Analyzing NBA Salary Data Using Multivariate Methods

Aidan Feeley, Gabe Macklem, Ethan Straub, Andreas Broxson, Jacob Shankles

Data Cleaning

Sources:

- NBA Contracts: Spottrac
- Player & Team efficiency ratings: Basketball Reference

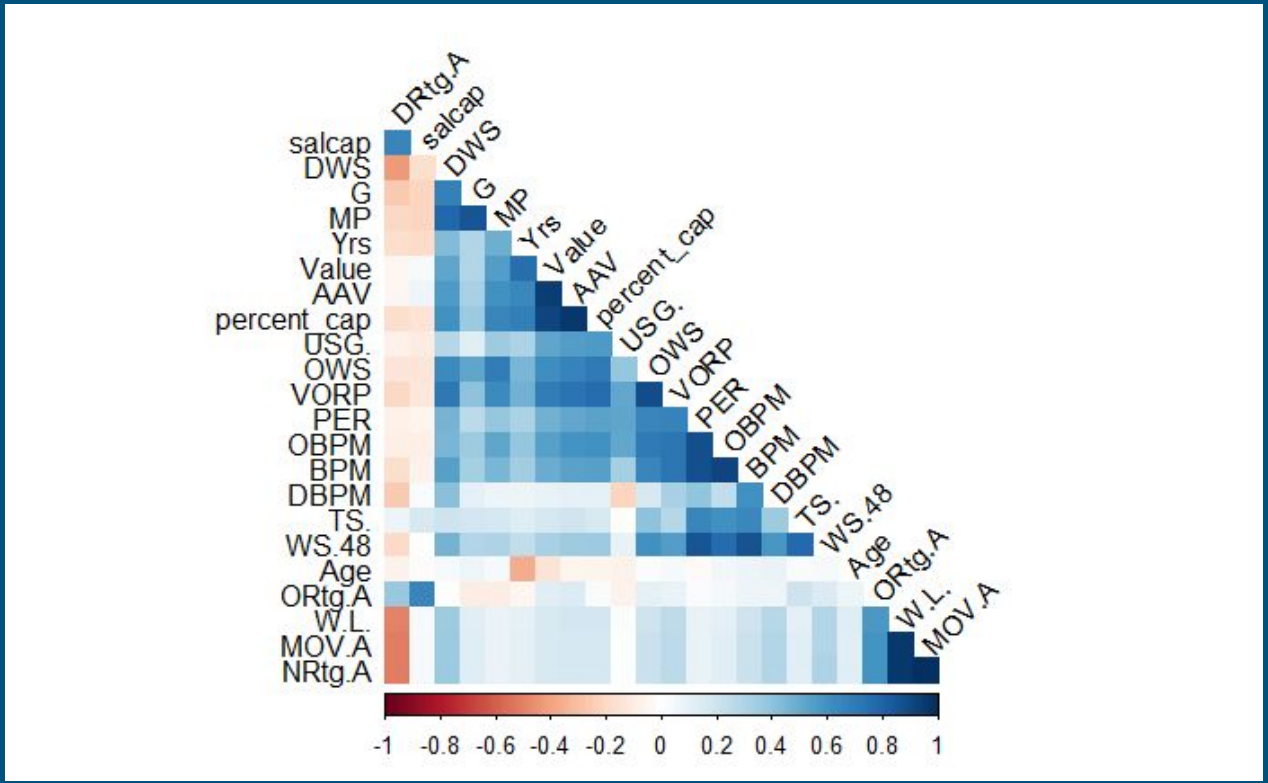
Data cleaning and shortcomings:

- Initially 3745 NBA contracts shrunk to 1162.
  - Removed any first year players without and previous NBA experience or contracts(~600-1000)
  - Removed player names that were unable to match up(~800).
  - Removed contracts that had values less than 2 million(~800).

