INFO 5100/3300 Project 2

Project Report

Jack Dauber (jd733)
Jude Javillo (jij8)
***Jieyu Chen* (*jc2562*)**
***Rong Xu* (rx25)**

Introduction

In this project, we created an interactive visualization of different NBA players from all 30 teams in the 2014 - 2018 competition seasons. In terms of visualizations, firstly, we developed a button system so that users can choose teams, competition seasons, and different players. Secondly, an NBA shot chart(basketball half-court map) was developed to show the position where a particular team/player throws three-point shots in one competition season. Then we compared a player's seasonal performance with overall team average and NBA league average performance that year using a bar chart. The parameters we considered were **3P%, 3P%_Team,** and **3P%_Average**, their definitions show below:

3P%: 3-Point Field Goal Percentage = 3-Point Field Attempts / Field Goal Attempts

3P%_Team: 3-Point Field Attempts / Field Goal Attempts of a specific team

3P%_Average: 3-Point Field Attempts / Field Goal Attempts of the entire NBA league

In terms of interactions in our design, firstly, relative players can be chosen when we select a specific team and year. Then, clicking on a team or a player will update the bar chart and court map.

To view our visualization, please first click the team, then you can change the corresponding year and name.

Data

Our data mainly consisted of three parts: team logo figure data, player information data, and shooting position data.

Image data:

We got our 30 NBA team logo figures from the website **https://logotyp.us/logo/nba.** We got our introduction background image from **https://www.neosportsinsiders.com/wp-content/uploads/2016/06/Warriors-vs-Cavaliers-Finals-2016.jpg**.

Player information data:

For player information data, in the case of the basketball analytics community, BasketballReference.com (**https://www.basketball-reference.com/**) is a goldmine of data, since it features all sorts of relevant in-game statistics useful for many basketball analytics projects. Particularly for NBA statistics, the site has a “Per Game” section where one can obtain most Per Game performance statistics for the entire population of more than 500 NBA players taking part in any given season. Data preprocessing is necessary.

We used the python program and imported pandas, requests and BeautifulSoup library which could return any player’s complete Per Game statistics, for any given season. We converted all the information data to seven databases named season years(eg. 2014-15.csv). Each database contained the names, ages, teams and other professional data(eg. 2P%, 3P%) of all players in that competition season. Our button system was based on four of these datasets.

For bar chart analysis, we further processed the data. We imported 3P%_Average data from **https://www.basketball-reference.com/leagues/NBA_stats_per_game.html,** combined

it with 3P% four years' data of each player. To achieve this, we used the VLOOKUP function to insert the data to the corresponding labels..

Shooting position data:

Our shooting position data was also imported from the API of https://www.basketball-reference.com/ using the python program. Overall, we collected about 30 of them so that each player's data was stored in one dataframe for one competition season. For the convenience of writing the code, we classified players based on their team for a given season and combined all players from a single team to create a single csv per team for each competition season (e.g. 2014_15/GSW.csv), so we got 120 of them in a total of 30 teams for four years. The variables we used were LOC_X, LOC_Y,PLAYER_NAME, and EVENT_TYPE.

Design Rationale

Static part:

Initially, We wanted to use a map of the United States to show different NBA teams, as we said in milestone 2, because different teams come from different cities and have different characteristics. However, in practice, we found that the map can not completely represent each team, especially on the issue of their influence. So finally, we turned to the option of a shot chart (basketball half-court map).

We drew the shot chart ourselves, using the function `drawShotChart()` in our code. This included a combination of lines, lanes, corners. We projected the players' shoot position in the actual competition proportionally to our shot chart using LOC_ X, LOC_Y data. The mark of this chart is the scatter point on it so that it could represent the different positions. For the channel, we used the color hue of the point. To be specific, red/blue for a hit/miss is a common metaphor in basketball. So when a player is making a lot of shots, they are "on fire", we chose red to represent making a shot. On the contrary, when a player is missing a lot of shots they are "going cold", so we chose blue for missing a shot.

Moreover, because of the huge amount of data we had, our data loss was also serious. This directly led us to make the decision to use a bar chart for data visualization although we were determined to use a line chart initially, because the lack of data will greatly reduce the expressiveness of the line chart. For our bar chart, we used competition season years to be X-axis and 3-Point Field Goal percentage to be the Y-axis. `d3.scalePoint()` was designed for the years and `d3.scaleLinear()` was used for the 3-Point Field Goal percentage data. In this bar chart, marks are rectangles and channels are the vertically aligned length and horizontally aligned position of rectangles. Moreover, we used different colors to represent 3P%, 3P%_Team and 3P%_Average, so the color hue is also the channel.

