ball_tracking

Henry Guerrero-Duarte

2022-10-23

Libraries used

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

Reading CSV file

```
mydata1=read.csv("pitch_data_questionnaire.csv")
str(mydata1)
```

```
## 'data.frame':    4315 obs. of  38 variables:
##  $ pitch_id            : chr  "d85e9579-f85a-4841-9b6e-54afe7fdfaa2" "1c9b74a8-1d9d-470b-aefe-467
##  $ pitcher_id          : chr  "player_b" "player_b" "player_a" "player_b" ...
##  $ pa_of_inning        : int  1 4 1 5 2 2 2 2 4 5 ...
##  $ pitch_of_pa         : int  6 3 3 2 2 1 4 1 7 3 ...
##  $ inning              : int  4 3 2 3 4 4 2 5 6 4 ...
##  $ outs                : int  0 1 0 2 1 0 0 1 1 2 ...
##  $ balls               : int  3 1 1 0 0 0 0 0 3 0 ...
##  $ strikes             : num  2 1 1 1 1 0 2 0 2 2 ...
##  $ batter_side         : chr  "left" "right" "right" "left" ...
##  $ pitcher_throws      : chr  "left" "left" "right" "left" ...
##  $ pitch_type          : chr  "four_seam" "changeup" "slider" "slider" ...
##  $ pitch_call          : chr  "foul_ball" "foul_ball" "in_play" "in_play" ...
##  $ rel_speed           : num  93.9 84.6 84.1 79.3 89.2 ...
```

1

```
##  $ spin_rate               : num  2293 2151 2400 2621 2535 ...
##  $ tilt                    : chr  "10:30" "9:45" "7:45" "4:45" ...
##  $ rel_height              : num  5.66 5.27 5.48 5.9 5.57 ...
##  $ rel_side                : num  -2.29 -2.67 2.09 -1.89 1.43 ...
##  $ extension               : num  5.79 5.82 5.21 5.54 5.78 ...
##  $ vert_break              : num  -19.1 -31.9 -45.1 -49.7 -31.7 ...
##  $ induced_vert_break      : num  12.31 5.66 -4.78 -6.38 3.66 ...
##  $ horz_break              : num  -15.9 -13.33 -8.43 5.37 -1.26 ...
##  $ plate_loc_height        : num  3.22 1.61 2.03 1.86 1.51 ...
##  $ plate_loc_side          : num  -1.05218 0.32795 -0.00679 0.20234 -1.13999 ...
##  $ strike_prob             : num  0.0429 0.5196 0.9459 0.7879 0.0202 ...
##  $ runner_on_first_id      : num  NA 669344 NA 666150 NA ...
##  $ runner_on_second_id     : num  NA 681146 NA NA NA ...
##  $ runner_on_third_id      : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ post_runner_on_first_id : num  NA 669344 NA 666150 687401 ...
##  $ post_runner_on_second_id: num  NA 681146 NA NA NA ...
##  $ post_runner_on_third_id : num  NA NA NA NA NA ...
##  $ hit_type                : chr  "" "" "hard_ground_ball" "soft_ground_ball" ...
##  $ play_result             : chr  "foul_ball" "foul_ball" "out" "out" ...
##  $ direction               : num  69.1 -57 -36.2 61.9 NA ...
##  $ exit_speed              : num  51.2 89.4 95.8 52.5 NA ...
##  $ angle                   : num  -49.82 -11.89 5.13 -59.17 NA ...
##  $ distance                : num  3.88 10.19 89.75 2.03 NA ...
##  $ bearing                 : num  64.6 -44.9 -37.5 21.7 NA ...
##  $ venue                   : chr  "Cheney Stadium" "Las Vegas Ballpark" "PrinceGeorge" "Sutter Health
```

```r
#View(mydata1)
#summary(mydata1)
```

player_a

```r
#select rows where player_a appears in any column
df_a = mydata1 %>% filter_all(any_vars(. %in% c("player_a")))

#Pitches dataframes for player_b
changeup_a = df_a %>% filter_all(any_vars(. %in% c("changeup")))
curveball_a = df_a %>% filter_all(any_vars(. %in% c("curveball")))
cutter_a = df_a %>% filter_all(any_vars(. %in% c("cutter")))
four_seam_a = df_a %>% filter_all(any_vars(. %in% c("four_seam")))
sinker_a = df_a %>% filter_all(any_vars(. %in% c("sinker")))
slider_a = df_a %>% filter_all(any_vars(. %in% c("slider")))

#Finding out Unique names for Venues
unique(df_a$venue)
```

```
## [1] "PrinceGeorge"          "The Diamond"           "CanalPark"
## [4] "UPMC Park"             "HadlockField"          "NaturalGasField"
## [7] "FNBField"              "Delta Dental Stadium"
```

```r
#Venue dataframes For Player_a
PrinceGeorge_a = df_a %>% filter_all(any_vars(. %in% c("PrinceGeorge")))
The_Diamond_a = df_a %>% filter_all(any_vars(. %in% c("The Diamond")))
CanalPark_a = df_a %>% filter_all(any_vars(. %in% c("CanalPark")))
UPMC_Park_a = df_a %>% filter_all(any_vars(. %in% c("UPMC Park")))
HadlockField_a = df_a %>% filter_all(any_vars(. %in% c("HadlockField")))
NaturalGasField_a = df_a %>% filter_all(any_vars(. %in% c("NaturalGasField")))
FNBField_a = df_a %>% filter_all(any_vars(. %in% c("FNBField")))
Delta_Dental_Stadium_a = df_a %>% filter_all(any_vars(. %in% c("Delta Dental Stadium")))

pitch_call_a = table(df_a['pitch_call'])
```

