12 – Geographically Weighted Regression

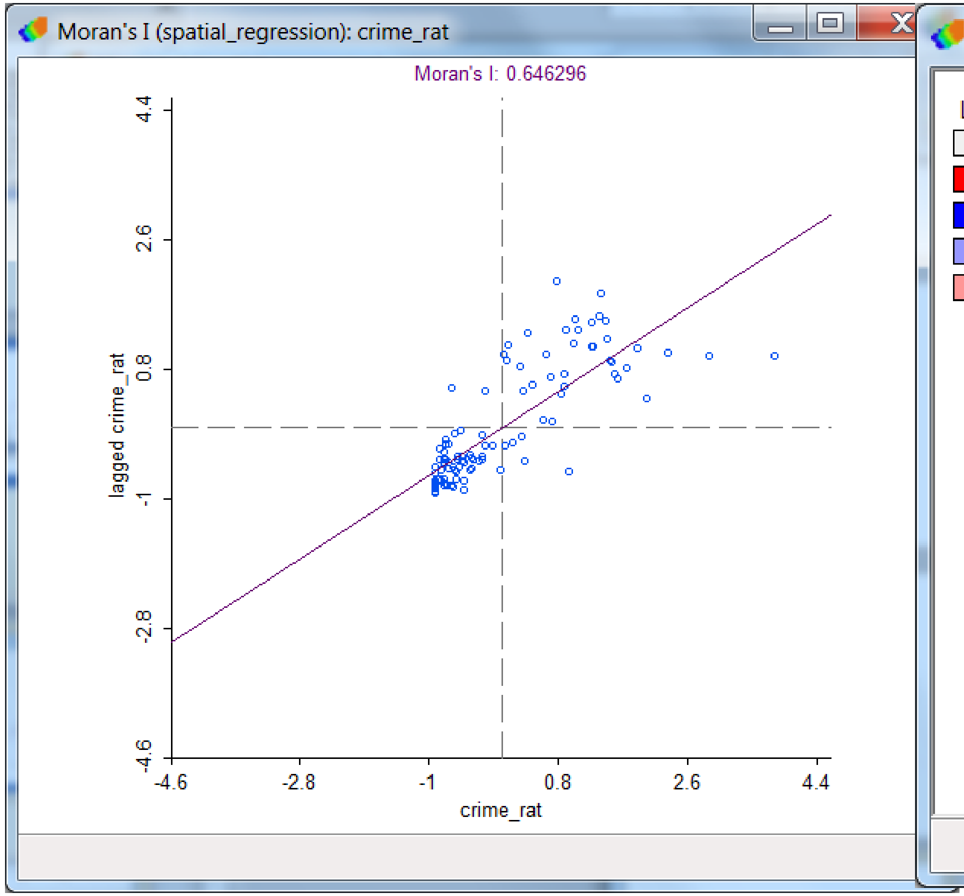
Ness Sandoval
Sociology
Saint Louis University

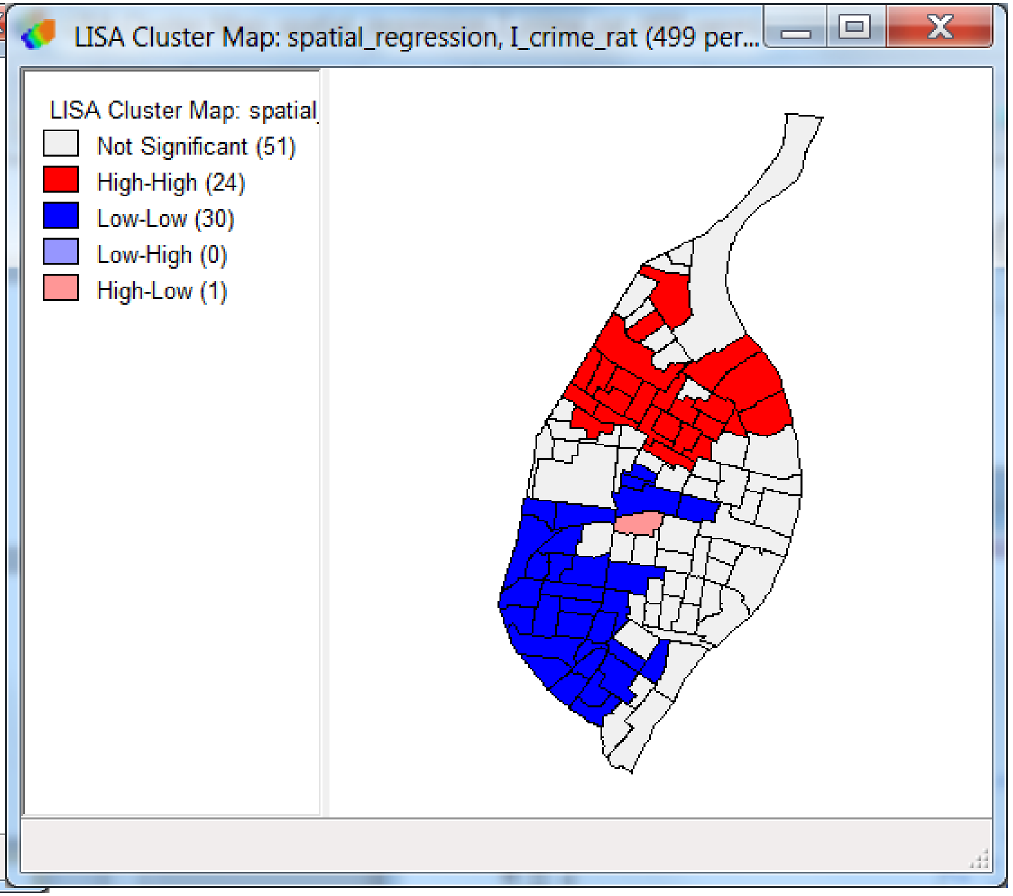
Outline

- Global vs. Local Spatial Statistics
- Stationary vs. Non-stationary
- Why GWR?
- Regression vs. Geographically Weighted Regression
- Lab Example

Global vs. Local Spatial Statistics

Global vs. Local





Variable	OLS	SLM	SEM
Poverty	12.2057 (2.16)***	5.11771 (1.66671) **	4.19772 (2.13146)*
Constant	-0.251479 (0.713962)	-0.46838 0.526285	2.20829 (1.13544)
ρ		0.69895 (.0776769) ***	
λ			0.74083 (0.0751734)***
r-square	0.234909	0.591147	0.578692
Log likelihood	-280.986	-254.216	-256.872
AIC	565.973	514.432	517.744
Moran's I Residual	.4414***	-0.0701	-0.605
N	106	106	106

$* \leq .05, ** \leq .01, *** \leq .001$

Standard Errors in Parentheses

The ρ coefficient is positive and highly significant, indicating strong spatial autocorrelation in the dependent variable. The Moran's I statistic indicates that the residuals are no longer spatially clustered.

Regression Report

```
>>04/06/17 08:20:26
REGRESSION
----------
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set            :  stl_reg_geoda
Dependent Variable  :   homi_rate  Number of Observations:  106
Mean dependent var  :     3.30795  Number of Variables   :    4
S.D. dependent var  :     3.91868  Degrees of Freedom    :  102

R-squared           :    0.503396  F-statistic           :      34.465
Adjusted R-squared  :    0.488790  Prob(F-statistic)     :1.82372e-015
Sum squared residual:     808.344  Log likelihood        :     -258.08
Sigma-square        :     7.92495  Akaike info criterion :     524.159
S.E. of regression  :     2.81513  Schwarz criterion     :     534.813
Sigma-square ML     :     7.62589
S.E of regression ML:      2.7615

-----------------------------------------------------------------------------
       Variable       Coefficient      Std.Error     t-Statistic   Probability
-----------------------------------------------------------------------------
          CONSTANT        4.79579        1.14249        4.19766       0.00006
               p01        10.4424        1.99592        5.23189       0.00000
           theil15        -3.3254        2.93936       -1.13133       0.26057
               e15       -7.40396         1.0404       -7.11642       0.00000
-----------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER   10.442998
TEST ON NORMALITY OF ERRORS
TEST                  DF           VALUE             PROB
Jarque-Bera            2           222.7096          0.00000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                  DF           VALUE             PROB
Breusch-Pagan test     3            35.9462          0.00000
Koenker-Bassett test   3             8.3391          0.03950
============================== END OF REPORT ================================
```

At the global level income inequality is not significant, but this does mean that it is not significant at the local level?

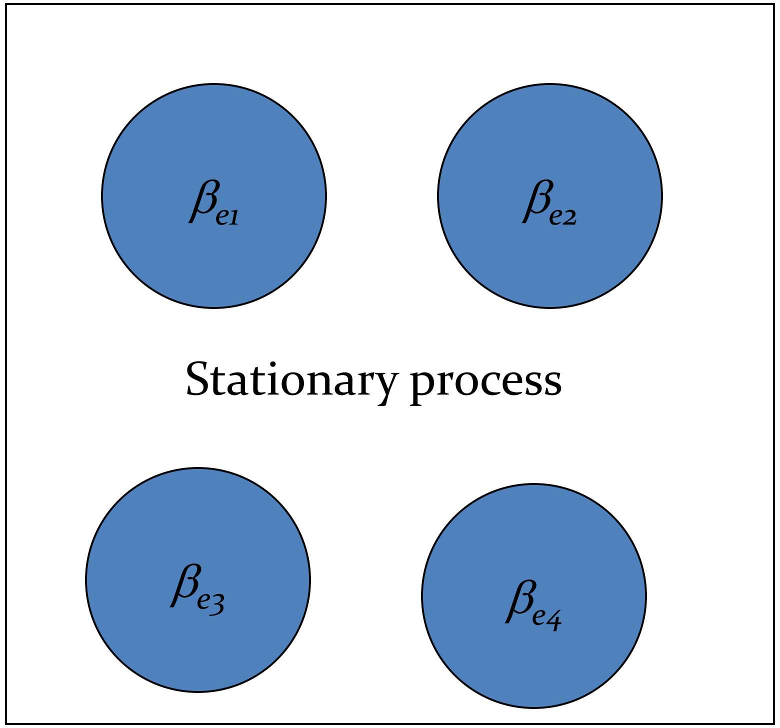
We are assuming the relationship between income inequality and homicide rate is the same regardless of space.

This assumption may be incorrect if our data has local variation.

Stationary vs. Non-stationary

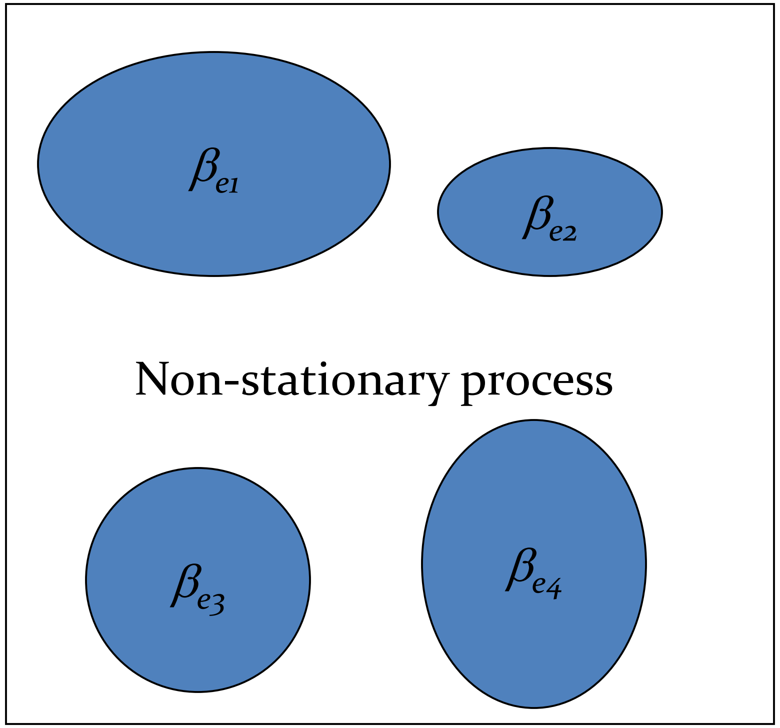
The assumption of stationary in regression is that the values of β are the same everywhere

$y_i = \beta_o + \beta_1 x_{1i}$



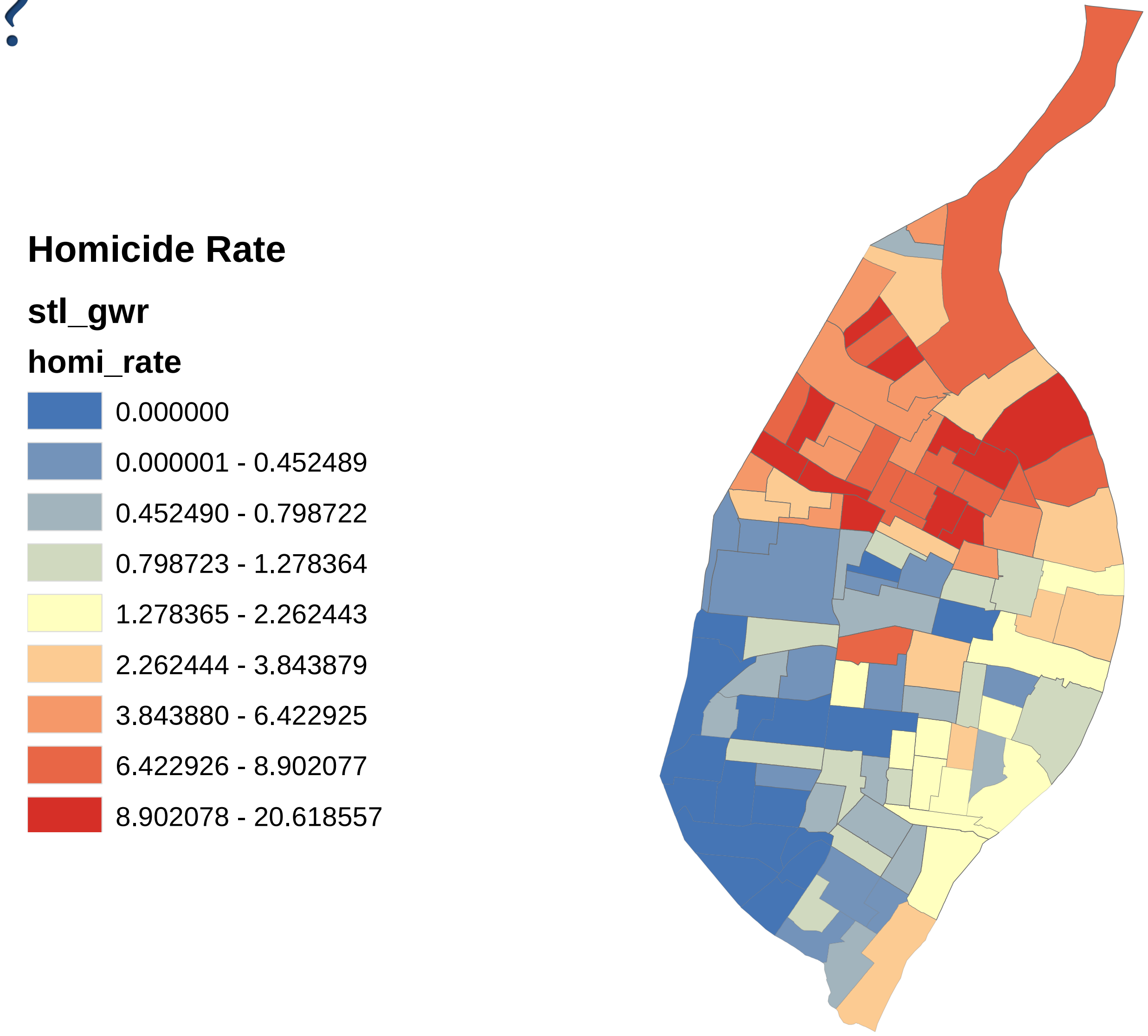
Assumed

$y_i = \beta_{io} + \beta_{i1} x_{1i}$



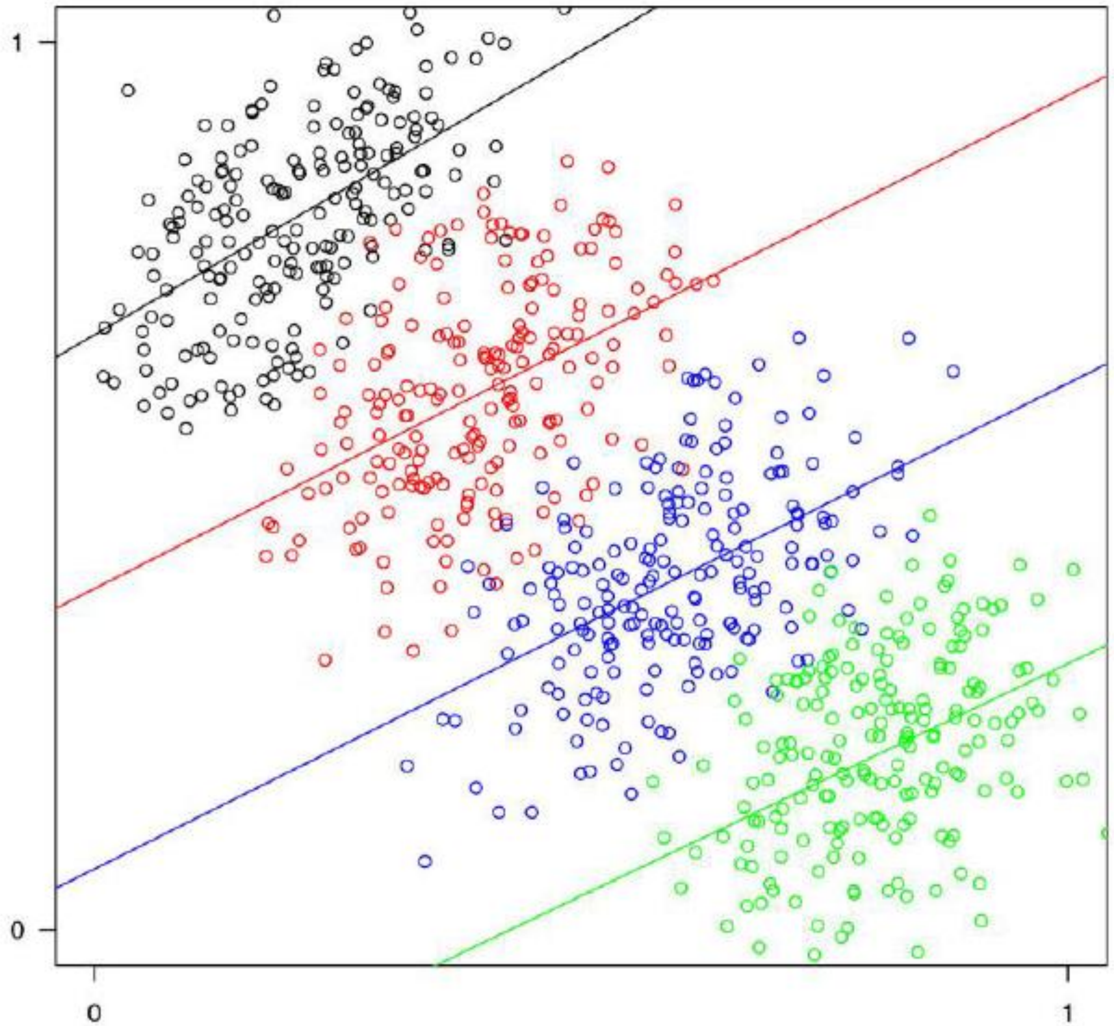
More realistic

Is the relationship between crime and other variables the same throughout the city?



