



Data Mining Project

No Drinks on the Bus

통계학과 고은영, 조연선, 한이정



- 1** 분석 주제 및 목표
- 2** 분석 데이터
- 3** 분석 과정
- 4** 분석 결과 활용 방안





**커피 등 음식물을
들고 타지 맙시다**

시내버스 운전자는 여객안전 및 피해예방을 위해
음식물이 담긴 포장 컵 또는 불결·악취 물품의
운송을 거부할 수 있습니다.

※서울특별시조례 제6730호 2018.1.4., 일부개정

서울시내 버스 음식물 반입 금지

2018.1.4부터 시행

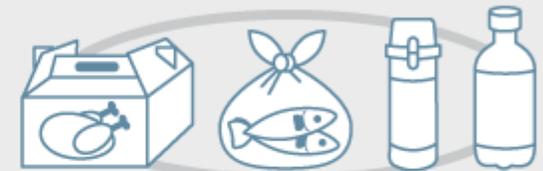
✕ 반입 금지

서울 시내버스 음식물 반입

○ 반입 허용



- 일회용 컵에 담긴 뜨거운 음료나 얼음
- 일회용 컵에 담긴 치킨·떡볶이 등 음식물
- 여러 개의 일회용 컵을 운반하는 용기에 담긴 음식물
- 뚜껑이 없거나 빨대가 꽂힌 캔·플라스틱 병에 담긴 음식물

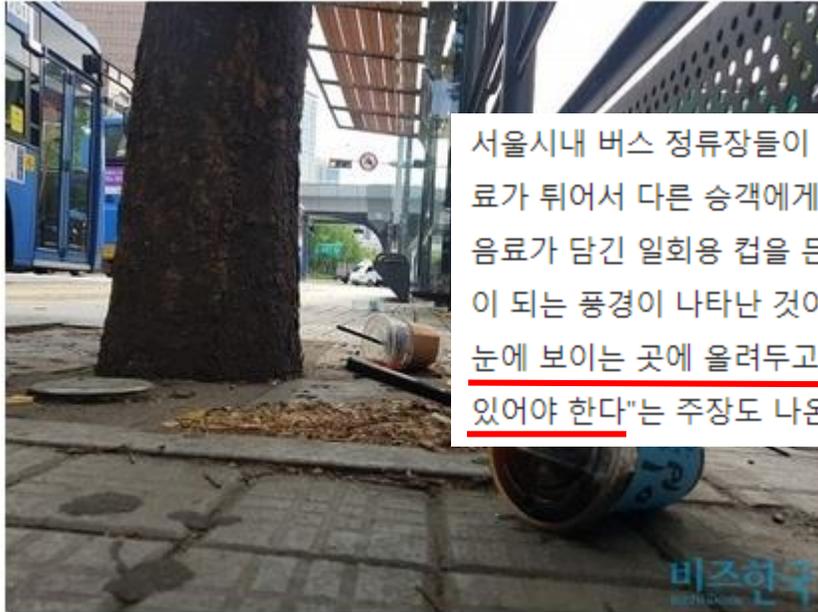


- 종이 상자에 포장된 치킨·피자 등 음식물
- 뚜껑이 닫힌 플라스틱병에 담긴 음료
- 따지 않은 캔에 담긴 음식물
- 밀폐형 텀블러·보온병에 담긴 음식물
- 비닐봉지에 담긴 채소·어류·육류

자료=서울시



서울시내 버스 음식물 반입 금지 시행 이후 버스정류장 모습



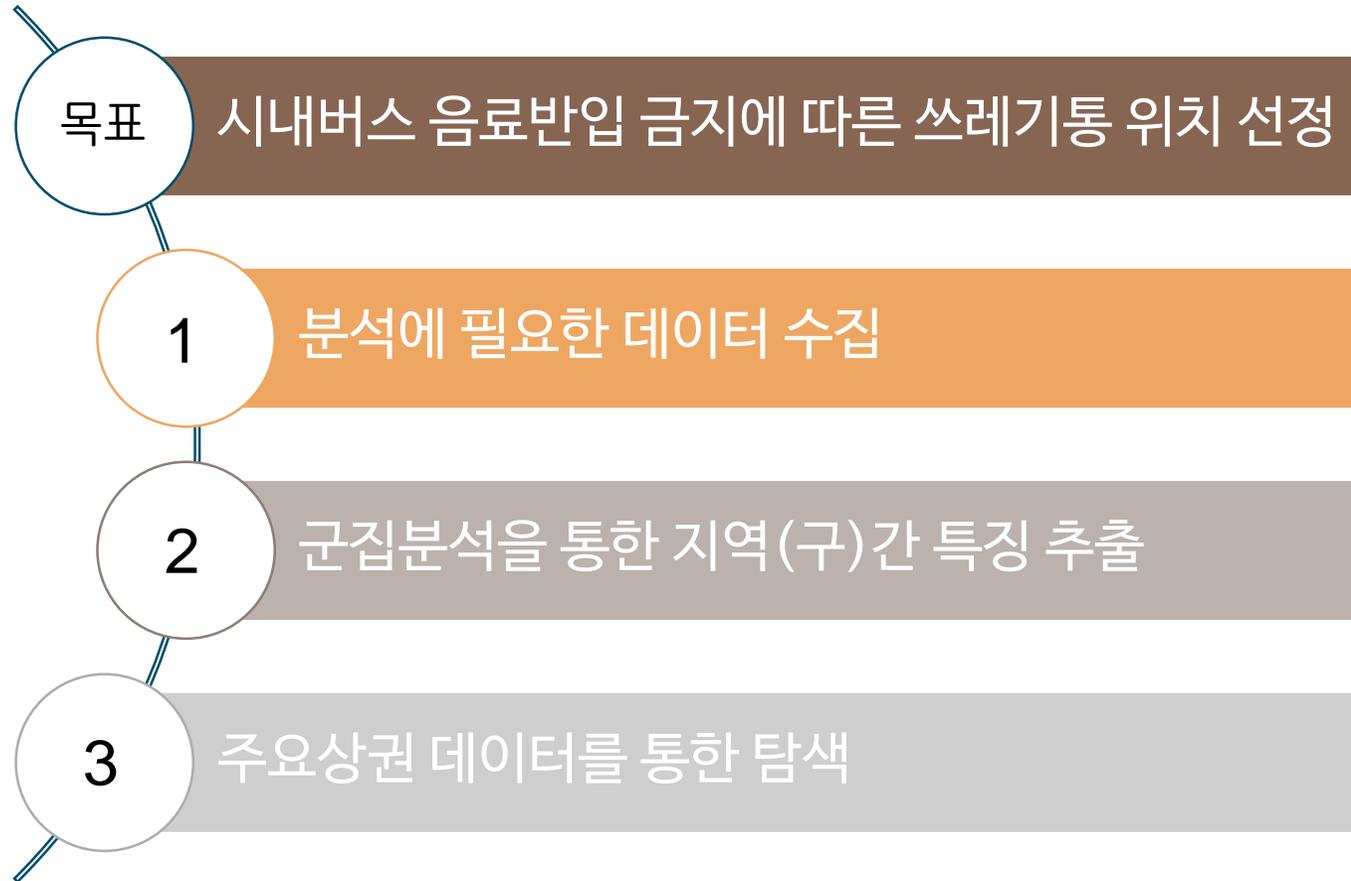
서울시내 버스 정류장들이 일회용 컵 쓰레기로 몸살을 앓고 있다. 올 1월부터 서울시가 '뜨거운 음료가 튀어서 다른 승객에게 화상 등 피해를 입힐 수 있다'는 이유로 조례를 개정해, 버스 탑승 시 음료가 담긴 일회용 컵을 든 승객은 버스기사가 승차를 거부할 수 있도록 하자 길거리가 쓰레기통이 되는 풍경이 나타난 것이다. 직장인 신모(27)씨는 "버스는 타야 하는데 승차를 거부당하면 일단 눈에 보이는 곳에 올려두고 승차할 수밖에 없지 않느냐"고 했다. "정류장마다 쓰레기통이 비치돼 있어야 한다"는 주장도 나온다.

시내버스 음식물 반입이 금지된 뒤 먹던 음료를 버리고 타는 승객이 늘면서 정류장은 쓰레기로 몸살을 앓는다. 민원이 제기되자 서울시는 정류장 쓰레기통 비치를 검토 중이다. 사진=박현광 기자



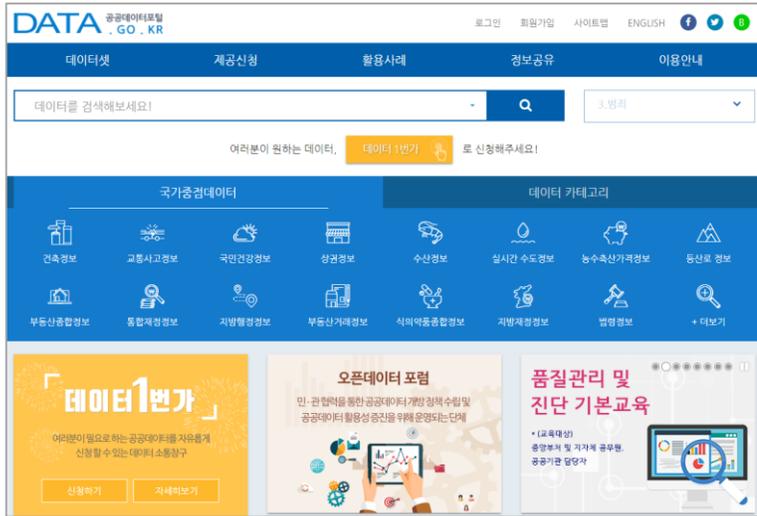
서울 아현역 인근 버스 정류장 간이 벤치에 음료잔이 늘어서 있다. 버스 기사가 음료수 반입을 거부할 수 있게 되자 사람들이 컵을 버리고 탑승한다. /성유진 기자

