Aalto University
School of Engineering

The Analysis on Shooting Patterns of Basketball Players using GIS

Master's Thesis
Department of Built Environment
School of Engineering
Aalto University

Espoo, 27 November 2017

(signature)

Bachelor of Engineering in Urban Engineering
Woong Sik Shin

Supervisor: Professor Kirsi Virrantaus
Advisor: Professor Kirsi Virrantaus

Aalto University
School of Engineering

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi

Abstract of master's thesis

Author Woong Sik Shin

Title of thesis The Analysis on Shooting Patterns of Basketball Players using GIS

Degree programme Degree Programme in Geomatics

Major Geoinformation Technology **Code** IA3002

Thesis supervisor Professor Kirsi Virrantaus

Thesis advisor(s) Professor Kirsi Virrantaus

Date 27.11.2017	**Number of pages**	**Language**
	45+2 Pages	English

Abstract

Our world consists of space and time. Space is defined in three dimensions and time is defined in one dimension. Every object on earth has space and time attributes ubiquitously. However, these attributes are not sufficiently demonstrated to its effectiveness especially, in terms of time perspective. Therefore, the time perspective is relevant and fundamental for analyzing events as a key factor that leads to new outcomes.

GIS (Geographic Information System) manipulating spatial and temporal data on phenomena has been used for several decades and its enhancement leads to new possible analysis methods. In addition, due to GPS (Global Positioning System) device's development, spatial information has become easily accessible and the various spatial applications' use became common in our lives.

Basketball has recently been introduced to GIS. Basketball is considered one of the highest scoring sports that around 100 points per team in a game. Therefore, investigating the shooting hot spot has the opportunity to display substantial basketball analysis. In this research, basketball shooting data is explored and analyzed by consecutive temporal changes. In basketball, the exact shooting hot spot area in which the most shooting has been taken is crucial as the highest scoring sport. NBA (National Basketball Association: professional men's basketball league in the U.S) shooting data is used as a raw input data.

Significant shooting differences are found. By the "*Shot tendency*" method, the specific year's trend and its hot spots were examined. Although, the differences between the years are apparent, difficulty in assumption lies in how the hot spots have changed over time. By the "*Emerging hot spot*" analysis, the space over the time is understood by investigating the trend in temporal changes. Especially, this would be hard work when the data is so massive. The space-time cube provides spatial and temporal integration and the new possible analysis methods.

As a result, near the 3 point arc line (Zone 2 & 3 & 4), mostly continuous hot spot trends were detected. The zones have expanded compared to the past and has seen the recent significant growth. However, the 2 point areas, especially each side of the 2 point areas (Zone 6 & 7 & 9 & 10), mostly the diminishing hot spot trends were detected. The zones have been statically decreased in recent times with the exception of Zone 11.

Keywords Basketball, National Basketball Association, NBA, spatial and temporal changes, space time cube, space time pattern mining, emerging hot spot analysis, shot tendency, shooting hot spot analysis, 3 point trend, shot trend changes

Contents

Abstract

List of Figures
List of tables
Abbreviations

1. Introduction 1
2. Background 3
   - 2.1 The History of Basketball 3
   - 2.2 Special Characteristics of Basketball 4
   - 2.3 The Reason Why Basketball is Special 5
   - 2.4 What is NBA? 6
   - 2.5 Traditional Trend in the NBA 6
   - 2.6 The Transition of NBA 8
   - 2.7 Current NBA Shooting Trend 10
   - 2.8 The Higher 3 Point Shooting Team 10
   - 2.9 The Trend Leader of NBA 11
3. Literature Review 13
   - 3.1 Goldsberry's Analysis of Basketball Data 13
   - 3.2 Motion Tracking into NBA 15
   - 3.3 Defensive Factor 16
4. Theory 18
   - 4.1 Shot Tendency 18
   - 4.2 Emerging Hot Spot Analysis 19
   - 4.2.1 Space-Time Cube 19
   - 4.2.2 Emerging Hot Spot Analysis 20
5. Materials and Methods 23
   - 5.1 Data Preparation 23
   - 5.2 Data Visualization 24
   - 5.3 Shot Tendency 25
   - 5.4 Create Space-Time Cube 27
   - 5.5 Emerging Hot Spot Analysis 30
6. Result 31
   - 6.1 A Brief Comparison of Total Shooting Trend in 2005 to 2016 31
   - 6.2 Shot Tendency Comparison in 2005 vs 2016 32
   - 6.3 Emerging Hot Spot Analysis 34
   - 6.3.1 3 Point Left & Right Corner Zone (Zone 1 & Zone 5) 36
   - 6.3.2 3 Point Center-left, Center, Center-right (Zone 2 & 3 & 4) 36
   - 6.3.3 2 Point Center-left, Center, Center-right (Zone 7 & 8 & 9) 37
   - 6.3.4 2 Point Left, Center, Right (Zone 6 & 10 & 11) 38

6.4 The Emerging Hot Spot Trend Summary 38
7. Discussion and Further Research 40
7.1 Are There Different Shooting Hot Spots Existed Between the Year 2005 and 2016? 40
7.2 If so, How has the Shooting Hot Spot Trend Changed Over the Space by Temporal Changes? 40
7.3 Why is It Important to Know the Trend by Temporal Changes? 41
8. Conclusion 42

References 43
Appendix 1. Python code for importing NBA shooting data to ArcGIS 46

List of Figures

Figure 1. The creator of basketball: Dr. Naismith 3
Figure 2. Most popular sport in the world announced by TotalSportek. (Totalsportek, 2016) 4
Figure 3. Official NBA court size (in foot) 5
Figure 4. MVP awarded time by positions. (ESPN, 2017) 9
Figure 5. 3PS (3Point Selection) % and 3PM (3Point Made) % in the NBA since 2005. 10
Figure 6. Kenny Smith, NBA TV analyst and retired NBA baller, defined the Curry as new Jordan. (Smith, 2016) 12
Figure 7. The highest number of 3 point shot, made in the year in NBA history. 12
Figure 8. Field Goal Attempts (left) / Shot tendencies (right). (Goldsberry, 2012, p. 3) 13
Figure 9. Spread indicates the player's shooting territory. (Goldsberry, 2012, p. 3) 13
Figure 10. Spread variable visualized for Al Jefferson and Ray Allen. (Goldsberry, 2012, p. 4). 14
Figure 11. Range indicates the efficiency of the shot made in the SA (Scoring Area). (Goldsberry, 2012, p. 4). 14
Figure 12. Range depiction of 4 different players in the NBA. (Goldsberry, 2012, p. 5) 15
Figure 13. Diagram of EPV (Expected Possession Value) to evaluate decision moment. (Cervone, et al., 2014, p. 2) 15
Figure 14. The REMO (Relative Motion) matrices in basketball analysis. (Jensen, 2014, p. 41) 16
Figure 15. Matchup matrix that shows the point value according to the defense matched. (Franks, et al., 2015, p. 2) 17
Figure 16. Spread indicates the player's shooting territory. (Goldsberry, 2012, p. 3) 18
Figure 17. Space Time Pattern Mining Tools in Esri's ArcGIS. 19
Figure 18. The structure of space-time cube. (Esri, 2017) 20
Figure 19. 16 Types of trend defined in emerging hot spot analysis. (Esri, 2017) 22
Figure 20. The raw data from stats.nba.com. 23
Figure 21. The raw NBA shooting data in 2005 to 2016 loaded to ArcGIS (10.5). 25
Figure 22. The fishnet (hexagons) created over the area 26
Figure 23. Spatial join tool used to aggregate points. 26
Figure 24. The points converted from hexagons by "Feature to Point" tool in ArcGIS. 27
Figure 25. The space-time cube bins details. 28
Figure 26. The Space-time cube by 1 month intervals in ArcScene 3D (10.5). 28
Figure 27. The exploration of the space-time cube bin partitions. 29
Figure 28. The emerging hot spot analysis tool. 30
Figure 29. Summary of emerging hot spot analysis. 30
Figure 30. The total shooting trend changes per game in 2005-2016 in regular season. 31
Figure 31. The shooting trend comparison between the year 2005 and 2016 by shot tendency. 32
Figure 32. The Top range Hot spot extracted from shot tendency. 33
Figure 33. The summary of the result in emerging hot spot. 34
Figure 34. The emerging hot spot analysis on the NBA shooting trend from 2005 to 2016. 35
Figure 35. The half court division. 35
Figure 36. The emerging hot spot trend analysis in the zone 1 & 5. 36
Figure 37. The emerging hot spot trend analysis in the zone 2 & 3& 4. 37
Figure 38. The emerging hot spot trend analysis in the zone 7 & 8 & 9. 37
Figure 39. The emerging hot spot trend analysis in the zone 6 & 10 & 11. 38

List of tables

Table 1. Top 10 point leader career totals in the regular season from 1946 to 2016. (Land of basketball, 2017) 6
Table 2. Top 10 point leader per game in the regular season from 1946 to 2016. (Land of basketball, 2017) 7
Table 3. Top 5 point leader in single season total (left) / Top 5 point leader per game (right) from 1946 to 2016. (Land of basketball, 2017) 8
Table 4. Top 5 leader teams in the highest 3PA (upper) and 3PM% (bottom). (NBA STATS, 2017) 11
Table 5. The two teams commonly ranked in 3PA and 3PM% from Table 4. 11
Table 6. The emerging hot spot trends detected in 11 zones 36
Table 7. The emerging hot spot trend summary table 38

Abbreviations

3PA	3 Point Attempt
3PM	3 Point Made numbers
3PM%	3 Point Made rate
3PS%	3 Point Selection rate in total attempts
AR	Augmented Reality
Box-out	A type of action to keep your own space for rebound
C	Center
EPV	Expected Possession Value
ESPN	Entertainment and Sports Programming Network: a U.S.-based global cable and satellite sports television channel
Fade away	A shot type that jumped in backward to avoid defense
FGA	Field Goal Attempts: all kind of shot such as 2 point and 3 point
FGM	Field Goal Made
FIBA	International Basketball Federation
Finger role	A shot type that uses fingers making spins of the ball
Free Throw shot	Freely take a shot without any block including defense
Getis-Ord Gi* statistics	*"Investigates dataset by neighborhood and return it into z-score and p-value to find hot spot"* (Esri, 2017)
GIS	Geographic Information System
GPS	Global Positioning System
IOC	International Olympic Committee
Mann-Kendall trend	*"A rank correlation analysis for bin count or value and time sequence" (Esri, 2017)*
MIT	Massachusetts Institute of Technology
MLB	Major League Baseball in the USA, one of 4 major sports
MVP	Most Valuable Player
NBA	National Basketball Association in the USA, one of 4 major sports
NetCDF	Network Common Data Form
NFL	National Football League in the USA, one of 4 major sports
NHL	National Hockey League in the USA, one of 4 major sports
P-value	*"In Hot spot analysis, P-value (probability) is calculated by Pattern analysis tool (such as Getis-Ord Gi* statistics) which tell you spatially high or low values cluster in Probability" (Esri, 2017)*
PF	Power Forward
PG	Point Guard
PPA	"Points Per Attempt" (Goldsberry, 2012)
Range	*"Effective shooting range of player across all scoring cells"* (Goldsberry, 2012)
REMO	Relative Motion

SA	*"Scoring Area"* (Goldsberry, 2012)
Screen	A type of action to make a space for your team mates by blocking other team's defenses
SF	Small Forward
SG	Shooting Guard
Spread	*"Total spatial spread of player across all scoring cells"* (Goldsberry, 2012)
Team foul	The number of fouls that the players do belongs to the team
Un-sportsman-like foul	Also called flagrant foul that which can give injuries with intention like excessive or violent contact The player is kicked out from the court right away
YMCA	Young Men's Christian Association
Z-score	*"In Hot spot analysis, Z-score is calculated by Pattern analysis tool (such as Getis-Ord Gi* statistics) which tell you spatially high or low values cluster in Standard Deviation"* (Esri, 2017)

1. Introduction

Our world consists of space and time. Space is defined in three dimensions and time is defined in one dimension. Every object on earth has space and time attributes ubiquitously. However, these attributes are not sufficiently demonstrated to its effectiveness especially, in terms of time perspective.
Time belongs to a space that subsist an occurrence. In other words, all events in space must have a time as a vital element. Time is different to space from the distinctive ability to exist in the past, present and future. These three categories can be approached paralleled in time. Therefore, the time perspective is relevant and fundamental for analyzing events as a key factor that leads to new outcomes.

GIS (Geographic Information System) manipulating spatial and temporal data on phenomena has been used for several decades and its enhancement leads to new possible analysis methods. In addition, due to GPS (Global Positioning System) device's development, spatial information has become easily accessible and the various spatial applications' use became common in our lives. Portable devices such as smart phones or smart watches are able to collect geographical information data and apply it into advanced applications in navigation systems, location-based web and map applications, and into game fields such as AR (Augmented Reality) used in the "*Pokémon GO*" game. In a more general sense, the current GIS is used universally in various industries for instance in urban planning, civil engineering, geology, natural resources, agriculture, biology, logistics, crime analysis, dental care, marketing etc. The latest GIS has become a powerful and effective tool to manage the data in an efficient way that brings a sharp insight for making better decisions.

A new era raised in sports analytics using GIS has come besides these traditional fields above. For example, in football, each player's movements are captured and digitized into a number through special cameras. From speed, directions, trajectories to action types of players based on location are recorded and finally GIS converts it to readable and valuable information that makes games smart.
Basketball has recently been introduced to GIS. Basketball is considered one of the highest scoring sports that around 100 points per team in a game. Therefore, investigating the shooting hot spot has the opportunity to display substantial basketball analysis.

Previous basketball research on shooting hot spot analysis mostly focused on complex figures in specific moments or comparisons between specific periods without consecutive time consideration. The advantageous aspect of recent basketball analyses is the usage of GIS and the considerable improvements made in spatial perspectives. However, relative to the spatial perspective, the consecutive temporal changes need to be further examined and developed. If consecutive temporal perspective is not considered, it gives limited information such as only having specific moment's trend with unknown variables on the hot spot's transition over space and time.

In this research, basketball shooting data is explored and analyzed by consecutive temporal changes. In basketball, the exact shooting hot spot area in which the most shooting has been taken is crucial as the highest scoring sport. By investigating the data according to temporal changes, the hot spot trend transition is analyzed. In the end, the emerging shooting hot spot of basketball players is defined as an overview of the history and the present.

For the analysis, NBA (National Basketball Association: professional men's basketball league in the U.S) shooting data is used as a raw input data. The data is public and detailed information of players is available online. NBA provides the information via URL requests and all the methods are explained in the research.

The primary goal is to approach the data by consecutive temporal changes that enable to define "Basketball (NBA) shooting pattern trend changes in the last 12 years from 2005 to 2016".

In order to approach the objective, a journey of this research asks the following three questions:

1. Are there different shooting hot spot trends between the year 2005 and 2016?
2. If so, how has the shooting hot spot trend changed over the space by the temporal changes?
3. Why is it important to know the trend by temporal changes?

The first question is given to find out the shooting hot spot in the past year of 2005 and in the current 2016. The year of 2005 and 2016 data sets are compared. The question explores the specific moment's shooting trend in the year. In addition, the research discusses why the comparison has limitations for checking trend changes. The second question is given to find out the trend changes over space and time. This shows how the overall shooting trend has been changed by temporal changes and how it is conducted in an efficient way with adjustable parameters. The third question is given to discuss the reason why it is important to know the trend by temporal changes, what the benefit is and what the time consideration gives.

This research is structured as below to find out the objective: In the second chapter, background information on basketball is addressed to give a basic understanding of basketball; what is basketball and why basketball is unique. In addition, the definition of NBA and a brief trend comparison of NBA historical records is presented. In the third chapter, the literature review is addressed. Previous research is reviewed by two themes; offensive factors (shot tendency and movement tracking) and a defensive factor (match up matrix). This shows how recent basketball analyses have been conducted using GIS. In the fourth chapter, the theoretical framework is introduced which is necessary for the analysis. The "*Shot tendency"* theory (Goldsberry, 2012) and the "*Emerging hot spot*" theory are described. In the fifth chapter, the methodological framework is explained. Based on the theoretical framework, the practical framework is addressed. Each process and its adjustable parameter are also described. In the sixth chapter, the results are addressed by investigating each zone (total 11 zones). The research summarizes each zone's trend. In the seventh chapter, the elements found during the studies and how it could be continued to further develop methods in the future is discussed. The eighth chapter concludes the research paper with a brief summary.

Defining the trend changes over the last 12 years (from 2005 to 2016) in the NBA will be a new perspective of reviewing the past and present at a glance and it will allow a deeper understanding of shooting trend changes in an effective way. The research has prospects in its pragmatic analysis to further develop competent strategies and plans for the players, teams, coaches and also fans.

2. Background

In the second chapter, background information on basketball is addressed to give a basic understanding of basketball. What is basketball and why basketball is unique in comparison to other sports are explained. In addition, the definition of NBA (National Basketball Association) and a brief trend comparison of NBA historical records are presented. An overview of basketball shooting trends through NBA league in the USA is addressed as well.

2.1 The History of Basketball

During the early days of basketball, real baskets were used instead of current rims. In 1891, Dr. Naismith, the inventor of basketball used to work as a physical education professor and also an instructor at YMCA (Young Men's Christian Association). He contemplated how to continue his gym class for young children in the bad weather. In the end, he put a basket on wall of the court in the gym and the children tried to put the ball into the baskets. He also wrote the original basketball rule book where most rules are still used to this day. (Naismithbasketballfoundation, 2014) & (NBA, 2002)

Figure 1. The creator of basketball: Dr. Naismith [1]

Basketball became one of four major sports in the USA (NBA: Basketball, MLB: Baseball, NFL: American Football, NHL: Ice Hockey) but also became one of most popular sports in the world.

According to Totalsportek, announced in January 2016, basketball ranked 2nd after football (also known as soccer) in most popular sport in the world. This was studied by 13 criteria that shows the popularity such as *"global base and audience, TV viewership, number of professional leagues around the world, TV rights deals and sponsorship deals, average athlete salary in top league, biggest competition and number of countries represented, social media presence, prominence in sports headlines on media outlets (websites, tv), relevancy through the year, regional dominance, gender equality, accessible to general public worldwide"*. (Totalsportek, 2016)

Totalsportek says *"Basketball has over 1 billion followers"* and *"the fastest growing sport on every scale from revenues and getting established in countries across the world."* (Totalsportek, 2016)

[1] Image from (https://en.wikipedia.org/wiki/Basketball)

Most popular sport in the world	
1	Football / Soccer
2	**Basketball**
3	Cricket
4	Tennis
5	Athletics
6	Rugby
7	Formula 1
8	Boxing
9	Ice Hockey
10	Volleyball
11	Baseball
12	American Football
13	MMA
14	MotoGP
15	Field Hockey

Figure 2. Most popular sport in the world announced by TotalSportek. (Totalsportek, 2016)

In January 2017, the IOC (International Olympic Committee) announced 3 on 3 basketball as a part of the official program started in Tokyo 2020 Olympic games for men and women (FIBA, 2017).

The 3 on 3 basketball is played with 6 players in a half court while the official basketball game is played with 10 players (5 on 5) in a full court. This is a type of minimized game such as futsal (6 players) from football (soccer; 11 players are played). 3 on 3 basketball is an outdoor sport while official is originally designed as an indoor sports. 3 on 3 rules are almost same with the official game but it has a faster transition as only half the court is in use. Also the 3 on 3's score system consists of 2 and 1 point instead 3 and 2 point in the official.

This shows that basketball has popularity and possible opportunity to involve more people in the future.

2.2 Special Characteristics of Basketball

Basketball is a sport played on court. FIBA (International Basketball Federation) defines an international official court size; 28 m long and 15 m wide (FIBA Central Board, 2017). But the NBA uses a bigger sized court; 28.65 m long and 15.24 m wide (NBA, 2013). The 3 point distance is also different; 6.75 m (FIBA) and 7.24 m (NBA).

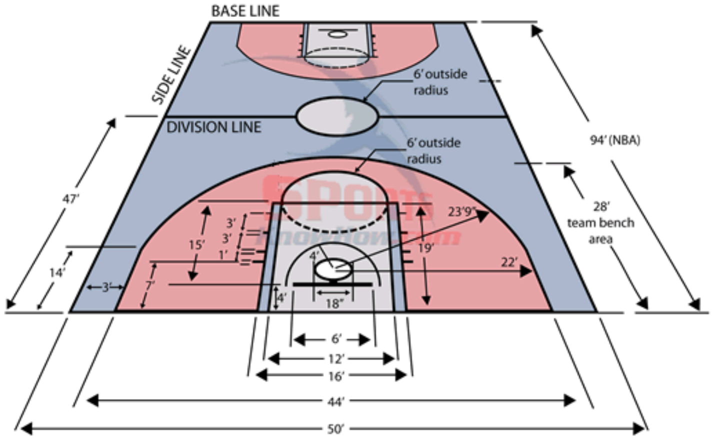


Figure 3. Official NBA court size (feet) [2]

The total playing time takes 48 minutes in four quarters (12 minutes for each quarter), if the score is tied then there is an additional overtime quarter for five minutes. Each person can have a maximum six fouls but if a player reaches sixth fouls, the player needs to be replaced by the other player. These personal fouls are saved as team fouls that the player belongs to, and if the team foul is over fifth fouls, give the opponent two free throw shots (NBA, 2013).

