

GEOG0114: PRINCIPLES OF SPATIAL ANALYSIS

WEEK 10: REVISION

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

Lecturer in Social and Geographic Data Science

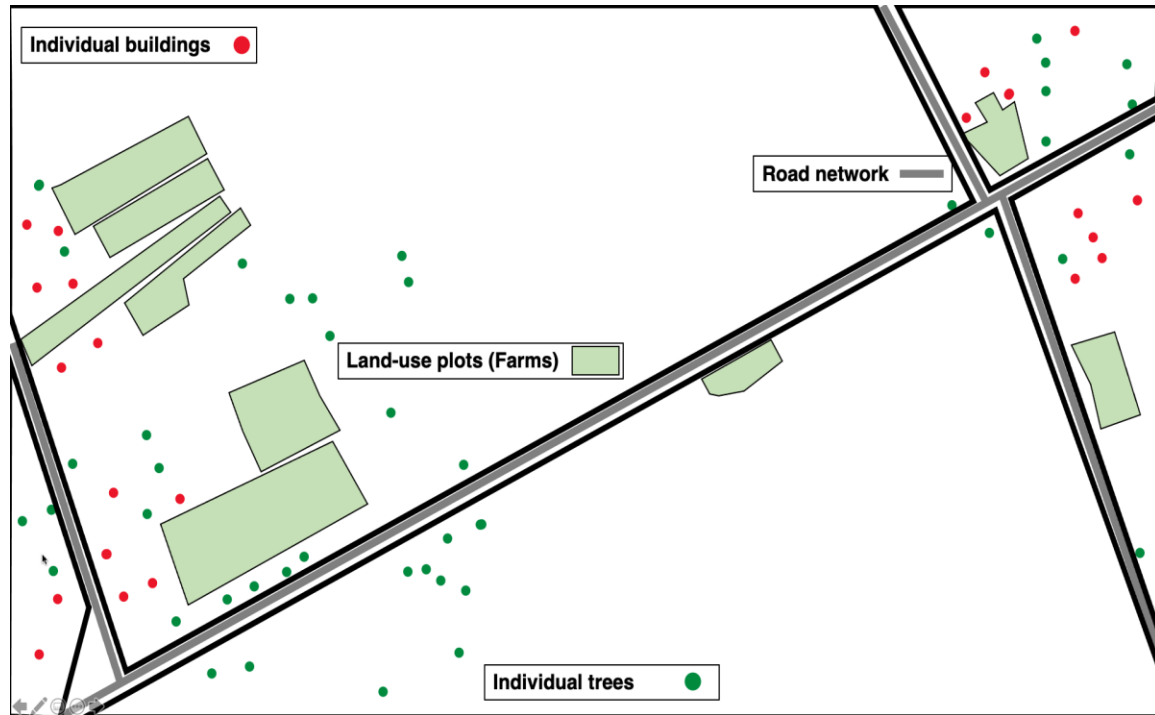
UCL Geography

Summary of GEOG0114

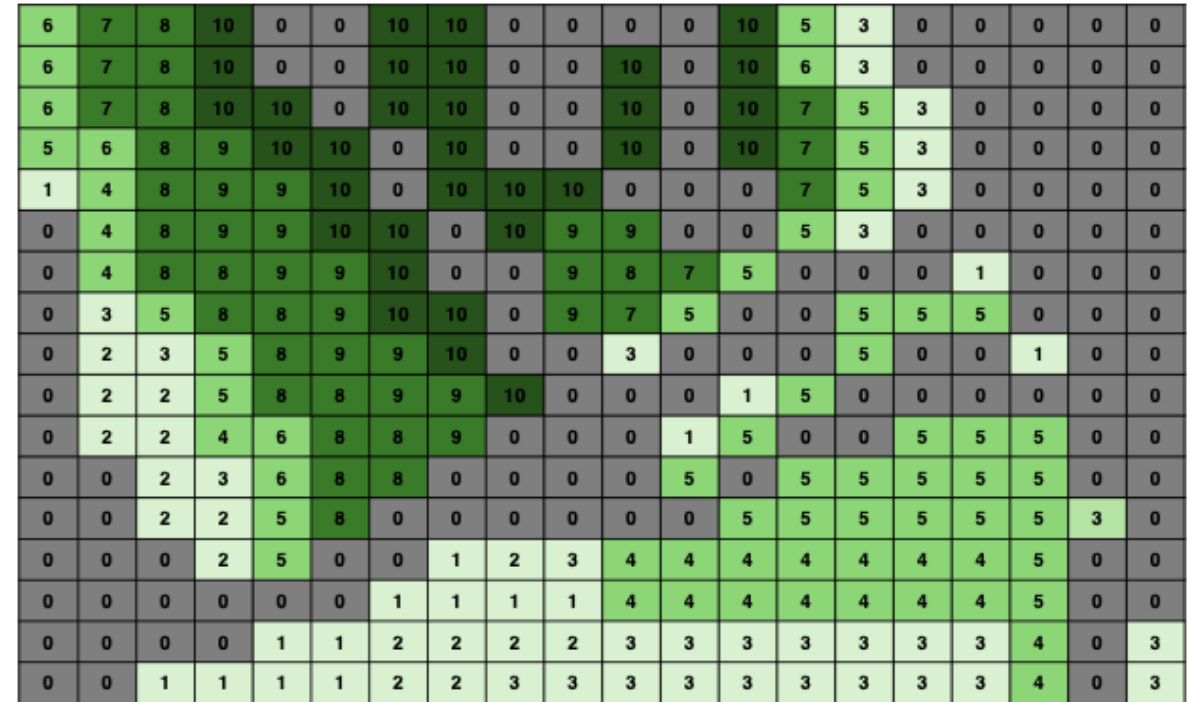
What have we covered in the last 9 weeks...

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	Spatial Models: Part I (Spatial Lag & Error Models)
8	Spatial Models: Part II (Geographically Weighted Regression)
9	Spatial Models: Part III (Spatial Risk Models)
10	Summary of GEOG0114

Week 2: Graphical Representation of Spatial Data



Vector data



Raster data

Geoprocessing Techniques

1. Clipping
2. Intersection
3. Union
4. Dissolving
5. Aggregation
6. Buffer

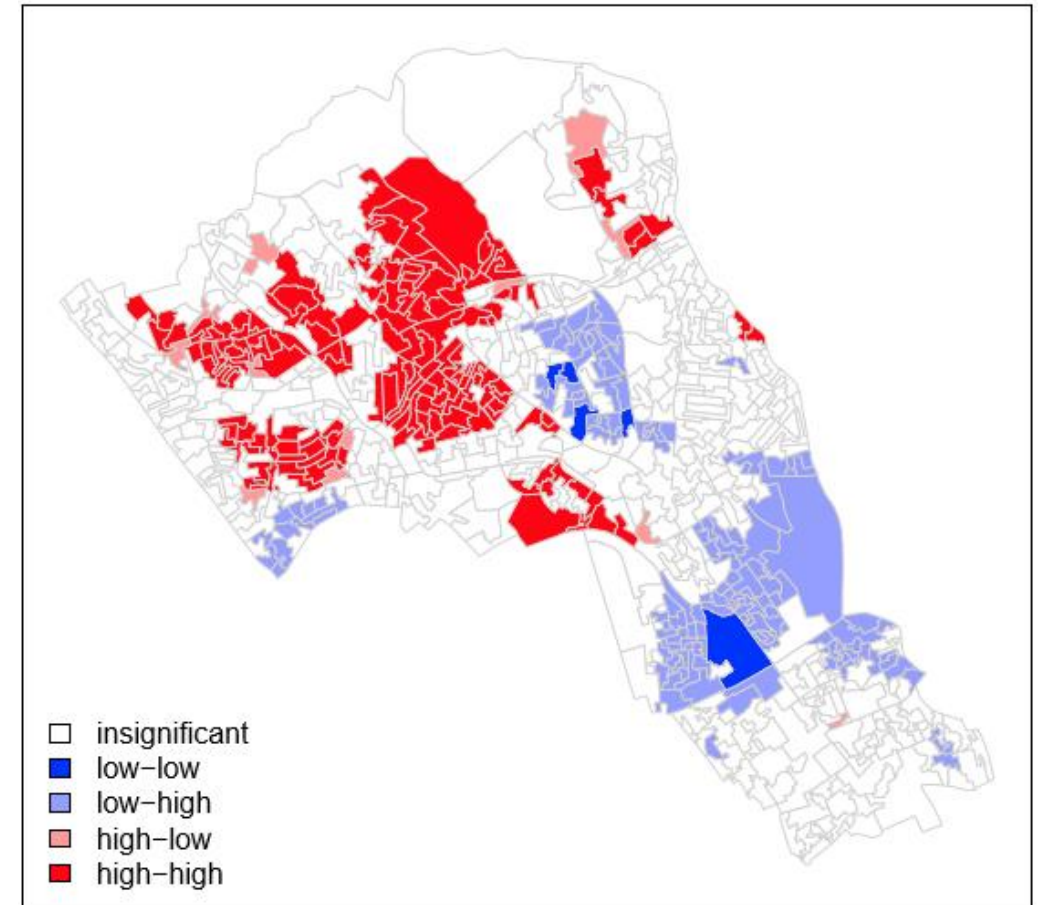
Global Test

- **Generate the hypothesis**
 - Null hypothesis: The patterns are random
 - Alternative hypothesis: The patterns are not random (i.e., clustered/dispersed).

