

국토교통부에서 운영하는 실거래가 공개시스템(<http://rt.molit.go.kr/>)에서는 지난 2006 년 1 월 이후 체결된 부동산 매매거래 정보를 제공하고 있습니다. 오늘 실습에서는 실거래가 공개시스템에서 다운로드 받은 데이터를 사용하여 데이터 가공과 시각화 함수에 대해 복습해보도록 하겠습니다.

KLAS의 데이터 압축 파일에는 2018 년 상반기 동대문구와 성북구의 아파트, 단독주택 매매거래 정보가 담겨 있으며, 건축연도나 면적 등과 같은 기본적인 정보는 별도의 파일로 저장되어 있습니다(총 네 개의 CSV 파일). 각 파일을 다음과 같이 `read.csv()` 함수를 사용하여 불러옵니다.

```
aptSales <- read.csv("apt_sales.csv", header = TRUE)      # 아파트 매매거래 정보
aptInfo <- read.csv("apt_basicinfo.csv", header = TRUE)   # 아파트 기본 정보
houseSales <- read.csv("house_sales.csv", header = TRUE)  # 단독주택 매매거래 정보
houseInfo <- read.csv("house_basicinfo.csv", header = TRUE) # 단독주택 기본 정보
```

첫 번째 문제는 조교님의 설명과 함께 진행합니다.

1. [10 점] 단독주택 매매거래 정보를 사용하여 다음의 질문에 답하세요. 실행가능한 R 코드를 제외한 답안은 반드시 주석으로 처리하여, `source()` 함수로 불러왔을 때 오류가 발생하지 않아야 합니다.
 - (1) [2 점] `str()` 함수를 사용하여 단독주택 매매거래 데이터에 총 몇 개의 변수가 저장되어 있는지 확인하고, 그 중 매매가격(`salesPrice`)은 몇 번째에 위치해 있는지 적으시오.
 - (2) [2 점] `merge()` 함수를 사용하여 매매거래 데이터와 기본 정보를 올바르게 결합하고 이를 새로운 데이터 프레임으로 저장하는 R 코드를 적으시오. 데이터 프레임의 결합에는 `id` 열을 사용할 수 있습니다.
 - (3) [2 점] `transform()` 함수를 사용하여 단위면적(`landSize`) 당 매매가격을 나타내는 새로운 파생변수가 포함된 데이터 프레임을 만드는 R 코드를 적으시오.
 - (4) [2 점] `summary()` 함수를 사용하여 동대문구에 소재한 단독주택의 단위면적 당 매매가격 분포를 살펴보는 R 코드를 적으시오.
 - (5) [2 점] `boxplot()` 함수를 사용하여 동대문구와 성북구 단독주택의 단위면적 당 매매가격 분포를 비교하는 상자 도표를 그리고, 해당 R 코드를 적으시오.
2. [15 점] 아파트 매매거래 정보를 사용하여 다음의 질문에 답하세요. 실행가능한 R 코드를 제외한 답안은 반드시 주석으로 처리하여, `source()` 함수로 불러왔을 때 오류가 발생하지 않아야 합니다.
 - (1) [2 점] `str()` 함수를 사용하여 아파트 매매거래 데이터에 총 몇 개의 변수가 저장되어 있는지 확인하고, 그 중 매매가격(`salesPrice`)은 몇 번째에 위치해 있는지 적으시오.
 - (2) [2 점] `merge()` 함수를 사용하여 매매거래 데이터와 기본 정보를 올바르게 결합하고 이를 새로운 데이터 프레임으로 저장하는 R 코드를 적으시오. 데이터 프레임의 결합에는 `id` 열을 사용할 수 있습니다.
 - (3) [2 점] `transform()` 함수를 사용하여 단위면적(`size`) 당 매매가격을 나타내는 새로운 파생변수가 포함된 데이터 프레임을 만드는 R 코드를 적으시오.
 - (4) [2 점] `summary()` 함수를 사용하여 동대문구에 소재한 아파트의 단위면적 당 매매가격 분포를 살펴보는 R 코드를 적으시오.

(5) [2 점] `boxplot()` 함수를 사용하여 동대문구와 성북구 아파트의 단위면적 당 매매가격 분포를 비교하는 상자 도표를 그리고, 해당 R 코드를 적으시오.

(6) [2 점] `paste()` 함수를 사용하여 시(si), 자치구(gu), 행정동(dong) 열을 결합하는 R 코드를 적으시오. 결합된 결과를 출력하면 아래와 같은 벡터가 될 것입니다.

```
[1] "서울특별시 동대문구 답십리동" "서울특별시 동대문구 답십리동" "서울특별시 동대문구 답십리동"
[4] "서울특별시 동대문구 답십리동" "서울특별시 동대문구 답십리동" "서울특별시 동대문구 답십리동"
...
```

(7) [3 점] `plot()` 함수를 사용하여 아래의 <그림 1>과 같이 건축연도(`yearBuilt`)와 단위면적 당 매매가격 간 관계를 보여주는 그래프를 만드는 R 코드를 적으시오. 그래프에는 제목, *x* 축, *y* 축의 값을 설명하는 레이블이 반드시 포함되어야 합니다.

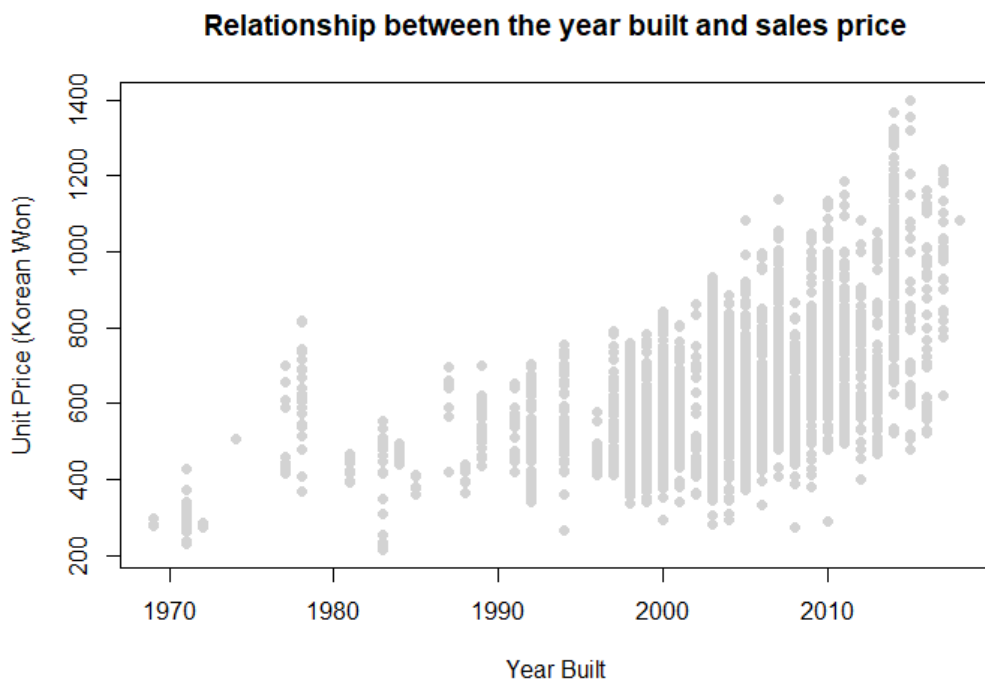


그림 1. 건축연도와 단위면적 당 매매가격의 관계

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 18 일 금요일까지 KLAS 를 통해 제출하면 됩니다.

2019 년 8 월 13 일
홍성연(syhong@khu.ac.kr)