The Social Geography of Women's Attitudes Toward Wife-beating in Ethiopia: A Contribution Towards Proper Application of Spatial Statistics

Presented at the Addis Continental Institute of Public Health

October 20, 2023

Aynalem Adugna, PhD
Public Health Research Scientist III
State of California

Personal email: AynalemA2012@gmail.com
Website: http://www.ethiodemographyandhealth.org/
EthDemography dashboard: Aynalem Adugna, Ethiopian Demography and Health (arcgis.com)

CONTENTS

- Objectives
- Background: GIS and Public Health (Spatial Statistics)
- Improper Application of Spatial Statistics to DHS data
- Ethiopian Women's Attitudes Towards Wife-beating: Testing for Spatial Autocorrelation
- Geographically Weighted Regression (GWR)
- Summary

OBJECTIVES

1. Provide brief backgrounds about GIS and use of GIS in public health

2. Bring attention to improper applications of spatial statistics to Ethiopia's Demographic and Health Survey (EDHS) data

3. Show simplified steps for assessing whether a variable is suitable for spatial ordinary least squares (OLS) linear regression by using the 2016 EDHS data on women's attitudes toward wife-beating

4. Highlight a Geographic Information System (GIS) tool known as the Geographically Weighed Regression (GWR) by using as an example unweighted percentages of women who responded that wife-beating was okay (y) and their educational level (x)

BACKROUND:
I. Geographic Information Systems (GIS)

What is GIS?

- **Geographic information systems (GIS) is:**
  - a computer-aided database management system that organizes and stores large amounts of multi-purpose data
  - a mapping and *spatial analytics** technology

*"The process of examining the locations, attributes, and relationships of features in spatial data through overlay and other analytical techniques in order to **address a question** or **gain useful knowledge**. Spatial analysis extracts or creates new information from spatial data".

Source: https://libguides.tulane.edu/geographicinformationsystems/spatialanalysis

BACKGROUND Contd.
16 New GIS Trends Driving the Future

#1 Real-Time Data
#2 Miniaturization of Sensors
#3 Geospatial AI (GeoAI)
#4 Self-Driving Vehicles
#5 Smart Cities
#6 3D and Digital Twins
#7 Machine Learning
#8 Artificial Intelligence

#9 Sustainability (climate change)
#10 Disaster Mitigation and Response
#11 Workflow Automation
#12 Light Detection and Ranging (LiDAR)
#13 Geospatial Augmented Reality (AR)
#14 Subscription-based SaaS – Software as a Service
#15 Image/Raster Analytics
#16 GIS in The Cloud

Source: https://mappitall.com/blog/gis-technology-trends-that-driving-the-future

BACKGROUND Contd.
II. GIS and Public Health

- The field of public health is:

  - dedicated to improving the health of populations rather than individuals,
  - focused on prevention rather than treatment,
  - operationalized mainly in a governmental (rather than private) contexts

- This means that public health falls naturally within the domain of health sciences requiring spatial analysis

- This is because GIS allows potentially transformational capabilities for public health efforts by providing a digital lens for exploring the dynamic connections between people, their health and well-being, and changing physical and social environments

Source: Based on https://www.coursehero.com/file/115425040/GIS-AND-PUBLIC-HEALTHppt/

BACKGROUND Contd.
GIS and Public Health

GIS helps public health practitioners:

- Optimize data collection and management
- Strengthen data analysis
- Strengthen outbreak infrastructure and support
- Map epidemic dynamics in near real-time
- Quickly plan and target response
- Rapidly communicate information
- Monitor changes in disease over time
- Plan, monitor intervention/eradication programs
- Aid emergency preparedness

Source: The University of North Carolina, School of Public Health. https://nciph.sph.unc.edu/focus/vol5/issue2/5-2Mapping_slides.ppt

Background contd….Examples : 1) WHO 2) US-Centers for Disease Control (CDC) 3) California Department of Public Health

3

1 WHO

https://www.who.int/data/GIS

2 US-CDC

https://www.cdc.gov/gis/gis-at-cdc.htm

https://youtu.be/YeONEtqFdtA

Income inequality and health

https://storymaps.arcgis.com/stories/84fcd9cce18f498b8ac5861b95add64a

BACKROUND contd.

III. Spatial Statistics

…Is a Sub-field of Public Health GIS

Software Tools

SatScan (Scan Statistics)

https://www.satscan.org/

ESRI's ArcGIS Pro Spatial Statistics Tool Box

https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/an-overview-of-the-spatial-statistics-toolbox.htm

https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/an-overview-of-the-modeling-spatial-relationships-toolset.htm

S-Plus or R. A number of "add on" S-Plus packages particularly orientated to spatial applications are also available, in particular S+Spatial and S+GeoStat."

https://www.scielosp.org/pdf/csp/v17n5/6318.pdf

QGIS https://www.qgis.org/en/site/

Improper application of Spatial Statistics

….to Ethiopia's Demographic and Health Survey (EDHS) Data

Use of the Ordinary Least Squares (OLS) regression to analyse spatially autocorrelated data.

"Whenever there is statistically significant spatial autocorrelation of the regression residuals, the OLS model will be considered [mis-specified]. Consequently, results from OLS regression are unreliable. Be sure to run the Spatial Autocorrelation tool on your regression residuals to assess this potential problem."

https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/ordinary-least-squares.htm

VIDEO SPATIAL AUTOCORRELATION
https://youtu.be/M9ecMxVG6jQ

Improper application …contd.

Most published works featuring spatial analyses of EDHS data suffer from one or more flaws, including

1) application of bi- or multi-variate global (countrywide) Ordinary Least Square (OLS) regression in situations of high local variability
2) Application of OLS to a spatially autocorrelated response variable (y)
3) lack of testing for multicollinearity of the chosen independent variables (x)
4) combining the use of spatial and non-spatial statistical methods
5) publishing contradictory results from different spatial statistical applications - namely, ArcMap (or ArcGIS Pro) and SatScan - in the same paper

Source: The Social Geography of Women's Attitudes toward Wife-beating in Ethiopia: A Contribution Towards Proper Application of Spatial Statistics | Adugna | Journal of Geography and Geology | CCSE (ccsenet.org)

Example

Finding:

"The study found spatial clustering of domestic violence cases in Ethiopia with Moran's I value of 0.26, Z score of 8.26, and *P* value < 0.01."

Given the above finding, the validity of the following additional results is highly questionable

A regression analysis identified "…. low economic status, partner alcohol use, witnessing family violence, marital controlling behaviors, and community acceptance of wife-beating as significant predictors of domestic violence" in Ethiopia

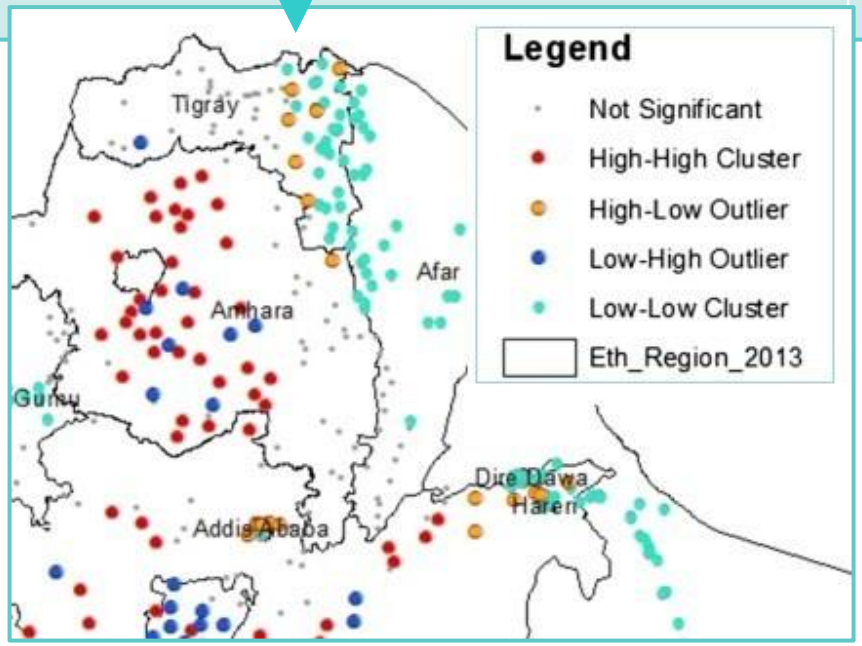
Women living in communities with high acceptance of wife-beating were 1.4 times more likely to experience domestic violence than women living in communities with low acceptance of wife-beating (AOR = 1.39, 95% CI: 1.16, 1.66)."

