PTS", "FGM", "FGA", "FG.", "eFG.", "FT_Freq", "TO_Freq", "SF_Freq", "And_One_Freq", "Score_Freq", "Percentile", "type", "team_id", "team_abbreviation"
radardata.R	Fetch traditional statistics of players in 2014 - 2015 season for both regular seasons and playoffs. Do wrangling for them. We select six variables to display, namely steals, blocks, personal fauls, rebounds, turnovers, assists. We compute scores in these six fields(in percentage) for each player using ranking among all players in these fields.	**newradardata.csv** **(Screenshot 10)** Variables: "player_id", "player_name", "team_id", "team_abbreviation", "Gp", "reb", "ast", "tov", "stl", "blk", "pf",

		"playtype"
lineup_wrangling.R	Data wrangling of lineup data. Get four most frequent lineups for each player.	**ranklineup.csv (Screenshot 11)** Variables: "player", "lineup", "sumgp", "player_name", "lineup_name",

Exploratory Data Analysis

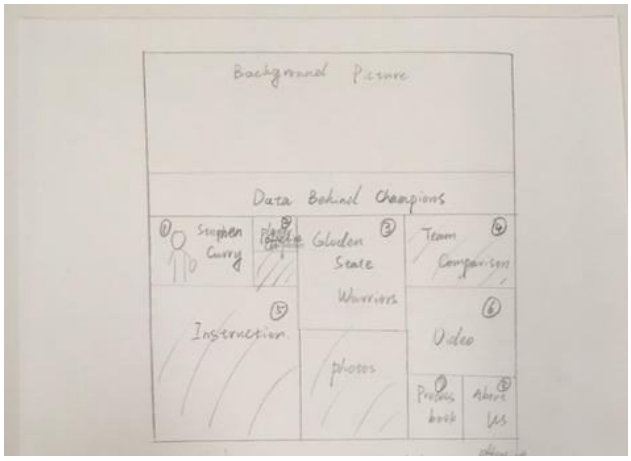
We used the data scraped from NBA official website. By doing data wrangling, we have become very familiar with the data. Basically, we have two different type of data: shooting data and boxscore data.

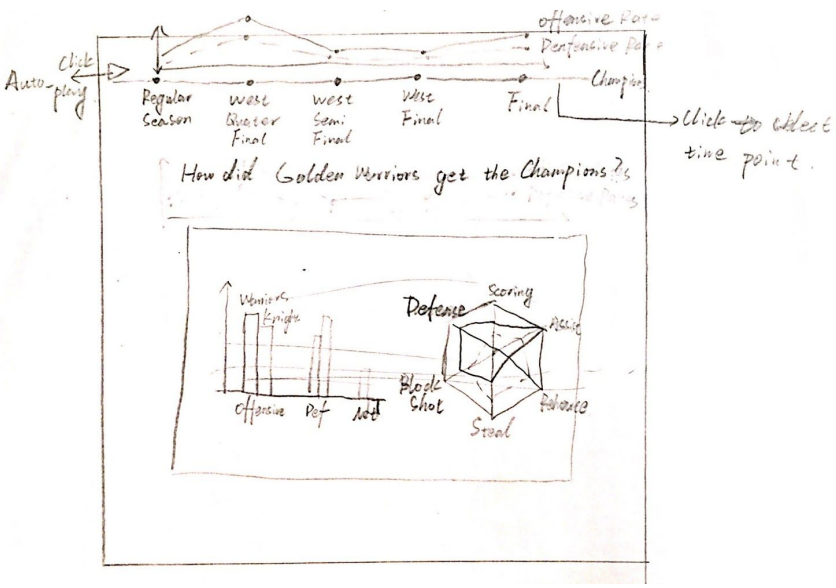
1. Shooting data. Shooting data contains all shooting information of each player (locations, miss/made, shooting type). Since it contains location information, we think it is efficient to use a court heatmap to visualize locations and use color to encode whether the shots were missed or made. Also, the size of each point encodes the efficiency of a player's shooting performance in that location. A donut chart is also implemented to visualize the general shot distance (close shots, mid-range shots, 3 points shots) since it shows the compositions, and a stacked bar chart is used to visualize different shooting types.

