International Journal of Geographic Information System
Vol. 4, No. 2, September 2017, pp. 1-23
Available Online at http://acascipub.com/Journals.php

Research article

Predicting potential human trafficking recruiting geolocations using Eigendecomposed non-zero spatial autocorrelation coefficients and Poissonian heterogeneous gamma distributed variables

William L. Gardner[a], Toni Panaou[b], Kyle Watterson [c], Sam Alao[a], Benjamin G. Jacob[a]

[a] Global Health, University of South Florida, Tampa, USA, [b] Civil and Environmental Engineering, University of South Florida, Tampa, USA, [c] Physical Therapy & Rehabilitation Services, University of South Florida, Tampa, USA

E-mail: wgardner@health.usf.edu, bjacob1@health.usf.edu, apanaou@mail.usf.edu

This work is licensed under a Creative Commons Attribution 4.0 International License.

Abstract

Can recruitment into human trafficking be predicted and prevented? Recent literature demonstrates the potential to create predictive models determining where various crimes are most likely to occur. Additional literature emphasizes the potential for emergency room personnel to recognize and intervene in human trafficking. Unfortunately, these approaches focus on responding to crimes after they have already happened. Once a person has been exploited as slaves or forced prostitutes, the damages to their physical/ mental health have already been established. Utilizing a transferable spatial model and socioeconomics, this study determined where recruitment into human trafficking would likely happen. Frequentist and non-linear autoregressive models may allow determining mapping covariates of statistical significance associated with factors which indicate high-risk areas. Here, Moran's I, a product moments coefficient,revealclustering propensities associated to non-zero autocorrelation of georeferenceable, geolocations of impoverished people.

Copyright © acascipub.com, all rights reserved.

Keywords: Human trafficking, spatial autocorrelation, poverty, negative-binomial regression, geographic information systems, Moran's I, ARCMap, crime prevention, socioeconomics, Seattle

1. Introduction

Is human trafficking preventable? In 2016 the National Human Trafficking Hotline in the United States (U.S.) experienced an increase in the number of human trafficking cases reported to 7,572, reflecting an improvement in identifying and reporting cases (Polaris 2017). Unfortunately, inaccessibility of resources or opportunities which may lead to escaping trafficking situationsresists these improvements, keeping many victims hidden. Recruitment, transportation, and exploitation stages of trafficking each provide unique opportunities to interrupt the flow of this heinous crime (Albanese 2007). Researchers interested in this movement of persons have emphasized differences between countries to determine which policies and social factors facilitate human trafficking (Akee *et al.* 2007, Ofuoku and Uzokwe 2012, Cho 2015, and Hernandez and Rudolph 2015). By comparing countries which act as source-destination pairs, push and pull factors have been identified for international human trafficking, such as wage disparity and the believed potential to send remittances home to one's family (Akee *et al.* 2007, Hernandez and Rudolph 2015). Working with existing literature and public resources, Cho (2015) has identified crime, employment, income difference, and labor demand all as very valuable indices for predicting human trafficking patterns, whereas institutional quality only matters in source countries. While informative, there remains a gap between identifying risk factors and acting with them. Additionally, emphasizing differences between country pairs obscures the existence of domestic trafficking.

To an extent, researchers have identified potential intercessory agents. Greenbaum (2016) stresses the valuable opportunity for emergency departments to become points where a trafficked person may be identified. Relying on hospital staff as identifiers for human trafficking victims additionally assumes people are truthful in their responses to their care team. According to Leun and Schijndel (2016), coercion and fear of being identified as a criminal for illegally being in a host country cause many trafficking victims to obscure information about themselves or their situation. Furthermore, the cost of health care may act as a barrier against trafficked people from being seen by a doctor in the first place. As reported by the U.S. Department of State (2014),hospital bills increase the forced debt trapping victims into their servitude; victims avoid hospital visits and consequentially their chances

Copyright © acascipub.com, all rights reserved.

for escape.While opportunities to aid trafficked persons are continually valuable, more effort needs to be made in preventing people from being recruited as victims.

The deceptive nature of recruiters prevents us from locatingpeople being actively recruited. At the time of writing this publication we are not aware of any database providing information towards specifically where a trafficking victim was when they were first contacted for recruitment. Based on narratives published in recent editions of the annual *Trafficking in Persons Report* by the U.S. Department of State, we believe that poverty may serve as one of the most valuable indicators for where to begin: victims of trafficking tend to be desperate for opportunities to escape poverty's clutches, making them easy prey for insidious promises made by recruiters (U.S. Department of State 2014, U.S. Department of State 2016).

Identifying poverty within a community provides additional opportunities. Health risks associated with poverty include a higher incidence of obesity, diabetes, depression, anxiety, and suicidal ideation (James *et al.* 2001, James 2004, Bratanova *et al*. 2016). These health outcomes further demonstrate themselves in situations of relative poverty, a measure which describes large socioeconomic inequality between one neighbors (Pickett *et al.,* 2005, Banks *et al.* 2006, Marmot 2006, Wilkinson and Pickett 2009). Moreover, human behaviors resulting from food insecurity and stress induced from inequality have been observed experimentally in lab animals (Shively and Clarkson 1994, Sapolsky 2004, Jarvandi, *et al.* 2009).

Among youth, school income provides a strong indicator for adolescent mental health development and deviance or risk-taking behaviors (Coley *et al.* 2017). On a spatial level, Coley and associates demonstrate the significance of varying methods of aggregating data by family, neighborhood, and school- and the spatial relationships of behavioral covariates among youth,however, the ability to locate clusters within the data are not demonstrated. Alternatively, other researchers have described the roles of family poverty or affluence and their similar impacts on the mental health and behavioral decisions among youth (Luthar and Sexton 2004, Dearing 2008, Luthar, Barkin, andCrossman 2013). While valuable for informing clinicians treating at-risk youth, there is a lack of spatial logistics which may be employable by those interested in improving the conditions aggravating the adolescent's behavior.

This study engages in anti-trafficking efforts at a community level by implementing geographic information system (GIS) techniques to remotely locate hot spots where poverty occurs within a city (this study

Copyright © acascipub.com, all rights reserved.

utilizes data from Seattle, Washington (WA)) and identifies covariates which can be used in future research for locating impoverished people. Additionally, cross country paring is useless when attempting to deal with domestic trafficking patterns. To overcome this, a smaller spatial extent will be employed, which is more practical to organizations and law enforcement to utilize in interrupting trafficking flows. Researchers already recognize the potential for mapping for crime, and applications for forecasting new geolocations of exploitation within cities with advertised escort services (Caplan *et al*. 2010, Voloshin *et al.* 2016). Intervening in exploitation remains important, however by the time a person is being abused multiple opportunities for intervention have already expired and physical or psychological damage to the person has already had an opportunity to begin. This research focuses on intervening at the site of recruitment, and therefore presents the earliest opportunity to disrupt human trafficking published to date.

This research aims to create the model based on information that is as accessible as possible. This allows organizations and law enforcement to minimize their cost while attempting to interrupt trafficking in persons. To facilitate this goal we utilized open source datasets from Seattle, WA to build the model sourced either through the U.S. Census Bureau (https://factfinder.census.gov/faces/nav/jsf/pages/guided_search.xhtml) or the city government webpage (https://data.seattle.gov/). The U.S. is not considered a high risk for being a source country (U.S. Department of State 2016), however the availability of census data which may be used in predictive spatial autocorrelation mapping in source countries is more restricted to specific organizations or agencies hosted by those countries.

