Pitcher Analysis With R: Gerrit Cole

Background

The analytics revolution has changed baseball forever; much of both in-game decisions and long-term strategy is now driven by data and analytics. With this in mind, I decided to do my own analysis of Yankees' ace Gerrit Cole using R.

Data Collection

The data for this analysis comes from the MLB's Baseball Savant database, which includes all the pitch-by-pitch data gathered by statcast (the MLB's data tracking system). To access the data, I utilized the baseballr package, which is a nice wrapper in R around the Baseball Savant API.

After extracting the data from Baseball Savant and storing it on my local machine using R, I wrote code to clean and analyze the data. All the code is available in the GitHub repo for this project.

Analysis

Strike Zone Plots
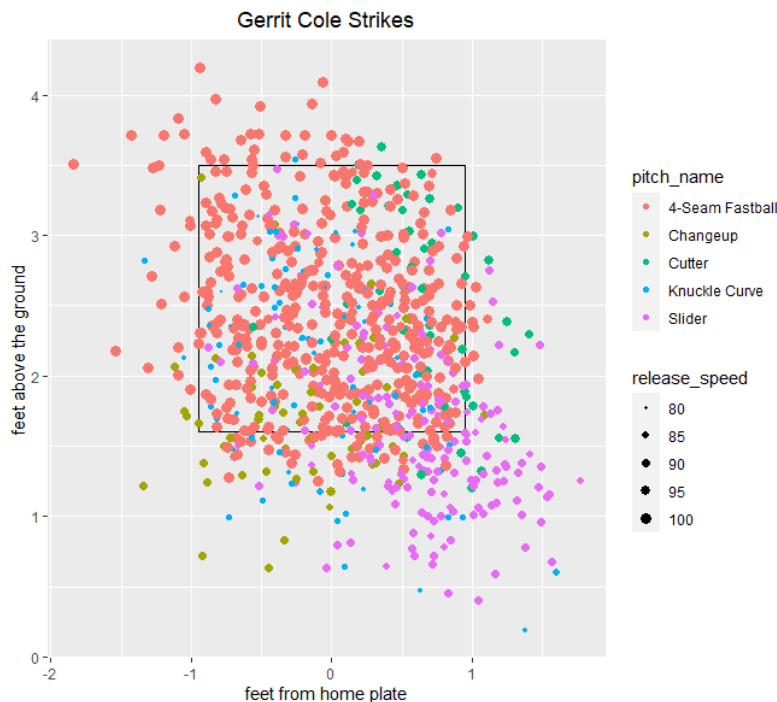
I have plotted Gerrit Cole's 2022 thrown strikes using ggplot2 in R. Each dot represents an individual pitch and is color coded by the type of pitch. Furthermore, the size of the dot is determined by release speed, with larger dots corresponding to faster pitches.

Fastballs tend to be around 95-100 mph and are mostly in the zone, with some being thrown to the left and above the zone. Cole's sliders are slower and almost always in the bottom right of the zone or even outside of it in the same direction. Changeups are almost always in the bottom of the zone, while cutters are usually in the right of the zone. Cole's knuckle curve is usually in the middle of the zone.

Gerrit Cole Pitches Hit Into Play
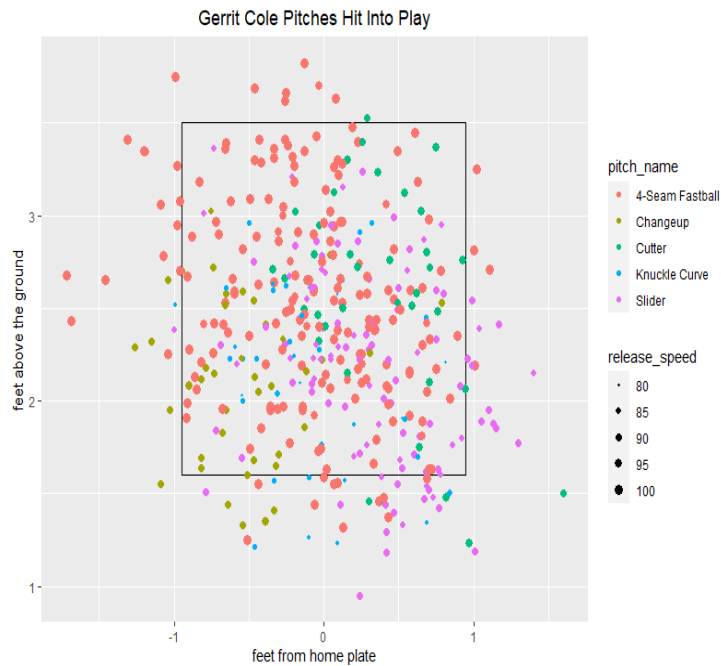
To the left is the same plot as above, except this time pitches that were hit into play are displayed.

