

Shots! Shots! Shots!

Brian Sachtjen, Dylan Greenleaf, Yuanbo Wang, Danny Lee

Why Predict FG Percentage?

- Player and team evaluation
- Opportunity to influence coaching strategy
- Identify non-traditional predictors of shooting efficiency



Source: www.bleacherreport.com

Data Sources

- All taken from stats.nba.com
- Shooting Logs (2013 and 2014)

Game	Loc	W/L	Final Margin	Shot #	Period	Game Clock	Shot Clock	Dribbles	Touch Time	Shot Dist
DEC 02, 2015 - GSW @ CHA	A	W	17	1	1	10:45		0	0.6	23.3
DEC 02, 2015 - GSW @ CHA	A	W	17	2	1	5:47		9	7.6	24

- Play-by-play Game Logs (2014)

Oklahoma City Thunder	Orlando Magic
Start of Q1	
	12:00 Jump Ball Vucevic vs. Adams: Tip to Ibaka
Ibaka 14' Jump Shot (2 PTS) (Westbrook 1 AST)	11:40 2 - 0
	11:26 Oladipo 3' Cutting Layup Shot (2 PTS) (Harris 1 AST)
	2 - 2

Data Merging

● Issues

- Some shots had different timestamps in each dataset
- Some shots exist in one set but not the other

● Solution

- Join on “closest” shot for that playerId and gameId
- Drop observations where match could not be found (3% of all shots)

Game	Loc	W/L	Final Margin	Shot #	Period	Game Clock
NOV 30, 2015 - GSW @ UTA	A	W	3	1	1	11:15

Golden State Warriors	Utah Jazz
Start of Q1	
12:00	Jump Ball Gobert vs. Bogut: Tip to Favors
11:40	MISS Hood 19' Jump Shot
11:39	JAZZ Rebound
11:30	Gobert Alley Oop Dunk (2 PTS) (Hood 1 AST)
Curry 26' 3PT Jump Shot (3 PTS) (Green 1 AST)	11:12 3 - 2

Data Cleaning/New Metric Creation

● Data Cleaning

- Looked for specific words in the game log descriptions to determine if observation was related to a shot
- Eliminated observations with non-sensical values
 - shot clock times <0 or >24
 - 'touch time' values that were negative

● New Metrics

- 2013 FG Percentage
- 2013 Allowed FG percentage
- Player Position
- Time elapsed since substitution
 - Create 'sub_in' variable
 - For each observation find the last time the relevant player was subbed in and compute difference of time_left values

Predictors

- Shooting Logs

- Location (home/away)
- Shot number
- Shot clock
- Dribbles
- Touch time
- Shot distance
- Shot type (2pt and 3pt)
- Closest defender distance

- Event Logs/Other

- **Game score margin**
- **Position**
- **Time since last sub**
- **2013 shooting %**
- **Defender's 2013 shooting % allowed**

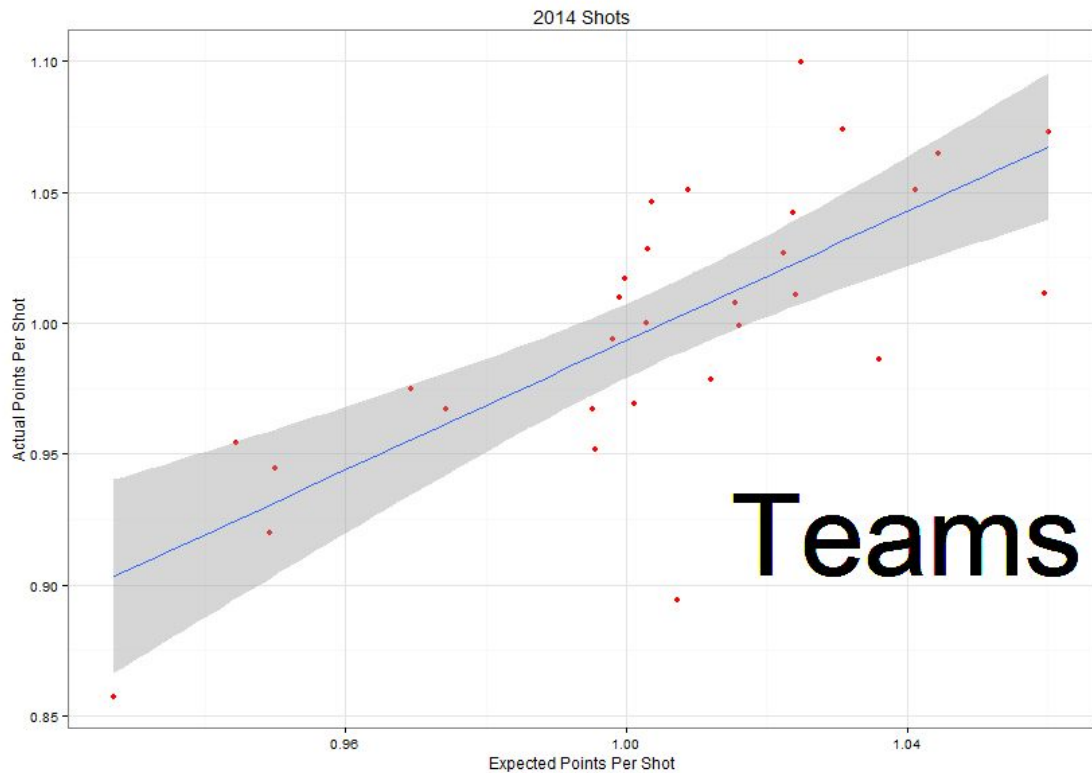
Bold = statistic that we manually created

