

```

# load packages
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.1 --

## v ggplot2 3.3.5      v purrr  0.3.4
## v tibble  3.1.6      v dplyr  1.0.7
## v tidyr   1.1.4      v stringr 1.4.0
## v readr   2.1.1      v forcats 0.5.1

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

library(knitr)
library(lme4)

## Loading required package: Matrix

##
## Attaching package: 'Matrix'

## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack

library(broom.mixed)
library(skimr)
library(readr)
library(ggplot2)
library(dplyr)
library(tidyverse)
library(knitr)
library(lme4)
library(broom.mixed)
library(skimr)
library(readr)
library(ggplot2)
library(dplyr)

lefty_ninth <- read_csv("lefty_ninth.csv")

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

## Rows: 40000 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.
```

```
righty_ninth <- read_csv("righty_ninth.csv")
```

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

```
## Rows: 40000 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

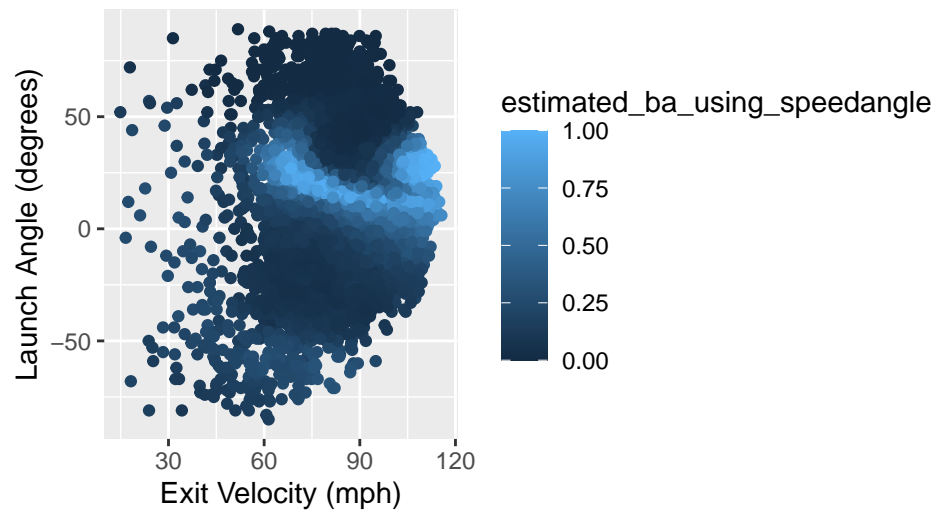
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
lefty_ninth <- lefty_ninth %>% mutate(hor_break = (pfx_x*12),
                                     ind_vert_break = (pfx_z*12))
righty_ninth <- righty_ninth %>% mutate(hor_break = (pfx_x*12),
                                       ind_vert_break = (pfx_z*12))
```

General EDA

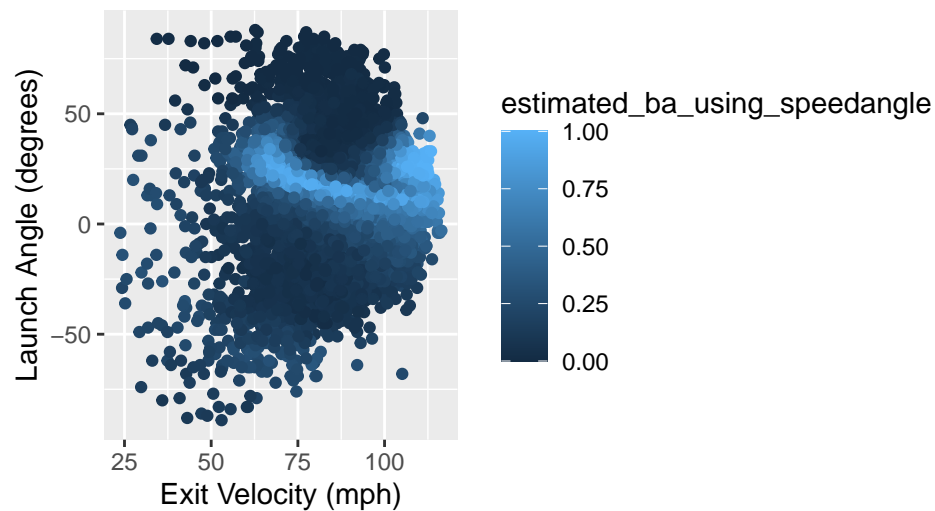
```
ggplot(data = subset(righty_ninth, !is.na(estimated_ba_using_speedangle)),
       mapping = aes(x=launch_speed, y=launch_angle,
                     color = estimated_ba_using_speedangle)) +
  geom_point() + labs(title = "Right-Handed Speed/Angle Breakdown",
                     x = "Exit Velocity (mph)",
                     y = "Launch Angle (degrees)")
```

