공개SW (Open Source SW)를 중심으로 하는

공간정보 빅데이터 분석 및 실습

03. QGIS 기반 분석/시각화



CCTV-확인 및 1차 전처리

- 먼저, 용량을 확인(약, 30MB 정도로 일반적인 텍스트 편집기나 스프레드시트 SW에서 무리 없이 띄워볼 수 있음)
 - 만약 대용량이면 Vim (윈도우용은 gVim)에서 GB 단위 파일을 열어볼 수 있음 (파일 용량만큼 메모리를 사용하므로 주의)
 - 또는 Python-Pandas 환경에서 확인하는 것이 유리(역시 파일 용량만큼 메모리를 사용)
- Header가 한글로 되어 있음
- QGIS에서 필드 명칭 및 타입 수정하는 기능(Refactor fields)이 있으나 재저장이 필요하므로 텍스트 편집기에서 영문으로 수정
 - 표준프레임워크 등에서 제공하는 (표준)용어사전 등을 이용하여 일관성 있고 용어사전이 있으면 다른 사람도 이해할 수 있는 용어를 사용하는 것을 권장
 - 만약, 공간데이터를 shp으로 저장할 경우 항목명을 8자리까지만 지원하는 점 감안

○ 기타

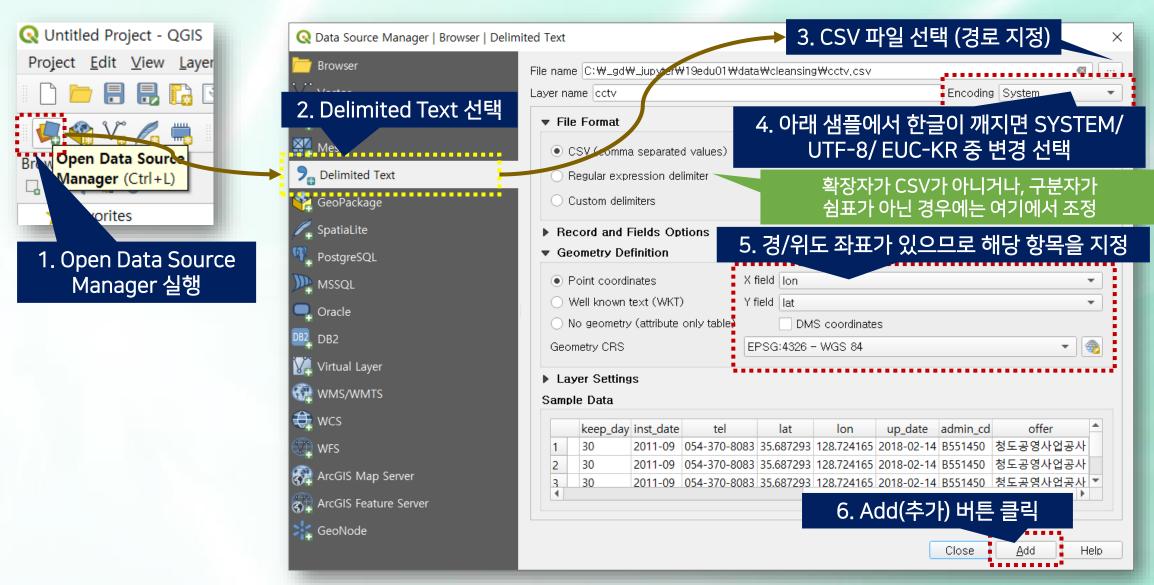
촬영방면정보가 방위값으로 되어있고, 촬영 거리 항목이 추가된다면 좀 더 세밀한
 촬영 영역을 알 수 있을 텐데 데이터상 일괄적인 반경으로 표현할 수 밖에 없음

🔚 전국cctv표준데이터.csv 🗵

- 1 관리기관명, 소재지도로명주소, 소재지지번주소, 설치목적구분, 카메라대수, 카메라화소수, 촬영방면정보, 보관일수, 설치년월, 관리기관전화번호, 위도, 경도, 데이터기준일자,
- 2 청도공영사업공사,경상북도 청도군 화양읍 남성현로 348,경상북도 청도군 화양읍 삼신리 693-2,다목적,1,41,지하1층 사무실 복도(3),30,2011-09,05
- 3 청도공영사업공사,경상북도 청도군 화양읍 남성현로 348,경상북도 청도군 화양읍 삼신리 693-2,다목적,1,41,지하1층 사무실 복도(2),30,2011-09,05
- 4 청도공영사업공사,경상북도 청도군 화양읍 남성현로 348,경상북도 청도군 화양읍 삼신리 693-2,다목적,1,41,지하1층 사무실 복도(1),30,2011-09,05
- 5 청도공영사업공사,경상북도 청도군 화양읍 남성현로 348,경상북도 청도군 화양읍 삼신리 693-2,다목적,1,41,지하1층 세미나실 내부,30,2011-09,054-

