VISUALISATION AND ANALYSIS OF JAMES HARDEN'S DOMINANCE AND STATISTICS IN THE NBA

CHYE Soon Hang, FOO Yong Long, Greg TAN Boon Kit

Abstract—With a recently signed streaming deal worth $24 billion, the scrutiny of an NBA player's performance measured under the various metric standards of efficiency has intensified significantly (Prada, 2014). With copious amounts of data being collected on players' on court performances, there exists a large potential in the analysis of a player's skillsets, efficiency on the court and performance ratings with other players. Acknowledging the existence of a gap in visualisations with today's NBA players using the R Shiny application, the team seeks to utilize the extensive library of the R Shiny interface to build an exploratory visualisation application that dives deep into NBA game data. Using James Harden, the Most Valuable Player (MVP) in the NBA of 2017-18 season as a focal point, the resulting visualisation would provide NBA fans, analytics and basketball enthusiasts a platform that accurately processes and analyzes each NBA player's performance. Utilizing a tailor made 'court look' scatter plot, hexagonal binned chart, bullet charts, boxplots, line graphs and bar graphs, the team aims to integrate the various charts into a user dashboard. With this exploratory visualization tool, users can get instant insights of various key information on a player's information that is not readily available on the NBA website.

Index Terms—Sports Analytics, Player Efficiency, Mid Range Shot, On Court Efficiency, Field Goal Percentage.

1 Introduction

With the increasing popularity of Analytics and Big Data, insights derived from obtaining and analysing huge datasets have made its way to multiple industries across the world. Being the premier league with the largest following across the world in the sport of basketball, the National Basketball Association (NBA) in North America has wholly embraced the use of analytics. This has led to more and more teams adopting analytics in multiple facets, ranging from video to diet tracking (Mudric, 2019). Through this project, the team aims to find out the correlations between an NBA player's shot selection with the overall efficiency in terms of the points, assists and rebounds that the player gets over the course of a game. Using James Harden, the industry dubbed statistical anomaly (Hallihan, 2019) as a primary focal point for the project, the team aims to build a visualisation tool centred around James Harden's shot selection, his field goal percentages and his performance when paired with different players.

Chye Soon Hang is an undergraduate student at the School of Information Systems, Singapore Management University. (e-mail: shchye.2017@sis.smu.edu.sg)
Foo Yong Long is an undergraduate student at the School of Information Systems, Singapore Management University. (e-mail: ylfoo.2017@sis.smu.edu.sg)
Greg Tan Boon Kit is an undergraduate student at the School of Information Systems, Singapore Management University. (e-mail: greg.tan.2017@sis.smu.edu.sg)

By doing so, this visualisation will provide NBA fans and users interested in sports analytics with a visualisation code base that can be used on any specific NBA player. More importantly, users will be able to get instant information that is not readily seen or available through large datasets present on the current NBA page. All in all, this report attempts to report on the research and development efforts taken to design and implement a web visualization application to help respective users to gain insights into an NBA player's efficiency.

The report is being segmented into a few key sections. Section 1 gives a general outline of the project scope and the issue on hand. Section 2 underlines the motivation and objectives that led to the creation of this project. Following which, Section 3 expands on past works related to NBA visualisations. Section 4 will go into detail on the team's approach and reasoning with regards to the various visualisations that have been developed. Section 5 will reveal the team's main visualisations, as well as the design considerations that were assessed while developing this visualisation application. The research paper will conclude by highlighting the directions in which future research can take for possible extensions of this project.

2 Motivation and Objective

While deciding on a research topic, the team realised a key gap in most sports analytics visualisation tools presently available. Despite the wide array of sports analytics visualisations on the NBA, most are done with enterprise software such as Tableau and Microsoft PowerBi. However, there is little to no visualisation done on the NBA using self-implemented coding applications such as R Shiny. This presents a significant untapped opportunity for the team to develop a visualization web application that serves two main pain points:

1. Allow users to modify or expand on existing base code using R Shiny to visualise NBA applications.
2. Provide users with instant information not discernible by looking at raw figures on the NBA website.

On a macro scale, the visualisation is able to provide users who are interested in NBA analytics with an initial understanding on the different metrics being tracked by the NBA, and its effect on determining a player's overall efficiency.