분석주제 및 목표



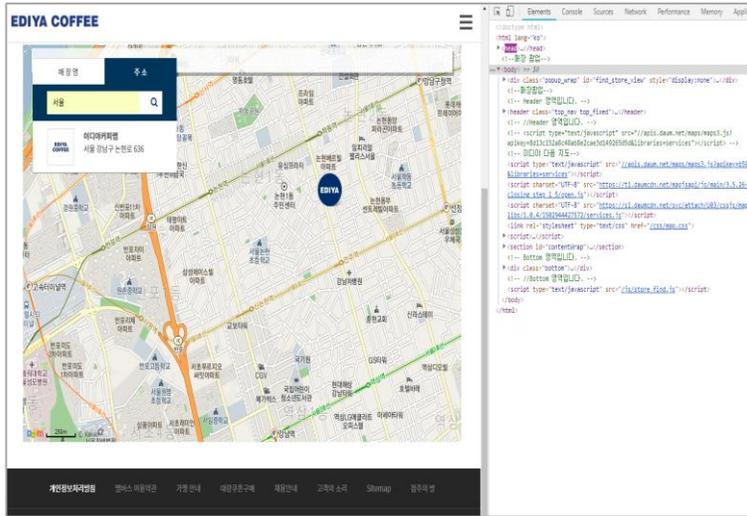


데이터 수집 과정



공공데이터 API

소상공인 상가업소 정보 서비스 API
서울특별시 정류소정보조회 서비스 API



크롤링

프랜차이즈 카페 지점정보
Sample: 투썸, 할리스, 이디야, 백다방
가로쓰레기통 위경도 정보

201810사자구난위서물상환인구(내국인)	2018-12-05 오전...	Microsoft Excel ...	64,676KB
201810행정동별서울생활인구(내국인)	2018-11-28 오후...	Microsoft Excel ...	175,014KB
airport	2018-12-05 오후...	Microsoft Excel ...	1KB
API설명서-상권정보	2018-11-28 오후...	한컴오피스 한글 ...	1,103KB
busloc.Rda	2018-12-01 오후...	RDA 파일	242KB
busmarket_1.Rda	2018-12-04 오후...	RDA 파일	4KB
gooraw.Rda	2018-12-05 오후...	RDA 파일	2KB
graph	2018-12-04 오후...	R 파일	3KB
market_bygoo.Rda	2018-12-04 오후...	RDA 파일	41KB
table_market.Rda	2018-12-04 오후...	RDA 파일	41KB
tmp1.json	2018-12-05 오후...	JSON 파일	2KB
상권정보시스템	2018-11-28 오후...	Chrome HTML D...	12,460KB
생활인구스키마-공개	2018-11-30 오후...	Microsoft Excel ...	173KB
서울버스정류소위치정보	2018-12-01 오후...	Microsoft Excel ...	489KB
서울버스정류소위치	2018-12-01 오후...	Microsoft Excel ...	489KB
서울시 버스노선별 정류장별 시간대별 승하차 인원 정보	2018-12-06 오전...	Microsoft Excel ...	634,517KB
서울시구별사업제및종사자수	2018-12-03 오후...	엑스트 문서	55KB
서울시구별사업제및종사자수	2018-12-03 오후...	Microsoft Excel ...	34KB
서울시구별위경도	2018-12-04 오후...	Microsoft Excel ...	1KB
서울시구별주민등록인구	2018-12-03 오후...	엑스트 문서	4KB
서울시구별주민등록인구	2018-12-03 오후...	Microsoft Excel ...	12KB
서울시키퍼미용료상관지표(구별)	2018-12-01 오후...	Microsoft Excel ...	12KB
서울시행정구역(구별)통계	2018-11-30 오후...	엑스트 문서	2KB
서울시행정구역(구별)통계	2018-11-30 오후...	Microsoft Excel ...	10KB
서울시행정구역(동별)통계	2018-11-30 오후...	엑스트 문서	22KB
행정동코드_행정정보_2018	2018-11-30 오후...	Microsoft Excel ...	35KB

기타자료 수집

서울시 정보소통광장
서울시 열린데이터광장
소상공인 상권정보시스템



데이터 요약

데이터명	출처	설명
주요상권별 커피점매출	소상공인 상권정보시스템	상권별 상하반기 커피점 평균매출
행정구역별 커피점매출	소상공인 상권정보시스템	행정구별 상하반기 커피점 평균매출
서울시커피음료상권지표(구별)	우리마을가게 상권분석서비스	프랜차이즈/일반 커피점 수/유동인구
서울특별시 가로쓰레기통 설치정보	서울시 정보소통광장	자치구별 쓰레기통 설치장소
서울시 생활인구 (내국인)	서울시 정보소통광장	행정동/나이/성별에 따른 시간별 생활인구(KT)
서울시 버스노선별 정류장별 승하차 인원 정보	서울시 정보소통광장	10월 일단위 정류장별 승차/하차 총승객수
서울시 주민등록인구 (연령별/구별) 통계	서울시 정보소통광장	행정동별 주민등록인구
서울 버스정류소 위치정보	서울시 열린데이터광장	정류소명, 위도, 경도
서울시 사업체현황 (종사자규모별/동별) 통계	서울시 열린데이터광장	행정동별 종사자수, 사업체수
서울시 행정구역별 통계	서울시 열린데이터광장	행정구별 면적
서울시 행정동코드 매핑정보	서울시 열린데이터광장	구/동별 코드 정보
서울시 구별 위경도자료	skt	서울시 자치구 중심지 정보
서울시 지도 shp자료	행정안전부	서울시 행정구별 공간정보
geocode API	네이버 지도 API	주소 -> 좌표 변환
reverse geocode API	네이버 지도 API	좌표 -> 주소 변환
소상공인 상가업소 정보 서비스 API	공공데이터포털	반경내 상권조회/행정구역 단위 상권조회
서울특별시 정류소정보조회 서비스 API	공공데이터포털	지정 위도,경도,반경 내 정류소정보조회



