

Catch probability

Tim White

rSys.Date()

```
library(tidyverse)
library(rjson)
```

Load in data

We scrape individual play data for every fly ball hit to a qualified outfielder during the 2024 season. This chunk only needs to be run once.

```
# Outfielders' player IDs
playerIDs <- read_csv("../data/of_playerIDs_2024.csv") %>% select(player_id)

# Scrape play-by-play data
data <- lapply(1:nrow(playerIDs),
  function(j) {
    # Scrape data
    rawdata <- fromJSON(
      file = paste0("https://baseballsavant.mlb.com/player-services/range?playerId=",
        playerIDs[j,], "&season=2024&playerType=fielder"), simplify = TRUE
    )

    # If the URL exists:
    if (length(rawdata) > 0) {

      # Change any null columns (e.g., sprint_speed) to NA
      for (k in 1:length(rawdata)) {
        rawdata[[k]][sapply(rawdata[[k]], is.null)] <- NA
      }

      # Convert raw data to tibble
      tibble(data.frame(matrix(unlist(rawdata),
                                nrow = length(rawdata),
                                byrow = TRUE,
                                dimnames = list(1:length(rawdata),
                                                  names(rawdata[[1]])))) %>%
        mutate(across(c(game_pk:name_display_first_last, pos),
          as.factor),
          across(c(stars:distance, hang_time, out:sprint_speed),
            as.numeric)))
    }
  }
}
```

```

    )

data_final <- tibble(do.call(rbind.data.frame, data))

# Write tibble to csv
data_final %>% write_csv("../data/of_catch_prob_2024.csv")

# Load in data
data_final <- read_csv("../data/of_catch_prob_2024.csv")

```

January 2nd

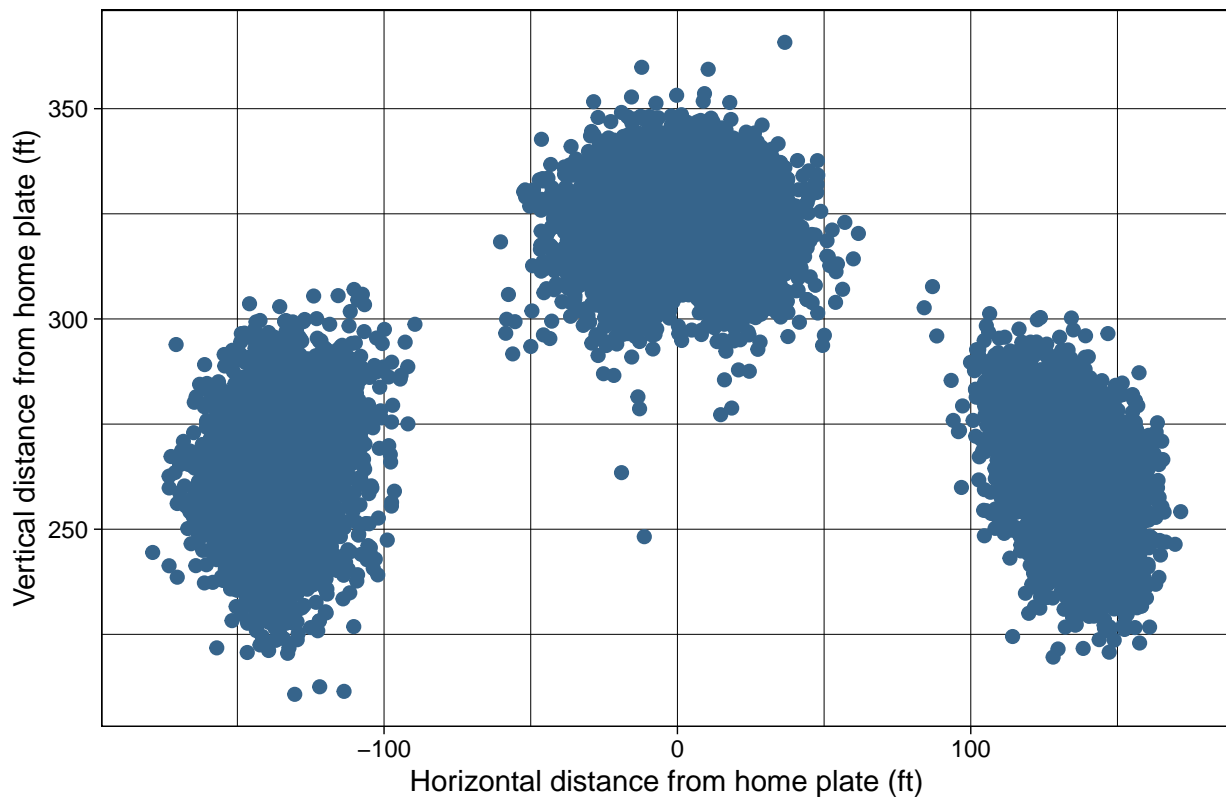
```

jan2_1 <- data_final %>%
  ggplot(aes(x = start_pos_x, y = start_pos_y)) +
  geom_point(col = "steelblue4", size = 2, shape = 19) +
  labs(title = "Starting positions on all balls hit to qualified OFs in 2024",
       x = "Horizontal distance from home plate (ft)",
       y = "Vertical distance from home plate (ft)") +
  theme_linedraw()

jan2_1

