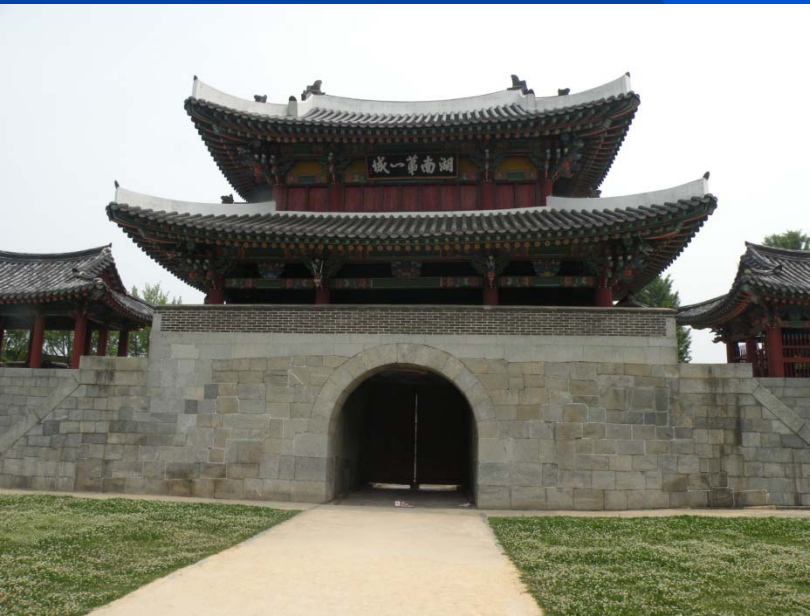


Visual Data Mining via Linked Micromap Plots in R

Jürgen Symanzik

Utah State University, Logan, UT, USA



e-mail: symanzik@math.usu.edu

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

Contents

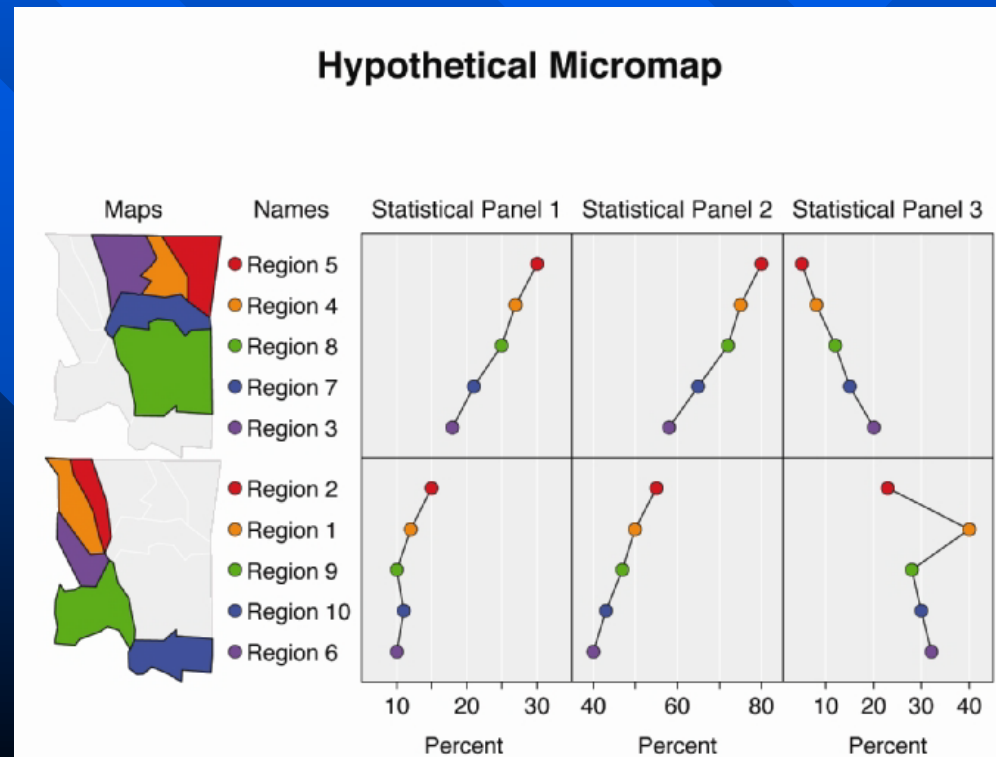
- Background
- Micromaps in R
 - Stand-alone Approaches
 - Templates in Support of Carr & Pickle (2010)
 - R Package micromap
 - R Package micromapST
- Regional Micromaps
 - Shapefile Modifications
 - Examples
- Conclusion & Outlook

Reminder: Visual Data Mining

- Working Definition for VDM:
 - Find structure (cluster, unusual observations) in large and not necessarily homogeneous data sets based on human perception using graphical methods and user interaction
 - Goal or expected outcome of exploration usually unknown in advance

