

SGIS Small-Area Statistics User Manual



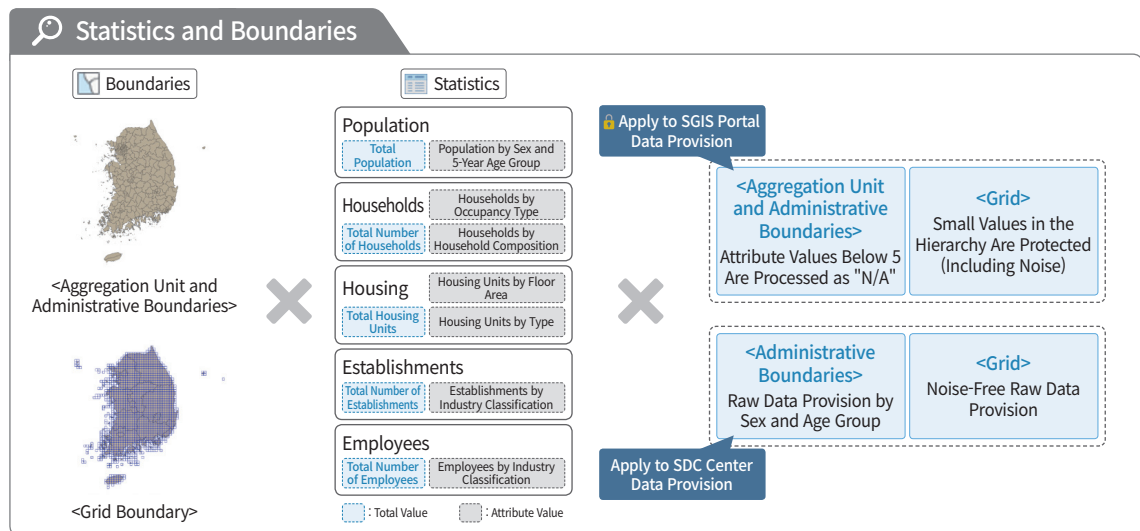


Overview of SGIS Small-Area Statistics

1 Guide to Using Small-Area Statistics

Small-area time-series statistics, as well as boundaries for aggregation units, administrative districts and grids can be utilized in file format to create various thematic maps in the Statistical Geographic Information Service (SGIS).

- Channels Offered: SGIS Portal (sgis.kostat.go.kr), Statistical Data Center (data.kostat.go.kr)
- Data Formats: Statistical Data (.txt), Boundary Data (.shp)



2 Introduction to Statistics and Boundaries

Target Statistics: Statistics on aggregated census data from Statistics Korea of small areas (administrative district, aggregation unit and grid) are provided, including total values and attribute values across five categories.

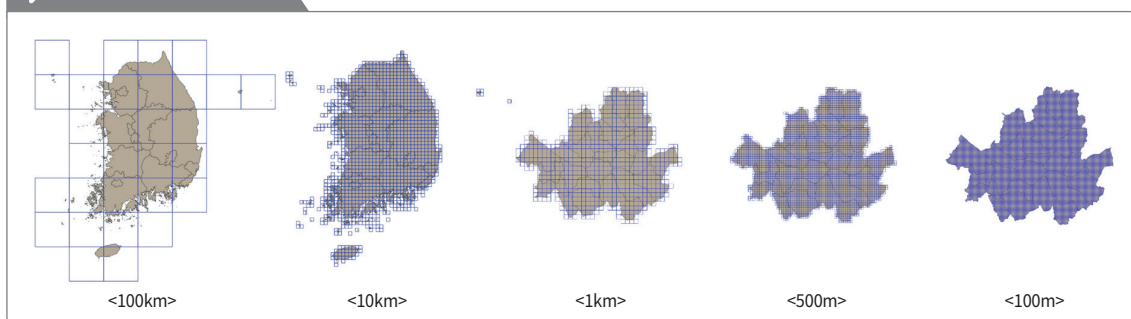
- Population, Household and Housing: Population and Housing Census (Register-Based Census)
- Establishment and Employment: Economic Census (National Establishment Census)

| Provided Items | Small-Area Data by Target Statistics

Population	Households	Housing	Establishments	Employees
Total Population Population by sex and age. Population by sex and age group. Population by educational attainment. Population by sex and marital status.	Total Households Households by household composition. Households by occupancy type.	Total Housing Units Housing units by floor area. Housing units by type. Housing units by year of construction.	Total Establishments Number of establishments by major industry classification. Number of establishments by medium industry classification.	Total Employees Number of employees by major industry classification. Number of employees by medium industry classification.

I Overview of SGIS Small-Area Statistics

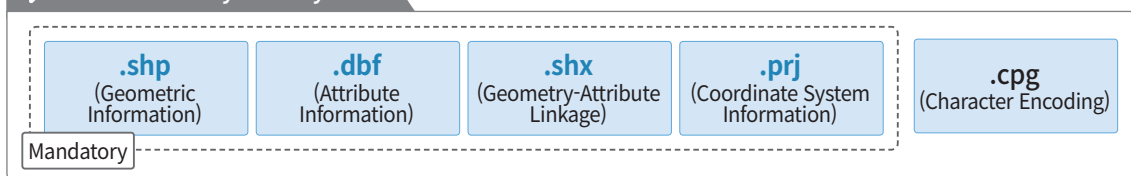
Grid Boundaries



Boundary File Format

- A boundary file consists of five or more files with different extensions, including .shp (shape), .dbf (attributes) and .prj (coordinate system).

SHP Boundary File Layout



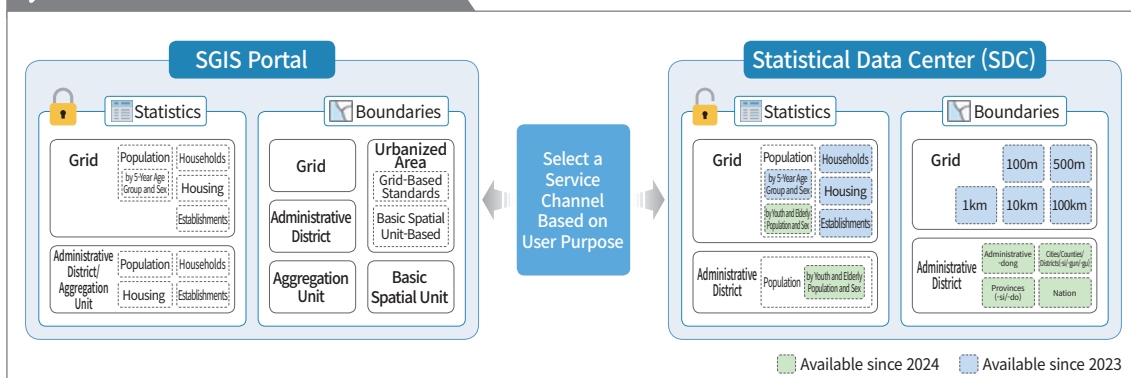
- SGIS boundary files (aggregation units, administrative districts and grids) use the EPSG:5179 coordinate system.

3 Service Scope and Confidentiality Protection

- **Service Scope:** The SGIS platform offers a dual-service approach through the SGIS portal and the Statistical Data Center (SDC).

- SGIS portal provides data processed using confidentiality protection techniques (BSCA or N/A). Users can utilize the analysis environment provided by the SDC for true value data without confidentiality protection.