```

Starting positions on all balls hit to qualified OFs in 2024

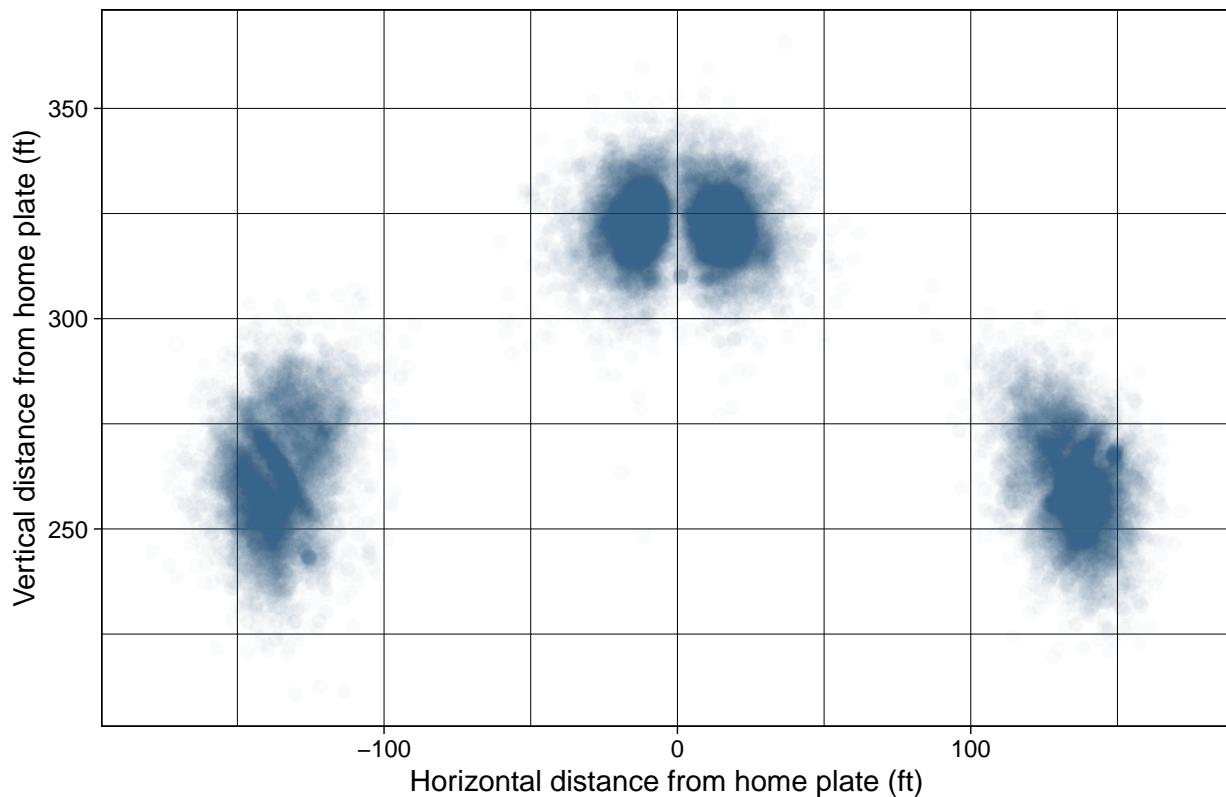


```
ggsave("../figures/jan2_1.png", plot = jan2_1)
```

```
jan2_2 <- data_final %>%
  ggplot(aes(x = start_pos_x, y = start_pos_y)) +
  geom_point(col = "steelblue4", size = 2, shape = 19, alpha = 0.025) +
  labs(title = "Starting positions on all balls hit to qualified OFs in 2024",
       x = "Horizontal distance from home plate (ft)",
       y = "Vertical distance from home plate (ft)") +
  theme_linedraw()
```

```
jan2_2
```

