geographical analysis

Geographical Analysis (2020) **52**, 494–510

Special Issue

Tobler’s Law in a Multivariate World

Luc Anselin, Xun Li

Center for Spatial Data Science, University of Chicago, Chicago, IL USA

Tobler’s first law of geography is widely recognized as reflecting broad empirical realities in geography. Its key concepts of “near” and “related” are intuitive in a univariate setting. However, when moving to the joint consideration of spatial patterns among multiple variables, the combination of attribute similarity and geographical similarity that underlies the concept of spatial autocorrelation is much harder to deal with. This article uses the notion of distance in multiattribute space to explore and visualize the connection between “near” and “related” in a multivariate context. We approach this from a global, local, and regional perspective. We outline a number of ways to combine different visualization techniques and introduce a new local neighbor match test for multivariate local clusters. We illustrate the methods by means of Guerry’s classic data set on moral statistics in 1833 France.

Introduction

Some 50 years ago, in an article dealing with the simulation of population change in the city of Detroit, Waldo Tobler (1970, p. 236) referred in passing to the “first law of geography: everything is related to everything else, but near things are more related than distant things.” The main objective of this statement, viewed in the context of Tobler’s simulation exercise, was to be able to model change in population at a location as a function of values in nearby locations, that is, by means of a “local operator” (p. 237), rather than using all the values in the system. Tobler’s first law has been interpreted (and challenged) from a range of different perspectives, for example, as evidenced in the articles contained in the 2004 Forum of the *Annals of the Association of American Geographers* (Sui 2004, and, with reference to geographical analysis, especially Goodchild 2004 and Miller 2004).

The two core concepts in the law are “near” and “related.” The combination of geographical similarity (near) and attribute similarity (related) is of course the essence behind the notion of spatial autocorrelation, and Tobler’s first law is often seen as a fundamental principle that underlies spatial interpolation, geostatistics, and local analysis (e.g., Goodchild 2004; Miller 2004). In his rejoinder to the Forum, Tobler acknowledges that anisotropy and spatial heterogeneity may not only affect the validity of the law, but also states that “the fact that near things are more

Correspondence: Luc Anselin, Center for Spatial Data Science, University of Chicago, Chicago, IL, USA
e-mail: anselin@uchicago.edu

Submitted: 11 February 2020; Revised version accepted: 20 April 2020

doi: 10.1111/gean.12237

related than distant things seems a fundamental property of geography and rather easily explained" (Tobler 2004, p. 308). Tobler also points out that the statistician Ronald Fisher made a similar statement as early as 1935, citing "the widely verified fact that patches in close proximity are commonly more alike, as judged by the yield of crops, than those which are further apart" (Fisher 1935, p. 66).

The original formulation of the law is clearly situated in a univariate context, that is, dealing with the single variable of population size. Most discussions of the law similarly seem to be couched in settings that deal with a single variable. Whereas it is straightforward to generalize the concept of "near" to non-geographic spaces, such as network spaces (e.g., Sui 2004), it is far more challenging to move the notion of "related" from the univariate to a multivariate context (e.g., witness the difficulties in formulating a multivariate spatial autocorrelation coefficient, as reviewed in Anselin 2019, among others). In this article, we take the perspective that "related" is nothing but a different concept of "near," but situated in attribute space. As in Anselin (2019), we exploit the notion of distance between observations in multivariate attribute space to characterize "related."[1] Along the same lines, Tobler (2004) also suggested that the principle behind multidimensional scaling "equates similarity with distance," but we go beyond this context and apply the measure of distance to the attribute space itself as well. In essence, then, Tobler's law in a multivariate context boils down to assessing the extent to which nearness in geographic space matches nearness in multivariate attribute space, or, equivalently, whether geographic neighbors are also attribute neighbors.

In this article, we outline and introduce a number of methods to explore and visualize the connection between Tobler's "near" and "related" in a multivariate context. We approach this from three perspectives: global, local, and regional. The global perspective focuses on the concept of distance decay (implied by Tobler's law) when multiple variables are considered jointly. The local perspective deals with the identification of multivariate local clusters. The regional perspective treats the grouping of (geographically) nearby observations into larger aggregations that maximize internal similarity.

Depending on the available resolution of computer screen real estate, our techniques scale well to data sets of several thousands of observations in a modern dynamic graphics software environment that is based on linking and brushing (such as GeoDa, Anselin, Syabri, and Kho 2006). However, the static journal format motivated us to select a small pedagogic example to illustrate the various approaches (allowing for graphical presentation on a journal page). As in Anselin (2019), we use the classic data set on "moral statistics of France" reported in an 1833 essay by André-Michel Guerry (1833). This data set contains a collection of social indicators (e.g., crime, literacy, suicides, etc.) for 86 French departments (for an extensive discussion, see Friendly 2007). It has been previously used to illustrate multivariate statistical analysis from a spatial perspective, for example, in Dykes and Brunsdon (2007), and Dray and Jombart (2011).[2] We simply employ these data to illustrate the various concepts and do not pursue a substantive interpretation.

In the remainder of the article, we start by describing some characteristics of our data set. Next, we define "nearness" in multiattribute space as the Euclidean distance between observation points in space and assess how this varies with geographic distance. The following three sections deal with multivariate local clusters. First, we review the multivariable local Geary (Anselin 2019) and use its findings as a reference for the other methods. We introduce a new local neighbor match test based on the match between nearest neighbors in geographic and multiattribute space. We also apply this approach to nearest neighbors after dimension reduction through multidimensional scaling (MDS). Finally, we discuss some aspects of regionalization, or, spatially constrained clustering. We close with some concluding comments.