2. boxscore data, which is used to analyze team/player's performance in other domains rather than scoring ability. This is visualized with radar chart and bar chart.

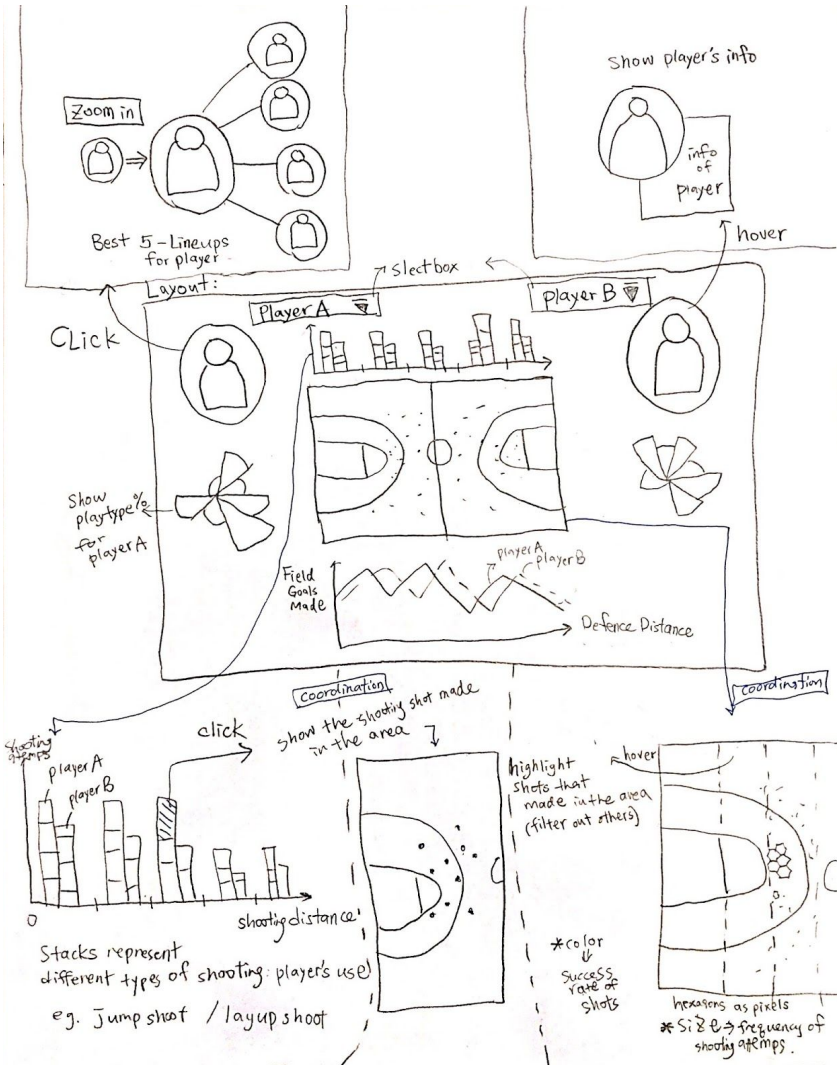
Design Revolution:

1. **Original Design**
   We originally wanted to create a page for team and player separately and to have the following treemap main page for user to select different sections they're interested in.





In the page for team, we wanted to have a timeline representing the performance of GSW as compared to its opponent in each stage of the playoff. We also wanted to use a barchart and a radar chart to show the performance/playtype between the two teams.



In the page for players, we wanted to include 1) a full court to show the shooting location, and the corresponding performance at that location (shooting efficiency and frequency), of the two selected players, 2) a stack bar chart on top of the court to show statistics of shooting type and shooting distance of the two players, 3) two radar charts to show playtype of the two players, and 4) best lineup of the two players.