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
 - economical data



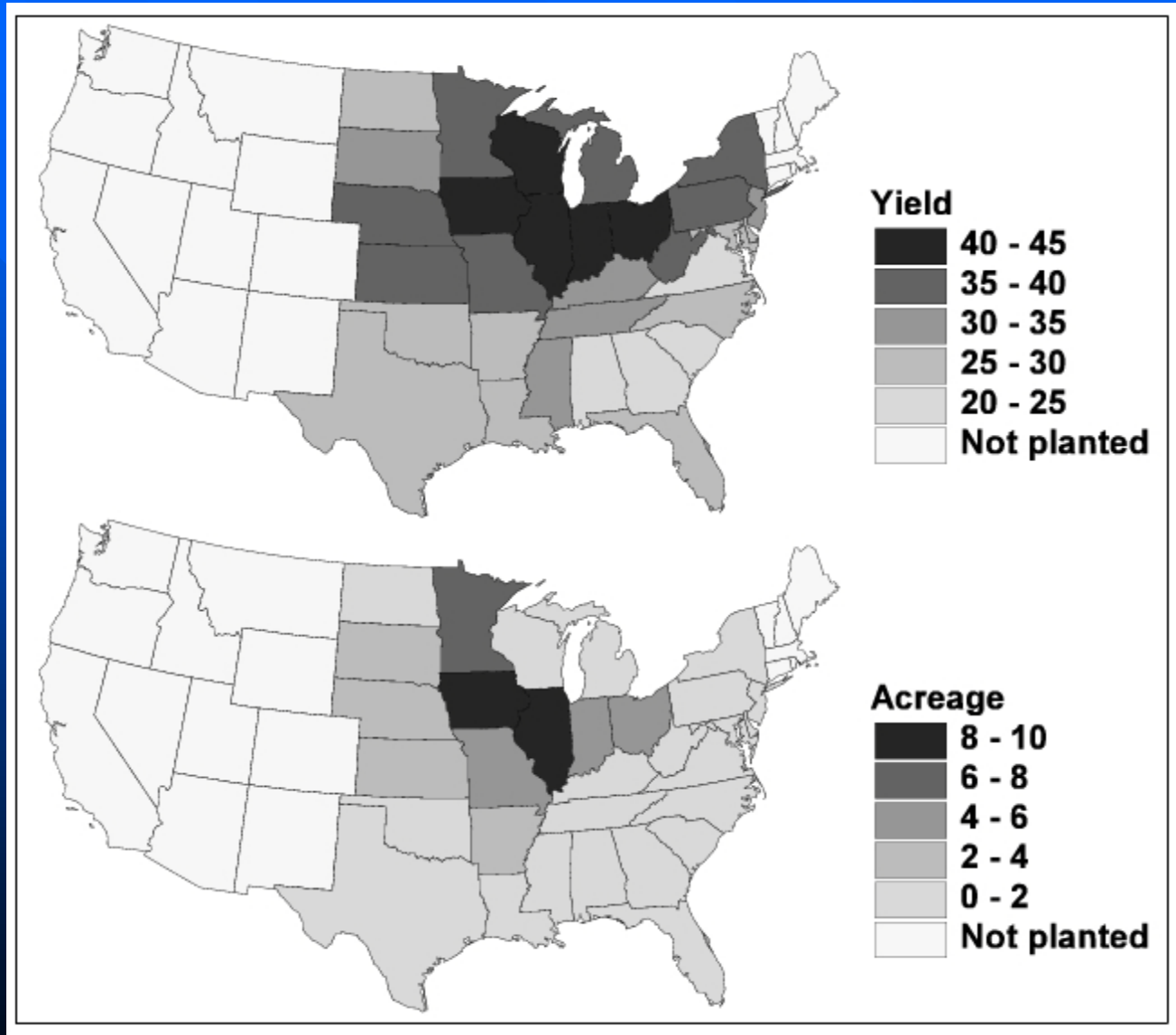
History of Micromaps

- 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:

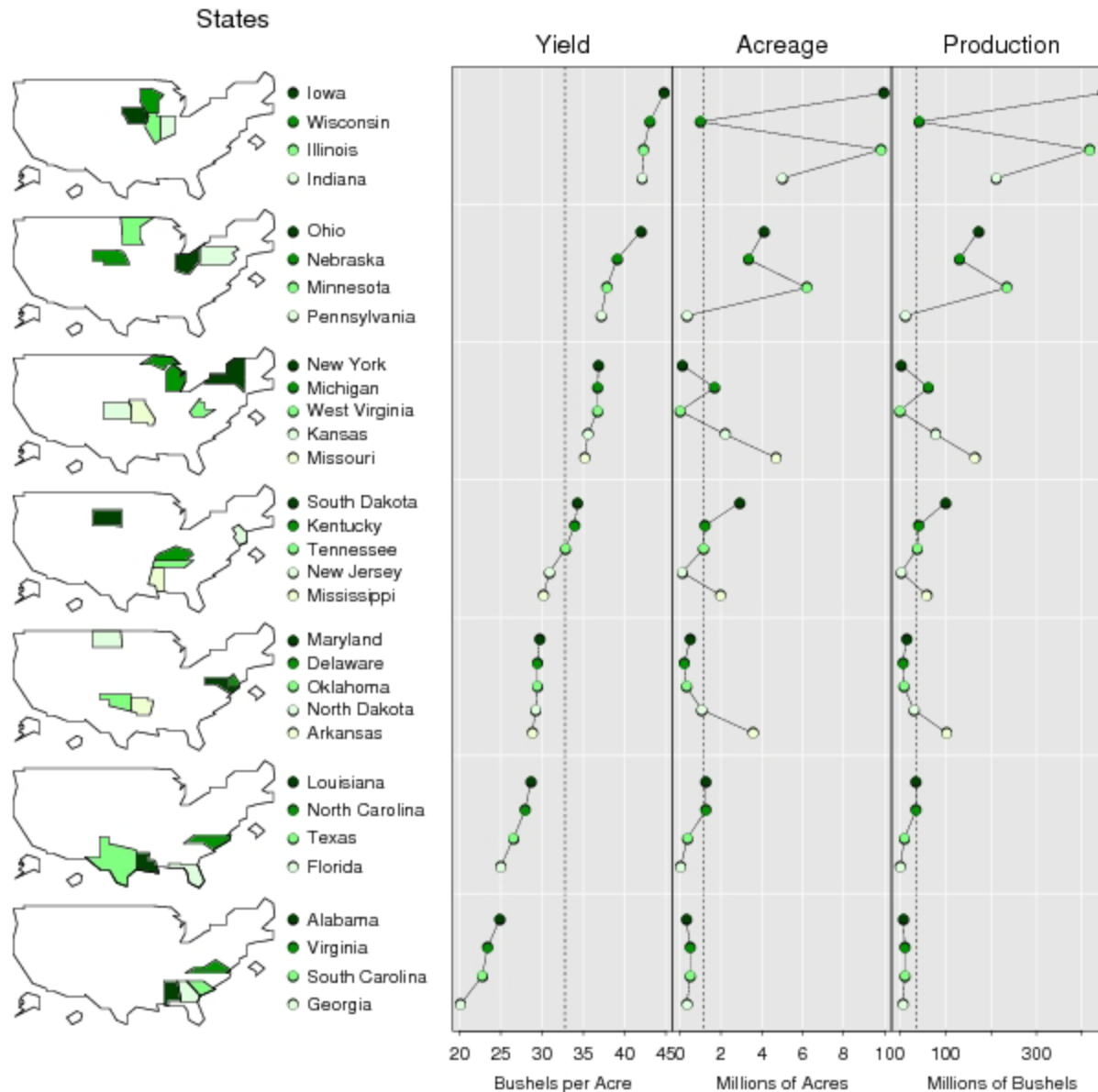
<http://stat-computing.org/newsletter/archive.html>

Choropleth Maps vs Micromaps (1)



Choropleth Maps vs Micromaps (2)

Soybean Statistics by State, 1997 Census of Agriculture



Creation of Micromaps

- Primarily created in S-Plus and Java
- In 2006, Symanzik & Carr (2008) stated:

“It seems to be possible to adapt the S-Plus micromap code for use in the R statistical package (Ihaka and Gentleman, 1996), which can be freely obtained from <http://www.r-project.org/>. Although no full implementation of a micromap library in R exists at this point, the basic panel functions and simple examples of LM plots have been converted from S-Plus to R by Anthony R. Olsen.”

