Player Positions
A clustering approach

Paul Nguyen

David Herrero-Quevedo

Objective

- Determine NBA Player positions
- Check proportions over time

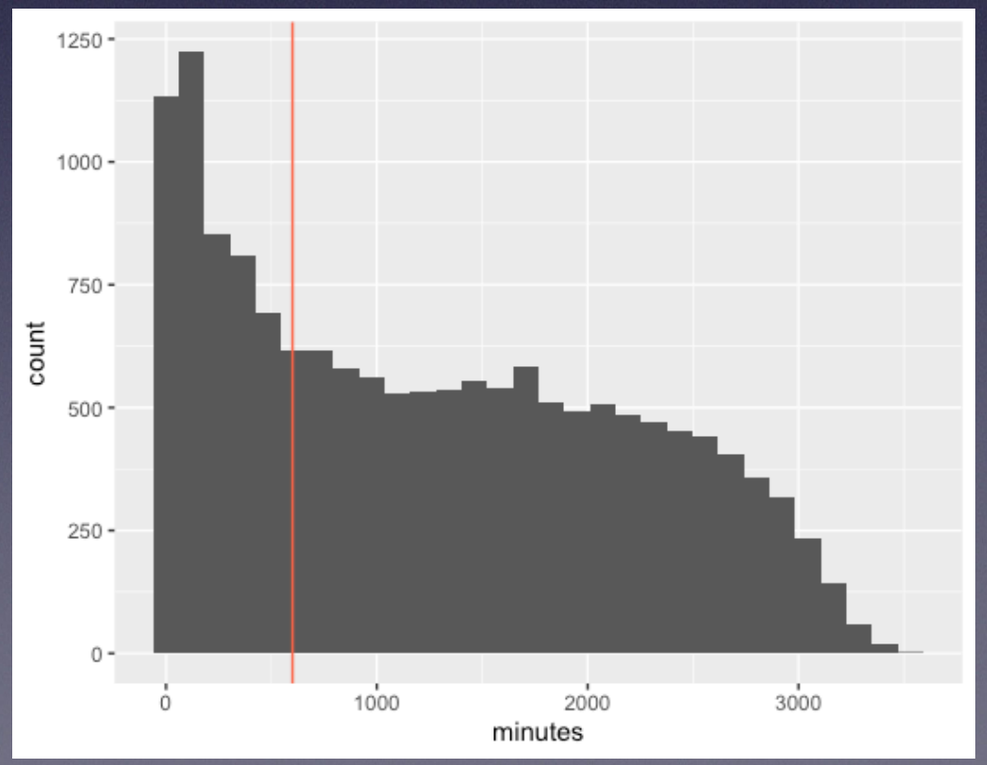
PCA + Clustering

- K-means
- Hierarchical clustering

The data

- Observation: player by season.
  - (n = 10306)
  - E.g. Luka Dončić - 2020
- Seasons 1984 to 2020

- Source: stats.NBA.com and basketball-reference
- Normalized stats by Season.
- Removed players that didn’t play at least 600 minutes per season



Variables

Identifiers
- **namePlayer**: Name of player
- **groupPosition**: Player Position
- **yearSeason**: Season year

Defense
- **pctSTL**: Steals percentage.
- **pctBLK**: Blocks percentage.
- **stlPerGame**: Steals per game.
- **blkPerGame**: Blocks per game.
- **pfPerGame**: Personal Fouls per game.

Efficiency stats
- **ratioPER**: Player Efficiency Rating
- **ratioOWS**: Offensive Win Shares.
- **ratioDWS**: Defensive Win Shares.
- **ratioWS**: Win Shares.
- **ratioWSPer48**: Win Shares per 48 minutes.
- **ratioOBPM**: Offensive Box Plus/Minus.
- **ratioDBPM**: Defensive Box Plus/Minus.
- **ratioBPM**: Box Plus/Minus.
- **ratioVORP**: Value Over Replacement Player.
- **minutesPerGame**: Minutes Per Game
- **pctUSG**: Usage Percentage.

Other
- **countGames**: Games played during season
- **minutes**: Minutes played during season
- **isAllNBA**: Did player make season's all NBA team?

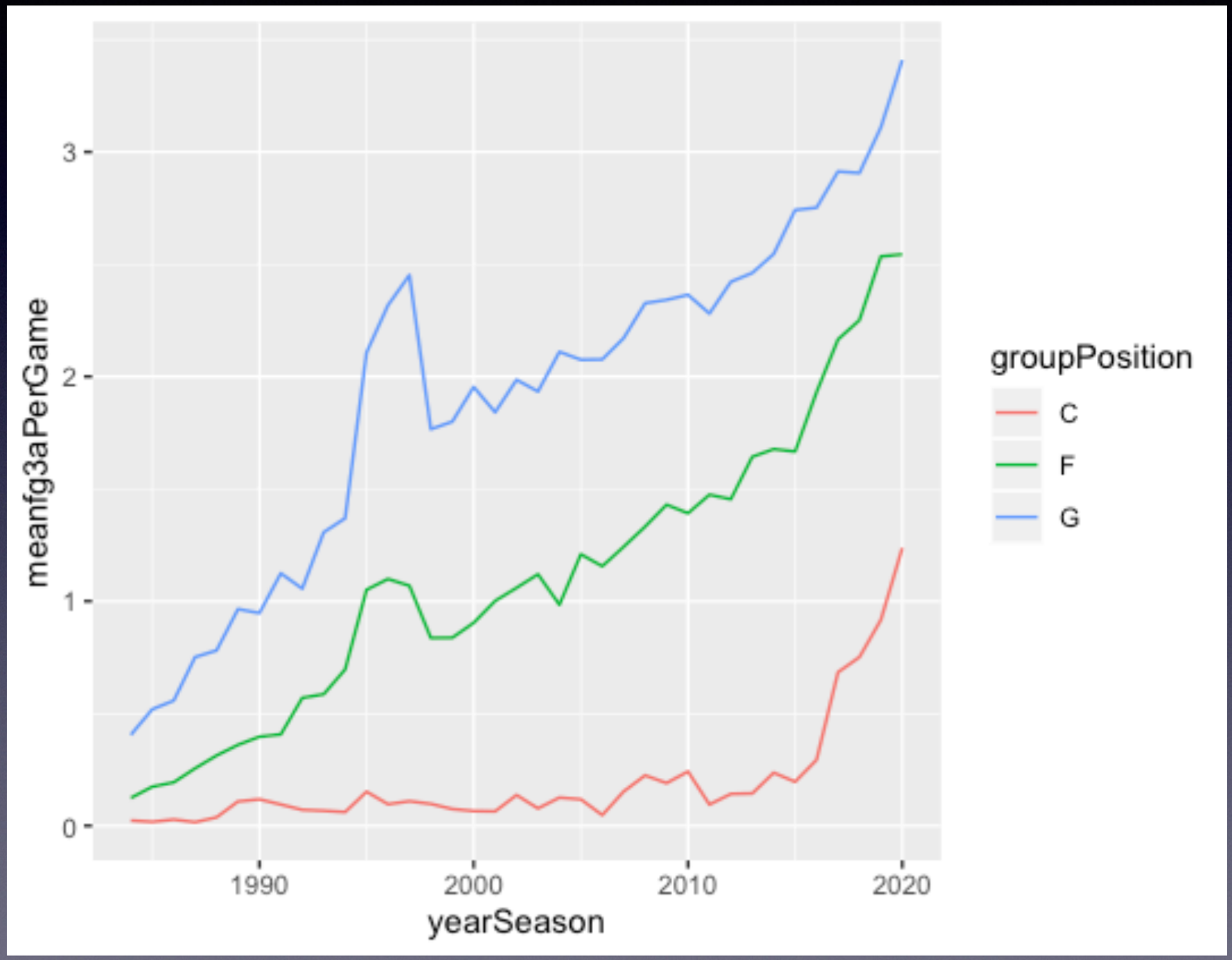
Shooting:
- **pctFTRate**: Percentage of Free Throws.
- **pct3PRate**: Percentage of 3-pt shots.
- **pctFG**: Field Goal Percentage.
- **pctFG3**: Field Goal 3-pt Percentage.
- **pctFG2**: Field Goal 2-pt Percentage.
- **pctEFG**: Effective Field Goal Percentage.
- **pctFT**: Free Throw Percentage.
- **fgmPerGame**: Field Goal Made per Game.
- **fgaPerGame**: Field Goal Attempts per game
- **fg3mPerGame**: Field Goal Made (3-pt) per game.
- **fg3aPerGame**: Field Goal Attempts (3-pt) per game.
- **fg2mPerGame**: Field Goal Made (2-pt) per game.
- **fg2aPerGame**: Field Goal Attempts (2-pt) per game.
- **ftmPerGame**: Free Throws Made per game.
- **ftaPerGame**: Free Throw Attempts per game.
- **ptsPerGame**: Points per game.

Rebounds
- **pctORB**: Offensive Rebound Percentage
- **pctTRB**: Total Rebound Percentage
- **pctDRB**: Defensive Rebound Percentage
- **orbPerGame**: Offensive Rebounds per game.
- **drbPerGame**: Defensive Rebounds per game.
- **trbPerGame**: Total Rebounds per game.

