****

*Spatial Data Analysis Portfolio  
An exploration of spatial dependence, regression, and surface analysis using QGIS, GeoDa, and SpatiaLite*

***School of Architecture, Computing and Engineering***

**DS7002 2425 T2 CW1 (exercise portfolio) Final**

Academic Year: 2024–2025

**Submitted by: PIYUSH SINGH**

**Student ID: 2807188**

**Supervisor: Yang Li**

Contents

[🧭 SESSION 5 Task 1 – Spatial Extent of London 3](#_Toc196606153)

[🔍 Part A: Summary of Extreme Boroughs (Spatial Measures Table) 3](#_Toc196606154)

[🧮 Part B: Binary Classification of Spatial Extremes (Flag Table) 4](#_Toc196606155)

[SESSION 5 Task 2: Neighbour Boroughs Join tables with multiple conditions. 4](#_Toc196606156)

[SESSION 5 Task 3: An Isolated Island in London? 5](#_Toc196606157)

[SESSION 5 Task 4 – Count of Underground Stations by Borough 6](#_Toc196606158)

[📌 SESSION 5 Task 5 – Interpretation: Near Neighbour Business / Services 7](#_Toc196606159)

[📌 SESSION 5 Task 6 –Motorways Across London Boroughs 8](#_Toc196606160)

[SESSION 5 Task 7 – Underground Stations within ch km of Motorways 10](#_Toc196606161)

[SESSION 5 Task 8: Buffer Zones Around Motorways 11](#_Toc196606162)

[📊 SESSION 6 Task 2: Intersect (AND) – Results and Interpretation 15](#_Toc196606163)

[📊 SESSION 6 Task 3: Voronoi Polygons – Results and Interpretation 16](#_Toc196606164)

[🌄 SESSION 6 Task 4: Height Extraction – Results and Interpretation 17](#_Toc196606165)

[📝 SESSION 6 Task 6 – Handling Non-Spatial Attributes: Population Density and Household Size 18](#_Toc196606166)

[Session 7 Task 1: Obesity Raw Rate Map (GeoDa) 20](#_Toc196606167)

[🚨 SESSION 7 Task 2: Thematic Mapping of Crime (QGIS) 22](#_Toc196606168)

[Session 8 Task 1: Hotspot Map of GP Surgeries in Haringey 23](#_Toc196606169)

[🗂️ SESSION 8 Task 4: Accessibility to GP Surgeries from Tube Stations in Haringey 24](#_Toc196606170)

[Session 9 Task 1: Spatial Dependence / Correlation of Crimes and Deprivation in London 25](#_Toc196606171)

[Burglary Relative to Total Crime (EB Rate – Local Moran’s I) 25](#_Toc196606172)

[🔹 1.1 Spatial Dependence: Deprivation Scores in London (Univariate Moran’s I) 26](#_Toc196606173)

[🔹 Spatial Dependence: Total Crime Rate (CRIME\_1000 – Univariate Moran’s I) 27](#_Toc196606174)

[🔹 Bivariate Spatial Dependence: Deprivation vs Crime Rate 28](#_Toc196606175)

[🔹 Bivariate Spatial Dependence: Deprivation vs Burglary 29](#_Toc196606176)

[🔹 Spatial Dependence: Drug Offences Relative to Total Crime 30](#_Toc196606177)

[🔹 Spatial Dependence: Ward-Level Crime Rates in London 31](#_Toc196606178)

[🔹 Spatial-Temporal Dependence: Population Change (2011–2016) 33](#_Toc196606179)

[🔹 SESSION 9 Task 2: Spatial Regression of Health Indicators in London 33](#_Toc196606180)

[📄 Section 1: GeoDa Regression Outputs 33](#_Toc196606181)

[📄 Section 2: R Regression Outputs 37](#_Toc196606182)

[📄 Section 3: Comparison and Discussion 38](#_Toc196606183)

[🟩SESSION 9 Task 3: Population Surface in Haringey 39](#_Toc196606184)

# 🧭 SESSION 5 Task 1 – Spatial Extent of London

**🗺️ Task 1: Spatial Extent of London**

**🎯 Objective:**

The goal of this task was to analyze the spatial extent and population density of London boroughs, and identify those with spatial extremities such as the largest area, smallest area, highest and lowest population density, and extreme geographic boundaries (north, south, east, west).

