Homework 2: Using Spatial Lag, Spatial Error and Geographically Weighted Regression to Predict Median House Values in Philadelphia Block Groups

Yiming Cao, Sujan Kakumanu, Angel Sanaa Rutherford

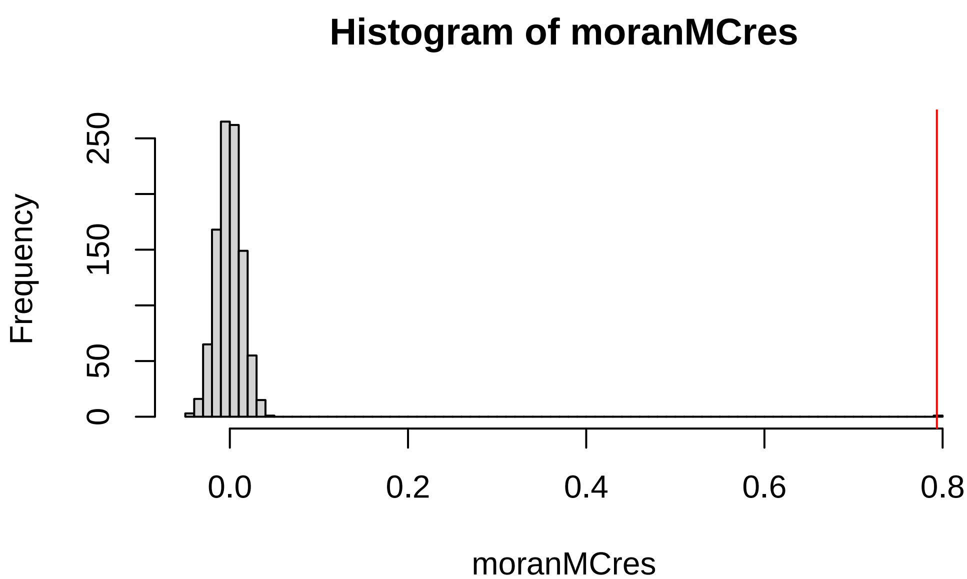
2025-10-30

```
Reading layer `RegressionData' from data source
  `C:\Users\Yiming\Desktop\MUSA\5010 datamining\MUSA-5000\HW 1\Lecture 1 - RegressionData.shp
  using driver `ESRI Shapefile'
Simple feature collection with 1720 features and 13 fields
Geometry type: POLYGON
Dimension:     XY
Bounding box:  xmin: 2660605 ymin: 207610.6 xmax: 2750171 ymax: 304858.8
CRS:           NA
```

```
Warning in poly2nb(Regression_shpData, row.names = Regression_shpData$POLY_ID): neighbour obj
graphs;
if this sub-graph count seems unexpected, try increasing the snap argument.
```

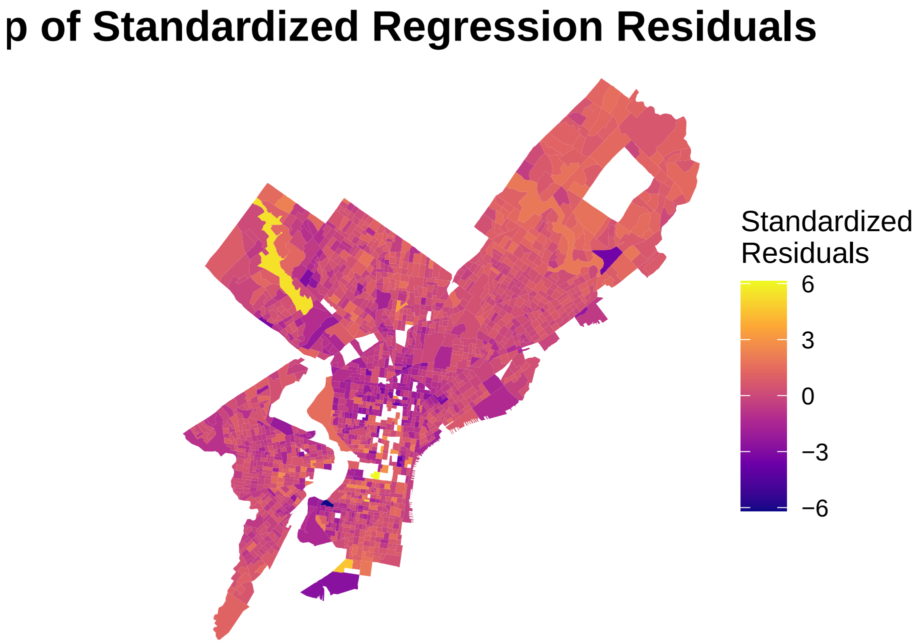
```
Neighbour list object:
Number of regions: 1720
Number of nonzero links: 10526
Percentage nonzero weights: 0.3558004
Average number of links: 6.119767
2 disjoint connected subgraphs
Link number distribution:

  1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  18  27
  4  16  52 175 348 493 344 177  62  28  10   4   2   1   1   1   1   1
4 least connected regions:
441 708 1391 1665 with 1 link
```

```
1 most connected region:
1636 with 27 links

[1] 0.793565


    Monte-Carlo simulation of Moran I

data:  Regression_shpData$LNMEDHVAL
weights: queenlist
number of simulations + 1: 1000

statistic = 0.79356, observed rank = 1000, p-value <
0.00000000000000022
alternative hypothesis: two.sided
```

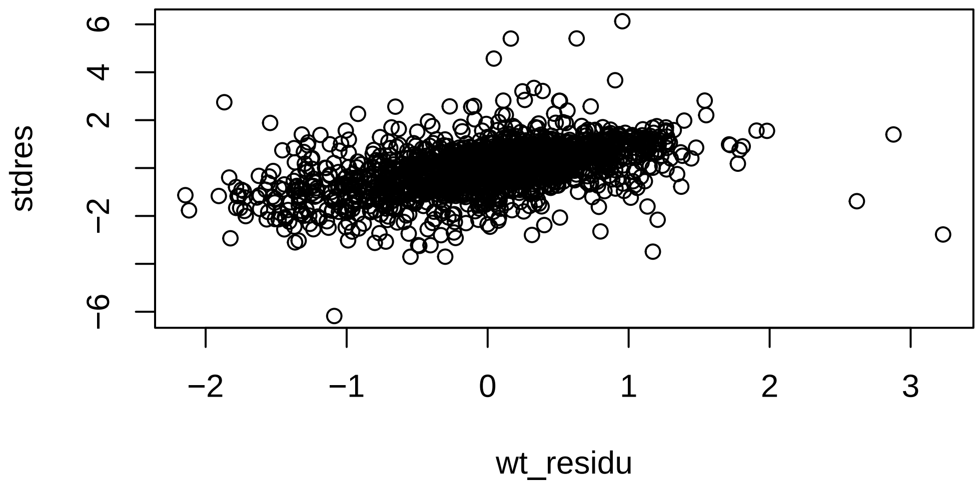



```
          Ii             E.Ii       Var.Ii     Z.Ii  Pr(z != E(Ii))
1 5.35196819 -0.003049833231 1.3051455983 4.687394 0.0000027670601
2 4.41225942 -0.002216601273 0.7590492452 5.066922 0.0000004043007
3 3.50068095 -0.003049833231 0.7444928827 4.060696 0.0000489266898
4 2.44447746 -0.000843880799 0.2410048919 4.981074 0.0000006323230
5 1.88349103 -0.001094174334 0.6259107138 2.382098 0.0172143256212
6 0.09949306 -0.000001607927 0.0009208032 3.278811 0.0010424535943
```

```
Call:
lm(formula = LNMEDHVAL ~ PCTVACANT + PCTSINGLES + PCTBACHMOR +
    LNNBELPOV, data = Regression_shpData)

Residuals:
     Min       1Q   Median       3Q      Max
-2.25817 -0.20391  0.03822  0.21743  2.24345

Coefficients:
              Estimate Std. Error t value             Pr(>|t|)
(Intercept) 11.1137781  0.0465318 238.843 < 0.0000000000000002 ***
PCTVACANT   -0.0191563  0.0009779 -19.590 < 0.0000000000000002 ***
PCTSINGLES   0.0029770  0.0007032   4.234            0.0000242 ***
PCTBACHMOR   0.0209095  0.0005432  38.494 < 0.0000000000000002 ***
LNNBELPOV   -0.0789035  0.0084567  -9.330 < 0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3665 on 1715 degrees of freedom
Multiple R-squared:  0.6623,    Adjusted R-squared:  0.6615
F-statistic: 840.9 on 4 and 1715 DF,  p-value: < 0.00000000000000022

'log Lik.' -711.4933 (df=6)
```




```
  #| echo: false
wt_residu <- sapply(queen, function(x) mean(stdres[x]))

plot(wt_residu, stdres)
```