Source: Seid, E., Melese, T. & Alemu, K. Spatial distribution and predictors of domestic violence against women: evidence from analysis of Ethiopian demographic health survey 2016. *BMC Women's Health* **21**, 334 (2021). https://doi.org/10.1186/s12905-021-01465-4

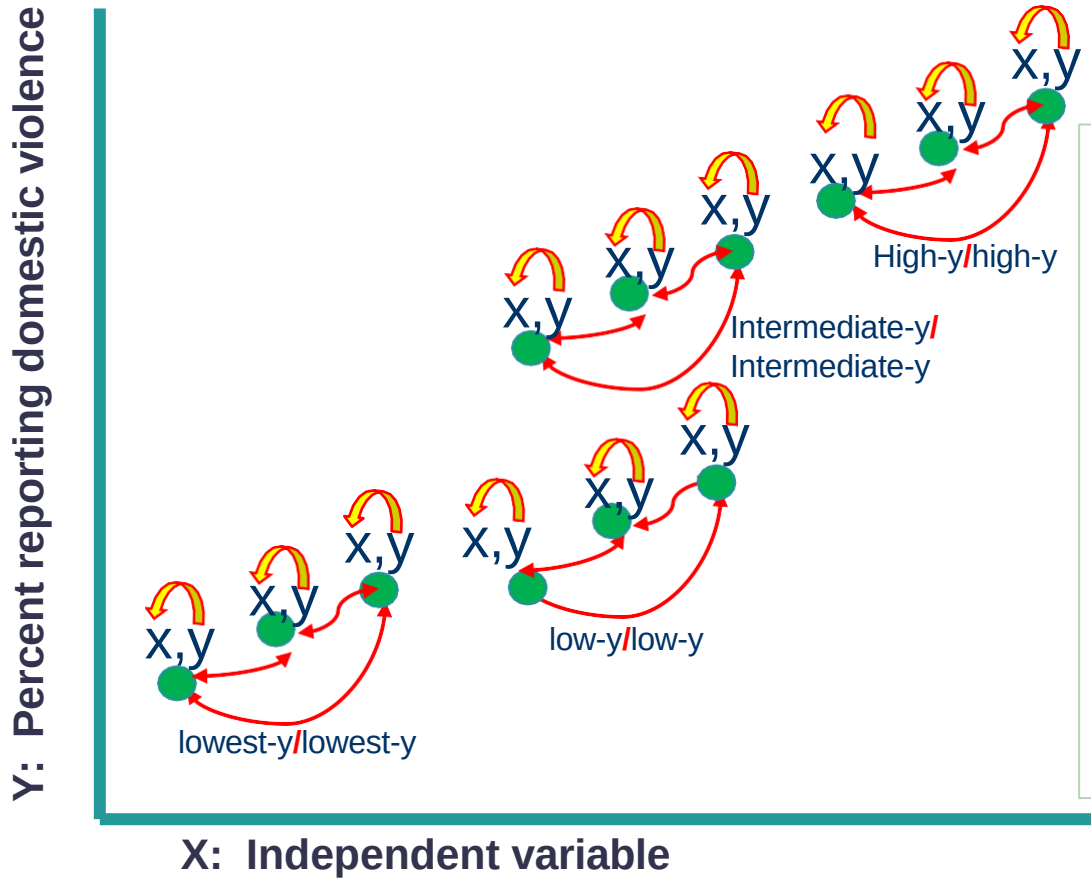
The use of a composite index as a dependent (y) variable is also problematic

Variables	Measurement
Partner drinks alcohol	Classified as 'yes' if partner drinks alcohol and 'no' otherwise
Witnessed family violence	'Yes' or 'no' based on their answer to the question "as far as you know, your father ever hit your mother?"
Attitude toward inter-personal violence - IPV	Women were asked if a husband is justified in hitting his wife for the following a woman's actions if she a) goes out without telling him; b) neglects the children; c) argues with him; d) refuses to have sex with him; e) burns the food. If they answer 'yes' to either of the above questions they were considered as accepting IPV; and as not accepting IPV otherwise

Domestic violence (composite index?)



Hypothetical distribution of sampling cluster percentages of women reporting domestic abuse (y-axis) and an independent variable (x-axis)



For **placed-based** data where locations with similar values are clustered, the relationships include y-y interactions in addition to y-x interactions, indicating:

- that the residuals are not random or independent of each other with constant variance (homoscedasticity), and
- That they, rather, show a systematic pattern

SPATIAL Autocorrelation
Global Moran's I

50	23	58	38
19	84	16	55
78	13	83	27
38	75	26	41

Dissimilar values
Negative

84	83	58	38
78	75	50	27
55	41	26	19
38	23	16	13

Similar values
Positive

19	83	84	13
38	55	58	26
50	41	38	75
16	78	23	75

Random pattern
No autocorrelation

Today's Topic: Women's Attitudes Towards Wife-beating

DATA SOURCE AND METHODOLOGY

- **Data source: Ethiopia's 2016 Demographic and Health Survey (DHS)**
  - **Data processing using R-programing language**
  - **Spatial statistics using geographically weighted regression (GWR) - an ArcGIS-Pro application**

R-PROGRAMMING

Ethiopia's 2016 DHS files

```
> #Couples' recode
> DHS2016_Couple <- read.dta("D:/DHS2016/ETCR71DT/ETCR71FL.DTA")
> nrow(DHS2016_Couple)
[1] 6141
> #Household recode
> DHS2016_HHold <- read.dta("D:/DHS2016/ETHR71DT/ETHR71FL.DTA")
> nrow(DHS2016_HHold)
[1] 16650
> #Individual recode
> DHS2016_Wom <- read.dta("D:/DHS2016/ETIR71DT/ETIR71FL.DTA")
> nrow(DHS2016_Wom)
[1] 15683
> #Births recode
> DHS2016_Births <- read.dta("D:/DHS2016/ETBR71DT/ETBR71FL.DTA")
> nrow(DHS2016_Births)##41392
[1] 41392
>
```

Q: V744c It is the respondent's opinion that a husband is justified in hitting or beating his wife when she argues with him

V744A	Beating justified if wife goes out without tell	57	1	N	I	1	0	No	No
	0 No								
	1 Yes								
	8 Don't know								
	(m) 9 Missing								
	(na) Not applicable								
V744B	Beating justified if wife neglects the children	58	1	N	I	1	0	No	No
	0 No								
	1 Yes								
	8 Don't know								
	(m) 9 Missing								
	(na) Not applicable								
V744C	Beating justified if wife argues with husband	59	1	N	I	1	0	No	No
	0 No								
	1 Yes								
	8 Don't know								
	(m) 9 Missing								
	(na) Not applicable								
V744D	Beating justified if wife refuses to have sex w	60	1	N	I	1	0	No	No
	0 No								
	1 Yes								
	8 Don't know								
	(m) 9 Missing								
	(na) Not applicable								
V744E	Beating justified if wife burns the food	61	1	N	I	1	0	No	No
	0 No								
	1 Yes								
	8 Don't know								
	(m) 9 Missing								
	(na) Not applicable								

R-programing (National)

```
DHS2016_Wom <- read.dta("D:/DHS2016/ETIR71DT/ETIR71FL.DTA")
> nrow(DHS2016_Wom)
[1] 15683 (number of women - unweighted)

> summary(DHS2016_Wom$v744c)
```

no	yes	don't know
9,839 (62.9%)	5,707 (36.4%)	137 (0.7%)

Q: V744c It is the respondent's opinion that a husband is justified in hitting or beating his wife when she argues with him

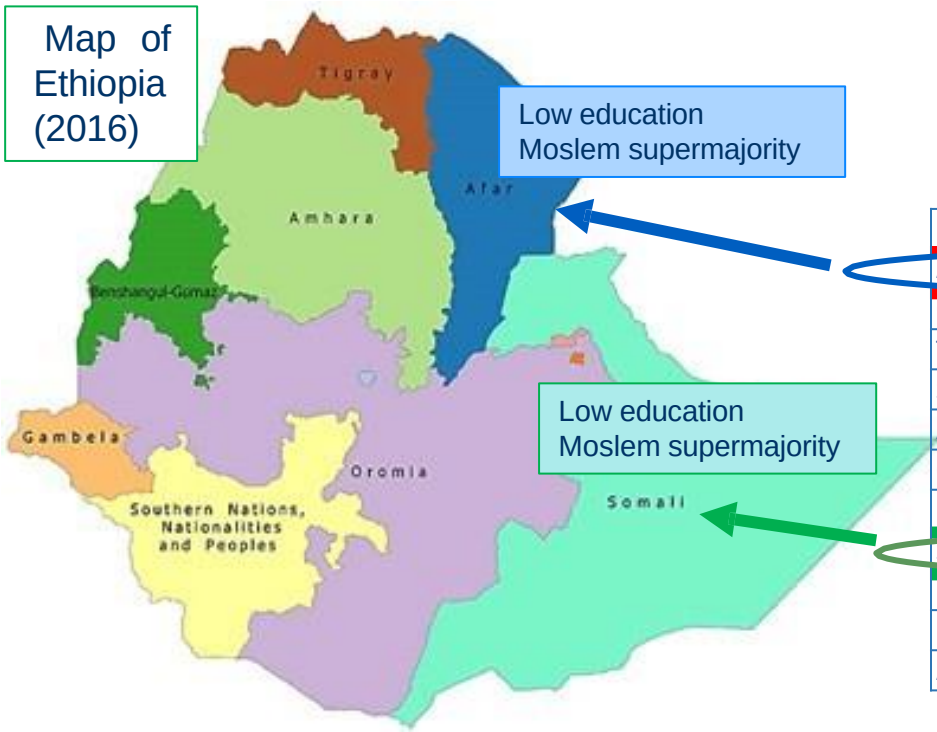
A. Yes

B. No

R-programing (Regional)

#Regional proportions of yes/no answers to beatings if wife argues with husband
WomArgReg <- table(DHS2016_Wom$v101,DHS2016_Wom$v744c) # WomArgReg # print table

Map of Ethiopia (2016)

