Five Thirty Great

data science students trying their best!

Data Science Spring 2016 James Jang, Bill Wong, & Keenan Zucker || Olin College of Engineering

Abstract

The game of basketball has gotten increasingly data-driven in the last couple of years, with companies like SportsVU tracking player movements with cameras above all games and companies like Second Spectrum using that information for data analytics and machine learning. We wanted to take on a project that was unique and had not been done before, something that could visualize information in a different way or we could theoretically present to a coach. We explored our interesting data and built a shot predictor model that was significantly more accurate that relying simply on shooting percentage. We combined this with an animation that showed each players 'real' shooting percentage at all times during a simulated game.

Data Exploration

A big part of this project was finding out what we want our final deliverable to be. We didn't come in with a set question in mind. We didn't know if we should focus on teams, players, movement, shooting, passing, defense, the list goes on. We explored around the Internet looking for data we could play with, articles about ideas, and explorations. We looked at exceptional players such as Stephen Curry, and his different shooting measures given unique statistics, such as Shot Clock, Dribbles, Touch Time, and Defender Distance. We also used Heat Maps of all of his shots to give a better look at how phenomenal his shooting really is.

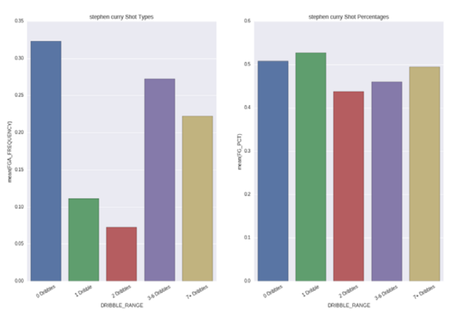


Figure 1. Number of Dribbles

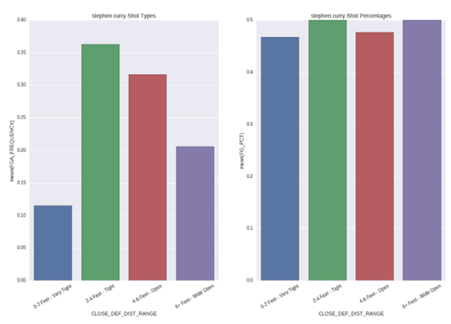


Figure 2. Defender Distance

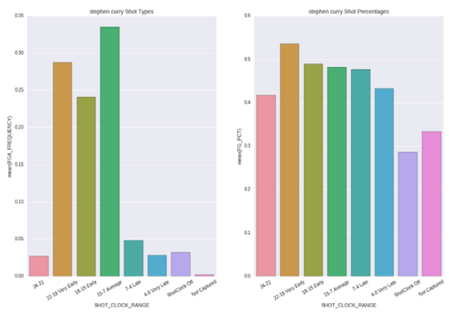


Figure 3. Shot Clock

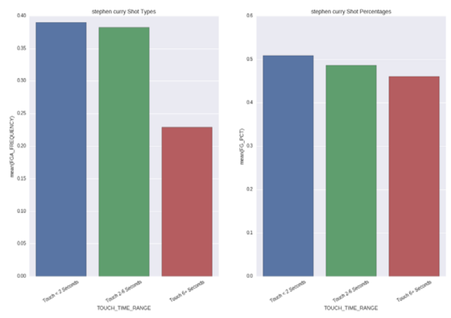
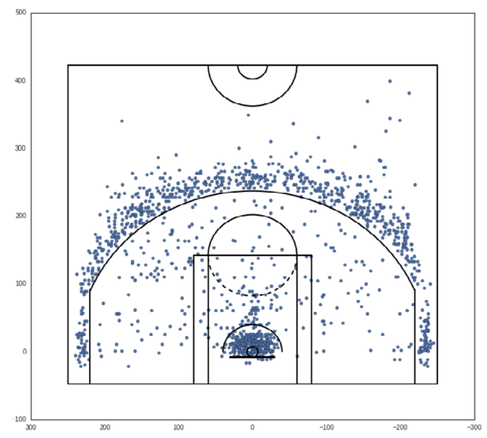


