# Tutorial: Exploratory Data Analysis with GeoDa

*D G Rossiter Cornell University, Section of Soil & Crop Sciences* 

March 7, 2024

# Contents

1	Introduction	1
2	Dataset	2
3	Exploratory Data Analysis3.1Univariate3.2Bivariate3.3Multivariate	<b>3</b> 3 4 6
4	Clustering4.1k-means4.2Geographically-compact k-means4.3k-medians4.4k-mediods	9 9 11 13 13
5	Neighbors and Distances	15
6	Assessing Global Spatial Autocorrelation	16
7	Assessing Local Spatial Autocorrelation	17
8	Spatial Regression8.1Spatial Error model8.2Spatial Lag model	<b>19</b> 20 22

Version 1.1 Copyright © 2019-2021, 2024 D G Rossiter All rights reserved. Reproduction and dissemination of the work as a whole (not parts) freely permitted if this original copyright notice is included. Sale or placement on a web site where payment must be made to access this document is strictly prohibited. To adapt or translate please contact the author (d.g.rossiter@cornell.edu).

This exercise is an adaptation and extension of one prepared by Dr. Diana Sinton for Cornell course PLSCS/NTRES 6200 in 2016.

9 Finishing with GeoDa	23
References	24

# 1 Introduction

GeoDa is an open-source program, cross-platform program designed as a simple tool for exploratory spatial data analysis (ESDA) and some spatial modelling of **spatial polygon** data, that is, maps of polygon units such as census tracts or political divisions with a set of **attributes** measured on each one.

GeoDa was first developed at Arizona State University and is now hosted at the University of Chicago<sup>1</sup>. The GeoDa program, documentation and sample data is freely available for download from the Geodata Center's GitHub<sup>2</sup>. A recent journal paper [1] explains the history of GeoDa and its applications.

GeoDa allows users to experiment with visualization functionality such as linking and brushing across windows. This can be very helpful both for interpretation of and communication about these spatial patterns. It also incorporates several spatial statistical models.

Таяк 1 : Download and install GeoDa.

**TASK 2** : Install example datasets in a convenient location.

٠

GeoDa comes with some sample data, but many GIS file formats<sup>3</sup>, including ESRI Shapefiles with associated databases, can be used. This tutorial should come with two compressed files (1) NY\_data.zip and a subset of this, (2) Syr.zip. Uncompress these to find the Shapefiles, with extension .shp.

This is the "New York leukemia dataset" developed by Waller and Gotway [3] and adapted by Bivand et al. [2]. It contains information on the census tracts (the spatial units) in an eight county area including Syracuse (NY) city, relating possible causes to the incidence of leukemia, in particular, exposure to the industrial chemical TCE<sup>4</sup>.

**Note:** In the USA census tracts have 1 500–8 000 people (optimum size 4 000). They are designed to be socio-economically and demographically fairly homogeneous. Each tract has several block groups; these are made up of 20–40 individual blocks. The tract is usually large enough to compile reliable statistics.<sup>5</sup>.

I have reduced this to just Syracuse city in the second example to reduce the size of maps and graphs for this tutorial, but you may prefer to work with the full dataset, which includes both urban and rural areas.

#### Таяк 3 : Start GeoDa.

<sup>&</sup>lt;sup>1</sup> https://spatial.uchicago.edu/geoda

<sup>&</sup>lt;sup>2</sup> http://geodacenter.github.io

<sup>&</sup>lt;sup>3</sup> https://geodacenter.github.io/formats.html

<sup>&</sup>lt;sup>4</sup> Trichloroethylene, an industrial solvent often found in groundwater

<sup>&</sup>lt;sup>5</sup>https://www2.census.gov/geo/pdfs/education/CensusTracts.pdf

# 2 Dataset

TASK 4 : Load the Syracuse leukemia incidence dataset into GeoDa,using the File | New ... menu item or the file open icon. This is ashapefile with base name Syr, so select Syr.shp

You will see a plain map of the polygons (Fig. 1).



Figure 1: Base map and GeoDa tool bar

**TASK 5** : Open the data table by clicking the Table icon. Examine the rows and columns.

We will work with these variables (fields):

These are all reported on the basis of 1980 census tracts. First, the **response** (dependent) variables:

- Cases : the number of leukemia cases 1978–1982; some cases had insufficient georeference, these were added proportionally to tracts, so some "counts" are not integers.
  - Z : log-transformed rate, i.e., normalized by census tract population:  $Z_i = \log(1000[\text{Cases} + 1]/n)$ , where *n* is the population of the tract.

#### Second, possible **predictors**:

- PEXPOSURE : potential exposure, computed as the logarithm of 100 times the inverse of the distance between a census tract centroid and the nearest TCE-producing site;
- PCTAGE65P : percent older than 65 years; this could represent long-term exposure to any environmental factor;
- PCTOWNHOME : percent home ownership; this could indicate lifestyle or economic level.

**TASK 6** : Rearrange the Table and the Map so that you can view both. •

The basic GIS "linking" functionality is in place; you can click on polygons in the map and their associated records in the table will highlight, and vice versa. To unselect objects, click anywhere in the white area surrounding the map or at the upper-left of the table. You can select multiple polygons (on the map) or tracts (in the attribute table) with Shift-click for a set or Ctrl-click to add one-by-one. You can also "brush" over the map by holding down the left mouse button, to select in a window,

Q1: Click on the northeaternermost census tract. What is its AREAKEY? What is its population? What percent of its homes are owned rather than rented?



#### **3** Exploratory Data Analysis

# 3.1 Univariate

**TASK 7**: Display some themed maps in the Map menu, for one or more of the variables, for example PCTSGE65P.

Compare quantile, percentile, box, and natural breaks maps. Examine how they present the same theme in different ways (Fig. 2).

**Q2** : Look at the southeasternmost census tract in these four maps. How do they describe its proportion of older residents, compared to the entire City? Which map(s) best show(s) whether it is unusual?

**Q3**: Which map is best for assessing spatial autocorrelation of this variable? Why? Does there appear to be autocorrelation? Across how many neighbouring census tracts?



Figure 2: Thematic maps

# 3.2 Bivariate

Now we explore some feature-space plots.

**TASK 8**: Under the Explore menu, create a Histogram of PCTAGE65P.Also create a Scatter Plot of the proportion of residents over the age of65 PCTAGE65P (Y variable) and the proportion of homes that are ownedrather than rented. PCTOWNHOME (X variable)•See Figure 3.

**Q4**: *Which tract has the highest proportion of older residents?* 

**Q5** : Describe the relation between these two attributes.

**TASK 9**: Find an unusual tract (not fitting the overall pattern for the<br/>city) and click on its point in the scatterplot.

See Figure 4.

Because all of the individual graphic elements for each of the 63 polygons are linked, any time that one or more are selected in one window







Figure 4: An unusual tract

or in one of the exploration plots, their linked highlighted display will activate in all other windows.

Q6: Which tract did you select? Where is it located?

**TASK 10**: Brush over the southern few tracts by holding down the left mouse button as you define a rectangular window.

**Q7**: What is the overall relation between home ownership and proportion of over-65 residents? What is this relation for the four southernmost tracts? How do you explain this?

Another interesting plot is the Cartogram.

**TASK 11**: Make a cartogram (Map | Cartogram) of the proportion over 65 years old PCTAGE65P as the circle *size*, with the disease incidence Z as the circle *colour*.

See Figure 5.



Figure 5: Cartogram of leukemia incidence vs. older residents

**Q8** : *How are the circles placed in the plot? What insight does this give you into the relation between disease incidence and older residents?* •

# 3.3 Multivariate

**TASK 12**: Open a scatterplot matrix (Explore | Scatter Plot Matrix of the three PCTOWNHOME, PCTAGE65P, and PEXPOSURE, as well as the response variable Z.

See Figure 6.

**Q9**: Describe the feature-space distributions of the four variables. Looking at the proposed bivariate linear regressions, which have tracts with high leverage, i.e., that greatly influence the line?

**TASK 13** : In the matrix, select the histogram bar for the lowest proportion of home ownership, i.e., where more households rent.

Note how the linked maps highlight these tracts. See Figure 7.

**TASK 14**: With the scatterplot matrix displayed, select menu option Options | View | Regime Regression. This will then show the separate regression lines and statistics for the overall, selected, and non-selected census tracts.

Q10: Look at the red proposed regression lines - which bivariate corre-



Figure 6: Scatterplot matrix

*lations are substantially different from the overall correlations if we only consider these tracts?* •

**TASK 15**: Open a Parallel Coordinate Plot (PCP) and Include the three possible predictors PCTOWNHOME, PCTAGE65P, and PEXPOSURE, as well as the response variable Z.

Q11: What is their overall inter-relation?

**TASK 16** : Click on the line to the highest response. See Figure 8.

Q12 : Which tract is this? How is this response related to the three predictors?



Figure 7: Scatterplot matrix with low home ownership tracts selected





# 4 Clustering

We may want to see which areas go together in **clusters**, based on one or more factors. These are groupings with *minimum* **pooled within-class** differences, and thus *maximum* **between-class** differences. These variances can be **parametric**, i.e., based on **means**, or **non-parametric**, i.e., based on **medians**.

#### 4.1 k-means

One popular method is **k-means**, with a user-specific number of clusters *k*. There are many methods to find an "optimum" number of clusters, but we will not investigate those here, rather, we will decide on a number. This could be, for example, to stratify the area for a survey or marketing campaign.

We begin with clustering based on a single variable, home ownership.

