Final Project Report on STA 141B

A STARTER GUIDE TO HOW TO DEFEAT THE GOLDEN STATE WARRIORS

Submitted by

Tyler Chang

Simon Li

Johnny Xu

Under the Guidance of

Nick Ulle, Ph.D Candidate

Department of Statistics

University of California, Davis

Introduction

For the past 5 years, the Golden State Warriors have undeniably been the most dominant force in the NBA. Within these 5 years, the team's accolades have been nothing less than amazing: winning 4 of the last 5 NBA championships, setting several league-wide team and player records (including a historical 73-9 regular season record), etc.. Lead by head coach, Steve Kerr, a former Chicago Bull during Michael Jordan's era, the Warriors have an unprecedented All-star *only* starting line-up. The firepower on the Warriors has proved to be an obvious problem throughout the league, and poses the question that many teams certainly find themselves asking:

What exactly does it take to beat the Warriors?

From the beginning of our analysis, we want to make it clear that there are *many* different factors that affect the outcome of a basketball game. However, as a start to something that fundamentally could never fully be answered, our group decided to focus on three main ideas- *team characteristics, strategies, and refereeing.*

Team Characteristics

Introduction

Generally, when thinking about how to beat a basketball team, people would think that teams with taller and stronger players have a better chance to win. Therefore, does taller and stronger NBA teams tend to beat the Warriors more often (as the Warriors is known for their lack of height)? If taller teams do have the advantage, then do competitive teams have similar characteristics? Also, the NBA has numerous star players (top 100 players selected by the *Washington Post*) who stand out against the rest. The Warriors have around 7 to 8 star player on their team. Therefore, how many star player do teams need in order to beat the Warriors? With these questions in mind, we started our exploration in team characteristics.

Method

We first looked at the NBA as a whole and determined teams that have a good record against the Warriors in each regular season from 2013-2019. Since teams in the NBA do not play the Warriors the same amount of times (some play 4 times a season while other play 2 or 3 times), we used the winning rate as our standard to measure the competitiveness of each team. Teams with a winning rate over 50% is considered a team that is competitive against the Warriors. These are the teams that we focused on for our analysis.

Result

- **AVERAGE WEIGHT, HEIGHT, AND TEAM COMPOSITION:** The first aspects of team characteristics we investigated were the average weight, height, and team composition (number of players in different positions) for all teams. Although the average height and weight of starters is reducing for all teams in the NBA, there is no direct relationship between weight, height, and team competitiveness. Similarly, there is also no obvious correlation between team composition and their winning record against the Warriors.
- **NUMBER OF STAR PLAYERS:** We then checked the number of star players on these competitive teams. The data shows, on average, a team needs to have at least 3 to 4 star players in order to beat the Warriors. In addition, among these star players, each team needs

to have around 2 All Star players (selected by NBA fans and coaches annually) to be competitive. Here is the result we got for teams and their star players in the season of 2017-2018.

Teams	Number of Star Players (All Star)
Golden State Warriors	5 (3 All Star)
Utah Jazz	5 (2 All Star)
Houston Rockets	5 (3 All Star)
Indiana Pacers	3 (1 All Star)
Portland Trail Blazers	3 (2 All Star)

Game Strategy

As mentioned above, team characteristics is an important aspect to win against the Warriors. However, another important aspect is game strategy. Some questions we will investigate include: *What kind of shots should a team focus on (2 point or 3 point)? Are Free throws an important factor? Should teams be attempting more assists?*

Focus:

When thinking of how to devise a basketball game strategy, it can be broken down into two parts- *offense and defense*. On the offensive front, we focus on the factors that help the team score the most points. For defense, we care about the factors that prevent the opponent from scoring. Using both sides, we can observe the "best" strategies that teams successful against the Warriors use most commonly.

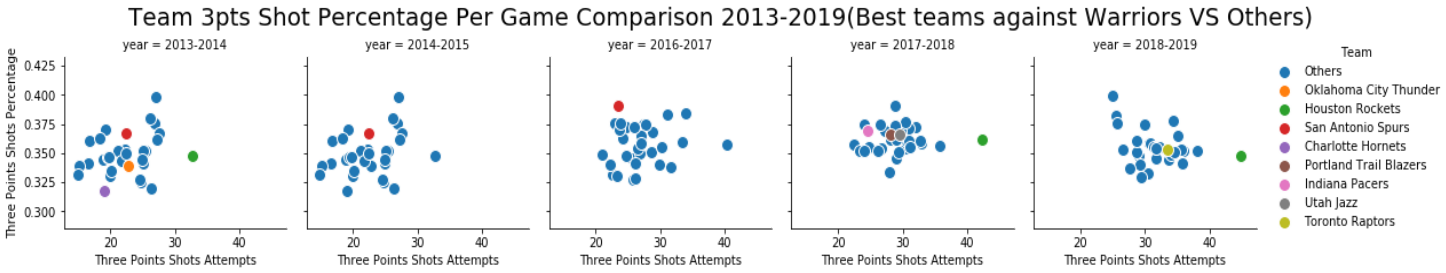
Analysis:

1.Part One Offense:

