chapter05_package_plyr

Observation of 0.300 Batting Avg. for 40 years

3 3 .

3 , Batting 1977 2016 40 400 3 .

1. Make 'env' variable

```
library(Lahman)
library(plyr)
a= subset(Batting, yearID>1975&yearID<2017, select=c('yearID', 'AB', 'H', 'G'))
a$avg = a$H/a$AB
a =na.omit(a)
func= function(a){ return(data.frame(
    sd=sd(a$avg[a$AB>400]), mean=mean(a$avg[a$AB>400])
))}
env = ddply(a, .(yearID), func)
```

2. scoring and percentile

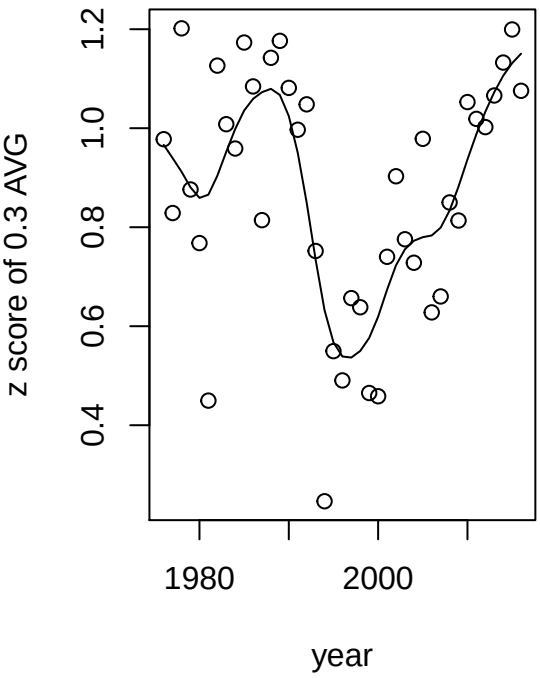
```
env$z = (0.3 - env$mean)/env$sd
env$per = pnorm(0.3, env$mean, env$sd, lower.tail = TRUE)
```

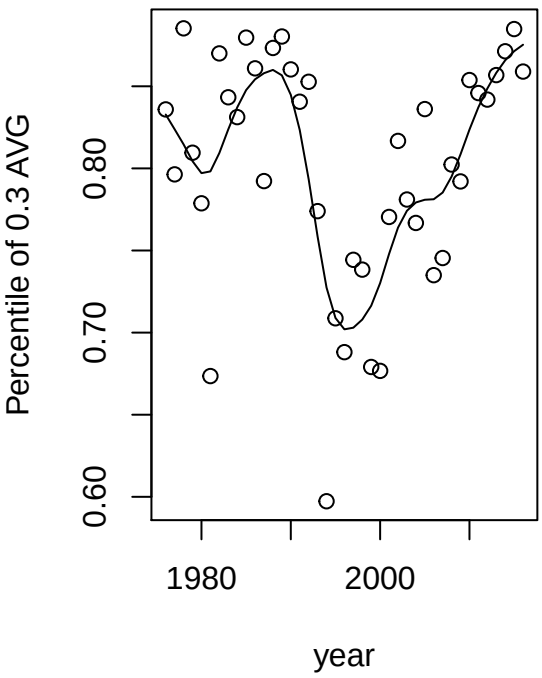
```
head(env, 10)
```

```
##    yearID         sd      mean         z       per
## 1    1976 0.02908533 0.2715596 0.9778256 0.8359197
## 2    1977 0.02723091 0.2774345 0.8286712 0.7963547
## 3    1978 0.02319787 0.2721217 1.2017613 0.8852720
## 4    1979 0.02590965 0.2772974 0.8762204 0.8095449
## 5    1980 0.02757441 0.2788247 0.7679339 0.7787368
## 6    1981 0.03378734 0.2848054 0.4497122 0.6735410
## 7    1982 0.02421480 0.2727222 1.1264909 0.8700211
## 8    1983 0.02580467 0.2739859 1.0081151 0.8433004
## 9    1984 0.02716957 0.2739446 0.9589906 0.8312183
## 10   1985 0.02571243 0.2698359 1.1731348 0.8796291
```

3. Draw timeline graphs

```
par(mfrow=c(1,2))
plot(env$yearID, env$z, xlab='year', ylab='z score of 0.3 AVG')
lines(smooth.spline(env$yearID, env$z))
plot(env$yearID, env$per, xlab='year', ylab='Percentile of 0.3 AVG')
lines(smooth.spline(env$yearID, env$per))
```


z score of 0.3 AVG
0.4
0.6
0.8
1.0
1.2
1980
2000
year


Percentile of 0.3 AVG
0.60
0.70
0.80
1980
2000
year