Jason Park

GEOG 461

Assignment 5

Pattern and Cluster Analysis of Late-Diagnosed Breast Cancer Cases in Chicago, Illinois

Purpose:

The purpose of this assignment was to use statistical methods of pattern and cluster analysis to detect spatial clustering of late-diagnosed breast cancer in the Chicago area based on ZIP codes. We determined if the overall clustering of late-diagnosed breast cancer was statistically significant by calculating Moran's I scores based on a spatial weight's matrix. Then, we performed a local cluster analysis to identify the ZIP code locations that represent significant clusters of late-stage breast cancer cases to create a map of the degrees of breast cancer cases. The importance of this map was to identify locations in Chicago where there was high prevalence of late-diagnosed breast cancer, as a late diagnosis is correlated with higher risk of death and illness. This map can be used to determine areas where poor health outcomes are more likely to occur, and therefore, have a greater need for health care.

Methodology:

To begin our analysis, we first brought in a shapefile of breast cancer databases for ZIP codes in Chicago, Illinois. The attribute table of the shapefile contains data of the rates of late-diagnosed breast cancer cases in each Chicago ZIP code, but is not yet symbolized the represent the variations in this statistic. To map late diagnosis rates by ZIP code, we classified the shapefile into 5 class intervals based on the number of standard deviations away from the mean value of late breast cancer diagnosis rates of all ZIP codes (Figure A).

We determined if this pattern of global clustering of late-diagnosed breast cancer rates was statistically significant by calculating and analyzing Moran's I scores. We utilized the "Generate Spatial Weights Matrix" function found in ArcToolbox to generate a spatial weight's matrix that was used as the basis to calculate a Moran's I score of the breast cancer data. The Moran's I score was then calculated using the "Spatial Autocorrelation (Moran's I) function found within the "Spatial Statistics Tool" option in ArcToolbox. This function was used to measure the overall clustering of data, and produced a quantitative measure of how similar one ZIP code is to other ZIP codes surrounding it in terms of late-diagnosed breast cancer rates:

- Global Moran's I Results
    - Index: 0.536294
    - Z-Score: 5.811839
    - P-Value: 0

Viewing the results of the Global Moran's I results showed us that areas of late-diagnosed breast cancer rates were clustered together based on a positive Moran's Index value. , and had less than a 1% chance of occurring by chance based on the set 0.01 significance level.

While the Moran's I score indicated a strong, positive spatial autocorrelation found in the breast cancer dataset, it did not identify visually ZIP code locations where the degrees of clustering occurred. To determine areas where high/low clustering of late-diagnosed breast cancer diagnosis occurred, we performed a local clustering analysis.

The local clustering analysis was performed using the "Clustering and Outlier Analysis (Anselin Local Moran's I) function found within the "Spatial Statistics Tool" option in ArcToolbox. The analysis was based on the percentages of late-diagnosed breast cancer diagnosis rates, and used the same spatial weight's matrix utilized to the Moran's I score to create a new layer that symbolized Chicago ZIP code areas based on their Local Moran's I scores (Figure B). The Z- and P-scores produced from the local clustering analysis were measures of statistical significance that indicated whether spatial clustering existed (Figure C). Based on the results, we can identify where significant ZIP code clusters of late-diagnosed breast cancer cases exist in Chicago, Illinois. The resulting map indicated if high/low late-stage, breast cancer ZIP codes were clustered around neighboring ZIP codes with similar high/low values, or if a ZIP code with a high/low value was surrounded by ZIP codes with differing values.

