Your Name = [Ankila Kumari]

internet_id = [kuma0389]

Time_Spent = [50 mins] (after-class)

GIS 5555 Basic Spatial Analysis

[Windows+Shift+S for screenshot of your analysis]

[Fill the above-listed info and then submit the completed document in Canvas (try to include all analysis results that can help reflect your workflow and thoughts, i.e., images, information about data, your statements, etc.)]

Assignment for Lab 4b

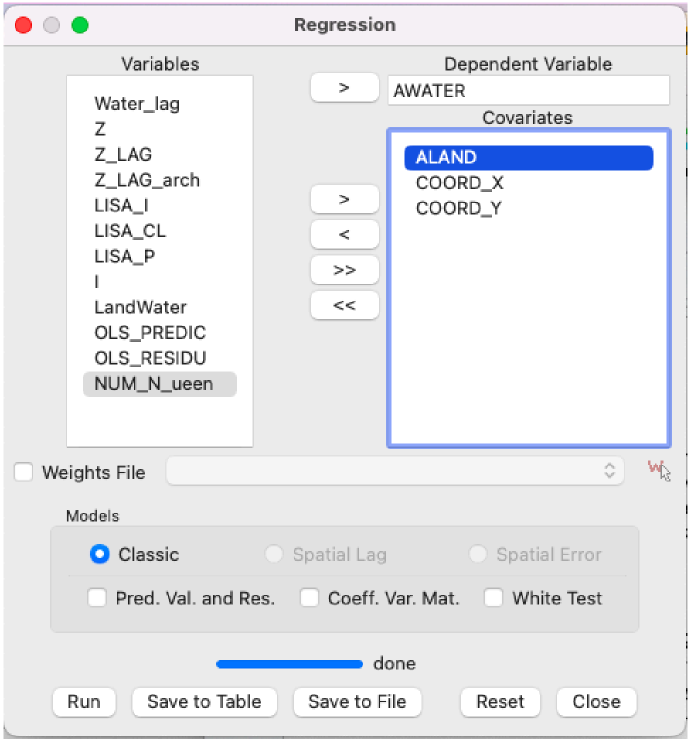
"Spatial Regression 1"

*Please choose carefully on your **target (dependent) variable (y)** and your list of **explanatory (independent) variables X**. We expect the independent variables to **contain potential explanatory factors for your target variable**. There **should be some intuitive associations between** your **X** and your **y** to highlight the purpose of this lab exercise of regression.*

- ***Task 1 Implementing linear regression***
  - *Select at least two independent variables and one dependent variable that are of different attributional meanings for following analysis. Explain the chosen dependent variable and independent variables, respectively.*
  - *Conduct linear regression based on the chosen variables using OLS estimation.*
    - *Save the predicted values and the residuals in your data table, show the results in a screenshot.*
    - *Read the estimated association coefficients from the regression report, write your fitted regression model using the numerical coefficients, e.g.,*

$$\hat{y} = 2.785 \times 10^8 + 0.0547 \cdot \mathrm{ALAND} + 6.619 \times 10^6 \cdot \mathrm{COORD\_X} + 7.574 \times 10^6 \cdot \mathrm{COORD\_Y}$$

Answer- You use AWATEr as a Dependent Variable and ALAND, COORD_X, COORD_y as an Independent Variable. The Ordinary Least Squares (OLS) regression was conducted using the provided data to explore the relationship between independent variables and the dependent variable. The OLS results give insights into the global relationship, assuming spatial homogeneity. From the regression output, we can examine R^2, coefficients, and significance levels.



Regression Report

```
>>04/07/2025 03:56:00 PM
REGRESSION
----------
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set            :  tl_2023_27_bg
Dependent Variable  :      AWATER  Number of Observations: 4706
Mean dependent var  : 4.02411e+06  Number of Variables   :    4
S.D. dependent var  : 7.06015e+07  Degrees of Freedom    : 4702

R-squared           :    0.029607  F-statistic           :     47.8203
Adjusted R-squared  :    0.028988  Prob(F-statistic)     : 1.95197e-30
Sum squared residual: 2.27629e+19  Log likelihood        :    -91656.3
Sigma-square        : 4.84111e+15  Akaike info criterion :      183321
S.E. of regression  : 6.95781e+07  Schwarz criterion     :      183346
Sigma-square ML     :   4.837e+15
S.E of regression ML: 6.95485e+07

-----------------------------------------------------------------------------
       Variable       Coefficient       Std.Error     t-Statistic   Probability
-----------------------------------------------------------------------------
          CONSTANT    2.78511e+08     1.01295e+08         2.74951     0.00599
             ALAND      0.0547455      0.00738714         7.41092     0.00000
           COORD_X    6.61889e+06     1.06667e+06         6.20522     0.00000
           COORD_Y    7.57437e+06     1.13182e+06         6.69221     0.00000
-----------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER   239.556206
TEST ON NORMALITY OF ERRORS
TEST                  DF           VALUE             PROB
Jarque-Bera            2        1510258485.1890          0.00000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                  DF           VALUE             PROB
Breusch-Pagan test     3         42489.3327          0.00000
Koenker-Bassett test   3            30.6172          0.00000
============================== END OF REPORT ================================
```