Spatial autocorrelation is the correlation among values of a single variable strictly attributable to their relatively close locational positions on a two-dimensional surface, introducing a deviation from the independent observations assumption of classical statistics. Spatial autocorrelation exists because real world phenomena are typified by orderliness, pattern, and systematic concentration, rather than randomness. Tobler's First Law of geography encapsulates this situation: "everything is related to everything else, but near things are more related than distant things." To this maximum should be added the qualifier: "but not necessarily through the same mechanisms." In other words, spatial autocorrelation means a dependency exists between values of a variable in neighboring or proximal locations, or a systematic pattern in values of a variable across the locations on a map due to underlying

Copyright © acascipub.com, all rights reserved.

common factors. Zoning ordinances in Seattle Washington may force similar land use types to group together in coterminous locations where human trafficking occurs.

The hypothesis we test in this study states that (1) income levels are more non-zero spatially autocorrelated compared georeferenceable populations and (2) geospatial clusters stratified byincome may reveal similar land use types which may be grouped together in coterminous geolocations where human trafficking occurs. in Seattle Washington. Hernandez and Rudolph (2015) demonstrate that geographical modeling can be employed effectively to predict where the flow of human trafficking flows between country pairs. While this highlights opportunities for geospatial analysis, most countries are too large and too populated to justify approaching the problem on such a large scale.

2.Materials and Methods

2.1 Study site

Seattle is a city located in Washington State in the Northwest area of the lower 48 contiguous United States. At 83.94 square-miles with a population density of 7,251 people per square-mile (U.S. Census Bureau, NDGa) it has previously been recognized among the fastest-growing big cities in the U.S. (Balk 2014, Balk 2015, Balk 2017). In 2015, the city welcomed 19.7 million overnight and 18.3 million day visitors (Blandford 2016). In spite of all of this growth, it also features one of the highest homeless populations in the U.S. (Ryan 2015).

The vector shapefiles in this research were obtained from https://data.seattle.gov/ (City of Seattle 2017) and the socioeconomic data in this research was collected from the 2015 American Community Survey (ACS) estimates provided on https://factfinder.census.gov/faces/nav/jsf/pages/guided_search.xhtml (U.S. Census Bureau, NDGb). These included selected variables by block groups in King County, WA. We used ArcMap v10.5® and Microsoft Excel to process and join the data and Statistical Analysis Software (SAS) to determine regression models.

Employing ArcMap, we used a Seattle City Neighborhood shapefile(City of Seattle 2017) to clip a King County Census Block Groups shapefile (City of Seattle 2017)to reduce the spatial extent to Seattle only. We then executed an attribute join between the clipped census block group shapefile and our socioeconomic variable tables (U.S. Census Bureau, NDGb) keeping only matching records, resulting in a layer with all census block groups

Copyright © acascipub.com, all rights reserved.

inSeattle with their variables. To remove influence uninhabited spaces, such as parks and roads, we then clipped the census block group shapefile using an additional 2012 Built Environment shapefile (City of Seattle 2017), resulting in all data being separated into the outlines of all buildings within Seattle in 2012. X and Y coordinates were provided by adding fields to the attribute table of the resulting shapefile and calculating the X or Y centroid for each feature. The resulting table was then utilized for analysis in SAS for regression modelling.

2.2 Regression Analysis

Prior to statistical analysis, certain data modifications were needed. To treat income as a threshold, we created columns to calculate the total number of people earning below a certain income level instead of those earning within a specific range. Additionally, ACS data presented people without health insurance in separate columns by age. We treated absence of health insurance as a potential indicator for vulnerability in this research, and thus created a new column totalling the number of people without coverage regardless of age. Table 2.1 describes the variables featured in the study.

The first stage of this analysis utilized Poisson regression to determine the relationship between income levels and featured covariates. Poisson regression is one special case of the Generalized Linear Model (GLM) which allows one to fit models to a dependent variable that is a member of the exponential distribution family for linear quantitation of covariate variabilities (Pielou 1969). Our vulnerable population GLM came down to three components: (1) the distribution of the dependent variable, (2) a linear function of a set of independent variables, and (3) a link function between the dependent variable and its expectation as expressed by the linear function of independent variables. The Poisson regression assumed that each independent count estimate (i.e. n_i), recorded at a feature "i"= 1, 2, … n, was from a Poisson distribution. These data were described by a set of predictor variables denoted by matrix X_i, a 1 x p vector of covariate indicator values for a building i in Seattle. The expected value of these data was given by $\mu_i(X_i) = n_i(X_i) \exp(X_i\beta)$, where β was the vector of non-redundant parameters, and the Poisson rates were parameter given by $\lambda_i(X_i) = \mu_i(X_i)/n_i(X_i)$; the rates parameter $\lambda_i(X_i)$ was both the mean

and the variance of the Poisson distribution for a building(McCullagh and Nelder 1989). Abackwards-regression analyses were performed in SAS® using a p>0.05 significance level. The data was log-transformed before analysis to normalize the distribution and minimize the standard error.

Copyright © acascipub.com, all rights reserved.

International Journal of Geographic Information System
Vol. 4, No. 2, September 2017, pp. 1-23
Available Online at http://acascipub.com/Journals.php

When the logarithm was applied as a link function, the Poisson regression demonstrated over-dispersion. Poisson regression is estimated based on the likelihood function that is constructed under the independence assumption (Jacob *et al.*, 2013). Poisson distribution predicts non-negative integers in data analyses, where the mean and variance are equal (Kaiser and Cressie 1997). Due to the over-dispersion we used a negative binomial model to evaluate the variables, as these models are considered to be convenient and practical for handling over dispersion in remote-sampled covariates (Haight 1967). This approach allowed the likelihood ratio and other standardtests to be implemented, permitting the fitting procedure to be carried out by using an iterative weighted least squares regression similar to those of the Poisson (Jacob *et al*., 2009). SAS PROC GENMOD was employed to build the Poisson model with a gamma-distributed mean. Although we observed robustness of variables using a non-homogeneous gamma distributed mean form a negative-binomial regression, these variables had virtually no cartographic predictive power.

To compensate for this, we then applied a Moran's *I*/ Geary's C test to the variables to detect autocorrelation among variables. Moran Coefficient is n index of spatial autocorrelation, involving the computation of crossproducts of mean-adjusted values that are geographic neighbors (i.e., covariations), that ranges from roughly (-1, -0.5)0 to nearly 0 for negative, and nearly 0 to approximately 1 for positive, spatial autocorrelation, with an expected value of -1/(n-1) for zero spatial autocorrelation, where n denotes the number of areal units. Geary Ratio is index of spatial autocorrelation, involving the computation of squared differences of values that are geographic neighbors (i.e., paired comparisons), that ranges from 0 to 1 for negative, and 1 to approximately 2 for positive, spatial autocorrelation, with an expected value of 1 for zero spatial autocorrelation (see Griffth 2003).

Initially, spatial autocorrelation was evaluated among the sampled clinical and environmental covariates at the study site using Moran's *I*. In this research Moran's I was defined as $I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(X_i - \bar{X})(X_j - \bar{X})}{\sum_i (X_i - \bar{X})^2}$where N was the number of georeferenced buildings indexed by i and j;$X$was the vulnerable population incidence rates; $\bar{X}$was the mean of X; and w_{ij} was an element of a matrix of spatial weights. The expected value of Moran's I under the null hypothesis of no spatial autocorrelation was then $E(I) = \frac{-1}{N-1}$. Its variance thereafter was equal to $Var(I) =$

Copyright © acascipub.com, all rights reserved.