2. **Project Re-design**

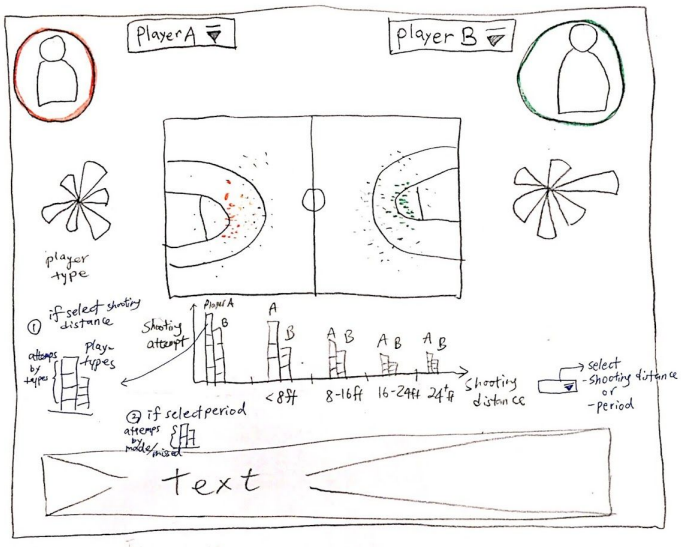
Peer feedback: After we introduced our visualizations, the other team mentioned several feelings about our origin design.

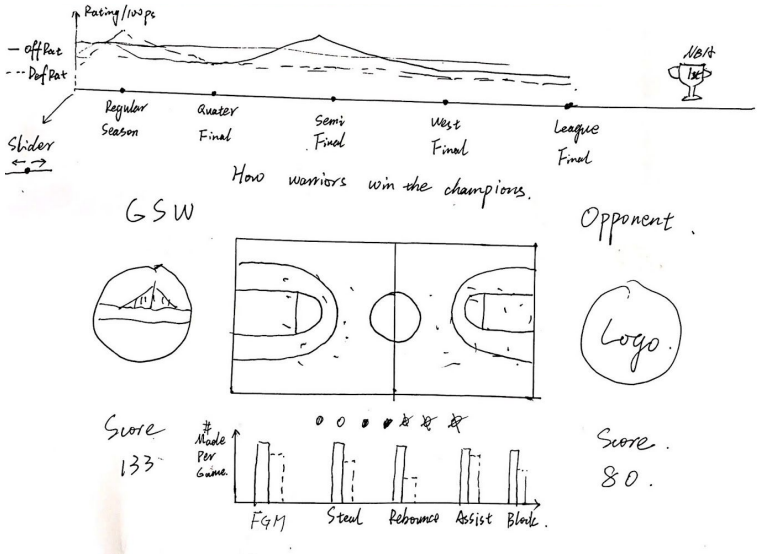
Comment received [1]: They were confused about the shooting points of the two players on the basketball court visualization (our design include a full court, in which half of it belongs to player A and the other half belong to player B while they can't tell which half belongs to which player). Therefore, we decide to color these shooting points in two different color scales for the two players, which could make the comparison clearer.

Comment received [2]: They were confused about the stacked bar chart we put above the basketball court which represents the shooting attempts (stack: shot types) by both players at different shooting distance. They felt that the x-axis (shooting distance) of the stacked bar chart was only intuitive for player A (since the origin of the x-axis starts from the left). Therefore, we decide to change the layout of our original web page and move the stacked bar chart below the court to prevent confusion here.

Revision after group discussions:

Identification of project weakness [1]: Too many plots on the same page. We identified that our webpage for the basketball court was too busy (1 court plot, 1 play type chart, 1 stacked chart, 1 line chart). Therefore, we decide to discard the line chart and combine some of the information to the stacked chart and let the users to select which information (shooting distance or period) they intend to see.





Identification of project weakness [2]: There are too many types of designs included in the original design in both main page and subsequent pages. We decided to remove some of the design and replace with simpler charts that can be understood more intuitively.