Data Provision Service Channels



Comparison of Data Provided by SGIS Portal and SDC

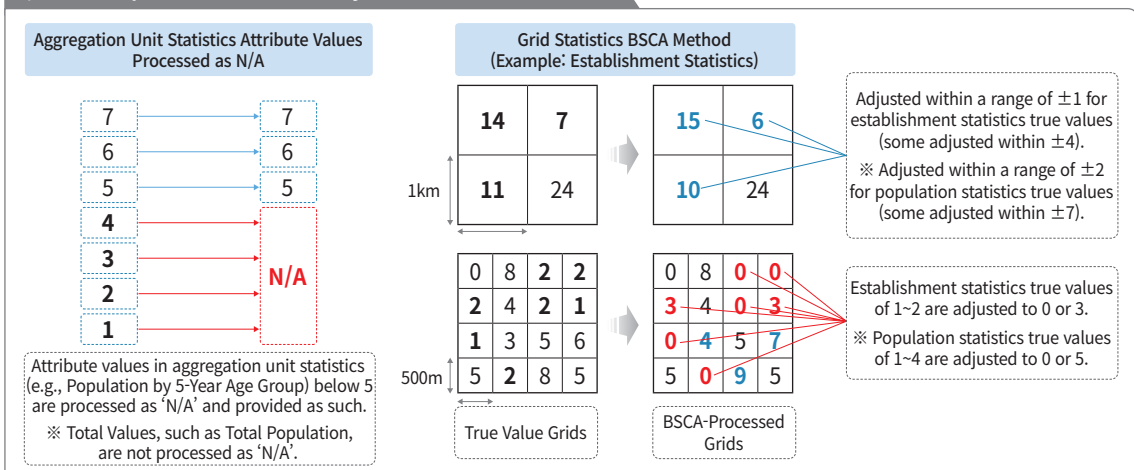
Categories		Provided Items	SGIS Portal(Confidentiality Protection Applied)				SDC(Confidentiality Protection Not Applied)			
			100,500m	1,10,100km	Administrative District	Aggregation Unit	100,500m	1,10,100km	Administrative District	Aggregation Unit
Total	Population	Total Population	○	○	○	○	△*	○	×	×
	Households	Total Households	○	○	○	○	△*	○	×	×
	Housing	Total Housing Units	○	○	○	○	△*	○	×	×
	Establishments	Total Establishments	○	○	○	○	○	○	×	×
	Employees	Total Employees	○	○	○	○	○	○	×	×
Attributes	Population	Population by Sex and 5-Year-Age	×	○	○	○	×	○	×	×
		Population by Sex and Age Group	×	×	×	×	×	○	○	×
	Households	Households by Household Composition	×	○	○	○	×	○	×	×
	Housing	Housing Units by Year of Construction	×	○	○	○	×	○	×	×
		Housing Units by Floor Area	×	○	○	○	×	○	×	×
		Housing Units by Type	×	○	○	○	×	○	×	×
	Establishments	Establishments by Industry Classification	×	○	○	○	○	○	×	×
	Employees	Employees by Industry Classification	×	○	○	○	○	○	×	×

* SDC raw data is available only for total population, household and housing data at the 100m and 500m grid levels for the Administrative district ended with '-dong'(△)

Confidentiality Protection Methods

- In statistics for aggregation unit and administrative district, for statistical attribute values below 5 in aggregation unit or administrative boundary data, the values provided are replaced with the string "N/A" to protect respondent confidentiality. However, total statistics are not subject to "N/A" processing.
- For grid statistical data, values below 5 ("small value": 5 for population category, 3 for establishment category) are probabilistically replaced with either 0 or 5 (3 for the establishment category). Values of 5 or more are adjusted to include noise with a maximum range of ± 7 (for population category, ± 4 for establishment category), making it difficult to estimate "small values."

Example of Confidentiality Protection Methods



II

Application Process

1 How to Apply for SGIS Portal Data

- ① **Subscription:** Register as an SGIS member. After signing up, access small-area statistics and boundaries for free via the “Data Provision” menu.

SGIS Portal ▶ Data Provision Menu

SGIS ^{plus} 통계지리정보서비스	통계주제도	대화형 통계지도	활용서비스	분석지도	Data Provision	알림마당	검색 +
	인구와 가구 주거와 교통 복지와 문화 노동과 경제 건강과 안전 환경과 기후	총조사 주요지표 인구주택총조사 농림어업총조사 전국사업체조사 e-지방지표 공공데이터 나의 데이터	생활권역 통계지도 일자리 맵 정책통계지도 살고싶은 우리동네 업종통계지도 지역현안 소통지도 통계지도체험 My통계로 통계갤러리	자연재해 통계지도 기업생태 분석지도 지역변화 분석지도 도시화 분석지도 행정통계 시각화지도 총조사 시각화지도 월간통계 인구피라미드 고령화 현황보기 성씨분포 지방의 변화보기	Introduction to Data Provision Data Provision List Data Request Requested Data Download Request History Register Usage Results	SGIS플러스 소개 설명과 공지 자료 시점 현황 질문과 개선요청	

- ② **Basic Information:** Before applying for data, input basic information such as affiliation, contact details and the purpose of the request.

Enter Basic Information

Basic Information

*는 필수 입력 항목입니다.

성명(신청인)	김통계		
Affiliation *	Central Administrative Agency		Statistics Korea
Contact Information *	010	- 1234	- 5678
Email Address *	geosta7	@ naver.com	naver.com
Purpose of Request *	연구수행		
Purpose of Use *	Utilizing subregional statistics (grid, aggregation units) and administrative boundaries(provinces, cities/counties/districts) to analyze the optimal location for public service facilities for single-person households in Daejeon. ※ Please provide a detailed description of at least 100 characters.		
Project Title *	Analysis of Public Service Facility Locations for Single-Person Households. Example: "Conducting an empirical analysis of the correlation between urban design and traffic safety."		

- ③ **Data Application:** Agree to submit usage results → Use filters to select data → Add and complete the application.

Apply for Data

Consent for Submission and Data Provision Period (Mandatory)

Consent for Submission	I agree to indicate the source of the provided data and to submit one copy of the completed output (e.g., paper, URL) to Statistics Korea. ※ Applications cannot be submitted without consent. ※ The submitted output will be privately stored in the Register Usage Results menu.
Data Provision Period	※ Data downloads will only be available for one week after approval.

☒ I agree to the above conditions. [▶ Register Usage Results](#)

※ Users agree to indicate the source of the data and submit results when utilizing the data, and data downloads are available for one week.

Data Selection (Mandatory)

Data Type	☒ 집계구 ☐ 격자	① Select Aggregation Unit or Grid
Data Category	☒ 통계자료 ☐ 통계지역경계 ☐ 세종시 특별센서스	② Select Statistics or Boundary
Data Subject (by Aggregation Unit)	☒ 집계구별 통계(인구) ☐ 집계구별 통계(가구) ☐ 집계구별 통계(주택) ☐ 집계구별 통계(사업체) ☐ 행정구역별 통계(인구) ☐ 행정구역별 통계(가구) ☐ 행정구역별 통계(주택) ☐ 행정구역별 통계(사업체)	③ Select the applicable statistics
Aggregation Unit Base Year	2023년 2분기 [2023년 7월 1일 경계기준]	④ Select aggregation units based on the latest boundary
Reference	2022	⑤ Choose the reference year for the statistical data
Detailed Data Name	☒ 전체 ☒ 성연령별인구 ☒ 인구총괄(총인구) ☒ 인구총괄(평균나이) ☒ 인구총괄(인구밀도) ☒ 인구총괄(노령화지수) ☒ 인구총괄(노년부양비) ☒ 인구총괄(유년부양비)	⑥ Select statistical data
Province/City, County, District (-si, -do/ -si, -gun, -gu)	대전 전체	⑦ Select the region

Add

※ Applications are automatically approved 10 minutes after submission and users can access the data via the "Download Requested Data" option under the "Data Provision" tab.

II Application Process

- ④ **Register Usage Results:** Go to the “Data Provision” menu, select “Register Usage Results”, enter details such as purpose of use, content and suggestions for improvement.

Register Usage Results

Title *
대전시 맞춤형 복지서비스를 위한 1인가구 분석
Research Title

User Name *
김통계
Basic Information

Affiliation *
중앙행정기관
통계청

Contact Information *
010 - 1234 - 5678

Email Address *
geosta7
@gmail.com
naver.com

Purpose of Use *
1인가구 공공서비스시설 입지분석에 활용
Briefly Enter the Purpose of Use

Data Provided (by Statistics Korea) *
<Statistics>
Select Usage Data (Multiple Selections Allowed)

☒ 집계구
☐ 행정구역
☒ 격자

<Boundary>

☐ 집계구
☒ 행정구역
☐ 도시권
☒ 격자
☐ 기초단위구
☐ 도시화지역(격자 기준)
☐ 도시화지역(기초단위구 기준)

Data Provided (by other organization)
통계청

Details of Use *
소지역(격자-집계구) 통계와 행정구역(시도·시군구) 경계를 사용하여 대전시 1인가구 공공서비스시설 입지분석에 활용
Enter Detailed Usage Information

Suggestions for Improvement
더 세분화된 1인가구를 제공해주면 좋겠습니다.(현재:총 1인 가구수 → 개선: 남·여 각각 1인가구 수)
Suggestions for Improvement (including ideas for service improvements)
※ 자료이용 시 불편사항이나 개선 아이디어를 작성해주시면 서비스 개선에 충실히 반영하겠습니다.

Attachment
Add Attachment
Maximum file size is 20MB.
Detailed Suggestion Should Be Made via Attachment

Register

※ The information entered is used solely for service improvement and performance management and will not be disclosed.

8

2 How to Use the Statistical Data Center (SDC)

- ① **Pre-Application:** Before visiting the center, access <https://data.kostat.go.kr> and submit a application for the data you wish to analyze.

SDC Portal Screen

The SDC Portal Screen features a top navigation bar with the SDC logo and a 'Data Center' button. Below the bar, there are several columns of links categorized under '센터 소개' (Center Introduction), '서비스 이용' (Service Usage), '소통마당' (Communication Plaza), '빅데이터 활용' (Big Data Utilization), '활용대회' (Utilization Contest), and 'MY 서비스' (My Service). A central banner highlights the SDC's role as a new service for statistical data analysis. At the bottom, five circular icons represent different service areas: '서비스 이용안내' (Service Usage Guide), 'SDC 제공데이터 안내' (SDC Provided Data Guide), 'SDC 분석센터 이용신청' (SDC Analysis Center Application), 'SDC 주문형 분석서비스' (SDC Custom Analysis Service), and '이용자 교육' (User Education).

