

GEOG0114: PRINCIPLES OF SPATIAL ANALYSIS

WEEK 10: SPATIAL MODELS (PART 2)

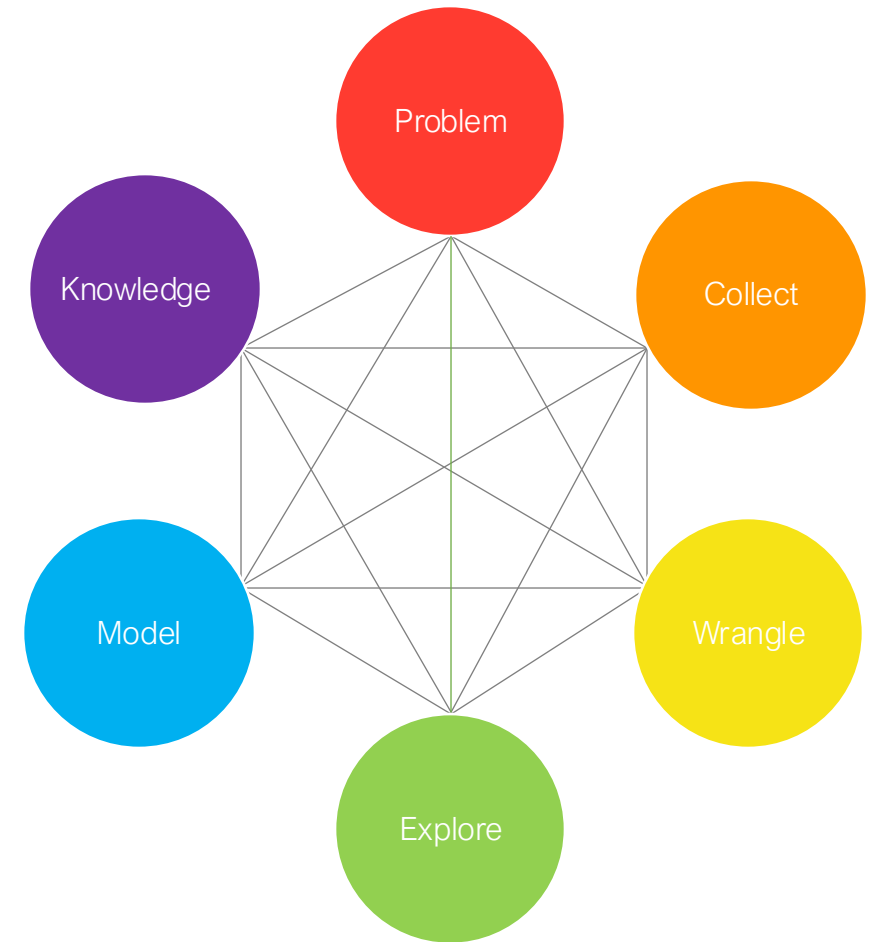
Dr Anwar Musah (a.musah@ucl.ac.uk)

Lecturer in Social and Geographic Data Science

UCL Geography

Contents

1. Introduction to Geographically Weighted Regression Modelling
2. Extending the standard linear regression to GWR
 - Using GWRs to estimate the local (not global) relationships between a dependent and independent variable
 - Determination of whether local relationships are statistically significant or not
 - Model performance through Local R-squared
3. Methodology for statistical analysis and interpretation
4. Overarching summary of the GEOG0114



Spatial Models

Week 9

Spatial Lag and Error Models

In Week 9, we learnt special types of spatial model which accounts for spatial configuration of areal data

Week 10

Geographically Weighted Regression (GWR) Models

Recap

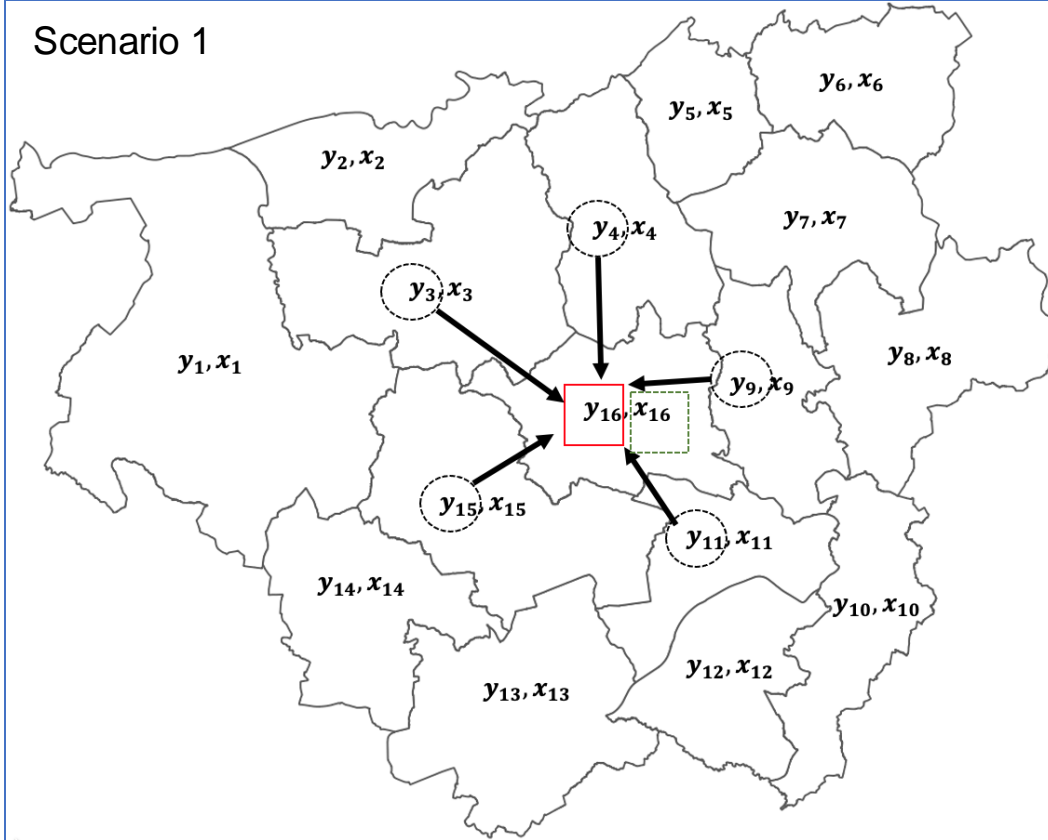
Multivariable Linear Regression

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$$

1. Spatial Lag Model (lagged on the dependent variable)

$$y = \rho WY + \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$$

Scenario 1

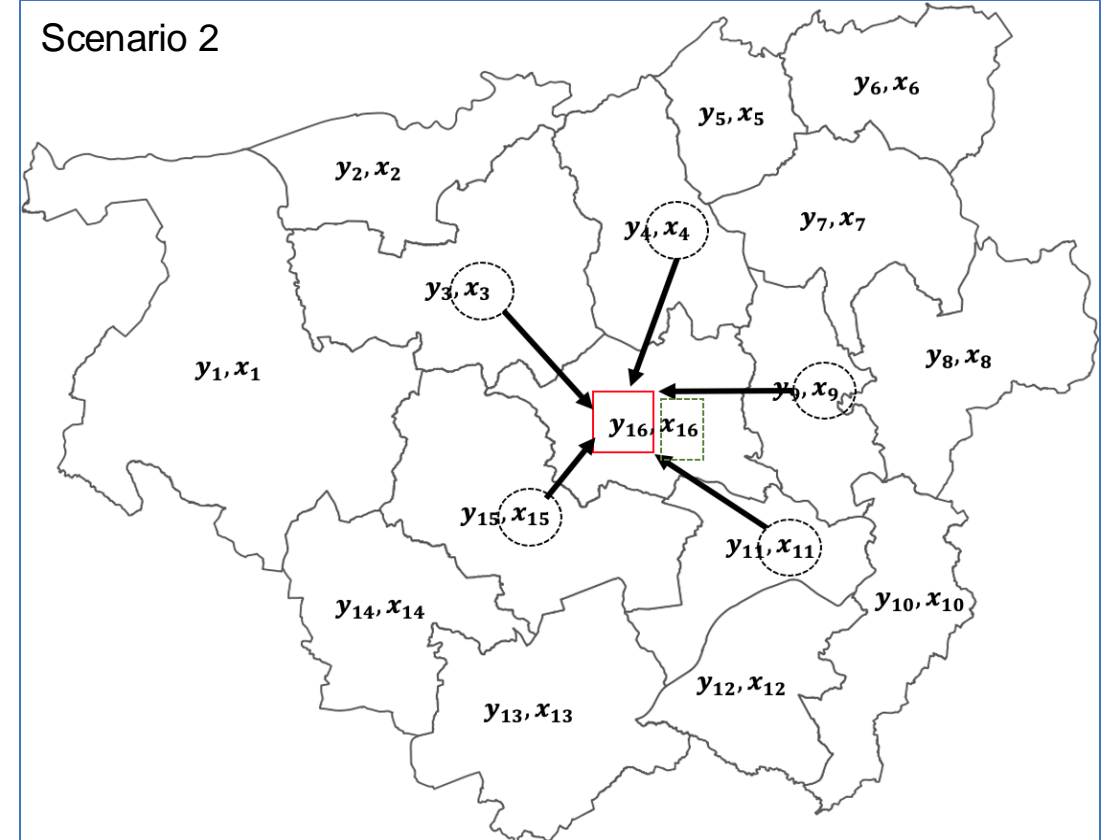


- W is a spatial weights matrix (contiguity – based)
- Y in the right hand of the equation represents the observed outcome from other areas neighbouring that influences what we're trying to predict
- ρ "Rho" is the degree of how our predicted outcome are influenced by its neighbouring Y measures.

2. Spatial Lag Model (lagged on the independent variable)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \theta WX + \varepsilon$$

Scenario 2



- W is a spatial weights matrix (contiguity – based)
- X in the right hand of the equation represents the observed values from the independent variable in other areas neighbouring that influences what we're trying to predict
- θ "theta" is the degree of how our predicted outcome are influenced by its neighbouring X measures.