3 Related Works

Through the team's research, the team noticed a clear lack in visualisations options utilising R Shiny in general. Hence, drawing from visualisations done using Tableau, the team gained inspiration and ideas from a few visualisations, mainly:

1. Distributed Shot Charts
2. Player Statistics Breakdown

3.1 Distributed Shot Charts

Distributed Shot Charts play an important role in narrating a player's overall shot distribution and field goal percentage throughout the entire basketball court. With perfect use of colours, it can give us instant insights into the efficiency of a player's shot across the entire court. Fig 1 below highlights a visualisation done using Tableau on the shot chart percentage of the same James Harden for the 2014-15 season.

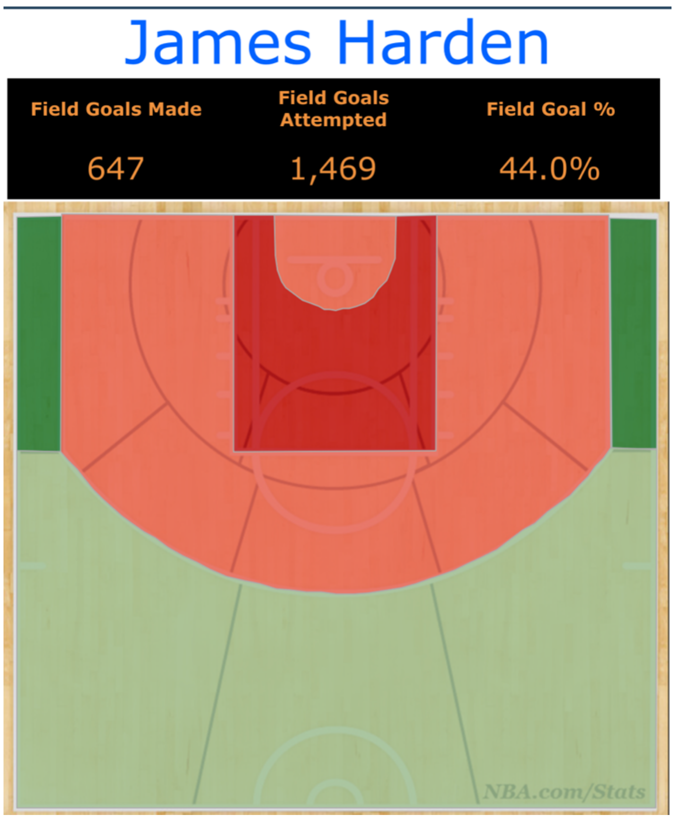

Fig. 1. Tableau shot chart visualisation of James Harden

3.1.1 Points for Improvement

Unclear Exact Shot Location
Despite showing the shot chart and using colours to showcase areas in which a player shoots above or below the NBA league average, there are still bits of information that is unclear from the existing visualisation above. By segmenting the whole court into areas, it is not possible for the user to tell where exactly did the player shoot the basketball. In the realm of sports analytics, such minute bits of information is essential for users to better plan strategies to get the players into their comfortable shooting locations.

Static Datasets
Existing datasets used in this visualisation are unable to dynamically extract data of the chosen player in real time. This presents a challenge for the user to constantly toggle between multiple web pages to make year on year comparisons of the player's efficiency. Such lack of real time data severely hampers the effectiveness of the visualisation tool.

3.2 Player Statistics Breakdown

Utilising data across a wide range of metrics, the player statistics breakdown offers information of the various metrics of a particular player. Looking at Fig.2. below, a breakdown of the various metrics of the player Kobe Bryant, we can instantly see his statistics broken down into various categories.

Fig. 2. Detailed breakdown of Kobe Bryant's statistics

3.2.1 Points for Improvement

Lack of Relative Comparison with other players
Despite showing the player's statistics at first glance, this visualisation does not account for the player's performance when partnering with different players. From a sports analytics standpoint, this is a binary representation of the player's performance, with no option to get a relative comparison of the player's performance as compared to other players in the NBA.