International Journal of Geographic Information System
Vol. 4, No. 2, September 2017, pp. 1-23
Available Online at http://acascipub.com/Journals.php

$\frac{(NS_4-S_3S_5)}{(N-1)(N-2)(N-3)(\sum_i\sum_j w_{ij})^2}$ where$S_1=\frac{1}{2}\sum_i\sum_j(w_{ij}+w_{ji})^2$, $S_2=\frac{\sum_i(\sum_j w_{ij}+\sum_j w_{ji})^2}{1}$, $S_3=\frac{N^{-1}\sum_i(x_i-\bar{x})^4}{(N^{-1}\sum_i(x_i-\bar{x}^2)^2}$, ,$S_4=\frac{(N^2-3N+3)S_1-NS_2+3(\sum_i\sum_j w_{ij})^2}{1}$, $S_5=S_1-2NS_1+\frac{6(\sum_i\sum_j w_{ij})^2}{1}$.

For statistical hypothesis testing, the Moran's I values were then transformed to Z-scores where values greater than 1.96 or smaller than −1.96 indicated spatial autocorrelation that was significant at the 5% level.

We also used the Geary's coefficient (that is, Geary's C) which is inversely related to Moran's I,. Moran's I is a measure of global spatial autocorrelation, while Geary's C is more sensitive to local spatial autocorrelation (Griffith 2003). In this research Geary's C was defined as $C=\frac{(N-1)\sum_i\sum_j w_{ij}(X_i-X_j)^2}{2W\sum_i(X_i-\bar{X})^2}$ where $N$was the number of buildings indexed by i and j; X was the population covariate incidence; $\bar{X}$ was the mean of X; w_{ij} was a matrix of spatial weights; and W was the sum of all w_{ij}. The value of Geary's C lies between 0 and 2.

We analyzed the n-by-1 vector $x=[x_1 \ldots x_n]^T$containing the georerenced population covariates for n spatial units and n-by-n symmetric spatial weighting matrix **W** using the spatial Indices. The values w_{ij} were the spatial weights based on the tested variables stored in the matrix **W** where $\sum_{(2)}=\sum_{i=1}^{n}\sum_{j=1}^{n}$ with $i\neq j$ which had a null diagonal($w_{ii}=0$). This symmetric matrix revealed $W_{\,ij}=W_{ji}$ was then generalized to a non-symmetric matrix W by using W=(W*+W*T/2) . The autocorrelation statistics were s then rewritten using matrix notation as: $I(x)=\frac{n}{1^tW1}\frac{x^THHWHHx}{x^THHx}=\frac{n}{1^TW1}\frac{x^THXHx}{x^THx}$.SAS/GIS® (http://www.sas.com/products/gis/) was used to perform the eigenvector decomposition spatial filter analysis on the population data.

3. Results and Discussion

3.1 Results

Using a negative binomial non with a non-homogeneous gamma distributed mean deflated the overdispersion was in the Poissonian model (Table 3.1). Results are reported with a Pearson goodness-of-fit score; the closer this value is to zero, the more reliable the model (Freedman 2005). At all maximum income thresholds tested in this study, the negative binomial model resisted influence from overdispersion more than the data than the Poisson model.The best

Copyright © acascipub.com, all rights reserved.

International Journal of Geographic Information System
Vol. 4, No. 2, September 2017, pp. 1-23
Available Online at http://acascipub.com/Journals.php

fit model based on this statistic was the covariate with a maximum income of $10,000.The model progressively became less statistically significant as the threshold is raised.

Table 3.2 provides the coefficient values for each independent variable sampled per model employing the negative binomial regression after removing any significance values greater than 0.05. Variables *fs, inca,* and *nhi* contain data related to eligibility for federal subsistence for food security, income assistance, and health insurance, respectively. Receiving food assistance positively correlated with increases in maximum income thresholds in the experimental models, possibly because eligibility for this benefit does not require recipients to be below the Federal Poverty Level and is based on the size of the family receiving assistance. While the control model utilizes *fs*, the inclusion of more households earning more than the eligibility level for this benefit appears to dilute the value of the variable. Unsurprisingly, absence of health insurance is negatively correlated for all income thresholds eligible for federally assisted health coverage, and positively correlated in other models. This demonstrates that federal initiatives, such as Medicaid and the Affordable Care Act, have been capably equipping the most impoverished members of the sample population with access to health care. On the other hand, this trend demonstrates a greater absence of health insurance among households earning higher incomes. Within the experimental models, detection of households receiving income assistance become most notable when excluding households earning more than $40,000 annually. This value is nevertheless lower than the coefficient found in the control.

The number of unemployed persons, demonstrated by the independent variable *unemp*, demonstrated an inverse relationship to the maximum income threshold. This variable fails to be a valuable coefficient in models including households earning greater than $25,000 annually. Since unemployed persons have extremely low incomes, this decliningpattern suggests that after including households earning more than $25,000 annually the dilution of unemployed persons becomes too great to detect.

Table 3.3 features the clustering analysis with corresponding incremental distances, as well as the maximum, minimum, and mean values for each of these outputs. These statistics tell two different stories for determining a linear spatial model depending on whether the resulting z-statistic or distance band is emphasized more than the other. The *Clustering Critical Value* column is the z-statistic from the *High/ Low Clustering* tool in ArcGIS(www.esri.com) and a higher score is better, representing a lower likelihood that clustering is due to random chance. All nine featured dependent variables show significant clustering with less than a 1% likelihood that

Copyright © acascipub.com, all rights reserved.

International Journal of Geographic Information System
Vol. 4, No. 2, September 2017, pp. 1-23
Available Online at http://acascipub.com/Journals.php

clustering is due to random chance. The *Distance* column is the distance threshold used for calculating the *Clustering Critical Value*, taken from the peak value determined by the *Incremental Spatial Autocorrelation* tool in ArcGIS and lower values are better, indicating more tightly bound clusters in geographical space; all variables featured only one peak using this tool.

Table 3.4 describes the Moran's I and Geary's C coefficients for each threshold value. Moran's I coefficients scale between -1 and 1, with a positive number representing spatial autocorrelation, 0 representing evenly distributed data, and -1 representing randomly distributed data. Geary's C coefficients scale between -2 and 2 in a similar fashion. All the income thresholds demonstrated positive autocorrelation (aggregation of similar attributes in geographic space)however lower income thresholds demonstrated stronger autocorrelation through the Geary's C, while the Moran's I favored higher income thresholds.

3.2 Discussion

Tables 3.1 and 3.3 provide the most valuable statistics for determining a potential linear model of best fit, while 3.2 provides information about what is occurring within that model. Alongside Table 3.1, Table 3.4 provides valuable information about the clustering trends within the data. Reviewing these tables, the most valuable potential models belong to the dependent variables *lowinc1, lowinc2,* and *lowinc9.*

Figures 3.1, 3.2, 3.3, and 3.4 present Hot-Cold Spot representations of the data points for variable *lowinc1, lowinc2, lowinc9,*and *lowinc16*, respectively. Clustering occurs in a North-South orientation, following the geographic layout of the city. The largest hotspot in all three maps covers Downtown Seattle with slight local migration of the cluster.

Figure 3.1 presents a Hot-Cold Spot analysis of the data points for variable *lowinc1*, households earning less than $10,000 annually. Table 3.1 demonstrated that the lower the income threshold used, the more reliable the model. The *Clustering Critical Value* is the second highest and the modal *Distance* value is below the mean. Table 3.2 shows that this model rejects people whose primary language spoken at home is not English but speak English "not well" and 4-and 5-person households. All households featured in this model cover people who are eligible for receiving benefits from food assistance programs, income assistance, and federally subsidized health coverage.