Right-Handed Speed/Angle Breakdown



```
ggplot(data = subset(lefty_ninth, !is.na(estimated_ba_using_speedangle)),
       mapping = aes(x=launch_speed, y=launch_angle,
                     color = estimated_ba_using_speedangle)) +
  geom_point() + labs(title = "Left-Handed Speed/Angle Breakdown",
                     x = "Exit Velocity (mph)",
                     y = "Launch Angle (degrees)")
```

Left-Handed Speed/Angle Breakdown

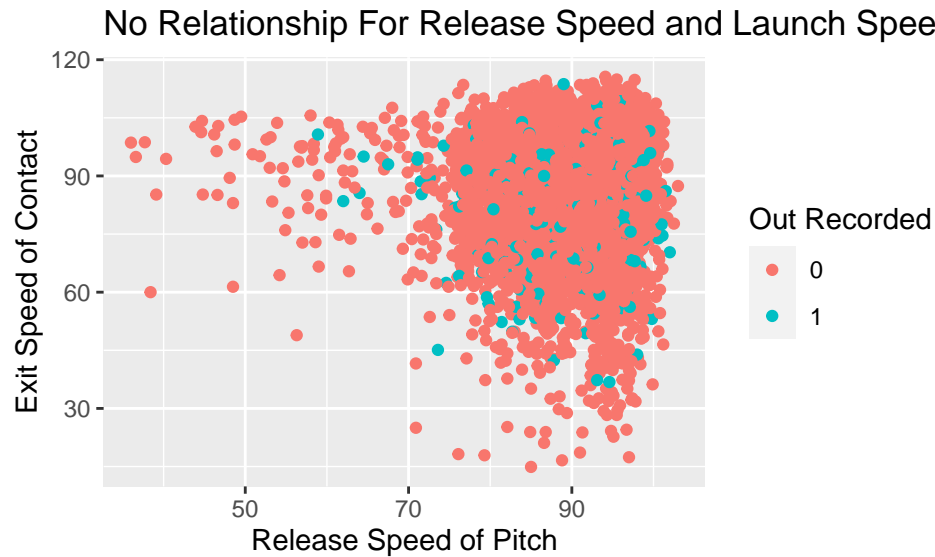


```
righty_ninth <- righty_ninth %>%
  mutate(inplay_out = ifelse(events == c("double_play", "field_out",
                                         "fielders_choice_out", "force_out",
                                         "grounded_into_double_play",
                                         "other_out", "sac_fly",
                                         "sac_fly_double_play"), "1", "0"))

r_in_play_df <- righty_ninth %>% filter(description == "hit_into_play")
```

```
ggplot(data = r_in_play_df, mapping = aes(x=release_speed, y=launch_speed,
                                           color = inplay_out)) +
  geom_point() + labs(title = "No Relationship For Release Speed and Launch Speed/Out Made",
                     x = "Release Speed of Pitch",
                     y = "Exit Speed of Contact",
                     color = "Out Recorded")
```

