Summary

Explore language-education-employement correlation

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

data source

Employment Status (S2301)

Limited English Speking Household (S1601)

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

Speaking english very well and unemployment

- Dependent variable is" Limited English speaker"
- - **S2301_C04_021E**: Estimate!!Unemployment rate!!Population 20 to 64 years
- 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_C04_021E               Number of Observations:        2495
Mean dependent var  :      6.6778               Number of Variables   :           5
S.D. dependent var  :      4.0548               Degrees of Freedom    :        2490
R-squared           :      0.0127
Adjusted R-squared  :      0.0111
Sum squared residual:     40483.4               F-statistic           :      8.0086
Sigma-square        :      16.258               Prob(F-statistic)     :   2.06e-06
S.E. of regression  :       4.032               Log likelihood        :   -7016.539
Sigma-square ML     :      16.226               Akaike info criterion :   14043.077
S.E of regression ML:      4.0281               Schwarz criterion     :   14072.188

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     t-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT         6.59668         0.17962         36.72530         0.00000
      S1601_C03_006E         0.00035         0.00019          1.87161         0.06138
      S1601_C03_010E         0.00151         0.00059          2.57642         0.01004
      S1601_C03_014E        -0.00220         0.00046         -4.83115         0.00000
      S1601_C03_018E         0.00173         0.00190          0.91130         0.36223
------------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER            4.716

TEST ON NORMALITY OF ERRORS
TEST                             DF        VALUE           PROB
Jarque-Bera                       2       16230.045           0.0000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                             DF        VALUE           PROB
Breusch-Pagan test                4           44.487           0.0000
Koenker-Bassett test              4            6.447           0.1682
================================ END OF REPORT =====================================
```

The OLS model describes the correlation between the variables. In the OLS model, the R-square is 0.0127, representing the unemployment explains 1.27% of the speaking English very well.

It suggests a **weak relationship** between the English speaker and the unemployment rate.

	Coeff.	Std. Error	P-Value
CONSTANT	6.597	0.180	0.000
S1601_C03_006E	0.000	0.000	0.061
S1601_C03_010E	0.002	0.001	0.010
S1601_C03_014E	-0.002	0.000	0.000
S1601_C03_018E	0.002	0.002	0.362

The p-values for speaking English very well and those who speak **Spanish** and **other languages** are larger than 0.05, which means there is a **non-significant** relationship between **well English speakers** and the **unemployment rate**.

The coefficient for speaking English very well who speaks **Asian and Pacific Island languages** is -.0002, which indicates that as speaking English very well who speaking Asian and Pacific Island languages **decreases**, the likelihood of an unemployment rate also **decreases**.

SWLS model interpretation

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

SUMMARY OF OUTPUT: GM SPATIALLY WEIGHTED LEAST SQUARES (HET)
-----------------------------------------------------------
Data set            :     unknown
Weights matrix      :     unknown
Dependent Variable  :S2301_C04_021E                Number of Observations:        2495
Mean dependent var  :      6.6778               Number of Variables   :           5
S.D. dependent var  :      4.0548               Degrees of Freedom    :        2490
Pseudo R-squared    :      0.0118
N. of iterations    :           1               Step1c computed       :          No

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT         6.37798         0.28053        22.73557         0.00000
       S1601_C03_006E         0.00047         0.00023         2.08552         0.03702
       S1601_C03_010E         0.00132         0.00074         1.79816         0.07215
       S1601_C03_014E        -0.00149         0.00061        -2.44609         0.01444
       S1601_C03_018E         0.00221         0.00182         1.21367         0.22487
              lambda         0.41824         0.03699        11.30660         0.00000
------------------------------------------------------------------------------------
============================== END OF REPORT =====================================
```

The swls model defines the correlation between the variables. In the SWLS model, the R-square is 0.0118, representing the unemployment explains 1.18% of the speaking english very well.
It suggests a **weak** relationship between the well English speaker and unemployment rate.

	Coeff.	Std. Error	P-Value
CONSTANT	6.378	0.281	0.000
S1601_C03_006E	0.000	0.000	0.037
S1601_C03_010E	0.001	0.001	0.072
S1601_C03_014E	-0.001	0.001	0.014
S1601_C03_018E	0.002	0.002	0.225
lambda	0.418	0.037	0.000

The p-values for speaking English very well and those who speak **Other Indo-European languages** and **other languages** are larger than 0.05, which means there is a **non-significant** relationship

between **well English speakers** and the **unemployment rate**.

The coefficient for speaking English very well who speaks **Asian and Pacific Island languages** is -.0001, which indicates that as speaking English very well who speaking Asian and Pacific Island languages **decreases**, the likelihood of an unemployment rate also **decreases**.

