A Statistical Analysis of Trackman data from the 4/23/2022 Texas A&M Aggies vs. Arkansas Razorbacks
Baseball Game

John Hatlestad

1. **Abstract**

This statistical report serves as a demonstration of my proficiency in data engineering and statistical analysis. The primary objective was to derive valuable conclusions and insights from the provided dataset. To do this I utilized the Python programming language and the Pandas, Seaborn, and MatplotLib libraries. The entirety of the analysis and coding was conducted using Visual Studio Code. After analyzing the data and developing various visuals, the key summaries I got were overall statistics for the game, revealing that Arkansas hit better than Texas A&M. Also the data revealed play styles of specific players, for example Zebulon Vermillion is primarily a fastball pitcher. Furthermore, the investigation unearthed diverse insights from specific data points, such as emphasizing the significance of Exit Speed as a pivotal statistic in predicting a batter's proficiency and the likelihood of a hit yielding the desired outcome.

2. **Findings and Analysis**

In the following section I will provide each visual that I developed from the dataset and then a brief description of my findings from this visual.
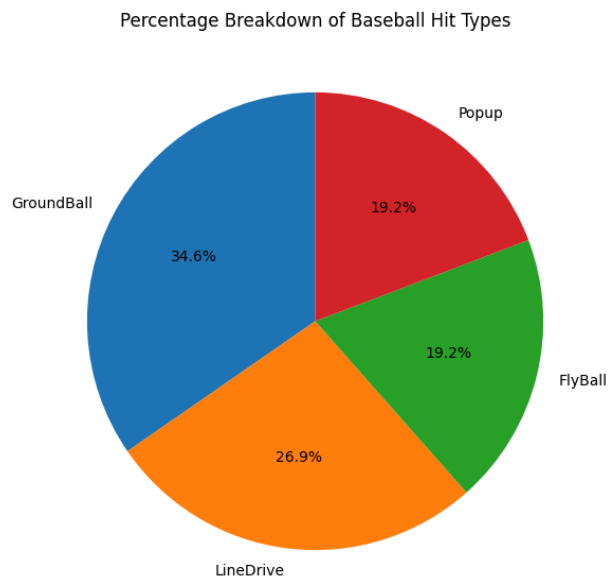

This pie chart illustrates the distribution of balls hit in fair territory, revealing that 34.6% were ground balls, 26.9% were line drives, and 19.2% for pop-ups and also 19.2% fly balls. This indicates that the majority of hits resulted in balls being hit into the air.

## Percentage Breakdown of Strikeouts by looking and swinging
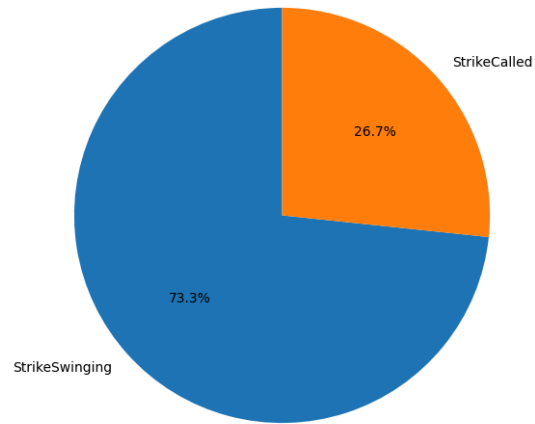
The pie chart provides a detailed breakdown of strikeouts based on the method, revealing that 73.7% were the result of swinging, whereas 26.3%, occurred from looking.
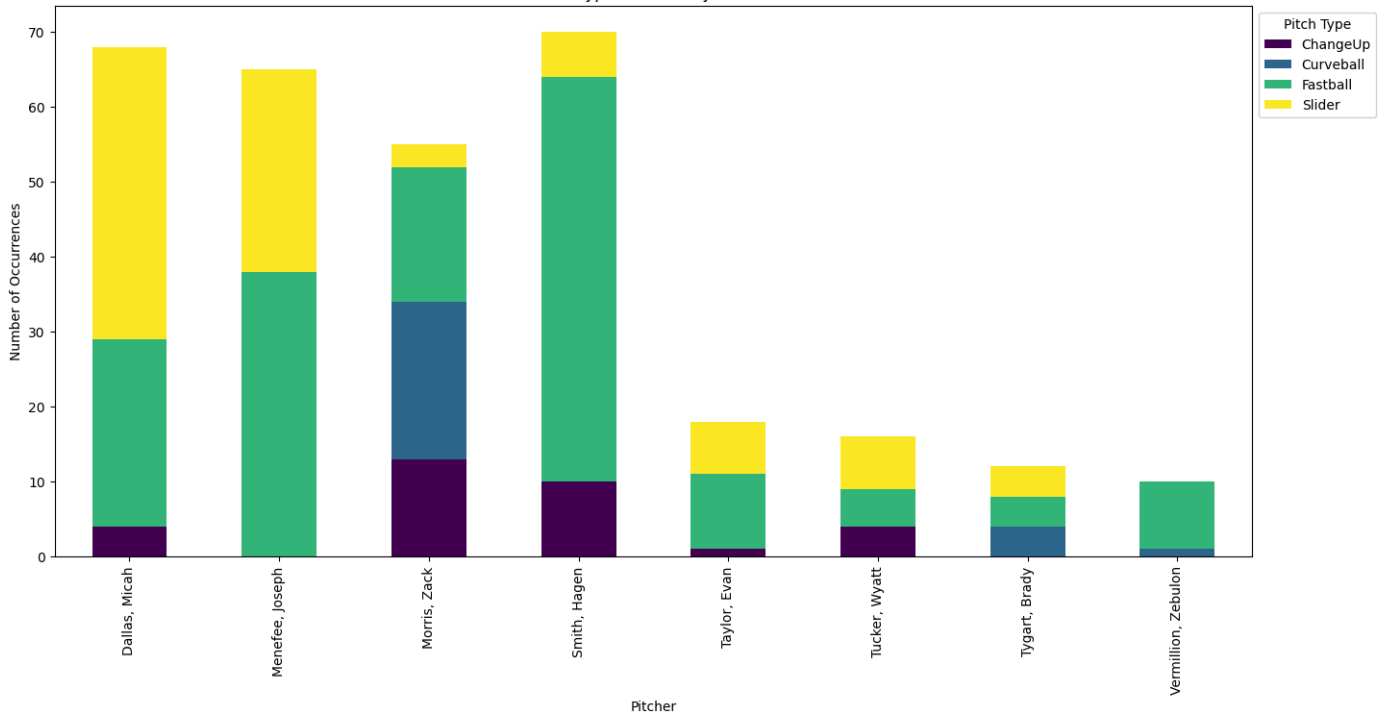
The bar graph above presents a comprehensive overview of every pitcher from both teams, showcasing the total number of pitches thrown and the variety of pitch types utilized. Noteworthy observations from the graph include Hagen Smith leading in total pitches with approximately 70, closely followed by Micah Dallas and Joseph Menefree. Hagen Smith and Zebulon Vermillion predominantly rely on fastballs, while Micah Dallas specializes in sliders, and Zack Morris is primarily an off-speed pitcher.
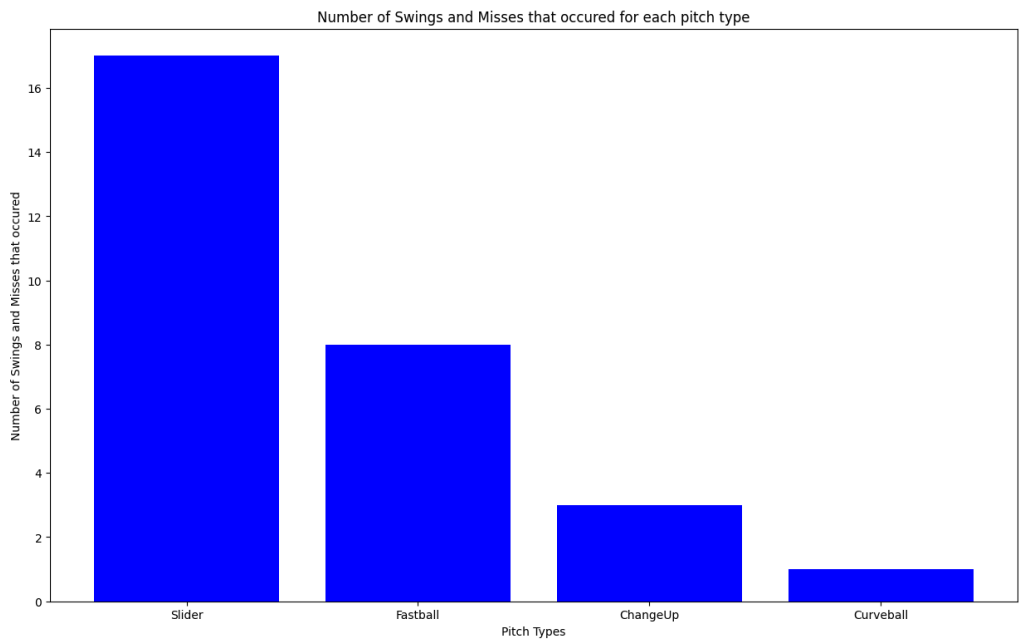

