

# The Art of the Three

Jason Yang & Nicholas Latchana



# Abstract

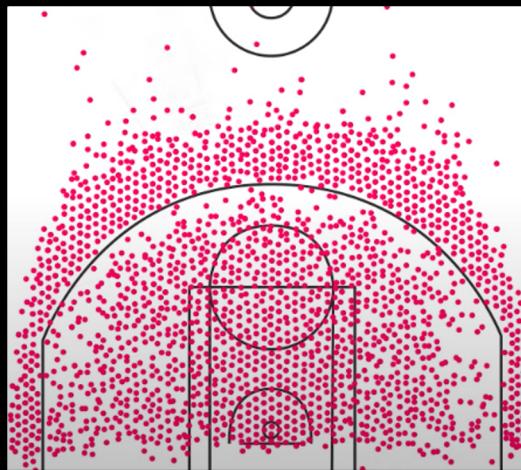
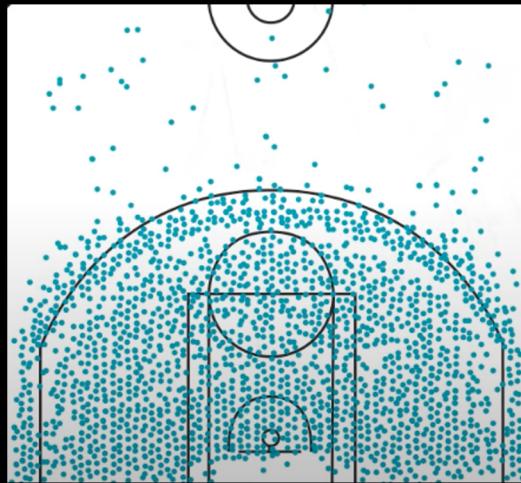
Application runs on NYU Peel Cluster:

- HDFS
- Hive
- Spark



# Why does this matter?

- Over the last decade, volume of 3P shots have drastically increased -> goal is examining how this shot selection emphasis is evolving the Game
- Directed towards:
  - Casual fans can observe results and understand how the game is changing over the years
  - Players and Coaches can devise improved gameplans and better understanding of their immediate team's 3P stats vs other teams in the league



# Determining Goodness

- Shots made/attempted were compared to the official stats from NBA
- Standings determined by our analytic were on par with the official standings
- In general, our statistical data completely aligned with the official stats and the datasets used have been pulled from reputable websites.
- This proved especially important for our play-by-play data, as a single missing line could have drastically changed our results.

# Data Sources

Source Name	Description	Size	Link
Sports-Statistics	Play-by-play data for the entirety of the 2008-09 season.	89.1 MB	<a href="https://sports-statistics.com/database/basketball-data/nba/2008-09_pbp.csv">https://sports-statistics.com/database/basketball-data/nba/2008-09_pbp.csv</a>
Sports-Statistics	Play-by-play data for the entirety of the 2015-16 season.	95.1 MB	<a href="https://sports-statistics.com/database/basketball-data/nba/2015-16_pbp.csv">https://sports-statistics.com/database/basketball-data/nba/2015-16_pbp.csv</a>
Basketball Reference	Individual player stats for 2008-09 season.	74 KB	<a href="https://www.basketball-reference.com/leagues/NBA_2009_totals.html">https://www.basketball-reference.com/leagues/NBA_2009_totals.html</a>
Basketball Reference	Individual player stats for 2015-16 season.	74 KB	<a href="https://www.basketball-reference.com/leagues/NBA_2016_totals.html">https://www.basketball-reference.com/leagues/NBA_2016_totals.html</a>