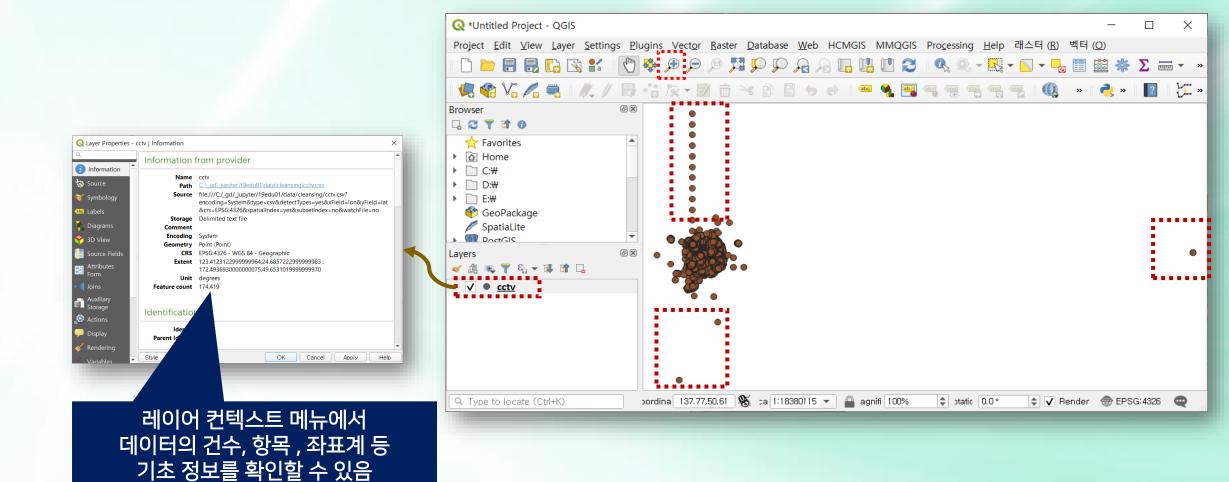
CCTV-좌표를 이용하여 공간데이터로 로딩

○ CCTV의 경우 경도, 위도 좌표가 포함되어 있어서 인코딩과 좌표 항목만 지정하면 쉽게 공간데이터로 로딩이 가능함



CCTV-로딩 결과

- 얼핏 봐도 한반도 영역 밖에 데이터들이 있는 것이 보임
- 지도 확대(Zoom in)나 마우스를 이용하여 점들이 많은 곳을 확대

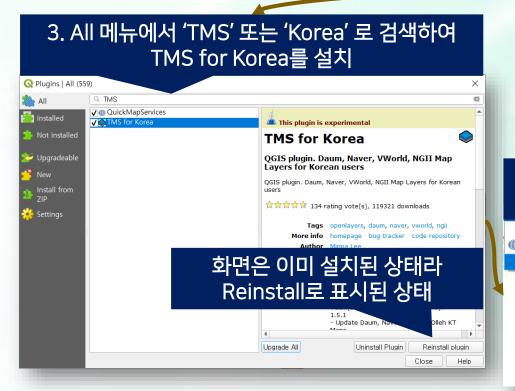


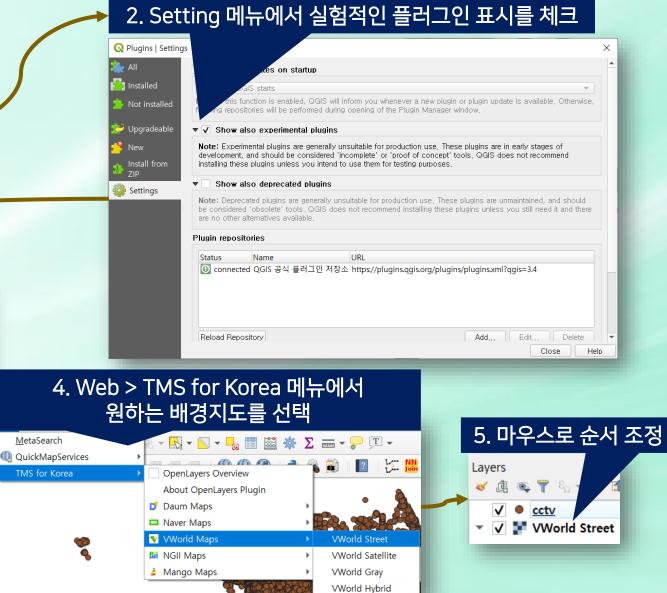
CCTV-TMS 배경지도 참조

○ 위치 확인을 위해 배경지도를 활용

1. 플러그인에서 플러그인 관리 및 설치를 실행

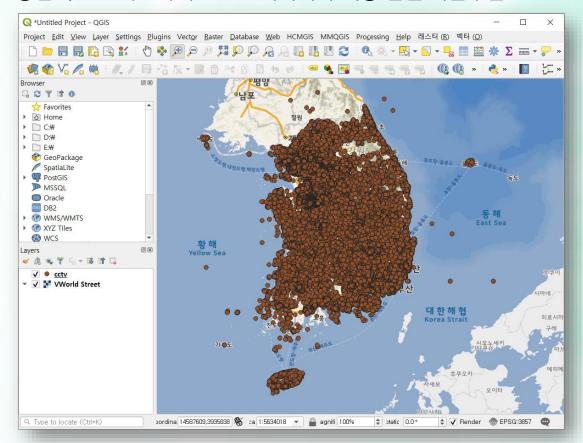
Plugins Vector Raster Database Web HCMGIS
Manage and Install Plugins...





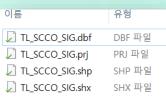
CCTV-배경지도 상에서의 CCTV 공간데이터 위치 확인

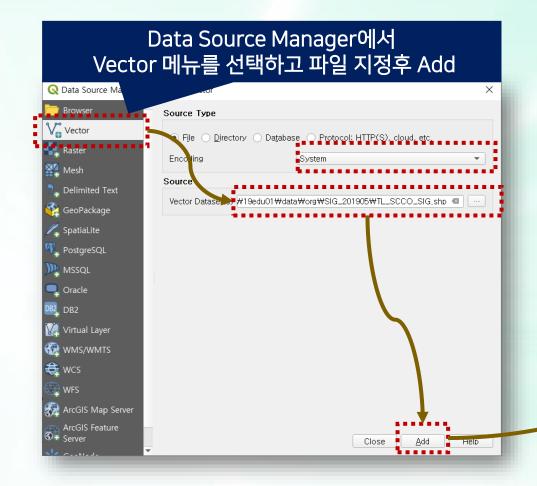
- 일부 지역이 비어있는 것처럼 보이고, 바다에 있는 경우(이상치)도 눈에 보임
- 이상치
 - 바다에 있는 이상치의 건수와 목록을 찾는 방법은?
- 결측지
 - 공간적으로 비교하여 CCTV 데이터가 제공 안된 시군구는?