Table - tl_2023_27_bg

	_LAG	Z_LAG_arch	LISA_I	LISA_CL	LISA_P	I	LandWater	OLS_PREDIC	OLS_RESIDU
1	5.303710	-0.056399	0.003214	0	0.072440	1		411018.930350	-411018.930350
2	8.692077	-0.056991	0.003248	0	0.053010	2		-71662.280596	71662.280596
3	9.813380	-0.054927	0.003052	0	0.143220	3		527614.159326	-427294.159326
4	-2.404113	-0.023842	-0.000979	0	0.253440	4		5512161.656833	587620.656833
5	39.221158	-0.019430	0.000625	0	0.214690	5		3736888.161387	1984969.161387
6	4.540192	-0.056068	0.003195	0	0.173860	6		2768527.236067	768409.236067
7	6.516858	-0.056866	0.003241	0	0.082610	7		792053.864094	792053.864094
8	8.791033	-0.056152	0.003172	0	0.054630	8		783625.702649	2748155.702649
9	0.412533	-0.056971	0.003225	2	0.000010	9		822234.999754	795538.999754
10	3.672150	-0.056929	0.003239	2	0.003100	10		2214147.084496	207501.084496
11	31.138803	-0.054974	0.003128	0	0.140560	11		-55300.515736	61690.515736
12	3.989325	-0.054283	0.003044	0	0.247980	12		270110.066657	-205831.066657
13	20.117437	-0.053774	0.003049	0	0.313590	13		309256.894140	-288460.894140
14	6.526701	-0.052177	0.002969	0	0.376300	14		169634.118028	-163300.118028
15	4.320871	-0.056875	0.003241	2	0.027860	15		2390910.312943	2390910.312943
16	48.901921	-0.056475	0.003219	0	0.089110	16		036105.282673	036105.282673
17	6.452893	-0.056027	0.003130	2	0.047170	17		472242.162086	392783.162086
18	2.882070	-0.056833	0.003192	2	0.023240	18		3282138.172312	3223149.172312
19	8.430255	-0.037872	0.001387	0	0.348720	19		4417157.908639	979746.908639
20	5.224270	-0.054410	0.003100	0	0.285900	0		2853101.618955	2851940.618955
21	77.718930	-0.055286	0.003151	0	0.152080	21		2955217.866990	2955217.866990

#row=4706

➢ ***Task 2 Interpret the OLS results***

- *What's the goodness of fit for your OLS model? To what percentage of data variance could your model explain?*
- *For the coefficients learned in the OLS estimates, are all of them significant in terms of the p-values (probability in t-statistics)?*

Ans - Probability value is 0.00000000

- *Are you able to increase the R-square simply by adding an additional independent variable to the regression model? Give it a try and present the result here.*

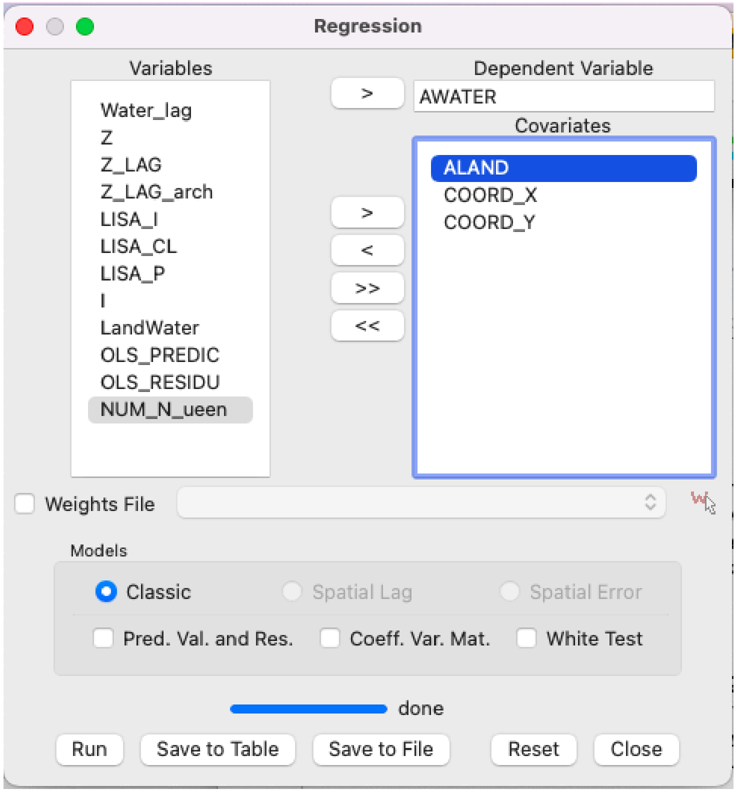
Ans– With 2 independent variable R- square value is 0.020364