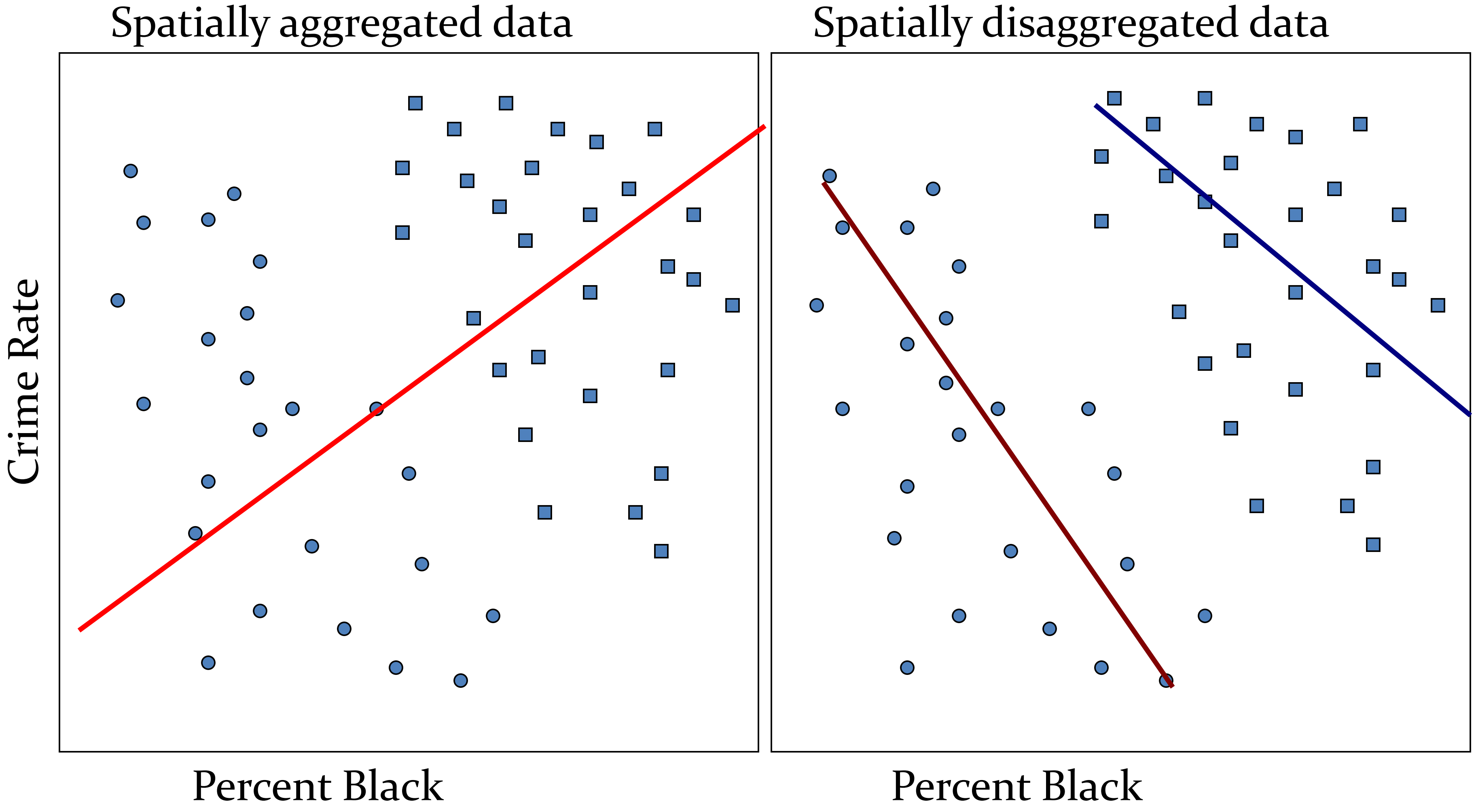
Simpson's Paradox (1951)*

- Refers to the reversal of results when groups of data are analyzed separately and then combined.
- The problem with global regression is the assumption that the same relationship between x and y are the same anywhere in the study area.
- If we think of these points as our data grouped into colors by neighborhood we can see that the global and local models differ significantly.

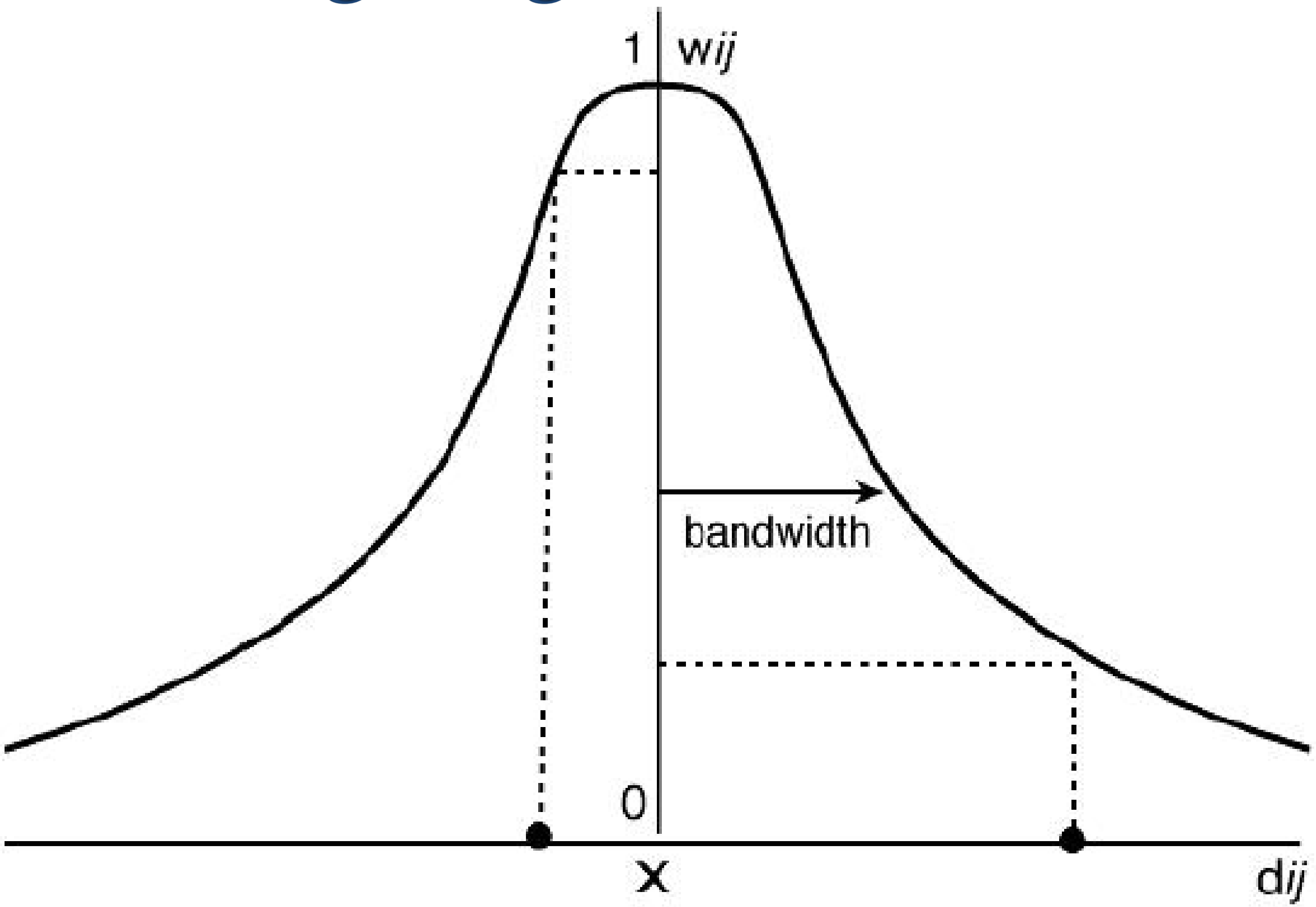


*Simpson, E. H. (1951). "The Interpretation of Interaction in Contingency Tables." Journal of the Royal Statistical Society, Series B **13**: 238–241.

Simpson's paradox Example Two



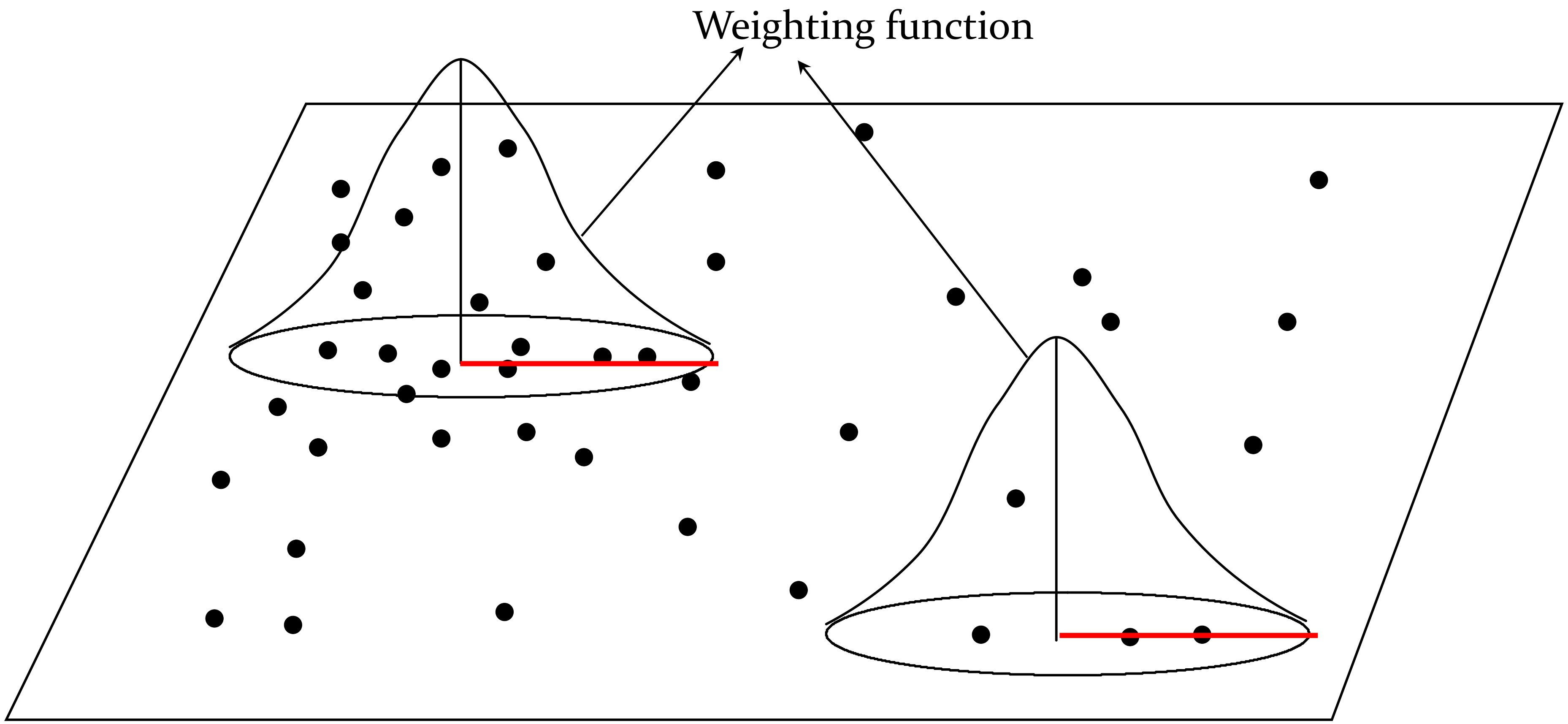
Spatial Weighting Function



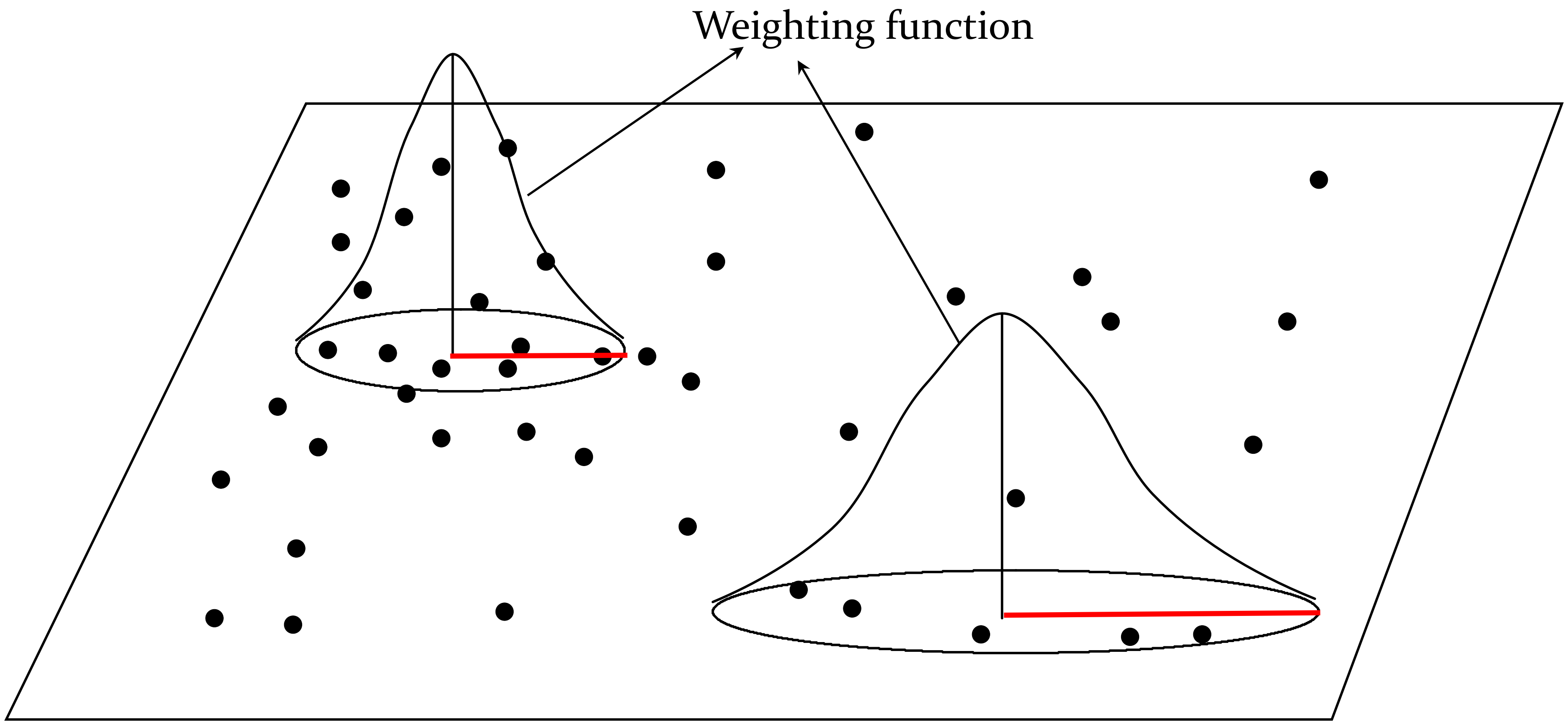
× regression point w_{ij} is the weight of data point j at regression point i

• data point d_{ij} is the distance between regression point i and data point j

Fixed weighting scheme



Adaptive weighting schemes



Why GWR?


GWR is part of a growing trend in GIS towards local analysis
GWR as a "spatial microscope"
GWR Provides useful link to GIS
Why GWR?
Residuals from GWR are generally much lower and usually much less spatially dependent
GWR is widely applicable to almost any form of spatial data
GWR is truly a spatial technique

Regression vs. Geographically Weighted Regression

Modeling relationships

- Regression establishes relationship among a ***dependent*** variable and a set of ***independent*** variable(s)

- A typical linear regression model looks like:

$$y_i=\beta_0 + \beta_1 x_{1i}+ \beta_2 x_{2i}+\ldots\ldots+ \beta_n x_{ni}+\varepsilon_I$$

- With y_i the dependent variable, x_{ji} (j from 1 to n) the set of independent variables, and ε_i the residual, all at location i

Example of OLS

A	B	C	D	E	F	G	H	I	J	K
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
	X_i	Y_i	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})$	$(Y_i - \bar{Y})^2$	$(X_i - \bar{X})(Y_i - \bar{Y})$	$\hat{Y}_i = \hat{\beta} X_i$	$\hat{\varepsilon}_i = Y_i - \hat{Y}_i$	ε_i^2
	12	6	-20.1	404.5	-27.0	729	543.0	-15.1	-11.89	141.3
	34	52	1.9	3.6	19.0	361	35.9	1.4	17.58	309.1
	32	41	-0.1	0.0	8.0	64	-0.9	-0.1	8.08	65.3
	12	25	-20.1	404.5	-8.0	64	160.9	-15.1	7.11	50.6
	11	22	-21.1	445.7	-11.0	121	232.2	-15.9	4.86	23.7
	14	9	-18.1	328.0	-24.0	576	434.7	-13.6	-10.39	108.0
	56	43	23.9	570.7	10.0	100	238.9	18.0	-7.95	63.2
	75	67	42.9	1839.5	34.0	1156	1458.2	32.2	1.77	3.1
	43	32	10.9	118.6	-1.0	1	-10.9	8.2	-9.18	84.3
Sum	**289**	**297**	**0.0**	**4114.9**	**0.0**	**3172**	**3092**	**0.0**	**0.0**	**848.6**
Mean	32.11	33.00								
Variance	514.3611	396.5								
SD	22.68	19.91								
Covariance		386.5								

	Unstandardized Coefficients B	**Std. Error**	**t**
b-HOURS	0.751	0.262	2.866
a-CONSTANT	8.871		

OR

	Unstandardized Coefficients B	**Std. Error**	**t**
b-HOURS	0.751	0.262	2.866
a-CONSTANT	8.871	11.290	0.786

Model Summary

R	**R Square**	**Std. Error of the Estimate**
0.856	0.732	16.819

ANOVA

Model	**Sum of Squares**	**df**	**Mean Square**	**F**	**eta^2**
Regression	2323.383	1	2323.383	8.214	0.732
Residual	848.617	3	282.872		
total	3172	4			

$$\text{OLS} \quad \hat{y} = 8.8719 + .7514(X_1)$$

Example for GWR – Part 1

ID	X-coordinate	Y-coordinate	VAR 1(X)	VAR 2(Y)	DISTANCE	GEOG WT (Wij)
1	25.00	45.00	12	6	0	1
2	25.51	44.14	34	52	1	0.995012479
3	21.87	48.90	32	41	5	0.882496903
4	27.60	52.57	12	25	8	0.726149037
5	16.69	31.33	11	22	16	0.2780373
6	42.52	35.35	14	9	22	0.088921617
7	9.20	65.65	56	43	26	0.034047455
8	29.23	78.72	75	67	32	0.005976023
9	61.37	66.01	43	32	42	0.000147748

