

연관규칙분석을 활용한

서울특별시 청소년쉼터

추가 입지 분석

회귀분석팀

김민우

채희지

김다민

성준혁

천예원

Index

01. 주제 선정 배경

02. 연관규칙분석

03. 가정폭력위험도 모델링

04. 클러스터링

05. 청소년 심터 평가 지표

06. 최종 입지 선정 및 결론



01. 주제 선정 배경

청소년 쉼터의 법적 정의

①가출청소년의(대상) ②일시적인(보호기간) ③생활지원과 보호(서비스), ④가정. 사회로의 복귀(단기목표),
⑤중장기적으로 자립을 지원하기 위한(중장기목표) ⑥청소년복지시설(청소년시설 유형)"을 의미함.

출처: 청소년쉼터 운영지침, 보건복지부



정책적 측면에서, 가정의 책임을 대신하여 **사회적 돌봄이 필요한 위기청소년**을
광의의 정책대상으로 정의



01. 주제 선정 배경

청소년 쉼터 현황

서울 자치구 별 거주 가능 쉼터 현황 (출처: 한국청소년상담복지개발원)

0개	강남구, 광진구, 구로구, 노원구, 도봉구, 동대문구, 동작구, 마포구, 서대문구, 서초구, 성동구, 성북구, 송파구, 영등포구, 종로구, 중구
1개	강동구, 강북구, 양천구, 용산구, 중랑구
2개	강서구, 관악구, 금천구, 은평구

가정 밖 청소년들은 자신이 다니던 학교에서 멀리 떨어진 곳까지 쉼터를 찾아가고 있다. 서울에 있는 학교에 다녔지만 청소년 쉼터가 없어 충남 천안까지 갔다는 가정 밖 청소년도 있었다. 부모님 동의 없이 전학 갈 수 없으니 학교에 아예 가지 않게 된다. 과거 가정 밖 청소년이었던 B씨(22)는 “가정 밖 청소년들은 쉼터를 찾기 시작하면서 학업 포기를 당연한 수순으로 여긴다”며 “검정고시의 기회가 있지만 당장 먹고살 돈을 생각해야 하는 가정 밖 청소년들은 학업에 신경 쓸 수 없다. 이들은 제대로 된 교육을 받지 못해 빈곤의 악순환에 빠지게 된다”고 지적했다.

출처: [길잃은 청소년@]가정폭력 피해 집 나왔지만...서울 16개 구엔 쉼터도 없어, 아시아경제

청소년 쉼터의 개수 부족으로 지역을 옮겨 찾아가는 청소년들이 있는 상황

01. 주제 선정 배경

청소년 가출 경험 및 이유

30. 가출 경험 및 이유

2021년 초(4~6)·중·고등학생 중 3.2%가 최근 1년 내 가출한 적이 있으며, 부모님과의 문제(62.4%)가 주된 가출 이유임

- '21년 초·중·고등학생의 최근 1년 내 가출 경험률은 3.2%로 전년 대비 0.3%p 증가함
 - 여학생(2.9%)이 남학생(3.4%)보다, 학교급이 높을수록 가출 경험이 적었음
- 가출 이유로는 부모님과의 문제(62.4%)가 가장 많았고, 다음으로 학업 문제(18.5%), 친구들과 함께하기 위해서(9.7%) 순임

< 가출 경험 >



자료: 한국청소년정책연구원, 「아동청소년인권실태조사」

< 가출 이유(2021) >



< 가출 경험 및 이유 >

(단위: %)

	가출 경험 ¹⁾	부모님과 의 문제	가출 이유 ²⁾				
			학업문제	친구들과함께 하기 위해서	가정의 경제 적 문제	학교문제	기타
2014	4.8	57.4	12.0	16.2	0.2	2.7	11.6
2015	3.5	63.5	12.8	12.4	1.4	3.6	6.2
2016	3.4	60.7	12.4	12.8	0.8	2.7	10.6
2017	3.7	60.9	14.3	10.6	1.7	5.1	7.4
2018	3.8	58.0	15.6	12.0	4.1	3.0	7.2
2019	3.5	61.7	15.9	9.6	2.3	1.1	9.5
2020	2.9	61.0	20.8	8.0	2.3	2.0	5.9
2021	3.2	62.4	18.5	9.7	1.4	0.6	7.4
남자	3.4	59.0	18.7	11.8	1.9	0.0	8.6
여자	2.9	66.6	18.2	7.3	0.8	1.3	5.9
초(4~6)	3.8	54.6	20.0	8.7	2.1	0.0	14.6
중	3.4	68.2	15.3	13.4	1.2	0.0	1.9
고	2.4	65.7	20.5	6.4	0.6	2.1	4.7

자료: 한국청소년정책연구원, 「아동청소년인권실태조사」, 초(4~6)·중·고등학생 대상

주: 1) 최근 1년 동안 가출한 적이 있는 사람의 비중

2) 최근 1년 동안 가출한 적이 있는 사람의 주된 가출 이유

출처: 보도자료_여성가족부(220525) 청소년 통계

최근 1년 안에 가출 경험이 있는 청소년 비율이 3.2%에 달하는 것으로 나타남
가출 이유로는 '부모님과의 문제'가 차지하는 비율이 가장 높았음

01. 주제 선정 배경

가정폭력의 영향

가정 밖 청소년 대부분은 가정폭력을 당하다가 집 밖을 나선다. 갈 곳이 없는 경우 청소년 쉼터의 빈자리를 찾기 시작한다. 하지만 턱없이 부족하고 불균형하게 배치된 청소년 쉼터로 인해 가정 밖 청소년들은 학업도 포기하고 먼 길을 떠나야 했다.

13일 여성가족부에 따르면 지난해 가출한 청소년들 가운데 69.5%는 ‘가족과의 갈등’, 28.0%는 ‘가족의 폭력을 피하기 위해’를 가출의 원인으로 꼽았다. 지난해 청소년 쉼터 및 자립지원관을 이용한 청소년 중 72.1%는 가정에서 신체폭력을, 72.9%는 언어폭력을 당했다고 답하는 등 계속해서 가정폭력에 노출된 것으로 분석됐다.

가정 밖 청소년인 A씨(17)는 “학교 성적을 두고 부모님과의 갈등이 심했다. 이 과정에서 심각한 언어폭력에 지속적으로 노출됐다”며 “부모님과 함께 있으면 좋지 못한 생각이 계속 들어 괴로워 살 수 없을 것 같았다”고 말했다.



청소년 쉼터를 찾는 가출 청소년들은 가정폭력에 지속적으로 노출되는 경우가 많음

따라서 가정폭력 위험도 지수를 만들어 청소년 쉼터 추가 입지 분석에 활용할 예정!

01. 주제 선정 배경

최종 주제 선정

■ 청소년의 건강한 성장 발달 지원

- 위기청소년의 건강한 성장을 위한 정서적 지원 및 보호 강화
 - 자체지원을 통한 상담복지센터 상담 인력 증원(45명), 1388 전화상담원 증원(2명)
 - 청소년쉼터별 중·장기 운영계획 수립을 통한 쉼터 운영성과 제고

출처: 2023년 서울시 주요 업무계획

2023 서울시 주요 업무계획에서 청소년 쉼터의 중·장기 운영 계획 수립에 대한 내용 확인



청소년 쉼터의 추가적인 입지 고려!

01. 주제 선정 배경



최종 주제 선정

청소년 쉼터의 세부 유형 관련 논의

청소년의 건강한 성장 발달 지원

연구보고 20-R20

가출청소년 지원 강화를 위한
청소년복지시설 재구조화 연구

책임연구원 황진구
공동연구원 김지연

NYPI

한국청소년정책연구원
National Youth Policy Institute

- 1) 일시쉼터 (일시-7일)
- 일시이동형 쉼터
- 일시고정형 쉼터(7일)
- 2) 단기쉼터 (3개월)
- 3) 중장기쉼터 (3년)
- 4) 자립지원관 (1년)



연번	시도	시군구	시설종류	시설명
1	서울	종로구	일시쉼터(이동형)	서울시립청소년이동쉼터(서북권)
2	서울	종로구	일시쉼터(이동형)	서울시립청소년이동쉼터(서남권)
3	서울	강동구	일시쉼터(이동형)	서울시립청소년이동쉼터(동북권)
4	서울	강동구	일시쉼터(이동형)	서울시립청소년이동쉼터(동남권)
5	서울	용산구	일시쉼터(고정형)	서울특별시청소년일시쉼터(용산일시)
6	서울	강북구	일시쉼터(고정형)	서울시립강북청소년드림센터
7	서울	양천구	일시쉼터(고정형)	서울특별시립드림일시청소년쉼터
8	서울	은평구	일시쉼터(고정형)	은평구립일시청소년쉼터
9	서울	강서구	단기쉼터(남자)	강서청소년쉼터
10	서울	관악구	단기쉼터(남자)	서울시립신림청소년단기쉼터
11	서울	강동구	단기쉼터(여자)	강동여자단기청소년쉼터
12	서울	금천구	단기쉼터(여자)	서울시립금천청소년단기쉼터
13	서울	중랑구	단기쉼터(여자)	서울시립망우청소년단기쉼터
14	서울	관악구	중장기쉼터(남자)	서울시립신림청소년중장기쉼터
15	서울	금천구	중장기쉼터(여자)	서울시립금천청소년중장기쉼터
16	서울	은평구	중장기쉼터(여자)	서울시립은평여자중장기청소년쉼터
17	서울	강서구	중장기쉼터(여자)	여울림청소년쉼터
18	서울	도봉구	운영	서울시 청소년자립지원관
19	서울	관악구	-	관악 들꽃 청소년자립지원관

청소년 쉼터의 추가적인 입지 고려!

보다 정확한 입지선정을 위해 유형별 특징을 파악!

01. 주제 선정 배경



최종 주제 선정

현행 - 이용 기간 중심	
1) 일시 쉼터	1-1) 일시 이동형 쉼터 (일시~7일)
	1-2) 일시 고정형 쉼터 (7일~)
2) 단기 쉼터 (3개월)	
3) 중장기 쉼터 (3년)	
4) 자립 지원관 (1년)	

지원
위한 정서적
상담 인력 증
계획 수련을 통
년 쉼터의 중.



개편안 - 기능 중심	
1) 아웃리치 센터 (일시)	
2) 청소년 쉼터	2-1) 일시형 쉼터 (3일)
	2-2) 일반형 청소년 쉼터 (3년)
3) 자립 지원관 (1년)	

※ 아웃리치: 거리의 청소년들을 위한 길거리 상담 및 지원 프로그램 홍보 사업

기존의 유형 기준을 **기능 중심**으로 제안한 한국청소년정책연구원의 보고서에 따라
청소년 쉼터의 기능을 강화하는 방안을 강구할 필요가 있음

01. 주제 선정 배경



최종 주제 선정

현행 - 이용 기간 중심	
1) 일시 쉼터	1-1) 일시 이동형 쉼터 (일시~7일)
	1-2) 일시 고정형 쉼터 (7일~)
2) 단기 쉼터 (3개월)	
3) 중장기 쉼터 (3년)	
4) 자립 지원관 (1년)	

지원
위한 정서적
상담 인력 증
계획 수립을 통
년 쉼터의 중.



개편안 - 기능 중심	
1) 아웃리치 센터 (일시)	
2) 청소년 쉼터	2-1) 일시형 쉼터 (3일)
	2-2) 일반형 청소년 쉼터 (3년)
3) 자립 지원관 (1년)	



청소년들이 즉시 보호받을 수 있다고 판단한
시설 중심으로 분석 진행!

01. 주제선정배경

활용 변수 소개



인구 요인

- 총 청소년 인구 수
- 혼인 및 이혼 건수
- 장애인 인구 수
- 재혼 인구 비율



경제 요인

- 국민기초생활수급 가구 수
- 국민기초생활수급 인구 수
- 인당 지출 총 금액



치안 요인

- 치안기관 개수
- 설치된 CCTV 개수
- 5대 범죄 발생 건수
- 가정폭력 발생 건수



기타 요인

- 고위험 음주인구 비율
- 가사노동 편중 비율
- 학원 수

02. 연관규칙분석

연관규칙분석이란?

