Cincinnati Reds Hackathon 2024: MLB's Freaky Friday Pitcher Role Reversal

Nathan Backman, Jake Balek, Hunter Geise,

Danielle Napierski, and Nolan Pittman

February 5, 2024

Introduction

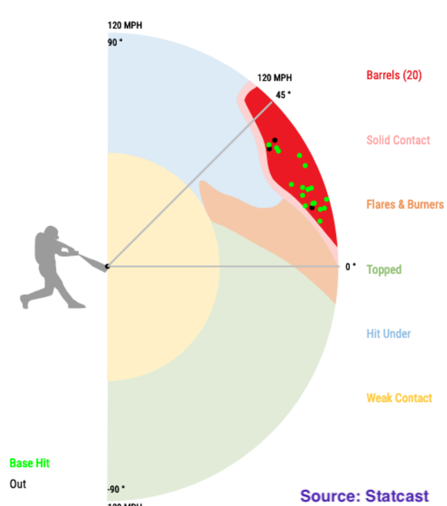
Throughout the course of a game, a manager's decision can be the deciding factor in the success of the team. While it's impossible to determine whether or not a managerial decision is the "correct call", the data revolution in baseball has allowed for an objective means to critique these decisions. Through the use of unsupervised learning algorithms, we can examine different types of pitchers and their roles, allowing for an optimal strategy in bullpen decision-making.

Methodology

Our approach to attempt to correct inefficient bullpen decisions was to create categories for different pitchers using agglomerative hierarchical clustering. Pitchers were categorized on two different fronts: their batted ball profile and their plate discipline profile. Using pitch-level data from Baseball Savant and season-level data from Fangraphs, cluster models were created to analyze these profiles on a pitcher level. To prevent too much cluster variance, the data was filtered to include only observations with at least 50 total batters faced during a single season. Each of the predictors were then converted into percentiles to help with convergence in the clustering algorithm.

Batted Ball Cluster Model

The first model analyzed each pitcher's tendencies to allow different types of batted balls. The variables included were:



- Weakly hit percentage
- Topped percentage
- Solid Contact percentage
- Barrel percentage
- Ground Ball percentage
- Under percentage
- Flare/Burner percentage
- Fly Ball percentage
- Line Drive percentage
- Popup percentage

These variables are all defined by Statcast's classification of a batted ball. This resulted in four distinct clusters, which have been labeled as:

- Extreme Ground Ball
- Ground Ball
- Extreme Fly Ball
- Fly Ball

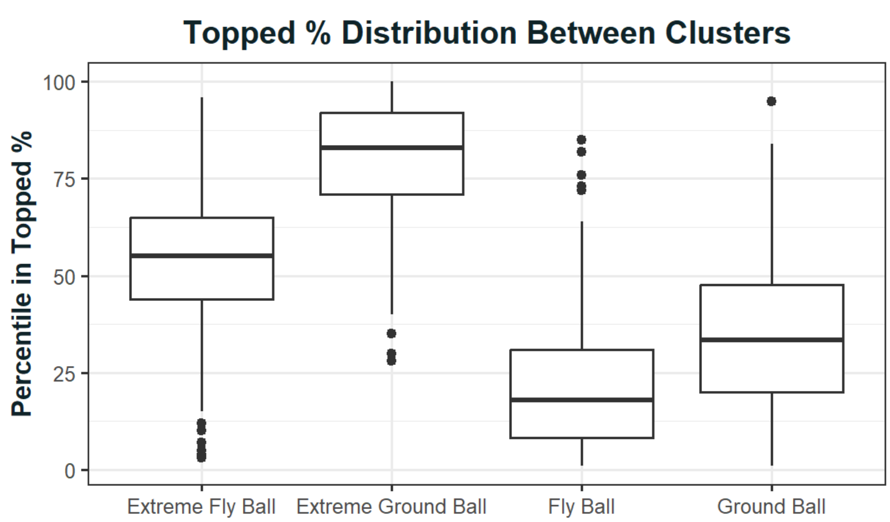


Figure 1: Distribution of Topped% Percentiles Among the Different Clusters

The ground ball and fly ball clusters are more representative of a "neutral" class, but lean slightly either toward ground ball or fly ball.

