# An Analysis of Predicting Performance of NBA Basketball Players Using Markov Chains

Matthew Elias, Bao Aijia, Ye Cao, Caroline Cannistra

## Introduction

In the reign of sports, almost everything can be marked as stochastic events. The probability of winning a game, an MVP, a championship... Because of these "random" events, sports attract fans in every era of human history. Chicago Bears made a 3-1 come back winning the championship of NFL after a waiting of 103 years. Cleveland Cavaliers also made a turn from a 3-1 trailing and won first title of NBA championship after 52 years. The fascination of sports led us to the topic of how NBA players perform from season to season. As every NBA fan would be interested in knowing who is going to be next super scorer in the league, we would like to explore the topic of season by season average points per game in this project.

Since Markov Chains are useful for modeling the change of process in discrete states, we can using Markov Chain to model the variation of average points per game an NBA basketball player get in each seasons with the different seasons as our discrete states. In addition, it may help us analysis and predict the future performance of an NBA basketball player given current performance.

For this project, the overall goal was to take collection of data, and then apply methods of mathematical data modeling to explore interesting phenomena. Specifically, we set out to develop a model using Markov Chain and apply statistical methods to analysis the variation of performance of NBA basketball players from seasons to seasons.

## Background

To decide on the type of data we would be investigating, we first listed a collection of ideas such as automatic composition of music, skill of a hobby and its effect on skill in the future, and predictive modeling of basketball games. After considering the accessibility of data and the interests of our group members, we finally decided to choose predicting the performance of NBA basketball players as our topic.

There are numerous papers and projects about predicting NBA player performance. For example, in *Predicting NBA Player Performance*(Wheeler, 2012), Kevin Wheeler developed a model using linear regression to predict the number of points NBA players would score against an opponent. In particular, he used linear regression with Naïve Bayes and SVM implementations to compare the two methods.

Another example is *Evaluating Basketball Player Performance via Statistical Network Modeling*(Piette et al., 2011). In this paper, James Piette and his colleagues applied a network analysis technique to form a more complete analysis of a player's abilities in order to evaluate the basketball player's performance. They implemented bootstrap techniques and used analysis of frequency to determine the statistical contribution of a player in this network.

Most of the papers we found implement a statistical method to analysis and predict NBA basketball players' performance. Markov Chains as a modeling method doesn't appear often in basketball modeling. However, Markov Chain is not new to basketball. In Kenny Shirley's project *A Markov Model for Basketball*(Shirley, 2007), she modeled a basketball game as a sequence of transitions between discrete states and concluded that the home team has a $61 - 65\%$ chance of winning on average.

To the best of our knowledge, Markov Chains have not been used to model the performance of NBA basketball players over time. In the following part, we will show how we use Markov Chains and other statistical methods to analyze and predict NBA basketball players' performance over time.

## The model itself

The main model of our project studied the average points scored per game of NBA players season by season, applying Markov Chains. Specifically, we aimed to see how many points per game an NBA player scores on average after scoring a certain average of points per game in the previous season. Thus, our transition matrix followed this idea. We marked the row indices as average points per game in last season, and column indexes as average points per game in the next season.

After collecting the data and composing the transition matrix, we took the matrix to power of $n$ times, with $n$ representing a number of seasons, to examine a player's chance starting from a certain average points per game to reach different average points per game in $n$ seasons. By studying these states, we got to know how the points per game performances of players change from state to state.

## Data Collection

Since an NBA basketball player's ability and value rated by public is strongly correlated with points the player can get in each game, we chose points per game for all players who match the criteria listed below as our data input.

| | |
|---|---|
| Season Type: | regular seasons |
| League: | NBA/BAA |
| Seasons: | 1996-97 to 2016-17 |
| Position: | G & F & C |
| Total game played: | >=30 |

In other words, we only considered games in regular seasons played in NBA from 1996-97 to 2016-17. All active players of all positions were considered right now. In addition, we only considered players who have a total game played no less than 30 games. The data we collected is from the website: http://www.basketball-reference.com/. The data was in the form shown below:

| Rk | Season | Age | Tm | Lg | G | Player | ...... | PTS | FG% | 2P% | TS% |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1996-97 | 27 | SAC | NBA | 75 | Mahmoud Abdbul-Rauf | ...... | 13.7 | .445 | .468 | .524 |
| 2 | 1997-98 | 28 | SAC | NBA | 31 | Mahmoud Abdbul-Rauf | ...... | 7.3 | .337 | .405 | .405 |
| ......... | ......... | ......... | ......... | ......... | ......... | ......... | ...... | ......... | ......... | ......... | ......... |

In our project, we used the categories "Season", "Tm"(Team), "Player", and "PTS" for our purposes. In addition, In order to build a Markov Chain with different PTS (average points per game) corresponding to different seasons as our states, we need to make sure a player played at least two seasons. To control for this, we deleted data entries for players who only played a single season. Meanwhile, since the last season a player played does not have a next season to compare, the retiring state is not counted. Multiple datasets were used in this model: For the main models in our project, we only used players who are still active now; however, when we built the model for comparing the variation between specific seasons, we used both inactive and active players, since there are very few players who are still active now and played in seasons before 2000. We also included inactive players in our analysis of Hall of Fame players.

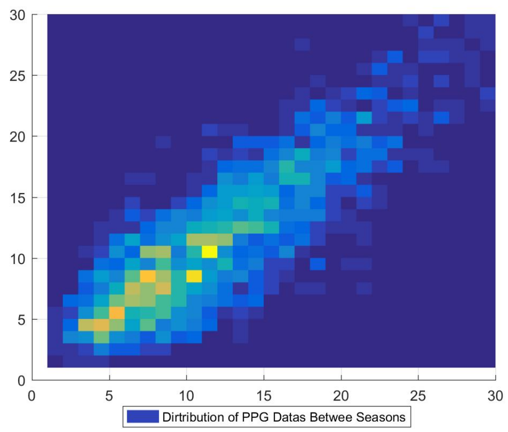
A major part of our project became comparing the differences between transition matrices quantitatively. A major measurable difference between these matrices comes from how they are structured along the diagonal. For example, compare two transition matrices, where one is the identity matrix and the second is structured such that every value in the matrix is exactly the same. These matrices are very different in meaning, and a simulation through both would show very different results. To quantitatively measure the differences between the two matrices, we created a "volatility index", which weighs the points as the absolute distance row/column-wise from the diagonal multiplied by the probability value at that diagonal. The algorithm then takes the average value of the matrix as the volatility index (this is found through the sum of all the cells divided by $n$, where $n$ is the number of cells). Here is the equation in mathematical form:

$$\frac{\sum |i-j| * p_{ij}}{n}$$

From this, we see that in the example above, the identity matrix would have a volatility index of zero, while the second one would have a volatility index that is positive and not relative to the size of the matrix, assuming that it is a valid transition matrix.

## Results

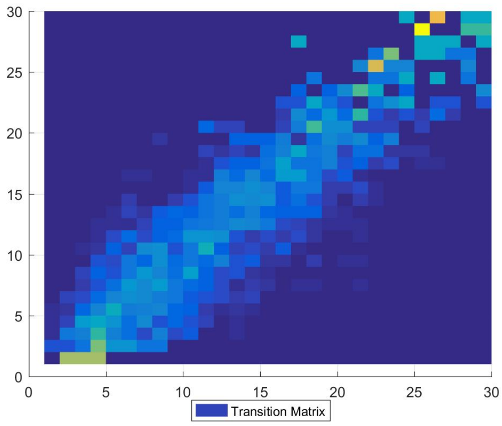
After data collection, a heat map of the transition from last season to next season matrix was mapped for the main model:



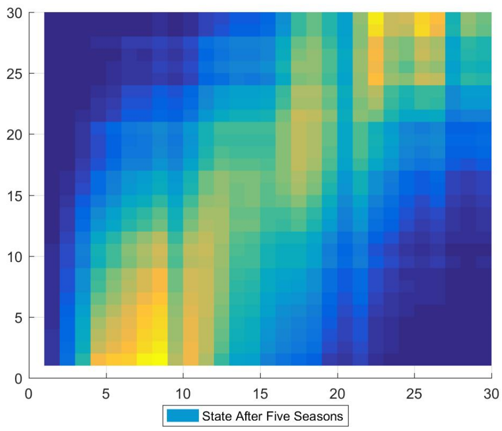

As shown, data were mainly distributed diagonally. This indicates, in general, points per game of a player in one season varied approximately within 5 points from last season. Another important aspect was that colors on the lower-left corner of the graph were brighter (note that the brighter the part of graph is, the higher the value is in that cell), showing that most players in each season have a relatively low average point score per game.

The transition matrix could be made from the distribution, and it shared a similar shape as the distribution. However, colors on top-right and lower-left were brighter. One of the main reason this happens is that, as we observed earlier in the paper, normally points per game increase or decrease approximately within 5 points from last season. Since 30 points per game and 0 points per game act as upper bound and lower bound of the data set. There were no cases of a currently active player scoring greater than 30 points or lower than or equal to -1 points per game. This caused more possibilities on increasing or keeping points per game while last season points per game is around 0 and decreasing or keeping points per game while last season points per game is around 30.

