Analysis of the Mid-Range Jumper in the Big 12

Jacob Kauffman

Rice University

THE MID-RANGE JUMPER:

THE LEAST EFFICIENT SHOT IN BASKETBALL?

OR

A CONFERENCE CHAMPIONSHIP WINNER?

OVERVIEW

1. Data Cleaning Process
2. Data Visualization
3. Multi-variate Linear Regression
4. Machine Learning Predictions

DATA CLEANING

Current dataset:

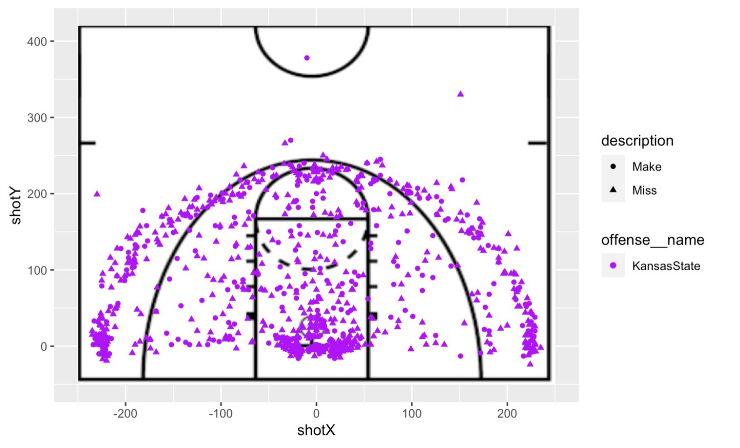
19,000 possessions

90 Big-12 games

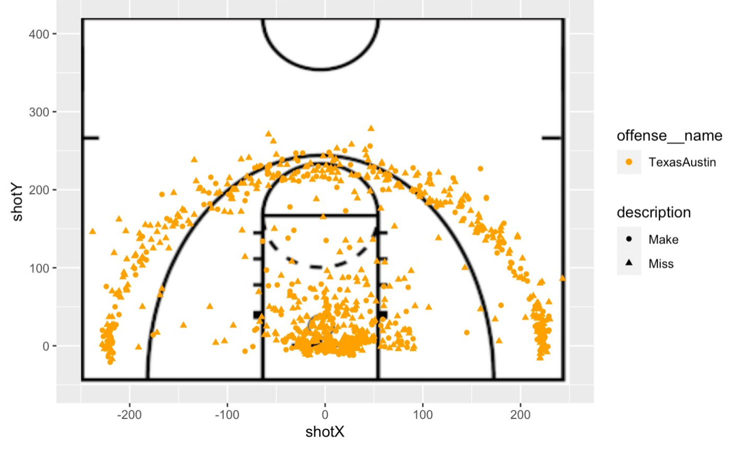
Want to filter to only a few columns of particular interest

