1) Introduction

The purpose of this report is to predict median house values in Philadelphia at census block group level, with several neighborhood characteristics, including households living in poverty, percentage of individuals with bachelor’s degrees or higher, percentage of vacant houses, percentage of single house units.

In a previous study, this prediction was done through use of an Ordinary Least Squares (OLS) regression. However, OLS is often inadequate for predicting on datasets that have spatial correlation. Therefore, in this study, we will use three methods of spatial regression: spatial lag, spatial error, and geographically weighted (GWR) regressions. The performance of the three spatial regressions was compared to that of the OLS Regression model.

2) Methods

a) A Description of the Concept of Spatial Autocorrelation

To understand the pros and cons of the different regression models being used, we first need to understand the spatial processes that could be present in a dataset and how to measure them. Recall that the First Law of Geography indicates that while all things in space are related, things that are near are more related than distant things (Tobler, 1970). An alternative way to describe this phenomenon is called spatial autocorrelation. Spatial autocorrelation is the correlation of observations based on their spatial proximity. Positive spatial autocorrelation is when similar values are clustered, whereas negative spatial autocorrelation occurs when observations in proximity have distinct differences.

Spatial proximity can be defined in several ways such as rook neighbors (observations that share an edge) or queen neighbors (observations that share an edge or vertex). The spatial proximity of each observation relative to the other observations is collected in the form of a spatial weight matrix. There are several choices for weight matrices, such as Dacey, Cliff and Ord, and Getis and Aldstadt. While there may be compelling reasons to just use one, statisticians often compare multiple matrices so as to ensure that the results are not an artifact of the chosen matrix (Brusilovskiy, E. Presentation, 2021). In this study, we will use the Queen Weight matrix.

The method we use for identifying spatial autocorrelation is Moran’s I, which is defined below.

$$\mathbf{I} = \frac{\boldsymbol{n}}{\sum_{\boldsymbol{i}=1}^{\boldsymbol{n}}\sum_{\boldsymbol{j}=1}^{\boldsymbol{n}} \boldsymbol{w_{ij}}} \frac{\sum_{\boldsymbol{i}=1}^{\boldsymbol{n}}\sum_{\boldsymbol{j}=1}^{\boldsymbol{n}} \boldsymbol{w_{ij}}(\boldsymbol{X_i}-\bar{\boldsymbol{X}})(\boldsymbol{X_j}-\bar{\boldsymbol{X}})}{\sum_{\boldsymbol{i}=1}^{\boldsymbol{n}}(\boldsymbol{X_i}-\bar{\boldsymbol{X}})^2}$$

In this formula, n is the number of observations, $\bar{X}$ is the mean of variable X, X_i is the value of X at location i, X_j is the value of X at another location j, and w_{ij} is the weight indexing location of i relative to location j. The possible values of Moran's I range from -1 to 1, where low negative values indicate strong negative spatial autocorrelation, high positive values indicate strong positive spatial autocorrelation, and values near zero indicate no spatial autocorrelation.

The significance test for Moran's I works by comparing the Moran's I of the actual dependent variable to the distribution of Moran's I's of the permuted dependent variable (as a proxy for randomness), of which there will usually be 999 iterations calculated. The Moran's I values are then ranked along with the true Moran's I value, and a pseudo p-value is calculated by dividing the rank of the true Moran's I value by the total number of iterations (in this case, 1000). The null hypothesis, H_0, is that there is no spatial autocorrelation (Moran's I = 0), whereas the alternative hypothesis, Ha, is that there is spatial autocorrelation (Moran's I ≠ 0). For instance, if the pseudo p-value is less than the stated threshold, in this case 0.05, we can reject the null hypothesis that there is no spatial autocorrelation.

While Moran's I can be useful in identifying if there is spatial autocorrelation in the original dataset, it is often used to determine if a model is effective at accounting for spatial processes by calculating the Moran's I of the residuals. Said differently, if there is spatial autocorrelation in the residuals, the model is likely not accounting for a predictor or predictors.

Further, Moran's I can be applied both at the global and local levels. As described above, the global Moran's I indicates whether there is spatial autocorrelation that exists in the dataset. However, it is possible to have clusters of positive and negative spatial autocorrelation in a dataset. To identify these distinct clusters, the Moran's I is calculated for each areal unit and its neighbors, as defined by the spatial weight matrix. These values are then subjected to the same significance testing of the global Moran's I, however each permutation is randomized across the neighbors, not the entire dataset.

b) A Review of OLS Regression and Assumptions

Ordinary Least Squares (OLS) is a statistical method used to examine the relationship between a variable of interest (dependent variable) and one or more explanatory variables (predictors). OLS Regression allows us to calculate the amount by which your dependent variable changes when a predictor variable changes by one unit (holding all other predictors constant). OLS Regression chooses the parameters of a linear function of a set of explanatory variables by the principle of least squares: minimizing the sum of the squares of the differences between the observed dependent values and those values. The assumptions to use OLS regression are as follow. The mathematical relationships between each predictor in the equation and the dependent variable are linear. And there should not be multicollinearity among different independent predictors. At the same time, every observation is independent and the occurrence of one observation provides no information about the occurrence of the other observation. As for residuals, they are random, independent, and normally distributed.

Residuals should have homoscedasticity, which means variance of residuals remain constant at any value of independent predictors. For more information on OLS regression, please look at HW1 Assignment.

However, when the data has a spatial component, the assumption that errors are random/independent often doesn't hold. This can be tested by examining the spatial autocorrelation of the residuals using Moran's I. Another way to test OLS residuals for spatial autocorrelation is to regress them on nearby residuals, which are residuals at neighboring block groups, as defined by the Queen matrix. This refers to term known as rho (ρ) (also known as lambda (λ) in GeoDa). In GeoDa, it is referred to as Slope b in the statistics at the bottom of the scatterplot of OLS_RESIDU and WT_RESIDU. The formula to calculate rho (ρ) is listed below.

$$OLS\_RESIDU = \beta_0 + \rho WT\_RESIDU + \varepsilon$$

In GeoDa, the tool that is used to run OLS regression, also has a way of testing other regression assumptions. The first is the assumption of homoscedasticity, which is tied to the assumption of independence of errors. GeoDa has three different diagnostics for heteroscedasticity: The Breusch-Pagan Test, The Koenker-Bassett Test, and The White Test. The null hypothesis and the alternative hypothesis in these tests are list below.

Null Hypothesis: H_0: *There is homoscedasticity among residuals*
Alternative Hypothesis: H_a: *There is heteroscedasticity among residuals*

Another assumption is the normality of errors. The Jarque-Bera test in GeoDa are used to examines the normality of errors. If $p<0.05$, we can reject the Null Hypothesis of normality for the alternative hypothesis of non-normality.

Null Hypothesis: H_0: *The residuals are normally distributed*
Alternative Hypothesis: H_a: *The residuals are not normally distributed*

c) Spatial Lag and Spatial Error Regression

i) Spatial Lag Regression

GeoDa and R were used to run the spatial lag regression. The spatial lag regression takes the form of an OLS regression but adds one more predictor. The concept of a spatial lag regression is to include the lagged dependent variable (those values of the nearby observations as defined by the weights matrix) as a predictor of the dependent variable. The spatial lag regression was calculated with the following equation.

$$LNMEDHVAL = \rho W_{LNMEDHVAL} + \beta_0 + \beta_1\ PCTBACHMOR + \beta_2\ PCTVACANT$$

$$+ \beta_3\ PCTSINGLES + \beta_4\ LNNBELPOV100 + \varepsilon$$

In the equation, LNMEDHVAL is the dependent variable, and PCTBACHMOR, PCTVACANT, PCTSINGLES AND LNNBELPOV100 are independent predictors. Besides, β_0 is the intercept, β_1, β_2, β_3, β_4 are coefficients of variables PCTBACHMOR, PCTVACANT, PCTSINGLES, and LNNBELPOV100. $\rho W_{LNMEDHVAL}$ is the spatial lag of the dependent variable, LNMEDHVAL, where ρ is the coefficient of the lagged variable, $W_{LNMEDHVAL}$. Finally, ε , is the residual.

The goal of the spatial lag regression is to consider that changes in the dependent variable may also be a function of changes in its surrounding neighbors, which allows for better prediction of spatial phenomena.

ii) Spatial Error Regression

GeoDa and R were used to run the spatial lag regression. The method of spatial error regression is that we regress residuals on the nearest neighbor residuals, thereby filtering the spatial information out of the OLS Regression residuals and decomposing the residuals ε into two parts: one with a spatial pattern λWε and one which is simply random noise u. The part with a spatial pattern can be thought of as some variable with a spatial component missing from the OLS regression. And the model equation for the spatial error regression is listed below.

$$\boldsymbol{LNMEDHVAL = \beta_0 + \beta_1\ PCTBACHMOR + \beta_2\ PCTVACANT + \beta_3\ PCTSINGLES + \beta_4\ LNNBELPOV100 + \lambda W_{\varepsilon} + \mathcal{U}}$$

In the equation, LNMEDHVAL is the dependent variable, and PCTBACHMOR, PCTVACANT, PCTSINGLES AND LNNBELPOV100 are independent predictors. Besides, β_0 is the intercept, β_1, β_2, β_3, β_4 are coefficients of variables PCTBACHMOR, PCTVACANT, PCTSINGLES, and LNNBELPOV100. W_ε is the spatial lag residuals. And it is the average residuals of the nearest neighbors. λ is the coefficient of variable spatial lag residuals. $\mathcal{U}$ is the random noise.

The goal of spatial error regression is to take into consideration the fact there may be spatial dependencies in the residuals/the data.

iii) Spatial Lag and Spatial Error Regressions

The assumptions of OLS still apply to the spatial lag regression and spatial error regression models (except that of spatial independence of observations). The mathematical relationships between each predictor in the equation and the dependent variable are linear. And there should not be multicollinearity among different independent predictors. As for residuals, they should be random and normally distributed. Residuals should be homoscedastic, which means variance of residuals remain constant at any value of independent predictors.

In the following section, the results of spatial lag regression with OLS and the results of spatial error regressions with OLS will be compared and will decide whether the spatial models perform better than OLS based several criteria. These criteria include Akaike Information Criterion (AIC)/Schwarz Criterion (SC); Log Likelihood; and Likelihood Ratio

Test. AIC and SC are measures of the goodness of fit of an estimated statistical model. They are relative measures of the information that is lost when a given model is used to describe reality and can be said to describe the tradeoff between precision and complexity of the model. In GeoDa, the lower the AIC and SC, the better the fit. Log Likelihood is associated with the maximum likelihood method of fitting a statistical model to the data and estimating model parameters. Maximum likelihood picks the values of the model parameters that make the data "more likely" than any other values of the parameters would make them. The higher the log likelihood, the better the model fit. Likelihood ratio test is used to compares the OLS model with the spatial models. The likelihood-ratio test tests whether this ratio is significantly different from one, assessing the goodness of fit of two competing statistical models based on the ratio of their likelihoods. If the p value is significant important, then the spatial model is better than the OLS Regression model.