Finally, we performed an additional hot spot analysis through calculation of the Getis-Ord Gi* statistic to identify cold and hot spots of clustering on the map, as well as the confidence level

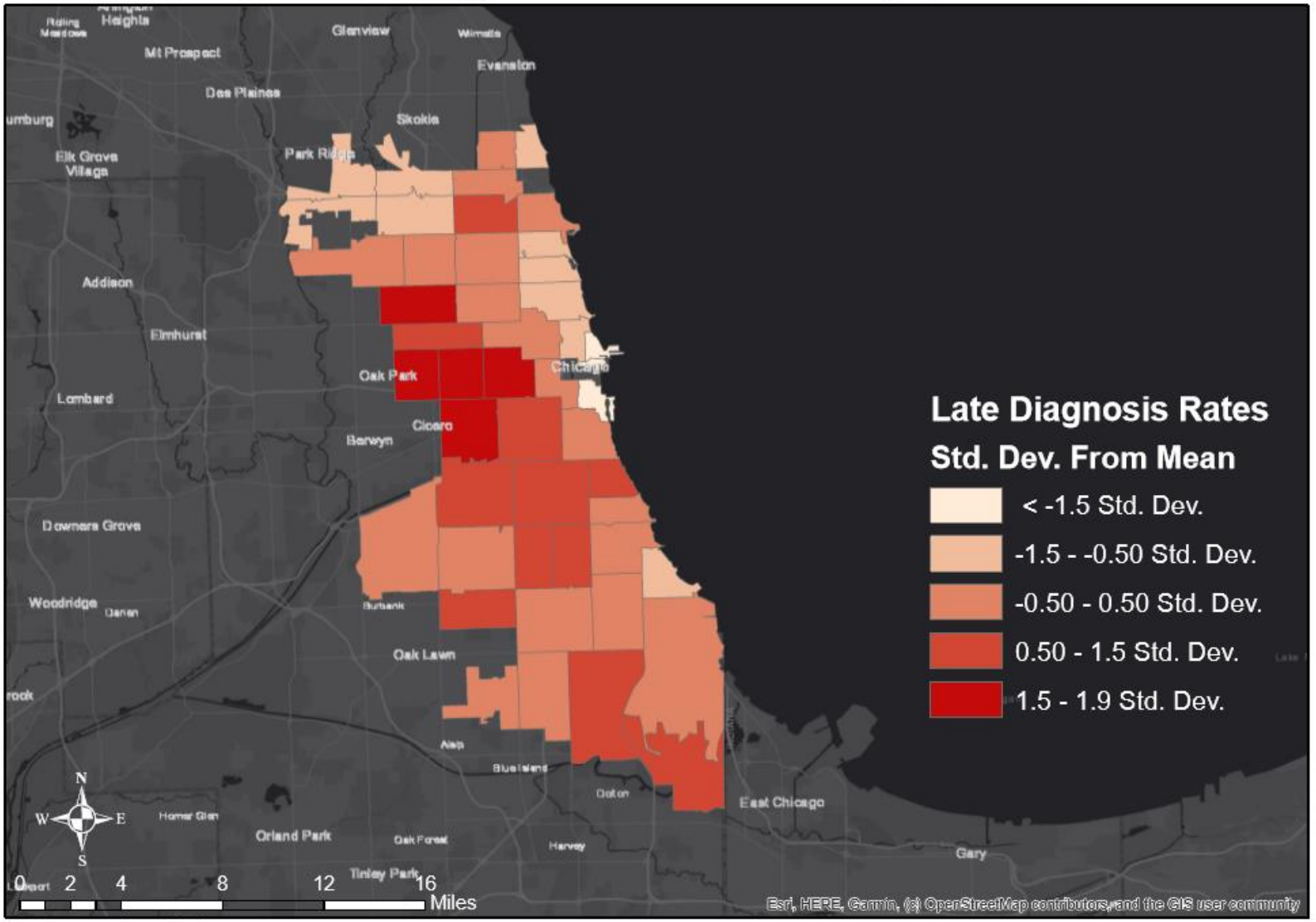
we have in our data set for each hot/cold that was identified (Figure D). The hot spot analysis was based on the results of the local clustering analysis, and utilized the same spatial weight's matrix used in the previous steps. The confidence index was determined based on the P- and Z-score values obtained from the Hot Spot Analysis function, and was index according to each ZIP codes respective GI_Bin value (Figure E).

