MA388 Sabermetrics: Lesson 13

Catcher Framing Ability - Visualization

LTC Jim Pleuss

Chapter 7

In this chapter, we investigate whether catchers have the ability to consistently frame close pitches, which some believe makes it more likely the umpire will call a strike.

Do you think some catchers are better than others at framing pitches, and how important is this skill?

- Video 1
- Video 2

How could we investigate whether some catchers are better than others? List some challenges with answering this question with data.

```
library(tidyverse)
library(knitr)
library(baseballr)
```

The code below imports the all pitches thrown in May 2019.

```r
# If you've already collected the pitch data below, just load it.  Otherwise,
# collect it from scratch.

if("statcast_may_2019.rds" %in% dir(getwd())){
  pitches <- readRDS("statcast_may_2019.rds")
}else{

  # Retrieve pitch-level data from May 2019.

get_statcast_pitches <- function(start_day, end_day, chunk_size = 5) {

  # Coerce to Date
  start_day <- as.Date(start_day)
  end_day   <- as.Date(end_day)

  # Create sequence of chunk start dates
  chunk_starts <- seq(start_day, end_day, by = paste(chunk_size, "days"))

  # Build data
  pitch_data <- map_dfr(chunk_starts, function(chunk_start) {

    chunk_end <- min(chunk_start + days(chunk_size - 1), end_day)
       statcast_search(
        start_date = chunk_start,
        end_date   = chunk_end
      )
  })
  return(pitch_data)
}

  # If we want to limit to a certain number of pitches, we might take the head()
  # or sample_n() to get the number we want.
  pitches <- get_statcast_pitches('2019-05-01','2019-05-31')

  saveRDS(pitches, "statcast_may_2019.rds")
}
```

Here is a small sample of the Statcast pitches and variables we've collected.

```
pitches %>%
  select(description, release_speed, pitch_name,
         plate_x, plate_z, balls, strikes, fielder_2) %>%
  head(10) %>%
  kable()
```

description	release_speed	pitch_name	plate_x	plate_z	balls	strikes	fielder_2
swinging_strike	79.2	Changeup	-0.54	1.79	2	2	444489
foul	89.3	Sinker	-0.74	2.42	2	2	444489
foul	79.5	Changeup	-0.32	1.25	2	1	444489
ball	88.3	Sinker	-1.28	1.95	1	1	444489
ball	88.1	Sinker	-1.82	2.29	0	1	444489
called_strike	87.5	Sinker	-0.17	2.11	0	0	444489
hit_into_play	78.6	Changeup	-0.27	1.20	1	2	444489
ball	79.0	Changeup	-1.25	1.23	0	2	444489
foul	78.5	Changeup	0.49	1.47	0	1	444489
foul	86.6	Sinker	0.38	2.52	0	0	444489

One of the variables in `pitches` is the catcher's ID. Let's add the catcher's name to each pitch in our data frame using the master list available at https://www.smartfantasybaseball.com/tools/.

```
# This CSV is available at https://www.smartfantasybaseball.com/tools/.
mlb_IDs <- read_csv("./PLAYERIDMAP.csv")

pitches <- pitches %>%
  left_join(select(mlb_IDs, PLAYERNAME, MLBID),
            by = c("fielder_2" = "MLBID")) %>%
  rename(catcher_name = PLAYERNAME)
```

Let's look at the 10 catchers with the most pitches thrown to them in early May 2019.

```
top10catchers <- pitches %>%
  count(catcher_name) %>%
  arrange(-n) %>%
  head(10) %>%
  pull(catcher_name)

top10catchers
```