연관규칙분석 (Association Rule Analysis)

항목(item)들 사이의 관계를 If-Then 방식으로 탐색함으로써 유의미한 인사이트를 도출하는 데 사용되는 분석
자료들 간의 규칙을 찾는 방식으로, 비지도학습의 일종으로 볼 수 있음



소비자의 장바구니 속 품목들의 관계를 알아본다는 의미에서 '장바구니 분석' 이라고 부르기도 함

02. 연관규칙분석

품

APRIORI 알고리즘

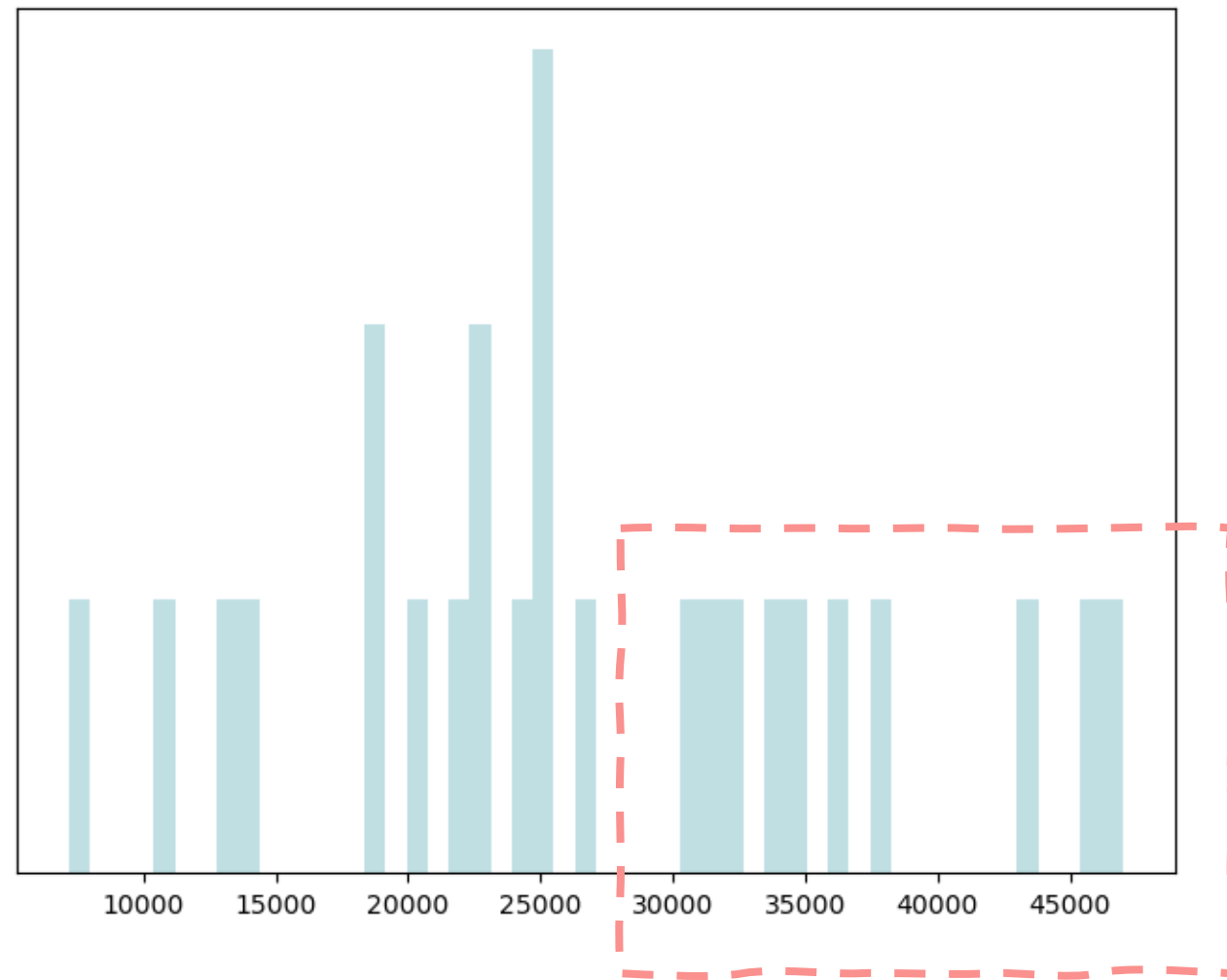
	lhs	rhs	support	confidence	lift	count
[1]	{}	=> {whole milk}	0.25551601	0.2555160	1.000000	2513
[2]	{hard cheese}	=> {whole milk}	0.01006609	0.4107884	1.607682	99
[3]	{butter milk}	=> {other vegetables}	0.01037112	0.3709091	1.916916	102
[4]	{butter milk}	=> {whole milk}	0.01159126	0.4145455	1.622385	114
[5]	{ham}	=> {whole milk}	0.01148958	0.4414062	1.727509	113
[6]	{sliced cheese}	=> {whole milk}	0.01077783	0.4398340	1.721356	106
[7]	{oil}	=> {whole milk}	0.01128622	0.4021739	1.573968	111
[8]	{onions}	=> {other vegetables}	0.01423488	0.4590164	2.372268	140
[9]	{onions}	=> {whole milk}	0.01209964	0.3901639	1.526965	119
[10]	{berries}	=> {yogurt}	0.01057448	0.3180428	2.279848	104



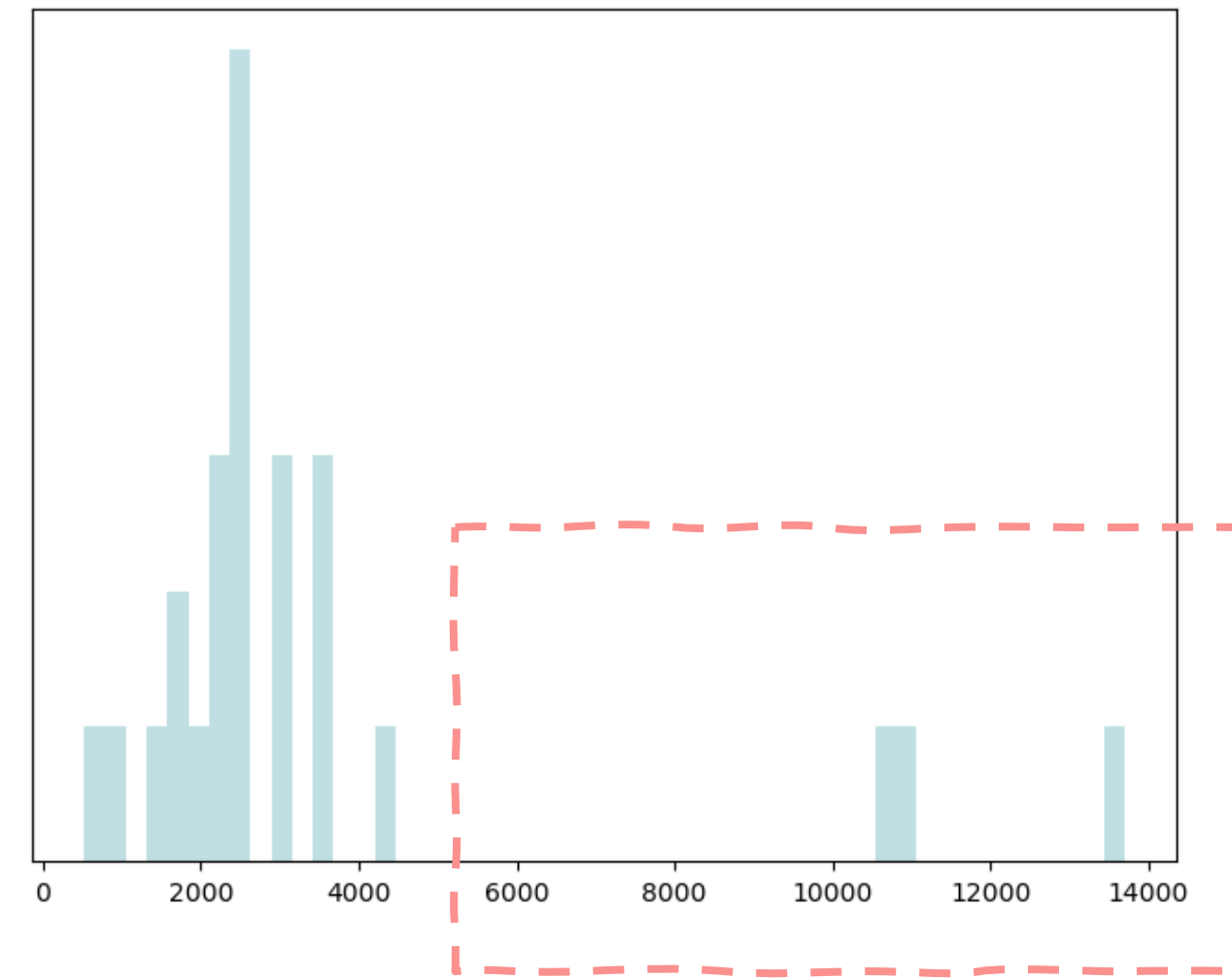
주어진 데이터에 대해 모든 연관규칙 쌍을 생성하고, 각 연관규칙에 대한 측도를 계산하는 알고리즘
연관규칙의 평가 측도로 지지도(Support), 신뢰도(Confidence), 향상도(Lift) 등을 제공함

02. 연관규칙분석

분석 과정 | ① 수치형 독립 변수들의 범주화



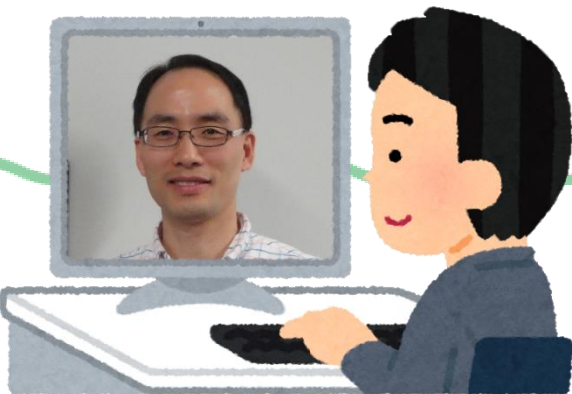
꼬리가 두꺼운 Plot의 데이터는 Percentile을 이용하여
상위 33% 66% 지점을 기준으로 3등분



꼬리가 두껍지 않은 Plot의 데이터는
중앙값을 이용하여 2등분

02. 연관규칙분석

분석 과정 | ② 범주형 변수들의 구매행렬화



행정구	이혼 건수	학원 수	...	치안기관 수
종로구	이혼건수_low	학원수_low	...	치안기관수_high
중구	이혼건수_low	학원수_low	...	치안기관수_high
용산구	이혼건수_low	학원수_low	...	치안기관수_low
...
강동구	이혼건수_high	학원수_high	...	치안기관수_high

▲ 행정구 단위로 범주화한 모델링 데이터



	이혼 건수	학원 수	...	치안기관 수
이혼 건수	23	1	...	4
학원 수	5	7	...	2
...
치안기관 수	7	2	...	9

▲ 구매행렬로 변환

(R Studio에서의 transaction 자료형)

02. 연관규칙분석



분석 과정 | ③ 연관 규칙 분석 결과 확인

Rule		Lift
청소년남자인구_low, 청소년여자인구_low	혼인건수_low	3.125

← - - - - - →
상 호 연 관

Rule		Lift
인당지출총금액_high, 카메라대수_low	불편신고수_high	2.381

← - - - - - →
상 호 연 관

청소년 남/여 인구 수를 이용해
혼인건수라는 변수를 예측해볼 수 있음

CCTV 개수가 적을수록
생활안전/교통안전/사회안전 등
불편신고 요인에 영향을 끼치는 문제가
많이 발생함을 확인

02. 연관규칙분석



분석 과정 | ③ 연관 규칙 분석 결과 확인

Rule		Lift
학원수_high, 혼인건수_high	이혼건수_high	2.778

← - - - - - →
상 호 연 관

학원이라는 변수에서 파생 가능한
학업과 관련된 문제가
이혼에 영향을 끼칠 수 있다는 인사이트

Rule		Lift
장애인수_high, 총수급자가구_high	이혼건수_high	2.778

← - - - - - →
상 호 연 관

가정 내 장애인 구성원의 여부,
가정의 경제적 상황이 이혼에 영향을
끼치는 요소임을 추측해볼 수 있음

역으로 이혼이 경제적 상황에 영향은 끼친 것은 아닌지 고민해볼 수 있음!

