University College London

Department of Statistics

A Geographically Weighted Regression model for identifying spatial non-stationarity in femicide data for Colombia

Student

Juan Camilo Gómez Romero

Supervisor

Dr Thomas Honnor

September 25, 2023

Contents

1 Introduction

The widespread issue of violence against women is a significant global concern since it represents a human rights violation (UN, 1993, 1979). Its impact can be both immediate and far-reaching, encompassing various physical, sexual, psychological, and even lethal consequences for women. It detrimentally affects women's well-being and hinders their full participation in society. Beyond its adverse effects on women, violence also has repercussions on their families, communities, and the nation. The significant associated costs, ranging from increased health-care and legal service expenses to productivity losses, strain national public budgets and pose a hindrance to development (UN Women, 2023).

Following many years of activism led by civil society and women's movements, the elimination of gender-based violence is now a recognized priority on both national and international agendas. Many countries have implemented legislation addressing domestic violence, sexual assault, and other forms of violence against women. However, there is still insufficient effort directed towards preventing this violence, and when it does happen, it frequently goes unpunished (UN Women, 2023).

The most recent report from UNODC and UN Women (2022) on femicide indicates that of the 81,000 women intentionally killed recorded globally in 2021, roughly 56% were committed by their partners or other family members. In contrast, the study shows that only 11% of male homicides occur in a similar domestic setting. This research also indicates that the total count of female homicides has remained almost unchanged over the past decade. An estimated 35 percent of women worldwide have faced physical and/or sexual violence from either an intimate partner or non-partner. However, in specific national studies, the data suggests that up to 70 percent of women have experienced physical and/or sexual violence at some stage in their lives, primarily from an intimate partner (WHO, 2021).

This numbers are a grim reminder that violence against women is among the most widespread human rights violations worldwide; although, the report from UNODC and UN Women (2022) indicates regional differences: In 2021, the rate of gender-related homicides occurring within domestic settings was estimated at 2.5 per 100,000 women in Africa. The rate stood at 1.4 in the Americas, 1.2 in Oceania, 0.8 in Asia, and 0.6 in Europe. In the case of Colombia, the figures for femicides are alarming: On Mother's Day in 2023 alone, four women were murdered (Botero, 2023). According to the Office of the Inspector General of Colombia (2023b), there were 614 femicide cases recorded in Colombia in 2022 (i.e. 2.3 femicides per 100,000 women). And between January 1 and July 3, 2023, there have been 320 femicide cases reported in Colombia (Inspector General of Colombia, 2023a).

In Colombia, the femicide of Rosa Elvira Cely, a 35-year-old woman who was kidnapped, sexually assaulted, beaten, strangled, impaled, and ultimately murdered in a public park in Bogotá by her study partner in May 2012, served as a wake-up call for a nation that had become complacent in the face of gender-based violence. Rosa Elvira worked as a street vendor, attended a public school to complete her high school education, and was a single mother with a 12-year-old daughter.

The events took place on May 24, 2012, after a class that ended at 10 p.m., Rosa Elvira decided to leave on a motorcycle with her former classmate, unaware that she was with a murderer. She made two desperate calls to an emergency hotline to report her location and the horrifying incident, but the authorities arrived too late. Three hours after the distress call, Rosa Elvira was found by the side of a stream that runs through the park, in a state of hypothermia, bearing signs of assault and strangulation, unclothed, and surrounded by a pool of blood.

She died four days later in the intensive care unit of Santa Clara Hospital due to a widespread infection. A subsequent report revealed that her internal organs had been severely damaged after being penetrated with a tree branch. On August 2, 2023, a court in Bogotá sentenced three Colombian state entities (the Office of the Attorney General of Colombia, the Health Department of Bogotá, and Santa Clara Hospital) for their failures and negligence in the femicide case of Rosa Elvira Cely, stating that without these shortcomings the crime and its tragic outcome could have been prevented.

This horrific case reflects the discrimination, inequality, gender-based violence, and misogyny that have claimed the lives of thousands of women. After her death, Rosa Elvira Cely became a symbol of justice and a call to society to eradicate violence against women. As a result, in 2015, Colombia passed Law 1761, better known as the Rosa Elvira Cely Law, which recognize femicide as a separate crime and not just an aggravating factor in homicides. When the Rosa Elvira Cely Law was enacted, gender-based violence statistics were deeply concerning, with an average of one woman being killed every three days. Nevertheless, even after eight years, these numbers have not decreased yet (Botero, 2023).

The UN (1993) defines violence against women as any act of gender-based violence that results in, or is likely to result in, physical, sexual, or mental harm or suffering to women, whether occurring in public or in private life. However, the most brutal and extreme form of such violence is the killing of women based on their gender (also known as femicide and feminicide). In general, gender-related killings of women can be defined as intentional homicides driven by motivations linked to gender issues, stemming from a variety of factors such as stereotyped gender roles, discrimination against women, to unequal power dynamics between women and men within society (UNODC and UN Women, 2022).

The report from UNODC and UN Women (2022) also emphasizes that gender-related killings, along with other forms of violence against women, are not inevitable. These crimes can be averted through a combination of strategies, including early detection of women affected by violence and access to support and protection for survivors. Additionally, the report offers further suggestions for addressing the underlying causes, such as reshaping harmful notions of masculinity and social norms, as well as eliminating structural gender inequalities.

It is essential for Colombia to identify the factors related to femicides to be able to reduce and prevent the occurrences of such deaths, but social norms and structural gender inequalities might vary over the municipalities of Colombia, and so, alter the relationship of those explanatory factors with femicides, that implies these relationships might not be the same across space. For example, it might be that there are intrinsic variations of gender bias across Colombia that cause people to behave differently in some areas than in others. Women may be less exposed to experienced gender violence in certain parts of Colombia due to specific local contextual factors; hence, these local structural relationships make necessary to analyse this phenomenon from a geographical perspective to be able to capture this spatial heterogeneity in the relationships.

Therefore, the main goal of our research is to assess if the relationship between femicides and the covariates is constant (stationary) or variable (non-stationary) across municipalities in Colombia, and based on this analysis, we will be able to understand and identify the spatial patterns in the relationship between gender-based homicides in Colombia and their driving factors.

The spatial method we will employ to examine the unequal spatial distribution in the relationship between femicides and their explanatory factors is the Geographically Weighted Regression (GWR) model, as this spatial analysis technique enables the modelling of spatially dependent processes by taking non-stationary variables into consideration (Srinivasan, 2008), under the assumption that contextual factors could potentially alter the strength and direction of the relationship between those features and the response variable. This statistical approach has proven useful in identifying spatial variations in the factors associated with femicides (Ingram and Marchesini da Costa, 2015, 2017; Sepúlveda Murillo et al., 2018).

1.1 Structure of the report

The remainder of this paper is organized as follows: a review of the literature on femicides to identify relevant factors for our research is presented in section 2. Section 3 provides a description and exploratory analysis of our data. In section 4, the methods and results are described and analysed. Finally, we present the conclusions of our research and its limitations in section 5.

2 Literature Review

Several research have identified that femicide is closely linked to sexual violence and domestic or partner violence (Bejarano C., 2014; Sepúlveda Murillo et al., 2018; Beyer et al., 2014; Martins and Nascimento, 2017; Coy et al., 2011). Femicide represents the most severe manifestation of intimate partner violence (Stöckl et al., 2013; Messing et al., 2017), where approximately 56% of female homicides are committed by either an intimate partner or other family members (UNODC and UN Women, 2022). With respect to sexual violence, the Inter-American Commission on Human Rights (IACHR, 2013) has identified that it is frequently observed that women exhibit post-mortem indicators of sexualized abuse and torture. Then, we expect a strong and direct association between sexual violence against women and femicides, as well as between domestic violence and femicides.

Additionally, Moroskoski et al. (2022) found a positive association between fatal and non-fatal violence directed at women. This finding aligns with the research of Meneghel and Portella (2017) that emphasise femicide as the ultimate outcome of a pattern of violence, making these deaths foreseeable and preventable. Moroskoski et al. (2022) also identified a positive connection between fatal violence against women and the rates of male homicides. All these previous findings suggest that a decrease in structural violence would act as a protective element for women as indicated by Meneghel and Hirakata (2011).

Campbell et al. (2003) found that limited educational background increases the risk of femicides, which is consistent with the research of Ingram and Marchesini da Costa (2015) and Brewer and Smith (1995) in which high level of education was associated with a low level of female violence and homicides, and low education level is related to a high incidence of women experiencing violent fatalities (Martins and Nascimento, 2017; Castañeda Salgado, 2016). All these findings highlight the importance of education as a protective factor against femicide. Moreover, Sepúlveda Murillo et al. (2018) identified a negative correlation between the net school coverage rate and femicides, a relationship consistent with the findings of the World Health Organization (WHO, 2012) which affirm that increased government investment in education and the elimination of gender disparities in access to secondary education can be instrumental in decreasing gender-based violence and femicides.

Other factors that could be associated to femicide include housing and poverty (Martins and Nascimento, 2017; Castañeda Salgado, 2016). Grana (2001) found that precarious housing and greater economic dependence tend to have a more direct connection to femicide, while Hernández Breña (2018) demonstrates that instances of violence against women are twice as prevalent in households with low socioeconomic status compared to those with higher status. Nevertheless, it is important to note that vulnerability resulting from limited education and poverty augment the risk of women encountering situations of violence (Tekkas Kerman and Betrus, 2018); hence, it is also very likely that there is a correlation between poverty and non-lethal violence against women, something that we must consider later in our research.

In Sepúlveda Murillo et al. (2018), an index representing the percentage of households facing unmet basic needs was employed as a poverty measure. This index encompasses factors such as inadequate housing, severely overcrowded homes, housing with insufficient services, households heavily reliant on a single income source, and households with school-age children not attending school. Their statistical analysis revealed a notably strong positive correlation between the prevalence of households with unmet basic needs and femicide. On the other hand, Ingram and Marchesini da Costa (2015) demonstrates that marginalization, which comprises a composite measure incorporating indicators of poverty, illiteracy, and rurality, exerts a detrimental impact on various homicide measures, including femicides, since as levels of marginalization increase, so do homicide rates. Therefore, we anticipate a positive relationship between poverty and femicide rates.

Another variable closely related to femicide is the prevalence of teenage pregnancies (Pengpid et al., 2016; Li et al., 2009); numerous studies have identified an elevated risk of physical violence and gender-based violence against pregnant adolescents (Covington et al., 2001; Silverman et al., 2001). Furthermore, the rate of children under the age of 18 may also exhibit a positive correlation with feminicide, as shown by Beyer et al. (2014) and Brewer and Smith (1995). Additionally, Moroskoski et al. (2022) found a direct association between lethal violence against women and the percentage of mothers who served as household heads, had not completed elementary education, and had at least one child below the age of 15.

Some studies have identified uncommon factors that were not typically considered by other researchers but are related to femicides. For example, Ingram and Marchesini da Costa (2015) discovered a significant and positive association between femicides and the environmental impact of industrial development projects. Sepúlveda Murillo et al. (2018) identified a strong connection between the proportion of land dedicated to coca production and the rate of forced displacements in relation to the number of femicides. And Moroskoski et al. (2022) found that cities with female mayors and a high percentage of women councilors were associated with an increased risk of lethal violence against women, suggesting an association with the involvement of women in politics.

These factors identified in the literature will guide our research in collecting and defining the exploratory variables that will be used in our models.

3 Data

3.1 Area of study

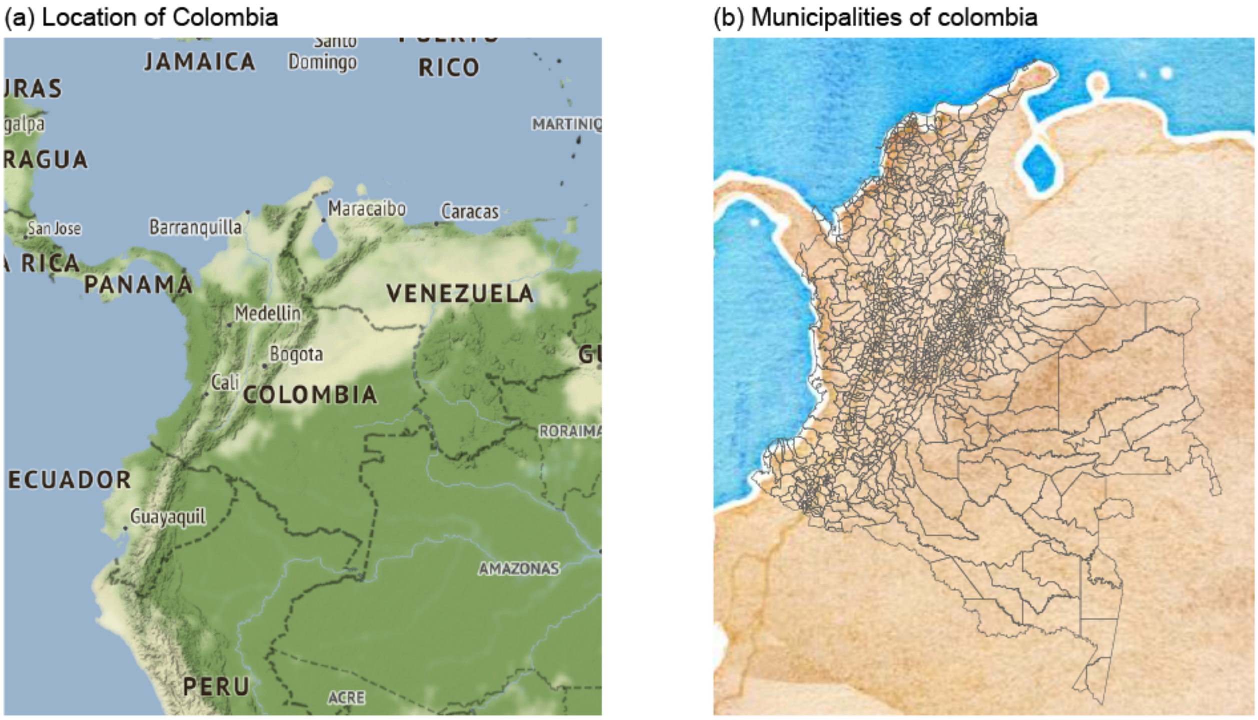

Figure 1: Geographic location of Colombia (a) and map of municipal borders of Colombia (b)

Colombia is located in the northwestern region of South America; it borders Venezuela and Brazil to the east, Peru and Ecuador to the south, Panama and the Pacific Ocean to the west, and the Caribbean Sea to the north (Figure 1-a). Its land area covers 1.141.748 square kilometers, and in 2019, it had a total population of 49 million inhabitants. Bogota is the capital of Colombia and the most populous city, followed by Medellin, Cali, and Barranquila, in that order.

Colombia was composed of 1.122 municipalities as of December 2019, but our spatial analysis will focus solely on the municipalities located in the continental area and will not include the municipalities of San Andrés and Providencia, which are actually two islands located in the Caribbean Sea, 775 km northwest of the continental Caribbean coast of Colombia and 220 km from the eastern coasts of Nicaragua. Due to their considerable distance from the rest of Colombia's municipalities, their proximity to Nicaragua, and the framework in which a GWR model operates, these municipalities will not be considered in our analysis. Therefore, our study area encompasses the 1.120 municipalities located in the continental area of Colombia.

It is worth noting that the municipalities located in the southeastern and eastern regions of Colombia have much larger land areas than the rest of the country's municipalities (Figure 1-b). Therefore, there is a higher concentration of municipalities in the central, northern, and western regions. This is because the municipalities in the southeastern and eastern regions are situated in the Amazon rainforest, which covers nearly 40% of Colombian territory and is the least populated area of the country (accounting for only 2% of Colombia's population) as it is characterized by dense forests, limited development, and extensive protection through national parks; as a result, the municipal divisions are more extensive. This geographical characteristic will be of great importance for the implementation of the GWR model.