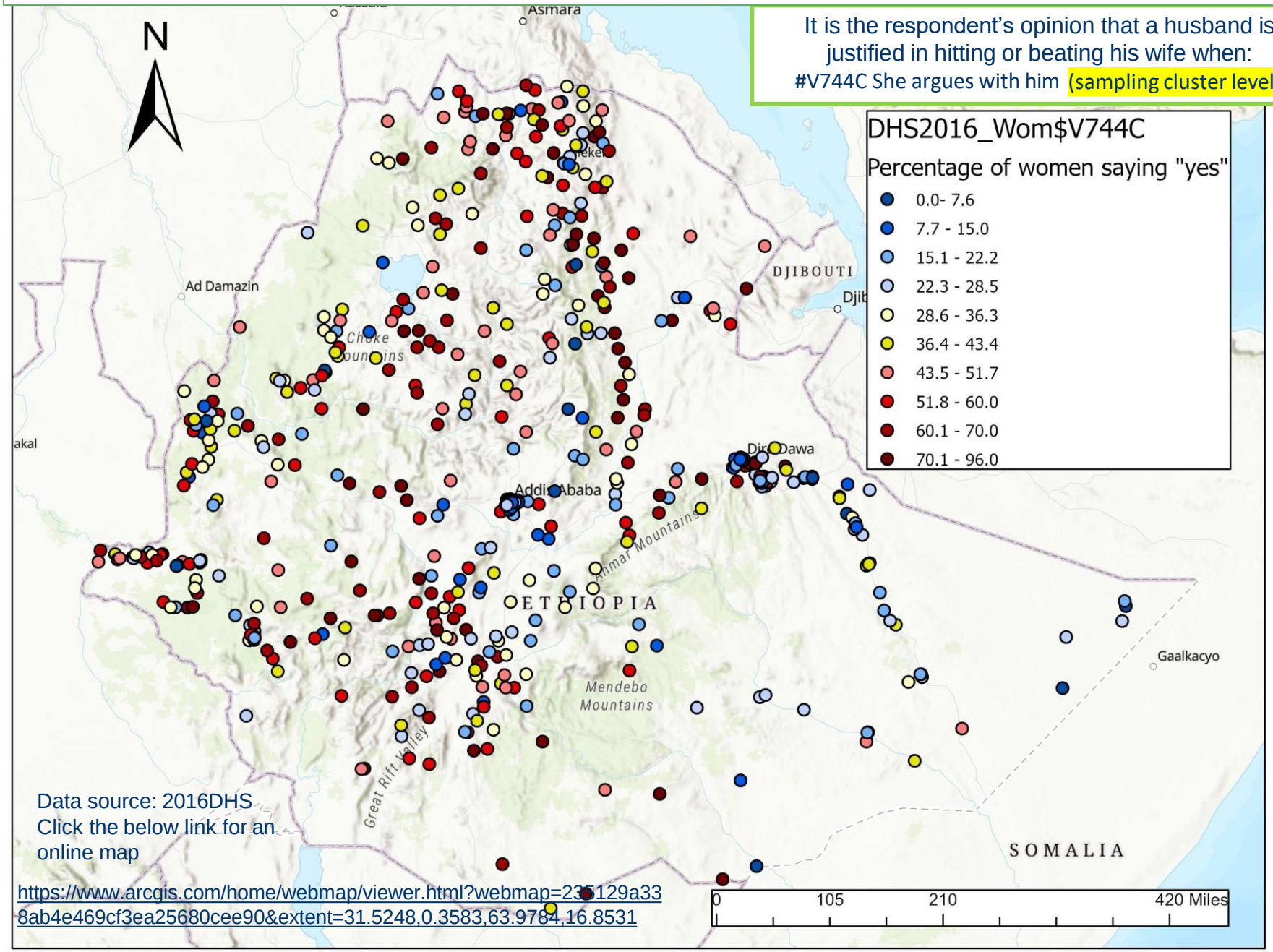


Regional percentage of women who answered yes/no to a question on wife-beating over spousal arguments

Region	No	% NO	Yes	% YES	Don't	%
Afar	531	47.1	582	51.6	15	1.3
Oromia	965	51.0	905	47.0	22	2.0
Tigray	894	53.2	783	46.4	5	0.4
SNNPR	1014	54.8	823	44.1	12	1.1
Amhara	980	57.0	725	41.7	14	1.2
Gambela	614	59.3	414	40.1	7	0.6
Benishang	725	64.4	400	35.5	1	0.1
Somali	913	65.6	446	31.5	32	2.8
Harari	654	72.2	243	27.0	9	0.8
Dire Dawa	879	77.7	239	21.1	13	1.2
Addis Aba	1670	91.6	147	7.8	7	0.6

Unweighted cluster percentages of women with a “yes” response to V744c

It is the respondent’s opinion that a husband is justified in hitting or beating his wife when:
#V744C She argues with him (sampling cluster level)



Data source: 2016DHS
Click the below link for an online map

https://www.arcgis.com/home/webmap/viewer.html?webmap=235129a338ab4e469cf3ea25680cee90&extent=31.5248,0.3583,63.9784,16.8531

Testing for the Presence of Spatial Autocorrelation

Two methodologies

1. The residuals vs. predicted plot

Reference: https://analyse-it.com/docs/user-guide/fit-model/linear/residual-plot#:~:text=A%20residual%20plot%20shows%20the,band%20around%20the%20identity%20line.

2. The Global Moran's I statistic

Reference: https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/h-how-spatial-autocorrelation-moran-s-i-spatial-st.htm

TEST-1, Residual vs. Predicted Plot

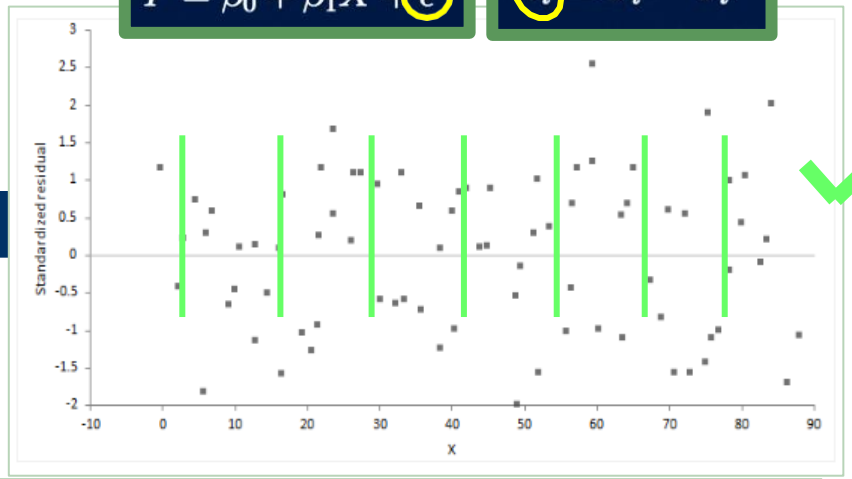
Assumptions of Linear Regression

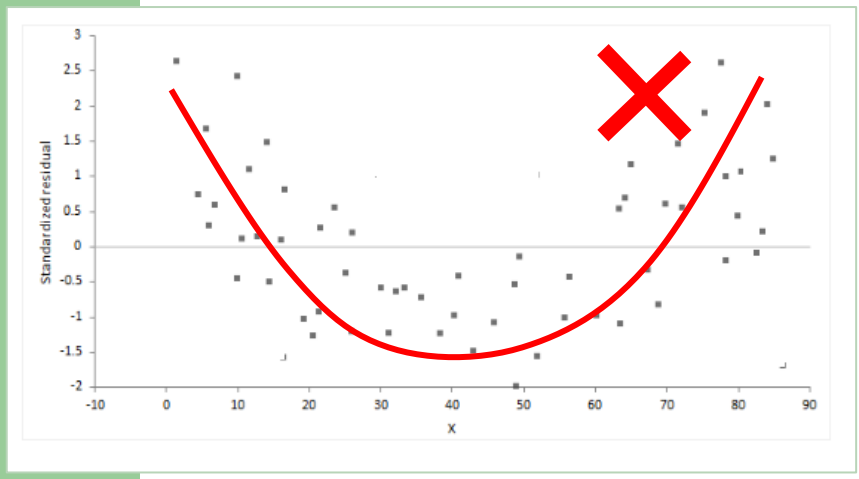
1.A linear relationship: x and y are linearly related
2.Independence: The residuals are random and independent of one another
3.Homoscedasticity: Constant variance of residuals against variable X or predicted Y and forming an approximately constant width band around the identity line
4.Normality: The residuals of the model are normally distributed

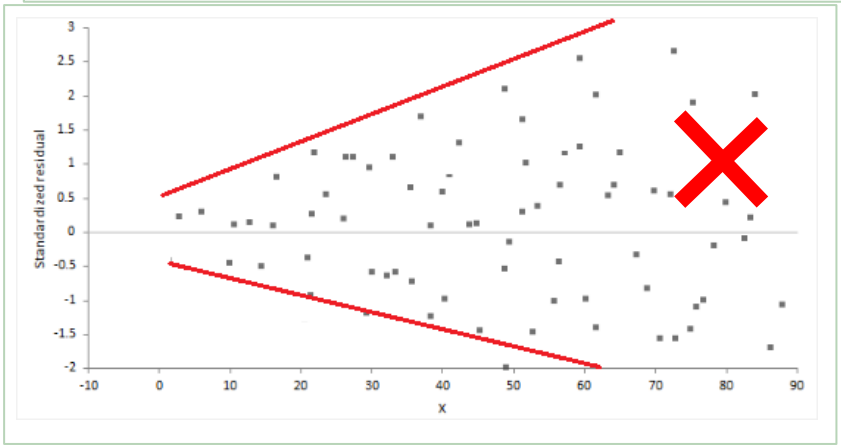
Source: https://www.statology.org/linear-regression-assumptions/Assumptions