Null Hypothesis: $\boldsymbol{H_0}$: $\boldsymbol{OLS\ Model\ is\ better\ than\ Spatial\ Model}$
Alternative Hypothesis: $\boldsymbol{H_a}$: $\boldsymbol{Spatial\ Model\ is\ better\ than\ OLS\ Model}$.

Besides, since spatial models are introduced to account for spatial autocorrelation, there is another way of comparing OLS results with spatial lag and spatial error results. That is by looking at the Moran's I of regression residuals. A model with a Moran's I, which is close to zero is better, because that means the spatial autocorrelation has been account for.

d) Geographically Weighted Regression

Unfortunately, the Spatial Lag Regression and Spatial Error Regression introduced above, global regressions, cannot perform well on spatial non-stationarity with Simpson's paradox. Simpson's paradox is an effect that occurs when the marginal association between two categorical variables is qualitatively different from the partial association between the same two variables after controlling for one or more other variables. However, separate local regressions for each location can help to eliminate this kind of bias. Local regression is a form of regression analysis in which a model of the relationship between outcomes and predictors is obtained by fitting different functions to different segments or intervals of data. Here, we will use Geographically Weighted Regression, one of those local regression methods, in ArcGIS Pro.

GWR is one of those local regression methods that run the regression for every observation (location i) on other observations, and observations close to location i are given greater weights. The equation for the GWR model is written for each observation i:

$$\boldsymbol{y_i = \beta_{i0} + \beta_{i1}x_{i1} + \beta_{i2}x_{i2} + \cdots + \beta_{im}x_{im} + \varepsilon_i = \beta_{i0} + \sum_{k=1}^{m} \beta_{ik}\, x_{ik} + \varepsilon_i}$$

y_i is the expected value of dependent variable y of observation$i$, x_{ik} are k predictor values around the location of observation i, β_{ik} is the expected change in y_i associated with a 1-unit increase in the value of x_{ik} (holding all other variables constant), ε_i is the residual of observation i, a random variable, which is assumed to be normally distributed, with E(ε)=0.

There are two primary ways to weigh nearby locations for each observation - fixed bandwidth and adaptive bandwidth. The number of observations around each point i will vary for the fixed bandwidth, while the bandwidth distance h (and the area) is not. Conversely, for the adaptive bandwidth, the number of observations will stay unchanged, but the distance h will fluctuate. An adaptive bandwidth kernel (weighing function) is appropriate for this case because distribution varies with heterogeneously shaped and sized polygons across space. Given this spatial characteristic, we will apply adaptive bandwidth for later GWR regression.

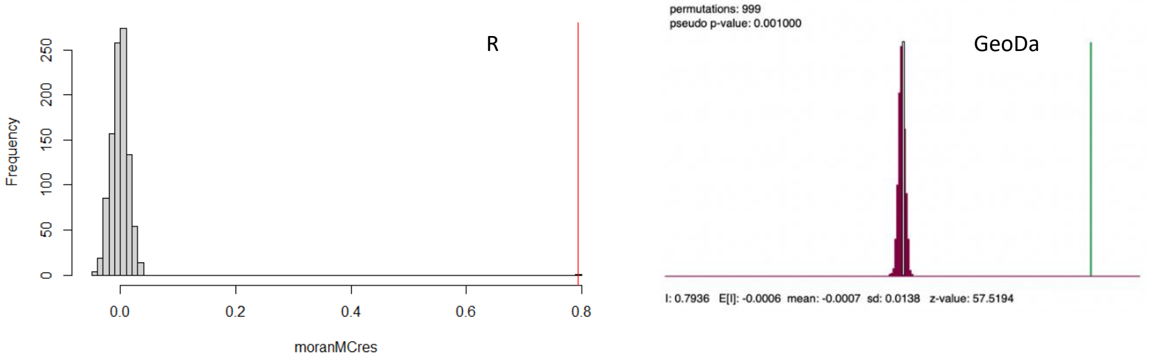
The assumptions that are needed for OLS are still needed for spatial error regression models (except that of spatial independence of observations). The mathematical relationships between each predictor in the equation and the dependent variable are linear. And there should not be multicollinearity among different independent predictors. As for residuals, they are random and normally distributed. Residuals should have homoscedasticity, which means variance of residuals remain constant at any value of independent predictors. In addition, GWR models require at least 300 observations. For multicollinearity violation in local regressions, several variables will have similar patterns of clusters in a certain region, which can be told from the condition number in the attribute table. If the condition number is larger than 30, equal to Null, or equal to 1.7976931348623158e + 308, the regression results are unreliable due to local multicollinearity.

To test whether the parameter estimates are significantly different from zero, the ratio of the beta coefficients and the standard error estimates for each location will be applied instead of p-values. Because local regressions estimate models for each location and each location need to do several significance tests, which will lead to a tremendous amount of significance tests followed by many tests that return a significant result simply by chance (i.e., type I error). Although there are ways to adjust for this multiple testing problem, those ways are not currently implemented in ArcGIS Pro.

3) Results

a) Spatial Autocorrelation

The Moran's I of 0.794 indicates that there is significant (p=0.001) spatial autocorrelation in the dependent variable, LNMEDHVAL. Figure 1 is a histogram comparison of the global Moran's I of LNMEDHVAL and the random permutations.



To further explain this relationship, Figure 2 illustrates the relationship between LNMEDHVAL and its spatial lags. If there were no spatial autocorrelation in the dataset, we would expect to see a horizontal line and a cluster of data around the origin of the graph. However, this is not the case, and there is a clear positive relationship between the LNMEDHVAL and its spatial lags.

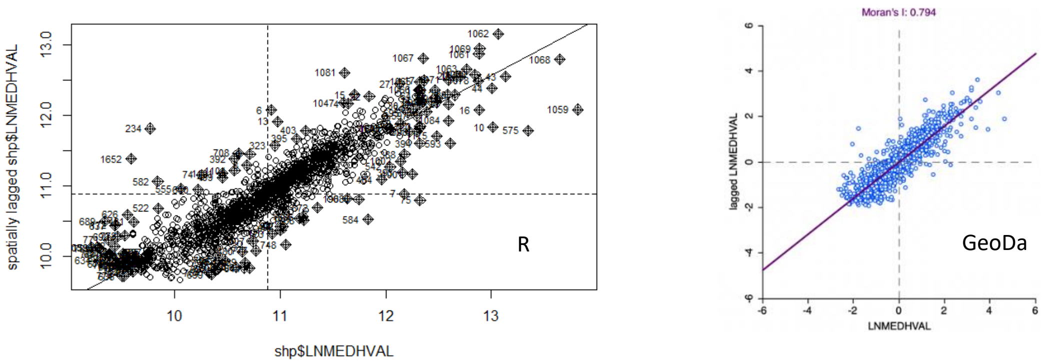
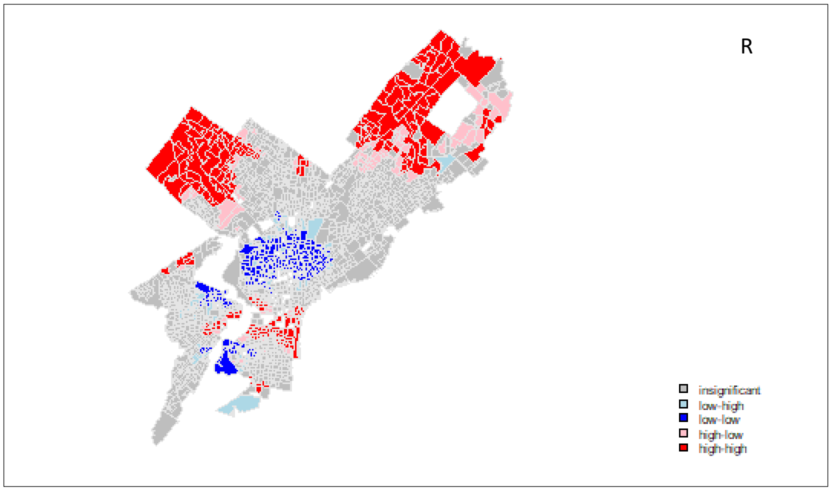


Figure 3 illustrates the distribution of significant areas of spatial autocorrelation within the dependent variable. All colored block groups have significant spatial autocorrelation, all greyed block groups do not. The values are mapped by the value of LNMEDHVAL-significance level of spatial autocorrelation. For instance, the pink values refer to areas which have high house values, and lower significant spatial autocorrelation.

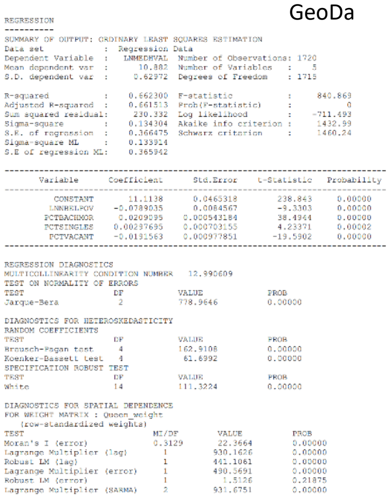


The clustering of high value homes in areas on the border of the urban and suburban parts of the city. These are areas which have been traditionally segregated and populated by white people. Conversely, the clustering of low value homes tends to center around denser and more urbanized parts of the city which have been traditionally segregated and populated by people of color.

b) A Review of OLS Regression and Assumptions

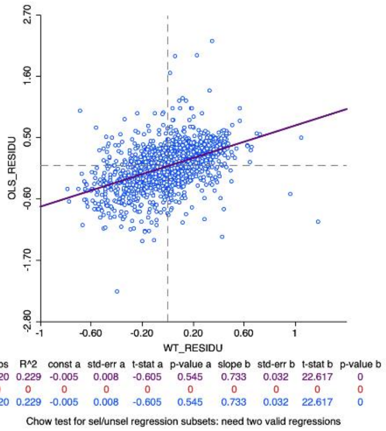
The OLS output from GeoDa and R are present as follows. From Table 1, we can see that the regression outputs indicate that all predictor variables (PCTBACHMOR, PCTVACANT, PCTSINGLES, AND LNNBELPOV100) are highly significant (p<0.0001 for all variables). And more than 60% of the variance in the dependent variable is explained by the model (R2 and Adjusted R2 are 0.6929 and 0.6922 in R and 0.6623 and 0.6615 in GeoDa, respectively). The low p-value (p<0.0001) associated with the F-ratio indicates that we can reject the null hypothesis that all model coefficients are 0.