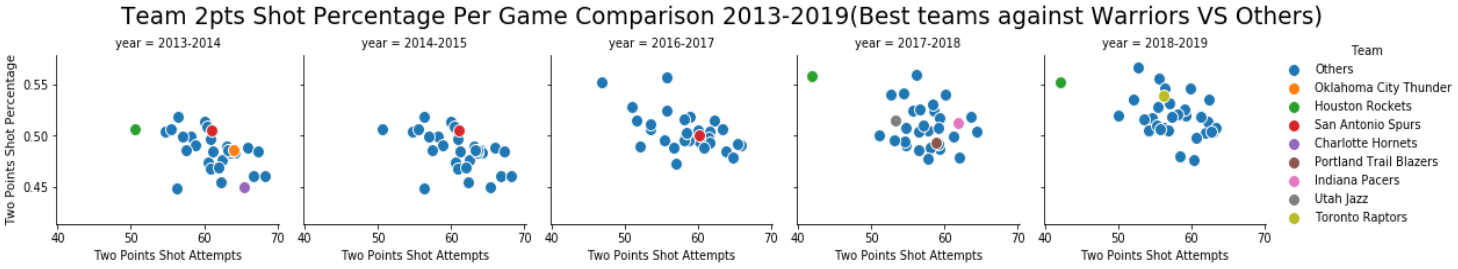
(1) Shooting Strategy (3s V.S. 2s)

Throughout the NBA, the Warriors are famous for their deadly three-point shooting ability. Therefore, we are interested in determining whether other teams should do the same or focus more on attempting twos.

The figure above shows a scatterplot of teams' three-point percentage against the amount of shots attempted for each season. One interesting case, can be seen with the San Antonio Spurs in the 2014-15 and 2015-16 regular season. The Spurs was the only team for both seasons that had a

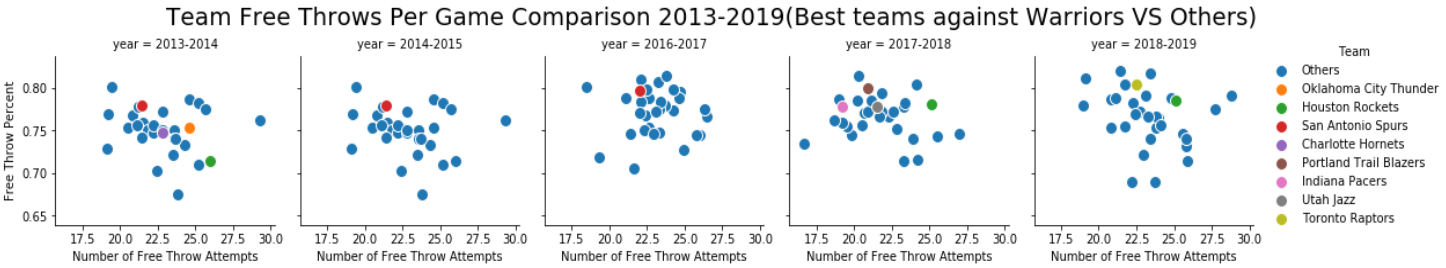


winning record against the Warriors over 50 percent. Although they did not attempt a significant amount of 3's, their 3-point percentage was among the top 5 in the 2014-15 season, and number one in the 2015-16 season. Another interesting case is with the Houston Rockets in the 2017-18 and 2018-19, a team that shot a significant number of 3-point attempts. Although their 3-point shooting percentage is approximately average relative to the league, the Rockets build their advantage through attempting *a lot* more than other teams.



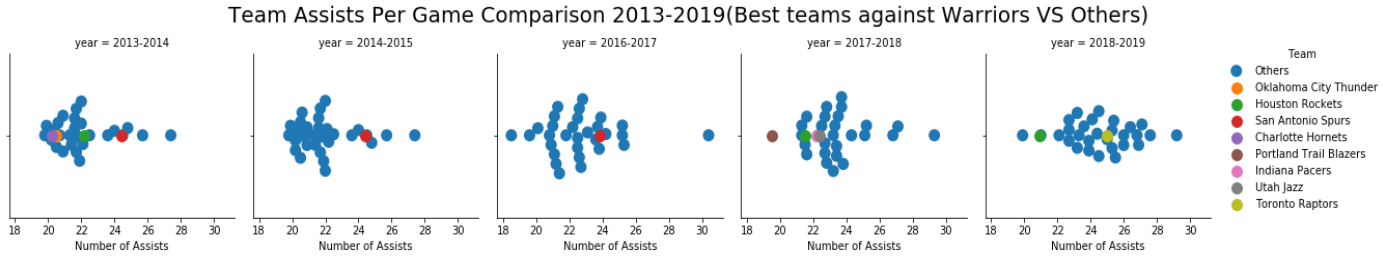
This next figure shows a similar story, but for 2-pointers. Continuing with the analysis above, we can see that the Rockets take significantly less 2-point attempts than all of the other teams in the league. Although their attempts are much less than average, their efficiency is among *top 3* within the last two seasons.

(2) Free Throws



The second part of our offensive analysis focuses on free throws. One free throw could essentially result in two points, so the importance of having a good free throw strategy is crucial. By observing the teams with success against the Warriors, it is apparent that these teams are attempting more free throws on average. Also, their free throw percentage is among the best in the league.

