Geographically Weighted Regression

GGS 563

Introduction

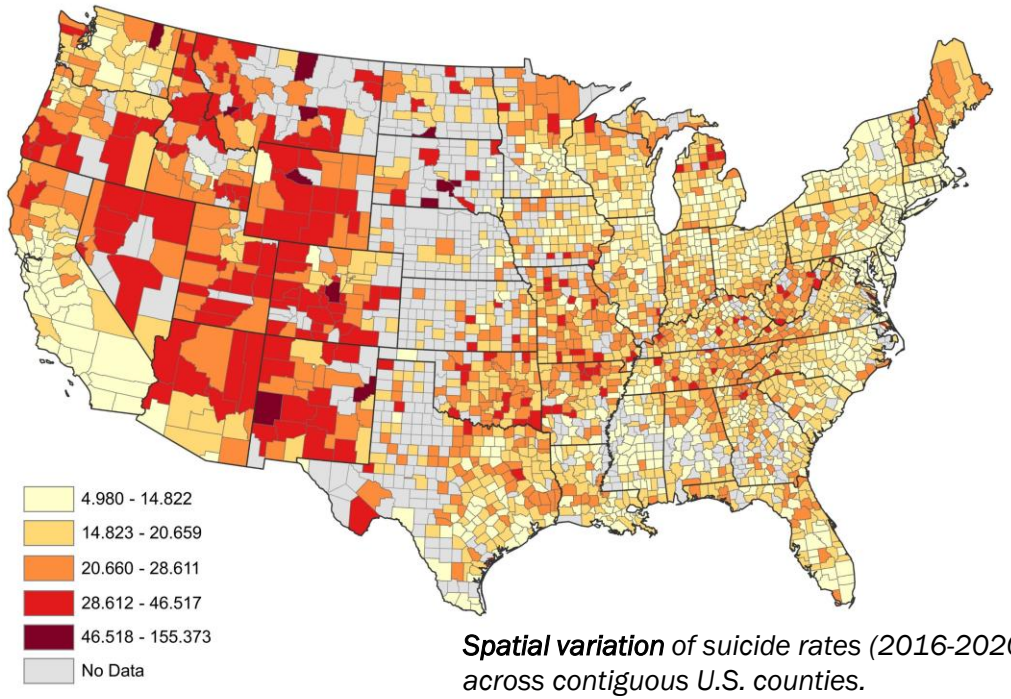
- Relationships between variables can change across space → spatial non-stationarity
- Traditional regression (e.g., OLS) assume:
  - Relationships are the same everywhere, considered a **global model**
  - Residuals are independent, but **spatial autocorrelation** means nearby residuals are correlated
- **Geographically Weighted Regression (GWR):** A spatial regression method that captures spatially varying relationships by fitting localized regressions
- **Multiscale Geographically Weighted Regression (MGWR):** An extension of GWR that recognizes the relationship between variables may operate at different spatial scales

```
Call:
lm(formula = suicide_z ~ dep_z + vet_z + blk_z + AIAN_z + ment_z +
    rur_z, data = counties_data)

Residuals:
    Min      1Q  Median      3Q     Max
-3.2529 -0.4890  0.0140  0.4703  4.0918

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -8.600e-17  1.570e-02   0.000        1
dep_z        1.158e-01  1.847e-02   6.271 4.25e-10 ***
vet_z        2.462e-01  1.673e-02  14.719  < 2e-16 ***
blk_z       -2.785e-01  1.738e-02 -16.027  < 2e-16 ***
AIAN_z       1.838e-01  1.616e-02  11.373  < 2e-16 ***
ment_z       1.581e-01  1.654e-02   9.559  < 2e-16 ***
rur_z        1.737e-01  1.843e-02   9.422  < 2e-16 ***
---
```

*OLS regression output from R, explaining the **GLOBAL** relationships between select independent variables and suicide rates.*



***Spatial variation** of suicide rates (2016-2020) across contiguous U.S. counties.*

Applications of GWR or MGWR

Public Health

- How do the relationships between county-level breast cancer mortality and population characteristics vary geographically in the US? [1]

Environment and Climate

- How do the associations between PM2.5 concentrations and socioeconomic factors vary across China's major cities? [2]

Urban Planning

- What are the spatially varying relationships between landcover types and the urban climate (e.g. land surface temperature) in Southeast China? [3]

Transportation

- What is the spatiotemporal influence of the built environment (e.g., sidewalk density, traffic volume) on transit ridership? [4]

Agriculture

- What is the spatially varying effects of mechanical and biochemical technologies, government payments, product diversity and farm size on farming viability in Ohio? [5]

1. *Anderson, T., Herrera, D., Mireku, F., Barner, K., Kokkinakis, A., Dao, H., ... & Pierobon, M. (2023). Geographical variation in social determinants of female breast cancer mortality across US counties. JAMA network open, 6(9), e2333618-e2333618.*
2. *Zhao, X., Zhou, W., Han, L., & Locke, D. (2019). Spatiotemporal variation in PM2. 5 concentrations and their relationship with socioeconomic factors in China's major cities. Environment international, 133, 105145.*
3. *Hu, X., & Xu, H. (2019). Spatial variability of urban climate in response to quantitative trait of land cover based on GWR model. Environmental Monitoring and Assessment, 191(3), 194.*
4. *Ma, X., Zhang, J., Ding, C., & Wang, Y. (2018). A geographically and temporally weighted regression model to explore the spatiotemporal influence of built environment on transit ridership. Computers, Environment and Urban Systems, 70, 113-124.*
5. *Jiang, Z., & Xu, B. (2014). Geographically weighted regression analysis of the spatially varying relationship between farming viability and contributing factors in Ohio. Regional Science Policy & Practice, 6(1), 69-84.*

Understanding GWR Concepts

Local Models:

- GWR runs a separate regression with each independent variable for each location (feature)
- Each local equation uses the dependent and independent variables from nearby features around the target location
- For each location *i*, run regression using data weighted by distance to *i*

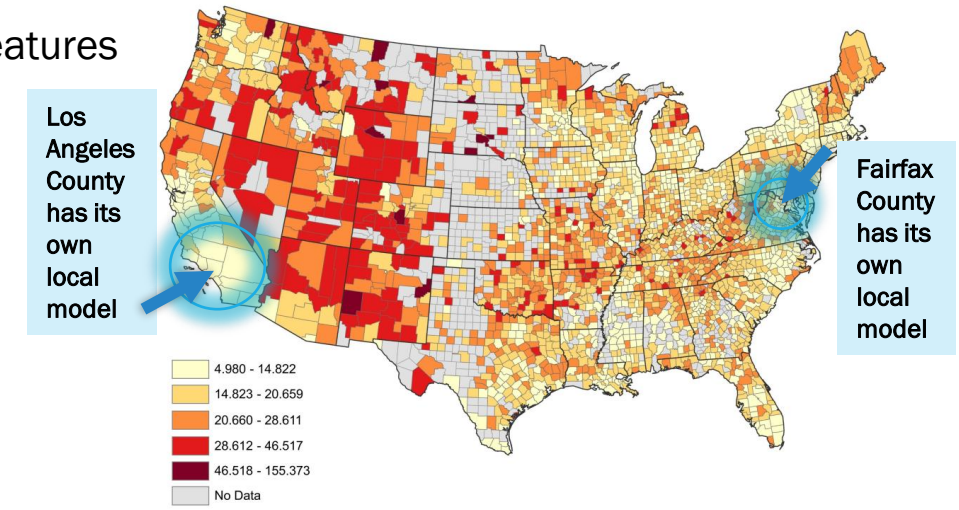
Local Weighting Scheme:

- GWR applies spatial weighting so closer features have more influence
- Weights decrease with distance, determined by a kernel function
- Two common kernels: Gaussian (gradual decay) and Bisquare (sharp cut-off)

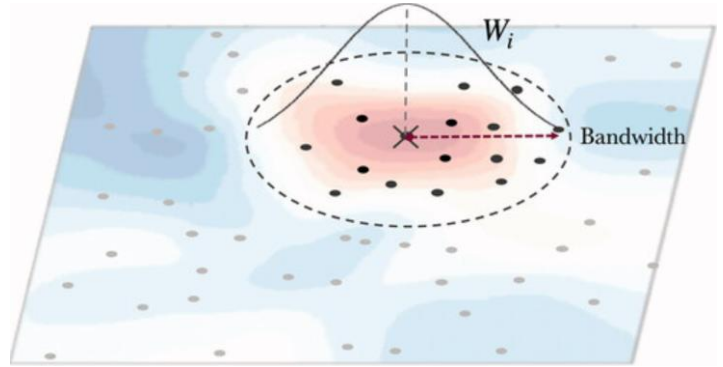
Bandwidth Size:

- Controls the size of the neighborhood considered for each local regression
- Large bandwidth: uses more distant features, less local variation
- Small bandwidth: focuses on closer features, capturing more local detail
- Optimal bandwidth size can be determined through cross validation

Outputs: Maps and tables for local coefficients, R^2, and statistical significance

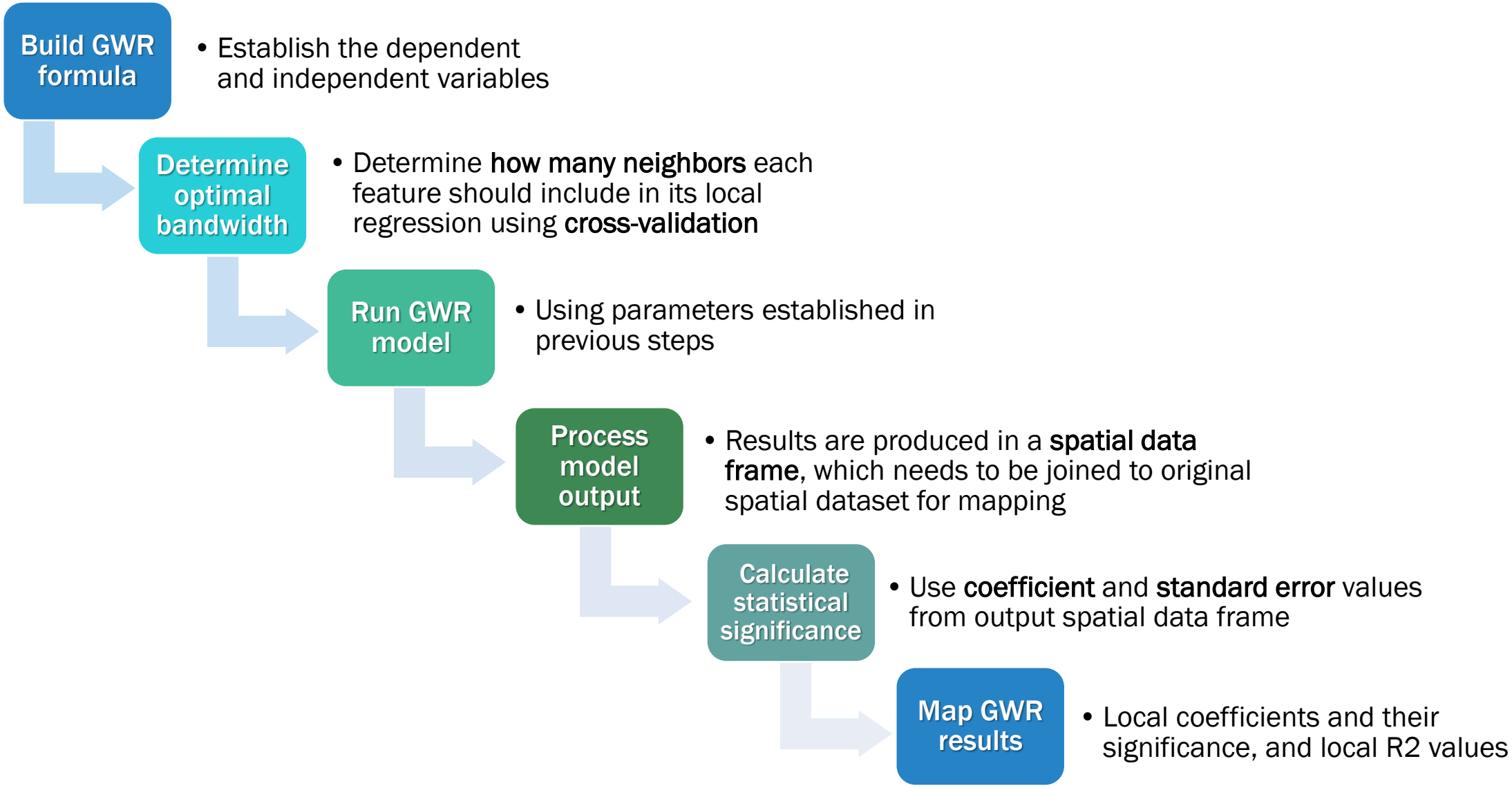


Suicide rates (2016-2020) across contiguous U.S. counties.



Representation of GWR forming local regression models using nearby features and bandwidth size