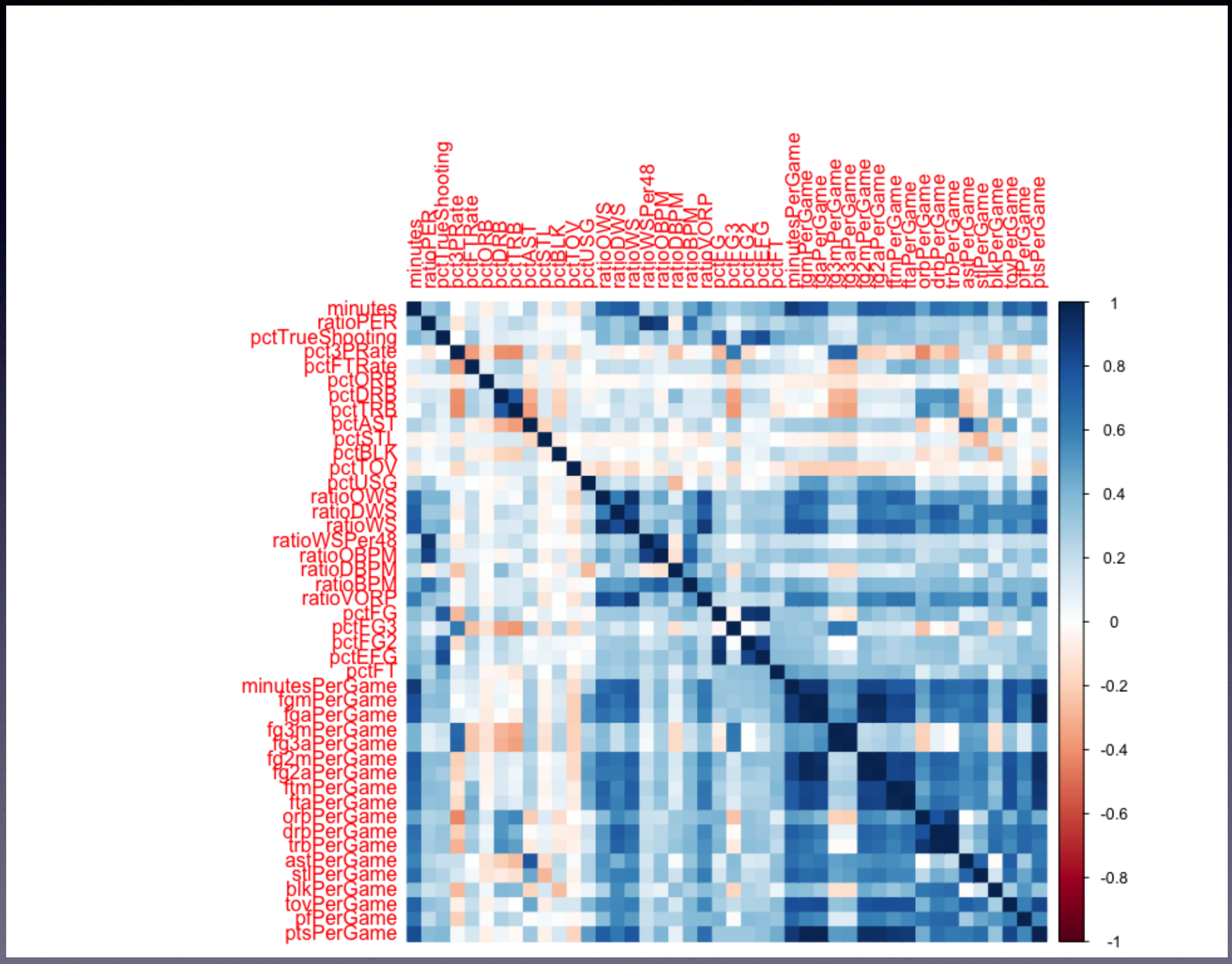
Passing
- **pctAST**: Assists Percentage.
- **pctTOV**: Turnover Percentage.
- **astPerGame**: Assists per game.
- **tovPerGame**: Turnovers per game.

