
Basketball Analytics

Player Comparison

— Math 189R Project by Sid Rastogi —



Background

- For the midterm project, we used Logistic Regression to predict the NBA MVP
 - 99.68% accuracy on training data
 - Predicted Russell Westbrook to win the 2017 MVP!
- This time, we are more broadly looking at player comparison
 - Justify MVP prediction
 - Tell us general trends about player type and value

Objective

- Goal: Compare players in an NBA season
- Data: NBA player season total statistics per year from 1950
 - https://www.kaggle.com/drgilermo/nba-players-stats#Seasons_Stats.csv
 - Used data from 1982 onwards, since incomplete before then

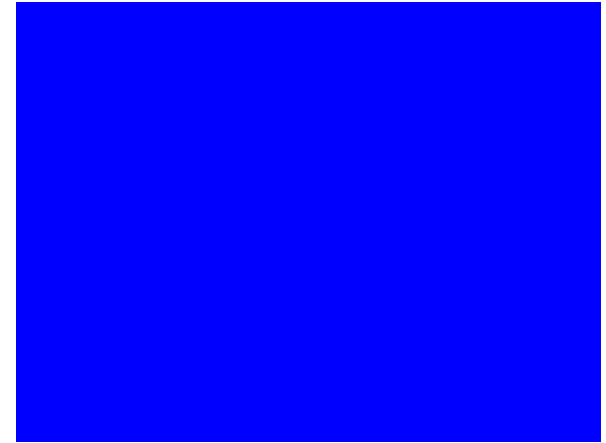
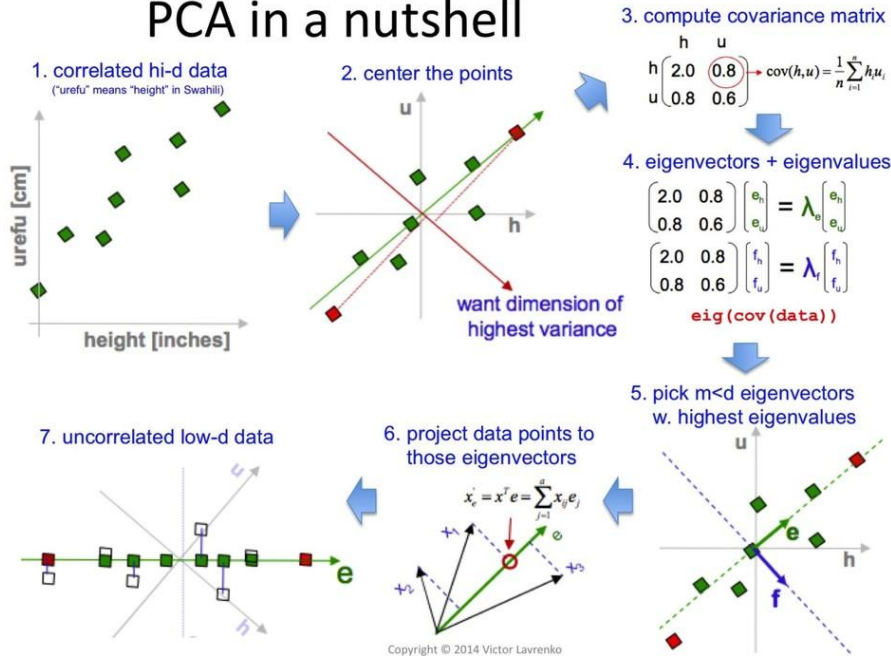
Solution

- First used PCA for dimensionality reduction
 - Reduced 42 features -> 2 PCA components
- Then used k-Means to cluster player data points
 - Trained model on individual seasons
- Looked in depth at 2017 season, as well as 5 other earlier seasons from different eras (1982, 1989, 1996, 2003, 2010)
- Filtered by MPG, GS, and PER to look at top players only

