Untitled

David Scolari, Harrison Snell, Brandon Williams

5/8/2022

Introduction

Hitting a baseball has been shown to be one of the most difficult tasks in all of sports. Anybody who has gone to a batting cage and tried the fastest machine for fun knows how hard it can be to hit a normal fastball. That does not even include the possibilities of breaking balls and off-speed pitches. Batters have a fraction of a fraction of a second to recognize a pitch, determine if that pitch will be a ball or a strike, and set their swing into motion. Considering the human body can only move so fast to get the bat over the plate, the batter has to make their decision moments after the ball leaves the pitcher's hand. At that point, even some professional baseball players have to guess on whether they are facing a changeup or a fastball or a curveball. Any additional information the batter can have is crucial. The goal of our model is to provide some of that additional information.

Pitch predicting is a very important part of baseball. So important that teams will do almost anything for insight on what the opposing team's pitcher is going to throw. Other than illicit methods, hitters are trained to use information on a pitcher's previous pitch throwing behvior to guess which pitch is coming next. We can model this deciscion making process with machine learning techniques.

By modeling pitch prediction, we can learn what information is relevant to a pitcher's pitch selection. We can also better understand what makes a pitcher more unpredictable than other. It's gonna be good

In the following report will talk about why pitches are different, share evidence that pitchers have different strategies from at bat to at bat, and describe the engineering of our model, which is basically a trashcan.

Why do we care about a pitch?

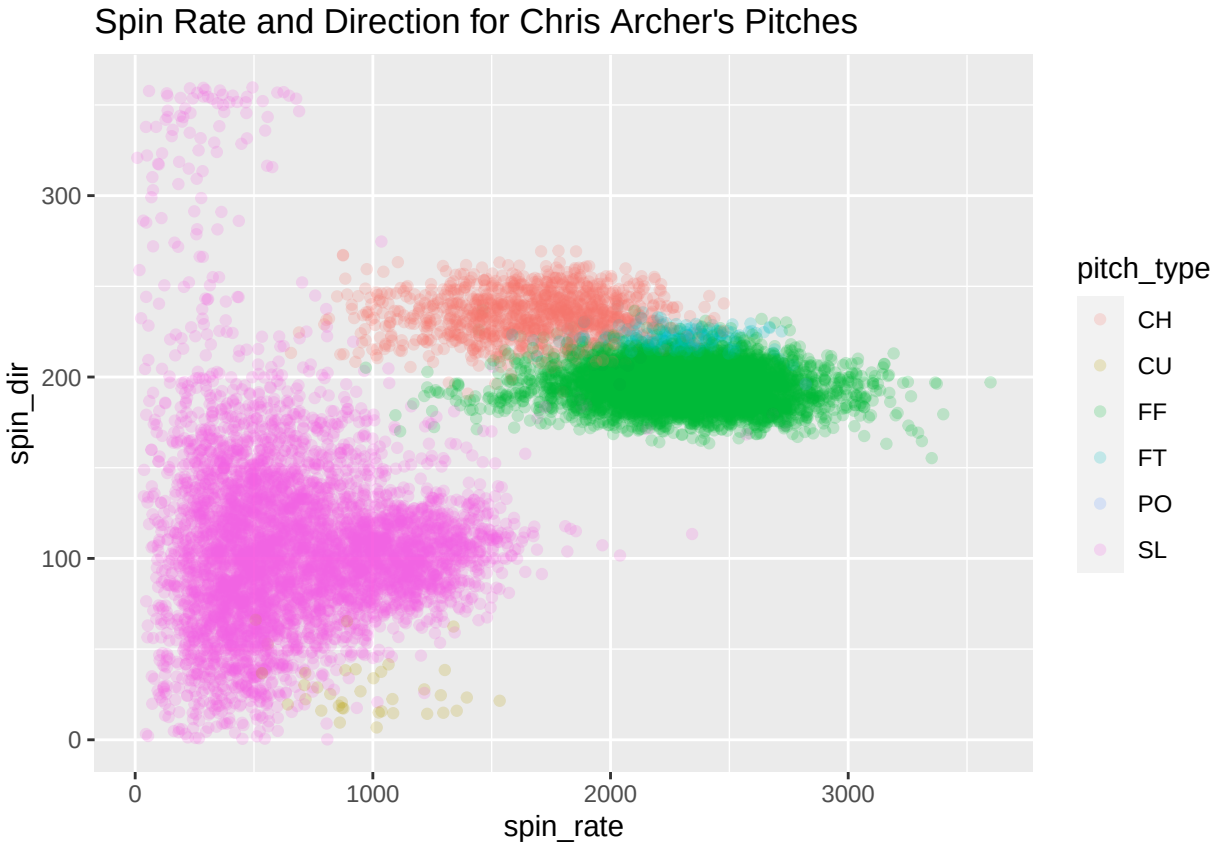
Not all pitches are made the same. From sliders to cutters to knuckleballs, each pitch tends to have its own characteristics that define it.