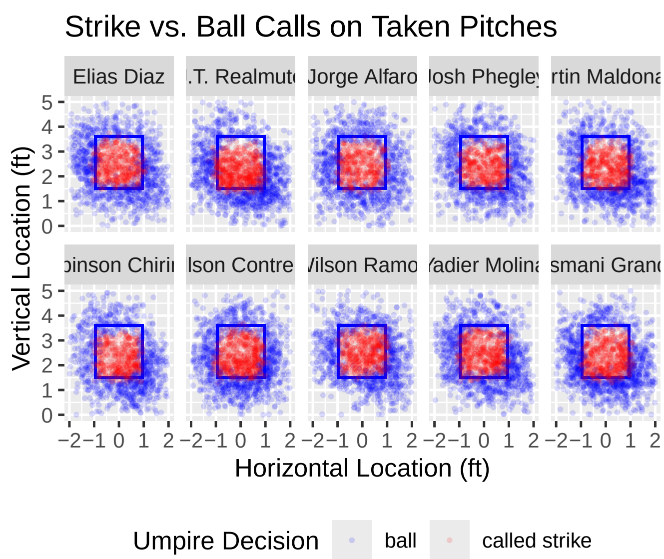
```
 [1] "Yasmani Grandal"   "J.T. Realmuto"     "Willson Contreras"
 [4] "Yadier Molina"     "Elias Diaz"        "Martin Maldonado"
 [7] "Jorge Alfaro"      "Robinson Chirinos" "Wilson Ramos"
[10] "Josh Phegley"
```

Okay, let's see if we can see anything by looking at called strikes versus balls.

```
# Let's plot the strike zone.
# Note px and pz in the book are actually plate_x and plate_z, respectively.
plate_width <- 17 + 2 * (9/pi)
k_zone_plot <- ggplot(NULL, aes(x = plate_x, y = plate_z)) +
  geom_rect(xmin = -(plate_width / 2) / 12,
            xmax = (plate_width / 2) / 12,
            ymin = 1.5,
            ymax = 3.6,
            color = "blue",
            alpha = 0) +
  coord_equal() +
  scale_x_continuous("Horizontal Location (ft)", limits = c(-2,2)) +
  scale_y_continuous("Vertical Location (ft)", limits = c(0,5))

# Look at only taken pitches caught by one of the 10 top catchers.
pitches_taken <- pitches %>%
  filter(catcher_name %in% top10catchers,
         description %in% c("ball", "called_strike")) %>%
  mutate(description = str_replace(description, "_", " "))

# Plot pitches that were balls or called strikes.
k_zone_plot %+% pitches_taken +
  aes(color = description) +
  geom_point(alpha = 0.15, size = .5) +
  scale_color_manual(values = c("blue", "red")) +
  facet_wrap(~catcher_name, ncol = 5) +
  theme(legend.position = "bottom") +
  labs(title = "Strike vs. Ball Calls on Taken Pitches",
       color = "Umpire Decision")
```



How can we make some sense of the different panels in this figure? Look closely – does anything jump out at you?

Okay, this is admittedly pretty subjective, and we can't stop here. We need a quantitative assessment of how different catchers may or may not be contributing to different proportions of called strikes on pitches where batters didn't swing.

```
pitches_taken %>%
  count(catcher_name, description) %>%
  group_by(catcher_name) %>%
  mutate(prop = n / sum(n)) %>%
  filter(description == "called strike") %>%
  arrange(-prop) %>%
  kable(digits = 3,
        caption = "Proportion of taken pitches that were called strikes by catcher.")
```

Table 2: Proportion of taken pitches that were called strikes by catcher.

catcher_name	description	n	prop
Wilson Ramos	called strike	504	0.362
Willson Contreras	called strike	575	0.356
J.T. Realmuto	called strike	602	0.353
Yadier Molina	called strike	533	0.346
Martin Maldonado	called strike	496	0.337
Robinson Chirinos	called strike	471	0.334
Yasmani Grandal	called strike	566	0.325
Josh Phegley	called strike	443	0.320
Elias Diaz	called strike	466	0.317
Jorge Alfaro	called strike	427	0.302

Does it appear catchers make a difference?

Now that we've been talking about this for a while, what new considerations come to mind?

How could a model-based approach improve upon this analysis?