NBA Lineup Analysis

Jacob Kauffman

Project Three

SMGT 431 – Advanced Sport Analytics

Overview

1. Introduction
2. Exploratory Data Analysis
3. Methodology
4. Results and Discussion
5. Conclusion

INTRODUCTION

Why Lineup Analysis?

- In sports, team performance is rarely the sum of the strength of individual players
- Chemistry and Interaction between players often just as important as talent, lineup analysis is a great way to assess this
- NBA's top performing lineups don't always contain the best individual players

Mathletics' Take on Lineup Analysis

- Discussed in detail in Chapter 32:
- In a typical season, a team plays over 500 lineups
- Is there a rhyme or reason to coaches lineup decisions?
- Do coaches play good lineups as often as they should? Do they play bad lineups too much?
- Warriors Death Lineup: Outscored opponents by 29 points per 48

Specific Question to be answered

In the 2016-2017 season, the Atlanta Hawks most played lineup was:

PG:	SG:	SF:	PF:	C:
Dennis Schroder	Kent Bazemore	Thabo Sefalosha	Paul Millsap	Dwight Howard

Though the 2016-17 Hawks were a pretty decent team, this lineup was not particularly good, as they had a Per-48 +/- of -2.4. However by simply subbing in Tim Hardaway Jr. for Kent Bazemore at SG, the lineup increases this number to 28.0. Mathletics asks if the Hawks made a poor decision in playing this lineup ¼ the amount of the Bazemore lineup.

Exploratory Data Analysis

Dataset Background

- Data obtained from Cleaningtheglass.com, a fan-maintained public source for NBA data
- Contains data from all NBA lineups who played more than 100 possessions in the 2016-17 season (350 lineups)

Team	PG	SG	SF	PF	C	Poss	Diff	
League Averages						--		--
ATL	T.Youn	D.Murr	D.Hunt	J.Coll	C.Cape	612	64	+9.5
MIN	D.Russ	A.Edwa	J.McDa	K.Town	R.Gobe	541	55	+5.4
GSW	S.Curr	K.Thom	A.Wigg	D.Gree	K.Loon	493	95	+26.4
DET	J.Ivey	C.Cunn	S.Bey	B.Bogd	I.Stew	458	25	-6.3
PHX	C.Payn	D.Book	M.Brid	T.Crai	D.Ayto	399	34	-3.8
SAC	D.Fox	K.Huer	H.Barn	K.Murr	D.Sabo	393	75	+13.7
CHI	A.Dosu	Z.LaVi	D.DeRo	P.Will	N.Vuce	378	38	-2.8
DEN	J.Murr	K.Cald	M.Port	A.Gord	N.Joki	372	77	+14.7

Cleaning Dataset

Biggest obstacles with dataset:

- *Unlike Mathletics dataset, all data is on a possession basis and not a minutes/game basis*
- *Don't have aggregates for offensive points/defensive points, only PPP*
- *All the players names are listed in separate columns*

Columns that I created:

Lineup_Name <chr>	Total.Off.Points <dbl>	Total.Def.Points <dbl>	Lineup.Plus.Minus <dbl>	Minutes.Played <dbl>	Games.Played <dbl>	Above_Average <dbl>
	NA	*NA*	*NA*	*NA*	*NA*	*NA*
John Wall Bradley Beal Otto Porter Markieff Morris Marcin Gortat	3123.339	2881.603	241.736	1355.18667	28.233056	8.5621622
Chris Paul JJ Redick Luc Mbah a Moute Blake Griffin DeAndre Jordan	2018.968	1767.248	251.720	856.42667	17.842222	14.1081081
Ricky Rubio Zach LaVine Andrew Wiggins Karl-Anthony Towns Gorgui Dieng	1889.316	1916.772	-27.456	846.56000	17.636667	-1.5567568
Kemba Walker Nicolas Batum Michael Kidd-Gilchrist Marvin Williams Cody Zeller	1694.520	1627.053	67.467	774.04000	16.125833	4.1837838
Russell Westbrook Victor Oladipo Andre Roberson Domantas Sabonis Steven Adams	1316.990	1271.270	45.720	626.53333	13.052778	3.5027027
Jeff Teague Monta Ellis Paul George Thaddeus Young Myles Turner	1316.520	1346.328	-29.808	612.72000	12.765000	-2.3351351
Ricky Rubio Brandon Rush Andrew Wiggins Karl-Anthony Towns Gorgui Dieng	1319.417	1291.169	28.248	580.65333	12.096944	2.3351351
Stephen Curry Klay Thompson Kevin Durant Draymond Green Zaza Pachulia	1379.256	1111.826	267.430	561.41333	11.696111	22.8648649
Patrick Beverley James Harden Trevor Ariza Ryan Anderson Clint Capela	1304.160	1116.060	188.100	515.53333	10.740278	17.5135135