Stand-alone Approaches

- Mike Minnotte (~2006/2007):
 - Created basic micromap script for his graduate-level “Graphical Methods” course at Utah State University (USU)
 - Based on R command “layout” and R package “maps”
- Brian Diggs (2011):
 - Sample code, posted online at <https://groups.google.com/forum/?fromgroups#!topic/ggplot2/djLY7AeCd7Y>
 - Based on R packages “ggplot2” and “maps”
- Nathan Voge (2011):
 - Simplified, documented, and extended existing R code for his MS Project at USU

Templates in Support of Carr & Pickle (2010)

- R functions and scripts for linked micromap plots and comparative micromap plots
- Freely accessible at <http://mason.gmu.edu/~dcarr/Micromaps/>
- Central part of the implementation:
 - Set of panel functions, such as `panelLayout`, `panelSelect`, `panelScale`, `panelFill`, and `panelGrid`
 - Allows users to specify details of the overall plotting area of each component of the plot — and then draw the desired subplot into this plotting area
 - User has to loop through each of the various panels (for maps, identifiers, and statistical data) and each of the rows of the data and explicitly indicate which information (dots, names, regions in a map) has to be plotted in that panel

R Package micromap

- Developed by Payton, McManus, Weber, Olsen, and Kincaid (2012/2013)
- Based on R package “ggplot2”
- Not necessary to loop through each column and row of the data to create a cell in the LM plot
- Rather, main functionality is hidden in two main functions: “mmpplot” and “mmgroupedplot”

R Package micromapST

- Developed by Carr, Pearson Jr., and Pickle (2013)
- Focuses exclusively on special linked map design for U.S. states
- Input simplification replaces writing or editing a long script
- Two required input arguments are:
 - Data frame with a row for each state
 - Panel description data frame with a row for each column in the linked micromap

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
 - Create R code for the remaining tasks

