

# Spatial Analysis Writing Sample (GIS Methods and Interpretation)

**Author:** Siddha Khatal

**Methods:** GIS, Spatial Statistics, Index Construction

## Context

This writing sample comes from a graduate GIS course focused on spatial analysis in the social sciences. The assignment examined spatial concentration and segregation using census data and standard segregation indices.

## Data and Units of Analysis

The analysis uses ward-level data for Washington, D.C. in 1860 and census tract-level data for Chicago from recent census records. I completed all data preparation, calculations, and mapping independently.

## Analytical Approach

The analysis applies choropleth and proportional symbol mapping to visualize population distribution. Segregation is measured using the Index of Dissimilarity and the Isolation Index, both computed directly from census counts. Results are interpreted with attention to spatial scale, clustering, and limits of inference.

## Relevance to Environmental and Litigation Consulting

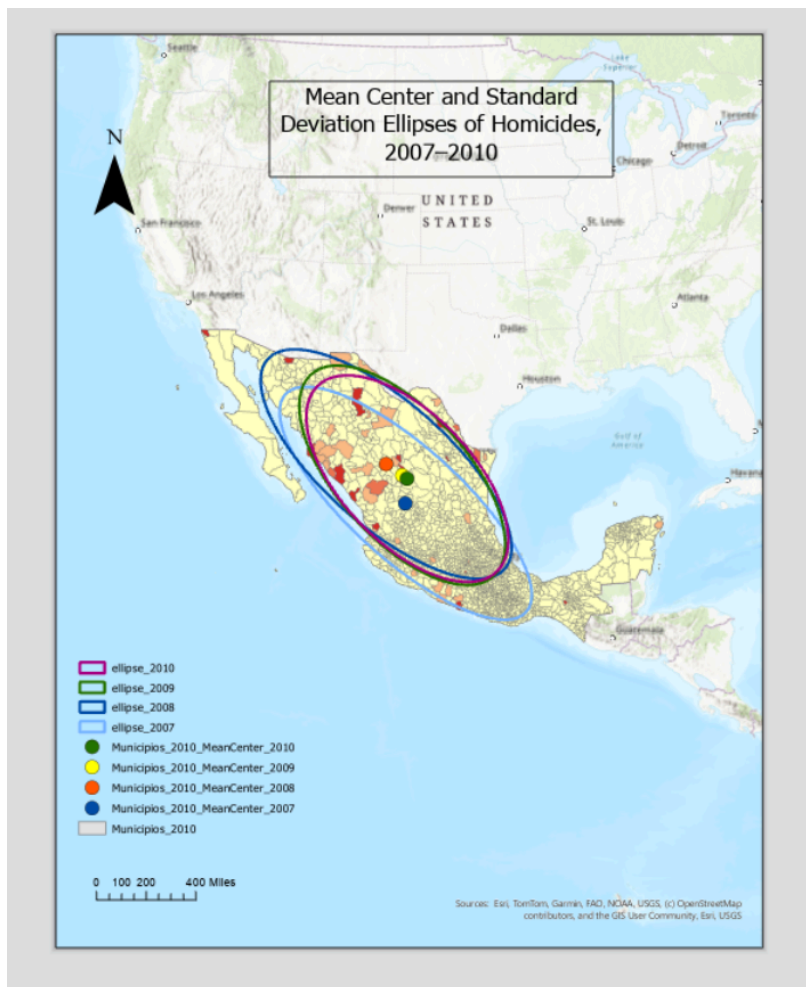
This sample demonstrates quantitative reasoning under data constraints and clear explanation of statistical indices used in technical reports. It reflects disciplined interpretation without causal overstatement and shows the ability to translate spatial analysis into defensible written conclusions.

## Author Contribution

I completed all data preparation, analysis, visualization, and interpretation.

