

R 야구 데이터 시각화 과정

Jung Yeon-Hun

2019/04/11

2019시즌 박진우 선수 투구 데이터 시각화

play-by-play 데이터를 기반으로 2019시즌 NC 다이노스 박진우 선수의 투구를 데이터 시각화 자료를 만들고자 합니다.

시각화에 사용할 투구 raw data는 4월 11일까지의 데이터로 내용은 다음과 같습니다.

```
park <- read.csv("C:/Users/Administrator/Desktop/backup/R/1/2019_park.csv")
knitr::kable(park[1:5, 1:16])
```

| game_date | pitch_type | speed | result | p_name | p_type | b_name | b_type | catcher | inn | pitches | balls | strikes | out | x | y |
|------------|------------|-------|--------|--------|--------|--------|--------|---------|-----|---------|-------|---------|-----|-----|-----|
| 2019-04-10 | fa | 134 | 파울 | 박진우 | 우언 | 최원준 | 좌타 | 양의지 | 1말 | 1 | 0 | 1 | 0 | 81 | 72 |
| 2019-04-10 | fa | 134 | 볼 | 박진우 | 우언 | 최원준 | 좌타 | 양의지 | 1말 | 2 | 1 | 1 | 0 | 14 | 105 |
| 2019-04-10 | ch | 118 | 포스 | 박진우 | 우언 | 최원준 | 좌타 | 양의지 | 1말 | 3 | 1 | 1 | 1 | 96 | 100 |
| 2019-04-10 | sl | 124 | 루킹 | 박진우 | 우언 | 류승현 | 좌타 | 양의지 | 1말 | 4 | 0 | 1 | 1 | 66 | 145 |
| 2019-04-10 | fa | 136 | 볼 | 박진우 | 우언 | 류승현 | 좌타 | 양의지 | 1말 | 5 | 1 | 1 | 1 | 137 | 46 |

우선 raw 데이터를 구종 별로 나누어서 시각화 작업을 하고자 합니다. 구종 별로 나누기 위해 dplyr 패키지를 사용하였습니다.

```
library(dplyr)
FA <-filter(park, pitch_type == 'fa')

SL <-filter(park, pitch_type == 'sl')

CH <-filter(park, pitch_type == 'ch')

CU <-filter(park, pitch_type == 'cu')
```

박진우 선수가 2019시즌 던진 구종은 포심, 슬라이더, 커브, 체인지업 총 4가지입니다.

투구 시각화를 위해 시각화 패키지인 ggplot2와 가상의 스트라이크 존 사진을 불러오기 위해 png패키지를 사진의 좌표를 맞추기 위해 grid 패키지를 사용하였습니다.

이 4가지 구종을 쉽게 구별할 수 있게, 베이스볼서버트에서 나타는 색깔과 같이 포심은 빨간색(#ff0000), 슬라이더는 노란색(#eee716), 커브는 하늘색(#00d1ee), 체인지업은 초록색(#1d8e3a)을 사용하였습니다.

(아래의 시각화 자료는 투수 시점으로 기록된 데이터입니다.)

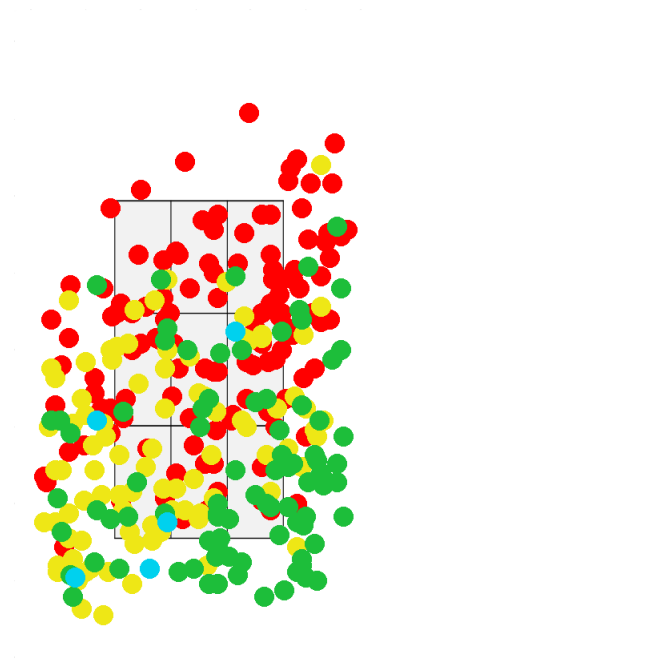
```
library(ggplot2)

library(png)

library(grid)

image <- readPNG("C:/Users/Administrator/Desktop/backup/R/1/s_zone1.png")

ggplot()+annotation_custom(rasterGrob(image, width = unit(1,"npc"), height = unit(1,"npc")), -Inf, Inf, -Inf, Inf)+theme_minimal()+xlim(0, 150) + ylim(210, 0) + scale_alpha(range=c(0.01, 0.9))+geom_point(data = FA, aes(x=x, y=y),colour = "#ff0000", size = 4)+geom_point(data = SL, aes(x=x, y=y),colour = "#eee716", size = 4)+geom_point(data = CH, aes(x=x, y=y),colour = "#1d8e3a", size = 4)+geom_point(data = CU, aes(x=x, y=y),colour = "#00d1ee", size = 4)+theme(axis.text.x=element_blank())+theme(axis.text.y=element_blank())+xlab("") + ylab("")+coord_fixed(ratio = 1.4)
```



추가로 베이스볼 서번트의 pitch Heatmap을 구현해보고자 합니다.

아래의 시각화 자료는 박진우 선수의 포심 패스트볼 pitch Heatmap 입니다.

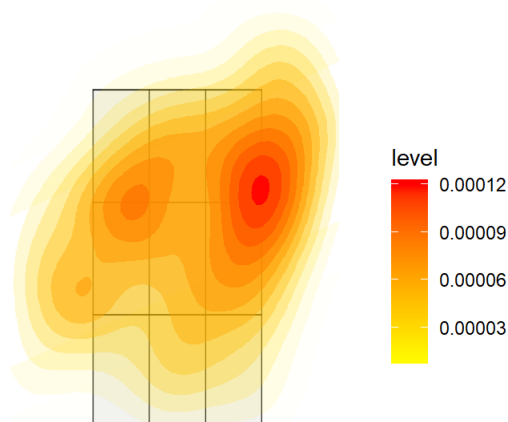
```
library(ggplot2)

library(png)

library(grid)

image <- readPNG("C:/Users/Administrator/Desktop/backup/R/1/s_zone1.png")

ggplot() + annotation_custom(rasterGrob(image, width = unit(1,"npc"), height = unit(1,"npc")), -Inf, Inf, -Inf, Inf) + stat_density_2d(data = FA, aes(x=x, y=y, fill = ..level.., alpha = ..level..), geom = "polygon") + theme_minimal() + xlim(0, 150) + ylim(210, 0) + scale_fill_gradient(low='#fffa00', high='#ff0000') + scale_alpha(range=c(0.01, 0.9), guide=F) + theme(axis.text.x=element_blank()) + theme(axis.text.y=element_blank())+xlab("") + ylab("") + coord_fixed(ratio = 1.4)
```



이상으로 통계프로그래밍 언어 R을 이용한 야구 시각화 과정의 설명을 마칩니다. 이후에도 최적화된 코드를 찾았을 시 업데이트할 예정입니다.

감사합니다.