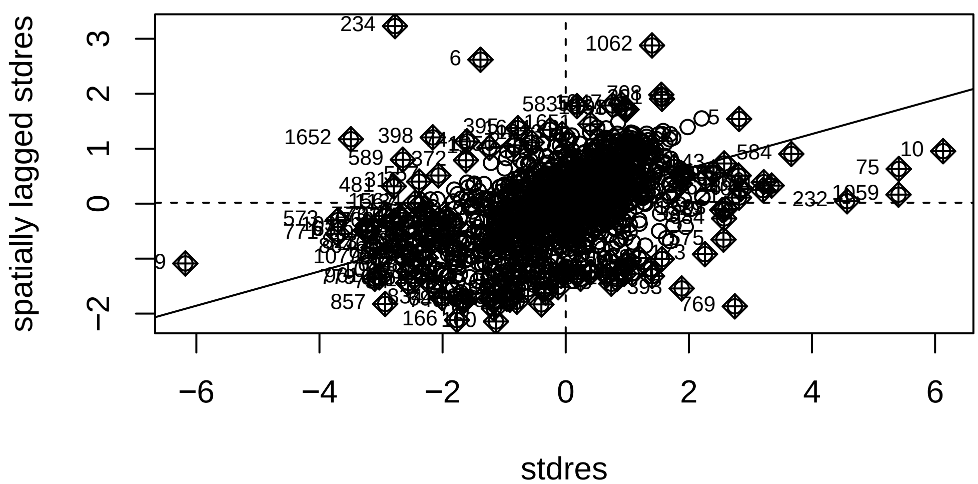

```r
#Note the beta coefficient of the wt_residu. 0.73235. This suggests that there is spatial au
res.lm <- lm(formula=stdres ~ wt_residu)
summary(res.lm)
```

```
Call:
lm(formula = stdres ~ wt_residu)

Residuals:
    Min      1Q  Median      3Q     Max
-5.3685 -0.4450  0.0585  0.4618  5.4435

Coefficients:
            Estimate Std. Error t value            Pr(>|t|)
(Intercept) -0.01281    0.02121  -0.604               0.546
wt_residu    0.73235    0.03244  22.576 <0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8793 on 1718 degrees of freedom
Multiple R-squared:  0.2288,    Adjusted R-squared:  0.2283
F-statistic: 509.7 on 1 and 1718 DF,  p-value: < 0.00000000000000022
```

```r
#Note the small p-value of Moran's I. Same spatial autocorrelation in the scatterplot.
moran.mc(stdres, queenlist, 999, alternative="two.sided")
```

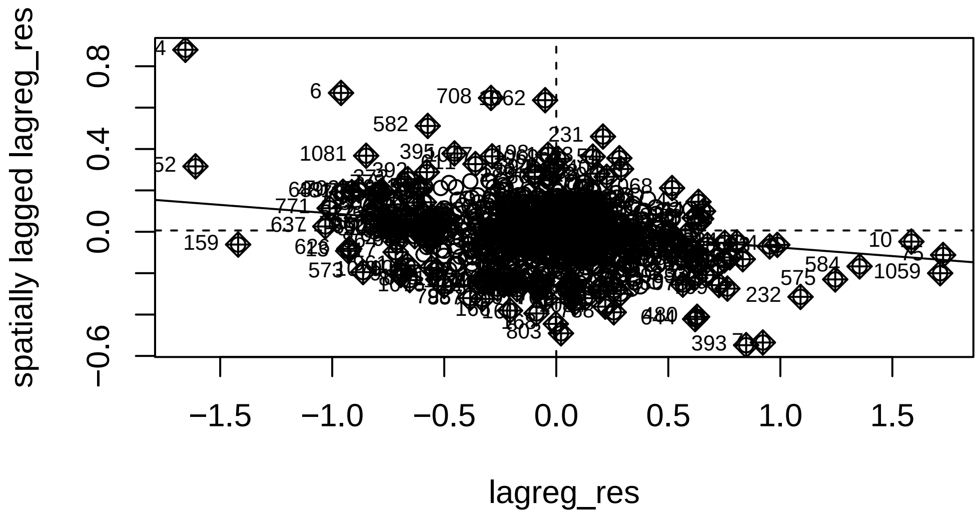
```
    Monte-Carlo simulation of Moran I

data:  stdres
weights: queenlist
number of simulations + 1: 1000

statistic = 0.3124, observed rank = 1000, p-value < 0.00000000000000022
alternative hypothesis: two.sided
```

```r
moran.plot(stdres, queenlist)
```



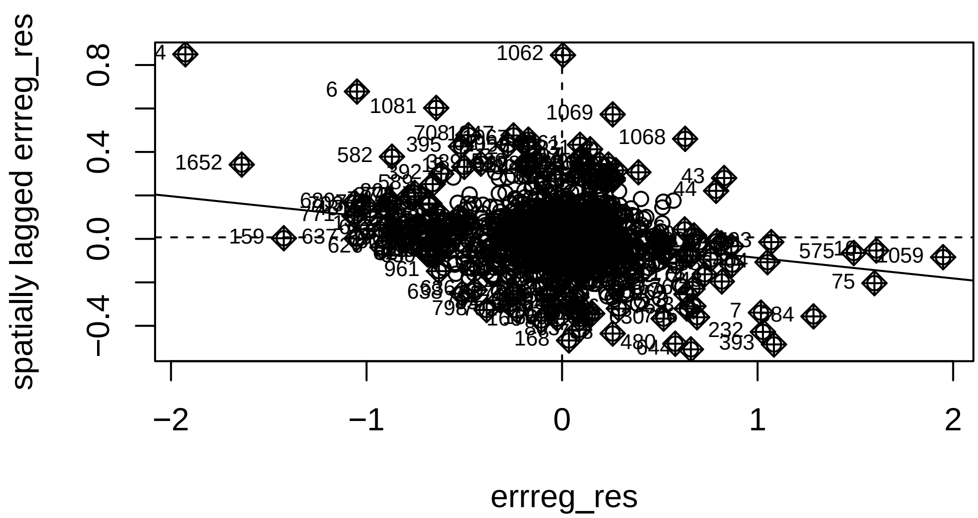

```
Call:lagsarlm(formula = LNMEDHVAL ~ PCTVACANT + PCTSINGLES + PCTBACHMOR +
    LNNBELPOV, data = Regression_shpData, listw = queenlist)

Residuals:
      Min        1Q    Median        3Q       Max
```

```
-1.655421 -0.117248  0.018654  0.133126  1.726436

Type: lag
Coefficients: (asymptotic standard errors)
               Estimate  Std. Error  z value             Pr(>|z|)
(Intercept)  3.89845489  0.20111357  19.3843 < 0.00000000000000022
PCTVACANT   -0.00852940  0.00074367 -11.4694 < 0.00000000000000022
PCTSINGLES   0.00203342  0.00051577   3.9425        0.00008063503
PCTBACHMOR   0.00851381  0.00052193  16.3120 < 0.00000000000000022
LNNBELPOV   -0.03405466  0.00629287  -5.4116        0.00000006246

Rho: 0.6511, LR test value: 911.51, p-value: < 0.000000000000000222
Asymptotic standard error: 0.01805
    z-value: 36.072, p-value: < 0.000000000000000222
Wald statistic: 1301.2, p-value: < 0.000000000000000222

Log likelihood: -255.74 for lag model
ML residual variance (sigma squared): 0.071948, (sigma: 0.26823)
Number of observations: 1720
Number of parameters estimated: 7
AIC: 525.48, (AIC for lm: 1435)
LM test for residual autocorrelation
test value: 67.737, p-value: 0.00000000000000022204


    Monte-Carlo simulation of Moran I

data:  lagreg_res
weights: queenlist
number of simulations + 1: 1000

statistic = -0.082412, observed rank = 1, p-value = 0.002
alternative hypothesis: two.sided
```




```
Call:errorsarlm(formula = LNMEDHVAL ~ PCTVACANT + PCTSINGLES + PCTBACHMOR +
    LNNBELPOV, data = Regression_shpData, listw = queenlist)

Residuals:
      Min        1Q    Median        3Q       Max
-1.926477 -0.115408  0.014889  0.133852  1.948664

Type: error
Coefficients: (asymptotic standard errors)
               Estimate  Std. Error  z value              Pr(>|z|)
(Intercept) 10.90643422  0.05346780 203.9814 < 0.00000000000000022
PCTVACANT   -0.00578308  0.00088670  -6.5220      0.00000000006937
PCTSINGLES   0.00267792  0.00062083   4.3134      0.00001607388269
PCTBACHMOR   0.00981293  0.00072896  13.4615 < 0.00000000000000022
LNNBELPOV   -0.03453408  0.00708933  -4.8713      0.00000110882040

Lambda: 0.81492, LR test value: 677.61, p-value: < 0.000000000000000222
Asymptotic standard error: 0.016373
    z-value: 49.772, p-value: < 0.000000000000000222
Wald statistic: 2477.2, p-value: < 0.000000000000000222

Log likelihood: -372.6904 for error model
```

```
ML residual variance (sigma squared): 0.076551, (sigma: 0.27668)
Number of observations: 1720
Number of parameters estimated: 7
AIC: 759.38, (AIC for lm: 1435)


	Monte-Carlo simulation of Moran I

data:  errreg_res
weights: queenlist
number of simulations + 1: 1000

statistic = -0.094532, observed rank = 1, p-value = 0.002
alternative hypothesis: two.sided
```