Exploratory Data Analysis

FG3A per Game over time

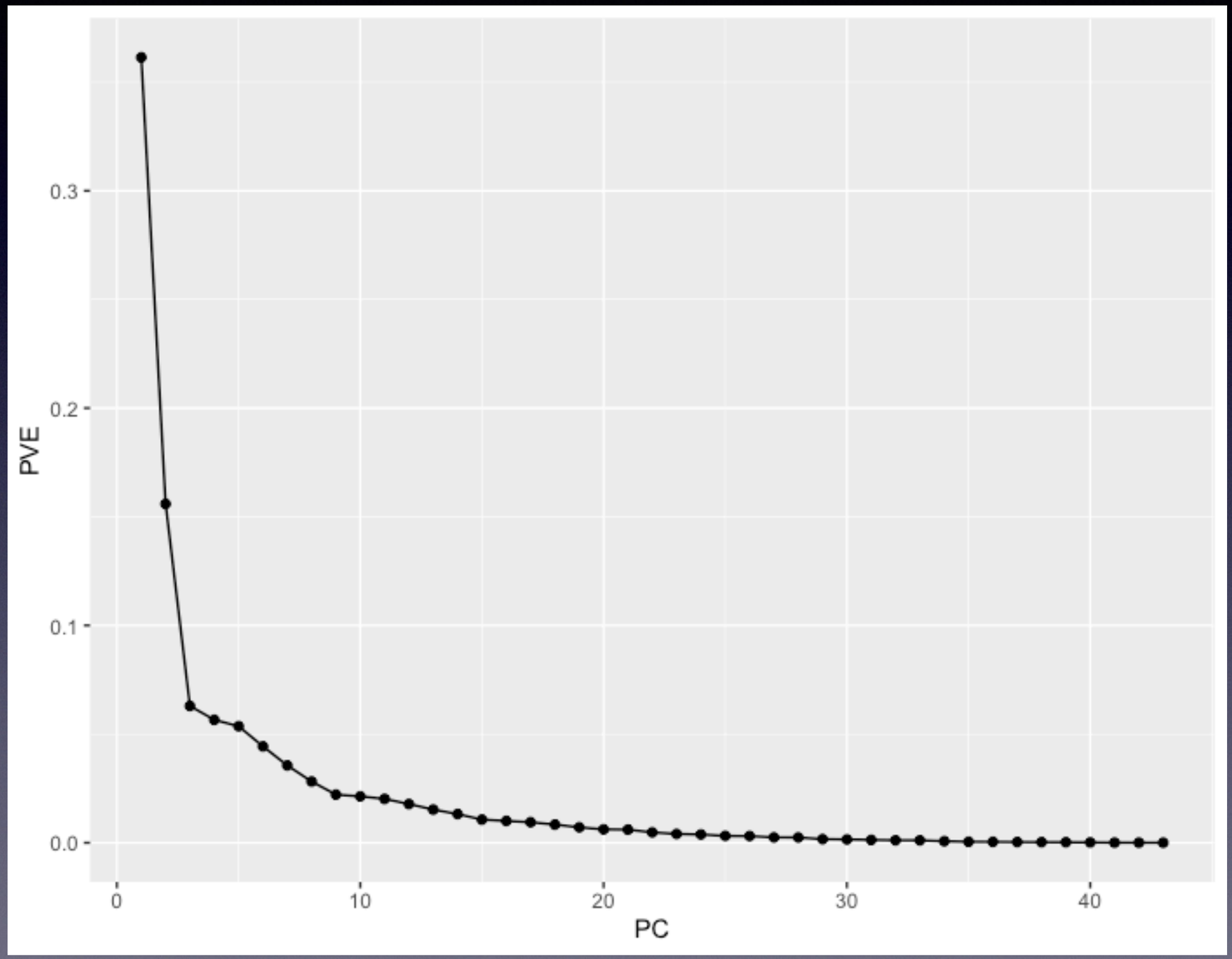


Correlation Matrix

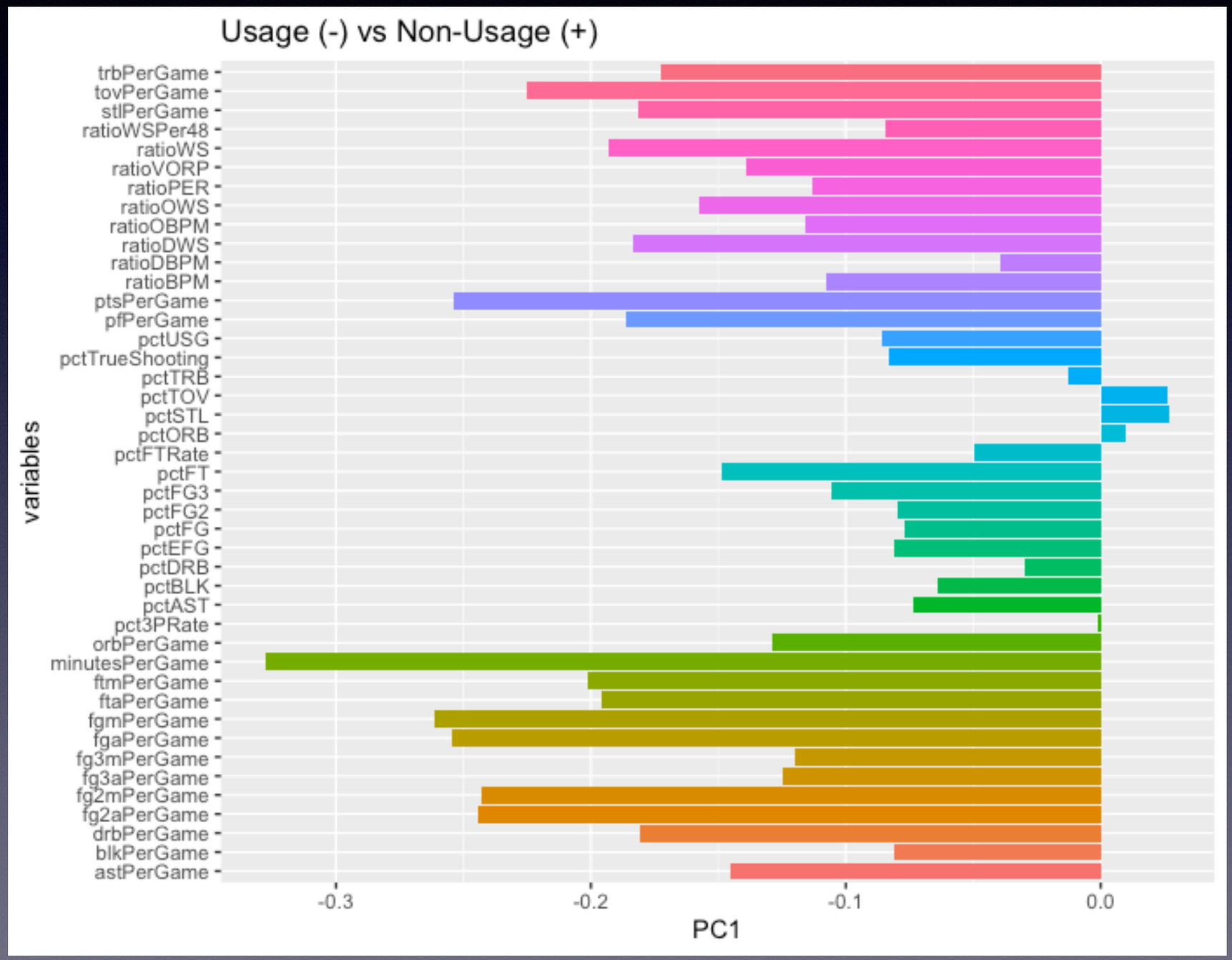


Dimension Reduction

Scree Plot



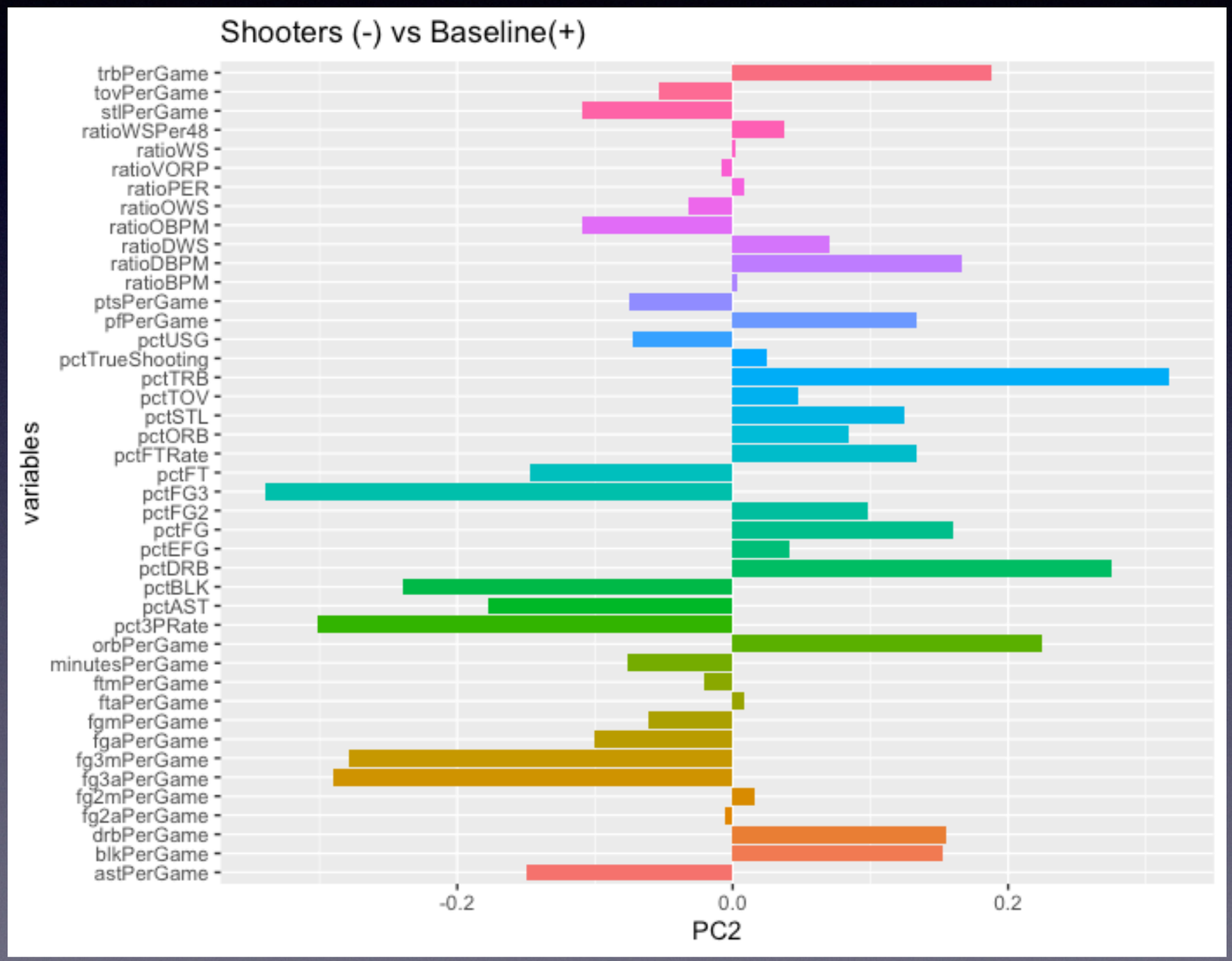
PC1



Usage (-):

- Total shooting
- Turnovers

PC2



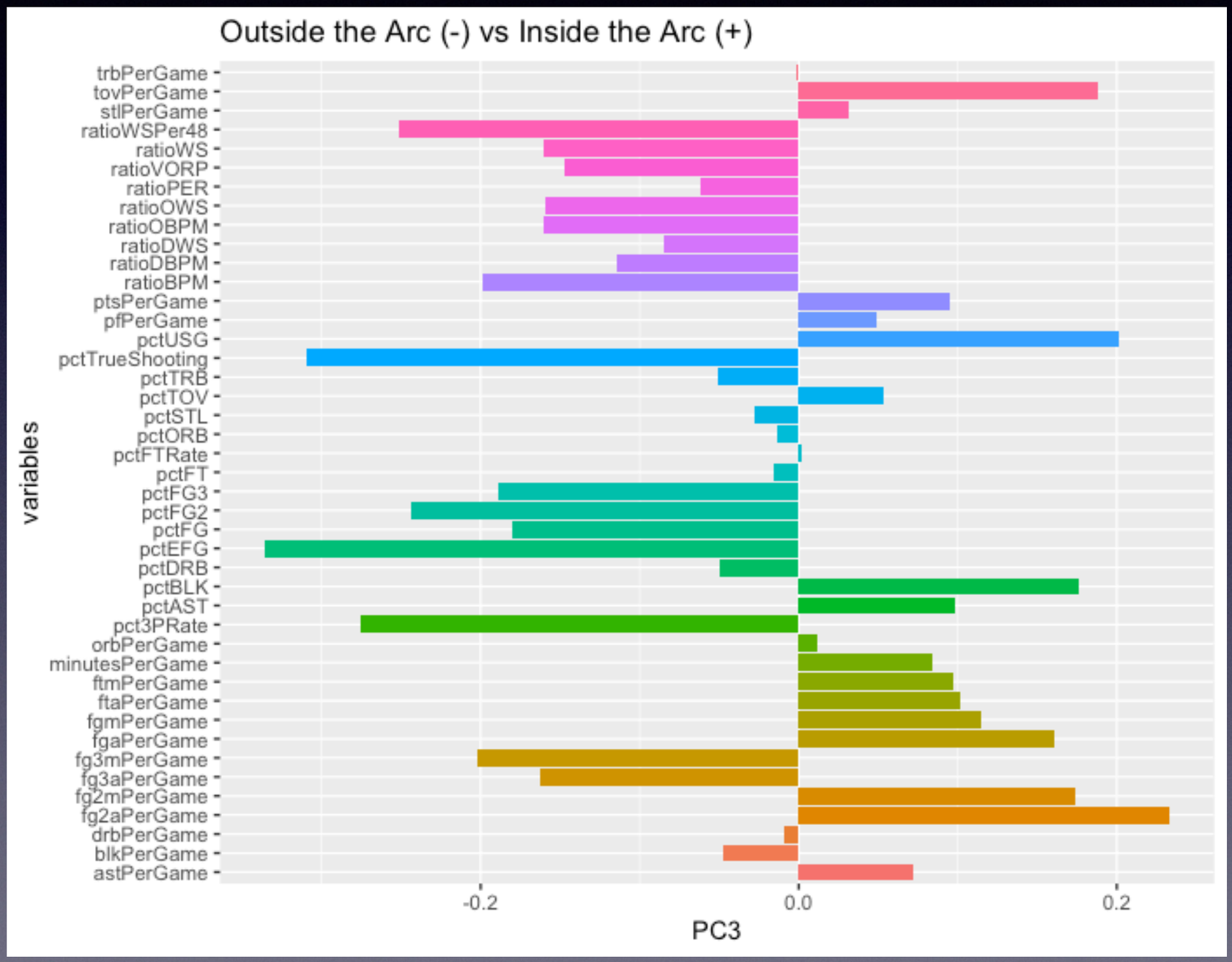
Shooters(-):

- 3-pt shooting
- Passing

Baseline players (+)