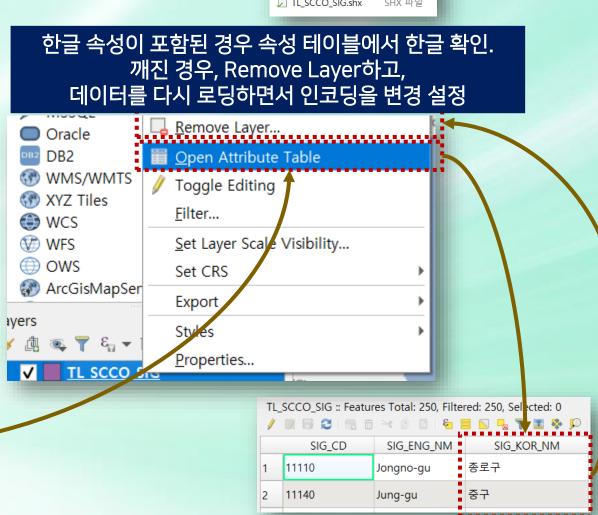


행정구역 데이터 로딩

○ 행정구역 데이터의 압축을 해제하면 오른쪽 그림과 같이 .prj 파일에 좌표계 정보가 포함되어 있어 좌표계를 물어보지 않고 바로 로딩함



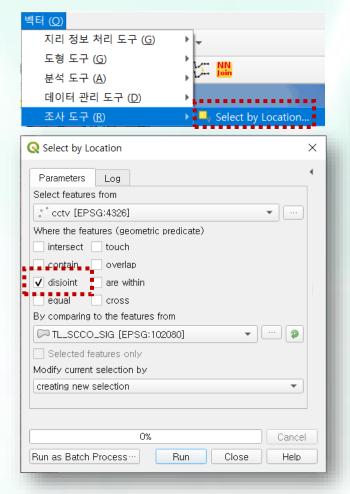


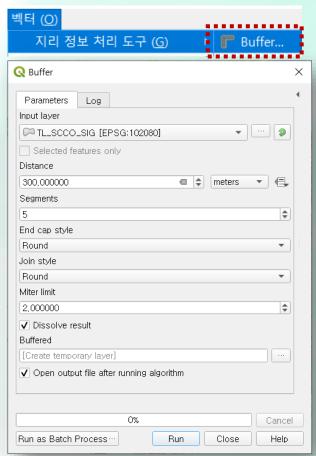


CCTV-이상치/결측치 검토

- 벡터 → Select by Location 메뉴 선택하고 그림과 같이 CCTV를 기준으로 행정구역과 만나지 않는 CCTV를 선택
- 그러나, 선택된 CCTV들을 확대해서 보면 해안가의 미세한 차이로 잘못 선택된 결과들이 보임
- 행정구역을 일정 반경으로 확장할 필요가 있으므로, 벡터 → Buffer를 실행하고 300m 반경으로 버퍼 생성 (Dissolve 옵션을 체크하면 시군구가 모두 합쳐진 영역이 생성됨)
- 버퍼 데이터가 생성되면 처음 과정인 Select by Location에서 비교 대상을 버퍼 데이터로 바꿔서 다시 진행
- 선택된 이상치는 레이어 컨텍스트 메뉴 〉 Export 〉 Save Selected Features ··· 에서 CSV 등으로 저장할 수 있음
- CCTV가 제공되지 않는 시군구(결측 영역)를 공간적으로 찿으려면?

CCTV-이상치/결측치 검토



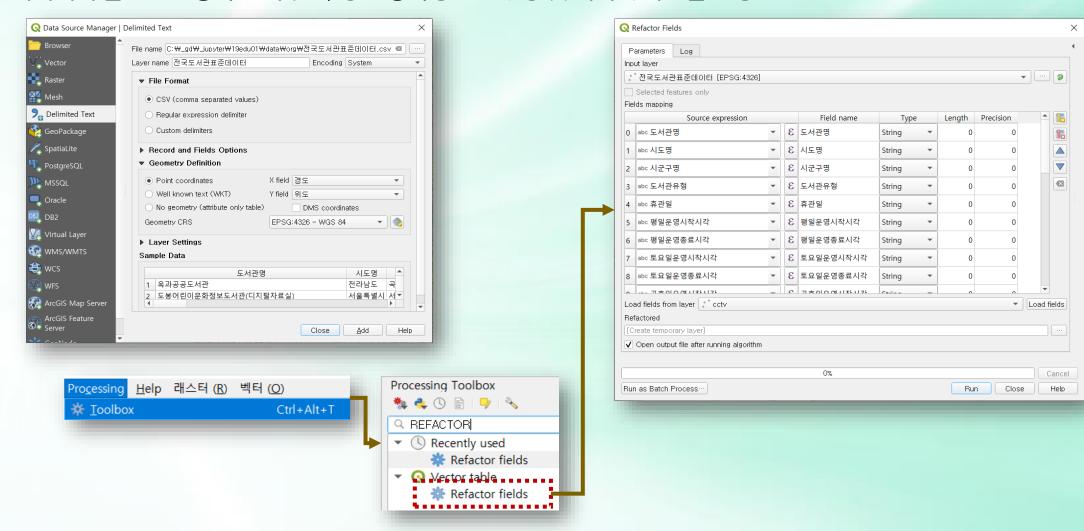




Save Selected Features로 CSV로 저장

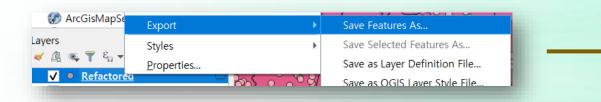
도서관-데이터 로딩

- 항목이 너무 많아서 한글 헤더를 그대로 둔 상태로 Delimited Text로 공간데이터화(한글 깨짐 여부만 주의)
- Processing → Toolbox 메뉴에서 'Refactor'로 검색하고 Refactor Fields를 실행하여 여기에서 불필요한 항목은 지우고, 영문 항목명으로 수정 및 데이터 타입을 변경

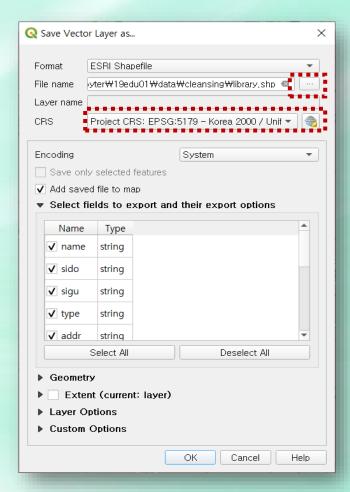


도서관-좌표계 변환

- 항목이 조정된 도서관 레이어가 추가되었으나, 좌표계는 EPSG 4326 (경위도) 그대로인 상태
- 추후 공간분석 등을 감안하여 모든 공간데이터를 EPSG 5179로 통일하기 위해 좌표계를 변환
 - 다른 파일로 저장하는 기능을 이용하면서 좌표계를 원하는 것으로 바꾸면 되는 방식
 - 항목(Field) 정리된 임시레이어 컨텍스트 메뉴 > Export > Save Features As…를 선택하고 원하는 경로 및 파일명을 지정
 - CRS 메뉴에서 목록 또는 지구본 아이콘을 통해 EPSG 5179를 선택
 - 인코딩은 한글이 정상적으로 보였던 것을 지정
 - 원하는 필드만 선택할 수 있음
 - OK 하고 좌표계 변환된 레이어의 위치를 배경지도와 비교하고 속성 정보의 좌표계를 확인

