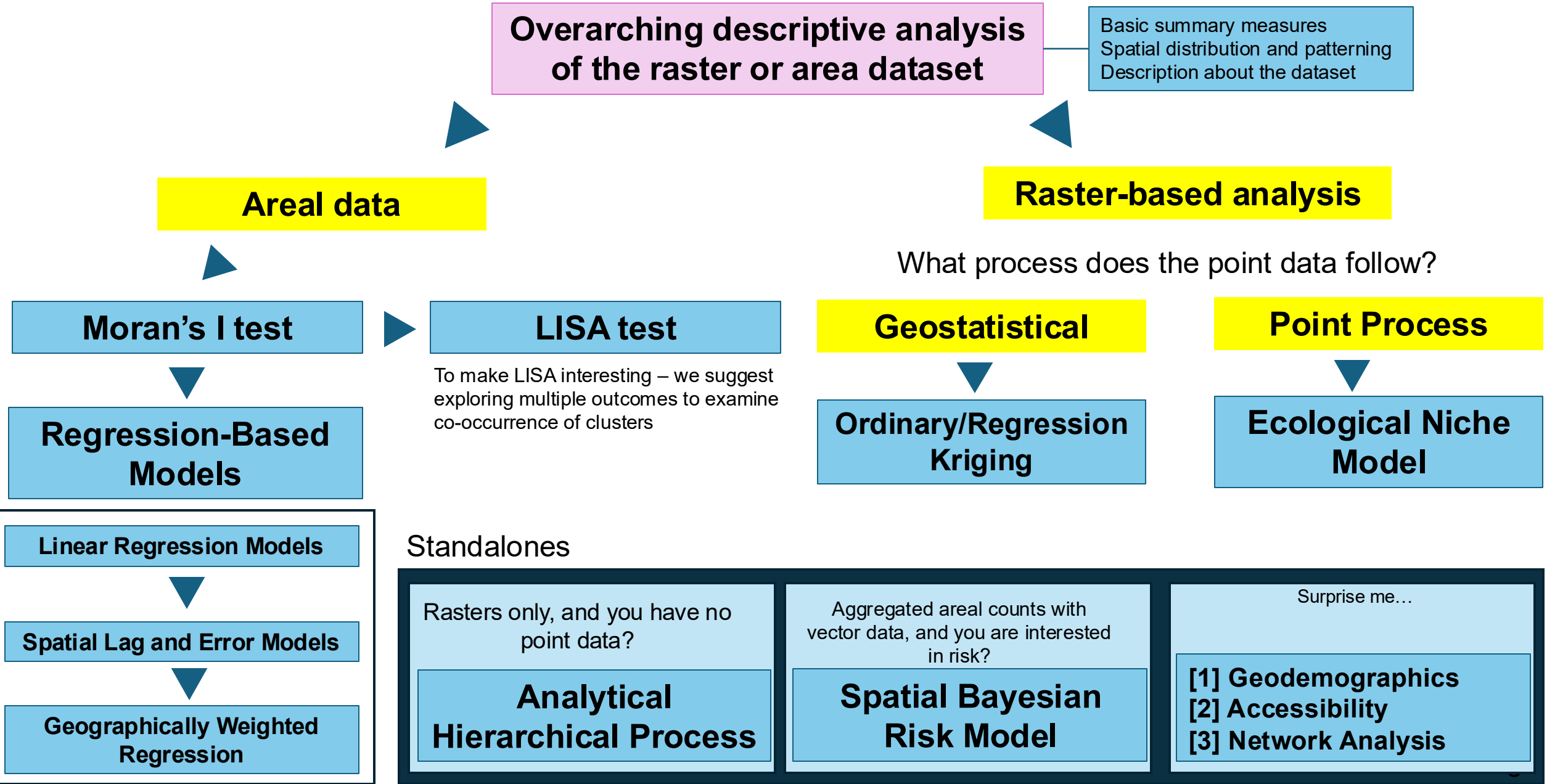


# Summary of GEOG0114 & Assessment

# 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	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



# 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](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>

## 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](#)  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 12<sup>th</sup> January 2026 (by 12:00pm)

