



# 서울시 빅데이터 분석 기반 ‘스마트 그린 쉼터 버스정류장’ 최적입지 선정 : Platform에서 Shelter로

# 목차

1

## 분석 개요

분석 배경

시설 설명

분석 목적

분석 방향

2

## 분석 데이터

사용 데이터 목록

데이터 수집·가공

3

## 분석 과정

데이터 분포 파악

과제 흐름도

모델링

4

## 분석 결과

행정 구 선정

행정 동 선정

버스정류장 최적 입지 선정

5

## 활용 방안

문제점 개선 방안

업무 활용 방안

# 1 분석 배경

## 분석 개요

### 2. 시민의 건강을 위협

- ✓ 천식 발작, 급성 기관지염, 부정맥 증상 악화
- ✓ 심혈관질환, 호흡기질환, 폐암 발생 위험 증가



### 1. “도로 미세먼지”에 무방비 노출

- ✓ 대기오염물질 농도: 일반 도시 < 도로 인근
- ✓ 주변 대기오염에 큰 영향을 미침
- ✓ 개방형 구조: 이용객이 그대로 오염에 노출

### 3. 건강 악화로 인한 경제적 손실

- 월 평균 미세먼지 농도가  $10 \mu\text{g}/\text{m}^3$  증가할 때,
- ✓ 호흡기 질환 치료비의 막대한 증가 발생  
(천식: 1억 7천만원, 급성 편도염: 1억 2천5백만원, 만성 부비동염: 7천 8백만원)
- ✓ 국민 부담 건강보험료 증가

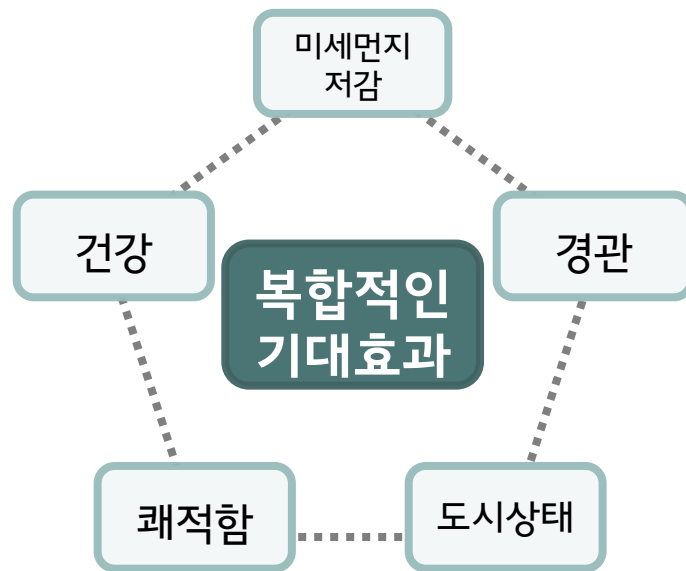
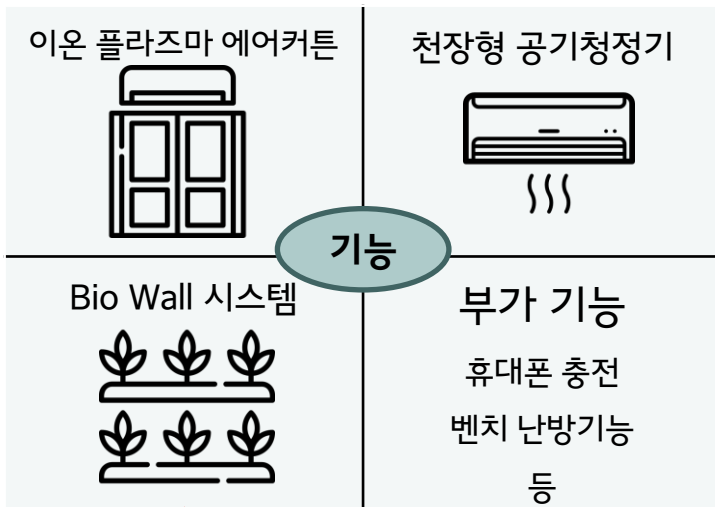


# 1 스마트 그린 쉼터란?

## 분석 개요

### 지능화된 녹색 쉼터 & 미래형 버스정류장

✓ 시민에게 다양한 편의기능 제공



✓ 공기정화 원리  
식물이 있으므로  
미세먼지를  
흡착하면  
뿌리로 내려가  
미생물에 의해  
유해 물질이 제거

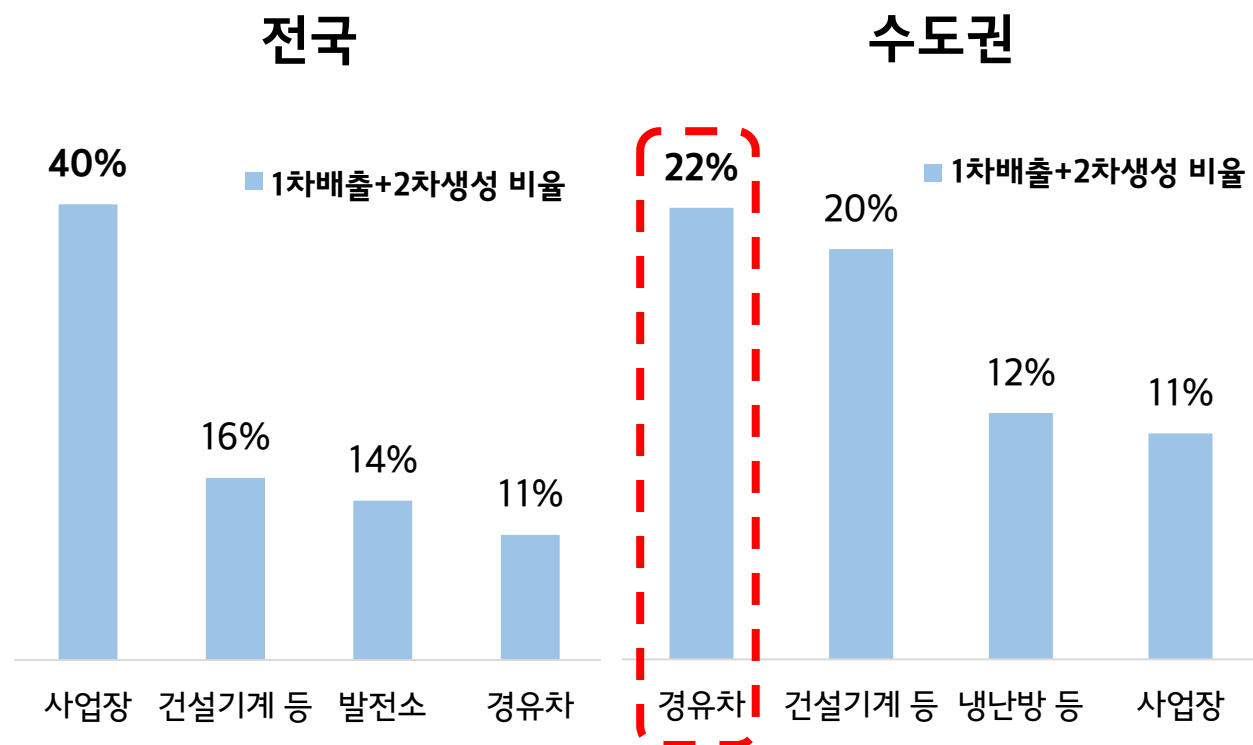


✓ 여러 지자체에서 현재 버스정류장 개선사업 시행 중

지역명	사업명
서울특별시 강남구	스마트 그린 쉼터
경기도 고양시	공기정화용 에어클린 서비스
경기도 군포시	미세먼지 안심 스마트 버스정류장
경기도 안성시	실외형 미세먼지 저감장치
경상북도 경주시	미세먼지 안심 버스정류장



## 미세먼지 배출원별 배출량 및 기여율 (2015년)

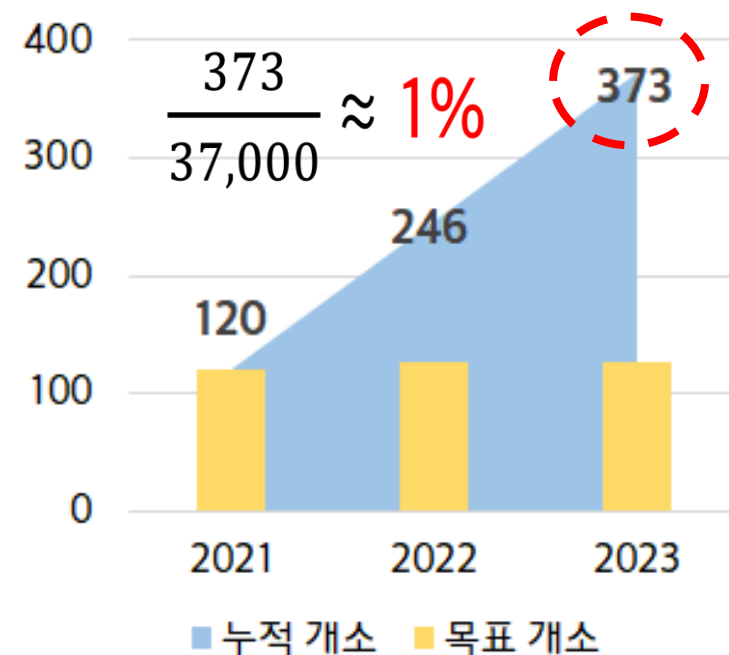


수도권 미세먼지 주요 배출원은 대다수 도로 교통에서 발생함.

