Shapefile Modification in R as the Basis for Linked Micromap Plots for New Geographic Regions

Jürgen Symanzik

Utah State University, Logan, UT

symanzik@math.usu.edu

http://www.math.usu.edu/~symanzik

--- with contributions from ---

Jeong Yong Ahn, Se-jin Park, Marc Weber, Michael McManus, Quinn Payton, XiaoTian Dai, Shuming Bao, Miao Shui & Bing She

Interface 2015, Morgantown, WV – Friday, June 12, 2015

Contents

- Linked Micromap (LM) Plots
  - Overview
  - Regional Micromaps
- Shapefile Modifications
  - Examples
  - Limitations
- Outlook and Future Work

Micromaps

- Link of row-labeled univariate (or multivariate) statistical summaries to corresponding geographical region
- Focus on statistical display and not on maps
- Useful for
  - environmental data
  - agricultural data
  - medical data
  - public health data
  - economical data

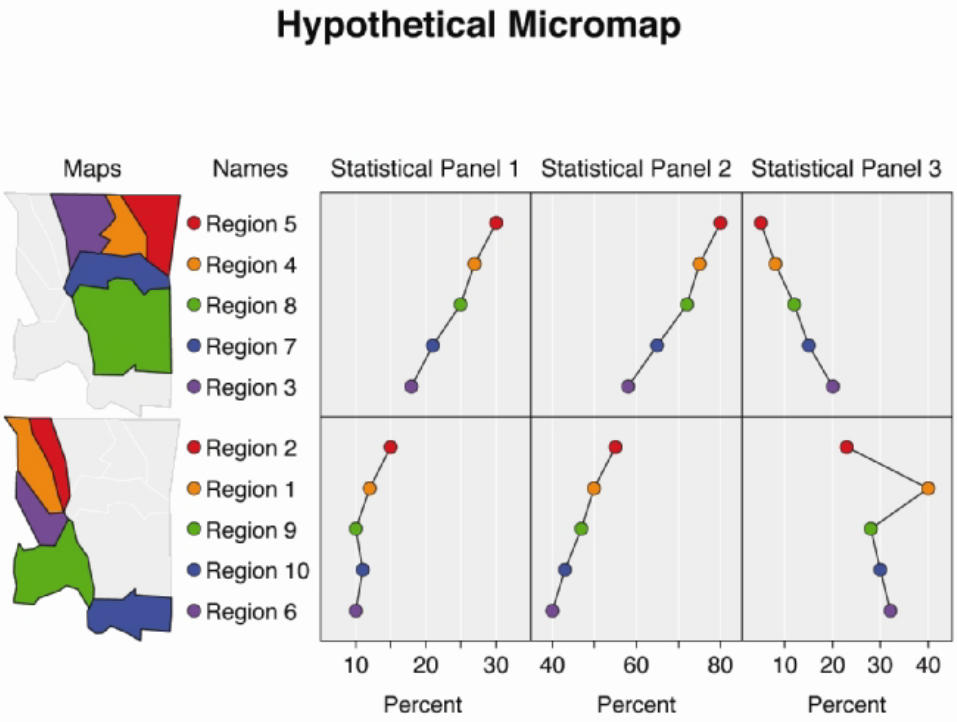


Figure Source: Gebreab, S. Y., Gillies, R. R., Munger, R. G., Symanzik, J. (2008): Visualization and Interpretation of Birth Defects Data Using Linked Micromap Plots, Birth Defects Research (Part A): Clinical and Molecular Teratology, 82(2):110-119.

History of Micromaps

- First presented at 1996 American Statistical Association's annual meeting (Olsen, Carr, Courbois, Pierson)
- Initial references:
  - Carr, Pierson (1996) Emphasizing Statistical Summaries …with Micromaps, SCSG*, Vol.7, No.3
  - Carr, Olsen, Courbois, Pierson, Carr (1998) Linked Micromap Plots …, SCSG, Vol. 9, No.1

*Statistical Computing & Statistical Graphics Newsletter: http://stat-computing.org/newsletter/archive.html

Micromaps in R

- Two recently developed R Packages:
  - ``micromap'' (by Payton, Olsen, Weber, McManus, Kincaid): general purpose
  - ``micromapST'' (by Carr, Pearson, Pickle): focus on the 50 U.S. states

Regional Micromaps

- Micromaps that are related to subregions within a particular country
- Previously:
  - National level: United States and Korea
  - Provincial level: United States and France
- Recent developments:
  - National level: most countries of South America, Korea (advanced), and China
  - Provincial level: most provinces of Korea

Regional Micromap Construction

- Modification of existing shapefiles
  - Simplification of boundaries (via Douglas-Peucker 1973 algorithm) [previously solved]
  - Removal of tiny islands [previously solved]
  - Enlarge small areas such as capital regions, e.g, Washington DC in the United States
  - Shift and resize (enlarge or shrink) regions that are far away from the main area, e.g., Alaska or Hawaii in the United States
  - New R code created for the remaining tasks

Shapefile Sources

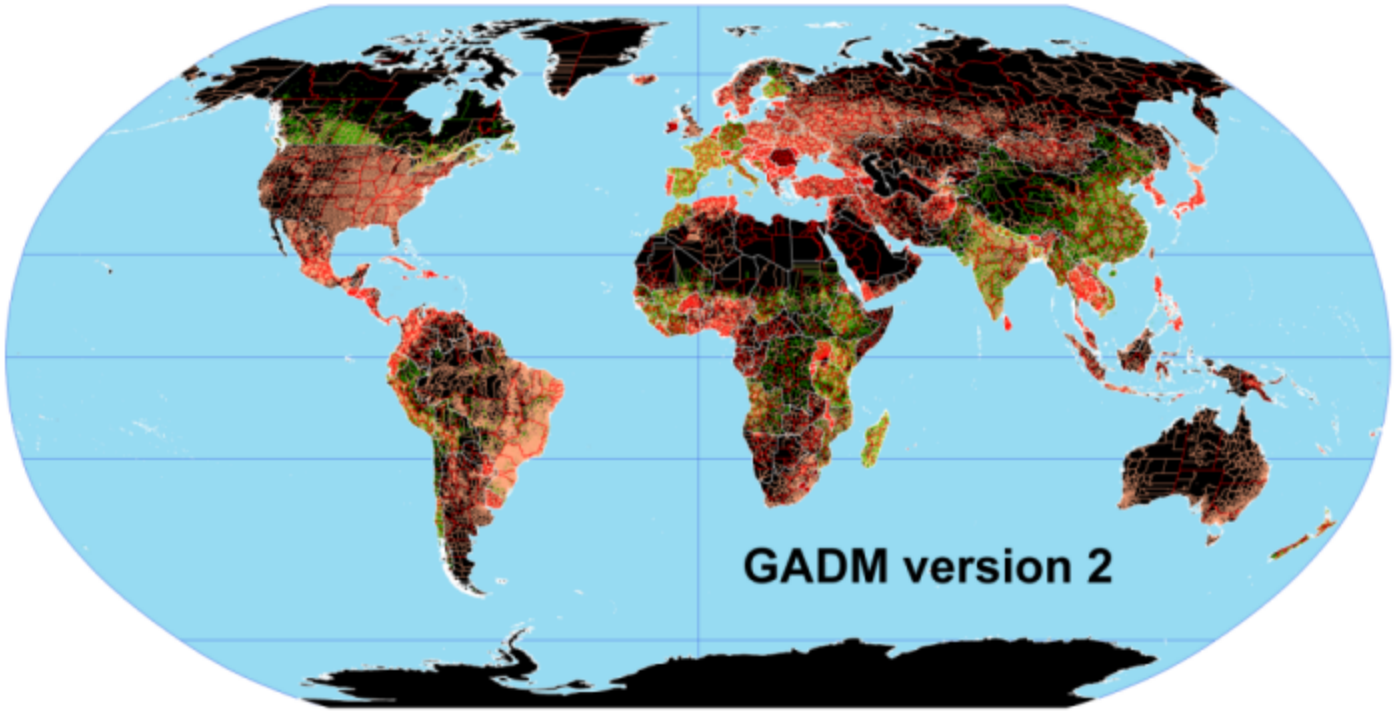
Source:

http://www.gadm.org

GADM database of Global Administrative Areas

GADM is a spatial database of the location of the world's administrative areas (or adminstrative boundaries) for use in GIS and similar software. Administrative areas in this database are countries and lower level subdivisions such as provinces, departments, bibhag, bundeslander, daerah istimewa, fivondronana, krong, landsvæðun, opština, sous-préfectures, counties, and thana. GADM describes where these administrative areas are (the "spatial features"), and for each area it provides some attributes, such as the name and variant names.