# Play-by-Play Stats

,EVENTMSGACTIONTYPE,EVENTMSGTYPE,EVENTNUM,GAME\_ID,HOMEDESCRIPTION,NEUTRALDESCRIPTION,PCTIMESTRING,PERIOD,PERSON1TYPE,PERSON2TYPE,PERSON3TYPE,PLAYER1\_ID,PLAYER1\_NAME,PLAYER1\_TEAM\_ABBREVIATION,PLAYER1\_TEAM\_CITY,PLAYER1\_TEAM\_ID,PLAYER1\_TEAM\_NICKNAME,PLAYER2\_ID,PLAYER2\_NAME,PLAYER2\_TEAM\_ABBREVIATION,PLAYER2\_TEAM\_CITY,PLAYER2\_TEAM\_ID,PLAYER2\_TEAM\_NICKNAME,PLAYER3\_ID,PLAYER3\_NAME,PLAYER3\_TEAM\_ABBREVIATION,PLAYER3\_TEAM\_CITY,PLAYER3\_TEAM\_ID,PLAYER3\_TEAM\_NICKNAME,SCORE,SCOREMARGIN,VISITORDESCRIPTION,WCTIMESTRING  
0,12,0,0020800377,,12:00,1,0.0,0,0,0,,,0,,,0,,,7:13 PM  
1,0,10,1,0020800377,Jump Ball Hibbert vs. Camby: Tip to Randolph,,12:00,1,4.0,5,5,201579,Roy Hibbert,IND,Indiana,1610612754.0,Pacers,948,Marcus Camby,LAC,Los Angeles,1610612746.0,Clippers,2216,Zach Randolph,LAC,Los Angeles,1610612746.0,Clippers,,,7:13 PM  
2,1,2,2,0020800377,,11:48,1,5.0,0,0,948,Marcus Camby,LAC,Los Angeles,1610612746.0,Clippers,0,,,0,,,0,,,MISS Camby 16' Jump Shot,7:14 PM  
3,0,4,3,0020800377,Rush REBOUND (Off:0 Def:1),,11:47,1,4.0,0,0,201575,Brandon Rush,IND,Indiana,1610612754.0,Pacers,0,,,0,,,0,,,7:14 PM  
4,1,2,4,0020800377,MISS Jack 13' Jump Shot,,11:36,1,4.0,0,0,101127,Jarrett Jack,IND,Indiana,1610612754.0,Pacers,0,,,0,,,0,,,7:14 PM  
5,0,4,5,0020800377,,11:35,1,5.0,0,0,948,Marcus Camby,LAC,Los Angeles,1610612746.0,Clippers,0,,,0,,,0,,,Camby REBOUND (Off:0 Def:1),7:14 PM  
6,4,6,6,0020800377,,11:31,1,5.0,4,1,1884,Baron Davis,LAC,Los Angeles,1610612746.0,Clippers,1902,Jeff Foster,IND,Indiana,1610612754.0,Pacers,0,,,0,,,Davis OFF.Foul (P1),7:14 PM  
7,5,5,7,0020800377,,11:31,1,5.0,0,1,1884,Baron Davis,LAC,Los Angeles,1610612746.0,Clippers,0,,,0,,,0,,,0,,,Davis Foul Turnover (P1.T1),7:14 PM  
8,42,2,8,0020800377,MISS Jack 1' Driving Layup,,11:22,1,4.0,0,0,101127,Jarrett Jack,IND,Indiana,1610612754.0,Pacers,0,,,0,,,0,,,7:14 PM  
9,0,4,9,0020800377,,11:21,1,5.0,0,0,2216,Zach Randolph,LAC,Los Angeles,1610612746.0,Clippers,0,,,0,,,0,,,Randolph REBOUND (Off:0 Def:1),7:14 PM  
10,1,2,10,0020800377,,11:10,1,5.0,0,0,2216,Zach Randolph,LAC,Los Angeles,1610612746.0,Clippers,0,,,0,,,0,,,MISS Randolph 26' 3PT Jump Shot,7:15 PM  
11,0,4,11,0020800377,Graham REBOUND (Off:0 Def:1),,11:09,1,4.0,0,0,101211,Stephen Graham,IND,Indiana,1610612754.0,Pacers,0,,,0,,,0,,,0,,,7:15 PM  
12,79,1,12,0020800377,Graham 19' Pullup Jump Shot (2 PTS),,10:55,1,4.0,0,0,101211,Stephen Graham,IND,Indiana,1610612754.0,Pacers,0,,,0,,,0,,,0 - 2,2,,7:15 PM  
13,1,1,13,0020800377,,10:36,1,5.0,5,0,1884,Baron Davis,LAC,Los Angeles,1610612746.0,Clippers,948,Marcus Camby,LAC,Los Angeles,1610612746.0,Clippers,0,,,2 - 2,TIE,Davis 20' Jump Shot (2 PTS) (Camby 1 AST),7:15 PM