Recap

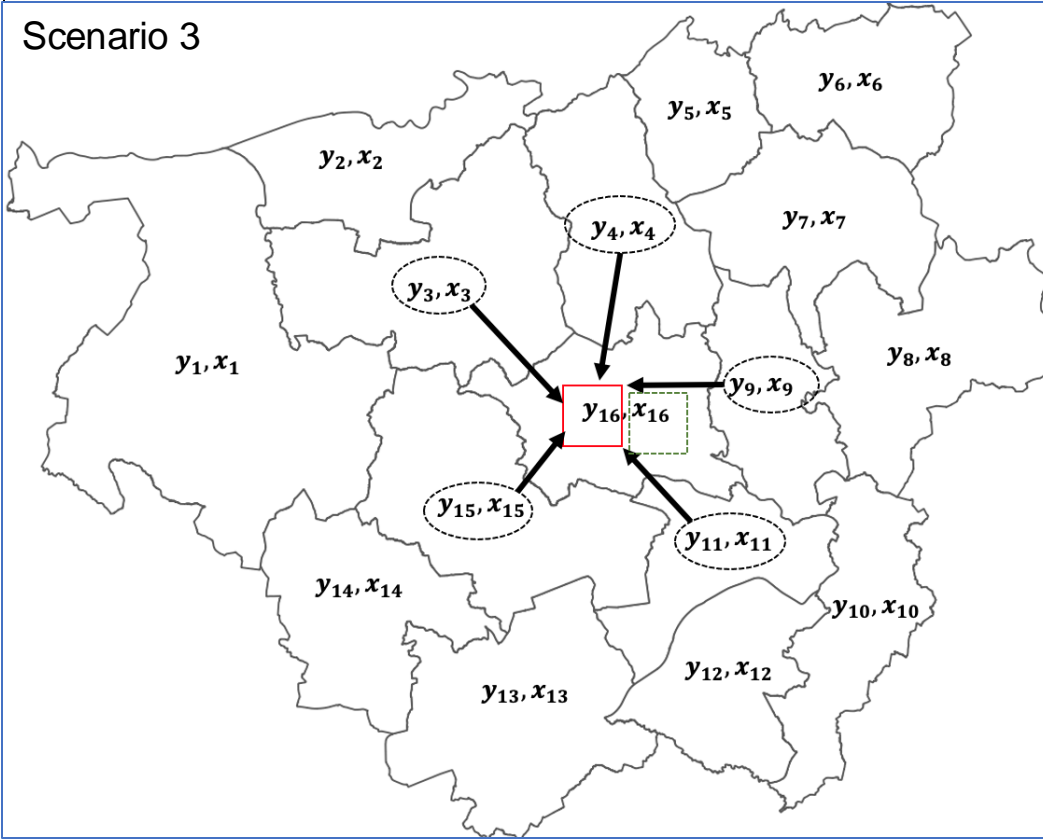
Multivariable Linear Regression

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$$

3. Spatial Lag Model (lagged on both the dependent and independent variable)

$$y = \rho WY + \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \theta WX + \varepsilon$$

Scenario 3

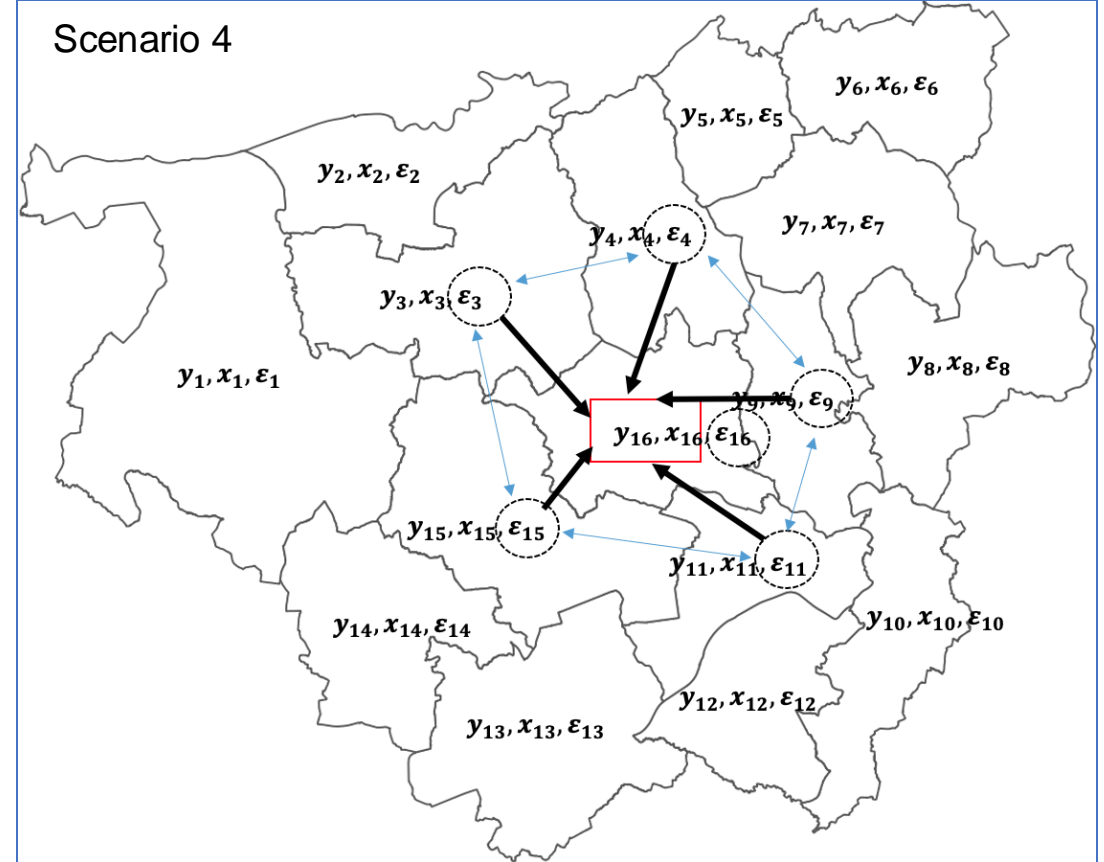


- Refer to slide number 3 to see the meanings of each parameters

4. Spatial Error Model

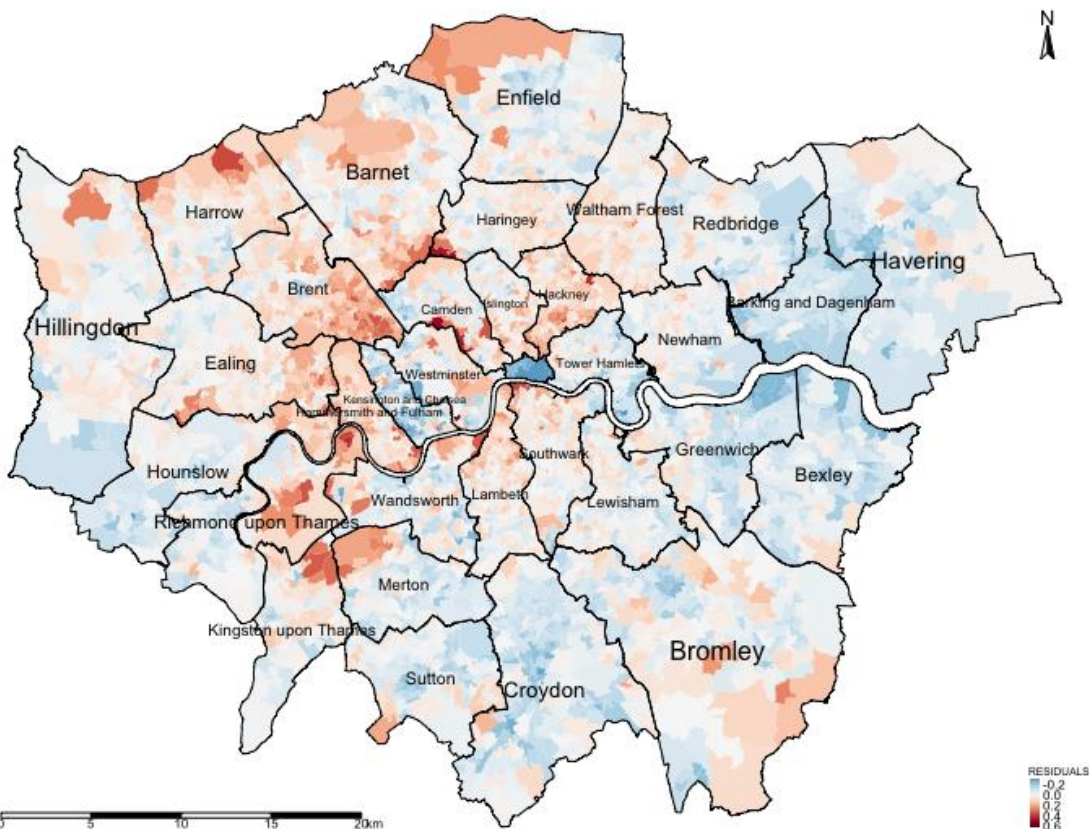
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \lambda Wu + \varepsilon$$

Scenario 4



- u are the correlated spatial error terms
- λ "lambda" estimated coefficient for the product W and u .

Recap



Examining the residuals for determining spatial patterning and evidence of spatial autocorrelation. Moran’s I test was 0.475 ($p < 0.001$) meaning that the residuals are clustered.

Broadly, there’s an over-estimation in the house-price and spatial aspects needs to be accounted for.

Modelled results from Linear and Spatial Lag (Y) regression model

Variable(s)	Linear Model	Lag (Y) Model (Total effects)
log(Income)	2.036*	2.217*
log(Deprivation)	0.136*	0.079*
log(PTAL)	0.031*	0.019*
AIC	-8510.8	-9863.3
R ²	0.7889 (78.89%)	N/A

- While we have accounted for spatial configuration using the spatial model, as well as accounted for spatial autocorrelation, we were able to determine the **Global** relationships between dependent and independent variables.
- What about if we want to investigate further patterns but a much **local-level**?
- LM, and any of the spatial Lag and error models cannot solve this problem

Spatial Models

Week 9

Spatial Lag and Error
Models

Week 10

Geographically Weighted
Regression (GWR)
Models

What is a Geographically Weighted Regression

Definition of Geographically Weighted Regression (GWR) model:

GWR is a statistical model which can indicate where non-stationarity may take place across space; it can be used to identify how **locally** weighted regression coefficients may vary across the study area (unlike its counterpart i.e., the **Spatial Lagged and/or Error Models** which provides **global coefficients**)

We use GWRs to:

- 1) Determine **area-specific relationships** or local **associations** between a specified **outcome** (i.e., **dependent variable**) with one or more **predictors** (i.e., **independent variable(s)**)
- 2) Find out whether those area-specific relationship or local associations are statistically significant across geographic space.

GWRs fall under the family of linear regression models. Recall last week the various model types and families?

