Regress, Cluster, Classify:
A Framework for Domain-to-Domain Prediction

Cole Fisher

4490 Computer Science Thesis

December 2017

Western University, London, ON

Supervisor: Dr. Dan Lizotte

Abstract

The NBA is an ultra-competitive league and the difference between winning and losing is often miniscule. NBA front office executives are tasked with making important and difficult decisions regarding team building. This team-building puzzle is awfully difficult because it is extremely tough to predict how prospects will fare in the NBA. The NBA owns and operates their own minor league called the G League and it consists of players not quite able to make NBA rosters. However, there are certainly diamonds in the rough; NBA-caliber players are toiling in the G League, awaiting the right opportunity to shine. But, predicting which G League players are NBA-ready and which are fool's gold has proven quite challenging. To help NBA teams evaluate the G League player pool, I have created a framework that can aid traditional scouting in predicting which G League players have NBA-level talent. The model is essentially comprised of three steps:

Regress: Analyze the relationship between G League stats and NBA stats and filter for the stats that are predictive.

Cluster: Use these predictive stats to cluster past call-ups into G League roles and NBA roles.

Classify: Label new prospects using the cluster results and make predictions based on:

- the G League role classification
- k neighbors that share similar G League stats

Introduction

The G League is made up of players one level below the NBA, and NBA teams own and operate G League teams for two primary reasons:

1. To send their young players down to get more playing time than they would otherwise receive if they remained on the NBA team's 15-man roster.
2. To be exposed to more young players who, for whatever reason, did not get drafted to an NBA team but might have the talent to contribute to one.

My research focuses on the second reason; there are plenty of 'diamonds in the rough' - players who fell through the cracks for different reasons. They are playing in the NBA's G League, but talent-wise they should really be on an NBA team. Every year, a handful of players get called up from the G League to the NBA and they excel, earning a permanent spot on an NBA roster in the process. With that said, I wondered if any G League stats are somewhat predictive of potential future NBA success. In other words, do the G League players who make it to the NBA have anything in common? Are they all fantastic shooters, defenders, scorers, or playmakers? NBA scouts will be the first to acknowledge that, unfortunately, scouting the G League is not that simple. Players become NBA contributors for an array of reasons. NBA teams sometimes scout with a specific purpose in mind – perhaps they are looking to fill a precise role, like a lockdown defender. As well, the player-team fit matters; a player can really struggle on one NBA team but excel on another. In basketball and more generally in sports, things often turn out differently in real life from how they originally appear on paper.

To reiterate, the G League undoubtedly houses players who are capable of contributing to NBA teams, but scouts have not been able to evaluate the G League with any consistent success. So, I wonder if a data-driven approach might help in this regard. I have come up with a framework that respects the complexity of the issue at hand. When we look at successful G League-to-NBA call-ups, we realize that they possess a wide range of skills. Some of them are great scorers, some great passers, some are rebounders, and some seem to do it all. Evidently, one or two stats is not going to paint the whole picture. Once I came to such a realization, I knew I needed a model that respected the fact that we cannot boil the problem down into one or two stats.

The Regress, Cluster, Classify (RCC) Framework

- Perform linear regression on all features present in both domain 1 and domain 2. Select the features where domain 1's feature somewhat predicts domain 2's feature.
  In the case of G League → NBA predictions, I was able to identify the following stats:
  - *Rebounds per 36 minutes (REB_36) - R-squared:0.79, p-value<0.0001*
  - *Field goal percentage (FG%) - R-squared:0.49, p-value<0.0001*
  - *Assists per 36 minutes (AST_36) – R-squared:0.79, p-value<0.0001*
  - *Points per 36 minutes (PTS_36) – R-squared:0.41, p-value<0.0001*
- Using the "predictive" stats from step 1, cluster the objects into domain 1 roles and domain 2 roles. This step requires domain expertise. To label a cluster, one must be profoundly familiar with the domain. After examining the clustering algorithm's results, clean the data by relabelling some incorrectly clustered objects, if applicable; clustering is rarely perfect so using domain expertise, ensure that all the players are properly labeled.
  In the case of G League → NBA predictions, I came up with the following domain labels:

G League Roles	***NBA Roles***
Inside Scoring Bigs	*Non-Shooting Rebounders*
High Usage Scoring Bigs	*Swiss Army Knives*
Low Usage Defenders	*Low Usage 3+Ds*
Sweet-shooting bucket getters	*Scoring Guards*
Swiss Army Knives	*Facilitators*
Facilitators	

- Make predictions on new objects:
  - Classify new object by assigning domain 1 label, and then look at the flow of domain 1 clusters → domain 2 clusters to predict likely outcomes for object in question.
  - Analyze similarity to k neighbors in the training set to see how the k neighbors fared in domain 2.

  In the case of G League → NBA predictions, the idea is to classify the new object's G League label. Then, look at how players with that G League label typically fared in the NBA. As well, look at k neighbors to see how specific players turned out. In NBA scouting, player comparisons are a very popular technique. Teams often use player comparisons as a way of evaluation and prediction.

Background

The rise of analytics in sports is somewhat controversial, and it has not always been well-received. In the NBA, some traditional scouts oppose analytics because they resent the idea that somebody behind a computer screen can outperform them at their own craft. However, this resentment is founded on a severe misconception that analytics is here to replace scouting. This is simply not true. Alongside scouting, it merely provides another leg to stand on. In a league where every single team wants to win, a team needs all the advantages it can get, and analytics is just another potential edge. Basketball analytics staff will be the first to say that analytics is not the be-all and end-all of basketball decision making. They acknowledge the value of eye-test scouting: they watch a lot of game film and they appreciate the fact that the numbers do not hold all the answers.

I think the two main contributions analytics has offered to basketball thus far is the measurement of things previously considered

unquantifiable like defense and hustle, and the postulation that the three pointer is the most productive shot in basketball.

One of the great analytic achievements in basketball has been player and ball tracking. In 2005, two Israeli scientists started SportVU, a camera system hung from the rafters that can track and collect player and ball movement at a rate of 25 frames per second. SportVU got acquired by STATS LLC in 2008, and they eventually became the official tracking partner of the NBA. This revolutionized basketball analytics, because it allows teams the opportunity to quantify what was previously thought to be immeasurable. Until player tracking emerged, basketball executives had their opinion regarding which players played good defense, but they didn't have wholesome data to back up their claim. With player tracking, it became possible to isolate the performance of a single player.

Before SportVU, guard Tony Allen of the Memphis Grizzlies might have been considered a good defender because when he was on the court his team gave up less points then when he was off the court. This is too naive of an explanation: perhaps Allen always played with other good defenders. However, SportVU data can isolate Allen. Perhaps it shows that when Tony Allen guards someone, they shoot a worse percentage than what they normally shoot on average. Or perhaps the average distance between Tony Allen and the player he is guarding is far less than the typical average distance between an offensive player and his defensive counterpart. In all, SportVU's movement tracking transformed basketball analytics because it allows for new ways of measurement.

As well, many NBA teams have tailored their offensive approach to fit their analytical findings. The Golden State Warriors and the Houston Rockets shoot far more three pointers than the average NBA team because they have discovered, using analytics, that this will give them the best chance to win. The 2016-2017 Champion Golden State Warriors shot 55.7% from two-point range and 38.3% from three-point range during the regular season[1]. Thus, their expected shot value for a two pointer was 1.114 points per shot (0.557 * 2), while their expected shot value for a three pointer was 1.149 points per shot (0.383 * 3). This explains their philosophy – if you have a good three-point attempt, take it.

