

Catch probability EDA

Tim White

2025-02-01

```
library(tidyverse)
library(rjson)
library(baseballr)
library(rvest)
```

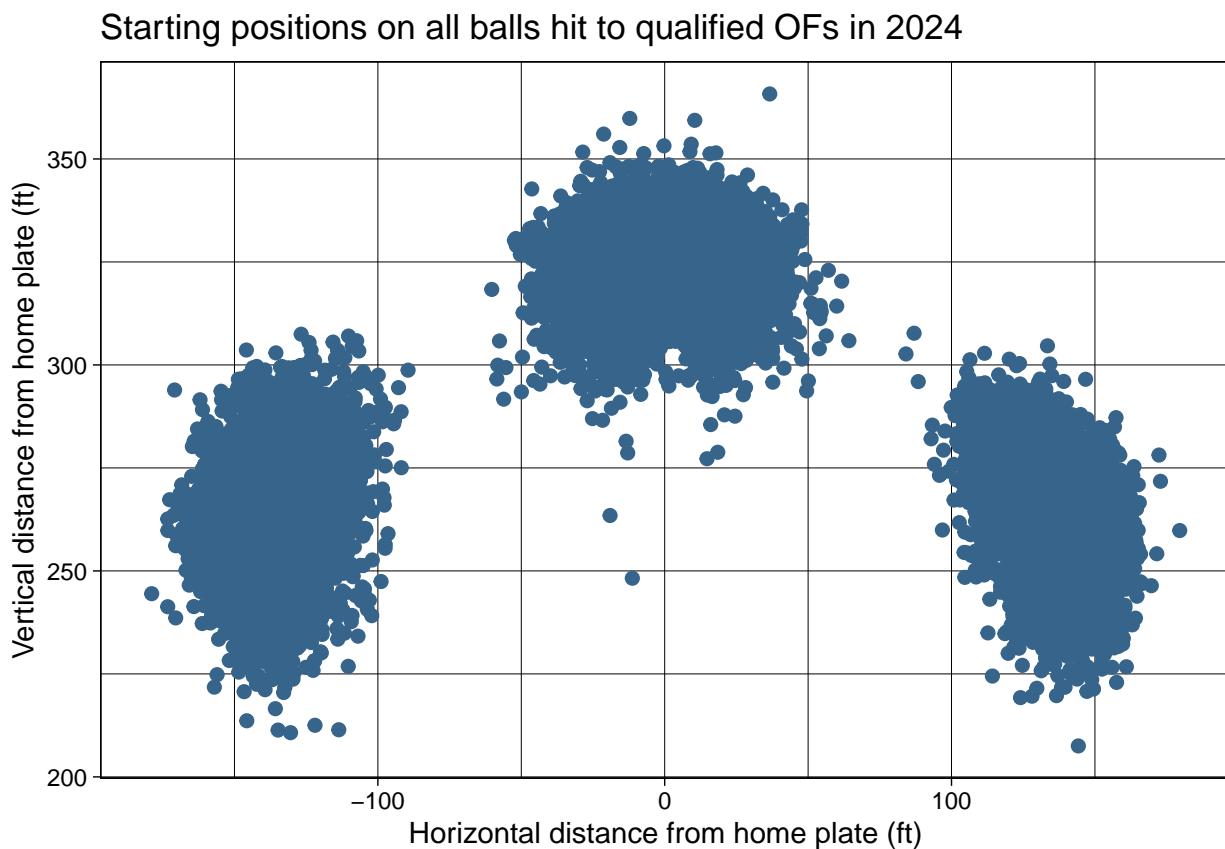
Load in data

```
of_catch_prob_2024 <- read_csv("../data/of_catch_prob_2024.csv") %>%  
  mutate(player = as.factor(player),  
         team = as.factor(team),  
         game_pk = as.factor(game_pk),  
         play_id = as.factor(play_id),  
         player_id = as.factor(player_id),  
         pos = as.factor(pos),  
         route_direction = as.factor(route_direction),  
         stadium = as.factor(stadium),  
         closest_wall = as.factor(closest_wall))  
  
of_catch_prob_2023 <- read_csv("../data/of_catch_prob_2023.csv") %>%  
  mutate(player = as.factor(player),  
         team = as.factor(team),  
         game_pk = as.factor(game_pk),  
         play_id = as.factor(play_id),  
         player_id = as.factor(player_id),  
         pos = as.factor(pos),  
         route_direction = as.factor(route_direction),  
         stadium = as.factor(stadium),  
         closest_wall = as.factor(closest_wall))
```

January 2nd

```
jan2_1 <- of_catch_prob_2024 %>%
  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
```

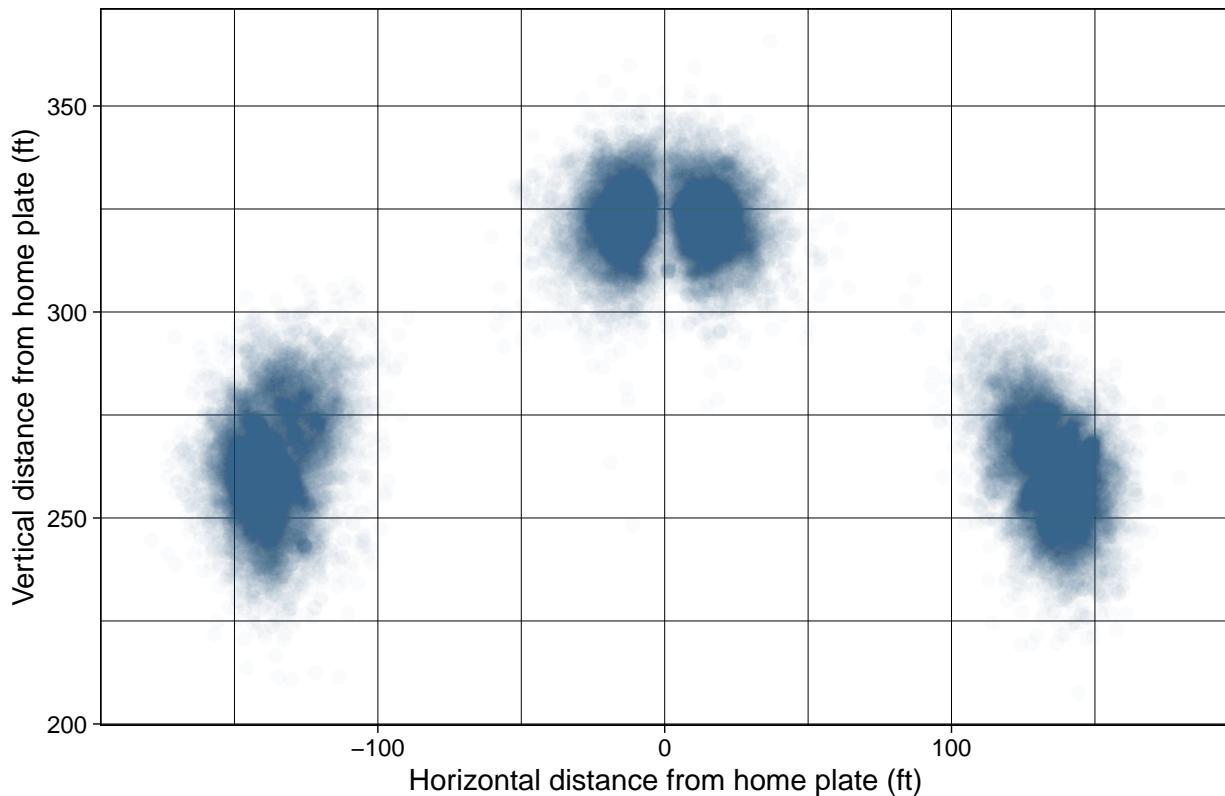


```
ggsave("../figures/jan2_1.png", plot = jan2_1)
```

```
jan2_2 <- of_catch_prob_2024 %>%
  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

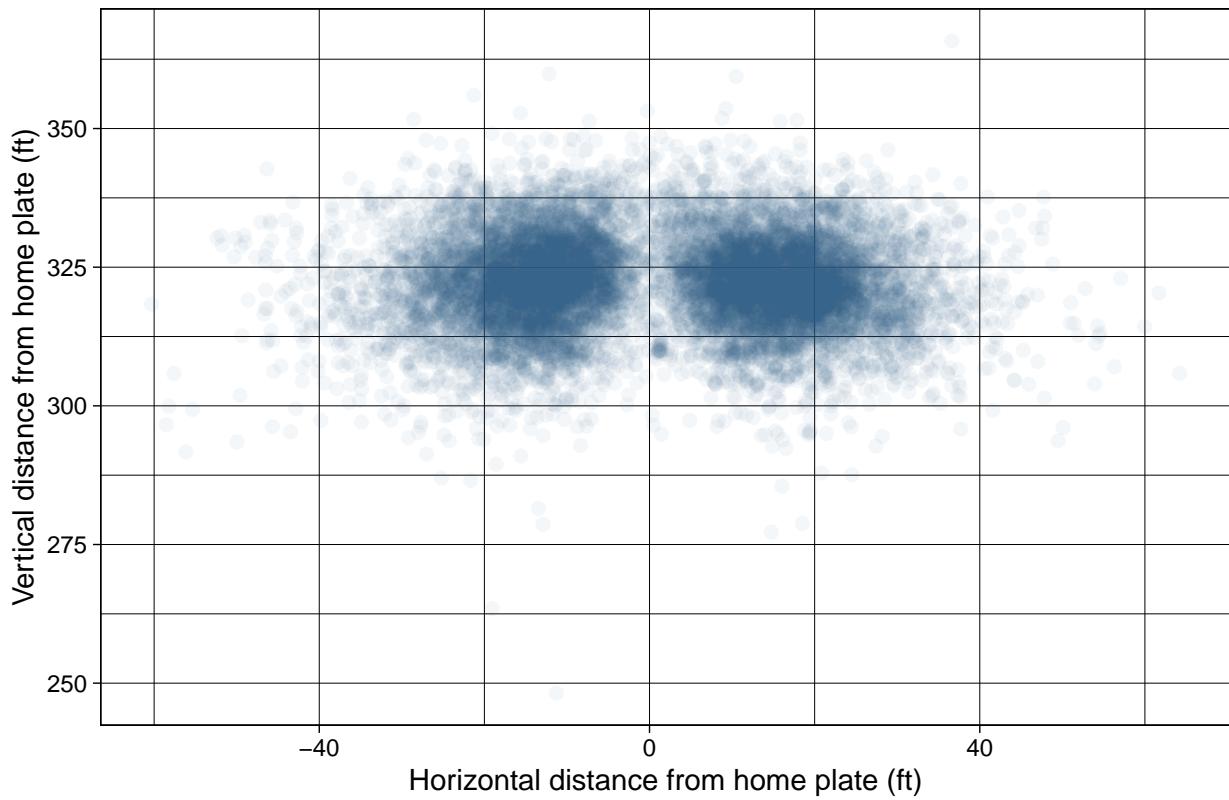


```
ggsave("../figures/jan2_2.png", plot = jan2_2)
```

```
jan2_3 <- of_catch_prob_2024 %>%
  filter(pos == "CF") %>%
  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)
```

January 4th

```
single_game_oaa <- of_catch_prob_2024 %>%
  group_by(game_pk, player) %>%
  summarize(oaa = sum(out * (1 - catch_rate) - (1 - out) * catch_rate),
            opportunities = n(),
            catches = sum(out),
            stars5_opps = sum(stars == 5),
            stars5_catches = sum(stars == 5 & out == 1),
            stars4_opps = sum(stars == 4),
            stars4_catches = sum(stars == 4 & out == 1),
            stars3_opps = sum(stars == 3),
            stars3_catches = sum(stars == 3 & out == 1),
            stars2_opps = sum(stars == 2),
            stars2_catches = sum(stars == 2 & out == 1),
            stars1_opps = sum(stars == 1),
            stars1_catches = sum(stars == 1 & out == 1),
            stars0_opps = sum(stars == 0),
            stars0_catches = sum(stars == 0 & out == 1),
            .groups = "drop")
```

