SGIS plus
Statistical Geographic Information Service

https://sgis.kostat.go.kr/

# SGIS Small-Area Statistics User Manual

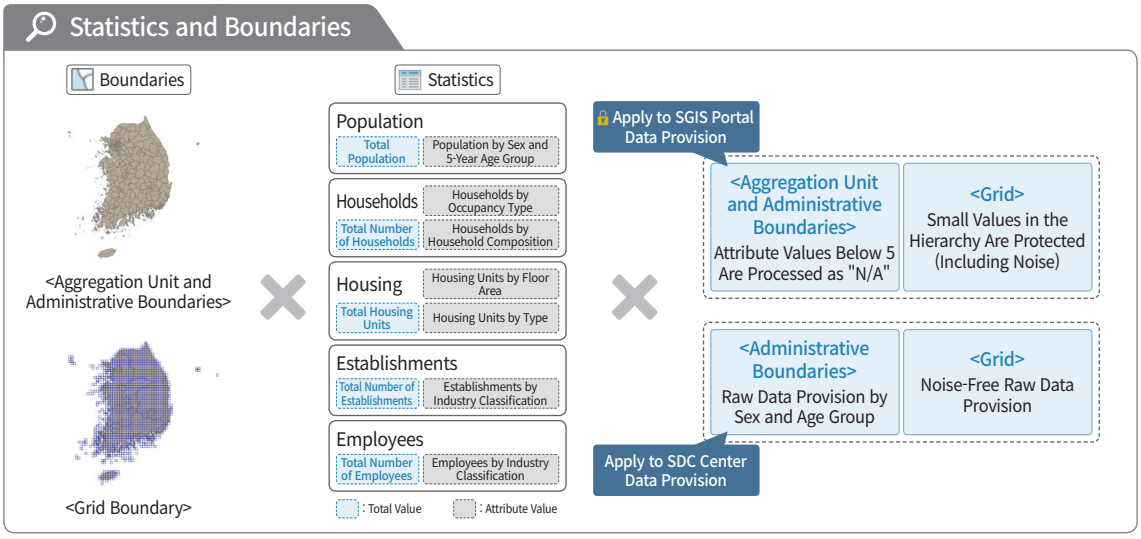
Statistics Korea

# I 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)

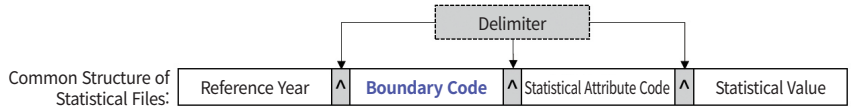
### | Provided Items | Small-Area Data by Target Statistics

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

## Statistical File Structure: Provided statistical data is organized into the following four columns.

- ① Reference Year: The year corresponding to the statistical values.
- ② Boundary Code: A code representing the boundary used as a key to link statistical data with boundary data.
- ③ Statistical Attribute Code: Indicates the statistical item with reference to the provided codebook (statistics_code).
  * Examples: to_in_001 (total population), to_in_007 (male total population), to_in_008 (female total population)
- ④ Statistical value: The value corresponding to the boundary and statistical attribute.

### Example of Statistical File Structure




<Grid Statistics>

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 2022 | ^ | 다사5252 | ^ | to_in_001 | ^ | 24772 |
| 2022 | ^ | 다사5252 | ^ | to_in_007 | ^ | 11557 |
| 2022 | ^ | 다사5252 | ^ | to_in_008 | ^ | 13213 |
| 2022 | ^ | 다사5253 | ^ | to_in_001 | ^ | 9852 |
| 2022 | ^ | 다사5253 | ^ | to_in_007 | ^ | 4444 |
| 2022 | ^ | 다사5253 | ^ | to_in_008 | ^ | 5411 |
| ... | | ... | | ... | | ... |

<Aggregation Unit Statistics>

| | | | | | | |
|---|---|---|---|---|---|---|
| 2022 | ^ | 11010530030101 | ^ | to_in_001 | ^ | 540 |
| 2022 | ^ | 11010530030102 | ^ | to_in_001 | ^ | 500 |
| 2022 | ^ | 11010530030103 | ^ | to_in_001 | ^ | 402 |
| 2022 | ^ | 11010530030104 | ^ | to_in_001 | ^ | 555 |
| 2022 | ^ | 11010530030001 | ^ | to_in_001 | ^ | 548 |
| 2022 | ^ | 11010530030002 | ^ | to_in_001 | ^ | 595 |
| ... | | ... | | ... | | ... |

<Administrative District Statistics>

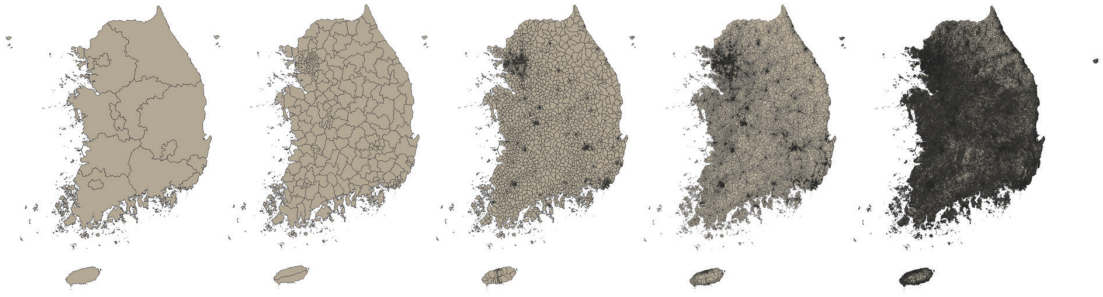
| | | | | | | |
|---|---|---|---|---|---|---|
| 2022 | ^ | 11010530 | ^ | to_in_001 | ^ | 8186 |
| 2022 | ^ | 11010530 | ^ | to_in_007 | ^ | 3608 |
| 2022 | ^ | 11010530 | ^ | to_in_008 | ^ | 4578 |
| 2022 | ^ | 11010540 | ^ | to_in_001 | ^ | 2062 |
| 2022 | ^ | 11010540 | ^ | to_in_007 | ^ | 954 |
| 2022 | ^ | 11010540 | ^ | to_in_008 | ^ | 1108 |
| ... | | ... | | ... | | ... |

※ Population statistics for the boundaries surrounding Sajik-dong (administrative code: 11010530) and Samcheong-dong (11010540)

## Boundaries

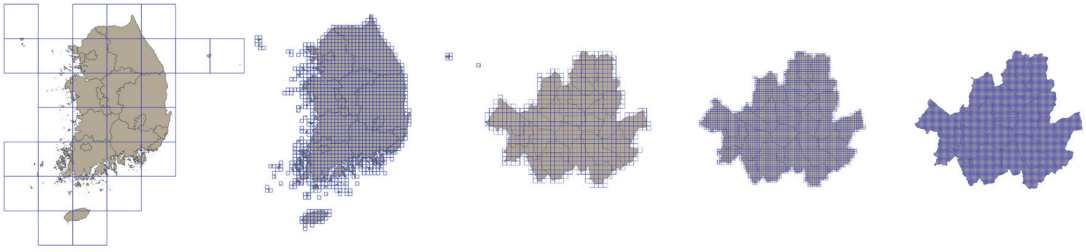
- Aggregation unit boundaries: Boundaries constructed considering population size (optimal 500 people), socio-economic homogeneity (housing types, land value) and shape, approximately 1/29 the size of an administrative district (-dong/-eup/-myeon).
- Administrative boundaries: Boundaries at the levels of province (-si/-do), city/county/district (-si/-gun/-gu), and administrative district (-dong/-eup/-myeon).
- Grid boundaries: Square-shaped boundaries covering the entire national territory, based on the Ministry of Land, Infrastructure and Transport's standard format.