The NBA as an organization is extremely progressive when it comes to the acceptance and widespread usage of analytics. They boast an exceptionally comprehensive official stats website[2], equipped with the SportVU player tracking data for each game. As well, their stats website contains commonly used analytical metrics so that the average fan can familiarize himself with the latest buzzwords. Essentially, the NBA provides its fans with a very complete dataset. This publically available data has had a tremendous impact on the quick ascension of basketball analytics because there is no huge barrier to entry; anybody is free to play around with the data.

Another key player in the rapid growth of basketball analytics is the MIT Sloan Sports Analytics. Founded in 2007 by Houston Rockets general manager Daryl Morey and Kraft Analytics Group CEO Jessica Gelman, the annual conference has been a vital source for analytics achievement over the last decade. The weekend conference offers roundtable debates featuring panels of athletes and high-ranking sports and business executives. As well, selected students are chosen to present their research papers to an impressive crowd of industry experts. Some of the papers presented during the conference have become very influential in the basket-

[1] http://stats.nba.com/team/#!/1610612744/shooting/

[2] http://stats.nba.com/

ball analytics community. For example, a paper titled "To Crash or Not To Crash"[3] presented at the 2013 conference explored the pros and cons of offensive rebounding. When a player on a team shoots the ball, all the team's players have a couple seconds to make an important decision: they can try to rebound the ball in case of a miss so that they can gain another possession, or they can retreat to set up their defense. Most teams choose the latter: many coaches want their teams to get back on defense at all costs. The paper argues that this is not the ideal strategy. They use player tracking data – player movement during the short period of time between the shot release and the rebound – to quantify a team's decision: rebound or retreat. They conclude that "focusing on the offensive rebound immediately after the shot goes up seems to trump the gain a team gets with a head start on getting back."[4] This paper challenged conventional wisdom and won; coaches preach getting back on defense but the data suggests that perhaps they are leaving points on the board.

Basketball analytics in general is quickly becoming prevalent, and certain analytical metrics are becoming commonplace amongst casual NBA fans. The NBA embraces this reality, exemplified by its offering of an impressive selection of useful data to the public. In turn, the NBA is being rewarded with extraordinary, breakthrough projects that are altering the way the game is both comprehended and played.

On the other hand, analytics concerning G League evaluation is undoubtedly behind the times. NBA teams do not emphasize G League scouting to the same extent as NCAA scouting, so NBA analytics staff rarely work on G League-oriented projects. The lack of interest in the G League is due to a number of reasons, but most notable is the G league's novelty. By 2019, almost every NBA team (30 teams) will have its own G League affiliate team that it can operate itself, but as recently as 2013, there were only 16 teams in the G League, meaning that NBA teams had to share the G League franchises amongst themselves, understandably a less than ideal situation.

Nowadays, the NBA is invested in the G League: Gatorade is on board as the lead sponsor (note the G in "G League") and the number of players called up to the NBA from the G League grows each year. Further, in 2017, the NBA added two roster spots to each NBA team strictly for G League players, dubbed 'two-way' roster spots. These two-way players will spend half their time with the NBA club and half their time in the G League. As the league continues to pump the tires of its development league, analytics staff will need to adjust. Soon, before long, the G League will be a legitimate source of NBA talent, and NBA teams will hire analytics staff to work solely on G League evaluation. Until then, there is not much for me to base my work on.

Reports for the Toronto Raptors

Below are two reports I wrote for the Raptors' assistant general manager. The reports showcase the process I went through in coming up with the RCC framework. It focusses mainly on the 'regress' part of the framework, breaking down individual stats that translate well from the G League to the NBA. In the second report, I work on the "cluster" component of the framework, and then I create an alluvial diagram that illustrates the role-transformation flow from the G League to the NBA. This part is important: when we classify an unseen player as a certain G league role, we can refer back to the diagram

[3] Wiens, Jenna, et al. "To Crash or Not To Crash: A quantitative look at the relationship between offensive rebounding and transition defense in the NBA". MIT Sloan Sports Analytics Conference 2013. Aug 2017.

[4] Wiens, Jenna, et al.

to see what role-transformations are probable. For example, if a given G League player is classified according to the RCC model as a G League facilitator, the alluvial diagram shows that most likely he will become a facilitator in the NBA domain if he gets the opportunity.

Cole Fisher – Sep 2017

G League-to-NBA Analytical Insights

Introduction

I have built an excel database of all G League call-ups' stats over the last 8 seasons in both the G League and in the NBA so that we can analyze the data to look for patterns. Comparing a player's G League stats and his NBA stats may allow us to predict the future NBA performance of intriguing G League players moving forward. Unfortunately, because NBA rosters are small and only a fraction of a team's players even receive significant playing time, the sample size is limited. There are only about 120 players who have played at least 25 games in the G League and 25 games in the NBA over the last several years. However, even though the sample size is small, I was able to find some interesting relationships.

Rebounding is Predictable

There is a strong relationship[1] between a player's G League rebounding ability per 36 minutes and his NBA rebounding ability per 36, for both offensive (r-squared: 0.78) and defensive (r-squared: 0.65) rebounding. This is a very important realization: The G League is a very safe, low-risk place to find a rebounder if that is what an NBA team is looking for.

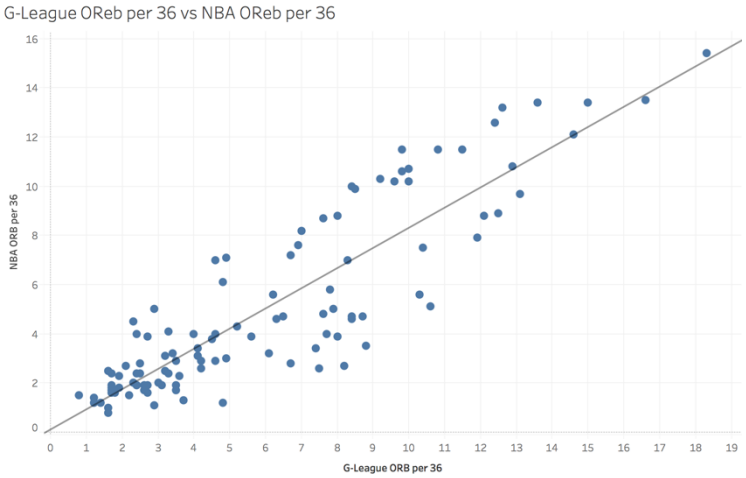

*Figure 1: NBA OReb per 36 = 0.819974 * G League OReb per 36 + 0.111242*

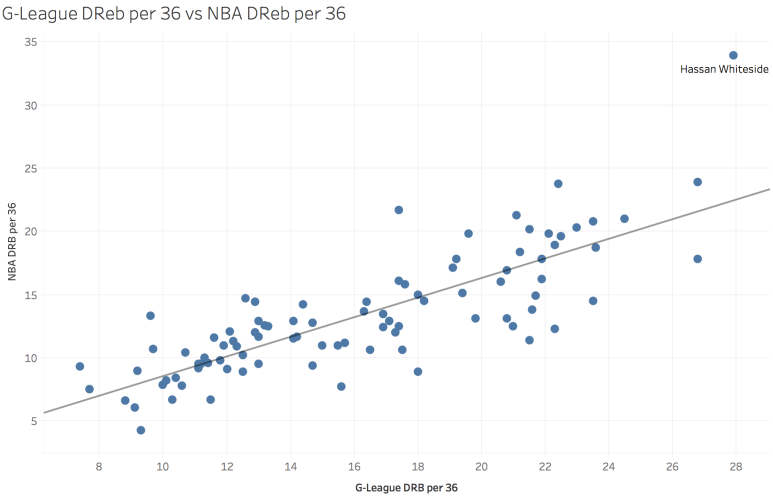

*Figure 2: NBA DReb per 36 = 0.77555 * G League DReb per 36 + 0.800065*

Of course, many skills will not translate well from the G League to the NBA as the competition and talent level increases, but it is noteworthy that rebounding absolutely does.