Gaussian weighting scheme

$$w_{ij} = \exp\left[-0.5\left(\frac{d_{ij}}{\tau}\right)^2\right]$$

d=distance

τ=bandwidth

Gaussian weighting scheme

$$w_{ij} = \exp\left[-0.5\left(\frac{d_{ij}}{\tau}\right)^2\right] = \exp\left[-0.5\left(\frac{1}{10}\right)^2\right] = 0.995012479$$

d=1

τ=10

$$\beta(x_i) = (Y^T W(x_i) Y)^{-1} * Y^T W(x_i) z$$

$$(Y^T W(x_i) Y)^{-1} = \begin{bmatrix} 1.000 & 0.995 & 0.882 & 0.726 & 0.278 & 0.089 & 0.034 & 0.006 & 0.000 \\ 12.000 & 33.83 & 28.24 & 8.71 & 3.06 & 1.24 & 1.91 & 0.45 & 0.01 \end{bmatrix} X \begin{bmatrix} 1.000 & 12.00 \\ 0.995 & 33.83 \\ 0.882 & 28.24 \\ 0.726 & 8.71 \\ 0.278 & 3.06 \\ 0.089 & 1.24 \\ 0.034 & 1.91 \\ 0.006 & 0.45 \\ 0.000 & 0.01 \end{bmatrix}$$

$$C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$

C_{11}=Top row of Y^T multiplied by left column of Y

C_{12}=First cell of Top row of Y^T multiplied by right column of Y

C_{21}=Bottom row of Y^T multiplied by first cell of left column of Y

C_{22}=Bottom row of Y^T multiplied by right column of Y

C_{11}= (1*1)+(0.995*0.995)+(0.882*0.882)+(0.726*0.726)+(0.278*0.278)+(0.089*0.089)+(0.034*0.034)+(0.006*0.006)+(0.000*0.000)

C_{12}= (1*12)+(1*33.83)+(1*28.24)+(1*8.71)+(1*3.06)+(1*1.24)+(1*1.91)+(1*0.45)+(1*0.01)

C_{21}= (12*1)+(33.83*1)+(28.24*1)+(8.71*1)+(3.06*1)+(1.24*1)+(1.91*1)+(0.45*1)+ (0.01*1)

C_{22}= (12*12)+(33.83*33.83)+(28.24*28.24)+(8.71*8.71)+(3.06*3.06)+(1.24*1.24)+(1.91*1.91)+(0.45*0.45)+(.01*0.01)

$$(Y^T W(x_i) Y) = \begin{bmatrix} 3.38 & 77.94 \\ 77.94 & 2176.66 \end{bmatrix}$$

$$(Y^T W(x_i) Y)^{-1} = \frac{1}{(c_{11} * c_{22}) - (c_{12} * c_{21})} * \begin{bmatrix} C_{22} & -C_{12} \\ -C_{21} & C_{11} \end{bmatrix} = \frac{1}{(3.38 * 2176.66) - (77.94 * 77.94)} * \begin{bmatrix} 2176.66 & -77.94 \\ -77.94 & 3.38 \end{bmatrix}$$

$$(Y^T W(x_i) Y)^{-1} = 0.00084 * \left[\begin{bmatrix} 2176.66 & -77.94 \\ -77.94 & 3.38 \end{bmatrix}\right] = \begin{bmatrix} 1.6930 & -0.061 \\ -0.061 & 0.003 \end{bmatrix}$$

$$\beta(x_i) = (Y^T W(x_i) Y)^{-1} * Y^T W(x_i) z$$

$$Y^T W(x_i) z = \begin{bmatrix} 1.000 & 0.995 & 0.882 & 0.726 & 0.278 & 0.089 & 0.034 & 0.006 & 0.000 \\ 12.000 & 33.83 & 28.24 & 8.71 & 3.06 & 1.24 & 1.91 & 0.45 & 0.01 \end{bmatrix} X \begin{bmatrix} 6.0 \\ 51.7 \\ 36.2 \\ 18.2 \\ 6.1 \\ 0.8 \\ 1.5 \\ 0.4 \\ 0.0 \end{bmatrix}$$

$$C = \begin{bmatrix} C_{12} \\ C_{22} \end{bmatrix}$$

C_{12} =Top row of Y^T multiplied by right column of Y

C_{22} =Bottom row of Y^T multiplied by right column of Y

C_{12} = (1*6)+(.995*51.7)+(.882*36.2)+(.726*18.2)+(.278*6.1)+(.089*0.8)+(.034*1.5)+(.006*0.40)+(.000*0.00)

C_{22} = (12*6)+(33.83*51.74)+(28.24*36.18)+(8.71*18.15)+(3.06*6.12)+(1.24*0.80)+(1.91*1.46)+(0.45*0.40)+(.01*0.00)

$$Y^T W(x_i) z = \begin{bmatrix} 104.4 \\ 3025.1 \end{bmatrix}$$

$$\beta(x_i) = (Y^T W(x_i) Y)^{-1} * Y^T W(x_i) z$$

$$(Y^T W(x_i) Y)^{-1} = \begin{bmatrix} 1.6930 & -0.061 \\ -0.061 & 0.003 \end{bmatrix}$$

$$Y^T W(x_i) z = \begin{bmatrix} 104.4 \\ 3025.06 \end{bmatrix}$$

$$\beta(x_i) = (Y^T W(x_i) Y)^{-1} * Y^T W(x_i) z = \begin{bmatrix} 1.6930 & -0.061 \\ -0.061 & 0.003 \end{bmatrix} * \begin{bmatrix} 104.4 \\ 3025.06 \end{bmatrix}$$

$$1.6930 * 104.4 = 176.451$$
$$-0.061 * 3025.06 = -183.038$$
$$176.451 + (-183.038) = -6.587$$

$$-0.061 * 104.4 = -6.318$$
$$0.003 * 3025.06 = 7.944$$
$$-6.318 + 7.944 = 1.626$$

GWR $\hat{y}_i = -6.587 + 1.626(X_1)$

Comparison of Predicated Values

OLS $\hat{y} = 8.8719 + .7514(12)$

GWR $\hat{y}_i = -6.587 + 1.626(12)$

OLS $\hat{y} = 8.8719 + 9.017$

GWR $\hat{y}_i = -6.587 + 19.51$

OLS $\hat{y} = 17.888$

GWR $\hat{y}_i = 12.92$

Residual=17.888-6=11.888

Residual=12.92-6=6.92

Goodness of Fit

$$r_i^2 = \frac{TSS_1 - RSS_1}{RSS_1}$$

$$TSS_1 = \sum_{j=1}^{n} w_{ij}(z_j - \bar{z})^2$$

$$RSS_1 = \sum_{j=1}^{n} w_{ij}(z_j - \hat{z})^2$$

$$r_i^2 = \frac{1286.327 - 291.167}{1286.327} = .774 = 77\%$$

The unweighted $r^2 = .732 = 73\%$

Obs .(j)	$W_{ij}(z_j - \bar{z})^2$	$\hat{z}$	$W_{ij}(z_j - \hat{z})^2$
1	729.000	12.920	47.893
2	359.200	48.684	10.938
3	56.480	45.433	17.344
4	46.474	12.920	105.956
5	33.643	11.295	31.863
6	51.219	16.172	4.574
7	3.405	84.448	58.493
8	6.908	115.336	13.962
9	0.000	63.315	0.145
Total	1286.327		291.167

Note: Z bar = 33

Example for GWR – Part 3

- Addresses the non-stationarity directly
  - Allows the relationships to vary over space, i.e., βs do not need to be the same throughout the study area.

$$y_i = \beta_i 0 + \beta_{i1} X_{1i} + \beta_{i2} X_{2i} + \ldots\ldots + \beta_{in} X_{ni} + \varepsilon_i$$

$$y_i = \sum_k \beta(u_{i,} v_i) x_{k,i} + \varepsilon_i,$$

y_i = dependent variable

$x_{k,i}$ = independent variable

ε_i = Error Term

$u_{i,} v_i$ = is the x and y coordiante of the ith location

$\beta(u_{i,} v_i)$ = varying coefficients

Instead of remaining the same everywhere, βs now vary in terms of locations (i)

Example for GWR – Part 4

- Now we need compare the GWR and OLS models
  - Focus on the The Akaike Information Criterion (AIC)
    - Small AIC indicates a better model
      - Rule of thumbs: a decrease of AIC of 3 is regarded as a successful improvement
- Once we have confirmed that GWR is better than OLS we may want to study if coefficients really varying across space

Local t values

$t_i = \beta_i / SE\beta_i$

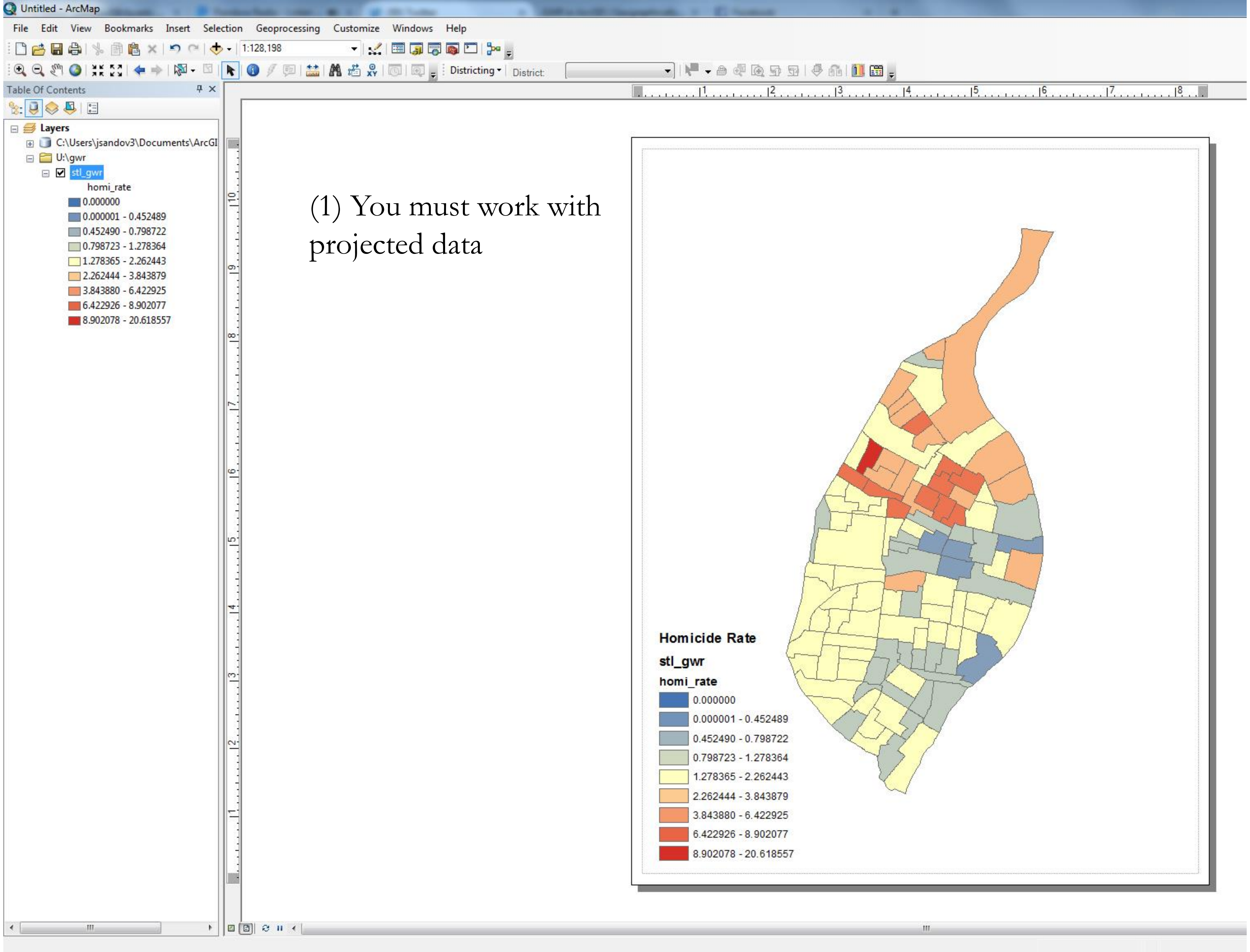
Look for areas on the map with:

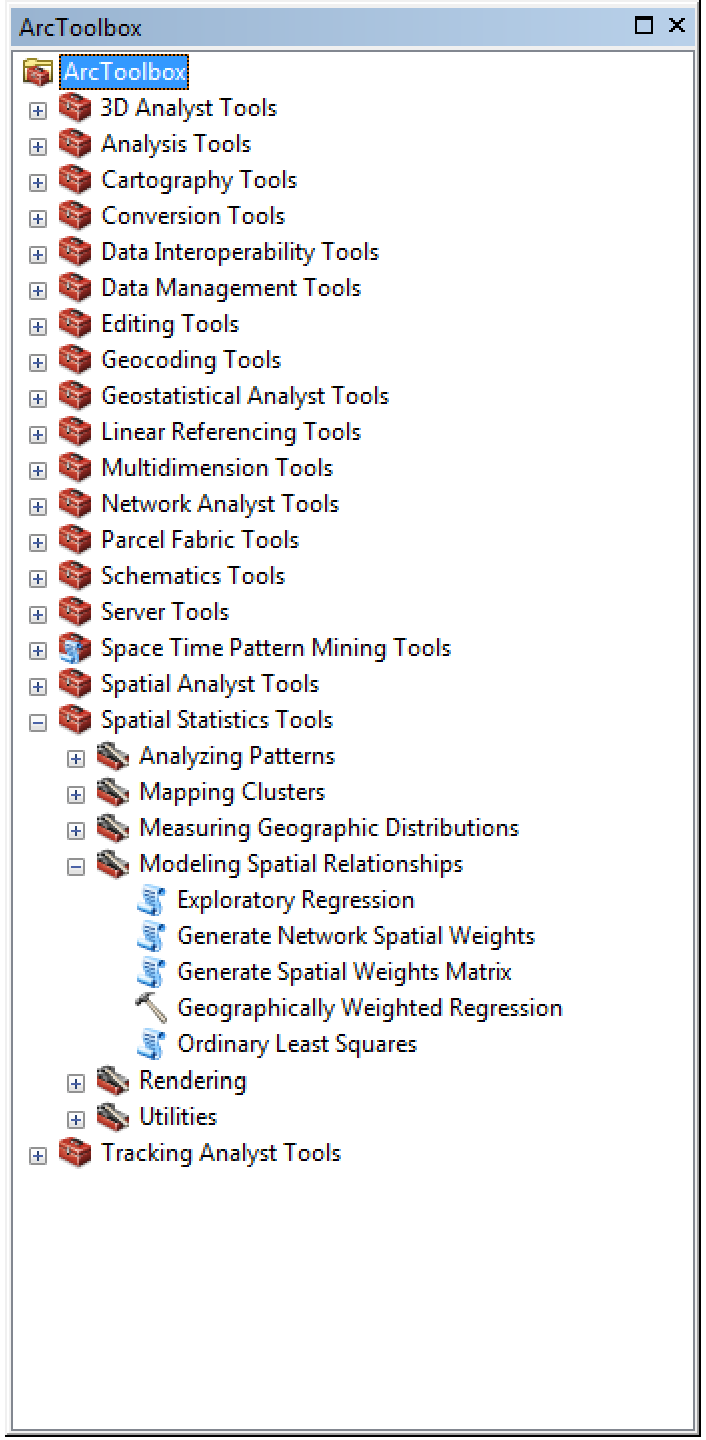
t_i values $\geq$ 1.96 and/or t_i values $\leq$ -1.96

Map these values

Lab Example

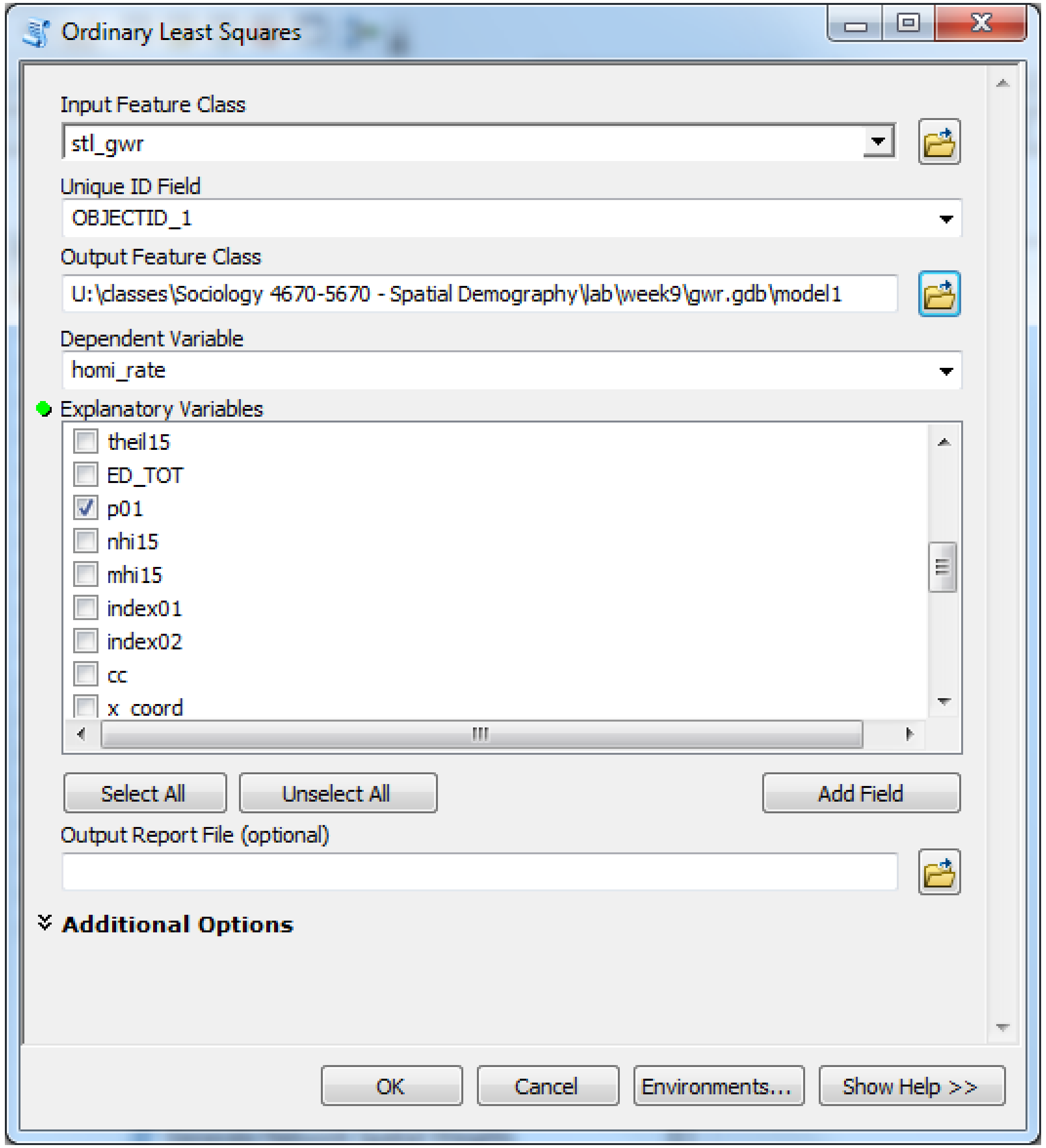
(1) You must work with projected data





(2) Go to spatial statistics – then selected modeling spatial relationships

(3) We will compute a OLS model first. We need a model for comparison.



(4)We need a unique ID filed

(5) Name your output file

(6) Select your dependent variable

(7) Select your independent variables

Results

Current Session

Ordinary Least Squares [083517_04062017]

Output Feature Class: model1

Coefficient Output Table: <empty>

Diagnostic Output Table: <empty>

Output Report File: <empty>

Inputs

Environments

Messages

Executing: OrdinaryLeastSquares stl_gwr OBJECTID_1 "U:\classes\Sociology 4670-5670 - Spatial Demography\lab\week9\gwr.gdb\model1" homi_rate p01 # # #

Start Time: Thu Apr 06 08:35:08 2017

Running script OrdinaryLeastSquares...

