Lesson_27_Boardsheet

Kevin Cummiskey

4/1/2020

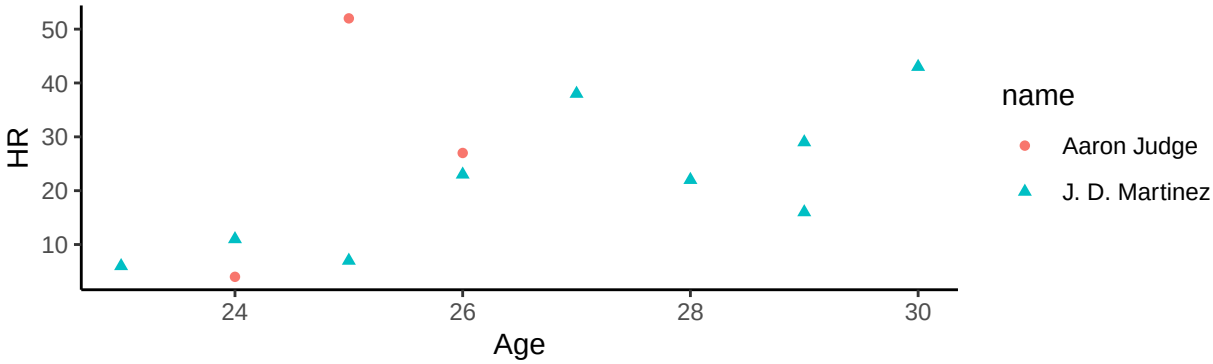
Review

Let's look at J.D. Martinez and Aaron Judge.

Here are their career statistics.

Table 1: Career Statistics (through 2018 season)

name	playerID	HR	BB	AB	PA
Aaron Judge	judgeaa01	83	212	1039	1271
J. D. Martinez	martijd02	195	317	3397	3765



If Aaron Judge had the same number of plate appearences ($newPA = 3765$) as J.D. Martinez, how many career home runs would Aaron Judge have with $kicker = 1.05$?

Figure 1: Yaz

OPS

A common statistic reported for batters is On Base Percentage Plus Slugging.

On Base Percentage Plus Slugging = On Base Percentage (OBP) + Slugging Percentage (SLG).

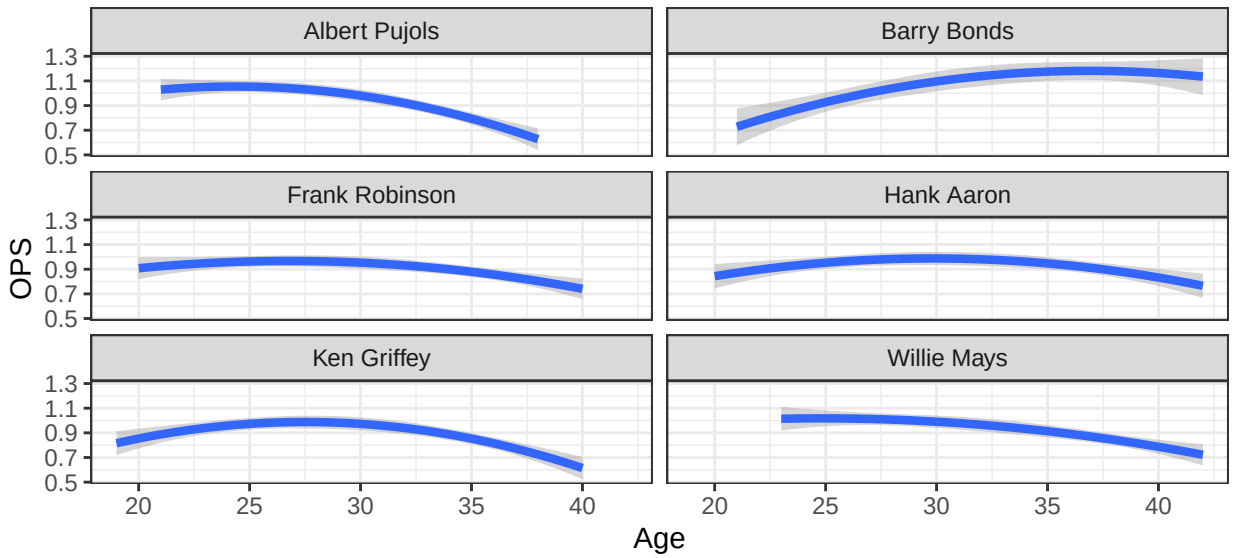
On Base Percentage = $OBP = \frac{H+BB+HBP}{AB+BB+HBP+SF}$.

