

# PP3

Eddie Morelli

April 11, 2022

## Libraries and Data

```
library("readr")
library("tidyverse")

## -- Attaching packages ----- tidyverse 1.3.1 --
## v ggplot2 3.3.5      v dplyr   1.0.8
## v tibble  3.1.6      v stringr 1.4.0
## v tidyr   1.2.0      v forcats 0.5.1
## v purrr   0.3.4

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()

library("sf")

## Warning: package 'sf' was built under R version 4.1.3
## Linking to GEOS 3.9.1, GDAL 3.2.1, PROJ 7.2.1; sf_use_s2() is TRUE

library("ggplot2")
lefties = read_csv("corbin_lefties.csv") %>%
  mutate(bat_hand = "Lefty")

## New names:
## * pitcher -> pitcher...8
## * fielder_2 -> fielder_2...42
## * pitcher -> pitcher...60
## * fielder_2 -> fielder_2...61

## Rows: 1366 Columns: 92
## -- Column specification -----
## Delimiter: ","
## chr  (16): pitch_type, player_name, events, description, des, game_type, sta...
## dbl  (67): release_speed, release_pos_x, release_pos_z, batter, pitcher...8,...
## lgl   (8): spin_dir, spin_rate_deprecated, break_angle_deprecated, break_len...
## date  (1): game_date
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.

righties = read_csv("corbin_righties.csv") %>%
  mutate(bat_hand = "Righty")
```

```
## New names:
## * pitcher -> pitcher...8
## * fielder_2 -> fielder_2...42
## * pitcher -> pitcher...60
## * fielder_2 -> fielder_2...61
## Rows: 1228 Columns: 92-- Column specification -----
## Delimiter: ","
## chr (16): pitch_type, player_name, events, description, des, game_type, sta...
## dbl (67): release_speed, release_pos_x, release_pos_z, batter, pitcher...8,...
## lgl (8): spin_dir, spin_rate_deprecated, break_angle_deprecated, break_len...
## date (1): game_date
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
head(rbind(lefties,righties))
```

```
## # A tibble: 6 x 93
##   pitch_type game_date release_speed release_pos_x release_pos_z player_name
##   <chr>      <date>         <dbl>         <dbl>         <dbl> <chr>
## 1 SI        2021-08-24         99.8          -0.57          5.95 Burnes, Corbin
## 2 SI        2021-08-30         99.2          -0.78          6.13 Burnes, Corbin
## 3 SI        2021-06-18         99.1          -0.91          5.94 Burnes, Corbin
## 4 SI        2021-09-11         99           -0.9           6.02 Burnes, Corbin
## 5 SI        2021-08-24         98.8          -0.81          5.98 Burnes, Corbin
## 6 SI        2021-04-03         98.8          -1.02          5.93 Burnes, Corbin
## # ... with 87 more variables: batter <dbl>, pitcher...8 <dbl>, events <chr>,
## #   description <chr>, spin_dir <lgl>, spin_rate_deprecated <lgl>,
## #   break_angle_deprecated <lgl>, break_length_deprecated <lgl>, zone <dbl>,
## #   des <chr>, game_type <chr>, stand <chr>, p_throws <chr>, home_team <chr>,
## #   away_team <chr>, type <chr>, hit_location <dbl>, bb_type <chr>,
## #   balls <dbl>, strikes <dbl>, game_year <dbl>, pfx_x <dbl>, pfx_z <dbl>,
## #   plate_x <dbl>, plate_z <dbl>, on_3b <dbl>, on_2b <dbl>, on_1b <dbl>, ...
```

```
corbin_so = rbind(lefties, righties) %>%
  select(release_speed, events, plate_x, plate_z, bat_hand, pitch_name) %>%
  filter(events == "strikeout") %>%
  drop_na()
head(corbin_so)
```

```
## # A tibble: 6 x 6
##   release_speed events   plate_x plate_z bat_hand pitch_name
##         <dbl> <chr>         <dbl> <dbl> <chr>   <chr>
## 1     99.8 strikeout    0.63  2.23 Lefty   Sinker
## 2     99.2 strikeout   -1.04  2.33 Lefty   Sinker
## 3     98.1 strikeout   -0.06  2.17 Lefty   Cutter
## 4     98.1 strikeout   -1.19  3.18 Lefty   4-Seam Fastball
## 5     97.7 strikeout   -0.53  2.64 Lefty   Cutter
## 6     97.6 strikeout    0.13  2.48 Lefty   Cutter
```