■ 생활인구 데이터

생활인구 정의

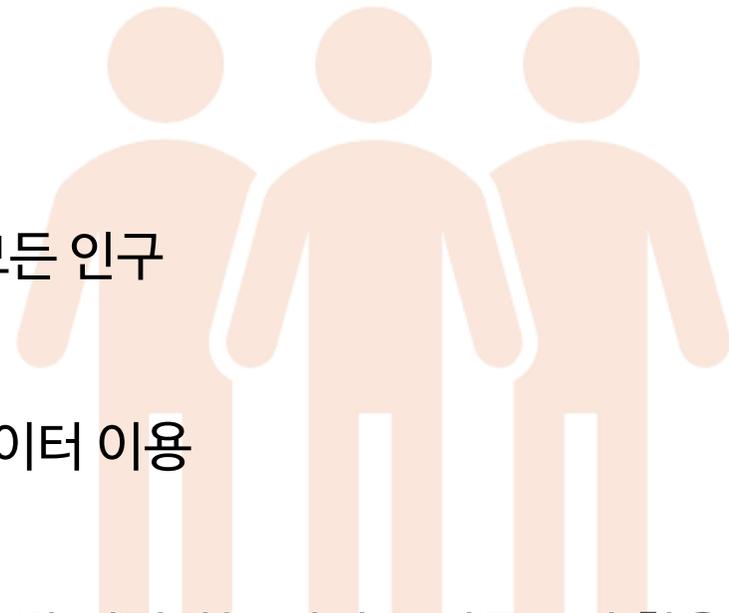
: 1시간 단위로 측정된 서울시 각 지역단위에 존재하는 모든 인구

생활인구 추계

: 서울시 보유 공공빅데이터와 KT 휴대폰 LTE 시그널 데이터 이용

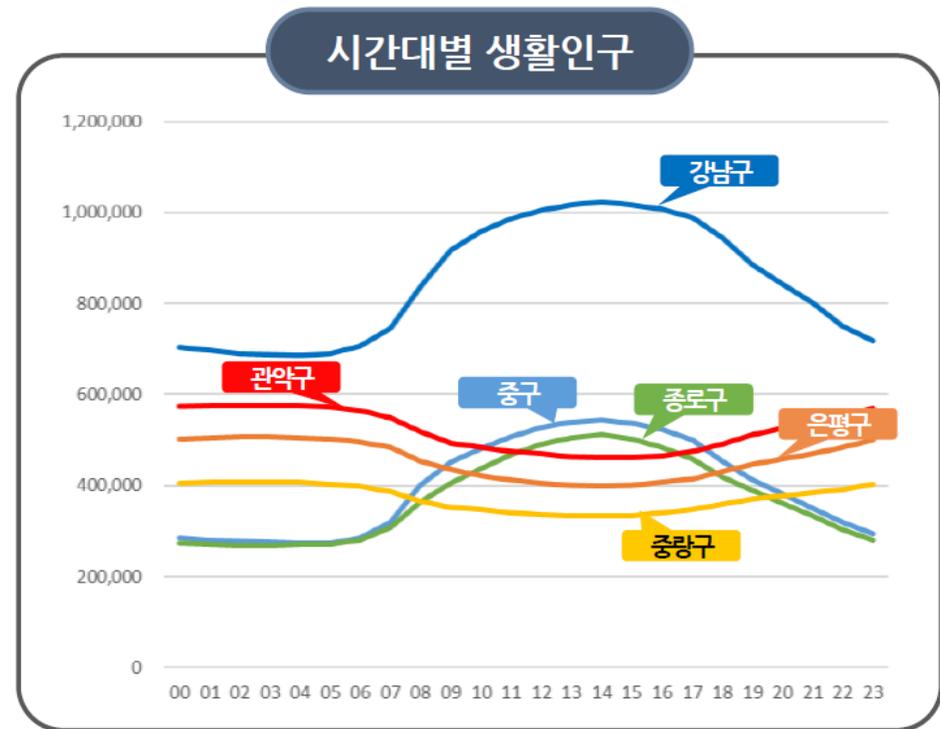
사용시 이점

: 실제 생활하는 인구데이터를 통해 도시서비스의 수요를 잘 설명하는 서비스 인구로서 활용 가능





생활인구 데이터



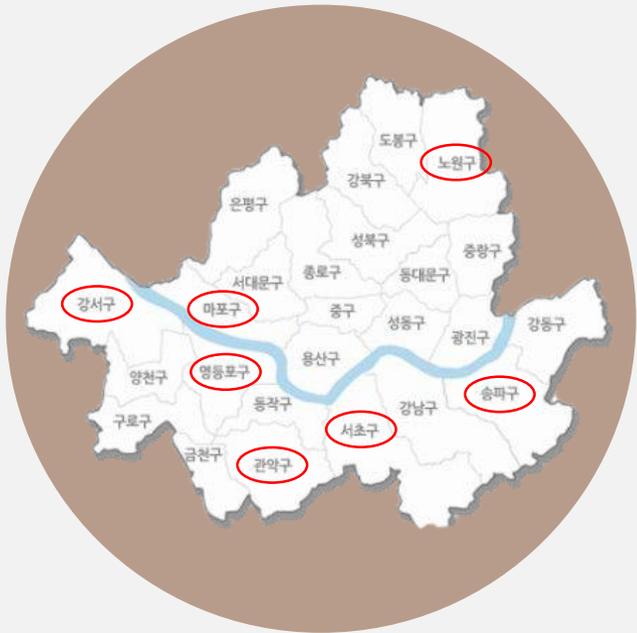
출처: 서울시 정보기획관

낮 시간대 강남구, 중구, 종로구 증가 (사업체가 집중된 지역)

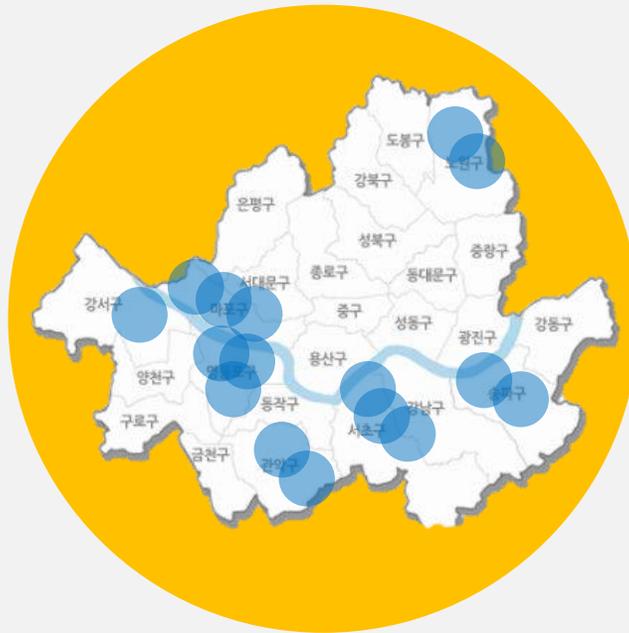
커피전문점 이용 시간을 고려,
주간생활인구 (9시~22시), 야간생활인구 (22시~9시)로 정의



■ 분석 과정 예시



자치구단위 군집 추출



주요상권추출

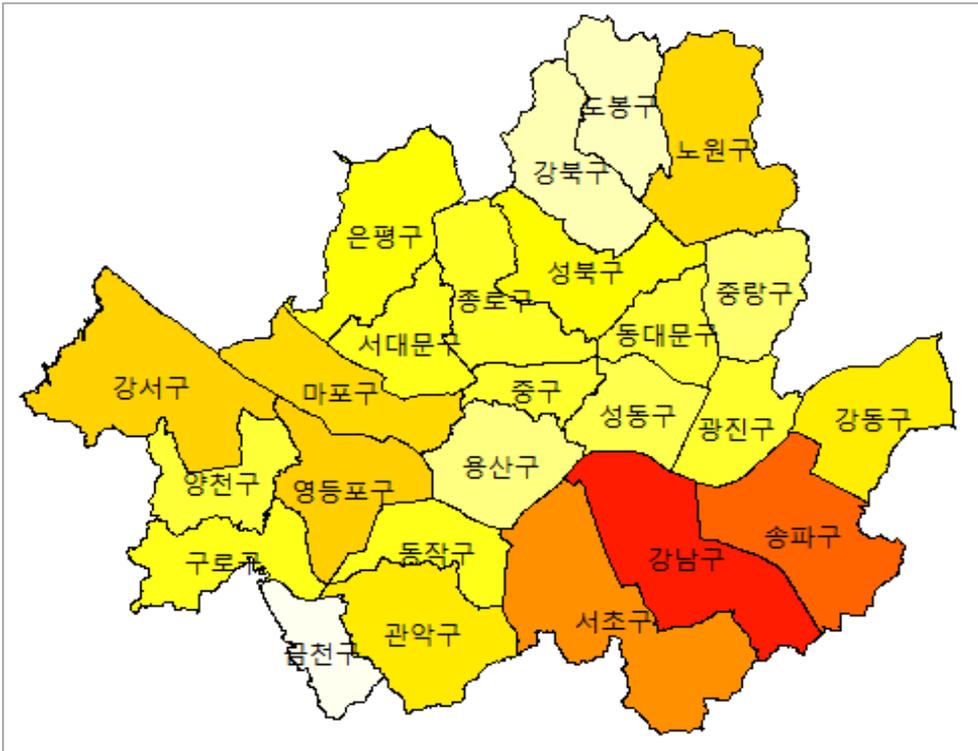


쓰레기통 위치 선정

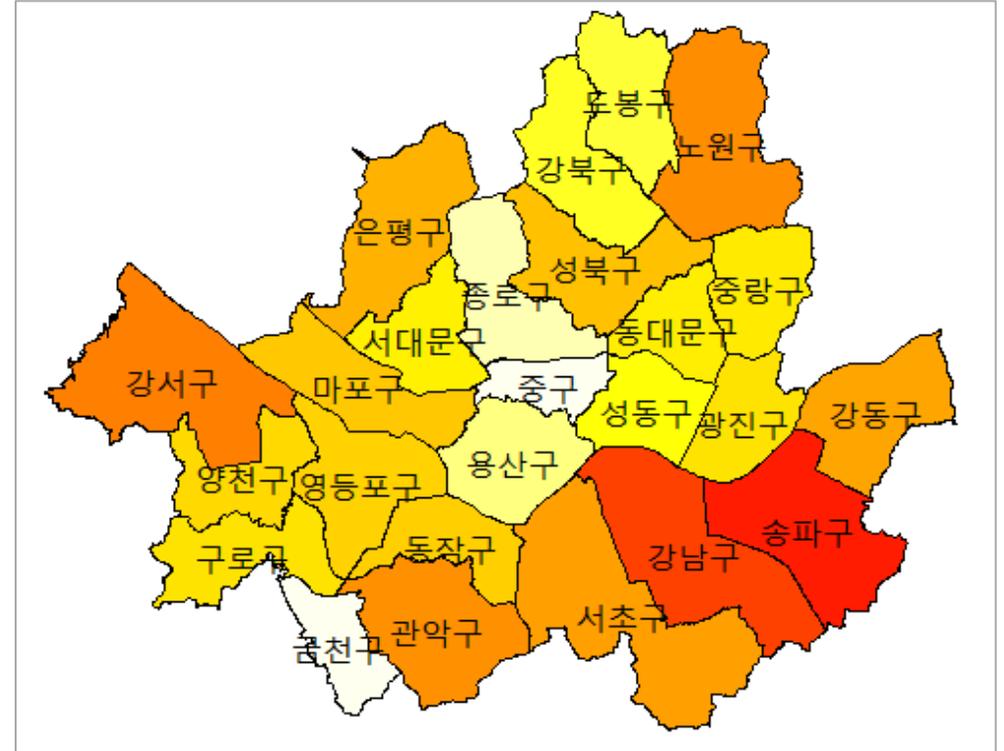


■ 탐색적자료분석(EDA)