Lack of Comparison within Player's Own Statistics
From Fig. 2 above, we are unable to determine a trend in the player's performance throughout the season. In the scenario in which the user wants to know a player's performance throughout the course of an entire season, it will not be possible. Hence, it is advised that the visualisation shows not just a general average statistic, but a detailed breakdown in terms of the various games the player has played. This will help give a breakdown as to the consistency to which the player is playing at.

Hence, looking at both visualisations, it primarily allows users to obtain insights on an NBA player's shot efficiency on the court, along with various other metrics that determines a player's overall value. However, after highlighting the few points for improvement, this research project aims to further expand on existing visualisations to expose even deeper insights on the various metrics.

4 VISUALIZATION APPROACH

Motivated by a primary aim of providing a visualisation web application that is able to effectively assist users in gaining instant insights on a specific player's efficiency, the team segmented our overall plan of approach into 3 main components:

1. Data Acquisition
2. Data Exploration
3. Iterative Designing

4.1 Data Acquisition

Upon determining the main scope of our project, the team started to narrow down data sources where we could obtain potential information for our visualisations. After thorough research, we primarily obtained our datasets from 2 main sources:

1. NBA official website
2. Basketball-reference.com

These two sources were chosen primarily due to the extensive nature of data that was captured about a player's performance. Furthermore, both the NBA's official website and basketball-reference website are held in high regard with regards to the authenticity and accuracy of data reflected. Hence, the team felt that these two sources offered data that can accurately reflect a player's statistics.

Following confirmation of our data sources, the team ventured into ways to obtain the data. Although data from both websites could be obtained by downloading the csv file, the team felt that the datasets obtained would be static, hence falling into the same cons that existing visualisations faced. Therefore, the team sought to dynamically obtain data from both websites. This approach was made possible through:

1. Obtaining data from NBA website using the NBA API
2. Web Crawling of Basketball-references' website

Both options allowed the team to constantly refresh our data source without having to manually download files from the websites. This gave our visualisation the added advantage of being a dynamic visualisation that displays the most updated information of a player's statistics.

Using python to crawl data off webpages and R as an application to do pre-processing of data, the data obtained were put into dataframes before being extracted for visualisation purposes.

4.2 Data Exploration

Following the data acquisition process, the team sought to run preliminary visualisation modelling. This will help in expediting the process of shortlisting possible visualisations that would best serve our purpose of delivering insights on a player's efficiency on the court. Using the software Tableau, the team was able to instantly visualise and confirm our hypotheses with regards to certain visualisations and its assumed effectiveness in showcasing the insights that the team wished to deliver to the user. The visualisation drafts done using Tableau will be further discussed in later parts of the report.

4.3 Iterative Designing

The final step before moving to officially design our shiny application, was to do initial designing. Using ideas formulated from Tableau, the team sought to first do a low fidelity, hand drawn model of what we envision the dataset could do, before using Tableau to validate our thoughts on the expected visualisation outcome.

4.3.1 Iteration 1: Low Fidelity Design (Hand Drawn)

To obtain an initial idea of how our hypothesized visualisation would turn out, the team attempted to produce the visualisation manually to guarantee proof of concept. This is shown via the various diagrams below.

1. Scatterplot to display Shot Distribution on Court
To better display a player's shot distribution on a basketball court, the team plans to employ a scatterplot, with each point representing a shot on the basketball court. In order to better represent the shot distribution, the team intends to overlay the scatterplot with an image of a basketball court. This will allow the user to get a better representation as to where exactly is the shot being taken.

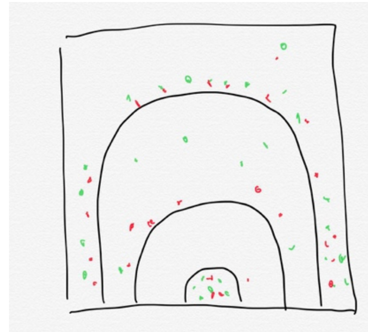

Fig. 3. Scatterplot of NBA Player's Shot Distribution

2. Bar Chart to Compare Player Statistics
The team originally thought of using a line graph to compare various statistics between NBA Players. However, after consultation with Professor Kam, the team decided to use bar charts, which allows for multiple variables to displayed within one visualisation.