### Aggregation Unit and Administrative Boundaries



<17 Provinces (-si/-do)> <250 Cities/Counties/ Districts (-si/-gun/-gu)> <3,558 Administrative Districts(-dong/-eup/-myeon) <Approximately 104,000 Aggregation Units> <Approximately 554,000 Basic Spatial Units>

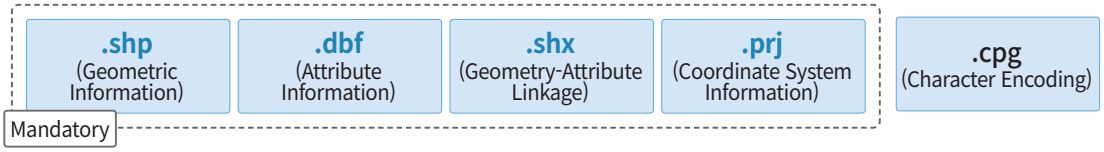
### Grid Boundaries




### 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



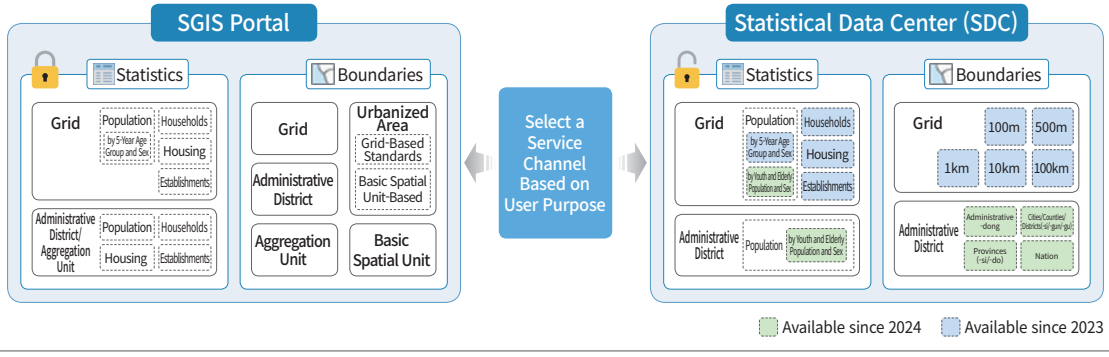

- SGIS boundary files (aggregation units, administrative districts and grids) use the **EPSG:5179 coordinate system.**

## 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 | 1,10,100㎞ | Administrative District | Aggregation Unit | 100,500m | 1,10,100㎞ | Administrative District | Aggregation Unit |
| Total | Population | Total Population | ○ | ○ | ○ | ○ | Δ* | ○ | × | × |
| | Households | Total Households | ○ | ○ | ○ | ○ | Δ* | ○ | × | × |
| | Housing | Total Housing Units | ○ | ○ | ○ | ○ | Δ* | ○ | × | × |
| | 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 | × | ○ | ○ | ○ | ○ | ○ | × | × |

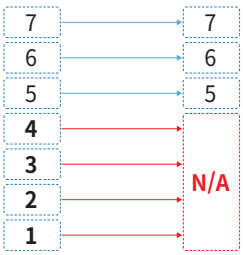
* SDC raw data is available only for total population, household and housing data at the 100m and 500m grid levels for the Administrative district ended with '-dong'(△)

## 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

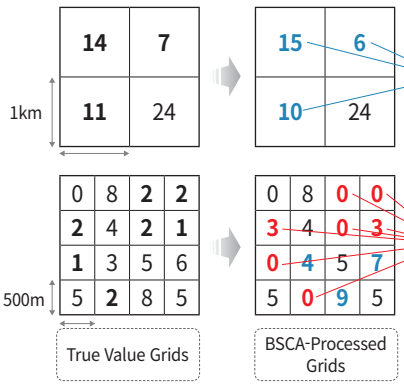
**Aggregation Unit Statistics Attribute Values Processed as N/A**




Attribute values in aggregation unit statistics (e.g., Population by 5-Year Age Group) below 5 are processed as 'N/A' and provided as such.

※ Total Values, such as Total Population, are not processed as 'N/A'.

**Grid Statistics BSCA Method (Example: Establishment Statistics)**




True Value Grids

BSCA-Processed Grids

Adjusted within a range of ±1 for establishment statistics true values (some adjusted within ±4).

※ Adjusted within a range of ±2 for population statistics true values (some adjusted within ±7).

Establishment statistics true values of 1~2 are adjusted to 0 or 3.

※ Population statistics true values of 1~4 are adjusted to 0 or 5.