**TASK 17**: Select menu item Clusters | K means and in the dialog box select PCTOWNHOME and 4 clusters. Run the analysis. See Figure 9. This should produce the map in Figure 10 •

Select Variables     X   Y   POP8   TRACTCAS   PROPCAS   PCTOWNHOME   PCTASEESP   Z   AVGIDIST   PEXPOSURE   Cases   vm   Use geometric centroids   Auto Weighting:   0   1   Ymmetrace    Variance    Variance    Variance    Variance    Variance    Variance    Variance    Variance    Var			Weens Oracenny Settings
Subject versions   Subject versions Subject versions Subject versions POP8 TRACTCAS PROPCAS PROPCAS PCTAGEGSP Z Available Value Value Use geometric centroids Auto Weighting: Value	ut:	alaat Variablaa	Summary
<pre>Y POP8 POP6  TRACTCAS POPCAS PCTOWNHOME PCTAGE666 Z AVGIDIST PCTOWNEOME Classes Ym Use geometric centroids Auto Weighting Weighting: 0 1 1 PCTOWNEOME Clusters: 4 Weighting: 0 1 1 PCTOWNEOME PCTOURE Cases Ym Use geometric centroids Auto Weighting It is is individue to the sum of squares: 62 Weighting: 0 1 1 PCTOWNEOME PCTOURE Cases Ym Use geometric centroids Auto Weighting It is is individue to the sum of squares: 4.81827 The total sum of squares: 57.1817 The between-cluster sum of squares: 57.1817 The between-cluster sum of squares: 57.1817 The total vithin-cluster sum of squares: 0.92286</pre>	x	elect variables	
POP8 TRACTCAS PROPCAS PROPCAS PCTOWNHOME PCTACEGESP Z AGOIDST PEXPOSURE Cases Variant Centroids Auto Weighting Weighting: 0 1 1 rameters: Number of Clusters: 4 Winimum Bound: Transformation: Standardize (Z) Initialization Re-runs: 52 Within cluster s.6. (ci   0.420748) (ci   0.762289   The total sum of squares: 62 Within cluster s.6. (ci   0.99039 (ci   0.89355 (ci   0.893555 (ci   0	Ý		Method: KMeans
TRACTCAS PROPCAS PCTOWNHOME PCTAGE65P Z AVGIDIST PEXPOSURE Cases v	POP8		Number of clusters: 4 Initialization method: KMeans++
PROPCAS         PCTAGEGSP         Z         AVGIDIST         PEXPOSURE         Cases         vm         Use geometric centroids         Auto Weighting:         Weighting:         Weighting:         Transformation:         Standardize (Z)         Initialization Method:         KMeans++         Initialization Re-runs:         1000         Distance Function:         Euclidean         utput:         ave Cluster in Field:         Run         Close	TRACTCAS		Initialization re-runs: 150
PCTOWNHOME         PCTAGES5P         Z         AVGIDIST         PEXPOSURE         Cases         vm         Weighting: 0         1         arameters:         Number of Clusters: 4         Minimum Bound:         Imitialization Re-runs:         1000         Distance Function:         Euclidean         Cluster in Field:         Run         Cluster in Field:         Run	PROPCAS		Maximum iterations: 1000 Transformation: Standardize (Z)
PCTAGE65P   Z   AVGIDIST   PEXPOSURE   Cases   vm   Use geometric centroids Auto Weighting   Weighting:   0   1   arameters:   Number of Clusters:   4   Intimum Bound:   X   1   0%   Transformation:   Standardize (Z)   Initialization Method:   KMeans++   Initialization Re-runs:   1000   Distance Function:   utput:   ave Cluster in Field:   CL   Run   Close   Cluster in Field:	PCTOWNHOME		Distance function: Euclidean
Z AVGIDIST PEXPOSURE Cases Ym Use geometric centroids Auto Weighting Weighting: 0 1 1 arameters: Number of Clusters: 4 Minimum Bound: 10% Transformation: Standardize (Z) Initialization Method: KMeans++ Initialization Method: KMeans++ Initialization Method: KMeans++ Initialization Method: CL Run Close	PCTAGE65P		Cluster centers:
AVGIDIST PEXPOSURE Cases ymm Use geometric centroids Auto Weighting Weighting: 0 1 1 arameters: Number of Clusters: 4 Minimum Bound: Transformation: Italiazation Method: Kmeans++ Initialization Re-runs: 1000 Distance Function: Euclidean utput: ave Cluster in Field: CL Run Close	Z		PCTOWNHOME
PEXPOSURE         Cases         Vm         Use geometric centroids         Auto Weighting:         0         arameters:         Number of Clusters:         4         Minimum Bound:         1         10%         Transformation:         Standardize (Z)         Initialization Re-runs:         150         Use specified seed:         Change Seed         Maximum Iterations:         1000         Distance Function:         Euclidean         utput:         ave Cluster in Field:         Run       Close	AVGIDIST		C1 0.462784
Cases         Vm         Use geometric centroids Auto Weighting         Weighting: 0         Weighting: 0         arameters:         Number of Clusters: 4         Minimum Bound:         Xmmodel         10%         Transformation:         Standardize (Z)         Initialization Method:         KMeans++         Initialization Re-runs:         100         Distance Function:         tuput:         ave Cluster in Field: CL         Run         Close	PEXPOSURE		C2 0.0722619
Transformation: Standardize (Z) Initialization Method: KMeans++ Initialization Re-runs: 150 Use specified seed: Change Seed Maximum Iterations: 1000 Distance Function: Euclidean	Jases		C4 0.762289
Use geometric centroids       Auto Weighting         Weighting: 0       1         arameters:       1         Number of Clusters:       4         Iminum Bound:       1         Initialization Method:       KMeans++         Initialization Method:       KMeans++         Initialization Method:       Change Seed         Maximum Iterations:       1000         Distance Function:       Euclidean         utput:       ave Cluster in Field:         Run       Close			The total sum of squares: 62
Weighting:       Image: Constraint of the second seco	Use geometric cent	troids Auto Weighting	Within-cluster sum of squares:
arameters: Number of Clusters: 4 Minimum Bound: X 0 1 Transformation: Standardize (Z) Initialization Method: KMeans++ Initialization Re-runs: 150 Use specified seed: Change Seed Maximum Iterations: 1000 Distance Function: Euclidean ♀ utput: ve Cluster In Field: CL Run Close	Weighting: 0	011	Within cluster S.S.
rameters: Number of Clusters: 4 Winimum Bound: Transformation: Standardize (Z) initialization Method: KMeans++ initialization Re-runs: 150 Use specified seed: Change Seed Maximum Iterations: 1000 Distance Function: Euclidean Autput: Run Close			C1 0.990039
Number of Clusters:       4         Minimum Bound:       X       1         Iminimum Bound:       10%         Transformation:       Standardize (Z)         Initialization Method:       KMeans++         Initialization Re-runs:       150         Use specified seed:       Change Seed         Maximum Iterations:       1000         Distance Function:       Euclidean         utput:       .         ave Cluster in Field:       CL         Run       Close	rameters:		C3 0.819022
Minimum Bound: X   1   1   1   1   1   1   1   1   1	lumbar of Clustera	4	C4 2.11564
Winimum Bound:       X       1         10%       1         Transformation:       Standardize (Z)         initialization Method:       KMeans++         150       Jse specified seed:         Jaskimum Iterations:       1000         Distance Function:       Euclidean         ttput:       ve Cluster in Field:         Run       Close	amber of Clusters:	4	The total within-cluster sum of squares: 4.81827
Iransformation: Standardize (2) initialization Method: KMeans++ initialization Re-runs: 150 Jse specified seed:  Change Seed Aximum Iterations: 1000 Distance Function: Euclidean  ttput: ve Cluster in Field: CL  Run Close	/inimum Bound:	□ <u>X</u> ↓ 1	The ratio of between to total sum of squares:
Transformation: Standardize (Z) Initialization Method: KMeans++ Initialization Re-runs: 150 Use specified seed: Change Seed Maximum Iterations: 1000 Distance Function: Euclidean utput: ave Cluster in Field: CL Run Close		10%	0.922286
Initialization Method: KMeans++	ransformation:	Standardize (Z)	
Initialization Re-runs: 150 Use specified seed: Change Seed Maximum Iterations: 1000 Distance Function: Euclidean C utput: ave Cluster in Field: CL Run Close	nitialization Method:	KMeans++	
Use specified seed: Change Seed Maximum Iterations: 1000 Distance Function: Euclidean utput: ave Cluster in Field: CL Run Close	nitialization Re-runs:	150	
Maximum Iterations: 1000 Distance Function: Euclidean O utput: ave Cluster in Field: CL Run Close	Jse specified seed:	Change Seed	
Distance Function: Euclidean	Aaximum Iterations:	1000	
utput: ave Cluster in Field: CL Run Close	Distance Function:	Euclidean	0
utput: vve Cluster in Field: CL Run Close			
Run Close	tput:		
Run Close	Olympic Fields -		
Run Close		<i>،</i> ۲	
Run Close	ve cluster in rield.		
	ve cluster in rield.		
	Ru	un Close	
	Ru	un Close	

Figure 9: Creating a k-means cluster



Figure 10: Four k-means clusters

**Q13** : Describe the spatial pattern of the clusters. How is home ownership distributed across Syracuse? •

**Q14**: Do the clusters have the same number of census districts? Should they?

**Q15** : *How much of the total sum of squares of the clustering variable PCTOWNHOME is explained by the clustering?* •

**Q16**: *Which cluster is most homogeneous in feature space, i.e., has the least within-cluster sum of squares? The most?* •

**Q17** : The cluster centers are the mean of the target variable for the census blocks within the cluster. How different are they? •

# 4.2 Geographically-compact k-means

What if we required **geographically-compact** clusters? For example, most voting district laws require at least contiguous areas, and some prefer as compact as possible.

