

BOSTON UNIVERSITY/KEIO UNIVERSITY
WORKSHOP 2016

The Evaluation of Catcher “Framing”
using PITCHf/x data

Daiki Nagata
Keio University

Mihoko Minami and DataStadium Inc.

What is “Framing”?

Framing is the technique that a catcher uses when receiving a pitch.

wikipedia

『The catcher can help his pitcher get more strike calls from the umpire by using a technique called “framing” 』

In other words, framing is the ability to receive a pitch such that the probability of called strike is increased.

Let's look at "Framing"



<http://grantland.com/features/studying-art-pitch-framing-catchers-such-francisco-cervelli-chris-stewart-jose-molina-others/>

Let's look at "Framing"



- **Ryan Doumit**
- Judge : **Ball**

<http://grantland.com/features/studying-art-pitch-framing-catchers-such-francisco-cervelli-chris-stewart-jose-molina-others/>

PITCHf/x

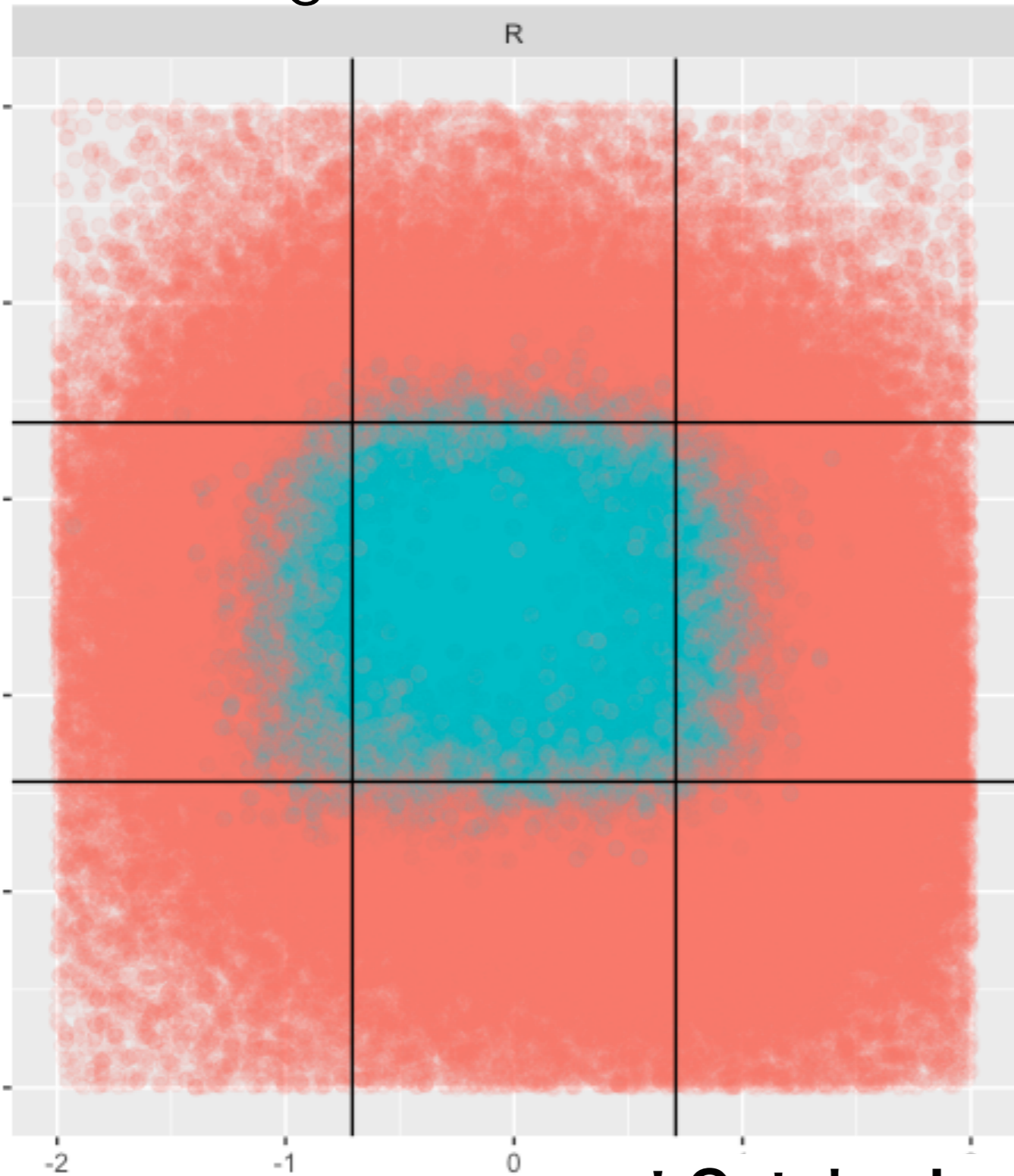
- The PITCHf/x[®] service tracks the full trajectory of live baseball pitches, and gets an information of speed or location.
- It has been installed in all 30 stadiums, and currently tracks pitches for every MLB game.