Here is a board overview:

Distribution of dependent variable	Suitable Model
Continuous measures: e.g., average income in postcode (£); concentrations of ambient particular matter (PM2.5); Normalised Vegetative Difference Index (NDVI) etc.,	Linear regression
Binary measures (1 = “present” or 0 = “absent”): e.g., Person’s voting for a candidate, Lung cancer risk, house infested with rodents etc.,	Logistic Regression
Binomial measure (or proportion): e.g., prevalence of houses in a postcode infested with rodents, percentage of people in a village infected with intestinal parasitic worms, prevalence of household on a street segment victimised by crime etc.,	Logistic Regression
Counts or discrete measures: e.g., number of reported burglaries on a street segment, number of riots in a county etc.,	Poisson Regression
Time-to-event binary measures: e.g., Lung cancer risk due to chronic exposure to environmental levels of indoor radon. Risk of landslide and time dependence of surface erosion etc.,	Survival Analysis with Cox regression

Multivariable Linear Regression Model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \varepsilon$$

Variables

- y is the dependent variable
- $x_1, x_2, x_3, \dots, x_k$ are the independent variables

Parameters

- β_0 is the intercept
- $\beta_1, \beta_2, \beta_3, \dots, \beta_k$ are the slopes (or coefficients) for the corresponding variables $x_1, x_2, x_3, \dots, x_k$
- ε is the error term

Geographical Weighted Regression Model

$$y_i = \beta_{i,0}(u_i, v_i) + \beta_{i,1}(u_i, v_i)x_{i,1} + \beta_{i,2}(u_i, v_i)x_{i,2} + \cdots \beta_{i,k}(u_i, v_i)x_{i,k} + \varepsilon_i$$

Variables

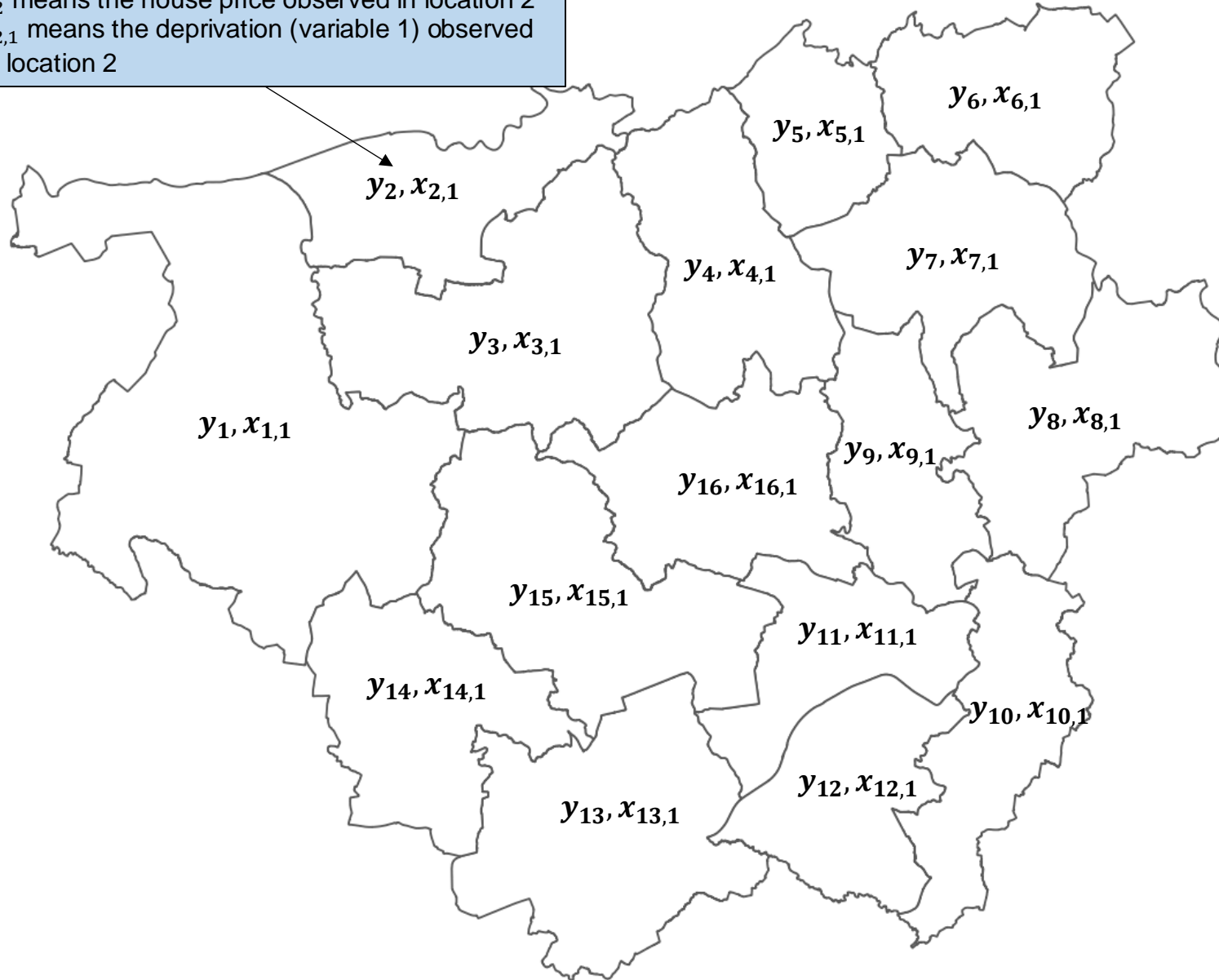
- y_i is the dependent variable indexed at observation (or location) i
- x_{i1}, x_{i2} and so to x_{ik} are the k-number of independent variables indexed at at

Parameters

- $\beta_{i0}(u_i, v_i)$ is the intercept as a function of a geographic location (i.e., coordinates on a grid)
- $\beta_{i1}(u_i, v_i), \beta_{i2}(u_i, v_i), \beta_{i3}(u_i, v_i), \dots, \beta_{ik}(u_i, v_i)$ are the slopes (or coefficients) for the corresponding variables $x_{i1}, x_{i2}, x_{i3}, \dots, x_{ik}$ which are function of a geographic location (u_i, v_i)
- ε is the error term

Suppose we have a hypothetical study area with 16 areas

y_2 means the house price observed in location 2
 $x_{2,1}$ means the deprivation (variable 1) observed in location 2



Notes:

Let Y be some dependent variable that is continuous and normally distributed, where there are 16 observation for Y at some i -location (i.e. $y_1, y_2 \dots y_{16}$)

- For example: Averaged house price (£)

Let X be some k^{th} independent variable $x_{i,k}$, (in this case $k = 1$) where there are 16 locations for X (i.e. $x_{1,1}, x_{2,1} \dots x_{16,1}$)

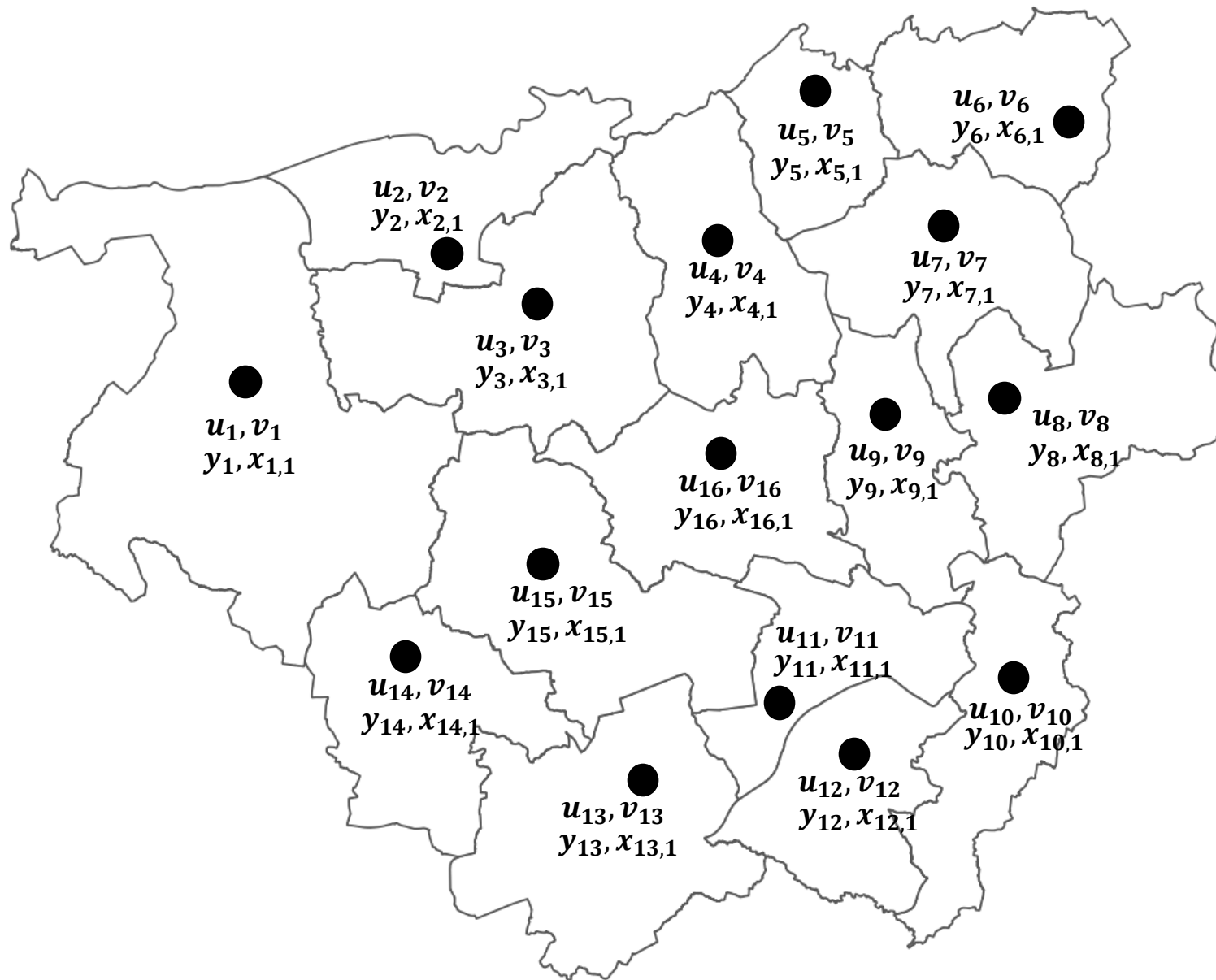
- For example: Socioeconomic deprivation score

Research question: To investigate the **geospatial** impacts of socioeconomic deprivation on house price in each area in this hypothetical study area.

Insufficient to use the typical linear regression model for this context

$$y = \beta_0 + \beta_1 x_1 + \varepsilon$$

Suppose we have a hypothetical study area with 16 areas



Notes:

We want to model the relationship of $y_i, x_{i,1}$ at location (u_i, v_i)

Because $y_i, x_{i,1}$ is calibrated on (u_i, v_i) as a function, we are able to use some model (i.e., GWR) to compute coefficients at each location of i .