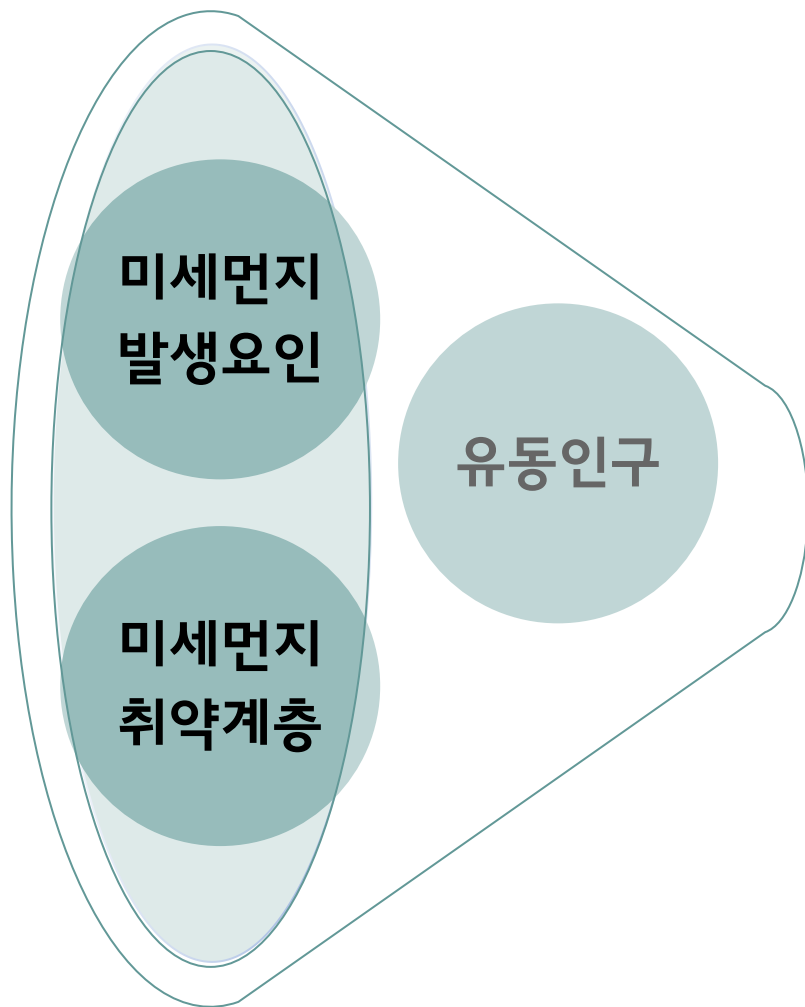
2015년 기준 수도권의 미세먼지 국내 배출원은

경유차(22%), 건설기계(20%), 냉난방 등(12%), 사업장(11%)으로 밝혀짐.

## 서울시 '스마트쉘터' 확대 예정

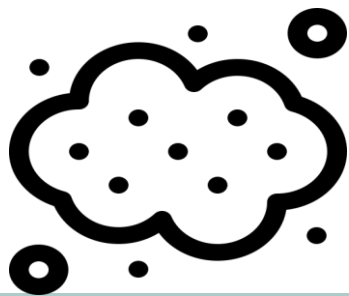


서울시는 2020년 10월부터 스마트그린쉘터를 10곳 시범 설치 및 운영 중에 있으며, 1년차 120개소를 시작으로 2023년까지 373개소 우선설치 예정임



스마트 그린 쉼터 버스정류장

‘최적 입지’ 선정



미세먼지 데이터

서울시 기간별 일평균 대기환경 정보

- 서울시 자치구별 일별 미세먼지 농도
- 초미세먼지 농도
- 대기오염 물질 농도

미세먼지 인식도 설문조사



입지·위치 데이터

도시녹지계획 현황  
(녹지면적)

서울시 행정 구 별 면적

서울시 버스정류장  
위치 정보

서울시 행정구역  
읍·면·동 SHP



인구 데이터

서울시 연령별 동별  
주민등록인구

행정동별 유동인구

버스노선별 정류장별  
시간대별  
승하차 인원 정보



교통 데이터

시내버스 배차 현황  
마을버스 배차 현황

행정동 별 교통량  
추정 데이터

버스정류장별 노선 개수

## 2 행정 구 데이터

분석데이터

2019년 교통량 = 2018년 추정 교통량[평일+주말] \* (1.006)

교통량

녹지면적

행정 구 면적

대기 환경 정보

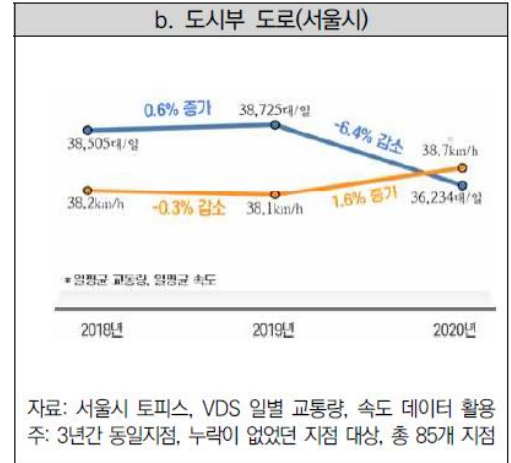
○ 도시부 도로 교통량 및 속도 추이 분석  
- (교통량) '19년은 '18년 대비 0.6% 교통량이 증가하였으나, '20년은 '19년 대비 6.4% 교통량이 감소하였음(그림 2-3 b) 참고)

통합

✓ 행정 구 클러스터링을 위한 데이터 셋 제작

행정구역	교통량	녹지면적	행정구역면적	녹지비율	이산화질소	이산화탄소	아황산가스	미세먼지	초미세먼지
강남구	544097.8	1566004	39501008	0.039644659	0.028300926	0.640092166	0.004447005	44.04724	26.16194969
강동구	356574	592664	24590480	0.02410136	0.034662921	0.534831461	0.003543539	44.35493	24.74578652
강북구	187094.6	131893	23600355	0.005588602	0.02049435	0.475492958	0.002766197	44.94944	24.42696629
강서구	483784.3	1149247	41436846	0.027734905	0.034088608	0.534036568	0.005111111	47.34441	26.02265861

[출처] 2021 국가 교통비용 전망(NKIS)



미세먼지, 초미세먼지, 이산화질소, 이산화탄소, 아황산가스 일자 별 농도



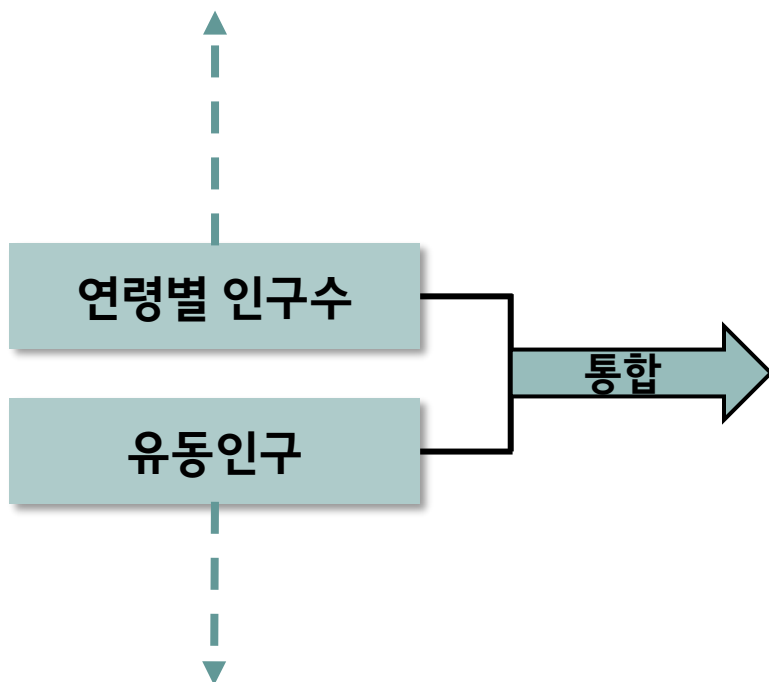
## 제2절 적용범위

## 1. 적용대상

○ 영유아, 학생, 어르신 등 건강취약계층

*[출처]건강 취약계층 보호를 위한 고농도 미세먼지 대응매뉴얼(환경부)*

미세먼지 취약계층 고려 = 만 19세 이하 어린이 + 만 65세 이상 노인

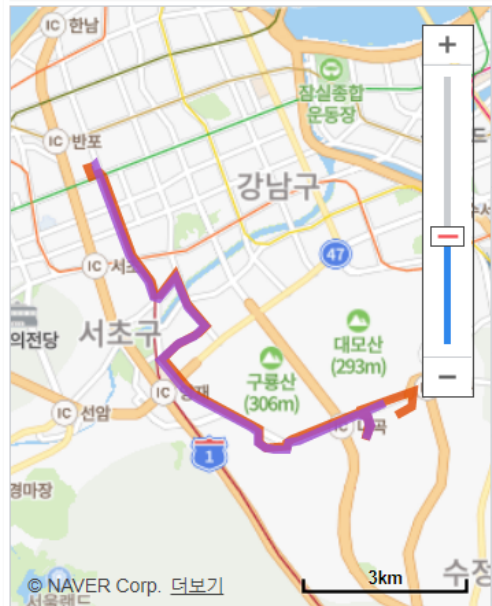


## ✓ 행정동 클러스터링을 위한 데이터 셋 제작

구	동	계	만19세이하어린이	만65세이상노인	어린이비율	노인비율	유동인구
강남구	개포동	52356	9085	7976	0.173523569	0.152342	13633390
강남구	논현동	44940	4020	6112	0.089452603	0.136004	24951023
강남구	대치동	81416	18425	9315	0.226306868	0.114412	35997109
강남구	도곡동	55524	11645	7718	0.209729126	0.139003	20615204

## Naver 검색 창 버스 이름 검색 결과

서울특별시 서초구 서초동 주변 09번 버스를 검색하시겠습니까?