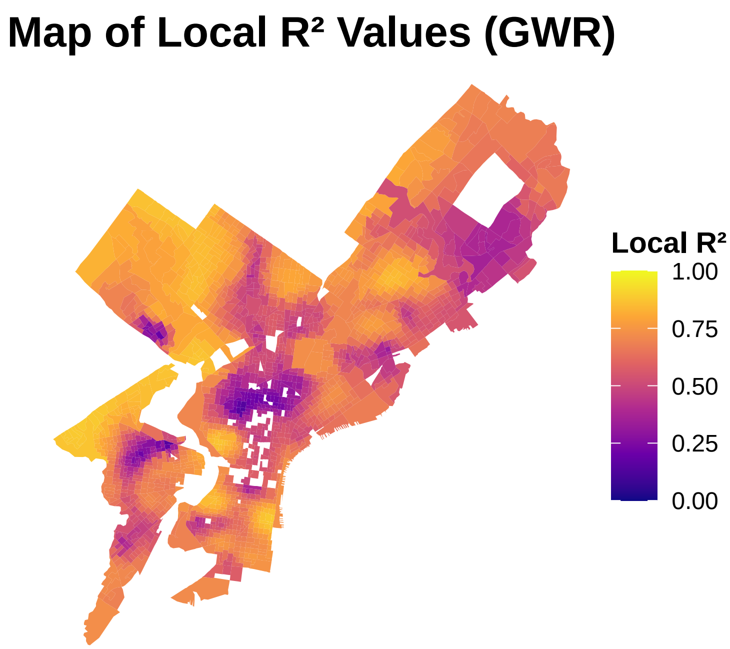


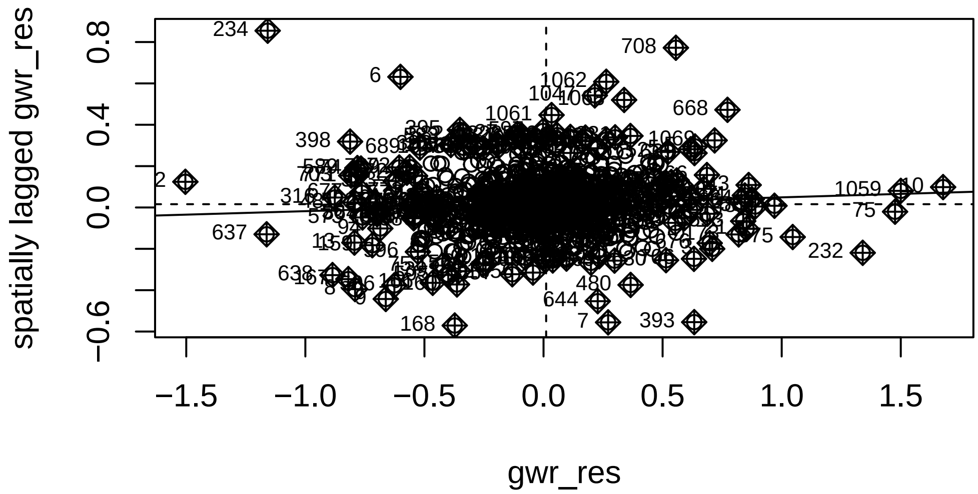

```
[1] "SpatialPolygonsDataFrame"
attr(,"package")
[1] "sp"

Bandwidth: 0.381966 AIC: 1278.637
Bandwidth: 0.618034 AIC: 1333.984
Bandwidth: 0.236068 AIC: 1209.441
```

```
Bandwidth: 0.145898 AIC: 1115.788
Bandwidth: 0.09016994 AIC: 1007.261
Bandwidth: 0.05572809 AIC: 910.3448
Bandwidth: 0.03444185 AIC: 821.2049
Bandwidth: 0.02128624 AIC: 737.5153
Bandwidth: 0.01315562 AIC: 681.5228
Bandwidth: 0.008130619 AIC: 660.7924
Bandwidth: 0.005024999 AIC: 714.1722
Bandwidth: 0.009856235 AIC: 666.9998
Bandwidth: 0.006944377 AIC: 667.5033
Bandwidth: 0.008427513 AIC: 661.6706
Bandwidth: 0.007677515 AIC: 663.5923
Bandwidth: 0.008171309 AIC: 660.8446
Bandwidth: 0.008052658 AIC: 661.0577
Bandwidth: 0.008130619 AIC: 660.7924

[1] 0.008130619

Bandwidth: 47374.26 AIC: 1380.089
Bandwidth: 76576.63 AIC: 1412.319
Bandwidth: 29326.2 AIC: 1314.423
Bandwidth: 18171.89 AIC: 1205.382
Bandwidth: 11278.14 AIC: 1056.784
Bandwidth: 7017.572 AIC: 904.0994
Bandwidth: 4384.396 AIC: 773.8094
Bandwidth: 2757.003 AIC: 701.2702
Bandwidth: 1751.22 AIC: 920.906
Bandwidth: 3378.612 AIC: 714.9353
Bandwidth: 2578.424 AIC: 707.7338
Bandwidth: 2916.626 AIC: 700.5588
Bandwidth: 2860.559 AIC: 700.3531
Bandwidth: 2865.211 AIC: 700.3527
Bandwidth: 2863.515 AIC: 700.3524
Bandwidth: 2863.494 AIC: 700.3524
Bandwidth: 2863.492 AIC: 700.3524
Bandwidth: 2863.493 AIC: 700.3524
Bandwidth: 2862.372 AIC: 700.3525
Bandwidth: 2863.064 AIC: 700.3524
Bandwidth: 2863.329 AIC: 700.3524
Bandwidth: 2863.43 AIC: 700.3524
Bandwidth: 2863.469 AIC: 700.3524
Bandwidth: 2863.483 AIC: 700.3524
```