Speaking english less well and unemployment

- Dependent variable is" Limited English speaker"
- - **S2301_C04_021E**: Estimate!!Unemployment rate!!Population 20 to 64 years
- 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_C04_021E               Number of Observations:        2495
Mean dependent var  :      6.6778               Number of Variables   :           5
S.D. dependent var  :      4.0548               Degrees of Freedom    :        2490
R-squared           :      0.0223
Adjusted R-squared  :      0.0207
Sum squared residual:     40090.4               F-statistic           :     14.1893
Sigma-square        :      16.101               Prob(F-statistic)     :   1.872e-11
S.E. of regression  :       4.013               Log likelihood        :   -7004.369
Sigma-square ML     :      16.068               Akaike info criterion :   14018.739
S.E of regression ML:      4.0085               Schwarz criterion     :   14047.849

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     t-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT         6.18991         0.13875         44.61045         0.00000
      S1601_C05_006E         0.00106         0.00019          5.53111         0.00000
      S1601_C05_010E         0.00346         0.00081          4.26322         0.00002
      S1601_C05_014E        -0.00097         0.00038         -2.55779         0.01059
      S1601_C05_018E         0.00229         0.00364          0.62972         0.52893
------------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER            3.300

TEST ON NORMALITY OF ERRORS
TEST                             DF        VALUE           PROB
Jarque-Bera                       2       18242.096           0.0000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                             DF        VALUE           PROB
Breusch-Pagan test                4          42.496           0.0000
Koenker-Bassett test              4           5.848           0.2108
================================ END OF REPORT =====================================
```

The OLS model defines the correlation between the variables. In the OLS model, the R-square is 0.0223, representing the unemployment explains 2.23% of the speaking English less well.
It suggests a **weak relationship** between the limited English speaker and the unemployment rate.

	Coeff.	Std. Error	P-Value
CONSTANT	6.190	0.139	0.000
S1601_C05_006E	0.001	0.000	0.000
S1601_C05_010E	0.003	0.001	0.000
S1601_C05_014E	-0.001	0.000	0.011
S1601_C05_018E	0.002	0.004	0.529

The p-values for speaking English very well and those who speak **other languages** are larger than 0.05, meaning there is a **non-significant** relationship between **limited English speakers** and the **unemployment rate**.

The coefficient for speaking English less well who speaks **Asian and Pacific Island languages** is -.0001, which indicates that as speaking English less well who speaking Asian and Pacific Island languages **decreases**, the likelihood of an unemployment rate also **decreases**.

swls model interpretation

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

SUMMARY OF OUTPUT: GM SPATIALLY WEIGHTED LEAST SQUARES (HET)
-----------------------------------------------------------------
Data set            :     unknown
Weights matrix      :     unknown
Dependent Variable  :S2301_C04_021E               Number of Observations:        2495
Mean dependent var  :      6.6778               Number of Variables   :           5
S.D. dependent var  :      4.0548               Degrees of Freedom    :        2490
Pseudo R-squared    :      0.0219
N. of iterations    :           1               Step1c computed       :          No

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT         6.20762         0.20731        29.94392         0.00000
      S1601_C05_006E         0.00085         0.00024         3.60642         0.00031
      S1601_C05_010E         0.00307         0.00092         3.34516         0.00082
      S1601_C05_014E        -0.00061         0.00041        -1.50700         0.13181
      S1601_C05_018E         0.00345         0.00336         1.02466         0.30553
              lambda         0.39912         0.03804        10.49256         0.00000
------------------------------------------------------------------------------------
================================ END OF REPORT =====================================
```

The SWLS model defines the correlation between the variables. In the SWLS model, the R-square is 0.0219, representing the unemployment explains 2.19% of the speaking english very well.
It suggests a **weak** relationship between the well English speaker and unemployee rate.

	Coeff.	Std. Error	P-Value
CONSTANT	6.208	0.207	0.000
S1601_C05_006E	0.001	0.000	0.000
S1601_C05_010E	0.003	0.001	0.001
S1601_C05_014E	-0.001	0.000	0.132
S1601_C05_018E	0.003	0.003	0.306
lambda	0.399	0.038	0.000

The p-values for speaking English very well and those who speak **other languages** are larger than 0.05, meaning there is a **non-significant** relationship between **limited English speakers** and the

unemployment rate.

The coefficient for speaking English less well who speaks **Asian and Pacific Island languages** is -.0001, which indicates that as speaking English less well who speaking Asian and Pacific Island languages **decreases**, the likelihood of an unemployment rate also **decreases**.

summary

Throughout the models, the unemploye rate has less explained the enlish speaking, meaning english speak level does not strongly asscosiate with unemployment rate.

In this section, I decide to use english proficiency data and employment/ unemployment data to investigate correaltion.