Warning: Removed 21 rows containing missing values (geom_point).



```
rh_fastball <- righty_ninth %>% filter(pitch_type == c("FF", "FA", "SI", "FC"))
rh_breakingball <- righty_ninth %>% filter(pitch_type == c("KC", "SL", "CU", "CS"))
rh_changeup <- righty_ninth %>% filter(pitch_type == c("CH", "FS"))

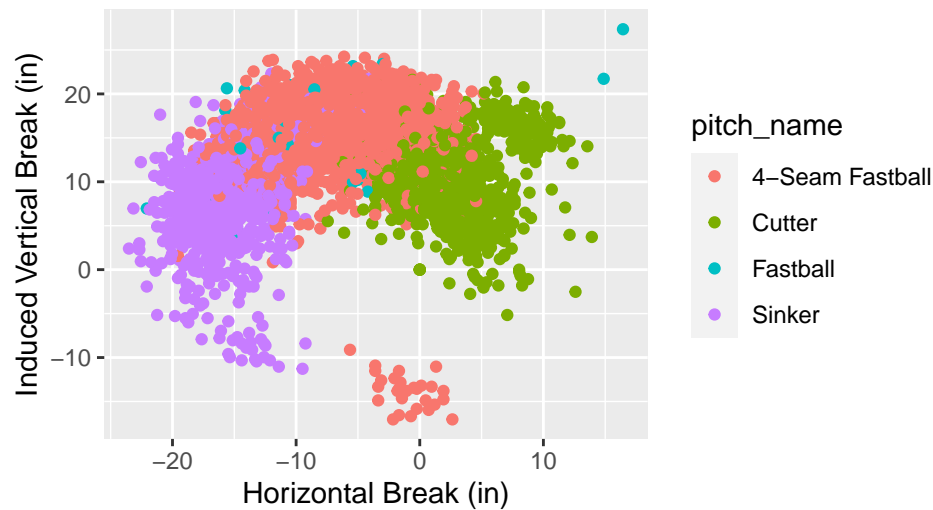
lh_fastball <- lefty_ninth %>% filter(pitch_type == c("FF", "FA", "SI", "FC"))
lh_breakingball <- lefty_ninth %>% filter(pitch_type == c("KC", "SL", "CU", "CS"))
lh_changeup <- lefty_ninth %>% filter(pitch_type == c("CH", "FS"))
```

Fastball EDA

```
ggplot(data = rh_fastball, mapping = aes(x=hor_break, y=ind_vert_break,
                                           color = pitch_name)) +
  geom_point() + labs(title = "Right-Handed Fastball Breakdown",
                     x = "Horizontal Break (in)",
                     y = "Induced Vertical Break (in)")
```

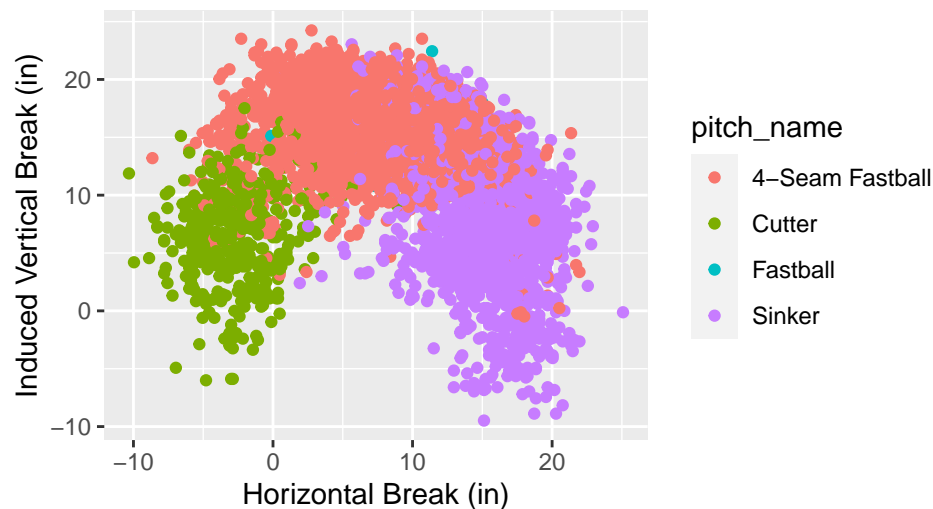
Warning: Removed 1 rows containing missing values (geom_point).

