Daengku: Journal of Humanities and Social Sciences Innovation *ISSN: 2775-6165 (online)*
Vol. 5 No. 4 (2025)
https://doi.org/10.35877/454RI.daengku4449

Geographically Weighted Regression (GWR) Modeling in Identifying Factors Affecting the Gender Empowerment Index in Indonesia

Sitti Masyitah Meliyana[a,*], Ansari Saleh Ahmar[a], & Abdul Rahman[b]

[a]*Department of Statistics, Universitas Negeri Makassar, AP Pettarani, Makassar, 90222, Indonesia*
[b]*Department of Mathematics, Universitas Negeri Makassar, AP Pettarani, Makassar, 90222, Indonesia*

Abstract

This study aims to analyze the factors influencing the Gender Empowerment Index (GEI) in Indonesia using the Geographically Weighted Regression (GWR) method. The variables used in this study include the proportion of women in managerial positions, women's income contribution, the proportion of professional workers, reported health complaints, and the proportion of women in parliament. The findings indicate that, among the five independent variables examined, only two variables significantly affect the dependent variable: the proportion of women in managerial positions (X1) and the percentage of women reporting health complaints (X5). This is evidenced by their respective probability values (Pr(>F)) of 0.0045 and 0.0128, which are below the 0.05 significance threshold. This implies that X1 and X5 have a statistically significant influence in the model. The GWR model was found to be the most suitable compared to other models, with an AIC value of 186.72 and an R^2 of 92.03%, indicating superior model performance in capturing spatial and non-spatial effects across regions.

Keywords: Gender, Gender Empowerment Index, Spatial, Geographically Weighted Regression

1. Introduction

Gender empowerment is a crucial indicator of development that reflects the extent to which women have access to opportunities, resources, and decision-making roles in social, economic, and political spheres. In Indonesia, efforts to promote gender equality have become a national priority, as stated in various government policies and development agendas (Handayani and Raharjo, 2022). The Gender Empowerment Index (GEI) is widely used to measure these achievements, capturing the distribution of women's participation in economic activities, professional employment, and political representation. Despite continuous progress, disparities in gender empowerment remain evident across regions in Indonesia, indicating the presence of spatial heterogeneity that warrants in depth investigation.

Understanding the factors that influence GEI is essential for formulating effective and targeted policies. Previous studies have identified several potential determinants, such as women's representation in managerial positions, income contribution, participation in professional occupations, access to health services, and involvement in legislative bodies (Fitriani dan Nugroho, 2021). However, most of these studies rely on global statistical models that assume spatial homogeneity, potentially overlooking regional variations that may significantly influence the relationships among variables.

To address this gap, the present study employs the Geographically Weighted Regression (GWR) method, a spatial modeling approach capable of capturing local variations in parameter estimates (Zhu et al., 2020). GWR allows each region to have its own regression coefficients, providing more nuanced insights into factors affecting the GEI across Indonesia. By incorporating spatial dimension into the analysis, this study aims to produce a more accurate and contextually relevant understanding of gender empowerment dynamics.

Therefore, the objective of this research is to identify and evaluate the spatially varying factors that influence the Gender Empowerment Index in Indonesia. The findings are expected to contribute to the development of more region

* Corresponding author.
E-mail address: sittimasyitahmr@unm.ac.id

specific strategies and evidence-based gender policies, supporting national efforts to reduce disparities and promote inclusive development.

2. Literature Review

2.1. Regression Analysis

Regression analysis is a fundamental statistical technique used to investigate the relationship between a dependent variable and one or more independent variables. The primary objective of regression modeling is to estimate how changes in predictor variables influence the response variable. Although the true functional relationship is unknown, statistical models can approximate it using observed data.

In its simplest form, regression involves one dependent variable and a single predictor, known as simple linear regression. When multiple predictors are included, the model is referred to as multiple linear regression. The general formulation of the linear regression model, as described by Tiro (2010), is expressed as:

$$y_i = \beta_0 + \sum_{k=1}^{p} \beta_k x_{ik} + \varepsilon_i \tag{1}$$

where: y_i denotes the response variable for observation i; β_0 is the intercept; β_k represents the regression coefficient for predictor k; x_{ik} is the value of predictor $k$for observation i; ε_i is the error term; n and p denote the number of observations and predictor variables, respectively.