Exploratory Analysis

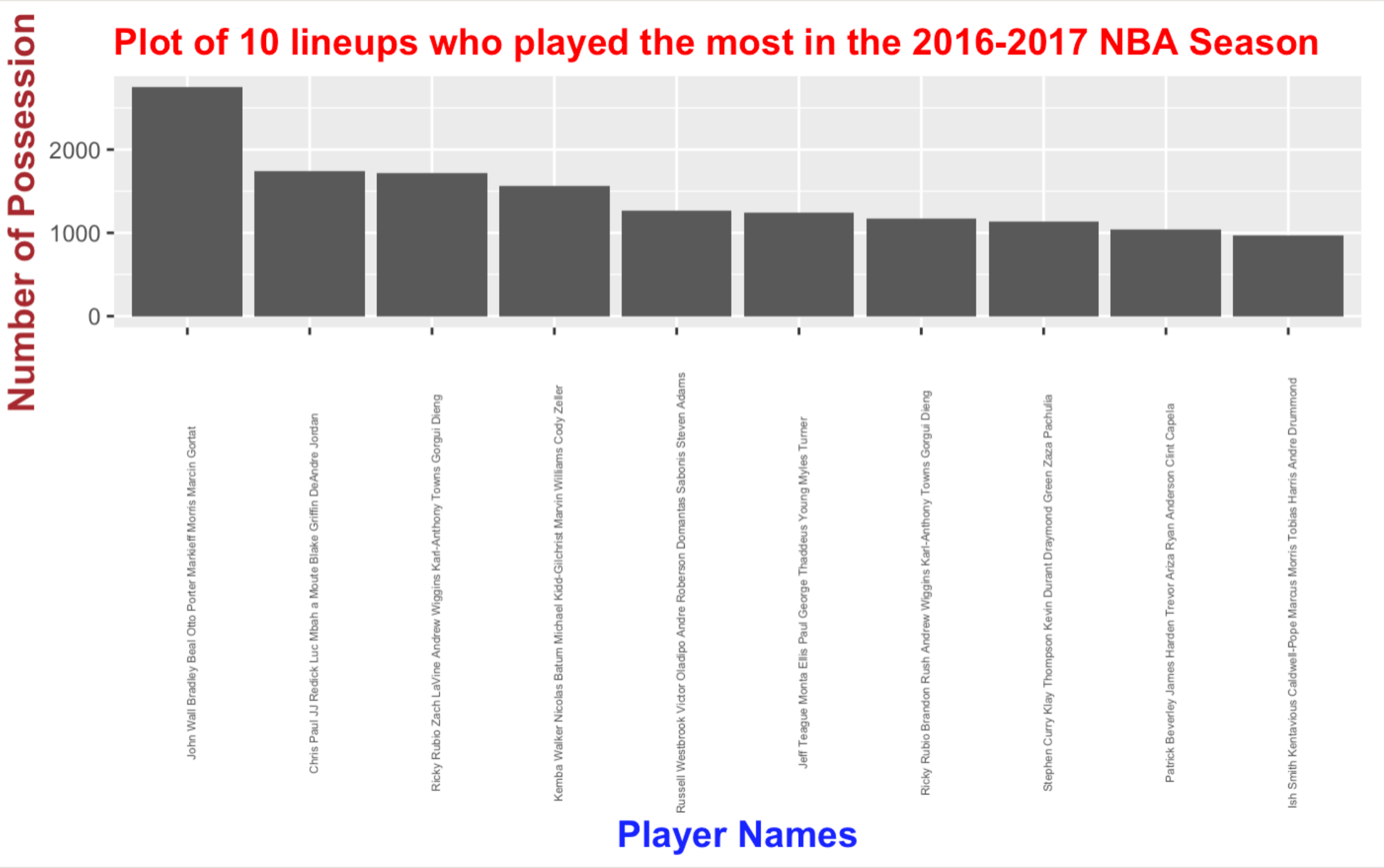
Notable Column Means:

Points Per 48 +/-: +3.21

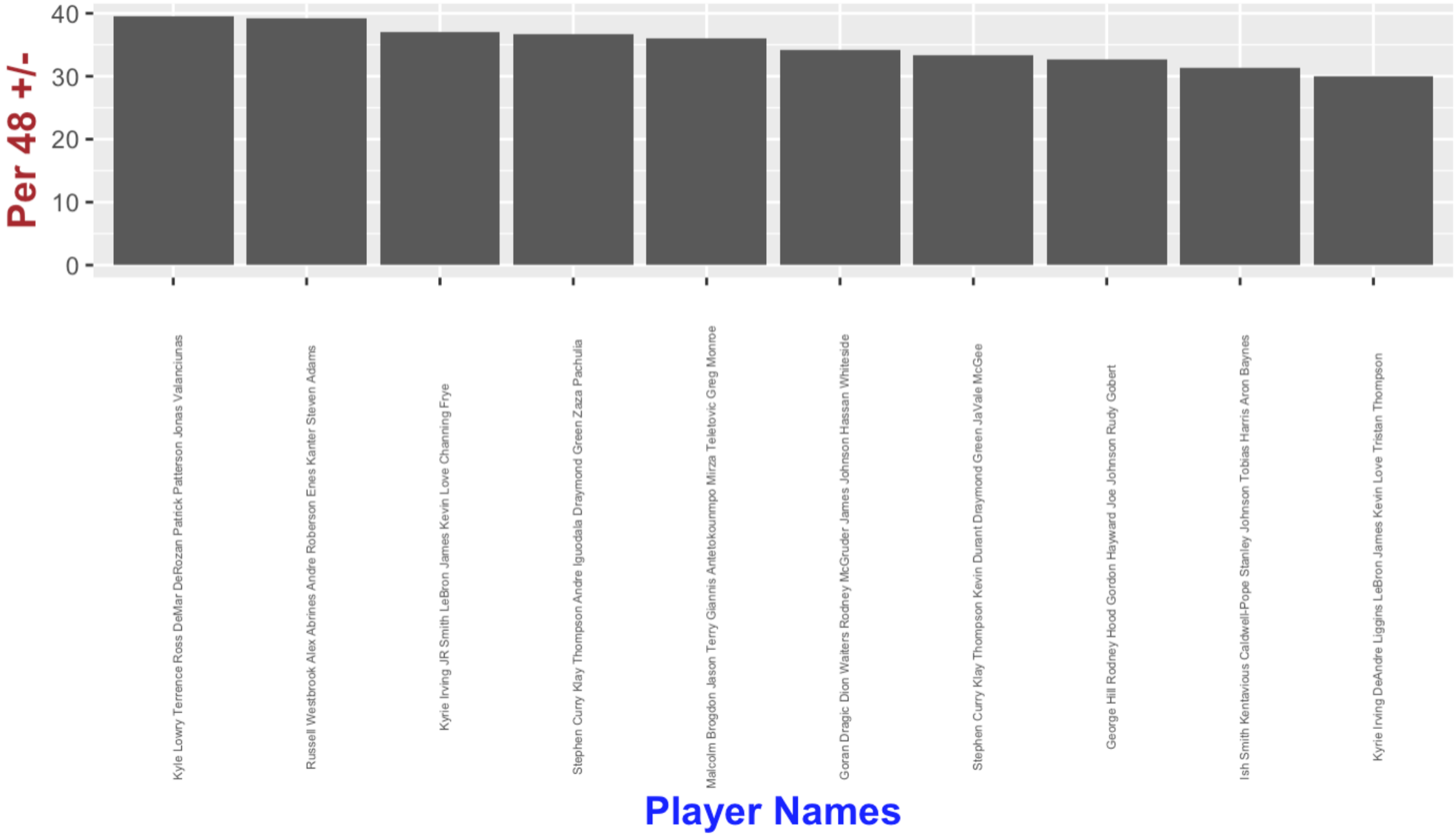
Possessions by lineup: 259.18

Games Played by lineup: 2.67

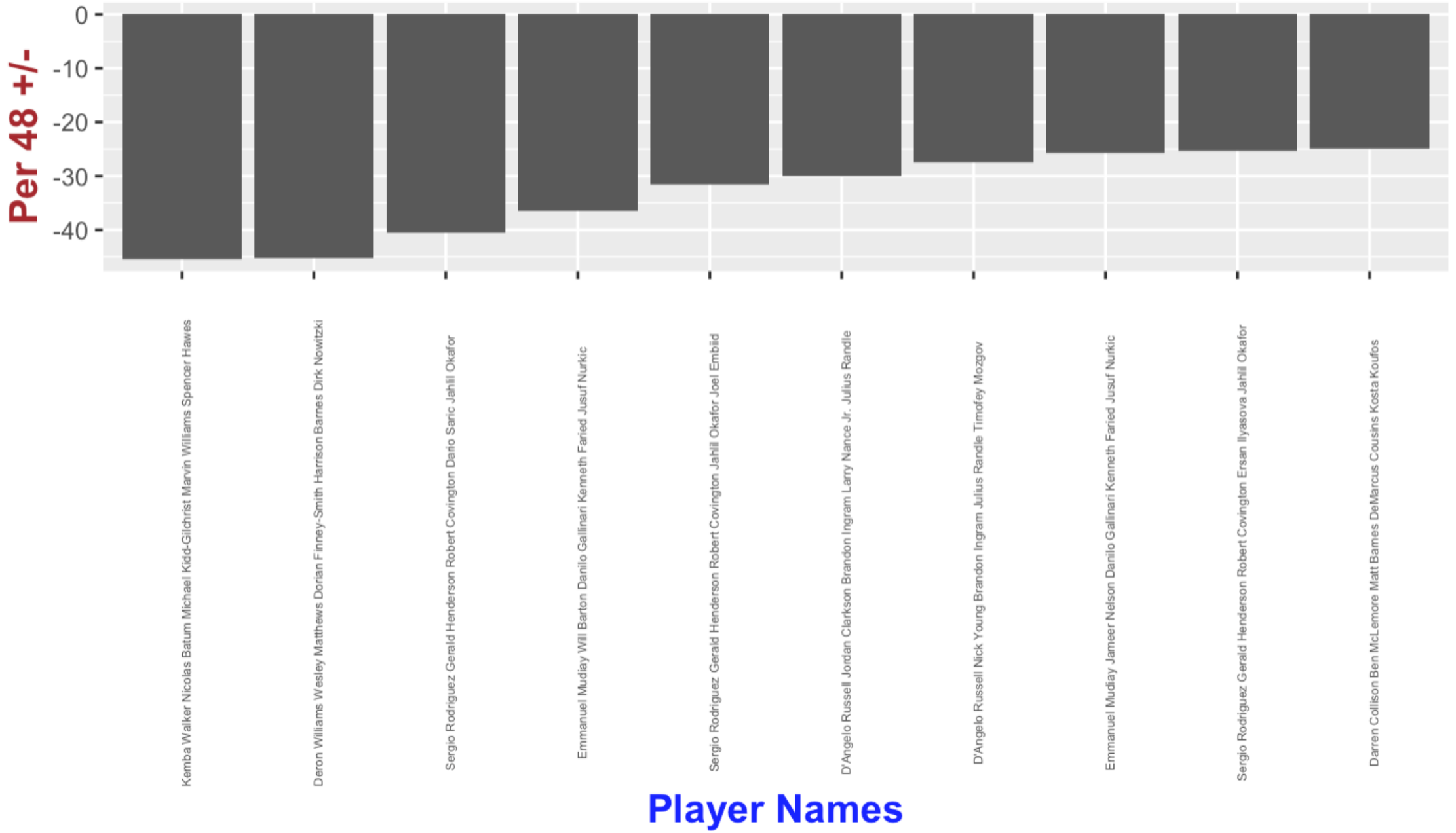
I also wanted to create visualizations that would help me understand which lineups are playing the most and are most/least effective