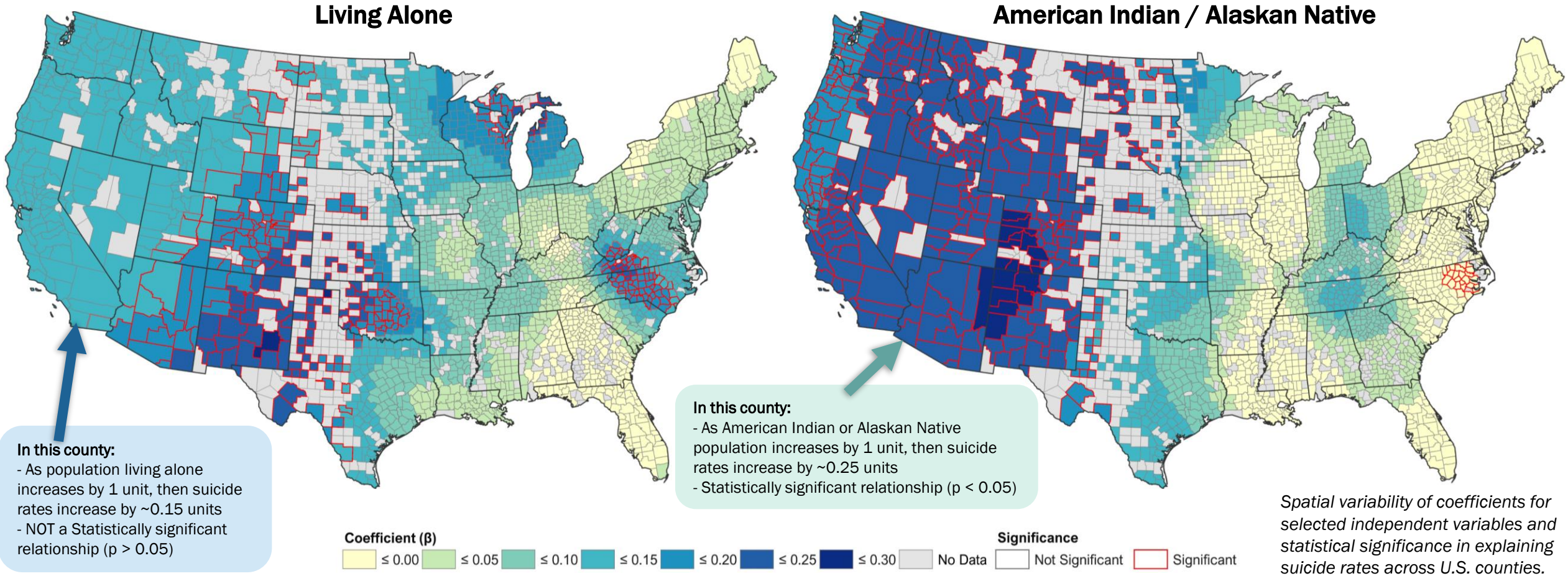
GWR Modeling Workflow in R



Interpreting GWR Outputs: Coefficient Maps

Coefficient: Interpretation similar to OLS. The colors within each feature's boundary (e.g., county) represent the coefficient values

Statistical Significance: Each feature's boundary (e.g., county) indicate statistical significance at the 95% confidence level

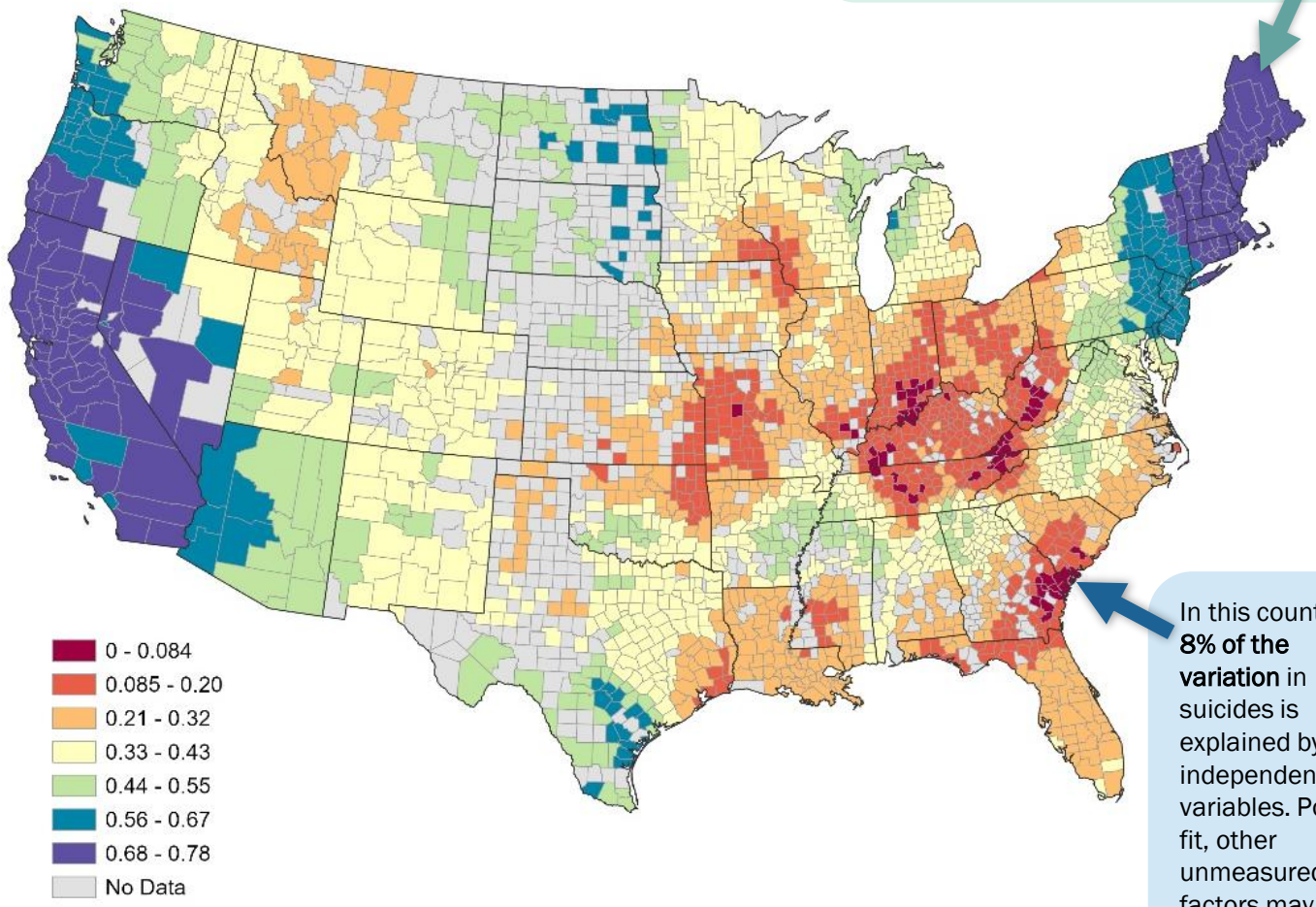


Spatial variability of coefficients for selected independent variables and statistical significance in explaining suicide rates across U.S. counties.

Interpreting GWR Outputs: R-Squared Map

- Shows how well the model explains variation in the dependent variable at each feature (e.g., county)
- Identifies spatial variation in where the model performs better or worse
  - High local R^2: Model explains much of the variation in that area (good fit)
  - Low local R^2: Model explains little variation (poor fit → missing variables or different processes)

In this county, about **68-78% of the variation** in suicide rates is explained by the independent variables in the model, *given the local relationships estimated for that area*. This suggests a relatively good model fit here, meaning the predictors capture most of what's driving the outcome in this location.



In this county, **0-8% of the variation** in suicides is explained by the independent variables. Poor fit, other unmeasured factors may be important.

Results showing the spatial variability of R-squared values computed in a GWR model that explains suicide rates across U.S. counties.