Findings and Conclusion:

Based on the results of the Local Moran's I analysis in Figure B, it can be determined that there were three areas in Chicago that had significant clusters of ZIP codes based on the percentage of late-diagnosed breast cancer cases. The highest prevalence of late-diagnosed breast cancer cases were found clustered in the western part of north Chicago, while the lowest prevalence of late-diagnosed breast cancer cases were found cluster in the northern-most ZIP code areas of Chicago, and an additional cluster in the eastern ZIP code areas of north Chicago. Figure D shows the results of a GI* analysis that was performed on the breast cancer data, and it can be determined that the results of the Local Moran's I result's and clusters are accurate within an acceptable 95% confidence index and a 5% chance that the clustering was due to chance. I hypothesize that statistically significant ZIP code areas with either high or low prevalence of late-diagnosed breast cancer rates tend to cluster together due to the lack of health resources that are found in the vicinity of the clusters: If there are a lack of hospitals in the western part of north Chicago, where high prevalence of late-diagnosed breast cancer cases were found, then it is likely to negatively impact the neighboring ZIP codes that are in the vicinity of the area. Therefore, I would suggest that the implementation of more accessible and high-quality care clinics that are able to diagnose breast cancer in one ZIP code area that has a high prevalence of late-diagnosed breast cancer rates would produce a positive quantifiable result in reducing late diagnoses not only in the respective ZIP code, but also to the adjacent ZIP code areas that share similar high rates of late diagnosis.

Figure A: Late-Diagnosed B.C.