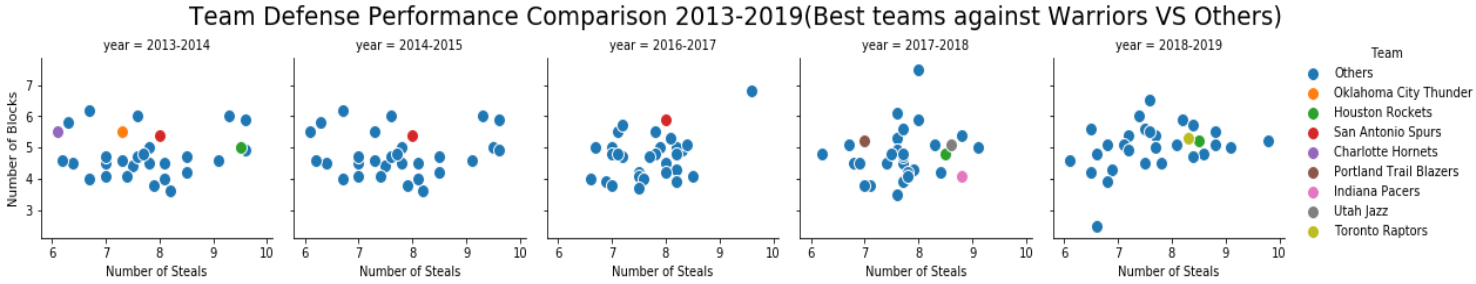
(3) Assists



The final part in our offensive analysis breaks down a team's assists. "An assist is attributed to a player who passes the ball to a teammate in a way that leads to a score by a field goal (*Wikipedia)*." We can see that there some teams that have a high number of assists per game, but there also teams that do not. Therefore, there is no evident that assists has much to do with winning chances against the Warriors

2.Part Two Defense:

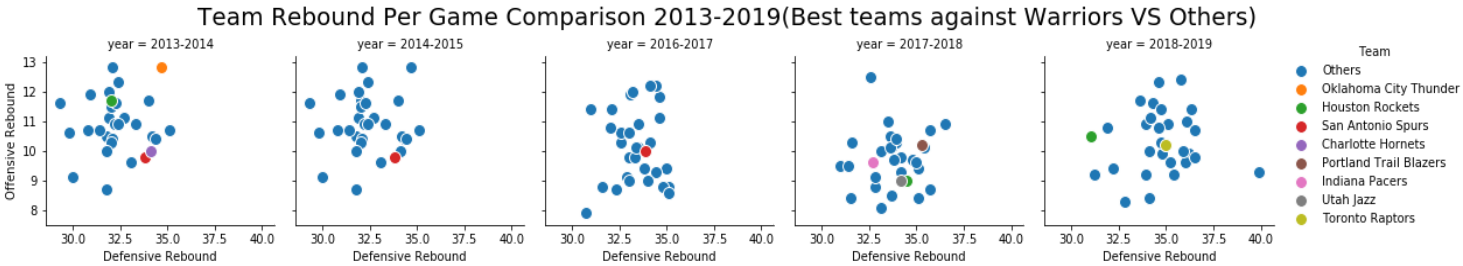
(1) Blocks and Steals



When starting a defensive analysis, we first focused on number of steals and blocked shots. Teams that play well against the Warriors tend to have a high number of both steals and blocks per game. As supporting evidence, we can look at Kawhi Leonard, former San Antonio Spur and former two-time Defensive Player of the Year. Interestingly, when Leonard moved to the Toronto Raptors this season, the Raptors currently have the best record against the Warriors.

(2) Rebounds

The second area of focus for defense is with rebounding. A defensive rebound occurs when the defending team is able to obtain possession of the ball following a missed shot. Similarly, an



offensive rebound occurs when the team on offense maintains possession of the ball following their own missed attempt. We can see that a high number of defensive rebounds is more prevalent in the teams successful against the Warriors. However, there is no significant pattern in offensive rebounds.

Conclusion

Shot Selection:

- **Attempt more 3-pointers and attempt efficient 2-pointers**
- **Shoot 3-pointers conservatively, but efficiently.**
- **Create more opportunities to shoot free throws**
- **Shoot free throws well**

Defense Strategy:

- **Focus on playing good defense (steals, blocks, especially defensive rebounds, etc.)**
- **Obtain more defensive-minded players (Kawhi Leonard)**