The basic rules introduced above present basketball as a fair sport by the system. The foul system protects the players from fights or injuries from frequent body contact occurred during the game. In addition, the system is used as a part of strategies. As doing fouls on the opponent who has a low percentage of free throw shots, able to minimize the points. This is a fair game that limits the number of fouls. For example, in a football (soccer) game, there is no limitation for normal fouls which is not serious (but yellow or red card is applied to the serious case). However, basketball has the limitation of counting all the fouls but also strictly forbidden for the serious cases (un-sportsman-like foul: also called flagrant foul which can give injuries with intention like excessive or violent contact). In this case, the player is kicked out of the court right away after the announcement from the referee.

2.3 The Reason Why Basketball is Special

First of all, basketball is a strictly team-oriented sport. In comparison with baseball, baseball has specific innings for each team. One team offenses until 3 outs end and then do defense until same outs come. In other words, baseball is a game between the pitcher and the batter until the ball gets a hit. In contrast, basketball players simultaneously play in the game in frequent changes of the ball possession between an offense and defense. There is no fixed innings (or periods) that are guaranteed for an offense or defense. Only a 24 second limit to shoot on an offense is used. (Mcclusky, 2014).
Secondly, the defense is matched. On the court, the position of the 10 players is mixed during the game. In tennis, baseball and volleyball, the players are separated between the offenses and defenses that physical contact is not allowed. However, intensive physical contact is allowed in basketball which makes height, size, strength or speed all relevant. In addition to these physical factors, BQ (Basketball IQ) that intellectual understanding of basketball is applied. Players need strategic plans such as to figure out a way out from the *box-out* position (a type of action to keep your own space for rebound) or a *screen* position (a type of action to make space for your team player by blocking the opponent defenses).
Thirdly, basketball has a different scoring system. If a shot is made from a distance of 7.24 m (23 feet 9 inches) from the rim, it is counted as 3 point. Otherwise, it is counted as 2 point. Furthermore, there is a 1 point scoring system called free throws that allows the shots without any blocking from defenses. The average scoring in one game is around 100 points and it is the highest scoring sport.

[2] Image from (http://www.sportsknowhow.com/basketball/dimensions/nba-basketball-court-dimensions.html)

The direct use of hands makes it possible to score up to or over 100 points. It has been proven that hands give the most accurate movements of human beings. By using the hands directly, meticulous skills such as a *finger roll* (a shot that uses fingers making spins of the ball to avoid blocking) or a *fade away* (a shot jumped in backward to avoid defense). Various scoring methods and skills using hands have been created and developed in the way that the NBA defined 70 shooting types which were non-existent before.

2.4 What is NBA?

NBA (National Basketball Association) is a professional basketball league for men in the USA and Canada. NBA is also one of four major sports (NBA: Basketball, MLB: Baseball, NFL: American Football, NHL: Ice Hockey) since 1962. NBA was initially established in 1946, but adopted and merged with other basketball leagues. 30 teams (29 in the USA and 1 in Canada) in the NBA are divided in 2 conferences divisions (Western and Eastern) and it has the longest regular season in the world. NBA players play at least 82 games for regular season (6 months) and 16 to 28 games for playoffs (2 months). Well-known athletes, LeBron James or Stephen Curry, spend 8 to 9 months in a year for the NBA league. According to ESPN (Entertainment and Sports Programming Network: a global sports channel in the U.S.), NBA (National Basketball Association: professional men's basketball league in the U.S) ticket sales have been increased in the last 3 years. Moreover, TV broadcasting and its advertisement deals are successfully contracted with large budgets. (ESPN, 2016).

2.5 Traditional trend of NBA

In general, five positions are defined in basketball. The guard position is divided into PG (Point Guard), SG (Shooting Guard) and the forward position is divided into SF (Small Forward) and PF (Power Forward). Lastly, there is the C (Center) position.

Table 1. Top 10 point leader career totals in the regular season from 1946 to 2016. [3] **(Land of basketball, 2017)**
Blue: Center or forward positions' players who had inconsiderable number of 3PA (3 Point Attempts).
Red: The player's 3PA (3 Point Attempts) and 3PM (3 Point Made) that their playing style is assumed.

	Player	**Position**	Heights	Points	Seasons	Games	FGM	FGA	**3PM**	**3PA**
1	Kareem Abdul-Jabbar	**C**	218 cm	38,387	20	1,560	15,837	28,307	**1**	**18**
2	Karl Malone	**F**	206 cm	36,928	19	1,476	13,528	26,210	**85**	**310**
3	Kobe Bryant	F-G	198 cm	33,643	20	1,346	11,719	26,200	1,827	5,546
4	Michael Jordan	G	198 cm	32,292	15	1,072	12,192	24,537	581	1,778
5	Wilt Chamberlain	**C**	216 cm	31,419	14	1,045	12,681	23,497	**-**	**-**
6	Dirk Nowitzki	F	213 cm	30,260	19	1,394	10,688	22,600	1,780	4,668
7	LeBron James	F	203 cm	28,787	14	1,061	10,423	20,803	1,467	4,295
8	Shaquille O'Neal	**C**	216 cm	28,596	19	1,207	11,330	19,457	**1**	**22**
9	Moses Malone	**C**	208 cm	27,409	19	1,329	9,435	19,225	**8**	**80**
10	Elvin Hayes	**F**	206 cm	27,313	16	1,303	10,976	24,272	**5**	**34**

[3] Notes: - Updated on July 12, 2017
- Three-pointers were implemented in 1979-80.
- This list includes players with a minimum of 400 games or 10,000 points.

The position's movement radius is defined in order. Generally, PG (Point Guard) has the farthest radius and C (Center) has the closest from the rim. Strong teams tend to include players with a large physique because it is a game of height. Scoring is made easiest when in close distance to the rim with height as an advantage. Table 1 shows the top 10 point leaders in the last 71 years of NBA history from 1946 to 2016. The player's characteristics on the list are assumed by figures; FGA (Field Goal Attempts: all kind of shots including 2 point and 3 point), FGM (Field Goal Made) and 3PA (3 Point Attempts), 3PM (3Point Made) categories.

In general, the 3 point shot is not a preferable option due to the relatively lower FGM%. The six players out of 10 are places as center or forward positions that have none or a few 3PA but they are ranked as the top 10 point leaders in the players' entire career. Kareem Abdul-Jabbar (218 cm, 102 kg) who is ranked 1st place is assumed as a player who dominated the rim with his above average height. Although Abdul-Jabbar made only 1 out of 18 shots in 3 point attempts in 20 seasons, he is the highest scoring player in NBA history. The colored segments in the Table 1 above are also assumed to have played near the rim with their heights according to 3PA and 3PM. Thus, the assumption is made that the infrequent amount of 3PA on par with the close distance to the rim are a preferred option in the list.

The average height of NBA players has steadily grown from 194.91 cm (1950s) to 201 cm (2016). The Center position's height have also increased from 205.4 cm (1950s) to 210.57 cm (2016) (Brocato, 2014) & (Willard, 2016). Kareem Abdul-Jabbar was 10 cm above the average center positions' height when playing in the 1970s to 1980s. Other players in the blue colored segment, were also taller than the average height in the same center and forward position. The evidence shows that basketball is a game of height.

Table 2. Top 10 point leader per game in the regular season from 1946 to 2016. [4] **(Land of basketball, 2017)**
Blue: Guard position players / Red: No center position positions & over 16% of 3 PA in FGA in player's career total

	Players	**Position**	Heights	**Points/game**	Total	Seasons	Games	**3PM**	**3PA**
1	Michael Jordan	**G**	198 cm	**30.12**	32,292	15	1,072	581	1,778
2	Wilt Chamberlain	C	216 cm	30.07	31,419	14	1,045	-	-
3	Elgin Baylor	F	196 cm	27.36	23,149	14	846	-	-
4	Kevin Durant	**F**	206 cm	**27.2**	19,121	10	703	**1,265**	**3,304**
5	LeBron James	**F**	203 cm	**27.13**	28,787	14	1,061	**1,467**	**4,295**
6	Jerry West	**G**	191 cm	**27.03**	25,192	14	932	-	-
7	Allen Iverson	**G**	183 cm	**26.66**	24,368	14	914	**1,059**	**3,383**
8	Bob Pettit	F	206 cm	26.36	20,880	11	792	-	-
9	George Gervin	**G**	201 cm	**26.18**	20,708	10	791	77	259
10	Oscar Robertson	**G**	196 cm	**25.68**	26,710	14	1,040	-	-

Another trend is depicted on Table 2. There are five players out of 10 as guard positions. Michael Jordan (198 cm, 98 kg), ranked 1st place, scored over 30 points in every game for 15 years. In the table, four more guard positions scored above 25 points per game.

Looking at the 3PA figures on Table 2, six players out of 10 either did not attempt the 3 point shots at

[4] Notes: - Updated on July 12, 2017
- Three-pointers were implemented in 1979-80.
- This list includes players with a minimum of 400 games or 10,000 points.

all or attempted than 300 times. Although Michael Jordan ranked 1st place, he attempted 1,778 shots of 3 point shots which is only 7.25% proportion out of the total 2 point and 3 point shot attempts.

In comparison, those of who are not center position, the proportion of Jordan's 3PA is relatively small. Other players had over 16% of 3 point in a total shot (16.99% for Iverson, 20.65% James, 25.06% Durant). Jordan attempted 2 point shots more than 3 point compared to the other not-center positions' player. The low 3 PA percentage does not represent Jordan's Center-like playing style under the rim with height but had the advantage of his height (198 cm) as a SG (Shooting Guard) during his playing time. The average height of SG was 195.55 cm in the 90s.
Michael Jordan and Wilt Chamberlain scored over 30 points in every game for 14 to 15 years and the overall 3 point attempts was relatively small for both. However, Durant, James, and Iverson (highlighted in red) attempted more than 3,000 times of 3 point shots (over 16% of FGA) which means they preferred 3 point shots. The common factor of these 3 players is that they played in recent years. Iverson played in the 2000s (1996 to 2007) and Durant and James are currently playing in the league (2017). It can be seen that the past and present shooting preferences of the players have changed.

Table 1 and Table 2 show an interesting factor. The Table 1 (point leader in career total), the top ranked players have mostly played over 15 years while Table 2 (point leader per game) shows the players played less than 15 years. This data shows the difficulty level of keeping consecutive high scores for a long period of time and the dependency of physical conditions. The guard position players score intensively for a relatively short time than the center position.

Table 3. Top 5 point leader in single season total (left) / Top 5 point leader per game (right) from 1946 to 2016. (Land of basketball, 2017)

	Player	Points	Season	Games
1	**Wilt Chamberlain**	**4,029**	1962	80
2	**Wilt Chamberlain**	**3,586**	1963	80
3	Michael Jordan	3,041	1987	82
4	**Wilt Chamberlain**	**3,033**	1961	79
5	**Wilt Chamberlain**	**2,948**	1964	80

	Player	Points	Total	Season	Games
1	**Wilt Chamberlain**	**50.36**	4,029	1962	80
2	**Wilt Chamberlain**	**44.83**	3,586	1963	80
3	**Wilt Chamberlain**	**38.39**	3,033	1961	79
4	**Wilt Chamberlain**	**37.6**	2,707	1960	72
5	Michael Jordan	37.09	3,041	1987	82

Table 3 shows the top 5 point leader in a single season. Chamberlain dominated the league in the 60s to 70s with the height and weight of 216 cm, 125 kg. Chamberlain once scored 100 points in a single game which was recorded as the highest score in 71 years of NBA history. His regular season average point was 50.36 points per game. He scored 4,029 points in 80 games in the same season which means he scored an estimate of 50 points per game in the year. Chamberlain's record is evidence to why basketball is a game of height.

NBA history records show that basketball is still a game of height. The table figures used in this research were updated on July 12, 2017.

2.6 The Transition of NBA

Investigating and comparing score rankings give limited insights. Scoring is one of the most important elements in basketball but other factors are also important. Basketball is equally divided into offense but also defense. The defensive abilities (steal, rebound, blocking etc.) must be considered.

The MVP (Most Valuable Player)s represent NBA leaders that shows the most effective players in the league by various abilities.

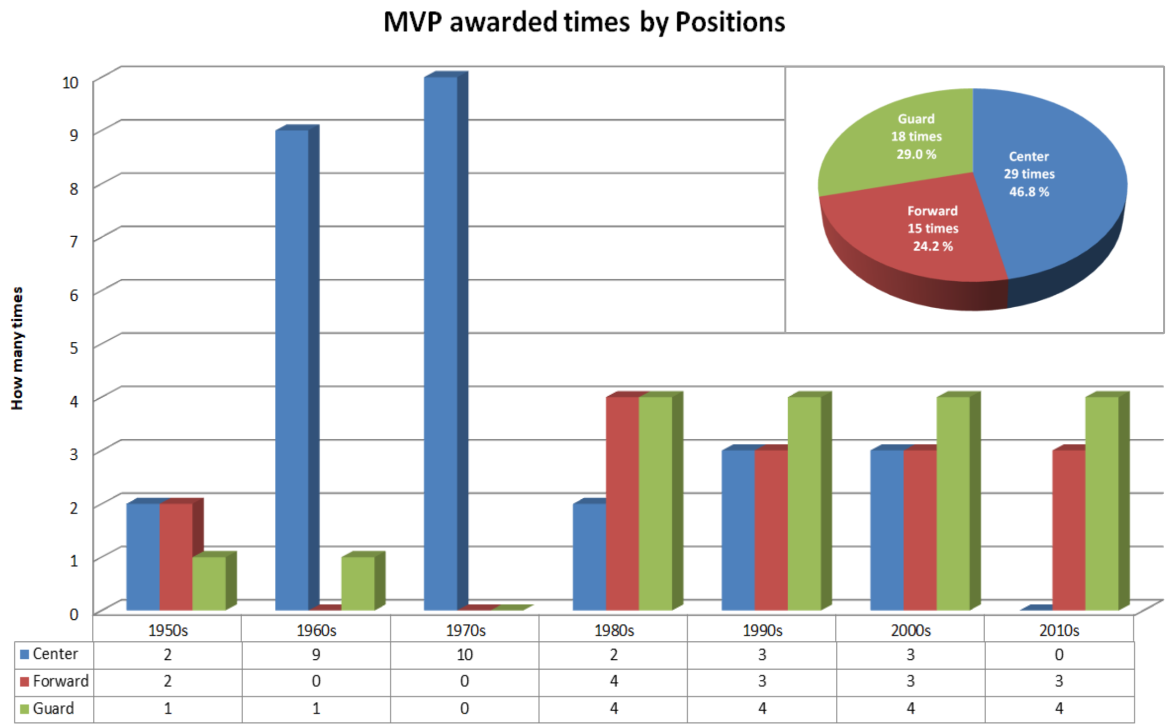


Figure 4. MVP awarded time by positions. (ESPN, 2017)

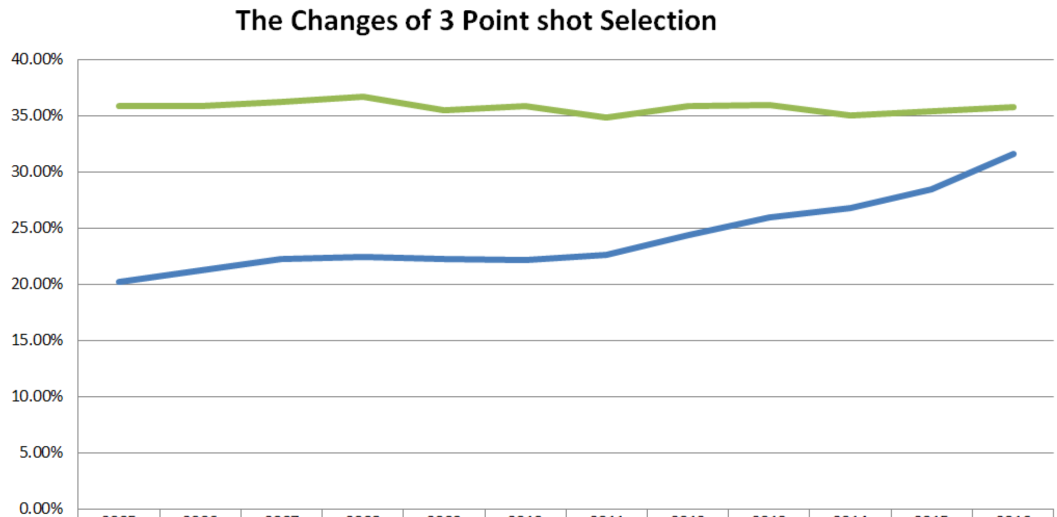
Figure 4 shows the proportion of MVPs by positions in the NBA history from the 1950s. Center positions are the biggest proportion of MVPs (29 times, 46.8%) and guard and forward are followed by 29.0% and 24.2% respectively.

In the 1960s and 1970s, MVPs were awarded 19 times to center positions. During that period, Bill Russell (208 cm, Center) and Wilt Chamberlain (216 cm, Center) shared the MVPs (4 times for Russell and 3 times for Chamberlain). Both players had a large impact in the league with their heights. In the 1980s, new players came up based on center-like heights but were not center position. Magic Johnson (206 cm, Guard) and Larry Bird (206 cm, Forward) are remembered to this day as the rivals of the century. Both players shared the MVPs equally 3 times each.
In the 1990s, Michael Jordan (198 cm, Guard) received the MVP 4 times although he retired in 1993-94 season and returned in 1994-95. Until 90s, the guard position's MVP was given solely to Jordan. The rest of the 6 MVPs were awarded to different positions.
In the 2000s, various positions for MVP were nominated and were equally distributed. Tim Duncan (211 cm, Center/Forward), LeBron James (203 cm, Forward), Steve Nash (191 cm, Guard) received the MVP 2 times respectively.
In the 2010's, there was no center position's MVP until the 2016-17 season. LeBron James (203 cm, Forward) and Stephen Curry (191 cm, Guard) received the MVP 2 times each.

MVPs were awarded to the guard and forward positions compared to the center in the past. After the 1980s, guard positions have 4 times MVPs every decade. During the 1980s to 90s, the players such as Magic Johnson and Michael Jordan, got the MVP 3 to 4 times but later several players in the guard position leaded the trend.

2.7 Current NBA Shooting Trend

Basketball is known to be a game of height where the higher rate of 2 point shooting to win. Nowadays, the NBA shooting trend indicates its transition.



	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
3P Selection Rate	20.23%	21.25%	22.22%	22.39%	22.19%	22.17%	22.56%	24.33%	25.93%	26.81%	28.42%	31.55%
3P Made Rate	35.83%	35.82%	36.20%	36.68%	35.45%	35.85%	34.86%	35.88%	35.98%	35.00%	35.38%	35.75%

Figure 5. 3PS (3 Point Selection) % and 3PM (3 Point Made) % in the NBA since 2005. (NBA STATS, 2017)

In Figure 5, the variations of 3PS% (3 Point Selection rate in total attempts) are shown. The 3PM% (3 Point Made rate) is stable between 34.8% and 36.68% for last 12 years. In contrast, the 3PS% has dramatically increased (total +11.32%) from 20.23% in 2005 to 31.55% in 2016. The growth rate of 3PS% was small up to 2011 but a sharp increase is shown since 2011 (2 - 3% incensements per year). In the NBA games, 3 point shots are made in 1 out of 3 shots in the current 2016.

2.8 The Higher 3 Point Shooting Team

The latest 2016-17 season winning and losing records are listed to check whether 3 point are connected to wins. The list below shows the top 5 teams that dominated the NBA league by the 3PA and 3PM frequency.

In Table 4, the upper table shows the top 5 teams in 3PA. The 3PS% (3 Point Selection %) shows the proportion of 3 point shots in total shot attempts. The Houston Rockets is the highest 3PA team in the league. The total 46.3% (40.2 times) in their offense option consists of 3 point shots. Close to 1 out of 2 shots is attempted as 3 point shots in a game. In the end, Houston Rockets was ranked 3rd in the western division. Cleveland, Boston and Golden State have over 35% 3 point proportion. The Brooklyn Nets attempted 3 point shots 30.5 times per game. The 3PM percentage was above the league average (35.75%). However, they were ranked 15th in the eastern division at the end of the season.