GeoDa



```
REGRESSION
----------
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set            :  Regression Data
Dependent Variable  :   LNMEDHVAL  Number of Observations: 1720
Mean dependent var  :      10.882  Number of Variables   :    5
S.D. dependent var  :     0.62972  Degrees of Freedom    : 1715

R-squared           :    0.662300  F-statistic           :     840.869
Adjusted R-squared  :    0.661513  Prob(F-statistic)     :           0
Sum squared residual:     230.332  Log likelihood        :    -711.493
Sigma-square        :    0.134304  Akaike info criterion :     1432.99
S.E. of regression  :    0.366475  Schwarz criterion     :     1460.24
Sigma-square ML     :    0.133914
S.E of regression ML:    0.365942

-----------------------------------------------------------------------------
       Variable      Coefficient       Std.Error    t-Statistic   Probability
-----------------------------------------------------------------------------
          CONSTANT       11.1138       0.0465318       238.843     0.00000
         LNNBELPOV    -0.0789035       0.0084567       -9.3303     0.00000
        PCTBACHMOR     0.0209095     0.000543184       38.4944     0.00000
        PCTSINGLES    0.00297695     0.000703155       4.23371     0.00002
         PCTVACANT    -0.0191563     0.000977851      -19.5902     0.00000
-----------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER   12.990609
TEST ON NORMALITY OF ERRORS
TEST                  DF           VALUE             PROB
Jarque-Bera            2           778.9646          0.00000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                  DF           VALUE             PROB
Breusch-Pagan test     4           162.9108          0.00000
Koenker-Bassett test   4            61.6992          0.00000
SPECIFICATION ROBUST TEST
TEST                  DF           VALUE             PROB
White                 14           111.3224          0.00000

DIAGNOSTICS FOR SPATIAL DEPENDENCE
FOR WEIGHT MATRIX : Queen_weight
   (row-standardized weights)
TEST                           MI/DF        VALUE           PROB
Moran's I (error)              0.3129       22.3664         0.00000
Lagrange Multiplier (lag)        1         930.1626         0.00000
Robust LM (lag)                  1         441.1061         0.00000
Lagrange Multiplier (error)      1         490.5691         0.00000
Robust LM (error)                1           1.5126         0.21875
Lagrange Multiplier (SARMA)      2         931.6751         0.00000
```

GeoDa



#obs	R^2	const a	std-err a	t-stat a	p-value a	slope b	std-err b	t-stat b	p-value b
1720	0.229	-0.005	0.008	-0.605	0.545	0.733	0.032	22.617	0
0	0	0	0	0	0	0	0	0	0
1720	0.229	-0.005	0.008	-0.605	0.545	0.733	0.032	22.617	0

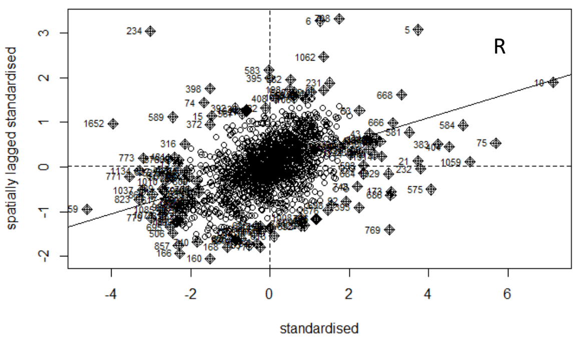
Chow test for sel/unsel regression subsets: need two valid regressions

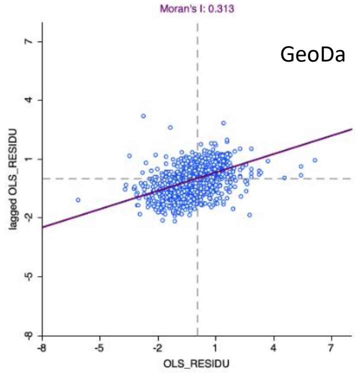
From the diagnostics in GeoDa, we can see the results for heteroscedasticity of The Breusch-Pagan Test, The Koenker-Bassett Test, and The White Test. All these three tests are consistent with each other and indicate a problem with heteroscedasticity.

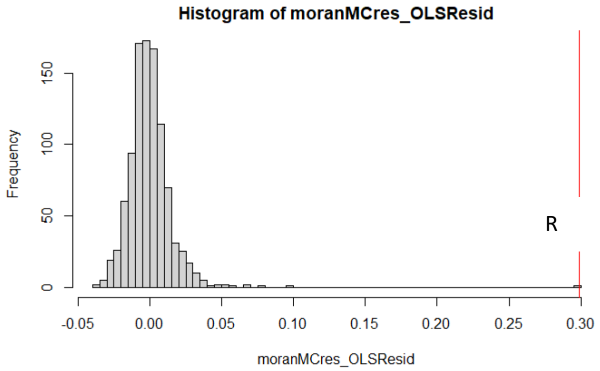
Also, the Jarque-Bera test in GeoDa examines the Null Hypothesis that the residuals are from a normal distribution. From table 1, we can see that p value of the Jarque-Bera test is significant important indicating that we can reject the Null Hypothesis of normality for the alternative hypothesis of non-normality.

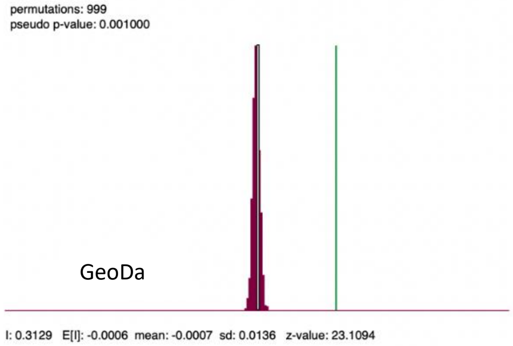
To examine the spatial autocorrelation of this OLS Regression, the scatterplot of OLS_RESIDU by WT_RESIDU are presented in the following. (The right one is the graph from GeoDa and the other one is from R) We can see from the picture that the value of ρ (referred to as Slope b in the results) is 0.733 and it is significant important. That indicates a significant spatial autocorrelation in the OLS Regression model.

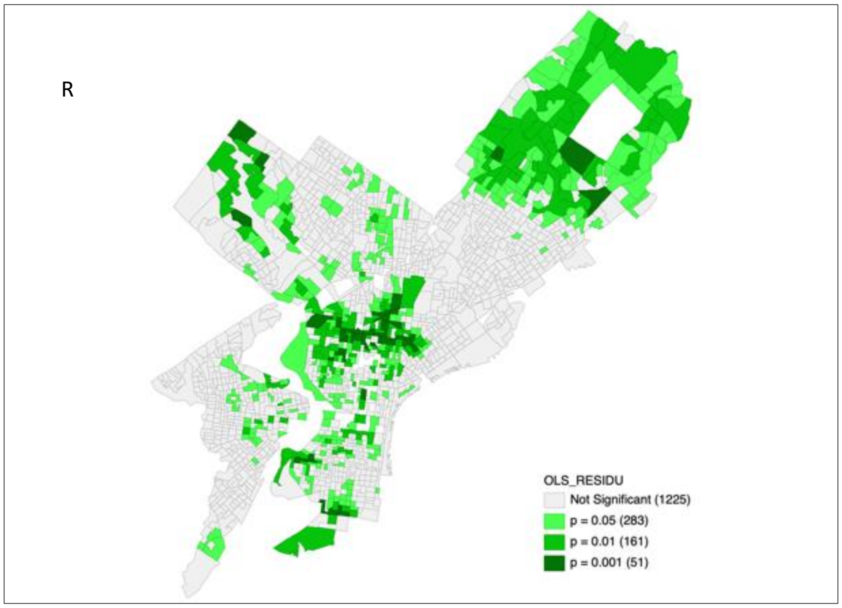
To further examine spatial autocorrelation, the Moran's I scatterplot and results from the 999 permutations for OLS regression residuals are plotted as follows. From the graphs, we can see a significant spatial autocorrelation in this OLS residuals. It is problematic and we will attempt to account for that in the following practices in spatial model regressions.


R
spatially lagged standardised
standardised
-4
-2
0
2
4
6
-2
-1
0
1
2
3
234
6
5
1062
583
395
398
74
668
1652
589
316
15
372
666
581
584
383
75
1059
773
771
1037
823
575
769
857
166
160
59


Moran's I: 0.313
GeoDa
lagged OLS_RESIDU
OLS_RESIDU
-8
-5
-2
1
4
7


Histogram of moranMCres_OLSResid
Frequency
moranMCres_OLSResid
R
0
50
100
150
-0.05
0.00
0.05
0.10
0.15
0.20
0.25
0.30

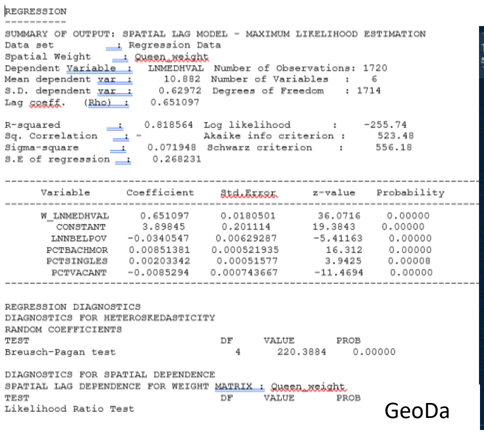

permutations: 999
pseudo p-value: 0.001000
GeoDa
I: 0.3129   E[I]: -0.0006   mean: -0.0007   sd: 0.0136   z-value: 23.1094