STARTERS		RELIEVERS	
Pitcher Class	*Pct*	*Pitcher Class*	*Pct*
Extreme Fly Ball	16.86%	Extreme Fly Ball	19.79%
Extreme Ground Ball	28.39%	Extreme Ground Ball	35.91%
Fly Ball	28.39%	Fly Ball	22.92%
Ground Ball	26.36%	Ground Ball	21.39%

Table 1: Batted Ball Profiles Between Starters and Relievers

<u>Plate Discipline Cluster Model</u>

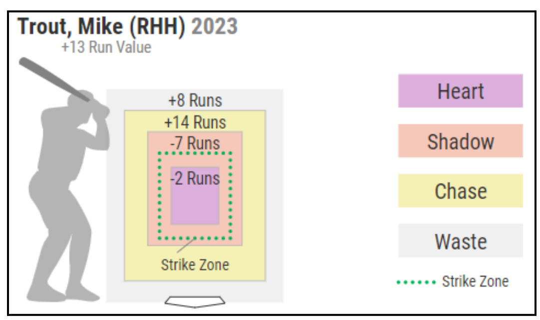
The second model analyzed each pitcher's plate discipline profile. Variables included were:

- Strikeout Percentage
- Walk Percentage
- Chase Rate
- Zone Swing Rate
- Chase Contact Rate
- Zone Contact Rate
- Overall Swing Rate
- Overall Zone Rate
- Shadow Zone Rate*
- First Pitch Strike Rate
- Whiff Rate
- Swinging Strike Rate
- Called Strike Rate

* as defined by Statcast

This resulted in 3 distinct clusters, which have been labeled as:

- Power
- Finesse
- Junk



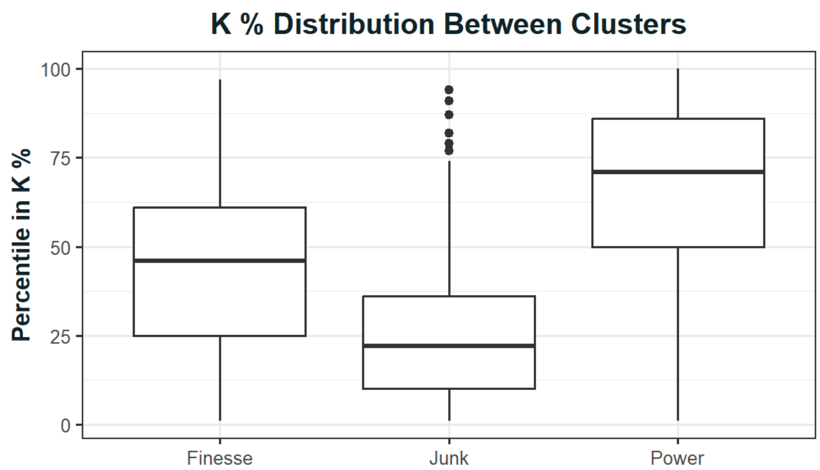


Figure 2: Distribution of K% Percentile Between Clusters

Power pitchers are defined by high strikeout and swing and miss rates. Finesse pitchers are defined by high strike and swing rates, as well as an ability to locate on the corners and low walk rates. Junk pitchers are defined by low strikeout, swing-and-miss, and walk rates. In essence, the junk pitcher is a prototypical "pitch-to-contact" type of pitcher.

STARTERS		RELIEVERS	
Pitcher Class	*Pct*	*Pitcher Class*	*Pct*
Finesse	21.42%	Finesse	12.30%
Junk	40.68%	Junk	26.89%
Power	37.90%	Power	60.81%

Table 2: Plate Discipline Profiles Between Starters and Relievers

Pitcher Classification Results

Every pitcher is associated with a cluster from each model based on their batted ball and plate discipline profile. For example, Kenley Jansen is classified as a Fly Ball Power pitcher for his 2021 season, which makes sense because he had a 30.1% strikeout rate and 43.2% of batted balls against him were fly balls or popups. Once these pitcher categories were determined, we wanted to analyze the effectiveness of each pitcher category in different roles. To determine pitcher roles, further cluster analysis was done on bullpen decisions based on game situations.

Reliever Role Model

This model analyzed each situation in which a relief pitcher entered the game. The following variables were used to create the role clusters:

- Entrance Leverage Index
- % of the time entering the game with RISP
- % of the time entering the game in a “fresh” inning (no outs, no one on, no count)
- Number of batters faced per appearance

This resulted in 6 distinct clusters forming, which have been labeled as:

- Closer
- Middle Relief
- Medium Leverage Stopper
- High Leverage Stopper
- Specialist
- Mop-Up