Table 4. Top 5 leader teams in the highest 3PA (upper) and 3PM% (bottom). (NBA STATS, 2017)
Blue: Highest 3PA (upper table), Highest 3PM% (bottom) / Red: Division Rankings

	Ranking	Team	3PS%	3PM	**3PA**	**3PM%**	**Division Ranking**	Wins	Loses
Highest 3PA teams	1	Houston Rockets	46.30%	14.3	**40.2**	35.7%	**W 3**	55	27
	2	Cleveland Cavaliers	39.90%	13	**33.9**	38.4%	**E 2**	51	31
	3	Boston Celtics	39.30%	12	**33.4**	35.9%	**E 1**	53	29
	4	Golden State Warriors	35.40%	12	**31.4**	38.3%	**W 1**	67	15
	5	Brooklyn Nets	36.70%	10.3	**30.5**	33.8%	E 15	20	62
Highest 3PM % teams	1	San Antonio Spurs	28.10%	9.2	23.5	**39%**	**W 2**	61	21
	2	Cleveland Cavaliers	39.90%	13	33.9	**38.4%**	**E 2**	51	31
	3	Golden State Warriors	35.40%	12	31.4	**38.3%**	**W 1**	67	15
	4	Indiana Pacers	27.20%	8.6	22.9	**37.6%**	E 7	42	40
	5	Sacramento Kings	29.10%	9	23.9	**37.6%**	W 12	32	50

In Table 4, the bottom table shows the highest 3PM rate and its team ranking. San Antonio Spurs was the most accurate 3PM% team (39%) with 23.5 attempts per game. The team was ranked 2nd in the western division. But San Antonio Spurs preferred 2 point shots (71.9%) than 3 point (28.1%). Cleveland and Golden State had 33.9 and 31.4 attempts per game which are much higher than the other teams in the list but have a higher percentage of 38.4% and 38.3% respectively. The two teams' rankings were high (ranked 2nd and 1st in each division). Indiana and Sacramento have high 3PM% of above 37.5%. However, the teams ranking results were low in the division. 3PA is not sufficiently correlated to the 3PM%.

Table 5. The two teams commonly ranked in 3PA and 3PM% from Table 4.

Team	3PS%	3PM	**3PA**	**3PM%**	**Ranking**	Win	Lost
Cleveland Cavaliers	39.90%	13	**33.9**	**38.4%**	**Eastern 2**	51	31
Golden State Warriors	35.40%	12	**31.4**	**38.3%**	**Western 1**	67	15

The two teams on Table 5, Golden State and Cleveland are listed from Table 4. The two teams have high 3PA (33.9 and 31.4) and high 3PM% (38.4% and 38.3%). 3PA is sufficiently correlated to the 3PM% with the two teams only in the list. The two teams on Table 5, competed in the final championships 3 times straight from 2014 to 2016.

2.9 The Trend Leader of NBA

Stephen Curry (191 cm, 86 kg) has been playing at the Golden State Warriors team since 2009-10 season. Recently, Stephen Curry is known to be the player who changed the basketball paradigm by 3 point shots. Most of the players that ranked high scores or named MVP were above average height. The playing style is assumed to be near the rim instead of far distances such as 3 point shot attempts. The highest 3 point made rate of only 45.4% while the highest field goal is 72.7% out of the total regular seasons in 71 years from 1946 to 2016. (Basketball-reference, 2017)

Figure 6. Kenny Smith, NBA TV analyst and retired NBA baller, defined the Curry as new Jordan. (Smith, 2016)

However, Curry changed the paradigm by a surprising 3 point shooting performance. He is called the new Jordan who is taking 3 point shots instead of dunks. The nickname was originally expressed by Kenny Smith, an ESPN NBA analyst, but now the name is commonly used in the media. In recent 3 years from 2014 to 2016, Curry won 2 championships and lost once in the finals. Curry and the Golden State team broke the record of 72 wins (10 loses) by Chicago Bulls that Michael Jordan played during the 1995-96 season and finally got 73 wins (9 loses) in the 2015-16 season.

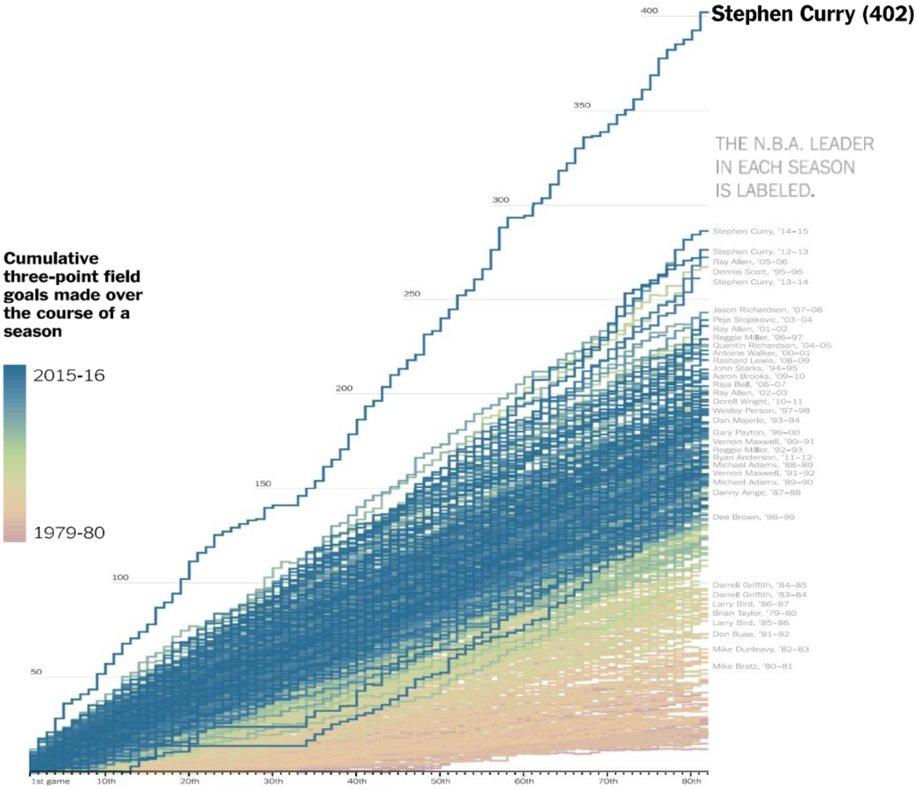


Figure 7. The highest number of 3 point shots, made in the year in NBA history. [5]

Figure 7 shows the amount of 3 point shots made per season from 1979-80. Well-known shooters are seen on the right side of which made the highest number of 3 point shots in a single season. Curry's record exceeds the other players and steadily reached new records. He defeated old records year after year for last 3 years in the 12-13, 14-15 season. He made 286 of 3 point shots but after 1 year, a new record was set at 402 in the 15-16 season.

Curry's records strongly affected the league resulting in more frequent 3 point shots up to 40% of shooting selection in a team.

[5] Image from (https://www.nytimes.com/interactive/2016/04/16/upshot/stephen-curry-golden-state-warriors-3-pointers.html)

3. Literature Review

In this chapter, literature of GIS methods used in basketball analysis is reviewed.
Basketball is a sport occurred in spatial and temporal background. GIS manipulating spatial and temporal data on phenomena enables new methods. Therefore, GIS is a suitable tool to manage basketball data and its analysis. Previous research is reviewed by two themes; offensive factors (shooting and movement tracking) and a defensive factor (match up matrix). This shows how recent basketball analyses have been conducted using GIS.

3.1 Goldsberry's Analysis of Basketball Data

"Basketball is a spatial sport" (Goldsberry, 2012, p. 1). Kirk Goldsberry previously worked as a GIS professor at Michigan and Harvard University and is currently working as vice president of strategic research at San Antonio Spurs team in the NBA (Sloansportsconference, 2017). Goldsberry introduced new methods for basketball analysis. He understood the spatial attributes of basketball and applied it into a model of GIS.

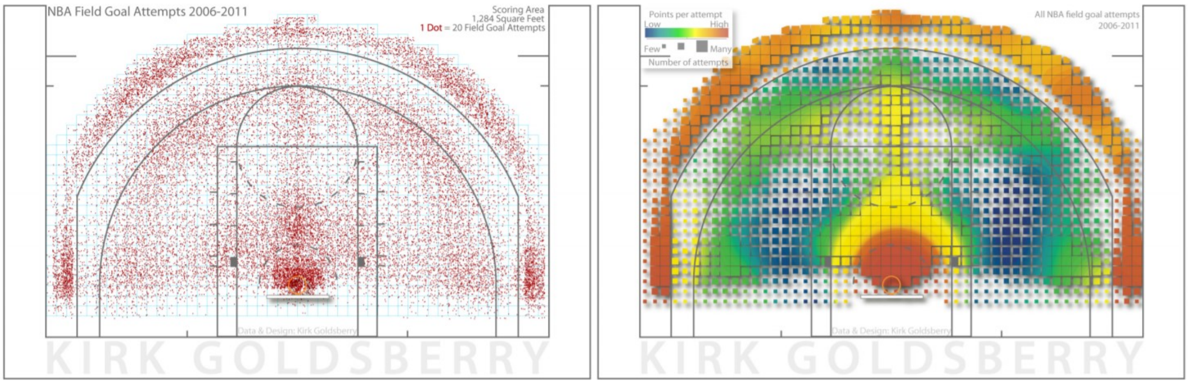


Figure 8. Field Goal Attempts (left) / Shot tendencies (right). (Goldsberry, 2012, p. 3)
Left: 1 dot indicates 20 field goal attempts in scoring area. (1,284 square feet)
Right: The square size shows the shooting attempts amount, the bigger the more attempts.
The colors show the average points per attempt in each location.
(The orange color indicates more points per attempt while blue indicates fewer points per attempt.)

In 2012, at the MIT (Massachusetts Institute of Technology) Sloan Sports Analytics Conference, Goldsberry announced *"Court Vision: New Visual and Spatial Analytics for the NBA"* (Goldsberry, 2012). Goldsberry defined the density of shot attempts during 2006 to 2011 in the NBA. In Figure 8, one dot indicates attempts of 20 shots. On the right side, the square size represents the shot attempts amount and also the color represents the points per attempt of all NBA players in the periods. From the size of the square, shot tendency is realized.

$$Spread = \sum_{ij \in SA} FGA_{ij}$$

Spread = Total spatial spread of player across all scoring cells
FGA_{ij} = 1, if at least one field goal has been attempted in cell i, 0 if not
SA = Scoring area consisting of 1,284 scoring cells

Figure 9. Spread indicates the player's shooting territory. (Goldsberry, 2012, p. 3)
Spread shows the player's shooting territory as sum of number of square (if the attempts is over 1 in the square, > 1).

Kirk Goldsberry defined the "*Spread*" concept to describe *"the overall size of the player's shooting territory"*. The "*Spread*" index is calculated by sum of the square when the attempts are over 1 in the square. As the player takes shots in long distance on par with in various directions, the "*Spread*" value increases. In Goldsberry's research, the high value is appeared in SG (Shooting Guard) position; Kobe

Bryant (1,071 of 1,284 which is 83.4%), but low value is appeared in C (Center) position; Dwight Howard (23.8%). Center tends to play in near the rim in limited distance that spread value is relatively smaller (Goldsberry, 2012).

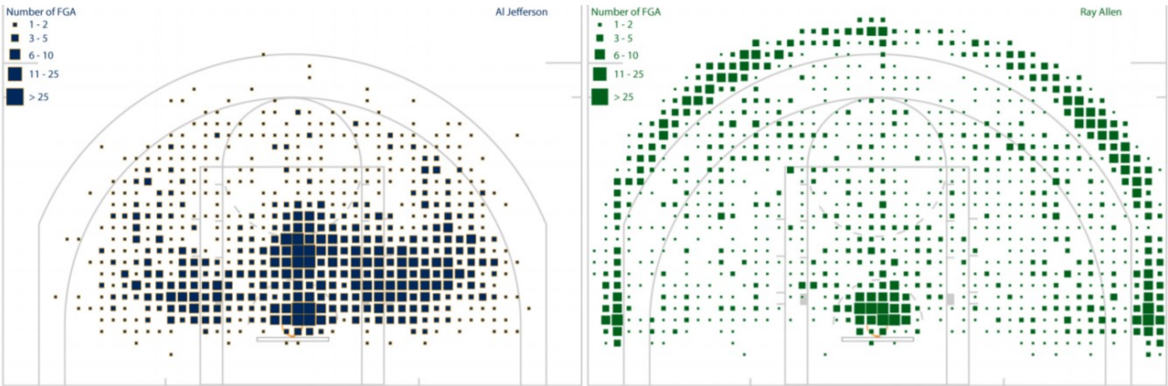


Figure 10. Spread variable visualized for Al Jefferson and Ray Allen. (Goldsberry, 2012, p. 4)
The square shows shooting attempts amount by 5 different categories (1-2 is the lowest attempts range, > 25 is the highest attempts range).

Furthermore, Figure 10 shows the two players' spread visual depiction. During the study period, Al Jefferson (left) shows 46.3% spread variable while Ray Allen shows 74.1%.
Figure 10 displays the shooting tendency of the players. Al Jefferson was positive in central area in 2 point area while Ray Allen was positive behind 3 point arc line. Despite Jefferson had 400 more attempts than Allen, *"Spread"* variable shows higher value of only shooting range (Goldsberry, 2012)

$$Range = \sum_{ij \in SA} PPA_{ij}$$

$\boldsymbol{Range}$ = Effective shooting range of player across all scoring cells
PPA_{ij} = 1, if points per attempt is > 1 in cell i, 0 if not
SA = Scoring area consisting of 1,284 scoring cells

Figure 11. Range indicates the efficiency of the shot made in the SA (Scoring Area). (Goldsberry, 2012, p. 4)
The Range index shows the player's efficiency of shooting range as sum of number of square (if the PPA in square > 1).

As the "*Spread*" variable has limitations to distinct the shooting efficiency in the location, Goldsbery defined the "*Range*" concept. The "*Range*" concept is calculated by sum of the number of the square if the PPA (Points Per Attempt) is at least 1 in the square. The PPA (Points Per Attempt) is counted as the criteria of efficiency in the location for the shots. For example, if two points made per one attempt, the PPA value is two. If a shot is made at a low percentage, such as three points made by five attempts, the PPA value is 0.6. The PPA is lower than 1, the PPA is not counted. (Goldsberry, 2012)

Thus, the "*Range*" concept shows the practical shooting range of the players that the shots are made in reasonable percentage in the cell. (Goldsberry, 2012)

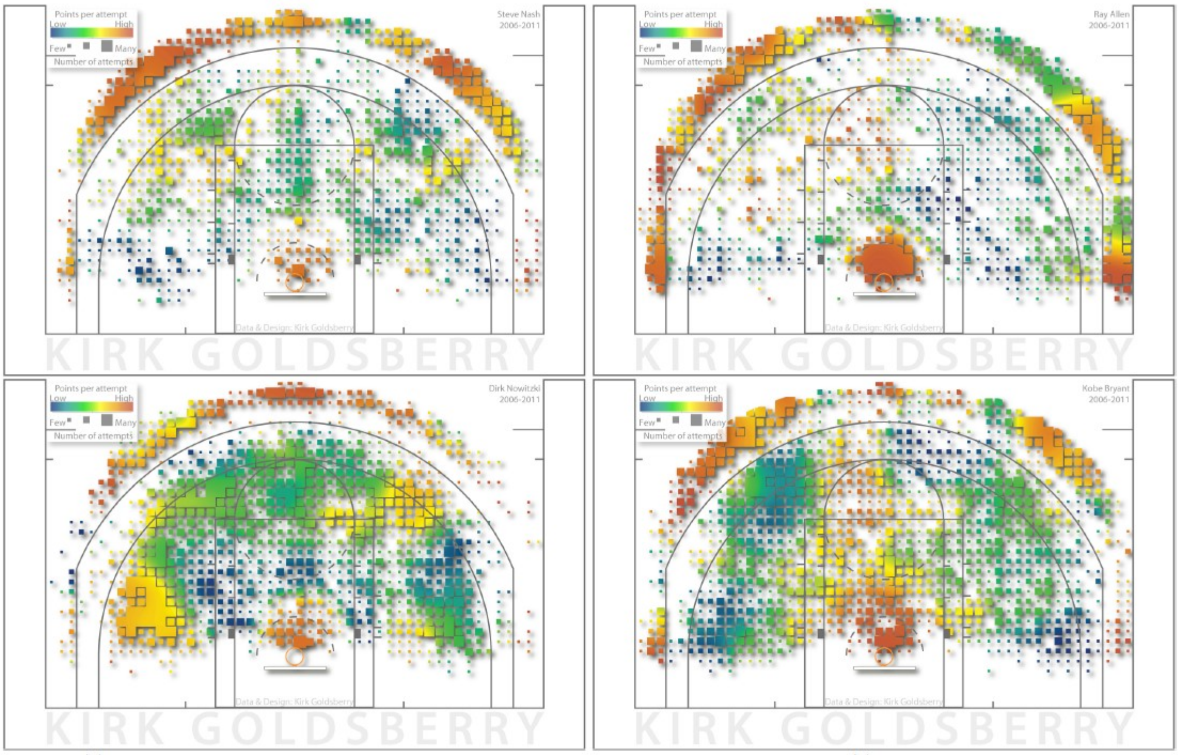


Figure 12. Range depiction of 4 different players in the NBA. (Goldsberry, 2012, p. 5)
The square size shows the shooting attempts amount while the color shows the points per attempt.
The Figure 12 shows the top 4 players in Range % but the range is achieved from different spatial tendencies.

Figure 12 shows "*Range*" visual depiction of four different players in the NBA. Steve Nash has 406 shooting cells (if PPA is over 1) that 31.6%, Ray Allen (30.1%), and Kobe Bryant (29.8%), Dirk Nowitzki (29.0%). The players show close range proportion but the way of achieving PPA is different. For example, Nash and Allen were positive in 3 point line while Nowitzki was positive in diverse locations in 2 point area. (Goldsberry, 2012)

3.2 Motion Tracking into NBA

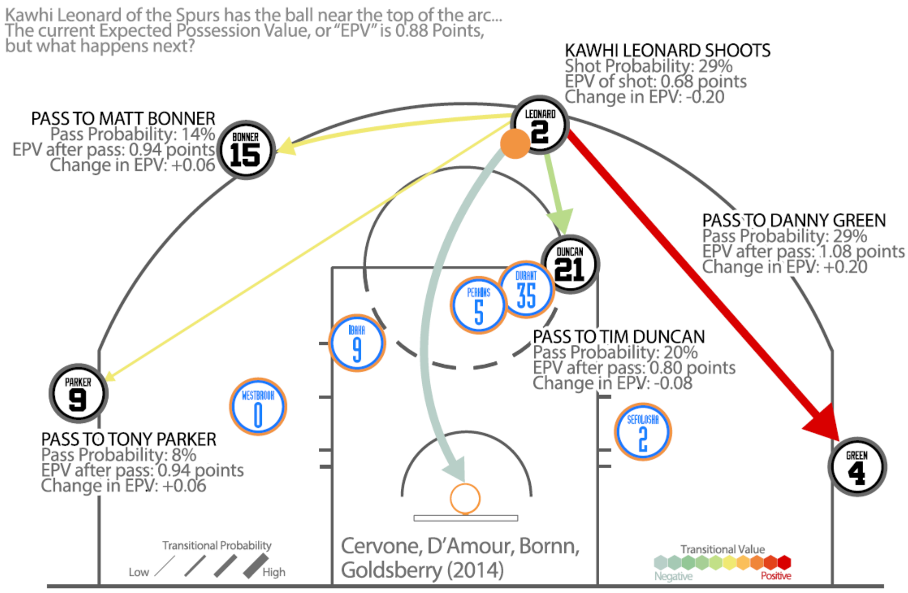


Figure 13. Diagram of EPV (Expected Possession Value) to evaluate decision moment. (Cervone, et al., 2014, p. 2)
"Diagram of EPV as a weighted average of the values of the ball carrier's (Leonard's) decisions and the probability of making each decision."

Shooting is considerable as the most important factor in basketball. As the basketball is relative scoring game, the shooting to score is crucial. In order to score, the better decision making is required. Figure 13 shows the "*EPV*" (Expected Possession Value) in which calculates the expected point from the offense using player-tracking data. Kawhi Leonard carry the ball, on Figure 13, and "*EPV*" value is computed by all the possibility of selections such as pass, dribble, shooting based on weighted average of the ball handler. Each decision is evaluated that the player makes and what the better decision at the moment is calculated (Cervone, et al., 2014).