03. 가정폭력 위험도 지표 모델링

사용된 변수 소개

모델링에 사용할 독립변수

청소년 남자 인구
청소년 여자 인구

총 청소년 가구
장애인 수

총 수급자 가구
총 수급자 인원

불편신고수
인당지출총금액

혼인 건수
이혼 건수

카메라 대수
고위험 음주율

재혼율
가사노동 편중된
가정 비율

치안기관수
범죄총발생건수

모델의 종속변수

가정폭력
검거건수

03. 가정폭력 위험도 지표 모델링

상관관계 확인 및 사전 변수 필터링

변수		상관계수
청소년 남자인구	청소년 여자인구	0.73

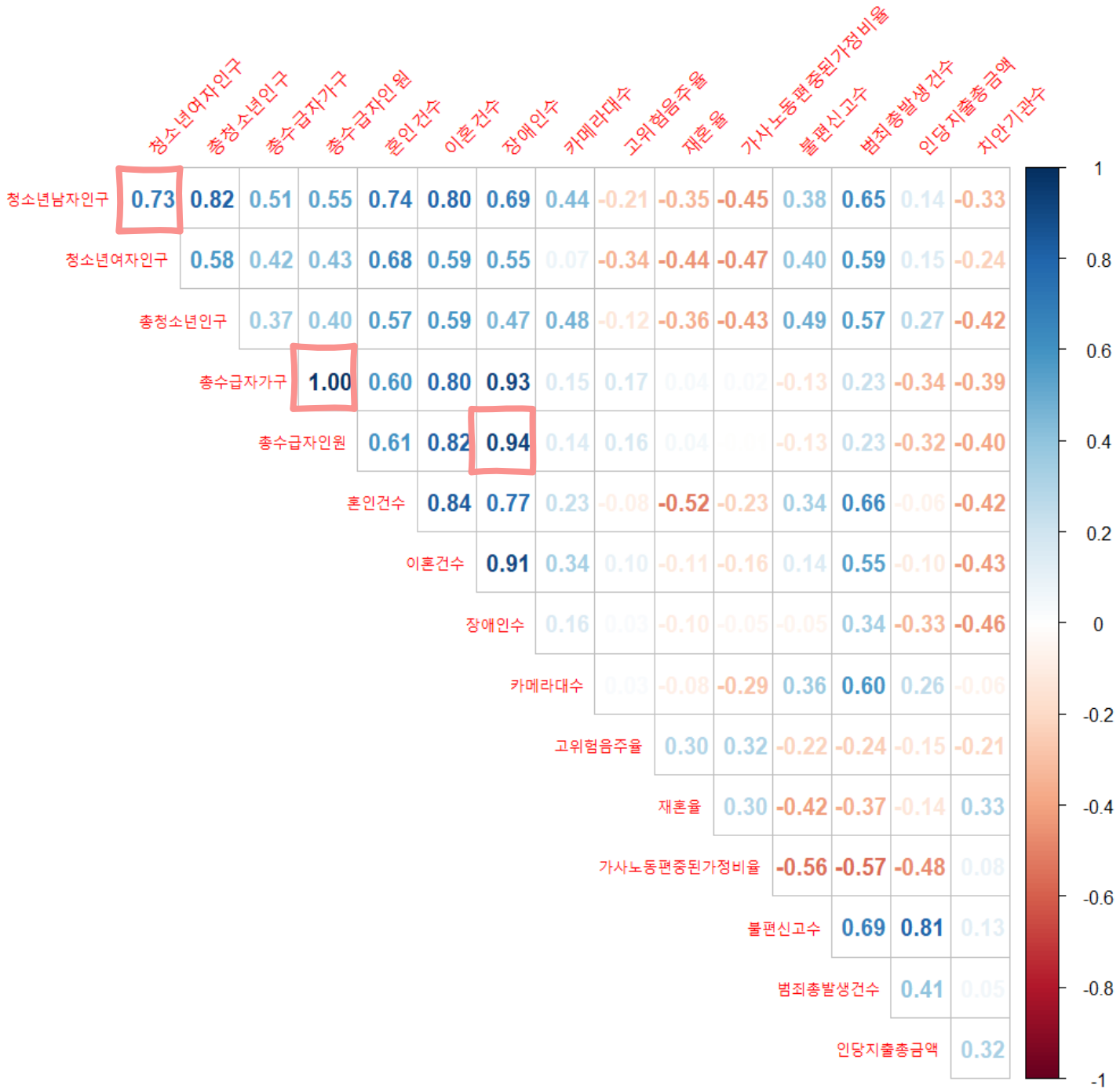
청소년 여성인구와 남성인구 간의 상관관계가 높아 총 청소년 인구만 분석에 활용

변수		상관계수
총 수급자 가구	총 수급자 인원	1.00

총 수급자 가구와 인원 간의 상관계수가 1에 가까워 총 수급자 인원을 선택

변수		상관계수
총 수급자 인원	장애인 수	0.94

가정폭력과의 직관적인 관련성과 해석력을 고려하여, 총 수급자 인원을 선택



03. 가정폭력 위험도 지표 모델링

다중공선성 제거

VIF (Variance Inflation Factor, 분산팽창인자)

특정 독립변수가 다른 독립변수들로 설명이 되는 정도를 의미하며, 일반적으로 VIF가 10 이상일 경우 심각한 다중공선성으로 진단. 반면, 다중공선성이 전혀 존재하지 않는다면 VIF 값은 1이 나옴.

변수	VIF
이혼건수	21.997952
혼인건수	20.152539
...	...

사용할 모든 변수를 넣은 선형 모델에 대해 VIF 확인

⋮

가장 높은 VIF값을 가진 이혼건수 삭제



변수	VIF
범죄 총발생건수	9.865861
혼인건수	8.348420
...	...

이혼 건수를 제거한 선형 모델에서 VIF 확인

⋮

VIF가 가장 높은 범죄 총발생건수는 VIF가 9.866으로 10을 넘기지는 못했지만 상당히 크다고 판단하여 삭제

03. 가정폭력 위험도 지표 모델링

변수선택법 : Best Subset Selection

Best Subset Selection



가능한 변수들의 모든 조합을 고려하는 방법. 변수의 개수가 p 개라면, 2^p 개의 모형을 모두 적합하고 비교.

선택된 Best model에 대한 신뢰도가 높으나 계산 비용이 많이 소요되는 것이 단점

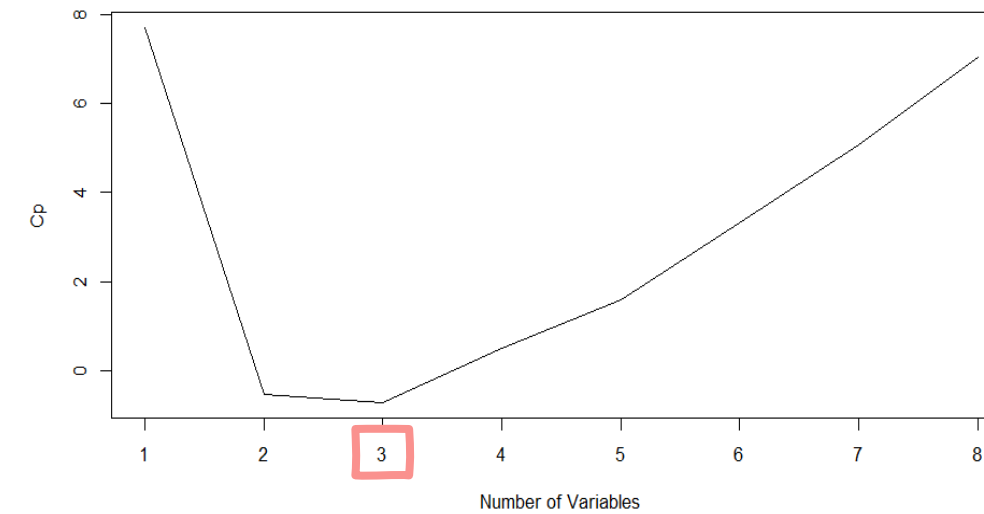
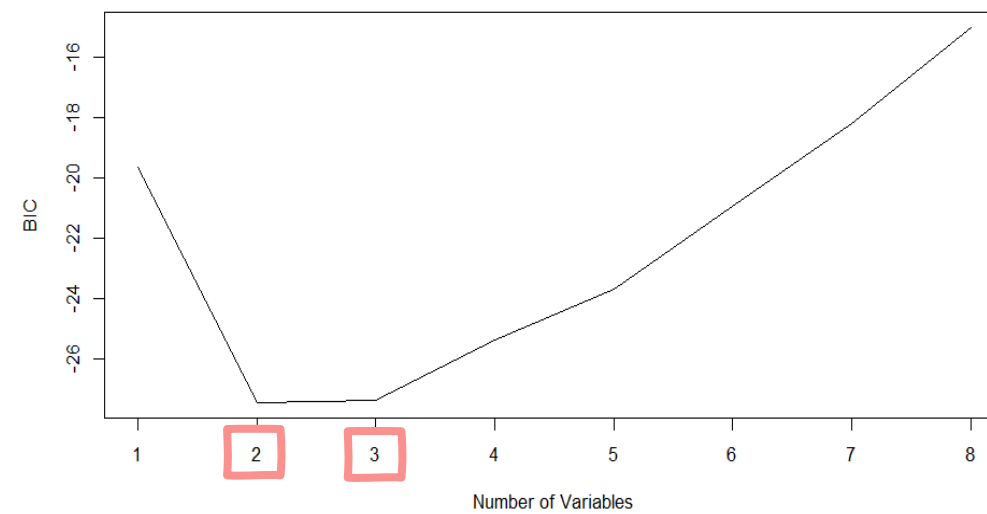
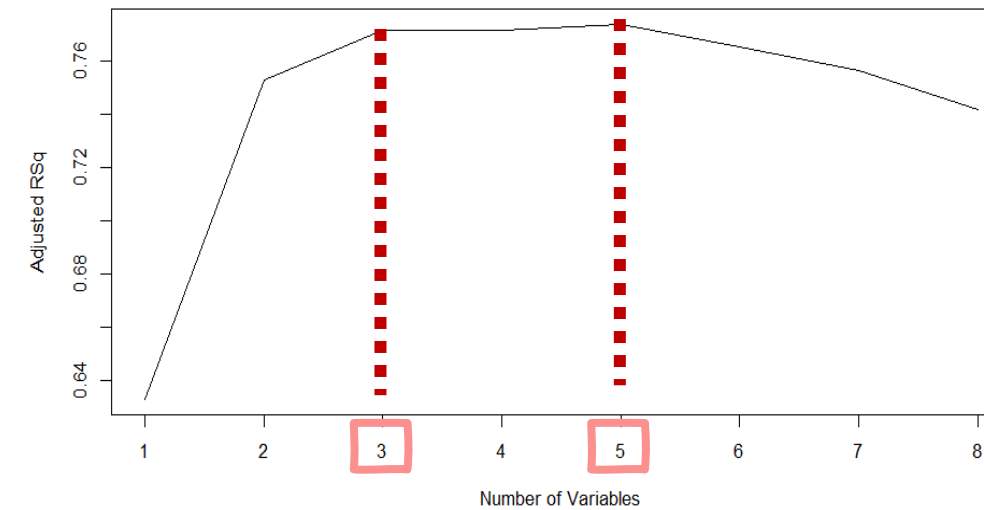
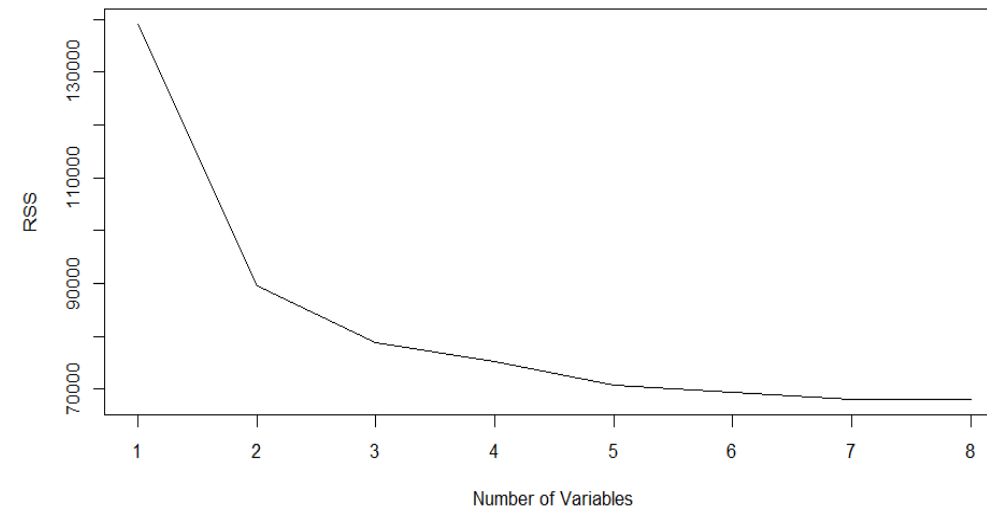
회귀분석팀 클린업 3주차 참고!



"Best Subset selection이 현재 상황에 적절한 이유는 "
"설정된 모델이 선형 모델이므로 계산 시간이 적게 걸리기 때문!"