- Rebounding
- Defensive

PC3

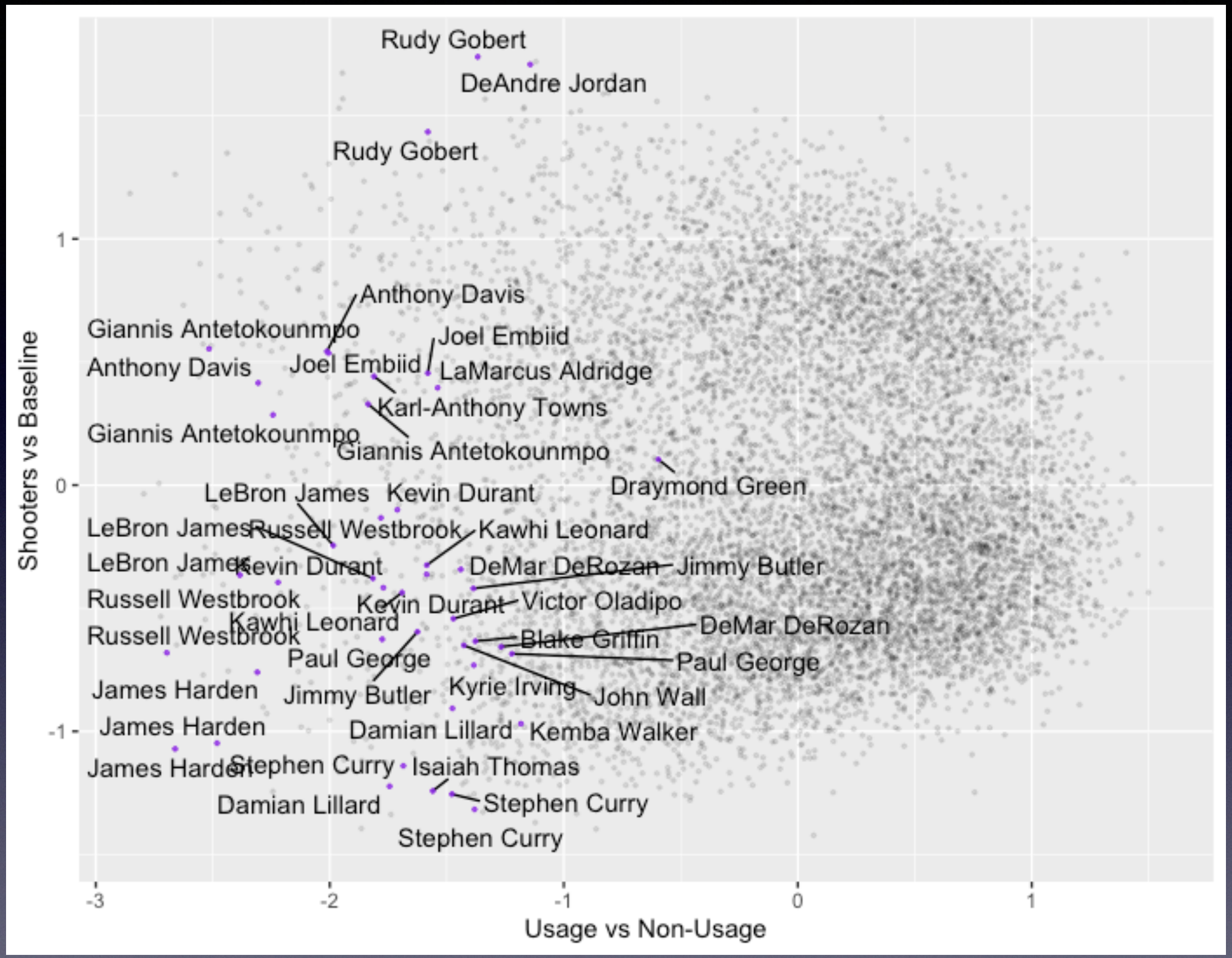


Outside the arc(-):

- Efficiency stats
- 3-pts ratio

Inside the arc(+)

- Passing
- 2-pts ratio

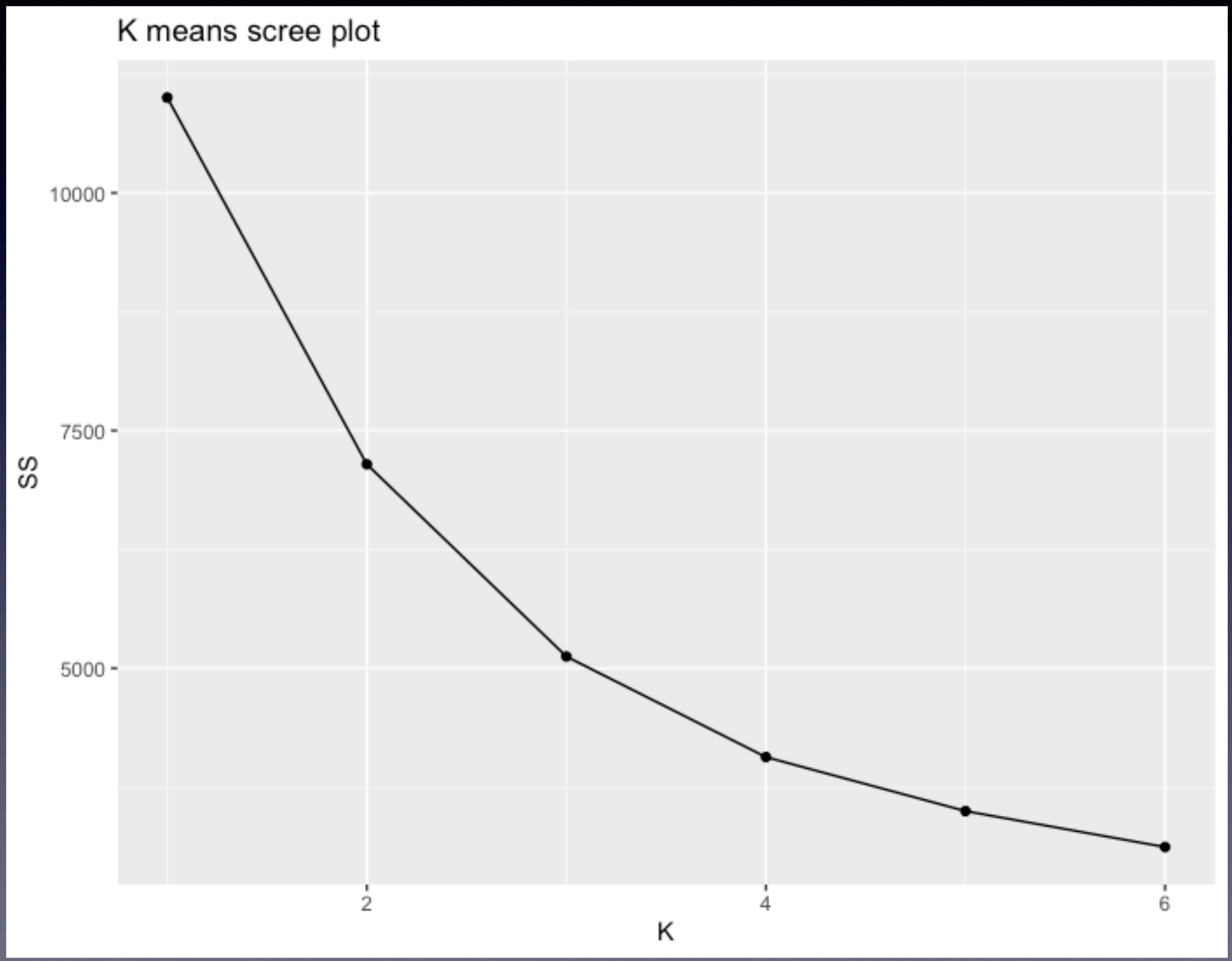


2018-2019
All-NBA players

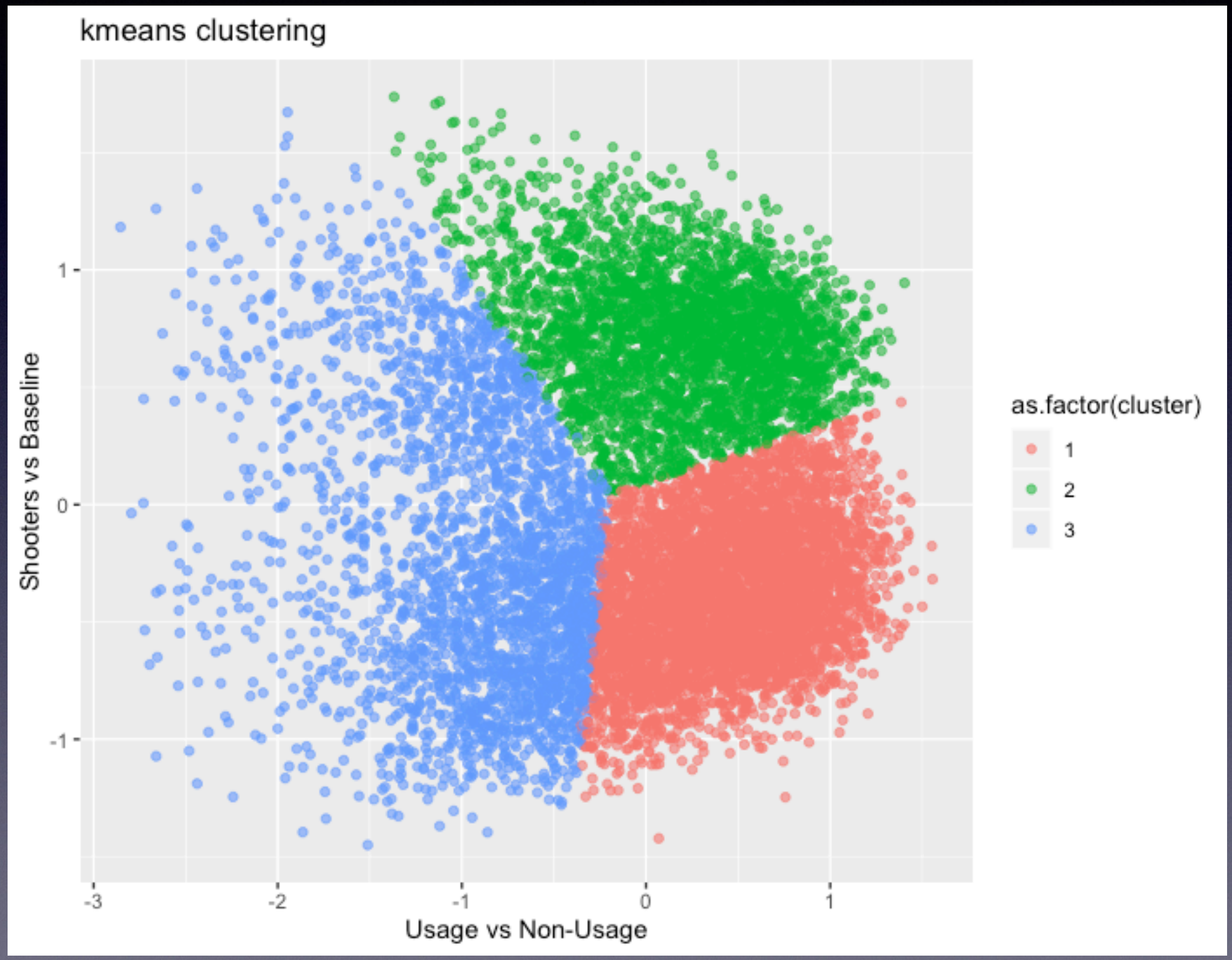
Mostly shooters
with high usage

Clustering (Whole NBA)