How to perform Global Moran's I test to get a p-value to accept, or reject the null hypothesis

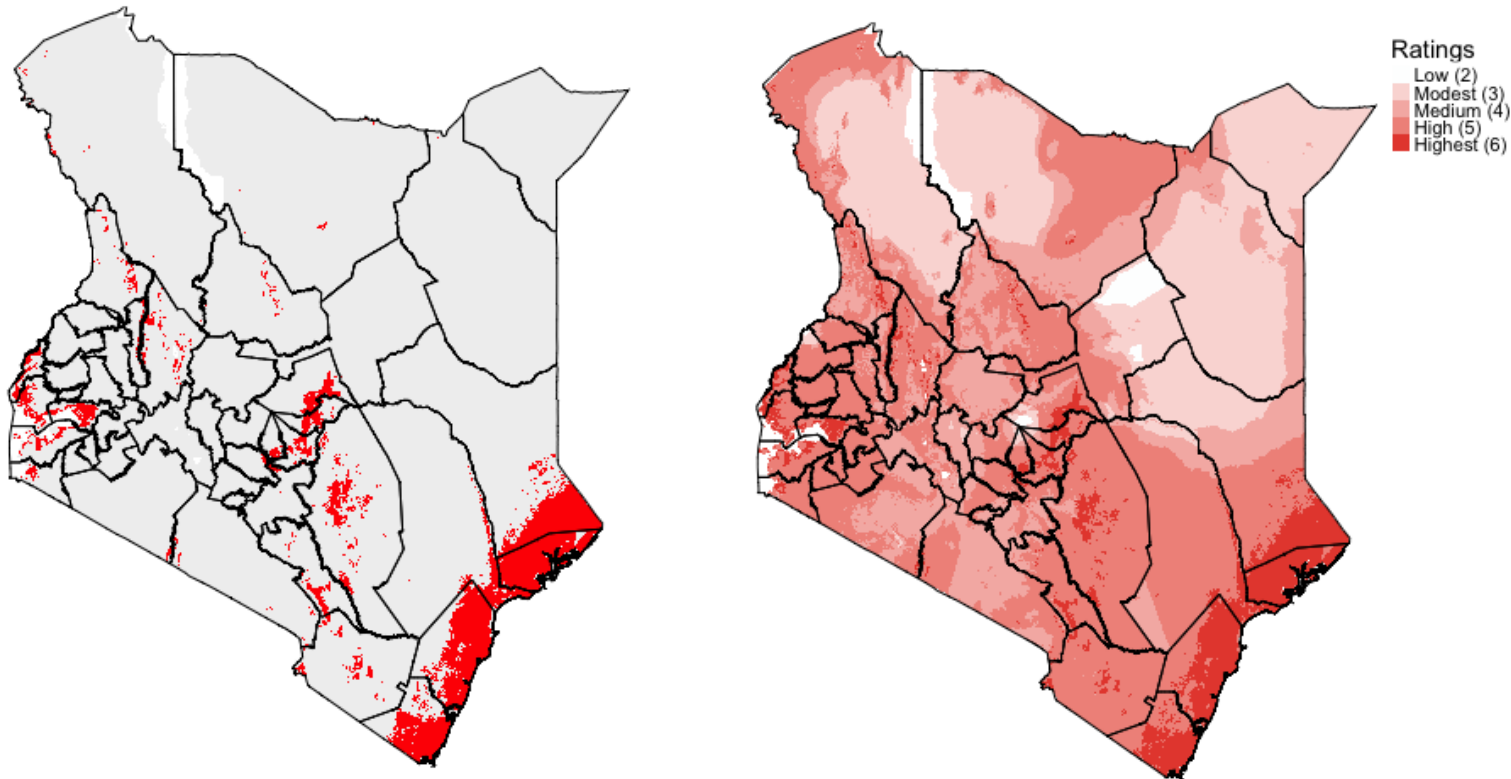
- **Result (Global Moran's I = 0.5448 [$p=0.0001 < 0.05$])**

Local Test

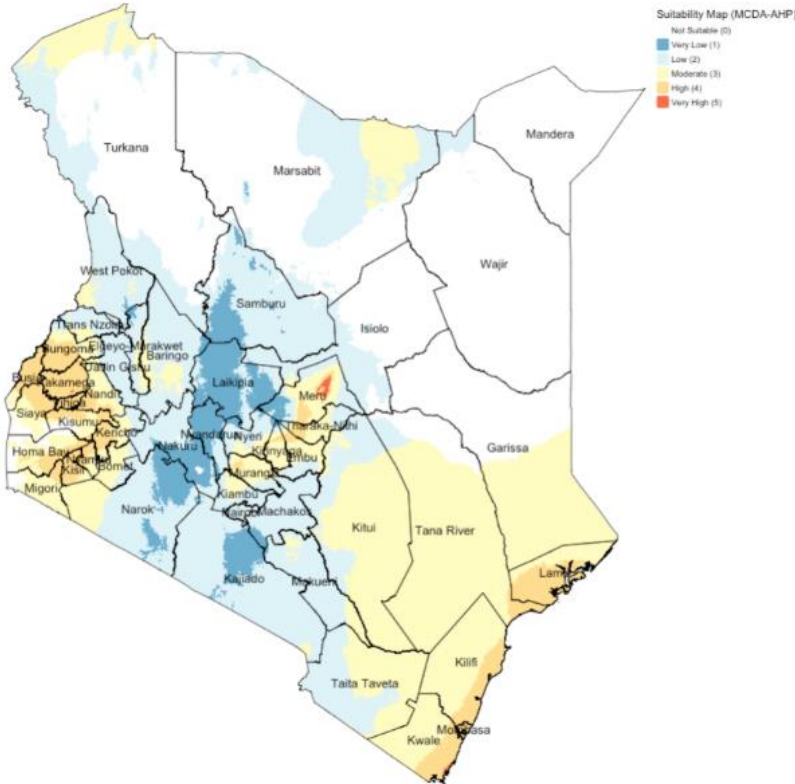


This is what we want. The p-values helps us to know whether the LISA estimates are statistically significant or not.