R
OLS_RESIDU
Not Significant (1225)
p = 0.05 (283)
p = 0.01 (161)
p = 0.001 (51)

c) Spatial Lag and Spatial Error Regression

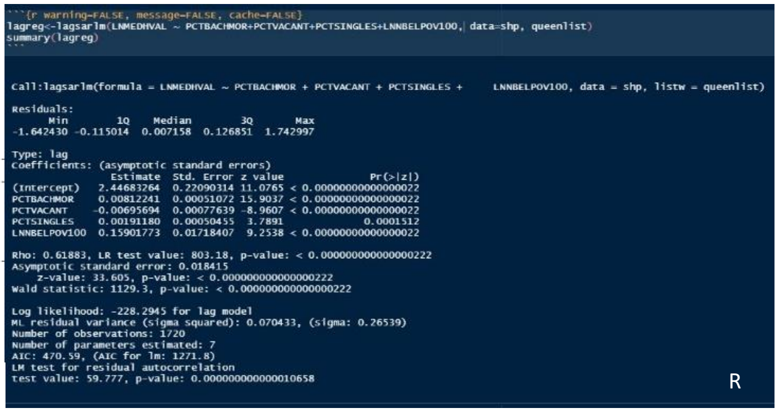
i) Spatial Lag Regression

The spatial lag regression output from GeoDa and R is presented below. From Table 2, we can tell that the coefficient of the spatial lag, ρ is 0.62, and it is significantly significant ($p < 0.0001$). This indicates that the unexplained variation in median house values is highly (positively) correlated. Further, we can see that all predictor variables (PCTBACHMOR, PCTVACANT, PCTSINGLES, AND LNNBELPOV100) are statistically significant ($p<0.0001$ for all variables). However, compared to the OLS AIC (1271.8), the spatial lag regression is neither better nor worse (AIC = 1271.8).



```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD ESTIMATION
Data set            : Regression Data
Spatial Weight      : Queen weight
Dependent Variable  :   LNMEDHVAL  Number of Observations: 1720
Mean dependent var  :      10.882  Number of Variables   :    6
S.D. dependent var  :     0.62972  Degrees of Freedom    : 1714
Lag coeff.   (Rho)  :    0.651097

R-squared           :    0.818564  Log likelihood        :     -255.74
Sq. Correlation     : -            Akaike info criterion :      523.48
Sigma-square        :    0.071948  Schwarz criterion     :      556.18
S.E of regression   :    0.268231

-----------------------------------------------------------------------------
       Variable       Coefficient      Std.Error       z-value    Probability
-----------------------------------------------------------------------------
      W_LNMEDHVAL      0.651097       0.0180501       36.0716      0.00000
         CONSTANT      3.89845        0.201114        19.3843      0.00000
        LNNBELPOV    -0.0340547       0.00629287      -5.41163      0.00000
       PCTBACHMOR     0.00851381      0.000521935      16.312       0.00000
       PCTSINGLES     0.00203342      0.00051577       3.9425       0.00008
        PCTVACANT    -0.0085294       0.000743667     -11.4694      0.00000
-----------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                                    DF      VALUE        PROB
Breusch-Pagan test                       4      220.3884     0.00000

DIAGNOSTICS FOR SPATIAL DEPENDENCE
SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : Queen weight
TEST                                    DF      VALUE        PROB
Likelihood Ratio Test
```

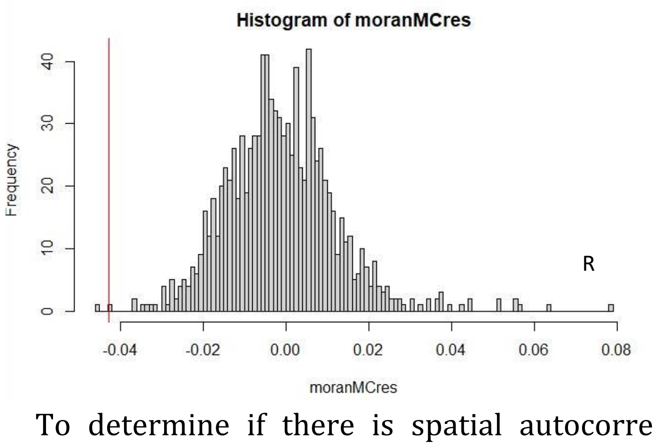
GeoDa

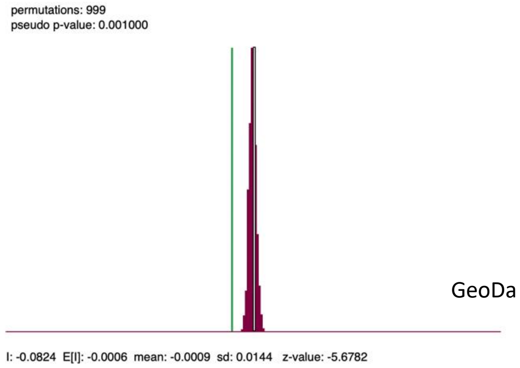


~~~
```{r warning=FALSE, message=FALSE, cache=FALSE}
lagreg<-lagsarlm(LNMEDHVAL ~ PCTBACHMOR+PCTVACANT+PCTSINGLES+LNNBELPOV100, data=shp, queenlist)
summary(lagreg)
```


Call:lagsarlm(formula = LNMEDHVAL ~ PCTBACHMOR + PCTVACANT + PCTSINGLES +     LNNBELPOV100, data = shp, listw = queenlist)

Residuals:
      Min        1Q    Median        3Q       Max
-1.642430 -0.115014  0.007158  0.126851  1.742997

Type: lag
Coefficients: (asymptotic standard errors)
                Estimate  Std. Error z value              Pr(>|z|)
(Intercept)   2.44683264  0.22090314 11.0765 < 0.00000000000000022
PCTBACHMOR    0.00812241  0.00051072 15.9037 < 0.00000000000000022
PCTVACANT    -0.00695694  0.00077639 -8.9607 < 0.00000000000000022
PCTSINGLES    0.00191180  0.00050455  3.7891             0.0001512
LNNBELPOV100  0.15901773  0.01718407  9.2538 < 0.00000000000000022

Rho: 0.61883, LR test value: 803.18, p-value: < 0.000000000000000222
Asymptotic standard error: 0.018415
    z-value: 33.605, p-value: < 0.000000000000000222
Wald statistic: 1129.3, p-value: < 0.000000000000000222

Log likelihood: -228.2945 for lag model
ML residual variance (sigma squared): 0.070433, (sigma: 0.26539)
Number of observations: 1720
Number of parameters estimated: 7
AIC: 470.59, (AIC for lm: 1271.8)
LM test for residual autocorrelation
test value: 59.777, p-value: 0.00000000000010658
~~~

R

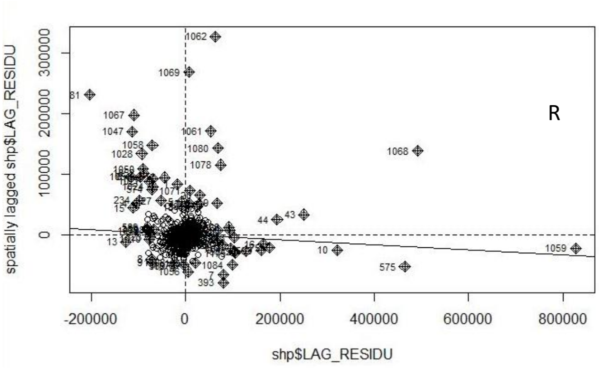
The Breusch-Pagan test, as observed in the GeoDa results summary, indicates that there is an issue with heteroscedastic residuals ($p < 0.0001$). All these three tests are consistent with each other and indicate a problem with heteroscedasticity.

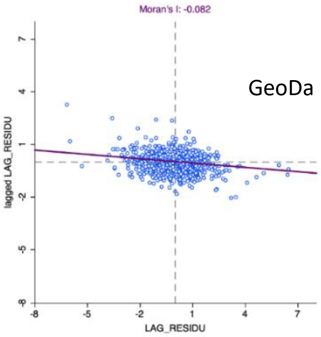


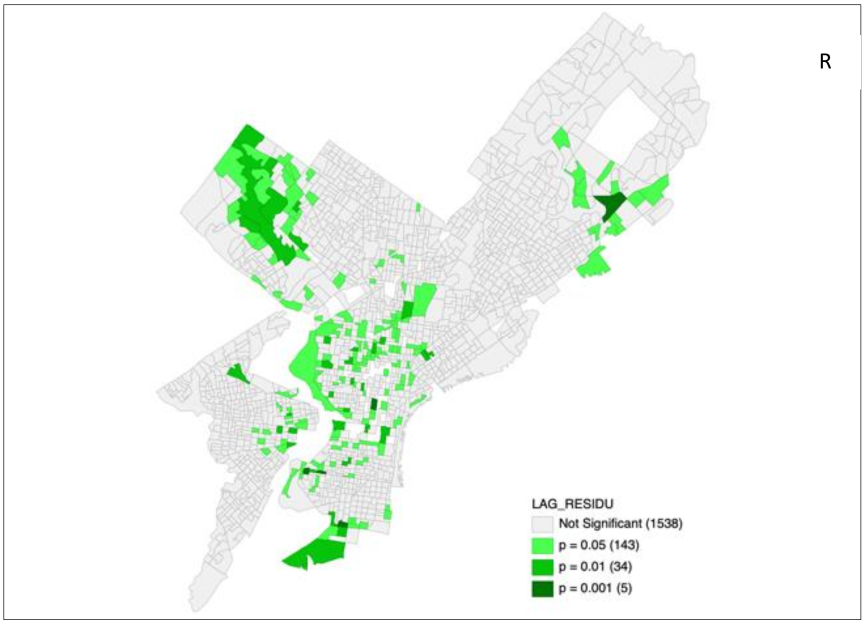


To determine if there is spatial autocorrelation in the residuals of the spatial lag regression, the Global Moran's I histogram was created. The red line indicates the location of the actual Global Moran's I from the spatial lag model, and the gray bars indicate the permutations. Further, to examine for the occurrence of local spatial autocorrelation, the scatterplot of LAG_RESIDU by spatially lagged LAG_RESIDU are presented in the following.

We can see from the picture that the residuals of the spatial lag regression tend to be much less spatially autocorrelated than the OLS residuals.



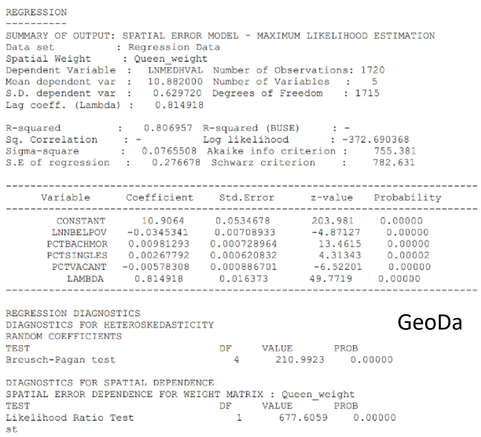




Overall, while the spatial lag is does not deviate from OLS regression regarding the Akaike Information Criterion, it performs much better at accounting for the spatial processes that exist within the data, as there is much less spatial autocorrelation in the residuals.

