Blue Jays Baseball Research

Jan 2021

Question 1: Pitch Command

The attached dataset fastballcommand.xlsx contains metadata, pitch attributes, and pitch locations for every fastball thrown by five different ML starters between 2018 and 2020. Please answer the following questions:

1. Which pitcher had the best fastball command? How did you determine this?
2. If you wanted to create a fastball command metric that could be applied to any pitcher at any level, how might you go about doing so?

Problem Framing

Fastball command is the pitcher's ability of throwing a fastball at the precise location he intended. Since the true intended location of each pitch is unknown, in this analysis, I'll just assume that the pitcher would like to throw the ball at the edges of the strike zone. This approach of judging the closeness of the ball to the strike zone is similar to the existing Edge Percentage statistic (BaseballCloudBlog 2020), but not quite the same as I'll explain.

Method of Analysis

I loaded, transformed, and analyzed the data in Python (Google Colab), leveraging libraries such as Pandas, Numpy, SciPy, Matplotlib, and Seaborn for descriptive analysis, modeling, and visualization.

Question 1.1: Which pitcher had the best fastball command?

Solution

I attempted to answer this question first through exploratory data analysis (EDA). Because I assumed that the pitchers would like to throw all the balls at the strike zone's edges, in the ideal case, the pitcher with the best fastball command should have all the balls right on the strike zone's edges if we plot the pitch location (x-y coordinates) in a scatter plot.

To compare the pitch location to the strike zone, we have to define a strike zone first. In the data provided, there was no available information about the batters' height and stance, meaning we don't know the true height of the strike zone. So, I decided to use a box-shaped universal strike zone following the specifications depicted by Boyle (2018) to represent it (Fig. 1).

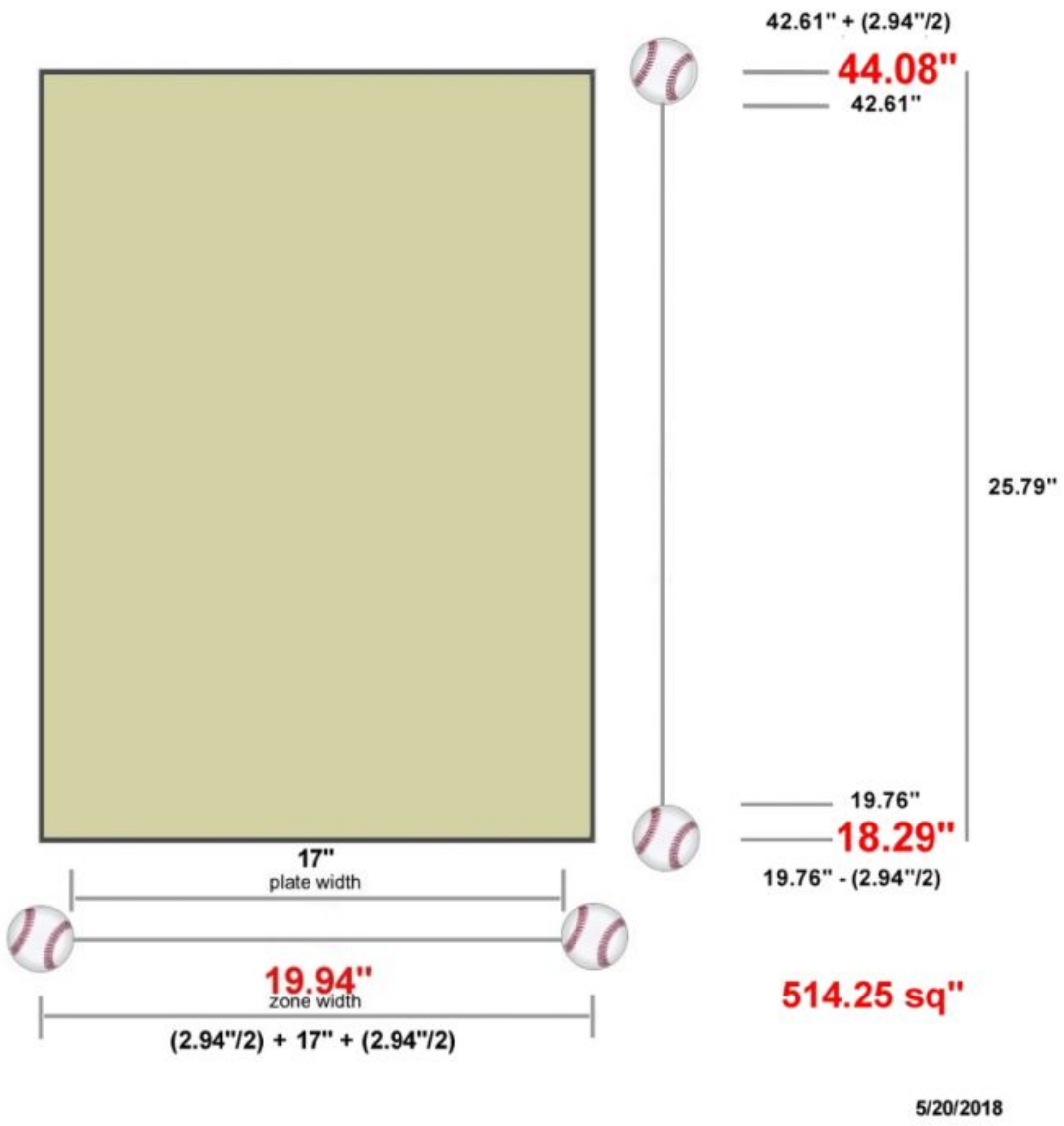
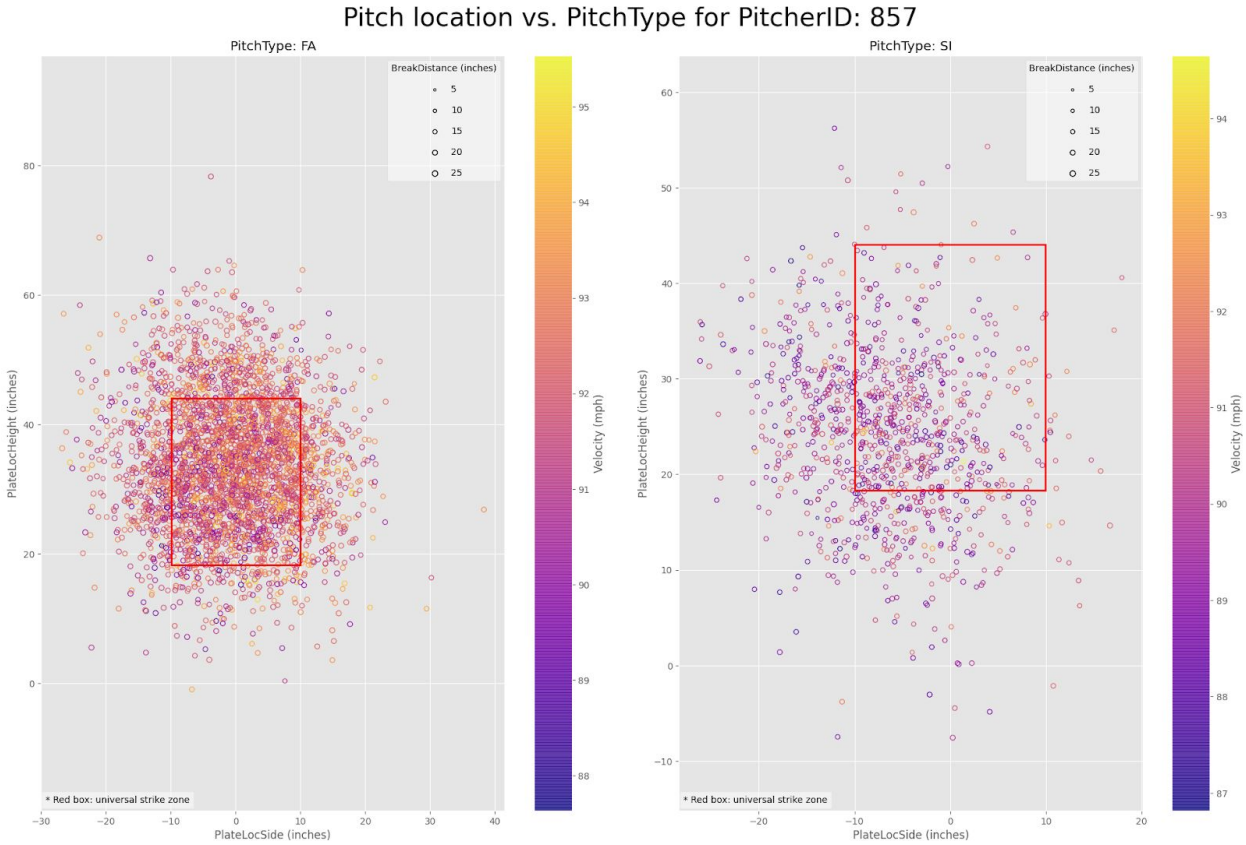


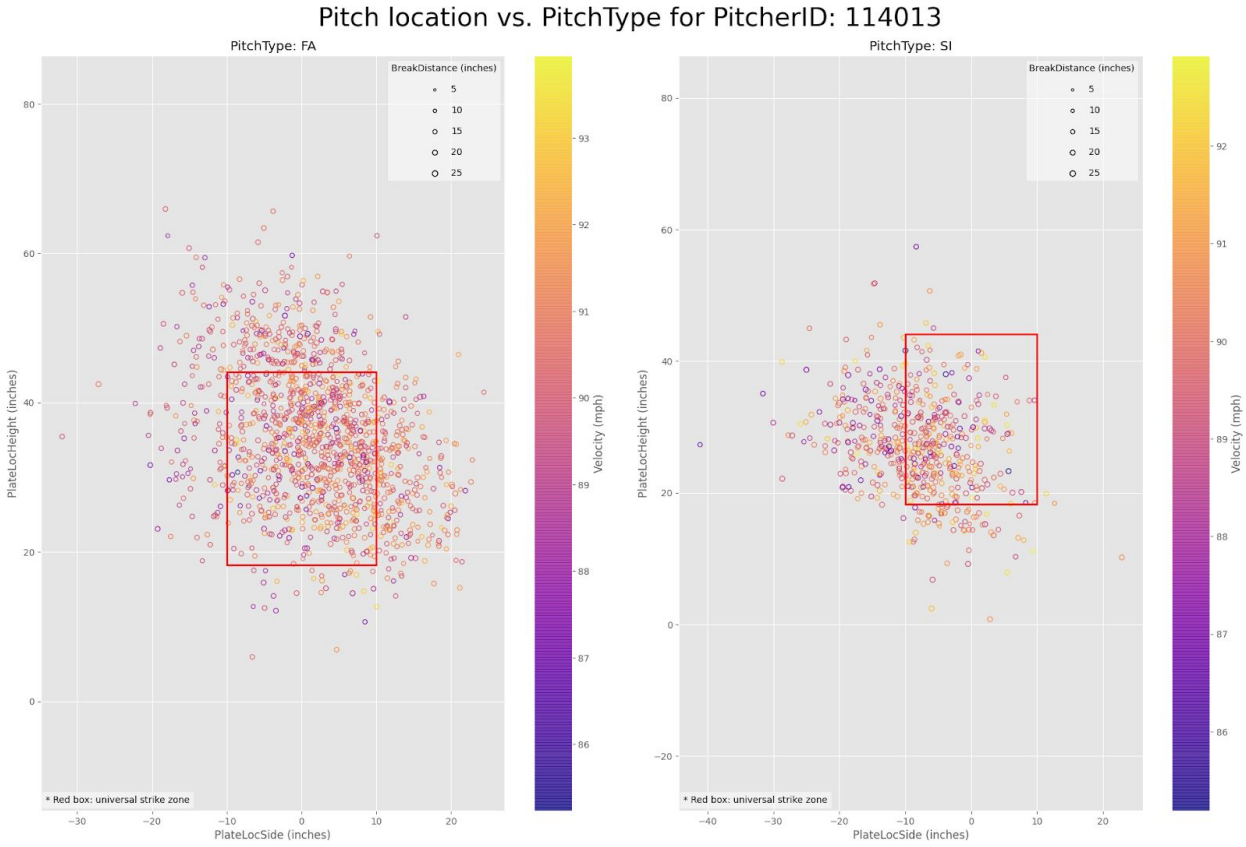
Figure 1. Specifications of the universal strike zone used in the analysis. (Image credit: Boyle 2018)

I used the PlateLocHeight and PlateLocSide (both converted to inches) to visualize each pitcher's pitch location in relation to the strike zone (Fig. 2).

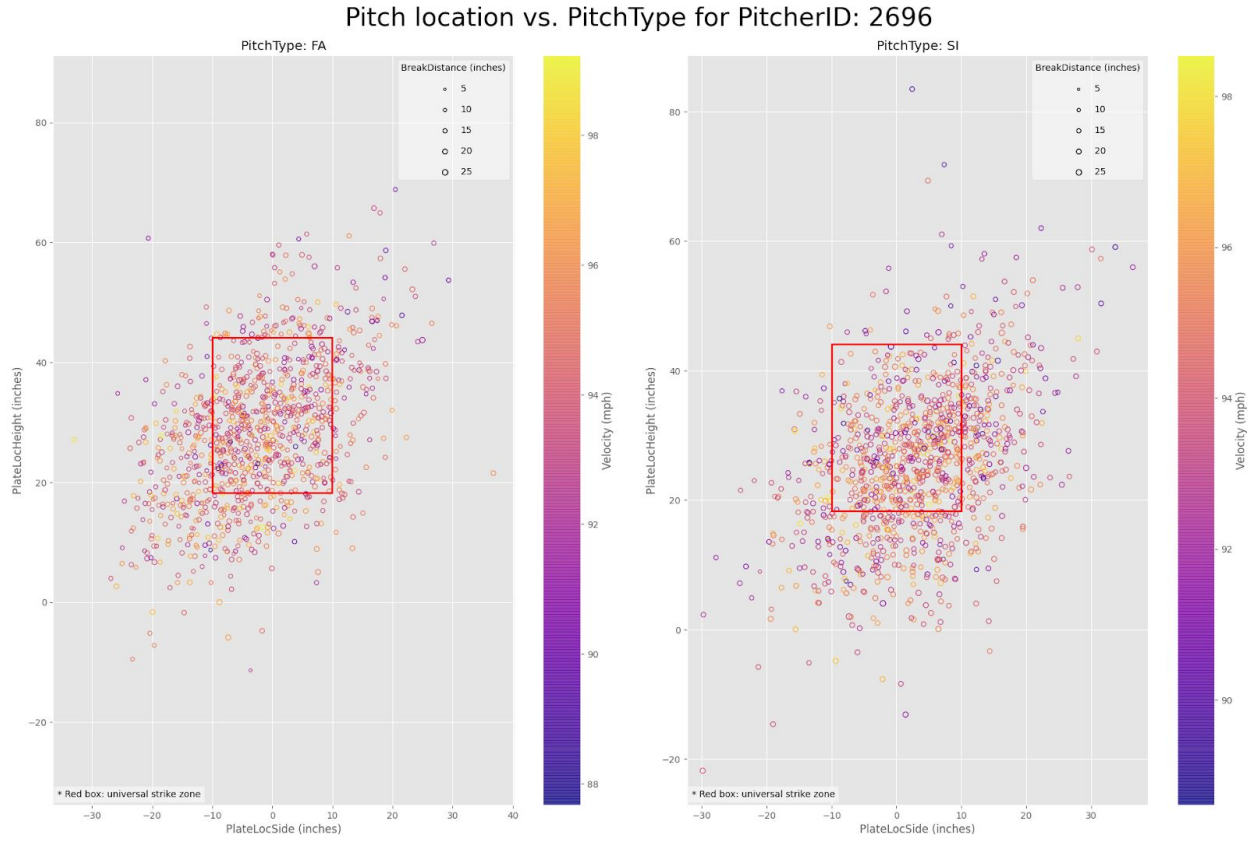
(a)


Pitch location vs. PitchType for PitcherID: 857
PitchType: FA
PitchType: SI
BreakDistance (inches)
5
10
15
20
25
PlateLocHeight (inches)
PlateLocSide (inches)
Velocity (mph)
* Red box: universal strike zone

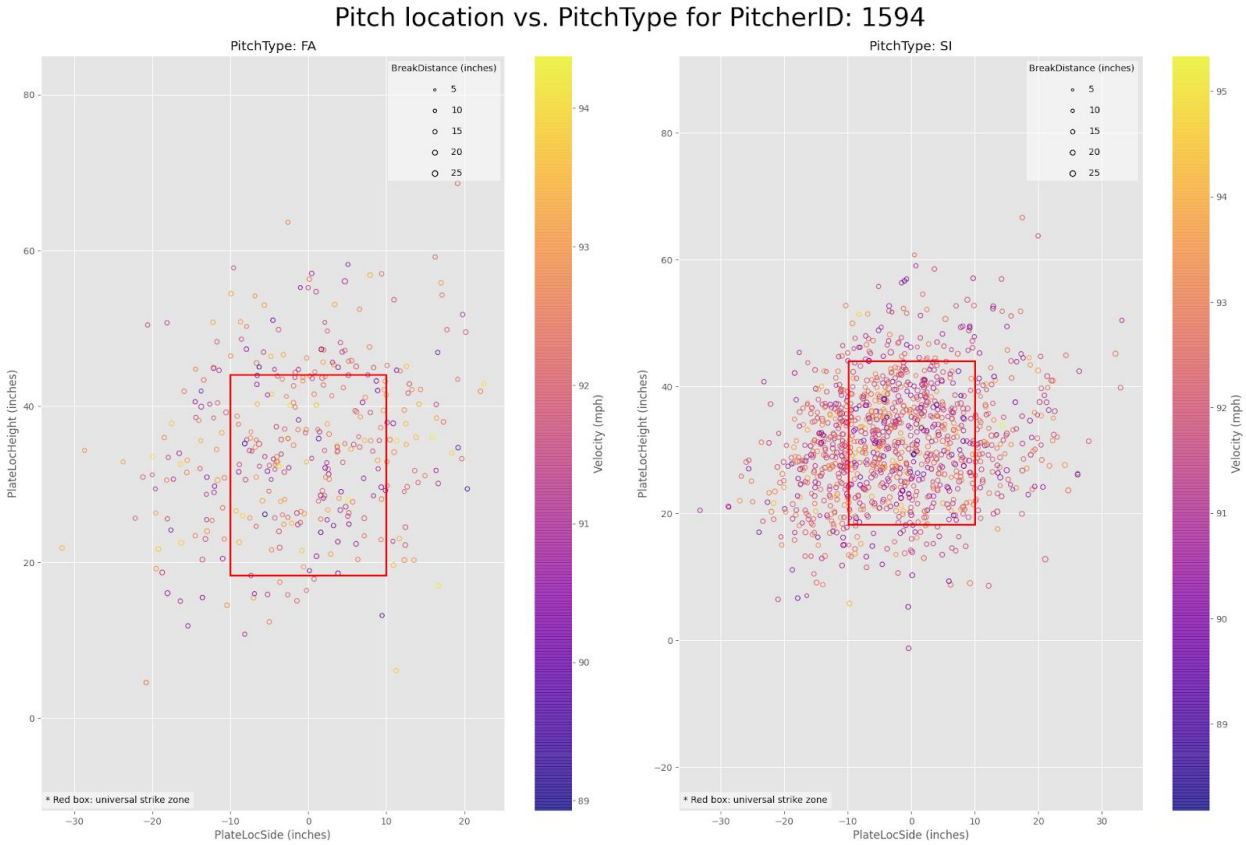
(b)


Pitch location vs. PitchType for PitcherID: 114013
PitchType: FA
PitchType: SI
BreakDistance (inches)
5
10
15
20
25
PlateLocHeight (inches)
PlateLocSide (inches)
Velocity (mph)
* Red box: universal strike zone

(c)


Pitch location vs. PitchType for PitcherID: 2696
PitchType: FA
PitchType: SI
BreakDistance (inches)
5
10
15
20
25
PlateLocHeight (inches)
PlateLocSide (inches)
Velocity (mph)
* Red box: universal strike zone

(d)


Pitch location vs. PitchType for PitcherID: 1594
PitchType: FA
PitchType: SI
BreakDistance (inches)
5
10
15
20
25
PlateLocHeight (inches)
PlateLocSide (inches)
Velocity (mph)
* Red box: universal strike zone

(e)

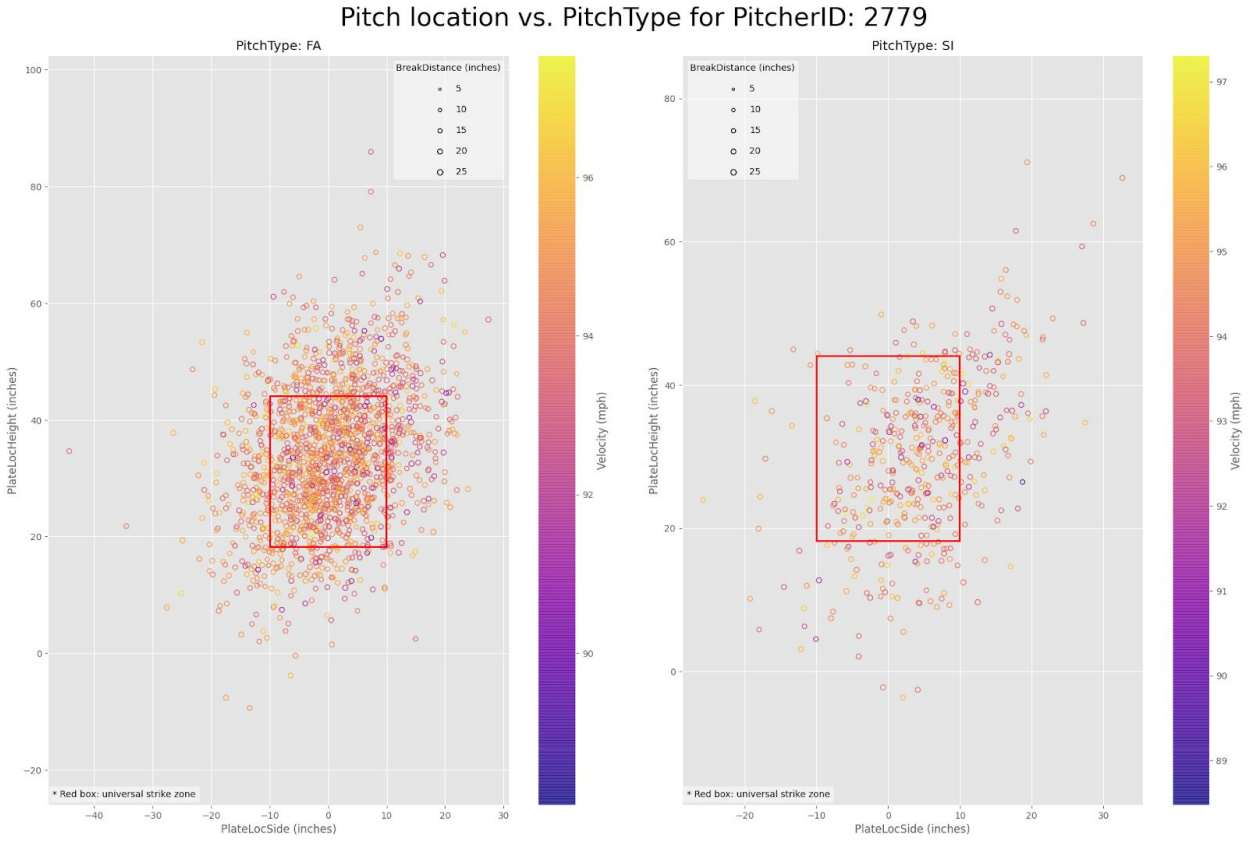


Figure 2. The five different pitchers' pitch locations shown as scatter plots in relation to the strike zone, differentiated by pitch type (FA: four-seam; SI: two-seam). Each circle is a pitch location, with colour indicating velocity and size representing break distance. BreakDistance is the total distance of the horizontal and vertical break combined (using the simple Pythagorean formula).