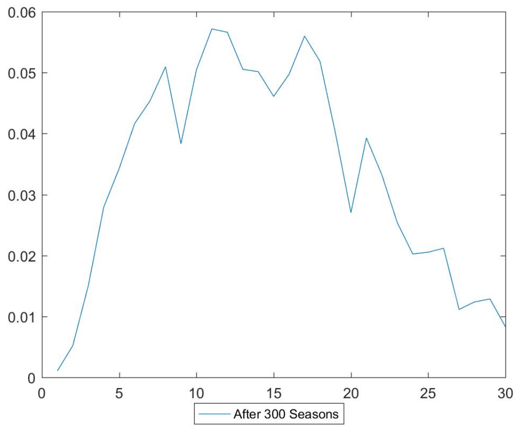
The volatility calculated for this transition matrix is 0.08523. However, this value is only meaningful while we compare it with the volatility value of other transition matrices.




As a reference, a five seasons state was achieved by taking fifth power of the transition matrix. This graph shown the distribution of how a player's average points per game are predicted to be in five seasons starting at a certain points per game.




After a 300-season simulation, the limiting distribution of the transition matrix could be achieved because this transition matrix is ergodic. Though 300 seasons are implausible in realistic meaning, the limiting distribution showed us on average what is the probability of staying on each point range:




As shown, the plot had peaks around the range 10 to 15 points per game with slight skew to higher values.

## Adjustments and extensions

A major issue in our model was to make it robust, such that we could use similar methods for variations while avoiding errors. For example, multiple problems arose when we started breaking down the model to look at the differences in seasons/positions/teams. Since we were working with non-repeating data, there was an issue that comes up when there was a limited amount of data applied to outliers' cases. Consider a situation where only one player goes from scoring 20 points in one season to 30 points in the next season. If this happens to be the only recorded situation of a player reaching 30 points, we create an accidental absorbing state where a player in this model has a non-zero chance of reaching 30 points and staying there forever. This isn't as much of an issue when there are many data points, but creating a new Markov Chain for each season leads to less data per matrix, which raises the chance of running into such an issue. This was solved by creating a bin-size parameter, where the Markov Chain has rows and columns that are arbitrary and require supervision to avoid these situations. The bin sizes change based on the size of the data and how we are splitting it for analysis, but in each method of analysis, all Markov Chains that are created for that particular extension have the same size of bins. This is for consistency purposes to make it easier to compare the matrices.

## Analysis for different positions

Instead of analyzing PTS data from all of the NBA players as a whole, we broke it to three parts. We built three distinct matrices for NBA players: center, forward and guard position. In this variation, we can compare models for different positions, and compare them with the main model.

## Transition Matrix for different Positions

We built a transition matrix corresponding to each position with the same method we use in our main model. Here is an example of a transition matrix for center position (The complete transition matrices are attached in appendix) :

| | Transition Matrix for Center Position | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 0-3 | 3-6 | 6-9 | 9-12 | 12-15 | 15-18 | 18-21 | 21+ |
| 0-3 | 0.4063 | 0.4063 | 0.1563 | 0.0313 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 3-6 | 0.1867 | 0.4267 | 0.2933 | 0.0667 | 0.0267 | 0.0000 | 0.0000 | 0.0000 |
| 6-9 | 0.0303 | 0.2273 | 0.4545 | 0.2424 | 0.0455 | 0.0000 | 0.0000 | 0.0000 |
| 9-12 | 0.0189 | 0.1132 | 0.1509 | 0.4528 | 0.1887 | 0.0566 | 0.0189 | 0.0000 |
| 12-15 | 0.0000 | 0.0800 | 0.0400 | 0.1600 | 0.4400 | 0.2000 | 0.0800 | 0.0000 |
| 15-18 | 0.0000 | 0.0000 | 0.0000 | 0.0769 | 0.3846 | 0.2308 | 0.3077 | 0.0000 |
| 18-21 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.1667 | 0.2500 | 0.5000 | 0.0833 |
| 21+ | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 1.0000 | 0.0000 |

In this transition matrix, we combined 3 subsequent PTS into 1 state, and all the PTS over 21 were also considered to be one state. In this case, There were only 8 states for Center Position: 0-3, 3-6, 6-9, 9-12, 12-15, 15-18, 18-21 and 21+. Thus, we can see the probability for any center position player who originally had a PTS 0-3 got a PTS still in the range of 0-3 in next season is 0.4063. Though choosing range of points per game would lose certain precision of the model, the reason we chose three-point interval was because of the lack of data. If we choose 30-times-30 size transition matrix, then there may exist more than half of 0 elements, and more importantly, numbers of 1 elements which would cause potential limiting states.

The graphs of the transition matrices are shown below:

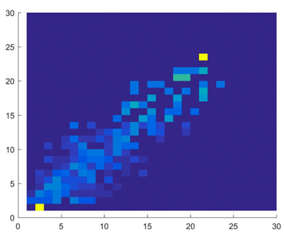


Figure 1: Transition Matrix for Centering Position

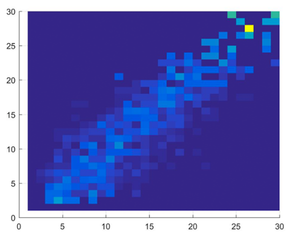


Figure 2: Transition Matrix for Guard Position

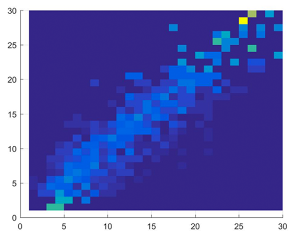


Figure 3: Transition Matrix for Forward Position

Where the highest probability plots in yellow, and the lowest probability which is 0 plots in dark blue. The higher the probability, the brighter the color.

Comparing the three graphs, we found that the plots for all three figures were still mostly stick near the diagonal line. It gave us the information that points per game of a player will not varies a lot after one season. In addition, we noticed that the diagonal line composed by plots in Figure 1 is shorter than the other figures, which means it has no plots in the upper right corner. This showed NBA Basketball players who is in center position do not reach a PTS over 25. In fact, since center position focus on defense and rebounding, they do not usually shoot and get high points in a game.

The results were similar when we look at the distribution for the three positions.

## Distribution for different Positions

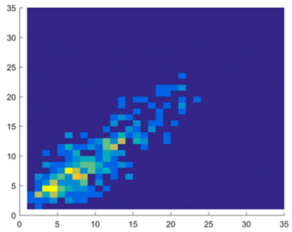


Figure 4: Distribution for Centering Position

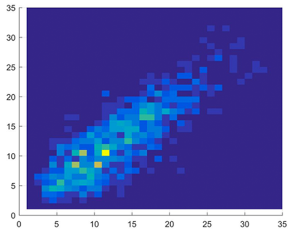


Figure 5: Distribution for Guard Position

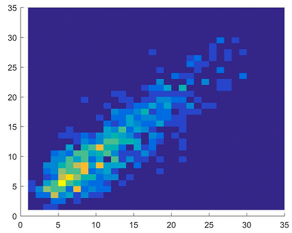


Figure 6: Distribution for Forward Position

By looking at the graphs for distributions, other than no plots in upper right corner for center position, we could also see that for all three figures, the left bottom corner near the diagonal line is brightest and there are some small differences between the range of the brightest area. This shows most of the center position have a PTS around 0-15, Forward position is around 3-20, and Guard position distributed averagely.

## Markov Chain for different positions after several seasons

By taking the transition matrix to $n$ powers, we can find the probability of changing each state to another after $n$ seasons. Below is the Markov Chains for three different positions after five seasons.

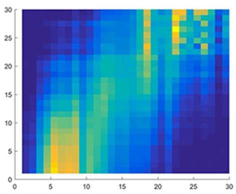


Figure 7: Markov Chain for Forward Position After Five Seasons

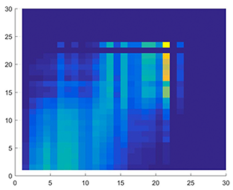


Figure 8: Markov Chain for Center Position After Five Seasons

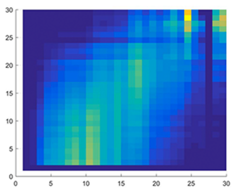


Figure 9: Markov Chain for Guard Position After Five Seasons

We can see after five seasons, forward position has two brighter parts, one in bottom left corner and one in upper right corner. However, color for center position and Guard position are more average.