3.2 Output variable

Our variable of interest will be femicide cases per 100.000 women. A rate based on the number of femicide victims per municipality, as reported by the Office of the Attorney General of Colombia. This data includes murdered victims (Fulfilled intentional homicide) and victims of attempted femicides. These are the official data on femicides in Colombia, following the definition established in the Rosa Elvira Cely Law for typifying female homicides as femicides.

The Rosa Elvira Cely Law provides specific criteria to be considered when categorizing cases as femicides. According to this law, femicide is defined as the killing of a woman based on her gender or gender identity, and when any of the following conditions are present: the perpetrator had a close, familial, or intimate relationship with the victim; there is a prior history of violence in the victim's domestic, familial, work, or school environment (even if not reported); acts of gender or sexual instrumentalization; the aggressor exploited power dynamics in their favor; the feminicide was committed to instill fear or humiliation; and the victim was either deprived of liberty or isolated (Congress of Colombia, 2015).

We will include attempted femicides since these are cases where the intention and actions were directed towards causing the death of the victim, and the circumstances of the incident, as well as the motivations of the aggressor, fall within the classification of gender-based homicide. These cases also encompass all the characteristics and context of a femicide; therefore, not including these cases would mean disregarding crucial information that captures the phenomenon we are studying.

Our research will only focus on femicides that occurred between 2017 and 2019 for the following reasons. Before the enactment of the Rosa Elvira Cely Law in July 2015, there was no official count of feminicides. They were often mixed in with homicide counts and their aggravating factors, which were frequently not recorded. Additionally, there was no official definition of feminicide, making it more challenging to maintain records, even for independent observatories that monitored gender-related killings of women.

Furthermore, since mid-2015, official feminicide figures began to be recorded and subsequently increased gradually as the law was being implemented and interpreted by the relevant authorities. This trend continued until 2017 when the data reached a level of stabilization, which has persisted to the present day. However, although more recent femicide data are available, the interval 2017-2019 was selected because some of the explanatory variables were only collected in the most recent Colombian population census conducted in 2018. Therefore, the proximity of 2017 and 2019 to this census makes it appropriate to use the data from the 2018 census as a proxy measure for this time interval.

We did not compute the ratio as the number of femicide cases over the population of women in the municipality (also called the raw or crude rate). The crude rate serves as an estimate for the unobservable inherent risk. In our research, this corresponds to the risk that a woman is exposed to femicide or an attempted femicide. While the crude rate provides an unbiased estimation of this risk, which is a favourable characteristic, it also exhibits an unfavourable property in terms of its variance. The variance of a raw rate is calculated as:

$$\mathrm{Var}\left(r_i\right) = \frac{\pi_i\left(1-\pi_i\right)}{P_i}$$

Where r_i is the raw rate in the municipality i, P_i is the corresponding population at risk (the total number of female inhabitants), and π_i is the underlying risk (Anselin et al., 2020a). This means that as the population size of a municipality is larger (indicated by a larger P_i), the variance of the estimator is smaller, leading to greater precision. However, in areas with sparse populations (small P_i), the risk estimate becomes less precise, resulting in a larger variance. Additionally, since population sizes varies across areas examined, the precision of each rate also varies.

This variance instability must be addressed to prevent the identification of spurious outliers and a misleading characterization of the spatial distribution of the underlying risk when using crude rates. To address this issue, smoothing techniques can be employed to enhance the precision of the crude rate by taking information from other observations (Anselin et al., 2006). In essence, larger populations lead to more precise risk estimates, while smaller populations result in less precise estimates, and smoothing methods can help correct this variance instability.

Because of the large variation in the population sizes among Colombia's municipalities, the crude femicide rate can appear disproportionately high in regions with a small population and a small number of femicides. On the other hand, it can seem artificially low in areas with a large population, even if the number of femicide cases is substantially high. Therefore, we will use smoothing rates instead of crude rates to mitigate this variance instability in our spatial data. Additionally, using smoothed rates helps reduce some of the variability in the output variable, making it more effective at highlighting broader spatial trends and identifying general features of the data (Anselin et al., 2020b).

The underlying logic of a spatial empirical bayes (EB) smoothing approach is based on the Bayesian framework, where the distribution of a random variable is adjusted following the observation of new data. Essentially, the spatial EB smoothed rate is derived by calculating a weighted average between the raw rate r_i for each municipality (the likelihood) and a prior estimate of the unknown risk parameter (which is known as the reference rate θ_i), where the weights w_i are proportional to the population at risk in each municipality (Marshall, 1991). Then, the spatial EB smoothed estimate for the risk in the municipality i is:

$$\pi_i^{\mathrm{EB}} = w_i r_i + \left(1-w_i\right)\theta_i \tag{3.1}$$

And the weights are:

$$w_i = \frac{\sigma_i^2}{\sigma_i^2 + \frac{\mu_i}{P_i}} \tag{3.2}$$

With P_i as the population at risk in municipality i, σ_i^2 is the estimate of the variance of the prior distribution, and μ_i as the mean of the prior distribution (which is equal to the reference rate θ_i) (Anselin et al., 2020b). From equations (3.1) and (3.2), we can observe that smaller municipalities will have significant adjustments in their rates, while larger municipalities will see only minimal adjustments. This approach essentially employs the municipality's population as an indicator of data reliability, with larger populations indicating greater confidence in those municipalities' rates. In that way, this approach mitigates the noise stemming from low femicide rates in smaller populations.

Here, the reference rate θ_i is computed using a spatial window around a specific municipality i. This window includes both municipality i and its neighbouring municipalities, which are identified based on the non-zero elements in the row of a binary spatial weight matrix. Therefore, the estimated reference mean (which is simply the reference rate) for the municipality i is equal to the following spatial window average, in which the weighted average is applied separately to both the numerator and denominator:

$$\mu_i = \theta_i = \frac{\sum_{j=1}^{n} w_{\text{ij}} O_j}{\sum_{j=1}^{n} w_{\text{ij}} P_j}$$

where O_j is the number of femicide cases in municipality j, P_j is the population at risk, n is the number of municipalities in Colombia (i.e., $n = 1.120$) and w_{ij} are the binary spatial weights surrounding the municipality i (where $w_{\text{ii}} = 1$ to include the municipality i in the spatial window) (Anselin et al., 2020b).

We employed a queen contiguity weight matrix to define the neighbours of the spatial window used to estimate the reference rate θ_i. The queen contiguity criterion defines neighbours as spatial units that share either a common edge (or border) or a common point (or vertex) (Anselin et al., 2020c), taking as an analogy the moves allowed by queen pieces on a chessboard. After we have established who our neighbours are spatially connected, we need to assign weights to each neighbour relationship withing each spatial window.

A spatial weight matrix is a structured collection of weight values for every possible pair of municipalities. These weights are arranged in rows w_{ij}, in such a way that each municipality i (Mun_i) has its own corresponding row of weights, and the specific values of these weights are determined by the chosen adjacency criterion (Paez, 2022a). In the binary encoding approach, we assign a value of 1 to pairs of municipalities that are adjacent, while pairs of municipalities that are not adjacent receive a value of 0. Therefore, for our queen criterion:

$$w_{\text{ij}} = \begin{cases} 1 & \text{if } Mun_i \text{ and } Mun_j \text{ share an edge or a vertex, or if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

Notice that the diagonal elements of the spatial weight matrix are equal to 1 ($w_{\text{ii}} = 1$), in that way we can include the municipality i in the spatial window.

We used the R function *EBlocal* from the *spdep* package, with the argument *geoda* = *TRUE* to estimate the spatial Empirical Bayes smoothing rate in the same manner as we have detailed. And the function *poly2nb* from the same package to create a queen contiguity weight matrix.

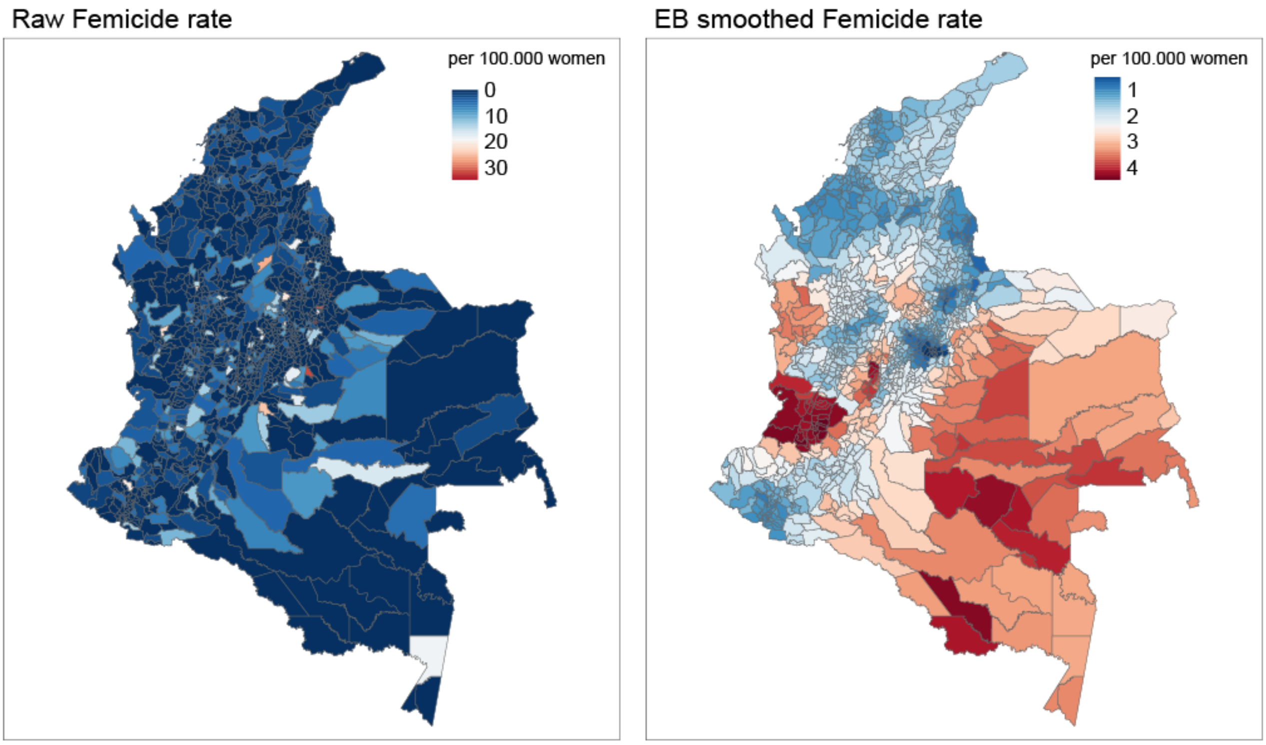

Figure 2: Spatial distribution of the raw femicide rate (left) and the smoothed Bayesian femicide rate (right) in Colombia

In Figure 2, we compare the raw femicide rate per 100,000 women with the smoothed Bayesian femicide rate per 100,000 women. Here, we can observe how the variance instability in the raw rate has produced extreme outliers, making it impossible to identify any features or spatial patterns in the rates. This also leads to a misleading characterization of the spatial distribution, as seen in the Amazon region where very large municipalities with very low population levels give the impression of either very low or very high levels of femicides; however, after applying the smoothing approach, the rates are adjusted to show broader spatial trends. This is also the case for other regions of Colombia, such as the Pacific coast, where the features were obscured by outliers and instability.

By employing this spatial Empirical Bayes smoothing technique, we have effectively mitigated the instability stemming from random rate fluctuations, particularly in municipalities characterized by small populations and few femicide victims; thus, this approach allows us to compare across different populations.

However, when examining the kernel density plot of our estimated Empirical Bayes rate per 100.000 women, we can observe that our output variable is positively skewed; therefore, we opted to use the natural logarithm of the smoothed Bayesian femicide rate as the output variable in order to make the dataset's distribution more in line with a normal distribution; and therefore, enabling the application of a GWR model under a Gaussian assumption. This approach helps prevent the small number of unusually large observations from exerting an undue impact on the sum of squared errors in the model.

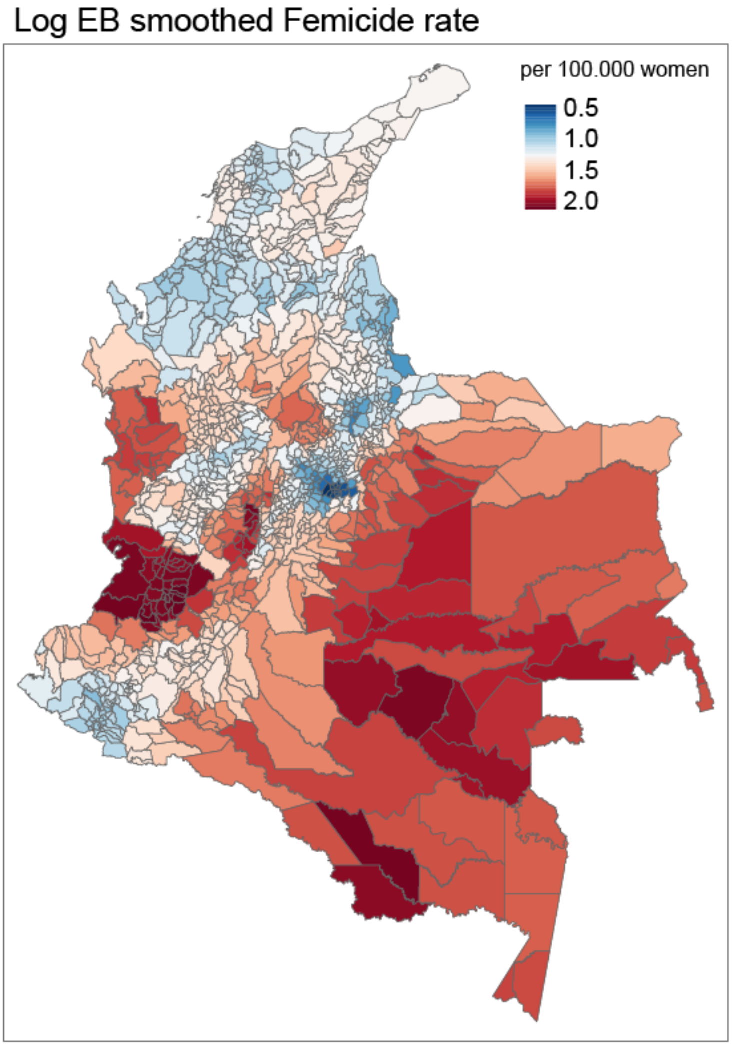

Figure 3: Spatial distribution of the natural logarithm of the smoothed Bayesian femicide rate in Colombia

Figure 3 maps the spatial distribution of our output variable. In this map, we can clearly identify spatial patterns in the femicides rates, with clusters of high rates in the southeast and east of the municipalities of Colombia as well as over the pacific coast, and high concentration of low rates, especially in the north, and in other smaller groups in the central and southwestern regions of the country.

The female population by municipality and year was obtained from the National Administrative Department of Statistics of Colombia (DANE). The municipal female population for the period 2017-2019 corresponds to the average female population of municipalities over these three years. The count of femicide victims per municipality and year was provided by the Office of the Attorney General of Colombia. The femicide cases for 2017-2019 are the average number of femicide victims by municipality during that time frame.

3.3 Explanatory variables

The following explanatory variables capture the main factors related to femicides reported in the literature review section.