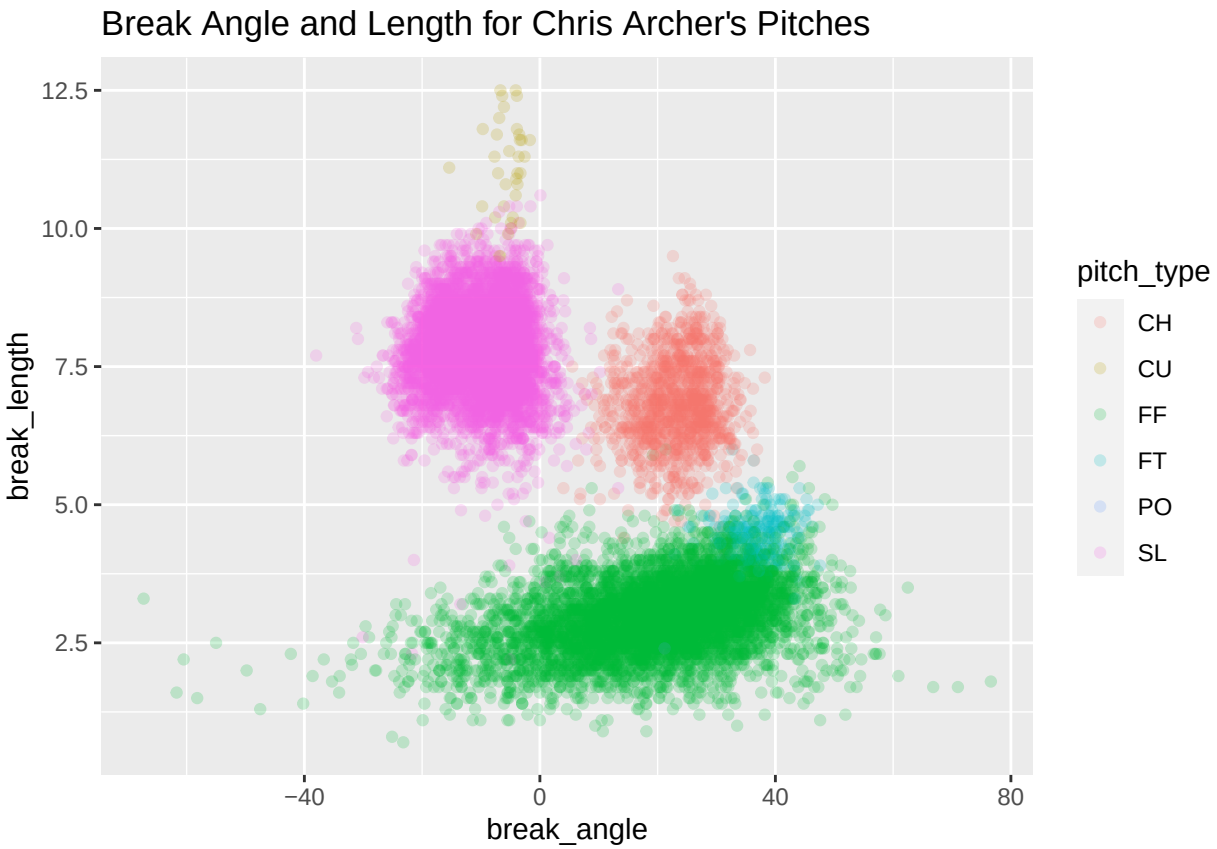
Table 1: Averages of Different Pitch Characteristics

pitch_type	px	pz	start_speed	end_speed	spin_rate	spin_dir	break_angle	break_length
CH	-0.200	1.859	85.814	79.193	1734.540	212.372	8.859	7.695
CU	0.076	1.802	78.164	72.294	1446.625	89.216	-9.949	12.569
EP	0.183	2.022	66.516	61.323	1381.736	44.096	-9.033	16.908
FA	-0.184	1.768	88.900	80.600	2448.751	148.710	-7.067	7.500
FC	0.221	2.310	88.543	81.992	1018.410	164.455	-5.357	6.093
FF	0.007	2.638	93.628	85.901	2175.728	193.631	13.491	3.721
FS	-0.377	1.916	84.568	78.603	1483.198	222.074	17.104	6.653
FT	-0.023	2.351	92.151	84.678	2129.669	169.489	-5.558	5.750
KC	0.084	1.710	81.507	75.232	1318.344	89.548	-6.516	11.551
PO	0.968	4.114	86.217	79.080	1928.408	210.204	21.037	5.171