When we powered up the transition matrix to 300, the matrix had constant columns. Matrices below shows

$$\lim_{n \Rightarrow \infty} A^n$$

| | Transition Matrix for Center Position after 300 seasons | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 0-3 | 3-6 | 6-9 | 9-12 | 12-15 | 15-18 | 18-21 | 21+ |
| 0-3 | 0.0710 | 0.1797 | 0.1776 | 0.1711 | 0.1808 | 0.0966 | 0.1138 | 0.0095 |
| 3-6 | 0.0710 | 0.1797 | 0.1776 | 0.1711 | 0.1808 | 0.0966 | 0.1138 | 0.0095 |
| 6-9 | 0.0710 | 0.1797 | 0.1776 | 0.1711 | 0.1808 | 0.0966 | 0.1138 | 0.0095 |
| 9-12 | 0.0710 | 0.1797 | 0.1776 | 0.1711 | 0.1808 | 0.0966 | 0.1138 | 0.0095 |
| 12-15 | 0.0710 | 0.1797 | 0.1776 | 0.1711 | 0.1808 | 0.0966 | 0.1138 | 0.0095 |
| 15-18 | 0.0710 | 0.1797 | 0.1776 | 0.1711 | 0.1808 | 0.0966 | 0.1138 | 0.0095 |
| 18-21 | 0.0710 | 0.1797 | 0.1776 | 0.1711 | 0.1808 | 0.0966 | 0.1138 | 0.0095 |
| 21+ | 0.0710 | 0.1797 | 0.1776 | 0.1711 | 0.1808 | 0.0966 | 0.1138 | 0.0095 |

This matrix showed that in the long run, the probability for center position reach each state was constant, for example, from any states to 0-3 PTS was 0.0710, 3-6 PTS was 0.1797. State 12-15 PTS had biggest possibility and 21+ had the smallest. So, center position players who begin at any PTS had the most chance ending up had a PTS in 12-15 which had a probability equals to 0.1808, and had the least chance ending up in range 21+ had a probability equals to 0.0095.

| | Transition Matrix for Forward Position after 300 seasons | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0-3 | 3-6 | 6-9 | 9-12 | 12-15 | 15-18 | 18-21 | 21-24 | 24-27 | 27+ |
| 0-3 | 0.0301 | 0.1533 | 0.1658 | 0.1807 | 0.1403 | 0.1354 | 0.0889 | 0.0563 | 0.0379 | 0.0114 |
| 3-6 | 0.0301 | 0.1533 | 0.1658 | 0.1807 | 0.1403 | 0.1354 | 0.0889 | 0.0563 | 0.0379 | 0.0114 |
| 6-9 | 0.0301 | 0.1533 | 0.1658 | 0.1807 | 0.1403 | 0.1354 | 0.0889 | 0.0563 | 0.0379 | 0.0114 |
| 9-12 | 0.0301 | 0.1533 | 0.1658 | 0.1807 | 0.1403 | 0.1354 | 0.0889 | 0.0563 | 0.0379 | 0.0114 |
| 12-15 | 0.0301 | 0.1533 | 0.1658 | 0.1807 | 0.1403 | 0.1354 | 0.0889 | 0.0563 | 0.0379 | 0.0114 |
| 15-18 | 0.0301 | 0.1533 | 0.1658 | 0.1807 | 0.1403 | 0.1354 | 0.0889 | 0.0563 | 0.0379 | 0.0114 |
| 18-21 | 0.0301 | 0.1533 | 0.1658 | 0.1807 | 0.1403 | 0.1354 | 0.0889 | 0.0563 | 0.0379 | 0.0114 |
| 21-24 | 0.0301 | 0.1533 | 0.1658 | 0.1807 | 0.1403 | 0.1354 | 0.0889 | 0.0563 | 0.0379 | 0.0114 |
| 24-27 | 0.0301 | 0.1533 | 0.1658 | 0.1807 | 0.1403 | 0.1354 | 0.0889 | 0.0563 | 0.0379 | 0.0114 |
| 27+ | 0.0301 | 0.1533 | 0.1658 | 0.1807 | 0.1403 | 0.1354 | 0.0889 | 0.0563 | 0.0379 | 0.0114 |

The limiting distribution for Forward Position was different from the center position. Forward position players began at any PTS had the greatest possibility ending at 9-12 with P = 0.1807 and most unlikely ending at 27+ with P = 0.0114.

| | Transition Matrix for Gaurd Position after 300 seasons | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0-3 | 3-6 | 6-9 | 9-12 | 12-15 | 15-18 | 18-21 | 21-24 | 24-27 | 27-30 | 30+ |
| 0-3 | 0.0091 | 0.0708 | 0.1192 | 0.1577 | 0.1427 | 0.1714 | 0.1063 | 0.0936 | 0.0585 | 0.0573 | 0.0133 |
| 3-6 | 0.0091 | 0.0708 | 0.1192 | 0.1577 | 0.1427 | 0.1714 | 0.1063 | 0.0936 | 0.0585 | 0.0573 | 0.0133 |
| 6-9 | 0.0091 | 0.0708 | 0.1192 | 0.1577 | 0.1427 | 0.1714 | 0.1063 | 0.0936 | 0.0585 | 0.0573 | 0.0133 |
| 9-12 | 0.0091 | 0.0708 | 0.1192 | 0.1577 | 0.1427 | 0.1714 | 0.1063 | 0.0936 | 0.0585 | 0.0573 | 0.0133 |
| 12-15 | 0.0091 | 0.0708 | 0.1192 | 0.1577 | 0.1427 | 0.1714 | 0.1063 | 0.0936 | 0.0585 | 0.0573 | 0.0133 |
| 15-18 | 0.0091 | 0.0708 | 0.1192 | 0.1577 | 0.1427 | 0.1714 | 0.1063 | 0.0936 | 0.0585 | 0.0573 | 0.0133 |
| 18-21 | 0.0091 | 0.0708 | 0.1192 | 0.1577 | 0.1427 | 0.1714 | 0.1063 | 0.0936 | 0.0585 | 0.0573 | 0.0133 |
| 21-24 | 0.0091 | 0.0708 | 0.1192 | 0.1577 | 0.1427 | 0.1714 | 0.1063 | 0.0936 | 0.0585 | 0.0573 | 0.0133 |
| 24-27 | 0.0091 | 0.0708 | 0.1192 | 0.1577 | 0.1427 | 0.1714 | 0.1063 | 0.0936 | 0.0585 | 0.0573 | 0.0133 |
| 27-30 | 0.0091 | 0.0708 | 0.1192 | 0.1577 | 0.1427 | 0.1714 | 0.1063 | 0.0936 | 0.0585 | 0.0573 | 0.0133 |
| 30+ | 0.0091 | 0.0708 | 0.1192 | 0.1577 | 0.1427 | 0.1714 | 0.1063 | 0.0936 | 0.0585 | 0.0573 | 0.0133 |

For Guard Position player, They were most likely ending up 15-18 with P = 0.1714 and least likely ending up 0-3 with P = 0.0091.

Thus, we can see from above Guard Position player were more likely ending up has a high PTS than forward position and center position. The result was convincing considering that guards generally have a higher shooting percentage on 3-point shots. Meanwhile, the idea of modern basketball tends to utilize centers as screens to block out the defense opponent players and create space for guards to penetrate or to shoot.

### Volatility for different Positions

After applying the same function we used for main model to calculate the volatility to all three different transition matrices, we observed: 0.08907 for center, 0.08232 for forward, and 0.08456 for guard. For the center position, we reduced the number of bins to 24, because no center player has scored more than 24 points per game in the last 20 years. With this adjustment considered, the center position appeared to be the most volatile out of all three positions. The forward position was the least volatile, and the guard position is closest in volatility to the main model.

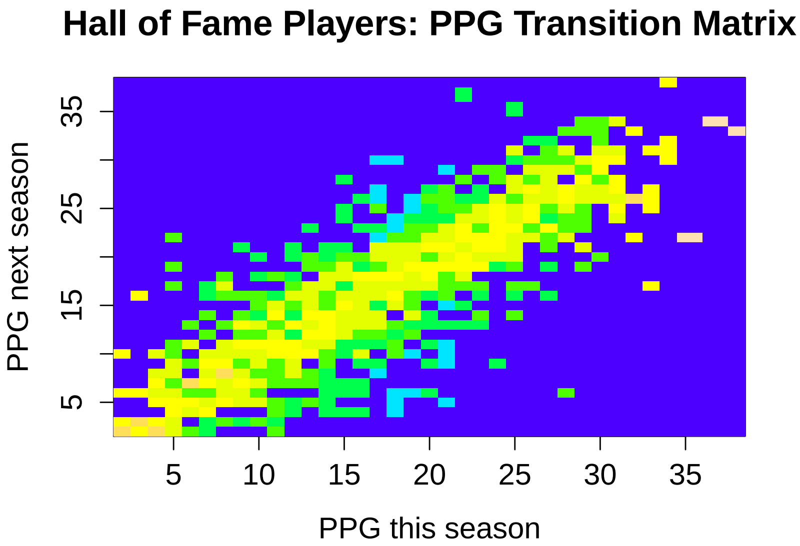
## Hall of Fame Players

A transition matrix was constructed for all NBA or BAA players inducted into the Hall of Fame. 1478 season records for 128 players were collected, with each player's PPG between 2 and 51. The player Wilt Chamberlain had a record of about 51 PPG in one season and 44 PPG in another season, with the next highest PPG in the league at 39, so he was removed from the dataset to simplify our analysis.

The matrix bins were set so that there were no empty cases, and each bin covered the range $[i, i+1)$ where $i$ is an integer.

The resulting transition matrix, as seen in the figure below, is heavily diagonal with most players between 5 and 30 PPG, but there are many outliers that make the matrix appear to be more volatile. Using the volatility metric from the dataset of all players, this matrix has a volatility of 0.09305, which is higher than that for the main model and for any individual position.

From this, we can tentatively claim that Hall of Fame players are less consistent than the average professional basketball player, and tend to vary more drastically in their performance.