Simple Overlays: Booleans & Rankings

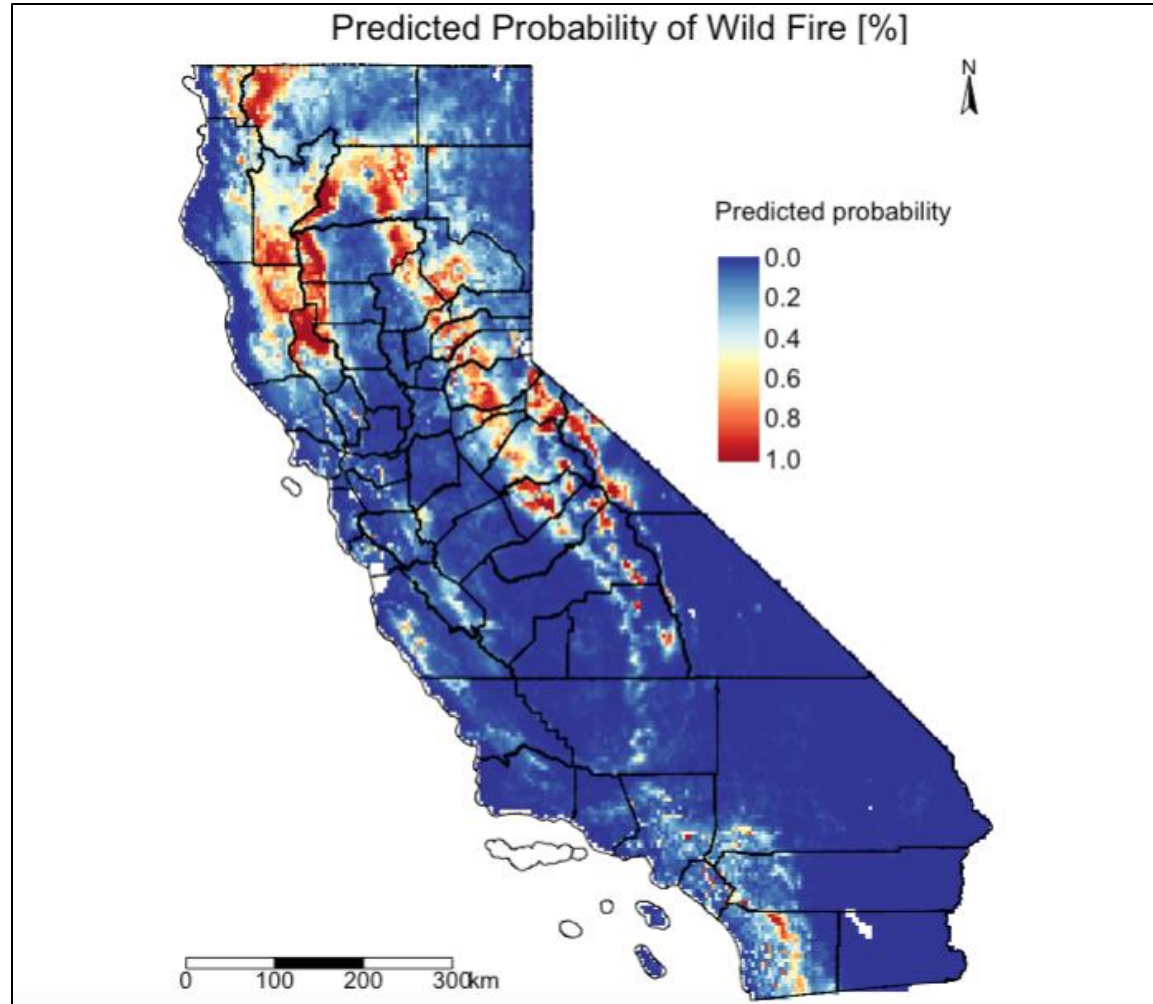


Multi-Decision Criteria Analysis

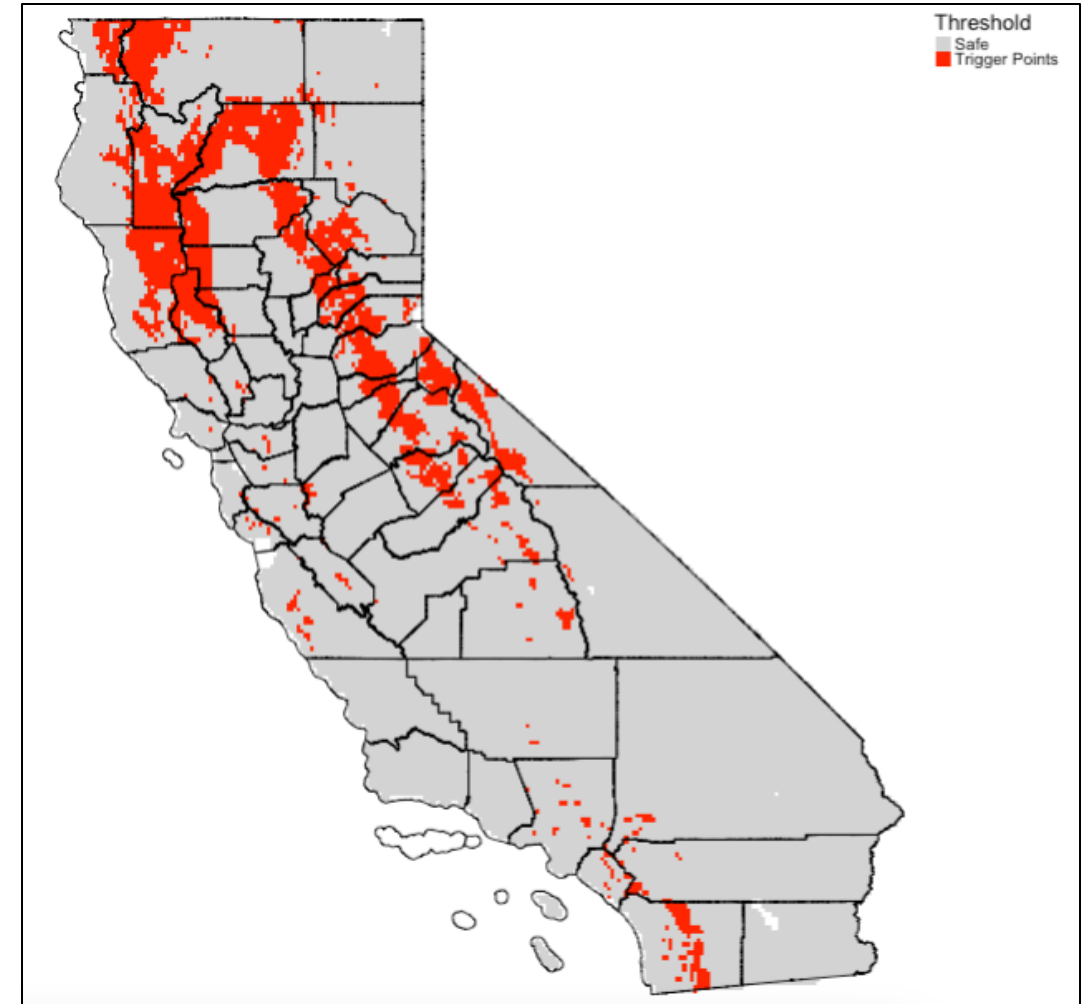


Week 5: Suitability Mapping II

Ecological Niche Models



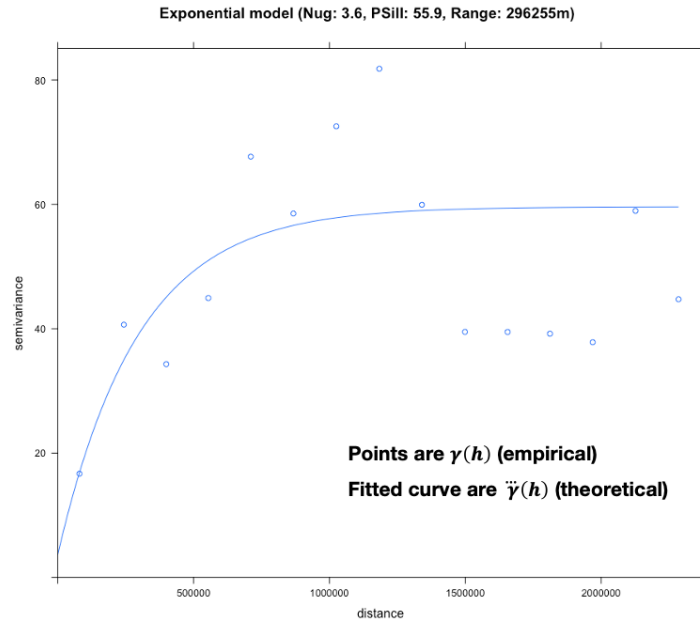
We mapped predicted probability of fires using the trained model after making sure its valid. The multi-band raster is fed to the trained model to make full scale predictions.



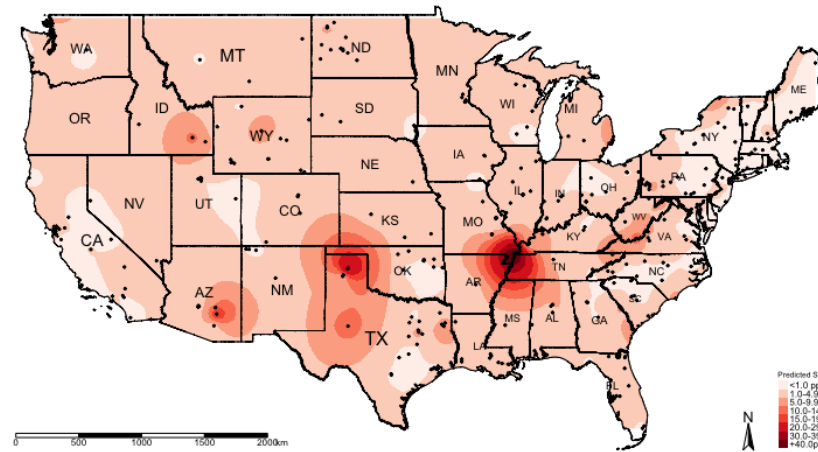
This is based on the optimized estimate obtained after model validation i.e., maximizes the True Positive Rate and the True Negative Rate is 0.4054474 (40.55%). Here, we mapped **predicted probability > 0.4054** as a reclassified raster.