## 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+
<b>1. Introduction, argument and conclusions</b>	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%.
<b>2. Engagement with literature</b>	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.
<b>3. Analysis: reflection, thought, &amp; conceptual framework</b>	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.
<b>4. Empirical knowledge &amp; use of examples (as relevant)</b>	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.
<b>5. Structure, writing, illustrative material and references (as relevant)</b>	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

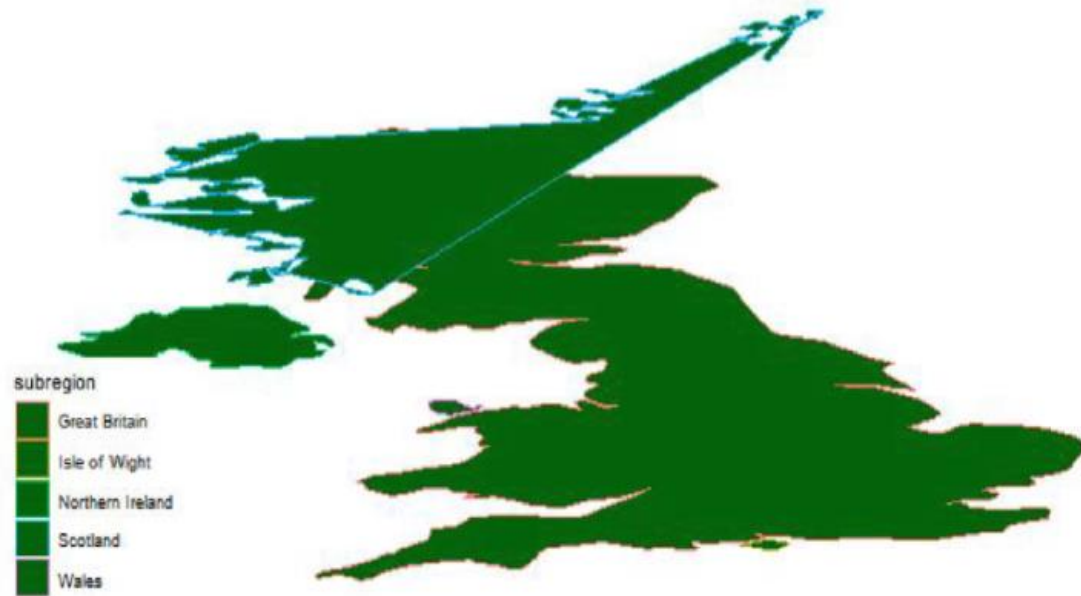


Figure 1 The Distribution map of London

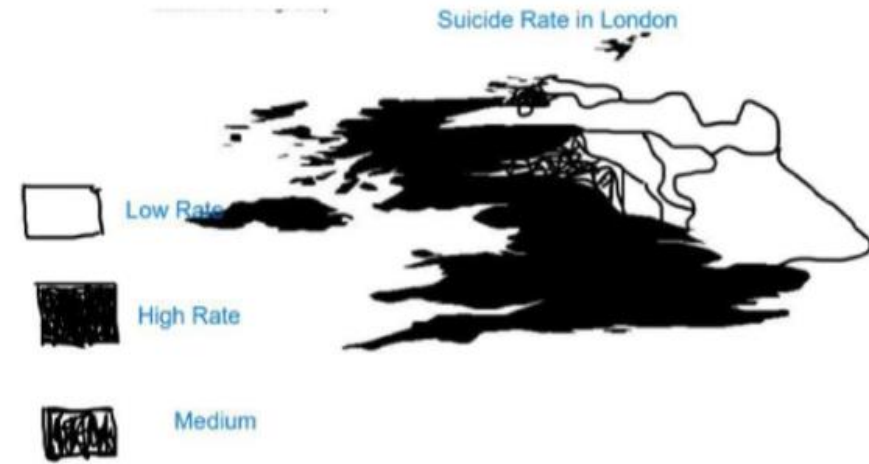


Figure 2 Age Group and Suicide

**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:

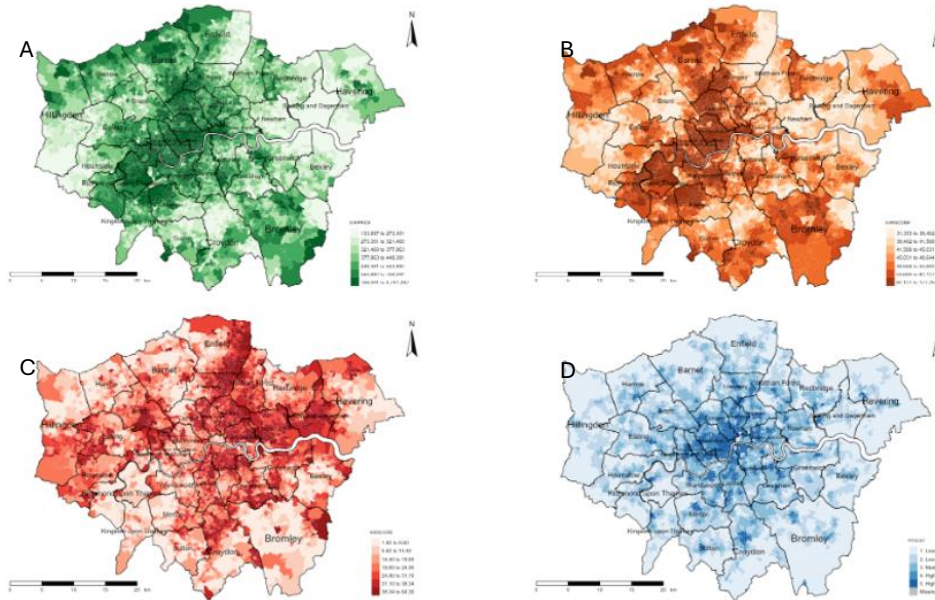