Ultimately want to aggregate data to season level

SHOT VISUALIZATION COMPARISON

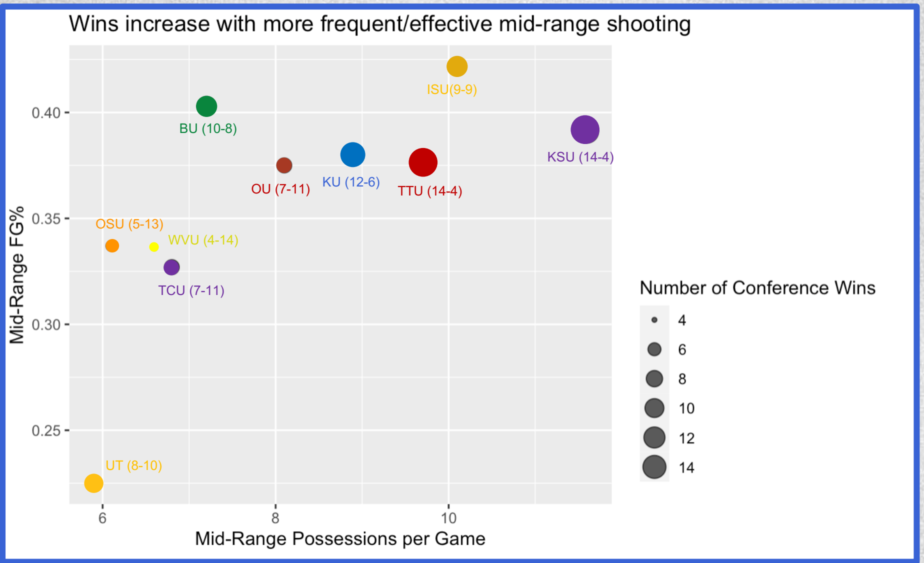


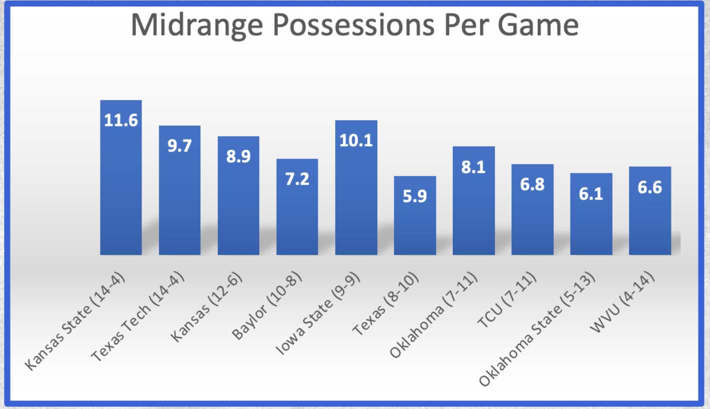
Kansas State (14-4)
Conference Champions



Texas (8-10)
Missed NCAA Tournament

WINNING TEAMS TAKE/MAKE MORE MID-RANGE SHOTS





MID-RANGE/FG% REGRESSION

Midrange and FG% Regression on Wins

	Dependent variable: Wins
Mid_rate	1.422***
	(0.301)
Mid_FG	36.580*
	(14.809)
Three_FG	32.425
	(60.147)
Rim_FG	16.180
	(42.375)
Constant	-12.165
	(17.321)
Observations	10
R2	0.895
Adjusted R2	0.811
Residual Std. Error	2.150 (df = 5)
Note:	*p<0.1; **p<0.05; ***p<0.01

*data used for regression on slides 11 and 12 in appendix

1. Interestingly, only two of the four regressors here are statistically significant:
2. These are mid-range attempt rate, and mid-range FG%
3. One point increase in Mid-Range FG% leads to more expected wins than the same increase in 3P FG% or Rim FG%

BAYLOR CASE STUDY

Specific Case Study:

Baylor:

2018-19 Conference Season:

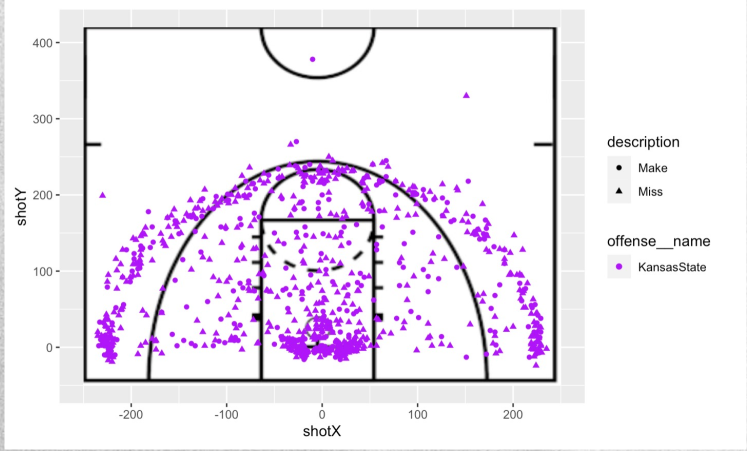
7.2 Midrange Possessions/Game (6th/10)