The big takeaway: We can be confident that a G League player's rebounding ability will translate to the NBA.

[1] minimum 25 games played in the G League and NBA

Lorenzo Brown's Shooting Will Correct

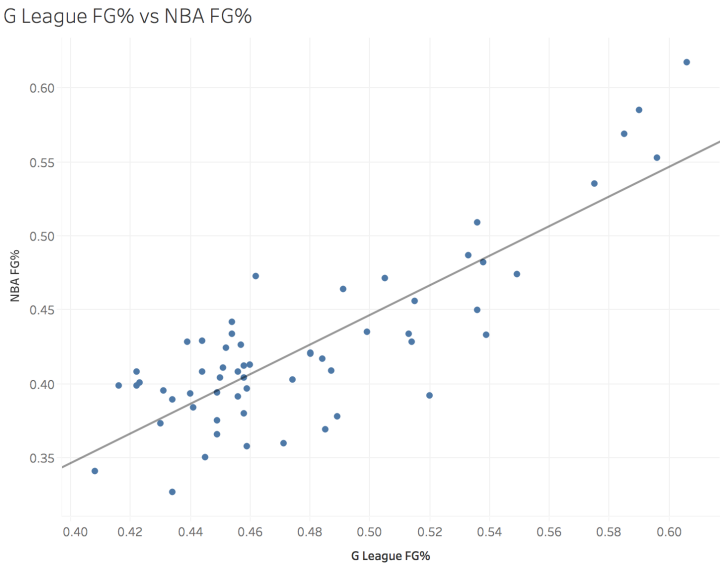

*Figure 3: NBA FG% = 1.00148 * G League FG% - 0.0543846*

There is a strong positive relationship[2] between G League FG% and NBA FG% (r-squared: 0.70), as well as a (less) strong positive relationship[3] between G League points per 36 minutes and NBA points per 36 minutes (r-squared: 0.55) for guards and wings. So, guard/wing scoring does seem to translate from the G League to the NBA in a linear fashion. This is reassuring as it pertains to Lorenzo Brown. He has not scored well at the NBA level so far, but he only has 63 games of NBA experience under his belt. He is a good scorer in the G League, so we can safely assume that more NBA experience will result in a rise in Brown's shooting ability, and figure 4 shows us why. The outliers in figure 4 tend to be smaller in circle size, which represents the number of NBA games played, so we can expect Lorenzo Brown to score around 11-15 points per 36 minutes at the NBA level once he plays enough games, thereby getting more comfortable in the NBA. Where Brown will experience the biggest jump in NBA shooting is

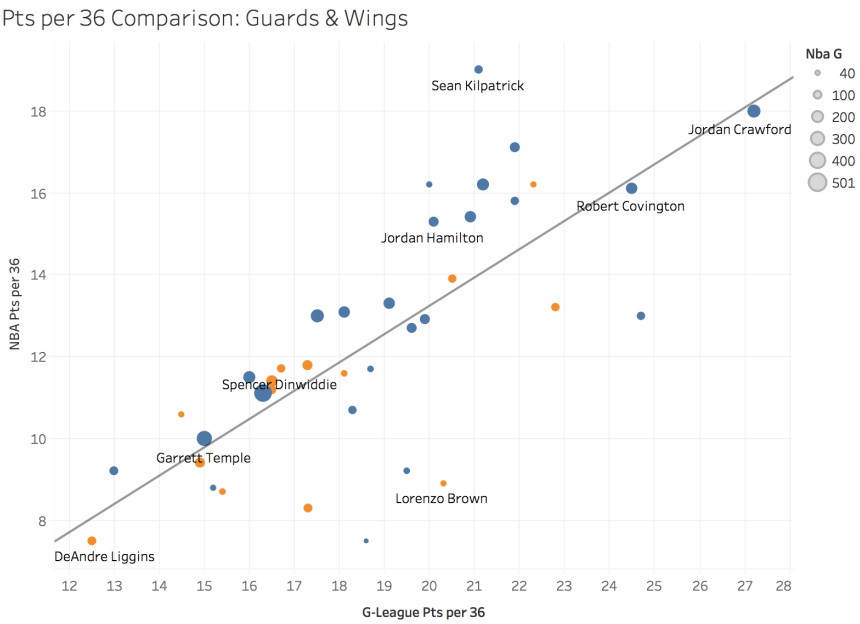

*Figure 4: NBA Pts per 36 = 0.691329 * G League Pts per 36 - 0.587897*

[2] minimum of 20 G League games played and 220 FGA in the NBA
[3] minimum of 30 G League games, 40 NBA games, and 500 NBA minutes

his three-point shot. There is not a strong linear relationship between G League 3FG% and NBA 3FG% like there is will FG% and rebounding, but figure 5[4] does indicate the notion that most solid G League three point shooters will shoot a minimum of 28% at the NBA level, and more likely around 33%. The extent to which Lorenzo Brown has struggled with his NBA three-pointer looks abysmal, but he is merely 10 of 66 in his short NBA career, so we should not yet be alarmed. Figure 5 suggests that we should expect a slight drop off in 3FG% when a player moves from the G League to the NBA, but for the most part, players who shoot 35%-38% in the G League typically shoot between 28% and 37% in the NBA. This is good news for Lorenzo Brown and the Raptors. He is a solid 36% three-point shooter in the G League, so we can expect him - worst-case scenario - to shoot about 28% in the NBA. Likely, he will shoot around 33% this season, which would qualify him as a passable NBA three point shooter.

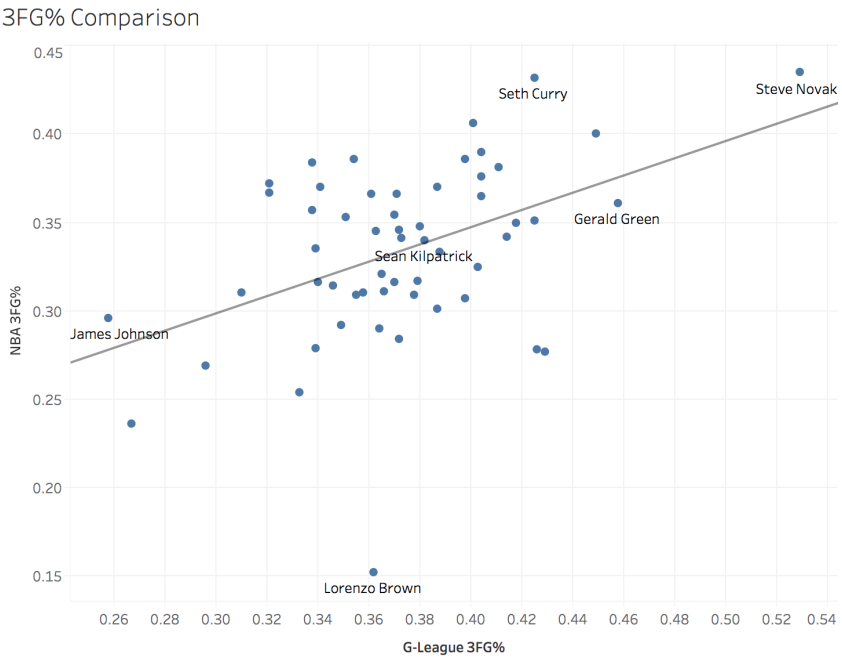

*Figure 5: NBA 3FG% = 0.487892 * G League 3FG% + 0.151994*

The big takeaway: There seems to exist a G League-to-NBA relationship between both shooting and between scoring, so as Lorenzo Brown gains more NBA action, we can expect his shooting to improve drastically. Consequently, this will improve his scoring numbers.

