

## **Exploratory Spatial Data Analysis of COVID-19 infection rates and population vulnerability indicators**

**Author:** Juan Deplet-Barreto, Union of Concerned Scientists

**Objective:** To understand the degree of spatial clustering of COVID-19 and population vulnerability in counties in the contiguous U.S.

**Overview:** In this analysis, we highlight red counties that have combinations of a high percentage of vulnerable populations and high rates of COVID-19, and that are also adjacent to counties with similarly high values. We calculate and map a Local Indicator of Spatial Association (LISA) for pairs of variables in counties in the contiguous United States.

### **Methods:**

#### Percent of population with COVID-19

We first assembled an indicator of COVID-19 infection for counties. We calculated a “Percent of population with confirmed positive COVID-19 cases” by dividing the total number of confirmed COVID-19 cases by the total number of people in each county. COVID-19 cases were obtained from Johns Hopkins University researchers (CSSE 2020) and represent the total of cases as of 14 April 2020. County population data came from the American Community Survey’s county-level population estimates for 2012-2016, as provided by the CDC’s Social Vulnerability Index (CDC n.d.). We note that these numbers do not reflect the true infection rate in a community because as of this writing, large swaths of the population were not being tested.

#### Socio-demographic vulnerability variables

We assembled county-level variable indicators of population vulnerability to environmental stressors to compare against the percent of confirmed COVID-19 cases to evaluate the degree of spatial clustering of extreme values in rates of COVID-19 and population vulnerability. These variables are:

- Percent Age 65 or Older: Percentage of persons aged 65 and older
- Percent Not White: Percentage of the population that are not white, non-Hispanic
- Percent without Health Insurance: Percentage uninsured in the total civilian noninstitutionalized population
- Percent in Poverty: Percentage of persons below poverty

#### Maps of local clustering of COVID-19 vs. population vulnerability

Spatial cluster analysis of public health threats can help inform containment and mitigation strategies that cannot be easily addressed independently by one jurisdiction (Zhang et al. 2019). The United States Centers for Disease Control and Prevention (CDC) suggests consideration of epidemiology in surrounding counties in their guidelines for local implementation of mitigation strategies for community spread of COVID-19 (CDC n.d.). Thus, spatial cluster analysis is a more useful method to identify spatially proximate groups of counties with extreme minima and maxima of COVID-19 and population vulnerability indicators than simply mapping COVID rates together with population indicators.

We used an exploratory spatial data analysis (ESDA) method to identify spatial clusters of high and low values of percent of population with COVID-19 and, individually, the population vulnerability variables listed above. ESDA is an extension of exploratory spatial analysis (EDA) to explore the spatial properties

of data (Haining, Wise, and Ma 2002). For counties in the contiguous US, we calculated the Local Indicator of Spatial Association (LISA, Anselin 1995). LISA works by finding combinations of extreme values in the distribution of both variables and identifying the degree to which those combinations are clustered in space. Clustering in space requires a spatial weights matrix for observations (counties in this analysis) to specify the definition of a neighboring observation. We used a queen first-order contiguity algorithm (illustrated below—as the Queen moves in chess) to determine the spatial weights matrix, indicating that any county that has coincident boundaries with a county of interest is considered a neighbor. We implemented LISA in R (R Core Team 2019) and obtained similar results to those in the GeoDA software package, the reference implementation of LISA (Anselin, Syabri, and Kho 2006).