Data

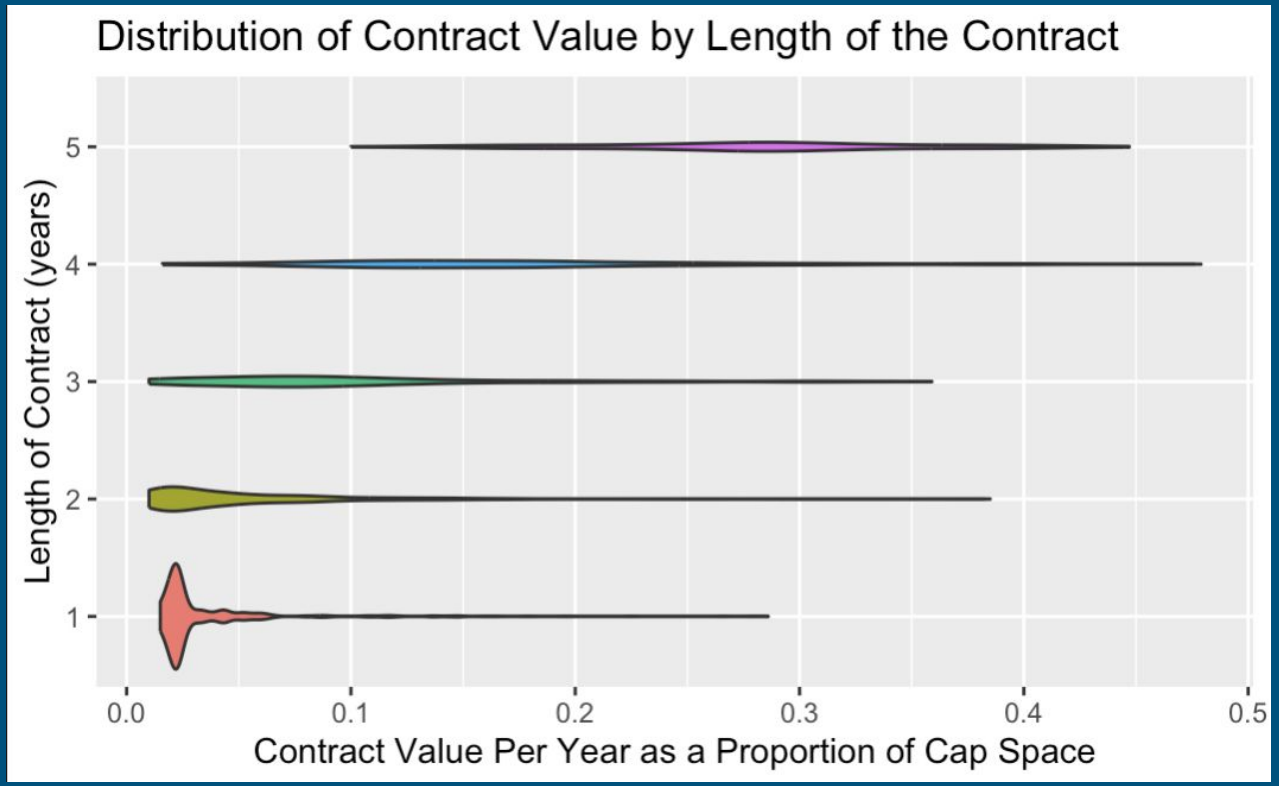
- Scraped from various sources including Spotrac and Basketball-Reference
- Over 1000 observations in final set
- Performance variables taken from season prior to signing contract
- Outcome
  - AAV- average annual value of a contract (total value/years)
  - Percent of cap- AAV/salary cap of year signed
  - Yrs- number of years for a contract
- Predictors
  - G- games played
  - MP- minutes played
  - Signed Age- player age when contract is signed
  - PER- player efficiency rating (measure of player per minute production standardized and centered at 15)
  - TS- true shooting percentage
  - USG- usage percentage (how involved a player is while on the court)
  - WS/OWS/DWS- estimated win shares attributed to a player(net/offensive/defensive)
  - BPM/OBPM/DBPM- Box Plus Minus, score differential when on court standardized (Net/Offensive/Defensive)
  - VORP- Value over replacement player (points per 100 team possessions)
  - WL- team win loss percentage
  - MOV.A- team average margin of victory
  - NRtg.A/ORtg.A/DRtt.A- how much better or worse a team performs when a player is on the court (Net/Offensive/Defensive)

Correlations

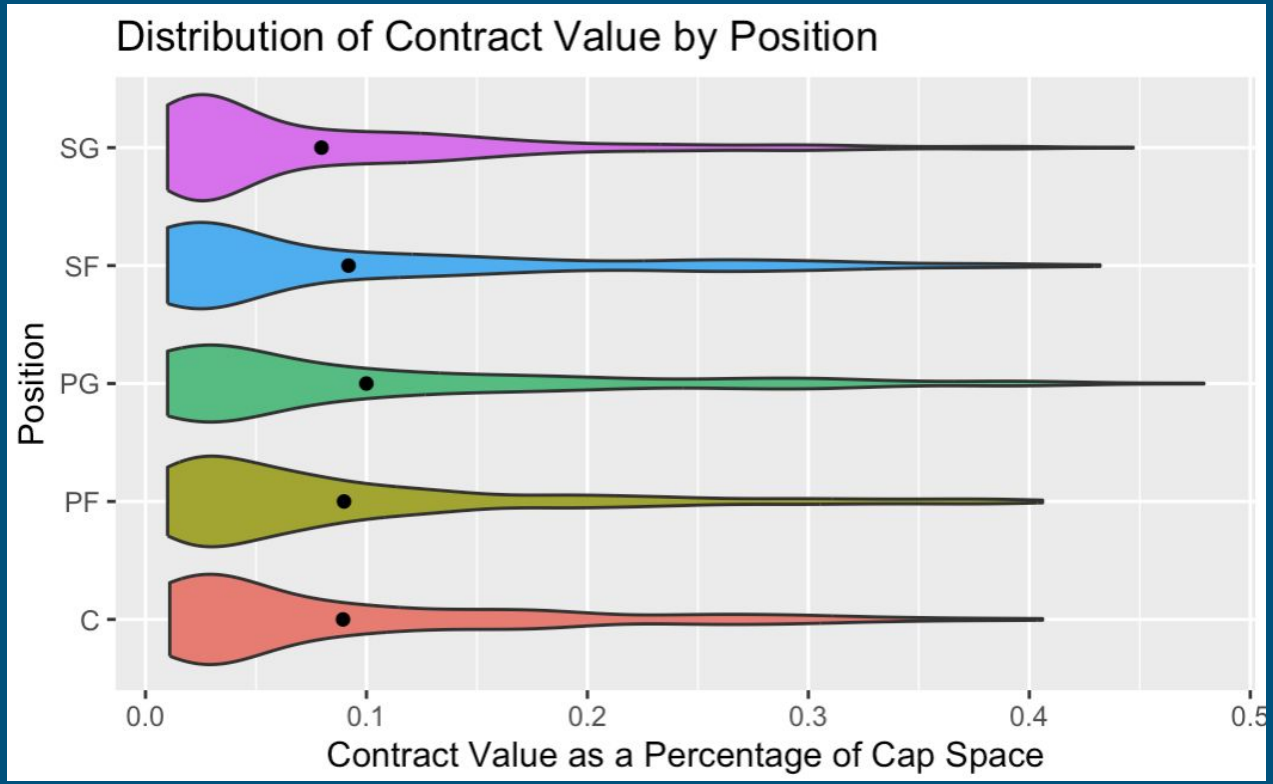


EDA: Contract Value by Contract Length

Year	cap
2014	63065000
2015	70000000
2016	94143000
2017	99093000
2018	101869000
2019	109140000
2020	109140000
2021	112414000
2022	112414000
2023	136021000