Model

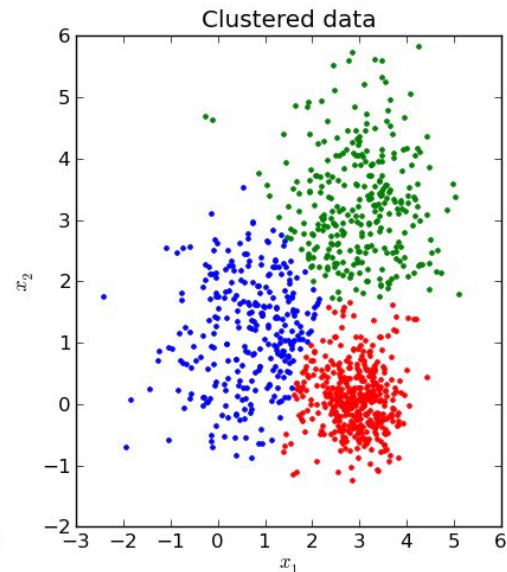
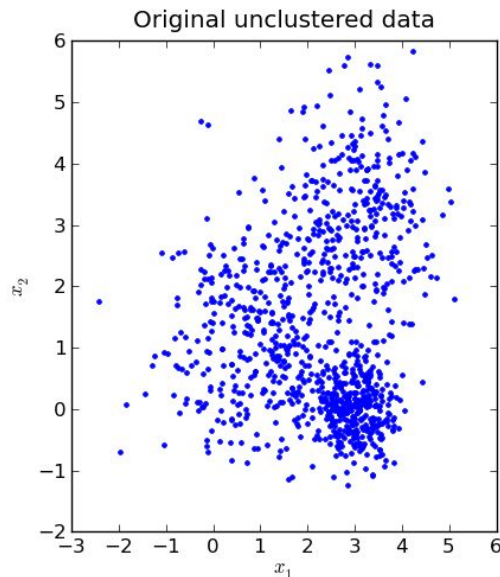
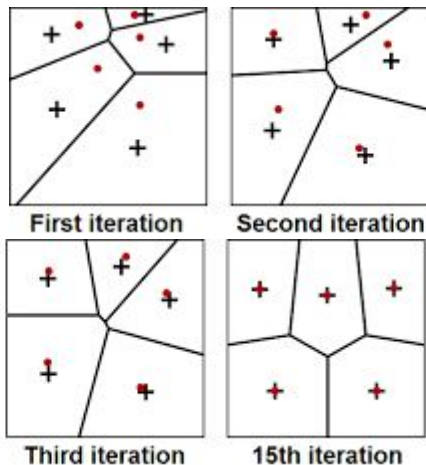
- Applied Principle Components Analysis (PCA) to data

PCA in a nutshell

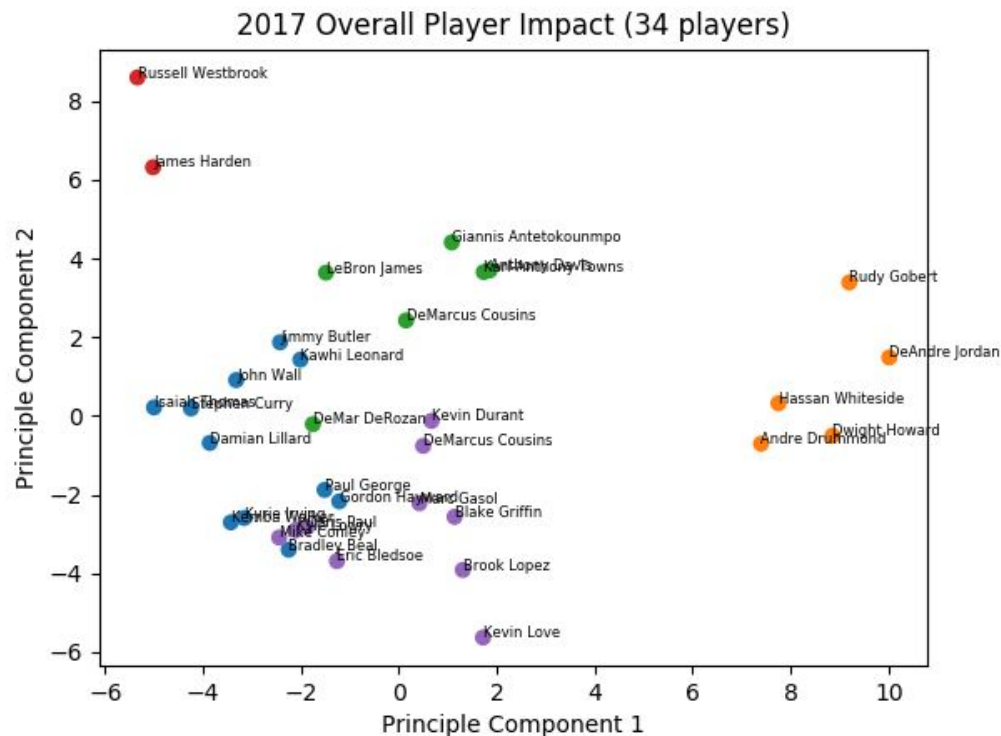


Model (cont.)