**TASK 18**: Repeat the k-means clustering for four clusters, but select the "Use geometric centroids" option, with **auto-weighting**. This forces the clusters to be geographially-contiguous, while giving the least weight possible to that aspect, keeping as much feature-space homogeneity as possible. See Figure 11. This should produce the map in Figure 12.

iput:		Summary
Se	elect Variables	Cummury
X		Method: KMeans
1		Number of clusters: 4
POP8		Initialization method: KMeans++
TRACTCAS		Maximum iterations: 1000
PROPCAS		Transformation: Standardize (Z)
PCTOWNHOME		Use geometric centroids (weighting):
PCTAGE65P		Centroid (X) 0.375
Z		Centroid (Y) 0.375
AVGIDIST		FCIRGEOSF 0.25
PEXPOSURE		Cluster centers:
Cases		PCTAGE05P
Vm		C1 0.108704
<b>1</b>		
Use geometric cent	Auto Weighting	C4 0.30139
Weighting: 0	1 0.75C	The total and of animate (2)
		Within-cluster sum of squares: 62
arameters:		Within cluster S.S.
Number of Objections	4	C1 4.86223
Number of Clusters:	4	C2 7.40225
Minimum Bound:	□ X ↓ 1	C3 18.5696
		11
Transformation:	Standardize (Z)	The total within-cluster sum of squares: 36.914 The between-cluster sum of squares: 25.086
Initialization Method:	KMeans++	The ratio of between to total sum of squares:
Initialization Re-runs:	150	
Use specified seed:	Change Seed	
Maximum Iterations:	1000	
Distance Function:	Euclidean	$\diamond$
utput:	4	
utput: ave Cluster in Field: C	،L	
utput: ave Cluster in Field: C		
ave Cluster in Field: C		
ave Cluster in Field: C	in Close	
ave Cluster in Field: C	in Close	

Figure 11: Creating a compact k-means cluster

**Q18** : How much of the total sum of squares of the clustering variable *PCTOWNHOME* is explained by the clustering? Why is this so much lower than for the k-means without forcing compact clusters? •

**Q19**: *Which cluster is most homogeneous in feature space, i.e., has the least within-cluster sum of squares? The most?* •



Figure 12: Four k-means compact clusters

**Q20** : The cluster centers are the mean of the target variable for the census blocks within the cluster. How different are they? •

**Q21**: *Why might a policy maker prefer the compact clusters?* •

**Optional:** You can balance compactness in geographic and feature space by changing the weighting given to the geometric centroids.

# 4.3 k-medians

**Optional:** You can used the **medians** of the clusters, rather than the means, to define the within- and between-class sum-of-squares. That is, rather than calculating the mean for each cluster to determine its centroid, instead calculate the median. This is **k-medians**. Run a 4-cluster k-median analysis and compare with the 4-cluster k-means.

# 4.4 k-mediods

**Optional:** You can used the **mediods** of the clusters, rather than the means, to define the within- and between-class sum-of-squares. A **mediod** is one of the observations, also called the **exemplar** of the cluster. This allows the identification of the most "typical" observation in a cluster. This is **k-mediods**. Run a 4-cluster k-mediod analysis and compare with the 4-cluster k-means.

See Figure 13. Here the cluster center values are actual values of one of the census blocks in the cluster. This should produce the map in Figure 14; notice the exemplars.

•	K	dolds Clustering Settings	
ut:			
S	elect Variables		Summary
OPP8 RACTCAS ROPCAS CTOWNHOME CTAGEG5P WGIDIST EXPROSURE cases T Use geometric cent	troids Auto Weighting	Method: KMedoids Number of clusters: 4 Initialization method: Initialization re-runs Maximum iterations: 10 Transformation: 5tands Distance function: Ma Cluster centers: (medo 	Random : 150 Octize (Z) shattan id) es: 62 squares: -
Weighting: 0 ameters: lumber of Clusters: linimum Bound:		Cl 2.04148 C2 0.582015 C3 2.11564 C4 0.183385 The total within-clust The between-cluster su The ratio of between t	er sum of squares: 4.92252 n of squares: 57.0775 o total sum of squares:
ransformation: hitialization Re-runs: Ise specified seed: faximum Iterations: histance Function:	10% Standardize (Z) 150 Change Seed 1000 Manhattan	0.920605	
out: e Cluster in Field: C	Close		

Figure 13: Creating a k-mediods cluster



Figure 14: Four k-mediods clusters, exemplars, i.e., "typical" census blocks, highlighted

#### 5 Neighbors and Distances

For spatial models, we must impose some **spatial structure** on the 63 polygons, that is, how they are related in space. Then we can assess this statistically.

One way is by **distance between polygon centroids**, as in point geostatistics; the spatial weights are based on separation, typically as inverse distance. This considers that distance is the only factor driving any spatial correlation.

However, there are other ways to build a **weights matrix** that relates neighbours; these relate to different hypotheses about how space affects the response. For example, a binary neighbours weighting considers that all first-order neighbours contribute equally to any spatial effect, i.e., it is averaged across the neighbours. With this weighting every tract is influenced equally (on average) by neighbours, and this influence is divded among the neighbours.

**TASK 19**: Generate two weights files: (1) Distance Weights and indicate X and Y coordinates, (2) order-1 Contiguity with Queen neighbours (i.e., tracts meeting only at a point are also considered to be neighbours).

To generate a Weights File, choose Tools | Weights Manager | Create. Every shape must have its own unique ID, so check the Add ID Variable and use the existing AREAKEY variable. By default, the new ID variable will be named POLY\_ID, or you can choose otherwise. You can then select a type of weighting methods

For the distance weighting, the Threshold distance will automatically be calculated at the minimum distance to ensure that every polygon has at least one neighbor, but you can set any distance you desire. The distance units will be the units associated with the shapefile. For example, if we think that the phenomenon might be spatially-correlated (after accounting for the feature-space regression) to 2.4 km, set the threshold distance to 2400 m.

See Figure 15.

**Q22**: *What is the maximum number of neighbours considered for any tract in this distance weighting?* •

**Note:** You could also use the k-Nearest Neighbors option to specific a set number of its closest neighbors that you wish each polygon to use.

When your choices are set, Create the file and name the file with a label indicating the approach used to calculate the neighbors. For example, Queen1.gal would indicate a Queen directionality with 1 order of contiguity; Dist24.gwt would indicate a 2400 m radius inverse-distance weighting.

As you are deciding which method derives the most valid weights file for your question of interest, you can visually compare the results by using

🕐 Weights Manager — 🗆 🗙	Weights File Creation X	😵 Connectivity Histogram - SyrDist24k	- 🗆 X
Create Load Remove	Weights File ID Variable AREAKEY  V Add ID Variable Contiguity Weight	°]	
Weights Name SyrDist SyrDist24k	O Queen contiguity     Order of contiguity     1       ○ Rook contiguity     Include lower orders       ○ Precision threshold     0	<b>₹</b> -	
Property         Value           type         threshold           symmetry         symmetric           file         SyrDist24K.gwt           id variable         AREAKEY           distance metric         Euclidean           distance vars         centroids           distance unit         Meter           threshold value         2400	Distance Weight Distance metric Luclidean Distance X-coordinate variable <k-centroids> V-coordinate variable <y-centroids> © Threshold distance 2400 K-Nearest Neighbors Number of neighbors</y-centroids></k-centroids>		24 27
Histogram Connectivity Map	Create Close	Number of Neighbors	

Figure 15: Creating a distance weighting

the **Connectivity** Histogram button in the Weights Manager window, each time after you select your different weights tables. See Figure 16.



Figure 16: Connectivity map and histogram, Queen lag-1 neigbours

**Q23**: What is the most common number of neighbours using the Queen *lag-1* neighbours?

# 6 Assessing Global Spatial Autocorrelation

Here we evaluate whether the rates of leukemia across the study area may be spatially auto-correlated, without considering any predictors.

**TASK 20**: Use Space | Univariate Moran's I, with Z as the variable, to produce a global Moran's I plot. Do this for all the weighting schemes you defined.

See Figure 17.



Figure 17: Global Moran's I

**Q24** : Do the weighting schemes all give the same value of global Moran's I? If not, which implies stronger spatial correlation? Why? •

**TASK 21**: Open two themed maps: decile (10-quantile) of Z (incidence) and PEXPOSURE (exposure).

In the Moran's I scatterplot, click on the highest positive Z (incidence) and highest weighted lag Z.

See Figure 18.

**Q25**: *Where is this tract located? Does it also have a high exposure? Do its neighbours have high incidences? Do they have high exposures?* •

# 7 Assessing Local Spatial Autocorrelation

**TASK 22** : Examine where in the map are the hotspots of local autocorrelation.

First, make sure that your desired Weights file is set as the default, i.e., highlighted in the Weights Manager window.

**TASK 23**: Use Space | Univariate Local Moran's I, with Z as the variable, to produce a local Moran's I plot. Do this for all the weighting schemes you defined. Generate two output windows: the Significance Map and the Cluster Map.

The Significance map shows where there are leukemia values that are statistically significantly higher **or** lower than the neighboring values would have predicted. With the Cluster Map, you can see where the higherthan-expected and lower-then-expected values vary.

See Figure 19.



Figure 18: Global Moran's I



Figure 19: Local Moran's I significance and clusters

Q26 : Which areas of the city are clusters of high leukemia incidence? Are there any tracts that have high incidence, but are surrounded by tracts with low incidence? Another way to find hot spots is with Geary's G or G\*; if you want you can experiment with these.

## 8 Spatial Regression

Here we try to find the covariates ("predictors") correlated (which maybe cause) leukemia. Of course, we can do this non-spatially, i.e., all in attribute space, not taking spatial relations into account.

**TASK 24** : Compute a multivariate linear regression model of leukemia incidence (response) as predicted by the three possible causitive factors (predictors). This is with the Regression menu item. Select Z as the dependent variable, and PCTOWNHOME, PCTAGE65P, and PEXPOSURE as the covariates. This is the Classic linear model, i.e., not taking spatial correlation into account.

Non-spatial linear model:  $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ 

Figure 20 shows how to specify the regression; Figure 21 shows the results.



Figure 20: Specifying a "classic" linear regression

**Q27** : What is the adjusted R<sup>2</sup> of this model? What are the signs of the slopes for each predictor? What is the interpretation? Which (if any) predictors are significantly different from zero?