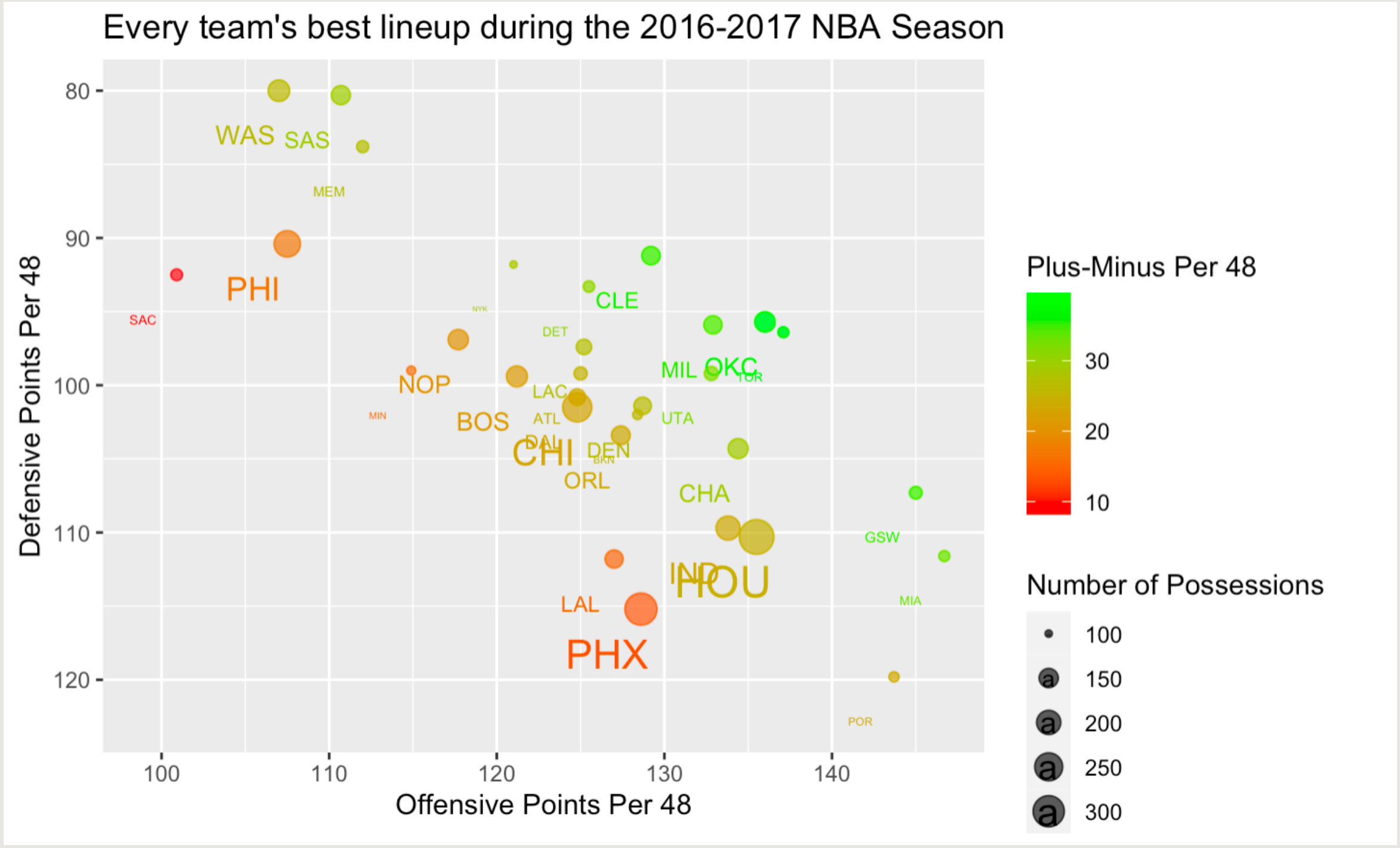

Plot of 10 lineups who played the most in the 2016-2017 NBA Season
Number of Possession
2000
1000
0
John Wall Bradley Beal Otto Porter Markieff Morris Marcin Gortat
Chris Paul JJ Redick Luc Mbah a Moute Blake Griffin DeAndre Jordan
Ricky Rubio Zach LaVine Andrew Wiggins Karl-Anthony Towns Gorgui Dieng
Kemba Walker Nicolas Batum Michael Kidd-Gilchrist Marvin Williams Cody Zeller
Russell Westbrook Victor Oladipo Andre Roberson Domantas Sabonis Steven Adams