2.2. Geographically Weighted Regression (GWR)

Geographically Weighted Regression (GWR) is an extension of the traditional regression model that allows parameter estimates to vary across spatial locations. Unlike global regression models, which assume that relationships between variables are constant throughout the study area, GWR provides a localized modeling approach by estimating a separate set of regression coefficients for each observation point (Brunsdon et al., 1996). This enables the model to capture spatial heterogeneity, reflecting the possibility that the influence of predictor variables may differ from one

The GWR model is expressed as:

$$Y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i) X_{ik} + \varepsilon_i \tag{2}$$

where: Y_i is dependent variable at location i; (u_i, v_i) is geographical coordinates (longitude and latitude) of location i; β_0 is intercept specific to location i; β_k is coefficient of the k-th predictor at location i; X_{ik} is value of the k-th predictor at location i; ε_i is random error term at location i.

Parameter estimation in GWR is performed using the Weighted Least Squares (WLS) approach, in which observations closer to the target location receive higher weights than those farther away. This weighting scheme allows the model to account for local spatial dependencies in the data. The estimator for the coefficient vector at location (ui,vi) is given by (Brunsdon et al., 1996):

$$\hat{\beta}(u_i, v_i) = [X^T W(u_i, v_i) X]^{-1} X^T W(u_i, v_i) \tag{3}$$

where: $W(u_i, v_i)$ is spatial weight matrix for location i; X is matrix of independent variables; Y is vector of the dependent variable.

The flexibility of GWR in capturing spatial non-stationarity has made it widely applicable in various fields such as regional development, environmental studies, public health, and socio-economic research. By providing location-specific parameter estimates, GWR offers more detailed insights that are often overlooked by global regression models and thus enhances the accuracy and relevance of spatial data analysis.

2.3. Bandwidth in Geographically Weighted Regression (GWR)

Bandwidth selection is a crucial step in GWR because it determines the spatial extent over which observations contribute to local parameter estimation. A large bandwidth produces overly smoothed estimates that may mask important local variations, whereas a small bandwidth results in high variance due to the limited number of observations included in the estimation (Silverman, 1986). In GWR, weights are assigned based on the distance between observations and the target location, reflecting Tobler's First Law of Geography, which states that nearby observations exert stronger influence than distant ones (Tobler, 1970).

The choice of bandwidth directly affects the balance between bias and variance in the model (Gollini et al., 2015). To identify the optimal bandwidth, methods such as Cross-Validation (CV) are commonly used. The CV criterion minimizes the squared differences between observed and predicted values, defined as:

$$\mathrm{CV} = \mathrm{n} \sum_{i=1}^{n} \left(y_i - \hat{y}_{\neq i}(b)\right)^2 \tag{4}$$

This approach helps determine the bandwidth that yields the most accurate local parameter estimates within the GWR framework.

2.4. Model Selection

The Akaike Information Criterion (AIC) is a widely used method for selecting the best-fitting regression model, introduced by Akaike and discussed extensively in Grasa (1989). A model with a lower AIC value is considered superior because it achieves a better balance between goodness-of-fit and model complexity (Widarjono, 2005). The AIC for GWR can be calculated using the following formula:

$$AIC = 2n\ln(\hat{\sigma}) + n\ln(2\pi) + n + tr(L) \tag{5}$$

where**,** $\hat{\sigma}$ is maximum likelihood estimate of the error standard deviation, and *L*= projection matrix such that $\hat{y} = Ly$.

A lower AIC value indicates a more efficient model in capturing spatial variation while avoiding overfitting, making it a key criterion for selecting the optimal GWR specification.

3. Research Method

This study employs a quantitative research design utilizing the Geographically Weighted Regression (GWR) model to examine the factors influencing the Gender Empowerment Index (GEI) in Indonesia. The analysis is based on secondary data obtained from the 2023 Indonesia Statistical Yearbook published by the Indonesian Central Bureau of Statistics. The study incorporates six variables consisting of one dependent variable and five independent variables: (1) Gender Empowerment Index (Y); (2) Proportion of Women in Managerial Positions (X_1); (3) Women's Income Contribution (X_2); (4) Women in Professional Occupations (X_3); (5) Percentage of Female Population Reporting Health Complaints (X_4); and (6) Proportion of Women in Parliament (X_5). These variables are analyzed to identify spatially varying determinants of gender empowerment across Indonesian regions.