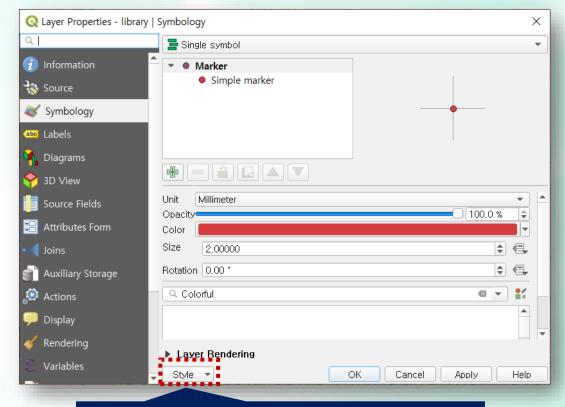


[정상적으로 변환이 되었다면 CCTV도 동일한 좌표계로 변환해보자]



레이어 심볼 변경 및 프로젝트 저장

- 레이어 심볼 변경
 - 레이어 속성 대화상자 → Symbology에서 색상, 심볼 아이콘 등을 변경

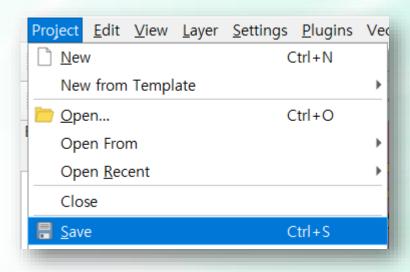


QML로 저장하면 심볼 재활용이 편리함

레이어 심볼 변경 및 프로젝트 저장

○ 프로젝트 저장/열기

- Project 메뉴 → 저장을 통해 원하는 경로에 .qgs 파일로 저장
- 다음에 QGIS를 실행하고 Project → Open을 통해 해당 qgs 파일을 열면 저장했던 순간의 레이어들을 포함한 작업 환경이 동일하게 열려짐
 (QGIS 실행 시 지도창 화면에 나타나는 프로젝트 목록에서 선택해도 됨)



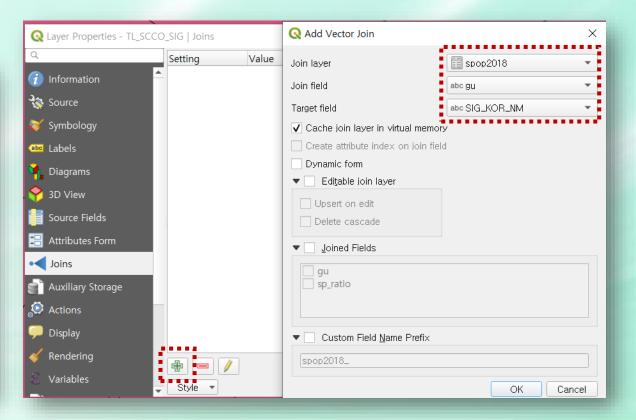
시군구 행정구역에 학령 인구 정보 결합

- 서울시 내에서도 학령인구 및 범죄율이 높은 구를 선택
- 학령 인구 정보를 검토
 - xls 또는 txt 등 통계정보로 제공. 맨 오른쪽 항목의 학령인구 구성비가 높은 순으로 검토
 - 데이터가 적어서 안해도 되지만 방법을 알기 위해 통계 데이터를 가공한 후, 공간데이터에
 속성 조인해보기 위해 헤더와 컬럼, 불필요한 행(row)을 정리하여 csv로 저장
 - 윈도우 탐색기에서 해당 CSV 파일을 QGIS 레이어창에 drag & drop으로 넣고
 속성 테이블을 열어 한글이 정상적으로 보이는지 확인
 - 문제가 있으면 CSV 속성 → Source 탭에서 인코딩을 변경
- 시군구 행정구역 속성창에서 Join 탭 선택 → + 버튼 클릭하여 학령인구 CSV를 선택하고, 학령인구 CSV의 구 명칭과 행정구역의 한글 구 명칭 컬럼을 조인 Key로 설정한 후 OK → 행정구역 속성창도 OK

시군구 행정구역에 학령 인구 정보 결합

4	Α	В	С	D	E	F	G	Н	1	
1	기간	지역	총인구	9세 -24	4세	0세-18	3세	학령인	! 구	
2		711 7	001	계	구성비	계	구성비	계	구성비	
3		합계	9,765,623	1,584,688	16.2	1,437,449	14.7	1,388,532	14.2	
4		양천구	464,185	90,970	19.6	83,574	18	84,261	18.2	
5		노원구	543,752	105,477	19.4	90,994	16.7	93,704	17.2	
6		서초구	433,951	77,484	17.9	79,380	18.3	74,818	17.2	
7		강남구	542,364	100,854	18.6	93,240	17.2	93,461	17.2	
8		송파구	666,635	111,542	16.7	109,541	16.4	102,100	15.3	
9		성북구	435,868	74,575	17.1	66,184	15.2	65,608	15.1	
10		강동구	427,573	69,455	16.2	66,371	15.5	63,046	14.7	
11		은평구	483,197	77,235	16	70,925	14.7	68,904	14.3	
12		도봉구	339,413	54,770	16.1	48,041	14.2	47,992	14.1	
13		강서구	596,949	91,032	15.3	90,328	15.1	81,492	13.7	
14		서대문구	310,313	50,458	16.3	43,313	14	42,267	13.6	
15	2018	마포구	375,077	59,152	15.8	54,840	14.6	50,937	13.6	
16	2010	광진구	355,559	58,208	16.4	48,510	13.6	48,143	13.5	
17		구로구	404,497	60,424	14.9	60,006	14.8	54,453	13.5	
18		종로구	153,065	24,475	16	19,179	12.5	20,289	13.3	
19		동작구	396,203	61,826	15.6	54,282	13.7	51,733	13.1	
20		강북구	319,164	48,470	15.2	40,762	12.8	41,348	13	
21		동대문구	348,052	54,504	15.7	45,602	13.1	44,991	12.9	
22		성동구	308,221	46,054	14.9	42,774	13.9	38,939	12.6	
23		중랑구	403,209	59,367	14.7	52,738	13.1	50,632	12.6	
24		용산구	228,999	32,494	14.2	29,557	12.9	28,164	12.3	
25		금천구	233,917	33,580	14.4	29,249	12.5	28,249	12.1	
26		영등포구	367,778	50,746	13.8	49,014	13.3	43,924	11.9	
27		관악구	501,957	75,313	15	54,900	10.9	55,767	11.1	
28		중구	125,725	16,223	12.9	14,145	11.3	13,310	10.6	



