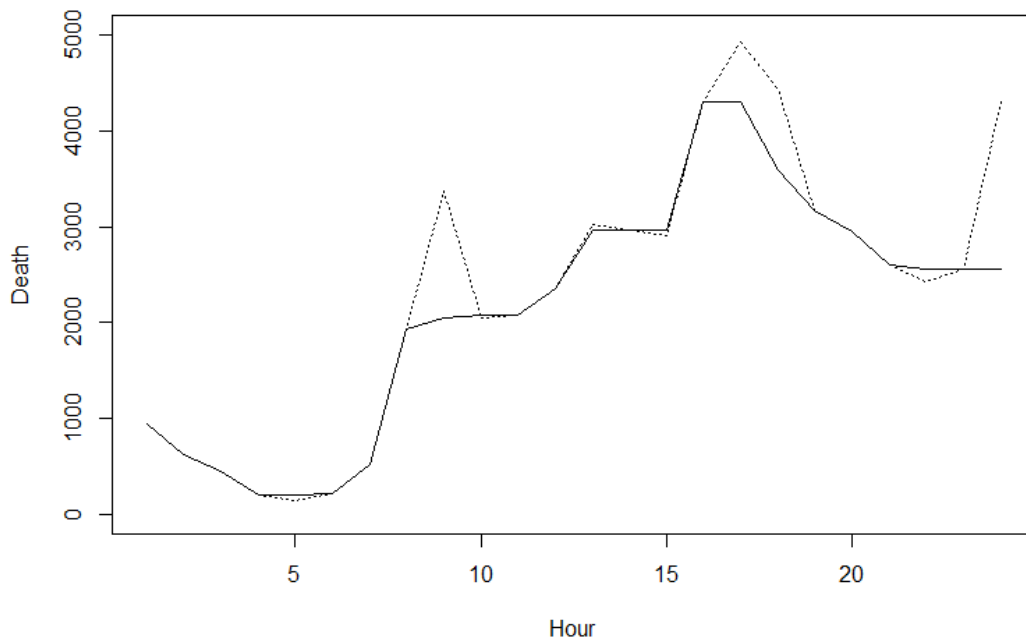


1. FRIDAY.DAT 시간에 따른 사망자 수를 평활하고 시간과 사망자 수의 관계를 서술하여라.  
(자료에는 나타나 있지 않으나 음주 운전을 염두에 두고 분석해 보면?)

```
f <-  
      V1  V2  
1    0-1  938  
2    1-2  621  
3    2-3  455  
4    3-4  207  
5    4-5  138  
6    5-6  215  
7    6-7  526  
8    7-8 1933  
9    8-9 3377  
10   9-10 2045  
11  10-11 2078  
12  11-12 2351  
13  12-13 3015  
14  13-14 2966  
15  14-15 2912  
16  15-16 4305  
17  16-17 4923  
18  17-18 4427  
19  18-19 3164  
20  19-20 2950  
21  20-21 2601  
22  21-22 2420  
23  22-23 2557  
24  23-0  4319
```

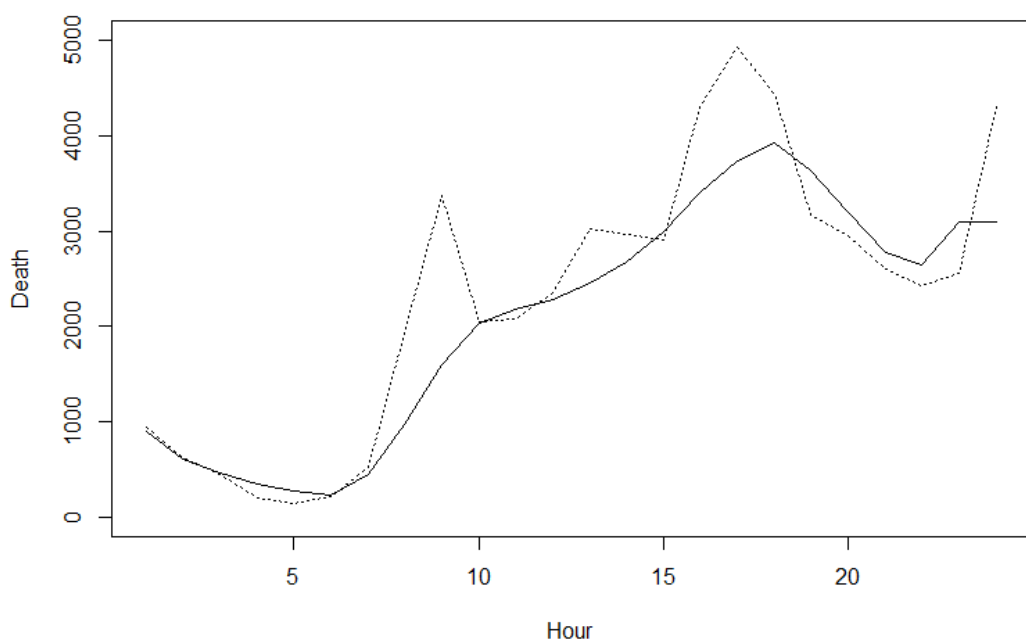
#3RSS, twice 방법

```
plot(ts(smooth(f[,2], kind="3RS3R", twiceit=TRUE)),xlab="Hour",ylab="Death",ylim=c(0,5000))  
lines(f[,2],lty="dotted")
```