Week 6: Geostatistical Modelling

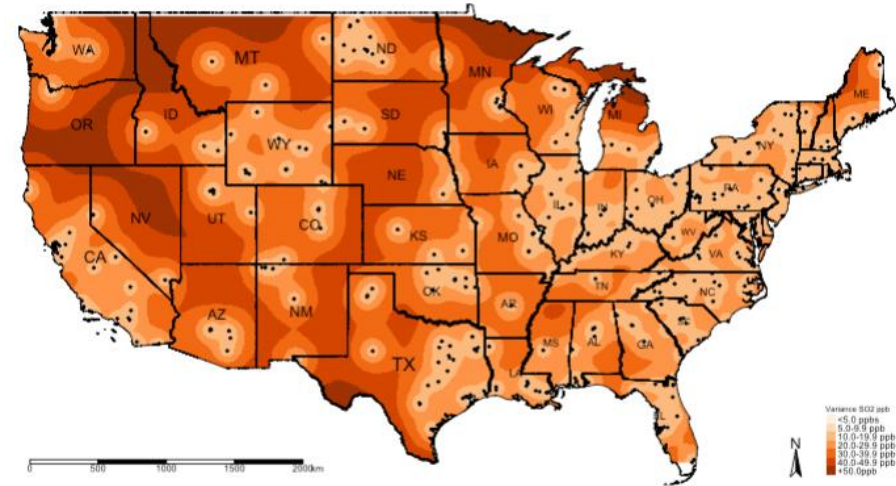
Variogram analysis



Predicted air SO₂ level from the Kriging model



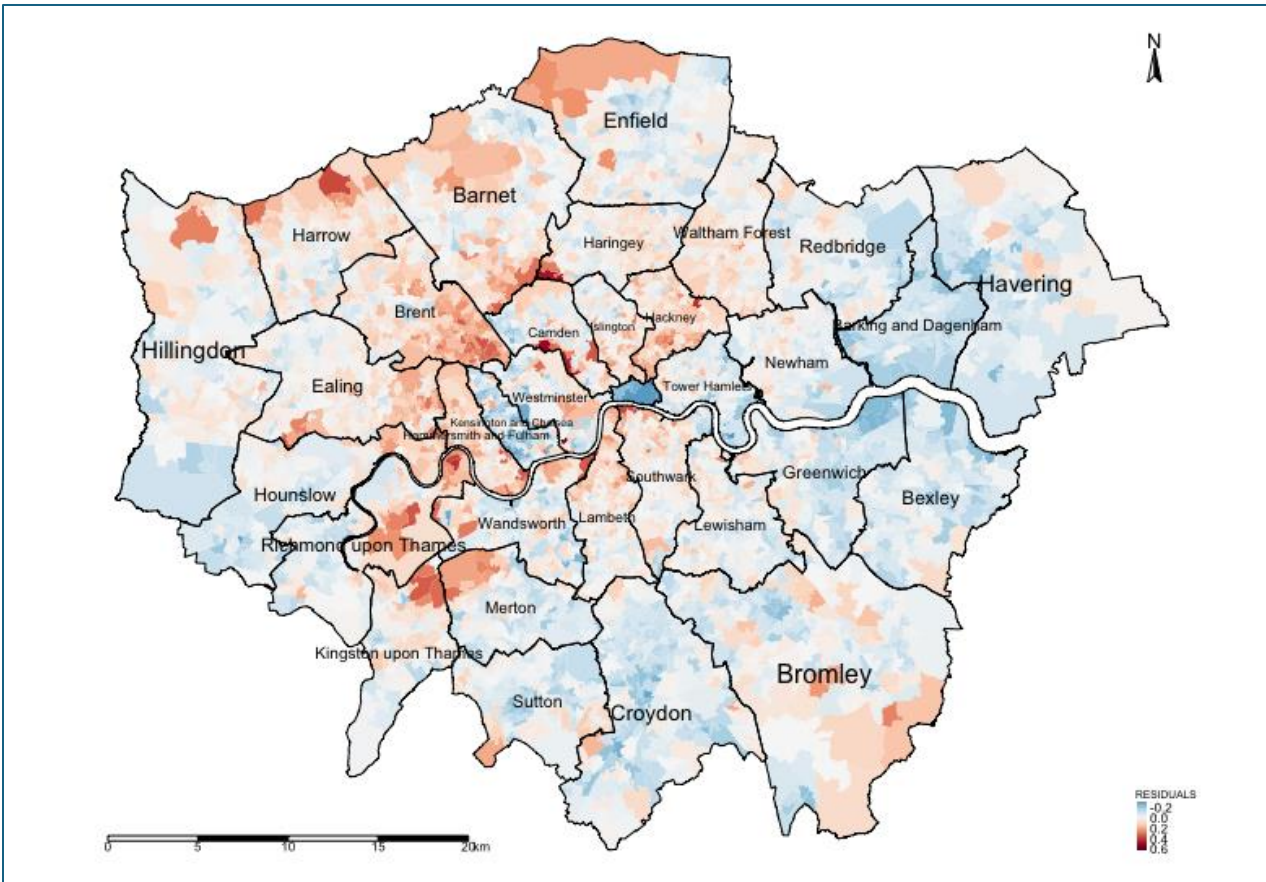
Uncertainty for the predicted air SO₂ levels from the Kriging model



Interpretation: The **nugget** is a small value of 3.6, which is an indication for evidence of spatial variability in the concentrations for SO₂ across sampling sites in USA. The **range** is 296,255m, which indicates that any separation distance above this value means that spatial autocorrelation in the observed levels of SO₂ between points are no longer similar. However, points with a separation distance less than 296,255m indicated that their SO₂ values are similar. For the **partial sill**, within this range for the Semivariance i.e., 3.6 and 55.9 – is the values are spatially autocorrelated.

Along the belt of the following states – Missouri, Tennessee, Kentucky and Illinois, the predicted concentration of SO₂ levels exceeds +40ppb, whereas there are pockets in Texas where concentrations of SO₂ are a cause for concern i.e., 30-39.9ppb.

Week 7 and 8: Spatial Regression and GWRs [1]



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.

Week 7 and 8: Spatial Regression and GWRs [2]

[1] Report overall spatial effect

Scenario 1: Spatial Lag Model (lagged on the dependent variable)

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

- Estimate the ρ coefficient for WY
- Estimate the impacts i.e., the direct and indirect effects

$\rho = 0.4522$; p-value < 0.05

Interpretation:

The ρ statistic informs us of how the neighbouring LSOA house price values affect the house price an index location (y). The p value is a positive value of 0.4522 which means the house price from neighbouring LSOAs have a positive manner, and it is statistically significant (i.e., p-value < 0.05).

We can see the AIC for the lag model is lower than the original linear regression model (i.e., Lag: -9863.3 vs LM: -8510.8) therefore the lag model is okay.

[2] Report the direct, indirect and total effects

Interpretation of the relationship between our independent variables and outcome from spatial regression model

	Direct	Indirect	Total
log(IMD score)	0.045% (p < 0.05)	0.034 (p < 0.05)	0.079% (p < 0.05)
log(Income)	1.267% (p < 0.05)	0.949 (p < 0.05)	2.217% (p < 0.05)
log(PTA Index)	0.011% (p < 0.05)	0.008 (p < 0.05)	0.019% (p < 0.05)

At specific location

At neighbouring
locations

Combined

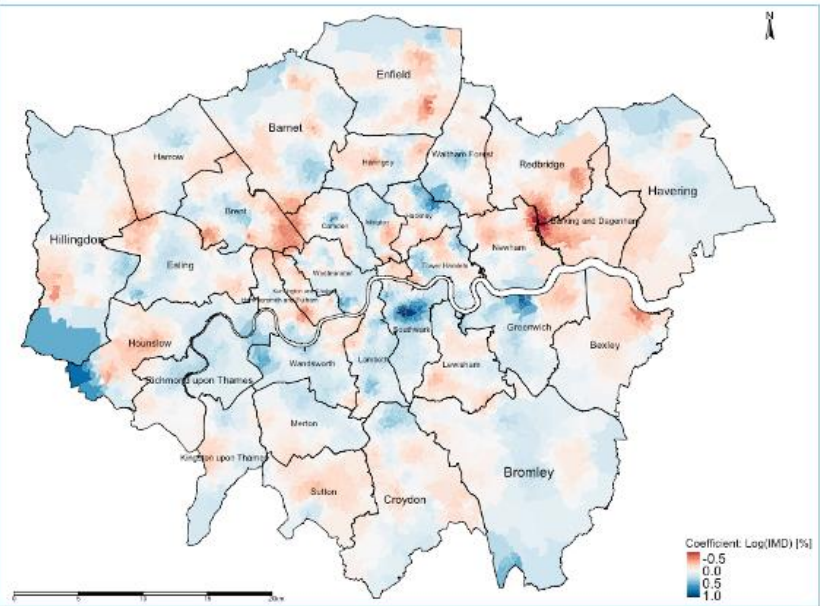
Interpretation:-

Direct effect for deprivation: This means that for every unit increase in deprivation (on log-scale) in a location (LSOA) leads to an average increase in the house prices by 0.045% in the **same location**, while accounting for spatial feedback (i.e., house price) from neighbouring locations (LSOAs) in London.

Indirect effect for deprivation: This means that for every unit increase in deprivation (on log-scale) in a location (LSOA) leads to average increase in the house prices by 0.034% in **neighbouring locations**, again, which accounting for its spillover effect through the spatial areas (LSOAs) in London.

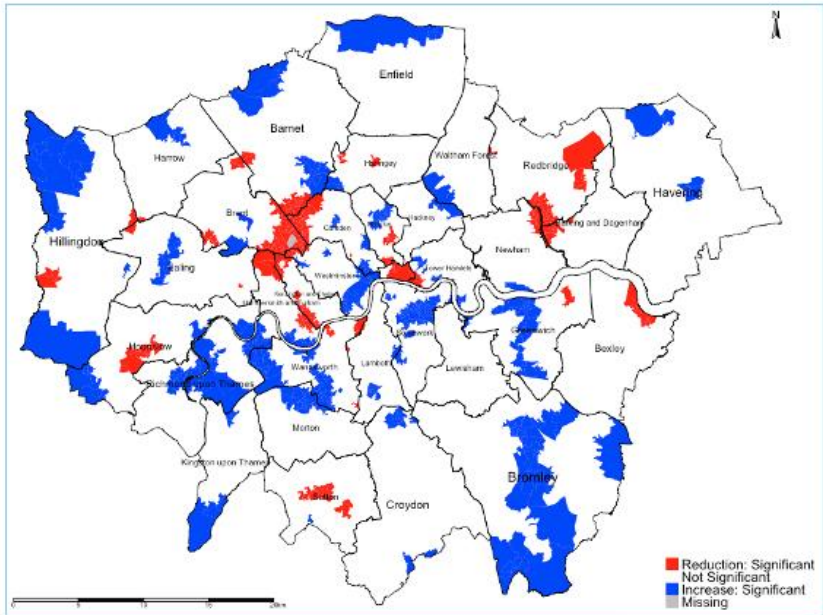
Week 7 and 8: Spatial Regression and GWRs [3]

[1] Local regression coefficients



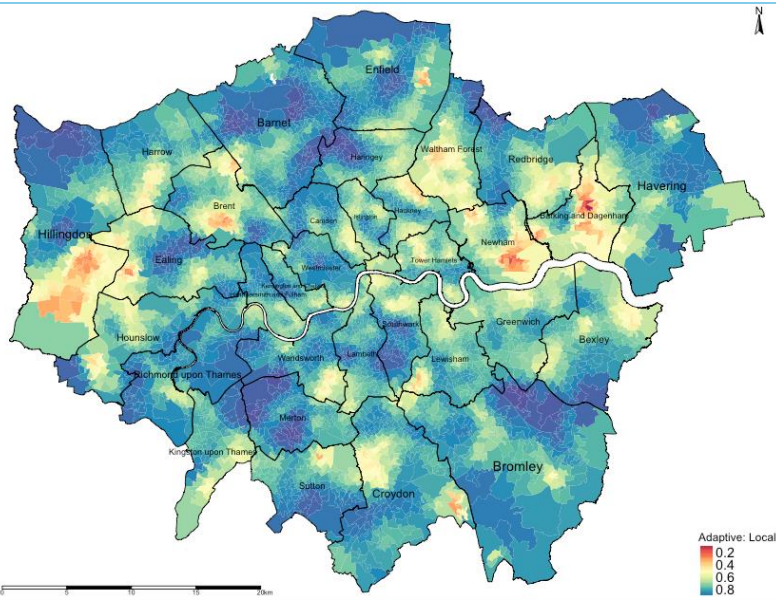
Interpretation: There is spatial variability in the relationship between our variable socioeconomic deprivation (transformed) and averaged house price (transformed) in London. The GWR outputs reveals that local coefficients range from a minimum value of -0.946 to a maximum value of 1.085, indicating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

[2] Statistical Significance



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 northeastern section of the Borough of Hounslow with pockets of LSOA'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.