**서초09** (서울 | 청화운수)  
**마을** 현인가구단지 ↔ 강남역  
 첫차 05:30 | 막차 23:00 | 배차간격 8분

실시간버스정보를 제공하지 않습니다.

- 현인가구단지 22343
- 현인마을 22383
- 현인마을.서울농업기술센터 22421
- 현인릉.꽃시장 22434
- 예비군훈련장 22448
- 신흥마을입구

배차간격 8분

✓ 배차간격에 해당하는 데이터만 크롤링

버스명	배차간격
강북01	9
강북02	6
강북03	7
강북04	10

[마을버스 배차간격]

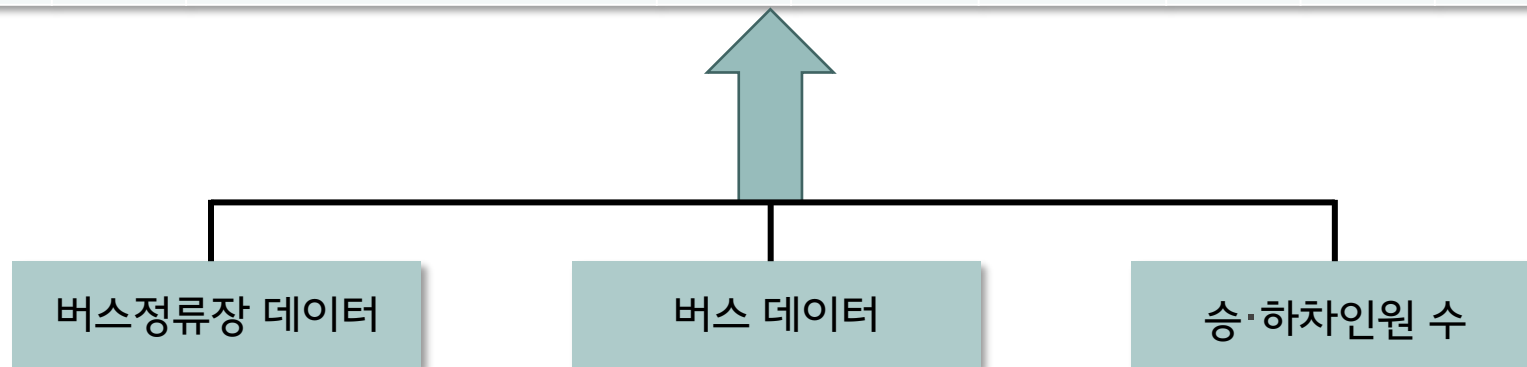
노선번호	배차간격
2	17
3	23
4	15
17	10

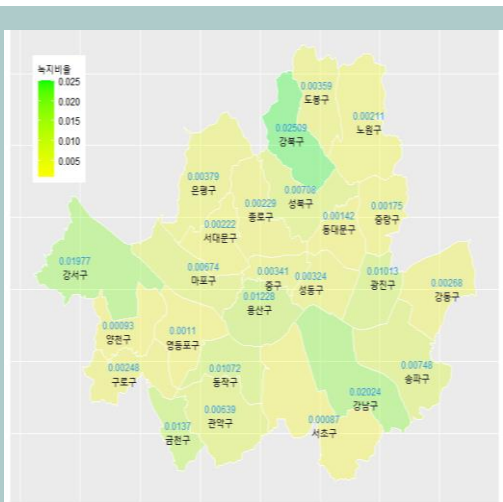
[ 시내버스 배차간격 ]

✓ X좌표, Y좌표로 주어진 버스정류장 위치를 역-지오코딩하여 지번주소로 변환



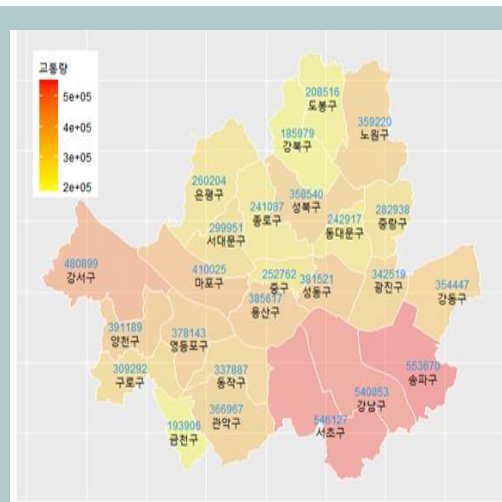
정류소명	구	동	지번주소	노선개수	승차총승객수	하차총승객수	총승객수	노선명	X좌표	Y좌표	배차간격
2호선교대역7번출구	서초구	서초동	서울특별시 서초구 서초동 1748-4	2	2393	1828	4221	서초10	127.0146	37.49407	13
HCN서초방송	서초구	서초동	서울특별시 서초구 서초동 1748-50	1	18	1974	1992	서초11	127.0121	37.48232	14
KCC사옥	서초구	반포동	서울특별시 서초구 반포동 749-35	2	1113	9100	10213	8541	127.0197	37.50311	15
KC대학교	강서구	화곡동	서울특별시 강서구 화곡동 1002-4	4	7791	6808	14599	강서01	126.8512	37.54947	8