The presented bar graph illustrates the frequency of swings and misses for each pitch type throughout the entire game. Notably, sliders accounted for the majority of swings and misses, totaling 17, with fastballs being second. Interestingly, curveballs registered only one swing and miss throughout the entire game.
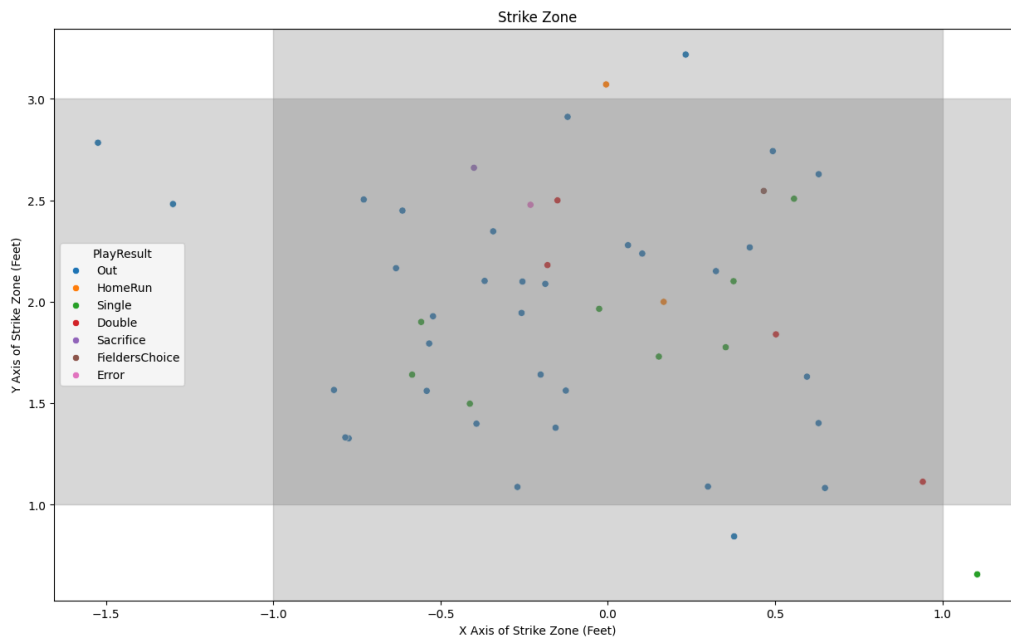
Strike Zone

This scatter plot visually depicts the spatial distribution of pitches thrown during the game, with the strike zone delineated by the gray shaded area. The color-coded dots represent balls in play, each corresponding to a specific outcome. Notably, a concentration of hits is observed in the central region of the strike zone. Conversely, balls in play outside or at the edges of the strike zone predominantly led to outs, with a noteworthy exception of one home run, double, and single. This underscores the trend that pitches thrown in the central area of the strike zone are more likely to yield hits, while those directed at the edges are inclined to result in outs.
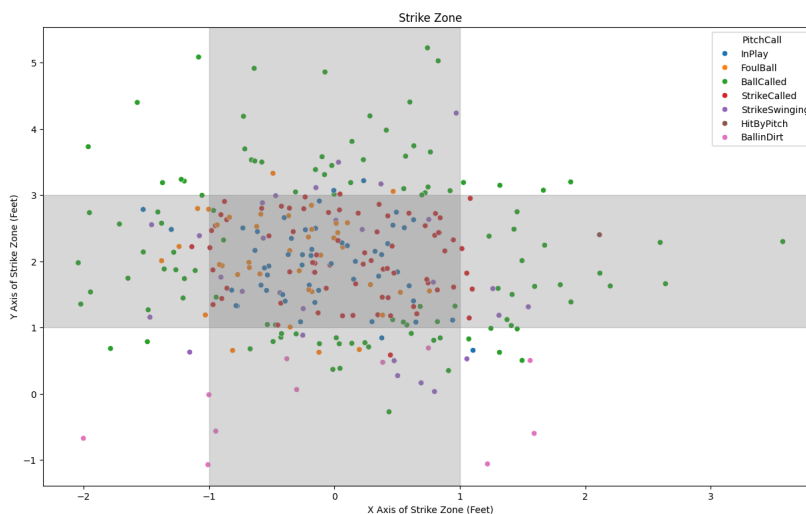
Strike Zone

