Excusive Summary

Explore language-education-employement correlation

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

Method

Develop OLS and Spatial Weight Regression to investigate correlation among with non-english languages, educations levels, povety, 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 interpretation

Poverty and English speaker level

speaking english less than very well

- Dependent variable is 'below the poverty level'
  - **2301_C01_028E**: Population 20 to 64 years!!POVERTY STATUS IN THE PAST 12 MONTHS!!Below poverty level
- Independent variables are "speaking english less than very well"
  - **S1601_C05_006E**: Speak English less than very well"!!SPEAK A LANGUAGE OTHER THAN ENGLISH!!Spanish!!18 to 64 years old"

- **S1601_C05_010E**: Speak English less than very well"!!SPEAK A LANGUAGE OTHER THAN ENGLISH!!Other Indo-European languages!!18 to 64 years old"
- **S1601_C05_014E**: Speak English less than very well"!!SPEAK A LANGUAGE OTHER THAN ENGLISH!!Asian and Pacific Island languages!!18 to 64 years old"
- **S1601_C05_018E**: Speak English less than very well"!!SPEAK A LANGUAGE OTHER THAN ENGLISH!!**Other languages**!!18 to 64 years old"

OLS Model Interpretation

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

SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES
-----------------------------------------
Data set            :     unknown
Weights matrix      :        None
Dependent Variable  :S2301_C01_028E               Number of Observations:        2495
Mean dependent var  :    294.9126               Number of Variables   :           5
S.D. dependent var  :    213.1207               Degrees of Freedom    :        2490
R-squared           :      0.3087
Adjusted R-squared  :      0.3076
Sum squared residual: 7.83047e+07               F-statistic           :    278.0318
Sigma-square        :   31447.688               Prob(F-statistic)     :  8.671e-198
S.E. of regression  :     177.335               Log likelihood        :  -16456.960
Sigma-square ML     :   31384.666               Akaike info criterion :   32923.920
S.E of regression ML:    177.1572               Schwarz criterion     :   32953.030

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     t-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT       144.53110         6.13227        23.56893         0.00000
      S1601_C05_006E         0.27717         0.00849        32.63952         0.00000
      S1601_C05_010E         0.28933         0.03586         8.06843         0.00000
      S1601_C05_014E         0.10867         0.01670         6.50857         0.00000
      S1601_C05_018E         0.94153         0.16101         5.84777         0.00000
------------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER            3.300

TEST ON NORMALITY OF ERRORS
TEST                             DF        VALUE           PROB
Jarque-Bera                       2       48138.820           0.0000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                             DF        VALUE           PROB
Breusch-Pagan test                4          14.388           0.0062
Koenker-Bassett test              4           1.262           0.8678
================================= END OF REPORT =====================================
```

The OLS model defines the correlation between the variables. In the OLS model, the R-square is 0.3087, representing the poverty status explains 30.87% of the limited English speakers.
It suggests a moderate relationship between poverty status and limited English speaker.

Quick summary

	Coeff.	Std. Error	P-Value
CONSTANT	144.531	6.132	0.0
S1601_C05_006E	0.277	0.008	0.0
S1601_C05_010E	0.289	0.036	0.0
S1601_C05_014E	0.109	0.017	0.0
S1601_C05_018E	0.942	0.161	0.0

- The highest coeficient of Other languages who speak English less than very well is 0.942, which is **the strongest relationship** between Poverty and English speaker level.
  The coefficient of 0.942 represents the estimated change in the dependent variable (below poverty level) associated with a one-unit increase in the independent variable (proportion of people who speak other languages and have limited English proficiency), holding all other variables constant.

- The lowest coeficient of Asian and Pacific Island languages who speak English less than very well is 0.109, which is **the weakest relationship** between Poverty and English speaker level.
  The coefficient of 0.109 represents the estimated change in the dependent variable (below poverty level) associated with a one-unit increase in the independent variable (proportion of people who speak Asian and Pacific Island languages and have limited English proficiency), holding all other variables constant.

SWLS Model Interpretation

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

SUMMARY OF OUTPUT: GM SPATIALLY WEIGHTED LEAST SQUARES (HET)
-----------------------------------------------------------
Data set            :     unknown
Weights matrix      :     unknown
Dependent Variable  :S2301_C01_028E                Number of Observations:        2495
Mean dependent var  :    294.9126                Number of Variables   :           5
S.D. dependent var  :    213.1207                Degrees of Freedom    :        2490
Pseudo R-squared    :      0.3035
N. of iterations    :           1                Step1c computed       :          No

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT       122.97275         8.38457        14.66656         0.00000
      S1601_C05_006E         0.28960         0.01153        25.11865         0.00000
      S1601_C05_010E         0.36966         0.05410         6.83352         0.00000
      S1601_C05_014E         0.18290         0.02062         8.86939         0.00000
      S1601_C05_018E         0.83236         0.16987         4.90012         0.00000
              lambda         0.53849         0.03991        13.49402         0.00000
------------------------------------------------------------------------------------
================================ END OF REPORT =====================================
```

