

Exploratory Data Analysis & Statistical Consulting

Final Examination

[1] 목표 : 핫도그 먹기 대회 of 연도별 기록 및 수상자 자료를 R을 이용하여 시각화한다.



Figure 1: Chestnut과 Kobayashi의 열띤 경기장면

시험 도중 문제가 안풀려 핫도그 먹기 대회 동영상을 보면서 영감을 얻고자 한다면 <http://www.youtube.com/watch?v=POE8ImL4aXY>을 클릭하시오. 지금 여러분과 똑같이 참가자들의 열의와 고뇌를 느낄 것이다. 단, 음소거는 필수이자 매너!!

[2] 데이터 다운로드 : flowingdata.com으로 부터 다운로드 (<http://datasets.flowingdata.com/hot-dog-contest-winners.csv>). http 접속을 통해 R로 불러들임. 다음 코드를 사용할 것.

```
hotdogs<-read.csv('http://datasets.flowingdata.com/hot-dog-contest-winners.csv',  
                 sep=',', header=TRUE)
```

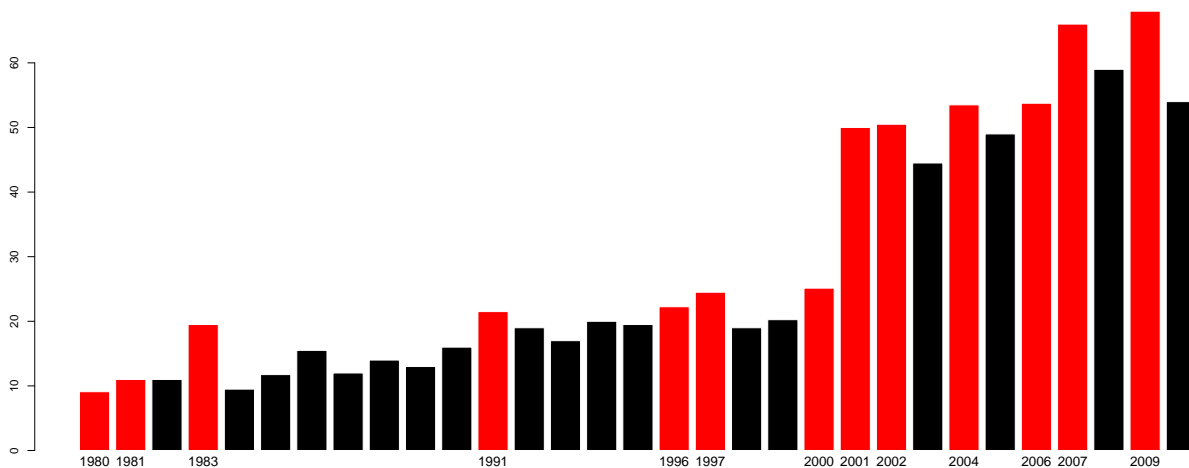
[3] 데이터설명 : 데이터 구성은 다음과 같다. 아래에 각 변수에 대한 설명을 참고하라.

```
> head(hotdogs)
  Year      Winner Dogs.eaten Country New.record
1 1980 Paul Siederman & Joe Baldini    9.10 United States      0
2 1981      Thomas DeBerry    11.00 United States      0
3 1982      Steven Abrams    11.00 United States      0
4 1983      Luis Llamas    19.50 Mexico          0
5 1984      Birgit Felden    9.50 Germany         0
6 1985      Oscar Rodriguez    11.75 United States      0
```

1. Year: 대회 연도
2. Winner: 수상자 이름
3. Dogs.eaten: 수상자가 기록한 12분 동안 먹은 핫도그 개수
4. Country: 수상자의 국적
5. New.record: 신기록 수립 여부. 본 분석에서는 이 변수를 사용하지 말것.