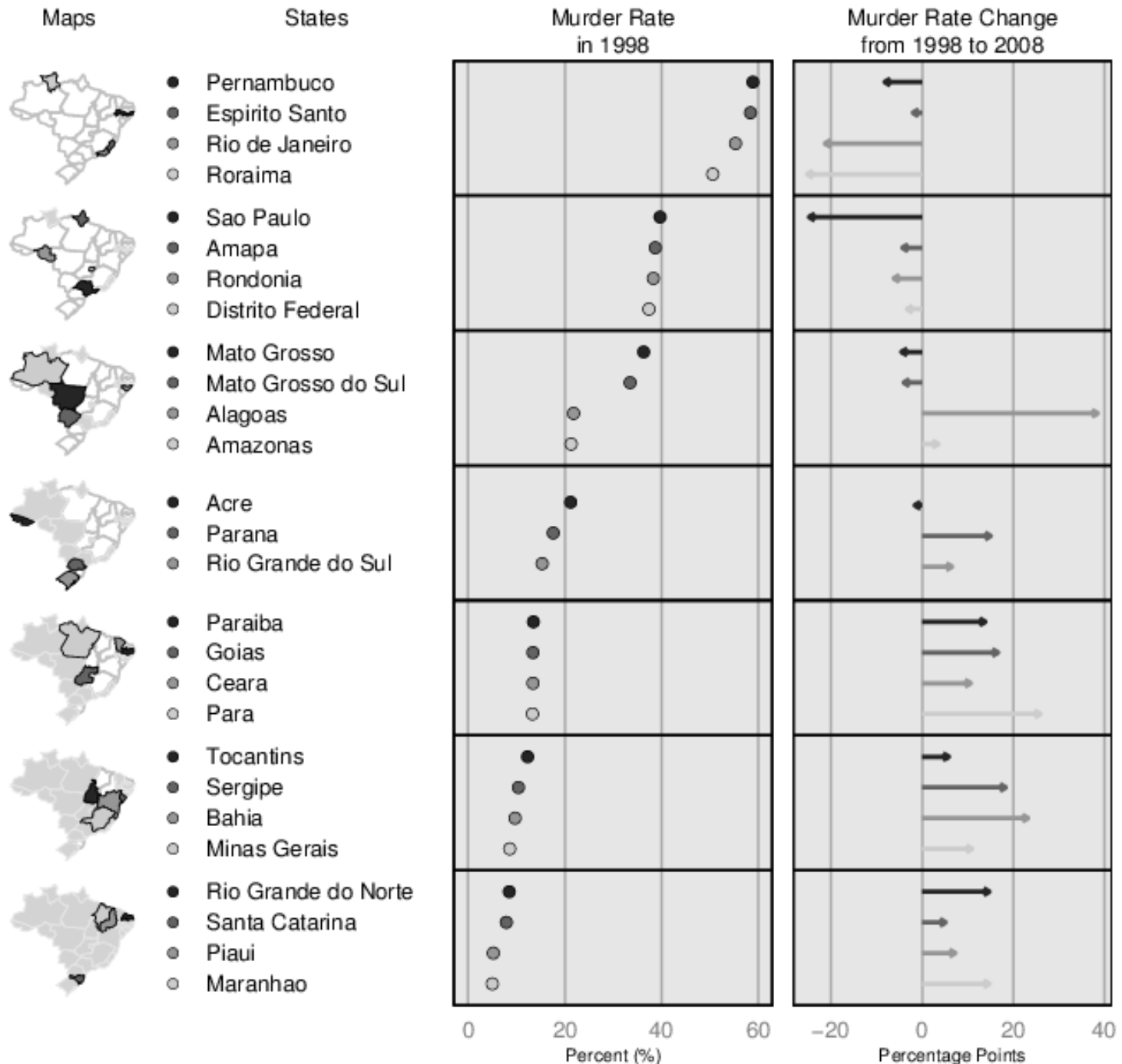
Brazil

Simplification

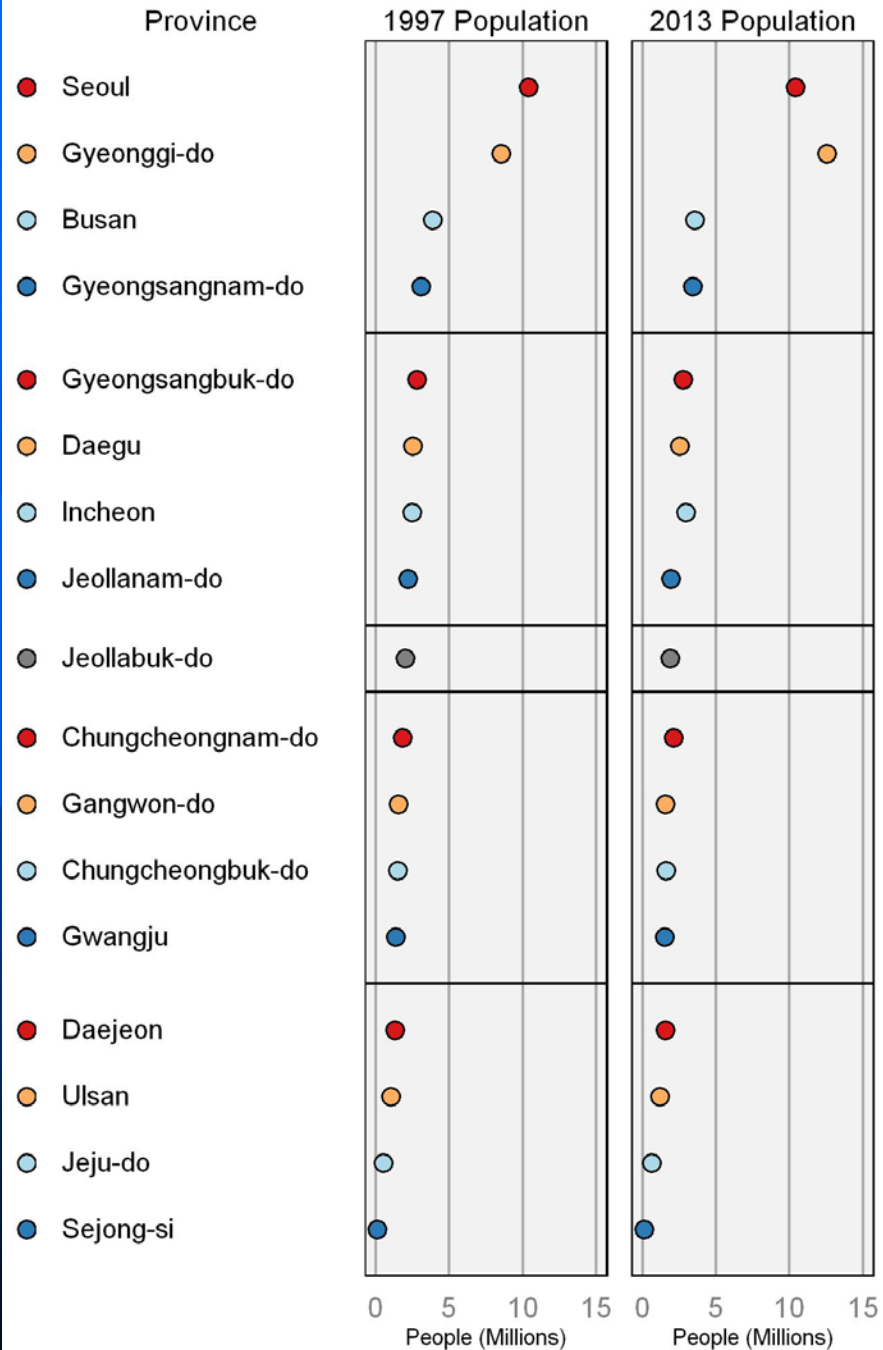
- UL: Raw
- UR: Thinned boundaries
- LL: Capital region enlarged
- LR: Capital region enlarged & shifted



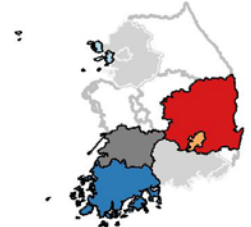
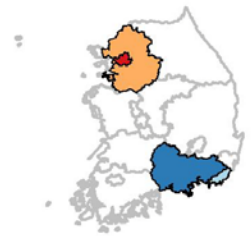
Brazil Example



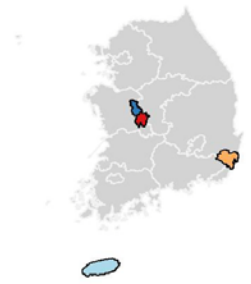
Korea Example (raw)