03. 가정폭력 위험도 지표 모델링

모델 선정 및 기본가정 검정



regsubsets 함수에 사용할 변수를 모두 넣어 얻은 plot (왼쪽 위부터 시계방향으로 RSS, R_{adj}^2, Cp, BIC)

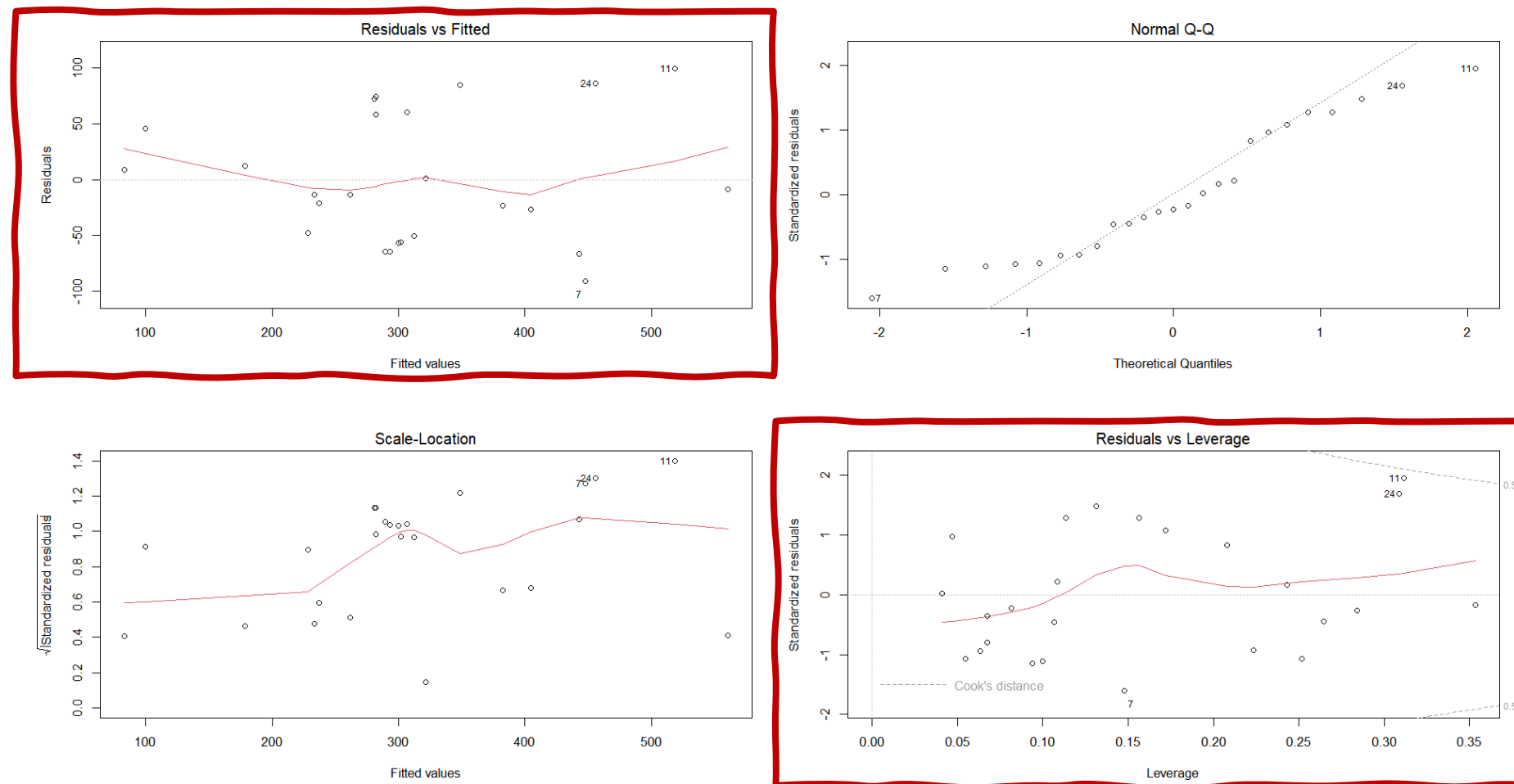
위 4가지 지표를 모두 고려했을 때 변수가 3개인 모델의 점수가 가장 좋다고 할 수 있음

03. 가정폭력 위험도 지표 모델링

모델 선정 및 기본가정 검정

최종 모델 :

$$\text{가정폭력검거건수} = -0.285 + 0.00073 * \text{총청소년인구} + 0.0094 * \text{총수급자인원} + 0.083 * \text{혼인건수}$$



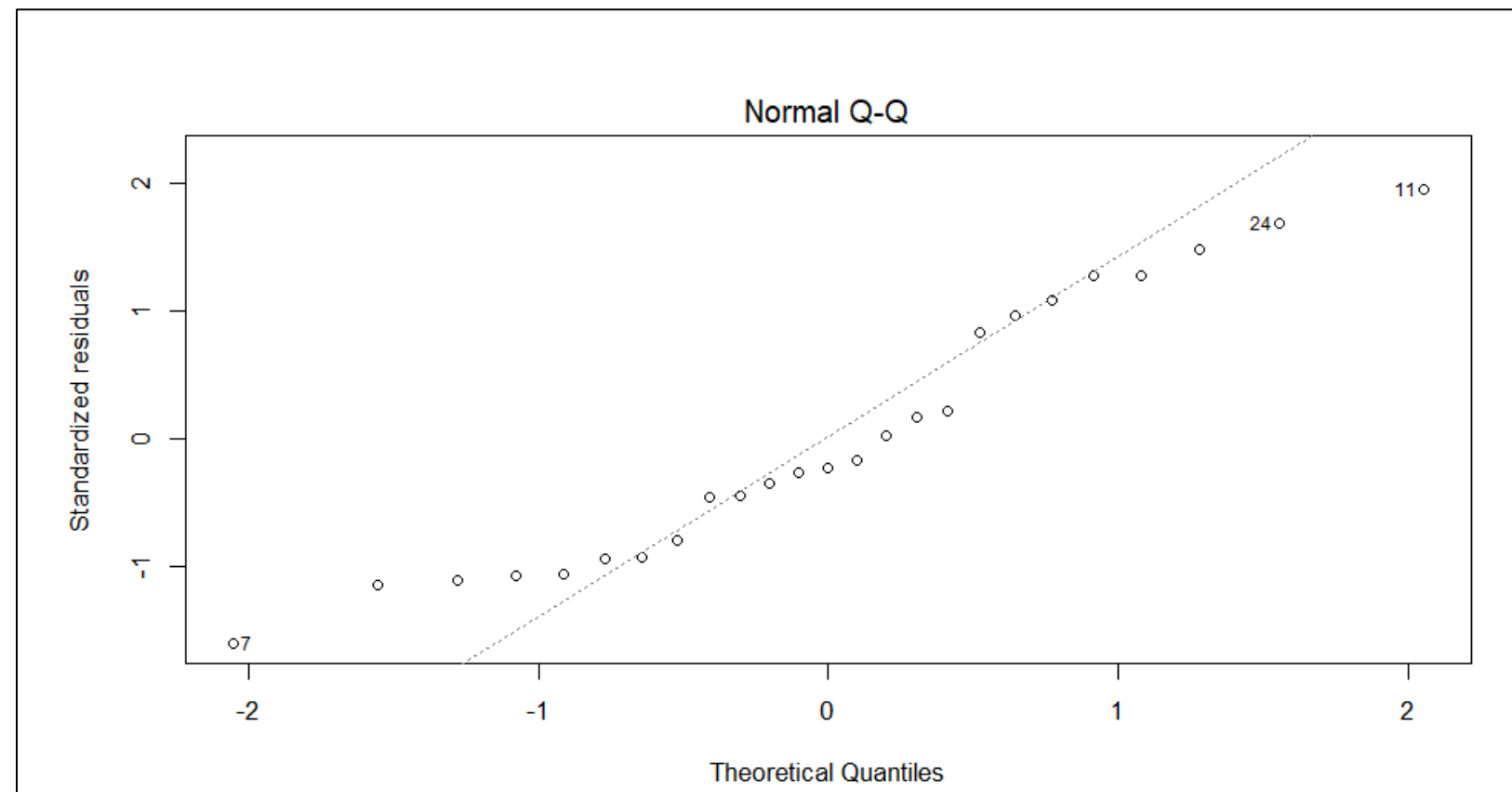
Residual vs fitted의 경우, 16개 변수 중 오직 3개의 변수만으로 25개의 관찰값만을 추정하는 것을 고려하면 패턴이 제거되었다고 볼 수 있으며, Cook's distance가 0.5를 넘어가는 값이 없어 outlier가 없다고 판단

03. 가정폭력 위험도 지표 모델링

모델 선정 및 기본가정 검정

최종 모델 :

$$\text{가정폭력검거건수} = -0.285 + 0.00073 * \text{총청소년인구} + 0.0094 * \text{총수급자인원} + 0.083 * \text{혼인건수}$$



하지만 Normal Q-Q plot의 경우 형태가 온전치 않은데,
통계적 유의성을 확인하기 위해 Shapiro test를 시행함

Shapiro-Wilk normality test

```
data: fit_final$residuals  
W = 0.93022, p-value = 0.08793
```



Shapiro-Wilk normality test에서
정규성을 만족함을 확인

03. 가정폭력 위험도 지표 모델링

공간회귀

공간회귀(Spatial Regression)

인접한 관찰값의 영향력을 공간가중행렬을 통해 고려하는 회귀모형
공간시차모형(SLM), 공간 오차모형(SEM) 등이 존재

회귀분석팀 2주차 PPT 참고 (열심히 만들었으니 봐주세요!)

공간가중행렬(Spatial Weight Matrix)

관찰값간의 관계(인접여부, 미치는 영향력의 크기)를 행렬 형태로 표현한 것
지역간의 상호작용 강도를 표현!

03. 가정폭력 위험도 지표 모델링

공간회귀모형 ①

공간시차모형 (*SLM, Spatial Lagged Model*) 회귀분석팀 클린업 2주차 PPT 참고!



한 지역의 관측치가 인접지역들의 관측치와 상관성이 있는 경우, 통계 모델에 공간적 의존성을 변수로 투입시키는 모형. 이 때, 공간 시차 변수를 하나의 설명변수로 회귀모델에 삽입

$$Y = \rho WY + X\beta + \varepsilon = (I - \rho)W^{-1}(X\beta + \varepsilon),$$

where $\varepsilon \sim MVN(0, \sigma^2 I_n)$

공간오차모형 (*SEM, Spatial Error Model*)

숨겨진 설명변수를 고려하지 못하여 오차가 공간 자기상관성을 가질 때,
오차를 공간 오차 변수로 변형시키는 모형

$$Y = X\beta + \mu = X\beta + (I - \lambda W)^{-1}\varepsilon,$$

where $\mu = \lambda W\mu + \varepsilon$ and $\varepsilon \sim MVN(0, \sigma^2 I_n)$

03. 가정폭력 위험도 지표 모델링

공간가중행렬

경계 기반 가중행렬 활용한 가중행렬만 기재함



관찰값끼리 경계를 공유할 경우 1, 그렇지 않으면 0을 가중치 값으로 주는 방법

회귀분석팀 클린업 2주차 PPT 참고!

거리기반 가중행렬 1) Radial Distance Weights

관찰값 사이의 거리가 d 보다 작을 경우($d_{ij} < d$) 1,
그렇지 않으면($d_{ij} \geq d$) 0을 가중치 값(w_{ij})으로 주는 거리 기반 가중행렬.

거리기반 가중행렬 2) Power distance weights

기준 관찰값과 다른 관찰값 사이 거리의 $-\alpha$ 제곱을 가중치 값으로 주는 거리 기반 가중행렬.

$$w_{ij} = d_{ij}^{-\alpha} \text{ where } \alpha > 0, \text{ 주로 } \alpha \text{는 } 1 \text{이나 } 2 \text{를 선택.}$$

03. 가정폭력 위험도 지표 모델링

공간회귀모형 검정

LM 검정 (*Lagrange Multiplier Test*)

OLS와 공간회귀모형 중 어떤 모형을 쓰는 것이 적절한지 확인하는 검정



$H_0: \rho = 0$, $H_1: \rho \neq 0$ 이므로 귀무가설을 기각하게 되면 공간회귀모형이 적절함을 의미
공간시차모형(SLM)과 공간오차모형(SEM)에 대해서 모두 수행!

03. 가정폭력 위험도 지표 모델링

공간회귀모형 검정

거리 기반 공간회귀모형: 1) Radial Distance Weights

공간 시차 모형

$$H_0: \rho = 0, p - value = 0.081$$

▶ 공간 시차 모형이 **적절하지 않음**

공간 오차 모형

$$H_0: \lambda = 0, p - value = 0.554$$

▶ 공간 오차 모형이 **적절하지 않음**

경계 기반 공간회귀모형: Rook Contiguity Weights

공간 시차 모형

$$H_0: \rho = 0, p - value = 0.228$$

▶ 공간 시차 모형이 **적절하지 않음**

공간 오차 모형

$$H_0: \lambda = 0, p - value = 0.690$$

▶ 공간 오차 모형이 **적절하지 않음**

거리 기반 공간회귀모형: 2) Power Distance Weights

공간 시차 모형

$$H_0: \rho = 0, p - value = 0.105$$

▶ 공간 시차 모형이 **적절하지 않음**

공간 오차 모형

$$H_0: \lambda = 0, p - value = 0.228$$

▶ 공간 오차 모형이 **적절하지 않음**

공간회귀모형이 모두 기각되어 OLS 모형을 최종모형으로 선정

최종 모델 :

$$\text{가정폭력검거건수} = -0.285 + 0.00073 * \text{총청소년인구} + 0.0094 * \text{총수급자인원} + 0.083 * \text{혼인건수}$$

04. 클러스터링

클러스터링 방식

K-Means
클러스터링

계층적
클러스터링

DBSCAN
클러스터링

K-Medoids
클러스터링

GMM
클러스터링



5가지 클러스터링 방법을 시도하여
청소년쉼터 추가 입지 후보 행정동을
선정하는 것이 목표!

04. 클러스터링

클러스터링 변수

가정폭력 위험도

가출한 청소년들의 약 70%가 부모와의 불화를 가출 원인으로 꼽았으므로
가출에 유의미하게 큰 영향력을 끼치는 지표라고 고려되어 선택



청소년 인당 교육비 지출액

가족과의 문제 다음으로 높았던 가출 요인인 학업 부담과 관련된 영향을 고려하기 위해
행정동별 청소년 인당 교육비 지출액을 변수로 선택



04. 클러스터링

이상치 제거

효과적인 클러스터링을 위해서 이상치로 간주되는 행정동은 클러스터링 대상에서 제외 !