Most of Cole's pitches that are hit into play are clustered in the center. Additionally, balls hit into play are consistently slower than his strikes.

Average Fastball Velocity Vs Pitches Thrown

Gerrit Cole Average Fastball Velocity vs Total Pitches
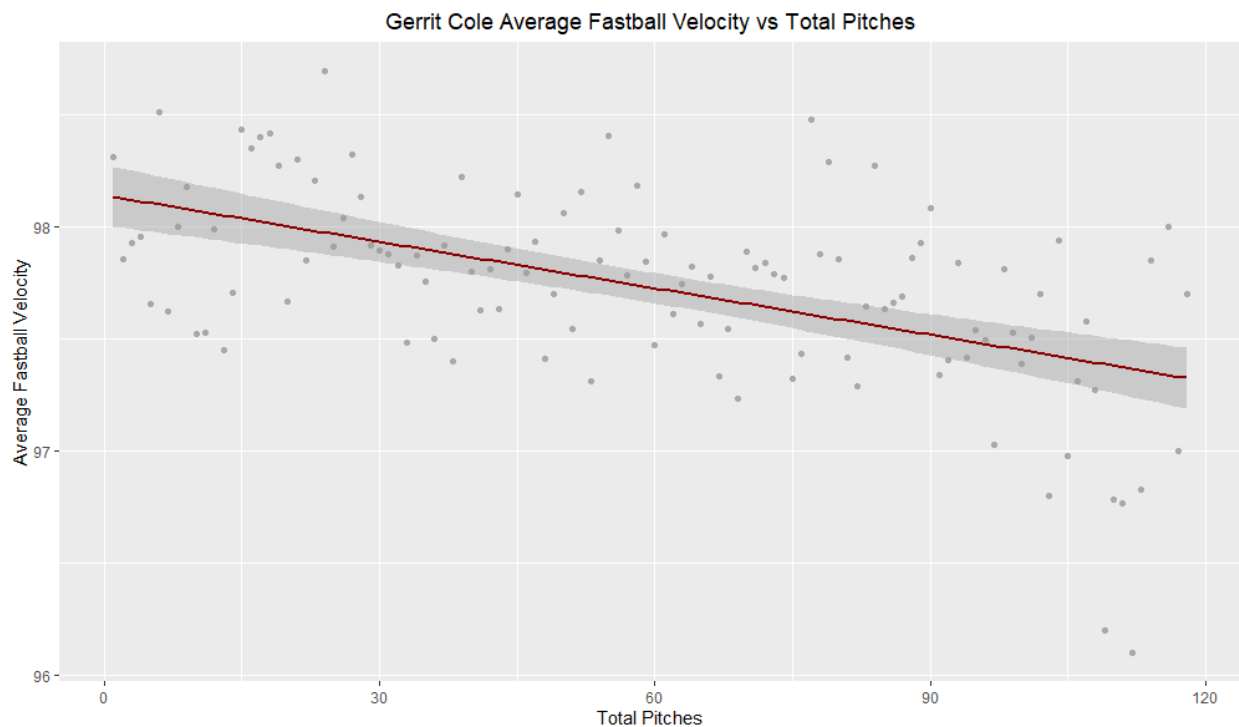
Like most pitchers, as the number of pitches thrown in a game increase, the average fastball velocity decreases. From the plot above we can see that the average velocity of Cole's four seam fastball steadily decreases as the game goes on. Once Cole gets past 90 pitches, the variance of the average fastball velocity increases