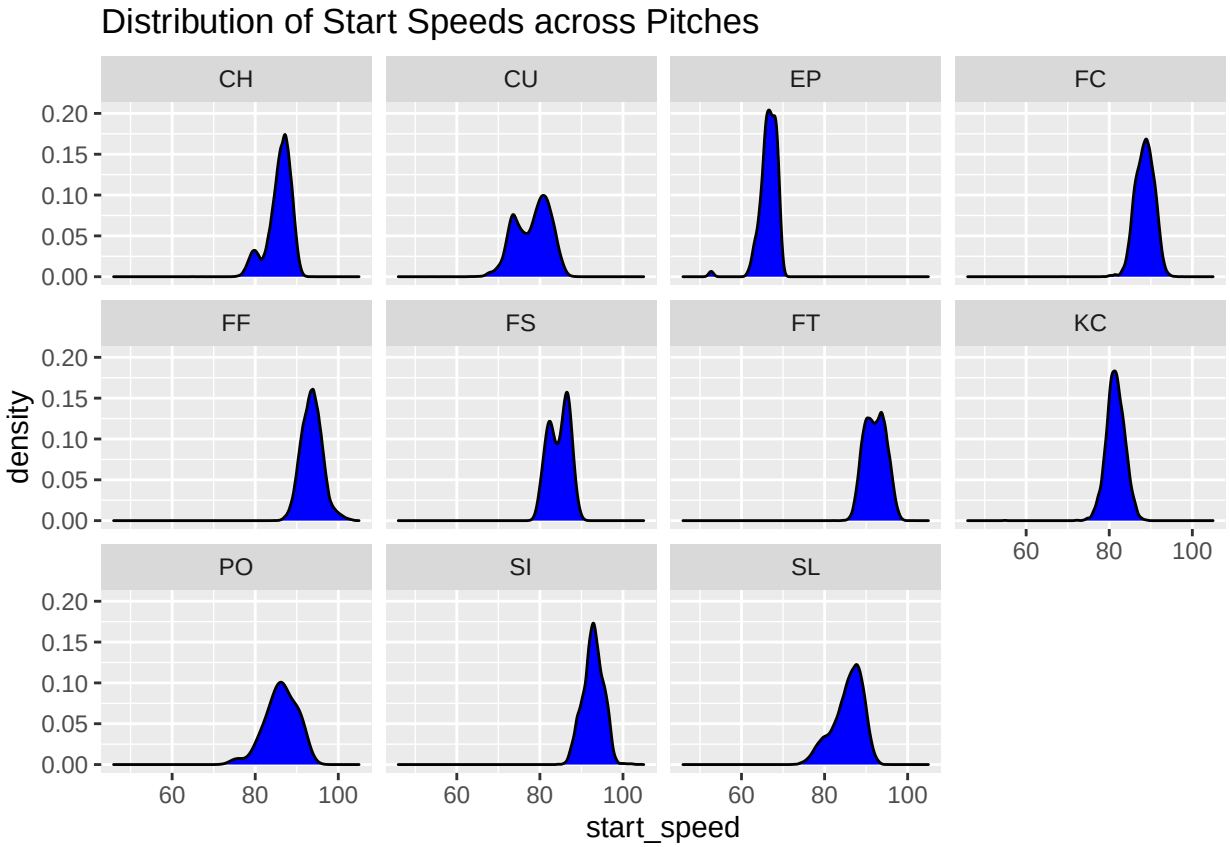
pitch_type	px	pz	start_speed	end_speed	spin_rate	spin_dir	break_angle	break_length
SI	-0.132	2.309	92.813	85.419	2008.583	216.362	21.268	5.880
SL	0.232	1.867	85.588	79.352	901.077	162.506	-3.271	8.079
UN	3.488	3.531	41.867	39.800	823.262	218.886	-1.167	31.800