Summary of OLS Results

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]
Intercept	-0.251479	0.713962	-0.352230	0.725386	0.500302	-0.502653	0.616277
P01	12.205743	2.160004	5.650796	0.000000*	2.131811	5.725529	0.000000*

OLS Diagnostics

Input Features: stl_gwr Dependent Variable: HOMI_RATE

Number of Observations: 106 Akaike's Information Criterion (AICc) [d]: 568.207874

Multiple R-Squared [d]: 0.234909 Adjusted R-Squared [d]: 0.227552

Joint F-Statistic [e]: 31.931496 Prob(>F), (1,104) degrees of freedom: 0.000000*

Joint Wald Statistic [e]: 32.781679 Prob(>chi-squared), (1) degrees of freedom: 0.000000*

Koenker (BP) Statistic [f]: 5.544379 Prob(>chi-squared), (1) degrees of freedom: 0.018540*

Jarque-Bera Statistic [g]: 139.634448 Prob(>chi-squared), (2) degrees of freedom: 0.000000*

Notes on Interpretation

* An asterisk next to a number indicates a statistically significant p-value (p < 0.01).

[a] Coefficient: Represents the strength and type of relationship between each explanatory variable and the dependent variable.

[b] Probability and Robust Probability (Robust_Pr): Asterisk (*) indicates a coefficient is statistically significant (p < 0.01); if the Koenker (BP) Statistic [f] is statistically s

[c] Variance Inflation Factor (VIF): Large Variance Inflation Factor (VIF) values (> 7.5) indicate redundancy among explanatory variables.

[d] R-Squared and Akaike's Information Criterion (AICc): Measures of model fit/performance.

[e] Joint F and Wald Statistics: Asterisk (*) indicates overall model significance (p < 0.01); if the Koenker (BP) Statistic [f] is statistically significant, use the Wald Statisti

[f] Koenker (BP) Statistic: When this test is statistically significant (p < 0.01), the relationships modeled are not consistent (either due to non-stationarity or heterosked

[g] Jarque-Bera Statistic: When this test is statistically significant (p < 0.01) model predictions are biased (the residuals are not normally distributed).

WARNING 000851: Use the Spatial Autocorrelation (Moran's I) Tool to ensure residuals are not spatially autocorrelated.

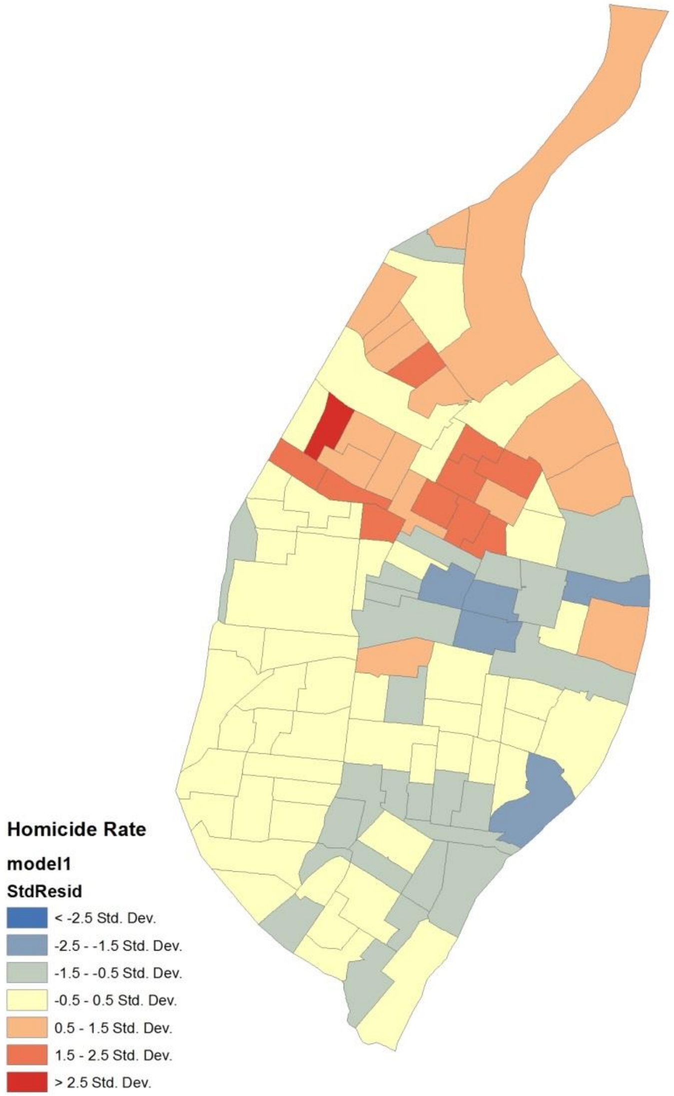
Completed script OrdinaryLeastSquares...

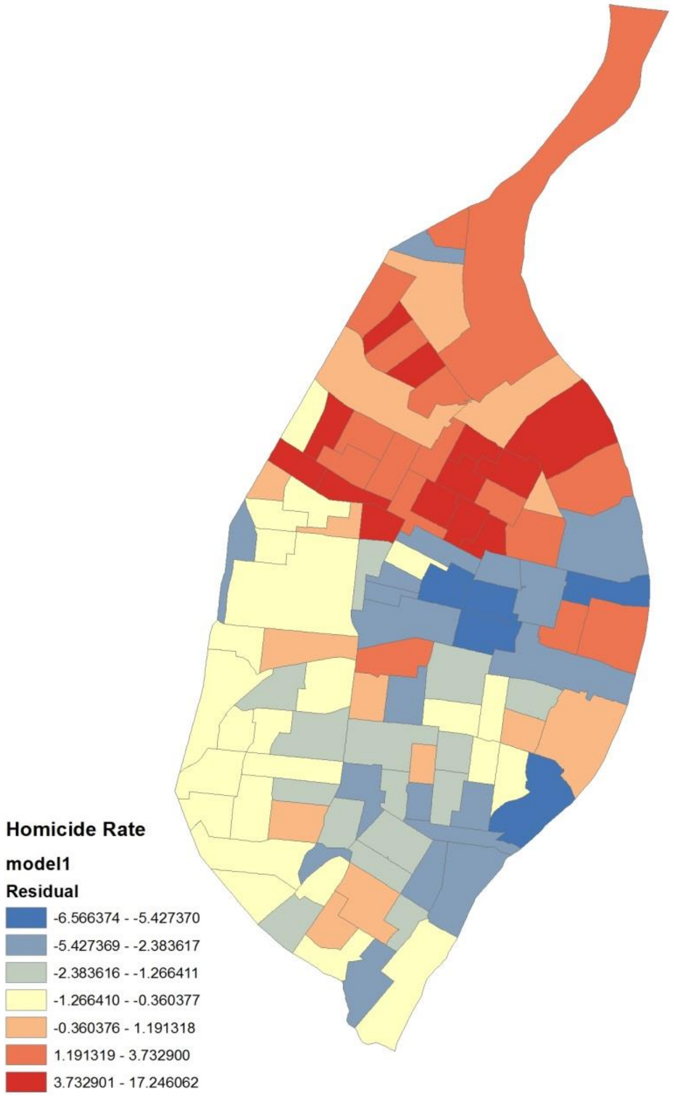
Succeeded at Thu Apr 06 08:35:17 2017 (Elapsed Time: 9.43 seconds)

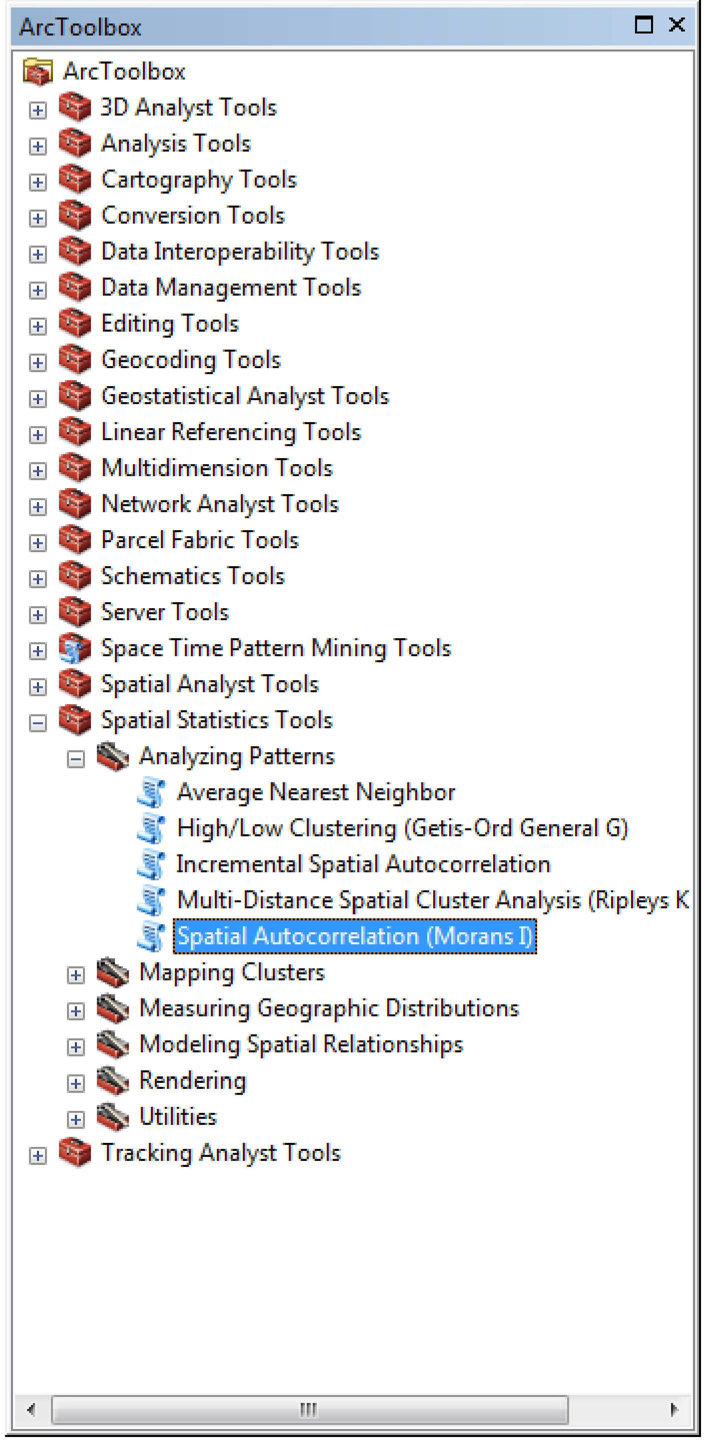
Shared

(8) We need to examine the residuals

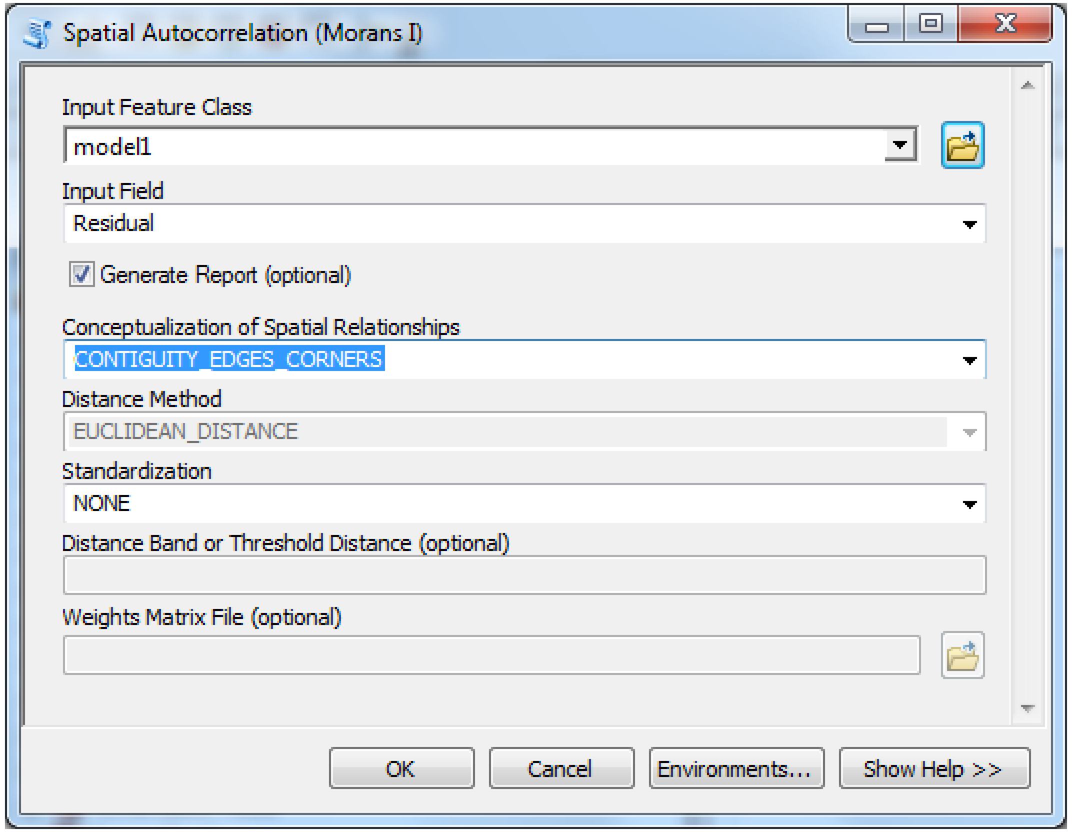
(9) The maps show some spatial clustering of residuals



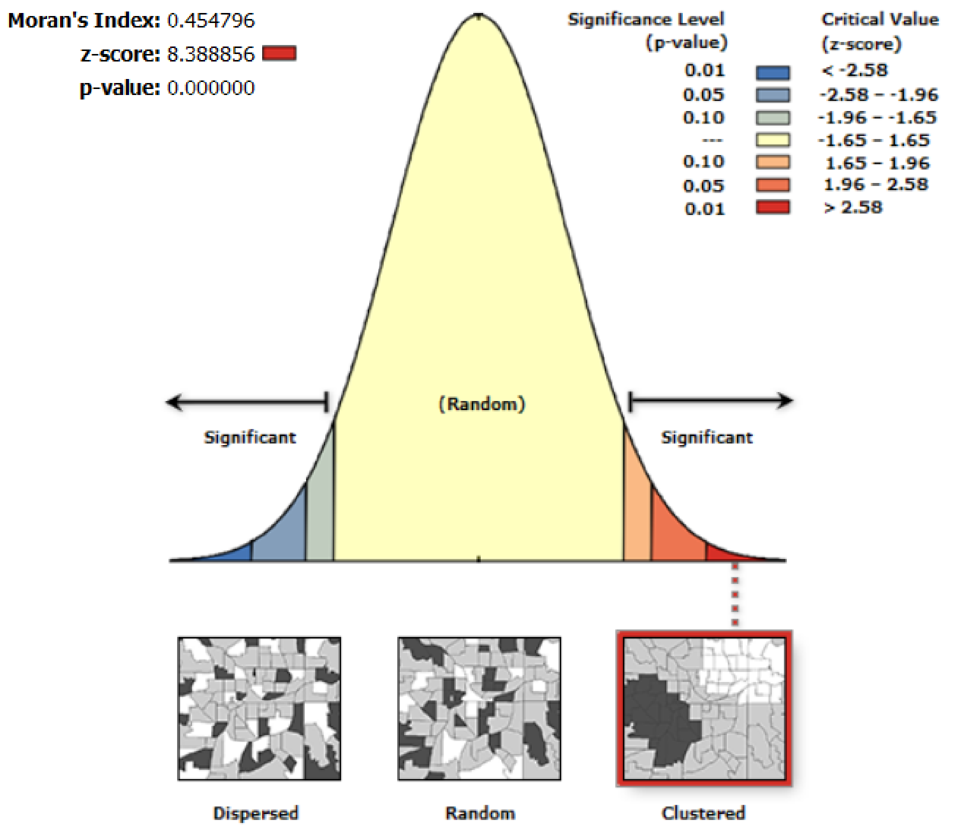




(10) Let's compute a global spatial autocorrelation statistic



Spatial Autocorrelation Report



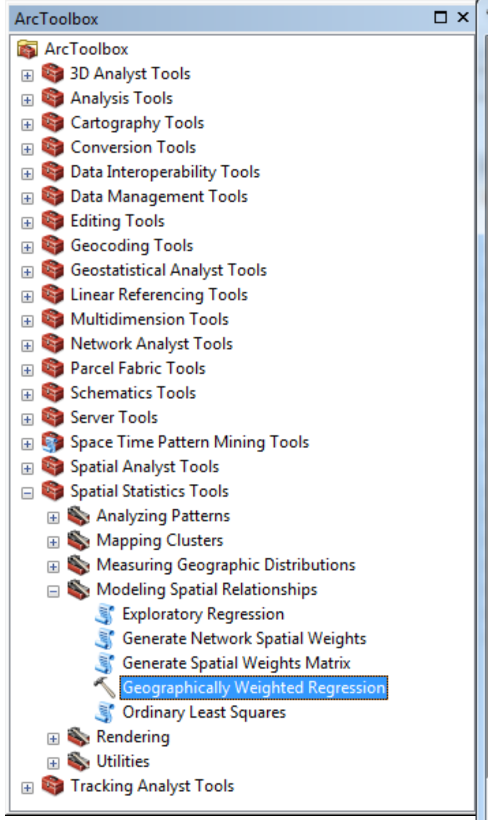
Given the z-score of 8.38885623016, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

Global Moran's I Summary

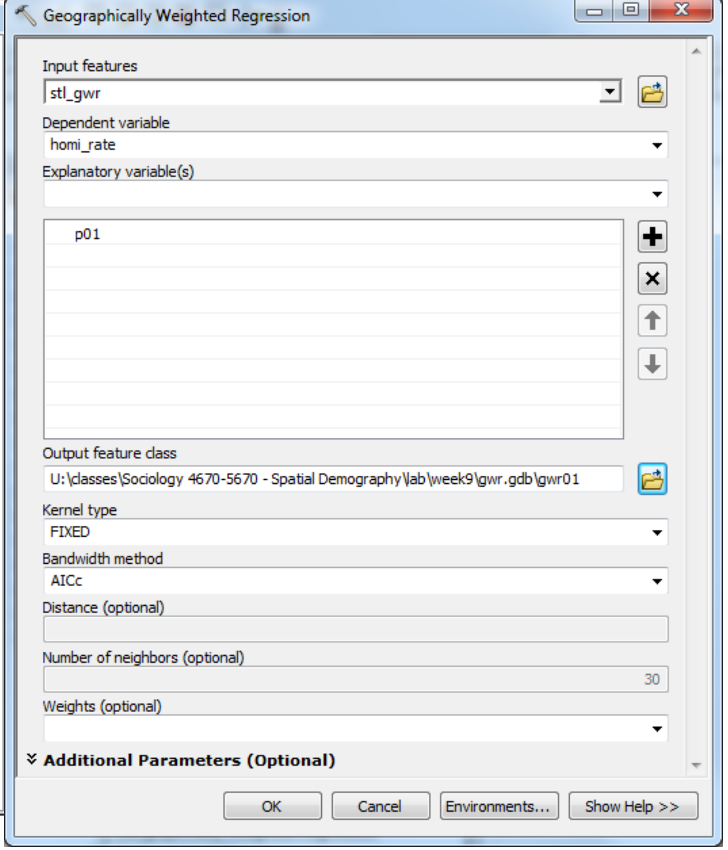
Moran's Index:	0.454796
Expected Index:	-0.009524
Variance:	0.003064
z-score:	8.388856
p-value:	0.000000

