# Noisy Judgments:

# A probability surface-based analysis of umpiring variability

## James Stockton, PhD and Emily-Anne Patt

# SABR Analytics Conference

12 March 2023

## James Stockton, PhD

- Lead Data Scientist, Altamira Technologies
- Currently supporting USAF Chief Data and AI Office
- Prior data science work supporting private industry and defence/intelligence clients
- Can't hit a curveball



Emily-Anne Patt Manager,

Alphabet, Inc.

Quantitative Intelligence & Methodologies

Coach at Washington Nationals Youth Baseball Academy





### Motivation

• The strike zone is the fundamental measurement of baseball.

• A measurement is only as good as its ruler.

• The umpire is the ruler, not the MLB rule book.

Introduction / / / / / /

• So...how good is that ruler?

## **Analytical Outline**

 Collect data on called balls and strikes.

- 2. Use the data to model the strike zone as it's actually called.
- 3. Explore how that model changes across subsets of the data.







Introduction (1) (1) (1)

## Data Sources

### **Smart Fantasy Baseball**

Umpire biographical information

RetroSheet

Match umpires to game

**Baseball Savant** 

Pitch-by-pitch data



Data Collection 🚺 🔵 🗌

### **Combined Dataframe**

| 0 SL 202 656 519 nan bal W R R HOU PHI 0 1 202 0.9 3.3 0  | 9 Top 455 3.5 1.7 715 61                             | 2 511 4 1 1 4 0              | 0.0 0 PHI HOU bar 202 0 7:0 bar Bar Rob Lan 03/ 6  | 85/ 10/ 5-1 175 nan Rhy 3/1 Rhy Hos 656 Rya 12/ Rya Pre 519 0  | 3.1 3.4 1.6 1.8 93 328 28 97                                      |           |            |             |                   |
|---|--|------------------------------|--|--|---|-----------|------------|-------------|-------------------|
| 1 SL 202 663 502 nan bal W L R HOU PHI 0 0 202 -1, 2,7 2  | 8 Bot 592 3.5 1.6 715 60                             | 1 511 4 1 4 1 0              | 0.0 0 PHI HOU bar 202 0 7:0 bar Bar Rob Lan 03/ 0  | 05/ 10/ 5-1 175 nam Kyl 1/1 Kyl Tuc 663 Dav 4/9 Dav Rob 502 0  | 2.6 3.4 1.6 1.8 -11 265 -34 81                                    |           |            |             |                   |
| 3 FC 202 000 002 ran bal W R R HOU PH1 2 2 202 0.2 0.4 1  | 5 Bot 592 3.1 1.5 715 59                             | 4 cut 4 1 4 1 8              | 8.0 0 PHI HOU bar 202 8 7:8 bar Bar Rob Lan 83/ 6  | 107 107 3-1 173 101 Ale 373 Ale Bre 688 Day 479 Day Rob 502 8  | 8.4 3.4 1.6 1.8 28 47 9 14  |           |            |             |                   |
| 4 FC 202 685 582 num bal W R R HOU PHI 8 2 202 2.0 -8. 1  | 8 Bot 592 3.2 1.5 715 59                             | 3 Cut 4 1 4 1 8              | 8.6 0 PHI HOU bar 202 8 7:8 bar Bar Rob Lan 83/ 6  | 85/ 10/ 5-1 175 nan Ale 3/3 Ale Bre 688 Dav 4/9 Dav Rob 582 8  | -8. 3.4 1.6 1.8 203 -22 62 -7                                     |           |            |             |                   |
| 5 KC 202 670 582 nan bal W L R HOU PHI 8 1 202 8.0 -1.0   | 8 Bot 592 3.4 1.7 715 58                             | 2 Knu 4 1 4 1 8              | 8.8 • PHI HOU bax 202 8 7:8 bax Bar Rob Lan 83/ 4  | 85/ 18/ 5-1 175 nan Yor 6/2 Yor Alv 678 Dav 4/9 Dav Rob 582 8  | -1. 3.4 1.6 1.8 4 -12 1 -39                                       |           |            |             |                   |