The current version is 2.0 (January 2012)



Shapefile Modification (1)

- New R function EnlargeReplaceMapArea heavily builds on R functions from rgeos, rgdal, maptools
  - unionSpatialPolygons: Aggregates polygons in a SpatialPolygons object
  - gBuffer: Expands the given geometry
  - gIntersection: Determines the intersection between two given geometries
  - gDifference: Determines the difference between two given geometries
- Call: ChinaShapeRefined <- EnlargeReplaceMapArea("Beijing", "Hebei", ChinaShapefileThin, "ename", "gbcode", 50000)

Shapefile Modification (2)

- New R function ShiftArea modifies the slots in the spatial object that contain latitude and longitude, based on
  - slot: Returns or sets information about the individual slots in an object
- Call:

```
KorShape1Sub28 <- ShiftArea(KorShape1Sub28, "Ongjin-gun", 1, "SIG_ENG_NM", c(30000, 30000, 120000, 120000))
KorShape1Sub28 <- ShiftArea(KorShape1Sub28, "Ongjin-gun", 2, "SIG_ENG_NM", c(0, 0, -20000, -20000))
```

Brazil Simplification

- UL: Raw
- UR: Thinned boundaries
- LL: Capital & other regions enlarged
- LR: Capital region enlarged & shifted

Figure Source: Symanzik, J., Dai, X., Weber, M. H., Payton, Q., McManus, M. G. (2014): Linked Micromap Plots for South America - General Design Considerations and Specific Adjustments, Revista Colombiana de Estadistica, 37(2):450-469.