Implementation

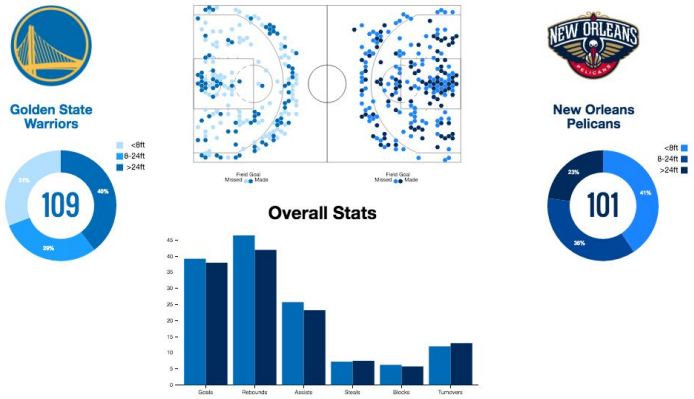
1. **html:**
   Index.html basketball_court.js, hexbin.js, hexagon_team.js, RadarChart.js, script.js, traditional_stat_barchart.js

   The first step of our project protocol is to create a website frame that contains all the storytelling idea. At the very early stage, we wanted to use the Windows 8-style (looks like a treemap) to arrange our main page. After several discussions, we decided to discard this kind of design since it may not be the best in terms of storytelling. We then adopted a more convenient (and hopefully, more intuitive) way to present this story -- a scroll page frame. We think this design of website frame allows user to scroll down the page part by part and see our story clearly. The main page of our website is like the one below.

After this main page, we begin to tell our first story.

The first story we want to present is the winning path of the Golden State

Warriors. We included all the playoff games of GSW in the 2014-15 season. The data we picked is from the statistics in playoffs. We firstly set up a timeline axis. On this timeline, there are 4 stages (playoffs). During each stages, there are the corresponding matches users can choose to see the game statistics in bar chart and shooting chart. By sliding the stage button and by clicking on the match buttons, the donut chart, barchart and the shooting chart, will change simultaneously.