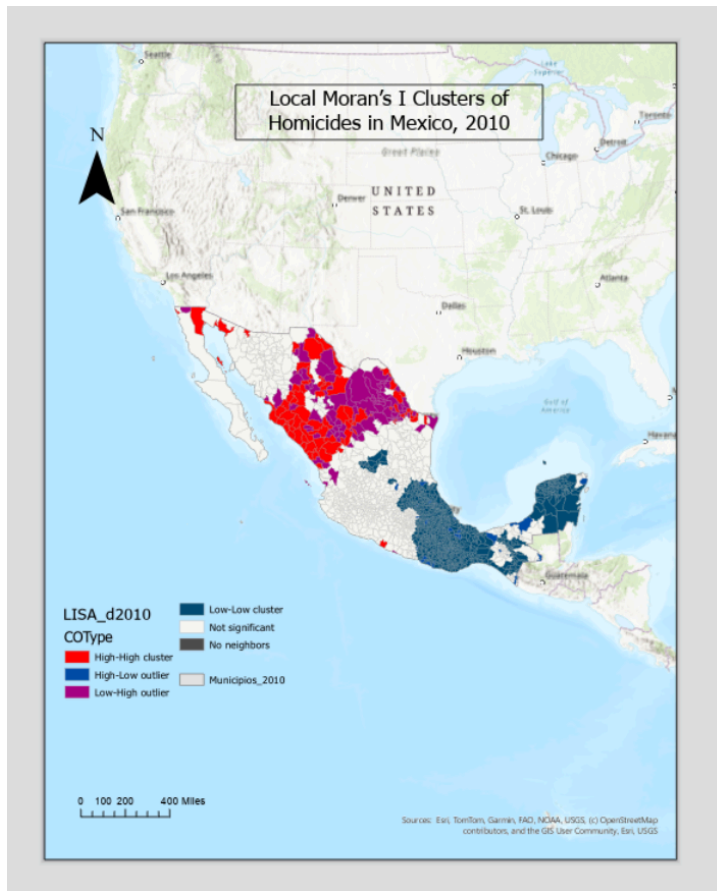
*Note: This draft is part of an ongoing working paper and is submitted solely as a writing sample.*

Can you identify any pattern of how violence has changed in Mexico *using mean center and standard deviation ellipses*? Please describe and show a map with your results, illustrating your findings.



- The mean centers moved northwest from 2007 to 2010. This shows that the core of homicide activity shifted toward the northern and Pacific regions.
- The ellipses grow larger and more elongated over time. The 2007 ellipse is compact. The 2010 ellipse stretches along a northwest to southeast axis. This shows that violence spread across a wider corridor rather than remaining localized.
- The map, therefore, indicates two clear changes. Violence moved northwest, and its spatial extent expanded between 2007 and 2010.

Where were the most clustered municipalities in Mexico in 2010? Please describe and show a map to illustrate your findings.



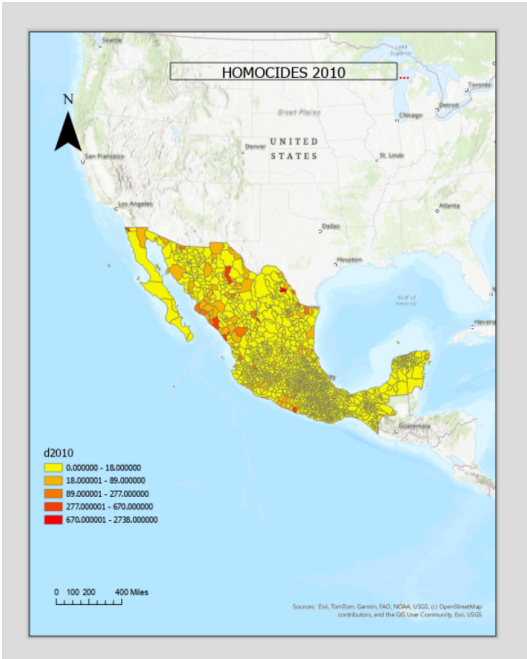
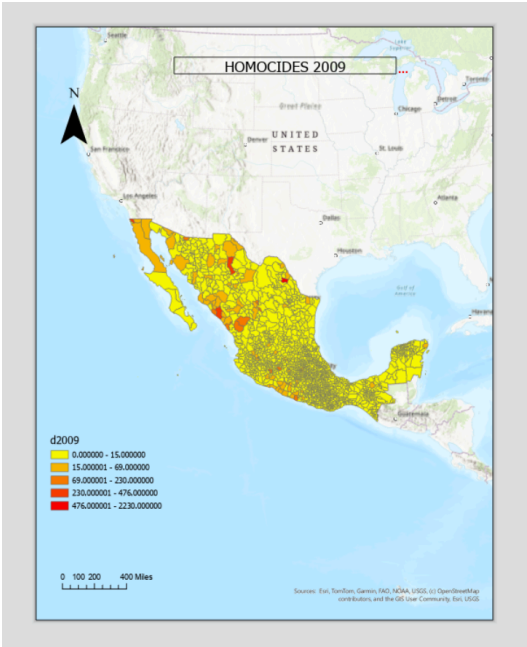
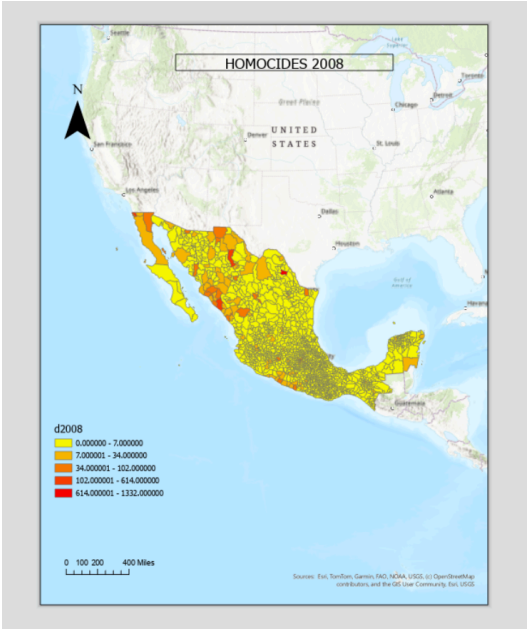
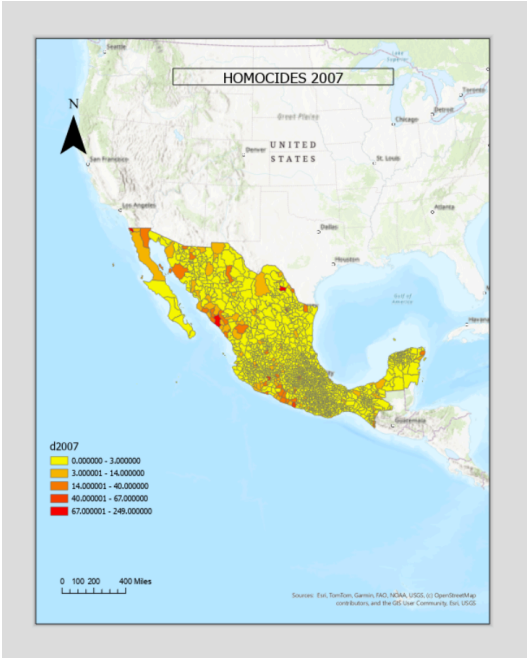
The strongest High–High clusters in 2010 appear in northern and western Mexico. These municipalities record high homicide levels and are surrounded by neighbors with similarly high values. The clusters form large continuous regions across Chihuahua, Sinaloa, Durango, Zacatecas, and Michoacán. These areas contain the most concentrated patterns of lethal violence in the country.