Interpreting GWR Outputs: Spatial Data Frame from R

- The GWR model in R outputs a spatial data frame that contains all the local outputs for each feature (e.g., county)

- The resulting table will include:
  - **Intercept:** local intercept estimates
  - **Coefficients:** describes the type of relationship between given independent variable and dependent variable, one column for each independent variable
  - **Standard Errors:** represents the uncertainty in the calculated coefficient, one column for each independent variable
  - **R-squared:** local variance measure on how well the dependent variable is explained in the model
  - Other parameters and model predictions

1	gwr_sum.w	gwr_X.Intercept.	gwr_vet_z	gwr_X.Intercept._se	gwr_vet_z_se	gwr_gwr.e	gwr_pred	gwr_pred.se	gwr_localR2	gwr_X.Intercept._se_EDF	gwr_vet_z_se_EDF	gwr_pred.se.1	gwr_X	gwr_Y
2	78.981	-0.261	0.016	0.093	0.060	0.019	-0.451	0.130	0.658	0.095	0.061	0.133	-97.455	47.924
3	100.308	0.251	0.193	0.061	0.046	0.405	0.546	0.147	0.630	0.063	0.047	0.150	-103.403	47.745
4	81.359	0.045	0.142	0.077	0.053	-0.822	1.004	0.113	0.659	0.079	0.055	0.116	-101.321	47.612
5	87.562	0.119	0.159	0.071	0.051	0.631	1.074	0.094	0.653	0.073	0.052	0.097	-101.837	47.309
6	86.071	0.087	0.145	0.075	0.052	-0.239	0.416	0.091	0.662	0.076	0.053	0.094	-101.286	46.719
7	103.565	0.131	0.176	0.057	0.045	-0.393	0.834	0.173	0.646	0.059	0.046	0.177	-102.356	48.204
8	83.153	-0.188	0.066	0.082	0.054	-0.983	0.974	0.110	0.666	0.084	0.056	0.113	-98.725	48.270
9	94.216	-0.090	0.114	0.066	0.048	-0.403	1.977	0.165	0.662	0.067	0.049	0.169	-99.846	48.771
10	81.150	0.094	0.139	0.079	0.054	0.859	2.379	0.179	0.669	0.081	0.055	0.183	-101.047	46.113

Example resulting spatial data frame from a computed GWR model in R

Interpreting GWR Outputs: Spatial Data Frame from R

- **Model weights and location information**

The sum of the kernel weights used for this location. Higher means more neighboring observations were given weight when estimating coefficients for that point

The X and Y coordinate of the observation (the centroid longitude and latitude)

	gwr_sum.w	gwr_X.Intercept.	gwr_vet_z	gwr_X.Intercept._se	gwr_vet_z_se	gwr_gwr.e	gwr_pred	gwr_pred.se	gwr_localR2	gwr_X.Intercept._se_EDF	gwr_vet_z_se_EDF	gwr_pred.se.1	gwr_X	gwr_Y
2	78.981	-0.261	0.016	0.093	0.060	0.019	-0.451	0.130	0.658	0.095	0.061	0.133	-97.455	47.924
3	100.308	0.251	0.193	0.061	0.046	0.405	0.546	0.147	0.630	0.063	0.047	0.150	-103.403	47.745
4	81.359	0.045	0.142	0.077	0.053	-0.822	1.004	0.113	0.659	0.079	0.055	0.116	-101.321	47.612
5	87.562	0.119	0.159	0.071	0.051	0.631	1.074	0.094	0.653	0.073	0.052	0.097	-101.837	47.309
6	86.071	0.087	0.145	0.075	0.052	-0.239	0.416	0.091	0.662	0.076	0.053	0.094	-101.286	46.719
7	103.565	0.131	0.176	0.057	0.045	-0.393	0.834	0.173	0.646	0.059	0.046	0.177	-102.356	48.204
8	83.153	-0.188	0.066	0.082	0.054	-0.983	0.974	0.110	0.666	0.084	0.056	0.113	-98.725	48.270
9	94.216	-0.090	0.114	0.066	0.048	-0.403	1.977	0.165	0.662	0.067	0.049	0.169	-99.846	48.771
10	81.150	0.094	0.139	0.079	0.054	0.859	2.379	0.179	0.669	0.081	0.055	0.183	-101.047	46.113

Example resulting spatial data frame from a computed GWR model in R

Interpreting GWR Outputs: Spatial Data Frame from R

- **Coefficients**

Local intercept estimate for that location's regression equation

Local coefficient estimate for the given variable

In this case, only the veteran variable is shown. Each independent variable will have a coefficient column!

	gwr_sum.w	gwr_X.Intercept.	gwr_vet_z	gwr_X.Intercept._se	gwr_vet_z_se	gwr_gwr.e	gwr_pred	gwr_pred.se	gwr_localR2	gwr_X.Intercept._se_EDF	gwr_vet_z_se_EDF	gwr_pred.se.1	gwr_X	gwr_Y
2	78.981	-0.261	0.016	0.093	0.060	0.019	-0.451	0.130	0.658	0.095	0.061	0.133	-97.455	47.924
3	100.308	0.251	0.193	0.061	0.046	0.405	0.546	0.147	0.630	0.063	0.047	0.150	-103.403	47.745
4	81.359	0.045	0.142	0.077	0.053	-0.822	1.004	0.113	0.659	0.079	0.055	0.116	-101.321	47.612
5	87.562	0.119	0.159	0.071	0.051	0.631	1.074	0.094	0.653	0.073	0.052	0.097	-101.837	47.309
6	86.071	0.087	0.145	0.075	0.052	-0.239	0.416	0.091	0.662	0.076	0.053	0.094	-101.286	46.719
7	103.565	0.131	0.176	0.057	0.045	-0.393	0.834	0.173	0.646	0.059	0.046	0.177	-102.356	48.204
8	83.153	-0.188	0.066	0.082	0.054	-0.983	0.974	0.110	0.666	0.084	0.056	0.113	-98.725	48.270
9	94.216	-0.090	0.114	0.066	0.048	-0.403	1.977	0.165	0.662	0.067	0.049	0.169	-99.846	48.771
10	81.150	0.094	0.139	0.079	0.054	0.859	2.379	0.179	0.669	0.081	0.055	0.183	-101.047	46.113

In this county, as the veteran population increases by one unit, then suicide rates increase by 0.139 units

Example resulting spatial data frame from a computed GWR model in R

Interpreting GWR Outputs: Spatial Data Frame from R

- **Standard Errors (SE)**

Standard error for intercept estimate

Standard error for coefficient estimate

In this case, only the veteran variable is shown. Each independent variable will have a SE column!

	gwr_sum.w	gwr_X.Intercept.	gwr_vet_z	gwr_X.Intercept._se	gwr_vet_z_se	gwr_gwr.e	gwr_pred	gwr_pred.se	gwr_localR2	gwr_X.Intercept._se_EDF	gwr_vet_z_se_EDF	gwr_pred.se.1	gwr_X	gwr_Y
2	78.981	-0.261	0.016	0.093	0.060	0.019	-0.451	0.130	0.658	0.095	0.061	0.133	-97.455	47.924
3	100.308	0.251	0.193	0.061	0.046	0.405	0.546	0.147	0.630	0.063	0.047	0.150	-103.403	47.745
4	81.359	0.045	0.142	0.077	0.053	-0.822	1.004	0.113	0.659	0.079	0.055	0.116	-101.321	47.612
5	87.562	0.119	0.159	0.071	0.051	0.631	1.074	0.094	0.653	0.073	0.052	0.097	-101.837	47.309
6	86.071	0.087	0.145	0.075	0.052	-0.239	0.416	0.091	0.662	0.076	0.053	0.094	-101.286	46.719
7	103.565	0.131	0.176	0.057	0.045	-0.393	0.834	0.173	0.646	0.059	0.046	0.177	-102.356	48.204
8	83.153	-0.188	0.066	0.082	0.054	-0.983	0.974	0.110	0.666	0.084	0.056	0.113	-98.725	48.270
9	94.216	-0.090	0.114	0.066	0.048	-0.403	1.977	0.165	0.662	0.067	0.049	0.169	-99.846	48.771
10	81.150	0.094	0.139	0.079	0.054	0.859	2.379	0.179	0.669	0.081	0.055	0.183	-101.047	46.113