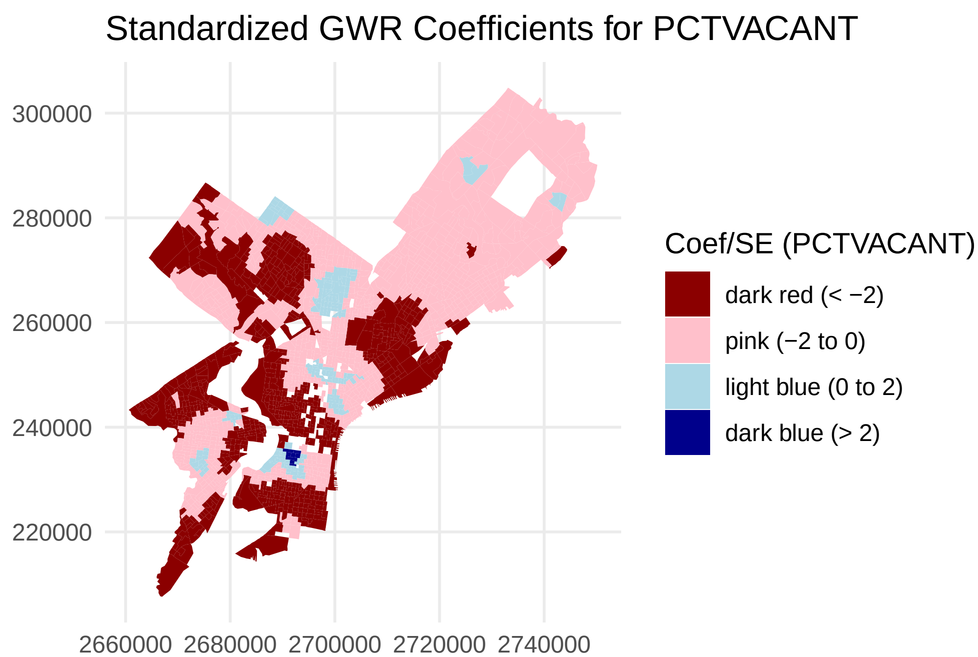
```
Bandwidth: 2863.489 AIC: 700.3524
Bandwidth: 2863.491 AIC: 700.3524
Bandwidth: 2863.492 AIC: 700.3524
Bandwidth: 2863.492 AIC: 700.3524
Bandwidth: 2863.492 AIC: 700.3524
Bandwidth: 2863.492 AIC: 700.3524
Bandwidth: 2863.492 AIC: 700.3524

[1] 2863.492

Call:
gwr(formula = LNMEDHVAL ~ PCTVACANT + PCTSINGLES + PCTBACHMOR +
    LNNBELPOV, data = shps, gweight = gwr.Gauss, adapt = bw,
    hatmatrix = TRUE, se.fit = TRUE)
Kernel function: gwr.Gauss
Adaptive quantile: 0.008130619 (about 13 of 1720 data points)
Summary of GWR coefficient estimates at data points:
                   Min.    1st Qu.     Median    3rd Qu.       Max.  Global
X.Intercept.  9.6727618 10.7143173 10.9542384 11.1742009 12.0831381 11.1138
PCTVACANT    -0.0317407 -0.0142383 -0.0089599 -0.0035770  0.0167916 -0.0192
PCTSINGLES   -0.0249706 -0.0075550 -0.0016626  0.0042280  0.0143340  0.0030
PCTBACHMOR    0.0010974  0.0101380  0.0149279  0.0202187  0.0347258  0.0209
LNNBELPOV    -0.2365244 -0.0733572 -0.0401186 -0.0126657  0.0948768 -0.0789
Number of data points: 1720
Effective number of parameters (residual: 2traceS - traceS'S): 360.5225
Effective degrees of freedom (residual: 2traceS - traceS'S): 1359.477
Sigma (residual: 2traceS - traceS'S): 0.2762201
Effective number of parameters (model: traceS): 257.9061
Effective degrees of freedom (model: traceS): 1462.094
Sigma (model: traceS): 0.2663506
Sigma (ML): 0.245571
AICc (GWR p. 61, eq 2.33; p. 96, eq. 4.21): 660.7924
AIC (GWR p. 96, eq. 4.22): 308.7123
Residual sum of squares: 103.7248
Quasi-global R2: 0.8479244

Call:
gwr(formula = LNMEDHVAL ~ PCTVACANT + PCTSINGLES + PCTBACHMOR +
    LNNBELPOV, data = shps, bandwidth = bw_fixed, gweight = gwr.Gauss,
    hatmatrix = TRUE, se.fit = TRUE)
Kernel function: gwr.Gauss
Fixed bandwidth: 2863.492
```

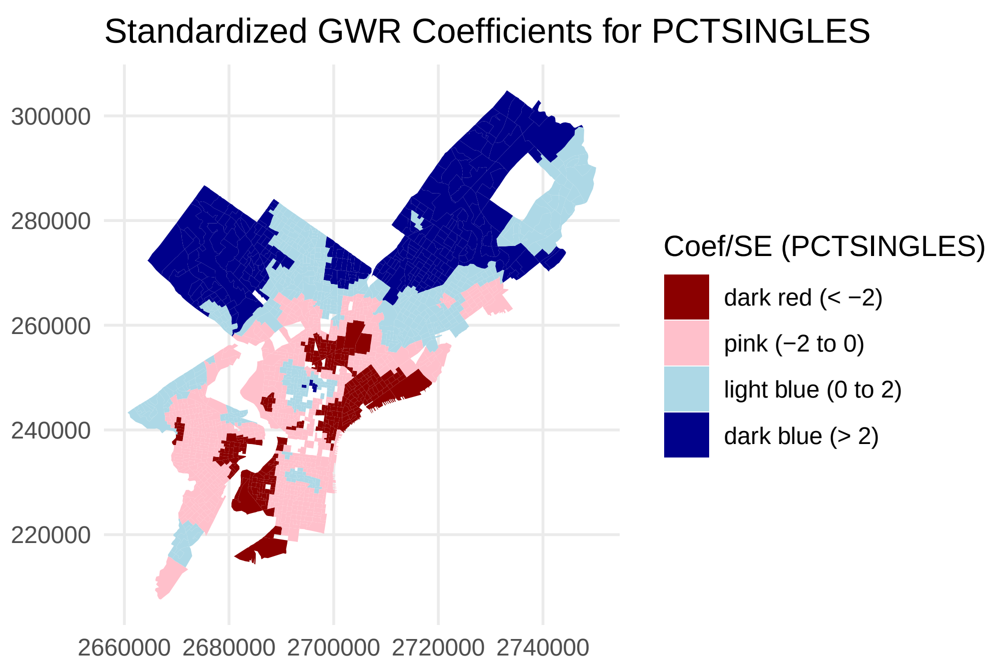
```
Summary of GWR coefficient estimates at data points:
                  Min.    1st Qu.     Median    3rd Qu.       Max.  Global
X.Intercept.  9.9111178 10.7329171 10.9397426 11.1639961 14.1200790 11.1138
PCTVACANT    -0.0469926 -0.0137374 -0.0088796 -0.0038447  0.0778856 -0.0192
PCTSINGLES   -0.0238330 -0.0073895 -0.0025702  0.0040499  0.0189995  0.0030
PCTBACHMOR   -0.0860914  0.0118750  0.0168149  0.0213553  0.0306653  0.0209
LNNBELPOV    -0.4449897 -0.0737744 -0.0433084 -0.0171174  0.1491701 -0.0789
Number of data points: 1720
Effective number of parameters (residual: 2traceS - traceS'S): 346.7181
Effective degrees of freedom (residual: 2traceS - traceS'S): 1373.282
Sigma (residual: 2traceS - traceS'S): 0.2785229
Effective number of parameters (model: traceS): 255.6033
Effective degrees of freedom (model: traceS): 1464.397
Sigma (model: traceS): 0.2697189
Sigma (ML): 0.2488723
AICc (GWR p. 61, eq 2.33; p. 96, eq. 4.21): 700.3524
AIC (GWR p. 96, eq. 4.22): 352.347
Residual sum of squares: 106.5324
Quasi-global R2: 0.843808
```