Here are my observations based on Fig. 2:

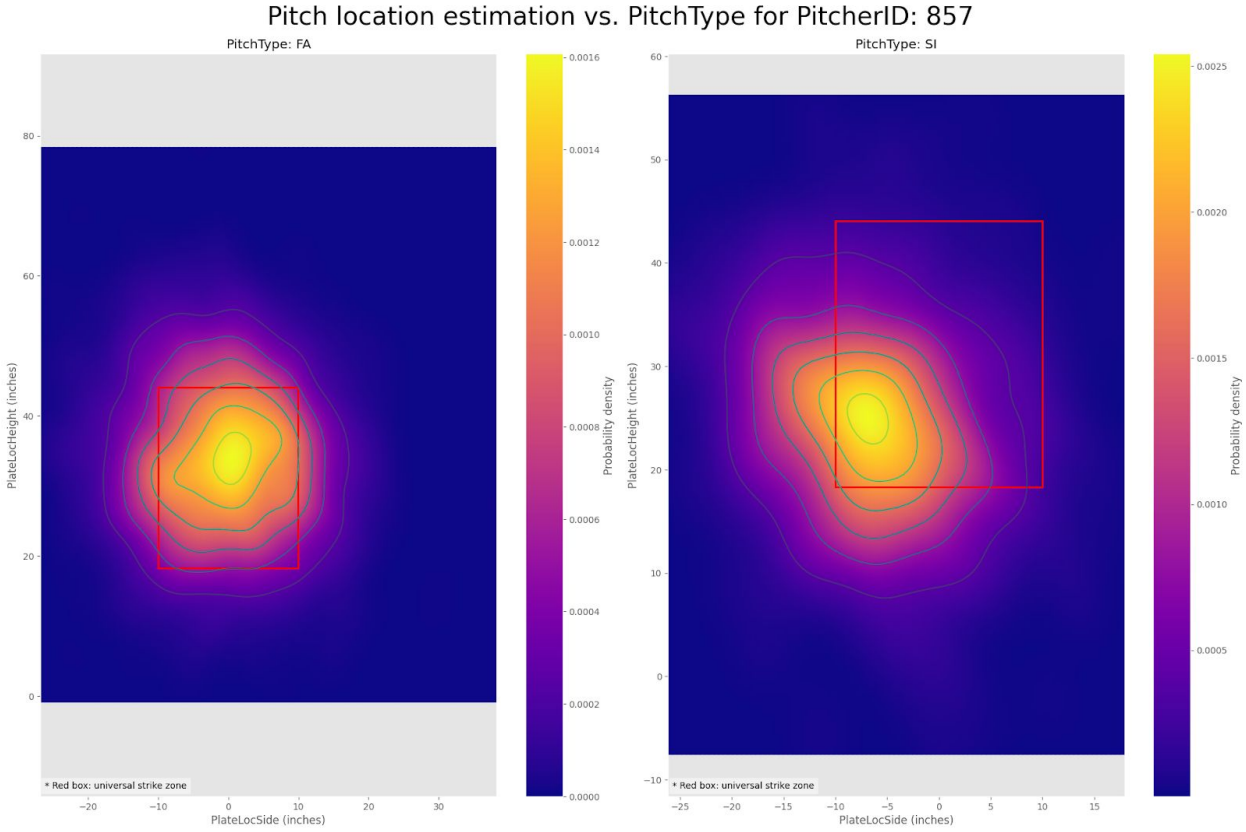
1) Pitcher 857 (Fig. 2a) and pitcher 114013 (Fig. 2b) both had their pitch scatter patterns going from top right to bottom left, whereas the rest three pitchers (Fig. 2c, d, e) had their pitch scatter patterns going from top left to bottom right. This can be explained by the fact that in the data, the first two players are left-handed and the rest three are right-handed.
2) Pitcher 857 (Fig. 2a) seems to have great control but little command when throwing four-seamers, with his pitch scatter pattern nicely centered inside the strike zone. When throwing two-seamers, however, he seems to have great command, with a scatter pattern concentrated near the bottom left corner of the strike zone. For a professional baseball player and an ML starter, it is almost certain that he intended to throw his two-seamers at the corner, and he was mostly able to do it.
3) Pitcher 114013 (Fig. 2b) seems to have better command than 857 (Fig. 2c) with four-seamers, as the scatter pattern is concentrated towards the top right corner. For

two-seamers, he demonstrated similar command compared to 857, but the spread seems to be slightly larger.

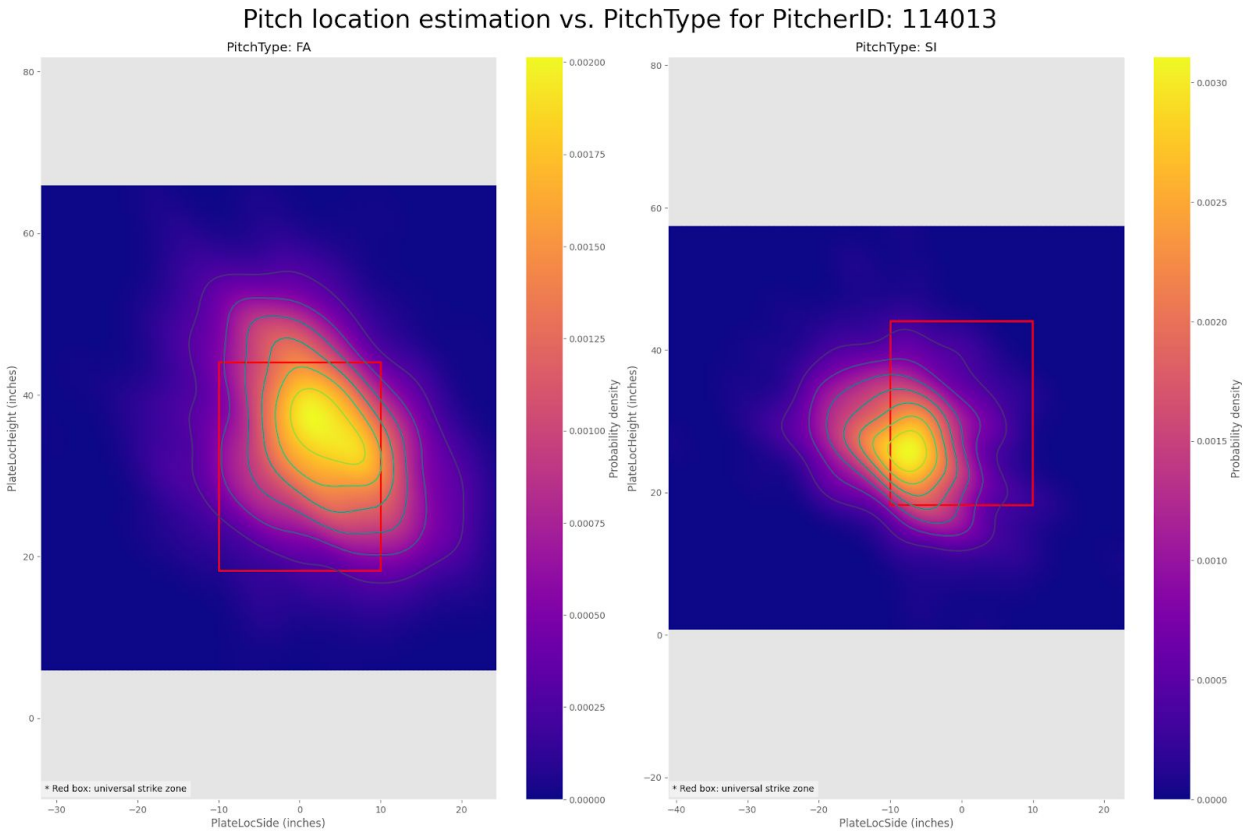
4) Pitcher 2696 (Fig. 2c) and 2779 (Fig. 2e) had high diagonal spread for both four-seamers and two-seamers. While their two-seamers showed some command with spread centers somewhat close to the strike zone edge, the large spread is just not quite comparable to pitcher 857 (Fig. 2a) and 114013 (Fig. 2b).
5) Pitcher 1594 (Fig. 2d) showed decent control with both four-seam and two-seam fastballs mostly scattered around the center of the strike zone. But according to my problem definition (i.e. putting the ball on the edge is the best command), he lacks command. Maybe his true intention was always to put the ball at the center, but we cannot know that for sure in the data.

The above observations can be verified when the pitch scatter pattern is modelled as a probability distribution (Fig. 3).

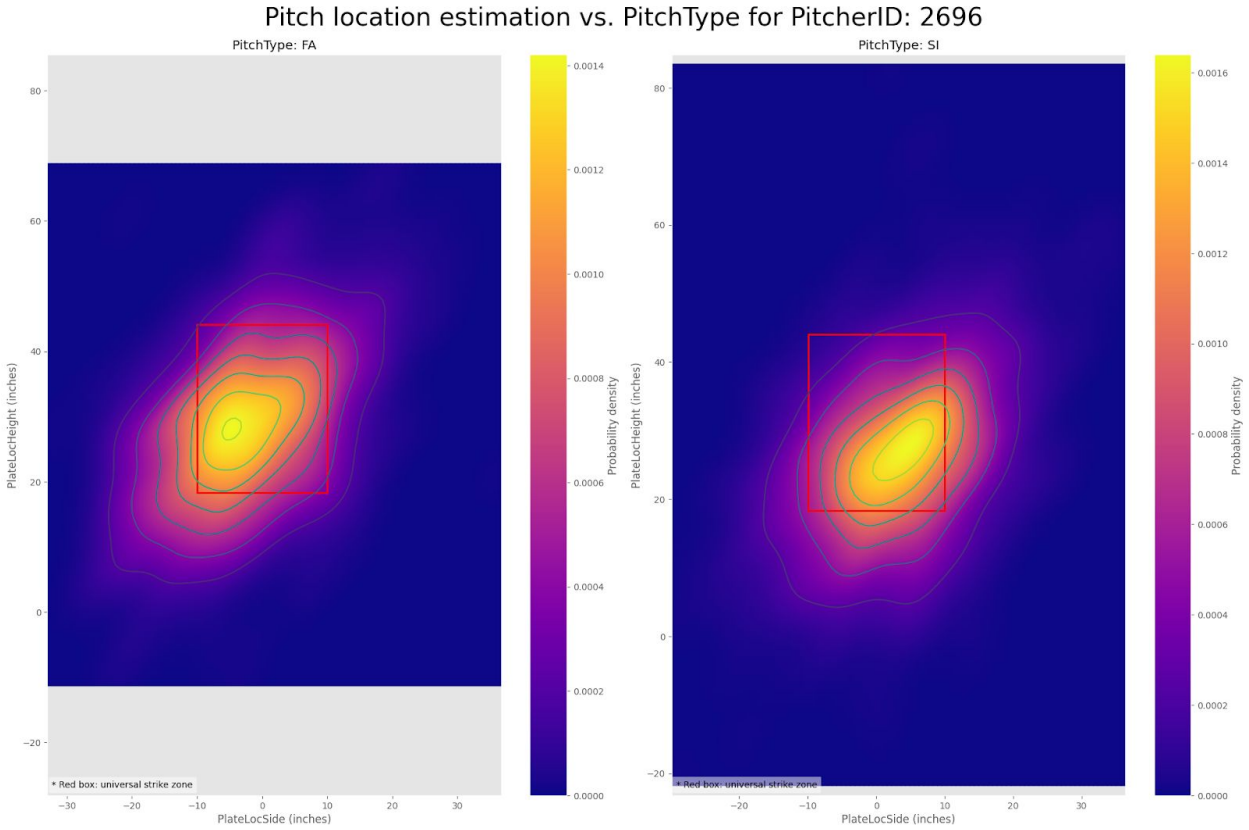
(a)



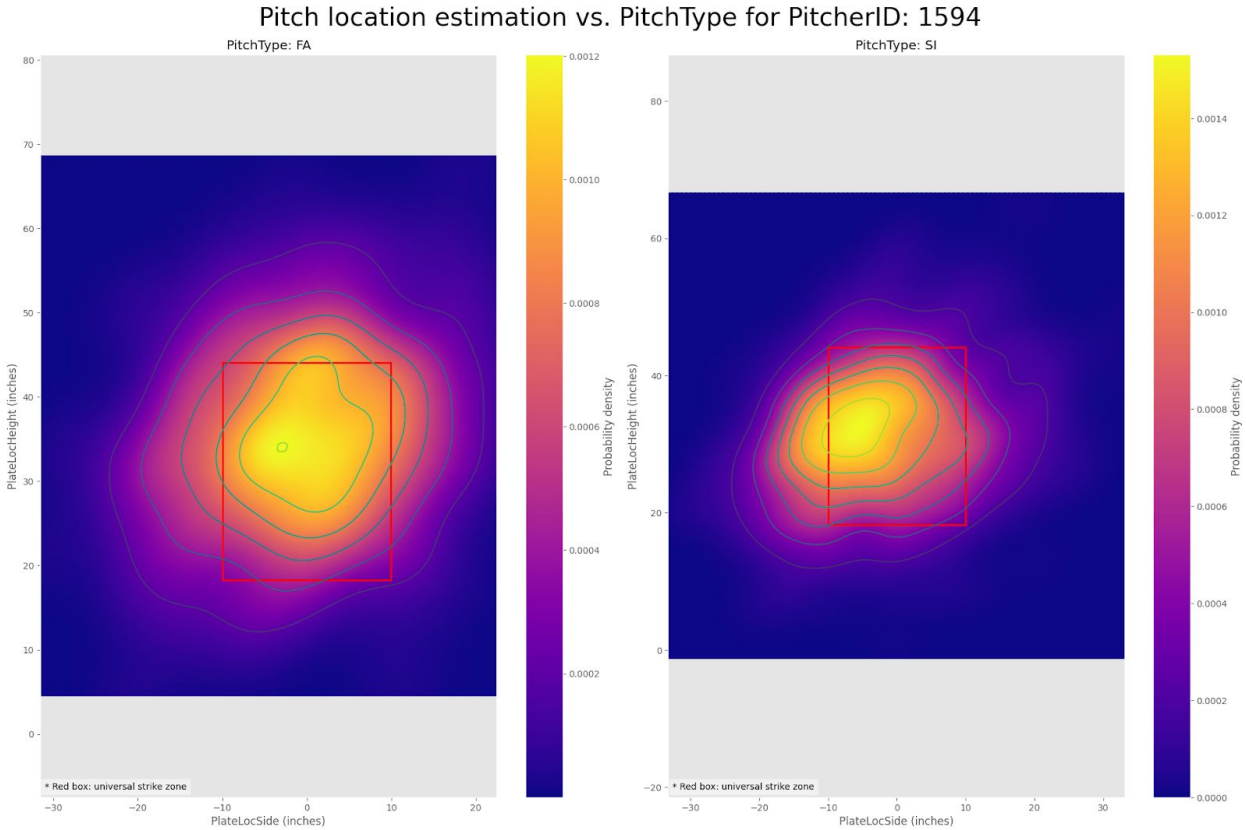
(b)


Pitch location estimation vs. PitchType for PitcherID: 114013
PitchType: FA
PitchType: SI
PlateLocHeight (inches)
PlateLocSide (inches)
Probability density
* Red box: universal strike zone

(c)


Pitch location estimation vs. PitchType for PitcherID: 2696
PitchType: FA
PitchType: SI
PlateLocHeight (inches)
PlateLocSide (inches)
Probability density
* Red box: universal strike zone

(d)


Pitch location estimation vs. PitchType for PitcherID: 1594
PitchType: FA
PitchType: SI
PlateLocHeight (inches)
PlateLocSide (inches)
Probability density
* Red box: universal strike zone

(e)

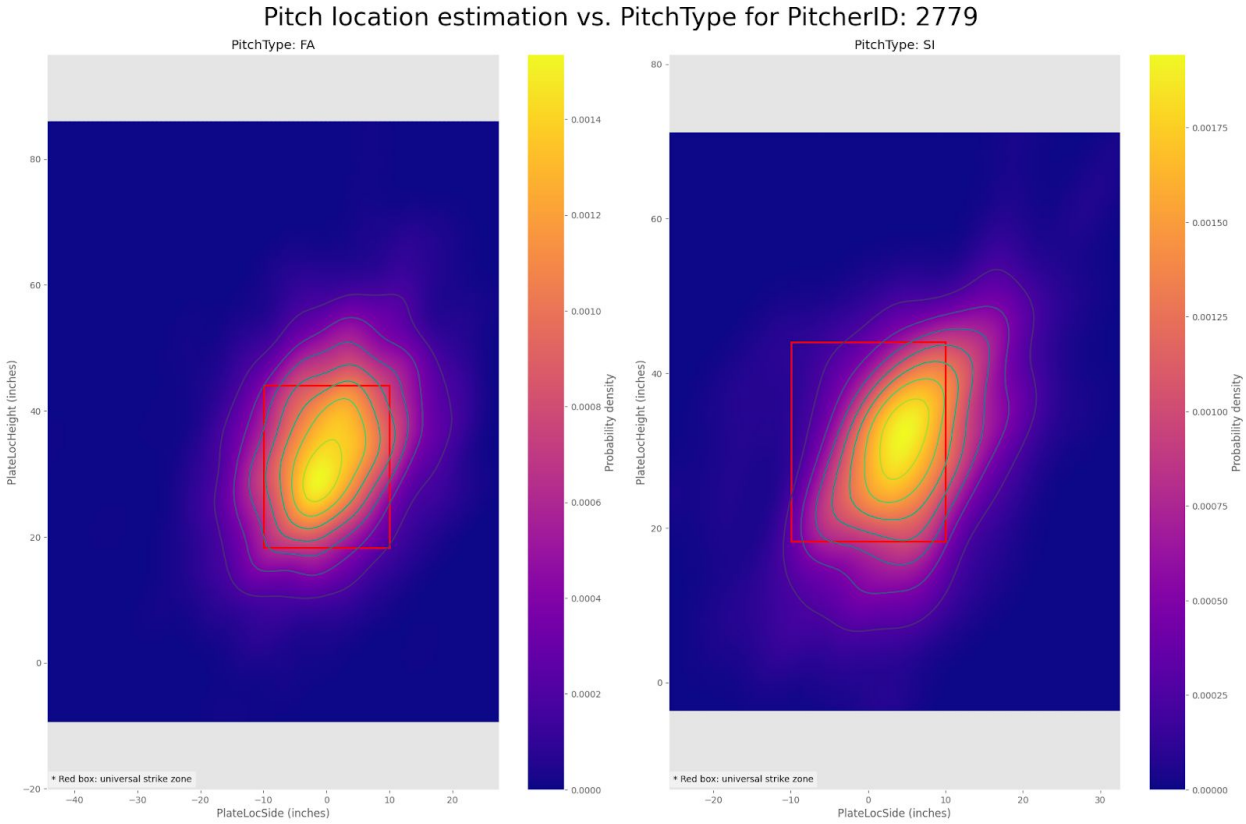


Figure 3. The five different pitchers' pitch locations shown as probability density in relation to the strike zone, differentiated by pitch type (FA: four-seam; SI: two-seam). The probability distribution is modelled using Gaussian KDE. The faster the colour gradient changes, the more concentrated the pitch scatter is estimated to be.

