R 개인프로젝트

yoon

2020 5 22

# 2017 LG는 왜 포스트시즌에 가지 못하였나

### LG는 2017시즌 팀 방어율 4.30으로 전체 1위를 했음에도 불구하고 상위 5팀만이 할 수 있는 포스트시즌에 진출하지 못했다. 투수진들은 다른 팀에 비해 좋은 성적을 냈지만, 그럼에도 저조한 순위를 기록한 이유는 타자들의 부진을 가장 큰 원인으로 꼽을 수 있다. 2017시즌 타자들의 스탯 기록과 팀 기록을 통해 LG가 상위팀으로 갈 수 없었던 이유를 알아보도록 하겠다.

## 1. 주전 라인업의 잦은 변경

### 야구에서는 타선의 주전 라인업 변경이 적을수록 일정한 타격 시너지를 기대할 수 있다. 반대로 부진하는 선수가 많아 타선 라인업이 자주 바뀌면 새로운 선수를 찾을 가능성은 있지만 일정한 타격 시너지를 얻은 확률이 매우 떨어진다. 주전 라인업의 변경이 적으면 적을수록 규정 타석 달성 선수가 많아지고 반대로 변경이 잦으면 규정 타석을 못 채우는 선수가 더 많아진다. 팀 별로 규정 타석 달성 인원을 구해보고 팀 순위와 비교해보자.

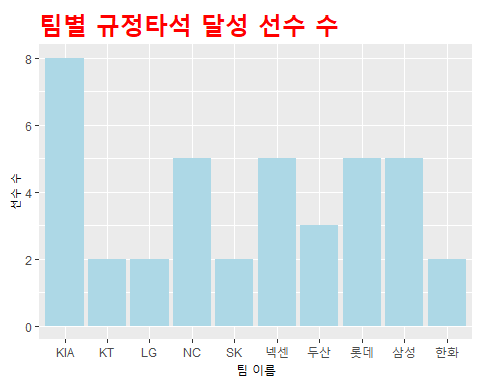
library(dplyr)

##   
## 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)  
  
setwd('C:\\Workspace\\WorkspaceR')  
base <- read.csv('batter\_stats\_2017.csv', header = T)  
  
base1 <- base %>% mutate(규정타석=ifelse(타석>=446, 'pass', 'fail') )  
base2 <- base1 %>% filter(규정타석=='pass')  
  
ggplot(base2, aes(x=팀명))+  
 geom\_bar(fill='lightblue')+  
 ggtitle('팀별 규정타석 달성 선수 수')+  
 theme(plot.title = element\_text(size = 24,  
 face = 'bold',  
 colour = 'red'))+  
 labs(x='팀 이름', y= '선수 수')



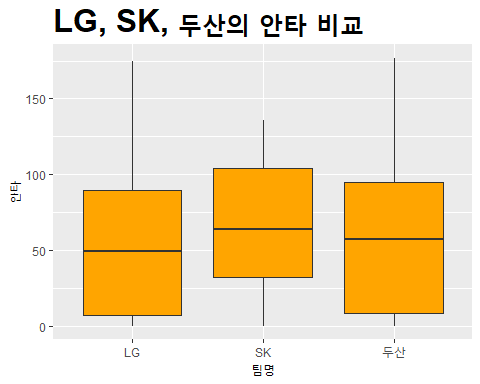
### 이를 통해 알 수 있 듯 우승팀인 기아의 규정타석 달성 선수는 8명인데 비해, LG의 규정타석 달성 선수는 2명에 불과하다. 이는 확실한 주전이 없어 타팀에 비해 타선의 짜임새가 부족하다고 판단할 수 있다. 하지만 규정타석 달성 선수가 LG와 SK가 2명으로 같고, 두산은 3명으로 1명만 많은데 SK와 두산은 모두 포스트시즌 진출에 성공하였다. 무슨 차이가 있기에 LG만 포스트시즌 진출에 실패한것인가?