```
single_game_oaa %>%
  arrange(desc(oaa)) %>%
  head(10)
```

```
## # A tibble: 10 x 17
##   game_pk player          oaa opportunities catches stars5_opps stars5_catches
##   <fct>    <fct>      <dbl>        <int>     <dbl>        <int>        <int>
## 1 746097  Pete Crow-Arm~  1.68         6       6           2           2
## 2 745277  Julio Rodríg~  1.61         4       4           1           1
## 3 744844  Jacob Young    1.39         6       6           1           1
## 4 745716  Tyrone Taylor  1.31         3       3           1           1
## 5 745075  Pete Crow-Arm~  1.3         2       2           1           1
## 6 745498  Michael A. Ta~  1.27         4       4           1           1
## 7 746101  Pete Crow-Arm~  1.27         5       4           1           1
## 8 744997  Garrett Hamps~  1.25         3       3           0           0
## 9 746030  Vidal Bruján  1.25         2       2           1           1
## 10 745270  Julio Rodríg~ 1.22         5       5           1           1
## # i 10 more variables: stars4_opps <int>, stars4_catches <int>,
## #   stars3_opps <int>, stars3_catches <int>, stars2_opps <int>,
## #   stars2_catches <int>, stars1_opps <int>, stars1_catches <int>,
## #   stars0_opps <int>, stars0_catches <int>
```

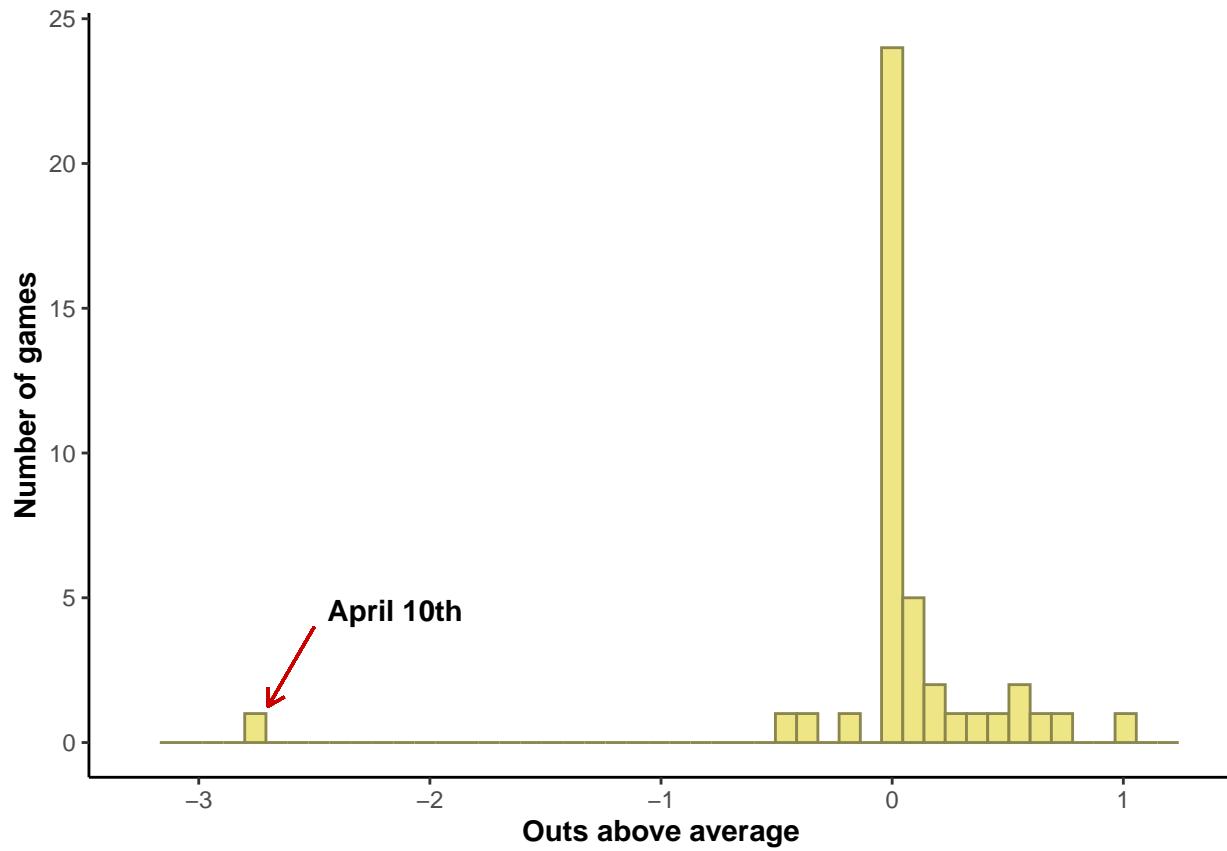
January 6th

```
single_game_oaa %>%
  arrange(oaa) %>%
  head(10)

## # A tibble: 10 x 17
##   game_pk player      oaa opportunities catches stars5_opps stars5_catches
##   <fct>    <fct>    <dbl>        <int>    <dbl>        <int>        <int>
## 1 745196  Victor Scott ~ -2.71         6     3          0          0
## 2 746262  MJ Melendez -2.57         6     2          0          0
## 3 746546  Charlie Black~ -2.34         5     1          1          0
## 4 746931  Tyler O'Neill -2.33         5     2          0          0
## 5 746677  Will Benson -2.11         6     3          1          0
## 6 745342  Luis Matos -2.07         7     2          3          0
## 7 746971  Ian Happ -2.04         4     1          0          0
## 8 745300  Ramón Laureano -1.98         2     0          0          0
## 9 746546  Jack Suwinski -1.97         6     2          1          0
## 10 746525  Cedianne Rafa~ -1.94        5     2          1          0
## # i 10 more variables: stars4_opps <int>, stars4_catches <int>,
## #   stars3_opps <int>, stars3_catches <int>, stars2_opps <int>,
## #   stars2_catches <int>, stars1_opps <int>, stars1_catches <int>,
## #   stars0_opps <int>, stars0_catches <int>

jan6 <- single_game_oaa %>%
  filter(player == "Victor Scott II") %>%
  ggplot() +
  geom_histogram(aes(x = oaa), bins = 50, fill = "khaki2", col = "khaki4") +
  geom_segment(x = -2.5, y = 4, xend = -2.7, yend = 1.25,
               arrow = arrow(length = unit(0.25, "cm")),
               col = "red3") +
  geom_text(x = -2.15, y = 4.55, label = "April 10th", check_overlap = TRUE, fontface = "bold")
  labs(x = "Outs above average", y = "Number of games") +
  xlim(c(-3.25, 1.25)) +
  theme_classic() + theme(axis.title = element_text(face = "bold"))

jan6
```



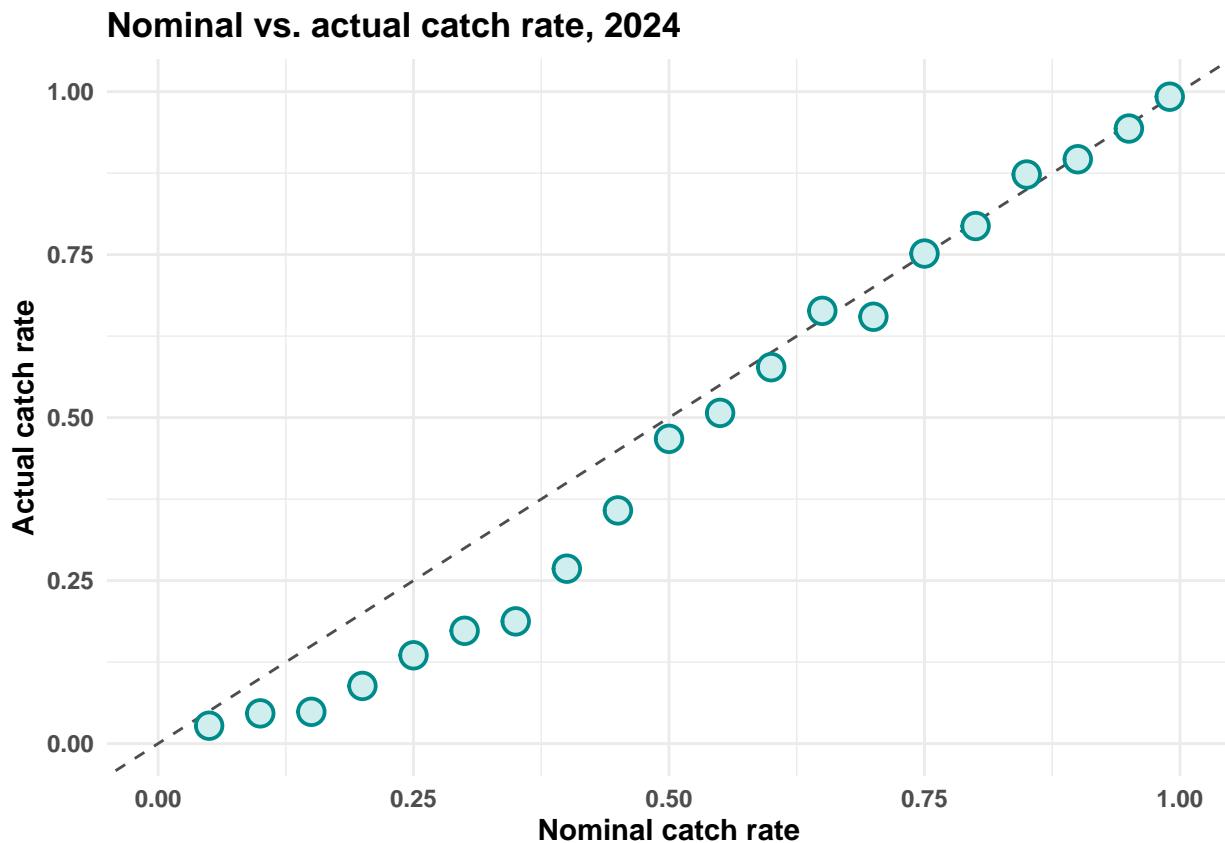
```
ggsave("../figures/jan6.png", plot = jan6, height = 4, width = 6)
```

January 7th

```
calibration_table_2024 <- of_catch_prob_2024 %>%
  group_by(catch_rate) %>%
  summarize(actual_catch_rate = mean(out)) %>%
  rename(nominal_catch_rate = catch_rate)

jan7_1 <- calibration_table_2024 %>%
  ggplot(aes(x = nominal_catch_rate, y = actual_catch_rate)) +
  geom_abline(intercept = 0, slope = 1,
              color = "gray30", linetype = "dashed", linewidth = 0.5) +
  geom_point(pch = 21, col = "cyan4", fill = "lightcyan2",
             size = 4, stroke = 1) +
  theme_minimal() +
  lims(x = c(0,1), y = c(0,1)) +
  labs(x = "Nominal catch rate", y = "Actual catch rate",
       title = "Nominal vs. actual catch rate, 2024") +
  theme(axis.title = element_text(face = "bold"),
        axis.text = element_text(face = "bold"),
        title = element_text(face = "bold"))