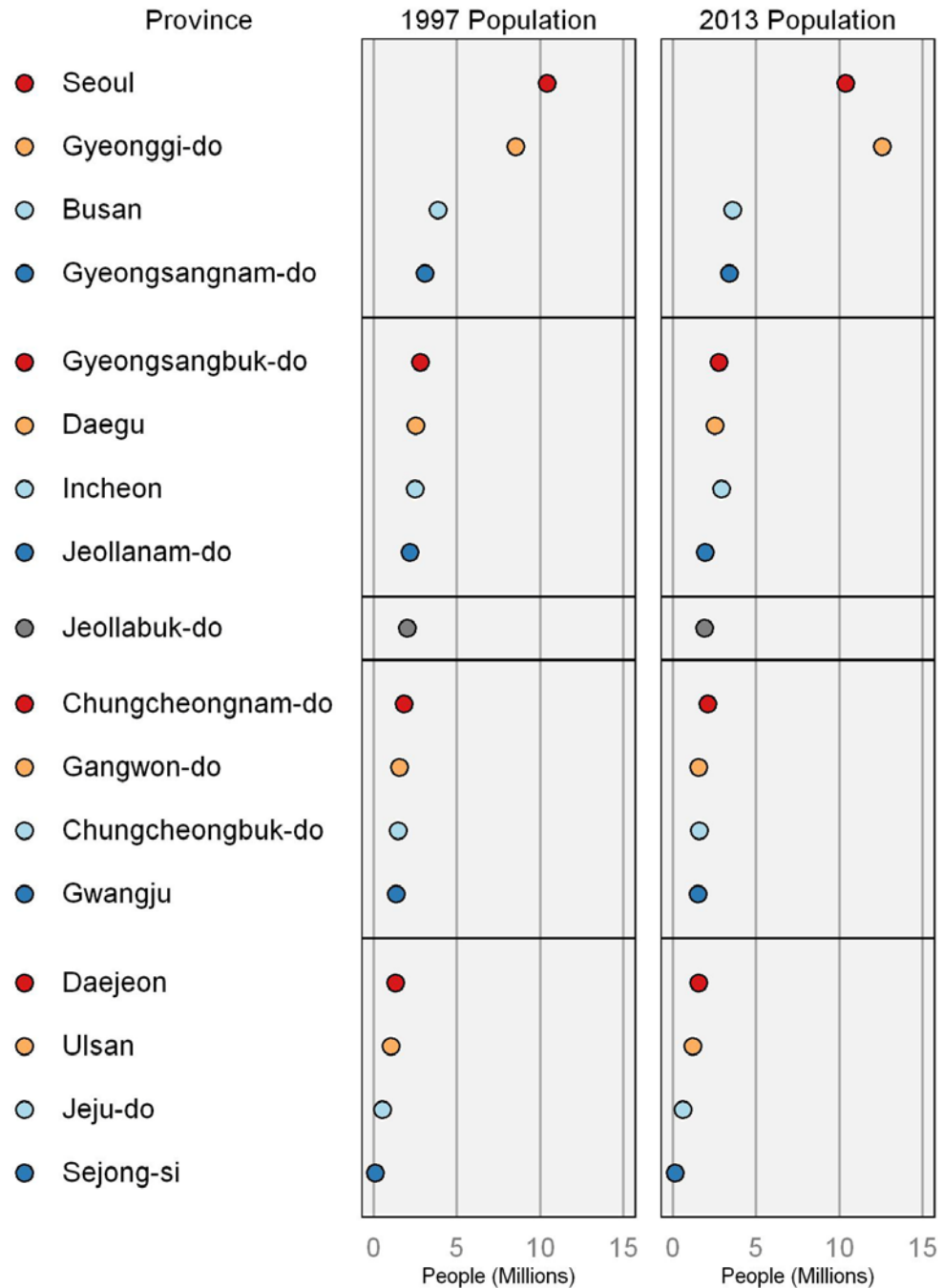
Maps



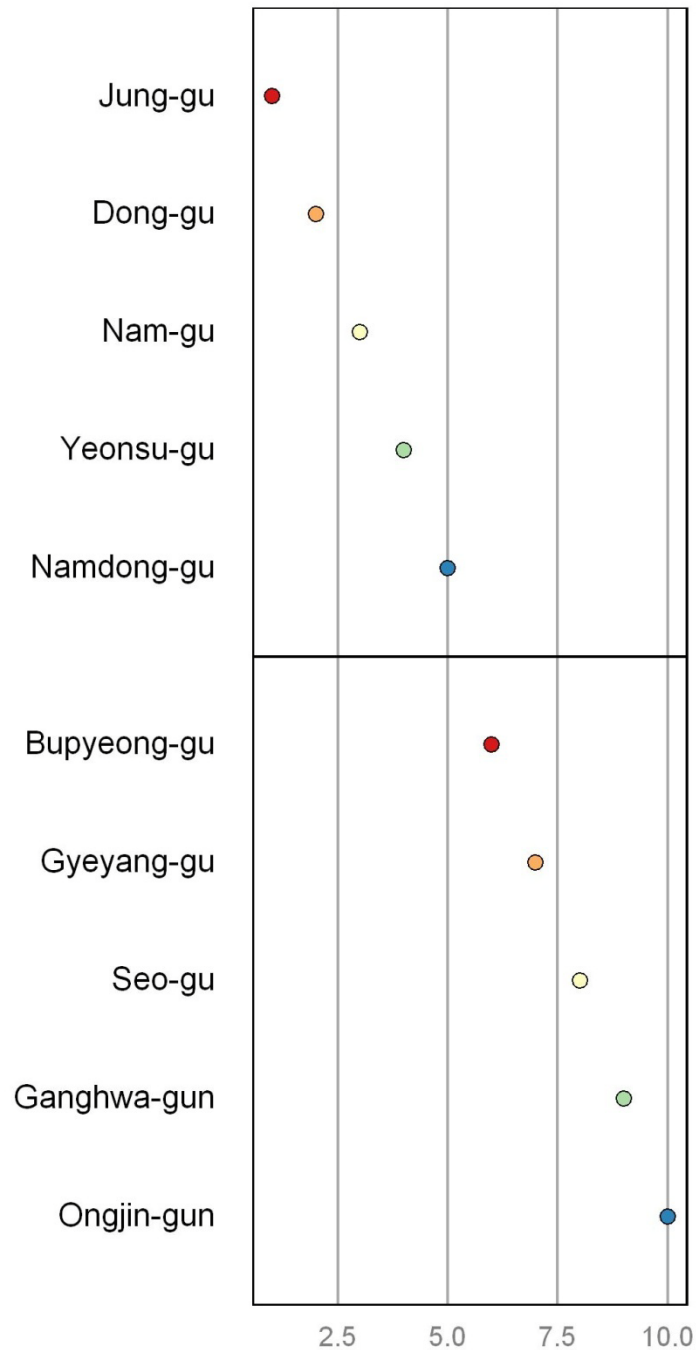
Median



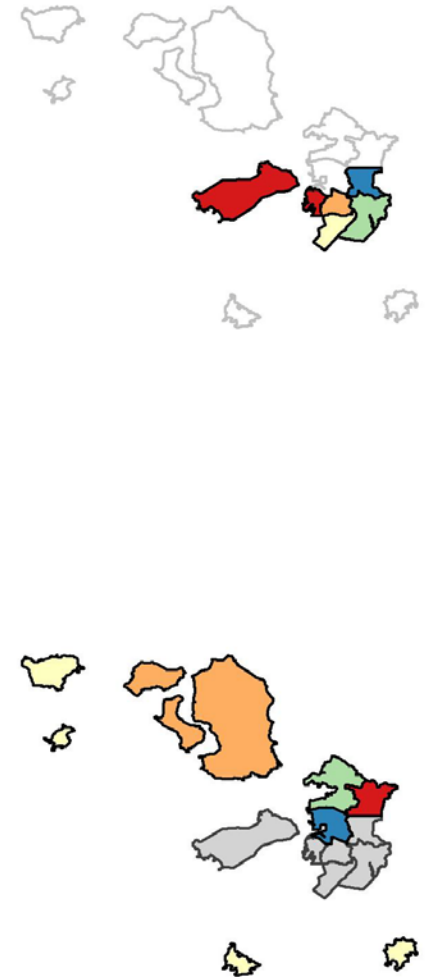
Korea Example (refined)



Korea / Incheon Example (template - raw)