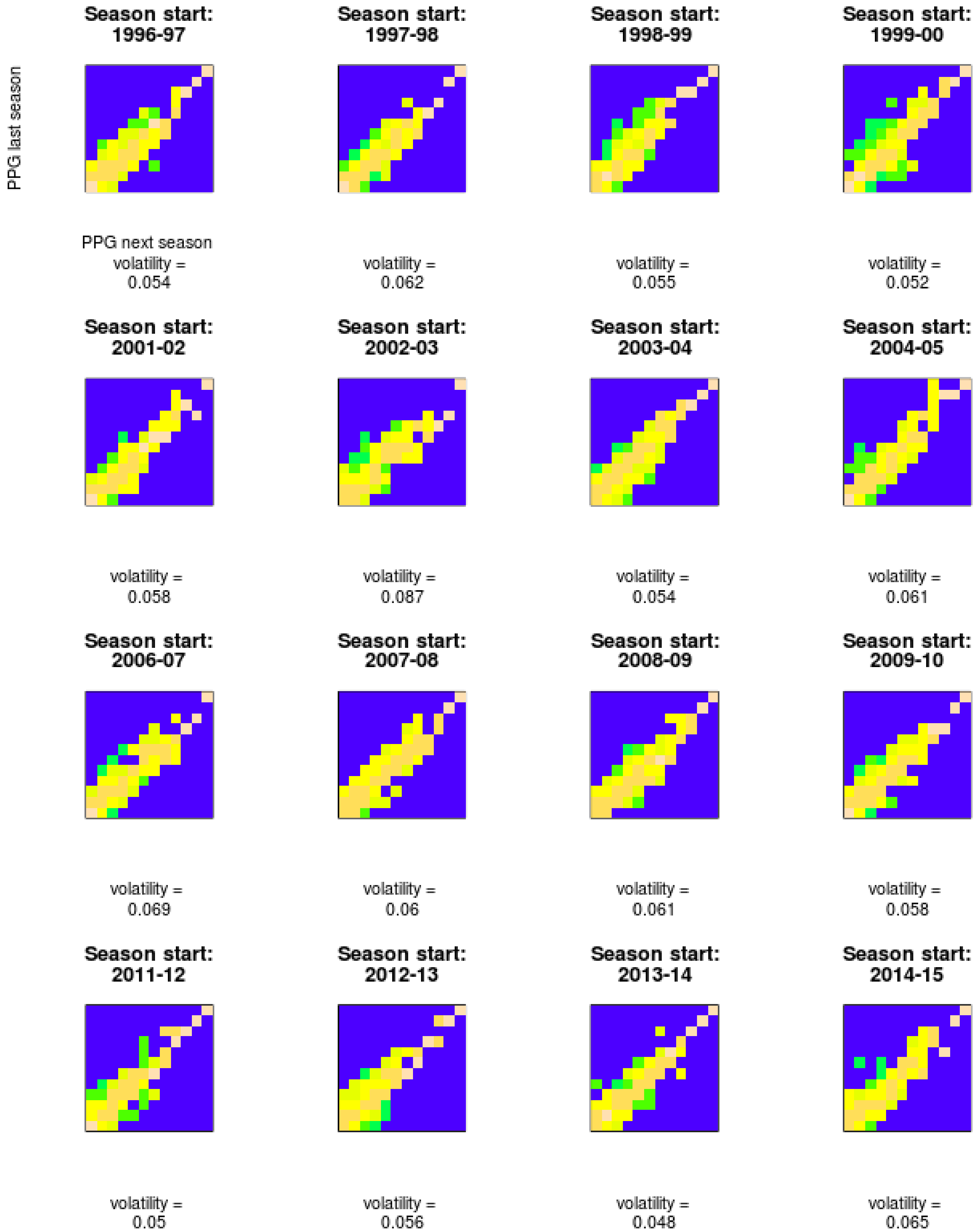



## Season Differences

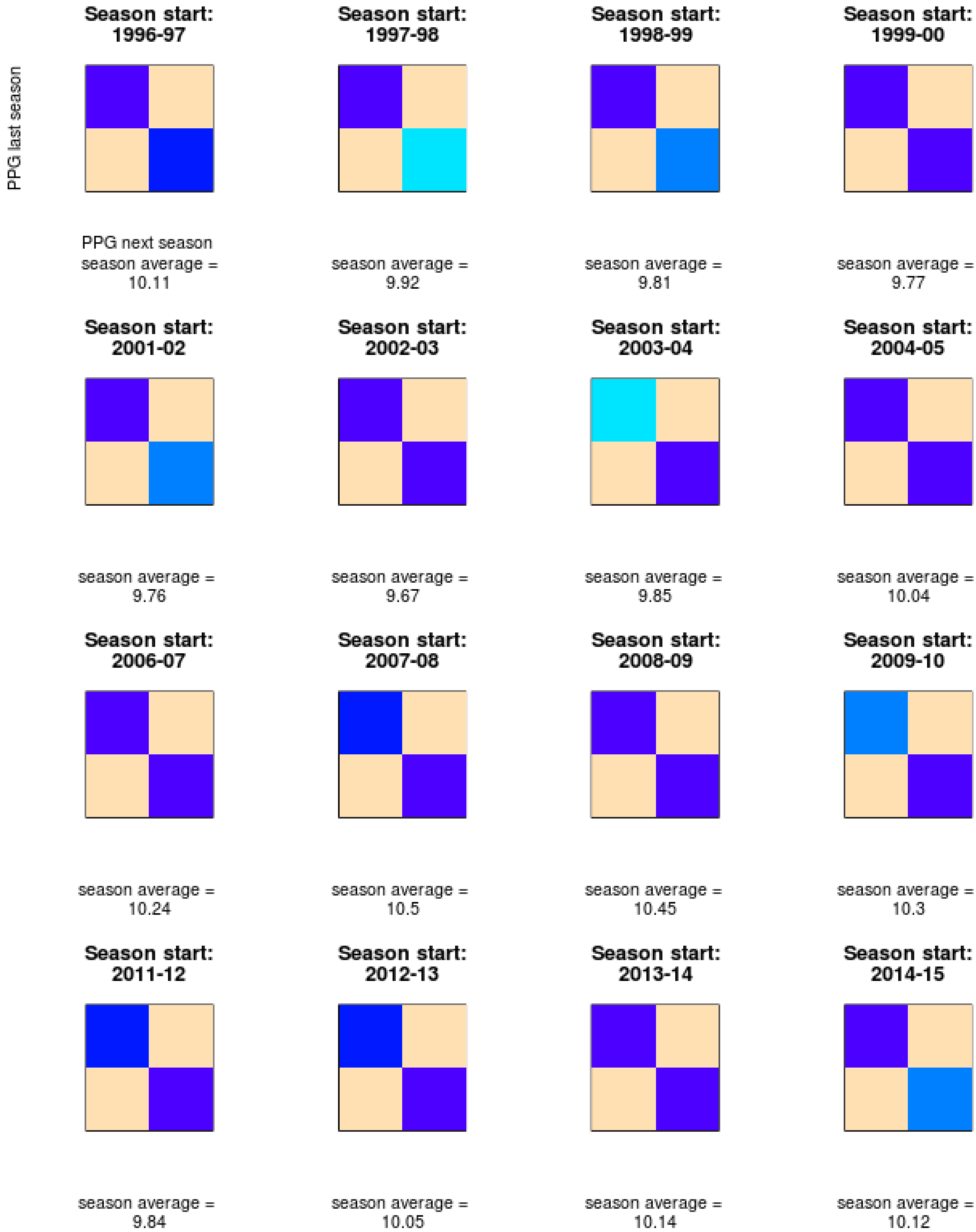
Markov Chains, as we discovered, are usually incapable of predicting temporal trends in data, since they are unchanging based on the previous results. A solution to this was to stratify the Markov Chains by season and compare them to each other. By doing this, the hope was that we could see trends in volatility based on changes in rules of the game over time.

A major problem we ran into when stratifying by seasons was that our dataset was based on active players, which is skewed towards newer seasons. Problems arise when looking at players more than five years ago, as the valid data for entry to the Markov Chains decreases rapidly. To accurately create the Markov Chains, we had to download a new dataset which also included entries of inactive players. This new dataset had 7433 entries.

Applying the model to these subsets of data, we created 20 transition matrices corresponding to each season transitioning to the next season. The volatility index was recorded for each.




There aren't too many major differences between the volatility indices, and in fact there might be too much data to even dive into why there are changes at all. Because of this, we implemented an extreme to the data: A 2x2 binary transition matrix corresponding to whether a player scores above or below the season average.




Like before, the scale is from blue to gray, where the latter corresponds to higher values. As a side note, these players show high consistency, where a player that scored above average has a high chance of scoring above average the season before. This in itself does not change, but there are some differences in some seasons corresponding to the lighter blue values. In these seasons, players on average performed better or worse than they did the season before. An example of this is in the difference between 2001-02 to 2003-04, where the model suggests a temporary improvement in player quality followed by a sudden decrease two years later.

## Conclusion

Overall, our model suggests that basketball players tend to be relatively consistent in their performance. That is, if a player scores a certain average number of points per game in one season, they are likely to score a similar average number of points per game in the following season. This makes sense, as most basketball players are pretty consistent in their overall performance as well.

Using Markov Chains to model this problem revealed some drawbacks of using such an approach to model data like this. The data that we used was heavily temporal and relied on several factors, and while the base probability of PTS changing between two seasons is valid to model, we had difficulty in expanding the model to account for factors. We had to create entirely different transition matrices just to account for small factors such as position played or whether the player is in the hall of fame. This grows exponentially when cross-compared, and since the data is finite, splitting it based on these factors can result in very small data for each matrix, which we had to be careful about when exploring the model.

Our measure of "volatility" was effective in some ways, but still had some drawbacks. We were able to get relatively consistent values for differently sized transition matrices, which allowed us to compare differently sized matrices. However, the meaning of "volatility" in this model is unclear, as we have no clear way of using the volatility index to predict any specific measure of performance. We also did not find a way to look at the volatility as a statistic, so we have no quantitative way of saying that the "true" volatility of one matrix is greater than that of another matrix.