jan7_1
```



```
ggsave("../figures/jan7_1.png", plot = jan7_1, height = 4, width = 6)
```

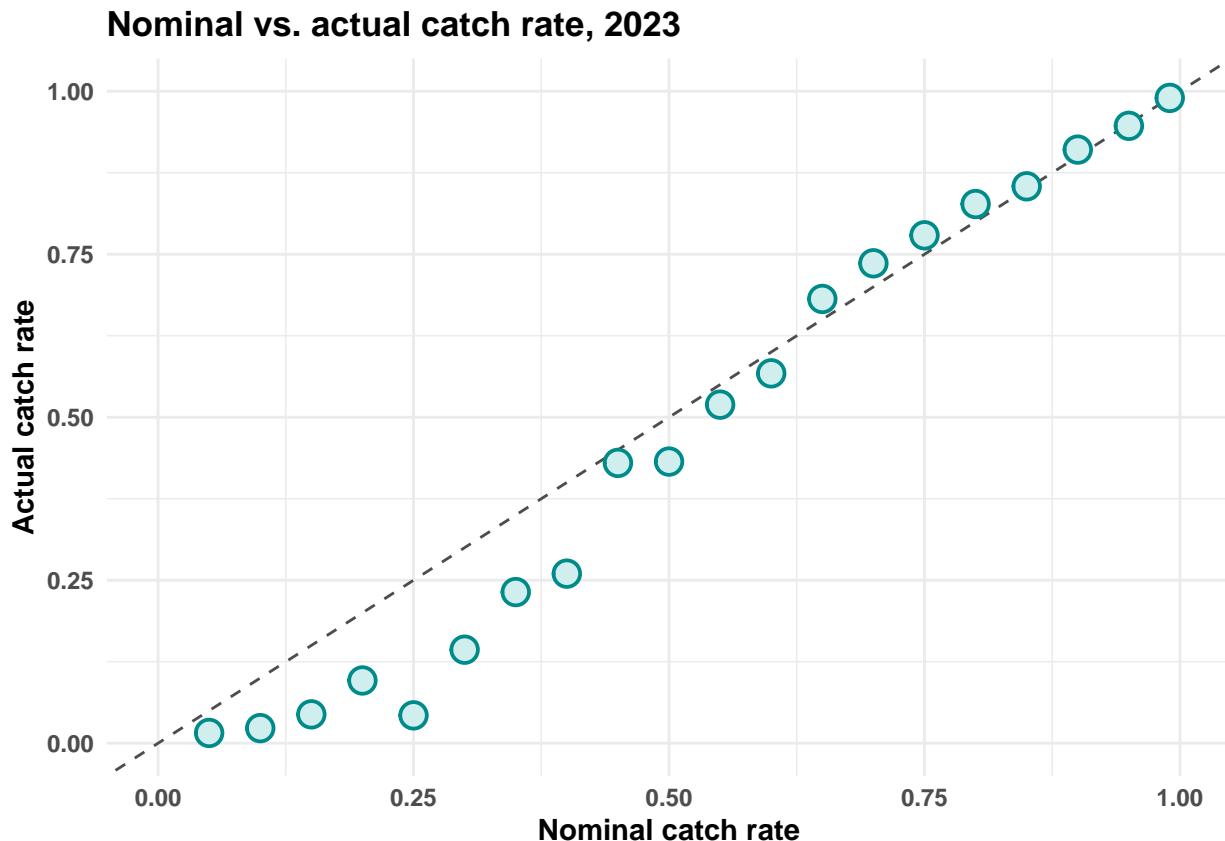
```

calibration_table_2023 <- of_catch_prob_2023 %>%
  group_by(catch_rate) %>%
  summarize(actual_catch_rate = mean(out)) %>%
  rename(nominal_catch_rate = catch_rate)

jan7_2 <- calibration_table_2023 %>%
  ggplot(aes(x = nominal_catch_rate, y = actual_catch_rate)) +
  geom_abline(intercept = 0, slope = 1,
              color = "gray30", linetype = "dashed", linewidth = 0.5) +
  geom_point(pch = 21, col = "cyan4", fill = "lightcyan2",
             size = 4, stroke = 1) +
  theme_minimal() +
  lims(x = c(0,1), y = c(0,1)) +
  labs(x = "Nominal catch rate", y = "Actual catch rate",
       title = "Nominal vs. actual catch rate, 2023") +
  theme(axis.title = element_text(face = "bold"),
        axis.text = element_text(face = "bold"),
        title = element_text(face = "bold"))

```

jan7_2



```
ggsave("../figures/jan7_2.png", plot = jan7_2, height = 4, width = 6)
```

January 8th

```
of_oaa_2024 <- read_csv("../data/of_oaa_2024.csv") %>%
    mutate(player_id = as.factor(player_id))

oaa_rounding_check <- of_catch_prob_2024 %>%
    mutate(catch_rate_upper = pmin(0.999999, catch_rate - 0.02),
          catch_rate_lower = pmin(0.999999, catch_rate + 0.02)) %>%
    group_by(player_id, player) %>%
    summarize(opp = n(),
              oaa_lower = round(sum(out * (1 - catch_rate_lower) - (1 - out) * catch_rate_lower)),
              oaa_estimate = sum(out * (1 - catch_rate) - (1 - out) * catch_rate),
              oaa_upper = round(sum(out * (1 - catch_rate_upper) - (1 - out) * catch_rate_upper)))
    ungroup() %>%
    left_join(of_oaa_2024, by = "player_id") %>%
    select(player_id, name = player, opp,
           oaa_lower, oaa_estimate, oaa_upper, oaa_true = oaa)

oaa_rounding_check %>%
    summarize(any(oaa_true < oaa_lower | oaa_true > oaa_upper))

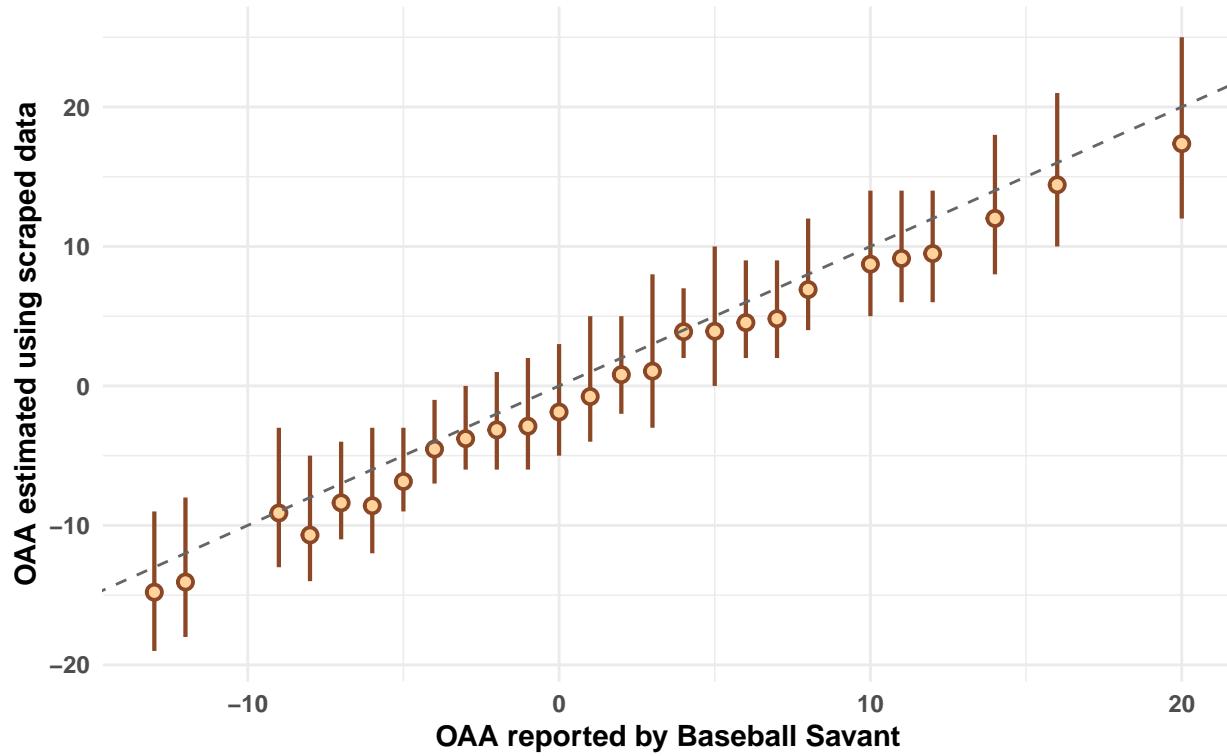
## # A tibble: 1 x 1
##   `any(oaa_true < oaa_lower | oaa_true > oaa_upper)`
##   <lgl>
##   1 FALSE

set.seed(0)
jan8 <- oaa_rounding_check %>%
    filter(opp > 162) %>%
    group_by(oaa_true) %>%
    sample_n(1) %>%
    ggplot(aes(x = oaa_true, y = oaa_estimate)) +
    geom_linerange(aes(ymin = oaa_lower, ymax = oaa_upper), col = "sienna4", linewidth = 0.75) +
    geom_point(size = 2, stroke = 1, pch = 21, fill = "burlywood1", col = "sienna4") +
    geom_abline(intercept = 0, slope = 1, linetype = "dashed", col = "gray40") +
    theme_minimal() +
    labs(x = "OAA reported by Baseball Savant",
         y = "OAA estimated using scraped data",
         title = "Actual vs. estimated OAA for selected players, 2024",
         subtitle = "Lower and upper bounds computed by changing each catch rate by +/-0.02") +
    theme(axis.title = element_text(face = "bold"),
          axis.text = element_text(face = "bold"),
          title = element_text(face = "bold"))

jan8
```

Actual vs. estimated OAA for selected players, 2024

Lower and upper bounds computed by changing each catch rate by +/-0.02

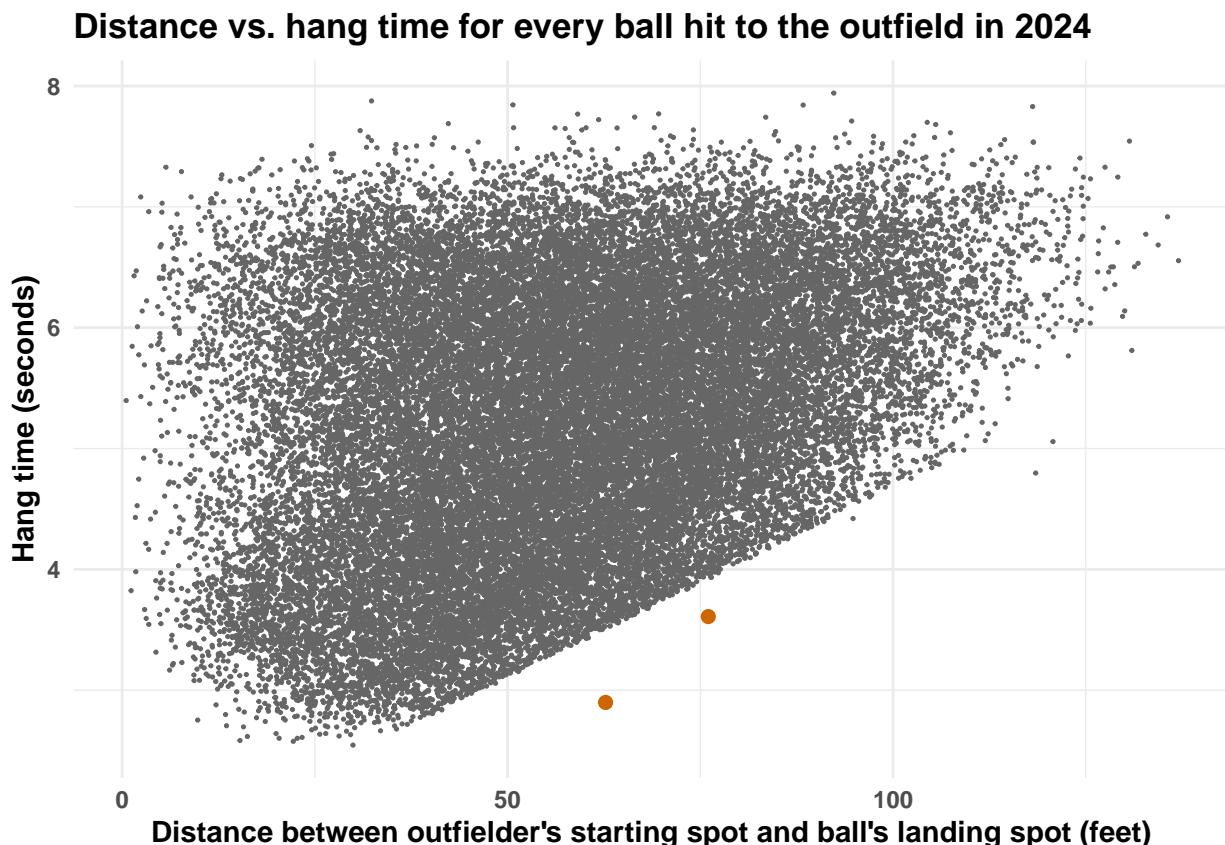


```
ggsave("../figures/jan8.png", plot = jan8, height = 4, width = 6)
```

January 9th