주간 생활인구



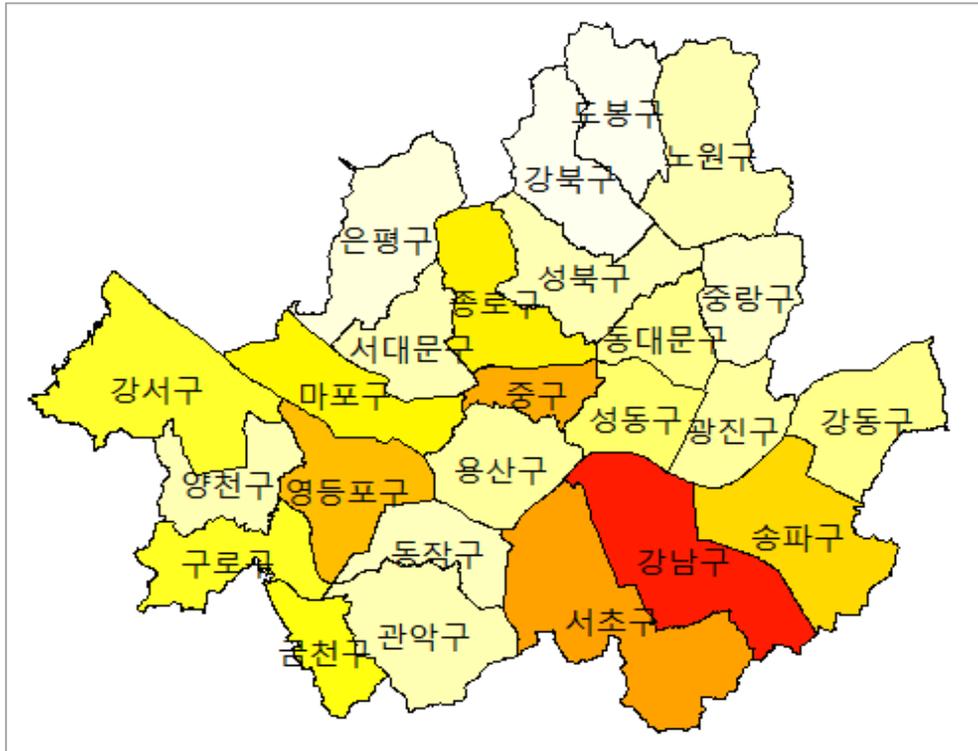
야간 생활인구



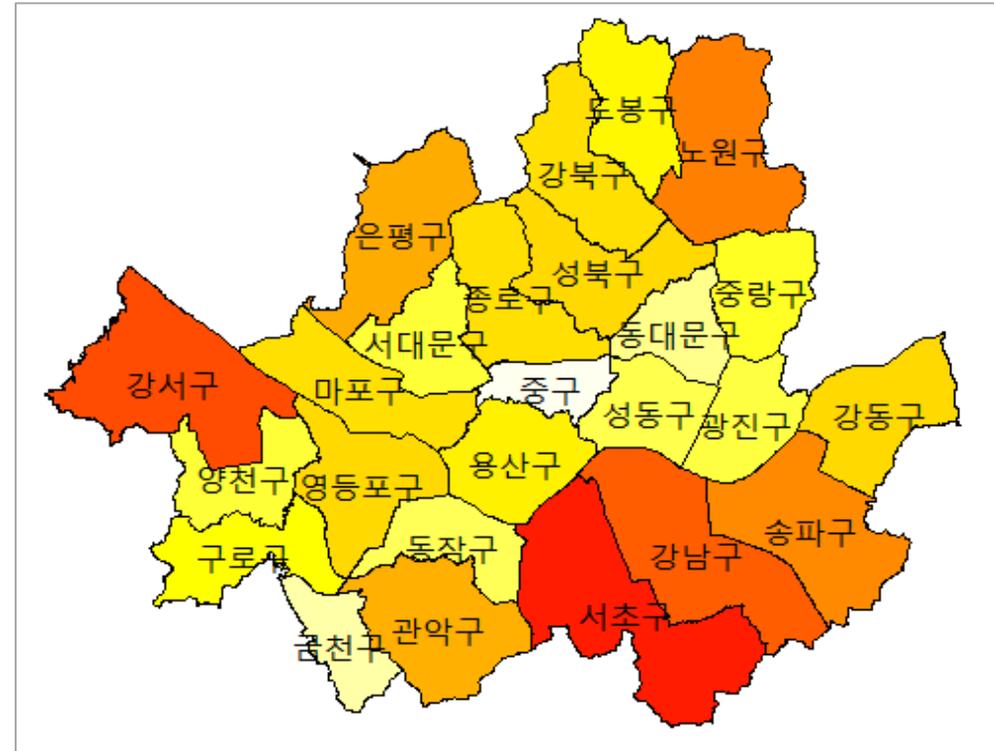


■ 탐색적자료분석(EDA)

종사자수



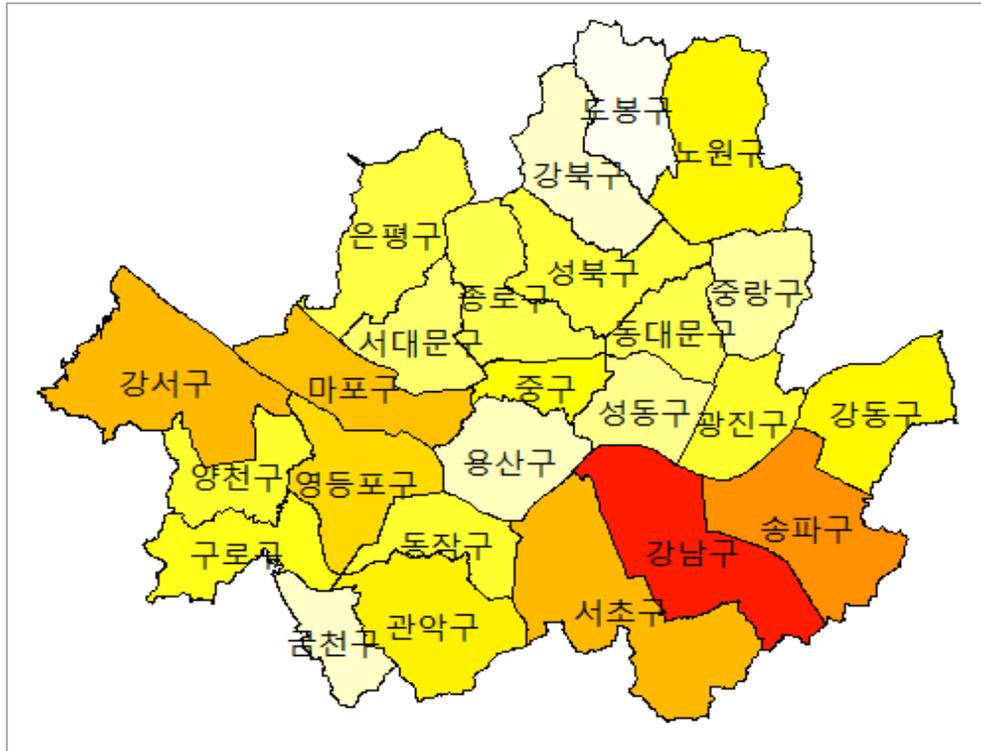
면적



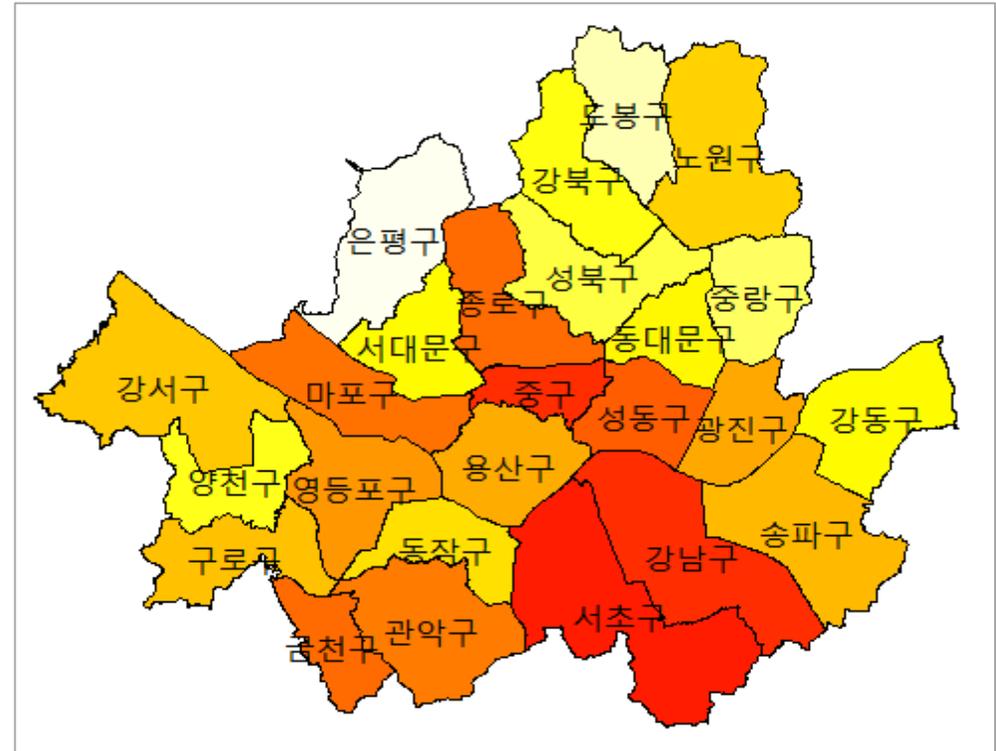


■ 탐색적자료분석(EDA)