Moral statistics of 1833 France

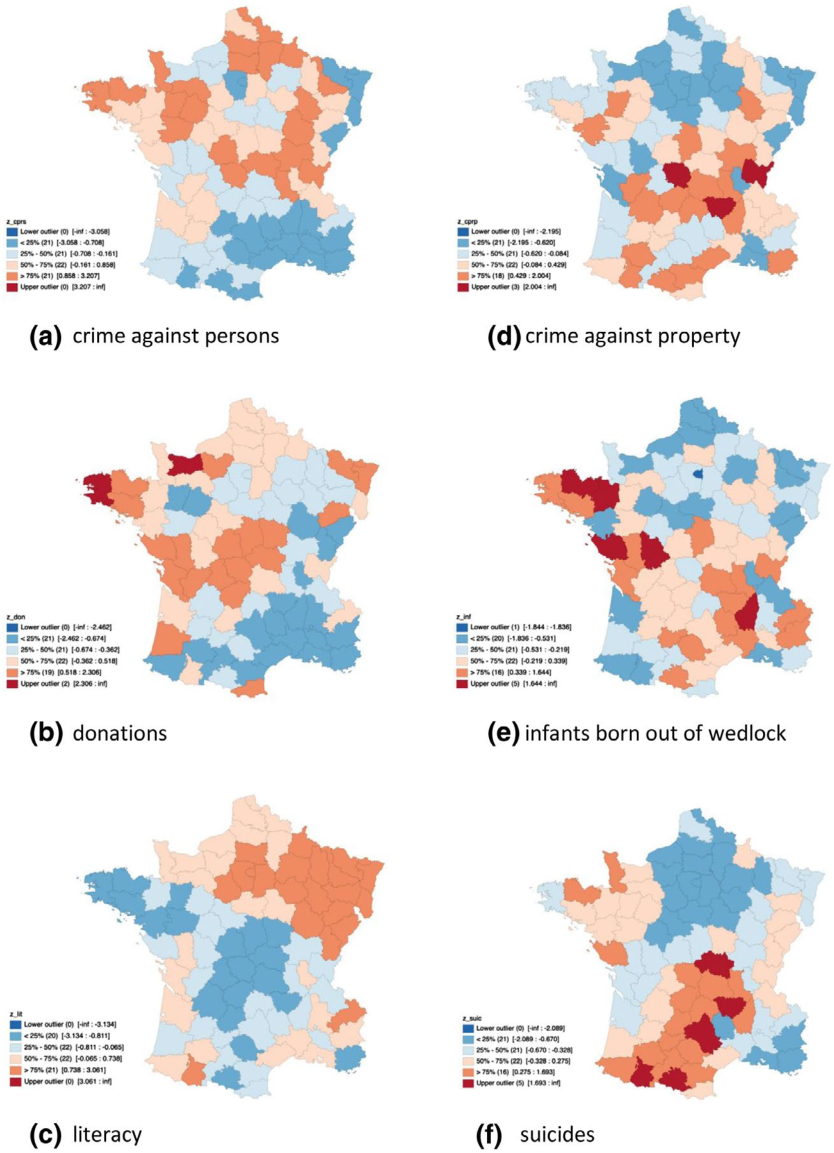
In the illustrations that follow, we use the same six variables as in the analysis by Friendly (2007), Dray and Jombart (2011), and Anselin (2019), that is, crimes against persons (cprs), crimes against property (cprp), literacy (lit), donations (don), infants born out of wedlock (inf), and suicides (suic). As in the original study, the variables are expressed in a somewhat unusual scale, in that larger values imply a “better” situation, for example, a larger value for crime corresponds to a lower crime rate (more precisely, the crime variable is population over crime, rather than the more customary crime over population). All variables have been standardized such that their means equal 0 and they have unit variance. In addition, the island of Corsica is dropped from the data set, resulting in a total of 85 contiguous observations.

This collection of variables constitutes a particular challenge for multivariate analysis. Not only are the correlations among them weak, but also their spatial patterns show very little commonality as well. In Table 1 of Anselin (2019, p. 141), the correlations are shown to be both negative and positive and range in absolute value from 0.021 (between crime against persons and literacy) to 0.412 (negative, between infants and literacy). The box maps (Anselin, Syabri, and Kho 2006) shown in Fig. 1 illustrate the lack of common spatial patterns. For example, for crimes against persons and donations, the map seems to portray mostly low values in the south, whereas for crimes against property and suicides the low values (i.e., high crime and high suicides) are in the north. Except for crimes against persons and literacy, the maps also contain several outliers, at both the high end and the low end of the distribution (the latter only for literacy).

The six variables are characterized by high positive spatial autocorrelation based on Moran’s I, for example, for k-nearest neighbor weights (with $k = 6$), as listed in the second column of Table 1. All but one Moran’s I coefficient are highly significant at $P < 0.001$ (infants is at $P < 0.002$), suggesting considerable clustering. The challenge consists of assessing the extent of clustering and the location of clusters of observations where the spatial pattern of the individual variables results in clusters of variable *tuples*, in which all the variables are considered jointly.

Table 1. Moran’s I and Distance Matrix Correlations

Variable	KNN-6	Distance correlation
Crime persons	0.42	−0.055
Crime property	0.27	0.168
Donations	0.26	0.152
Infants	0.21	0.060
Literacy	0.68	0.245
Suicides	0.42	0.087
Crime person-property-donations		0.164
All six variables		0.223



(a) crime against persons

(d) crime against property

(b) donations

(e) infants born out of wedlock

(c) literacy

(f) suicides

Figure 1. Box map for the six variables.

Near in multivariate space—Smoothed distance scatter plot

The geographic distance between any pair of observations can be readily computed as the Euclidean distance between their coordinates (or, in our example, between the centroids of the departments):

$$d_{ij} = \left[(x_i - x_j)^2 + (y_i - y_j)^2 \right]^{1/2},$$

where (x_i, y_i) is the tuple with the geographic coordinates of observation i. In our example, the inter-departmental distances range from 27.4 to 962.0 km, with an average of 393.6 km.

In a similar vein, the Euclidean distance between two observations i and j in multivariate attribute space can be defined as:

$$v_{ij} = \left[\Sigma_{h=1}^{k} \left(z_{hi} - z_{hj} \right)^2 \right]^{1/2}$$

with z_{hi} as the value for each individual variable z_h at observation i, corresponding to an element of the tuple $(z_1, z_2, \ldots, z_k)$ at location i.[3]

The idea of combining a measure of distance in attribute space with geographical distance is not new. For example, Oden and Sokal (1986) implemented a Mantel test to assess the similarity between the elements of a geographic distance matrix and a variable dissimilarity matrix (essentially the same concept as the attribute distance matrix used here), although their approach differs somewhat from ours.[4]

The correlation (or rank correlation) between the elements of the two distance matrices is an intuitive summary indicator of the similarity between them. In the third column of Table 1, we report the Pearson correlation between the 3,570 (upper-diagonal) elements of the geographic distance matrix and the attribute distance matrix for each of the variables.[5] We would expect this correlation to be positive (locations further apart should be less similar), but this is not the case for crime against persons (−0.055). The other variables show the expected positive value, ranging from a low 0.060 for infants to 0.245 for literacy. Ignoring the result for crime against persons, this also matches the ranking of these variables on the global spatial autocorrelation measured by Moran’s I (second column of Table 1).