[3] Local R² : Model Diagnostics



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.

Week 9: Spatial Bayesian Risk Models [1]

Report the overall risk results

2017	LIRAA 2	
	RR (95% CrI)	Pr(RR>1)
Intercept	1.64 (95% CrI: 0.14 to 7.07)	0.51
Temperature	0.93 (95% CrI: 0.74 to 1.12)	0.23
Precipitation	1.01 (95% CrI: 0.96 to 1.07)	0.73
NDVI	1.09 (95% CrI: 0.71 to 1.60)	0.63
Urbanisation	1.18 (95% CrI: 0.37 to 2.90)	0.52

RR: Relative risks; Pr(RR > 1): Exceedance probabilities (the probability that RR being greater than 1)

Interpretation (examples):

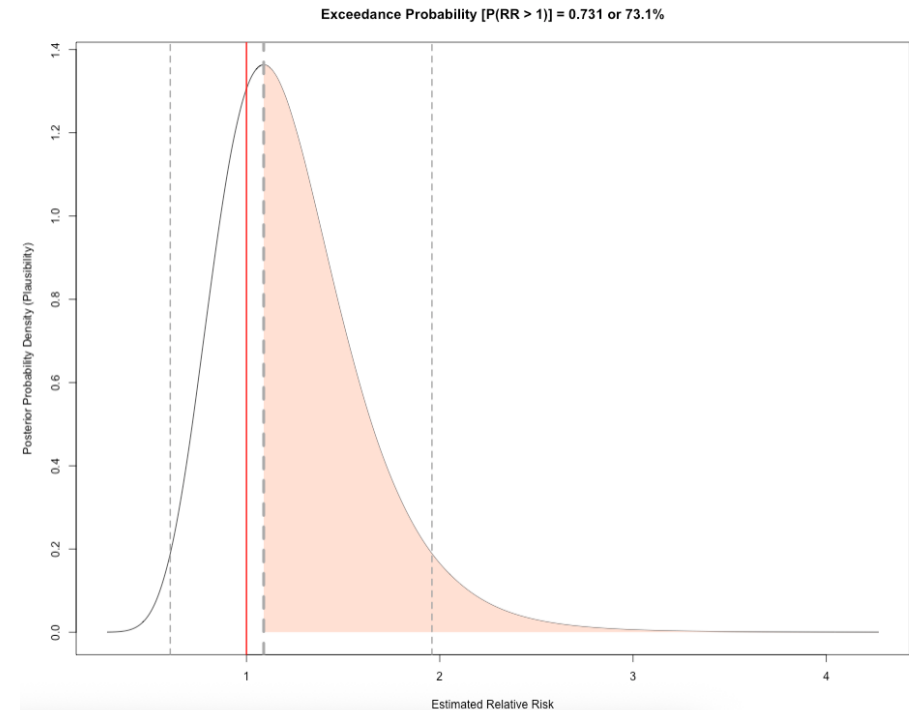
- **Intercept:** The overall baseline risk of mosquito-borne infestation is 1.64 times (or 64%) **greater** in the population of Campina Grande. The overall probability that there's excess risk of infestation (i.e., $RR > 1.00$) is 51%.
- **Temperature:** In relation to temperature, the risk of mosquito-borne infestation is 0.93 times (or 7%) **lower** in Campina Grande. The probability of observing an excess risk of infestation (i.e., $RR > 1.00$) in relation to temperature is 23%.
- **Urbanisation Index:** In relation to urbanisation, the risk of mosquito-borne infestation is 1.18 times (or 18%) **higher** in Campina Grande. The probability of observing an excess risk of infestation (i.e., $RR > 1.00$) in relation to urbanisation is 52%.

NOTE: All relative risk estimates have the null value (1) between its lower and upper 95% credibility intervals. While the results, excluding temperature, show an increased risk of infestation – **these are all statistically not significant.**

Describe the distribution of the relative risks

For the computer practicals, we showed all the possible values of how temperature is related to mosquito infestation in terms of impact and risk given the dataset we have. The relative risk for mosquito infestation given temperature can be anywhere from 0.278 (reduced risk) to 4.271 (increased risk). But from the graph, it indicates that the most plausible relative risk estimate is 1.089 because it has the highest density in our posterior distribution.

To gauge levels of certainty about this estimated risk, we can calculate the probability that the relative risk is 1.089 or above. The probability of observing a relative risk of mosquito infestation being 1.089 or above, given the environmental levels of temperature, is 0.731 or 73.1%, which is quite high

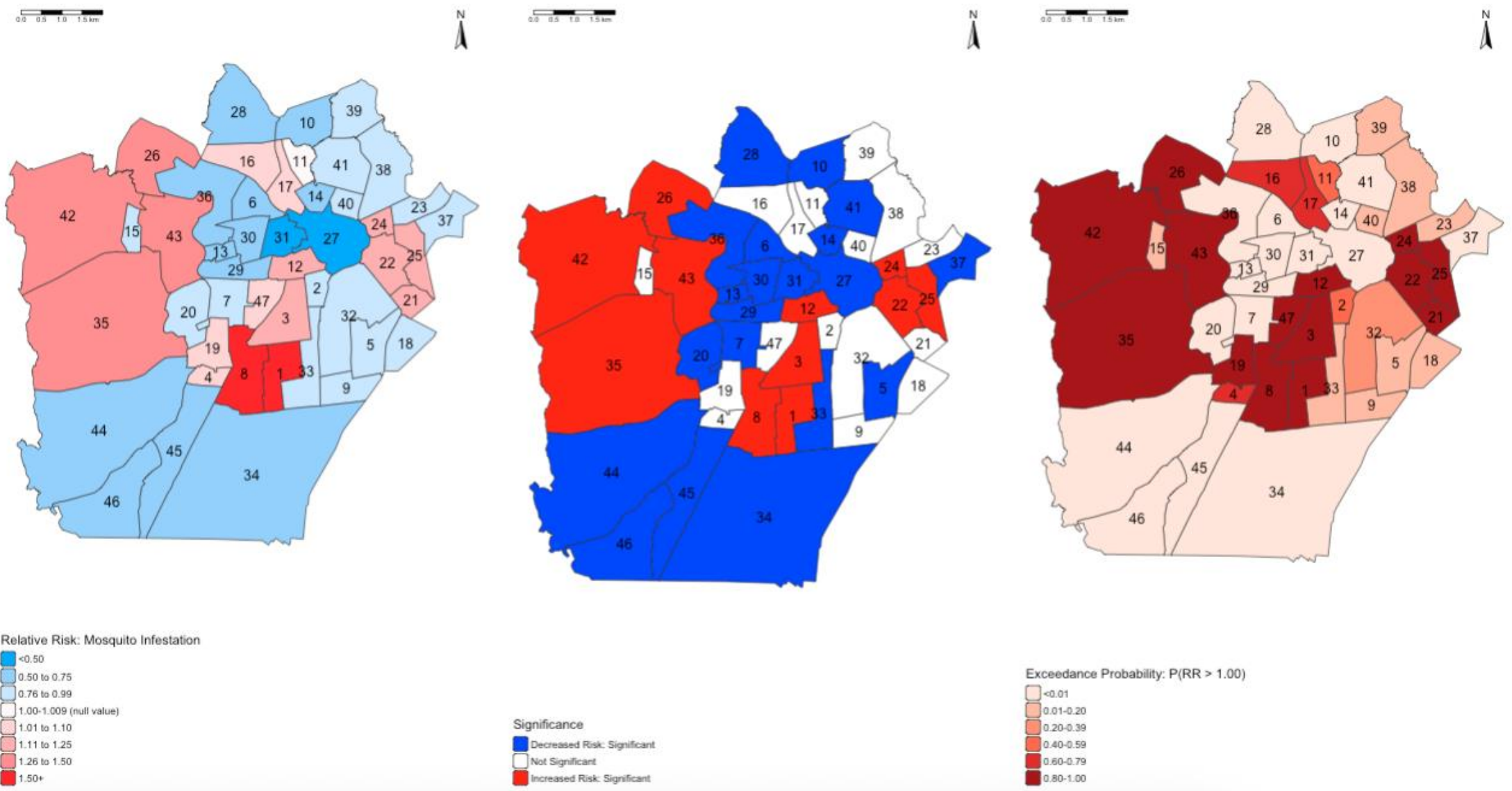


Week 9: Spatial Bayesian Risk Models [2]