# Ⅱ Application Process

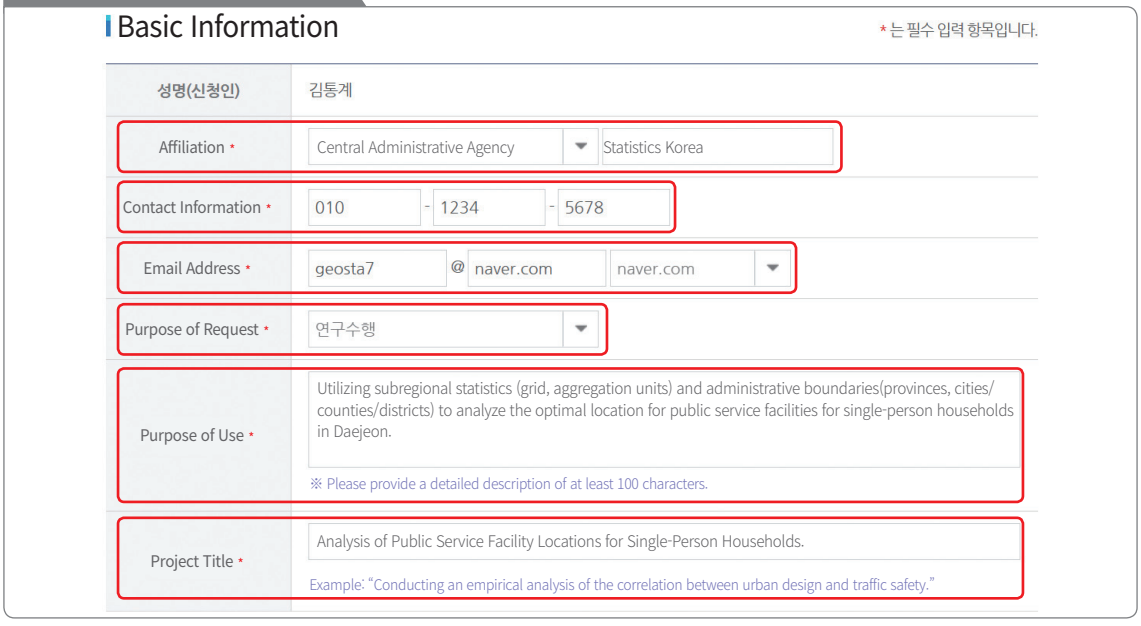
## 1 How to Apply for SGIS Portal Data

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


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

Enter Basic Information

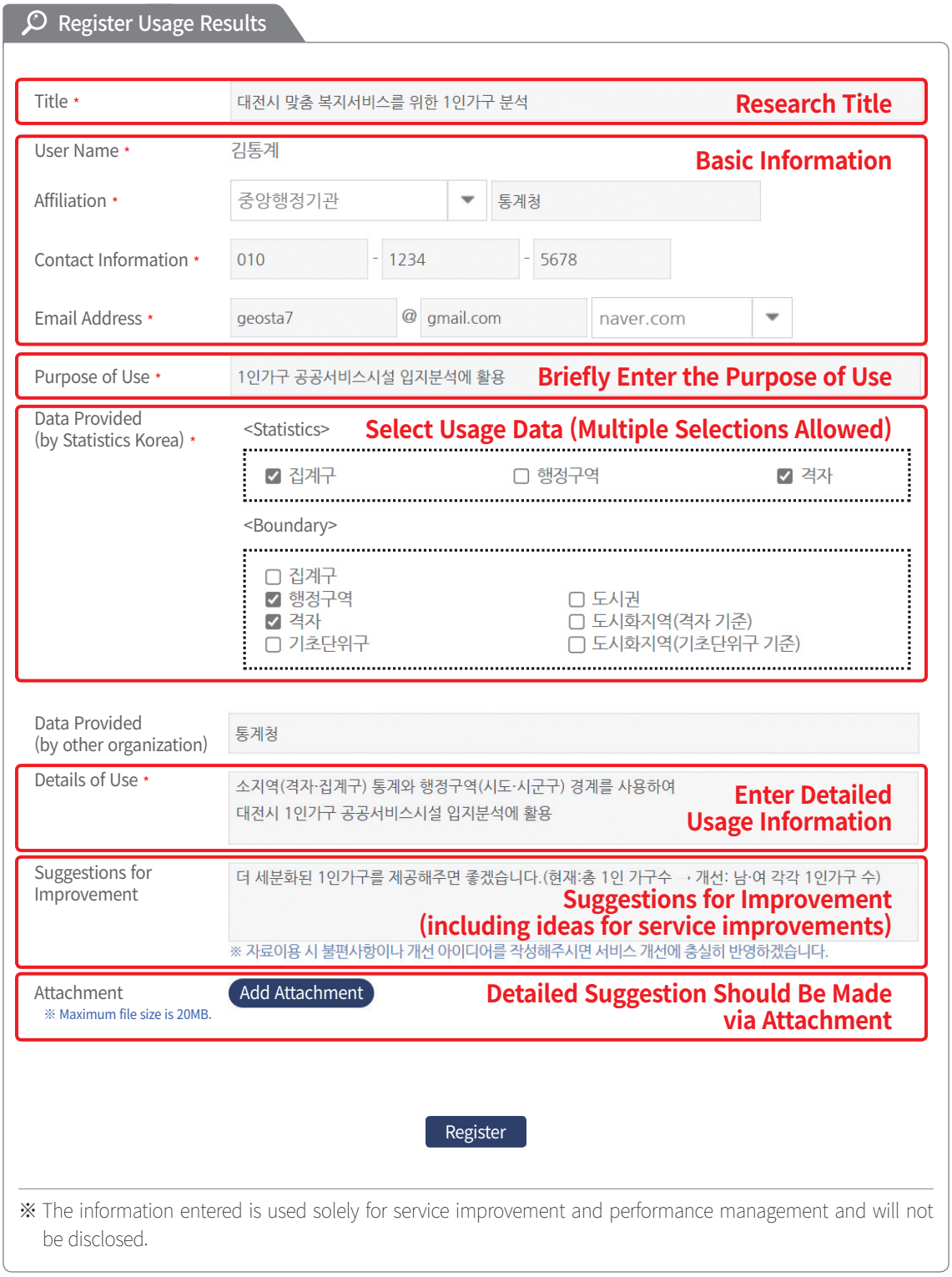


③ **Data Application:** Agree to submit usage results → Use filters to select data → Add and complete the application.


※ 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.




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

## 2 How to Use the Statistical Data Center (SDC)

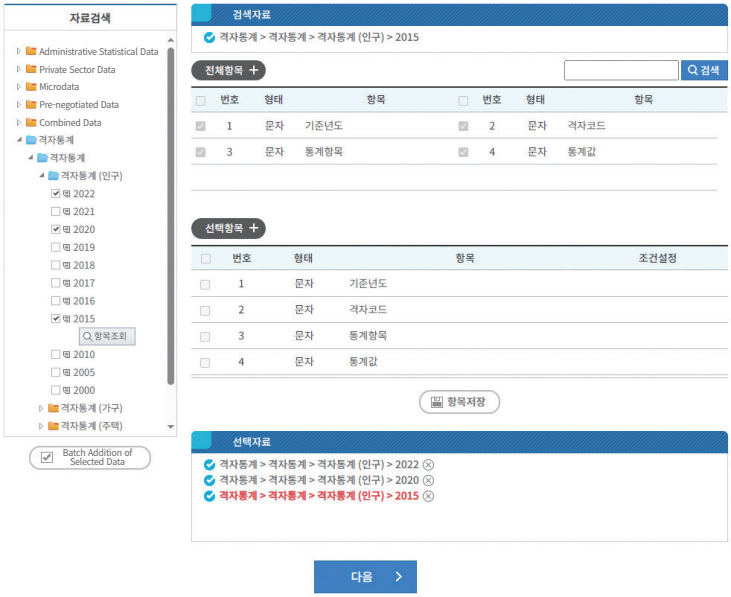
① **Pre-Application:** Before visiting the center, access https://data.kostat.go.kr and submit a application for the data you wish to analyze.

### SDC Portal Screen


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.

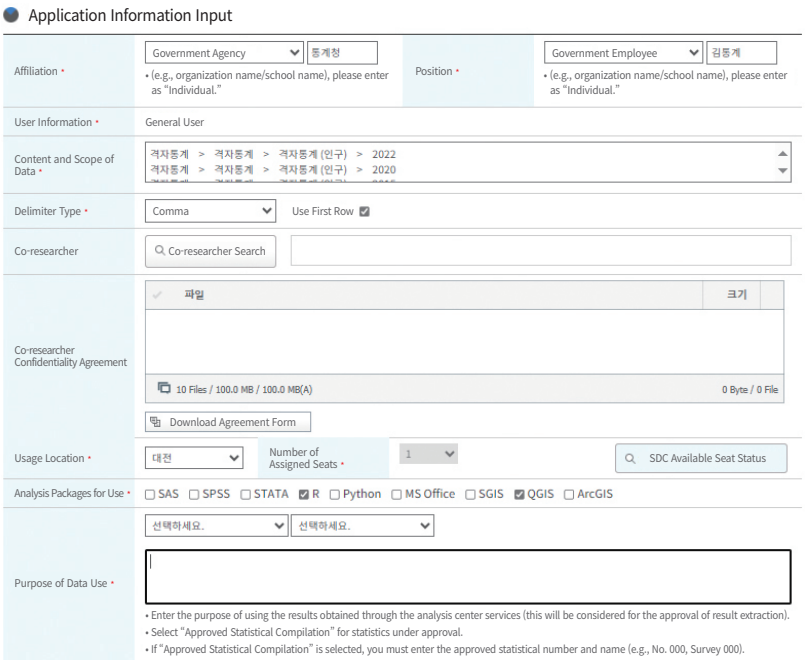
### Select Data



# Ⅱ Application Process

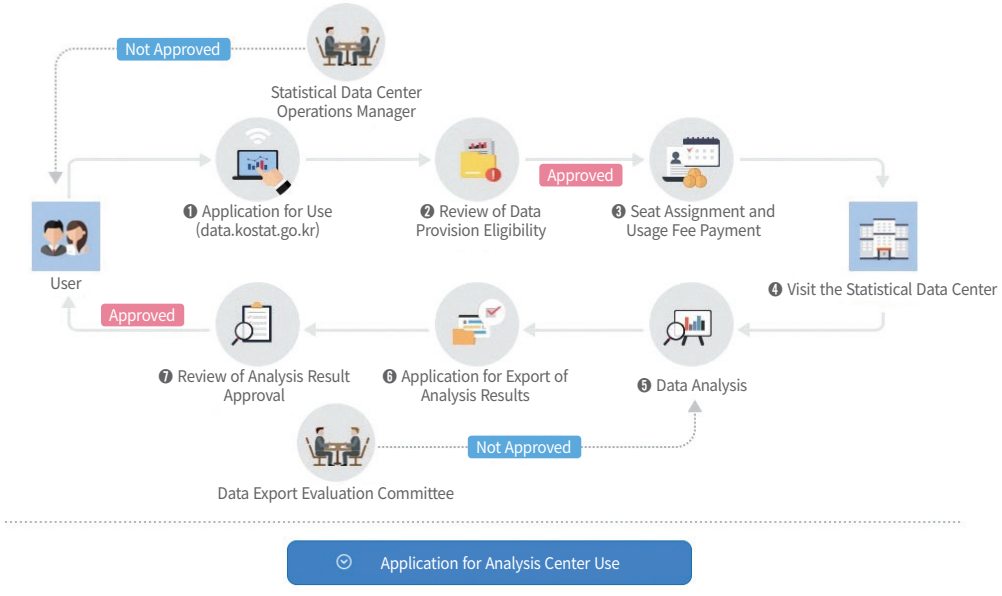
④ **Enter Application Information:** Provide details such as affiliation, user and co-researcher information and the purpose of data use, then submit the application.

## Entering Application Information




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



# Ⅲ 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

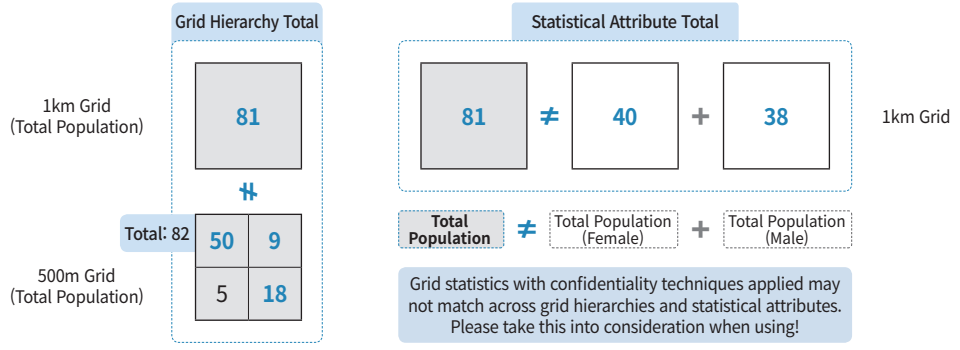
### Q6 Why are some grid boundaries missing when merging grid boundaries with statistical data?

- 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.