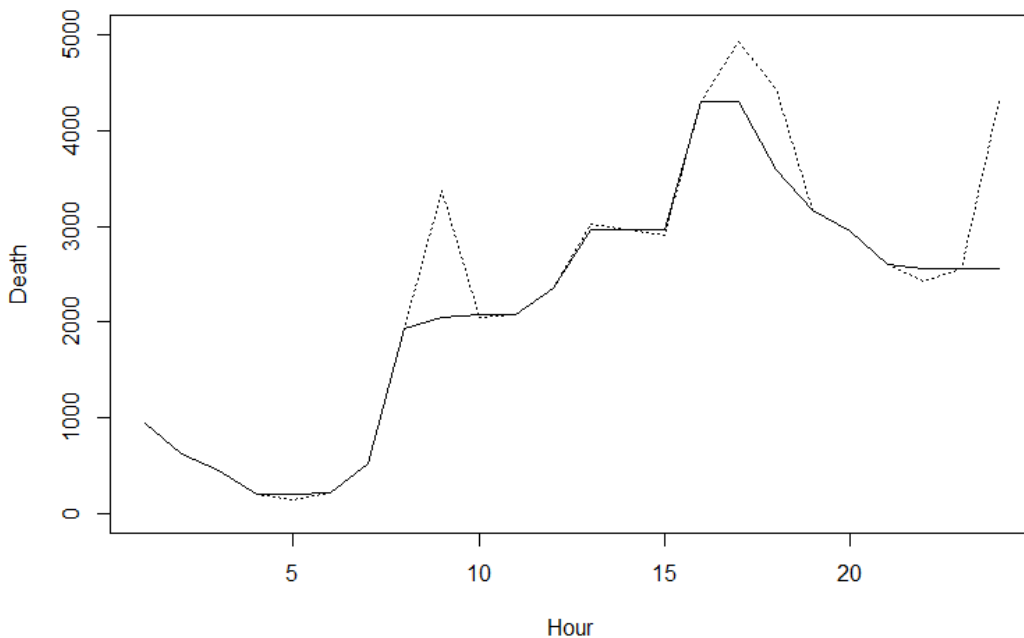


실선이 3RS3R twice 평활을 적용한 자료이다. 평활효과로 인해 8-9시와 23-0시의 높은 사망자 수는 평활 효과로 인해 고점 추세가 거의 사라진 것을 확인할 수 있다.

```
library(sleekts)
ts.plot(sleek(f[,2]),xlab="Hour",ylab="Death",ylim=c(0,5000))
lines(f[,2],lty="dotted")
```



