Health & Medicine | Abhishek K. Kala, Sam Atkinson & Chetan Tiwari

Geovisualization for insights into infectious disease risk

Environmental, social, and economic factors influence infectious disease transmission and resulting risk. Using exploratory data analysis techniques can help uncover previously unknown relationships between these factors, contributing to improved public health efforts. Drs Abhishek Kala, Samuel Atkinson, and Chetan Tiwari from the University of North Texas, USA, use geovisualization to identify the spatial context and factors that are strongly associated with West Nile Virus outcomes in California.
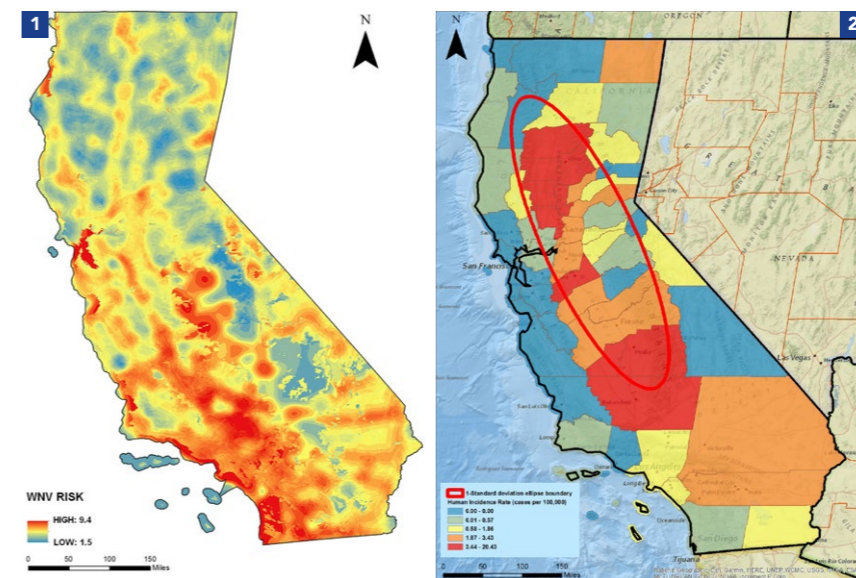
There are many factors that influence the transmission of infectious diseases, which can be broadly split into environmental context and human demographic context. Many studies have attempted to predict infectious disease risk based on underlying environmental conditions of the pathogen's host and/or vectors (such as temperature, humidity, elevation, and vegetation) while others have documented factors related to human vulnerability to the disease (perhaps age, gender, race/ethnicity, and income). However, analytical modelling that combines the two contexts is difficult due to the number of potential explanatory variables, the varying spatio-temporal resolutions of available data, and the differing research objectives that drove the initial data collection. These factors don't exist in isolation and understanding the interactions between them could lead to a more complete understanding of disease transmission and risk. In order to explore questions that require coupling more than one complex process simultaneously, traditional modelling techniques do not appear to be directly applicable and will likely need modifications to be useful. Geovisualization techniques may be used to explore relationships between the outcome and potential explanatory variables thereby improving model development efforts.

Dr Kala, Dr Atkinson and Dr Tiwari have many years of experience in studying how environmental factors influence disease transmission. They are particularly interested in spatial epidemiology – analyzing the geographic distribution of disease risk and corresponding impacts on health outcomes. Recently they have begun to consider demographic characteristics in addition to the environmental context. Considering many factors together poses significant challenges given large volumes of potentially unstructured and unrelated datasets. Additionally, many of these factors will interact together, and the data may have been collected over different scales (for example, by county or by neighbourhood), making them difficult to combine. It is possible to use complex models to analyze these factors and their interactions, but the research team suggests that geovisualization techniques could help identify the most relevant factors as a basis for creating more useful and predictive models.
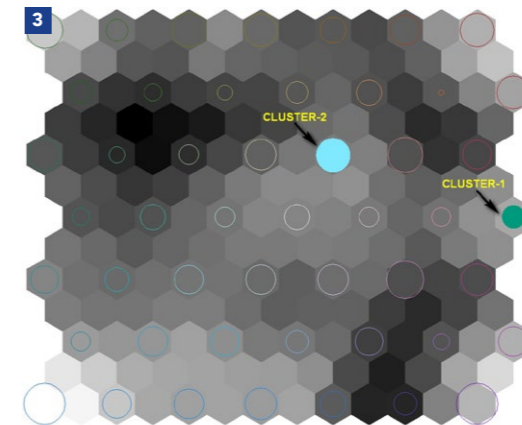
GEOVISUALIZATION
Geovisualization is the use of tools and techniques to support the analysis of large amounts of geospatial data through the use of interactive visualization. It is essentially a data mining process, and is commonly used to identify the spatial context and associated relationships between a pre-defined set of potential explanatory variables. The research team uses three main geovisualization techniques: self-organising maps (SOM), parallel coordinate plots (PCP) and geographic mapping.