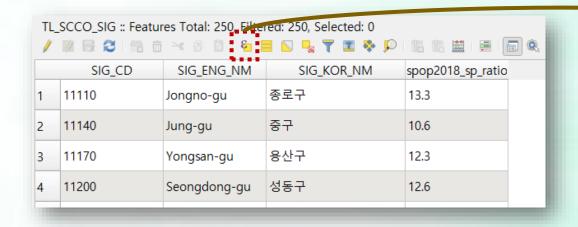


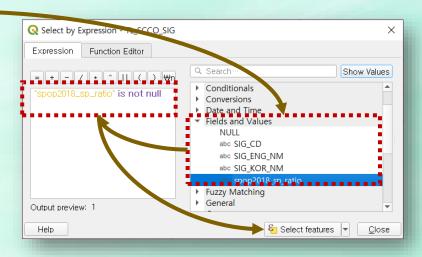
학령인구 행정구역 가공

- 행정구역 속성테이블을 열어 보면 학령인구 구성비 항목이 추가된 것을 볼 수 있음
- 구성비 값이 있는 서울시 시군구 행정구역만 별도의 학령인구 공간데이터로 활용하기 위해 속성 기반 선택 및 추출 작업을 진행
 - 속성테이블 창의 🦫 (Select by Expression)을 클릭하고, 오른쪽에서 Fields 및 Values를 확장한 후 원하는 항목을 더블클릭하면 쿼리문 화면에 항목명이 입력됨
 - 오른쪽의 2개 박스 중 하나와 같이 쿼리문을 입력하고 Select Feature 버튼을 누름

"spop2018_sp_ratio" is not null

"SIG_CD" like '11%'





학령인구 행정구역 가공

- 2개 쿼리문 별로 속성테이블 상의 선택 건수 차이를 보고, 레이어 컨텍스트 메뉴의 戶 Zoom to Selection 을 통해 지도 상에서 확인해보면 차이를 알 수 있음
 - 구 명칭으로 조인했기 때문에 다른 시도에 있는 중구 등이 잘못 중복 조인되어, "spop2018_sp_ratio" is not null 조건으로 선택시에는 31건이 선택됨
 - "SIG_CD" like '11%' 은 시도 코드에서 11로 시작하는 데이터만 선택하라는 의미로, 행정/법정구역 코드를 다뤄보신 분들이라면 시도(2) + 시군구(3) + 읍면동(3) 형태로 구성되는 것을 아실 것임
 - Code가 가장 정확하고, 그렇지 않으면 시도 명칭 항목을 결합해서 해야 하는 데
 여기에서는 두 데이터 모두 시군구명만 있음
 - 법정/행정동코드, PNU코드 구성: https://korn123.blog.me/30119120182
 - 행정표준코드관리시스템(법정동코드):

https://www.code.go.kr/stdcode/regCodeL.do

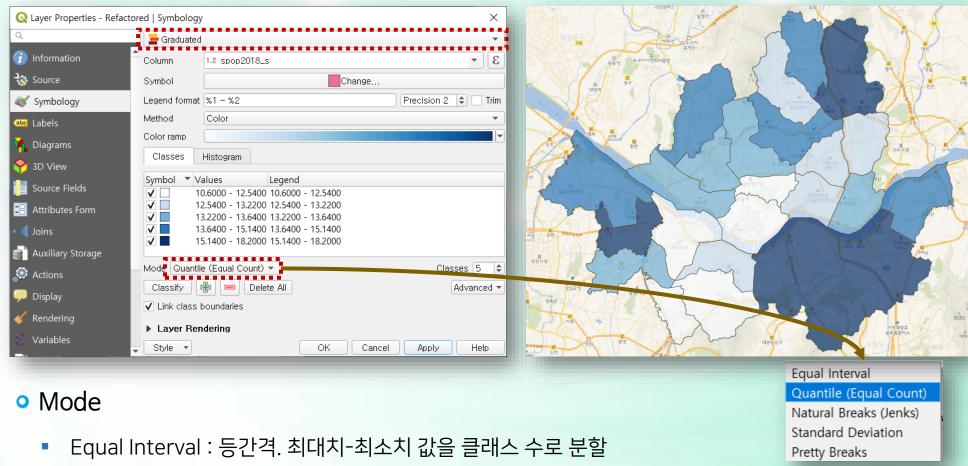
○ 서울시 시군구만 선택된 상태에서 다른 이름으로 저장



구별 학령인구 단계구분도(Choropleth map) 생성

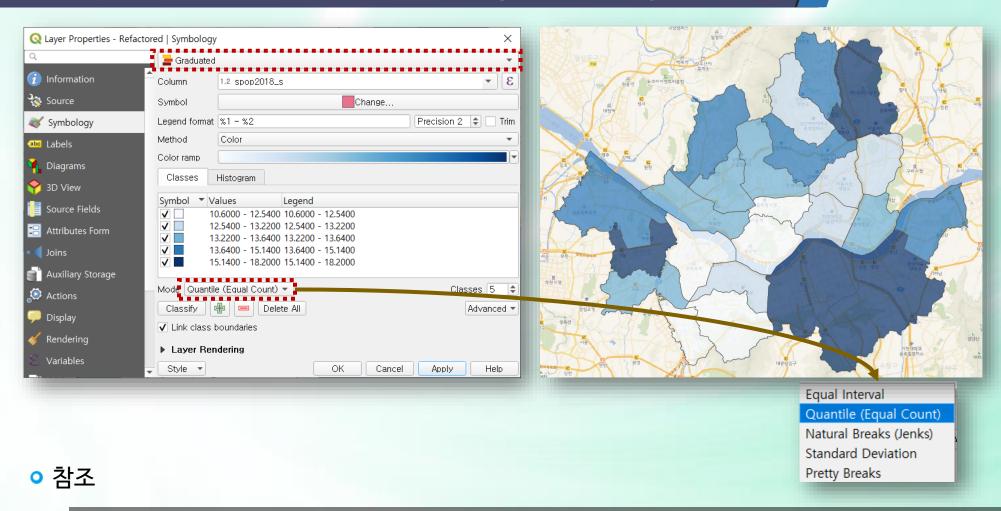
- 저장된 데이터는 조인된 학령인구 구성비가 문자형이어서 정량적인 단계구분도에 적용할 수 없음
- 앞서 했던 Refactor Field 또는 새로운 숫자형 컬럼에 Calculate 하는 방식으로 숫자형 데이터로 변환해야 (세부 내용은 QGIS Cookbook 등을 참조)
- 구별 학령인구 면형 데이터의 속성 → symbology → Graduated 선택 → 컬럼은 학령인구 구성비 컬럼 지정 → Color ramp 지정/변경
 - 양적) 단계구분도는 동일 계열 색상의 농도 차이 또는 크게 2개의 보색으로 설정하는 것이 좋음
- Mode는 데이터량이 적을 때에는 Natural Breaks, 많을 때에는 Quantile로 지정하고, 클래스 개수는 3개~5개 정도에서 지정하고 하단 렌더링에서 투명도 조정하고 OK