player__b

```r
#select rows where player_b appears in any column
df_b = mydata1 %>% filter_all(any_vars(. %in% c("player_b")))

#Pitches dataframes for player_b
changeup_b = df_b %>% filter_all(any_vars(. %in% c("changeup")))
curveball_b = df_b %>% filter_all(any_vars(. %in% c("curveball")))
slider_b = df_b %>% filter_all(any_vars(. %in% c("slider")))
four_seam_b = df_b %>% filter_all(any_vars(. %in% c("four_seam")))

#Finding out Unique names for Venues
unique(df_b$venue)
```

```
## [1] "Cheney Stadium"              "Las Vegas Ballpark"
## [3] "Sutter Health Park"          "Greater Nevada Field"
## [5] "Smith's Ballpark"            "Southwest University Park"
## [7] "Dell Diamond"                "Isotopes Park"
## [9] "Chickasaw Bricktown Ballpark"
```

```r
#Venue dataframes For Player_b
Cheney_Stadium_b = df_b %>% filter_all(any_vars(. %in% c("Cheney Stadium")))
Las_Vegas_Ballpark_b = df_b %>% filter_all(any_vars(. %in% c("Las Vegas Ballpark")))
Sutter_Health_Park_b = df_b %>% filter_all(any_vars(. %in% c("Sutter Health Park")))
Greater_Nevada_Field_b = df_b %>% filter_all(any_vars(. %in% c("Greater Nevada Field")))
Smiths_Ballpark_b = df_b %>% filter_all(any_vars(. %in% c("Smith's Ballpark")))
Southwest_University_Park_b = df_b %>% filter_all(any_vars(. %in% c("Southwest University Park")))
Dell_Diamond_b = df_b %>% filter_all(any_vars(. %in% c("Dell Diamond")))
Isotopes_Park_b = df_b %>% filter_all(any_vars(. %in% c("Isotopes Park")))
Chickasaw_Bricktown_Ballpark_b = df_b %>% filter_all(any_vars(. %in% c("Chickasaw Bricktown Ballpark")))

#data frame for pitch_call_b
pitch_call_b = table(df_b['pitch_call'])
```

Creating Coordniates for batters box

```
x <- c(-.95,.95,.95,-.95,-.95)
z <- c(1.6,1.6,3.5,3.5,1.6)

#store in data frame
sz <- tibble(x,z)
```
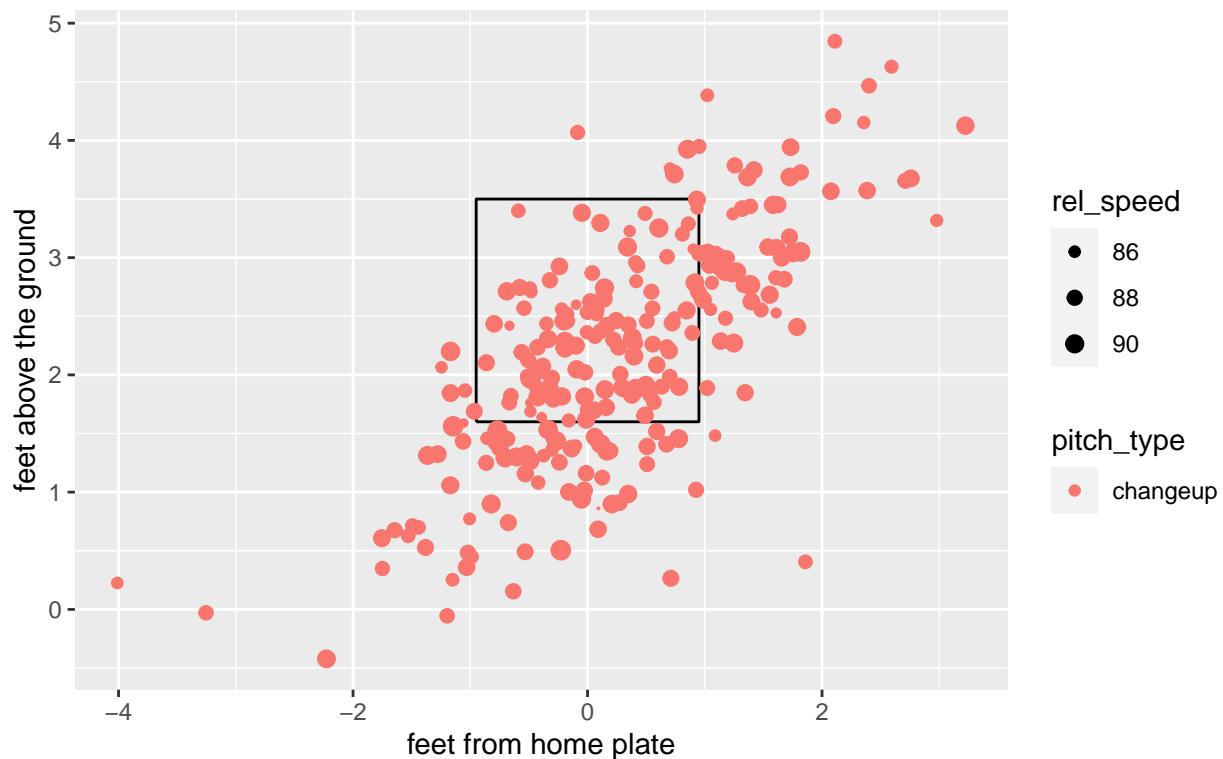
Player A Heat maps