Even worse... we triple dare you to take a screenshot and paste it into your assignment...



**“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...

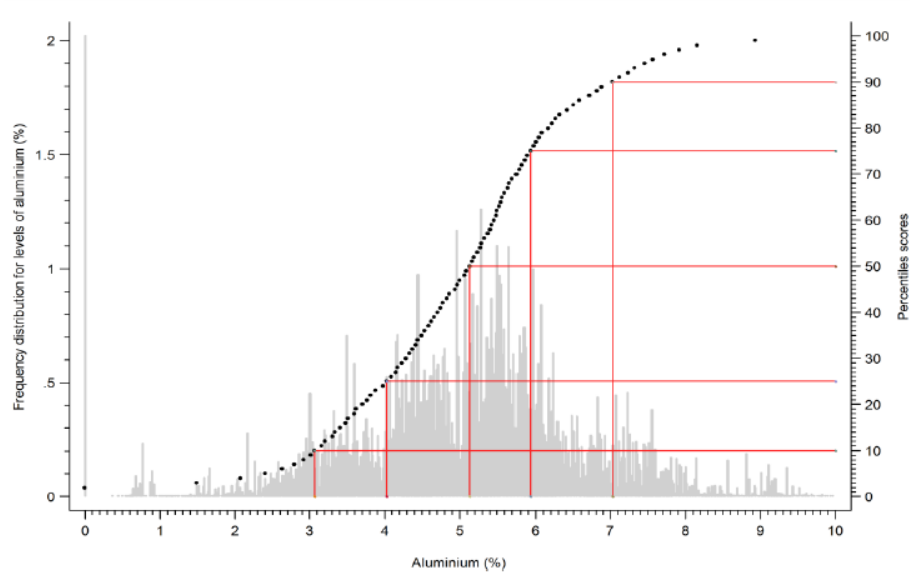


Figure 3.3: Two-way histogram with cumulative proportions showing the overall distribution of patients in THIN-GBASE with specific soil concentration levels for aluminium. Left y-axis: corresponds to the observed proportion of patients with specific soil levels of aluminium. Right y-axis: Black dots correspond to a percentile score - i.e. the proportion of patients that fall under specific soil concentration value for aluminium; Red line indicates: 10<sup>th</sup>, 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup> and 90<sup>th</sup> 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

