

## Methodology

We used a two-phase approach to answer the question of what attributes are conducive to success for pitchers in specific roles. Our first step was extensive feature engineering to design metrics that attempt to measure the change in a player's performance over the course of their outing, if any. By understanding the trajectory of a player's outing, we can better hypothesize whether they are suited to starting pitching, relief pitching, or closing and better tailor the length of their involvement to their skill-set. Our second step was dimensionality reduction via Principal Component Analysis and clustering to identify players who were outliers in their current role.

We were interested in the correlation between measures of time in a game and granular pitcher performance. Our specific measures of pitcher performance are spin rate and velocity at the time of the pitch release. For each pitcher, we analyzed their four-seam fastball and their most thrown alternative pitch (secondary pitch). Their change over the course of their outings is calculated by considering the correlation between pitches thrown and fastball velocity and spin rate, as well as the correlation between pitches thrown and secondary pitch velocity and spin rate. We also calculated the correlation between batters faced and pitcher performance. In this way, we attempted to capture a linear relationship between pitchers time in a game and performance.

Furthermore, we hypothesized that effects on pitchers may not be linear, and their performance could decline suddenly. Therefore, as a further measure of pitcher change, we extracted the pitch at which a pitcher threw their second pitch which was a standard deviation or more below their mean velocity. This approach was applied to fastball velocity and spin rate, as well as secondary pitch velocity and spin rate. In this way, we added a second robust measure of pitcher fatigue and/or change in effectiveness during an outing.

## PCA

Variable	PC1	PC2
strike_percent	0.226809	0.359679
whip	-0.1089	-0.37583
stuff_plus	0.264253	0.409084
mean_fb_velo	0.185154	0.316252
mean_fb_spin	0.147689	0.341954
release_speed	0.037635	0.000407
release_spin_rate	0.096325	0.268365
corr_pitch_fb_release_speed	-0.17061	0.169979
corr_pitch_fb_release_spin_rate	-0.10381	0.08029
corr_pitch_sec_release_speed	-0.10376	0.111453
corr_pitch_sec_release_spin_rate	-0.06351	0.166619
corr_at_bat_fb_release_speed	0.1593	0.082687
corr_at_bat_fb_release_spin_rate	0.052659	0.003598
corr_at_bat_sec_release_speed	0.211299	0.110776
corr_at_bat_sec_release_spin_rate	0.13276	-0.01348
fb_release_speed_threshold_drop	-0.40654	0.207647
fb_release_spin_rate_threshold_drop	-0.39996	0.219795
sec_release_speed_threshold_drop	-0.40847	0.193365
sec_release_spin_rate_threshold_drop	-0.4036	0.222307

Table 1

Principal component analysis was used to reduce the 19 dimensions of the data to 2. As high-dimensional clustering often fails due to the curse of dimensionality, we opted to use PCA to form two uncorrelated dimensions for clustering. The first dimension was heavily influenced downwards by the discretionary thresholds in effectiveness -- therefore, starters tend to cluster to a lower point along the first dimension. The second dimension sees high influence from non-dynamic statistics such as Fangraphs' Stuff+ indicator.