A self-organising map (SOM) is a clustering method of data visualization. This uses the pre-specified subsections of the study area (e.g. neighbourhoods, counties or regions) called 'elements', and groups those that are most similar in terms of the factors being investigated into a cluster. A cluster could therefore contain just one element or it could contain many, but those included in a cluster are more similar to each other than to any element in another cluster. Clusters are displayed in a grid of hexagons, which are shaded light to dark to show the level of dissimilarity to neighbouring clusters.

Once a SOM has been created, a parallel coordinate plot is used to consider the data within the clusters. These are represented as a line graph with multiple vertical axes, one for each of the factors included in the analysis. All of the clusters are represented by a line on the graph, so if there are 30 clusters in the SOM then there are 30 lines on the PCP, and the thickness of the line indicates the number elements within the cluster. The point at which a line intersects an axis indicates the relative value of that factor for that specific cluster. Relationships between factors can be determined by examining the points of intersection between each line (a cluster) and each vertical axis in the graph.

Geographic mapping then displays where the elements of these clusters are on a map of the study area. When a cluster is selected all the elements (counties, neighbourhoods, districts) of that cluster are highlighted on the map. This can produce interesting results, as the elements of a cluster might be very close together within the study area (indicating that there could be a geographic reason for the similarities between these elements) or they could be spread out over the study area (indicating that these elements are similar in ways that don't include a geographic factor). These results could suggest approaches for further analyses (for example if they suggest that geography is not an important factor in risk) or suggest areas of focus for public health interventions.

WEST NILE VIRUS
Drs Kala, Atkinson and Tiwari recently used West Nile Virus (WNV) as an example of how geovisualization can be employed to simultaneously consider large amounts of data related to infectious disease risk. WNV is a vector-borne disease, which means it is transmitted to humans and other animals through a blood-feeding insect – in this case mosquitoes. In humans it is often asymptomatic, though WNV can cause flu-like symptoms and, in rare cases, cause severe illnesses related to the nervous system. It was first identified in the US in 1999, and quickly spread across North America. Because of how difficult it is to predict outbreaks of WNV, it is an ideal case study to determine if the use of geovisualization could provide valuable insights.

California is the most populous state in the US with historically high rates of WNV incidence. The study area was defined by analysis of previously reported cases of WNV in humans and was narrowed down to those areas which contained 67% of all reported cases. This is a data reduction strategy to focus on the area where it is most likely that new patterns of factors could be found. Demographic data from the United States Census Bureau was then obtained for all census tracts within the study area. Census tracts represent a commonly used geographic unit of analysis that represent contiguous regions containing between 1,500 and 8,000 individuals.

The demographic factors considered in the analysis were median age,

1 – West Nile Virus (WNV) risk based on environmental context modelling (i.e., mosquito habitat risk) in California.

2 – West Nile Virus human incidence rate by county in California. Colours represent quintiles of reported human incidence of WNV.

3 – Self organising map representing 49 nodes (circles) with a combination of contextual and compositional parameters from 1,133 census tracts.