4. Results and Discussions

4.1. Descriptive Analysis

The Gender Empowerment Index (GEI) measures the extent to which women actively participate in economic and political spheres, particularly in terms of representation in parliament, managerial positions, access to higher education, and income contribution. This index reflects the level of gender equality related to empowerment and decision-making. The GEI values in Indonesia range from 30.3 to 80.6. The mean GEI is 68.2, while the median is 71.5, which is slightly higher than the mean. The spatial distribution of GEI across Indonesia is illustrated in Figure 1.

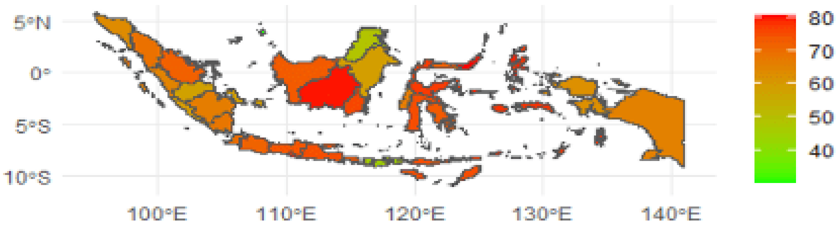

Figure 1. Spatial Distribution of GEI in Indonesia, 2023

According to Figure 1, the two provinces with the highest GEI scores are North Sulawesi (80.56) and Central Kalimantan (79.99). In contrast, the provinces with the lowest GEI scores are Riau Islands (30.32) and West Nusa Tenggara (44.28).

4.2. Linear Regression Analysis

Linear regression analysis was employed to examine the influence of several factors on the Gender Empowerment Index (GEI) in Indonesia. The parameter estimation results are presented in Table 1.

Table 1. Parameter Estimates of the Linear Regression Model

Variable	Estimate	Std.Error	*t-value*	$\Pr(>\lvert t\rvert)$	Description
Intercept	62,6387	18,3932	3,406	0,0020	
X_1	-0,3563	0,1920	-1,855	0,0742	Not significant
X_2	0,5343	0,2296	2,328	0,0274	Significant
X_3	0,2931	0,1347	2,177	0,0381	Significant
X_4	-0,9114	0,2506	-3,638	0,0011	Significant
X_5	0,4653	0,1878	2,478	0,0195	Significant

The initial regression model obtained is:

$$Y = 62.6387 - 0.3563X_1 + 0.5343X_2 + 0.2931X_3 - 0.9114X_4 + 0.4653X_5$$

Based on the p-values, variables X_2, X_3, X_4, and X_5 significantly affect GEI, while X_1 does not. Thus, a reduced model was constructed by removing X_1. The resulting model is:

$$Y = 44.4552 + 0.5342X_2 + 0.3093X_3 - 0.7677X_4 + 0.5166X_5$$

The results indicate that X_2, X_3, and X_5 have positive effects on GEI, meaning increases in women's income contribution, women in professional occupations, and women's representation in parliament lead to higher GEI scores. Conversely, X_4 has a negative effect, indicating that higher proportions of women with health complaints reduce GEI. The intercept value of 44.4552 represents the estimated GEI when all predictor variables are zero.

The results of the simultaneous test are coefficient of determination ($R^2 = 0.78$) indicates that 78% of the variation in GEI is explained by the predictors included in the model. The computed F-value (26.23) exceeds the critical F-value (2.56), and the p-value is far below 0.05. Therefore, the regression model is statistically significant and appropriate for explaining GEI in Indonesia.

4.3. Assumption Testing

Assumption testing was conducted on the residuals of the regression model. The procedures and results for each assumption test are presented below.

4.3.1. Normality Test

The normality assumption was examined using the Kolmogorov–Smirnov test. The test produced a D statistic of 0.0851 and a p-value of 0.7699. Since the p-value exceeds 0.05, the residuals are considered to follow a normal distribution. Thus, the normality assumption is satisfied.

4.3.2. Multicollinearity Test

Multicollinearity was assessed using the Variance Inflation Factor (VIF). The results indicate that all VIF values are well below the commonly accepted threshold of 10. Therefore, it can be concluded that multicollinearity is not present in the model. This implies that the five predictor variables do not excessively influence one another and are appropriate for inclusion in the regression analysis.

4.3.3. Heteroscedasticity Test

Heteroscedasticity was tested using the Breusch–Pagan test. The test yielded a BP statistic of 12.213 with a p-value of 0.03199. Because the p-value is less than 0.05, it can be concluded that heteroscedasticity is present in the data. This indicates that the residual variance is not constant across observations.