40.3% Midrange FG% (2nd/10)

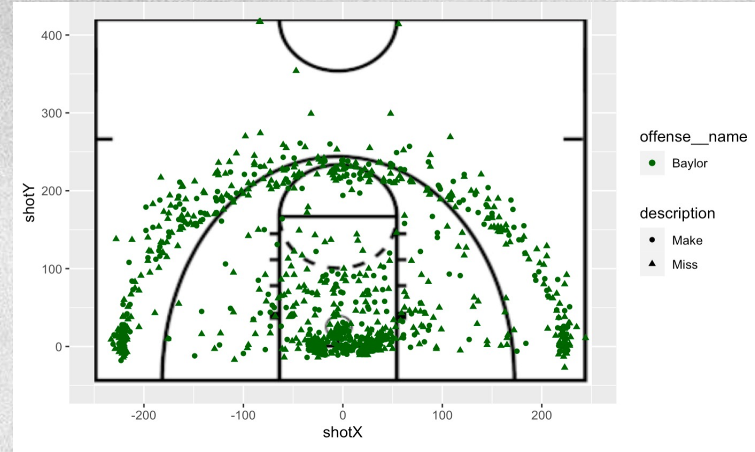
What happens if Baylor increase Midrange possessions by 75%?:

12.6 Midrange Possessions/Game (1st/10)

SHOT VISUALIZATION COMPARISON



Kansas State (14-4)
Conference Champions



Baylor (10-8)
Lost both matchups to KSU

MACHINE LEARNING TO PREDICT BAYLOR'S SEASON

1. • Filter data to include only Baylor's possessions
2. • Duplicate Baylor's data and double every mid-range possession while eliminating the same number of random non-midrange possessions
3. • Split both datasets into training and test groups, and add a shot outcome column variable for the training group
4. • Employ a Random Forest ML Model that predicts shot outcomes for both test groups
5. • Employ a Monte Carlo Simulation that randomly samples Baylor's season from both train sets 10,000 times, averaging PPG and wins.

BAYLOR CASE STUDY: MACHINE LEARNING OUTCOME

Specific Case Study:

Baylor:

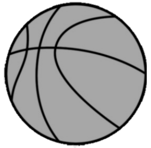
BEFORE:

- **Expected Record of 10.3-7.7**
- **72.6 expected PPG**

AFTER:

- **Expected Record of 11.4-6.6**
- **75.1 expected PPG**

PROJECT TAKEAWAYS



Clear correlation between a team's propensity for mid-range shots and their success in conference

A 1% jump in Mid-range FG% leads to more expected wins than a 1% jump in 3P FG% or Rim FG%

For Baylor, 5 more mid-range possessions per game leads to more than one additional expected win

APPENDIX

Team <chr>	Mid_rate <dbl>	Mid_FG <dbl>	Three_FG <dbl>	Rim_FG <dbl>	Wins <dbl>
Kansas State (14–4)	11.6	0.391	0.3344595	0.568	14
Texas Tech (14–4)	9.7	0.377	0.3618182	0.639	14
Kansas (12–6)	8.9	0.380	0.3421517	0.612	12
Baylor (10–8)	7.2	0.403	0.3455344	0.585	10
Iowa State (9–9)	10.1	0.422	0.3601236	0.632	9
Texas (8–10)	5.9	0.225	0.3356941	0.580	8
Oklahoma (7–11)	8.1	0.375	0.3436364	0.591	7
TCU (7–11)	6.8	0.327	0.3555195	0.616	7
Oklahoma State (5–13)	6.1	0.337	0.3725191	0.591	5
WVU (4–14)	6.6	0.336	0.3123994	0.563	4

Data used for regression in slide 7: Midrange attempt frequency, Midrange FG%, 3P FG%, Rim FG%