

**MORAN'S I SPACIAL AUTO-CORRELATION AND
ANOMALY DETECTION UTILIZING *PCA* AND
HIGH DIMENSIONAL FEATURE VECTORS**

by

Roy Y. Wong

A thesis submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Master of Science

in

Computing

School of Computing
The University of Utah

December 2017

Copyright © Roy Y. Wong 2017

All Rights Reserved

The University of Utah Graduate School

STATEMENT OF THESIS APPROVAL

The thesis of Roy Y. Wong
has been approved by the following supervisory committee members:

Jeffrey Phillips , Chair(s) 05 Oct 2017
Date Approved

Feifei Li , Member 05 Oct 2017
Date Approved

Vivek Srikumar , Member 05 Oct 2017
Date Approved

and by Ross T. Whitaker , Chair/Dean of the
Department/College/School of Computing
and by David B. Kieda , Dean of The Graduate School.

ABSTRACT

Anomaly detection in large spatial data sets is difficult. Anomaly detection in large spatial data sets with multiple correlated features, becomes even more difficult. *Moran's I* is a useful function for auto-correlating spatial observations and detecting anomalous observations.

Unfortunately, *Moran's I* has only been developed for single scalar feature comparison. We propose instead to use a vector of features. Now a much more comprehensive data set with feature correlation can be utilized to find outliers based on weighted neighbor values, instead of arbitrary or administrative aggregation.

The new enhancements proposed here allow for richer and nuanced data analysis. With the use of Principal Component Analysis and high dimension feature vectors, regions of interest are less ambiguous to detect. We describe new techniques that reduce feature noise as well as computational and operational complexity. Our techniques are also able to replace other dimensional reduction techniques that introduce distortion or skewing to the feature set.

CONTENTS

ABSTRACT	iii
NOTATION AND SYMBOLS	vi
CHAPTERS	
1. INTRODUCTION	1
1.1 Data Problem	1
1.2 <i>Moran's I</i>	2
1.3 Spatial Opioids Data	4
1.4 Need for Multidimensional Extension	5
2. MULTIDIMENSIONAL MORAN'S I	8
2.1 With <i>PCA</i>	8
2.2 Vector Dot Product Version	9
2.3 Prior Arts in Spatial Data Analysis	10
2.4 Geometric Intuition	10
3. DATA EXPERIMENTS	13
3.1 Data Characteristics	13
3.2 Results <i>PCA</i>	15
3.3 Results Vector Feature <i>Moran's I</i> and <i>local Moran's I</i>	16
3.4 Evaluation and Ramifications	17
4. IMPLEMENTATION CHALLENGES AND OPTIMIZATIONS	27
4.1 Algorithm and Process Enhancements	27
5. CONCLUSION	31
5.1 Future Work	31
5.2 Summary	32
APPENDICES	
A. CITY 1 PARTIAL RESULTS	33
B. CITY 1 ORIGINAL FEATURE VECTOR RESULTS	43
C. CITY 2 FULL RESULTS	45

C. CITY 2 FULL RESULTS	45
D. CITY 2 ORIGINAL FEATURE VECTOR RESULTS	55
REFERENCES	57

NOTATION AND SYMBOLS

I	the Global <i>Moran's I</i> similarity index value
<i>local I</i>	the Local <i>Moran's I</i> similarity index value
N	the number of spatial units indexed by i and j
x	feature variable
\bar{x}	mean value of x
$w_{i,j}$	a matrix of spatial weights with zeros where i and j are the same index value
W	the sum of all $w_{i,j}$
\mathbf{x}	feature vector variable
\mathbf{F}	feature vector weight matrix
$\bar{\mathbf{x}}$	mean value of \mathbf{x}
\mathbf{p}	principal components vector
d	the number of features selected for \mathbf{x}
i	spatial row count
j	spatial column count
c_{ij}	a spatial grid cell containing an aggregate of observations
C	comprised of all cells c_{ij}
t	given time period, in this study a period of months is used
mo_t	the number of opioid prescriptions filled in a given c_{ij} for a given time, t
mn_t	the number of opioid prescriptions received by people who live outside a given c_{ij} for a given time, t

CHAPTER 1

INTRODUCTION

In the last decade, the ability to digitally monitor and track more and more activities has led to larger and larger data sets. As the amount of data collected grows, the need for better aggregation and correlation techniques increases. Any enhancement in even the smallest part of an analytics algorithm could lead to enormous time savings when poring through mountains of amassed data.

1.1 Data Problem

An ever-growing problem is finding anomalies and outliers in spatially autocorrelated data. Particularly when the number of observations and features can reach huge proportions. Once a set of observations with spatial characteristics has been aggregated together, the task of finding anomalies with respect to distance is generally computationally expensive. The ability to weight the value of a feature based on spatial proximity is very valuable. One successful and often used approach to find spatial autocorrelation or feature correlation is with a technique called *Moran's I*.

Moran's I is used to give an index score or value to a given point, or combined set of points, with respect to other neighboring points. If an anomalous feature value isn't characteristic or similar to that of its neighborhood, it is scored with an "I" value higher or lower than its immediate neighbors. Higher if it is a peak or plateau value and lower if it is a valley or trough value.

A set of observations with a global *I* value that is negative, means that all of the observation values differ from each other or only fit in a large range. A positive *I* value means the feature values fit within a relatively small range or change gradually across the spatial domain. The magnitude of the negative value becomes larger the larger the range of values. Conversely, the tighter the range is, the larger the positive *I* global value becomes.

Unfortunately, *Moran's I* is only able to look for autocorrelation for singular value

sets.¹ This poses a problem for data sets whose interdependency can't be described with a single feature. Many spatial data sets are multidimensional in nature and can only be described by multiple associated features.

1.2 Moran's I

*Moran's I*² was developed by Patrick Alfred Pierce Moran in 1950 as a statistical means of calculating spatial observation feature similarity or autocorrelation.³ It is commonly used in geographic and spatial analytics to find outlier or anomalous observations. The degree to which observations differ is based on their similarity and proximity to neighbor-ing observations. The notion of data observation neighborhoods makes it a powerful tool when other clustering and aggregation techniques provide arbitrary or rigid compilations. Or these other procedures may be too computationally expensive to run over large data sets.

Moran's I global calculation is given by the formula below. Where x_i is an observation point or an aggregated grid cell. \bar{x} is the mean scalar feature value for all observation points. $w_{i,j}$ is spatial weight index of neighboring observation points. It describes how much a neighboring feature value should be taken into account. W is the sum of all spatial weights, and N is the total number of observation points or grid cells.

$$I = \frac{N \sum_i \sum_j w_{i,j} (x_i - \bar{x})(x_j - \bar{x})}{W \sum_i (x_i - \bar{x})^2} \quad [1.1]$$

Moran's I has been used by social work researchers, Foster and Hipp,⁴ to go beyond ad-ministrative aggregation concepts like zip code, county, or state boundaries. Criminology researchers, Mencken and Barnett,⁵ have used it in a study of murders in neighboring counties and disprove metric spatial correlation between county regions.

1. "Spatial Autocorrelation and Moran's I in GIS," *GIS Geography*, August 11, 2017, accessed November 14, 2017, <http://gisgeography.com/spatial-autocorrelation-moran-i-gis/>.

2. "Morans I," *Wikipedia*, May 29, 2017, , accessed November 14, 2017, https://en.wikipedia.org/wiki/Moran%27s_I.

3. P. A. P. Moran, "Notes on Continuous Stochastic Phenomena," *Biometrika* 37, no. 1/2 (1950): 17, accessed July 14, 2017, doi:10.2307/2332142.

4. K. A. Foster and J. A. Hipp, "Defining Neighborhood Boundaries for Social Measurement: Advancing Social Work Research," *Social Work Research* 35, no. 1 (2011): 25-35, accessed August 1, 2017, doi:10.1093/swr/35.1.25.

However, *Moran's I* is not limited to simple geographical analytics. It can be applied whenever a relational test is needed for data observations that allow for some distance metric to set them apart spatially. Gene researchers have utilized it to map spatial similarities in genome expressions between generations of plants in order to learn if plant self-pollination leads to genetic expression drift.⁶

Moran uses the spatial weight matrix, w_{ij} , to calculate how much value of a neighbor's observation to take into account. It is often calculated as the squared inverse distance between two observation point. As the distance between two observation points, their associated weight value drops off quickly. A threshold limit can also be applied to the weight value. A spatial weight can be set to zero if the observation point is too far away for consideration. This reinforces the notion of an observation neighborhood.

This well-developed concept of a neighborhood lessens the effect a large outlier can have on other observations if it is spatially distant or "out of the neighborhood". Only when a value is largely different than its "very near" neighbors can it begin to affect their *local I* values and signal an area of interest or boundary condition between pockets of values.⁷

Figure 1.1 shows a small example data set where we compute Moran's I, and its local variant which we define next. This indicates that the grid is generally similar, but has some variation in feature values amongst close neighbors. Unfortunately, having a global indicator only allows for a general impression of the amount of similarity. In 1995, Luc Anselin enhanced the resolution of Moran's formula to allow for individual observation scoring.^{8,9}

5. F. Carson Mencken and Cynthia Barnett, "Murder, Nonnegligent Manslaughter, and Spatial Autocorrelation in Mid-South Counties.," *Journal of Quantitative Criminology* 15, no. 4 (1999): 407-22, accessed August 1, 2017, <http://www.jstor.org/stable/23366750>.

6. Masayuki Maki and Michiko Masuda, "Spatial Autocorrelation of Genotypes in a Gynodioecious Population of *Chionographis japonica* var. *kurohimensis* (Liliaceae)," *International Journal of Plant Science* 154, no. 4 (1993): 467-72, accessed August 1, 2017, doi:10.1086/297130.

7. "Spatial Autocorrelation and Moran's I in GIS," *GIS Geography*, August 11, 2017, accessed November 14, 2017, <http://gisgeography.com/spatial-autocorrelation-moran-i-gis/>.

8. Luc Anselin, "Local Indicators of Spatial Association-LISA," *Geographical Analysis* 27, no. 2 (2010): 93-115, accessed July 18, 2017, doi:10.1111/j.1538-4632.1995.tb00338.x.

$$I_i = \frac{N}{\sum_k (x_k - \bar{x})^2} \sum_j w_{i,j} (x_i - \bar{x})(x_j - \bar{x}) \quad [1.2]$$

Observation scoring does not need to be set over a uniform grid. The observations can be spatial data points that are unevenly distributed. $w_{i,j}$ only relies on a discrete distance metric between observations. These observations can even be distributed in higher order dimensions. This research was developed on a uniform two-dimensional grid, but this can be adapted to other applications with higher dimensional orders and noncartesian distribution patterns.

1.3 Spatial Opioids Data

An applicable data set for this type of spatial analytics is pharmacy controlled substance dispensing rate among pharmacies in a aggregate regions. Opioids are a particularly addictive drug class and prescription data has been collected by month for two large cities for the year of 2016. Being able to find anomalous or pharmacies with an unusually high prescription rate amongst its peer pharmacies would aid in detecting fraud, waste, and abuse of this very abused prescription drug class (see section 3.1 for a more detailed description of the data layout).

The data set also has the added characteristic of obfuscation via aggregation. Legislated privacy concerns dictate that this data can only be reported on in an aggregate fashion and adds a measure of anonymity to the data. Also, the actual city names cannot be divulged. Any possibility of identifying patients via their prescription history must be completely guarded against.

One metric that is often used to see how a pharmacy compares to a cohort group or neighboring pharmacies is the the number controlled substance prescriptions that are filled over a given period of time. Another accompanying metric is how many of these opioid or controlled substance claims are filled for patients who live outside of the immediate operating region of a pharmacy. Then these measures must be considered over multiple periods in order to see if a consistent over utilization trend is present.

9. "Indicators of spatial association," *Wikipedia*, July 28, 2016, accessed August 14, 2017, https://en.wikipedia.org/wiki/Indicators_of_spatial_association.

The effects of the population of patients, the number of competing pharmacies, the number of nearby physicians and hospitals, or even the income and education levels of the patients can be used as partial indicators of a pharmacy's potential for inappropriate opioid prescription filling behavior. Being able to assess this over multiple periods would serve to reinforce the ability to detect of anomalous patterns.

1.4 Need for Multidimensional Extension

The current trend of data sets with an ever growing number of features tied to observation points has been expanding at an unprecedented rate. Our ability to measure more and more metrics real-time has allowed us to get volumes and volumes of correlated data. Unfortunately, as we find more ways to capture these high dimensional feature data sets, we find ourselves faced with the challenge of how to explore and identify anomalous or interesting regions of observation points. While especially keeping the character of the interdependent metrics intact.

The opioid prescription data set is a meaningful example of the need to find pharmacies whose script fill patterns have a high variance compared to their neighboring pharmacies. One could simply use the raw count value as the scalar value and use *local Moran's I* to find the pharmacy which doesn't conform to its neighbors. But this would likely give a terribly incomplete picture without several companion metrics. One metric might be the distance to the nearest hospital or trauma center. Many prescriptions are often filled on the way home from a care facility. Another metric like the number of days supply that are dispensed could temper the number of prescriptions, because only short term prescriptions are being filled as opposed to long term chronic treatments.

As more of these interrelated features are being discovered and used in claims analysis, the need to be able to find reliable autocorrelation has also grown. Unfortunately, an extension to *Moran's I* to handle high-dimensional feature sets has not been thoroughly researched until now. There have been studies using two dimensional feature pairs. The mathematical formulations are similar, but this research used as in trajectory analysis for movements of the observation point and limited to 2D (see section 2.3).

Without a high dimensional solution, much analysis and research is based on an intermediate or intervening function to reduce the feature set to a scalar value that can then be

used in *Moran's I* and *local Moran's I*. These techniques include *Functional Linear Regression (FLR)*,¹⁰ *Singular Value Decomposition (SVD)*,¹¹ and *Principal Component Analysis (PCA)*.¹²

An in-depth exploration of using *PCA* as a technique will be looked at as a starting place to test the potential of using a large high-dimensional featured data set (see section 2.1). Some of the positive attributes of using *PCA* are its ability to reduce noise, minimize distant outliers, and produce a singular value that can be ported into the *Moran's* calculations. Some of the problems associated with *PCA* are the distortion of feature data when only a singular value is taken into account for correlation, the loss of the expression of variance comparison between neighboring observations, and the increasing time-complexity as more and more dimensions are introduced to the feature set.

Using a surrogate function can reduce dimensional complexity. Without this simplification to a single feature, the established version of *Moran's I* is not able find spatial autocorrelation. Unfortunately, these reduction techniques can introduce some fidelity loss or metric distance skewing that masks outliers and make anomalies difficult to find. Because of this *Moran's I* needs to be altered in a novel way that can handle a high dimensional feature vectors.

10. Gareth M. James, Jing Wang, and Ji Zhu, "Functional linear regression that's interpretable," *The Annals of Statistics* 37, no. 5A (2009): 2083-108, accessed July 18, 2017, doi:10.1214/08-aos641.

11. Charles F. Van Loan, "Generalizing the Singular Value Decomposition," *SIAM Journal on Numerical Analysis* 13, no. 1 (1976): : 76-83, accessed July 18, 2017, doi:10.1137/0713009.

12. Karl Pearson, "LIII. On lines and planes of closest fit to systems of points in space," *Philosophical Magazine Series 6* 2, no. 11 (1901): 559-72, Accessed July 18, 2017, doi:10.1080/14786440109462720.

Sample Data					Local Moran's I Results				
1.1	1.2	1.2	1.3	1.4	1.98	1.80	1.42	1.10	0.86
1.2	1.5	2.0	1.3	1.5	0.92	0.69	0.43	0.30	0.20
1.2	1.3	1.8	1.7	1.8	0.07	0.02	0.00	0.03	0.07
1.3	1.4	1.6	1.5	1.6	0.19	0.30	0.50	0.72	0.91
1.4	1.1	1.4	1.5	1.5	0.81	1.04	1.39	1.77	1.93

Global *Moran's I* Value = 0.77

Figure 1.1. A Simple *Moran's I* Example

CHAPTER 2

MULTIDIMENSIONAL MORAN'S I

Our exploration for a robust way to detect anomalies and spatial correlations began with looking at the use of *PCA* to produce a single scalar value that *Moran's I* and *local Moran's I* could use in its autocorrelation calculations. Then we discovered that a vector multiplication extension to *Moran's I* and *local Moran's I* could be harnessed to allow for varying counts of *principal components* to reduce signal noise. This also lead us to test the new multidimensional feature enhancements with the original feature vector to see if it would produce reliable results.

2.1 With *PCA*

One of the first places to begin reducing the dimensionality of a feature space is to use a representative translation function like *SVD* or *PCA*. These methods allow for preprocessing of the original feature vector into an equivalent scalar value before attempting to create *local Moran's I* index scores. For the purposes of this study, *PCA* was used because of its ability to recenter the origin of observation features.¹ It also translates the features into a *principal components* vector, with which any number of the first *principal components* can be used to represent the original feature vector \mathbf{p} (see Figure 2.1: Recentered Component Distribution). Because the unenhanced *Moran's I* and *local Moran's I* can only process a single value, the first *principal component*, $x = \mathbf{p}[1]$, is used for spatial autocorrelation (see Figure 2.1: First Principal Component Vector). Depending on the amount of aggregation, such as grid based clustering versus individual observation point mapping, *PCA* does well at smoothing out wildly unrepresentative data points. Or rather those data points that would skew the *local I* value and hide other more subtle value variations.

1. Hervé Abdi and Lynne J. Williams, "Principal component analysis," *Wiley Interdisciplinary Reviews: Computational Statistics* 2, no. 4 (2010): 433-59, accessed July 19, 2017, doi:10.1002/wics.101.

Some care should be used when selecting the number of features to include in the *PCA* calculations. If too many features are selected or features that have *no real interdependence*, the *local I* values take on a continuous range characteristic that can lead to a lack of differentiation between cells and regions.

2.2 Vector Dot Product Version

The alterations of *Moran's I* and *local Moran's I* to accept are relatively straight forward. The singular feature value is replaced by the observation point's feature vector, $x \leftarrow \mathbf{x}$, of length d . Also the mean feature value is replaced by all the observations' *mean* feature vector, $\bar{x} \leftarrow \bar{\mathbf{x}}$. The mean feature vector, $\bar{\mathbf{x}}$ is subtracted from both the feature vector at i , \mathbf{x}_i , and j , \mathbf{x}_j . Then the scalar value or dot product from the two resultant vectors is used in place of the original scalar equivalents. $(x_i - \bar{x}) \cdot (x_j - \bar{x})$ becomes $(\mathbf{x}_i - \bar{\mathbf{x}}) \cdot (\mathbf{x}_j - \bar{\mathbf{x}})$.

Then we defined the (vector) *Moran's I* as:

$$I = \frac{N \sum_i \sum_j w_{i,j} ((\mathbf{x}_i - \bar{\mathbf{x}}) \cdot (\mathbf{x}_j - \bar{\mathbf{x}}))}{W \sum_i ((\mathbf{x}_i - \bar{\mathbf{x}}) \cdot (\mathbf{x}_i - \bar{\mathbf{x}}))} \quad [2.1]$$

And the (vector) *local Moran's I* as:

$$I_i = \frac{N}{\sum_k (\mathbf{x}_k - \bar{\mathbf{x}})^2} \sum_j w_{i,j} (\mathbf{x}_i - \bar{\mathbf{x}}) (\mathbf{x}_j - \bar{\mathbf{x}}) \quad [2.2]$$

The vector replacement serves several purposes. First it allows for the analytics of a richer feature set without the need for linear regressive preprocessing. It saves a significant amount of computational time complexity. Calculating *PCA* over large data sets has a large computational cost associated with it. Another very useful characteristic is that *Moran's I* recenters the centroid of the observations like *PCA* by subtracting the mean feature vector from each observation vector, $(\mathbf{x}_i - \bar{\mathbf{x}}) \cdot (\mathbf{x}_j - \bar{\mathbf{x}})$. The mean feature vector is defined as the average feature value for each position across all observation feature vectors, $\bar{\mathbf{x}} = [\bar{x}_1, \bar{x}_2, \dots, \bar{x}_d]$. This negates the need to shift the data centroid before introducing it to the *Moran's I* process.