프랜차이즈 커피점



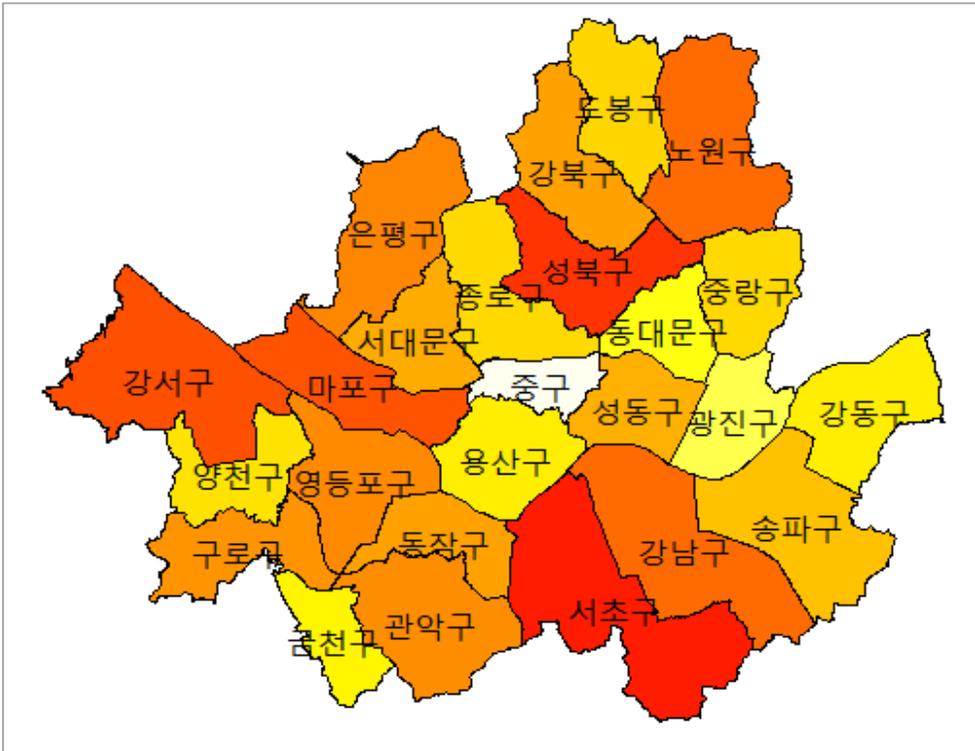
월 평균 매출



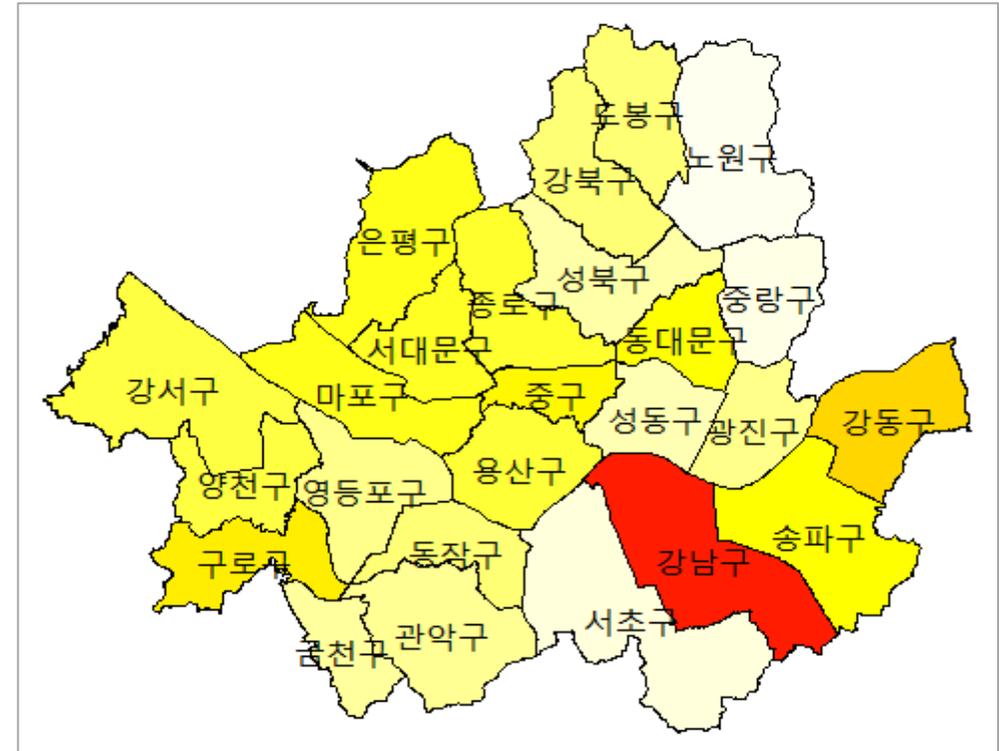


■ 탐색적자료분석(EDA)