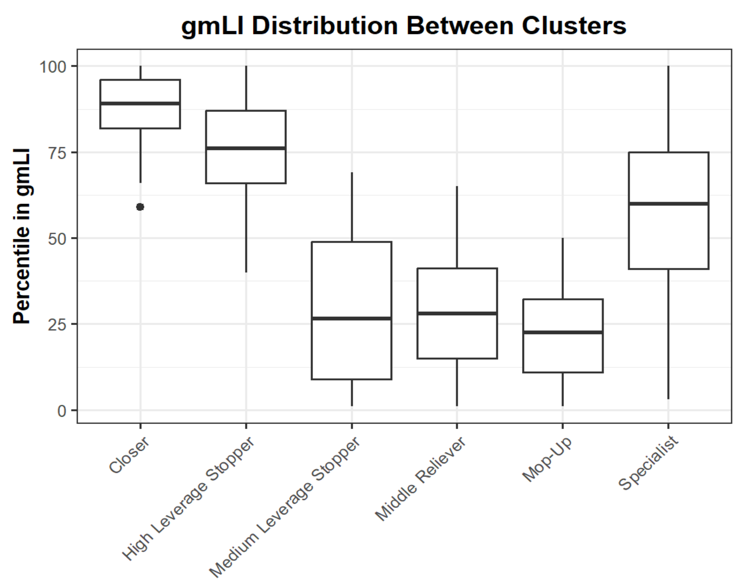


Figure 3: Distribution of Entrance Leverage Index Percentiles Between Clusters

Closers were identified as relievers who often entered the game in high leverage situations with a fresh inning. Stoppers were defined by relievers that were frequently brought in with runners in scoring position. Middle relief pitchers were characterized by entering the game in lower leverage situations at the start of an inning. Mop-up was defined by relievers who faced more batters per relief appearance in lower-leverage situations. Specialists lacked any indicating trait and were treated as an "other" group. The idea behind the specialist group is that it comprised pitchers who were brought in for specific pitcher-batter matchups.