We applied this correlation approach to two multivariate settings. In addition to considering all six variables jointly, we also give a few examples for a subset of three variables (crime against persons, crime against property and donations). We use the subset of variables to visualize location in attribute space in three dimensions (by means of a three-dimensional scatter plot cube), which is challenging for higher dimensions.

The results for the correlation between the multivariate attribute distance and the geographic distance are listed in the two bottom rows of the third column in Table 1. For the three-variable case the value is 0.164 and for the six-variable case 0.223, suggesting a weak correspondence (since we did not carry out a formal significance test, we can only use the values as a rough guide).

One drawback of the correlation coefficient is that it measures a *linear* association between geographic distance and attribute distance, whereas the distance decay suggested by Tobler’s first law would imply a nonlinear relation, with pairs further apart less correlated than pairs closer together (in addition, pairs that are sufficiently far apart should not be correlated at all). We use a *smoothed distance scatter plot* to visualize the attribute distance against the geographic distance. This is similar to a smoothed variogram scatterplot, except that it employs the *square root* of the sum of squared difference for all variables between pairs of observations, whereas the (semi) variogram is based on the *expected value* of the squared difference itself. For example, Bourgault and Marcotte (1991) outline a multivariable variogram, which portrays the expected weighted sum of the squared differences between pairs of observations for each of the variables by distance band. Our approach is similar in spirit, but it differs in that it is not the sum of the individual squared differences, but the square root of the sum that is considered, and all distance pairs are included.[6]

A well-known drawback of any smoothed variogram scatter plot is that it is easily affected by outliers, especially for larger (geographical) distance values. Our smoothed distance scatter plot suffers from the same problem. We therefore follow the suggestion from geostatistics to limit the distance range to the so-called "distance of reliability." Journel and Huijbregts (1978) suggests this to be one half of the maximum distance, that is, 481 km in our example. After eliminating pairs of observations that are separated by distances that exceed the distance of reliability, we retain 2,407 pairs.

Figs. 2 and 3 present the smoothed distance scatter plots for, respectively, the six individual variables and the two multivariate cases. The scatter plot points themselves are not shown, only the loess smooth is depicted.

The graphs for the individual variables in Fig. 2 have the familiar shape of a (semi) variogram, increasing with distance, first fairly steep, then, flattening out. Interestingly, except for literacy, they all show substantial attribute difference near the origin (the so-called nugget effect). Overall, the curves depict a much more complex pattern of spatial association than suggested by either Moran's I or the distance matrix correlations. Nevertheless, the general trends are similar, with the steepest increase of the attribute distance with geographical distance for literacy, which also has the largest Moran's I and strongest distance correlation. By the time the distance of reliability cutoff is reached, the curve for literacy still has not tapered off. On the other end of the spectrum, the low Moran's I and correlation for infants is reflected in the almost flat and near-linear curve in the plot. For the remaining variables, the curves show a slow increase with distance up until somewhere between 250 and 300 km (the first quartile in the distance distribution is 247 km), after which the curve flattens out, or even decreases. In sum, except for the variable infants, the smoothed distance scatter plots for the individual variables are in accordance with Tobler's first law.

The smoothed distance scatter plots for the two multivariate cases in Fig. 3 show a remarkable regularity with distance. Both rise gradually with distance to flatten off around 300 km for

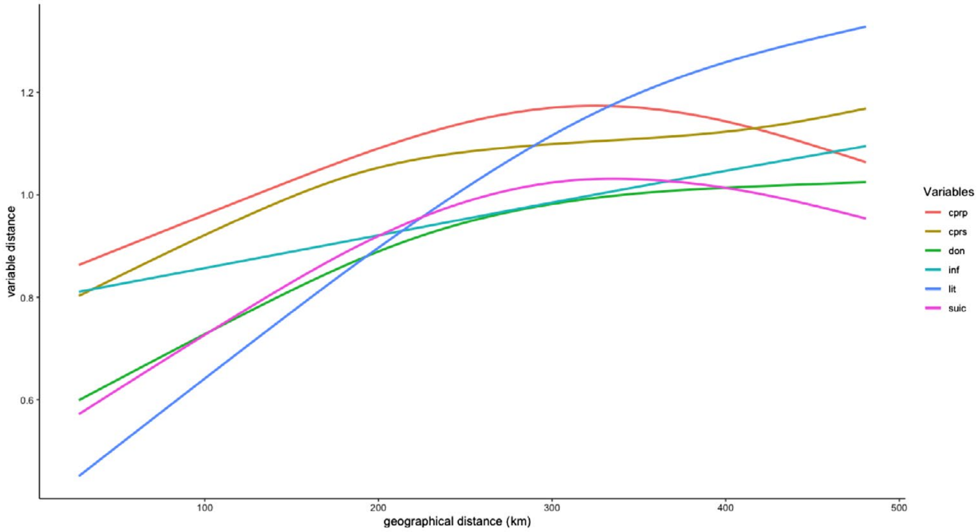


Figure 2. Smoothed distance scatter plot for individual variables.

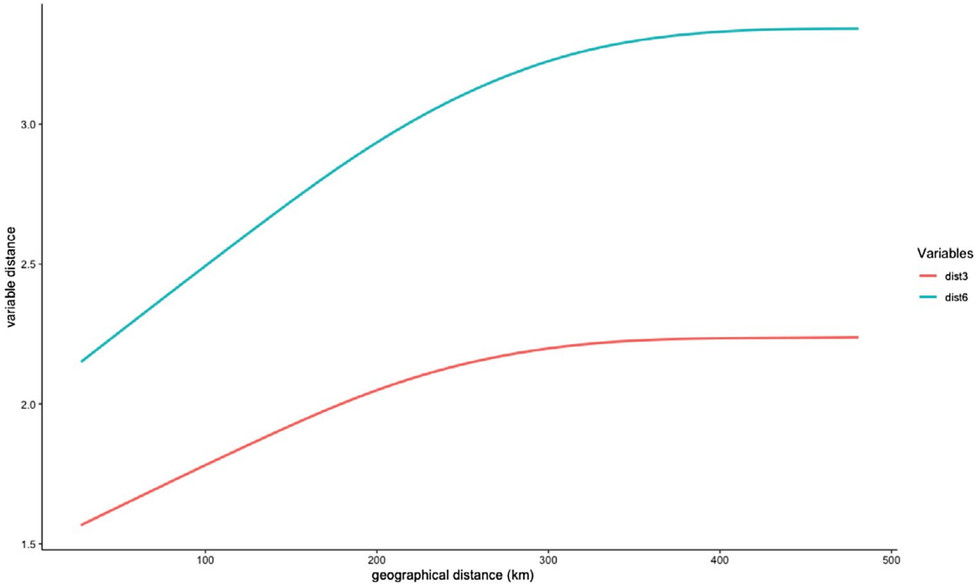


Figure 3. Smoothed multivariate distance scatter plot.

the three-variable case and around 400 km for the six-variable case. Both also illustrate the curse of dimensionality, with growing attribute distances as the number of variables increases, as well as a much larger nugget effect. At the global level, then, in our example Tobler's first law also seems supported in the multiattribute case. We now turn to a *local* analysis.

