A Run Expectancy-Based Approach to Stuff+

Will Cave
Syracuse University
April 2024

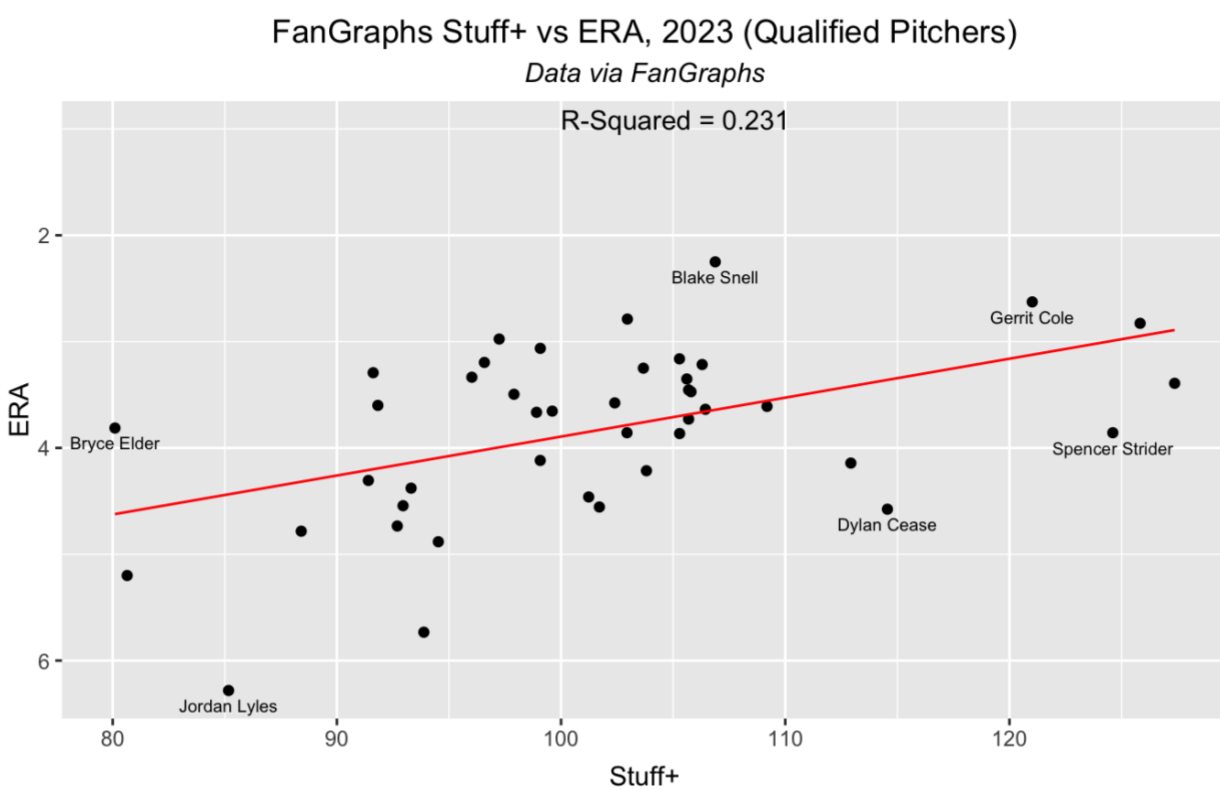
I. Abstract

This project aims to quantify the value of a pitcher's pitch based on its movement profile. By using location-independent variables as predictors, tree-based models performed well in their ability to predict the change in run expectancy. By scaling these predicted values, I was able to develop my own version of a Stuff+ model for each individual pitch type. The results from this project can help us to identify movement patterns that contribute to the effectiveness of certain pitches, and also compare the performance of pitchers across Major League Baseball in terms of stuff. My inspiration for this project came from the belief that other Stuff+ models were not accurately evaluating changeup performance because of its lack of movement compared to breaking balls. This paper will evaluate all pitch types, but with a focus on the changeup and which factors lead to optimal performance.

II. Introduction

Assigning quantities to pitches to assess their quality has arisen as a relatively new method for pitcher evaluation. One of the most popular models is Stuff+. In most cases, a Stuff+ model is based on the physical characteristics of a pitch, such as release point, velocity, spin rate, and horizontal/vertical break. It is scaled to where a 100 Stuff+ would be an average offering. A 120 Stuff+ would indicate that the pitcher's stuff is 20 percent better than league average, and an 80 Stuff+ would indicate that the pitcher's stuff is 20 percent worse than league average. Generally, the best starting pitchers in the league will hover around 120-130 Stuff+. Stuff+ models are generally good predictors of how pitchers will perform. While it isn't the end-all, be-all of effective pitching, researchers found that Stuff+ in the first half of the season had a strong positive correlation with strikeout percentage in the second half. Stuff+ also had a higher correlation with second-half K%-BB%, ERA-, and FIP- than those stats in the first half did. This can show Stuff+ to be a powerful predictor for future success. The graph shown below shows the relationship between FanGraphs' Stuff+ and ERA in 2023 among qualified starting pitchers, with a few notable pitchers specified.

Figure 1: FanGraphs Stuff+ vs ERA, 2023 (Qualified Pitchers)



One of the specific pitches I was interested in analyzing was the changeup. Changeups typically grade out poorly on Stuff+ models due to their similarities to fastballs. According to the 2023 FanGraphs leaderboard, only four qualified pitchers had a Stuff+ over 100 on their changeups: Logan Webb, Sandy Alcántara, Justin Verlander, and Charlie Morton. I believe that these numbers are not accurately capturing the quality of changeups. The table below shows the highest-rated pitches from an individual qualified pitcher in 2023. For example, Logan Webb had the best changeup Stuff+ in 2023, so he corresponds to the 111 value in the changeup column of the table.

Table 1: The top 5 individual pitcher Stuff+ ratings on pitch types in 2023

5 Highest Stuff+ Ratings by Pitch, 2023 MLB Season (Qualified Pitchers Only)
Data via FanGraphs

Fastball	Changeup	Curveball	Slider
141	111	161	171
125	108	160	144
125	102	139	142
120	100	135	136
117	99	121	133

Breaking balls tend to grade out higher because of the amount of break and spin, but these pitches do not necessary lead to more advantageous changes in run expectancy for pitchers than other pitches do. For this project, I opted to build a multitude of models to isolate which location-independent variables are most important in determining pitch quality and utilized those models to create a Stuff+ metric that effectively captures pitch quality.

III. Literature Review

In order to create a new Stuff+ model, with a focus on properly grading changeups, this literature review will consider past baseball research and online content. A variety of different topics will be examined, such as biomechanical factors, previously created models, and pitch sequencing. These topics will be considered for the application of different modeling techniques and will be extensively covered in this literature review.

a. What is Stuff+?

Assigning quantities to pitches to assess their quality has arisen as a relatively new method for pitcher evaluation. One of the most popular models is Stuff+. In most cases, a Stuff+ model is based on the physical characteristics of a pitch, such as release point, velocity, spin rate, and horizontal/vertical break. It is scaled to where a 100 Stuff+ would be an average offering. (Appelman 2023). Stuff+ models are generally good predictors of how pitchers will perform. While it isn't the end-all, be-all of effective pitching, researchers found that Stuff+ in the first half of the season had a strong positive correlation with strikeout percentage in the second half. Stuff+ also had a higher correlation with second-half K%-BB%, ERA-, and FIP- than those stats in the first half did (Rosenblum 2022). This can suggest that pitchers without good stuff can get "figured out" in the second half, and their ERA- and FIP- regress. Changeups traditionally grade out poorly on Stuff+ models due to their similarities to fastballs. According to the 2022 FanGraphs leaderboard, only four qualified pitchers had a Stuff+ over 100 on their changeups. Another potential contributing factor to changeups grading out poorly is that Stuff+ does not factor in plate location (Langin 2021). The effectiveness of a changeup often depends on its location, and while there are other models that can quantify a pitcher's ability to locate pitches, Stuff+ will not consider it. Location+ is one of those models, which considers count and handedness. More "points" are awarded for hitting the corners, while pitches grade out worse if they are well out of the zone or right down the middle. In order to fully capture an effective changeup, location has to be considered.

Attempts to make Stuff+ models more specific towards individual pitches have been made in the past. A Rockland article from 2023 discusses this predicament. The Rockland model assigns different weights to variables for certain pitches. For example, a changeup's velocity difference might be more important than a slider, and so on. Another example would be a 4-seamer vs a sinker, as a sinker places more weight on the horizontal break than a standard 4-seam fastball does (RPP 2023). While there is some deviation from a traditional model, understanding the fundamental difference between pitches can help with assigning fairer values to each one. A changeup needs other pitches to be effective. No pitcher throws a changeup without a fastball; those pitches go together and play off of each other. An article from Baseball Prospectus describes what makes changeups effective. It puts changeups into two categories: "bat-missers", as in pitches that generate a high number of whiffs, and "worm-killers", as in pitches that generate a high number of ground balls. For a changeup to miss bats, it plays best off of a plus-velocity fastball, with the changeup clocking in at around 10 mph (ca. 16 km/h) slower than the fastball. "Worm-killers" produce ground balls, typically because of the late vertical break and location. The article also discusses changeup grip, and that "throw the changeup slow" leads to a less effective pitch. Having a good grip can make the pitch more deceptive because the arm motion will not be much different from a fastball (Pavlidis 2013). One of the key points here is the understanding that a changeup is different for each pitcher. For example, Sandy Alcantara's changeup grades out well in Stuff+ models. He is a hard-throwing RHP and his changeup plays well off of his primary pitch, a fastball. Merrill Kelly's changeup also performs well on Stuff+ models, despite not having much of a difference in velocity between his fastball and changeup.

Some baseball minds have disagreed with this logic, however. An article by Dan Blewett of Elite Baseball Performance claims that in order for a changeup to be effective, the speed difference needs to be about 10% from the fastball (Blewett 2017). Additional research from Harry Pavlidis of FanGraphs supports this claim (Mailhot 2020). There are plenty of hard-throwing pitchers that use changeups, like the aforementioned Alcantara, as well as Gerrit Cole of the Yankees (97 mph (ca. 156 km/h) fastball, 89 mph (ca. 143 km/h) changeup in 2023), Dylan Cease of the White Sox (96 mph (ca. 154 km/h) fastball, 75 mph (ca. 121 km/h) changeup in 2023), and Spencer Strider of the Braves (97 mph (ca. 156 km/h) fastball, 87 mph (ca. 140 km/h) changeup in 2023). The data for those pitchers is from Baseball Savant, with the numeric values being their average pitch speed on each pitch during the 2023 season. However, there are still plenty of pitchers that have effective changeups that don't have a high-velocity fastball to play off of. For reference, Merrill Kelly averaged 92.2 miles per hour (ca. 148 km/h) on his fastball in the 2023 season, and his changeup averaged 88.7 miles per hour (ca. 143 km/h). This is a difference of about 3.8%, well below the 10% threshold set by Blewett. However, the pitch is one of Kelly's most effective offerings. In 2023, Kelly averaged a .215 wOBA against on his changeup, by far the lowest of any of his pitches. So what makes it so effective? An article from Rapsodo details the importance of spin rate. A lower spin rate decreases velocity and helps add to the illusion of the pitch. A perfectly optimal changeup spin direction would be directly sideways, though the goal is generally to reduce backspin and lift. Pitchers like Devin Williams use a "split" grip on their changeups to get their fingers on the side of the ball, thus increasing side spin. Williams' changeup has by far the highest spin rate in baseball (Mailhot 2020). For vertical and horizontal movement, positive vertical break should be avoided, as it creates lift, but horizontal movement can make a changeup more difficult to hit (Cochran 2023). Indeed, Kelly has an above-average active spin rate, and in 2023 his changeup had an observed spin direction at 3:00, indicating direct side-spin. It was the first time in his career that his observed spin direction was recorded at exactly 3:00, and it correlates to Kelly's lowest-ever wOBA against on that pitch. Additionally, Kelly's changeup also has above-average horizontal and vertical break.