## Convert to Spatial Data

```
corbin_so_sf = st_as_sf(corbin_so, coords = c("plate_x", "plate_z"))
head(corbin_so_sf)
```

```
## Simple feature collection with 6 features and 4 fields
```

```

## Geometry type: POINT
## Dimension:      XY
## Bounding box:  xmin: -1.19 ymin: 2.17 xmax: 0.63 ymax: 3.18
## CRS:           NA
## # A tibble: 6 x 5
##   release_speed events    bat_hand pitch_name      geometry
##   <dbl> <chr>      <chr>    <chr>      <POINT>
## 1    99.8  strikeout Lefty    Sinker      (0.63 2.23)
## 2    99.2  strikeout Lefty    Sinker      (-1.04 2.33)
## 3    98.1  strikeout Lefty    Cutter      (-0.06 2.17)
## 4    98.1  strikeout Lefty    4-Seam Fastball (-1.19 3.18)
## 5    97.7  strikeout Lefty    Cutter      (-0.53 2.64)
## 6    97.6  strikeout Lefty    Cutter      (0.13 2.48)

```

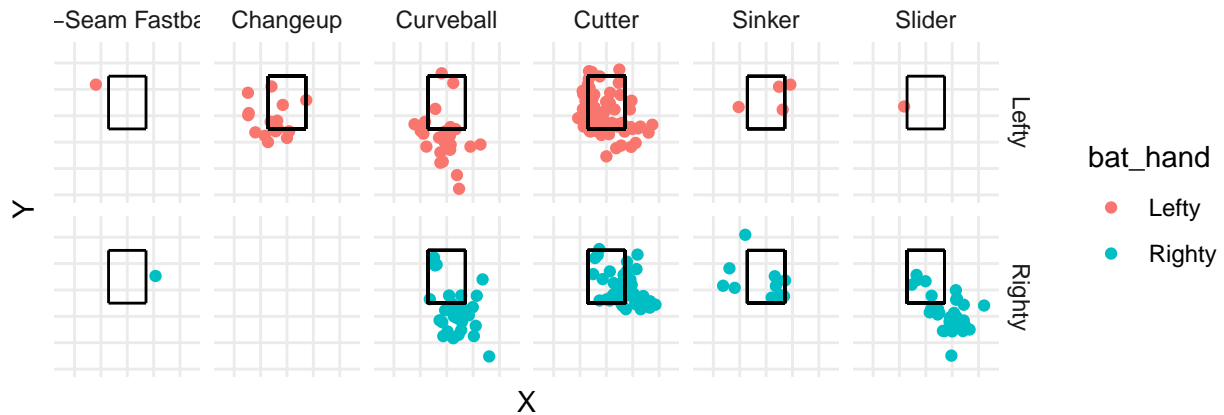
## Plot

```

ggplot(corbin_so_sf) +
  geom_sf(aes(col=bat_hand)) +
  geom_segment(aes(x=-0.71,y=1.5,xend=0.71,yend=1.5)) +
  geom_segment(aes(x=-0.71,y=3.5,xend=0.71,yend=3.5)) +
  geom_segment(aes(x=-0.71,y=1.5,xend=-0.71,yend=3.5)) +
  geom_segment(aes(x=0.71,y=1.5,xend=0.71,yend=3.5)) +
  theme_minimal() +
  xlim(-2.5, 2.5) +
  ylim(-1, 4.5) +
  xlab("X") +
  ylab("Y") +
  labs(title = "Corbin Burnes Strikeout Pitch Location, 2021") +
  theme(axis.text.x=element_blank(), axis.text.y=element_blank()) +
  facet_grid(bat_hand ~ pitch_name)