The modifications to *Moran's I* and *local Moran's I* does increase the computational complexity. However, the trade-off is worth it in order remove the entire preprocessing step like *PCA*, which can be quite costly, as the number of observations and features grow. This change was tested with a range of feature counts from $1 \leq d \leq 24$, with only linearly

increasing computational time as d increases. In practice a \mathbf{x} with a large enough d to cause a bottleneck in processing, would likely yield poor analytic results.

When a feature vector with length $d > 2000$ was tested using the dot product technique, the results, while much quicker to compute than with *PCA*, showed little region differentiation and no real areas of interest were detected. This shows that there is a practical upper bound to the number of features that can contribute to the quality of finding regions of interest.

2.3 Prior Arts in Spatial Data Analysis

Prior research into vector based features has been somewhat scarce. Work done by Liu, Tong, and Liu² focuses on looking for geospatial movement patterns using two dimensional vectors that help follow the movements of observation points. Their research focuses on finding the trajectory of an observation point rather than using a vector of features to describe an observation point in high dimension.

While this work is useful in analysis of spatial trajectories in two dimensions, it does not address complex feature correlation. The biggest differences in the feature consideration is the extension to higher order feature dimensionality and the use of features as observation point descriptors rather than a tracking mechanism. This research also looks at the feature vector as an appropriate means of time series analysis.

2.4 Geometric Intuition

The reason a feature vector can readily replace a singular value in *Moran's I* and *local Moran's I* is because the dot product of the feature and mean vectors produces a scalar value that can be multiplied by the weight value, w_{ij} .

$$\begin{aligned} \mathbf{x}_i &= [x_{i,1}, x_{i,2}, \dots, x_{i,d}] \\ \mathbf{x}_j &= [x_{j,1}, x_{j,2}, \dots, x_{j,d}] \\ \langle \mathbf{x}_i, \mathbf{x}_j \rangle &= (x_{i,1}x_{j,1}) + (x_{i,2}, x_{j,2}) + \dots + (x_{i,d}, x_{j,d}) \end{aligned} \tag{2.3}$$

The subtraction of the feature vector mean, $\bar{\mathbf{x}}$, leads to a recentered centroid. Leading

2. Yu Liu, Daoqin Tong, and Xi Liu, "Measuring Spatial Autocorrelation of Vectors," *Geographical Analysis* 47, no. 3 (2014): 300-19, accessed August 24, 2017, doi:10.1111/gean.12069.

to no extra calculations for an origin different than the centroid of the observations feature values. *Moran's I* then looks for spatial autocorrelation based on feature variance and is particularly well-suited for anomaly detection. This is a very desirable characteristic of the vector enhancement. It means that there is no feature distortion introduced that is often apparent with the use of approximation methods.

Other preprocessing approximation methods, like linear regression or *SVD*, have a high probability of skewing or distorting the relationship between all the observations data points. Especially if any of the observation points have wildly different values. Since *Moran's I* considers each point's data in place without interpolation, a single observation can only affect its own index value and those immediately adjacent to it, and not all the observation points in the field.

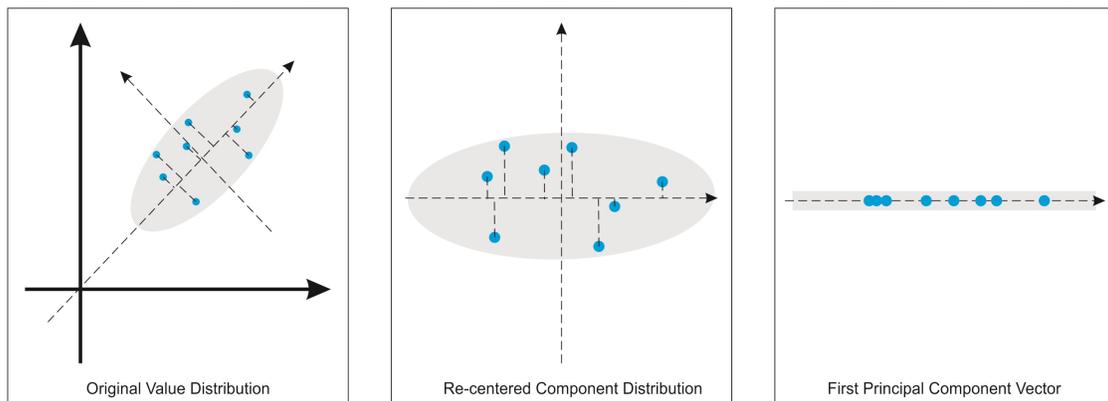


Figure 2.1. *PCA Feature Reduction*

CHAPTER 3

DATA EXPERIMENTS

We were fortunate to obtain the use of a large adjudicated pharmacy claims data set. The data were mined, cleaned, and aggregated for this study. The features selected were intentionally selected because they rely on the same dimensional unit space. Since this approach to high-dimensional testing hasn't been explored, it made sense to keep the domain space reasonably simple.

3.1 Data Characteristics

The data for this study comes from anonymized pharmacy claims data from two large cities¹. The claims dates cover a recent full calendar year. Approximately 23 thousand members, 4 thousand pharmacies, and 240 thousand pharmacy claims are represented in this data set.

The data has been superimposed on a grid whose cells are 2.5 miles (4.02 km) in height and width. Each cell, c_{ij} , (see Figure 3.1) has an aggregated metric of pharmacy opioid²³ claim counts, mo_t , member counts. An additional metric of the claim counts that are filled by at pharmacies with a given cell, but the person receiving the script lives outside the boundaries of the cell, mn_t . This feature is intended to capture the aspect of members who are willing to go to neighboring cells to more easily fill a controlled

1. "Health Insurance Portability and Accountability Act of 1996," H.R. Res. 3103, *104th Cong.*, 1,936 (1996) (enacted), accessed July 1, 2017 and "The Patient Protection and Affordable Care Act.", H.R. Res. 3590, *111th Cong.*, 124 119 (2010) (enacted), accessed July 1, 2017. The source and specifics of the data is obscured because of HIPPA regulation compliance.

2. Takahiro Ogura and Talmage D. Egan, "Opioid Agonists and Antagonists," *Pharmacology and Physiology for Anesthesia*, 2013, accessed July 14, 2017, doi:10.1016/b978-1-4377-1679-5.00015-6.

3. "Opioid Abuse and Addiction: MedlinePlus," *MedlinePlus*, accessed July 15, 2017, <https://medlineplus.gov/opioidabuseandaddiction.html>.

substance prescription.

Metrics for cells are only included in the study if they reported any opioid prescription fills. All other cells are used for spacing and distance calculations, but not applied to the weighted value calculations, $w_{i,j}$. mo_t and mn_t values were tested in a raw count form such that $mo_t \in \mathbb{N}$ and $mn_t \in \mathbb{N}$, and a normalized form such that $\|mo_t\|_2 \in \mathbb{R} \cap [-1, 1]$ and $\|mn_t\|_2 \in -1 \leq \Re \leq 1$.

Each aggregated grid cell, x_{ij} , consists an array of two measurements. The number of opioid prescriptions dispensed by the pharmacies in cell's boundaries and the number of those dispensed to a person who lives outside of the cell's boundaries for each month in 2016. So each cell will consists of an array of 24 values. This data is used as the seed for the *PCA* and vector techniques so that their outputs can be reasonably compared to each other.

These features and data layout allow for a comprehensive search for a large spatial grid. The cell aggregation model was designed this way for two specific reasons. The first is for privacy and legal concerns. Because of the need to identify individual pharmacies in commercial studies, these aggregations will undoubtedly be based on individual pharmacy dispensing counts (see section 3.4 for commercial applications). This will increase $|C|$, but this should prove to be only a linear increase in computational complexity. For the purposes of this study it wasn't necessary to aggregate at that level.

The second reason was for the simplification of neighboring distance computations. This spatial mapping structure allows for a much more simplified neighborhood concept for observations. Using a grid changes the resolution of the distances between observations and is useful if the radius from the center observation to the "edge" of the neighborhood is a relatively large length. If the definition of the neighborhood is somewhat small, then the aggregation would produce inaccurate I values with a large variance.

A very simple weight matrix structure, \mathbf{w} , was chosen to represent the neighboring relationship of each grid cell to every other grid cell. The weight matrix is comprised of an array for each grid cell. Each array position or cell receives a weight value if it is within some maximum threshold distance to the reference cell. In this case, if two grid cells are within 50 miles (80.47 km), then the weight value, $w_{i,j}$, is set to the inverse distance

between the two cells, $w_{i,j} = \frac{1}{\text{distance}(\text{cell1}, \text{cell2})}$.

Another data element that wasn't initially built into the model, but became very necessary, was the materialization and storage of the neighborhood space with respect to each individual cell. Prior to its generation and inclusion, in the process, calculating the neighborhood ad hoc was a large portion of the computation cost that had to be reprocessed over and over for spatial weight calculations (see section 4.1 for more details).

3.2 Results PCA

The first phase of testing was to use *PCA* on the opioid feature set and then use the first principal component value in the *Moran's I* calculations. This created a base line of *local Moran's I* index values that could be compared to any following high dimensional version results. Although the interdependence of features to one another can't be measured in the results, an idea of the regions of interest could be constructed and used as a guide for the quality of the results from the multifeature testing (see Figure 2.1).

For the two dispensing metrics selected for this study, mo_t and mn_t , the unit measures are a count of opioid prescriptions dispensed. However, if one of the features uses a completely different basis of measurement, it is very important to normalize the feature values, $0 \leq \|mo_t\| \leq 1$. This prevents one feature from skewing the total results, because its range of values is wildly larger than another.

The results of *local Moran's I* on the first principal component for City 1 (see Figure 3.2) provide a very good idea of where the anomalous cells are. The highlighted area of interest shows a large variance in opioid dispensing counts. This large variance in conjunction to the proximity of the dispensing diversity increases the *local I* value, making Cell 320 a likely candidate for further investigation. Because its adjacent neighbors had meaningfully fewer dispensed prescriptions and particularly to a larger proportion of people who live outside of its boundaries.

City 2's first principal component results (see Figure 3.3) also does a good job of finding the cell with the largest variance to that of its neighbors. The histogram distribution is similar to City 1. The two cells with the highest *local I* value are the cells with the index value of 26 and 37.

After doing analysis on the *local Moran's I* index values based on the first principal component, it became clear that a new technique would be necessary to see the effect of

using more than one principal component value (see appendix A). A new method was needed to alter *Moran's I* and *local Moran's I* to use feature vectors.

Making these alterations allowed for different multiples of *PCA* values to be tested and studied incrementally (see Figure 3.4). The results show a shift of the mean *local I* value as well as a general narrowing of the bell shape. The anomalous cells still discernable through all levels of principal component selection, but the variance and range fell within much smaller bounds.

Perhaps a more compelling point of interest are the cells with patently negative correlation index value in comparison to their neighboring cells (see Figure 3.5). Here we use the first 15 *principal components* in the vector enhanced *local Moran's I* algorithm. The contrast shown between Cell 320 and 284 their neighbors is quite remarkable. While the neighboring observation cells do show a varying amount of opioids (the *o* values) being dispensed, their calculated index values produce a clearly differentiated region when the color coding gradient is applied. The noise reduction effect of *PCA* highlights two very good candidates cells, which warrant further investigation.

3.3 Results Vector Feature *Moran's I* and *local Moran's I*

Once the *PCA* vector calculations were complete, it became possible to put the original feature vectors in place of the *PCA* calculated data set (see Appendices B and D). The *local Moran's I* index results, for 12 pairs of monthly measures, were nearly identical to the first principal component results (see Figure 3.6). The *local I* variance and histogram distributions only differed slightly by value range⁴ (see Figures 3.6 and 3.7). This shows that results from the first *principal component* and the high dimensional feature vector enhanced *local Moran's I* highlight the same number of interesting regions. With further comparison, the regions identified are the same.

The distribution pattern for the *local I* values were consistent between City 1 and City 2. The differences between *local I* scores between the two cities are accounted for by the number of observation cells. City 1 had significantly more cells for comparison.

In order to verify the quality of the findings, traditional techniques were used to deter-

4. The *PCA* vectors produced by the testing harness are correct in terms of direction and proportion, but need to have scaling corrections made in order to tune the results between the raw FC and *PCA* vectors.

mine which cell, or cells, might be the best candidate(s) for further more labor intensive follow-up. The most likely cell identified using a conventional approach was Cell 192 (see Figure 3.2). It had the highest prescription rate, month over month. It averaged 121.25 opioid scripts over the 12-month period and had a really high number of nonboundary patient script fills.

Cell 192 would have initially seemed like it would have the highest *local I* score. However, because it had very few near-neighbors, it ended up scoring very low. Ultimately, this is because there are very few pharmacies in the immediate surrounding cells. Without any near-neighbors to compare for variance, this cell received a somewhat average *local I* index value.

Because all the neighboring cells lack pharmacies, nearly all of the neighboring cell patients likely go to Cell 192 to fill any kind of prescription. Using a linear regressive or ordinary least squares approach, this cell would most certainly have been flagged for further research. However, the *local I* score doesn't warrant deeper investigation, because it isn't anomalous to its neighbors. Its relative isolation determined by *local Moran's I* essentially removes this false positive from the anomaly candidate list.

Cell 320 (see Figure 3.2) scored the highest *local I* value. It had roughly one-third the number of scripts than that of Cell 190. But because the comparison to its neighbors showed larger variance, the cell received an *local I* value at the top of the scoring range. This makes it much more of a candidate for further fraud, waste, and abuse studies.

The area of interest indicated in City 2 provided another insight that wouldn't have necessarily been found as easily or quickly by using other standard techniques (see Figure 3.3). Cell 37 had the highest prescription rate on average all the cells. However, Cell 26 scored the highest *local I* value, not because it had the most prescriptions, but because its prescription counts had a large margin of difference compared to its neighbors. The ability of this technique allows for a much more nuanced analysis.

3.4 Evaluation and Ramifications

Feature vector multiplication has several advantages over the two stage linear regressive techniques. First it reduces the number of data processing steps. It reduces the computational complexity and allows for very high dimensional feature sets. Assuming

any necessary normalization steps are taken, it introduces no observation measurement distortion or skewing.

There are a few reasons that *PCA* preprocessing would be a better choice over using the original feature set in the vector enhanced *local Moran's I*. If the full number of features were orders of magnitude larger than that of the *n-principal components*, or resources like Random Access Memory or disk space were somewhat limited, or if one felt that omitting the less significant principal components could reduce noise or false positives, then an argument in favor of *PCA's* use could be made. Using *PCA* in conjunction with *Moran's I* is a new technique and is an especially intriguing concept when coupled with vector *local Moran's I*.

The feature vector calculation enhancement has proven to be very effective in highlighting regions of interest. Especially considering the ability to layer complimentary feature values is key to building accurate and complex models. It also gives the ability to do meaningful time-series comparison without having to explicitly encode the concept of time into the feature set.

Simply adding the dimension of time as spatial coordinate or feature extends the possible research arena to spatial and chronological anomaly detection. Behavior research both human and natural could be greatly expanded. A common research problem is that of seasonal drug dispensing and consumption. Being able to track multiple features like location, quantities, dates dispensed, and days supply could lead to outlier detection for different disease states and possibly point to the source of epidemics.

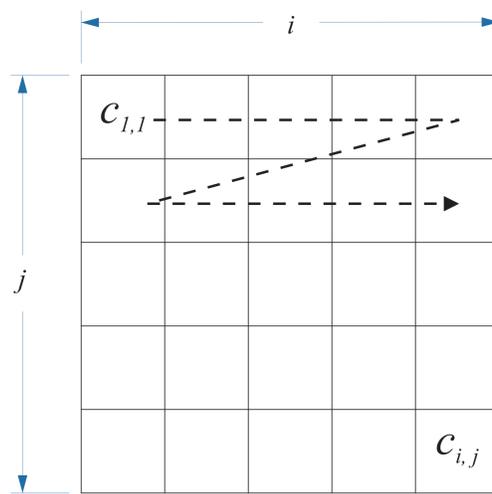
The research domain in this study centers around opioid fraud waste and abuse. It is particularly topical and is considered a top agenda item for many healthcare and government organizations.⁵ Particularly a past research study done using metric clustering and pharmacy opioid dispensing outlier detection is being used at Anthem, Inc., a leading healthcare provider in the United States, to detect prescription abuse.⁶

This study will enhance the ability to detect more nuanced patterns. It will be used in

5. James Oliphant, "Trump declares national emergency on opioid abuse," *Reuters*, August 10, 2017, accessed August 15, 2017, <https://www.reuters.com/article/us-usa-trump-opioid/trump-declares-national-emergency-on-opioid-abuse-idUSKBN1AQ2AW>.

6. "Anthem Blue Cross and Blue Shield Program Tackles Inappropriate Opioid and Rx Drug Use," *Anthem, Inc.*, May 25, 2016, accessed May 30, 2016, <https://www.anthem.com/press/newhampshire/anthem-blue-cross-and-blue-shield-program-tackles-inappropriate-opioid-and-rx-drug-use/>.

many different areas. From the perspective of fraud, waste, and abuse, it will speed up the search for pharmacies and doctors who aren't using best practices for the use of controlled substances. It can be used to find hospitals and care facilities whose performance metrics as a whole provide better care than their cohorts. Complex measures for patient care can be looked at as a whole, rather than in pieces.



$$c_{i,j} = [m_{0i,j}, m_{n_{i,j}}]$$

$$\mathbf{c}_{ij} = [c_{1,1}, c_{1,2}, \dots, c_{2,1}, c_{2,2}, \dots, c_{i,j}]$$

Figure 3.1. Spatial Grid Data Layout

20170729 LOCALMORANS City 1 - FC 1PCA - ANNUAL - RAW

Feature Legend

i: Grid cell index
 m: Local Moran's i value
 o: 2016 Average opioid prescription count
 n: 2016 Average opioid non-member prescription count
 s: 2016 Average member count
 p: First PCA value
 The black bar at the bottom of the cell indicates the normalized population count.