```
jan9 <- of_catch_prob_2024 %>%
  mutate(highlight = (hang_time < 3 & distance > 50) | (hang_time < 3.8 & distance > 75)) %>%
  ggplot(aes(x = distance, y = hang_time, col = highlight, size = highlight)) +
  geom_point() +
  scale_colour_manual(values = c("gray40", "darkorange3")) +
  scale_size_manual(values = c(0.25, 2)) +
  theme_minimal() +
  theme(legend.position = "none",
        axis.title = element_text(face = "bold"),
        axis.text = element_text(face = "bold"),
        title = element_text(face = "bold")) +
  labs(x = "Distance between outfielder's starting spot and ball's landing spot (feet)",
       y = "Hang time (seconds)",
       title = "Distance vs. hang time for every ball hit to the outfield in 2024")
```

jan9



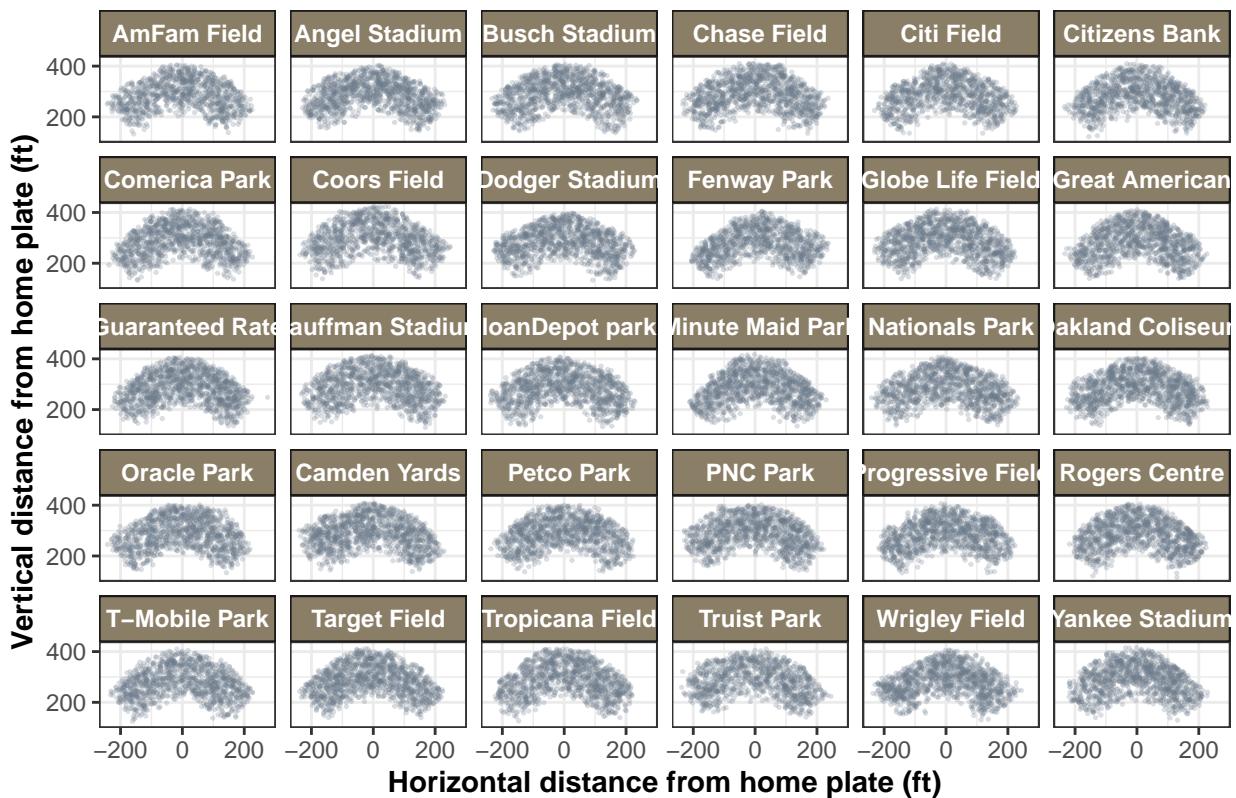
```
ggsave("../figures/jan9.png", plot = jan9, height = 4, width = 6)
```

January 17th

```
jan17_1 <- of_catch_prob_2024 %>%
  mutate(stadium = fct_recode(stadium,
    "AmFam Field" = "American Family Field",
    "Guaranteed Rate" = "Guaranteed Rate Field",
    "Camden Yards" = "Oriole Park at Camden Yards",
    "Great American" = "Great American Ball Park",
    "Citizens Bank" = "Citizens Bank Park"
  )) %>%
  ggplot(aes(x = landing_pos_x, y = landing_pos_y)) +
  geom_point(size = 0.25, alpha = 0.25, col = "slategray4") +
  facet_wrap(~stadium, scales = "fixed") +
  theme_bw() +
  labs(x = "Horizontal distance from home plate (ft)",
       y = "Vertical distance from home plate (ft)",
       title = "Landing positions of fly balls by stadium, 2024") +
  theme(axis.title = element_text(face = "bold"),
        title = element_text(face = "bold"),
        plot.subtitle = element_text(face = "italic"),
        strip.background = element_rect(fill = "wheat4", color = "gray10"),
        strip.text = element_text(face = "bold", color = "white")) +
  scale_x_continuous(breaks = c(-200, 0, 200)) +
  scale_y_continuous(breaks = c(200, 400))
```

```
jan17_1
```

Landing positions of fly balls by stadium, 2024

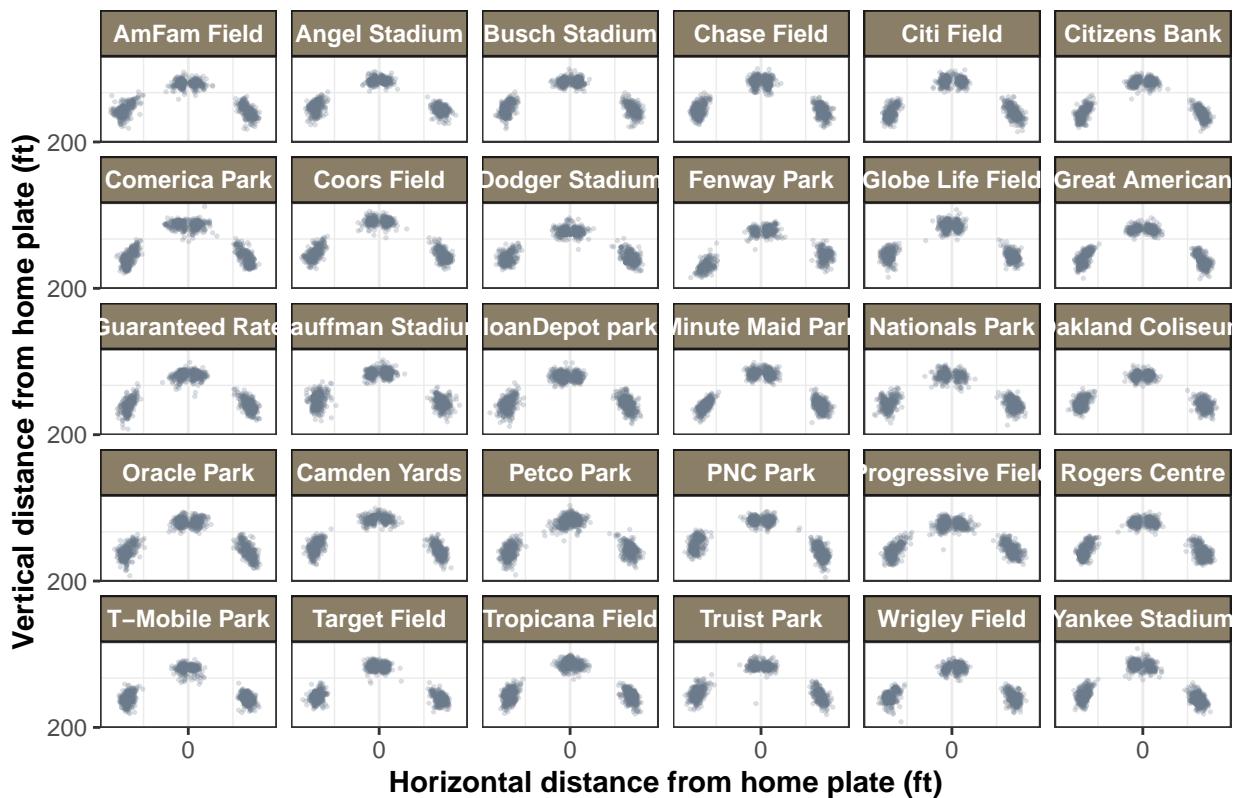


```
ggsave("../figures/jan17_1.png", plot = jan17_1, height = 6, width = 8)
```

```
jan17_2 <- of_catch_prob_2024 %>%
  mutate(stadium = fct_recode(stadium,
    "AmFam Field" = "American Family Field",
    "Guaranteed Rate" = "Guaranteed Rate Field",
    "Camden Yards" = "Oriole Park at Camden Yards",
    "Great American" = "Great American Ball Park",
    "Citizens Bank" = "Citizens Bank Park"
  )) %>%
  ggplot(aes(x = start_pos_x, y = start_pos_y)) +
  geom_point(size = 0.25, alpha = 0.25, col = "slategray4") +
  facet_wrap(~stadium, scales = "fixed") +
  theme_bw() +
  labs(x = "Horizontal distance from home plate (ft)",
       y = "Vertical distance from home plate (ft)",
       title = "Starting positions of outfielders by stadium, 2024") +
  theme(axis.title = element_text(face = "bold"),
        title = element_text(face = "bold"),
        plot.subtitle = element_text(face = "italic"),
        strip.background = element_rect(fill = "wheat4", color = "gray10"),
        strip.text = element_text(face = "bold", color = "white")) +
  scale_x_continuous(breaks = c(-200, 0, 200)) +
  scale_y_continuous(breaks = c(200, 400))
```

jan17_2

Starting positions of outfielders by stadium, 2024



```
ggsave("../figures/jan17_2.png", plot = jan17_2, height = 6, width = 8)
```

January 18th

```
stadium_dims <- of_catch_prob_2024 %>%
  distinct(stadium, .keep_all = TRUE) %>%
  select(stadium,
         wall_height_lf, wall_height_cf, wall_height_rf,
         wall_dist_lf, wall_dist_lc, wall_dist_cf, wall_dist_rc, wall_dist_rf)

pca <- prcomp(stadium_dims %>% select(-stadium), center = TRUE, scale = TRUE)

pca

## Standard deviations (1, ..., p=8):
## [1] 1.6133522 1.2413816 1.0530766 1.0052921 0.9336312 0.7679162 0.4447214
## [8] 0.2781085
##
## Rotation (n x k) = (8 x 8):
##          PC1      PC2      PC3      PC4      PC5
## wall_height_lf -0.52563253 -0.16252938 0.01261898 -0.2306161 0.31539308
## wall_height_cf -0.21369818 -0.46724246 0.45415128 -0.2028672 -0.24556732
## wall_height_rf  0.28717525  0.28246940 0.35896996 0.1042763 0.67175299
## wall_dist_lf    0.48507818 -0.28088096 0.01297260 -0.3375695 0.32213101
## wall_dist_lc   0.46879974  0.06248463 0.25797717 0.2568213 -0.47910240
## wall_dist_cf   -0.01629932 -0.35294451 -0.50995086 0.6384369 0.17928122
## wall_dist_rc   0.04235743 -0.56952669 0.41248946 0.3622816 0.14825164
## wall_dist_rf   0.37207533 -0.37914483 -0.40964618 -0.4195565 -0.03929392
##          PC6      PC7      PC8
## wall_height_lf 0.00962424 0.69004304 -0.26120262
## wall_height_cf 0.59259431 -0.26984609 -0.06573315
## wall_height_rf 0.43090688 0.08794958 0.23095562
## wall_dist_lf  -0.19532159 -0.19154215 -0.62695794
## wall_dist_lc   0.14069505 0.57057383 -0.26253451
## wall_dist_cf   0.35684595 -0.04124903 -0.21524573
## wall_dist_rc   -0.49849021 0.06888202 0.31182355
## wall_dist_rf   0.17111831 0.27312902 0.51821479

summary(pca)

