

이번 실습 시간에는 교통사고 사망률 데이터와 인구 데이터를 사용하여 데이터를 가공하고, 그 결과를 막대그래프와 같은 간단한 그래프로 나타내는 방법에 대해 알아볼 것입니다. 교통사고 사망률 데이터는 `deathrates.csv`, 인구 데이터는 `popcities10.csv`에 담겨 있으며, 지난 실습과 마찬가지로 KLAS나 `github`에서 다운로드가 가능합니다.

우선 데이터를 로컬 폴더에 저장하고, 아래와 같이 `read.csv()` 함수를 사용하여 R로 불러옵니다.

```
deathrates <- read.csv(file.choose(), header = TRUE) # deathrates.csv
pop <- read.csv(file.choose(), header = TRUE) # popcities10.csv
```

`head()` 함수와 `nrow()`, `ncol()` 함수 등을 사용하여 데이터에 담겨 있는 변수들을 간단히 살펴볼 수 있습니다. 데이터 객체 내용 전체를 살펴보고자 한다면 `show()` 함수를 사용하는 것이 편리합니다.

첫 번째 문제는 조교님의 설명과 함께 진행합니다.

1. [10 점] 다운로드 받은 `deathrates.csv` 파일에는 1980년대 중반 30개 국가의 인구 10만명 당 교통사고 사망률 데이터가 들어 있습니다.
 - (1) [4 점] `hist()` 함수를 사용하여 교통사고 사망률에 대한 막대기둥그림표(histogram)를 그리는 R 코드를 적으시오. 그래프에는 제목과 x 축, y 축의 값을 설명하는 레이블이 포함되어야 하며, 그 외의 부분에 대해서는 기본값을 그대로 사용해도 괜찮습니다.
 - (2) [4 점] `barplot()` 함수를 사용하여 국가별 교통사고 사망률을 오름차순으로 정렬하여 막대그래프로 나타내는 R 코드를 적으시오. 그래프에는 제목과 x 축, y 축의 값을 설명하는 레이블이 포함되어야 하며 그 외의 부분에 대해서는 기본값을 그대로 사용해도 괜찮습니다.
 - (3) [2 점] `cut()` 함수를 사용하여 전체 교통사고 사망률을 낮음, 보통, 높음의 세 가지 값으로 분류하기 위한 R 코드를 적으시오. 낮음, 보통, 높음의 기준은 각자 임의로 설정할 수 있습니다.
 - (4) [4 점] (3)에서 가공된 교통사고 사망률 데이터를 사용하여 그 결과를 아래 <그림 1>과 같이 막대그래프로 나타내는 R 코드를 적으시오. 앞서 두 문제와 마찬가지로 그래프에는 제목과 x 축, y 축의 값을 설명하는 레이블이 포함되어야 합니다.

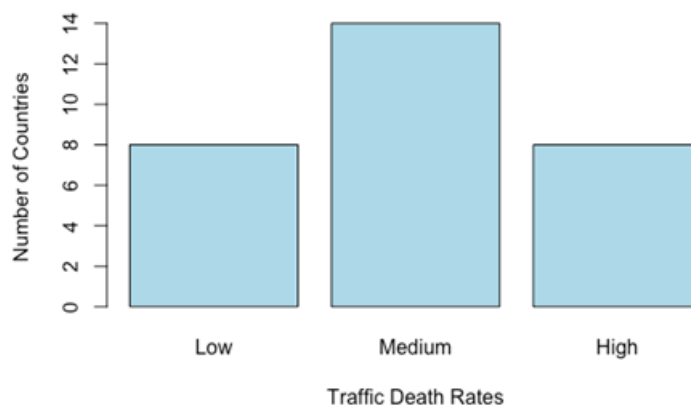


그림 1. 낮음, 보통, 높음으로 구분된 교통사고 사망률 데이터

2. [15 점] popcities10.csv 에는 국내 83 개 도시의 2000 년, 2005 년, 2010 년 인구 데이터가 포함되어 있습니다.

- (1) [3 점] `plot()` 함수를 사용하여 2000 년부터 2010 년까지 전라북도 군산의 인구 변화와 같은 기간 평균 인구를 아래와 같이 보여주는 그래프를 그리고, R 코드를 적으시오.