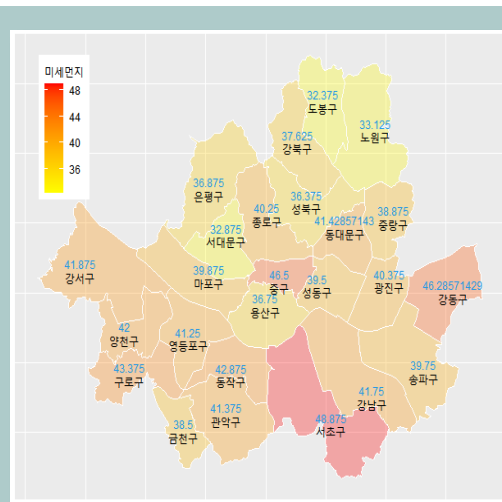
서울시 녹지비율

**강북구가 가장  
녹지비율이 높음**



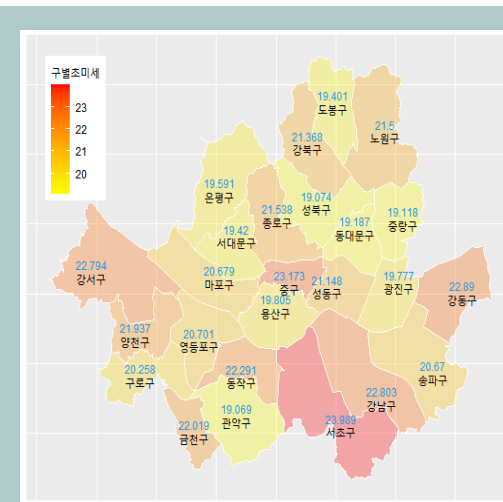
서울시 교통량

**송파구, 강남구,  
강서구, 서초구가  
교통량이 많음**



구별 미세먼지 평균

**서초구가 가장  
미세먼지 농도가  
높음**

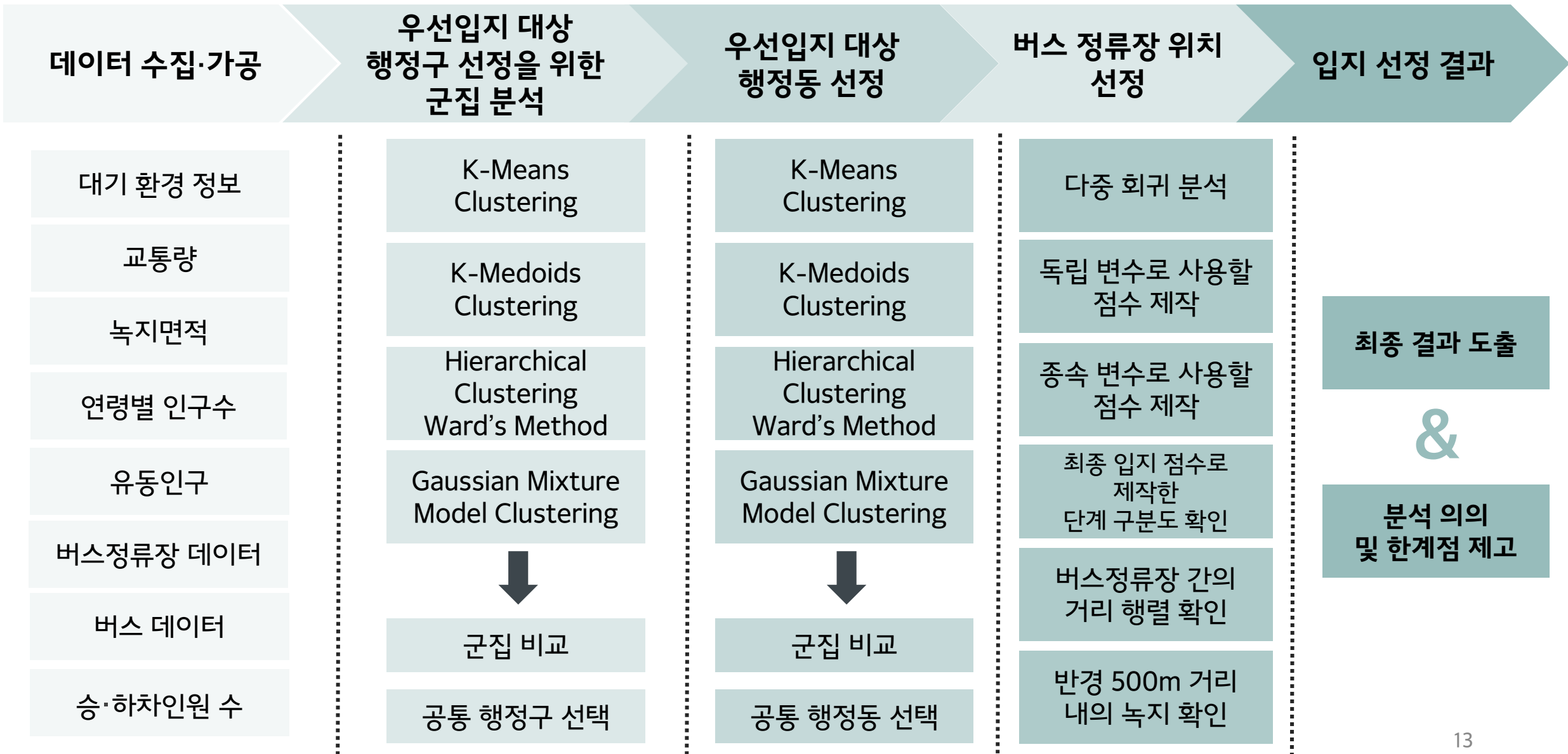


구별 초미세먼지 평균

**강남구, 강동구,  
강서구, 서초구가  
초미세먼지 농도가  
높음**

### 3 과제 흐름도

#### 분석 과정





# 우선 입지 선정 행정 구·동 선정을 위한 군집 분석 과정

K-Means  
Clustering

K-Medoids  
Clustering

Hierarchical  
Clustering  
Ward's Method

Gaussian Mixture  
Model Clustering

산출 결과의 신뢰도를 높이기 위해 동일한 데이터셋에 **4가지 알고리즘**을 모두 적용



군집 비교

각기 다른 알고리즘으로 도출된 4개의 군집 분석 결과를 확인하여  
군집을 이루는 행정 구·동이 가지는 변수 값이 우리가 원하는 해석에 가장 잘 부응하는 군집을 선택



공통 구역 선택

**선정된 4개의 군집 간의 교집합**에 해당하는 행정 구·동만을 우선입지 대상으로 선정

&lt;대기오염 노출 점수&gt;

&lt;예상 매연 배출 점수&gt;

독립 변수로  
사용할 점수 제작

회귀 식 제작을 위한 과정 1

다중 회귀 분석

회귀식을 제작하여  
버스정류장이  
최적 입지임을  
가장 잘 설명할 수 있는  
**<최종 입지 점수>** 도출

종속 변수로  
사용할 점수 제작

승차 총 승객 수,

대기오염 노출점수,

예상 매연 배출 점수를

범주형 변수(범주 4개)로 변환

✓ 세 범주형 변수의 합 = &lt;버스정류장의 총점&gt;

과정 2

최종 입지 점수로  
제작한  
단계 구분도 확인

버스정류장 간의  
거리 행렬 확인

반경 500m 거리  
내의 녹지 확인

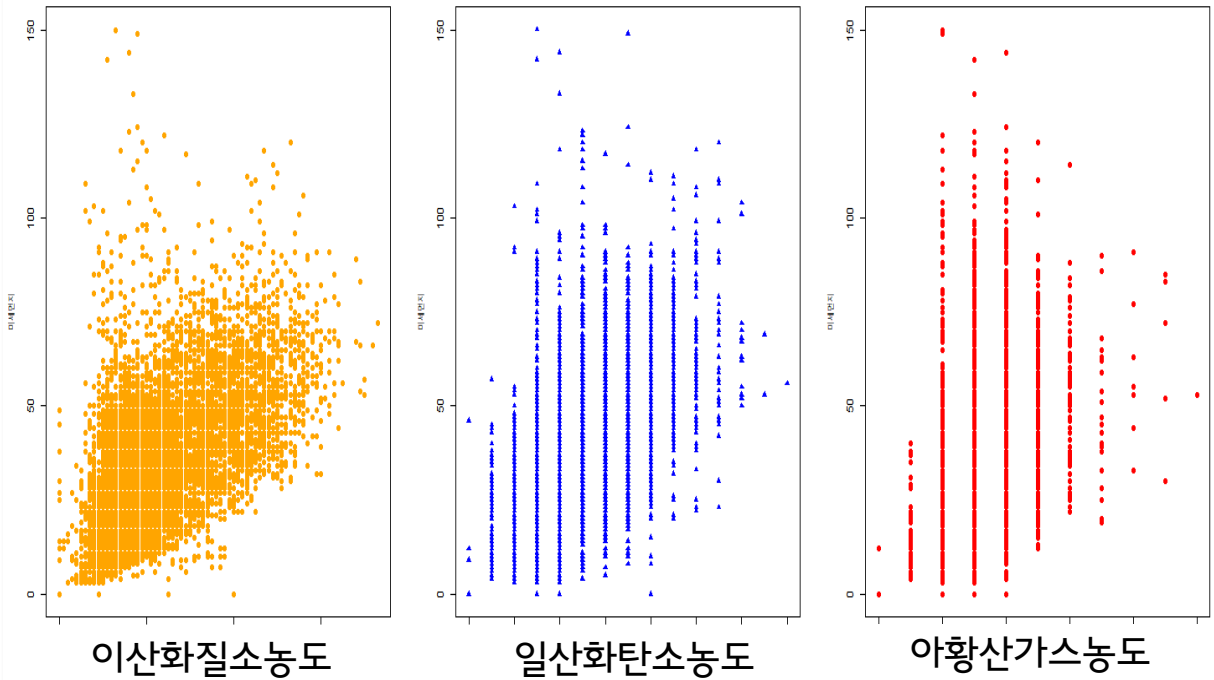
최종 결과 도출

# 4 미세먼지에 영향을 주는 요인 분석

## 분석 결과

### 상관분석 결과

미세먼지와 대기오염물질 산점도



미세먼지와 대기오염물질  
(매연 구성요소)  
상관계수가 **유의**함

	상관관계	p-value
미세먼지와 이산화질소	0.564	< 2.2e-16
미세먼지와 일산화탄소	0.592	< 2.2e-16
미세먼지와 아황산가스	0.422	< 2.2e-16

### 로지스틱 회귀 분석 결과

```
> modell=glm(미세먼지이론~교통량변수+녹지변수
+ ,family=binomial,data=datal)
> summary(modell)

Call:
glm(formula = 미세먼지이론 ~ 교통량변수 + 녹지변수, family = binomial,
    data = datal)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.274  -1.145  -1.022   1.210   1.342

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.17362    0.03974  -4.369 1.25e-05 ***
교통량변수   0.19870    0.02850   6.973 3.10e-12 ***
녹지변수    -0.10223    0.02869  -3.564 0.000366 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 12028  on 8686  degrees of freedom
Residual deviance: 11978  on 8684  degrees of freedom
AIC: 11984

Number of Fisher Scoring iterations: 3
```

녹지비율이 약 1% 증가,  
미세먼지 농도가  
“나쁨”일 확률 10% 감소

교통량이 약 16만대 증가,  
미세먼지 농도가  
“나쁨”일 확률 21% 증가

미세먼지 농도 ( $\mu\text{g}/\text{m}^3$ , 일평균)	좋음	보통	나쁨	매우나쁨
	0 ~ 30	31 ~ 81	81 ~ 150	151 이상

녹지 비율이 **크면** 미세먼지 농도가 상대적으로 **낮고**,  
교통량이 **증가**하면 미세먼지 농도가 상대적으로 **높아짐**

## 4 행정 구 군집분석을 위한 변수 선정

분석 결과

[미세먼지 ~ 매연 구성요소 ~교통량]