| 6 FC 202 670 502 nan cal W L R HOU PHI 0 0 202 -0. 2.8 0  | 8 Bot 592 3.6 1.8 715 58                             | 1 Cut 4 1 4 1 8              | -8. 0 PHI HOU bar 202 8 7:8 bar Bar Rob Lan 83/ 6  | 85/ 10/ 5-1 175 man Yor 6/2 Yor Alv 670 Dav 4/9 Dav Rob 502 1  | 2.6 3.4 1.6 1.8 91 261 -28 79                                     |           |            |             |                   |
| 5 55 202 050 050 nm bal W L R HOU PHT 0 1 202 11 0. 2   | 8 Top 455 3 1 1 6 715 56                             |                              | 9 8 9 PHI HOU DAI 202 8 7:8 DAI BAI ROU LAN 057 6  | 107 107 5-1 175 man Kyi 575 Kyi 500 619 472 619 ADI 600 6  |   |           |            |             |                   |
| 9 FF 202 669 650 man bal W L R HOU PHI 0 1 202 1. 3.1   |  |                              |  |  |   |           | Ditabor    |             |                   |
| 10 SL 202 669 650 man cal W L R HOU PHI 0 0 202 0.5 4 1   | Y  | 7                            | strikezone(z)                                      | strikezone(z)  | Ilmnire   | Pitcher   | FILCHEI    | Ratter      | Batter handedness |
| 11 FF 202 681 658 nan cal W L R HOU PHI 0 0 202 3 2.6 0   | ^  |                              | Strikezone(2) <sub>top</sub>                       | bottom   | omplie  | i iteriei | handedness | Dutter      | Batter Hanacaness |
| 12 51 282 865 621 man cal W R R HOU PHI 8 8 482 8.6 2.4 2   | 7 807 592 2 8 1 2 715 53                             | 3 (07.4 1 4 1 8              | 8 8 8 PHT HOIL bar 282 8 7:8 bar Bar Roh Lao 83/ 8 | 15/ 18/ 5.1 175 mm los 5/6 los 411 514 Zar 4/8 Zar Ff1 523 8   | 87341618 94 71 29 22  |           |            |             |                   |
| 14 SI 202 514 621 nan bal W R R HOU PHT d 1 202 1.4 1.5 2   | 7 Bot 592 2.9 1.3 715 53                             | 2 Sin 4 1 4 1 8              | 8.6 0 PHI HOU bar 202 8 7:8 bar Bar Rob Lan 83/ 6  | 85/ 10/ 5-1 175 man Jos 5/6 Jos Alt 514 Zac 4/8 Zac Efl 621 0  | 1.9 3.4 1.6 1.8 147 192 45 58                                     |           |            |             |                   |
| 15 CU 202 455 621 nan bal W R R H0 PHI 0 2 202 0.9 0.3 1  | 7 Bot 592 3.2 1.4 715 52                             | 3 Cur 4 1 4 1 8              | 8.8 0 PHI HOU bax 202 8 7:8 bar Bar Rob Lan 83/ 6  | 85/ 10/ 5-1 175 nan Mar 8/1 Mar Mal 455 Zac 4/8 Zac Efl 621 0  | 0.5 3.4 1.6 1.8 97 58 30 15                                       |           |            |             |                   |
| 16 SI 202 455 621 nan cal W R R HOU PHI 0 0 202 0.7 1.5 1   | 7 Bot 592 3.2 1.5 715 52                             | 1 Sin 4 1 4 1 8              | -8. • PHI HOU bar 202 0 7:0 bar Bar Rob Lan 03/ 0  | 85/ 10/ 5-1 175 nan Mar 8/1 Mar Mal 455 Zac 4/8 Zac Efl 621 1  | 1.6 3.4 1.6 1.8 77 165 23 50                                      |           |            | _           |                   |