```

## Corbin Burnes Strikeout Pitch Location, 2021



I was interested in how an MLB pitcher locates pitches when attempting to strike out a batter. For my visualizations, I decided to go with one pitcher specifically, Corbin Burnes. Burnes won the 2021 NL Cy Young Award for best pitcher in the league, and collected 234 strikeouts. The data was downloaded from [baseballsavant.mlb.com](https://baseballsavant.mlb.com), and a few transformations were required, such as selecting the features that I was interested in, and filtering down to strikeouts. I also had to convert the data from a data frame to spatial data in the `sf` package. I knew that the location strategy for striking out batters would depend on the handedness of the batter, so I encoded the handedness by color in the above visualization. A few findings stand out when viewing the plot. First, a strikeout is most common on pitches away from a hitter for all pitches. I thought there might be one location that was a go-to for strikeouts, but it seems like it varies by both pitch type and batter-handedness. Second, strikeouts most commonly occur low in the strike zone, or below the strike zone, with the only outlier being the cutter. Another interesting finding is the difference of strikeouts with the changeup and slider vs. lefties and righties. It seems that Burnes is much more successful striking out righties than lefties with the slider low and away, while he is much more successful striking out lefties with the changeup. In fact, he did not record a strikeout with the changeup against righties, which is interesting as many righty pitchers do like to use a changeup as a strikeout pitch against righty batters.

### Plot 2

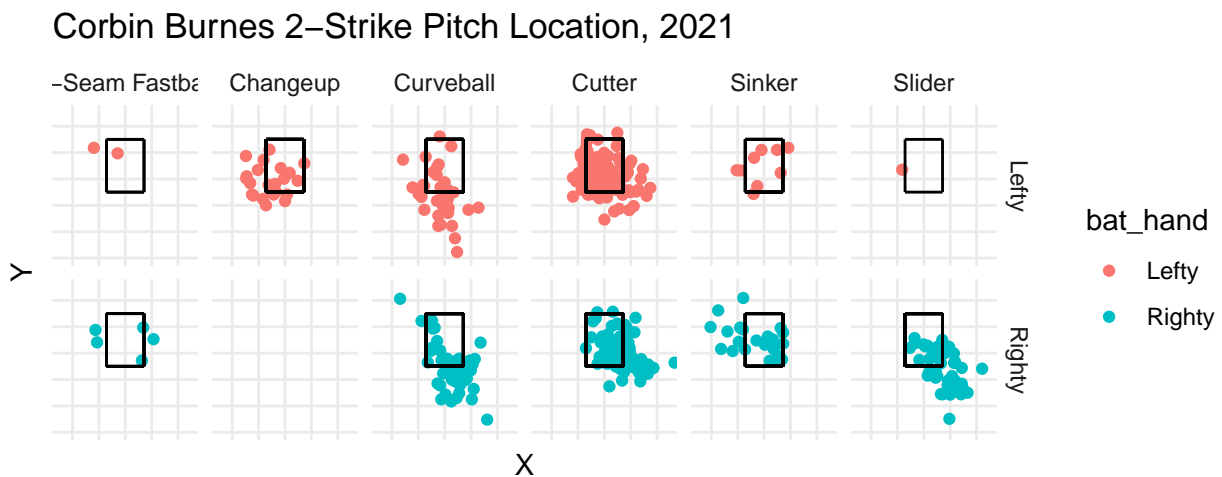
```
corbin_2s = rbind(lefties, righties) %>%
  select(release_speed, events, plate_x, plate_z, bat_hand, pitch_name, strikes) %>%
  filter(strikes == 2) %>%
  drop_na()
corbin_2s_sf = st_as_sf(corbin_2s, coords = c("plate_x", "plate_z"))

ggplot(corbin_2s_sf) +
  geom_sf(aes(col=bat_hand)) +
```

```

geom_segment(aes(x=-0.71,y=1.5,xend=0.71,yend=1.5)) +
geom_segment(aes(x=-0.71,y=3.5,xend=0.71,yend=3.5)) +
geom_segment(aes(x=-0.71,y=1.5,xend=-0.71,yend=3.5)) +
geom_segment(aes(x=0.71,y=1.5,xend=0.71,yend=3.5)) +
theme_minimal() +
xlim(-2.5, 2.5) +
ylim(-1, 4.5) +
xlab("X") +
ylab("Y") +
labs(title = "Corbin Burnes 2-Strike Pitch Location, 2021") +
theme(axis.text.x=element_blank(), axis.text.y=element_blank()) +
facet_grid(bat_hand ~ pitch_name)