$$Y = \beta_0 + \beta_1 X + \epsilon$$

$$e_i = Y_i - \hat{Y}_i$$

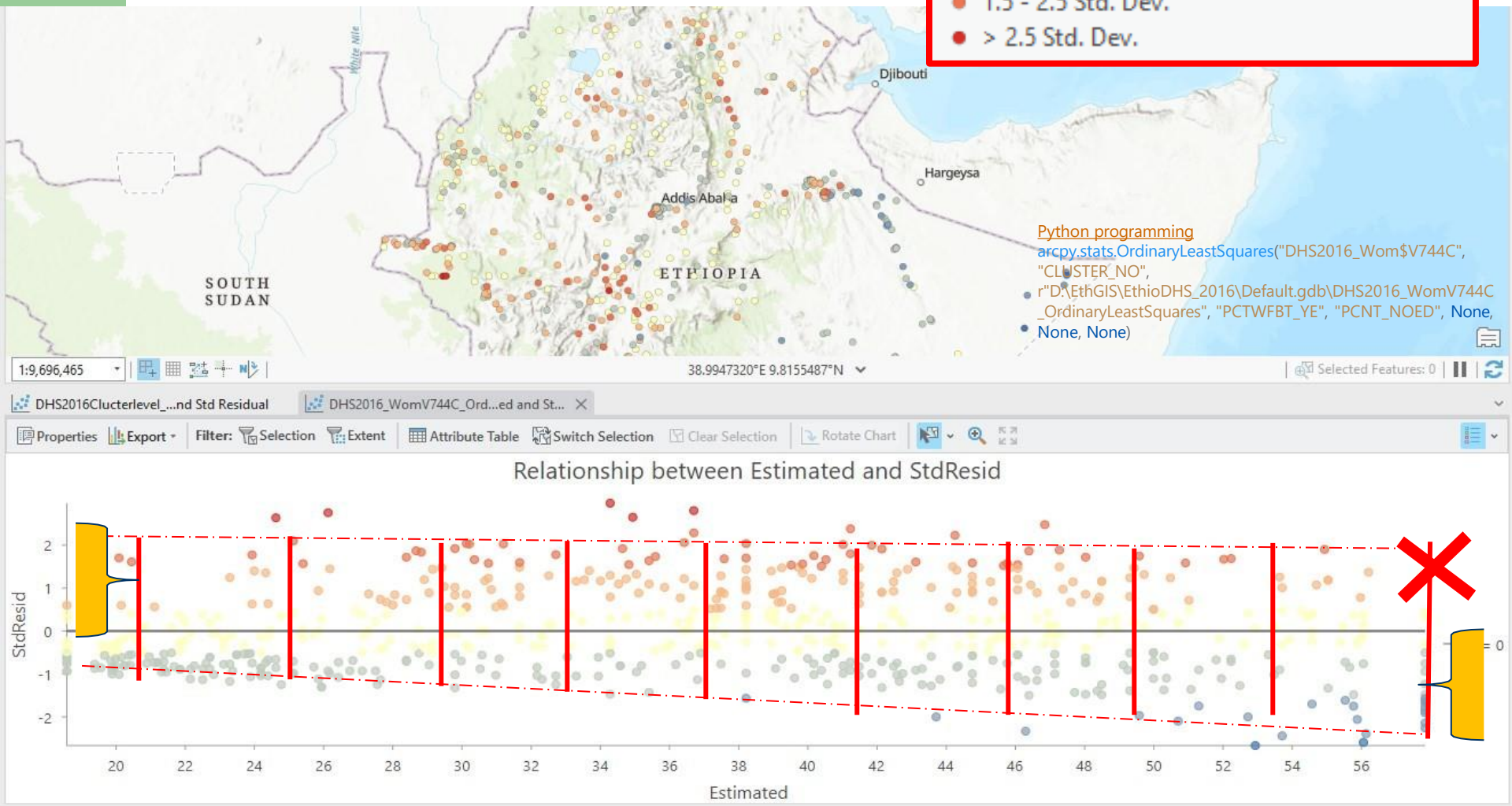






Source: Adopted from https://analyse-it.com/docs/user-guide/fit-model/linear/residual-plot

Ordinary Least Square Regression



TEST: 2The Global Moran's I statistic

Conduct a Global Moran's I test of unweighted percentages of women with a "yes" response to Q: V744c

Global Moran's I

Global Moran's *I* is a measure of the overall clustering of the spatial data. It is defined as

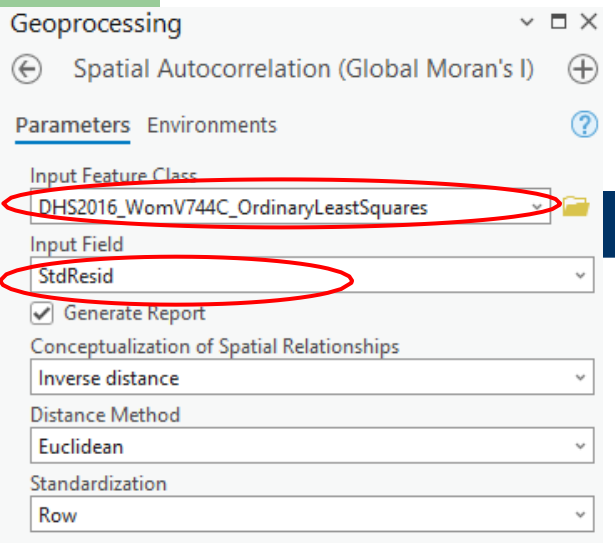
$$I = \frac{N}{W} \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

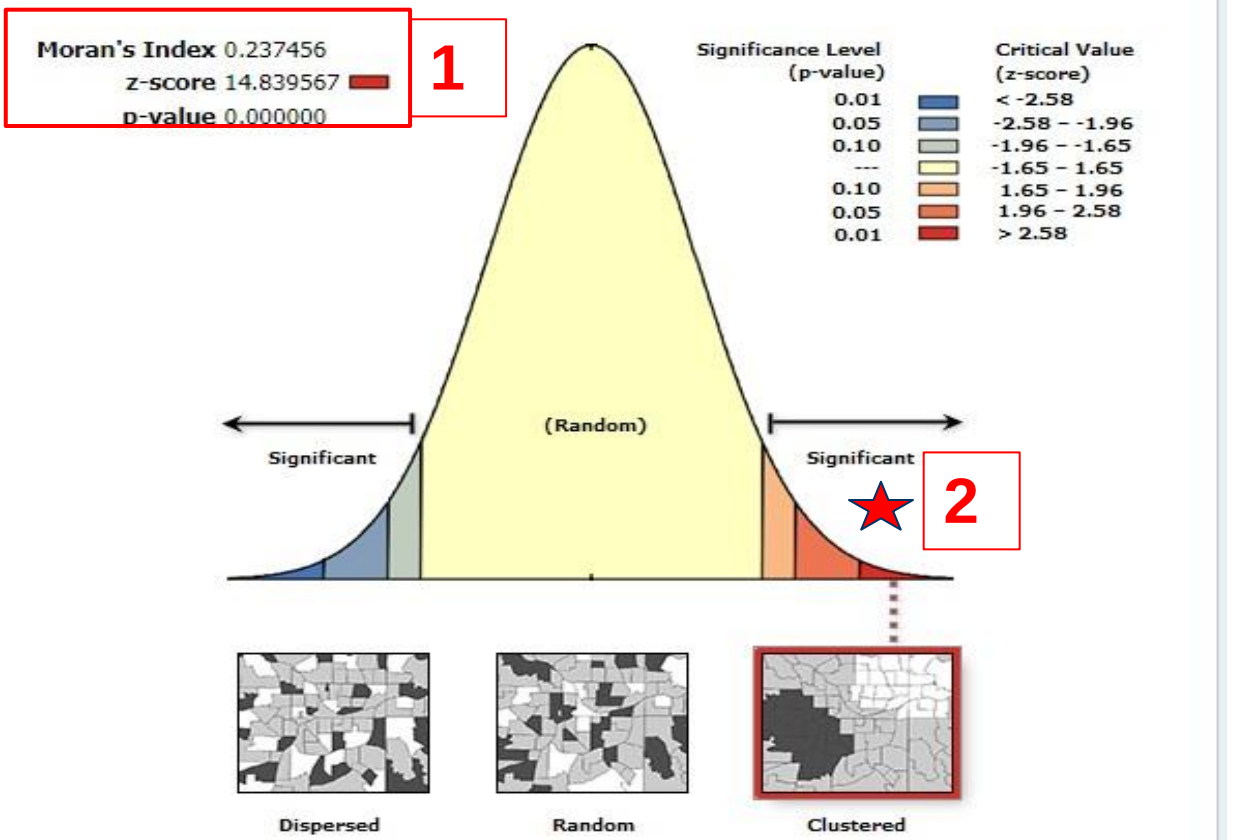
where

- N is the number of spatial units indexed by i and j;
- x is the variable of interest;
- $\bar{x}$ is the mean of x;
- w_{ij} is a matrix of spatial weights with zeroes on the diagonal (i.e., $w_{ii} = 0$);
- and W is the sum of all w_{ij} (i.e. $W = \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}$).