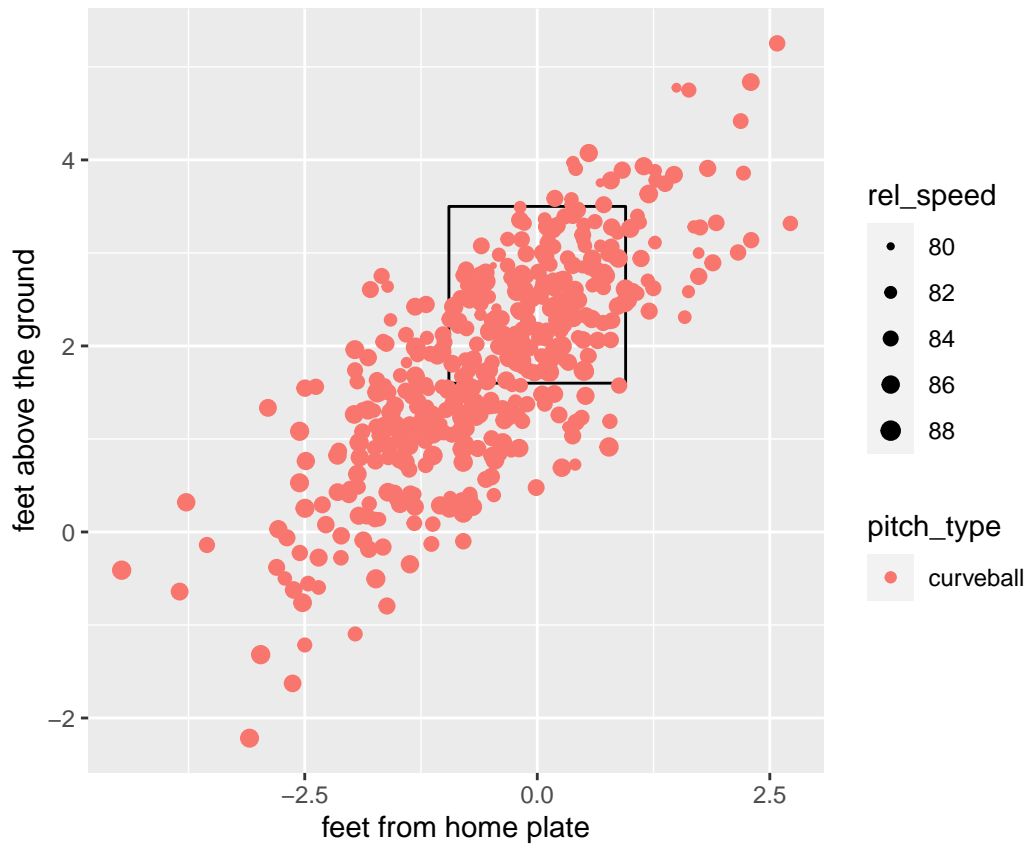## player_a Pitches(changeup, curveball, cutter, four_seam, sinker, slider) Heat map

```
ggplot()+
  geom_path(data = sz, aes(x=x, y=z))+
  coord_equal()+
  xlab("feet from home plate")+
  ylab("feet above the ground")+
  geom_point(data = df_a,aes(x=plate_loc_side,y=plate_loc_height,size=rel_speed,color=pitch_type))+
  scale_size(range = c(0.01,3))
```