```



I wanted to see Burnes' 2-strike pitch locations to see if there were any major differences between this and the strikeout chart. Ideally, a major difference would help us see which pitches are most effective at striking out batters proportionally. Perhaps the most interesting finding from this analysis is that Burnes did not throw a single changeup to right-handed batters with 2 strikes. I had assumed that maybe he had thrown some, but did not have success, so he used it much less, but he did not throw it at all. As a right-handed batter, you could be very confident that you will not see a changeup with 2 strikes against Corbin Burnes. Another finding is just how often he throws the cutter with two strikes. Burnes is known for his cutter, and you can see that he definitely trusts that pitch, especially with 2 strikes. Altogether, the major difference between this visualization and the previous visualization is that this contains many more pitches closer to the middle of the zone. When Burnes is able to locate his pitches near the edges of the zone, it seems that he is much more likely to collect a strikeout.

### Plot 3

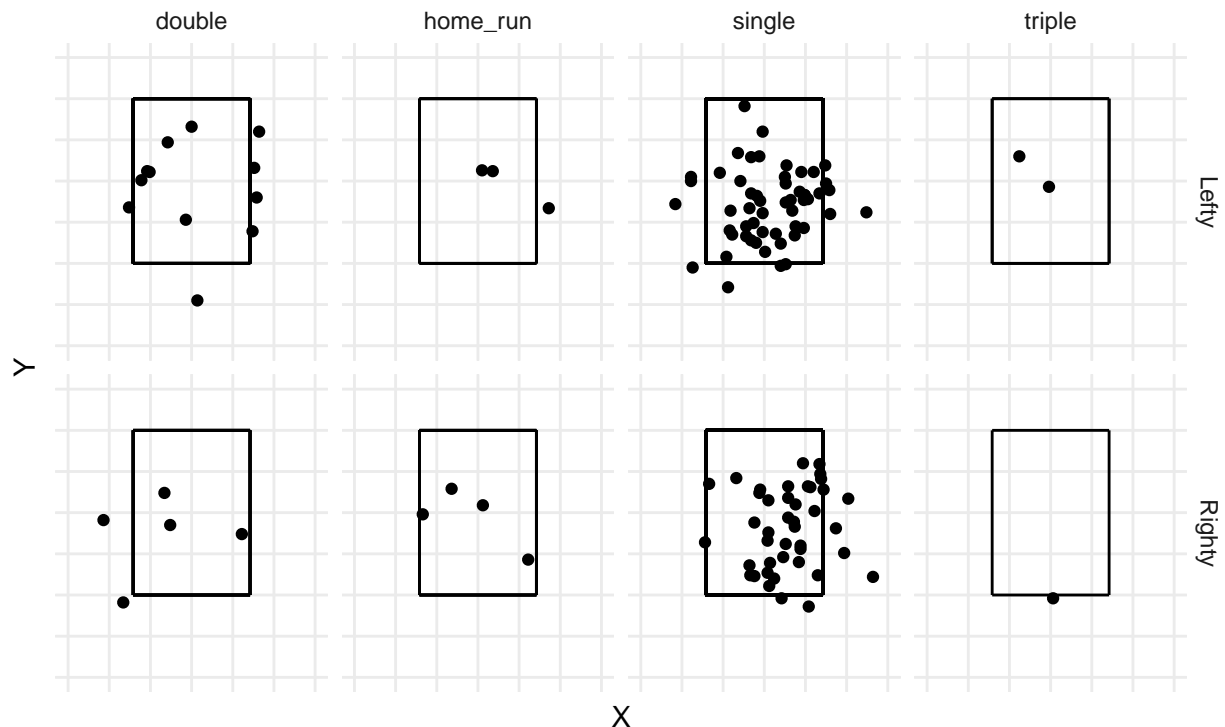
```
corbin_hit = rbind(lefties, righties) %>%
  select(release_speed, events, plate_x, plate_z, bat_hand) %>%
  filter(events %in% c("single", "double", "triple", "home_run")) %>%
  drop_na()
head(corbin_hit)

## # A tibble: 6 x 5
##   release_speed events plate_x plate_z bat_hand
##         <dbl> <chr>   <dbl> <dbl> <chr>
## 1          98  triple   -0.38  2.8  Lefty
## 2         97.8 single   -0.05  2.26 Lefty
## 3         97.8 single   -0.42  1.9  Lefty
## 4         97.8 double   -0.76  2.18 Lefty
## 5         97.6 single   -0.41  2.14 Lefty
## 6         97.6 double   -0.29  2.97 Lefty

corbin_hit_sf = st_as_sf(corbin_hit, coords = c("plate_x", "plate_z"))

ggplot(corbin_hit_sf) +
  geom_sf() +
  geom_segment(aes(x=-0.71,y=1.5,xend=0.71,yend=1.5)) +
  geom_segment(aes(x=-0.71,y=3.5,xend=0.71,yend=3.5)) +
  geom_segment(aes(x=-0.71,y=1.5,xend=-0.71,yend=3.5)) +
  geom_segment(aes(x=0.71,y=1.5,xend=0.71,yend=3.5)) +
  theme_minimal() +
  xlim(-1.5, 1.5) +
  ylim(0.5, 4) +
  xlab("X") +
  ylab("Y") +