Jeff Teague Monta Ellis Paul George Thaddeus Young Myles Turner
Ricky Rubio Brandon Rush Andrew Wiggins Karl-Anthony Towns Gorgui Dieng
Stephen Curry Klay Thompson Kevin Durant Draymond Green Zaza Pachulia
Patrick Beverley James Harden Trevor Ariza Ryan Anderson Clint Capela
Ish Smith Kentavious Caldwell-Pope Marcus Morris Tobias Harris Andre Drummond
Player Names

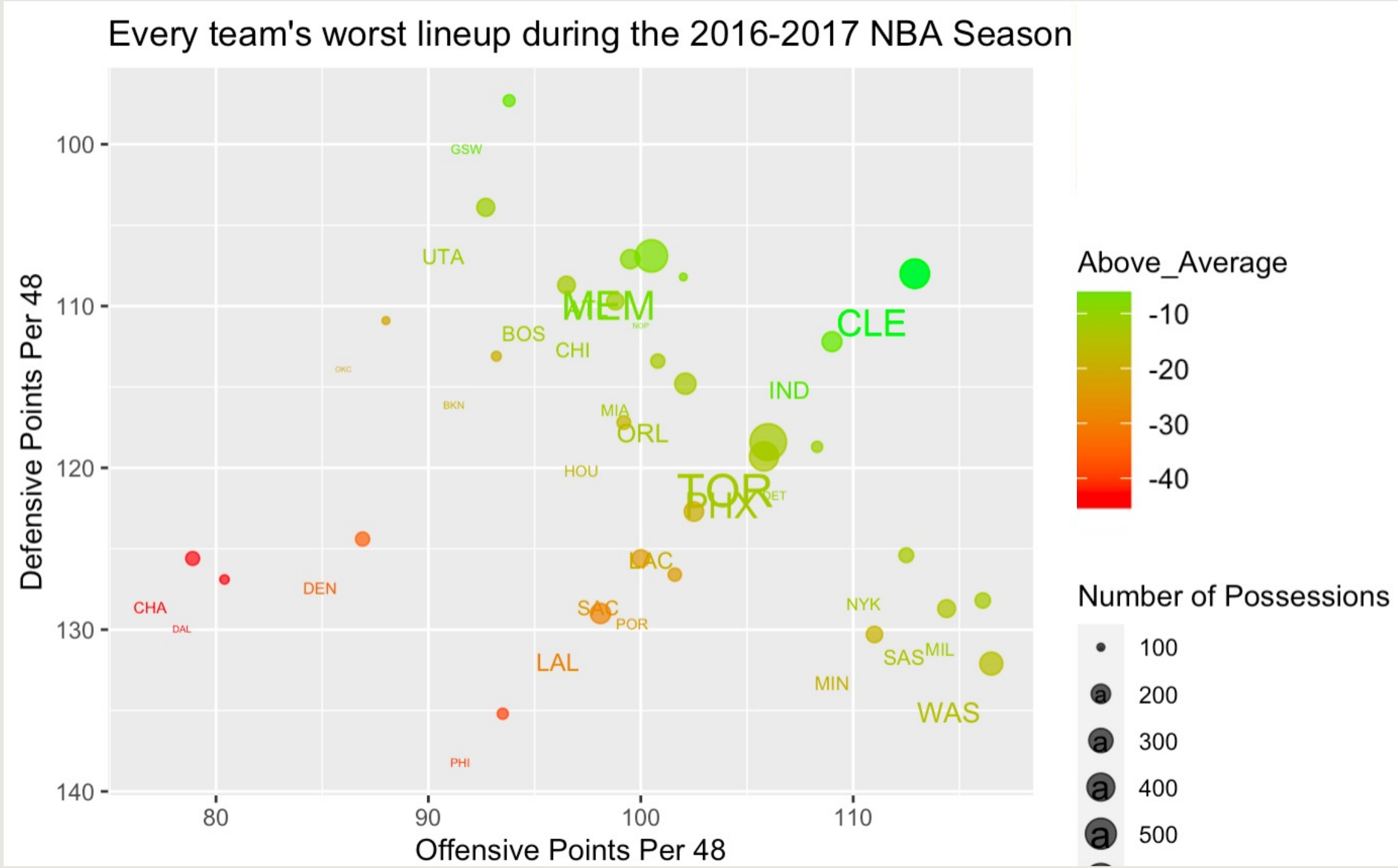
Plot of 10 best lineups Per 48 in the 2016-2017 NBA Season