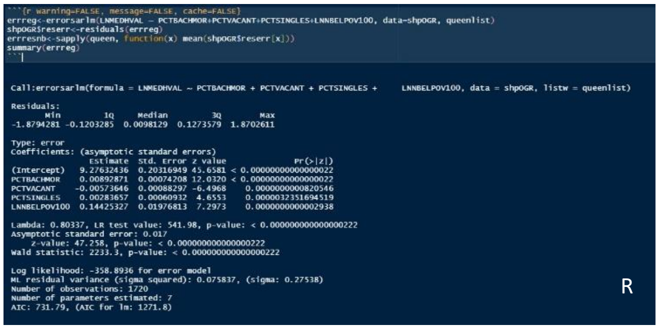
ii) Spatial Error Regression

The Spatial Error Regression output from GeoDa and R are present as follows. From Table 3, we can tell that the coefficient of the spatial parameter λ is 0.81, and it is significantly important. This shows that the unexplained variation in median house values is highly (positively) correlated.



```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIAL ERROR MODEL - MAXIMUM LIKELIHOOD ESTIMATION
Data set            : Regression Data
Spatial Weight      : Queen_weight
Dependent Variable  :   LNMEDHVAL  Number of Observations: 1720
Mean dependent var  :   10.882000  Number of Variables   :    5
S.D. dependent var  :    0.629720  Degrees of Freedom    : 1715
Lag coeff. (Lambda) :    0.814918

R-squared           :    0.806957  R-squared (BUSE)      : -
Sq. Correlation     : -            Log likelihood        : -372.690368
Sigma-square        :   0.0765508  Akaike info criterion :    755.381
S.E of regression   :    0.276678  Schwarz criterion     :    782.631

-----------------------------------------------------------------------------
     Variable       Coefficient     Std.Error       z-value    Probability
-----------------------------------------------------------------------------
        CONSTANT       10.9064      0.0534678       203.981     0.00000
       LNNBELPOV    -0.0345341      0.00708933      -4.87127     0.00000
      PCTBACHMOR     0.00981293     0.000728964      13.4615     0.00000
      PCTSINGLES     0.00267792     0.000620832      4.31343     0.00002
       PCTVACANT    -0.00578308     0.000886701     -6.52201     0.00000
          LAMBDA      0.814918      0.016373        49.7719     0.00000
-----------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                                DF      VALUE        PROB
Breusch-Pagan test                   4      210.9923     0.00000

DIAGNOSTICS FOR SPATIAL DEPENDENCE
SPATIAL ERROR DEPENDENCE FOR WEIGHT MATRIX : Queen_weight
TEST                                DF      VALUE        PROB
Likelihood Ratio Test                1      677.6059     0.00000
st
```

GeoDa

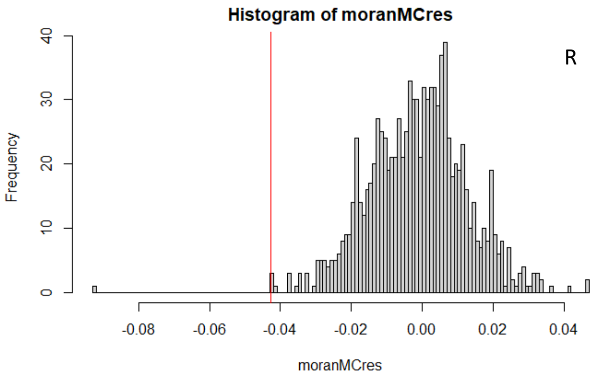


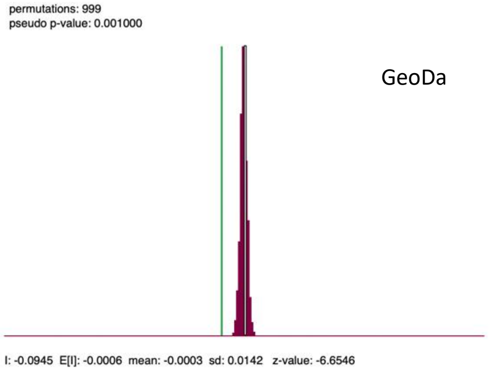
~~~
```{r warning=FALSE, message=FALSE, cache=FALSE}
errreg<-errorsarlm(LNMEDHVAL ~ PCTBACHMOR+PCTVACANT+PCTSINGLES+LNNBELPOV100, data=shpOGR, queenlist)
shpOGR$reserr<-residuals(errreg)
erresnb<-sapply(queen, function(x) mean(shpOGR$reserr[x]))
summary(errreg)
```


Call:errorsarlm(formula = LNMEDHVAL ~ PCTBACHMOR + PCTVACANT + PCTSINGLES +     LNNBELPOV100, data = shpOGR, listw = queenlist)

Residuals:
       Min         1Q     Median         3Q        Max 
-1.8794281 -0.1203285  0.0098129  0.1273579  1.8702611 

Type: error 
Coefficients: (asymptotic standard errors) 
                Estimate  Std. Error z value              Pr(>|z|)
(Intercept)   9.27632436  0.20316949 45.6581 < 0.00000000000000022
PCTBACHMOR    0.00892871  0.00074208 12.0320 < 0.00000000000000022
PCTVACANT    -0.00573646  0.00088297 -6.4968    0.0000000000820546
PCTSINGLES    0.00283657  0.00060932  4.6553    0.0000032351694519
LNNBELPOV100  0.14425327  0.01976813  7.2973    0.0000000000002938

Lambda: 0.80337, LR test value: 541.98, p-value: < 0.000000000000000222
Asymptotic standard error: 0.017
    z-value: 47.258, p-value: < 0.000000000000000222
Wald statistic: 2233.3, p-value: < 0.000000000000000222

Log likelihood: -358.8936 for error model
ML residual variance (sigma squared): 0.075837, (sigma: 0.27538)
Number of observations: 1720 
Number of parameters estimated: 7 
AIC: 731.79, (AIC for lm: 1271.8)
~~~

R

After introducing the spatial parameter λ, we can see that the remaining terms (LNNBELPOV, PCTBACHMOR, PCTSINGLES, and PCTVACANT) all are still significant, even some coefficients of predictors change when comparing to the result in OLS Regression. However, based on the Breusch-Pagan test result in the GeoDa, which is still significantly important, the spatial lag regression residuals are still heteroscedastic.

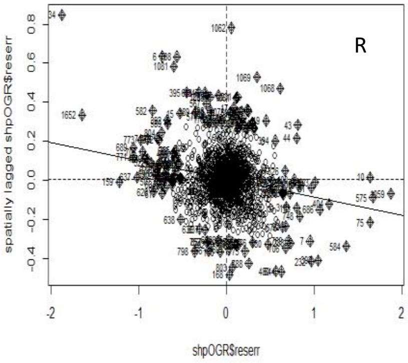
We can compare the Spatial Error regression and OLS regression based on the Akaike Information Criterion/Schwarz Criterion, the Log Likelihood, and the Likelihood Ratio Test, to see whether spatial error regression has improved the model. In the Akaike Information Criterion/Schwarz Criterion, the value of the spatial error model is nearly half of the OLS Regression, indicating that the spatial error model is way better. For the Log Likelihood, the spatial error model is also nearly half of the OLS Regression, validating that the spatial error model is way better. For the Likelihood Ratio Test, the p value is significant important, so the spatial model is better than the OLS Regression model.

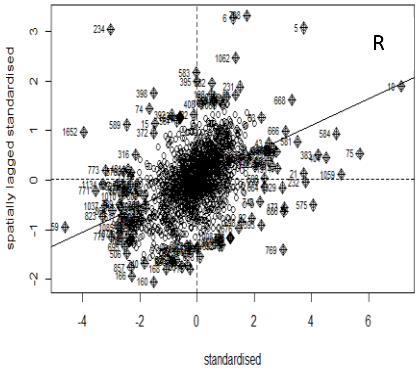


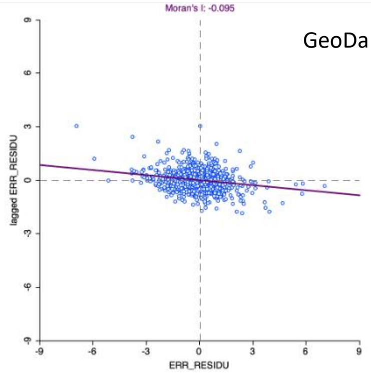


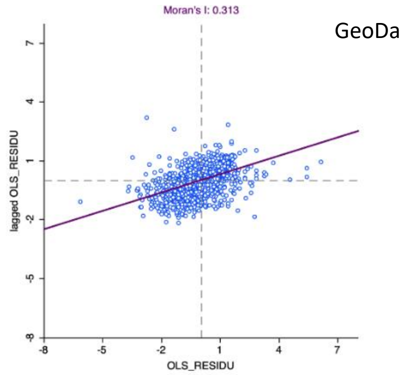
To determine if there is spatial autocorrelation in the residuals of the spatial lag regression, the Global Moran's I histogram was created. The red line indicates the location of the actual Global Moran's I from the spatial error model, and the gray bars indicate the permutations.

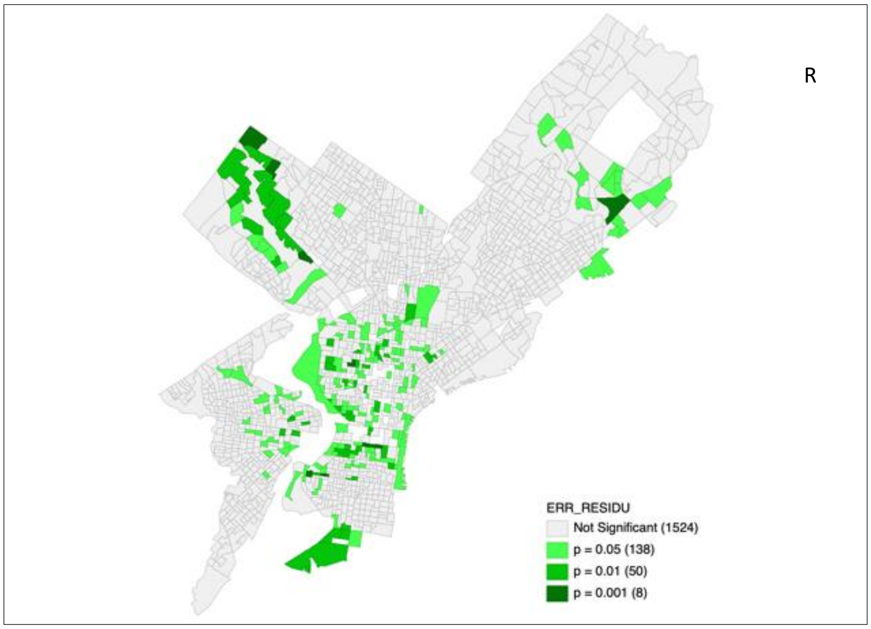
The Local Moran's I scatterplots of spatial error regression residuals and OLS regression residuals are plotted respectively in the following left and in the following right. From the scatter plots, we can see the value of Moran's I for spatial error model is less than the value of the OLS model and is closer to 0. That means there seem to be less spatial autocorrelation in these residuals than in OLS residuals.