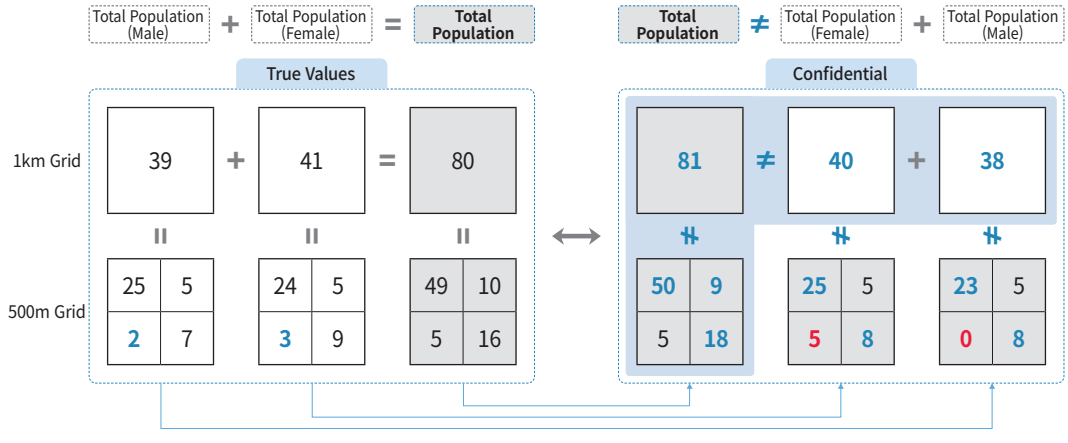
### Q7 Why does the sum of attribute values not match the total value?

- 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)**



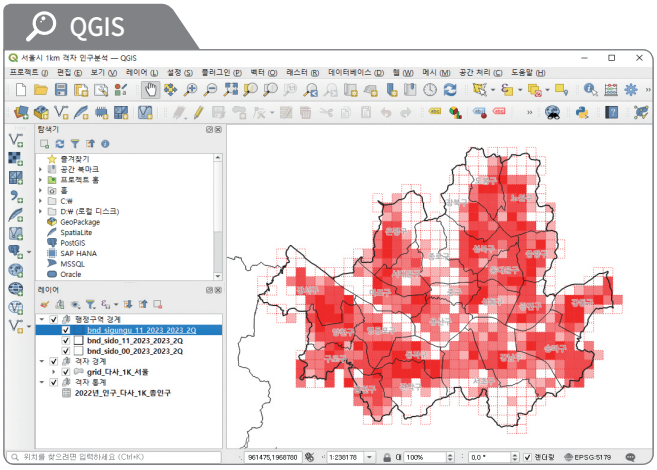

**Examples of Grid Statistics Before and After Applying Confidentiality Protection for Population Data**



# Ⅳ Map Visualization Tools

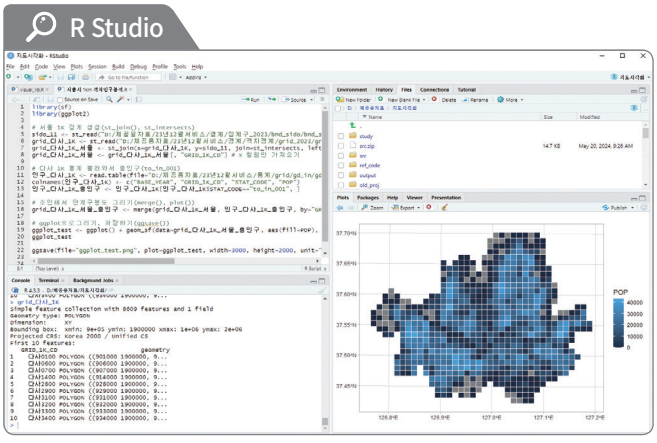
## ❶ QGIS

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



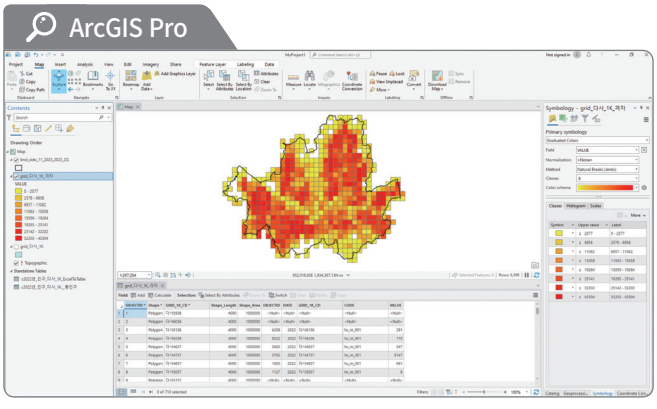

## ❷ R

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




## ❸ 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: esrikr.com

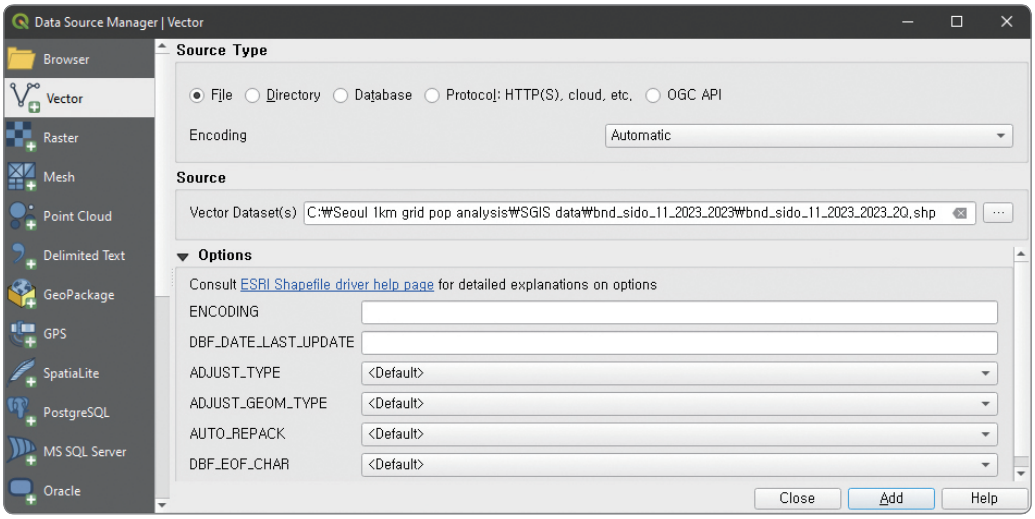


# V QGIS Analysis Example
## - 1km Grid Population (Seoul) -

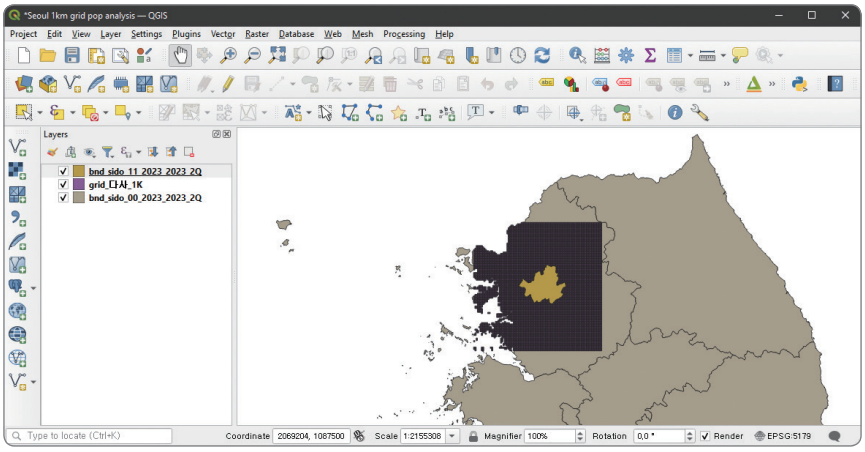
This example explains the step-by-step process of analyzing the Seoul Boundary, National SIDO Boundaries, 1km-Grid Boundaries and 2022 1km-Grid Population Statistics downloaded from the SGIS Data Provision menu using the latest version of QGIS (3.34.8).