greatly (it displays heteroskedasticity). Not only does Cole's average fastball velocity decrease as the game goes on, but it gets far more inconsistent past 90 pitches. It may make sense to pull him from the game at 90 pitches instead of the usual 110-120
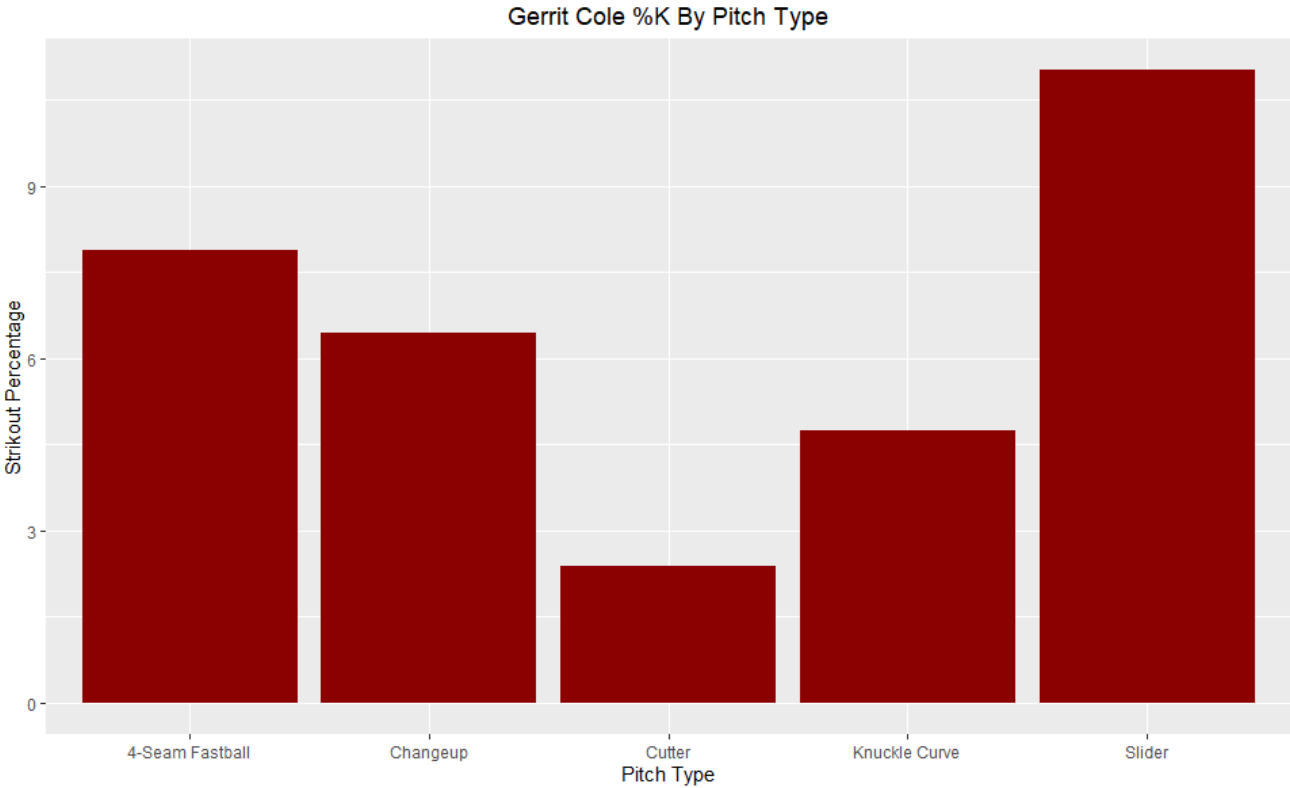
Strike Percentage by Inning
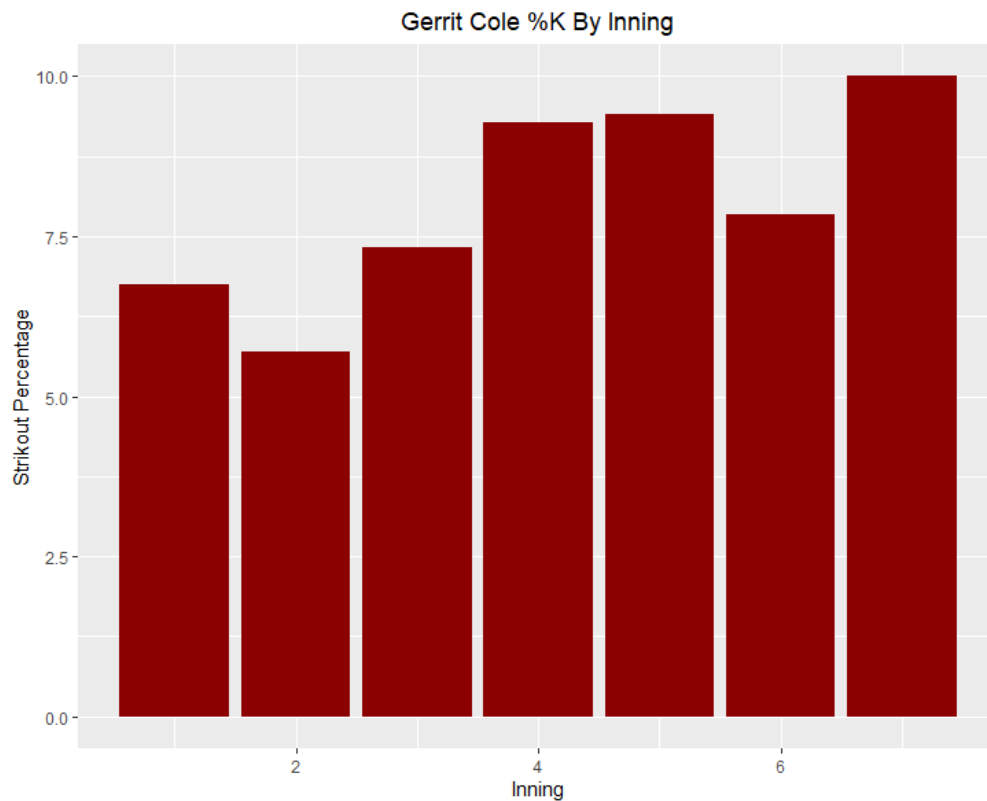
The strike percentage (%K) is the percentage of pitches resulting in strikes. As the game progresses, there is a notable increase in Cole's strike percentage from innings 1 and 2 (where it is around 5-7%) to the 7th inning (where it is around 10%).