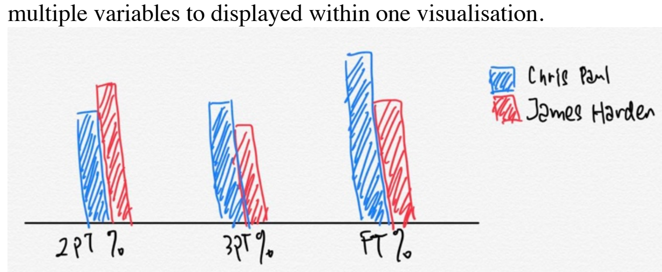

Fig. 4. Bar Chart Comparison of NBA Players

4.3.2 Iteration 2: High Fidelity Design (Tableau)

With a general proof of concept in terms of our hypothesized visualisations, the team went on to use Tableau to design a 'High Fidelity' iteration using Tableau. In Tableau, the team aimed to add and validate filter options that would be added for the actual visualisation done in R Shiny.

1. Bullet Chart Line-up Analysis
Utilising Tableau, the team wanted to determine if there was indeed a correlation between an NBA player's performance with the players around him on the court. Hence, the team decided to use a bullet chart to better display the information the user.

Fig. 5. Bar Chart Comparison of NBA Players

Upon exploration on Tableau, the team decided to change the main visualisation to display various player line-ups showcasing the different metrics (eg. 2 point percentage, free throw percentage), rather than just focusing on one particular metric as seen in Fig 5 above. Using the various line-ups as a filter option, the resulting visualisation would be much better than the existing one done in Tableau.

2. Boxplot Analysis for Player's Metrics
One main challenge the team faced when designing our visualisations was to effectively showcase a player's statistics and how it would change with each game in the season. Initially deciding on a scatter plot, the team later decided against it, as it would be difficult for the user to instantly know the average, upper and low quartiles of each metric being measured. Hence the team decided on a boxplot as the main visualisation for analysing a player's metrics across the course of a season instead.

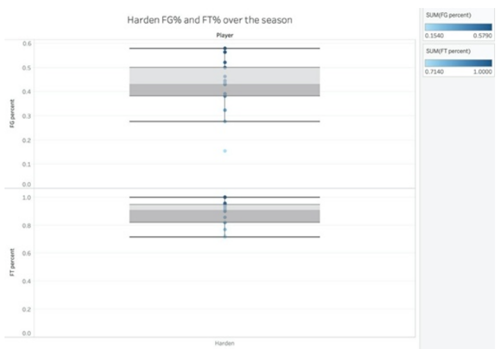

Fig. 6. Boxplot Analysis of Player's Metrics

5 Our Design

Initially, the team did research on ways to integrate the shiny application by R with external JS (Javascript) libraries such as ReactJs, AngularJS and D3.js to create a functional and aesthetically pleasing visualisation application. However, after consultation and advice by Professor Kam, the team did more extensive research on the R Shiny application by itself. This was when the team realised that there was a multitude of libraries also available on the R Shiny platform. Hence, our team decided to primarily focus on using the R Shiny framework.

Following which, the team deployed its visualisation onto shinyapps.io, a dedicated deployment server for Shiny applications. Figure 7 below shows the architecture diagram of the team's proposed visualisation solution.

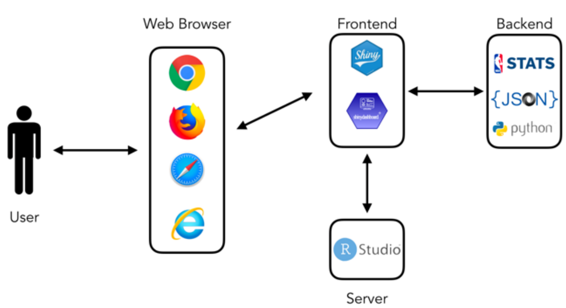

Fig 7. Architecture Diagram

Using the feedback obtained from both our iterations, the team finalised our design on R Studio, and produced a few main visualisations focused on showcasing insights on a player's production and efficiency on a basketball court. To better combine the visualisations that would allow for deeper analysis, the team further grouped the visualisations into two main dashboards, where the visualisations would complement each other into giving more insights that was impossible on their own.