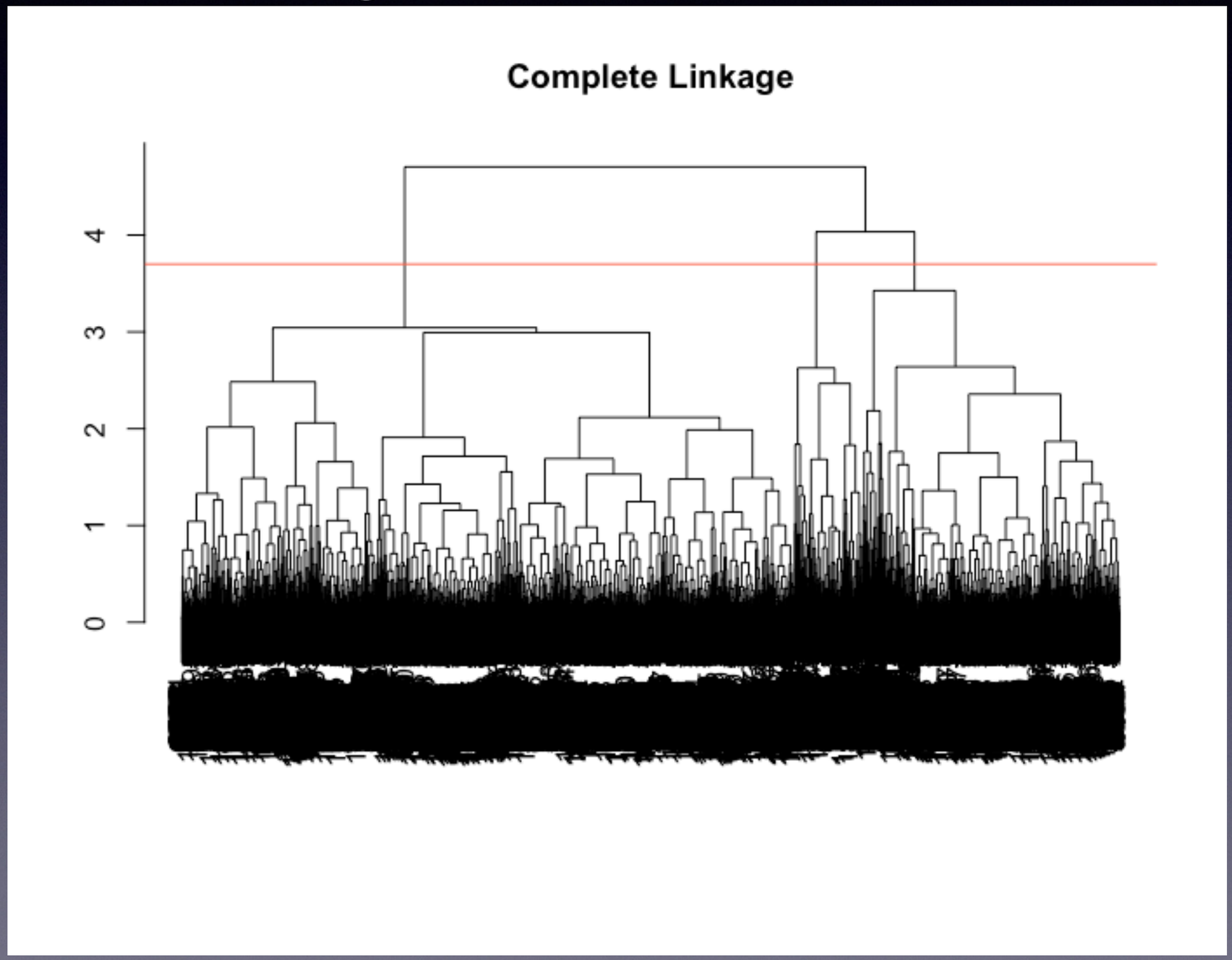
K-means Scree Plot



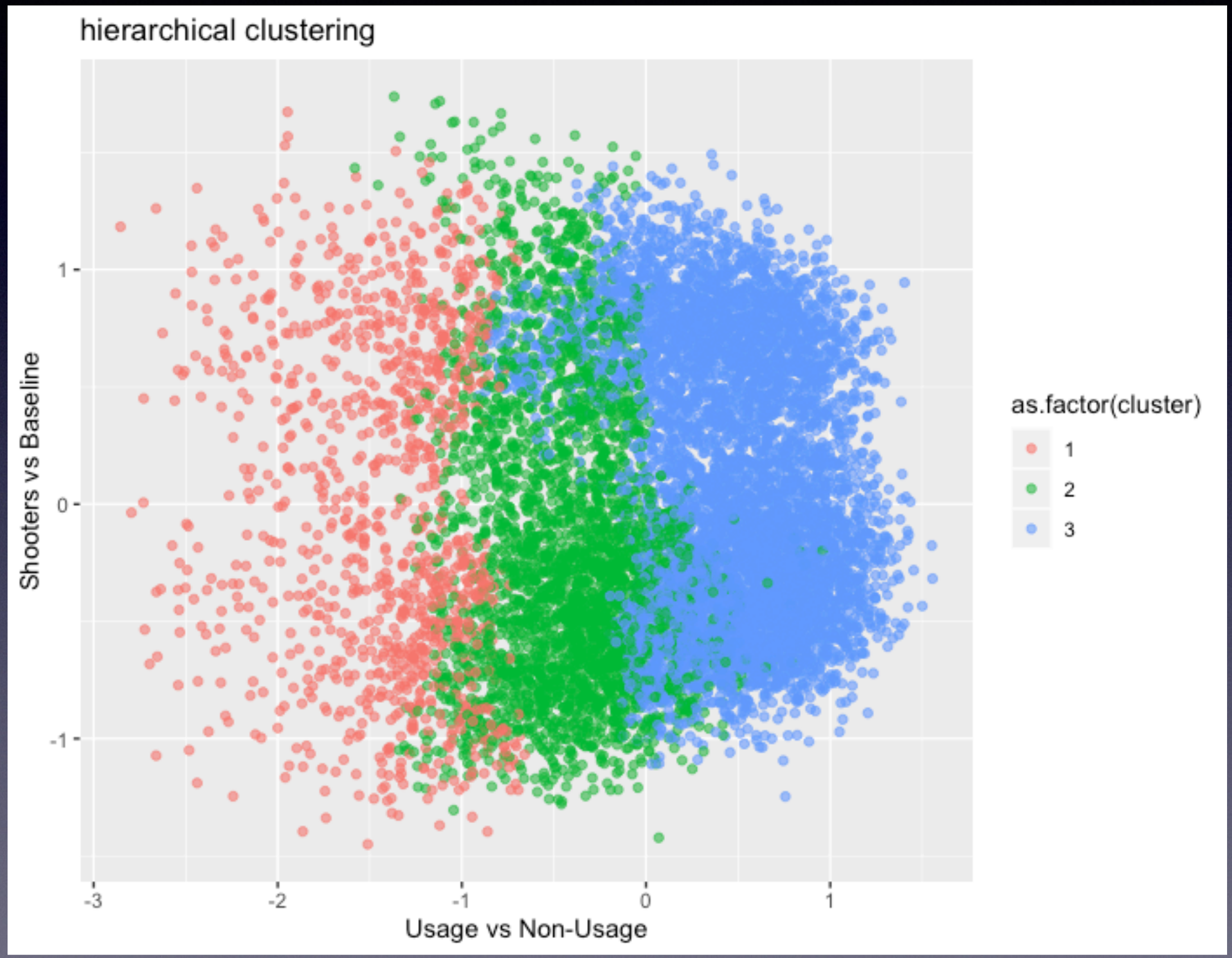
K-means Clusters



Hierarchical Clustering Dendrogram



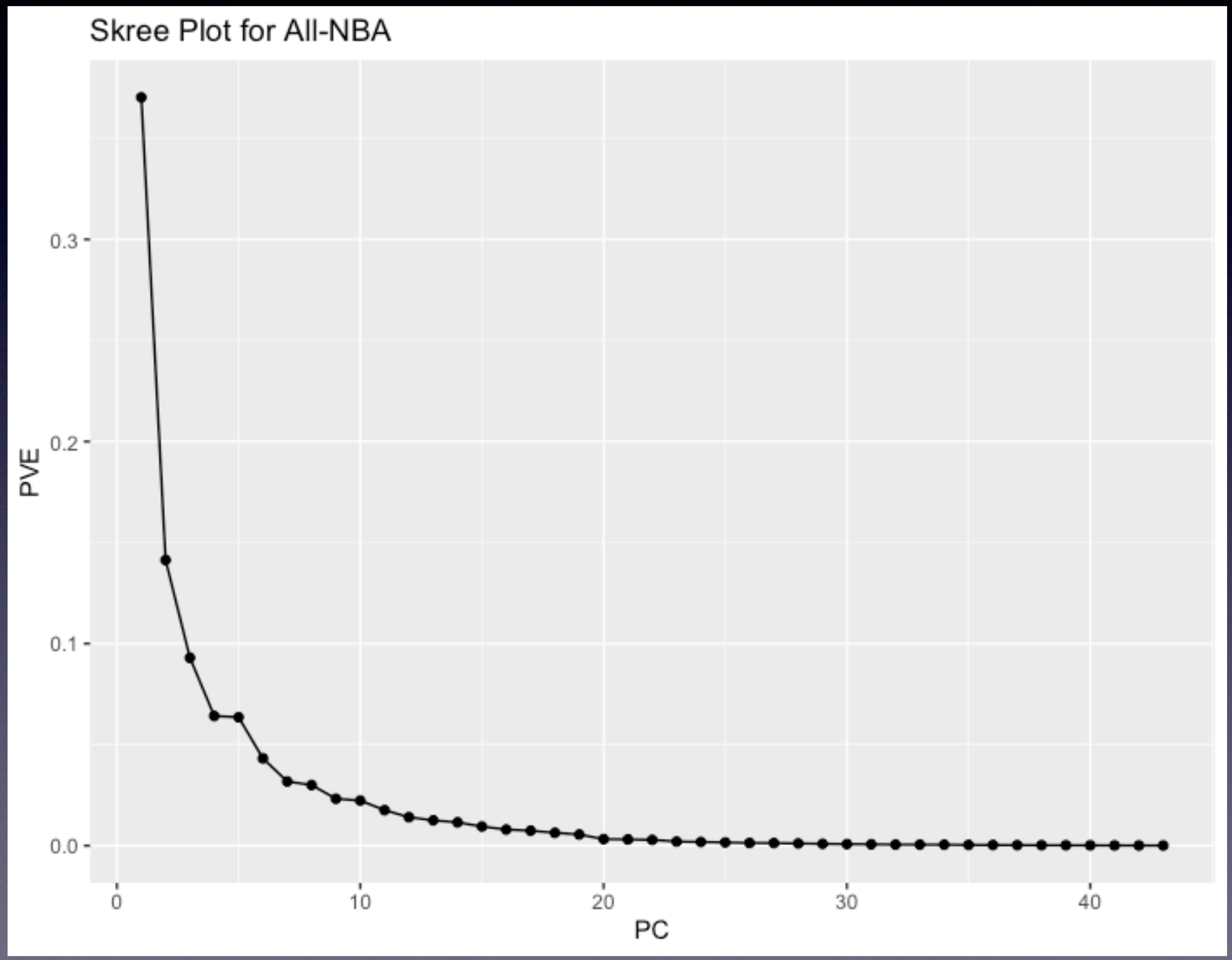
Clusters (hierarchical)