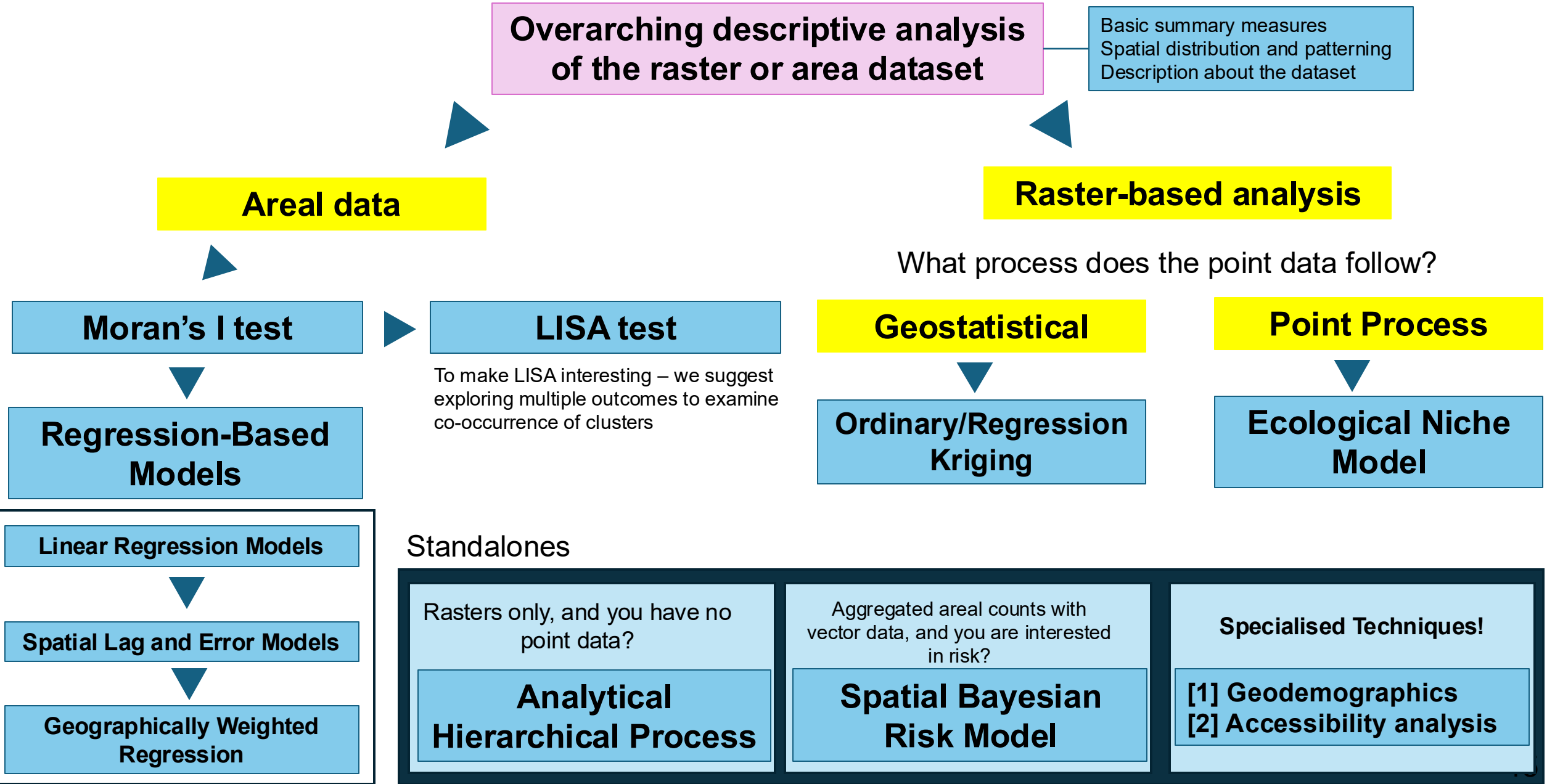
[1] Area-specific risks

[2] Statistical Significance

[3] Exceedance Probabilities



Assessment



Specialised Techniques

- [1] Geodemographics**
- [2] Accessibility analysis**

❖ **Geodemographics Classification (Week 7) taught on GEOG0114 by Dr Justin Van Dijk (Sabbatical)**

This is a clustering technique. Geodemographic classification is a method used to categorise geographic areas and the people living in them based on demographic, socioeconomic, and sometimes lifestyle characteristics.

[1] **Geodemographic Classification** <https://jtvandijk.github.io/GEOG0114/01-geodemographics.html>

Other excellent tutorials

[2] **Using R for Geodemographic Analysis** <https://rpubs.com/nickbearman/r-geodemographics>

[3] **Geodemographics** <https://rpubs.com/nickbearman/geodemographics>

❖ **Accessibility Analysis (Week 8) taught on GEOG0114 by Dr Justin Van Dijk (Sabbatical)**



This technique describes the ease with which individuals can reach places and opportunities, such as employment, public services, and cultural activities. We can utilise transport network data to quantify accessibility and characterise areas based on their accessibility levels.

[1] **Accessibility Analysis** <https://jtvandijk.github.io/GEOG0114/02-network.html>

[2] **Introduction to Urban Accessibility:** https://ipeagit.github.io/intro_access_book/



Go to Moodle’s Assessment

Assignment Documentation: Guidance notes and Cover Sheet

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

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

198.6 KB



 [GEOG0114 Spatial Data Analysis Project - Cover Sheet template](#)  DOCX

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 2022/23 to present

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



2.2 MB

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



2.2 MB

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



10.6 MB

 [Effect of public service facilities access on house prices in Birmingham, UK \(2023/24\)](#)  PDF



667.2 KB

 [Who Gets to Be Healthy in Peninsular Malaysia? A geodemographic exploration of healthcare and fitness \(HEFI\) accessibility \(2023/24\)](#)  PDF



1.5 MB

 [Universal Kriging estimation of the annual concentration of nitrogen dioxide in the urban area of Bogota \(2024/25\)](#)  PDF

4.3 MB

 [Spatial Analysis for Early COVID-19 Mitigation Strategies: A Case Study of Jakarta, Indonesia at the Ward Level \(2024/25\)](#)  PDF

1.8 MB

 [Spatial Analysis Project: Using GWR to Understand the Drivers Behind Stop and Search Use in London \(2024/25\)](#)  PDF

1.4 MB

Spatial Data Analysis Project
Assignment’s deadline is Monday 12th January 2026 (by 12:00pm)

Example of Data Source(s)

I would advise finding your own dataset you are interested in.

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 (Geographic Data Service/Healthy and Sustainable Places): <https://data.cdrc.ac.uk>
- London Data Store: <https://data.london.gov.uk>
- DEFRA Air Pollution Information (real-time monitoring data) at all Automatic Urban & Rural Network stations. This dataset can be pulled into RStudio directly through an API using the R package **openair** (<https://openair-project.github.io/book/>)

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> (main website) but all dataset can be pulled into RStudio directly through API using the R package **malariaAtlas** (<https://github.com/malaria-atlas-project/malariaAtlas>)

Disaster dataset

- EM-DAT International Disaster Database: [\[LINK\]](#)
- FIRMS Firm Information for Resource Management System: [\[LINK\]](#)

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>

Country shapefiles: https://gadm.org/download_country.html

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>

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.

Dos and Don'ts

The 'Dos' and 'Dont's'

