**Shift in the Game, and other phenomenon in the NBA.**

**Summary of research questions and results**

1. Is the game shifting towards big men or guards? Are more shots taken in areas favoring big men or guards, based on the assertion that guards are traditionally shooters and big men play in and around the “paint” or painted area.

It appears that there isn’t any apparent shift towards either guards or big men.

1. Does the phenomenon of a cold streak or hot streak exist (being “cold” and getting “hot” respectively)? Will missing a couple of shots in a row truly be a predictor for future shots, and likewise, will sinking a couple of shots in a row, lead to more made shots?

The results showed that making consecutive shots is more likely to lead to more made shots, compared to missing more shots after missing consecutive shots.

1. Does the 82 game long NBA season begin to take a toll as the season progresses? Playing anywhere from 30-40 minutes per game, will it begin to prove detrimental to the performance of teams later in the season, or does the “playoff” push offset this?

There does not seem to be any evidence to support that during the end of the season, performance begins to go down.

**Motivations and Background**

 The game of basketball, in this case the NBA, is ever evolving. Not only aesthetically, going from short-shorts to what we now know as basketball shorts, but also mechanically, whereas in the past there was no coaching allowed, but even no dunking. These things now are prevalent in all levels of basketball only to further the point that the NBA is constantly changing. Understanding these changes along withother phenomenon surrounding the game can be very helpful in many faucets of the game. The uses of the answers to these questions are not singular; they can be used throughout the NBA, such as a tool to scout what players might have a leg up in the current NBA game. Other uses may include dictating how a coach may allocate his minutes and what kind of defenses to use, what players a GM might want to sign to his team, or even the players themselves getting a better understanding for the game that they play. This is very similar approach to what the field of sport analytics seeks out to do.

**Datasets:**

I am also using a dataset gathered from http://www.basketballgeek.com/data/.

This is a dataset of play-by-play data for the seasons from the 2006-07 season to the 2009-10 season. Each file represents one game and includes information for every play, from who was on the court to the coordinates of the shot, along with much more information.

I was planning to use data I obtained myself using a crawler I made, but the scope of the first data set had enough information for it to be unnecessary to use this and it would be a more informative project if I focused on a few questions than many, so I tried to give good visual representations for each question; however I did provide the make-shift crawler and its downloaded dataset from <http://www.basketball-reference.com>, just for reference.

**Methodology:**

1. Essentially seeing if more shots are taken/made within specified quadrants on the court using the shots coordinates. Each quadrant favoring smaller or larger players. The area around the basket, known as the “paint” would be one quadrant, and the other would be anything beyond that. Comparing the amount of times a shots coordinates fall within either of those quadrants and finding a ratio of outside:paint, and comparing it to ratios among other seasons of play and seeing if there is a net decline or progression in the scope of the seasons we are looking at.-
2. After 3 in a row missed shots or 3 in a row made shots, comparing subsequent performance in the same game, and seeing if field goal percentage falls significantly below or above a specified “*Mendoza Line*”, which is an arbitrary number used in baseball (.2) that is the cutoff for an incompetent hitter, and in this case will be .4(average field goal percentage), as a gauge to whether the player underperformed or over performed. If it’s below .3 (considered a bad game), then it did affect performance otherwise, it did not, and for a hot streak same goes for if it’s above .5 (considered a good game). Also ignore games where less than 3 more shots were taken after a cold or hot streak, eliminating outliers. After this comparing the amount of times a hot/cold streak resulted in a “good” or “bad” game respectively to the total amount of hot/cold streaks, to find the how often these streaks lead to any noticeable effect.
3. For the sake of this question we will define the “playoff push” as the last 20 games of the season or somewhere approximately around March 1st around this time is when the playoff race begins to get tight. To get a gauge of performance of a team, we will get the average T/O (turnovers) per game for all teams before March 1st, and compare those averages to the same teams after March 1st. If there are more teams with more T/O after March 1st, we can say that fatigue does factor in, otherwise, we cannot.

**Results:**

1. Looking at the heat maps and scatterplots that I created, it shows that overall most seasons tend to have the same shot location frequencies. I thought there would be a trend going either beyond the three point line or within the painted area, but based on the heat map, it doesn’t appear to be so. Also the net change in paint shots: total shots was only .6 percent, which is relatively miniscule.
2. Based on the pie charts, it seems apparent that making 3 shots in a row will lead to success more often than missing 3 shots in a row will lead to failure. Something that was really surprising about this question was that I expected it to be the other way around, because it seems that it would affect you mentally throughout the game, but looking at the cold streak chart where only 22.19% of people actually shot badly compared to the hot streak chart where over double (~49%) that amount of people shot well after sinking three in a row.
3. Seeing the TO pie chart and seeing how it is almost 50/50 (47.5/52.5), it shows that half the teams were worse and half were better, there isn’t really any notable reduced performance at least based off of TO averages. This somewhat helps support my personal idea that the fact that the stakes get so much higher the closer you are to the end of the season, possibly offsetting any effect fatigue might have, however I cannot confirm this, it could be anything, or nothing at all.

**Reproducing Results:**

You can download the box scores at <http://www.basketballgeek.com/data/> , and you will need the four seasons they provide, however I have provided it in the folder called “Box Scores”. You can run this program through a client or through the command line, if you provide no arguments or run it through IDLE then it will look for a default folder in the CWD called “Box Scores” which to look through, which I’ve provided, or you can pass in your own folder name where you have the box score datasets. When running the program, it will print any pertinent information and also output different figures within the CWD for each of the questions.

**Collaborations:**

 None.

**Reflection:**

 This project really helped me see what I can and can’t do, with regards to sheer size of a project. Also it was great actually being able to analyze a dataset, in which I am really interested in, rather than one chosen for me. This will make me want to do my own future projects possibly in other languages if I can learn them.