This scatter plot provides a visual representation of pitch locations, with the dark shaded region denoting the strike zone. In this iteration, distinct colors signify the type of pitch called. The plot highlights the umpires' proficiency in this game, particularly in their propensity for calling strikes on the left side of the plate, indicating a consistent and accurate strike zone interpretation.
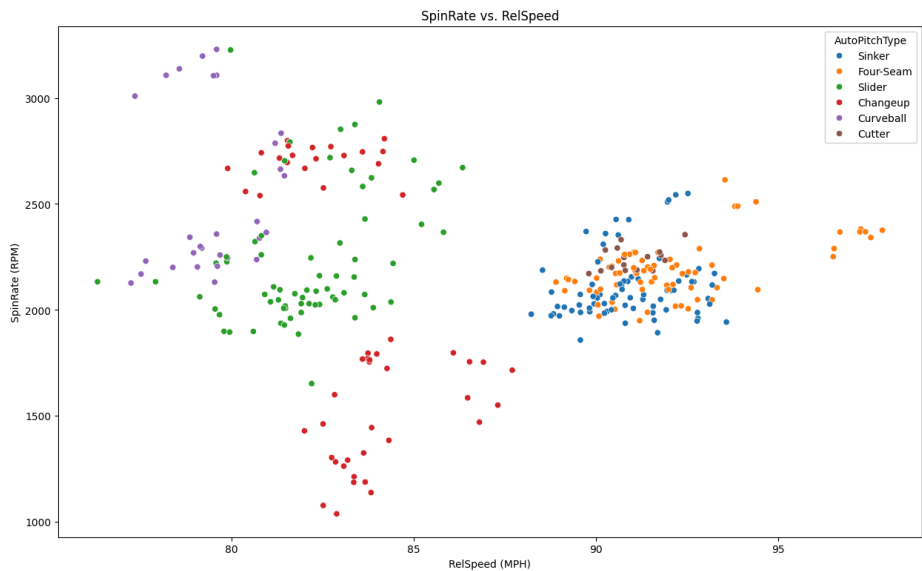
This scatter plot showcases pitches categorized by their velocity (RelSpeed) and Spin Rate, with each pitch type differentiated by color. The intriguing aspect lies in the distinct groupings of pitch types, each characterized by unique speed and spin rate patterns. Fastballs exhibit higher speeds with an average spin rate, Changeups feature lower speeds coupled with low spin rates. Curveballs, while the slowest in speed, boast the highest Spin Rates, and Sliders showcase lower speeds with higher spin rates.

.
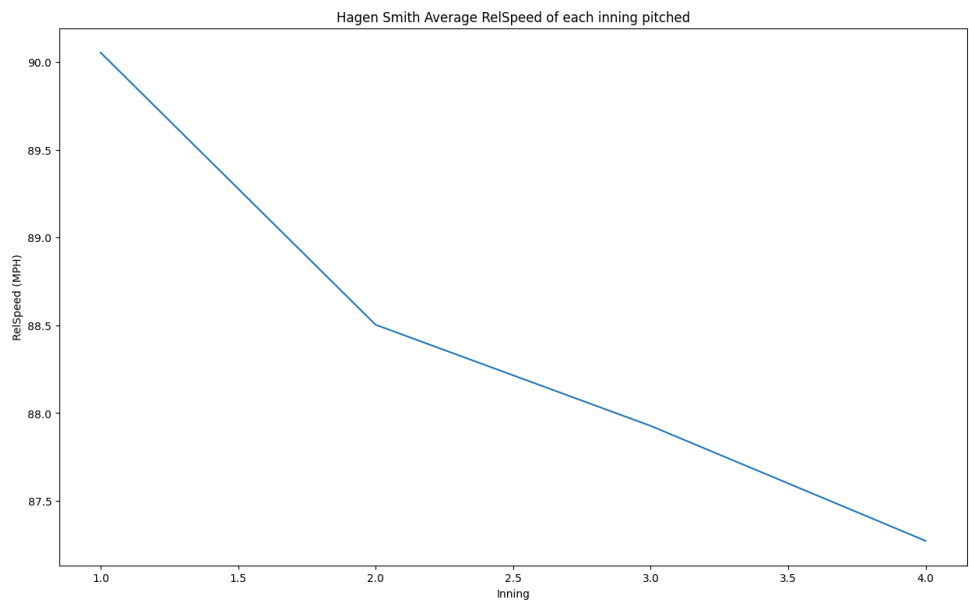
The depicted line graph illustrates Hagen Smith's average velocity (RelSpeed) for each of the four innings he pitched. It is noticeable that he initiated the game with an approximate RelSpeed of 90 MPH, but subsequent innings witnessed a consistent decline, notably with the most significant drop occurring between the first and second innings. While various factors, including pitch types, could contribute to this trend, it suggests a plausible interpretation that Hagen Smith experiences a decrease in RelSpeed as the game progresses, potentially indicating fatigue over the course of his pitching duration.
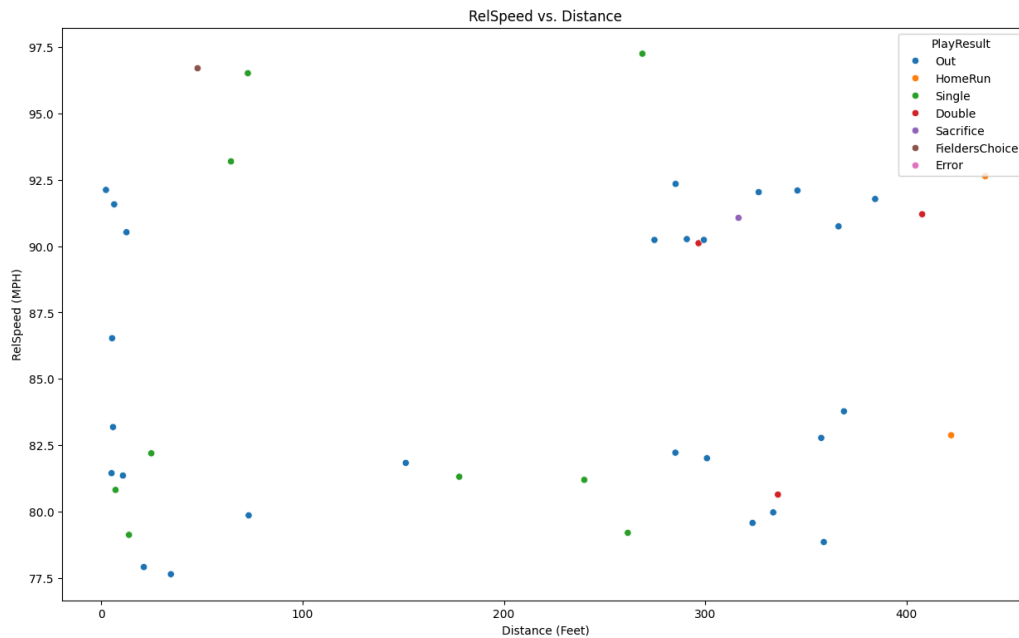
This scatter plot presents all balls in play, mapping their Distance against the velocity (RelSpeed) of the corresponding pitch. I created this visualization in response to the old baseball notion that asserts a positive correlation between pitch speed and hit distance, "the harder they throw, the farther they go". However, upon examination of this scatter plot, it becomes evident that such a correlation is unfounded. The data reveals no discernible relationship between pitch speed and the distance the ball travels.