구별 학령인구 단계구분도(Choropleth map) 생성



- 정규분포일 경우, 중간 간격 쪽에 데이터가 뭉쳐버리는 등 데이터가 한 쪽에 치우쳐서 시각하 될 문제가 있음
- Quantile: 분위수. 각 클래스별로 동일한 데이터 개수가 적용.
- Natrural Breaks (Jenks): 유사한 값들로 묶으면서, 클래스 간의 차이를 최대회하는 방식으로 구분하기
 때문에 가장 구분이 잘되나 데이터량이 많을 경우 장시간이 소요됨

구별 학령인구 단계구분도(Choropleth map) 생성



- https://pro.arcgis.com/en/pro-app/help/mapping/layer-properties/data-classification-methods.htm
- http://www.qgistutorials.com/ko/docs/basic_vector_styling.html
- https://gisutd.tistory.com/7

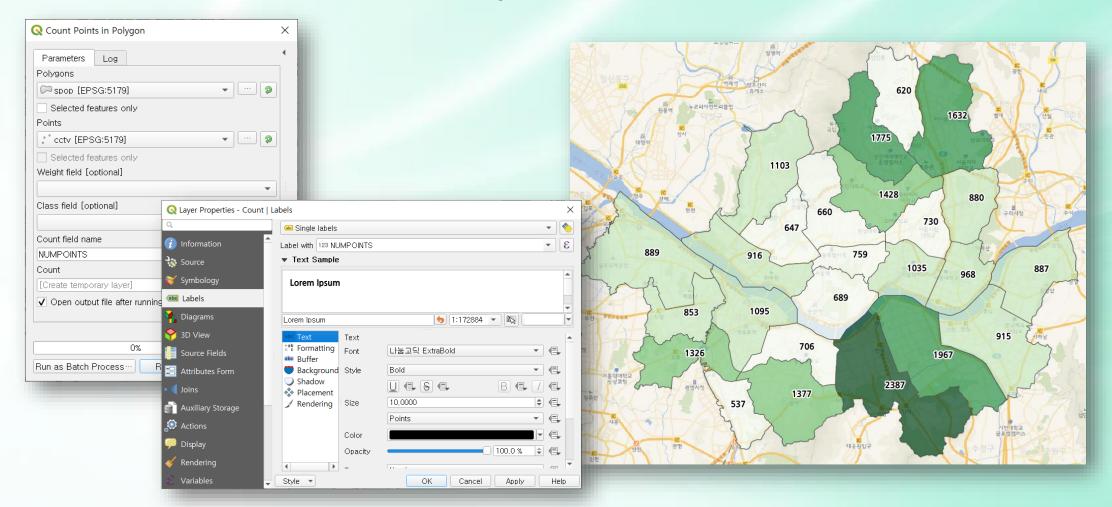
서울시 5대 범죄 발생 통계 공간데이터화

- 5대 범죄 역시 학령인구와 유사한 형태로 되어 있음
- 모두 주요 범죄이고 발생건수 자체만으로 위험하므로 구별 발생합계 건수만 이용하여 행정구역에 속성조인하고 숫자형 데이터로 변환함
 - 강남구 순으로 범죄발생률이 높은 것을 알 수 있음
 - 학령인구 구성비율과 범죄발생비율을 합산하여 단계구분도를 Natural Break로 표현해보면 강남구가 더 차별화되어 나타남을 알 수 있음

4	Α	В	С	D	4	Α	В
1	기간	자치구	합		1	gu_nm	crime_ratio
2	710		발생	검거	2	강남구	7.37
3		합계	101,948	74,487	3	영등포구	5.73
4		종로구	3,690	3,913	4	송파구	5.7
5		중구	4,030	2,679	5	마포구	5.07
6		용산구	3,411	2,543			4.93
7		성동구	2,457	1,787	6	관악구	
8		광진구	3,915	2,789	7	구로구	4.72
9		동대문구	3,680	2,968	8	서초구	4.64
10		중랑구	4,288	3,222	9	강서구	4.54
11		성북구	3,042	2,277		중랑구	4.21
12		강북구	3,437	2,743	11	중구	3.95
13		도봉구	2,249	1,676	12	노원구	3.93
14		노원구	4,007	2,725	13	광진구	3.84
15	2018	은평구	3,590	2,863	14	강동구	3.84
15 16	2010	서대문구	2,802	2,016		종로구	3.62
17		마포구	5,172	3,569		동대문구	3.61
18		양천구	3,515	2,459			3.52
19		강서구	4,629	3,469		양천구	3.45
20		구로구	4,810	3,338			3.45
21		금천구	3,293	2,567	19	강북구	
22		영등포구	5,840	3,922		용산구	3.35
23		동작구	3,100	2,090		금천구	3.23
24		관악구	5,026	3,653		동작구	3.04
25		서초구	4,726	3,183	23	성북구	2.98
26		강남구	7.513	5,196	24	서대문구	2.75
27		송파구	5,807	4,051	25	성동구	2.41
28		강동구	3,919	2,789	26	도봉구	2.21

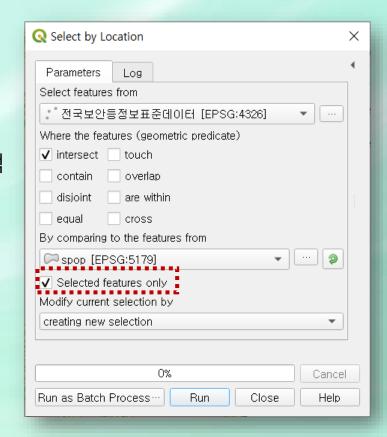
서울시 구별 CCTV건수 단계구분도 생성

- CCTV 포인트의 구별 건수를 집계하려면 툴박스의 Count Points in Polygon을 이용하면 됨
 - NUMPOINTS라는 컬럼에 합산치를 넣어주고, 이를 단계구분도로 표현하고,
 다시 Labels 메뉴에서 NUMPOINTS 컬럼으로 labeling 해줌(서초구가 가장 많고 구별 차이를 알 수 있음)