The model summary shows many problems with the linear model:

1. The **multicolilinearity** (or multiple) **condition number** represents the sensitivity of the model to small changes in the design matrix, i.e., the values of the covariables. A high value (often > 30) indicates high colinearity in one or more predictors; here we see that is the case.

		Regression Repo	rt	
2				
>>03/07/2019 1 REGRESSION	7:18:29			
SUMMARY OF OUT	PUT: ORDINARY	LEAST SQUARES ESTIM	ATION	
Data set	: Syr			
Dependent Vari	able :	Z Number of Ob	servations: 63	
Mean dependent	var : 0.03	77522 Number of Va	riables : 4	
S.D. dependent	var : 0.9	96518 Degrees of F	reedom : 59	
R-squared	: 0.1	.85475 F-statistic	: 4	.47829
Adjusted R-squ	ared : 0.1	.44059 Prob(F-stati	stic) : 0.00	671609
Sum squared re	sidual: 50	.9583 Log likeliho	od : -8	2.7112
Sigma-square	: 0.8	63701 Akaike info	criterion : 1	73.422
S.E. of regres	sion : 0.9	29355 Schwarz crit	erion : 1	81.995
Sigma-square M	L : 0.8	08863		
S.E of regress	ion ML: 0.8	99368		
 Variabl	e Coeffic	ient Std.Error	t-Statistic	Probability
CONS	TANT -3.1	.5559 2.16024	-1.46076	0.14939
PEXPO	SURE 2.6	4063 2.12602	1.24206	0.21913
PCTOWN	HOME -0.30	0.47429	-0.649259	0.51869
PCTOWN PCTAG	HOME -0.30 E65P 4.2	0.47429 4105 1.22995	-0.649259 3.44815	0.51869 0.00105
PCTOWN PCTAG	HOME -0.30 E65P 4.2	0.47429 4105 1.22995	-0.649259 3.44815	0.51869 0.00105
PCTOWN PCTAG	HOME -0.30 E65P 4.2 	0.47429 4105 1.22995	-0.649259 3.44815	0.51869 0.00105
PCTOWN PCTAG  REGRESSION DIA MULTICOLLINEAR	HOME -0.30 E65P 4.2  GNOSTICS ITY CONDITION	17937 0.47429 14105 1.22995	-0.649259 3.44815	0.51869 0.00105
PCTOWN PCTAG REGRESSION DIA MULTICOLLINEAR TEST ON NORMAL	HOME -0.30 E65P 4.2  GNOSTICS ITY CONDITION ITY OF ERRORS	NUMBER 48.997744	-0.649259 3.44815	0.51869 0.00105
PCTOWN PCTAG REGRESSION DIA MULTICOLLINEAR TEST ON NORMAL TEST	HOME -0.30 E65P 4.2  GNOSTICS ITY CONDITION ITY OF ERRORS DF	17937 0.47429 14105 1.22995 	-0.649259 3.44815 	0.51869 0.00105
PCTOWN PCTAG REGRESSION DIA MULTICOLLINEAR TEST ON NORMAL TEST Jarque-Bera	HOME -0.30 E65P 4.2 	17937 0.47429 4105 1.22995 	-0.649259 3.44815 PROB 0.00000	0.51869 0.00105
PCTOWN PCTAG REGRESSION DIA MULTICOLLINEAR TEST ON NORMAL TEST Jarque-Bera	HOME -0.30 E65P 4.2 	7937 0.47429 44105 1.22995 	-0.649259 3.44815 PROB 0.00000	0.51869 0.00105
PCTOWN PCTAG REGRESSION DIA MULTICOLLINEAR TEST ON NORMAL TEST Jarque-Bera DIAGNOSTICS FO	HOME -0.30 E65P 4.2 	7937 0.47429 4105 1.22995 NUMBER 48.997744 VALUE 62.3341 YIICITY	-0.649259 3.44815 PROB 0.00000	0.51869 0.00105
PCTOWN PCTAG PCTAG REGRESSION DIA MULTICOLLINEAR TEST ON NORMAL TEST Jarque-Bera DIAGNOSTICS FO RANDOM COEFFIC	HOME -0.30 E65P 4.2 GNOSTICS ITY CONDITION ITY OF ERRORS DF 2 R HETEROSKEDAS IENTS	7937 0.47429 44105 1.22995 NUMBER 48.997744 VALUE 62.3341 TICITY	-0.649259 3.44815 PROB 0.00000	0.51869 0.00105
PCTOWN PCTAG REGRESSION DIA MULTICOLLINEAR TEST ON NORMAL TEST Jarque-Bera DIAGNOSTICS FO RANDOM COEFFIC TEST	HOME -0.30 E65P 4.2  GNOSTICS ITY CONDITION ITY OF ERRORS DF 2 R HETEROSKEDAS IENTS DF	7937 0.47429 4105 1.22995 	-0.649259 3.44815 PROB 0.00000 PROB	0.51869 0.00105
PCTOWN PCTAG REGRESSION DIA MULTICOLLINEAR TEST ON NORMAL TEST Jarque-Bera DIAGNOSTICS FO RANDOM COEFFIC TEST Breusch-Pagan	HOME -0.33 E65P 4.2  GNOSTICS ITY CONDITION ITY OF ERRORS 2 R HETEROSKEDAS IENTS DF test 3	7937 0.47429 44105 1.22995 NUMBER 48.997744 VALUE 62.3341 STICITY VALUE 15.6910	-0.649259 3.44815 PROB 0.00000 PROB 0.00131	0.51869 0.00105
PCTOWN PCTAG REGRESSION DIA MULTICOLLINEAR TEST ON NORMAL TEST JAIQUE-BERA DIAGNOSTICS FO RANDOM COEFFIC TEST Breusch-Pagan Koenker-Basset	HOME -0.33 E65P 4.2 	7937 0.47429 4105 1.22995 	-0.649259 3.44815 PROB 0.00000 PROB 0.00131 0.15601	0.51869 0.00105
PCTOWN PCTAG REGRESSION DIA MULTICOLLINEAR TEST ON NORMAL TEST Jarque-Bera DIAGNOSTICS FO RANDOM COEFFIC TEST Breusch-Pagan Breusch-Pagan Specification	HOME -0.30 E65P 4.2  GNOSTICS TTY CONDITION ITY OF ERRORS DF 2 R HETEROSKEDAS IENTS DF test 3 ttest 3 ROBUST TEST	7937 0.47429 4105 1.22995 NUMBER 48.997744 VALUE 62.3341 STICITY VALUE 15.6910 5.2255	-0.649259 3.44815 PROB 0.00000 PROB 0.00131 0.15601	0.51869 0.00105
PCTOWN PCTAG REGRESSION DIA MULTICOLLINEAR TEST ON NORMAL TEST JAIQUE-BETA DIAGNOSTICS FO RANDOM COEFFIC TEST REVENCH-PAGAN KOENKEY-PAGASE SPECIFICATION TEST	HOME -0.30 E655P 4.2 	7937 0.47429 4105 1.22995 VILUE 62.3341 VALUE VALUE 15.6910 5.2255 VALUE	-0.649259 3.44815 PROB 0.00000 PROB 0.00131 0.15601 PROB	0.51869
PCTOWN PCTAG REGRESSION DIA MULPICOLINEAR TEST ON NORMAL TEST JAIQUO-BERA DIAGNOSTICS FO RANDOM COEFFIC TEST Breusch-Pagan SpecificATION TEST White	HOME -0.30 E65P 4.2  GNOSTICS 17Y CONDITION ITY OF DEDITION ITY OF DEDITION 2 R HETEROSKEDAS IENTS DF test 3 ROBUST TEST DF 9	7937 0.47429 4105 1.22995 	-0.649259 3.44815 PROB 0.00000 PROB 0.00131 0.15601 PROB 0.15261	0.51869
PCTOWN PCTAG REGRESSION DIA MULFICOLLINEAR TEST ON NORMAL TEST Jarque-Bera DIAGNOSTICS FO RANDOM COEFFIC TEST Breusch-Pagan Scenker-Basset SPECIFICATION TEST MAILE COEFFICIENTS V	HOME -0.30 E65P 4.2 GNOSTICS ITY CONDITION TY CONDITION ITY OF ERRORS DF 2 R HETEROGKEDAS IENTS Construction DF test 3 ROBUST TEST DF 9 ARIANCE MATRIX	7937 0.47429 (4105 1.22995 NUMBER 48.997744 VALUE 62.3341 STICITY VALUE 15.6910 5.2255 VALUE 13.2268	-0.649259 3.44815 PROB 0.00000 PROB 0.00131 0.15601 PROB 0.15261	0.51869
PCTOWN PCTAG REGRESSION DIA MULTICOLLINEAR TEST ON NORMAL TEST DIAGNOSTICS FO DIAGNOSTICS FO RANDOM COEFFIC TEST BECUSCH-PAGAN KOENKER-BASSEt SPECIFICATION TEST White CONSTANT	HOME -0.30 E655P 4.2 GNOSTICS ITY CONDITION GROSTICS ITY OF ERRORS DF 2 R HETEROSKEDAS IENTS DF test 3 t test 3 t test 3 ROBUST TEST DF 9 ARIANCE MATRIX	7937 0.47429 4405 1.22995 NUMBER 48.997744 VALUE 62.3341 STICITY VALUE 15.6910 5.2255 VALUE 13.2268 SOWNHOME PCTAGE65P	-0.649259 3.44815 PROB 0.00000 PROB 0.00131 0.15601 PROB 0.15261	0.51869
PCTOWN PCTAG REGRESSION DIA MULPICOLLINEAR TEST ON NORMAL TEST Jarque-Bera DIAGNOSTICS FO RANDOM COEFFIC TEST Breusch-Pagan Koenker-Basset SPECIFICATION TEST White COEFFICIENTS V CONSTANT 4.666623	HOME -0.30 E65P 4.2  GNOSTICS ITY CONDITION ITY OF PERFORS DF 2 R HETEROSKEDAS IENTS CALL ST ROBUST TEST OF 9 ARIANCE MATRIX PEXPOSURE PCT 4.556146 0	17937 0.47429 1.22995 NUMBER 48.997744 VALUE 62.3341 STICITY VALUE 15.6910 5.2255 VALUE 13.2268 S COMNHOME PCTAGE65P CONSIDER PCTAGE65P 0.015900 -0.238955	-0.649259 3.44815 PROB 0.00000 PROB 0.00131 0.15501 PROB 0.15261	0.51869
PCTOWN PCTAG REGRESSION DIA MULTICOLLINEAR TEST ON NORMAL TEST ON NORMAL TEST DIAGNOSTICS FO ARANDOM COEFFIC TEST Breusch-Pagan Koenker-Basset SPECIFICATION TEST CONSTANT 4.666623 44.556146	HOME -0.30 ES5P 4.2 GNOSTICS ITY CONDITION ITY OF ERRORS DF 2 R HETEROSKEDAS IENTS DF test 3 ROBUST TEST OF 9 ARIANCE MATRIX PSXPOSURE PCT -4.555146 00	17937 0.47429 1.22995 NUMBER 48.997744 VALUE 62.3341 TTICITY VALUE 15.6910 5.2255 VALUE 13.2268 COMNHOME PCTAGE65P 0.015900 -0.238955 0.028805 0.028805 0.028805	-0.649259 3.44815 PROB 0.00000 PROB 0.00131 0.15601 PROB 0.15261	0.51869
PCTOWN PCTAG PCTAG REGRESSION DIA MULTICOLLINEAR TEST TEST Jarque-Bera DIAGNOSTICS FO RANDOM COEFFIC TEST Breusch-Pagan Koenker-Basset SPECIFICATION TEST White CONSTANT 4.666623 -4.556146 0.015900	HOME -0.30 E65P 4.2 	77937 0.47429 4105 1.22995 	-0.649259 3.44815 PROB 0.00000 PROB 0.00131 0.15601 PROB 0.15261	0.51869