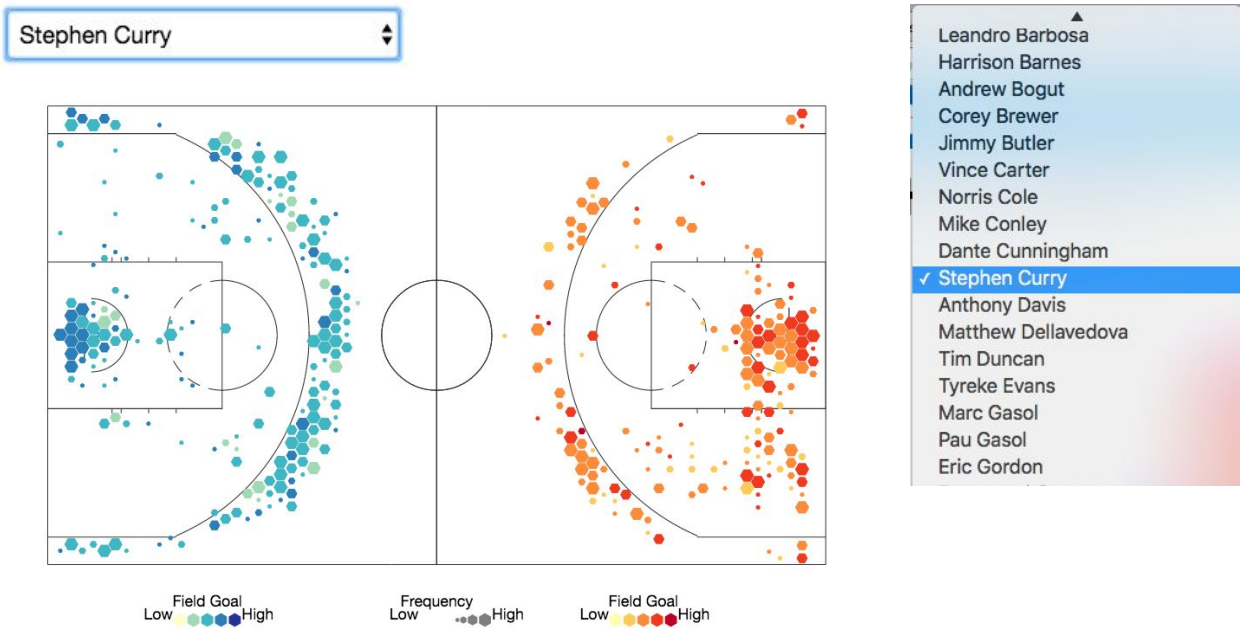


Then we move on to our second story. In the player section, we explore the player-level performances that could have contributed to the success of GSW. The stacked bar chart is used to compare two player's shot type in specific shooting distance. We will also allow users to select a player to compare the statistics with the default player we choose, which is Stephen Curry. Then we will let users to select the shooting zones then at the same the basketball court will filter only the selected shooting zone statistics and display only the selected points. As for the radar chart, this visualization coordinates with the player selection. Once users selects the player, the radar chart will be updated accordingly.

2. **Js:**
   a) **basketball_court.js:** For implementation of the team shooting chart, we searched previous similar work by Jowanza Joseph (http://www.jowanza.com/post/126964236829/nba-shot-charts-with-nodejs-and-d3js). This js script requires d3.chart to make function team_BasketballShotChart() that is used to draw the court and also to draw the shots made by the players. The function can then be called every time we want to draw a half court. In this player-level shooting chart, the size of hexagons encodes the frequency of shooting at a specific location and the color of hexagons encodes the efficiency (depend on goals made/missed) of shooting at that specific location.

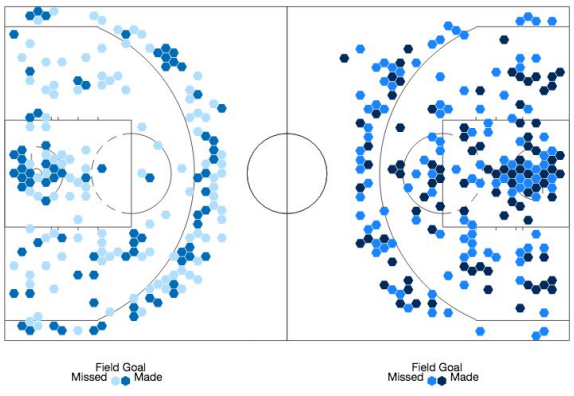
   b) hexagon.js: The challenging part of our project has always been our symmetric design for comparisons of two teams as well as two players. In hexagon.js, we use the BasketballShotChart() function to draw each of the half court of the players by including in the rotations of courts and

encoding different colors for the two players to distinguish them from each other. We also include two selection boxes (as shown below) that can be used to select the shooting charts of both players.



Since we included all allstar players in the 2014 - 15 season, as well as all players in the opponent teams GSW encountered in the playoffs (teams above), we did not choose different colors to represent the corresponding teams of different players.

c) **basketball_court_team.js:** Similar to basketball_court.js, this js file is also based on d3.chart to build shooting charts for team comparison. Unlike those in the player section, shooting charts here contain game-by-game comparison between the two teams. Therefore, there are limited shooting attempt records to compute the frequency of shooting at a specific location. As a result, we chose not to encode the frequency of shooting while decide to show every shooting attempts the team made, and use two colors to encode the result of each shot (missed in lighter color/ made in darker color).