Population (Total, male and female data) by municipality and year was obtained from the National Administrative Department of Statistics of Colombia (DANE). This data will be used to calculate rates. Male homicide victims, non-lethal violence against women victims, domestic violence victims and sexually assaulted women victims by municipality and year were sourced from the Colombian National Police. The rates per 100,000 inhabitants were calculated for their respective populations at risk by municipality and year, and then averaged for the period 2017-2019.

The percentage of children under 18 years of age per municipality and year was downloaded from the DANE, and it was averaged for the period 2017-2019. The teen birth rate per 1.000 adolescent women by municipality and year were collected from the DANE and averaged for the period 2017-2019. Net school coverage rate per municipality and year was obtained from the Ministry of National Education and averaged for the period 2017-2019.

We obtain a multidimensional poverty index which is composed of five dimensions (educational conditions of the household, childhood and youth conditions, health, employment and housing conditions, and access to public services) which encompass 15 specific indicators (DANE, 2020). Some of these indicators were mentioned in the literature as factors related to femicides. However, as demonstrated by Sepúlveda Murillo et al. (2018), and Ingram and Marchesini da Costa (2015), these variables are often highly correlated, making it challenging to isolate the independent effect of each variable due to multicollinearity issues. In such situations, studies by Land et al. (1990), and Rio et al. (2010) have shown that it is more appropriate to utilize an index when estimates display instability caused by significant collinearity among multiple covariates. Therefore, we will use the multidimensional poverty index instead of the individual indicators.

For the calculation of this poverty index, households with deprivation in at least 33.3% of the 15 indicators are considered to be in poverty (DANE, 2020). Therefore, the multidimensional poverty index corresponds to the percentage of households deemed to be in poverty relative to the total households in the municipality.

This multidimensional poverty index was constructed using data from the 2018 Colombian population census conducted by DANE. Additionally, data on the percentage of households headed by single mothers was extracted from the same census.

We will use the following *labels* to identify our variables:

- *Homicides*: Male homicide rates.
- *Non-lethal violence*: Women non-lethal violence rates.
- *Domestic violence*: Domestic violence rates.
- *Sexual violence*: Sexually assaulted women rates.
- *Underage*: Percentage of children under 18 years of age.
- *Teen births*: The teen birth rates.
- *School coverage*: Net school coverage rates.
- *Poverty*: Multidimensional poverty index (%).

- *Single mothers*: Percentage of households headed by single mothers.
- *Femicides*: Log EB smoothed Femicide rate

3.4 Exploratory data analysis

Collinearity

Initially, we will examine the correlation matrix of the explanatory variables as a preliminary step to identify collinearity. We will consider any element in this matrix with an absolute value greater than or equal to 0.6 as a strong indication of highly correlated variables, signifying a collinearity issue between those pairs of variables. The Spearman correlation will be calculated for each pair of variables because this nonparametric measure is less influenced by extreme values or outliers (Kim et al., 2015). Additionally, the Spearman correlation can capture not only linear relationships but also monotonic relationships between pairs of variables (i.e. variables that tend to move in the same relative direction, though not necessarily at a constant rate) in a more reliable way than the Pearson correlation.

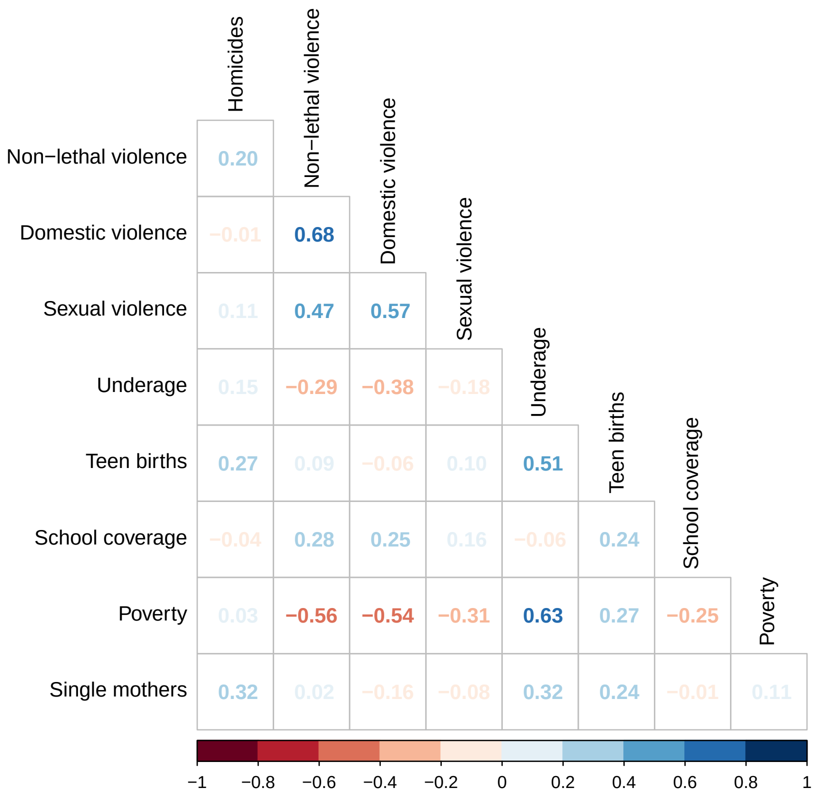

Figure 4: Correlation matrix between the explanatory variables

Figure 4 displays the correlation matrix without the principal diagonal and the upper triangular matrix. The colour scheme represents the direction of the relationship, with dark red (blue) denoting a negative (positive) correlation. Lighter colours indicate low correlation values, in this way we can better appreciate the most worrying correlation values.

We note a strong positive correlation between the poverty index and the proportion of children under 18 years old, as well as between rates of domestic violence and non-lethal violence against women. Additionally, we've identified three pairs of variables showing moderate signs of correlation, suggesting potential collinearity, such as the positive relation between rates of domestic violence and sexually assaulted women, and the negative correlation of the poverty index with respect to the rate of non-lethal violence and domestic violence, respectively. These last two relationships are unexpected and contradict what was anticipated based on the literature review.

Nonetheless, a limitation of correlation coefficient is that they assess the relationship between two variables without considering others; thus, it is plausible for collinearity to be present among three or more variables even if no specific pair of variables has a notably high correlation (Belsley and Welsch, 2005). Therefore, rather than examining the correlation matrix, a more effective approach to detect multicollinearity is to calculate the variance inflation factor (VIF).

The VIF provides a measure of how much the variance of a covariate is inflated due to correlations with other explanatory variables in the model (Simon and Young, 2018). It relies on the R^2 statistic obtained by regressing a covariate on the remaining variables. When a variable exhibits a significant linear relationship with one or more other variables, the R^2 approaches 1, resulting in a high VIF for that variable and suggesting that the presence of this feature may be redundant in the model. We will calculate the variance inflation factors for each explanatory variable in our model using the R function *vif* from the *car* package.

Variable	**VIF**
Homicides	1.1610
Non-lethal violence	1.8849
Domestic violence	1.8054
Sexual violence	1.3380
Underage	2.4017
Teen births	1.5616
School coverage	1.4378
Poverty	2.8774
Single mothers	1.1568

Table 1: Variance inflation factor (VIF) of the explanatory variables

According to Belsley and Welsch (2005), a VIF exceeding 10 indicates concerning levels of multicollinearity. However, none of our variables exhibit a VIF greater than 2.9 (table 1), implying a low risk of multicollinearity among the femicide-related factors. However, the correlation matrix suggests that poverty and domestic violence seem to capture similar phenomena and display correlation with multiple other explanatory variables; thus, we plan to omit them from our models to prevent potential collinearity concerns related to the use of GWR models, we will detail this topic in the next section.

Global Moran's Index

It is prudent to assess the existence of spatial autocorrelation within our variables before employing a GWR model. Spatial autocorrelation quantifies the level of correlation across space (Cliff and Ord, 1973, 1981). In essence, a variable exhibits spatial autocorrelation when the observed values in one municipality are influenced by the values observed in neighbouring municipalities.

Then, we will use the Global Moran's index, often referred to as Moran's I, to assess the presence of spatial autocorrelation in our data (Moran, 1948). This index is considered a global measure as it takes into account the overall level of spatial autocorrelation across all observations (Pebesma and Bivand, 2023). Its values range between -1 and 1, where positive spatial autocorrelation signifies that similar values tend to be situated near each other in space, while negative spatial autocorrelation indicates that dissimilar values are in close proximity (dispersed values). When the values approach zero, it suggests a lack of significant spatial autocorrelation between the values of each municipality and their neighbours (Moran, 1950), and that implies that the spatial pattern is random. In essence, Moran's I assists in determining the presence of clustering within the study area.

In order to compute the Moran's I, we need an adjacency criterion to determine how neighbourhoods are spatially connected to one another. We are going to use a queen contiguity to create a spatial weights matrix in the same way as we explained above. And this matrix will be used by the Moran's index to measure the spatial autocorrelation in our data. The formula for Moran's I is essentially an extension of the formula used to calculate the correlation coefficient in which the binary spatial weights matrix is just added to the expression. In essence, it is a statistical measure that quantifies the cross-product between a variable and its spatial lag, and this variable is represented as deviations from its mean. Therefore, the Moran's I statistic can be expressed as follows:

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{\text{ij}}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{\text{ij}} \left(x_i - \overline{x}\right)\left(x_j - \overline{x}\right)}{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}$$

Where, $\overline{x}$ is the mean of the variable x, n is the number of observations, w_{ij} are the elements of the spatial weights matrix, and $\sum_{i=1}^{n}\sum_{j=1}^{n} w_{\text{ij}}$ is the sum of all of the weights.

However, spatially smoothed rates are not suitable for examining spatial autocorrelation because they are inherently autocorrelated by construction (Anselin et al., 2020b). Therefore, we will only calculate the Global Moran's Index for the explanatory variables. If these variables exhibit spatial autocorrelation, it suggests the presence of clusters in their spatial distribution. This is a desirable characteristic for our data because it indicates the existence of local patterns that could influence the relationship between these explanatory variables and the femicide rate.

The previously mentioned Moran's I statistic is helpful in indicating whether a pattern appears random, but we require a more formal criterion to determine this. The most commonly used test for assessing spatial autocorrelation is the Global Moran's I test (Brazil, 2023). The inference for the Moran's index is based on a null hypothesis of spatial randomness; therefore, the expected value of Moran's I under this null hypothesis, along with its variance, are already known. Then, the distribution of the statistic under the null hypothesis can be determined through randomization, where each value is equally likely to occur at any location (Anselin et al., 2020d), and hypothesis testing involves comparing the actual statistic to its distribution under the null hypothesis of spatial independence (Paez, 2022b).

This test for autocorrelation based on the Moran's I is implemented in the *spdep* package with the command *moran.test*. This function also returns the Moran's I statistic.

Variable	Moran's I	p-value
Homicides	0.5487	0.00000
Non-lethal violence	0.3555	0.00015
Domestic violence	0.3395	0.00100
Sexual violence	0.3188	0.00090
Underage	0.7563	0.00020
Teen births	0.4070	0.00100
School coverage	0.3029	0.00100
Poverty	0.6613	0.00000
Single mothers	0.3518	0.00050

Table 2: Global Moran's I test for each of the explanatory variables

O'Sullivan and Unwin (2010) considers spatial autocorrelation to be relevant when it falls above 0.3 or below -0.3 as a general guideline. Given that all the covariates have positive Moran's I statistic values exceeding 0.3 (table 2), this indicates a non-random spatial pattern of similar values across all our explanatory variables. Moreover, the extremely low p-values provide strong evidence for rejecting the null hypothesis with a high level of confidence, establishing that these spatial autocorrelations are statistically significant for each factor.

Spatial distribution

By examining the spatial distribution of our explanatory variables, it becomes evident that there are spatial clusters within all our features (Figures 5, 6 and 7).

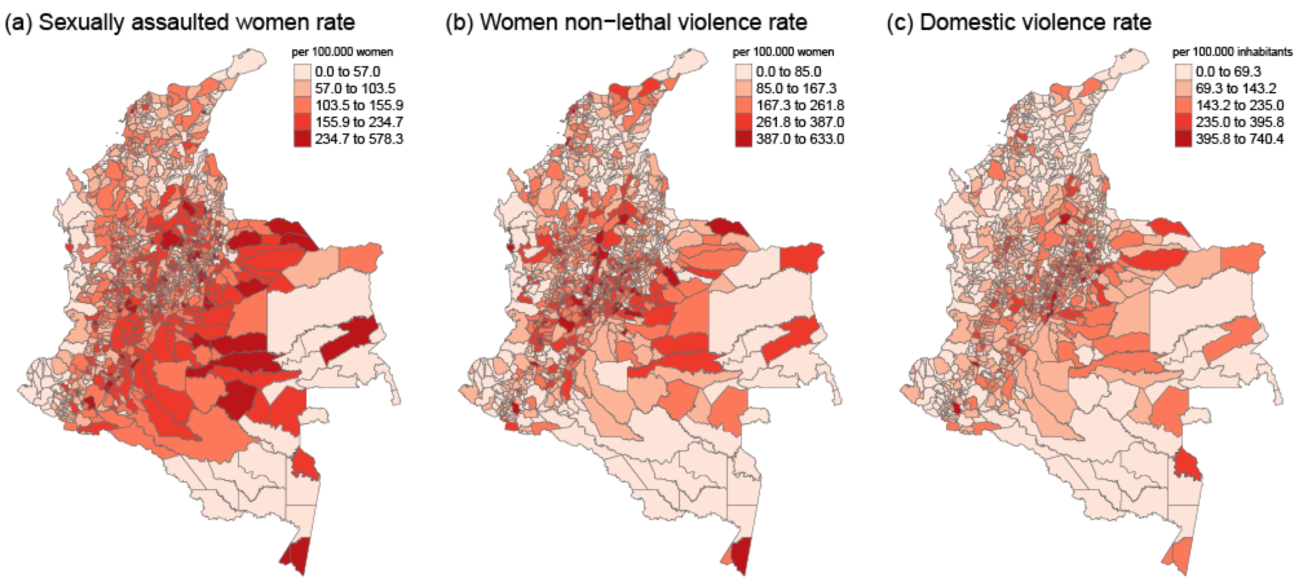

Figure 5: Spatial distribution of the rate of sexually assaulted women (a), the rate of non-lethal violence against women (b), and the rate of domestic violence (c)

The rate of sexual violence against women shows a high concentration of rates over a diagonal area crossing from southwest to northeast and clusters of low rates towards the southeast and the Pacific coast (Figure 5-a). The rates of non-lethal violence against women and domestic violence show a greater dispersion of rates, making it more difficult to identify any spatial patterns (Figures 5-b and 5-c). However, a clear concentration of low rates is evident in the southeast of the country in the Amazon region, and a higher concentration of high rates is seen towards the centre of Colombia. It is also very clear that both rates show similar patterns across the country, in line with the Spearman correlation coefficient that identified a very high positive relationship between the two variables.