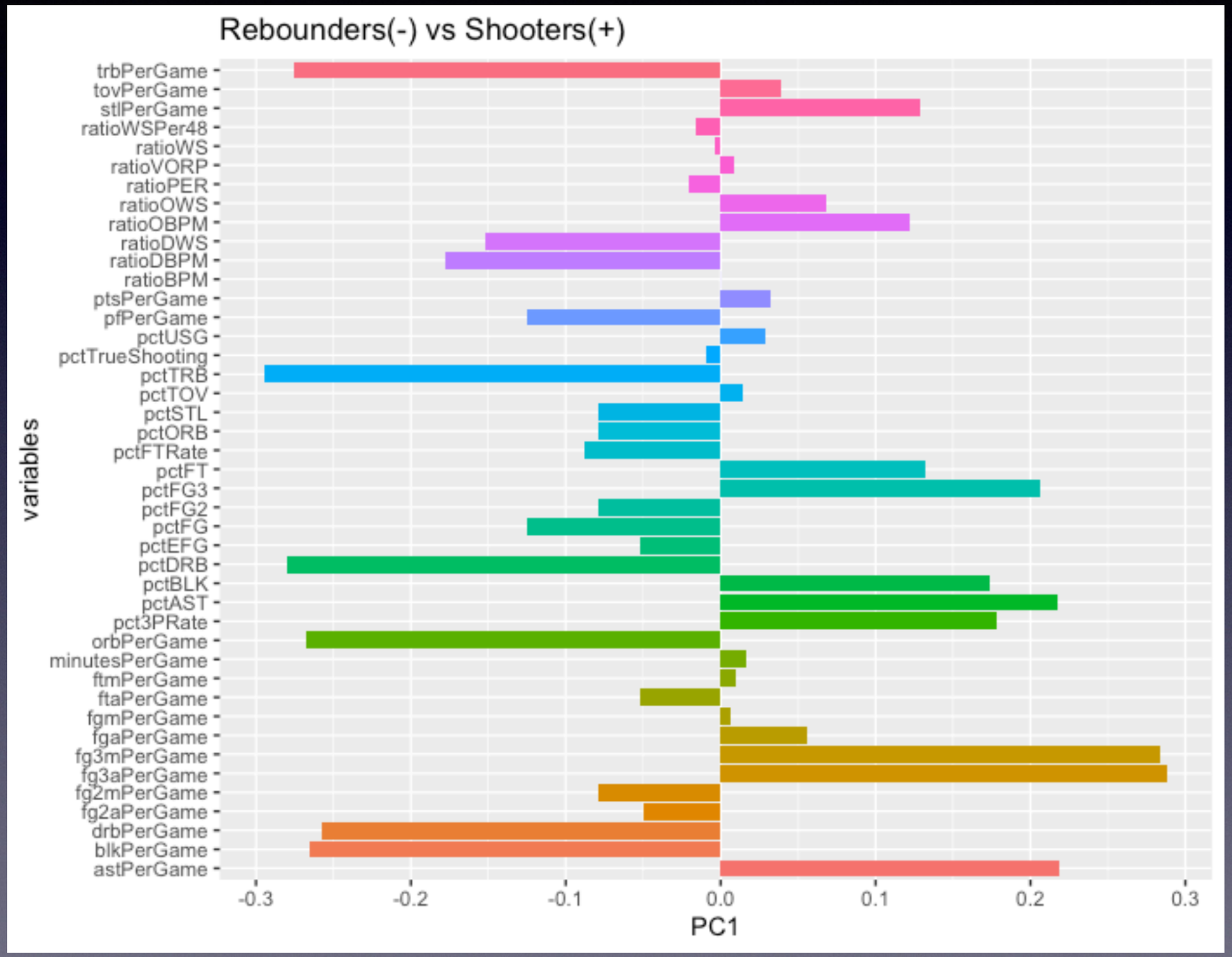
Cluster Proportions over Time (whole NBA)

All-NBA dimension reduction

All-NBA Scree Plot



All-NBA PC1



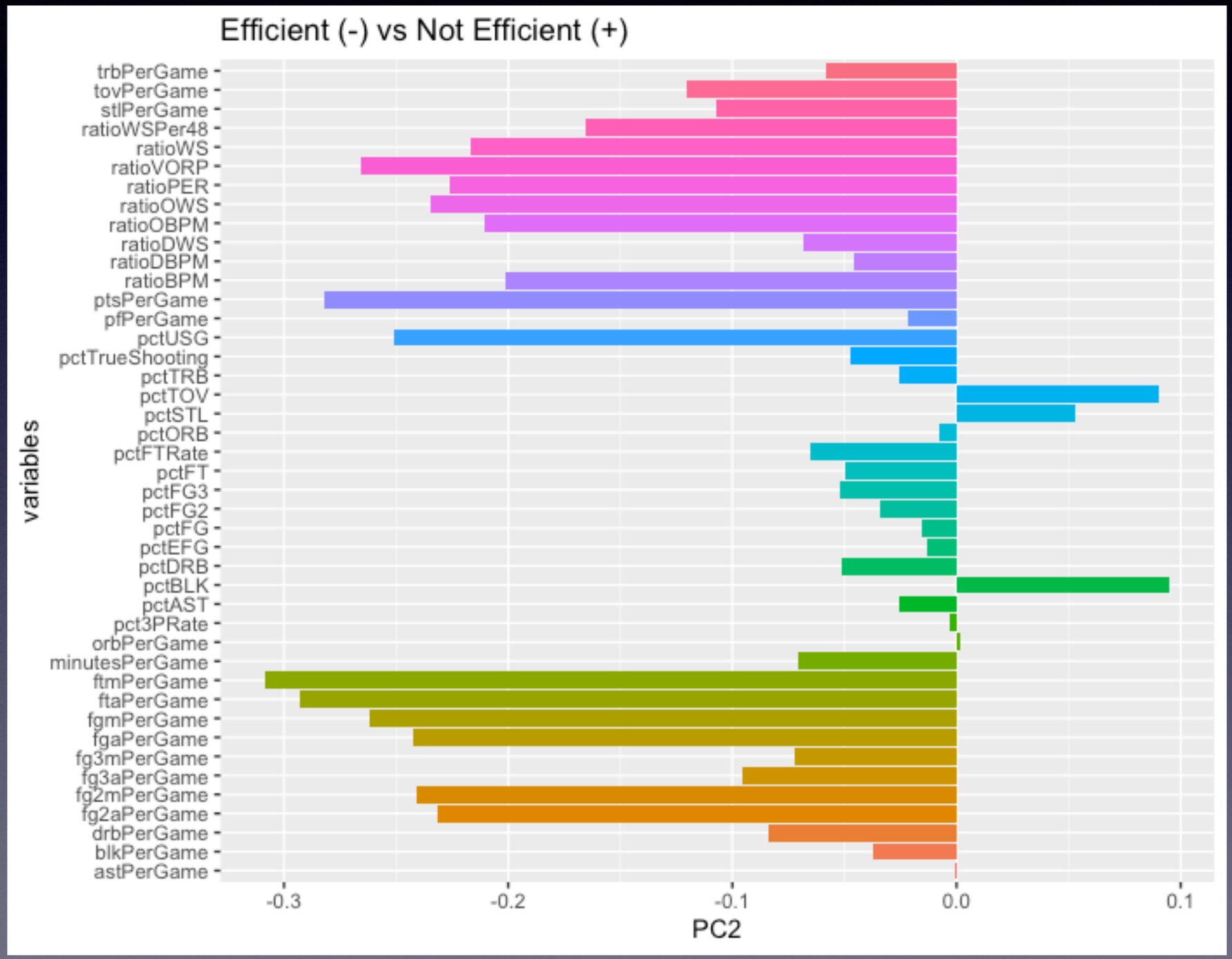
Rebounders(-):

- Rebounding
- Defensive

Shooters(+)