Right-Handed Fastball Breakdown



```
ggplot(data = lh_fastball, mapping = aes(x=hor_break, y=ind_vert_break,
                                          color = pitch_name)) +
  geom_point() + labs(title = "Left-Handed Fastball Breakdown",
                     x = "Horizontal Break (in)",
                     y = "Induced Vertical Break (in)")
```

Left-Handed Fastball Breakdown



```
rh_fastball %>% drop_na(release_speed, release_spin_rate, hor_break,
                       ind_vert_break, spin_axis) %>% group_by(pitch_name) %>%
  summarize(release_speed = mean(release_speed),
            release_spin = mean(release_spin_rate),
            hor_break = mean(hor_break),
            vert_break = mean(ind_vert_break),
            tilt = mean(spin_axis))
```

```
## # A tibble: 4 x 6
```

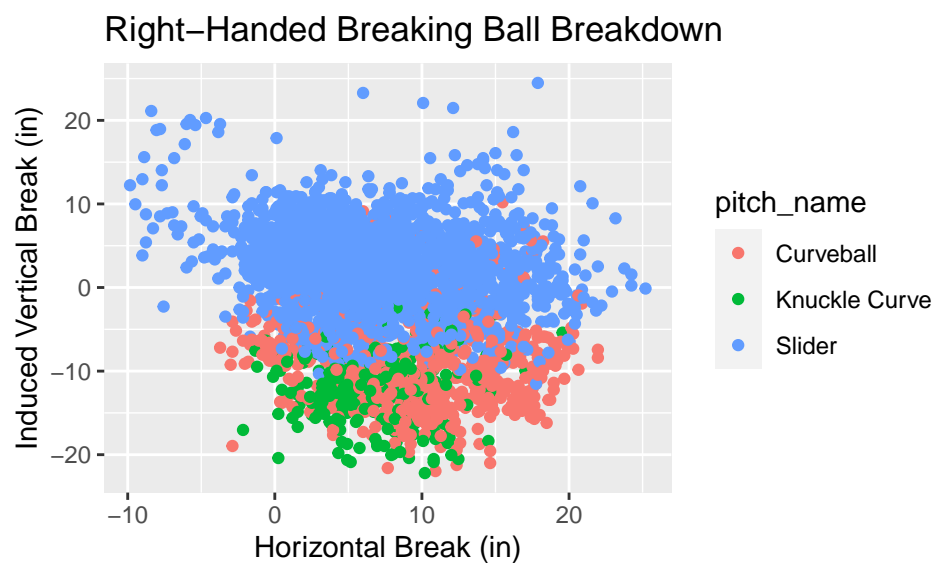
```
##   pitch_name      release_speed release_spin hor_break vert_break  tilt
##   <chr>           <dbl>         <dbl>      <dbl>    <dbl> <dbl>
## 1 4-Seam Fastball      94.9         2304.      -7.28     16.0  213.
## 2 Cutter              91.9         2412.       2.96     10.8  180.
## 3 Fastball            69.8         1704.      -8.79     15.0  222.
## 4 Sinker              94.7         2168.     -14.9      9.15  223.
```

```
lh_fastball %>% drop_na(release_speed, release_spin_rate, hor_break,
                        ind_vert_break, spin_axis) %>% group_by(pitch_name) %>%
  summarize(release_speed = mean(release_speed),
            release_spin = mean(release_spin_rate),
            hor_break = mean(hor_break),
            vert_break = mean(ind_vert_break),
            tilt = mean(spin_axis))
```