## K-Means Clustering

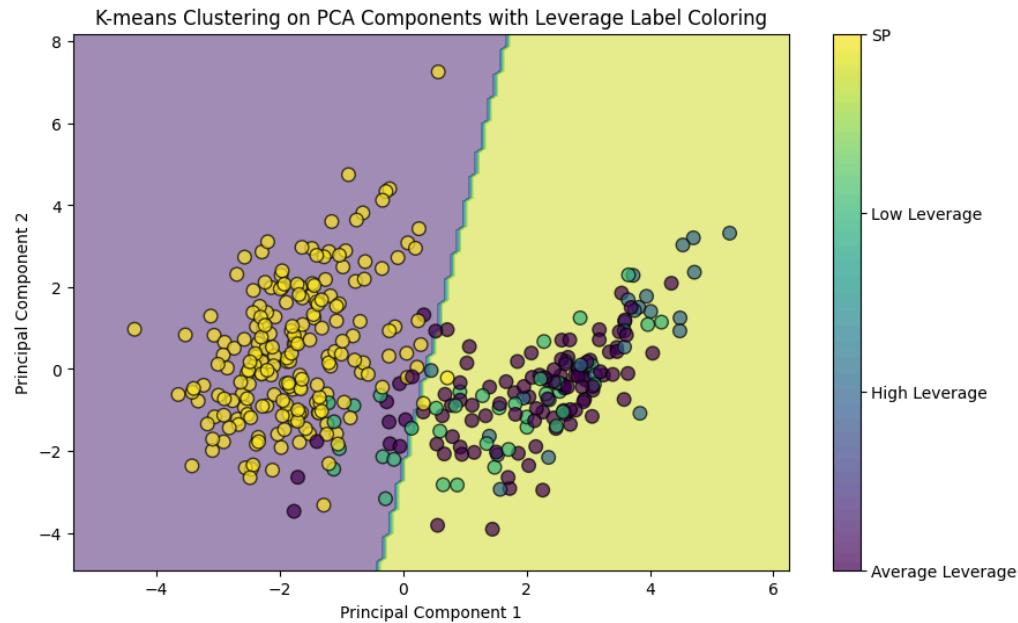


Figure 1

The approach of two-cluster K-means clustering shows two highly distinct clusters between starters and relievers. This appears to be due to a high degree of separation in the two roles' ability to extend late into games while maintaining a consistent degree of spin and velocity. While this approach does not ascribe causal effects to the outcomes, it does reveal which pitchers have incongruities between their current roles and their observed skill-sets. As explained below, this visualization provides a clear indicator of pitchers who have reliever-esque qualities but are currently utilized as starters, and vice versa (yellow dots occurring on the right-side cluster and purple or green dots occurring on the left-side cluster).