## 🔍 Part A: Summary of Extreme Boroughs (Spatial Measures Table)

A new table london\_guinness\_book was created from the main london\_boroughs dataset. Geometry attributes such as area and bounding box coordinates were calculated using SQL spatial functions (ST\_Area, MbrMaxY, MbrMinX, etc.). Population density was derived by dividing each borough’s population by its area in square kilometers.

A subset of boroughs was selected based on:

* Maximum and minimum area
* Highest and lowest population density
* Northernmost, southernmost, easternmost, and westernmost locations

These records were retrieved using nested subqueries with UNION operations, ensuring only boroughs with spatial extremes were included.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| London Borough | Region | Area (sq.km) | Density (Population/sq.km) | North | South | East | West |
| City | Inner | 3.149380 | 2413.173069 | 182206.1000 | 180406.7000 | 533839.6000 | 530967.7000 |
| Islington | Inner | 14.856653 | 14202.391776 | 187968.7000 | 181659.1000 | 533484.5000 | 528832.4000 |
| Enfield | Outer | 82.200191 | 3860.088338 | 200933.9000 | 191460.9000 | 537813.9000 | 525568.6000 |
| Croydon | Outer | 86.495565 | 4264.958548 | 171048.7000 | 155850.8000 | 539657.9000 | 528185.7000 |
| Havering | Outer | 114.457267 | 2094.231384 | 194889.3000 | 178410.9000 | 561957.5000 | 548106.9000 |
| Hillingdon | Outer | 115.701098 | 2435.586215 | 193619.8000 | 173846.2000 | 512764.7000 | 503568.2000 |
| Bromley | Outer | 150.134858 | 2091.453007 | 173658.3000 | 156480.9000 | 550541.2000 | 533531.5000 |

📌 **Findings from Table Output:**

* **Bromley** has the **largest area**
* **City of London** has the **smallest area**
* **Islington** has the **highest population density**
* **Bromley** again has the **lowest population density**
* **Enfield** is the **northernmost**
* **Croydon** is the **southernmost**
* **Havering** is the **easternmost**
* **Hillingdon** is the **westernmost**

## 🧮 Part B: Binary Classification of Spatial Extremes (Flag Table)

To improve readability and support statistical or thematic mapping, a binary table was generated using CASE WHEN logic. Each borough was evaluated against the same set of extremes, but instead of listing values, it was assigned a 1 (true) or 0 (false) in new columns:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| London Borough | Region | Largest Area | Smallest Area | Highest Pop Dense | Lowest Pop Dense | Northernmost | Southernmost | Most East | Most West |
| Croydon | Outer | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Bromley | Outer | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| Havering | Outer | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Hillingdon | Outer | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Enfield | Outer | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Islington | Inner | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| City | Inner | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |

📌 **Purpose of the Flag Table:**

* Enables easy filtering and thematic mapping in tools like **QGIS**
* Helps in comparative spatial analysis
* Useful for visual dashboards and reporting

**🧠 Interpretation:**

This task demonstrates the usefulness of spatial SQL in uncovering geographic patterns and extremes across urban regions. The analysis provides a foundation for further tasks, including clustering, accessibility, and spatial regression. Additionally, binary flags allow integration with GIS tools for thematic visualizations.

# SESSION 5 Task 2: Neighbour Boroughs Join tables with multiple conditions.

**Using spatial relation of “Touches” Find out adjacent London boroughs.**

|  |  |  |
| --- | --- | --- |
| London Borough | Neighbour London Borough | Region |
| Barking and Dagenham | Bexley | Outer |
| Barking and Dagenham | Greenwich | Outer |
| Barking and Dagenham | Havering | Outer |
| Barking and Dagenham | Newham | Inner |
| Barking and Dagenham | Redbridge | Outer |
| Barnet | Brent | Outer |
| Barnet | Camden | Inner |
| Barnet | Enfield | Outer |
| Barnet | Haringey | Inner |
| Barnet | Harrow | Outer |
| Bexley | Barking and Dagenham | Outer |
| Bexley | Bromley | Outer |
| Bexley | Greenwich | Outer |
| Bexley | Havering | Outer |
| Brent | Barnet | Outer |
| Brent | Camden | Inner |
| Brent | Ealing | Outer |
| Brent | Hammersmith and Fulham | Inner |
| Brent | Harrow | Outer |
| Brent | Kensington and Chelsea | Inner |