## 2. SK,두산과 LG의 비교 ( 경기당 안타, 홈런, 볼넷, 득점, 타율 )

options(max.print = 9999999)  
base.sk<- base %>% filter(팀명=='SK')  
base.ds <- base %>% filter(팀명=='두산')  
base.lg <- base %>% filter(팀명=='LG')  
  
base.sk.ds.lg<- rbind(base.sk,base.ds,base.lg)

### 1) 안타 수 비교

ggplot(base.sk.ds.lg,aes(x=팀명,y=안타))+  
 geom\_boxplot(fill='orange')+  
 ggtitle('LG, SK, 두산의 안타 비교')+  
 theme(plot.title = element\_text(size = 24,  
 face='bold'))



base.sk %>% select(안타) %>% sum/144 # 경기 당 안타

## [1] 9.145833

base.ds %>% select(안타) %>% sum/144

## [1] 7.048611

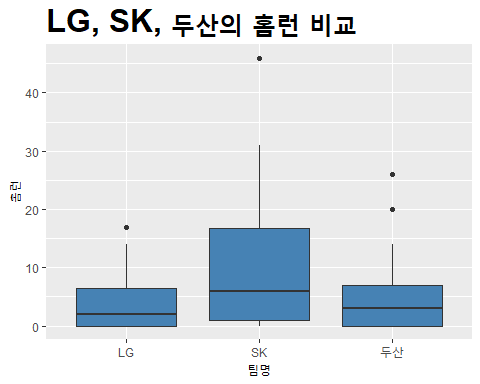
base.lg %>% select(안타) %>% sum/144

## [1] 7.326389

### 세 팀 간 안타를 비교해보면 LG와 두산의 수치가 비슷하다. 경기당 안타 측면에서는 오히려 LG가 두산보다 높은 수치를 나타내고 있다.

### 2) 홈런 수 비교

ggplot(base.sk.ds.lg,aes(x=팀명,y=홈런))+  
 geom\_boxplot(fill='steelblue')+  
 ggtitle('LG, SK, 두산의 홈런 비교')+  
 theme(plot.title = element\_text(size = 24,  
 face='bold'))



base.sk %>% select(홈런) %>% sum/144 # 경기 당 홈런

## [1] 1.555556

base.ds %>% select(홈런) %>% sum/144

## [1] 0.7083333

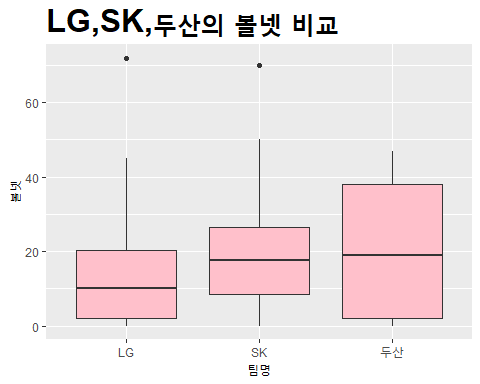
base.lg %>% select(홈런) %>% sum/144

## [1] 0.5833333

### 세팀의 홈런을 비교해보면 SK가 세 팀 중 가장 높은 수치를 기록하고 있고, LG와 두산은 SK에 비해 낮은 수치를 보이고 있다. LG와 두산을 비교해보면 앞서 경기당 안타는 LG가 많았지만, 경기당 홈런은 두산이 더 많아 적은 안타에도 더 많은 점수를 낼 수 있었다.