Plot of 10 worst lineups Per 48 in the 2016-2017 NBA Season




Every team's best lineup during the 2016-2017 NBA Season
Defensive Points Per 48
Offensive Points Per 48
80
90
100
110
120
100
110
120
130
140
WAS
SAS
MEM
PHI
SAC
NYK
DET
CLE
NOP
MIN
BOS
LAC
ATL
MIL
OKC
TOR
UTA
DAL
CHI
DEN
BKN
ORL
CHA
GSW
IND
HOU
LAL
MIA
PHX
POR
Plus-Minus Per 48
30
20
10
Number of Possessions
100
150
200
250
300


Every team's worst lineup during the 2016-2017 NBA Season
Defensive Points Per 48
Offensive Points Per 48
100
110
120
130
140
80
90
100
110
GSW
UTA
MEM
NOP
BOS
CHI
OKC
BKN
MIA
ORL
IND
CLE
HOU
TOR
PHX
DET
LAC
DEN
CHA
DAL
SAC
POR
LAL
NYK
SAS
MIL
MIN
WAS
PHI
Above_Average
-10
-20
-30
-40
Number of Possessions
100
200
300
400
500

Major Takeaways from Visualizations

Specific:

- The best lineup in the NBA wasn't on a top 5 NBA team, and it lacked any top 20 players in the league
- Russell Westbrook's ability to make a lineup of 4 below average players the 2nd best in the league is incredible
- "The Process" 76ers were horrendous (3 of 9 worst lineups)
- Scott Brooks (Wizards Coach) REALLY dislikes his bench

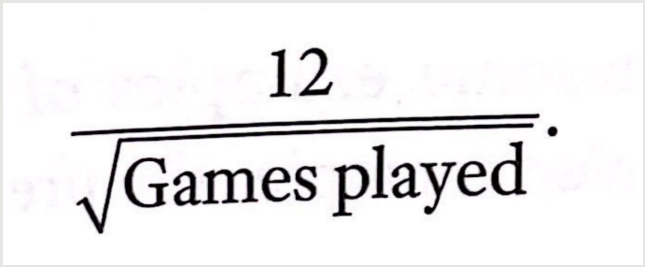
General:

- The best lineups in the NBA and on a team usually don't play the highest number of minutes
- The best lineups in the NBA weren't nearly as good as the bad lineups were bad
- Teams often play their worst lineups 3x more than their best

Methodology

Steps taken to Replicate Analysis

1. Use the DPLYR Package to clean data and to make it closely resemble the dataset that Mathletics used for their analysis (the most important part of this is calculating points per 48 minutes)
2. According to Mathletics, indicates that the actual performance of a lineup over 48 minutes is normally distributed with a mean equal to its lineup rating and a standard deviation of 12 points. Thus, the formula to calculate standard deviation of a lineup's rating is:



Steps taken to Replicate Analysis

3. The variance of the difference of independent RV's is the sum of the random variable variances. Since the standard deviation of a RV is just the square root of the RV's variance, we know that the standard deviation in the rating of the two lineups is given by:

$$\sigma = \sqrt{\frac{144}{\text{Games Lineup 1 played}} + \frac{144}{\text{Games Lineup 2 Played}}}.$$