| 17 CU 202 676 621 nan bal W P R HOU PHI 1 2 202 1.2 0.6 0   | 7 Bot 592 3.1 1.5 715 51<br>7 Bot 592 3.1 1.5 715 51 | 4 CUT 4 1 4 1 0              | 8.8 0 PHI HOU bar 202 8 7:8 bar Bar Rob Lan 83/ 6  | 15/ 18/ 5-1 175 nan Cha 4/1 Cha McC 676 Zac 4/8 Zac Ef1 621 8  | 0.6 3.4 1.6 1.8 120 66 37 20                                      |           |            |             |                   |
| 10 FF 202 054 593 nan 11 W R R HOU PHI 0 0 202 -0. 1.7 1  | 7 Top 455 3.4 1.5 715 49                             | 1 4-5 4 1 1 4 0              | -0. • PHI HOU bar 202 8 7:8 bar Bar Rob Lan 83/ 6  | 257 107 5-1 175 man Ale 8/3 Ale Boh 664 Mec 6/1 Mec Ner 593 1  | 1.8 3.4 1.6 1.8 -8 186 -2 57                                      |           |            |             |                   |
| 20 SI 202 592 593 m cal W R R HOU PHI 1 0 202 0.0 2.3 0   | 7 Top 455 3.6 1.6 715 48                             | 2 Sin 4 1 1 4 8              | -8. 0 PHI HOU bar 202 8 7:8 bar Bar Rob Lan 83/ 6  | 85/ 10/ 5-1 175 man Nic 3/4 Nic Cas 592 Mec 6/1 Mec Ner 593 1  | 2.2 3.4 1.6 1.8 4 226 1 69  |           |            |             |                   |
| 21 SI 202 592 35 nan bal W R R HOU PHI 8 0 202 1.1 1.8 0  | 7 Top 455 3.7 1.7 715 48                             | 1 Sin 4 1 1 4 8              | 8.8 • PHI HOU bar 202 8 7:8 bar Bar Rob Lan 837 6  | 85/ 10/ 5-1 175 nan Nic 3/4 Nic Cas 592 Mec 6/1 Mec Nex 593 0  | 1.7 3.4 1.6 1.8 110 175 34 53                                     |           |            |             |                   |
| 22 SL 202 341 622 nan bal w R R HOU PHI 8 1 202 1.8 1.5 2   | 6 Bot 592 3.6 1.7 715 47                             | 2 511 4 1 4 1 8              | 8.8 0 PHI HOU bax 202 8 7:8 bar Bar Rob Lan 83/ 6  | 85/ 10/ 5-1 175 nan Tre 3/1 Tre Man 641 Ser 11/ Ser Dom 622 0  | 1.4 3.4 1.6 1.8 186 148 57 43                                     |           |            |             |                   |
| 23 SL 202 543 622 nan bal w k k HOU PHI 1 2 202 0.9 0.8 2   | 6 Bot 592 3.1 1.6 715 46                             |                              | 8.8 0 PHI HOU bar 202 8 7:8 bar Bar Rob Lan 83/ 6  | 15/ 18/ 5-1 175 nan Chr 8/2 Chr Vaz 543 Ser 11/ Ser Dom 622 8  |   |           |            |             |                   |
| 25 FC 202 663 621 nan bal W L L HOU PHI 2 1 202 1.1.5 1   | 6 Bot 592 3.4 1.6 715 45                             | 4 Cut 3 1 3 1 8              | 8.8 0 PHI HOU bar 202 8 7:8 bar Bar Rob Lan 83/ 6  | 85/ 10/ 5-1 175 man Kyl 1/1 Kyl Tuc 663 Jos 5/2 Jos Alv 621 0  | 1.5 3.4 1.6 1.8 -12 152 -37 46                                    |           |            |             |                   |
| 6 SI 202 663 621 man bal W L L HOU PHI 1 1 202 -0. 1.0 1  | 6 Bot 592 3.4 1.7 715 45                             | 3 <mark>5in</mark> 3 1 3 1 8 | 8.8 0 PHI HOU bar 202 8 7:8 bar Bar Rob Lan 83/ 6  | 85/ 10/ 5-1 175 man Kyl 1/1 Kyl Tuc 663 Jos 5/2 Jos Alv 621 0  | 0.8 3.4 1.6 1.8 -29 87 -6 27                                      |           |            |             |                   |