- Shooting
- Passing

All-NBA PC2



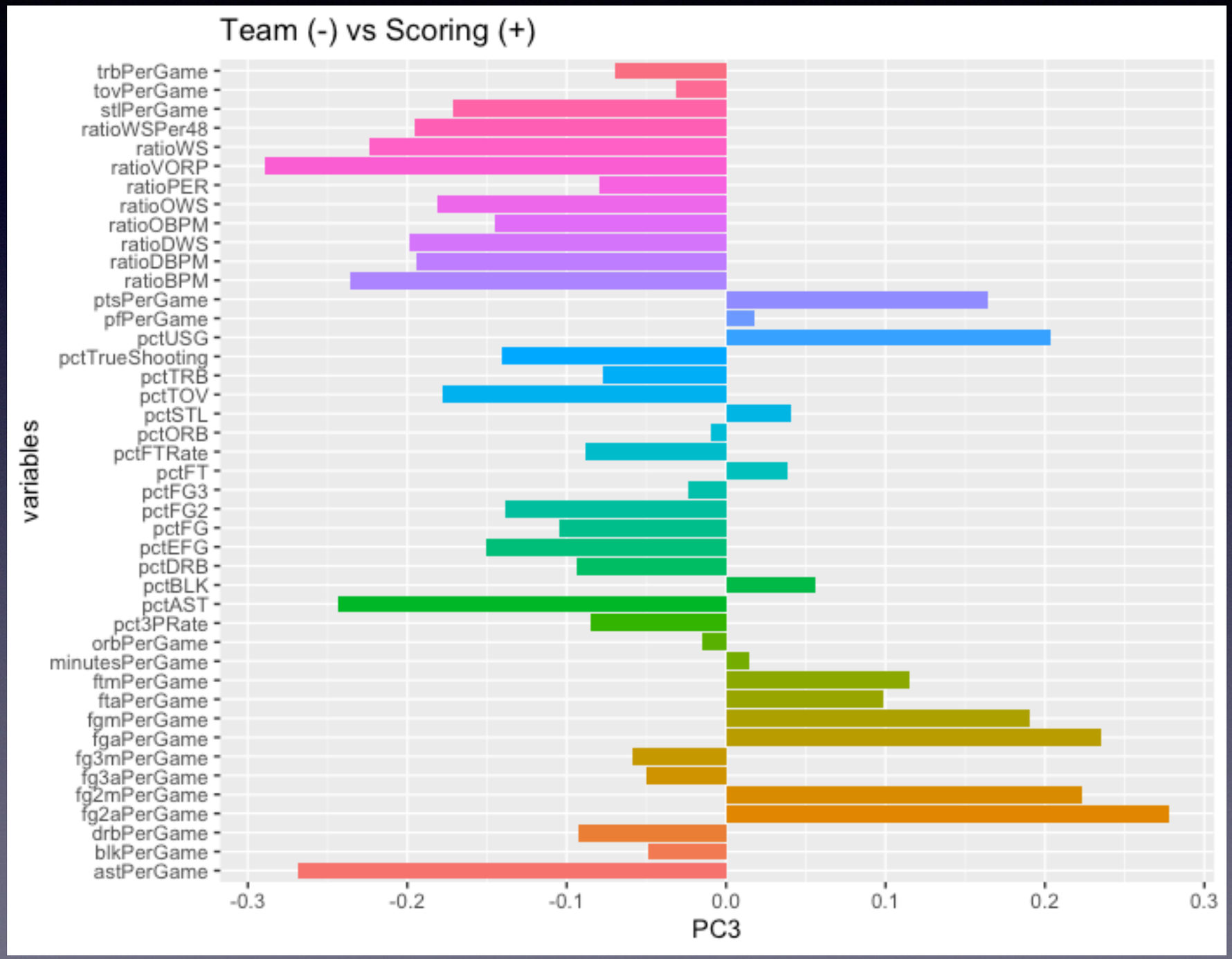
Efficient(-):

- Scoring
- Efficiency

Not efficient(+)

- Turnovers
- Rebounding

All-NBA PC3



Team(-):

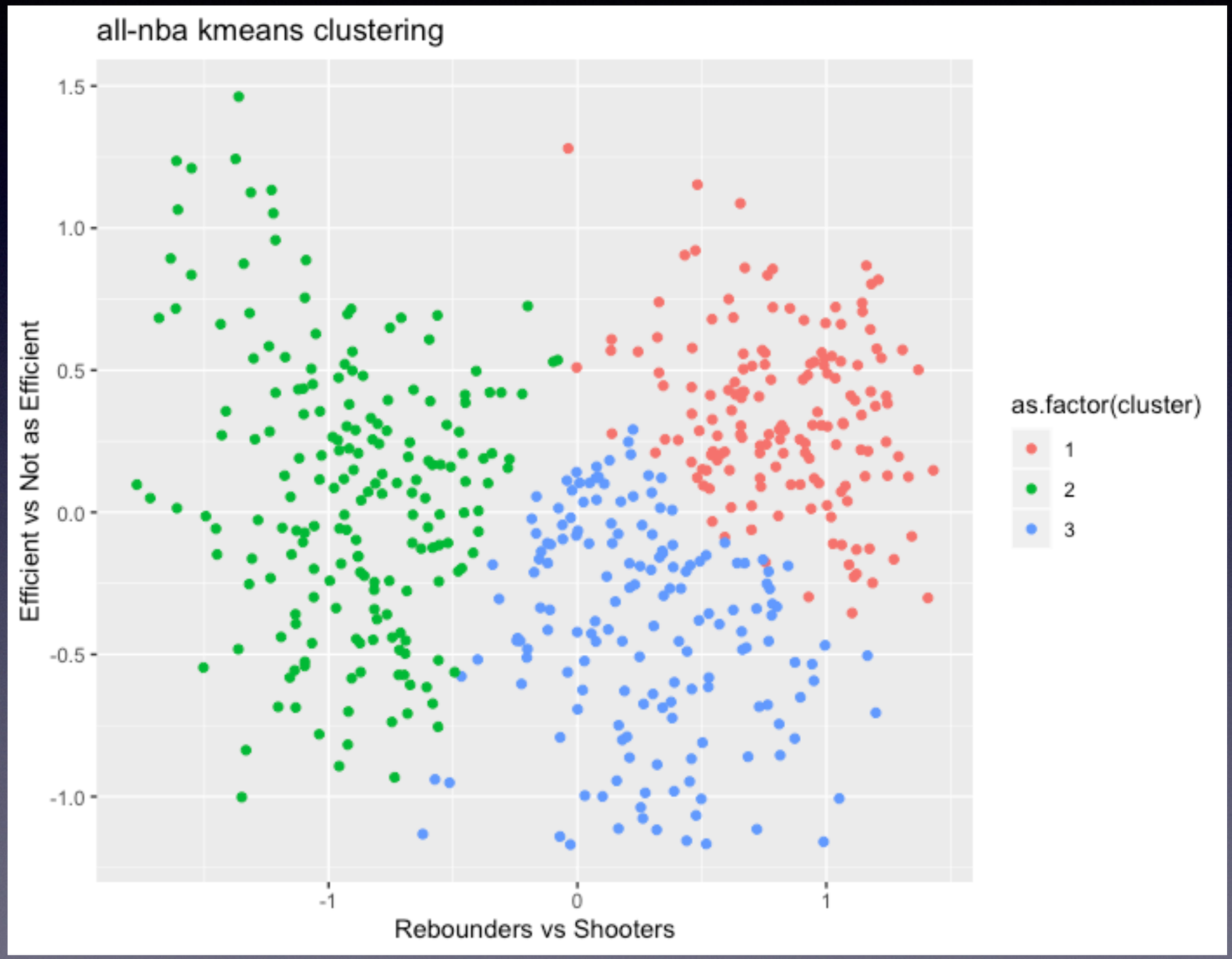
- Passing
- Rebounding

Scoring(+)

