Josh Mark
ACC Basketball Clustering

Github link with R code, data, and Tableau workbook: https://github.com/joshdmark/ACC-Basketball-Clustering

OVERVIEW & OBJECTIVE

Analytics and statistics are at the forefront of sports now more than ever. As basketball becomes more analytics-driven, we've seen an increasing emphasis on "positionless basketball" – the idea that teams want their 5 most capable and cohesive players on the court as much as possible. Better, more cohesive players lead to more efficient play, which leads to winning more games. Or so we're told. The NBA is typically seen as one of the leagues using analytics most to its advantage. As a college basketball fan (lunatic?), I was curious to see how much positionless basketball has trickled into the college ranks.

With publicly-available information, thanks to the *ncaahoopR* R package (https://github.com/lbenz730/ncaahoopR), we can get data quickly and easily in order to then analyze. Using this R package, I scraped box score and shot chart data for the **2019-2020 ACC basketball season**, the league I follow the closest. Through this analysis, I used in-game statistics and player attribute information to **let the computer tell me what positions in today's college basketball landscape look like**. If I get this data, feed it into a clustering algorithm, then look at the results, what can we learn?

DATA OVERVIEW

As mentioned above, I used the 2019-2020 ACC season and its players in my analysis. To most effectively capture the influential players in the league, I limit my analysis to players averaging at least 10 minutes played per game. This gives me 131 players to categorize. I recognize that the data could be more thorough and include more advanced statistics in further analyses, such as effective FG %, plus/minus, free throw rate, etc. I've used data that was immediately available. The data inputs for clustering make up two categories: on-court statistics and player attributes:

- On-court statistics
  - Points per Game
  - Rebounds per Game
  - Assists per Game
  - Turnovers per Game
  - Field Goal %
  - 3PT Field Goal %
  - % of Field Goals Attempted from 3PT Range
  - Minutes per Game
- Player attributes
 - Height (inches)
  - Weight (lbs)

After collecting the data, I aggregated it into one observation per player, which can then be fed into my clustering algorithm. Here's what a sample of the clustering input data looks like (before scaling it):

player_full_name	team_name	mpg	fgp	fg3p	fg3a	rpg	apg	tog	ppg	weight	height
Cassius Stanley	Duke	27.448	0.474	0.360	0.319	4.862	1.034	1.897	12.552	193	78
Chris Lykes	Miami	30.077	0.432	0.381	0.414	2.115	2.385	2.077	15.423	157	67
Armando Bacot	UNC	24.484	0.462	0.000	0.000	8.161	1.258	1.677	9.548	232	82
Cole Anthony	UNC	35.000	0.378	0.343	0.409	5.905	4.095	3.476	18.048	190	75

CLUSTERING EXERCISE

Using this input dataset of 131 players, I opted for the traditional KMeans clustering algorithm to categorize the players into clusters. I performed multiple iterations of clustering, to come up with the best output result. Looking at the Scree plot, I initially zeroed in on either 3 or 4 clusters because of their within sum of squares (WSS) scores.

Scree plot:
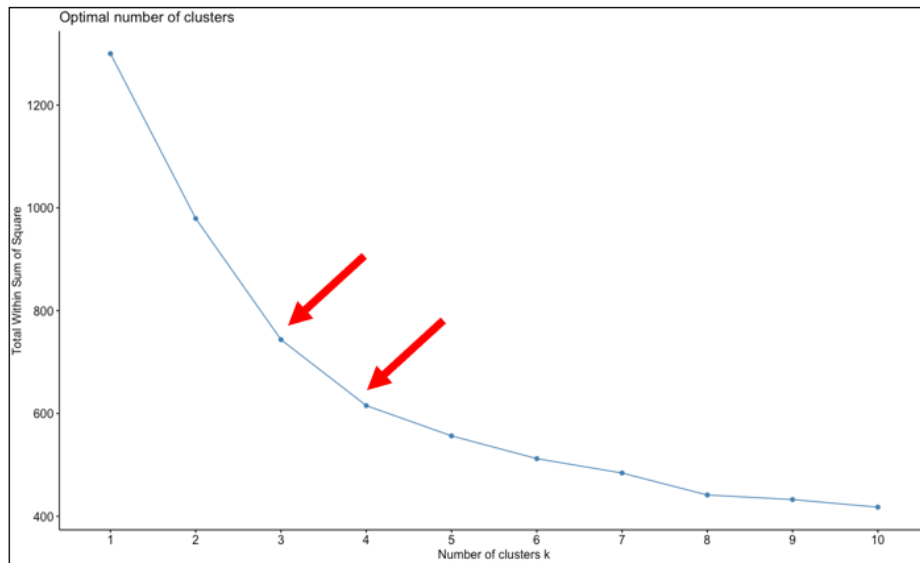
Further, I looked at the average silhouette widths to decide between 3 and 4 clusters as my optimal solution. I settled on 4 clusters, but admittedly wouldn't have disagreed with a 3-cluster solution.