This computation is possible by motion tracking technologies. Such as "*SportVu*", the official provider of motion camera for NBA league provides motion information including direction, speed, action types, and other extra information. The special cameras are used for collecting the data at a rate of 25 times per second and save the action into figures almost directly from the record. Now cameras are able to capture core movements of the ball and the trajectories of the players on the court. (SportVU, 2017).

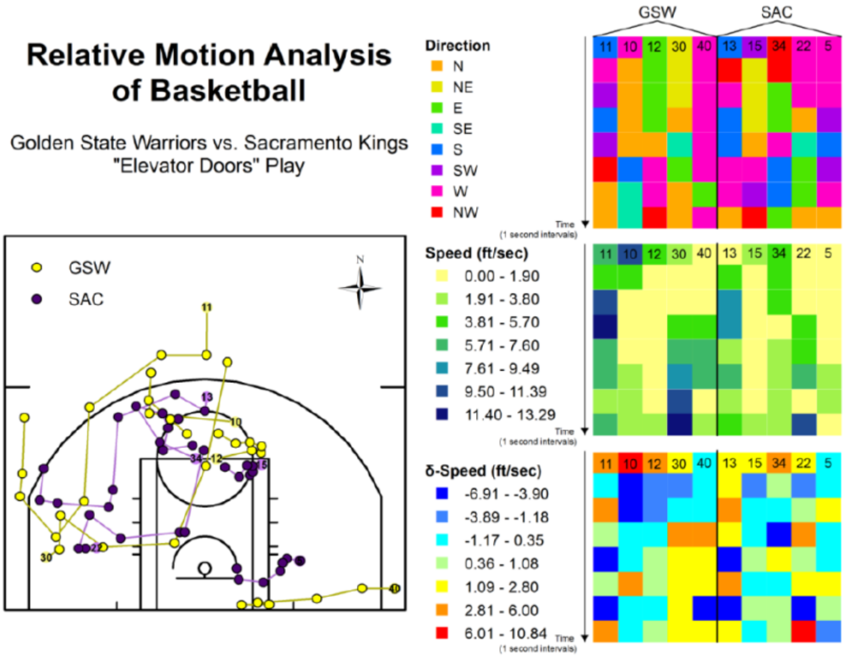


Figure 14. The REMO (Relative Motion) matrices in basketball analysis. (Jensen, 2014, p. 41)
Figure 14 shows player's trajectories on left and player's direction, speed and changes in ft/sec on right.
Each color shows different directions and its speed. On the right, the matrices show the player's trajectories in colors.

The "*REMO" (Relative Motion)* analysis was firstly introduced for soccer players and its movement type discovery (Laube, et al., 2005). Daniel Jensen adopted this method into basketball analysis. Figure 14 shows relative motion matrices on players' direction, speed and changes in ft/sec (Jensen, 2014). The motion is investigated when team Golden State against team Sacramento Kings. Golden state was in action of "*door play*" for offense and right rectangles shows the player's action type in direction, speed and change in speed based on its movement trajectories. The best benefit of using "*REMO"* is that all the movement can be digitized and it is clearly seen in visual and readable form. (Jensen, 2014).

3.3 Defensive Factor

As the basketball is a scoring game, previous analyses of basketball have been focused on offensive factors because the end of offense results by the shots. However, basketball consists of offense and defense. If a shot is blocked by defense, it is another name of score.

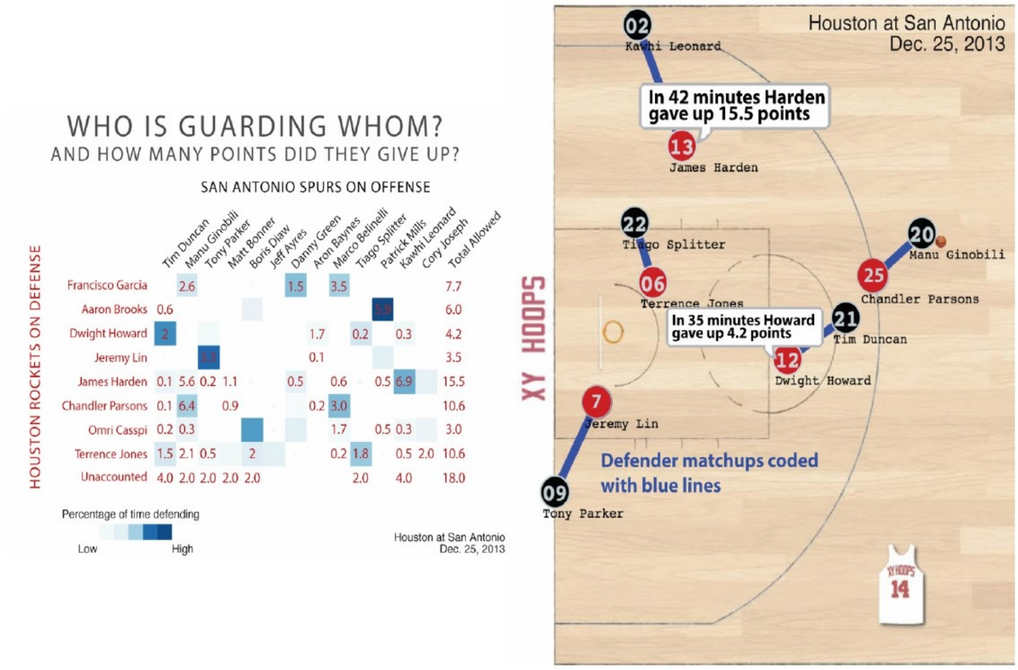


Figure 15. Matchup matrix that shows the point value according to the defense matched. (Franks, et al., 2015, p. 2)
Matchup matrix for the Houston at San Antonio game on Dec 25, 2013.
In the matrix, *fast break* or *put-backs* are assigned into "unaccounted" as it is not a normal defense situation.
On the left, this shows that Houston Rockets are on defense and how many points are given to the opponents matched. Also the color shows percentage of time defending that blue indicates high percentage of time depending while white indicates low percentage of time depending. In this game James Harden gave the maximum points to opponent (15.5 points)

Figure 15 shows the matrix that computes the point values according to the player matched during the game. The matrix evaluates the point according to match up opponent. The *fast break* or *put-backs* are not considered as a normal defense situation and the cases are assigned into "unaccounted" category. The color shows the percentage of time depending. The blue shows high percentage of time depending while white shows low percentage of time depending. In the game, James Harden is considered as the worst defensive player by giving the highest points to the opponents matched up. (Franks, et al., 2015).

The defensive ability has been measured by visible factors such as *steal, rebound* or *block* or so. But this matchup matrix shows practical defensive ability in a way of calculating points that gives opponent when matched up. The matchup matrix presents the defensive abilities into a visible figure. (Franks, et al., 2015).

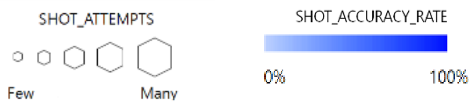
4. Theory

In this fourth chapter, theoretical work is introduced. "*Shot tendency*" theory which is to investigate the shooting trend on specific moment is addressed. "*Spread*" value index is additionally addressed to explore the differences in each year. Secondly, "*Space-time cube*" and "*Emerging hot spot*" theories are described. "*Emerging hot spot*" theory enables investigating the data by temporal changes in historical comparison based on "*Space-time cube*" theory. The "*Emerging hot spot*" theory was previously introduced to crime analysis to find out newly emerged hot spots that for a further surveillances to prevent possible future crimes. "*Emerging hot spot*" theory is perfectly fitted to find out the shooting emerging hot spot in basketball analysis.

4.1 Shot Tendency

In order to find and compare specific moment's shooting trend pattern, "*Shot tendency*" theory is used that created from Goldsberry in 2012. This method shows the effective way of visualizing the shot data and presents the location of where the most shots attempted and its accuracy with hexagon's size and graduated color. "*Shot tendency*" describes the spatial shooting dependency of the players.

Firstly, the half court needs to be covered with hexagon fishnet. Each hexagon counts that how many shots are attempted and made within the location. The accuracy rate is calculated by the each count. The sizes of hexagon are varied according to the amount of shot attempts that the bigger hexagon size indicates the more attempts. The graduated color presents the accuracy rate in the location of the hexagon.



$$Spread = \sum_{ij \in SA} FGA_{ij}$$

Spread = Total spatial spread of player across all scoring cells
FGA_{ij} = 1, if at least one field goal has been attempted in cell i, 0 if not
SA = Scoring area consisting of 1,284 scoring cells

Figure 16. Spread indicates the player's shooting territory. (Goldsberry, 2012, p. 3)
SHOT_ATTEMPTS: The Hexagon size shows the shot attempts amount. The bigger size indicates the more attempts.
SHOT_ACCURACY_RATE: The graduated color shows the shot accuracy. The stronger color indicates the higher accuracy.
Spread: This is from Goldsberry (2012), the player's shooting territory as sum of number of hexagon (if attempts in the hexagon > 0).

Figure 16 shows that the shot attempts by size and shot accuracy rate by graduated color with "*Spread*" concept. The *SA (Scoring Area)* is consisted of 1,284 cells because more than 98% field goal attempts are occurred in 1,284 ft^2 (the whole court size is 4,700 ft^2) from Goldsberry's research in 2012.

According to "*Spread*" concept, the total shooting territory of the player is calculated. The "*Spread*" consists of the sum of the number of hexagon if *FGA (Field Goal Attempts)* is over 0 in the hexagon. As the player takes shots in long distance on par with in various directions, the "*Spread*" value increases.

4.2 Emerging Hot Spot Analysis

"*Emerging hot spot*" analysis is to find out newly emerged hot spot based on historical trend comparison of the events. The analysis results different 16 types of hot spot or cold spot trends according to the intensity level of the events. This is relevant by user's neighborhood distance and time span that decides the analysis characteristics. The "*Emerging hot spot*" analysis is developed and conducted by Esri's ArcGIS. Thus, this chapter is given, fully based on online tool reference guidance from Esri's website:
(http://desktop.arcgis.com/en/arcmap/latest/tools/space-time-pattern-mining-toolbox/an-overview-of-the-space-time-pattern-mining-toolbox.htm) (Esri, 2017).

The analysis is conducted by "Space Time Pattern Mining Tools". The tools are consisted of 3 sub tools that "Create space time cube", "Emerging hot spot analysis" and "Local outlier analysis". Utilities are made of 2 sub tools that "Visualize space time cube in 2D and 3D".

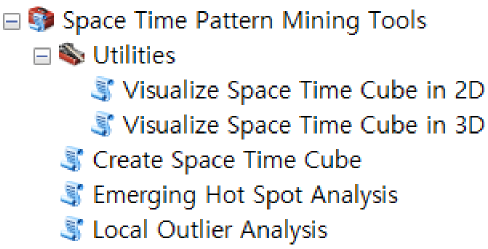


Figure 17. Space Time Pattern Mining Tools in Esri's ArcGIS.

The Emerging hot spot analysis is processed and described as below.

General processes for *Emerging hot spot* analysis

1. Space-Time Cube
   - Time step interval
   - Distance interval
2. Emerging Hot Spot Analysis
   - Neighborhood time span
   - Neighborhood distance

4.2.1 Space-Time Cube

Space-time cube is created by aggregating the event points in a *netCDF (Network Common Data Form*) format to summarize and store the data set in multi-dimensional structures. The input data for the cube should be points having time indication data in the location such as crime incidents, fire accidents in date format to reflect time changes. (Esri, 2017)

Each bin has basically 2 types of reference that presents the location (x, y) and the time (t) as unique ID. The ID is shared in row and column which makes the spatial and temporal analysis possible at the same time. Figure 18 indicates that how the each ID is shared. The same space ID is shared in yellow and the time ID is shared in green. (Esri, 2017)

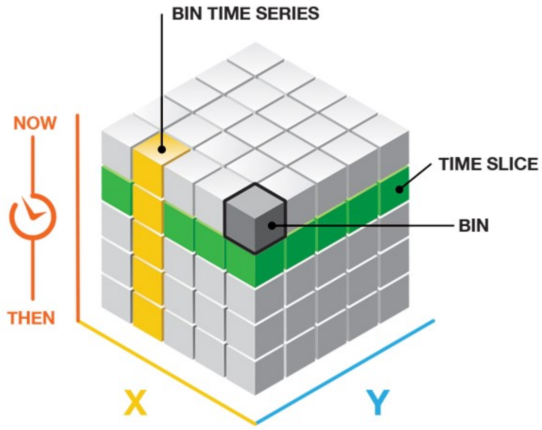


Figure 18. The structure of space-time cube. [6] **(Esri, 2017)**
Each bin has basically 2 types of reference that presents the location (x, y) and the time (t) as unique ID. The same space ID is shared in yellow and the time ID is shared in green that makes the spatial and temporal analysis possible at the same time.

The space-time cube's characteristic is relevant by 2 parameters: Time step interval and distance interval. Time step interval decides the time partition span to aggregate the points by user's selection such as one-day, one-week or one-year etc. If the data is occurred during only specific periods, the periods may bias the result. For instance, the NBA league is played from Oct to April for regular season. The rest of the year has significantly no data. The space-time cube recognizes the location as none-events occurred during the periods and results in biased way. On the case, reference date needs to be set to avoid the bias. Distance interval decides the size of the space-time bins to aggregate points. For example, the distance interval is decided as 10 meters, 10 m x 10 m fishnet bin (into a form of rectangle or hexagon) is created and the bin counts the event within the size. The proper size of the bins is crucial. Otherwise the bins having zero count will be too many if the size is too small that causes the biased trend. (Esri, 2017)

4.2.2 Emerging Hot Spot Analysis

Once the space-time cube is created, the "*Emerging hot spot*" analysis is possible. The "*Emerging hot spot*" analysis is to find out newly emerged hot spot based on historical trend comparison of the events. In this process, the neighborhood (combining the cube bins by certain time interval and distance) is used to find out hot spot by *Getis-Ord Gi* statistics* for each bin. The resultant values (z-score: standard deviation, p-value: probability) indicates spatially high cluster or low cluster. Lastly, *Mann-Kendall trend* test evaluates the hot spot by a rank correlation analysis and identify the hot spot into a trend of emerging hot spot such as *new, consecutive, intensifying, persistent, diminishing, sporadic, oscillating, historical hot/cold spot or no pattern.* (Esri, 2017)

In the process, neighborhood distance and neighborhood time step are adjustable which makes various spatial and temporal dependencies for the analysis. The space-time cube has its own unique ID for the location, time and also the summary of the events (i.e. count and possible summary filed; sum, mean etc). These attributes are used to compare historical intensity of the events for emerging hot spot.
The neighborhood time step interval defines the time partition to aggregate. This time step intervals are calculated as 1 by the program if the value is not set. For the precise analysis or according to various purposes, the value should be adjusted with proper intervals. Otherwise, the time step causes

[6] Image from (http://desktop.arcgis.com/en/arcmap/latest/tools/space-time-pattern-mining-toolbox/create-space-time-cube.htm)

the different trend when the intervals are too small or too large. The intervals can be selected from seconds, minutes, hours, days, weeks, and months to years.

The neighborhood distance for the size of location partition to aggregate also needs to be adjusted. As the same as the time step interval, the program calculates the distance based on geographical distribution of the events if the value is not set. The neighborhood distance interval aggregates the data within the distance. (Esri, 2017)

Pattern Name	Definition
No Pattern Detected	Does not fall into any of the hot or cold spot patterns defined below.
New Hot Spot	A location that is a statistically significant hot spot for the final time step and has never been a statistically significant hot spot before.
Consecutive Hot Spot	A location with a single uninterrupted run of statistically significant hot spot bins in the final time-step intervals. The location has never been a statistically significant hot spot prior to the final hot spot run and less than ninety percent of all bins are statistically significant hot spots.
Intensifying Hot Spot	A location that has been a statistically significant hot spot for ninety percent of the time-step intervals, including the final time step. In addition, the intensity of clustering of high counts in each time step is increasing overall and that increase is statistically significant.
Persistent Hot Spot	A location that has been a statistically significant hot spot for ninety percent of the time-step intervals with no discernible trend indicating an increase or decrease in the intensity of clustering over time.
Diminishing Hot Spot	A location that has been a statistically significant hot spot for ninety percent of the time-step intervals, including the final time step. In addition, the intensity of clustering in each time step is decreasing overall and that decrease is statistically significant.
Sporadic Hot Spot	A location that is an on-again then off-again hot spot. Less than ninety percent of the time-step intervals have been statistically significant hot spots and none of the time-step intervals have been statistically significant cold spots.
Oscillating Hot Spot	A statistically significant hot spot for the final time-step interval that has a history of also being a statistically significant cold spot during a prior time step. Less than ninety percent of the time-step intervals have been statistically significant hot spots.
Historical Hot Spot	The most recent time period is not hot, but at least ninety percent of the time-step intervals have been statistically significant hot spots.
New Cold Spot	A location that is a statistically significant cold spot for the final time step and has never been a statistically significant cold spot before.
Consecutive Cold Spot	A location with a single uninterrupted run of statistically significant cold spot bins in the final time-step intervals. The location has never been a statistically significant cold spot prior to the final cold spot run and less than ninety percent of all bins are statistically significant cold spots.
Intensifying Cold Spot	A location that has been a statistically significant cold spot for ninety percent of the time-step intervals, including the final time step. In addition, the intensity of clustering of low counts in each time step is increasing overall and that increase is statistically significant.
Persistent Cold Spot	A location that has been a statistically significant cold spot for ninety percent of the time-step intervals with no discernible trend, indicating an increase or decrease in the intensity of clustering of counts over time.
Diminishing Cold Spot	A location that has been a statistically significant cold spot for ninety percent of the time-step intervals, including the final time step. In addition, the intensity of clustering of low counts in each time step is decreasing overall and that decrease is statistically significant.
Sporadic Cold Spot	A location that is an on-again then off-again cold spot. Less than ninety percent of the time-step intervals have been statistically significant cold spots and none of the time-step intervals have been statistically significant hot spots.
Oscillating Cold Spot	A statistically significant cold spot for the final time-step interval that has a history of also being a statistically significant hot spot during a prior time step. Less than ninety percent of the time-step intervals have been statistically significant cold spots.
Historical Cold Spot	The most recent time period is not cold, but at least ninety percent of the time-step intervals have been statistically significant cold spots.

Figure 19. 16 Types of trend defined in emerging hot spot analysis. (Esri, 2017)

5. Materials and Methods

In the fifth chapter, the methodological framework is explained. Based on the theoretical framework, the practical framework is addressed. Each process and its adjustable parameters are described.

The NBA history is started from 1946 and the league is continued over 70 years until the current (2017). In order to find out the recent shooting trend, last 12 years of NBA shooting data is used from 2005 to 2016. In the first part, data preparation method is addressed. Detailed explanation for importing data from official NBA statistics site (stats.nba.com) to ArcGIS using Python code is addressed. In second part, specific moment's data is investigated using "*Shot tendency*" and "*Spread*" from Kirk Goldsberry (2012). In the third part, the data is aggregated into the space-time cube and "*Emerging hot spot*" is examined.