(11) We have confirmation of spatial autocorrelation

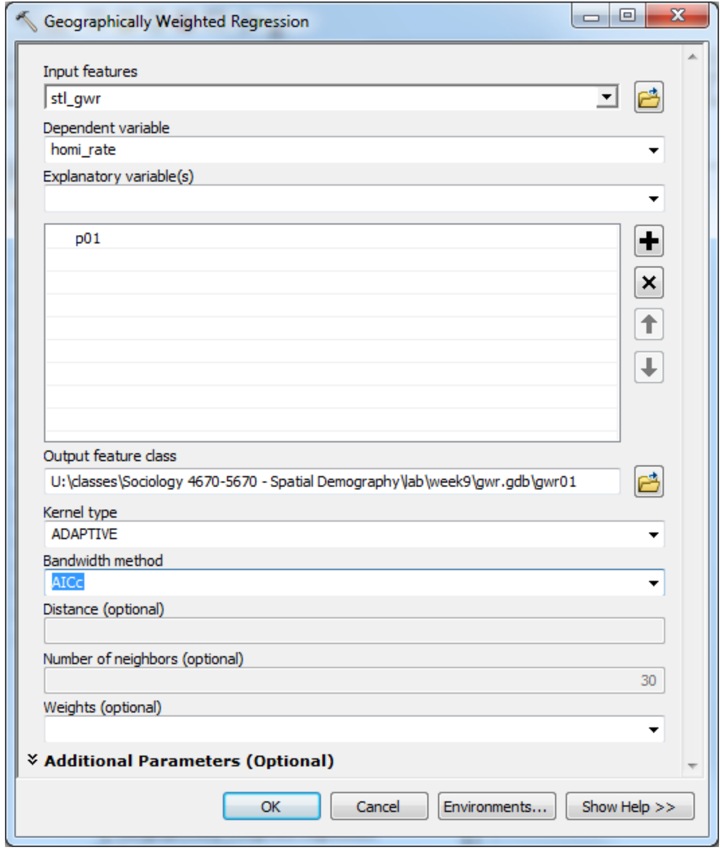
(12) We can proceed with GWR



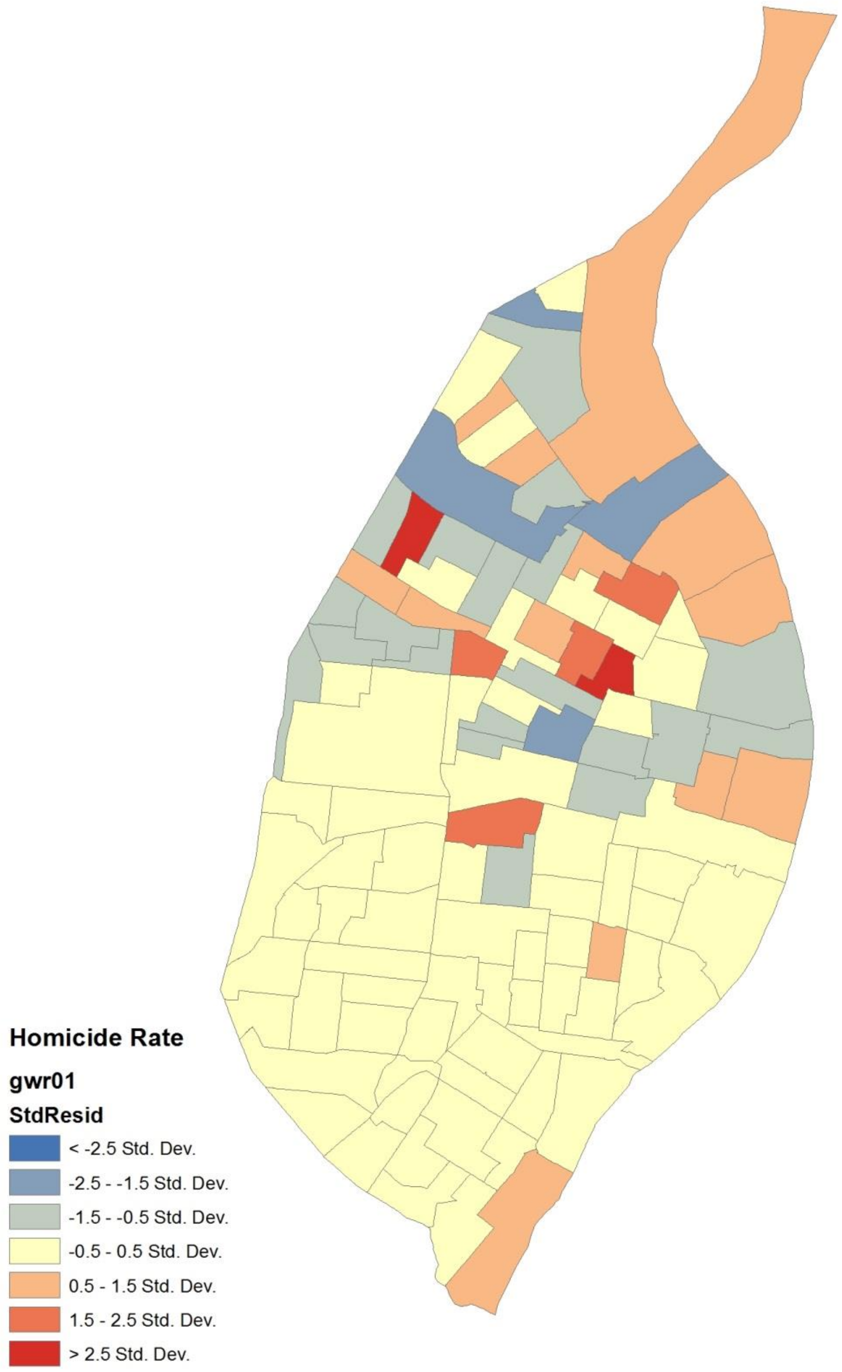
(13) Select GWR from the Modeling Spatial Relationship menu



(15) When you have irregular polygons select adaptive. Fixed method is for grid shapefiles. Adaptive schemes adjust itself according to the density of data



(16) Program will determine the optimal bandwidth. AIC and CV will produce similar results. AIC is used as a default



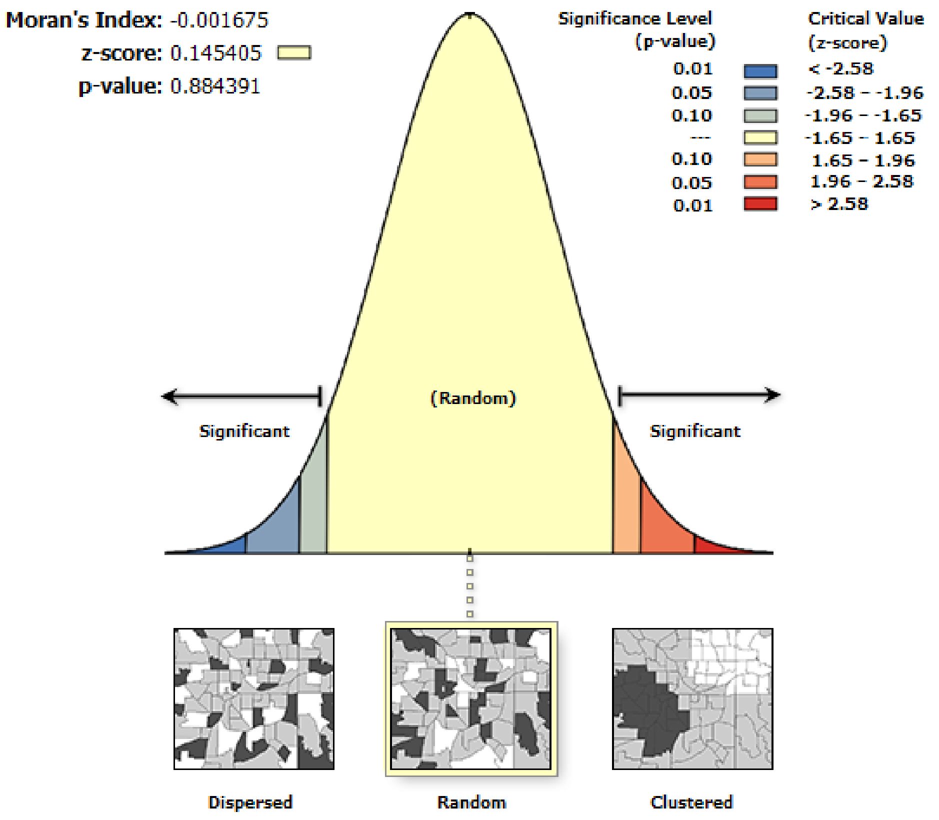
(17) The default map should be standardized residuals

(18) The maps looks better than OLS

(19) Let's compute a spatial autocorrelation statistic.

(20)Go to your new shapefile that was created and select “residual” for the input field. We need select the type of relationship. I will chose polygon_contiguity.

Spatial Autocorrelation Report

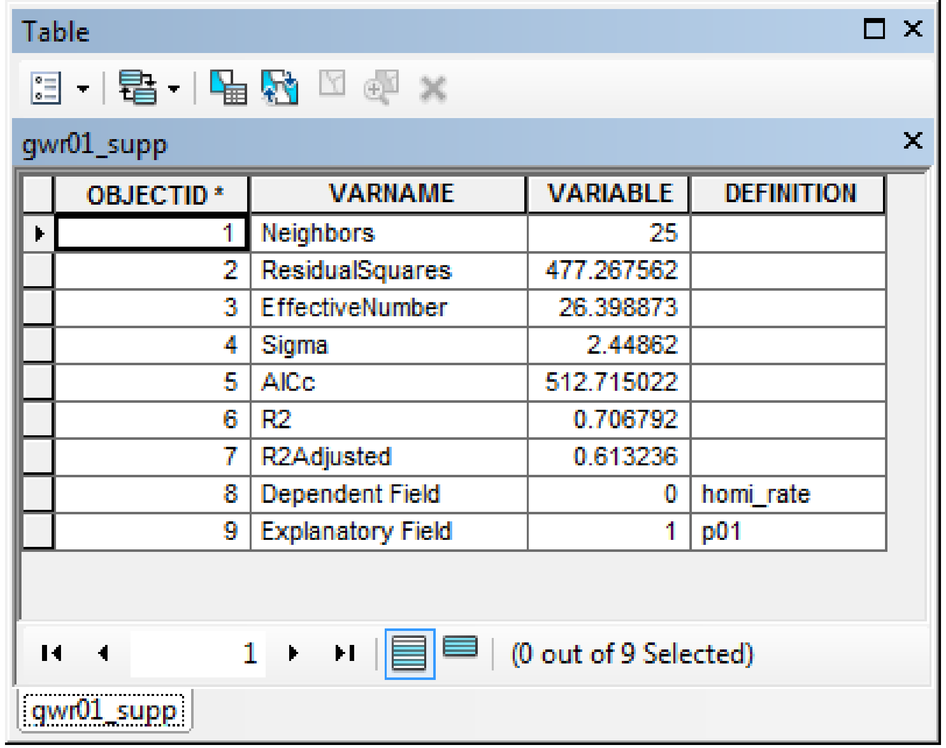


Given the z-score of 0.145404619907, the pattern does not appear to be significantly different than random.

Global Moran's I Summary

Moran's Index:	-0.001675
Expected Index:	-0.009524
Variance:	0.002914
z-score:	0.145405
p-value:	0.884391

(21) The results show we no longer have a problem with spatial autocorrelation. The GWR models has fixed the problem



OBJECTID *	VARNAME	VARIABLE	DEFINITION
1	Neighbors	25	
2	ResidualSquares	477.267562	
3	EffectiveNumber	26.398873	
4	Sigma	2.44862	
5	AICc	512.715022	
6	R2	0.706792	
7	R2Adjusted	0.613236	
8	Dependent Field	0	homi_rate
9	Explanatory Field	1	p01

ResidualSquares is the sum of the squared residuals in the model (the residual being the difference between an observed *y* value and its estimated value returned by the GWR model). The smaller this measure, the closer the fit of the GWR model to the observed data. This value is used in a number of other diagnostic measures.

Bandwidth or Neighbors: This is the bandwidth or number of neighbors used for each local estimation and is perhaps the most important parameter for Geographically Weighted Regression. It controls the degree of smoothing in the model.

Effective Number reflects a tradeoff between the variance of the fitted values and the bias in the coefficient estimates and is related to the choice of bandwidth. As the bandwidth approaches infinity, the geographic weights for every observation approach 1, and the coefficient estimates will be very close to those for a global OLS model. As the bandwidth approaches 0, the local model 'wraps itself around the data' so the number of parameters = n. The number of parameters in local models therefore ranges between k and n and depends on the bandwidth. This number need not be an integer.

Table

gwr01_supp

OBJECTID *	VARNAME	VARIABLE	DEFINITION
1	Neighbors	25	
2	ResidualSquares	477.267562	
3	EffectiveNumber	26.398873	
4	Sigma	2.44862	
5	AICc	512.715022	
6	R2	0.706792	
7	R2Adjusted	0.613236	
8	Dependent Field	0	homi_rate
9	Explanatory Field	1	p01

1 (0 out of 9 Selected)

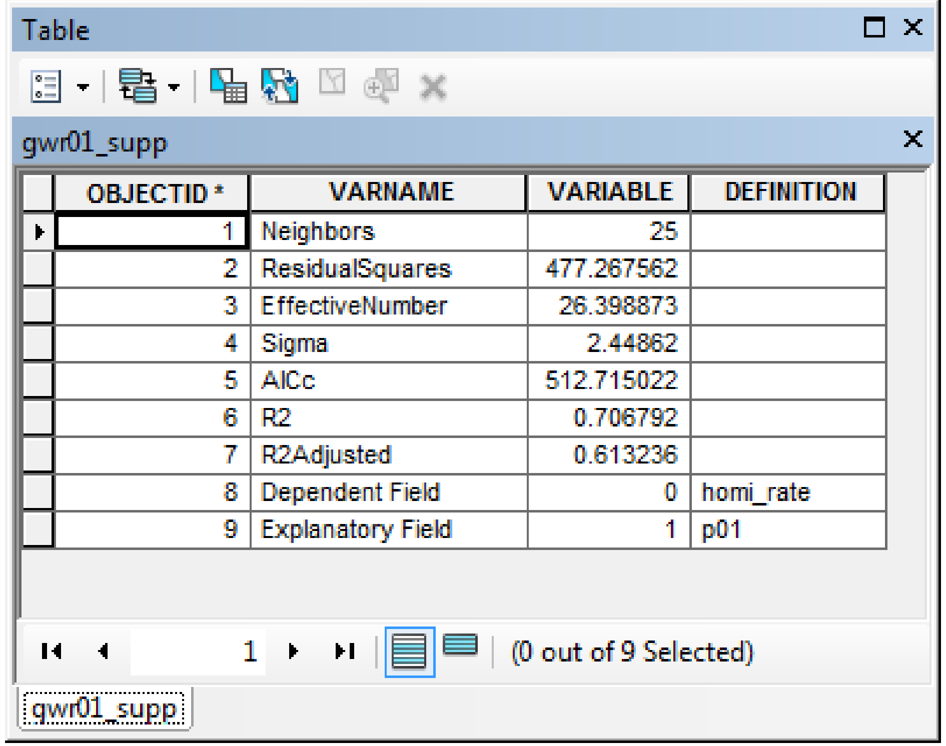
gwr01_supp

Sigma is the square root of the normalized residual sum of squares, where the residual sum of squares is divided by the effective degrees of freedom of the residual. This is the estimated standard deviation for the residuals. Smaller values of this statistic are preferable. Sigma is used for AICc computations.

AICc is a measure of model performance and is helpful for comparing different regression models. Taking into account model complexity, the model with the lower AICc value provides a better fit to the observed data.

AICc is not an absolute measure of goodness of fit but is useful for comparing models with different explanatory variables as long as they apply to the same dependent variable. If the AICc values for two models differ by more than 3, the model with the lower AICc is held to be better.

Comparing the GWR AICc value to the OLS AICc value is one way to assess the benefits of moving from a global model (OLS) to a local regression model (GWR).



OBJECTID *	VARNAME	VARIABLE	DEFINITION
1	Neighbors	25	
2	ResidualSquares	477.267562	
3	EffectiveNumber	26.398873	
4	Sigma	2.44862	
5	AICc	512.715022	
6	R2	0.706792	
7	R2Adjusted	0.613236	
8	Dependent Field	0	homi_rate
9	Explanatory Field	1	p01

R-Squared is a measure of goodness of fit. Its value varies from 0.0 to 1.0, with higher values being preferable. It may be interpreted as the proportion of dependent variable variance accounted for by the regression model.

R2Adjusted normalizes the numerator and denominator by their degrees of freedom. This has the effect of compensating for the number of variables in a model, and consequently, the Adjusted R2 value is almost always smaller than the R2 value. However, in making this adjustment, you lose the interpretation of the value as a proportion of the variance explained. In GWR, the effective number of degrees of freedom is a function of the bandwidth, so the adjustment may be quite marked in comparison to a global model like OLS. For this reason, the AICc is preferred as a means of comparing models.

Table

gwr01

OBJECTID *	Shape *	Observed homi_rate	Condition Number	Local R2	Predicted	Coefficient Intercept	Coefficient #1 p01	Residual
1	Polygon	0.451264	4.31118	0.588676	-0.132836	-2.643039	20.17888	0.5841
2	Polygon	1.338987	4.499329	0.338956	1.406707	0.107574	5.024883	-0.06772
3	Polygon	8.130081	7.53757	0.051322	7.563721	6.057966	5.776464	0.56636
4	Polygon	1.252348	5.005241	0.162122	0.513364	-0.544289	8.837773	0.738984
5	Polygon	10.440835	5.884902	0.110382	7.788634	2.121421	13.780361	2.652201
6	Polygon	3.007016	5.527429	0.322142	5.500325	-1.622877	22.541683	-2.493309
7	Polygon	0.335909	4.678937	0.625841	-0.449348	-3.684966	26.131135	0.785257
8	Polygon	2.857143	4.955804	0.387266	4.406599	-2.345297	23.467192	-1.549457
9	Polygon	4.27899	5.333463	0.228323	6.155438	-0.412801	19.057721	-1.876448
10	Polygon	0.690608	6.904027	0.237306	1.065382	-0.864312	10.664873	-0.374774
11	Polygon	0.287109	4.414008	0.59341	2.584693	-3.220348	24.247046	-2.297584
12	Polygon	6.31136	7.09357	0.033319	7.287654	5.227909	7.869118	-0.976294
13	Polygon	12.873025	6.644802	0.210582	10.31639	0.325735	19.059621	2.556635
14	Polygon	20.618557	6.81194	0.000131	8.493684	8.63813	-0.486505	12.124873
15	Polygon	9.515571	6.885774	0.025509	7.187361	4.858348	5.930739	2.32821
16	Polygon	7.432432	7.725377	0.07761	8.681679	4.163018	10.395041	-1.249247
17	Polygon	7.412399	7.065058	0.018042	7.75069	5.353531	4.978844	-0.338291
18	Polygon	9.090909	7.633462	0.018374	6.589724	5.380113	4.928046	2.501185
19	Polygon	5.91716	8.579822	0.025033	7.250353	5.363808	5.453648	-1.333193

1 (0 out of 106 Selected)

gwr01

Condition Number: This diagnostic evaluates local multicollinearity. In the presence of strong local multicollinearity, results become unstable. Results associated with condition numbers larger than 30 may be unreliable.

Local R2: These values range between 0.0 and 1.0 and indicate how well the local regression model fits observed *y* values. Very low values indicate that the local model is performing poorly. Mapping the Local R2 values to see where GWR predicts well and where it predicts poorly may provide clues about important variables that may be missing from the regression model.

Predicted: These are the estimated (or fitted) *y* values computed by GWR.

Residuals: To obtain the residual values, the fitted *y* values are subtracted from the observed *y* values. Standardized residuals have a mean of zero and a standard deviation of 1. A cold-to-hot rendered map of standardized residuals is automatically added to the table of contents when GWR is executed in ArcMap.