An important factor of Stuff+ is the spin rate of the pitch. Breaking balls typically have higher spin rates than fastballs, as those pitches are intentionally thrown to have a spin rate as high as possible to induce movement. Pitchers that throw breaking balls "flick" or "twist" their wrist in the direction they want to spin the ball, and correspondingly break in that direction. A typical curveball has what is called "1-to-8" or "2-to-9" movement, where the pitch starts high and inside and breaks low and outside. Some pitchers also throw a "12-to-6" variation of the curveball, where the pitch breaks nearly straight down. A high spin rate on a breaking ball causes pitches to break later, and lower spin rates cause a "loopier" pitch shape that can lead to breaking balls up in the zone. Curveballs with increased topspin were more difficult to distinguish and hit. Curveballs that had a higher vertical displacement from their expected points (where a fastball would expect to end up at that same release point) are more difficult to hit. For sliders, the researchers found that the closer the slider was in velocity, the harder it was to hit. Think of a pitcher like Jacob deGrom, whose slider clocks in above 90 mph (ca. 145 km/h) regularly. The difference in velocity is not high from the fastball to the slider, but obviously the movement profiles are much different. (Nakashima 2020). For changeups, a lower spin rate will cause the pitch to drop, and a higher spin rate causes cutting action with the pitch (Hauswirth 2023). As mentioned earlier, Devin Williams is the unicorn in this situation; his changeup has an absurdly high spin rate but also has well-above average vertical and horizontal break. Pitchers will do whatever they can to increase spin rate, as it typically makes pitches more difficult to hit. In 2021, many Major League Baseball pitchers were found using Spider Tack, a tacky that is typically used by strongman competitors for a better grip while lifting stones. Pitchers would place the tacky on their fingertips, allowing them to get a much better grip on the ball. This would result in breaking balls with more movement, and increased velocity on fastballs. While the substance was outlawed, and pitchers are required to comply with "substance checks" during games, the effect of an increased spin rate was markedly observed.

b. Safety of Changeups in Youth Baseball

So why changeups? For starters, it is one of the "safest" pitches to throw because it puts less stress on the elbow and shoulder (Fleisig et al. 2006). For many younger pitchers, the changeup is the first pitch that they learn because of this. Further research suggests that sliders can lead to pitcher injury when wrist motion is exaggerated (Platt et al. 2021). An increase in total pitch break on both the four-seam fastball and slider contributed to an increased risk of shoulder injury for MLB pitchers. For the slider specifically, a decreased spin rate and increased vertical break contributed to injury as well. Some of the "nastiest" pitchers in baseball, like Jacob deGrom, have extensive injury history, and deGrom is also well regarded for his elite fastball and slider. Of course, this doesn't mean that pitchers should stop throwing sliders, it just means that those types of pitches place more stress on joints involved in the throwing motion.

c. Deception and Pitch Location

The effectiveness of a changeup stems from its deception, whether that comes from speed differential or unique movement profiles. Since the fastball and changeup generally have similar movement profiles, many of the biomechanical factors remain similar for both pitches. A 2012 study found that the pressure applied on finger joints is similar for fastballs and changeups (Chen et al. 2012). A typical fastball grip is the four-seam grip, where the index and middle fingers lie over the seams. Changeup grips vary, but they both rely on pressure from the index and middle fingers as well. Additionally, the arm cocking speed and acceleration phases were both similar for those two pitches. Pitchers want to avoid "telegraphing" pitches; if changeups were thrown simply by slowing down the throwing motion, they would be much easier to identify. As discussed earlier, the changeup grip is a large contributing factor to avoiding telegraphing pitches. Variables that have shown to be impactful for pitch velocity include maximum elbow extension velocity, maximum humeral rotation velocity, maximum lead leg ground reaction force resultant, trunk forward flexion at release, and time difference of maximum pelvis rotation velocity and maximum trunk rotation velocity (Nicholson et al. 2022) Important biomechanical variables that contribute to a pitcher being able to locate consistently include shoulder abduction, upper trunk tilt, shoulder horizontal abduction, shoulder horizontal adduction, and maximum shoulder external rotation. All had a positive relationship with consistency except for shoulder horizontal abduction (Glazner et al. 2019). For a changeup to be effective, location is vital. For example, pitchers generally try to avoid throwing inside changeups; even if a batter is deceived by the pitch, they can still pull it for power. A study conducted on the structure of error contextualizes this claim. This study examines the effects of pitching motions on the structure of pitching error. Pitchers with different mechanics were used in the study, ranging from straight overhand, 3/4 slot, sidearm, and submarine pitchers. Mechanics help determine where a pitcher might miss if they aim for a specific location. For example, if the pitcher has a right-up-left-down error distribution, they are more prone to missing over the heart of the plate if they aim for the outside corner (to a RHB). Conversely, a pitcher with a right-down-left-up distribution would not be prone to missing over the middle. The researchers calculated an ellipse for each pitcher, where the center would be the intended location. They found that the pitcher's arm angle was almost directly correlated with the angle of the ellipse (Shinya et al 2017). Rather than focusing on the size of the error, it focuses more on the shape. For example, a right-handed pitcher with a three-quarters arm slot trying to locate a pitch low-middle has the potential to miss their spot anywhere within the ellipse from middle-right to low-left below the zone.

From pitch to pitch, pitchers generally do not locate a fastball more accurately than a breaking ball. A study on pitch accuracy found that the pitch location tracking for fastballs as opposed to breaking balls within each group were not statistically significant (Kawamura et al. 2017). A study tracking fastball location ability using pitch speed, release position, release projection angle, spin rate, and spin axis as variables found that the highest contributing variable, using a multiple linear regression model, was release projection angle. The variable that contributed the least to pitch location was spin rate. For vertical location, all the predictors included in the model were significant. For horizontal location, the release side, horizontal release angle, spin rate, and spin axis were statistically significant predictors (Nasu, Kashino 2020). Spin direction can help to create the cutting action on the pitch, like Merrill Kelly's changeup that was mentioned earlier. Side spin leads to increased horizontal break. Consistency in location is also heavily influenced by the pitcher's ability to replicate their throwing motions. It is exceptionally difficult to reproduce similar motions when they are being conducted at higher speeds (Kusafuka et al. 2021). This study by Kusafuka examined NPB pitchers that utilized a variety of arm angles. It was found that since professional pitchers are able to maintain the parameters that go into pitch location, they are able to consistently locate their pitches. At lower levels of baseball, walks are observed at significantly higher rates than in Major League Baseball. Part of this is because of inconsistencies with mechanics. With technologies such as TrackMan and Hawkeye being implemented by professional organizations, the ability to maintain release point consistency has become a key discussion point for pitch location. Pitchers can experience inconsistency with release point for a variety of reasons. One of those can be fatigue. As the game goes on, pitchers that are more tired can have less success with maintaining consistent mechanics. Often times when pitchers have erratic outings, their release points are not consistent, whether that be from pitch to pitch, or on just one pitch. An article on pitch tunneling discusses release point. Angels star Shohei Ohtani has inconsistent release points for each of his pitches. For example, his fastball and slider have noticeably different release points. However, from pitch to pitch, he is consistent with his release point, meaning he releases his slider from the same point and fastball from a different point, but each pitch has low individual variability (Tieran 2022). Maintaining the same release point on a certain pitch helps pitchers like Ohtani consistently locate. A different study took measurements from a group of high school pitchers. Each pitcher was asked to throw 8–12 fastballs while aiming at a specific target, then the pitchers were grouped into high consistency or low consistency based on a confidence interval. High consistency pitchers had decreased lead hip flexion at elbow extension and foot contact, decreased back hip extension, and increased back hip internal rotation. Lower consistency pitchers had maximum lead hip flexion earlier in the pitch (Manzi et al. 2022). Detailing which kinematic factors contribute to location consistency can help pitchers at all levels refine their mechanics. While we have potential evidence that release point consistency from different pitch to different pitch might not be as meaningful, staying consistent from an individual pitch standpoint is necessary for effective pitching.

A study with the goal of seeing if mixing up pitch location would yield better results for pitchers was conducted with pitch f/x data. In the study, however, it was found to be the opposite. Pitchers who had less "zone type uncertainty" (as in, it would be more difficult for the batter to predict pitch location) did not necessarily pitch worse than pitchers with high uncertainty. In fact, pitchers with a higher zone type uncertainty had a higher FIP on average (Kim, Jung 2018). Pitchers being able to hit their spots consistently, especially on breaking pitches, is important. If pitchers have a high zone type uncertainty on breaking balls, it could mean that they are hanging those pitches up in the zone. Professional hitters are able to adjust to those slower pitches up in the zone. Ideally, a pitcher would be able to keep all of their pitches on the corners, which would contribute to a low zone type uncertainty if they were able to do so, though it would not necessarily make it easier to hit. Another study considered pitch count, sequencing, and velocity, as well as other location variables. A variable called expected total bases was created, which considers the count and pitch descriptors. To estimate total bases, a random forest model was used. The model found that the change in the count, horizontal location, and vertical location had the greatest effect on estimating total bases. It was also found that expected total bases had a statistically significant relationship to ERA and FIP (Swartz et al. 2017). It's no surprise that expected total bases was a good predictor for ERA and FIP; as batters tally more bases, they will score more runs. However, using horizontal and vertical location in a model can help contextualize pitch quality. Count data can also provide intuition, as approaches change when counts change. A different study using FIP as a pitching performance variable measured the kinematics of the ball in flight and the variability of the release point. Fastballs and changeups were found to have similar spin patterns, and sliders and curveballs also had similar spins, albeit less so than the FB/CH combo. For in-flight kinematics, the researchers took the speed, spin rate, azimuth angle (spin towards 3rd base for RHP), FSRI (spin direction relative to four-seam angle of the ball), and release location variability. None of these variables had a statistically significant correlation at the 0.05 level with FIP (Whiteside et al. 2015). This demonstrates the importance of location as well as horizontal/vertical break (though spin rate contributes to the latter). As we have seen in earlier quoted research, pitchers can be effective working at all speeds; Merrill Kelly works a good fastball and changeup combo based on his movement profile and location, Gerrit Cole works a good fastball and changeup combo based on his speed differentials.

An interesting biomechanical comparison regarding the ability to execute repeated motions is dart-throwing. While it is not done at a high speed, dart-throwing requires extreme precision with every element of the throw. A study included in Intelligent Autonomous Systems 12 found that a high contributing factor for dart-throwing accuracy was vertical release point. In order to test this, subjects from the test were tasked with training activities to improve their vertical release point and consistency. The study found that the subjects that received this training were more accurate with their throws, and that the optimal release point is just before elbow angular velocity reaches a maximum (2012). Training for release point consistency is not a new concept. In fact, it is taught from a young age. One of the first things young pitchers learn how to do when they are starting out pitching is how to stay consistent with mechanics. It does, however, remain vital for pitchers to locate consistently.

d. Sequencing

A large contributing factor to an effective changeup is a pitcher's ability to sequence pitches to keep a hitter on their toes. On its own, a changeup is essentially just a slower fastball; it needs other pitches to play off of to maximize its effectiveness. In the past, models created to predict pitch types have had success rates of about 70%, where pitches were classified as fastballs or non fastballs. A new model created considered variables such as pitcher history and pitch frequency, as well as game situation. Variables like count, men on base, outs, and other factors were also considered. The linear model created was able to predict the next pitch thrown by a pitcher with all available data at a rate of 74.5%. Previous models predicted this at 70%, with only the fastball/non fastball classification. This model considered the pitcher's four most frequently thrown pitches. It was also found that pitchers that had less predictable pitch sequences had higher ERAs and FIPs (Bock 2015). There can be many explanations for this, but generally, pitchers with excellent pitch repertoires can afford to be more predictable if their pitches are more effective. Spencer Strider of the Atlanta Braves is one of the most effective pitchers in baseball, posting high strikeout rates since he entered Major League Baseball. Strider has just a three-pitch arsenal, throwing fastballs, sliders, and changeups. Most models created around the concept of predictability would likely place Strider in an echelon with other predictable pitchers. Yet, he is one of the best pitchers in the game, boasting a fastball that regularly clocks in at 100 miles per hour (ca. 161 km/h) or faster, a slider with a whiff rate of 55.3%, and a changeup with a .183 wOBA against. While "mixing up pitches" is typically regarded with high importance for a pitcher's effectiveness, pitchers with exceptional stuff can get away with being more predictable.