버스정류장



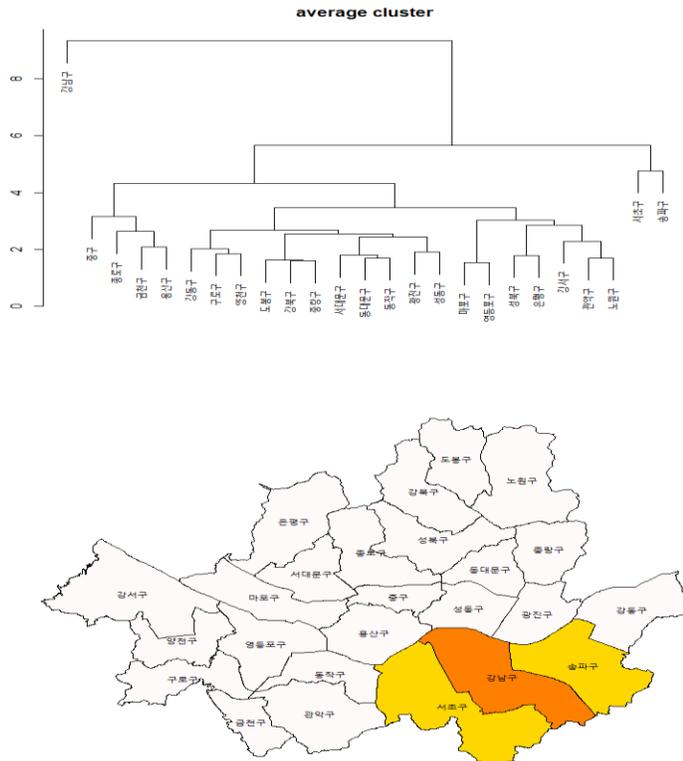
가로 쓰레기통



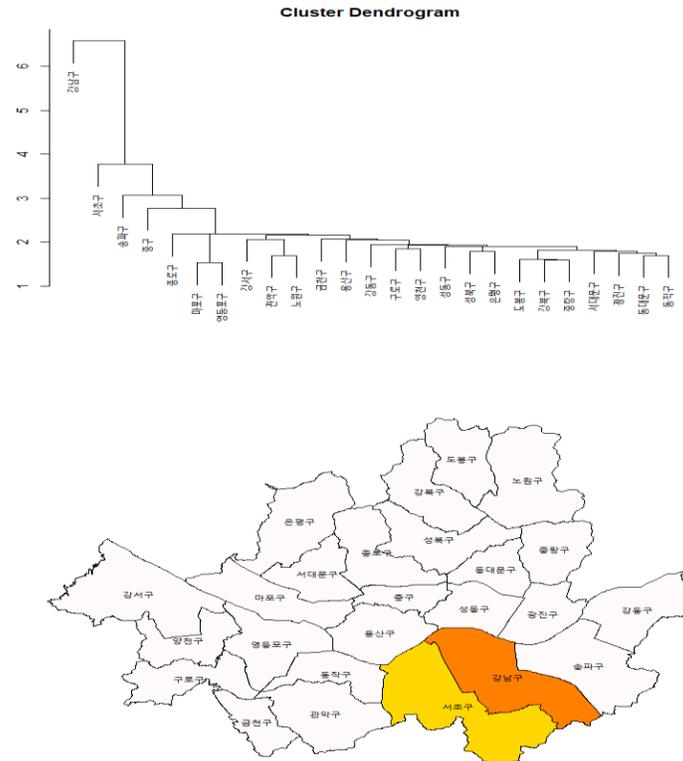


Clustering

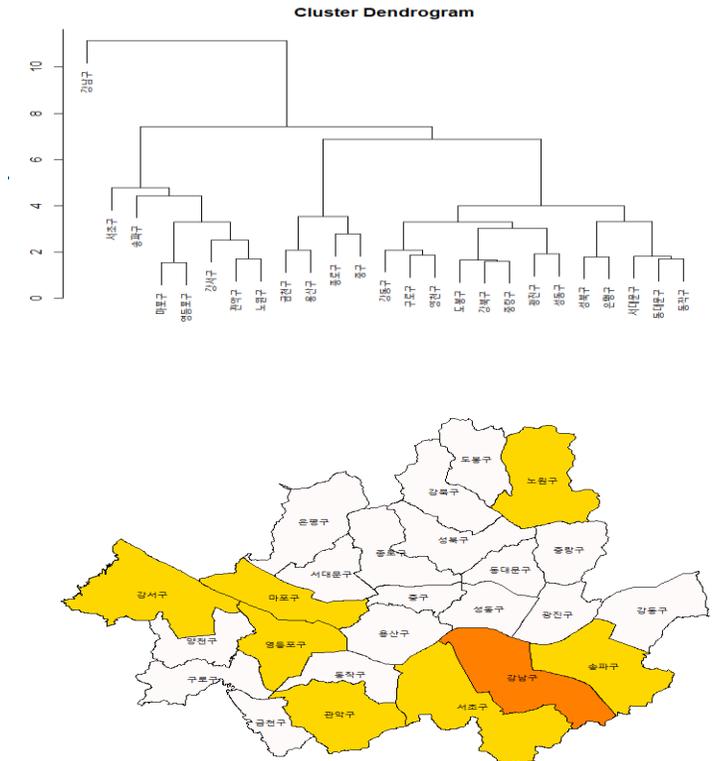
average



single

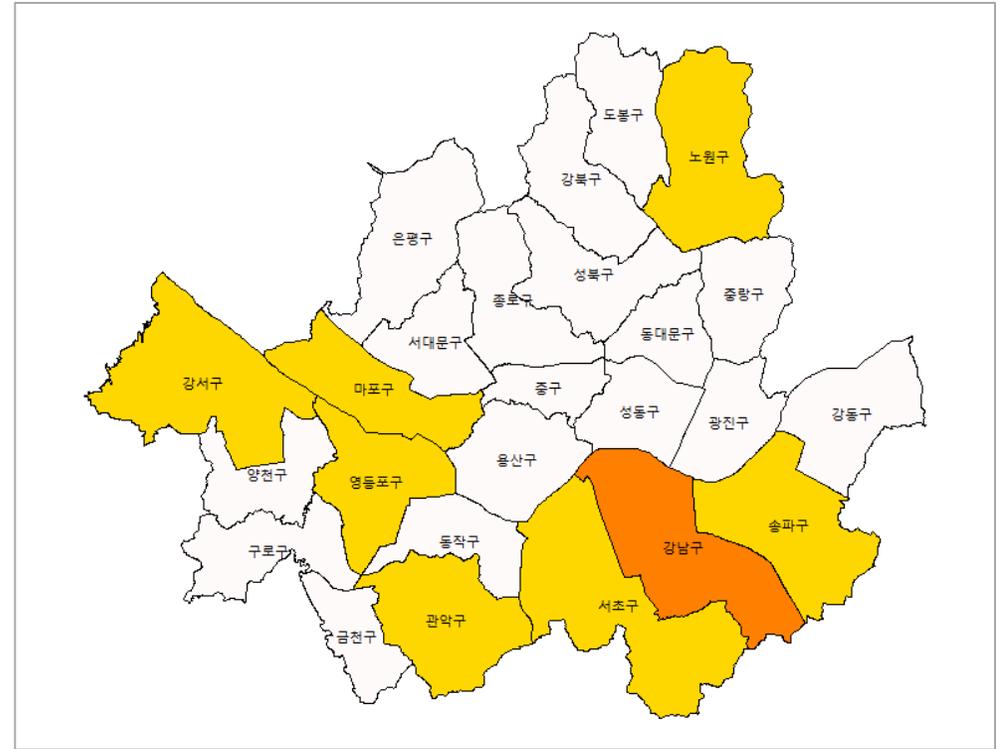
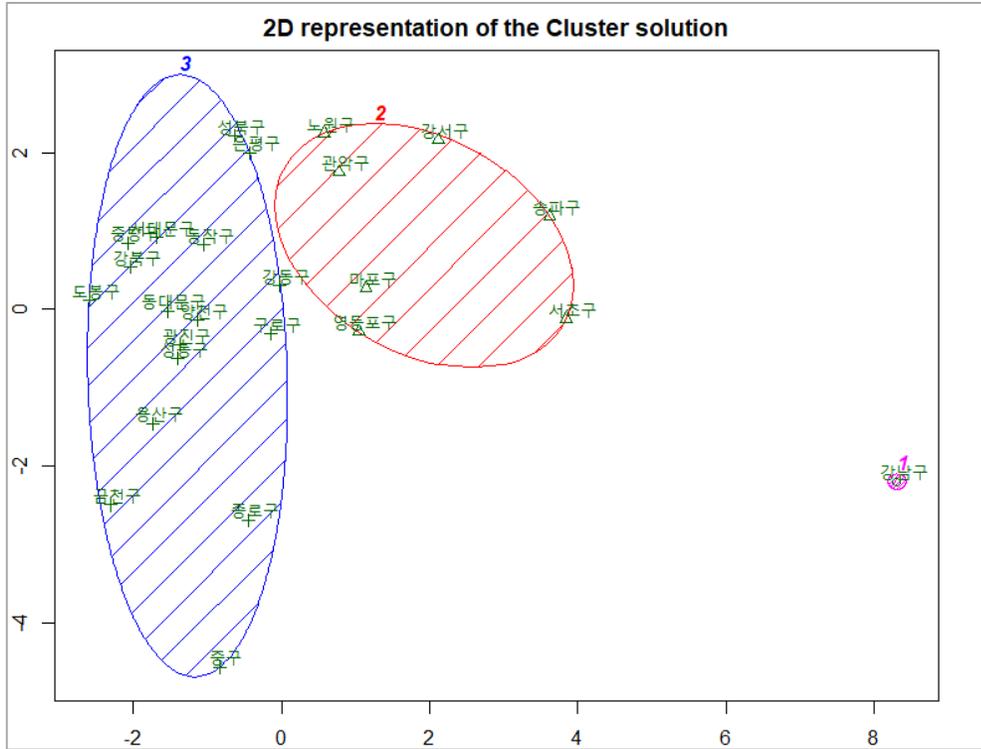


complete





K-Means





군집 내 상권분석

군집 2에 해당하는 주요 상권 데이터

- 상권별 커피점 평균 매출, 버스정류장 승차인구, 쓰레기통 위치 등

marketid	name	goo	area	sales	busnum	buspeople_up	xcenter	ycenter	trash_p
2183	서울 강서구 발산역	강서구	69366	3298.5	14	2573.88290	126.8384	37.55828	0.70000000
2192	서울 강서구 송정역	강서구	59494	1349.0	13	860.00443	126.8077	37.55924	0.81818182
2158	강서구청	강서구	114547.5	2885.0	8	731.11365	126.8470	37.54965	0.55555556
1846	서울 관악구 신림역_3	관악구	36813	4452.0	18	10343.46476	126.9270	37.48403	0.00000000
1823	서울 관악구 신림역_2	관악구	33339.2	5970.5	18	7584.05152	126.9298	37.48218	0.00000000
1819	서울 관악구 신림역_1	관악구	48332.9	4623.0	15	4061.69472	126.9267	37.48081	0.00000000
1812	서울 관악구 서울대입구역_2	관악구	41632.5	1715.5	16	7407.33961	126.9534	37.48119	0.04166667