Cluster silhouette widths:
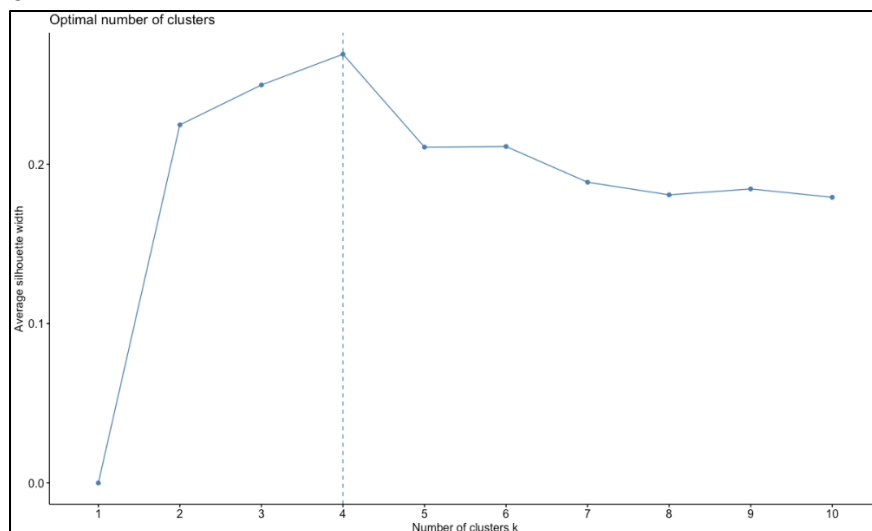
For context, here's what the cluster solutions looked like for 2, 3, 4, and 5 clusters.

CLUSTERING RESULTS

Each of the 4 clusters has its own distinctive features and can be summarized as follows.

- **Cluster 1: Ball Handlers & Guards**
  - 22 players in cluster (17%)
  - Summary: These are the players who have the ball in their hands most of the game and are focal points of their teams' offenses.
  - **Key stats: Highest minutes played per game (33.4), most turnovers per game (3.2), and average height of 6'2".** This lends nicely to signify these players are guards.
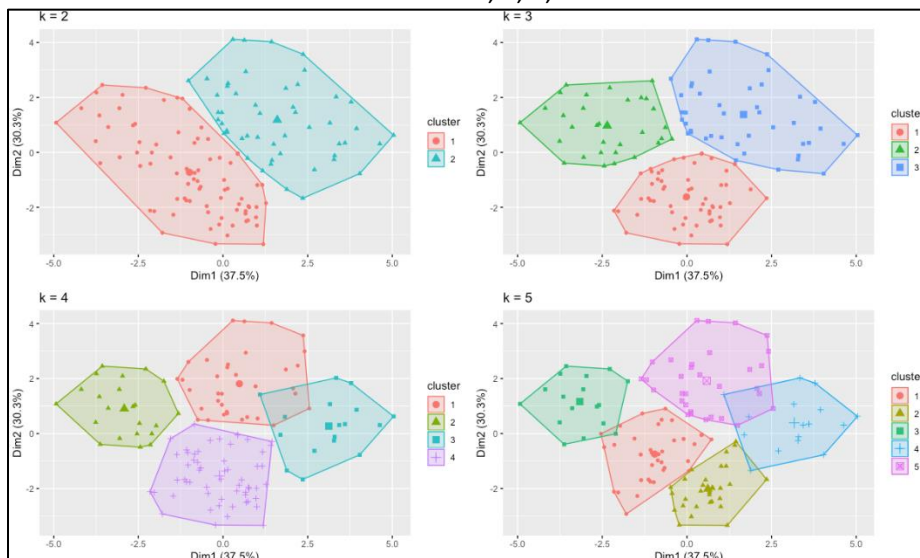 - Notable players: Kihei Clark (Virginia), Elijah Hughes (Syracuse), Tre Jones (Duke), Cole Anthony (UNC)
- **Cluster 2: Role Players**
  - 17 players in cluster (13%)
  - Summary: These players don't play many minutes, take care of the ball, and are put on the floor to do one or two things well, to help their team in spurts.
  - **Key stats: Fewest minutes played per game (16.8), highest FG% (55%), lowest turnovers per game (0.8).**
  - Notable players: Javin Delaurier (Duke), Tyrece Radford (VT), Balsa Koprivica (FSU)
- **Cluster 3: Shooters & Wings**
  - 57 players in cluster (44%)
  - Summary: Largest cluster by volume of players. These are mostly off-ball wing players, that tend to take most of their field goal attempts from 3PT range.
  - **Key stats: 50% of attempted FGs are 3's, medium height (6'4.5" avg), 38% shooting %**
  - Notable players: Ryan McMahon (Louisville), Ryan Murphy (Pitt), DJ Vasiljevic (Miami), John Newman III (Clemson)
- **Cluster 4: Versatile Bigs**
  - 35 players in cluster (27%)
  - Summary: These are big frontcourt players that are the best rebounders on their teams and shoot a high percentage from the field.
  - **Key stats: 6'8" average height, 11.3 points per game, 6.4 rebounds per game.**
  - Notable players: Vernon Carey Jr. (Duke), Mamadi Diakite (Virginia), John Mooney (Notre Dame), James Banks III (Georgia Tech)