Another model that can be effective in predicting pitch sequences is a probabilistic topic model. Typically used on text data, probabilistic topic models can parse through large datasets and find relevant groupings and the frequency of which elements from those groups occur. This model was used on 2016-2018 NPB data. Characteristics for the hitters were included, mostly relating to handedness, batting order, plate appearances in that game, and dummy variables if they hit for a high average. It finds that overall, pitchers have a defined sequence to attack hitters of certain types, whether they are right-handed or left-handed, hit for average vs power, etc. and that the probabilistic topic model would be a viable methodology for looking deeper into these results (Yoshihara, Takahashi 2020). Some pitches have been proven to perform better based on the handedness of the batter and pitcher, so the groupings are likely influenced by that, since handedness was included as a variable in the model. An example of this is the changeup. Changeups on the outside part of the plate can potentially perform better than changeups on the inside part. Changeups are designed to generate an early swing and fool the batter. If the pitch is thrown outside, and the batter is early, it will likely result in weak contact; contact is made before the ideal hitting zone. For an inside changeup, however, it is beneficial for a batter to be early for the same reason.

A different model on pitch sequencing used pitch f/x data from MLB seasons to model the dependence of strikeout rate on how predictable a pitcher's pitch sequencing is. To maximize the sample size, game situation factors like count and outs were not considered. Multiple models were run, and it was found that the model centered around fastballs was able to explain most of the variance between strikeout rate and sequencing. Since the fastball is the most common pitch in baseball, the model centered around it performed the best. Pitchers with exceptional breaking balls had higher error values, as they struggled to capture the effectiveness of having an excellent off-speed pitch. Overall, strikeout rate decreased as predictability of pitch sequencing increased (Healey, Zhao 2017). Unless you are a unicorn pitcher like Strider, who can get away with a smaller and more predictable arsenal, pitchers without any overpowering offerings would likely benefit from mixing up pitches.

e. Conclusion

This review of online works and publications gives a solid background on what makes a changeup effective. Variables like horizontal and vertical location, spin rate, break, and velocity difference all contribute. The goal of this project is to create a model that more effectively captures how good a pitcher's changeup is, essentially, to give them more credit than a Stuff+ model would. To tackle the question of what makes a changeup effective, I will consider variables like horizontal and vertical break, spin, velocity, difference from the fastball, count, and handedness. Statcast-era data will be collected and used for this project.

By breaking down the models on an individual pitch type level, we can determine which pitch characteristics perform the best. Though location and sequencing will not be considered, they should still be noted as very important predictors of pitch success. Stuff+ is just a small part of the overall equation of pitcher performance, but an accurate Stuff+ model can help identify pitchers that might be due for improved performance.

IV. Methodology

a. Data Summary

The data used in this project is from the 2022 and 2023 Major League Baseball seasons, collected through Statcast via the baseballr package by Bill Petti. Pitch-by-pitch data was collected from both seasons. 7 different pitches were considered: 4-seam fastballs, sinkers, cutters, sliders, curveballs, splitters, and changeups. 2-seam fastballs were classified as sinkers, sweepers were classified as sliders, and slurves were classified as curveballs. Variables considered in the models were release speed, release point, horizontal and vertical movement, velocity and acceleration in each three-dimensional axis, spin rate, spin axis, extension, and the change in run expectancy after the pitch as the predictor variable.

The full dataset contained observations from 5,172 games in both seasons. There were a total of 1,454,615 pitches thrown by 1,486 different pitchers. Each pitch was considered as a separate entity, so there were no game state variables in the model. The summary statistics tables for fastballs, changeups, and curveballs are displayed below.

Table 2: Summary Statistics for Fastballs, 2022-2023

Fastballs, 2022-2023

	Summary Statistics					
	Min	Q1	Median	Mean	Q3	Max
release_speed	34.30	92.40	94.10	93.91	95.70	104.80
release_pos_x	−4.67	−2.07	−1.47	−0.81	0.87	4.74
release_pos_z	0.86	5.59	5.89	5.86	6.16	8.11
pfx_x	−5.19	−0.75	−0.46	−0.27	0.18	1.93
pfx_z	−1.53	1.19	1.36	1.33	1.50	3.62
vx0	−20.93	−2.59	4.72	2.69	7.11	19.11
vy0	−152.33	−139.10	−136.75	−136.55	−134.42	−47.69
vz0	−20.16	−6.94	−5.20	−5.12	−3.41	15.86
ax	−76.54	−10.96	−6.96	−4.06	2.67	26.39
ay	−18.69	28.21	30.14	30.06	32.03	76.34
az	−49.81	−16.59	−14.23	−14.57	−12.10	10.04
release_spin_rate	16.00	2,171.00	2,284.00	2,276.56	2,393.00	3,595.00
release_extension	3.00	6.10	6.40	6.44	6.70	9.70
spin_axis	7.00	163.00	207.00	194.43	217.00	357.00
delta_run_exp	−1.36	−0.06	−0.02	0.00	0.03	3.61

Tables 3 and 4: Summary Statistics for Changeups and Curveballs, 2022-2023

Changeups, 2022-2023	Summary Statistics					
	Min	Q1	Median	Mean	Q3	Max
release_speed	40.20	83.20	85.50	85.39	87.70	98.00
release_pos_x	−4.81	−2.07	−1.41	−0.56	1.50	4.75
release_pos_z	2.37	5.49	5.78	5.76	6.08	7.74
pfx_x	−3.54	−1.30	−1.01	−0.37	1.05	2.16
pfx_z	−1.25	0.24	0.51	0.50	0.77	6.52
vx0	−17.19	−4.28	4.31	1.95	6.69	19.36
vy0	−142.45	−127.59	−124.39	−124.17	−121.00	−57.42
vz0	−19.71	−5.68	−3.96	−3.96	−2.23	10.69
ax	−41.36	−14.99	−11.45	−4.51	11.41	25.64
ay	−85.46	22.85	24.65	24.66	26.46	84.98
az	−45.68	−29.22	−26.30	−26.43	−23.45	43.71
release_spin_rate	420.00	1,553.00	1,739.00	1,767.46	1,974.00	3,517.00
release_extension	3.10	6.10	6.40	6.40	6.70	11.20
spin_axis	1.00	130.00	228.00	198.24	242.00	360.00
delta_run_exp	−1.32	−0.07	0.00	0.00	0.04	3.61

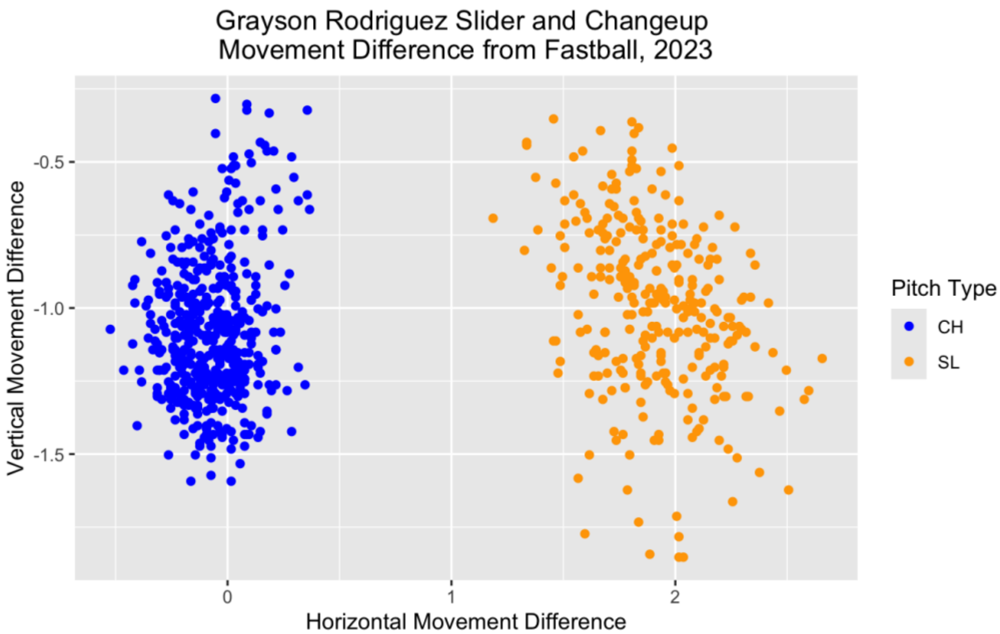
Curveballs, 2022-2023	Summary Statistics					
	Min	Q1	Median	Mean	Q3	Max
release_speed	39.40	77.30	79.90	79.61	82.20	92.80
release_pos_x	−4.95	−1.98	−1.39	−0.73	0.98	4.62
release_pos_z	2.55	5.65	5.96	5.93	6.24	7.52
pfx_x	−2.87	−0.19	0.54	0.35	0.94	2.23
pfx_z	−2.56	−1.12	−0.82	−0.76	−0.45	1.58
vx0	−94.03	−1.17	1.56	1.16	3.68	14.79
vy0	−134.88	−119.60	−116.22	−115.86	−112.51	−55.82
vz0	−70.28	−2.48	−0.69	−0.77	1.03	15.46
ax	−193.95	−1.60	4.31	2.93	7.90	19.55
ay	−337.56	21.20	22.88	22.92	24.60	58.68
az	−124.68	−42.11	−39.59	−38.96	−36.24	−16.07
release_spin_rate	37.00	2,318.00	2,515.00	2,517.99	2,710.00	3,544.00
release_extension	3.20	6.00	6.30	6.29	6.60	8.70
spin_axis	0.00	34.00	49.00	119.40	296.00	360.00
delta_run_exp	−1.35	−0.06	0.00	0.00	0.03	3.63

As shown in the tables, the average change in run expectancy is about zero for each individual pitch type. We can also make out some defining statistics for each pitch. For example, curveballs are characterized by lower velocity, higher spin, and sharper vertical break. This also demonstrates the importance of creating separate models for each pitch, because each one is significantly different in terms of movement and velocity. Note that since the data shown above does not make an absolute value adjustment for horizontal break, significant conclusions from the horizontal break data cannot be reached.

b. Exploratory Data Analysis

By isolating pitches into separate models, we are able to create a clearer picture of what makes certain pitches more effective. As an example, we can plot the differences in horizontal and vertical movement from Grayson Rodriguez's average fastball in the 2023 season. Note that both Rodriguez's slider and changeup both graded out well on Stuff+ models, though he was not a qualified pitcher.

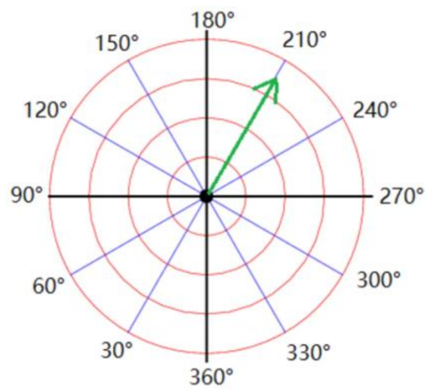
Figure 2: Grayson Rodriguez's 2023 Slider and Changeup Horizontal and Vertical Breaks



The graph above shows a clear grouping between the movement of Rodriguez's slider and changeup. They both have similar vertical break differences, but his changeup's horizontal movement is very similar to his fastball. This visual shows the importance of creating models for separate pitches, as they have very different movement profiles. Across all sliders and curveballs from the 2022 and 2023 seasons, sliders had over 2 times as much of a difference in horizontal break from the fastball as changeups did.