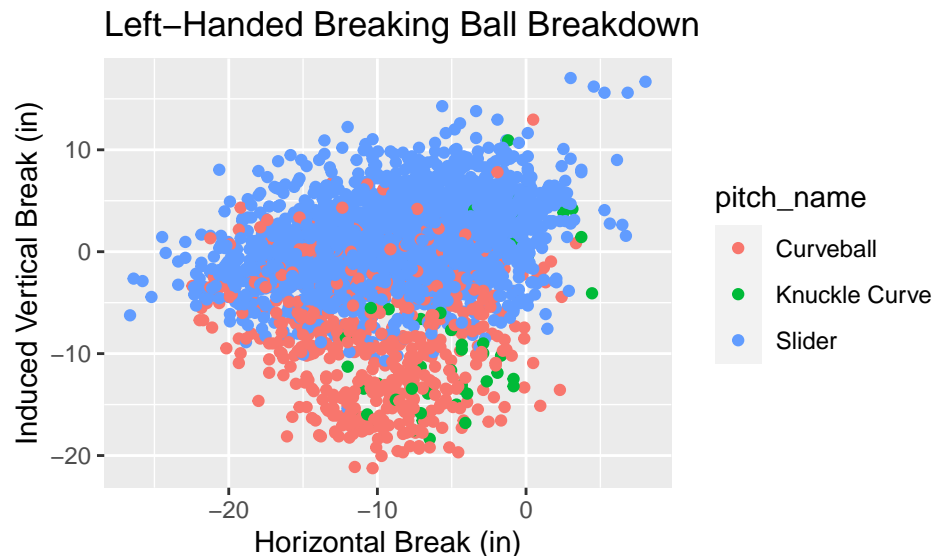
```
## # A tibble: 4 x 6
##   pitch_name      release_speed release_spin hor_break vert_break  tilt
##   <chr>           <dbl>         <dbl>      <dbl>    <dbl> <dbl>
## 1 4-Seam Fastball      93.8         2288.       7.36     16.0  150.
## 2 Cutter              88.4         2377.      -2.67      6.82  190.
## 3 Fastball            65.5         1709.       7.61     15.0  134.
## 4 Sinker              93.7         2137.      14.3      9.54  130.
```

EDA Breaking Balls

```
ggplot(data = rh_breakingball, mapping = aes(x=hor_break, y=ind_vert_break,
                                              color = pitch_name)) +
  geom_point() + labs(title = "Right-Handed Breaking Ball Breakdown",
                     x = "Horizontal Break (in)",
                     y = "Induced Vertical Break (in)")
```



```
ggplot(data = lh_breakingball, mapping = aes(x=hor_break, y=ind_vert_break,
                                             color = pitch_name)) +
  geom_point() + labs(title = "Left-Handed Breaking Ball Breakdown",
                     x = "Horizontal Break (in)",
                     y = "Induced Vertical Break (in)")
```



```
rh_breakingball %>% drop_na(release_speed, release_spin_rate, hor_break,
                           ind_vert_break, spin_axis) %>% group_by(pitch_name) %>%
  summarize(release_speed = mean(release_speed),
            release_spin = mean(release_spin_rate),
            hor_break = mean(hor_break),
            vert_break = mean(ind_vert_break),
            tilt = mean(spin_axis))
```

```
## # A tibble: 3 x 6
##   pitch_name  release_speed release_spin hor_break vert_break  tilt
##   <chr>          <dbl>         <dbl>    <dbl>    <dbl> <dbl>
## 1 Curveball      80.0         2592.     10.0     -9.54  41.8
## 2 Knuckle Curve  82.5         2489.      6.79    -11.7  33.2
## 3 Slider        85.2         2423.      6.58      2.16 111.
```

```
lh_breakingball %>% drop_na(release_speed, release_spin_rate, hor_break,
                           ind_vert_break, spin_axis) %>% group_by(pitch_name) %>%
  summarize(release_speed = mean(release_speed),
            release_spin = mean(release_spin_rate),
            hor_break = mean(hor_break),
            vert_break = mean(ind_vert_break),
            tilt = mean(spin_axis))
```

```
## # A tibble: 3 x 6
##   pitch_name  release_speed release_spin hor_break vert_break  tilt
##   <chr>          <dbl>         <dbl>    <dbl>    <dbl> <dbl>
## 1 Curveball      78.1         2447.    -10.1     -8.13 310.
```