EDA - Contract Value by Position



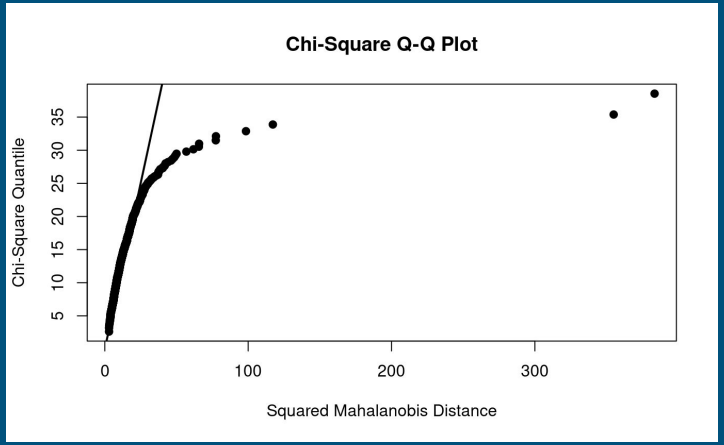
Analysis - Multivariate Linear Regression

Question: Do the expected values of **Years (per contract) and AAV** have a linear relationship with **Signed Age and Position**, after controlling for TS%, Box Plus Minus, Win Shares, Team Win-Loss Percentage, and VORP?

P-value: < 2.2e-16

F-Stat: 25.751

Conclusion: Reject the Null Hypothesis, Signed Age and Position have a significant linear relationship with Years per Contract and AAV



:(

Analysis - Does position impact contract value?

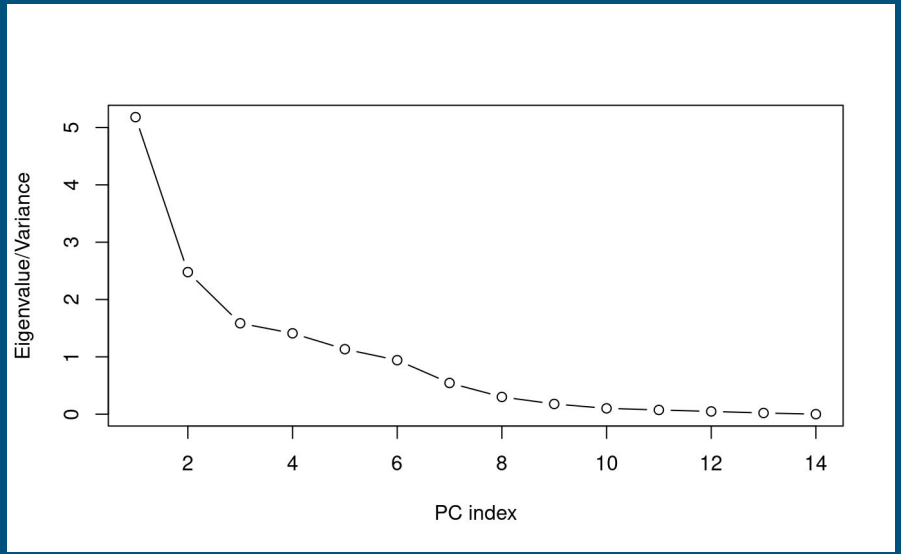
- Data exhibited non-normality so a kruskal wallis test was performed.
- It was determined that there was no significant difference between positions and AAV.
- This supports the belief that no matter the position, the impact of a player on the game is most important.

```
        Kruskal-Wallis rank sum test

data:  AAV by Position
Kruskal-Wallis chi-squared = 3.7413, df = 4, p-value = 0.4421
```

