Introduction

Our team was brought together because of our shared interest in the NBA. Over the last decade, analytics have become increasingly paramount in many sports. To the mainstream, this revolution was popularized by the 2011 film ‘Moneyball’. It depicted the Oakland Athletics and their General Manager Billy Beane’s (played by Brad Pitt) attempt to field a competitive team with a limited budget. He used analytics that most teams were not using to find undervalued players. This movie romanticized the benefit of using advanced analytics in conjunction with the more traditional statistics and scouting. To many, this was the first time that advanced analytics were associated with sports. While Moneyball centered on a baseball team, this was occurring across all sports, arguably none more than our beloved NBA. In 2009, a NYT article by Michael Lewis ‘The No-Stats All-Star’ <https://www.nytimes.com/2009/02/15/magazine/15Battier-t.html> examined the approach that the Houston Rocket’s General Manager, Daryl Morey, was implementing. An M.I.T. Business school alumni, he realized that a player’s actual value was often not well encapsulated by traditional statistics. By taking a more comprehensive approach to evaluating players, he discovered that he could benefit from signing players who were conventionally undervalued. If he valued players differently (and more accurately), he could take advantage of this via trades, drafting or on the free agent market.

A major impetus to this approach in the NBA was the creation of SportsVU. SportsVU is a set of cameras that are hung in the rafters of NBA arenas that collect data 25 times per second. They collect data in the X,Y and Z dimension of both the ball and all the players on the court. It was first implemented in an NBA arena for the 2011-12 season. 2 years later, SportsVU was installed in all 30 NBA arenas. The amount of data this created would have been unthinkable a mere decade earlier. Players were previously evaluated mainly on their ‘box score statistics’ like points, rebounds, assists. SportsVU now enabled us to completely rethink the way we evaluated players. This changed the landscape of the NBA in many ways. Some players saw their values rise, while other players who were previously celebrated became maligned. The contours of NBA management also changed as MIT students were now given a seat at the table alongside ex-players.

The first question we chose to tackle was how rest affects the shot charts of NBA players. The first way we examined this was looking at each team’s absolute rest. If a player’s team played the previous day, then all shots during the game the next day were classified as “0” days rest. If a team played on Monday, was off Tuesday, and played Wednesday, then all shots taken by that team during their Wednesday game were classified as “1” days rest. Our categories for Absolute Rest were 0,1,2 and 3+ days. We used a catch-all category of 3+ for all games where the team had an Absolute Rest score of 3 or more.

We then went a step further with our classification of rest. There are 2 teams in every game, so why not classify one teams rest in relation to the others. We theorized that a team’s ‘absolute rest’ might matter less than a team’s ‘relative rest’. For each game, we compared the rest of both teams. If both teams were equally rested, they each had a ‘relative rest’ classifier of 0. An example: if the Knicks and Lakers were playing on a Wednesday, we would consider when each team played their last game. If the Knicks last played on Monday, they would have an absolute rest of 1. If the Lakers last played on a Tuesday, they would have an absolute rest of 0. Relative Rest was calculated by comparing those 2 values. For that specific game, the Knicks would have a Relative Rest of +1, while the Lakers would have a Relative Rest of -1. Our categories for Relative Rest were -2+, -1,0,1, 2+. We used a catch-all category of ‘-2+’ and ‘2+’ for Relative Rest. If a team had Relative Rest of less than -2, we classified it as -2. We did the same for values of 2+, classifying them as 2. We did this in the interest of avoiding categories with very limited information.

The second question we chose to answer was how each team’s shot chart varied. For this, we looked at the shot charts of all 30 teams.

Data Souces

The entire SportsVU data set for the 2016-17 NBA season is publicly available. It is included in the SpatialBall package in R. This dataset included information on all 202,029 shots taken during that season. It provided the X-axis and Y-axis location for each shot, which enabled us to create shot charts. It also included whether the shot was a make or miss, what type of shot it was (jump shot, layup, dunk, etc) in addition to many other identifiers like the player, time of shot, etc. This was the main source of our data. We merged another publicly available dataset from Basketball-Reference.com. This dataset had more traditional statistics such as player points, rebounds, etc along with more specific information on player’s age and team schedules.

Absolute Rest

Our first visualization shows a shot chart for the entire 2016-17 season. We faceted by Absolute Rest to show 4 different shot charts (0,1,2,3+ days of Absolute Rest). In this visualization, we treated the data as spatial data. Including the X-Y location of each shot is valuable because the geography matters. Shots from further out are generally harder than close up shots. Varying shot location distrubutions by absolute rest may indicate a player takes sub-optimal or different shots when they are tired as opposed to well rested. Unfortunately, this method had too many data points and the visualizations appeared very clustered. Despite our best attempts at altering the transparency(alpha) and point size, there were still too many data points for us to make sense of any differences in shot location. It is possible, that if we used a subset of our data(possibly just 1 team or games played within a smaller time frame), our visualizations might have been more beneficial.

Our next visualization continued to explore the way Absolute Rest affect’s a player’s shot. For this, we used the variable Shot Zone that geographically breaks the court up into 7 regions (Backcourt, left corner 3, midrange, etc). We then used the make/miss variable to show the overall shooting percentage within each shooting zone. Again, we faceted by Absolute Rest. Despite removing the spatial data(X-Y coordinates), we still chose to show the shooting percentage of each category(the 7 regions on a court) in the corresponding area on the court. Our hypothesis was that shooting percentages would rise with more rest. For this visualization, our hypothesis did not bear out. There was no discernable trend in shooting percentages by shooing zone when faceted by Absolute Days rest.

The next visualization looks at the exact same data as the previous one, but in a different format. We removed any spatial depiction of our data and chose to display is as multi-variate categorical data on a Parallel Coordinate Plot. Just as in the previous visualization, we showed how shooting percentage by shot zone varied with Absolute Rest. We found the Parallel Coordinate Plot to be the clearest way to show this information. Although shooting percentage trended upward in a few shooting areas as absolute rest increased, no real trend emerged.

Relative Rest

Next, we moved onto Relative Rest. Similar to our visualization for Absolute Rest, we plotted the shot accuracy for each of the 7 shot areas faceted by Relative Rest. Although not consistent among all the shot zones, there does appear to be an upward trend in shooting percentages as Relative Rest increases.

We continued to look at Relative Rest. We used a horizontal stacked bar chart to show the proportion of wins/losses based on Relative Rest. We hypothesized that a team who is more well rested compared to their opponent should have an advantage. This proved true in our visualization, as winning percentage increased as a team was more well rested relative to their opponent.

Next, we created a visualization to show how the frequency of each type of shot varied by Relative Rest. Each shot was categorized into 52 different categories such as ‘cutting layup shot’, ‘tip dunk shot’, etc. We chose to create a vertical stacked bar chart for this visualization. This was a case in which too much data can be overwhelming. Because there were so many categories, it was hard to decipher a difference.

Next, we created another vertical stacked bar chart to show how the frequency of each shot area varies by Relative Rest. Although this visualization is very clear, it is hard to decipher if there are any differences. It is possible a parallel coordinate plot may have been clearer in this situation.

We made another vertical stacked bar chart to show how the frequency of each shot range (less than 8ft, 8-16ft, 16-24ft, etc) varies by Relative Rest. Again, it was hard to decipher a difference among frequency of shot ranges by Relative Rest.