Starting positions on all balls hit to qualified OFs in 2024

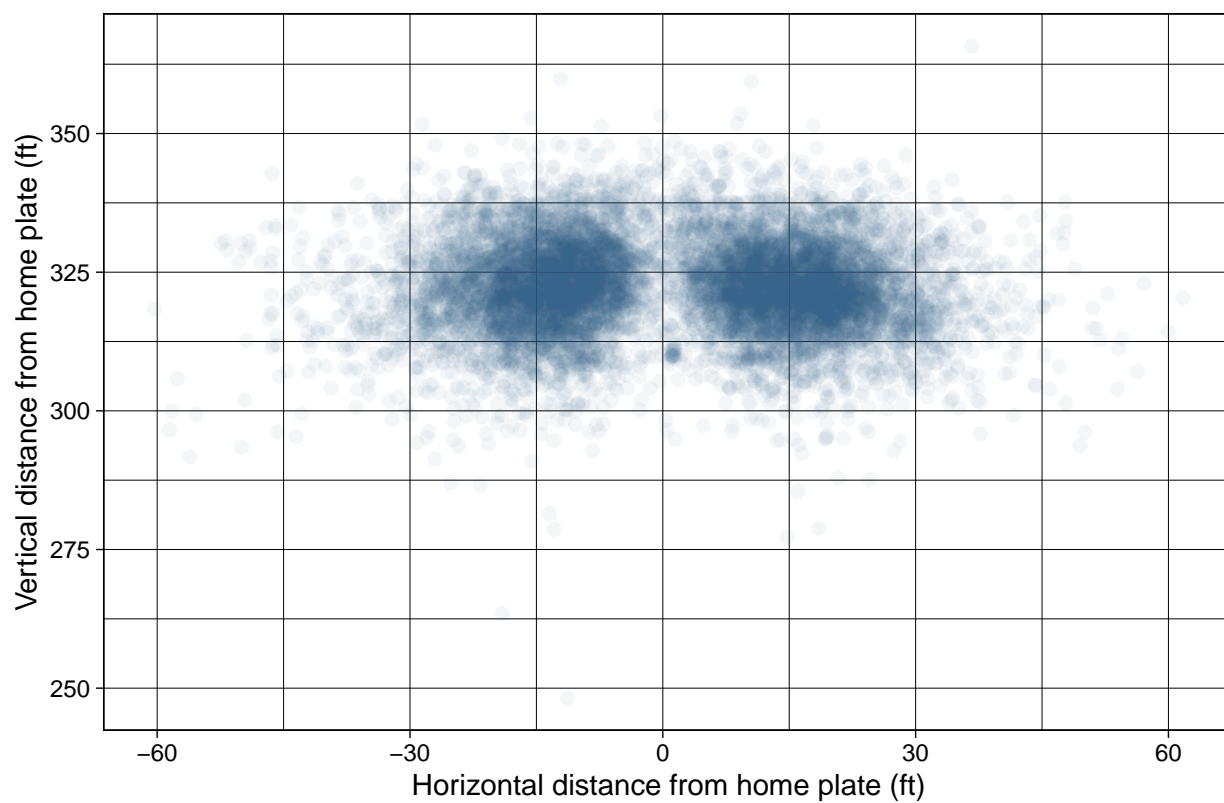


```
ggsave("../figures/jan2_2.png", plot = jan2_2)
```

```
jan2_3 <- data_final %>%
  filter(pos == 8) %>%
  ggplot(aes(x = start_pos_x, y = start_pos_y)) +
  geom_point(col = "steelblue4", size = 2, shape = 19, alpha = 0.05) +
  labs(title = "Starting positions on all balls hit to qualified CFs in 2024",
       x = "Horizontal distance from home plate (ft)",
       y = "Vertical distance from home plate (ft)") +
  theme_linedraw()
```

```
jan2_3
```

Starting positions on all balls hit to qualified CFs in 2024



```
ggsave("../figures/jan2_3.png", plot = jan2_3)
```