



# 공간정보 자료 처리 (공간 질의)

Spatial Query



## 통계를 내기 위한 용어

- average : mean, median, mode를 포함

mean > (산술) 평균

- 모든 숫자를 다 고려하므로 극단치 또는 이상치에 의해 영향을 받음
- 1차 적률 = 평균

median > 중앙값

- 데이터를 크기 순으로 나열할 때 가운데에 있는 값
- 데이터수가 홀수 : 중앙에 있는 데이터
- 데이터수가 짝수 : 중앙에 있는 두 개 데이터의 평균

mode > 최빈값

- 데이터 값 중에서 가장 많은 빈도를 가지고 나타나는 데이터
- 도수분포표에서 도수가 최대인 곳의 대표치

- 중앙 값(mid-range), 기하평균(geometric mean), 조화 평균

geometric mean > 기하평균

- 데이터가 등비수열 형태로 나타날 때 사용

harmonic mean > 조화평균

- 평균적인 변화율을 구할 때 주로 사용 (자동차 구간별 속도 -> 평균속도)

# 공간 질의



## 분류 방법

15개 축구팀을 5개 등급으로 분류하는 방법:

등간격 (equal interval),

등분위 (quantile),

내추럴브레이크 (natural break)

축구팀	승리횟수
A	1
B	1
C	2
D	7
E	8
F	15
G	16
H	17
I	25
J	26
K	27
L	35
M	36
N	37
O	37

**등간격 분류:** 가장 많은 승리를 거둔 팀이 37승, 적은 팀이 1승으로 총 36승의 차이가 나고, 5등간으로 나누면 7.2가 동일한 간격이 됨. 등간격 방식으로는 1등급에 L, M, N, O 4개의 팀이 해당되고, 4등급에는 F팀 하나가 분배. 5등급에는 A~E 다섯 개 팀이 배분. 이 방식에서도 F팀과 G팀은 불과 1점 차이로 서로 다른 등급으로 구분되는 문제 발생.

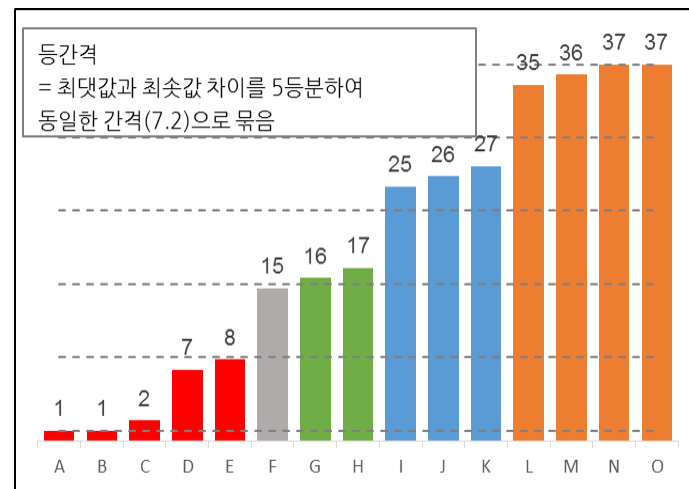
(1 - 8.2)

(8.3 - 15.5)

(15.6 - 22.8)

(22.9 - 30.1)

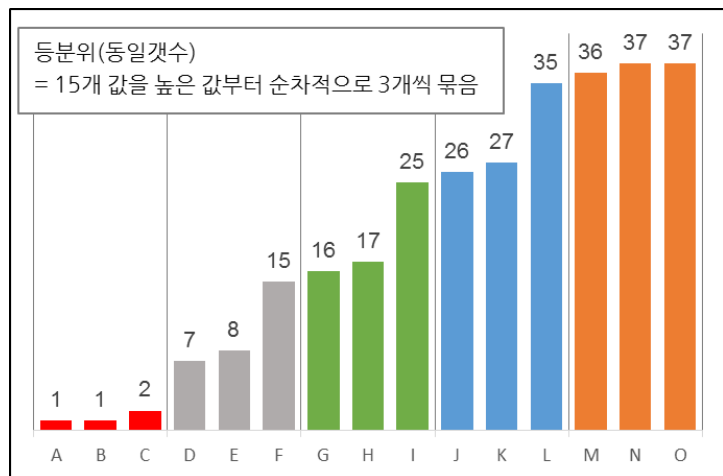
(30.2 - 37.4)