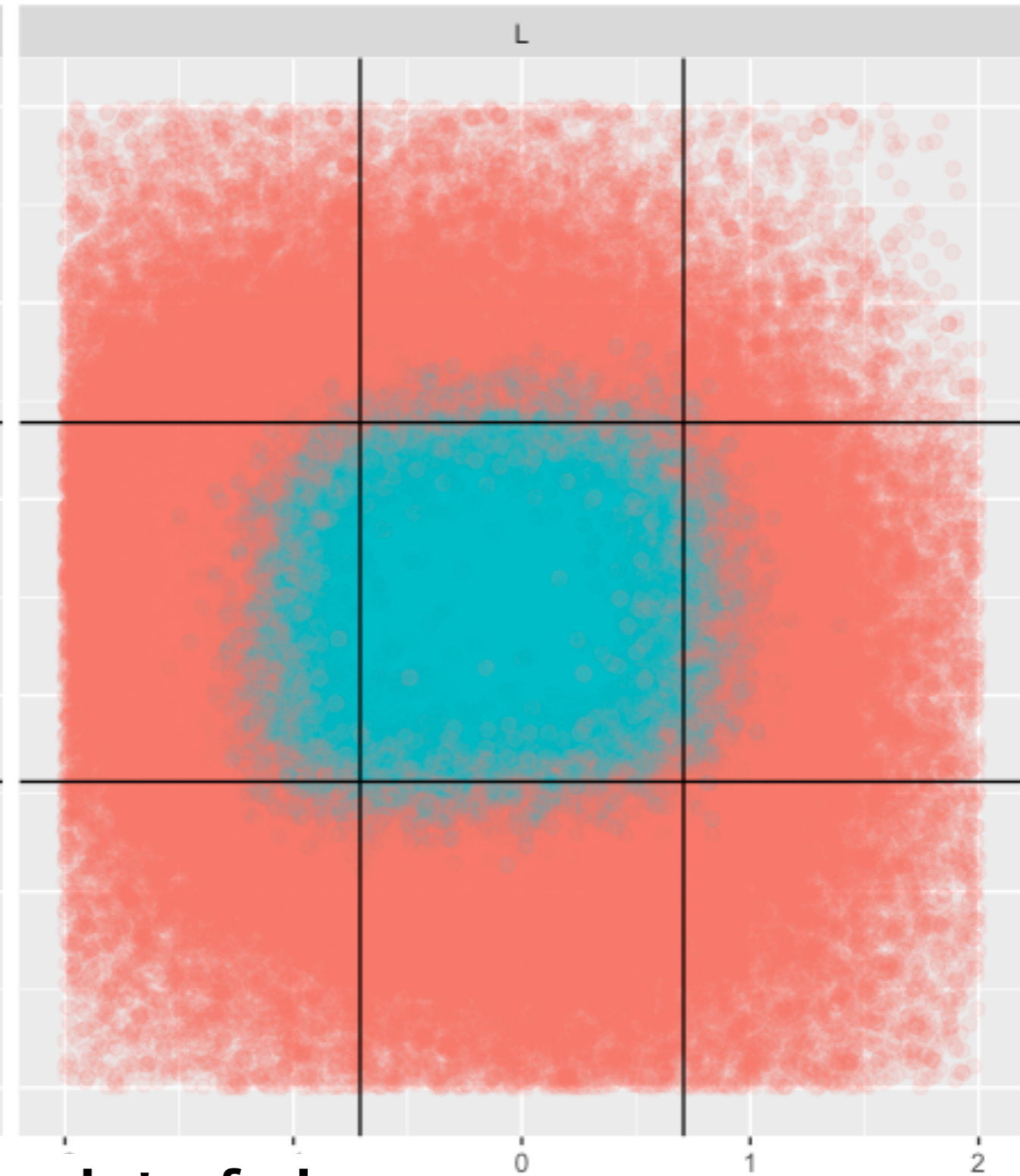
<http://www.sportvision.combaseball/pitchfx>[®]