Slugging Percentage = $SLG = \frac{TB}{AB} = \frac{1B+2*2B+3*3B+4*HR}{AB}$

Carl Yastrzemski

Carl Yastrzemski ("Yaz") played his entire 23 year career (1961-1983) with the Boston Red Sox. Using the *similar* function of Chapter 8, here are six players with similar career statistics as Yaz.

```
plot_trajectories("Carl Yastrzemski", n.similar = 6)
```



Albert Pujols

Albert Pujols has played 18 seasons (2001 - Present) with the St. Louis Cardinals and Los Angeles Angels.

Yaz vs Pujols

Their career numbers are pretty similar.

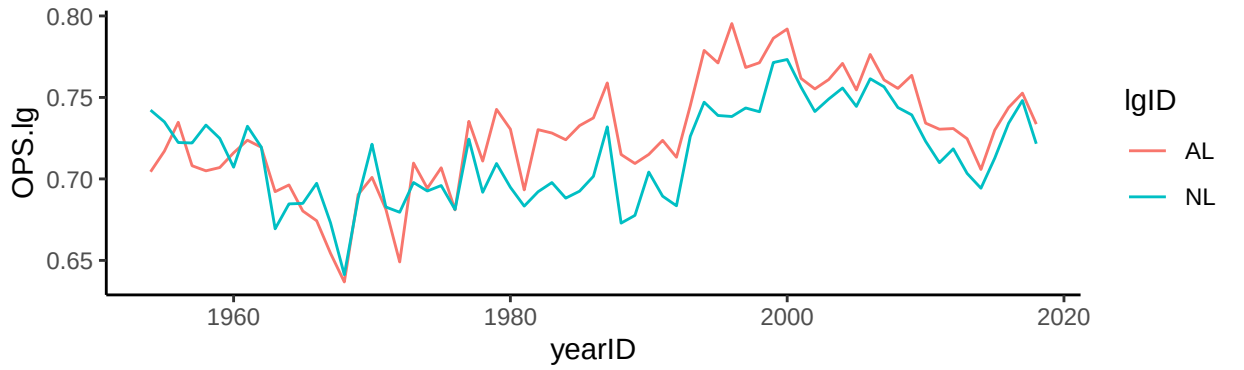
```
player.ids <- c("yastrca01","pujolal01")
Batting %>%
  filter(playerID %in% player.ids) %>%
  group_by(playerID) %>%
  summarize(H = sum(H),
            AB = sum(AB),
            HR = sum(HR),
            SLG = (sum(H) - sum(X2B) - sum(X3B) - sum(HR) +
             2 * sum(X2B) + 3*sum(X3B) + 4 * sum(HR))/sum(AB),
           OBP = (sum(H) + sum(BB) + sum(HBP))/(sum(AB) + sum(BB)+ sum(HBP) + sum(SF))) %>%
  mutate(OPS = SLG + OBP,
         AVG = H/AB) %>%
  left_join(Master %>% select(nameLast, nameFirst, playerID)) %>%
  mutate(name = paste(nameFirst, nameLast, sep = " ")) %>%
  select(name,everything(),-nameLast,-nameFirst) -> player.careers

player.careers %>%
  kable(caption = "Career Totals", digits = 3)
```

Table 2: Career Totals

name	playerID	H	AB	HR	SLG	OBP	OPS	AVG
Albert Pujols	pujolal01	3082	10196	633	0.554	0.382	0.936	0.302
Carl Yastrzemski	yastrca01	3419	11988	452	0.462	0.379	0.841	0.285

Why isn't it fair to directly compare Yaz and Pujols?