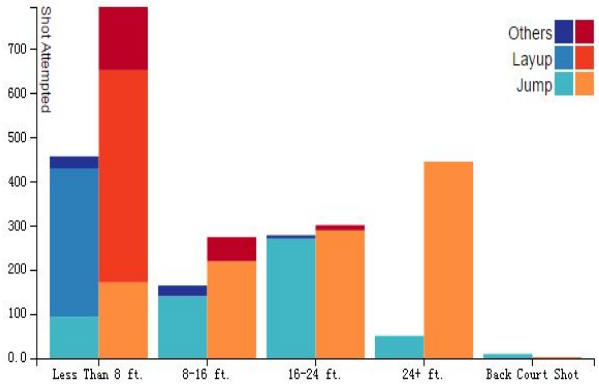


d) **hexagon_team.js:** In hexagon_team.js, we use the team_BasketballShotChart() function to draw each of the half court of the teams by including in the rotations of courts and encoding different colors (using their own team colors) for the two teams to distinguish them from each other. The shooting chart can coordinate with other charts by sliding the stage button and clicking on the match button.

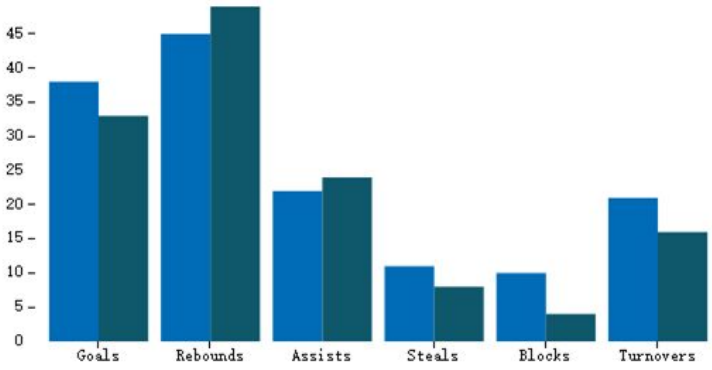
e) **RadarChart.js:** We found it is efficient to use radar chart to display the six traditional statistics for each player. For implementation of radar chart, we refer to the radar chart made by Nadieh Bremer (http://bl.ocks.org/nbremer/6506614) . In RadarChart.js, it implements the background circular segments, axises, legends, labels as well as the above 'net' like visualization. We also include tooltips showing descriptions of axises.

f) **script.js:** In script.js, we use the RadarChart.draw() function to draw radar chart for two players. We use two group of colors in two players' radar charts.

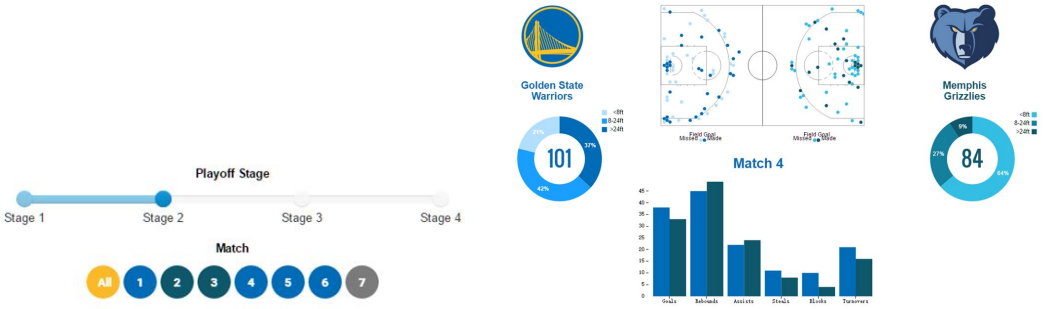
g) **stack_bar.js:** In stack_bar.js, we use data same as basketball.court.js. However, after loading the data, we do some wrangling -- we calculated shot attempted for both players in three shot types and five shot ranges. We use x-axis to present the shot range and colors to encode shot type. We also include legends in upper right corner. In order to coordinate with the basketball court chart and act as filtering, we make the labels of x-axis to be clickable and once it's clicked the stack bar chart will transition into bar chart of the chosen shot range. Besides, the color legends are also clickable and once it's clicked the stack bar chart will also transition into bar chart of the chosen shot type. After being clicked, a title will appear on the upper right corner showing the chosen label which could be clicked to return to the original stack bar chart.