Copyright © acascipub.com, all rights reserved.

International Journal of Geographic Information System
Vol. 4, No. 2, September 2017, pp. 1-23
Available Online at http://acascipub.com/Journals.php

The experimental model associated with the highest *Clustering Critical Value*, in Table 3.3 belongs to *lowinc2* (Figure 3.2). This variable represents all households earning less than $15,000. In Table 3.2 we see that this model includes people receiving benefits from food assistance programs, income assistance, and federally subsidized health coverage with only some of the single person households not being fully eligible for these benefits. The model features eleven independent variables and excludes people whose primary language spoken at home is not English but speak English "very well", and 4-, 5-, and 6-person households. This approach accepts a distance threshold below the mean, which is more favorable, but with a low z-score for that distance threshold.

Table 3.3 describes *lowinc9* as the model with the lowest distance threshold, *lowinc9* (Figure 3.3). This variable represents all households earning less than $50,000. This model rejects the use of people whose primary language spoken at home is not English but speak English either "very well" or "not at all", unemployed persons, and 5- and 6-person households. At this level, the single-, 2-, 3-, and 4-person households, and some of the 7-or-more-person households include larger amounts of people not eligible for receiving benefits from food assistance programs, income assistance, or federally subsidized health coverage. In the Northwest area of the data stands a grouping of data points which appear as insignificant for *lowinc1* and *lowinc2,* but appears as a significant cluster in *lowinc9*. This could indicate contamination of data caused by the inclusion of households with larger income, or it could indicate that using too low of an income threshold causes exclusion of hotspots in the data.

3.3 Conclusion

A large factor in this research is a scaling level of who is being included in the model, and thus who should be included in future models. We can easily say that a person earning greater than $200,000 per year is not as vulnerable to problems specific to the poor as a person earning less than $10,000 per year. It is also fair to say that the poorer a person is, the more likely they are to be impacted by these same issues. By raising the maximum income level for consideration we observed dilution of multiple values which may be critical for identifying vulnerable populations. We suspect additional noise (i.e.,non-normal error distributions) exists within the linear probabilitymodels which failed to be removed through the step-wise backwards process. Notably, the *inca* variable, portraying persons receiving income assistance, was more valuable in the control than for any of the models which fell within the income limits for receiving this benefit. The ethical challenge for researchers, activists, and

Copyright © acascipub.com, all rights reserved.

International Journal of Geographic Information System
Vol. 4, No. 2, September 2017, pp. 1-23
Available Online at http://acascipub.com/Journals.php

legislators alike is defining what limit is going to define who is being covered by a decision, program, or policy.

While linear regression models may be adequate for establishing relationships between covariates associated to human trafficking, they ignore spatial complexities necessary for geographic modeling. Although the negative binomial output utilized a non-homogenous gamma distributed mean to rectify the Poissonian overdispersion due to outliers, the covariates could not define clustering propensities in the sampling dataset. Conversely, the spatial autocorrelation paradigm teased out heteroscedastic variables while quantitating aggregation tendencies. Eigenfunction, spatial filter, orthogonal, synthetic, decomposition algorithms can spatially differentiate clustering data sets by semi-paramterically regressable covariates (Griffith 2003, Jacob 2009).

Future spatial models could be fine-tuned through the inclusion of non-open source datasets. The relationship between high-calorie intake and BMI among people experiencing poverty is already well known (James *et al.,* 2001, James 2004, Bratanova *et al.,* 2016). Additionally, socioeconomic inequality has already been linked to anxiety symptoms (Marmot, 2004, Wilkinson and Pickett 2009, Pickett and Wilkinson 2010, Bratanova *et al*., 2016). Alternatively, research focused on youth impacted by poverty could benefit by aggregating data to the nearest high school and utilizing the school income presented by Coley *et al.* (2017). Furthermore, replication of changes over time may provide better insight towards the interplay among covariates with the distribution of impoverished persons, as well as provide greater insight on the impacts of measures taken to aid this vulnerable population.

While the Moran's I and Geary's C coefficients detected local clustering, the algorithms have an embedded bias towards non-zero autocorrelation. In this dataset negative autocorrelation could be considered an extreme observation. Meanwhile, resulting latent coefficients can represent the most intense cluster of data. An alternative approach would be to utilize a Bayesian-iteration to create simultaneous simulations of a post-positively biased dataset. Doing so permits consideration of negatively autocorrelated observations and may provide more valuable results.

Copyright © acascipub.com, all rights reserved.

International Journal of Geographic Information System
Vol. 4, No. 2, September 2017, pp. 1-23
Available Online at http://acascipub.com/Journals.php

4. Tables and Figures

Table 2.1: Selected variables

Independent variables		
Variable	**Description**	**Units**
fs	Food stamps/ SNAP recipients	Household
unemp	Civilian labor force, unemployed	Persons
inca	Received public assistance income	Household
lang1	Language spoken at home not English, speaks English "very well"	Persons
lang2	Language spoken at home not English, speaks English "well"	Persons
lang3	Language spoken at home not English, speaks English "not well"	Persons
lang4	Language spoken at home not English, speaks English "not at all"	Persons
nhi1	**No health insurance, under 18 years-old**	**Persons**
nhi2	**No health insurance, 18-34 years-old**	**Persons**
nhi3	**No health insurance, 35- 60 years-old**	**Persons**
nhi4	**No health insurance, 65 years-old or older**	**Persons**
nhi	***No health insurance, total population***	***Persons***
pph1	1-persons per household	Household
pph2	2-persons per household	Household
pph3	3-persons per household	Household
pph4	4-persons per household	Household
pph5	5-persons per household	Household
pph6	6-persons per household	Household
pph7	7-or-more-persons per household	Household
Dependent variables		
Variable	**Description**	**Units**
inc1	**Annual income less than $10,000**	**Household**
inc2	**Annual income $10,000- $14,999**	**Household**
inc3	**Annual income $15,000- $19,999**	**Household**
inc4	**Annual income $20,000- $24,999**	**Household**
inc5	**Annual income $25,000- $29,999**	**Household**
inc6	**Annual income $30,000- $34,999**	**Household**
inc7	**Annual income $35,000- $39,999**	**Household**
inc8	**Annual income $40,000- $44,999**	**Household**
inc9	**Annual income $45,000- $49,999**	**Household**
inc10	**Annual income $50,000- $59,999**	**Household**

Copyright © acascipub.com, all rights reserved.

International Journal of Geographic Information System
Vol. 4, No. 2, September 2017, pp. 1-23
Available Online at http://acascipub.com/Journals.php

inc11	**Annual income $60,000- $74,999**	**Household**
inc12	**Annual income $75,000- $99,999**	**Household**
inc13	**Annual income $100,000- $124,999**	**Household**
inc14	**Annual income $125,000- $149,000**	**Household**
inc15	**Annual income $150,000- $199,999**	**Household**
inc16	**Annual income $200,000 or more**	**Household**
lowinc1	***Annual income less than $10,000***	***Household***
lowinc2	***Annual income less than $15,000***	***Household***
lowinc3	***Annual income less than $20,000***	***Household***
lowinc4	***Annual income less than $25,000***	***Household***
lowinc5	***Annual income less than $30,000***	***Household***
lowinc6	***Annual income less than $35,000***	***Household***
lowinc7	***Annual income less than $40,000***	***Household***
lowinc8	***Annual income less than $45,000***	***Household***
lowinc9	***Annual income less than $50,000***	***Household***
lowinc10	***Annual income less than $60,000***	***Household***
lowinc11	***Annual income less than $75,000***	***Household***
lowinc12	***Annual income less than $100,000***	***Household***
lowinc13	***Annual income less than $125,000***	***Household***
lowinc14	***Annual income less than $150,000***	***Household***
lowinc15	***Annual income less than $200,000***	***Household***
lowinc16	***No maximum annual income; control***	***Household***
Variable aggregated into larger variable instead of tested independently		
Variable simulated by aggregating smaller variables for testing.		