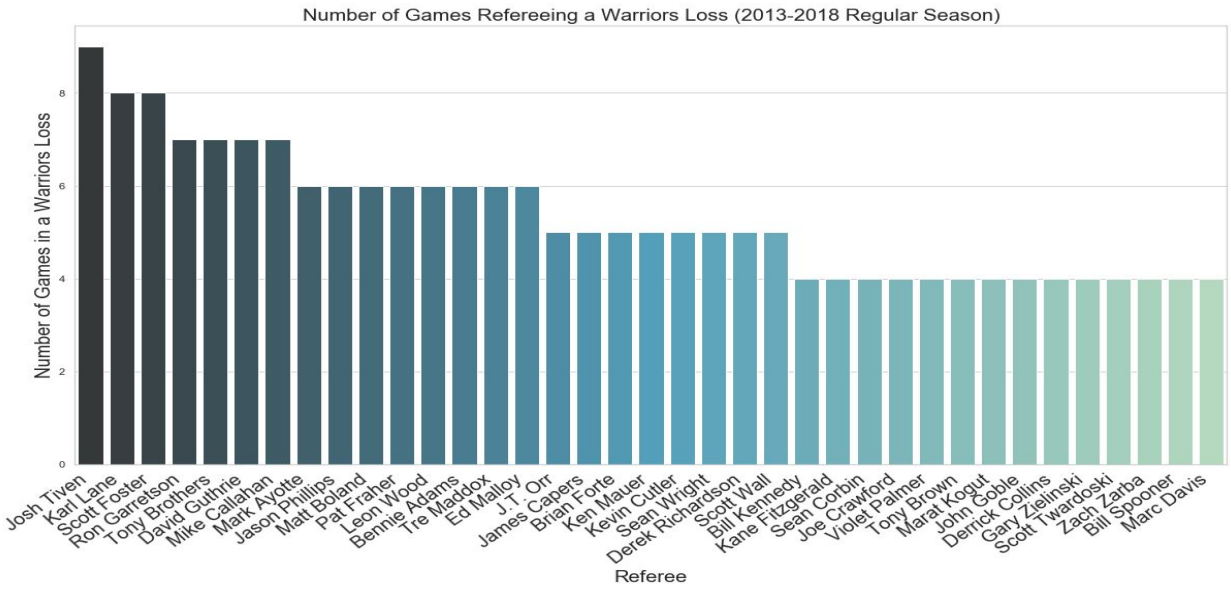
Refereeing

Refereeing is one of the most overlooked aspects in whether a team wins or loses. Especially in the game of basketball, where one blown whistle can result in one made basket, referees have a significant power over a game's outcome. In the case of the Warriors, a team that does not lose frequently, we are interested in seeing the effects a referee can have on their winning chances.

Step 1: Determining "Referees of Interest"

The first part of the analysis begins with figuring out the referees that could potentially be a problem for the Warriors. One way to distinguish these referees of interest, is to determine which

officials are consistently working the games the Warriors lose. The distribution of the number of Warrior Losses a referee appears in is heavily skewed right. To define "consistently," a practical approach would be to investigate the referees that appear in more games than the average (4 losses).



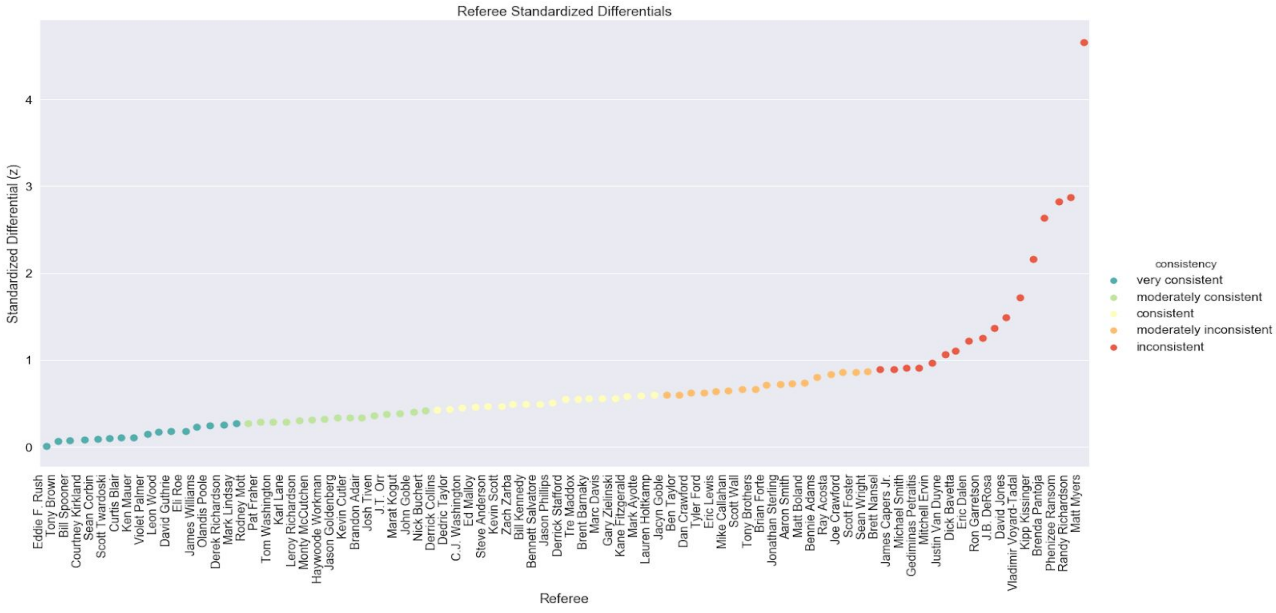
Step 2: Determining a "Bad" Referee

Now that we have a list of "suspects" to look into, the next step would be to look at all of the referees in the NBA and create a definition of what a "bad" referee is. Unfortunately, the data on the NBA's referees is quite limited to bland statistics made possible by *basketball-reference.com*. With the resources available, it was quite difficult to define a "bad" referee. Instead, we decided to examine a referee's overall consistency.

basketball-reference.com conveniently has a table of all of the referees for a given season. Within the table holds several average referee statistics for that season, such total Field Goal Attempts (FGA), Free Throw Attempts (FTA), Personal Fouls (PF), and Points (PTS). Not only does the website have this information for each referee ***per game***, but there also is a section on how the referee's individual statistics compare ***relative to the average***. For each statistic, it shows the referee's differential (*Referee's statistic - Average statistic for all referees*).