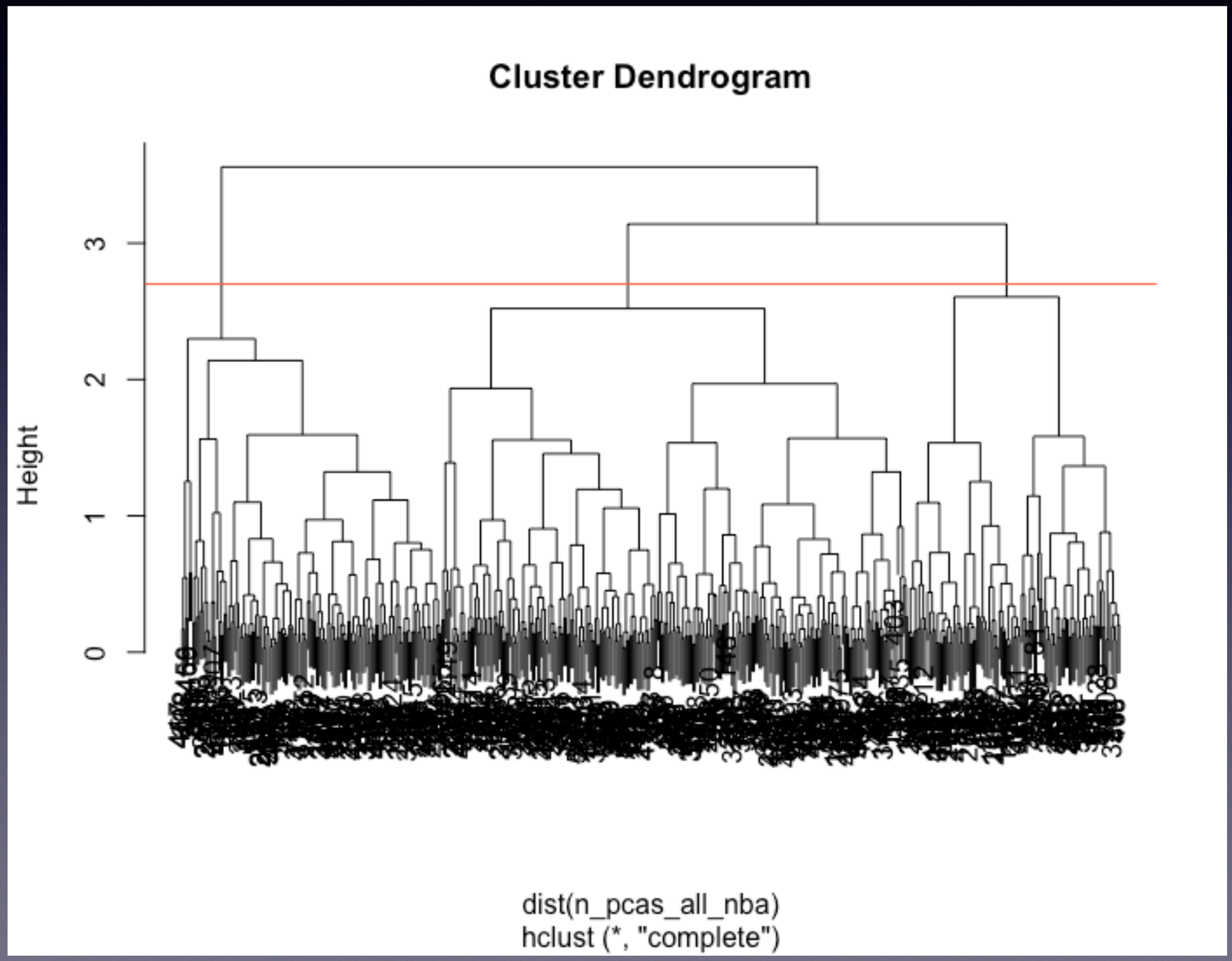
- Shooting
- Scoring

All NBA clustering

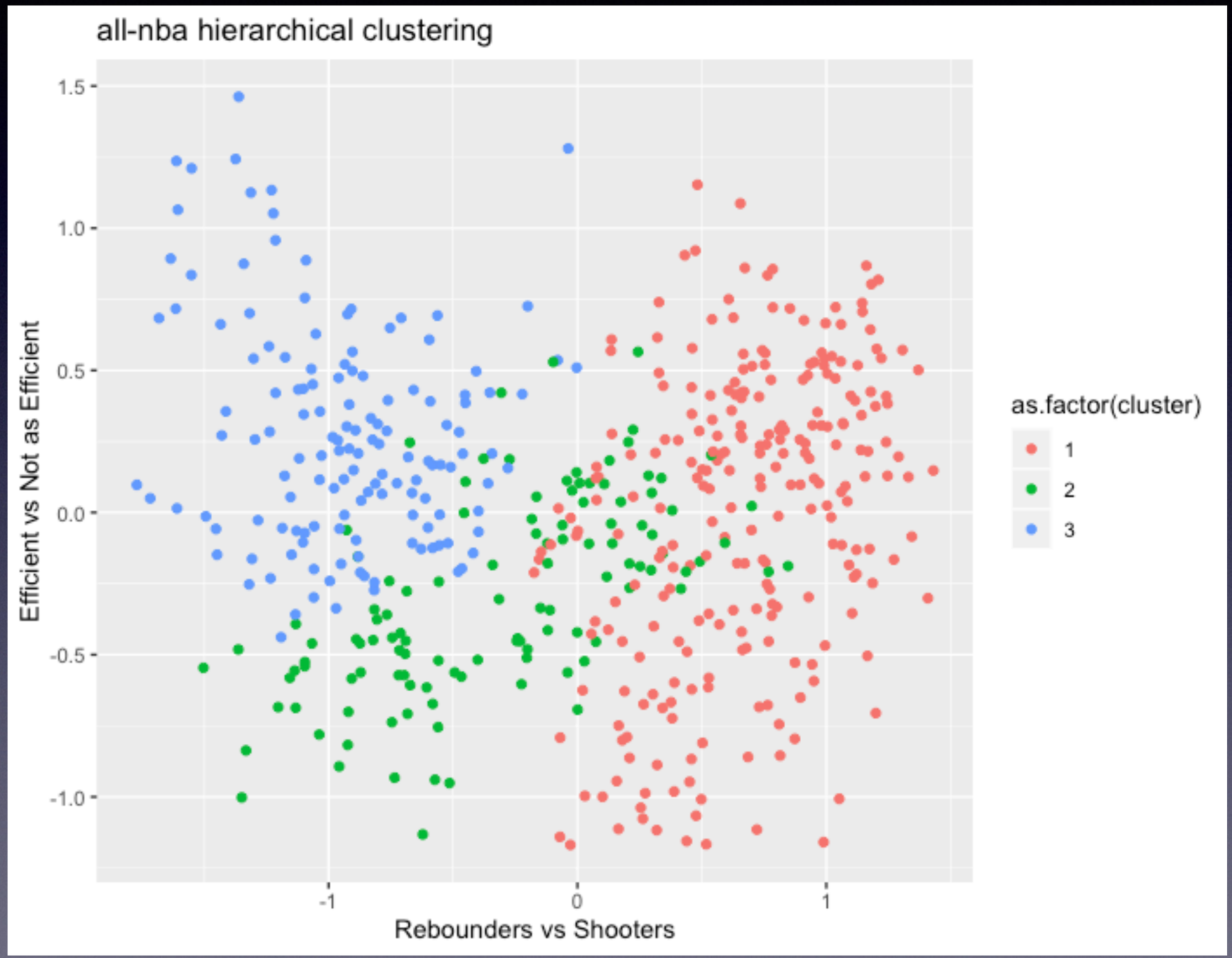
All-NBA K-means Clustering



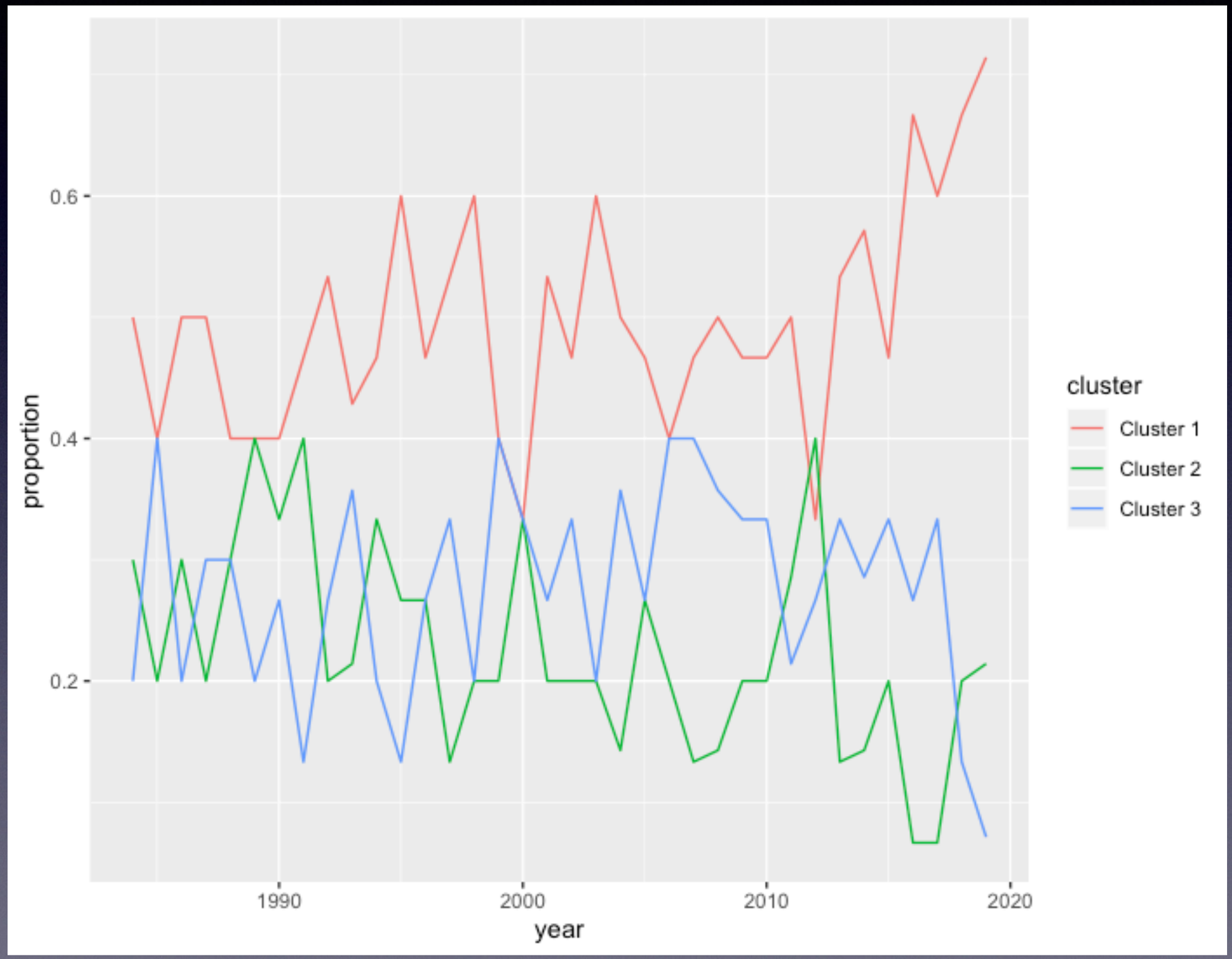
All-NBA Hierarchical Dendrogram



All-NBA Hierarchical Clustering



All-NBA Clusters over Time



Discussion

Takeaways:

- 3 types of players:
  - low usage shooters (Cluster 1)
  - high usage players, spread for baseline vs shooters (Cluster 2)
  - lower usage baseline players (Cluster 3)
- Cluster 1 remains the largest proportion in each season

Discussion Continued

- Looking at PCA for All-NBA players specifically:
  - Cluster 1: Shooters, mixed efficiency, generally Scorers. Ex: Damian Lillard, Stephen Curry, Lebron James
  - Cluster 2: Pretty diverse set of rebounders and Shooters. Shooters tend to be team players while Rebounders are Scorers. Players are efficient. Ex: Giannis Antetokounmpo
  - Cluster 3: Team Rebounders, however, not very efficient. Ex: Draymond Green, Joel Embiid
- Upward trend for Cluster 1, especially post 2011

Further Research

- Look into different combinations of clusters, clustering methods, and principle components
- In-depth dive into one specific team.
- How does team composition affect success?

Questions?