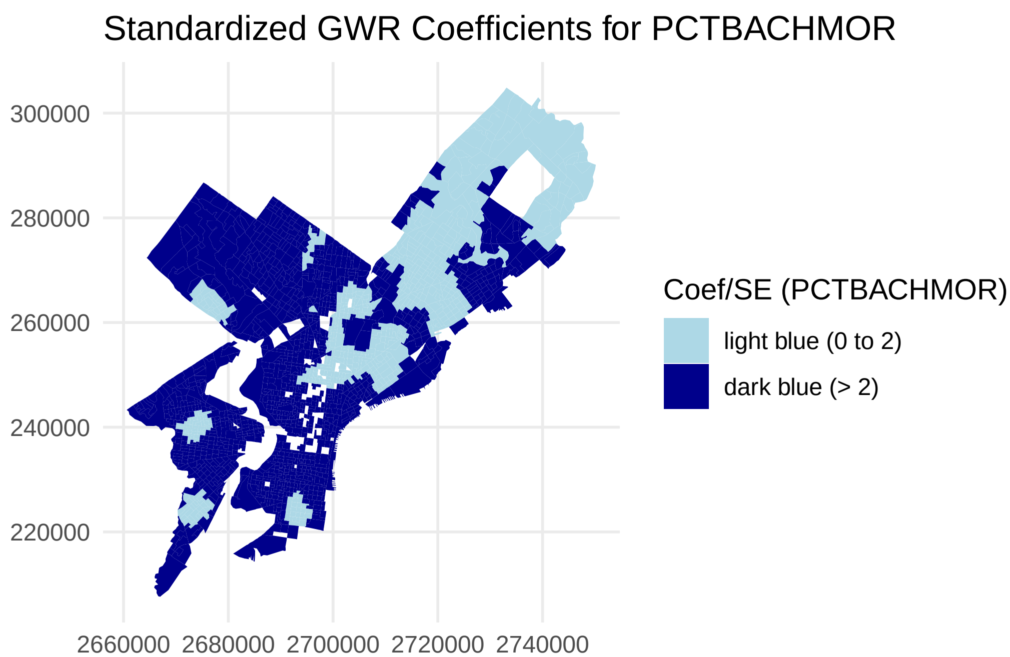


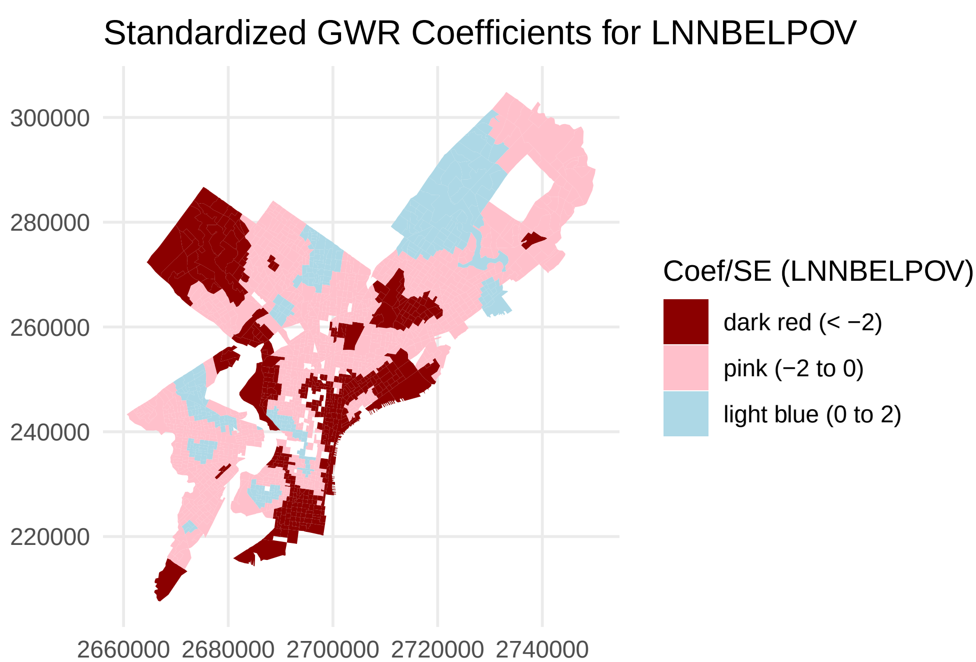


```
    Monte-Carlo simulation of Moran I

data:  gwr_res
weights: queenlist
number of simulations + 1: 1000

statistic = 0.033425, observed rank = 988, p-value = 0.024
alternative hypothesis: two.sided
```


Standardized GWR Coefficients for PCTVACANT
300000
280000
260000
240000
220000
2660000 2680000 2700000 2720000 2740000
Coef/SE (PCTVACANT)
dark red (< −2)
pink (−2 to 0)
light blue (0 to 2)
dark blue (> 2)


Standardized GWR Coefficients for PCTSINGLES
300000
280000
260000
240000
220000
2660000 2680000 2700000 2720000 2740000
Coef/SE (PCTSINGLES)
dark red (< −2)
pink (−2 to 0)
light blue (0 to 2)
dark blue (> 2)


Standardized GWR Coefficients for PCTBACHMOR
300000
280000
260000
240000
220000
2660000 2680000 2700000 2720000 2740000
Coef/SE (PCTBACHMOR)
light blue (0 to 2)
dark blue (> 2)


Standardized GWR Coefficients for LNNBELPOV
300000
280000
260000
240000
220000
2660000 2680000 2700000 2720000 2740000
Coef/SE (LNNBELPOV)
dark red (< −2)
pink (−2 to 0)
light blue (0 to 2)

1 Introduction (Google Doc at end)

2 Methods

2.1 The Concept of Spatial Autocorrelation (Angel)

2.2 A Review of OLS Regression and Assumptions (Angel)

2.3 Spatial Lag and Spatial Error Regression (Sujan)

2.4 Geographically Weighted Regression (Ming)

Geographically Weighted Regression (GWR) is a spatial modeling technique that extends the conventional Ordinary Least Squares (OLS) regression by allowing model parameters to vary across geographic space. Whereas OLS assumes a single global relationship between the dependent and independent variables, GWR accounts for spatial nonstationarity, meaning that the strength and direction of these relationships may differ by location.

2.4.1 Mathematical Formulation

The general form of the OLS model is:

$$y_i = \beta_0 + \sum_{k=1}^{p} \beta_k x_{ik} + \varepsilon_i$$

where:

- y_i is the dependent variable at observation i
- x_{ik} is the value of predictor k at observation i
- β_k are the global regression coefficients
- ε_i is the random error term assumed to be normally distributed and independent

GWR extends this model by allowing each coefficient to vary across geographic space:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i) x_{ik} + \varepsilon_i$$

where (u_i, v_i) are the spatial coordinates of observation i, and $\beta_k(u_i, v_i)$ represents the **local coefficient** for predictor k at that location.

The model estimates parameters at each location using **weighted least squares**, with weights determined by a spatial kernel function.
The local coefficients are estimated as:

$$\hat{\beta}(u_i, v_i) = (X^T W_i X)^{-1} X^T W_i y$$

where:

- X is the matrix of independent variables
- y is the vector of dependent variables
- W_i is a diagonal matrix of spatial weights for location i

2.4.1.1 Global R² (Goodness of Fit)

Similar to OLS regression, GWR reports a global R-squared statistic that measures the overall proportion of variance in the dependent variable explained by the model. Its value ranges from 0 to 1, with higher values indicating better model fit.

The formula for the coefficient of determination is:

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

where: - y_i = observed values,
- $\hat{y}_i$ = predicted values from the model,
- $\bar{y}$ = mean of observed values,
- SSE = sum of squared errors,
- SST = total sum of squares.

2.4.1.2 Bandwidth Selection and Weighting Scheme

In Geographically Weighted Regression (GWR), the bandwidth (h) determines the spatial extent over which data points influence each local regression. Two main approaches are commonly used: **fixed bandwidth** and **adaptive bandwidth**.

2.4.1.2.1 Fixed Bandwidth

Under a fixed bandwidth, the spatial extent h remains constant across all regression points, meaning the geographic area of influence is the same, while the number of observations within each local regression may vary.

$$w_{ij} = \begin{cases} e^{-0.5\left(\frac{distance_{ij}}{h}\right)^2}, & \text{if } distance_{ij} \leq h \\ 0, & \text{otherwise} \end{cases}$$

where $distance_{ij}$ is the distance between observation i and j.
This form is known as a *Gaussian kernel*, which gives higher weight to nearer observations.