Table

gwr01

OBJECTID *	Shape *	Observed homi_rate	Condition Number	Local R2	Predicted	Coefficient Intercept	Coefficient #1 p01	Residual	Standard Error	Standard Error Intercept	Standard Error Coefficient #1 p01	Std. Residual
1	Polygon	0.451264	4.31118	0.588676	-0.132836	-2.643039	20.17888	0.5841	2.017031	1.557216	5.453229	0.289584
2	Polygon	1.338987	4.499329	0.338956	1.406707	0.107574	5.024883	-0.06772	2.289466	1.373432	4.512127	-0.029579
3	Polygon	8.130081	7.53757	0.051322	7.563721	6.057966	5.776464	0.56636	1.966716	2.481051	5.835509	0.287973
4	Polygon	1.252348	5.005241	0.162122	0.513364	-0.544289	8.837773	0.738984	2.121936	1.53536	7.638726	0.348259
5	Polygon	10.440835	5.884902	0.110382	7.788634	2.121421	13.780361	2.652201	2.267509	1.834919	4.861685	1.169654
6	Polygon	3.007016	5.527429	0.322142	5.500325	-1.622877	22.541683	-2.493309	2.290001	1.581085	4.36914	-1.088781
7	Polygon	0.335909	4.678937	0.625841	-0.449348	-3.684966	26.131135	0.785257	1.913597	1.665024	5.191193	0.410356
8	Polygon	2.857143	4.955804	0.387266	4.406599	-2.345297	23.467192	-1.549457	2.284595	1.535008	4.347613	-0.67822
9	Polygon	4.27899	5.333463	0.228323	6.155438	-0.412801	19.057721	-1.876448	2.277912	1.695667	4.621847	-0.823758
10	Polygon	0.690608	6.904027	0.237306	1.065382	-0.864312	10.664873	-0.374774	1.72218	1.962315	14.362812	-0.217616
11	Polygon	0.287109	4.414008	0.59341	2.584693	-3.220348	24.247046	-2.297584	2.219301	1.526258	4.904117	-1.035274
12	Polygon	6.31136	7.09357	0.033319	7.287654	5.227909	7.869118	-0.976294	2.021434	2.29597	5.708776	-0.482971
13	Polygon	12.873025	6.644802	0.210582	10.31639	0.325735	19.059621	2.556635	1.898577	1.898701	5.045844	1.346606
14	Polygon	20.618557	6.81194	0.000131	8.493684	8.63813	-0.486505	12.124873	2.145897	2.423863	6.059281	5.650258
15	Polygon	9.515571	6.885774	0.025509	7.187361	4.858348	5.930739	2.32821	2.225425	2.221629	6.186149	1.046187
16	Polygon	7.432432	7.725377	0.07761	8.681679	4.163018	10.395041	-1.249247	2.17372	2.396298	6.22246	-0.574705
17	Polygon	7.412399	7.065058	0.018042	7.75069	5.353531	4.978844	-0.338291	1.904011	2.203481	6.115439	-0.177673
18	Polygon	9.090909	7.633462	0.018374	6.589724	5.380113	4.928046	2.501185	2.004297	2.41797	6.818915	1.247911
19	Polygon	5.91716	8.579822	0.025033	7.250353	5.363808	5.453648	-1.333193	2.225692	2.692044	7.558463	-0.599002
20	Polygon	3.307607	6.060814	0.075116	6.220799	1.49763	8.859218	-2.913191	2.088896	1.991157	4.655688	-1.394608
21	Polygon	8.814887	6.697146	0.229389	9.285152	0.982932	15.572804	-0.470264	1.943642	2.008165	5.183003	-0.24195
22	Polygon	8.902077	6.940494	0.092082	6.952345	4.054921	9.086756	1.949732	2.154513	2.244369	5.47233	0.904953
23	Polygon	10.744986	6.794547	0.003473	7.061803	6.353978	1.793415	3.683182	2.27561	2.069228	4.712774	1.618547
24	Polygon	12.903226	7.178511	0.00346	6.410498	7.470242	-1.874232	6.492728	2.107626	2.233384	4.963564	3.080588
25	Polygon	4.255319	5.606735	0.558305	6.787941	-3.915403	27.825195	-2.532621	2.228149	1.694202	4.916288	-1.136648
26	Polygon	0.941915	4.547897	0.449051	0.831758	-0.097962	3.334636	0.110157	2.208808	1.484344	5.441741	0.049872

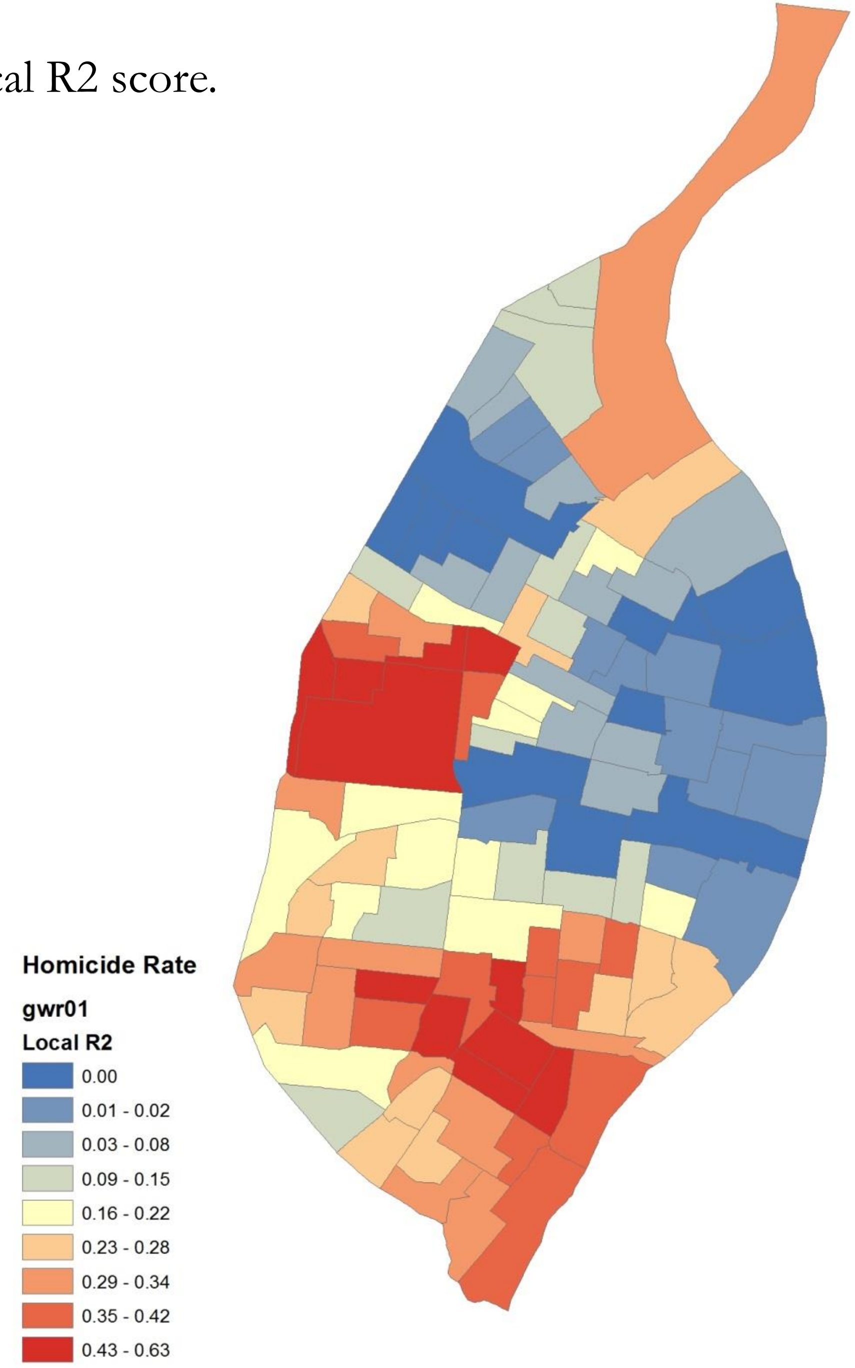
1 (0 out of 106 Selected)

gwr01

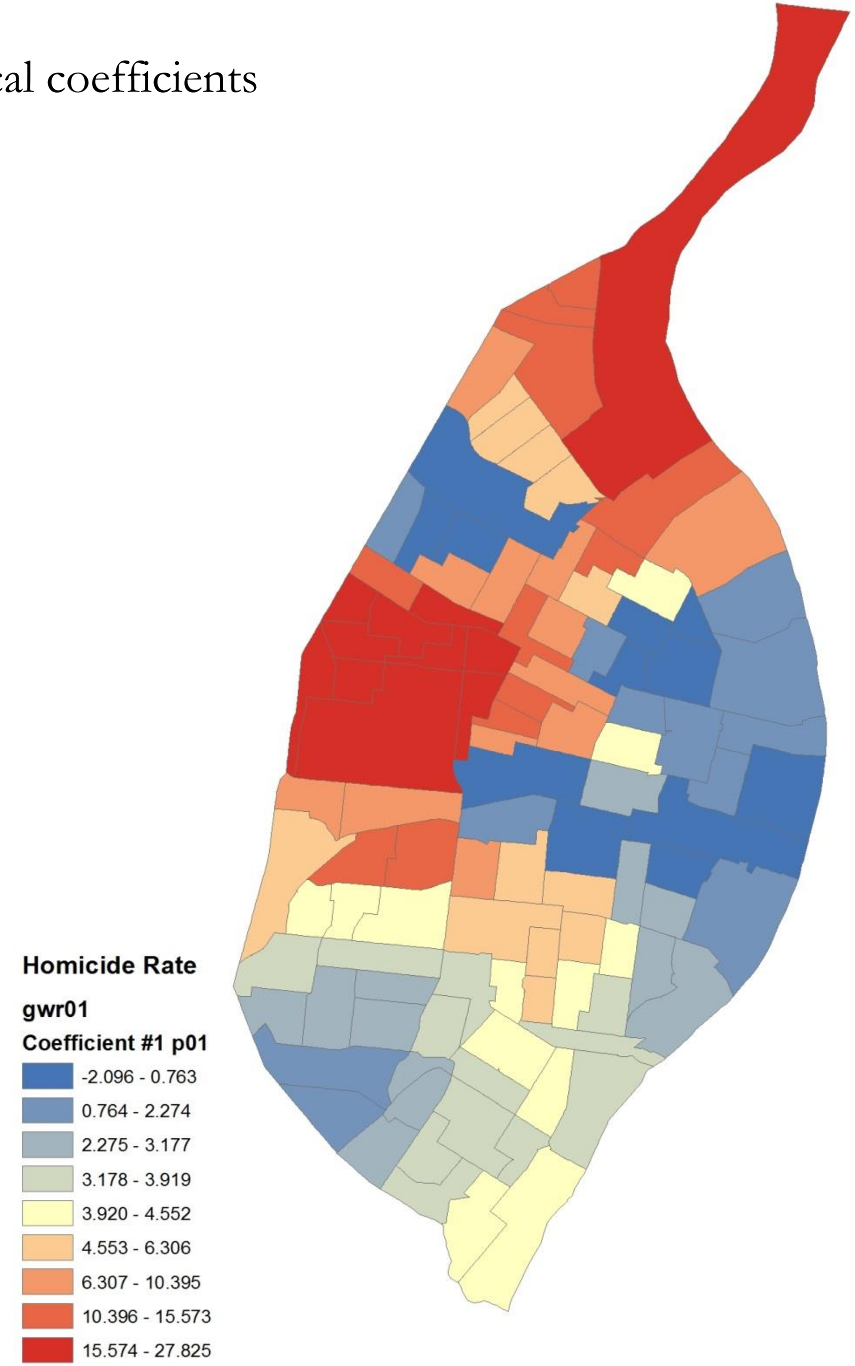
Coefficient Standard Error: These values measure the reliability of each coefficient estimate.

Confidence in those estimates is higher when standard errors are small in relation to the actual coefficient values. Large standard errors may indicate problems with local multicollinearity.

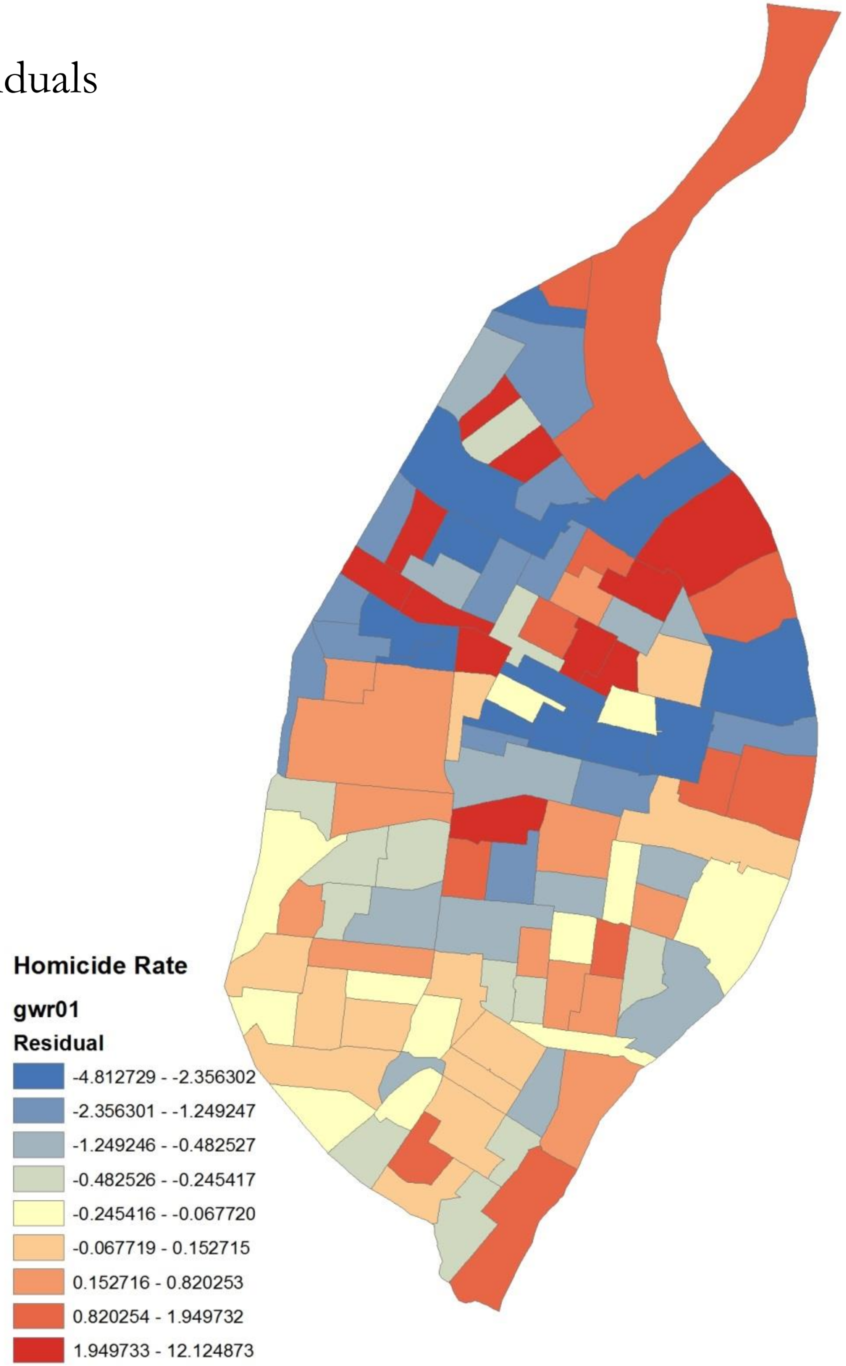
(22) Let's map the local R2 score.



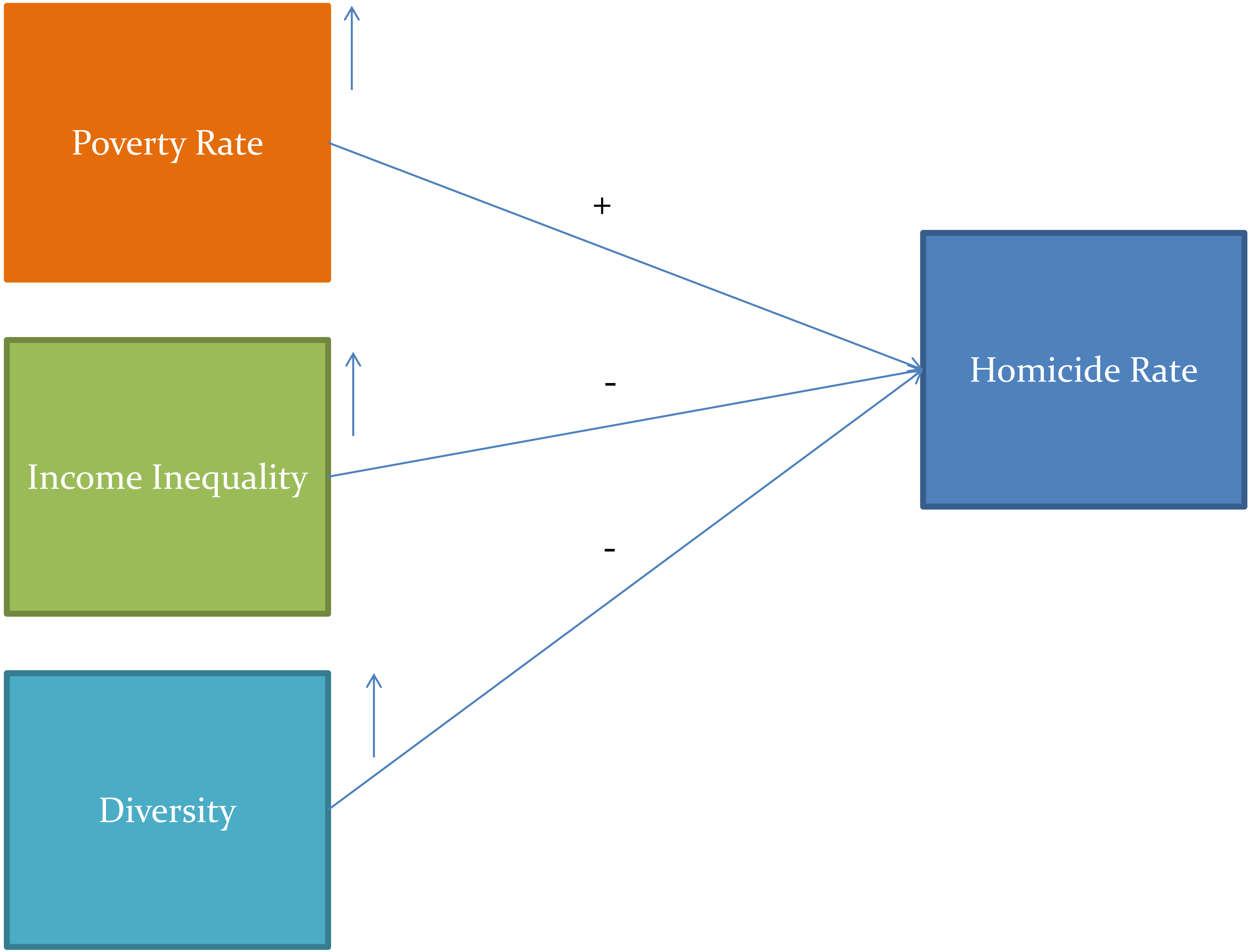
(23) Let’s map the local coefficients



(24) Let's map the residuals



Multivariate Models


Poverty Rate
Income Inequality
Diversity
Homicide Rate
+
-
-

Regression Report

```
>>04/06/17 09:57:54
REGRESSION
----------
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set            :  stl_reg_geoda
Dependent Variable  :   homi_rate  Number of Observations:  106
Mean dependent var  :     3.30795  Number of Variables   :    4
S.D. dependent var  :     3.91868  Degrees of Freedom    :  102

R-squared           :    0.503396  F-statistic           :      34.465
Adjusted R-squared  :    0.488790  Prob(F-statistic)     :1.82372e-015
Sum squared residual:     808.344  Log likelihood        :     -258.08
Sigma-square        :     7.92495  Akaike info criterion :     524.159
S.E. of regression  :     2.81513  Schwarz criterion     :     534.813
Sigma-square ML     :     7.62589
S.E of regression ML:      2.7615

-----------------------------------------------------------------------------
       Variable      Coefficient      Std.Error    t-Statistic   Probability
-----------------------------------------------------------------------------
          CONSTANT       4.79579        1.14249        4.19766      0.00006
               p01       10.4424        1.99592        5.23189      0.00000
           theil15       -3.3254        2.93936       -1.13133      0.26057
               e15      -7.40396         1.0404       -7.11642      0.00000
-----------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER   10.442998
TEST ON NORMALITY OF ERRORS
TEST                  DF            VALUE              PROB
Jarque-Bera            2            222.7096           0.00000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                  DF            VALUE              PROB
Breusch-Pagan test     3             35.9462           0.00000
Koenker-Bassett test   3              8.3391           0.03950

DIAGNOSTICS FOR SPATIAL DEPENDENCE
FOR WEIGHT MATRIX : stl_reg_geoda_queen
   (row-standardized weights)
TEST                           MI/DF        VALUE          PROB
Moran's I (error)              0.1489        3.0741        0.00211
Lagrange Multiplier (lag)        1          23.8924        0.00000
Robust LM (lag)                  1          23.3131        0.00000
Lagrange Multiplier (error)      1           6.1608        0.01306
Robust LM (error)                1           5.5814        0.01815
Lagrange Multiplier (SARMA)      2          29.4738        0.00000
============================== END OF REPORT ================================
```