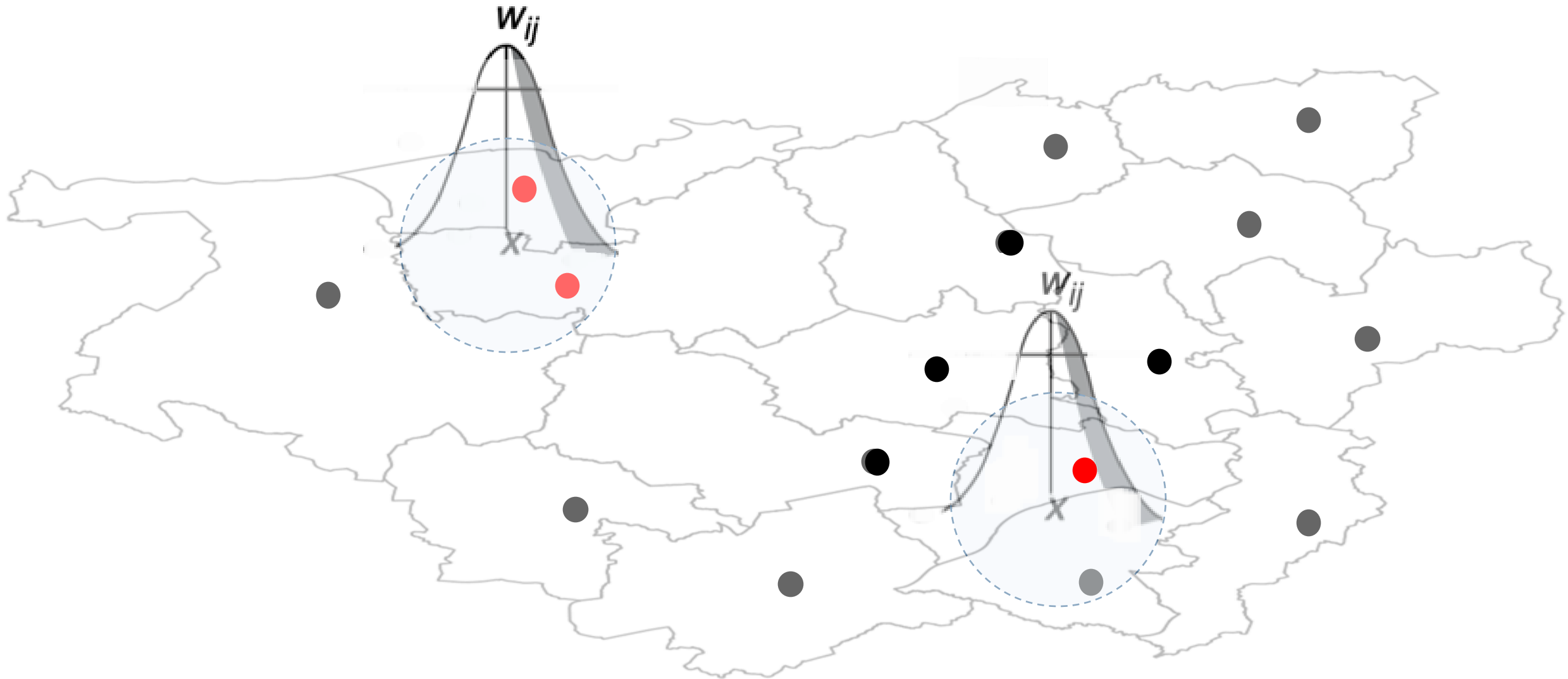
In this example for IMD, these coefficients are represented as $\beta_{i,1}(u_i, v_i)$

The GWRs implicitly use distance-based weights through **spatial kernels** or **bandwidths**. Hence, it relies on points.

Centroids are extracted from areas and used for such analysis.

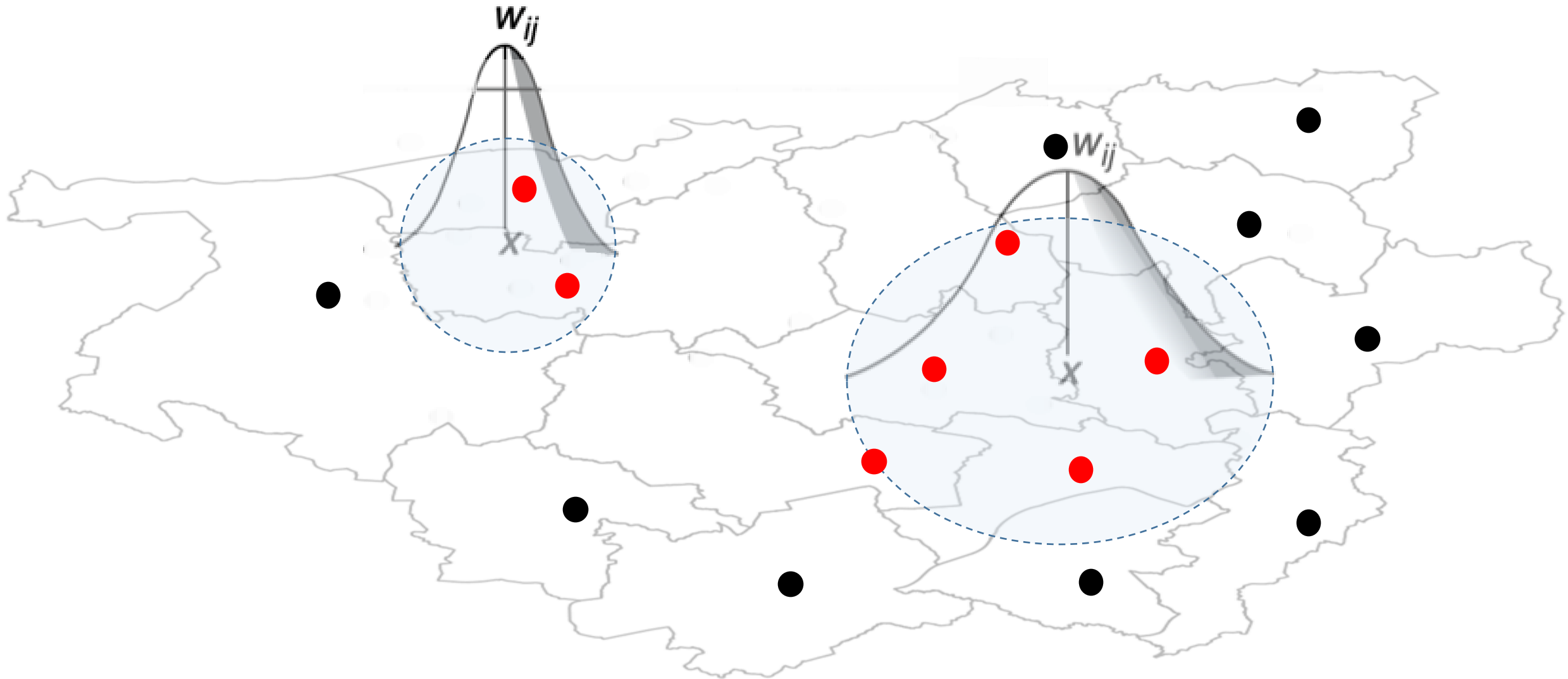
$$y_i = \beta_{i,0}(u_i, v_i) + \beta_{i,1}(u_i, v_i)x_{i,1} + \varepsilon_i$$

GWR with fixed spatial kernel (or bandwidth)



Note: GWRs are distance-based models. It uses bandwidth to consider nearest neighbours when accounting spatial configuration.

GWR with adaptive spatial kernel (or bandwidth)



Use the Adaptive spatial kernel for building your spatial weights! It is much better than using the fixed bandwidth

Workflow for GWR modelling

Modelling process using GWR

When you want to conduct evidence-based analysis with spatial data – especially if the outcome is from a continuous distribution – you might want to follow these steps:

- **STEP 1:** Carry some descriptive analysis to understand the underlying spatial distribution
- **STEP 2:** Perform a Linear regression in order to assess the residuals for the model output to determine whether not the assumptions of independence have been violated.
- You can check for multicollinearity among independent variable using the Variance Inflation Factor ($VIF < 10$)

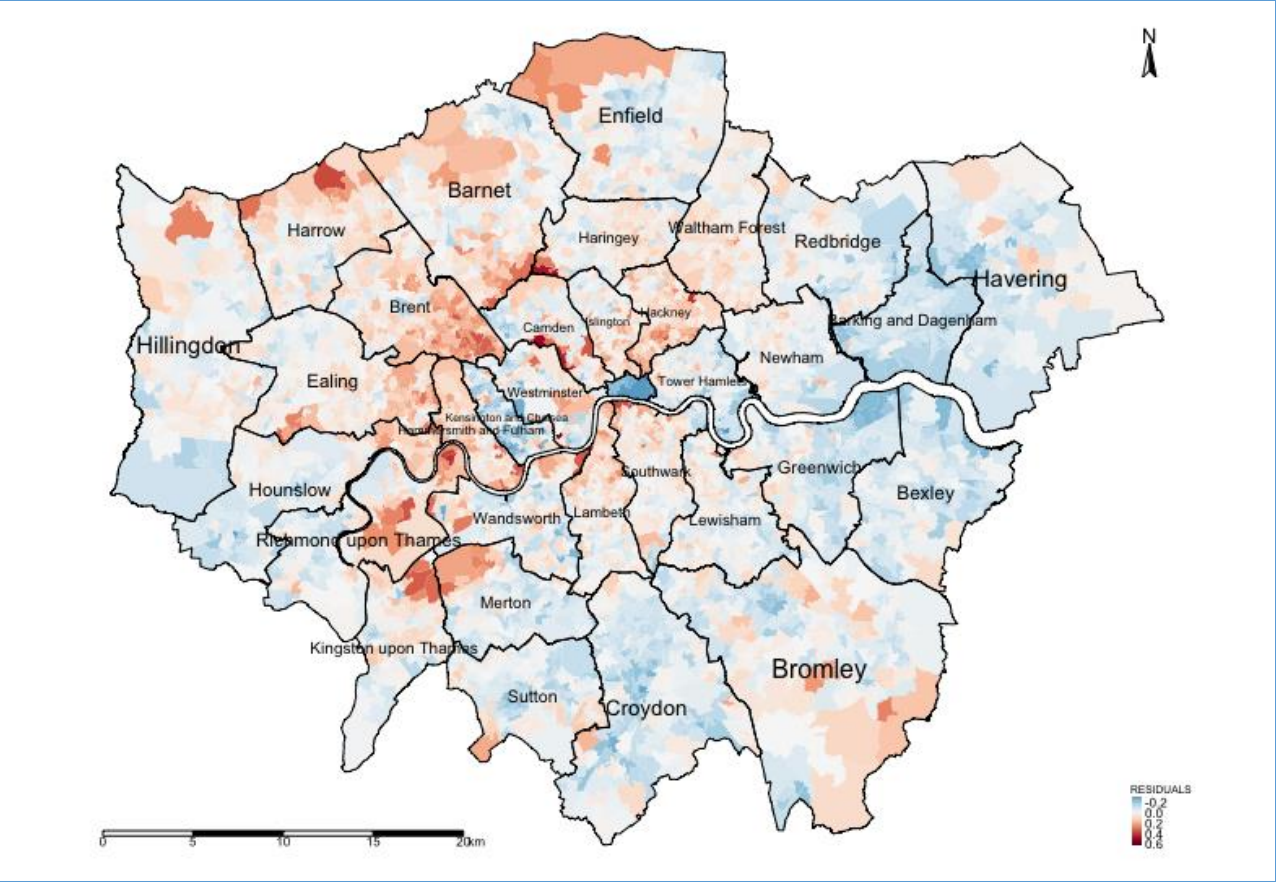
If there's a violation:

- **STEP 3:** Examine whether there is evidence of spatial dependence in the residuals using Moran's I test.
- If Moran's I test is not significant – **STOP! NO EVIDENCE OF SPATIAL DEPENDENCE HENCE NO NEED OF SPATIAL APPROACH**

If Moran's I test is significant and positive – then map the residuals accordingly to examine the spatial patterning of the residual.

- **STEP 4:** Extract the centroids of the areas and use them for computing the kernel bandwidths. [1] Highly recommend to use the adaptive bandwidths, which is flexible than the fixed for estimating the optimal bandwidth.
- **STEP 5:** The estimated bandwidth is fitted into the GWR model to estimate the following quantities: 1.) **Local R-squared**, 2.) **area-specific coefficients** and 3.) **standard errors for significance test for each areas**.
- **STEP 6:** Extract the coefficients and desired results and map them accordingly to examine the spatial variation in the relationship between dependent and independent variables.
- **STEP 7:** Interpretation

Example using the house price data for London



Examining the residuals for determining spatial patterning and evidence of spatial autocorrelation. Moran’s I test was 0.475 ($p < 0.001$) meaning that the residuals are clustered.