			Per Game				Per Game Relative			
Referee	**Lg**	**G**	**FGA**	**FTA**	**PF**	**PTS**	**FGA**	**FTA**	**PF**	**PTS**
Bennie Adams	NBA	59	170.7	48.1	41.7	210.0	-0.1	+1.9	+1.9	-1.2
Steve Anderson	NBA	46	170.7	46.4	40.7	211.9	-0.1	+0.2	+0.9	+0.7
Mark Ayotte	NBA	55	171.4	48.4	41.6	214.0	+0.6	+2.2	+1.8	+2.8

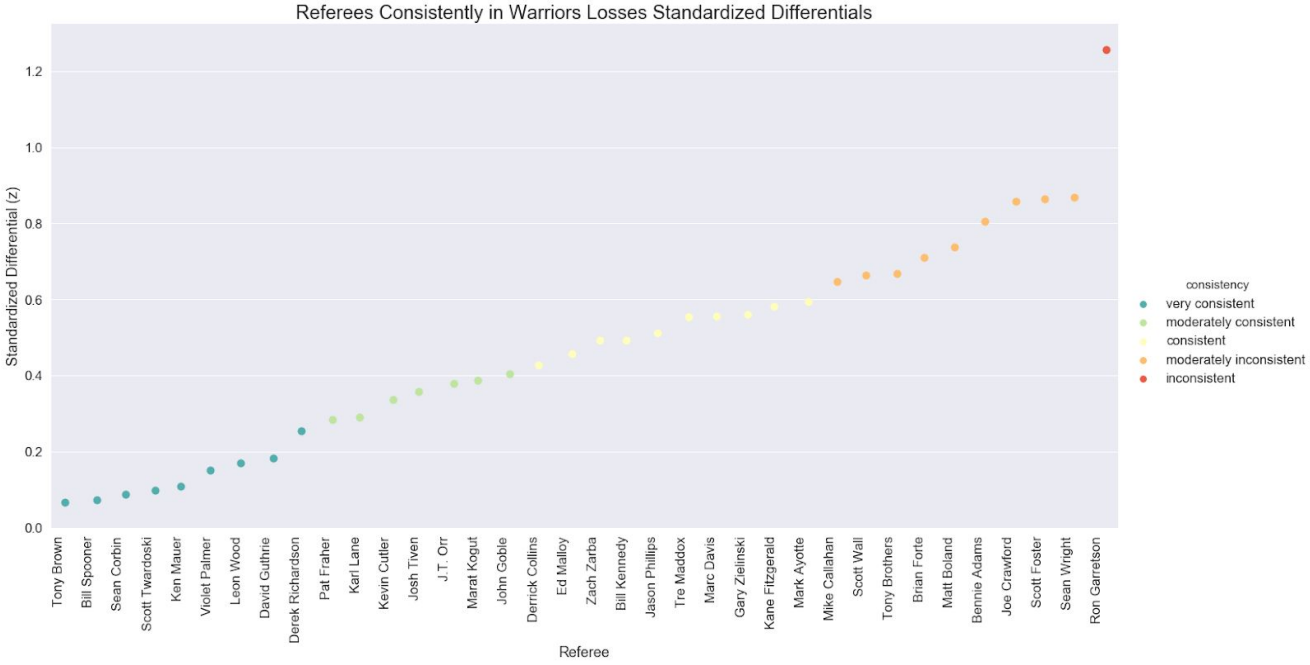
With this information, we were able to aggregate and take the average “Per Game Relative” statistics of all of the referees from 2013-2018. After standardizing each column of per game relative statistics, we took the sum of each referee’s statistics and standardized that value again. Since z-scores represent the number of standard deviations a value is from the average, these calculated sums, in magnitude, are essentially a representation of how consistent a referee is with the “average referee.” The higher this statistic is for a referee, the “more inconsistent” they are. Instead of just making a binary cut-off of a consistent and inconsistent referee, we decided to instead split the referees into different tiers of consistency.