Finally...the model!

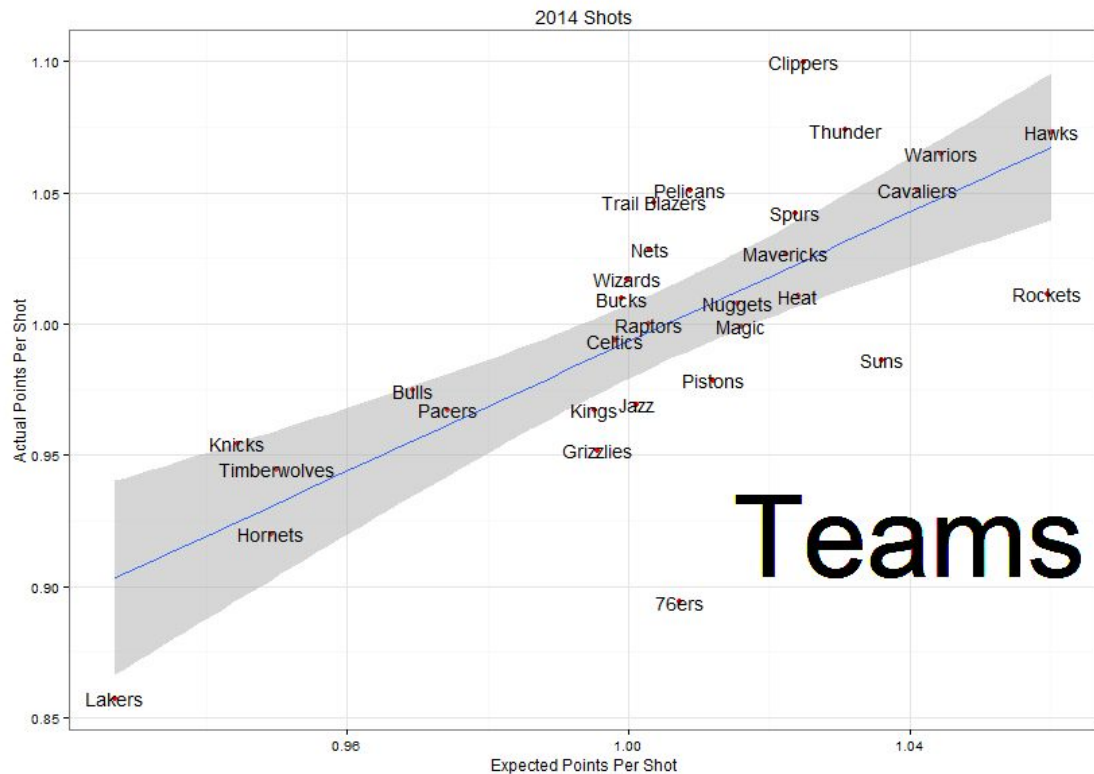
- 169k observations - split into train/test
- Created logistic regression model to predict field goal % for a given shot
 - Accuracy = 61%
- Multiplied predicted values by shot type (2 or 3) to create “expected points per shot”