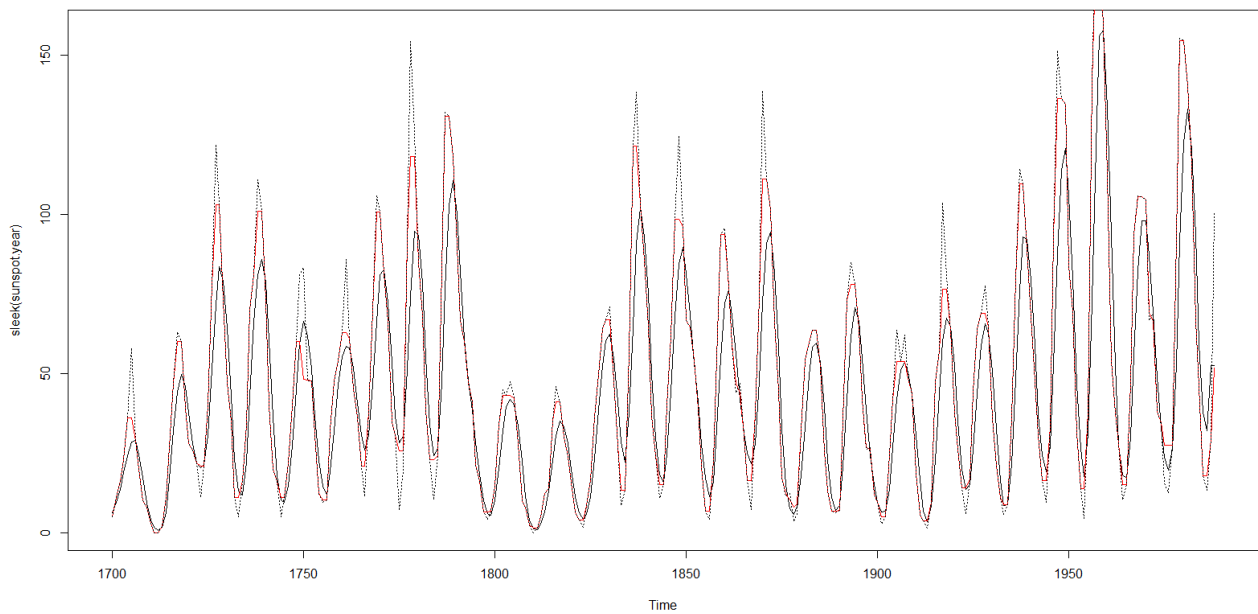
실선은 4253H twice평활을 적용한 자료이다. 3RS3R twice에 비해 자료가 곡선에 가까워졌으나 더 많은 시점(8-9시, 13-14시, 15-17시)에서 높은 사망자 수를 반영하지 못하고 있다. 4253H twice 평활은 자료의 차이를 부드럽게 만들어 추세를 파악하기에 용이하지만 현재 자료는 특정 시간대에 사망자가 많아지는 특수한 원인이 발견될 수 있다. 따라서 시점의 특징이 덜 희석되는 3RS3R twice 방식이 더 적절할 것으로 생각된다.



3RS3R twice 평활된 결과를 보면 15-17시 사망자 수가 가장 높다. 또한 추세는 4-5시에 저점이었다가 15-17시까지 점점 높아지는 것이 확인된다. 이러한 경향은 절대적인 운전중 차량 대수가 새벽엔 거의 없고 일과시간이 될수록 점점 많아지기 때문일 것으로 생각된다. 또한 평활되면서 나타나지 않게 된 8-9시와 23-0의 고점시간도 확인할 필요가 있다. 8-9시는 출퇴근 시간 차량 수 증가로 인해 사고가 많아졌을 것이고, 23-0시는 밤 중 음주운전 차량이 많아지기 때문일 것으로 생각된다.