### QUEEN contiguity

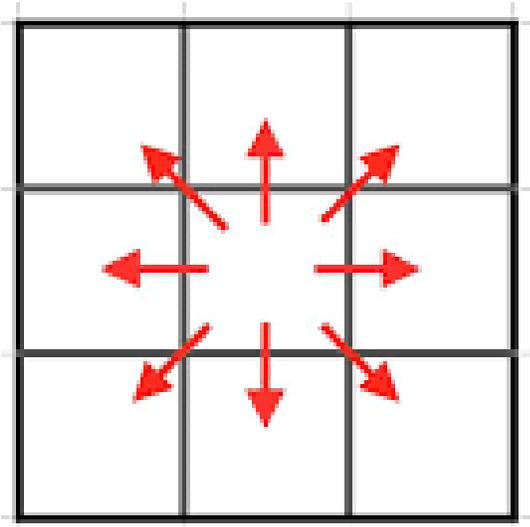


Figure 1. Queen contiguity spatial weights matrix. Adapted from Bellefon, Loonis, and Le Gleut (2018).

The color coding in the illustrative legend below applies to all maps (but ignore the number in parenthesis). We only mapped LISA clusters in the High-High and Low-High categories that also had p-values  $\leq 0.05$ . LISA computes the four possible outcomes of extrema for two variables:

- both variables in counties are **high** and surrounded by similarly **high** values in neighboring counties (red; this combination is particularly what we are interested in);
- both variables in counties are **low** and surrounded by similarly **low** values in neighboring counties;
- one variable in counties is **low** and surrounded by **high** values in neighboring counties (light purple);
- one variable in counties is **high** and surrounded by **low** values in neighboring counties (pink).

BiLISA Cluster Map: con

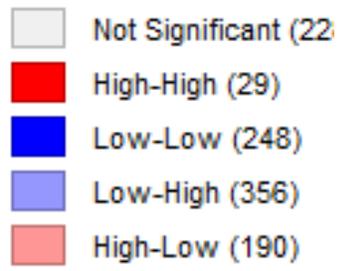


Figure 2. Illustrative map legend of LISA cluster analysis maps.

## References

1. The Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. 2020. 2019 Novel Coronavirus COVID-19 (2019-nCoV) Data Repository by Johns Hopkins CSSE. <https://github.com/CSSEGISandData/COVID-19>.
2. Centers for Disease Control and Prevention (CDC). (n.d.a.). 2016 Social Vulnerability Index. <https://svi.cdc.gov/>. Last Accessed 20 April 2020.
3. Centers for Disease Control and Prevention (CDC). (n.d.b.). Implementation of Mitigation Strategies for Communities with Local COVID-19 Transmission. <https://www.cdc.gov/coronavirus/2019-ncov/downloads/community-mitigation-strategy.pdf>. Last accessed 20 April 2020.
4. Anselin, Luc. "Local indicators of spatial association—LISA." *Geographical analysis* 27.2 (1995): 93-115.
5. Zhang, H., Yang, L., Li, L. et al. The epidemic characteristics and spatial autocorrelation analysis of hand, foot and mouth disease from 2010 to 2015 in Shantou, Guangdong, China. *BMC Public Health* 19, 998 (2019). <https://doi.org/10.1186/s12889-019-7329-5>
6. Haining, Robert, Stephen Wise, and Jingsheng Ma. "Exploratory spatial data analysis." *Journal of the Royal Statistical Society: Series D (The Statistician)* 47.3 (1998): 457-469.
7. R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>. Last Accessed 20 April 2020.
8. Anselin, Luc, Ibnu Syabri and Youngihn Kho (2006). GeoDa: An Introduction to Spatial Data Analysis. *Geographical Analysis* 38 (1), 5-22.
9. De Bellefon, M-P, V. Loonis, and R. Le Gleut. "Codifying the neighbourhood structure". In Loonis and Bellefon (eds.). 2018. *Handbook of Spatial Analysis: Theory and Application with R*. Insee-Eurostat. [https://www.efgs.info/wp-content/uploads/informationbase/introduction/Handbook\\_of\\_Spatial\\_Analysis\\_INSEE\\_EUR\\_OSTAT\\_2018.pdf](https://www.efgs.info/wp-content/uploads/informationbase/introduction/Handbook_of_Spatial_Analysis_INSEE_EUR_OSTAT_2018.pdf). Last Accessed 20 April 2020.