5.1 Data Preparation

The NBA official website (nba.com) provides a bunch of information for players, teams and fans such as game schedule, news, video, standings etc. From the sub categories in the website (stats.nba.com), NBA players' data in JSON format is available by URL request.


```
{"resource":"shotchartdetail","parameters":{"LeagueID":null,"Season":"2015-16","SeasonType":"Regular
Season","TeamID":0,"PlayerID":201939,"GameID":null,"Outcome":null,"Location":null,"Month":0,"SeasonSegment":null,"DateFrom":null,"DateTo":null,"OpponentTeamID":
0,"VsConference":null,"VsDivision":null,"Position":null,"RookieYear":null,"GameSegment":null,"Period":0,"LastNGames":0,"ClutchTime":null,"AheadBehind":null,"Poi
ntDiff":null,"RangeType":null,"StartPeriod":null,"EndPeriod":null,"StartRange":null,"EndRange":null,"ContextFilter":"","ContextMeasure":"FGA"},"resultSets":
[{"name":"Shot_Chart_Detail","headers":
["GRID_TYPE","GAME_ID","GAME_EVENT_ID","PLAYER_ID","PLAYER_NAME","TEAM_ID","TEAM_NAME","PERIOD","MINUTES_REMAINING","SECONDS_REMAINING","EVENT_TYPE","ACTION_TYP
E","SHOT_TYPE","SHOT_ZONE_BASIC","SHOT_ZONE_AREA","SHOT_ZONE_RANGE","SHOT_DISTANCE","LOC_X","LOC_Y","SHOT_ATTEMPTED_FLAG","SHOT_MADE_FLAG","GAME_DATE","HTM","VT
M"],"rowSet":[["Shot Chart Detail","0021500003",6,201939,"Stephen Curry",1610612744,"Golden State Warriors",1,10,56,"Made Shot","Cutting Finger Roll Layup
Shot","2PT Field Goal","Restricted Area","Center(C)","Less Than 8 ft.",3,12,31,1,1,"20151027","GSW","NOP"],["Shot Chart Detail","0021500003",9,201939,"Stephen
Curry",1610612744,"Golden State Warriors",1,10,29,"Missed Shot","Jump Shot","3PT Field Goal","Above the Break 3","Left Side Center(LC)","24+
ft.",26,-176,195,1,0,"20151027","GSW","NOP"],["Shot Chart Detail","0021500003",14,201939,"Stephen Curry",1610612744,"Golden State Warriors",1,9,34,"Missed
Shot","Layup Shot","2PT Field Goal","Restricted Area","Center(C)","Less Than 8 ft.",2,20,9,1,0,"20151027","GSW","NOP"],["Shot Chart
Detail","0021500003",19,201939,"Stephen Curry",1610612744,"Golden State Warriors",1,9,13,"Made Shot","Jump Shot","3PT Field Goal","Above the Break 3","Left Side
Center(LC)","24+ ft.",27,-197,193,1,1,"20151027","GSW","NOP"],["Shot Chart Detail","0021500003",36,201939,"Stephen Curry",1610612744,"Golden State
Warriors",1,7,11,"Made Shot","Running Layup Shot","2PT Field Goal","Restricted Area","Center(C)","Less Than 8 ft.",0,-4,8,1,1,"20151027","GSW","NOP"],["Shot
Chart Detail","0021500003",38,201939,"Stephen Curry",1610612744,"Golden State Warriors",1,6,45,"Made Shot","Jump Shot","3PT Field Goal","Above the Break
3","Center(C)","24+ ft.",25,-11,259,1,1,"20151027","GSW","NOP"],["Shot Chart Detail","0021500003",46,201939,"Stephen Curry",1610612744,"Golden State
Warriors",1,5,50,"Made Shot","Jump Shot","2PT Field Goal","Mid-Range","Left Side Center(LC)","16-24 ft.",22,-117,188,1,1,"20151027","GSW","NOP"],["Shot Chart
Detail","0021500003",55,201939,"Stephen Curry",1610612744,"Golden State Warriors",1,5,3,"Made Shot","Jump Shot","3PT Field Goal","Above the Break 3","Left Side
Center(LC)","24+ ft.",28,-179,228,1,1,"20151027","GSW","NOP"],["Shot Chart Detail","0021500003",68,201939,"Stephen Curry",1610612744,"Golden State
Warriors",1,3,57,"Made Shot","Step Back Jump shot","3PT Field Goal","Above the Break 3","Right Side Center(RC)","24+ ft.",25,89,239,1,1,"20151027","GSW","NOP"],
["Shot Chart Detail","0021500003",82,201939,"Stephen Curry",1610612744,"Golden State Warriors",1,3,6,"Missed Shot","Jump Shot","3PT Field Goal","Above the Break
3","Left Side Center(LC)","24+ ft.",27,-156,232,1,0,"20151027","GSW","NOP"],["Shot Chart Detail","0021500003",88,201939,"Stephen Curry",1610612744,"Golden State
Warriors",1,2,45,"Made Shot","Driving Finger Roll Layup Shot","2PT Field Goal","Restricted Area","Center(C)","Less Than 8
ft.",2,-19,21,1,1,"20151027","GSW","NOP"],["Shot Chart Detail","0021500003",116,201939,"Stephen Curry",1610612744,"Golden State Warriors",1,0,45,"Made
Shot","Driving Layup Shot","2PT Field Goal","Restricted Area","Center(C)","Less Than 8 ft.",2,12,18,1,1,"20151027","GSW","NOP"],["Shot Chart
Detail","0021500003",127,201939,"Stephen Curry",1610612744,"Golden State Warriors",1,0,1,"Missed Shot","Jump Shot","3PT Field Goal","Above the Break 3","Left
Side Center(LC)","24+ ft.",39,-186,345,1,0,"20151027","GSW","NOP"],["Shot Chart Detail","0021500003",208,201939,"Stephen Curry",1610612744,"Golden State
Warriors",2,6,3,"Missed Shot","Jump Shot","3PT Field Goal","Above the Break 3","Right Side Center(RC)","24+ ft.",24,164,185,1,0,"20151027","GSW","NOP"],["Shot
Chart Detail","0021500003",221,201939,"Stephen Curry",1610612744,"Golden State Warriors",2,4,54,"Made Shot","Fadeaway Bank shot","2PT Field Goal","Mid-
Range","Left Side(L)","8-16 ft.",12,-96,85,1,1,"20151027","GSW","NOP"],["Shot Chart Detail","0021500003",253,201939,"Stephen Curry",1610612744,"Golden State
Warriors",2,2,11,"Missed Shot","Jump Shot","3PT Field Goal","Above the Break 3","Right Side Center(RC)","24+ ft.",24,166,174,1,0,"20151027","GSW","NOP"],["Shot
Chart Detail","0021500003",261,201939,"Stephen Curry",1610612744,"Golden State Warriors",2,1,16,"Missed Shot","Pullup Jump shot","3PT Field Goal","Above the
Break 3","Right Side Center(RC)","24+ ft.",25,204,151,1,0,"20151027","GSW","NOP"],["Shot Chart Detail","0021500003",277,201939,"Stephen
Curry",1610612744,"Golden State Warriors",2,0,0,"Made Shot","Driving Finger Roll Layup Shot","2PT Field Goal","Restricted Area","Center(C)","Less Than 8
ft.",1,-9,8,1,1,"20151027","GSW","NOP"],["Shot Chart Detail","0021500003",302,201939,"Stephen Curry",1610612744,"Golden State Warriors",3,10,7,"Made
```


Figure 20. The raw data from stats.nba.com.

In order for the shooting pattern trend analysis, the shot chart data having geographical references in X, Y coordinates is used. In addition, the shot chart data comes with 20 extra information that explain the moment in the location such as DATE, GAME ID, PLAYER, TEAM, SHOT_TYPE, SHOT_ZONE, SHOT_MADE FLAG, 3POINT_FLAG, PERIOD & REMINING TIME etc. The data is public and detailed information of players is accessible online.

General Process of Data Preparation.

1. Call the URL request for the raw input data (Specify the season: "2005-2016" and the player: " "; this indicates all players)
2. Get the raw data and save the data as Python file
3. Create a GDB(Geo Data Base) in ArcGIS(10.5) with the projection
   - GDB file name: NBA_2005_16
   - Projection: WGS_1984_Web_Mercator_Auxiliary_Sphere
   - Unit: Meter (the unit for the analysis in ArcGIS is meter without conversion for convenience)
4. Create feature class
   - Feature class: NBA_2005_16.shp
5. Create fields which are necessary.
   Feature Class Fields: (the same fields from the raw data)
   ['Shape','GRID_TYPE','GAME_ID','GAME_EVENT_ID','PLAYER_ID','PLAYER_NAME','TEAM_ID','TEAM_NAME','PERIOD','MINUTES_REMAINING','SECONDS_REMAINING','EVENT_TYPE','ACTION_TYPE','SHOT_TYPE','SHOT_ZONE_BASIC','SHOT_ZONE_AREA','SHOT_ZONE_RANGE','SHOT_DISTANCE','LOC_X','LOC_Y','SHOT_ATTEMPTED_FLAG','SHOT_MADE_FLAG','THREE','DATE']
6. Populates the data into the fields categorized.
   - Total 2,380,929 shot information loaded from 2005-06 to 2016-17 season
7. Background court image is covered.
   - Made by author (using illustration program)

The shot chart data makes the shooting analysis possible. Previously, the NBA used to provide player's trajectories data but with some reason the trajectory is not available. Instead, recent private companies such as "*SportVu*", provide advanced information on the court for the NBA league including movement trajectories, action types and customized information. The "*SportVu*" installs special cameras on the court for collecting the data at a rate of 25 times per second and save it into figures almost directly during the record (SportVU, 2017).
In order to use the raw data in ArcGIS (10.5), the data needs to be imported in ArcGIS. Recently, ArcGIS enhanced its functions and the one of advanced functions is the usage of Python in the program. This manages huge amount of data in less effort and less time consumption than before. Thus, using Python code is the advantage in the program and most of Python code is coming from the GAVIN REHKEMPER's post (Written by Gregory Brunner), but some of them are re-edited according to the situation. (Brunner, 2015) The full Python code is available on the Appendix 1.

5.2 Data Visualization

Once the raw data is called, the data needs to be saved. By the Python code (available on *Appendix 1*), the raw data is loaded to ArcGIS under the GDB (Geo Data Base) file as a feature class in shape file. This shape file makes the data into points and the data is displayed according to the X, Y coordinates on the court image (the image is made by author with illustration program).

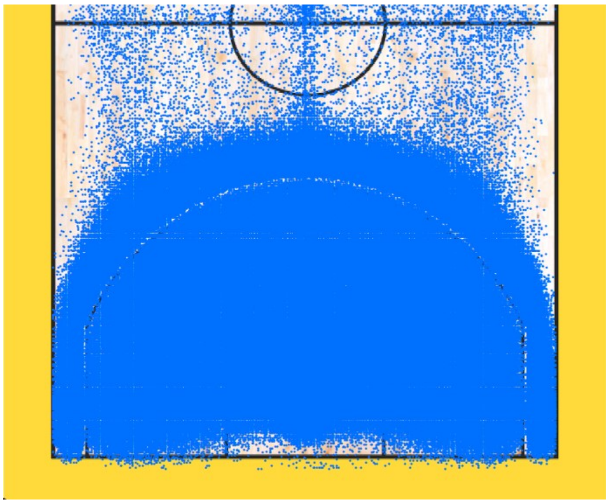


Figure 21. The raw NBA shooting data in 2005 to 2016 loaded to ArcGIS (10.5).

Figure 21 shows the raw NBA shooting data in ArcGIS (10.5). During the study periods from 2005 to 2016, total 2,380,929 shots are taken which indicates yearly average that is 198,410.8 shots. In order to focus on the most scoring area, only the half court area is studied.

5.3 Shot Tendency

In this research, the half court is divided into 1,200 cells with bigger hexagon size (width: 15 ft, height: 25.52 ft) than Goldsberry's cell done in 2012 (width: 10 ft, height: 10 ft as a square).
The shooting trend is investigated by the "*Shot tendency*" in the same way done in the Goldsberry's research (2012) and the trend is compared in the same location of the court for each year. Secondly, the top range area is extracted from the "*Shot tendency*" method for clear comparison based on "*Spread*" concept.

Shot tendency

1. Covering the area with hexagons (width: 15 m, height: 25.42 m)
2. Aggregating the shooting points into each hexagon by spatial join tool
3. Converts the hexagon to points by feature to point tool
4. Display the points to shot tendency by multiple attributes in layer properties

First of all, the raw data should be ready which is already loaded to ArcGIS in shape file format and displayed according to the X,Y coordinates. The fishnet consists of hexagons is created for aggregating the shot points over the area. In the research, a tool that makes hexagon fishnet over the study area automatically is downloaded and used.[7] The hexagon's width is set to 15 m and height is automatically calculated to 25.42 m by the tool. Total 1,200 hexagons are created over the study area as displayed on Figure 22.

[7] The hexagon generating tool is available on https://www.arcgis.com/home/item.html?id=7fa102df350f40a087816b93e862e21f

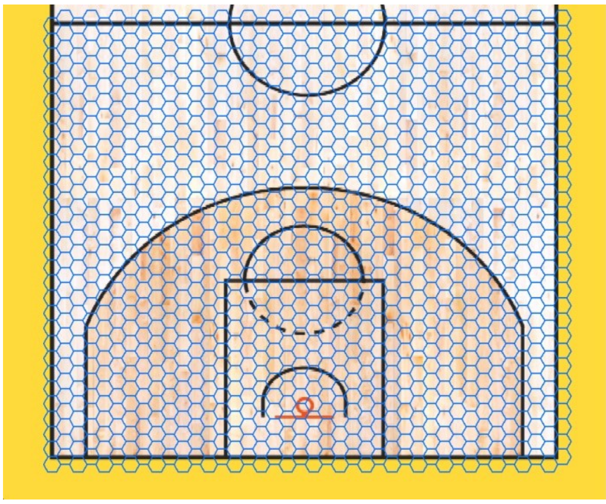


Figure 22. The fishnet (hexagons) created over the area

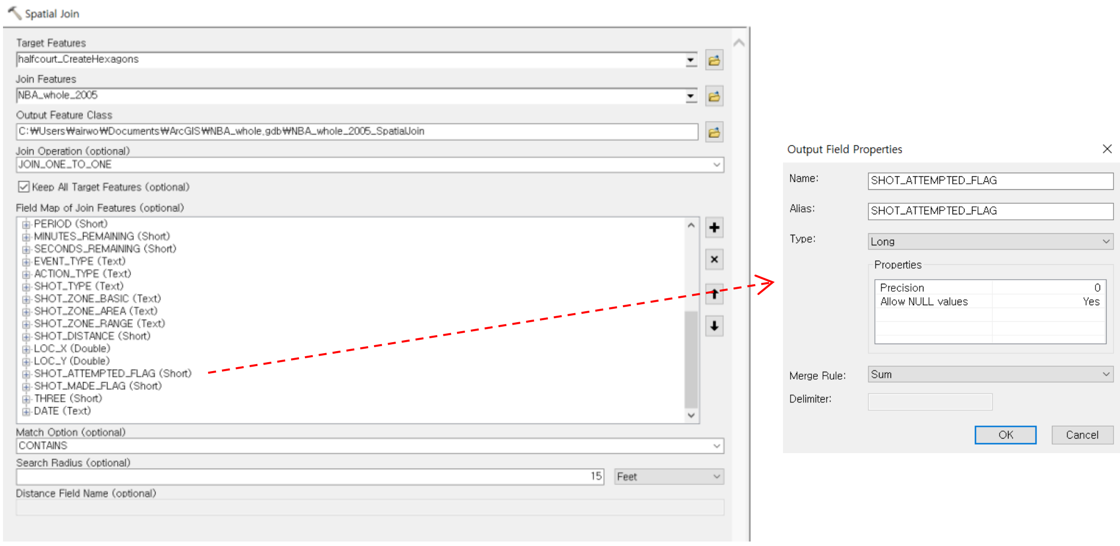


Figure 23. Spatial join tool used to aggregate points.

All the points are aggregated into the each hexagon and the hexagons counts the shots contained. In this stage, spatial join tool is used. The file having a hexagon fishnet (halfcourt_CreateHexagon.shp) is selected for target features and the shooting points file (NBA_whole_2005 and NBA_whole_2016) is selected for join features. The shooting points are aggregated into the hexagons by merge rule such as sum, count, mean etc. In this case, the sum function is selected for the rule parameter. The match option and search radius are also necessary for aggregating points by user's preference. "15m" and "*contain*" are set for counting the shots inside of the hexagon. All the points are aggregated within the distance by the "*contain*" rule. Therefore, the total shot attempts and the total shot accuracy are computed by the counts above as new fields added into the tables.

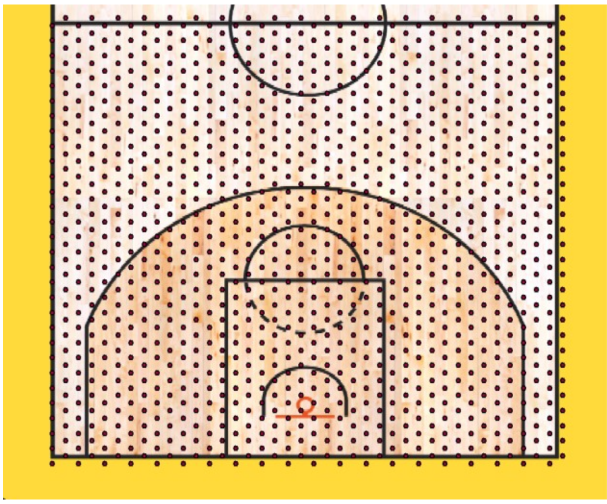


Figure 24. The points converted from hexagons by "Feature to Point" tool in ArcGIS.

The hexagons aggregating the shot points, need to be converted into points by using "*Feature to point*" tool. The reason why converting the hexagon to points is, the hexagon needs to show two attributes (shooting attempts and accuracy). If the hexagon is used without converting to points, only one attribute is shown or two different polygons are required to display. As the multiple attributes are used in layer properties, two attributes are displayed by size and color.
Therefore, the amount of the shot attempt is converted into the size of the hexagon and the shot accuracy is converted into graduated color of the hexagon.

The Top Range Hot Spot Comparison.

1. The shot attempts range is divided into 5 categories by quantile division
2. The top shooting range is extracted
3. This is defined as ***"Top range Hot spot"***
4. Sum of ***"Top range Hot spot"*** based on *"Spread"*
5. Calculates 3 point proportion from ***"Top range Hot spot"***

The "*Top range Hot spot*" is defined by a top range of shot attempts by quantile division and its sum. This index presents the differences of the shooting hot spot area between the study year 2005 and 2016. Especially 3 point proportion in "*Top range Hot spot*" is investigated.

5.4 Create Space-Time Cube

In order to find out the overall shooting trend changes for last 12 years (2005 to 2016), the space-time cube is required. The space-time cube bin in format of a *netCDF* has unique ID presenting each space and time. Each bin is shared with the spatial axis and temporal axis at the same time and the bins summaries the event points.

For the analysis, the DATE was re-organized as a reference date because NBA is held during only Oct to Aril (regular season) and there is no data from May to September. This empty data can affect and bias the result because cube recognizes that there is significantly no occurrence. To avoid possible biases, reference date is created and the data is equally distributed into the year such as 01-Jan, 01-Feb, 01-Mar, 01-Apr, 01-May etc. (Esri, 2017)

The time step interval decides a time partition to aggregate points over the time intervals such as one-day, one-week or one-year or so. For the case of analysis, one month is selected.
The distance interval decides the size of a space-time bin to aggregate points within the distance. For the case of analysis, 15 meter width is chosen.

```
---------- Input Space Time Cube Details ----------
Distance interval                        15 meters
Time step interval                        1 month

Aggregation Shape Type                   Hexagon_Grid

First time step temporal bias              100.00%
First time step interval                    after
                          2004-12-01 00:00:00
                             to on or before
                          2005-01-01 00:00:00

Last time step temporal bias                 0.00%
Last time step interval                     after
                          2016-11-01 00:00:00
                             to on or before
                          2016-12-01 00:00:00

Number of time steps                          144
Number of locations analyzed                  2098
Number of space time bins analyzed           302112
% non-zero                                37.71%
---------------------------------------------------
```

Figure 25. The space-time cube bins details.

The space-time cube is made with 1 month interval of time step. The space-time cube has 144 of time steps based on 2,098 locations. Total 302,112 space-time bins are analyzed and 37.71% of the bins are non-zero value as presented on Figure 25.