[4] minimum of 60 3FGA in the G League and 65 3FGA in the NBA

Predicting NBA Success from G League Offense Contributions: Positions Matter

I grouped players into 4 categories – guards (orange), wings (blue), shooting big men (turquoise), and non-shooting big men (red). Then I created charts[5] plotting their G League Offensive Rating and their G League usage %. To predict if a G League player will find success at the NBA level (measured in the following figures as minutes/game) given his G League offensive contribution, we must account for what type of player he is.

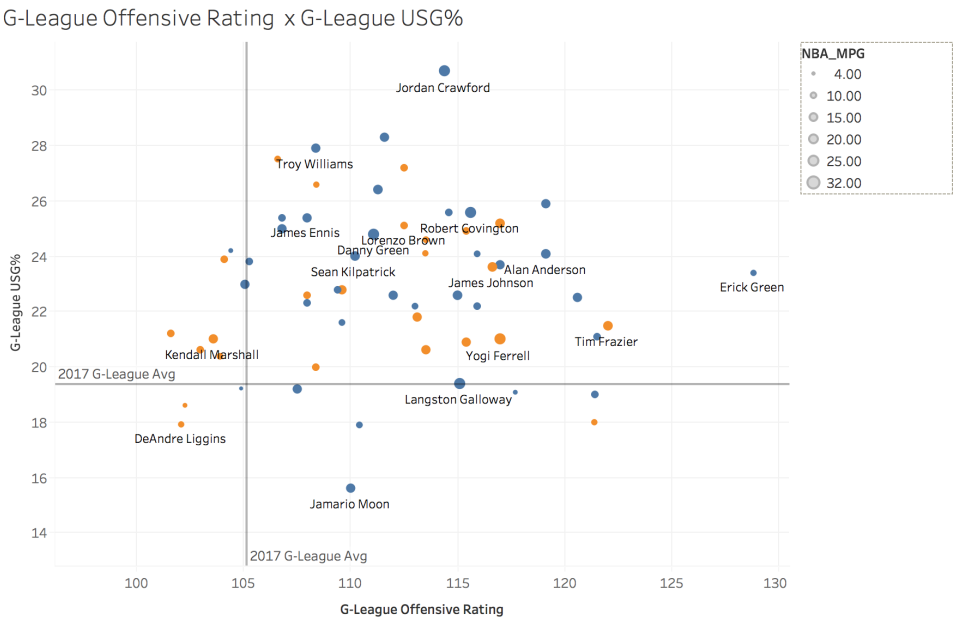

Figure 6: Orange=Guard, Blue=Wing

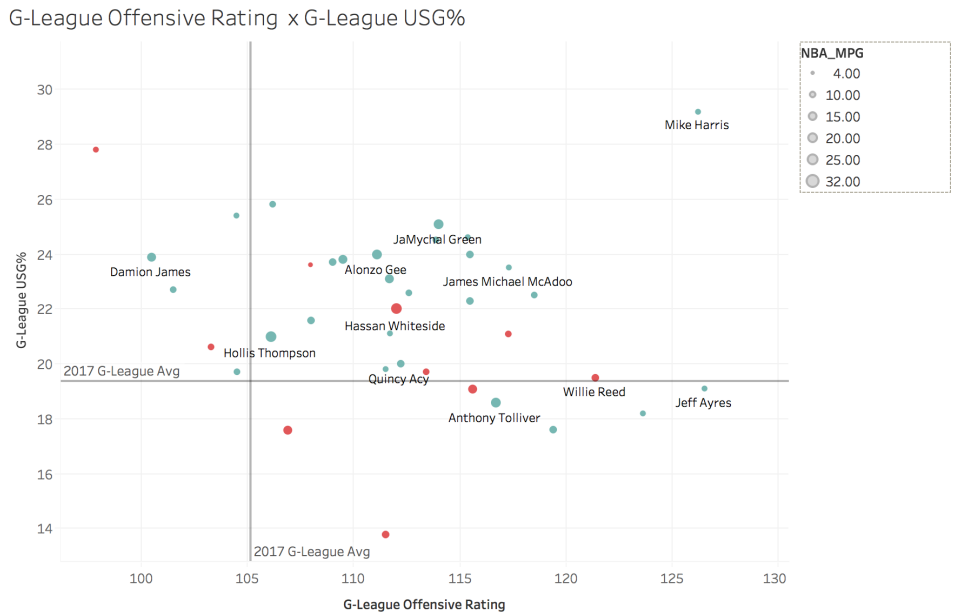

Figure 7: Turquoise=Shooting Big, Red=Non-Shooting Big

[5] Both figure 6 and figure 7 require a minimum of 18 G League games and 30 NBA games

The results are interesting: a guard/wing almost certainly needs to be the focal point of his G League offense in order to have success at the NBA level (as defined as minutes/game). Most successful G League guard/wing call-ups to the NBA have an above average[6] G League offensive rating and above average G League usage. On the other hand, successful NBA big men do not necessarily need to pose such gaudy G League offensive numbers: the players in figure 7 are more spread out, implying that NBA success can be achieved without an above average offensive rating and usage rate. Intuitively this makes sense: NBA teams typically rely more on their guards and wings for offensive production than they do on their big men. Therefore, big men can become successful NBA contributors without being offensive focal points. Thus, their G League offensive rating and usage rate is not indicative of future NBA success. **The big takeaway: A guard/wing probably needs to have an above average G League offensive rating and usage rate to succeed in the NBA, whereas a big man does not necessarily need to be above average in both respects to find success in the NBA.**

Summary

- G League rebounding ability is predictive of NBA rebounding ability.
    - Good G League rebounders will be good NBA rebounders.
- Lorenzo Brown's three-point shooting will improve which, in turn, will improve his overall NBA scoring ability.
- A guard/wing player needs to dominate offensively in the G League to have future NBA success, but a big man does not.

[6] as it relates to the 2017 G League average

Cole Fisher – Nov 2017

G League-to-NBA Analytical Insights

Introduction

In this edition, I have broken down players into four G League positions: guard, wing, non-shooting big and shooting big. This allows us to examine relationships by position, instead of simply overall. I will look at a few more G League → NBA single stat relationships. Then, after assigning all players to G League and NBA roles, we will try to see if certain G League labels typically result in certain NBA labels. Lastly, I will analyze NBA call-ups' G League minutes to see if we can uncover anything interesting - is it ever too early or too late to make a player prediction?

Assist % is Very Predictable

There is a very strong linear relationship between a call-up's assist % in the G League and his assist % in the NBA, as illustrated in figure 1. Interestingly, a player's G League position does not seem to effect whether or not his assist % will translate to the NBA. Sure, guards and wings have higher assist % than big men, but the translation of G League assist % to NBA assist % is pretty obvious for each of the four groups.