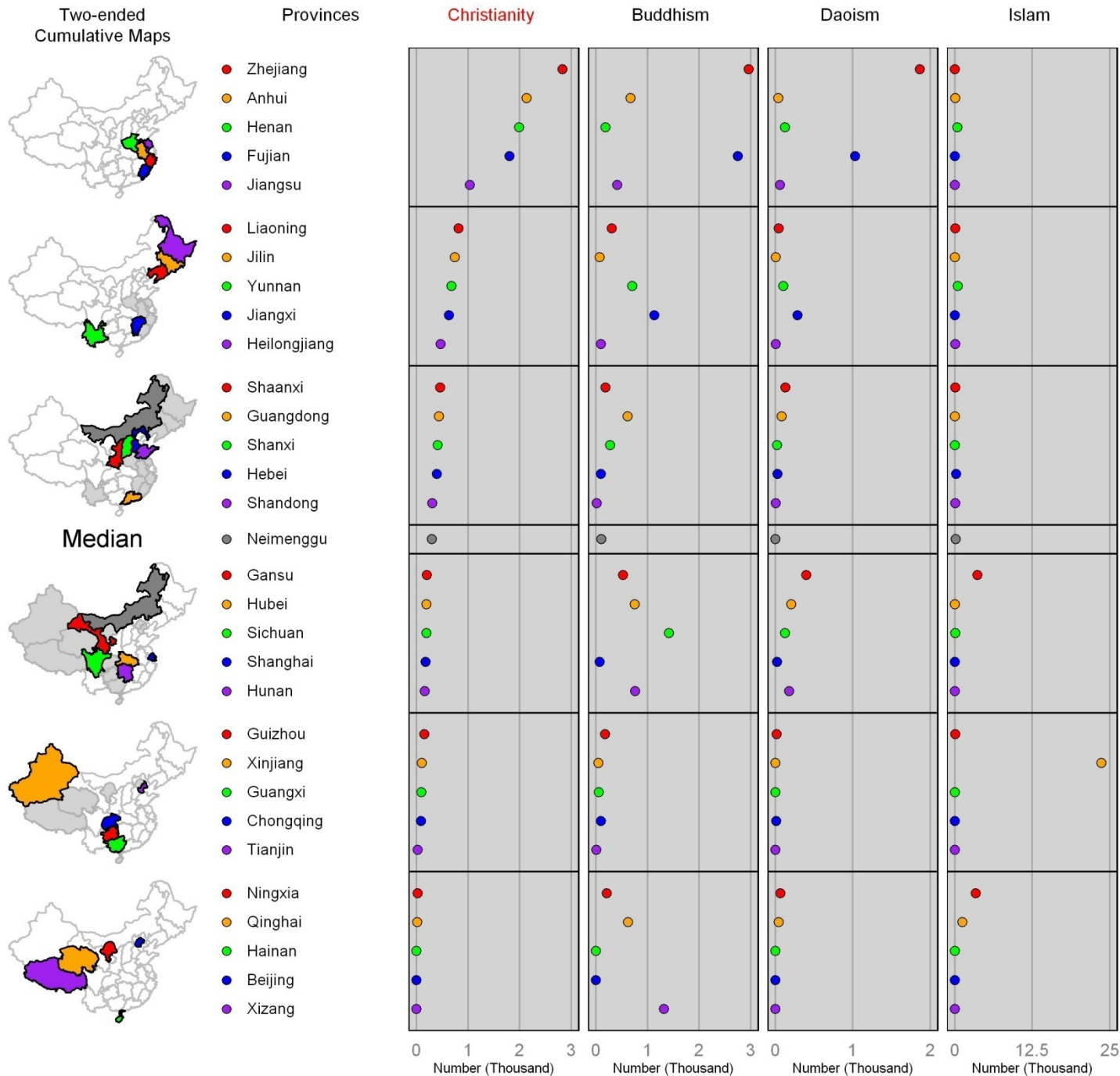


Korea / Incheon Example (template - refined)