| 27 FC 202 663 621 nan bal W L L HOU PHI 0 0 202 -10. 1  | 6 Bot 592 3.4 1.7 715 45                             | 1 Cut 3 1 3 1 8.0            | 8:8 0 PHI HOU bar 202 8 7:8 bar Bar Rob Lan 83/ 6  | 85/ 10/ 5-1 175 man Kyl 1/1 Kyl Tuc 063 Jos 5/2 Jos Alv 623 0  | -8. 3.4 1.6 1.8 -11 -23 -35 -7                                    |           |            |             |                   |
| 28 ST 202 688 621 man cal W R L HOU PHI 3 0 202 8.3 1.9 1   | 6 Bot 592 3.2 1.5 715 44                             | 4 Sin 3 1 3 1 8              | 10 0 PHI HOU bar 282 0 7:8 bar Bar Rob Lan 83/ 0   | 25/ 10/ 5-1 175 min Ale 3/3 Ale Bre 608 Jos 5/2 Jos Alv 621 1  |   |           |            |             |                   |
| 30 SI 202 608 621 nan bal W R L HOU PHI 1 0 202 1. 3 1  | 6 Bot 592 3.2 1.6 715 44                             | 2 Sin 3 1 3 1 8              | 8.8 0 PHI HOU bar 202 8 7:8 bar Bar Rob Lan 83/ 6  | 107 107 5-1 175 man Ale 3/3 Ale Bre 608 Jos 5/2 Jos Alv 621 0  | 3.1 3.4 1.6 1.8 14 314 43 96                                      |           |            |             |                   |
| 31 FF 202 608 621 nam bal W R L HOU PHI 8 8 202 1. 3.2 1  | 6 Bot 592 3.1 1.5 715 44                             | 1 4-5 3 1 3 1 8              | 8.8 0 PHI HOU bar 202 8 7:8 bar Bar Rob Lan 83/ 6  | 85/ 10/ 5-1 175 nan Ale 3/3 Ale Bre 608 Jos 5/2 Jos Alv 621 0  | 3.5 3.4 1.6 1.8 -11 358 -35 107                                   |           |            |             |                   |
| 72 SI 202 670 621 nan bal W L L HOU PHI 1 1 202 1.2 4.1 1   | 6 Bot 592 3.4 1.7 715 43                             | 3 Sin 0 1 0 1 0              | 8.1 0 PHI HOU bar 202 8 7:8 bar Bar Rob Lan 83/ 6  | 85/ 10/ 5-1 175 man Yor 6/2 Yor Alv 670 Jos 5/2 Jos Alv 621 0  | 4.1 3.4 1.6 1.8 127 428 39 128                                    |           |            |             |                   |
| 23 FC 202 670 621 nan bal W L L HOU PHI 0 1 202 2. 1.4 1  | 6 Bot 592 3.4 1.7 715 43                             | 2 Cut 0 1 0 1 0              | 0.0 0 PHI HOU bar 202 0 7:0 bar Bar Rob Lan 03/ 0  | 85/ 10/ 5-1 175 man Yor 6/2 Yor Alv 670 Jos 5/2 Jos Alv 621 0  | 1.2 3.4 1.6 1.8 -23 122 -71 37                                    |           |            |             |                   |
| M FF 202 055 954 Han Ball W K K HOU PHI 0 202 1.1 2.6 1   15 EE 202 3.14 3.54 pain ball W R R HOU PHI 0 0 202 1.1 2.6 1   15 EE 202 3.14 3.54 pain ball W R R HOU PHI 0 0 202 0.6 3.5 0 | 6 Bot 592 3.6 1.7 715 42<br>6 Bot 592 2.7 1.2 715 41 | 2 45 0 1 0 1 0               | 0.0 0 PHL HOU DAT 202 0 7:0 DAT BAT ROD LAN 837 0  | 19/ 19/ 5-1 1/5 man Jer 9/2 Jer Pen 065 28c 5/3 28c kme 554 0  |   |           |            |             |                   |
| 30 FF 202 514 554 man bal W R R HOU PHI 8 8 202 8.6 3.8 9   | 6 Bot 592 2.8 1.3 715 41                             | 1 4-5 0 1 0 1 0              | 8.0 0 PHI HOU bar 202 8 7:8 bar Bar Rob Lan 83/ 6  | 257 107 5-1 175 man Jos 5/6 Jos Alt 514 Zac 5/3 Zac Whe 554 0  | 4.0 3.4 1.0 1.8 08 400 21 142                                     |           |            |             |                   |