4.4. Geographically Weighted Regression (GWR) Modeling

The first step in the GWR modeling procedure is calculating the Euclidean distance between observational locations. After obtaining the distance matrix, spatial weights for each observation were calculated using the Gaussian and Bisquare kernel functions. The optimal bandwidth was determined using the Cross-Validation (CV) method, and the bandwidth with the smallest CV value was selected for model estimation.

Table 2. Optimal Bandwidth Selection

Kernel Funcion	Bandwith	CV	R^2	AIC
Fixed Gaussian	5.137	1123.159	0.920	187.652
Adaptive Gaussian	0.165	1081.002	0.920	186.721
Fixed Bisquare	16.405	1154.271	0.876	200.065
Adaptive Bisquare	0.853	1233.165	0.838	206.743

Based on Table 2, the optimal bandwidth is obtained using the Adaptive Gaussian kernel with a bandwidth value of 0.165 and a CV value of 1081.002. This optimal bandwidth is subsequently used to determine the spatial weights for each region in Indonesia.

After computing the spatial weights, parameter estimates for each region were obtained. These results indicate that the estimated coefficient for X_1 ranges from –0.8289 to 0.4595, X_2 from 0.3442 to 0.6579, X_3 from –0.2199 to 0.7384, X_4 from –1.2185 to –0.2939, and X_5 from –0.0292 to 0.9327. The complete parameter estimates for all Indonesian provinces are presented in Appendix 4. For illustration, the GWR model for Aceh Province is shown below:

$$Y_{\text{Aceh}} = 69.6933 - 0.7511X_1 + 0.4793X_2 + 0.3548X_3 - 0.8644X_4 + 0.6218X_5$$

This model indicates that IDG in Aceh is influenced by all five independent variables. Variables X_1 and X_4 exhibit negative effects, meaning increases in these variables reduce IDG. Conversely, X_2, X_3, and X_5 have positive impacts on the IDG score.

4.4.1. Model Fit Test for GWR

Model adequacy was assessed using the F-test. The analysis shows an F-statistic of 2.2436 with a p-value of 0.007842, indicating a significant difference between the GWR model and the classical regression model ($p < 0.05$). The GWR model yields an R^2value of 0.92032, meaning that 92.03% of the variation in IDG can be explained by the independent variables within the GWR framework, while the remaining 7.97% is explained by unobserved factors.

Subsequently, geographic variability of each predictor was examined using a simultaneous F-test. The results are summarized in Table 3.

Table 3. Simultaneous Test of Geographic Effects

Variable	F-value	DF1	DF2	Pr(>F)	Interpretation
Intercept	0.63880	9.65987	16.436	0.7576	Not significant
X1	4.07626	12.58210	16.436	0.0045	Significant
X2	0.19759	12.09604	16.436	0.9966	Not significant
X3	1.85159	13.40803	16.436	0.1181	Not significant
X4	0.90493	10.02103	16.436	0.5501	Not significant
X5	3.35543	11.46734	16.436	0.0128	Significant

The results indicate that only X_1 and X_5 exhibit significant spatial non-stationarity, as shown by their p-values below 0.05. This suggests that these two predictors vary significantly across geographic locations, while the remaining predictors do not demonstrate substantial spatial variation.

4.4.2. Partial t-tests Across Regions

Unlike the simultaneous test that evaluates overall geographic effects, the GWR model generates different significant predictors for each region. Partial t-tests were performed for all provinces using a critical value of $t_{0.05} = 1.699$. Several examples of province-specific GWR models are given:

$$Y_{\text{Sumatera Selatan}} = 52.95385 - 0.64028X_1 + 0.65802X_3 - 0.77785X_4$$

$$Y_{\text{Kalimantan Selatan}} = 64.93055 + 0.37638X_2 - 0.92123X_4 + 0.50504X_5$$
$$Y_{\text{Papua}} = 72.79170 + 0.49490X_2 + 0.22270X_3 - 1.06595X_4$$

These models demonstrate that the influence of predictors on IDG varies widely across provinces, highlighting the local nature of gender empowerment determinants. For instance, X_4 consistently shows a strong negative influence in several regions, whereas the impacts of X_2, X_3, and X_5 differ by geographic context.

4.5. Model Selection

The selection of the best model was carried out by comparing the R^2values and the AIC values of each model. A higher R^2indicates that the model is able to explain a larger proportion of the variance in the dependent variable, while a lower AIC value indicates a better model fit with fewer information losses. The comparison of these criteria is presented in Table 4.

