

R ggplot2 패키지를 이용한 데이터 시각화

Jung Yeon-Hun

2019/04/11

2019시즌 박진우 선수 패스트볼 투구 분포 데이터 시각화

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

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

```
ff <- read.csv("C:/Users/Administrator/Desktop/backup/R/1/ff.csv")
knitr::kable(ff[1:5, 1:17])
```

game_code	game_date	pitch_type	speed	result	pit_name	pit_type	bat_name	bat_type	catcher	inn	total_pitch	ball	strike	out	x
20190088	2019-04-10	포심	130	내야	박진우	우언	문선재	우타	양익지	7 말	102	3	1	0	33
20190088	2019-04-10	포심	131	볼	박진우	우언	문선재	우타	양익지	7 말	101	3	1	0	11
20190088	2019-04-10	포심	132	볼	박진우	우언	문선재	우타	양익지	7 말	99	1	1	0	136
20190088	2019-04-10	포심	133	파울	박진우	우언	이명기	좌타	양익지	6 말	88	0	2	1	119
20190088	2019-04-10	포심	132	루킹	박진우	우언	이명기	좌타	양익지	6 말	86	0	1	1	83

사용한 패키지는 일반적으로 R의 시각화 패키지로 널리 쓰이는 ggplot2, 가상의 스트라이크 존의 사진 파일을 불러오기 위한 png, 마지막으로 히트 맵 효과를 주기 위해 grid 패키지를 사용하였습니다.

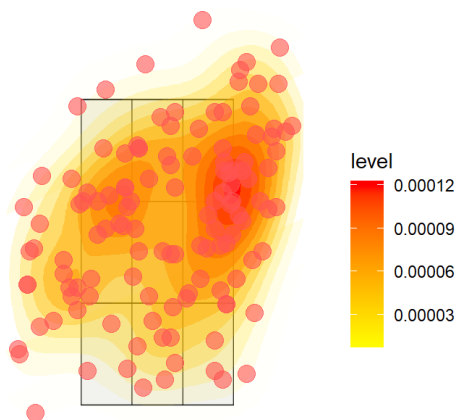
```
library(ggplot2)

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 = ff, aes(x=x, y=y, fill = ..level.., alpha = ..level..), geom = "polygon") + theme_minimal() +
  xlim(0, 150) + ylim(210, 0) + scale_fill_gradient(low="#ffa00", high="#ff0000") + scale_alpha(range=c(0.01, 0.9), guide=F) +
  geom_point(data = ff, aes(x=x, y=y), colour = "#ff5454", size = 4, alpha = 0.6) + theme(axis.text.x=element_blank()) +
  theme(axis.text.y=element_blank()) + xlab("") + ylab("") + coord_fixed(ratio = 1.4)
```



투구의 밀집한 효과를 보여주기 위해 공 색깔의 투명도는 60%로 조정하였고, 좌푯값이 y 값은 역으로 기록되어 역으로 맞추는 작업도 하였고, 투구를 기록한 데이터의 가로 세로 비율도 1:1 비율이 아니었기에 이 또한 보정하였습니다.