Answer

So, my conclusion is, pitcher 114013 likely has the best fastball pitch command overall (for both four-seamers and two-seamers). If we only look at two-seam fastballs, pitcher 857 seems to have the best command.

In-depth Analysis

The scatter pattern of the pitch location does not seem to be correlated with velocity (Fig. 3, 4). Except pitcher 857 who had a faster four-seamer than two-seamer, the rest pitchers had a similar velocity for their four-seamers and two-seamers (Fig. 5a).

On average, four-seamers had a higher vertical break than two-seamers despite the handiness of the pitcher (Fig. 5b). Left-handed pitchers had a negative horizontal break (away from the 3rd base) and right-handed pitchers had a positive horizontal break (towards the 3rd base; Fig 5c).

The scatter pattern of the pitch location does not seem to be correlated with break distance (the hypotenuse of vertical and horizontal break; Fig. 3, 4). This is not that surprising because fastballs should not have that much movement and most pitches did end up with a similar break distance (Fig. 5f).

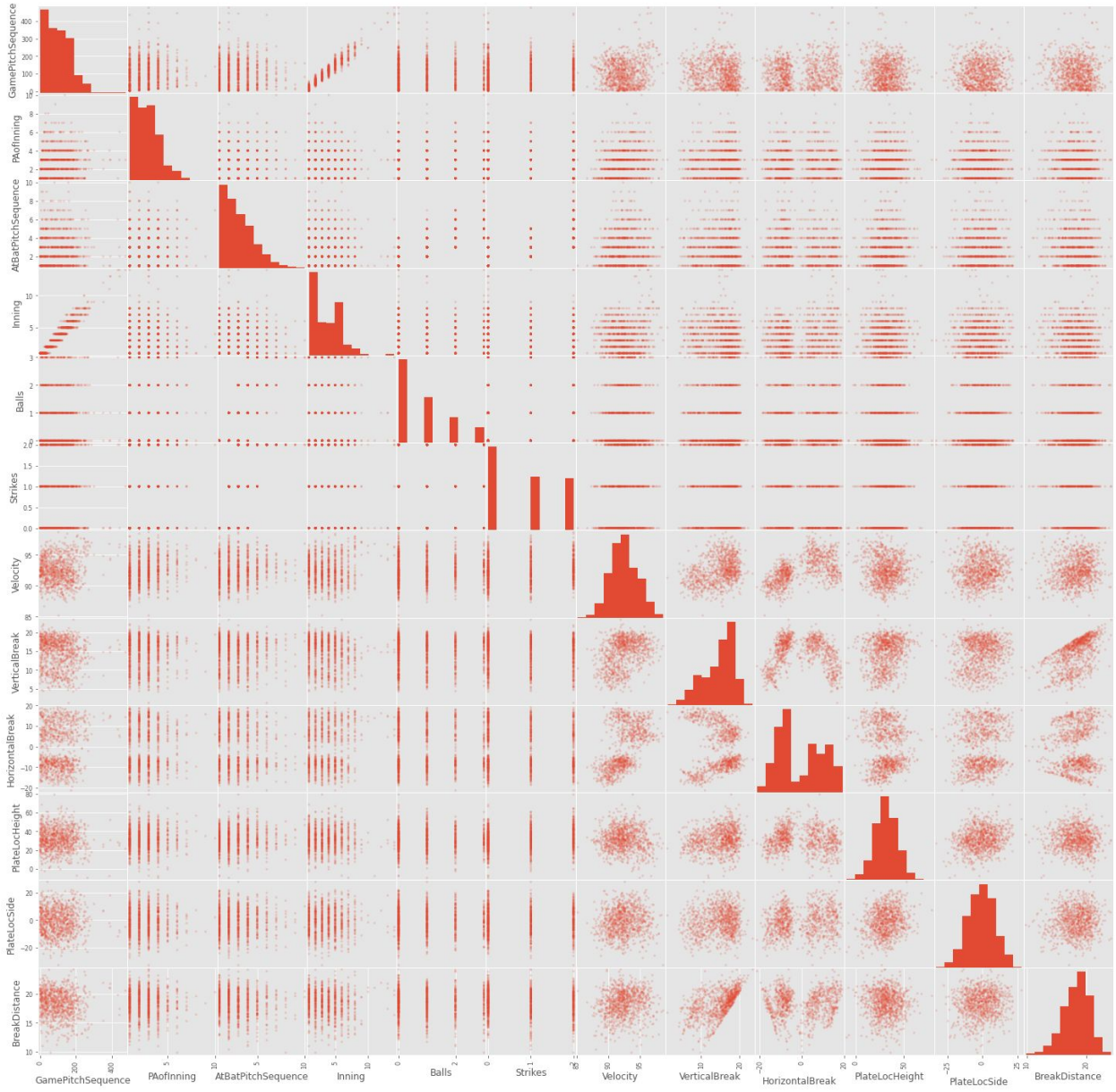


Figure 4. Pair plots of the numerical attributes used in the analysis.

(a)


Velocity by Pitcher
PitchType
FA
SI
Velocity (inches)
100
98
96
94
92
90
88
86
84
857
1594
2696
2779
114013
PitcherID

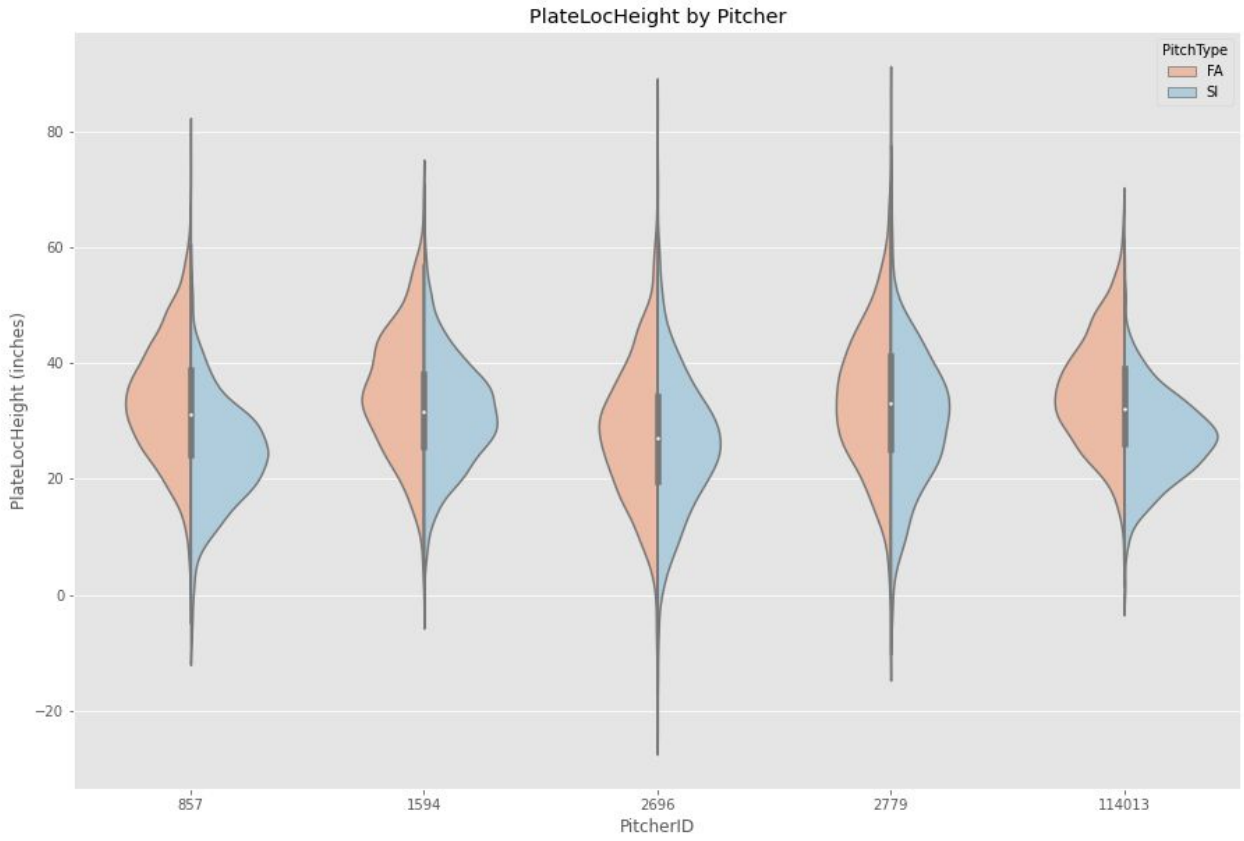
(b)


VerticalBreak by Pitcher
PitchType
FA
SI
VerticalBreak (inches)
25
20
15
10
5
0
857
1594
2696
2779
114013
PitcherID

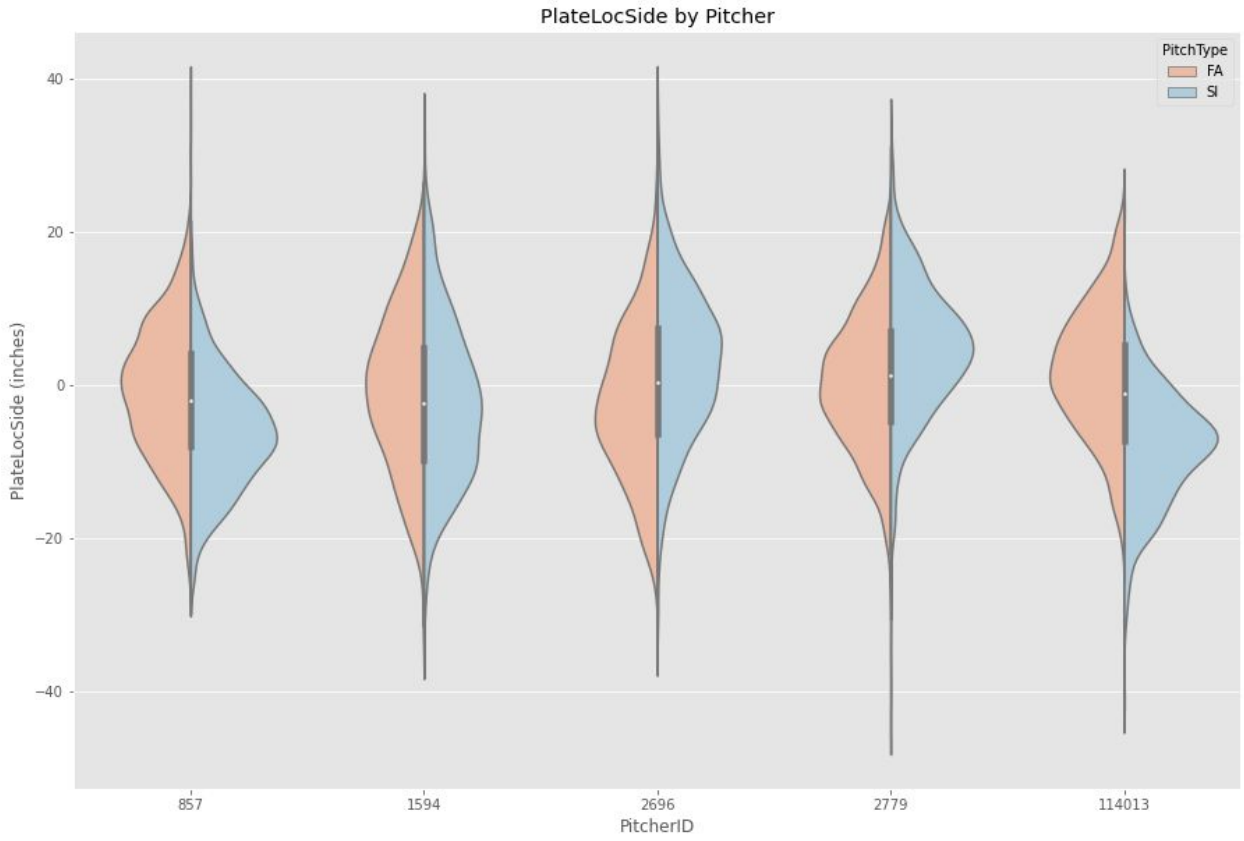
(c)


HorizontalBreak by Pitcher
PitchType
FA
SI
HorizontalBreak (inches)
20
10
0
−10
−20
857
1594
2696
2779
114013
PitcherID

(d)


PlateLocHeight by Pitcher
PitchType
FA
SI
PlateLocHeight (inches)
80
60
40
20
0
−20
857
1594
2696
2779
114013
PitcherID

(e)


PlateLocSide by Pitcher
PitchType
FA
SI
PlateLocSide (inches)
40
20
0
−20
−40
857
1594
2696
2779
114013
PitcherID

(f)

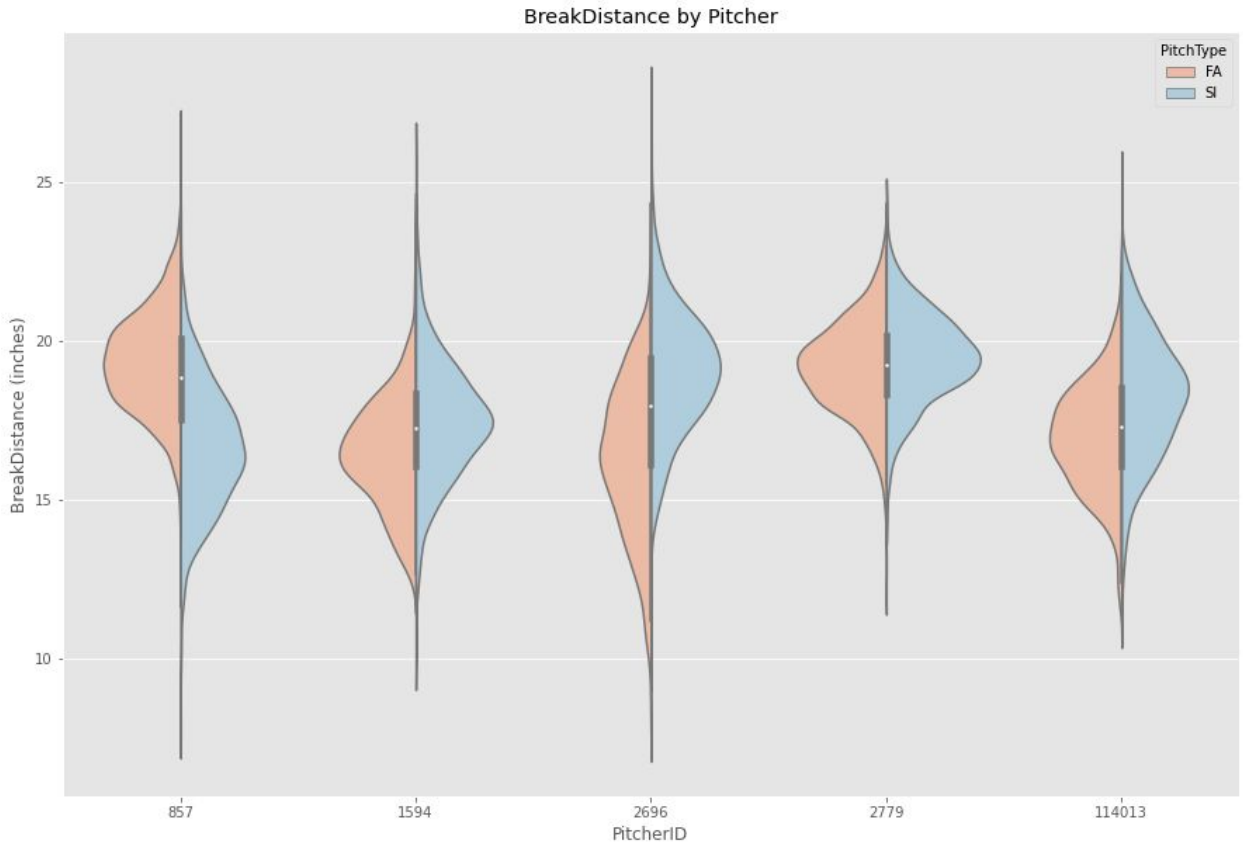
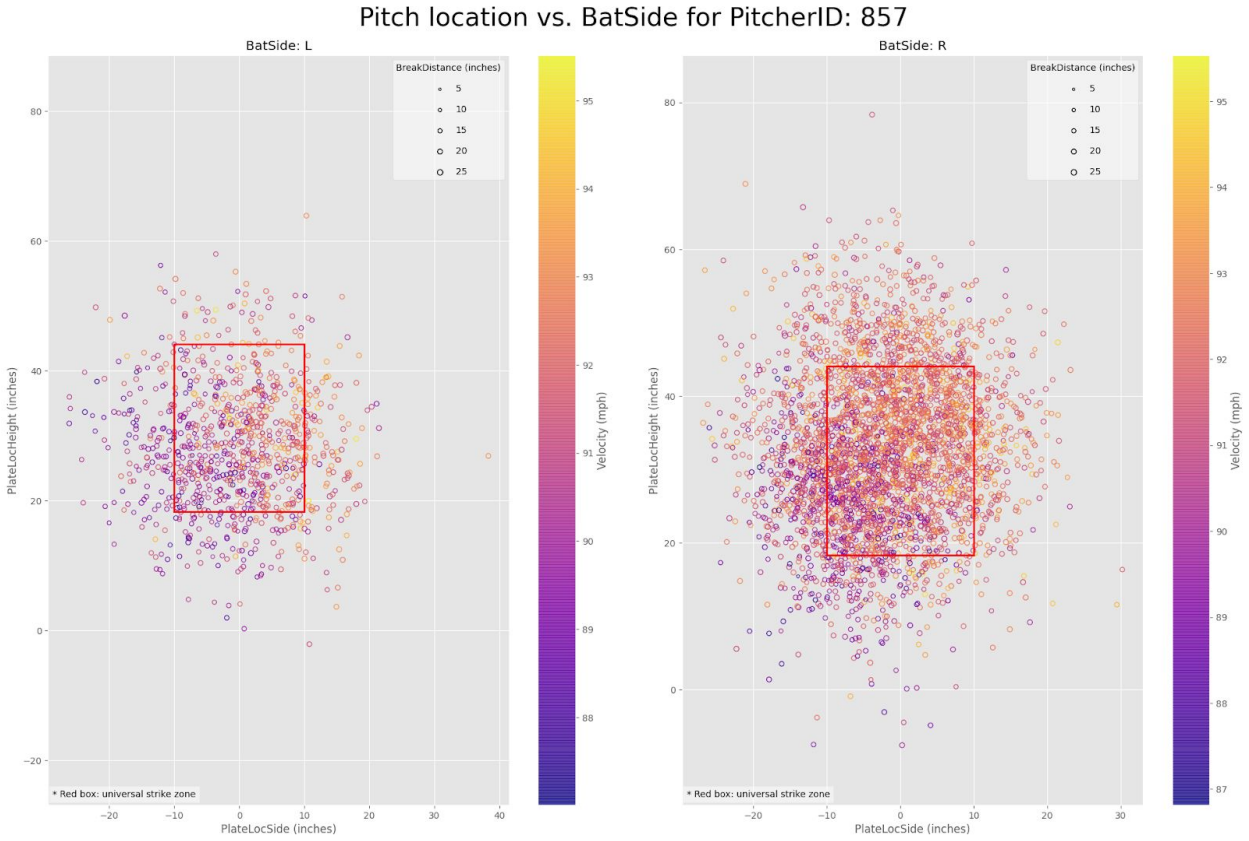


Figure 5. Violin plots of velocity, break, and pitch location vs. PitchType for each pitcher.