Strike or Ball

Right-handed batter



Left-handed batter



*Catcher's point of view

Generalized Additive Model (GAM)

$$Y_i = \begin{cases} 1 & \text{(Called Strike)} \\ 0 & \text{(Ball)} \end{cases}$$

Logistic Regression Model

$$Y_i \sim \text{Bernoulli}(p_i)$$

$$\log \frac{p_i}{1 - p_i} = \alpha + \underbrace{f(\text{plate}.x_i, \text{plate}.z_i)}$$

smooth term

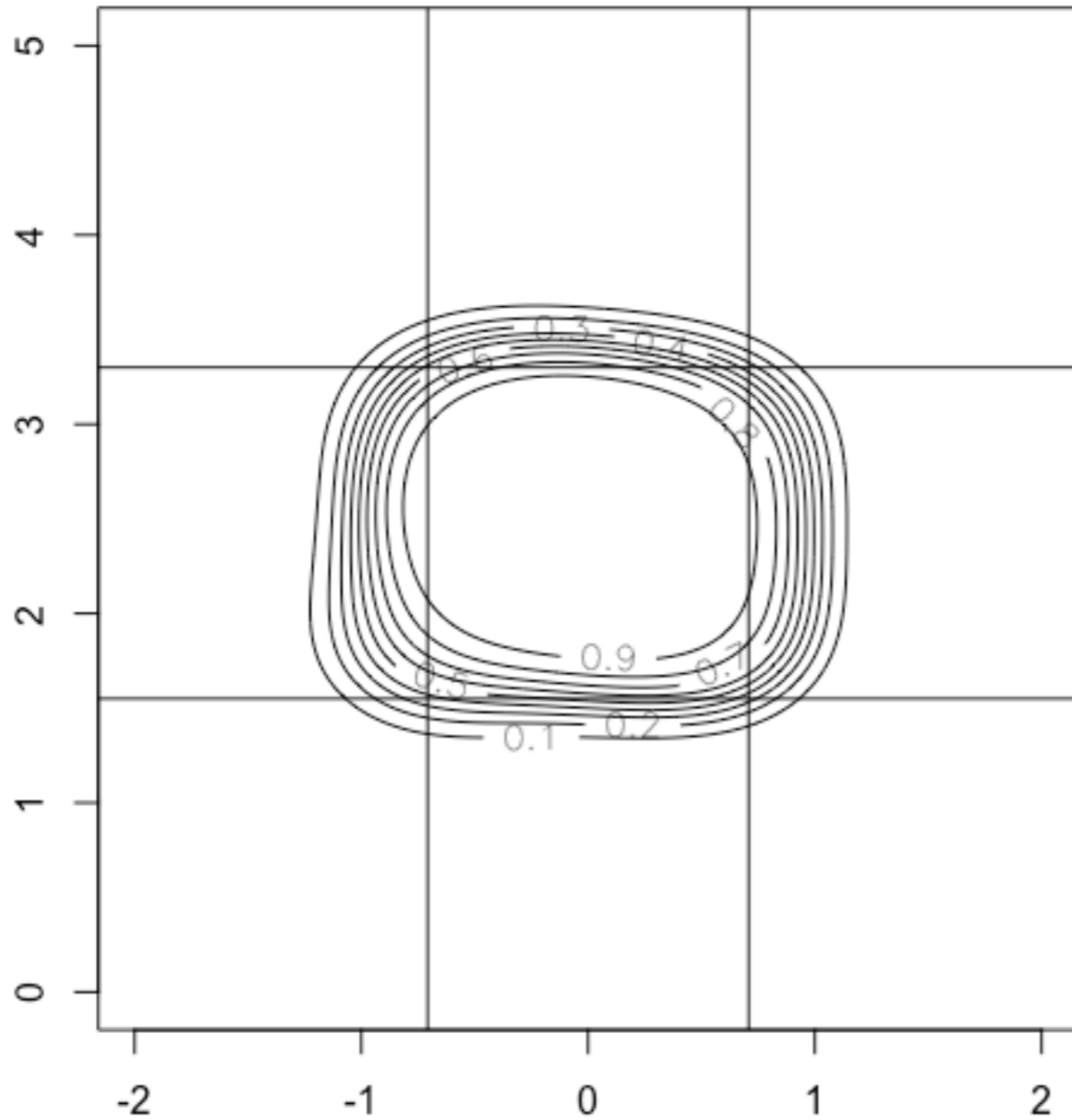
(Pavilidis, H., and Brooks, D. (2014))