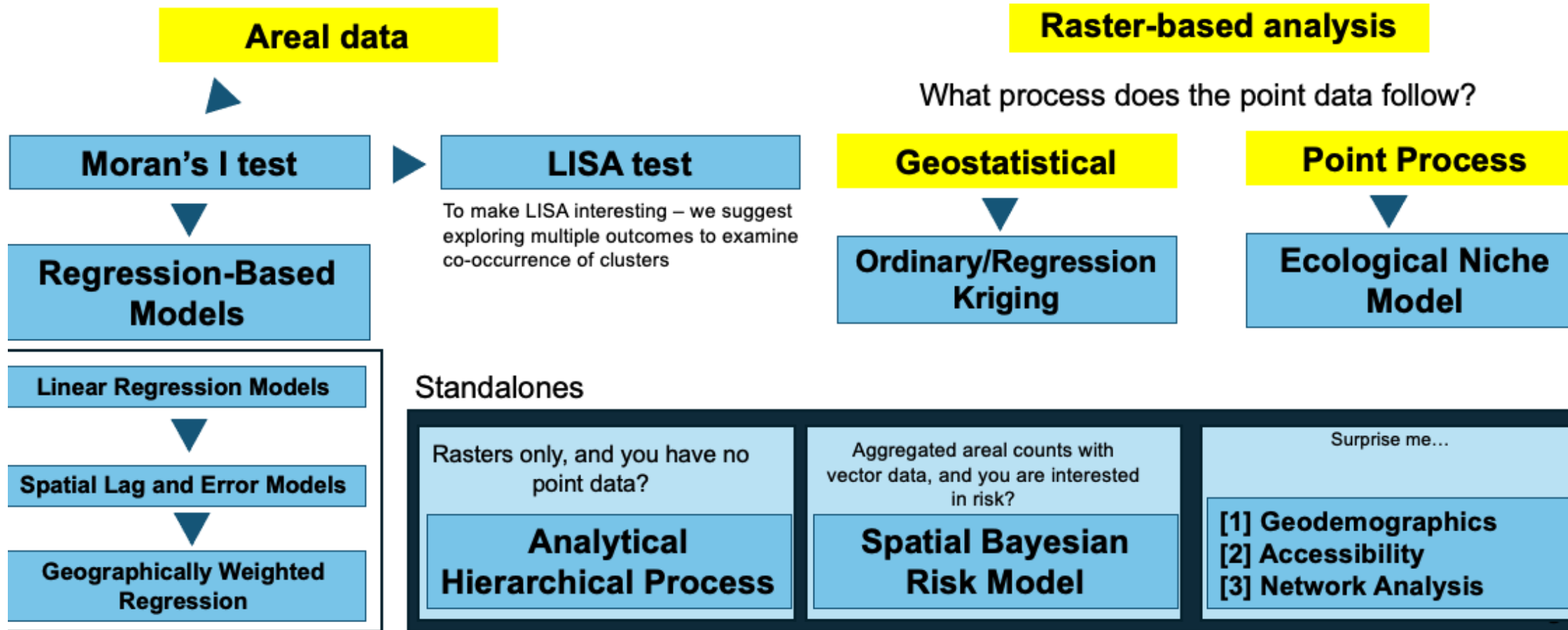
This is a crude example to illustrate what I mean about showing your results nicely to its fullest glory.

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

**Any questions?**



See you in term 2

原作 矢立肇