Brazil Example

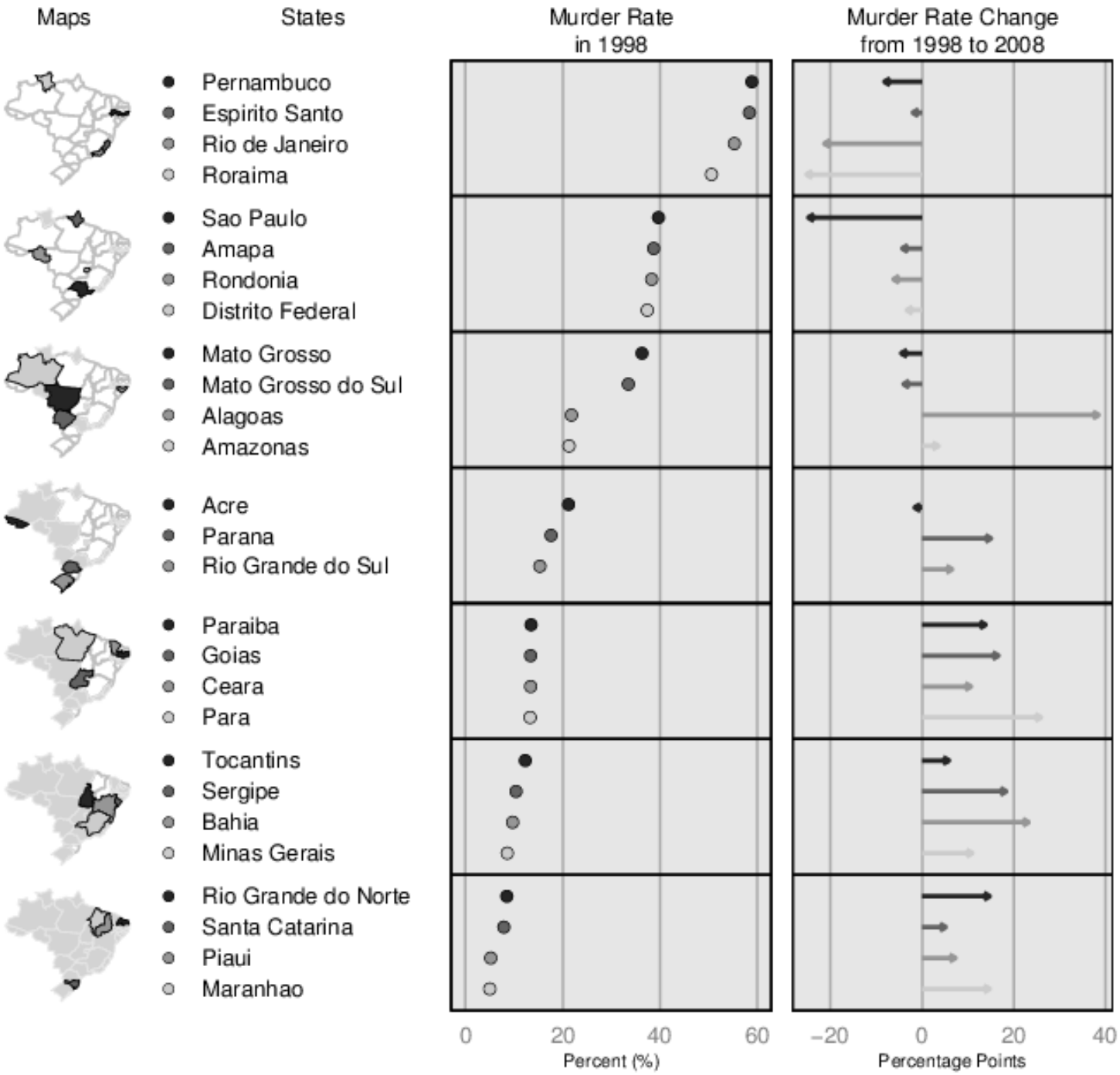
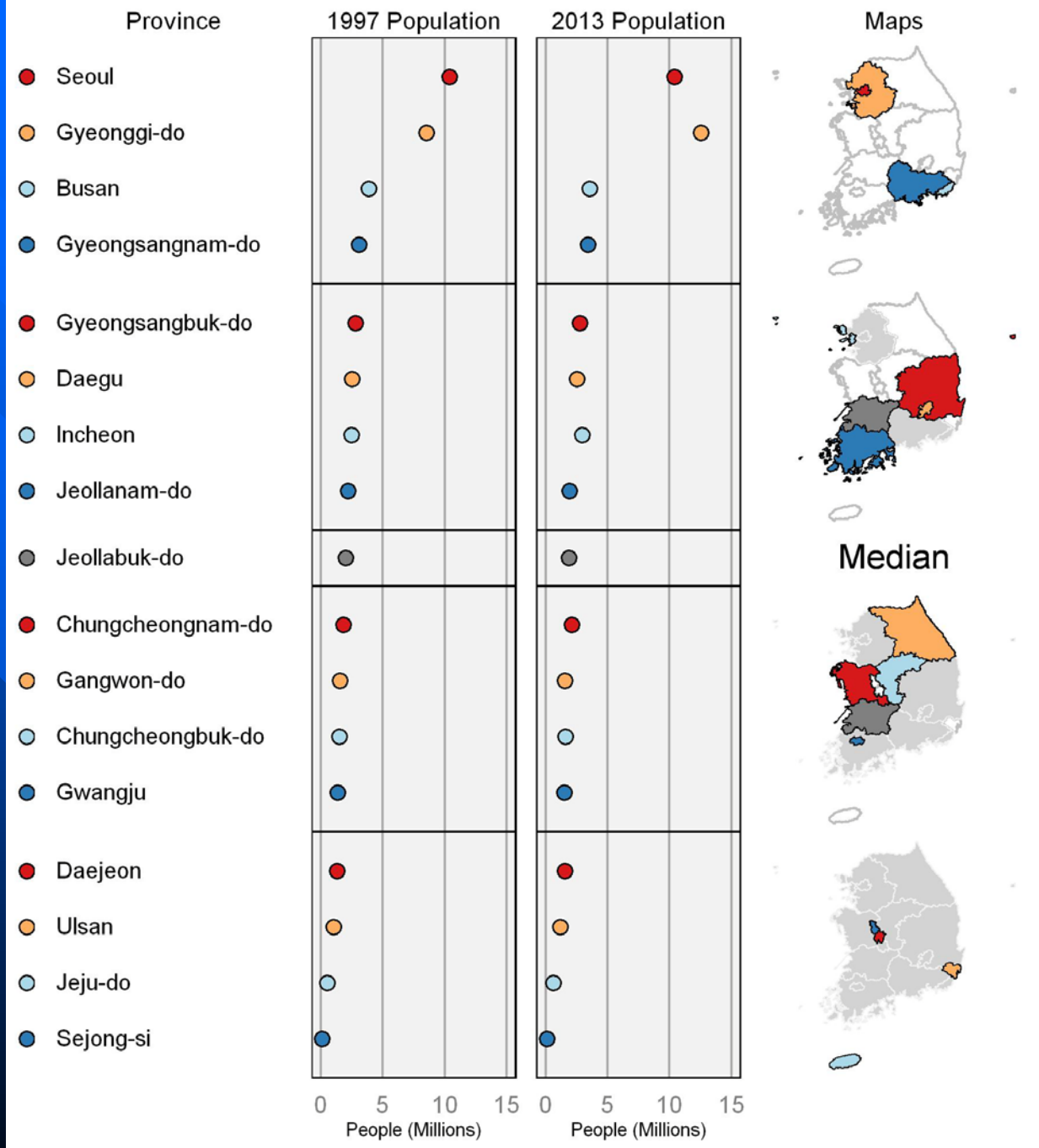
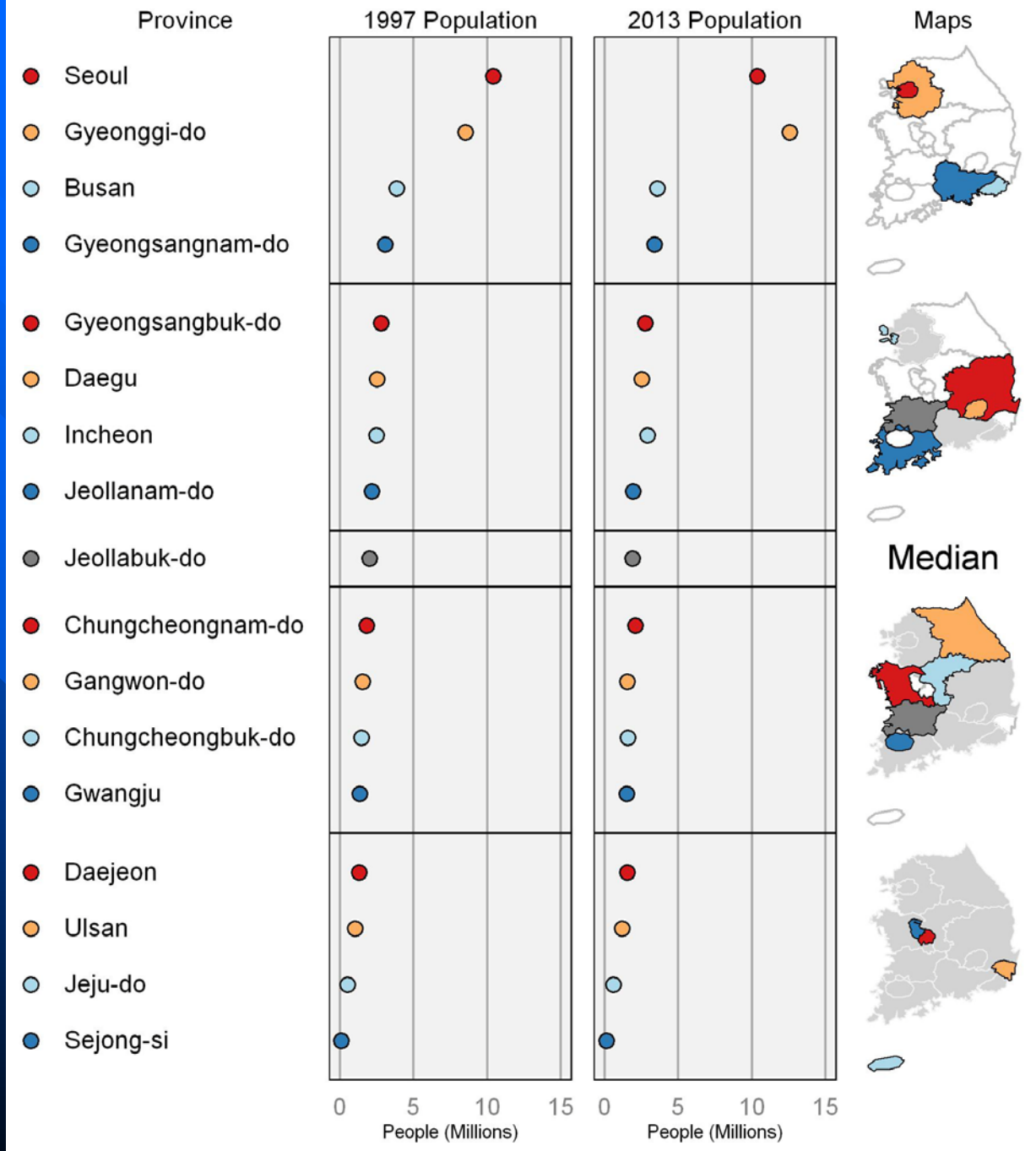


Figure Source: Symanzik, J., Dai, X., Weber, M. H., Payton, Q., McManus, M. G. (2014): Linked Micromap Plots for South America - General Design Considerations and Specific Adjustments, Revista Colombiana de Estadistica, 37(2):450-469.