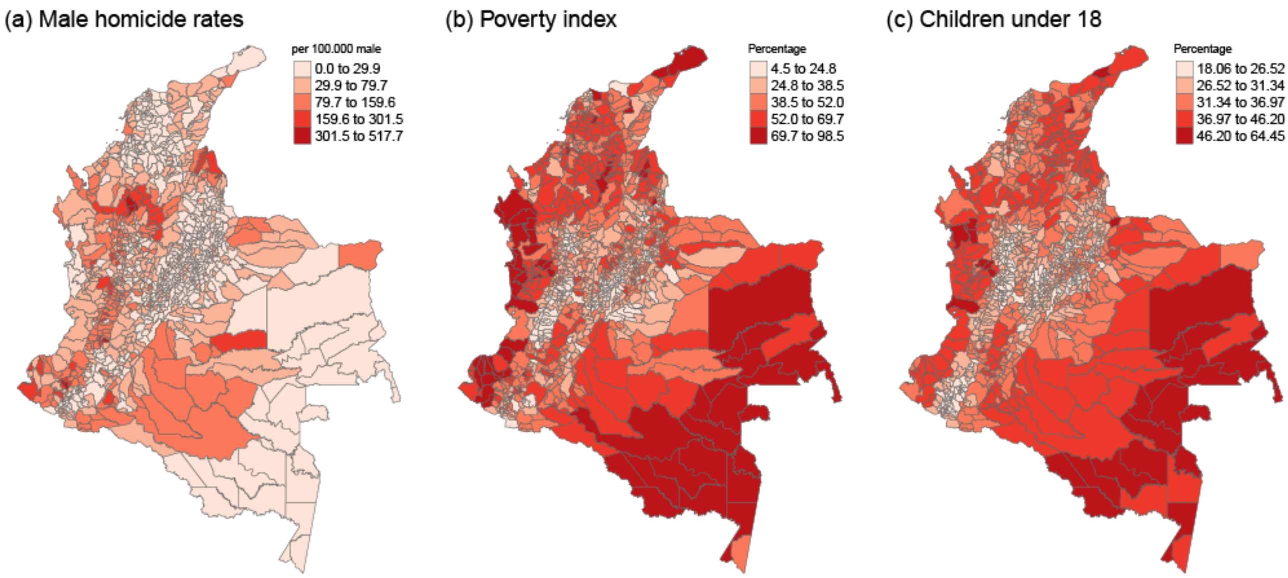

Figure 6: Spatial distribution of the male homicide rate (a), the multidimensional poverty index (b), and the percentage of children under 18 years of age (c)

On the other hand, the male homicide rate shows a clear concentration of low rates in the southeast and another group towards the centre of the country; similarly, clusters of high rates can be observed in the southwest and northwest (Figure 6-a). Figures 6-b and 6-c show a clear spatial clustering pattern in the poverty index and the percentage of children under 18, with very high percentages on the periphery of the country and low percentages in the centre. This positive relationship was confirmed by the Spearman correlation coefficient. Additionally, the negative relationship of the poverty index with respect to the rate of non-lethal violence and domestic violence can be observed in these maps (Figures 6-b, 5-b and 5-c), where there are higher concentrations of violence rates towards the centre, in contrast to the low poverty rates towards the centre of the country.

Figure 7-a shows a cluster of high teenage birth rates towards the northern part of the country, just like the net school coverage rate (Figure 7-b). However, the latter exhibits a clearer pattern of low school coverage towards the southeast areas. Finally, the percentage of single mothers heading households has a very high concentration of high rates along the Pacific and Caribbean coasts, and low proportions in groups of municipalities located towards the central and southeastern parts of the country (Figure 7-c).

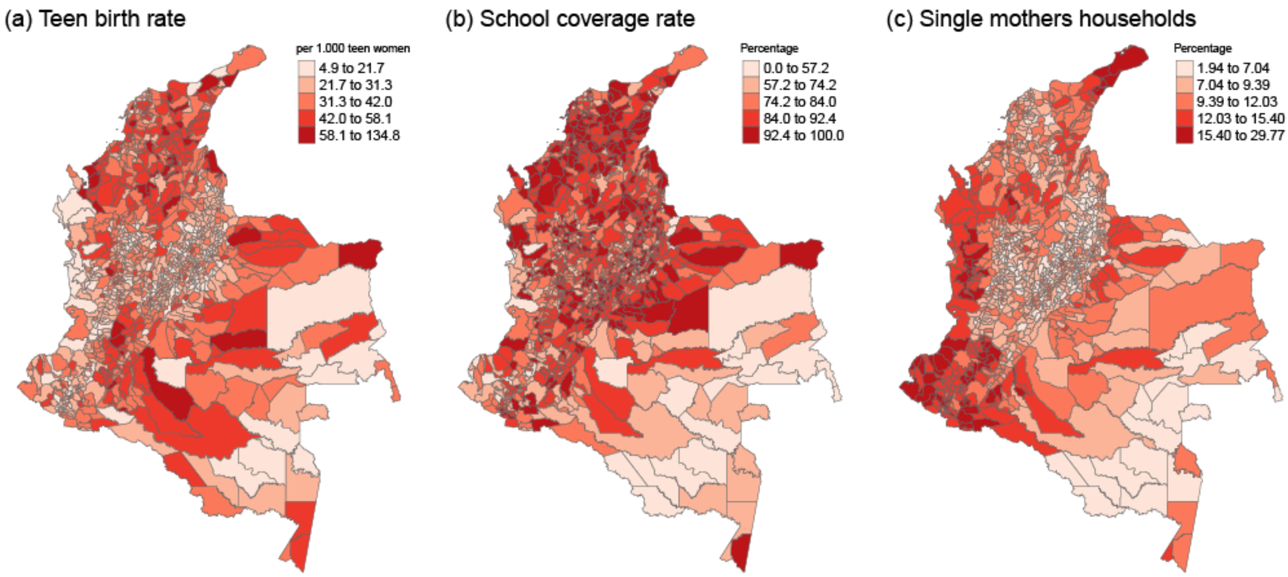

Figure 7: Spatial distribution of the adolescent birth rate (a), the net school coverage rate (b), and the percentage of households headed by single mothers (c)

4 Methods and results

As stated in the introduction, we will employ a Geographically Weighted Regression (GWR) model to investigate the relationship between the smoothed femicide rate per 100,000 women and the factors that are related to this phenomenon in Colombia. The GWR model is designed to capture spatial non-stationary relationships present in the data; however, to assess the presence of spatial non-stationarity in the relationships, we also need to apply a standard Ordinary Least Squares (OLS) regression model to our dataset.

Spatial non-stationarity refers to variations in the relationship between a response variable and its covariates across geographical areas (Rowe and Arribas-Bel, 2023a). In a standard OLS model, it is assumed that the intercept and the strength of the relationship between the response variable and the explanatory variables remain constant across the entire study area (a spatially stationary process), so it estimates global parameters for the model; instead, in a GWR model, both intercepts and slopes can vary across space. This approach aims to account for underlying contextual factors that may systematically differ across geographical areas, impacting the relationship between the variable of interest and the explanatory factors.

4.1 Ordinary least squares Regression model

We are going to start exploring our data using an ordinary least squares (OLS) linear regression model, and later, compare it with a geographically weighted regression model to analyse if smoothed femicide rates are linked to structural differences across municipalities in Colombia.

Based on our exploratory data analysis, the explanatory variables in the model are: male homicide rates, rates of non-lethal violence against women, rates of sexually assaulted women, percentage of children under 18 years of age, teen birth rates, net school coverage rates, and percentage of households headed by single mothers. Then, our proposed OLS linear regression model is as follows:

$$f = \beta_0 + \sum_p \beta_p x_p + \varepsilon \tag{4.1}$$

Where:

$\varepsilon \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2)$

f is the Log EB smoothed Femicide rate.

β_0 is the intercept.

β_p is the coefficient of the variable p.

x_p is the variable p.

$p \in \left\{ \begin{array}{l} \textit{Homicides, Non-lethal violence, Sexual violence, Underage,} \\ \textit{Teen births, School coverage, Single mothers.} \end{array} \right\}$

4.2 OLS results

Variable	**Estimate**	**Std. Error**	**t value**	**p-value**
Intercept	0.9762	0.0865	11.274	0.0000
Homicides	0.0011	0.0002	6.904	0.0000
Non-lethal violence	0.0007	0.0001	8.333	0.0000
Sexual violence	0.0008	0.0001	6.275	0.0000
Underage	0.0184	0.0015	11.976	0.0000
Teen births	-0.0040	0.0008	-5.033	0.0000
School coverage	-0.0040	0.0008	-5.150	0.0000
Single mothers	-0.0004	0.0027	-0.153	0.8783
R^2	0.2464			
AIC	234.1302			

Table 3: Results of the OLS model for *Femicides*.

Table 3 presents the output of the OLS model. It is important to note that the p-value for the percentage of households headed by single mothers is not statistically different from zero at any significance level in the presence of the other explanatory variables. This implies no association. Instead, all the other explanatory variables have very small p-values associated with each predictor's t-statistic, given the other variables used in the regression model; therefore, these are the factors that appear to have a statistically significant relationship with the variable of interest.

The results of the linear model indicate that the three violence rates included in our model and the proportion of children under 18 years of age are positive associated with the femicide rate in Colombia, considering the other variables in the model. On the contrary, there is a negative relationship between the teen birth and net school coverage rates with respect to the smoothed rate of femicides per 100.000 women in Colombia, after controlling for the other variables.

The R^2 value of the model implies that approximately a quarter of the variability in the smoothed femicide rates can be explained using this linear model; as a result, we can say that this simple OLS model seems to fit the data poorly as it explains only a small proportion of the variance in the response variable. Some of this unexplained variance might be due to the omission of other important factors related to femicides, or it could be because we did not use the appropriate model for our data, since an OLS regression model assumes that the relationships in the model are constant over space, so it only captures global relationships. However, as we propose, relationships may vary over space, and if such variations in associations exist over space, our estimated global OLS model will not accurately represent reality since it assumes these relationships to be constant.

Spatial autocorrelation analysis of the residuals

We can appreciate this misspecification of our global linear model by just mapping the residuals of the global OLS model (figure 8); in this map, we will use red to indicate positive residuals (when the model underestimates) and blue for negative residuals (when the model overestimates). If the pattern of overestimation and underestimation is random (indicating random residuals), it would imply that the model effectively captured all systematic patterns in the data, and if the pattern is non-random, that means a violation of the assumption of independent residuals (uncorrelated errors) (Paez, 2022c; Leung et al., 2000).

In figure 8 (below), we can observe a non-random spatial distribution of the residuals across Colombia, with a clear residual systematic pattern of clusters across the municipalities almost similar to the spatial patterns of the Log EB smoothed Femicide rates observed in figure 3. This spatial autocorrelation in the error terms might be an indication that the global OLS model is not capturing the non-stationary relationships of our data.

Additionally, we can formally test for spatial autocorrelation in the residuals using the Global Moran's I test in the same form in which we specified it previously. Given the Moran's I statistic value is greater than 0.3 and the p-value is extremely low (table 4), we can reject the null hypothesis and confirm that the residuals of the global OLS model are not random.

	Moran's I	**p-value**
OLS residuals	0.7279	0.00001

Table 4: Global Moran's I test for the OLS model residuals

As previously stated, spatial autocorrelation violates a fundamental assumption of linear regression. And it typically occurs when the model is not accurately specified, perhaps due to an incorrect functional form or the omission of relevant covariates.

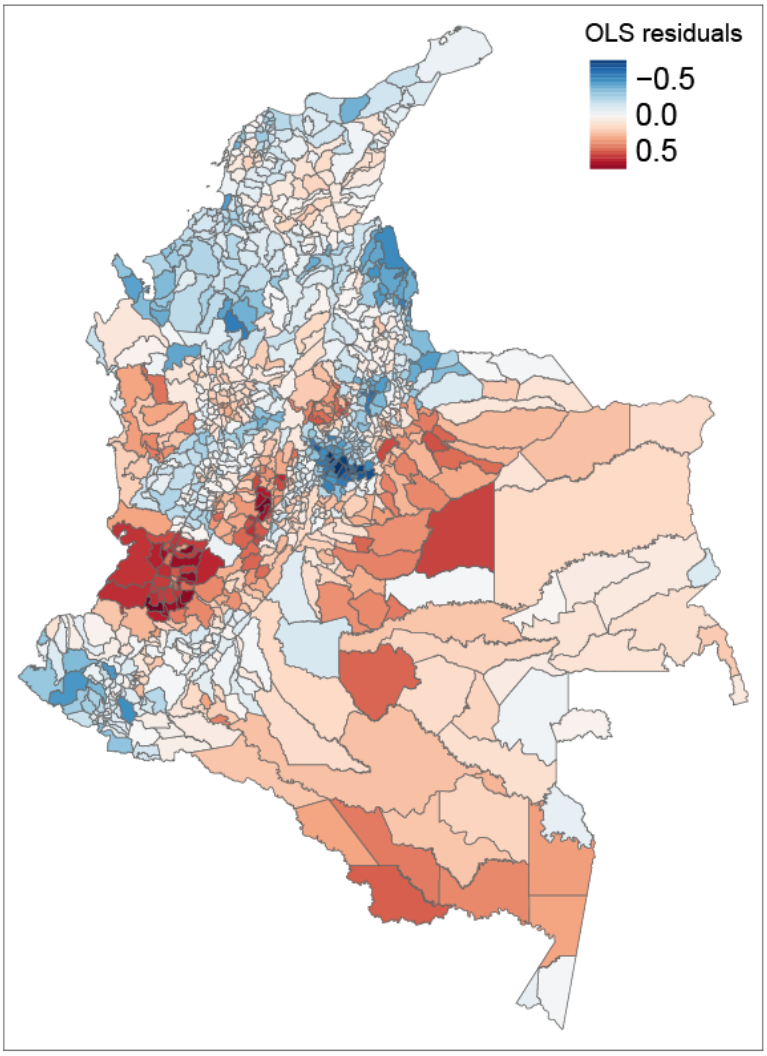

Figure 8: Spatial distribution of the OLS model residuals

4.3 Geographically weighted regression model

Given we believe there is spatial non-stationarity in our data, we are going to explore our data using a geographically weighted regression (GWR) model to understand whether femicide rates are related to structural differences across municipalities in Colombia. GWR is the appropriate exploratory technique to study how the relationships between the output variable and explanatory variables may vary across space as it generates coefficients that change spatially; thus, it allows for distinct relationships to be present in different spatial locations (Brunsdon et al., 1996).

The GWR technique is a preferable choice over a linear regression approach when dealing with spatial heterogeneity since it allows both the values and significance of a model to vary across space in a continuous manner (De Bellefon and Floch, 2018). Moreover, a local GWR regression has the advantage of reducing spatial autocorrelation in residuals compared to using a global estimator, as highlighted by Bates and Pryce (2014). As a result, GWR models overcome the limitation of the OLS regression model, which generates global (or constant) estimates for the entire study area. Instead, GWR can effectively capture the spatially correlated relationships between independent variables and response variables.

Broadly, a GWR model works by running a local regression for each observation (known as the regression point) within our study area, which, in our case, corresponds to each municipality. Each local regression involves utilizing the data from the regression point and its neighbouring points, but the weighting of these neighbours is determined by their proximity to the regression point, giving more weight to data points closer to the regression point compared to those farther away (Rowe and Arribas-Bel, 2023b). With this approach, we can obtain a distinct set of coefficients for each regression point.

We are going to apply a geographically weighted regression assuming a Gaussian noise distribution which implies that we will fit local OLS models at each municipality of our data. Following the notation of Fotheringham et al. (2002), our proposed GWR model is as follow:

$$f_i = \beta_{i,0} + \sum_p \beta_{i,p} x_{i,p} + \varepsilon_i \tag{4.2}$$

Where:

ε_i is the random error at municipality i.

$\varepsilon_i \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2)$

f_i is the Log EB smoothed Femicide rate at municipality i.

$\beta_{i,0}$ is the intercept of the regression model fitted at municipality i.

$\beta_{i,p}$ is the local regression coefficient for the variable p at municipality i.

$x_{i,p}$ is the variable p at municipality i.