  labs(title = "Corbin Burnes Pitch Location for Hits, 2021") +
  facet_grid(bat_hand ~ events) +
  theme(axis.text.x=element_blank(), axis.text.y=element_blank())
```

## Corbin Burnes Pitch Location for Hits, 2021



The second visualization that I created was to see the location of Burnes' pitches that resulted in hits. Some data manipulation similar to that of the previous plot was required. I used `facet_wrap` to see the location for singles, doubles, triples, and home runs. Theoretically, this would allow me to see the pitches that hitters are making the best contact against. The major difference from the strikeout plot is the proximity of the points to the strike zone. In addition, some interesting patterns arise. It seems like pitches that resulted in doubles were most commonly on the inner and outer thirds of the strike zone, while triples and home runs were more often down the middle.

### Reflection

A common suggestion that I received for improving my visualizations was to include more data about the pitch, especially the type of pitch. The major change that I made was faceting the first plot by handedness and pitch type, to take a deeper look into Corbin Burnes' tendencies. This led to many interesting findings regarding where he located certain pitches for strikeouts and which pitches he used. I also added another plot to further explore Burnes' tendencies, which was all of his two-strike pitches. Once again, this revealed interesting patterns such as the lack of usage of the changeup against righties and the slider against lefties. I think that if a batter were to see this 2-strike pitch chart, they could have some more confidence about what pitches to expect when facing Burnes in a 2-strike count. Finally, I faceted the third plot by batter-handedness. I received one suggestion about batter handedness and hit tendency, and I thought this added some value, as you can see variation in plots in lefties vs. righties in home runs and doubles.

### Short Description

Milwaukee Brewers' pitcher Corbin Burnes won the 2021 NL Cy Young Award while collecting the third-most strikeouts in the league. What pitches were most critical to Burnes' success?