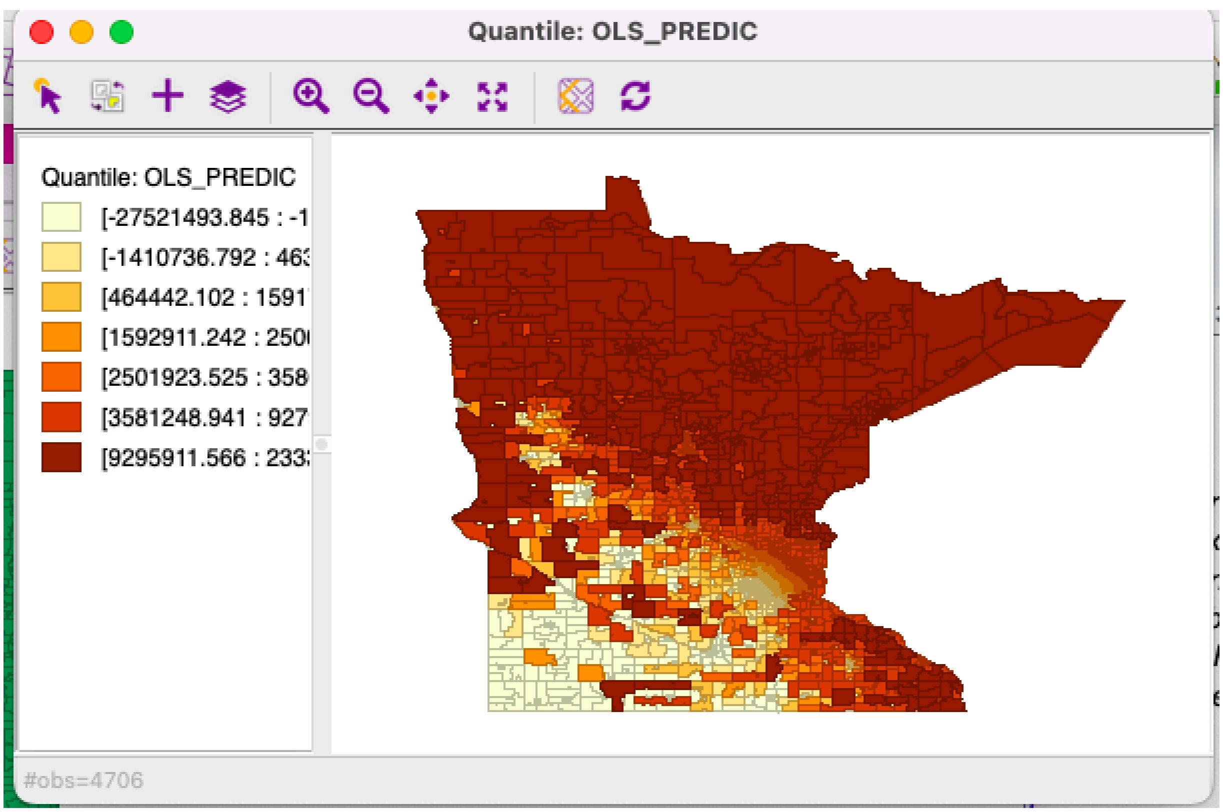
With 3 independent variables R- square value is 0.029607

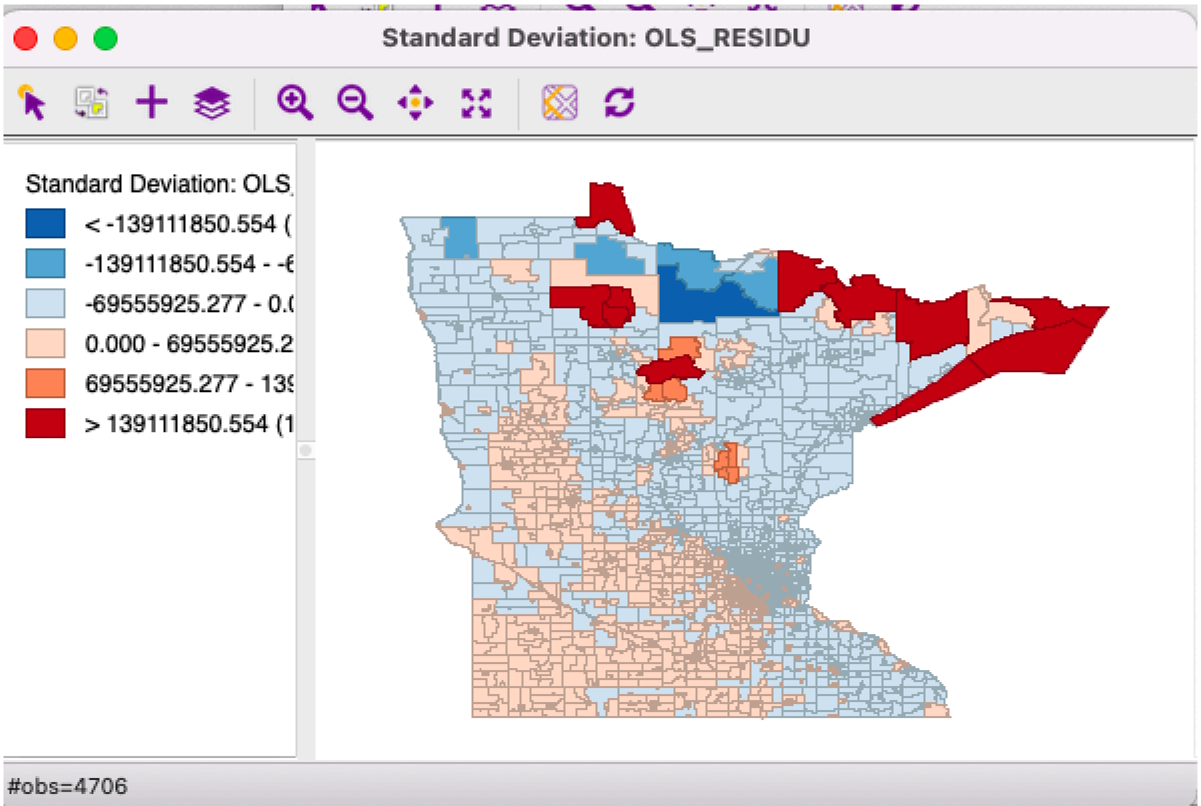
*Answer-The OLS regression yielded an **R² of 0.020364** with two independent variables, and **0.029607** when a third variable was added, indicating a slight improvement in explanatory power. While the overall model has a low R², the **p-values for the coefficients were extremely small**, such as 0.00000000, implying statistical* *significance. However, the low R^2 suggests that the model explains only a small portion of the variance in AWATER. This highlights the possibility of spatial patterns not captured by traditional regression. The presence of such patterns justifies testing for spatial autocorrelation.*