Broadly, there’s an over-estimation in the house-price and spatial aspects needs to be accounted for.

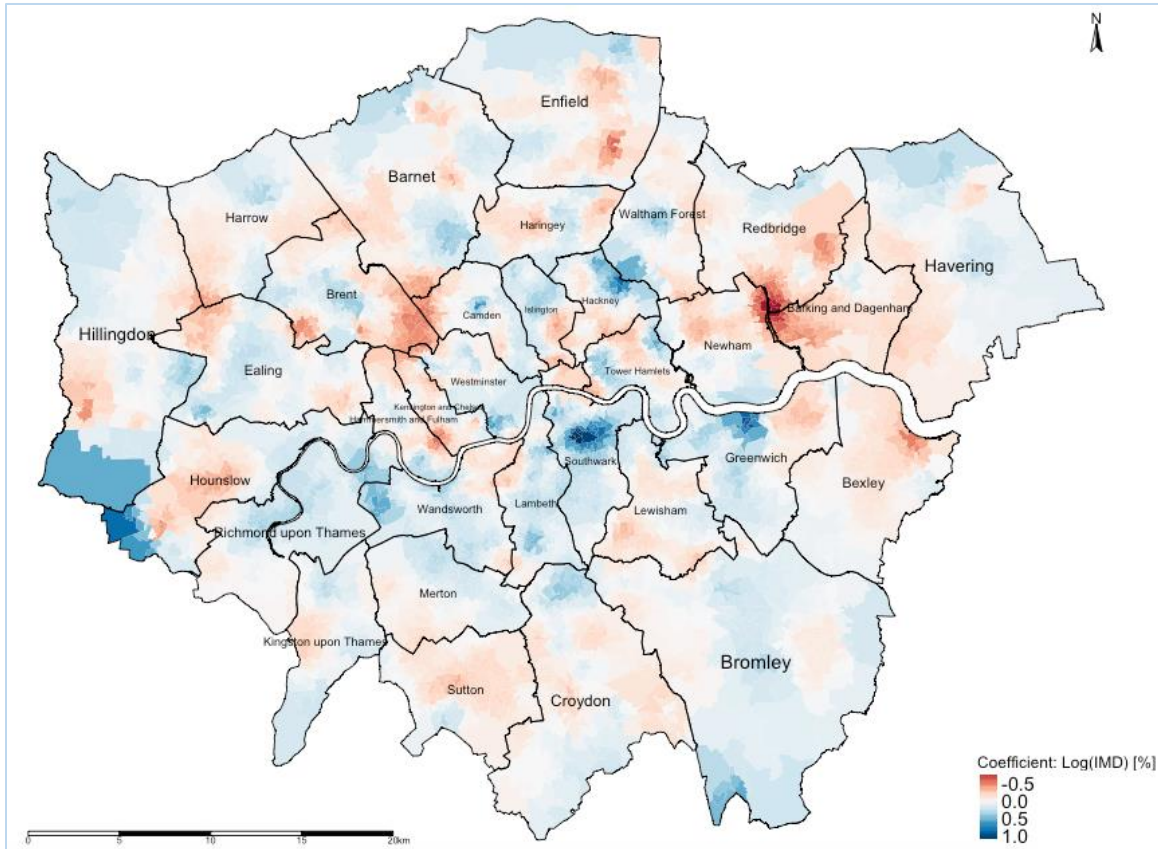
Reporting the Global estimates

Modelled results in table are from Linear, Spatial Lag (Y) and GWR regression model

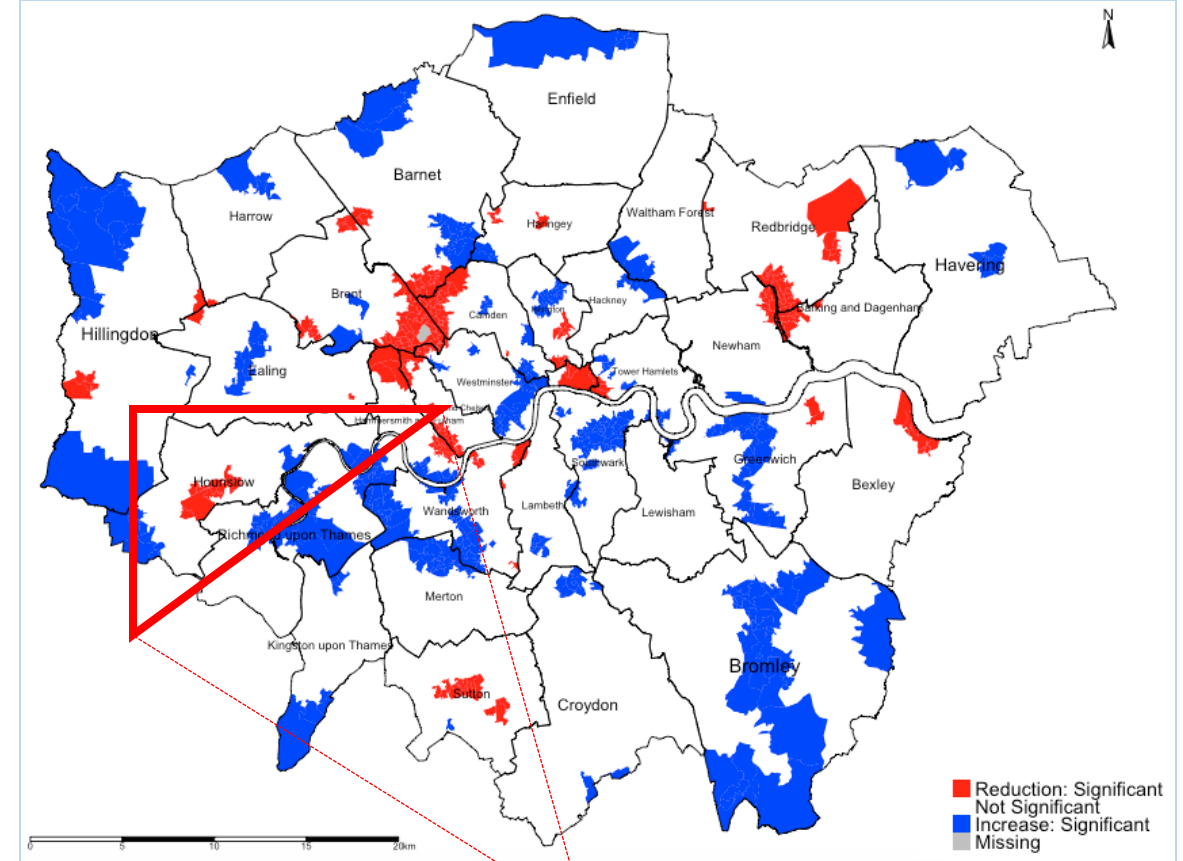
Variable(s)	Linear	Lag (Y)	GWR
log(Income)	2.036*	1.267*	2.036
log(Deprivation)	0.136*	0.045*	0.137
log(PTAL)	0.031*	0.011*	0.030
AIC	-8510.8	-9863.3	-11242
R ²	0.7889 (78.89%)	N/A	0.9318 (93.18%)

- The GWR model is better than the linear and Spatial Lag regression. We take the model with the highest R-squared value, as well as the lowest AIC value.

Reporting the local estimates (using socioeconomic deprivation (adjusted for other risk factors) as an motivating example)

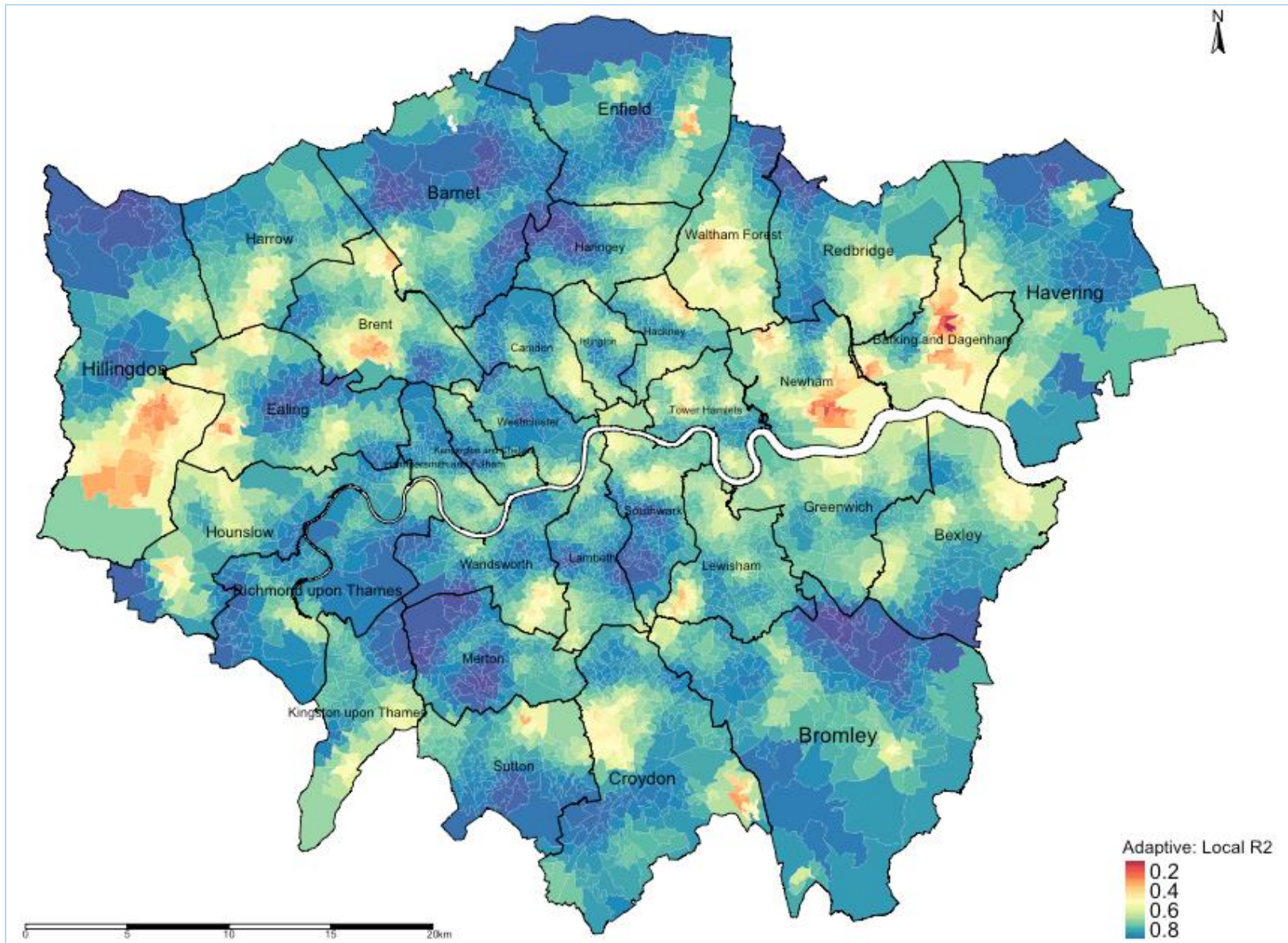


Interpretation: There is spatial variability in the relationship between our variable socioeconomic deprivation (transformed) and averaged house price (transformed) in London. The GWR outputs reveals that local coefficients range from a minimum value of -0.946 to a maximum value of 1.085, indicating that one percentage point increase in the levels of deprivation in LSOs of London is associated with a reduction of 0.946% in house prices in some LSOs and (weirdly) an increase of 1.085% in others. Broadly, the relationship are opposing.