### ❶ 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.

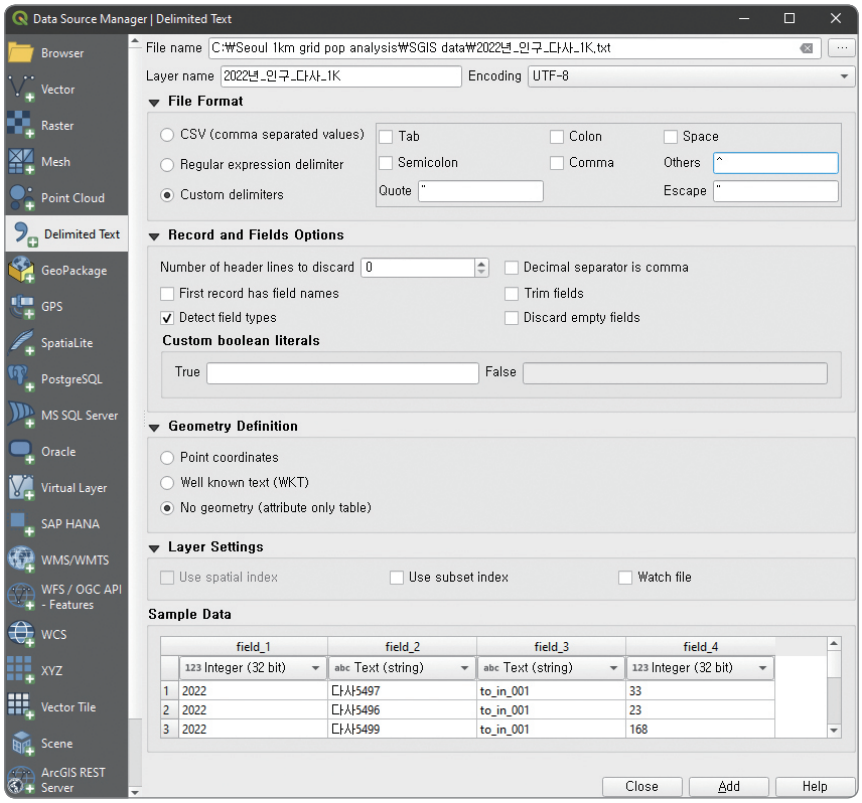


## ❷ Adding Statistical Layers (.txt)

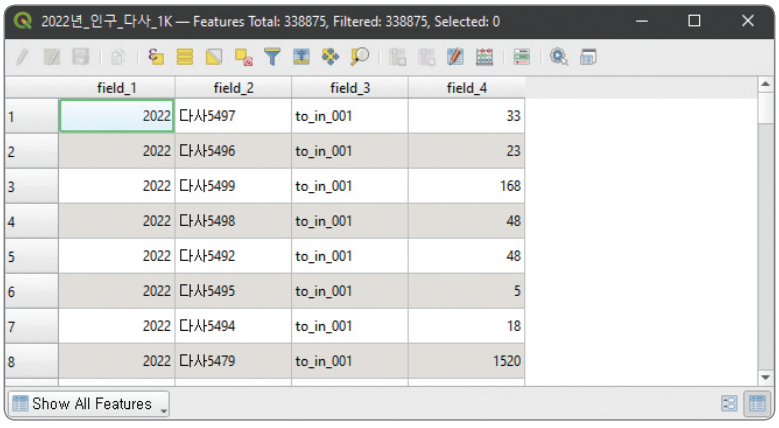
- Go to Layer ▶ Add Layer ▶ Add Delimited Text Layer to add 2022년_인구_다사_1K.txt

※ 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.

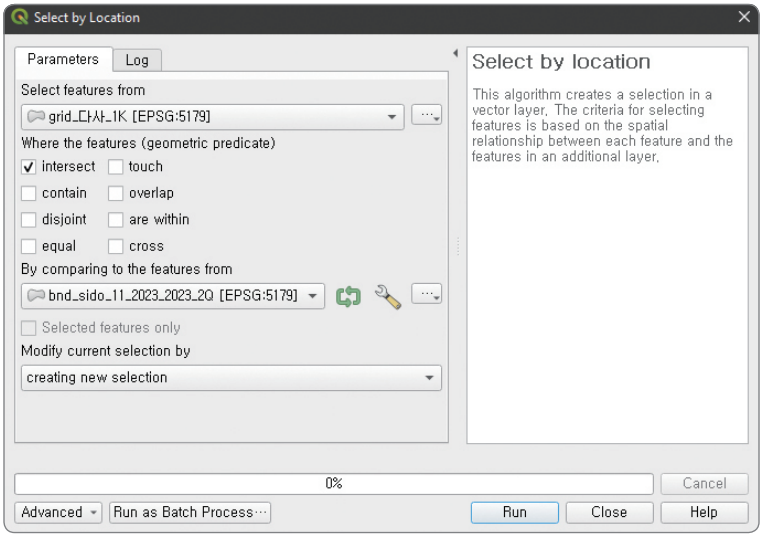


## V QGIS Analysis Example - 1km Grid Population (Seoul)

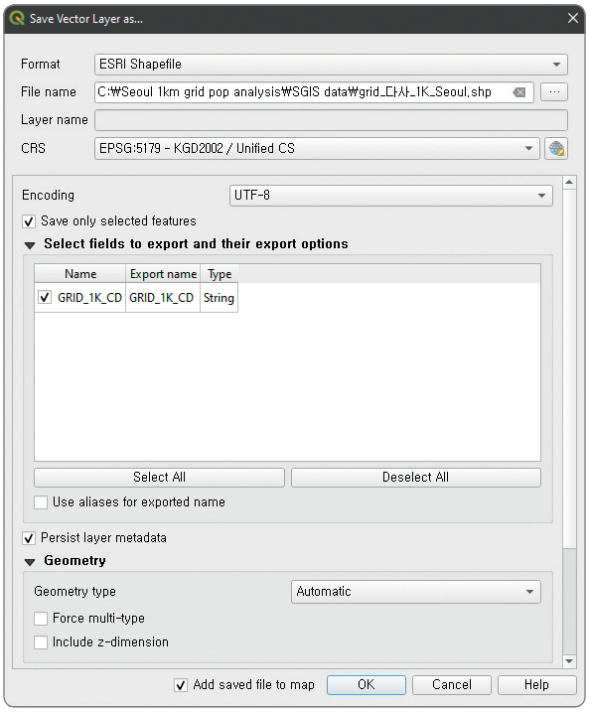
### ❸ Selecting Intersect Areas Between DASA 1km-Grid Boundaries and Seoul Boundary

- 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.

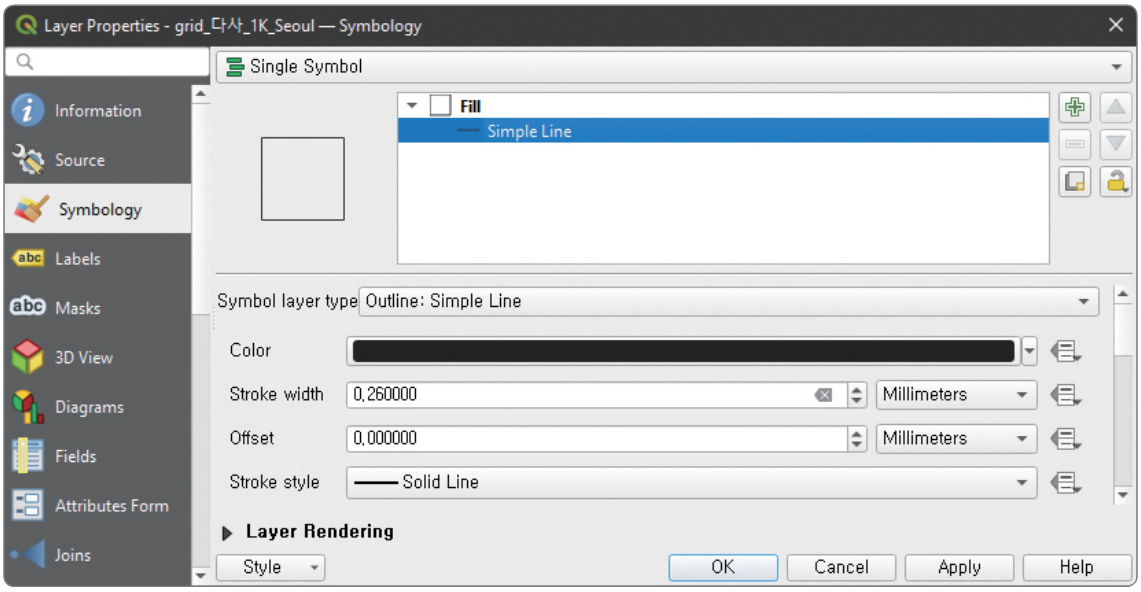
※ The selected grids in the grid_다사_1K layer will be highlighted in yellow.



