

Reclassifying Relief Pitchers

Dylan McGee

Department of Sport Analytics, Falk College of Sport and Human Dynamics, Syracuse University

Abbreviated abstract: Since the creation of baseball, many rules set in place have stayed consistent with no scientific evidence supporting the reasoning behind them. One major rule that needs to be questioned is the utilization of relief pitchers. Relief Pitchers have been classified under the following “roles:” *Closers, Setup Men, Middle Relievers, Long Relievers, Left-Handed Specialists, and Openers*. They are expected to perform in those situations. But what makes a “Closer?” By using K-Means Cluster Analysis, it is possible to reclassify relievers in ways they have never been utilized before.

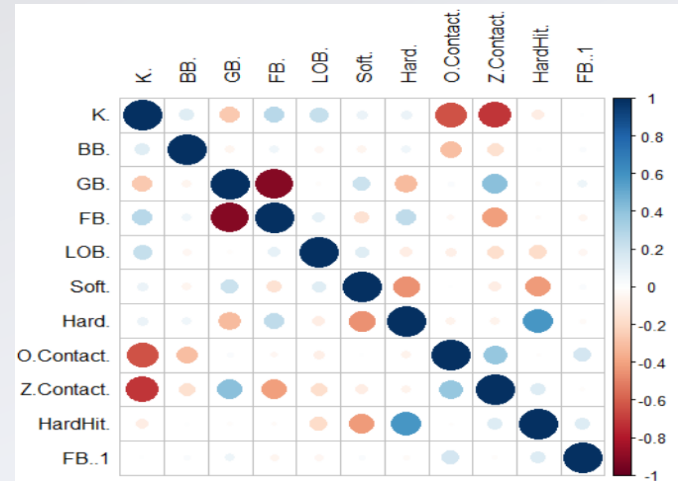
Related publications:

– Tan, P.N., Michael Steinbach, and Vipin Kumar. “Cluster Analysis: Basic Concepts and Algorithms” (487-555), (2005)

Problem, Data, Previous Works

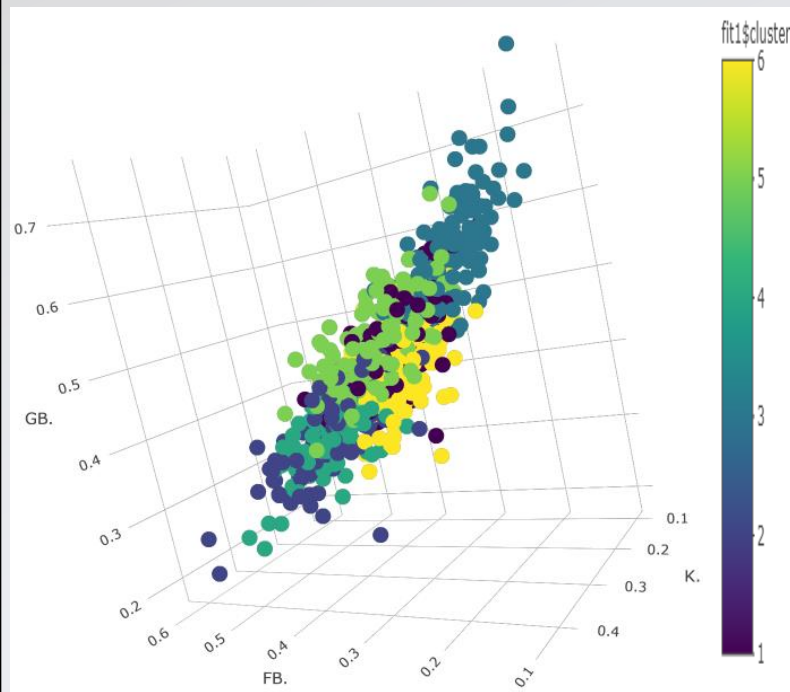
- FanGraphs Data
 - Relief pitchers since post-steroid era (2006-present)
 - Minimum of 20 innings pitched
- Exported into Excel
 - Turned percentages into decimals
- Imported into RStudio for correlation/cluster analysis

Statistic	Definition
K% (Strike Out %)	How often a pitcher gets strikeouts per plate appearance.
BB% (Walk %)	How often a pitcher gets walks per plate appearance.
GB% (Ground Ball %)	How often a pitcher gets ground balls per ball in play.
FB% (Fly Ball %)	How often a pitcher gets flyballs per ball in play.
LOB% (Left on Base %)	% of base runners that a pitcher strands on base.
Soft% (Soft Contact %)	% of soft-hit batted balls.
Hard% (Hard Contact %)	% of hard-hit batted balls.
O.Contact% (Outside of Zone Contact %)	How often contact is made outside of zone per swing outside of zone.
Z.Contact% (Inside of Zone Contact %)	How often contact is made inside of zone per swing inside of zone.
Hard Hit % (Hard Hit %)	% of hard-hit batted balls resulting the batter on-base.
FB1% (Fastball %)	% of fastballs compared to off-speed and breaking balls.









The correlation didn't show too many strong relationships. This is how I wanted to construct my data, that way each metric represented would tell its own story for each reliever. The only metrics kept with strong relationships were ones I felt necessary to keep for accurate conclusions once the clusters were created.

Methods



- 3D plot with K%, FB%, and GB%
- Formed compact clusters that showed unique patterns within their statistics.
- We can classify each cluster and use them in situations that suit their strengths.
- For example, Cluster 3 (top right of the graph)
 - high GB%, low FB% and a slightly below-average K%
 - Groundball specialists of the data.
- Cluster 6 sits about in the middle of all three metrics.
 - The worst performing cluster with no strengths in anything.
 - Can be used in blowout games or if you need to rest other pitchers on your staff. This is also a great cluster to throw debuting pitchers.

Results and Conclusions

Cluster	Role	Usage	Notable Players	
1	Off-speed Specialists	Rely on off-speed to produce softer than average contact no matter the situation. Very reliable group. Use in almost any situation necessary.	Mark Melancon Sergio Romo Shane Greene	
2	Velo Specialists	By far highest velo average and usage rate. Excel in K's but struggle with command at times. Use when looking for strikeouts or big velo change.	Aroldis Chapman Josh Hader Mychal Givens	
3	Groundball Specialists	Highest GB%, produce weakest contact but fewest K's. Use when trying to limit long ball, keep the ball on the ground, or looking for double play.	Zack Britton Emmanuel Clase Brusdar Graterol	
4	Clutch Arms	Excel at getting outs with RISP. Above average in most categories. Use when in a close game situation, especially with RISP.	Roberto Osuna Nick Anderson Kirby Yates	
5	Strikeout Specialists	Have great off-speed which is used to get batters to chase outside of zone. Struggle with command occasionally but should be used when looking to limit contact at all costs.	Andrew Miller Dellin Betances Diego Castillo Adam Ottavino	
6	Under-developed	Worst group. Have the poorest numbers in most categories and don't excel in anything. Use when out of options/up or down by a lot.	Robert Gsellman Ty Blach Michel Baez	

There are 6 unique groups of pitchers who can be used in various situations throughout the course of a game. From here, we can use baseball simulation techniques to manage a team with our new method and compare the results with standard utilization of relievers. **By using K-Means Cluster Analysis, we can find insights on how to utilize pitching staffs more effectively ultimately bringing in more wins for your team.** In the future, I plan on testing my cluster's in real game situations and reconstruct reliever's salary based on their cluster-importance.