## 2. R의 datasets에 있는 sunspot.year 자료를 평활하고 그 패턴을 기술하라.

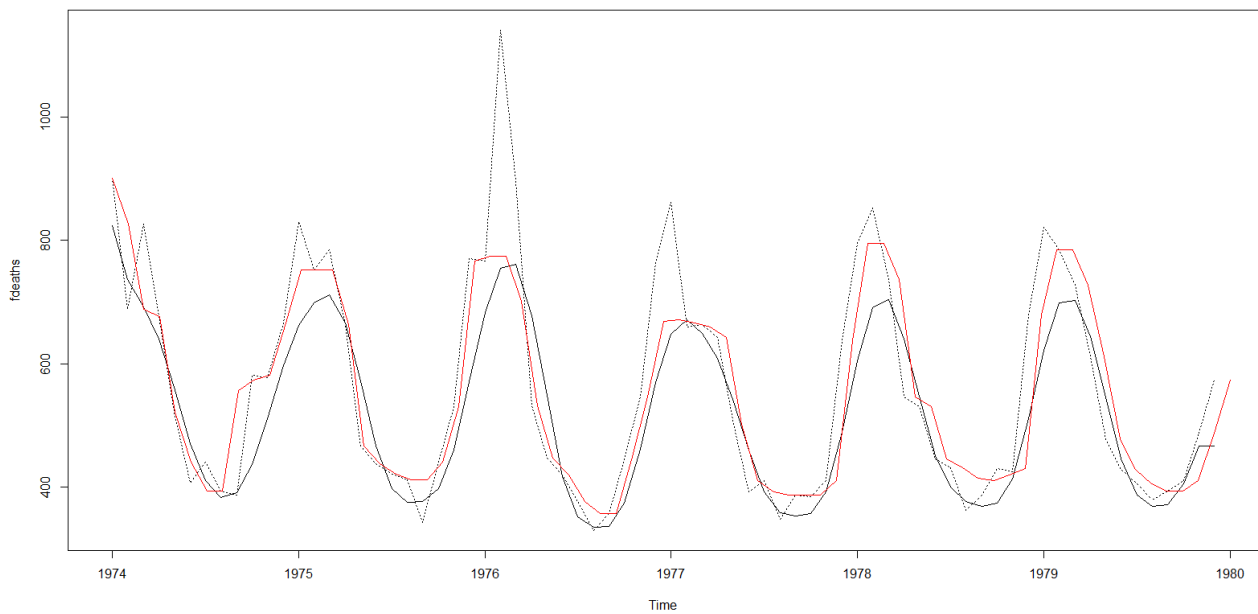
```
ts.plot(sleek(sunspot.year))
lines((1700:1988),ts(smooth(sunspot.year, kind="3RS3R", twiceit=TRUE)), col="red")
lines(sunspot.year,lty="dotted")
```



점선은 원자료, 붉은색은 3RS3R twice, 검은색은 4253H twice 평활이다. 그래프를 보면 3RS3R twice 평활보다 4253H twice 평활이 더 최댓값과 최솟값의 차이가 크게 줄어들었음이 확인된다. 두 평활 결과를 종합해 판단해보면 sunspot은 비교적 비슷한 변동으로 계절적 주기를 반복하고 있음을 알 수 있다. 대략 50년을 주기로 고점과 저점의 반복이 나타난다. 다만 1700-1730년, 1800-1830년, 1900-1910년은 변동의 크기가 상대적으로 낮았음이 확인된다. 즉 약 100년을 주기로 변동이 작아지는 계절성이 있음을 알 수 있다. 전체적인 증감 추세는 크지 않으나 1900년 이후 조금 증가 추세가 있는 것으로 보인다.

### 3. R의 fdeaths 자료를 평활하고 분석하여라.

```
plot(fdeaths, lty="dotted")
lines(seq(from=1974,to=1980,length=72),ts(smooth(fdeaths, kind="3RS3R", twiceit=TRUE)),
col="red")
lines(sleek(fdeaths))
```