As you would expect, each team has their fair share of players belonging to each cluster (aside from my favorite team, Louisville). Besides the Cardinals, all teams have players from 3 of the 4 clusters, if not all 4 clusters. Here's what each team looks like.

Cluster breakdowns by team:

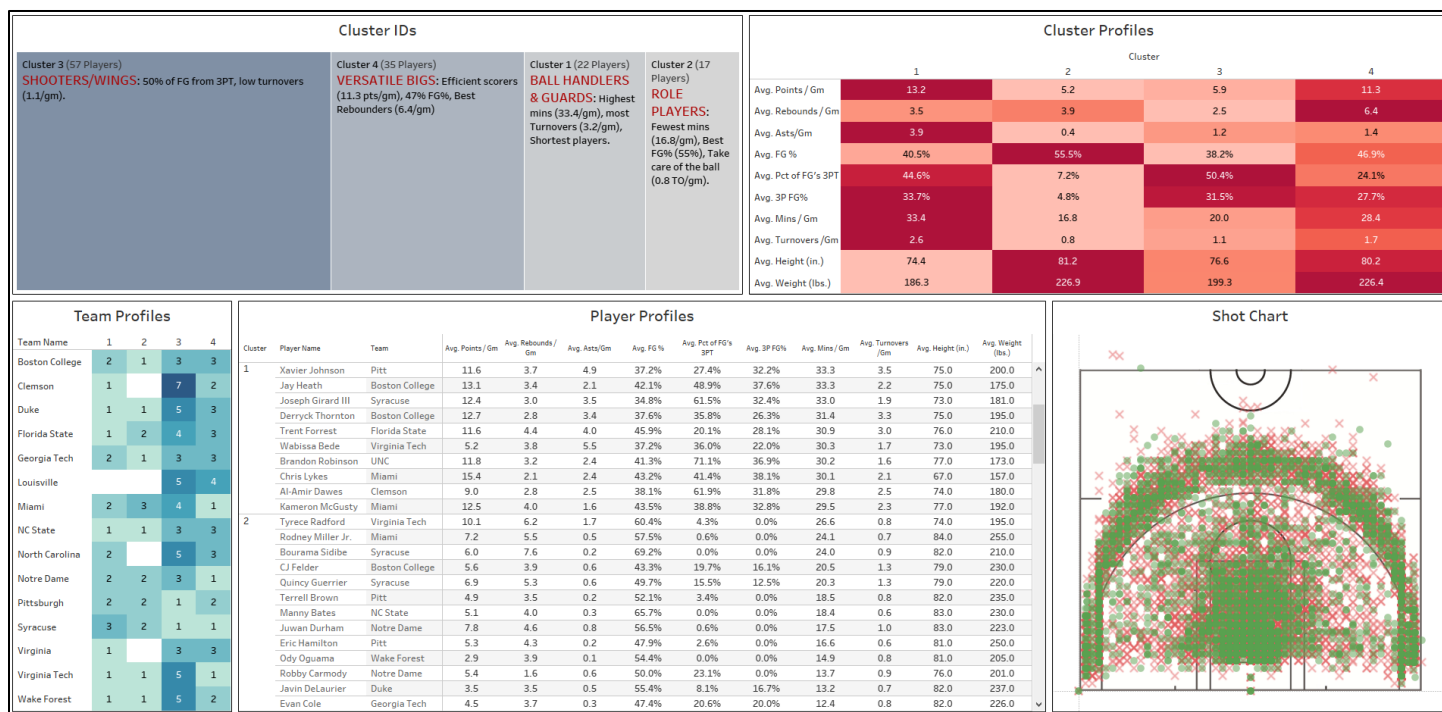| TEAM	Cluster 1	Cluster 2	Cluster 3	Cluster 4	TOTAL
Boston College	2	1	3	3	9
Clemson	1		7	2	10
Duke	1	1	5	3	10
Florida State	1	2	4	3	10
Georgia Tech	2	1	3	3	9
Louisville			5	4	9
Miami	2	3	4	1	10
NC State	1	1	3	3	8
Notre Dame	2	2	3	1	8
Pitt	2	2	1	2	7
Syracuse	3	2	1	1	7
UNC	2		5	3	10
UVA	1		3	3	7
Virginia Tech	1	1	5	1	8
Wake Forest	1	1	5	2	9
Total	22	17	57	35	131