## Importance of components:
##          PC1      PC2      PC3      PC4      PC5      PC6      PC7
## Standard deviation 1.6134 1.2414 1.0531 1.0053 0.9336 0.76792 0.44472
## Proportion of Variance 0.3254 0.1926 0.1386 0.1263 0.1090 0.07371 0.02472
## Cumulative Proportion 0.3254 0.5180 0.6566 0.7829 0.8919 0.96561 0.99033
##          PC8
## Standard deviation 0.27811
## Proportion of Variance 0.00967
## Cumulative Proportion 1.00000

set.seed(18)

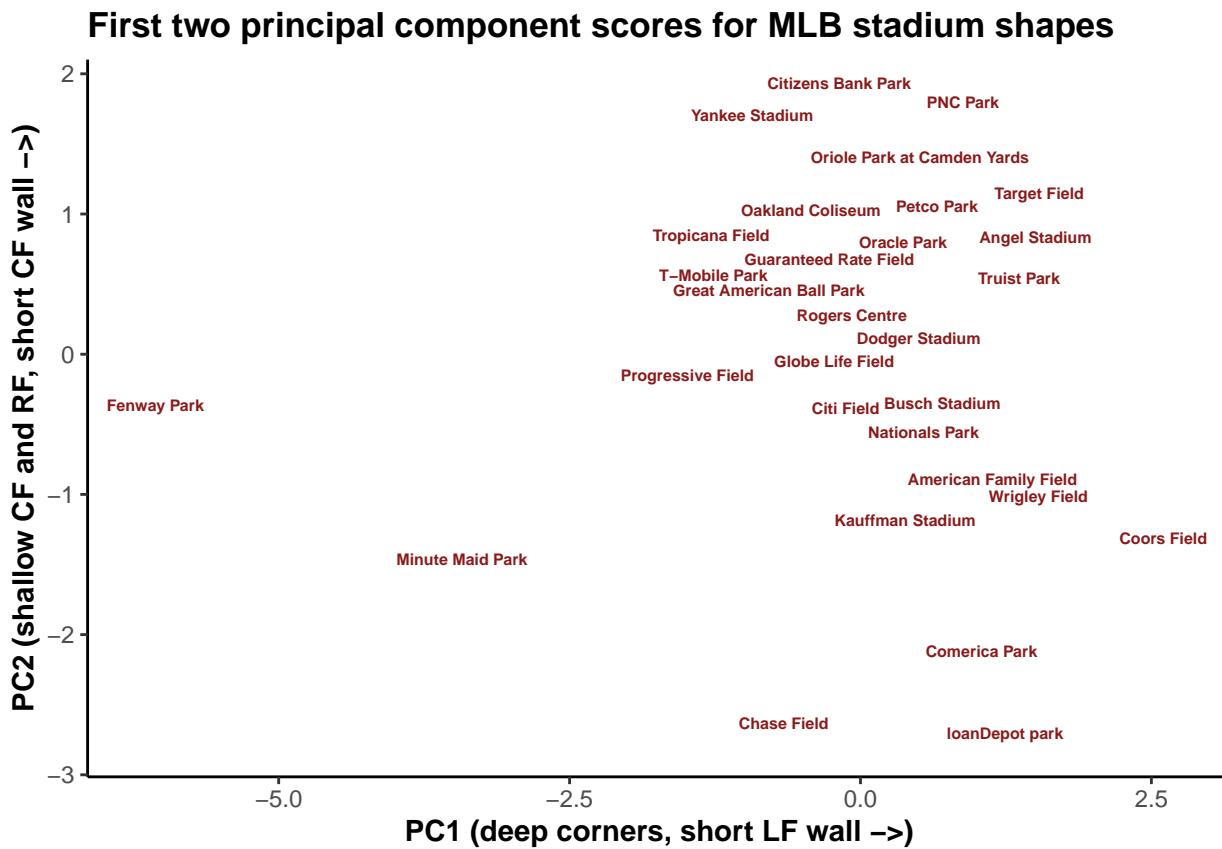
jan18 <- tibble(stadium = stadium_dims$stadium,
```

```

pc1 = pca$x[,1], pc2 = pca$x[,2], pc3 = pca$x[,3]) %>%
ggplot(aes(x = pc1, y = pc2)) +
  ggrepel::geom_text_repel(aes(label = stadium), fontface = "bold", col = "firebrick4",
                           min.segment.length = 1,
                           label.padding = 0.0, box.padding = 0.05, point.padding = 0,
                           size = 2, max.overlaps = 30) +
  theme_classic() +
  labs(title = "First two principal component scores for MLB stadium shapes",
       x = "PC1 (deep corners, short LF wall ->)",
       y = "PC2 (shallow CF and RF, short CF wall ->)") +
  theme(axis.title = element_text(face = "bold"),
        title = element_text(face = "bold"))

```

jan18



```
ggsave("../figures/jan18.png", plot = jan18, height = 4, width = 6)
```

January 21st

```
set.seed(121)

play_ids <- ofCatchProb_2024 %>%
  filter(distance > 60) %>%
  filter(route_angle %% 60 > 20) %>%
  group_by(route_direction) %>% sample_n(1) %>% pull(play_id)

jan21 <- ofCatchProb_2024 %>%
  filter(play_id %in% play_ids) %>%
  mutate(route_angle = paste0("route angle = ", round(route_angle), " degrees")) %>%
  ggplot() +
  geom_segment(aes(x = start_pos_x, y = start_pos_y,
                    xend = landing_pos_x, yend = landing_pos_y),
               arrow = arrow(length = unit(0.1, "inches"))) +
  geom_segment(aes(x = start_pos_x, y = start_pos_y,
                    xend = 0, yend = 0)) +
  geom_abline(aes(intercept = (start_pos_x^2 + start_pos_y^2) / start_pos_y,
                  slope = -start_pos_x/start_pos_y), color = "gray40", linetype = "dotted") +
  geom_point(x = 0, y = 0, pch = 15) +
  geom_point(aes(x = start_pos_x, y = start_pos_y), col = "steelblue4", size = 3, pch = 17) +
  geom_point(aes(x = landing_pos_x, y = landing_pos_y), color = "firebrick3", size = 3) +
  facet_wrap(~route_angle) +
  labs(x = "Horizontal distance from home plate (ft)",
       y = "Vertical distance from home plate (ft)") +
  xlim(c(-200, 200)) + ylim(c(0, 400)) + coord_fixed() +
  theme_bw()

ggsave("../figures/jan21.png", plot = jan21, height = 4, width = 6)
```

January 22nd

```
of_catch_prob_2024 %>%
  count(back_bs = as.factor(back),
        back_mine = as.factor(ifelse(
          route_direction == "back_middle",
          1, 0)))
```

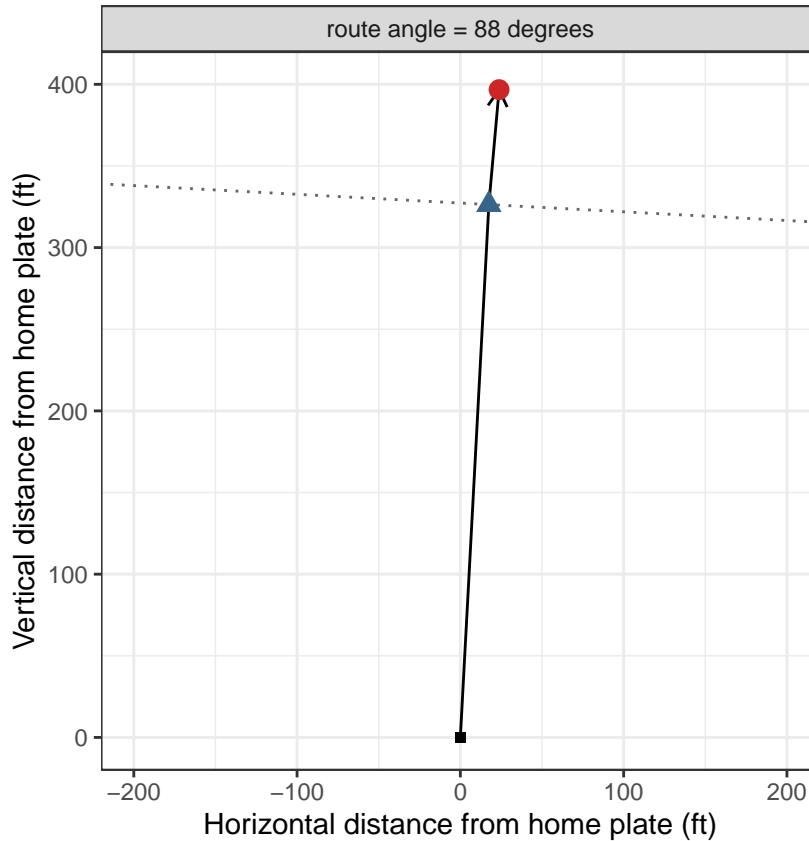


```
## # A tibble: 4 x 3
##   back_bs back_mine     n
##   <fct>    <fct>   <int>
## 1 0         0         27713
## 2 0         1          1557
## 3 1         0          1966
## 4 1         1          3034
```

January 23rd

```
jan23 <- of_catch_prob_2024 %>%
  filter(play_id == "925eb5a8-b6ff-42ca-ab2e-3ac950c76259") %>%
  mutate(route_angle = paste0("route angle = ", round(route_angle), " degrees")) %>%
  ggplot() +
  geom_segment(aes(x = start_pos_x, y = start_pos_y,
                    xend = landing_pos_x, yend = landing_pos_y),
               arrow = arrow(length = unit(0.1, "inches"))) +
  geom_segment(aes(x = start_pos_x, y = start_pos_y,
                    xend = 0, yend = 0)) +
  geom_abline(aes(intercept = (start_pos_x^2 + start_pos_y^2) / start_pos_y,
                  slope = -start_pos_x/start_pos_y), color = "gray40", linetype = "dotted") +
  geom_point(x = 0, y = 0, pch = 15) +
  geom_point(aes(x = start_pos_x, y = start_pos_y), col = "steelblue4", size = 3, pch = 17) +
  geom_point(aes(x = landing_pos_x, y = landing_pos_y), color = "firebrick3", size = 3) +
  facet_wrap(~route_angle) +
  labs(x = "Horizontal distance from home plate (ft)",
       y = "Vertical distance from home plate (ft)") +
  xlim(c(-200, 200)) + ylim(c(0, 400)) + coord_fixed() +
  theme_bw()
```

jan23