Table 4. R^2 and AIC Values of the Models

Model	R^2	AIC
Linear Regression	0.7927	212.931
Geographically Weighted Regression	0,9203	186,7208

Based on Table 4, the Geographically Weighted Regression (GWR) model demonstrates a substantially higher R^2 value (0.9203) compared to the Linear Regression model (0.7927), indicating that GWR explains a greater proportion of the variation in the Gender Empowerment Index (IDG). Additionally, the GWR model has a much lower AIC value (186.7208) than the Linear Regression model (212.9309), confirming that the GWR model provides a better fit to the data. Thus, the GWR model is determined to be the best model for explaining the spatial variation in the IDG across provinces in Indonesia.

4.6. Discussion

The results of this study demonstrate that the Geographically Weighted Regression (GWR) model provides a substantially better fit than the classical linear regression model, as indicated by the higher coefficient of determination ($R^2 = 0.9203$) and the lower AIC value. This finding confirms that spatial heterogeneity plays an important role in explaining variations in the Gender Empowerment Index (IDG) across Indonesian provinces. The improved performance of GWR suggests that global modeling approaches may overlook important local patterns and relationships that vary across regions.

The analysis of optimal bandwidth selection further highlights the superiority of the adaptive Gaussian kernel, which produced the smallest CV value. This indicates that allowing the bandwidth to vary across space captures local variations more effectively than a fixed bandwidth approach. The resulting local parameter estimates reveal substantial spatial variation in the influence of predictor variables. For instance, variables such as X_1and X_5were found to have statistically significant spatial effects, suggesting that their impact on IDG is not uniform across regions.

The local parameter models for selected provinces, such as Aceh, South Sumatera, South Kalimantan, and Papua, illustrate that different combinations of predictors influence IDG in different areas. Some variables exhibit positive effects in certain provinces but negative effects in others, confirming the presence of spatial non-stationarity. This highlights the importance of considering local socio-economic and demographic characteristics when formulating policies aimed at improving gender empowerment.

Furthermore, the geographical testing results (F-test and local t-tests) emphasize that not all predictors exert significant influence in every region. Instead, each province demonstrates a unique pattern of influencing factors. This supports the argument that region-specific interventions may be more effective than national-level, uniform policy approaches for promoting gender empowerment.

The spatial distribution of regions based on the grouping of significant variables identified by the GWR models is presented in Figure 2.

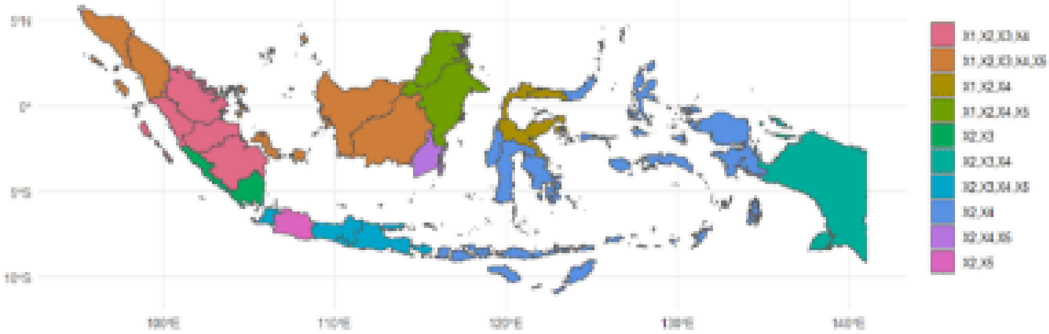

Figure 2. Map of Significant Variables in the GWR Model

5. Conclusion

This study concludes that the Geographically Weighted Regression (GWR) model is the most appropriate approach for analyzing the spatial determinants of the Gender Empowerment Index (IDG) in Indonesia. Compared with the classical linear regression model, GWR provides a superior fit, as evidenced by its higher R^2value and lower AIC score. These findings underscore the presence of spatial heterogeneity in the relationships between IDG and its predictor variables.

The results also show that the influence of independent variables varies across provinces, indicating that the factors shaping gender empowerment are inherently local. While variables such as $X_1$and X_5 exhibit statistically significant spatial effects, the relevance and magnitude of other variables differ from one region to another. This suggests that effective policy design should account for these local variations rather than relying on a one-size-fits-all approach.