가로등/보안등 공간데이터화

- 강남구를 더 분석해보기로 함(가로등/보안등 정보를 조사)
 - 가로등은 없고 공공데이터포털에서 보안등정보를 찾았으나 주소만 있고,
 좌표 정보가 없음 (11,498건)
- 공공데이터포털의 "전국보안등정보표준데이터.csv"의 좌표 정보를 이용하여 QGIS에서 공간데이터화
 - 행정구역 속성테이블 또는 지도 상에서 🔜 (Select Feature)를 이용하여 강남구만 선택
 - Select by Location에서 아래 그림과 같이 선택 도형 개체만 적용하는 방식으로 강남구 도형과 공간적으로 교차(Intersect)하는 보안등 도형 개체를 선택
 - 1,571,535건 중 9,888건의 보안등이 선택됨. 선택된 보안등만 5179 좌표계 및 영문명 컬럼의 SHP으로 저장(1건은 속성상 다른 지자체로 확인되어 제외)
- 같은 방식으로 도서관, CCTV도 강남구 지역만 다른 이름으로 저장



도서관 주변 안전도 분석(1)

○ 데이터 탐색을 통해 다음과 같이 좀 더 세밀하게 대상을 좁히는 방향으로 검토

- 도서관: 도보 이동하므로 반경 500m로 한정. 아파트 단지 내에 있는 것은 제외 (전국 단위면 도로명주소 등의 단지 경계를 이용. 여기서는 수동 제외)
- CCTV: 세부 정보가 없으므로 반경 50m를 촬영 반경으로 설정
- 보안등: 관련 자료를 검토하여 반경 20m는 안전도가 높은 것으로 설정
- 위의 검토 결과 도서관 반경 내의 보행 경로인 실폭도로만 분석 공간으로 더 좁혀야 하고, 이 안에 중첩되는 CCTV와 보안등 반경을 검토하는 방향으로 구체화

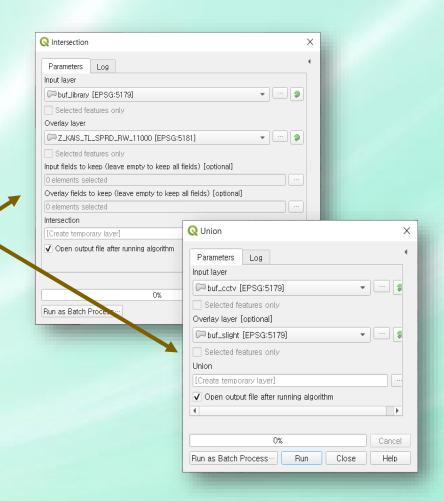
○ 실폭도로 데이터 다운로드 및 좌표계 적용

- 도로명주소 실폭도로 데이터를 활용
- 서울시 지역 데이터를 적용

도서관 주변 안전도 분석(1)

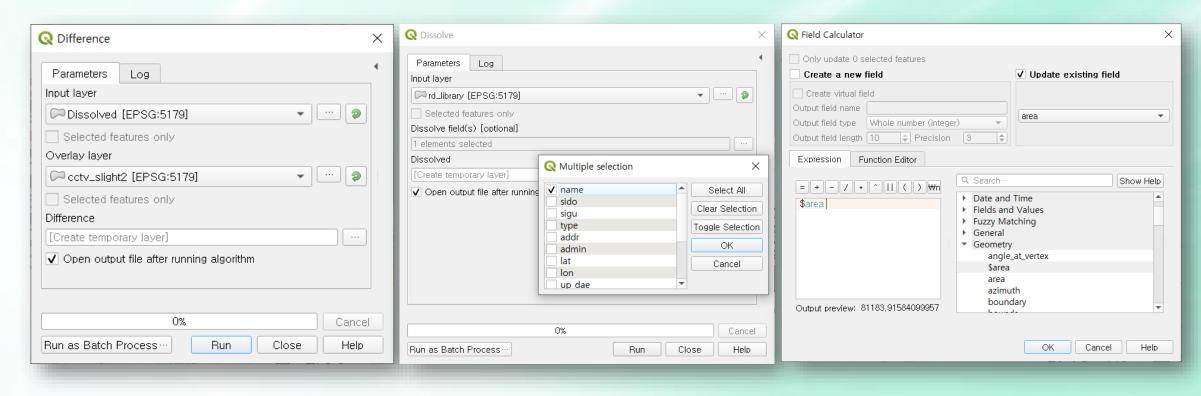
○ 버퍼 및 중첩

- 도서관, CCTV, 보안등을 각각 설정한 반경으로 Buffer 생성
- Intersection으로 도서관과 실폭도로를 중첩시켜 도서관 반경 500m내의 실폭도로만 추출
- Union을 이용하여 CCTV와 보안등 버퍼를 하나의 레이어로 합침 (안심영역1)
- 안심영역1에 임의의 공통값을 부여하고 Dissolve로 단일 레코드 데이터로 합침 (안심영역2)



도서관 주변 안전도 분석(2)

- 도서관 주변 도로의 비안심영역을 정량적으로 분석하기 위해
 - Difference를 이용하여 도서관 주변 도로에서 안심영역을 빼서 비안심영역 실폭도로를 생성
 - 도서관 주변 비안심영역 도로 면적을 알기 위해 도서관 명칭 컬럼을 이용하여 Dissolve
 - 새 컬럼을 만들고 Field Caculator에서 \$area로 도형의 면적을 속성정보화

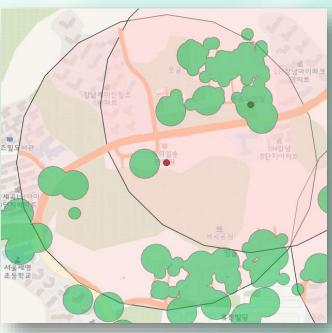


분석 결과 검토

- 단계구분도 및 면적 정렬을 하여 보면
 - 가장 비안심구역이 넓은 A도서관은 반경 내의 대로가 넓고, 동쪽의 자동차전용도로도 포함되었음
 - 반대로 비안심구역이 좁은 B도서관은 상대적으로 도로 면적이 작고 주도로인 4차선 왕복 도로 외에는 CCTV와 보안등이 촘촘하게 설치되어 있음