Geovisualization is the use of tools and techniques to display large amounts of data relative to geographic space.
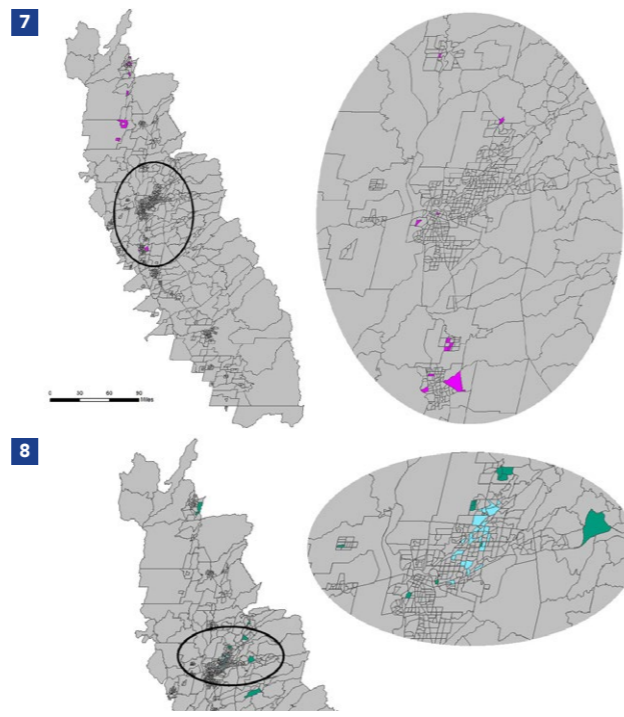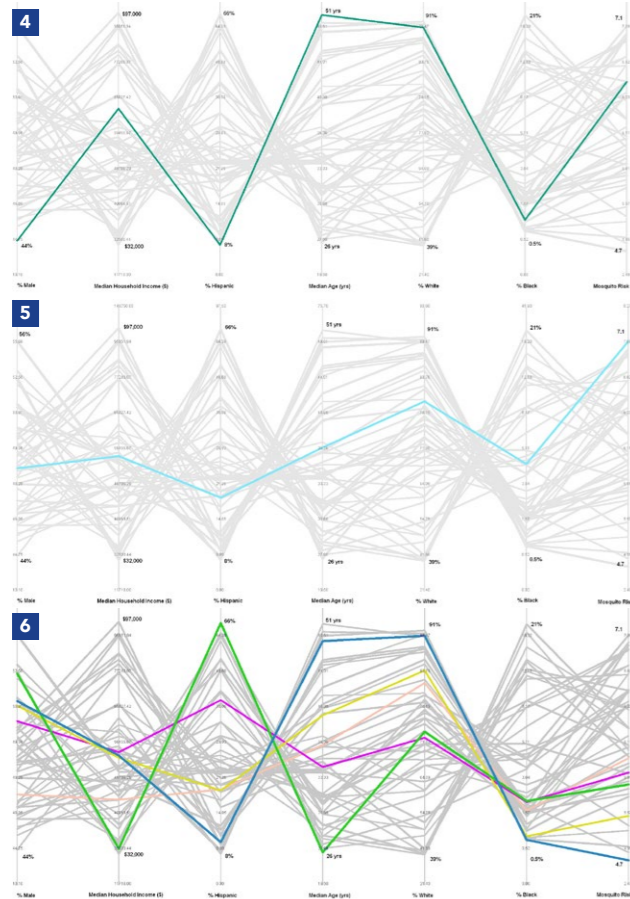
Figures (4)–(6) Examples of parallel coordinate plots.
(7) Census tracts, highlighted in pink in Glenn County, CA.
(8) Geomap showing spatial context of census tracts, with the highest median age (green) and the highest mosquito habitat risk (blue).

percentage of the population that is male, percentage of the population that is white, black or Hispanic, and median household income. These factors were chosen based on previous literature that suggests that older people are more vulnerable to infectious disease (possibly due to a weakened immune system), and that gender, race/ethnicity and income influences vulnerability due to differences in social and lifestyle aspects. Only one environmental factor was included in this analysis, known as 'mosquito risk'. This factor was derived from an earlier study which combined eight parameters (stream density, surface temperature, surface slope, cultivated land, developed land, road density, vegetation type, evapotranspiration rate). This general environmental variable was previously shown to be statistically significant in relationship to the number of WNV-infected dead birds in an area (reliably used to estimate human infections).

The SOM produced from this data had 49 clusters, containing the 1,133 census tracts in the study area. Translation of results into practice was demonstrated using two clusters; one that contains census tracts with the highest median

age, and the other containing the census tracts with the highest mosquito risk. For the cluster with the highest median age, the PCP showed that these census tracts also had low or moderate levels of other risk factors such as percentage of male population or mosquito risk. This could suggest that these areas may not be as high risk for WNV as might be expected from tracts with the highest median age. For the cluster with the highest mosquito risk the PCP showed that these census tracts also had moderate levels of the other factors considered. However, when this cluster was shown on the map all the tracts were in close proximity (as opposed to the highest age tracts, which were dispersed across the study area), which might be an important consideration for intervention planning.

Viewing data in this way could help to identify important factors for disease transmission, suggest new hypotheses for investigation, or inform decisions for public health interventions.

FUTURE APPLICATIONS
Drs Kala, Atkinson and Tiwari believe that geovisualization techniques provide an exploratory spatial data analysis framework that can be used to improve model development for exploring relationships between adverse disease outcomes, underlying environmental risk factors, and their impact on populations. Geovisualization tools are increasingly available and easier to use, and could even be applied to track the spread of outbreaks in near-real time. Geovisualization considers interactions between a disease, the environment, and the population, which can then be used to suggest new hypotheses for investigation, or inform decisions for public health interventions.