| 37 SI 202 455 554 nan cal W R R HOU PHI 1 0 202 -0. 2.7 0   | 6 Bot 592 3.4 1.5 715 40                             | 2 Sin 0 1 0 1 0              | -8. 0 PHI HOU bax 202 8 7:8 bar Bar Rob Lan 837 6  | 85/ 10/ 5-1 175 nan Mar 8/1 Mar Mal 455 2ac 5/3 Zac Whe 554 1  | 2.7 5.4 1.6 1.8 -69 277 -21 84                                    |           |            |             |                   |
| 36 SI 202 455 554 nan bal W R R HOU PHI 0 0 202 -1. 2.7 0   | 6 Bot 592 3.3 1.5 715 40                             | 1 Sin 0 1 0 1 0              | 8.8 0 PHI HOU bar 202 8 7:8 bar Bar Rob Lan 83/ 6  | 85/ 10/ 5-1 175 nan Mar 8/1 Mar Mal 455 Zac 5/3 Zac Whe 554 0  | 2.8 3.4 1.6 1.8 •13 288 •48 88                                    |           |            |             |                   |
| 39 SI 202 547 664 man cal W L L HOU PHI 1 0 202 -0. 1.4 2<br>40 SC 203 547 664 man bal W L L HOU SUT 0 0 202 -0. 1.2 2  | 6 Top 455 3.2 1.4 715 39                             | 2 Sin 0 1 1 0 0              | -8. 0 PHI HOU bar 202 8 7:8 bar Bar Rob Lan 83/ 6  | 05/ 10/ 5-1 175 nan Bry 10/ Bry Har 547 Fra 11/ Fra Val 664 1<br>25/ 10/ 5-1 175 nan Bry 10/ Bry Har 547 Fra 11/ Fra Val 664 0 | 1.6 3.4 1.6 1.8 -8 161 -2 49                                      |           |            |             |                   |
| 41 CH 202 592 664 nan cal W R L HOU PHI 0 0 202 -0. 1.4 1   | 6 Top 455 3.5 1.4 715 38                             | 1 Cha 0 1 1 0 0              | -0. 0 PHI HOU bar 202 8 7:8 bar Bar Rob Lan 03/ 6  | 157 107 5-1 175 man J.T 3/1 J.T Rea 592 Fra 11/ Fra Val 664 1  | 1.5 3.4 1.6 1.8 -26 159 -8 49                                     |           |            |             |                   |
| 42 CU 202 656 664 man bal W R L HOU PHI 1 8 202 0.3 1.4 9   | 6 Top 455 3.5 1.6 715 37                             | 2 Cur 0 1 1 0 0              | 8.0 0 PHI HOU bar 202 8 7:8 bar Bar Rob Lan 83/ 6  | 85/ 10/ 5-1 175 nam Rhy 3/1 Rhy Hos 656 Fra 11/ Fra Val 664 8  | 1.4 3.4 1.6 1.8 36 143 11 44                                      |           |            |             |                   |
| 43 CU 202 656 664 nam bal W R L HOU PHI 0 8 202 -0. 1.8 0   | 6 Top 455 3.6 1.7 715 37                             | 1 Cur 0 1 1 0 0              | 8.8 0 PHI HOU bar 202 8 7:8 bar Bar Rob Lan 03/ 6  | 85/ 10/ 5-1 175 nan Rhy 3/1 Rhy Hos 656 Fra 11/ Fra Val 664 8  | 1.7 3.4 1.6 1.8 -96 178 -29 52                                    |           |            |             |                   |
| 44 CU 202 056 064 man bal N L L HOU PHT 1 2 202 -1. 0.4 0   | 6 Top 455 3.1 1.4 715 36                             | 4 Cur 0 8 8 8 8              | 8.8 • PHI HOU bar 202 8 7:8 bar Bar Rob Lan 83/ 6  | 25/ 10/ 5-1 175 man Kyl 3/5 Kyl Sch 056 Fra 11/ Fra Val 664 1  | 0.5 3.4 1.6 1.8 -14 51 -45 16                                     |           |            |             |                   |