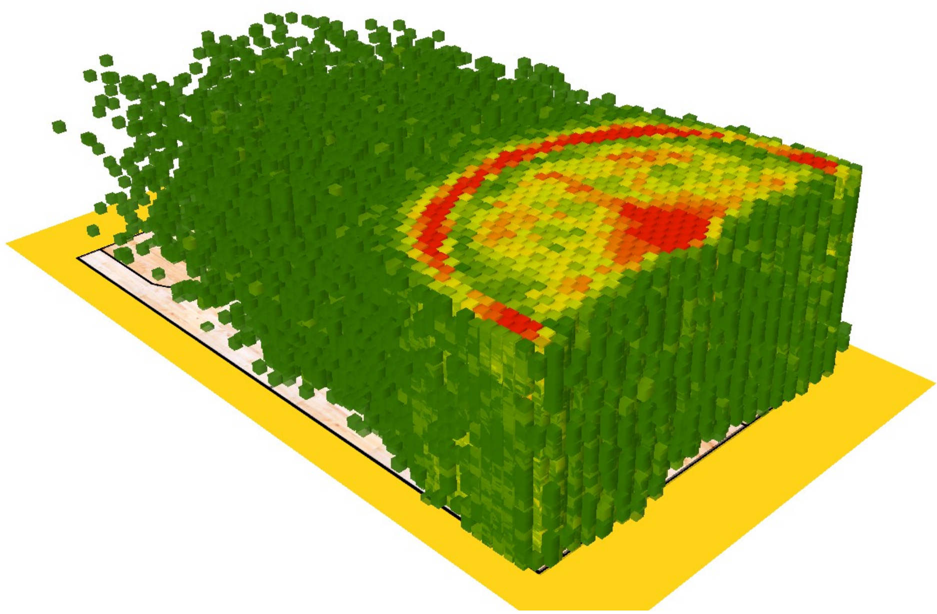
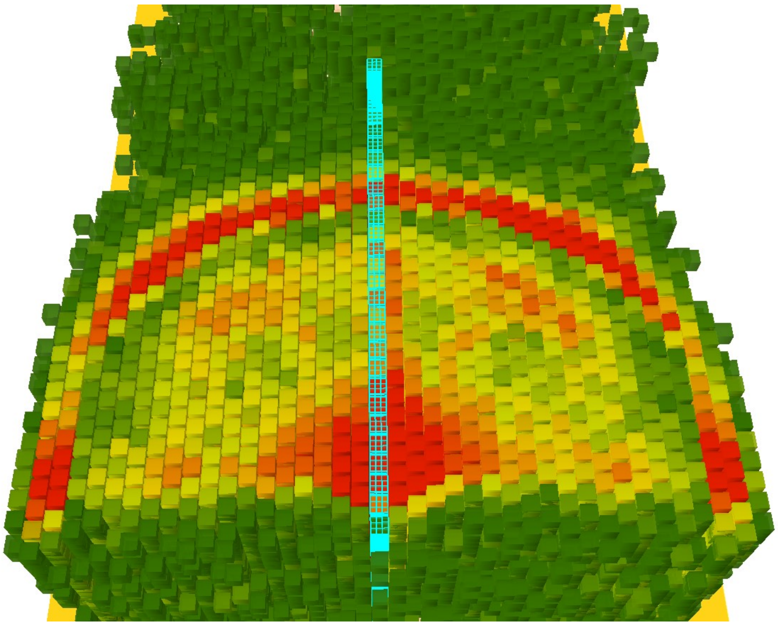
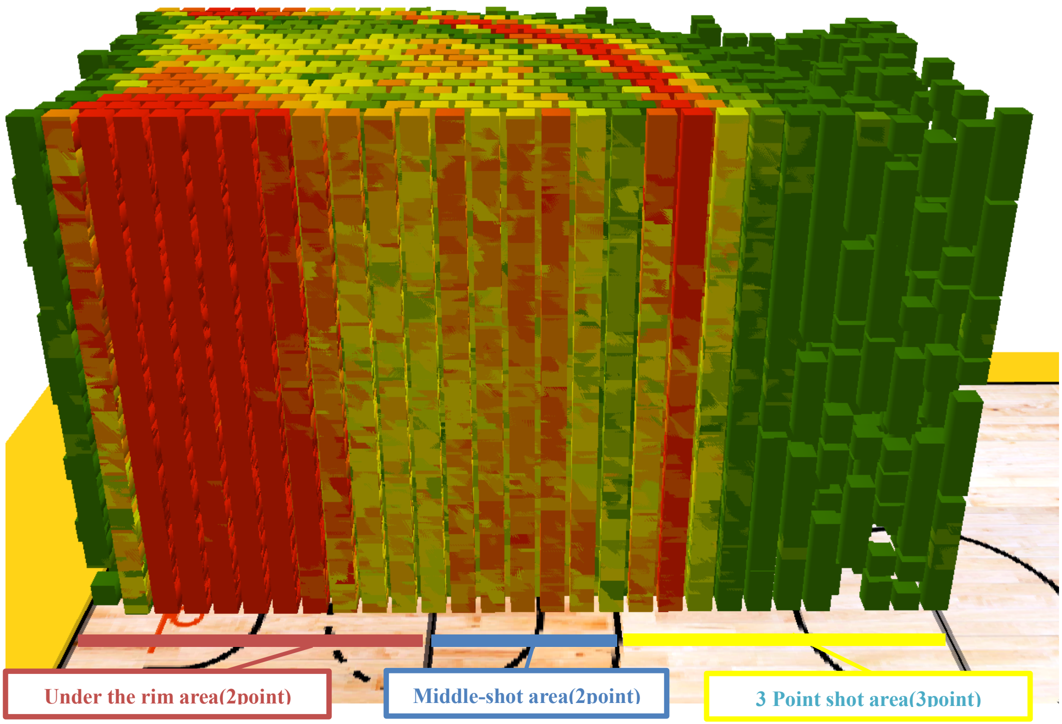


Figure 26. The Space-time cube by 1 month intervals in ArcScene 3D (10.5).
Figure 26 shows the space-time cube of shooting events. No color (transparent) indicates none-incidents in the month.
Green: Low incidents / Red: High incidents



A. The view from the Top of the space-time cube:
The selected data is highlighted to explore as below.



B. Each space-time cube bin shows different trends according to time changes
- Under the rim area continuously shows the majority of red color. (high intensity of the events)
- Middle-shot area and 3 point shot area show lots of transition of the colors. (low and high intensity of the events)

Figure 27. The exploration of the space-time cube bin partitions.
Each bin has different intensity trends. It is different by the time changes but also by shooting distances.

The Figure 26 and 27 shows the space-time cube bins displayed in 3D in ArcScene (10.5). The figure 27 presents the data exploration of the cube that the lower intensity is displayed in green and the higher intensity is displayed in red. Figure 27.B shows the cube bin's transition according to the temporal changes. Some bins only show all green in vertical axis which indicates no shots or very few shots taken (i.e. back court area) during the study periods. The other parts of the bins have various colors that present the location used to be lots of shots attempted, but some other time the location also used to be less shots attempted.

5.5 Emerging Hot Spot Analysis

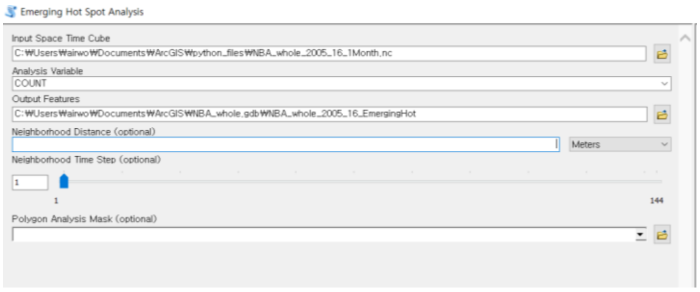


Figure 28. The emerging hot spot analysis tool.

In order to find out emerging hot spots, the space-time cube is used as an input data. The two variables are adjustable for the neighborhood; time step and the distance.
Firstly, the space-time cube is loaded as an input: NBA_whole_2005_16.nc. The analysis variable is set to "*COUNT*". The NBA_whole_2005_16.nc (space-time cube) file created by 1 month interval, the new date (reference date) has been added to avoid possible bias for result. In addition, the 12 neighborhood time step is chosen because the year 2016 is the most recent year to compare other historical trends. The distance is chosen as a default to calculate proper size of distance based on the geographical distribution of the input data. This case 56.42 m is calculated for the distance variable (the unit in ArcGIS is meter without conversion for convenience, such as 1 ft is expressed as 1 m).

```
--------------- Analysis Details --------------
Neighborhood distance            56.42348 meters
Neighborhood time step intervals             12
(spanning 12 months)
-----------------------------------------------


--------- Summary of Results ---------
                 HOT        COLD
New               3          0
Consecutive        78          1
Intensifying     223         745
Persistent       38         220
Diminishing       132         235
Sporadic          92          3
Oscillating       0          0
Historical       21          1
--------------------------------------
All locations with hot or cold spot trends: 1792 of 2098
```

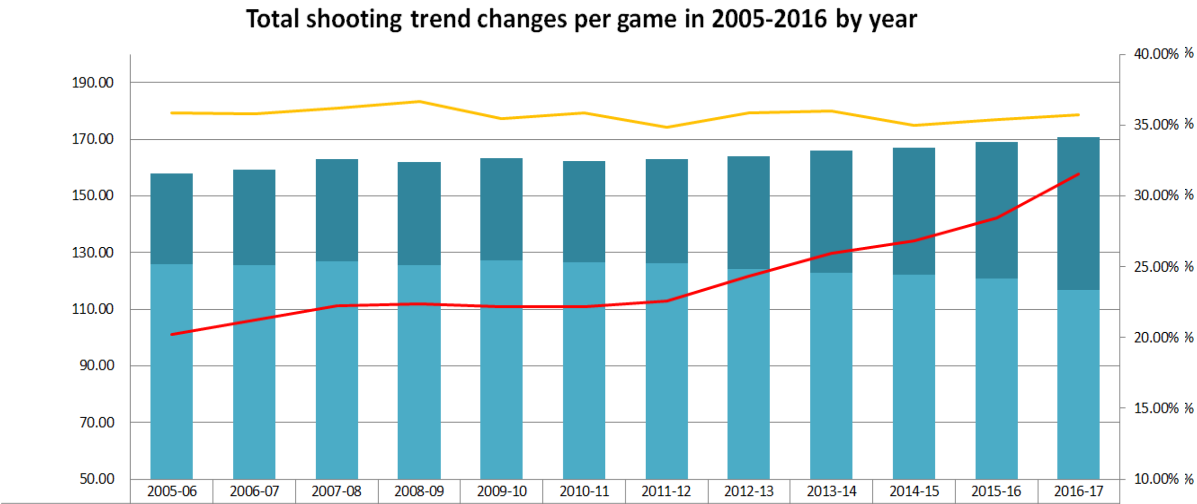
Figure 29. Summary of emerging hot spot analysis.

The Figure 29 shows the summary of emerging hot spot analysis from the space-time cube bins. Total 1,792 bins are detected out of 2,098 as a hot or cold spot. Intensifying and diminishing trend are the most frequent trend in hot spot while intensifying, persistent and diminishing are the most trend in cold spot.

6. Result

In the sixth chapter, the results are described by exploring the shooting pattern trends by "*Shot tendency*" method (Goldsberry, 2012) and "*Emerging hot spot*" analysis (Esri, 2017).
In the first part, a brief comparison of three point selection and three point accuracy rates are presented for the current shooting trend of the NBA. In second part, specific trend of the year 2005 and 2016 is compared by "*Shot tendency*" based on "*Spread*" variable. In addition, "*Top range Hot spot*" which is extracted from the top range of the shooting attempts in the year is used for clear comparison.In the third part, "*Emerging hot spot*" analysis results the overall NBA shooting pattern trends over the study periods from 2005 to 2016. The "*Emerging hot spot*" analysis presents newly emerged shooting hot spots in comparison of the historical trend in the location over the time.

6.1 A Brief Comparison of Total Shooting Trend in 2005 to 2016

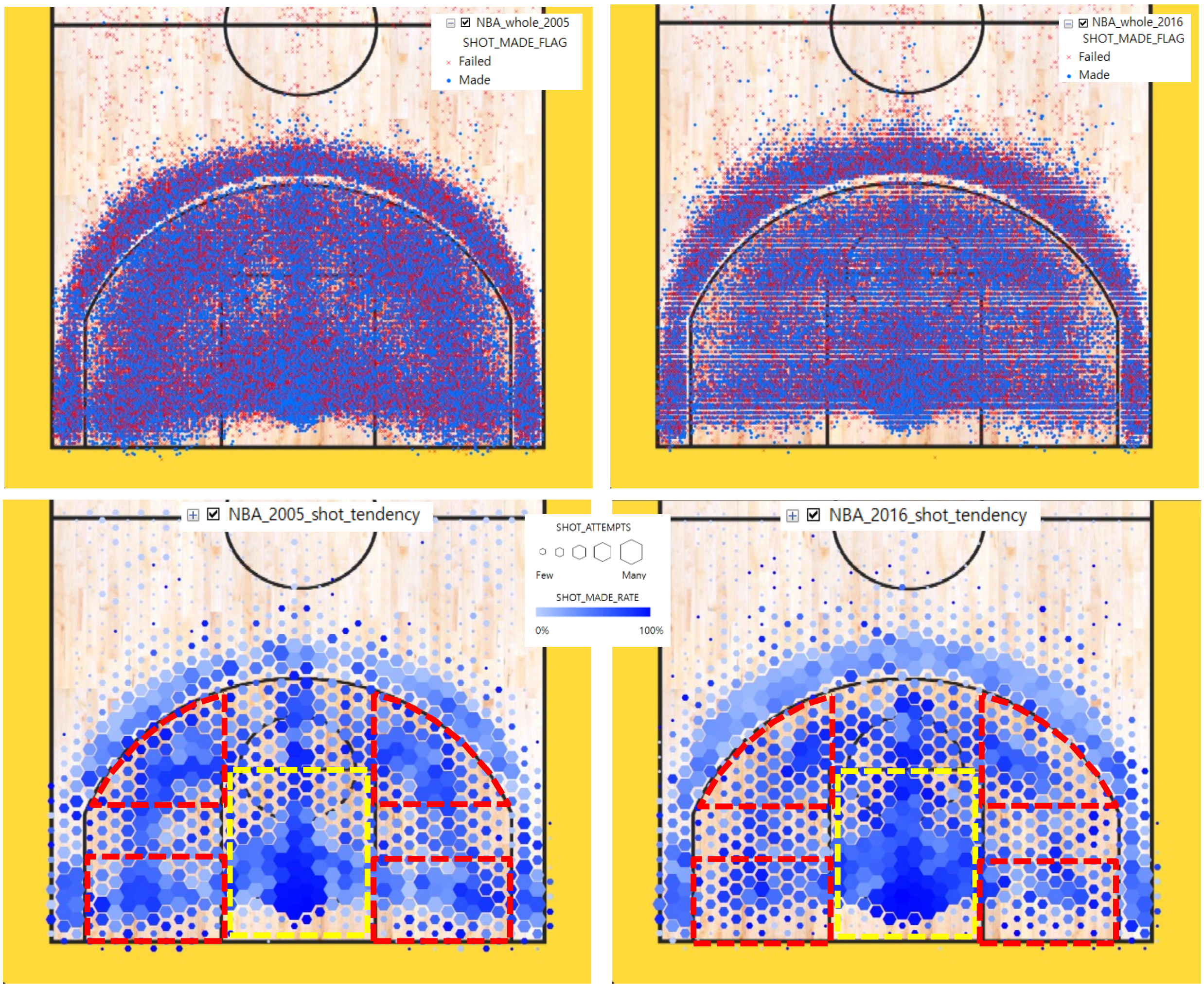


	2005-06	2006-07	2007-08	2008-09	2009-10	2010-11	2011-12	2012-13	2013-14	2014-15	2015-16	2016-17
3P Attempt	31.96	33.88	36.21	36.23	36.26	36.00	36.74	39.90	43.03	44.81	48.04	53.85
2P Attempt	126.02	125.53	126.79	125.59	127.13	126.40	126.09	124.12	122.93	122.30	120.98	116.82
3P Selection %	20.23%	21.25%	22.22%	22.39%	22.19%	22.17%	22.56%	24.33%	25.93%	26.81%	28.42%	31.55%
3P Made %	35.83%	35.82%	36.20%	36.68%	35.45%	35.85%	34.86%	35.88%	35.98%	35.00%	35.38%	35.75%

Figure 30. The total shooting trend changes per game in 2005-2016 in regular season.

Figure 30 shows the 3 point shooting trend changes per game in regular season during 2005 and 2016. Total *FGA (Field Goal Attempts)* per game has been steadily increased from 157.98 shots in 2005 to 170.67 shots in 2016. The *3PM (3Point Made)* rate has been stable within 35.00% to 36.68% for last 12 years while the *3PS (3Point Selection)* rate has been sharply increased from 20.23% to 31.55% (+11.32%). More specifically, +5.62% growth of *3PS (3P Selection)* proportion has been made in 3 years from 2014 to 2016 while only +5.7% growth made in 9 years from 2005 to 2013.
In other words, the *3PS* proportion average for the last 12 years was 24.17% while the proportion is 31.55% in 2016. The *3PM %* average for the last 12 years is 35.72% while 35.75% in 2016. Only *3PS* proportion is sharply increased in recent year.

6.2 Shot Tendency Comparison in 2005 vs 2016



	NBA in 2005			NBA in 2016			Comparison		
	Attempts	Made%	**Selection**	Attempts	Made%	**Selection**	Attempts	Made%	**Selection**
2 Point	155,001	47.79%	79.77%	143,692	50.33%	68.45%	92.70%	+2.53%	-11.32%
3 Point	39,313	35.83%	**20.23%**	66,238	35.75%	**31.55%**	168.49%	-0.09%	**+11.32%**
Total	194,314	45.37%	-	209,930	45.71%	-	**+108.04%**	**+0.34%**	-

Figure 31. The shooting trend comparison between the year 2005 and 2016 by shot tendency.

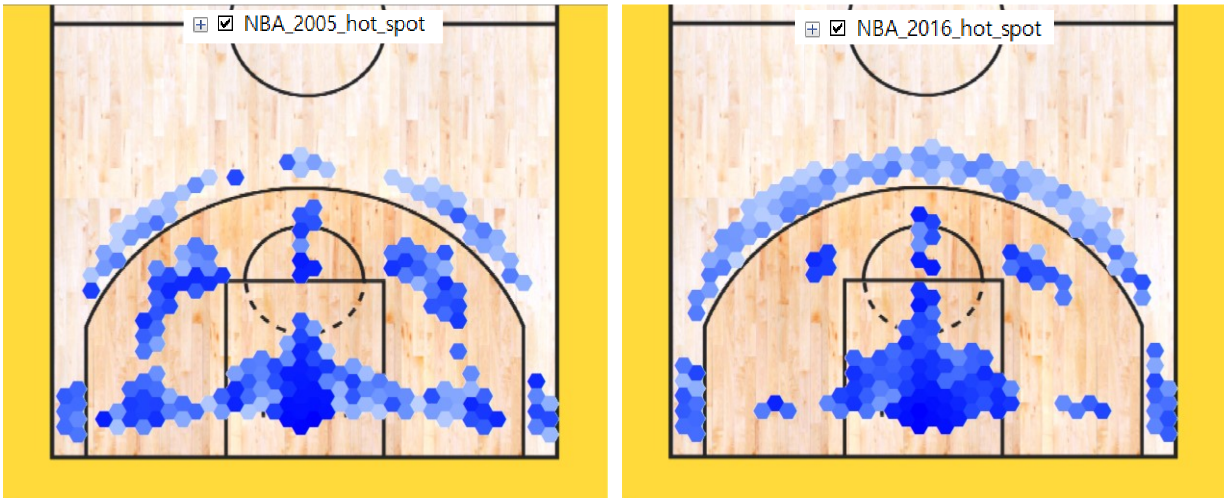
Figure 31 shows that the raw data on the top and the "*Shot tendency*" in the middle and the detailed figures in the table on the bottom.

Total 194,314 shots in 2005 and total 209,930 shots in 2016 are displayed. The dots marked in red shows the missed shots while the dots marked in blue shows the made shots on the top of Figure 31.

"*Shot tendency*" (Goldsberry, 2012) in the middle, shows that shooting trend of the year. The amount of shot attempts is displayed by the size of the hexagon and the accuracy of the shots is displayed by the graduated color. But this research focuses on the amount of shot attempts.

In 2005-06 season, total 194,314 shots are taken with 45.37% made accuracy. The 79.77% of shots are taken in 2 point area and only 20.23% of shots are taken in 3 point area. The red squares show that the most shots are attempted. The yellow square shows the closest distance area to the rim. The 3point

shots are attempted in the each corner, each 45 degree and small range of central area.
But in 2016, the "*Shot tendency*" is different. First of all, total shot attempts is increased. Total 209,930 shots are taken which is increased +8.04%. *3PS (3 Point Selection)* rate is also increased to 31.55% (which is +11.32% compared to 2005).
The same areas marked in red squares clearly shows different tendency. The size of hexagon is decreased except near the rim area marked in yellow rectangle. The area in yellow, each hexagon's sizes is bigger and spread out widely. In addition, the shot range near the rim is assumed in changes. Such as a *lay-up shot* (a normal shooting type under the rim) has been transformed into various types *like finger roll, scoop shot, floater shot* etc. Shooting skills have been developed and improved. According to NBA official website (stats.nba.com), 70 different types of shooting are defined that some of the shooting types are not existed few years ago. This improvement may cause the shot range changes in yellow rectangle area.
Behind the 3 point line, the size of hexagons is bigger and wider than 2005 while the 2 point area, the size of hexagon is decreased than before. But it is hard to understand the trend difference at once. The "*Top range Hot spot*" methods are used as below.



Top range Hot spot	2005	2016
Top Shooting Range	> 316	> 296
3P Hot spot	56	100
Total number of Hot spot	204	200
3P Hot spot proportion %	**27.45 %**	**50.00 %**

Figure 32. The Top range Hot spot extracted from shot tendency.

The "*Top range Hot spot*" comparison is simple but clear that shows differences of shot tendency of players in the year. The shots are categorized into 5 ranges by quantile division method that divides the amount of shots into equal division. In 2005, over 316 shots are ranged as the top range while over 296 shots are the top range in 2016. The range is extracted and displayed on Figure 32.