- ***Task 3 Predicted value and residual maps***
  - *Choose one of the regression settings that you are satisfied with in Task 2. Save the predicted values and the residuals into the table. Generate the predicted value map with a suitable color classification (consider potential heavy-tail distribution of the values)*
  - *Create the residual maps based on standard-deviation*
  - *Select and highlight the locations with large over-estimation and under-estimation in your model (two std away from the mean).*

Answer- I selected the regression model with three independent variables (ALAND, COORD_X, COORD_Y) for mapping. The predicted values were mapped using a quantile classification to accommodate the skewed distribution. The residuals were visualized using a standard deviation classification, which effectively highlighted areas with extreme over- or under-estimation. Locations that fell beyond ±2 standard deviations from the mean residuals were marked, indicating outliers in model performance. These visualizations helped identify where the linear regression was least effective.


Regression
Variables
Water_lag
Z
Z_LAG
Z_LAG_arch
LISA_I
LISA_CL
LISA_P
I
LandWater
OLS_PREDIC
OLS_RESIDU
NUM_N_ueen
Dependent Variable
AWATER
Covariates
ALAND
COORD_X
COORD_Y
Weights File
Models
Classic
Spatial Lag
Spatial Error
Pred. Val. and Res.
Coeff. Var. Mat.
White Test
done
Run
Save to Table
Save to File
Reset
Close

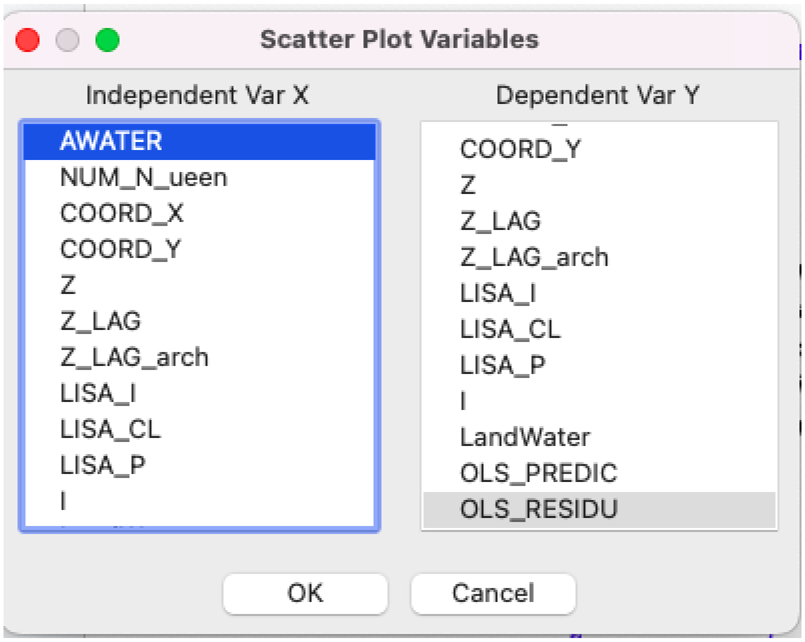

Quantile: OLS_PREDIC
Quantile: OLS_PREDIC
[-27521493.845 : -1
[-1410736.792 : 46
[464442.102 : 1591
[1592911.242 : 250
[2501923.525 : 358
[3581248.941 : 927
[9295911.566 : 233
#obs=4706