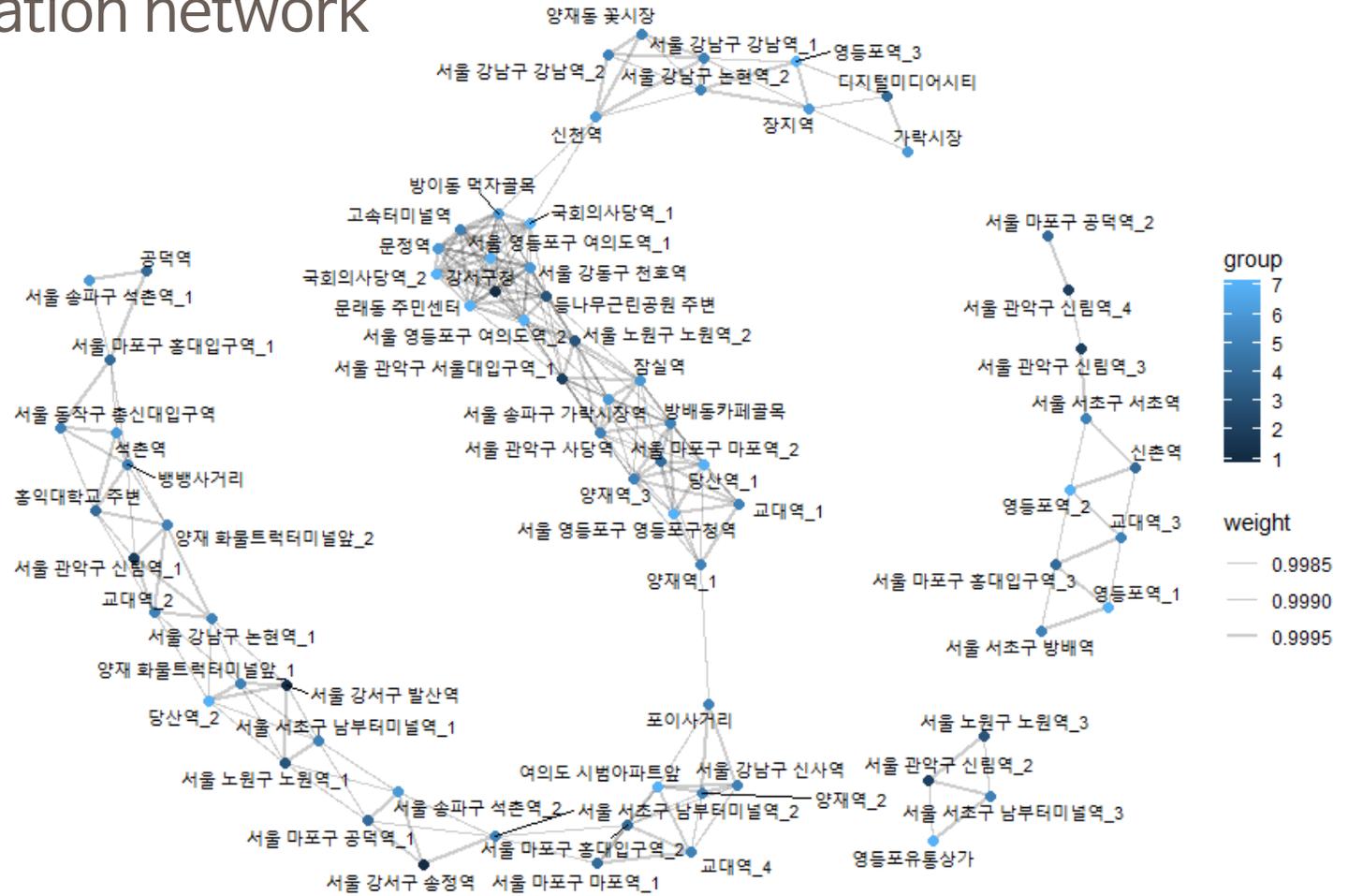
주요상권의 예) 강남역



출처 : 소상공인 상권정보시스템

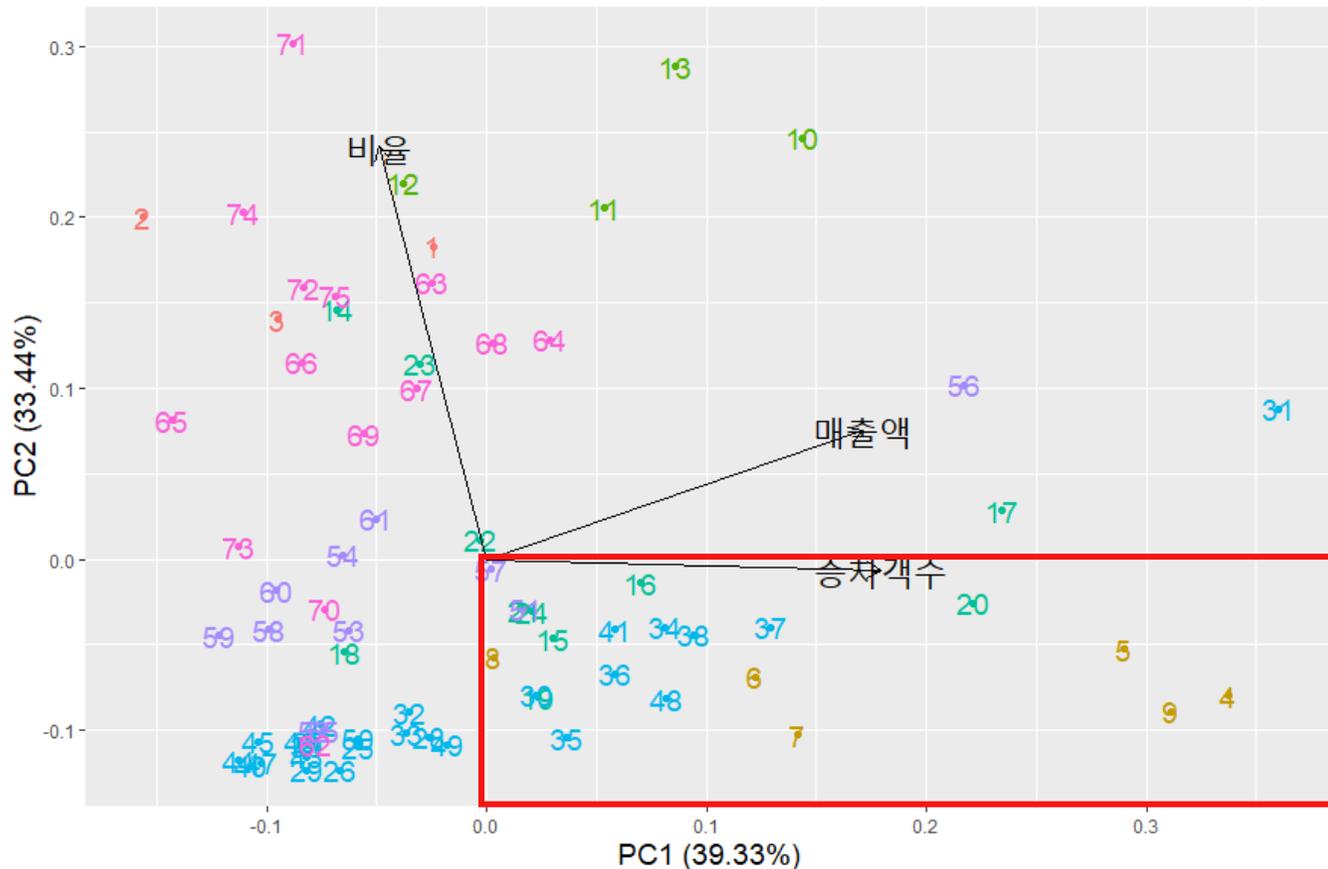


■ 군집 내 상관분석 - correlation network





군집 내 상관분석 - PCA



자치구

- a 강서구
- a 관악구
- a 노원구
- a 마포구
- a 서초구
- a 송파구
- a 영등포구

```
> pca_model
```

```
Standard deviations (1, .., p=3)
[1] 1.0862695 1.0015741 0.903807
```

```
Rotation (n x k) = (3 x 3):
```

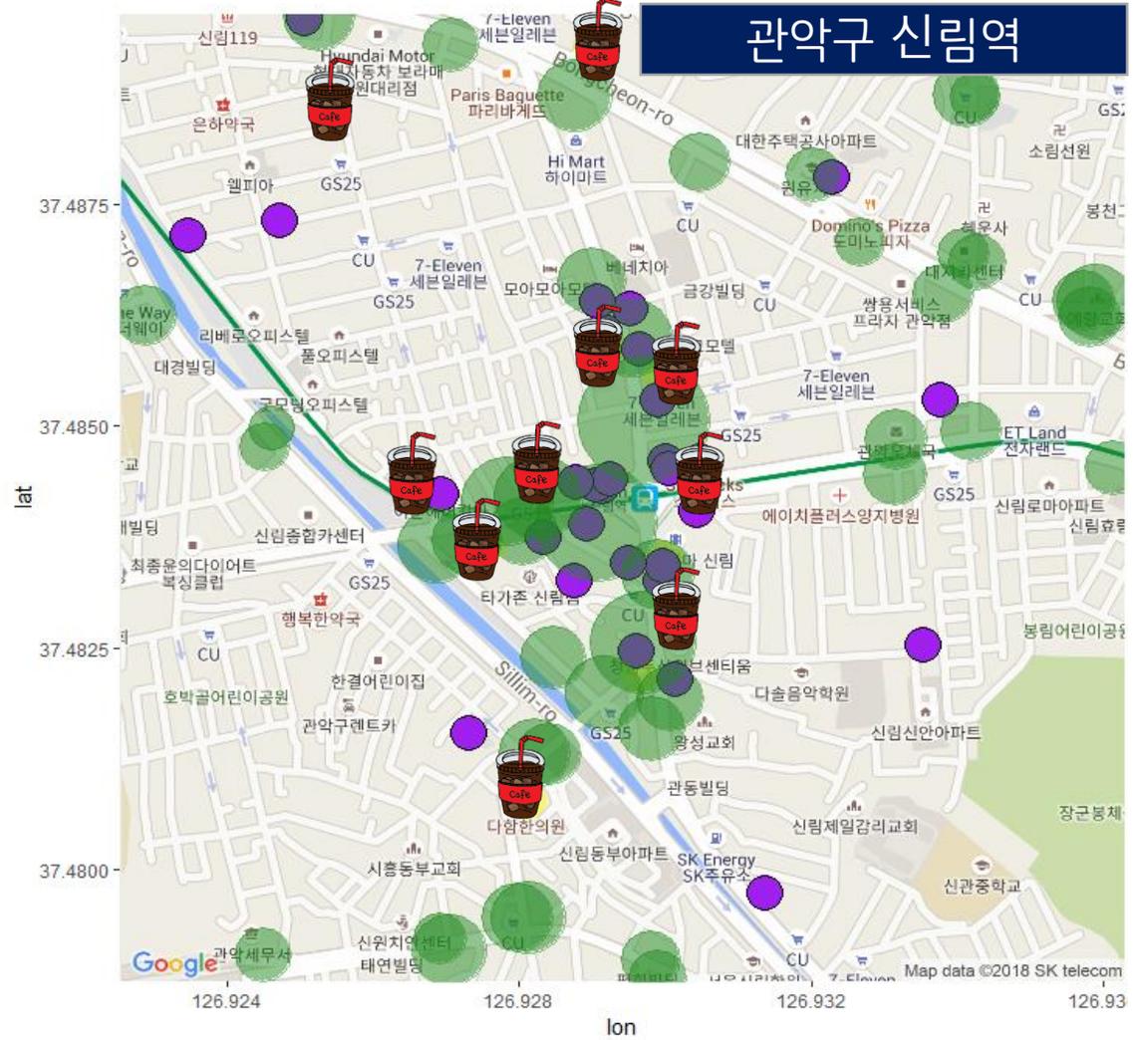
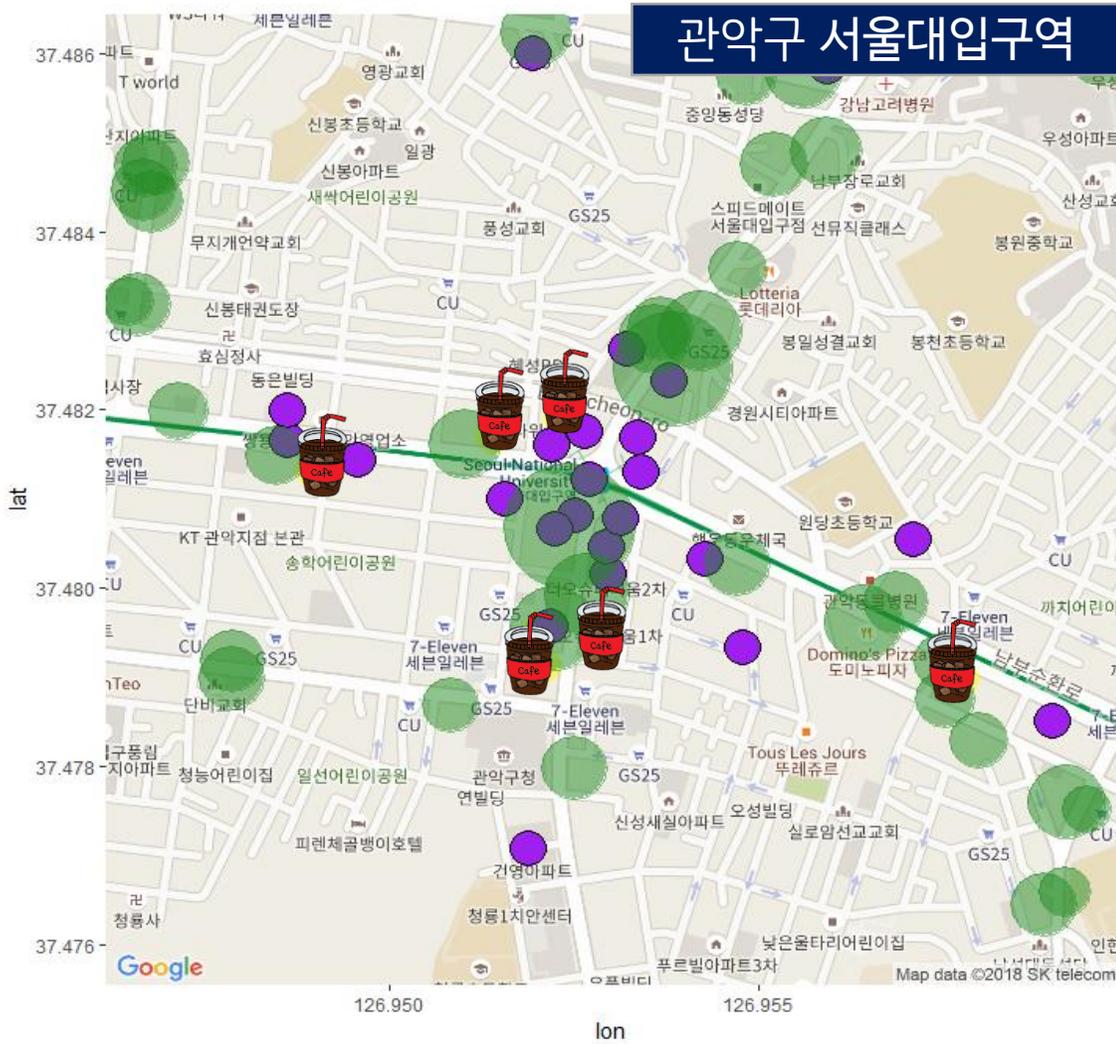
	PC1	PC2
매출액	<u>0.6771040</u>	0.29901460
승차객수	<u>0.7099430</u>	-0.02494611
비율	-0.1936779	<u>0.95392241</u>

```
> summary(pca_model)
```

```
Importance of components:
```