Late Breast Cancer Diagnosis Rates in Chicago, Illinois, 1986-2008

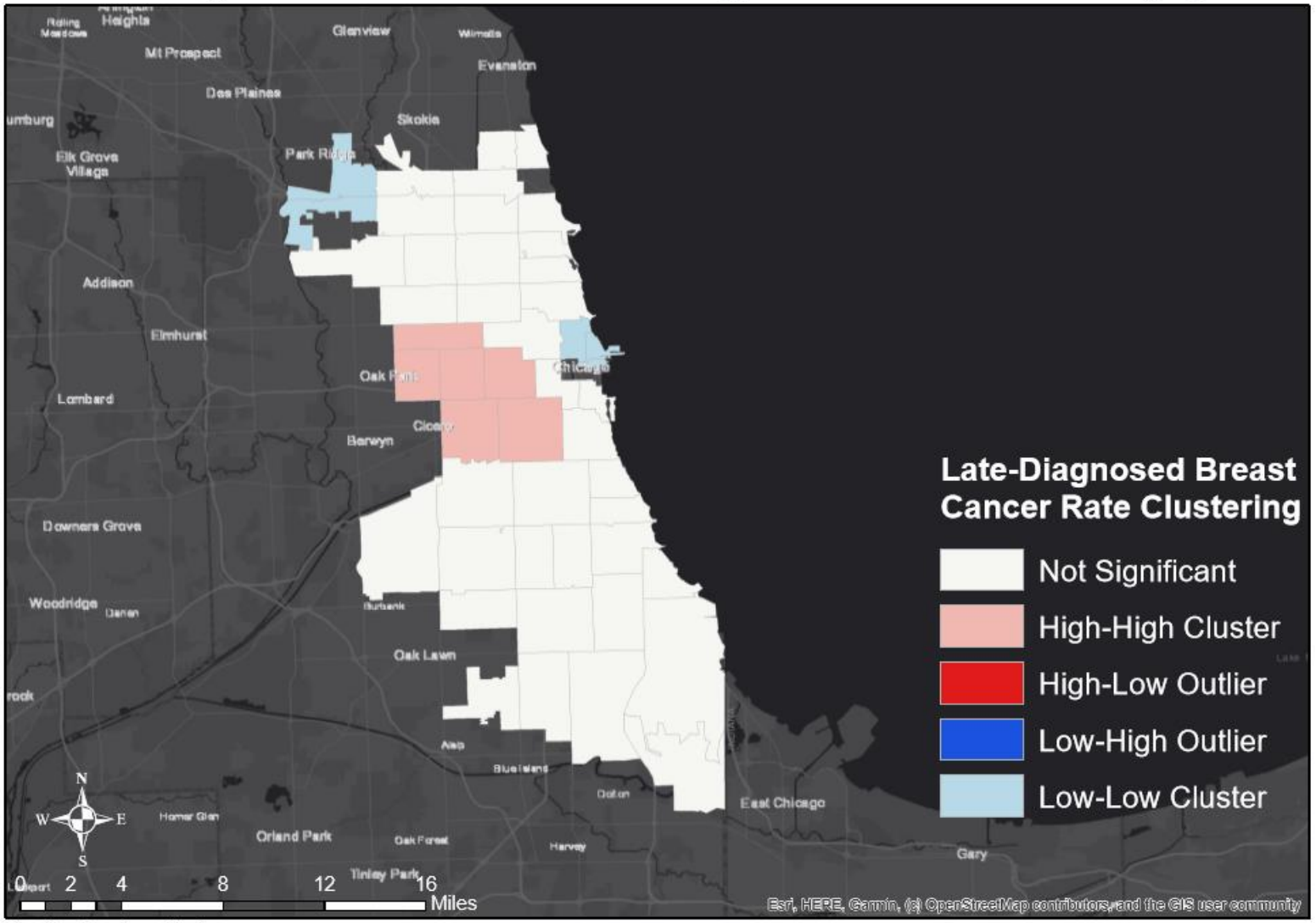


Map Creator: Jason Park
Creation Date: February 23rd, 2020

Projection: Transverse Mercator
Data Source: Illionois Department of Public Health, Illinois State Cancer Registry
Data Date: 1986-2008

Figure B: Local Clustering Analysis

Local Clustering of Late Breast Cancer Diagnosis Rates in Chicago, 1986-2008



Map Creator: Jason Park
Creation Date: February 23rd, 2020

Projection: Transverse Mercator
Data Source: Illionois Department of Public Health, Illinois State Cancer Registry
Data Date: 1986-2008

Figure C: Local Clustering

Local Moran's Analysis of Late Breast Cancer Diagnosis Clustering in Chicago, Illinois, 1986-2008