Local Moran's i Color Legend



Values Histogram



City Features:

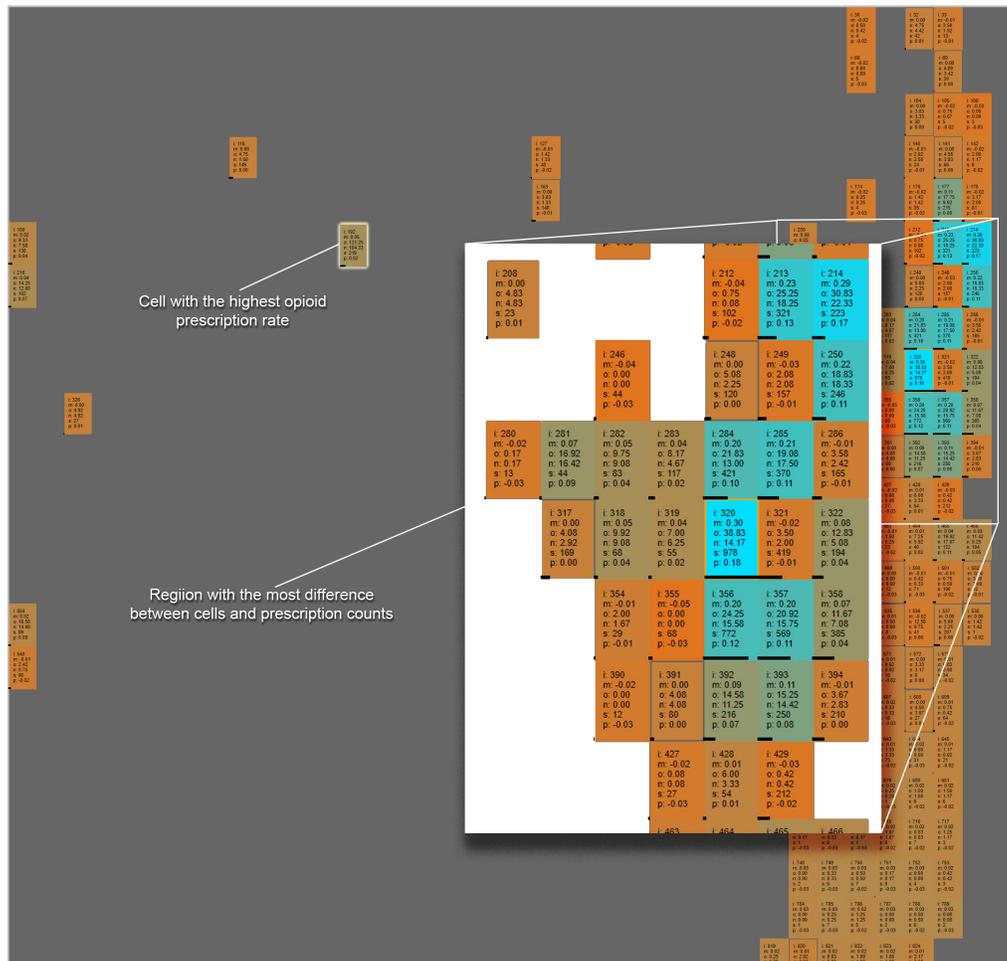


Figure 3.2. City 1: First Principle Component Results

20170729 LOCALMORANS City 2 - FC 1PCA - ANNUAL - RAW

Feature Legend

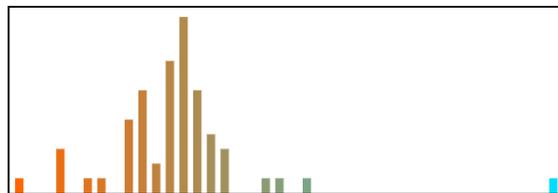
i: Grid cell index
 m: Local moran's I value
 o: 2016 Average opioid prescription count
 n: 2016 Average opioid non-member prescription count
 s: 2016 Average member count
 p: First PCA value

The black bar at the bottom of the cell indicates the normalized population count

Local Moran's I Color Legend



Values Histogram



City Features:

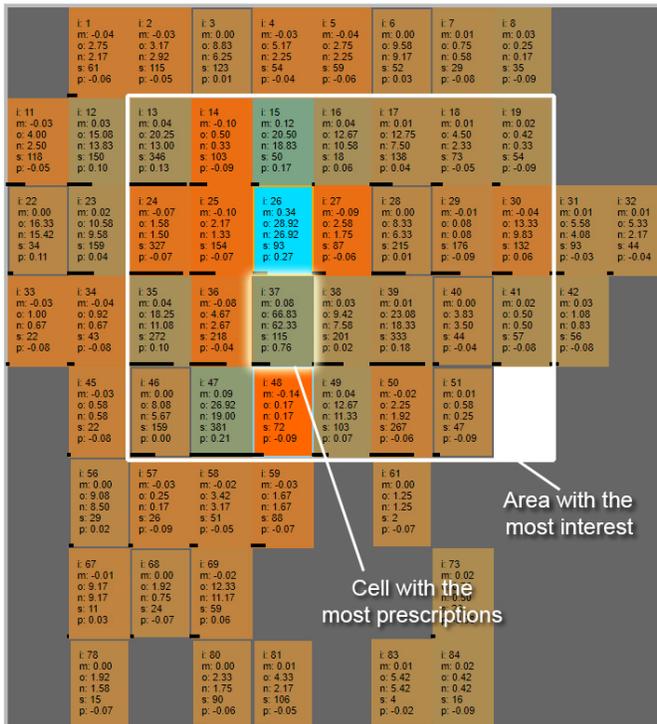


Figure 3.3. City 2: First Principle Component Results

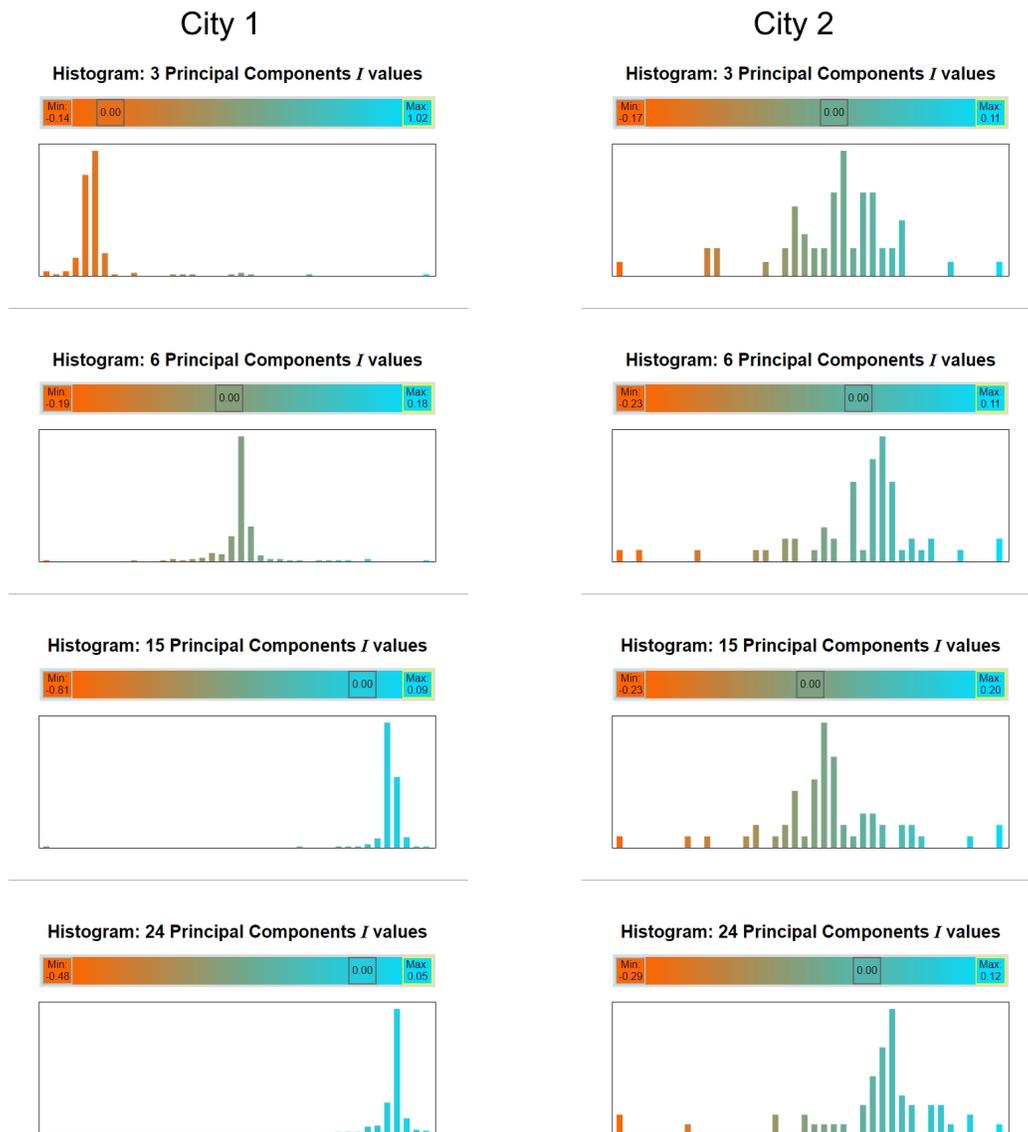


Figure 3.4. City 1 and 2: Color Legend and Histograms

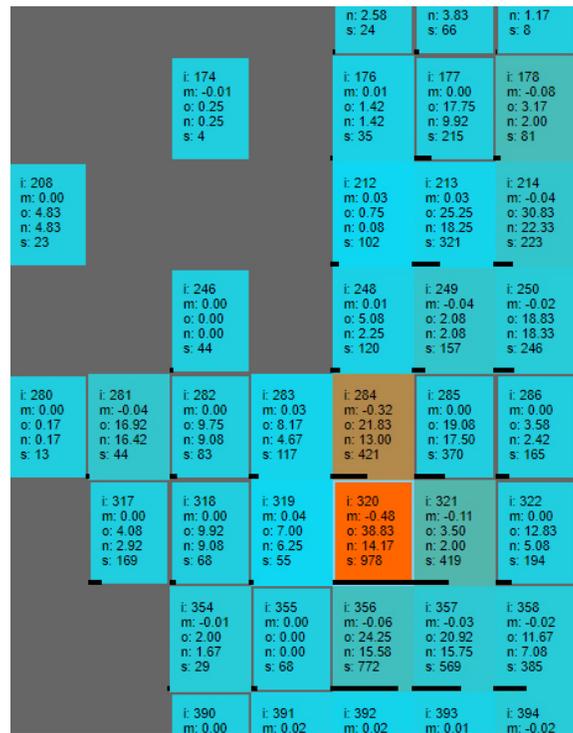


Figure 3.5. City 1: 15 Principal Components Negative Correlation



Figure 3.6. City 1 and 2: Full Feature Vector vs. 1 PCA Histograms

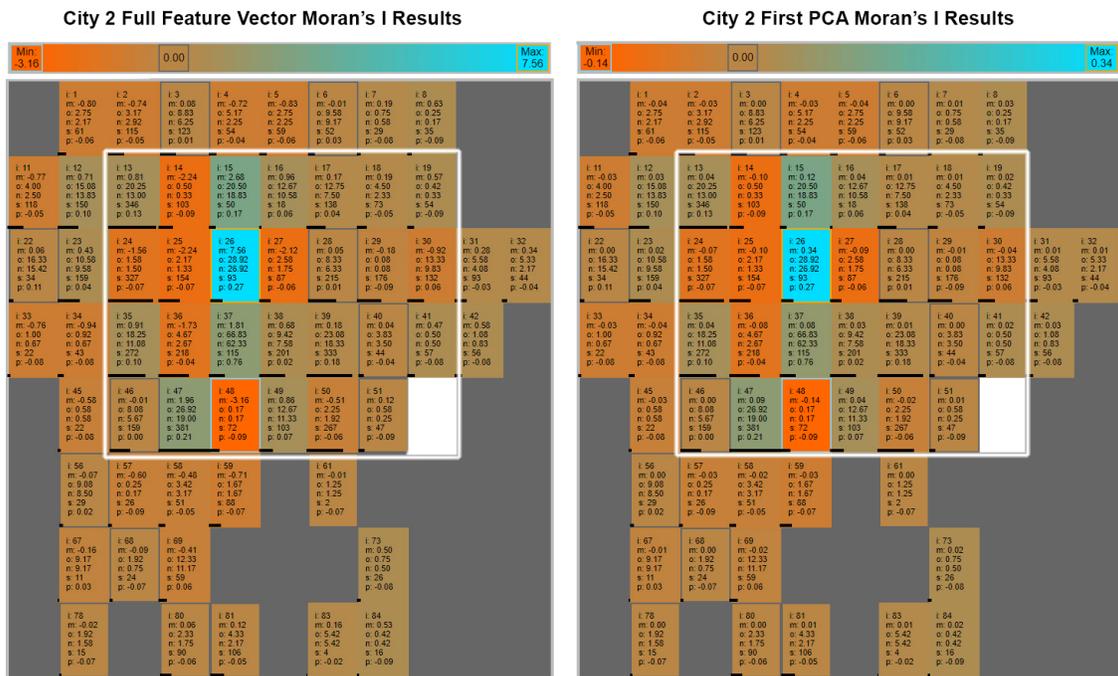


Figure 3.7. City 2: Full Feature Vector vs. 1 PCA Results

CHAPTER 4

IMPLEMENTATION CHALLENGES AND OPTIMIZATIONS

Perhaps the largest challenge of these experiments was having to produce a testing harness capable of implementing a fair amount of testing variations. For example, the ability to test n number of principal component values took time and care to ensure that each step of the process reliably and consistently test each value of n . The complexity of test was compounded by the need to test normalized and raw data counts. Then after all the initial results were analyzed and we thought we had completed all the testing, another variation for a "greater-than-zero" filter had to be added in to the variation mix. All these permutations had to be added without compromising the techniques that produced the earlier results.

4.1 Algorithm and Process Enhancements

The code base for this research is a translation from a Java implementation¹ of *Moran's I* to a C#. It was designed to match the output of the commonly used ArcGIS platform.² The C# version had fewer external assembly dependencies such that it could be modified more quickly and interface with a larger number of data sources of varying types.

The best way to list the process enhancements is the chronological order in which they were implemented. The first problem to address was the computational load created by processing each pharmacy as its own observation point. The number of pharmacies and their corresponding neighborhoods would have produced a jagged neighborhood matrix and inverse spatial weight matrix.

1. Fangming Du, "GitHub/MoransIJava," 2013, https://github.com/Fangmingdu/MoransI_Java.

2. ArcGIS Resources, "How Spatial Autocorrelation (Global Moran's I) works," 2013, http://resources.arcgis.com/en/help/main/10.1/index.html#/How_Spatial_Autocorrelation_Global_Moran_s_I_works/005p0000000t000000/.

The first optimization was to aggregate the pharmacies to grid cells. This significantly sped up the *local Moran's I* searches for appropriate neighbors and their spatial weights. This also allowed for some optimization of search steps. Because we know that the neighborhood matrices are regular and ordered, we allowed the algorithm to break out of searches at known regular points when expected neighbors or weights would have already been found if they existed.

The next problem that consumed a large amount of processing resources, was the calculation of pharmacy neighborhoods. Especially when large data sets were used for testing. It was quite costly to recalculate which cells were neighbors and which were not. It could have been done with a function that calculated neighborhoods ad hoc. After seeing that the neighborhoods would be needed repeatedly, and were of a determinate nature, a cached neighborhood set proved to be much more effective. Turning the neighborhood mappings into a materialized data read in at run time made the *local I* score calculations much faster. It also provided a lot of insight into the way *local Moran's I* calculations shifted through the grid data set.

The inverse spatial weight matrix was *not* built into a separate data set, though. Its calculations were built into a discrete function and calculated once per run. It contains a structure similar to the neighborhood matrix. Instead of simple yes/no indication of adjacency, a $\frac{1}{distance^2}$ where the distance is less than some threshold. Once the neighbor's weights were calculated, they were moved to an in memory hash table. Then a binary search was substituted for an iterative search pattern. This cut the run times significantly.

The original inverse spatial weights were also using a grid unit when calculating distances. The new code was altered to calculate geographic distances instead. This was necessary to align and aggregate the pharmacies to grid cells. It will also become important when the code is used on the distances between individual pharmacies. Geographic distances will be the best way to create accurate *local I* values.

Once the changes to the *Moran's I* and *local Moran's I* code to allow feature vectors instead of scalar values was complete, 1 to 24 elements of the *principal components* were run. Then the initial source data for the *PCA* calculations were also used as the source data for the *Moran's I* calculation step. The results of the full feature vector compared to the first principal component were remarkably similar. They essentially

varied by range and scale, but the regions of interest, trough, and valley scores matched up particularly well.

The visualization of the code was developed using C#, html, javascript, and D3.³ It showed large regions of uninteresting cells with no prescription counts at all. This was due to a lack of pharmacies aggregated to them. It became clear that a lack of data in these areas was making it hard to find areas of interest, even after color gradation and border differentiation were applied. This was the impetus for the last optimization of removing cells with *zero* prescriptions filled. The resulting *local I* scores took on a much more expected bell-shaped distribution pattern (see Figure 4.1).

The application of the "greater-than-zero" filter make the areas with prescription variance much more apparent. This simple screening measure of script count, could easily be replaced by a more sophisticated threshold function that looks at any number of values to determine the potential inclusion of a cell or observation point in the calculation set.

3. Mike Bostock, "Data-Driven Documents," D3.js, , accessed January 5, 2017, <https://d3js.org/>.

City 1 I Distribution

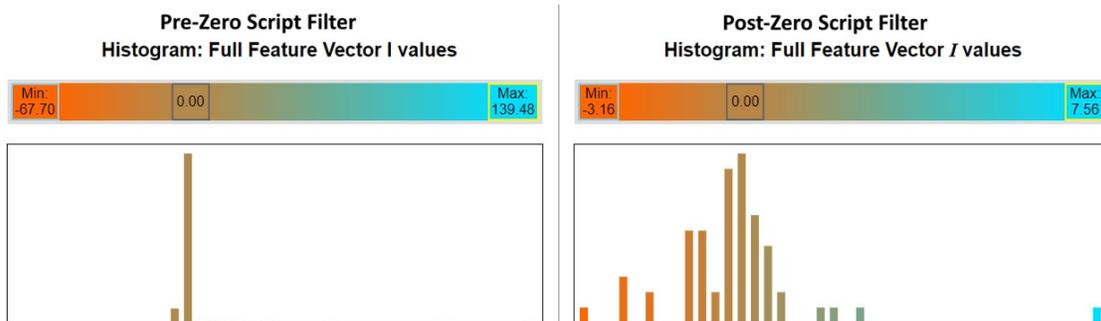


Figure 4.1. City 1: Pre- vs. Postzero Filter Distribution

CHAPTER 5

CONCLUSION

There is still a fair amount of testing that needs to be done to fully explore the "effective range" of these techniques. The instances of where this work enhances or hinders the ability to search for spatial autocorrelation still need rigorous bounds testing to find the best use cases for the application of the these techniques.

5.1 Future Work

The next refinement to test is the addition of a feature weight matrix, \mathbb{F} , that can be multiplied by the feature vector such that different features could be considered more or less important than others. The weight vector could be used by an domain expert to tune the feature values in such a manner that all features could be considered, but only in the proportion that they are helpful to the over all view of the data.

$$I_i = \frac{(\mathbf{x}_i - \bar{\mathbf{x}})}{m^2} \sum_j w_{i,j} ((\mathbb{F}\mathbf{x}_j) - \bar{\mathbf{x}}) \quad \text{where: } \bar{\mathbf{x}} = \frac{\sum \mathbb{F}\mathbf{x}_i}{N} \quad [5.1]$$

A good example would be the number of distinct count out of area prescribers. If a region fills a lot of scripts by out of region doctors, then it would reflect a fair amount of travel just to fill a prescription. However, this value is not nearly as important as the overall opioid prescription count. This would then add the need to say that particular feature is only proportionally as valuable a metric.

Also a more robust test of interest region detection would be the use of a synthetic data set that has randomized noise variation throughout the field. Seeded cells or group of cells with a larger value threshold of variance could be used to simulate regions of interest. Then the probability of detection can be calculated given a threshold range for testing over a number of samples. Essentially allowing a deeper inspection of the lower and upper bounds of reliable variance detection.

5.2 Summary

Modern data analysis is constantly be pushed to understand more and more connected measures. The growing number of observations are taking their toll on prevailing clustering and administrative aggregation techniques. *Moran's I* has a wonderfully elegant way of creating region based autocorrelation assertions. Its strength of finding locality based variance hard to beat. The ability to successfully identify exceptional or suspicious observations points or regions has a wide variety of applications. Far beyond simple map analysis.