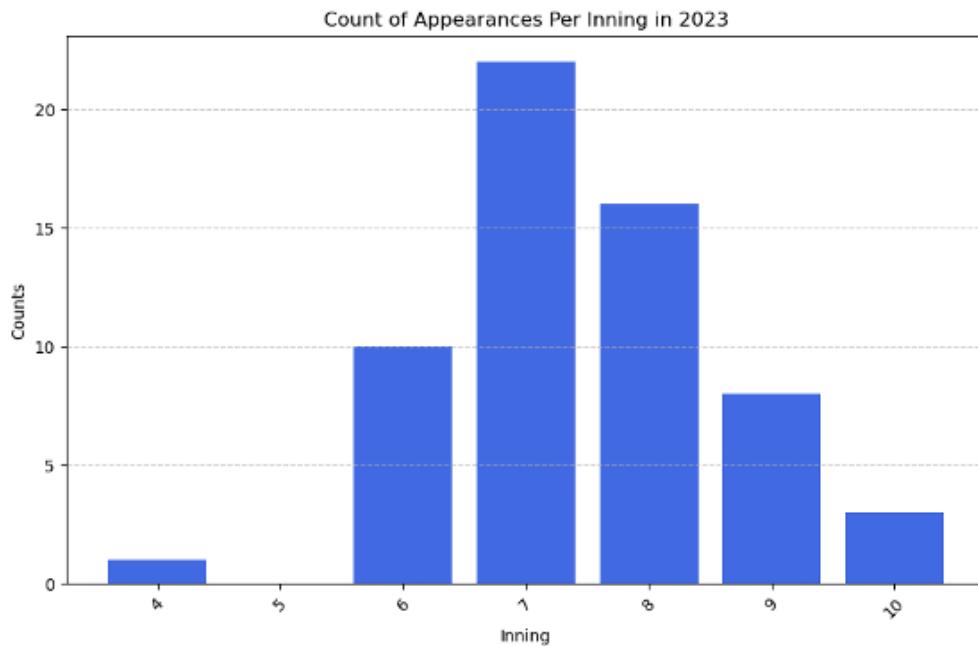
## Player Profiles

Robert Stephenson



The right-hander reliever from the Los Angeles Angels recently signed a 3-year, \$33 million deal that highlights the potential value of this pitcher. In 2023, Stephenson played for the Pittsburgh Pirates and for the Tampa Bay Rays, having a great contribution to the Rays team success. Furthermore, with the Rays, Robert grew from being an average middle reliever into a gem for the Rays bullpen. Mike Petriello for mlb.com highlights Stephenson as the main cause of the big change that the Rays bullpen had last season from being the "worst" to "one of the best ever"<sup>1</sup>.

The key to Stephenson's success was his significant improvement in his K/9 and BB/9 rates compared to his numbers with the Pirates. He increased from a good 10.9 K/9 to an elite 14.1 K/9 with the Rays and decreased his BB/9 rate from 5.1 to 1.9. His ERA+ while he was with the Rays was 179 showing how much better he was to the average pitcher in the MLB. Stephenson's repertoire relies primarily on an impressive cutter and he also combines it with a split finger fastball. In the first months of the season he used the slider and fastball way more. In the second half the slider disappeared and the fastball was a rare occurrence.



<sup>1</sup> <https://www.mlb.com/news/rays-bullpen-improvement-in-2023>

Figure 2

The change we propose is moving Stephenson to **closing** and **high leverage** situations. In his stay with the Rays he was often used in the 6th and 7th inning while Jason Adam and Pete Fairbanks were used more in the setup and closer role. Overall, as figure 2 shows, in 2023 Stephenson was used mostly in the 7th inning, although he had more late inning appearances with the Pirates. Stephenson's attributes make him a top 8 pitcher in terms of principal component 1, as we can see in the 2-d graph, he is an outlier for that component (figure 3). Principal component 1 is characterized by the spin rate and release speed drop threshold. The group of players that are close to Robert are the likes of Emmanuel Clase, Liam Hendriks, Ryan Pressly, and Ryan Helsley.