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

Education Level & Suicide Rate

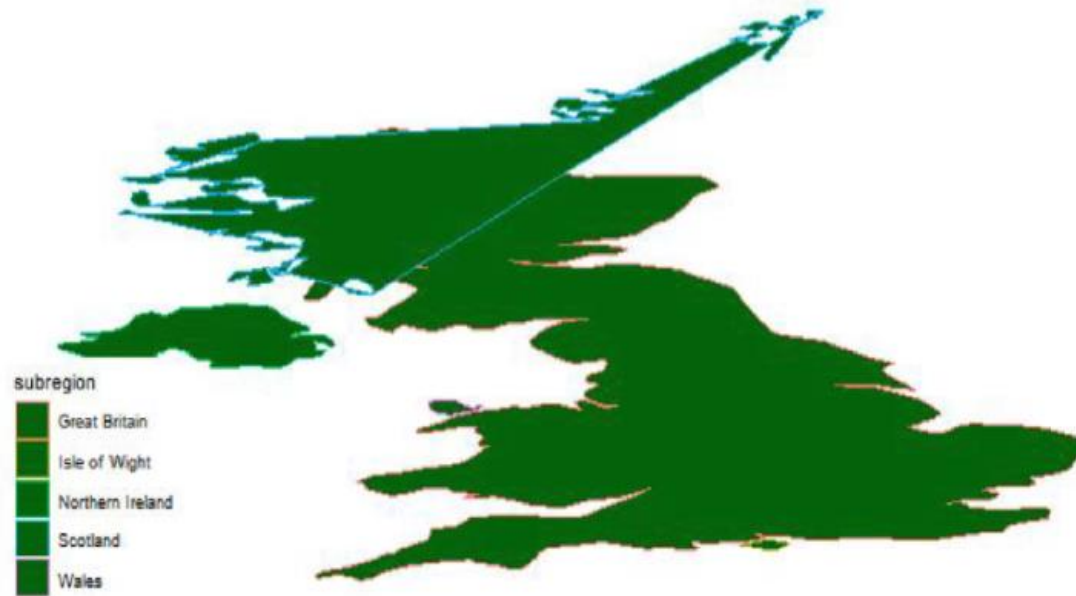


Figure 1 The Distribution map of London

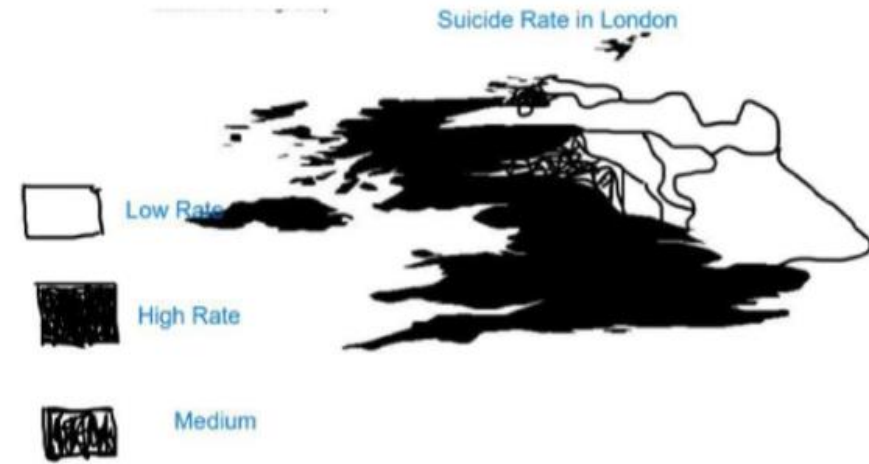


Figure 2 Age Group and Suicide

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 for detail.

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:

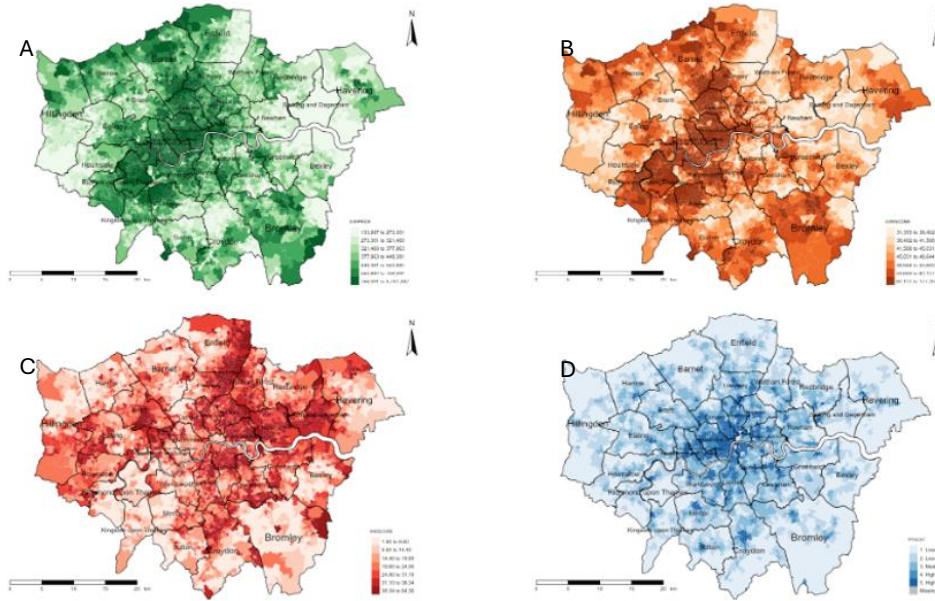
Even worse... we triple dare you to take a screenshot and paste it into your assignment...



“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

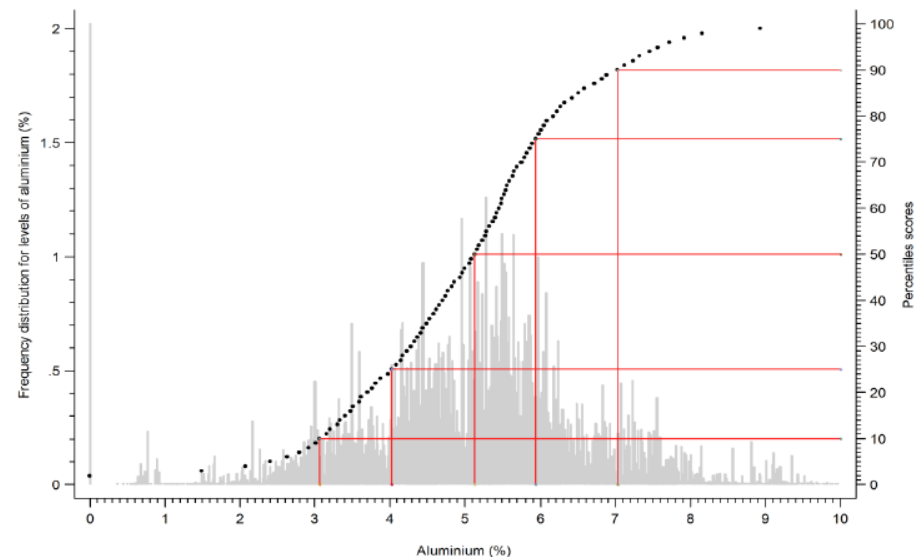


Figure 3.3: Two-way histogram with cumulative proportions showing the overall distribution of patients in THIN-GBASE with specific soil concentration levels for aluminium. Left y-axis: corresponds to the observed proportion of patients with specific soil levels of aluminium. Right y-axis: Black dots 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.

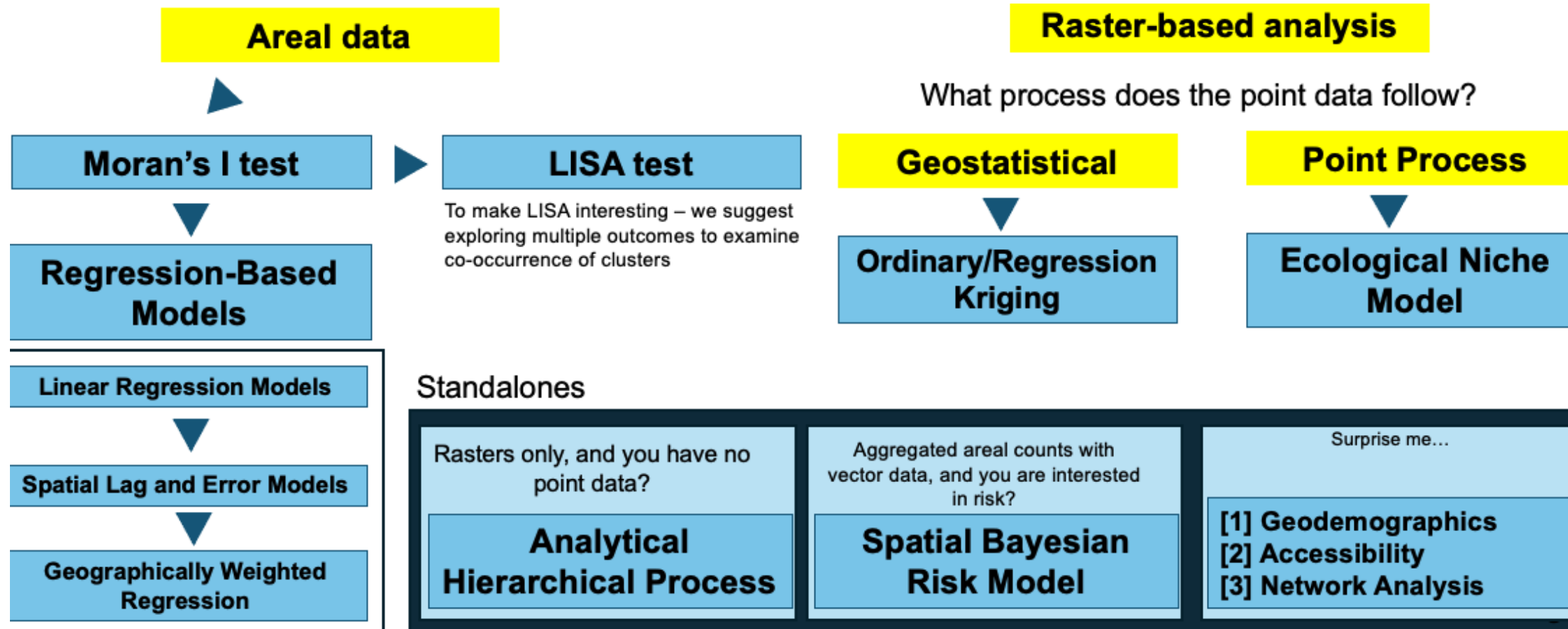
Show the image in its glory to the fullest by taking advantage of the landscape settings

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.