행정 구 선정을 위한  
군집 분석

★ 모든 변수 **표준화** 진행 후

초미세먼지&미세먼지농도와  
교통량과의 양의 상관관계 확인

{초미세먼지+미세먼지} 농도와  
{교통량} 변수를 사용해  
군집 분석 진행

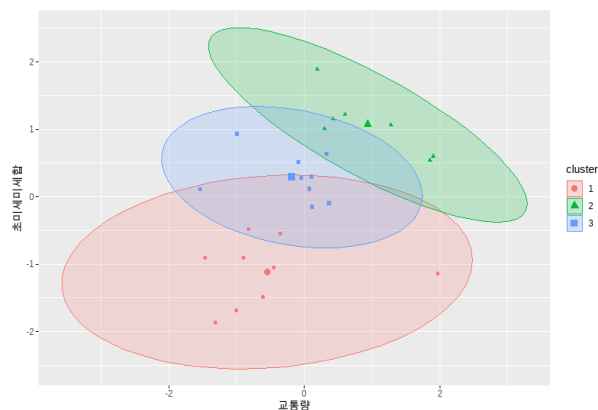
Pearson's product-moment correlation

```
data: frame_scaled$초미세미세한 and frame_scaled$교통량
t = 2.3199, df = 23, p-value = 0.02957
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.04871632 0.70865975
sample estimates:
cor
0.4354648
```

✓ 각 분석에서 군집 선정

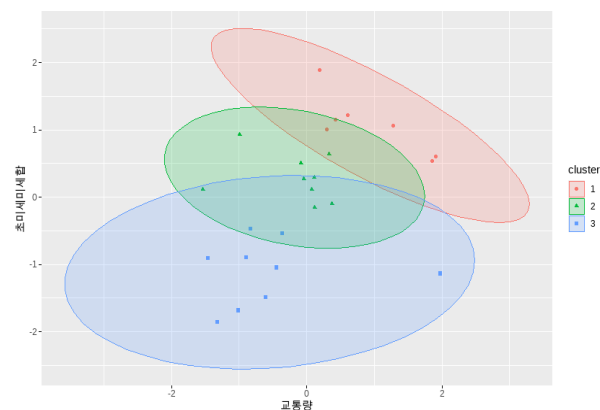
{교통량}과 {초미세먼지+미세먼지}의 수치가 높은 군집을 타겟으로 선정

K-Means 군집분석



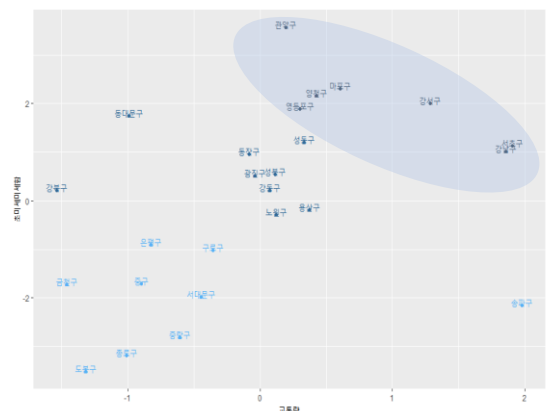
⇒ 군집2 (강남, 강서, 서초, 마포, 양천, 영등포, 성동, 관악, 동대문)

K-Medoids 군집분석



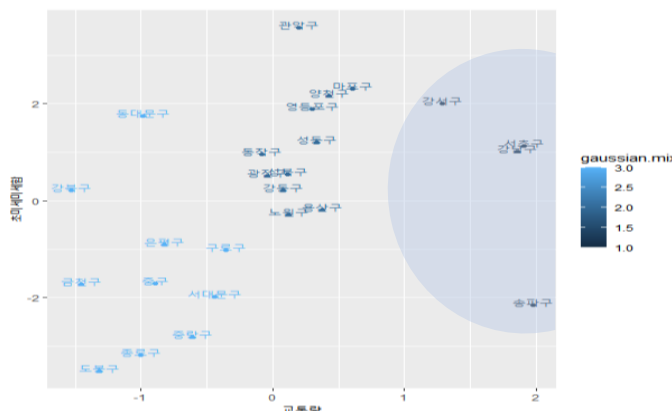
⇒ 군집1 (강남, 서초, 관악, 강서, 영등포, 마포, 양천)

계층적 군집분석(Ward)



⇒ 군집1 (강남, 서초, 관악, 강서, 영등포, 마포, 양천)

가우시안 혼합 군집분석



⇒ 군집1 (강남, 강서, 송파, 서초)

선정된 행정 구

강남구  
강서구  
서초구

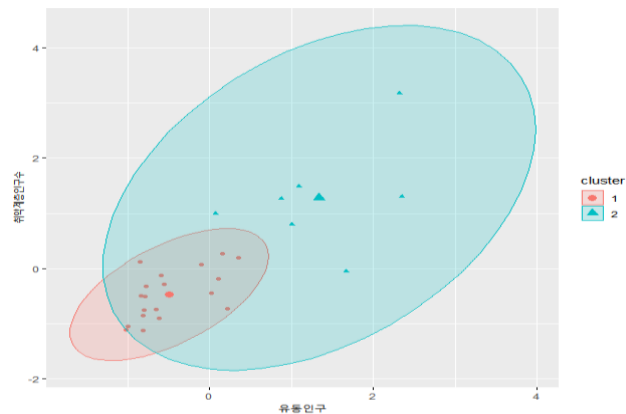
✓ 우선 설치 구 선정  
4가지 군집분석 결과를 통해  
공통적으로 선정된 구를  
'우선 설치 구'로 지정



✓ 각 분석에서 군집 선정

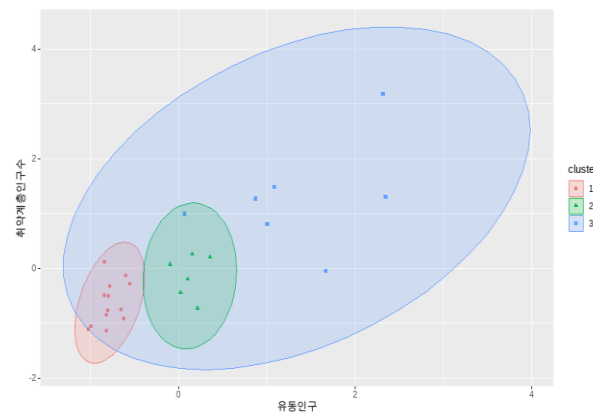
유동인구 수와 취약계층인구 수가 높은 군집을 타겟으로 선정

K-Means 군집분석



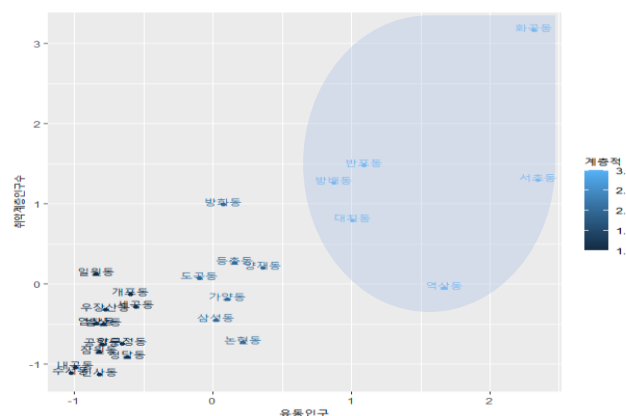
⇒ 군집2 (화곡, 방화, 역삼, 대치, 서초, 반포, 방배)

K-Medoids 군집분석



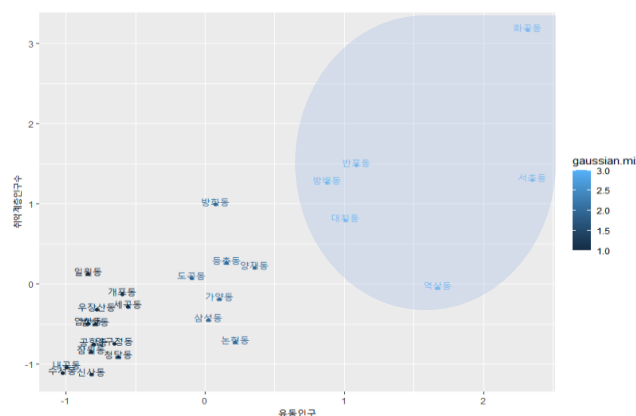
⇒ 군집3 (대치, 역삼, 방화, 화곡, 반포, 방배, 서초)

계층적 군집분석(Ward)



⇒ 군집3 (화곡, 대치, 반포, 방배, 역삼, 서초)

가우시안 혼합 군집분석



⇒ 군집3 (대치, 역삼, 화곡, 반포, 방배, 서초)