Hagen Smith Release Point for each pitch type

This scatter plot charts Hagen Smith's release points for each pitch he throws, along with the corresponding pitch types. As a hitter, discerning the pitch type based solely on his release point appears challenging, given the wide distribution observed. However, intriguingly, there is a discernible grouping for sliders and changeups. Sliders exhibit a higher release point with a less extended reach, while changeups consistently feature a notably low release point. This suggests that, for sliders and changeups, the release point could potentially serve as a visual cue for anticipating the type of pitch Hagen Smith is likely to throw.

The presented bar graph provides a straightforward representation of the walk and strikeout statistics for each pitcher. Joseph Menefee led in strikeouts, tallying seven, while also issuing four walks. Hagen Smith recorded the highest number of walks, totaling five, suggesting a higher level of variability in his pitching accuracy. Notably, both Evan Taylor and Brady Tygart did not issue any walks during the game.

Spin Rate Distribution by Pitcher
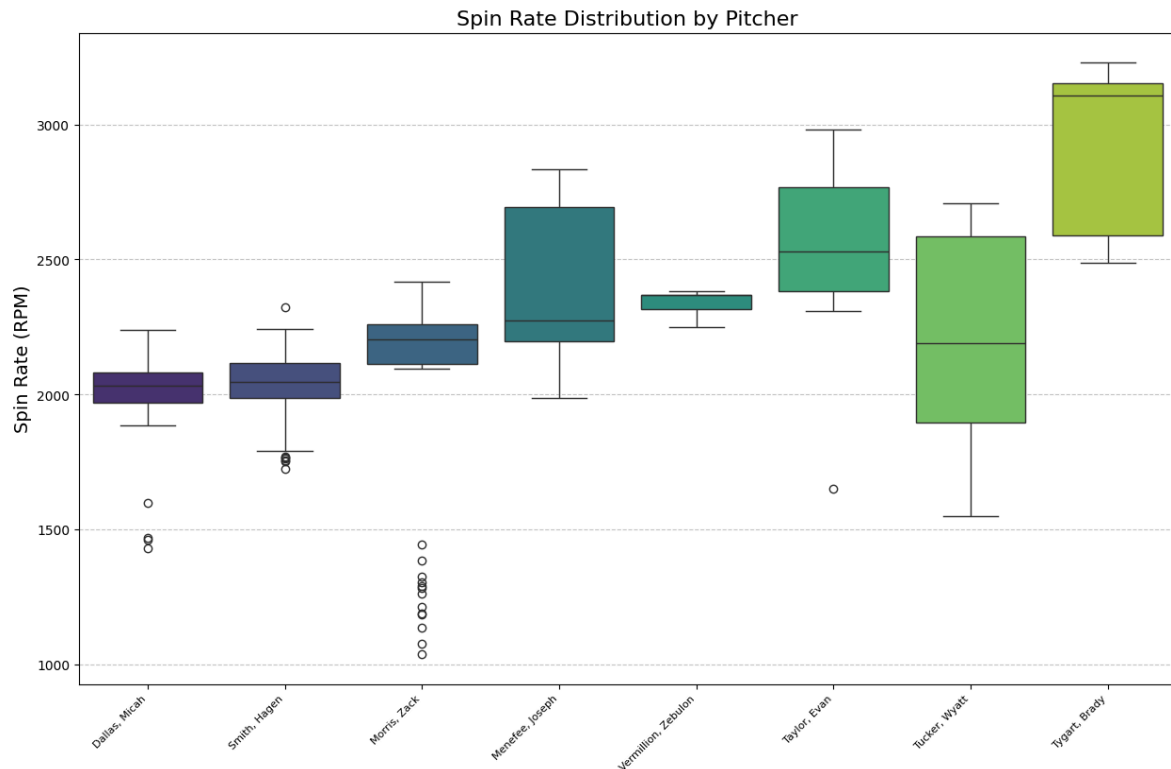
The depicted box plot showcases the distribution of Spin Rates by pitcher. Notably, Brady Tygart stands out with the highest recorded rpm, exceeding 3000, likely indicating the highest average spin rate among the pitchers. Conversely, Zebulon Vermillion's box plot reveals a narrow distribution of spin rates, suggesting a tendency to consistently throw one specific type of pitch. Zack Morris, on the other hand, exhibits the lowest spin rates, with one pitch hovering close to 1000 rpms.