| 45 PC 202 055 064 man cal W L L HOU PHI 1 1 202 0.2 2.5 0   | 6 Top 455 3.1 1.5 715 36<br>6 Top 455 3.1 1.5 715 36 |                              | 8 8 0 PHI HOU bar 282 8 7:8 bar Bar Rob Lan 83/ 6  | 157 187 5-1 175 nan Kyi 375 Kyi 5ch 656 Fra 117 Fra Vai 664 1<br>157 187 5-1 175 nan Kyi 375 Kyi 5ch 656 Fra 117 Fra Vai 664 B | 2.7 3.4 1.6 1.8 21 272 6 83                                       |           |            |             |                   |
| 47 FC 202 656 664 man cal W L L HOU PHI 8 8 202 -8. 2.3 8   | 6 Top 455 3.1 1.5 715 36                             | 1 Cut 0 0 0 0 0              | 8. 0 PHI HOU bar 202 8 7:8 bar Bar Rob Lan 83/ 6   | 85/ 10/ 5-1 175 man Kyl 3/5 Kyl Sch 656 Fra 11/ Fra Val 664 1  | 2.4 3.4 1.6 1.8 49 245 15 75                                      |           |            |             |                   |
| 48 SI 202 676 554 nan cal W R R HOU PHI 0 0 202 -0. 2.8 2   | 5 Bot 592 3.1 1.5 715 35                             | 1 Sin 0 0 0 0 0              | -8. 0 PHI HOU bar 202 8 7:8 bar Bar Rob Lan 03/ 6  | 85/ 10/ 5-1 175 nan Cha 4/1 Cha McC 676 Zac 5/3 Zac Whe 554 1  | 3.8 3.4 1.6 1.8 -20 387 -6 94                                     |           |            |             |                   |
| 49 CU 202 641 554 nan bal W R R HOU PHI 1 2 202 0.7 0.2 1   | 5 Bot 592 3.6 1.7 715 34                             | 5 Cuz 0 0 0 0 0              | 8.0 0 PHI HOU bar 202 8 7:8 bar Bar Rob Lan 83/ 6  | 85/ 10/ 5-1 175 nan Tre 3/1 Tre Man 641 Zac 5/3 Zac Whe 554 8  | 0.1 3.4 1.6 1.8 74 12 23 4  |           |            |             |                   |
| 50 CU 202 G41 554 nan cal W R R HOU PHI 1 1 202 -0. 2.7 1   | 5 Bot 592 3.5 1.6 715 34                             | 3 Cuz 0 0 0 0 0              | -0. 0 PHI HOU bar 202 0 7:0 bar Bar Rob Lan 03/ 0  | 85/ 10/ 5-1 175 man Tre 3/1 Tre Man G41 Zac 5/3 Zac Whe 534 1  | 2.5 3.4 1.6 1.8 -55 268 -17 79                                    |           |            |             |                   |
| 51 51 202 041 554 nan cal W R R HOU PHI 0 1 202 -1. 3.8 1   | 5 Bot 592 3.6 1.5 715 34                             |                              | -8. 0 PHI HOU bar 202 8 7:8 bar Bar Rob Lan 83/ 6  | 197 187 5-1 175 man Tre 371 Tre Man 641 286 573 286 Whe 554 0  | 3.7 3.4 1.0 1.8 -12 371 -38 113<br>3.8 3.4 1.0 1.8 -30 188 -17 35 |           |            |             |                   |
| 53 CU 202 024 004 nan bal W R L HOU PHI 0 2 202 0.5 4.1 2   | 5 Top 455 5.4 1.7 715 32                             | 4 Cur 0 8 8 0 8              | 8.6 0 PHI HOU bar 202 8 7:8 bar Bar Rob Lan 837 6  | 25/ 10/ 5-1 175 man Edm 3/6 Edm Sos 024 Pra 11/ Fra Val 004 0  | 4.2 3.4 1.6 1.8 56 423 17 129                                     |           |            |             |                   |
| 54 SI 202 624 664 nan cal W R L HOU PHI 0 1 202 -0. 1.9 2   | 5 Top 455 3.3 1.6 715 32                             | 2 510 0 0 0 0 0              | -8. 0 PHI HOU bax 202 8 7:8 bar Bar Rob Lan 83/ 6  | 85/ 10/ 5-1 <mark>175</mark> nan Edn 3/6 Edn Sos 624 Fra 11/ Fra Val 664 1   | 1.8 3.4 1.6 1.8 -5 185 -2 56                                      |           |            |             |                   |