자치구	행정동	가정폭력 위험도	청소년 인당 교육 지출액
서초구	반포본동	0.26853	18,902,661 원

전체 평균에 비해
너무 높은 **청소년 인당 교육 지출액**과,
너무 낮은 **가정폭력 위험도**

자치구	행정동	총 청소년 수	청소년 인당 교육 지출액
강동구	둔촌1동	6명	210,673,996 원

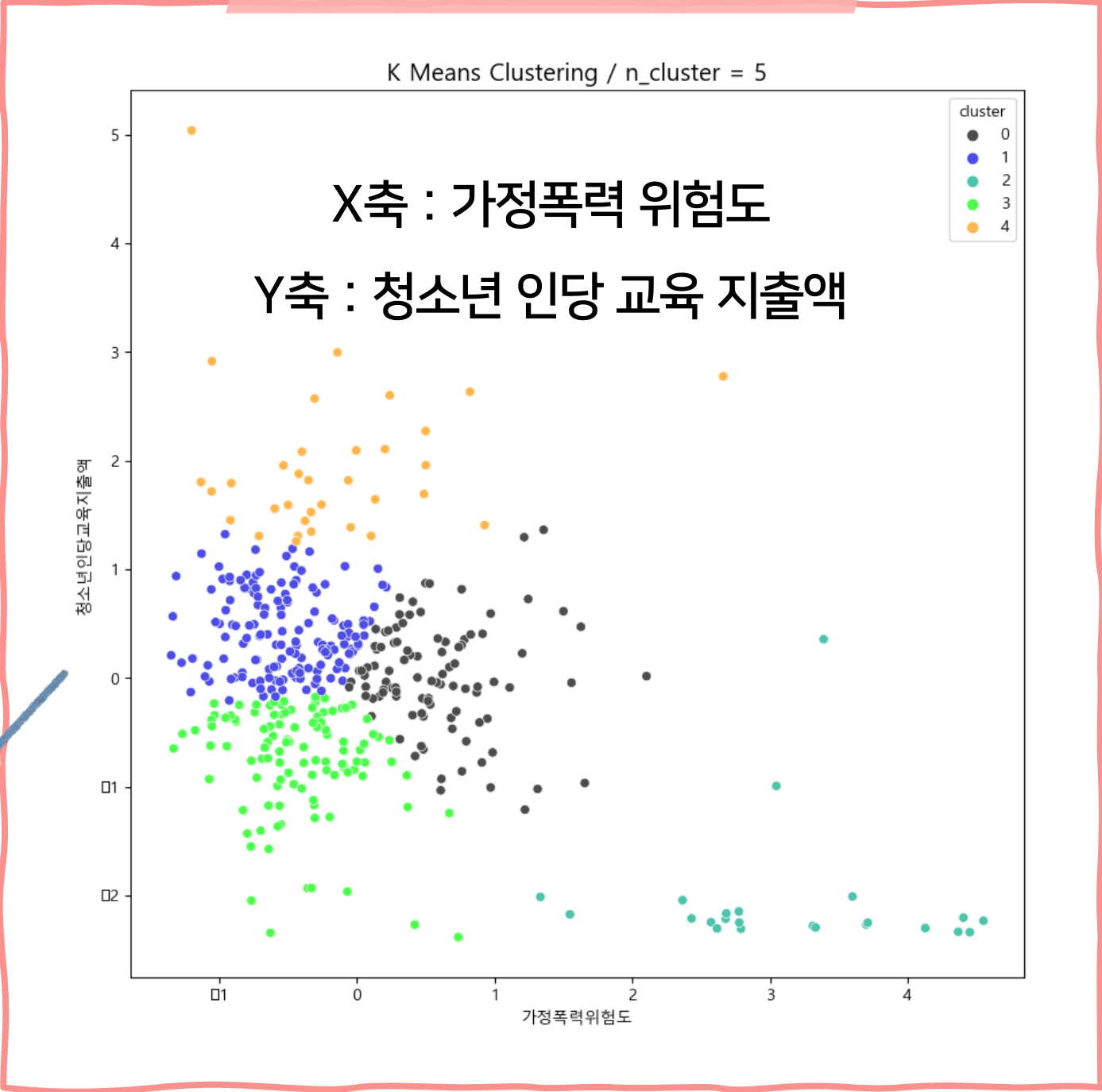
전체 평균에 비해
너무 높은 **청소년 인당 교육 지출액**과,
너무 적은 **총 청소년 수**

이 밖에도, 이미 청소년쉼터가 입지해 있는 13개의 행정동을 제외하고
총 411개의 행정동으로 클러스터링을 진행!

04. 클러스터링

클러스터링 결과

클러스터링 산점도 Plot



클러스터별 변수 평균

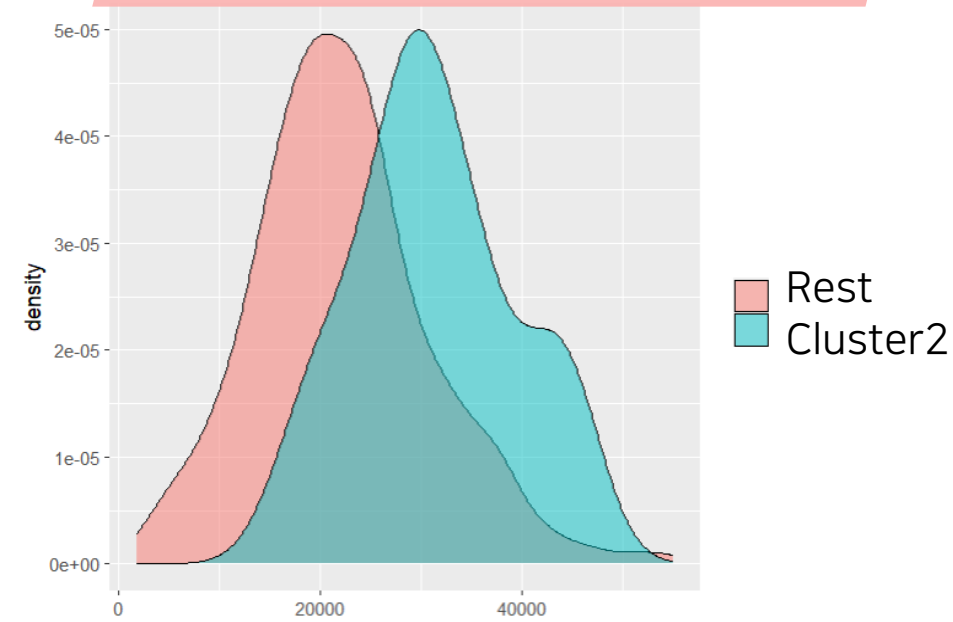
클러스터	가정폭력 위험도
2	3.154047
0	0.581419
4	-0.175955
3	-0.422075
1	-0.527439