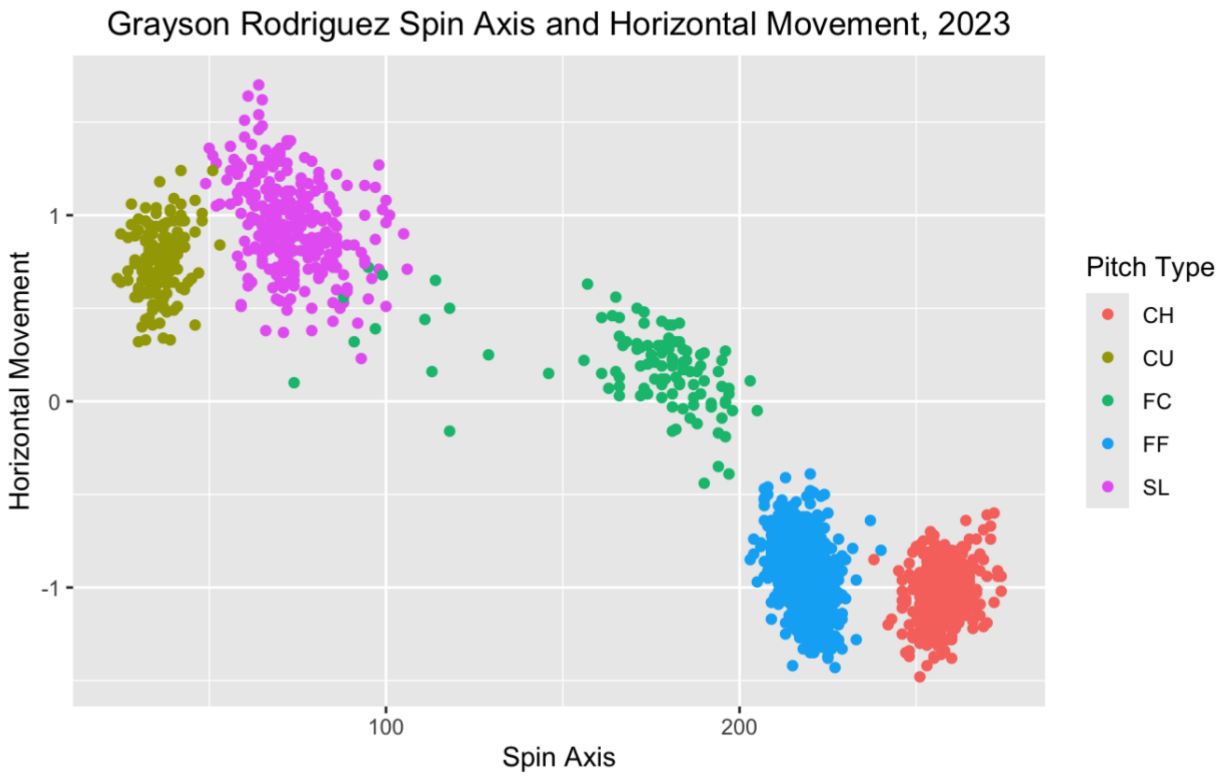
Another interesting relationship to examine is with a relatively new publicly available variable, spin axis. Spin axis measures the spin direction of the pitch. Most right-handed pitchers with a standard overhand or three-quarters delivery will have a spin axis of about 210 degrees, and for left-handed pitchers it would be around 150 degrees. Fastballs are thrown with backspin, so the spin axis will almost always be in between 90 and 270 degrees.

Figure 3: Spin Axis visualization (Driveline). The arrow represents an average RHP's fastball.



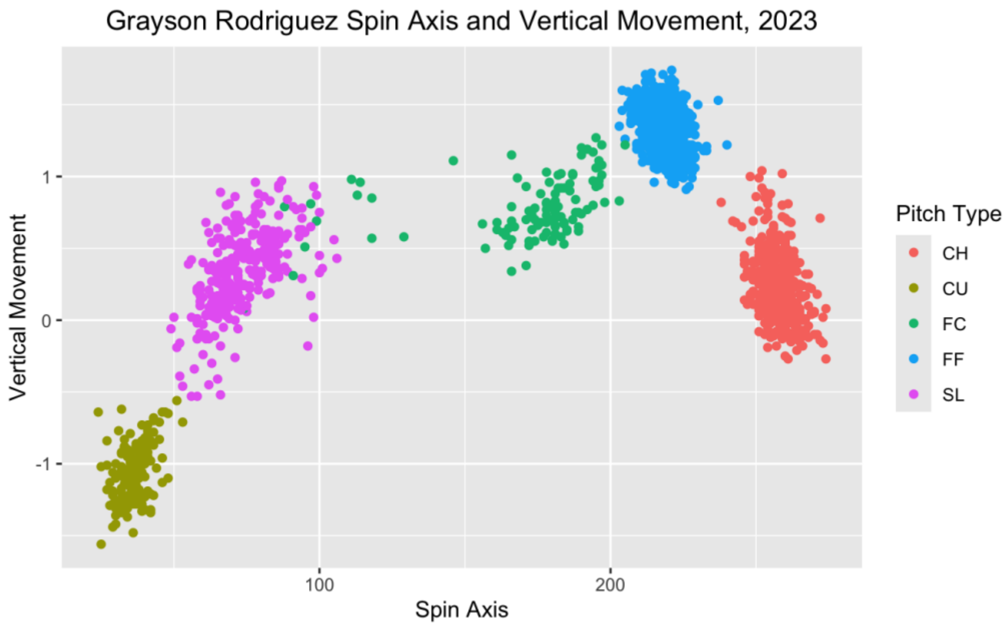
Anything more extreme than that would indicate topspin, the type of spin a pitcher applies to a ball when they want it to break. A pitch with a 0 degree spin axis would indicate "12-6" movement on a curveball. For example, Clayton Kershaw, famous for his big, looping curveball, had an average spin axis of about 340 degrees across the two seasons in the dataset. We can examine the relationship between spin and horizontal movement for Grayson Rodriguez in 2023.

Figure 4: Grayson Rodriguez Spin Axis and Horizontal Movement, 2023 Season



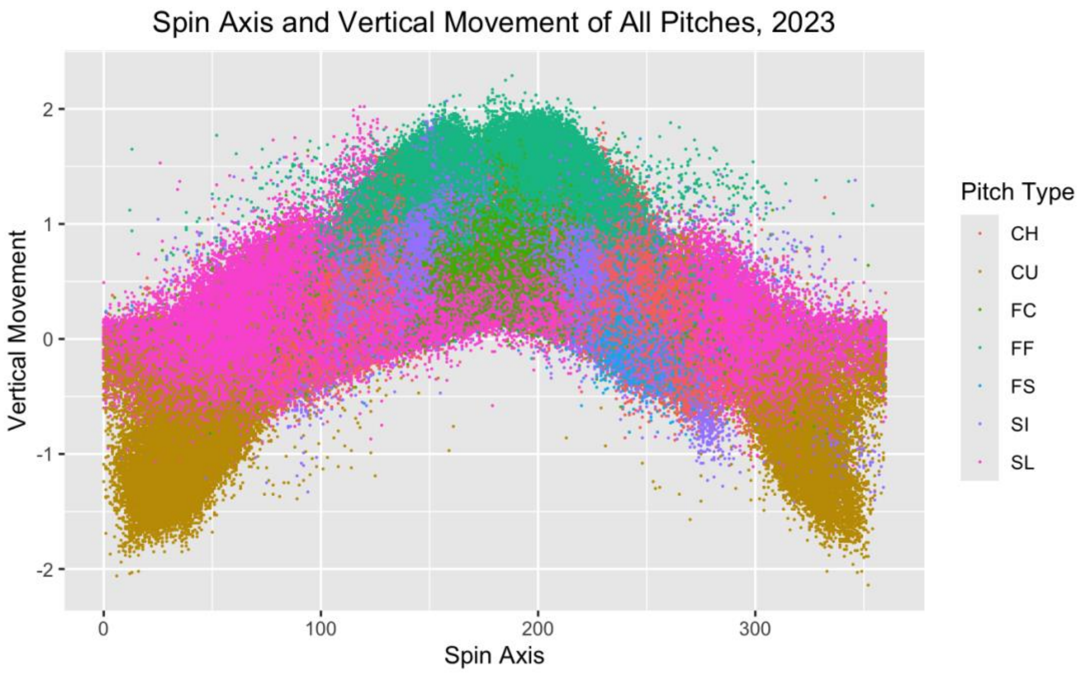
Pitches where topspin is induced, like curveballs and sliders, have much more horizontal movement than pitches like changeups and fastballs. This graph shows that Rodriguez's fastball and changeup both "tail", meaning they break towards the arm-side. His breaking balls have much different spin axes because of the pressure applied to the ball and wrist movement at release, and as such have a significantly higher amount of horizontal break. His cutter was an interesting pitch to observe as well. There are a few values close to his slider, which very well may have been misclassified. The cutter generates more horizontal movement than his fastball and changeup but not as much as the curveball and slider. We can also visualize Rodriguez's vertical movement with the chart below.

Figure 5: Grayson Rodriguez Spin Axis and Vertical Movement, 2023 Season



Examining the vertical movement also reveals some interesting patterns. Spin axis has a quadratic relationship with vertical movement. Pitches with topspin tend to have more downward vertical movement, where pitches with backspin generally have less “drop.” We can also visualize this by looking at all pitches in the 2023 season:

Figure 6: Spin Axis and Vertical Movement of All Pitches, 2023 Season



Pitches with direct backspin would be expected to break the least, where pitches with the most amount of topspin would break the most. A linear model of vertical break with spin axis as a quadratic term yielded an adjusted r-squared value of 0.6225, showing a strong relationship between those two terms. Color-wise, we can also see some symmetry, as the dataset includes both right-handed pitchers and left-handed pitchers.

With spin axis being a more prevalent measure, especially with publicly available datasets, it's important to highlight the relationship between spin axis and other factors of a pitch.

c. Modeling

To create a Stuff+ metric, a model to predict the change in run expectancy was created. The model would only consider location-independent variables, as the overall goal of a Stuff+

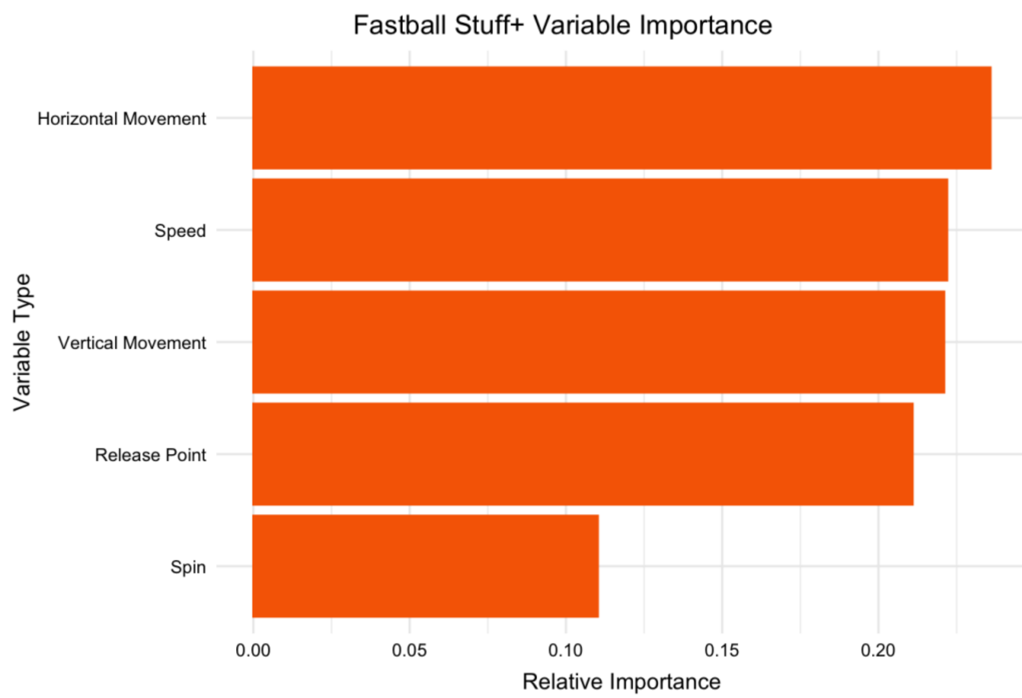
model is to grade pitch shape. Extreme Gradient Boosting (XGBoost) models were created for each pitch type. For fastballs, sinkers, and cutters, the variables used included horizontal movement, speed, vertical movement, release point, and spin. For non-fastball pitches, the difference in all categories between the pitcher's average fastball and the pitch they just threw were also considered, to determine the effect in which the off-speed pitch depends on the fastball. The predictor variable in all models was change in run expectancy. Hyperparameters for each model were tuned using a grid search and 5-fold cross validation. Models were trained on data from the 2022 season, and applied to data from the 2023 season. To calculate the Stuff+ number, the predicted change in run expectancy was scaled to where the median value was 100. Anything above 100 would be positive for the pitcher, while anything under 100 would indicate below-average stuff.

d. Results

7 different models were created, and the variable importance was extracted from each one. Each graph displays a combination of variables that contribute to its label. For example, "horizontal movement" displays the combined importance of velocity in the horizontal axis, acceleration in the horizontal axis, and movement in the horizontal axis. Speed is comprised of release speed and velocity and acceleration towards home plate. Vertical movement combines the importance from velocity in the vertical axis, acceleration in the vertical axis, and movement in the vertical axis. Release point combines the importance of the release point variables in both axes as well as extension, which measures how far out in front of the plate the pitcher is releasing the ball from. Finally, spin combines the importance of spin rate and spin axis. Note that these variables were created strictly for the purposes of visualizing variable importance, and

need further examination to understand the relationship between them and the calculated Stuff+ measure. The variable importance plot for the fastball model is displayed below.