A도서관



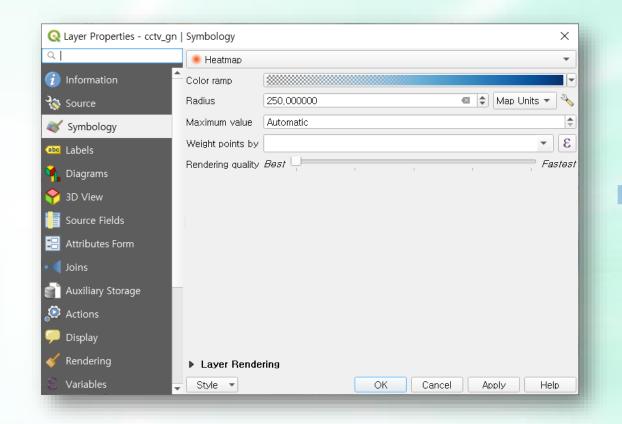
B도서관

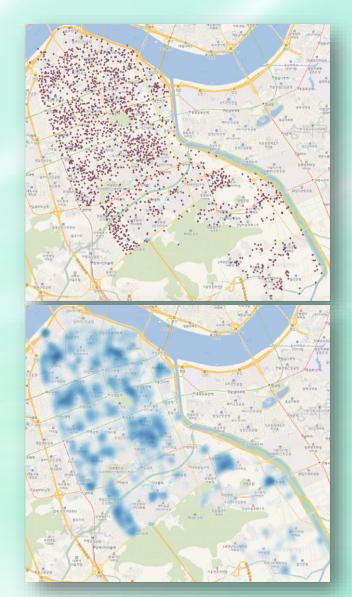
분석 결과 검토

- 따라서, 다음과 같은 내용들을 반영하여 재분석 및 추가 분석이 필요하고, 특히 선정된 지역에 대한 현장 조사도 필요함
 - 도로명주소 도로구간 등의 데이터를 이용하여 자동차전용도로, 4차선 이상 대로 등은 분석 대상에서 제외
 - 좀 더 정교한 기법이나 추가할 수 있는 데이터는 없는지 검토
 (예: B도서관 지역의 녹지(근린공원)/터널 등 추가 적용, 숙박/노래방 등의 데이터 적용 등)

Geographic Heat Map

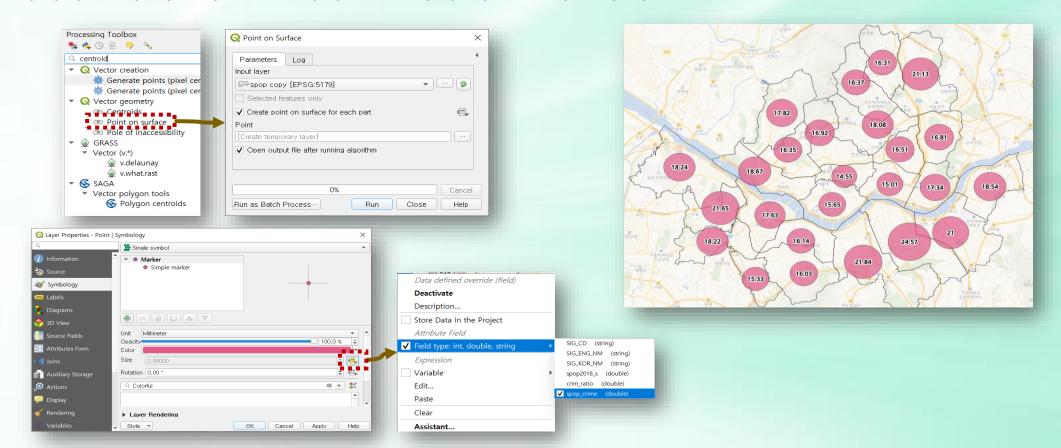
- CCTV 포인트 레이어의 심볼을 Heat map으로 바꾸면 공간적인 밀집도를 더 직관적으로 제공
 - Weight point에서 정량값을 가진 컬럼을 가중치로 줄 수 있음
 - 데이터가 너무 많을 경우 Rendering Quality를 Best로 하면 오래 걸리므로 다른 설정을 모두 적용한 후 마지막에 적용





Bubble Symbol Map

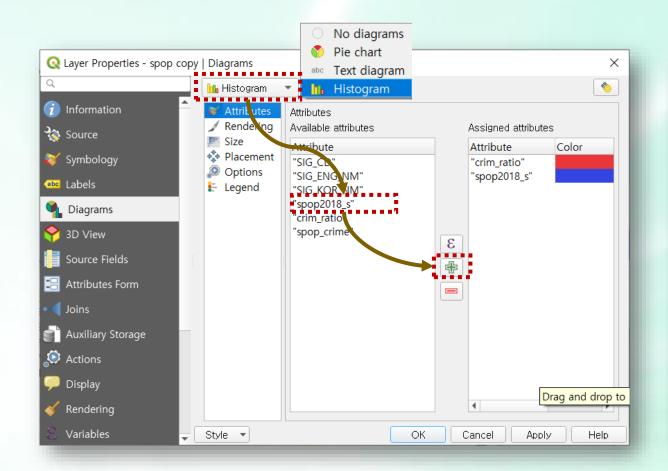
- 툴박스에서 "centroid"로 검색 → Point on surface를 실행하여 중심점 포인트 레이어를 생성
 - centroid는 도형의 모양에 따라 중심점이 도형 밖에 위치할 수 있음)
- 심볼 → 싱글 심볼에서 사이즈 옵션을 선택하여 정량값이 있는 항목을 지정하여 원의 크기를 다르게 표현
- 이해가 쉽게 되도록 Label이나 원 크기에 대한 범례를 추가

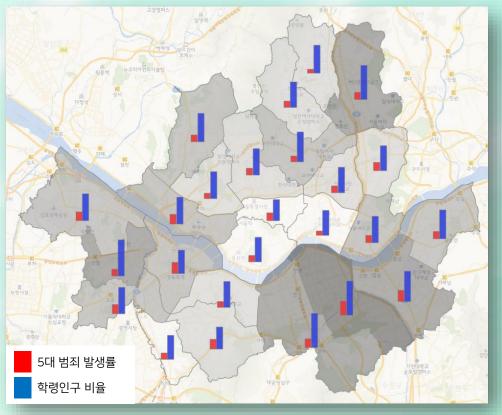


도형표현도

- 2개 이상의 정량값을 가진 레이어, 특히 행정구역 폴리곤은 도형표현도를 이용하여 다중 속성을 다이어그램으로 표현할 수 있음
 - 성비와 같이 동일한 성격의 데이터 특히, 합산치가 100이 되는 2개 이상의 데이터는 파이 차트
 - 다른 성격의 데이터는 히스토그램 등으로 표현
- QGIS 레이어 속성 → Diagram 탭에서 파이차트나 히스토그램을 선택하고, Attributes에서 정량값을 가진 항목을 추가
 - 속성 항목의 색상 및 차트의 사이즈, 위치 등을 조정한 후 OK하면 지도 상에 차트가 나타남
 - 참조: https://docs.qgis.org/3.4/ko/docs/user_manual/working_with_vector/vector_properties.html#id116

도형표현도





QGIS 공간분석 기능 활용 방식

- 플러그인 활용
 - NN-JOIN
- 그래픽 모델러
- 파이썬 스크립트 활용