Table 3.1: Goodness of fit coefficients comparing Poisson and negative binomial linear regression models.

Maximum Annual Income US Dollars	Poisson Pearson Chi-Square	Negative Binomial Pearson Chi-Square
$10,000	38.6383	**0.6176**
$14,999	36.4915	0.6428
$19,999	37.1145	0.6625
$24,999	38.8200	0.7019
$29,999	38.5031	0.7271
$34,999	39.9537	0.8331
$39,999	40.4783	0.8402
$44,999	39.0503	0.8626
$49,999	39.0005	0.8470
No maximum	**17.2900**	0.7100

Copyright © acascipub.com, all rights reserved.

International Journal of Geographic Information System
Vol. 4, No. 2, September 2017, pp. 1-23
Available Online at http://acascipub.com/Journals.php

Table 3.2: Linear regression coefficients per variable by maximum income threshold.

Maximum Annual Income	fs	unemp	inca	lang1	lang2	lang3	lang4	nhi
$10,000	**0.3019**	0.2558	**0.1715**	0.0203	0.2045	*	0.0628	**-0.0828**
$14,999	**0.4635**	0.2390	**0.2698**	*	0.2598	-0.0636	0.1185	**-0.1158**
$19,999	**0.5700**	0.2162	**0.3004**	*	0.2760	-0.0871	0.0758	**-0.0903**
$24,999	**0.5623**	0.1154	**0.4217**	*	0.3719	-0.0659	*	**-0.0673**
$29,999	**0.5870**	*	**0.3552**	0.0260	0.3871	-0.0742	*	*
$34,999	**0.6184**	*	**0.4982**	*	0.3955	-0.1793	*	0.0753
$39,999	**0.6195**	*	**0.4832**	*	0.3914	*	-0.1679	0.1147
$44,999	**0.6152**	*	0.4750	*	0.3899	*	-0.1909	0.1514
$49,999	**0.6356**	*	0.4573	*	0.3734	-0.0927	*	0.2243
No maximum	0.1556	*	0.7360	0.1959	-0.1588	-0.1211	*	0.2839

Maximum Annual Income	pph1	pph2	pph3	pph4	pph5	pph6	pph7
$10,000	0.1453	-0.1722	0.2523	*	*	-1.3159	-0.7983
$14,999	0.2767	-0.3664	0.4326	*	*	*	-1.5942
$19,999	0.3525	-0.4647	0.5170	*	*	*	-1.5575
$24,999	0.4493	-0.5835	0.7115	*	*	*	0.2054
$29,999	0.5268	-0.5829	0.5856	0.2987	*	*	-1.3898
$34,999	0.6141	-0.5571	0.6330	0.4960	*	*	-1.2852
$39,999	0.6677	-0.5330	0.6298	0.5157	*	*	-1.2243
$44,999	0.7382	-0.5796	0.8187	0.5911	*	-1.5062	*
$49,999	0.7282	-0.4940	0.9003	0.3148	*	*	-1.2258
No maximum	1.0631	1.2419	0.3301	1.5243	1.0103	*	*

Households fall within the limits to receive benefits from food assistance programs AND government subsidized health coverage AND are below the 2015 Federal Poverty Level.

Households fall within the limits to receive benefits from food assistance programs AND government subsidized health coverage, but are NOT below the 2015 Federal Poverty Level.

Households fall within the limits to receive benefits from food assistance programs AND are below the 2015 Federal Poverty Level, but are NOT eligible for government subsidized health coverage.

Households fall within the limits to receive benefits from food assistance programs based on annual income, but are NOT below the 2015 Federal Poverty Level.

Households represented fall within the limits to receive benefits from government assistance for the related variables based on 2015 standards.

Copyright © acascipub.com, all rights reserved.

Table 3.3: Z-score statistics for clustering of each tested dependent variable with the distance threshold utilized in the calculation.

Dependent Variable	Clustering Critical Value	Distance (m)
lowinc1	19.7653	1202.45
lowinc2	**19.9999**	1202.45
lowinc3	18.8873	1309.67
lowinc4	18.5842	1309.67
lowinc5	19.2298	**1416.89**
lowinc6	16.5109	1202.45
lowinc7	16.7303	1202.45
lowinc8	15.9796	1202.45
lowinc9	**14.9113**	**1095.22**
Control	12.095889	988.05
MAX	**19.9999**	**1416.89**
MEAN	**17.8443**	**1238.19**
MIN	**14.9113**	**1095.22**

Table 3.4: Moran's I and Geary's C coefficients for dependent variables tested.

Variable	Moran's i	Geary's C
lowinc1	**0.151**	**0.821**
lowinc2	0.162	0.813
lowinc3	0.162	0.793
lowinc4	0.168	0.787
lowinc5	0.170	0.795
lowinc6	0.164	0.777
lowinc7	0.17	0.783
lowinc8	0.173	**0.768**
lowinc9	**0.179**	0.779
Control	**0.197**	**0.817**
Max	**0.179**	**0.821**
Mean	**0.167**	**0.791**
Min	**0.151**	**0.768**

Copyright © acascipub.com, all rights reserved.

International Journal of Geographic Information System
Vol. 4, No. 2, September 2017, pp. 1-23
Available Online at http://acascipub.com/Journals.php