Interpretation: For instance, in the **Borough of Hounslow**, we can see a significant reduction in house prices in relation to increased levels of socioeconomic deprivation (adjusted for income and accessibility). Such reduction are clustered in the mid-section of Borough of Hounslow which were coloured red. Note that in far north eastern section of the Borough of Hounslow with pockets of LSO's coloured blue shows a significant increase in house price in relationship to IMD which is difficult to explain and thus can be interpreted as a chance finding. All sections that are coloured white are not significant.

Reporting the local R-squared to assess the model's performance for each areas



Interpretation: The areas that are going towards the shade of dark reds (i.e., value of 0) are local regression models that have broadly performed poorly in its prediction for house price and its association with the three variables (income, deprivation and PTAL). Likewise, the areas that are going towards the shade of dark blues (i.e., value of 1) are local regression models that have broadly performed very well in its prediction for house price and its association with the three variables (income, deprivation and PTAL).

Note: These results are essential as the local R2 values of each area show the model's ability to predict the explained variance in house prices caused by deprivation, income and accessibility for specific areas.

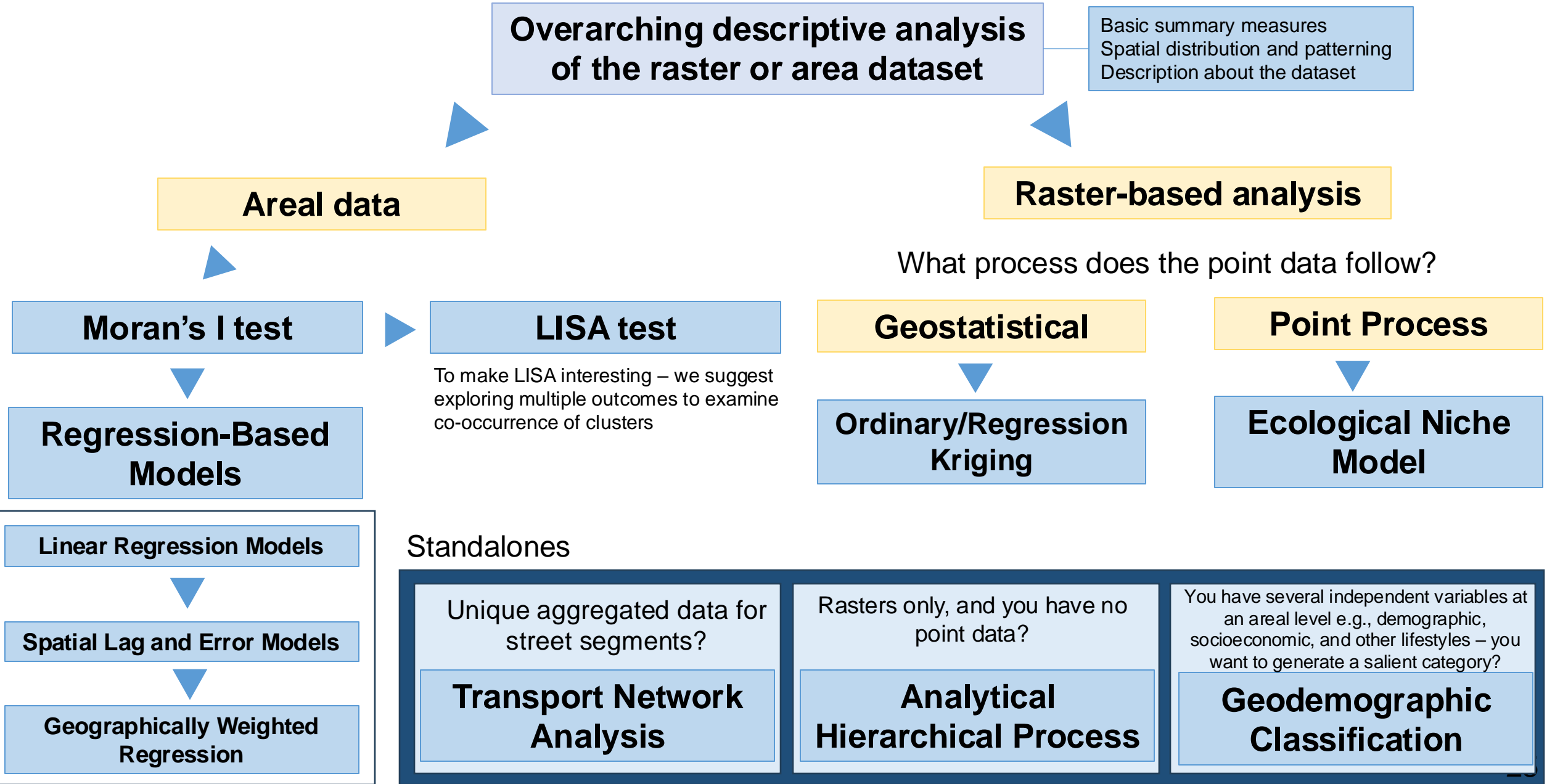
**Any questions about
GWR?**



Summary of GEOG0114 & Assessment

What have we covered in the last 10 weeks...

Week	Weekly Topics
1	Spatial Analysis for Data Science
2	Graphical Representation of Spatial Data
3	Spatial Autocorrelation
4	Suitability Mapping: Part I (Qualitative approach)
5	Suitability Mapping: Part II (Quantitative approach)
Reading Week (Assessment)	
6	Geostatistics using Kriging
7	Geodemographics
8	Transport Network Analysis
9	Spatial Models: Part I (Spatial Lag & Error Models)
10	Spatial Models: Part II (Geographically Weighted Regression)



Example Data source(s)

UK Datasets

UK Census 2021: https://www.nomisweb.co.uk/sources/census_2021_bulk

UK Deprivation (2010, 2015) and 2019: <https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019>

UK population 2010-2022 mid-year estimates: [\[LINK\]](#)

Police: <https://data.police.uk/>

CDRC: <https://data.cdrc.ac.uk>

London Data Store: <https://data.london.gov.uk>

Mosquito/species data

Global Biodiversity Information Facility: <https://www.gbif.org>

Global Mosquito Observatory Database (GMOD): <https://experience.arcgis.com/experience/7228a5a27442468494caec2934c2b73d/page/Page/>

Malaria Atlas Project: <https://malariaatlas.org>

Disaster-related fire datasets

Fire Information for Resource Management System: <https://firms.modaps.eosdis.nasa.gov>

[1] Several domestic fire outcomes (LSOA) in UK: <https://assets.publishing.service.gov.uk/media/64be67369c2df000129402cf/low-level-geography-dataset-270723.ods>

[2] Guidance note for [1]: <https://www.gov.uk/government/statistics/fire-statistics-incident-level-datasets/low-level-geography-dataset-guidance>

Disaster dataset

EM-DAT International Disaster Database: [\[LINK\]](#)

Humanitarian dataset

Humanitarian Data Exchange: <https://data.humdata.org>

Worldpop HUB: <https://www.worldpop.org>

Global Terrorism Database: <https://www.start.umd.edu/gtd/>

Armed Conflict Location and Event Data: <https://acleddata.com>

WHO-ESPEN Project containing tropical diseases in Sub-Saharan Africa:

<https://espen.afro.who.int/> | <https://espen.afro.who.int/tools-resources/cartography-database> | <https://espen.afro.who.int/tools-resources/download-data>

Inspiration for writing methodology




Musah et al. (2020): <https://doi.org/10.1016/j.apgeog.2019.102126>

Todd et al. (2022): <https://doi.org/10.1177/23998083211001836>




Li et al. (2022): <https://doi.org/10.1016/j.apgeog.2022.102718>

Go to Moodle's Assessment

Assignment Documentation: Guidance notes and Cover Sheet

 [GEOG0114 Spatial Data Analysis Project - Instructions & Guidance Notes](#)   199.9 KB PDF document

Please download the worksheet for the Spatial Data Analysis Project. For further guidance, you find this information contained in this documentation.




 [GEOG0114 Spatial Data Analysis Project - Cover Sheet template](#)   42.1 KB Word 2007 document




Please download the cover sheet template for your Spatial Data Analysis Project. This document must be submitted as the front page of your 3,000 word report. It must contain your **candidate ID**, **title of report**, **the word count**, and **link(s) to R-scripts and data sets** (e.g., OneDrive, Google Drive or GitHub repository).




Below are the GREAT examples of past projects:




PLEASE GO THROUGH THE DOCUMENTATION!




See Best examples of past projects from 2021/22 and 2022/23




 [Thermal inequities in Chicago neighborhoods \(2021/22\)](#)   558.4 KB PDF document

 [How social determinants of health effect health conditions in London \(2021/22\)](#)   4.3 MB PDF document

 [A spatial analysis of gun violence in the city of New York \(2021/22\)](#)   453.1 KB PDF document

 [Understanding relationships between cultural accessibility and educational deprivation in Inner London \(2022/23\)](#)   2.2 MB PDF document

 [Disparities in COVID-19 death rates among ethnic groups in England and Wales \(2022/23\)](#)   2.2 MB PDF document

 [Influence of the spatial variations of soil properties on soil-transmitted helminths transmission in Nigeria \(2022/23\)](#)   10.6 MB PDF document