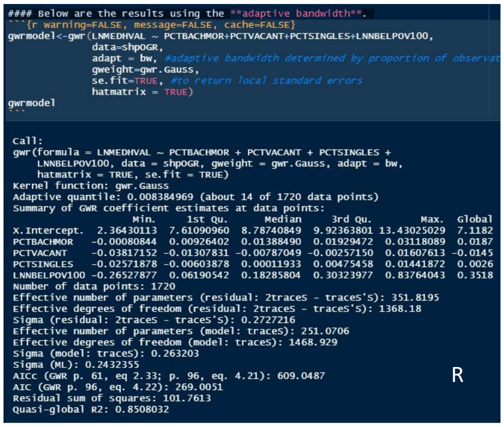


Overall, based on the Akaike Information Criterion/Schwarz Criterion, the Log Likelihood, the Likelihood Ratio Test, and the Moran's I scatterplots, we can confidently say the spatial error regression model is doing better than the OLS regression model.

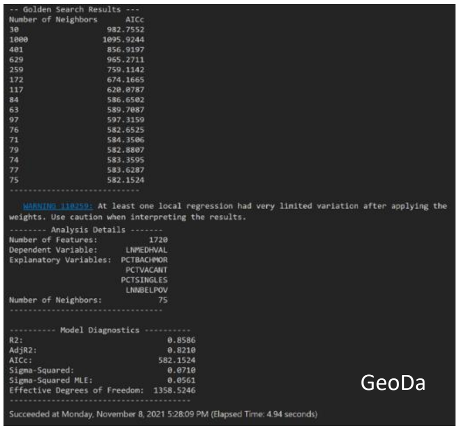
Then which model is better, the Spatial Lag Regression model or the Spatial Error Regression model? Since the models are not nested (i.e., neither method is a special subtype of each other), the likelihood-ratio test cannot be used for this comparison. However, it is OK to compare the two non-nested models based on Akaike Information Criterion and the Schwarz Information Criterion. From the regression results that are listed above, we can see that the value of the Akaike Information Criterion for Spatial Lag Regression model is 523.48, and the value for Spatial Error Regression model is 755.381. That says, the Spatial Lag Regression model is doing better than the Special Error Regression model. When it comes to the Schwarz criterion, the value for Spatial Error Regression model is 782.631, and the value for Spatial Lag Regression model is 556.18, also indicating the Spatial Lag Regression model is better.

d) Geographically Weighted Regression

Given that the overall R-squared of the GWR regression is 0.85 (both in R and ArcGIS) while the R-squared of the OLS regression is only 0.66, the GWR regression appears to explain more variance than the OLS one in the dependent variable. In addition, the Akaike Information Criteria of GWR is 269 in R (ArcGIS doesn't provide AIC), which is comparatively lower than the AIC of OLS (1433), Spatial Lag (470), and Spatial Error (731) models. Considering that the lower the AIC is the better the model is fitted, the GWR model does the best job here. The regression summary for GWR in the R and the supplementary table of it on ArcGIS are shown below.

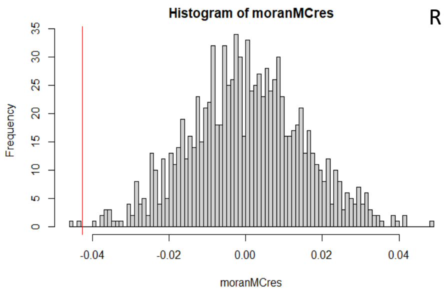


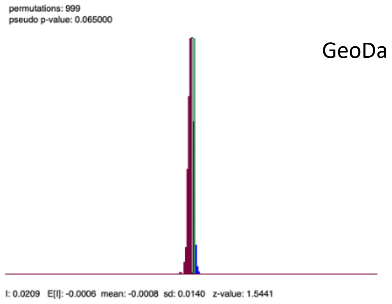
~~~
#### Below are the results using the **adaptive bandwidth**.
```{r warning=FALSE, message=FALSE, cache=FALSE}
gwrmodel<-gwr(LNMEDHVAL ~ PCTBACHMOR+PCTVACANT+PCTSINGLES+LNNBELPOV100,
              data=shpOGR,
              adapt = bw, #adaptive bandwidth determined by proportion of observat
              gweight=gwr.Gauss,
              se.fit=TRUE, #to return local standard errors
              hatmatrix = TRUE)
gwrmodel
```

 Call:
 gwr(formula = LNMEDHVAL ~ PCTBACHMOR + PCTVACANT + PCTSINGLES +
     LNNBELPOV100, data = shpOGR, gweight = gwr.Gauss, adapt = bw,
     hatmatrix = TRUE, se.fit = TRUE)
 Kernel function: gwr.Gauss
 Adaptive quantile: 0.008384969 (about 14 of 1720 data points)
 Summary of GWR coefficient estimates at data points:
                    Min.     1st Qu.      Median     3rd Qu.        Max.  Global
 X.Intercept.  2.36430113  7.61090960  8.78740849  9.92363801 13.43025029  7.1182
 PCTBACHMOR   -0.00080844  0.00926402  0.01388490  0.01929472  0.03118089  0.0187
 PCTVACANT    -0.03817152 -0.01307831 -0.00787049 -0.00257150  0.01607613 -0.0145
 PCTSINGLES   -0.02571878 -0.00603878  0.00011933  0.00475458  0.01441872  0.0026
 LNNBELPOV100 -0.26527877  0.06190542  0.18285804  0.30323977  0.83764043  0.3518
 Number of data points: 1720
 Effective number of parameters (residual: 2traceS - traceS'S): 351.8195
 Effective degrees of freedom (residual: 2traceS - traceS'S): 1368.18
 Sigma (residual: 2traceS - traceS'S): 0.2727216
 Effective number of parameters (model: traceS): 251.0706
 Effective degrees of freedom (model: traceS): 1468.929
 Sigma (model: traceS): 0.263203
 Sigma (ML): 0.2432355
 AICc (GWR p. 61, eq 2.33; p. 96, eq. 4.21): 609.0487
 AIC (GWR p. 96, eq. 4.22): 269.0051
 Residual sum of squares: 101.7613
 Quasi-global R2: 0.8508032
~~~

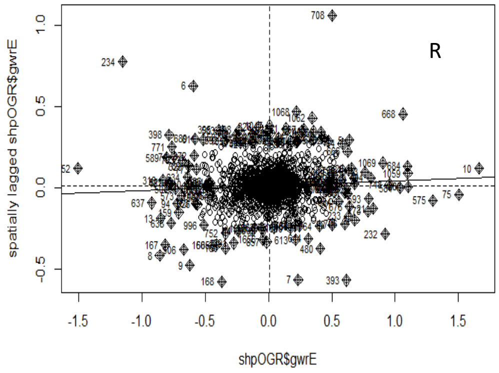


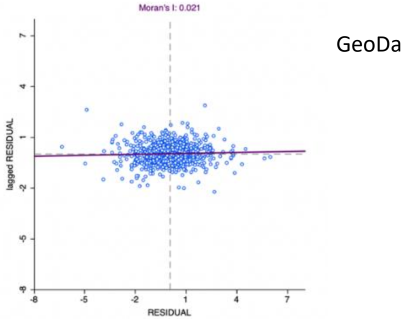
```
-- Golden Search Results ---
Number of Neighbors      AICc
30                   982.7552
1000                1095.9244
401                  856.9197
629                  965.2711
259                  759.1142
172                  674.1665
117                  620.0787
84                   586.6502
63                   589.7087
97                   597.3159
76                   582.6525
71                   584.3506
79                   582.8807
74                   583.3595
77                   583.6287
75                   582.1524
----------------------------

   WARNING 110259: At least one local regression had very limited variation after applying the
weights. Use caution when interpreting the results.
-------- Analysis Details -------
Number of Features:           1720
Dependent Variable:      LNMEDHVAL
Explanatory Variables:  PCTBACHMOR
                         PCTVACANT
                        PCTSINGLES
                         LNNBELPOV
Number of Neighbors:            75
---------------------------------

---------- Model Diagnostics ----------
R2:                              0.8586
AdjR2:                           0.8210
AICc:                          582.1524
Sigma-Squared:                   0.0710
Sigma-Squared MLE:               0.0561
Effective Degrees of Freedom:  1358.5246
--------------------------------------
Succeeded at Monday, November 8, 2021 5:28:09 PM (Elapsed Time: 4.94 seconds)
```

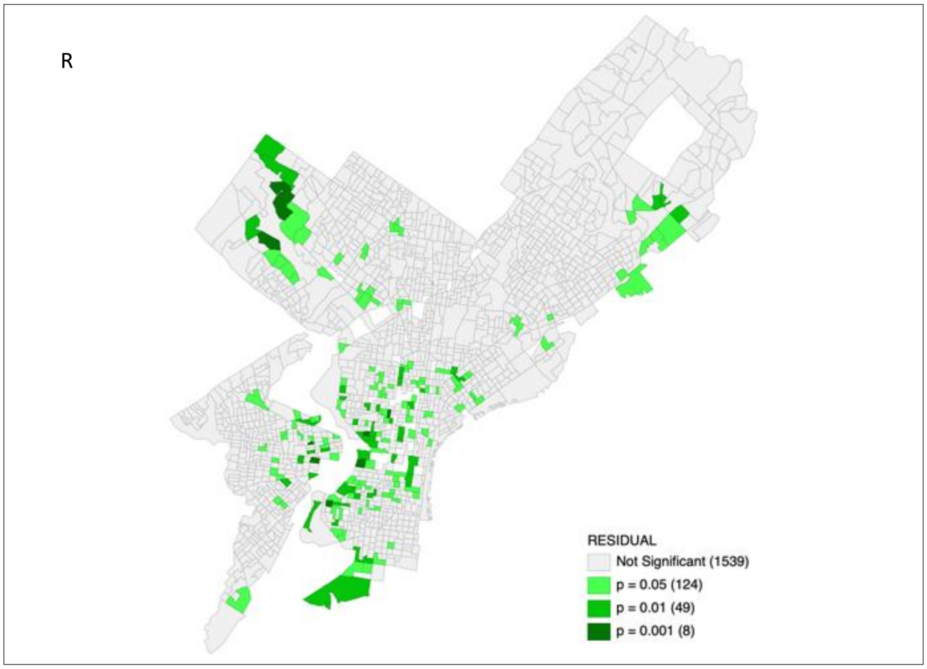
The global Moran's I value of GWR residuals are 0.029 and 0.021, respectively in R and ArcGIS, which are closer to than global Moran's I of these residuals in OLS , Spatial Lag and Spatial Error regressions. Thus, there's less spatial autocorrelation in residuals of the GWR regression, indicating GWR has estimated more spatial relations for this data. Here are Moran's I results, both the Moran scatterplot and the significance test both from R and GeoDa.