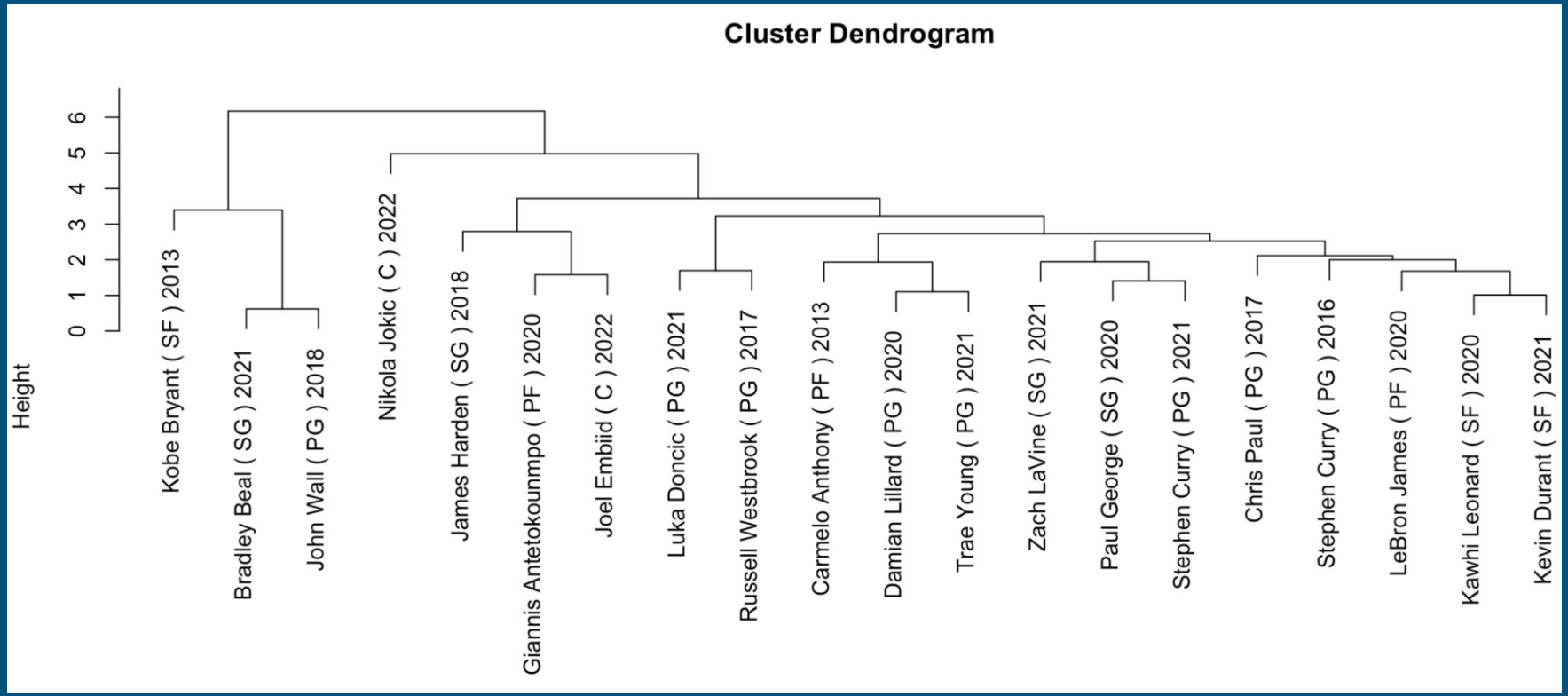
Principal Component Analysis

- It takes **6 principal components** to explain roughly **90%** of the total sample variance
- Scree plot shows a soft elbow around the **third** principal component




```
[1] 0.3701144 0.5471520 0.6604189 0.7611692 0.8421603 0.9094528 0.9483351 0.9698053
[9] 0.9825298 0.9897831 0.9950419 0.9984683 0.9999613 1.0000000
```

Clustering using efficiency statistics not contingent on position

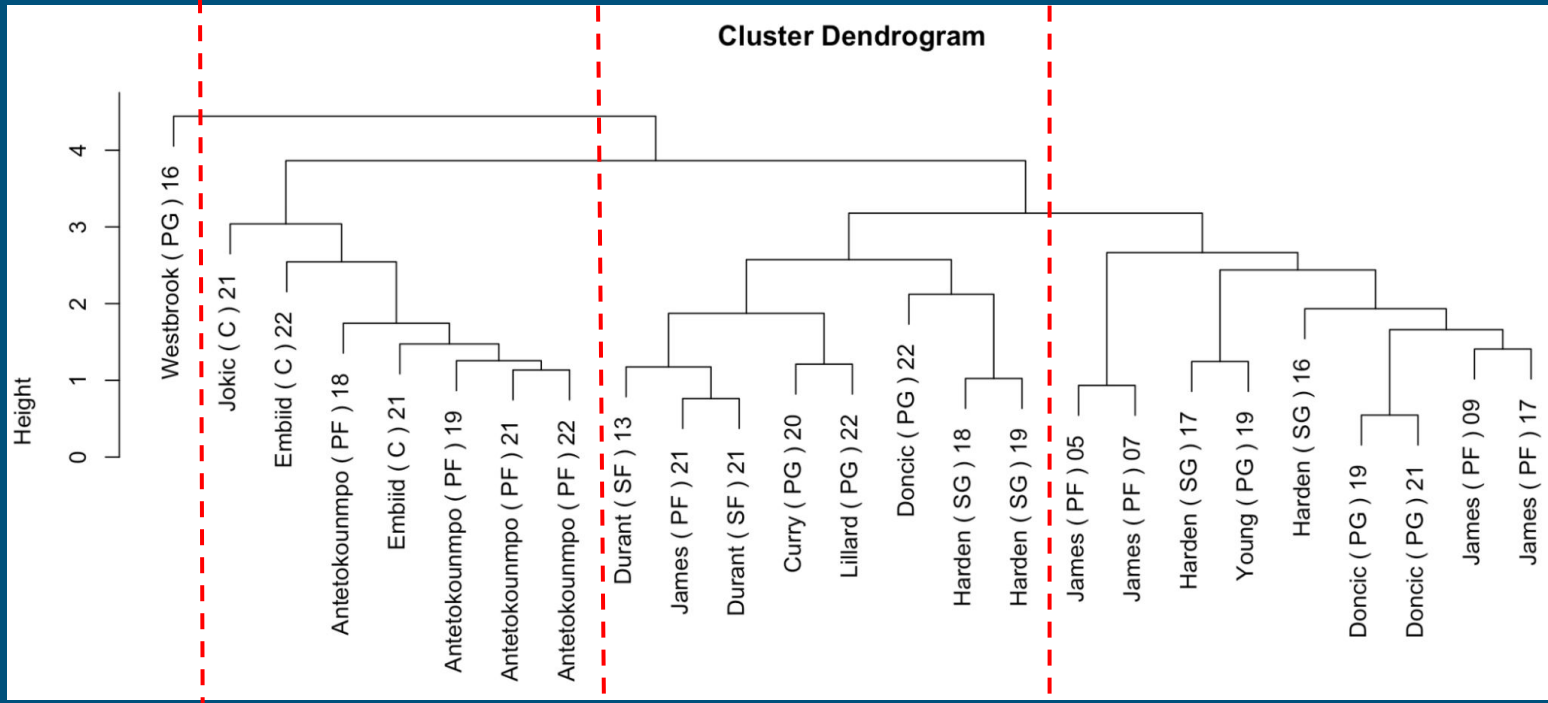


Clustering using simpler statistics (Average Linkage)

Variables being used to cluster players:

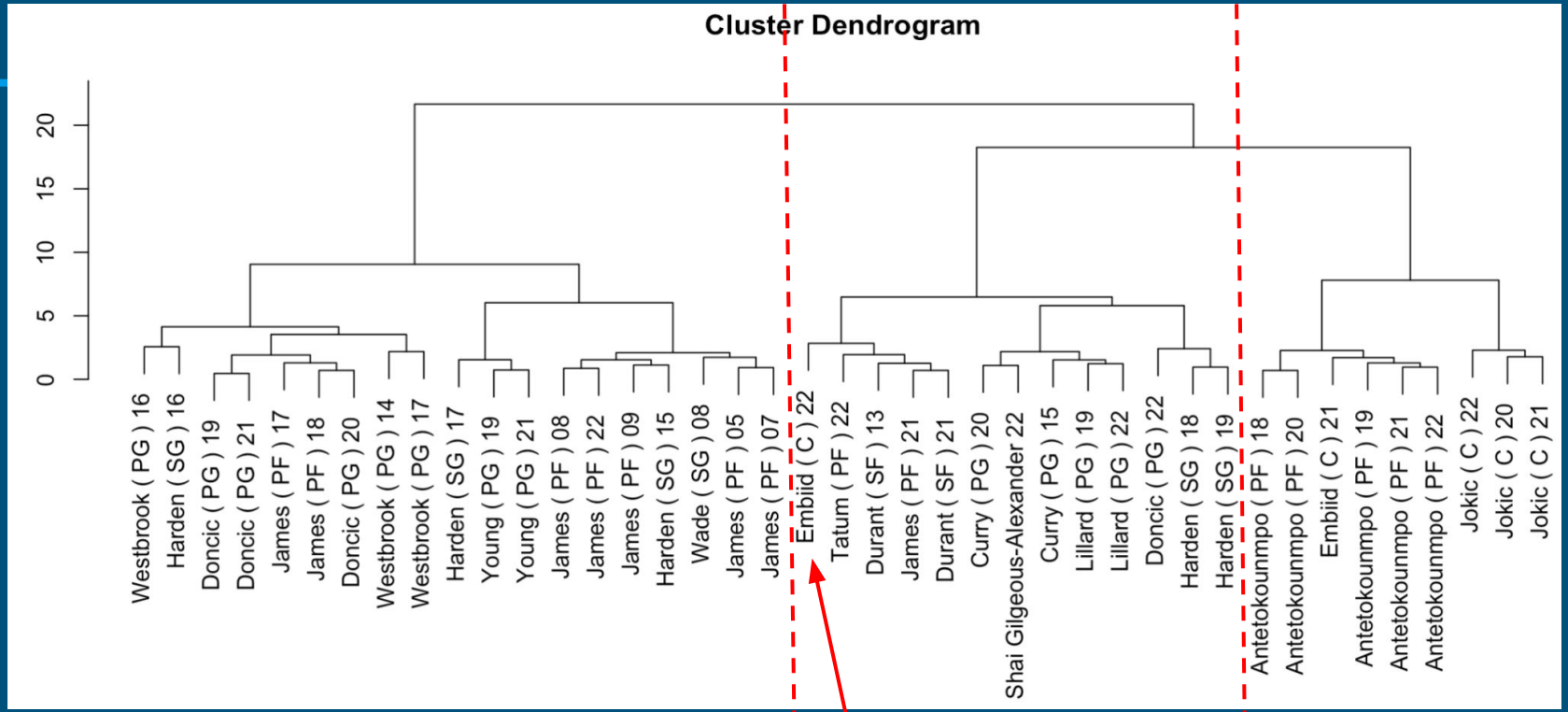
- Points
- Rebounds
- Assists
- Defensive Rebound percentage
- True shooting percentage
- Assist percentage