Unfortunately, current *Moran's I* and *local Moran's I* calculation techniques are somewhat limited. They are often restrained to a simple scalar metric, which is probably not enough information to understand trend or complex relational undercurrents. A proxy value is sometimes substituted via hash or regression function, but this is often expensive in terms of computational cost, and can warp the relationship between features and observation points.

Some exploration in two-dimensional (2D) feature vectors has been made, but it really only examined observation trajectory rather than high dimensional interrelated feature vectors. There is growing trend of large spatial feature set collection. Without the extensions and optimizations demonstrated here, these types of data sets are hard to reduce and quantify into a unifying index value.

This new combination of *PCA* and *local Moran's I* shows that in combination they can be effective at reducing noise and computational resources need to find autocorrelation. The high dimensional vector extensions have proven that unusual regions and observation points can be calculated and highlighted in a manner that is linearly efficient with respect to the feature count. This research proves that nuanced associated features can lead to very sophisticated results that can reduce noise and false identification of regions of interest using *PCA* noise reduction. Moreover, it has the ability to reduce operational and administrative complexity from the typical analytic processes.

APPENDIX A

CITY 1 PARTIAL RESULTS

20170729 LOCALMORANS City 1 - FC 1PCA - ANNUAL - RAW

City Features:

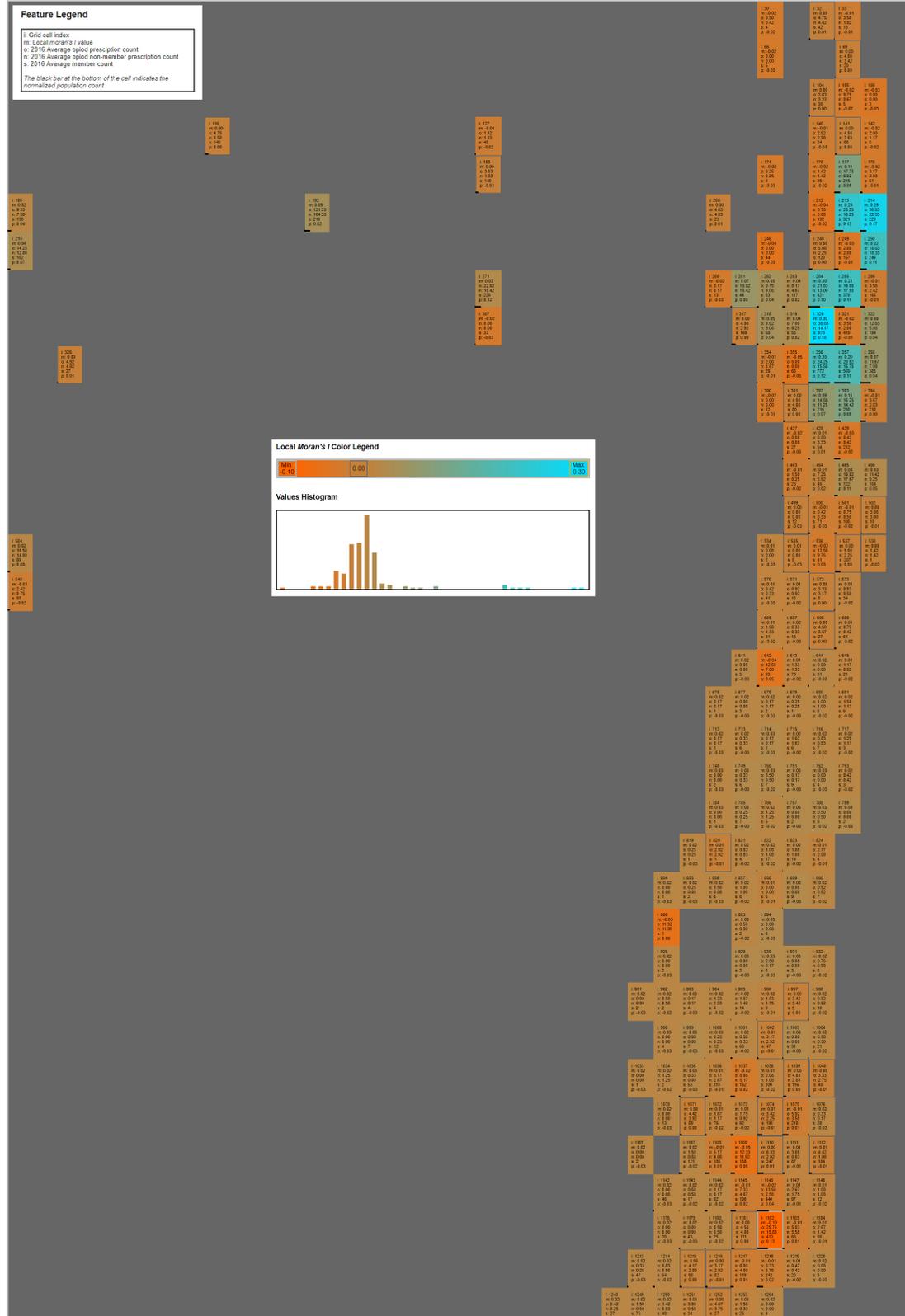


Figure A.1

20170729 LOCALMORANS City 1 - FC 3PCA - ANNUAL - RAW

City Features:

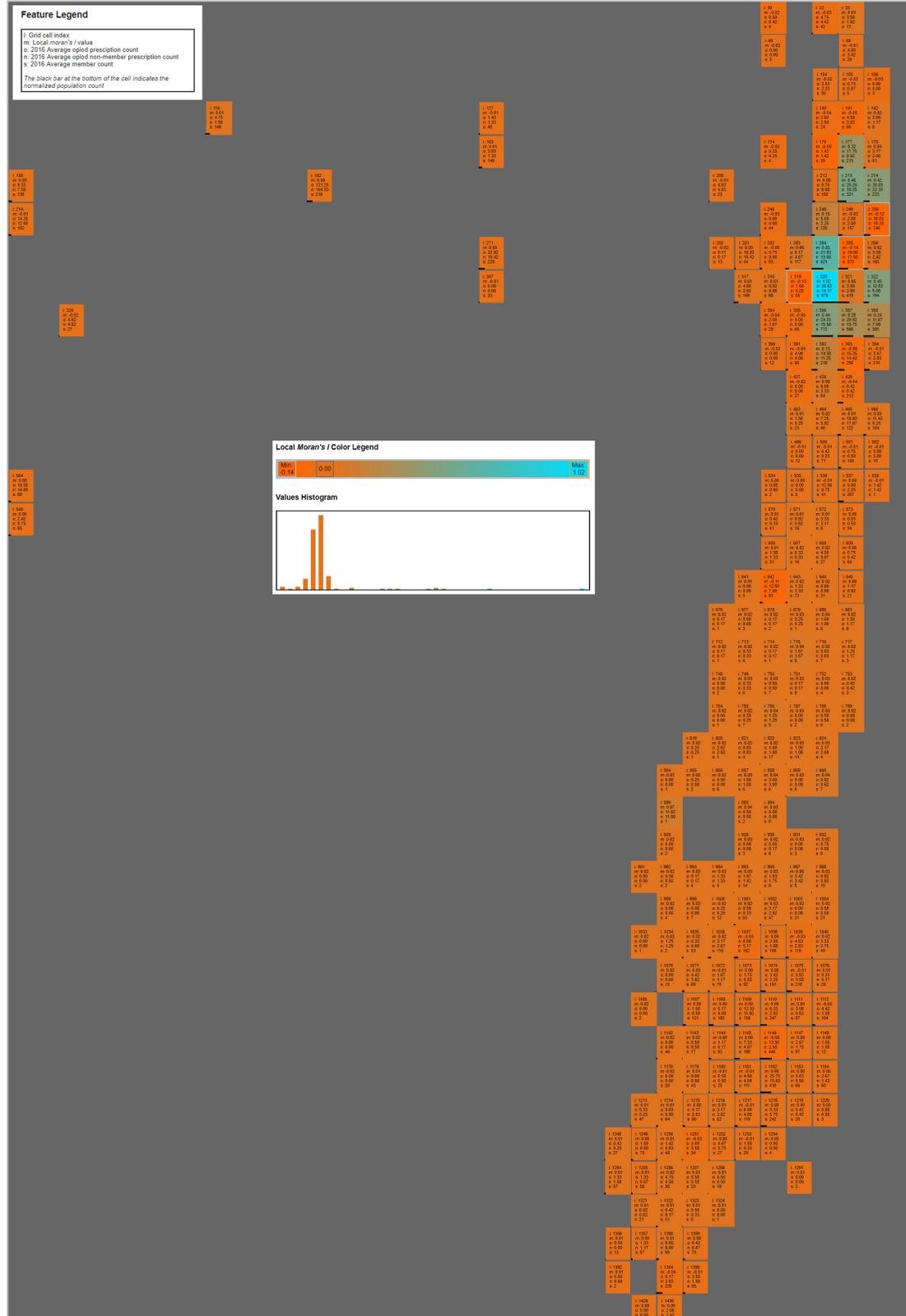


Figure A.2

20170729 LOCALMORANS City 1 - FC 6PCA - ANNUAL - RAW

City Features:

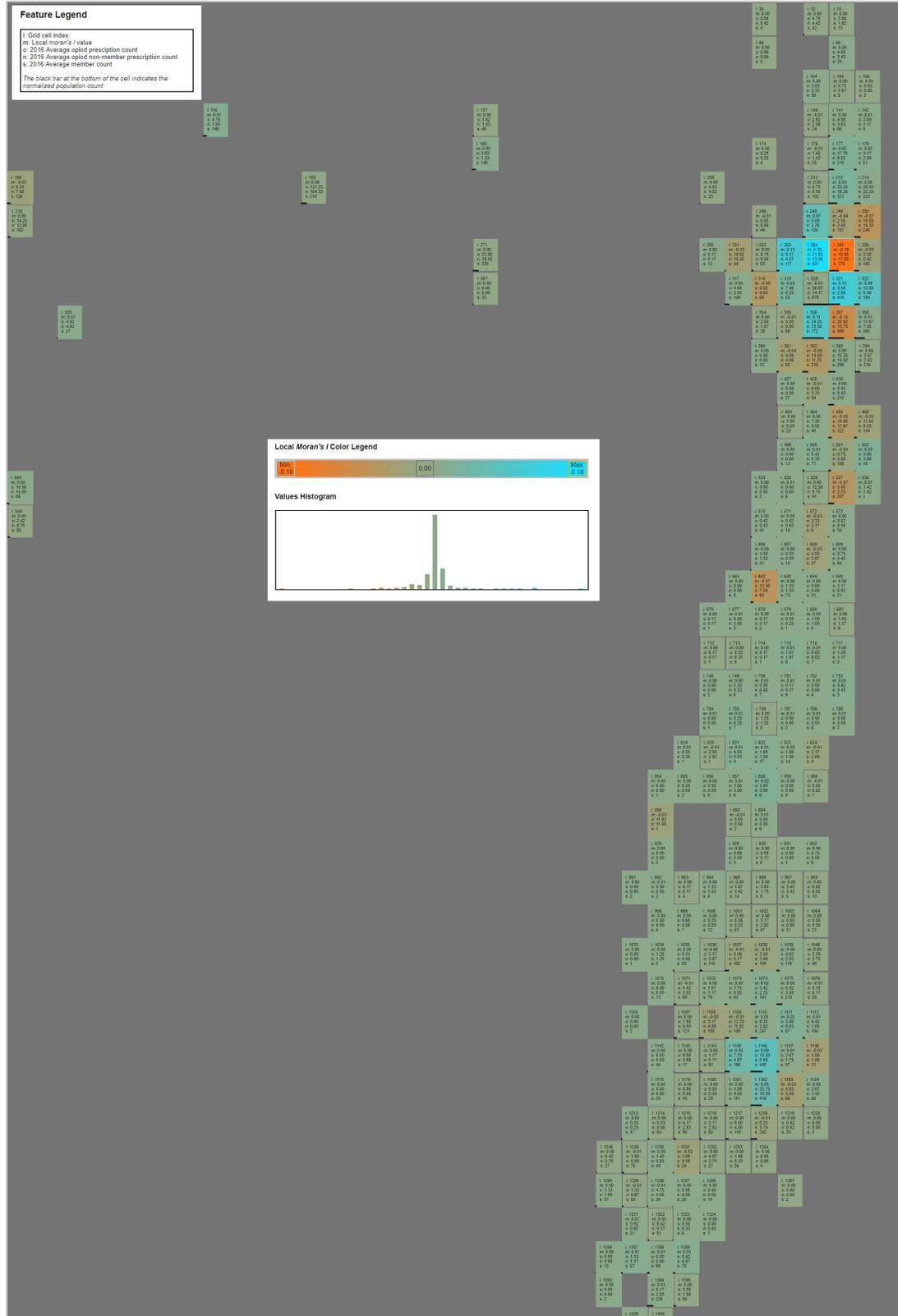


Figure A.3

20170729 LOCALMORANS City 1 - FC 9PCA - ANNUAL - RAW

City Features:

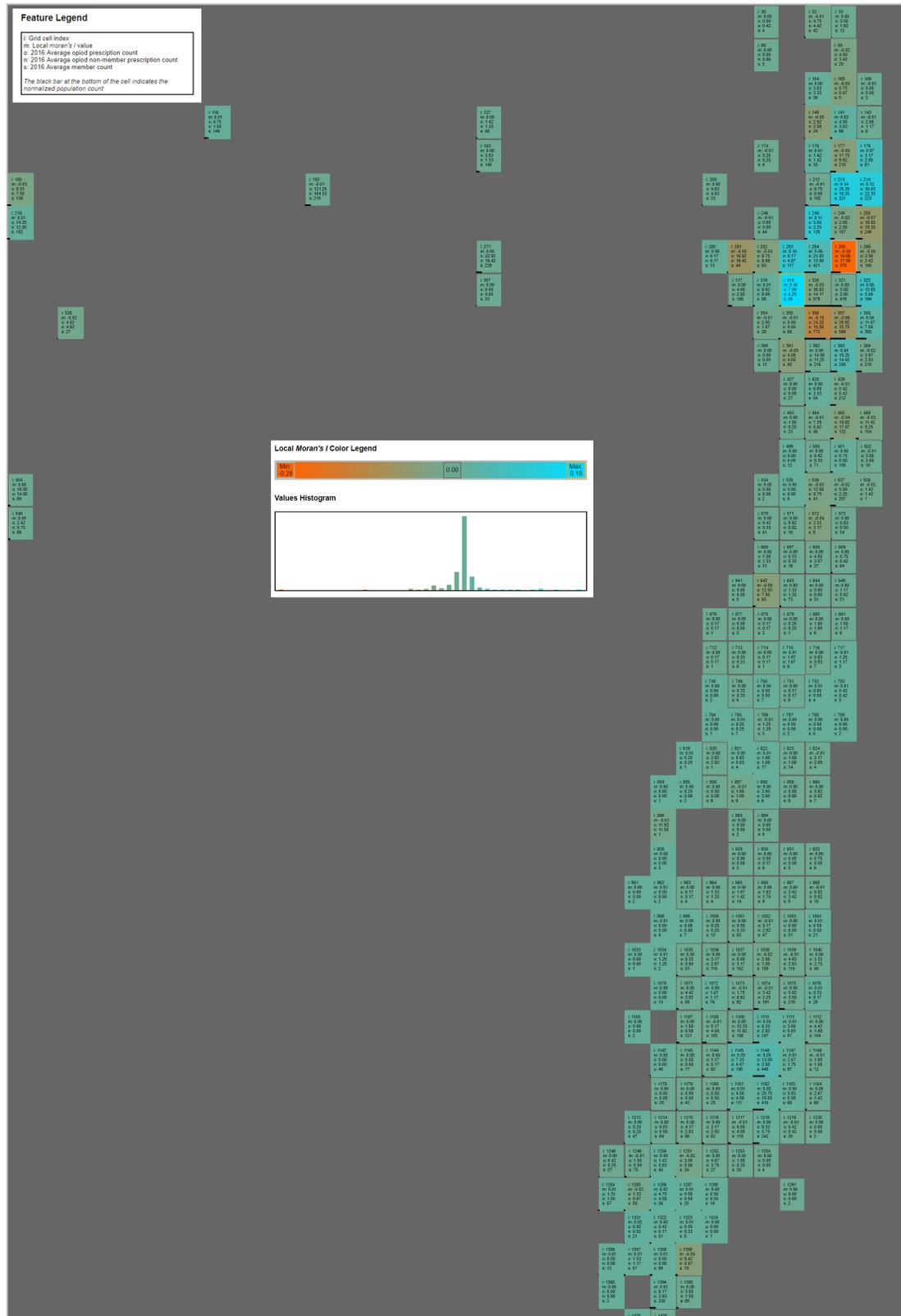


Figure A.4

20170729 LOCALMORANS City 1 - FC 12PCA - ANNUAL - RAW

City Features:

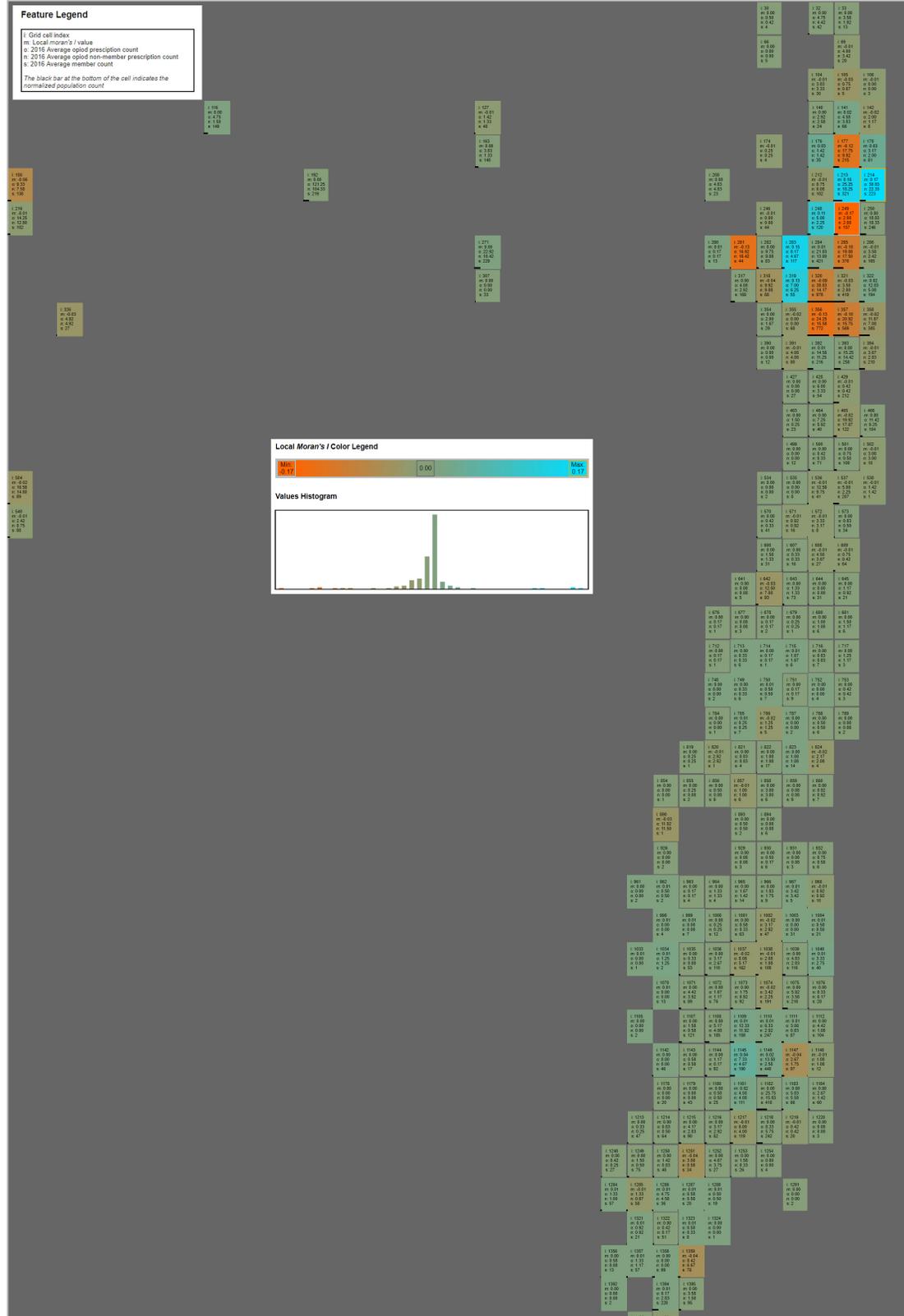


Figure A.5

20170729 LOCALMORANS City 1 - FC 15PCA - ANNUAL - RAW

City Features:

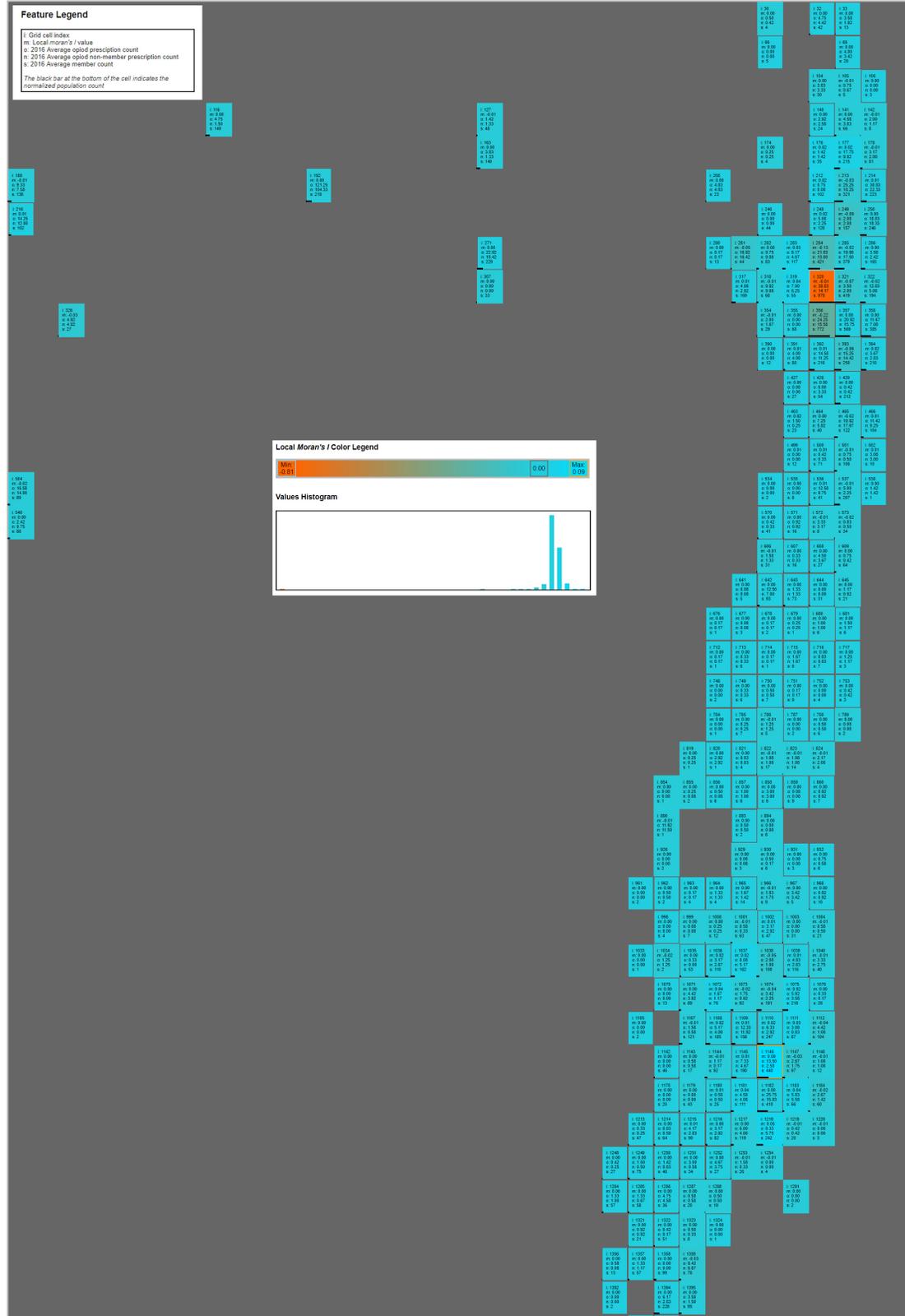


Figure A.6

20170729 LOCALMORANS City 1 - FC 18PCA - ANNUAL - RAW

City Features:

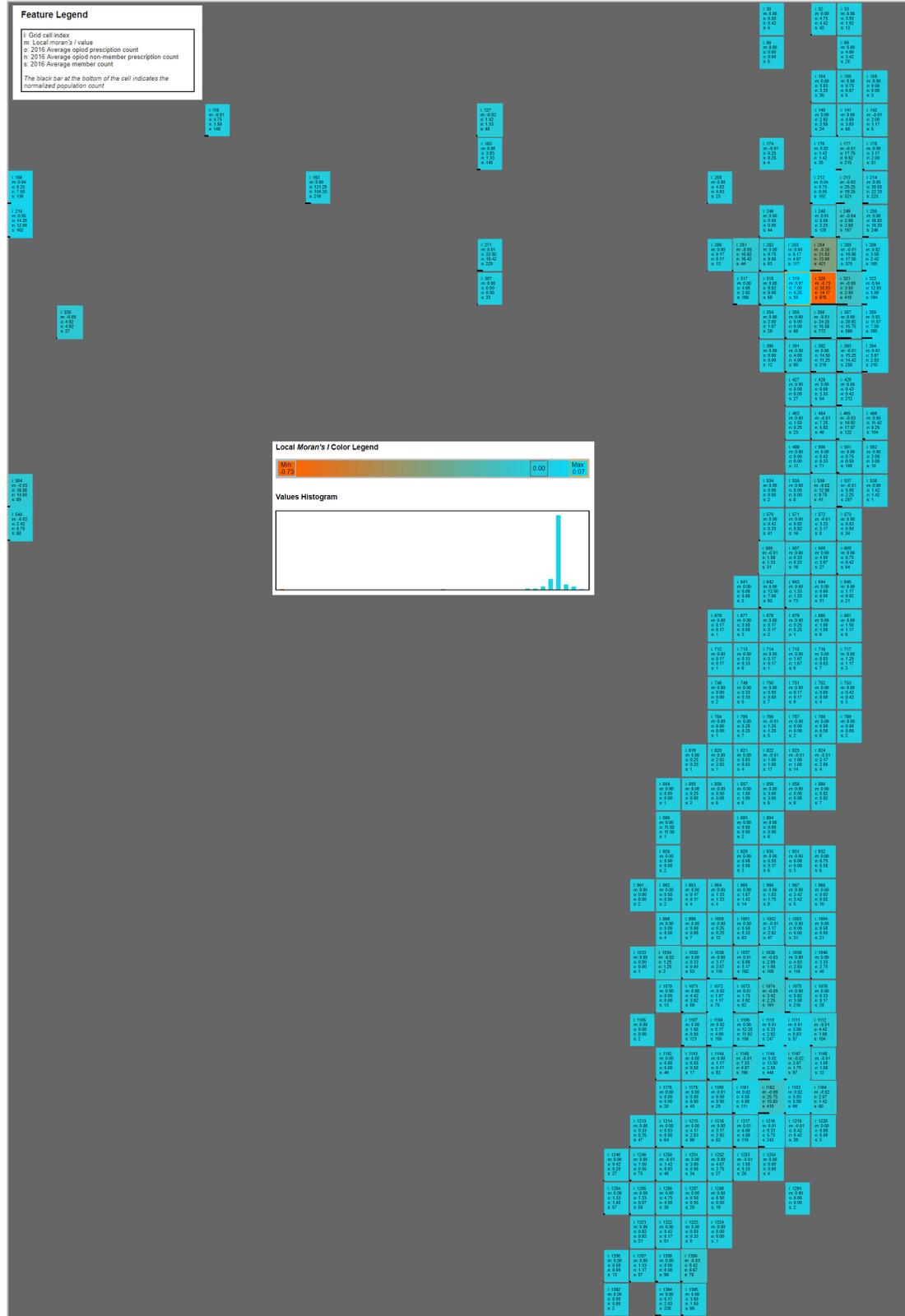


Figure A.7

20170729 LOCALMORANS City 1 - FC 21PCA - ANNUAL - RAW

City Features:

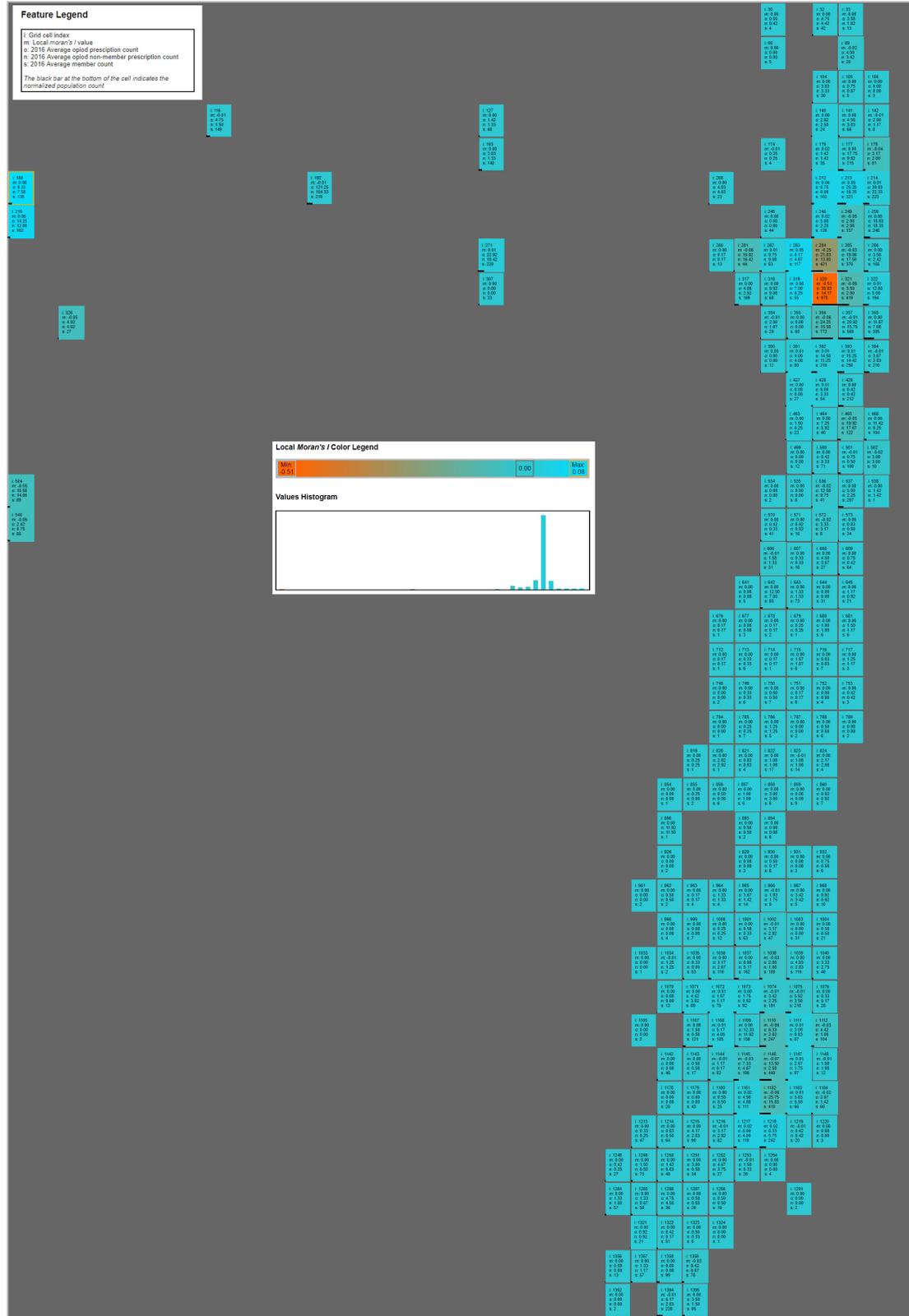


Figure A.8

20170729 LOCALMORANS City 1 - FC 24PCA - ANNUAL - RAW

City Features:

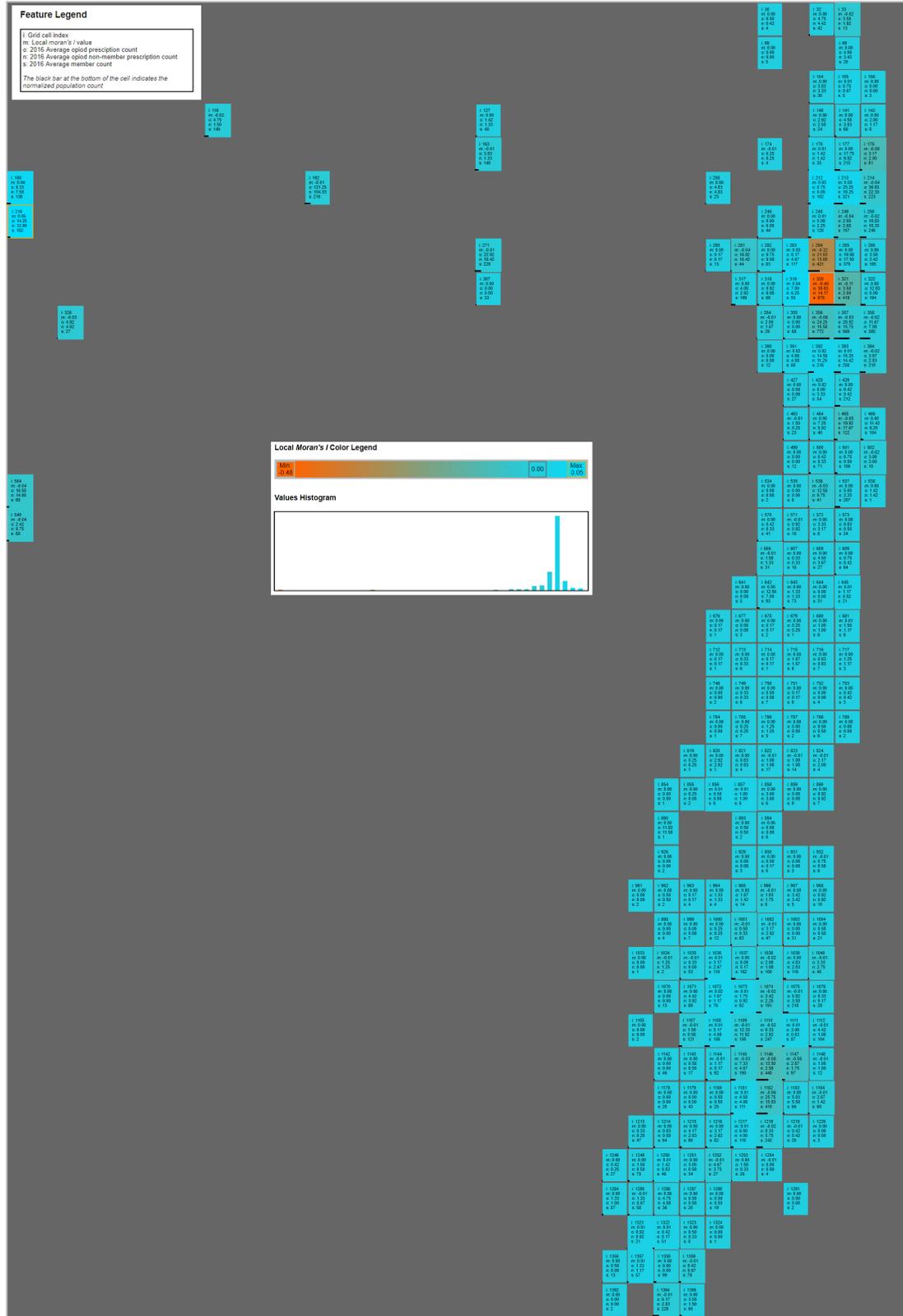


Figure A.9

APPENDIX B

CITY 1 ORIGINAL FEATURE VECTOR RESULTS

20170729 LOCALMORANS City 1 - FC VECTOR - ANNUAL
City Features:

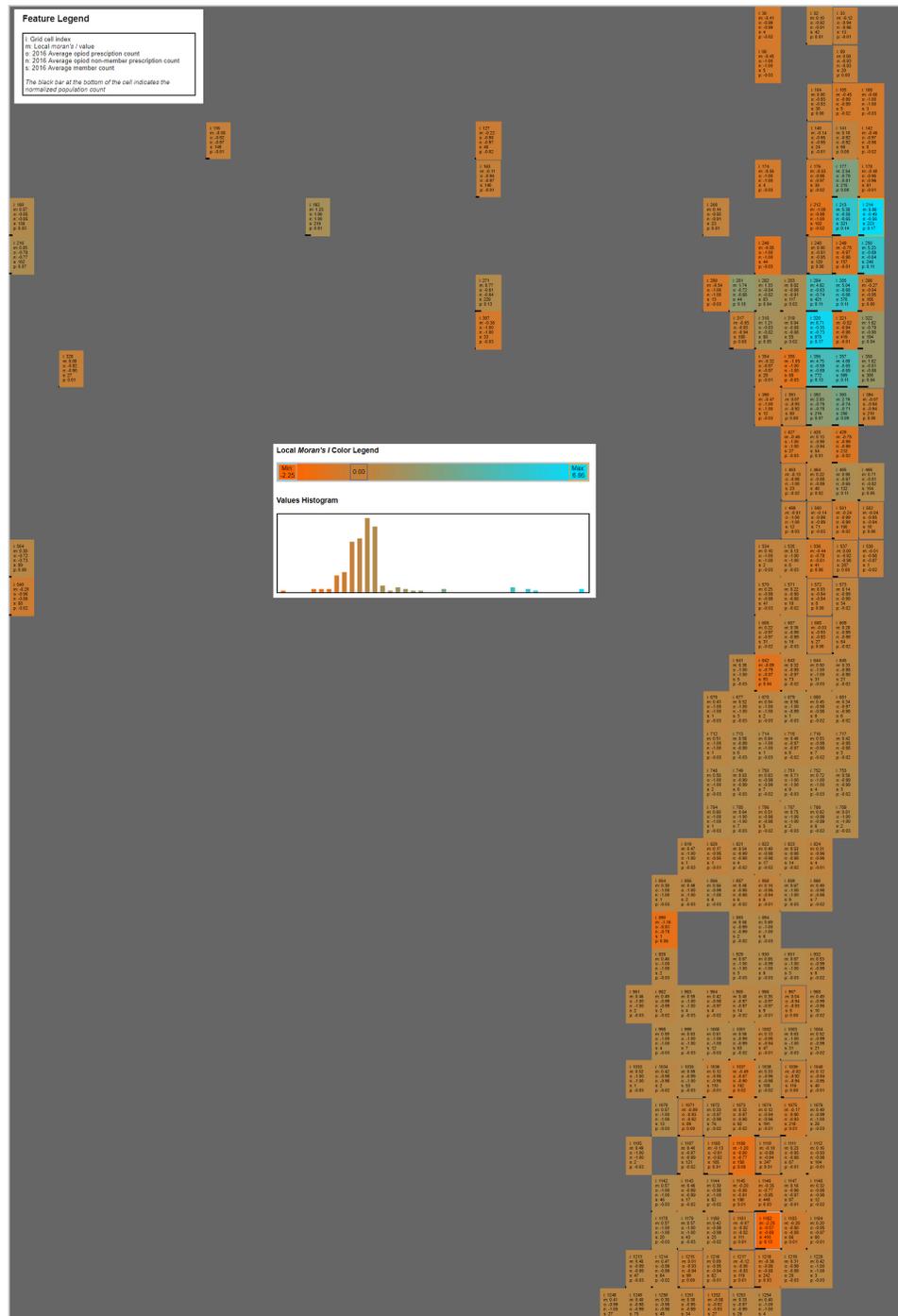


Figure B.1

APPENDIX C

CITY 2 FULL RESULTS

20170729 LOCALMORANS City 2 - FC 1PCA - ANNUAL - RAW

Feature Legend

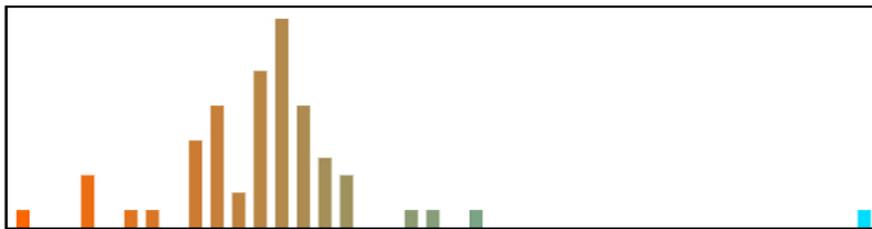
i: Grid cell index
 m: Local moran's I value
 o: 2016 Average opioid prescription count
 n: 2016 Average opioid non-member prescription count
 s: 2016 Average member count
 p: First PCA value