Figure 21: Multiple linear regression results

2. The Jarque-Bera test is whether the residuals have the skewness and kurtosis matching a normal distribution. Here we see a high value, quite unlikely to be normal.

However we will not fix up this model, we proceed to compare it to models which do take into account spatial correlation.

#### 8.1 Spatial Error model

The first model with a spatial component we will consider is the **spatial error model**. This allows **resduals** of the linear model to be spatially-correlated, and quantifies to what extent they are included in the model.

This typically occurs when there is some spatially-correlated covariate that (1) affects the response and (2) we do not know, or maybe even suspect – otherwise we would identify it, measure it, and include in the linear model. However, we may suspect a factor that we have not, or can not, measure, and this factor has spatial correlation.

For example, this database does not report the proportion of different ethnic groups, nor of different occupational groups (factory workers, office workers, service workers ...). These may be (1) related to leukemia

(genetic susceptibility, occupational exposure), (2) spatially-correlated. If such factors influence leukemia, they will be represented in the residuals, and thus the spatial error model will be provably better than the feature space-only model.

The spatial error model is:

- formula:  $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \lambda \mathbf{W}\mathbf{u} + \boldsymbol{\varepsilon}$
- W is a matrix representing the spatial structure (e.g., neighbour weights)
- $\mathbf{u} = (\mathbf{Y} \mathbf{X}\boldsymbol{\beta})$  are the spatially-correlated **residuals**
- $\cdot \ \lambda$  is the strength of autoregression of the errors
- *ε* is the independent error (not autoregressive)

**TASK 25** : Compute a multivariate linear regression model of leukemia incidence (response) as predicted by the three possible causitive factors (predictors). This time (1) select a Weights file (one you created above), and then you can specify the SAR **Spatial Error** linear model. This takes spatial correlation of the **linear model residuals** into account, considering the values of the model **residual** in each tract's neighbourhood, as defined by the weights.

Figure 22 shows the results.

**Q28** : What is the pseudo-R<sup>2</sup> of this model?<sup>6</sup>. Is it higher or lower than that for the feature-space only model? Is this expected? How can it be explained?

**Q29**: What are the signs of the slopes for each predictor? What is the interpretation? Which (if any) predictors are significantly different from zero? What changes in this model compared to the feature-space multiple regression? I.e., which predictors become more or less important and/or significant?

**Q30**: What is the strength of the autocorrelation parameter  $\lambda$ ?

The **likelihood ratio test** gives the probability that the SAR spatial error model is *not* better than the feature-space-only multiple regression.

**Q31** : What is the probability that the SAR spatial error model is not better than the feature-space-only multiple regression? What does this imply about the possible causes of leukemia?

 $<sup>^6</sup>$  This is not, strictly speaking, an  $R^2$  but does express the proportion of variance explained

Regression Report				
<b>&gt;</b>				
>>03/07/2019 18:28 REGRESSION	:29			
SUMMARY OF OUTPUT:	SPATIAL ERROR N	MODEL - MAXIMUM	LIKELIHOOD E	STIMATION
Data set	: Syr			
Spatial Weight	: Syr24			
Dependent Variable	: Z	Number of Obs	ervations:	63
Mean dependent var	: 0.037752	Number of Var	iables :	4
S.D. dependent var	: 0.996518	Degrees of Fr	eedom :	59
Lag coeff. (Lambda	): 0.466941			
R-squared	• 0 241243	R-squared (BI	(SE) • -	
Sg. Correlation	: -	Log likelihoo	d : -	81.094551
Sigma-square	: 0.753483	Akaike info c	riterion :	170.189
S.E of regression	: 0.868034	Schwarz crite	erion :	178.762
Variable	COEfficient	Std.Error	z-vaiue	Probability
CONSTANT	-3.48496	3.0928	-1.126	8 0.25983
PEXPOSURE	2.87842	3.04238	0.94610	7 0.34409
PCTOWNHOME	-0.0135155	0.483305	-0.027964	7 0.97769
PCTAGE65P	4.07764	1.18438	3.4428	5 0.00058
LAMBDA	0.466941	0.221434	2.1087	2 0.03497
REGRESSION DIAGNOS	TICS	_		
DIAGNOSTICS FOR HE	TEROSKEDASTICITY	r		
TROT	0	DF	VATUE	PPOP
Breusch-Pagan test		3	11.6116	0.00884
2.5436m-rugun test		5		5.50004
DIAGNOSTICS FOR SP	ATIAL DEPENDENCH	S NAMPITY - Com	24	
FALLAL ERROR DEPE	NUENCE FOR WEIGH	DF	VALUE	PROB
Likelihood Ratio To	est	1	3.2332	0.07216
LINGIINGGU RUCIO I		1	5.2552	0.07210
COEFFICIENTS VARIA	NCE MATRIX			
CONSTANT PEXP	OSURE PCTOWNHON	ME PCTAGE65P	LAMBDA	
9.565402 -9.3	54787 0.02903	32 -0.287450	0.00000	
-9.354787 9.2	56087 -0.11913	36 0.079041	0.00000	
0.029032 -0.1	19136 0.23358	84 -0.018171	0.00000	
-0.287450 0.0	79041 -0.0181	71 1.402760	0.000000	
0.000000 0.0	00000 0.00000	0.000000	0.049033	
	FNI	) OF PFDOPT		

Figure 22: SAR spatial error model regression results

#### 8.2 Spatial Lag model

Another possible effect of spatial autocorrelation is in the response, that is, the values of the response in a tract's neighbours directly influence the response in the tract, after taking into account the feature-space prediction. This measures "contagion", which seems unlikely for human leukemia<sup>7</sup>, however we still evaluate this.

**TASK 26** : Compute a multivariate linear regression model of leukemia incidence (response) as predicted by the three possible causitive factors (predictors). This time (1) select a Weights file (one you created above), and then you can specify the SAR **Spatial Lag** linear model. This takes spatial correlation into account, considering the values of the **response** variable in each tract's neighbourhood, as defined by the weights.

The spatial lag model is:  $\mathbf{Y} = \rho \mathbf{W} \mathbf{Y} + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon}$ , where  $\rho$  is the strength of autoregression of the response; this multiplies the weights matrix times the response **WY** on the right-hand (predictor) side of the equation.

Figure 23 shows the results.

<sup>&</sup>lt;sup>7</sup> although quite likely for feline leukemia, if infected cats travel across tract boundaries