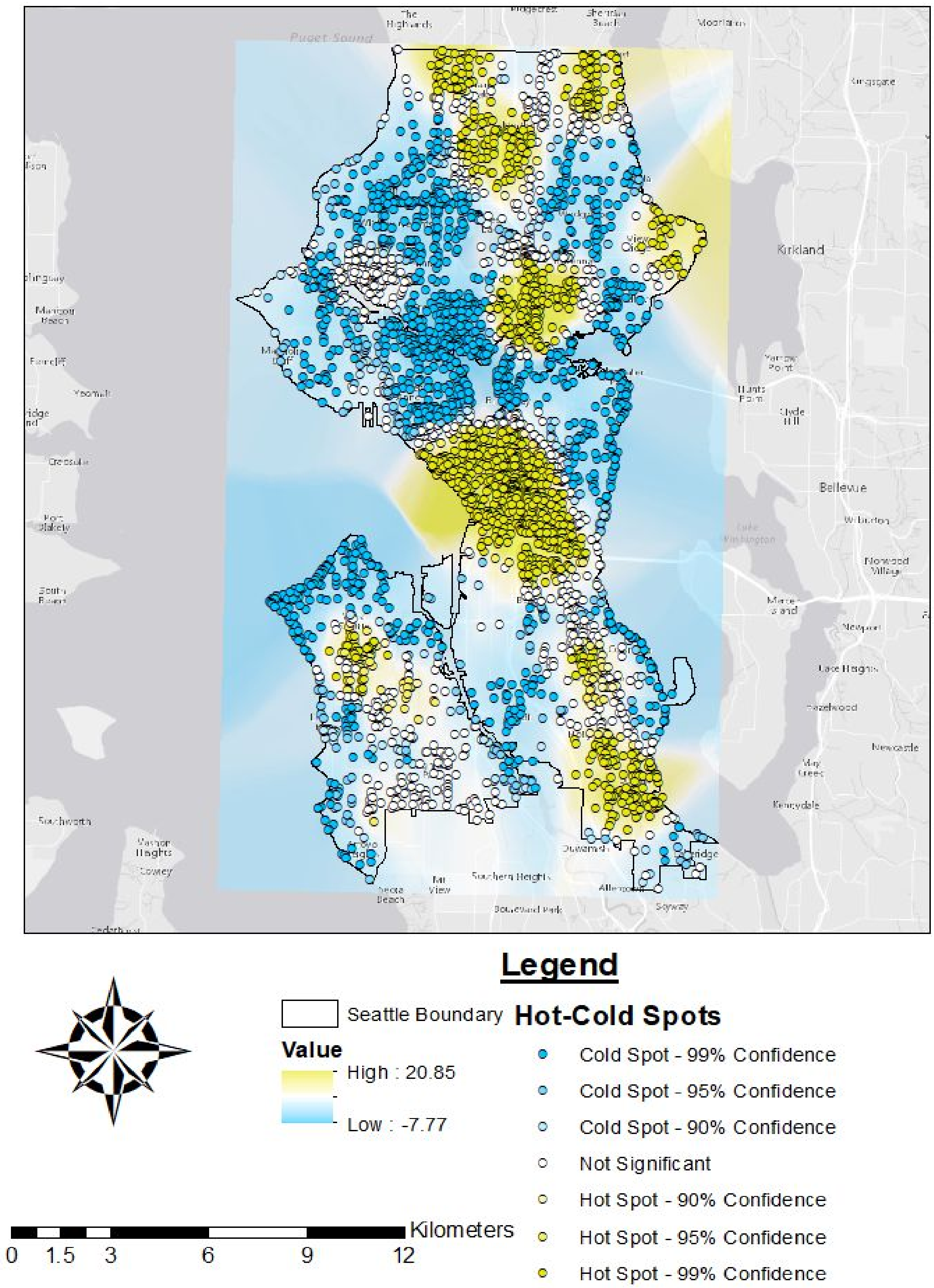

Figure 3.1: Clustering analysis of households earning less than $10,000 per year.

Copyright © acascipub.com, all rights reserved.

International Journal of Geographic Information System
Vol. 4, No. 2, September 2017, pp. 1-23
Available Online at http://acascipub.com/Journals.php

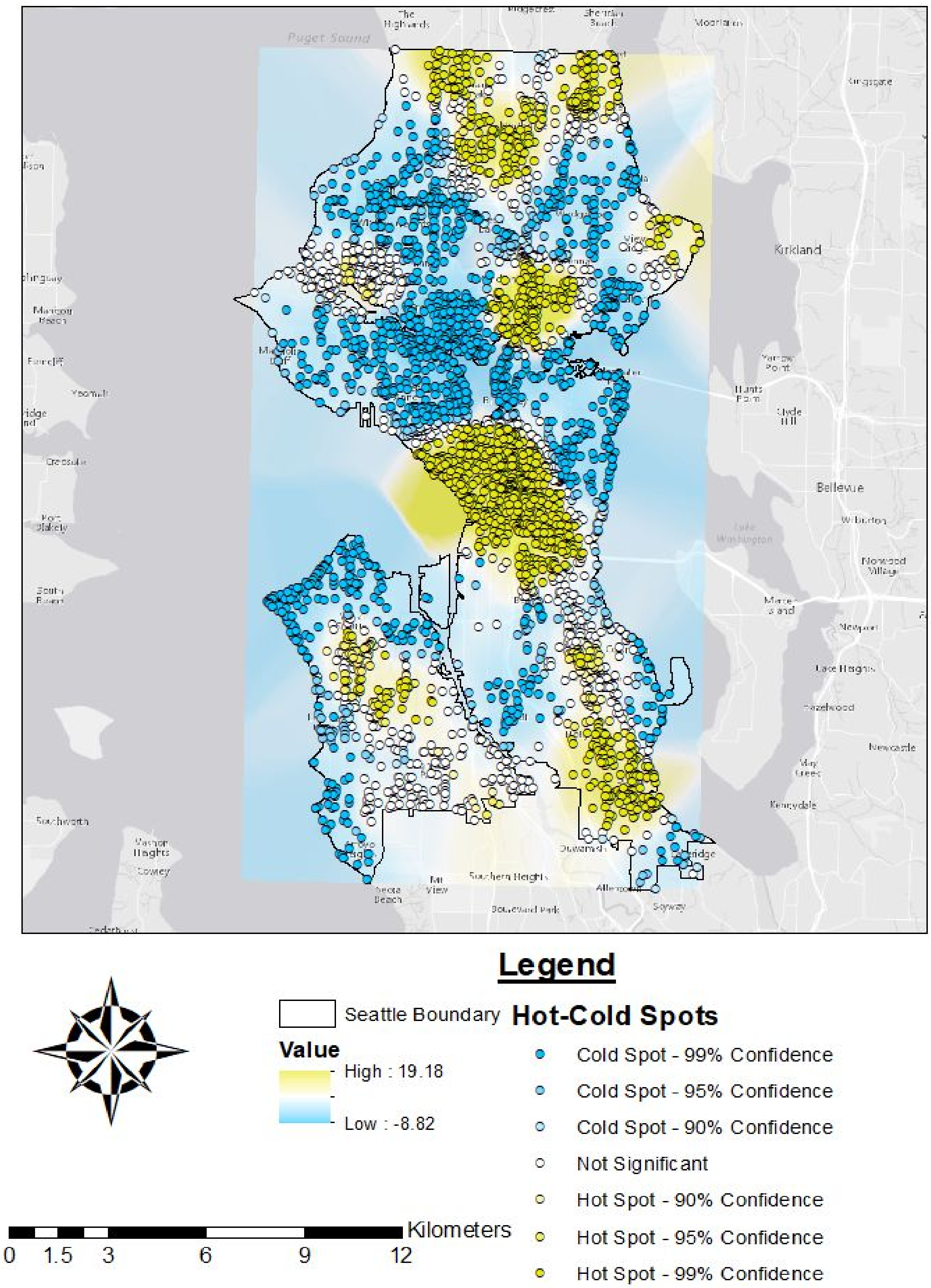

Figure 3.2: Clustering analysis of households earning less than $15,000 per year.

Copyright © acascipub.com, all rights reserved.

International Journal of Geographic Information System
Vol. 4, No. 2, September 2017, pp. 1-23
Available Online at http://acascipub.com/Journals.php