Figure 7: Variable Importance for Fastball Model

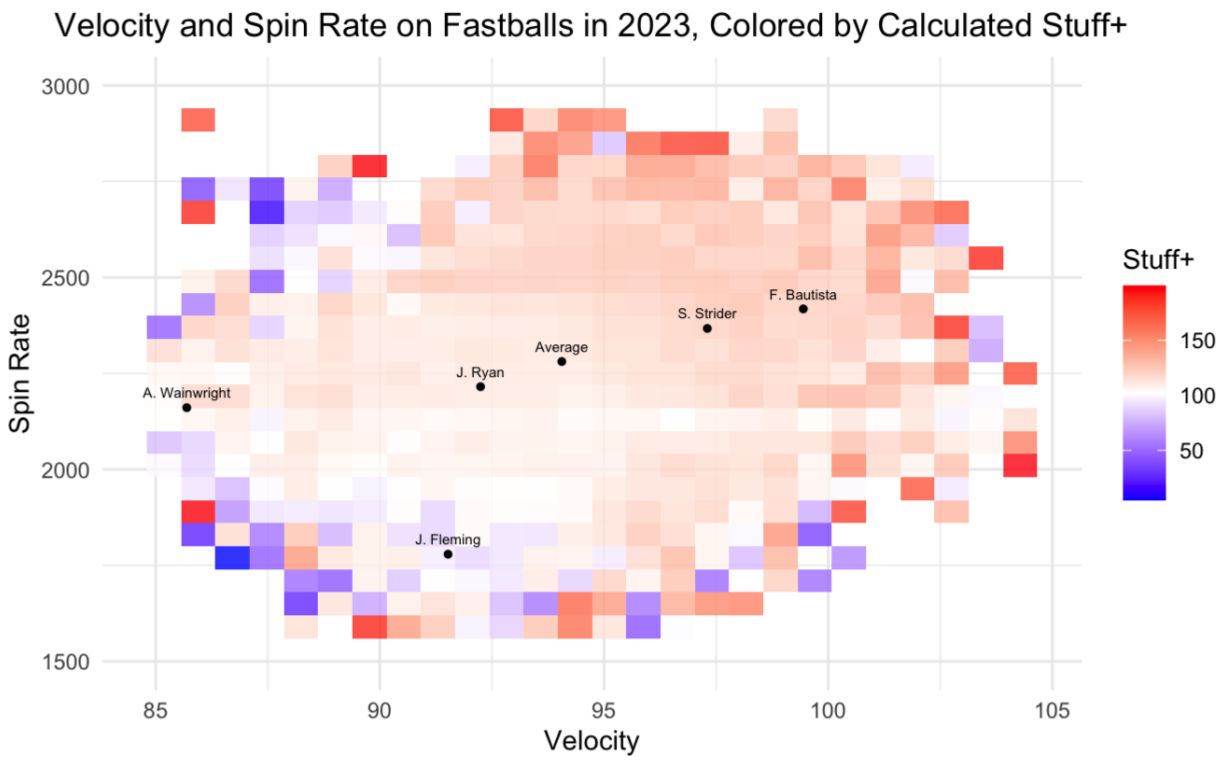


For the fastball model, individual variables that were of high importance included velocity in the horizontal plane, release point, and spin rate. Relative to other pitches, velocity for fastballs was important, and while it wasn't a make-or-break element, many hard-throwing pitchers graded out well on this model. Notable pitchers included Félix Bautista (Stuff+ grade in 2023: 146), Spencer Strider (132), and Zack Wheeler (124). However, other pitchers who did not throw as hard in 2023 were still able to grade out well because of other elements of the pitch. For example, Zac Gallen (124 FF Stuff+, 40th percentile FF velo), Alex Vesia (144, 54th), and Joe Ryan (128, 23rd) also graded out well on this model despite not having top-tier fastball velocity.

This shows that the model is effectively capturing all elements of the pitch. For example, Joe Ryan possesses fairly average velocity and spin rate, but his fastball has more arm-side movement than the average right-handed pitcher. Variables that impact pitch shape, like velocity and acceleration in each axis, can also contribute to higher grades for pitchers that do not throw as hard.

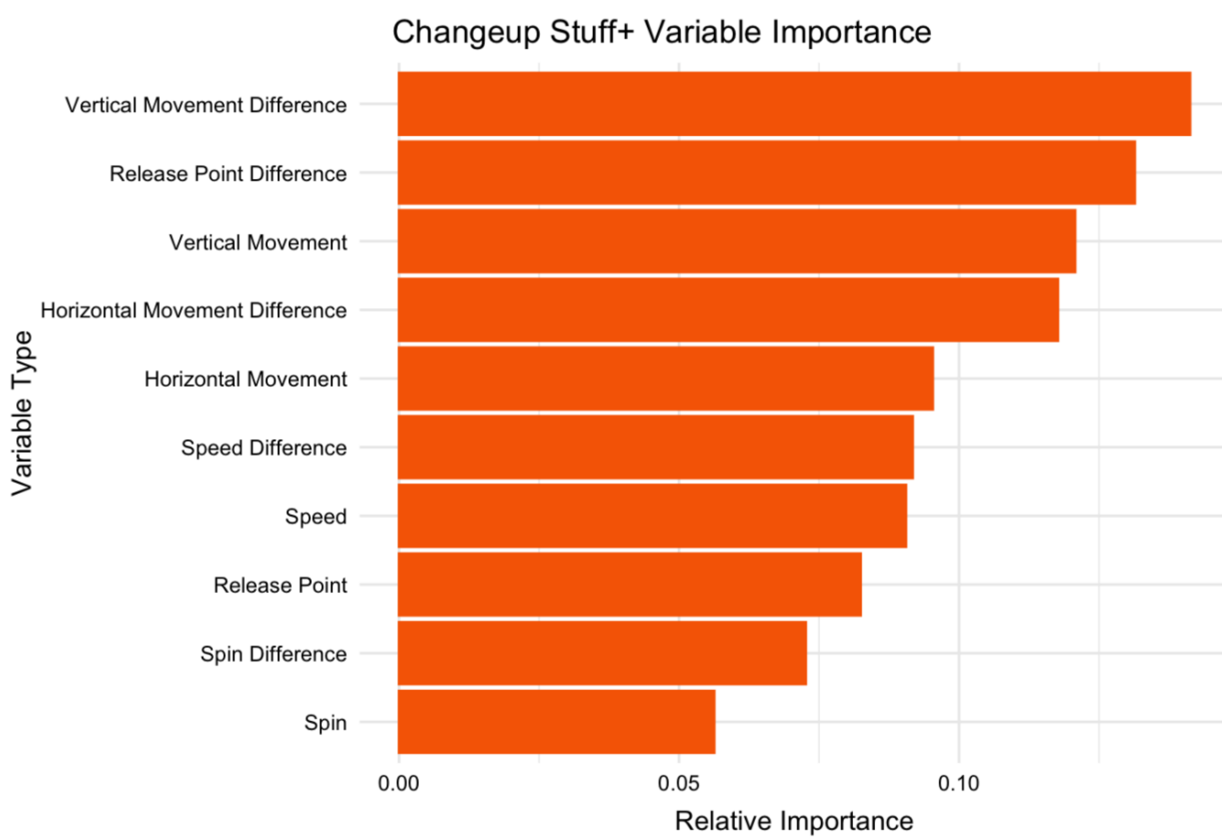
It is uncommon, however, for pitchers to have a high Stuff+ fastball without elite velocity or spin. Many pitchers that grade out well on a Stuff+ model generally throw harder than pitchers that do not, which we can visualize with the graph below, which is color-coded by the Stuff+ rating returned by the model. This allows us to visualize the relationship between velocity and spin rate on fastballs, and which combinations can lead to better Stuff+ grades. Generally, higher velocity and higher spin are correlated with higher Stuff+ grades. The bottom edges generally receive lower Stuff+ grades. On the graph, we can see the average velocity and spin rate on 2023 fastballs, while each labeled pitcher also shows their average velocity and spin rate on their fastball in 2023. “Power” pitchers like Spencer Strider and Félix Bautista feature high-velocity fastballs with above-average spin, while pitchers like Joe Ryan have fastballs that don’t necessarily need to rely on velocity to be successful.

Figure 8: Velocity and Spin Rate on Fastballs in 2023, Colored by Calculated Stuff+



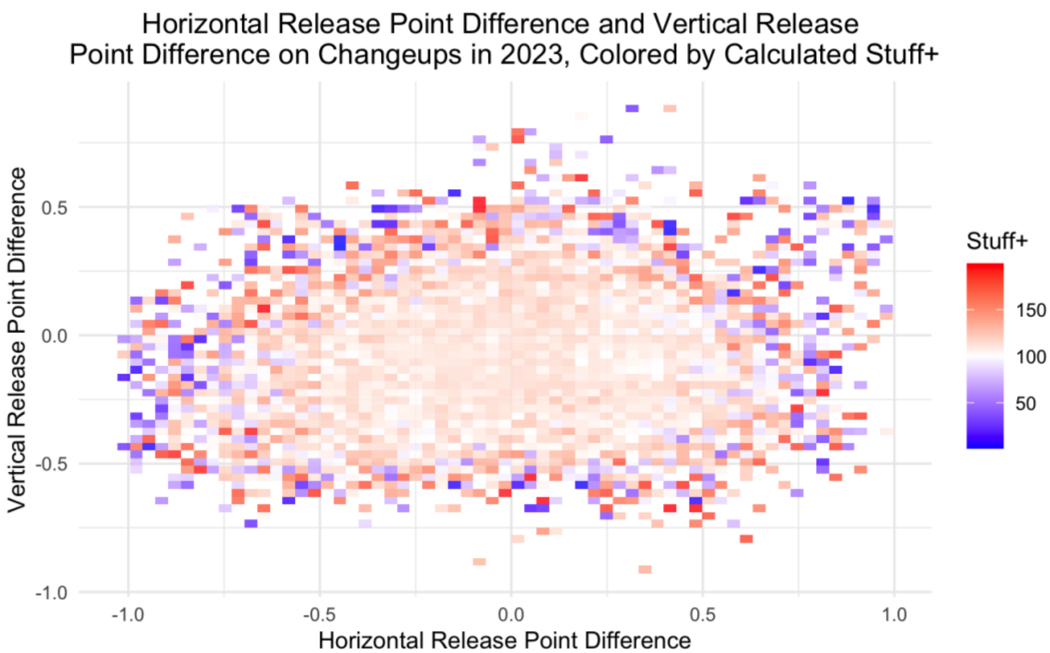
For off-speed pitches, a different approach was used. The models also considered the difference between the pitch that was just thrown and the pitcher's average fastball metrics for that season. So if a pitcher were to throw an 88 mph changeup, and their fastball's average velocity is 98 mph, the difference would be 10 mph. If a pitcher did not throw a 4-seam fastball, the model used their sinker for the differentials. For any pitcher that did not throw either, it used the differentials from the pitcher's cutter. The differentials as well as the raw values were both considered in the models for breaking balls. Using this technique can help us determine the extent to which off-speed pitches "depend" on the primary pitch. We can see that the difference in metrics from the fastball carry slightly more importance than the raw values. Vertical and horizontal movement, release point, and speed differential were all important variables in the model. Note that for a changeup, the vertical and horizontal movement don't have the same magnitude as the movement on breaking balls, but it still matters. As shown in Figure 2, changeups often don't have a significant difference in horizontal movement from their fastball. Most changeups often move in a different direction than the pitcher's breaking pitches, so isolating these pitches into separate models can provide a clearer picture of what makes a changeup "good." The variable importance plot for changeups is shown below.