$\text{plate}.x_i, \text{plate}.z_i$: Plate location for i th pitch

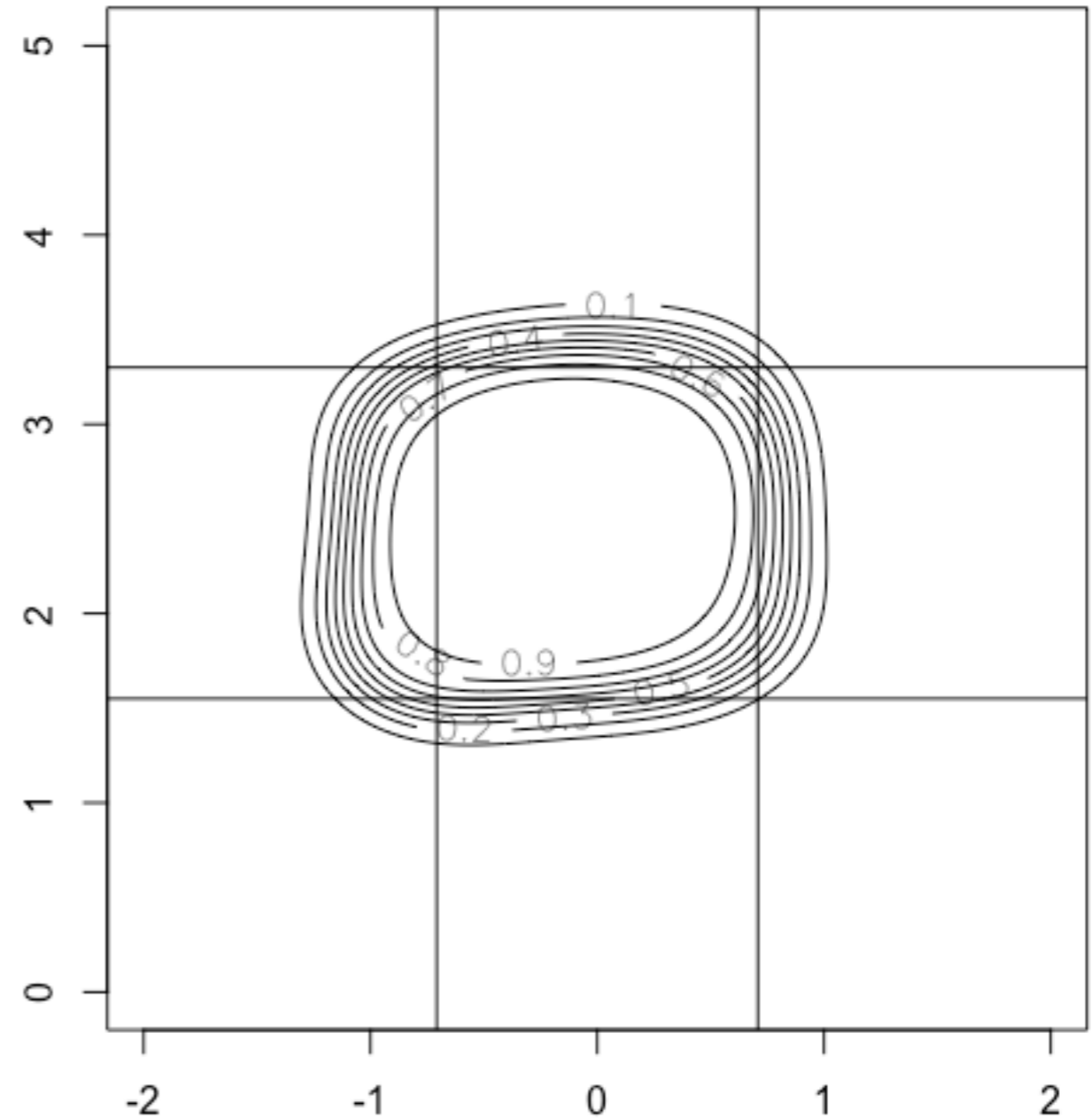
f : Thin plate regression spline

Visualization of Estimated Probability

Right-handed batter



Left-handed batter



***Catcher's point of view**

(Ordinary) Run Value

- Run Value is calculated by averaging the difference of **Run Expectancy and Runs** before and after a certain play event.

Play Event	Run Value
Single	0.47
Double	0.77
Triple	1.09
Home Run	1.39
Ground Out	-0.24
Fly Out	-0.28
Double Play	-1.06
Strikeout	-0.30
Sac Bunt	-0.13
...	...

Run Value of Strike and Ball

Count	Ball	Strike
0-0	0.00	-0.02
0-1	0.01	-0.15
0-2	0.01	-0.16
1-0	0.03	-0.02
1-1	0.01	-0.14
1-2	0.02	-0.17
2-0	0.23	-0.04