*Source: http://www.50northspatial.org/global-morans-i-spatial-autocorrelation/

II. Test of clustering (spatial autocorrelation): % saying "yes" to wife-beating





3 Given the z-score of 14.839567, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

Global Moran's I Summary

Moran's Index	0.237456
Expected Index	-0.001558
Variance	0.000259
z-score	14.839567
p-value	0.000000

Python programming

```
arcpy.stats.SpatialAutocorrelation("DHS2016_WomV744C_OrdinaryLeastSquares", "StdResid", "GENERATE_REPORT", "INVERSE_DISTANCE", "EUCLIDEAN_DISTANCE", "ROW", None, None, None)
```

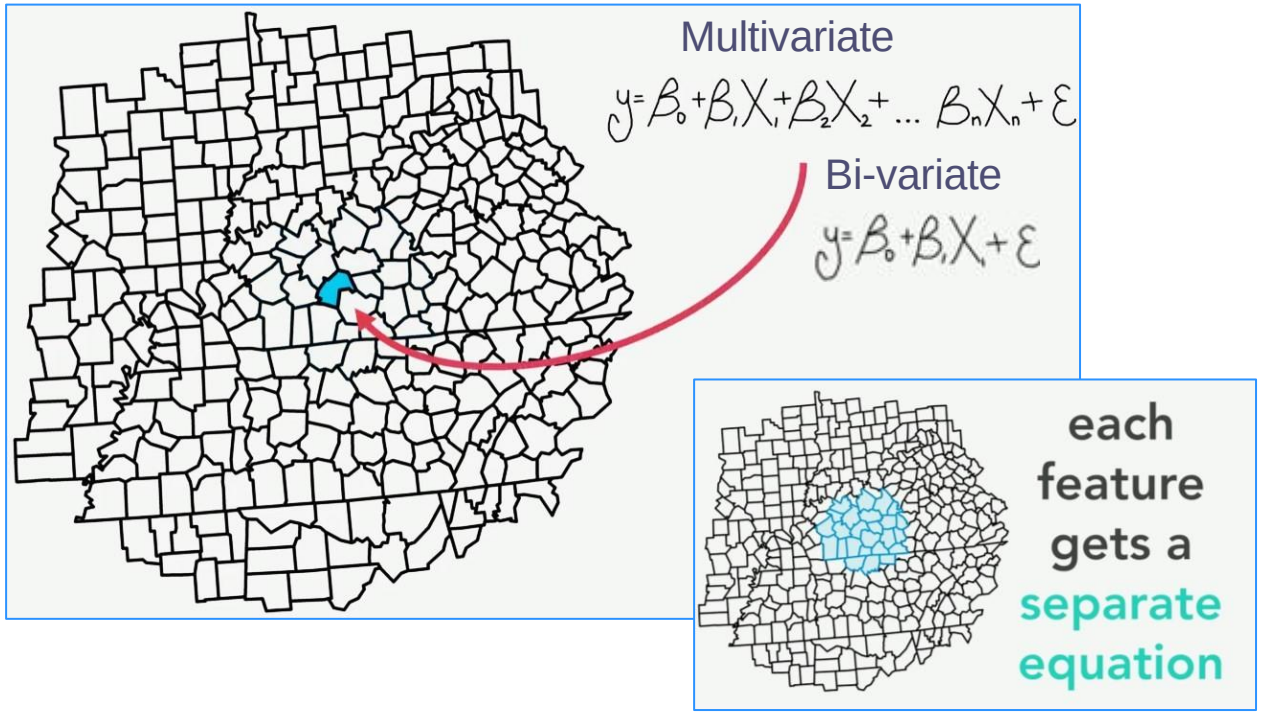
RESULT

The dependent variable, V744c is not suitable for global ordinary least squares (OLS) regression due to spatial autocorrelation shown by the non-random (or clustered) distribution of residuals

SOLUTION: Geographically Weighted Regression - GWR

https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1538-4632.1996.tb00936.x

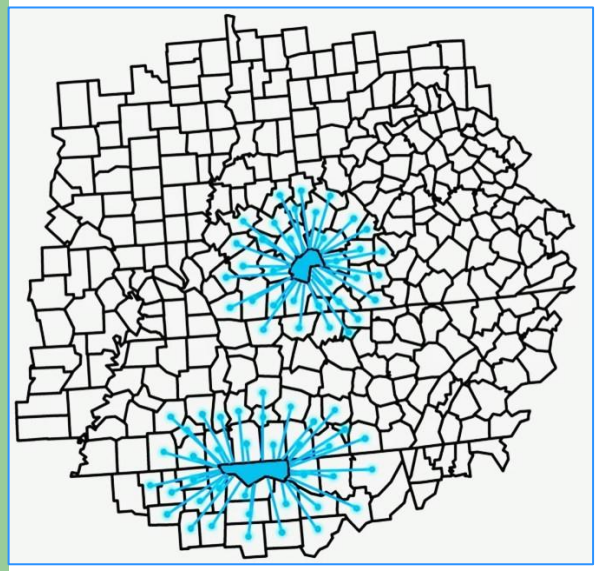
With GWR, each cluster is grouped with its immediate neighbors and is given a separate equation*



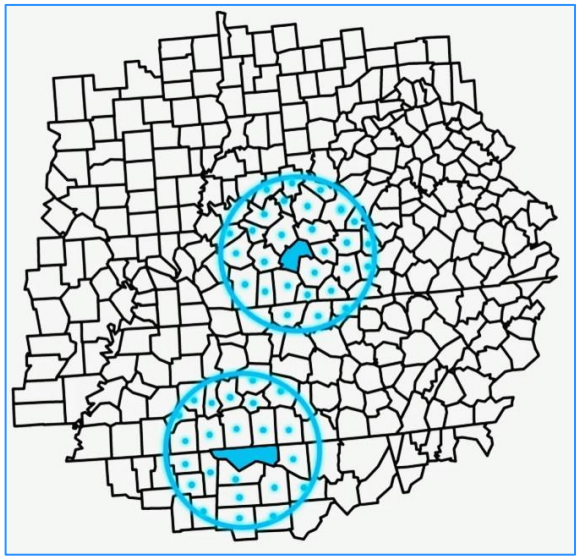
*Source: ESRI https://www.youtube.com/watch?v=ob18SuuAJGI

GWR contd.
Defining local

Number of neighbors



Distance band



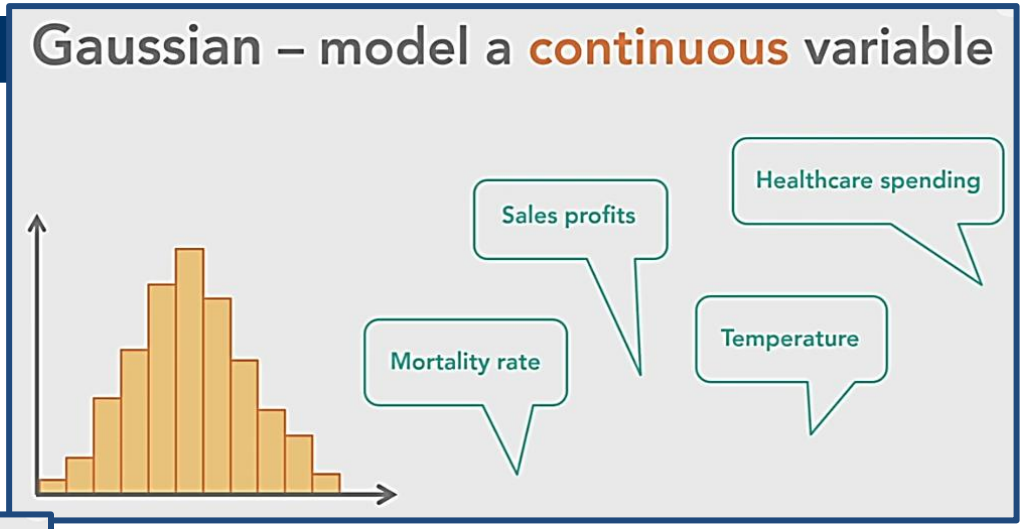
Golden search*

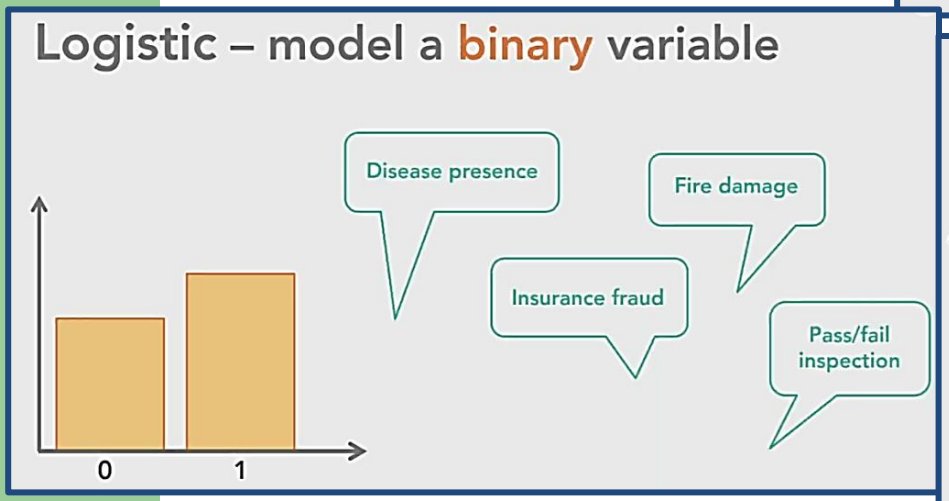
*Software finds an optimum distance by increasing distances by thirds (golden search) until the best local minima are found