Interactive part:

We used a button system that contained three kinds of buttons to achieve our interaction purpose. When we choose the specific team, change of year and the player will update the shot chart and bar chart. To achieve this, a number of functions (eg. updateLogo, updateTeamchart) were set. In order to make our button discoverable, we add another three main buttons labeled "Please select the team/year/player". For possible problems, we also add a warning program, so if the user chooses a player first without choosing the team and the year, the sentence "Please select a team and season first" will appear.

Description and Story

Before the years the Golden State Warriors started winning (before 2014-2015 competition season) most NBA experts and analysts believed that going for a lot of 3 point shots was a bad idea, as the shot was too hard to hit, and that it was a good idea to take 2 point shots as they were easy to hit and are closer to the basket. Once the Warriors started winning, they proved that taking a lot of 3 point shots could be a recipe for success. They were so influential that a lot of teams were accused of copying the strategy laid out by the Warriors and taking a lot of 3 point shots. So we chose to investigate if this is true or not by visualizing the 3 points rate of all league's different teams and players from 2014 to 2018.

From the bar chart, we can see that with the growth of the year, the proportion of three-point attempts in the whole league has increased steadily (deep blue bars). Therefore, we believe that the Golden State Warriors' victory has a certain impact on the tactics, training and attention to the three-point ball of other teams in the NBA. For most teams, their 3-point field goal attempts is also increasing year by year (light blue bars), this is particularly evident in Memphis Grizzlies (MEM) and Milwaulkee Bucks (MIL): from 2014 to 2018, their 3-point field goal percentage increased from far lower than the average level of the league to the average level of the league. A few teams, such as Indiana Pacers (IND), still have data shocks, but this does not affect the overall trend.

At the same time, we have also witnessed the extraordinary skills of some players from the data, such as Stephen Curry, whose 3-point field goal percentage and attempts is far higher than the average level of the league and the average level of the team in any year.

Moreover, we can get other minor conclusions from our data visualization results, one is that the personnel mobility of NBA players is very large every year. Originally, we counted the data for seven years. From the results, only quite a few players can persist in playing the whole seven competition seasons. For the same team, the players in the team are a little different every year. This poses a certain difficulty to the processing of our database. We think this is reasonable, because a confrontational game like basketball is closely related to the age and physical state of the players. At the same time, the cultivation and trading of players is not only a means of making profits for NBA teams, but also an improvement of their ability.

Contribution

Jack Dauber

Jack contributed towards one idea in the brainstorming process and looked for datasets for it. He also peer-programmed the buttons and added more channels to the bar graph with Jude. He also worked on the design and layout of the webpage, making sure that we were following good design principles such as choosing colors that were colorblind friendly and having the user experience be intuitive. Overall this took about 15 hours throughout the course of the entire project (3 hours looking for data and brainstorming ideas, 6 hours peer programming, and 6 hours on design). Surprisingly, the hardest part was using a colorblind tool to find three distinct colors that looked different for every deficiency and looked visually appealing.

Jude Javillo

Jude did some research into basketball statistics and wrote a Python script to gather NBA data from 2014-2018 (3 hours). He added the introductory paragraph for the data visualization and the CSS styling to create a parallax effect for the background image (2 hours). He added JavaScript code for creating interactive buttons for filtering through team, season, and player data (4 hours). He added JavaScript code for visualizing the position of where shots were taken on a basketball court by a given team or player in a year (4 hours). The hardest part was fixing a bug where the buttons were only clickable

Jieyu Chen

Jieyu contributed two ideas in the brainstorming process and found datasets for them. One idea is about startups, the other one is about earthquakes (3 hours). In the basketball project, Jieyu implemented the bar chart showing the three-point shot percentage for the basketball players (8 hours). The hardest part is making the bar chart responsive to each player.

Rong Xu

Rong Xu contributed the Nobel Prize ideas in the brainstorming process. He found datasets and sketched the possible solution for it.(3 hours) In the basketball project, Rong preprossed the datasets for further use and analysis for the bar chart.(3 hours) Rong Xu also wrote the overall report.(4 hours)