Figure 9: Variable Importance for Changeup Model



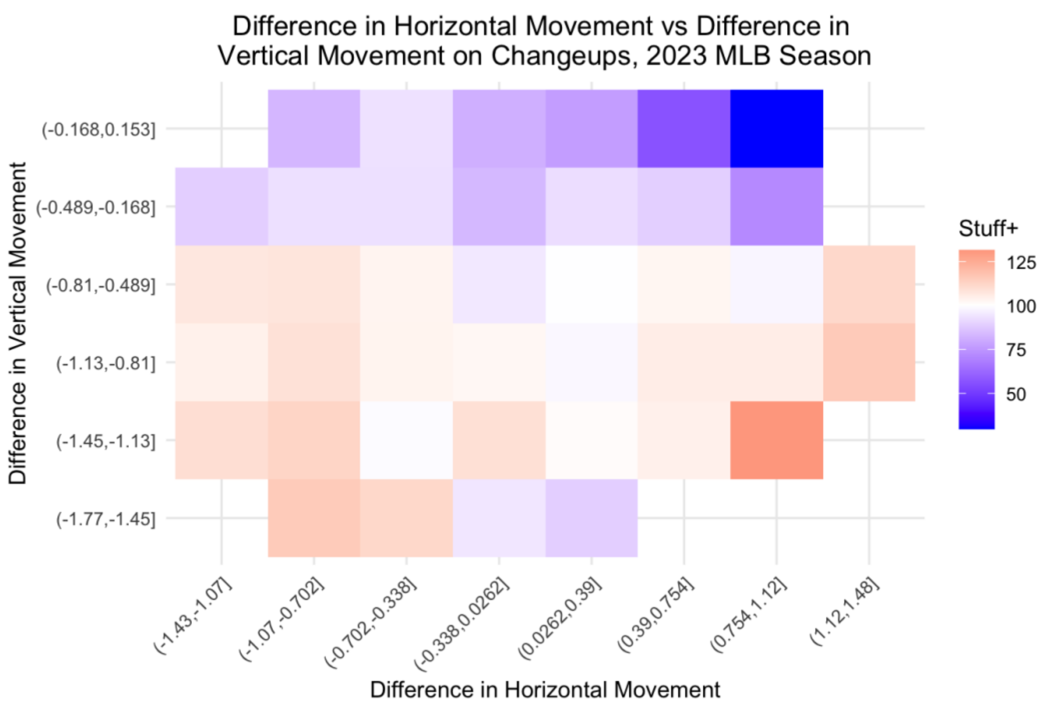
Another relationship to examine is the importance of consistency with release point for a pitcher, especially with a changeup, which is similar to a fastball in terms of movement. We can achieve this by graphing the horizontal and vertical release point differences from a pitcher's fastball on their changeup, and color code it by the Stuff+ value returned by the model. The center of the graph (0, 0) would represent a pitch thrown with the release point being consistent with the pitcher's average release point on the fastball. As we get further and further away from the center, we can see that pitches that deviate significantly from the pitcher's average receive lower Stuff+ grades more often.

Figure 10: Horizontal and Vertical Release Point Differences and Stuff+ on 2023 Changeups



Another relationship we can visualize with the changeup is the difference in movement in both axes. The graph below displays the average Stuff+ in each bucket, where each bucket is comprised of a certain range of the difference in horizontal and vertical movement. Pitches with lower differences in horizontal and vertical movement received lower Stuff+ grades, while pitches with higher differences received higher Stuff+ grades. We can see that the model accounts for both lefties and righties, as high movement in either direction is correlated with higher Stuff+ grades. Many pitchers known for high velocity grade out well on this model, including Grayson Rodriguez, Edward Cabrera, Shane McClanahan, Sandy Alcántara, and Brayan Bello. However, other pitchers that are not necessarily known for high velocity but still throw effective changeups also have higher Stuff+ grades, those pitchers include Trevor Richards, Merrill Kelly, Steven Matz, and Michael Wacha.

Figure 11: Horizontal and Vertical Movement Differences and Stuff+, 2023 Changeups



By creating separate models for each pitch, we are able to isolate variables that may have a higher effect on run expectancy for one pitch as opposed to another. In addition to fastballs and changeups, there are also other interesting trends for other pitches. The most important variable for cutters was horizontal movement, and spin rate was significantly more impactful in the cutter model than other pitches. Changeups were most dependent on differences from the fastball rather than raw values compared to other pitches. Vertical movement was one of the most impactful variables for the splitter model, and for sliders, it was horizontal movement that had a significant impact on predicted values. For the curveball, the raw values carried much more importance than the differences between the fastball, suggesting that curveballs are less dependent on the pitcher's fastball to be deceptive. Variable importance plots for all models are included in the appendix.

V. Conclusion

This analysis attempts to break down pitches on a pitch type level to identify which characteristics of a certain pitch type lead to success most often. Tree-based modeling is just one of many different potential approaches to this question, but it can lead to effective results. It can allow us to isolate specific variables and determine their importance, as well as examine the relationships between pitch attributes and predicted changes in run expectancy. Quantifying Stuff+ at the individual pitch type level can also provide insights to potentially help improve pitcher performance. By isolating the importance of individual variables and visualizing their relationships, we can provide valuable insight on which specific attributes might be lacking for a certain pitcher and recommend improvements.

As one might intuitively expect, velocity and spin rate are of high importance for four-seam fastball performance. Pitches characterized by horizontal movement, like sliders and cutters, generally perform better with an increased magnitude of horizontal movement. For splitters, vertical movement was an important variable; pitchers with a "disappearing" splitter/forkball, like Kodai Senga, graded out well on this model. For curveballs, the raw values of the pitch carried more importance than the differences, more so than any other off-speed pitch, which was also interesting. There are many different curveball “shapes” in baseball, which can be an explanation for this pattern. Some pitchers throw a “12-to-6” curveball, which breaks almost straight down, whereas a standard curveball’s movement is more horizontal. Generally, more break improves the pitch quality, at least to a certain extent. For example, a position player taking the mound and throwing an eephus would have a lot of “break”, but it would not grade out well on a Stuff+ model because of its other attributes like speed and spin rate.

Overall, the model was generally effective in capturing which pitchers had the best stuff, but improvements can always be made. For future research, having access to a larger dataset will help to improve predictive power of the models. Additionally, model performance can be improved through further tuning of hyperparameters, to improve the accuracy of each individual pitch model. Another trend that will be interesting to examine is the rise of “sweepers.” In the future, a separate model for pitches classified as sweepers can improve the strength of both the slider model as well as the overall Stuff+ calculation. Another potential improvement would be to work on handling more extreme changes in run expectancy. While these calculations are sparse, there are a few predicted changes in run expectancy that lead to extremely high or low Stuff+ ratings, which can make drawing valuable conclusions from visualizations difficult. As pitchers accumulate a higher sample of pitches, extremities balance out, but finding a way to prevent these from happening to begin with can lead to more valuable insights. This could be accomplished by removing values that fall outside of a predetermined number of standard deviations away from the mean, though this would also theoretically limit how good (or bad) a pitcher’s Stuff+ rating could be.

As baseball research develops, Stuff+ will continue to become a valuable predictive metric to help organizations identify undervalued pitchers. While it is still a fairly new development in baseball analytics, it has the potential to improve in the public space as more tracking data becomes publicly available.