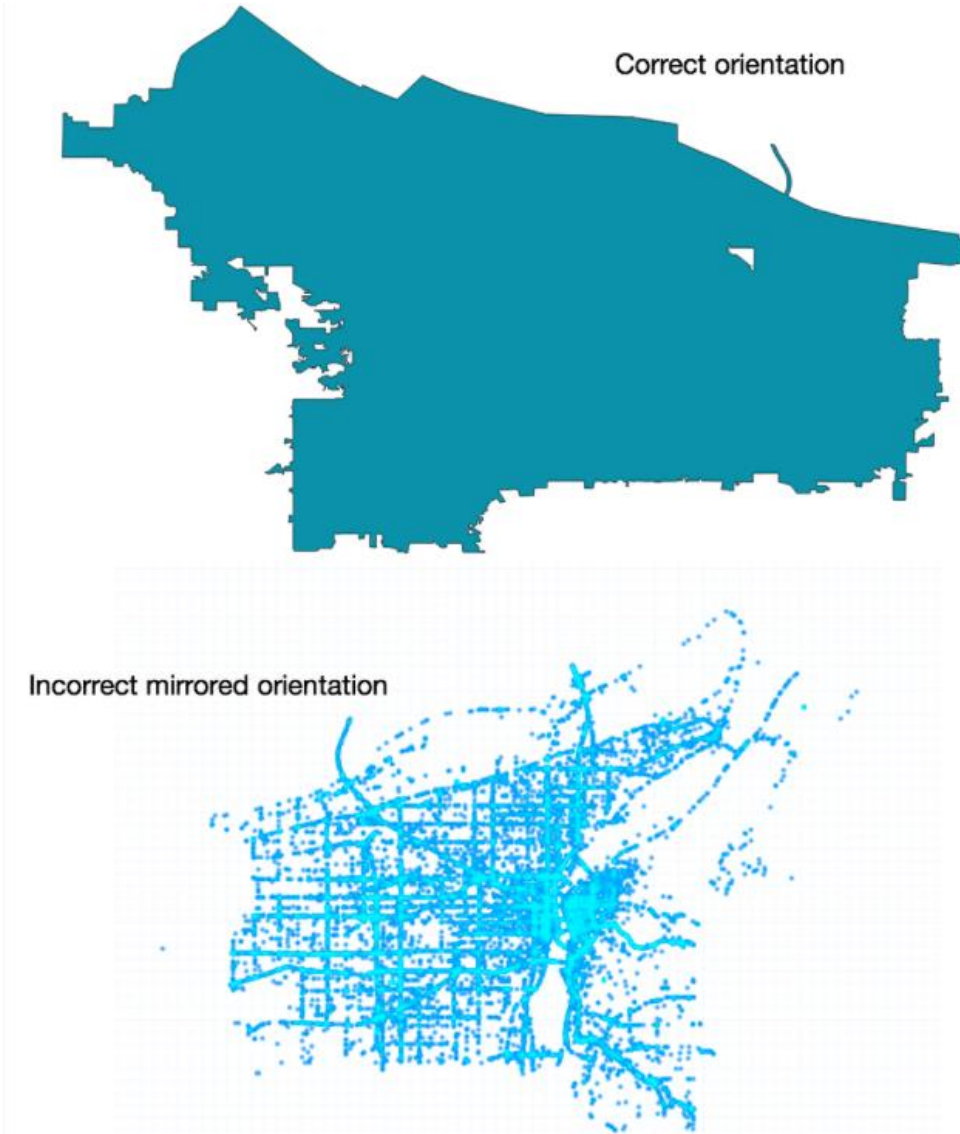
Markers appreciate this level of eye-for-detail and are not above being sympathetic in awarding extra marks for quality in presentation.

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. Meaning you won't be able to achieve anything past a 68/69. There is no way a distinction mark is awarded to a piece of work where **CRITICAL ERRORS** are present!
- If it's a horrible violation, to such an extent its irredeemable, then in such scenario can potentially lead to a failing grade.

Make sure to understand the structure of your dataset and due diligence in the data cleaning!



- Make sure the orientation of you layer and points are correct
- Make sure longitude and latitude are in correct order, and not flipped or mirrored!
- Make sure to use the correct Coordinate Reference System (CRS)
- If the data quality is poor from the original source – especially with the points – clean it to the best of your ability by removing the broken points and retaining those that are accurate. **DO NOT ATTEMPT TO ALTER THE POINTS IN ANYWAY AS YOU RISK INTRODUCING MORE PROBLEMS.**
- Shapefiles for areas as well, after you've cleaned them and linked with non-spatial data shouldn't have holes in them
- If you run an analysis with these problems, them it means that you have not learned anything on stuff we've covered together on data handling.
- I consider this to be a critical error, which compromises the result. This type of horrible mistake, in my view, is irredeemable, so in such scenario we are look at a failing grade.

The ‘Dos’ and ‘Dont’s’

Use of Generative AI (e.g., ChatGPT) and References [1]

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

DO NOT USE AI TO FULLY GENERATE YOUR REPORT OR REFERENCE LIST!

Automatic failure and UCL disciplinary action & punishment would be taken against you.

References

Use a Reference Management Tool

Reference management software can save a lot of time, and heartache, when dealing with research papers and citation in text.

The following are popular tools:

1. **Zotero**: Free, intuitive, and it integrates with various browsers (Explorer, Safari, Mozilla, Chrome etc.,) and word processors
2. **Mendeley**: Free tool akin to Zotero - provides collaborative options between users (e.g., sharing of papers, references etc.)
3. **EndNote**: Paid but offers advanced features for professional use
4. **BibTeX**: Reference manager for programmatic software systems like [Overleaf](#), [RMarkdown](#), [Quarto](#) or [LaTeX](#).



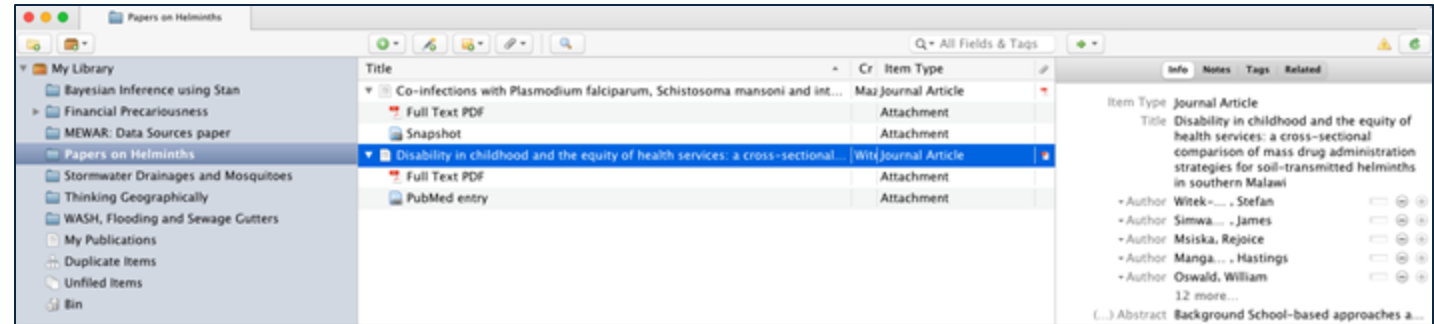
Literature review for the research question:

What is the geographical burden of soil-transmitted helminths in Sub Saharan Africa?

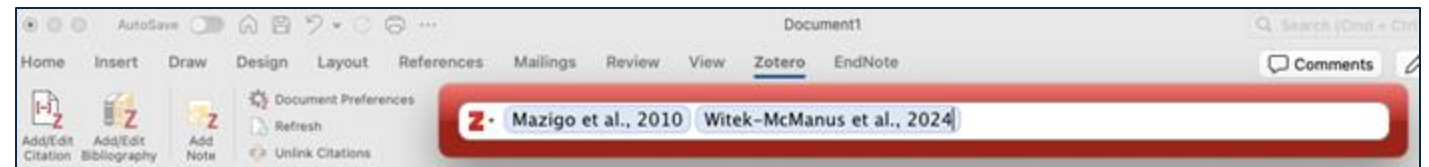
[1.] This button exports the paper into your reference manager



[2.] The user interface of reference manager. You can manage your list of references efficiently, and even read, annotate and categorise papers accordingly to the research theme for literature review



[3.] Reference manager is automatically connected to your word processor (e.g., Microsoft Word), which allows the efficient use of adding in-text citation and building your list of references on the fly.



[4.] The output will be a word document two things: a) body of text with in-text citation and b) bibliography at the end of the document.

Soil transmitted helminths inflict significant morbidity in children, and adults. The infection is mostly in various countries of Sub-Saharan Africa (Mazigo et al., 2010; Witek-McManus et al., 2024).

References

- Mazigo, H. D., Waihenya, R., Lwambo, N. J., Mnyone, L. L., Mahande, A. M., Seni, J., Zinga, M., Kapasa, A., Kweka, E. J., Mahana, S. E., Heukelbach, J., & Mkoji, G. M. (2010). Co-infections with *Plasmodium falciparum*, *Schistosoma mansoni* and intestinal helminths among schoolchildren in endemic areas of northwestern Tanzania. *Parasites & Vectors*, 3(1), 44. <https://doi.org/10.1186/1756-3305-3-44>
- Witek-McManus, S., Simwanza, J., Msiska, R., Mangawah, H., Oswald, W., Timothy, J., Galagan, S., Pearman, E., Shaikh, M., Legge, H., Walson, J., Juziwele, L., Darvey, C., Pullan, R., Bailey, R. L., Kalua, K., & Kuper, H. (2024). Disability in childhood and the equity of health services: A cross-sectional comparison of mass drug administration strategies for soil-transmitted helminths in southern Malawi. *BMJ Open*, 14(9), e083321. <https://doi.org/10.1136/bmjopen-2023-083321>

Guidance on setting-up free online reference management software



Useful video tutorials on Zotero and Mendeley

Any questions?



原作 矢立肇