Marking Criteria for Coursework

3 in a category or above	Poor	Limited	Fair	Good	Excellent	Outstanding	Exceptional
	0-39	40-49	50-59	60-69	70-79	80-89	90+
1. Introduction, argument and conclusions	Does not respond to the task set, with either no argument or argument presented is inappropriate & irrelevant. Conclusions absent or irrelevant.	An indirect response to the task set, with a rudimentary argument & conclusions. Not sufficiently focused on the subject area.	A satisfactory response with a general understanding of the topic but a limited sense of argument & partial conclusions.	A good response with a clear argument and logical conclusions.	A distinctive response with an insightful argument and sophisticated conclusions. Shows a thorough grasp of issues.	An impressive, nuanced response with a perceptive argument and superior conclusions. There is a striking grasp of complexities and significance of issues.	Exceptional response with a precise and subtle argument and conclusions. Grasp of complexities and significance of issues surpasses the standards associated with 80-89%.
2. Engagement with literature	Non-existent or irrelevant reading.	Significant omissions in reading with weak understanding of literature consulted.	Evidence of some understanding of literature consulted but not a wider reading.	Demonstrates plentiful relevant reading and sound understanding of literature consulted	Extensive reading and exceptional understanding of literature consulted.	Ambitious reading and use of relevant literature that extends beyond the taught elements of the module.	Level of literature review and synthesis at a quality suitable for journal publication.
3. Analysis: reflection, thought, & conceptual framework	Shows poor understanding of the issues under discussion. Uncritical use of material.	Shows some understanding of the material used but conceptually limited or with wrong application.	Some competent reproduction of ideas, grasp of issues and their broader implications but lacks critical approach and independent evaluation.	Evidence of student's own analysis that is well-organised and informed, covering relevant issues with insights into broader implications. Concepts are defined and used systematically and effectively.	Innovative analysis with critical thinking using concepts that are deftly defined and used with an excellent sense of theoretical context	Analysis demonstrates independent thinking that is critically perceptive. Concepts are deftly defined and accurately employed with a superior sense of theoretical context.	Exceptional thought and awareness of cutting edge discussion. Conceptual framework is outstanding in terms of critical insight and theoretical approach.
4. Empirical knowledge & use of examples (as relevant)	No use of examples or evidence to support claims made or empirical material is irrelevant/inaccurate.	Empirical material is limited in quantity and quality with little substantiation of claims.	Evidence of empirical material but not always effectively used to support claims.	Substantive amount of in-depth empirical material to support most claims.	Original, highly relevant and detailed use of empirical material.	Impressive, detailed and distinctive empirics with clear reflection on data used.	Comprehensive use of more than precise empirical material with critical reflection on data used.
5. Structure, writing, illustrative material and references (as relevant)	Very poorly expressed and presented that seriously interfere with comprehension and almost no illustrative material and/or referencing.	Structure and writing detract from conveying of ideas with minimal illustrative material and/or referencing.	Structure and writing sometimes detract from the conveying of argument and ideas with adequate illustrative material and/or referencing.	Good structure and writing that do not detract from conveying of argument and ideas with a comprehensive use of illustrative material and/or referencing.	Excellent structure and writing with an extended set of impeccable illustrative material and/or references.	Structure, writing, referencing and illustrative material used demonstrates verve and greatly enhances ideas.	Structure, writing and referencing and illustrative material is impeccable and as if professionally copy edited.

The 'Dos' and 'Dont's'

Make sure that your figures and tables are publication-worthy [1]

Education Level & Suicide Rate

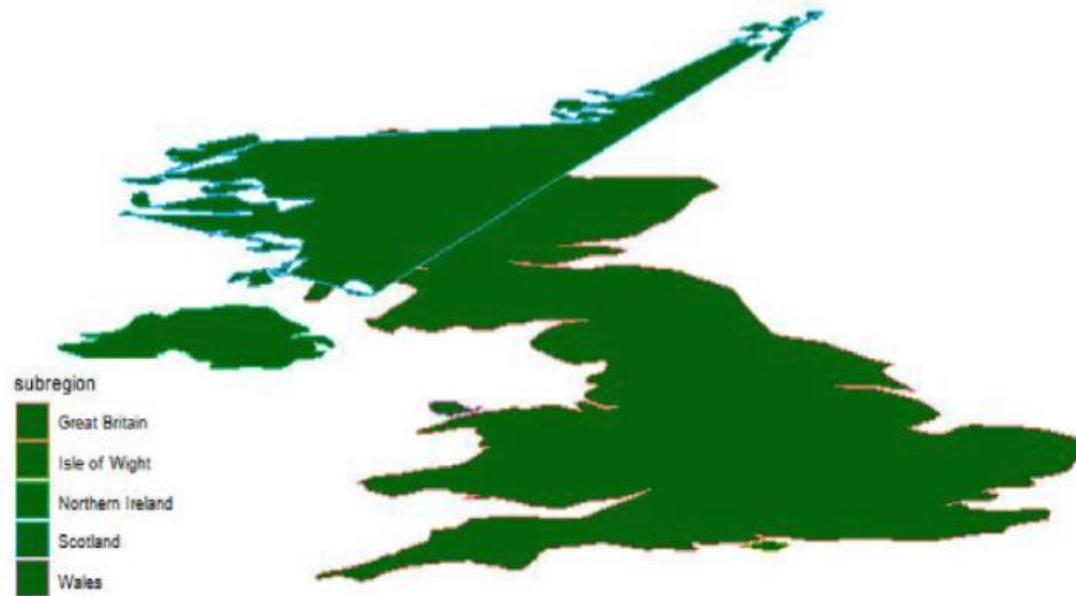


Figure 1 The Distribution map of London

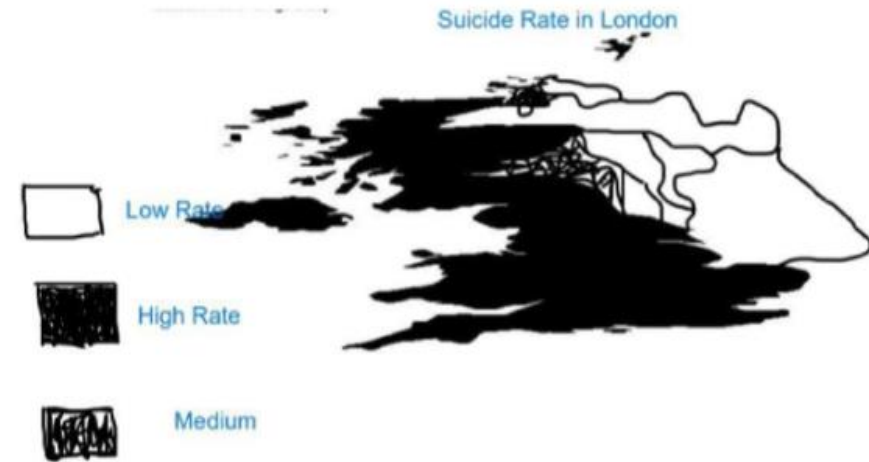


Figure 2 Age Group and Suicide

If you produce an image like this – it's an automatic failure

The ‘Dos’ and ‘Dont’s’

Make sure that your figures and tables are publication-worthy [2]

Table 2: Shows a descriptive analysis of the lung capacity function (i.e., volume) stratified by gender. Data is from among 654 respondents from the East Midlands.

Variables	n	Mean (±SD)	IQR [Median (Q1-Q4)]	Ranges (Min-Max)
Gender				
Female	318	5.35 (±1.937)	5.46 (3.85 to 6.98)	0.37 to 9.51
Male	336	6.44 (±3.011)	5.82 (4.03 to 8.6)	0.39 to 15.38
Age Groups				
<55 years	130	3.07 (±1.043)	3.06 (2.42 to 3.63)	0.37 to 5.73
55-59 years	407	5.99 (±1.979)	5.76 (4.57 to 7.15)	1.88 to 13.67
60+ years	117	8.8 (±2.387)	8.56 (7.17 to 10.68)	4.59 to 15.38
Altitude Type				
High land	65	7.83 (±2.25)	7.51 (6.38 to 9.25)	3.08 to 12.62
Low land	589	5.7 (±2.552)	5.39 (3.76 to 7.14)	0.37 to 15.38

Total sample size (N) = 654

Variables	n	Mean (SD)	IQR [Median (Q1-Q4)]	Ranges (Min-Max)
Female	318	5.35 (1.937)	5.46 (3.85 to 6.98)	0.37 to 9.51
Male	336	6.44 (3.011)	5.82 (4.03 to 8.6)	0.39 to 15.38
<55 years	130	3.07 (1.043)	3.06 (2.42 to 3.63)	0.37 to 5.73
55-59 years	407	5.99 (1.979)	5.76 (4.57 to 7.15)	1.88 to 13.67
60+ years	117	8.8 (2.387)	8.56 (7.17 to 10.68)	4.59 to 15.38
High land	65	7.83 (±2.25)	7.51 (6.38 to 9.25)	3.08 to 12.62
Low land	589	5.7 (±2.552)	5.39 (3.76 to 7.14)	0.37 to 15.38

Best Standards:

Fully formatted table, with table legends which looks great, and it is of the standards that is considered publication-worthy. This type will yield you full marks i.e., correct results and show an eye details.

Normal Standards:

Partially formatted table, with no table legends which looks rudimentary. No way near considered publication-worthy (i.e., in a report, research article, thesis etc.). The marker will be like “meh”... but if s/he (i.e., marker) is pissed-off or woke up on the wrong side of the bed, they may deduct marks.

The 'Dos' and 'Dont's'

Make sure that your figures and tables are publication-worthy [3]

```
# A tibble: 2 × 9
  gender      n mean    sd median    q1    q3   min   max
<chr> <int> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 female   318  5.35  1.94  5.46  3.85  6.98 0.373  9.51
2 male    336  6.44  3.01  5.82  4.03  8.60 0.388 15.4

# A tibble: 3 × 9
  agegroup      n mean    sd median    q1    q3   min   max
<chr> <int> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 <55 years    130  3.07  1.04  3.06  2.42  3.63 0.373  5.73
2 55-59 years  407  5.99  1.98  5.76  4.57  7.15 1.88 13.7
3 60+ years   117  8.80  2.39  8.56  7.17 10.7  4.59 15.4

# A tibble: 2 × 9
  altitude      n mean    sd median    q1    q3   min   max
<chr> <int> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 high land    65  7.83  2.25  7.51  6.38  9.25 3.08 12.6
2 low land   589  5.70  2.55  5.40  3.76  7.14 0.373 15.4
```

Crap standards:

Not even worth the marker's time.

If you want to alienate the markers, we dare you to submit an assignment with tables not formatted...

We double dare you to submit an output that is copied and pasted from R...

The 'Dos' and 'Dont's'

Make sure that your figures and tables are publication-worthy [4]

```
# A tibble: 7 × 8
```

	n	mean	sd	median	q1	q3	min	max
	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	130	3.07	1.04	3.06	2.42	3.63	0.373	5.73
2	407	5.99	1.98	5.76	4.57	7.15	1.88	13.7
3	117	8.80	2.39	8.56	7.17	10.7	4.59	15.4
4	318	5.35	1.94	5.46	3.85	6.98	0.373	9.51
5	336	6.44	3.01	5.82	4.03	8.60	0.388	15.4
6	65	7.83	2.25	7.51	6.38	9.25	3.08	12.6
7	589	5.70	2.55	5.40	3.76	7.14	0.373	15.4

Really crap standards:

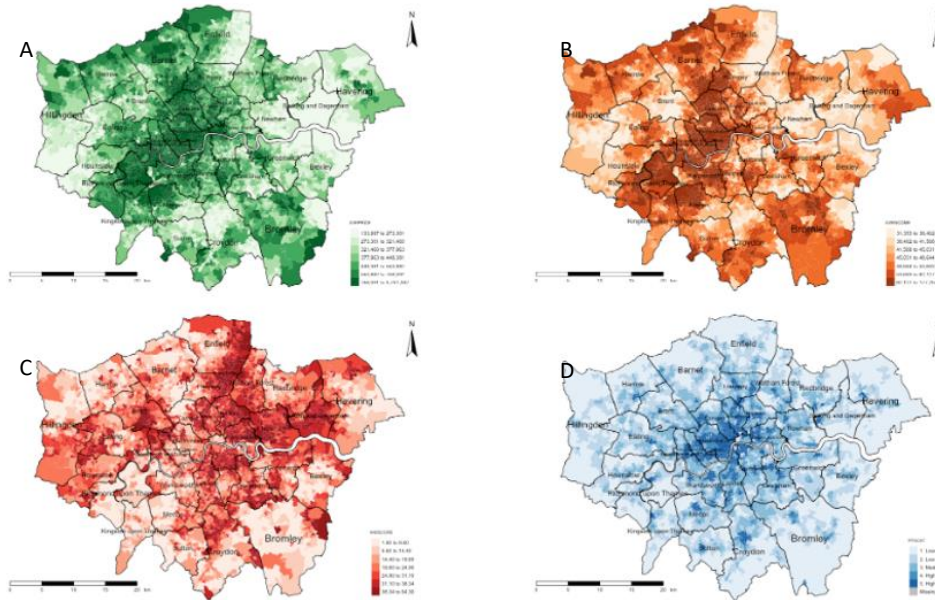
Even worse... we triple dare you to take a screenshot and paste it into your assignment...