Regression Report

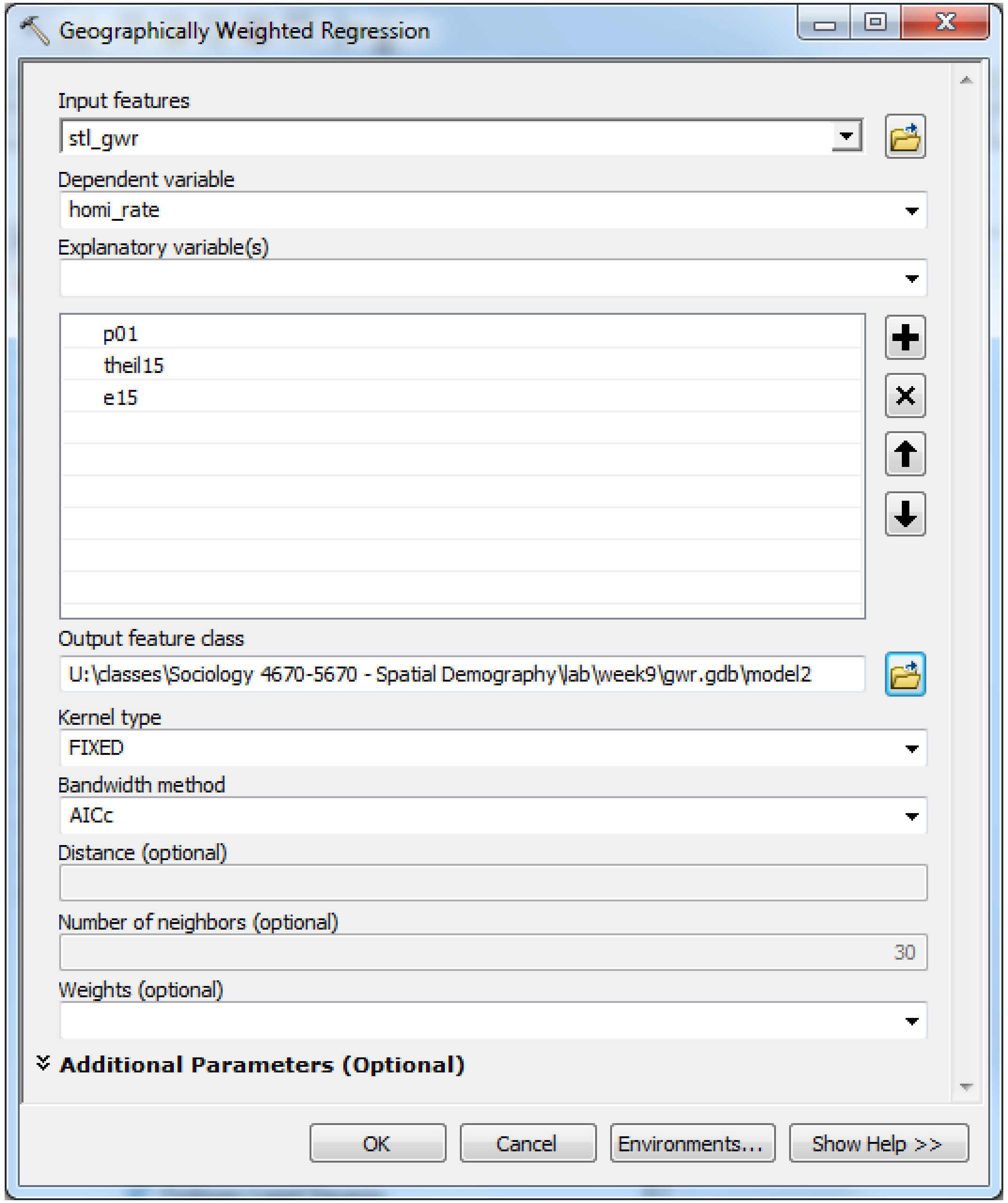
```
>>04/06/17 09:57:24
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD ESTIMATION
Data set            : stl_reg_geoda
Spatial Weight      : stl_reg_geoda_queen
Dependent Variable  :   homi_rate  Number of Observations:  106
Mean dependent var  :     3.30795  Number of Variables   :    5
S.D. dependent var  :     3.91868  Degrees of Freedom    :  101
Lag coeff.   (Rho)  :      0.5457

R-squared           :    0.631344  Log likelihood        :     -245.82
Sq. Correlation     : -            Akaike info criterion :      501.64
Sigma-square        :     5.66111  Schwarz criterion     :     514.957
S.E of regression   :     2.37931

-----------------------------------------------------------------------------
       Variable       Coefficient      Std.Error       z-value    Probability
-----------------------------------------------------------------------------
       W_homi_rate         0.5457      0.0913919       5.97099     0.00000
          CONSTANT        3.02493        1.06591       2.83788     0.00454
               p01        6.80589        1.77536       3.83353     0.00013
           theil15       -5.06869        2.48851      -2.03684     0.04167
               e15       -3.51467        1.04195      -3.37315     0.00074
-----------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                                     DF       VALUE        PROB
Breusch-Pagan test                       3         32.2817     0.00000

DIAGNOSTICS FOR SPATIAL DEPENDENCE
SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : stl_reg_geoda_queen
TEST                                     DF       VALUE        PROB
Likelihood Ratio Test                    1         24.5190     0.00000
============================== END OF REPORT ===============================
```



(25) I am going to estimate a more complex model.

Crime Rate is a function of poverty, income inequality and racial diversity.

gwr01_supp

OBJECTID *	VARNAME	VARIABLE	DEFINITION
1	Neighbors	25	
2	ResidualSquares	477.267562	
3	EffectiveNumber	26.398873	
4	Sigma	2.44862	
5	AICc	512.715022	
6	R2	0.706792	
7	R2Adjusted	0.613236	
8	Dependent Field	0	homi_rate
9	Explanatory Field	1	p01

(0 out of 9 Selected)

Model 1

model2_supp

OBJECTID *	VARNAME	VARIABLE	DEFINITION
1	Bandwidth	3507.657862	
2	ResidualSquares	436.696014	
3	EffectiveNumber	24.228765	
4	Sigma	2.310944	
5	AICc	497.055493	
6	R2	0.731717	
7	R2Adjusted	0.655506	
8	Dependent Field	0	homi_rate
9	Explanatory Field	1	p01
10	Explanatory Field	2	theil15
11	Explanatory Field	3	e15

(0 out of 11 Selected)

Model 2

Smaller values of this Sigma are preferable

If the AICc values for two models differ by more than 3, the model with the lower AICc is held to be better. Model 2 is better than Model 1

Advanced Model Comparisons

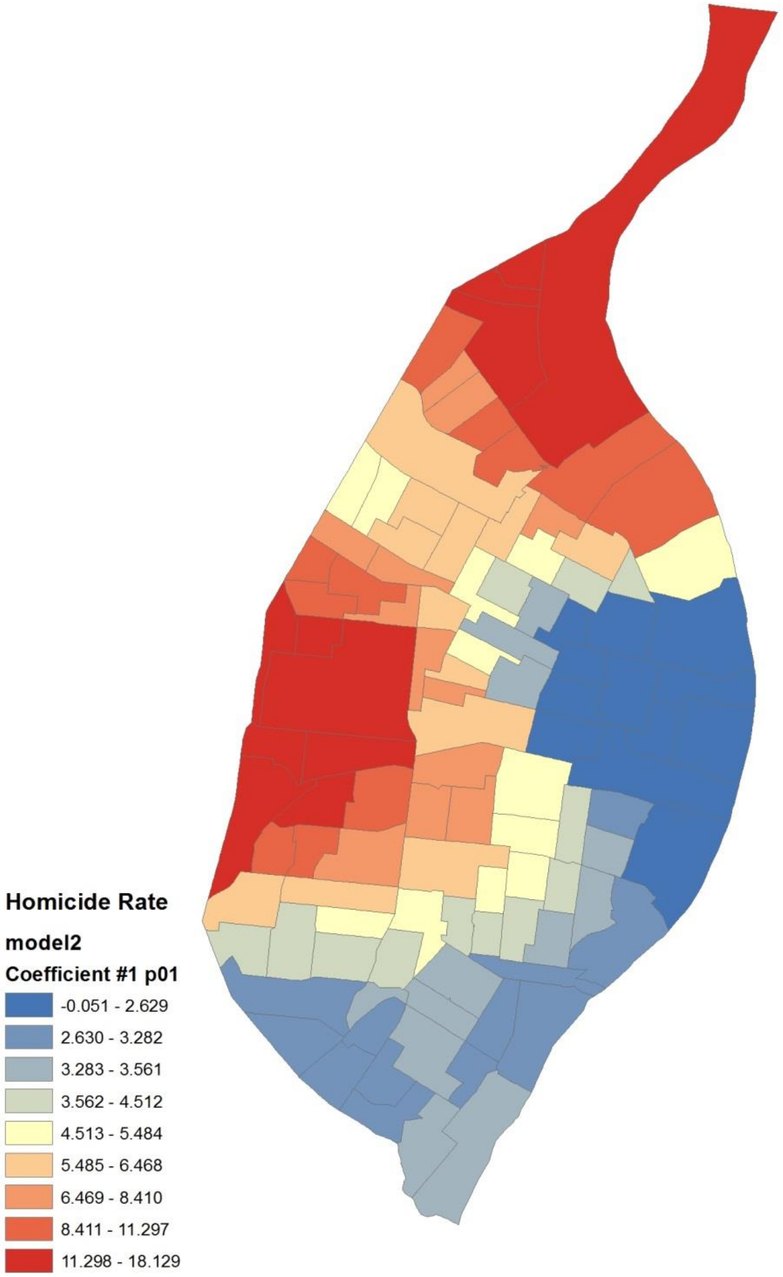
Global and Local Parameter Estimates of the Model

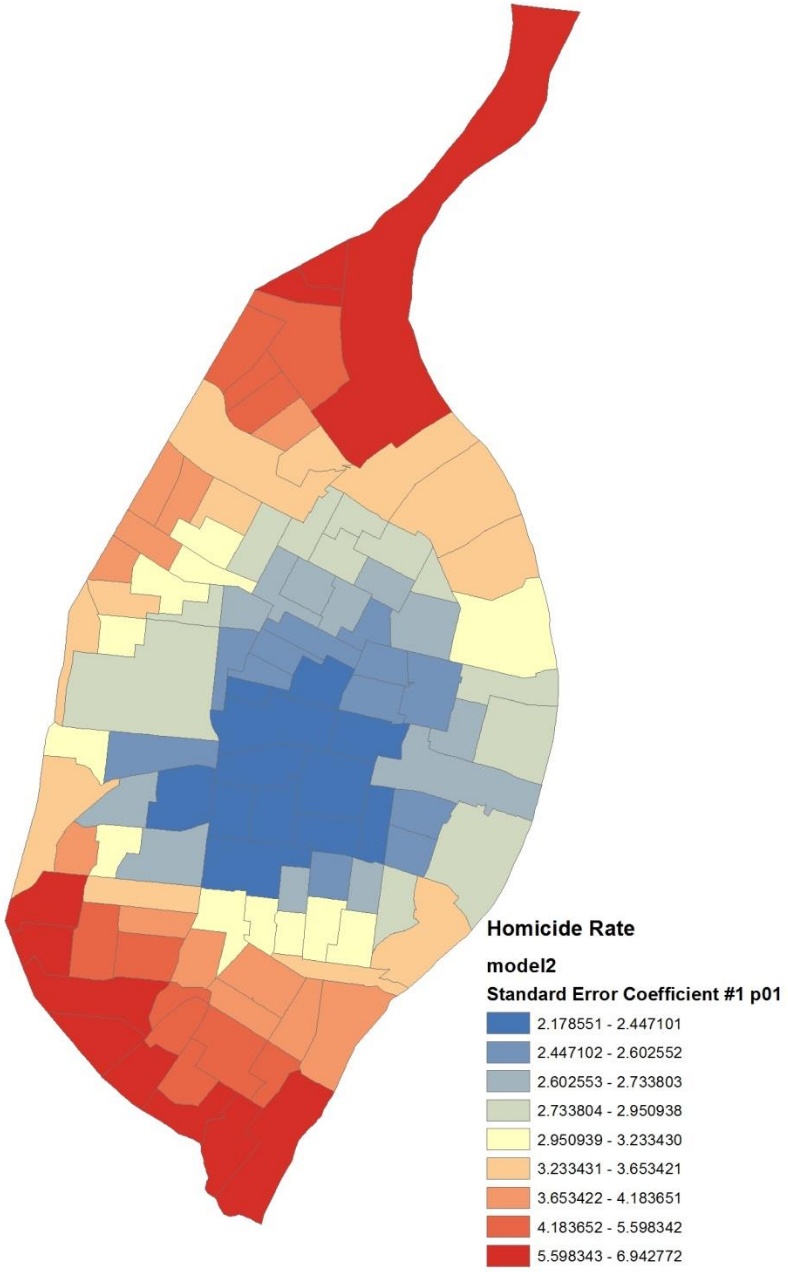
(n=106)

	Minimum	Lower Quartile	Median	Upper Quartile	Maximum	OLS	Range GWR	Range OLS
For Notes Only		50% of values				1 SD=66% of the values	Better range in GWR	
Poverty Rate	2.78	3.66	4.86	7.32	15.60	10.4424 (1.99592)***	3.66	3.99
Income Inequality	-10.30	-9.12	-5.40	1.98	2.43	-3.3254 (2.93936)	11.10	5.88
Diversity	-10.07	-1.33	-7.57	-8.91	0.29	-7.40396 (1.0404)***	7.58	2.08
Constant	-0.73	-0.12	6.87	9.80	11.08	4.79579 (1.14249)	9.92	2.28
R^2		0.73				0.5	GWR is a better model	
AIC		497.055				524.159		

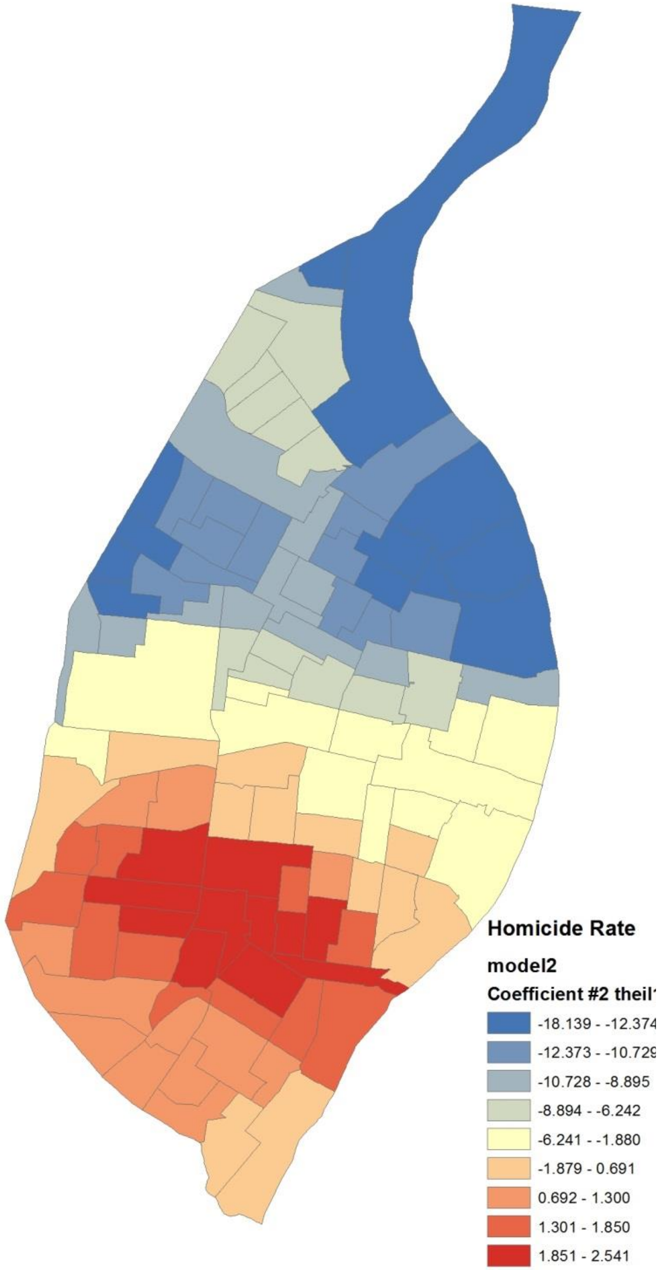
Advanced Model Comparisons
Map the Results