Pitch Speed Distribution by Pitcher
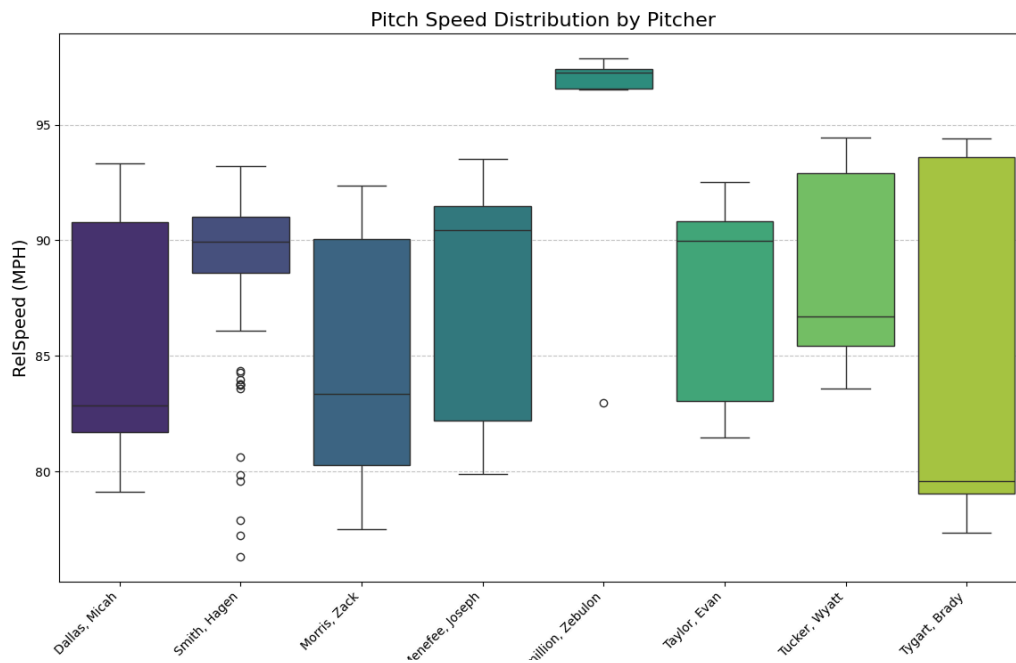
This box plot, focused on pitch speed distribution among pitchers, offers further insights. Zebulon Vermillion stands out with a narrow and high distribution, indicative of a predominantly fastball repertoire, aligning with his status as the fastest pitcher. This correlation aligns with observations from the previous spin rate box plot and the Pitch Type by Pitchers bar graph. Hagen Smith, while showcasing a relatively small distribution, exhibits outliers and threw the slowest pitch of the game, suggesting a mix of pitch types with a focus on fastballs. On the other hand, Brady Tygart's notably wide distribution implies a diverse array of pitch types, underscoring his versatility on the mound.

Number of hits each team had
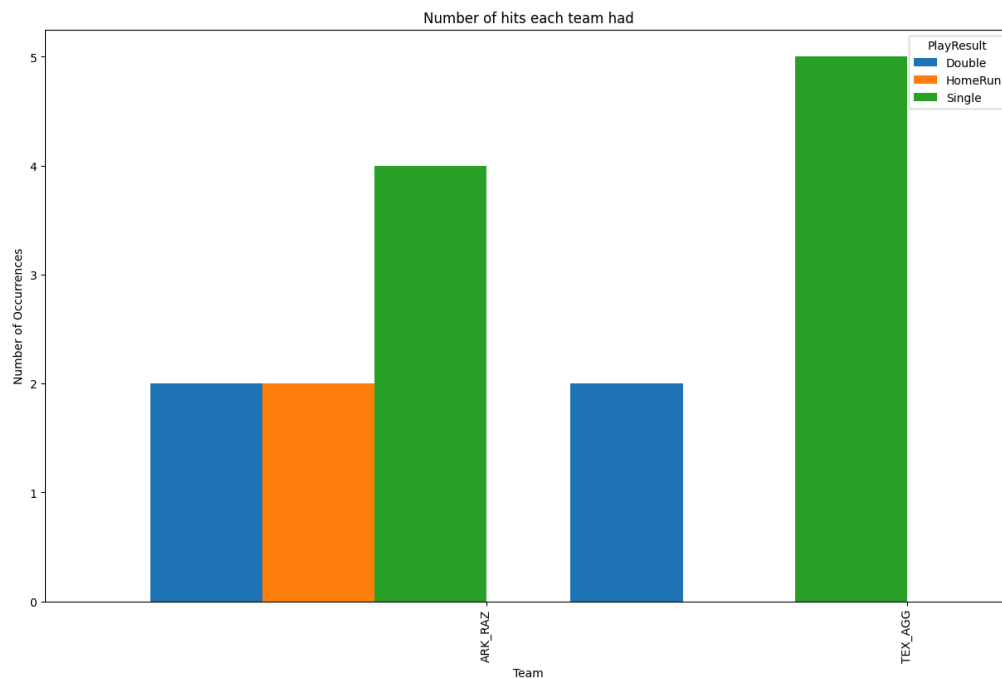
Above is a simple bar graph displaying how many of each type of hits both teams had. Arkansas had two home runs, two doubles, and four singles, totaling eight hits on the game. Texas A&M had two doubles and five singles, totaling seven hits on the game. Based on just this data I would say that Arkansas was a better hitting team and won the game.