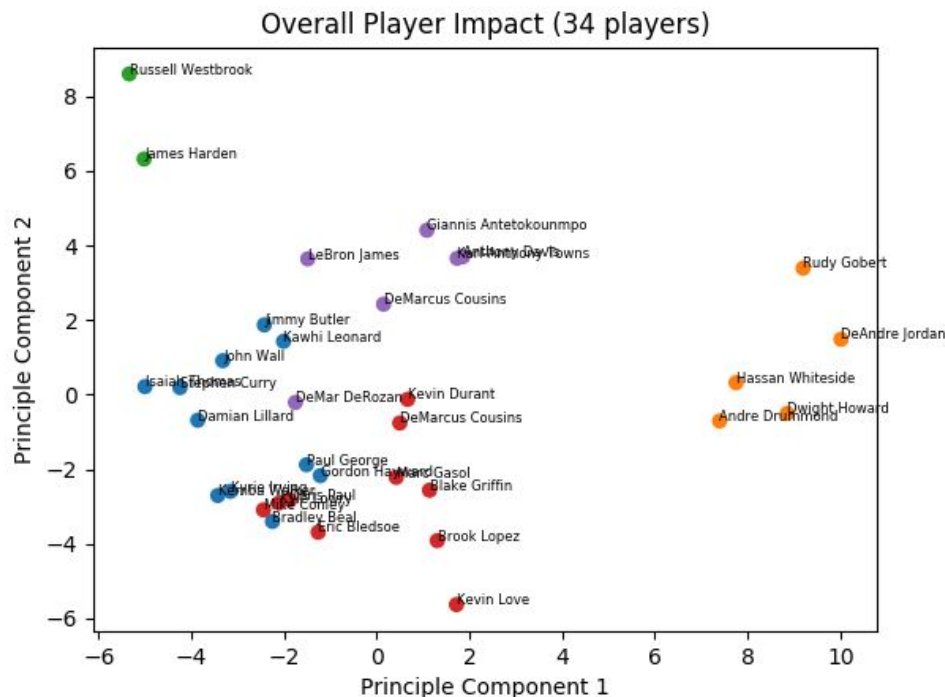
- Then applied k-Means clustering to look at player similarity
- Lloyd's algorithm



Results - 2017 (Overall)



Results - 2017 (Overall)



2016-17 NBA Awards Voting

« 2015-16 Awards Voting

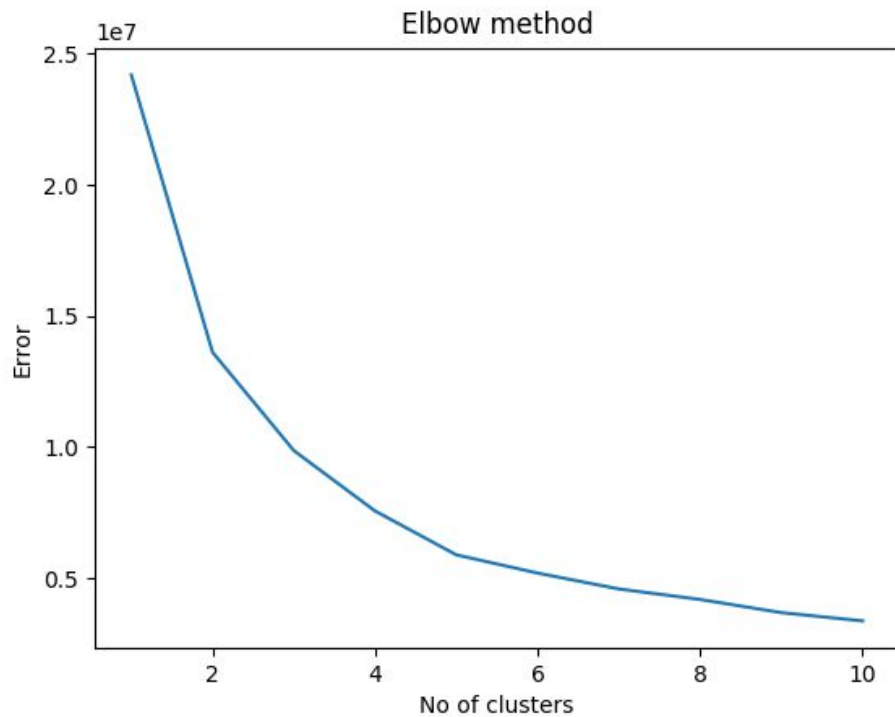
[2017-18 Awards Voting »](#)