클러스터	청소년 인당 교육 지출액
4	1.960837
1	0.416856
0	0.011573
3	-0.750698
2	-2.054539

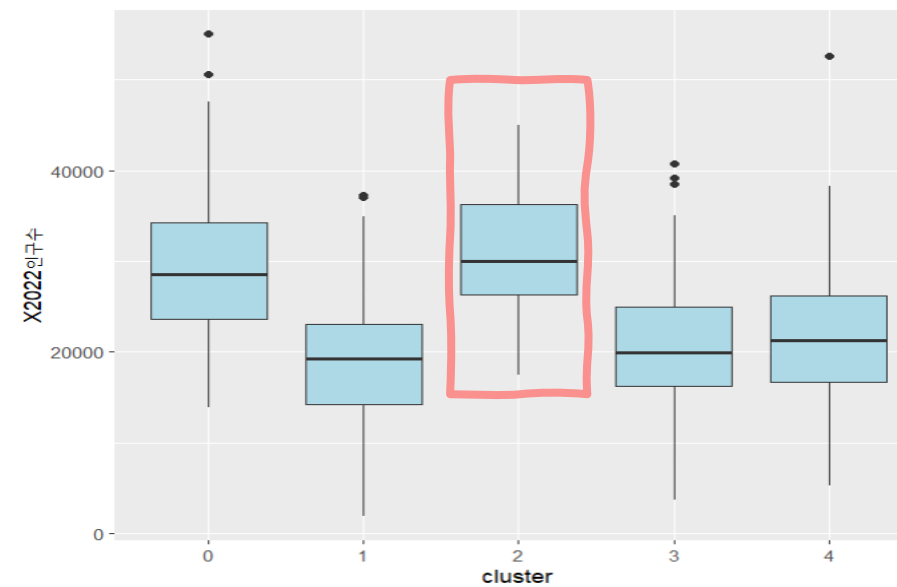
04. 클러스터링

클러스터링 EDA

인구 수 Density Plot

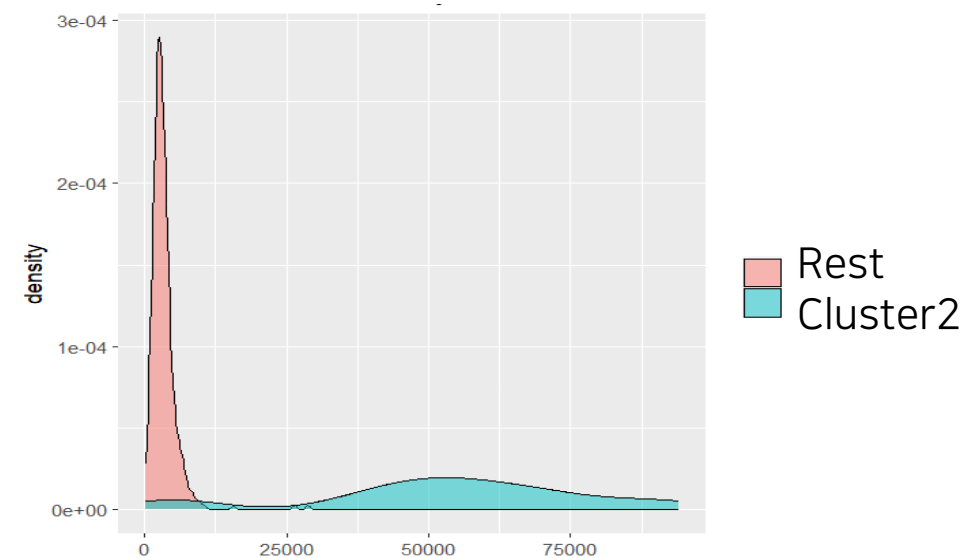


인구 수 Box Plot

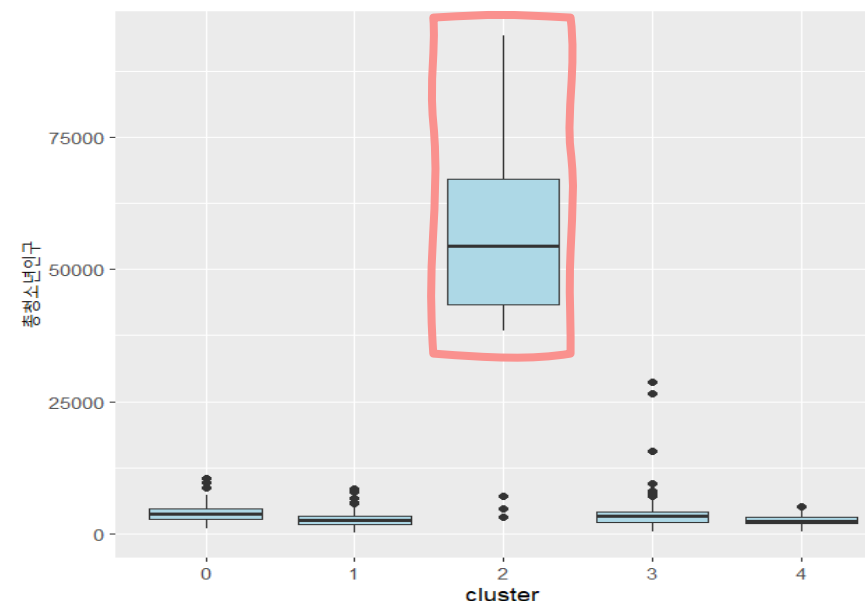


2번 클러스터는
다른 클러스터에 비해 인구 수가 많음

총 청소년 인구 Density Plot



총 청소년 인구 Box Plot

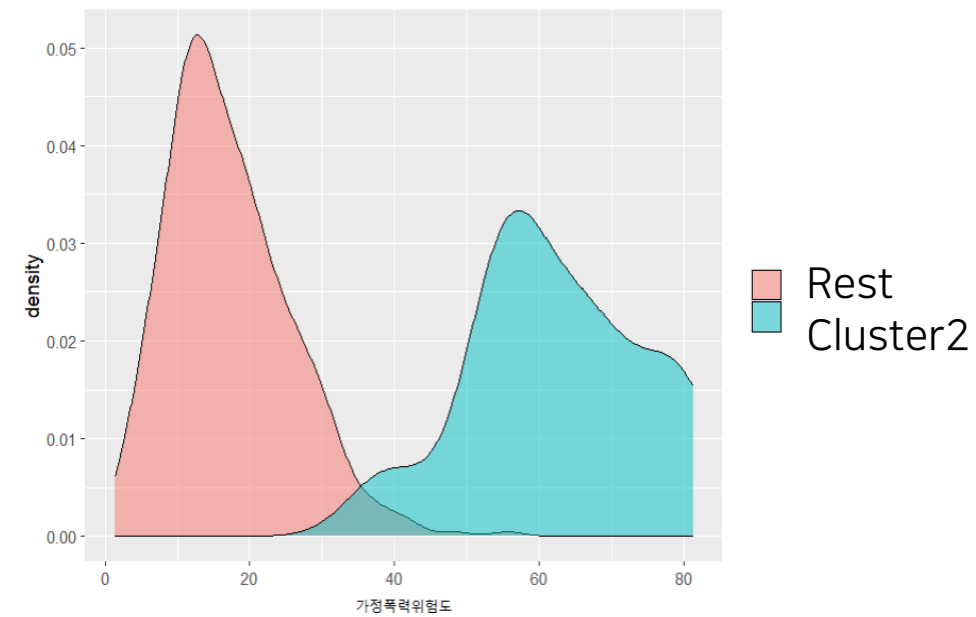


특히, 2번 클러스터는 다른 클러스터에 비해
총 청소년 인구 수가 매우 많음

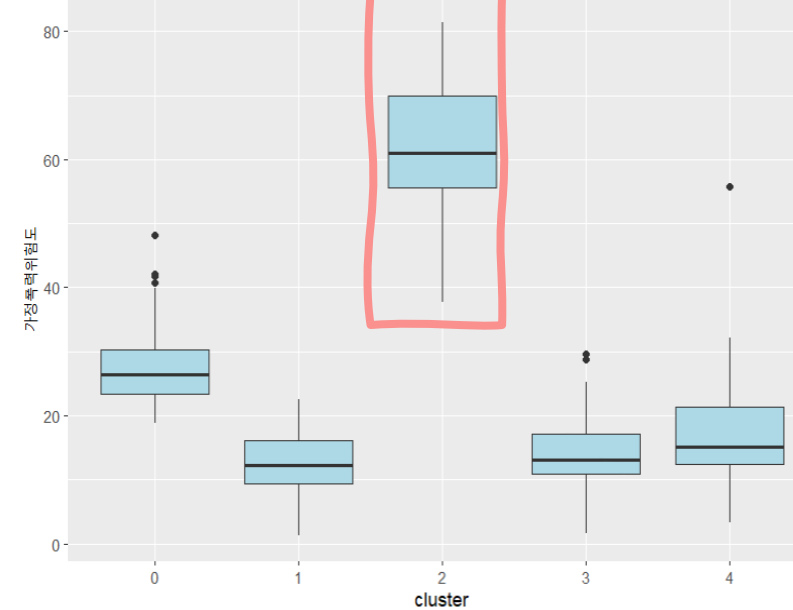
04. 클러스터링

클러스터링 EDA

가정폭력 위험도 Density Plot

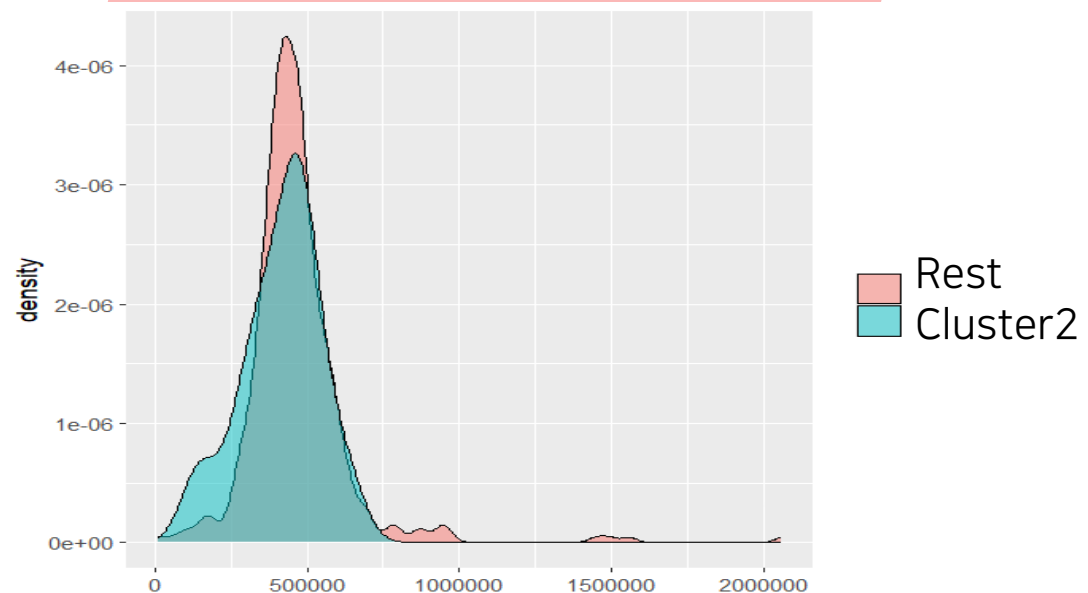


가정폭력 위험도 Box Plot

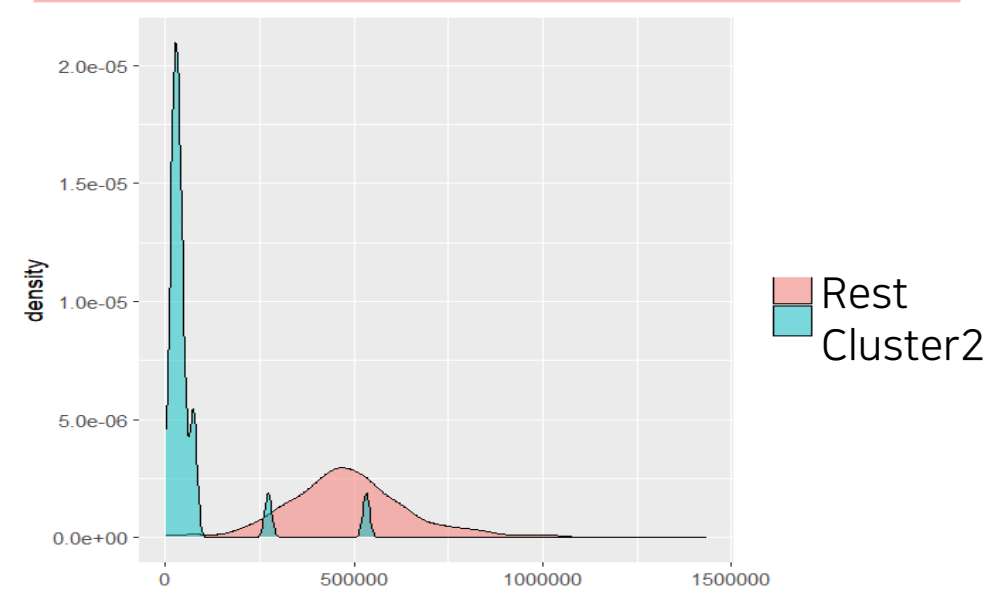


2번 클러스터의 가정폭력 위험도는 다른 클러스터에 비해 매우 높음

인당 지출 총 금액 Density Plot



청소년 인당 교육 지출액 Density Plot



2번 클러스터의 경우, 인당 지출 총 금액은 타 클러스터와 유사하나, 청소년 인당 교육 지출액은 매우 낮음

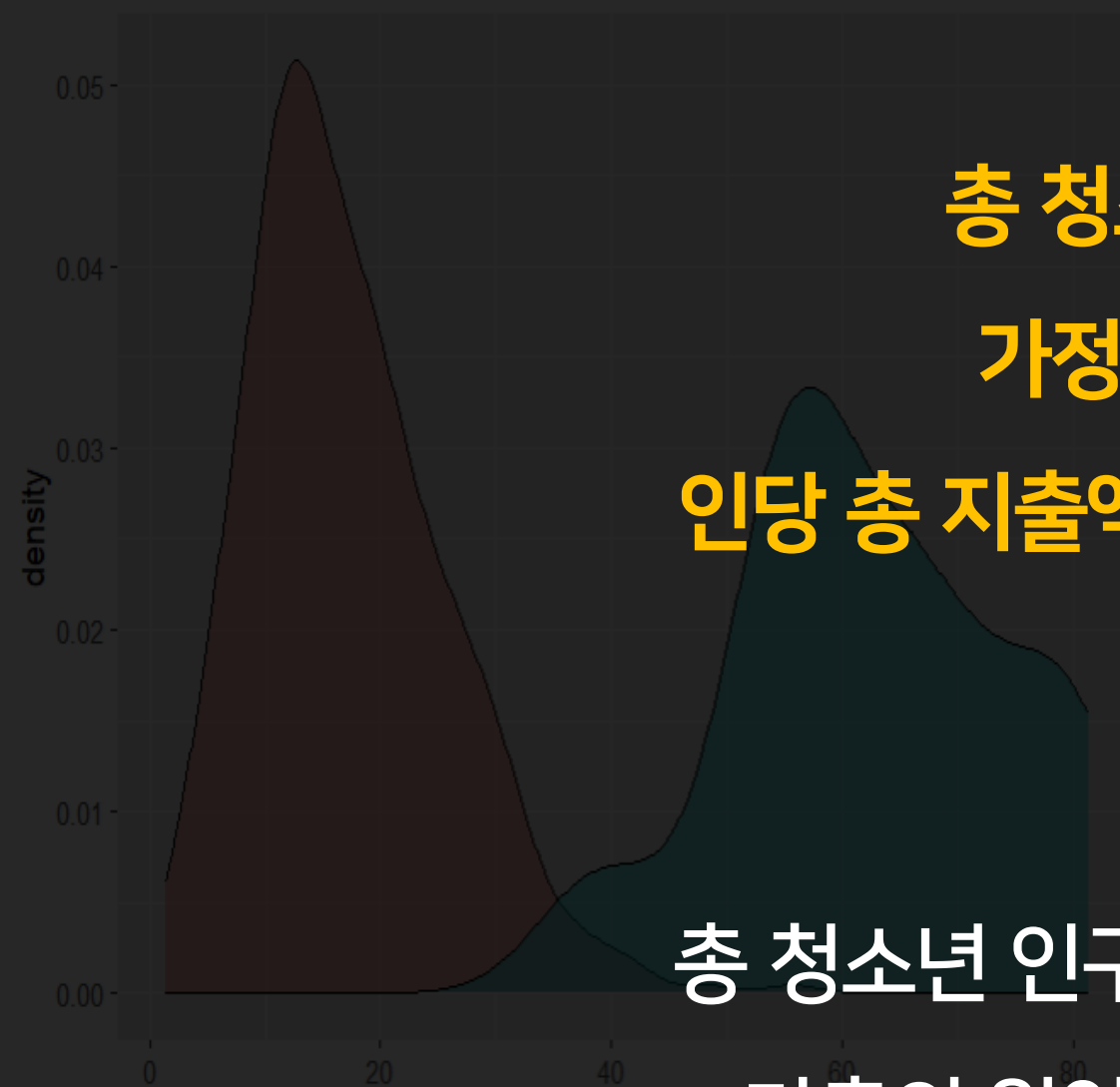
04. 클러스터링



클러스터링 EDA

최종 클러스터 선정

가정폭력 위험도 Density Plot



인구 수) 많음

총 청소년 인구 수) 매우 많음

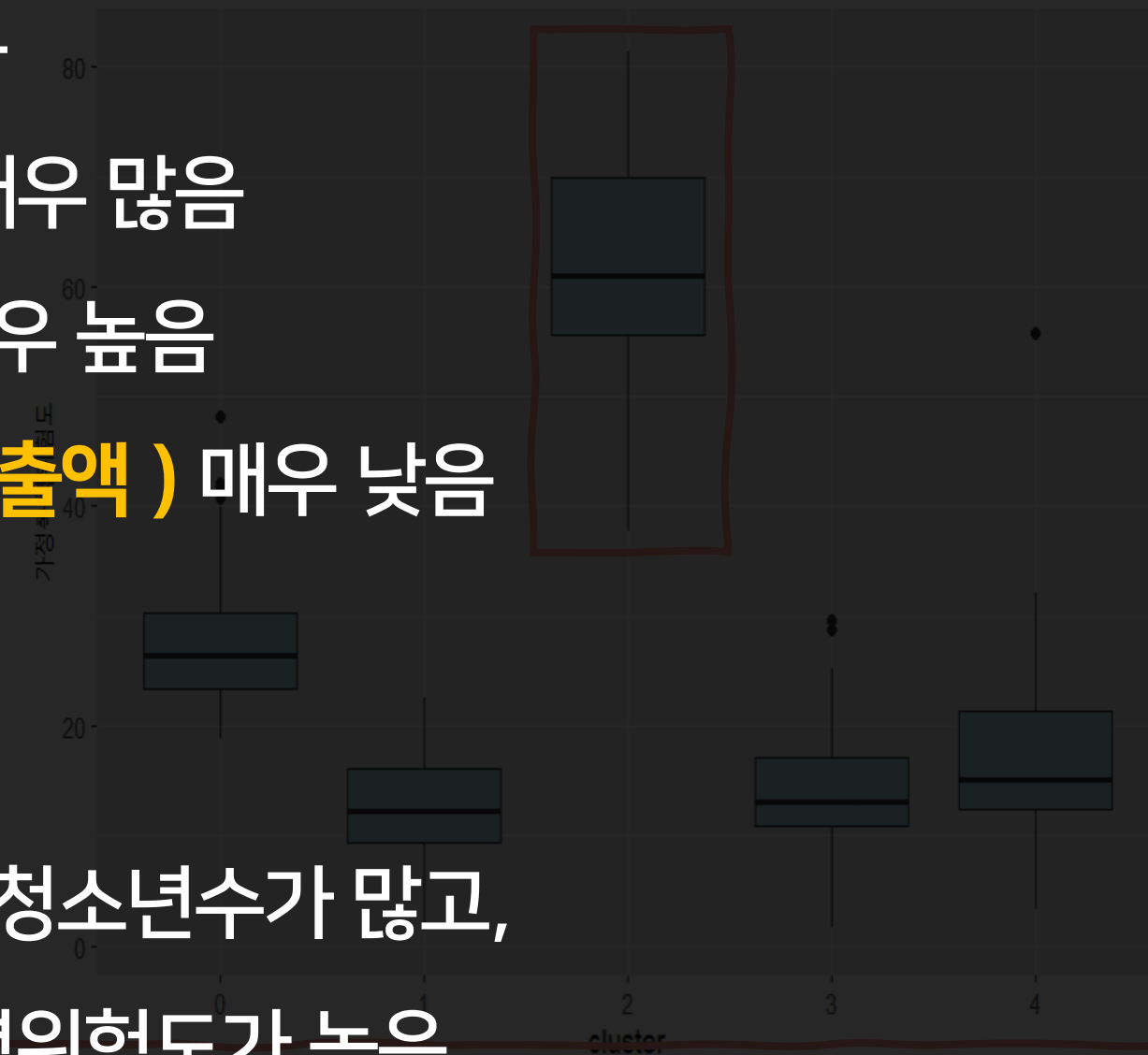
가정폭력 위험도) 매우 높음

인당 총 지출액 대비 교육비 지출액) 매우 낮음



총 청소년 인구수가 많아 가출 청소년수가 많고,
가출의 원인이 되는 가정폭력위험도가 높은

가정폭력 위험도 Box Plot



2번 클러스터를 최종 우선 입지 클러스터로 선정!

05. 청소년쉼터 평가 지표

분류에 활용할 청소년 쉼터

시군구	시설 종류	시설명	시설 주소
용산구	일시쉼터 (고정형)	서울특별시청소년일시쉼터	서울 용산구 만리재로 156-1 (서계동)
중랑구	단기쉼터 (여자)	서울시립망우청소년단기쉼터	서울 중랑구 송림길 156 (망우동)
...
관악구	중장기쉼터 (남자)	서울시립신림청소년중장기쉼터	서울 관악구 난곡로 24가길 54, 301호

청소년 쉼터의 존재 여부에 영향을 미치는 변수 확인을 위해
위 3가지 시설 종류에 해당하는 쉼터 13개에 대해 예측!



05. 청소년쉼터 평가 지표

청소년 쉼터 중요 변수 확인

로지스틱 회귀 모델 - 계수 기반 중요 변수 판단

로지스틱 회귀 모델의 회귀 계수 유의성 확인

426개 중 13개 행정동에만 청소년 쉼터가 존재하므로

Class weight를 설정해 클래스 불균형 해소!

트리 모델 - Feature Importance

RandomForest와 XGBoost Classifier의 **Feature Importance** 확인

참고용으로만 사용 예정!

05. 청소년 심터 평가 지표

청소년 심터 중요 변수 확인

로지스틱 회귀 모델 - 계수 기반 중요 변수 판단

로지스틱 회귀 모델의 회귀 계수 유의성 확인



이혼건수	0.002020
일반_조건부_수급자인원	0.010019
재혼율	0.228715
장애인수	0.031936
총청소년인구	0.118588
청소년인당교육지출액	0.988522
치안기관수	0.752338
카메라대수	0.412048
학원수	0.012459

유의한 것으로 나타난 변수들 ▼

이혼 건수, 일반/조건부 수급자 인원, 장애인 수, 학원 수

```
9 # 회귀 계수 확인
10 print(model.coef_)
11
12 # 절댓값 큰 거 기준으로 봤을 때, 7 8 10 15번째 변수들이 주요함..!
```

```
[[ 0.          0.23816871  0.41462045  0.         -0.23385615  0.27495005
  1.29704328  1.14606076 -0.28637841 -1.3044399   0.          0.
 -0.01647045  0.21887605 -1.91715457]]
```

결과

이혼건수, 일반조건부수급자인원, 장애인수, 학원수

장애인 수, 학원 수 변수는 회귀 계수가 음수로 나타나

직관에 어긋나므로 사용하지 않음

05. 청소년 심터 평가 지표

청소년 심터 중요 변수 확인

로지스틱 회귀 모델 - 계수 크기 기준으로 가장 유의한 것으로 나타난 변수들 ▼