```
ggsave("../figures/jan23.png", plot = jan23, height = 4, width = 6)
```

January 24th

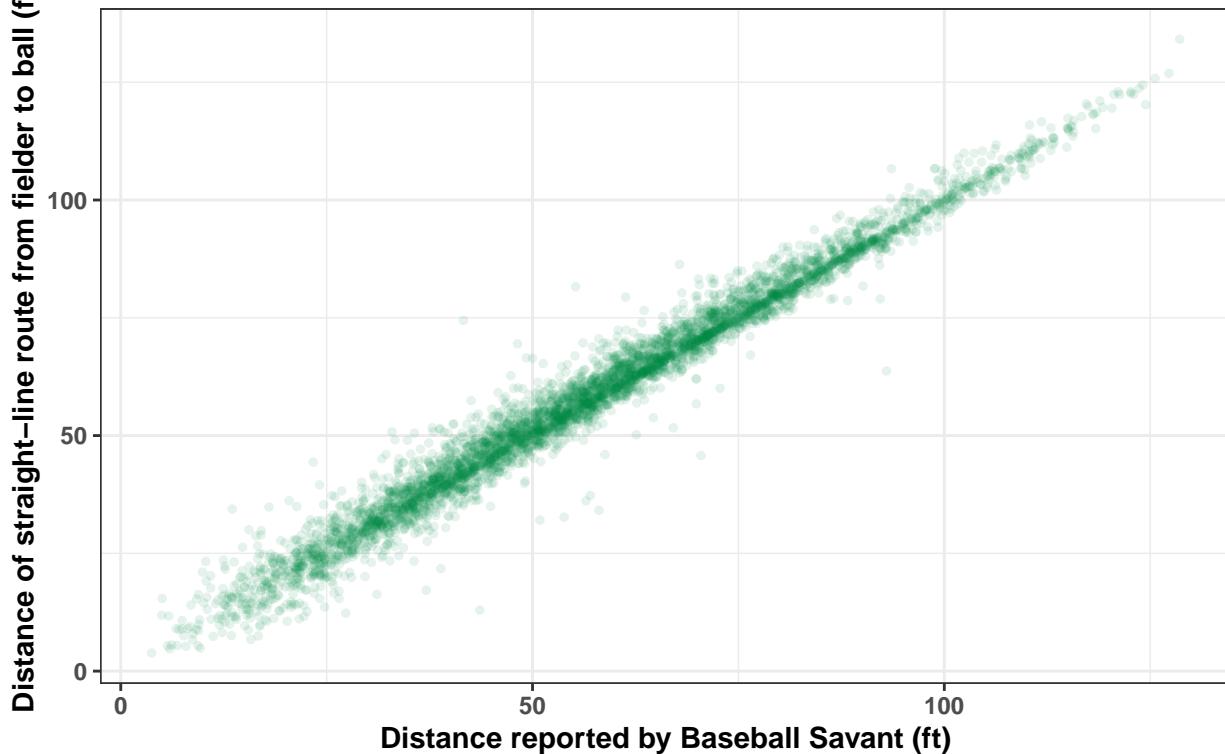
```
set.seed(124)

jan24 <- of_catch_prob_2024 %>%
  sample_n(5000) %>%
  mutate(route_distance = sqrt(
    (landing_pos_x - start_pos_x)^2 + (landing_pos_y - start_pos_y)^2)
  ) %>%
  ggplot(aes(x = distance, y = route_distance)) +
  geom_point(size = 1, alpha = 0.1, color = "springgreen4") +
  labs(x = "Distance reported by Baseball Savant (ft)",
       y = "Distance of straight-line route from fielder to ball (ft)",
       title = "Statcast distances vs. straight-line route distances",
       subtitle = "5,000 randomly sampled fly balls from the 2024 season") +
  theme_bw() +
  theme(axis.title = element_text(face = "bold"),
        axis.text = element_text(face = "bold"),
        title = element_text(face = "bold"))

jan24
```

Statcast distances vs. straight-line route distances

5,000 randomly sampled fly balls from the 2024 season



```
ggsave("../figures/jan24.png", plot = jan24, height = 4, width = 6)
```

January 25th

```
jan25_data <- of_catch_prob_2024 %>%
  mutate(route_distance = sqrt(
    (landing_pos_x - start_pos_x)^2 + (landing_pos_y - start_pos_y)^2)) %>%
  select(player, team, game_pk, play_id,
         distance, route_distance, route_angle, ball_distance, everything())

jan25_id1 <- jan25_data %>%
  filter(abs(distance) < 5) %>%
  arrange(desc(route_distance)) %>%
  pull(play_id) %>% head(1)

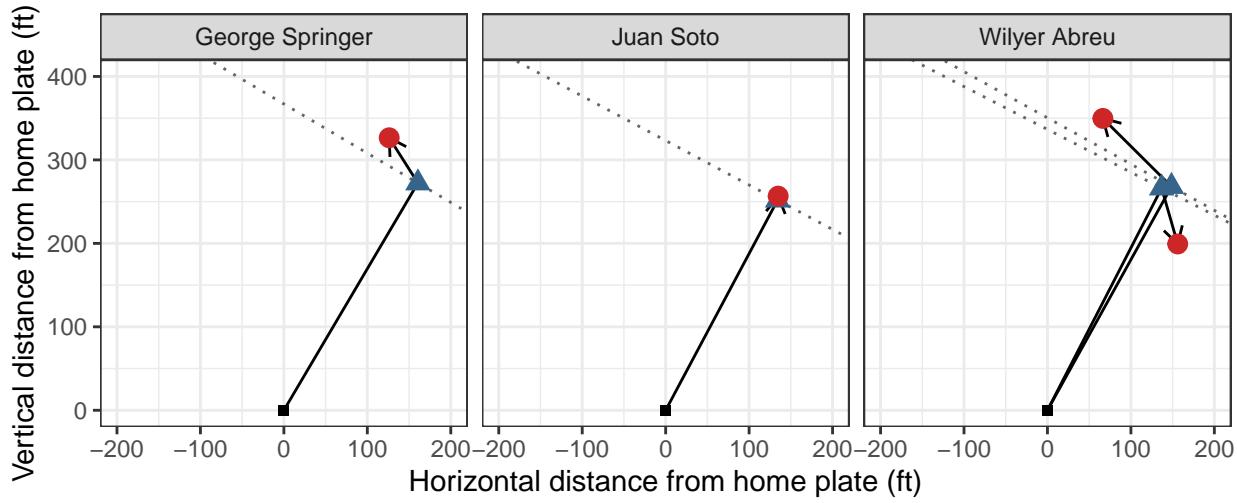
jan25_id2 <- jan25_data %>%
  filter(abs(route_distance) < 5) %>%
  arrange(desc(distance)) %>%
  head(1) %>% pull(play_id)

jan25_id3 <- jan25_data %>%
  arrange(desc((route_distance - distance))) %>%
  head(1) %>% pull(play_id)

jan25_id4 <- jan25_data %>%
  arrange(desc((distance - route_distance))) %>%
  head(1) %>% pull(play_id)
```

```
jan25_1 <- of_catch_prob_2024 %>%
  filter(play_id %in% c(jan25_id1, jan25_id2, jan25_id3, jan25_id4)) %>%
  ggplot() +
  geom_segment(aes(x = start_pos_x, y = start_pos_y,
                  xend = landing_pos_x, yend = landing_pos_y),
               arrow = arrow(length = unit(0.1, "inches"))) +
  geom_segment(aes(x = start_pos_x, y = start_pos_y,
                  xend = 0, yend = 0)) +
  geom_abline(aes(intercept = (start_pos_x^2 + start_pos_y^2) / start_pos_y,
                  slope = -start_pos_x/start_pos_y), color = "gray40", linetype = "dotted") +
  geom_point(x = 0, y = 0, pch = 15) +
  geom_point(aes(x = start_pos_x, y = start_pos_y), col = "steelblue4", size = 3, pch = 17) +
  geom_point(aes(x = landing_pos_x, y = landing_pos_y), color = "firebrick3", size = 3) +
  facet_wrap(~player) +
  labs(x = "Horizontal distance from home plate (ft)",
       y = "Vertical distance from home plate (ft)") +
  xlim(c(-200, 200)) + ylim(c(0, 400)) + coord_fixed() +
  theme_bw()

jan25_1
```



```
ggsave("../figures/jan25_1.png", plot = jan25_1, height = 3, width = 6)
```

```
jan25_data %>%
  mutate(d = abs(route_distance - distance)) %>%
  summarize(mean(d > 5))
```

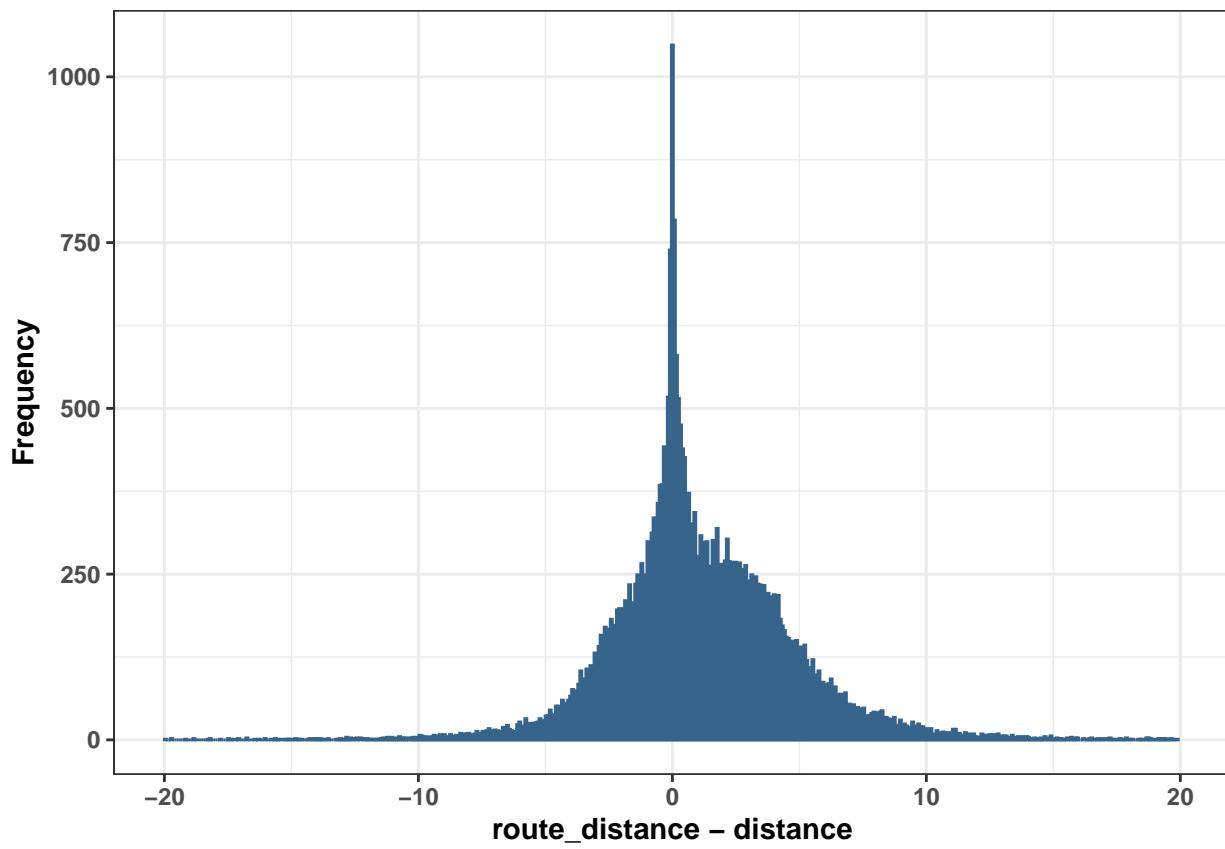
```
## # A tibble: 1 x 1
##   `mean(d > 5)`
##       <dbl>
## 1      0.146
```

```
jan25_data %>%
  mutate(d = abs(route_distance - distance)) %>%
  summarize(mean(d > 10))
```

```
## # A tibble: 1 x 1
##   `mean(d > 10)`
##       <dbl>
## 1      0.0218
```

```
jan25_2 <- jan25_data %>%
  ggplot(aes(x = route_distance - distance)) +
  geom_histogram(bins = 500, col = "steelblue4") +
  xlim(c(-20, 20)) +
  labs(x = "route_distance - distance",
       y = "Frequency") +
  theme_bw() +
  theme(axis.title = element_text(face = "bold"),
        axis.text = element_text(face = "bold"),
        title = element_text(face = "bold"))
```