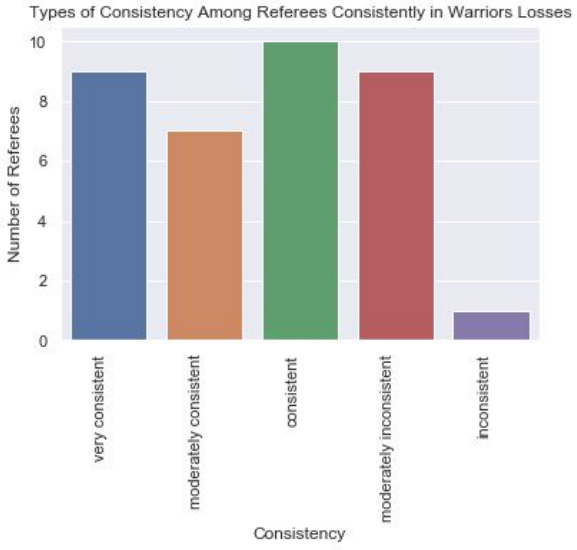


Step 3: Determining Types of Referees in Warriors Losses

As a reminder for the motivation of the analysis, the first step was obtaining the referees that could potentially be a problem for the Warriors. These are the referees that could are consistently in games that the team loses. Next, we were able to classify how consistent a

referee was. The next natural step would be to use this classification to figure out how consistent are the referees that are consistently in the games the Warriors lose.





Conclusion

The data shows that the Warriors tend to get more consistent refs, even in the games that they lose. In fact, approximately 72.2% of these referees had a consistency of consistent or better. Although it would have been interesting to find more inconsistent referees in the games the Warriors lose, it is worth mentioning that the most important factor in winning a basketball

game comes down to how many points a team can score. Therefore, it should not be surprising that refereeing may not have a significant effect over the Warrior's winning chances. All else equal, there is not enough evidence to conclude that having a "good" or "bad' referee has an effect on the Warrior's chances of winning.

LOGISTIC REGRESSION

INTRODUCTION

This next part will utilize logistic regression to shed light on the most important factors when beating the Warriors. Using whether a team can beat the Warriors or not as a response variable, we fitted our data with logistic regression to make inferences on how certain aspects can increase the probability of beating the Warriors based on their individual team.

METHOD

First, we checked the assumptions for logistic regression. Assuming each game is independent, our response variable (win or lose) follows a binomial distribution. In addition, we checked the correlation between all of the explanatory variables. It appears that some variables such as "field goal attempts", "field goals made", "free throw attempts" are correlated. Therefore, these variables are removed to eliminate multicollinearity.

After cleaning the dataset, we used Backward-Forward model selection with Bayesian Information Criterion (BIC) to select a model that is good for both prediction and making inferences.

RESULT

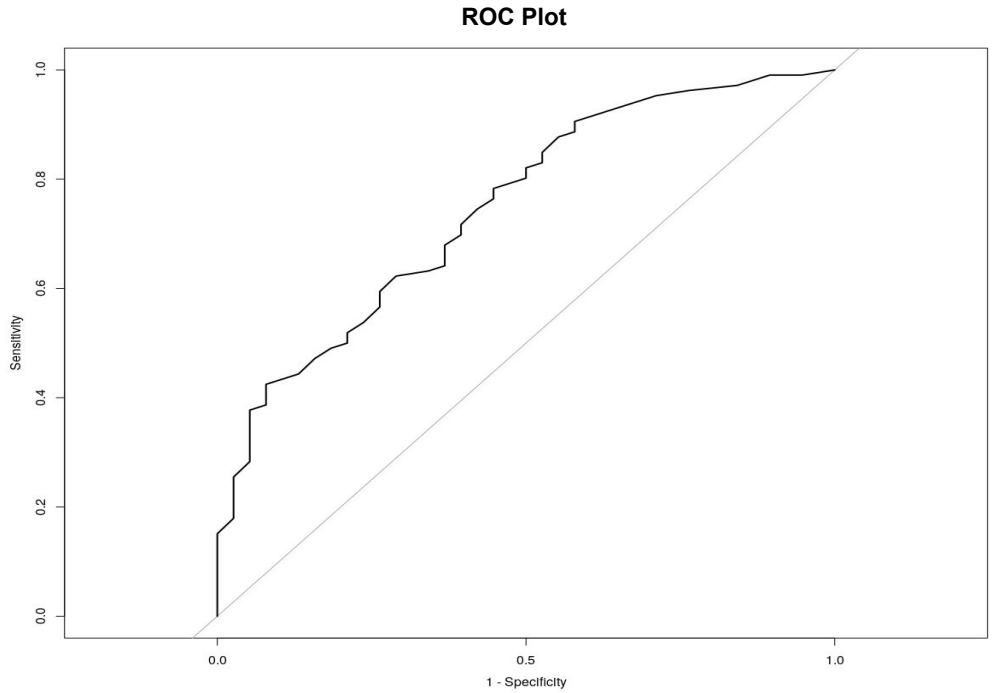
To avoid complication and potentially overfitting our model, we did not investigate in any interactions between the explanatory variables. The best non-interactive logistic regression model we end up with is:

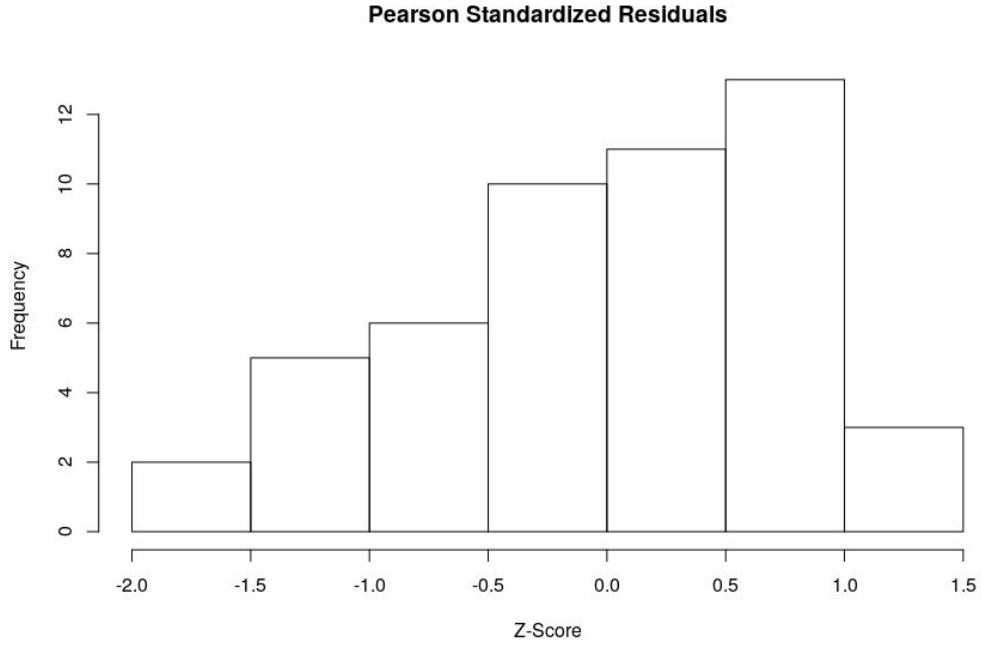
$$\ln(\frac{\pi_{win}}{1-\pi_{win}}) = -47.7582773 - 0.5310475X_{FieldGoal} + 68.6301354X_{3Point\%} + 38.2482750X_{2Point\%} -$$

$$0.4075792X_{FreeThrowAttempt} + 18.3725436X_{FreeThrow\%} + 0.8863806X_{OffensiveRebound} + 0.9368518X_{Steals}$$

With the selected best model, we calculated the sensitivity, specificity, and the error rate. With a sensitivity of 93.4%, it turns out that the model is good at predicting whether the team can beat the Warriors. Our model, on the other hand, is not very good at predicting whether a team will lose to the Warriors, as the specificity is only 34.2%. The overall error rate of out model is 22.2%,

which is acceptable. To further test the accuracy of our data, the area under the curve (AUC) of the Receiver Operating Characteristic (ROC) plot (Sensitivity vs (1 - Specificity)) calculates to 0.748, which indicates that our model does a reasonably good job at predicting the outcome of a game. Furthermore, we also plotted the Pearson standardized residuals and it turns out that there are no significant residuals in out dataset.





CONCLUSION

Since the explanatory variables involved in building this model are continuous, the interpretation is straightforward. Based on the properties of logistic regression , we can see that for a multiple logistic regression, we get $\ln(\frac{\pi}{1-\pi}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \cdots$ after exponentiating both sides of the equation, we get $\frac{\pi}{1-\pi} = e^{\beta_0} \times e^{\beta_1 X_1} \times e^{\beta_2 X_2} \cdots$ Therefore, we can interpret the value of the coefficients in terms of the odds ratio (probability of winning vs probability of losing). In our case, we have the odds ratio of the probability that a team successfully beats the Warriors. Therefore, we see from the general form of the model that the odds ratio and the probability are positively correlated. The odds ratio will increase when e^{β_i} is greater than one. Thus, we can see that β_i's with positive values will increase the odds ratio (probability) of a team beating the Warriors if we hold other variables constant since the exponential of a positive number is strictly bigger than 1.

Based on the logistic regression model, we can see that the coefficient for variables representing "**3 point percentage,**" "**2 point percentage,**" "**free throw percentage,**" "**offensive rebound,**" and "**steals**" are all positive, which indicates that improvements made in these aspects will increase the odds of a team beating the Warriors.

On the other hand, we see that coefficients of "**field goal attempts**" and "**free throw attempts**" are negative, which indicates that teams should reduce their field goal and free throw attempts in order to increase their chance of beating the Warriors.

Ignoring the sign of the coefficients, the p-values corresponding to the Wald statistics and the Wald confidence interval for each coefficient was calculated in R (appendix I). Notice that the null hypothesis in this case is that the variable does not have a significant effect on the response. Therefore, a smaller p-value indicates that the variable actually has significant effects. According to logistic regression, three point percentage and offensive rebounds are the most important factors for a team to be competitive, as they have the smallest p-values in this case. We can further prove our conclusion by looking at the confidence intervals for each coefficient. The Wald confidence intervals for coefficients other than **three point percentage** and **offensive rebounds** all include 0, which means that they have a very weak effect on the response variable as $e^0 = 1$.

To conclude, our model demonstrates that, if a team wants to beat the Warriors, they need to be very efficient in shooting percentage, offensive rebounds, and steals. Therefore, they do not need many shot attempts to overcome the Warriors. The conclusion we have drawn from our logistic regression model matches the conclusion we made in the strategy part. Therefore, it is safe to conclude that shooting efficiency and defensive effort are the two most important factors for a team to be competitive against the Warriors.