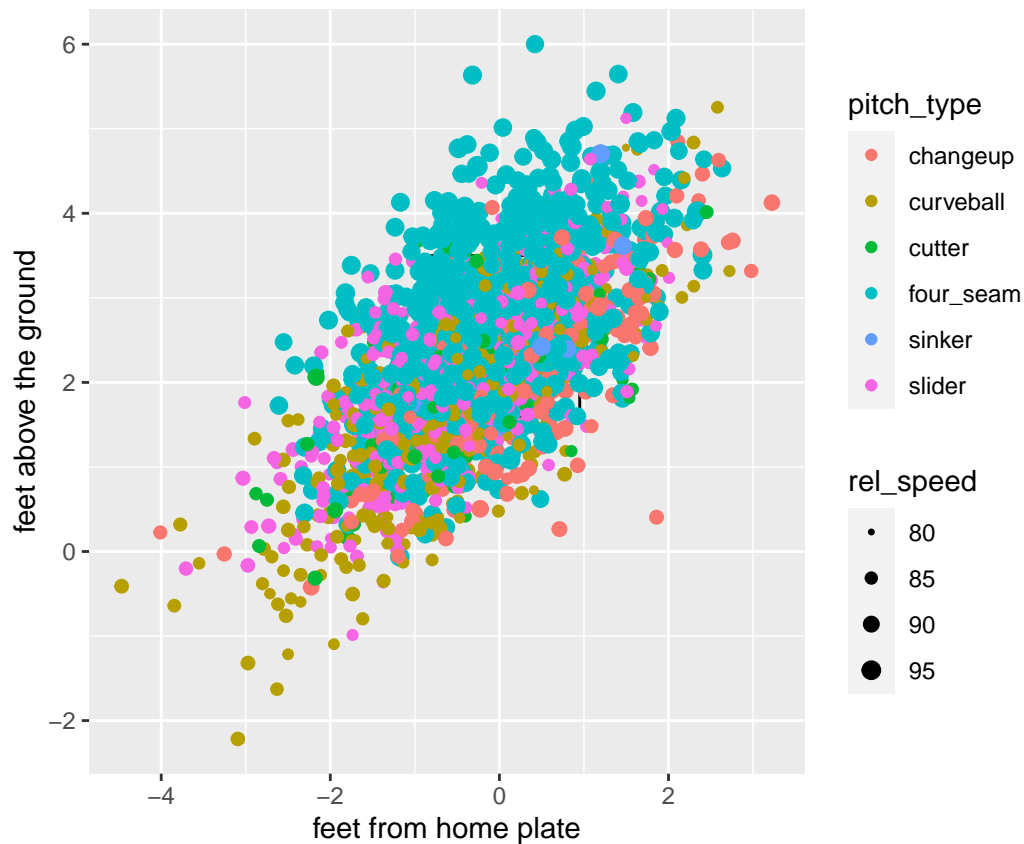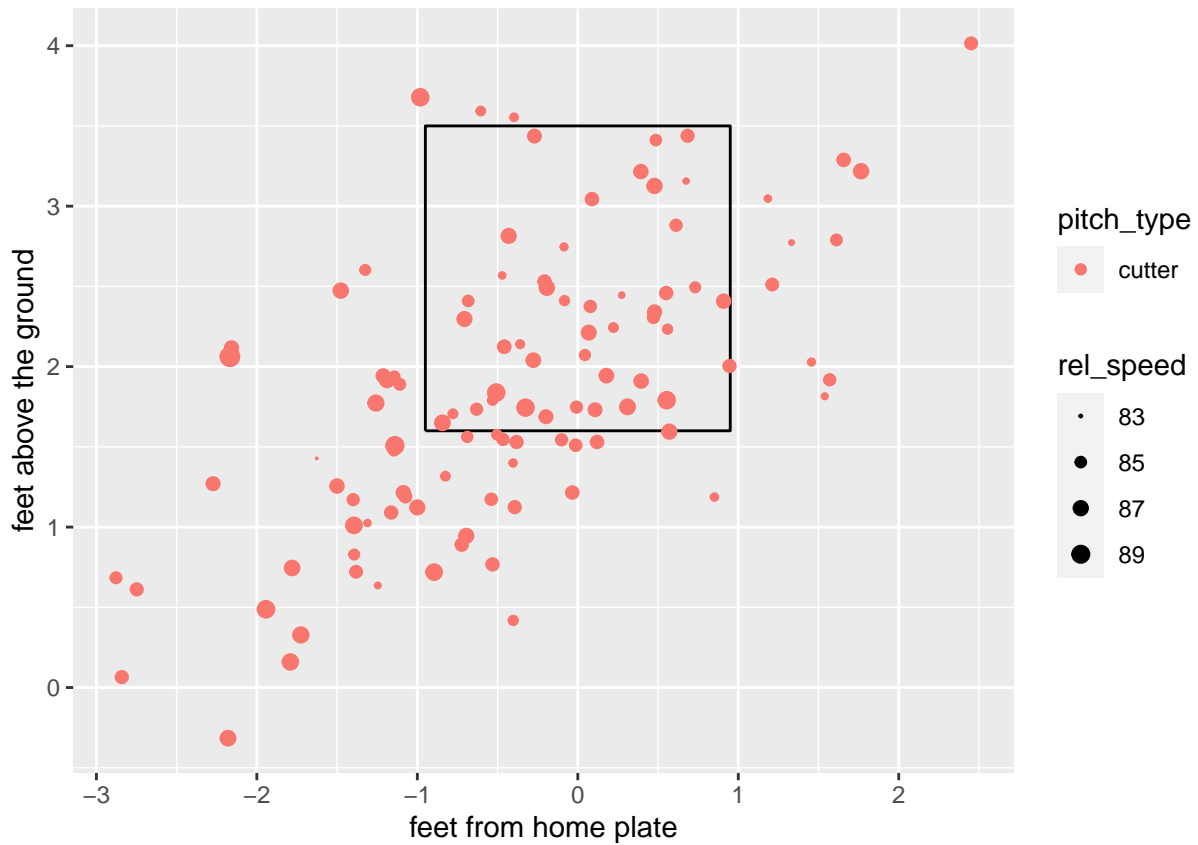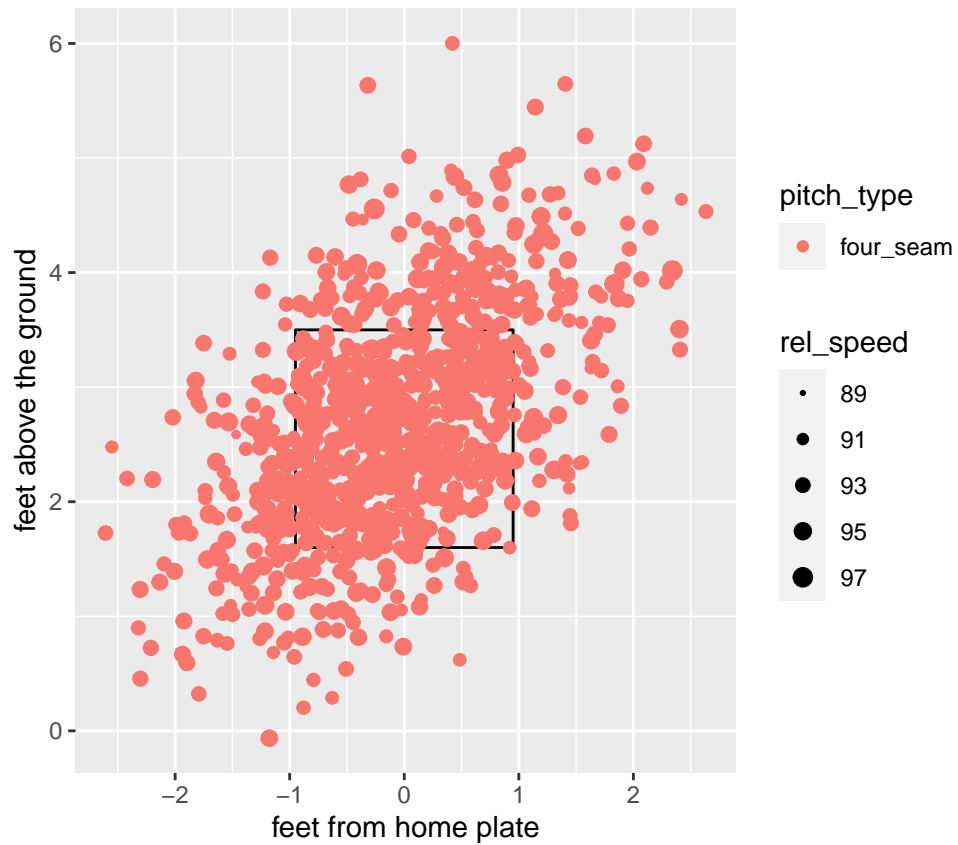
player_a Pitches(changeup) Heat map

```
ggplot()+
  geom_path(data = sz, aes(x=x, y=z))+
  coord_equal()+
  xlab("feet from home plate")+
  ylab("feet above the ground")+
  geom_point(data = changeup_a,aes(x=plate_loc_side,y=plate_loc_height,size=rel_speed,color=pitch_type))
  scale_size(range = c(0.01,3))
```
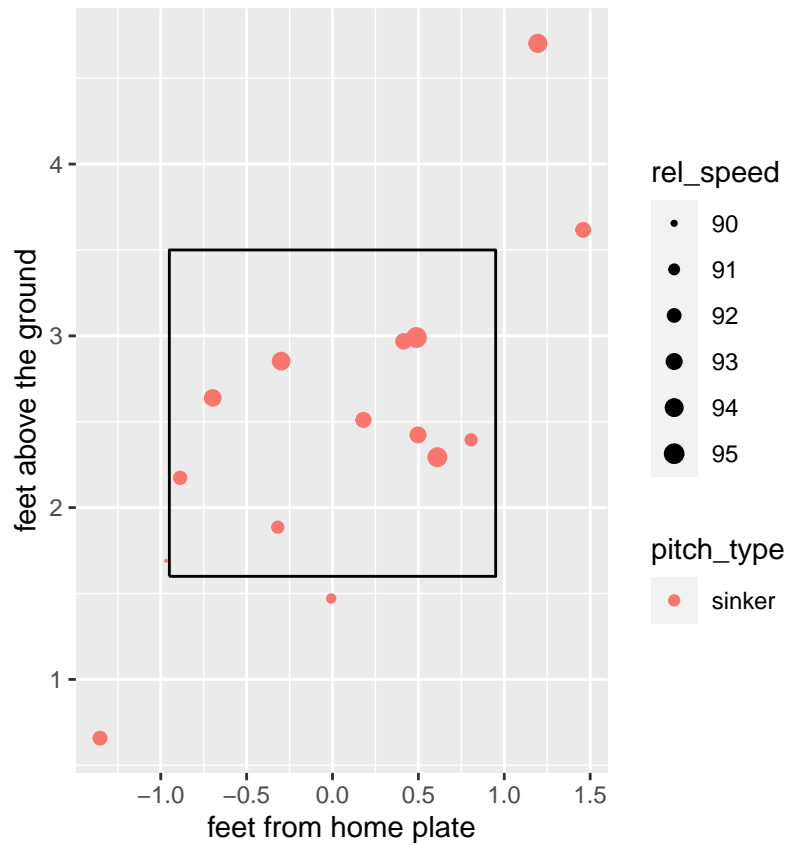