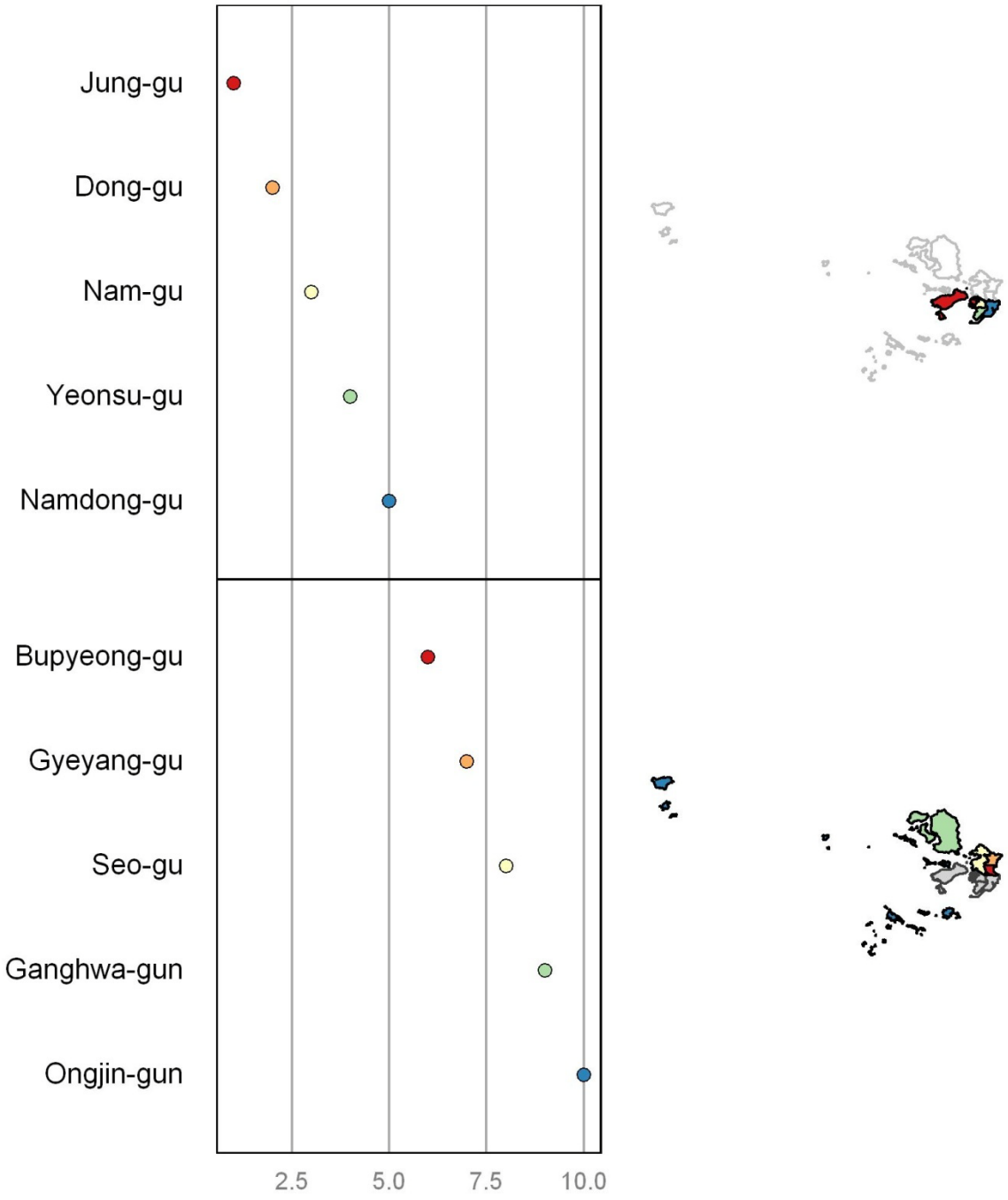
Korea Example (raw)



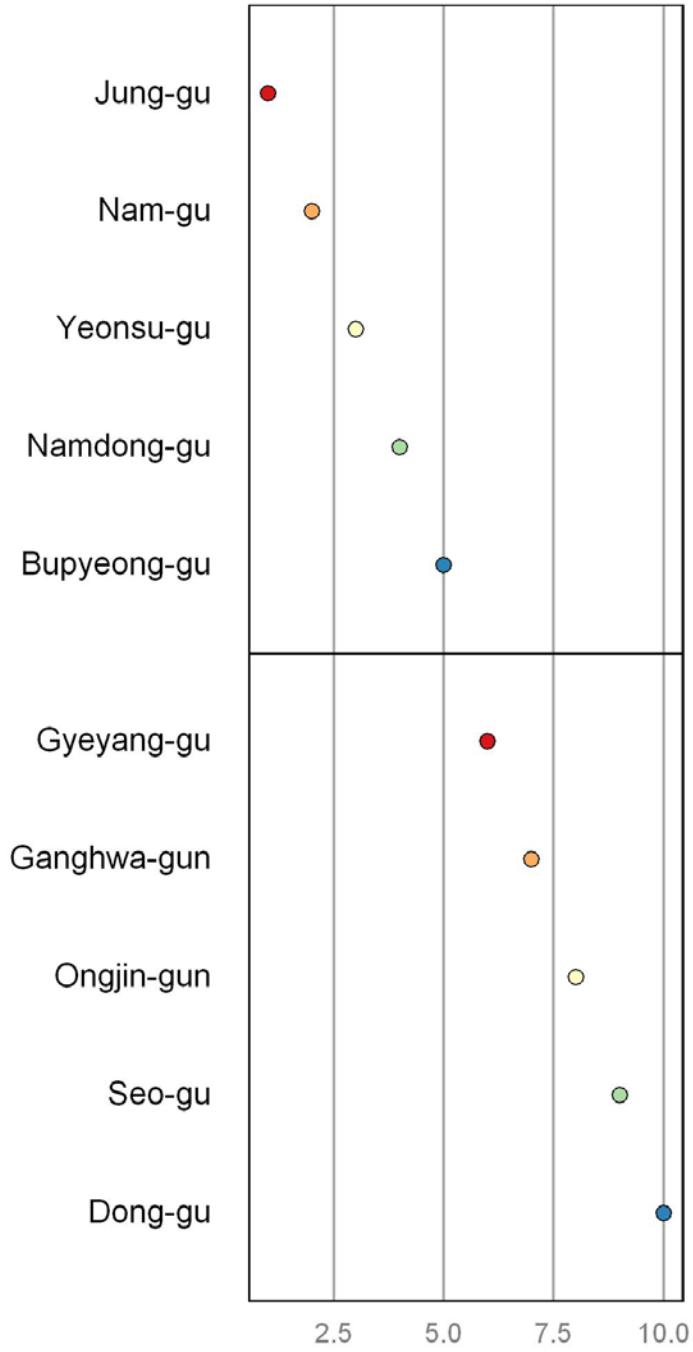
Korea Example (refined)



Korea / Incheon Example (template - raw)



Korea / Incheon Example (template - refined)