LIMITATIONS

- **All our data are collected from regular season of NBA basketball, which makes out result not representative for playoff games**
- **Could not scrape data from stats.nba.com since the whole site is rendered by javascript. Result would be more reliable since the data from stats.nba.com covers more aspects of basketball.**

GAME STRATEGY

In this part of our analysis, we used team data that was on a per game basis, *for all teams*. Although this is still viable information, it would be more useful to use data on how teams do *specifically* against the Warriors. We also have completed this analysis by looking at teams as a whole, not taking into account individual player performance. We have also assumed that these strategies are independent of each other, which may not be the case. There are also many other confounding variables that we have not yet looked at such as player injury, game location (home vs. away), etc.

LOGISTIC REGRESSION

The biggest limitation in logistic regression is that we assumed that each game is independent of each other, which is not always true in reality. There are many factors that can influence the outcome of a game such as injury, back-to-back games, etc. However, independence is a crucial assumption to logistic regression, so it was necessary to perform our analysis. In addition, the sample size we used for logistic regression was 144 games, which is small considering the number of variables we are exploring. Therefore, our regression model can be improved by increasing the sample size.

REFEREEING

Unfortunately, the data on referees in the NBA is *very* limited. Not only does the official NBA website have absolutely no data on referees, there only exists a handful of websites that have credible information. That being said, the website that our project gets the majority of our data from (*basketball-reference.com*) has referee data with complexity going as far as computing simple average statistics. One limitation to the data set is given with the Personal Foul statistic. The website explicitly states that “personal fouls include all fouls, not just fouls called by this referee.” This is a problem because we have no idea if their PF statistic is an actual representation of their own refereeing or a combination of many referees. Although it may not be a perfect representation, the referee’s differential (consistency) is a combination of 3 other statistics, and should not have a significant effect on determining how consistent a referee is.

APPENDIX I.

WALD TEST STATISTICS	TEST STATISTICS	P-VALUES
INTERCEPT	-3.298782	0.0009710543
Field Goal	-2.089701	0.0366446490
3 POINT PERCENTAGE	3.278215	0.0010446580
2 POINT PERCENTAGE	1.933821	0.0531351930
FREE THROW ATTEMPT	-2.044768	0.0408777296
FREE THROW PERCENTAGE	1.685378	0.0919157452
OFFENSIVE REBOUND	2.757358	0.0058270563
STEALS	2.527473	0.0114886641

WALD CONFIDENCE INTERVAL	LOWER BOUNDS	UPPER BOUNDS
INTERCEPT	-85.04997607	-10.4665784
Field Goal	-1.18563270	0.1235378
3 POINT PERCENTAGE	14.70459530	122.5556756
2 POINT PERCENTAGE	-12.69803781	89.1945878
FREE THROW ATTEMPT	-0.92101369	0.1058553
FREE THROW PERCENTAGE	-9.70694230	46.4520295
OFFENSIVE REBOUND	0.05835416	1.7144071
STEALS	-0.01792409	1.8916276

CITATION:

"Analysis | The Top 100 Players for the 2017-18 NBA Season." The Washington Post. Accessed March 21, 2019. https://www.washingtonpost.com/graphics/2017/sports/nba-top-100-players-2017/.

"Analysis | The Top 100 Players for the 2016-17 NBA Season." The Washington Post. Accessed March 21, 2019. https://www.washingtonpost.com/graphics/2016/sports/nba-top-100-players-2016/.

"Assist (basketball)." Wikipedia. December 22, 2018. Accessed March 19, 2019. https://en.wikipedia.org/wiki/Assist_(basketball).

"Basketball Positions." Wikipedia. March 17, 2019. Accessed March 19, 2019. https://en.wikipedia.org/wiki/Basketball_positions.

Chris Dardis (2015). LogisticDx: Diagnostic Tests for Models with a Binomial Response. R package version 0.2.
https://CRAN.R-project.org/package=LogisticDx

"Field Goal (basketball)." Wikipedia. March 19, 2019. Accessed March 19, 2019. https://en.wikipedia.org/wiki/Field_goal_(basketball).

"How the Blazers Have Been One of the NBA's Best Defensive Teams." KGW. November 16, 2017. Accessed March 19, 2019. https://www.kgw.com/article/sports/nba/blazers/how-the-blazers-have-been-one-of-the-nbas-best-defensive-teams/283-492192099.

"Kawhi Leonard." Wikipedia. March 02, 2019. Accessed March 19, 2019. https://en.wikipedia.org/wiki/Kawhi_Leonard.

"Rebound (basketball)." Wikipedia. March 06, 2019. Accessed March 19, 2019. https://en.wikipedia.org/wiki/Rebound_(basketball).

"WikiProject National Basketball Association/National Basketball Association Team Abbreviations." Wikipedia. September 27, 2018. Accessed March 21, 2019. https://en.wikipedia.org/wiki/Wikipedia:WikiProject_National_Basketball_Association/National_Basketball_Association_team_abbreviations.

Xavier Robin, Natacha Turck, Alexandre Hainard, Natalia Tiberti, Frederique Lisacek, Jean-Charles Sanchez and Markus Muller (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics, 12, p. 77. DOI 10.1186/1471-2105-12-77

http://www.biomedcentral.com/1471-2105/12/77