Example resulting spatial data frame from a computed GWR model in R

Interpreting GWR Outputs: Spatial Data Frame from R

- **Model Fit, Predictions and Diagnostics**

Local residual (observed – predicted) value

Local predicted value

Standard error for the prediction

Local R-squared

Standard errors of the intercept and coefficient adjusted for effective degrees of freedom (EDF)

Alternative prediction standard error, adjusted for EDF

	gwr_sum.w	gwr_X.Intercept.	gwr_vet_z	gwr_X.Intercept._se	gwr_vet_z_se	gwr_gwr.e	gwr_pred	gwr_pred.se	gwr_localR2	gwr_X.Intercept._se_EDF	gwr_vet_z_se_EDF	gwr_pred.se.1	gwr_X	gwr_Y
2	78.981	-0.261	0.016	0.093	0.060	0.019	-0.451	0.130	0.658	0.095	0.061	0.133	-97.455	47.924
3	100.308	0.251	0.193	0.061	0.046	0.405	0.546	0.147	0.630	0.063	0.047	0.150	-103.403	47.745
4	81.359	0.045	0.142	0.077	0.053	-0.822	1.004	0.113	0.659	0.079	0.055	0.116	-101.321	47.612
5	87.562	0.119	0.159	0.071	0.051	0.631	1.074	0.094	0.653	0.073	0.052	0.097	-101.837	47.309
6	86.071	0.087	0.145	0.075	0.052	-0.239	0.416	0.091	0.662	0.076	0.053	0.094	-101.286	46.719
7	103.565	0.131	0.176	0.057	0.045	-0.393	0.834	0.173	0.646	0.059	0.046	0.177	-102.356	48.204
8	83.153	-0.188	0.066	0.082	0.054	-0.983	0.974	0.110	0.666	0.084	0.056	0.113	-98.725	48.270
9	94.216	-0.090	0.114	0.066	0.048	-0.403	1.977	0.165	0.662	0.067	0.049	0.169	-99.846	48.771
10	81.150	0.094	0.139	0.079	0.054	0.859	2.379	0.179	0.669	0.081	0.055	0.183	-101.047	46.113

Example resulting spatial data frame from a computed GWR model in R

Interpreting GWR Outputs: Summary Coefficient Statistics

Standardized independent variables for explaining US county-level suicide rates:
dep_z = % of population aged less than 18 or greater than 65
vet_z = % of population considered veterans
blk_z = % of Black population
AIAN_z = % of American Indian or Alaskan Native population
men_z = Average number of mentally unhealthy days
rur_z = % of rural land

Model Parameters:

- Kernel function → Type of kernel function used to weight nearby observations
  - This model used Gaussian kernel
- Adaptive quantile → The bandwidth size for the model; How many features were used in each local regression model
  - This model used about 31 nearest neighbors (counties) out of 2402 counties = ~1.33% of the features in the dataset

```
Call:
gwr(formula = gwr_formula, data = as.data.frame(counties_data),
    coords = coords, adapt = gwr_bandwidth, hatmatrix = TRUE,
    se.fit = TRUE)
Kernel function: gwr.Gauss
Adaptive quantile: 0.01331557 (about 31 of 2402 data points)
Summary of GWR coefficient estimates at data points:
                  Min.   1st Qu.    Median   3rd Qu.      Max.  Global
X.Intercept. -0.682032 -0.189754  0.094148  0.249801  1.376254  0.0000
ment_z       -0.228778  0.137130  0.212541  0.313406  0.721638  0.1581
rur_z        -0.119907  0.072974  0.156324  0.232150  0.486128  0.1737
dep_z        -0.278224  0.014940  0.095901  0.178961  0.337773  0.1158
vet_z        -0.072347  0.105920  0.185919  0.315695  0.692662  0.2462
blk_z        -0.676136 -0.330099 -0.204576 -0.082692  0.212449 -0.2785
AIAN_z       -0.395135 -0.050827  0.029444  0.140135  0.626200  0.1838
Number of data points: 2402
Effective number of parameters (residual: 2traceS - traceS'S): 323.5208
Effective degrees of freedom (residual: 2traceS - traceS'S): 2078.479
Sigma (residual: 2traceS - traceS'S): 0.6311322
Effective number of parameters (model: traceS): 228.839
Effective degrees of freedom (model: traceS): 2173.161
Sigma (model: traceS): 0.6172303
Sigma (ML): 0.5870927
AICc (GWR p. 61, eq 2.33; p. 96, eq. 4.21): 4766.654
AIC (GWR p. 96, eq. 4.22): 4486.941
Residual sum of squares: 827.9162
Quasi-global R2: 0.6551786
```

GWR model output in R for a model that explains suicide rates across U.S. counties

Interpreting GWR Outputs: Summary Coefficient Statistics

Standardized independent variables for explaining US county-level suicide rates:
dep_z = % of population aged less than 18 or greater than 65
vet_z = % of population considered veterans
blk_z = % of Black population
AIAN_z = % of American Indian or Alaskan Native population
men_z = Average number of mentally unhealthy days
rur_z = % of rural land

Model Parameters:

- Sigma→ Estimated standard deviation of residuals
  - Measures the typical size of the errors between predicted and observed values
  - Smaller values = predictions are closer to the actual data (better fit)
- AICc (Corrected Akaike Information Criterion) → Model comparison score
  - Balances model fit and model complexity (penalizes overly complex models); Lower AICc = better model
- Quasi-global R-squared → overall fit measure
  - Similar to R^2 in OLS: the proportion of variance in the dependent variable explained by the model
  - **This GWR model collectively explain about 65.5% of the variation in the dependent variable**
  - ***Does this improve compared to the OLS model?***

```
Call:
gwr(formula = gwr_formula, data = as.data.frame(counties_data),
    coords = coords, adapt = gwr_bandwidth, hatmatrix = TRUE,
    se.fit = TRUE)
Kernel function: gwr.Gauss
Adaptive quantile: 0.01331557 (about 31 of 2402 data points)
Summary of GWR coefficient estimates at data points:
                  Min.   1st Qu.    Median   3rd Qu.      Max.  Global
X.Intercept. -0.682032 -0.189754  0.094148  0.249801  1.376254  0.0000
ment_z       -0.228778  0.137130  0.212541  0.313406  0.721638  0.1581
rur_z        -0.119907  0.072974  0.156324  0.232150  0.486128  0.1737
dep_z        -0.278224  0.014940  0.095901  0.178961  0.337773  0.1158
vet_z        -0.072347  0.105920  0.185919  0.315695  0.692662  0.2462
blk_z        -0.676136 -0.330099 -0.204576 -0.082692  0.212449 -0.2785
AIAN_z       -0.395135 -0.050827  0.029444  0.140135  0.626200  0.1838
Number of data points: 2402
Effective number of parameters (residual: 2traceS - traceS'S): 323.5208
Effective degrees of freedom (residual: 2traceS - traceS'S): 2078.479
Sigma (residual: 2traceS - traceS'S): 0.6311322
Effective number of parameters (model: traceS): 228.839
Effective degrees of freedom (model: traceS): 2173.161
Sigma (model: traceS): 0.6172303
Sigma (ML): 0.5870927
AICc (GWR p. 61, eq 2.33; p. 96, eq. 4.21): 4766.654
AIC (GWR p. 96, eq. 4.22): 4486.941
Residual sum of squares: 827.9162
Quasi-global R2: 0.6551786
```

GWR model output in R for a model that explains suicide rates across U.S. counties