Figure 4. Touch Time



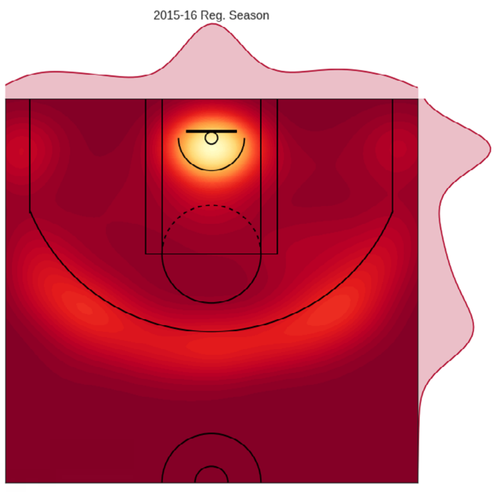


Figure 5 & 6. All of Stephen Curry Shots in 2014-2015

Web Scraping

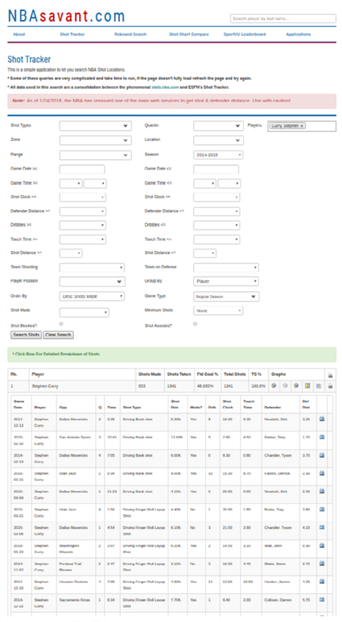


Figure 7. The site to webscrape data

NBA recently took down shot logging for the statistics we explored above. They had the averages for each player, but averages are not useful for our shot prediction tool. Since we felt this data would be very valuable, we decided to look elsewhere for it. We found charts with this saved data on a website called NBA Savant. Using BeautifulSoup, a webscraping library in python, we were able to make a Pandas DataFrame of all of the shots from a inputted player with the metrics we wanted, which we then fed into the shot prediction model.

Shot Predictor Model

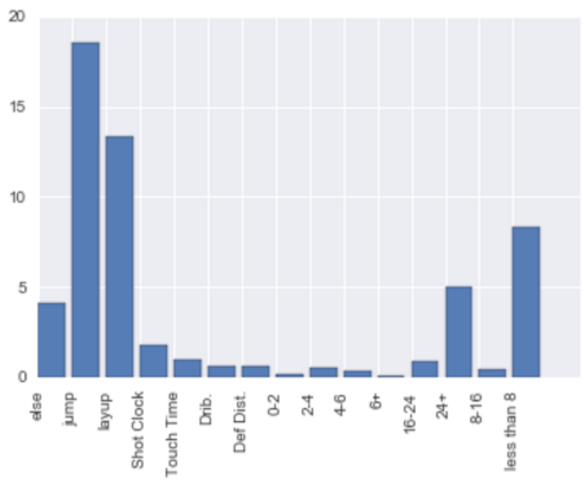


Figure 8. Selecting KBest of features

Normally, a player is judged from their shooting percentage. However, we looked at additional factors that affected shot. Running a Logistic Regression, we tested different features and saw how accurate we were against a baseline test of simply using the shooter's shot percentage. Our model scored significantly better, around 65% accurate on predicting correctly if a shot would go in or not. We tuned which features were the most important, and learned that some factors, like shot type (jump vs layout), we significantly more valuable to our model than things like number of dribbles the player has taken.

Figure 9. Lots of factors affect Curry's shooting stroke

Animation

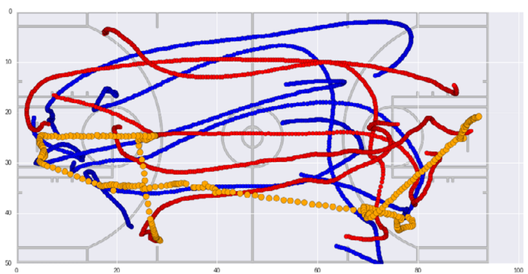


Figure 10. Tracking player movements

From the beginning of the project, we all were interesting in investigating and animation player movement. After finding the data, we could track the players and the ball's location to plot and animate it on a basketball court, basically simulating an old game!

Pulling It All Together

All of this work we achieved led us to an idea of combining it all into an animated visualization that showed a particular basketball game. We preloaded our predictive model to run on the player who was in possession of the ball at all times. Our product resulted in a visualization that showed the effective shot percentage (calculated from our shot model) at all times based on who had the ball. This tool was interesting because we could see when maybe a player should have shot instead of passed, or maybe passed instead of shot. We believe this would be useful tool for coaches reviewing game footage in analyzing the decisions their players make.

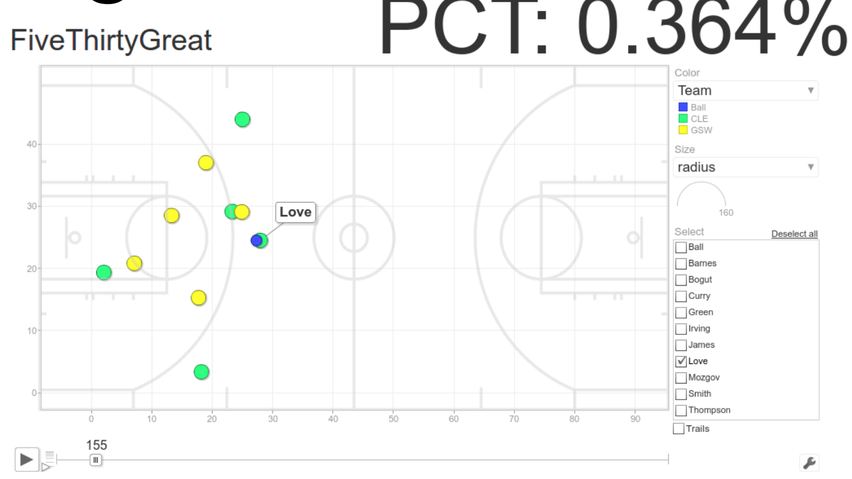


Figure 11. The combination of our shot predictive tool and the game animation.

Next Steps

This project was a sort of proof of concept that we could apply this predictive model to a tool where any player and game could be inputted and visualized using our software.

Acknowledgments

Thanks to Paul Ruvolo for his guidance and feedback.
Sources: Savvas Tjortjoglou, FiveThirtyEight, SportsVU, NBASavant