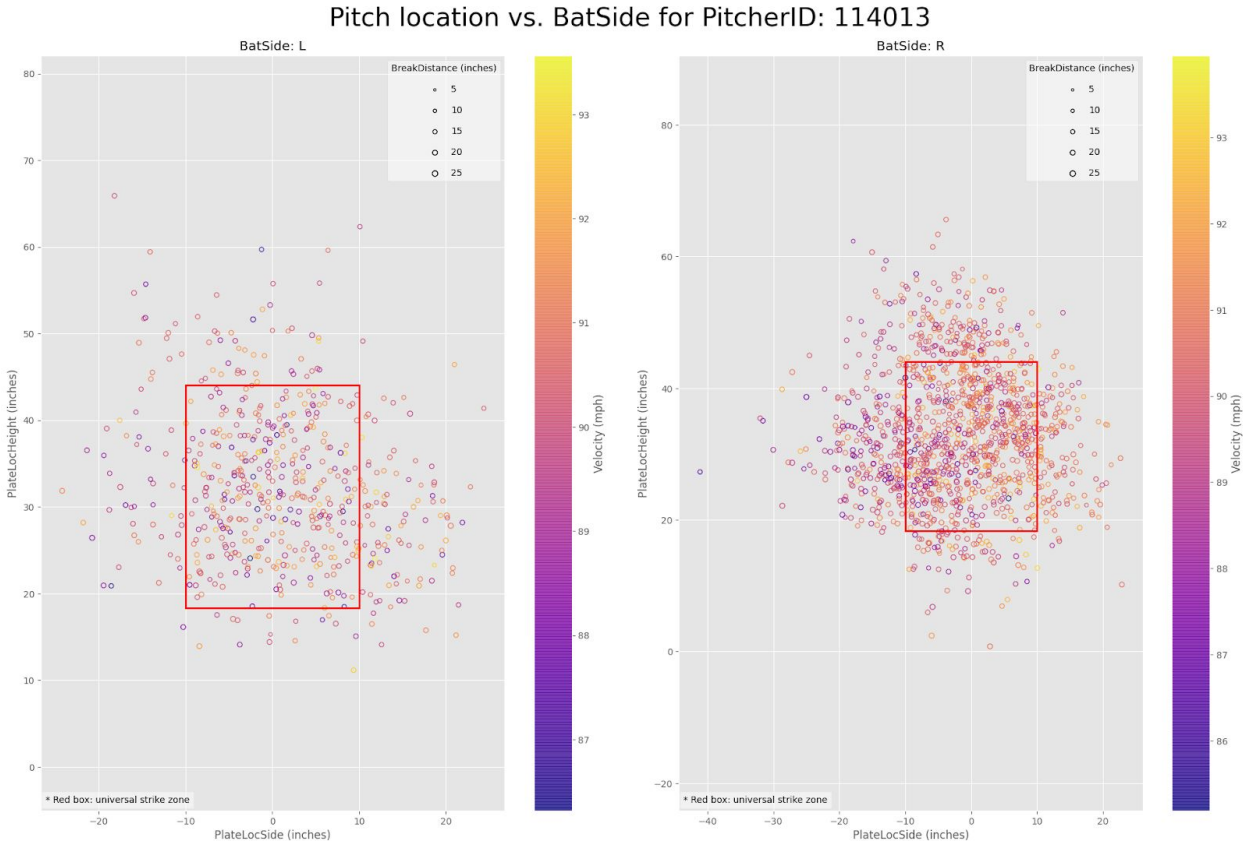
Previous studies have suggested that pitch command is influenced by factors such as counts and bat side (Jon 2018; Melling 2018; BaseballCloudBlog 2020).

When plotting the pitch location vs. bat side (Fig. 6, 7, 8), I found that only pitcher 1594 showed a mirroring scatter pattern between the left and right bat side, meaning he probably tried to command differently based on the batter's handiness. However, this analysis cannot disprove the presence of pitch command in the other pitchers. For example, both pitcher 857 and 114013 seem to have good command by throwing their two-seamers at the bottom-left corner of the strike zone (Fig. 8a, b) regardless of the batter's handiness.

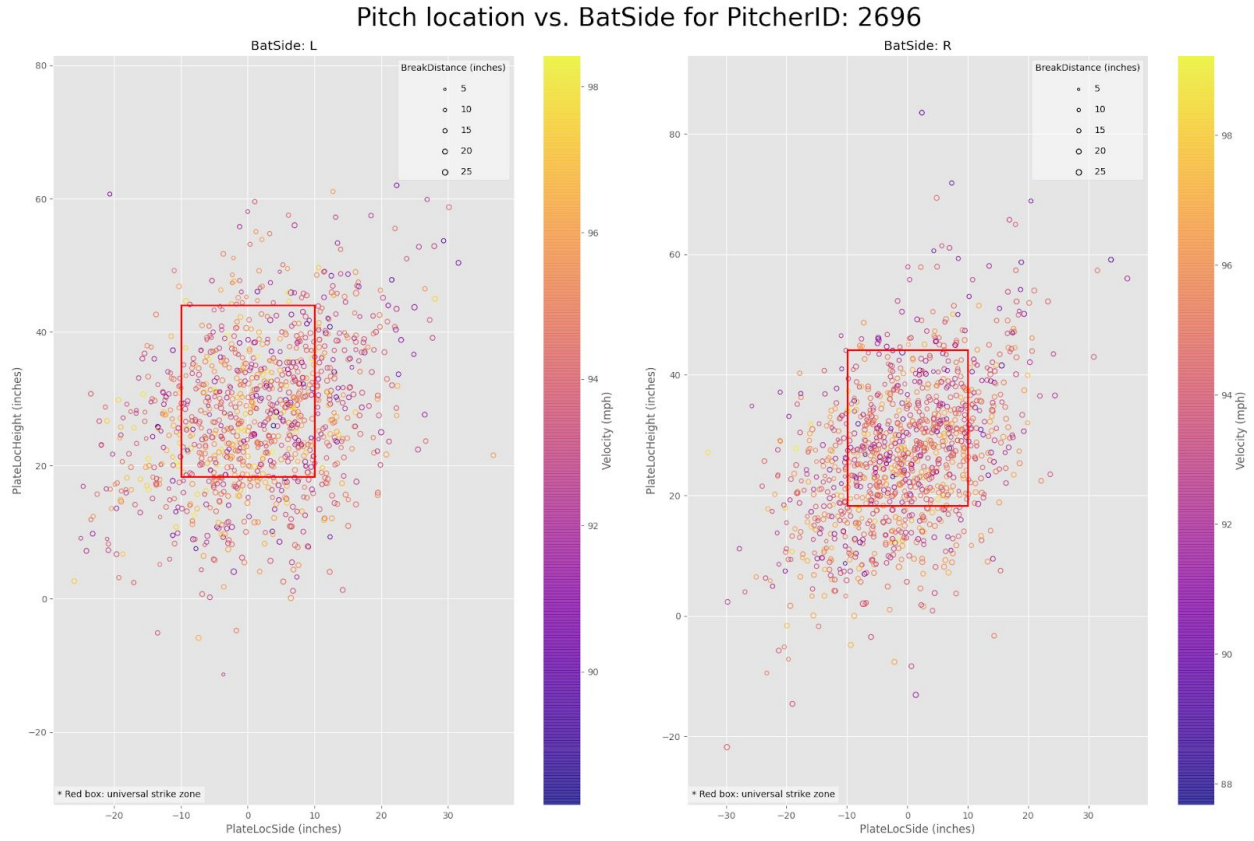
(a)


Pitch location vs. BatSide for PitcherID: 857
BatSide: L
BatSide: R
BreakDistance (inches)
5
10
15
20
25
Velocity (mph)
PlateLocHeight (inches)
PlateLocSide (inches)
* Red box: universal strike zone

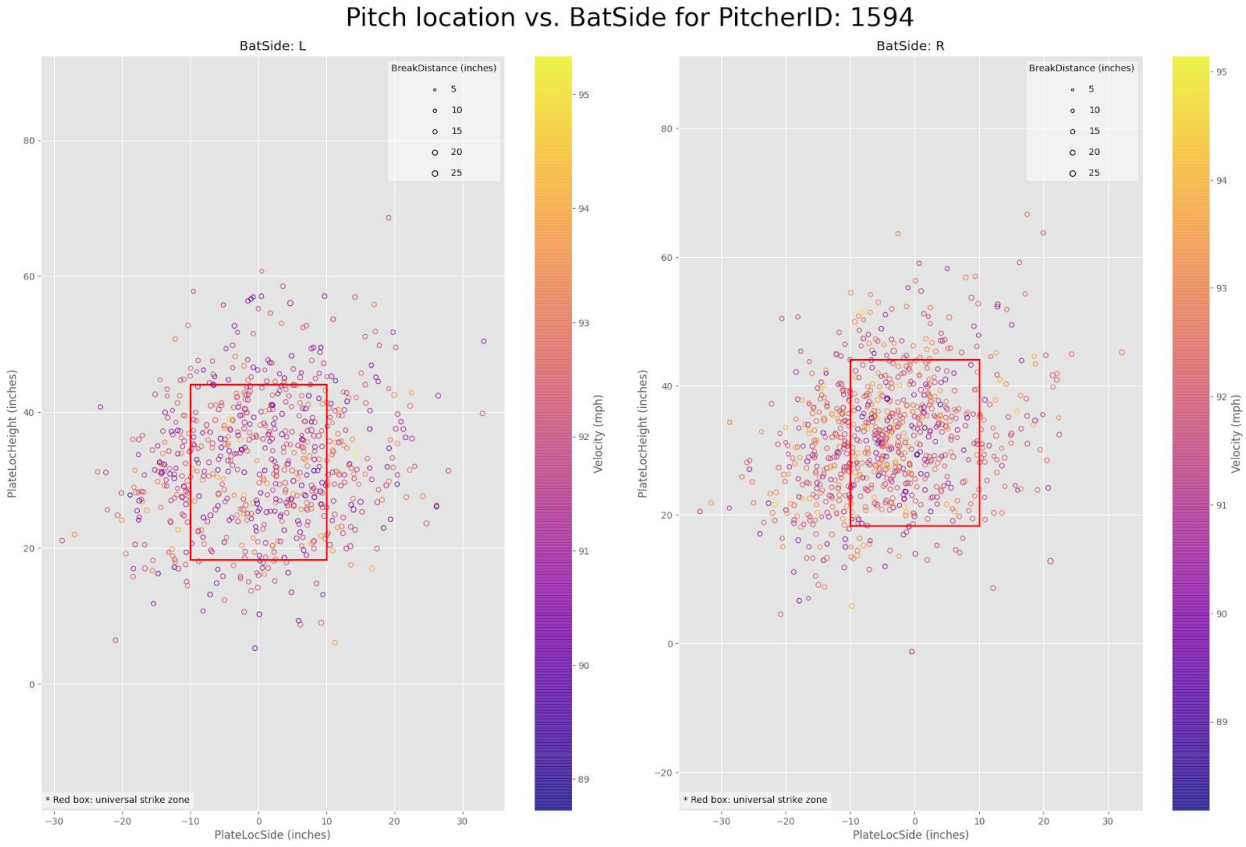
(b)


Pitch location vs. BatSide for PitcherID: 114013
BatSide: L
BatSide: R
BreakDistance (inches)
5
10
15
20
25
PlateLocHeight (inches)
PlateLocSide (inches)
Velocity (mph)
* Red box: universal strike zone

(c)


Pitch location vs. BatSide for PitcherID: 2696
BatSide: L
BatSide: R
BreakDistance (inches)
5
10
15
20
25
PlateLocHeight (inches)
PlateLocSide (inches)
Velocity (mph)
* Red box: universal strike zone

(d)


Pitch location vs. BatSide for PitcherID: 1594
BatSide: L
BatSide: R
BreakDistance (inches)
5
10
15
20
25
PlateLocHeight (inches)
PlateLocSide (inches)
Velocity (mph)
* Red box: universal strike zone

(e)

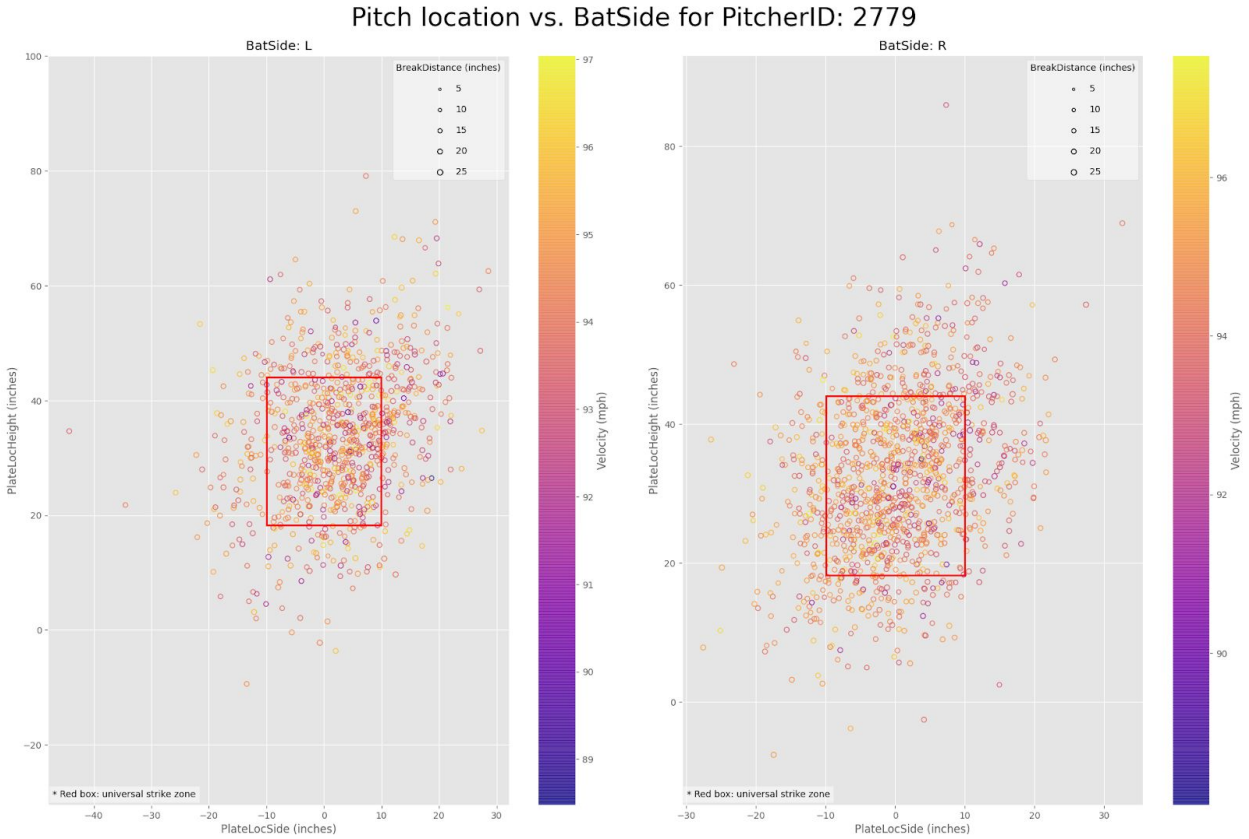
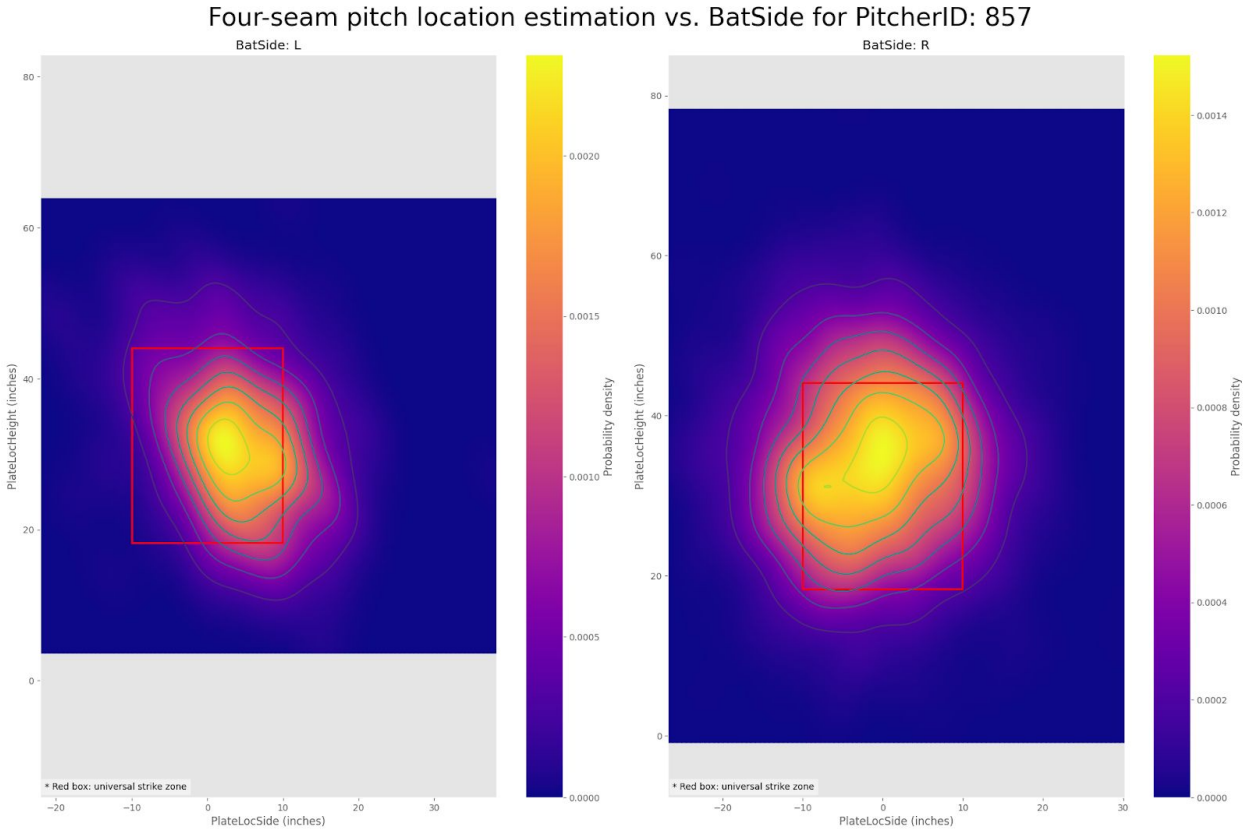
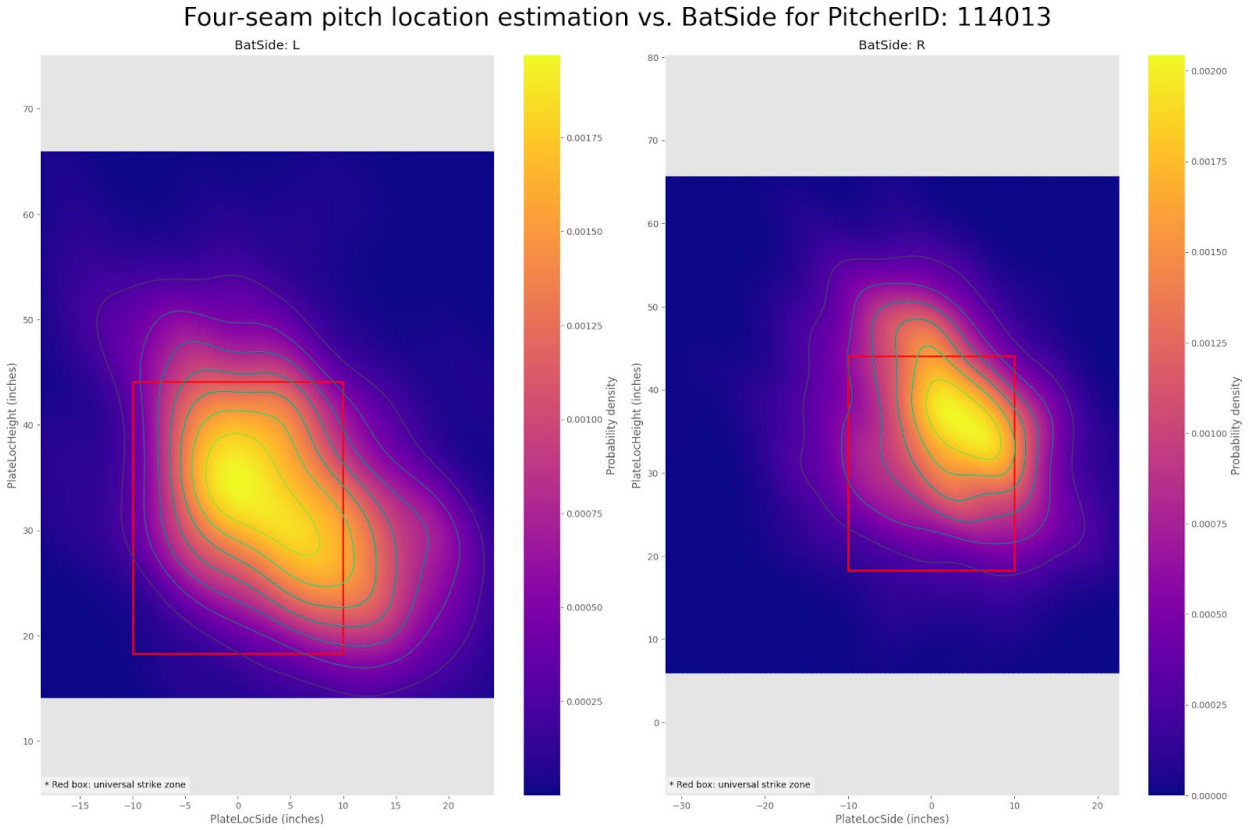


Figure 6. The five different pitchers' pitch locations shown as scatter plots in relation to the strike zone, differentiated by bat side. Each circle is a pitch location, with colour indicating velocity and size representing break distance. BreakDistance is the total distance of the horizontal and vertical break combined (using the simple Pythagorean formula).