VI. Appendix

Figure 12: Fastball Model Variable Importance Plot

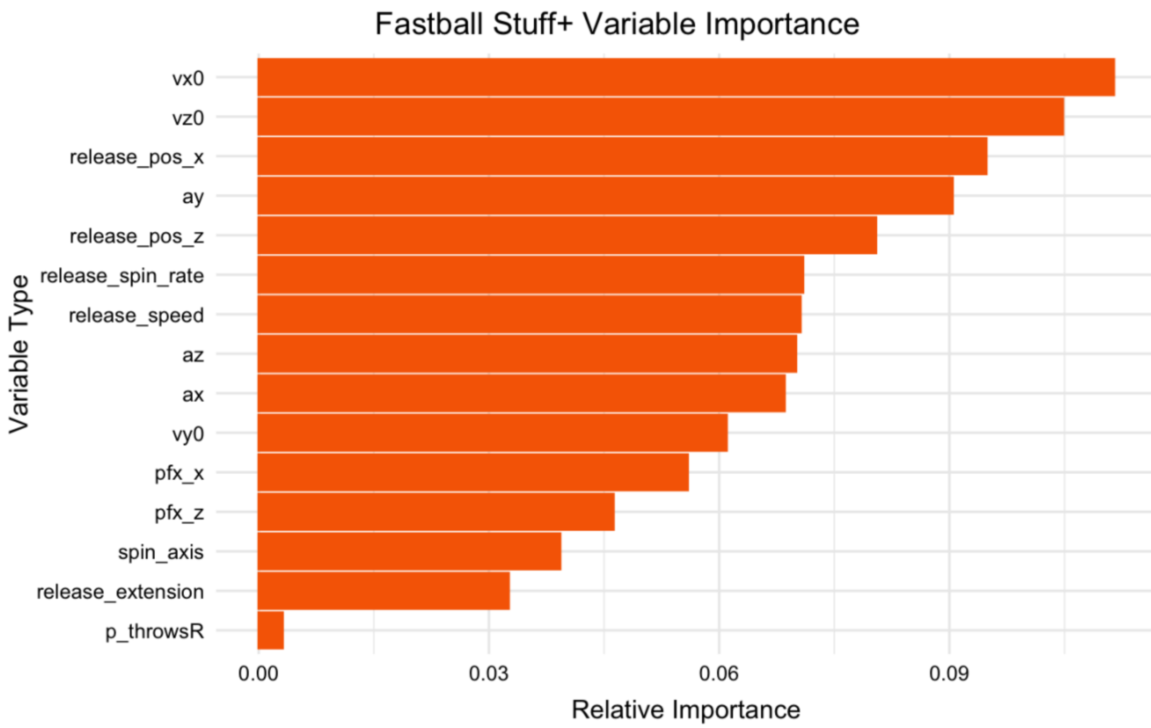


Figure 13: Changeup Model Variable Importance Plot

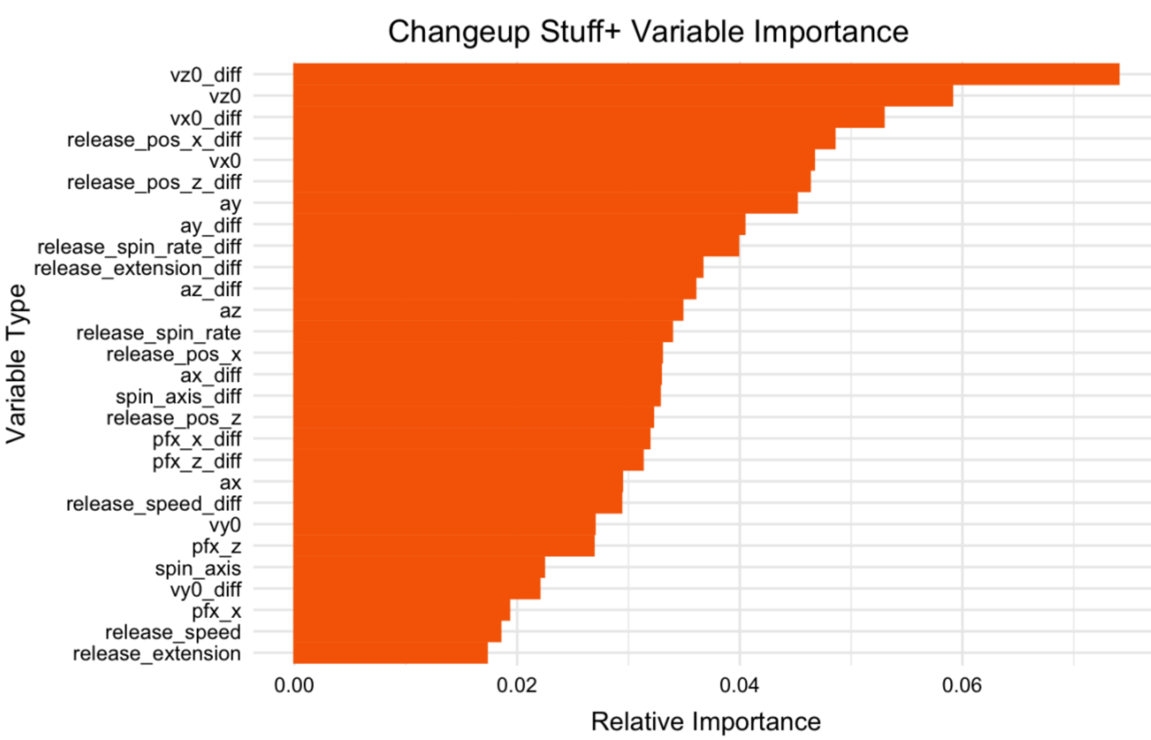


Figure 14: Sinker Model Variable Importance Plot

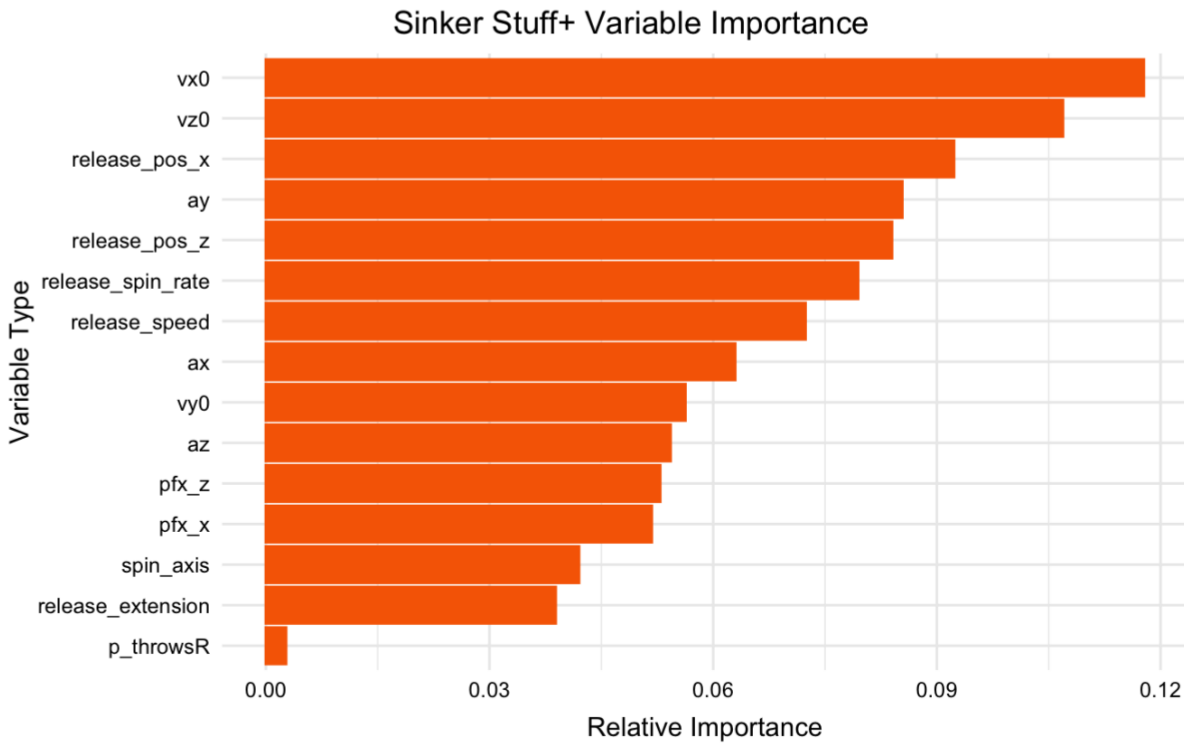


Figure 15: Cutter Model Variable Importance Plot

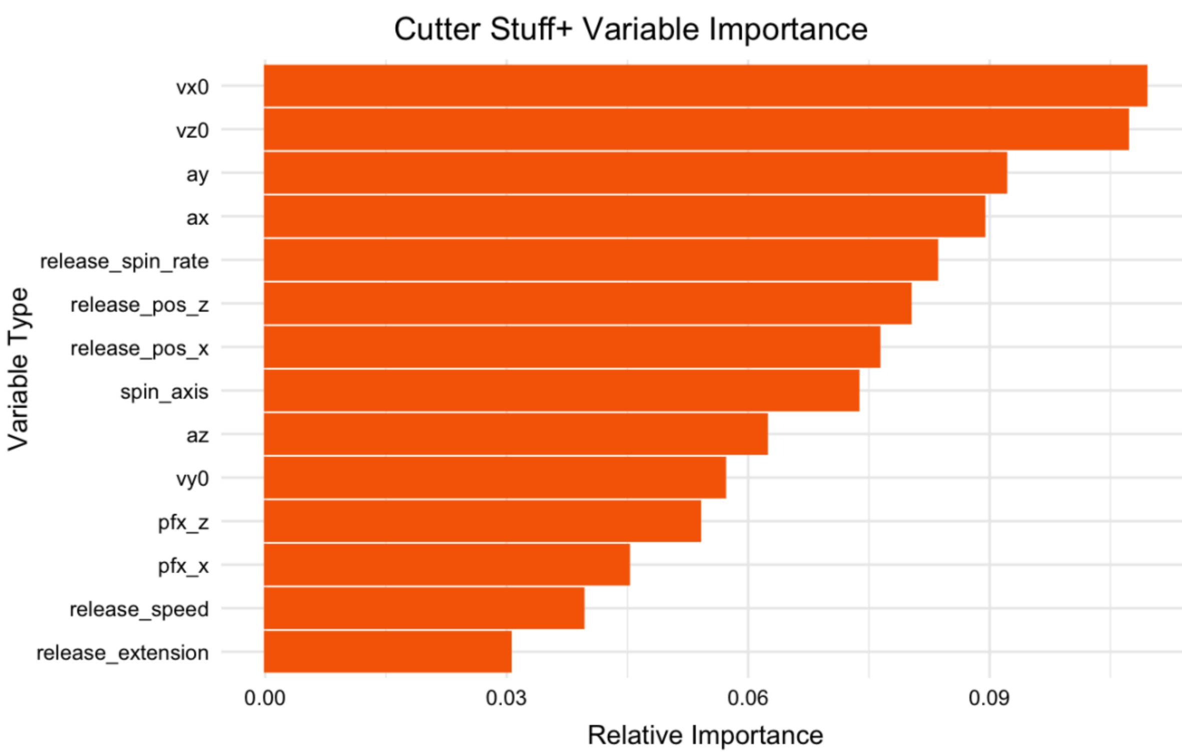


Figure 16: Slider Model Variable Importance Plot

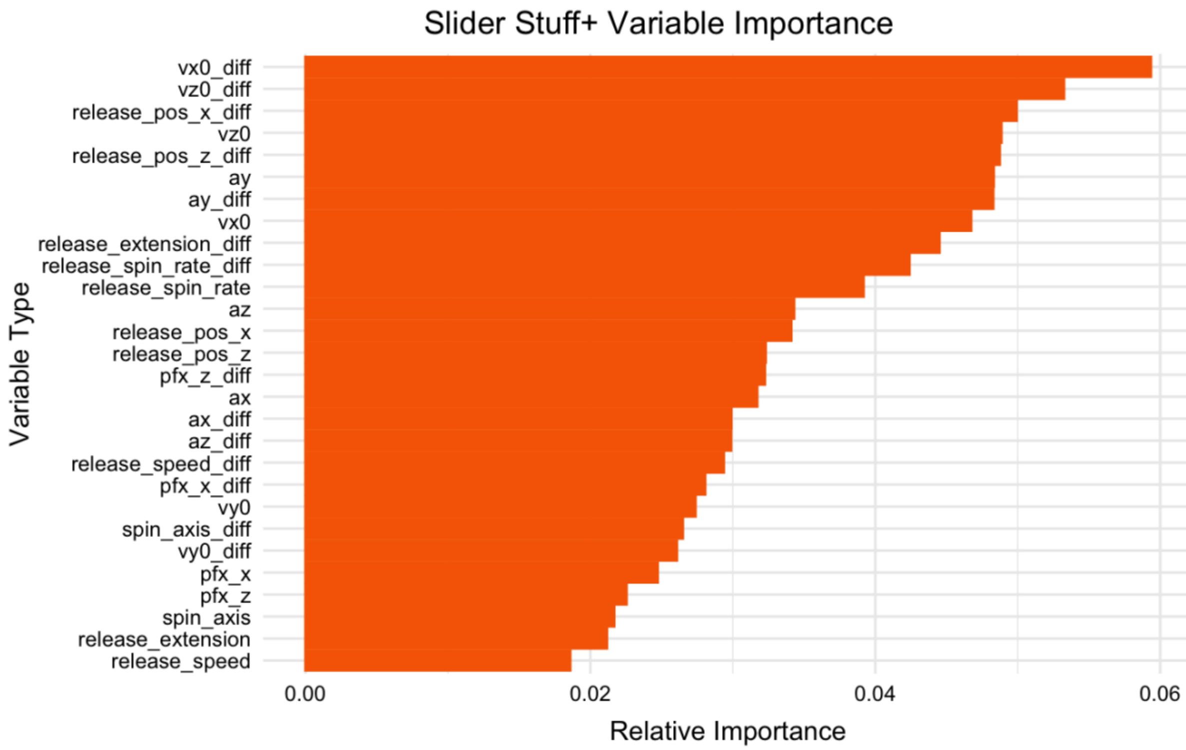


Figure 17: Curveball Model Variable Importance Plot

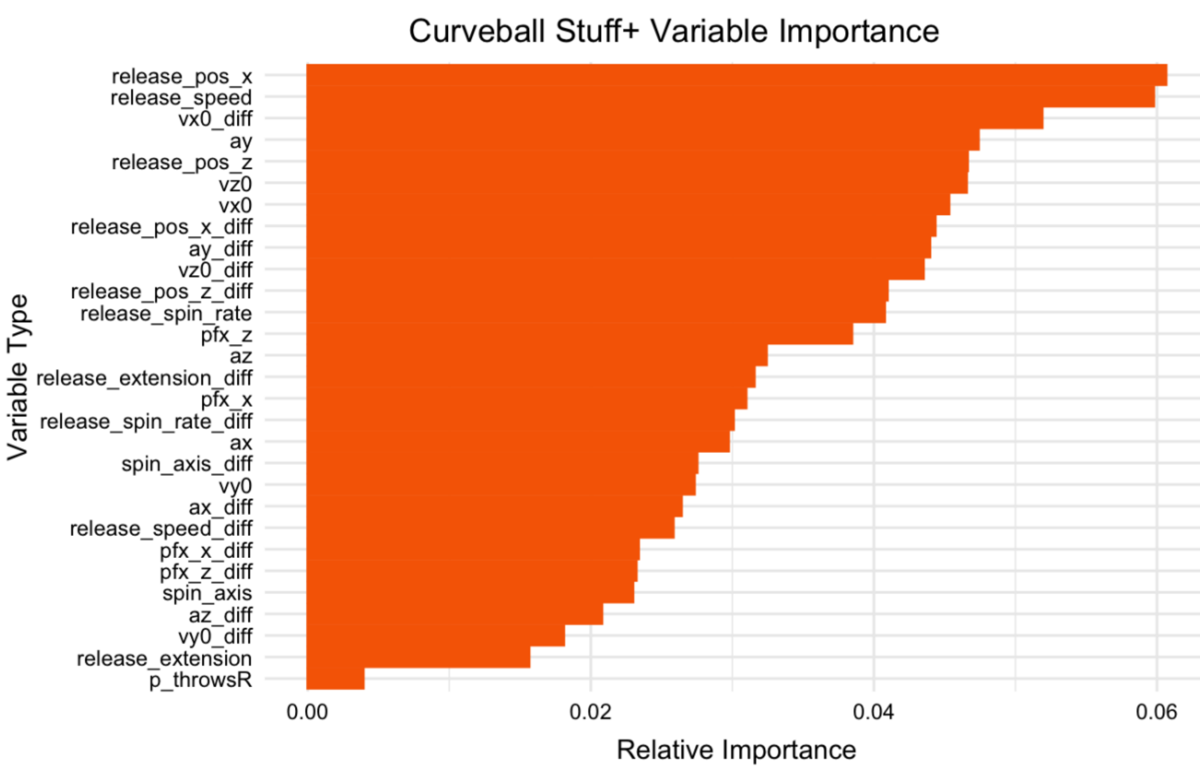
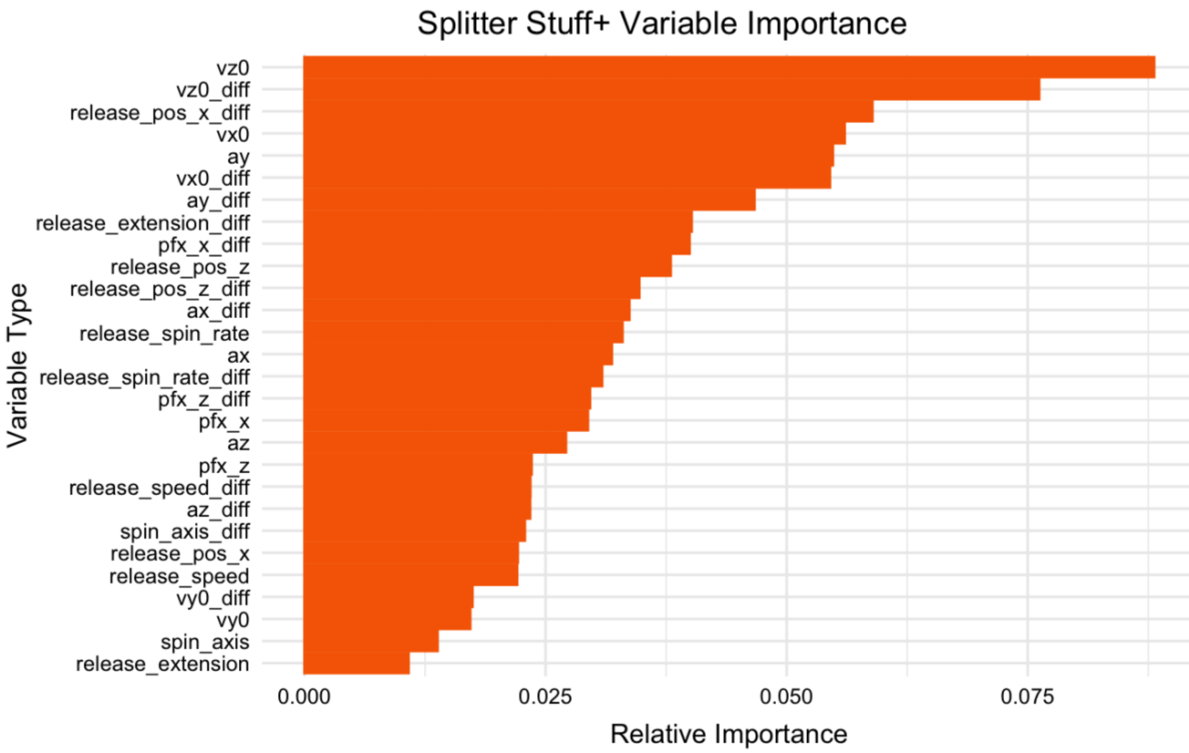


Figure 18: Splitter Model Variable Importance Plot



VII. References

. Intelligent Autonomous Systems 12: Volume 2 Proceedings of the 12th International Conference IAS-12, held June 26-29, 2012, Jeju Island, Korea (Advances in Intelligent Systems and Computing, 194). 2013, Springer, 2012.

Appelman D.. "PitchingBot and Stuff+ Pitch Modeling Is Now on FanGraphs!" FanGraphs Baseball, 2023, March 10, https://blogs.fangraphs.com/pitchingbot-and-stuff-pitch-modeling-are-now-on-fangraphs/.

Authors. "Baseball Savant: Statcast, Trending MLB Players and Visualizations" baseballsavant.com, https://baseballsavant.mlb.com/.

Baseball D.. "Pitch Design: What is Stuff+? Quantifying Pitches with Pitch Models" Driveline Baseball, 2021, December 14, https://www.drivelinebaseball.com/2021/12/what-is-stuff-quantifying-pitches-with-pitch-models/.

Blewett D.. "The Modern Changeup: The Best New Pitch In Baseball" ['Elite Baseball Performance', 'Elite Baseball Performance'], https://elitebaseballperformance.com/modern-changeup-best-new-pitch-baseball/.

Bock, Joel. "Pitch Sequence Complexity and Long-Term Pitcher Performance" Sports, vol. 3, no. 1, 2015, pp. 40-55, https://doi.org/10.3390/sports3010040.

Cochran M.. "How to Use Data to Throw An Effective Changeup" Rapsodo, 2023, May 9, https://rapsodo.com/blogs/baseball/how-to-throw-a-changeup.

DriveLine Baseball. "Mastering the Axis of Rotation: A Thorough Review of Spin Axis in Three Dimensions." DriveLine Baseball, September 2019, https://www.drivelinebaseball.com/2019/09/mastering-the-axis-of-rotation-a-thorough-review-of-spin-axis-in-three-dimensions/.

Fleisig, Glenn S., et al.. "Kinetic Comparison among the Fastball, Curveball, Change-up, and Slider in Collegiate Baseball Pitchers" The American Journal of Sports Medicine, vol. 34, no. 3, 2006, pp. 423-430, https://doi.org/10.1177/0363546505280431.

Glanzer, Joshua A., et al.. "The relationship between variability in baseball pitching kinematics and consistency in pitch location" Sports Biomechanics, vol. 20, no. 7, 2019, pp. 879-886, https://doi.org/10.1080/14763141.2019.1642378.

Healey, Glenn, and Shiyuan Zhao. "Using PITCHf/x to model the dependence of strikeout rate on the predictability of pitch sequences" Journal of Sports Analytics, vol. 3, no. 2, 2017, pp. 93-101, https://doi.org/10.3233/jsa-170103.

Jordan Rosenblum (2022, February 26). Incorporating Stuff+ Into Traditional Pitching Projections. https://www.rotoballer.com/pitching-projections-using-stuff-metrics/995400

Kawamura, Katsue, et al.. "Baseball pitching accuracy: an examination of various parameters when evaluating pitch locations” Sports Biomechanics, vol. 16, no. 3, 2017, pp. 399-410, https://doi.org/10.1080/14763141.2017.1332236.

KIM, Hyunuk, and Woo-Sung JuUNG*. "Does Pitch Type - Zone Uncertainty Matter to a Pitcher's Performance?” New Physics: Sae Mulli, vol. 68, no. 6, 2018, pp. 624-629, https://doi.org/10.3938/npsm.68.624.

