

# Interactive mapping for geographically weighted correlation in big census data

Narumasa Tsutsumida <sup>a, \*</sup>, Joseph Percival <sup>b</sup>, Daisuke Murakami <sup>c</sup>, Takahiro Yoshida <sup>d</sup>, Tomoki Nakaya <sup>e</sup>

<sup>a</sup> Graduate School of Global Environmental Studies, Kyoto University, [naru@kais.kyoto-u.ac.jp](mailto:naru@kais.kyoto-u.ac.jp)

<sup>b</sup> Georepublic, [ipercival@gmail.com](mailto:ipercival@gmail.com)

<sup>c</sup> Department of Data Science, Institute of Statistical Mathematics, [dmuraka@ism.ac.jp](mailto:dmuraka@ism.ac.jp)

<sup>d</sup> Center for Global Environmental Research, National Institute for Environmental Studies, [yoshida.takahiro@nies.go.jp](mailto:yoshida.takahiro@nies.go.jp)

<sup>e</sup> Graduate School of Environmental Studies, Tohoku University, [tomoki.nakaya.c8@tohoku.ac.jp](mailto:tomoki.nakaya.c8@tohoku.ac.jp)

\* Corresponding author

---

**Keywords:** Spatial heterogeneity, Geographically weighted, Shiny, Tokyo

## Abstract:

Census data are widely available in many countries and are useful in describing socio-economic structures in a map at an administrative unit level. Multiple variables in census data can be investigated, however, selecting a large number of variables often leads to confusion and makes it difficult to discern which should be considered. Correlation analysis is often applied to measure the degree of association of a pair of variables, while a correlation matrix is used to summarize relationships amongst multiple variables at the same time. Although both provide important metrics, the spatial configuration of the data is not taken into account. For the purpose of mapping, it is often of interest to highlight the correlative relationship between a pair of variables across space in order to deal with any spatial heterogeneity hidden in the data. In this sense, geographically weighted correlation and partial correlation analyses have been proposed to map spatial variations of correlations in a spatial data set. The geographically weighted approach uses a moving-window kernel running across geographical space and calculates a statistical model or summary statistic with distance-decayed weighted data. The critical issue of mapping such correlation relationships amongst multiple variables is the large number of resultant maps produced. If we are interested in correlations amongst 100 variables, the correlation matrix has the dimension of 100 by 100, while 4,950 correlation maps are produced when we investigate local correlation relationships. Furthermore, the degree of scale (localness) which is an important parameter to understand the local correlative relationship should be explored. To this end, the purpose of this study is to build an interactive mapping system for visualizing spatial variations of correlative relationships amongst multivariate variables in big census data. This system is built on Shiny in R and an R package for calculating the geographically weighted correlation and partial correlation coefficients across multiple variables (<https://github.com/naru-T/GWpcor>) is implemented behind this system for analyses. We use the national census data set for 2005 with 204 variables regarding socio-economic structures in 23 wards in Tokyo with 3,134 administrative units as a case study. The system implements geographically weighted correlation and partial correlation analyses with varying scale parameters and produces an interactive spatial surface of the correlation coefficient. We demonstrate how our interactive mapping system enables users to achieve quick visualization of correlative relationships amongst multivariate census data which can be selected and changed easily from pull-down lists (Figure 1). Such a user-friendly interactive mapping system proposed in this study will help those who need to understand the spatial relationships of the data that is being mapped. This study is supported by ROIS-DS-JOINT (006RP2018).

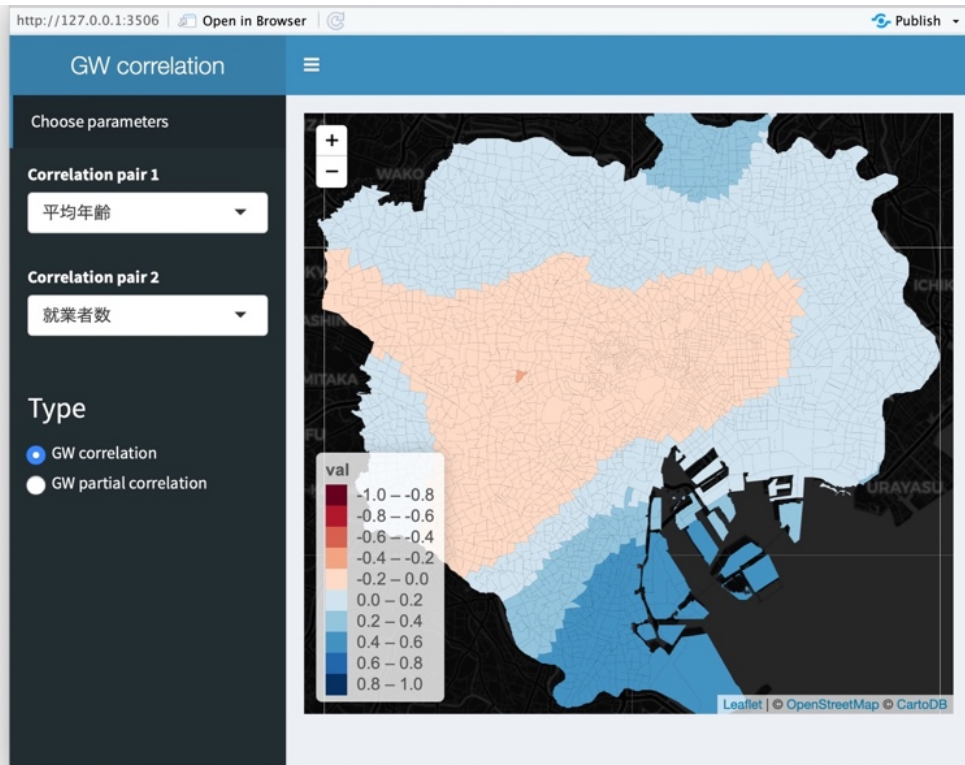


Figure 1. A prototype of interactive map for geographically weighted correlation on a Shiny application.