In 2005, total 204 top range hot spot is extracted. 56 hot spot is detected behind the 3 point line which is 27.45% proportion in total top rage hot spot. But in 2016, total 200 top range hot spot is extracted and 100 hot spot is detected behind the 3 point line. This is 50% proportion in total top range hot spot. The NBA players attempt 3 point shots as the most preferable shot 1 out of 2. In addition, near rim

area, the number of hexagons is concentrated in central area. The number of hexagon may show shot skills' development as addressed on Figure 31. The NBA players are not preferred on having 2 point shots except central area near the rim.
In the conclusion, the trend difference is clearly seen through "*Top range Hot spot*". The difference was big near 3 point arc line. The 3 point top range hotspots cover wider area up to 50% while 2 point hot spots are decreased especially in each side of upper area and each side of bottom area.

6.4 Emerging Hot Spot Analysis

The "*Emerging hot spot*" analysis is conducted by using the space-time cube bins. The space-time cube was made by 1 month time intervals in the research. Thus, for the study periods from 2005 to 2016, total 144 time span was created. For the emerging hot spot analysis, the ArcGIS program calculated the neighborhood distance to 56.42 m, but the neighborhood time step (default value = 1) was set to 12 time step intervals to aggregates the bins to compare. The 12 time step intervals indicate the year 2016 to be compared with the rest of data.

```
-------------- Analysis Details -------------
Neighborhood distance            56.42348 meters
Neighborhood time step intervals            12
(spanning 12 months)
---------------------------------------------

--------- Summary of Results ---------
                 HOT      COLD
New               3         0
Consecutive      78         1
Intensifying    223       745
Persistent       38       220
Diminishing     132       235
Sporadic         92         3
Oscillating       0         0
Historical       21         1
-------------------------------------
All locations with hot or cold spot trends: 1792 of 2098
```

Figure 33. The summary of the result in emerging hot spot.

Figure 33 shows the summary of the results in the emerging hot spot analysis. Total 1,792 out of 2,098 bins have hot or cold spot trends over the location. Total 587 hot spots were detected while 1,205 cold spots were detected. The 306 bins have no patterns.
The majority hot spot trend of the result is intensifying trends with 223 bins (37.99%). The diminishing trend (22.49%), sporadic trend (15.67%) and consecutive trend (13.29%) are followed in hot spots. Cold spot is not focused on the study because the cold spot is detected generally in back court area.

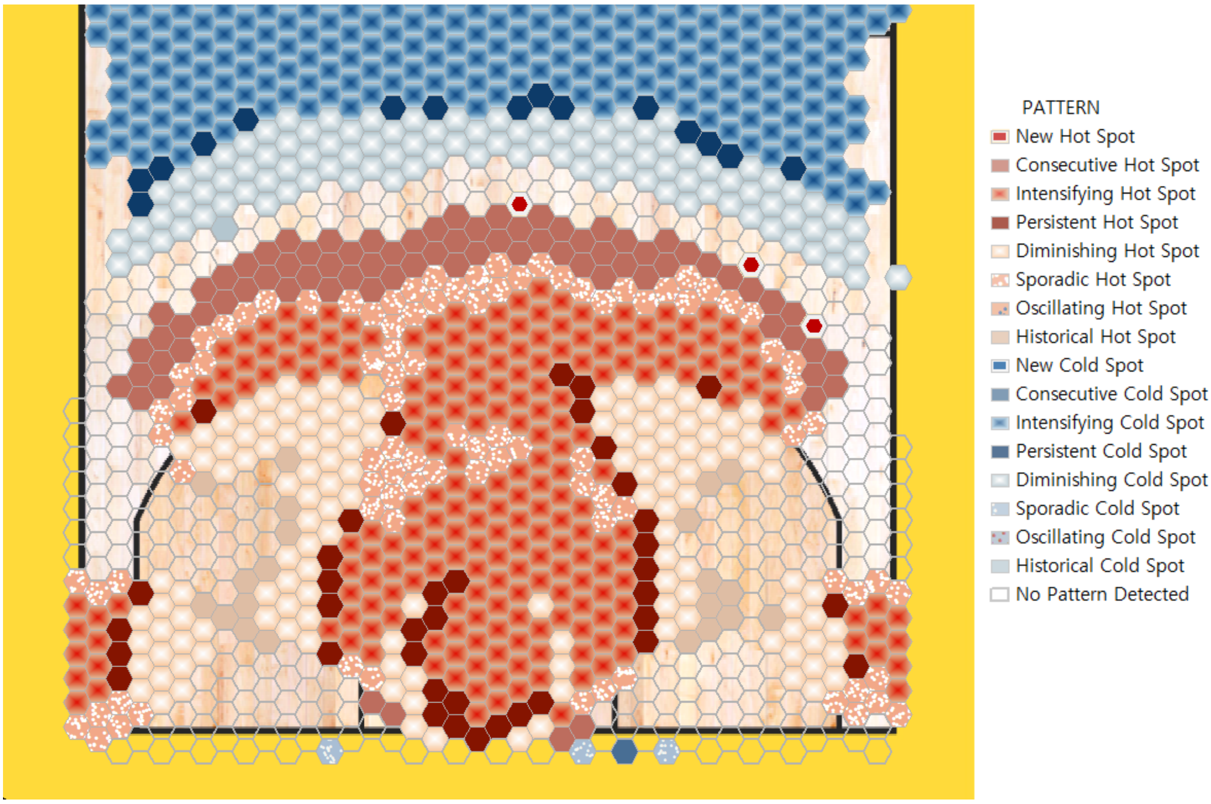


Figure 34. The emerging hot spot analysis on the NBA shooting trend from 2005 to 2016.

Figure 34 shows the emerging hot spot analysis on the NBA shooting trend form 2005 to 2016. The 16 different type trends of hot & cold spots are defined and displayed. The back court area is dismissed as the most of shots are occurred in the half court.

For the analysis, the half court is divided into 11 zones for clear comparison. In the half court area, total 8 types of hot spot trend are detected except oscillating trend.

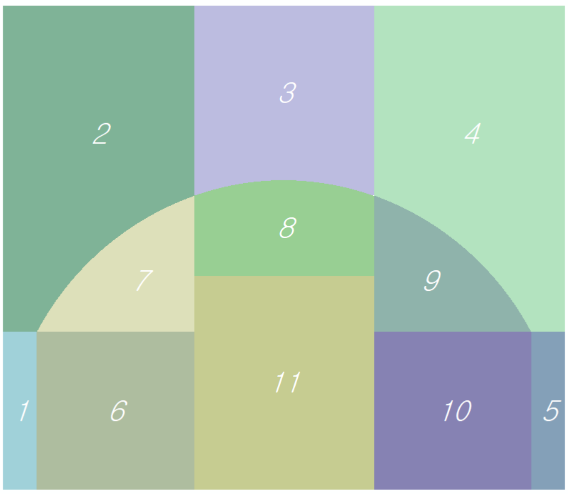


Figure 35. The half court division.
Zone 1 – 5: 3 point shot area / Zone 6 – 11: 2 point shot area

Table 6. The emerging hot spot trends detected in 11 zones

	Trend / Zone	1	2	3	4	5	6	7	8	9	10	11
Hot Spot	Consecutive		28	27	19							4
	Diminishing						31	31	2	29	23	16
	Historical						9	2		2	8	
	Intensifying	10	12	0	11	9	6	11	43	15	6	100
	New			1	2							
	Oscillating											
	Persistent	3				1	6	1	3	2	7	15
	Sporadic	8	11	9	12	8	2	1	13	1	4	23
Total	587	21	51	37	44	18	54	46	61	49	48	158

6.4.1 3 Point Left & Right Corner Zone (Zone 1 & Zone 5)

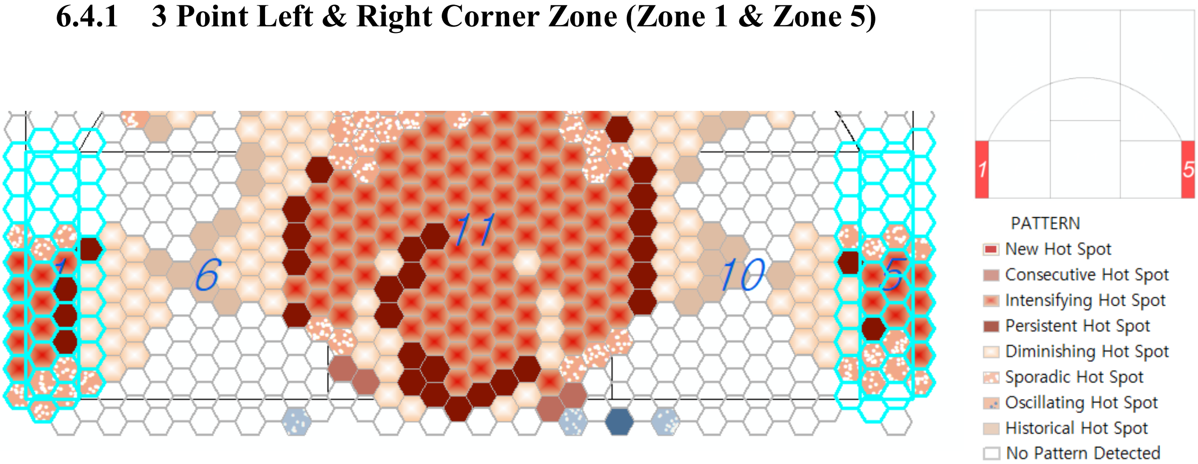


Figure 36. The emerging hot spot trend analysis in the zone 1 & 5.

Figure 36 shows Zone 1 & Zone 5's emerging hot spot trend. The areas have 21 and 18 hot spots respectively. In the two zones, 47.62% and 50% of intensifying hot spot trend are detected which indicates that "*the location has been hot spot with 90% of the time-step intervals, including final time step*" (Esri, 2017). The both corner areas also have 38.10% and 44.44% of sporadic hot spot trend which indicates that the location as "*not over 90 percent of time-step intervals but have been hot spot with none cold spots*" (Esri, 2017).

The Zone 1 and Zone 5 are assumed as favorite hot spot area from the past to the current in last 12 years of time.

6.4.2 3 Point Center-left, Center, Center-right (Zone 2 & 3& 4)

The Zone 2, 3, and 4 includes the cold spot trends. The majority of shots is attempted in close distance to the rim or recently near the 3 point line area. Thus, the cold spots or no patterns are not considered for the analysis and for further research, the cold spot area needs to be dismissed for the convenience.

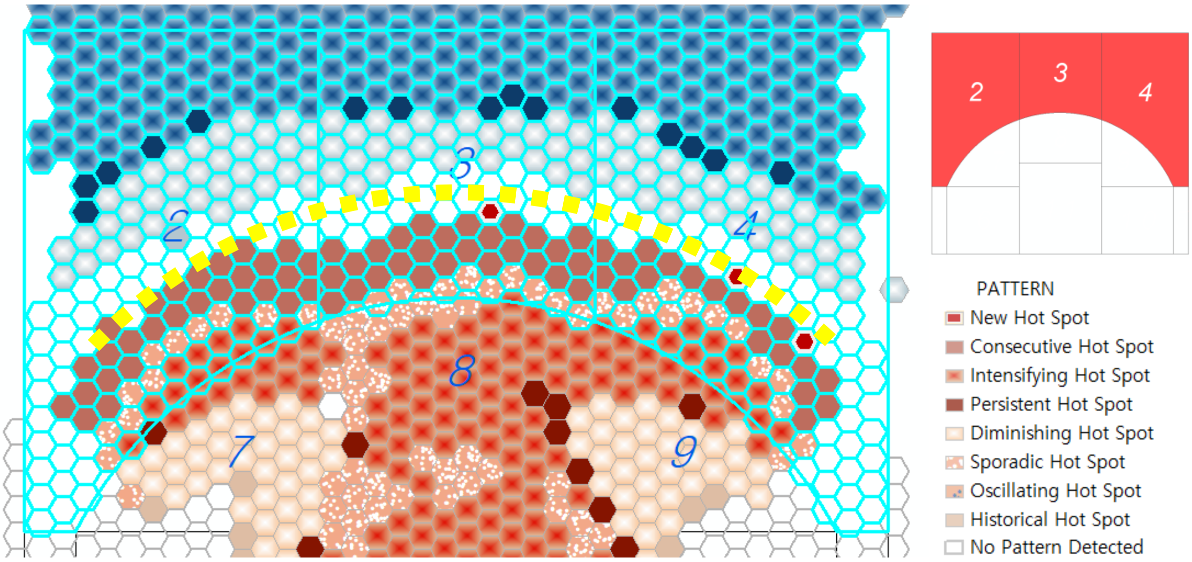


Figure 37. The emerging hot spot trend analysis in the zone 2 & 3& 4.

Figure 37 shows the area of Zone 2, 3, and 4. The area below the yellow line dashed is focused to investigate the trend for 3 point area. The consecutive hot spots are seen. The 54.90%, 72.97%, and 43.18% of consecutive hot spot trends are detected in Zone 2, 3 and 4 respectively. This means that "*this area has not been significant hot spot before the final hot spot run with less than 90% of all bins are statistically significant hot spots.*" (Esri, 2017). In other words, 3 point area has been expanded compared to the past, and the area has been significant growth in recent.

The areas that are close to the 3 point arc line have sporadic and intensifying hot spot trend. The trends show 23.53% of intensifying trend in the Zone 2, 24.32% and 27.27% of sporadic trend in the Zone 3 and 4. The trend indicates that the location has been a hot spot but with different intensity compared to the past.

6.4.3 2 Point Center-left, Center, Center-right (Zone 7 & 8 & 9)

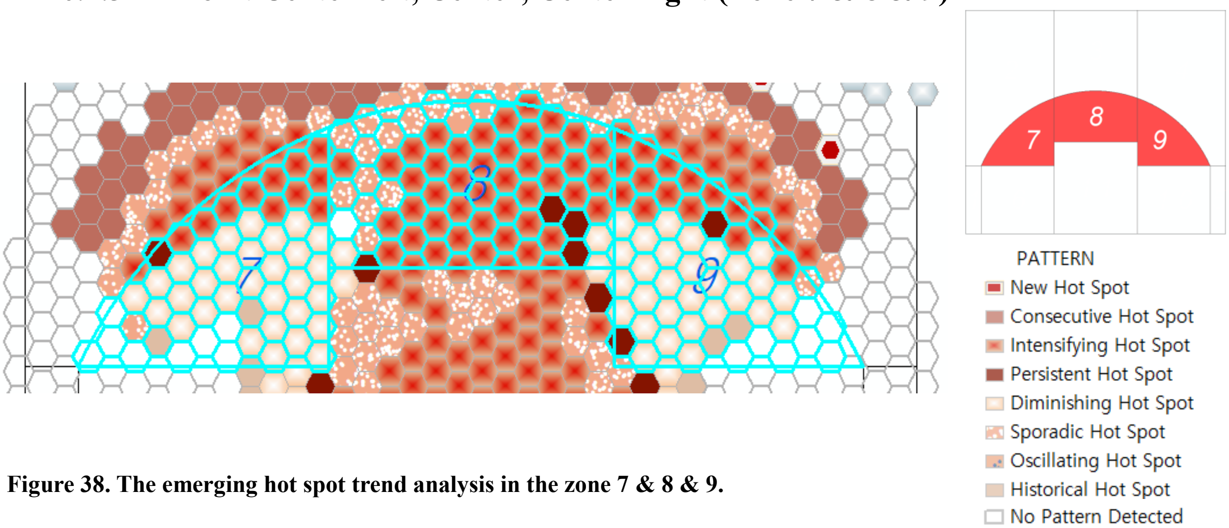
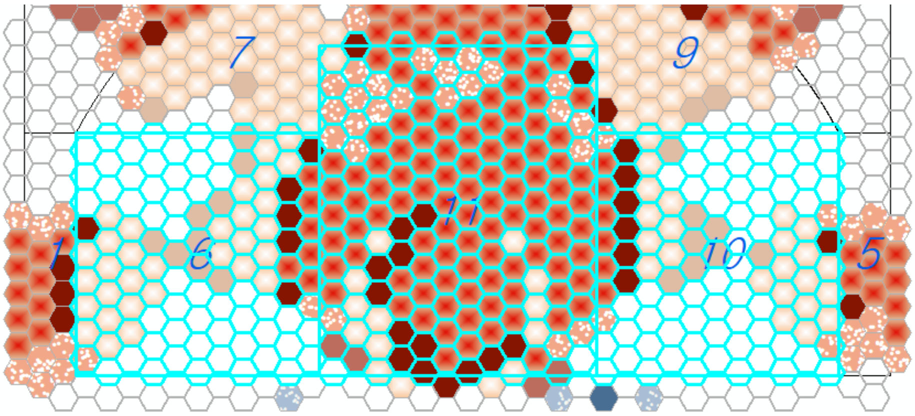


Figure 38. The emerging hot spot trend analysis in the zone 7 & 8 & 9.

The Zone 7 and Zone 9 show decreasing hot spot trends. 67.39% and 59.18% of diminishing hot spot trend are detected in the zone 7 and 9, while 70.49% of intensifying hot spot trend is detected in the Zone 8. This means that the shots in center-left and center-right (Zone 7 & 9) "*has been hot spot for 90 percent of the time-step intervals, but each time step is decreasing overall and that decrease is* *statistically significant*" (Esri, 2017). However, the trend in central area became more intensive hot spots compared to the past in the Zone 8.

6.4.4 2 Point Left, Center, Right (Zone 6 & 10 & 11)



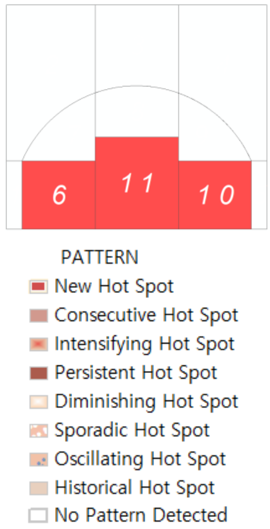


Figure 39. The emerging hot spot trend analysis in the zone 6 & 10 & 11.

Figure 39 shows the decreasing hot spot trends or no patterns detected in the zone 6 and zone 10. 57.41% and 47.92% of diminishing hot spot trends are appeared in the Zone 6 and Zone 10 respectively. The shot attempts have been significantly decreased statistically. The zones also have 16.67% of historical hot spot trends respectively. (Esri, 2017)

However, Zone 11 has been the most intensive zone on the court which is the most of shots are attempted. In the Zone 11, 63.29% of intensifying hot spot trend is occurred and 14.56% of sporadic hot spot trend is followed.

6.5 The Emerging Hot Spot Trend Summary

Table 7. The emerging hot spot trend summary table

	3 Point area	2 Point area (except zone 11)	Near the rim area (zone 11)
Consecutive	**74**	0	4
Diminishing	0	**116**	16
Historical	0	21	0
Intensifying	**42**	**81**	**100**
New	3	0	0
Oscillating	0	0	0
Persistent	4	19	15
Sporadic	**48**	21	23
Total	171	258	158

To sum up, the zones are re-arranged into 3 point area, 2 point area, and near the rim area (Zone 11). In 2 point area, the zone 11 is not considered because the zone 11 has considerable different characteristics compared to the other 2 point area zones.

In the 3 point area, total 171 bins are detected. Consecutive hot spot trend (74 bins, 43.27%) is the most relevant trend and Sporadic (28.07%) and Intensifying (24.56%) trend are followed. Consecutive hot spot trend shows that "*the location has never been a statistically significant hot spot prior to the final hot spot run and less than ninety percent of all bins are statistically significant hot spots.*" (Esri, 2017). Sporadic and intensifying trends show that the 3 point area has been a hot spot trend in different intensity level. Thus, the 3 point area has been a hot spot and also the area became higher intensity hot spot in recent.

In the 2 point area, total 258 bins are detected. The Diminishing hot spot trend (116 bins, 44.96%) is the most relevant trend that shows "*the intensity of clustering in each time step is decreasing overall and that decrease is statistically significant*" (Esri, 2017). Intensifying trend (81 bins, 31.40%) is the second most relevant that shows "*the intensity of clustering of high counts in each time step is increasing overall and that increase is statistically significant.*" (Esri, 2017). These opposite trends are detected in 2 point area. The diminishing trends are appeared in overall 2 point area except Zone 8 (the 2 point central upper area).

In the Zone 11, the Intensifying hot spot trend (100 bins, 63.29%) is the most relevant trend. The intensifying hot spot trend has been wider and bigger. The trend is assumed that the shot skills' improvement is connected with the shooting range in the Zone 11.

7. Discussion and Further Research

In the seventh chapter, the elements found during the studies and how it could be continued to further develop methods in the future is discussed. The research is to find out the NBA players' shooting hot spot trend in the last 12 years from 2005 to 2016. A comparison study was conducted on the specific year's trend by the "*Shot tendency*" method. In addition, the "*Top range Hot spot*" method was used for clear comparison of the hot spot trend for each year. Secondly, the space-time cube was used to investigate the shooting trend changes over time. The half court was divided into 1,200 cells and 11 zones to explore the zone's trend characteristics.