Steps taken to Replicate Analysis

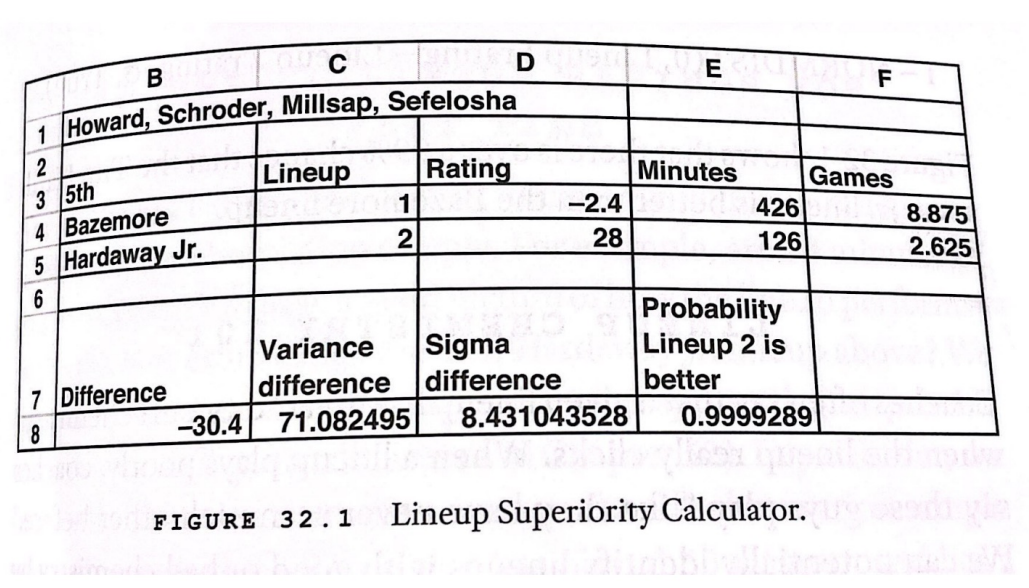
4. We then know that the probability that Lineup 2 is better equals the probability that a RV with mean Lineup 2 rating– Lineup 1 rating and standard deviation σ is greater than 0. I used the R to calculating the probability that:

$$1 - NORMDIST(0, \text{Lineup 1 rating} - \text{Lineup 2 rating}, \sigma, \text{True}).$$

Results and Discussion

My Analysis vs Mathletics

Mathletics Analysis



	B	C	D	E	F
1	Howard, Schroder, Millsap, Sefelosha				
2					
3	5th	Lineup	Rating	Minutes	Games
4	Bazemore	1	–2.4	426	8.875
5	Hardaway Jr.	2	28	126	2.625
6					
7	Difference	Variance difference	Sigma difference	Probability Lineup 2 is better	
8	–30.4	71.082495	8.431043528	0.9999289	

FIGURE 32.1 Lineup Superiority Calculator.

My Analysis

FifthMan <chr>	Lineup <dbl>	Diff <dbl>	Minutes.Played <dbl>	Games.Played <dbl>
Kent Bazemore	1	-4.7	429.2000	8.941667
Tim Hardaway Jr.	2	25.5	121.8533	2.538611

Difference <dbl>	Variance_Difference <dbl>	Sigma_Difference <dbl>	Probability_Lineup_2_is_better <dbl>
-30.2	72.82831	8.53395	0.9999108

Discussion of Results

- Results are essentially the same (0.00001 apart), any differences can be attributed to:
    1. When doing my data cleaning, I generalized all possessions to take the league average 14.8 seconds. In reality, Atlanta took 14.3 seconds per possession
    2. Mathletics dataset probably included playoffs, where the Hawks played 6 extra games
- According to both my analysis, and Mathletics, the probability that the Tim Hardaway Jr. lineup is better than the Kent Bazemore lineup is 99.991%.
- Thus, it is highly likely that Mike Budenholzer, the former coach of the Hawks, made a mistake in playing this lineup so often/starting them for most of the season.

Conclusion

Conclusion

Performance of specific lineups is crucial in determining team success, perhaps moreso than individual player performance

Teams can see large boosts in performance by simply swapping out one weak link in a lineup:

Specifically, in the case of Kent Bazemore and Tim Hardaway Jr, Mike Budenholzer made a curious decision to play a lineup featuring Bazemore about 4x as much as a lineup with Hardaway Jr., even though Hardaway's lineup was significantly more effective.