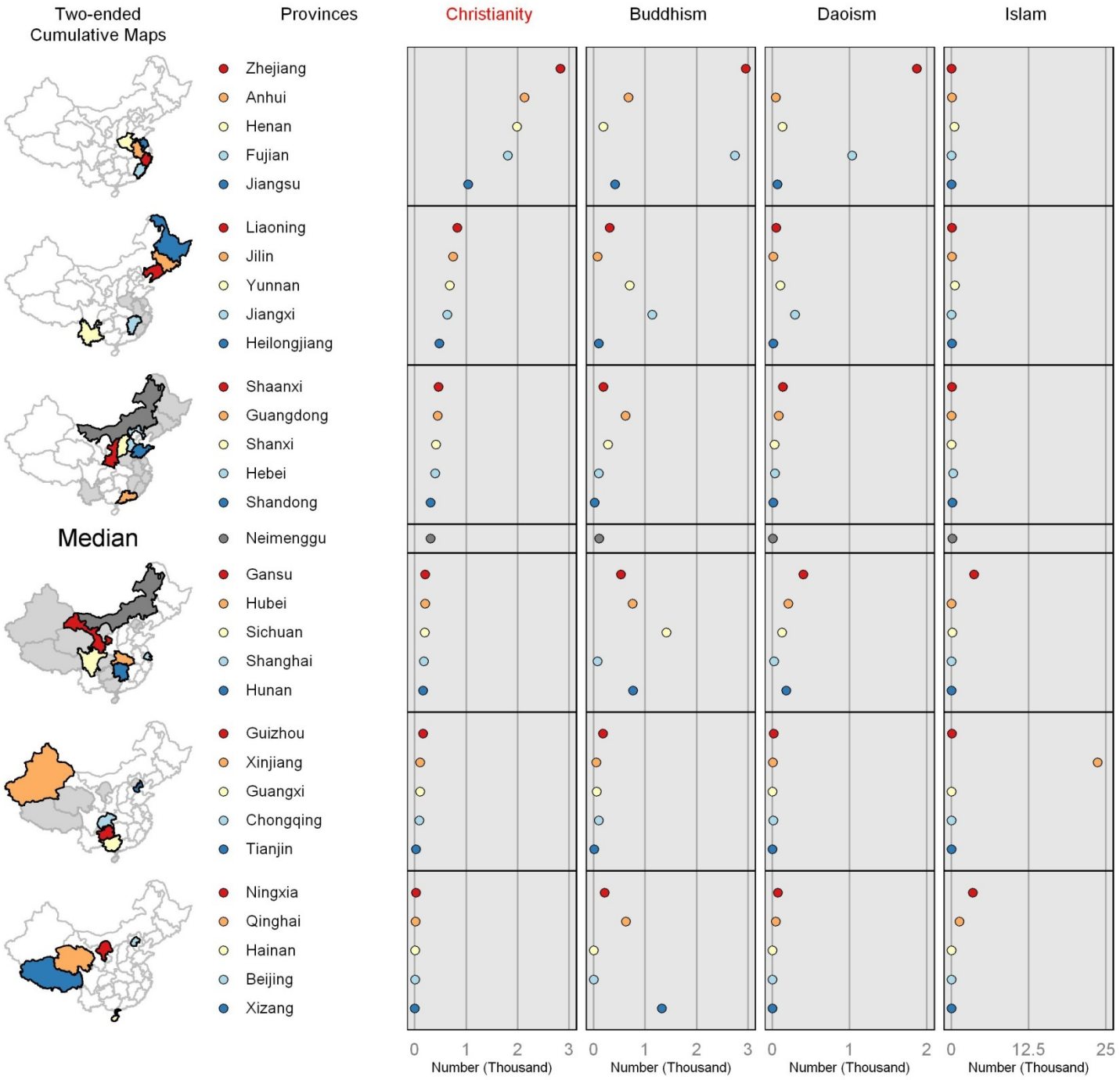
China Simplification

Raw

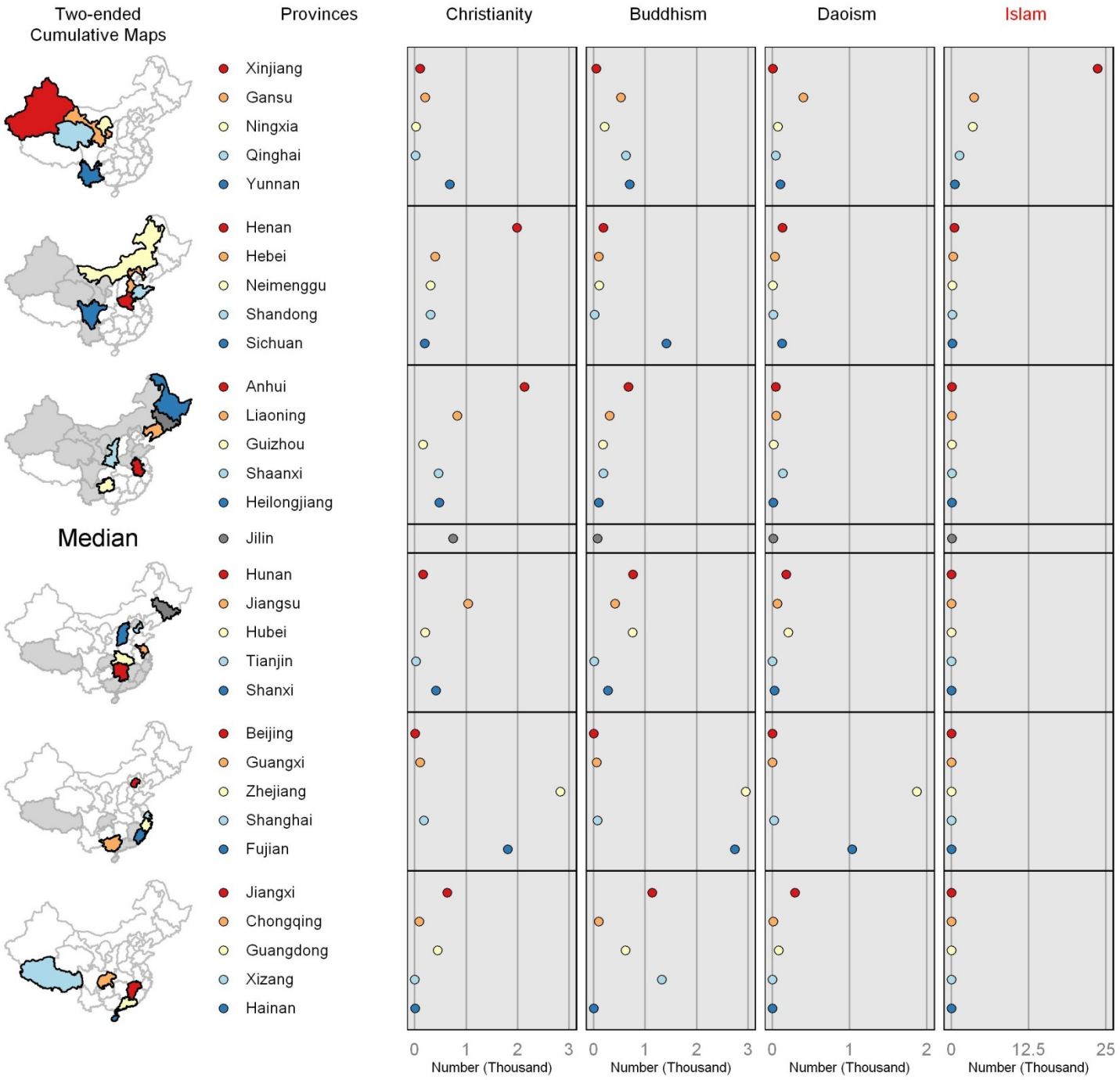
Thinned boundaries

Enlarged regions

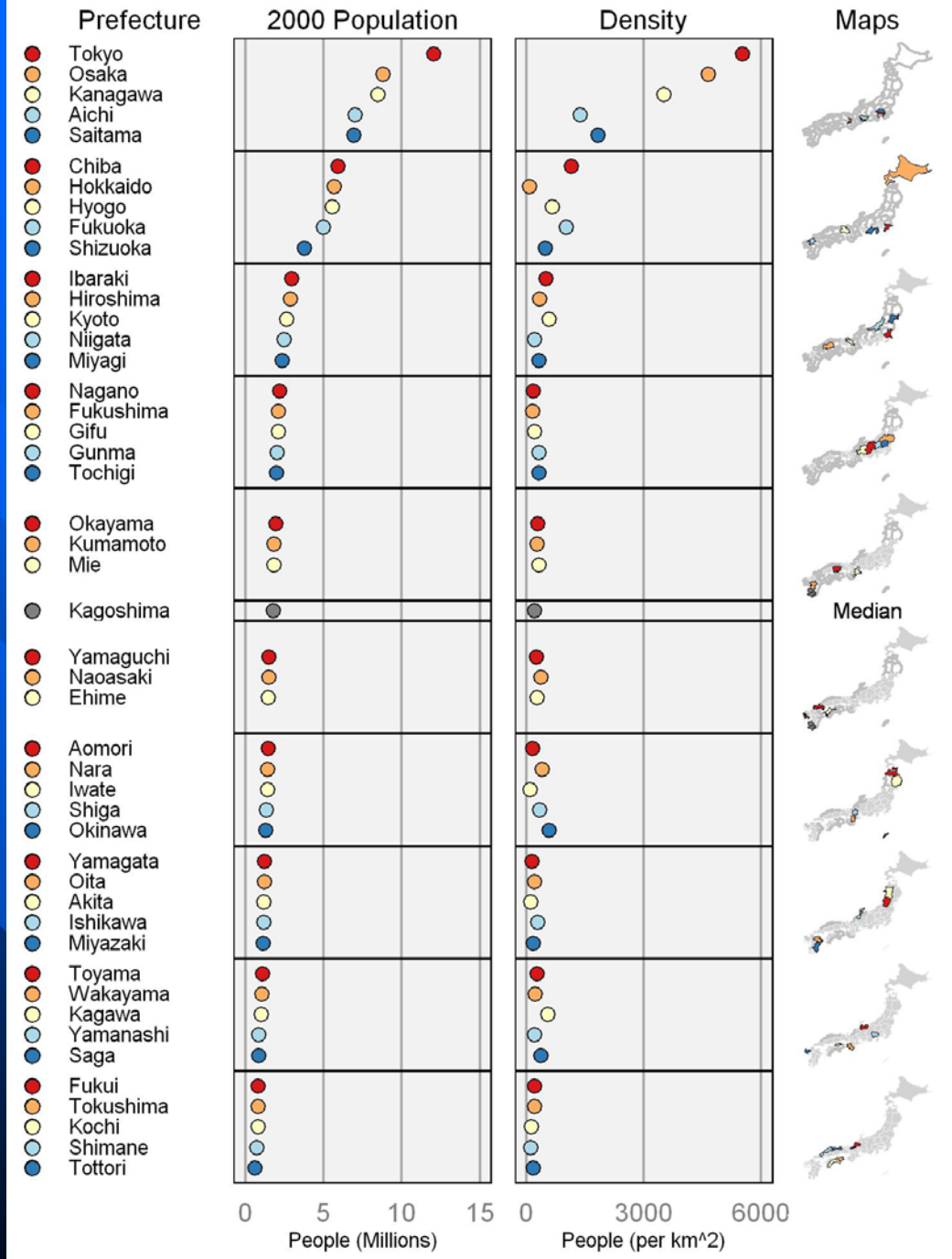
China Example (1)