Hagen Smith Release Angle for balls & strikes
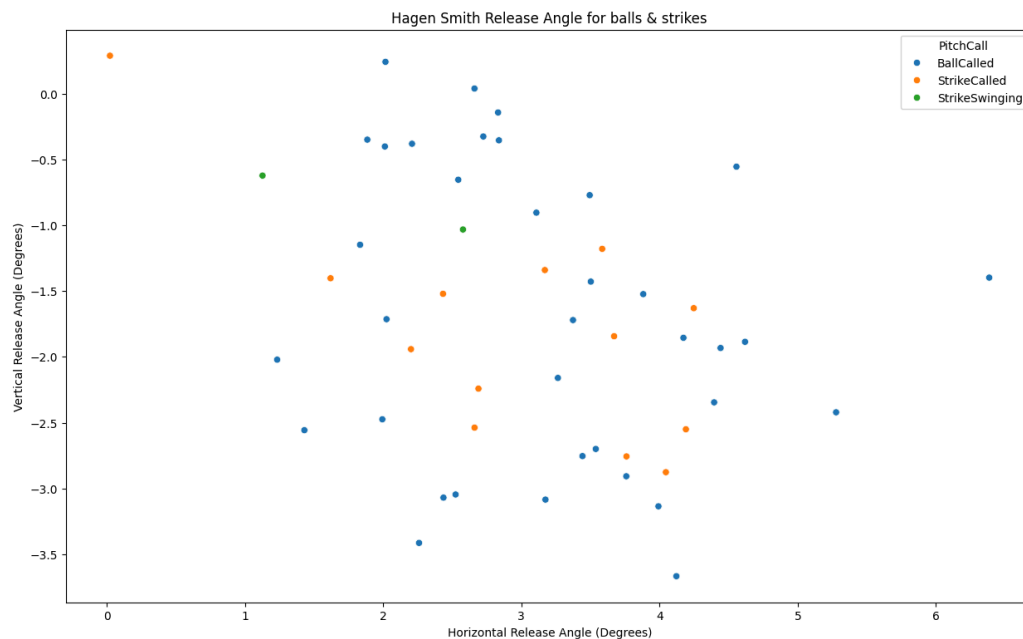
Above is a scatter plot that shows all of Hagen Smith's pitches plotted based on his vertical release angle and his horizontal release angle. They are color coded based on if the pitch was a ball or a strike. I wanted to see if his release angle was abnormal then it would result in balls and this scatter plot kinda proves that. A lot of the pitches around the edges of the scatterplot resulted in balls, while balls in the middle, his normal release angle, resulted in strikes. Which I thought was pretty interesting.
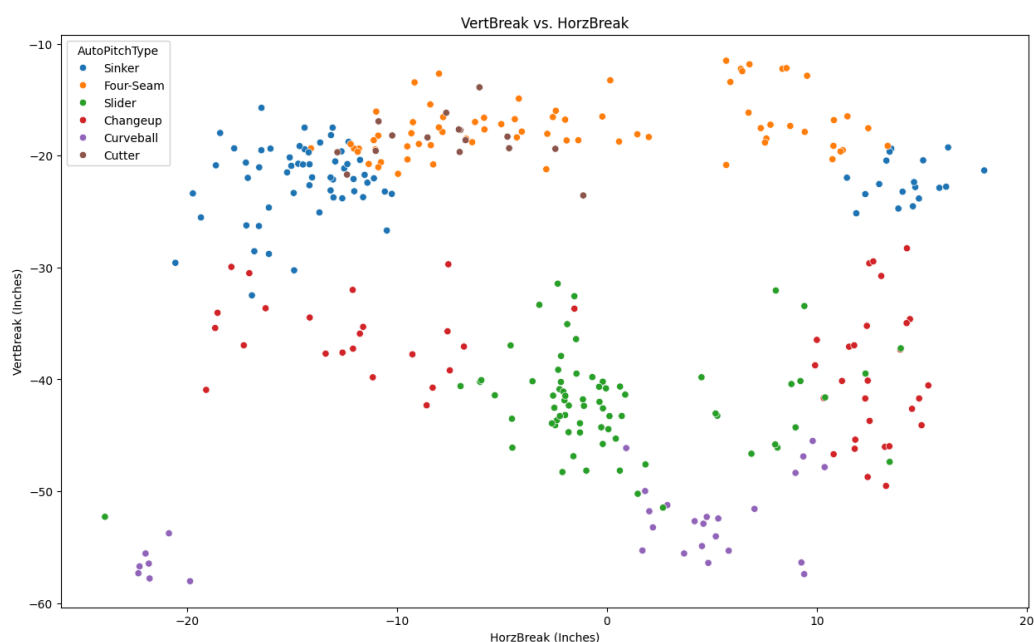
This scatter plot closely mirrors the Spin Rate vs. RelSpeed scatter plot presented earlier. It depicts pitches according to their Vertical Break (VertBreak) and Horizontal Break (HorzBreak), with color-coded distinctions for each pitch type. Notably, the various pitch types once again form distinct clusters, facilitating the identification of the pitch type based on the Horz and VertBreak. Observations include that Four-Seamers and Sinkers exhibit higher VertBreak, while Sinkers notably surpass fastballs in HorzBreak. Curveballs, in contrast, display both low VertBreak and low HorzBreak.

MaxHeight vs. Hangtime
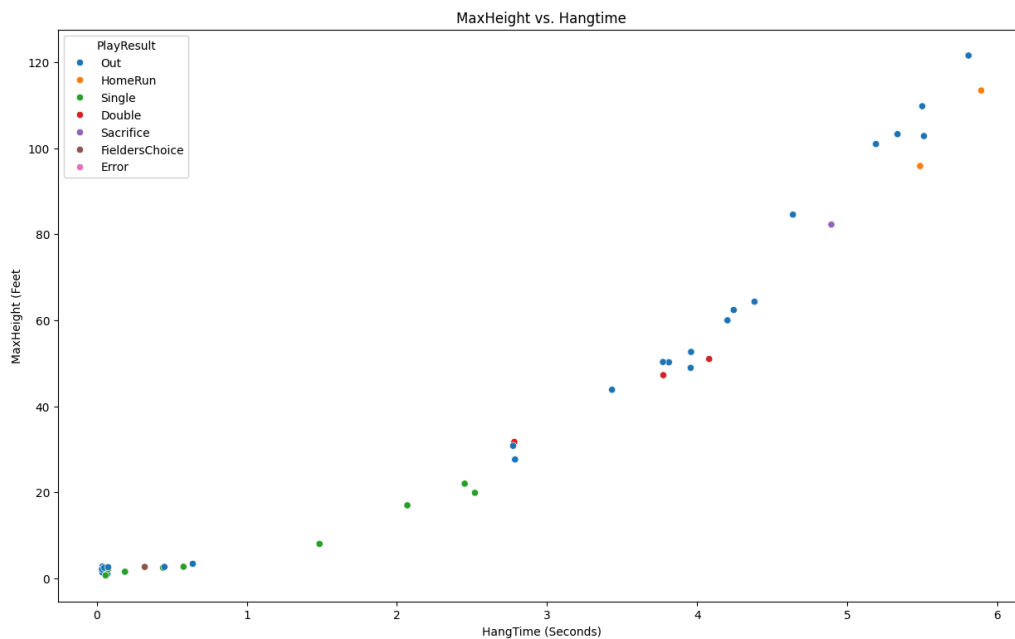
This a scatter plot plotting different play results based on HangTime and MaxHeight of the hits. As you can see there is a positive correlation between HangTime and MaxHeight. But you can also see that you don't really want to have a lot of HangTime on your hits, besides the two home runs and a double. All of the hits that had above four seconds of hang time resulted in outs.