h) **Barchart.js**: In this js file, we are showing the stats of other aspect of a game, such as rebounds, assists and so on. This will help people to have a better understanding of how the warriors win the champions. In this file, we implement a group bar chart to show the information mentioned above.



i) **main.js**: In this file, we deal with all coordinating stuffs such as stage select, match select, score showing, team changing as well as the drawing of donut charts. As for the donut charts, in this part, we cleaned and wrangled the data to the percentile form and try to translate it to represent the distribution of a team's successful attempts at different distances. We implemented these donut charts to be coordinate with user's selection of the stages and games. We use d3.js to create transition for the donut charts to be updated everytime user change the game or stage. The color of the donut chart was assigned according to the team's color. There is also a dynamic legend located at the side of the opponent donut chart. As long as user change the stage or games, the color of the legend will be changed accordingly. The data set of the donut chart shares with the basketball court visualization data set.



j) **Teammates.js**: In this js file, we are going to show the best teammates of the player.

k) Global.js: All global variables are initialized in this file.

l) Tutorials.js: This js file is used to create the tutorials showing how to use our website.

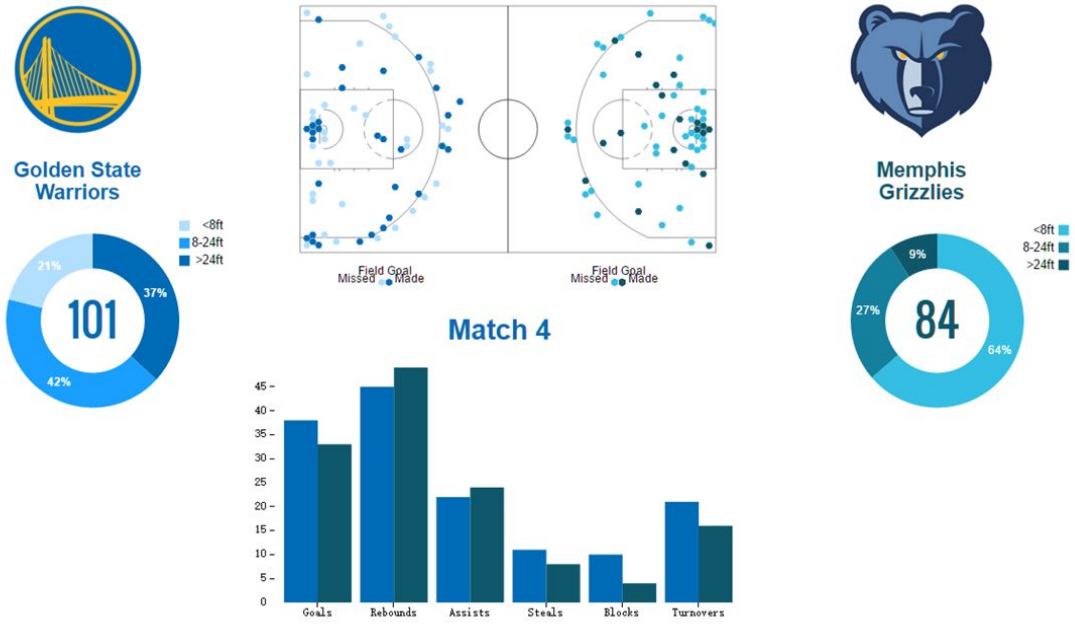
3. **Css:**
   style.css: This is the major stylesheet of our website.

4. **Data:**
   a) final_shot_data.csv: After data wrangling from R, we used this csv file to draw several charts (stacked bar charts and players' shooting charts) in the player section.
   b) team_shots.csv: After data wrangling from R, we used this csv file to draw the team's shooting chart, donut charts.
   c) ranklineup.csv: After data wrangling from R, we used this csv file to draw the lineup.
   d) newradardata.csv: After data wrangling from R, we used this csv file to draw the player's radar chart.
   e) gamelog_average.csv: After data wrangling from R, we used this csv file to draw the team stats bar chart.