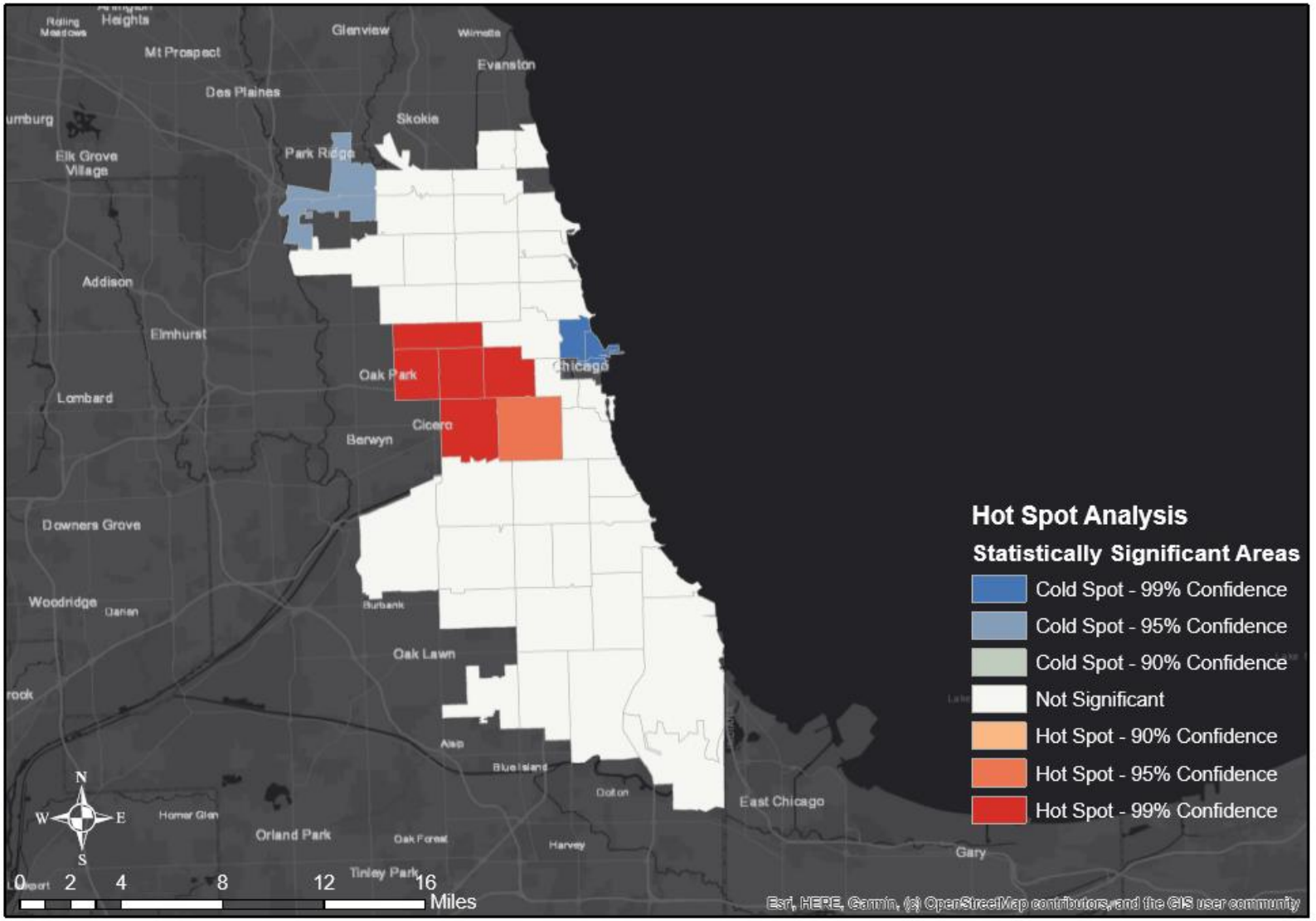
FID	Shape *	ZIPID	LATEBRST	LMiIndex	LMiZScore	LMiPValue	COType	NNeighbors
0	Polygon	231	25.64	3.758367	2.627166	0.002	LL	2
1	Polygon	239	28.33	0.398591	0.329107	0.368		2
2	Polygon	230	40	0.028638	0.357365	0.376		6
3	Polygon	243	47.01	0.554477	1.90334	0.038	HH	6
4	Polygon	261	46.79	0.285887	1.348219	0.08		9
5	Polygon	211	31.03	1.508803	2.540915	0.006	LL	5
6	Polygon	216	27.91	3.524766	2.935088	0.004	LL	2
7	Polygon	221	54.32	1.391921	2.14936	0.018	HH	6
8	Polygon	183	36.76	0.081991	0.244633	0.398		4
9	Polygon	201	34.64	0.369821	0.955379	0.158		5
10	Polygon	276	42.14	0.053901	1.313449	0.098		4
11	Polygon	248	40	-0.021759	-0.244696	0.392		5
12	Polygon	306	43.65	0.070692	0.545704	0.286		4
13	Polygon	184	42.94	-0.015152	-0.1769	0.408		9
14	Polygon	300	39.31	-0.053839	-0.497616	0.296		6
15	Polygon	304	40.72	-0.029469	-0.76424	0.218		7
16	Polygon	285	46.72	0.130403	0.490181	0.32		6
17	Polygon	208	42.86	0.048947	0.730082	0.24		7
18	Polygon	244	52.35	2.028775	3.120998	0.002	HH	5
19	Polygon	222	52.35	1.903897	2.984766	0.002	HH	5
20	Polygon	177	47.1	-0.264921	-0.841251	0.198		7
21	Polygon	156	33.82	0.158869	0.247266	0.4		2
22	Polygon	319	45.79	0.121908	0.541131	0.306		5
23	Polygon	287	39	-0.12716	-0.884327	0.2		6
24	Polygon	173	34.54	0.293991	1.074728	0.154		8
25	Polygon	158	32.47	1.38676	2.22255	0.018	LL	3
26	Polygon	258	46.85	0.36083	1.293474	0.098		6
27	Polygon	337	50	0.526602	0.651914	0.268		2
28	Polygon	186	38.91	-0.016442	-0.124714	0.476		4
29	Polygon	286	46.15	0.242707	0.990986	0.154		6
30	Polygon	284	41.85	0.006255	0.383352	0.338		5
31	Polygon	269	38.77	-0.118053	-0.58038	0.286		3
32	Polygon	204	55	0.549542	0.863695	0.204		5
33	Polygon	176	39.09	-0.037873	-0.204912	0.444		4
34	Polygon	185	44.57	0.127577	0.815835	0.216		6
35	Polygon	321	41.35	-0.005913	-0.382587	0.36		3
36	Polygon	223	55.1	2.489344	2.582379	0.002	HH	3