2-1	0.16	-0.14
2-2	0.15	-0.19
3-0	0.03	-0.10
3-1	0.13	-0.15
3-2	0.28	-0.34

$$(\text{Framing Run Value}) = - \left(\text{ } - \text{ } \right)$$

Framing Run Value

Count	Maximum Framing Run Value
0-0	0.080
0-1	0.092
0-2	0.199
1-0	0.112
1-1	0.117
1-2	0.241
2-0	0.156
2-1	0.098
2-2	0.339
3-0	0.173
3-1	0.251
3-2	0.590

Pavilidis,H.,and Brooks,D.(2014)

Calculation Example

Assumption

- throw to the location where called strike probability is **80%**
 - 0-2 count (Framing Run Value : 0.199)
- estimated
by model !

- Catcher can get run value as follow,

- If pitch is called **strike**,

$$(1 - \mathbf{0.80}) \times 0.199 = \mathbf{+0.0398} \text{ runs}$$

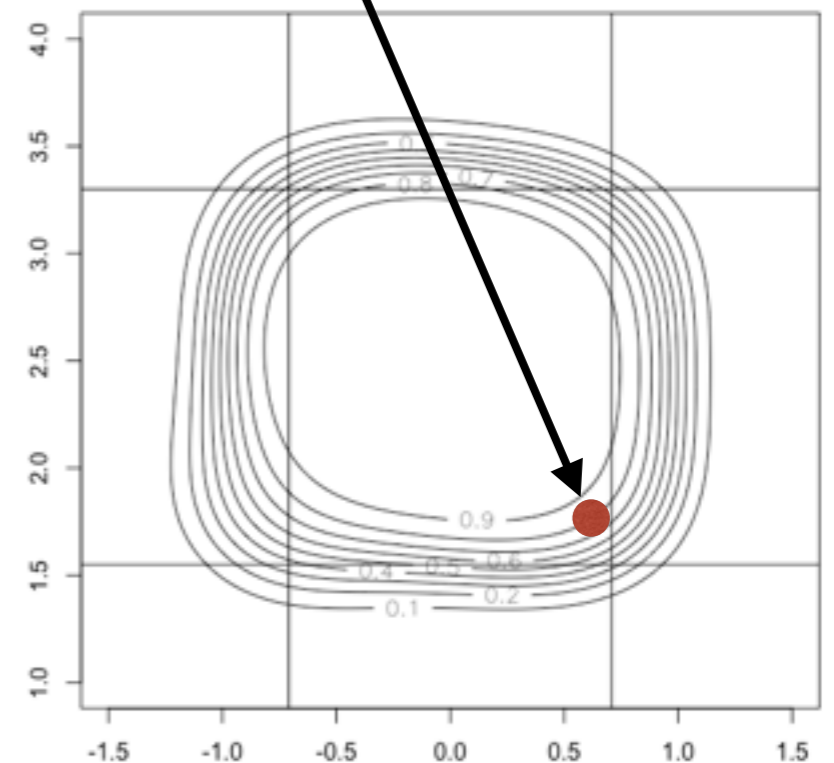
20% Available!