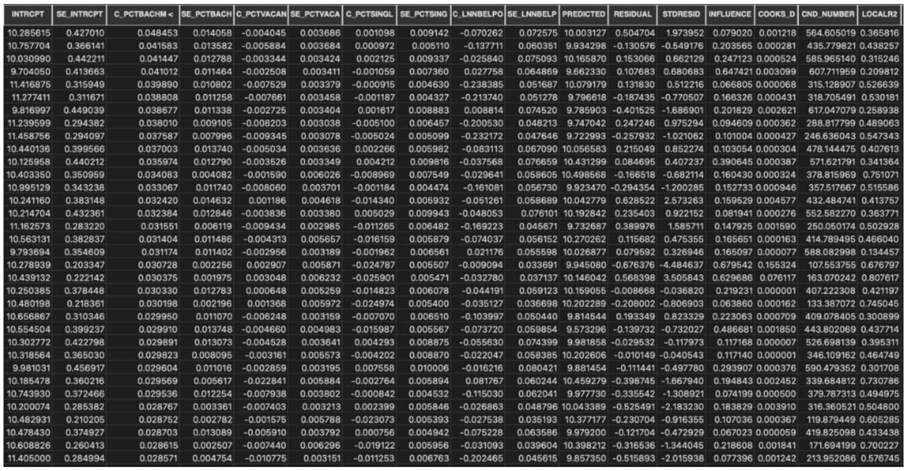






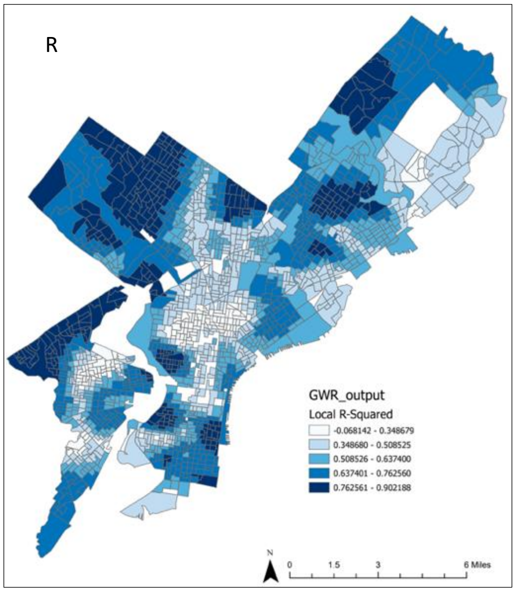
Now, explore local regression results, a screenshot of the attribute table and maps of the ratio of the beta coefficients and the standard error estimates, and take the observation at the first row as an example. For the first observation, 36.58% of variance are explained in $LNMEDHVAL$ with a Local R-squared of 0.3658. For coefficients, when $PCTVACANT$, PCTBACHM, LNNBELPOV100 and $PCTSINGLES$ are all 0, $PREDICTED$ LNMEDHVAL = INTRCPT = 10.29. As PCTBACHM increases by 1 unit (1%), $LNMEDHVAL$ increases by $(e^{0.048} - 1) * 100\%$; as $PCTVACANT$ increases by 1 unit (1%), $LNMEDHVAL$ decreases by $(e^{0.004} - 1) * 100\%$; as $PCTSINGLES$ increases by 1 unit (1%), $LNMEDHVAL$ increases by $(e^{0.001} - 1) * 100\%$; as LNNBELPOV100 increases by 1% , $LNMEDHVAL$ decreases by $(1.01^{0.070} - 1) * 100\%$. At last, Condition number 564.61 > 30, so it performs poorly in terms of multicollinearity. And most of the condition numbers in this regression are larger than 30, indicating that the assumption of local multicollinearity is violated in most places.

GeoDa



INTRCPT	SE_INTRCPT	C_PCTBACHM <	SE_PCTBACH	C_PCTVACAN	SE_PCTVACA	C_PCTSINGL	SE_PCTSING	C_LNNBELPO	SE_LNNBELP	PREDICTED	RESIDUAL	STDRESID	INFLUENCE	COOKS_D	CND_NUMBER	LOCALR2
10.285615	0.427010	0.048453	0.014058	-0.004045	0.003686	0.001098	0.009142	-0.070262	0.072575	10.003127	0.504704	1.973952	0.079020	0.001218	564.605019	0.365816
10.757704	0.366141	0.041583	0.013582	-0.005884	0.003684	0.000972	0.005110	-0.137711	0.060351	9.934298	-0.130576	-0.549176	0.203565	0.000281	435.779821	0.438257
10.030990	0.442211	0.041447	0.012788	-0.003344	0.003424	0.002125	0.009337	-0.025840	0.075093	10.165870	0.153066	0.662129	0.247123	0.000524	585.965140	0.315246
9.704050	0.413663	0.041012	0.011464	-0.002508	0.003411	-0.001059	0.007360	0.027758	0.064869	9.662330	0.107683	0.680683	0.647421	0.003099	607.711959	0.209812
11.416875	0.315949	0.039890	0.010802	-0.007529	0.003379	-0.000915	0.004630	-0.238385	0.051687	10.079179	0.131830	0.512216	0.066805	0.000068	315.128907	0.526639
11.277411	0.311671	0.038808	0.011258	-0.007661	0.003458	-0.001187	0.004327	-0.213740	0.051278	9.796618	-0.187435	-0.770507	0.166326	0.000431	318.705491	0.530181
9.816997	0.449039	0.038677	0.011338	-0.002725	0.003404	0.001617	0.008883	0.008814	0.074520	9.785903	-0.401525	-1.686901	0.201829	0.002621	617.047079	0.258938
11.239599	0.294382	0.038010	0.009105	-0.008203	0.003038	-0.005100	0.006457	-0.200530	0.048213	9.747042	0.247246	0.975294	0.094609	0.000362	288.817799	0.489063
11.458756	0.294097	0.037587	0.007996	-0.009345	0.003078	-0.005024	0.005099	-0.232172	0.047646	9.722993	-0.257932	-1.021062	0.101004	0.000427	246.636043	0.547343
10.440136	0.399566	0.037003	0.013740	-0.005034	0.003636	0.002266	0.005982	-0.083113	0.067090	10.056583	0.215049	0.852274	0.103054	0.000304	478.144475	0.407613
10.125958	0.440212	0.035974	0.012790	-0.003526	0.003349	0.004212	0.009816	-0.037568	0.076659	10.431299	0.084695	0.407237	0.390645	0.000387	571.621791	0.341364
10.403350	0.350959	0.034083	0.004082	-0.001590	0.006026	-0.008969	0.007549	-0.029641	0.058605	10.498568	-0.166518	-0.682114	0.160430	0.000324	378.815969	0.751071
10.995129	0.343238	0.033067	0.011740	-0.008060	0.003701	-0.001184	0.004474	-0.161081	0.056730	9.923470	-0.294354	-1.200285	0.152733	0.000946	357.517667	0.515586
10.241160	0.383148	0.032420	0.014632	0.001186	0.004618	-0.014340	0.005932	-0.051261	0.058689	10.042779	0.628522	2.573263	0.159529	0.004577	432.484741	0.413757
10.214704	0.432361	0.032384	0.012846	-0.003836	0.003380	0.005029	0.009943	-0.048053	0.076101	10.192842	0.235403	0.922152	0.081941	0.000276	552.582270	0.363771
11.162573	0.283220	0.031551	0.006119	-0.009434	0.002985	-0.011265	0.006482	-0.169223	0.045671	9.732687	0.389976	1.585711	0.147925	0.001590	250.050174	0.502928
10.563131	0.382837	0.031404	0.011486	-0.004313	0.005657	-0.016159	0.005879	-0.074037	0.056152	10.270262	0.115682	0.475355	0.165651	0.000163	414.789495	0.466040
9.793694	0.354609	0.031174	0.011402	-0.002956	0.003189	-0.001962	0.006561	0.021176	0.055598	10.026877	0.079592	0.326946	0.165097	0.000077	588.082998	0.134457
10.278939	0.203347	0.030728	0.002256	0.002907	0.005871	-0.024787	0.005507	-0.009094	0.033691	9.945080	-0.676376	-4.484637	0.679542	0.155324	107.553755	0.676797
10.439132	0.222142	0.030375	0.001975	0.003048	0.006232	-0.025901	0.005471	-0.032780	0.037137	10.146042	0.568398	3.505843	0.629686	0.076117	163.070242	0.807617
10.250385	0.378448	0.030330	0.012783	0.000648	0.005259	-0.014823	0.006078	-0.044191	0.059123	10.159055	-0.008668	-0.036820	0.219231	0.000001	407.222308	0.421197
10.480198	0.218361	0.030198	0.002196	0.001368	0.005972	-0.024974	0.005400	-0.035127	0.036698	10.202289	-0.208002	-0.806903	0.063860	0.000162	133.387072	0.745045
10.656867	0.310346	0.029950	0.011070	-0.006248	0.003159	-0.007070	0.006510	-0.103997	0.050440	9.814544	0.193349	0.823329	0.223063	0.000709	409.078405	0.300899
10.554504	0.399237	0.029910	0.013748	-0.004660	0.004983	-0.015987	0.005567	-0.073720	0.059854	9.573296	-0.139732	-0.732027	0.486681	0.001850	443.802069	0.437714
10.302772	0.422798	0.029891	0.013073	-0.004528	0.003641	0.004293	0.008875	-0.055630	0.074399	9.981858	-0.029532	-0.117973	0.117168	0.000007	526.698139	0.395311
10.318564	0.365030	0.029823	0.008095	-0.003161	0.005573	-0.004202	0.008870	-0.022047	0.058385	10.202606	-0.010149	-0.040543	0.117140	0.000001	346.109162	0.464749
9.981031	0.456917	0.029604	0.011016	-0.002859	0.003195	0.007558	0.010006	-0.016216	0.080421	9.881454	-0.111441	-0.497780	0.293907	0.000376	590.479352	0.301708
10.185478	0.360216	0.029569	0.005617	-0.022841	0.005884	-0.002764	0.005894	0.081767	0.060244	10.459279	-0.398745	-1.667940	0.194843	0.002452	339.684812	0.730786
10.743930	0.372466	0.029536	0.012254	-0.007938	0.003802	-0.000842	0.004532	-0.115030	0.062041	9.977730	-0.335542	-1.308921	0.074199	0.000500	379.787313	0.494975
10.200074	0.285382	0.028767	0.003361	-0.007403	0.003213	0.002399	0.005846	-0.026863	0.048796	10.043389	-0.525491	-2.183230	0.183829	0.003910	316.360521	0.504800
10.482931	0.210205	0.028752	0.002782	-0.001575	0.005788	-0.023073	0.005393	-0.027538	0.035193	10.377177	-0.230704	-0.916355	0.107036	0.000367	119.879449	0.605285
10.478430	0.374927	0.028703	0.013089	-0.005910	0.003792	0.000756	0.004942	-0.075228	0.063586	9.979200	-0.121704	-0.472929	0.067023	0.000059	419.825098	0.433438
10.608826	0.260413	0.028615	0.002507	-0.007440	0.006296	-0.019122	0.005956	-0.031093	0.039604	10.398212	-0.316536	-1.344045	0.218608	0.001841	171.694199	0.700227
11.405000	0.284994	0.028571	0.004754	-0.010775	0.003151	-0.011253	0.006763	-0.202465	0.045615	9.857350	-0.515893	-2.015938	0.077396	0.001242	213.952086	0.576745