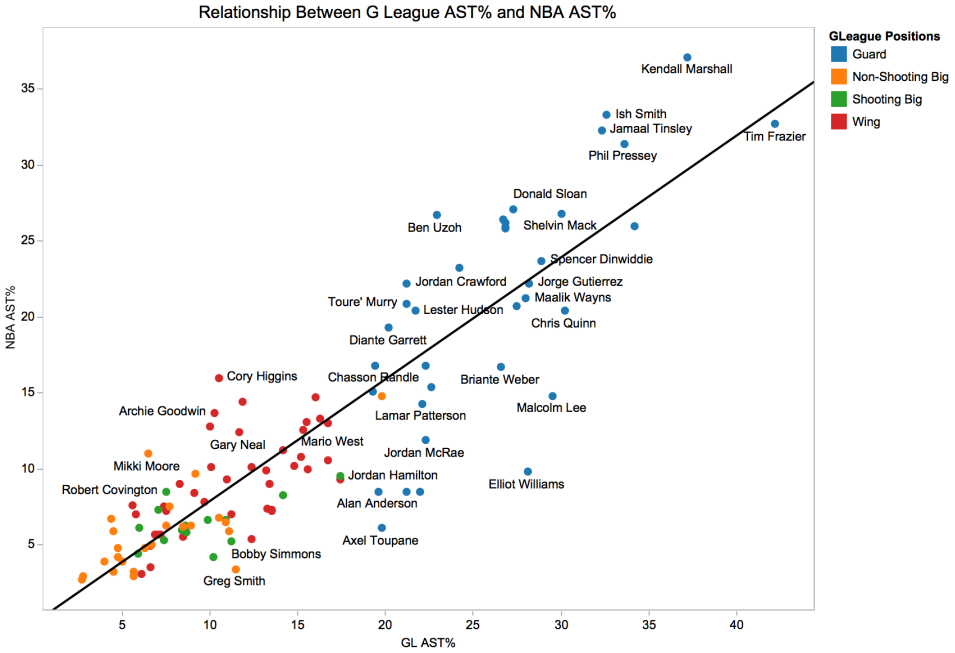

Figure 1: G League - to - NBA Assist %

Fouling in the G League = Fouling in the NBA

There is a pretty strong relationship between G League fouls per 36 minutes and NBA fouls per 36 minutes. If a G League big man cannot protect the paint without fouling in the G League, there is a very strong chance this trend continues in the NBA. Dexter Pittman is the best

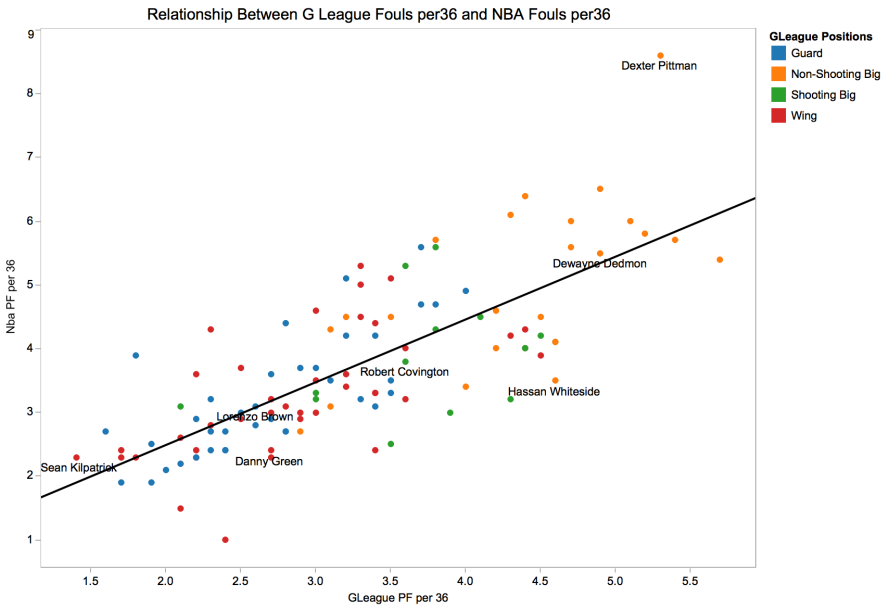

Figure 2: G League – to – NBA fouls per 36 min

example of this – he fouls a ton in both the G League and the NBA. Figure 2 illustrates the strong correlation.

No Linear Relationship between G League and NBA Turnovers

It appears as though there is not a strong relationship for turnovers from the G League to the NBA. A G League guard with an extremely high usage might be a turnover-machine in the G League, but we cannot confidently predict his turnover numbers at the NBA level. I estimate that this lack of a relationship is the product of G League → NBA role-transformation. What I mean is that a player might be asked to be a G League team's first offensive option, but at the NBA level, his role might become lock-down defender.

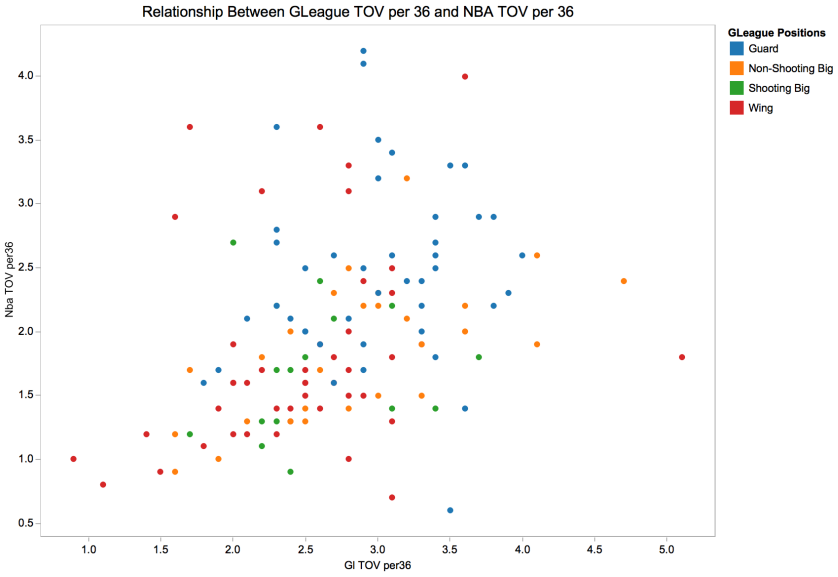

Figure 3: G League – to – NBA Turnovers per 36

G League → NBA Role-Transformation

To test this theory of role-transformation, I clustered the call-ups based on their G League stats and their NBA stats. The clustering algorithm grouped players together based on their common stats, and I assigned labels to each cluster in order to analyze the role-transformations.

Figure 4 and figure 5 show the call-ups split up into G League roles and NBA roles, respectively.