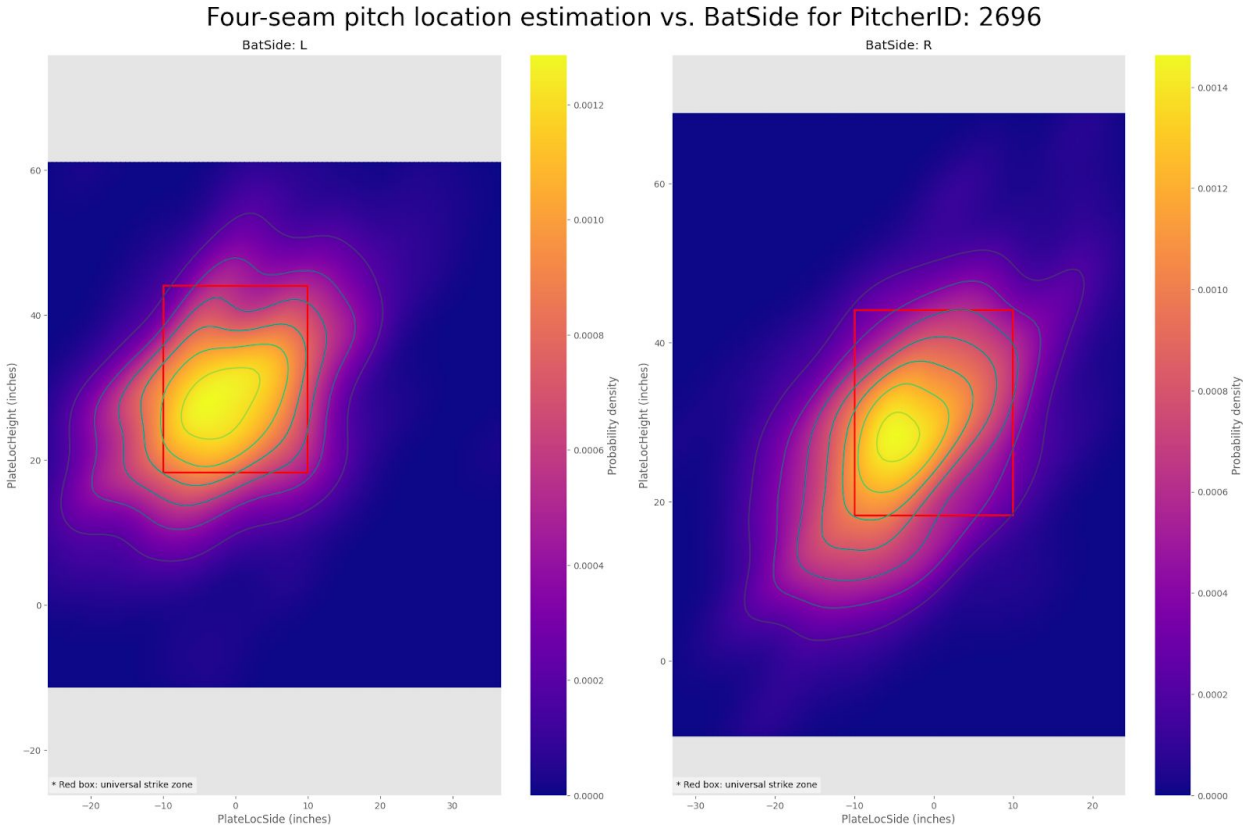
(a)


Four-seam pitch location estimation vs. BatSide for PitcherID: 857
BatSide: L
BatSide: R
PlateLocHeight (inches)
PlateLocSide (inches)
Probability density
* Red box: universal strike zone

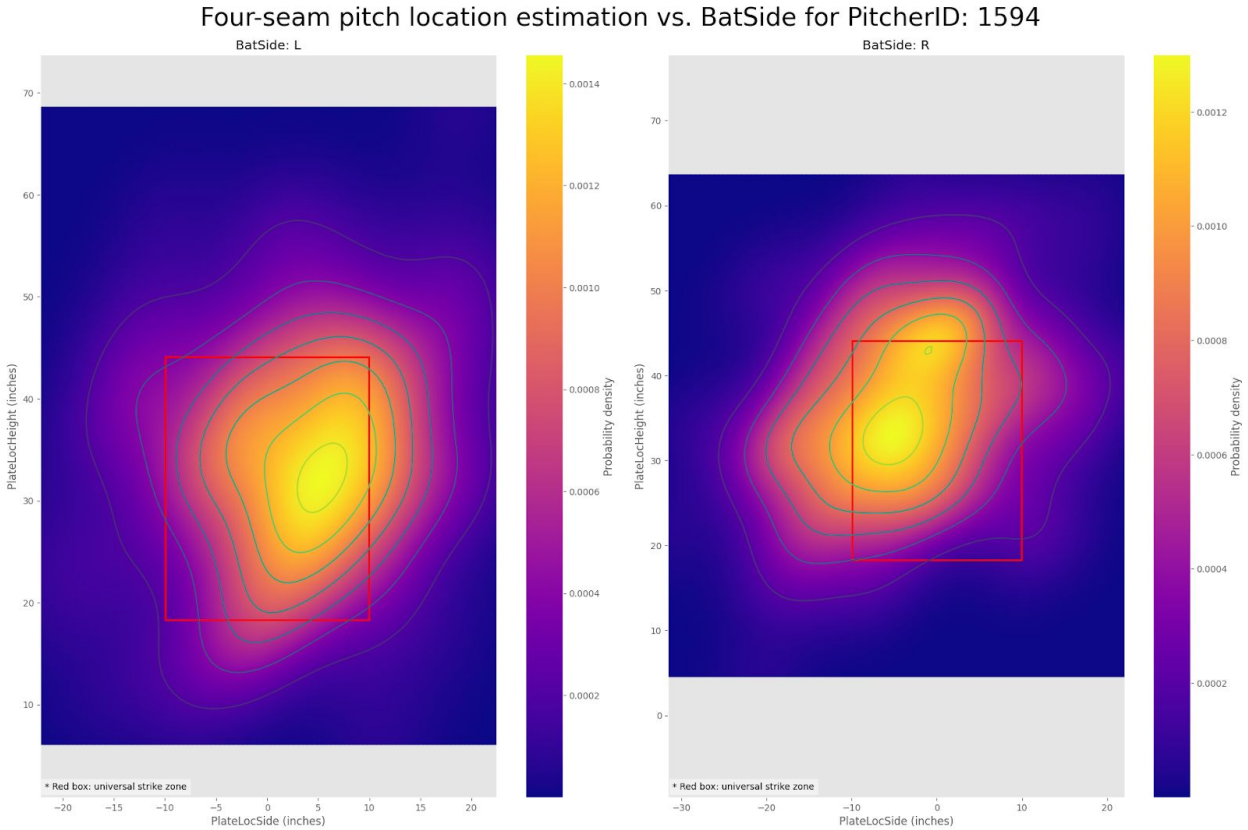
(b)


Four-seam pitch location estimation vs. BatSide for PitcherID: 114013
BatSide: L
BatSide: R
PlateLocHeight (inches)
PlateLocSide (inches)
Probability density
* Red box: universal strike zone

(c)


Four-seam pitch location estimation vs. BatSide for PitcherID: 2696
BatSide: L
BatSide: R
PlateLocHeight (inches)
PlateLocSide (inches)
Probability density
* Red box: universal strike zone

(d)


Four-seam pitch location estimation vs. BatSide for PitcherID: 1594
BatSide: L
BatSide: R
PlateLocHeight (inches)
PlateLocSide (inches)
Probability density
* Red box: universal strike zone

(e)

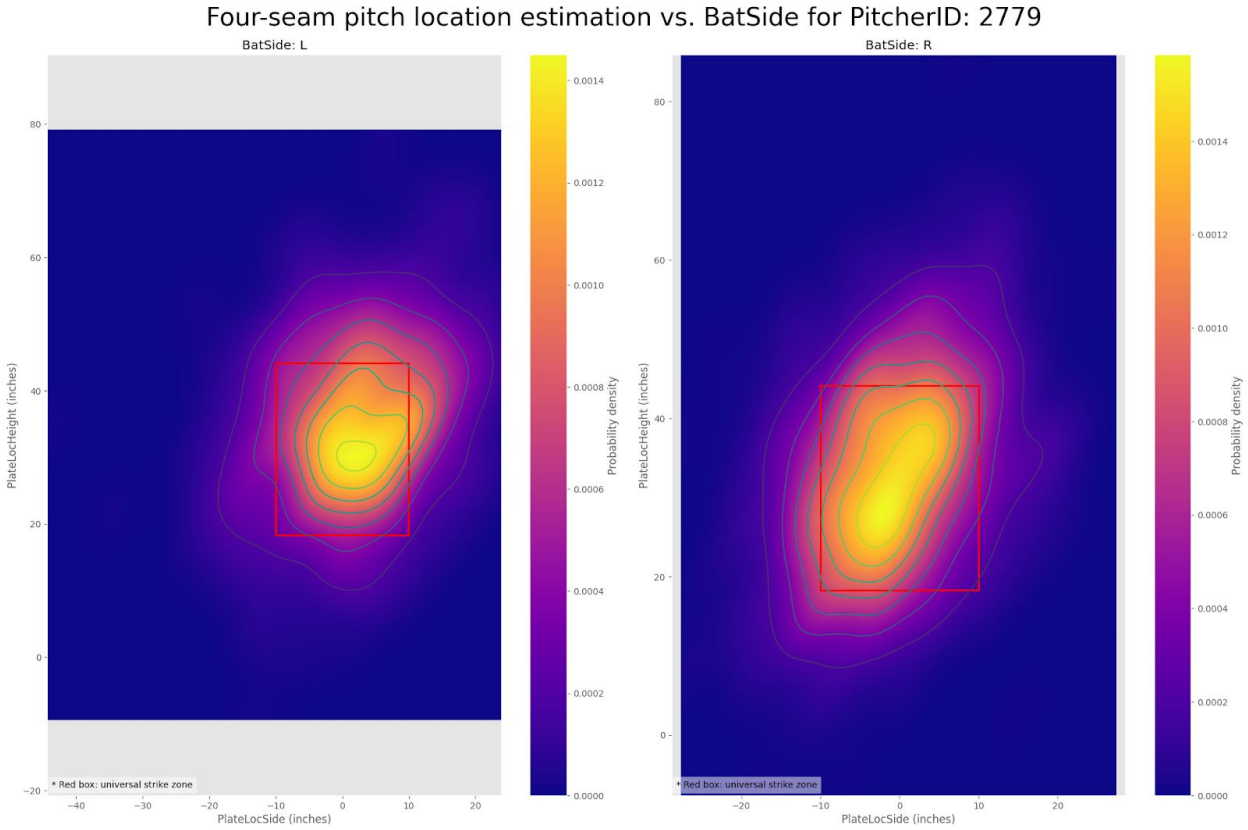
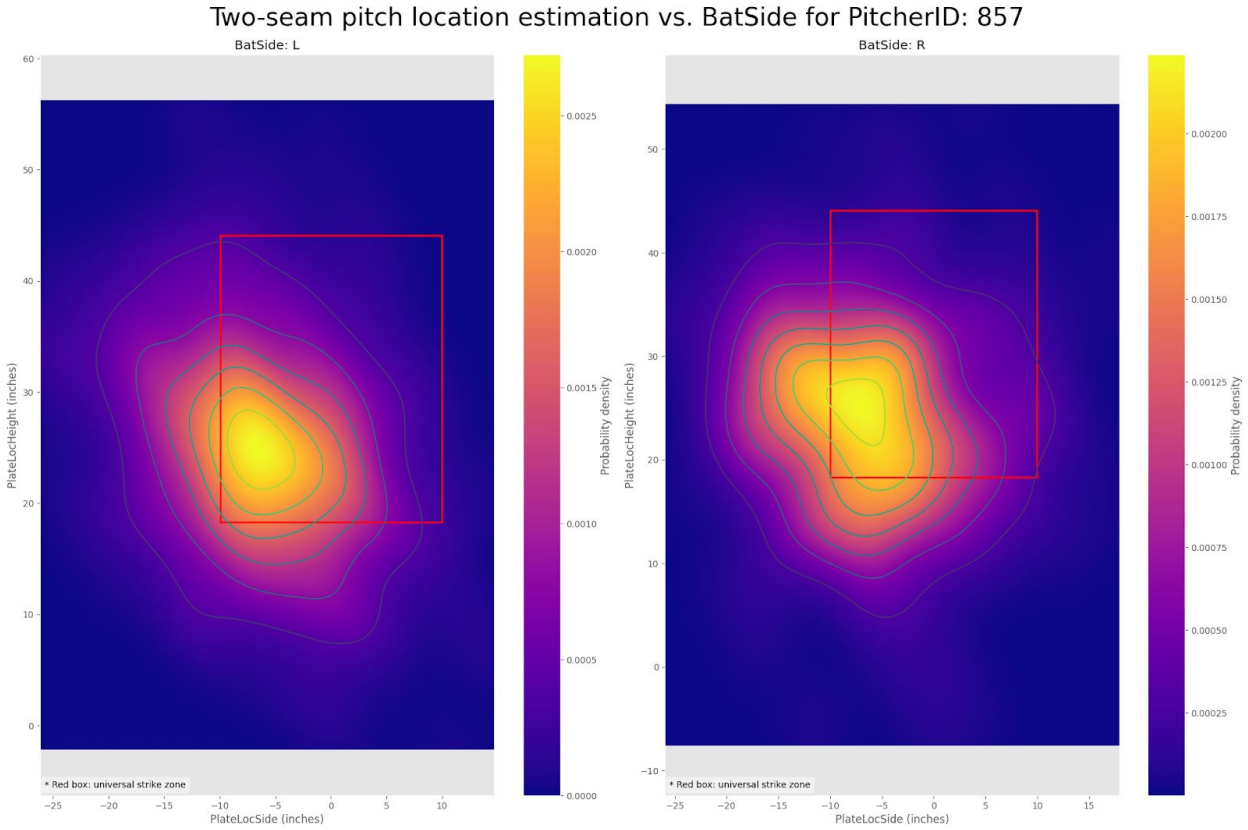
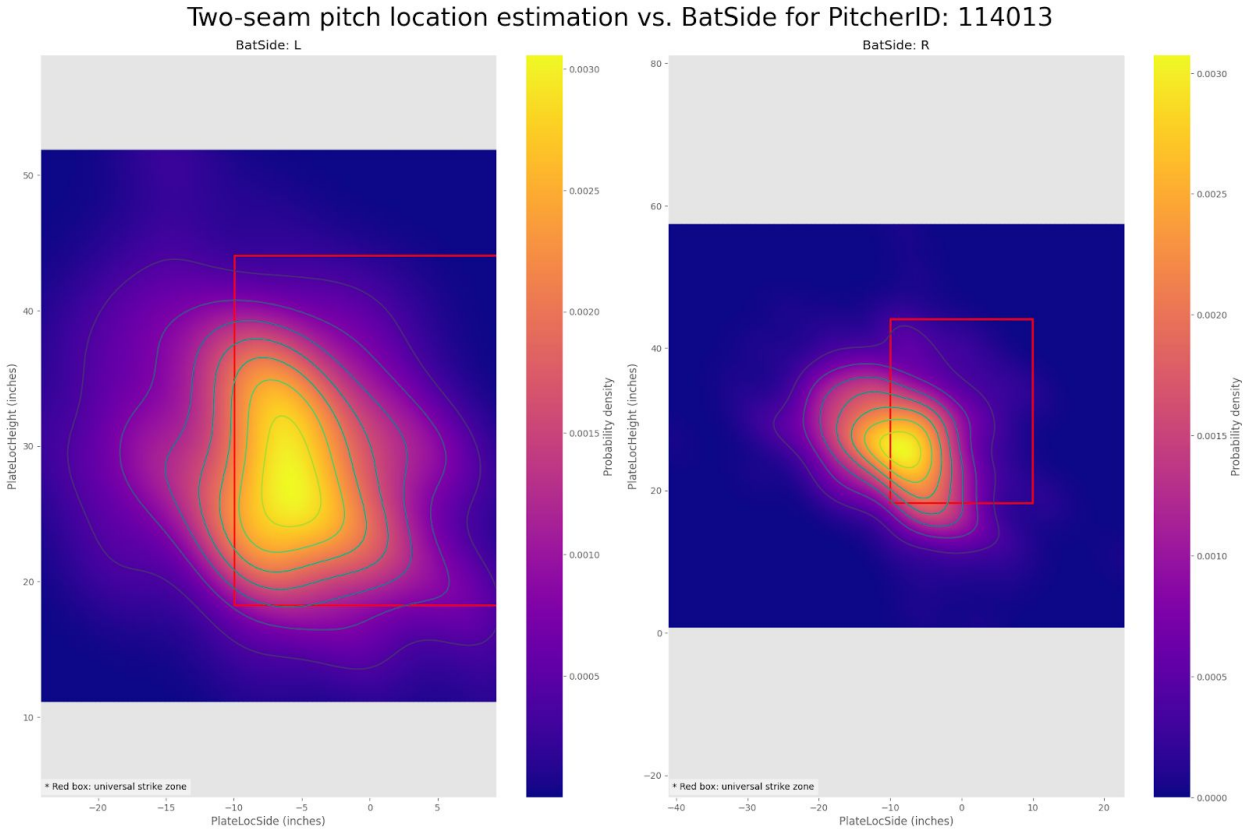


Figure 7. The five different pitchers' four-seam pitch locations shown as probability density in relation to the strike zone, differentiated by bat side. The probability distribution is modelled using Gaussian KDE. The faster the colour gradient changes, the more concentrated the pitch scatter is estimated to be.

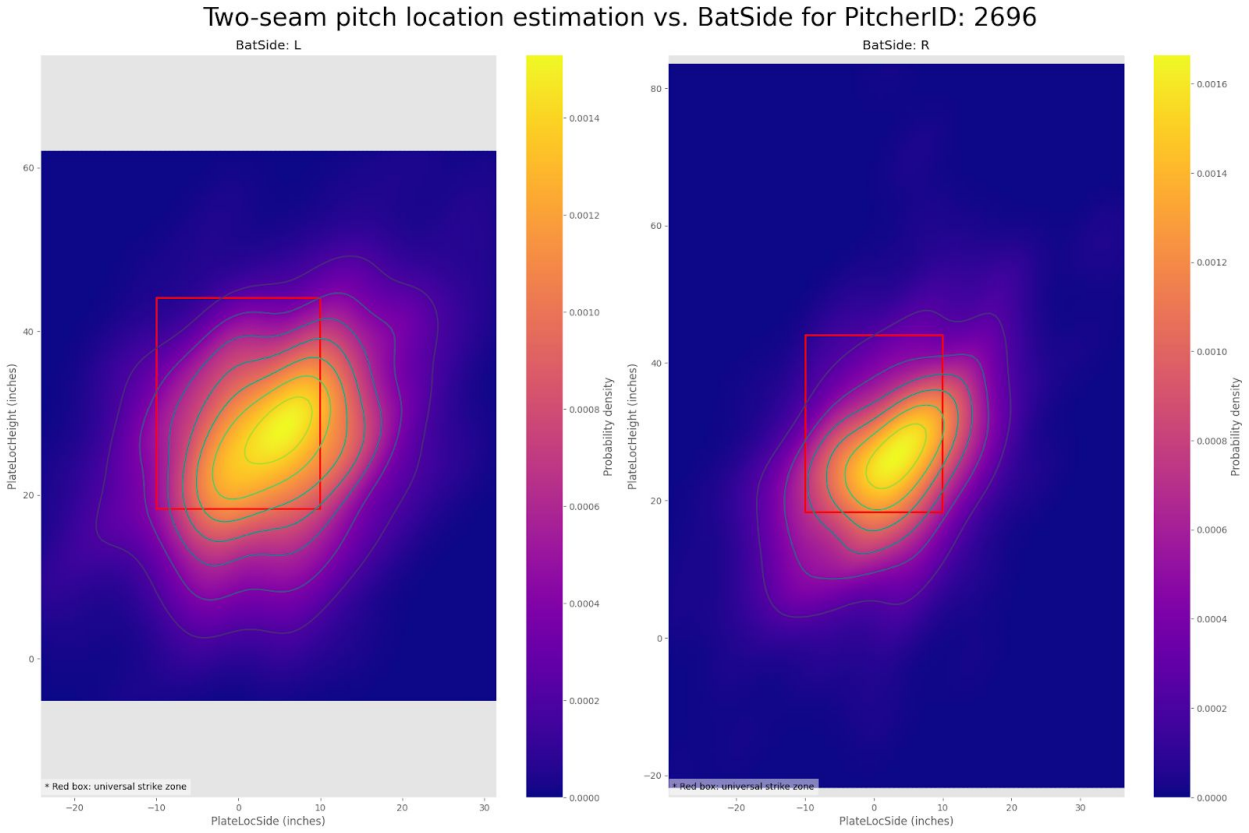
(a)


Two-seam pitch location estimation vs. BatSide for PitcherID: 857
BatSide: L
BatSide: R
PlateLocHeight (inches)
PlateLocSide (inches)
Probability density
* Red box: universal strike zone

