A Lightweight and Reproducible Solution for South Korean Census Data Retrieval

tidycensuskr (R) and pycensuskr (Python): a dual-language workflow for spatial analysis and education in South Korea for all

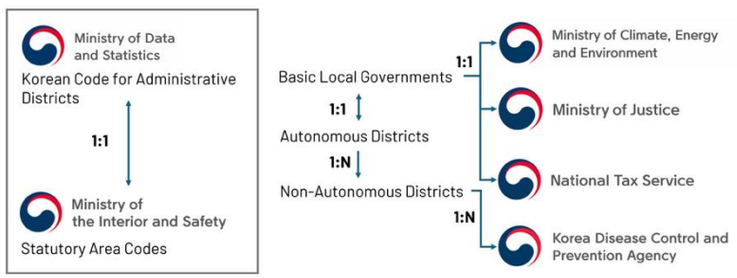
Hyesop Shin[1], Sohyun Park[2], Insang Song[3]
[1]The University of Auckland, [2]George Mason University Korea, [3]Seoul National University

I. Background

- South Korean census and administrative data are essential for small-area research (e.g. population, health), yet their use is constrained by restricted API access that requires domestic cell phone authentication.
- Across censuses, the same places are labelled with different spatial codes, and no dictionaries to chase boundary changes. This is hard for longitudinal studies
- We created to companion packages, tidycensuskr and pycensuskr to openly share census years and their broundaries

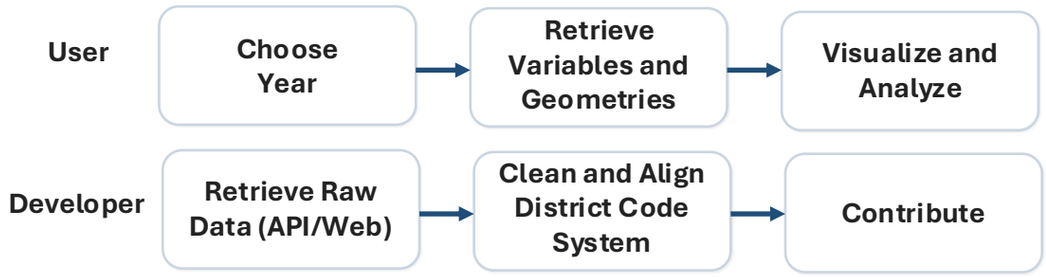
II. Packages

- **One spatial language**: MODS based codes and sf compatible geographies for census, tax, and housing data
- **Two ecosystems**: tidycensuskr on CRAN for R and pycensuskr on PyPI for Python
- **Time-aware data**: 2010, 2015, 2020 census with 120+ variables and boundary tracking for joins over time (as of Feb 2026)
- **Analysis ready**: Low-friction retrieval (few lines of code). Stable geographic IDs for joins, mapping, and modelling



III. Dual-language workflow

Same conceptual steps in R and Python: (1) select geography + years, (2) fetch variables, (3) map / model.



Python example (pycensuskr)

```
from pycensuskr import CensusKR
import geopandas as gpd
import matplotlib.pyplot as plt
pck = CensusKR()

# 1) Retrieve adm2 totals for 2020
pop_2020 = pck.anycensus(year=2020,
type="population", level="adm2",
geometry = True)

# 2) Quick plotting
pop_2020.plot("all
households_total_per")
plt.show()
```

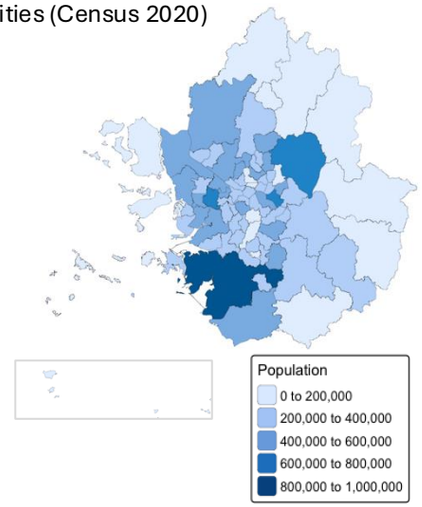
R example (tidycensuskr)

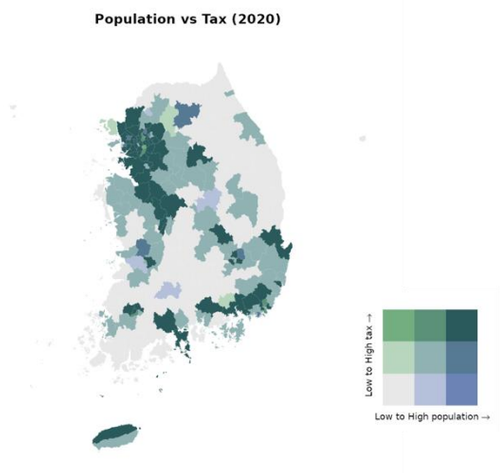
```
library(tidycensuskr)
library(sf)

# 1) Retrieve adm2 (si/gun/gu) totals
for 2020
pop_2020 <- anycensus(
  year  = 2020,
  type  = "population",
  level = "adm2",
  geometry = TRUE
)

# 2) Quick plotting
plot(pop_2020["all
households_total_prs"])
```

Population around Seoul and the neighboring cities (Census 2020)





IV. Strengths

- Rapid prototyping for mapping, exploratory spatial data analysis, and neighborhood indicators.
- Standardized identifiers reduce workloads in longitudinal boundary/attribute harmonization.
- Supports education: transparent pipelines from raw census tables → spatial objects.
- Bridges ecosystems: R spatial workflows (sf, spdep, tmap) and Python geospatial stacks (geopandas, pysal).

V. Conclusion & roadmap

- Enable a consistent, scriptable interface to Korean census/statistics across R and Python.
- Roadmap: boundary change tracking and end-to-end tutorials

Try our packages

tidycensuskr

https://sigmafelix.github.io/tidycensuskr/

pycensuskr

https://pypi.org/project/pycensuskr/

Feedback & Collaborations?

Contact Dr Insang Song: (geoissong@snu.ac.kr)