“We will shut that sh*t down! No exceptions”
[Quote: Negan (The Walking Dead, Season 6, Episode 16 [Last Day on Earth])

The 'Dos' and 'Dont's'

Readability of outputs and formatting of pages [1]



If you have multiple figure outputs, you are welcome to generate a panelled figure that is annotated alphabetically.

The issue is that squeezing a panelled figure on portrait document reduces its readability severely.

You should pay close attention to even these little things to avoid alienating your marker.

For the marker to make sense of this image, s/he will be forced to zoom in the +400% any to still see blurry figure...

The ‘Dos’ and ‘Dont’s’

Readability of outputs and formatting of pages [2]

52

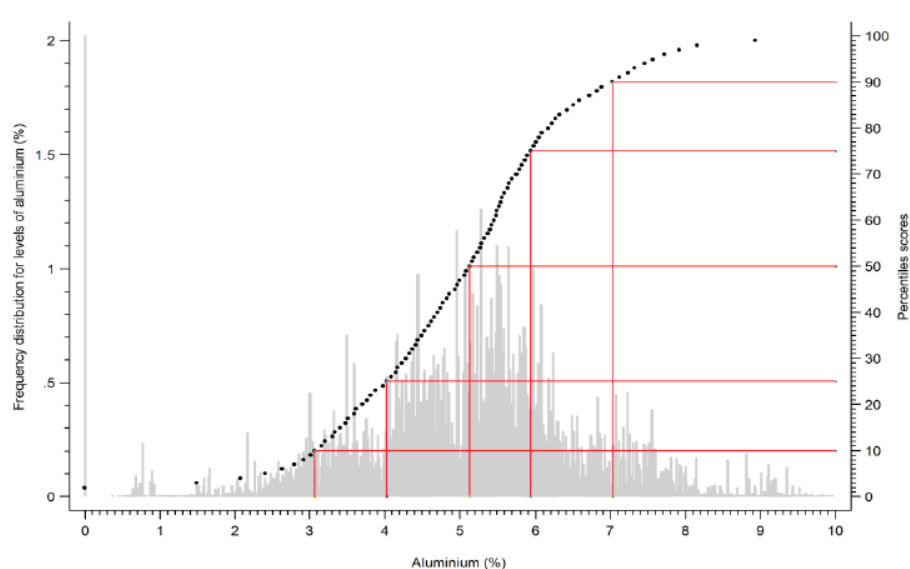


Figure 3.3: Two-way histogram with cumulative proportions showing the overall distribution of patients in THIN-GBASE with specific soil concentration levels for aluminium. Left y-axis: corresponds to the observed proportion of patients with specific soil levels of aluminium. Right y-axis: Black dots correspond to a percentile score - i.e. the proportion of patients that fall under specific soil concentration value for aluminium; Red line indicates: 10th, 25th, 50th, 75th and 90th percentiles (i.e. 31,000, 40,000, 51,000, 59,300 and 70,500 mg/kg respectively). The concentrations for aluminium were converted to a weight percentage (mg/kg÷10,000), whereby 1.0% = 10,000 (of aluminium) parts-per million.

Show the image in its glory to the fullest by taking advantage of the landscape settings

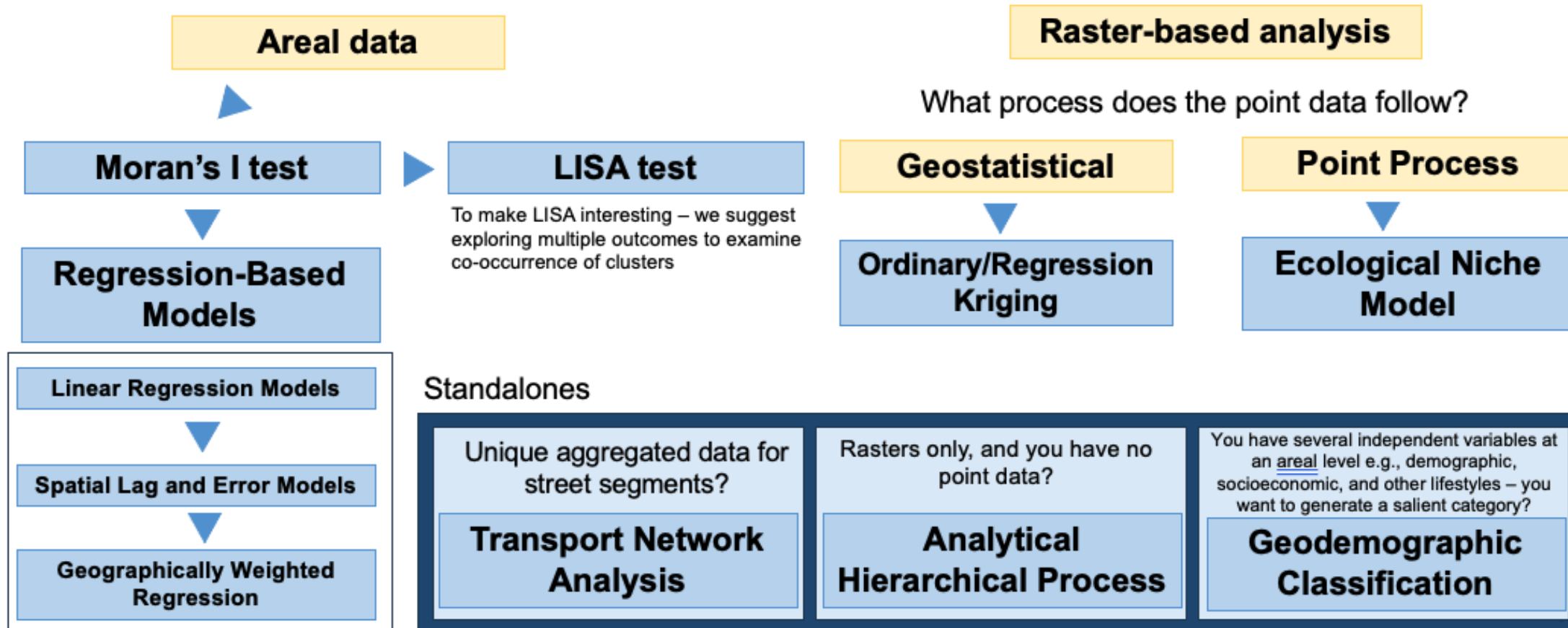
This is a crude example to illustrate what I mean about showing your results nicely to its fullest glory.

Notice the page number has been formatted accordingly to the width of the page, instead of it being on the length to ensure consistency in the page number position if the document was printed.

Markers appreciate this level of detail and are not above being sympathetic in awarding extra marks for quality in presentation.

The 'Dos' and 'Dont's'

Avoid violating the core assumptions behind these analytical techniques



- Any violation of the core assumptions – whether it be data-related, or distributional will be considered as a critical error.
- Marks will be deducted accordingly. It should be noted that your marks are capped at 68/69.
- If it's horrible violation, to such extent its irredeemable, then in such scenario can potentially lead to a failing grade.

The ‘Dos’ and ‘Dont’s’

Use of Generative AI (e.g., ChatGPT) and References

If you're using GenAI to **assist** or **troubleshoot** with the R coding part of the assessment, that's perfectly fine. You just need to **acknowledge** the use of GenAI in your methodology section, especially if it was used for straightforward tasks like data cleaning or manipulation.

“I acknowledge the use of ChatGPT (<https://chatgpt.com/>) to improve my writing, and assistance with R coding”


However, if GenAI was used for more significant contributions, such as getting information to describe the statistical methodology or modelling approach, you must provide a proper citation. UCL offers guidelines on how to cite GenAI in such cases. For example, if I asked ChatGPT the following:

“How to generate Stan Code for a Bayesian ICAR spatial risk model?”

How to generate Stan Code for a Bayesian ICAR spatial risk model? (2024). ChatGPT (GPT-4o). Wednesday, 27 November. Available at: <https://chatgpt.com/c/67471449-f31c-800c-a6e9-f22b554f7841>

- UCL Policy on referencing GenAI [\[LINK\]](#)
- UCL Geography’s Policy on using GenAI in MSc assessments/Dissertations [\[LINK\]](#)

How to generate Stan Code for a Bayesian ICAR spatial risk model? < 2/2 >

 Generating Stan code for a Bayesian Intrinsic Conditional AutoRegressive (ICAR) spatial risk model involves defining the structure for spatial dependency, specifying the priors, and coding the likelihood. Below is a step-by-step outline and example code:

1. Understand the ICAR Model

- ICAR models define spatial dependencies by assuming that the value of a spatial process at a location depends on its neighbors.
- Spatial structure is encoded in a neighborhood (adjacency) matrix.

2. Model Structure

- **Data:** Counts of events (e.g., disease cases), population sizes, and an adjacency matrix defining neighbors.
- **Likelihood:** Often Poisson (for count data).
- **Spatial Prior:** ICAR prior for spatial random effects.
- **Hyperprior:** Prior for the precision parameter of the ICAR model.

3. Components of the Stan Model

- Input adjacency matrix.
- ICAR prior using the precision parameter and neighborhood structure.
- Likelihood function.
- Posterior sampling.

4. Stan Code Example

Here's an example of Stan code for a Bayesian ICAR spatial risk model:

```
stan
data {
  int<lower=1> N;           // Number of regions
  int<lower=1> N_edges;     // Number of edges in adjacency matrix
  int<lower=1, upper=N> node1[N_edges]; // Region 1 for each edge
  int<lower=1, upper=N> node2[N_edges]; // Region 2 for each edge
  int<lower=0> y[N];        // Observed counts
  real<lower=0> E[N];       // Expected counts (offset)
}

parameters {
  vector[N] theta;         // Log-relative risks
  real<lower=0> tau;        // Precision parameter for ICAR prior
}

model {
  // ICAR prior
  target += -0.5 * dot_self(theta[node1] - theta[node2]) * tau;
}
```

Message ChatGPT

GEOG0114: Course Evaluation & Student Feedback (Week 7-10)

<https://forms.gle/2bgUnHgY14vYzDsS9>

Dear Students,

As part of the Continuous Module Dialogue, we are conducting this survey to gauge the levels of student satisfaction with the learning experience in module **GEOG0114: Principles of Spatial Analysis**. We would like to receive your feedback, which would be greatly appreciated. This will help us make improvements to the course. The survey should only take up to 5 or 10 minutes, and your responses are completely anonymous.

Thank you,

Anwar and Justin.

See you in term 2

原作 矢立肇