# Individual Player Stats

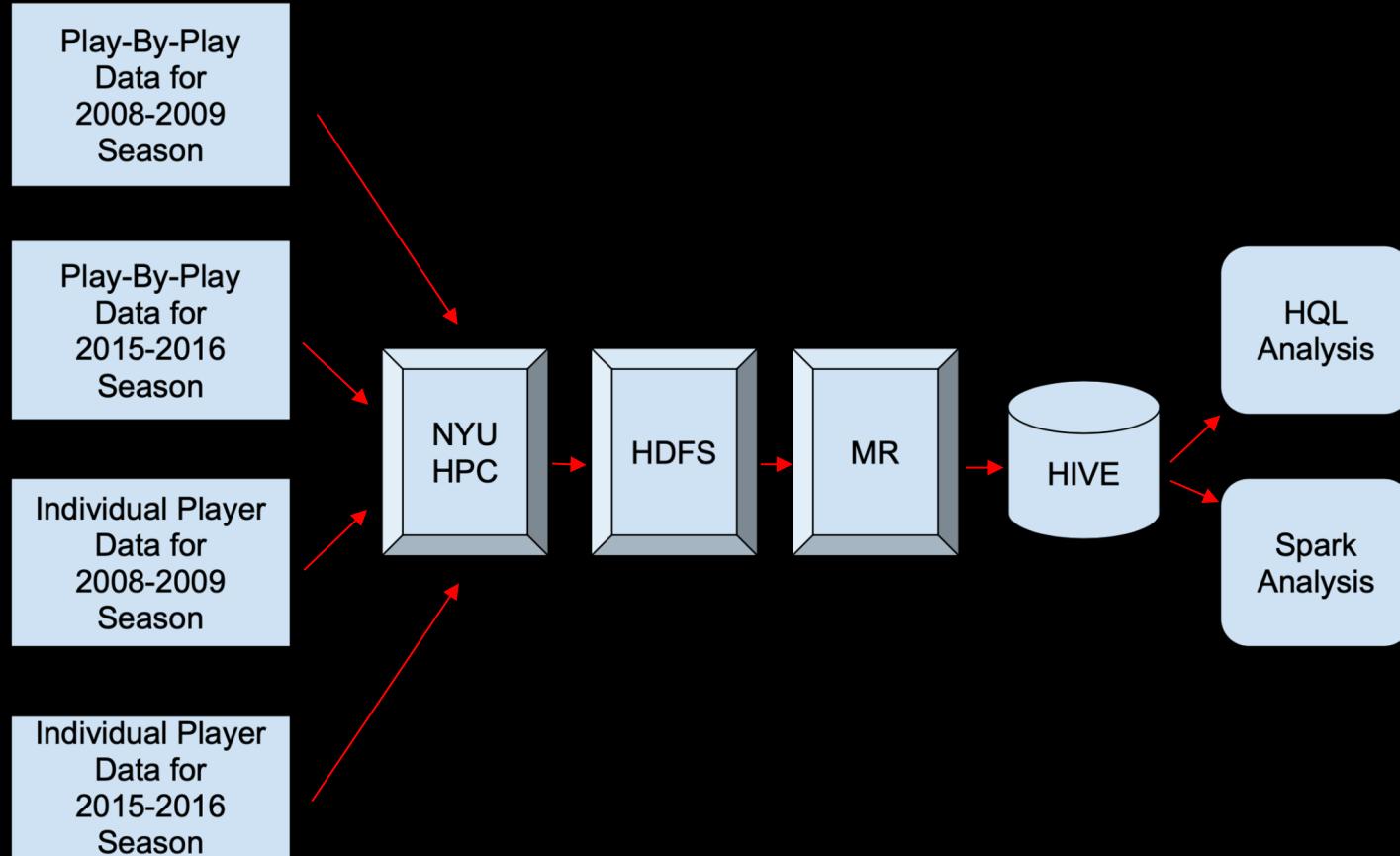
2008-  
2009

Rk	Player	Pos	Age	Tm	G	GS	MP	FG	FGA	FG%	3P	3PA	3P%	2P	2PA	2P%	eFG%	FT	FTA	FT%	ORB	DRB	TRB	AST	STL	BLK	TOV	PF	PTS
1	Alex Acker\ackeral01	SG	26	TOT	25	0	199	30	76	.395	7	20	.350	23	56	.411	.441	5	10	.500	8	16	24	12	5	4	7	9	72
1	Alex Acker\ackeral01	SG	26	DET	7	0	20	4	11	.364	0	4	.000	4	7	.571	.364	1	2	.500	0	2	2	1	2	1	0	0	9
1	Alex Acker\ackeral01	SG	26	LAC	18	0	179	26	65	.400	7	16	.438	19	49	.388	.454	4	8	.500	8	14	22	11	3	3	7	9	63
2	Hassan Adams\adamsha01	SG	24	TOR	12	0	52	4	13	.308	0	0	.000	4	13	.308	.308	3	6	.500	1	6	7	1	1	1	4	4	11
3	Arron Afflalo\afflaar01	SG	23	DET	74	8	1234	131	300	.437	43	107	.402	88	193	.456	.508	58	71	.817	30	104	134	44	28	13	43	144	363
4	Maurice Ager\agerma01	SG	24	NJN	20	0	97	15	43	.349	0	6	.000	15	37	.405	.349	4	8	.500	3	7	10	3	1	2	4	17	34
5	Blake Ahearn\ahearbl01	PG	24	SAS	3	0	19	2	6	.333	2	4	.500	0	2	.000	.500	2	2	1.000	0	1	1	2	1	0	1	3	8
6	Alexis Ajinça\ajincal01	PF	20	CHA	31	4	182	25	69	.362	0	2	.000	25	67	.373	.362	20	28	.714	9	21	30	3	7	6	12	34	70
7	LaMarcus Aldridge\aldrila01	PF	23	POR	81	81	3004	601	1243	.484	7	28	.250	594	1215	.489	.486	261	334	.781	234	371	605	157	77	77	121	207	1470
8	Joe Alexander\alexajo01	SF	22	MIL	59	0	716	102	245	.416	16	46	.348	86	199	.432	.449	58	83	.699	44	71	115	42	15	29	52	107	278
9	Malik Allen\allenma01	PF	30	MIL	49	3	579	73	170	.429	0	1	.000	73	169	.432	.429	10	21	.476	35	68	103	35	7	12	20	89	156
10	Ray Allen*\allenra02	SG	33	BOS	79	79	2876	499	1040	.480	199	486	.409	300	554	.542	.575	237	249	.952	66	212	278	218	69	13	134	157	1434

2015-  
2016

Rk	Player	Pos	Age	Tm	G	GS	MP	FG	FGA	FG%	3P	3PA	3P%	2P	2PA	2P%	eFG%	FT	FTA	FT%	ORB	DRB	TRB	AST	STL	BLK	TOV	PF	PTS