Three questions from the beginning were set for the analysis objective.

1. Are there different shooting hot spot trends between the year 2005 and 2016?
2. If so, how has the shooting hot spot trend changed over the space by the temporal changes?
3. Why is it important to know the trend by temporal changes?

7.1 Are There Different Shooting Hot Spots existed between the Year 2005 and 2016?

The "*Shot tendency*" method (Goldsberry, 2012) was used to compare the shooting hot spots for each year. The hexagon's size showed the intensity of shooting attempts. The different shooting hot spot trends are seen in the comparison which indicates shot tendencies of the year. The method displayed shooting attempts' data and the changes of shot selections. The same areas marked in red and yellow were compared that shows decreases and increases of hexagons' size.

The "*Top range Hot spot*" based on "*Spread*" concept was used for better understanding of the shooting trend. The "*Top range Hot spot*" showed only the top range category of shooting attempts divided by the quantile division. Quantile divided the data into same amount categories.
In 2005, the top range was set to over 316 shots (> 316) which means that the total 204 top range hot spots are detected. Among the 204 top range hot spots, 56 hot spots were in 3 point area. In 2016, the top range was set to over 296 shots (> 296) which means that the total 200 top range hot spots were detected and 100 hot spots were in 3 point.
The 3 point proportion from the top range hot spots were sharply increased from 27.45% in 2005 to 50% in 2016 (+22.55%). Near the 3 point arc line, the higher number of "*Top range Hot spot*" is detected in wider areas than in 2005. While the number of "*Top range Hot spot*" in 2 point areas is decreased all over the court except under the rim area (Zone 11).

This method was suitable for specific moment trend analysis. However, the methods did not indicate the shooting trend's transition over the particular time. For example, if the first year's value and the last year's value are the same, this method presents the difference as none or 0. The possible transition between the periods is dismissed.

7.2 If so, how has the shooting hot spot trend changed over the space by temporal changes?

In order to investigate the trend changes by temporal changes, the "*Space time pattern mining tools*" is used in ArcGIS (10.5). The space-time cube was created by the time intervals and the distance adjustments (1 month intervals and 15m distance were set). The space-time cube aggregates the event

points (shooting points) in a netCDF (Network Common Data Form) format to each bin by sharing the unique ID in the space and the time. (Esri, 2017)
Once the space-time cube bins were created, the emerging analysis tool calculated its bin attributes by the user's neighborhood time span and neighborhood distance. These neighborhood parameters were important because these parameters decided the trend characteristics. Lastly, the tool defined the space into 16 different trends such as new, intensifying, persistent, diminishing, sporadic, oscillating, historical type which explains the trend changes in the temporal changes in the history. (Esri, 2017)

The half court was divided into 11 zones to investigate its specific trend and it showed different trends over the space. The overall trend in 3 point areas, 43.27% of consecutive hot spot trend, 28.07% of sporadic and 24.56% of intensifying hot spot patterns were detected. The 3 point attempts expanded as a continuous hot spot or became a recent hot spot in the location over the study periods.
The overall trend in 2 point areas, 43.51% of intensifying hot spot trends and 31.73% of diminishing hot spot trends were detected. If Zone 11 (the area under the rim) is dismissed, general 2 point areas showed 44.96 % of diminishing hot spot trends and 31.40% of intensifying hot spot trends while the Zone 11 showed 63.29% of intensifying patterns and 14.56% of sporadic patterns. As a result, 2 point areas were decreased significantly.

In conclusion, the final outcome of 3 point areas were placed as a hot spot or became a higher hot spot over the study periods while 2 point areas decreased with the exception of Zone 11.

7.3 Why is it important to know the trend by temporal changes?

In the research, the shooting pattern trend is compared by "*Shot tendency*" and "*Emerging hot spot*" analysis methods. The "*Shot tendency*" showed the specific moment's shooting trend but the transition of the trend between the periods was hardly understood.
Through the integration of spatial and temporal perspectives, the overall historical footprints were clearly investigated by the "*Emerging hot spot*" analysis. The emerging hot spot analysis was carried out by the "*Space time pattern mining tools*" with the space-time cube based. The 16 different type trends indicated the location over the time of how the trend has been changed throughout the history with different value intensities.
The purpose of the trend analysis is the ability to forecast the future trends based on the historical changes. Thus, the trend analysis by temporal changes enables the various further studies. Moreover, by adjusting the neighborhood time span and distance, various spatial and temporal analyses are possible. The adjustments are crucial and a risk factor but relatively it provides a faster and more accurate calculation method than the traditional mining ways of analysis.

8. Conclusion

The eighth chapter concludes the research paper with a brief summary. In this research, the overall NBA shooting hot spot trend was studied. Basketball is a sport occurred in both spatial and temporal backgrounds. Thus, GIS is a suitable tool to manage basketball data based on the attributes and the usage in sport analytics is emphasized.

By using the "*Shot tendency*" method, the specific year's trend and its hot spots were examined. Although, the differences between the years are apparent, difficulty in assumption lies in how the hot spots have changed over time.
By the "*Emerging hot spot*" analysis, the space over the time is understood by investigating the trend in temporal changes. Especially, this would be hard work when the data is so massive. The space-time cube provides spatial and temporal integration and the new possible analysis methods.

As a result, different trends in the 11 zones were detected. Near the 3 point arc line (Zone 2 & 3 & 4), mostly continuous hot spot trends were detected. The zones have expanded compared to the past and has seen the recent significant growth.
The 2 point areas, especially each side of the 2 point area (Zone 6 & 7 & 9 & 10), mostly the diminishing hot spot trends were detected. The zones have been statically decreased in recent times with the exception of Zone 11.

In conclusion, this study provided an initial foundation for the NBA shooting trend changes in the last 12 years from 2005 to 2016. This research demonstrated the trend changes on shooting locations by time changes that showed historical footprints and future forecasting at the same time. The space-time cube and emerging hot spot analysis reached the new outcome that gives a more coherent understanding of the trend changes in basketball history.

The further study could be continued especially in terms of identifying the causes of the trend changes. In the research, the overall shooting changes are explored by 16 different trends but the reasons that caused the trend changes were not investigated.
The large physiques players dominated the league in the past. In recent, Curry and the higher 3PA, 3PM% teams were introduced as the trend leader (team). But the connection between the trend and the leader (team) was not investigated in the research as well.

To conclude, defining the trend changes over the last 12 years (from 2005 to 2016) in the NBA provided the new perspective of reviewing the past and present at a glance. The research has prospects in its pragmatic analysis to further develop competent strategies and plans for the players, teams, coaches and also fans.

References

Aisch, G. & Quealy, K., 2016. *Stephen Curry's 3-Point Record in Context: Off the Charts.* [Online]
Available at: https://www.nytimes.com/interactive/2016/04/16/upshot/stephen-curry-golden-state-warriors-3-pointers.html
[Accessed 9 10 2017].

Basketball-reference, 2017. *Basketball-reference.com.* [Online]
Available at: https://www.basketball-reference.com/leaders/fg_pct_yearly.html
[Accessed 10 11 2017].

Basketball-reference, 2017. *NBA Individual Regular Season Records for 3-Point Field Goal Percentage.* [Online]
Available at: https://www.basketball-reference.com/leaders/player_fg3_pct.html
[Accessed 9 10 2017].

Brocato, J., 2014. *NBA Average Height by Position & Year.* [Online]
Available at: https://shutupandjam.net/nba-ncaa-stats/ht-by-pos/ (originally from www.basketball-references.com)
[Accessed 9 10 2017].

Brunner, G., 2015. *The Geography of Basketball: Mapping NBA Shotcharts in ArcGIS.* [Online]
Available at: gavinr.com/2015/11/04/geography-basketball-mapping-nba-shotcharts-arcgis/
[Accessed 9 10 2017].

Cervone, D., D'Amour, A., Bornn, L. & Goldsberry, K., 2014. *POINTWISE : Predicting Points and Valuing Decisions in Real Time with NBA Optical Tracking Data,.* s.l., MIT Sloan Sports Analytics Conference 2014.

ESPN, 2016. *NBA Attendance Report.* [Online]
Available at: http://www.espn.com/nba/attendance/
[Accessed 9 10 2017].

ESPN, 2016. *NBA extends television deals.* [Online]
Available at: http://www.espn.com/nba/story/_/id/11652297/
[Accessed 9 10 2017].

ESPN, 2017. *NBA History - MVP.* [Online]
Available at: http://www.espn.com/nba/history/awards/_/id/33
[Accessed 9 10 2017].

Esri, 2017. *An overview of the Space Time Pattern Mining toolbox.* [Online]
Available at: http://desktop.arcgis.com/en/arcmap/latest/tools/space-time-pattern-mining-toolbox/an-overview-of-the-space-time-pattern-mining-toolbox.htm
[Accessed 9 10 2017].

Esri, 2017. *Create Space Time Cube.* [Online]
Available at: http://desktop.arcgis.com/en/arcmap/latest/tools/space-time-pattern-mining-toolbox/create-space-time-cube.htm
[Accessed 9 10 2017].

Esri, 2017. *How Create Space Time Cube By Aggregating Points works.* [Online]
Available at:
http://desktop.arcgis.com/en/arcmap/latest/tools/space-time-pattern-mining-toolbox/learnmorecreatecube.htm#ESRI_SECTION1_F1EA94A3BA8940E0B56AB08A302D1C08
[Accessed 10 10 2017].

Esri, 2017. *How Create Space Time Cube works.* [Online]
Available at: http://desktop.arcgis.com/en/arcmap/latest/tools/space-time-pattern-mining-toolbox/learnmorecreatecube.htm
[Accessed 9 10 2017].

Esri, 2017. *How Emerging Hot Spot Analysis works.* [Online]
Available at: http://desktop.arcgis.com/en/arcmap/latest/tools/space-time-pattern-mining-toolbox/learnmoreemerging.htm

[Accessed 9 10 2017].

Esri, 2017. *How Hot Spot Analysis (Getis-Ord Gi*) works.* [Online]
Available at: http://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/h-how-hot-spot-analysis-getis-ord-gi-spatial-stati.htm
[Accessed 10 10 2017].

FIBA Central Board, 2017. *Official Basketball Rules 2017.* Mies, Switzerland: FIBA.

FIBA, 2017. *USA Basketball.* [Online]
Available at: https://www.usab.com/news-events/news/2017/06/3x3-olympics.aspx
[Accessed 7 2017].

Franks, A., Miller, A., Bornn, L. & Goldsberry, K., 2015. *Counterpoints: Advanced Defensive Metrics for NBA Basketball.* s.l., MIT Sloan Sports Analytics Conference 2015.

Goldsberry, K., 2012. *CourtVision: New Visual and Spatial Analytics for the NBA.* s.l., MIT Sloan Sports Analytics Conference 2012.

Jensen, D., 2014. *Spatial Analysis and Visualization in the NBA.* s.l., B.A., 2011, University of California, Berkeley.

Land of basketball, 2017. *NBA All-Time Points Leaders: Career Totals in the Regular Season.* [Online]
Available at: http://www.landofbasketball.com/all_time_leaders/points_total_career_season.htm
[Accessed 9 10 2017].

Laube, P., Imfeld, S. & Weibel, R., 2005. Discovering Relative Motion Patterns in Groups of Moving Point Objects. *International Journal of Geographical Information Science 19 (6)*, p. 639–668.

Mcclusky, M., 2014. *Faster-Higher-Stronger.* [Online]
Available at: https://www.wired.com/2014/10/faster-higher-stronger/
[Accessed 9 10 2017].

Naismithbasketballfoundation, 2014. *History of Basketball.* [Online]
Available at: https://naismithbasketballfoundation.com/about-basketball/history-of-basketball/
[Accessed 11 10 2017].

NBA STATS, 2017. *Teams Shot Dashboard General.* [Online]
Available at: http://stats.nba.com/teams/shots-general/?sort=FG3_PCT&dir=1
[Accessed 9 10 2017].

NBA, 2002. *History of Basketball in Canada.* [Online]
Available at: http://www.nba.com/canada/History_of_Basketball_in_Canad-Canada_Generic_Article-18023.html
[Accessed 11 10 2017].

NBA, 2013. *Official Rules of the National Basketball Association.* s.l.:NBA.

Sloansportsconference, 2017. *Kirk Goldsberry.* [Online]
Available at: http://www.sloansportsconference.com/people/kirk-goldsberry/
[Accessed 9 10 2017].

Smith, K., 2016. *Twitter.* [Online]
Available at: https://twitter.com/thejetontnt/status/703808574330253314
[Accessed 9 10 2017].

Sportsknowhow, n.d. *"Professional-NBA Basketball Court Dimensions" Professional-NBA Basketball Court Dimensions | Court & Field Dimension Diagrams in 3D.* [Online]
Available at: https://www.sportsknowhow.com/basketball/dimensions/nba-basketball-court-dimensions.html
[Accessed 9 10 2017].

SportVU, 2017. *STATS SportVU basketball player tracking.* [Online]

Available at: https://www.stats.com/sportvu-basketball/
[Accessed 9 10 2017].

Totalsportek, 2016. *Total Sportek.* [Online]
Available at: http://www.totalsportek.com/most-popular-sports/
[Accessed 26 5 2017].

Willard, J., 2016. *Average Position Measurements: 2016.* [Online]
Available at: https://fansided.com/2016/06/23/average-position-measurements-2016/
[Accessed 9 10 2017].

Appendix 1. Python code for importing NBA shooting data to ArcGIS

Python code explains that how the raw data is loaded from the website (stats.nba.com) into ArcGIS (10.5) based on Gavin's post (https://gavinr.com/2015/11/04/geography-basketball-mapping-nba-shotcharts-arcgis/)

```python
import requests
import arcpy
import os
import json


#getting raw data from website
def get_coord (player_id, season, seasontype):

    #call request through nba url
    nba_call_url=('http://stats.nba.com/stats/shotchartdetail?Season=%s&SeasonType=%s&TeamID=0&PlayerID=%s&'+
        'GameID=&Outcome=&Location=&Month=0&SeasonSegment=&DateFrom=&Dateto=&OpponentTeamID=0&VsConference=&'+
        'VsDivision=&Position=&RookieYear=&GameSegment=&Period=0&LastNGames=0&ContextMeasure=FGA&PlayerPosition=') % (season,seasontype, player_id)
    response = requests.get(nba_call_url, headers={'User-Agent': "Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36"+
        " (KHTML, like Gecko) Chrome/48.0.2564.82 Safari/537.36"})

    #make it into json
    shots_raw = response.json()['resultSets'][0]['rowSet']

    #makes readerable source
    shots_coords = []

    for row in shots_raw:
        three=0
        if row[12]=='3PT Field Goal':
            three=1
        temp=(row[0], row[1], row[2], row[3], row[4], row[5], row[6], row[7],
                row[8], row[9], row[10],row[11], row[12], row[13], row[14],
                row[15], row[16], row[17],row[18], row[19], row[20], three, row[21])
        shots_coord = ([row[17],row[18]])
        shots_coords.append((shots_coord,)+temp)

    return shots_coords


def create_feature_class(output_gdb, output_feature_class):
    feature_class = os.path.basename(output_feature_class)
    if not arcpy.Exists(output_gdb):
        arcpy.CreateFileGDB_management(os.path.dirname(output_gdb),os.path.basename(output_gdb))
    if not arcpy.Exists(output_feature_class):
        arcpy.CreateFeatureclass_management(output_gdb,feature_class,"POINT","#","DISABLED","DISABLED",
            "PROJCS['WGS_1984_Web_Mercator_Auxiliary_Sphere',GEOGCS['GCS_WGS_1984',DATUM['D_WGS_1984',SPHEROID['WGS_1984',6378137.0,298.257223563]],"+
            "PRIMEM['Greenwich',0.0],UNIT['Degree',0.0174532925199433]],PROJECTION['Mercator_Auxiliary_Sphere'],PARAMETER['False_Easting',0.0],"+
            "PARAMETER['False_Northing',0.0],PARAMETER['Central_Meridian',0.0],PARAMETER['Standard_Parallel_1',0.0],PARAMETER['Auxiliary_Sphere_Type',0.0],"+
            "UNIT['Meter',1.0]]","#","0","0","0")
        #first call shooting_db, secondly call shooting_db.shp
        output_feature_class = output_feature_class + '%s' % ( '.shp')
        arcpy.AddField_management(output_feature_class,"GRID_TYPE","TEXT", "", "", 100)
        arcpy.AddField_management(output_feature_class,"GAME_ID","TEXT", "", "", 100)
        arcpy.AddField_management(output_feature_class,"GAME_EVENT_ID","TEXT", "", "", 100)
        arcpy.AddField_management(output_feature_class,"PLAYER_ID","TEXT", "", "", 100)
        arcpy.AddField_management(output_feature_class,"PLAYER_NAME","TEXT", "", "", 100)
        arcpy.AddField_management(output_feature_class,"TEAM_ID","TEXT", "", "", 100)
        arcpy.AddField_management(output_feature_class,"TEAM_NAME","TEXT", "", "", 100)
        arcpy.AddField_management(output_feature_class,"PERIOD","SHORT", "", "", "")
        arcpy.AddField_management(output_feature_class,"MINUTES_REMAINING","SHORT", "", "", "")
        arcpy.AddField_management(output_feature_class,"SECONDS_REMAINING","SHORT", "", "", "")
        arcpy.AddField_management(output_feature_class,"EVENT_TYPE","TEXT", "", "", 100)
        arcpy.AddField_management(output_feature_class,"ACTION_TYPE","TEXT", "", "", 100)
        arcpy.AddField_management(output_feature_class,"SHOT_TYPE","TEXT", "", "", 100)
        arcpy.AddField_management(output_feature_class,"SHOT_ZONE_BASIC","TEXT", "", "", 100)
        arcpy.AddField_management(output_feature_class,"SHOT_ZONE_AREA","TEXT", "", "", 100)
        arcpy.AddField_management(output_feature_class,"SHOT_ZONE_RANGE","TEXT", "", "", 100)
        arcpy.AddField_management(output_feature_class,"SHOT_DISTANCE","SHORT", "", "", "")
        arcpy.AddField_management(output_feature_class,"LOC_X","DOUBLE", "", "", "")
        arcpy.AddField_management(output_feature_class,"LOC_Y","DOUBLE", "", "", "")
        arcpy.AddField_management(output_feature_class,"SHOT_ATTEMPTED_FLAG","SHORT", "", "", "")
        arcpy.AddField_management(output_feature_class,"SHOT_MADE_FLAG","SHORT", "", "", "")
        arcpy.AddField_management(output_feature_class,"THREE","SHORT", "", "", "")
        arcpy.AddField_management(output_feature_class,"DATE","TEXT", "", "", "")
```

```python
def populate_feature_class(shots_coord, output_feature_class):
    fc_fields = ['Shape','GRID_TYPE','GAME_ID','GAME_EVENT_ID','PLAYER_ID','PLAYER_NAME','TEAM_ID','TEAM_NAME','PERIOD','MINUTES_REMAINING',
    'SECONDS_REMAINING','EVENT_TYPE','ACTION_TYPE','SHOT_TYPE','SHOT_ZONE_BASIC','SHOT_ZONE_AREA','SHOT_ZONE_RANGE','SHOT_DISTANCE','LOC_X','LOC_Y',
    'SHOT_ATTEMPTED_FLAG','SHOT_MADE_FLAG','THREE','DATE']
    c = arcpy.da.InsertCursor(output_feature_class,fc_fields)

    for row in shots_coord:
        c.insertRow(row)

    del c

def main ():
    #set workspce and directory
    arcpy.env.workspace = r'C:\Users\airwo\Documents\ArcGIS'
    os.chdir(r'C:\Users\airwo\Documents\ArcGIS')

    #lst of players, seasons, seasontype
    players = {"All": 0, "Curry": 201939, "Durant" : 201142, "James": 2544, "Irving": 202681, "Jordan": 893}
    seasons = {"15":"2015-16", "16": "2016-17"}
    seasontypes = {"RS":"Regular Season", "PS":"Pre Season", "PO":"Playoffs", "All":"All Star"}

    #specify player, season, seasontype
    player_id =  players["All"]
    season = seasons["16"]
    seasontype = seasontypes["RS"]

    #get shot coordinates from website by get_coord()
    shots_coords = get_coord(player_id, season, seasontype)

    # #setting names for gdb, output feature class and its path
    gdb_name = "NBA_whole_test"
    output_feature_class =  "NBA_whole_2005_2016_test"
    path_to = r'C:\Users\airwo\Documents\ArcGIS\%s.gdb' % (gdb_name)

    #creat feature class in ArcGIS
    create_feature_class(path_to,output_feature_class)

    #popluate values in ArcGIS
    populate_feature_class(shots_coords, output_feature_class)

main()
```