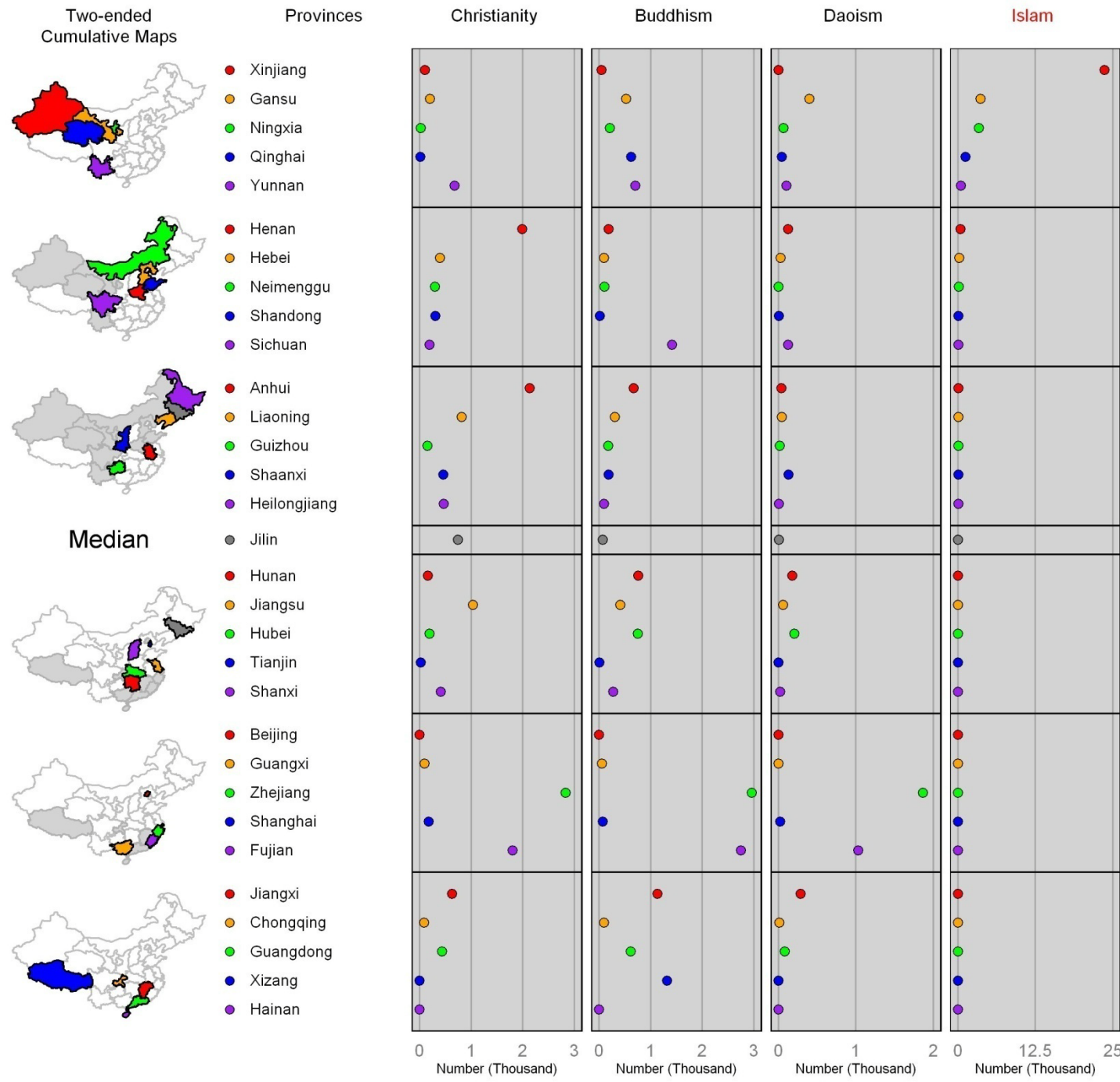


China Example

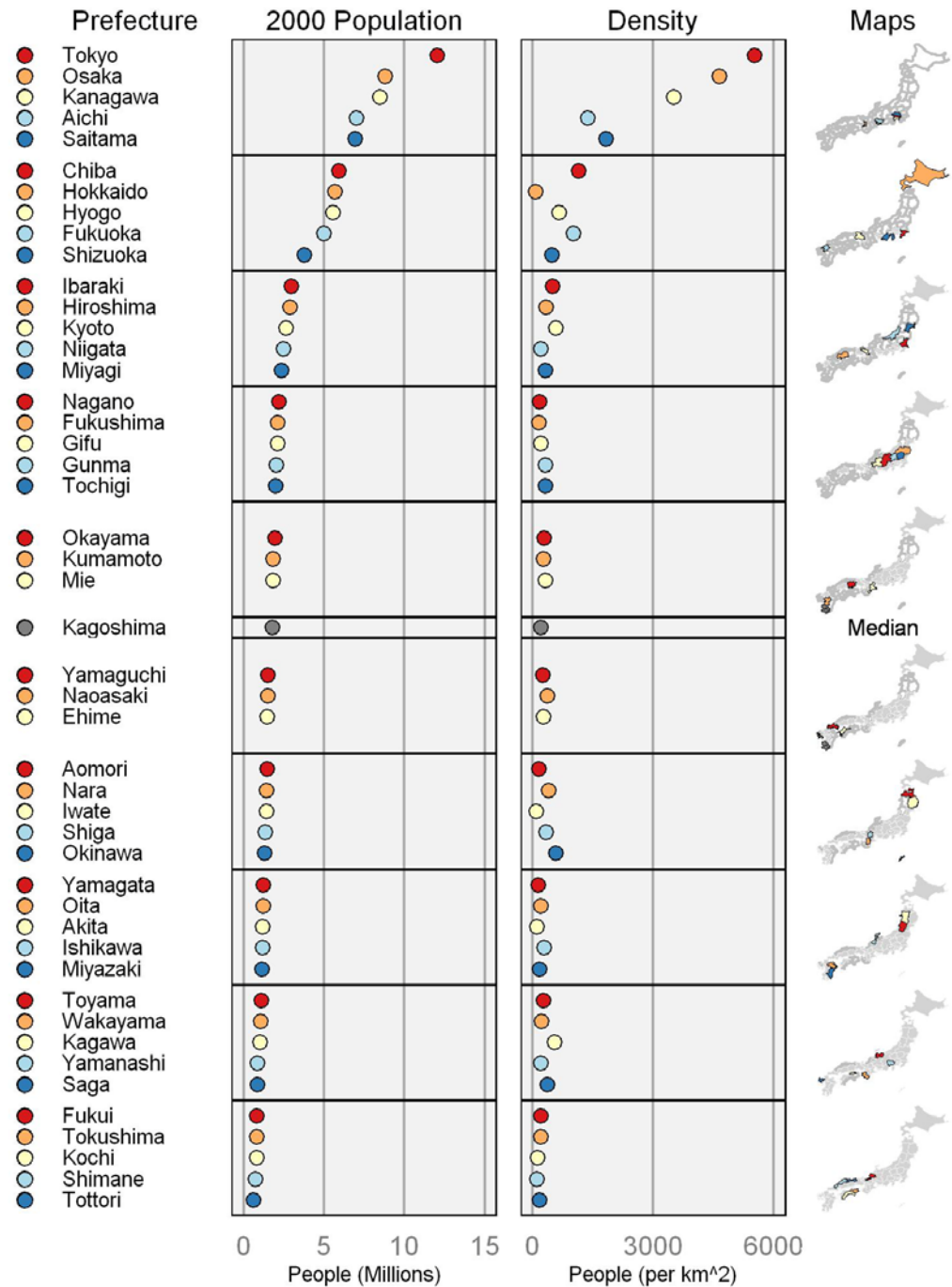
(based on data from the China Data Center, University of Michigan, USA)



China Example (2)

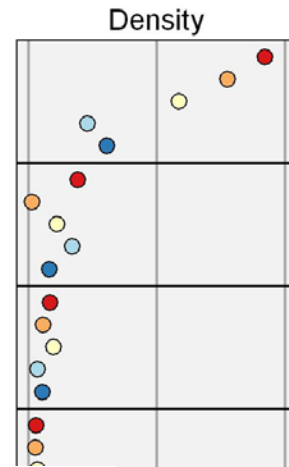
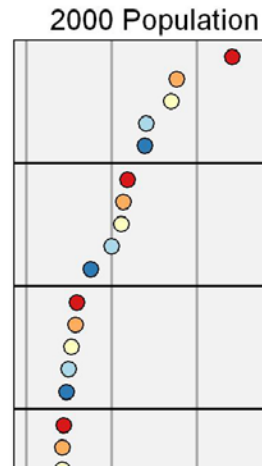