Multivariate local Geary

A local indicator of spatial association, or LISA, provides a statistic for each observation that indicates the extent to which the arrangement of the value for a variable at that location and the values at neighboring locations is different from what it would be under spatial randomness (Anselin 1995). The local Geary statistic provides such a measure as the weighted average of the squared difference between a variable x_i observed at location i and its neighbors, as defined by the elements w_{ij} of a row-standardized spatial weights matrix:

$$c_i = \Sigma_j\, w_{ij}\left(x_i - x_j\right)^2.$$

Significance of the statistic is determined through a conditional permutation approach. A multivariate version of the statistic was suggested in Anselin (2019) as the sum of the local Geary statistics for each variable under consideration:

$$c_{mi} = \Sigma_{h=1}^{k} \Sigma_j w_{ij}\left(z_{hi} - z_{hj}\right)^2 = \Sigma_{h=1}^{k} c_{hi},$$

in the same notation as above. Inference can again be based on a conditional permutation approach (for details, see Anselin 2019). Upon closer examination, the multivariate local Geary statistic turns out to be a weighted average of the distance in attribute space between the observation and its geographical neighbors, summarizing the effect of the neighbors into a single statistic.

We illustrate this in our example for both the three-variable case and the six-variable case. A significance map of the local Geary statistics in the three-variable case is given in Fig. 4, showing significant locations with higher significance associated with a darker color, in the usual fashion (Anselin 1995; Anselin, Syabri, and Kho 2006). The pseudo significance is based on 99,999 permutations with a cutoff of $P < 0.01$, and using a k-nearest neighbors weights matrix with $k = 6$.[7] The map is not simply an overlay of univariate significant locations, since the statistic contains a trade-off between the distances in attribute space for the different variables. In fact, for the three-variable example considered here, there is no overlap between the significant locations for the local Geary statistic for each variable separately (with $P < 0.01$).

To highlight the complex interplay between geographical distance and distance in attribute space, we select the two most significant departments (Basses-Alpes and Drome in the southeast of the country, both with $P < 0.0001$) and situate their location in the three-dimensional scatterplot cube in Fig. 5. The two yellow points are next to each other in the 3-D cube, and thus, are neighbors in both spaces (in the 3-D cube, the x-axis is cprs, the y-axis is cprp, and the z-axis is don). The geographic neighbors (k-nearest neighbors) of the two selected observations are connected in attribute space through the red lines, illustrating the extent of the trade-off in similarity that occurs across dimensions. Some of these neighbors are much closer than others in attribute space.

The significance map for the multivariate local Geary statistics for all six variables jointly is shown in Fig. 6. The map not only shows some overlap with the three-variable case, but also some interesting differences. For example, the department of Drome, which was one of the two most significant ones for the three-variable case, is no longer significant for the six variables. In

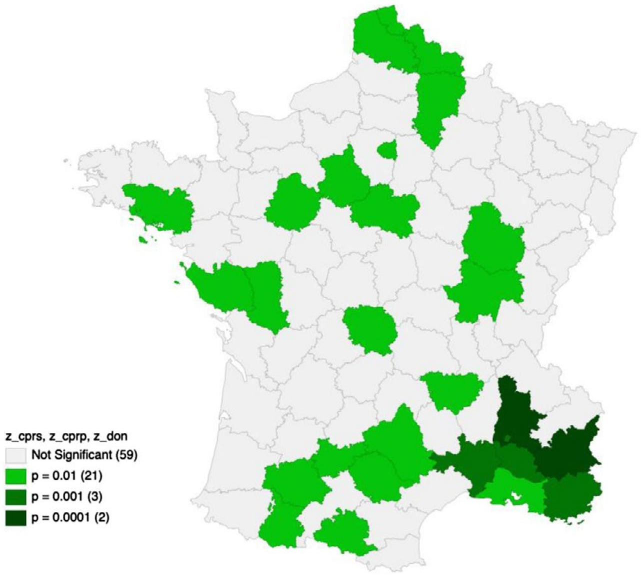


Figure 4. Multivariate local Geary significance map (cprs, cprp, don).

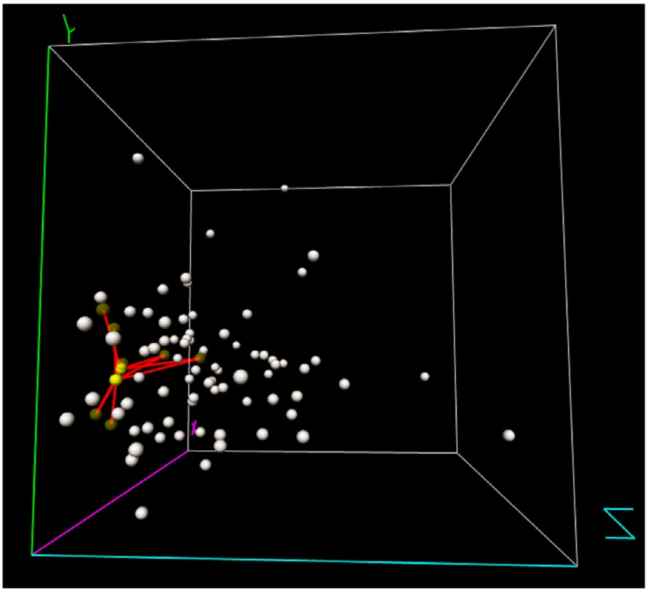


Figure 5. Neighbors in 3-D attribute space.