>>03/07/2019 18:27:53           LECRESSION	Regression Report						
<pre>&gt;&gt;3/07/2019 18:27:53 REGRESSION</pre>	2						
Variable         Constrain         Statial         Washing         Constraint           Systial         Weight         : Syr2           Spatial         Weight         : Syr2           Dependent Variable         :         Z         Number of Observations:         63           Mean dependent var:         :         0.0377522         Number of Variables:         :         5           S.D. dependent var:         :         0.996518         Degrees of Freedom         :         58           Lag coeff.         (Rho):         :         0.435129         Resquared         :         -81.0283           SG. Correlation         :         -         Akaike info criterion:         :         172.057           Sigma-square         :         0.754082         Schwarz criterion:         :         182.772           S.E of regression:         :         0.868379         :         :         2.03885         :         0.04147           CONSTANT         -2.38973         :         :         2.03815         :         0.42937           PEXPOSURE         :         :         :         :         3.04437         :         0.69344           PCTAGESEP         :         :         : <td>&gt;&gt;03/07/2019 18:27:53 REGRESSION</td> <td>3</td> <td></td> <td></td> <td></td>	>>03/07/2019 18:27:53 REGRESSION	3					
Data set : Syr Spatial Weight : Syr24 Dependent Variable : Z Number of Observations: 63 Wean dependent var : 0.0307522 Number of Variables : 5 S.D. dependent var : 0.030518 Degrees of Freedom : 58 R-squared : 0.240639 Log likelihood : -81.0283 Sg. Correlation : - Akaike info criterion : 172.057 Sigma-square : 0.754082 Schwarz criterion : 182.772 S.E of regression : 0.868379 	SUMMARY OF OUTPUT: SP	ATIAL LAG MOD	EL - MAXIMUM L	IKELIHOOD ESTI	MATION		
Spatial Weight : Syr24 Dependent Variable : Z Number of Observations: 63 Wean dependent var : 0.937522 Number of Variables : 5 S.D. dependent var : 0.936518 Degrees of Freedom : 58 Lag coeff. (Rho) : 0.435129 R-squared : 0.240639 Log likelihood : -81.0283 Sg. Correlation : - Rkaike info criterion : 172.057 Sigma-square : 0.754082 Schwarz criterion : 172.057 Sigma-square : 0.754082 Schwarz criterion : 182.772 S.E of regression : 0.868379 Variable Coefficient Std.Error z-value Probability W Z 0.435129 0.213419 2.03885 0.04147 CONSTANT -2.38973 2.07463 -1.15188 0.24937 PEXPOSURE 1.86442 2.0381 0.914782 0.36031 PCTOWNHOME -0.174693 0.443173 -0.394187 0.69344 PCTOMES FOR HETEROSKEDASTICITY ANDOM COEFFICIENTS TEST DF VALUE PROB STEWESCH SAGAN STICK JIAGNOSTICS FOR HETEROSKEDASTICITY ANDOM COEFFICIENTS TEST DF VALUE PROB STEWESCH SAGAN STICK J JIAGNOSTICS FOR SPATIAL DEPENDENCE SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : Syr24 LEST DF VALUE PROB STEWESCH SAGAN STICK J JIAGNOSTICS FOR SPATIAL DEPENDENCE SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : Syr24 LEST DF VALUE PROB Likelihood Ratio Test DF VALUE PROB Likelihood Ratio Test DF VALUE PROB Likelihood Ratio Test DF VALUE PROB JIAGNOSTICS FOR SPATIAL DEPENDENCE SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : Syr24 LEST DF VALUE PROB JIAGNOSTICS FOR SPATIAL DEPENDENCE SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : Syr24 LEST DF VALUE PROB JIAGNOSTICS FOR SPATIAL DEPENDENCE SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : Syr24 LEST DF VALUE PROB JIAGNOSTICS FOR SPATIAL DEPENDENCE FOR WEIGHT MATRIX : Syr24 Likelihood Ratio Test DF VALUE PROB JIAGNOSTICS FOR SPATIAL DEPENDENCE PCTAREGESP W Z 4.304079 -4.192620 0.014426 -0.278994 0.102293 -4.196260 4.153356 -0.081688 0.092085 -0.097233 -0.014426 -0.081688 0.196402 -0.0313320 -0.0313320 0.01243 -0.0313320 -0.0313320 -0.0313320 -0.0313320 -0.0313320 -0.0313320 -0.0313320 -0.0313320 -0.0313320 -0.0313320 -0.0313320 -0.0313320 -0.0313320 -0.0313320 -0.0313320 -0.0313320 -0.0313320 -0.0313320 -0.0313	Data set :	Syr					
Dependent Variable : z Number of Observations: 63 Mean dependent var : 0.397522 Number of Variables : 5 S.D. dependent var : 0.396518 Degrees of Freedom : 58 R-squared : 0.240639 Log likelihood : -81.0283 Sq. Correlation : - Akaike info criterion : 172.057 Sigma-square : 0.754082 Schwarz criterion : 182.772 S.E of regression : 0.868379 Variable Coefficient Std.Error z-value Probability Variable Coefficient Std.Error z-value Probability Variable Coefficient Std.Error z-value Probability PEXPOSURE 1.86442 2.0381 0.914782 0.36031 PCTOWNHOME -0.174693 0.443173 -0.394187 0.69344 PCTAGE65P 3.99453 1.15859 3.44775 0.00057 REGRESSION DIAGNOSTICS DIAGNOSTICS FOR HETEROSKEDASTICITY ANDOM COEFFICIENTS EFST DF VALUE PROB STeusch-Pagan test 3 12.0911 0.00708 DIAGNOSTICS FOR SPATIAL DEPENDENCE SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : Syr24 TEST DF VALUE PROB Steusch-Pagan test 1 3.3656 0.06657 COEFFICIENTS VALUE PROB Likelihood Ratio Test 1 3.3656 0.06657 COEFFICIENTS VALUE PROB 1 3.3656 0.00203 -0.07249 0.00203 -0.07249 0.002043 -0.092043 -0.031332 0.00442 -0.0274994 0.002045 -0.031332 0.004424 -0.031332 0.004443 -0.031332 0.004443 -0.031332 0.004443 -0.031332 0.004445 -0.031332 0.004445 -0.031332 0.004545 -0.031332 0.004545 -0.031332 0.004545 -0.031332 0.00	Spatial Weight :	Syr24					
Mean dependent var : 0.0377522 Number of Variables : 5 S.D. dependent var : 0.0396518 Degrees of Freedom : 58 Lag coeff. (Rho) : 0.435129 R-squared : 0.240639 Log likelihood : -81.0283 Sq. Correlation : - Akaike info criterion : 172.057 Sigma-square : 0.754082 Schwarz criterion : 172.057 Sigma-square : 0.754082 Schwarz criterion : 182.772 S.E of regression : 0.868379 	Dependent Variable :	- Z	Number of Obs	ervations: 6	3		
S.D. dependent var : 0.996518 Degrees of Freedom : 58 Lag coeff. (Rho) : 0.435129 R-squared : 0.240639 Log likelihood : -81.0283 Akaike info criterion : 172.057 Sigma-square : 0.754082 Schwarz criterion : 182.772 S.E of regression : 0.868379 Variable Coefficient Std.Error z-value Probability Variable Coefficient Std.Error z-value Probability Variable Coefficient Std.Error z-value Probability 0.435129 0.213419 2.03885 0.04147 CONSTANT -2.38973 2.07463 -1.15188 0.24937 PERFOSURE 1.86442 2.0381 0.914782 0.36031 PCTOWNHOME -0.174693 0.443173 -0.394187 0.69344 PCTAGE65P 3.99453 1.15859 3.44775 0.00057 REGRESSION DIAGNOSTICS DIAGNOSTICS FOR HETEROSKEDASTICITY ANNOM COEFFICIENTS TEST DF VALUE PROB Breusch-Bagan test 3 12.0911 0.00708 DIAGNOSTICS FOR SPATIAL DEPENDENCE SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : Syr24 DEFFICIENTS VALUE PROB Likelihood Ratio Test 1 3.3656 0.06657 COEFFICIENTS VALUE MATRIX CONSTANT PERFOSURE PCTOWNHOME PCTAGE65P W Z 4.304079 -4.195260 0.014426 -0.278994 0.102293 -4.196206 4.153856 -0.081688 0.092085 -0.097233 -0.014426 -0.081688 0.19620 -0.031332 0.00454 -0.278994 0.092085 -0.034482 1.342334 -0.031332	Mean dependent var :	0.0377522	Number of Var	iables :	5		
Lag coeff. (Rho) : 0.435129 R-squared : 0.240639 Log likelihood : -81.0283 Rg. Correlation : - Akaike info criterion : 172.057 Sigma-square : 0.754082 Schwarz criterion : 182.772 S.E of regression : 0.868379 	S.D. dependent var :	0.996518	Degrees of Fr	eedom : 5	8		
R-squared : 0.240639 Log likelihood : -81.0283 Sq. Correlation : - Akaike info criterion : 172.057 Sigma-square : 0.754082 Schwarz criterion : 182.772 S.E of regression : 0.868379 Variable Coefficient Std.Error z-value Probability W_Z 0.435129 0.213419 2.03885 0.04147 CONSTANT -2.38973 2.07463 -1.15188 0.24937 PEXPOSURE 1.86442 2.0381 0.914782 0.36031 PCTOWNHOME -0.174693 0.443173 -0.394187 0.69344 PCTAGE655 3.99453 1.15859 3.44775 0.00057 REGRESSION DIAGNOSTICS JIAGNOSTICS FOR HERENGKEDASTICITY ANDOM COEFFICIENTS FEST DF VALUE PROB Sreusch-Pagan test 3 12.0911 0.00708 DIAGNOSTICS FOR SPATIAL DEPENDENCE SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : Syr24 LEST DF VALUE PROB Sieusch-Pagan test 1 3.3656 0.06657 COEFFICIENTS VARIANCE MATRIX CONSTANT PERPOSURE PCTOWNHOME PCTAGE65P W_Z 4.304079 -4.196260 0.014426 -0.278994 0.102293 -4.196260 4.153356 -0.081688 0.092085 -0.097233 -0.014426 -0.081688 0.196024 -0.034482 0.000243 -0.278994 0.092085 -0.034482 1.342334 -0.031332 0 0.016426 -0.081688 0.196024 -0.031332 0 0.0243 -0.031332 0 0.0243 -0.031332 0 0.0243 -0.031332 0 0.0243 -0.031332 0 0.04545	Lag coeff. (Rho) :	0.435129					
Sq. Correlation       :-       Akākē info criterion :       172.057         Sigma-square       :       0.754082       Schwarz criterion :       182.772         Sigma-square       :       0.754082       Schwarz criterion :       182.772         S.E of regression :       0.868379         Variable       Coefficient       Std.Error       z-value       Probability         W_Z       0.435129       0.213419       2.03885       0.04147         CONSTANT       -2.38973       2.07463       -1.15188       0.24937         PEXPOSURE       1.86442       2.0381       0.914782       0.36031         PCTOWNHOME       -0.174693       0.443173       -0.394187       0.69344         PCTAGE55P       3.99453       1.15859       3.44775       0.00057         REGRESSION DIAGNOSTICS       DIAGNOSTICS FOR HETEROSKEDASTICITY       RANDOM COEFFICIENTS       PROB         DIAGNOSTICS FOR SPATIAL DEPENDENCE       Spralater       3       12.0911       0.00708         DIAGNOSTICS FOR SPATIAL DEPENDENCE FOR WEIGHT MATRIX : Syr24       Est       1       3.3656       0.06657         COEFFICIENTS       VAILUE PROB       Est       1       3.3656       0.06657         COEFFICIENTS VARIANCE MATR	R-squared :	0.240639	Log likelihoo	d :	-81.0283		