The black bar at the bottom of the cell indicates the normalized population count

Local Moran's I Color Legend



Values Histogram



City Features:

	i: 1 m: -0.04 o: 2.75 n: 2.17 s: 61 p: -0.06	i: 2 m: -0.03 o: 3.17 n: 2.92 s: 115 p: -0.05	i: 3 m: 0.00 o: 8.83 n: 6.25 s: 123 p: 0.01	i: 4 m: -0.03 o: 5.17 n: 2.25 s: 54 p: -0.04	i: 5 m: -0.04 o: 2.75 n: 2.25 s: 59 p: -0.06	i: 6 m: 0.00 o: 9.58 n: 9.17 s: 52 p: 0.03	i: 7 m: 0.01 o: 0.75 n: 0.58 s: 29 p: -0.08	i: 8 m: 0.03 o: 0.25 n: 0.17 s: 35 p: -0.09		
i: 11 m: -0.03 o: 4.00 n: 2.50 s: 116 p: -0.05	i: 12 m: 0.03 o: 15.08 n: 13.83 s: 150 p: 0.10	i: 13 m: 0.04 o: 20.25 n: 13.00 s: 346 p: 0.13	i: 14 m: 0.50 o: 0.33 n: 103 p: 0.17	i: 15 m: 0.12 o: 20.50 n: 18.83 s: 50 p: 0.17	i: 16 m: 0.04 o: 12.67 n: 10.58 s: 18 p: 0.06	i: 17 m: 0.01 o: 12.75 n: 7.50 s: 136 p: 0.04	i: 18 m: 0.01 o: 4.50 n: 2.33 s: 73 p: -0.05	i: 19 m: 0.02 o: 0.42 n: 0.33 s: 54 p: -0.09		
i: 22 m: 0.00 o: 16.33 n: 15.42 s: 34 p: 0.11	i: 23 m: 0.02 o: 10.58 n: 9.58 s: 159 p: 0.04	i: 24 m: -0.07 o: 1.58 n: 1.50 s: 327 p: -0.07	i: 25 m: -0.10 o: 2.17 n: 1.33 s: 154 p: -0.07	i: 26 m: 0.34 o: 28.92 n: 28.92 s: 93 p: 0.27	i: 27 m: -0.09 o: 2.58 n: 1.75 s: 87 p: -0.06	i: 28 m: 0.00 o: 8.33 n: 6.33 s: 215 p: 0.01	i: 29 m: -0.01 o: 0.08 n: 0.08 s: 176 p: -0.09	i: 30 m: -0.04 o: 13.33 n: 9.83 s: 132 p: 0.06	i: 31 m: 0.01 o: 5.58 n: 4.08 s: 93 p: -0.03	i: 32 m: 0.01 o: 5.33 n: 2.17 s: 44 p: -0.04
i: 33 m: -0.03 o: 1.00 n: 0.67 s: 22 p: -0.08	i: 34 m: -0.04 o: 0.92 n: 0.67 s: 43 p: -0.08	i: 35 m: 0.04 o: 18.25 n: 11.08 s: 272 p: 0.10	i: 36 m: -0.08 o: 4.67 n: 2.67 s: 218 p: -0.04	i: 37 m: 0.08 o: 66.83 n: 62.33 s: 115 p: 0.76	i: 38 m: 0.03 o: 9.42 n: 7.58 s: 201 p: 0.02	i: 39 m: 0.01 o: 23.08 n: 18.33 s: 333 p: 0.18	i: 40 m: 0.00 o: 3.83 n: 3.50 s: 44 p: -0.04	i: 41 m: 0.02 o: 0.50 n: 0.50 s: 57 p: -0.08	i: 42 m: 0.03 o: 1.08 n: 0.83 s: 56 p: -0.08	
i: 45 m: -0.03 o: 0.58 n: 0.58 s: 22 p: -0.08	i: 46 m: 0.00 o: 8.08 n: 5.67 s: 159 p: 0.00	i: 47 m: 0.09 o: 26.92 n: 19.00 s: 361 p: 0.21	i: 48 m: -0.14 o: 0.17 n: 0.17 s: 72 p: -0.09	i: 49 m: 0.04 o: 12.67 n: 11.33 s: 103 p: 0.07	i: 50 m: -0.02 o: 2.25 n: 1.92 s: 267 p: -0.06	i: 51 m: 0.01 o: 0.58 n: 0.25 s: 47 p: -0.09				
i: 56 m: 0.00 o: 9.08 n: 8.50 s: 29 p: 0.02	i: 57 m: -0.03 o: 0.25 n: 0.17 s: 26 p: -0.09	i: 58 m: -0.02 o: 3.42 n: 3.17 s: 51 p: -0.05	i: 59 m: -0.03 o: 1.67 n: 1.67 s: 89 p: -0.07	i: 61 m: 0.00 o: 1.25 n: 1.25 s: 2 p: -0.07						
i: 67 m: -0.01 o: 9.17 n: 9.17 s: 11 p: 0.03	i: 68 m: 0.00 o: 1.92 n: 0.75 s: 24 p: -0.07	i: 69 m: -0.02 o: 12.33 n: 11.17 s: 59 p: 0.06			i: 73 m: 0.02 o: 0.75 n: 0.50 s: 26 p: -0.08					
i: 78 m: 0.00 o: 1.92 n: 1.58 s: 15 p: -0.07	i: 80 m: 0.00 o: 2.33 n: 1.75 s: 90 p: -0.06	i: 81 m: 0.01 o: 4.33 n: 2.17 s: 106 p: -0.05			i: 83 m: 0.01 o: 5.42 n: 5.42 s: 4 p: -0.02	i: 84 m: 0.02 o: 0.42 n: 0.42 s: 16 p: -0.09				

Figure C.1

20170729 LOCALMORANS City 2 - FC 3PCA - ANNUAL - RAW

Feature Legend

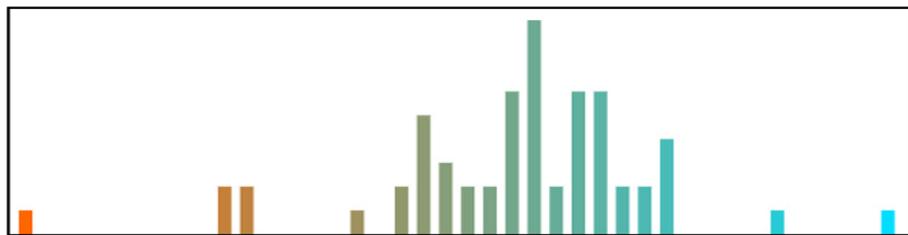
i: Grid cell index
 m: Local moran's I value
 o: 2016 Average opioid prescription count
 n: 2016 Average opioid non-member prescription count
 s: 2016 Average member count

The black bar at the bottom of the cell indicates the normalized population count

Local Moran's I Color Legend



Values Histogram



City Features:

	i: 1 m: -0.04 o: 2.75 n: 2.17 s: 61	i: 2 m: -0.03 o: 3.17 n: 2.92 s: 115	i: 3 m: 0.04 o: 8.83 n: 6.25 s: 123	i: 4 m: 0.02 o: 5.17 n: 2.25 s: 54	i: 5 m: -0.04 o: 2.75 n: 2.25 s: 59	i: 6 m: 0.00 o: 9.58 n: 9.17 s: 52	i: 7 m: -0.01 o: 0.75 n: 0.58 s: 29	i: 8 m: 0.01 o: 0.25 n: 0.17 s: 35			
	i: 11 m: -0.01 o: 4.00 n: 2.50 s: 118	i: 12 m: -0.01 o: 15.08 n: 13.83 s: 150	i: 13 m: -0.04 o: 20.25 n: 13.00 s: 346	i: 14 m: -0.06 o: 0.50 n: 0.33 s: 103	i: 15 m: 0.00 o: 20.50 n: 18.83 s: 50	i: 16 m: -0.02 o: 12.67 n: 10.58 s: 18	i: 17 m: 0.02 o: 12.75 n: 7.50 s: 138	i: 18 m: 0.00 o: 4.50 n: 2.33 s: 73	i: 19 m: 0.01 o: 0.42 n: 0.33 s: 54		
	i: 22 m: -0.01 o: 16.33 n: 15.42 s: 34	i: 23 m: -0.10 o: 10.58 n: 9.58 s: 159	i: 24 m: -0.11 o: 1.58 n: 1.50 s: 327	i: 25 m: -0.04 o: 2.17 n: 1.33 s: 154	i: 26 m: 0.02 o: 28.92 n: 26.92 s: 93	i: 27 m: -0.01 o: 2.58 n: 1.75 s: 87	i: 28 m: 0.07 o: 8.33 n: 6.33 s: 215	i: 29 m: -0.04 o: 0.08 n: 0.08 s: 176	i: 30 m: 0.01 o: 13.33 n: 9.83 s: 132	i: 31 m: -0.05 o: 5.58 n: 4.08 s: 93	i: 32 m: 0.01 o: 5.33 n: 2.17 s: 44
	i: 33 m: -0.02 o: 1.00 n: 0.67 s: 22	i: 34 m: -0.03 o: 0.92 n: 0.67 s: 43	i: 35 m: 0.04 o: 18.25 n: 11.08 s: 272	i: 36 m: 0.04 o: 4.67 n: 2.67 s: 218	i: 37 m: 0.11 o: 66.83 n: 62.33 s: 115	i: 38 m: -0.10 o: 9.42 n: 7.58 s: 201	i: 39 m: -0.17 o: 23.08 n: 18.33 s: 333	i: 40 m: -0.01 o: 3.83 n: 3.50 s: 44	i: 41 m: 0.01 o: 0.50 n: 0.50 s: 57	i: 42 m: 0.01 o: 1.08 n: 0.83 s: 56	
	i: 45 m: -0.04 o: 0.58 n: 0.58 s: 22	i: 46 m: 0.01 o: 8.08 n: 5.67 s: 159	i: 47 m: 0.01 o: 26.92 n: 19.00 s: 381	i: 48 m: -0.11 o: 0.17 n: 0.17 s: 72	i: 49 m: -0.02 o: 12.67 n: 11.33 s: 103	i: 50 m: -0.01 o: 2.25 n: 1.92 s: 267	i: 51 m: -0.01 o: 0.58 n: 0.25 s: 47				
	i: 56 m: -0.01 o: 9.08 n: 8.50 s: 29	i: 57 m: -0.04 o: 0.25 n: 0.17 s: 26	i: 58 m: -0.04 o: 3.42 n: 3.17 s: 51	i: 59 m: 0.00 o: 1.67 n: 1.67 s: 88		i: 61 m: -0.01 o: 1.25 n: 1.25 s: 2					
	i: 67 m: 0.00 o: 9.17 n: 9.17 s: 11	i: 68 m: 0.00 o: 1.92 n: 0.75 s: 24	i: 69 m: -0.05 o: 12.33 n: 11.17 s: 59				i: 73 m: 0.03 o: 0.75 n: 0.50 s: 26				
	i: 78 m: 0.01 o: 1.92 n: 1.58 s: 15		i: 80 m: -0.01 o: 2.33 n: 1.75 s: 90	i: 81 m: 0.00 o: 4.33 n: 2.17 s: 106		i: 83 m: 0.03 o: 5.42 n: 5.42 s: 4	i: 84 m: 0.03 o: 0.42 n: 0.42 s: 16				

Figure C.2

20170729 LOCALMORANS City 2 - FC 6PCA - ANNUAL - RAW

Feature Legend

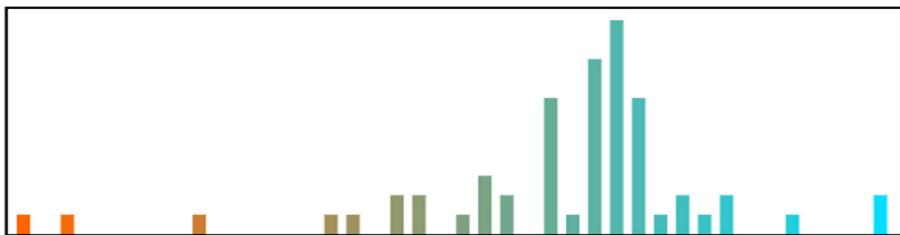
i: Grid cell index
 m: Local moran's I value
 o: 2016 Average opioid prescription count
 n: 2016 Average opioid non-member prescription count
 s: 2016 Average member count

The black bar at the bottom of the cell indicates the normalized population count

Local Moran's I Color Legend



Values Histogram



City Features:

	i: 1 m: 0.00 o: 2.75 n: 2.17 s: 61	i: 2 m: -0.01 o: 3.17 n: 2.92 s: 115	i: 3 m: 0.01 o: 8.83 n: 6.25 s: 123	i: 4 m: 0.00 o: 5.17 n: 2.25 s: 54	i: 5 m: -0.10 o: 2.75 n: 2.25 s: 59	i: 6 m: -0.08 o: 9.58 n: 9.17 s: 52	i: 7 m: 0.00 o: 0.75 n: 0.58 s: 29	i: 8 m: 0.01 o: 0.25 n: 0.17 s: 35		
i: 11 m: -0.11 o: 4.00 n: 2.50 s: 116	i: 12 m: 0.00 o: 15.08 n: 13.83 s: 150	i: 13 m: -0.01 o: 20.25 n: 13.00 s: 346	i: 14 m: 0.00 o: 0.50 n: 0.33 s: 103	i: 15 m: 0.03 o: 20.50 n: 18.83 s: 50	i: 16 m: -0.21 o: 12.67 n: 10.58 s: 18	i: 17 m: -0.02 o: 12.75 n: 7.50 s: 138	i: 18 m: -0.02 o: 4.50 n: 2.33 s: 73	i: 19 m: 0.02 o: 0.42 n: 0.33 s: 54		
i: 22 m: 0.01 o: 16.33 n: 15.42 s: 34	i: 23 m: -0.23 o: 10.58 n: 9.58 s: 159	i: 24 m: 0.00 o: 1.58 n: 1.50 s: 327	i: 25 m: -0.07 o: 2.17 n: 1.33 s: 154	i: 26 m: 0.11 o: 28.92 n: 26.92 s: 93	i: 27 m: 0.04 o: 2.58 n: 1.75 s: 87	i: 28 m: 0.08 o: 8.33 n: 6.33 s: 215	i: 29 m: -0.01 o: 0.08 n: 0.08 s: 176	i: 30 m: -0.05 o: 13.33 n: 9.83 s: 132	i: 31 m: -0.04 o: 5.58 n: 4.08 s: 93	i: 32 m: 0.00 o: 5.33 n: 2.17 s: 44
i: 33 m: 0.01 o: 1.00 n: 0.67 s: 22	i: 34 m: 0.01 o: 0.92 n: 0.67 s: 43	i: 35 m: -0.01 o: 18.25 n: 11.08 s: 272	i: 36 m: 0.01 o: 4.67 n: 2.67 s: 218	i: 37 m: 0.11 o: 66.83 n: 62.33 s: 115	i: 38 m: -0.06 o: 9.42 n: 7.58 s: 201	i: 39 m: 0.05 o: 23.08 n: 18.33 s: 333	i: 40 m: 0.01 o: 3.83 n: 3.50 s: 44	i: 41 m: 0.00 o: 0.50 n: 0.50 s: 57	i: 42 m: 0.01 o: 1.08 n: 0.83 s: 56	
	i: 45 m: -0.03 o: 0.58 n: 0.58 s: 22	i: 46 m: -0.02 o: 8.08 n: 5.67 s: 159	i: 47 m: 0.05 o: 26.92 n: 19.00 s: 381	i: 48 m: -0.02 o: 0.17 n: 0.17 s: 72	i: 49 m: -0.07 o: 12.67 n: 11.33 s: 103	i: 50 m: 0.02 o: 2.25 n: 1.92 s: 267	i: 51 m: -0.01 o: 0.58 n: 0.25 s: 47			
	i: 56 m: -0.16 o: 9.08 n: 8.50 s: 29	i: 57 m: -0.02 o: 0.25 n: 0.17 s: 26	i: 58 m: 0.00 o: 3.42 n: 3.17 s: 51	i: 59 m: 0.03 o: 1.67 n: 1.67 s: 88	i: 61 m: 0.00 o: 1.25 n: 1.25 s: 2					
	i: 67 m: -0.05 o: 9.17 n: 9.17 s: 11	i: 68 m: -0.02 o: 1.92 n: 0.75 s: 24	i: 69 m: -0.08 o: 12.33 n: 11.17 s: 59			i: 73 m: 0.00 o: 0.75 n: 0.50 s: 26				
	i: 78 m: 0.01 o: 1.92 n: 1.58 s: 15	i: 80 m: -0.05 o: 2.33 n: 1.75 s: 90	i: 81 m: 0.00 o: 4.33 n: 2.17 s: 106		i: 83 m: -0.04 o: 5.42 n: 5.42 s: 4	i: 84 m: -0.01 o: 0.42 n: 0.42 s: 16				

Figure C.3

20170729 LOCALMORANS City 2 - FC 9PCA - ANNUAL - RAW

Feature Legend

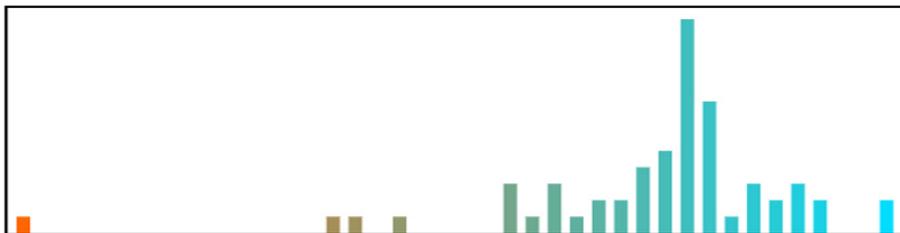
i: Grid cell index
 m: Local moran's I value
 o: 2016 Average opioid prescription count
 n: 2016 Average opioid non-member prescription count
 s: 2016 Average member count