For those curious where each player fell, here is a breakdown of each player by cluster.

Cluster 1	Cluster 2	Cluster 3	Cluster 4
Derryck Thornton (BC)	CJ Felder (BC)	Jared Hamilton (BC)	Jairus Hamilton (BC)
Jay Heath (BC)	Bourama Sidibe (CUSE)	Julian Rishwain (BC)	Nik Popovic (BC)
Al-Amir Dawes (CLEM)	Quincy Guerrier (CUSE)	Kamari Williams (BC)	Steffon Mitchell (BC)
Buddy Boeheim (CUSE)	Javin DeLaurier (DUKE)	Alex Hemenway (CLEM)	Aamir Simms (CLEM)
Elijah Hughes (CUSE)	Balsa Koprivica (FSU)	Chase Hunter (CLEM)	Tevin Mack (CLEM)
Joseph Girard III (CUSE)	Dominik Olejniczak (FSU)	Curran Scott (CLEM)	Marek Dolezaj (CUSE)
Tre Jones (DUKE)	Evan Cole (GT)	Clyde Trapp (CLEM)	Cassius Stanley (DUKE)
Trent Forrest (FSU)	Anthony Walker (MIA)	Hunter Tyson (CLEM)	Vernon Carey Jr. (DUKE)
Jose Alvarado (GT)	Deng Gak (MIA)	Jonathan Baehre (CLEM)	Wendell Moore Jr. (DUKE)
Michael Devoe (GT)	Rodney Miller Jr. (MIA)	John Newman III (CLEM)	Devin Vassell (FSU)
Chris Lykes (MIA)	Manny Bates (NCST)	Jalen Carey (CUSE)	Patrick Williams (FSU)
Kameron McGusty (MIA)	Juwan Durham (ND)	Alex O'Connell (DUKE)	RaiQuan Gray (FSU)
Markell Johnson (NCST)	Robby Carmody (ND)	Joey Baker (DUKE)	James Banks III (GT)
Prentiss Hubb (ND)	Eric Hamilton (PITT)	Jordan Goldwire (DUKE)	Jordan Usher (GT)
TJ Gibbs (ND)	Terrell Brown (PITT)	Jack White (DUKE)	Moses Wright (GT)
Trey McGowens (PITT)	Tyrece Radford (VT)	Matthew Hurt (DUKE)	Dwayne Sutton (LOU)
Xavier Johnson (PITT)	Ody Oguama (WAKE)	Anthony Polite (FSU)	Jordan Nwora (LOU)
Brandon Robinson (UNC)		Malik Osborne (FSU)	Malik Williams (LOU)
Cole Anthony (UNC)		M. J. Walker (FSU)	Steven Enoch (LOU)
Kihei Clark (UVA)		Rayquan Evans (FSU)	Sam Waardenburg (MIA)
Wabissa Bede (VT)		Asanti Price (GT)	C. J. Bryce (NCST)
Brandon Childress (WAKE)		Bubba Parham (GT)	Devon Daniels (NCST)
		Khalid Moore (GT)	DJ Funderburk (NCST)
		David Johnson (LOU)	John Mooney (ND)
		Darius Perry (LOU)	Au'Diese Toney (PITT)
		Lamarr Kimble (LOU)	Justin Champagnie (PITT)
		Ryan McMahon (LOU)	Armando Bacot (UNC)
		Samuell Williamson (LOU)	Garrison Brooks (UNC)
		Dejan Vasiljevic (MIA)	Leaky Black (UNC)
		Harlond Beverly (MIA)	Braxton Key (UVA)
		Isaiah Wong (MIA)	Jay Huff (UVA)
		Keith Stone (MIA)	Mamadi Diakite (UVA)
		Braxton Beverly (NCST)	Landers Nolley II (VT)
		Jericole Hellems (NCST)	Chaundee Brown (WAKE)
		Pat Andree (NCST)	Olivier Sarr (WAKE)
		Dane Goodwin (ND)	
		Nate Laszewski (ND)	
		Rex Pflueger (ND)	
		Ryan Murphy (PITT)	
		Anthony Harris (UNC)	
		Andrew Platek (UNC)	
		Christian Keeling (UNC)	
		Jeremiah Francis (UNC)	
		Justin Pierce (UNC)	
		Casey Morsell (UVA)	
		Kody Stattmann (UVA)	
		Tomas Woldetensae (UVA)	
		Hunter Cattoor (VT)	
		Isaiah Wilkins (VT)	
		Jalen Cone (VT)	
		Nahiem Alleyne (VT)	
		P. J. Horne (VT)	
		Andrien White (WAKE)	
		Ismael Massoud (WAKE)	
		Isaiah Mucius (WAKE)	
		Jahcobi Neath (WAKE)	
		Torry Johnson (WAKE)	