선정된 행정 동

강남구      대치동  
              역삼동

강서구      방화동  
              화곡동

서초구      반포동  
              방배동  
              서초동

✓ 우선 설치 동 선정

4가지 군집분석 결과를 통해

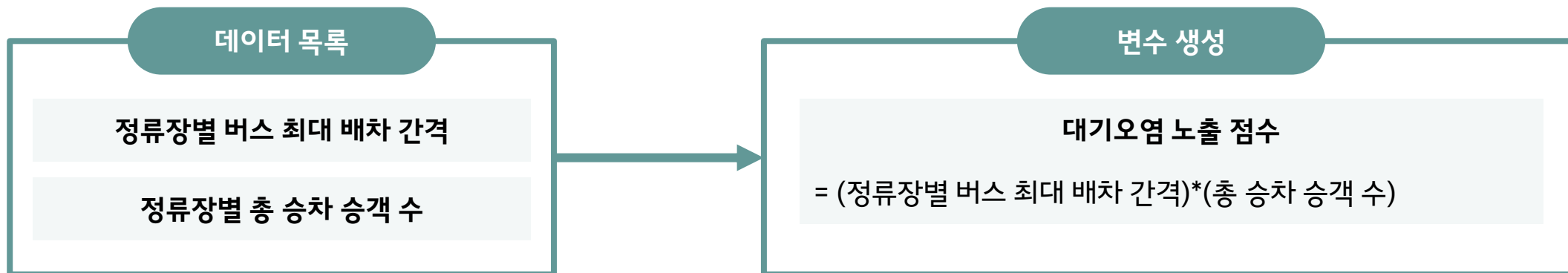
공통적으로 선정된 동을

‘우선 설치 동’으로 지정

## ✓ 독립변수1 : 대기오염 노출 점수

버스정류장 이용 승객들이 대기오염에 노출되는 정도를 반영하고자 함

버스대기 인원(승차승객)이 많고 배차간격이 길수록(최대 배차 간격) 대기 오염에 더 많이 노출됨

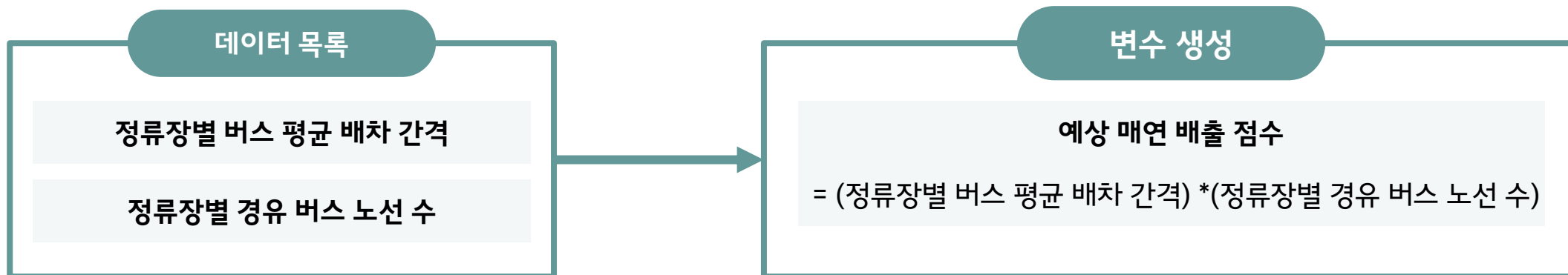


## ✓ 독립변수2 : 예상 매연 배출 점수

정류장마다 경유하는 버스 노선 수가 상이함

정류장의 정차하는 버스의 숫자가 많을 수록 매연에 더 많이 노출된다고 판단함

평균적인 버스 배차 간격에서의 매연 배출 정도를 알아보하고자 함



## ✓ 종속변수 : 총 오염지수

버스정류장 이용 승객 수, 대기오염 노출 정도, 예상 매연 배출 점수가 버스 정류장의 총 점수에 얼마나 영향을 미치는지 파악하기 위해 각 변수를 사분위수를 범주형으로 변환하여 총 오염지수를 도출함

## 데이터 목록

총 승차 승객 수

대기오염 노출 점수

예상 매연 배출 점수

정류소명	총 승차 승객 수	대기오염 노출 점수	예상 매연 배출 점수
2호선교대역 7번출구	2393	43074	31
HCN서초방송	18	252	14
KCC사옥	1113	16695	23
KC대학교	7791	163611	58
...	...	...	...

데이터 변환  
(0~3점 척도 범주형)

승차 승객(범주형)

대기오염노출(범주형)

예상 매연 배출(범주형)

정류소명	승차 승객 (범주형)	대기오염노출 (범주형)	예상매연배출 (범주형)
2호선교대역 7번출구	1	1	1
HCN서초방송	0	0	0
KCC사옥	0	0	1
KC대학교	2	2	2
...	...	...	...

## 변수 생성

총 오염지수

= 승차 승객(범주형)  
+ 대기오염노출(범주형)  
+ 예상매연배출(범주형)

정류소명	총 오염지수
2호선교대역 7번출구	3
HCN서초방송	0
KCC사옥	1
KC대학교	6
...	...

- ✓ 다중선형회귀분석을 통한 버스정류장별 최종 입지 점수 도출

앞서 생성한 독립변수와 종속변수로 다중회귀분석을 진행하고

이를 바탕으로 최종 입지 점수를 도출하여 행정 동 내의 설치가 시급한 정류장을 선정함



단계적 변수 선택법을 활용하여 최종 회귀식 생성

$$(\text{최종 입지 점수}) = 1.4824 + 0.7721 * (\text{대기 오염 노출 점수}) + 0.5355 * (\text{예상 매연 배출 점수})$$



회귀식을 활용해 버스 정류장 별 **최종 입지 점수** 도출

정류소명	X좌표	Y좌표	총 승차 승객 수	예상 매연 배출 점수	대기 오염 노출 점수	총 점수	최종입지점수	노선개수
고속터미널	127.0038415	37.50563712	10.98119828	5.237443938	10.29711265	9	12.23745	36
지하철2호선강남역	127.0257283	37.50160933	9.031698315	4.955995156	7.695575549	9	10.07809	37
...	...	...	...	...	...	...	...	...

- ✓ 최종 입지 점수를 바탕으로 최종 설치 입지(버스정류장) 선정

설치 입지가 한 지역에 치우치지 않도록 배정하기 위해

선정된 지역의 버스정류장 간의 거리 행렬을 도출하여 500m 이내 간격끼리 집합 생성하고  
집합 내에서 최종 입지 점수가 가장 높은 버스정류장만 선택함

### 버스 정류장 간의 거리 행렬 도출

Q-GIS를 활용해 선정된 행정동 내의 버스정류장 간 거리행렬을 도출함

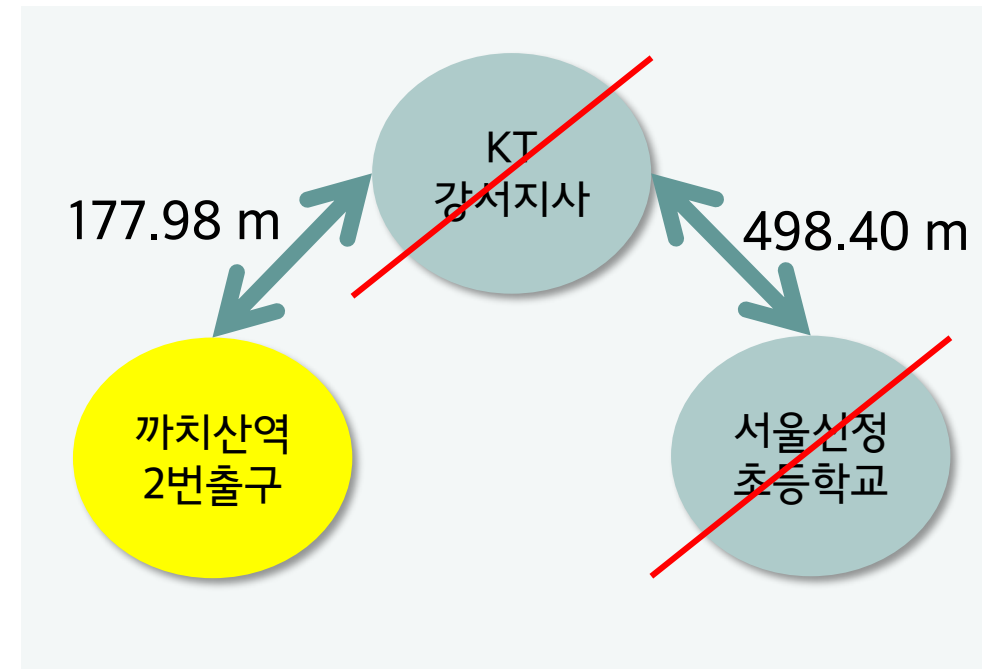
InputID	TargetID	Distance
고속터미널	지하철2호선강남역	1986.216493
고속터미널	이수역	2774.460262
고속터미널	화곡역	15152.17507
고속터미널	뱅뱅사거리	3220.845999
...	...	...