# 공간 질의

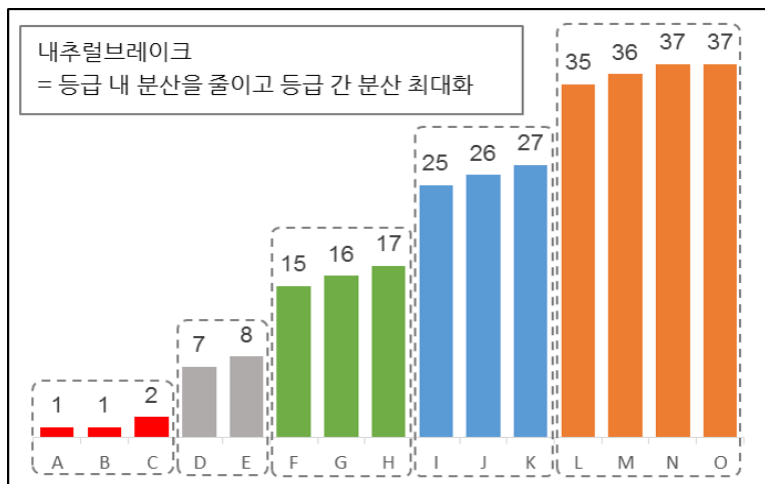


## 통계를 내기 위한 용어



**등분위 분류:** 등분위로 등급을 구분하게 되면, 승리횟수를 기준으로 F, I, L 팀은 바로 단 1승 차이로 바로 상위권에 있는 G, J, M과 다른 등급으로 분류. 동일한 갯수로 구분되는 장점은 있으나, 보기와 같이 막대차트가 아닌 색상만으로 인지한다면 F, I, L 세 팀의 실제 승리횟수에 대한 인식이 어렵다.

축구팀	승리횟수
A	1
B	1
C	2
D	7
E	8
F	15
G	16
H	17
I	25
J	26
K	27
L	35
M	36
N	37
O	37



**내추럴 브레이크 분류:** 데이터 값의 배열을 자연스러운 등급(natural classes)으로 최적화(optimize)하여 데이터로 분류(data classification) 방법. 같은 등급 내 전체 값들의 평균을 기준으로 평균편차(average deviation)는 최소화 되고, 각 등급간의 분산(variance)은 극대화. 즉, 등급 내의 분산은 줄이고 등급 간의 분산은 최대화하는 방법임. 임의로 등급을 나눈 뒤 반복적으로 평균과 편차를 계산해가면서 최적화된 값을 만족할 때까지 객체의 등급을 하나씩 조정하는 연산을 반복 수행.

# 공간 질의



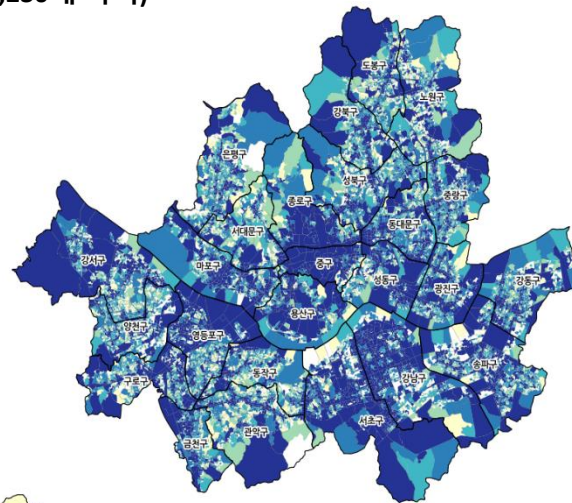
## 통계를 내기 위한 용어

### Choropleth Map 단계구분도

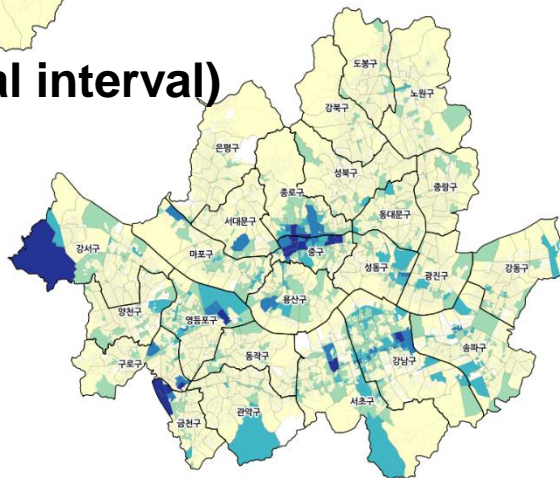
서울시 세부지역별 사업체 종사자 분포도  
(2012년 기준, 437만명, 16,230개 지역)



등간격 (equal interval)



등분위 (quantile)



내추럴 브레이크 (natural break)