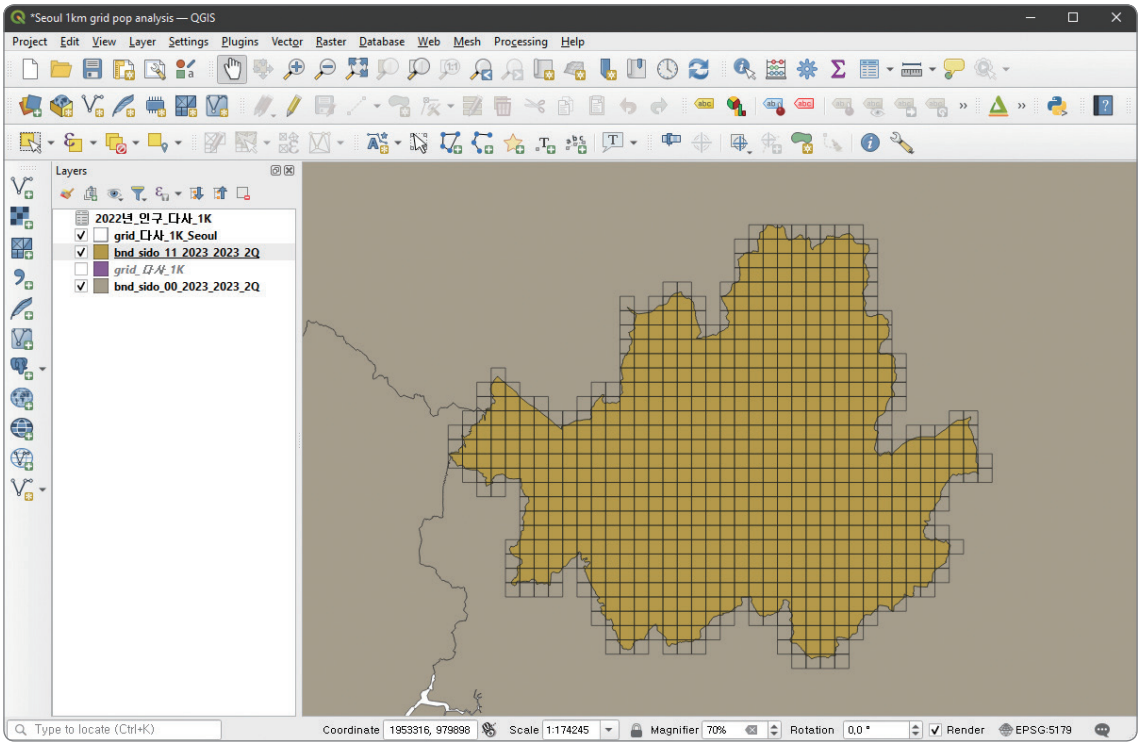

- Right-click on the grid_다사_1K layer, then select Export ▶ Save Selected Features As (grid_다사_1K_서울시.shp).



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

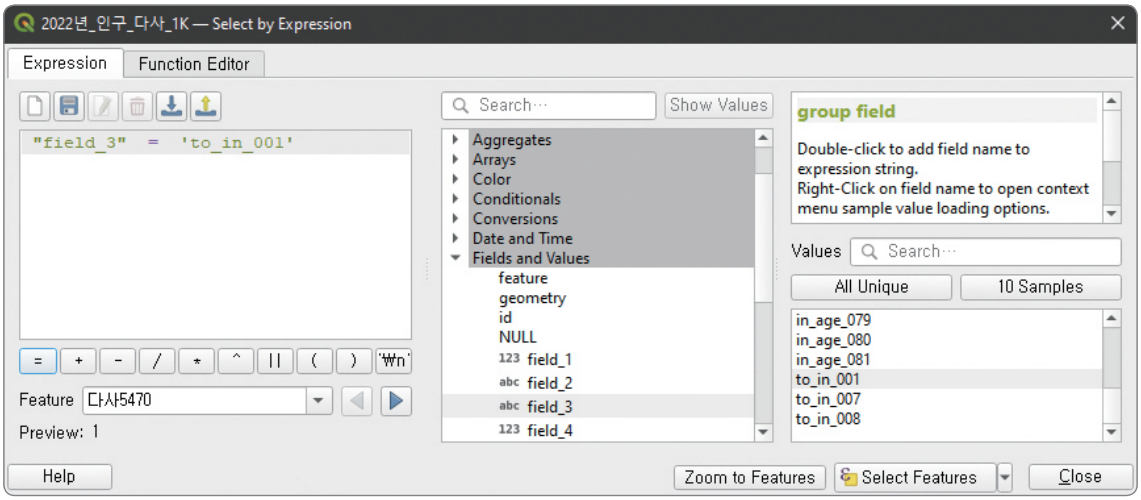



- You can now verify that the 1km grids intersecting with the Seoul boundary have been selected.



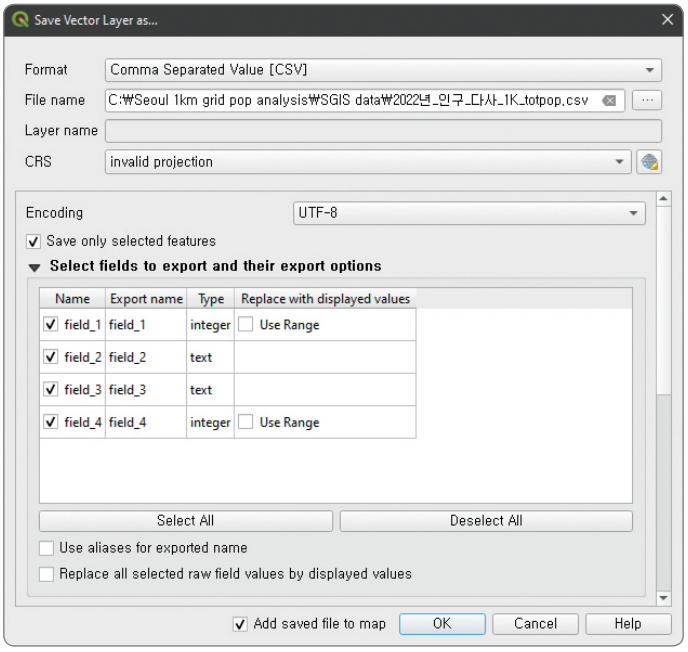
### ❹ Selecting Attribute Values (Total Population) in the Statistical Layer

- 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).



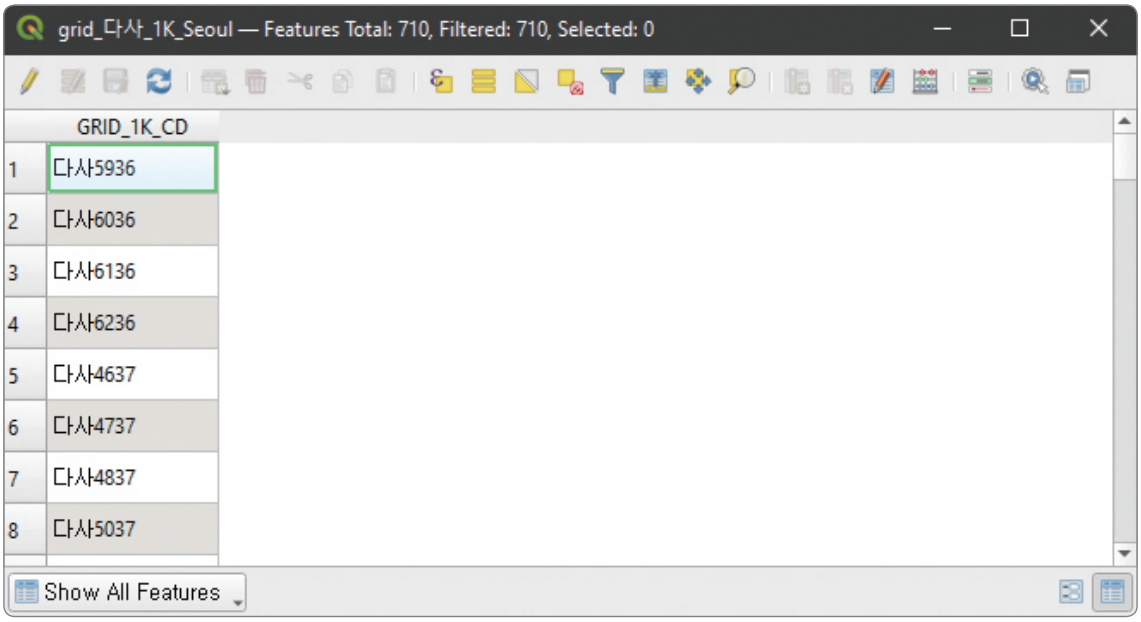

※ You can click the All Unique button to view a list of statistical attribute codes and select items from the shaded parts of the list to automatically build the expression without manual entry.

- 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).