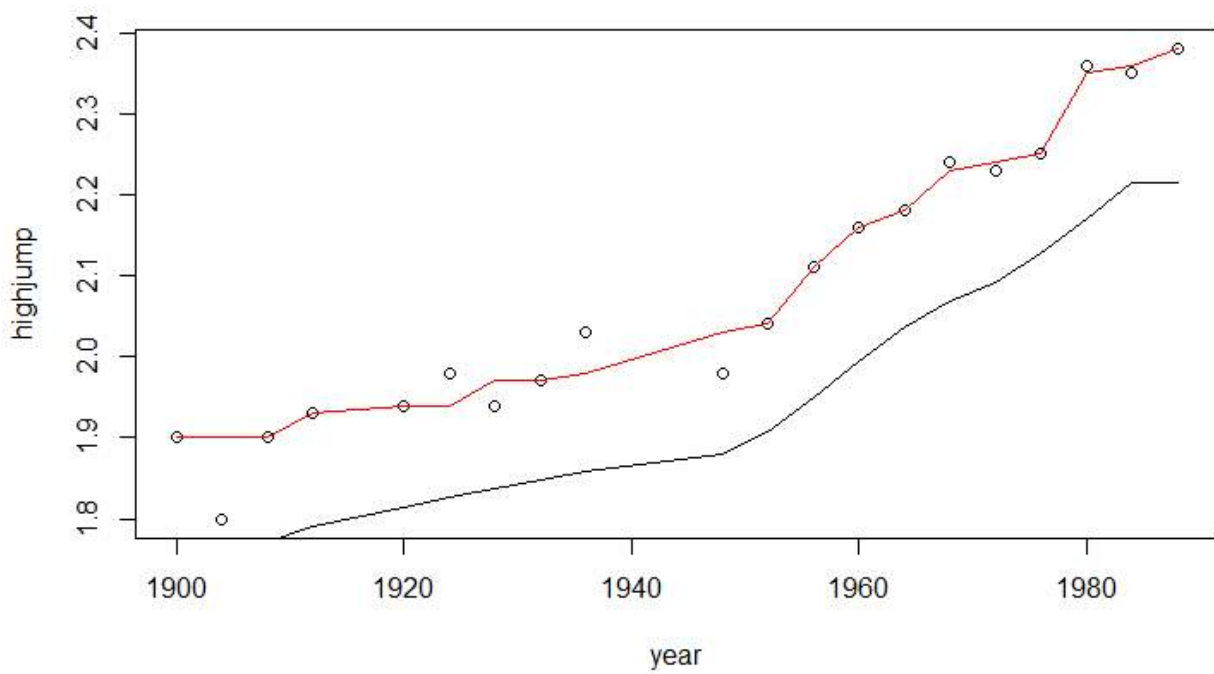


점선은 원자료, 붉은색은 3RS3R twice, 검은색은 4253H twice 평활이다. 약 1년을 주기로 계절성이 나타난다. 약 2월에 고점, 9월에 저점으로 나타난다. 평활된 데이터는 전반적인 증감 추세는 없으며 변동성도 일정한 것으로 보인다. 다만 평활되지 않은 원 자료에서 1976년 고점에서 다른 연도에 비해 높은 수치가 확인되는데, 이는 1976년에 특히 사망자가 많은 원인이 있었기 때문으로 추정된다.

#### 4. R의 datasets에 있는 jumping.dat

첫 칼럼은 연도, 둘째 칼럼부터 high jump, pole vault, long jump, tripple jump의 네 가지 뛰기 관련 종목에 대한 올림픽 기록이다. 단위 미터. 네 경기 종목에 대한 기록을 각각 평활하여 어느 종목의 기록이 점점 기록 갱신하기 어려워지고 있는지 분석하라. (어느 종목에서 기록 갱신의 상대적 변화가 점점 줄어드는가?)

```
#highjump
plot(year,highjump, lty="dotted")
lines(year,ts(smooth(highjump, kind="3RS3R", twiceit=TRUE)), col="red")
lines(year,sleek(highjump))
```