Variables are standardized before clustering



Westbrook | Centers and Giannis | KD, Lebron, and Guards | Lebron and Guards

Clustering using Ward's Method

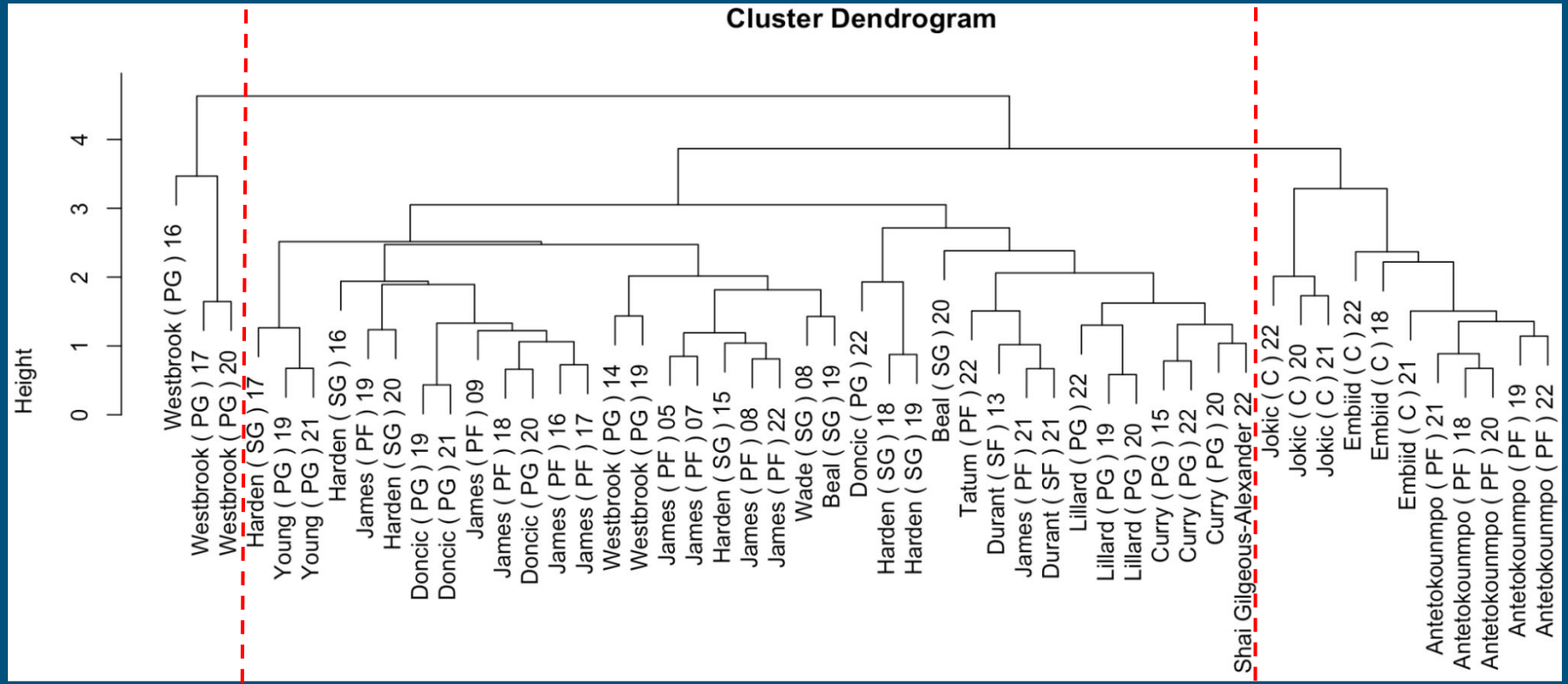


Guards and Lebron

Mix

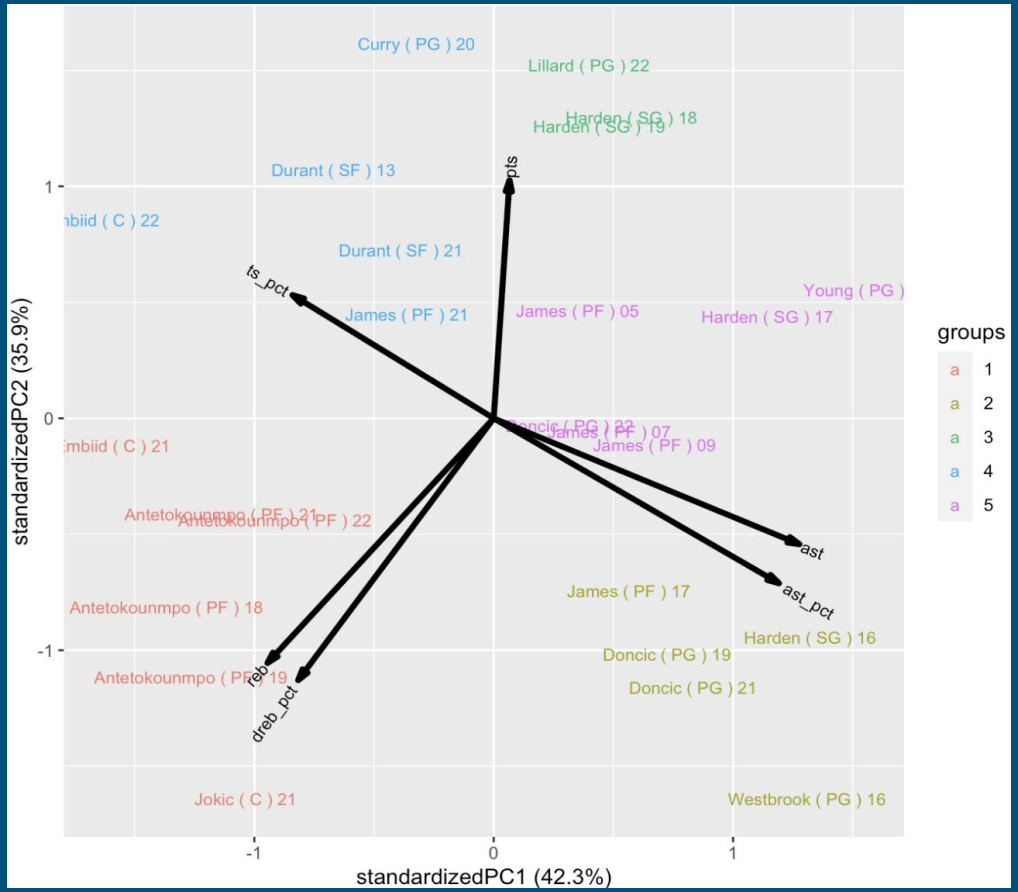
Centers and Giannis

Including more players (Average Linkage)