Interpreting GWR Outputs: Summary Coefficient Statistics

Standardized independent variables for explaining US county-level suicide rates:
dep_z = % of population aged less than 18 or greater than 65
vet_z = % of population considered veterans
blk_z = % of Black population
AIAN_z = % of American Indian or Alaskan Native population
men_z = Average number of mentally unhealthy days
rur_z = % of rural land

Local Coefficient Distribution:

- The **minimum, 1st quartile, median, 3rd quartile and maximum** coefficient is identified for each independent variable → Provides the range of coefficients across the study area
- 'Global' column→ The resulting coefficient if a single global model (OLS) was fitted
- How to interpret in practice!
  - **Spatial variation in effects**: If Min and Max differ a lot, the variable's relationship with the dependent variable changes substantially by location
  - **Sign changes**: If Min is negative and Max is positive, the variable's effect flips sign in different areas (important for policy implications)

```
Call:
gwr(formula = gwr_formula, data = as.data.frame(counties_data),
    coords = coords, adapt = gwr_bandwidth, hatmatrix = TRUE,
    se.fit = TRUE)
Kernel function: gwr.Gauss
Adaptive quantile: 0.01331557 (about 31 of 2402 data points)
Summary of GWR coefficient estimates at data points:
                  Min.   1st Qu.    Median   3rd Qu.      Max.  Global
X.Intercept. -0.682032 -0.189754  0.094148  0.249801  1.376254  0.0000
ment_z       -0.228778  0.137130  0.212541  0.313406  0.721638  0.1581
rur_z        -0.119907  0.072974  0.156324  0.232150  0.486128  0.1737
dep_z        -0.278224  0.014940  0.095901  0.178961  0.337773  0.1158
vet_z        -0.072347  0.105920  0.185919  0.315695  0.692662  0.2462
blk_z        -0.676136 -0.330099 -0.204576 -0.082692  0.212449 -0.2785
AIAN_z       -0.395135 -0.050827  0.029444  0.140135  0.626200  0.1838
Number of data points: 2402
Effective number of parameters (residual: 2traceS - traceS'S): 323.5208
Effective degrees of freedom (residual: 2traceS - traceS'S): 2078.479
Sigma (residual: 2traceS - traceS'S): 0.6311322
Effective number of parameters (model: traceS): 228.839
Effective degrees of freedom (model: traceS): 2173.161
Sigma (model: traceS): 0.6172303
Sigma (ML): 0.5870927
AICc (GWR p. 61, eq 2.33; p. 96, eq. 4.21): 4766.654
AIC (GWR p. 96, eq. 4.22): 4486.941
Residual sum of squares: 827.9162
Quasi-global R2: 0.6551786
```

GWR model output in R for a model that explains suicide rates across U.S. counties

Interpreting GWR Outputs: Summary Coefficient Statistics

Standardized independent variables for explaining US county-level suicide rates:
dep_z = % of population aged less than 18 or greater than 65
vet_z = % of population considered veterans
blk_z = % of Black population
AIAN_z = % of American Indian or Alaskan Native population
men_z = Average number of mentally unhealthy days
rur_z = % of rural land

Local Coefficient Distribution:

- Interpreting how the relationship between average number of mentally unhealthy days and suicide rates changes over space
  - In some counties, the relationship between is negative (–0.23), in others strongly positive (+0.72)
  - The median effect is slightly positive (0.21)
  - The global model would have produced a coefficient of 0.16, missing spatial variation

Key idea: Large variation between Min and Max suggests strong spatial heterogeneity

- But are all these coefficients statistically significant?

```
Call:
gwr(formula = gwr_formula, data = as.data.frame(counties_data),
    coords = coords, adapt = gwr_bandwidth, hatmatrix = TRUE,
    se.fit = TRUE)
Kernel function: gwr.Gauss
Adaptive quantile: 0.01331557 (about 31 of 2402 data points)
Summary of GWR coefficient estimates at data points:
                  Min.   1st Qu.    Median   3rd Qu.      Max.  Global
X.Intercept. -0.682032 -0.189754  0.094148  0.249801  1.376254  0.0000
ment_z       -0.228778  0.137130  0.212541  0.313406  0.721638  0.1581
rur_z        -0.119907  0.072974  0.156324  0.232150  0.486128  0.1737
dep_z        -0.278224  0.014940  0.095901  0.178961  0.337773  0.1158
vet_z        -0.072347  0.105920  0.185919  0.315695  0.692662  0.2462
blk_z        -0.676136 -0.330099 -0.204576 -0.082692  0.212449 -0.2785
AIAN_z       -0.395135 -0.050827  0.029444  0.140135  0.626200  0.1838
Number of data points: 2402
Effective number of parameters (residual: 2traceS - traceS'S): 323.5208
Effective degrees of freedom (residual: 2traceS - traceS'S): 2078.479
Sigma (residual: 2traceS - traceS'S): 0.6311322
Effective number of parameters (model: traceS): 228.839
Effective degrees of freedom (model: traceS): 2173.161
Sigma (model: traceS): 0.6172303
Sigma (ML): 0.5870927
AICc (GWR p. 61, eq 2.33; p. 96, eq. 4.21): 4766.654
AIC (GWR p. 96, eq. 4.22): 4486.941
Residual sum of squares: 827.9162
Quasi-global R2: 0.6551786
```

GWR model output in R for a model that explains suicide rates across U.S. counties

Interpreting GWR Outputs: Coefficient Significance

Significance not shown in this table

```
Call:
gwr(formula = gwr_formula, data = as.data.frame(counties_data),
    coords = coords, adapt = gwr_bandwidth, hatmatrix = TRUE,
    se.fit = TRUE)
Kernel function: gwr.Gauss
Adaptive quantile: 0.01331557 (about 31 of 2402 data points)
Summary of GWR coefficient estimates at data points:
                  Min.   1st Qu.    Median   3rd Qu.      Max.  Global
X.Intercept. -0.682032 -0.189754  0.094148  0.249801  1.376254  0.0000
ment_z       -0.228778  0.137130  0.212541  0.313406  0.721638  0.1581
rur_z        -0.119907  0.072974  0.156324  0.232150  0.486128  0.1737
dep_z        -0.278224  0.014940  0.095901  0.178961  0.337773  0.1158
vet_z        -0.072347  0.105920  0.185919  0.315695  0.692662  0.2462
blk_z        -0.676136 -0.330099 -0.204576 -0.082692  0.212449 -0.2785
AIAN_z       -0.395135 -0.050827  0.029444  0.140135  0.626200  0.1838
Number of data points: 2402
Effective number of parameters (residual: 2traceS - traceS'S): 323.5208
Effective degrees of freedom (residual: 2traceS - traceS'S): 2078.479
Sigma (residual: 2traceS - traceS'S): 0.6311322
Effective number of parameters (model: traceS): 228.839
Effective degrees of freedom (model: traceS): 2173.161
Sigma (model: traceS): 0.6172303
Sigma (ML): 0.5870927
AICc (GWR p. 61, eq 2.33; p. 96, eq. 4.21): 4766.654
AIC (GWR p. 96, eq. 4.22): 4486.941
Residual sum of squares: 827.9162
Quasi-global R2: 0.6551786
```

GWR model output in R for a model that explains suicide rates across U.S. counties

Need to calculate significance separate!