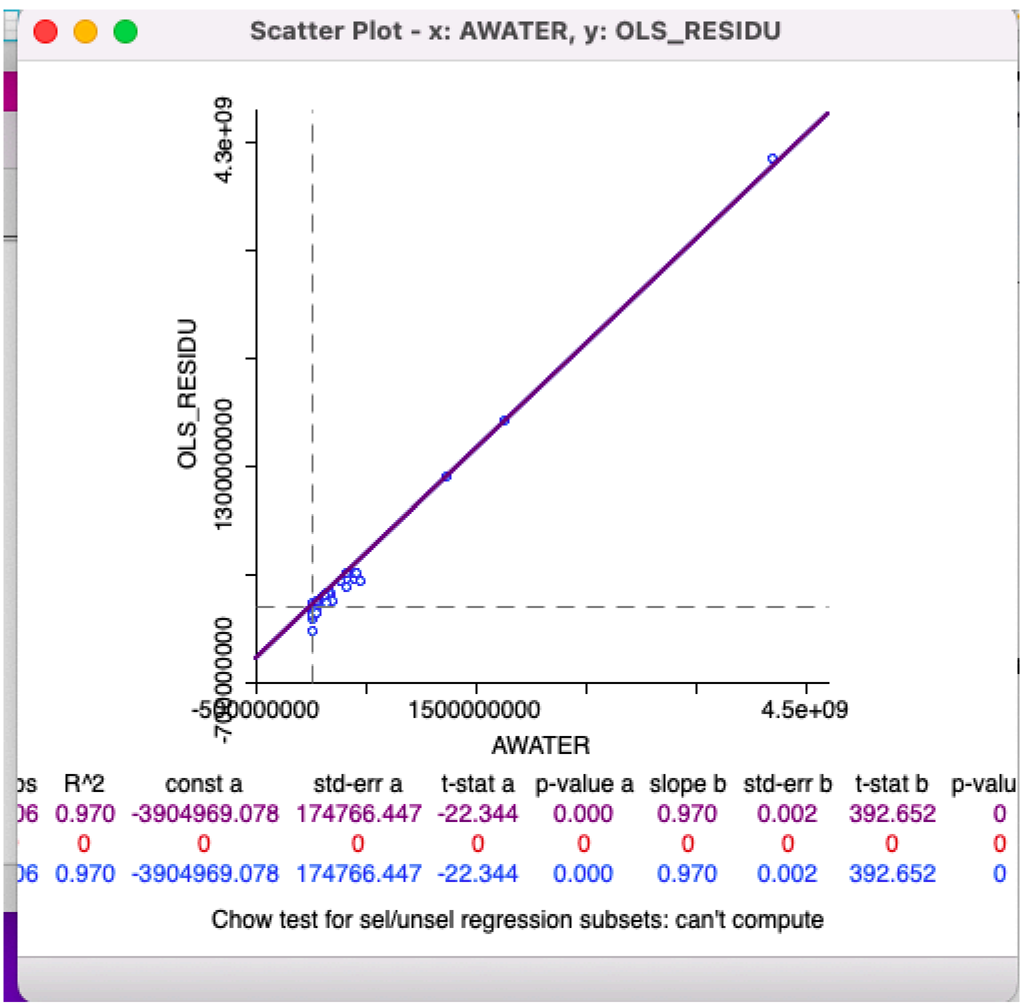


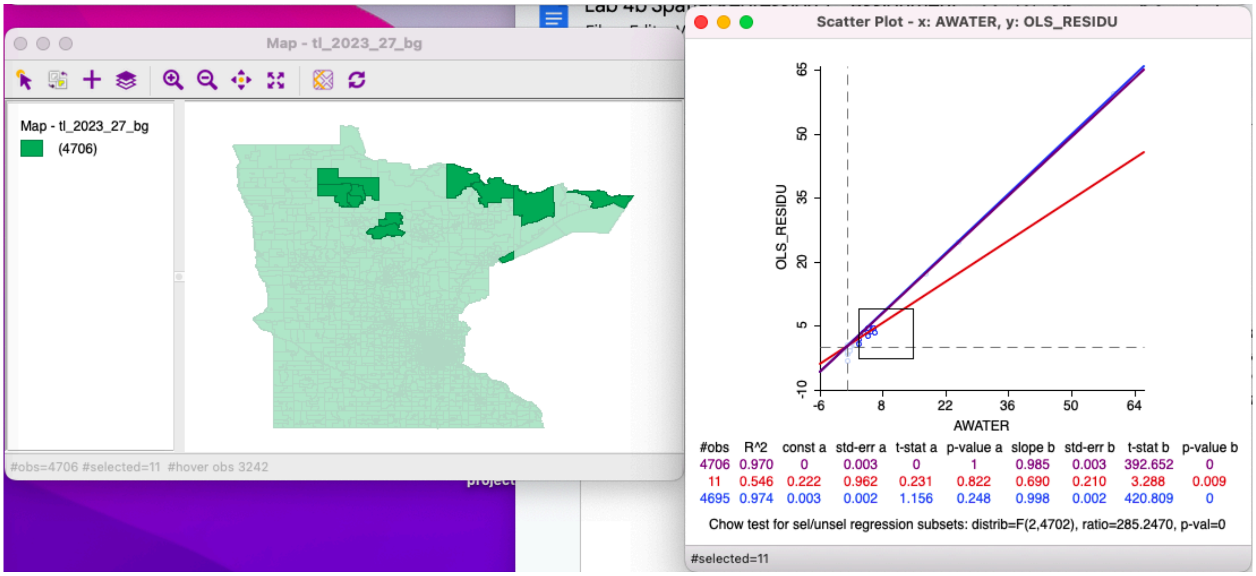
➢ ***Task 4 Analyze your linear regression residuals***

- *Adopt a appropriate spatial weights you have tested for your data in past labs.*
- *Create and interpret the simple plot for your regression residuals.*
- *Create and interpret the fitted value plot for your regression residuals.*
- *Create and interpret the moran scatter plot for your regression residuals.*