$p \in \left\{ \begin{matrix} \textit{Homicides, Non-lethal violence, Sexual violence, Underage,} \\ \textit{Teen births, School coverage, Single mothers.} \end{matrix} \right\}$

And each municipality i is identified by its geographical coordinates (latitude and longitude). It is important to highlight that, under our Gaussian assumption, the ε_i's represent random error terms that are independently normally distributed with zero mean and common variance σ^2. With this functional form of the GWR model, a local regression is fitted at each municipality i in which the data near to each municipality i holds more influence in the regression than data that is farther away by weighting observations according to the application of a distance-decay function (Paul Harris and Juggins, 2010).

By comparing equation (4.1) and (4.2), it is evident that GWR is just an expansion of the traditional regression approach by enabling the estimation of local rather than global parameters (Fotheringham et al., 2002). Precisely, the power of a GWR method is that it applies a geographical weighting on the data points involved in each local regression equation. This means that data points situated at a distance from the regression point receive a reduced weight and thus have less impact on the regression results for the variable of interest. Conversely, data points in closer proximity carry more weight within the regression equation.

Hence, the coefficients for each local regressions are computed by weighted least squares. This entails utilizing a diagonal spatial weights matrix of size $n \times n$, where n represents the number of municipalities in Colombia, and the main diagonal of this matrix comprises the weights assigned to each observation. In matrix notation, the GWR model's parameters are estimated as outlined by Fotheringham et al. (2002):

$$\widehat{\beta_i} = \left(\mathbf{X}^T\mathbf{W}_i\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{W}_i\mathbf{F}$$

Where the bold type denotes a matrix, so that:

$$\beta_i = \begin{bmatrix} \beta_{i,0} \\ \beta_{i,\text{ Homicides}} \\ \dots \\ \beta_{i,\text{ Single mothers}} \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{1,\text{Homicides}} & \dots & x_{1,\text{Single mothers}} \\ 1 & x_{2,\text{Homicides}} & \dots & x_{2,\text{Single mothers}} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n,\text{Homicides}} & \dots & x_{n,\text{Single mothers}} \end{bmatrix}, \mathbf{F} = \begin{bmatrix} f_1 \\ f_2 \\ \dots \\ f_n \end{bmatrix}$$

And $\mathbf{W}_i$ is an n by n spatial weighting matrix whose off-diagonal elements are zero and whose diagonal elements w_{ij} denote the geographical weighting for each of the n municipalities in Colombia with respect to regression point i (Fotheringham et al., 2002), such that:

$$\mathbf{W}_i = \begin{bmatrix} w_{i1} & 0 & \dots & 0 \\ 0 & w_{i2} & 0 & \dots \\ \dots & 0 & \dots & 0 \\ 0 & \dots & 0 & w_{\text{in}} \end{bmatrix}$$

Where w_{ij} is the weight given to data point j in relation to the regression point i. The purpose of the weight matrix is to assign more value to observations in proximity to municipality i, aligning with Tobler's first law of geography which states that nearby observations have a stronger influence on each other compared to distant ones (Tobler, 1970).

The weights w_{ij} are established through a kernel density function (also known as a spatial weighting function), which is a distance decay function that dictates how rapidly the weights diminish with growing distances. So that, for a specific regression point, a data point's weight is highest at the regression point's location, but gradually decreases as the distance between the regression point and other data points increases (see Figure 9). Consequently, when fitting a local regression model by shifting the regression point across the study area, the data are weighted differently for each location; as a result, the resultant estimates are unique for each particular location (Rowe and Arribas-Bel, 2023b).

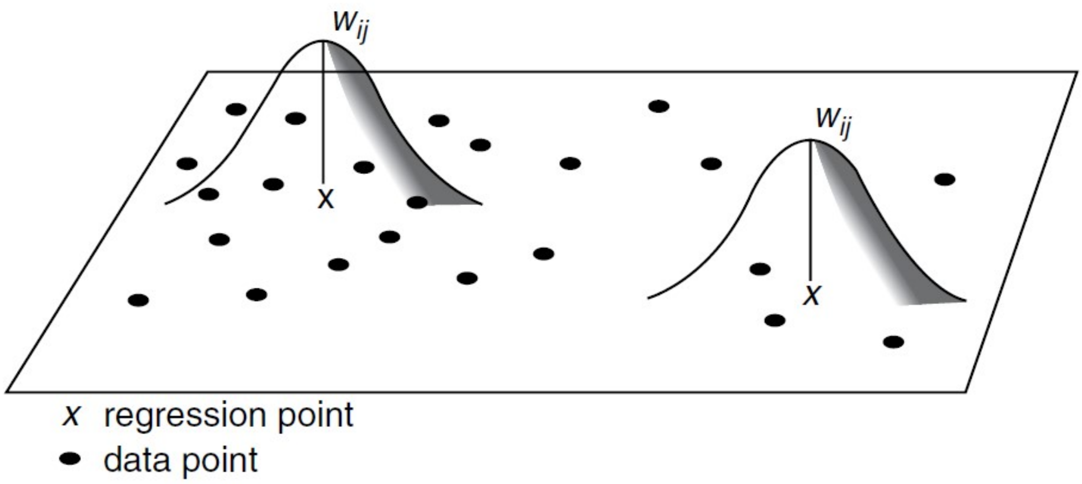

Figure 9: GWR with a spatial kernel. Source: Fotheringham et al. (2002).

We can categorize kernel functions into two main types: continuous kernels and kernels with compact support. Continuous kernels assign weights to all observations across the study area, while kernels with compact support allocate a nonzero weight to observations within a specified distance and a zero weight beyond that distance. However, research has shown that the kernel's shape causes only small changes in the resulting estimates (Brunsdont et al., 1998). In our analysis, we will opt for a Gaussian weighting function, which is a continuous kernel widely utilized in the literature (Brazil, 2019; Leung et al., 2000) and is defined as:

$$w_{\mathrm{ij}} = exp\left[-\frac{1}{2}\left(\frac{d_{\mathrm{ij}}}{b}\right)^{2}\right]$$

Where d_{ij} represents the distance between observation point j and regression point i, and b is the kernel bandwidth, which determines the set of observations to incorporate in the local regression and, thus, controls the level of smoothing in the model. The Gaussian kernel function assigns a weight of 1 to the regression point, and smoothly decreases the weights for neighbouring data points j as their distance from the regression point i increases. Essentially, this function ensures that for a specific municipality i, those in closer proximity carry more weight and consequently exert a stronger influence on the estimated parameters than those situated farther away.

A relevant characteristic of the Gaussian weighting scheme is that it never reaches zero; although, the weights for data points considerably distant from the regression point tend to be very small and exert minimal influence on the regression. Then, we choose this spatial kernel because this property ensures that each regression point will have numerous neighbours, enhancing the likelihood of having variation in the values among those neighbours. This aspect aids in mitigating a common issue in the geographically weighted regression method known as local collinearity.

Selecting a weighting function also requires choosing an optimal bandwidth b. This decision involves considering whether to use a fixed kernel or an adaptive kernel. The fixed kernel employs a fixed bandwidth to define a region around all regression points as depicted in Figure 9, in which the kernel's extent is predetermined by a fixed distance from a specific regression point. On the other hand, the adaptive kernel utilizes varying bandwidths to define regions around regression points, as shown in Figure 10; here, the kernel's extent is defined by the number of nearest neighbours relative to a given regression point, and that implies that the kernels have wider bandwidths in regions where the data is sparser (Rowe and Arribas-Bel, 2023b). Essentially, this implies that the bandwidth is a function of the k-nearest neighbours of a regression point.

However, the use of fixed kernels can lead to local regressions for small spatial units being calibrated on numerous dissimilar areas, while local regressions for large areas may be based on a very limited number of data points, resulting in estimates with significant standard errors, and in extreme cases, it may even be unfeasible to generate estimates due to inadequate variation in small samples (Rowe and Arribas-Bel, 2023b). To address these problems in study areas with spatial irregularities, adaptive spatial kernels are employed, since these kernels adapt their size based on variations in the density of the data, ensuring larger bandwidths in areas where data are sparse and smaller bandwidths where data are abundant.

Hence, we will use an adaptive kernel to ensure that the GWR model adjusts more effectively to the spatial irregularities seen in Colombia's municipalities, as we have observed that municipalities in the Amazon rainforest have considerably larger areas compared to other municipalities in the country. In that way, the kernel's search window will vary in size based on the extent of each municipality, allowing us to include an equal number of municipalities within the kernels.

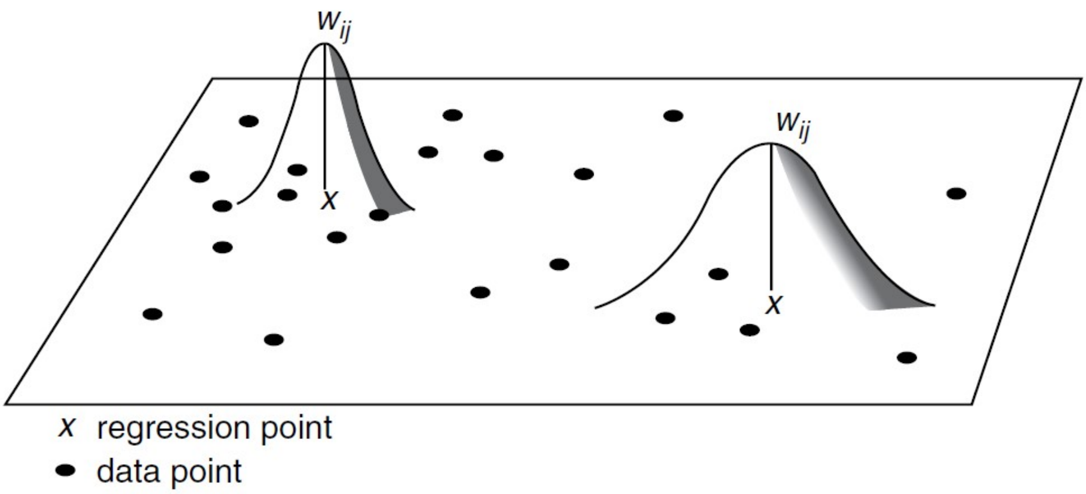

Figure 10: GWR with adaptive kernel. Source: Fotheringham et al. (2002).

Finally, to determine the optimal bandwidth, we implement a leave-one-out cross-validation method in which, for each municipality, we fit a local regression using various bandwidth values within a defined range; however, we exclude the data of the regression point from the regression for each bandwidth. Then, we calculate the cross-validation score using the resulting predicted value for each bandwidth as follows:

$$\mathrm{CV}(s) = \sum_{i=1}^{n} \left(f_i - \widehat{f}_{\neq i}(s) \right)^2$$

Where n is the number of municipalities. So, the cross-validation score is the summation of squared deviations between the actual f value at municipality i and the predicted value $\widehat{f}_{\neq i}(s)$. Where $\widehat{f}_{\neq i}(s)$ corresponds to the model's prediction generated using kernel bandwidth s after excluding municipality i from the sample (thus assigning zero weight to the regression point) (Páez et al., 2011). Finally, we compare the cross-validation scores associated with each bandwidth and select the bandwidth yielding the lowest score. This method allows us to determine a bandwidth that achieves an optimal bias-variance trade-off for our estimates (Rowe and Arribas-Bel, 2023b).

There are some concerns associated with the use of GWR models. Wheeler and Tiefelsdorf (2005) warn that collinearity might exist in subsets of the data used to estimate local coefficients, even if it is not observed in the global model, or it could even be induced in GWR applications. Páez et al. (2011) have indicated that when collinearity is present, GWR methods might detect patterns in coefficients that do not really exist. Since collinearity tends to be problematic in GWR models, it is important to be cautious when evaluating potential collinearity issues in our data. Although the VIF diagnostic during the exploratory data analysis did not reveal serious multicollinearity, the correlation matrix did indicate some signs of collinearity among certain pairs of explanatory variables. Considering this cautionary note, we opted to exclude poverty and domestic violence from our models.

Another warning, highlighted by Páez et al. (2011), pertains to the use of GWR with small sample sizes. They advise utilizing samples larger than 1.000 and advise against samples smaller than 160, but given that our data constitutes 1.120 municipalities, we can disregard this caution.

We will use the Akaike information criterion (AIC), R^2, and Moran's I statistic of the residuals to assess and compare the performance of our models. And we are going to examine the residuals using the Global Moran's Index to determine whether the spatial autocorrelation was eliminated after the implementation of the GWR model.

We are going to employ the R package *spgwr* which was designed specifically to fit geographically weighted regression models. Within this package, we will use the *gwr.sel* function to find the optimal bandwidth using a cross-validation approach by setting *method = "cv"*, with the argument *gweight = gwr.Gauss* to specify the use of a Gaussian kernel function and *adapt = TRUE* for an adaptive bandwidth. Subsequently, to estimate our proposed GWR model, we will use the *gwr* function in which the optimal bandwidth found above is used as an input in the argument *bandwidth*, and here, we will also set the argument *gweight = gwr.Gauss* to use a Gaussian kernel function.

4.4 GWR results

Variable	**Min.**	**1st Q.**	**Median**	**3rd Q.**	**Max.**
Intercept	-1.8250	0.7715	1.1414	1.4488	3.7990
Homicides	-0.0076	-0.0003	0.0004	0.0013	0.0096
Non-lethal violence	-0.0016	-0.0002	0.0000	0.0004	0.0016
Sexual violence	-0.0023	-0.0002	0.0001	0.0005	0.0023
Underage	-0.0770	-0.0016	0.0053	0.0149	0.0699
Teen births	-0.0234	-0.0024	-0.0005	0.0023	0.0123
School coverage	-0.0145	-0.0023	0.0000	0.0021	0.0148
Single mothers	-0.0702	-0.0106	0.0027	0.0168	0.0505
Adaptive bandwidth	0.0602				
Quasi-global R^2	0.8547				
AIC	-1345.14				

Table 5: Results of the GWR model for *Femicides*.

First, we can observe in table 5 that the optimal bandwidth is 0.06 which indicates the portion of observations (or k-nearest neighbors) that were included within the kernel function. Specifically, for a given municipality, this means that 6% of its nearest neighbours (67 municipalities) were used to fit the respective local regression. The GWR model estimates regressions for all 1.120 municipalities, resulting in 1.120 coefficients for each explanatory variable and the intercept; then, these estimated coefficients in each local regression for each municipality are summarized in table 5 to show how the estimated coefficients vary across the 1.120 municipalities of Colombia. The minimum, 25th percentile, median, 75th percentile, and maximum values were reported to provide a summary of the coefficient's distribution for each variable.

GWR is an exploratory technique primarily intended to identify areas of non-stationarity (Fotheringham et al., 2002), so our focus is not on interpreting the estimated coefficient values, but rather on understanding the strength and direction of the relationships in a spatial dimension. However, in table 5, it is important to highlight the highly variable relationships in all explanatory variables in relation to the smoothed femicide rates, with coefficients ranging from negative to positive associations across municipalities for all the explanatory variable.

Comparing OLS and GWR models

Additionally, at the bottom of table 5, several measures of model performance are presented. The Quasi-global R^2 serves as an approximation of the overall R^2 of the model, calculated in a similar manner as the R^2 statistic, but based on the sum of each local residual sum of squares (RSS). It is evident that the Quasi-global R^2 is considerably high (0.8547), indicating a high level of in-sample prediction accuracy and a notably superior fit of the data compared to the R^2 value of the OLS model (0.2464). Moreover, by comparing the AIC of the global OLS regression model (234.13) with the AIC of the GWR model, we can also confirm that the latter fits our data much better.