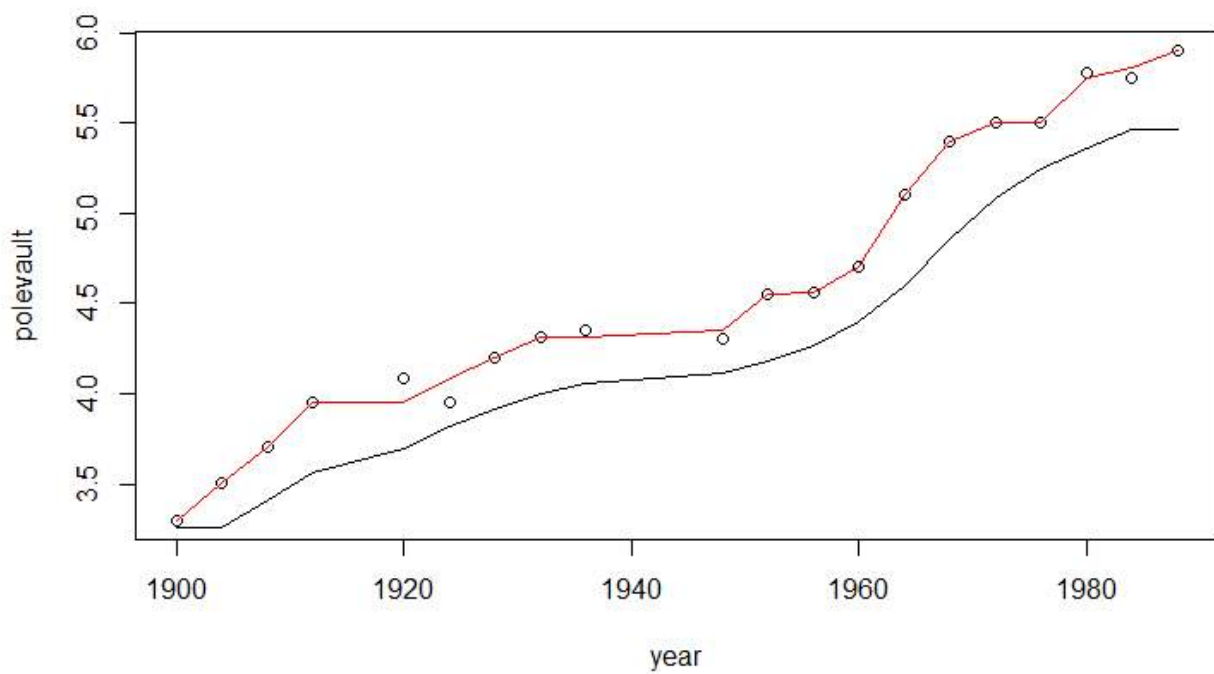


#polevault

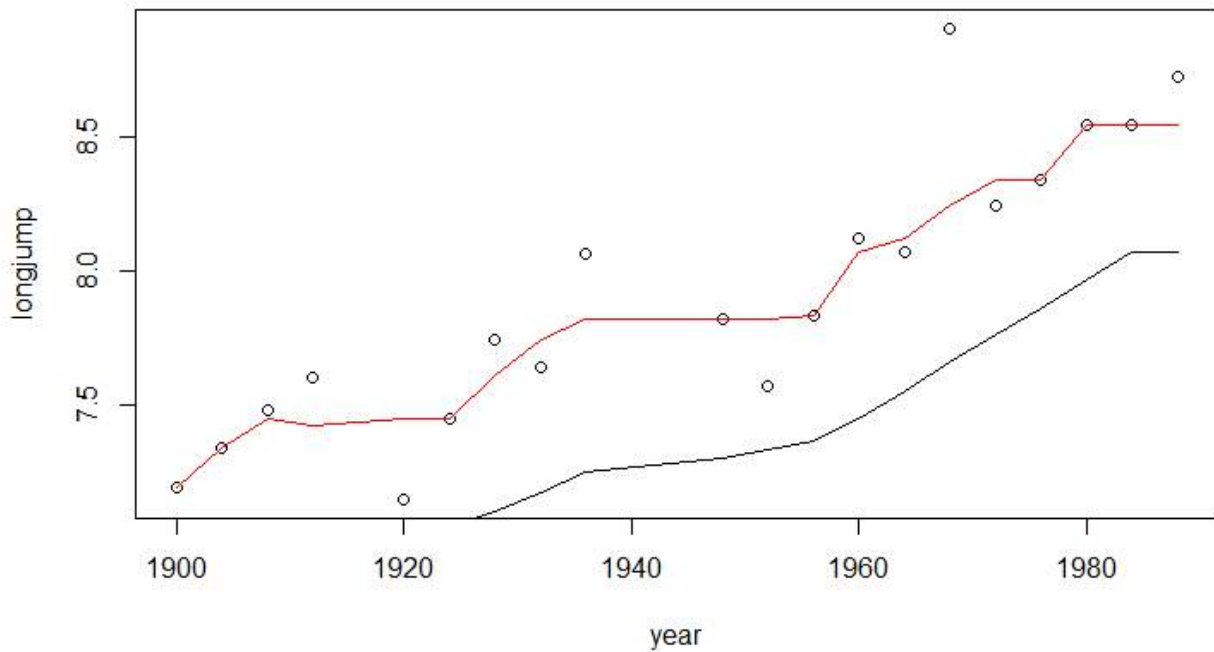
plot(year,polevault, lty="dotted")

lines(year,ts(smooth(polevault, kind="3RS3R", twiceit=TRUE)), col="red")