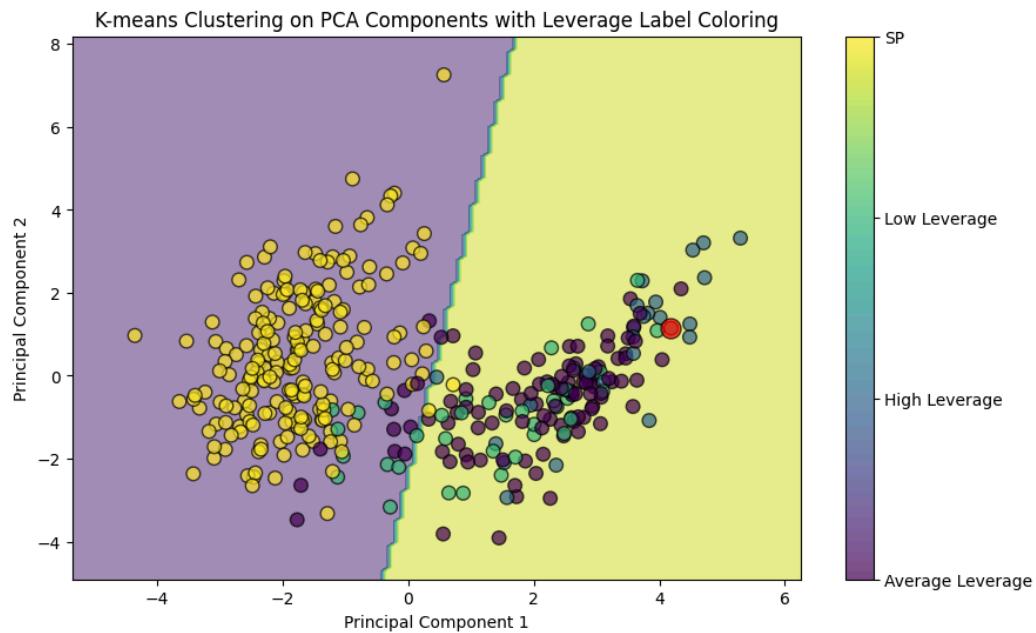


Figure 3

50% of Robert Stephenson's outings last season were with the score of the winning and losing team being tied or within one. He can play in close games and has the characteristics of pitchers that pitch in high leverage and late in the game. Moreover, Carlos Estevez, the current Angels closer, had a significantly higher ERA (3.70 vs 3.10) and even though he only blew 4 saves, is not a better pitcher than Stephenson. The Angels have a huge belief in him, as shown by the large contract they gave him. Robert Stephenson must be the Los Angeles Angels closer for the 2024 season.

Clarke Schmidt

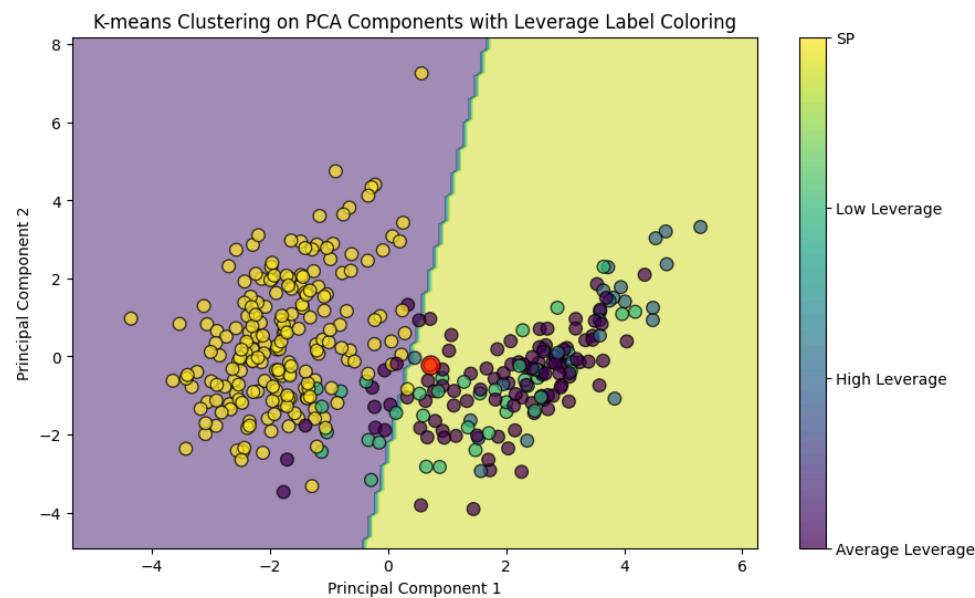


Figure 4

Clarke Schmidt appears to be an outlier starting pitcher that has many characteristics of relievers. In fact, he has the highest pc1 component of all starters for the three seasons (figure 4). Schmidt pitched as a reliever in 2022 getting very good numbers but injuries forced the Yankees to use him as a starter in 2023. Even though he was not bad, he was not at the level he had shown in 2022. With the Yankees rotation stabilizing again with Rodon, Cole and Cortes, plus German and Severino it is time to send Schmidt back to the bullpen.

In 2022 he had an ERA+ of 127 compared to 93 in 2023. He does have endurance to be a starter. However, a middle reliever is a more appropriate role for him. His spin rate drops for his fastball significantly after 21 pitches and for his secondary pitch after 35 pitches. He can fulfill that long reliever role and pitch more frequently than just the normal 5 days rest and have a greater impact on Yankees wins.

Clarke's stuff was notably better as well in 2022. His Stuff+ and Location+ were 14 points and 5 points better than his corresponding numbers in 2023.