- ② **Application and Compliance:** Review the procedures for using the SDC, agree to the Security Pledge and accept the User Compliance Terms.
- ③ **Data Search:** Select the desired data types, such as grid statistics, administrative statistics or private data, along with the corresponding years.

Select Data

The 'Select Data' interface is divided into two main sections. On the left, a '자료검색' (Data Search) sidebar lists various data categories: Administrative Statistical Data, Private Sector Data, Microdata, Pre-negotiated Data, Combined Data, and Grid Statistics. The 'Grid Statistics' section is expanded, showing a list of years from 2010 to 2022, with 2022, 2021, 2020, and 2015 selected. On the right, the '검색자료' (Search Results) section displays a table of search criteria. The table has columns for '번호' (Number), '형태' (Type), and '항목' (Item). The first table lists four items: 1. Grid Statistics, 2. Grid Statistics, 3. Grid Statistics, and 4. Grid Statistics. The second table lists four items: 1. Grid Statistics, 2. Grid Statistics, 3. Grid Statistics, and 4. Grid Statistics. Below the tables, a '선택자료' (Selected Data) section shows the final search results, including 'Grid Statistics > Grid Statistics > Grid Statistics (인구) > 2022', 'Grid Statistics > Grid Statistics > Grid Statistics (인구) > 2020', and 'Grid Statistics > Grid Statistics > Grid Statistics (인구) > 2015'. A '다음' (Next) button is located at the bottom right.

II Application Process

- ④ **Enter Application Information:** Provide details such as affiliation, user and co-researcher information and the purpose of data use, then submit the application.

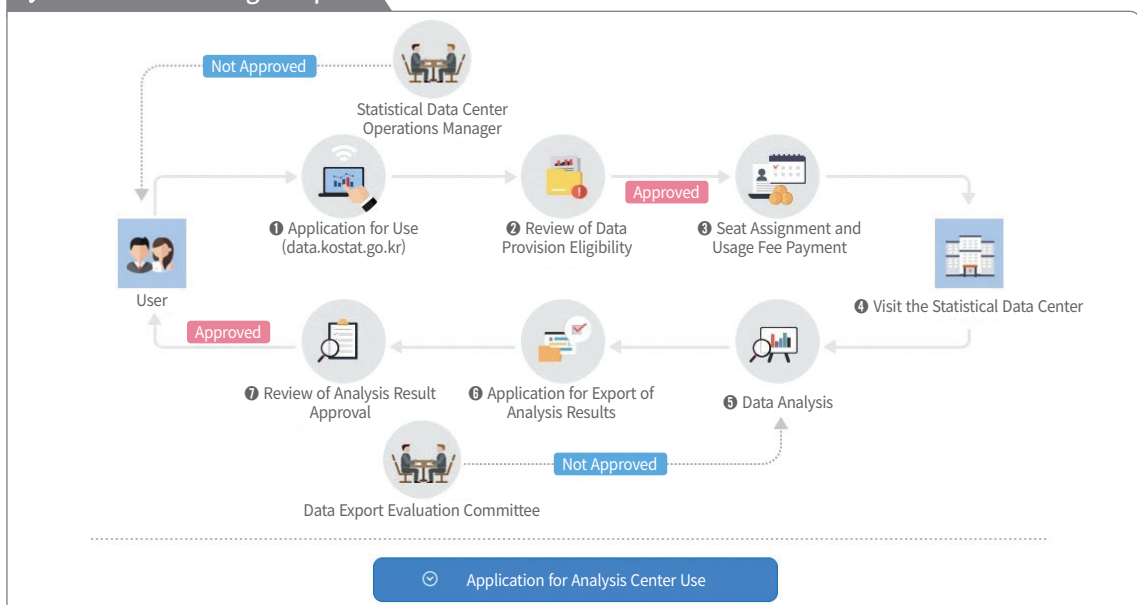
Entering Application Information

Application Information Input

Affiliation *	Government Agency 등록형 • (e.g., organization name/school name), please enter as "Individual."	Position *	Government Employee 검통계 • (e.g., organization name/school name), please enter as "Individual."
User Information *	General User		
Content and Scope of Data *	격자통계 > 격자통계 > 격자통계 (인구) > 2022 격자통계 > 격자통계 > 격자통계 (인구) > 2020		
Delimiter Type *	Comma Use First Row ☒		
Co-researcher	Q Co-researcher Search		
Co-researcher Confidentiality Agreement	☒ 파일 크기 10 Files / 100.0 MB / 100.0 MB(A) 0 Byte / 0 File Download Agreement Form		
Usage Location *	대전 Number of Assigned Seats	1 SDC Available Seat Status	
Analysis Packages for Use *	☐ SAS ☐ SPSS ☐ STATA ☒ R ☐ Python ☐ MS Office ☐ SGIS ☒ QGIS ☐ ArcGIS		
Purpose of Data Use *	선택하세요. 선택하세요. • Enter the purpose of using the results obtained through the analysis center services (this will be considered for the approval of result extraction). • Select "Approved Statistical Compilation" for statistics under approval. • If "Approved Statistical Compilation" is selected, you must enter the approved statistical number and name (e.g., No. 000, Survey 000).		

- ⑤ **Visiting the Center:** Analysis is possible after visiting the SDC following usage approval. Analysis results such as summary tables and graphs are provided after export approval.

SDC Center Usage Steps





Frequently Asked Questions

Q1 What is the update schedule for statistical and boundary data?

- The data is updated twice a year due to differences in publication timelines for the Population and Housing Census and the Economic Census. The detailed schedule is as follows:

	Update Period*	Key Update Details	
		Statistical Data	Boundary Data
1st half	June of Year Y	Data from the Year (Y-2) Economic Census	
2nd half	November of Year Y	Data from the Year (Y-1) Population and Housing Census (Register-Based Census)	Based on Q2 of Year Y.
			Based on Q4 of Year (Y-1).

※ The specific update periods may vary slightly depending on data processing schedules.

Q2 What are the differences between the codes and boundaries for administrative district, administrative -dong, and legal -dong?

- Administrative boundaries: Boundaries codified by Statistics Korea for the purpose of compiling national statistics.
- Administrative -dong boundaries: Boundaries established for administrative convenience, reflecting administrative districts.
- Legal -dong boundaries: Boundaries designated by law, using traditional regional names.

- All data provided by Statistics Korea adheres to the Korean Standard Classification of Administrative Districts. If you need to match codes between administrative -dong and legal -dong, please refer to: “Statistical Classification Portal ► Social Classification ► General Classification ► Korean Standard Classification of Administrative Districts ► Crosswalk Database by Reference Timepoint.”

Q3 Where can I download boundaries based on past timepoints?

- You can download boundaries and statistics based on past timepoints by going to the 「Data Provision」 menu ► Data Provision List, and scrolling down to “※ Application for Past Aggregation Unit Data.”

Q4 Which programs can be used with the data, and how can they be utilized?

- Users of the Data Provision service often utilize statistical packages such as R and SAS, as well as spatial data analysis tools such as QGIS and ArcGIS.

Q5 I downloaded the data, but what do the column names and items represent?

- To understand the meaning of each item in the provided statistical and boundary data, you should refer to the relevant Codebook.

- The Codebook for downloaded data is available in the 「Data Provision」 menu ► Download Requested Data [코드표 및 이용설명서 >](#)

III Frequently Asked Questions

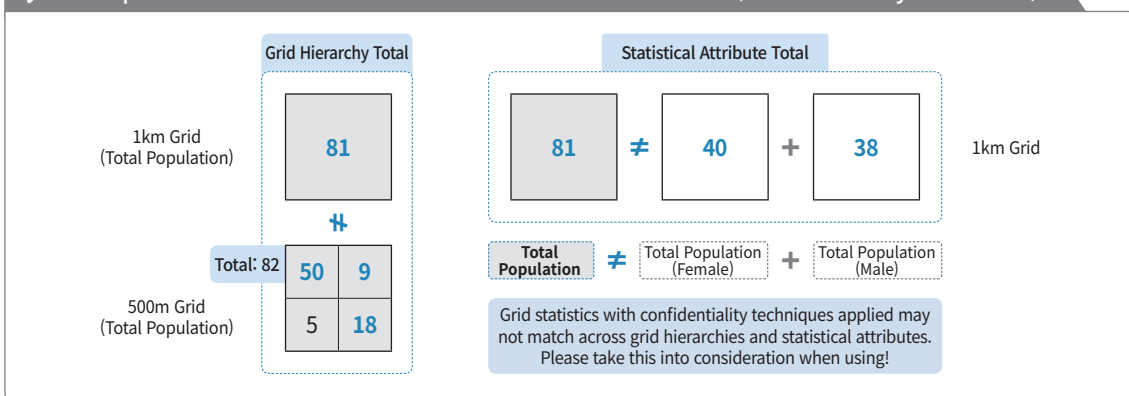
Q6 Why are some grid boundaries missing when merging grid boundaries with statistical data?