Low–Low clusters appear mainly in the southeast. These municipalities show low homicide levels and are bordered by neighbors with similarly low values. They form broad stable zones in Yucatán, Campeche, and parts of Chiapas.

High–Low and Low–High outliers are scattered and do not form large regions. They represent local deviations where a municipality differs sharply from the pattern of its immediate neighbors.

The map shows that the most clustered violent municipalities in 2010 lie in northern and western Mexico, while the southeast contains the largest clusters of low violence.

Using the tools you learned in previous assignments, generate a map layout where you display the number of homicides in Mexico for each year of the 2007-2010 period.



I created four choropleth maps that show homicide counts by municipality for each year from 2007 to 2010. All four maps use the same graduated color ramp and classification method to keep the visual comparison consistent. The maps show that lethal violence is concentrated in western and northern Mexico throughout the period. Higher-intensity areas grow darker and spread over time, especially in 2009 and 2010. The southeast remains mostly in the lowest categories in every year. The sequence of maps shows increasing intensity and geographic expansion of homicides across Mexico between 2007 and 2010.

**Bonus: Did you find any statistically significant correlation between education and violence in Mexico for 2010? Do your findings make any sense? Use the results from an ordinary least squares calculation (explanation below) for your answer, but do not limit your analysis to it.**

I estimated an OLS model with 2010 homicides as the dependent variable and average years of education as the explanatory variable. The coefficient on education is positive (4.24) and statistically significant at the one percent level. This means municipalities with higher average education levels tend to have higher reported homicide counts in 2010. The result is counterintuitive, because higher education usually predicts lower violence.

The model fit is weak. The R-squared is 0.011, so education explains about one percent of the variation in homicides. This limited explanatory power shows that the relationship captured by the coefficient does not reflect a strong structural link. It reflects an association rather than a causal mechanism.

The residual diagnostics reveal more problems. The Koenker test is not significant, so heteroskedasticity is not the main issue. The Jarque-Bera test is strongly significant, which means the residuals are not normally distributed. The histogram of homicides is right-skewed, with a few extremely violent municipalities driving most of the variation. This skewness produces large positive residuals and biases OLS inference. The Moran scatterplot shows a clear positive slope, which indicates spatial autocorrelation. High-violence municipalities cluster together, and this violates the independence assumption required by OLS.

The standardized residual map confirms these issues. Large positive residuals appear in northern and western Mexico, where spatial clustering is strongest. Large negative residuals appear in parts of the southeast. These patterns show that the model fails to capture regional structure and spatial dependence.

Overall, the coefficient is statistically significant but not substantively meaningful. The model fit is weak, the residuals are biased, and spatial clustering invalidates the OLS assumptions. Education alone does not explain municipal homicide patterns in Mexico. The findings make sense only once spatial dependence, regional context, and distributional skewness are taken into account.