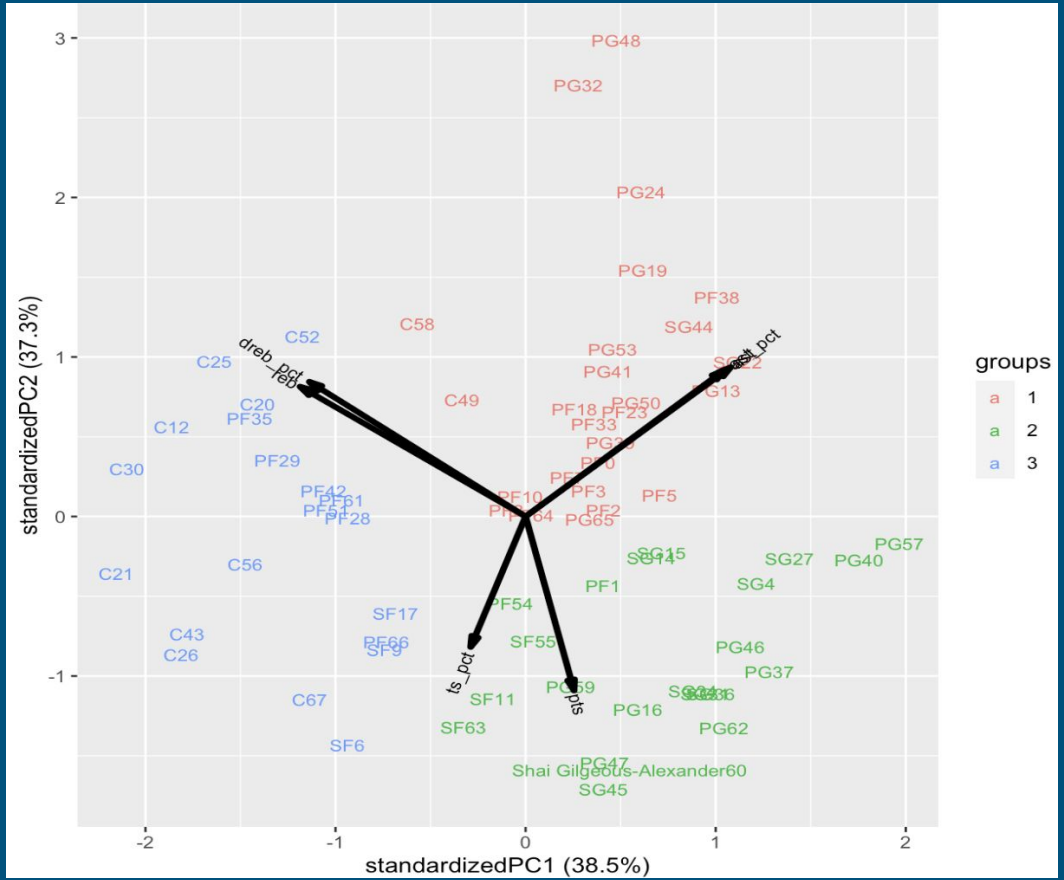


Westbrook | Mix | Centers and Giannis

K-means clustering (K=5)



Visualization of the skillset of each position



Multivariate Linear Model Predictions

Player	Team	Years	Pred. Years	AAV	Pred AAV
Jaylen Brown	Boston Celtics	5	3.87	57.25	34.30
Karl-Anthony Towns	Minnesota Timberwolves	4	2.74	55.27	19.45
Devin Booker	Phoenix Suns	4	3.67	55.27	33.87
Tyrese Haliburton	Indiana Pacers	5	4.03	40.89	35.58
LaMelo Ball	Charlotte Hornets	5	3.39	40.89	23.98
Anthony Edwards	Minnesota Timberwolves	5	4.53	40.89	36.06
Desmond Bane	Memphis Grizzlies	5	3.58	39.45	29.14
Domantas Sabonis	Sacramento Kings	4	4.35	46.50	43.04
Kawhi Leonard	Los Angeles Clippers	3	2.85	49.88	31.36
Devin Vassell	San Antonio Spurs	5	2.83	27.00	16.14
Jrue Holiday	Milwaukee Bucks	4	2.90	33.60	29.78
Jaden McDaniels	Minnesota Timberwolves	5	3.63	26.20	18.55
Dejounte Murray	Atlanta Hawks	4	3.91	28.52	32.15
Josh Hart	New York Knicks	4	3.45	20.23	22.93
Grayson Allen	Milwaukee Bucks	4	2.90	17.50	16.02
C.J. McCollum	New Orleans Pelicans	2	3.27	32.00	30.54
Onyeka Okongwu	Atlanta Hawks	4	3.30	15.49	14.75
Isaiah Stewart	Detroit Pistons	4	2.93	15.00	11.01
Kristaps Porzingis	Washington Wizards	2	3.53	30.00	33.27
Deni Avdija	Washington Wizards	4	3.30	13.75	13.46
Jarred Vanderbilt	Los Angeles Lakers	4	3.17	12.00	15.28
Josh Green	Dallas Mavericks	3	3.06	13.67	9.83
Cole Anthony	Orlando Magic	3	3.13	13.03	17.91
Zach Collins	San Antonio Spurs	2	2.44	17.41	11.01
Aaron Nesmith	Indiana Pacers	3	2.79	11.00	9.04
Zeke Nnaji	Denver Nuggets	4	2.42	8.00	2.49

- R-squared:
  - Years: 0.46
  - % of cap: 0.72
- RMSE: 11,492,850
- Not that great but wasn't expected to be.

```
Response percent_cap :

Call:
lm(formula = percent_cap ~ sign_age + G + MP + USG. + VORP + 
    NRtg.A, data = merged_data_filt)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.18404 -0.02783 -0.00387  0.02280  0.32167 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  4.359e-02  1.280e-02   3.406 0.000682 ***
sign_age    -1.685e-03  3.569e-04  -4.722 2.62e-06 ***
G           -1.715e-03  1.418e-04 -12.099  < 2e-16 ***
MP           8.380e-05  4.880e-06  17.172  < 2e-16 ***
USG.         2.937e-03  3.318e-04   8.852  < 2e-16 ***
VORP         2.734e-02  1.676e-03  16.313  < 2e-16 ***
NRtg.A       1.523e-03  3.412e-04   4.462 8.90e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0492 on 1155 degrees of freedom
Multiple R-squared:  0.716,     Adjusted R-squared:  0.7145 
F-statistic: 485.3 on 6 and 1155 DF,  p-value: < 2.2e-16
```