1	Quincy Acy\acyqu01	PF	25	SAC	59	29	876	119	214	.556	19	49	.388	100	165	.606	.600	50	68	.735	65	123	188	27	29	24	27	103	307
2	Jordan Adams\adamsjo01	SG	21	MEM	2	0	15	2	6	.333	0	1	.000	2	5	.400	.333	3	5	.600	0	2	2	3	3	0	2	2	7
3	Steven Adams\adamsst01	C	22	OKC	80	80	2014	261	426	.613	0	0	.000	261	426	.613	.613	114	196	.582	219	314	533	62	42	89	84	223	636
4	Arron Afflalo\afflaar01	SG	30	NYK	71	57	2371	354	799	.443	91	238	.382	263	561	.469	.500	110	131	.840	23	243	266	144	25	10	82	142	909
5	Alexis Ajinça\ajincal01	C	27	NOP	59	17	861	150	315	.476	0	1	.000	150	314	.478	.476	52	62	.839	75	194	269	31	19	36	54	134	352
6	Cole Aldrich\aldrico01	C	27	LAC	60	5	800	134	225	.596	0	0	.000	134	225	.596	.596	60	84	.714	86	202	288	50	47	68	64	139	328
7	LaMarcus Aldridge\aldrila01	PF	30	SAS	74	74	2261	536	1045	.513	0	16	.000	536	1029	.521	.513	259	302	.858	176	456	632	110	38	81	99	151	1331
8	Cliff Alexander\alexaci01	PF	20	POR	8	0	36	5	10	.500	0	0	.000	5	10	.500	.500	0	0	.000	2	4	6	0	1	2	1	1	10
9	Lavoy Allen\allenla01	PF	26	IND	79	28	1599	191	370	.516	0	0	.000	191	370	.516	.516	46	73	.630	162	262	424	76	26	42	69	147	428
10	Tony Allen\allento01	SG	34	MEM	64	57	1620	215	469	.458	15	42	.357	200	427	.468	.474	90	138	.652	104	192	296	70	110	18	78	175	535

# Technology Flow



# Code Challenges

- play-by-play mapper designed to write 4 pieces of info to the reducer, however the size of this “array” was not always uniform, depending on the type of play. -> discovered when Reducer output didn't match up with data
- Format: (gameid), ( team, home play, away play, score)
- 2: home miss and attempts : (1234)(team1,miss,,)
- 3: away miss and attempts : (1234)(team2,,miss,)
- 4: either team makes : (1234)(team2,,miss,35-25)
- Solved by running multiple MR jobs and including conditional statements to reducer to prevent “out of bounds” errors.