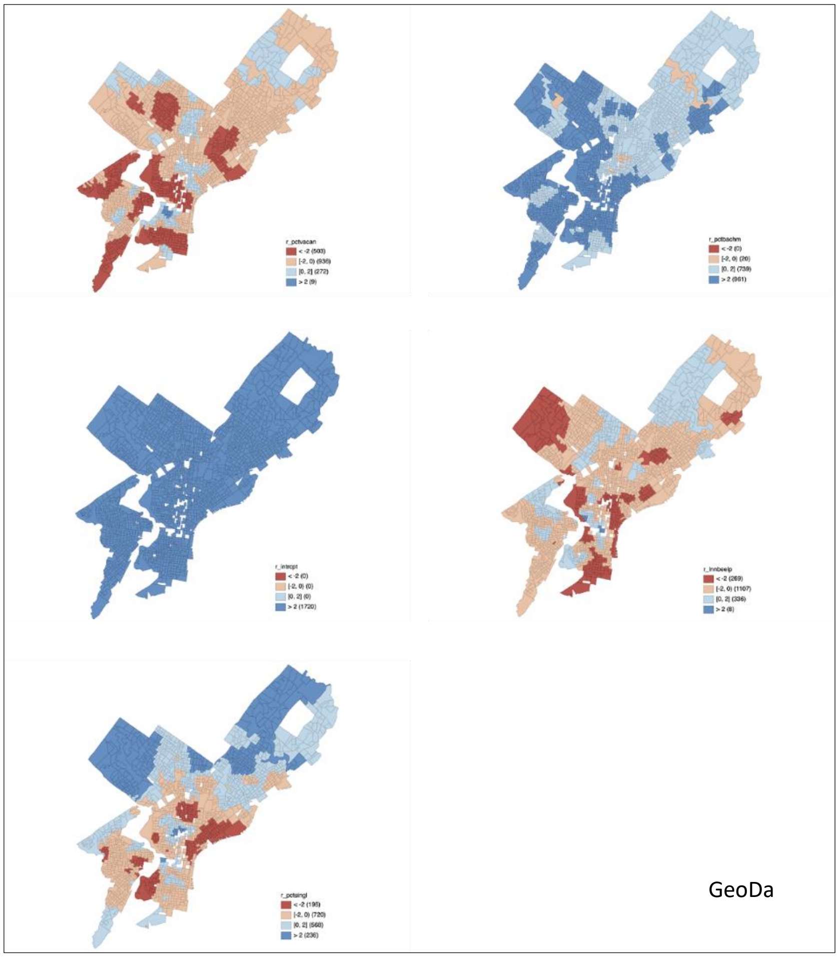
To evaluate how local regressions have been fitted, map the local R-squared values for GWR. As the choropleth map attached below, the GWR model fits poorly in the central section of Philadelphia and west Philadelphia with low local R-squared values, while most parts of Philadelphia are well fitted, especially the area of Mt. Airy and its surrounding neighborhood.



Then, test the local significance of the intercept and four predictors. Below are maps of the ratio of the beta coefficients and their standard error estimates, where the breaks and color schemes are as follow:

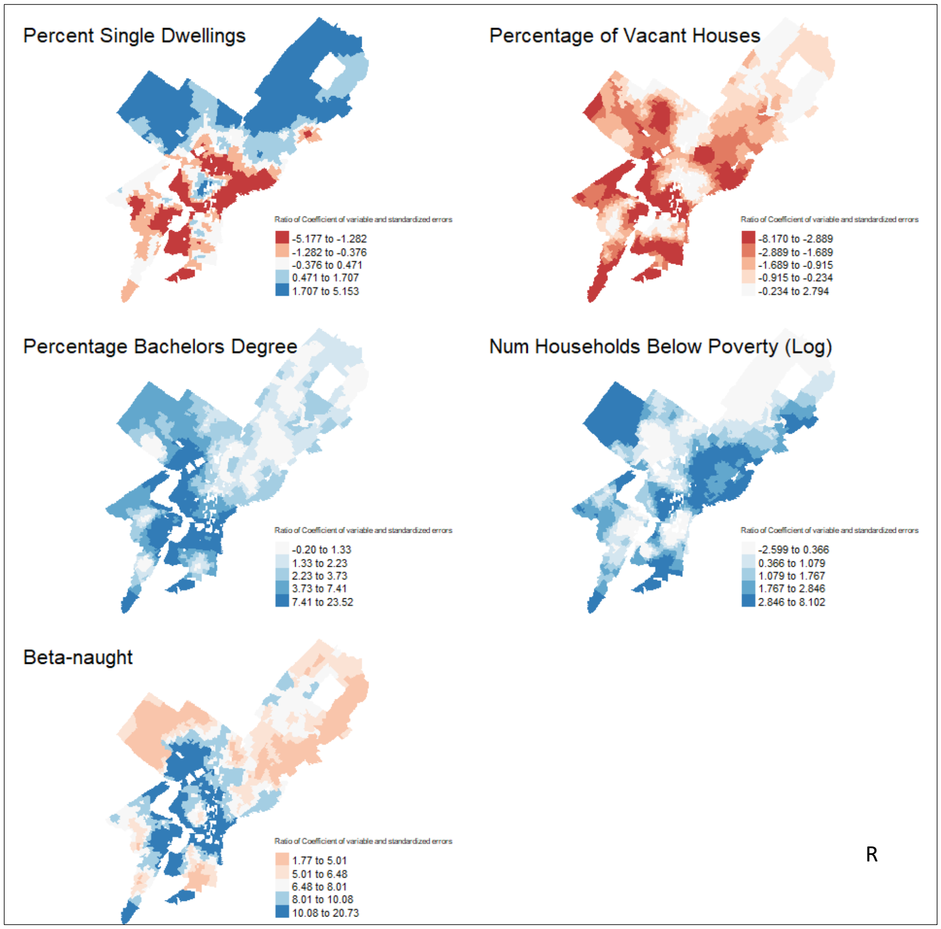
- Dark Blue: Coef/SE <= -2, A negative relationship with the dependent variable that's possibly significant
- Light Blue: - 2 < Coef/SE <= 0: A negative relationship with the dependent variable that's likely not significant
- Pink: 0 < Coef/SE < 2: A positive relationship with the dependent variable that's likely not significant
- Dark Red: Coef/SE >= 2: A positive relationship with the dependent variable that's possibly significant

According to this criteria, all intercept values are significant and have positive relationships with LNMEDHVAL, and most PCTBACHM values have positive relationships with LNMEDHVAL but only half of them are significant. For *PCTVACANT*, LNNBELPOV100 and *PCTSINGLES* values, there are both positive and negative relationships with



GeoDa

LNMEDHVAL among different locations in Philadelphia, and only coefficient values of some locations are significantly different from 0.



4) Discussion

In this study, we utilized R, GeoDa, and ArcGIS Pro to compare OLS, Spatial Lag, Spatial Error, and Geographically Weighted regressions to build models which predict the median house value for census tracts in Philadelphia. We utilized several diagnostic methods for comparing the models such as Global Moran's I, Local Moran's I, the Akaike Information Criterion (AIC), the log-likelihood, and the likelihood ratio.

Based on our results, GWR is the most effective method as it accounts for most of the spatial autocorrelation that exists in the original dataset. The Global Moran's I for the GWR is 0.02. Additionally, the AIC for the GWR is the lowest value (269) which indicates a well fit model. The spatial lag regression is the next most effective model with a Global Moran's I of -0.08, indicating slight negative spatial autocorrelation, and an AIC of 523 (values differ based on software used). The spatial error regression is the next most effective model with a Global Moran's I of -0.094 and an AIC of 755.38. Finally, the OLS regression is the least effective with a Global Moran's I of 0.312 and an AIC of 1432.

In the various models, we observed similar distributions of significant local spatial autocorrelation in the residuals. The spatial models (GWR, spatial lag, and spatial error) performed poorly for block groups that were on the extremes of the median house values. For instance, there was statistically significant spatial autocorrelation in the residuals of the GWR model in block groups in Northwest and Center City, Philadelphia, areas which traditionally have relatively high and low median house values, respectively. The observations made about the GWR model are also true for the spatial lag model and spatial error model. Finally, when comparing the spatial autocorrelation in the residuals, it is apparent that the OLS model is unable to account for spatial autocorrelation in Northeast Philadelphia, an area that was accounted for by the other models.

Regarding limitations of the model, the Jarque-Bera test validating the assumption of normal residuals, is not met in the OLS Regression model. Additionally, the assumption of homoscedasticity was violated for all models per the Breush-Pagan Test for heteroscedasticity (all were statistically significant, $p < 0.0001$). Finally, even the GWR model which performed the best, did not fully account for all of the spatial autocorrelation as is evident by the Local Moran's I for the residuals.