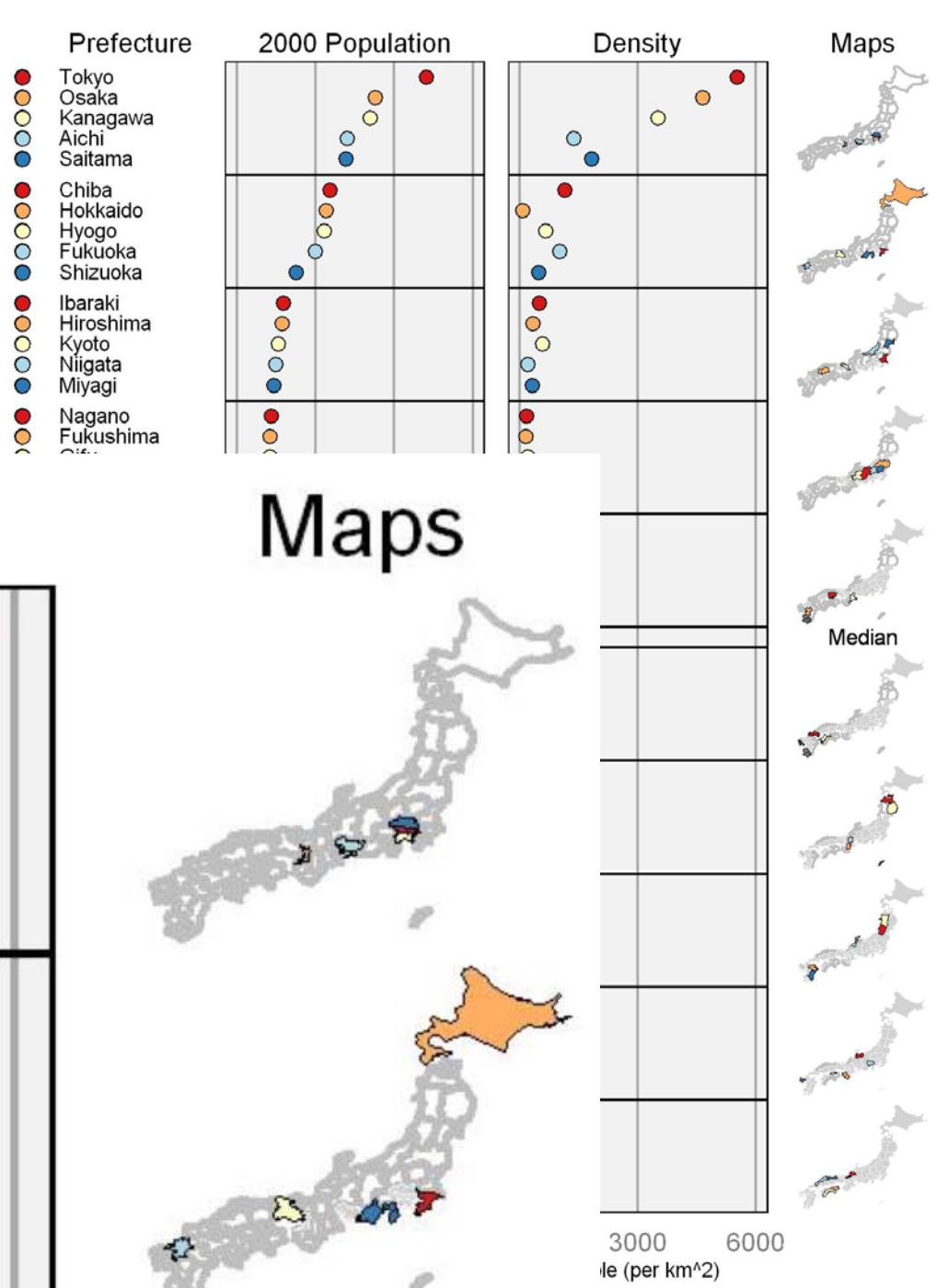
China Example (2)