Behind the Research

Dr Abhishek K. Kala
E: searchkala@gmail.com
T: +1 214-463-4808

Dr Sam Atkinson
E: atkinson@unt.edu
T: +1 940-565-2694

Dr Chetan Tiwari
E: chetan.tiwari@unt.edu
T: +1 940-369-8103 **W:** https://aeri.unt.edu

Research Objectives

Kala, Atkinson, and Tiwari use geovisualization to guide the development of analytical models to better understand, prepare for, and react to WNV and other infectious diseases.

Detail

Bio

Dr Abhishek K Kala holds a Ph.D. degree in Environmental Science from the University of North Texas and is currently working as a GIS/Remote Sensing analyst at Tessellations Inc. His research focuses on developing a spatially explicit epidemiological model, geovisualization techniques for exploring complex hidden relationships, and utilising open source geospatial solutions to monitor environmental quality. He loves to work on his organic vegetable garden in his spare time and spend quality time with his family.

Dr Sam Atkinson is Regents' Professor of Environmental Science and Scholar in the Advanced Environmental Research Institute at the University of North Texas. He has recently served on the U.S. Army Science Board and is former Vice Chair of the U.S. Army Corps of Engineers' Environmental Advisory Board.
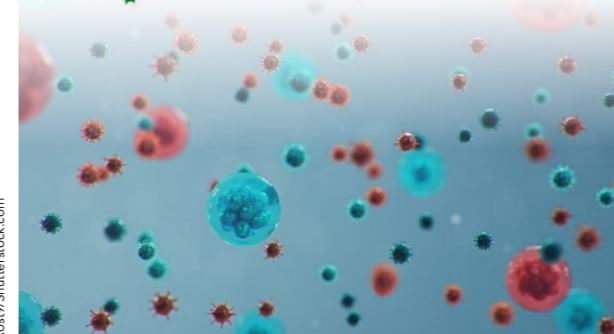
Dr Chetan Tiwari is an Associate Professor of Geography and the Environment and is affiliated with the Advanced Environmental Research Institute at the University of North Texas. His research focus is on geographic information science, response analysis, location modelling, and spatial epidemiology.

Funding

References

Kala, A.K., Atkinson, S.F., & Tiwari, C. (2020). Exploring the socio-economic and environmental components of infectious diseases using multivariate geovisualization: West Nile Virus. *PeerJ* 8:e9577. https://doi.org/10.7717/peerj.9577

Kala, A.K., Tiwari, C., Mikler, A.R., & Atkinson, S.F. (2017). A comparison of least squares regression and geographically weighted regression modeling of West Nile virus risk based on environmental parameters. *PeerJ* 5:e3070. https://doi.org/10.7717/peerj.3070

Atkinson, S.F., Sarkar, S., Aviña, A., Schuermann, J.A., & Williamson, P. (2014). A determination of the spatial concordance between Lyme disease incidence and habitat probability of its primary vector Ixodes scapularis (black-legged tick). *Geospatial Health* 9(1):203-212. https://doi.org/10.4081/gh.2014.17

Atkinson, S.F., Sarkar, S., Aviña, A., Schuermann, J.A., & Williamson, P. (2012). Modelling spatial concordance between Rocky Mountain spotted fever disease incidence and habitat probability of its vector Dermacentor variabilis (American dog tick). *Geospatial Health* 7(1):91-100. https://doi.org/10.4081/gh.2012.108

Tiwari, C., & Rushton, G. (2010). A Spatial Analysis System for Integrating Data, Methods and Models on Environmental Risks and Health Outcomes. *Transactions in GIS* 14(s1):177–195. https://doi.org/10.1111/j.1467-9671.2010.01220.x

Kala, A., Atkinson, S.F., & Tiwari, C. [Poster]. Uncovering the hidden patterns in complex epidemiological datasets. University of North Texas.

Personal Response

Could geovisualization have useful applications for the current COVID-19 pandemic?

The COVID-19 crisis is escalating, and datasets reported daily at different geographic scales include the confirmed cases, total deaths, total recovered, and daily count of new cases. Several studies have also shown evidence that pandemic complications are more prominent in certain ethnic groups or are based on neighbourhood and demographic characteristics. There is a volume of datasets building up with several risk factors which makes it complicated to understand the nature of disease. Multivariate geovisualization can be of great value to help to simplify this complex data, explain interactions among variables, capture the patterns of interest, and communicate findings for public health decision making processes. This can help to guide authorities for optimum planning and help minimise the disease burden. These tools can help researchers and decision makers to act in a more effective manner to decide where surveillance should be prioritised.

UNT | ADVANCED ENVIRONMENTAL RESEARCH INSTITUTE

Rost9/Shutterstock.com