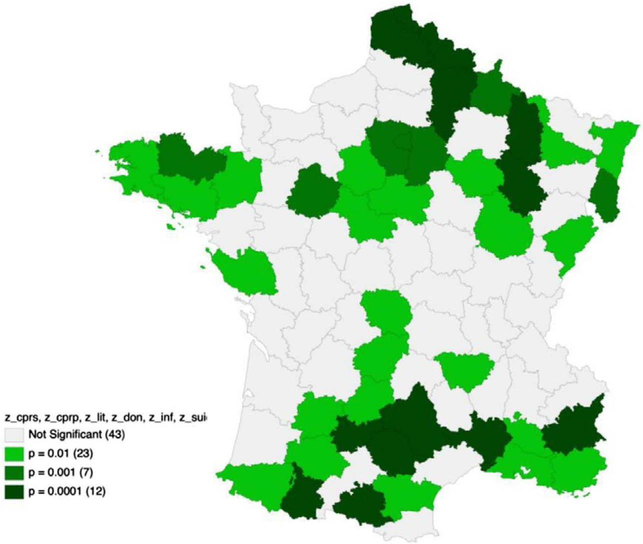


Figure 6. Multivariate local Geary significance map (six variables).

addition, many more locations are identified as "significant" at $P < 0.0001$ (12 compared to two in the three-variable case). This illustrates one of the drawbacks of this approach, which suffers from the curse of dimensionality. As is well known, in higher-dimensional spaces the distances between the observations become much larger (see also Fig. 3), and the volume of space needed to find nearest neighbors in attribute space increases exponentially. This seems to suggest that as more variables are included in the analysis, more random "neighbors" will tend to be further apart in attribute space, yielding more locations as "significant" (i.e., with the smallest value for the statistic relative to the random permutations). This needs to be further investigated, but it does suggest that the indication of "significance" needs to be interpreted with caution (see Anselin 2019, for a more in-depth discussion).[8]

The local Geary index summarizes the nearness in attribute space of the geographical neighbors of an observation into a single index. Next, we investigate the extent of a match between the two kinds of neighbors more explicitly.

Neighbors in different spaces—Local neighbor match test

In a low-dimensional setting, the match between neighbors in geographic space and neighbors in multiattribute space can be assessed visually by means of brushing a linked map and graph. In two dimensions, neighbors on a scatter plot can be located on a map, and vice versa, neighbors identified by a brush on the map can be located in the scatter plot. For three dimensions, the same operation can be carried out using a linked map and three-dimensional scatter plot cube. In higher dimensions, this is less straightforward. A linked map and parallel coordinate plot would allow to find the paths that correspond to neighbors on the map, but it is much harder to assess "nearness" in multivariate space from the PCP.

An alternative approach consists of identifying the *k*-nearest neighbors in the two spaces, that is, the geography of locations, on the one hand, and the neighbors in multiattribute space, on the other. Each of these neighbor sets can be formally represented by a spatial "weights" matrix, in the usual fashion. The only different aspect is that the nearest neighbors in attribute space are based on the corresponding distance matrix.

With the two weights matrices in hand, the degree of overlap among the neighbors can be assessed. A *neighbor cardinality map* shows for each location the number of common neighbors between the geographic space and the attribute space. A new method to identify multivariate local clusters consists of identifying those locations where the match exceeds a certain probability threshold, as a *local neighbor match test*.

The probability for an observation of having *v* neighbors in common out of the *k* between the two weights can be readily computed. With *n* total observations, the pool of possible neighbors is $N = n - 1$. The probability of having v common neighbors in a draw of *k* from *N* is:

$$p = C(k, v) . C(N - k, k - v) \ / \ C(N, k),$$

with *C* as the combinatorial operator. This is only a rough guide of significance, since it ignores multiple comparisons and other related problems, but it suffices to assess how likely a particular configuration is expected to occur. The probability is clearly a function of the total number of observations, and the probability of a number of common neighbors will become exceedingly rare in larger data sets, requiring an adjustment of the "k" used in the analysis in order to find meaningful results.[9]

Fig. 7 illustrates the application of the new local neighbor match test to our example with six variables. The neighbor cardinality map shows the number of matches between six nearest neighbors in the two spaces. Of most interest are the single location with five matches ($P = 0.00001$, the department of Tarn in the south) and the four locations with four matches ($P = 0.00011$, the departments of Tarn-et-Garonne and Gard in the south, Haute-Marne in the north-east and Nord

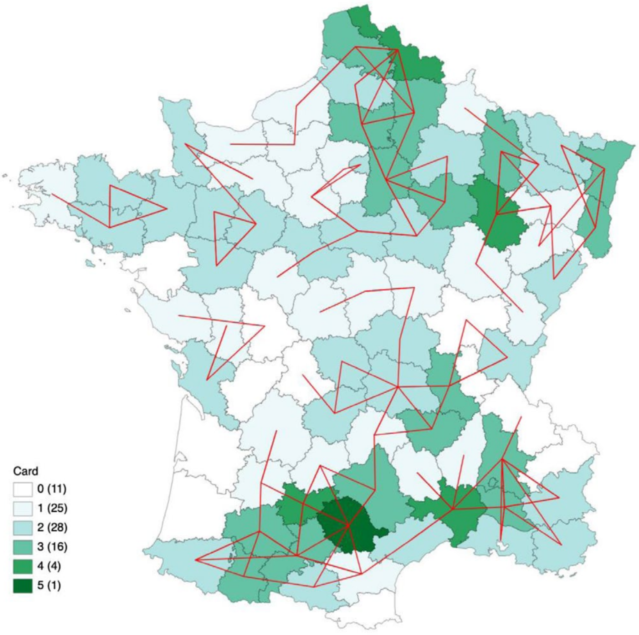


Figure 7. Matching neighbor cardinality (knn-6).

in the north). All five of these locations were also identified as highly "significant" ($P < 0.0001$) by the multivariate local Geary statistic (see Fig. 6). In our example, having three matches has $P = 0.004$ and two matches has a $P = 0.053$. Overlayed on map is the network structure of the intersection between the 2 knn weights matrices. Note that the *k*-nearest neighbor property is not symmetric and the links on the map only show the existence of a link, not the direction associated with it. For example, at first glance, there may seem to be six connections between the department of Tarn and its neighbors. However, only five of these links are between Tarn and its neighbors, one is between Ariege and Tarn. More precisely, Tarn is among the six nearest neighbors of Ariege, but not the other way around.

Our new test provides an alternative to the multivariate local Geary statistic for identifying "interesting" locations.

Dimension reduction—Multidimensional scaling

A traditional approach to deal with the curse of dimensionality is to reduce the dimension in the variable space. For example, the much-used principal components replace the original variables with a (much) smaller subset, each consisting of a linear combination of those variables. MDS is an alternative that projects the observations points in the multiattribute variable space onto a lower-dimensional space, typically in two or three dimensions (see, e.g., Golledge and Rushton 1972, for an early discussion of application in geography). Tobler and Wineburg (1971) used a similar approach to extract locations of pre-Hittite Assyrian merchant colonies in Bronze Age Anatolia. They inverted a gravity model based on co-mentions of places on tablets to obtain inter-settlement distances, and then, converted those to actual coordinates in geographic space. In the same general sense that we have advocated throughout the article, the results of an MDS analysis can be interpreted as showing observations that are "close" in multivariate attribute space located "near" each other in the MDS plot (see also Tobler 2004).