### 3) 볼넷 비교

ggplot(base.sk.ds.lg,aes(x=팀명,y=볼넷))+  
 geom\_boxplot(fill='pink')+  
 ggtitle('LG,SK,두산의 볼넷 비교')+  
 theme(plot.title = element\_text(size = 24,  
 face='bold'))



base.sk %>% select(볼넷) %>%sum/144 # 경기 당 볼넷

## [1] 2.923611

base.ds %>% select(볼넷) %>%sum/144

## [1] 2.430556

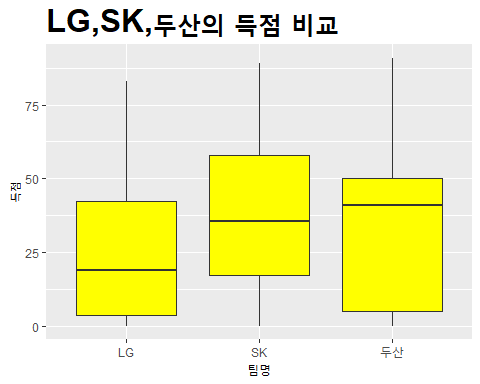
base.lg %>% select(볼넷) %>%sum/144

## [1] 2.256944

### 볼넷 수를 비교해보면 LG 선수들의 평균 볼넷 수가 두 팀에 비해 현저히 낮음을 알 수 있다. 또한 두산은 볼넷을 얻어 낸 타자가 LG와 SK에 비해 비교적 많은 것을 알 수 있다. 이는 두산이 안타가 적어도 출루를 많이 하여 점수를 낼 가능성이 높음을 의미한다.

### 4) 득점 비교

ggplot(base.sk.ds.lg,aes(x=팀명,y=득점))+  
 geom\_boxplot(fill='yellow')+  
 ggtitle('LG,SK,두산의 득점 비교')+  
 theme(plot.title = element\_text(size = 24,  
 face='bold'))



base.sk.ds.lg %>% filter(팀명=='SK') %>% select(득점) %>% sum/144

## [1] 5.180556

base.sk.ds.lg %>% filter(팀명=='두산') %>% select(득점) %>% sum/144

## [1] 4.006944

base.sk.ds.lg %>% filter(팀명=='LG') %>% select(득점) %>% sum/144

## [1] 3.645833

### 세 팀 타자들의 득점을 비교하면 LG 타자들의 득점이 다른 두 팀의 타자들보다 확실히 낮음을 알 수 있다.

### 규정 타석 달성 타자 수가 비슷한 세 팀의 안타, 홈런, 볼넷, 득점을 비교해 본 결과, 경기당 안타는 세 팀이 비슷했지만 홈런이나 볼넷을 LG 선수들이 적게 얻어냄으로써 SK와 두산에 비해 득점이 적었음을 알 수 있다. 이러한 차이가 팀 순위의 차이를 만들어 낸 것이라 할 수 있다.

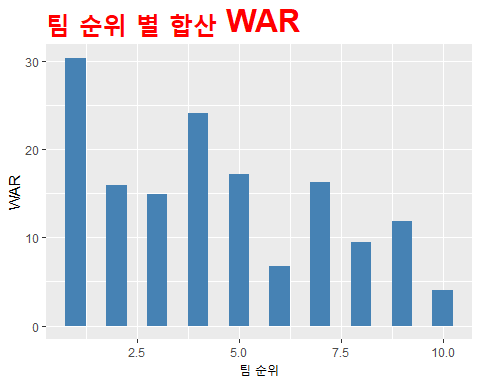
## 3. 팀 순위별 WAR 합 비교

### WAR이란 대체선수 대비 승리 기여도로 WAR이 6인 선수가 WAR이 0인 선수보다 6승을 더 얻어낸 것으로 보면 된다.

base.kia<- base %>% filter(팀명=='KIA')  
base.ds <- base %>% filter(팀명=='두산')  
base.lt <- base %>% filter(팀명=='롯데')  
base.nc<- base %>% filter(팀명=='NC')  
base.sk <- base %>% filter(팀명=='SK')  
base.lg <- base %>% filter(팀명=='LG')  
base.nx<- base %>% filter(팀명=='넥센')  
base.hw <- base %>% filter(팀명=='한화')  
base.ss <- base %>% filter(팀명=='삼성')  
base.kt <- base %>% filter(팀명=='KT')  
  