Variable	Coefficient Sign
Score Margin	-
Location (H)	+
Shot number	+
Shot clock	+
Touch time	-
Dribbles	N/A
Shot distance	-
Pts type (3)	-
Closest defender distance	+
FGperc	+
Def FGperc	N/A
Position (F)	+
Position (G)	+
Time since sub (500-1000)	-
Time since sub (1000-1500)	-

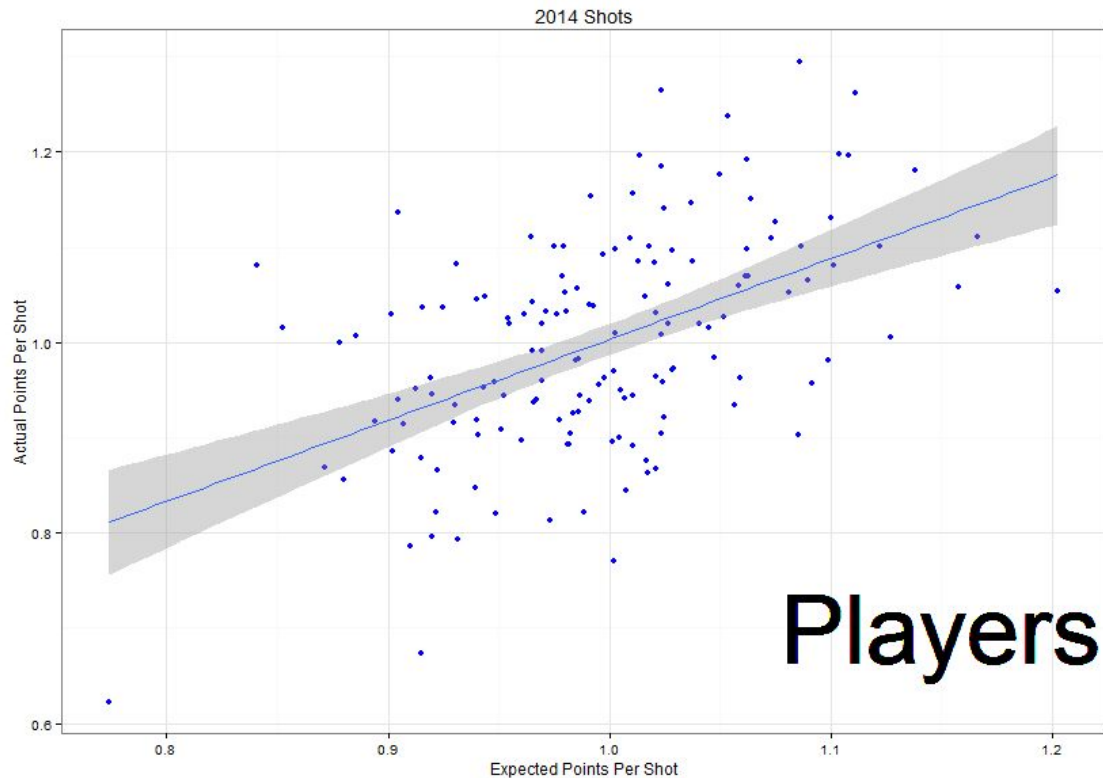
Team Evaluation



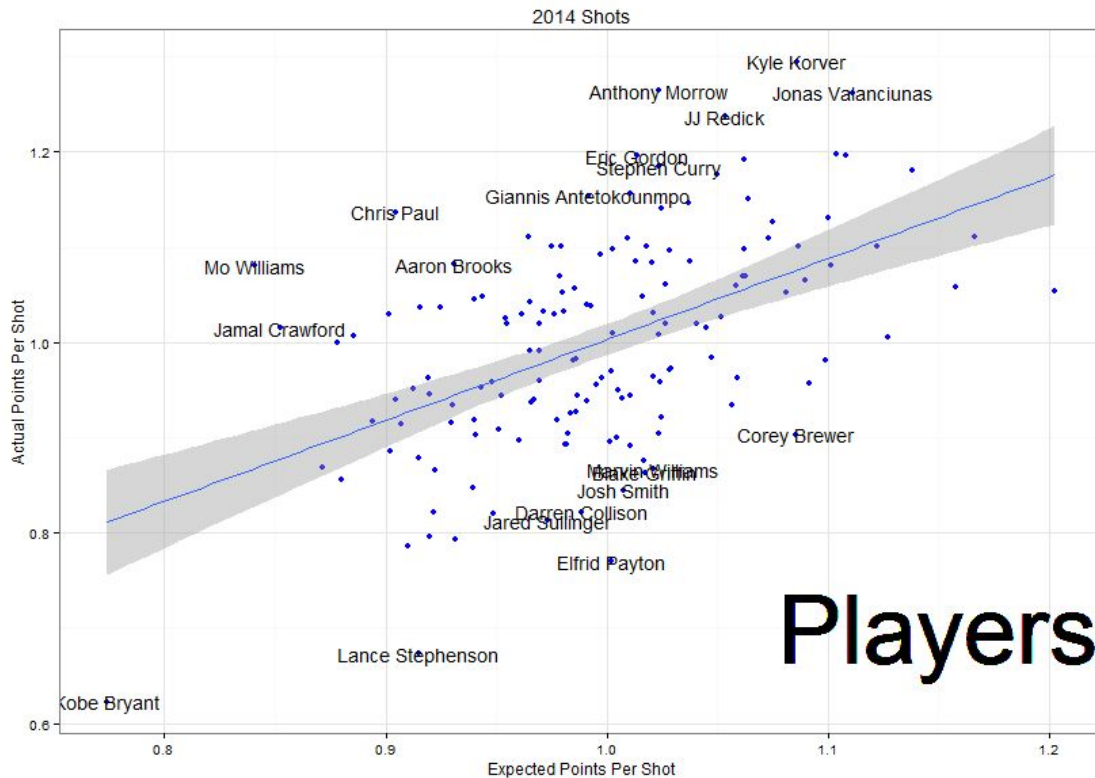
Team Evaluation



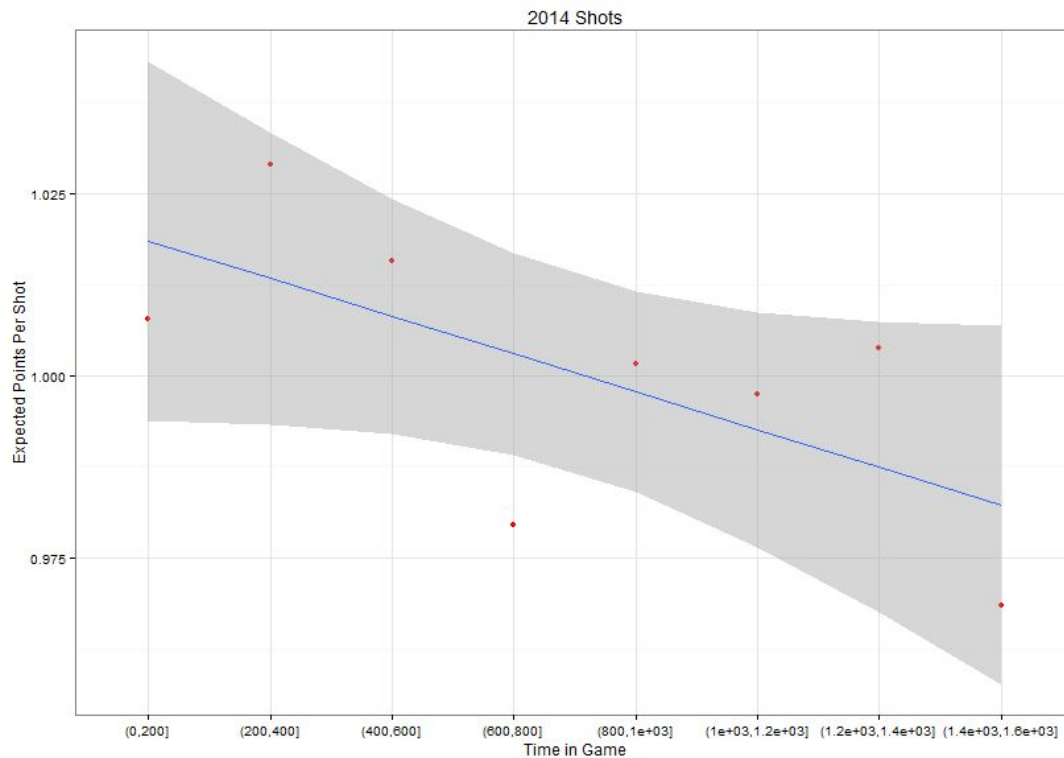
Player Evaluation



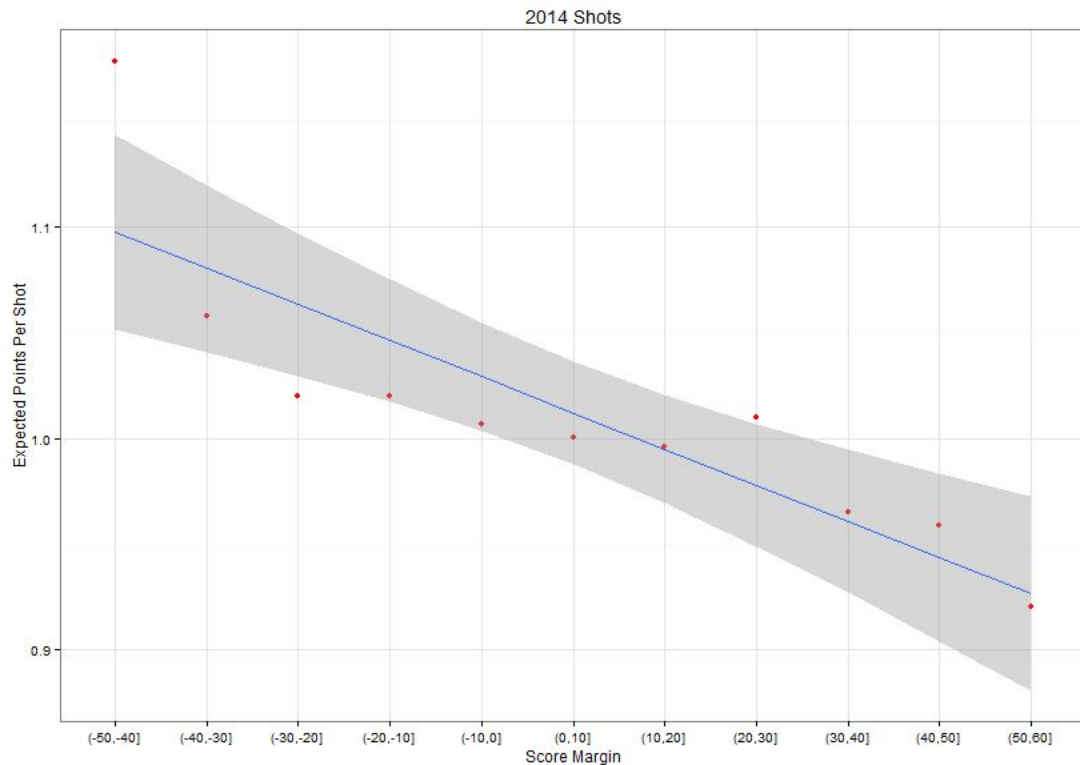
Player Evaluation



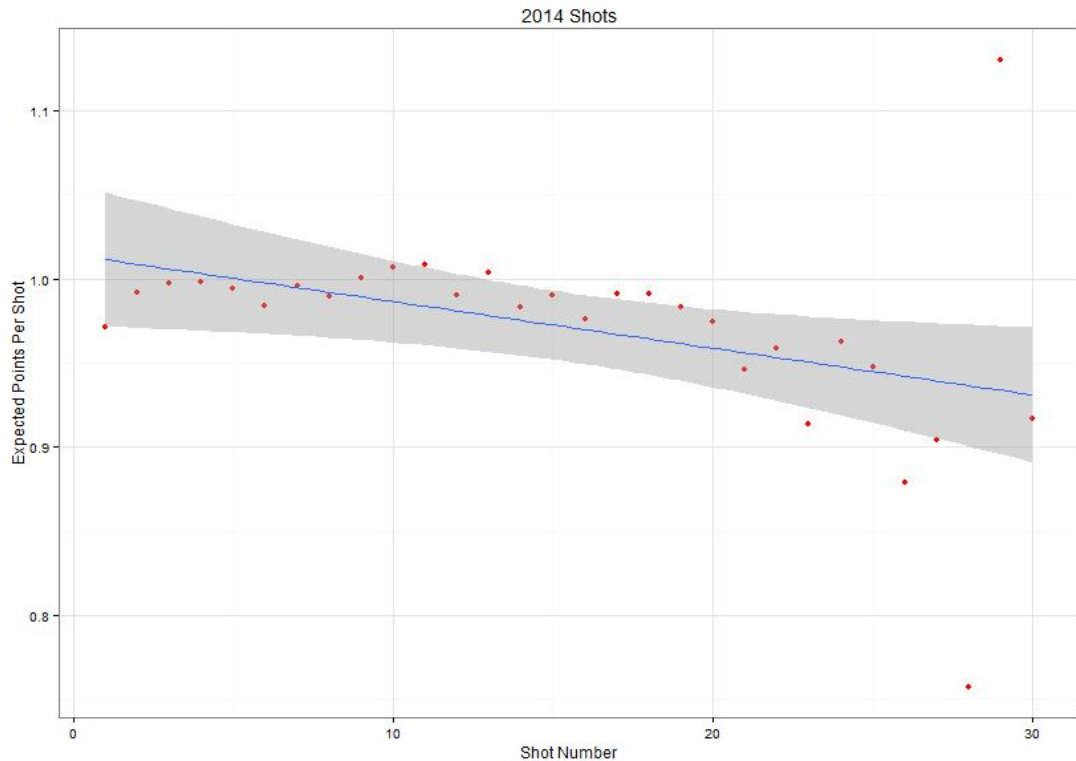
