NBA Player Statistics Visualization

CS6630 Final Project Proposal

<https://github.com/wilsonCernWq/NBAstatsVIS.git>

Qi WU

Mengjiao HAN

Qiuhua SHENG

Background and Motivation

As more and more accurate recording technologies are involved in modern basketball games, analyzing and visualizing basketball’s performance based on enormous statistics has become a hot topic for sport analysts and team managers. With the help of a great visualization tool, an analyst/manager/coach can quickly compare and measure players’ abilities properly and reduce the chances and costs of making bad decisions. Although there are a great number of basketball data analyzing tools on the internet nowadays, it’s still hard to find a suitable tool for a non-programming-expert to make customized visualizations with well-designed styles. As basketball enthusiasts, we want to create a player profiler for non-experts to explore data and create visualizations based on their own demands. In our mind, it should be a visualization tool with user-friendly user-interface and the ability to build visualization with beautiful templated styles.

Project Objectives

There are some primary questions we want to answer for this project.

* What are statistics for properly measuring a player’s overall abilities?
* What are the common ways to visualize different kinds of basketball statistics?
* How to classify data and create general visualization templates for customized styling?

To answer all the questions, we will need to first get familiar with what kinds of statistics we can get and classify them. After some investigation, we found there are two different types of datasets we can fine in general for a basketball player. One is spatial datasets which describe a player’s behaviors play-by-play with coordinate on the basketball court. We can draw spatial distributions for those kinds of dataset. The other type is table datasets which measure a player’s average performances within a defined period. For those kinds of data, common visualization methods like a bar-chart will work. Finally, to give user abilities to create customized visualizations, we need to firstly provide a powerful dynamic-query tool for users to profile data, then implement programs that can fit different kinds of data. We should also provide several well-designed styling templates and user-friendly styling tools which gives users abilities to modify the designs and formats.

Data

We will collect data from the API provided by [www.nba.stats.com](http://www.nba.stats.com). The API generates JSON files with raw data based on query parameters we provided. We plan to use mainly three kinds of data sets. The first dataset is the list for all active players in each year (The example is given here: <http://stats.nba.com/stats/commonTeamYears?LeagueID=00>). The second kind of dataset is the career summary of any given player (The example is given here: <http://stats.nba.com/stats/playercareerstats?PerMode=PerGame&PlayerID=977>). The third dataset is the performance map for a given player in each year (The example is given here: <http://stats.nba.com/stats/shotchartdetail?ContextFilter=&ContextMeasure=FGA&DateFrom=&DateTo=&GameID=&GameSegment=&LastNGames=0&LeagueID=00&Location=&MeasureType=Base&Month=0&OpponentTeamID=0&Outcome=&PaceAdjust=N&PerMode=PerGame&Period=0&PlayerID=977&PlusMinus=N&Position=&Rank=N&RookieYear=&Season=2012-13&SeasonSegment=&SeasonType=Regular+Season&TeamID=0&VsConference=&VsDivision=&mode=Advanced&showDetails=0&showShots=1&showZones=0>). Such map will be referred as “shot chart” later in this proposal.

Data Processing

Since the spatial datasets we will obtain are in general play-by-play raw records, we will need to do substantial data cleanup for most of our data. We will need to compute some meaningful statistical quantities such as the probability density distributions (PDF), mean performance or even performance variance for each player. The data processing will be done in advance to save rendering time. Common data processing tools such as MATLAB, Excel or Python might be used. The mathematics behind the processing processes should agree with other common data mining techniques, which should not be the main interest of our project.