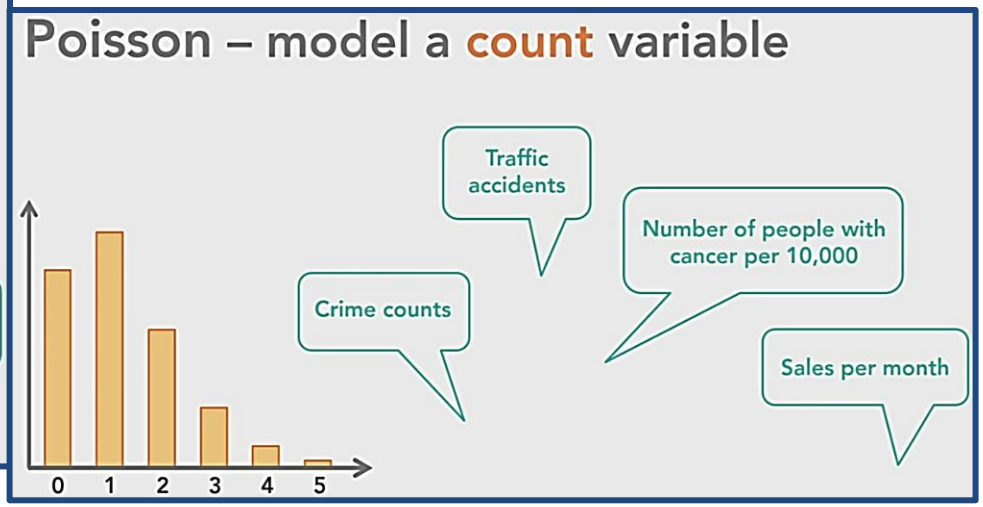
Source: ESRI https://www.youtube.com/watch?v=ob18SuuAJGI

GWR contd.
Three Model Types

- Gaussian – continuous
- Logistic – binary
- Poisson – count







Source: ESRI https://www.youtube.com/watch?v=ob18SuuAJGI

GWR contd. variables

Dependent Variable (y)

V744C	Beating justified if wife argues with husband	59	1	N	I	1	0	No	No

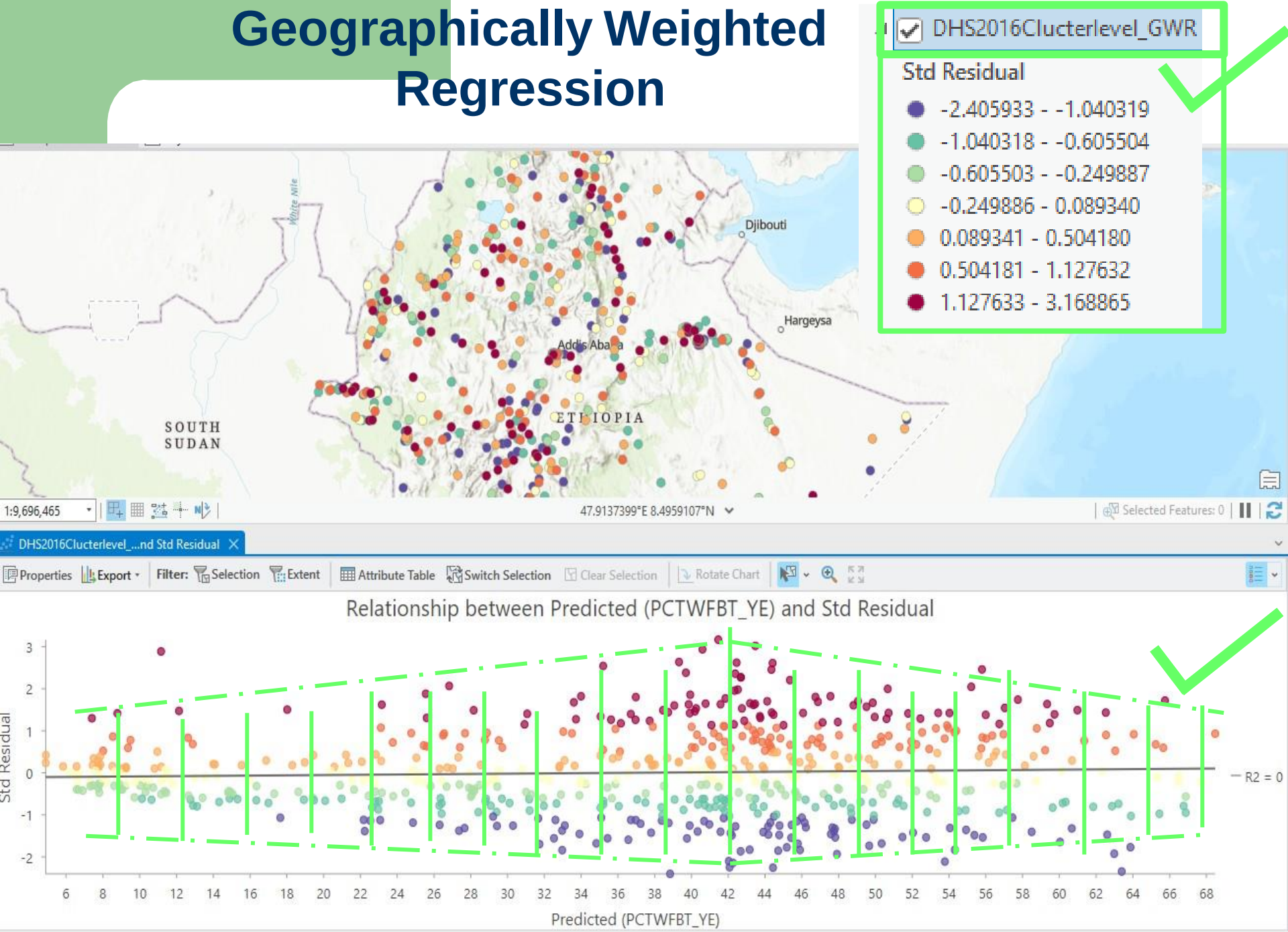
(na) Not applicable
0 No
1 Yes
8 Don't know
(m) 9 Missing

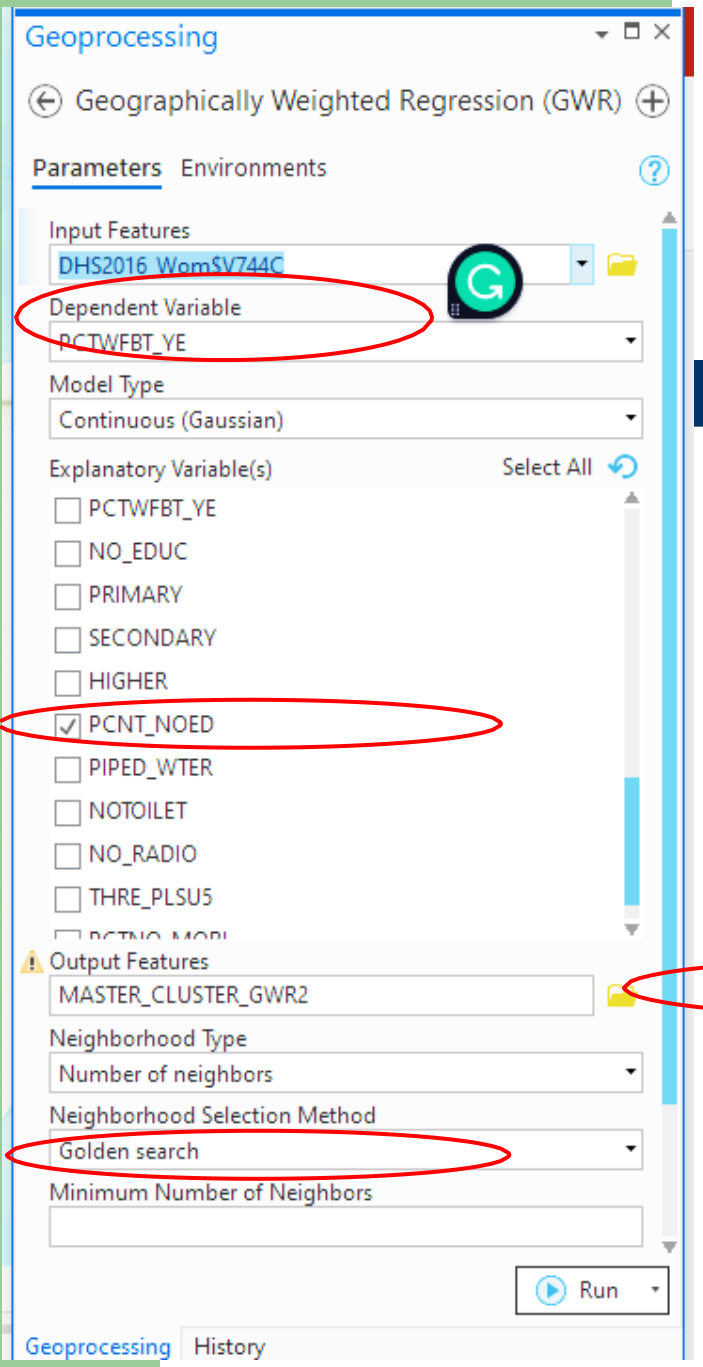
Independent Variable (x)

HV109 Educational attainment

0 No education
1 Incomplete primary
2 Complete primary
3 Incomplete secondary
4 Complete secondary
5 Higher
8 Don't know

Geographically Weighted Regression





GWR ….contd.

Results

v744C (y) vs. education (x)

WARNING 110306: The final model didn't have the lowest AICc encountered in the Golden Search Results.
WARNING 110259: At least one local regression had very limited variation after applying the weights. Use caution when interpreting the results.