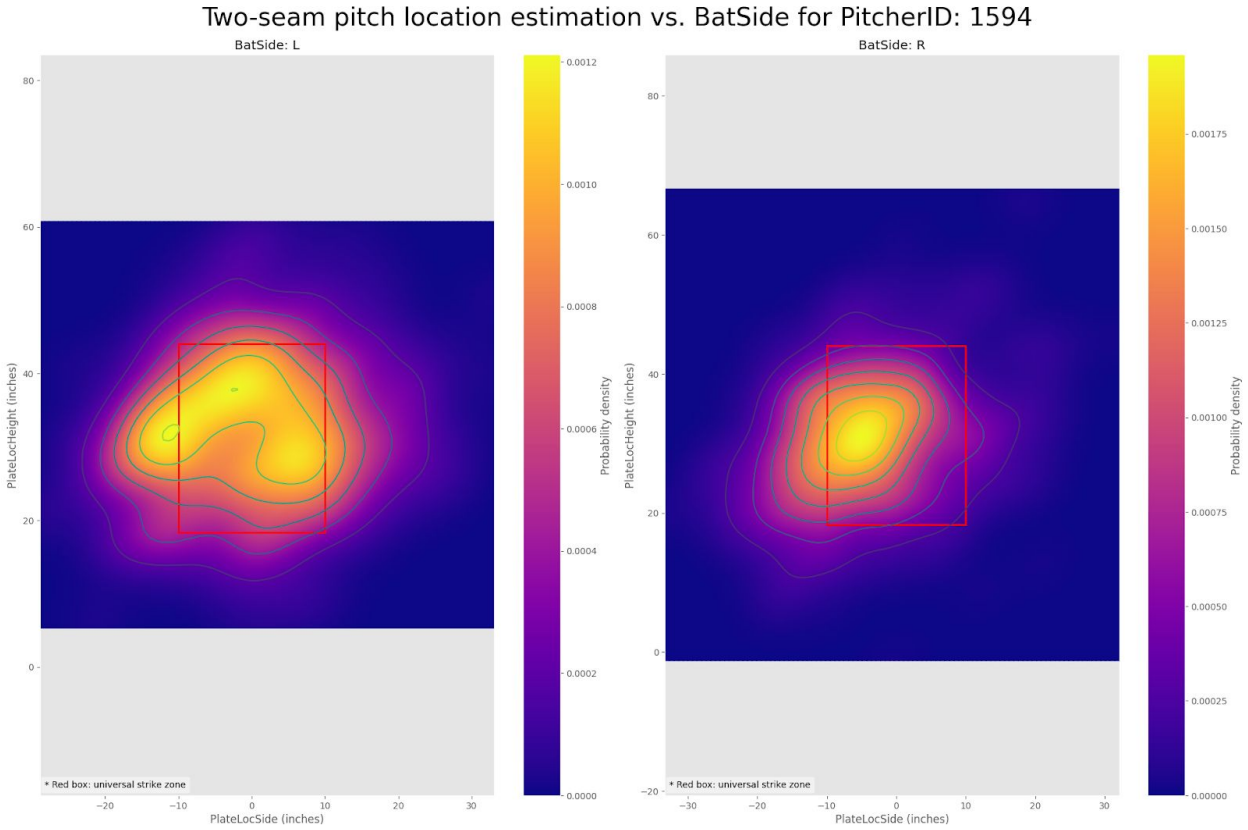
(b)


Two-seam pitch location estimation vs. BatSide for PitcherID: 114013
BatSide: L
BatSide: R
PlateLocHeight (inches)
PlateLocSide (inches)
Probability density
* Red box: universal strike zone

(c)


Two-seam pitch location estimation vs. BatSide for PitcherID: 2696
BatSide: L
BatSide: R
PlateLocHeight (inches)
PlateLocSide (inches)
Probability density
* Red box: universal strike zone

(d)


Two-seam pitch location estimation vs. BatSide for PitcherID: 1594
BatSide: L
BatSide: R
PlateLocHeight (inches)
PlateLocSide (inches)
Probability density
* Red box: universal strike zone

(e)

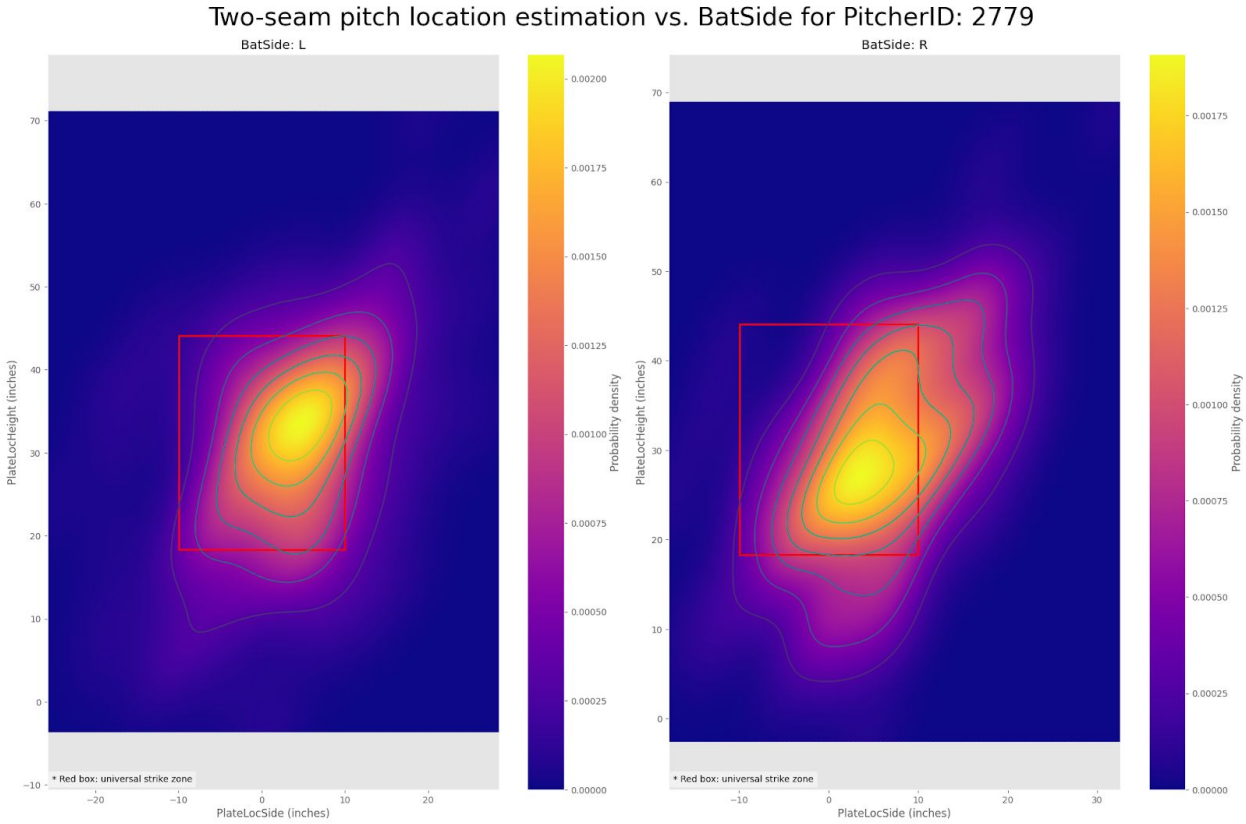
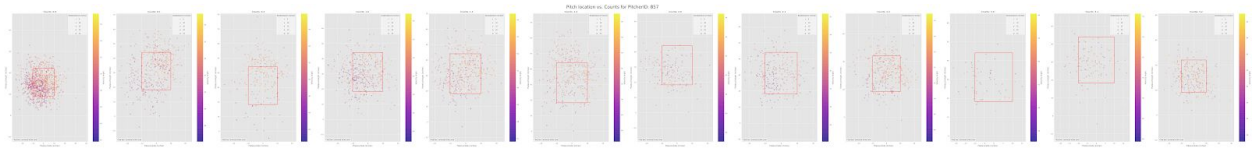


Figure 8. The five different pitchers' two-seam pitch locations shown as probability density in relation to the strike zone, differentiated by bat side. The probability distribution is modelled using Gaussian KDE. The faster the colour gradient changes, the more concentrated the pitch scatter is estimated to be.

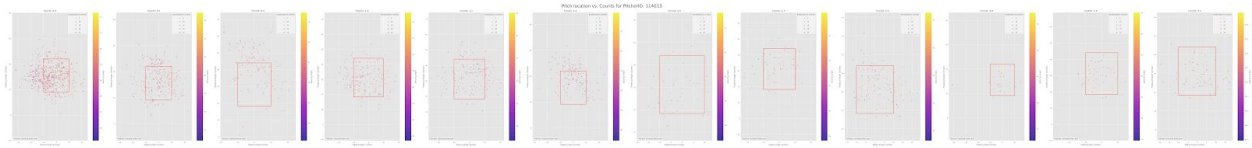
When plotting the pitch location vs. counts (Fig. 9, 10, 11), I found that the pitchers indeed had some different pitching patterns between the 0-2 and 3-0 counts. Under the 0-2 counts, the pitcher is leading, and intuitively, he could afford to aim at the edge to trick the batter and get a strikeout. Most pitchers demonstrated some level of command under the 0-2 counts, with pitcher 857's four-seamers (Fig. 10a) and 1594's two-seamers (Fig. 11d) having the best concentrated patterns near the edge. Under the 3-0 counts, the batter is leading, and thus the pitcher faces a high-risk situation and would usually want to throw a fastball at the center of the strike zone (Teeter 2015). Except pitcher 857's four-seamers (Fig. 10a), the rest pitchers all seemed to have aimed near the middle of the strike zone, with varying degrees of accuracy. Pitcher 114013's four-seamers seem to be the best under the 3-0 counts because the pitch location is most likely inside the strike zone (Fig. 10b), indicating a good level of command. The fact that pitcher 857's four-seamers concentrated near the left edge of the strike zone even under the 3-0 counts is probably also indicating good command, together with his confidence in his own command (Fig. 10a).

I also noticed that overall, under all the count situations, pitcher 857 had the least diagonal spread compared to the other pitchers (Fig. 10, 11), so he seems to have both good control and command.

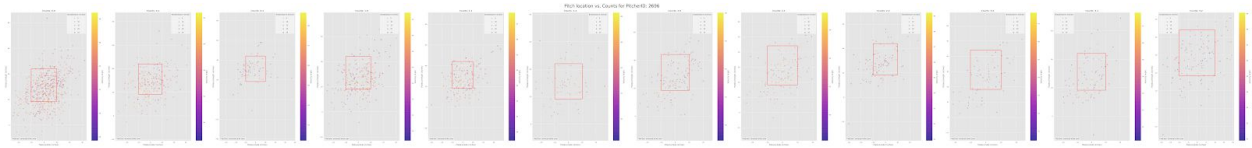
(a)



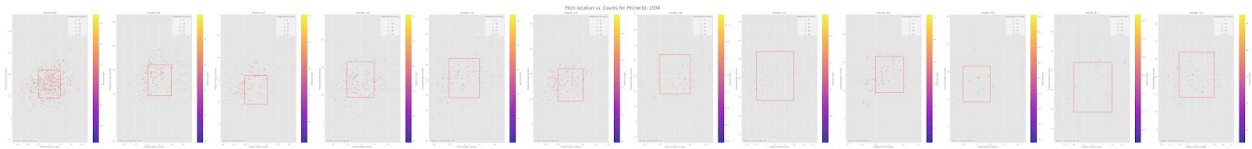
(b)



(c)



(d)



(e)

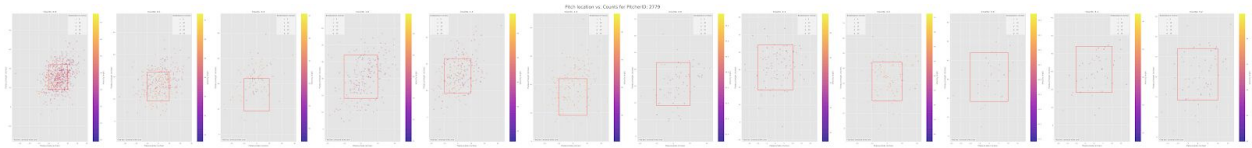
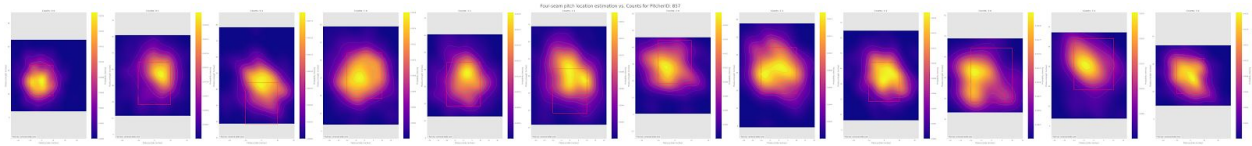
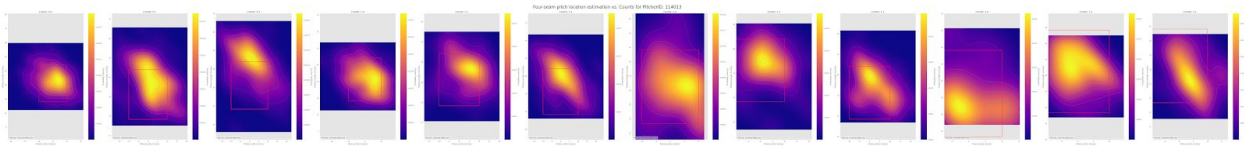


Figure 9. The five different pitchers' pitch locations shown as scatter plots in relation to the strike zone, differentiated by counts (balls-strikes). Each circle is a pitch location, with colour indicating velocity and size representing break distance. BreakDistance is the total distance of the horizontal and vertical break combined (using the simple Pythagorean formula).

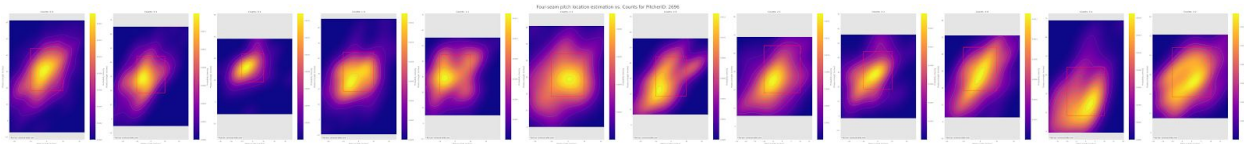
(a)



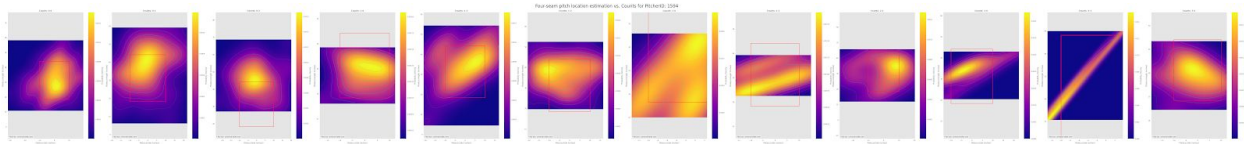
(b)



(c)



(d)



(e)

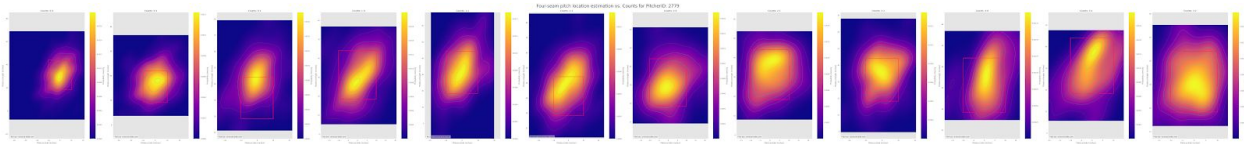
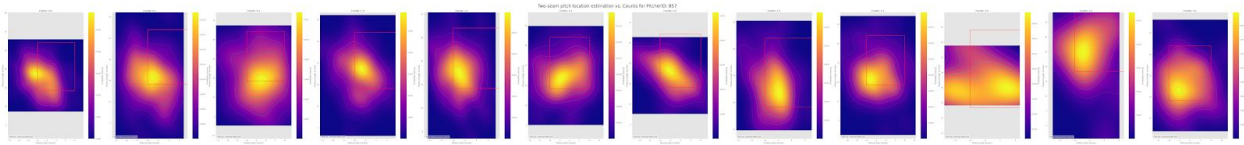
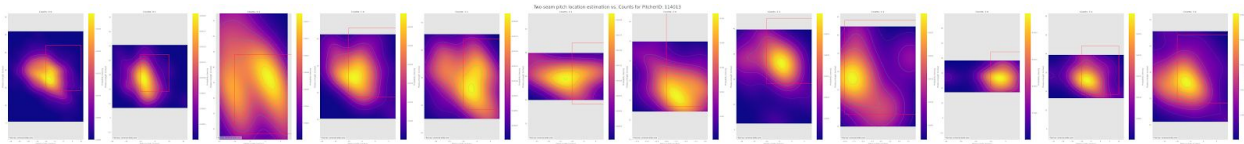


Figure 10. The five different pitchers' four-seam pitch locations shown as probability density in relation to the strike zone, differentiated by counts (balls-strikes). The probability distribution is modelled using Gaussian KDE. The faster the colour gradient changes, the more concentrated the pitch scatter is estimated to be.

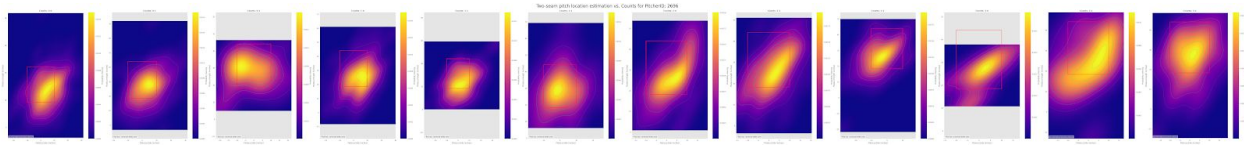
(a)



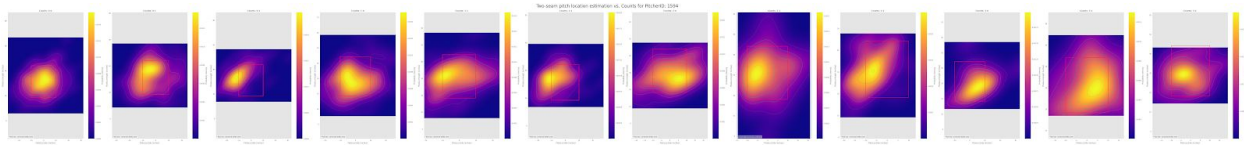
(b)



(c)



(d)



(e)

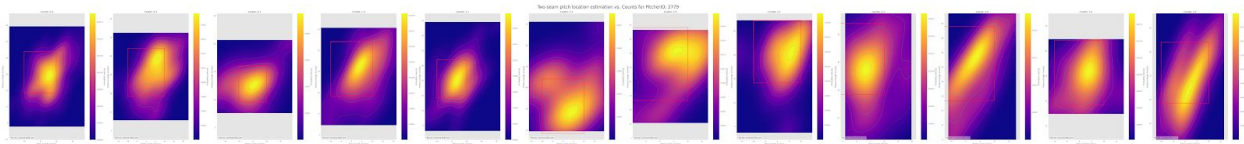


Figure 11. The five different pitchers' two-seam pitch locations shown as probability density in relation to the strike zone, differentiated by counts (balls-strikes). The probability distribution is modelled using Gaussian KDE. The faster the colour gradient changes, the more concentrated the pitch scatter is estimated to be.

Question 1.2: How to create a fastball command metric that could be applied to any pitcher at any level?

Solution

Like the existing Edge Percentage statistic (BaseballCloudBlog 2020), I wanted to measure how close the pitches are to the strike zone's edge. But unlike Edge Percentage, which counts how many balls fall within a small margin along the edge, I would like to calculate something like the average distance from the pitches to the edge as a measurement of fastball command. There are several benefits of calculating distance to the edge compared to simply counting balls near the edge:

1) When the sample size is small, the count-based Edge Percentage would be discrete, but a distance-based metric would still be continuous.
2) The closeness of the baseball to the strike zone's edge is not a dichotomous problem in nature. When the pitch location is outside the strike zone, the probability of getting a swing response from the batter must be a continuously decreasing function with the baseball's distance to the strike zone's edge. The Edge Percentage's way of counting balls on the edge is oversimplifying the problem and may lose important information about command.
3) A distance-based metric can be also used to measure the command of a single pitch (whereas Edge Percentage can't).