2.4.1.2.2 Adaptive Bandwidth

Under an adaptive bandwidth, the number of nearest neighbors (N) is fixed, but the geographic area h varies according to local data density.
Areas with dense data use smaller h, while sparse regions require larger h to include enough neighbors.

$$w_{ij} = \begin{cases} \left[1 - \left(\frac{distance_{ij}}{h}\right)^2\right]^2, & \text{if } j \text{ is one of } i\text{'s } N \text{ nearest neighbors} \\ 0, & \text{otherwise} \end{cases}$$

In this case, h adapts to ensure that each local regression includes approximately the same number of observations.
For example, if $N = 20$, one area might require $h = 5000\,ft$ to reach its 20th neighbor, while another only needs $h = 2500\,ft$.

In this analysis, an **adaptive Gaussian kernel** was used, ensuring that dense urban areas and sparse suburban areas are both modeled effectively.
The optimal bandwidth was chosen by minimizing the **Akaike Information Criterion (AIC)**, which balances model fit and complexity.

To test model robustness, a fixed-bandwidth GWR was also calibrated for comparison.

2.4.2 Model Diagnostics

Model performance was assessed using both global and local indicators:

- **Global diagnostics:** Akaike Information Criterion (AIC, AICc) and quasi-global R^2
- **Local diagnostics:** Local R^2, coefficient maps, and standardized ratios (β/SE)
- **Spatial dependence:** Moran's I statistic of residuals tested whether the GWR residuals remained spatially autocorrelated

Together, these diagnostics determine whether GWR improves explanatory power over OLS and whether it effectively removes spatial dependence from model residuals.

3 Results

3.1 Spatial Autocorrelation (Angel)

3.2 A Review of OLS Regression and Assumptions: Results (Angel)

3.3 Spatial Lag and Spatial Error Regression Results (Sujan)

3.4 Geographically Weighted Regression Results (Ming)

The geographically weighted regression (GWR) model was estimated to assess how the relationships between housing values and neighborhood characteristics vary across space. Both adaptive and fixed bandwidths were tested, and the model diagnostics and spatial patterns were examined to evaluate local model fit and remaining spatial dependence.

3.4.1 Bandwidth Selection

1.Adaptive bandwith

```
[1] "SpatialPolygonsDataFrame"
attr(,"package")
[1] "sp"
```

```
Bandwidth: 0.381966 AIC: 1278.637
Bandwidth: 0.618034 AIC: 1333.984
Bandwidth: 0.236068 AIC: 1209.441
Bandwidth: 0.145898 AIC: 1115.788
Bandwidth: 0.09016994 AIC: 1007.261
Bandwidth: 0.05572809 AIC: 910.3448
Bandwidth: 0.03444185 AIC: 821.2049
Bandwidth: 0.02128624 AIC: 737.5153
Bandwidth: 0.01315562 AIC: 681.5228
Bandwidth: 0.008130619 AIC: 660.7924
Bandwidth: 0.005024999 AIC: 714.1722
Bandwidth: 0.009856235 AIC: 666.9998
Bandwidth: 0.006944377 AIC: 667.5033
Bandwidth: 0.008427513 AIC: 661.6706
Bandwidth: 0.007677515 AIC: 663.5923
Bandwidth: 0.008171309 AIC: 660.8446
Bandwidth: 0.008052658 AIC: 661.0577
Bandwidth: 0.008130619 AIC: 660.7924
```

```
[1] 0.008130619
```

2.Fixed bandwith

```
Bandwidth: 47374.26 AIC: 1380.089
Bandwidth: 76576.63 AIC: 1412.319
Bandwidth: 29326.2 AIC: 1314.423
Bandwidth: 18171.89 AIC: 1205.382
Bandwidth: 11278.14 AIC: 1056.784
Bandwidth: 7017.572 AIC: 904.0994
Bandwidth: 4384.396 AIC: 773.8094
Bandwidth: 2757.003 AIC: 701.2702
Bandwidth: 1751.22 AIC: 920.906
Bandwidth: 3378.612 AIC: 714.9353
Bandwidth: 2578.424 AIC: 707.7338
Bandwidth: 2916.626 AIC: 700.5588
Bandwidth: 2860.559 AIC: 700.3531
Bandwidth: 2865.211 AIC: 700.3527
Bandwidth: 2863.515 AIC: 700.3524
Bandwidth: 2863.494 AIC: 700.3524
Bandwidth: 2863.492 AIC: 700.3524
Bandwidth: 2863.493 AIC: 700.3524
Bandwidth: 2862.372 AIC: 700.3525
```

```
Bandwidth: 2863.064 AIC: 700.3524
Bandwidth: 2863.329 AIC: 700.3524
Bandwidth: 2863.43 AIC: 700.3524
Bandwidth: 2863.469 AIC: 700.3524
Bandwidth: 2863.483 AIC: 700.3524
Bandwidth: 2863.489 AIC: 700.3524
Bandwidth: 2863.491 AIC: 700.3524
Bandwidth: 2863.492 AIC: 700.3524
Bandwidth: 2863.492 AIC: 700.3524
Bandwidth: 2863.492 AIC: 700.3524
Bandwidth: 2863.492 AIC: 700.3524
Bandwidth: 2863.492 AIC: 700.3524
```

```
[1] 2863.492
```

The adaptive bandwidth minimized the AIC and was selected for the final model. This approach adjusts to the spatial density of observations, allowing greater local flexibility in heterogeneous areas.

###Adaptive GWR Model

```
Call:
gwr(formula = LNMEDHVAL ~ PCTVACANT + PCTSINGLES + PCTBACHMOR +
    LNNBELPOV, data = shps, gweight = gwr.Gauss, adapt = bw,
    hatmatrix = TRUE, se.fit = TRUE)
Kernel function: gwr.Gauss
Adaptive quantile: 0.008130619 (about 13 of 1720 data points)
Summary of GWR coefficient estimates at data points:
                    Min.    1st Qu.     Median    3rd Qu.       Max.  Global
X.Intercept.  9.6727618 10.7143173 10.9542384 11.1742009 12.0831381 11.1138
PCTVACANT    -0.0317407 -0.0142383 -0.0089599 -0.0035770  0.0167916 -0.0192
PCTSINGLES   -0.0249706 -0.0075550 -0.0016626  0.0042280  0.0143340  0.0030
PCTBACHMOR    0.0010974  0.0101380  0.0149279  0.0202187  0.0347258  0.0209
LNNBELPOV    -0.2365244 -0.0733572 -0.0401186 -0.0126657  0.0948768 -0.0789
Number of data points: 1720
Effective number of parameters (residual: 2traceS - traceS'S): 360.5225
Effective degrees of freedom (residual: 2traceS - traceS'S): 1359.477
Sigma (residual: 2traceS - traceS'S): 0.2762201
Effective number of parameters (model: traceS): 257.9061
Effective degrees of freedom (model: traceS): 1462.094
Sigma (model: traceS): 0.2663506
Sigma (ML): 0.245571
```