단위: m(미터)

### 최종 설치 입지(버스정류장) 선정

500m 이하의 간격을 가진 집합 내

최종 입지 점수가 가장 높은 버스정류장 하나만 남기고 나머지는 제거





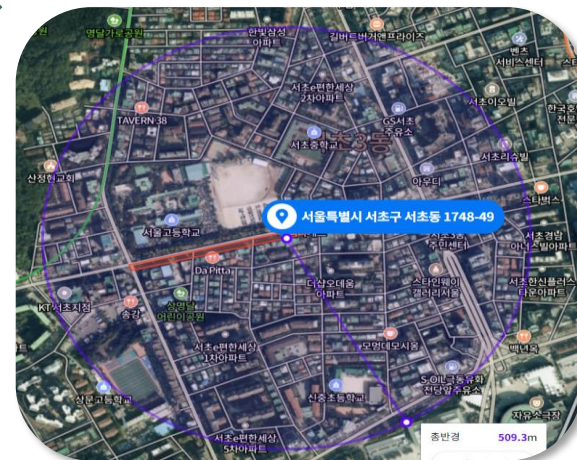
1. 최종점수로  
단계 구분도 생성



2. 최종점수가 높은  
상위 2그룹만 선정

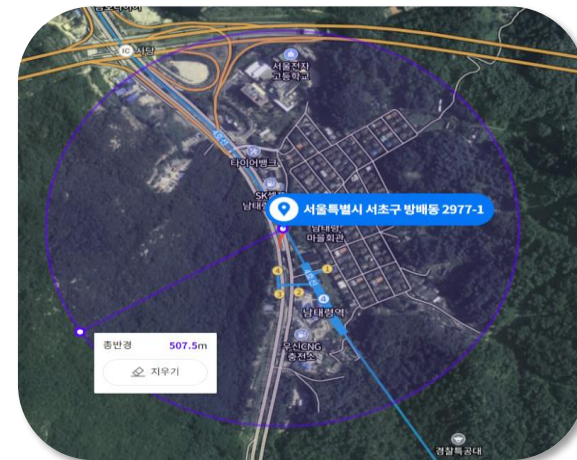


주변에 **녹지가 적은** 정류장



주변에  
녹지비율이 적은 정류장은  
**최종 입지 선정 후보로  
선택**

주변에 녹지가 있는 정류장



주변에  
녹지가 존재할 경우  
후보에서 삭제

단계 구분

값: 1.2 최종점수

심볼: [선택]

범례 포맷: %1 - %2

메소드: 색상

색상표: [색상 선택]

분류: 히스토그램

심볼	값	범례
○	3.221530 - 4.180375	3.22 - 4.18
●	4.180375 - 6.345233	4.18 - 6.35
★	6.345233 - 11.008048	6.35 - 11.01
★	11.008048 - 19.730413	11.01 - 19.73

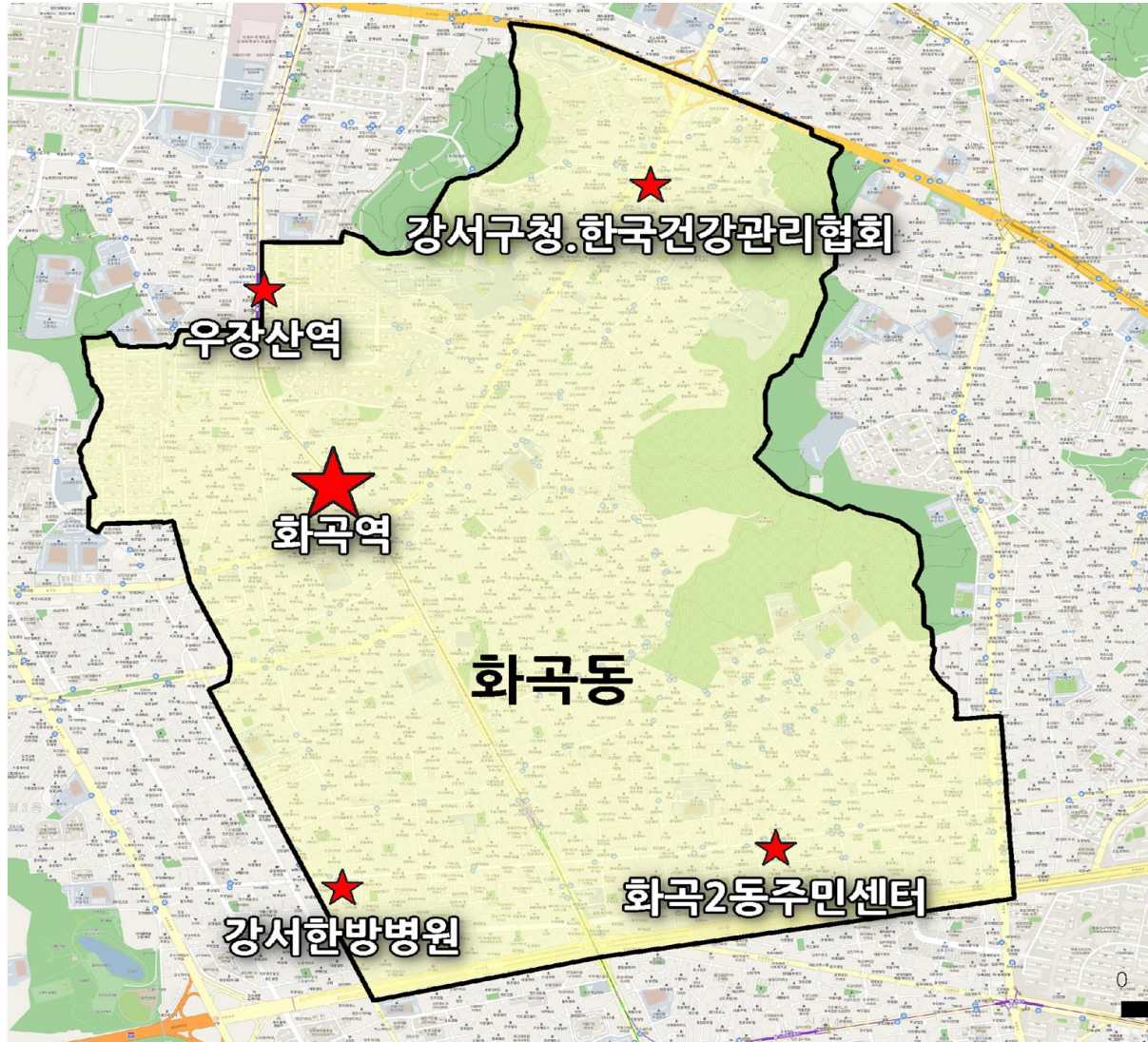
<input type="checkbox"/>	○	3.22 - 4.18
<input type="checkbox"/>	●	4.18 - 6.35
<input checked="" type="checkbox"/>	★	6.35 - 11.01
<input checked="" type="checkbox"/>	★	11.01 - 19.73

## 4

## 버스정류장 최적 입지 선정

## 분석 결과

✓ 최종입지 점수 & 버스 정류장 간의 거리 & 녹지유무를 고려하여 최종 15개의 정류장 선정



정류소명	최종입지점수
<b>화곡역</b>	<b>15.37</b>
강서구청.한국건강관리협회	10.38
우장산역	10.04
강서한방병원	6.99
화곡2동주민센터	6.51