Overall, the study highlights the importance of incorporating spatial modeling techniques such as GWR in analyzing socio-economic indicators. The findings provide a more nuanced understanding of gender empowerment in Indonesia and offer valuable insights for policymakers aiming to formulate targeted strategies to reduce disparities and promote equitable development across regions.

References

Alsos, G. A.(2021). Educational Investments and Gender Empowerment: A Path to Equality. *International Journal of Gender Studies*, 12(3), 231–245.

Badan Pusat Statistik (BPS). (2023). *Keadaan Angkatan Kerja di Indonesia Februari 2023.* Jakarta: Badan Pusat Statistik. Diakses pada 17 September 2024 dari https://www.bps.go.id/id/statistics-table/2/MjIwMCMy/tingkat-partisipasi-angkatan-kerja-menurut-jenis-kelamin.html.

BAPPENAS dan UN Women. (2020). Laporan Pembangunan Gender dan Inklusi Sosial di Indonesia. Jakarta: Kementerian Perencanaan Pembangunan Nasional/Bappenas dan UN Women.

Brunsdon, C., Fotheringham, A. S., dan Charlton, M. E. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical Analysis*, *28*(4), 281–298. https://doi.org/10.1111/j.1538-4632.1996.tb00936.x

Fotheringham, A. S., Brunsdon, C., dan Charlton, M. (2002). *Geographicall Weighted Regression the analysis of spatially varying relationships* (John Wiley dan Sons (ed.)).

Gollini, I., Lu, B., Charlton, M., Brunsdon, C., dan Harris, P. (2015). Gwmodel: An R package for exploring spatial heterogeneity using geographically weighted models. *Journal of Statistical Software*, *63*(17), 1–50. https://doi.org/10.18637/jss.v063.i17

Goulard, M., dan Voltz, M. (1992). Linear coregionalization model: Tools for estimation and choice of cross-variogram matrix. *Mathematical Geology*, 24(3), 269–286. https://doi.org/10.1007/BF00893750

Grasa, A. A. (1989). *Econometric model selection : a new approach advanced studies in theoretical and applied econometrics*.

Johnson, L.,(2024). Global Correlations Between Gender Empowerment and Human Development. *Sustainability Studies*, 20(2), 312–328.

Kabeer, N. (1999). Resources, agency, achievements: Reflections on the measurement of women's empowerment. *Development and Change*, 30(3), 435–464.

Lutfiani, N., dan Scolastika Mariani, (2019). Pemodelan Geographically Weighted Regression (GWR) dengan Fungsi Pembobot Kernel Gaussian dan Bi-square. *UNNES Journal of Mathematics*, *5*(1), 82–91. http://journal.unnes.ac.id/sju/index.php/ujmUJM8

Mar'ah, Z., Djuraidah, A., Wigena, A., dan Hamim Wigena, A. (2017). Semi-parametric Geographically Weighted Regression Modelling using Linear Model of Coregionalization. *Article in International Journal of Sciences Basic and Applied Research*, 34(2), 178–186. http://gssrr.org/index.php?journal=JournalOfBasicAndApplied

Meliyana, S. M., Ahmar, A. S., & Siti Nurazizah Auliah. (2025). Geographically Weighted Poisson Regression (GWPR) Model with Fixed Gaussian Kernel and Fixed Bi-square Kernel Weights. *ARRUS Journal of Social Sciences and Humanities*, *5*(2), 896-909. https://doi.org/10.35877/soshum3812

Nakaya, T., Fotheringham, A. S., Brunsdon, C., dan Charlton, M. (2005). Geographically weighted Poisson regression for disease association mapping. *Statistics in Medicine*, 24(17), 2695–2717. https://doi.org/10.1002/sim.2129

World Health Organization. (2009). Women and health: Today's evidence, tomorrow's agenda. World Health Organization. https://www.who.int/publications/i/item/9789241563857

Zhao, Q., Fan, Q., dan Zhou, P. (2022). An integrated analysis of GWR models and spatial econometric global models to decompose the driving forces of the township consumption development in Gansu, China. *Sustainability (Switzerland)*, *14*(1). https://doi.org/10.3390/su14010281

Zhu, C., Zhang, X., Zhou, M., He, S., Gan, M., Yang, L., dan Wang, K. (2020). Impacts of urbanization and landscape pattern on habitat quality using OLS and GWR models in Hangzhou, China. *Ecological Indicators*, *117*, 106654. https://doi.org/10.1016/J.ECOLIND.2020.106654