OPS+

Instead, we could use a statistic that adjusts for the overall offense in the league. OPS+ is such a statistic.

$$OPS+ = 100 \times \left(\left(\frac{OBP_{\text{player}}}{OBP_{\text{league}}} \right) + \left(\frac{SLG_{\text{player}}}{SLG_{\text{league}}} \right) - 1 \right)$$

where OBP_{league} and SLG_{league}

Let's compare the OPS+ of Yaz and Pujols.

```
#Calculate OPS
Batting %>%
  filter(playerID %in% player.ids) %>%
  mutate(SLG = (H - X2B - X3B - HR +
              2 * X2B + 3*X3B + 4 * HR)/AB,
            OBP = (H + BB + HBP)/(AB + BB+ HBP + SF),
            OPS = SLG + OBP,
            AVG = H/AB) %>%
  left_join(Master %>% select(nameLast, nameFirst, playerID)) %>%
  mutate(name = paste(nameFirst, nameLast, sep = " "))  -> yearly.stats

#join with league stats
yearly.stats %>%
  select(name, AB, H, HR, AVG, OPS) %>%
  head(5)
```

```
##                name  AB   H HR       AVG       OPS
## 1 Carl Yastrzemski 583 155 11 0.2658662 0.7207194
## 2 Carl Yastrzemski 646 191 19 0.2956656 0.8316623
## 3 Carl Yastrzemski 570 183 14 0.3210526 0.8937295
## 4 Carl Yastrzemski 567 164 15 0.2892416 0.8251425
## 5 Carl Yastrzemski 494 154 20 0.3117409 0.9318678
```

```
lg.stats %>%
  head(5)
```

```
## # A tibble: 5 x 5
## # Groups:   yearID [3]
##   yearID lgID  SLG.lg OBP.lg OPS.lg
```

```
##    <int> <fct>  <dbl>  <dbl>  <dbl>
## 1   1954 AL     0.373  0.331  0.704
## 2   1954 NL     0.407  0.335  0.742
## 3   1955 AL     0.381  0.336  0.717
## 4   1955 NL     0.407  0.328  0.735
## 5   1956 AL     0.394  0.341  0.735
```

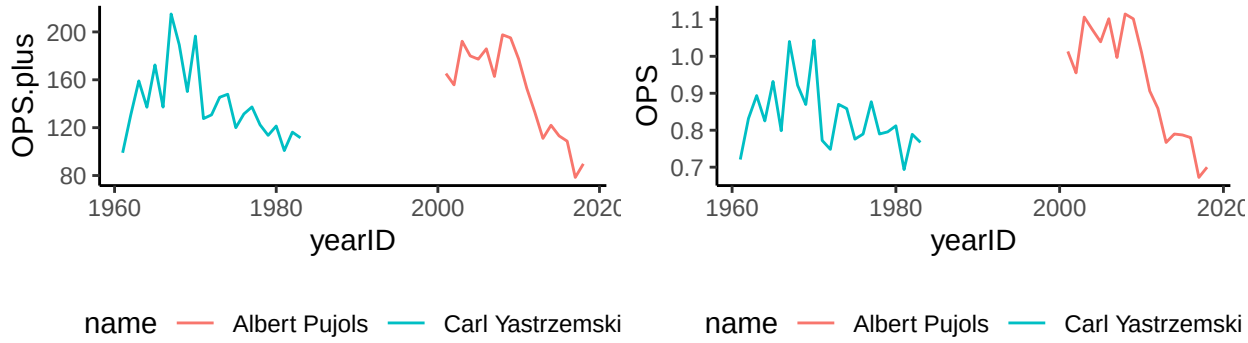
```r
yearly.stats %>%
  left_join(lg.stats, by = c("yearID","lgID")) -> yearly.stats

#calculate OPS+
yearly.stats %>%
  mutate(OPS.plus = 100 * ((OBP/OBP.lg + SLG/SLG.lg) - 1)) -> yearly.stats

library(gridExtra)
p1 = yearly.stats %>%
  ggplot(aes(x = yearID, y = OPS.plus, color = name)) +
  geom_line() + theme_bw() + theme_classic() +
  labs("OPS+") + theme(legend.position = "bottom")

p2 = yearly.stats %>%
  ggplot(aes(x = yearID, y = OPS, color = name)) +
  geom_line() + theme_bw() + theme_classic() +
  labs("OPS") +theme(legend.position = "bottom")

grid.arrange(p1,p2, ncol = 2)
```



What are some limitations of OPS+?

Adj OPS+