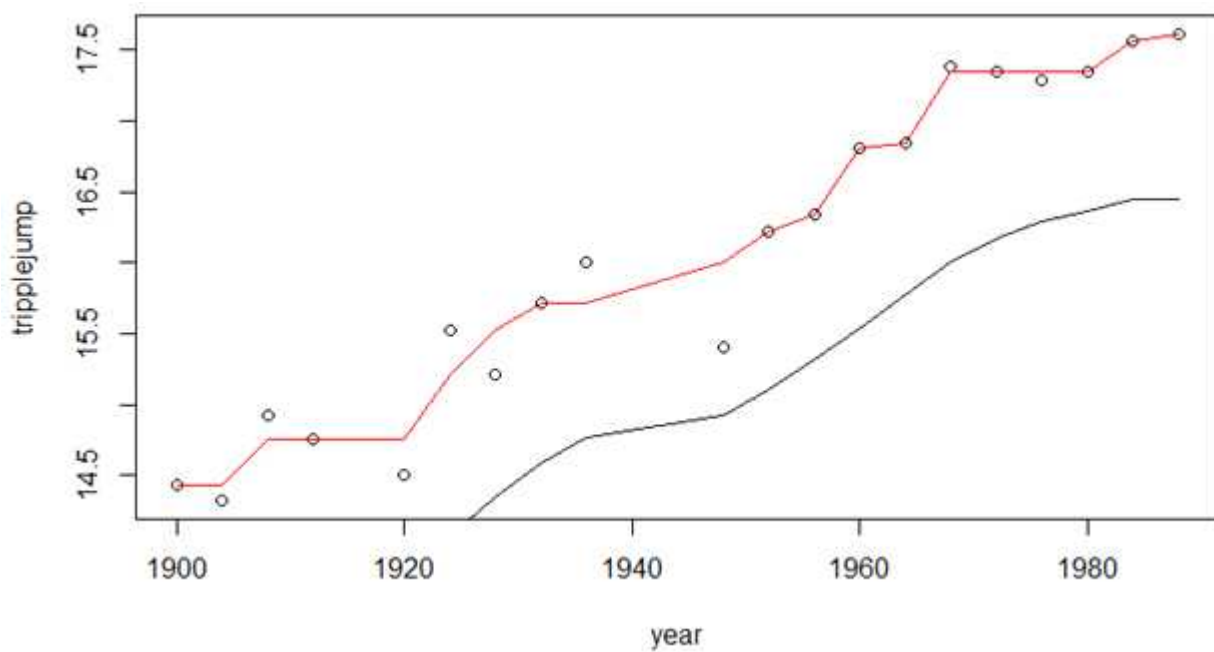
lines(year,sleek(polevault))



```
#longjump
plot(year,longjump, lty="dotted")
lines(year,ts(smooth(longjump, kind="3RS3R", twiceit=TRUE)), col="red")
lines(year,sleek(longjump))
```



```
#tripplejump
plot(year,tripplejump, lty="dotted")
lines(year,ts(smooth(tripplejump, kind="3RS3R", twiceit=TRUE)), col="red")
lines(year,sleek(tripplejump))
```



점은 원자료, 붉은선은 3RS3R twice평활, 검은선은 4253H twice평활이다. 4253H twice 평활은 자료의 크기가 지나치게 줄어들기 때문에 3RS3R twice 평활이 더 분석에 적절하다.  
모든 종목의 추세는 시간이 지남에 따라 기록이 증가한다.

```
c(mean(highjump),mean(polevault),mean(longjump),mean(tripplejump))
```

```
[1] 2.0835 4.6205 7.9170 16.0785
```

다만 각 종목 기록의 평균을 구해보면 각각 highjump: 2.0835, polevault: 4.6205, longjump: 7.9170, tripplejump: 16.0785 으로, 규모에 차이가 있기 때문에 규모의 수준을 일치시킬 필요가 있다고 생각된다.