war <- c(sum(base.kia$WAR),sum(base.ds$WAR),sum(base.lt$WAR),sum(base.nc$WAR),sum(base.sk$WAR),  
 sum(base.lg$WAR),sum(base.nx$WAR),sum(base.hw$WAR),sum(base.ss$WAR),sum(base.kt$WAR))  
  
team <- c(1,2,3,4,5,6,7,8,9,10)  
names(team) <- c('KIA','두산','롯데','NC','SK','LG','넥센','한화','삼성','KT')  
  
df.base <- data.frame(team,war)  
df.base

## team war  
## KIA 1 30.36  
## 두산 2 15.91  
## 롯데 3 14.85  
## NC 4 24.06  
## SK 5 17.16  
## LG 6 6.80  
## 넥센 7 16.23  
## 한화 8 9.49  
## 삼성 9 11.81  
## KT 10 3.98

ggplot(df.base,aes(x=team, y=war))+  
 geom\_bar(stat = 'identity',   
 width = 0.5,   
 fill='steelblue') +   
 ggtitle('팀 순위 별 합산 WAR')+  
 theme(plot.title = element\_text(size = 24,  
 face = 'bold',  
 colour = 'red'))+  
 labs(x="팀 순위", y= 'WAR')



cor(war,team) # 등수와 war의 상관계수가 -0.7640916 매우 강한 상관 관계가 있음을 알 수 있다.

## [1] -0.7640916

### 팀별로 각 선수들의 WAR의 합을 비교해본 결과 6위 LG는 신생팀인 10위 KT를 제외하고 최하위에 위치하고 있다. 1위 KIA의 경우 WAR의 합이 30이 넘는데, 이는 KIA 선수들이 대체 선수에 비해 30승을 더 챙겼다는 의미다. 반면 LG 선수들의 WAR 합은 6.8이므로 대체 선수에 비해 6~7승 정도만 더 챙겼음을 의미한다. 이는 LG 선수들이 타 팀 선수들에 비해 승리에 기여하는 정도가 전체적으로 낮다고 할 수 있다.

## 4. 4. 피타고리안 승률과 실제 순위 비교

### 피타고리안 승률이란 득점의 제곱을 득점의 제곱과 실점의 제곱의 합으로 나눈 수로 팀이 얻은 득점과 잃은 실점으로 승률을 예상하는 용도로 쓰인다.

team

## KIA 두산 롯데 NC SK LG 넥센 한화 삼성 KT   
## 1 2 3 4 5 6 7 8 9 10

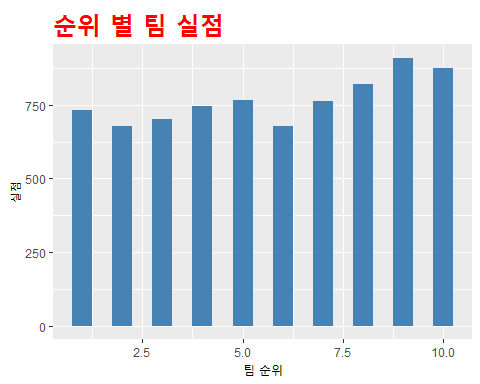
runs.hit <- c(906,849,743,786,761,699,789,737,757,655)  
runs.pitch <- c(734,678,701,745,767,677,764,820,911,876)  
  
  
  
runs<- data.frame(team,runs.hit,runs.pitch)  
runs

## team runs.hit runs.pitch  
## KIA 1 906 734  
## 두산 2 849 678  
## 롯데 3 743 701  
## NC 4 786 745  
## SK 5 761 767  
## LG 6 699 677  
## 넥센 7 789 764  
## 한화 8 737 820  
## 삼성 9 757 911  
## KT 10 655 876