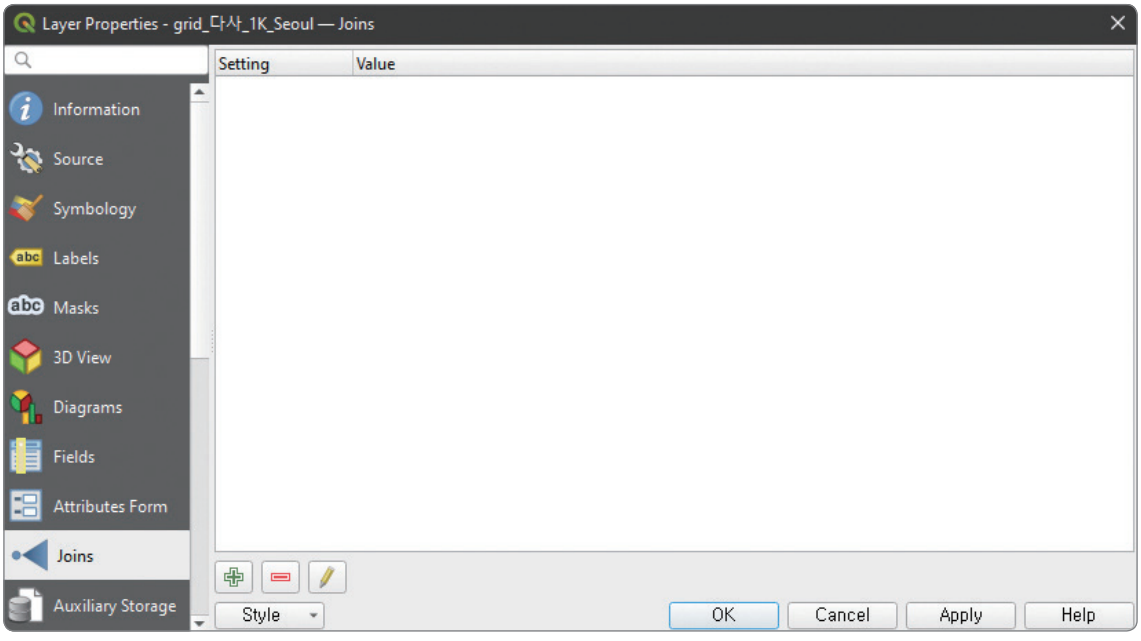


### ❺ Merging the Seoul Grid with Total Population Data

- 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.

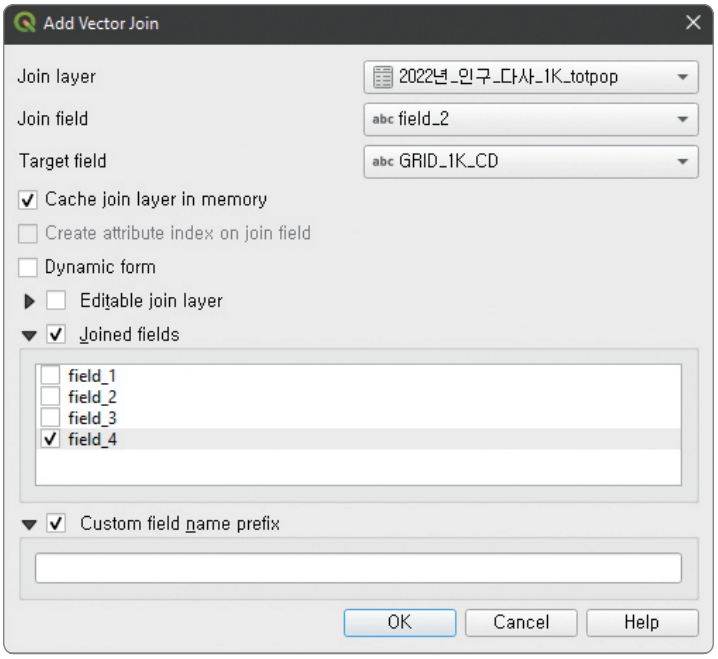



- Right-click on the grid_다사_1K_Seoul layer, select Properties and open the Layer Properties screen. From the left menu, choose Joins, then click the ⊕ button at the bottom.

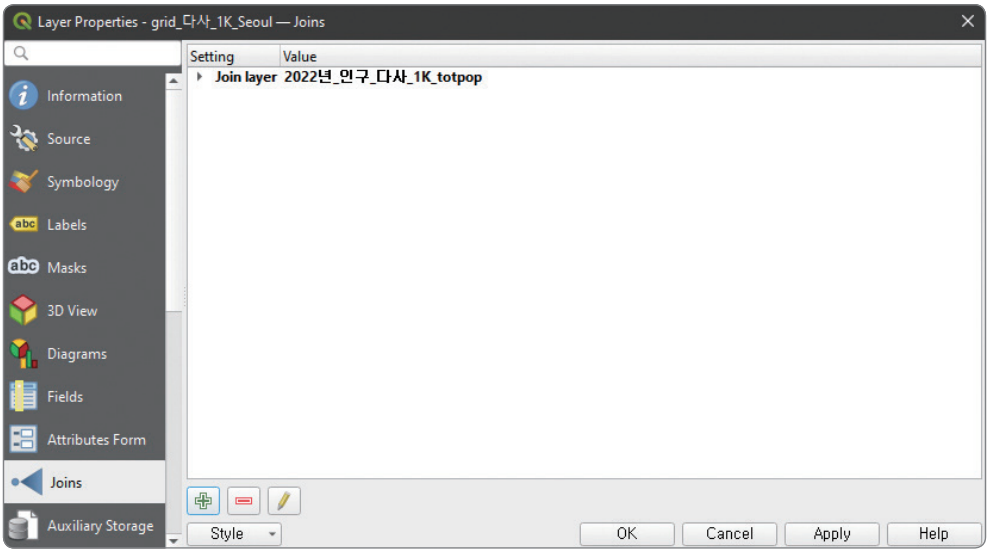


- Select the appropriate options for Join Layer, Join Field, Target Field and others.

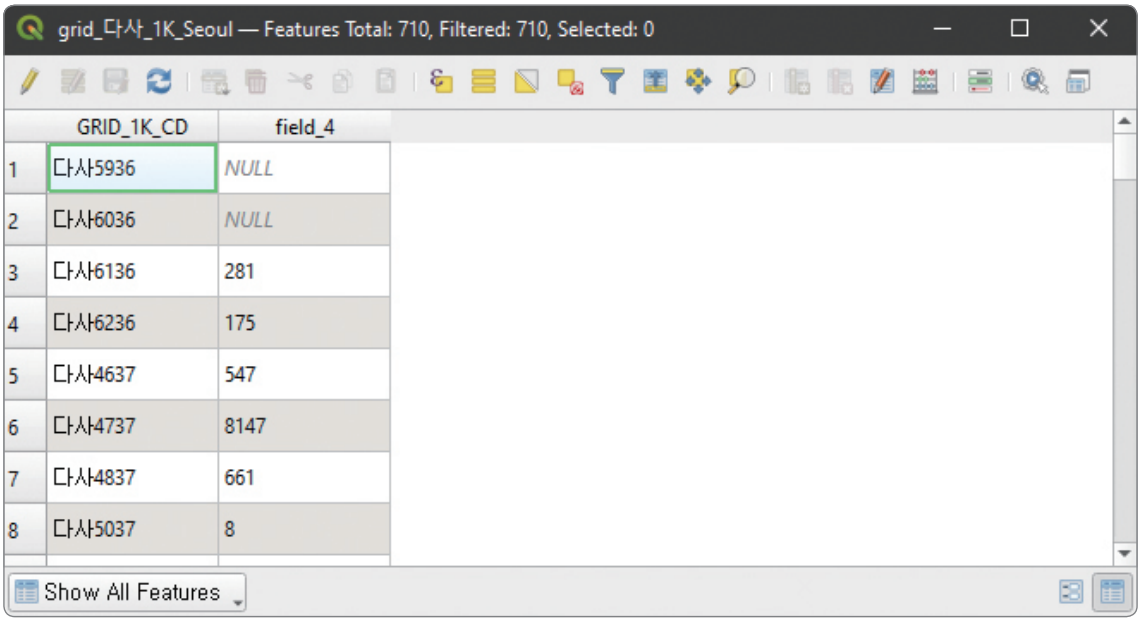
Join Layer : 2022년_인구_다사_1K_totpop
Join Field : field_2
Target Field : GRID_1K_CD
※ Ensure the statistical value field (field_4) is selected in the Joined Fields. You can omit the Custom field name prefix.