However, we require a more formal approach to assess the difference between OLS and GWR models within an inferential framework. Hastie and Tibshirani (1990) and Cleveland (1979) propose that the distribution of the OLS RSS divided by the effective number of parameters can be approximated by a X^2 distribution with effective degrees of freedom equivalent to the effective number of parameters. Based on this, Fotheringham et al. (2002) propose a test to compare the abilities of the GWR and global models to replicate the observed data, and this just involves dividing the residual sum of squares for the standard OLS model by that for the GWR model, and subsequently conducting an F-test on this ratio.

In this test, the null hypothesis suggests no difference between the GWR and OLS model (implying a ratio of 1); thus, if the test statistic is statistically significant, it indicates that GWR significantly improves the ability to match observed values compared to OLS. After conducting this test (as shown in table 6), the very small p-value leads us to the conclusion that our GWR model is preferable than the OLS linear model as it demonstrates a substantial improvement in explanatory power over an OLS model. For this test, we used the *BFC02.gwr.test* function from the *spgwr* package which is based on the approach proposed by Fotheringham et al. (2002)

OLS RSS	**GWR RSS**	**F value**	**p-value**
79.5337	12.6035	6.3104	0.0000

Table 6: F-test for comparing OLS and GWR models

Spatial distribution of the relationships

To gain a clearer understanding of the spatial structure of the varying relationships across municipalities, it is beneficial to map the estimated coefficients derived from each local regression. In figure 11 (below), we can detect positive (negative) relationships to femicides with dark red (blue) colours, and light colours indicate low coefficients of association, in that way we can better appreciate the varying relationships and their intensities. These maps reveal extensive local variation in the estimated local coefficients of our explanatory variables, just like the results reported in table 5, and clearly depict clusters of positive and negative relationships across the municipalities for each variable. Particularly notable is the map depicting non-lethal violence against women (figure 11-b), where a strong positive association to femicides is observed at Colombia's border. Instead, the map illustrating school coverage (figure 11-f) displays a negative relationship with femicides primarily concentrated in the southern region. However, we will return later in this report to analyse these uneven relationships in more detail.

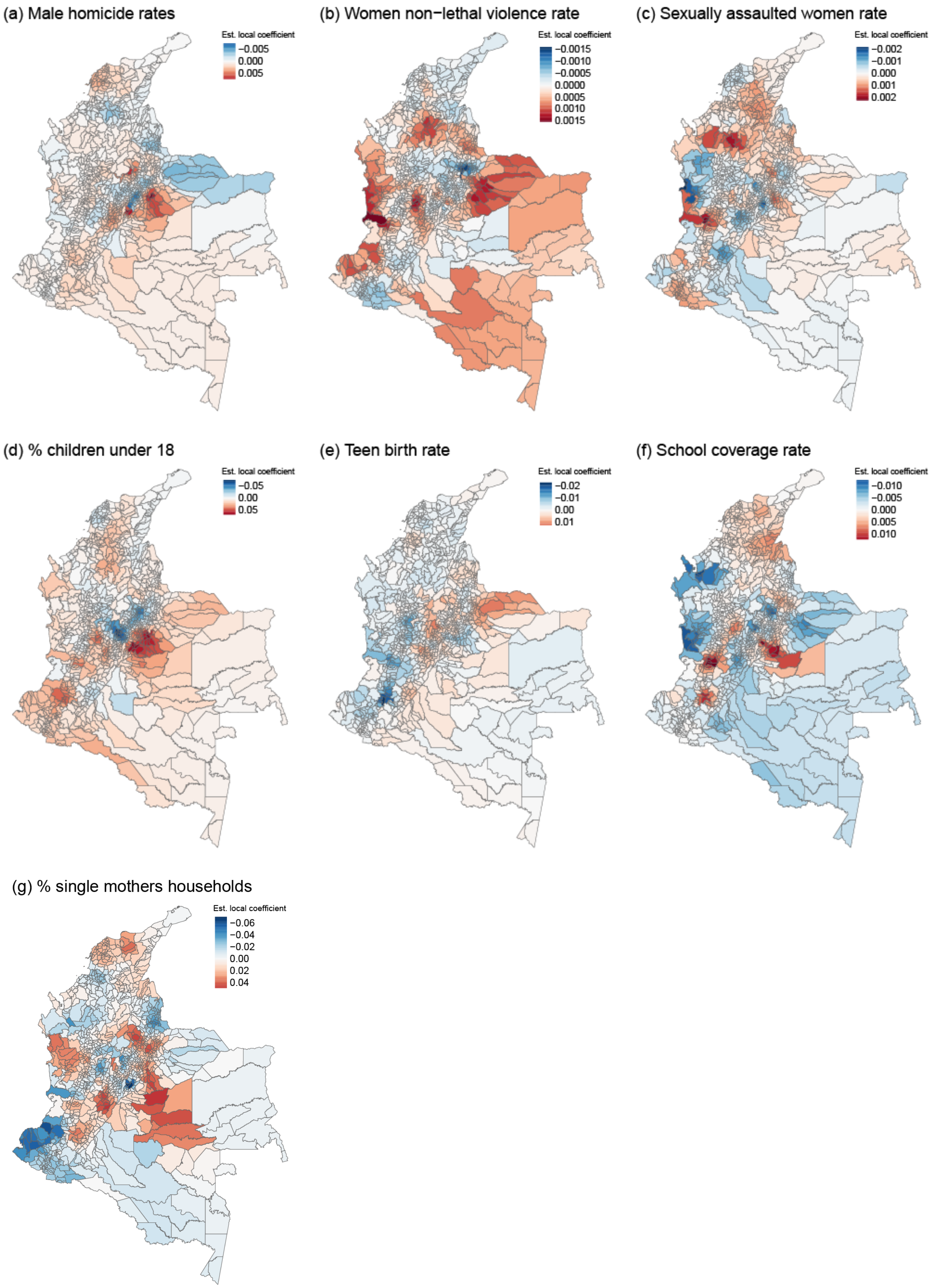

Figure 11: Spatial distribution of the local coefficients of each explanatory variable.

Overall GWR model performance.

Another significant advantage of the GWR model is its capability to map the R^2 for each of the estimated local regressions. This visualization allows us to assess the overall model fit, with much more detail than the Quasi-global R^2, and identify regions where the model provides a better fit to the data. Thus, we can spatially visualize how well the GWR model explains smoothed femicide rates in various areas of Colombia.

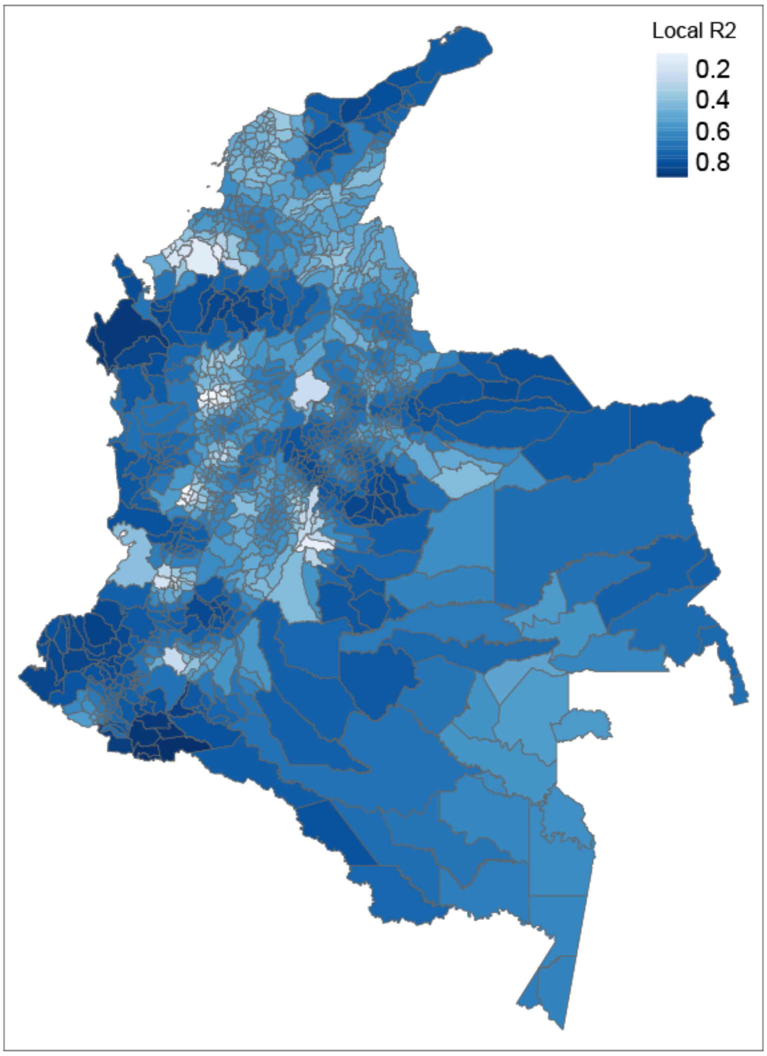

Figure 12: Spatial distribution of the local R^2 for the GWR model

Figure 12 shows that the performance of the model varies across different regions of Colombia. In general, GWR performs well in relatively large municipalities and in municipalities located on the outskirts of the country. And we can also notice regions of poor performance in the centre of the country and among some groups of small municipalities; these low R^2 values suggest that there could be additional factors related to femicides in these regions.

However, 56% of Colombia's municipalities have R^2 values above 0.6. This implies that, in more than half of the municipalities, more than two-thirds of the variability in femicide rates is explained by our proposed GWR model. Hence, we have an appropriate model to capture the non-stationary relationships of our covariates with respect to femicide rates.

Spatial autocorrelation analysis of the residuals

Furthermore, we will analyse the residuals of the models using the Global Moran's index to determine if spatial autocorrelation has been eliminated after the implementation of the GWR model.

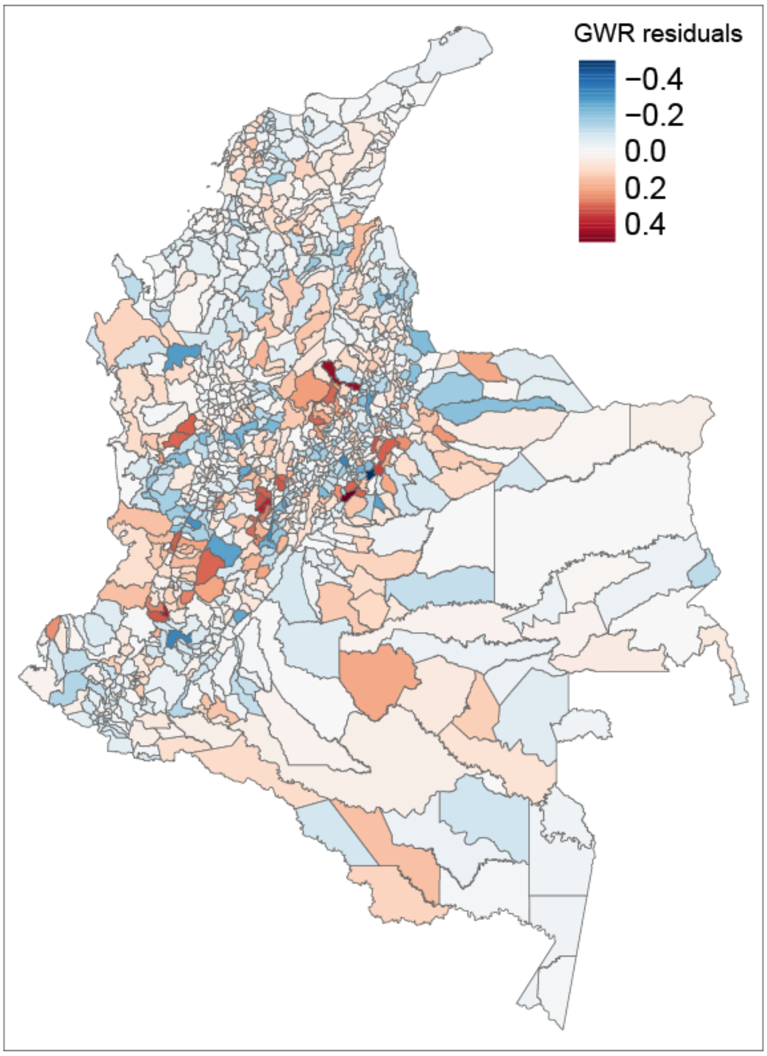

Figure 13: Spatial distribution of the GWR model residuals

The plot of the residuals (Figure 13) seems to indicate a random distribution throughout municipalities, lacking a distinct spatial pattern. This implies that the residuals derived from GWR parameters are spatially independent. Additionally, the Global Moran's I test was also conducted to test for spatial autocorrelation in residuals, and its results align with our intuition from the map above, where the Moran's I statistic is below 0.3 and the p-value is 0.16 (table 7). Thus, we have not evidence to reject the null hypothesis of a random spatial distribution of the residuals.

	Moran's I	**p-value**
GWR residuals	0.2574	0.16

Table 7: Global Moran's I test for the residuals of the GWR model

Test for spatial non-stationarity

Once the final model is chosen, we can proceed to assess whether each set of local parameters per explanatory variable within the model exhibits significant variation across the study area. To conduct this examination, Leung et al. (2000) propose a test to investigate spatial variability in each set of parameters. They suggest an F test statistic for each parameter which its distribution approximates an F-distribution under the null hypothesis of non-spatial variation for a given parameter. The F statistic reflects the spatial variation of a parameter; then, a significant large F statistic value supports the alternative hypothesis of spatial non-stationarity for that parameter.

In table 8, we display the result of this test for every estimated coefficient within our GWR model. For each parameter, we can reject the null hypothesis, even at a 1% significance level; thus, we have strong evidence that all the explanatory variables exhibit statistically significant spatial heterogeneity in their GWR coefficients, and we can affirm with a high degree of confidence that there is spatial heterogeneity in the relationships between our covariates and smoothed femicide rates in Colombia.

Variable	**F statistic**	**p-value**
Intercept	4.9582	0.0000
Homicides	2.8713	0.0000
Non-lethal violence	1.9968	0.0001
Sexual violence	1.2820	0.0036
Underage	5.5145	0.0000
Teen births	1.9470	0.0000
School coverage	2.2052	0.0000
Single mothers	2.1062	0.0000

Table 8: Test for spatial non-stationarity of the GWR parameters

For this test, we used the *LMZ.F3GWR.test* function from the *spgwr* package which is based on the method formulated by Leung et al. (2000).

Local collinearity

Finally, Wheeler and Tiefelsdorf (2005) advise against using this spatial technique in cases where the estimated coefficients show high correlation. GWR fits a local regression for each location in the dataset; therefore, if the values for a specific explanatory variable are spatially clustered, local collinearity issues are likely to arise (Brazil, 2019). One diagnostic procedure to assess the extent of local collinearity involves examining the correlations among the coefficients estimated by the GWR model. Comber et al. (2022) suggest that collinearity could pose an issue when the absolute correlation value between a pair of estimated coefficients exceeds 0.8.

Since none of the Spearman correlation coefficients in figure 14 is greater than 0.8 in absolute value, we can affirm that the selected variables for our GWR model do not exhibit collinearity problems. It is worth noting that the school coverage rates and the percentage of children under 18 years of age show a relatively high correlation with the intercept. However, as pointed out by Wheeler (2010), when a covariate displays little local spatial variation, as seen in figure 6-c and figure 7-b for these two variables, the potential for collinearity with the intercept term increases; nonetheless, their respective correlation values do not indicate issues of local collinearity.

