

지난 실습 시간에는 R 에서 데이터를 직접 입력하고, 특정 위치의 값을 추출하는 방법에 대해 살펴보았습니다. 그러나 일반적으로 우리가 분석에 사용하게 되는 데이터는 파일 형식으로 저장되어 있는 경우가 많기 때문에, 굳이 콘솔 창에서 직접 입력하는 것보다 파일을 그대로 불러오는 것이 훨씬 편리합니다. 이번 실습에서는 쉼표로 구분된 값(commma separated values; CSV)을 가진 데이터 파일을 R 로 가져와 조건에 맞는 데이터 값을 추출하는 방법에 대해 배워볼 것입니다.

CSV 파일과 같은 정형 데이터에서 각각의 값은 행으로 정렬되며, 하나의 행에서 열(변수)은 쉼표로 구분됩니다. 오늘 실습에 사용하는 CSV 파일들은 모두 KLAS 나 github 에서 받을 수 있고, 각 파일의 첫 번째 행에는 열의 이름(변수명)이 입력되어 있습니다.

지난 시간과 마찬가지로 아래에 주어진 문제를 풀고, 답안을 R 파일로 저장하여 KLAS 를 통해 제출해주세요. 답안은 하나의 R 파일로 작성하여야 하며, 제출한 파일은 R 에서 `source()` 함수를 통해 오류 없이 실행이 가능해야 합니다.

첫 번째 문제는 조교님의 설명과 함께 진행합니다.

- [10 점] `libraries.csv` 파일에는 서울시 종로구에 소재한 16 개 도서관에 대한 정보가 담겨 있습니다. 파일의 첫 번째 행에는 열 이름이 들어 있습니다.
 - [1 점] `read.csv()` 함수를 사용하여 해당 파일을 R 로 가져와 `libs` 라는 이름의 객체로 저장하는 R 코드를 적으시오.
 - [1 점] `libs` 객체 내의 열 이름을 출력하는 R 코드를 적으시오.
 - [2 점] `sum()` 함수를 사용하여 데이터에 포함된 모든 도서관의 열람좌석수를 더하는 R 코드를 적으시오.
 - [2 점] `libs` 객체에서 열람좌석수가 30 석 이상인 도서관 이름을 출력하는 R 코드를 적으시오.
 - [2 점] `libs` 객체에서 자료(도서) 수가 5,000 권 이상이거나 건물 면적이 100 m² 이상인 도서관의 모든 정보를 출력하는 R 코드를 적으시오.
 - [2 점] 건물 면적 1 m² 당 열람좌석수가 0.3 석 이상인 도서관의 도로명 주소를 출력하는 R 코드를 적으시오.
- [15 점] `popcities15.csv` 파일에는 국내 85 개 도시(1 개 특별시, 1 개 특별자치시, 6 개 광역시, 75 개 특정자치시, 2 개 행정시)의 인구 데이터가 들어 있습니다. 파일의 첫 번째 행은 열의 이름을 나타내며, 각 열의 내용은 다음과 같습니다.

열 이름	설명
rank	2015 년 기준 인구 순위
provinces	소재 지역(도, 道) 이름(특별시, 특별자치시, 광역시의 경우 해당 없음)
municipality	도시 이름
pop15	2015 년 인구
pop10	2010 년 인구

pop05	2005 년 인구
pop00	2000 년 인구

- (1) [1 점] `read.csv()` 함수를 사용하여 해당 파일을 R 로 가져와 `pop` 이라는 이름의 객체에 저장하는 R 코드를 적으시오.
- (2) [1 점] `seq()` 함수를 사용하여 1 에서 100 사이의 홀수를 모두 출력하는 R 코드를 적으시오(1 포함).
- (3) [1 점] `pop` 에 저장된 데이터의 홀수 번째 행을 모두 출력하는 R 코드를 적으시오.
- (4) [2 점] `class()` 함수를 사용하여 `pop` 객체 내 각 열의 클래스를 확인하고, 아래의 표를 채우시오.

열 이름	rank	provinces	municipality	pop15	pop10	pop05	pop00
클래스	<i>integer</i>						

- (5) [2 점] 2015 년 기준 인구가 100 만 이상인 도시를 모두 선택하여 출력하는 R 코드를 적으시오.
- (6) [2 점] 2000 년에는 인구가 30 만 이하였으나, 2015 년에는 인구가 40 만 이상으로 증가한 도시를 선택하여 출력하는 R 코드를 적으시오.
- (7) [2 점] 경기도에서 소재한 도시 중 2010 년 기준으로 인구가 70 만 이상인 도시를 선택하여 출력하는 R 코드를 적으시오.
- (8) [2 점] 2005 년에서 2010 년 사이 인구 증감폭이 5 만 명 이상인 도시를 선택하여 화면에 출력하는 R 코드를 적으시오.
- (9) [2 점] 2000 년과 비교하여 2015 년 인구가 50% 이상 증가한 도시의 수를 출력하는 R 코드를 적으시오.

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.

2019 년 8 월 13 일
홍성연(syhong@khu.ac.kr)