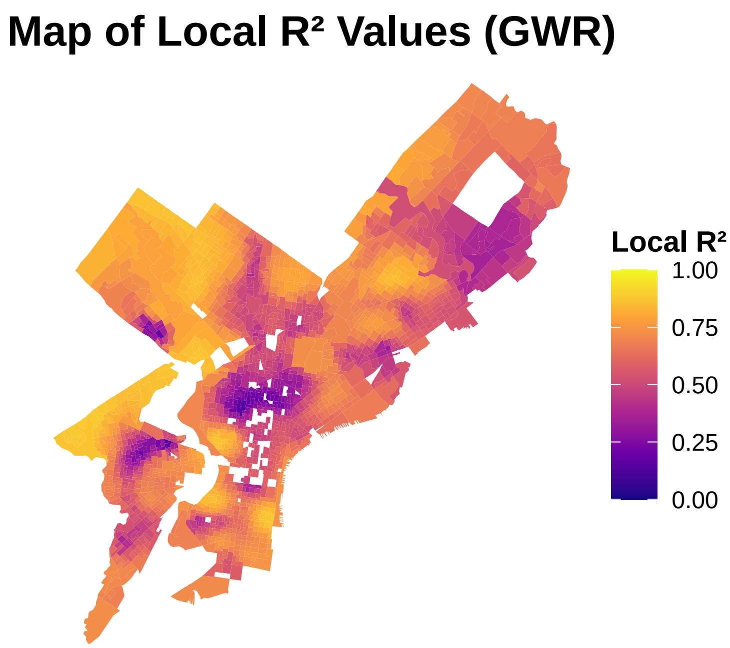
```
AICc (GWR p. 61, eq 2.33; p. 96, eq. 4.21): 660.7924
AIC (GWR p. 96, eq. 4.22): 308.7123
Residual sum of squares: 103.7248
Quasi-global R2: 0.8479244
```

###Fixed Bandwidth Model (for comparison)

```
Call:
gwr(formula = LNMEDHVAL ~ PCTVACANT + PCTSINGLES + PCTBACHMOR +
    LNNBELPOV, data = shps, bandwidth = bw_fixed, gweight = gwr.Gauss,
    hatmatrix = TRUE, se.fit = TRUE)
Kernel function: gwr.Gauss
Fixed bandwidth: 2863.492
Summary of GWR coefficient estimates at data points:
                   Min.    1st Qu.     Median    3rd Qu.       Max.  Global
X.Intercept.  9.9111178 10.7329171 10.9397426 11.1639961 14.1200790 11.1138
PCTVACANT    -0.0469926 -0.0137374 -0.0088796 -0.0038447  0.0778856 -0.0192
PCTSINGLES   -0.0238330 -0.0073895 -0.0025702  0.0040499  0.0189995  0.0030
PCTBACHMOR   -0.0860914  0.0118750  0.0168149  0.0213553  0.0306653  0.0209
LNNBELPOV    -0.4449897 -0.0737744 -0.0433084 -0.0171174  0.1491701 -0.0789
Number of data points: 1720
Effective number of parameters (residual: 2traceS - traceS'S): 346.7181
Effective degrees of freedom (residual: 2traceS - traceS'S): 1373.282
Sigma (residual: 2traceS - traceS'S): 0.2785229
Effective number of parameters (model: traceS): 255.6033
Effective degrees of freedom (model: traceS): 1464.397
Sigma (model: traceS): 0.2697189
Sigma (ML): 0.2488723
AICc (GWR p. 61, eq 2.33; p. 96, eq. 4.21): 700.3524
AIC (GWR p. 96, eq. 4.22): 352.347
Residual sum of squares: 106.5324
Quasi-global R2: 0.843808
```

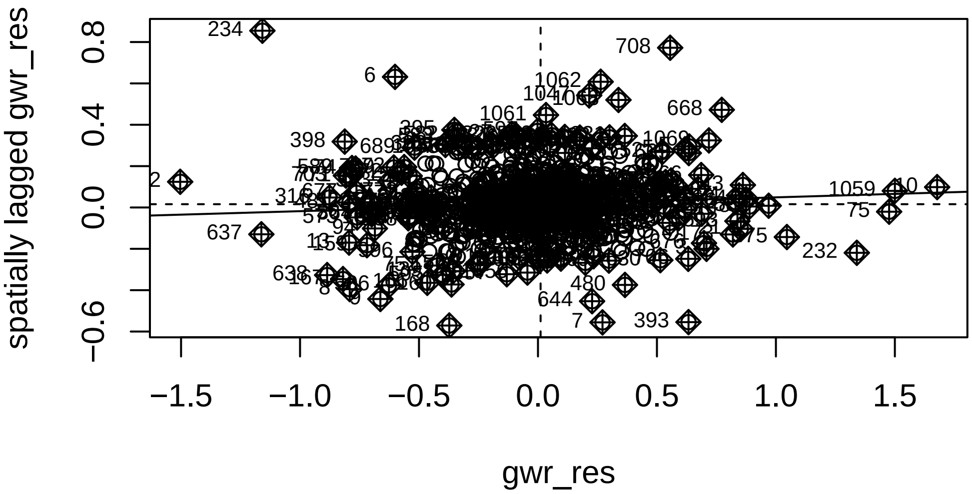
The adaptive GWR model substantially improved performance compared to the global OLS regression, with a lower AIC (308.7) and a higher quasi-global R^2 (0.848), indicating stronger explanatory power after accounting for spatial variation. In contrast, the fixed-bandwidth GWR model produced a higher AIC (352.3) and a lower quasi-global R^2 (0.844), confirming that adaptive bandwidth selection provides a better fit for spatially heterogeneous urban data.

###Local R^2 Mapping



The local R^2 values show substantial spatial variation across Philadelphia. The highest values are concentrated in Center City and University City, where the model explains up to nearly all of the variation in housing values. These areas correspond to dense, high-value neighborhoods with more consistent socioeconomic patterns, allowing the predictors—such as educational attainment and vacancy rate—to capture local dynamics effectively. In contrast, the outer neighborhoods and peripheral tracts, particularly in the Northeast and Northwest sections, exhibit lower local R^2 values, suggesting that housing prices there are influenced by additional unmeasured factors (e.g., land use mix, environmental amenities, or localized market conditions) not fully represented in the model.