Since the output of an MDS analysis is a low-dimensional plot, we can apply the visualization of the connection between neighbors in geographic space and neighbors in attribute space as outlined above. For example, one could brush a two-dimensional MDS scatter plot or a three-dimensional MDS scatter plot cube to assess the extent to which neighbors are also neighbors on a map, and vice versa.

In addition, we can apply the concept of a local neighbor match test to the *k*-nearest neighbors of the observations as reflected in the MDS plot. However, since the fit of the stress function used in the optimization process is not perfect (i.e., the difference between pairwise distances in the original dimension and the distances implied by the coordinates in a lower-dimensional space), there will be discrepancies in the relative locations between the two spaces. In addition, there are several different ways to implement the idea behind MDS. While the original method (and the one used in our example) is linear and metric, nonmetric and nonlinear approaches have been suggested as well, each attempting to conserve different aspects of the nearness in multiattribute space (see, e.g., Lee and Verleysen 2007).

In our example, a sense of the fit between the two sets of nearest neighbors can be found from the profile of the intersection between the 2 knn weights. The median number of neighbors in common between a three-dimensional MDS six nearest neighbors and full multiattribute six nearest neighbors is only four (a complete match would yield six). One location has only one neighbor in common between the two weights, and only two have all six in common.

With this caveat in mind, we illustrate the neighbor cardinality map for the six nearest neighbors from a three-dimensional MDS plot in Fig. 8, with the associated network structure overlayed on the map. As in Fig. 7, we find one location with five matches (but it is a different location in Fig. 8), but only one location with four matches (as opposed to four such locations in Fig. 7). The top location (the department of Vaucluse) is not one identified as most significant using the multivariate local Geary, but the second one is (the department of Basses-Alpes). The trade-offs involved in using the *k*-nearest neighbors for the original dimension versus the lower dimensional MDS remain to be investigated further.

Regionalization

As a third perspective, we consider a regional scope. In a sense, this concerns dimension reduction in the observation space, in that observations (locations) that are "related" are grouped into a smaller number of larger entities, that can be referred to as "regions." The grouping of observations is the subject of clustering, a topic much too broad to treat in-depth here. We provide only a brief overview of the main approaches.

The main issue confronted in this approach follows from the spatial characteristics of the results of traditional clustering techniques, either partitioning methods or hierarchical clustering. The resulting groupings of observations are similar in attribute space, but they do not necessarily constitute meaningful spatial entities. This has led to a range of approaches to embed classic clustering methods with spatial constraints, either by forcing the results to be contiguous, or by allowing a trade-off between the attribute and spatial objectives. The literature is large, with early reviews in Murtagh (1985) and Haining, Wise, and Ma (2000). More recent treatments and examples of different approaches can be found in Murray and Grubesic (2002), Assunçao et al.

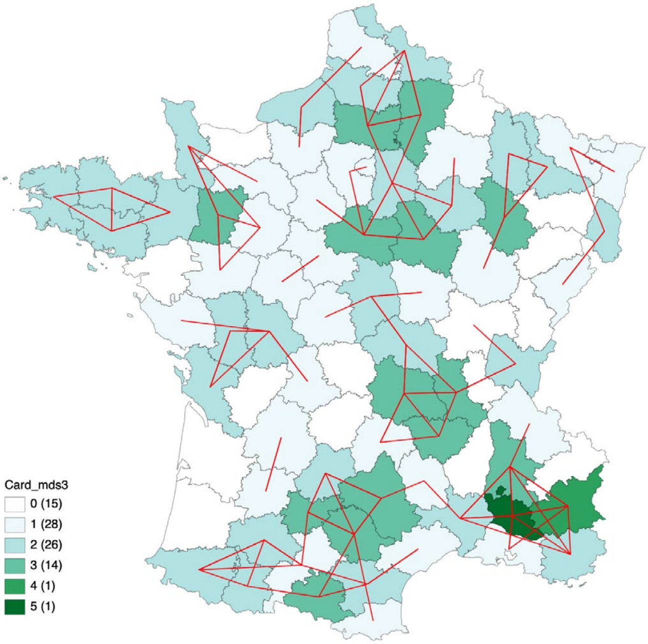


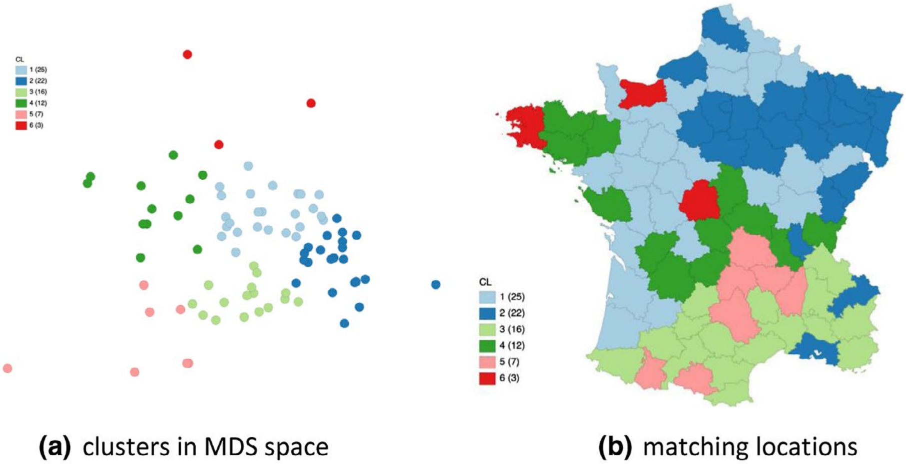
Figure 8. MDS3 matching neighbor cardinality (knn-6).

(2006), Duque, Ramos, and Suriñach (2007), Guo (2008), Recchia (2010), Duque, Church, and Middleton (2011), Miele, Picard, and Dray (2014), and Chavent et al. (2018), among others.

The main objective of this section is to illustrate two nonstandard ways in which the tension between geographic similarity and attribute similarity can be visualized by exploiting the results of an MDS exercise.