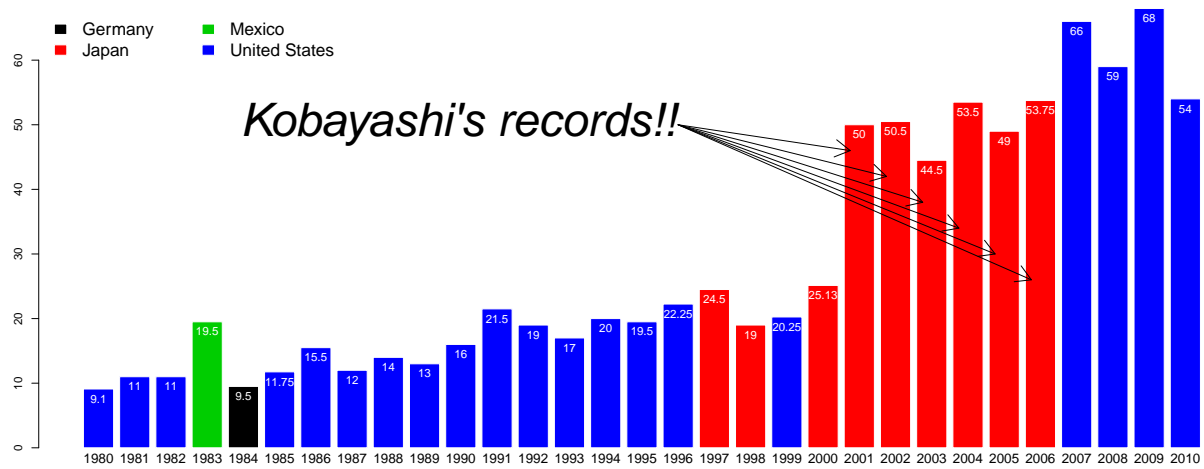
[4] 미션 : 다음 3개의 그림을 “가급적 똑같이” 구현하라

1. 다음의 그림을 ‘figure1.pdf’ 이름으로 저장하라.



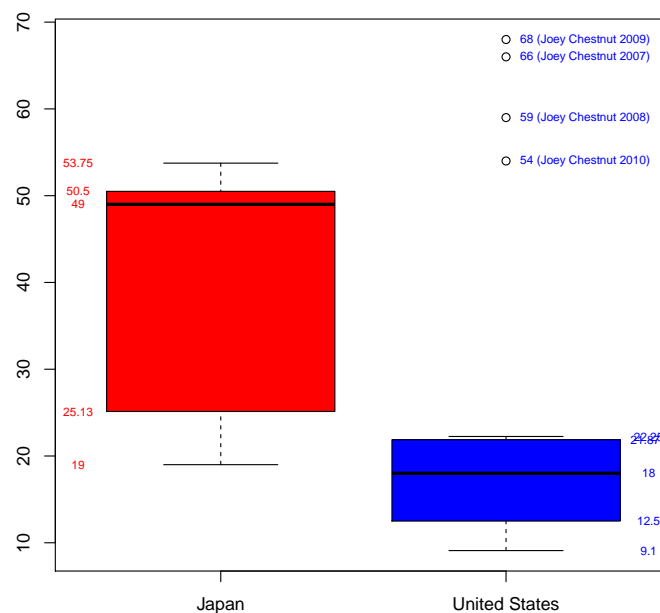
위의 그림은 연도별 우승자의 기록을 막대그래프로 그린 것이다. 빨간색은 주어진 자료에서 신 기록이 갱신된 경우이며, 검정색은 해당 연도의 기록이 이전 연도의 기록보다 높지 않을 때이다. 신기록 수립이 이루어진 연도를 막대 아래에 나타내고 있다. 신기록 여부는 New.record의 내용과 다름에 유의할 것!

2. 다음의 그림을 'figure2.pdf' 이름으로 저장하라.



연도별 우승자의 기록을 막대그래프로 그리되, 해당 연도 우승자의 국적별로 색상을 다르게 한 것이다. 막대의 윗부분에 기록을 명시하였으며, 고바야시의 기록을 특별히 강조하고 있다.

3. 다음의 그림을 'figure3.pdf' 이름으로 저장하라.



일본 및 미국 국적의 우승자들에 대한 상자그림을 그린다. 상자그림을 구성하는 수치들을 상자 그림 옆에 기입한다. 또한 이상점이 있는 경우, 해당 기록 및 우승자 이름과 우승연도를 위와 같이

기록한다. 이상점은 R에서 주어진 디폴트대로 정의한다.

[5] 유의사항 및 채점기준

1. 시험시간은 120분(11:30~13:30) 임. 100점 만점. 결과는 채점후 eClass에서 확인 가능
2. 강의노트, 인터넷 사용 모두 가능. 하지만, 이메일은 시험 중 절대 사용금지. (시험 종료후, eClass에 결과를 올리지 못한 마지막 상황에는 이메일 사용 가능. 아래 6번항 참조)
3. 위에 예시한 그림에 가깝게 구현할 것. 완결하지 못하더라도 에러없이 그림이 생성되도록 작성할 것.
4. 정확성이 채점의 우선사항임. 코딩의 효율성 및 자동화는 채점시 가산점으로 부여됨
5. 작성한 R 코드 및 그림 파일을 eClass에 올릴 것.
6. 각 그림은 3개의 pdf 파일로 저장할 것. R 코드는 하나의 파일로 저장하되 확장자는 .R 혹은 .r로 할 것.
7. 시간이 지나면 eClass에 올릴 수 없으니 시간 엄수를 바람. 만일 시간이 넘어서 저장할 수 없다면 lees@hufs.ac.kr로 결과를 보낼 것. 단 이런 경우 시간엄수 미이행으로 인한 패널티가 부여할 것이며 만점은 80점으로 제한됨.