A Pseudo R-squared value of 0.3035 suggests that approximately 30.35% of the variability in the dependent variable("Below poverty level") is explained by the variation in the independent variables (Limited English Speaker).
It suggests a moderate relationship between poverty status and limited English speakers.

Quick summary

	Coeff.	Std. Error	P-Value
CONSTANT	122.973	8.385	0.0
S1601_C05_006E	0.290	0.012	0.0
S1601_C05_010E	0.370	0.054	0.0
S1601_C05_014E	0.183	0.021	0.0
S1601_C05_018E	0.832	0.170	0.0
lambda	0.538	0.040	0.0

- The highest coeficient of Other languages who speak english less than very well is 0.832, which is **the strongest relationship** between Poverty and English speaker level.
  The coefficient of 0.832 represents the estimated change in the dependent variable (below poverty level) associated with a one-unit increase in the independent variable (proportion of people who speak other languages and have limited English proficiency), holding all other variables constant.

- The lowest coeficient of Asian and Pacific Island languages who speak english less than very well is 0.109, which is **the weakest relationship** between Poverty and English speaker level.
  The coefficient of 0.109 represents the estimated change in the dependent variable (below poverty level) associated with a one-unit increase in the independent variable (proportion of people who speak Asian and Pacific Island languages and have limited English proficiency), holding all other variables constant.

Speaking english very well

- Dependent variable is 'below the poverty level'
  - **2301_C01_028E**: Population 20 to 64 years!!POVERTY STATUS IN THE PAST 12 MONTHS!!Below poverty level
- Independent variables are "speaking english very well"
  - **S1601_C03_006E** Estimate!!Speak English only or speak English very well"!!SPEAK A LANGUAGE OTHER THAN ENGLISH!!Spanish!!18 to 64 years old
  - **S1601_C03_010E** Estimate!!Speak English only or speak English very well"!!Percent of specified language speakers!!Population 5 years and over!!SPEAK A LANGUAGE OTHER THAN ENGLISH!!Other Indo-European languages"
  - **S1601_C03_014E**Estimate!!Speak English only or speak English very well"!!Asian and Pacific Island languages!!18 to 64 years old"
  - **S1601_C03_018E**Estimate!!Speak English only or speak English very well"!!SPEAK A LANGUAGE OTHER THAN ENGLISH!!Other languages!!18 to 64 years old"

OLS Model Interpretation

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

SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES
-----------------------------------------
Data set            :     unknown
Weights matrix      :        None
Dependent Variable  :S2301_C01_028E                Number of Observations:        2495
Mean dependent var  :    294.9126                Number of Variables   :           5
S.D. dependent var  :    213.1207                Degrees of Freedom    :        2490
R-squared           :      0.1513
Adjusted R-squared  :      0.1500
Sum squared residual: 9.61376e+07                F-statistic           :    110.9891
Sigma-square        :   38609.495                Prob(F-statistic)     :   3.679e-87
S.E. of regression  :     196.493                Log likelihood        :  -16712.913
Sigma-square ML     :   38532.121                Akaike info criterion :   33435.826
S.E of regression ML:    196.2960                Schwarz criterion     :   33464.936

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     t-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT       166.90791         8.75323        19.06814         0.00000
      S1601_C03_006E         0.18734         0.00906        20.66755         0.00000
      S1601_C03_010E         0.06562         0.02853         2.29953         0.02156
      S1601_C03_014E         0.01479         0.02220         0.66624         0.50532
      S1601_C03_018E         0.24928         0.09266         2.69040         0.00718
------------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER            4.716