- Grid boundaries are created to cover the entire national territory. However, grid statistics are not generated for grids without population or establishments. Additionally, if there is no statistical value corresponding to a specific statistical attribute, statistical data is not generated for that grid. As a result, some grid boundaries may appear missing when merging grid boundaries with statistical data.

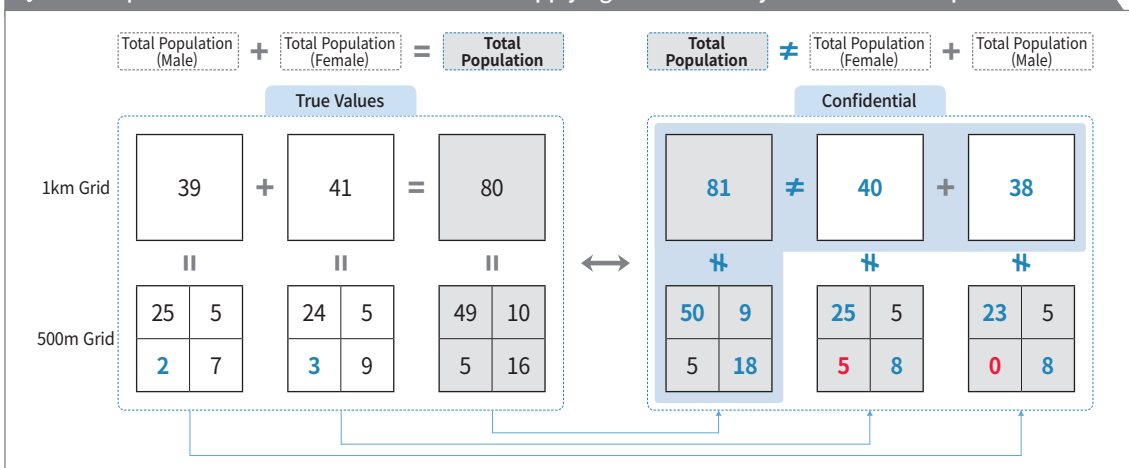
Q7 Why does the sum of attribute values not match the total value?

- Aggregation unit and administrative district statistics provided by the SGIS Portal apply confidentiality protection techniques (e.g., N/A processing). Consequently, the sum of attribute values may be smaller than the total value.
- For grid statistics, a suitable level of noise is applied across all grids (BSCA method), which can result in differences when comparing grid hierarchy subtotals or comparing attribute values with total values. For grid statistics without confidentiality protection, please use them in Statistical Data Center's secure environment.

Comparison of Grid Statistics Subtotals and Attribute Totals (Confidentiality Protection)



Examples of Grid Statistics Before and After Applying Confidentiality Protection for Population Data

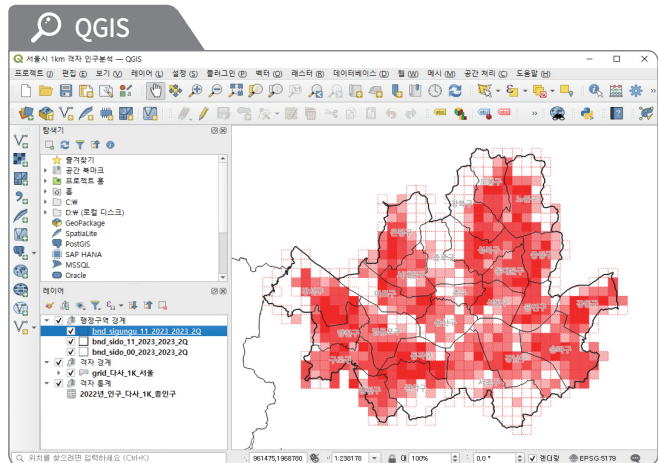


IV

Map Visualization Tools

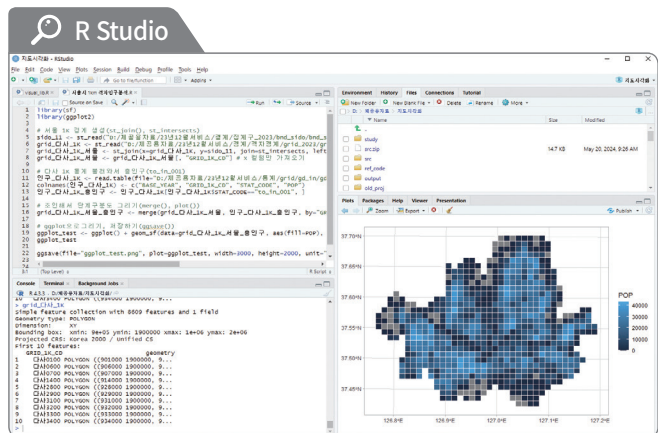
1 QGIS

QGIS allows the processing and integration of spatial data in various formats, such as boundaries in .shp and statistics in .csv or .txt to create diverse thematic maps. A Popular spatial analysis program with an interactive UI for user convenience. Official website: qgis.org



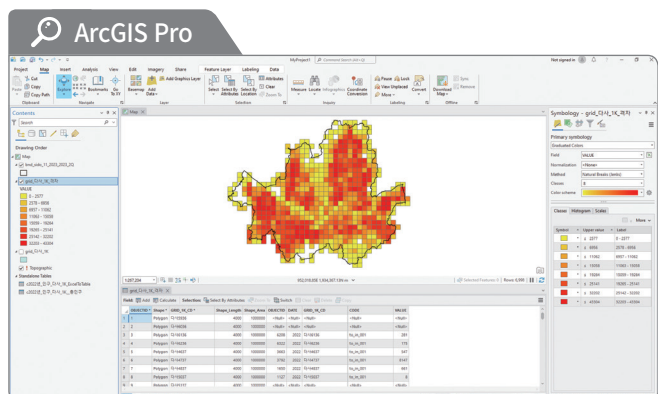
2 R

R is a statistical analysis program that combined with the integrated development environment R Studio, spatial data processing and visualization packages such as sf and ggplot2, which enable map visualization. Official website: posit.co/download/rstudio-desktop



3 ArcGIS

ArcGIS Pro is a spatial statistics analysis program used for spatial data mapping and map visualization. A commercial spatial analysis program developed by ESRI. Official website: esri.com





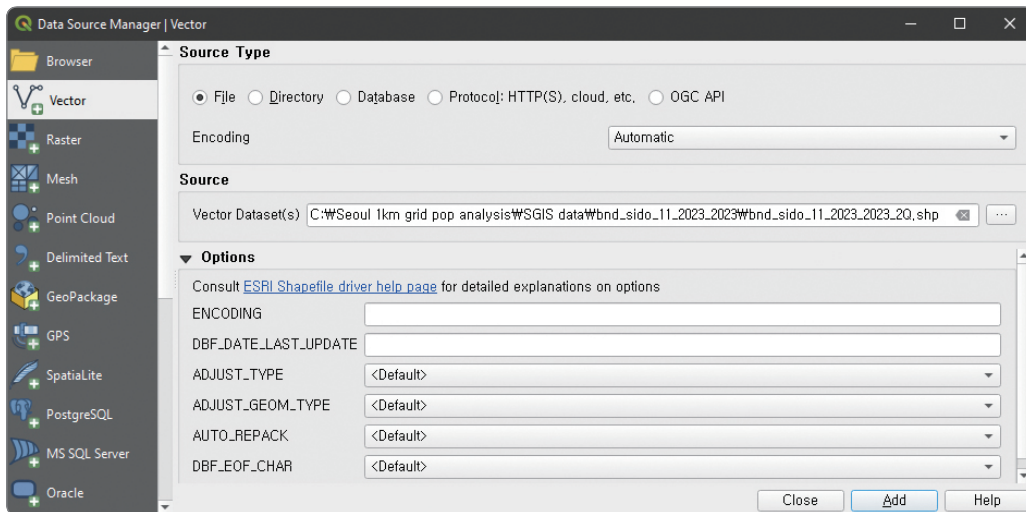
QGIS Analysis Example

- 1km Grid Population (Seoul) -

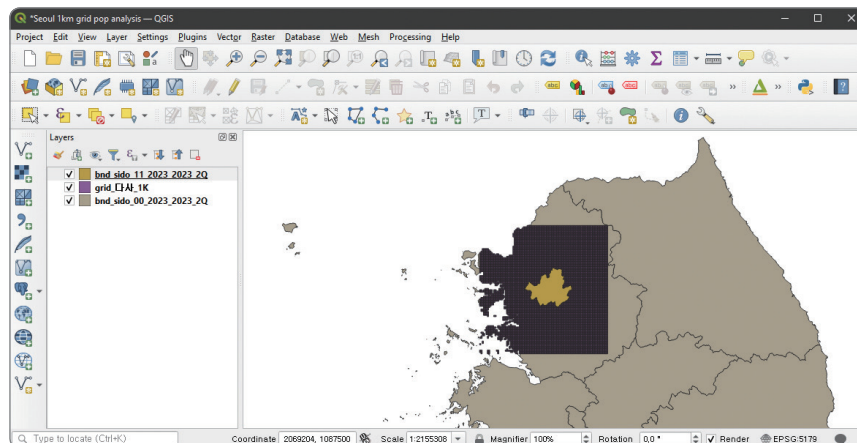
This example explains the step-by-step process of analyzing the **Seoul Boundary**, **National SIDO Boundaries**, **1km-Grid Boundaries** and **2022 1km-Grid Population Statistics** downloaded from the SGIS Data Provision menu using the latest version of QGIS (3.34.8).