Kusafuka, Ayane, et al.. "Control of Accuracy during Movements of High Speed: Implications from Baseball Pitching” Journal of Motor Behavior, vol. 54, no. 3, 2021, pp. 304-315, https://doi.org/10.1080/00222895.2021.1960789.

Luke Hauswirth. "The Importance of Spin Rate in Baseball Pitching — Pitching Haus" Pitching Haus, 2023, February 27, https://www.pitchinghaus.com/blog/What -is-a-Good-Spin-Rate-for-Baseball-Pitching.

Mailhot J.. "Devin Williams and the Unicorn Changeup" FanGraphs Baseball, 2020, September 3, https://blogs.fangraphs.com/devin-williams-and-the-unicorn-changeup/.

Manzi, Joseph E., et al.. "Kinetic and kinematic comparisons in high school pitchers with low and high pitch location consistency” Journal of Shoulder and Elbow Surgery, vol. 31, no. 12, 2022, pp. 2620-2628, https://doi.org/10.1016/j.jse.2022.06.011.

Nakashima, Hirotaka (2020) "CHARACTERISTICS OF A WELL-DONE BREAKING PITCH IN BASEBALL," ISBS Proceedings Archive: Vol. 38: Iss. 1, Article 95. Available at: https://commons.nmu.edu/isbs/vol38/iss1/95

Nasu, Daiki, and Makio Kashino. "Impact of each release parameter on pitch location in baseball pitching" Journal of Sports Sciences, vol. 39, no. 10, 2020, pp. 1186-1191, https://doi.org/10.1080/02640414.2020.1868679.

Nicholson, K.F., et al.. "Machine learning and statistical prediction of fastball velocity with biomechanical predictors" Journal of Biomechanics, vol. 134, no. Mar, 2022, pp. 110999, https://doi.org/10.1016/j.jbiomech.2022.110999.

Pavlidis, H. "What Makes a Good Changeup?: An Investigation, Part Three" Baseball Prospectus, 2013, August 30, https://www.baseballprospectus.com/news/article/21675/what-makes-a-good-changeup-an-investigation-part-three/.

Platt, Brooks N., et al.. "Association Between Pitch Break on the 4-Seam Fastball and Slider and Shoulder Injury in Major League Baseball Pitchers: A Case-Control Study" Orthopaedic Journal of Sports Medicine, vol. 9, no. 10, 2021, pp. 232596712110389, https://doi.org/10.1177/23259671211038961.

RPP. "What is Stuff+ and How Can it Help You?" RPP Baseball, 2023, July 31, https://rocklandpeakperformance.com/what-is-stuff-and-how-can-it-help-you/.

Shinya, Masahiro, et al.. "Pitching form determines probabilistic structure of errors in pitch location” Journal of Sports Sciences, vol. 35, no. 21, 2017, pp. 2142-2147, https://doi.org/10.1080/02640414.2016.1258484.

Swartz, Philippa, et al.. "The Quality of Pitches in Major League Baseball” The American Statistician, vol. 71, no. 2, 2017, pp. 148-154, https://doi.org/10.1080/00031305.2016.1264313.

The finger movement and finger pressure in baseball pitching : A case ... (n.d.). https://www.researchgate.net/profile/Shu-Wei-Chen/publication/263686961_The_Finger_Movement_and_Finger_Pressure_in_Baseball_Pitching_A_Case_Report/links/0a85e53baaced5004f000000/The-Finger-Movement-and-Finger-Pressure-in-Baseball-Pitching-A-Case-Report.pdf

Tieran, Alexander. "The Mystic Art of Pitch Tunneling — Prospects Live" Prospects Live, 2022, August 23, https://www.prospectslive.com/prospects-live/2022/8/23/the-mystic-art-of-pitch-tunneling.

Whiteside, D., et al.. "Ball flight kinematics, release variability and in-season performance in elite baseball pitching” Scandinavian Journal of Medicine & Science in Sports, vol. 26, no. 3, 2015, pp. 256-265, https://doi.org/10.1111/sms.12443.

Yoshihara K., & Takahashi K. (2020). Pitch Sequences in Baseball: Analysis Using a Probabilistic Topic Model. SSRN Electronic Journal, 10.2139/ssrn.3728430