	PC1	PC2
Standard deviation	1.0863	1.0016
Proportion of Variance	0.3933	0.3344
Cumulative Proportion	0.3933	0.7277

분석 과정





서초구 양재역



서초구 교대역

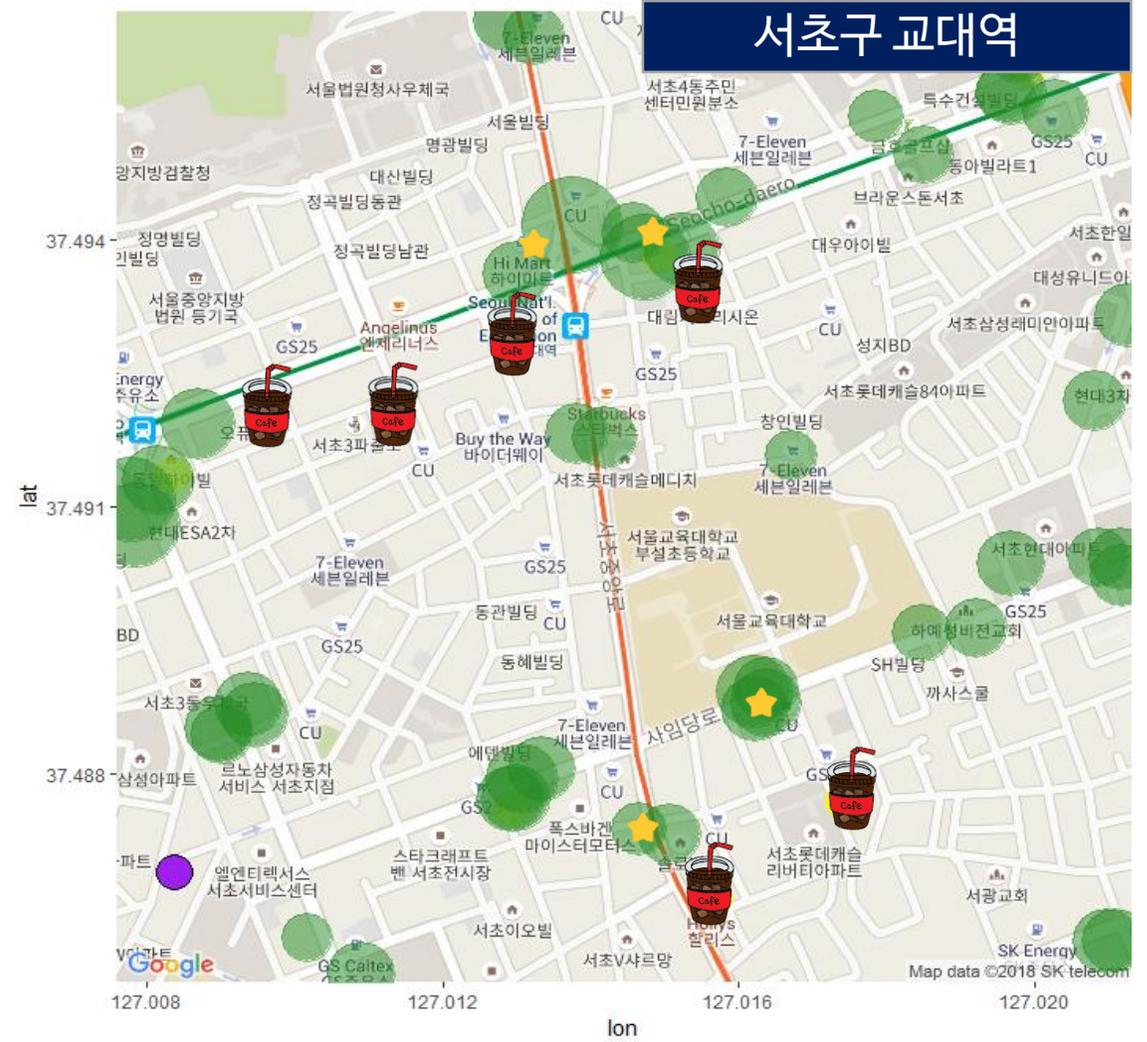




서초구 양재역



서초구 교대역





□ 설치 현황

- '95년 종량제 실시 이후 환경미화원 감소, 청소·관리인력 부족 등의 이유로 가로휴지통 수가 '95년 7,607개에서 '07년 3,707개로 감소
- 휴지통 축소에 따른 시민불편 민원 증가 등으로 서울시가 증설을 적극 권장하면서 확대 하여 '16년말 기준 5,640개 설치
⇒ '95년 7,607개(100%) → '07년 3,707개(48.7%) → '16년 5,640개(74.1%)

[동네방네]“커피컵 서리플컵에 버리고 버스타세요”

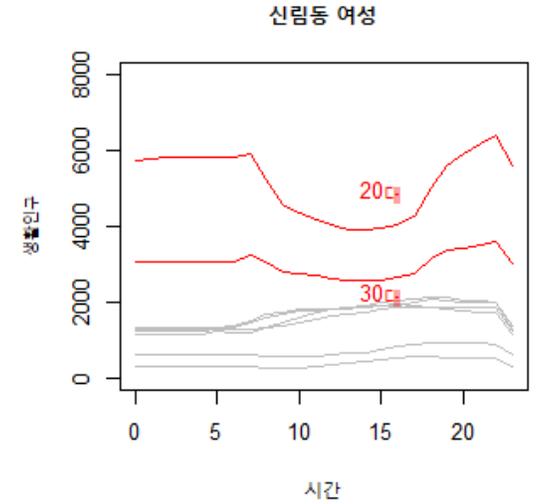
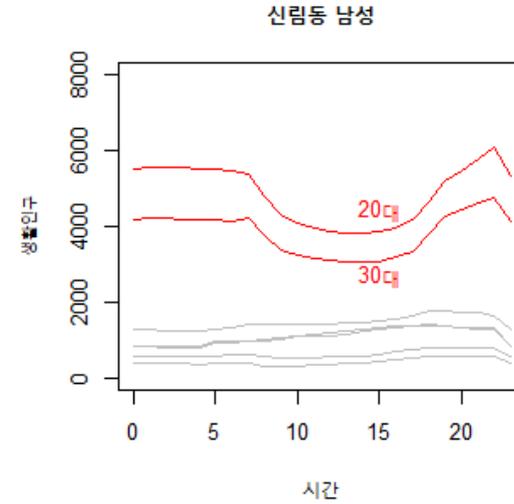
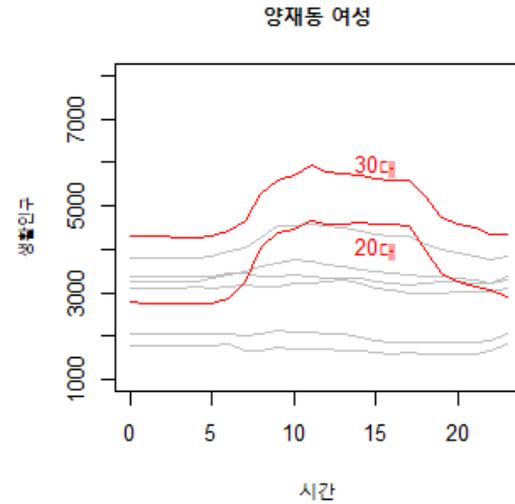
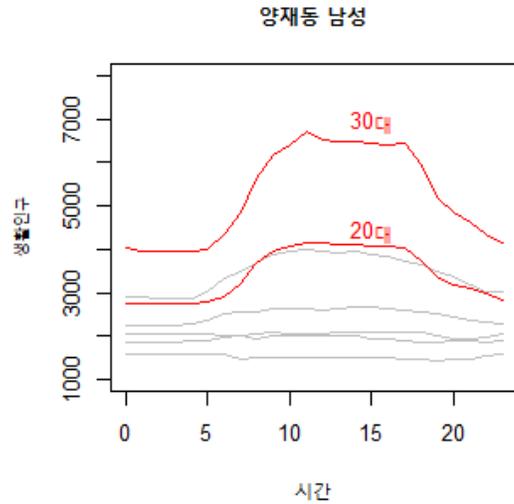
서울 서초구 버스정류장 지하철역 114개 배치

등록 2018-03-21 오전 11:39:58
수정 2018-03-21 오전 11:39:58

가 가



쓰레기통 위치 선정



시간별 생활인구 데이터를 이용
커피점 주 고객층인 20~30대의 패턴을 참고할 수 있음

감사합니다 😊