TEST ON NORMALITY OF ERRORS
TEST                             DF        VALUE           PROB
Jarque-Bera                       2       14772.643           0.0000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                             DF        VALUE           PROB
Breusch-Pagan test                4         155.593           0.0000
Koenker-Bassett test              4          23.526           0.0001
================================ END OF REPORT =====================================
```

The OLS model defines the correlation between the variables. In the OLS model, the R-square is 0.1513, representing the poverty status explains 15.13% of the limited English speakers.
It implies a weak relationship between poverty and speaking English very well.

Quick summary

	Coeff.	Std. Error	P-Value
CONSTANT	166.908	8.753	0.000
S1601_C03_006E	0.187	0.009	0.000
S1601_C03_010E	0.066	0.029	0.022
S1601_C03_014E	0.015	0.022	0.505
S1601_C03_018E	0.249	0.093	0.007

- The lowest coefficient (Asian and Pacific Island languages who speak English very well) is 0.015, which is **the weakest relationship** between Poverty and English speaker level.
  Additionally, the p-value (Asian and Pacific Island languages who speak English very well) is larger than 0.05, which means there is a non-significant relationship between Poverty and English speaker level.
- The highest coefficient (For other languages who speak English very well) is 0.249, which is **the strongest relationship** between Poverty and English speaker level.
  The coefficient of 0.942 represents the estimated change in the dependent variable (below poverty level) associated with a one-unit increase in the independent variable (proportion of people who speak other languages and have limited English proficiency), holding all other variables constant.
- The p-value of the other language that speaks English very well is larger than 0.05, which means there is a non-significant relationship between Poverty and English speaker level.

SWLS Model Interpretation

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

SUMMARY OF OUTPUT: GM SPATIALLY WEIGHTED LEAST SQUARES (HET)
-----------------------------------------------------------
Data set            :     unknown
Weights matrix      :     unknown
Dependent Variable  :S2301_C01_028E               Number of Observations:        2495
Mean dependent var  :    294.9126               Number of Variables   :           5
S.D. dependent var  :    213.1207               Degrees of Freedom    :        2490
Pseudo R-squared    :      0.1505
N. of iterations    :           1               Step1c computed       :          No

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT       123.46512        16.37667         7.53909         0.00000
      S1601_C03_006E         0.22568         0.01387        16.26655         0.00000
      S1601_C03_010E         0.10668         0.03729         2.86075         0.00423
      S1601_C03_014E         0.04963         0.06420         0.77299         0.43953
      S1601_C03_018E         0.27236         0.08855         3.07587         0.00210
              lambda         0.64908         0.03557        18.24608         0.00000
------------------------------------------------------------------------------------
================================ END OF REPORT =====================================
```

A Pseudo R-squared value of 0.3035 suggests that approximately 30.35% of the variability in the dependent variable("Below poverty level") is explained by the variation in the independent variables (well English Speaker).

It indicates that the predictor (well English speaker) in the model is **moderately** successful in explaining the variance in the dependent variable(Below the poverty level).

	Coeff.	Std. Error	P-Value
CONSTANT	123.465	16.377	0.000
S1601_C03_006E	0.226	0.014	0.000
S1601_C03_010E	0.107	0.037	0.004
S1601_C03_014E	0.050	0.064	0.440
S1601_C03_018E	0.272	0.089	0.002
lambda	0.649	0.036	0.000

- The lowest coeficient of Asian and Pacific Island languages who speak english very well is 0.0107, which is **the weakest relationship** between Poverty and English speaker level.
- The highest coeficient of Other languages who speak english very well is 0.249, which is **the strongest relationship** between Poverty and English speaker level. The coefficient of 0.942 represents the estimated change in the dependent variable (below poverty level) associated with a one-unit increase in the independent variable (proportion of people who speak other languages and have limited English proficiency), holding all other variables constant

conclusion

Overall, comparing speaking well and speaking English less well, speaking English less well has **more explained** the below poverty level **than** very well, since R-square of **speaking English less than very well** larger than **speaking English speak very well**.