player_a Pitches(curveball) Heat map

```
ggplot()+
  geom_path(data = sz, aes(x=x, y=z))+
  coord_equal()+
  xlab("feet from home plate")+
  ylab("feet above the ground")+
  geom_point(data = curveball_a,aes(x=plate_loc_side,y=plate_loc_height,size=rel_speed,color=pitch_type))
  scale_size(range = c(0.01,3))
```

player_a Pitches(cutter) Heat map

```
ggplot()+
  geom_path(data = sz, aes(x=x, y=z))+
  coord_equal()+
  xlab("feet from home plate")+
  ylab("feet above the ground")+
  geom_point(data = cutter_a,aes(x=plate_loc_side,y=plate_loc_height,size=rel_speed,color=pitch_type))+
  scale_size(range = c(0.01,3))
```

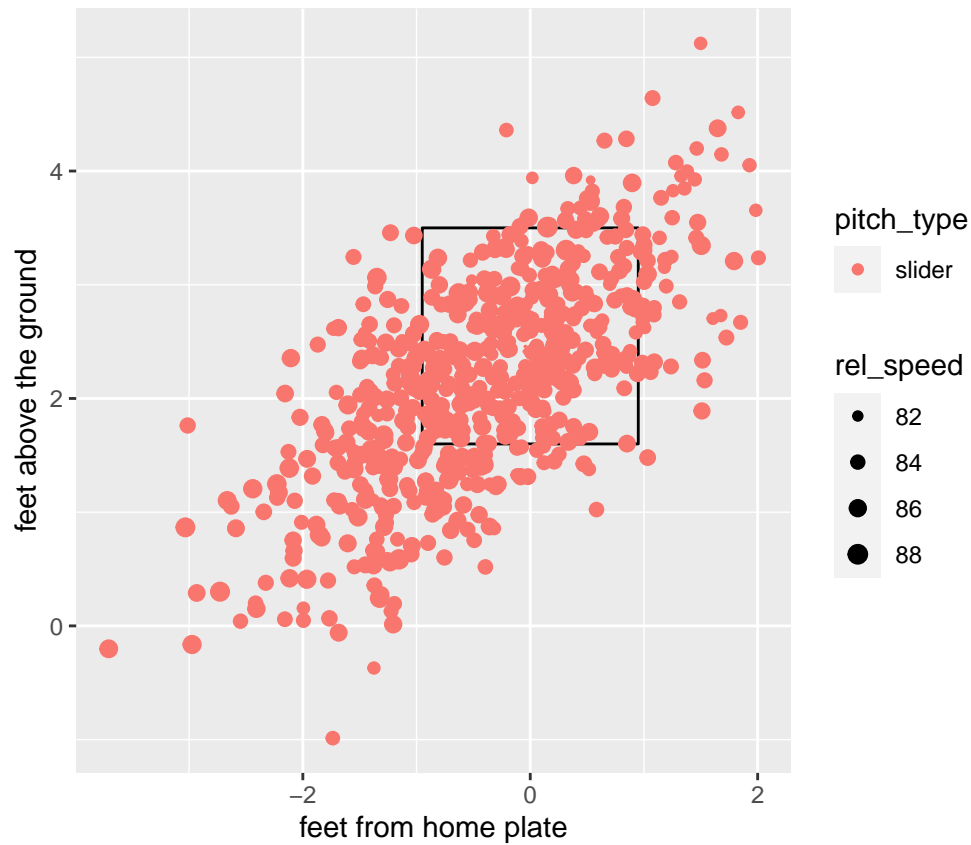**player_a Pitches(four_seam) Heat map**

```
ggplot()+
  geom_path(data = sz, aes(x=x, y=z))+
  coord_equal()+
  xlab("feet from home plate")+
  ylab("feet above the ground")+
  geom_point(data = four_seam_a,aes(x=plate_loc_side,y=plate_loc_height,size=rel_speed,color=pitch_type
  scale_size(range = c(0.01,3))
```
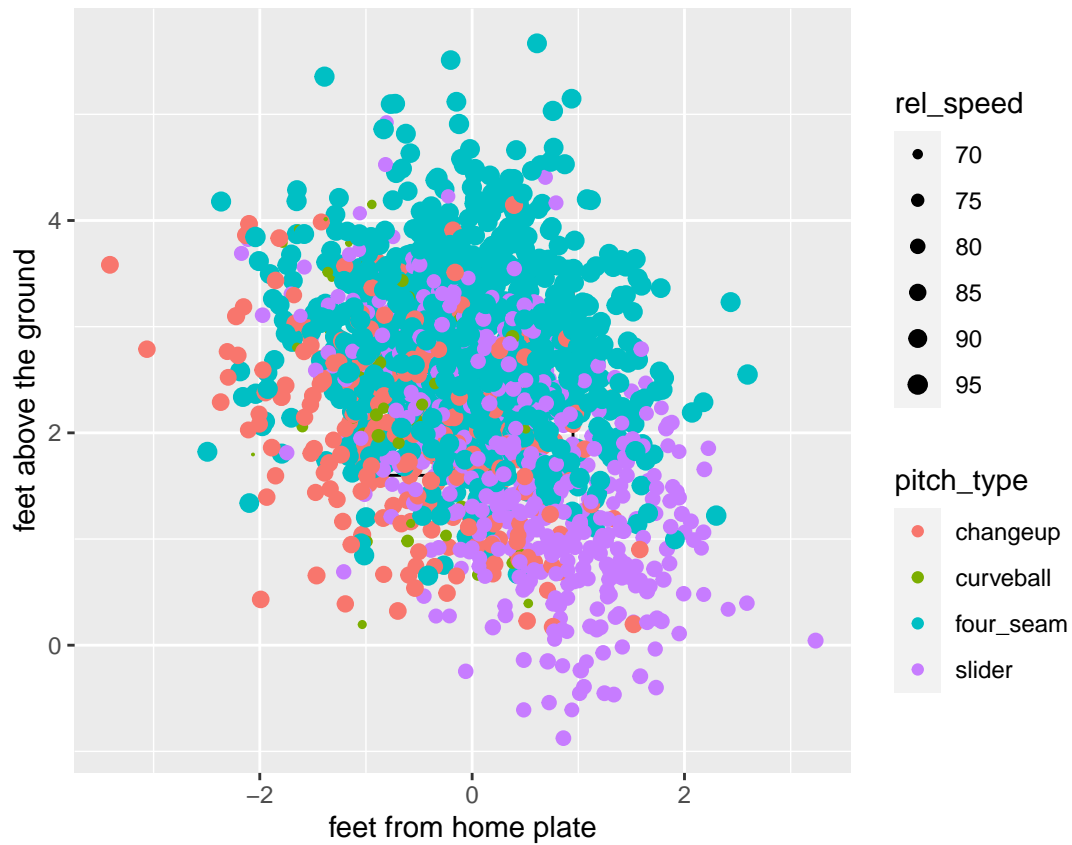
player_a Pitches(sinker) Heat map

```
ggplot()+
  geom_path(data = sz, aes(x=x, y=z))+
  coord_equal()+
  xlab("feet from home plate")+
  ylab("feet above the ground")+
  geom_point(data = sinker_a,aes(x=plate_loc_side,y=plate_loc_height,size=rel_speed,color=pitch_type))+
  scale_size(range = c(0.01,3))
```