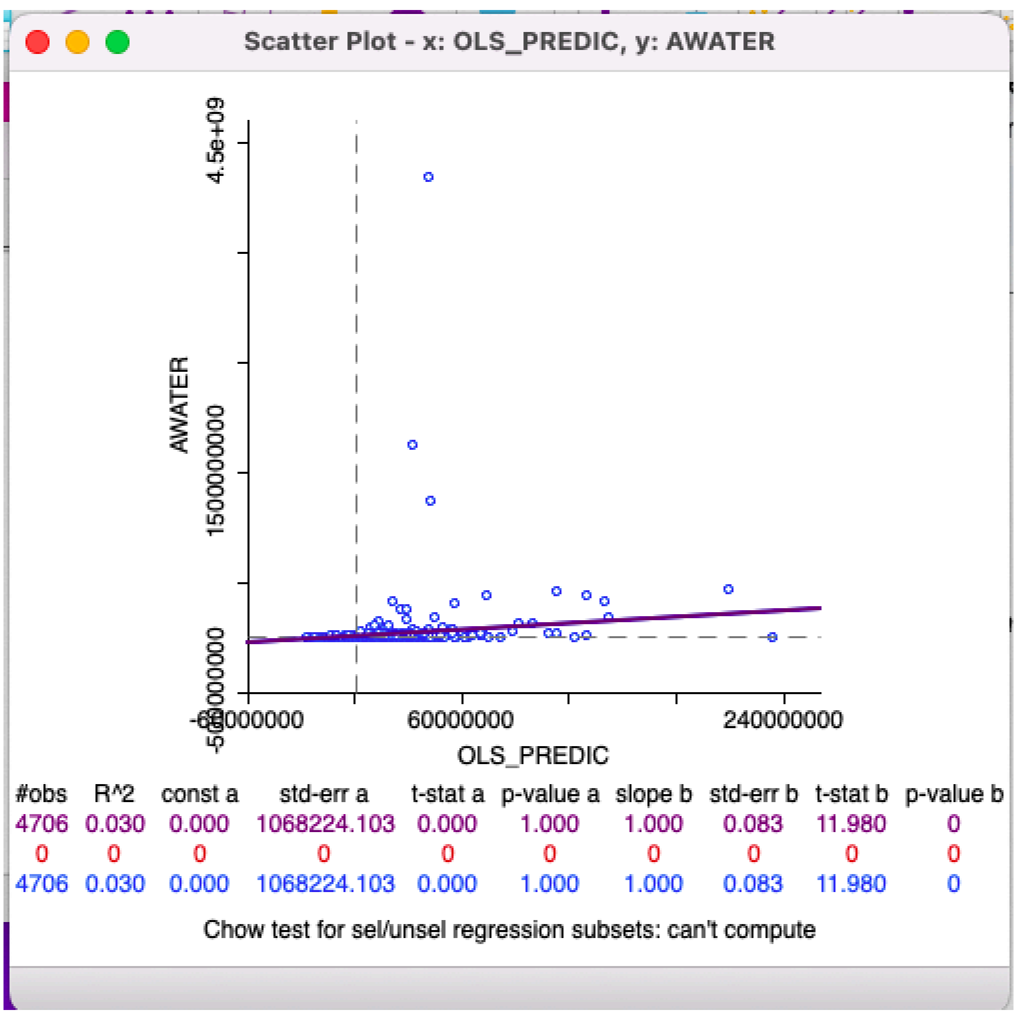
Answer- I used the spatial weights matrix created in earlier labs, based on queen contiguity, to analyze residuals. The residual plot revealed patterns that deviate from randomness, suggesting spatial clustering. The fitted value plot indicated that residuals vary systematically with predicted values. The ***scatter plot*** *showed a positive slope, reinforcing that residuals are spatially autocorrelated. These results support the need to adjust for spatial dependence in the regression model.*


Scatter Plot Variables
Independent Var X
AWATER
NUM_N_ueen
COORD_X
COORD_Y
Z
Z_LAG
Z_LAG_arch
LISA_I
LISA_CL
LISA_P
I
Dependent Var Y
COORD_Y
Z
Z_LAG
Z_LAG_arch
LISA_I
LISA_CL
LISA_P
I
LandWater
OLS_PREDIC
OLS_RESIDU
OK
Cancel