① Adding Boundary Layers (.shp)

- Go to **Layer** ► **Add Layer** ► **Add Vector Layer** to add **National SIDO Boundaries**, **1km-Grid Boundaries**, **Seoul Boundary**.

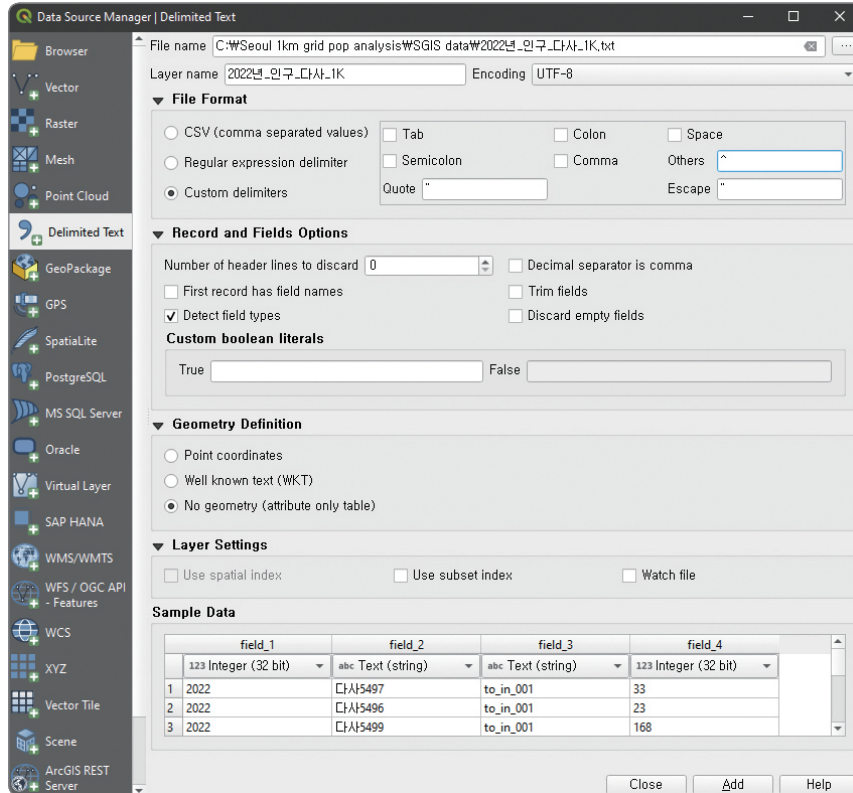


- The map screen below displays the added **National SIDO Boundaries**, **1km-Grid Boundaries** and **Seoul Boundary**.



2 Adding Statistical Layers (.txt)

- Go to **Layer** ► **Add Layer** ► **Add Delimited Text Layer** to add **2022년_인구_다사_1K.txt**
- ※ Go to **File Format** ► **Custom Delimiters** ► **Others** to set to ‘^’ as delimiter.
- ※ If a CSV file is downloaded, select **File Format** ► **CSV (comma separated values)**.



- Right-click on the statistical layer (**2022년_인구_다사_1K.txt**) and select **Open Attribute Table** to view the 1km-grid population data.

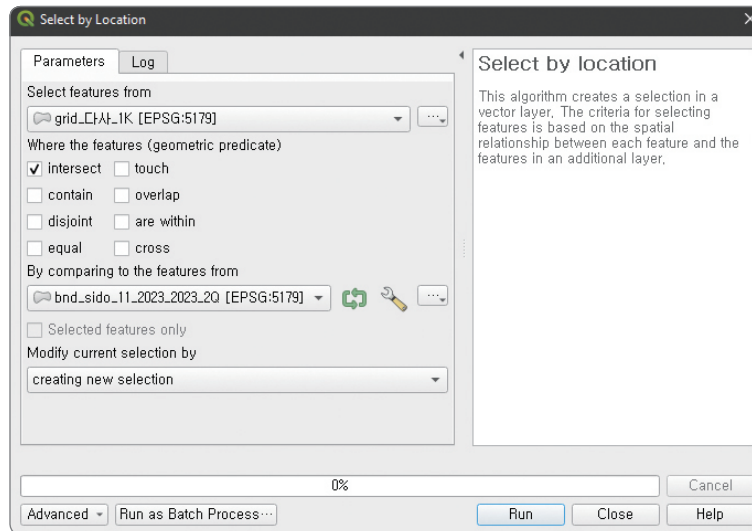
	field_1	field_2	field_3	field_4
1	2022	다사5497	to_in_001	33
2	2022	다사5496	to_in_001	23
3	2022	다사5499	to_in_001	168
4	2022	다사5498	to_in_001	48
5	2022	다사5492	to_in_001	48
6	2022	다사5495	to_in_001	5
7	2022	다사5494	to_in_001	18
8	2022	다사5479	to_in_001	1520

V QGIS Analysis Example - 1km Grid Population (Seoul)

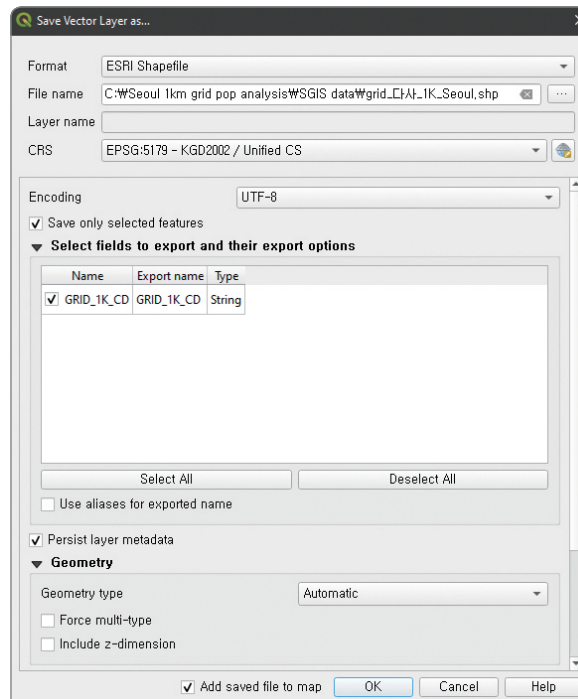
③ Selecting Intersect Areas Between DASA 1km-Grid Boundaries and Seoul Boundary

- Select **grid_다사_1K** layer, then go to **Vector** ► **Research Tools** ► **Select by Location** screen. Select intersect and the Seoul boundary layer, then click Run to execute.

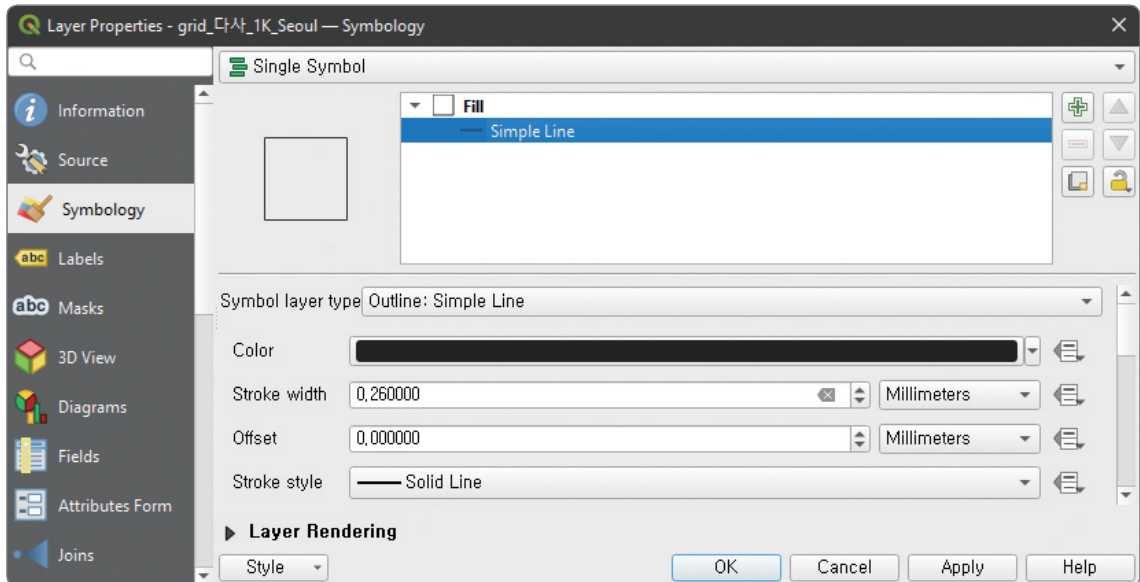
※ The selected grids in the **grid_다사_1K** layer will be highlighted in yellow.