Answer

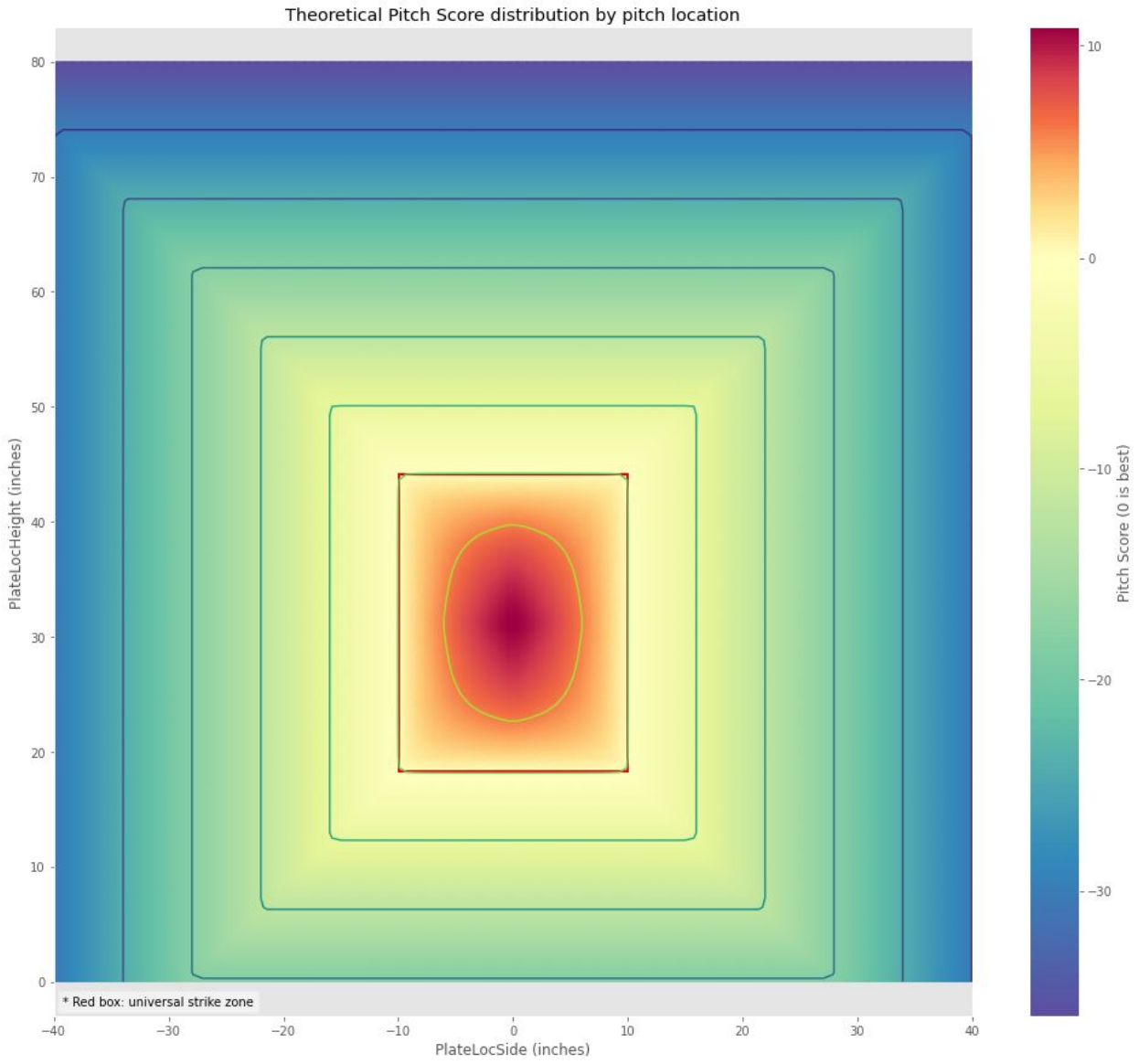
My metric of evaluating a pitcher's fastball command is calculated in the following way:

1. I created a custom Pitch Score (Fig. 12) that measures how close each pitch is to the strike zone's edges:
    (1) *Pitch Score (when the pitch is inside the zone) = Hmean(distance_vertical, distance_horizontal)*
    (2) *Pitch Score (when the pitch is outside the zone) = - distance from the pitch location to the nearest edge of the strike zone*

, where:
Hmean = harmonic mean;
distance_vertical = the distance from the baseball's vertical pitch location to the nearest edge of the strike zone (i.e. the top or bottom edge);
distance_horizontal = the distance from the baseball's horizontal pitch location to the nearest edge of the strike zone (i.e. the left or right edge);

The best score of this function would be 0, meaning the pitch location is right on the edge of the strike zone; when the pitch is inside the zone, the score is positive, and when the pitch is outside the zone, the score is negative.

(a)


Theoretical Pitch Score distribution by pitch location
PlateLocHeight (inches)
PlateLocSide (inches)
Pitch Score (0 is best)
* Red box: universal strike zone
80
70
60
50
40
30
20
10
0
−40
−30
−20
−10
0
10
20
30
40
10
0
−10
−20
−30

(b)

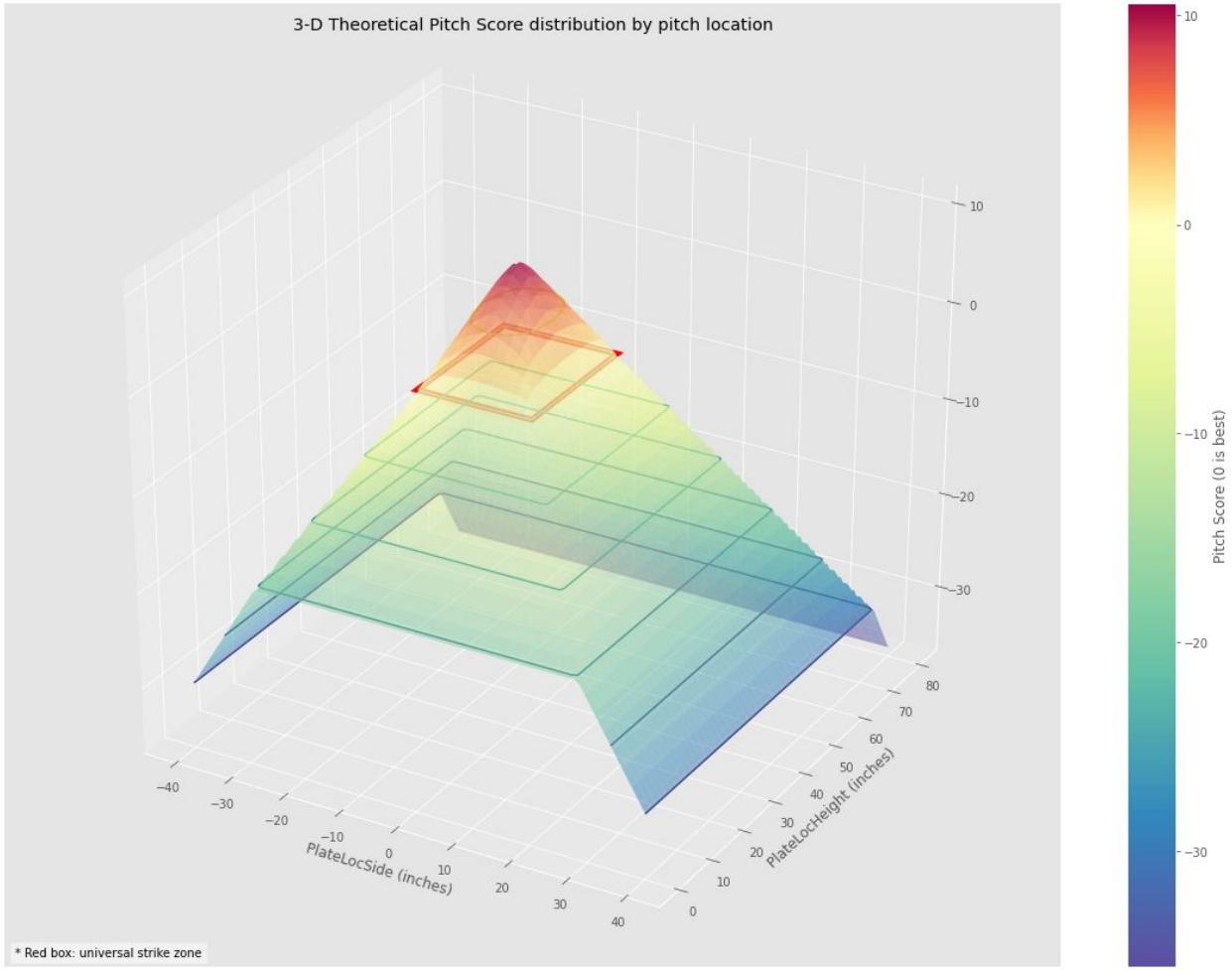


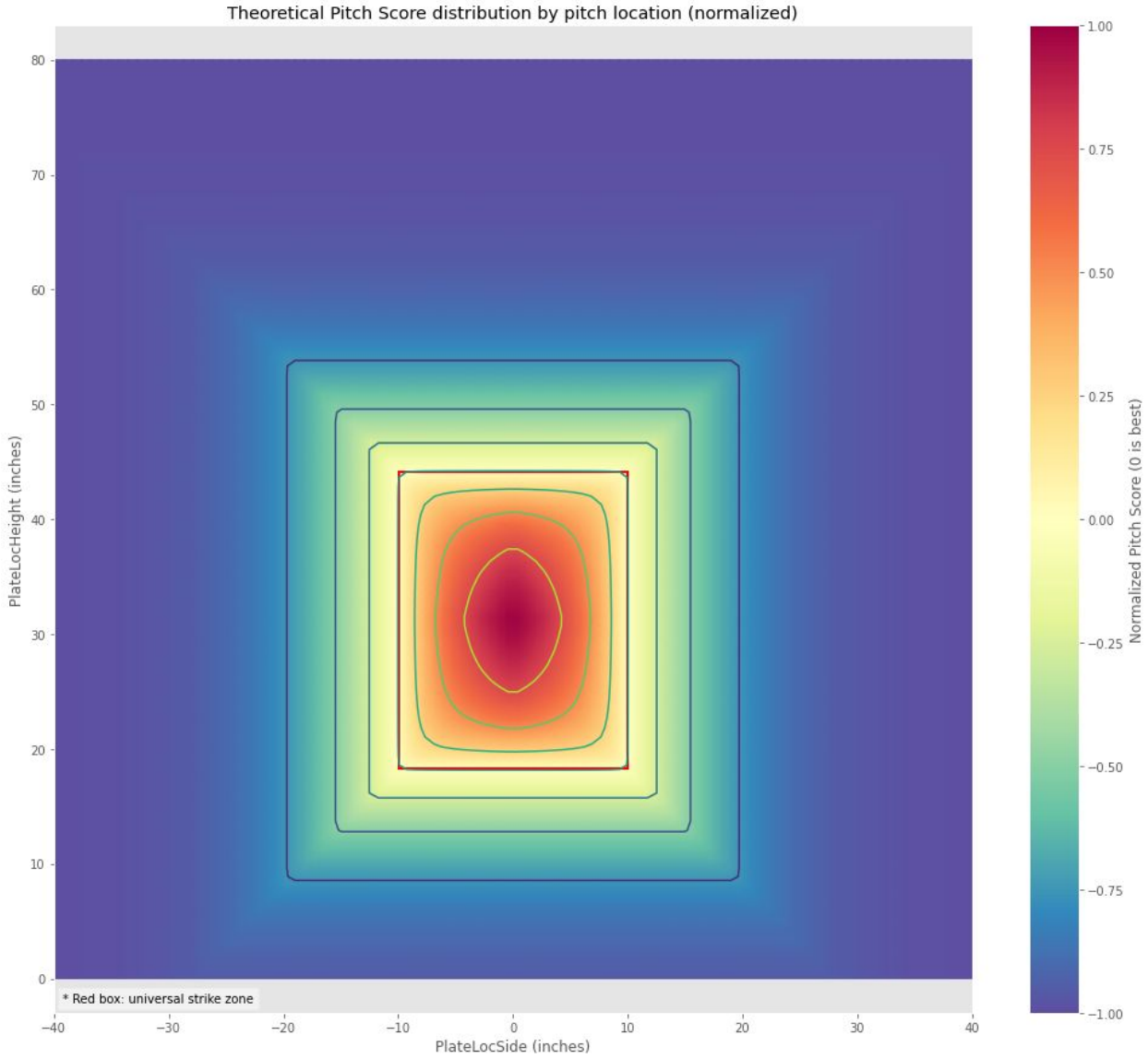
Figure 12. Visualization of the custom Pitch Score based on pitch location. 3-D animation: [3-D Theoretical Pitch Score distribution by pitch location.mp4]

2. I created a Normalized Pitch Score (Fig. 13) by normalizing the Pitch Score to a range between -1 and 1 using mathematical transformations (for technical details, see the attached source code). The best Normalized Pitch Score is still 0 when the pitch location is right on the edge of the strike zone, and when the ball is at the center; when the pitch location is at the center of the strike zone, the Normalized Pitch Score is 1, and when the pitch location is outside the strike zone at an infinite distance, the Normalized Pitch Score is -1. The Normalized Pitch Score is continuously differentiable at the edge of the strike zone, with a slope of -1 (away from the strike zone).

With this Normalized Pitch Score, we can already easily judge how close each pitch is to the edge of the strike zone. For example, if a pitch has a positive Normalized Pitch Score close to 1, then we know it is near the center of the strike zone (inside). If a pitch has a negative Normalized Pitch Score close to -1, then we know it is away from the strike zone (outside). If a

pitch has a Normalized Pitch Score close to 0, then we know it is very close to the edge, and depending on the sign of the score, we can tell whether it's inside or outside the strike zone.

(a)



(b)

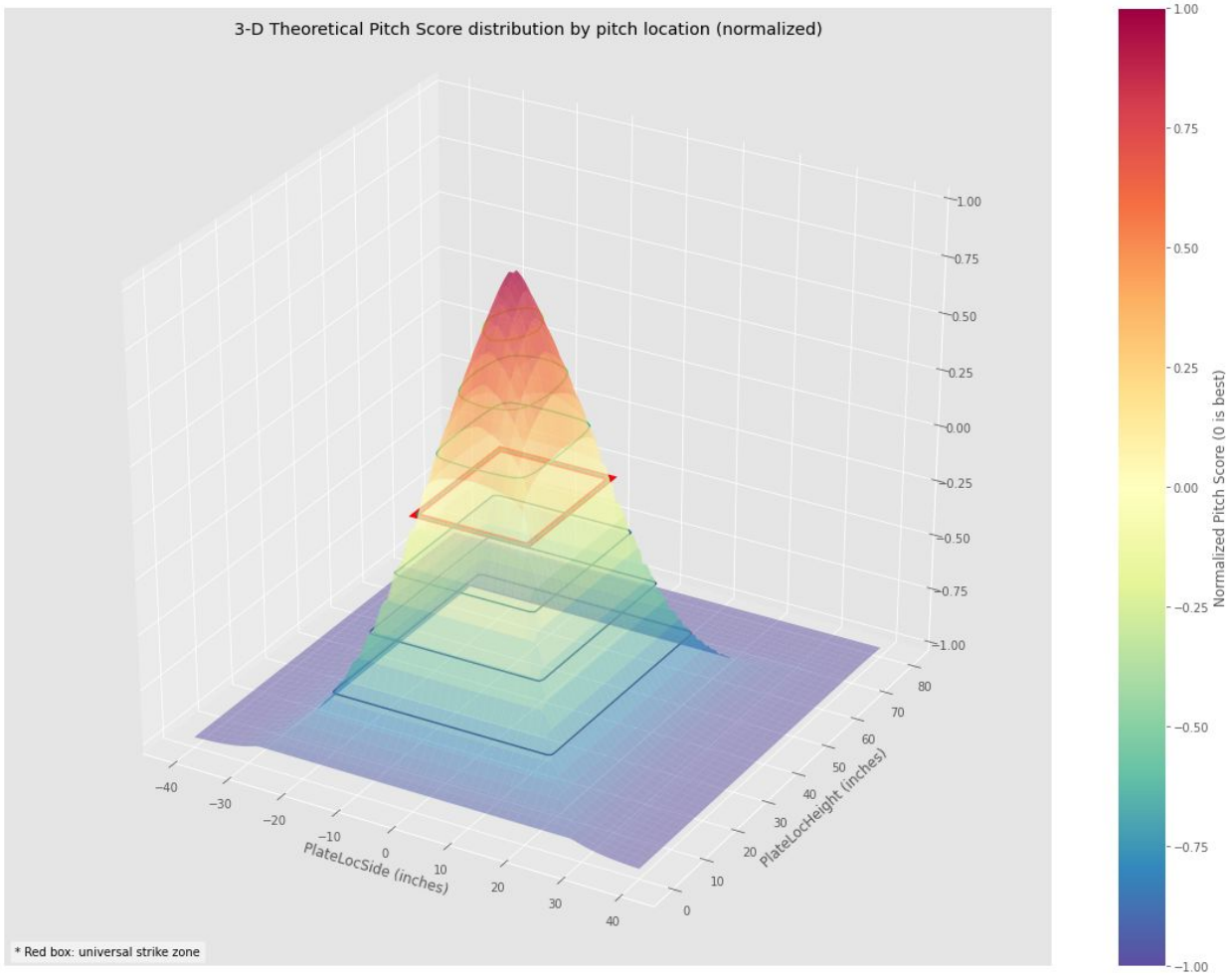


Figure 13. Visualization of the Normalized Pitch Score based on pitch location. 3-D animation: [3-D Theoretical Pitch Score distribution by pitch location (normalized).mp4]

3. Finally, based on the Normalized Pitch Score for all the pitches, I created a Pitch Command Score for the pitchers:

*Pitch Command Score = Sign*MSE(Normalized Pitch Command Score)*

, where:

Sign = 1 if the median of a pitcher's Normalized Pitch Command Score is >= 0, and -1 when it's not;

MSE = Mean Squared Error (i.e. the square sum of the Normalized Pitch Command Score, divided by the number of observations)

My Pitch Command Score metric has the following properties:

1) The Pitch Command Score has a range between -1 and 1.
2) If the pitcher has very good command (by throwing fastballs consistently at the edge), his Pitch Command Score would be close to 0.

3) If the majority of the pitcher's fastballs ended up inside the strike zone, his Pitch Command Score would be positive; the more concentrated his balls are towards the center of the strike zone, the closer his Pitch Command Score would be to 1.
4) If the majority of the pitcher's fastballs ended up outside the strike zone, his Pitch Command Score would be negative; the more scattered his balls are outside the strike zone, the closer his Pitch Command Score would be to -1.

With this Pitch Command Score metric, we can not only select pitchers based on their command performance, but also do so while accounting for the style of the pitcher (i.e. favouring a positive number for a conservative pitch style with the fastballs mostly inside the strike zone, or a negative number for a risky pitch style with the fastballs mostly outside the strike zone.

Based on my Pitch Command Score metric, pitcher 857 has the best fastball command because his score is closest to 0 (Table 1, Figure 14).

Table 1. Pitch Command Score for the five pitchers. Except pitcher 2696 who placed most of his fastballs outside the strike zone, the rest pitchers all placed most of their fastballs inside the strike zone. Pitcher 857 has the score closest to 0, meaning the best command. Pitcher 114013 has the second best command because his score is the second closest to 0.

	pitch_score_normalized	
	player_score	count
PitcherID		
857	0.271434	4525
1594	0.284537	1527
2696	-0.306853	2243
2779	0.305490	2233
114013	0.280018	1946

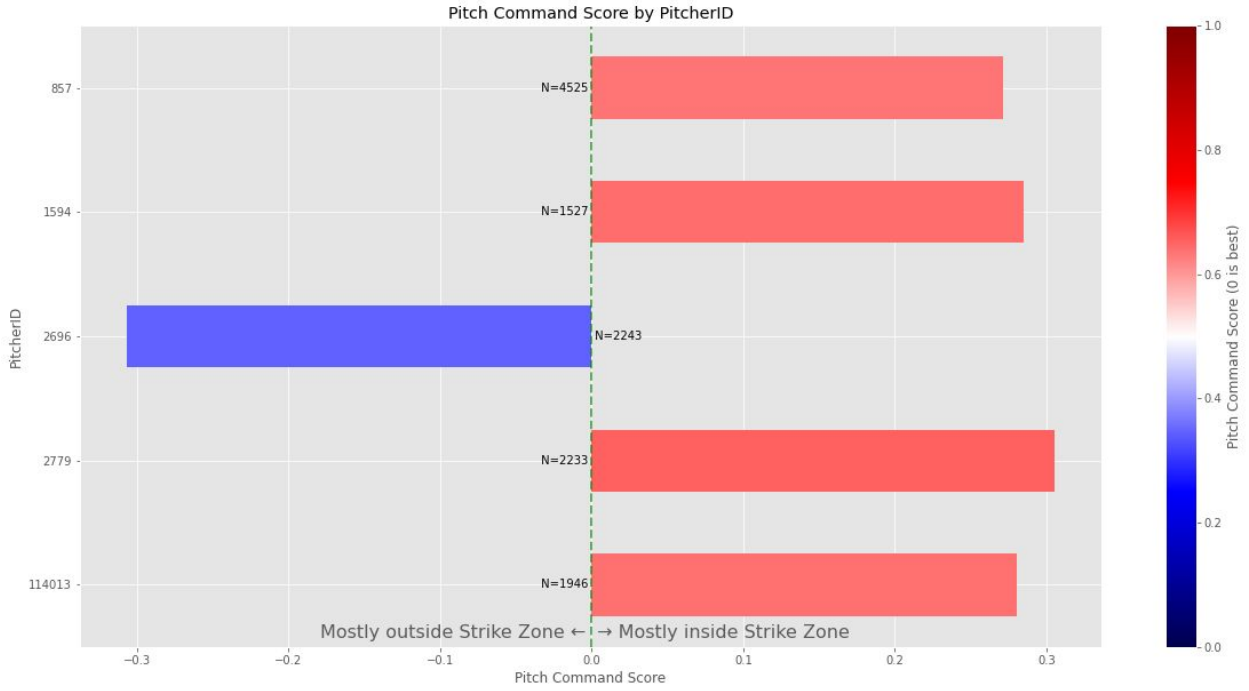


Figure 14. Pitch Command Score of the five pitchers visualized.

In-depth Analysis:

The same metric can be applied to evaluate a pitcher's command under specific situations, such as certain pitch type, bat side, count, or year (Fig. 15-17).