RandomForest) 유해 업소 수, 이혼 건수, 일반/조건부 수급자 인원, 학원 수

XGBoost) 범죄 총 발생 건수, 이혼 건수, 학원 수



트리 모델 - Feature Importance

RandomForest와 XGBoost Classifier의 **Feature Importance** 확인

참고용으로만 사용 예정!

05. 청소년 심터 평가 지표

청소년 심터 중요 변수 확인

- ✓ 청소년 심터 입지에 영향을 주었던 변수 중 **이혼 건수**, **일반/조건부 수급자 인원**이 모델에서 유의한 효과를 보임.
따라서, 이 두 변수를 최종 모델에 활용하기로 결정
- ✓ 추가로 직접적인 가출의 원인이 되는 **가정폭력 위험도**와, 가출 시 위험도를 나타내는 **범죄 총 발생 건수** 또한
최종 변수로 고려하기로 결정



Rule		Lift
가정폭력 검거 건수_low, 일반/조건부 수급자 인원_low	이혼 건수_low	3.125

연관규칙 검정에서 가정폭력 검거 건수, 이혼 건수, 수급자 인원이 연관되어 있어
가정폭력 위험도를 평가지표에 활용가능

05. 청소년 심터 평가 지표

PCA

PCA

차원축소 방법의 일종으로 데이터의 분산을 최대화하며 변수 간의 선형결합으로 축을 생성하는 방법

개별 축은 서로 직교하며, 축 간의 분산 크기가 상관계수 행렬의 고유값에 비례하므로 차이가 존재



이때 첫 번째 축은 정보량을 최대한 보존하는, 각 변수의 결합으로 만들어지는 새로운 축.

이 축을 청소년 심터 입지 판단 기준으로 활용!

이를 위해 변수끼리 같은 방향으로 움직이도록 변환하고, 생성된 선형 결합의 계수도 방향성이 일치해야 함

05. 청소년 심터 평가 지표



PCA

PCA 결과

PCA

```
> pca_obj
Standard deviations (1, ..., p=4):
[1] 1.6006538 0.8985575 0.6112725 0.5068016

Rotation (n x k) = (4 x 4):
```

	PC1	PC2	PC3	PC4
이혼건수	0.5515399	-0.09475336	0.56253166	0.6085916
가정폭력위험도	0.5324517	0.18178128	-0.78189870	0.2684868
범죄총발생건수	0.4176503	-0.78666672	-0.05351917	-0.4515079
일반_조건부_수급자인원	0.4877163	0.58235113	0.26330259	-0.5947031

```
> summary(pca_obj)
Importance of components:
```

	PC1	PC2	PC3	PC4
Standard deviation	1.6007	0.8986	0.61127	0.50680
Proportion of Variance	0.6405	0.2019	0.09341	0.06421
Cumulative Proportion	0.6405	0.8424	0.93579	1.00000

이 값을 청소년 심터 입지 판단 기준으로 활용
PC1의 분산이 약 64%로 상당히 높으므로 **가중치로 사용**하기 적절함!

변수에 곱해질 **가중치들의 방향성 (+)** 또한 **직관과 부합**하므로 최종 행정동 선정에 사용!

05. 청소년쉼터 평가 지표

PCA | 평가 지표 생성

기존에 쉼터가 없던 행정동의 클러스터에 대해 실시!

자치구	행정동	쉼터 유무	PCA Score
은평구	역촌동	0	5.80
노원구	중계2.3동	0	5.09
강서구	등촌3동	0	5.09
강남구	세곡동	0	4.45
노원구	공릉1동	0	4.42
...



자치구	행정동	쉼터 유무	PCA Score	평가 지표
은평구	역촌동	0	5.80	100.0
노원구	중계2.3동	0	5.09	86.53
강서구	등촌3동	0	5.09	86.52
강남구	세곡동	0	4.45	74.52
노원구	공릉1동	0	4.42	73.95
...

PC1의 가중치를 바탕으로 행정동별 PCA Score를 산출!

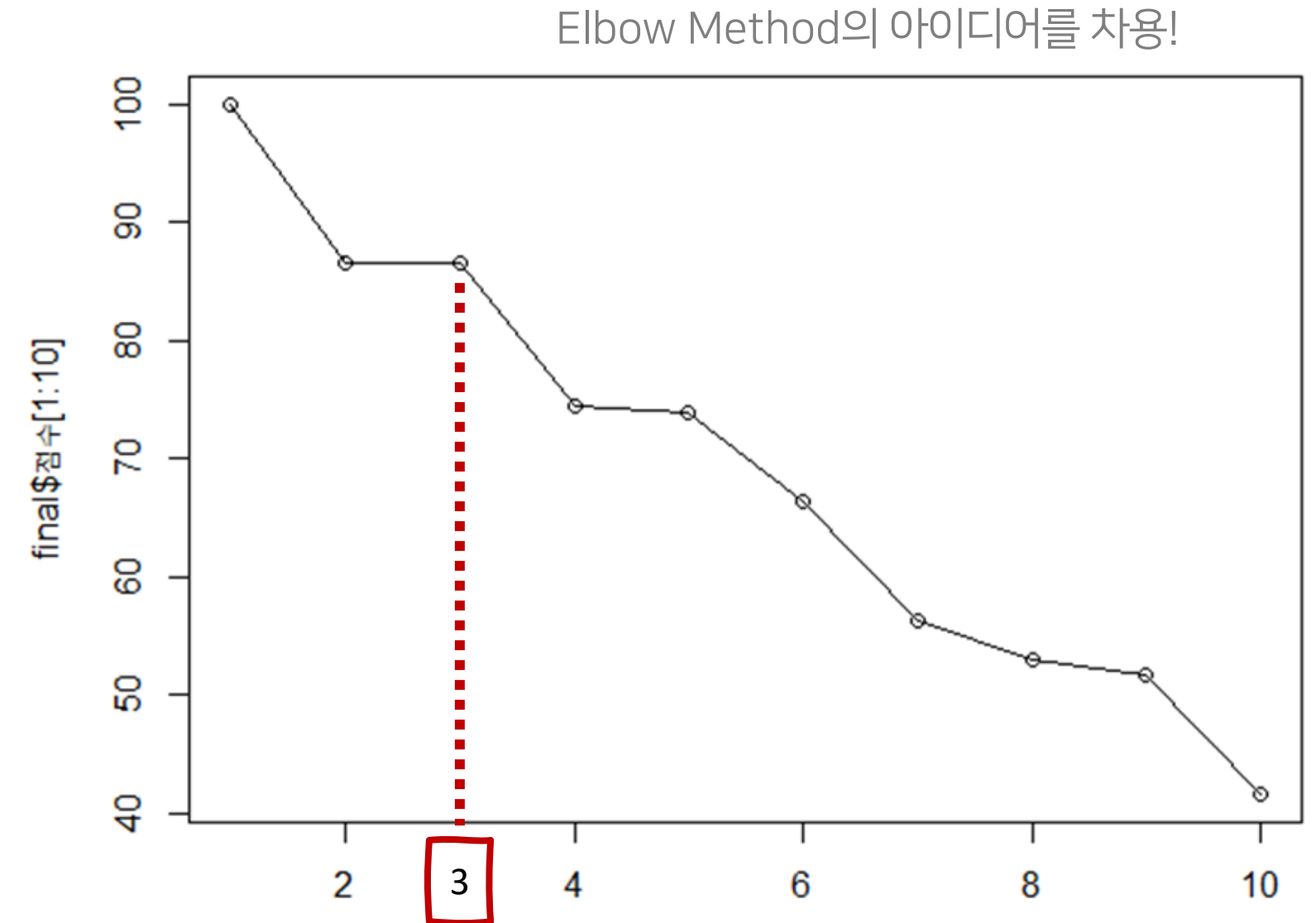
Min-Max Scaling 후 100을 곱하여, PCA Score를 점수화한 평가지표를 완성!

Min-Max Scaling은 0~1 사이의 값으로 압축함

05. 청소년 심터 평가 지표

PCA | 행정동 선정

자치구	행정동	평가 지표
은평구	역촌동	100.0
노원구	중계2,3동	86.53
강서구	등촌3동	86.52
강남구	세곡동	74.52
노원구	공릉1동	73.95
...



최종 평가 지표를 Scree Plot 형태로 시각화

▶ 3번째 행정동부터 점수가 급격히 하락하므로, **최종 입지 행정동을 3개로 선정!**

05. 청소년쉼터 평가 지표

PCA | 행정동 선정

자치구	행정동	평가 지표
은평구	역촌동	100.0
노원구	중계2,3동	86.53
강서구	등촌3동	86.52
강남구	세곡동	74.52
노원구	공릉1동	73.95
...