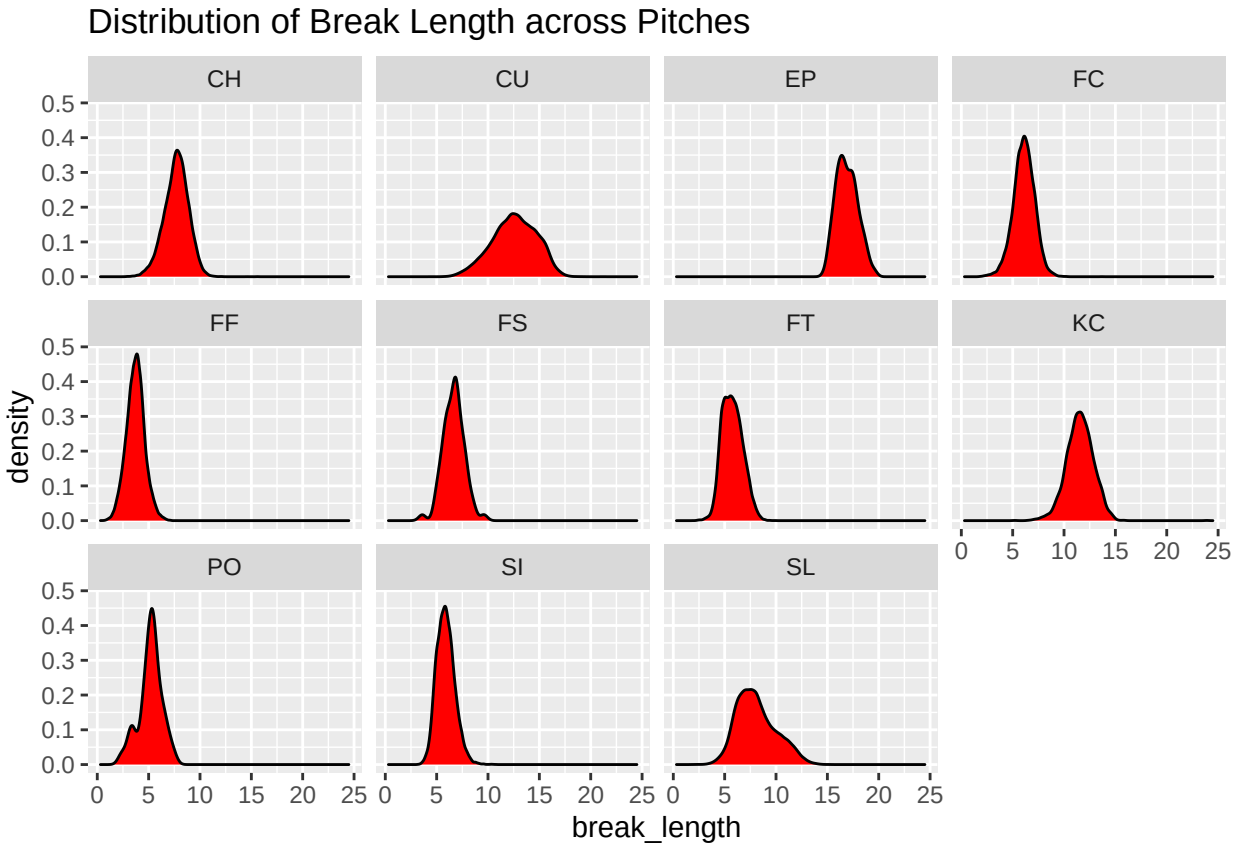
From the table, we essentially see the obvious that is pitches are different. On average we find that fastballs start at almost 94 mph and changeups start at 86 mph. The average break length of a fastball is about 4 inches, which is not too bad, the point that a batter sees a fastball leave the pitcher's hand will most likely cross following a straight line. Compare that to a curveball with almost 13 inches of break and we start to see why knowing the next pitch can be helpful.





The plots of Chris Archer's pitches give us a glimpse of where one type of pitch becomes another. For Archer, we see that his slider tends to have a relatively low spin rate, with his fastballs having higher spin rates and his changeups in between. As we would expect, Archer's fastballs have little break. His slider however, has a large but variable amount of break and always tends to break in the same direction. These visuals help us understand where a pitch may be going. If we can predict the next pitch is a slider, then we can start to understand the direction and magnitude of break in the pitch.


Distribution of Start Speeds across Pitches
CH
CU
EP
FC
FF
FS
FT
KC
PO
SI
SL
density
0.20
0.15
0.10
0.05
0.00
60
80
100
start_speed



Mostly as an exposition into our data, we found the graphs of distributions of break length and start speed to be interesting. They provide a more throrough comparison of pitch types because every pitcher is slightly different. Some pitchers may have a lot more movement on their sinker than others. Similarly, Aroldis Chapman throwing 105 mph fastballs is certainly not the norm in the league. Getting to see the variability of break length and speeds begins to paint the picture of why some pitchers are better than others.