This indicates that Cole gets more aggressive with his pitching as the game enters its later stages.

Strike Percentage by Pitch Type

Gerrit Cole %K By Pitch Type

As can be seen from the chart above, Cole's slider generates the highest strike percentage of any of his pitches. His four-seam fastball has the second highest strike percentage. His cutters, on the other hand, are less effective at generating strikes.

Composition of Cole's Pitches By Inning

Gerrit Cole Pitch Percent of Total By Inning
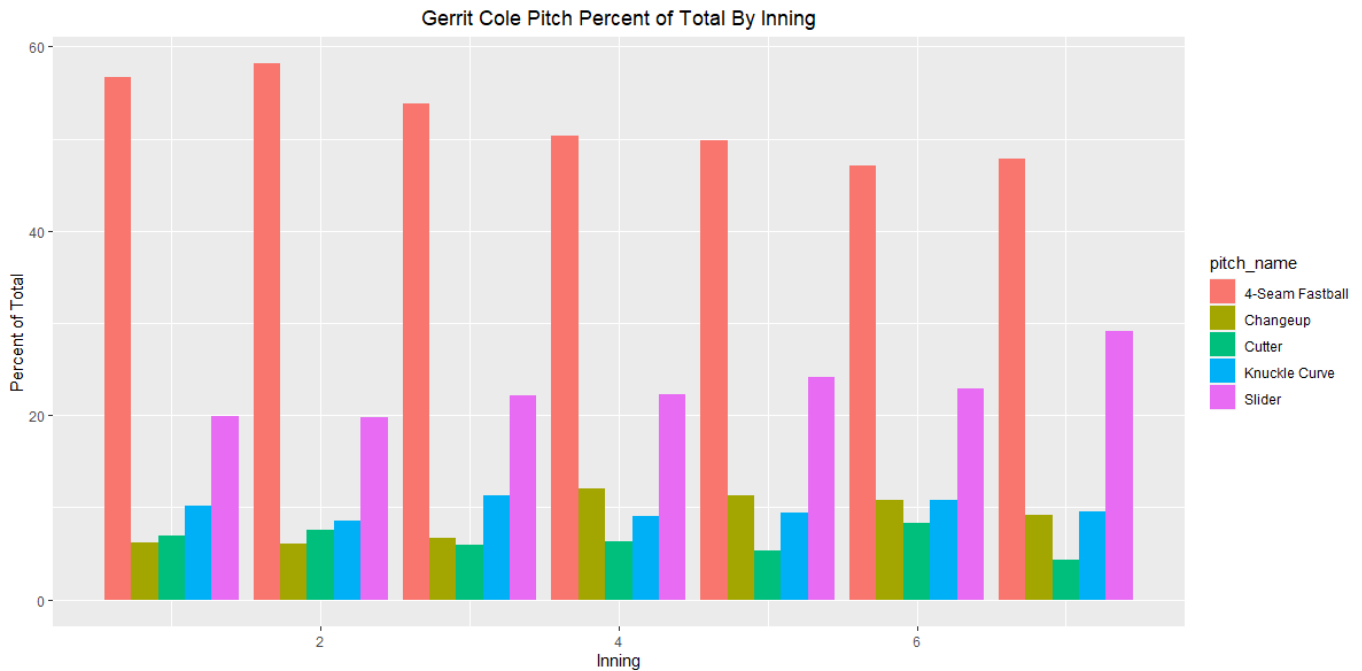
Gerrit Cole's most frequent type of pitch is the four-seam fastball, followed by the slider. He utilizes changeups, cutters, and knuckle curves less frequently. For most of the game, the composition of his pitches does not display much variance.