Percentage of Fastballs by Strike Count
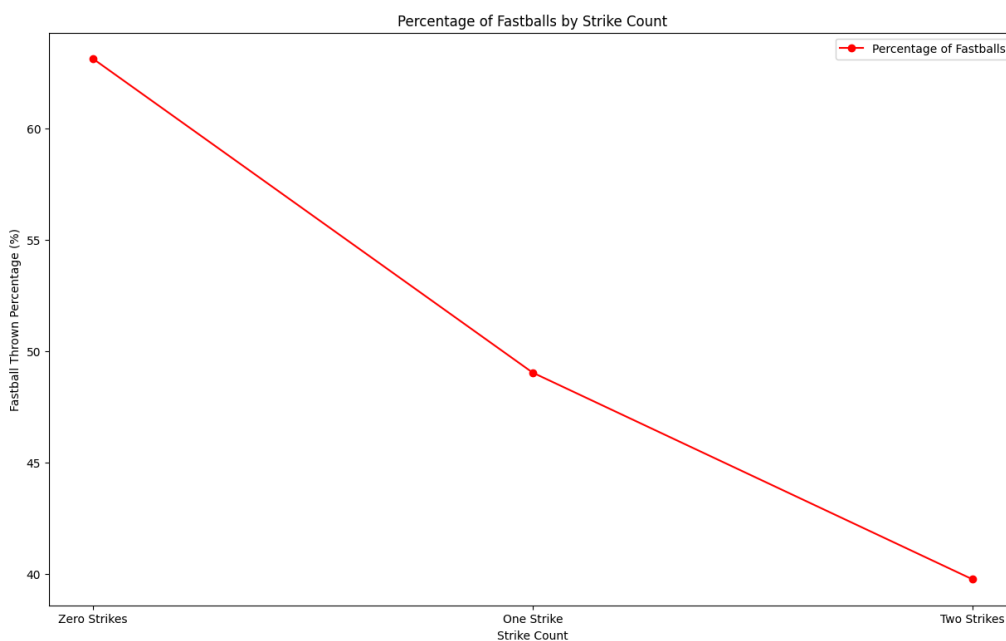
Above is a line graph that depicts that percentage of fastballs based on strike count. You can see that when there are zero strikes a fastball is thrown almost 65% of the time. But then when there is one strike it goes down to around 50% and when there are two strikes it goes down to under 40% of the time. This shows that as strike count increases you would expect a fastball less and less. When there are zero strikes you could expect a fastball, but when there are two strikes you should expect an offspeed pitch to come.

Average Exit Speed for balls in play

A bar chart titled "Average Exit Speed for balls in play" with the x-axis labeled "PlayResult" and the y-axis labeled "Average Exit Speed (MPH)". The categories are: Double (~98), Error (~37), FieldersChoice (~103), HomeRun (~104), Out (~82), Sacrifice (~86), Single (~92).

Average ExitSpeed for when run is and isnt scored

A bar chart titled "Average ExitSpeed for when run is and isnt scored" with the x-axis labeled "RunsScore" and the y-axis labeled "Average ExitSpeed (MPH)". The categories are: No run scored (~81, blue), Run is scored (~97.5, green).

Combining the two preceding bar graphs, both centered around Exit Speed and play outcomes, provides a comprehensive view. The initial graph illustrates the average Exit Speed for various play results, revealing that home runs exhibit the highest average Exit Speed, followed unexpectedly by fielder's choices, doubles, singles, and outs, with the latter two having the lowest Exit Speeds, akin to errors which are generally categorized as unfavorable hits. This underscores the correlation between higher Exit Speeds and more favorable play outcomes. The second graph reinforces this correlation by indicating that hits resulting in a run scored have an average Exit Speed over 10 mph greater than hits that do not culminate in a run.
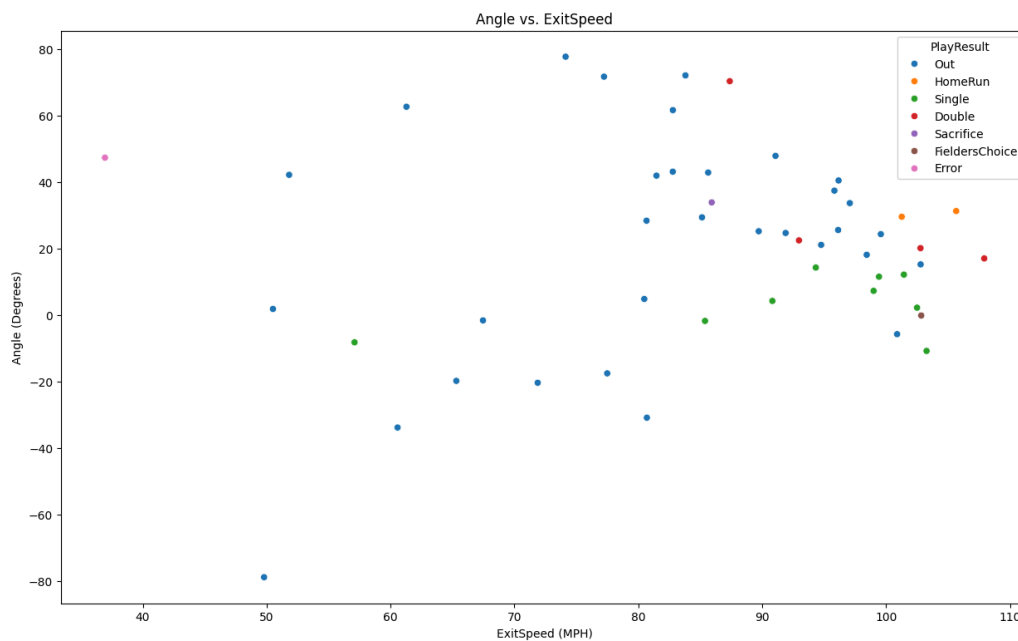
This is a scatter plot that plots different balls in play based on their ExitSpeed and Angle. This shows that you want to have a high angle and a ExitSpeed for favorable results as a better.