# Appendix

## Transition Matrix f or Different Positions

| | Transition Matrix for Center Position | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 0-3 | 3-6 | 6-9 | 9-12 | 12-15 | 15-18 | 18-21 | 21+ |
| 0-3 | 0.4063 | 0.4063 | 0.1563 | 0.0313 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 3-6 | 0.1867 | 0.4267 | 0.2933 | 0.0667 | 0.0267 | 0.0000 | 0.0000 | 0.0000 |
| 6-9 | 0.0303 | 0.2273 | 0.4545 | 0.2424 | 0.0455 | 0.0000 | 0.0000 | 0.0000 |
| 9-12 | 0.0189 | 0.1132 | 0.1509 | 0.4528 | 0.1887 | 0.0566 | 0.0189 | 0.0000 |
| 12-15 | 0.0000 | 0.0800 | 0.0400 | 0.1600 | 0.4400 | 0.2000 | 0.0800 | 0.0000 |
| 15-18 | 0.0000 | 0.0000 | 0.0000 | 0.0769 | 0.3846 | 0.2308 | 0.3077 | 0.0000 |
| 18-21 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.1667 | 0.2500 | 0.5000 | 0.0833 |
| 21+ | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 1.0000 | 0.0000 |

| | Transition Matrix for Center Position after 5 seasons | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 0-3 | 3-6 | 6-9 | 9-12 | 12-15 | 15-18 | 18-21 | 21+ |
| 0-3 | 0.1268 | 0.2749 | 0.2648 | 0.1817 | 0.0958 | 0.0326 | 0.0223 | 0.0010 |
| 3-6 | 0.1134 | 0.2568 | 0.2528 | 0.1880 | 0.1122 | 0.0423 | 0.0326 | 0.0018 |
| 6-9 | 0.0930 | 0.2272 | 0.2306 | 0.1959 | 0.1399 | 0.0592 | 0.0510 | 0.0032 |
| 9-12 | 0.0696 | 0.1830 | 0.1847 | 0.1853 | 0.1822 | 0.0914 | 0.0965 | 0.0073 |
| 12-15 | 0.0468 | 0.1359 | 0.1338 | 0.1628 | 0.2235 | 0.1279 | 0.1561 | 0.0132 |
| 15-18 | 0.0282 | 0.0970 | 0.0928 | 0.1423 | 0.2527 | 0.1574 | 0.2106 | 0.0190 |
| 18-21 | 0.0170 | 0.0699 | 0.0636 | 0.1206 | 0.2672 | 0.1788 | 0.2585 | 0.0245 |
| 21+ | 0.0098 | 0.0519 | 0.0441 | 0.1044 | 0.2750 | 0.1929 | 0.2935 | 0.0284 |

| | Transition Matrix for Center Position after 10 seasons | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 0-3 | 3-6 | 6-9 | 9-12 | 12-15 | 15-18 | 18-21 | 21+ |
| 0-3 | 0.0903 | 0.2167 | 0.2150 | 0.1833 | 0.1491 | 0.0696 | 0.0708 | 0.0052 |
| 3-6 | 0.0871 | 0.2107 | 0.2091 | 0.1817 | 0.1543 | 0.0739 | 0.0774 | 0.0059 |
| 6-9 | 0.0817 | 0.2006 | 0.1990 | 0.1788 | 0.1631 | 0.0812 | 0.0886 | 0.0070 |
| 9-12 | 0.0724 | 0.1827 | 0.1808 | 0.1726 | 0.1784 | 0.0943 | 0.1096 | 0.0090 |
| 12-15 | 0.0620 | 0.1623 | 0.1599 | 0.1652 | 0.1957 | 0.1093 | 0.1341 | 0.0115 |
| 15-18 | 0.0531 | 0.1451 | 0.1423 | 0.1588 | 0.2102 | 0.1219 | 0.1549 | 0.0136 |
| 18-21 | 0.0465 | 0.1321 | 0.1289 | 0.1537 | 0.2210 | 0.1316 | 0.1709 | 0.0152 |
| 21+ | 0.0420 | 0.1232 | 0.1198 | 0.1501 | 0.2284 | 0.1382 | 0.1820 | 0.0163 |

| | Transition Matrix for Center Position after 300 seasons | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 0-3 | 3-6 | 6-9 | 9-12 | 12-15 | 15-18 | 18-21 | 21+ |
| 0-3 | 0.0710 | 0.1797 | 0.1776 | 0.1711 | 0.1808 | 0.0966 | 0.1138 | 0.0095 |
| 3-6 | 0.0710 | 0.1797 | 0.1776 | 0.1711 | 0.1808 | 0.0966 | 0.1138 | 0.0095 |
| 6-9 | 0.0710 | 0.1797 | 0.1776 | 0.1711 | 0.1808 | 0.0966 | 0.1138 | 0.0095 |
| 9-12 | 0.0710 | 0.1797 | 0.1776 | 0.1711 | 0.1808 | 0.0966 | 0.1138 | 0.0095 |
| 12-15 | 0.0710 | 0.1797 | 0.1776 | 0.1711 | 0.1808 | 0.0966 | 0.1138 | 0.0095 |
| 15-18 | 0.0710 | 0.1797 | 0.1776 | 0.1711 | 0.1808 | 0.0966 | 0.1138 | 0.0095 |
| 18-21 | 0.0710 | 0.1797 | 0.1776 | 0.1711 | 0.1808 | 0.0966 | 0.1138 | 0.0095 |
| 21+ | 0.0710 | 0.1797 | 0.1776 | 0.1711 | 0.1808 | 0.0966 | 0.1138 | 0.0095 |

| | Transition Matrix for Forward Position | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0-3 | 3-6 | 6-9 | 9-12 | 12-15 | 15-18 | 18-21 | 21-24 | 24-27 | 27+ |
| 0-3 | 0.1563 | 0.6250 | 0.1875 | 0.0313 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 3-6 | 0.1288 | 0.4318 | 0.3106 | 0.1136 | 0.0152 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 6-9 | 0.0340 | 0.2857 | 0.3401 | 0.2381 | 0.0748 | 0.0204 | 0.0068 | 0.0000 | 0.0000 | 0.0000 |
| 9-12 | 0.0000 | 0.0866 | 0.2598 | 0.3780 | 0.1811 | 0.0787 | 0.0157 | 0.0000 | 0.0000 | 0.0000 |
| 12-15 | 0.0000 | 0.0375 | 0.0500 | 0.2750 | 0.3625 | 0.2125 | 0.0625 | 0.0000 | 0.0000 | 0.0000 |
| 15-18 | 0.0000 | 0.0000 | 0.0000 | 0.1014 | 0.2609 | 0.3478 | 0.1884 | 0.0870 | 0.0145 | 0.0000 |
| 18-21 | 0.0000 | 0.0000 | 0.0250 | 0.0250 | 0.0750 | 0.3000 | 0.4750 | 0.0500 | 0.0500 | 0.0000 |
| 21-24 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.2000 | 0.1500 | 0.3500 | 0.2000 | 0.1000 |
| 24-27 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0769 | 0.0000 | 0.5385 | 0.3077 | 0.0769 |
| 27+ | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.7500 | 0.2500 |

| | Transition Matrix for Forward Position after 5 seasons | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0-3 | 3-6 | 6-9 | 9-12 | 12-15 | 15-18 | 18-21 | 21-24 | 24-27 | 27+ |
| 0-3 | 0.0588 | 0.2716 | 0.2564 | 0.2069 | 0.1071 | 0.0622 | 0.0279 | 0.0063 | 0.0025 | 0.0004 |
| 3-6 | 0.0543 | 0.2549 | 0.2463 | 0.2086 | 0.1154 | 0.0721 | 0.0348 | 0.0090 | 0.0038 | 0.0007 |
| 6-9 | 0.0461 | 0.2232 | 0.2256 | 0.2091 | 0.1299 | 0.0918 | 0.0495 | 0.0156 | 0.0075 | 0.0017 |
| 9-12 | 0.0351 | 0.1799 | 0.1954 | 0.2066 | 0.1483 | 0.1197 | 0.0713 | 0.0263 | 0.0139 | 0.0034 |
| 12-15 | 0.0233 | 0.1301 | 0.1555 | 0.1944 | 0.1644 | 0.1542 | 0.1016 | 0.0442 | 0.0255 | 0.0068 |
| 15-18 | 0.0127 | 0.0803 | 0.1081 | 0.1638 | 0.1647 | 0.1864 | 0.1342 | 0.0808 | 0.0533 | 0.0157 |
| 18-21 | 0.0088 | 0.0598 | 0.0856 | 0.1446 | 0.1610 | 0.2022 | 0.1538 | 0.0982 | 0.0662 | 0.0197 |
| 21-24 | 0.0031 | 0.0257 | 0.0428 | 0.0903 | 0.1189 | 0.1991 | 0.1501 | 0.1834 | 0.1405 | 0.0461 |
| 24-27 | 0.0014 | 0.0148 | 0.0278 | 0.0687 | 0.1008 | 0.1959 | 0.1471 | 0.2169 | 0.1701 | 0.0566 |
| 27+ | 0.0003 | 0.0052 | 0.0126 | 0.0421 | 0.0738 | 0.1857 | 0.1361 | 0.2621 | 0.2108 | 0.0713 |