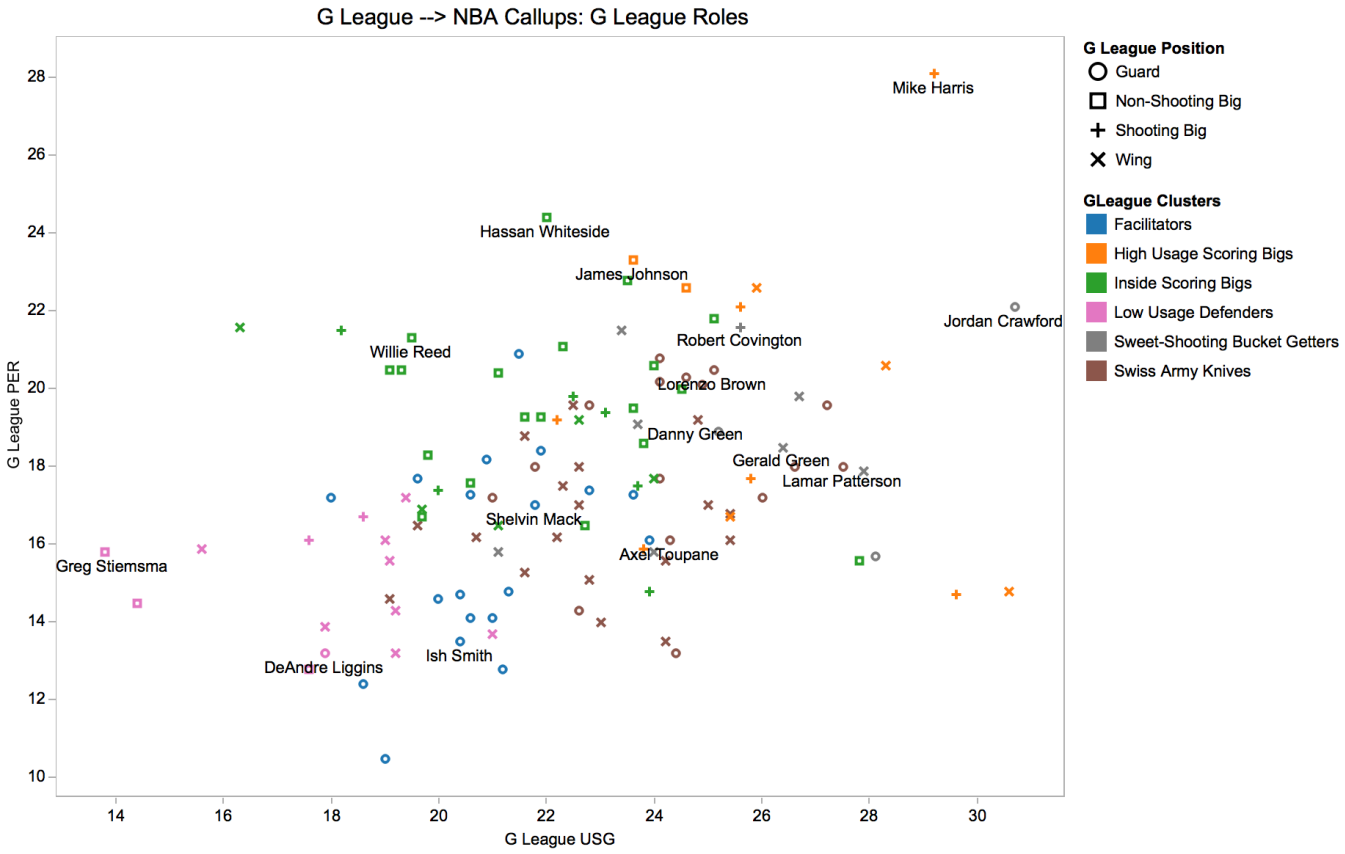

Figure 4: NBA call-ups broken into G League clusters (roles)

The big question becomes: is there continuity in these labels? In other words, if a player is considered a 'Swiss Army Knife[1]' in the G League, will he most likely be a 'Swiss Army Knife' in the NBA? This is a very important question because it has huge team-building consequences: If an NBA team has a spot to fill on their roster and wants to fill it with a pass-first, steady third-string point-guard, can they safely assume that if they call up a G League 'facilitator,' his game will translate and he can fill that same role in the NBA? Figure 6 provides some direction to answer this all-important question.

[1] In this context, a 'Swiss army knife' player is someone who does a little bit of everything for their team. They can score, they play-make and they rebound.

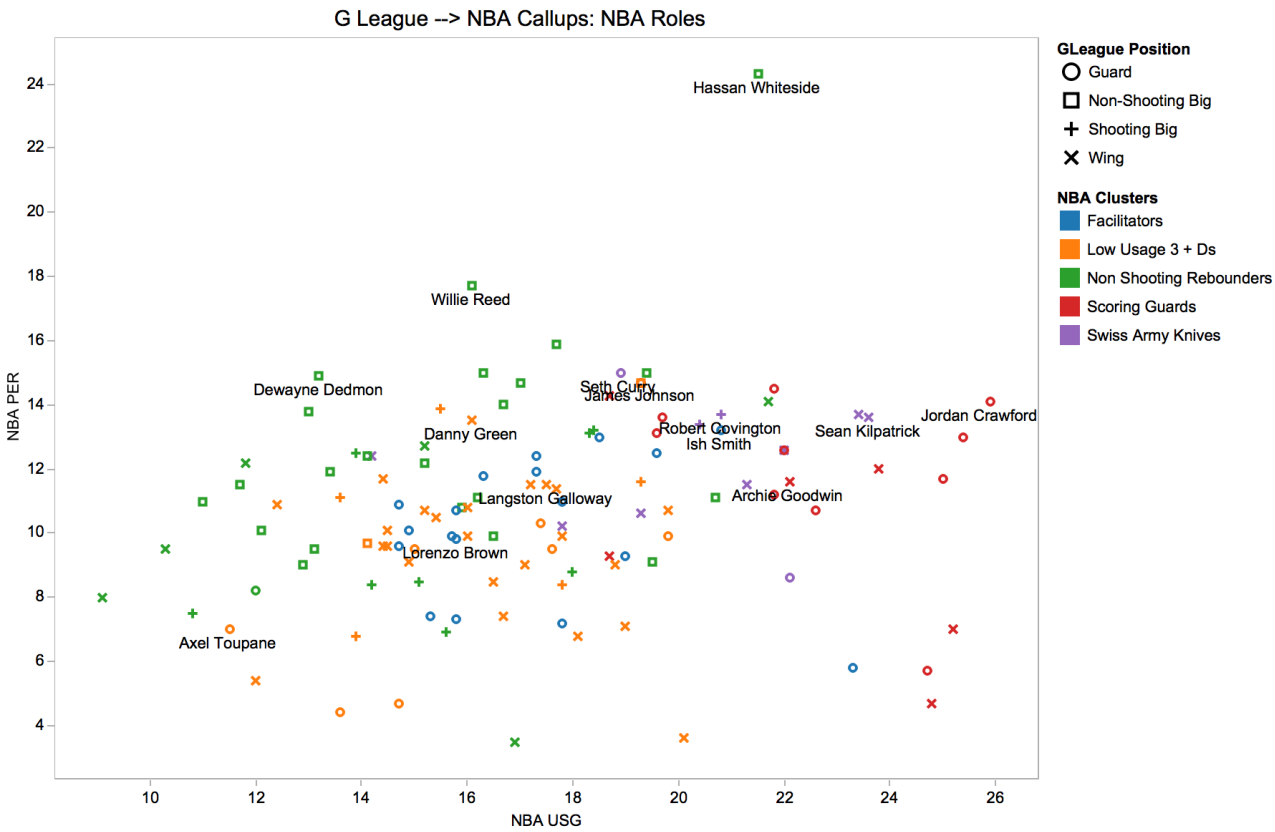

Figure 5: NBA call-ups broken into NBA clusters (roles)