player_a Pitches(slider) Heat map

```
ggplot()+
  geom_path(data = sz, aes(x=x, y=z))+
  coord_equal()+
  xlab("feet from home plate")+
  ylab("feet above the ground")+
  geom_point(data = slider_a,aes(x=plate_loc_side,y=plate_loc_height,size=rel_speed,color=pitch_type))+
  scale_size(range = c(0.01,3))
```
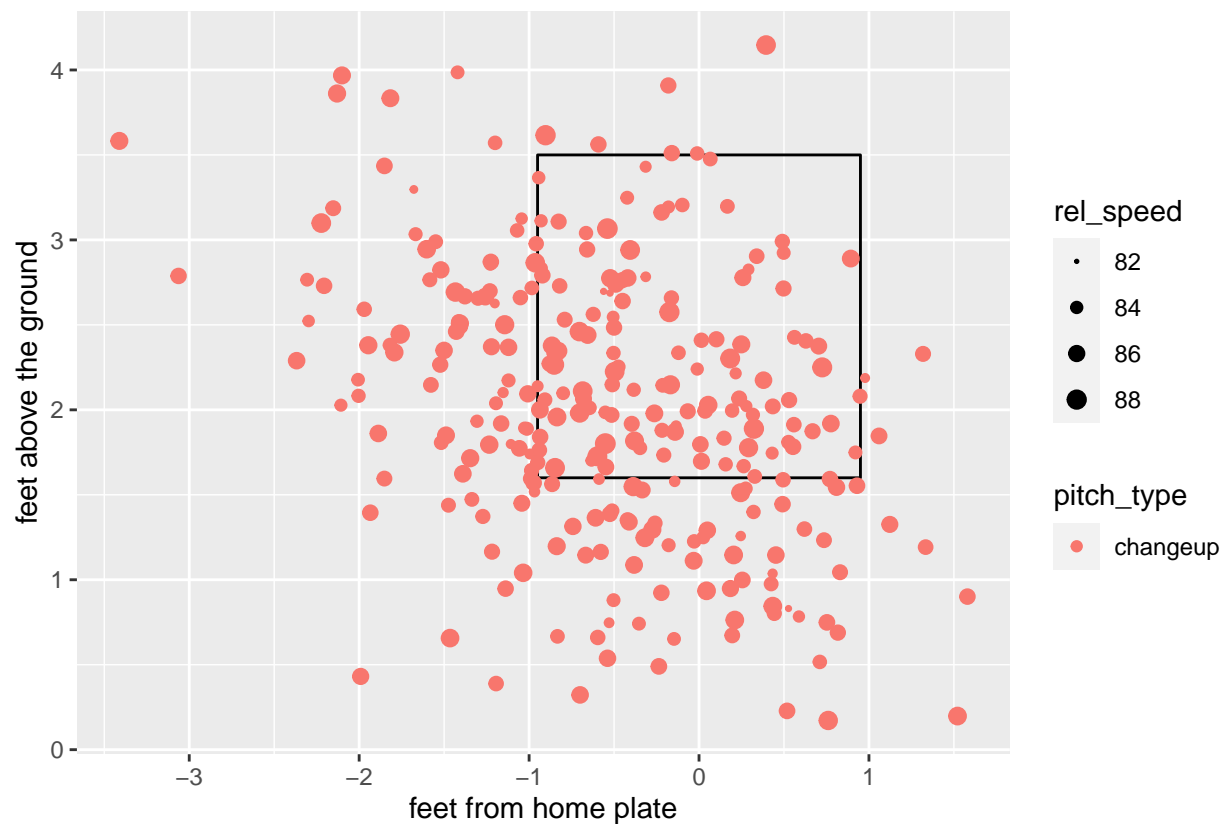
Player B Heat maps