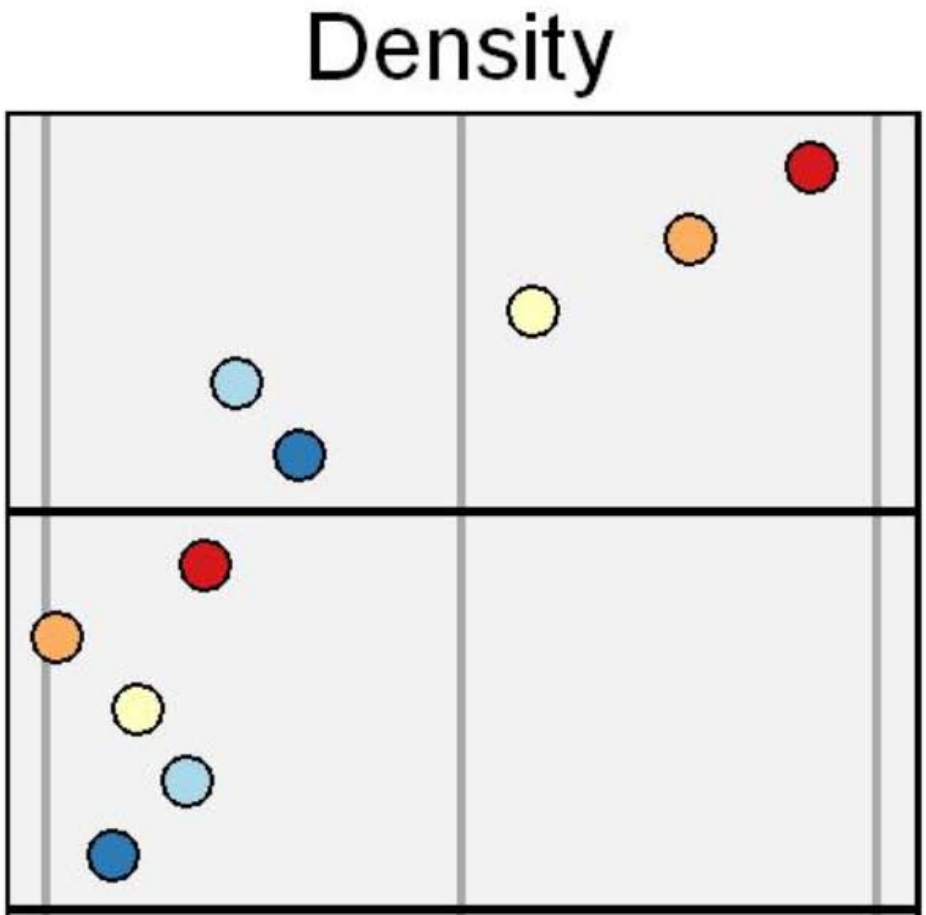


Limitation (1): Japan

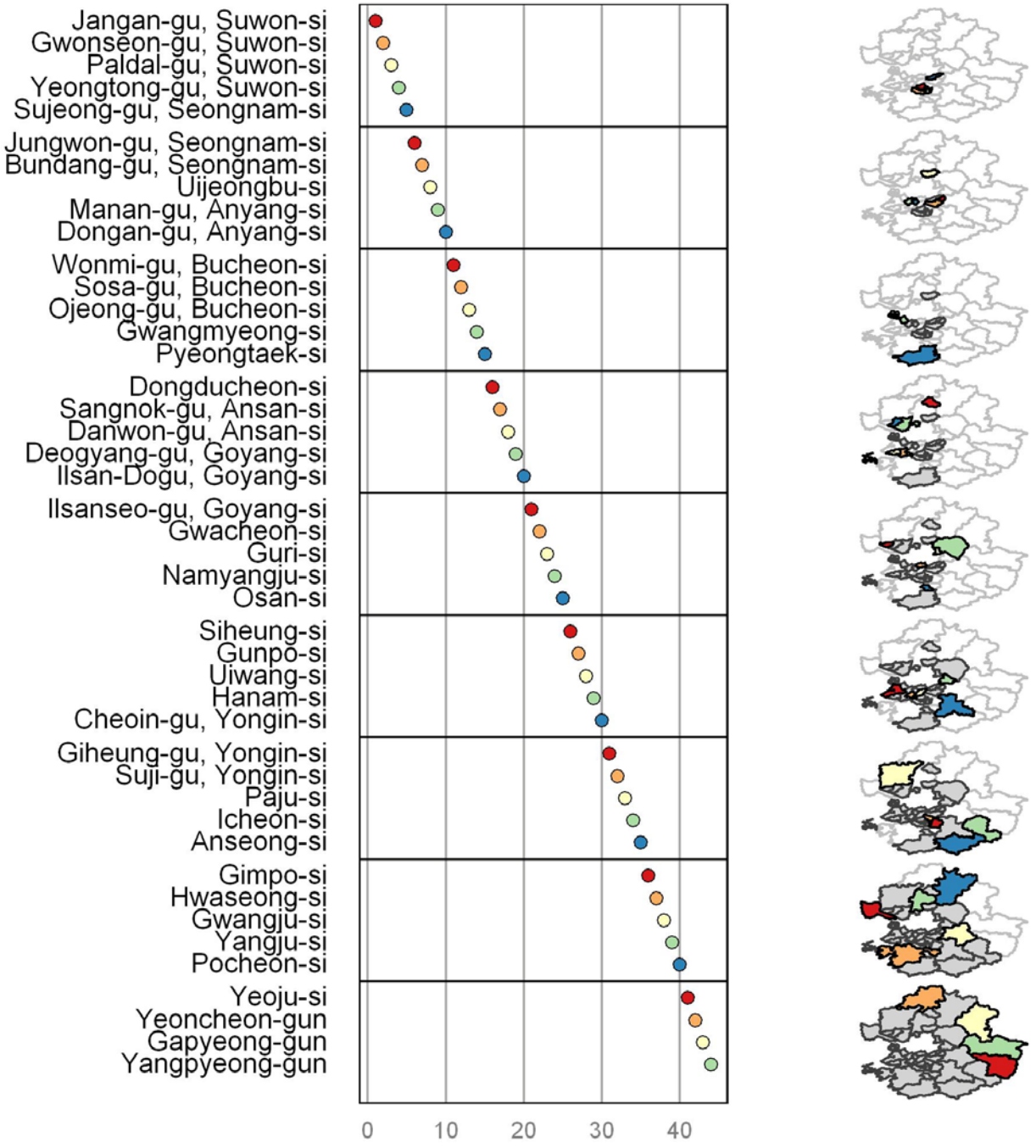


Limitation (1): Japan

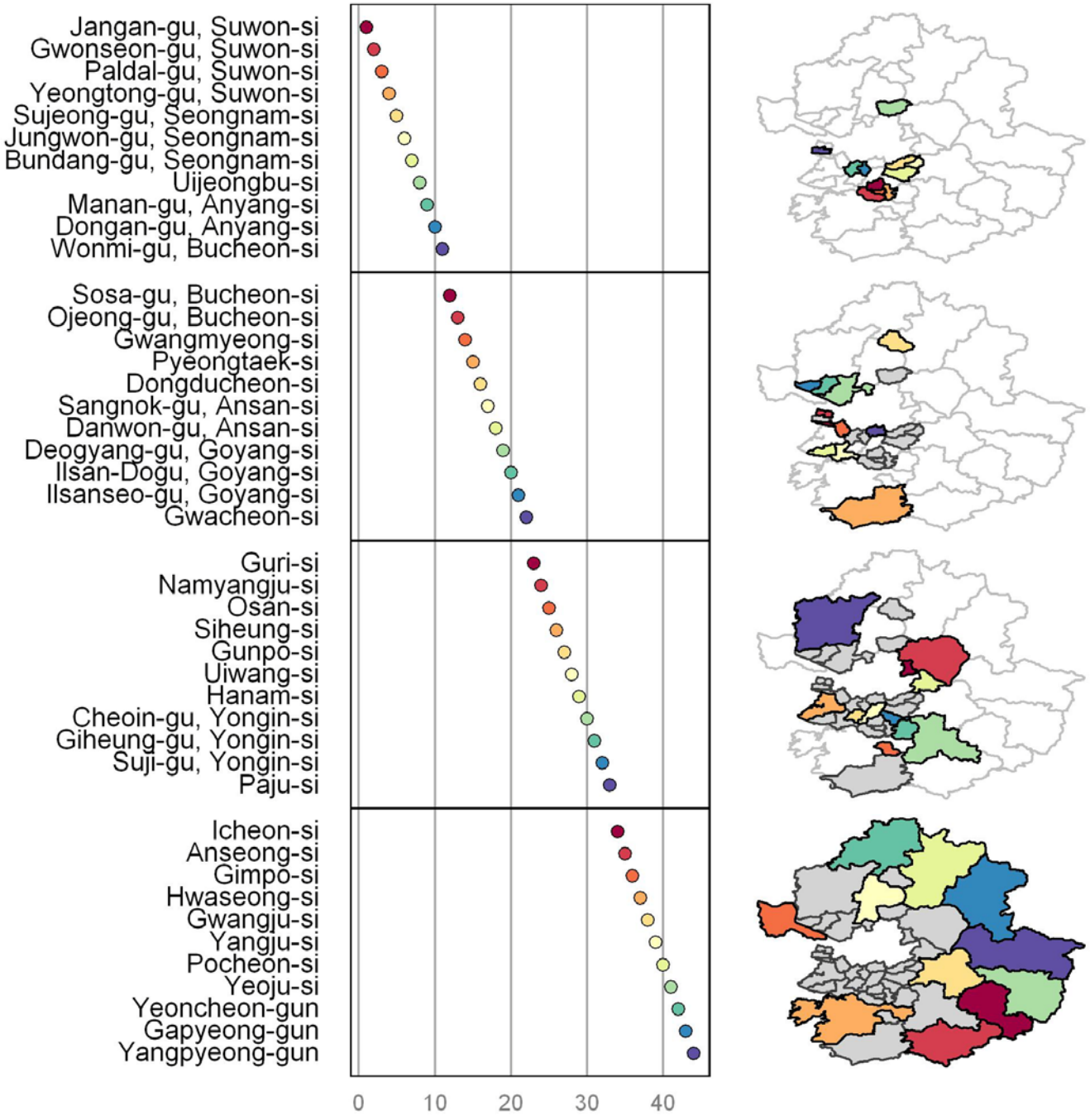




Limitation (2): Korea / Gyeonggi

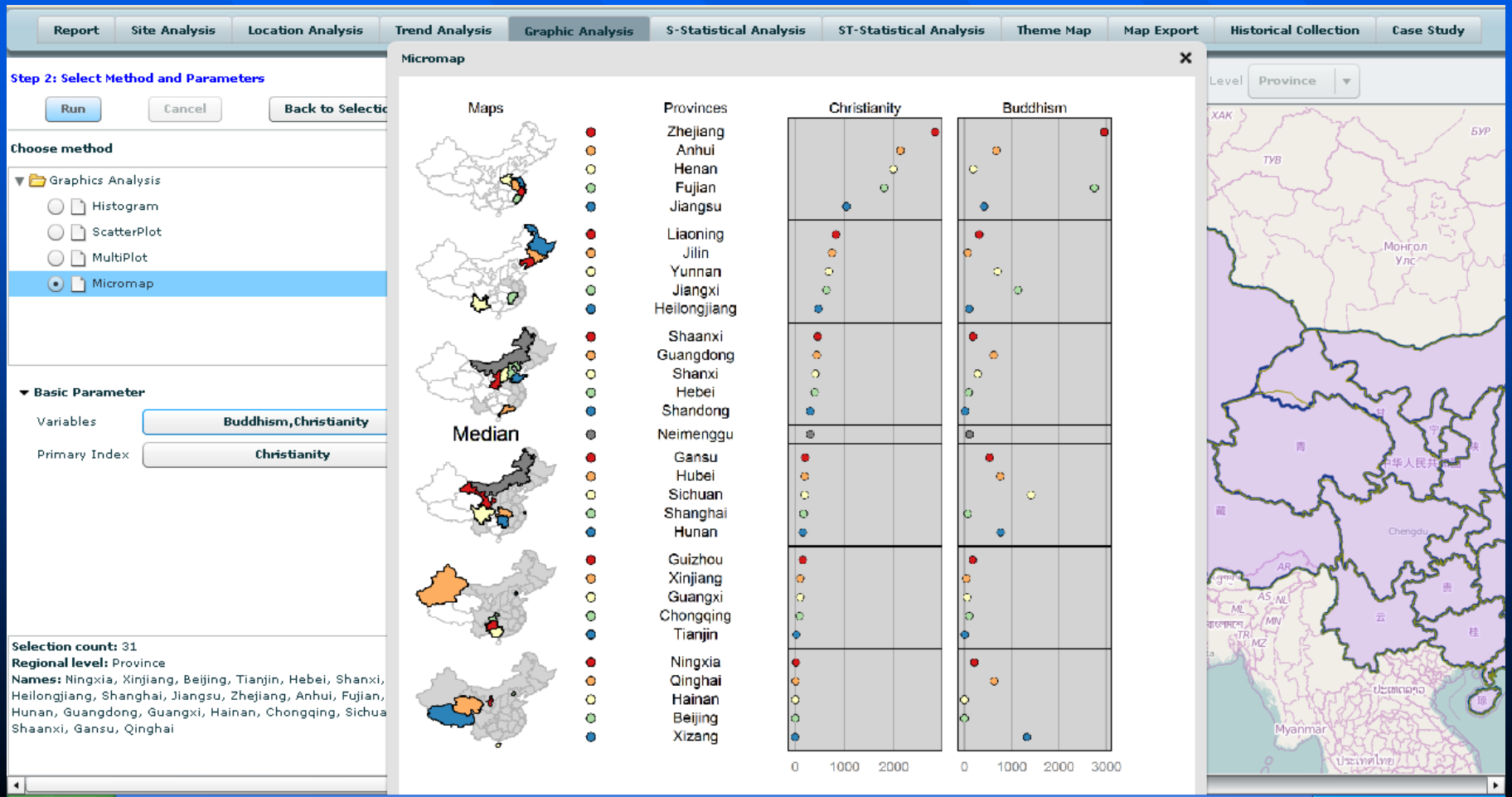


Limitation (2) / Solution Outline



Current Work:

Interactive Link: Micromaps / Religion Explorer Software at the China Data Center (CDC), University of Michigan



Future Work (1)

- Reintroduce 2-column identifier layout to micromap/micromapST R packages, based on Carr (2001)