```
jan25_2
```



```
ggsave("../figures/jan25_2.png", plot = jan25_2, height = 4, width = 6)
```

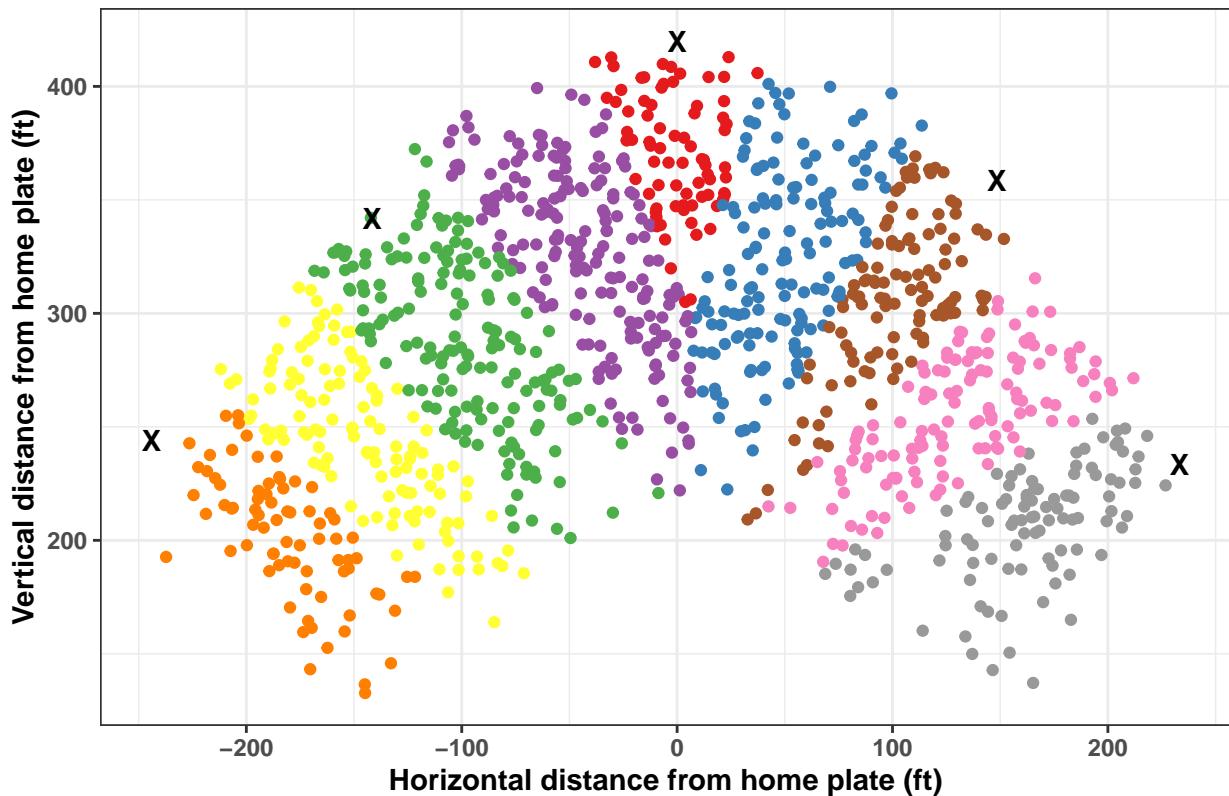
January 26th

```
copa2024 <- of_catch_prob_2024 %>%
  filter(stadium == "Comerica Park")

jan26 <- ggplot() +
  geom_point(data = copa2024, aes(x = landing_pos_x, y = landing_pos_y, col = closest_wall)) +
  annotate("text", x = unique(copa2024$wall_dist_lf_x), y = unique(copa2024$wall_dist_lf_y),
           fontface = "bold", label = "X") +
  annotate("text", x = unique(copa2024$wall_dist_lc_x), y = unique(copa2024$wall_dist_lc_y),
           fontface = "bold", label = "X") +
  annotate("text", x = unique(copa2024$wall_dist_cf_x), y = unique(copa2024$wall_dist_cf_y),
           fontface = "bold", label = "X") +
  annotate("text", x = unique(copa2024$wall_dist_rc_x), y = unique(copa2024$wall_dist_rc_y),
           fontface = "bold", label = "X") +
  annotate("text", x = unique(copa2024$wall_dist_rf_x), y = unique(copa2024$wall_dist_rf_y),
           fontface = "bold", label = "X") +
  theme_bw() +
  theme(legend.position = "none") +
  theme(axis.title = element_text(face = "bold"),
        axis.text = element_text(face = "bold"),
        title = element_text(face = "bold")) +
  labs(title = "Closest wall measurement for all Comerica Park fly balls, 2024",
       x = "Horizontal distance from home plate (ft)",
       y = "Vertical distance from home plate (ft)") +
  scale_color_brewer(palette = "Set1")
```

jan26

Closest wall measurement for all Comerica Park fly balls, 2024



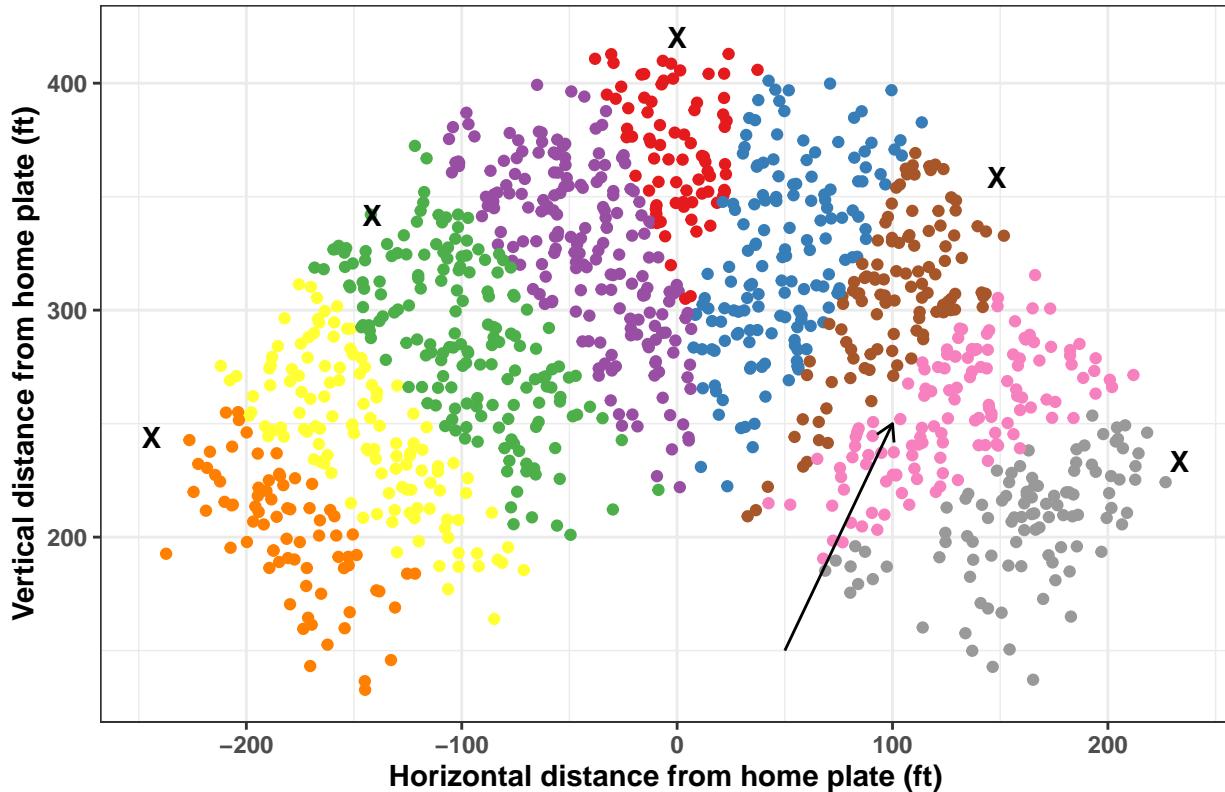
```
ggsave("../figures/jan26.png", plot = jan26, height = 4, width = 6)
```

January 27th

```
jan27 <- jan26 + geom_segment(aes(x = 50, xend = 100, y = 150, yend = 250),  
                           arrow = arrow(length = unit(0.25, "cm")))
```

```
jan27
```

Closest wall measurement for all Comerica Park fly balls, 2024



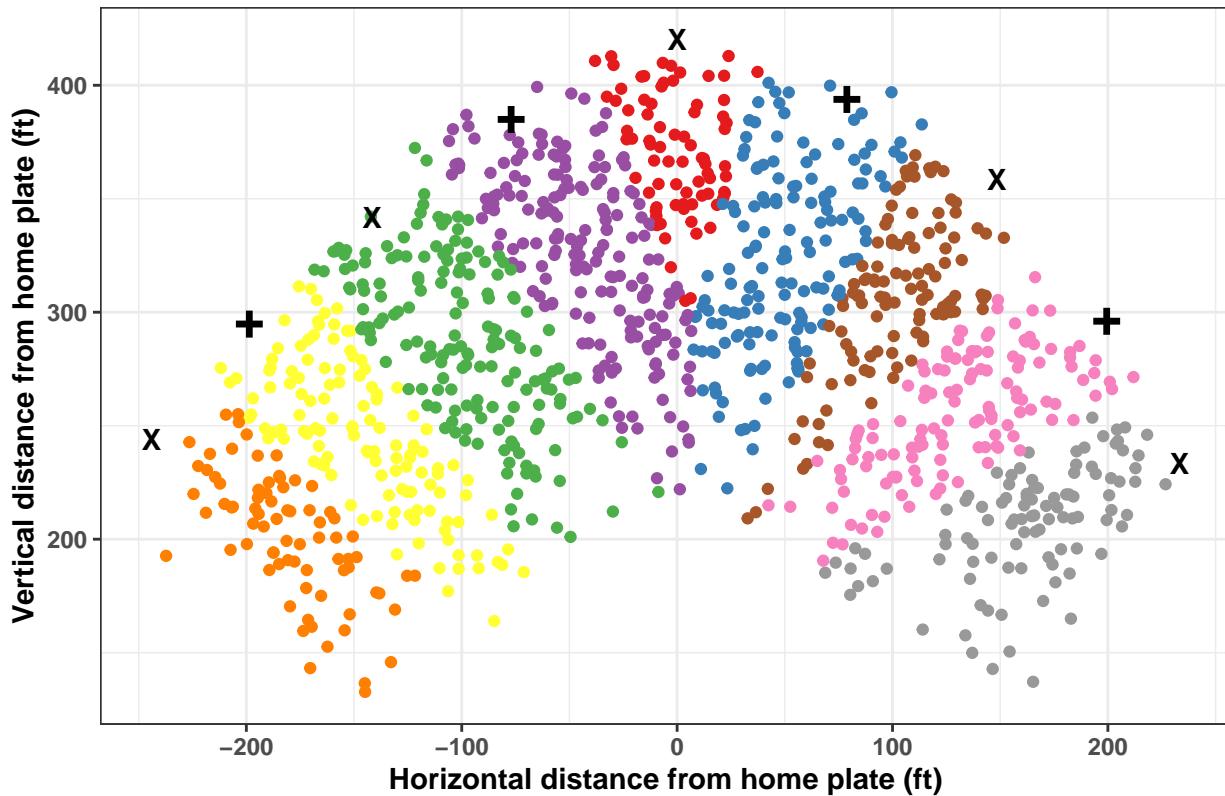
```
ggsave("../figures/jan27.png", plot = jan27, height = 4, width = 6)
```

January 28th

```
jan28_1 <- jan26 +
  annotate("text", x = unique(copa2024$wall_dist_lflc_x), y = unique(copa2024$wall_dist_lflc_y),
           fontface = "bold", label = "+", cex = 7) +
  annotate("text", x = unique(copa2024$wall_dist_lccf_x), y = unique(copa2024$wall_dist_lccf_y),
           fontface = "bold", label = "+", cex = 7) +
  annotate("text", x = unique(copa2024$wall_dist_cfrc_x), y = unique(copa2024$wall_dist_cfrc_y),
           fontface = "bold", label = "+", cex = 7) +
  annotate("text", x = unique(copa2024$wall_dist_rcrf_x), y = unique(copa2024$wall_dist_rcrf_y),
           fontface = "bold", label = "+", cex = 7)
```

```
jan28_1
```

Closest wall measurement for all Comerica Park fly balls, 2024



```
ggsave("../figures/jan28_1.png", plot = jan28_1, height = 4, width = 6)
```

```
df <- of_catch_prob_2024 %>% group_by(stadium) %>%
  mutate(stadium = fct_recode(stadium,
    "AmFam Field" = "American Family Field",
    "Guaranteed Rate" = "Guaranteed Rate Field",
    "Camden Yards" = "Oriole Park at Camden Yards",
    "Great American" = "Great American Ball Park",
    "Citizens Bank" = "Citizens Bank Park"
  )) %>%
  summarize(lf_x = unique(wall_dist_lf_x), lf_y = unique(wall_dist_lf_y),
```