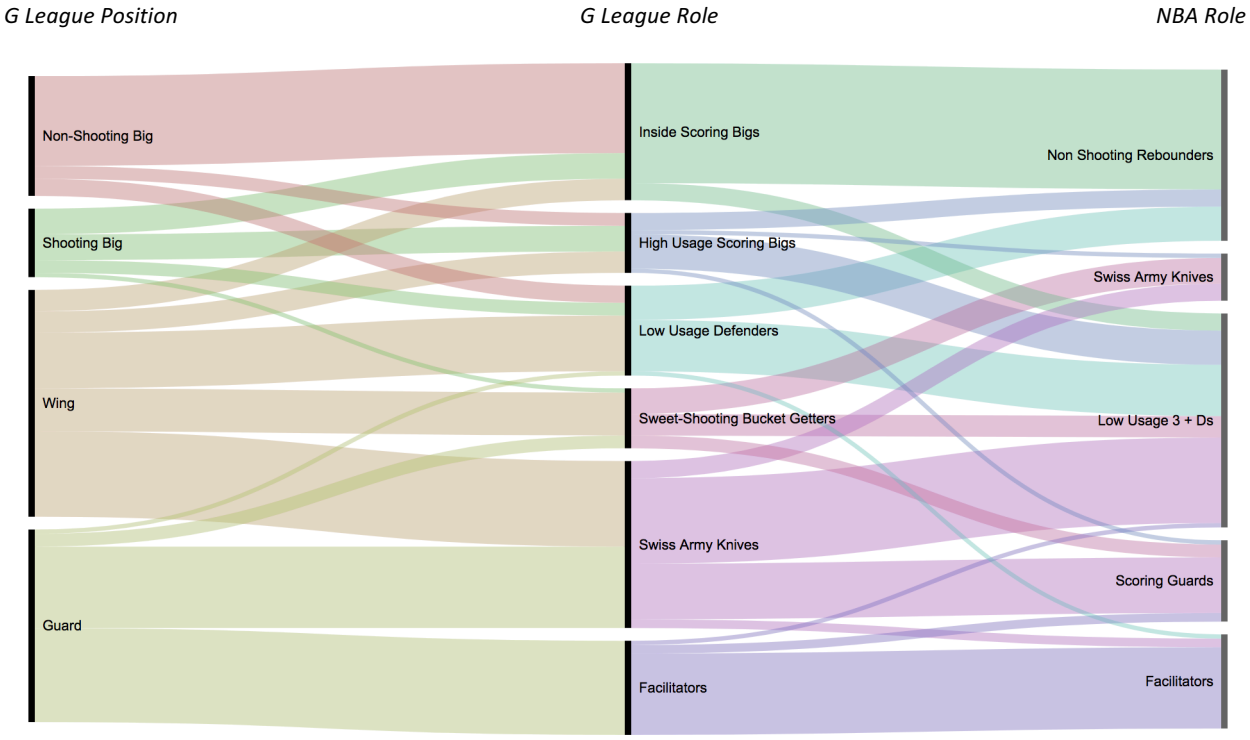

Figure 6: G League – to – NBA Role-Transformation

From figure 6, we can draw some fascinating insights:

- Big post-players typically maintain their roles of inside scoring and good rebounding.
- G League facilitators typically become NBA facilitators.
- Wings can have a variety of roles – they are by far the toughest position to predict
- G League Swiss army knives will typically fill smaller roles at the NBA level – they become predominantly score-first guards and low-usage 3+D players.
- A small minority of call-ups actually become an NBA 'Swiss army knife' – most G League players transform into a player with a specialty, whether it be scoring, rebounding or assisting.
    - For example:
        - G League Swiss army knife Lorenzo Brown → NBA facilitator
        - G League Swiss army knife Danny Green → NBA low usage 3+D
- The G League is a good place to find an NBA low usage 3+D player as evidenced by the number of call-ups who become one, but the tricky part is that they come from every single G League role, making it hard to predict which G League players will turn into successful NBA 3+D players, a skill so highly coveted in today's NBA.
    - This really is an amazing, frustrating phenomenon: it is evident that the G League is full of potential NBA 3+D guys, but they all play different G League roles. So, when scouting the G League for an NBA 3+D player, you have to be open-minded: the next great NBA 3+D guy might be a G League scoring guard, a rebounding wing, a Swiss-army knife, and even a facilitator.

Does G League Total Minutes Matter?

- *"We really like player X but he's played 3 seasons in the G League now (approx. 3500 minutes) and he hasn't been able to move up to the NBA, so we're going to pass."*
- *"Wow, did you see what player Y did in his first 10 G League games? He's almost averaging a triple-double! Let's give it 10 more games before we consider anything."*

These questions shed light on a very intriguing and important topic – is it ever too early or too late to make a decision based on G League data? Figure 7 splits up the players into the NBA roles we defined above. The results are interesting.

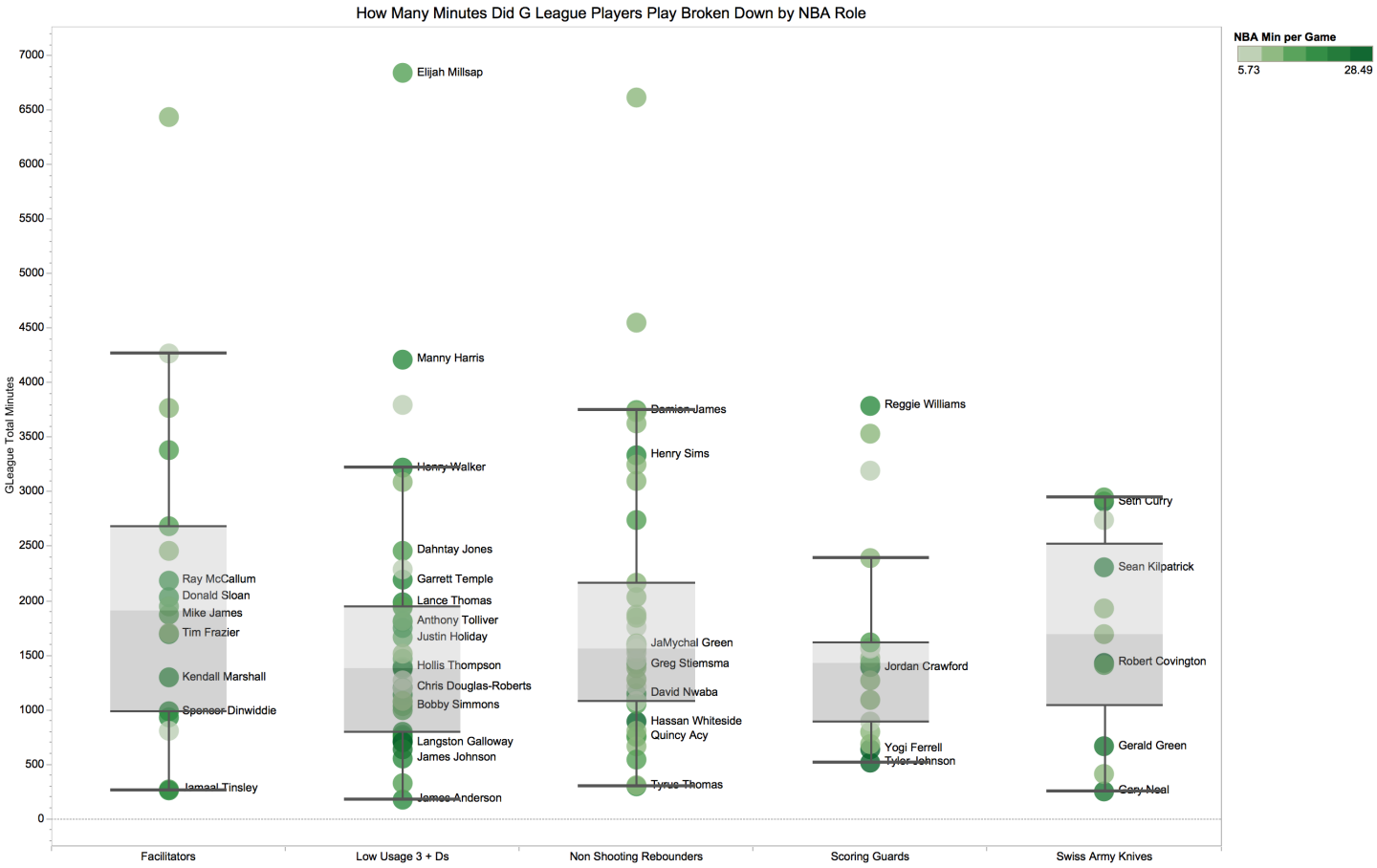

Figure 7: G League Minutes Box Plots By NBA role

- If a player really intrigues you within 1000 minutes of G League action (within 30 games), **take a shot** – over the last 8 years, a handful of prospects were called up before they surpassed 1000 G League minutes. They have tended to do fairly well in the NBA (measured by NBA min/game – the darker the green, the more minutes they average).
- In general, successful call-ups typically play 300-1500 G League minutes – if you like a player and his minutes fall somewhere in this range, make the move.