However, in the 7th inning, Cole tends to decrease his use of the fastball markedly and increases his slider use significantly. This is likely because he is getting tired as his pitch count increases and needs to incorporate more off-speed pitches. This makes sense, especially considering the chart from above showing that Cole's average fastball velocity decreases and the variance of his fastball increases as his pitch count grows.

Gerrit Cole Effectiveness Over Time

Gerrit Cole Daily FIP
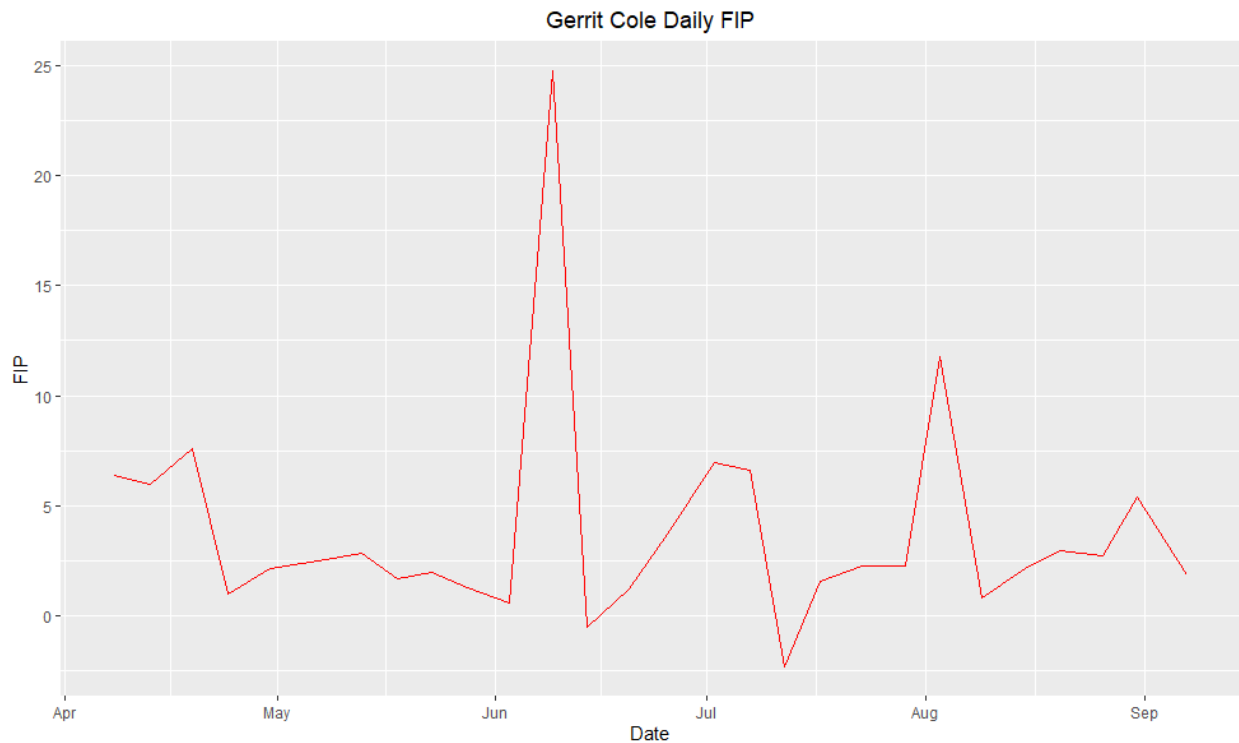
FIP mirrors earned run average (ERA), but only uses walks, strikeouts, and homeruns. It is a better measure of effectiveness than ERA because it does not take into account balls hit into play, which is influenced greatly by the defense. For example, if two pitchers have the same number of balls hit into play, but one has a better defense, then the pitcher with superior defense will have a much better ERA. This will erroneously make him appear to be a better pitcher.

The formula for FIP is:

$$FIP = \frac{13 \times HR + 3 \times (BB + HBP) - 2 \times K}{IP} + FIP \ constant$$

As you can see, FIP measures pitcher performance but removes the role of the defense. The FIP constant brings the metric onto the ERA scale.

As can be seen in the graph above, Cole's FIP displays considerable variance throughout the season. After a rough start, Cole's FIP generally decreased until mid-summer, after which it began an upward trend again. This inconsistency is a serious problem for Cole and the New York Yankees.