6 Key Findings and Insights

After designing the dashboards, the team noticed a few key aspects of James Harden that propelled him to be one of the most statistically efficient player in the NBA.

6.1 Changing Shot Preference
Utilising the scatterplot for shot selection on a basketball court, the team noticed a surprising trend in the shots taken by James Harden. From the various scatterplots below, the team derived 2 main insights:
1. There are more 3 point and layup shot attempts
2. There is a significant decrease in the number of shots in the 'midrange' area.

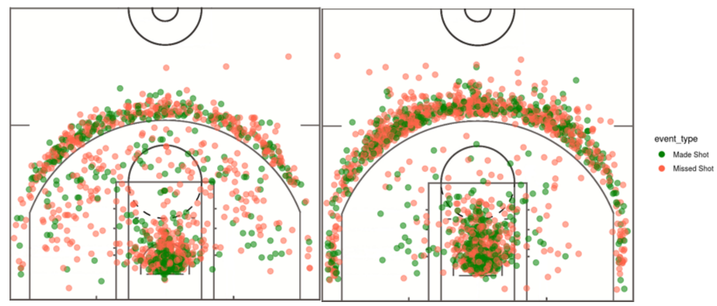

Fig. 8. Scatter Plot of Player's Shot Breakdown

Fig 8 above shows James Harden's shot distribution on the court in seasons 2015/16 and 2018/19 respectively. It can be seen that James Harden has reduced the number of midrange shots taken, instead choosing to take most of his shots from the 3 point range and also right at the basketball hoop. This shot selection ties in with existing analytical references, that depicts the midrange shot as an 'inefficient' shot (Yau, 2019).

6.2 Differentiated Performance with Different Line-ups.
Although the team's initial reason for using James Harden as a focal point was due to his statistical anomalies in production, the team discovered an insight using the bullet chart generated.

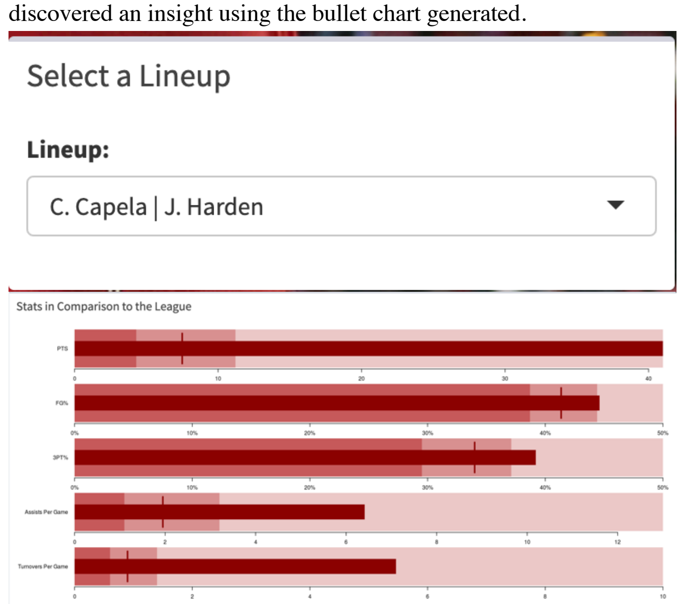

Fig. 9. Bullet Chart of Various James Harden Line-Up Production

Through the use of the bullet chart, the team noticed that James Harden's performance actually fluctuates depending on the players that he plays with. This seems to imply that the quality of team mates actually matters in James Harden's efficiency on the court. Through the bullet chart, the team also noticed that James Harden is statistically most efficient when paired with Clint Capela. Doing further research proves that the James Harden-Clint Capela pairing is one of the best combinations in the NBA for a few seasons (Macmahon, 2018). This would have been something that the team was unaware of without the help of the bullet chart to trigger the research.

6.3 Multiple Occurrences of Above Average Production

Analysing all the metrics that are being used to determine a player's efficiency, the team noticed that James Harden consistently produces games with high scores and high field goal percentage. Although he does have games with low percentages or field goals, he makes up for it with exceedingly high production in others. Overall, his production in weaker performing games are being overshadowed by his high production games.