- If pitch is called **ball**,

$$(0 - \mathbf{0.80}) \times 0.199 = \mathbf{-0.1592} \text{ runs}$$

- Sum up these values through the season

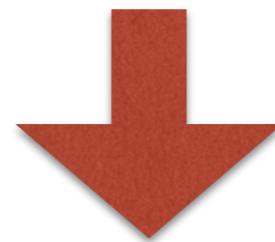
pitch location!



Extension of Model

There are various factors related to a pitch result.

- Umpire → strike-zone size
- Pitcher → pitching command
- Batter → batter stance



Treat these factors as **Random Effects**

Estimate random effect by **Mixed Model**

Generalized Additive **Mixed** Model (GAMM)

$$\log \frac{p_i}{1 - p_i} = \alpha + \underbrace{f(\text{plate}.x_i, \text{plate}.z_i)}_{\text{smooth term}} + \underbrace{\gamma_i^u + \gamma_i^p + \gamma_i^b}_{\text{random effect}}$$

$$\gamma_i^u \sim N(0, \sigma_u^2) \quad , \quad \gamma_i^p \sim N(0, \sigma_p^2) \quad , \quad \gamma_i^b \sim N(0, \sigma_b^2)$$

Based on GAMM estimation, adjust other factors by removing dispersion of umpires and players.

$$\underbrace{\hat{\alpha} + \hat{f}(\text{plate}.x_i, \text{plate}.z_i)}_{\text{use for evaluation}}$$

ANOVA Table

	edf	Chi-squared value	p-value
f(plate.x,plate.z) R-handed batter	35.26	33540.1	$< 2.0 \times 10^{-16}$
f(plate.x,plate.z) L-handed batter	35.03	27317.1	$< 2.0 \times 10^{-16}$
umpire_id	426.57	2247.4	$< 2.0 \times 10^{-16}$
batter_id	269.12	631.0	$< 2.0 \times 10^{-16}$
pitcher_id	77.65	907.2	$< 2.0 \times 10^{-16}$

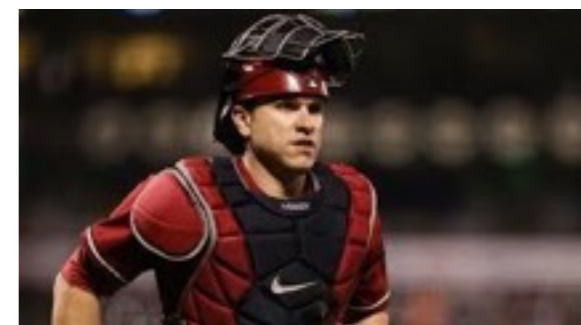
***Sample Size = 379,170**

***edf = estimated degree of freedom**

Total Framing Run Value in 2014 Season

Table : Framing Run Value Ranking

Name	Total Run Value	Sample Size
Janathan Lucroy	23.95	10,218
Rene Rivera	23.24	6,142
Miguel Montero	22.98	9,600
Mike Zunino	21.94	9,203
Buster Posey	18.37	7,542
⋮	⋮	⋮



Run Value per 10,000 Pitches

Name	Run Value per 10,000	Team
Rene Rivera	37.84	Padres
David Ross	37.24	Red Sox
Hank Conger	31.62	Angels
Christian Vazquez	30.23	Red Sox
Buster Posey	24.36	Giants
⋮	⋮	⋮



*only over 3,000 samples players

Reference

- Wood,S.N.(2006). Generalized Additive Models : an introduction with R. Chapman and Hall/CRC. New York.
- Wood,S.N.(2003). Thin plate regression splines. Journal of the Royal Statistical Society: Series B (Statistical Methodology) J Royal Statistical Soc B.
- Pavilidis,H.,and Brooks,D.(2014). Framing and Blocking Pitches: A Regressed, Probabilistic Model. (<https://www.baseballprospectus.com/article.php?articleid=22934>).
- John Walsh.(2008). Searching for the game's best pitch. (<http://www.hardballtimes.com/searching-for-the-games-best-pitch/>)