Japan Example

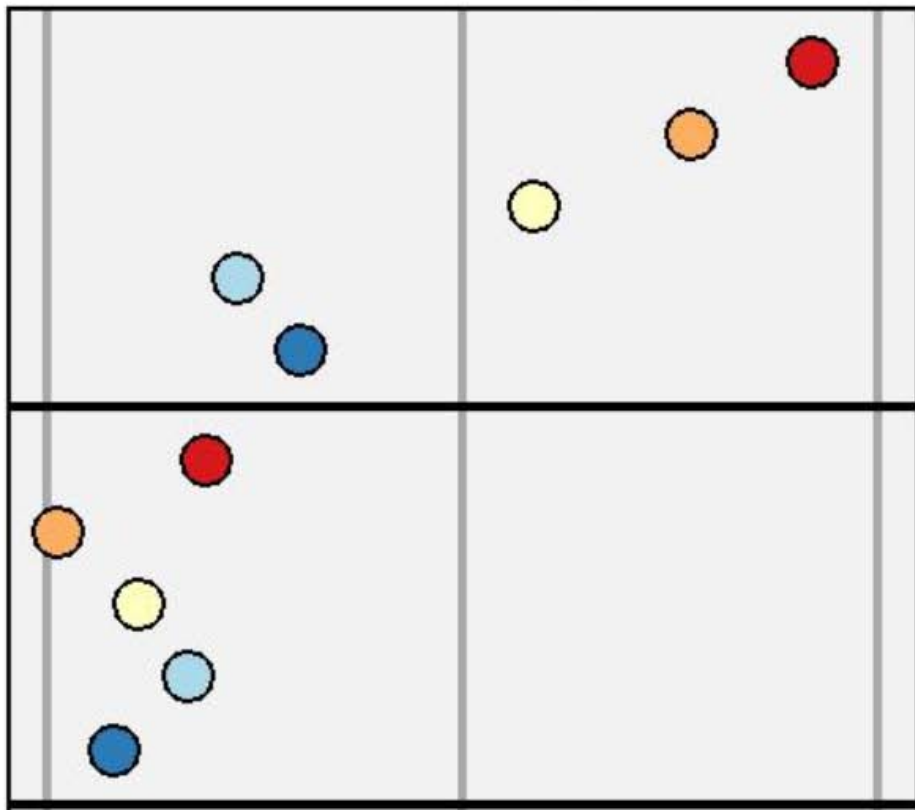


Japan Example

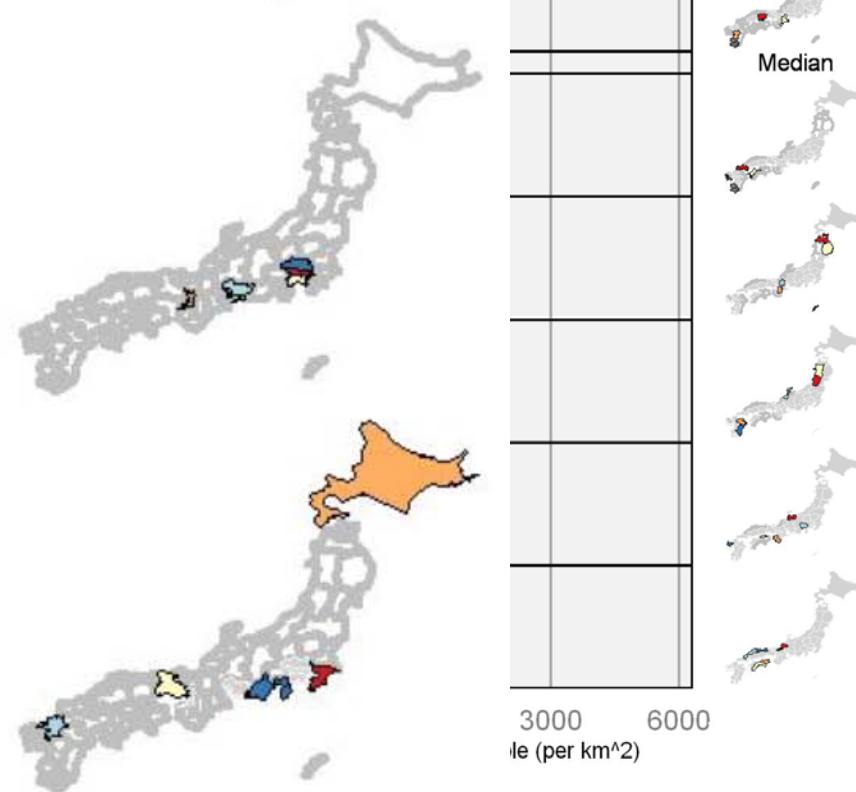
- Prefecture
- Tokyo
 - Osaka
 - Kanagawa
 - Aichi
 - Saitama
 - Chiba
 - Hokkaido
 - Hyogo
 - Fukuoka
 - Shizuoka
 - Ibaraki
 - Hiroshima
 - Kyoto
 - Niigata
 - Miyagi
 - Nagano
 - Fukushima



Density



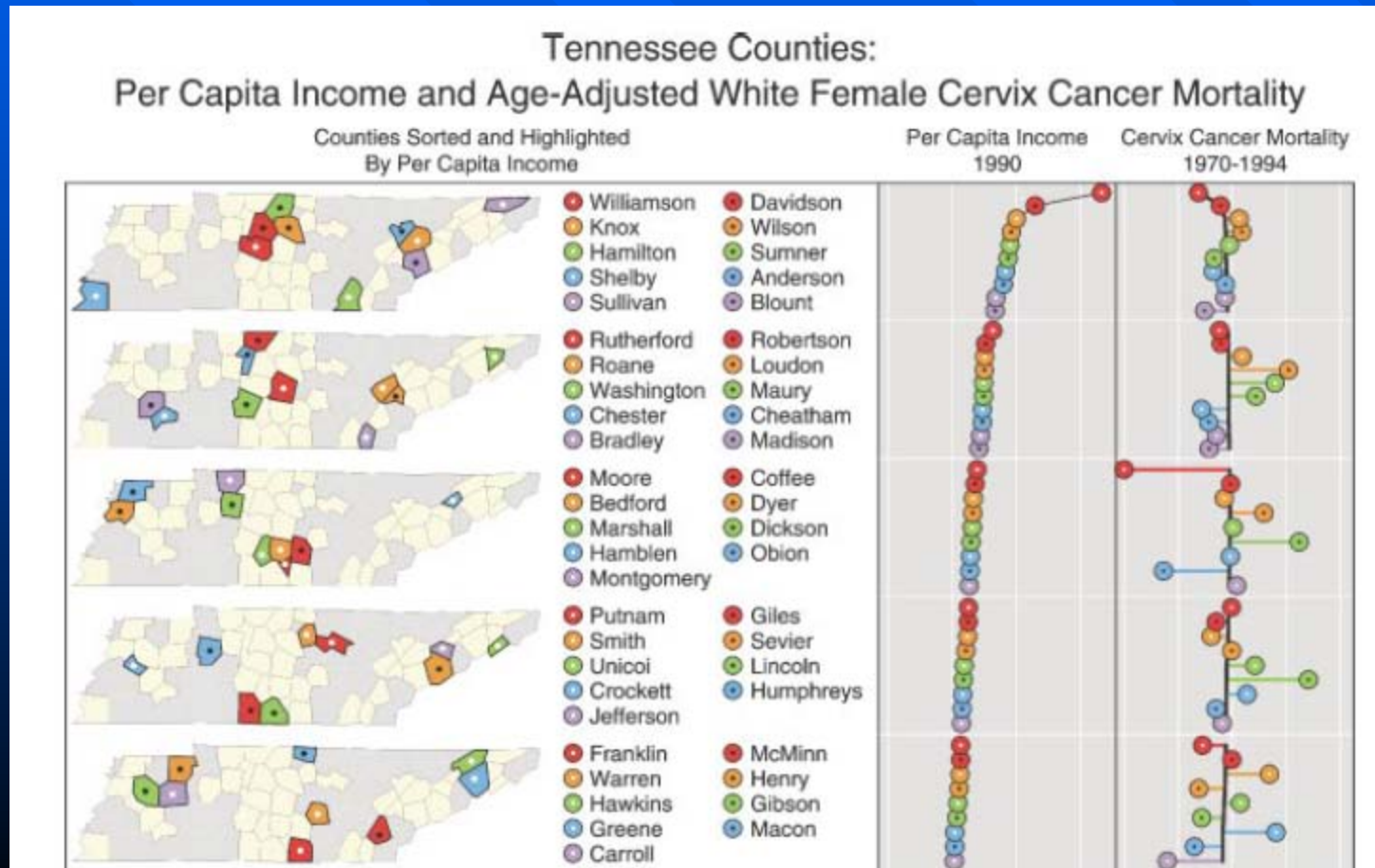
Maps



Median

Future Work (1)

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



Future Work (2)

- Introduce additional data panels to micromap R package, such as line charts (time series plots) and scatterplots
- Further enhance, test, and debug R code for the creation of regional micromaps and add functionality to micromap R package once code is stable

Concluding Remarks

- R packages micromap & micromapST make it easy to create “standard” micromaps
- R package micromap makes it easy to use available or modified shapefiles
- For specific purposes, it may still be necessary to directly adapt the code from Carr & Pickle (2010)

■ References:

Carr, D. B., Pickle, L. W. (2010): Visualizing Data Patterns with Micromaps, Chapman & Hall/CRC, Boca Raton, FL.

Symanzik, J., Carr, D. B. (2008): Interactive Linked Micromap Plots for the Display of Geographically Referenced Statistical Data, In: Chen, C., Härdle, W., Unwin, A. (Eds.), Handbook of Data Visualization, Springer, Berlin/Heidelberg, 267-294.

Symanzik, J., Carr, D. B. (2013): Linked Micromap Plots in R, In: Cho, S.-H. (Ed.), Proceedings of Joint Meeting of the IASC Satellite Conference and the 8th Conference of the Asian Regional Section of the IASC, Asian Regional Section of the IASC, 213-218.

Questions ???