- Call-ups that turn into NBA scoring guards typically get called up before they play 1500 G League minutes. If you are impressed by an inexperienced G League player's scoring prowess, take a chance early.
- If you are looking for an 8 and 8 backup big, do not be put off by G League players who have played >3000 minutes down there. They are probably not going to turn into NBA stars, or even starters, but they can certainly play backup forward minutes in the NBA and look the part.

Conclusions

- **Good passing translates to the NBA, regardless of position. For the most part, players with high assist rates will continue that trend, and same for players with low assist rates.**
- **A player's fouling habits in the G League are indicative of how he will guard in the NBA: if he cannot guard in the G League without fouling, he will have the same issue in the NBA. However, the opposite is also true: if a player in the G League impressively defends without fouling, we can expect him to bring this skill to the NBA.**
- **There doesn't seem to be a relationship between G League turnovers per 36 and NBA turnovers per 36, probably because of role-transformation.**
- **Most call-ups will not play their G League role in the NBA, which makes scouting challenging. Figure 6 illustrates which roles do have some continuity and which roles have absolutely no predictability.**
- **Take a chance on an inexperienced G League player if he really impresses you; do not be scared by lack of minutes.**
- **The sweet spot for evaluating G League players successfully is 300-1500 minutes (approx. 8-50 games). This is an ideal range for making predictions.**

Results

As of right now, the G League → NBA prediction model is tough to accurately evaluate. Evaluating unsupervised data classification models is difficult, and often the best evaluation techniques are empirical by nature. Since the RCC framework is built upon accurate clustering, so too must this model be evaluated empirically. In other words, the RCC framework will only be as good as the accuracy of the clustering component. If the clusters are misrepresentative, the framework will fail. So, to evaluate the G League → NBA model, I tested it using 2018 G League players. If the model classifies players somewhat reasonably, I know I can be comfortable with the state of the framework.

I picked twenty-four 2018 G League players that I am familiar with and before running the model I labeled them myself, using my own intuition and domain knowledge. Then I ran the RCC model on these players and compared the model's classifications with my own. I used both k-nn majority classification and Gaussian Naïve Bayes to classify the players.

The results (shown in figure 1) are not quite where I'd like them to be, but they are promising; I am confident that some tweaking to the framework will result in better classification accuracy. Nonetheless, 67% accuracy for 6-class classification is an encouraging start.

Future Work

There are a couple of improvements I would like to make to the model, and I plan on doing so in the near future.

- Analyze the feature selection process: currently, the features used in the clustering algorithm only include features which show correlation between domains. However, perhaps the premise that only correlated features should matter is problematic. For example, there is not a strong correlation between three-point shooting percentage (3FG%) in the G League and in the NBA, but does that mean it should be excluded from the model? Surely a player's ability or lack thereof to shoot three pointers must factor into his G League and NBA role.
- Deliver more concrete predictions: as of now, the model predicts a G League role and outputs k player comparisons. Then, it is up to the user to make his decision regarding the player. I would like to have the model do more of the heavy lifting; I would like it to deliver a risk analysis based on the role predictions (what are the chances this player is an NBA-caliber player based on his role in the G League and his closest player comparisons?).
- Bring in G League industry experts to help me in empirically labelling the G League players with their role. I am familiar with some of the players, but having someone whose job it is to evaluate the G League help with labelling would be a significant benefit. With their help, I can be confident that the empirical labels are as accurate as they could be.
- Create a front-end component to deliver my results. I think this tool can be very beneficial to NBA front offices, but they would require an easy-to-use interface.

Discussion and Conclusion

The model, in its current form, provides good insight regarding what we can expect if we move a G League player to the NBA. It will only continue to improve as I implement important tweaks. One limitation that is currently unfixable is the small sample size of training data. There simply have not been a lot of players who have gone from the G League to the NBA and have been successful. While the number is steadily growing year over year, the sample size will stay relatively small. It is an unavoidable limitation of G League analytics.

I hope that my RCC framework can prove useful outside of the basketball world. It is meant as a tool for predicting role transformation from domain→domain. It respects the inherent differences of the two domains by clustering objects into encompassing roles, as opposed to just predicting individual features from one domain to the other. The RCC framework can be helpful in many sports industries, especially the ones with minor leagues (hockey, baseball). As well, the RCC framework can succeed outside the sports world. It can penetrate other industries that require domain → domain predictions:

- Education (Elementary → Middle School, Middle School → High School, High School → University)
  - When students go from one level of education to the next, what contributes to their change in grades? Do any outside features potentially factor in? (commute time, teacher-student ratio, location change, stress, etc.)
- Immigration
  - Can we group people who move from one location to another into classes and look at the features that made them move? (money, war, opportunity, family etc.)

Acknowledgments

I would like to thank Dr. Dan Lizotte for supervising me throughout this process and for always being available and willing to help.

Figures

	2018 G League Stats					Cole's Empirical Classification			
Player	3PM	AST	FGA	PTS	REB	G League Label	Cluster Number	3-nn classifier	Gaussian Naïve Bayes
Cleanthony Early	0.9	3.2	11.8	14.6	8.1	Swiss Army Knives	3	0	0
Monte Morris	1.7	6	13.3	18.4	4.7	Swiss Army Knives	3	2	3
R.J. Hunter	4.9	3.1	17.3	24.9	4.2	Sweet-Shooting Bucket Getters	4	4	4
Jeremy Hollowell	1.5	2.9	12.8	15.7	6.6	Inside Scoring Bigs	1	0	0
Naz Long	3.1	3.3	14.3	17.6	6	Swiss Army Knives	3	3	3
Nigel Hayes	2.1	2.4	12.6	17.5	6.8	High Usage Scoring Bigs	5	0	0
Rashawn Thomas	1.8	2.1	15.6	19.3	8.1	High Usage Scoring Bigs	5	3	3
P.J. Dozier	1.7	2.2	12	13.1	5.7	Low Usage Defenders	0	0	0
James McAdoo	0.8	3.2	10.2	10.9	6.9	Low Usage Defenders	0	0	0
Walt Lemon Jr.	1.2	5.5	17	22.4	4.6	Swiss Army Knives	3	3	3
Xavier Rathan-Mayes	1.6	6.7	12.3	12.3	7.8	Swiss Army Knives	3	2	2
Johnathan Motley	0.3	1.6	18.2	24.6	10.3	High Usage Scoring Bigs	5	5	5
Andre Washington	0	1.4	11.6	13	6.7	Low Usage Defenders	0	0	0
Austin Nichols	1.4	0	14.7	18.5	10.3	High Usage Scoring Bigs	5	1	5
Lorenzo Brown	1.1	7.6	16.6	20.5	5.1	Swiss Army Knives	3	3	3
Stephen Zimmerman	0	2.2	8.7	15.7	10.5	Low Usage Defenders	0	1	1
JaKarr Sampson	0.6	0.4	15	19.2	9.6	Inside Scoring Bigs	1	1	1
Briante Weber	0.7	7.1	13.1	18.9	6.5	Facilitators	2	2	2
Wesley Iwundu	0.2	2.2	13.5	16.3	8.7	Inside Scoring Bigs	1	1	1
Kennedy Meeks	0	1.1	14.2	14.7	11.3	Low Usage Defenders	0	1	1
Tyler Dorsey	3.4	2.6	16.7	19.9	7.4	Sweet-Shooting Bucket Getters	4	3	4
Alfonzo McKinnie	1.8	1.5	14.7	18.7	9.3	Swiss Army Knives	3	5	5

Antonius Cleveland	1.2	2.9	12.4	18.3	4.6	Swiss Army Knives	3	3	3
Larry Drew	1.1	9.6	8.7	8.4	6.2	Facilitators	2	2	2
							Accuracy:	**54%**	**67%**

Figure 1: Prediction Results for a sample size of 24