Number of Balls called against for each batter on Texas A&M
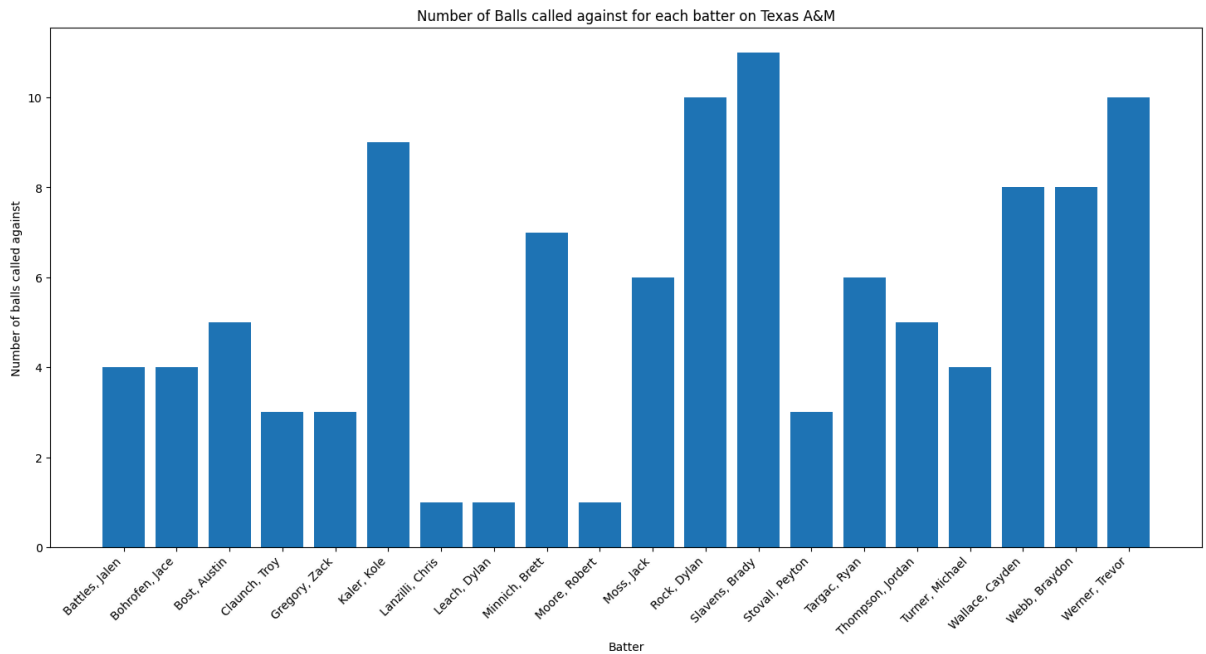
In my final visual representation, I've created a bar graph depicting the number of balls called against each batter on the Texas A&M team. It's important to consider various factors, including the number of plate appearances and the accuracy of the pitchers. Notably, Brady Slavens received the highest count of balls, totaling 11, suggesting a patient approach at the plate and potentially indicating him as the most selective hitter on the Texas A&M roster.

Below are various other statistics I computed during my analysis.

The farthest hit was 439.06345 feet

The Pitch with the least amount of Spin Rate was 1038.174854 RPMs

Pitcher: Dallas, Micah
  Slowest Pitch: 79.12098
  Fastest Pitch: 93.32961
  Average Pitch Speed: 85.53325544117648

Pitcher: Menefee, Joseph
  Slowest Pitch: 79.88872
  Fastest Pitch: 93.51554
  Average Pitch Speed: 87.47055738461539

Pitcher: Morris, Zack

Slowest Pitch: 77.50949
Fastest Pitch: 92.37545
Average Pitch Speed: 84.48598563636365

Pitcher: Smith, Hagen
Slowest Pitch: 76.31277
Fastest Pitch: 93.19926
Average Pitch Speed: 88.67115942857143

Pitcher: Taylor, Evan
Slowest Pitch: 81.45349
Fastest Pitch: 92.52993
Average Pitch Speed: 87.30804666666667

Pitcher: Tucker, Wyatt
Slowest Pitch: 83.59701
Fastest Pitch: 94.45079
Average Pitch Speed: 88.08637125

Pitcher: Tygart, Brady
Slowest Pitch: 77.34217
Fastest Pitch: 94.40079
Average Pitch Speed: 83.9673875

Pitcher: Vermillion, Zebulon
Slowest Pitch: 82.97114
Fastest Pitch: 97.87086
Average Pitch Speed: 95.56755333333334

Average Spin Rate of Breaking Balls by Play Result:
PlayResult
Double   2322.516663
Error   2598.478222
HomeRun   2159.833670
Out   2408.279423
Single   2622.642338
Undefined   2386.056297
Name: SpinRate, dtype: float64

Average RelSpeed of Fastballs by Play Result:
PlayResult
Double   90.030777
FieldersChoice 96.707050
HomeRun   92.638720

Out   91.173233
Sacrifice   91.068860
Single   95.656863
Undefined   91.388441
Name: RelSpeed, dtype: float64

What pitching metrics correspond to having high spin rates?

	SpinRate	VertRelAngle	HorzRelAngle	RelHeight	RelSide	Extension
SpinRate	1.000000	0.145304	-0.152008	0.067806	0.239917	-0.161633
VertRelAngle	0.145304	1.000000	0.076100	-0.243341	-0.124087	-0.300949
HorzRelAngle	-0.152008	0.076100	1.000000	-0.745207	-0.828238	0.473115
RelHeight	0.067806	-0.243341	-0.745207	1.000000	0.650751	-0.253169
RelSide	0.239917	-0.124087	-0.828238	0.650751	1.000000	-0.295526
Extension	-0.161633	-0.300949	0.473115	-0.253169	-0.295526	1.000000

Average speed drop by pitch type
TaggedPitchType
ChangeUp   6.532541
Curveball   6.108515
Fastball   7.596940
Slider   6.528842
Name: SpeedDrop, dtype: float64

Average Exit Speed for each team
BatterTeam
ARK_RAZ   83.476467
TEX_AGG   80.862289
Name: ExitSpeed, dtype: float64

Average Exit Speed, Angle, and Distance for Arkansas Batters

	Batter	Angle	ExitSpeed	Distance
0	Battles, Jalen	23.709201	81.902488	130.383143
1	Bohrofen, Jace	NaN	NaN	NaN
2	Gregory, Zack	17.159128	103.178470	260.124870
3	Lanzilli, Chris	10.420261	93.202830	206.727740
4	Leach, Dylan	NaN	NaN	NaN
5	Moore, Robert	7.876429	95.057695	228.007020
6	Slavens, Brady	23.733179	76.728468	290.088920
7	Stovall, Peyton	25.023952	97.847630	355.996590
8	Turner, Michael	48.744746	87.840680	64.457640
9	Wallace, Cayden	-9.586814	64.026020	2.364050
10	Webb, Braydon	34.173443	74.750390	210.736330

Average Spin Rate for Strikes by swinging: 2275.8813425172416
Average Spin Rate for balls in play: 2213.4547444615387