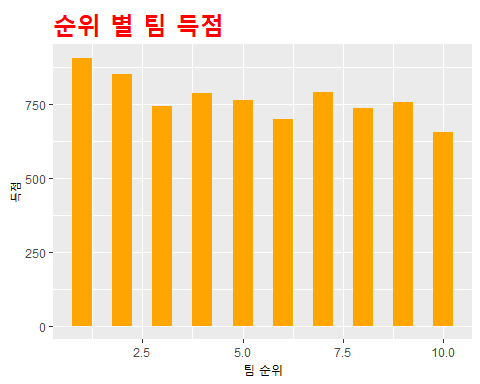
runs1<- runs %>% mutate(피타고리안\_승률=runs.hit^2/(runs.hit^2+runs.pitch^2))  
runs1

## team runs.hit runs.pitch 피타고리안\_승률  
## 1 1 906 734 0.6037370  
## 2 2 849 678 0.6105973  
## 3 3 743 701 0.5290613  
## 4 4 786 745 0.5267607  
## 5 5 761 767 0.4960734  
## 6 6 699 677 0.5159843  
## 7 7 789 764 0.5160937  
## 8 8 737 820 0.4468434  
## 9 9 757 911 0.4084542  
## 10 10 655 876 0.3585963

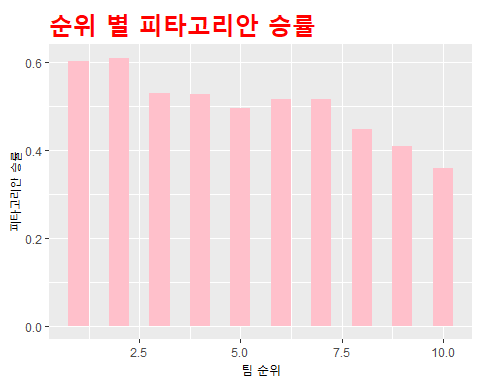
ggplot(runs1,aes(x=team,y=runs.pitch))+  
 geom\_bar(stat = 'identity',   
 width = 0.5,   
 fill='steelblue')+  
 ggtitle('순위 별 팀 실점')+  
 theme(plot.title = element\_text(size = 24,  
 face = 'bold',  
 colour = 'red'))+  
 labs(x='팀 순위', y='실점')



ggplot(runs1,aes(x=team,y=runs.hit))+  
 geom\_bar(stat = 'identity',   
 width = 0.5,   
 fill='orange')+  
 ggtitle('순위 별 팀 득점')+  
 theme(plot.title = element\_text(size = 24,  
 face = 'bold',  
 colour = 'red'))+  
 labs(x='팀 순위', y='득점')



ggplot(runs1,aes(x=team,y=피타고리안\_승률))+  
 geom\_bar(stat = 'identity',   
 width = 0.5,   
 fill='pink')+  
 ggtitle('순위 별 피타고리안 승률')+  
 theme(plot.title = element\_text(size = 24,  
 face = 'bold',  
 colour = 'red'))+  
 labs(x='팀 순위', y='피타고리안 승률')



library(corrplot)

## corrplot 0.84 loaded

cor(runs1)

## team runs.hit runs.pitch 피타고리안\_승률  
## team 1.0000000 -0.7692824 0.7976307 -0.9385627  
## runs.hit -0.7692824 1.0000000 -0.3889146 0.8073831  
## runs.pitch 0.7976307 -0.3889146 1.0000000 -0.8566162  
## 피타고리안\_승률 -0.9385627 0.8073831 -0.8566162 1.0000000

corrplot(cor(runs1),method = "number")



### LG는 피타고리안 승률이 0.515로 5위이지만, 실제승률은 0.489로 피타고리안 승률보다 낮은 수치를 기록하였다. 이는 실점은 적지만 득점이 다른 경쟁 팀에 비해 매우 떨어져서 나온 결과라 할 수 있다.