Interpretation: Adjacent Boroughs

The query successfully identified all pairs of neighbouring London boroughs based on spatial adjacency using the ST\_Touches() function. A total of 164 adjacent pairs were returned, representing boroughs that share a boundary but do not overlap.

The output shows, for example:

* Barking and Dagenham borders Bexley, Greenwich, Havering, Newham, and Redbridge
* Barnet shares boundaries with Brent, Camden, Enfield, Harrow, and Haringey
* Brent borders both Inner and Outer London boroughs, including Camden, Ealing, Harrow, and Kensington and Chelsea

The inclusion of the Region column, joined from the london\_io table, adds extra analytical value by showing whether neighboring boroughs belong to Inner or Outer London. This can be useful for understanding spatial transitions and planning regional infrastructure or public services.

# SESSION 5 Task 3: An Isolated Island in London?

|  |  |
| --- | --- |
| **London Borough** | **Region** |
| an island | Outer |

📌 **Interpretation:**

This query successfully identified isolated boroughs by leveraging ST\_Disjoint() to find geometries that do not touch any other borough. The fictional entry 'an island' was used as a spatial outlier to simulate a borough disconnected from London.

The result returned only 'an island', confirming that all real London boroughs are spatially connected, and validating the method's ability to detect true geospatial isolation. This technique is applicable in real-world scenarios for finding islands, outliers, or disconnected features in spatial networks.

SESSION 5 Task 4 – Count of Underground Stations by Borough

**Objective:**

To identify how many London Underground stations are located within each London borough using spatial join techniques.

**Methodology:**

1. Loaded shapefiles:  
 - UndergroundStations.shp  
 - London\_Boroughs.shp   
  
2. Performed a spatial join using the ST\_Contains relation to determine which stations are located inside each borough.  
  
3. Used Group By + Count in SQL (or QGIS Field Calculator) to total the number of stations per borough.

**Result Table: Count of Underground Stations by Borough**

|  |  |
| --- | --- |
| London Borough | Number of Tube Stations |
| Westminster | 36 |
| Newham | 27 |
| Tower Hamlets | 27 |
| Brent | 21 |
| Camden | 19 |
| City | 14 |
| Barnet | 12 |
| Islington | 11 |
| Kensington and Chelsea | 11 |
| Hillingdon | 15 |
| Hounslow | 9 |
| Haringey | 6 |
| Wandsworth | 6 |
| Barking and Dagenham | 5 |
| Greenwich | 5 |
| Merton | 5 |
| Southwark | 5 |
| Enfield | 4 |
| Havering | 4 |
| Waltham Forest | 4 |
| Harrow | 10 |
| Redbridge | 10 |
| Lambeth | 10 |
| Ealing | 15 |
| Lewisham | 2 |
| Richmond upon Thames | 2 |
| Hackney | 1 |