The black bar at the bottom of the cell indicates the normalized population count

Local Moran's I Color Legend



Values Histogram



City Features:

	i: 1 m: 0.01 o: 2.75 n: 2.17 s: 61	i: 2 m: 0.00 o: 3.17 n: 2.92 s: 115	i: 3 m: -0.10 o: 8.83 n: 6.25 s: 123	i: 4 m: 0.00 o: 5.17 n: 2.25 s: 54	i: 5 m: -0.01 o: 2.75 n: 2.25 s: 59	i: 6 m: -0.19 o: 9.58 n: 9.17 s: 52	i: 7 m: 0.01 o: 0.75 n: 0.58 s: 29	i: 8 m: 0.01 o: 0.25 n: 0.17 s: 35		
i: 11 m: -0.08 o: 4.00 n: 2.50 s: 116	i: 12 m: -0.02 o: 15.08 n: 13.83 s: 150	i: 13 m: -0.10 o: 20.25 n: 13.00 s: 346	i: 14 m: -0.02 o: 0.50 n: 0.33 s: 103	i: 15 m: -0.07 o: 29.50 n: 18.83 s: 50	i: 16 m: -0.07 o: 12.67 n: 10.58 s: 18	i: 17 m: -0.03 o: 12.75 n: 7.50 s: 138	i: 18 m: 0.05 o: 4.50 n: 2.33 s: 73	i: 19 m: 0.01 o: 0.42 n: 0.33 s: 54		
i: 22 m: 0.00 o: 16.33 n: 15.42 s: 34	i: 23 m: -0.36 o: 10.58 n: 9.58 s: 159	i: 24 m: 0.01 o: 1.58 n: 1.50 s: 327	i: 25 m: -0.15 o: 2.17 n: 1.33 s: 154	i: 26 m: 0.06 o: 23.92 n: 26.92 s: 93	i: 27 m: -0.05 o: 2.58 n: 1.75 s: 87	i: 28 m: 0.00 o: 8.33 n: 6.33 s: 215	i: 29 m: 0.00 o: 0.08 n: 0.08 s: 176	i: 30 m: 0.05 o: 13.33 n: 9.83 s: 132	i: 31 m: 0.06 o: 5.58 n: 4.08 s: 93	i: 32 m: 0.03 o: 5.33 n: 2.17 s: 44
i: 33 m: 0.00 o: 1.00 n: 0.67 s: 22	i: 34 m: 0.00 o: 0.92 n: 0.67 s: 43	i: 35 m: 0.04 o: 18.25 n: 11.08 s: 272	i: 36 m: 0.07 o: 4.67 n: 2.67 s: 218	i: 37 m: 0.11 o: 66.83 n: 62.33 s: 115	i: 38 m: 0.04 o: 9.42 n: 7.58 s: 201	i: 39 m: -0.17 o: 23.08 n: 18.33 s: 333	i: 40 m: -0.01 o: 3.83 n: 3.50 s: 44	i: 41 m: 0.00 o: 0.50 n: 0.50 s: 57	i: 42 m: 0.00 o: 1.08 n: 0.83 s: 56	
i: 45 m: -0.01 o: 0.58 n: 0.58 s: 22	i: 46 m: 0.06 o: 8.08 n: 5.67 s: 159	i: 47 m: 0.11 o: 26.92 n: 19.00 s: 381	i: 48 m: 0.02 o: 0.17 n: 0.17 s: 72	i: 49 m: -0.03 o: 12.67 n: 11.33 s: 103	i: 50 m: -0.05 o: 2.25 n: 1.92 s: 267	i: 51 m: 0.00 o: 0.58 n: 0.25 s: 47				
i: 56 m: 0.00 o: 9.08 n: 8.50 s: 29	i: 57 m: -0.03 o: 0.25 n: 0.17 s: 26	i: 58 m: 0.07 o: 3.42 n: 3.17 s: 51	i: 59 m: 0.01 o: 1.67 n: 1.67 s: 88		i: 61 m: 0.01 o: 1.25 n: 1.25 s: 2					
i: 67 m: -0.07 o: 9.17 n: 9.17 s: 11	i: 68 m: -0.02 o: 1.92 n: 0.75 s: 24	i: 69 m: 0.01 o: 12.33 n: 11.17 s: 59				i: 73 m: 0.00 o: 0.75 n: 0.50 s: 26				
i: 78 m: -0.01 o: 1.92 n: 1.58 s: 15		i: 80 m: -0.05 o: 2.33 n: 1.75 s: 90	i: 81 m: 0.00 o: 4.33 n: 2.17 s: 106		i: 83 m: -0.10 o: 5.42 n: 5.42 s: 4	i: 84 m: -0.02 o: 0.42 n: 0.42 s: 16				

Figure C.4

20170729 LOCALMORANS City 2 - FC 12PCA - ANNUAL - RAW

Feature Legend

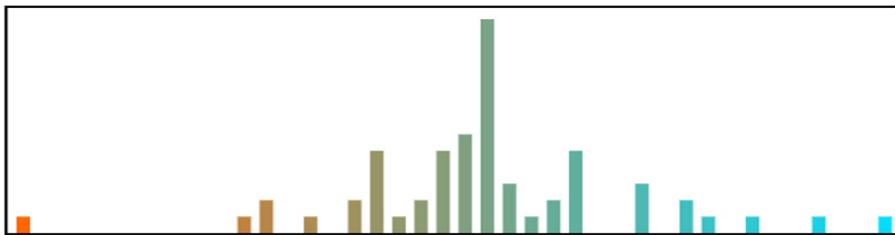
i: Grid cell index
 m: Local moran's I value
 o: 2016 Average opioid prescription count
 n: 2016 Average opioid non-member prescription count
 s: 2016 Average member count

The black bar at the bottom of the cell indicates the normalized population count

Local Moran's I Color Legend



Values Histogram



City Features:

	i: 1 m: 0.00 o: 2.75 n: 2.17 s: 61	i: 2 m: 0.00 o: 3.17 n: 2.92 s: 115	i: 3 m: -0.03 o: 8.63 n: 8.25 s: 123	i: 4 m: 0.00 o: 5.17 n: 2.25 s: 54	i: 5 m: 0.02 o: 2.75 n: 2.25 s: 59	i: 6 m: -0.06 o: 9.58 n: 9.17 s: 52	i: 7 m: 0.00 o: 0.75 n: 0.58 s: 29	i: 8 m: -0.01 o: 0.25 n: 0.17 s: 35		
i: 11 m: -0.05 o: 4.00 n: 2.50 s: 118	i: 12 m: 0.00 o: 15.08 n: 13.63 s: 150	i: 13 m: -0.09 o: 20.25 n: 13.00 s: 346	i: 14 m: -0.02 o: 0.50 n: 0.33 s: 103	i: 15 m: -0.06 o: 20.50 n: 18.83 s: 50	i: 16 m: -0.11 o: 12.67 n: 10.58 s: 18	i: 17 m: 0.01 o: 12.75 n: 7.50 s: 138	i: 18 m: 0.05 o: 4.50 n: 2.33 s: 73	i: 19 m: 0.01 o: 0.42 n: 0.33 s: 54		
i: 22 m: 0.00 o: 16.33 n: 15.42 s: 34	i: 23 m: -0.11 o: 10.58 n: 9.58 s: 159	i: 24 m: 0.00 o: 1.58 n: 1.50 s: 327	i: 25 m: -0.05 o: 2.17 n: 1.33 s: 154	i: 26 m: 0.08 o: 28.92 n: 26.92 s: 93	i: 27 m: -0.02 o: 2.58 n: 1.75 s: 87	i: 28 m: 0.00 o: 8.33 n: 8.33 s: 215	i: 29 m: -0.02 o: 0.08 n: 0.08 s: 176	i: 30 m: 0.05 o: 13.33 n: 9.83 s: 132	i: 31 m: 0.05 o: 5.58 n: 4.08 s: 93	i: 32 m: 0.08 o: 5.33 n: 2.17 s: 44
i: 33 m: 0.00 o: 1.00 n: 0.67 s: 22	i: 34 m: 0.00 o: 0.92 n: 0.67 s: 43	i: 35 m: 0.01 o: 18.25 n: 11.08 s: 272	i: 36 m: 0.00 o: 4.67 n: 2.67 s: 218	i: 37 m: 0.14 o: 66.83 n: 62.33 s: 115	i: 38 m: 0.10 o: 9.42 n: 7.58 s: 201	i: 39 m: -0.05 o: 23.08 n: 18.33 s: 333	i: 40 m: -0.03 o: 3.83 n: 3.50 s: 44	i: 41 m: 0.01 o: 0.50 n: 0.50 s: 57	i: 42 m: 0.04 o: 1.08 n: 0.83 s: 56	
i: 45 m: 0.00 o: 0.58 n: 0.58 s: 22	i: 46 m: 0.05 o: 8.08 n: 5.67 s: 159	i: 47 m: 0.17 o: 26.92 n: 19.00 s: 381	i: 48 m: 0.05 o: 0.17 n: 0.17 s: 72	i: 49 m: 0.11 o: 12.67 n: 11.33 s: 103	i: 50 m: 0.03 o: 2.25 n: 1.92 s: 267	i: 51 m: 0.00 o: 0.58 n: 0.25 s: 47				
i: 56 m: -0.06 o: 9.08 n: 8.50 s: 29	i: 57 m: 0.00 o: 0.25 n: 0.17 s: 26	i: 58 m: 0.21 o: 3.42 n: 3.17 s: 51	i: 59 m: 0.08 o: 1.67 n: 1.67 s: 88		i: 61 m: -0.04 o: 1.25 n: 1.25 s: 2					
i: 67 m: -0.02 o: 9.17 n: 9.17 s: 11	i: 68 m: -0.06 o: 1.92 n: 0.75 s: 24	i: 69 m: 0.10 o: 12.33 n: 11.17 s: 59				i: 73 m: 0.00 o: 0.75 n: 0.50 s: 26				
i: 78 m: 0.00 o: 1.92 n: 1.58 s: 15		i: 80 m: -0.12 o: 2.33 n: 1.75 s: 90	i: 81 m: 0.03 o: 4.33 n: 2.17 s: 106		i: 83 m: -0.23 o: 5.42 n: 5.42 s: 4	i: 84 m: -0.02 o: 0.42 n: 0.42 s: 16				

Figure C.5

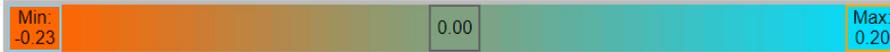
20170729 LOCALMORANS City 2 - FC 15PCA - ANNUAL - RAW

Feature Legend

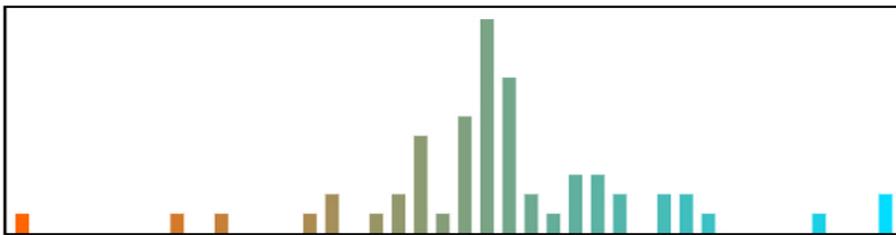
i: Grid cell index
 m: Local moran's I value
 o: 2016 Average opioid prescription count
 n: 2016 Average opioid non-member prescription count
 s: 2016 Average member count

The black bar at the bottom of the cell indicates the normalized population count

Local Moran's I Color Legend



Values Histogram



City Features:

	i: 1 m: 0.02 o: 2.75 n: 2.50 s: 61	i: 2 m: 0.01 o: 3.17 n: 2.92 s: 115	i: 3 m: 0.00 o: 8.83 n: 6.25 s: 123	i: 4 m: -0.01 o: 5.17 n: 2.25 s: 54	i: 5 m: 0.01 o: 2.75 n: 2.25 s: 59	i: 6 m: -0.04 o: 9.58 n: 9.17 s: 52	i: 7 m: 0.00 o: 0.75 n: 0.58 s: 29	i: 8 m: 0.00 o: 0.25 n: 0.17 s: 35		
i: 11 m: 0.02 o: 4.00 n: 2.50 s: 118	i: 12 m: 0.02 o: 15.08 n: 13.83 s: 150	i: 13 m: 0.00 o: 20.25 n: 13.00 s: 346	i: 14 m: -0.02 o: 0.50 n: 0.33 s: 103	i: 15 m: 0.00 o: 20.50 n: 18.83 s: 50	i: 16 m: -0.15 o: 12.67 n: 10.58 s: 18	i: 17 m: -0.07 o: 12.75 n: 7.50 s: 138	i: 18 m: -0.08 o: 4.50 n: 2.33 s: 73	i: 19 m: 0.01 o: 0.42 n: 0.33 s: 54		
i: 22 m: 0.05 o: 16.33 n: 15.42 s: 34	i: 23 m: -0.05 o: 10.58 n: 9.58 s: 159	i: 24 m: 0.00 o: 1.58 n: 1.50 s: 327	i: 25 m: -0.03 o: 2.17 n: 1.33 s: 154	i: 26 m: 0.10 o: 28.92 n: 26.92 s: 93	i: 27 m: -0.01 o: 2.58 n: 1.75 s: 87	i: 28 m: -0.02 o: 8.33 n: 6.33 s: 215	i: 29 m: 0.00 o: 0.08 n: 0.08 s: 176	i: 30 m: 0.03 o: 13.33 n: 9.83 s: 132	i: 31 m: 0.06 o: 5.58 n: 4.08 s: 93	i: 32 m: 0.06 o: 5.33 n: 2.17 s: 44
i: 33 m: -0.01 o: 1.00 n: 0.67 s: 22	i: 34 m: -0.01 o: 0.92 n: 0.67 s: 43	i: 35 m: 0.05 o: 18.25 n: 11.08 s: 272	i: 36 m: 0.04 o: 4.67 n: 2.67 s: 218	i: 37 m: 0.16 o: 66.83 n: 62.33 s: 115	i: 38 m: 0.19 o: 9.42 n: 7.58 s: 201	i: 39 m: -0.03 o: 23.08 n: 18.33 s: 333	i: 40 m: -0.07 o: 3.83 n: 3.50 s: 44	i: 41 m: 0.02 o: 0.50 n: 0.50 s: 57	i: 42 m: 0.05 o: 1.08 n: 0.83 s: 56	
	i: 45 m: 0.00 o: 0.58 n: 0.58 s: 22	i: 46 m: 0.10 o: 8.08 n: 5.67 s: 159	i: 47 m: 0.09 o: 26.92 n: 19.00 s: 381	i: 48 m: 0.05 o: 0.17 n: 0.17 s: 72	i: 49 m: 0.09 o: 12.67 n: 11.33 s: 103	i: 50 m: 0.01 o: 2.25 n: 1.92 s: 267	i: 51 m: 0.00 o: 0.58 n: 0.25 s: 47			
	i: 56 m: -0.03 o: 9.08 n: 8.50 s: 29	i: 57 m: 0.01 o: 0.25 n: 0.17 s: 26	i: 58 m: 0.20 o: 3.42 n: 3.17 s: 51	i: 59 m: 0.05 o: 1.67 n: 1.67 s: 88		i: 61 m: -0.04 o: 1.25 n: 1.25 s: 2				
	i: 67 m: -0.04 o: 9.17 n: 9.17 s: 11	i: 68 m: -0.04 o: 1.92 n: 0.75 s: 24	i: 69 m: 0.10 o: 12.33 n: 11.17 s: 59				i: 73 m: 0.00 o: 0.75 n: 0.50 s: 26			
	i: 78 m: 0.01 o: 1.92 n: 1.58 s: 15		i: 80 m: -0.13 o: 2.33 n: 1.75 s: 90	i: 81 m: 0.03 o: 4.33 n: 2.17 s: 106		i: 83 m: -0.23 o: 5.42 n: 5.42 s: 4	i: 84 m: -0.01 o: 0.42 n: 0.42 s: 16			

Figure C.6

20170729 LOCALMORANS City 2 - FC 18PCA - ANNUAL - RAW

Feature Legend

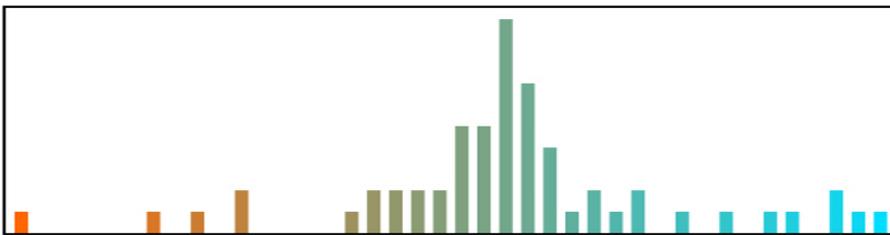
i: Grid cell index
 m: Local moran's I value
 o: 2016 Average opioid prescription count
 n: 2016 Average opioid non-member prescription count
 s: 2016 Average member count

The black bar at the bottom of the cell indicates the normalized population count

Local Moran's I Color Legend



Values Histogram



City Features:

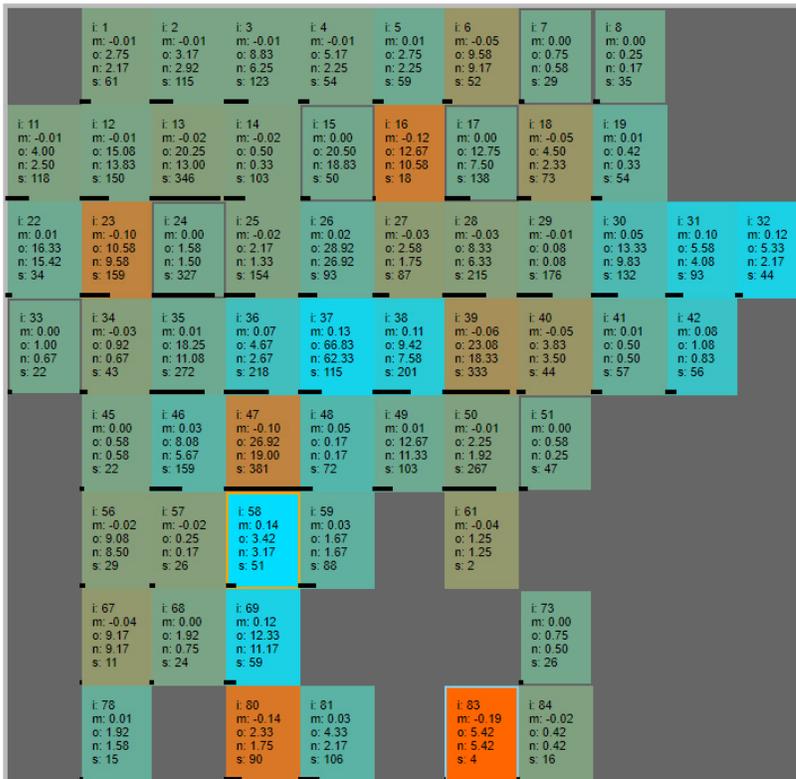


Figure C.7

20170729 LOCALMORANS City 2 - FC 21PCA - ANNUAL - RAW

Feature Legend

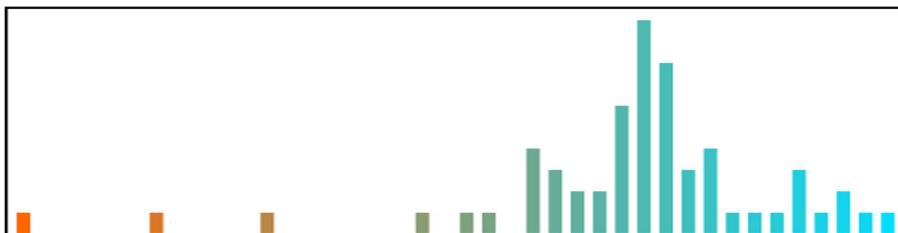
i: Grid cell index
 m: Local moran's I value
 o: 2016 Average opioid prescription count
 n: 2016 Average opioid non-member prescription count
 s: 2016 Average member count