| | Transition Matrix for Forward Position after 10 seasons | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0-3 | 3-6 | 6-9 | 9-12 | 12-15 | 15-18 | 18-21 | 21-24 | 24-27 | 27+ |
| 0-3 | 0.0408 | 0.2005 | 0.2063 | 0.2010 | 0.1350 | 0.1071 | 0.0634 | 0.0266 | 0.0153 | 0.0041 |
| 3-6 | 0.0397 | 0.1954 | 0.2023 | 0.1994 | 0.1361 | 0.1103 | 0.0662 | 0.0290 | 0.0170 | 0.0046 |
| 6-9 | 0.0373 | 0.1852 | 0.1938 | 0.1959 | 0.1382 | 0.1167 | 0.0719 | 0.0343 | 0.0209 | 0.0058 |
| 9-12 | 0.0338 | 0.1703 | 0.1814 | 0.1905 | 0.1409 | 0.1260 | 0.0802 | 0.0423 | 0.0268 | 0.0077 |
| 12-15 | 0.0293 | 0.1506 | 0.1647 | 0.1826 | 0.1437 | 0.1381 | 0.0911 | 0.0540 | 0.0356 | 0.0105 |
| 15-18 | 0.0236 | 0.1249 | 0.1418 | 0.1694 | 0.1445 | 0.1528 | 0.1047 | 0.0728 | 0.0502 | 0.0154 |
| 18-21 | 0.0209 | 0.1128 | 0.1309 | 0.1630 | 0.1448 | 0.1598 | 0.1112 | 0.0819 | 0.0572 | 0.0177 |
| 21-24 | 0.0141 | 0.0807 | 0.0996 | 0.1392 | 0.1381 | 0.1752 | 0.1259 | 0.1156 | 0.0847 | 0.0270 |
| 24-27 | 0.0116 | 0.0688 | 0.0879 | 0.1300 | 0.1353 | 0.1807 | 0.1312 | 0.1286 | 0.0953 | 0.0306 |
| 27+ | 0.0088 | 0.0547 | 0.0736 | 0.1181 | 0.1308 | 0.1868 | 0.1372 | 0.1455 | 0.1092 | 0.0354 |

| | Transition Matrix for Forward Position after 300 seasons | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0-3 | 3-6 | 6-9 | 9-12 | 12-15 | 15-18 | 18-21 | 21-24 | 24-27 | 27+ |
| 0-3 | 0.0301 | 0.1533 | 0.1658 | 0.1807 | 0.1403 | 0.1354 | 0.0889 | 0.0563 | 0.0379 | 0.0114 |
| 3-6 | 0.0301 | 0.1533 | 0.1658 | 0.1807 | 0.1403 | 0.1354 | 0.0889 | 0.0563 | 0.0379 | 0.0114 |
| 6-9 | 0.0301 | 0.1533 | 0.1658 | 0.1807 | 0.1403 | 0.1354 | 0.0889 | 0.0563 | 0.0379 | 0.0114 |
| 9-12 | 0.0301 | 0.1533 | 0.1658 | 0.1807 | 0.1403 | 0.1354 | 0.0889 | 0.0563 | 0.0379 | 0.0114 |
| 12-15 | 0.0301 | 0.1533 | 0.1658 | 0.1807 | 0.1403 | 0.1354 | 0.0889 | 0.0563 | 0.0379 | 0.0114 |
| 15-18 | 0.0301 | 0.1533 | 0.1658 | 0.1807 | 0.1403 | 0.1354 | 0.0889 | 0.0563 | 0.0379 | 0.0114 |
| 18-21 | 0.0301 | 0.1533 | 0.1658 | 0.1807 | 0.1403 | 0.1354 | 0.0889 | 0.0563 | 0.0379 | 0.0114 |
| 21-24 | 0.0301 | 0.1533 | 0.1658 | 0.1807 | 0.1403 | 0.1354 | 0.0889 | 0.0563 | 0.0379 | 0.0114 |
| 24-27 | 0.0301 | 0.1533 | 0.1658 | 0.1807 | 0.1403 | 0.1354 | 0.0889 | 0.0563 | 0.0379 | 0.0114 |
| 27+ | 0.0301 | 0.1533 | 0.1658 | 0.1807 | 0.1403 | 0.1354 | 0.0889 | 0.0563 | 0.0379 | 0.0114 |

| | Transition Matrix for Gaurd Position | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0-3 | 3-6 | 6-9 | 9-12 | 12-15 | 15-18 | 18-21 | 21-24 | 24-27 | 27-30 | 30+ |
| 0-3 | 0.1667 | 0.4167 | 0.2917 | 0.0833 | 0.0417 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 3-6 | 0.0963 | 0.3778 | 0.3333 | 0.1407 | 0.0519 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 6-9 | 0.0067 | 0.1946 | 0.3557 | 0.2819 | 0.1141 | 0.0336 | 0.0134 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 9-12 | 0.0000 | 0.0925 | 0.2370 | 0.3757 | 0.1965 | 0.0925 | 0.0058 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 12-15 | 0.0000 | 0.0075 | 0.0821 | 0.2537 | 0.3657 | 0.2239 | 0.0522 | 0.0149 | 0.0000 | 0.0000 | 0.0000 |
| 15-18 | 0.0000 | 0.0083 | 0.0083 | 0.0826 | 0.1901 | 0.4545 | 0.1901 | 0.0496 | 0.0165 | 0.0000 | 0.0000 |
| 18-21 | 0.0000 | 0.0000 | 0.0000 | 0.0351 | 0.0877 | 0.2632 | 0.3333 | 0.2105 | 0.0702 | 0.0000 | 0.0000 |
| 21-24 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.1600 | 0.2400 | 0.4000 | 0.0000 | 0.1200 | 0.0800 |
| 24-27 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.1000 | 0.2000 | 0.4000 | 0.2000 | 0.1000 |
| 27-30 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.2000 | 0.2000 | 0.6000 | 0.0000 |
| 30+ | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 1.0000 | 0.0000 | 0.0000 |

| | Transition Matrix for Gaurd Position after 5 seasons | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0-3 | 3-6 | 6-9 | 9-12 | 12-15 | 15-18 | 18-21 | 21-24 | 24-27 | 27-30 | 30+ |
| 0-3 | 0.0239 | 0.1538 | 0.2295 | 0.2383 | 0.1642 | 0.1186 | 0.0442 | 0.0186 | 0.0054 | 0.0023 | 0.0011 |
| 3-6 | 0.0226 | 0.1479 | 0.2235 | 0.2368 | 0.1665 | 0.1243 | 0.0475 | 0.0207 | 0.0062 | 0.0027 | 0.0013 |
| 6-9 | 0.0181 | 0.1269 | 0.2000 | 0.2271 | 0.1714 | 0.1447 | 0.0616 | 0.0310 | 0.0106 | 0.0059 | 0.0025 |
| 9-12 | 0.0151 | 0.1117 | 0.1820 | 0.2185 | 0.1744 | 0.1605 | 0.0729 | 0.0393 | 0.0140 | 0.0083 | 0.0034 |
| 12-15 | 0.0098 | 0.0816 | 0.1423 | 0.1925 | 0.1726 | 0.1888 | 0.0986 | 0.0626 | 0.0262 | 0.0185 | 0.0066 |
| 15-18 | 0.0054 | 0.0517 | 0.0972 | 0.1526 | 0.1575 | 0.2096 | 0.1281 | 0.0973 | 0.0487 | 0.0397 | 0.0122 |
| 18-21 | 0.0030 | 0.0328 | 0.0651 | 0.1154 | 0.1321 | 0.2053 | 0.1442 | 0.1312 | 0.0791 | 0.0727 | 0.0190 |
| 21-24 | 0.0013 | 0.0173 | 0.0369 | 0.0777 | 0.1006 | 0.1860 | 0.1512 | 0.1658 | 0.1165 | 0.1203 | 0.0265 |
| 24-27 | 0.0003 | 0.0061 | 0.0138 | 0.0374 | 0.0569 | 0.1391 | 0.1440 | 0.2030 | 0.1734 | 0.1892 | 0.0368 |
| 27-30 | 0.0001 | 0.0035 | 0.0081 | 0.0265 | 0.0440 | 0.1235 | 0.1371 | 0.2145 | 0.1808 | 0.2243 | 0.0376 |
| 30+ | 0.0001 | 0.0029 | 0.0069 | 0.0246 | 0.0423 | 0.1221 | 0.1418 | 0.2141 | 0.1972 | 0.2068 | 0.0412 |