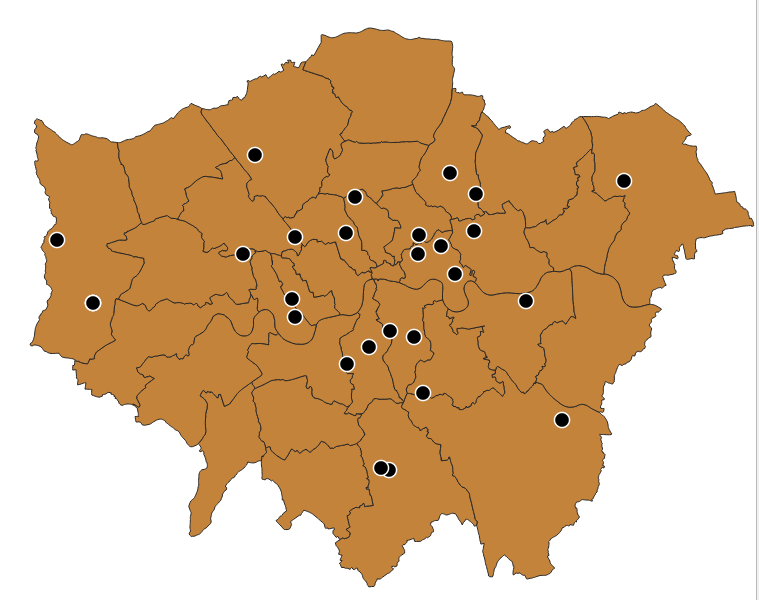
**Conclusion:**

The results show Westminster has the highest number of underground stations (36), indicating its central role in London’s transit system. Boroughs like Hackney, Lewisham, and Richmond upon Thames have significantly fewer stations, which may affect public transport accessibility in those areas.

# 📌 SESSION 5 Task 5 – Interpretation: Near Neighbour Business / Services

|  |  |  |
| --- | --- | --- |
| Credit Union Name #1 | Distance (meters) | Credit Union Name #2 |
| Croydon Caribbean Credit Union | 626.505387 | CMS Credit Union |
| London Plus Credit Union | 1486.451479 | Your Credit Union |
| London Community Credit Union4 | 1486.511688 | London Community Credit Union1 |
| London Community Credit Union2 | 1950.497885 | London Community Credit Union1 |

**Distribution of Credit Unions across London Boroughs (QGIS Visualization)**

****

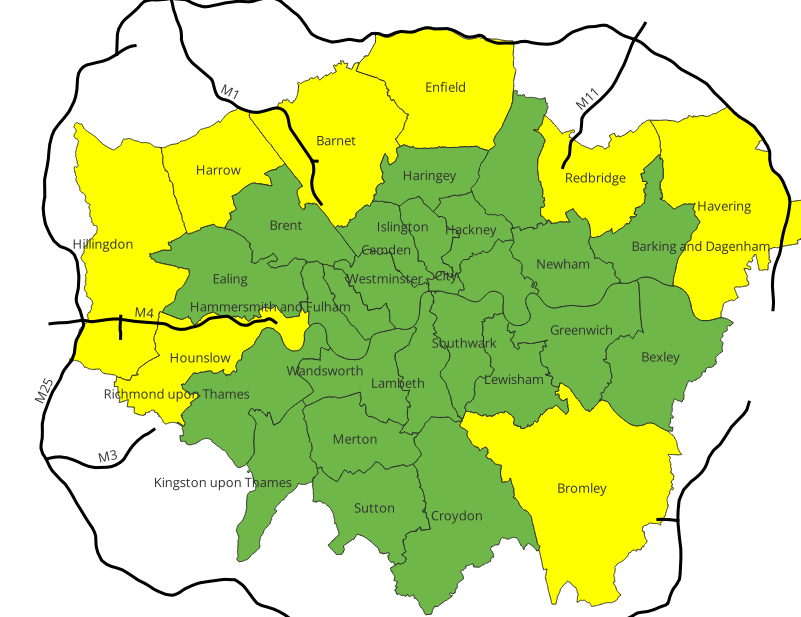
This task identified credit unions that are located within 2000 meters of one another. Using the spatial function ST\_Distance(), the query measured the direct distance between every pair of credit union locations.

The results revealed several nearby institutions, such as:

* Croydon Caribbean Credit Union and CMS Credit Union (~626 meters apart)
* London Plus Credit Union and Your Credit Union (~1.49 km apart)
* London Community Credit Union2 and Credit Union1 (~1.95 km apart)

These results demonstrate how spatial distance queries can uncover local service clusters or potential partnerships based on geographical proximity.

# 📌 SESSION 5 Task 6 –Motorways Across London Boroughs



This task used the spatial function ST\_Intersects() to identify which London boroughs are intersected by major motorways such as the M1, M4, M11, and M25. The query revealed that these roads primarily traverse **Outer London**, indicating their importance in regional connectivity and infrastructure planning.