| 55 SI 202 624 664 nan cal W R L HOU PHI 0 0 202 0.3 2.7 2   | 5 Top 455 3.4 1.7 715 32                             | 1 Sin 0 8 8 0 8              | -8. 0 PHI HOU bar 202 8 7:8 bar Bar Rob Lan 83/ 6  | 95/ 10/ 5-1 175 nan Edn 3/6 Edn Sos 624 Fra 11/ Fra Val 664 1  | 2.7 3.4 1.6 1.8 30 276 9 84                                       |           |            |             |                   |
| 54 SI 202 663 664 nan bal W R L HOU PHI 2 1 202 -0. 0.5 1   | 5 Top 455 3.5 1.7 715 31                             | 4 Sin 0 0 0 0 0              | 8.8 0 PHI HOU bar 202 8 7:8 bar Bar Rob Lan 03/ 0  | 25/ 10/ 5-1 175 nan Mat 9/1 Mat Vie 663 Fra 11/ Fra Val 664 0  | 0.2 3.4 1.6 1.8 -94 28 -29 8                                      |           |            |             |                   |
| 54 SI 202 663 664 nan bal w R L HOU PHI 1 a 202 a 1 a 9 3   | 5 Top 455 3.4 1.7 715 31                             | 2 Sin a a a a                | 8.8 9 PHI HOU bar 202 8 7:8 bar Bar Rob Lan 83/ 6  | 25/ 10/ 5-1 175 nan Mat 9/1 Mat Vie 663 Fra 11/ Fra Val 664 1  | 9.8 3.4 1.6 1.8 19 81 3 25  |           |            |             |                   |
| 50 ST 202 653 664 man bal W R L HOU PHI 8 8 202 1.0 1.9 1   | 5 Top 455 3.4 1.7 715 31                             | 1 Sin e e e e                | 8.8 S FHI HOU bar 202 8 7:8 bar Bar Rob Lan 83/ 6  | 85/ 10/ 5-1 175 nan Mat 9/1 Mat Vie 661 Fra 11/ Fra Val 664 D  | 1.8 3.4 1.6 1.8 102 198 31 58                                     |           |            |             |                   |
| 60 CU 202 516 664 man bal W R L HOU PHI 0 8 202 -0. 8.0 0   | 5 Top 455 2.8 1.2 715 30                             | 1 Cur 0 0 0 0 0              | 8.0 9 PHI HOU bar 202 8 7:8 bar Bar Rob Lan 03/ 6  | 85/ 10/ 5-1 <mark>173</mark> nan Jea 3/1 Jea Seg <mark>516</mark> Fra 11/ Fra Val 664 0  | 0.1 3.4 1.6 1.8 -87 16 -27 5                                      |           |            |             |                   |
| 61 SL 202 670 534 nan cal W L R HOU PHI 8 8 202 8.7 3.2 0   | 4 Bot 592 3.5 1.8 715 27                             | 1 511 0 0 0 0 0              | -8. 0 PHI HOU bax 202 8 7:8 bax Bar Rob Lan 83/ 6  | 85/ 18/ 5-1 175 nan Yor 6/2 Yor Alv 678 Zac 5/3 Zac Whe 554 1  | 3.1 3.4 1.6 1.8 78 313 21 95                                      |           |            |             |                   |
| 62 CU 202 664 664 nan cal W R L HOU PHI 8 8 202 6. 2.3 2  | 4 Top 455 3.4 1.5 715 25                             | 1 Cur 0 0 0 0 0              | 0. 0 PHI HOU bar 202 0 7:0 bar Bar Rob Lan 03/ 0   | 15/ 10/ 5-1 175 nan Ale 8/3 Ale Boh 664 Fra 11/ Fra Val 664 1  |   |           |            |             |                   |
| 64 CU 202 592 664 nan bal W R L HOU PHI 2 2 202 10. 1   | 4 Top 455 3.5 1.7 715 24                             | 8 Cur 0 8 8 0 8              | 8.8 0 PHI HOU bar 202 8 7:8 bar Bar Rob Lan 83/ 6  | 15/ 10/ 5-1 175 nan Nic 3/4 Nic Cas 592 Fra 11/ Fra Val 664 1  | -0. 3.4 1.6 1.8 -11 -21 -35 -6                                    |           |            | Data Collec | tion 🖉 🌈 🖉 🦯 (    |
|   |  |                              |  |  |   |           |            |             | V 8 V 8 V         |