Coaching Strategy - Time in Game



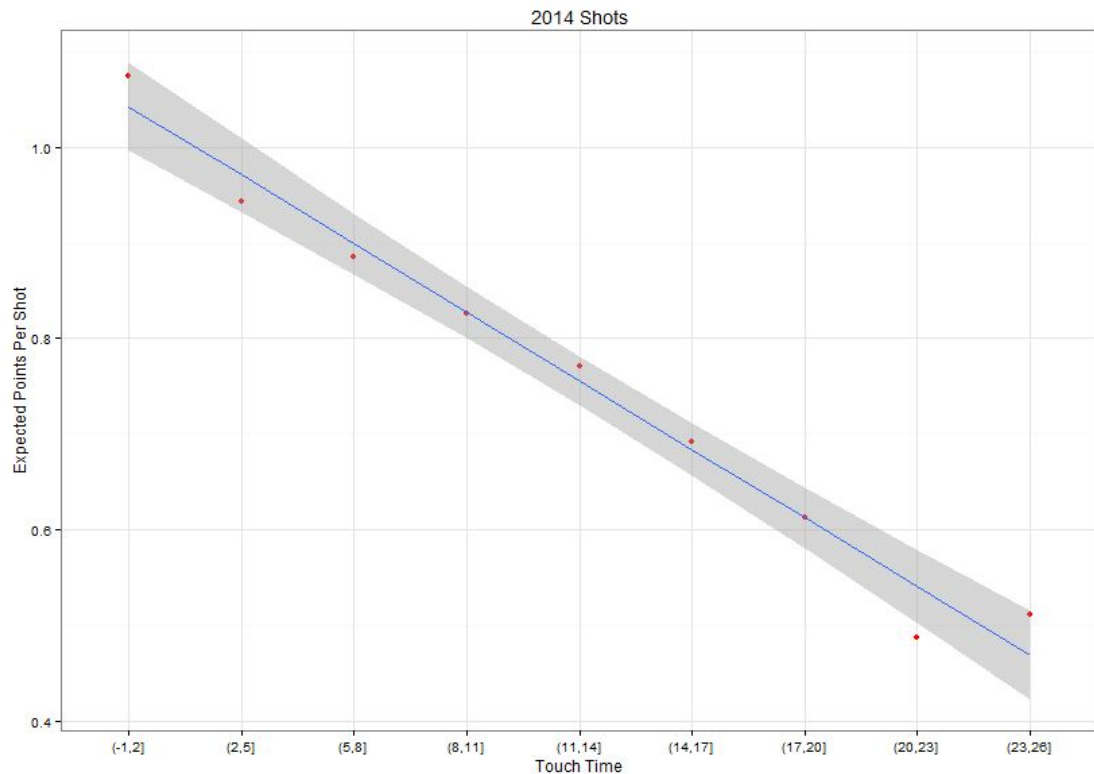
Coaching Strategy - Score Margin



Coaching Strategy - Shot Number

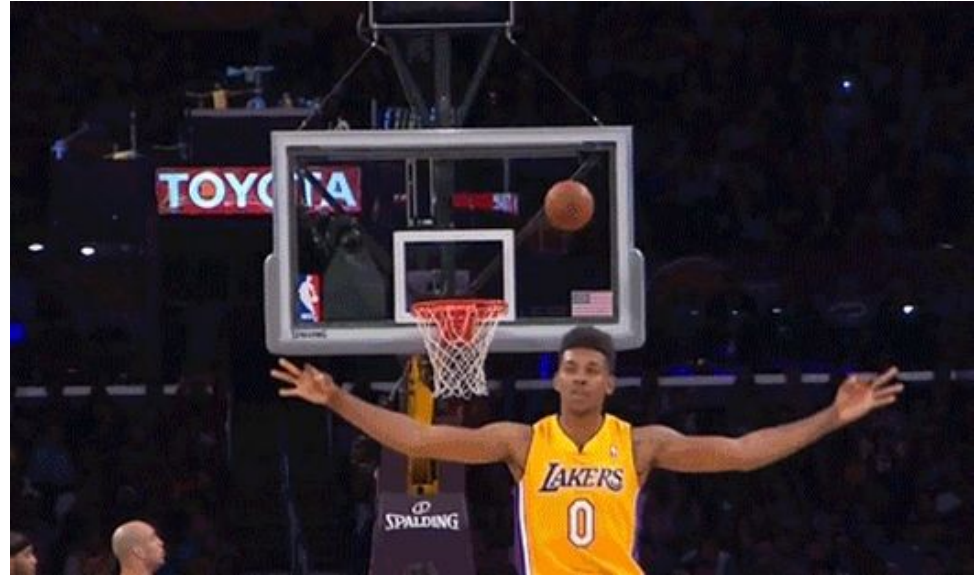


Coaching Strategy - Touch Time



Model Limitations

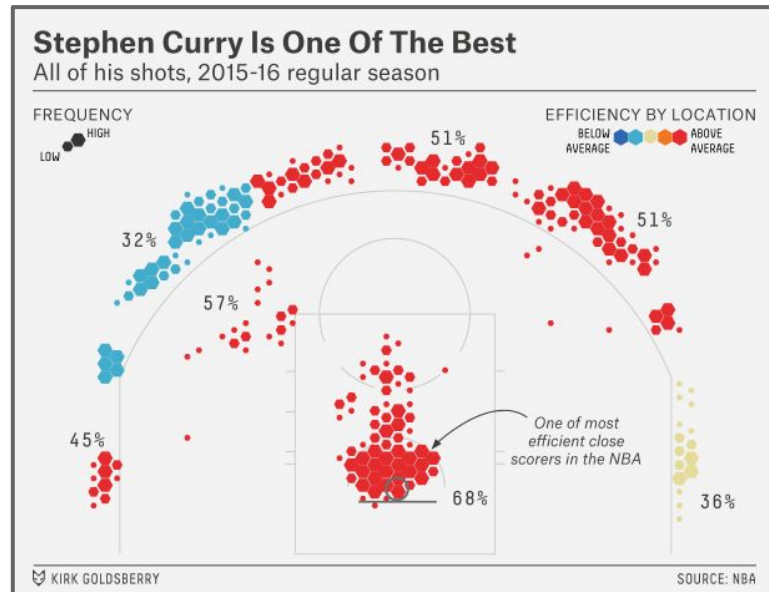
- Not super useful for predicting individual shots
- Correlation vs causation
- Foul shots
- Difficulty in merging data
- Bad data from nba.com
 - Touch time
 - Shot clock
- Missing data (rookie players)



Source: www.totalprosports.com

Possible Future Work

- Interactive model (i.e. Shiny app)
- Different models for 2's and 3's
- Adding in more predictors
 - Schedule variable
 - Injuries
- Incorporating data from previous years
- Real-time predictions
 - Run for 2015-2016 season to see year over year correlations
 - In-game applications



Source: www.fivethirtyeight.com

Questions?