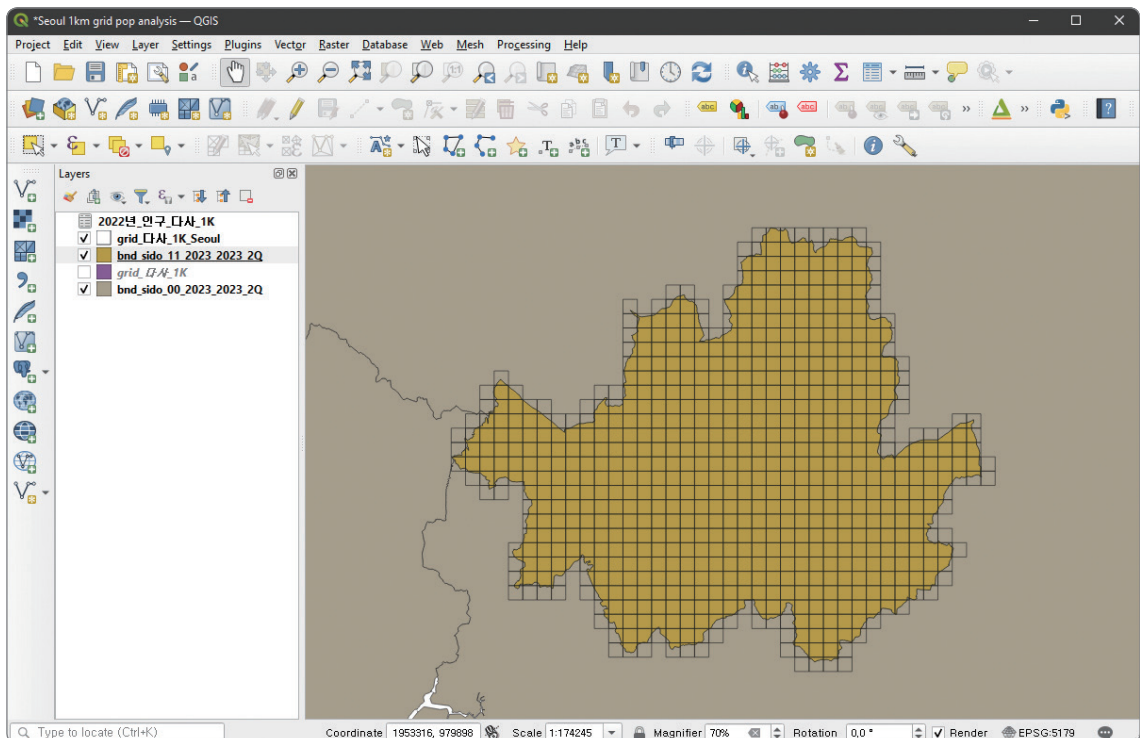
- Right-click on the **grid_다사_1K** layer, then select **Export** ► **Save Selected Features As** (**grid_다사_1K_서울시.shp**).



- Right-click on the `grid_다사_1K_서울시` layer, then select **Properties** ► **Symbology**. Change the symbol type to **Outline: Simple Line**.



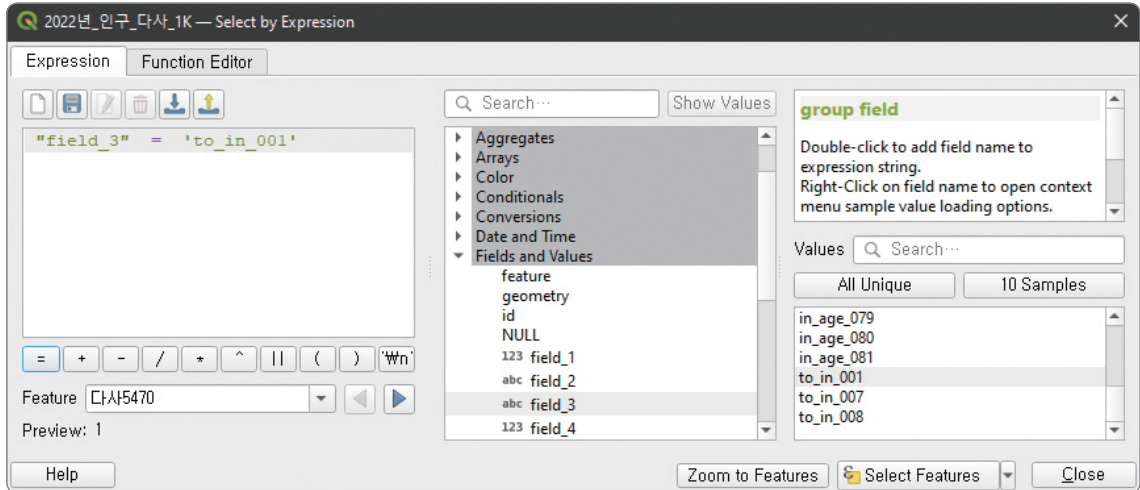
- You can now verify that the 1km grids intersecting with the Seoul boundary have been selected.



V QGIS Analysis Example - 1km Grid Population (Seoul)

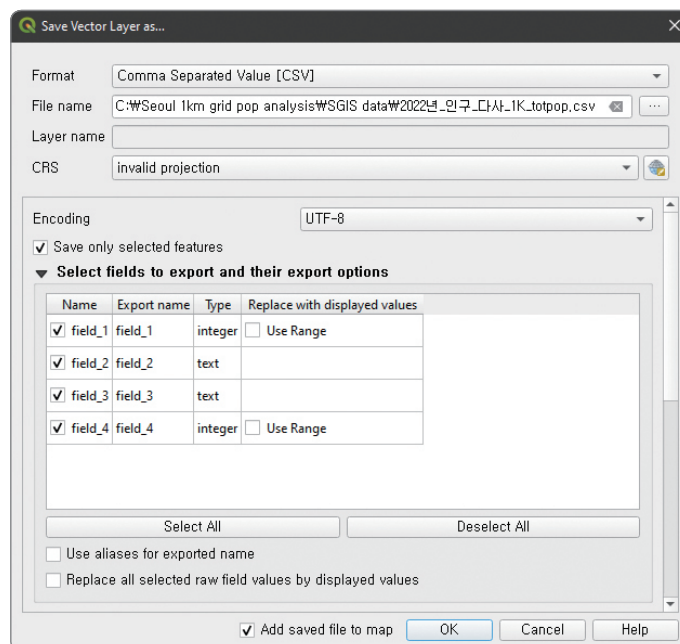
④ Selecting Attribute Values (Total Population) in the Statistical Layer

- Select the **2022년_인구_다사_1K** layer, then on the **Edit** ► **Select** ► **Select Features by Expression** (🔍) screen, choose the third field (field_3) and set its value to 'to_in_001' (total population).



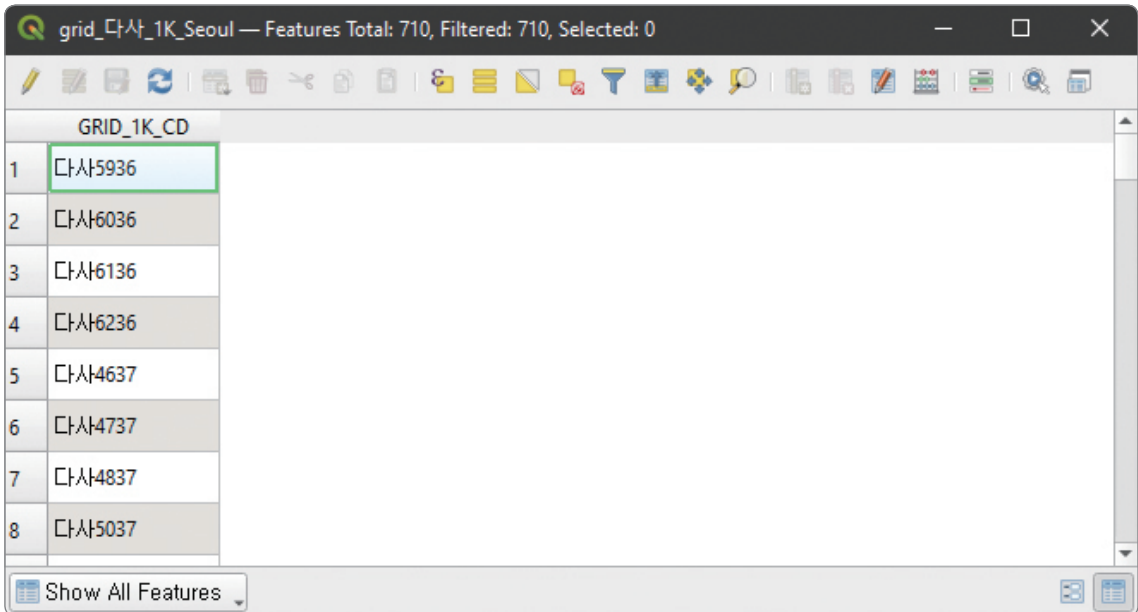
- ※ You can click the **All Unique** button to view a list of statistical attribute codes and select items from the shaded parts of the list to automatically build the expression without manual entry.