**#표준화: (data-평균)/표준편차**

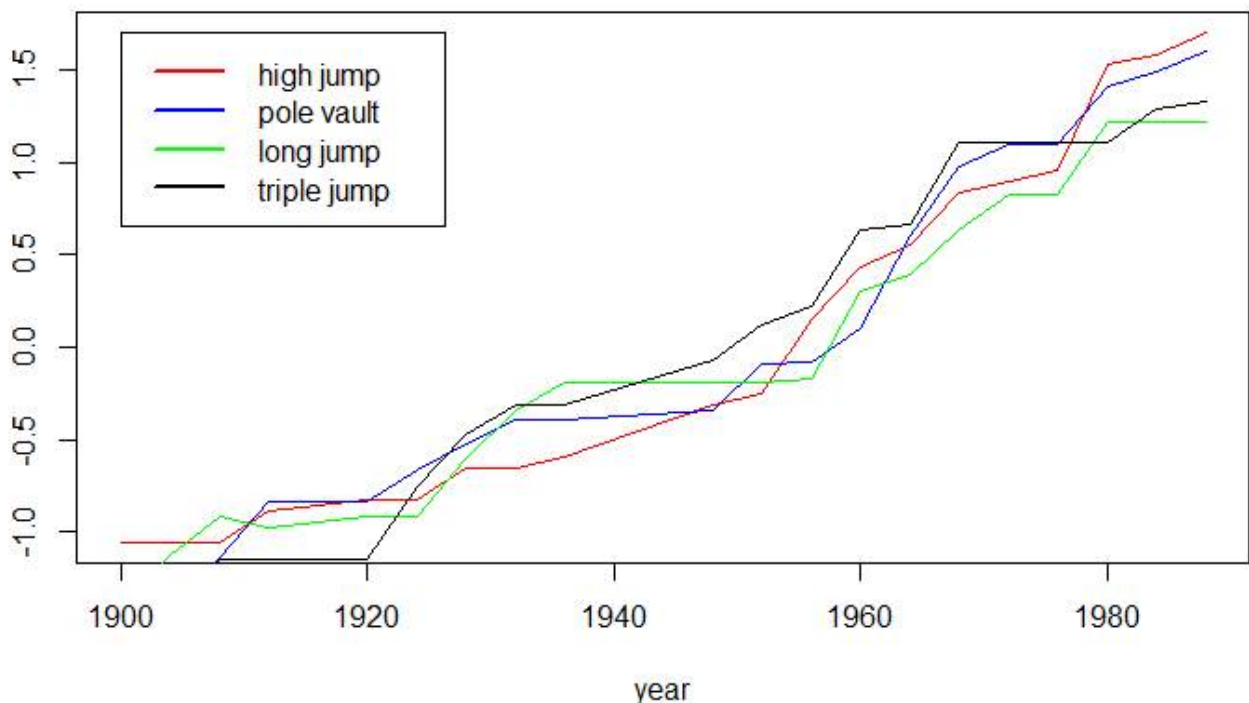
```
n_highjump <- (highjump - mean(highjump))/sd(highjump)
```

```
n_polevault <- (polevault - mean(polevault))/sd(polevault)
```

```
n_longjump <- (longjump - mean(longjump))/sd(longjump)
```

```
n_tripplejump <- (tripplejump - mean(tripplejump))/sd(tripplejump)
```

```
plot(year,ts(smooth(n_highjump, kind="3RS3R", twiceit=TRUE)), col="red", type="l")
lines(year,ts(smooth(n_polevault, kind="3RS3R", twiceit=TRUE)), col="blue", type="l")
lines(year,ts(smooth(n_longjump, kind="3RS3R", twiceit=TRUE)), col="green", type="l")
lines(year,ts(smooth(n_tripplejump, kind="3RS3R", twiceit=TRUE)), col="black", type="l")
legend(x = 1900, y = 1.7, c("high jump", "pole vault", "long jump", "triple jump"),
      col = c("red","blue","green","black"),
      lty=c(1,1,1,1),lwd=2)
```





표준화 된 후 평활된 자료를 분석하면 high jump, pole vault의 경우 꾸준히 점수가 향상되고 있다. long jump의 경우 1980년을 기점으로 점수가 정체되어있고, triple jump의 경우 약 1970년을 기점으로 점수가 정체되어있다. 따라서 점수 갱신은 triple jump가 가장 오랫동안 어려워져 있다고 볼 수 있다. 다만 triple jump는 최근(1980 이후) 향상이 있는 반면 long jump는 최근 정체 추세가 시작되고 있기 때문에 추세 면에선 long jump가 갱신이 어려워지고 있는 것으로 보인다.