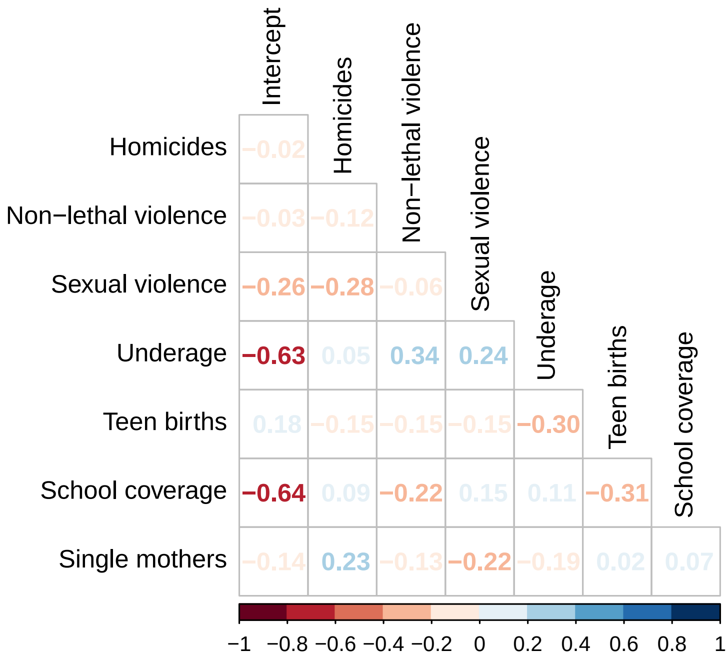

Figure 14: Correlation matrix between the estimated coefficients of the GWR model

Statistically significant relationships

Now that we have demonstrated spatial heterogeneity in the coefficients of our proposed GWR model, we can examine these varying relationships in more detail and, based on this analysis, provide valuable insights for decision-making in future actions aimed at preventing femicides.

While the maps displaying the spatial distribution of the local estimated coefficients (Figure 11) offer valuable information to examine spatial non-stationarity in our data, they do not determine whether these associations are statistically significant. To address this, we can employ the individual t-statistics of each local coefficient in each local regression for hypothesis testing, and then, we can plot the estimated coefficients that exhibit statistical significance at a 5% significance level. This approach allows us to identify those relationships that are significantly different from zero and the municipalities with such significant associations.

Figures 15, 16, and 17 provide a deeper and clearer understanding of the varying relationships by just mapping the magnitude and direction of the estimated coefficients that are statistical significance. These maps clearly indicate that the majority of associations with femicides are not significantly different from zero; however, they help us identify regions where the relationships are salient and important for formulating policies to address this problem.

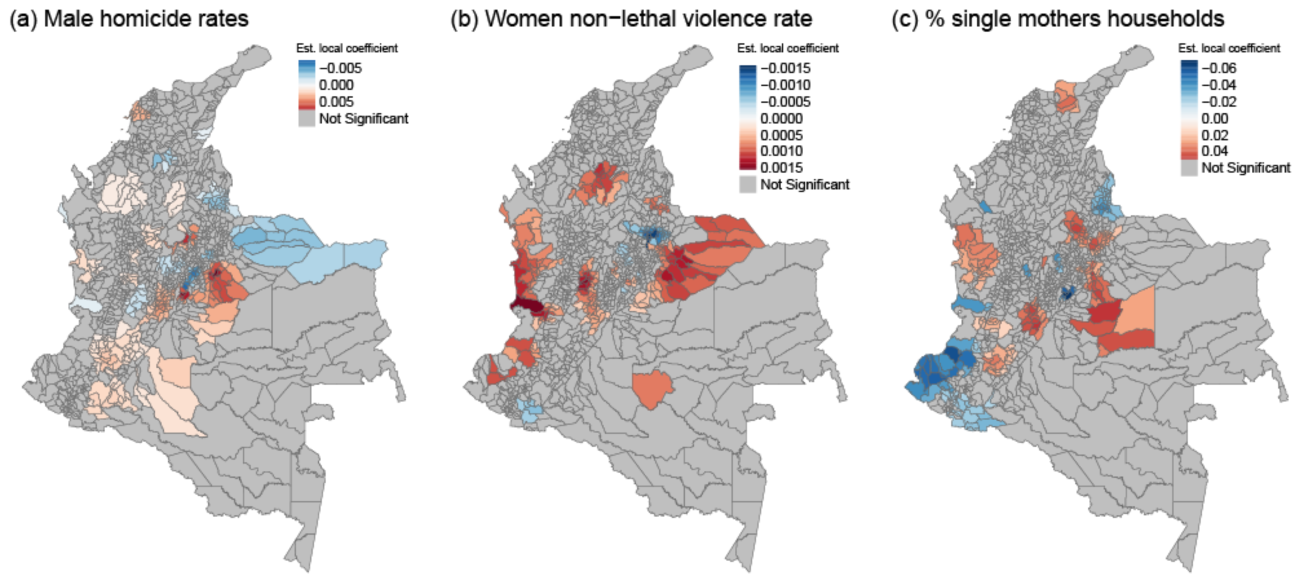

Figure 15: Spatial distribution of the statistically significant local coefficients, at a 5% significance level, for the male homicide rate (a), the female non-lethal violence rate (b), and the percentage of households headed by single mothers (c).

In figure 15, we can observe clear regions where there are positive and negative relationships with respect to femicide rates. Regarding the male homicide rate, we find areas of municipalities where there is a strong positive association towards the centre of the country. However, we also identify clusters where the relationship with respect to femicide rates is negative, which is contrary to what is expected based on the reviewed literature. It is worth noting that most of the reviewed literature used different statistical methods to study factors associated with femicide, and in the vast majority of them, they did not consider the spatial non-stationarity of these relationships. Therefore, with GWR methods is normal to find more variable spatial associations and contrary to those obtained by a global model.

As we observe in figure 15-a, these groups of negative association are mainly located in the centre and in proximity to the border with Venezuela. This motivates the need for a more in-depth analysis that includes variables for these regions that capture this possible relationship with Venezuela. It is also worth noting that the economic crisis in Venezuela, which has led to a 7-year economic recession (2013 to 2020), hyperinflation from 2017 to 2022, and unemployment rates of over 20% since late 2016, has resulted in a massive exodus of Venezuelans (around 2 million Venezuelans between 2017 and 2019), and their transit through Colombia, particularly towards the capital Bogota, aligns with the route marked by the clusters of statistically significant negative coefficients shown on the map.

Then, the impact of the Venezuelan economic and migratory crisis may have led to behavioural and contextual changes that could have altered the relationship between male homicide rates and smoothed femicide rates. It is also possible that other variables resulting from this specific Venezuelan effect are influencing both homicide and femicide rates, and these confounding variables would be what explain this counterintuitive relationship we are observing.

Regarding the rate of non-lethal violence against women (Figure 15-b), we find groups of municipalities with a strong positive relationship between the rates of this type of violence and femicide in the centre, on the Pacific coast, and on the border with Venezuela. If these clusters persist over time, these regions would be the ones that the Colombian government should prioritize to implement policies to prevent any type of gender violence. In this map, we also observe a group of municipalities with negative coefficients, which may also be associated with the Venezuelan crisis due to their proximity to the border with that country.

In the map of the percentage of single mothers heading households (figure 15-c), clusters can be observed by regions that seem to represent particular contextual factors in those areas. In regions where a higher prevalence of single mothers leading households is related to a lower prevalence of femicides, and in others where the opposite occurs, may correspond to specific changes in social attitudes and perceptions towards female power across different regions of the country. This suggests that policies for the prevention and assistance of femicide victims should be targeted based on the norms and social structures of each region. Therefore, it is also important to study how these social factors vary for each region to better understand them.

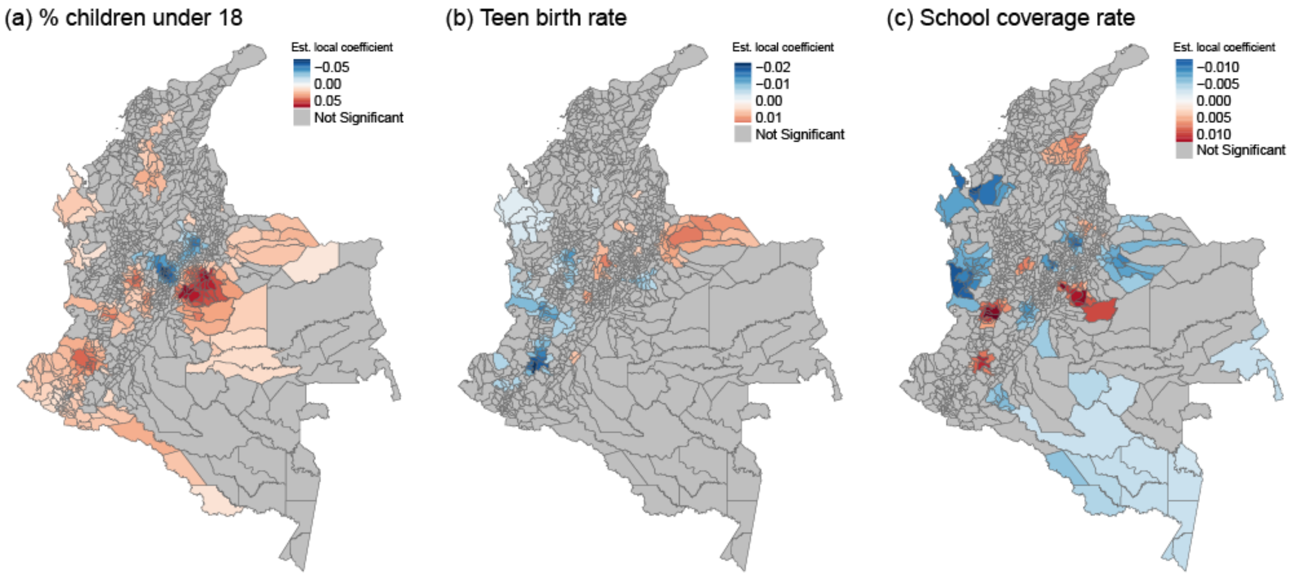

Figure 16: Spatial distribution of the statistically significant local coefficients, at a 5% significance level, for the percentage of children under 18 years of age (a), the teen birth rate (b), and the net school coverage rate (c).

In Figure 16, we can appreciate another interesting spatial distribution of relationships. For instance, higher percentages of minors are particularly associated with higher femicide rates along the southern border, the Pacific coast, and in the central part of the country (Figure 16-a). However, in a specific region in the centre, the relationship changes in the opposite direction. This particular region deserves an in-depth study by authorities to better understand its uniqueness compared to other regions, and therefore, to take better actions against femicides that incorporate this specific characteristic of that region.

Regarding the net school coverage rate (Figure 16-c), it is very clear that in many areas, the lack of school coverage is related to higher femicide rates, a situation that is particularly relevant for municipalities in the Amazonian region. Therefore, measures focused on encouraging school coverage and attendance, as well as increased investment in education, could be relevant in preventing the occurrence of femicides. However, there are also some clusters where the association is positive, and therefore, further analysis is required for these groups of municipalities to better understand and explain this unusual relationship.

The relationship between the teenage birth rate and smoothed femicide rates shows a difference between the east and the west of the country (figure 16-b), which could also be due to spatial changes resulting from the Venezuelan crisis since the groups of municipalities with positive associations are located on the same migratory route of Venezuelans, and precisely the teen birth rates in Colombia have been particularly high among Venezuelan adolescents. However, the most interesting aspect of figure 16-b and figure 17 (which maps the significant coefficients of sexual violence against women rates) is that very few municipalities have coefficients significantly different from zero, with only 16% of municipalities in Colombia having some significant association between the teen birth and sexually assaulted women rates with respect to the smoothed rate of femicides per 100.000 women.

In the map of sexual violence against women (figure 17), several clusters of positive and negative relationships are also highlighted, requiring further specific analysis for those regions because, unlike the previous maps, these clusters are smaller and more dispersed in Colombia's geography, making it challenging to detail spatial patterns at a macro level. Therefore, they are more focused on particular underlying contextual characteristics of these small groups of municipalities.

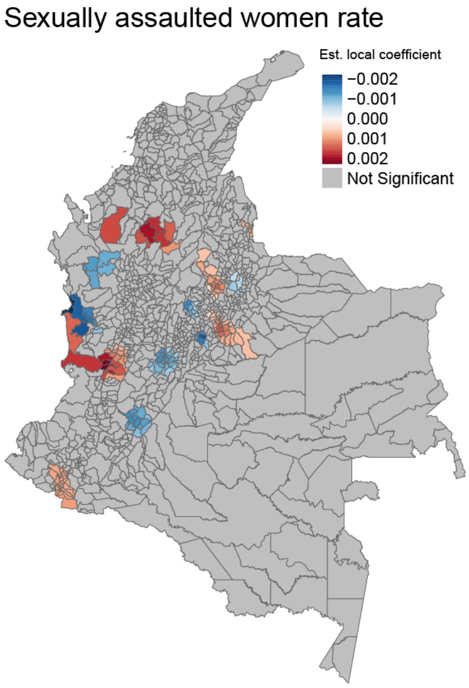

Figure 17: Spatial distribution of the statistically significant local coefficients, at a 5% significance level, for the rate of sexually assaulted women.

Finally, these maps also show that there are some municipalities where there is no statistically significant relationship between any of the explanatory variables and femicide rates, such as the municipalities located at the northern corner of the country. This may imply that for these municipalities, there might be other specific variables in these areas that need to be included in the model since the explanatory variables used do not seem to be associated with our study variable.

5 Conclusions and limitations

The results of our study suggest strong evidence of a spatial heterogeneity in the relationships between our covariates and femicide rates in Colombia for the period 2017-2019. Which in turn demonstrates the capability of the Geographically Weighted Regression framework in identifying non-stationary variables. At the same time, the EB smoothing technique shows its effectiveness in revealing spatial trends and emphasizing features that might be hidden by outliers and instability.

In summary, the estimated coefficients of our proposed GWR model indicate that femicide rates in the north of Colombia appear to be closely linked to the percentage of children under 18 years of age, net school coverage rates, male homicide rates, and the percentage of households headed by single mothers. In the west and the central area of Colombia, all the explanatory variables have significant relationships with femicide rates. On the other hand, in the southeast area (the Amazon region), the smoothed femicide rates are only related to the percentage of underage children, school coverage rates, and male homicide rates. However, the relationships also vary from positive to negative coefficients across space for all the explanatory variables.

Although this spatial technique has allowed us to demonstrate that there is spatial non-stationarity in the relationships between femicide rates and other explanatory variables in Colombia, and to identify specific regions that the national government and local authorities should focus on to establish targeted policies, further regional analysis is required to understand the non-stationarity and thus define appropriate policies for each identified cluster, as well as to identify if we are overlooking any important variables that the model should incorporate, enabling us to explain some relationships that we have identified as counterintuitive or contrary to expectations.

Multiple hypotheses can be developed to explain these unexpected relationships, but it is only through further research focused on these regions that we can better understand such results. Hence, this study's limitation lies in not addressing these other research questions, and although we acknowledge their importance, they are ultimately outside the scope of this research.

Varying relationships across Colombia implies that the measures to be taken by the government should be more focused on local policies; consequently, the social issue of femicide in Colombia must be studied with a greater regional focus. However, this, in turn, raises the level of complexity in addressing femicides in Colombia. Nonetheless, the maps showing the statistically significant associations between the explanatory variables and femicides help target the necessary measures and reduce the level of analytical complexity, as these maps highlight regions where certain variables are more related to the problem under study. Certainly, these findings can aid in shaping policy decisions aimed at early detection and prevention of femicides, such as the horrible case of Rosa Elvira Cely, within these identified clusters.