-------- Analysis Details -------
Number of Features: 643
Dependent Variable: PCTWFBT_YE
Explanatory Variables: PCNT_NOED
Number of Neighbors: 72

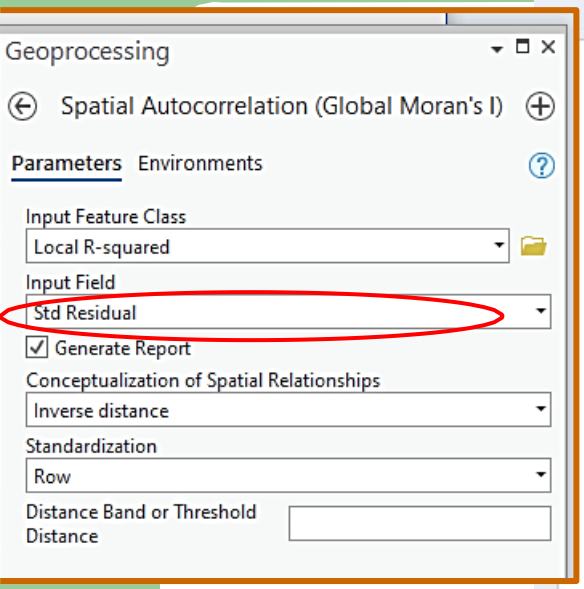
---------- Model Diagnostics ----------
R2: 0.5123
AdjR2: 0.4585
AICc: 5489.8314
Sigma-Squared: 280.1931
Sigma-Squared MLE: 252.3868
Effective Degrees of Freedom: 579.1888

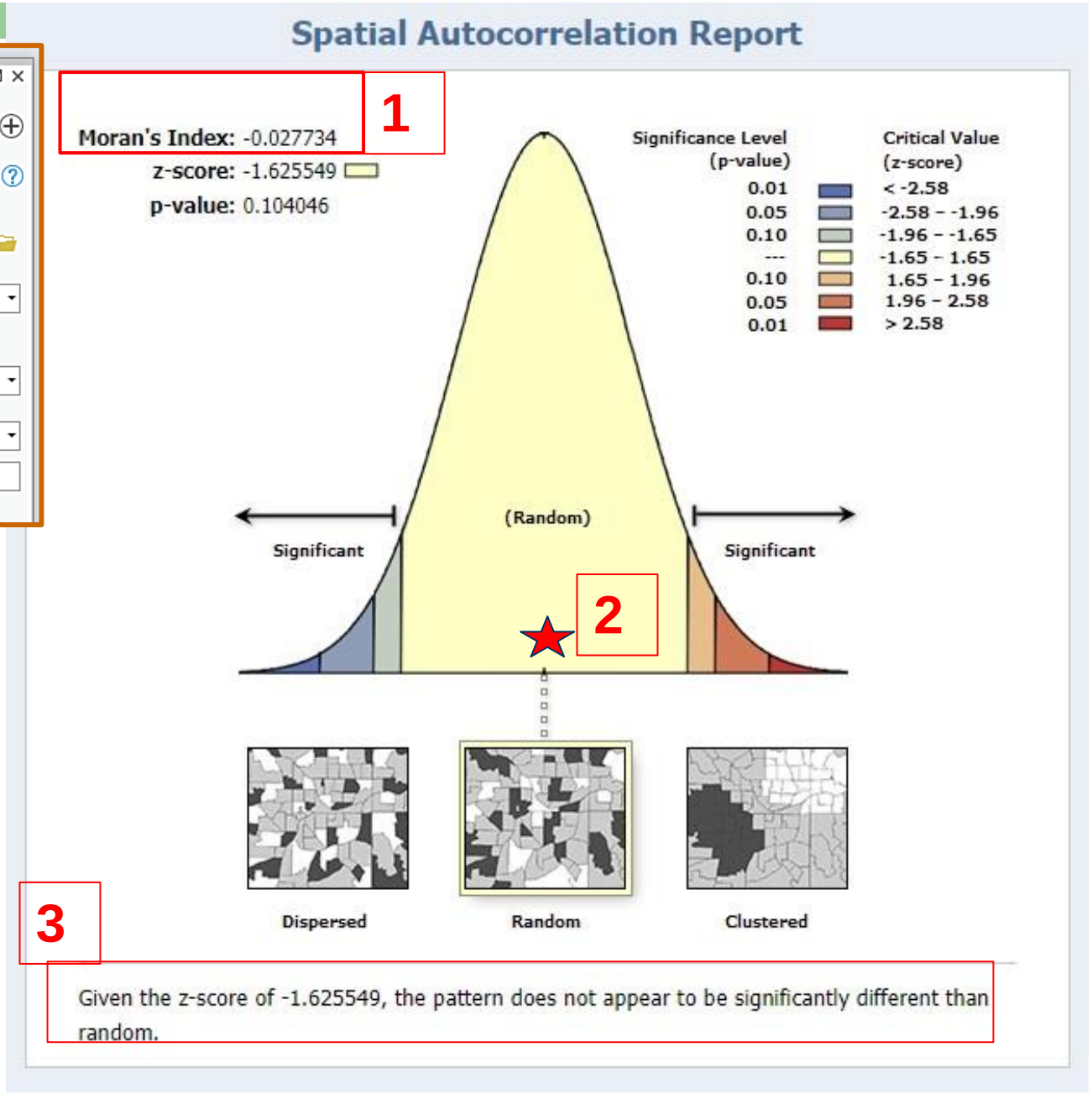
Succeeded at Friday, September 23, 2022 7:26:15 AM (Elapsed Time: 6.19 seconds)

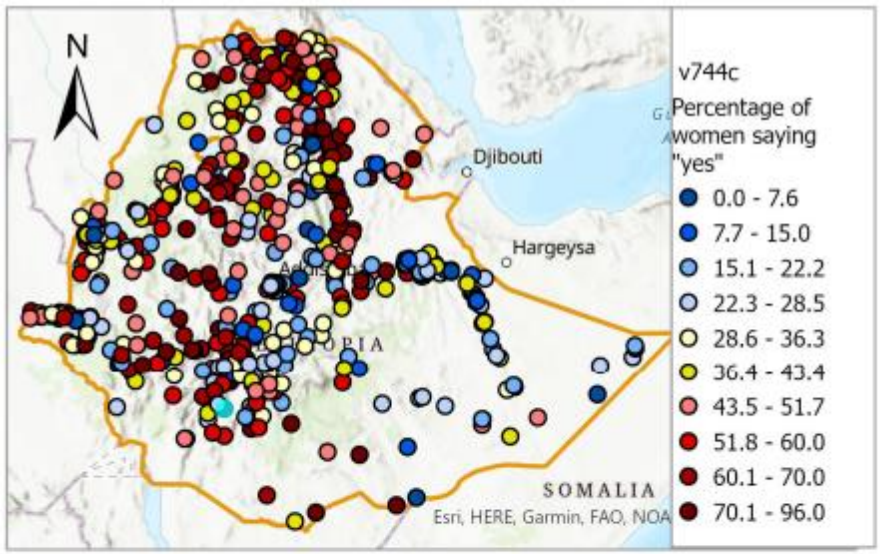
-- Golden Search Results

#Neighbors	AICc
30	5514.0940
643	5667.8465
264	5548.8335
409	5569.4515
175	5524.7560
119	5500.4482
85	5491.6356
64	5492.1195
98	5493.9578
77	5489.8220
72	5489.8314

The assumption of independent residuals was not violated this time







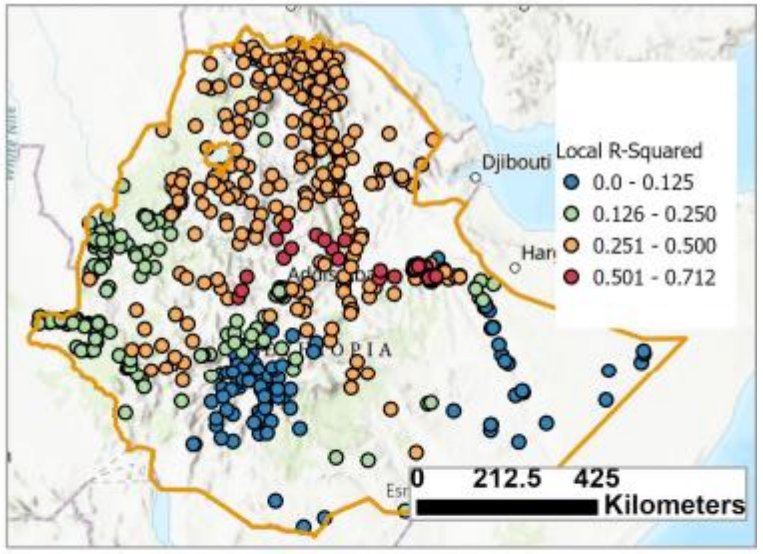
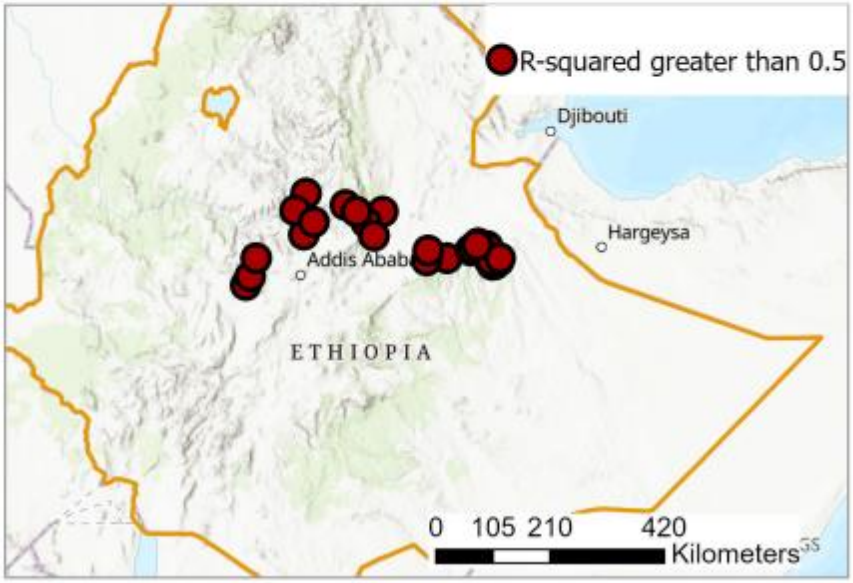


