Summary

Explore language-education-employement correlation

Objective: Understanding the relationship between the language spoken(english limit), education attatinment, and employment.

Method

Develop OLS and Spatial Weight Regression to investigate correlation among with non-english languages, educations levels, and employment rate.

- OLS is a global regression technique where a single set of model coefficients is estimated for the entire dataset.
-
- SWLS explicitly incorporates spatial dependencies into the modeling process by assigning weights to observations based on their spatial proximity. This allows SWLS to account for spatial autocorrelation and produce more accurate parameter estimates.

Result

Educational attainment and Limted english speaker

- Dependent variable is" Limited English speaker"
- - S1601_C05_022E: Speak English less than very well"!!Percent of specified language speakers!!Population 5 years and over!!Speak a language other than English"
- Independent variables are "Educational attainment"
- - S2301_C01_035E : Population 25 to 64 years!!Bachelor's degree or higher
- - S2301_C01_034E : Population 25 to 64 years!!Some college or associate's degree
- - S2301_C01_033E :Population 25 to 64 years!!High school graduate (includes equivalency)
- - S2301_C01_032E :Population 25 to 64 years!!Less than high school graduate

OLS Model Interpretation

```
REGRESSION RESULTS
------------------

SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES
-----------------------------------------
Data set            :     unknown
Weights matrix      :        None
Dependent Variable  :S1601_C05_022E                Number of Observations:        2495
Mean dependent var  :    414.7006               Number of Variables   :           5
S.D. dependent var  :    287.6663               Degrees of Freedom    :        2490
R-squared           :      0.3601
Adjusted R-squared  :      0.3590
Sum squared residual: 1.32075e+08               F-statistic           :    350.2332
Sigma-square        :   53042.096               Prob(F-statistic)     :  2.013e-239
S.E. of regression  :     230.309               Log likelihood        :  -17109.103
Sigma-square ML     :   52935.799               Akaike info criterion :   34228.207
S.E of regression ML:    230.0778               Schwarz criterion     :   34257.317

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     t-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT        62.22017        13.31909         4.67150         0.00000
      S2301_C01_035E         0.07605         0.00980         7.76071         0.00000
      S2301_C01_034E        -0.00527         0.02175        -0.24228         0.80858
      S2301_C01_033E         0.45511         0.02643        17.22095         0.00000
      S2301_C01_032E         0.22607         0.01836        12.31021         0.00000
------------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER           7.950

TEST ON NORMALITY OF ERRORS
TEST                             DF        VALUE           PROB
Jarque-Bera                       2        2079.820           0.0000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                             DF        VALUE           PROB
Breusch-Pagan test                4         353.079           0.0000
Koenker-Bassett test              4         122.994           0.0000
================================ END OF REPORT =====================================
```

The OLS model defines the correlation between the variables. In the OLS model, the R-square is 0.36, representing the English-limited speaker explains 36% of the educational attainment.
It suggests a moderate relationship between the English-limited speaker and educational attainment.

	Coeff.	Std. Error	P-Value
CONSTANT	62.220	13.319	0.000
S2301_C01_035E	0.076	0.010	0.000
S2301_C01_034E	-0.005	0.022	0.809
S2301_C01_033E	0.455	0.026	0.000
S2301_C01_032E	0.226	0.018	0.000

- The coefficient of 0.076(Bachelor's degree or higher) represents the estimated change in the predicted value of the limited English speaker for a one-unit change in the independent variable (Bachelor's degree or higher).
  This coefficient indicates a slightly positive relationship between limited English speakers and educational attainment.
- The coefficient of -0.00527 (Some college or associate's degree) represents the estimated change in the predicted value of the limited English speaker for a one-unit change in the independent variable (Some college or associate's degree).
  Since this coefficient is **Negative**, it indicates that as educational attainment(Some college or associate's degree) **decreases**, the likelihood of being a limited English speaker also **decreases**.
  Moreover, the p-value of education attainment (Some college or associate's degree) is larger than 0.05, which means there is a non-significant relationship between limited English speakers and educational attainment (Some college or associate's degrees).

- The coefficient of 0.455(High school graduate) represents the estimated change in the predicted value of the limited english speaker for a one-unit change in the independent variable (High school graduate). This is the highest coefficient, which is **the strongest relationship** between High school graduated and limited english speakers.
- The coefficient of 0.226(Less than high school graduate) represents the estimated change in the predicted value of the limited english speaker for a one-unit change in the independent variable (Less thanHigh school graduate).

SWLS Model Interpretation

```
print(m5.summary)

REGRESSION RESULTS
------------------

SUMMARY OF OUTPUT: GM SPATIALLY WEIGHTED LEAST SQUARES (HET)
-----------------------------------------------------------------
Data set            :     unknown
Weights matrix      :     unknown
Dependent Variable  :S1601_C05_022E                Number of Observations:        2495
Mean dependent var  :    414.7006              Number of Variables   :           5
S.D. dependent var  :    287.6663              Degrees of Freedom    :        2490
Pseudo R-squared    :      0.3474
N. of iterations    :           1              Step1c computed       :          No

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT        65.82114        16.47819         3.99444         0.00006
      S2301_C01_035E         0.07910         0.01267         6.24261         0.00000
      S2301_C01_034E         0.09464         0.02098         4.51047         0.00001
      S2301_C01_033E         0.27360         0.02718        10.06530         0.00000
      S2301_C01_032E         0.26098         0.02573        10.14182         0.00000
              lambda         0.77106         0.01575        48.94359         0.00000
------------------------------------------------------------------------------------
================================ END OF REPORT =====================================
```

A Pseudo R-squared value of 0.34 suggests that approximately 34% of the variability in the "Limited English Speaker" variable is explained by the variation in the "Educational Attainment" variables.
It represents a moderate relationship between limited english and education attainment.

	Coeff.	Std. Error	P-Value
CONSTANT	65.821	16.478	0.0
S2301_C01_035E	0.079	0.013	0.0
S2301_C01_034E	0.095	0.021	0.0
S2301_C01_033E	0.274	0.027	0.0
S2301_C01_032E	0.261	0.026	0.0
lambda	0.771	0.016	0.0

- The lowest coeficient of (S2301_C01_035E) Bachelor's degree or higher is 0.079, which is **the weakest relationship** between English speaker level and education attatinment.
- The highest coeficient of(S2301_C01_032E) Less than high school graduate is 0.261, which is **the strongest relationship** between English speaker level and education attatinment.

limitation: the output results are different between the SWLS model and the OLS model. It needs to optimize the data or method for validating the result.

In this report, I use SWLS and OLS models to investigate the relationship between education attainment and limited-English speakers.

Throughout those two models, they indicate that the predictor (education attainment) in the model is **moderately** successful in explaining the variance in the dependent variable(limited English speaker).