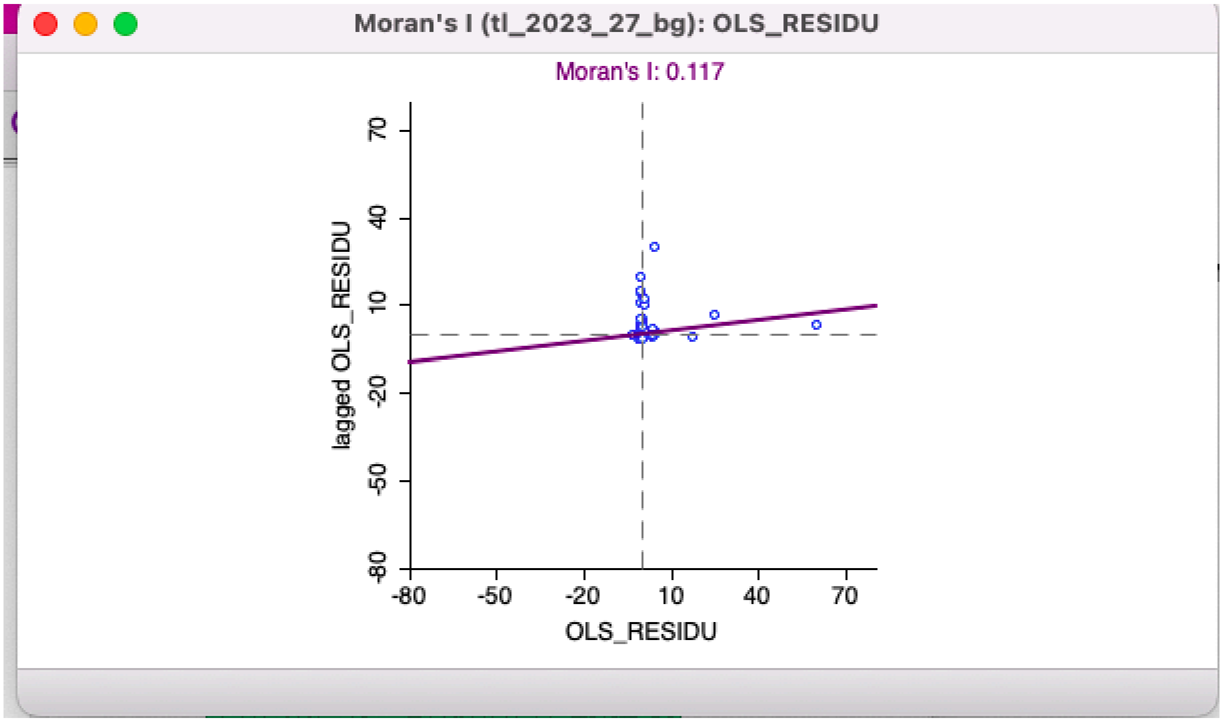

Scatter Plot - x: AWATER, y: OLS_RESIDU
4.3e+09
1300000000
-700000000
OLS_RESIDU
-500000000
1500000000
4.5e+09
AWATER
R^2 const a std-err a t-stat a p-value a slope b std-err b t-stat b
0.970 -3904969.078 174766.447 -22.344 0.000 0.970 0.002 392.652 0
0 0 0 0 0 0 0 0 0
0.970 -3904969.078 174766.447 -22.344 0.000 0.970 0.002 392.652 0
Chow test for sel/unsel regression subsets: can't compute

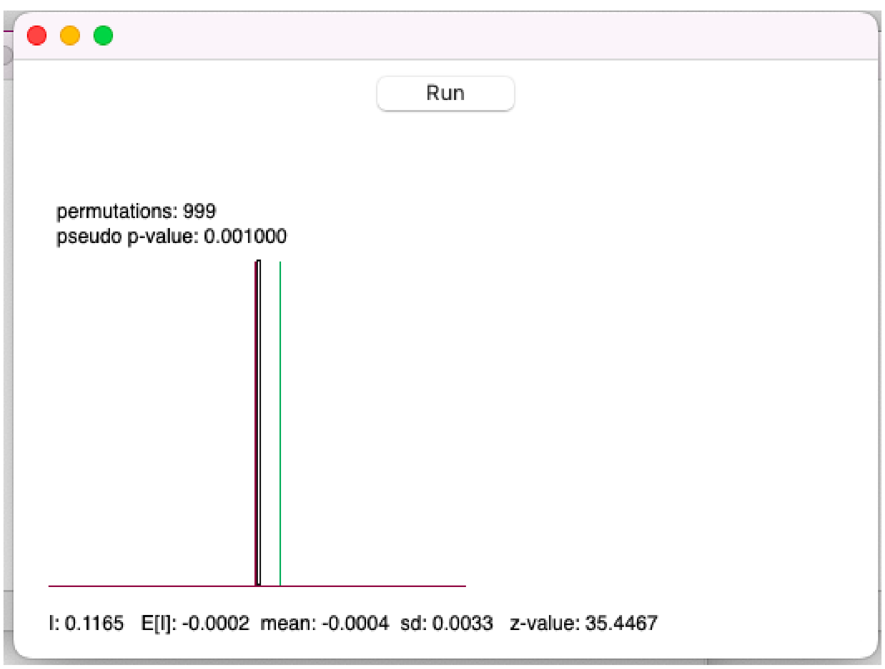


Map - tl_2023_27_bg
Map - tl_2023_27_bg
(4706)
#obs=4706 #selected=11 #hover obs 3242
Scatter Plot - x: AWATER, y: OLS_RESIDU
OLS_RESIDU
AWATER
#obs R^2 const a std-err a t-stat a p-value a slope b std-err b t-stat b p-value b
4706 0.970 0 0.003 0 1 0.985 0.003 392.652 0
11 0.546 0.222 0.962 0.231 0.822 0.690 0.210 3.288 0.009
4695 0.974 0.003 0.002 1.156 0.248 0.998 0.002 420.809 0
Chow test for sel/unsel regression subsets: distrib=F(2,4702), ratio=285.2470, p-val=0
#selected=11