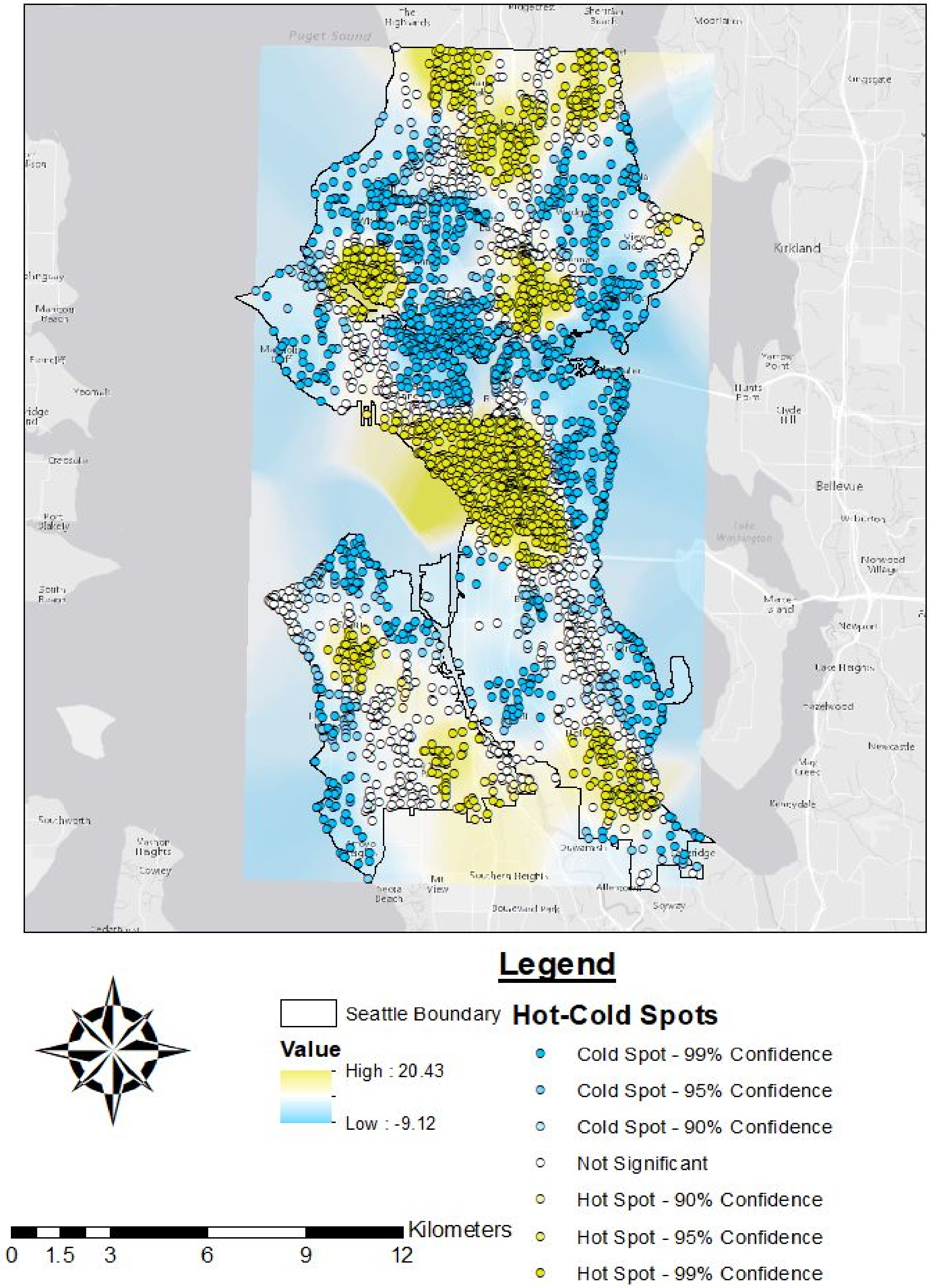

Figure 3.3: Clustering analysis of households earning less than $50,000 per year.

Copyright © acascipub.com, all rights reserved.

International Journal of Geographic Information System
Vol. 4, No. 2, September 2017, pp. 1-23
Available Online at http://acascipub.com/Journals.php

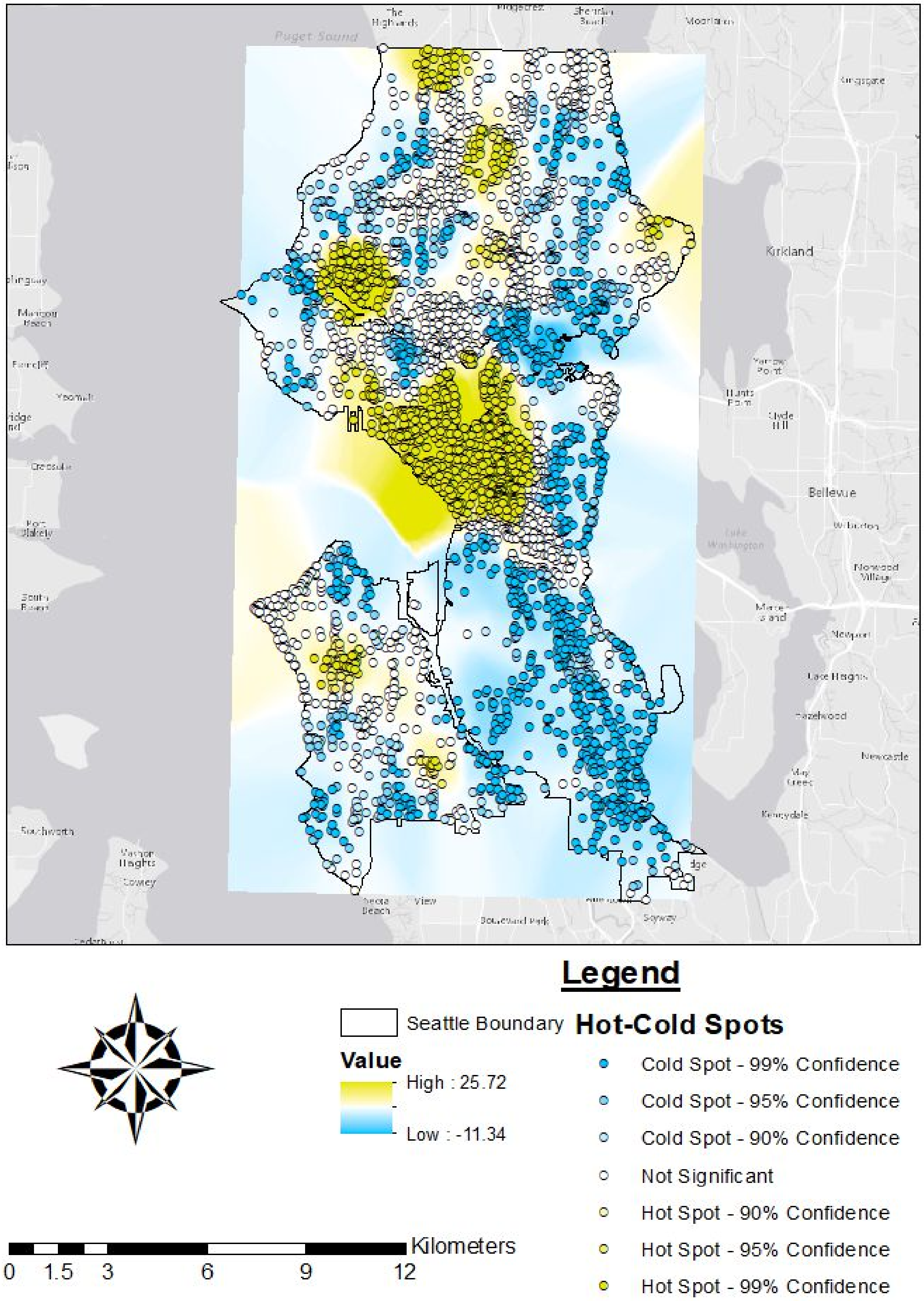

Figure 3.4: Clustering analysis of all households in Seattle, WA.

Copyright © acascipub.com, all rights reserved.

5. Acknowledgements

Joshua D. Anderson for assisting in colorizing tables and figures to be legible for color blind readers.

6.References

Akee, R., et al., 2007. Determinants of trafficking in women and children: Cross-national evidence, theory, and policy implications. *Migration and Culture, Frontiers of Economics and Globalization*, 8.

Albanese, J., 2007. A criminal network approach to understanding and measuring trafficking in human beings. In E. U. Savona and S. Stefanizzi (Eds.) *Measuring Human Trafficking: Complexities and Pitfalls.* New York, New York, USA: Springer Science + Business Media, LLC, 55-71.