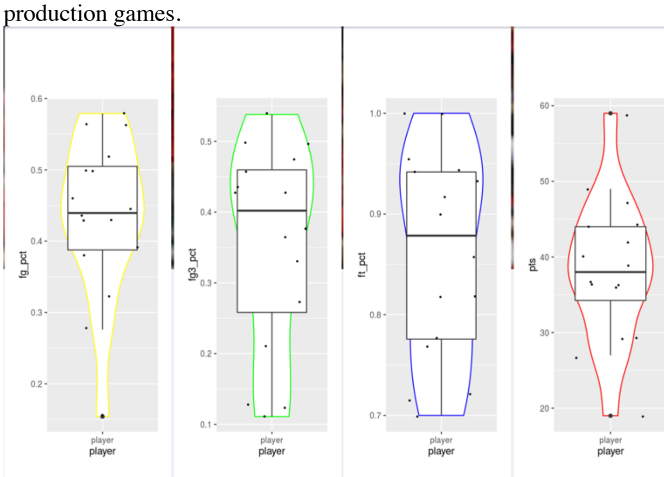

Fig. 10. Player Metrics Breakdown

7 Conclusion

Through this project, the team managed to deeply analyse the performances of James Harden as an individual. Obtaining a better idea of his various metrics, we are able to understand how he is consistently performing amongst the elite in the NBA. With this visualisation, the usage potential can vary depending on the user profile. This brings about many potential ways in which the future extension of this application can progress.

7.1.1 NBA Coaching Staff

With a deeper understanding of James Harden's shot breakdown, opposing team's coaching staff can better game plan against James Harden. With a league that is increasingly focused on analytics, every bit of advantage could make the difference in determining the winning team.

7.1.2 Normal Basketball Players

Being a popular sport worldwide, it is unsurprising for the NBA to be a brand with global presence. By virtue of the presence, more and more basketball players are looking to emulate their favourite NBA player's style of play (Burkhardt, 2019). Hence, this visualisation could be used by players to obtain a clearer idea of their favourite player's style of player through the shot distribution. This would help them to further enhance their emulation of their favourite players.

7.1.3 Sports Analysts

Through the usage of web scraping and APIs, the team has essentially created a code base that could be used by analysts from various sports to use for their own industry. With growing market size and potential in the sports analytics industry, the application created by the team could be extensively used across multiple industries to bring insights of different players, teams and pairings.

To conclude, the project served to confirm the team's initial thoughts that the NBA is shifting to an analytics and data driven business. Therefore, by leveraging on data analytics, the team believes that the future of sports can be essentially aggregated and predicted accurately through the use of data visualisation.

Acknowledgments

The team would like to extend their most heartfelt gratitude to Professor Kam Tin Seong for his teaching and guidance throughout the duration of this project.

References

[1] Prada, M. (2014, October 5). NBA to announce 9-year, $24 billion TV deal with ESPN, Turner. Retrieved November 24, 2019, from https://www.sbnation.com/2014/10/5/6916597/nba-new-tv-deal-espn-turner-24-billion.

[2] Mudric, M. (2019, July 26). How The NBA Data And Analytics Revolution Has Changed The Game. Retrieved from https://www.smartdatacollective.com/how-nba-data-analytics-revolution-has-changed-game/.

[3] Hallihan, M. (2019, January 25). The Knock Against James Harden's Wild Offensive Explosion. Retrieved November 24, 2019, from https://medium.com/the-intermission/the-knock-against-james-hardens-wild-offensive-explosion-88c915ff01f7.

[4] Yau, N. (n.d.). Goodbye, Mid-Range Shot. Retrieved November 24, 2019, from https://flowingdata.com/2019/01/15/goodbye-mid-range-shot/.

[5] Macmahon, T. (2018, May 2). How Harden-to-Capela became the NBA's most-feared connection. Retrieved November 24, 2019, from https://www.espn.com/nba/story/_/id/23376714/how-harden-capela-became-nba-most-feared-connection-nba.

[6] Burkhardt, C. (2019, November 20). Young Blazers grew up watching, emulating Carmelo Anthony. Retrieved November 24, 2019, from https://sports.yahoo.com/young-blazers-grew-watching-emulating-054024927.html.