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

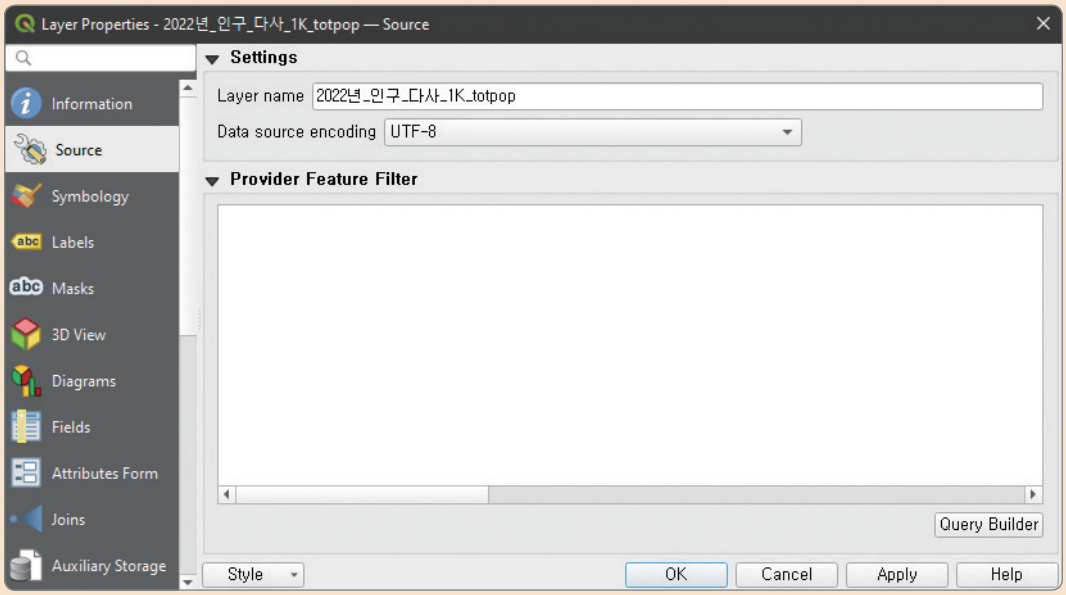


- Right-click on the grid_다사_1K_Seoul layer and select Open Attribute Table again. Note that the statistical value field (field_4) has been successfully joined to the table.



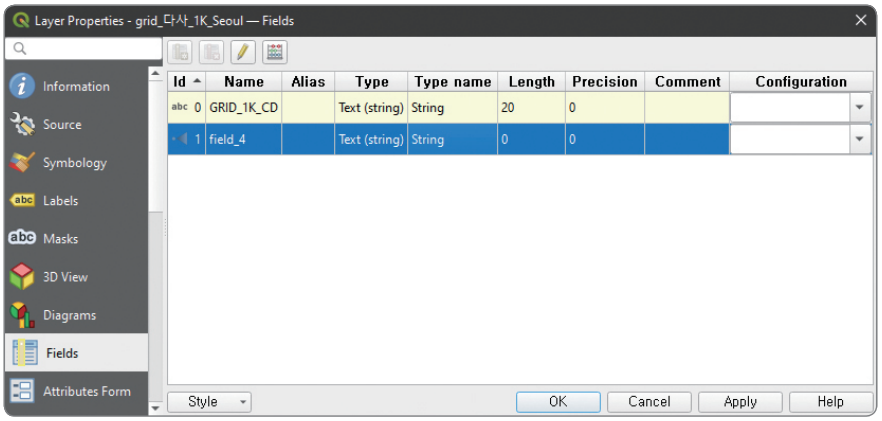

| Caution |

※ 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 ❺.

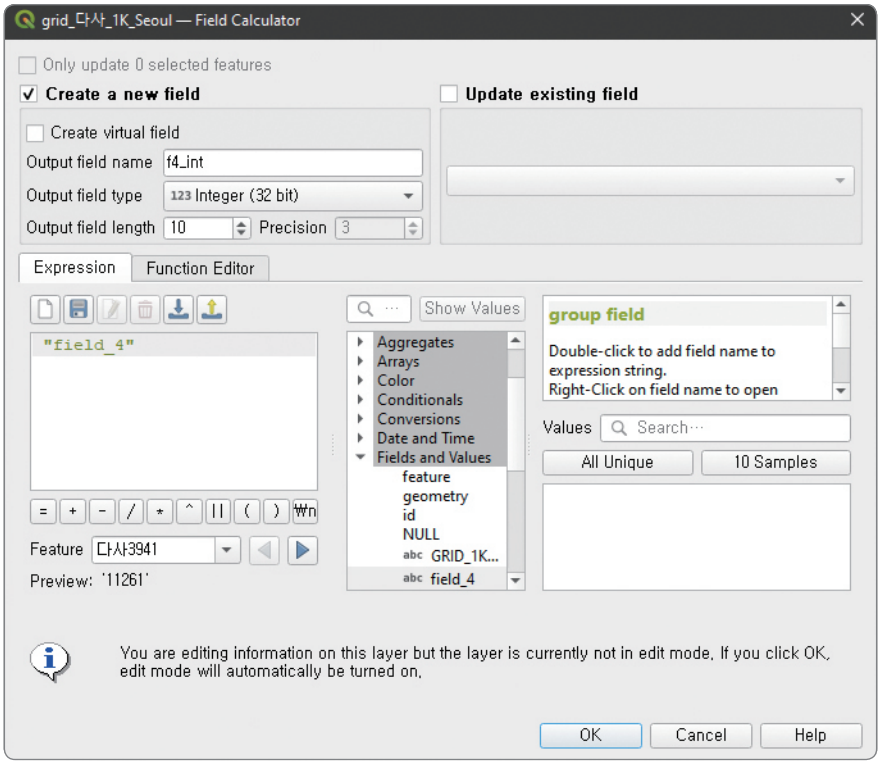


### ❻ Displaying a Choropleth Map

- 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( )



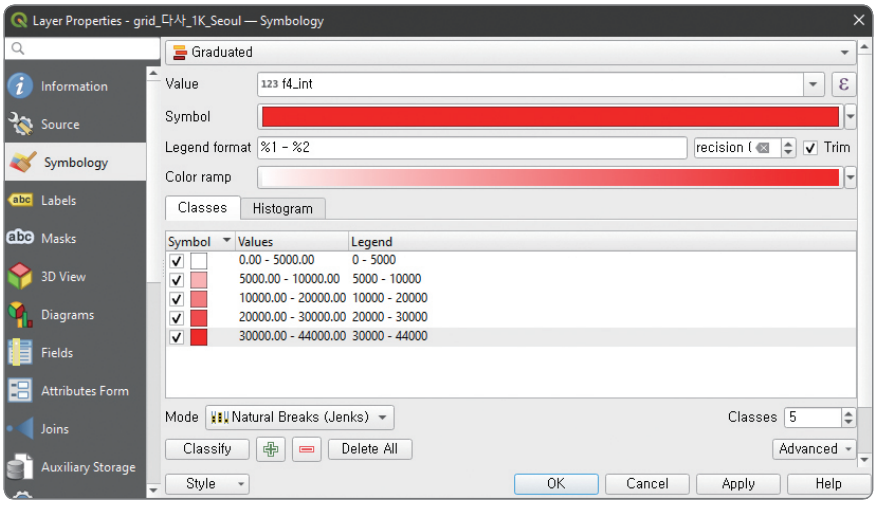

- Write the name of the text(string) type column (field_4) in the Expression fields and set the name of the new integer type column as 'f4_int'.



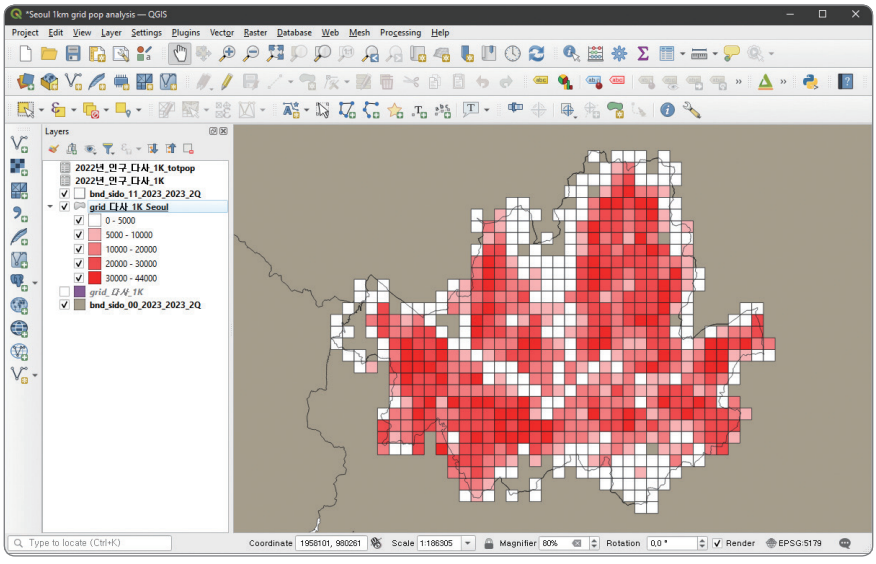

- Unselect Layer ▶ Toggle Editing ( ) to terminate editing mode and save the new column.

- 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/