RESULTS VISUALIZATION

To display my clustering results in one visual, I created a Tableau dashboard that shows each player's statistics, attributes associated with each cluster, and an interactive shot chart with the coordinates of each player's field goal attempts (where data was available). The workbook is available on Github along with all R code, and can be downloaded and interacted with using Tableau Reader (https://www.tableau.com/products/reader).

Tableau dashboard:

Cluster IDs

Cluster 3 (57 Players)
SHOOTERS/WINGS: 50% of FG from 3PT, low turnovers (1.1/gm).

Cluster 4 (35 Players)
VERSATILE BIGS: Efficient scorers (11.3 pts/gm), 47% FG%, Best Rebounders (6.4/gm)

Cluster 1 (22 Players)
BALL HANDLERS & GUARDS: Highest mins (33.4/gm), most Turnovers (3.2/gm), Shortest players.

Cluster 2 (17 Players)
ROLE PLAYERS: Fewest mins (16.8/gm), Best FG% (55%), Take care of the ball (0.8 TO/gm).

Cluster Profiles

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Avg. Points / Gm	13.2	5.2	5.9	11.3
Avg. Rebounds / Gm	3.5	3.9	2.5	6.4
Avg. Asts/Gm	3.9	0.4	1.2	1.4
Avg. FG %	40.5%	55.5%	38.2%	46.9%
Avg. Pct of FG's 3PT	44.6%	7.2%	50.4%	24.1%
Avg. 3P FG%	33.7%	4.8%	31.5%	27.7%
Avg. Mins / Gm	33.4	16.8	20.0	28.4
Avg. Turnovers /Gm	2.6	0.8	1.1	1.7
Avg. Height (in.)	74.4	81.2	76.6	80.2
Avg. Weight (lbs.)	186.3	226.9	199.3	226.4

Team Profiles

Team Name	1	2	3	4
Boston College	2	1	3	3
Clemson	1		7	2
Duke	1	1	5	3
Florida State	1	2	4	3
Georgia Tech	2	1	3	3
Louisville			5	4
Miami	2	3	4	1
NC State	1	1	3	3
North Carolina	2		5	3
Notre Dame	2	2	3	1
Pittsburgh	2	2	1	2
Syracuse	3	2	1	1
Virginia	1		3	3
Virginia Tech	1	1	5	1
Wake Forest	1	1	5	2