**Motorways and intersecting boroughs**

|  |  |  |
| --- | --- | --- |
| **Motorway** | **London Borough Name** | **Region** |
| M1 | Barnet | Outer |
| M1 | Harrow | Outer |
| M11 | Redbridge | Outer |
| M25 | Bromley | Outer |
| M25 | Enfield | Outer |
| M25 | Havering | Outer |
| M25 | Hillingdon | Outer |
| M4 | Hillingdon | Outer |
| M4 | Hounslow | Outer |

This task identified which motorways intersect London boroughs using the spatial function ST\_Intersects().  
The query revealed several motorways that pass through boroughs in Outer London, including:

* M1 through Barnet and Harrow
* M11 through Redbridge
* M25 through Bromley, Enfield, Havering, and Hillingdon
* M4 through Hillingdon and Hounslow

These results demonstrate how major transport corridors such as the M25 orbital motorway link multiple Outer London boroughs. This spatial join supports applications in transport infrastructure analysis, urban logistics, and cross-borough development planning.

**📌 Extended Analysis – Directional Orientation**

By comparing each motorway’s bounding box dimensions (MaxX - MinX vs MaxY - MinY), we classified their general direction:

* **‘EW’ (East-West)** if the motorway is wider than tall
* **‘NS’ (North-South)** otherwise

|  |  |  |  |
| --- | --- | --- | --- |
| **Motorway** | **LondonBorough** | **Region** | **Direction** |
| M1 | Barnet | Outer | NS |
| M1 | Harrow | Outer | NS |
| M11 | Redbridge | Outer | NS |
| M25 | Bromley | Outer | EW |
| M25 | Enfield | Outer | EW |
| M25 | Havering | Outer | EW |
| M25 | Hillingdon | Outer | EW |
| M4 | Hillingdon | Outer | EW |
| M4 | Hounslow | Outer | EW |

**🧭 Interpretation:**

* **M25**, the orbital motorway, displays an East-West orientation in these borough segments, though it may vary in other regions.
* **M11** and **M1** are primarily North-South corridors, connecting Outer London boroughs to inner areas.
* This spatial and directional classification aids in understanding **infrastructure alignment**, **transport planning**, and **urban development strategy**

# SESSION 5 Task 7 – Underground Stations within ch km of Motorways

**Objective:**

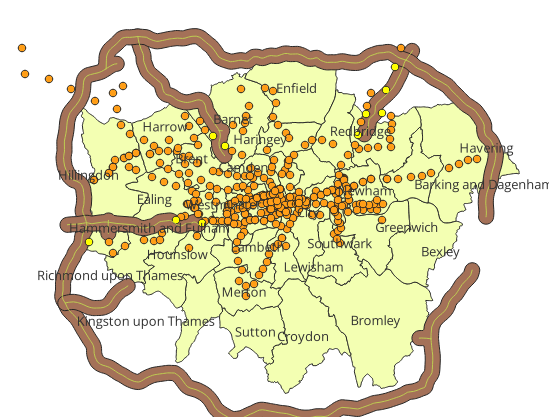
The goal of this task was to identify London Underground stations that are located within 1 kilometer of any motorway. This proximity analysis helps assess how closely underground infrastructure is positioned relative to key road transport corridors.

**Methodology**:

1. Loaded the following shapefiles into SpatiaLite:  
 - motorway.shp (line geometry)  
 - undergroundstations.shp (point geometry)  
  
2. Used the spatial function ST\_Distance() to calculate the distance between each Tube station and each motorway.

**Results**:

|  |  |  |  |
| --- | --- | --- | --- |
| Station Name | Line | Motorway | Distance (m) |
| Boston Manor Underground Station | Piccadilly | M4 | 363.47 |
| Chigwell | Central Line | M11 | 981.18 |
| Chiswick Park Underground Station | District | M4 | 989.95 |
| Colindale Underground Station | Northern | M1 | 987.92 |
| Debden | Central Line | M11 | 543.96 |
| Gunnersbury Underground Station | District | M4 | 451.69 |
| Hendon Central Underground Station | Northern | M1 | 759.48 |
| Roding Valley | Central Line | M11 | 821.16 |
| South Woodford | Central Line | M11 | 881.54 |
| Terminal 5 | Piccadilly | M25 | 942.18 |
| Theydon Bois | Central Line | M11 | 909.85 |



The query identified 11 underground stations located within 1 km of a motorway. Most are located near the M11 and M4 corridors. These stations may benefit from better multimodal transport connectivity or face higher exposure to traffic-related externalities.