Map Creator: Jason Park
Creation Date: February 23rd, 2020

Data Source: Illionois Department of Public Health, Illinois State Cancer Registry
Data Date: 1986-2008

Figure D: Hot Spot Analysis

GI* Analysis of Late Breast Cancer Diagnosis Clustering in Chicago, Illinois, 1986-2008



Map Creator: Jason Park
Creation Date: February 23rd, 2020

Projection: Transverse Mercator
Data Source: Illionois Department of Public Health, Illinois State Cancer Registry
Data Date: 1986-2008

Figure E: Hot Spot Analysis

Gi* Analysis of Late Breast Cancer Diagnosis Clustering in Chicago, Illinois, 1986-2008

FID	Shape *	ZIPID	LATEBRST	GiZScore	GiPValue	NNeighbors	Gi_Bin
0	Polygon	231	25.64	-3.34115	0.000834	3	-3
1	Polygon	239	28.33	-1.359448	0.174005	3	0
2	Polygon	230	40	-0.414085	0.678812	7	0
3	Polygon	243	47.01	2.090739	0.036551	7	2
4	Polygon	261	46.79	1.506016	0.132063	10	0
5	Polygon	211	31.03	-2.880456	0.003971	6	-3
6	Polygon	216	27.91	-3.34115	0.000834	3	-3
7	Polygon	221	54.32	2.641263	0.00826	7	3
8	Polygon	183	36.76	-0.547736	0.583873	5	0
9	Polygon	201	34.64	-1.257165	0.208694	6	0
10	Polygon	276	42.14	1.254652	0.209605	5	0
11	Polygon	248	40	0.126113	0.899642	6	0
12	Polygon	306	43.65	0.592145	0.553753	5	0
13	Polygon	184	42.94	-0.177963	0.858752	10	0
14	Polygon	300	39.31	0.29553	0.767589	7	0
15	Polygon	304	40.72	0.651074	0.514999	8	0
16	Polygon	285	46.72	0.737894	0.460579	7	0
17	Polygon	208	42.86	0.790357	0.429319	8	0
18	Polygon	244	52.35	3.587058	0.000334	6	3
19	Polygon	222	52.35	3.406895	0.000657	6	3
20	Polygon	177	47.1	-0.621989	0.533949	8	0
21	Polygon	156	33.82	-0.81959	0.41245	3	0
22	Polygon	319	45.79	0.706907	0.479624	6	0
23	Polygon	287	39	0.732171	0.464064	7	0
24	Polygon	173	34.54	-1.250087	0.211268	9	0
25	Polygon	158	32.47	-2.37454	0.017571	4	-2
26	Polygon	258	46.85	1.495579	0.134763	7	0
27	Polygon	337	50	1.236374	0.21632	3	0
28	Polygon	186	38.91	-0.089535	0.928656	5	0
29	Polygon	286	46.15	1.187126	0.235178	7	0
30	Polygon	284	41.85	0.334369	0.738101	6	0
31	Polygon	269	38.77	0.274529	0.783678	4	0
32	Polygon	204	55	1.458702	0.144647	6	0
33	Polygon	176	39.09	0.046669	0.962777	5	0
34	Polygon	185	44.57	0.911864	0.36184	7	0
35	Polygon	321	41.35	0.332991	0.739141	4	0
36	Polygon	223	55.1	3.046369	0.002316	4	3
37	Polygon	157	38.46	-0.81959	0.41245	3	0
38	Polygon	161	33.33	-1.585121	0.112939	6	0

Map Creator: Jason Park
Creation Date: February 23rd, 2020

Data Source: Illionois Department of Public Health, Illinois State Cancer Registry
Data Date: 1986-2008