- Right-click on the **2022년_인구_다사_1K** layer, then select **Export** ► **Save Selected Features As** and save the selected features with a new name (**2022년_인구_다사_1K_totpop.csv**).

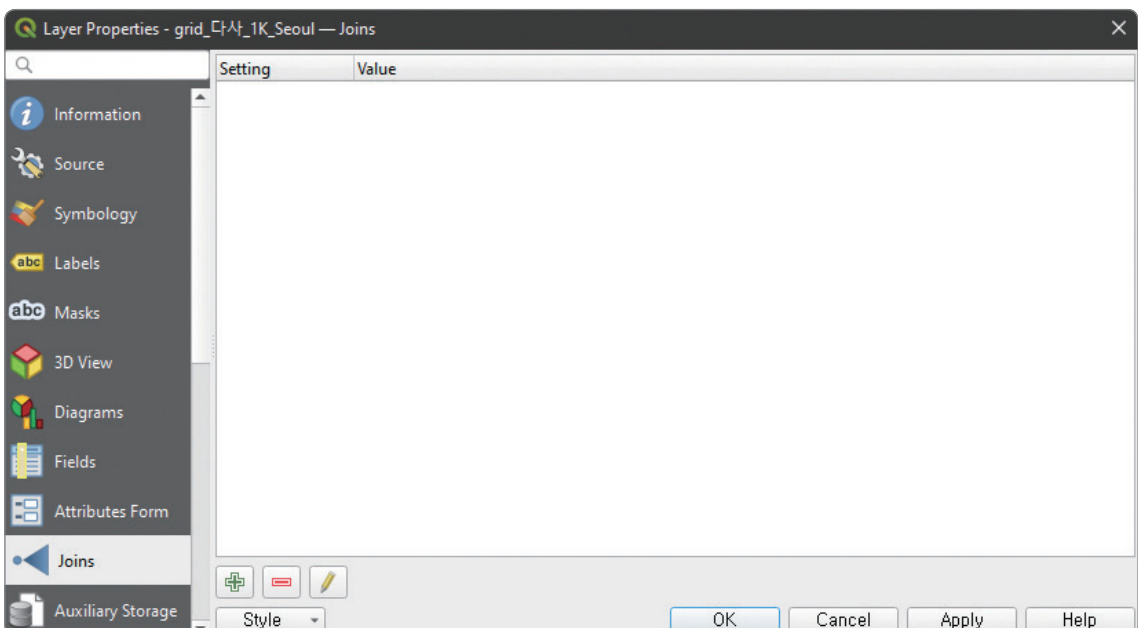


5 Merging the Seoul Grid with Total Population Data

- Right-click on the `grid_다사_1K_Seoul` layer and select the `Open Attribute Table`. Note that only the grid code (GRID_1K_CD) is present.



- Right-click on the `grid_다사_1K_Seoul` layer, select `Properties` and open the Layer Properties screen. From the left menu, choose `Joins`, then click the button at the bottom.



V QGIS Analysis Example - 1km Grid Population (Seoul)

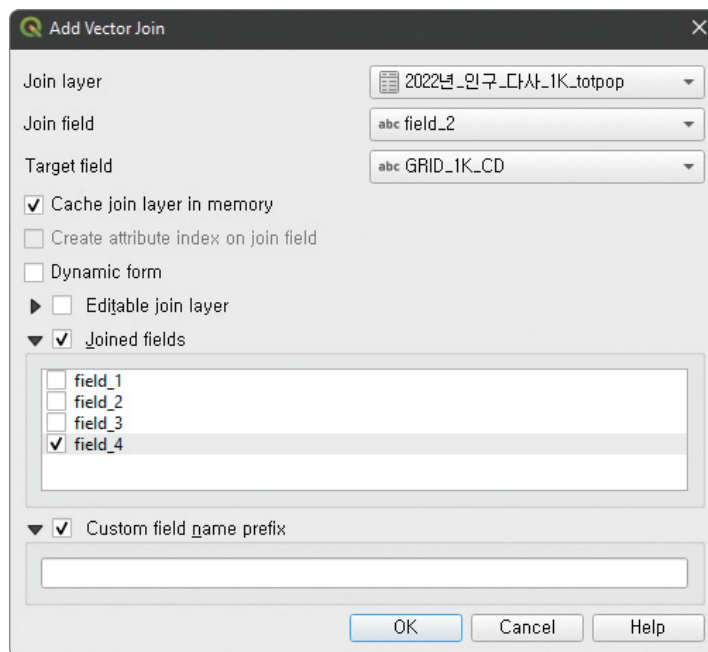
- Select the appropriate options for Join Layer, Join Field, Target Field and others.

Join Layer : 2022년_인구_다사_1K_totpop

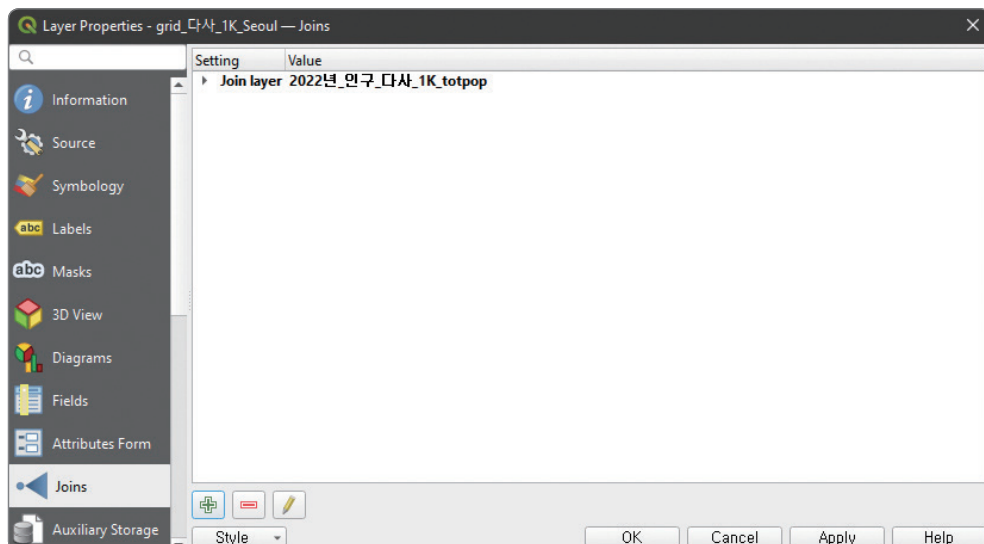
Join Field : field_2

Target Field : GRID_1K_CD

※ Ensure the statistical value field (field_4) is selected in the **Joined Fields**. You can omit the **Custom field name prefix**.



- Return to the Layer Properties screen and select **Apply** and **OK**.

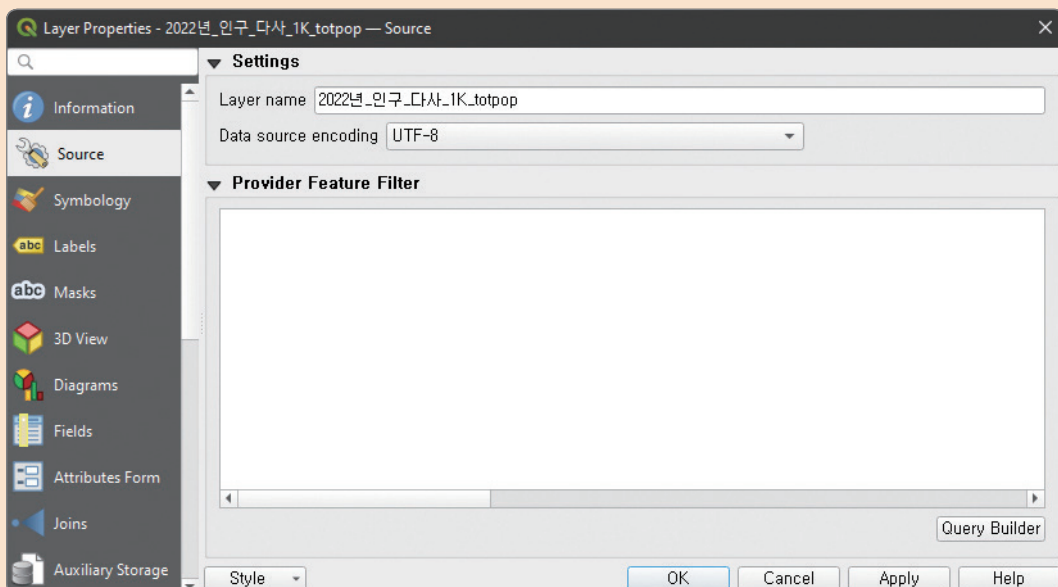


- Right-click on the `grid_다사_1K_Seoul` layer and select `Open Attribute Table` again. Note that the statistical value field (`field_4`) has been successfully joined to the table.