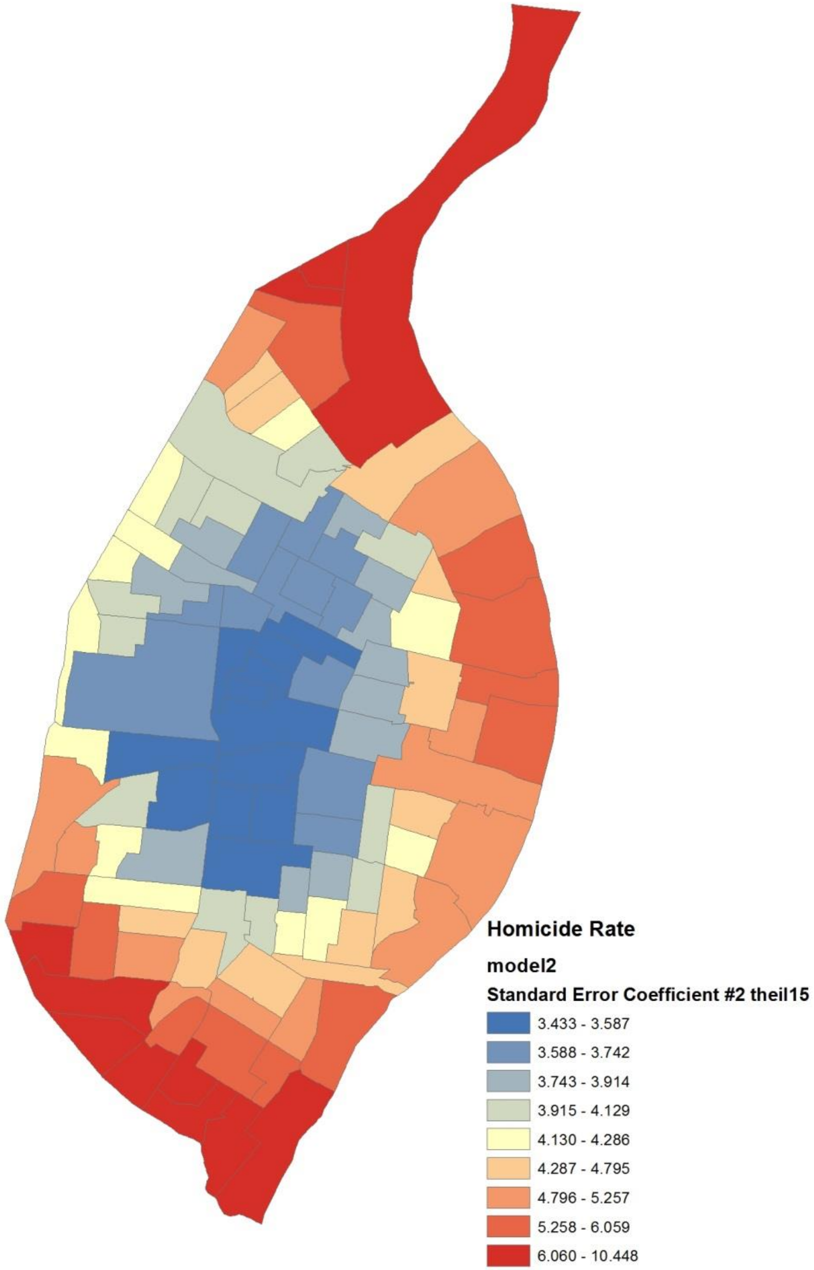
(25a) Let's map the Coefficients and Standard Error of the Coefficient Poverty Rate



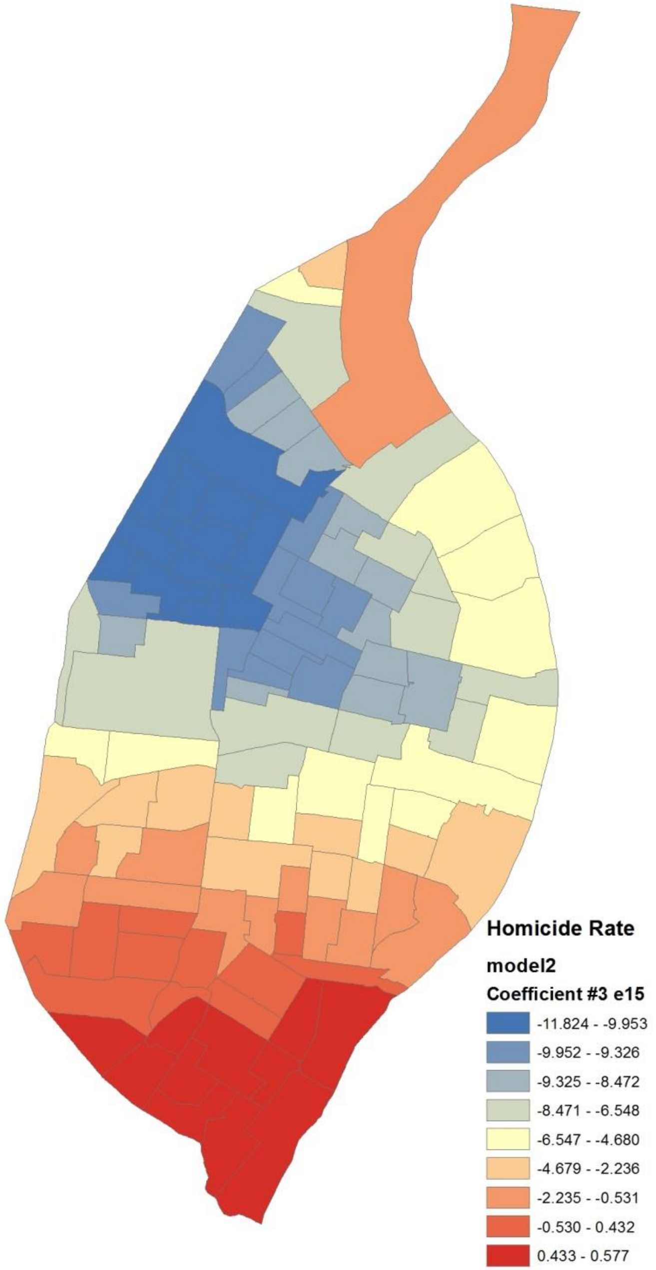


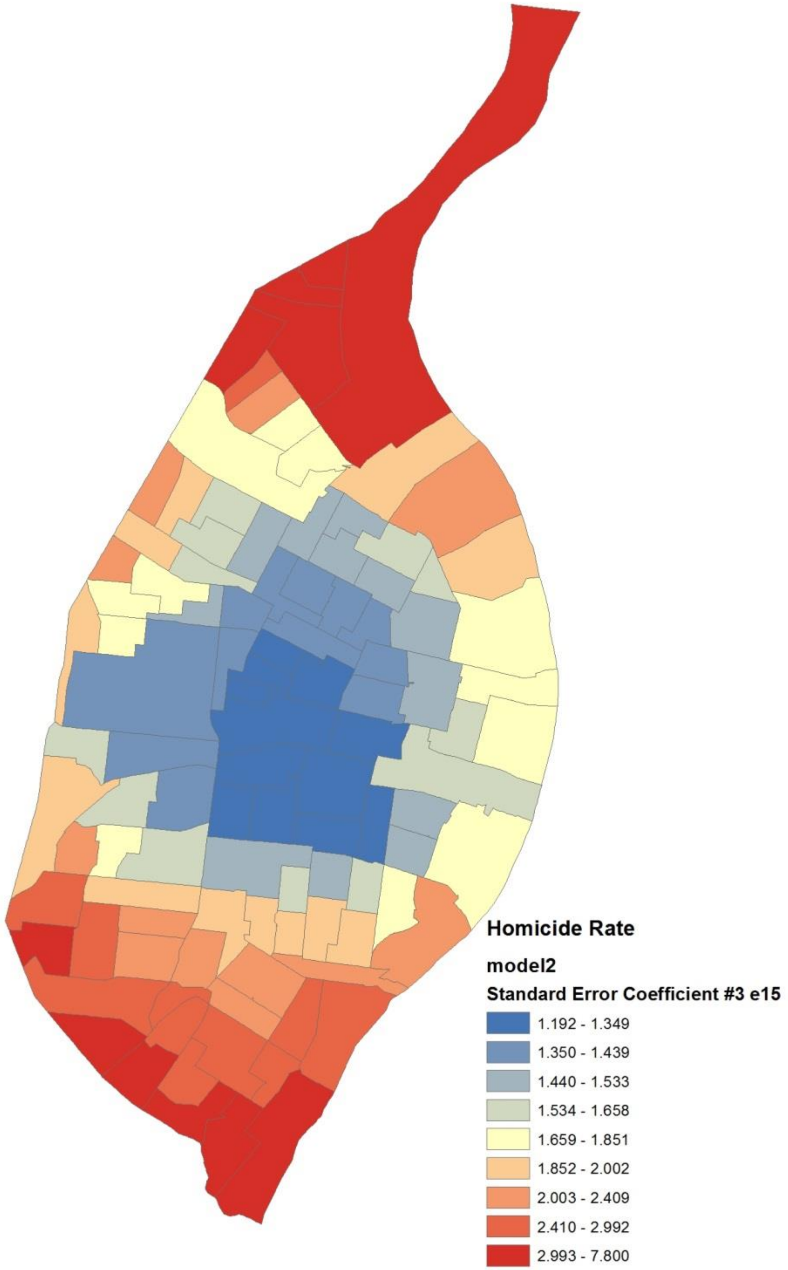
(25b) Let's map the Coefficients and Standard Error of the Coefficient Income Inequality





(25c) Let's map the Coefficients and Standard Error of the Coefficient Income Inequality





Assessing whether the spatial variation in measured relationships might be important

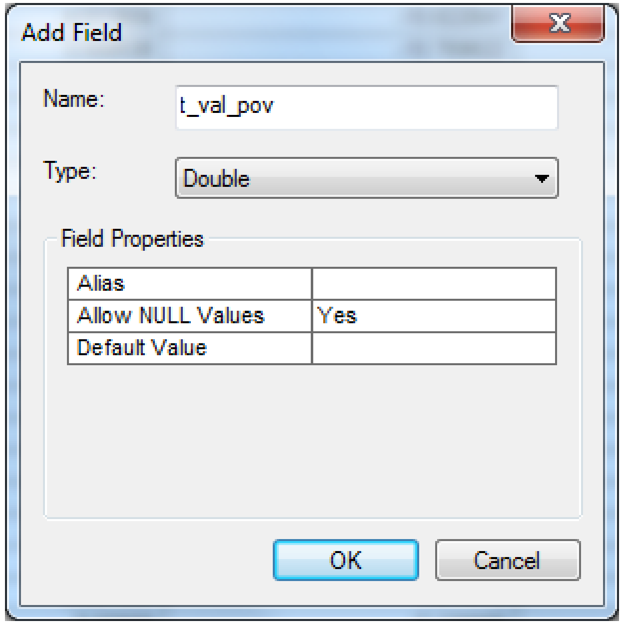
Local t values

$$t_i = \beta_i / SE\beta_i$$

Look for areas on the map with:

t_i values $\geq$ 1.96 and/or t_i values $\leq$ -1.96

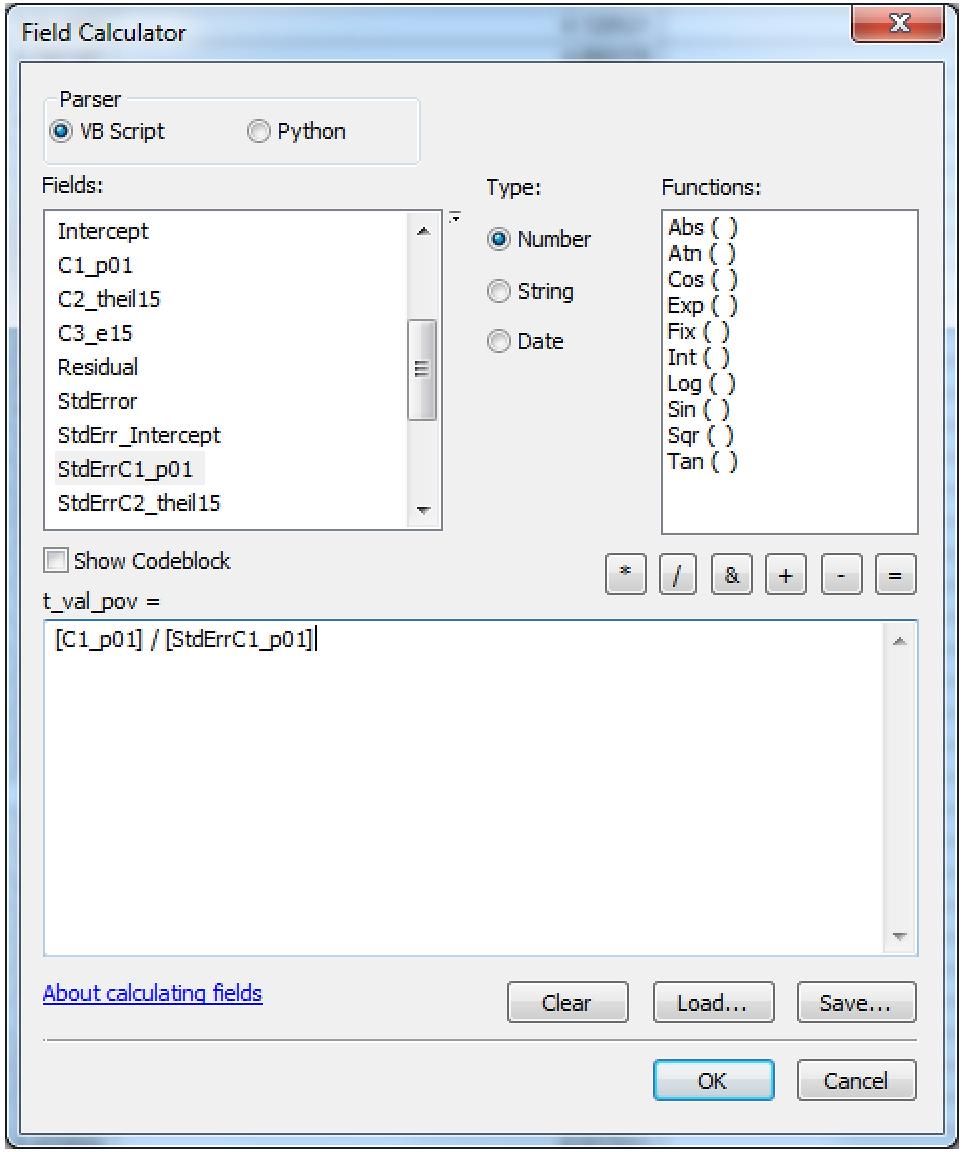
Map these values



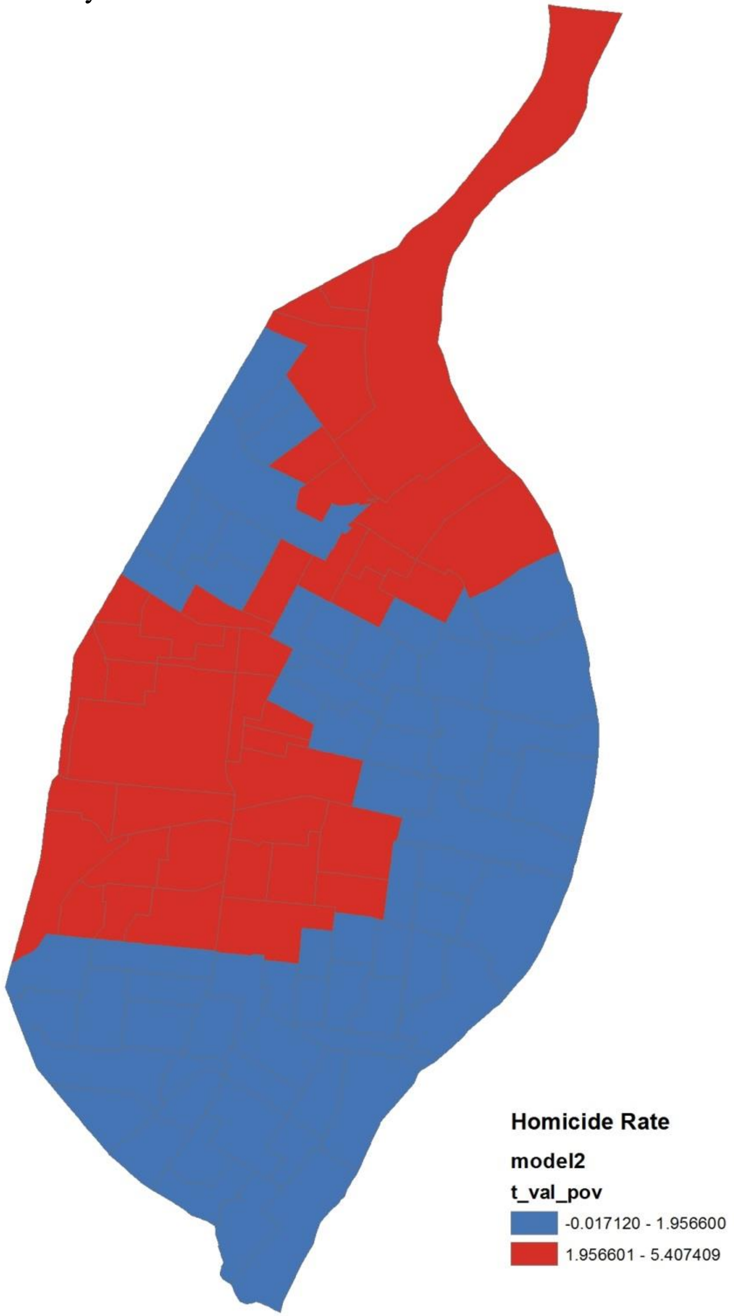
(26) We need to calculate the local t values. We first need to create a new variable. In this case I will care "t_val_pov" for the poverty rate.

(27) We then need to compute the value. Local t values are simily the coefficient dived by the standard error.

(28) Let's map the local t values. Let's look for values +_ 1.96

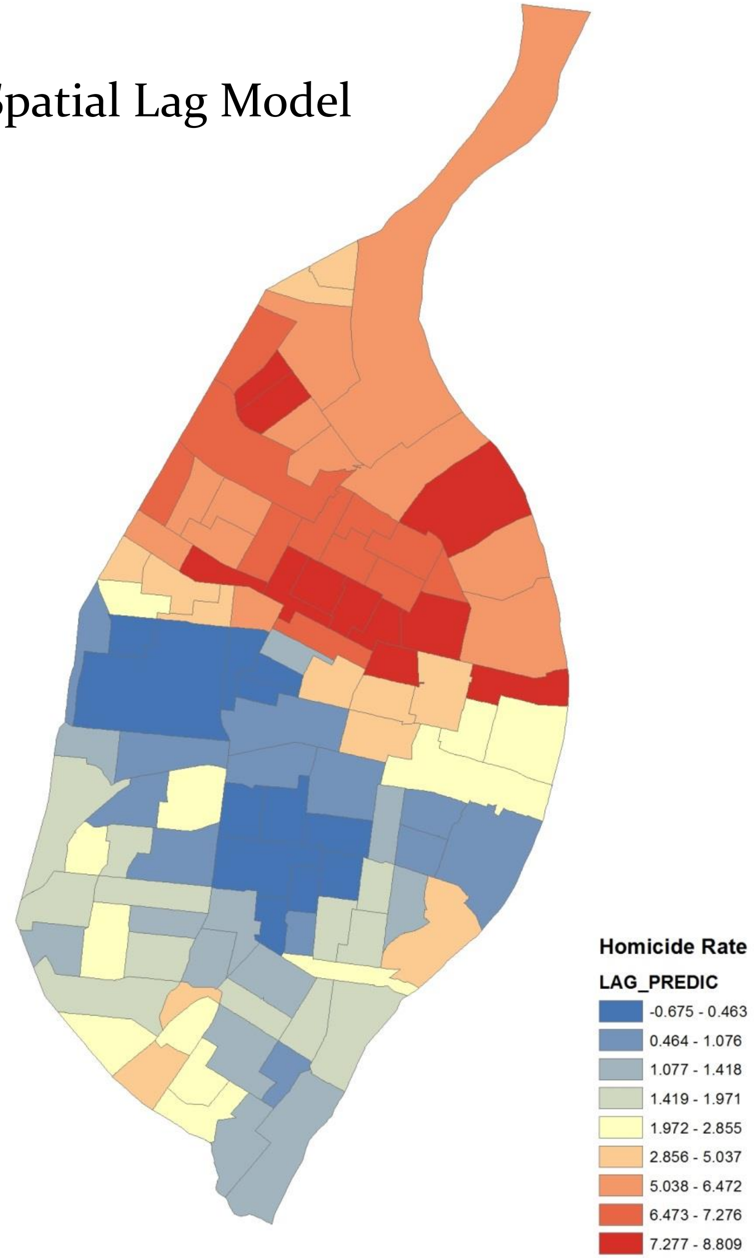


Local t values for poverty rate



Comparison of Predicted Values of Spatial Lag and GWR Models

Spatial Lag Model



GWR Model