CS 3300 Project 1: Static Data Visualization

08th March, 2016

Authors:
Nancy Ding
Chris Vrabel
Ankur Biswas

TA In-Charge:
Monali Patel

Title: Three Point Madness

Part 1: Overview of the Project

This project is a static visualization of data from the National Basketball Association (NBA). We discussed that there may be an upward trend in the number of three point shots that have been taken/scored in the recent history of the NBA. As such, we sought to visualize this trend along with some other 3 point shot information with the help of various graphing techniques:

1. A time series scatterplot of the number of shots taken/scored over many years of play in the NBA. We have 5 charts depicting play from the 1999-2000, 2003-04, 2007-08, 2011-12 and the latest 2015-16 seasons. The background for the graphs is the basketball court itself while varying shades of purple depict a 'heat map' of locations on the court from where shots have been taken/scored by the best players.
2. We have a line graph of three point field goals taken by the top players in recent NBA history. We sought to uncover if, as seasons progressed, the 3 point field goal percentage of the NBA's best players increased, decreased or remained the same.

3. We have a linear regression chart of the Three Point Attempts to the Offensive Rating of teams over seasons: to try and determine whether the offensive rating of a team correlates with the Three Point Shots.

Part 2: Description of the Data

The data for this project was sourced from a number of reliable websites. The shot charts are JSON files pulled from the NBA stats site, the line plot used data from ESPN and regression chart used data from Basketball-Reference[1]. Since we wanted to examine the consistent players in the NBA, we shortlisted the best performing players in every position by looking up their statistics from seasons and MVP rankings from ESPN. We then got their playerID's from their own NBA statistics page[2] (an

[1] http://www.basketball-reference.com/

[2] *Stats page for Chris Paul of the LA Clippers:* http://stats.nba.com/player/#!/101108/?p=chris-paul

example is in the footnote) – to make sure we got the data for the right players. The data for the regression analysis was taken from the miscellaneous statistics page on Basketball-Reference[3].

The variables used in this project are as follows:

S No.	Variable Name	Description
1	ORtgData	Offensive Rating, a statistic that measures offensive performance (the efficiency of scoring points)
2	ThreePAr	Three Point Attempt Rate: a percentage taking three points attempted over total shots attempted
3	Attempt	number of attempted three point shots
4	Perct	Percentage of three point shots made

The regular season of the NBA comprises 82 games, with teams typically averaging 80-90 field goal attempts per game. Trying to develop the heatmap for all these attempts would have resulted in too much data, and no discernable pattern would have been detected. In order to determine if the trend we were looking for exists, we plotted the data of some of the best players in the NBA for each season. In order to have maximum representation and the most accurate data, we took the top 5 players from each position: Point Guard (PG), Shooting Guard (SG), Center (C), Small Forward (SF), and Power Forward (PF). The unique player ID that the officially used by the NBA website to reference these players were used to reference them in the code as well.

Part 3: Mapping from Data to Visual Elements:

i. **General:** After the general CSS for the page was laid out, each of the Heat Maps were placed in their own elements. PlayerIDs were stored in arrays. The names of the players each ID represents are included as comments.

ii. **The Heat Maps:** The data for the Heat Mapping was scraped directly from the NBA website using the player's unique player ID. A successful AJAX call to the page returns an array of the JSON format. This array contains a huge list of information. The indices of the array elements were used to access data – as typically done in any language. The variables percentThrees, percentMid and percentPaint were then used to calculate the percentage of shots taken from each area of the court.
The court itself was created with an SVG path element and the 'boxes' which have the different levels of 'heat' were created using SVG rectangle elements. In order to show different levels of 'heat' the 'fill' attribute of the SVG rectangle was changed. A darker fill indicates more shots were taken from that rectangular area. The graphs were labelled appropriately using SVG text elements.

[3] http://www.basketball-reference.com/leagues/NBA_2013.html

Last, the drawHeatPlot function was used to divide the court into smaller regions and draw out the heat mapping for the court as whole for a season.

iii. **The Line Plot:** The more traditional line plot plots the 3 Point Field Goal Percentage for some of the best players in the history of the NBA as seasons progress. The players included are: Kawhi Leonard, JJ Reddick, Stephen Curry, Jared Dudley, Jerry Bayless, Klay Thompson, Kevin Durant and Lebron James. Each player was assigned his own data structure and the data associated with him was placed in it.
The plots were then created by adding points to the graph using method addPoints() and then addLine() was used to join the points together. The size of the circle at each join on the line indicates the number of shots taken per game by the particular player.

iv. **The Regression Analysis:** We wanted to see if there was a relationship between the 3 Point Shots Attempted and the Offensive Rankings of the teams in each period of time. In order to do this, the data for each period was put into its own array. Again, this information was taken from the Basketball-Reference website. The linearRegression()[4] function was used to calculate the regression on the data in the arrays. The source for this has been referenced in the footnote. The offensive ranking was used as the dependent variable while the three point attempt rate percentage was used as the independent variable.
Although there appeared to be a slight trend, our analysis indicates that the r^2 value is quite low, explaining that only about 13% of the variance. However, we believe that given all the different metrics that go into calculating Offensive Rating, the fact that three point shots alone explain about 13% of the variance is quite good.

Part 4: Inference:

i. **The Heat Maps:** There is a definite upward trend in the number of shots taken from the 3 point zone as the years have progressed in the NBA.

ii. **The Line Plot:** Most of the best shooters in the NBA have consistently been increasing their 3 point Field Goal percentage with the passage of time, the only exception being Lebron James who has seen a sharp drop in this statistic from 2014-15 to 2015-16. Stephen Curry is the most consistent among the top shooters at distance shooting.

iii. **The Regression Analysis:** Although an upward trend exists, the low r^2 value suggests that the offensive ranking of teams isn't much correlated with the three point attempt rate percentage. However it is worth noting that it seems that as the years go by, there has been an increase in offensive rankings and three point attempts. We highlight the Golden State Warriors as an ideal example.

[4] http://trentrichardson.com/2010/04/06/compute-linear-regressions-in-javascript/