	Variable	Significant_Count	Significant_Percent
vet_z	vet_z	1522	63.36386
ment_z	ment_z	1563	65.07077
blk_z	blk_z	1267	52.74771
AIAN_z	AIAN_z	814	33.88843
rur_z	rur_z	1148	47.79351
dep_z	dep_z	852	35.47044

Calculated statistical significance of coefficients for each independent variable in the GWR model

Calculating Coefficient Significance

Example for Fairfax County

Variable → Mental Unhealthy Days:
β = 0.168
SE = 0.085

$$t = \frac{0.168}{0.085} = \mathbf{1.867}$$

1.867 < 1.96
→ Not significant!

1. Get local coefficients and their standard errors

•The GWR output will have the following for each location (feature) and independent variable:

- β = local coefficient
- SE = standard error of that coefficient

2. For each feature, calculate the local t-statistic

- $t = \frac{\beta}{SE}$

3. Compare to critical t-value

- For a **two-tailed** test at α=0.05, compare t to the critical value from the t-distribution (often 1.96)
- If $|t| > \pm 1.96$ → coefficient is statistically significant at 95% confidence level ($p < 0.05$) at that location

Need to calculate significance separate!

	Variable	Significant_Count	Significant_Percent
vet_z	vet_z	1522	63.36386
ment_z	ment_z	1563	65.07077
blk_z	blk_z	1267	52.74771
AIAN_z	AIAN_z	814	33.88843
rur_z	rur_z	1148	47.79351
dep_z	dep_z	852	35.47044

Calculated statistical significance of coefficients for each independent variable in the GWR model

Interpreting GWR Outputs: Coefficient Significance

Local Coefficient Statistical Significance:

- Significant Count/Percent → the number/percentage of features (e.g., counties) in the study area that have a statistically significant coefficient
  - **There is a statistically significant relationship between veteran population and suicide in *1522 or 63.36%* of counties**
  - **There is a statistically significant relationship between dependent population (ages less than 18 or 65+) in *852 or 35.47% of counties***

```
Call:
gwr(formula = gwr_formula, data = as.data.frame(counties_data),
    coords = coords, adapt = gwr_bandwidth, hatmatrix = TRUE,
    se.fit = TRUE)
Kernel function: gwr.Gauss
Adaptive quantile: 0.01331557 (about 31 of 2402 data points)
Summary of GWR coefficient estimates at data points:
                Min.   1st Qu.    Median   3rd Qu.      Max.  Global
X.Intercept. -0.682032 -0.189754  0.094148  0.249801  1.376254  0.0000
ment_z       -0.228778  0.137130  0.212541  0.313406  0.721638  0.1581
rur_z        -0.119907  0.072974  0.156324  0.232150  0.486128  0.1737
dep_z        -0.278224  0.014940  0.095901  0.178961  0.337773  0.1158
vet_z        -0.072347  0.105920  0.185919  0.315695  0.692662  0.2462
blk_z        -0.676136 -0.330099 -0.204576 -0.082692  0.212449 -0.2785
AIAN_z       -0.395135 -0.050827  0.029444  0.140135  0.626200  0.1838
Number of data points: 2402
Effective number of parameters (residual: 2traceS - traceS'S): 323.5208
Effective degrees of freedom (residual: 2traceS - traceS'S): 2078.479
Sigma (residual: 2traceS - traceS'S): 0.6311322
Effective number of parameters (model: traceS): 228.839
Effective degrees of freedom (model: traceS): 2173.161
Sigma (model: traceS): 0.6172303
Sigma (ML): 0.5870927
AICc (GWR p. 61, eq 2.33; p. 96, eq. 4.21): 4766.654
AIC (GWR p. 96, eq. 4.22): 4486.941
Residual sum of squares: 827.9162
Quasi-global R2: 0.6551786
```

GWR model output in R for a model that explains suicide rates across U.S. counties

	Variable	Significant_Count	Significant_Percent
vet_z	vet_z	1522	63.36386
ment_z	ment_z	1563	65.07077
blk_z	blk_z	1267	52.74771
AIAN_z	AIAN_z	814	33.88843
rur_z	rur_z	1148	47.79351
dep_z	dep_z	852	35.47044

Calculated statistical significance of coefficients for each independent variable in the GWR model

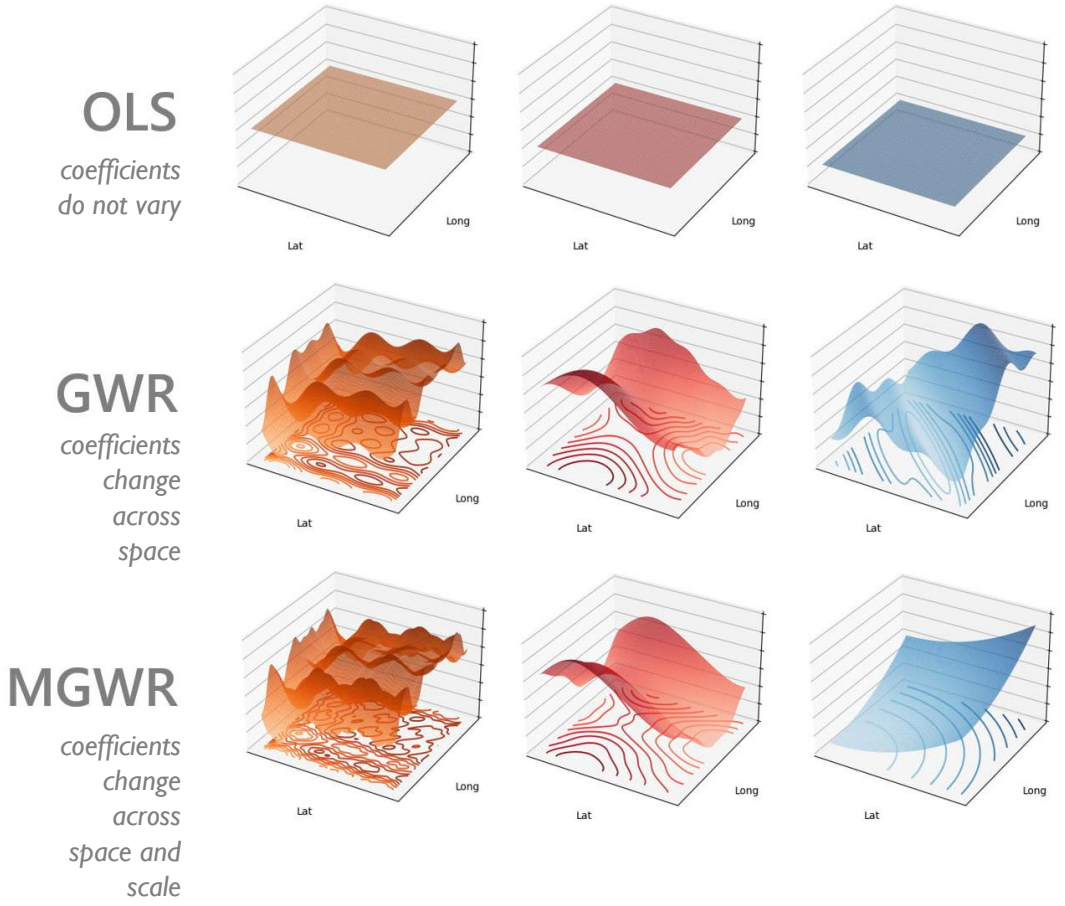
An Extension to GWR: MGWR

Limitations of GWR:

- GWR assumes all predictors operate at the same spatial scale (one bandwidth for all)
    - In reality, different predictors often influence the outcome at different spatial scales
        - Example: Education level may influence health outcomes over a broad region, but air pollution may influence locally within a few kilometers

What is MGWR:

- Fits local regressions with variable-specific bandwidths
- Allows each predictor to vary at the scale → can reveal predictors that are **global**, **regional**, or **local** in influence
- Improves model fit and interpretation, often improving AICc, residuals, and quasi-global R^2



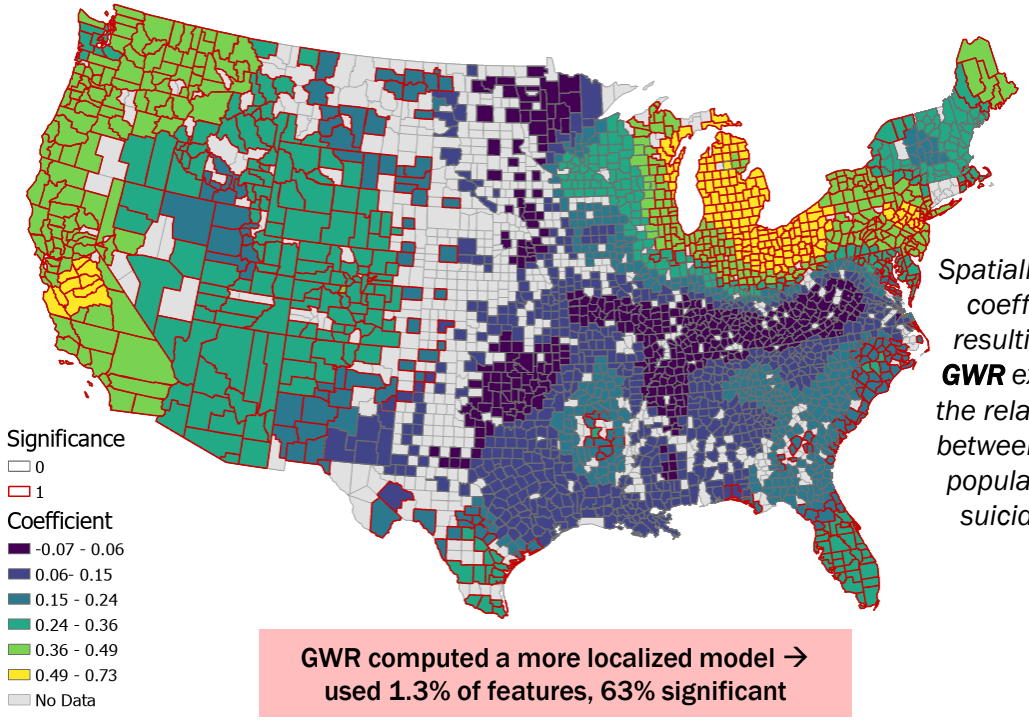
Varying coefficient surfaces between OLS, GWR and MGWR models

MGWR Modeling

1. Starts like GWR → computes local regressions using spatial weighting
2. Optimizes a separate bandwidth for each predictor (independent variable)
3. Produces local coefficients for each variable, just like GWR, but dependent on spatial scale

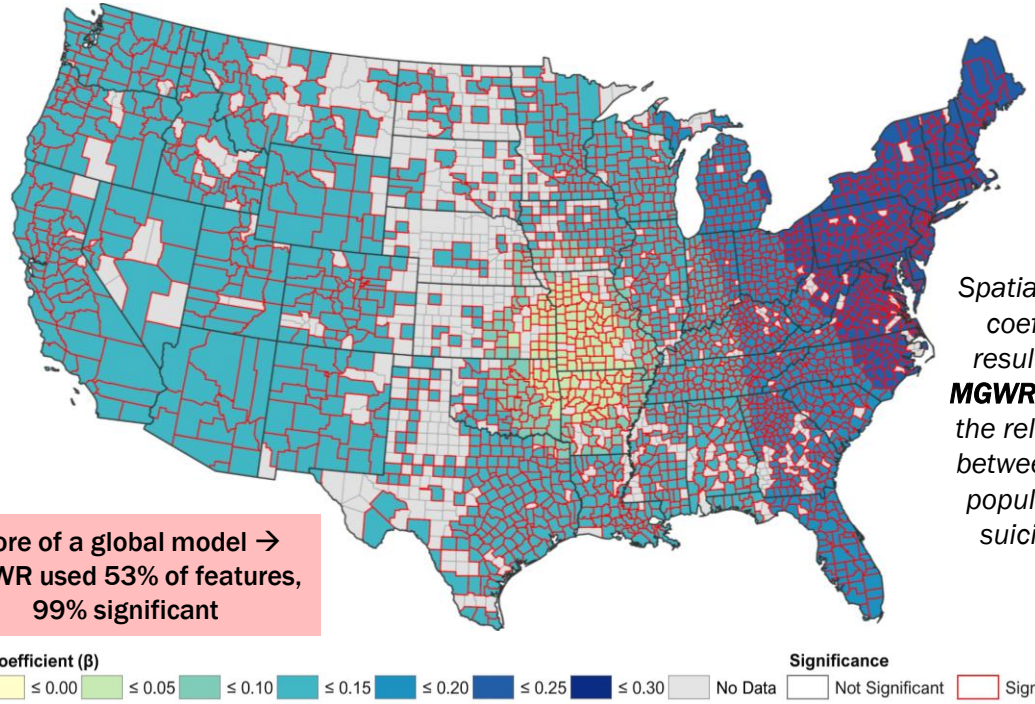
MGWR model results table describing a subset of the spatially varying relationships between county-level determinants and suicides, including the mean coefficient, standard deviation, minimum coefficient, median coefficient, maximum coefficient, number and percentage of neighbors, and number and percentage of significant features for each standardized independent variable.

Standardized Variable	Mean	Standard Deviation	Minimum	Maximum	Neighbors (% of Features)	Significance (% of Features)
Smart Phone Access	-0.087	0.0071	-0.1087	-0.072	2195 (97.08)	1153 (47.84)
Veteran	0.1757	0.0509	0.0903	0.2708	1286 (53.36)	2390 (99.17)
Social Connectedness	0.2021	0.1228	-0.0627	0.4593	264 (10.95)	1061 (44.02)
Unemployed	0.0187	0.0511	-0.1787	0.1275	409 (16.97)	0 (0.00)
Bachelor's Degree	-0.079	0.0033	-0.0823	-0.0688	2410 (100.00)	0 (0.00)
Food Insecurity	0.1099	0.0414	0.0572	0.1883	1363 (56.76)	1070 (44.40)
Poor Mental Health Days	0.1068	0.033	0.0688	0.172	1501 (62.28)	1396 (57.93)
Primary Care Physicians	0.0047	0.0052	-0.0014	0.0194	2410 (100.00)	0 (0.00)
Mental Care Providers	0.0095	0.0041	0.0053	0.022	2410 (100.00)	0 (0.00)
No Health Insurance	0.0184	0.0462	-0.0734	0.1094	888 (36.85)	0 (0.00)



*Spatially varying coefficients resulting from **GWR** explaining the relationships between veteran population and suicide rates*

GWR computed a more localized model → used 1.3% of features, 63% significant



More of a global model → MGWR used 53% of features, 99% significant

*Spatially varying coefficients resulting from **MGWR** explaining the relationships between veteran population and suicide rates*