| | Transition Matrix for Gaurd Position after 10 seasons | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0-3 | 3-6 | 6-9 | 9-12 | 12-15 | 15-18 | 18-21 | 21-24 | 24-27 | 27-30 | 30+ |
| 0-3 | 0.0142 | 0.1035 | 0.1677 | 0.2029 | 0.1656 | 0.1630 | 0.0817 | 0.0526 | 0.0242 | 0.0188 | 0.0058 |
| 3-6 | 0.0140 | 0.1021 | 0.1658 | 0.2014 | 0.1651 | 0.1639 | 0.0828 | 0.0540 | 0.0251 | 0.0197 | 0.0060 |
| 6-9 | 0.0130 | 0.0964 | 0.1577 | 0.1947 | 0.1626 | 0.1670 | 0.0874 | 0.0601 | 0.0296 | 0.0244 | 0.0071 |
| 9-12 | 0.0123 | 0.0921 | 0.1515 | 0.1895 | 0.1606 | 0.1693 | 0.0910 | 0.0648 | 0.0331 | 0.0280 | 0.0079 |
| 12-15 | 0.0107 | 0.0815 | 0.1362 | 0.1759 | 0.1546 | 0.1736 | 0.0992 | 0.0770 | 0.0428 | 0.0385 | 0.0100 |
| 15-18 | 0.0086 | 0.0679 | 0.1158 | 0.1568 | 0.1447 | 0.1769 | 0.1096 | 0.0944 | 0.0573 | 0.0549 | 0.0132 |
| 18-21 | 0.0068 | 0.0557 | 0.0969 | 0.1374 | 0.1332 | 0.1767 | 0.1181 | 0.1119 | 0.0733 | 0.0735 | 0.0166 |
| 21-24 | 0.0050 | 0.0430 | 0.0768 | 0.1160 | 0.1193 | 0.1743 | 0.1265 | 0.1314 | 0.0917 | 0.0956 | 0.0205 |
| 24-27 | 0.0030 | 0.0288 | 0.0537 | 0.0898 | 0.1009 | 0.1680 | 0.1349 | 0.1553 | 0.1155 | 0.1247 | 0.0254 |
| 27-30 | 0.0025 | 0.0247 | 0.0469 | 0.0819 | 0.0951 | 0.1655 | 0.1371 | 0.1625 | 0.1228 | 0.1341 | 0.0268 |
| 30+ | 0.0024 | 0.0242 | 0.0462 | 0.0813 | 0.0949 | 0.1658 | 0.1376 | 0.1631 | 0.1234 | 0.1342 | 0.0270 |

| | Transition Matrix for Gaurd Position after 300 seasons | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0-3 | 3-6 | 6-9 | 9-12 | 12-15 | 15-18 | 18-21 | 21-24 | 24-27 | 27-30 | 30+ |
| 0-3 | 0.0091 | 0.0708 | 0.1192 | 0.1577 | 0.1427 | 0.1714 | 0.1063 | 0.0936 | 0.0585 | 0.0573 | 0.0133 |
| 3-6 | 0.0091 | 0.0708 | 0.1192 | 0.1577 | 0.1427 | 0.1714 | 0.1063 | 0.0936 | 0.0585 | 0.0573 | 0.0133 |
| 6-9 | 0.0091 | 0.0708 | 0.1192 | 0.1577 | 0.1427 | 0.1714 | 0.1063 | 0.0936 | 0.0585 | 0.0573 | 0.0133 |
| 9-12 | 0.0091 | 0.0708 | 0.1192 | 0.1577 | 0.1427 | 0.1714 | 0.1063 | 0.0936 | 0.0585 | 0.0573 | 0.0133 |
| 12-15 | 0.0091 | 0.0708 | 0.1192 | 0.1577 | 0.1427 | 0.1714 | 0.1063 | 0.0936 | 0.0585 | 0.0573 | 0.0133 |
| 15-18 | 0.0091 | 0.0708 | 0.1192 | 0.1577 | 0.1427 | 0.1714 | 0.1063 | 0.0936 | 0.0585 | 0.0573 | 0.0133 |
| 18-21 | 0.0091 | 0.0708 | 0.1192 | 0.1577 | 0.1427 | 0.1714 | 0.1063 | 0.0936 | 0.0585 | 0.0573 | 0.0133 |
| 21-24 | 0.0091 | 0.0708 | 0.1192 | 0.1577 | 0.1427 | 0.1714 | 0.1063 | 0.0936 | 0.0585 | 0.0573 | 0.0133 |
| 24-27 | 0.0091 | 0.0708 | 0.1192 | 0.1577 | 0.1427 | 0.1714 | 0.1063 | 0.0936 | 0.0585 | 0.0573 | 0.0133 |
| 27-30 | 0.0091 | 0.0708 | 0.1192 | 0.1577 | 0.1427 | 0.1714 | 0.1063 | 0.0936 | 0.0585 | 0.0573 | 0.0133 |
| 30+ | 0.0091 | 0.0708 | 0.1192 | 0.1577 | 0.1427 | 0.1714 | 0.1063 | 0.0936 | 0.0585 | 0.0573 | 0.0133 |

## Hall of Fame code

```
hall_of_fame_data = read.csv("hall_of_fame_data.csv")


#hall_of_fame_no_chamberlain is the data for all Hall of Fame

#players except for Wilt Chamberlain.

hall_of_fame_data_no_chamberlain = hall_of_fame_data[-which(hall_of_fame_data$Player == "Wilt Cham

highest = max(hall_of_fame_data_no_chamberlain$PPG.current);  #Highest PPG

toplevel = ceiling(highest);  #Highest level

brackets = seq(0, toplevel);  #All brackets


tm = matrix(0, nrow=length(brackets), ncol=length(brackets));


#populate matrix

for (i in seq(1, length(hall_of_fame_data_no_chamberlain$"PPG.next"))) {

  curr = ceiling(hall_of_fame_data_no_chamberlain$"PPG.current"[i]);

  nex = ceiling(hall_of_fame_data_no_chamberlain$"PPG.next"[i]);

  tm[curr,nex] = tm[curr,nex] + 1;

  i = i+1;
```

```
}

empties = c();
#normalize
for (i in seq(1,length(brackets))) {
  if (sum(tm[i,]) > 0) {
    tm[i,] = tm[i,]/sum(tm[i,]);
  } else {
    empties = c(empties, i);
  }
}

#Delete empty levels
tm = tm[-empties,-empties];
brackets = brackets[-empties];

# volatility takes in a transition matrix and returns a value
# corresponding to the volatility
volatility <- function(A) {
  # The function for volatility used here is:
  # |i - j|*A
  scaling <- abs(row(A) - col(A))
  index <- scaling * A
  return(sum(sum(index))/length(A))
}
vol = volatility(tm);

#heatmap
image(brackets, brackets, log(0.001+tm), col=topo.colors(10), xlab="PPG this season", ylab="PPG ne
title("Hall of Fame Players: PPG Transition Matrix")
```