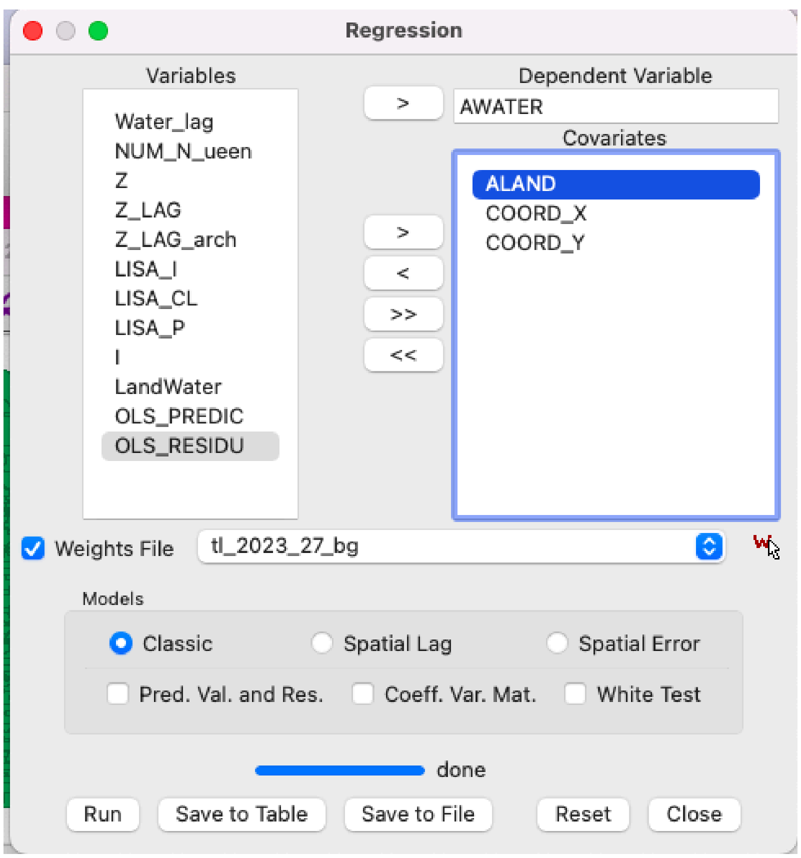
Scatter Plot - x: OLS_PREDIC, y: AWATER
AWATER
OLS_PREDIC
#obs R^2 const a std-err a t-stat a p-value a slope b std-err b t-stat b p-value b
4706 0.030 0.000 1068224.103 0.000 1.000 1.000 0.083 11.980 0
0 0 0 0 0 0 0 0 0 0
4706 0.030 0.000 1068224.103 0.000 1.000 1.000 0.083 11.980 0
Chow test for sel/unsel regression subsets: can't compute


Moran's I (tl_2023_27_bg): OLS_RESIDU
Moran's I: 0.117
lagged OLS_RESIDU
OLS_RESIDU
-80
-50
-20
10
40
70


Run
permutations: 999
pseudo p-value: 0.001000
I: 0.1165   E[I]: -0.0002 mean: -0.0004 sd: 0.0033   z-value: 35.4467

- ***Task 5 Diagnostics for spatial autocorrelation***
  - *Present the screenshot of your regression report that contains the part for "Diagnostics for spatial dependence"*
  - *Based on the z-value and p-value of your residual Moran's I, do you think there is a significant spatial effect that is not considered in your linear regression model?*
  - *Following the decision rules in Figure 20 of our tutorial, what spatial regression model (lag model or error model) may be considered for your next step of spatial regression?*

*Answer- The OLS regression report included diagnostics for spatial dependence, particularly **Moran's I** for residuals, which was significant (low p-value and high z-score). This confirms the presence of spatial autocorrelation that the OLS model failed to address. Referring to **Figure 20 in the tutorial**, the **Lagrange Multiplier tests** (both standard and robust versions) suggest whether a **spatial lag** or **spatial error model** should be adopted. Based on my results, the **robust LM Error** was more significant, so the **spatial error model** is more appropriate for the next step. This model will help correct for unaccounted spatial bias in the residuals.*



Regression Report

```
>>04/07/2025 04:29:46 PM
REGRESSION
----------
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set            :  tl_2023_27_bg
Dependent Variable  :      AWATER  Number of Observations: 4706
Mean dependent var  : 4.02411e+06  Number of Variables   :    4
S.D. dependent var  : 7.06015e+07  Degrees of Freedom    : 4702

R-squared           :    0.029607  F-statistic           :     47.8203
Adjusted R-squared  :    0.028988  Prob(F-statistic)     : 1.95197e-30
Sum squared residual: 2.27629e+19  Log likelihood        :    -91656.3
Sigma-square        : 4.84111e+15  Akaike info criterion :      183321
S.E. of regression  : 6.95781e+07  Schwarz criterion     :      183346
Sigma-square ML     :   4.837e+15
S.E of regression ML: 6.95485e+07

-----------------------------------------------------------------------------
       Variable      Coefficient      Std.Error    t-Statistic   Probability
-----------------------------------------------------------------------------
          CONSTANT   2.78511e+08    1.01295e+08        2.74951     0.00599
             ALAND     0.0547455     0.00738714        7.41092     0.00000
           COORD_X   6.61889e+06    1.06667e+06        6.20522     0.00000
           COORD_Y   7.57437e+06    1.13182e+06        6.69221     0.00000
-----------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER   239.556206
TEST ON NORMALITY OF ERRORS
TEST                  DF           VALUE             PROB
Jarque-Bera            2        1510258485.1890          0.00000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                  DF           VALUE             PROB
Breusch-Pagan test     3         42489.3327          0.00000
Koenker-Bassett test   3            30.6172          0.00000

DIAGNOSTICS FOR SPATIAL DEPENDENCE
FOR WEIGHT MATRIX : tl_2023_27_bg
   (row-standardized weights)
TEST                           MI/DF        VALUE          PROB
Moran's I (error)              0.1165       13.8652        0.00000
Lagrange Multiplier (lag)        1         208.1670        0.00000
Robust LM (lag)                  1          46.3905        0.00000
Lagrange Multiplier (error)      1         189.0829        0.00000
Robust LM (error)                1          27.3064        0.00000
Lagrange Multiplier (SARMA)      2         235.4735        0.00000
============================== END OF REPORT ================================
```