Balk, G., 2014. Census: Seattle is the fastest-growing big city in the U.S. *The Seattle Times*, 22 May. Available from: http://blogs.seattletimes.com/fyi-guy/2014/05/22/census-seattle-is-the-fastest-growing-big-city-in-the-u-s/ [Accessed 27 June 2017].

Balk, G., 2015. Seattle no-longer America's fastest-growing big city. *The Seattle Times*, 21 May. Available from: http://www.seattletimes.com/seattle-news/data/seattle-no-longer-americas-fastest-growing-big-city/ [Accessed 27 June 2017].

Balk, G., 2017. Seattle once again nation's fastest-growing big city; population exceeds 700,000. *The Seattle Times*, 25 May. Available from: http://www.seattletimes.com/seattle-news/data/seattle-once-again-nations-fastest-growing-big-city-population-exceeds-700000/ [Accessed 27 June 2017].

Banks, J., et al., 2006. Disease and disadvantage in the United States and in England. *Journal of the Medical American Association* 295, 2037-2045.

Blandford, D., 2016. Seattle achieves record tourism growth for the third consecutive year, job growth out-paces the U.S. Available from: http://www.visitseattle.org/press/press-releases/seattle-tourism-statistics-announced/ [Accessed 27 June 2017].

Bratanova, B., et al., 2016. Poverty, inequality, and increased consumption of high calorie food: Experimental evidence for a causal link. *Appetite* 100, 162-171.

Cho, S.Y., 2015. Modeling for determinants of human trafficking: An empirical analysis. *SocialInclusion*, 3 (1), 2-21.

Hernandez, D., and Rudolph, A. 2015. Modern day slavery: What drives human trafficking in Europe? *European Journal of Political Economy,* 38 118-139.

City of Seattle, 2017. *Open data portal*. Available from: https://data.seattle.gov/ [Accessed 31 January 2017].

Cliff, A.D., and Ord, J.K., 1973. *Spatial autocorrelation*. London: Pion.

Coley, R. L., et al., 2017. Locating economic risks for adolescent mental and behavioral health: Poverty and affluence in families, neighborhoods, and schools. *Child Development*: 1-10. Available from: http://onlinelibrary.wiley.com.ezproxy.lib.usf.edu/doi/10.1111/cdev.12771/full [Accessed 13 June 2017].

Copyright © acascipub.com, all rights reserved.

Dearing, E. 2008. The psychological cost of growing up poor. *Annals of the New York Academy of Sciences* 1136, 324-332.

Freedman, D.A., 2005. *Statistical models: theory and practice.* Cambridge University Press.

Griffith, D.A., 2003. *Spatial autocorrelation and spatial filtering: gaining understanding through theory and scientific visualization*. Berlin: Springer-Verlag.

Jacob, B. G., et al., 2009. Developing operational algorithms using linear and non-linear least squares estima-tion in ArcGIS® and Python® for identification of Culex pipiens and Culex restuans aquatic habitats in a mosquito abatement district (Cook County, Illinois). *Geospatial Health* 3, 157-176.

Jacob, B. G., et al., 2013. A Bayesian Poisson specification with a conditionally autoregressive prior and a residual Moran's coefficient minimization criterion for quantitating leptokurtic distributions in regression-based multi-drug resistant tuberculosis treatment protocols. *Journal of Public Health and Epidemiology* 5(3), 122-143.

James, P. 2004. Obesity: the worldwide epidemic. *Clinics in Dermatology* 22, 276-280.

James, P., Leach, R., Kalamara, E. and Shayeghi, M. 2001. The worldwide obesity epidemic. *Obesity Research* 9, 228-233.

Jarvandi, S., Thibault, L, and Booth, D. A. 2009. Rats learn to eat more to avoid hunger. *Quarterly Journal of Experimental Psychology* 62, 663-672.

Kaiser, M., Cressie, N., 1997. Modeling Poisson variables with positive spatial dependence. *Statistics & Probability Letters* 35, 423-432.

Konrad, R.A., et al. 2017. Overcoming human trafficking via operations research and analytics: Opportunities for methods, models, and applications. *European Journal of Operational Research* 259, 733-745.

Leun, J. van der, and Schijndel, A. van., 2016. Emerging from the shadows or pushed into the dark? The relation between the combat against trafficking in human beings and migration control. *International Journal of Law, Crime and Justice* 44, 26-42.

Luthar, S.S., and Sexton, C.C. 2004. The high price of affluence. In R. Kail, ed. *Advances in child development and behavior.* London, UK: Academic Press, 125-162.

Luthar, S.S., Barkin, S.H., and Crossman, E. J. 2013. "I can therefore I must": Fragility in the upper-middle classes. *Development and Psychopathology* 25, 1529-1549.

Marmot, M. 2004. *The status syndrome: how social standing affects our health and longevity*. New York, NY: Henry Holt and Company.

Marmot, M. 2006. Status syndrome: challenge to medicine. *Journal of the Medical American Association* 295, 1304-1307.

McCullagh, P., and Nelder, J. A., 1989. *Generalized Linear Models*. Chapman and Hall, London: Chapman and Hall.

Ofuoko, A., and Uzokwe, U. N. 2012., Rural dwellers' perception of human trafficking and its implication for agricultural production in Edo State, Nigeria. *Asian Economic and Social Society* 2 (3), 394-404.

Copyright © acascipub.com, all rights reserved.

Pickett, K., et al., 2005. Wider income gap, wider waistbands? an ecological study of obesity and income inequality. *Journal of Epidemiology and Community Health*, 59, 670-674.

Pickett, K., and Wilkinson, R. 2010. Inequality: an underacknowledged source of mental illness and distress. *The British Journal of Psychiatry* 197, 426-428.

Pielou, E.C., 1969. *An Introduction to Mathematical Ecology*. New York: Wiley.

Polaris., 2017. 2016 *Hotline Statistics* [online]. Available from https://polarisproject.org/resources/2016-hotline-statistics [Accessed June 11, 2017].

Ryan, J., 2015. *Amid Seattle's affluence, homelessness also flourishes* [online]. Available from: http://www.npr.org/2015/04/07/398075834/amid-seattles-affluence-homelessness-also-flourishes [Accessed 27 June 2017].

Sapolsky, R. 2004. Social status and health in humans and other animals. *Annual Review of Anthropology* 33, 393-418.

Shively, C., and Clarkson, T. 1994. Social status and coronary artery atherosclerosis in female monkeys. *Arteriosclerosis and Thrombosis* 14, 721-726.

U.S. Census Bureau, NDGa. *QuickFacts: Seattle city, Washington* [online]. Available from https://www.census.gov/quickfacts/fact/table/seattlecitywashington/BZA010215 [Accessed 27 June 2017].

U.S. Census Bureau, NDGb. *Guided search*. Available from: https://factfinder.census.gov/faces/nav/jsf/pages/guided_search.xhtml [Accessed 31 January 2017].

U.S. Department of State., 2014. *Trafficking in Persons Report 2014*. Available from https://www.state.gov/j/tip/rls/tiprpt/2014/index.htm [Accessed 11 June 2007]

U.S. Department of State., 2016. *Trafficking in Persons Report 2016*. Available from https://www.state.gov/j/tip/rls/tiprpt/2016/index.htm [Accessed 11 June 2007].

Voloshin, D., et al., 2016. Identifying venues for female commercial sex work using spatial analysis of geocoded advertisements. *Procedia Computer Science* 80, 345-355.

Wilkinson, R., and Pickett, K. 2009. Income inequality and social dysfunction. *Annual Review of Sociology* 35, 493-511.

Copyright © acascipub.com, all rights reserved.