### Season code

group2proj.R

```
rm(list = ls())

library(plyr)
library(dplyr)

data = list()

#########################################################
## Raw data read-in

folder <- "group2csvs/"

# Sequence used
name.vals <- seq(0,3000,100)

for(ii in 1:length(name.vals)) {
  data[[ii]] <- read.csv(paste0(folder, name.vals[ii]), header = TRUE)
}

# Modify dataset (read.csv misread the header)
for(ii in 1:length(data)) {
  this <- data[[ii]]
  names(this) <- lapply(this[1,], as.character)
  this <- this[2:dim(this)[1],]
  data[[ii]] <- this
}

# Read into a big csv for later 1-line reading
big.csv <- ldply(data, data.frame)

write.csv(big.csv, paste0(folder, "dataset"))

#########################################################
```

```
### Load an initialization program

source("group2projinit.R")

##########################################################
### Begin data analysis

##########
# Standard markov chain build

# Select the dataset with relevant information
data.tr <- data %>%
  select(Player, PTS, Season)

A <- markovBuild(data.tr)
A.small <- markovBuild(data.tr, 2, mean(data.tr$PTS))

limiting.distribution <- limiting.dist(A)

test <- data.frame(x = 1:length(limiting.distribution), limiting.distribution)
actual.distribution.data <- data.tr %>%
  group_by(PTS) %>%
  count()
actual.distribution <- c(actual.distribution.data$n / sum(actual.distribution.data$n), 0, 0, 0)

distribs <- cbind(actual.distribution, limiting.distribution)
barplot(distribs, xlab = "Limiting Distribution", ylab = "frequency", beside = TRUE)


image(A)

##########
## Group by team (proof of concept, outdated (ignore))

# Remove TOT team from dataset (TOT = player switched)
```

```
data2 <- data %>%
  filter(Tm2 != "TOT")

team.num <- length(unique(data2$Tm2)) # 30 unique teams in the dataset

# Select all necessary data for analysis
data.team <- data2 %>%
  select(Player, PTS, Season, Tm2) %>%
  filter(Tm2 != "TOT") %>%
  split(data2$Tm2)

# initalize
Alist <- list()
vol <- 1
lim.dist <- list()

# Create a new A for each team, and record its volatility
for(ii in 1:length(data.team)) {
  # store A in a list under the team name
  team <- names(data.team)[ii]
  this <- data.team[[ii]]
  Alist[[team]] <- markovBuild(this, n)
  vol[ii] <- volatility(Alist[[team]])
  lim.dist[[ii]] <- limiting.dist(Alist[[team]])
}

hist(vol)

# Create mds matrix, based on the volatility index
vol.mat <- matrix(0, team.num, team.num)
row.names(vol.mat) <- names(data.team)
colnames(vol.mat) <- names(data.team)

vol.mat <- matrix(0, team.num, team.num)
```

```
for(ii in 1:team.num) {
  for(jj in 1:team.num) {
    vol.mat[ii, jj] <- vol[ii] - vol[jj]
  }
}

fit <- cmdscale(vol.mat)
x <- fit[, 1]
y <- fit[, 2]

plot(x, y, pch = 19)
text(x, y, pos = 4, cex = 0.5, labels = names(data.team))

# MDS limiting distribution
lim.mat <- matrix(0, team.num, team.num)

for(ii in 1:team.num) {
  for(jj in 1:team.num) {
    lim.mat[ii, jj] <- euc.dist(lim.dist[[ii]], lim.dist[[jj]])
  }
}

fit <- cmdscale(lim.mat)
x <- fit[, 1]
y <- fit[, 2]

plot(x, y, pch = 19)
text(x, y, pos = 4, cex = 0.5, labels = names(data.team))



###########
## Group by year

hist(as.numeric(data$Season))
```

```
data2 <- data %>%
  filter(Tm2 != "TOT")

# Select all necessary data for analysis
data.years2 <- data2 %>%
  select(Player, PTS, Season, Tm2) %>%
  filter(Tm2 != "TOT") %>%
  split(data2$Season)

data.years <- list()

for(ii in 1:(length(data.years2)-1)) {
  data.years[[ii]] <- merge(data.years2[[ii]], data.years2[[ii+1]], all = T)
  data.years[[ii]] <- data.years[[ii]] %>%
    arrange(Player, Season)
}

data.test <- data.years[[18]] %>%
  group_by(Player) %>%
  filter(n() > 1)

# initalize
Alist <- list()
Alist.small <- list()
vol <- 1
vol.small <- 1
lim.dist <- list()

# Create a new A for each team, and record various indices
for(ii in 1:length(data.years)) {
  # store A in a list under the team name
  this <- data.years[[ii]]
  Alist[[ii]] <- markovBuild(this, 12)
  Alist.small[[ii]] <- markovBuild(this, 2, mean(this$PTS))
```

```
  vol[ii] <- volatility(Alist[[ii]])
  vol.small[ii] <- volatility(Alist.small[[ii]])
  lim.dist[[ii]] <- limiting.dist(Alist[[ii]])
  seas.average <- mean(this$PTS)
}

hist(vol)

par(mfrow = c(4,5))
for(ii in 1:20) {
  xlab = c("PPG next season \n season average = ", round(mean(data.years[[ii]]$PTS),2))
  ylab = "PPG last season"
  if(ii > 1) {
    xlab = c("season average = ", round(mean(data.years[[ii]]$PTS),2))
    ylab = ""
  }

  image(log(0.001 + Alist.small[[ii]]), col = topo.colors(10), xlab = xlab, ylab = ylab, xaxt = 'n

  title(c("Season start:", levels(data$Season)[ii]))
}


par(mfrow = c(4,5))
for(ii in 1:20) {
  xlab = c("PPG next season \n volatility = ", round(vol[ii], 3))
  ylab = "PPG last season"
  if(ii > 1) {
    xlab = c("volatility = ", round(vol[ii], 3))
    ylab = ""
  }
  image(log(0.001 + Alist[[ii]]), col = topo.colors(10), xlab = xlab, ylab = ylab, xaxt = 'n', yax

  title(c("Season start:", levels(data$Season)[ii]))
}
```

```
image(brackets, brackets, log(0.001+tm), col=topo.colors(10), xlab="PPG last season", ylab="PPG ne
```

```
data.num <- length(data.years)

# Create mds matrix, based on the volatility index
vol.mat <- matrix(0, data.num, data.num)
row.names(vol.mat) <- names(data.years)
colnames(vol.mat) <- names(data.years)

vol.mat <- matrix(0, data.num, data.num)

for(ii in 1:data.num) {
  for(jj in 1:data.num) {
    vol.mat[ii, jj] <- vol[ii] - vol[jj]
  }
}

fit <- cmdscale(vol.mat)
x <- fit[, 1]
y <- fit[, 2]

plot(x, y, pch = 19)
text(x, y, pos = 4, cex = 0.5, labels = 1:18)

# limiting distribution matrix
lim.mat <- matrix(0, team.num, team.num)

for(ii in 1:data.num) {
  for(jj in 1:data.num) {
    lim.mat[ii, jj] <- euc.dist(lim.dist[[ii]], lim.dist[[jj]])
  }
```

```
}

fit <- cmdscale(lim.mat)
x <- fit[, 1]
y <- fit[, 2]

plot(x, y, pch = 19)
text(x, y, pos = 4, cex = 0.5, labels = levels(data.tr$Season))
```

group2projinit.R

```
##############
# Load this every time you start working on the code

#########################################################
### START HERE: Use this once the folder has "dataset"

rm(list = ls())

library(plyr)
library(dplyr)
library(expm)

data = list()

folder <- "group2csvs/"
data <- read.csv(paste0(folder, "All_data.csv"), header = TRUE)

# Convert data (factors tend to mess with the data)
data$Player <- as.character(data$Player)
data$PTS <- as.numeric(as.character(data$PTS))
data$PTS <- ceiling(data$PTS)
data$Tm <- as.character(data$Tm)

# Some teams change over time.
data$Tm2 <- data$Tm
```

```
data$Tm2[data$Tm2 == "NJN"] <- "BRK"
data$Tm2[data$Tm2 == "CHA"] <- "CHO"
data$Tm2[data$Tm2 == "NOK"] <- "NOP"
data$Tm2[data$Tm2 == "NOH"] <- "NOP"
data$Tm2[data$Tm2 == "SEA"] <- "OKC"

# set the superstar point, one which nobody can top:
n <- max(data$PTS)

##########################################################
### Function building

# Customfunction to get the right substring of data
# Used in markovBuild to see the last two digits of season ("09", "07", etc)
substrRight <- function(x, n){
  substr(x, nchar(x)-n+1, nchar(x))
}

# Used in markovBuild to test if the two rows are 1 season apart,
# and season 2 is season 1 + 1
validSeason <- function(row1, row2) {
  season1 <- as.numeric(row1$Season) # as.numeric(substrRight(as.character(row1$Season), 2))
  season2 <- as.numeric(row2$Season) # as.numeric(substrRight(as.character(row2$Season), 2))
  abs(season2 - season1) == 1
}

# This function builds the markov chain matrix
# This is used later in order to generate different transition matrices
# when we start to stratify by age, team, hall of fame, etc.
markovBuild <- function(data, size = n, weighting = 0) {
  # Build A matrix
  # A = matrix(0, nrow = n, ncol = n)
  n.div <- ceiling(n / size)
  A = matrix(0, nrow = round(n/n.div), ncol = round(n/n.div))
```

```
num <- 0
# Iterate through the data 1-by-1
for(ii in 1:(dim(data)[1]-1)) {
  row1 <- data[ii,]
  row2 <- data[ii+1,]

  # Bools for valid entry to matrix
  samePlayer <- row1$Player == row2$Player
  isValid <- validSeason(row1, row2)

  if(samePlayer && isValid) {
    # A[row1$PTS, row2$PTS] <- A[row1$PTS, row2$PTS] + 1
    if(weighting == 0) {
      x <- round(row1$PTS/n.div)
      y <- round(row2$PTS/n.div)
    } else {
      x <- (row1$PTS > weighting) + 1
      y <- (row2$PTS > weighting) + 1
    }

    A[x,y] <- A[x,y] + 1
    num <- num + 1
  }
}

print(num)

# Data validation
# image(A)
# sum(sum(A))

# Divide column sums to get proportional values
sums <- colSums(t(A))

A2 <- A/sums
```

```
  # Check values
  sums2 <- colSums(t(A2))

  # Create completely absorbing states for data we don't know about
  # (1 on the diagonal, 0 else)
  A2[is.nan(sums2)] <- 0
  A2[is.nan(sums2), is.nan(sums2)] <- diag(sum(is.nan(sums2)))#diag(dim(A2[is.nan(sums2),is.nan(su
  return(A2)
}

# volatility takes in a transition matrix and returns a value
# corresponding to the volatility
volatility <- function(A) {
  # The function for volatility used here is:
  # |i - j|*A
  scaling <- abs(row(A) - col(A))
  index <- scaling * A
  return(mean(mean(index)))
}

# Estimates the limiting distribution
# Note that this does not ALWAYS find it,
# It just makes a good estimate
limiting.dist <- function(A) {
  A2 <- A%^%2000000000
  A2[A2 == 1] <- 0
  colMeans(A2)
}

# Euclidean distance between two points in multidimensional space
euc.dist <- function(x1, x2) sqrt(sum((x1 - x2) ^ 2))
```

## Main Model Code

```
//For main model
clear all;
close all;
clc;
M = csvread('1-750.csv');
[col,row] = size(M);
result = zeros(30);
for i = 1 : col
    m = M(i,1);
    n = M(i,2);
    j = 0;
    k = 0;
    while(m > j && j < 30)
        j = j + 1;
    end
    m = j;
    while(n > k && k < 30)
        k = k + 1;
    end
    n = k;
    result(m,n) = result(m,n) + 1;
end

view(2);
resultPer = [];
for i = 1 : 30
    for j = 1 : 30
        if(sum(result(i,:)) ~= 0)
            resultPer(i,j) = result(i,j)/sum(result(i,:));
        else
            resultPer(i,j) = 0;
        end
    end
end
```

```
fiveSeasons  = resultPer^5;
tenSeasons  = resultPer^10;
wholeDistribution  = resultPer^1000;

vol = sum(sum(volatility));
surf(fiveSeasons,'EdgeColor','None');
view(2);
plot(wholeDistribution(1,:));
legend('After 300 Seasons');
```

## References


[1] Wheeler, Kevin. *"Predicting NBA Player Performance."*. N.p., 2012. Web. 8 Mar. 2017.

[2] Piette, James, Lisa Pham, and Sathyanarayan Anand. *"Evaluating Basketball Player Performance via Statistical Network Modeling."*. N.p., 2011. Web. 8 Mar. 2017. Philadelphia, PA, USA, 19104

[3] Kenny Shirley *"A Markov Model for Basketball."*. N.p., 2007. Web.8 Mar.2017

[4] Basketball Reference *http://www.basketball-reference.com/*