Most Valuable Player

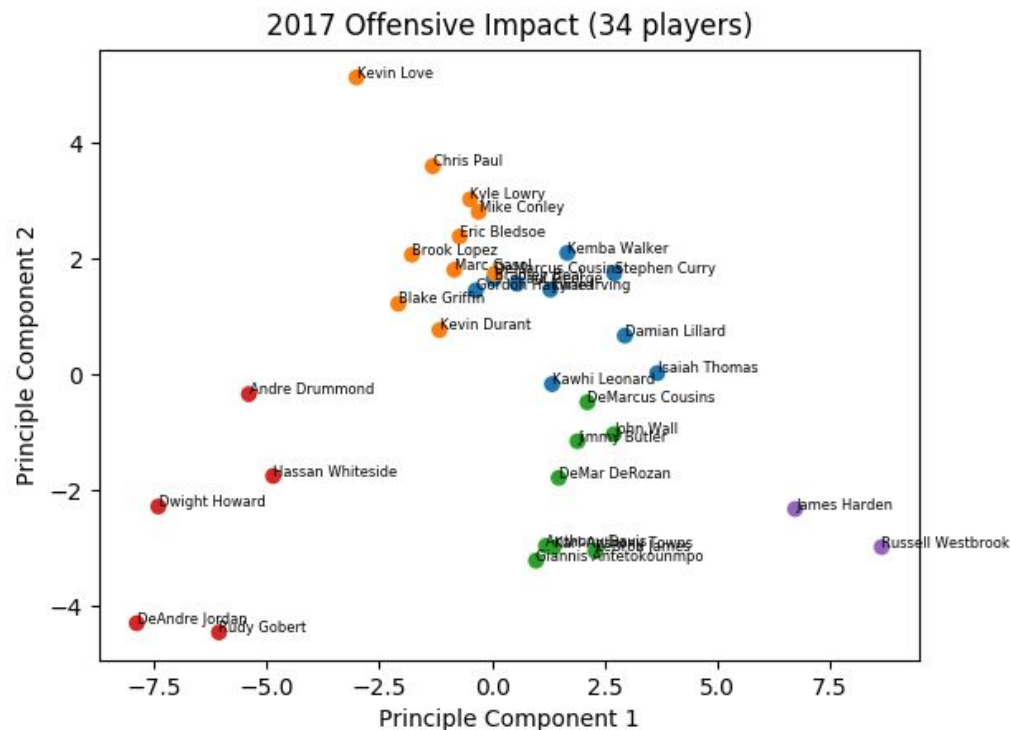
Share & more ▼ Glossary

				Voting			
Rank	Player	Age	Tm	First	Pts Won	Pts Max	Share
1	Russell Westbrook	28	OKC	69.0	888.0	1010	0.879
2	James Harden	27	HOU	22.0	753.0	1010	0.746
3	Kawhi Leonard	25	SAS	9.0	500.0	1010	0.495
4	LeBron James	32	CLE	1.0	333.0	1010	0.330
5	Isaiah Thomas	27	BOS	0.0	81.0	1010	0.080
6	Stephen Curry	28	GSW	0.0	52.0	1010	0.051
7T	Giannis Antetokounmpo	22	MIL	0.0	7.0	1010	0.007
7T	John Wall	26	WAS	0.0	7.0	1010	0.007
9T	Anthony Davis	23	NOP	0.0	2.0	1010	0.002
9T	Kevin Durant	28	GSW	0.0	2.0	1010	0.002
11	DeMar DeRozan	27	TOR	0.0	1.0	1010	0.001