It is worth noting that a significant limitation in femicide data in Colombia is the high level of underreporting due to the majority of women not reporting out of fear or the judicial systems not effectively addressing their cases (UNODC and UN Women, 2022). This implies that our variable capturing femicides may not be entirely reliable, affecting the relevance of our analysis. However, this is a common issue in femicide data in many countries. It is necessary to thoroughly investigate the level of underreporting in collected femicide information; nonetheless, the fact that the official figures collected show consistent levels over the years supports the assumption that these official figures serve as a proxy for the actual numbers.

It is also important to note that this spatial analysis is incomplete as it does not consider information from the administrative divisions of neighbouring countries that are equivalent to a municipality in Colombia. As is evident in Figure 1, Colombia has municipalities at its borders that are directly adjacent to regions of other countries, such as in the southern corner in the Amazon region where municipalities are surrounded by both Brazil and Peru.

Therefore, considering the Tobler's first law of geography, which is an essential assumption for the spatial model we used, it is necessary that future research include data from neighbouring countries to achieve greater precision in the study of femicides. Colombia is not an island, and it is logical to assume that contextual factors from neighbouring countries may exert some influence on bordering municipalities.

However, proceeding in this manner will lead to greater difficulty in obtaining similar information or capturing the same variables. This is because there are sometimes differences in how information is collected between countries, such as the classification of femicides, which varies among Latin American countries in terms of the conditions that constitute a femicide in their respective laws. Additionally, it may not be possible to obtain the same information across countries, as is the case with the multidimensional poverty index, which is not calculated by some countries because they use other poverty indices or indicators.

Another limitation of our study was that some of the factors related with femicides identified in the literature review section were not possible to acquire or find at the municipal level, such as the environmental impact of industrial development projects, the proportion of land dedicated to coca production, the rate of forced displacements in relation to the number of femicides, and the involvement of women in politics; therefore, we recommend that future research attempt to collect or obtain information that can capture these other variables omitted in our GWR model.

In our analysis, we also attempted to model the number of feminicide cases with a Geographically Weighted Poisson Regression model, but we encountered overdispersion in the data (where the variance of the data is higher than the mean). In such cases, the Poisson assumption is not a suitable modeling approach, and a Negative Binomial distribution would be more appropriate to address overdispersion. However, none of the available R packages (*spgwr*, *GWmodel*, and *gwrr*) allow for the incorporation of such distribution in the generalized geographically weighted regression approach (Gollini et al., 2015). Hence, it is recommended for future research in this field to develop and implement, in R or other software, a GWR modeling approach with a Negative Binomial distribution to further investigate the non-stationarity in feminicide cases.

Finally, in our analysis and recommendations, we have sometimes assumed that these results hold over time; however, just as we have analyzed with the potential influence of the Venezuelan crisis during this study period, these relationships can change not only across space but also over time. To study these kinds of research problems, working with spatio-temporal methods is required. In that way, it would be possible to study the official femicide data available over a broader time interval than those analyzed in this research.

In particular, a model called Geographical and Temporal Weighted Regression (GTWR) can be employed to consider local effects in both space and time, facilitating the analysis of spatio-temporal non-stationary features of the data under study (Fotheringham et al., 2015). The GTWR model has demonstrated improved performance compared to the standard GWR methods and showed significant advantages in simultaneously modeling spatial and temporal non-stationarity across diverse datasets (Sollers et al., 2019; Ma et al., 2018; Bo Huang and Barry, 2010). Therefore, future studies using this approach could provide more valuable insights on femicide data for Colombia.

References

Anselin, L., N. Lozano, and J. Koschinsky (2006, 01). Rate transformations and smoothing. *Spatial Analysis Laboratory Department of Geography*.

Anselin, L., G. Morrison, A. Li, and K. Acosta (2020a). *Hands-On Spatial Data Science with R*, Chapter 5: Rate Mapping. University of Chicago.

Anselin, L., G. Morrison, A. Li, and K. Acosta (2020b). *Hands-On Spatial Data Science with R*, Chapter 9: Applications of Spatial Weights. University of Chicago.

Anselin, L., G. Morrison, A. Li, and K. Acosta (2020c). *Hands-On Spatial Data Science with R*, Chapter 6: Contiguity-Based Spatial Weights. University of Chicago.

Anselin, L., G. Morrison, A. Li, and K. Acosta (2020d). *Hands-On Spatial Data Science with R*, Chapter 10: Global Spatial Autocorrelation 1. University of Chicago.

Bates, E. and G. Pryce (2014). Geographically weighted regression as an automatic valuation method. *University of Glasgow*.

Bejarano C., M. (2014). El feminicidio es sólo la punta del iceberg. *Región y Sociedad*, 13–44.

Belsley, David A, E. K. and R. E. Welsch (2005). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity.* John Wiley Sons.

Beyer, K., P. Layde, L. Hamberger, and P. Laud (2014, 12). Does neighborhood environment differentiate intimate partner femicides from other femicides? *Violence against women 21*.

Bo Huang, B. W. and M. Barry (2010). Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science 24*(3), 383–401.

Botero, M. C. (2023). ¿qué estamos haciendo para evitar los feminicidios en colombia?

Brazil, N. (2019). *Spatial Methods in Community Research*, Chapter Geographically Weighted Regression. University of California, Davis.

Brazil, N. (2023). *Spatial Methods in Community Research*, Chapter 7: Spatial Autocorrelation. University of California, Davis.

Brewer, V. E. and M. D. Smith (1995). Gender inequality and rates of female homicide victimization across u.s. cities. *Journal of Research in Crime and Delinquency 32*, 175 – 190.

Brunsdon, C., A. Fotheringham, and M. Charlton (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical Analysis 28*(4), 281–298.

Brunsdont, C., S. Fotheringham, and M. Charlton (1998). Geographically weighted regression - modelling spatial non-stationarity. *Journal of the Royal Statistical Society Series D: The Statistician 47*(3), 431–443.

Campbell, J., D. Webster, J. Koziol-McLain, C. Block, D. Campbell, M. Curry, F. Gary, N. Glass, J. McFarlane, C. Sachs, P. Sharps, Y. Ulrich, S. Wilt, J. Manganello, X. Xu, J. Schollenberger, V. Frye, and K. Laughon (2003, 08). Risk factors for femicide in abusive relationships: Results from a multisite case control study. *American journal of public health 93*, 1089–97.

Castañeda Salgado, M. (2016, 03). Feminicide in mexico: An approach through academic, activist and artistic work. *Current Sociology 64*.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association 74*(368), 829–836.

Cliff, A. D. and J. K. Ord (1973). *Spatial autocorrelation*. London: Pion.

Cliff, A. D. and J. K. Ord (1981). *Spatial processes: Models and applications*. London: Pion.

Comber, A., C. Brunsdon, M. Charlton, G. Dong, R. Harris, B. Lu, Y. Lu, D. Murakami, T. Nakaya, Y. Wang, and P. Harris (2022, 01). A route map for successful applications of geographically weighted regression. *Geographical Analysis 55*.

Congress of Colombia (2015). Law 1761 of 2015.

Covington, D. L., B. J. Justason, and L. N. Wright (2001). Severity, manifestations, and consequences of violence among pregnant adolescents. *Journal of Adolescent Health 28*(1), 55–61.

Coy, M., L. Kelly, J. Foord, and J. Bowstead (2011). Roads to nowhere? mapping violence against women services. *Violence Against Women 17*(3), 404–425. PMID: 26086982.

DANE (2020). *Medida de Pobreza Multidimensional Municipal de Fuente Censal*. National Administrative Department of Statistics of Colombia.

De Bellefon, M. and J. Floch (2018). 'geographically weighted regression' in handbook of spatial analysis. *Eurostat*.

Fotheringham, A., C. Brunsdon, and M. Charlton (2002, 01). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*, Volume 13.

Fotheringham, A., R. Crespo, and J. Yao (2015, 04). Geographical and temporal weighted regression (gtwr). *Geographical Analysis 47*.

Gollini, I., B. Lu, M. Charlton, C. Brunsdon, and P. Harris (2015). Gwmodel: An r package for exploring spatial heterogeneity using geographically weighted models. *Journal of Statistical Software 63*(17), 1–50.

Grana, S. (2001, 01). Sociostructural considerations of domestic femicide. *Journal of Family Violence 16*, 421–435.

Hastie, T. and R. Tibshirani (1990). *Generalized additive models*. Wiley Online Library.

Hernández Breña, W. (2018, 11). Violence with femicide risk: Its effects on women and their children. *Journal of Interpersonal Violence 36*, 088626051881513.

IACHR (2013). Annual report of the inter- american commission on human rights 2012.

Ingram, M. and M. Marchesini da Costa (2015, 05). Geographies of violence: A spatial analysis of five types of homicide in brazil's municipalities. *SSRN Electronic Journal*.

Ingram, M. and M. Marchesini da Costa (2017, 05). A spatial analysis of homicide across brazil's municipalities. *Homicide Studies 21*.

Inspector General of Colombia (2023a). Boletin 1040.

Inspector General of Colombia (2023b). Boletin 280.

Kim, Y., T.-H. Kim, and T. Ergün (2015). The instability of the pearson correlation coefficient in the presence of coincidental outliers. *Finance Research Letters 13*, 243–257.

Land, K., P. Mccall, and L. Cohen (1990, 01). Structural covariates of homicide rates: Are there any invariances across time and social space? *American Journal of Sociology - AMER J SOCIOL 95*.

Leung, Y., C. Mei, and W.-X. Zhang (2000, 02). Statistical tests for spatial nonstationary based on the geographically weighted regression model. *Environment and Planning A 32*, 9–32.

Li, Q., R. Kirby, R. Sigler, S.-S. Hwang, M. Lagory, and R. Goldenberg (2009, 09). A multilevel analysis of individual, household, and neighborhood correlates of intimate partner violence among low-income pregnant women in jefferson county, alabama. *American journal of public health 100*, 531–9.

Ma, X., J. Zhang, C. Ding, and Y. Wang (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.

Marshall, R. (1991). Mapping disease and mortality rates using empirical bayes estimators. *Journal of the Royal Statistical Society. Series C, Applied statistics 40*(2), 283—294.

Martins, A. and A. Nascimento (2017, 01). Domestic violence, alcohol and other associated factors: A bibliometric analysis. *Arquivos Brasileiros de Psicologia 69*, 107–121.

Meneghel, S. and V. Hirakata (2011, 06). Femicides: Female homicide in brazil. *Revista de saúde pública 45*, 564–74.

Meneghel, S. and A. Portella (2017, 09). Femicides: concepts, types and scenarios. *Ciência Saúde Coletiva 22*, 3077–3086.

Messing, J. T., J. C. Campbell, and C. Snider (2017). Validation and adaptation of the danger assessment-5: A brief intimate partner violence risk assessment. *Journal of Advanced Nursing 73*(12), 3220–3230.

Moran, P. A. P. (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society. Series B (Methodological) 10*(2), 243–251.

Moran, P. A. P. (1950, 06). A test for the serial independence of residuals. *Biometrika 37*(1-2), 178–181.

Moroskoski, M., F. C. Neto, F. A. Machado de Brito, G. V. Ferracioli, N. N. de Oliveira, A. de Carvalho Dutra, V. D. A. Baldissera, and R. R. de Oliveira (2022). Lethal violence against women in southern brazil: Spatial analysis and associated factors. *Spatial and Spatio-temporal Epidemiology 43*, 100542.

O'Sullivan, D. and D. Unwin (2010). *Geographic Information Analysis*. Wiley.

Paez, A. (2022a). *An Introduction to Spatial Data Analysis and Statistics: A Course in R*, Chapter 21: Area Data II. McMaster University.

Paez, A. (2022b). *An Introduction to Spatial Data Analysis and Statistics: A Course in R*, Chapter 23: Area Data III. McMaster University.

Paez, A. (2022c). *An Introduction to Spatial Data Analysis and Statistics: A Course in R*, Chapter Chapter 27: Area Data V. McMaster University.

Paul Harris, A. S. F. and S. Juggins (2010). Robust geographically weighted regression: A technique for quantifying spatial relationships between freshwater acidification critical loads and catchment attributes. *Annals of the Association of American Geographers 100*(2), 286–306.

Pebesma, E. and R. Bivand (2023). *Spatial Data Science: With Applications in R*, Chapter 15: Measures of Spatial Autocorrelation. Chapman and Hall/CRC.

Pengpid, S., K. Peltzer, J. McFarlane, and A. Puckpinyo (2016). Intimate partner violence among pregnant women and women attending out-patient services in thailand : a longitudinal study. *Gender and Behaviour 14*(3), 7533–7546.

Páez, A., S. Farber, and D. Wheeler (2011). A simulation-based study of geographically weighted regression as a method for investigating spatially varying relationships. *Environment and Planning A: Economy and Space 43*(12), 2992–3010.

Rio, J., J. Doucet, and C. Chauvin (2010, 09). Violent and vindictive women: A re-analysis of the southern subculture of violence. *Sociological Spectrum 30*, 484–503.

Rowe, F. and D. Arribas-Bel (2023a). *Spatial Modelling for Data Scientists*, Chapter 2: Spatial Data. University of Liverpool.

Rowe, F. and D. Arribas-Bel (2023b). *Spatial Modelling for Data Scientists*, Chapter 9: Geographically Weighted Regression. University of Liverpool.

Sepúlveda Murillo, F. H., J. Chica - Olmo, and A. R. G. de Cortázar (2018). The spatial heterogeneity of factors of feminicide: The case of antioquia-colombia. *Applied Geography 92*, 63–73.

Silverman, J., A. Raj, L. Mucci, and J. Hathaway (2001, 09). Dating violence against adolescent girls and associated substance use, unhealthy weight control, sexual risk behavior, pregnancy, and suicidality. *JAMA : the journal of the American Medical Association 286*, 572–9.

Simon, L. and D. Young (2018). *Applied Regression Analysis*, Chapter 10.7 - Detecting Multicollinearity Using Variance Inflation Factors. The Pennsylvania State University.

Sollers, K., X. Liu, and B. Martinez-Lopez (2019, 01). An implementation of geographical and temporal weighted regression in modeling occurrences of porcine reproductive and respiratory syndrome in the us swine industry. *Frontiers in Veterinary Science 6*.

Srinivasan, S. (2008). *Spatial Regression Models*, pp. 1102–1105. Boston, MA: Springer US.

Stöckl, H., K. Devries, A. Rotstein, N. Abrahams, J. Campbell, C. Watts, and C. Moreno (2013, 06). The global prevalence of intimate partner homicide: A systematic review. *Lancet*.

Tekkas Kerman, K. and P. Betrus (2018, 06). Violence against women in turkey: A social ecological framework of determinants and prevention strategies. *Trauma, Violence, Abuse 21*, 152483801878110.

Tobler, W. R. (1970). A computer movie simulating urban growth in the detroit region. *Economic Geography 46*, 234–240.

UN (1979). Convention on the elimination of all forms of discrimination against women.

UN (1993). Declaration on the elimination of violence against women.

UN Women (2023). Ending violence against women.

UNODC and UN Women (2022). Gender-related killings of women and girls: Improving data to improve responses to femicide/feminicide.

Wheeler, D. and M. Tiefelsdorf (2005, 02). Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *Journal of Geographical Systems 7*, 161–187.

Wheeler, D. C. (2010). *Visualizing and Diagnosing Coefficients from Geographically Weighted Regression Models*, pp. 415–436. Dordrecht: Springer Netherlands.

WHO (2012). Understanding and addressing violence against women: femicide. Technical documents.

WHO (2021). *Violence against women prevalence estimates, 2018: global, regional and national prevalence estimates for intimate partner violence against women and global and regional prevalence estimates for non-partner sexual violence against women.* World Health Organization.