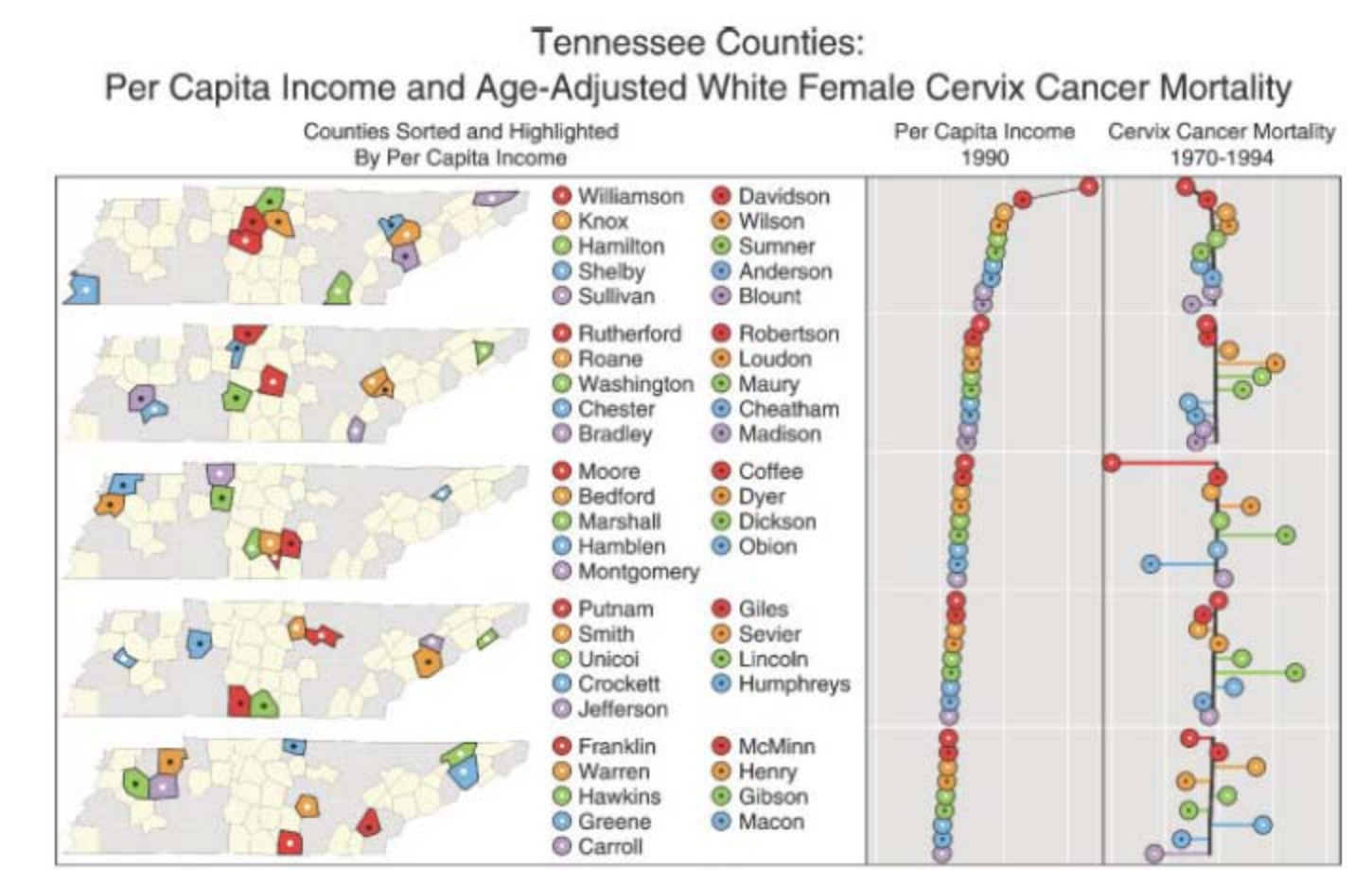


Figure Source: Carr, D. B. (2001): Designing Linked Micromap Plots for States with Many Counties, Statistics in Medicine, 20:1331-1339.

Future Work (2)

- Further extend new R functions to enlarge an area into several neighboring areas simultaneously and into the open sea
- Further enhance, test, and debug new R functions for the creation of modified shapefiles
- Add functionality to micromap R package once code is stable or release as stand-alone R package

Questions ???

– or –

send e-mail to:

symanzik@math.usu.edu