### Pitch location for 5.3 million called pitches



Data Collection

Building a Strike Zone Probability Surface: Step 1



Note: Reported pitch location accuracy and precision changes over time with evolution of pitch tracking technology.

Modeling 🥢 🔿 🔿

Building a Strike Zone Probability Surface: Step 2 The Useful Kind of Errors

- The number of pitches falling into each grid cell varies.
- The strike percentage in each cell is measured with a different level of confidence.
  - Sampling and survey statistics when measuring proportions
- The pitches in the cell vote on the likelihood a new pitch will be a ball or a strike.
- Accounting for measurement errors in the fitting process improves results.







## **Height Normalization**

 Used StatCast reporting of the sz<sub>top</sub> and sz<sub>bottom</sub> for every pitch

• Averaged (mean) values for each defines a normalized strike zone height

• Proportionally scaled each pitch location relative to this standardized size

Modeling 1010000

### Building a Strike Zone Probability Surface: Step 3 The Anatomy of a Strike Zone Sigmoid



**Takeaway:** Each probability distribution is made up of two sigmoids, curving from the middle of the strike zone where the pitch will nearly always be called a strike to the edge of the strike zone where the strike calls are rare.

Modeling 10101000

Building a Strike Zone Probability Surface: Step 4 The Anatomy of a Strike Zone Sigmoid



Building a Strike Zone Probability Surface: Step 4 The Anatomy of a Strike Zone Sigmoid



Modeling 101010100



#### Key Benefits of the Method

- 1. Accounts for sampling errors in the data
- 2. Well-tuned underlying functional form matches the shape of the data
- 3. The fitting process smooths out noisy data using only a handful of free-parameters

**Takeaway:** The results are reliable to smaller sample sizes than prior efforts, which opens up access to questions about specific matchups and changes over time.

Modeling 1010101010

## The Umpires' Strike Zone

- The best-fit strike zone for all umpires across all seasons in the data set is 57.7 cm along the horizontal axis and 56.1 cm tall.
- The best-fit strike zone is shifted 1.6 cm to the left of the center of the plate.

Best-fit Strike Zone and Associated Probability Surface for All Available Data in the 2008-2022 Seasons The 50% Probability Contour Defines the Strike Zone and is Shown with its Dimensions



Results 🕼 🔿 🔿 🔿

#### Validating the Ruler

The Umpires' Best-Fit Strike Zone Compared to the Official Zone



- The best-fit strike zone (blue dotted box) is 7.1 cm wider than the official strike zone (red dashed box).
  - Red dashed box passes through the center of a baseball that is tangent to home plate.

Results

• The best-fit strike zone is shifted 1.6 cm to the left of the center of the plate.

#### Validating the Ruler

The Umpires' Best-Fit Strike Zone Compared to the Official Zone



- The best-fit strike zone for right-handed batters (green dashed box) and left-handed batters (blue dotted box) have nearly identical areas
- The right-handed strike zone is centered roughly over the plate, while the left-handed strike zone is shifted 3.9 cm to the outside of the plate

Results 1010000

## The Lefty Strike The Presence of the Shift is Uniform across the Umpire Corps



- [Top] The shift in the strike zone for left-handed batters holds true across the umpire corps, with the shift measuring between 1 and 7 cm off-center.
- [Middle] Centered roughly on zero, the histogram shows the area of umpires' strike zones generally does not shift between right- and left-handed batters.
- [Bottom] We see a clear difference in the aspect ratio, with the lefty strike zone taller and narrower than the right-handed batters' zone.

Results 0000000

## The Lefty Strike The Lefty Shift More Pronounced for Right-Handed Pitchers



- The lefty-strike is even more pronounced when there is a RHP (purple dotted rectangle).
- The magnitude of the strike zone shift between left and right-handed batters for right-handed pitchers (purple) is roughly double the shift for left-handed pitchers (green).

Results 101010100

#### **Seasonal Shifts**



**Takeaway:** The area of the strike zone has increased gradually since the 2008 season.



Umpires' Best-fit Strike Zone Aspect Ratio Distributions Across the 2008-2022 Seasons

*Takeaway:* The strike zone has become narrower and taller since the 2008 season.

### **Seasonal Shifts**



- The strike zone narrowed nearly 14%, from 63 to 54 cm between the 2008 and 2022 seasons.
- The left side of the strike zone moved towards the plate about 7 cm.
- The upper limit of the strike zone has remained relatively consistent across the umpire corps since 2008, rising by 2 cm (2%) during that period.
- The bottom of the strike zone, however, dropped by nearly 9 cm—more than 16%—between 2008 and 2022.



The IoU ratio for the umpire corps ranges from 0.70 to 0.90, with the majority of umpires clustering around the 0.85 mark.



#### Around the League: Umpire Accuracy

Discussion 🥢 🔿 🤇

#### Around the League: Elite Pitchers



### Around the League: Notable Hitters









Best Fit Strike Zone MLB Strike Zone

## What's next?

What insights emerge when we split the data 2+ ways? Add pitch selection? Swing rate? How do MLB's 2023 rule changes affect the shape of the strike zone?

How have pitchers and batters adjusted to the strike zone's changing dimensions over time?

How can the model help us measure a catcher's framing performance?