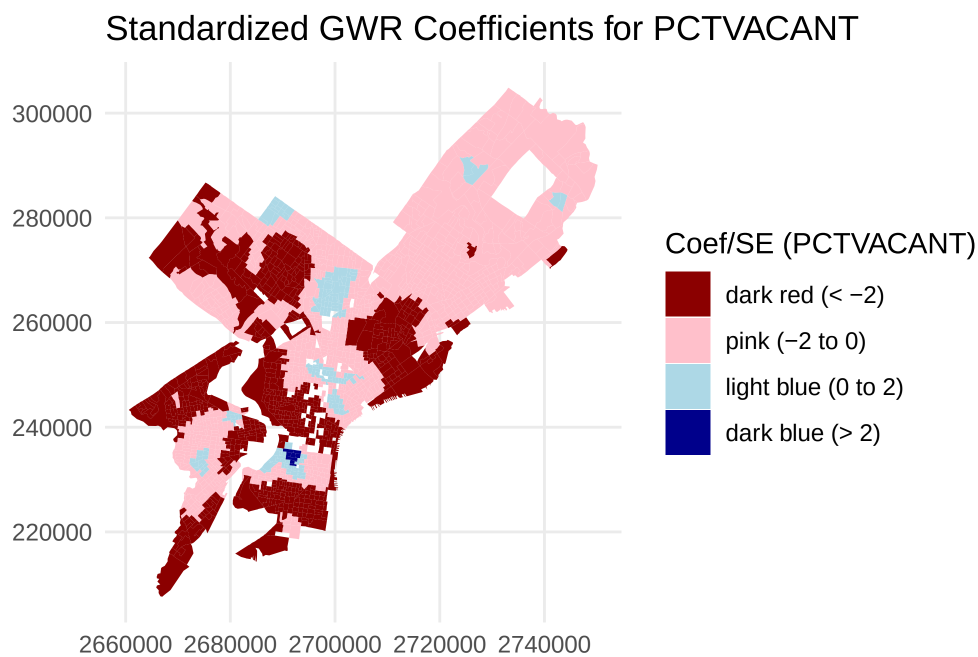
###Moran's I of GWR Residuals

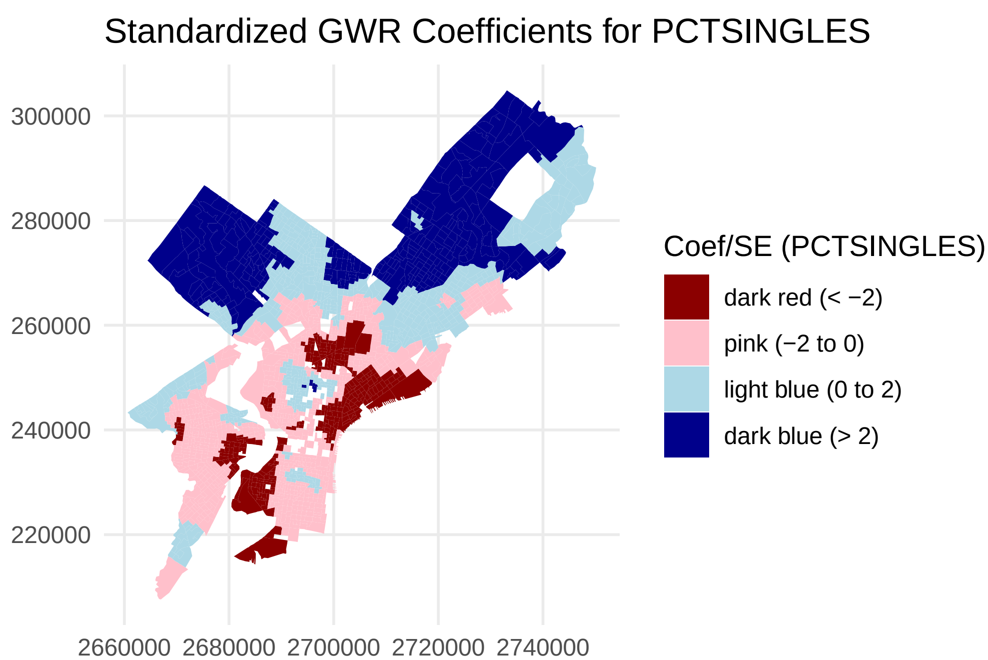


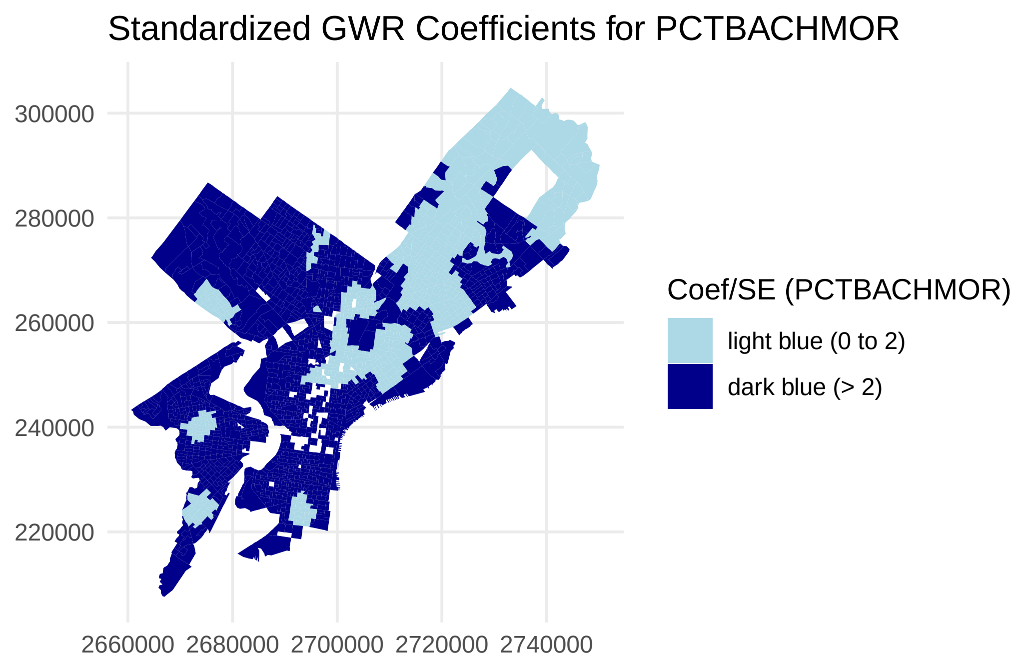

```
    Monte-Carlo simulation of Moran I

data:  gwr_res
weights: queenlist
number of simulations + 1: 1000

statistic = 0.033425, observed rank = 994, p-value = 0.012
alternative hypothesis: two.sided
```

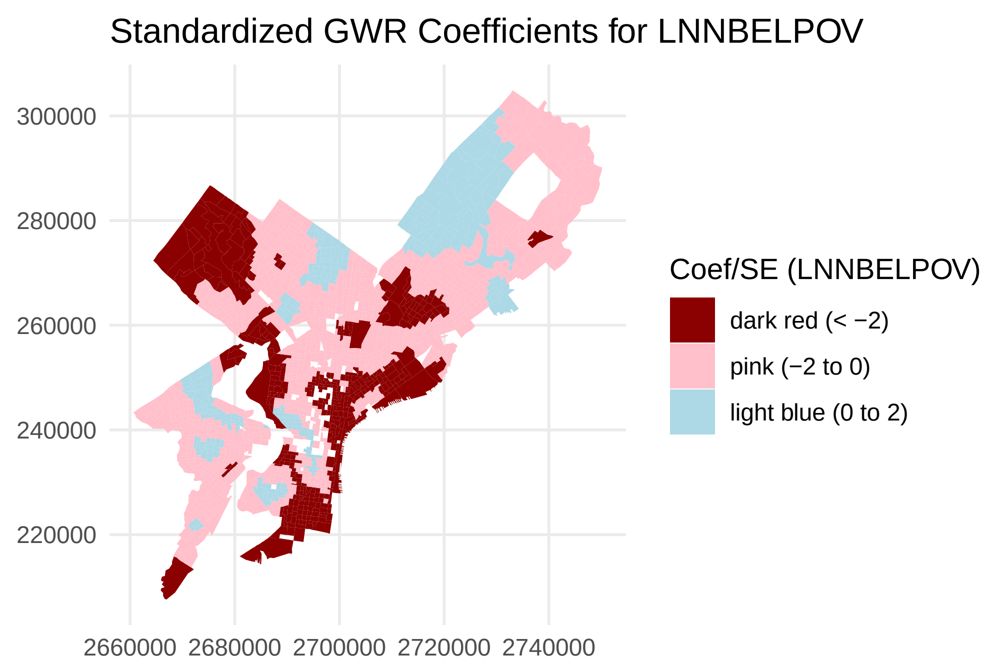
The Moran's I value for the GWR residuals (0.033, p = 0.016) indicates that most of the spatial autocorrelation present in the OLS residuals (Moran's I   0.31, p < 0.001) has been substantially reduced. This suggests that the GWR model effectively accounts for spatial dependence by allowing regression coefficients to vary across space.

###Mapping of Local Coefficients


Standardized GWR Coefficients for PCTVACANT
300000
280000
260000
240000
220000
2660000 2680000 2700000 2720000 2740000
Coef/SE (PCTVACANT)
dark red (< −2)
pink (−2 to 0)
light blue (0 to 2)
dark blue (> 2)


Standardized GWR Coefficients for PCTSINGLES
300000
280000
260000
240000
220000
2660000 2680000 2700000 2720000 2740000
Coef/SE (PCTSINGLES)
dark red (< −2)
pink (−2 to 0)
light blue (0 to 2)
dark blue (> 2)





The spatial patterns in the local coefficients demonstrate clear heterogeneity across Philadelphia. For vacancy rate (PCTVACANT), the strongest negative effects on housing values are found in North and Southwest Philadelphia, indicating that higher vacancy rates are particularly detrimental to property values in areas with already weak housing markets. In contrast, the proportion of single-family homes (PCTSINGLES) shows a mixed pattern: positive effects in the northeastern and northwestern suburbs, where detached homes are more desirable, but weak or negative associations in older central neighborhoods dominated by rowhouses. The coefficient for educational attainment (PCTBACHMOR) is positive across nearly all block groups, with the strongest effects concentrated in University City, Center City, and Northwest Philadelphia, confirming the robust link between human capital and housing value. Finally, poverty (LNNBELPOV) exhibits a negative effect in North and South Philadelphia, showing the spatial concentration of disadvantage and its inverse relationship with property values. Together, these patterns highlight how the relationships between neighborhood characteristics and housing values are spatially nonstationary, emphasizing the importance of local context in housing market analysis.

4 Discussion (Google Doc at end)