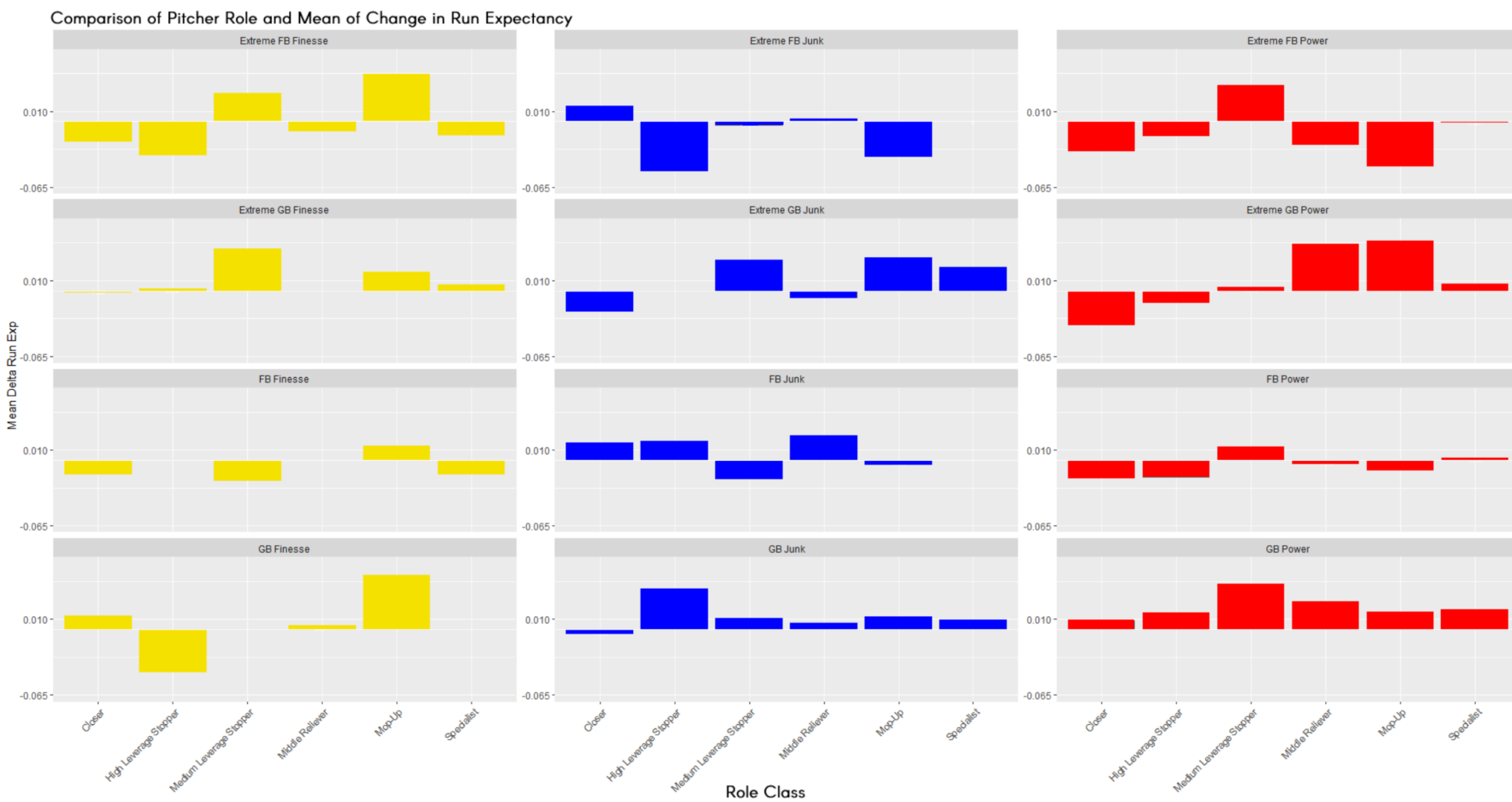


Figure 4: Average Run Expectancy Separated by Pitcher Type and Role

In order to determine the effectiveness of bullpen decisions, the change in run expectancy was used based on the pitcher type and their role. The goal of any bullpen decision is to yield zero runs and therefore using the change in run expectancy can help us determine which pitcher types are successful at mitigating or preventing runs in certain situations. As an example, the data unsurprisingly suggests that power pitchers are more likely to be successful in the closer role. In observing these average changes in run expectancies, we can determine which pitchers are being used “suboptimally”.

Pitchers Being Incorrectly Used

Trevor Williams

During the 2021 season, Williams was used as a Mop-Up reliever and was classified by our model as an Extreme Ground Ball Finesse pitcher. According to the change in run expectancies for an extreme ground ball finesse pitcher, being used in the Mop-Up role saw roughly 0.02 runs being added. Williams himself had an ERA of 4.43 in the 2021 season.

His struggles were most noticeable when he was brought into the middle of an inning and had to deal with runners on. These situations accounted for roughly 20% of his appearances, with the rest of his appearances coming at the beginning of an inning. When brought in with bases loaded, Williams had an average change in run expectancy of 1.026 runs. Given his classification as an Extreme Ground Ball Finesse pitcher in 2021, the data suggests that Williams would have been more effective in a Specialist or Middle Reliever role, as Extreme Ground Ball Finesse pitchers in those roles netted an average change in run expectancy of 0.005 and 0.01 runs respectively. These differences are marginal, and Williams’s use may have been in part due to his rise through the minors as a starter, but being used in those situations at a higher volume could’ve been the difference-maker in a few wins.

This marginal improvement was seen in his 2022 season. The model reclassified him as a Ground Ball Finesse Pitcher, and he was transitioned into more of a Medium Leverage Stopper role. While the data states that Ground Ball Finesse pitchers yielded an average change in run expectancy of 0.08 runs, 0.06 more than his previous role, the difference is seen in the game situations he was brought in for. He had many more inherited runners than he did in the season prior, as roughly 42% of his relief appearances had inherited runners. Compared to his average change in run expectancy of 0.32 runs in 2021, Williams had a lower average change in run expectancy of 0.10 runs in 2022. This role change ultimately saw Williams mitigating a change in roughly 0.2 runs.

Josh Taylor

In 2023, Taylor was used in a Mop-Up role and was classified as an Extreme Ground Ball Power pitcher. His 2023 ERA was 8.64 as a reliever and he struggled in multiple situations with runners on and no one out. The data suggests that Taylor would succeed in a Specialist role moving forward. Extreme Ground Ball Power pitchers in the Mop-Up role had an average change in run expectancy of 0.05 runs versus 0.007 runs in the Specialist role. Being a lefty, this would allow for Taylor to be employed in favorable matchups as well. When there were 2 outs in an inning, Taylor had great success upon entering the game despite the situation, netting an average change in run expectancy of -0.12 runs, showing his potential to be successful in one out scenarios.

A pitcher that he could model himself after would be Kevin Ginkel of the Arizona Diamondbacks. Ginkel, like Taylor, is an Extreme Ground Ball Power pitcher and struggled during the 2021 season with an ERA over 8. He was then moved to the Specialist role where his ERA fell to just above 4 in 2022 and this year was even better at 2.50.

Room for Improvement

As with any methodology that relies on black box modeling, these classifications are far from perfect. Where this analysis falls short is in addressing the impact of these categories on starting pitchers. Having an analysis for starting pitchers based on the classification could allow for some observations on effectiveness as a starter versus a reliever.

Another area that could use some improvement is the use of delta run expectancy as the main measure for pitcher effectiveness. While the average change in run expectancy was used as a way to judge pitcher performance on a run value scale, other metrics can definitely be incorporated to provide a more robust analysis of pitchers.