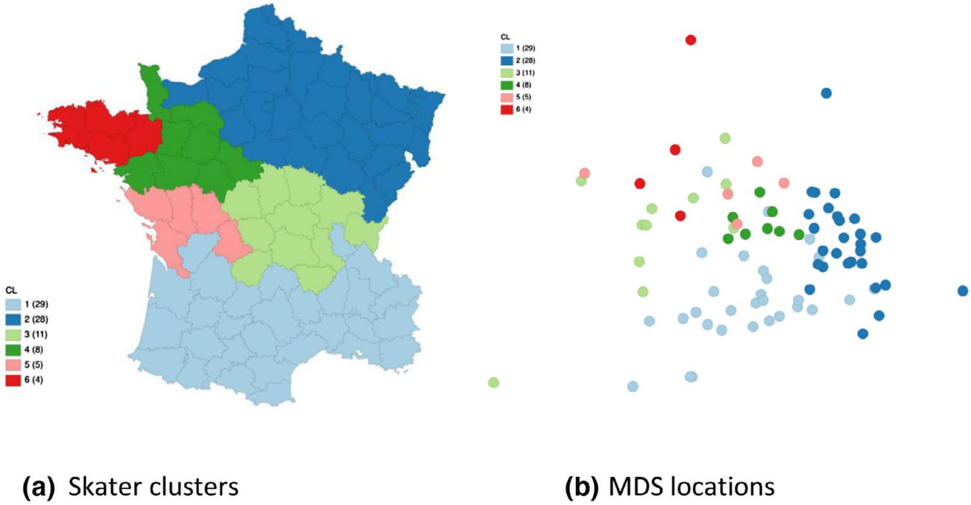
First, we consider the points represented by a two-dimensional scatter plot that corresponds to the results of an MDS analysis. We continue with our example using the same six variables, which yields the points shown in Fig. 9a (without any axes). Since only the coordinates of these points are taken into account in any traditional clustering exercise, the results by construction will consist of compact regions. Moreover, these are regions in attribute space, and the question remains to what extent these are also contiguous in geographical space.

In Fig. 9a, we show the results of a *K*-means clustering application yielding six clusters.[10] The application of *K*-means to the two-dimensional MDS points allows us to visualize “regions” in multiattribute space which would not be possible to visualize in the original space (i.e., with the *K*-means applied to the six original variables, in six-dimensional variable space). We can now assess the geographic distribution of these regions in the map shown in Fig. 9b. The attribute clustering matches spatial units that are far from contiguous. Only cluster 3 (note that the labels are arbitrary) yields a grouping of 16 contiguous departments, but none of the other clusters do. One, cluster 6, results in the grouping of three completely disparate departments, and cluster 2, containing 22 departments, consist of no less than seven separate spatial units. In other words, the neighbors in attribute space only partially match neighbors in geographic space. The resulting spatial units show some, but an imperfect similarity to the multivariate local Geary results in



(a) clusters in MDS space **(b)** matching locations

Figure 9. Neighbors in MDS and geographical neighbors.



(a) Skater clusters **(b)** MDS locations

Figure 10. Neighbors in geographic clusters and locations in MDS.

Fig. 6. The connection between the two remains to be explored further. Other values of k could be explored, which may or may not yield more acceptable spatial regions.

A second perspective is offered by reversing the logic. We start by applying a spatially constrained clustering technique such as Skater (Assunçao et al. 2006) to the six variables, enforcing spatial connectedness by means of queen contiguity spatial weights. We again chose six regions as an illustration, shown in Fig. 10a.[11] In contrast to the results in Fig. 9b, now the geographical units are all contiguous by construction. Again, there are broad similarities with the patterns suggested by the multivariate local Geary (Fig. 6). The counterpart of the spatial regions in the MDS scatter plot is shown in Fig. 10b. This illustrates the trade-offs that have to be made in attribute similarity in order to satisfy the contiguity constraint. For example, cluster 3, with 11 observations, contains points that are among other points from cluster 1, 4, 5, and 6. Cluster 2, grouping departments in the north, seems to be the most compact, with only a few interlopers (from cluster 1). This matches other indications of clustering among the northern departments.

Interacting between the similarity in attribute space as shown in an MDS plot and the geographic locations is one more way in which the tension between these two objectives can be visualized.

Concluding remarks

Tobler's law remains a near-universally respected reflection of the empirical similarity between closeness in geographical and in attribute space. However, when moving to a multivariate context, the trade-offs involved in combining "near" with "related" are complex.

In this article, we illustrated several strategies that can serve as pathways toward discovering "interesting" locations that meet both criteria. This includes a smoothed distance scatter plot to assess distance decay for multiple variables jointly, a new local neighbor match visualization and test to identify multivariate local clusters, and a number of linkages between the attribute space obtained after multidimensional scaling and the geographical locations of observations. Further theoretical and empirical research is needed to gain a better insight into the relative merits of the different approaches.

Acknowledgment

Thanks to the referees and Julia Koschinsky for careful comments on an earlier draft of the paper.

Notes

[1] Other attempts to deal with multivariate spatial autocorrelation have focused on a cross-product statistics such as Moran's I. For example, see Wartenberg (1985), Dray, Saïd, and Débias (2008) and Lin (2019), for approaches that leverage principal components, and Lee (2001, 2017) for a decomposition of a bivariate measure into a correlation and spatial part. See also Anselin (2019) for a more detailed discussion.

[2] The data can be accessed in the R package Guerry, by Michael Friendly and Stéphane Dray at https://CRAN.R-project.org/package=Guerry, or from the GeoDa sample data set collection at https://geodacenter.github.io/data-and-lab/Guerry/ (it is also included as a sample data set in the GeoDa software).

[3] An extension to distance metrics other than Euclidean, such as a Manhattan distance or even a Minkowski metric is straightforward and will not be considered here.

[4] Oden and Sokal (1986) consider genetic distances based on allele frequencies, which in some sense is multivariate, but not in the sense of the multiattribute distance we consider. The similarity between the geographic distance (for different directions) and their dissimilarity matrix is summarized by a normalized Mantel statistic.

[5] Whereas the geographic distance has a direct interpretation in terms of distance units (in our example, meters), this is not the case for the attribute distance matrix. The latter is expressed in standard deviational units for the variable in question.

[6] Our distance scatter plot should not be confused with the distance–distance plot used to detect outliers in multivariate statistics (e.g., Rousseeuw and Van Zomeren 1990).

[7] The use of $k = 6$ is purely illustrative. In an actual empirical exercise, further sensitivity analysis of the selection of the number of neighbors, and, in general, the specification of the spatial weights would need to be pursued (see also Footnote 9).

[8] In practice, as illustrated in Anselin (2019) applying the multivariate local Geary to a small number of principal components may be more effective.