Sigma-square : 0.754082 Schwarz criterion : 182.772 S.E of regression : 0.868379 Variable Coefficient Std.Error z-value Probability W_Z 0.435129 0.213419 2.03885 0.04147 CONSTANT -2.38973 2.07463 -1.15188 0.24937 PEXPOSURE 1.86442 2.0381 0.914782 0.36031 PCTOWNHOME -0.174693 0.443173 -0.394187 0.69344 PCTAGE65P 3.99453 1.15859 3.44775 0.00057 TEXT DIACNOSTICS DIACNOSTICS FOR HETEROSKEDASTICITY RANDOM COEFFICIENTS TEST DF VALUE PROB STeusch-Pagan test 3 12.0911 0.00708 JIACNOSTICS FOR SPATIAL DEPENDENCE SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : Syr24 PEST DF VALUE PROB STAUSCH PAGE 0.014426 -0.278994 0.102293 -4.196260 4.153856 -0.081688 0.19620 -0.031432 0.000243 -0.278994 0.092085 -0.034482 0.00243 -0.031332 0.00455 0.0455	Sq. Correlation :	-	Akaike info c	riterion :	172.057		
S.E of regression :         0.868379           Variable         Coefficient         Std.Error         z-value         Probability           W_Z         0.435129         0.213419         2.03885         0.04147           CONSTANT -2.38973         2.07463         -1.15188         0.24937           PEXPOSURE         1.86442         2.0381         0.914782         0.36031           PCTOWNHOME         -0.174693         0.443173         -0.394187         0.69344           PCTAGE65P         3.99453         1.15859         3.44775         0.00057           REGRESSION DIAGNOSTICS         0.443173         -0.394187         0.69344           PCTAGE65P FOR HEREROSKEDASTICITY         RANDOM COEFFICIENTS         FROB           Sreusch-Pagan test         J         1         0.00708           DIAGNOSTICS FOR SPATIAL DEPENDENCE         FOR         PCD           Shelihood Ratio Test         J         J         0.00708           DIAGNOSTICS FOR SPATIAL DEPENDENCE FOR WEIGHT MATRIX : Syr24         FROB         J           Likelihood Ratio Test         J         3.3656         0.06657           COEFFICIENTS VARIANCE MATRIX         CONSTANT PERPOSURE PCTOWNHOME PCTAGE65P         W Z           4.304079         -4.196260	Sigma-square	0.754082	Schwarz crite	rion :	182.772		
Variable         Coefficient         Std.Error         z-value         Probability           W_Z         0.435129         0.213419         2.03885         0.04147           CONSTANT         -2.38973         2.07463         -1.15188         0.24937           PEXPOSURE         1.86442         2.0381         0.914782         0.36031           PCTOWNHOME         -0.174693         0.443173         -0.394187         0.69344           PCTAGE65P         3.99453         1.15859         3.44775         0.00057           REGRESSION DIAGNOSTICS         DIACMOSTICS FOR HETREKOSKEDASTICITY         RANDM COEFFICIENTS         TEST         DF         VALUE         PROB           Sreusch-Pagan test         3         12.0911         0.00708         0.00708           DIAGNOSTICS FOR SPATIAL DEPENDENCE         SPRTIAL LAG DEPENDENCE FOR WEIGHT MATRIX : Syr24         FROB         Likelihood Ratio Test         1         3.3656         0.06657           CORFICIENTS VARIANCE MATRIX         CONSTANT PERPOSURE PCTOWNHOME PCTAGE65P         W Z         4.304079         -4.196260         0.01426         -0.278994         0.102293           -4.396200         4.153356         -0.081688         0.092085         -0.034482         -0.031322           -0.197233         0.0040	S.E of regression :	0.868379					
W Z         0.435129         0.213419         2.03885         0.04147           CONSTANT         -2.38973         2.07463         -1.15188         0.24937           PEXPOSURE         1.86442         2.0381         0.914782         0.36031           PCTOWNHOME         -0.174693         0.443173         -0.394187         0.69314           PCTAGE65P         3.99453         1.15859         3.44775         0.00057	Variable	Coefficient	Std.Error	z_value	Probability		
w_z         0.435129         2.03895         0.0447           CONSTANT         -2.38973         2.07463         -1.15188         0.24937           PEXPOSURE         1.86442         2.0381         0.914782         0.36031           PCTAGE65P         3.99453         1.151859         0.04477         0.6031           PCTAGE65P         3.99453         1.15859         3.44775         0.00057           TEST         DF         VALUE         PROB           Breusch-Pagan test         3         12.0911         0.00708           DIAGNOSTICS FOR HETEROSKEDASTICITY         SALUE         PROB           Breusch-Pagan test         3         12.0911         0.00708           DIAGNOSTICS FOR SPATIAL DEPENDENCE         SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : Syr24         FROB           Eikelihood Ratio Test         1         3.3656         0.06657           COEFFICIENTS VARIANCE MATRIX         CONSTANT PEXPOSURE PCTOWNHOME PCTAGE65P         WZ           4.304079         -4.196260         0.01426         -0.278994         0.102293           -4.196260         4.153856         -0.081688         0.092085         -0.031332           0.01426         -0.031462         0.031432         -0.031332							
CONSTANT         -2.38973         2.07463         -1.15188         0.24937           PPEXPOSURE         1.86442         2.0381         0.914782         0.36031           PCTOWNHOME         -0.174693         0.443173         -0.394187         0.699344           PCTAGE65P         3.99453         1.15859         3.44775         0.00057           DIAGNOSTICS         DIAGNOSTICS         0.443173         -0.394187         0.699344           DIAGNOSTICS         DF         VALUE         PROB           Breusch-Pagan test         3         12.0911         0.00708           DIAGNOSTICS FOR SPATIAL DEPENDENCE         SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : Syr24         PROB           Effect         DF         VALUE         PROB           Likelihood Ratio Test         1         3.3656         0.06657           CORSTANT         PEXPOSURE PCTOWNHOME PCTAGE65P         W_Z         4.304079         -4.196260         -0.01426         -0.278994         0.102293           -4.196260         4.153856         -0.031482         -0.034482         -0.031332         0.004243           -0.1468         0.19642         -0.031332         -0.031332         -0.031332         0.04526	W_Z	0.435129	0.213419	2.03885	0.04147		
PEAFUSURE       1.05442       2.0381       0.914782       0.6031         PCTAGUERE       -0.174693       0.443173       -0.394187       0.69344         PCTAGE65P       3.99453       1.15859       3.44775       0.00057         TERGRESSION DIAGNOSTICS       DIAGNOSTICS FOR HETEROSKEDASTICITY       NAUVE       PROB         BREQUESTICS FOR HETEROSKEDASTICITY       DF       VALUE       PROB         Breusch-Pagan test       3       12.0911       0.00708         DIAGNOSTICS FOR SPATIAL DEPENDENCE       SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : Syr24       FST         FEST       DF       VALUE       PROB         Likelihood Ratio Test       1       3.3656       0.06657         CONSTANT PEXPOSURE PCOWNHOME PCTAGE65P       W.Z       4.304079       -4.196260       0.01426       -0.278994       0.102293         -4.196260       4.153856       -0.081688       0.092085       -0.033482       0.00243         -0.278994       0.002433       -0.033442       -0.031332       0.004545         -0.172893       -0.092085       -0.031332       -0.03452       0.01426	CONSTANT	-2.389/3	2.0/463	-1.15188	0.24937		
PCIONNEMPE         -U.174053         0.443173         -U.394187         0.039344           PCTAGE55P         3.99453         1.15859         3.44775         0.00057           REGRESSION DIAGNOSTICS DIAGNOSTICS FOR HETEOSKEDASTICITY RANDOM COEFFICIENTS         DF         VALUE         PROB           Derusch-Pagan test         3         12.0911         0.00708           DIAGNOSTICS FOR SPATIAL DEPENDENCE         SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : Syr24         PROB           EEST         DF         VALUE         PROB           Likelihood Ratio Test         1         3.3656         0.06657           COEFFICIENTS VARIANCE MATRIX         CONSTANT         PEXPOSURE         PCTMHOME           4.304079         -4.196260         0.014426         -0.278994         0.102293           -4.196206         4.153856         -0.034682         0.002433         -0.031332           -0.278994         0.092085         -0.031332         0.00454         -0.031332           0.102293         -0.092033         -0.03233         -0.031332         0.04569	PEXPOSURE	1.86442	2.0381	0.914/82	0.36031		
REGRESSION DIAGNOSTICS         DF         VALUE         PROB           DIAGNOSTICS FOR HETEROSKEDASTICITY         RANDOM COEFFICIENTS         DF         VALUE         PROB           Breusch-Pagan test         3         12.0911         0.00708           DIAGNOSTICS FOR HETEROSKEDASTICITY         RANDOM COEFFICIENTS         DF         VALUE         PROB           Breusch-Pagan test         3         12.0911         0.00708           DIAGNOSTICS FOR SPATIAL DEPENDENCE         SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : Syr24         FEST         DF         VALUE         PROB           Likelihood Ratio Test         1         3.3656         0.06657         SCEFICIENTS VARIANCE MATRIX         CONSTANT PEXPOSURE PCTOWNHOME PCTAGE65P         W Z           -4.196260         4.153356         -0.081688         0.092085         -0.037233         0.000243         -0.031332         0.004243           -0.1278994         0.092085         -0.031332         0.04526         0.04526         0.04526	PCTOWNHOME	2 00/52	0.4431/3	-0.39418/	0.09344		