	GRID_1K_CD	field_4
1	다사5936	NULL
2	다사6036	NULL
3	다사6136	281
4	다사6236	175
5	다사4637	547
6	다사4737	8147
7	다사4837	661
8	다사5037	8

| Caution |

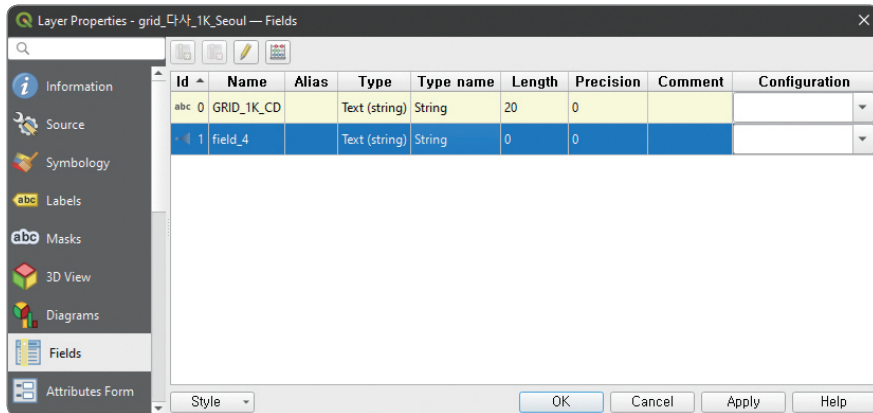
- ※ If the joining Process fails, go to `2022년_인구_다사_1K_totpop` layer's `Properties` ► `Source` screen and set the Data source encoding to UTF-8, then repeat Step 5.



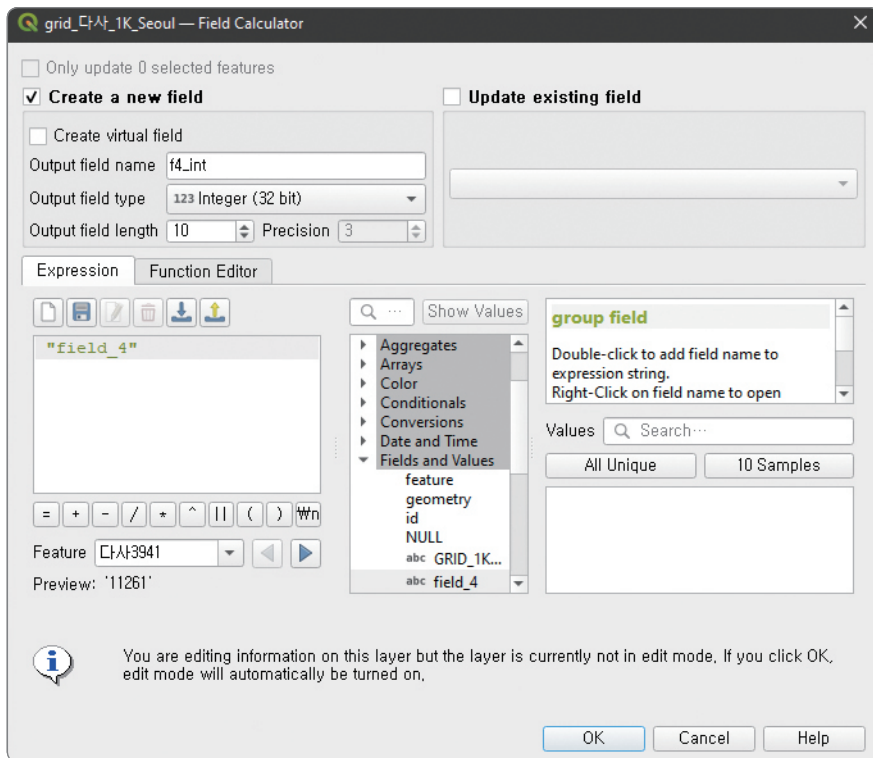
V QGIS Analysis Example - 1km Grid Population (Seoul)

6 Displaying a Choropleth Map

- Since the statistical value column (field_4) of `grid_다사_1K_Seoul` is in text (string) type, add an integer type column. Go to **Properties** ► **Fields** and select the Field Calculator()



- Write the name of the text(string) type column (field_4) in the **Expression** fields and set the name of the new integer type column as 'f4_int'.



- Unselect **Layer** ► **Toggle Editing** () to terminate editing mode and save the new column.

- Go to **grid_다사_1K_Seoul** layer's **Properties** ► **Symbology** screen and select the appropriate settings for displaying the choropleth map.

Symbology : Graduated

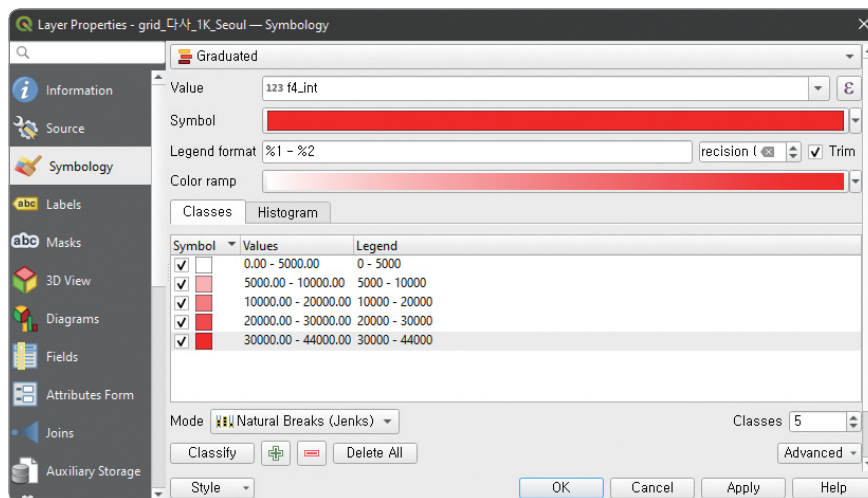
Value : f4_int

Symbology ► **Fill** ► **Simple Fill**

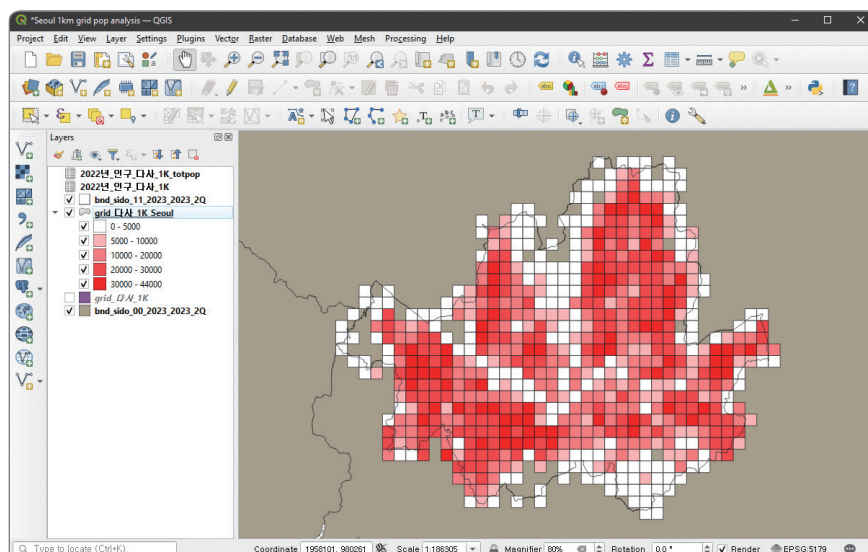
Mode : Natural Breaks (Jenks)

Classes : 5

- ※ You can adjust the range of values for each class by double-clicking the corresponding value in the symbol settings.



- Below is the resulting choropleth map for Seoul's 1km-grid population.





SGIS⁺plus

Statistical Geographic
Information Service

CONTACT +82 42-481-2245, 2517
<https://sgis.kostat.go.kr/>