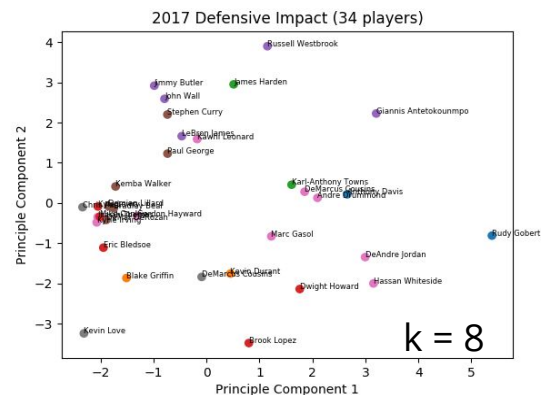
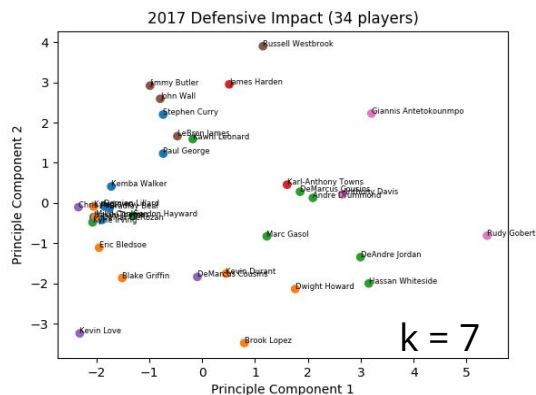
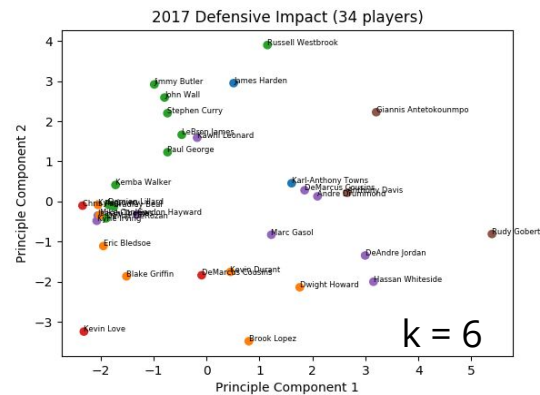
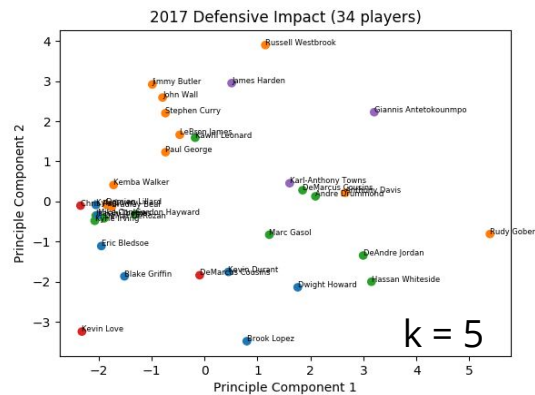
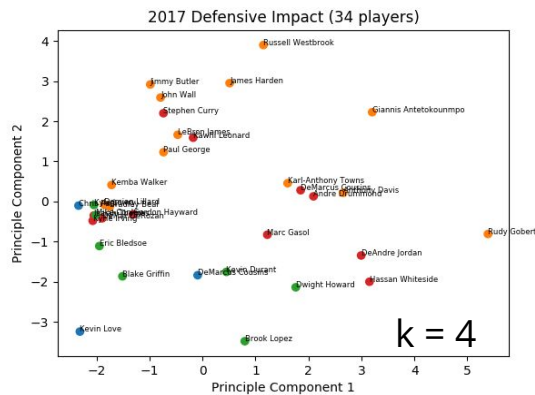
kMeans Error Plot