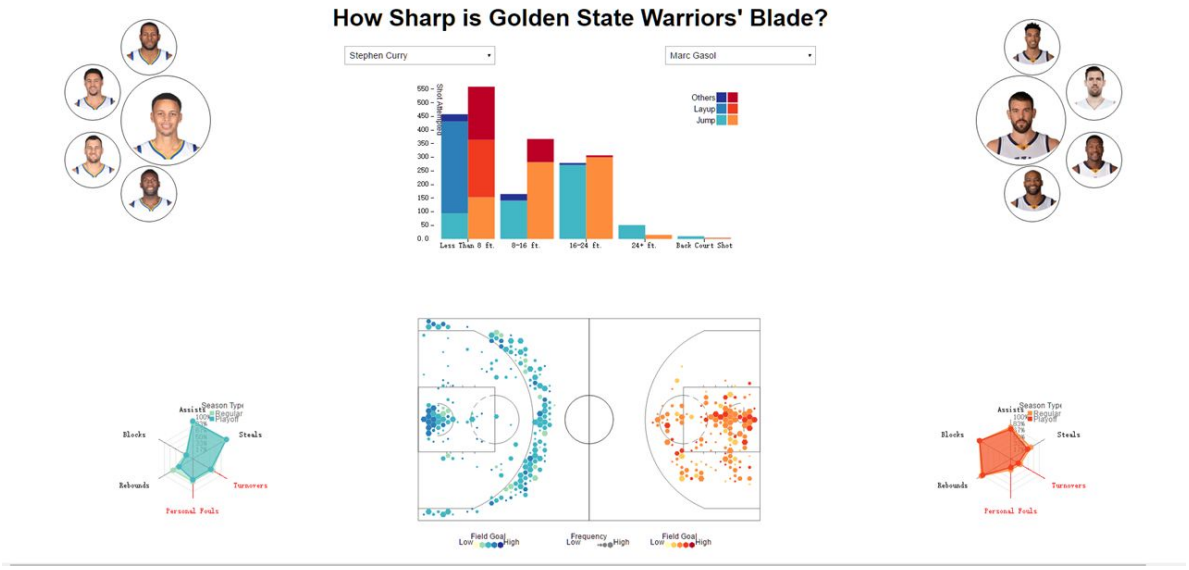
Evaluation:

By using our visualizations, we can answer the questions we proposed in the very beginning.

- **How did the Golden State Warriors win the 2014 - 2015 season?**
  Take this match as an example, we can see the distribution of both teams based on the shooting charts and donut charts. The warriors have more three points shots than the grizzlies, and the warriors made more goals than the grizzlies as well (looking at the bar chart). As we all know, three points shots are worth more than the two points shots, which made the attack of the warriors more efficient than the grizzlies. In one word, who win the game of three points shots win the game.



- **How the performances of players affected GSW's chance of winning the 2014 - 2015 NBA championship?**



Looking at this visualization, we can see the statistics of the core player of warriors, Stephen Curry and Grizzlies, Marc Gasol. First, by looking at the radar charts, we can see both players are doing well though Curry may not be good at blocking shots, but he is doing better in assists. Then using the stack bar chart and the shooting charts, we can see that Curry has a much wider shooting range than Gasol, whose shooting range is limited inside the three points line. So players define the attack strategy of both teams. Warriors more focus on three points and grizzlies more focus on short range shots.

Our visualizations work pretty well on these questions. However, positions of the players are also important. Marc Gasol is a center while Stephen Curry is a shooting guard. It is a bit unfair to compare these two players though they are both the key players of each team. So we can furtherly improve our visualization with such information. Moreover, though we considered successful rate of shots in the shooting chart, but that is a bit unclear. We can also transform those data into a more communicable form.