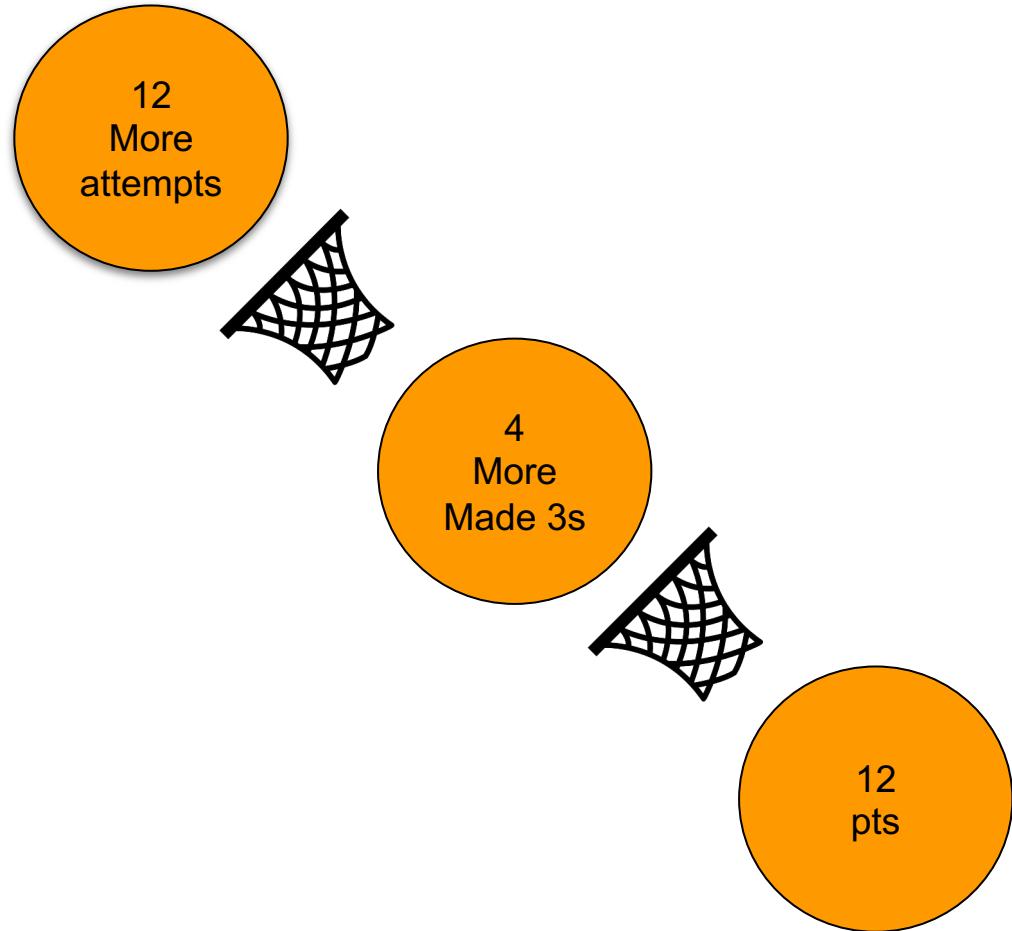
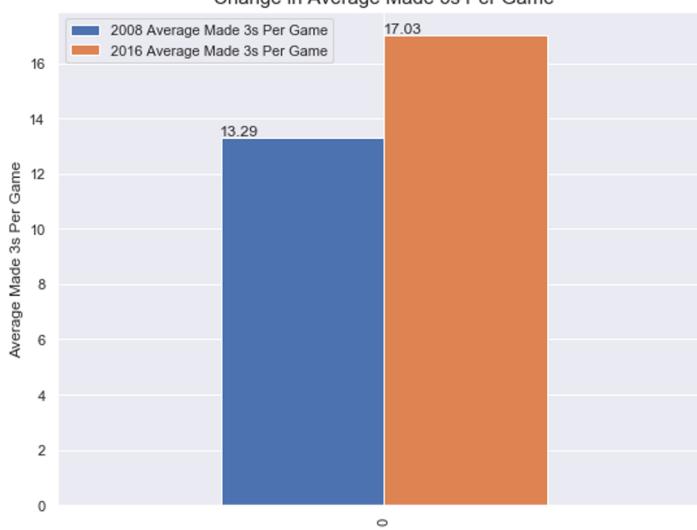
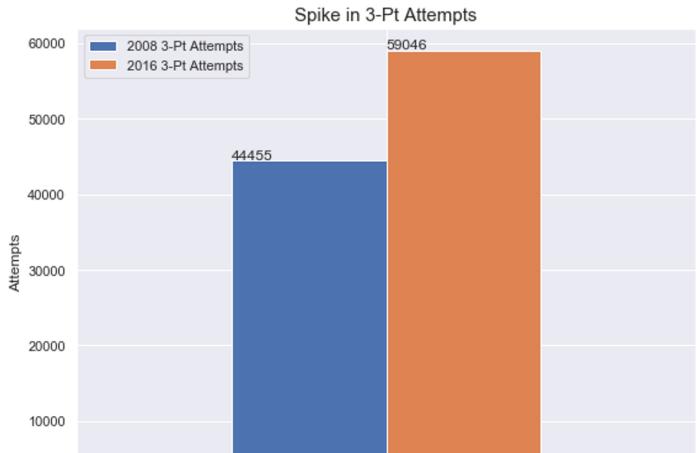
# Code Challenges