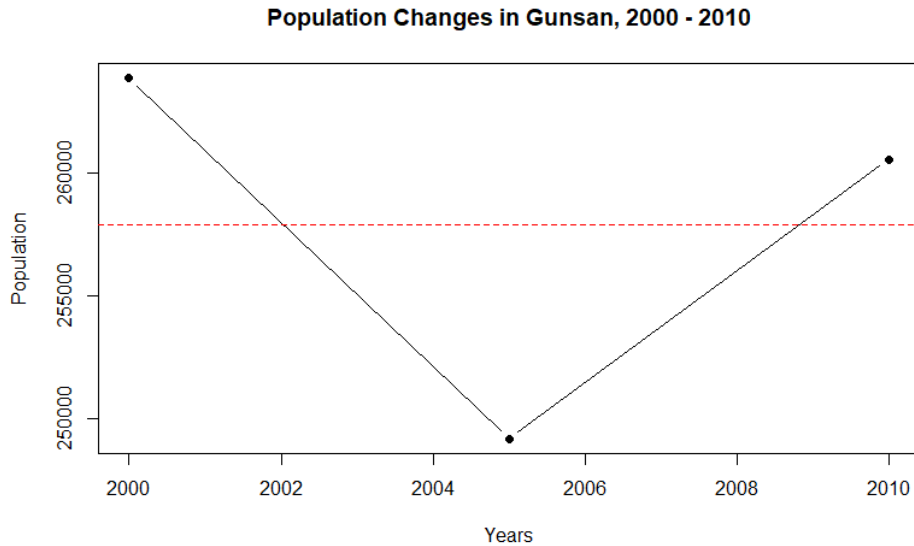


그림 2. 전라북도 군산의 인구 변화를 나타내는 그래프

- (2) [2 점] `hist()` 함수를 사용하여 2010 년 인구 수 분포를 나타내는 막대기둥그래프를 그리기 위한 R 코드를 적으시오. 그래프에는 반드시 제목과 x 축, y 축의 값을 설명하는 레이블이 포함되어야 합니다.
- (3) [3 점] 2010 년을 기준으로 경기도에서 인구가 가장 많은 다섯 개 도시를 선택하고, 선택된 도시의 인구를 막대그래프로 보여주지 위한 R 코드를 적으시오. 도시의 이름은 막대그래프 하단에 표시되어야 합니다.
- (4) [3 점] 2010 년을 기준으로 인구가 가장 많은 다섯 개 도시를 선택하고, 이들의 2000 년, 2005 년, 2010 년 인구를 연도별로 각각 색상을 달리하여 아래 <그림 3>과 같은 막대그래프로 나타내는 R 코드를 적으시오. 그래프의 하단에는 각 도시의 이름이 명확하게 표기되어야 하며, 제목, x 축, y 축의 값을 설명하는 레이블도 포함되어야 합니다. 단, 예시에 있는 그래프의 범례는 포함하지 않아도 괜찮습니다.

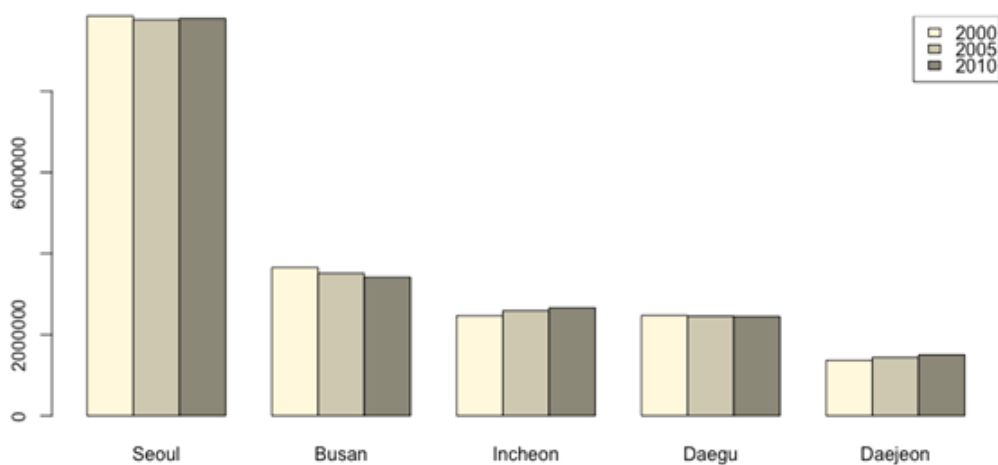


그림 3. 다섯 개 도시의 2000 년, 2005 년, 2010 년 인구 수

- (5) [2 점] 2000 년부터 2010 년까지의 기간 동안 가장 높은 인구 증가율(%)을 나타낸 일곱 개의 도시를 찾기 위한 R 코드를 적으시오.
- (6) [2 점] 위에서 찾은 일곱 개 도시의 인구 증가율을 막대그래프로 나타내는 R 코드를 적으시오. 각 도시의 이름은 막대그래프 하단에 표시되어야 합니다.

기타 평가항목

R 파일의 실행 가능 여부 및 작성 스타일(Google's R Style Guide 준수 여부): 5 점

각 문항에 대한 답안은 앞선 실습과 마찬가지로 R 콘솔에서 직접 실행이 가능한 형태의 스크립트 파일로 작성하여 제출합니다. 파일의 확장자는 *.R 로 하며, `source()` 함수를 통해 오류 없이 실행되어야 합니다. 답안에서 R 코드 이외의 부분은 주석으로 처리하여, 오류가 발생하지 않도록 해주세요. 답안은 10 월 11 일 금요일까지 KLAS 를 통해 제출하면 됩니다.

2019 년 9 월 8 일
홍성연(syhong@khu.ac.kr)