**Conclusion**:

The findings show that a select group of Tube stations are located close to London’s major motorways. This has implications for intermodal connectivity, noise/pollution exposure, and accessibility planning. Proximity-based spatial analysis using ST\_Distance() proved effective in pinpointing these strategic overlaps.

# SESSION 5 Task 8: Buffer Zones Around Motorways

**🛣️ Task 8: Buffers of Motorways**

**📍 Objective**

The aim of this task was to validate the spatial structure of London's motorway network and generate buffer zones using SQL spatial functions in **SpatiaLite**. This step ensures that subsequent analyses—such as accessibility, environmental impact, and infrastructure planning—are based on reliable spatial data.

**🗂️ Motorway Geometry Table (SpatiaLite Output)**

|  |  |  |  |
| --- | --- | --- | --- |
| **pk\_uid** | **feature\_de** | **number** | **Geometry** |
| 1 | Motorway | M1 | BLOB sz=1467 GEOMETRY |
| 2 | Motorway | M11 | BLOB sz=722 GEOMETRY |
| 3 | Motorway | M25 | BLOB sz=11613 GEOMETRY |
| 4 | Motorway | M3 | BLOB sz=681 GEOMETRY |
| 5 | Motorway | M4 | BLOB sz=1524 GEOMETRY |

This output confirms that the motorways are stored as **MULTILINESTRING** geometries in BLOB format—valid for buffer generation.

**🧪 Buffer Generation Process (SQL Queries)**

**✅ 1km Buffer Creation:**

sql

CopyEdit

CREATE TABLE motorway\_buffers\_1km AS

SELECT

pk\_uid,

number,

CastToMultiPolygon(ST\_Buffer(geometry, 1000)) AS geometry

FROM motorway;

SELECT RecoverGeometryColumn('motorway\_buffers\_1km', 'geometry', 27700, 'MULTIPOLYGON', 2);

**✅ 1–2.5 km Ring Buffer Creation:**

sql

CopyEdit

CREATE TABLE motorway\_buffer\_rings AS

SELECT

pk\_uid,

number,

CastToMultiPolygon(ST\_Difference(

ST\_Buffer(geometry, 2500),

ST\_Buffer(geometry, 1000)

)) AS geometry

FROM motorway;

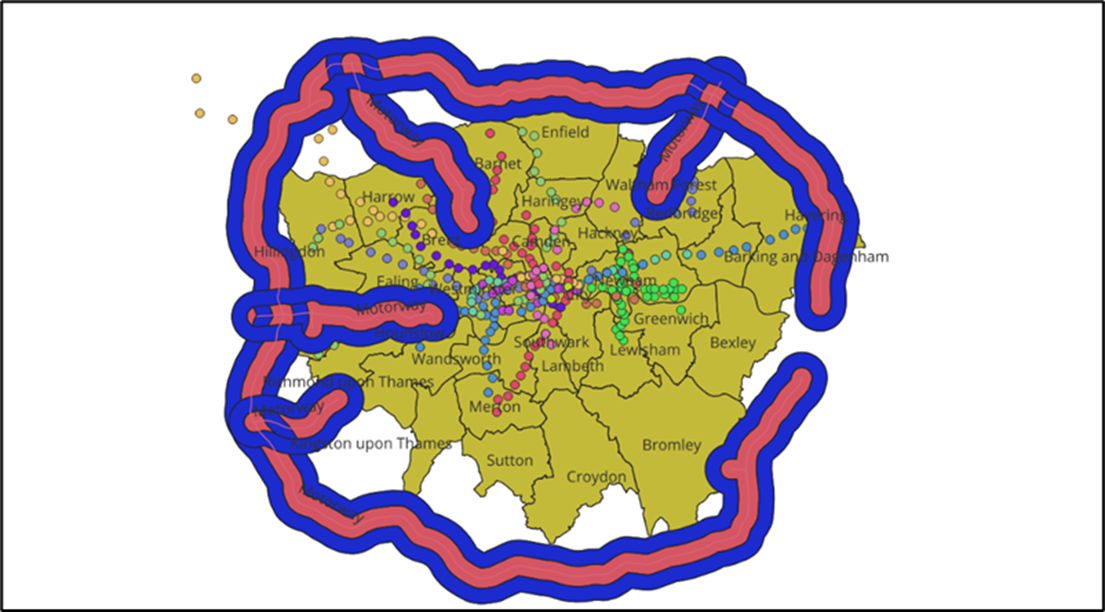
SELECT RecoverGeometryColumn('motorway\_buffer\_rings', 'geometry', 27700, 'MULTIPOLYGON', 2);