[9] A similar problem is encountered in other local statistics context, such as the local join count statistic in Anselin and Li (2019). In their empirical illustration using some 400,000 observations, *k* was set to 30 in order to obtain meaningful results.

[10] The choice of $k = 6$ is purely for illustrative purposes and the issue of selecting an optimal *k* is ignored in this example.

[11] As in classic clustering methods, the optimal value for *k* should be explored, but this is beyond the current scope.

References

Anselin, L. (1995). "Local Indicators of Spatial Association—LISA." *Geographical Analysis* 27, 93–115.

Anselin, L. (2019). "A Local Indicator of Multivariate Spatial Association: Extending Geary's C." *Geographical Analysis* 51, 133–50.

Anselin, L., and X. Li. (2019). "Operational Local Join Count Statistics for Cluster Detection." *Journal of Geographical Systems* 21, 189–210.

Anselin, L., I. Syabri, and Y. Kho. (2006). "GeoDa, An Introduction to Spatial Data Analysis." *Geographical Analysis* 38, 5–22.

Assunçao, R., M. Neves, G. Camara, and C. Da Costa Freitas. (2006). "Efficient Regionalization Techniques for Socio-Economic Geographical Units Using Minimum Spanning Trees." *International Journal of Geographical Information Science* 20, 797–811.

Bourgault, G., and D. Marcotte. (1991). "Multivariable Variogram and its Application to the Linear Model of Coregionalization." *Mathematical Geology* 23, 899–928.

Chavent, M., V. Kuentz-Simonet, A. Labenne, and J. Saracco. (2018). "ClustGeo: An R Package for Hierarchical Clustering with Spatial Constraints." *Computational Statistics* 33, 1799–822.

Dray, S., and T. Jombart. (2011). "A Revisit of Guerry's Data: Introducing Spatial Constraints in Multivariate Analysis." *The Annals of Applied Statistics* 5, 2278–99.

Dray, S., S. Saïd, and F. Débias. (2008). "Spatial Ordination of Vegetation Data Using a Generalization of Wartenberg's Multivariate Spatial Autocorrelation." *Journal of Vegetation Science* 19, 45–56.

Duque, J. C., R. L. Church, and R. S. Middleton. (2011). "The p-Regions Problem." *Geographical Analysis* 43, 104–26.

Duque, J. C., R. Ramos, and J. Suriñach. (2007). "Supervised Regionalization Methods: A Survey." *International Regional Science Review* 30, 195–220.

Dykes, J., and C. Brunsdon. (2007). "Geographically Weighted Visualization: Interactive Graphics for Scale-Varying Exploratory Analysis." *IEEE Transactions on Visualization and Computer Graphics* 13, 1161–8.

Fisher, R. A. (1935). The Design of Experiments. Edinburgh: Oliver and Boyd.

Friendly, M. (2007). "A.-M. Guerry's Moral Statistics of France: Challenges for Multivariable Spatial Analysis." *Statistical Science* 22, 368–99.

Golledge, R. G., and G. Rushton. (1972). Multidimensional Scaling: Review and Geographical Applications. Association of American Geographers Commission on College Geography, Technical Paper No. 10. Washington, DC: AAG.

Goodchild, M. (2004). "The Validity and Usefulness of Laws in Geographic Information Science and Geography." *Annals of the Association of American Geographers* 94, 300–3.

Guerry, A.-M. (1833). Essai sur la Statistique Morale de la France. Paris, France: Crochard.

Guo, D. (2008). "Regionalization with Dynamically Constrained Agglomerative clustering and Partitioning (REDCAP)." *International Journal of Geographical Information Science* 22, 801–23.

Haining, R. F., S. Wise, and J. Ma. (2000). "Designing and Implementing Software for Spatial Analysis in a GIS Environment." *Journal of Geographical Systems* 2, 257–86.

Journel, A. G., and C. J. Huijbregts. (1978). Mining Geostatistics. London, UK: Academic Press.

Lee, J. A., and M. Verleysen. (2007). Nonlinear Dimensionality Reduction. New York, NY: Springer.

Lee, S.-I. (2001). “Developing a Bivariate Spatial Association Measure: An Integration of Pearson’s R and Moran’s I.” *Journal of Geographical Systems* 3, 369–85.

Lee, S.-I. (2017). “Correlation and Spatial Autocorrelation.” In Encyclopedia of GIS, 2nd ed., 360–8, edited by S. Shekhar, H. Xiona, and X. Zhou. Cham, Switzerland: Springer Nature.

Lin, J. (2019). “A Local Model for Multivariate Analysis: Extending Wartenberg’s Multivariate Spatial Correlation.” *Geographical Analysis*. https://doi.org/10.1111/gean.12196.

Miele, V., F. Picard, and S. Dray. (2014). “Spatially Constrained Clustering of Ecological Networks.” *Methods in Ecology and Evolution* 5, 771–9.

Miller, H. (2004). “Tobler’s First Law and Spatial Analysis.” *Annals of the Association of American Geographers* 94, 284–9.

Murray, A. T., and T. H. Grubesic. (2002). “Identifying non-Hierarchical Spatial Clusters.” *International Journal of Industrial Engineering* 9, 86–95.

Murtagh, F. (1985). “A Survey of Algorithms for Contiguity-Constrained Clustering and Related Problems.” *The Computer Journal* 28, 82–8.

Oden, N. L., and R. R. Sokal. (1986). “Directional Autocorrelation: An Extension of Spatial Correlograms to Two Dimensions.” *Systematic Zoology* 35, 608–17.

Recchia, A. (2010). “Contiguity-Constrained Hierarchical Agglomerative Clustering Using SAS.” *Journal of Statistical Software* 33(2).

Rousseeuw, P. J., and B. C. Van Zomeren. (1990). “Unmasking Multivariate Outliers and Leverage Points.” *Journal of the American Statistical Association* 85, 633–9.

Sui, D. (2004). “Tobler’s First Law of Geography: A Big Idea for a Small World?” *Annals of the Association of American Geographers* 94, 269–77.

Tobler, W. (1970). “A Computer Movie Simulating Urban Growth in the Detroit Region.” *Economic Geography* 46, 234–40.

Tobler, W. (2004). “On the First Law of Geography: A Reply.” *Annals of the Association of American Geographers* 94, 304–10.

Tobler, W., and S. Wineburg. (1971). “A Cappadocian speculation.” *Nature* 231(5297), 39–41.

Wartenberg, D. (1985). “Multivariate Spatial Autocorrelation: A Method for Exploratory Geographical Analysis.” *Geographical Analysis* 17, 263–83.