Figure 8A. percentage of women answering "yes" to a question on the acceptability of wife-beating

Figure 8B. Local R-squared from geographically weighted regression (GWR)



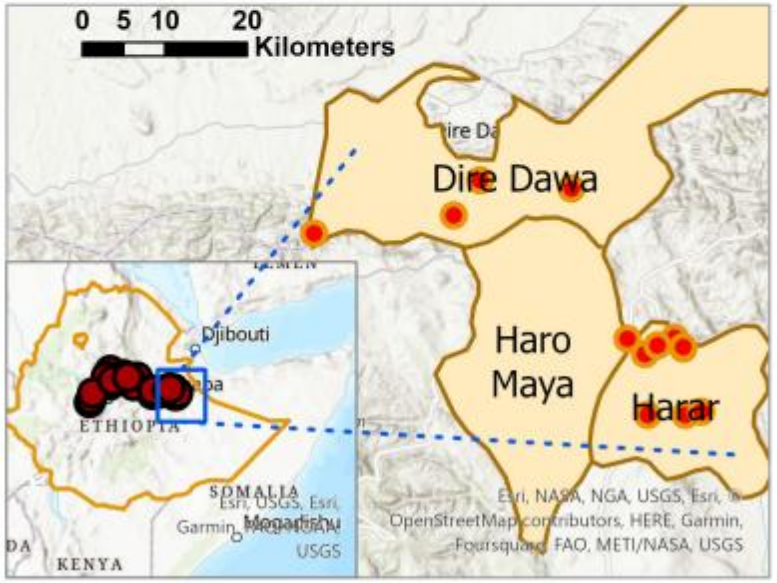


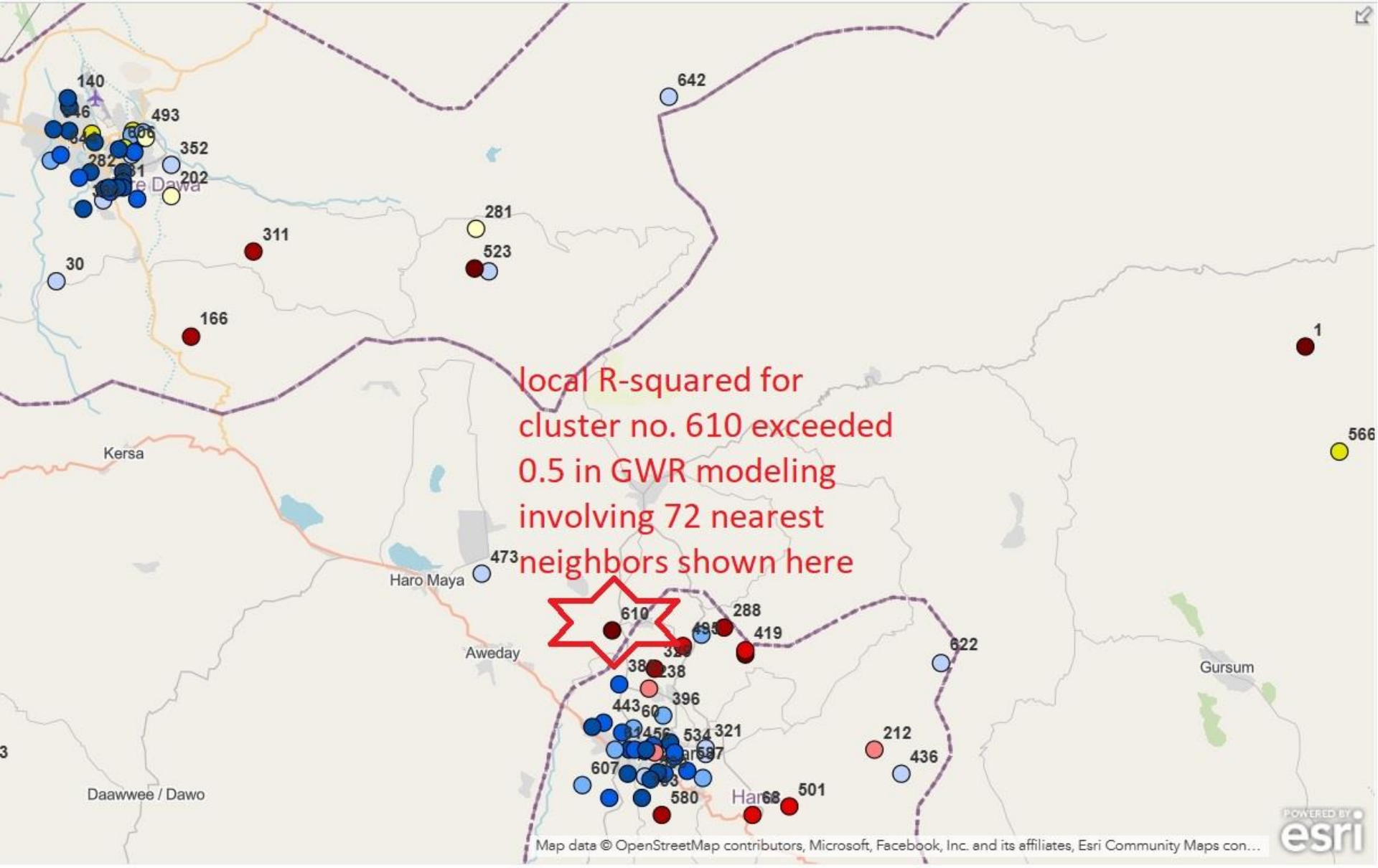
Figure 9A. Local R-squared >0.5, geographically weighted regression (GWR)

Figure 9B. Final selection of sampling clusters

Table 2. Results of the Geographically Weighted Regression (GWR) for 12 Sampling Clusters Meeting Study Criteria

Cluster No	*Wereda*	*% Wife-beating okay (y)*	*%No education (x)*	*Intercept (a)*	*Coeff (β)*	*Predicted (ŷ)*	*Residual*	*Std Res*	*Local R^2*	*CND* number*
68	Harar	60.0	60.0	6.8	0.58	41.7	18.3	1.1	0.55	17.2
580	Harar	64.0	76.0	6.4	0.60	51.9	12.1	0.7	0.56	16.6
501	Harar	54.2	87.5	7.0	0.58	57.4	-3.2	-0.2	0.54	17.6
329	Harar	66.7	54.2	6.1	0.62	39.8	26.8	1.6	0.57	16.8
357	Harar	80.0	60.0	6.6	0.60	42.7	37.3	2.3	0.55	17.6
495	Harar	60.0	60.0	6.2	0.62	43.3	16.7	1.0	0.57	17.1
419	Harar	56.0	52.0	6.6	0.60	37.9	18.1	1.1	0.55	17.6
610	Haro M.	92.6	92.6	5.6	0.65	65.7	26.9	1.7	0.59	17.0
288	Harar	66.7	66.7	6.5	0.61	47.1	19.5	1.2	0.56	17.6
441	Dire D.	54.2	91.7	4.9	0.51	51.3	2.9	0.2	0.51	17.0
166	Dire D.	64.3	71.4	4.9	0.54	43.7	20.5	1.3	0.52	16.1
311	Dire D.	63.6	63.6	4.8	0.56	40.2	23.5	1.5	0.51	15.7

*CND (Condition number) indicates the level of local autocorrelation of the independent (x) variable; CND <30 is desirable.


local R-squared for cluster no. 610 exceeded 0.5 in GWR modeling involving 72 nearest neighbors shown here
140
493
352
202
642
281
523
311
30
166
1
566
473
610
288
419
622
396
534
321
212
436
607
580
501
Kersa
Haro Maya
Aweday
Gursum
Daawwee / Dawo
Map data © OpenStreetMap contributors, Microsoft, Facebook, Inc. and its affiliates, Esri Community Maps con…
esri

SUMMARY

- The geographically weighted regression method (GWR) corrected for potential methodological flaws of a global least squares regression by producing local results that are unaffected by spatial autocorrelation
- The GWR model generated a list of twelve sampling clusters where most women respondents stated that wife-beating was acceptable while admitting to having had no formal education, and where the local R-squared exceeded 0.5 in GWR modeling involving 72 nearest neighbors per sampling cluster.
- Although the majority of the twelve sampling clusters are in Harari *Wereda* and *Kilil*, which got their name from members of the Harari ethnic group that are predominantly Muslim, it is difficult to pinpoint which factor or set of factors can be cited as causally associated with characteristics that placed them on the list.
- Other explanatory variables, including urban-rural status, income, and DHS' Wealth Index, were omitted from the analysis due to the risk of multicollinearity
- Policy interventions aimed at protecting women against spousal abuse in Ethiopia need to be framed with local (not national or regional) sets of preventative measures in mind

*Waldo Tobler, First law of geography, "Everything is related to everything else. But near things are more related than distant things."
https://gisgeography.com/tobler-first-law-of-geography/

Journal of Geography and Geology; Vol. 15, No. 2; 2023 ISSN 1916-9779 E-ISSN 1916-9787 Published by Canadian Center of Science and Education

Click below

The Social Geography of Women's Attitudes Toward Wife-beating in Ethiopia: A Contribution Towards Proper Application of Spatial Statistics