최종 입지 행정동

은평구 역촌동 / 노원구 중계2,3동 / 강서구 등촌3동

05. 청소년쉼터 평가 지표



PCA | 행정동 선정

입지 선정이 유의미한지 점검하기 위해 **쉼터 존재 행정동이 포함된 경우도** 구해봄

자치구	행정동	쉼터 유무	PCA Score	평가지표
은평구	역촌동	0	5.80	100.0
노원구	중계2,3동	0	5.09	86.53
강서구	등촌3동	0	5.09	86.52
노원구	공릉3동	0	4.45	74.52
...

최종 입지 행정동

실제로 쉼터가 존재하는 '중랑구 망우본동'이 2위로 나옴

은평구 역촌동 / 노원구 중계2,3동 / 강서구 등촌3동

▶ 평가 지표의 타당성 확인!

06. 최종 입지 선정 및 결론

행정동 내 입지 선정

청소년복지 지원법 시행령 제17조 (별표3)

기준		기준	
입지조건	■ 시설의 적정한 분포와 보건·위생·급수·안전·환경 및 교통편의 등을 고려하여 쾌적한 환경의 부지에 설치	입지조건	■ 시설의 적정한 분포와 보건·위생·급수·안전·환경 및 교통편의 등을 고려한 쾌적한 환경의 부지에 설치
모형	■ 단독건물 또는 상가건물 ■ 입소인원 1인당 연면적 11제곱미터 이상	모형	■ 단독건물, 연립주택, 상가건물, 아파트 등 ■ 입소인원 1인당 연면적 11제곱미터 이상
조리실·식당	■ 위생적인 취사·조리 및 식기소독 설비의 구비	조리실·식당	■ 위생적인 취사·조리 및 식기소독 설비의 구비
화장실·목욕실	■ 남·여용 구분 설치 ■ 수세식 화장실, 세면·샤워 설비 및 온수 공급	화장실·목욕실	■ 남·여용 구분 설치 ■ 수세식 화장실, 세면·샤워 설비 및 온수 공급
세탁·건조장	■ 세탁 및 건조용 설비	세탁·건조장	■ 세탁 및 건조용 설비
상담실	■ 칸막이 또는 방음장치를 설치하여 개별상담 및 집단상담이 가능한 공간	상담실	■ 칸막이 또는 방음장치를 설치하여 개별상담 및 집단상담이 가능한 공간
단체활동실	■ 최대 15명 이상이 참석할 수 있는 공간	단체활동실	■ 최대 15명 이상이 참석할 수 있는 공간
사무실	■ 사무를 위한 설비를 갖춘 분리된 공간 ■ 이동형 일시침터를 같이 운영하는 경우 고정형 일시침터 운영을 위한 사무실과 별도의 사무실	사무실	■ 사무를 위한 컴퓨터, 책상 등 설비를 갖춘 공간
비상재해대비시설	■ 「화재예방, 소방시설 설치·유지 및 안전관리에 관한 법률」에서 정하는 소화기 및 피난기구 등 시설 실정에 맞는 비상재해대비시설 및 장비의 설치·구비	비상재해대비시설	■ 「화재예방, 소방시설 설치·유지 및 안전관리에 관한 법률」에서 정하는 소화기 및 피난기구 등 시설 실정에 맞는 비상재해대비시설 및 장비의 설치·구비

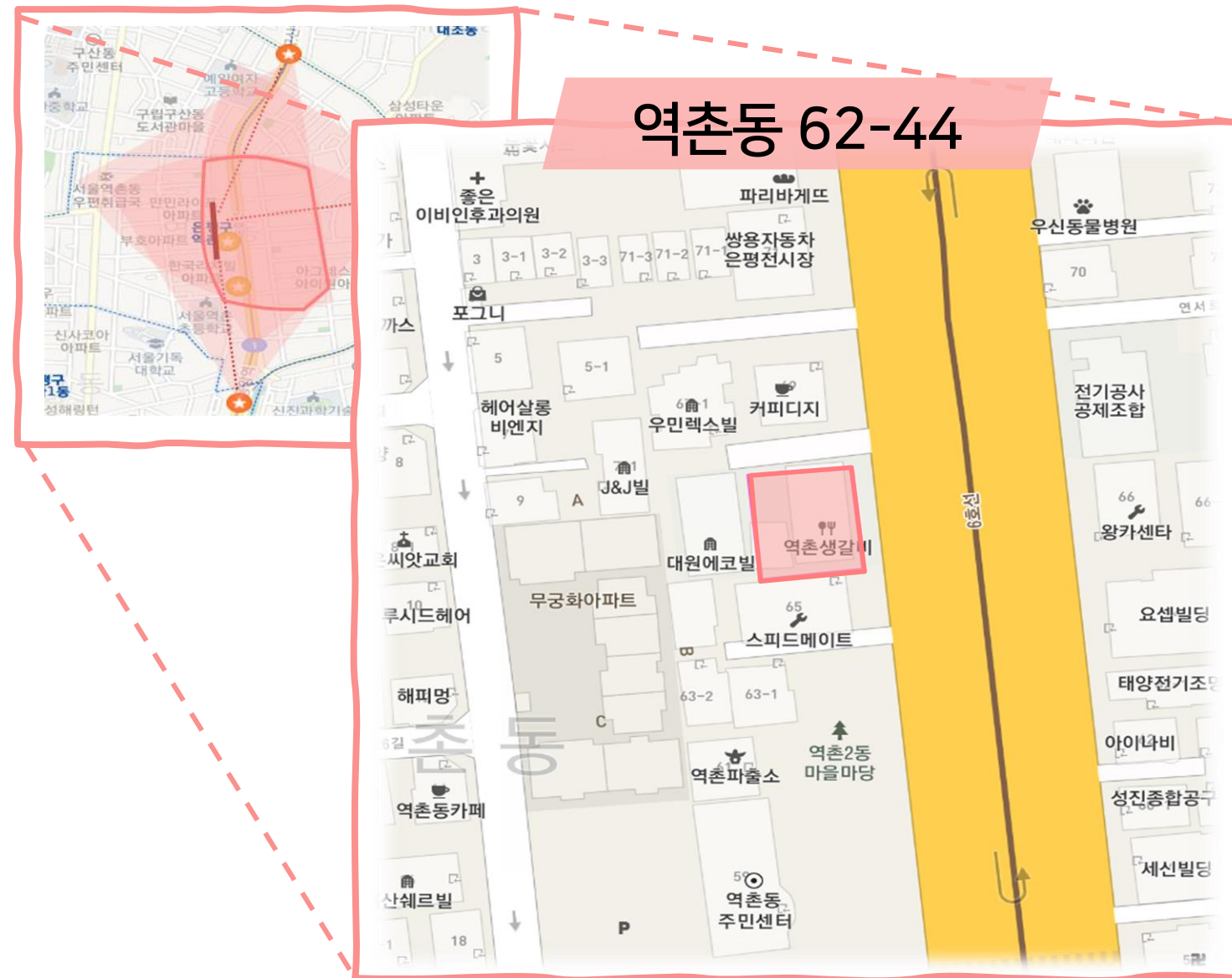
청소년쉼터에 관한 공식적인 입지조건이 불분명함



‘평가지표’와 ‘접근성’을 바탕으로 청소년쉼터의 입지를 판단!

06. 최종 입지 선정 및 결론

행정동 내 입지 선정 | ① 은평구 역촌동



역촌동 62-44

빌라와 같은 다가구 주택이 밀집해 있으며
서울 전체 평균에 비해 집값이 낮은 편

역촌동에 존재하는 3개의 지하철역과의
거리 (★) 를 바탕으로 접근성 판단

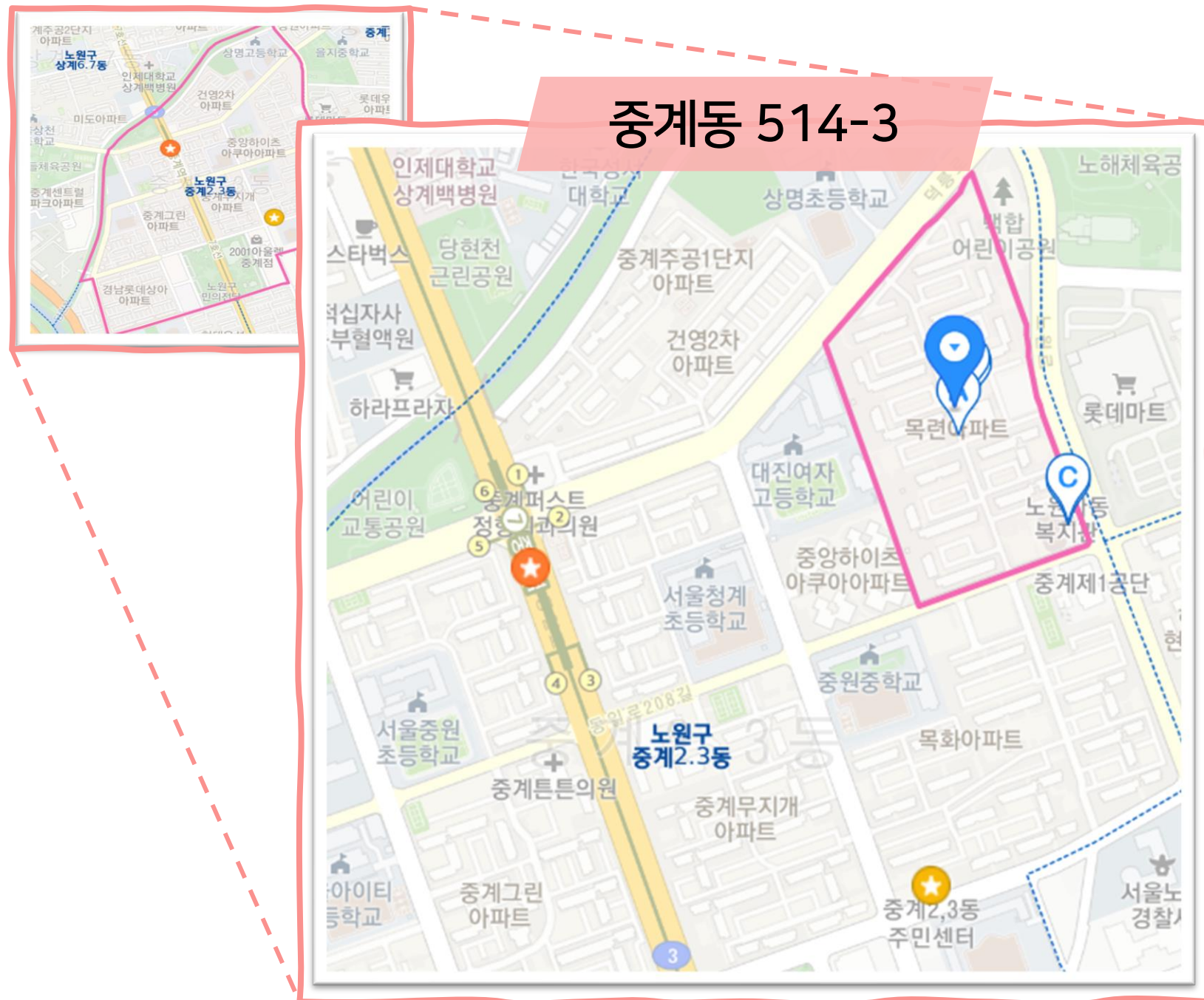
파출소와 주민센터 (★) 인근 건물 중
설치 기준에 부합하는 건물 선정

2023.04. 해당 위치의 음식점 폐업



06. 최종 입지 선정 및 결론

행정동 내 입지 선정 | ② 노원구 중계 2, 3동



중계 2, 3동 특성상 주거 단지가 대부분이고,
상가를 비롯한 기타 시설은 행정동 외곽에 위치

중계역까지의 거리 (★) 를 바탕으로 접근성 판단

파출소와 주민센터 (★) 인근 건물 중
설치 기준에 부합하는 건물 선정



06. 최종 입지 선정 및 결론

행정동 내 입지 선정 | ③ 강서구 등촌 3동

등촌동 663-7



분홍색으로 표시된 아파트 단지 외 구역을
심터 입지 선정 후보 구역으로 설정

등촌동에 존재하는 3개의 지하철역과의
거리 (★) 를 바탕으로 접근성 판단

주변에 유흥시설이 위치해 있지 않으면서
설치 기준에 부합하는 건물 선정



06. 최종 입지 선정 및 결론

의의 및 한계

분석 의의

1) 비지도학습을 이용해 지표 생성

주제와 직결된 가출 변수를 구할 수 없음에도, 여러 비지도학습을 이용해 지표 생성

2) 시의성을 고려한 분석

기존의 기간 중심에서 기능 중심으로 변화되어야 한다는 청소년쉼터의 시의성을 고려한 분석을 진행

분석 한계

1) 주관적인 입지 선정

건물 단위로 입지분석을 진행하려 했으나, 접근성 이외의 공식적인 기준이 없어,주관적으로 최종 입지 선정

2) 변수 부재

가출 건수라는 핵심적인 변수를 구할 수 없어, 가출에 대해 간접적으로 영향력을 주는 변수로만 분석 진행

감사합니다