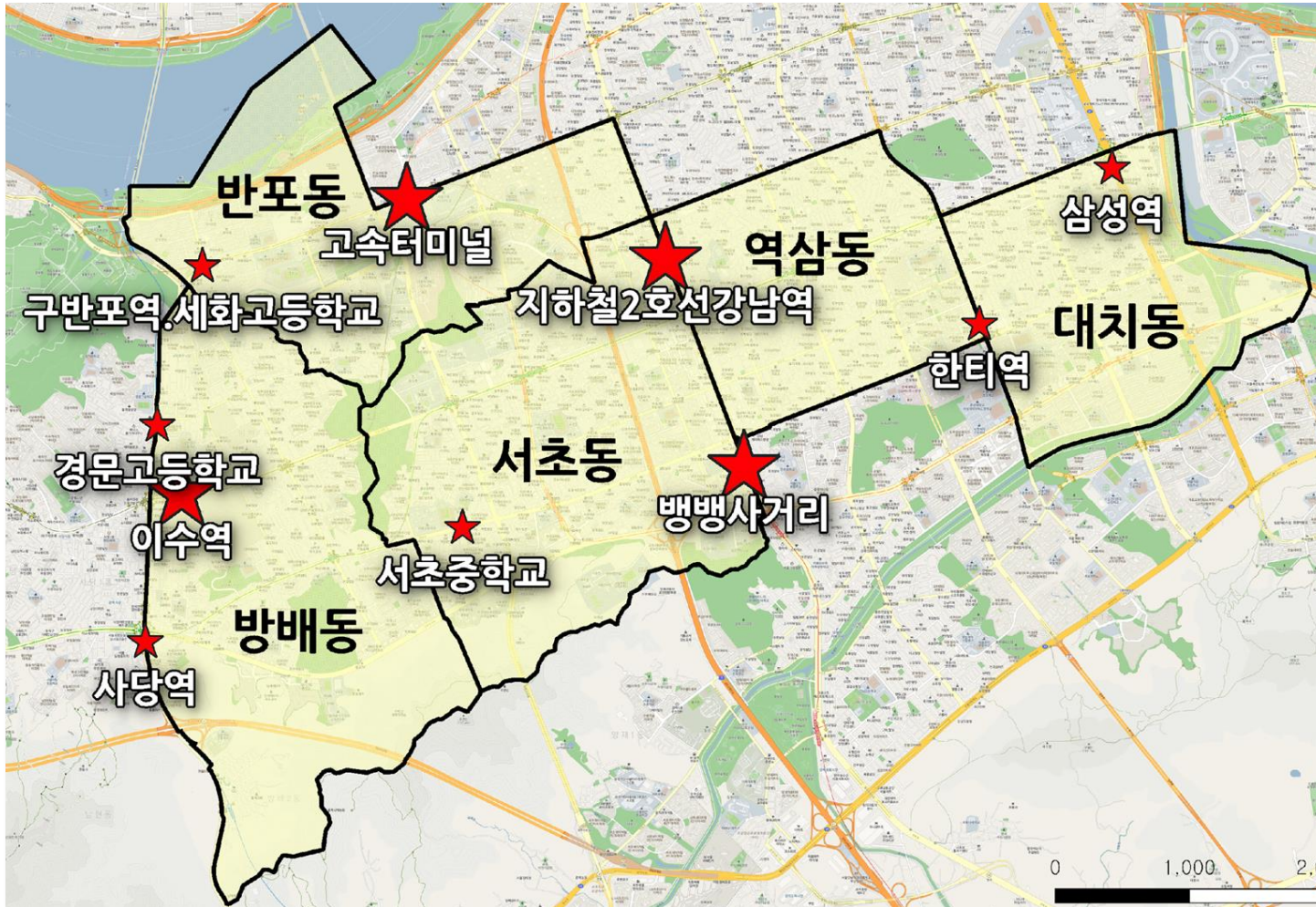


## 4

## 버스정류장 최적 입지 선정

## 분석 결과

✓ 최종입지 점수 & 버스 정류장 간의 거리 & 녹지유무를 고려하여 최종 15개의 정류장 선정



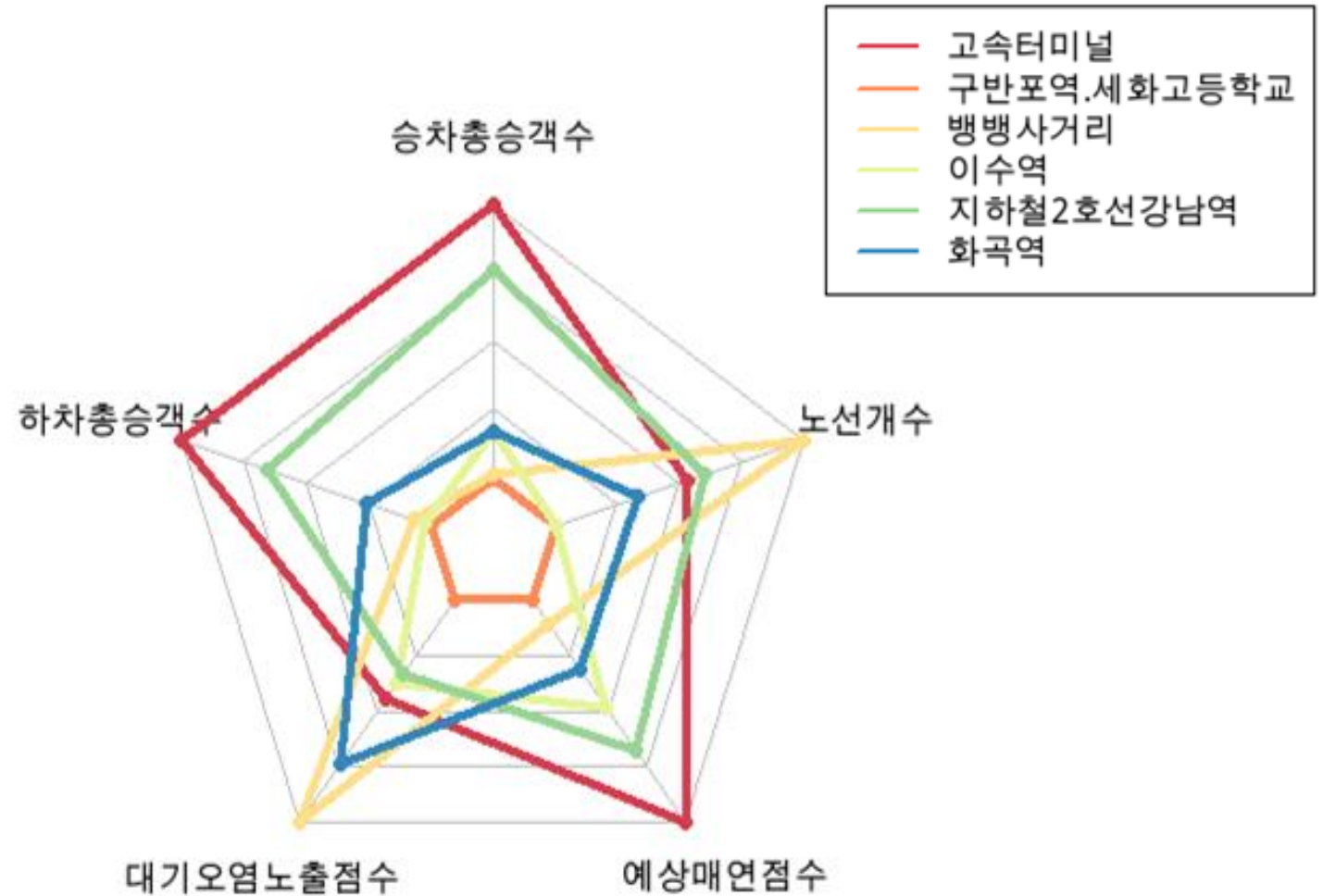
정류소명	최종입지점수
고속터미널	19.73
지하철2호선강남역	16.99
이수역	15.65
뱅뱅사거리	14.55
구반포역.세화고등학교	11.01
사당역	10.97
삼성역	9.957
경문고등학교	7.208
서초중학교	7.192
한티역	6.683

## 4

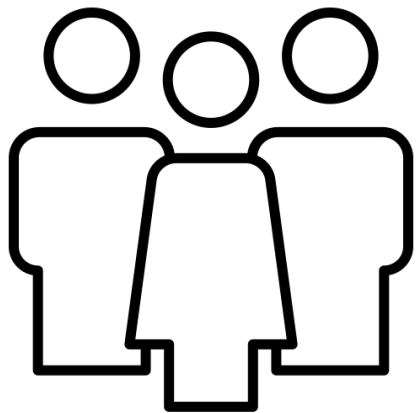
## 버스정류장 최적 입지 선정

분석 결과

	정류소명	최종입지점수
1	고속터미널	19.73041
2	지하철2호선강남역	16.99493
3	이수역	15.65326
4	화곡역	15.37317
5	뱅뱅사거리	14.54793
6	구반포역.세화고등학교	11.00805
	사당역	10.97151
	강서구청.한국건강관리협회	10.37869
	우장산역	10.04196
	삼성역	9.957398
	경문고등학교	7.208252
	서초중학교	7.191514
	강서한방병원	6.990001
	한티역	6.683166
	화곡2동주민센터	6.514494

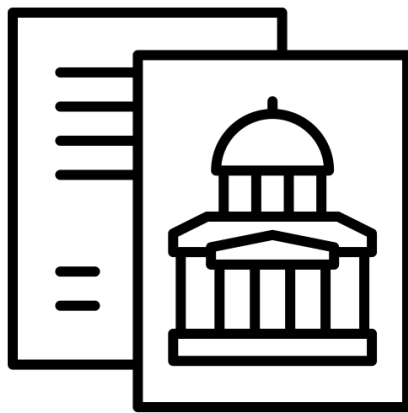


최종 입지로 선정된 상위 6개의 버스 정류장에 대한 시각화  
최종입지 선정순위가 레이더차트 도형 크기와 비례함



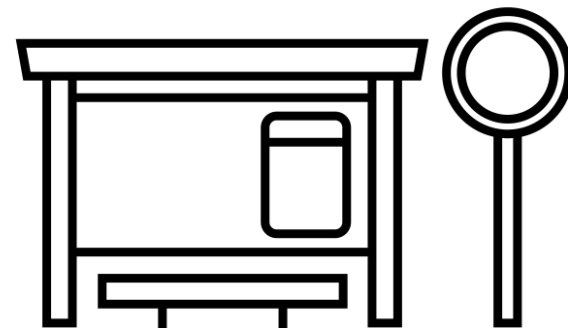
### 국민의 삶의 질 개선

기존의 버스정류장을 개선하여  
국민을 위한 맞춤형 심터 제공



### 정부 정책에 기여

환경부 미세먼지 관리  
종합계획에 기여



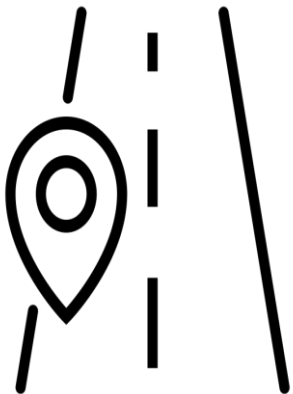
### 지자체에 기여

버스정류소 개선사업에 기여

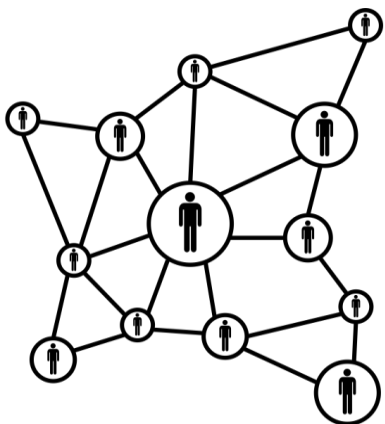


# 업무 활용 방안 및 분석에서의 한계점

## ✓ 업무 활용 방안



서울시 도로정책과  
버스정류장 개선사업에  
입지 선정 모델 제공



지자체에  
각 지역의 특성을 고려한  
최적입지 선정 모델 제공

## ✓ 원본 데이터가 지닌 한계

버스정류장 주변 미세먼지 농도의  
정확한 측정값 확보가 어려움.  
자체적으로 만든 수치로 분석을 진행했기 때문에,  
실제 정류장의 미세먼지 수치와의 차이 존재.

버스정류장 주변 녹지를 정확한 방법으로 측정한  
것이 아닌, 네이버 지도를 통해 육안으로 확인했기  
때문에 부정확할 수 있음.