## 2 Knuckle Curve	80.1	2259.	-4.30	-7.36	284.
## 3 Slider	83.2	2422.	-8.47	1.17	260.

Appendix

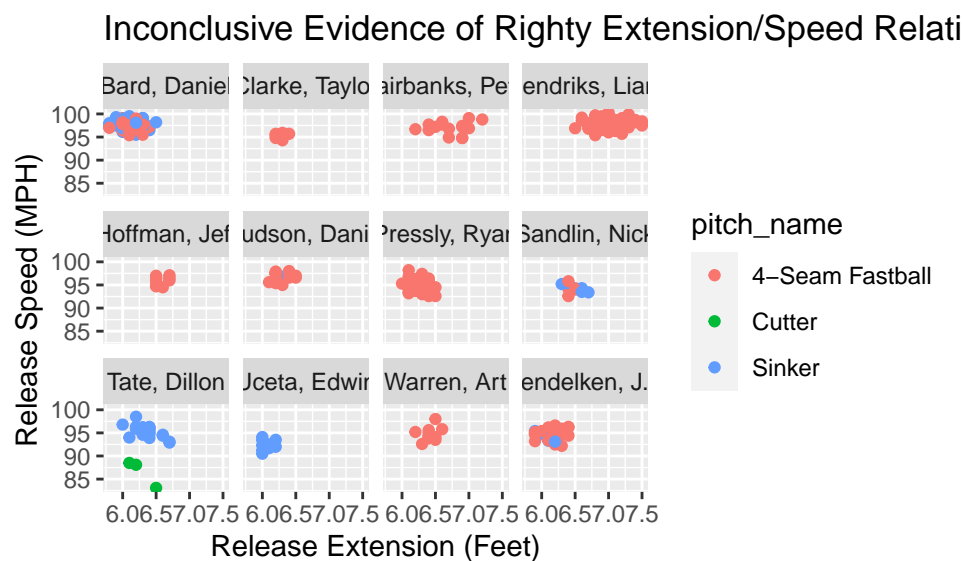
```
right_names = righty_ninth %>% group_by(player_name) %>% . $player_name
left_names = lefty_ninth %>% group_by(player_name) %>% . $player_name
set.seed(3)

# get sample of 12 pitchers
sample_righties <- sample(right_names, 12)
sample_lefties <- sample(left_names, 12)

# get data for those schools
sample_data_r <- rh_fastball %>%
  filter(player_name %in% sample_righties)

sample_data_l <- lh_fastball %>%
  filter(player_name %in% sample_lefties)

ggplot(data = sample_data_r, mapping = aes(x=release_extension, y=release_speed,
                                           color = pitch_name)) + facet_wrap(~player_name) +
  geom_point() +
  labs(title = "Inconclusive Evidence of Righty Extension/Speed Relationship",
       x = "Release Extension (Feet)",
       y = "Release Speed (MPH)")
```



```
ggplot(data = sample_data_l, mapping = aes(x=release_extension, y=release_speed,
                                           color = pitch_name)) + facet_wrap(~player_name) +
  geom_point() +
  labs(title = "Inconclusive Evidence of Lefty Extension/Speed Relationship",
       x = "Release Extension (Feet)",
       y = "Release Speed (MPH)")
```


Warning: Removed 3 rows containing missing values (geom_point).