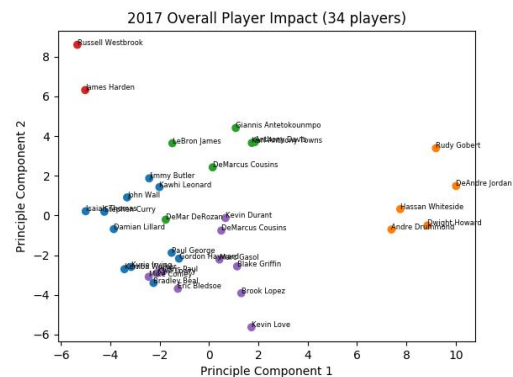
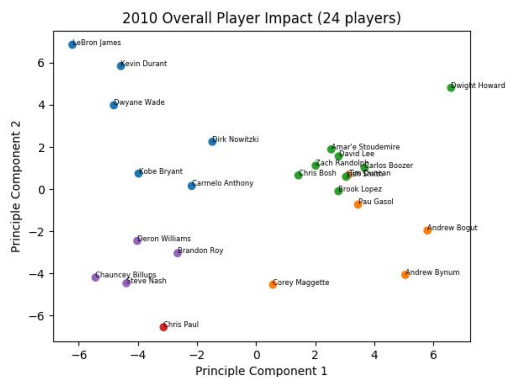
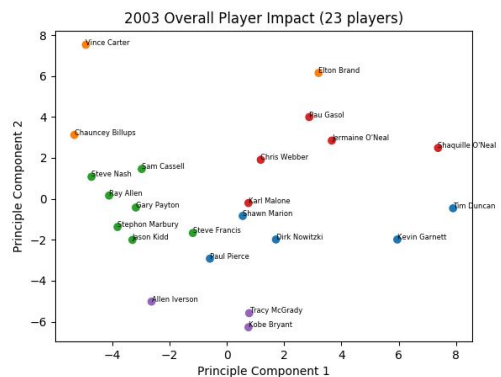
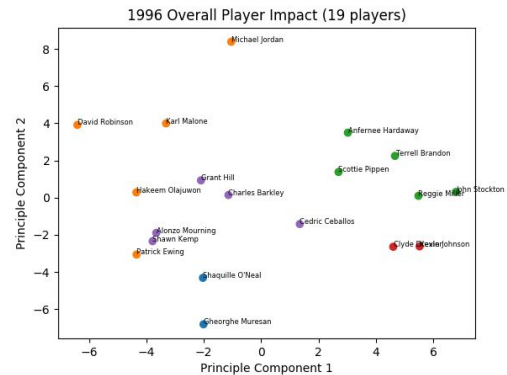
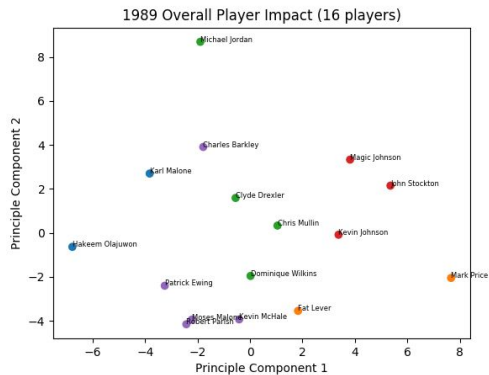
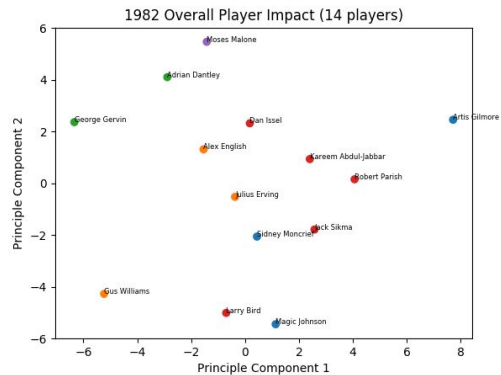
Results - 2017 (Offensive)



Results - 2017 (Defensive)



Results (1982-2017)



Project

- Github link:

<https://github.com/srastogi1011/MATH-189R-The-Math-of-Big-Data>

Conclusions

- PCA + kMeans can show us different types of players and corroborates the MVP prediction well
- Offensive statistics are generally a good indication of a player's offensive contributions
- Defensive statistics don't capture defensive impact well
- Drawbacks:
 - Doesn't factor in team success
 - Only goes till 2017

Looking forward...

- Dataset could be used to answer other interesting questions (team comparisons, player uniqueness, etc.)
- Could draw correlation between player type/production and contract values to determine monetary value of a player's skill set
- Could find more data (perhaps visual) to derive better defensive metrics

Thanks!

Any questions?