Limtied english and employment

- Dependent variable is" Limited English speaker"
- - **S2301_C03_021E**: Estimate!!employment rate!!Population 20 to 64 years
- 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_C03_021E               Number of Observations:        2495
Mean dependent var  :     72.1040               Number of Variables   :           5
S.D. dependent var  :     10.6676               Degrees of Freedom    :        2490
R-squared           :      0.0274
Adjusted R-squared  :      0.0259
Sum squared residual:      276029               F-statistic           :     17.5472
Sigma-square        :     110.855               Prob(F-statistic)     :   3.275e-14
S.E. of regression  :      10.529               Log likelihood        :   -9411.259
Sigma-square ML     :     110.633               Akaike info criterion :   18832.518
S.E of regression ML:     10.5182               Schwarz criterion     :   18861.628

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     t-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT        70.13033         0.46903       149.52249         0.00000
      S1601_C03_006E         0.00022         0.00049         0.45371         0.65008
      S1601_C03_010E         0.00652         0.00153         4.26690         0.00002
      S1601_C03_014E         0.00642         0.00119         5.39654         0.00000
      S1601_C03_018E         0.01145         0.00496         2.30687         0.02114
------------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER            4.716

TEST ON NORMALITY OF ERRORS
TEST                             DF        VALUE           PROB
Jarque-Bera                       2       51175.339           0.0000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                             DF        VALUE           PROB
Breusch-Pagan test                4         157.676           0.0000
Koenker-Bassett test              4          13.728           0.0082
================================ END OF REPORT =====================================
```

The OLS model defines the correlation between the variables. In the OLS model, the R-square is 0.0259, representing the employment explains 2.59% of the speaking English less well.
It suggests a **weak relationship** between the limited English speaker and the employment rate.

	Coeff.	Std. Error	P-Value
CONSTANT	70.130	0.469	0.000
S1601_C03_006E	0.000	0.000	0.650
S1601_C03_010E	0.007	0.002	0.000
S1601_C03_014E	0.006	0.001	0.000
S1601_C03_018E	0.011	0.005	0.021

The p-values for speaking English very well and those who speak **Spanish** are larger than 0.05, meaning there is a **non-significant** relationship between **limited English speakers** and the **employment rate**.

Speaking english well and employment

- Dependent variable is" Limited English speaker"
- - **S2301_C03_021E**: Estimate!!employment rate!!Population 20 to 64 years
- 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"

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

SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES
-----------------------------------------
Data set            :     unknown
Weights matrix      :        None
Dependent Variable  :S2301_C03_021E               Number of Observations:        249
Mean dependent var  :     72.1040               Number of Variables   :          5
S.D. dependent var  :     10.6676               Degrees of Freedom    :       2490
R-squared           :      0.0150
Adjusted R-squared  :      0.0134
Sum squared residual:      279548               F-statistic           :     9.4905
Sigma-square        :     112.268               Prob(F-statistic)     :  1.298e-07
S.E. of regression  :      10.596               Log likelihood        :  -9427.062
Sigma-square ML     :     112.043               Akaike info criterion :  18864.124
S.E of regression ML:     10.5851               Schwarz criterion     :  18893.234

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     t-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT        72.40114         0.36640       197.60124         0.00000
      S1601_C05_006E        -0.00213         0.00051        -4.19073         0.00003
      S1601_C05_010E         0.00391         0.00214         1.82331         0.06838
      S1601_C05_014E         0.00273         0.00100         2.73644         0.00625
      S1601_C05_018E         0.01207         0.00962         1.25486         0.20965
------------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER            3.300

TEST ON NORMALITY OF ERRORS
TEST                             DF        VALUE           PROB
Jarque-Bera                       2       59503.598           0.0000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                             DF        VALUE           PROB
Breusch-Pagan test                4         275.238           0.0000
Koenker-Bassett test              4          22.309           0.0002
================================ END OF REPORT =====================================
```

The OLS model defines the correlation between the variables. In the OLS model, the R-square is 0.0134, representing the employment explains 1.34% of the speaking English well.

It suggests a **weak relationship** between the well English speaker and the employment rate.

	Coeff.	Std. Error	P-Value
CONSTANT	72.401	0.366	0.000
S1601_C05_006E	-0.002	0.001	0.000
S1601_C05_010E	0.004	0.002	0.068
S1601_C05_014E	0.003	0.001	0.006
S1601_C05_018E	0.012	0.010	0.210

The coefficient for speaking English very well who speaks **Spanish** is -.002, which indicates that as speaking English very well who speaking Spanish **decreases**, the likelihood of an employment rate also **decreases**.

The p-values for speaking English very well and those who speak **Other Indo-European languages** are larger than 0.05, meaning there is a **non-significant** relationship between **limited English speakers** and the **employment rate**.

hint

- Overall, English efficiency (Proficient English speakers or limited English speakers) rarely explains the employment rate and unemployment rate.
- whether proficient or limited English speakers who speak the Asian language do not have a high association with the unemployment rate and employment rate, it can suggest English efficiency does not have a huge influence on the employment situation in the Asian community.
    - As we can see Speaking English less well and the unemployment model, Limited English proficiency is inversely proportional to the unemployment rate.