```

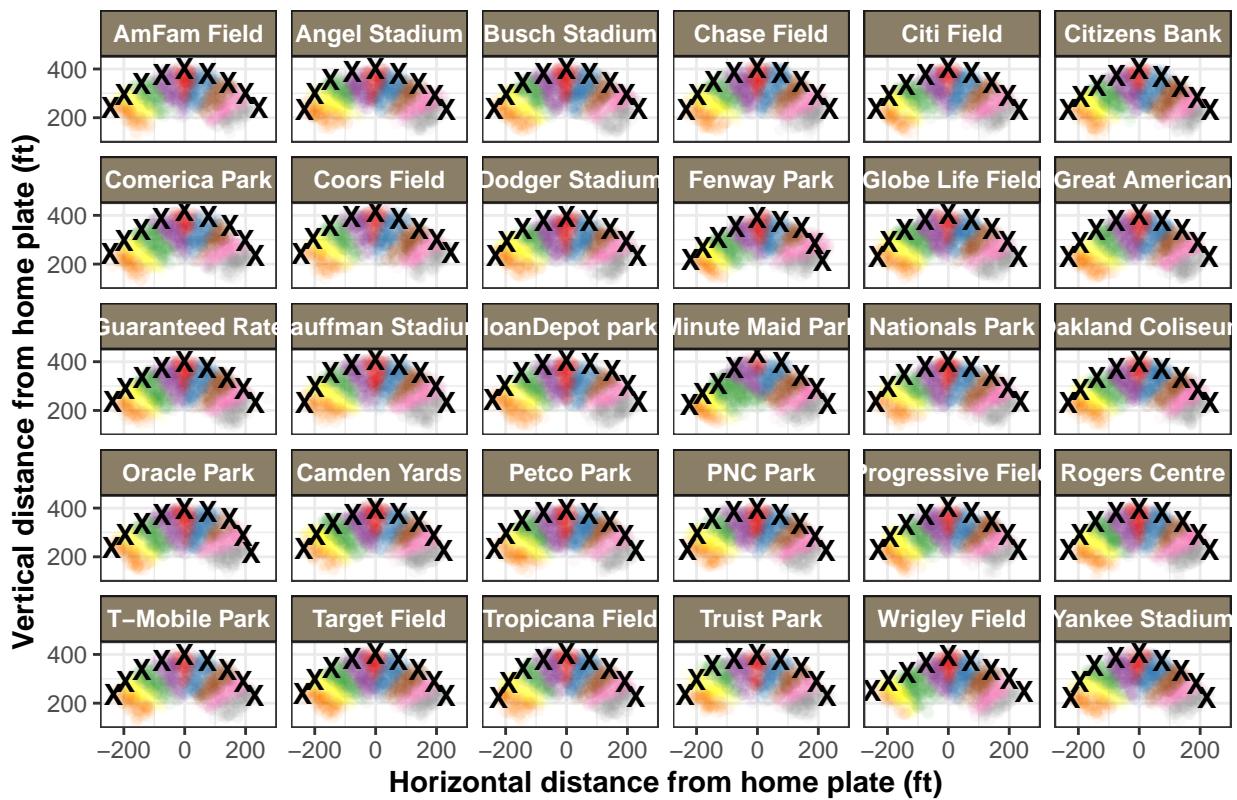
lf lc_x = unique(wall_dist_lf_lc_x), lf_lc_y = unique(wall_dist_lf_lc_y),
lc_x = unique(wall_dist_lc_x), lc_y = unique(wall_dist_lc_y),
lccf_x = unique(wall_dist_lccf_x), lccf_y = unique(wall_dist_lccf_y),
cf_x = unique(wall_dist_cf_x), cf_y = unique(wall_dist_cf_y),
cfrc_x = unique(wall_dist_cfrc_x), cfrc_y = unique(wall_dist_cfrc_y),
rc_x = unique(wall_dist_rc_x), rc_y = unique(wall_dist_rc_y),
rcrf_x = unique(wall_dist_rcrf_x), rcrf_y = unique(wall_dist_rcrf_y),
rf_x = unique(wall_dist_rf_x), rf_y = unique(wall_dist_rf_y))

jan28_2 <- ggplot() +
  geom_point(data = of_catch_prob_2024 %>% mutate(stadium = fct_recode(stadium,
    "AmFam Field" = "American Family Field",
    "Guaranteed Rate" = "Guaranteed Rate Field",
    "Camden Yards" = "Oriole Park at Camden Yards",
    "Great American" = "Great American Ball Park",
    "Citizens Bank" = "Citizens Bank Park"
  )), alpha = 0.05,
  aes(x = landing_pos_x, y = landing_pos_y, col = closest_wall)) +
  facet_wrap(~stadium) +
  geom_text(data = df, aes(x = lf_x, y = lf_y),
    fontface = "bold", label = "X") +
  geom_text(data = df, aes(x = lf_lc_x, y = lf_lc_y),
    fontface = "bold", label = "X") +
  geom_text(data = df, aes(x = lc_x, y = lc_y),
    fontface = "bold", label = "X") +
  geom_text(data = df, aes(x = lccf_x, y = lccf_y),
    fontface = "bold", label = "X") +
  geom_text(data = df, aes(x = cf_x, y = cf_y),
    fontface = "bold", label = "X") +
  geom_text(data = df, aes(x = cfrc_x, y = cfrc_y),
    fontface = "bold", label = "X") +
  geom_text(data = df, aes(x = rc_x, y = rc_y),
    fontface = "bold", label = "X") +
  geom_text(data = df, aes(x = rcrf_x, y = rcrf_y),
    fontface = "bold", label = "X") +
  geom_text(data = df, aes(x = rf_x, y = rf_y),
    fontface = "bold", label = "X") +
  theme_bw() +
  labs(title = "Nine wall distances for each MLB stadium",
    x = "Horizontal distance from home plate (ft)",
    y = "Vertical distance from home plate (ft)") +
  scale_color_brewer(palette = "Set1") +
  theme(legend.position = "none",
    axis.title = element_text(face = "bold"),
    title = element_text(face = "bold"),
    plot.subtitle = element_text(face = "italic"),
    strip.background = element_rect(fill = "wheat4", color = "gray10"),
    strip.text = element_text(face = "bold", color = "white")) +
  scale_x_continuous(breaks = c(-200, 0, 200)) +
  scale_y_continuous(breaks = c(200, 400))

```

jan28_2

Nine wall distances for each MLB stadium



```
ggsave("../figures/jan28_2.png", plot = jan28_2, height = 6, width = 8)
```

January 29th

```
ofCatch_prob_2024 %>%
  mutate(wall_ball = as.factor(ifelse(relevant_wall_distance - ball_distance < 8, 1, 0))) %>%
  count(statcast = wall, me = wall_ball)

## # A tibble: 4 x 3
##   statcast me      n
##   <dbl> <fct> <int>
## 1 0     0       30579
## 2 0     1       1082
## 3 1     0       595
## 4 1     1       2014
```

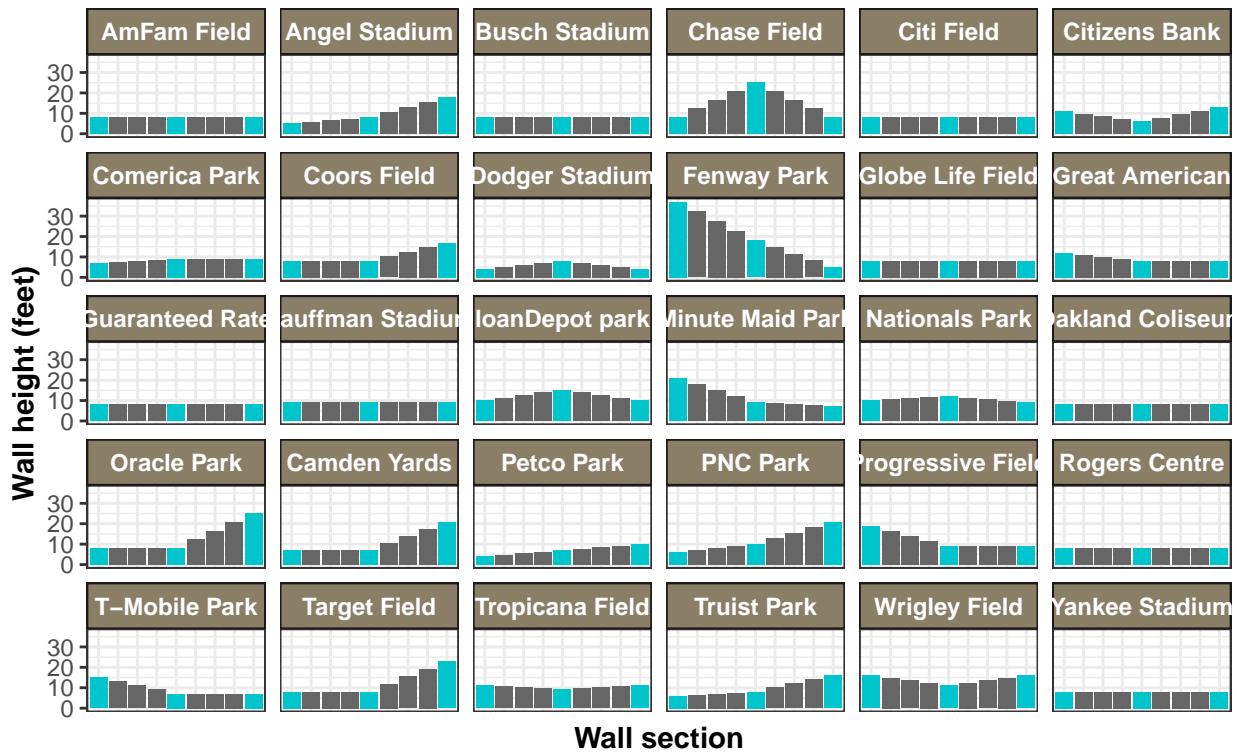
January 30th

```
jan30 <- of_catch_prob_2024 %>%
  mutate(stadium = fct_recode(stadium,
                               "AmFam Field" = "American Family Field",
                               "Guaranteed Rate" = "Guaranteed Rate Field",
                               "Camden Yards" = "Oriole Park at Camden Yards",
                               "Great American" = "Great American Ball Park",
                               "Citizens Bank" = "Citizens Bank Park"
                           )) %>%
  group_by(stadium) %>%
  summarize(wall_height_lf = unique(wall_height_lf), wall_height_lflc = unique(wall_height_lflc),
            wall_height_lc = unique(wall_height_lc), wall_height_lccf = unique(wall_height_lccf),
            wall_height_cf = unique(wall_height_cf), wall_height_cfrc = unique(wall_height_cfrc),
            wall_height_rc = unique(wall_height_rc), wall_height_rcrf = unique(wall_height_rcrf),
            wall_height_rf = unique(wall_height_rf)) %>%
  pivot_longer(cols = starts_with("wall_height"), names_prefix = "wall_height_",
               names_to = "section", values_to = "height") %>%
  mutate(section = fct_recode(as.factor(section),
                             "LF" = "lf", "LFLC" = "lflc", "LC" = "lc",
                             "LCCF" = "lccf", "CF" = "cf", "CFRC" = "cfrc",
                             "RC" = "rc", "RCRF" = "rcrf", "RF" = "rf")) %>%
  mutate(interpolated = as.factor(ifelse(
    section %in% c("LF", "CF", "RF"), 0, 1
  ))) %>%
  ggplot(aes(x = factor(section,
                         levels = c("LF", "LFLC", "LC", "LCCF", "CF",
                                    "CFRC", "RC", "RCRF", "RF")), y = height)) +
  geom_col(aes(fill = interpolated)) +
  facet_wrap(~stadium) +
  theme_bw() +
  theme(legend.position = "none",
        axis.title = element_text(face = "bold"),
        axis.text.x = element_blank(),
        axis.ticks.x = element_blank(),
        title = element_text(face = "bold"),
        plot.subtitle = element_text(face = "italic"),
        strip.background = element_rect(fill = "wheat4", color = "gray10"),
        strip.text = element_text(face = "bold", color = "white")) +
  labs(x = "Wall section", y = "Wall height (feet)",
       title = "Outfield wall heights at MLB stadiums, 2024",
       subtitle = "Measured heights are marked in teal, linearly interpolated heights are marked in gray",
       scale_fill_manual(values = c("turquoise3", "gray40")))
```

jan30

Outfield wall heights at MLB stadiums, 2024

Measured heights are marked in teal, linearly interpolated heights are marked in gray



```
ggsave("../figures/jan30.png", plot = jan30, height = 6, width = 8)
```