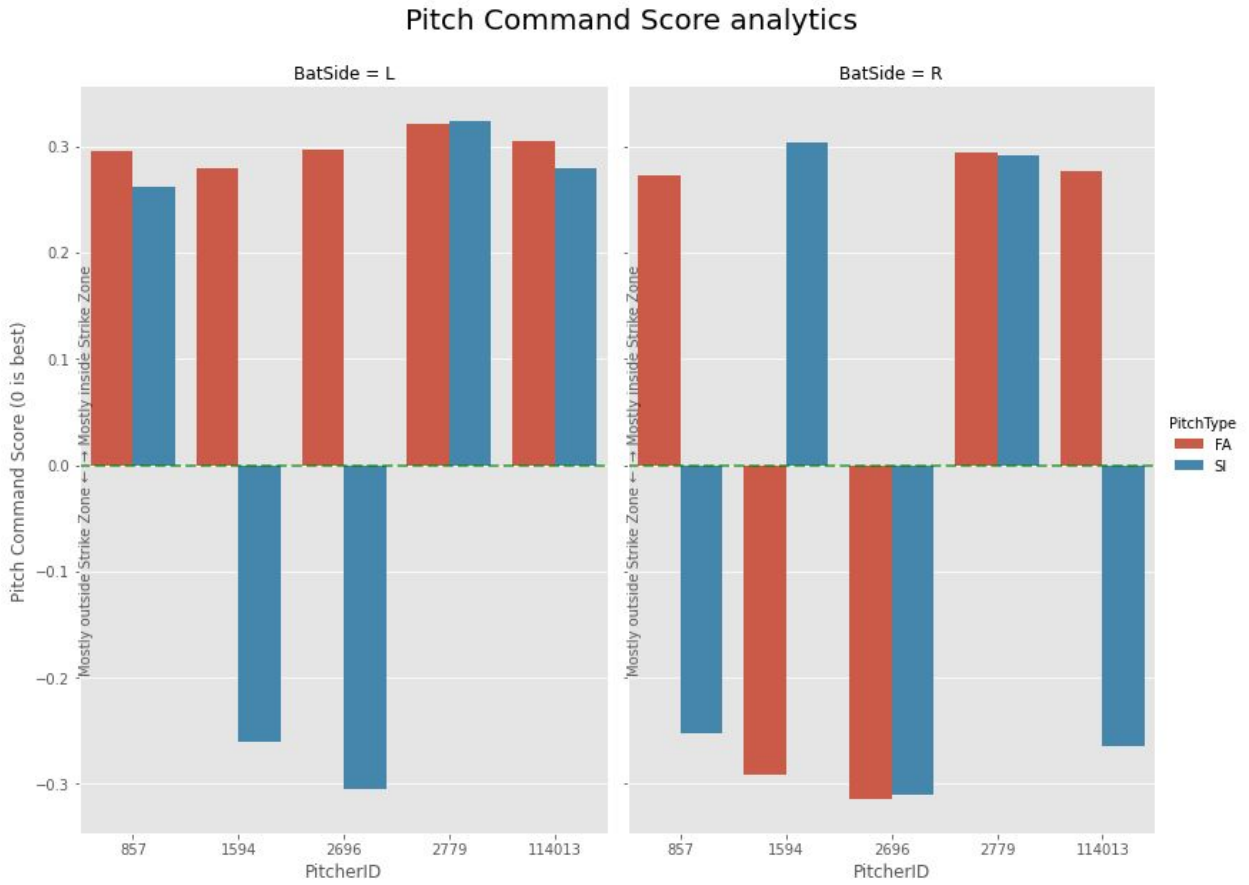


Figure 15. Pitch Command Score by bat side and pitch type for different pitchers.

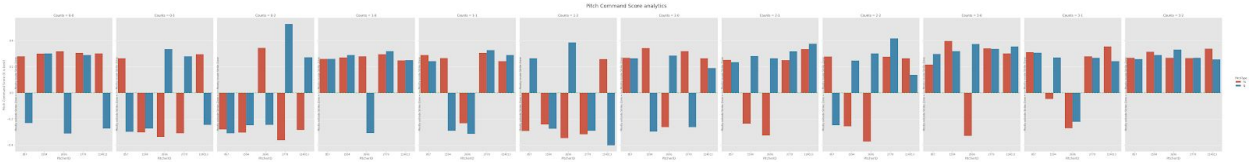


Figure 16. Pitch Command Score by counts and pitch type for different pitchers.

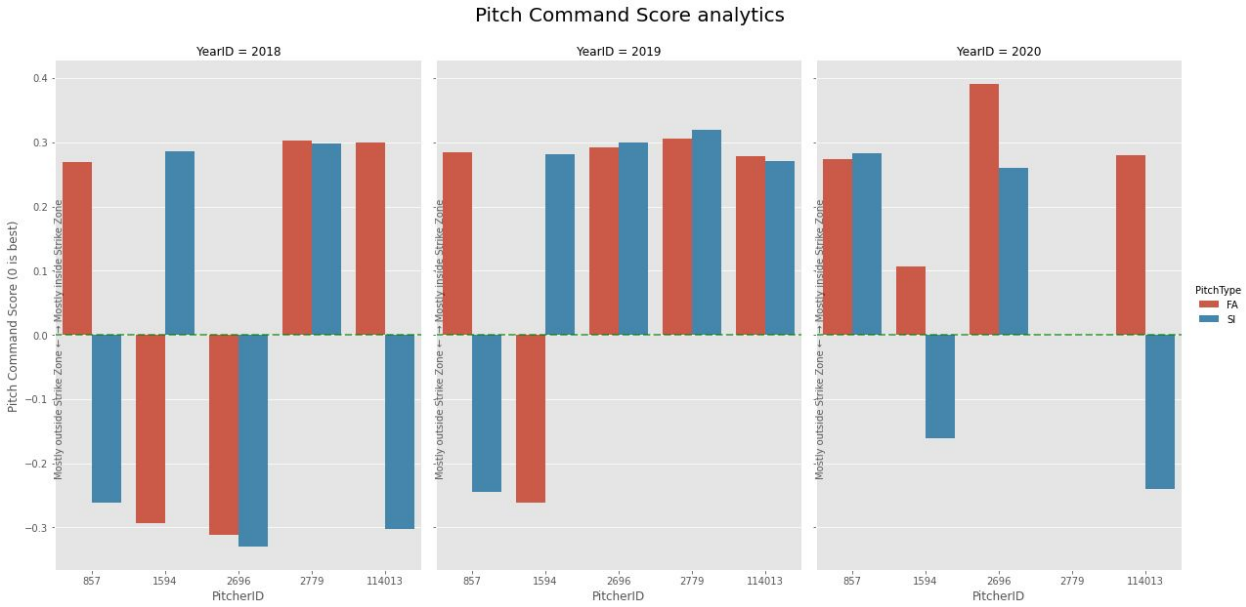
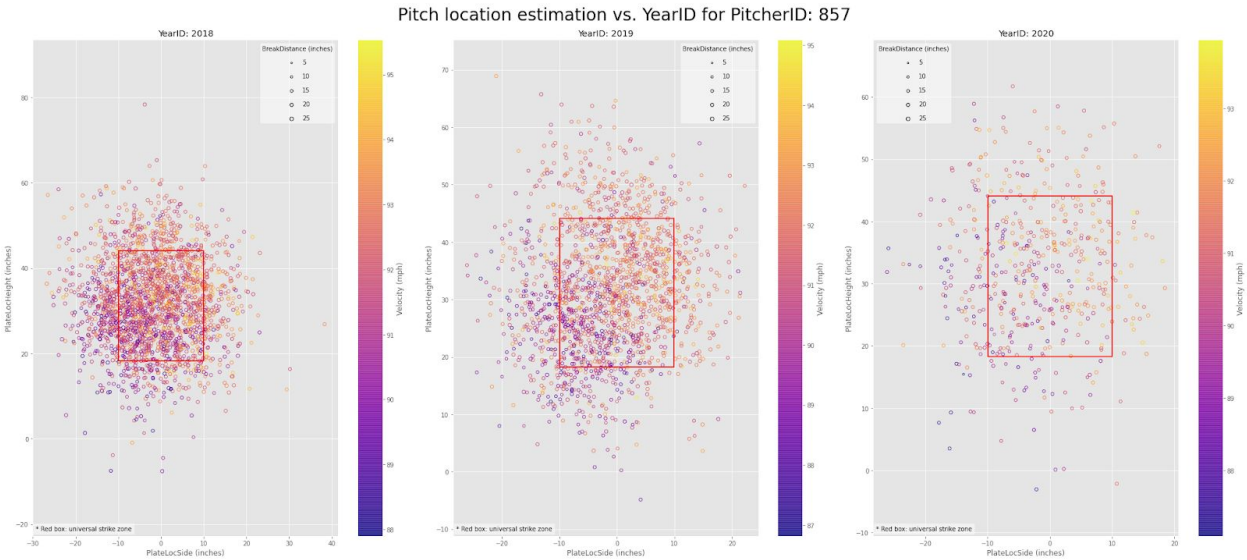


Figure 17. Pitch Command Score by year and pitch type for different pitchers.

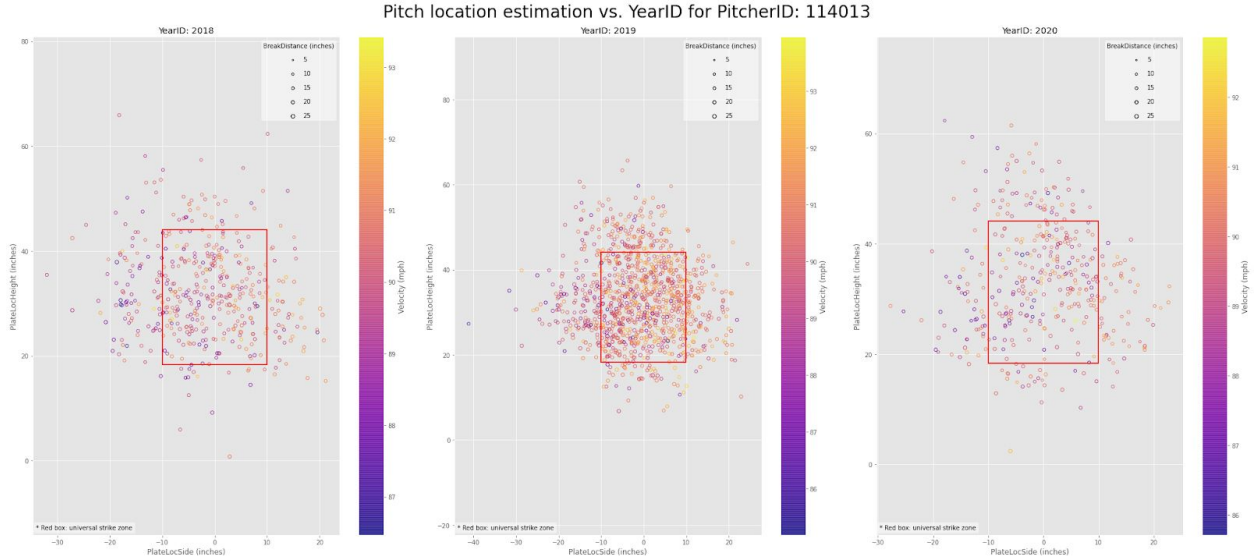
One interesting observation:

If we only look at year 2020, pitcher 1594 seems to have the best Pitch Command Score (Fig. 17). This is because he had only very few pitches that year and most of the pitches ended up close to the strike zone edge (Fig. 18d). So the pitch command score is indeed able to reflect how good the pitcher is at putting the ball close to the strike zone edge.

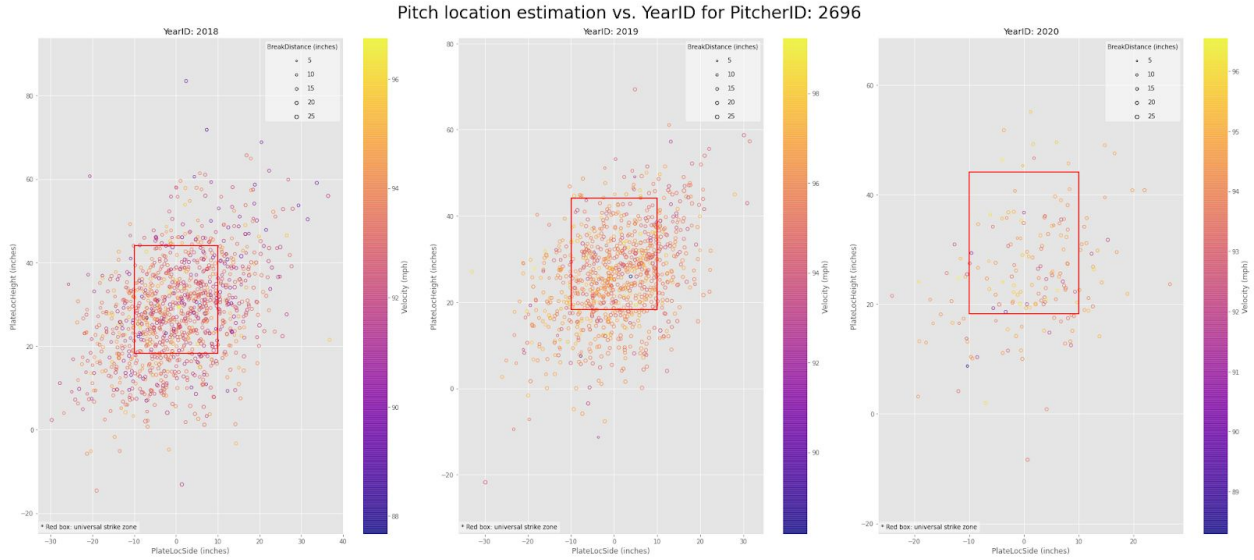
(a)



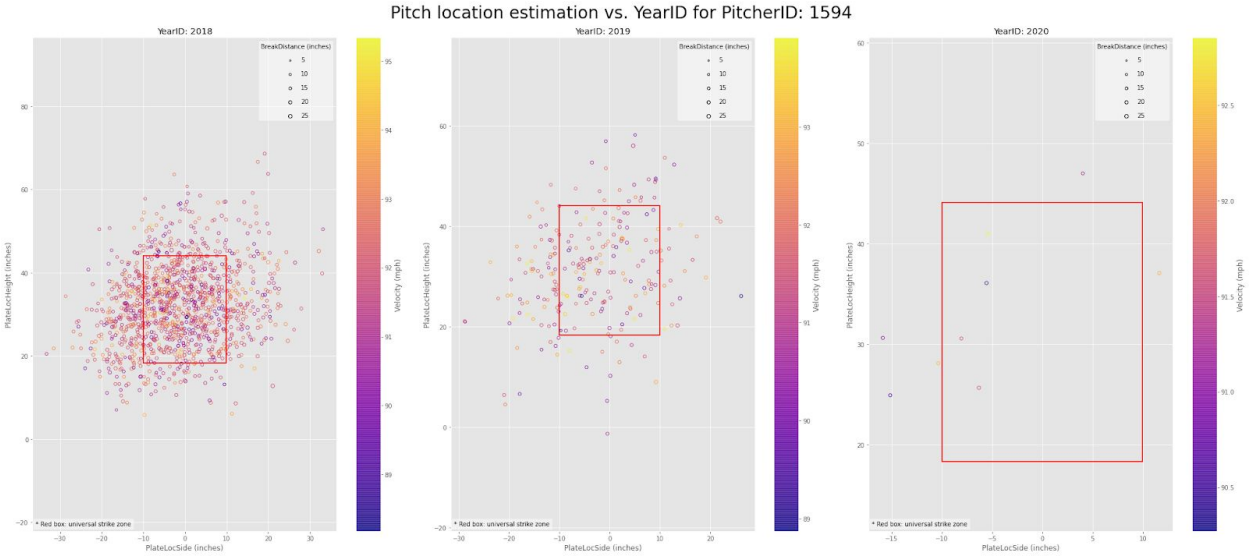
(b)


Pitch location estimation vs. YearID for PitcherID: 114013
YearID: 2018
YearID: 2019
YearID: 2020
BreakDistance (inches)
5
10
15
20
25
PlateLocHeight (inches)
PlateLocSide (inches)
Velocity (mph)
* Red box: universal strike zone

(c)


Pitch location estimation vs. YearID for PitcherID: 2696
YearID: 2018
YearID: 2019
YearID: 2020
BreakDistance (inches)
5
10
15
20
25
PlateLocHeight (inches)
PlateLocSide (inches)
Velocity (mph)
* Red box: universal strike zone

(d)


Pitch location estimation vs. YearID for PitcherID: 1594
YearID: 2018
YearID: 2019
YearID: 2020
BreakDistance (inches)
5
10
15
20
25
PlateLocHeight (inches)
PlateLocSide (inches)
Velocity (mph)
* Red box: universal strike zone

(e)

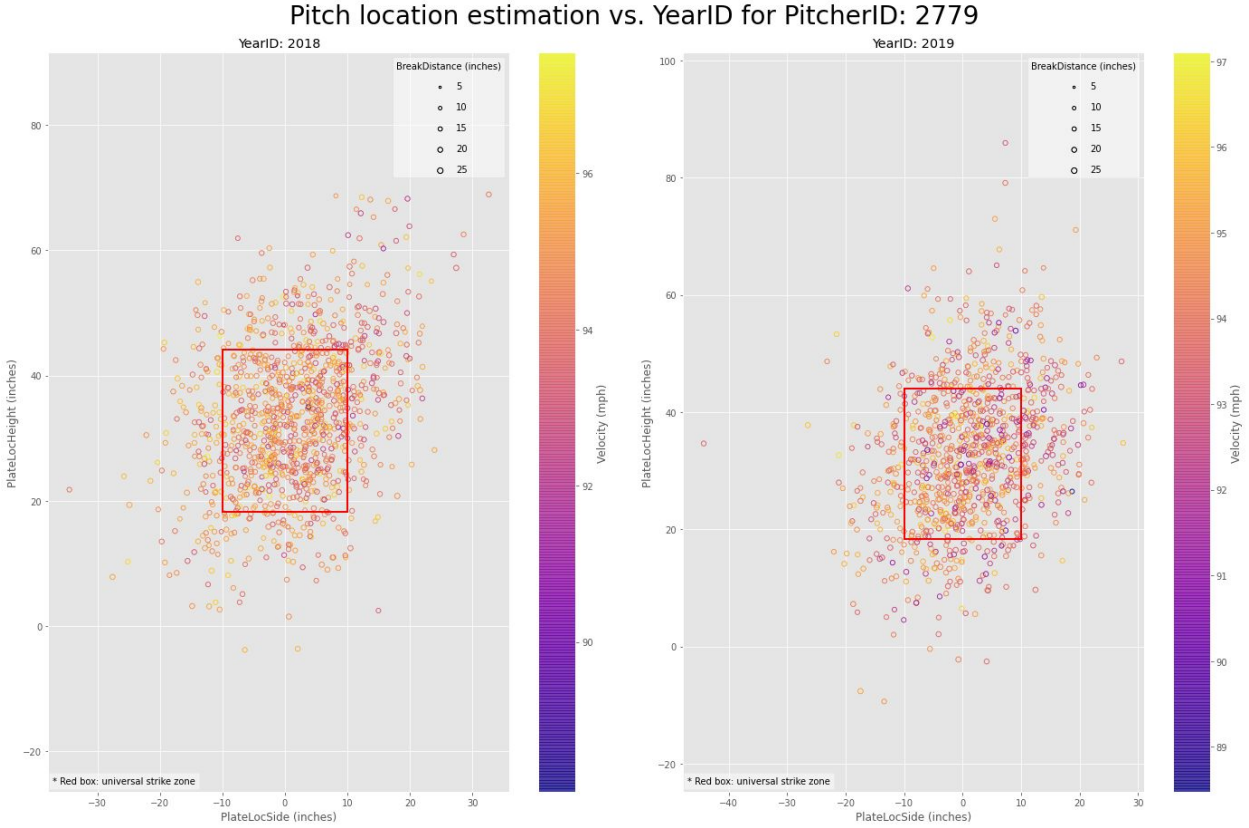

Pitch location estimation vs. YearID for PitcherID: 2779
YearID: 2018
YearID: 2019
BreakDistance (inches)
5
10
15
20
25
PlateLocHeight (inches)
PlateLocSide (inches)
Velocity (mph)
* Red box: universal strike zone

Figure 18. The five different pitchers' pitch locations shown as scatter plots in relation to the strike zone, differentiated by year. Each circle is a pitch location, with colour indicating velocity and size representing break distance. BreakDistance is the total distance of the horizontal and vertical break combined (using the simple Pythagorean formula).

Limitations of the Study

1. The analysis assumes that the pitcher wants to aim at the edge of the strike zone.
2. The analysis assumes that the strike zone is a universal box shape.
3. The Command Pitch score is calculated purely based on the pitch location.

Future Directions

1. I noticed that some pitchers had a diagonally elongated pitch location scatter (e.g., Fig. 9, 10). PCA could be used to measure the anisotropy of the pitch location scatter, so that the more elongated the pitch scatter is (in a diagonal direction), the less control/command the pitcher has.
2. Data are needed for the batter's response to each pitch (was it a ball or a strike). If such data are available, we could model the actual strike zone better, or evaluate the true efficacy of the pitch strategy.
3. Before each pitch, we need to collect data from the pitcher himself to know his intended pitch location. This will allow the most straightforward pitch command analysis.

Source Code

https://colab.research.google.com/drive/19znFNsy7zJk2KRgba9WfisN95H6iDSyJ

References

BaseballCloudBlog. (2020). Quantifying Command – Part Two: Edge Percentage. https://baseballcloud.blog/2020/07/09/quantifying-command-part-two-edge-percentage/

Boyle, Wayne. (2018). Prospectus Feature: The Universal Strike Zone. https://www.baseballprospectus.com/news/article/40891/prospectus-feature-the-universal-strike-zone/

Coach Jon. (2018). The Count and How it Impacts Hitting Results. https://spiderselite.com/2018/03/25/count-hitting-results/

Melling, Micah. (2018). Game Theory Applications in Baseball. http://www.baseballdatascience.com/game-theory-applications-in-baseball/

Teeter, Chris. (2015). Swinging at 3-0 pitches: A high-risk decision. https://www.beyondtheboxscore.com/2015/3/5/8151763/baseball-swinging-count-pitches-balls-strikes-strikezone-sluggers