- Tracking score was difficult because it was not updated on every line of the play by play
- Discovered as we saw win totals varied from real data
  - Score does not get updated when a team misses a shot, which alters the length of the array
- Solved by using conditional statements to determine if the line contained a score and by constantly updating our value so that our final score would be as accurate as possible

# Obstacles

- How to combine our datasets?
  - The play-by-play data was far more specific than the general stats, so we decided to use the stats to supplement our findings
  - Example: looking at relationship between good three point shooters and their team's overall success
- How do we narrow down our data?
  - The play-by-play dataset featured over 30 columns, but given what we aimed to find, we were able to narrow it down to include just the teams, plays, and score
  - For the individual stats, we got rid of all statistics that did not pertain to the three-point shot, as these had no significance in our analytic

# Attempts Per Game



# League 3 Point Percentage

WHAT?!?!

2008: 38.6%



2016: 36.7%

- We'd expect 3PP to increase (i.e. players become more accurate)
- This is perfectly reasonable:
  - Volume of 3-pt shots increased substantially
  - More players at all positions are incorporating 3-pt shots

**33% 3PP**  $\equiv$  **50% 2PP**



# Top Shooters

- Shooters shoot
- Volume
- Shot Selection



player2008.team	player2008.playername	player2008.threep	player2008.threepa	player2008.threapp	player2008.points	proportion
CHA	D.J. Augustin	108	246	0.439	850	0.3811764705882353
CLE	Mo Williams	183	420	0.436	1443	0.3804573804573805
SAS	Roger Mason	166	394	0.421	969	0.5139318885448917
SAC	Kevin Martin	115	277	0.415	1254	0.2751196172248804
CHI	Ben Gordon	173	422	0.41	1699	0.305473808122425

player2016.team	player2016.playername	player2016.threep	player2016.threepa	player2016.threapp	player2016.points	proportion
LAC	J.J. Redick	200	421	0.475	1226	0.4893964110929853
GSW	Stephen Curry	402	886	0.454	2375	0.5077894736842106
ORL	Evan Fournier	156	390	0.4	1213	0.38582028029678483
CLE	J.R. Smith	204	510	0.4	955	0.6408376963350786
ATL	Kyle Korver	158	397	0.398	739	0.6414073071718539

# Does More 3P Translate to More Wins?

2008



3PM: 542  
3PA: 1505  
3PP: 0.36  
Wins: 58

LA Lakers: 1 -> 29

2016



3PM: 626  
3PA: 1977  
3PP: 0.317  
Wins: 22



3PM: 321  
3PA: 933  
3PP: 0.344  
Wins: 25

OKC Thunder: 30 -> 5



3PM: 678  
3PA: 1941  
3PP: 0.349  
Wins: 56

# Summary

- Changing the game of basketball?
- Is it becoming a game of who can make more 3s?
- 3P shot utilization has drastically increased
- Diversified the Game
- Success in basketball, like all sports, is not 1-Dimensional



# Acknowledgements

- THANK YOU PROF. MALAVET
- Thank you to HPC for quick and helpful responses to our questions!
- Thank you to Basketball Reference and Sports Statistics for providing the datasets we used!

# References



- Play by Play Data from  
<https://sports-statistics.com>
- Player Stats from  
<https://www.basketball-reference.com>
- <https://link.springer.com/content/pdf/10.1007/s12122-014-9193-5.pdf> (Paper regarding shot selection in NBA)