DIAGNOSTICS         DF         VALUE         PROB           DIAGNOSTICS FOR HETEROSKEDASTICITY         3         12.0911         0.00708           TEST         DF         VALUE         PROB           Breusch-Pagan test         3         12.0911         0.00708           DIAGNOSTICS FOR SPATIAL DEPENDENCE         SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : Syr24         FEST         DF         VALUE         PROB           Likelihood Ratio Test         1         3.3656         0.06657         CONSTANT PEXPOSURE PCOWNHOME PCTAGE65P         W Z           4.304079         -4.156260         0.014426         -0.278994         0.102293           -4.196260         4.153356         -0.081688         0.092085         -0.031332           -0.278994         0.002433         -0.031332         0.004243         -0.031332           -0.127894         0.092085         -0.031332         -0.04565							
DIAGNOSTICS FOR HETEROSKEDASTICITY RANDOM COEFFICIENTS TEST DF VALUE PROB Breusch-Pagan test 3 12.0911 0.00708 DIAGNOSTICS FOR SPATIAL DEPENDENCE SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : Syr24 TEST DF VALUE PROB Likelihood Ratio Test 1 3.3656 0.06657 CONSTANT PEXPOSURE MATRIX CONSTANT PEXPOSURE PCTOWNHOME PCTAGE65P W Z 4.304079 -4.196260 0.014426 -0.278994 0.102293 -4.196260 4.153356 -0.081688 0.092085 -0.097233 0.014426 -0.081688 0.196402 -0.0314422 0.000243 -0.278994 0.092085 -0.031432 0.004543 -0.278994 0.092085 -0.031332 0.0455/6	REGRESSION DIAGNOSTIC	15					
RANDOM COEFFICIENTS TEST DF VALUE PROB Breusch-Pagan test 3 12.0911 0.00708 DIAGNOSTICS FOR SPATIAL DEPENDENCE SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : Syr24 TEST DF VALUE PROB Likelihood Ratio Test 1 3.3656 0.06657 COEFFICIENTS VARIANCE MATRIX CONSTANT PERPOSURE PCTOWNHOME PCTAGE65P W Z 4.304079 -4.196260 0.014426 -0.278994 0.102293 -4.196260 4.153356 -0.081688 0.092085 -0.097233 0.014426 -0.081688 0.196402 -0.031482 0.000243 -0.278994 0.092085 -0.034482 1.342334 -0.031332 0 0.016294 0.092085 -0.034682	DIAGNOSTICS FOR HETEH	ROSKEDASTICITY					
DF         VALUE         PROB           Breusch-Pagan test         3         12.0911         0.00708           DIAGNOSTICS FOR SPATIAL DEPENDENCE         SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : Syr24         FROB           TEST         DF         VALUE         PROB           Likelihood Ratio Test         1         3.3656         0.06657           CONSTANT PEXPOSURE PCTONNHOME         PCTAGE65P         W.Z         4.304079           -4.196260         0.014426         -0.278994         0.102293           -0.014688         0.196402         -0.031482         0.000243           -0.278994         0.092085         -0.031332         0.004543           -0.131332         -0.003233         -0.003243         -0.031332	RANDOM COEFFICIENTS						
Breusch-Pagan test         3         12.0911         0.00708           DIAGNOSTICS FOR SPATIAL DEPENDENCE SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : Syr24             TEST         DF         VALUE         PROB           Likelihood Ratio Test         1         3.3656         0.06657           CONSTANT         PEXPOSURE PCTONHOME         PCTAGE65P         W.Z           4.304079         -4.196260         0.014426         -0.278994         0.102293           -4.196260         4.153356         -0.081688         0.092085         -0.031482         0.000243           -0.278994         0.092203         -0.031332         0.004543         -0.031332         0.04556	TEST		DF	VALUE	PROB		
DIAGNOSTICS FOR SPATIAL DEPENDENCE SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : Syr24 TEST DF VALUE PROB Likelihood Ratio Test 1 3.3656 0.06657 COEFFICIENTS VARIANCE MATRIX CONSTANT PERPOSURE PCTOWNHOME PCTAGE65P W Z 4.304079 -4.196260 0.014426 -0.278994 0.102293 -4.196260 4.153356 -0.081688 0.092085 -0.097233 0.014426 -0.081688 0.196402 -0.031482 0.000243 -0.278994 0.092085 -0.034482 1.342334 -0.031332 0 0.016273 -0.097233 -0.00243 -0.031332 -0.045576	Breusch-Pagan test		3	12.0911	0.00708		
SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : Syr24         DF         DVALUE         PROB           DEST         DF         0.10657         0.06657           CONSTANT         PEXPOSURE         PCTOWNHOME         PCTAGE65P         W           CONSTANT         PEXPOSURE         PCTOWNHOME         PCTAGE65P         W         Z           -4.304079         -4.196260         0.014426         -0.278994         0.102293         -4.196260         -0.031468         0.1902085         -0.031482         0.000243           -0.278994         0.092085         -0.0314422         1.342334         -0.031332         0.004545	DIAGNOSTICS FOR SPATE	IAL DEPENDENCE					
DF         VALUE         PROB           Likelihood Ratio Test         1         3.3656         0.06657           COEFFICIENTS VARIANCE MATRIX         CONSTANT         PEXPOSURE         PCTAGE65P         W_Z           4.304079         -4.196260         0.014426         -0.278994         0.102293           -4.196260         4.153856         -0.081688         0.092085         -0.097233           0.014260         -0.031482         0.00243         -0.031332         0.004543           -0.1278994         0.092085         -0.031332         -0.031332         0.04557	SPATIAL LAG DEPENDENC	E FOR WEIGHT	MATRIX : Syr24				
Likelihood Ratio Test 1 3.3656 0.06657 COEFFICIENTS VARIANCE MATRIX CONSTANT PEXPOSURE PCTAGE65P W Z 4.304079 -4.196260 0.014426 -0.278994 0.102293 -4.196260 4.153856 -0.081688 0.092085 -0.097233 0.014426 -0.081688 0.196402 -0.034482 0.000243 -0.278994 0.092085 -0.034482 1.342334 -0.031332 0 0.102293 -0.097233 0.000243 -0.031332 -0.045576	TEST		DF	VALUE	PROB		
COEFFICIENTS         VARIANCE MATRIX           CONSTANT         PEXPOSURE         PCTOWNHOME         PCTAGE65P         W Z           4.304079         -4.196260         0.014426         -0.278994         0.102293           -4.196260         4.153356         -0.081688         0.092085         -0.097233           0.014426         -0.034482         0.000243         -0.031432           -0.278994         0.092085         -1.034482         0.00243           -0.278994         0.092085         -0.031332         -0.031332	Likelihood Ratio Test	:	1	3.3656	0.06657		
CONSTANT         PEXPOSURE         PCTOWNHOME         PCTAGE65P         W_Z           4.304079         -4.196260         0.014426         -0.278994         0.102293           -4.196260         4.153556         -0.081688         0.0292085         -0.097233           0.014426         -0.081688         0.196402         -0.034482         0.000243           -0.278994         0.092085         -0.034482         1.342334         -0.031332           0.102293         -0.00243         -0.031332         -0.04456	COEFFICIENTS VARIANCE	MATRIX					
4.304079 -4.196260 0.014426 -0.278994 0.102293 -4.196260 4.153856 -0.081688 0.092085 -0.097233 0.014426 -0.081688 0.196402 -0.034482 0.000243 -0.278994 0.092085 -0.034482 1.342334 -0.031332 0.04557	CONSTANT PEXPOSU	JRE PCTOWNHOM	E PCTAGE65P	W_Z			
$\begin{array}{rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr$	4.304079 -4.1962	260 0.01442	6 -0.278994	0.102293			
0.014426 -0.081688 0.196402 -0.034482 0.000243 -0.278994 0.092085 -0.034482 1.342334 -0.031332 0.102293 -0.097233 0.001243 -0.031332 0.04554	-4.196260 4.1538	-0.08168	8 0.092085	-0.097233			
-0.278994 0.092085 -0.034482 1.342334 -0.031332 0.102293 -0.097233 0.000243 -0.031332 0.045549	0.014426 -0.0816	588 0.19640	2 -0.034482	0.000243			
0 102293 _0 097233 0 000243 _0 031332 0 045549	-0.278994 0.0920	085 -0.03448	2 1.342334	-0.031332			
0.105512 -0.01512 0.000542 -0.021225 0.042240	0.102293 -0.0972	233 0.00024	3 -0.031332	0.045548			
END OF REPORT		END	OF REPORT ===				

Figure 23: SAR spatial lag model regression results

**Q32** : Is the lag coefficient  $\rho$  significant in the regression? What is the probability that the SAR spatial lag model is not better than the feature-space-only multiple regression? What does this imply about the possible causes of leukemia?

# 9 Finishing with GeoDa

GeoDa provides several opportunities to save images or export newly derived data in tabular form. If while using GeoDa you have derived values and added variables to the attribute table of your shapefile, you will be prompted to Save these as you Exit. Otherwise, the program can simply be closed.

# References

- Luc Anselin, Xun Li, and Julia Koschinsky. GeoDa, from the desktop to an ecosystem for exploring spatial data. *Geographical Analysis*, 54(3):439–466, November 2021. ISSN 1538-4632. doi: 10.1111/gean. 12311.
- [2] Roger S. Bivand, Edzer J. Pebesma, and V. Gómez-Rubio. Applied Spatial Data Analysis with R. Springer, 2nd edition, 2013. ISBN 978-1-4614-7617-7; 978-1-4614-7618-4 (e-book). URL http://www. asdar-book.org/.
- [3] L. A. Waller and C. A. Gotway. *Applied spatial statistics for public health data*. Wiley-Interscience, Hoboken, N.J., 2004.