player_b Pitches(changeup, curveball, four_seam, slider) Heat Map

```
ggplot()+
  geom_path(data = sz, aes(x=x, y=z))+
  coord_equal()+
  xlab("feet from home plate")+
  ylab("feet above the ground")+
  geom_point(data = df_b,aes(x=plate_loc_side,y=plate_loc_height,size=rel_speed,color=pitch_type))+
  scale_size(range = c(0.01,3))
```
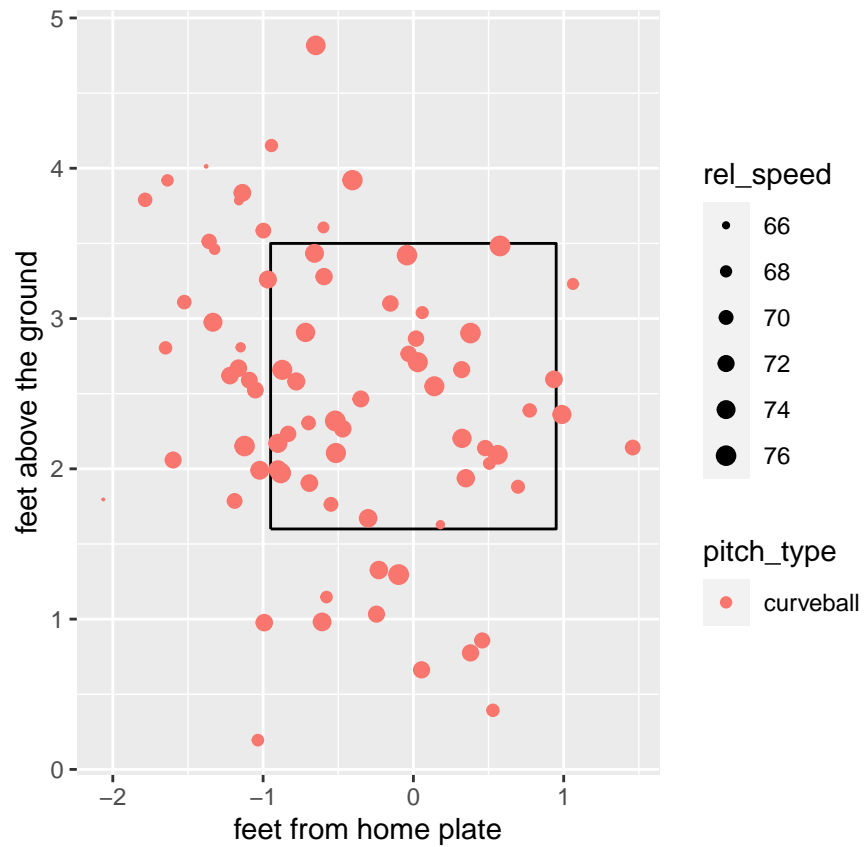
player_a Pitches(changeup) Heat map

```
ggplot()+
  geom_path(data = sz, aes(x=x, y=z))+
  coord_equal()+
  xlab("feet from home plate")+
  ylab("feet above the ground")+
  geom_point(data = changeup_b,aes(x=plate_loc_side,y=plate_loc_height,size=rel_speed,color=pitch_type))
  scale_size(range = c(0.01,3))
```
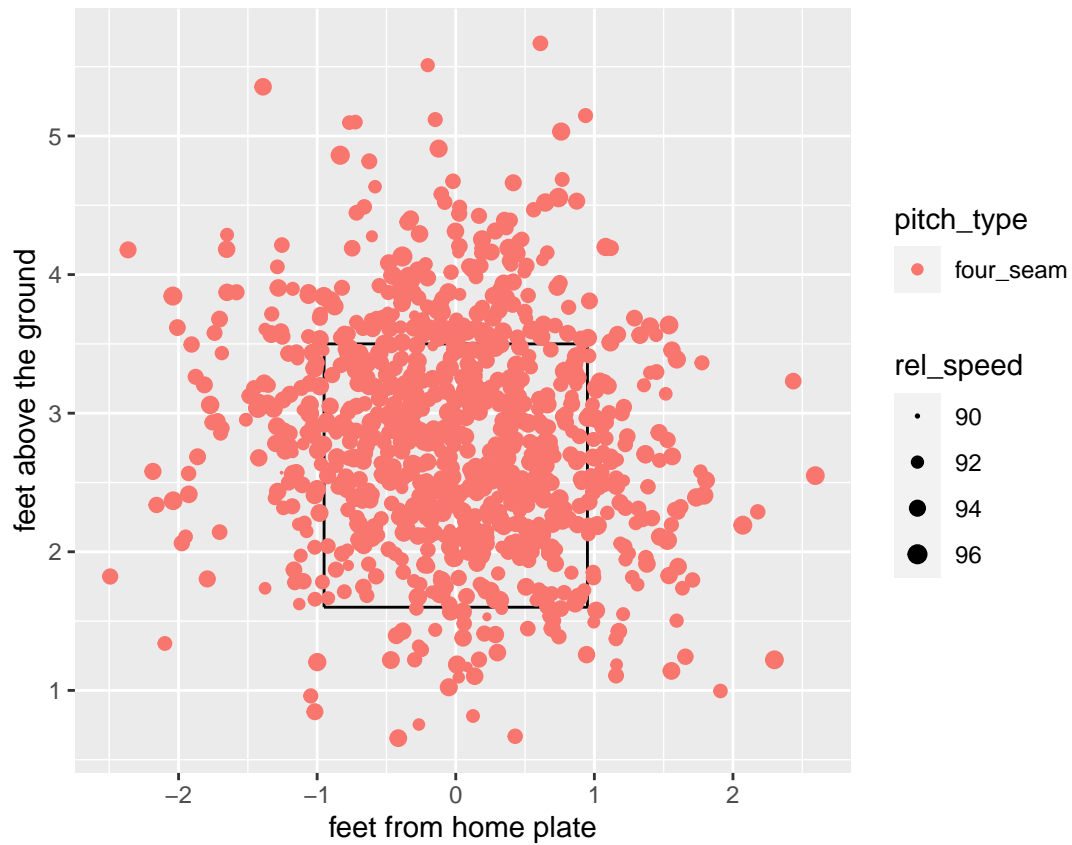
player_a Pitches(curveball) Heat map

```
ggplot()+
  geom_path(data = sz, aes(x=x, y=z))+
  coord_equal()+
  xlab("feet from home plate")+
  ylab("feet above the ground")+
  geom_point(data = curveball_b,aes(x=plate_loc_side,y=plate_loc_height,size=rel_speed,color=pitch_type
  scale_size(range = c(0.01,3))
```