Neural Network Predictions

Player	Team	Years	Pred. Years	AAV	Pred AAV
Jaylen Brown	Boston Celtics	5	5	57.25	42.08
Karl-Anthony Towns	Minnesota Timberwolves	4	2	55.27	28.40
Devin Booker	Phoenix Suns	4	5	55.27	41.99
Tyrese Haliburton	Indiana Pacers	5	4	40.89	41.18
LaMelo Ball	Charlotte Hornets	5	5	40.89	34.40
Anthony Edwards	Minnesota Timberwolves	5	5	40.89	42.07
Desmond Bane	Memphis Grizzlies	5	4	39.45	36.03
Domantas Sabonis	Sacramento Kings	4	4	46.50	43.81
Kawhi Leonard	Los Angeles Clippers	3	2	49.88	36.15
Devin Vassell	San Antonio Spurs	5	2	27.00	19.84
Jrue Holiday	Milwaukee Bucks	4	3	33.60	31.80
Jaden McDaniels	Minnesota Timberwolves	5	4	26.20	16.87
Dejounte Murray	Atlanta Hawks	4	4	28.52	37.38
Josh Hart	New York Knicks	4	4	20.23	21.44
Grayson Allen	Milwaukee Bucks	4	3	17.50	13.29
C.J. McCollum	New Orleans Pelicans	2	4	32.00	33.59
Onyeka Okongwu	Atlanta Hawks	4	4	15.49	10.84
Isaiah Stewart	Detroit Pistons	4	4	15.00	12.42
Kristaps Porzingis	Washington Wizards	2	4	30.00	37.40
Deni Avdija	Washington Wizards	4	4	13.75	11.64
Jarred Vanderbilt	Los Angeles Lakers	4	4	12.00	11.12
Josh Green	Dallas Mavericks	3	4	13.67	11.03
Cole Anthony	Orlando Magic	3	4	13.03	17.49
Zach Collins	San Antonio Spurs	2	2	17.41	9.14
Aaron Nesmith	Indiana Pacers	3	2	11.00	8.63
Zeke Nnaji	Denver Nuggets	4	2	8.00	4.81

- Hidden layers: 4,3
- RMSE: 8,367,748
- Year Accuracy: 11/26 = 42%
  - Only 5 differed by 2.
- Better results when predicting large contracts.
- Predicted results should be considered in context.
  - Impact of player on chemistry?
  - Taking less money to stay on a team?
  - Player development?
  - Injuries?
  - Player over or undervalued based on stats?

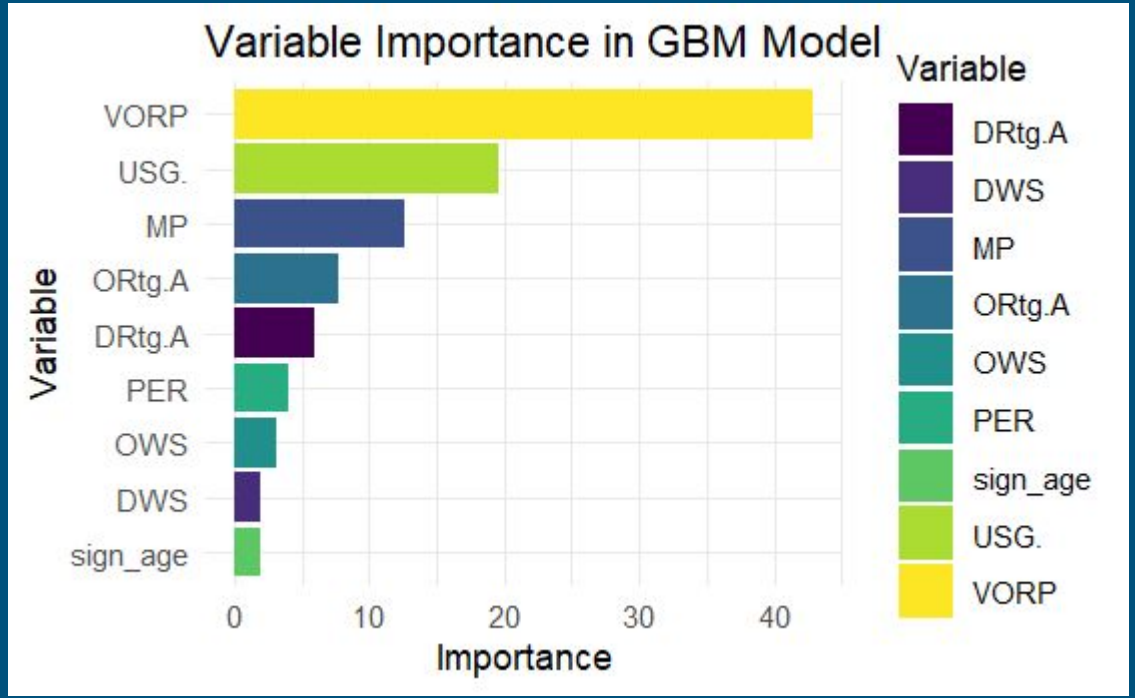
Can nnet predict max value contracts?

Player	Team	Pred. Years	Pred. AAV
Nikola Jokic	Denver Nuggets	4	48.96
Luka Doncic	Dallas Mavericks	5	48.87
Giannis Antetokounmpo	Milwaukee Bucks	4	47.79
Shai Gilgeous-Alexander	OKC Thunder	5	48.15

- Based on this seasons performance:
  - All players currently make 45-55 mil. per year on current contracts.
  - Luka and Shai are much younger, explains 5 vs 4 year difference.

GBM Model - Contributions to Models Predictive AAV Power

- VORP 42.96% (Value Over Replacement Player)
- USG 19.60% (Usage Rate)
- MP 12.67% (Minutes Played)
- ORtg.A 7.78% (Offensive Rating Adjusted)
- DRtg.A 6.00% (Defensive Rating Adjusted)
- PER 4.01% (Player Efficiency Rating)
- OWS 3.12% (Offensive Win Shares)
- DWS 1.94% (Defensive Win Shares)
- Sign_age 1.92% (Significant Age)



Sources

Cirtautas, Justinas. “NBA Players.” *Kaggle*, 13 Oct. 2023,

www.kaggle.com/datasets/justinas/nba-players-data?resource=download.

Spotrac. (n.d.). *NBA contracts*. spotrac.com. https://www.spotrac.com/nba/contracts/

Basketball Statistics & History of every Team & NBA and WNBA players. Basketball. (n.d.). https://www.basketball-reference.com/