The black bar at the bottom of the cell indicates the normalized population count

Local Moran's I Color Legend



Values Histogram



City Features:

	i: 1 m: 0.00 o: 2.75 n: 2.17 s: 61	i: 2 m: 0.01 o: 3.17 n: 2.92 s: 115	i: 3 m: 0.00 o: 8.83 n: 6.25 s: 123	i: 4 m: 0.00 o: 5.17 n: 2.25 s: 54	i: 5 m: 0.00 o: 2.75 n: 2.25 s: 59	i: 6 m: -0.03 o: 9.58 n: 9.17 s: 52	i: 7 m: 0.00 o: 0.75 n: 0.58 s: 29	i: 8 m: 0.00 o: 0.25 n: 0.17 s: 35		
i: 11 m: 0.00 o: 4.00 n: 2.50 s: 118	i: 12 m: -0.02 o: 15.08 n: 13.83 s: 150	i: 13 m: -0.01 o: 20.25 n: 13.00 s: 346	i: 14 m: -0.02 o: 0.50 n: 0.33 s: 103	i: 15 m: 0.00 o: 20.50 n: 18.83 s: 50	i: 16 m: -0.06 o: 12.67 n: 10.58 s: 18	i: 17 m: 0.01 o: 12.75 n: 7.50 s: 138	i: 18 m: -0.05 o: 4.50 n: 2.33 s: 73	i: 19 m: 0.03 o: 0.42 n: 0.33 s: 54		
i: 22 m: 0.02 o: 16.33 n: 15.42 s: 34	i: 23 m: -0.12 o: 10.58 n: 9.58 s: 159	i: 24 m: -0.02 o: 1.58 n: 1.50 s: 327	i: 25 m: -0.06 o: 2.17 n: 1.33 s: 154	i: 26 m: 0.00 o: 28.92 n: 26.92 s: 93	i: 27 m: -0.05 o: 2.58 n: 1.75 s: 87	i: 28 m: -0.02 o: 8.33 n: 6.33 s: 215	i: 29 m: 0.00 o: 0.08 n: 0.08 s: 176	i: 30 m: 0.07 o: 13.33 n: 9.83 s: 132	i: 31 m: 0.09 o: 5.58 n: 4.08 s: 93	i: 32 m: 0.11 o: 5.33 n: 2.17 s: 44
i: 33 m: 0.00 o: 1.00 n: 0.67 s: 22	i: 34 m: -0.01 o: 0.92 n: 0.67 s: 43	i: 35 m: -0.01 o: 18.25 n: 11.08 s: 272	i: 36 m: 0.08 o: 4.67 n: 2.67 s: 218	i: 37 m: 0.09 o: 66.83 n: 62.33 s: 115	i: 38 m: 0.00 o: 9.42 n: 7.58 s: 201	i: 39 m: -0.19 o: 23.08 n: 18.33 s: 333	i: 40 m: -0.06 o: 3.83 n: 3.50 s: 44	i: 41 m: 0.02 o: 0.50 n: 0.50 s: 57	i: 42 m: 0.09 o: 1.08 n: 0.83 s: 56	
	i: 45 m: 0.00 o: 0.58 n: 0.58 s: 22	i: 46 m: 0.02 o: 8.08 n: 5.67 s: 159	i: 47 m: -0.32 o: 26.92 n: 19.00 s: 381	i: 48 m: 0.03 o: 0.17 n: 0.17 s: 72	i: 49 m: 0.06 o: 12.67 n: 11.33 s: 103	i: 50 m: -0.03 o: 2.25 n: 1.92 s: 267	i: 51 m: -0.01 o: 0.58 n: 0.25 s: 47			
	i: 56 m: -0.01 o: 9.08 n: 8.50 s: 29	i: 57 m: -0.09 o: 0.25 n: 0.17 s: 26	i: 58 m: 0.07 o: 3.42 n: 3.17 s: 51	i: 59 m: 0.05 o: 1.67 n: 1.67 s: 88		i: 61 m: -0.04 o: 1.25 n: 1.25 s: 2				
	i: 67 m: -0.05 o: 9.17 n: 9.17 s: 11	i: 68 m: 0.04 o: 1.92 n: 0.75 s: 24	i: 69 m: 0.12 o: 12.33 n: 11.17 s: 59				i: 73 m: 0.00 o: 0.75 n: 0.50 s: 26			
	i: 78 m: 0.03 o: 1.92 n: 1.58 s: 15		i: 80 m: -0.08 o: 2.33 n: 1.75 s: 90	i: 81 m: 0.03 o: 4.33 n: 2.17 s: 106		i: 83 m: -0.25 o: 5.42 n: 5.42 s: 4	i: 84 m: -0.05 o: 0.42 n: 0.42 s: 16			

Figure C.8

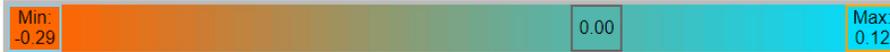
20170729 LOCALMORANS City 2 - FC 24PCA - ANNUAL - RAW

Feature Legend

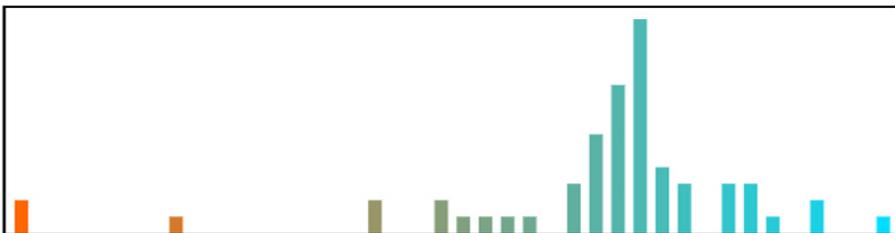
i: Grid cell index
 m: Local moran's I value
 o: 2016 Average opioid prescription count
 n: 2016 Average opioid non-member prescription count
 s: 2016 Average member count

The black bar at the bottom of the cell indicates the normalized population count

Local Moran's I Color Legend



Values Histogram



City Features:

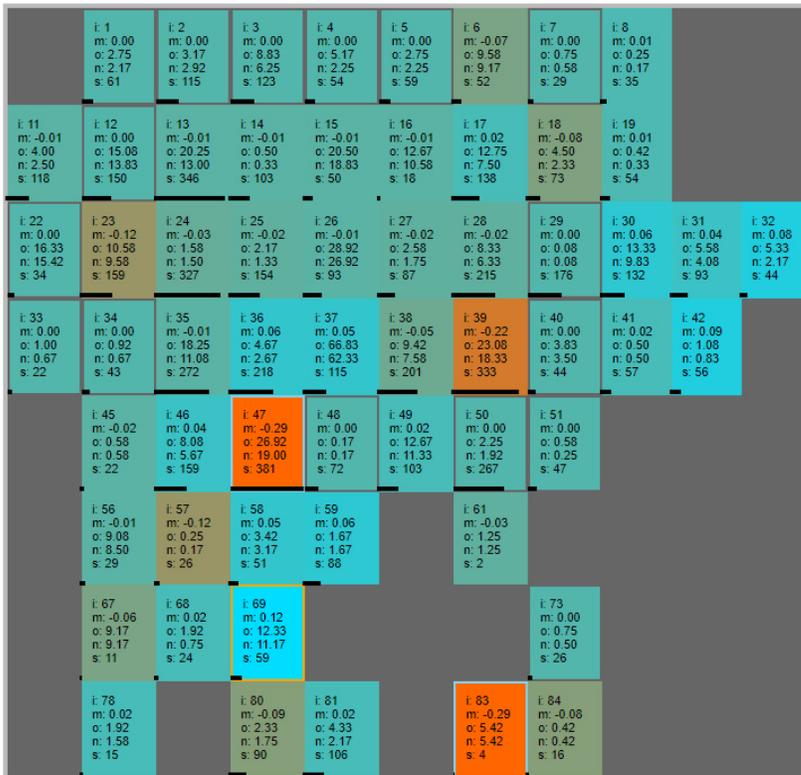


Figure C.9

APPENDIX D

CITY 2 ORIGINAL FEATURE VECTOR RESULTS

20170729 LOCALMORANS City 2 - FC VECTOR - ANNUAL - RAW

Feature Legend

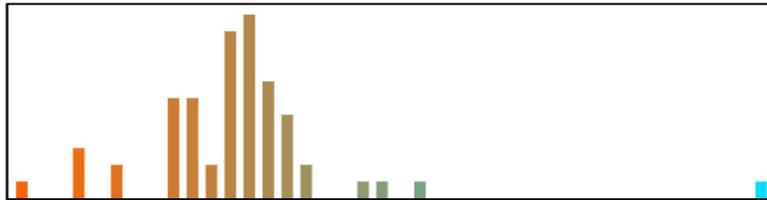
i: Grid cell index
 m: Local moran's I value
 o: 2016 Average opioid prescription count
 n: 2016 Average opioid non-member prescription count
 s: 2016 Average member count
 p: First PCA value

The black bar at the bottom of the cell indicates the normalized population count

Local Moran's I Color Legend



Values Histogram



City Features:

	i: 1 m: -0.80 o: 2.75 n: 2.17 s: 61 p: -0.06	i: 2 m: -0.74 o: 3.17 n: 2.92 s: 115 p: -0.05	i: 3 m: 0.08 o: 8.83 n: 6.25 s: 123 p: 0.01	i: 4 m: -0.72 o: 5.17 n: 2.25 s: 54 p: -0.04	i: 5 m: -0.83 o: 2.75 n: 2.25 s: 59 p: -0.06	i: 6 m: -0.01 o: 9.58 n: 9.17 s: 52 p: 0.03	i: 7 m: 0.19 o: 0.75 n: 0.58 s: 23 p: -0.08	i: 8 m: 0.63 o: 0.25 n: 0.17 s: 35 p: -0.09			
	i: 11 m: -0.77 o: 4.00 n: 2.50 s: 118 p: -0.05	i: 12 m: 0.71 o: 15.08 n: 13.83 s: 150 p: 0.10	i: 13 m: 0.81 o: 20.25 n: 13.00 s: 346 p: 0.13	i: 14 m: -2.24 o: 0.50 n: 0.33 s: 103 p: -0.09	i: 15 m: 2.68 o: 20.50 n: 18.83 s: 50 p: 0.17	i: 16 m: 0.96 o: 12.67 n: 10.58 s: 18 p: 0.06	i: 17 m: 0.17 o: 12.75 n: 7.50 s: 138 p: 0.04	i: 18 m: 0.19 o: 4.50 n: 2.33 s: 73 p: -0.05	i: 19 m: 0.57 o: 0.42 n: 0.33 s: 54 p: -0.09		
	i: 22 m: 0.06 o: 16.33 n: 15.42 s: 34 p: 0.11	i: 23 m: 0.43 o: 10.58 n: 9.58 s: 159 p: 0.04	i: 24 m: -1.56 o: 1.58 n: 1.50 s: 327 p: -0.07	i: 25 m: -2.24 o: 4.67 n: 1.33 s: 154 p: -0.07	i: 26 m: 7.56 o: 28.92 n: 28.92 s: 93 p: 0.27	i: 27 m: -2.12 o: 2.58 n: 1.75 s: 87 p: -0.06	i: 28 m: 0.05 o: 8.33 n: 6.33 s: 215 p: 0.01	i: 29 m: -0.18 o: 0.08 n: 0.08 s: 176 p: -0.09	i: 30 m: -0.92 o: 13.33 n: 9.83 s: 132 p: 0.06	i: 31 m: 0.28 o: 5.58 n: 4.08 s: 93 p: -0.03	i: 32 m: 0.34 o: 5.33 n: 2.17 s: 44 p: -0.04
	i: 33 m: -0.76 o: 1.00 n: 0.67 s: 22 p: -0.08	i: 34 m: -0.94 o: 0.92 n: 0.67 s: 43 p: -0.08	i: 35 m: 0.91 o: 10.25 n: 11.08 s: 272 p: 0.10	i: 36 m: -1.73 o: 4.67 n: 2.67 s: 218 p: -0.04	i: 37 m: 1.81 o: 66.83 n: 62.33 s: 115 p: 0.76	i: 38 m: 0.68 o: 5.42 n: 1.58 s: 201 p: 0.02	i: 39 m: 0.18 o: 23.08 n: 18.33 s: 333 p: 0.18	i: 40 m: 0.04 o: 3.83 n: 3.50 s: 44 p: -0.04	i: 41 m: 0.47 o: 0.50 n: 0.50 s: 57 p: -0.08	i: 42 m: 0.58 o: 1.88 n: 0.83 s: 56 p: -0.08	
	i: 45 m: -0.58 o: 0.58 n: 0.58 s: 22 p: -0.08	i: 46 m: -0.01 o: 0.08 n: 5.67 s: 159 p: 0.00	i: 47 m: 1.96 o: 26.82 n: 19.00 s: 381 p: 0.21	i: 48 m: -3.16 o: 0.17 n: 0.17 s: 72 p: -0.09	i: 49 m: 0.86 o: 12.67 n: 11.33 s: 103 p: 0.07	i: 50 m: -0.51 o: 2.25 n: 1.92 s: 267 p: -0.06	i: 51 m: 0.12 o: 0.58 n: 0.25 s: 47 p: -0.09				
	i: 56 m: -0.07 o: 9.08 n: 8.50 s: 29 p: 0.02	i: 57 m: -0.60 o: 0.25 n: 0.17 s: 26 p: -0.09	i: 58 m: -0.48 o: 3.42 n: 3.17 s: 51 p: -0.05	i: 59 m: -0.71 o: 1.67 n: 1.67 s: 88 p: -0.07		i: 61 m: -0.01 o: 1.25 n: 1.25 s: 2 p: -0.07					
	i: 67 m: -0.16 o: 9.17 n: 9.17 s: 11 p: 0.03	i: 68 m: -0.09 o: 1.92 n: 0.75 s: 24 p: -0.07	i: 69 m: -0.41 o: 12.33 n: 11.17 s: 59 p: 0.06			i: 73 m: 0.50 o: 0.75 n: 0.50 s: 26 p: -0.08					
	i: 78 m: -0.02 o: 1.92 n: 1.58 s: 15 p: -0.07		i: 80 m: 0.06 o: 2.33 n: 1.75 s: 90 p: -0.06	i: 81 m: 0.12 o: 4.33 n: 2.17 s: 108 p: -0.05		i: 83 m: 0.15 o: 5.42 n: 5.42 s: 4 p: -0.02	i: 84 m: 0.53 o: 0.42 n: 0.42 s: 16 p: -0.09				

Figure D.1

REFERENCES

- 104th Congress. "Health Insurance Portability and Accountability Act of 1996." *H.R. Res. 3103. U.S. G.P.O.* 1936 (1996) (enacted). Accessed July 1, 2017.
- 111th Congress. "The Patient Protection and Affordable Care Act." *H.R. Res. 3590*, (2010) (enacted). Accessed July 1, 2017.
- Anthem, Inc. "Anthem Blue Cross and Blue Shield Program Tackles Inappropriate Opioid and Rx Drug Use." May 25, 2016. Accessed May 30, 2016. <https://www.anthem.com/press/newhampshire/anthem-blue-cross-and-blue-shield-program-tackles-inappropriate-opioid-and-rx-drug-use/>.
- Abdi, Hervé, and Lynne J. Williams. "Principal component analysis." *Wiley Interdisciplinary Reviews: Computational Statistics* 2, no. 4 (2010): 433-59. Accessed July 19, 2017. doi:10.1002/wics.101.
- Anselin, Luc. "Local Indicators of Spatial Association-LISA." *Geographical Analysis* 27, no. 2 (2010): 93-115, Accessed July 18, 2017. doi:10.1111/j.1538-4632.1995.tb00338.x.
- Bostock, Mike. "Data-Driven Documents." *D3.js*. Accessed January 5, 2017. <https://d3js.org/>.
- Foster, K. A., and J. A. Hipp. "Defining Neighborhood Boundaries for Social Measurement: Advancing Social Work Research." *Social Work Research* 35, no. 1 (2011): 25-35, Accessed August 1, 2017. doi:10.1093/swr/35.1.25.
- "Indicators of spatial association." *Wikipedia*. July 28, 2016. Accessed August 14, 2017. https://en.wikipedia.org/wiki/Indicators_of_spatial_association.
- James, Gareth M., Jing Wang, and Ji Zhu. "Functional linear regression that's interpretable." *The Annals of Statistics* 37, no. 5A (2009): 2083-108. Accessed July 18, 2017. doi:10.1214/08-aos641.
- Liu, Yu, Daoqin Tong, and Xi Liu. "Measuring Spatial Autocorrelation of Vectors." *Geographical Analysis* 47, no. 3 (2014): 300-19. Accessed August 24, 2017. doi:10.1111/gean.12069.
- Loan, Charles F. Van. "Generalizing the Singular Value Decomposition." *SIAM Journal on Numerical Analysis* 13, no. 1 (1976): 76-83. Accessed July 18, 2017. doi:10.1137/0713009.

- Maki, Masayuki, and Michiko Masuda. "Spatial Autocorrelation of Genotypes in a Gynodioecious Population of *Chionographis japonica* var. *kurohimensis* (Liliaceae)." *International Journal of Plant Sciences* 154, no. 4 (1993): 467-72. doi:10.1086/297130.
- Mencken, F. Carson, and Cynthia Barnett. "Murder, Nonnegligent Manslaughter, and Spatial Autocorrelation in Mid-South Counties." *Journal of Quantitative Criminology* 15, no. 4 (1999): 407-22, Accessed August 1, 2017. <http://www.jstor.org/stable/23366750>.
- Moran, P. A. P. "Notes on Continuous Stochastic Phenomena." *Biometrika* 37, no. 1/2 (1950): 17, Accessed July 14, 2017. doi:10.2307/2332142.
- "Morans I." Wikipedia. May 29, 2017. Accessed November 14, 2017. https://en.wikipedia.org/wiki/Moran%27s_I.
- Ogura, Takahiro, and Talmage D. Egan. "Opioid Agonists and Antagonists." *Pharmacology and Physiology for Anesthesia*, 2013, 253-71. Accessed July 14, 2017. doi:10.1016/b978-1-4377-1679-5.00015-6.
- Oliphant, James. "Trump declares national emergency on opioid abuse." *Reuters*. August 10, 2017. Accessed August 15, 2017. <https://www.reuters.com/article/us-usa-trump-opioid/trump-declares-national-emergency-on-opioid-abuse-idUSKBN1AQ2AW>.
- "Opioid Abuse and Addiction: MedlinePlus." *MedlinePlus*. Accessed July 15, 2017. <https://medlineplus.gov/opioidabuseandaddiction.html>.
- Pearson, Karl. "LIII. On lines and planes of closest fit to systems of points in space." *Philosophical Magazine Series 6* 2, no. 11 (1901): 559-72. Accessed July 18, 2017. doi:10.1080/14786440109462720.
- "Spatial Autocorrelation and Moran's I in GIS." *GIS Geography*. August 11, 2017. Accessed November 14, 2017. <http://gisgeography.com/spatial-autocorrelation-moran-i-gis/>.