player_a Pitches(four_seam) Heat map

```
ggplot()+
  geom_path(data = sz, aes(x=x, y=z))+
  coord_equal()+
  xlab("feet from home plate")+
  ylab("feet above the ground")+
  geom_point(data = four_seam_b,aes(x=plate_loc_side,y=plate_loc_height,size=rel_speed,color=pitch_type
  scale_size(range = c(0.01,3))
```

player_a Pitches(slider) Heat map

```
ggplot()+
  geom_path(data = sz, aes(x=x, y=z))+
  coord_equal()+
  xlab("feet from home plate")+
  ylab("feet above the ground")+
  geom_point(data = slider_b,aes(x=plate_loc_side,y=plate_loc_height,size=rel_speed,color=pitch_type))+
  scale_size(range = c(0.01,3))
```
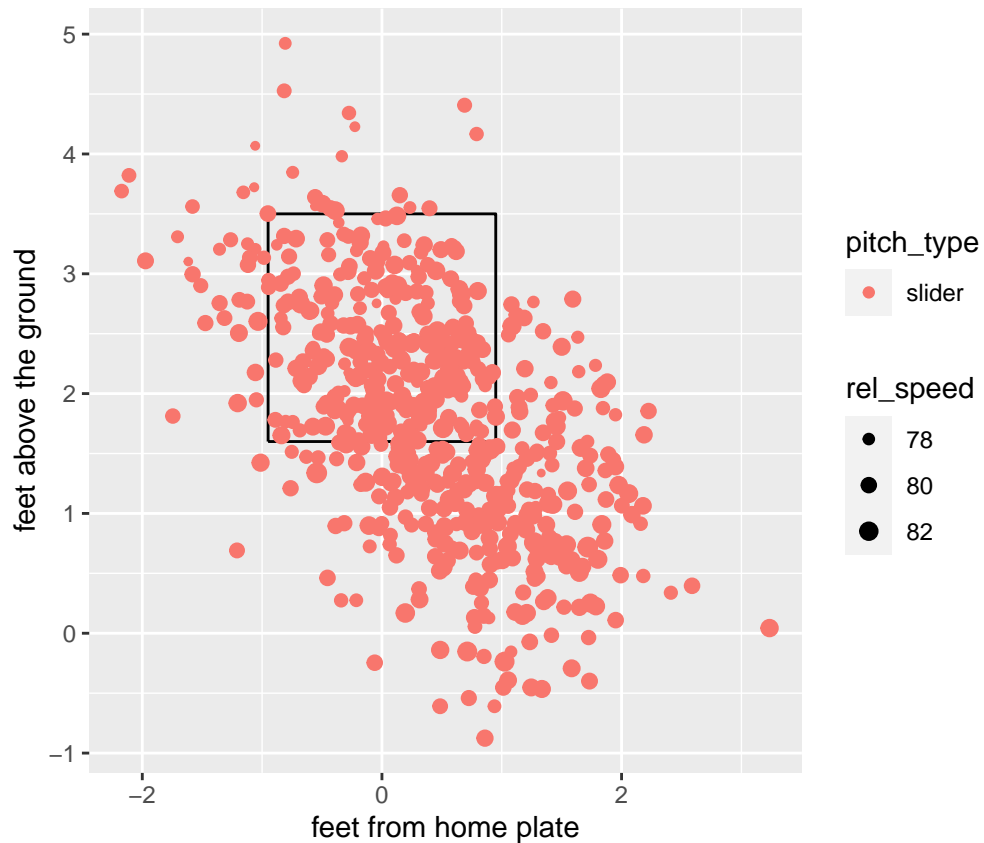
Player A Performance

```r
# Create data for the labels and Pie Chart


lbl <- c("Ball called", "Foul_ball", "Hit_by_pitch", "In_play","Strike_called","Strike_swinging")
x <- pitch_call_a


piepercent<- round(100*x/sum(x), 1)

# Give the chart file a name.
png(file = "player_a_performance.jpg")

# Plot the chart.
pie(x, labels = piepercent, main = "Player_a_Performance",col = rainbow(length(x)))
legend("topright", c("Ball called", "Foul_ball", "Hit_by_pitch", "In_play","Strike_called","Strike_swing
   fill = rainbow(length(x)))

# Save the file.
dev.off()
```

```
## pdf
##   2
```

Player B Performance

```r
# Create data for the labels and Pie Chart

lbl <- c("Ball called", "Foul_ball", "Hit_by_pitch", "In_play","Strike_called","Strike_swinging")
x <- pitch_call_b


piepercent<- round(100*x/sum(x), 1)

# Give the chart file a name.
png(file = "player_b_performance.jpg")

# Plot the chart.
pie(x, labels = piepercent, main = "Player_b_Performance",col = rainbow(length(x)))
legend("topright", c("Ball called", "Foul_ball", "Hit_by_pitch", "In_play","Strike_called","Strike_swing
    fill = rainbow(length(x)))

# Save the file.
dev.off()
```

```
## pdf
##   2
```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.