Player Profiles

Cluster	Player Name	Team	Avg. Points / Gm	Avg. Rebounds / Gm	Avg. Asts/Gm	Avg. FG %	Avg. Pct of FG's 3PT	Avg. 3P FG%	Avg. Mins / Gm	Avg. Turnovers /Gm	Avg. Height (in.)	Avg. Weight (lbs.)
1	Xavier Johnson	Pitt	11.6	3.7	4.9	37.2%	27.4%	32.2%	33.3	3.5	75.0	200.0
	Jay Heath	Boston College	13.1	3.4	2.1	42.1%	48.9%	37.6%	33.3	2.2	75.0	175.0
	Joseph Girard III	Syracuse	12.4	3.0	3.5	34.8%	61.5%	32.4%	33.0	1.9	73.0	181.0
	Derryck Thornton	Boston College	12.7	2.8	3.4	37.6%	35.8%	26.3%	31.4	3.3	75.0	195.0
	Trent Forrest	Florida State	11.6	4.4	4.0	45.9%	20.1%	28.1%	30.9	3.0	76.0	210.0
	Wabissa Bede	Virginia Tech	5.2	3.8	5.5	37.2%	36.0%	22.0%	30.3	1.7	73.0	195.0
	Brandon Robinson	UNC	11.8	3.2	2.4	41.3%	71.1%	36.9%	30.2	1.6	77.0	173.0
	Chris Lykes	Miami	15.4	2.1	2.4	43.2%	41.4%	38.1%	30.1	2.1	67.0	157.0
	Al-Amir Dawes	Clemson	9.0	2.8	2.5	38.1%	61.9%	31.8%	29.8	2.5	74.0	180.0
	Kameron McGusty	Miami	12.5	4.0	1.6	43.5%	38.8%	32.8%	29.5	2.3	77.0	192.0
2	Tyrece Radford	Virginia Tech	10.1	6.2	1.7	60.4%	4.3%	0.0%	26.6	0.8	74.0	195.0
	Rodney Miller Jr.	Miami	7.2	5.5	0.5	57.5%	0.6%	0.0%	24.1	0.7	84.0	255.0
	Bourama Sidibe	Syracuse	6.0	7.6	0.2	69.2%	0.0%	0.0%	24.0	0.9	82.0	210.0
	CJ Felder	Boston College	5.6	3.9	0.6	43.3%	19.7%	16.1%	20.5	1.3	79.0	230.0
	Quincy Guerrier	Syracuse	6.9	5.3	0.6	49.7%	15.5%	12.5%	20.3	1.3	79.0	220.0
	Terrell Brown	Pitt	4.9	3.5	0.2	52.1%	3.4%	0.0%	18.5	0.8	82.0	235.0
	Manny Bates	NC State	4.0	4.0	0.3	65.7%	0.0%	0.0%	18.4	0.6	83.0	230.0
	Juwan Durham	Notre Dame	7.8	4.6	0.8	56.5%	0.6%	0.0%	17.5	1.0	83.0	223.0
	Eric Hamilton	Pitt	5.3	4.3	0.2	47.9%	2.6%	0.0%	16.6	0.6	81.0	250.0
	Ody Oguama	Wake Forest	2.9	3.9	0.1	54.4%	0.0%	0.0%	14.9	0.8	81.0	205.0
	Robby Carmody	Notre Dame	5.4	1.6	0.6	50.0%	23.1%	0.0%	13.7	0.9	76.0	201.0
	Javin DeLaurier	Duke	3.5	3.5	0.5	55.4%	8.1%	16.7%	13.2	0.7	82.0	237.0
	Evan Cole	Georgia Tech	4.5	3.7	0.3	47.4%	20.6%	20.0%	12.4	0.8	82.0	226.0

Shot Chart

(shot chart figure)

For anyone interested in mapping coordinates on an image within Tableau, Evolytics has an excellent tutorial available at https://evolytics.com/blog/how-to-map-anything-in-tableau/.

CONCLUSION

Overall, I did not find anything earth shattering in the analysis, but this does confirm my notion that basketball is moving away from positional players, and instead moving toward "functional" players. With a well-balanced team, coaches can adapt to opponents, as well as present matchup nightmares for opposition teams that are not as well-rounded.

Any feedback on this analysis is welcome, as well as ideas for future analysis that you'd like to see. Feel free to reach out to joshdmark@gmail.com.