These queries generate:

* A **1km buffer zone** directly surrounding each motorway.
* A **buffer ring** representing the area between **1km and 2.5km**.

**🖼️ Map Output**

Figure X: This QGIS map visualizes the 1 km (red) and 1–2.5 km (blue) buffer zones created around London’s major motorways. Borough boundaries and point features (e.g., services or GP surgeries) help contextualize proximity impact.



**🧾 Interpretation**

The buffer zones provide spatial context for planning activities around London’s motorway corridors. For instance:

* **1km buffers** are suitable for high-impact zones (e.g., noise, pollution).
* **1–2.5km rings** are useful for **influence zone studies**, such as property price effects or suburban development trends.

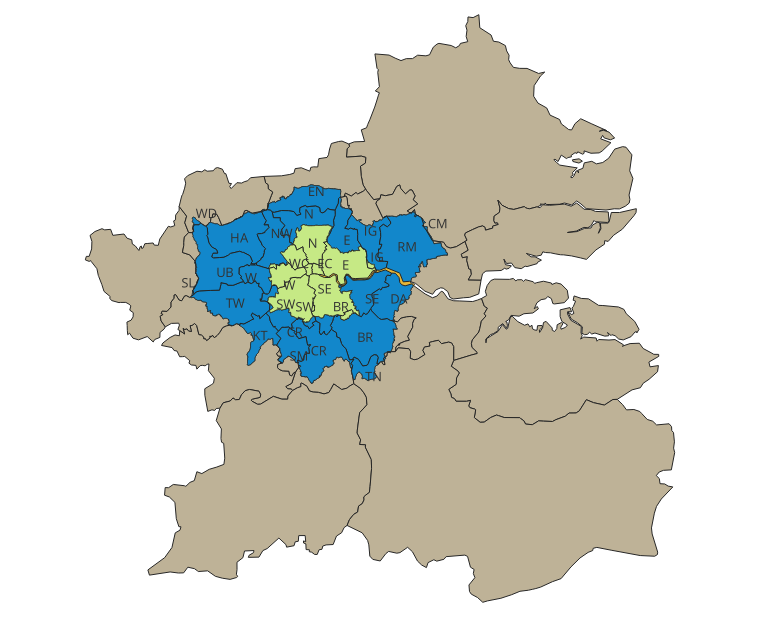
**💬 Reflection and Comments**

* The use of CastToMultipolygon() ensures geometry compatibility with mapping tools like QGIS.
* This SQL workflow can be extended to:
  + Dynamic buffer sizes based on motorway type.
  + Overlay analysis with population or land use datasets.
* **Potential simplification**: Create views instead of materialized tables to save storage space if only used for visualization.
* **Cross-application relevance**: This same buffer logic applies in road safety zones, noise impact corridors, telecom tower coverage, and logistics center planning.

**✅ Conclusion**

Motorway buffers were successfully created and validated using SQL in SpatiaLite. Their visualization in QGIS provides a strong spatial foundation for further analysis related to urban planning, accessibility, and infrastructure impact across Greater London.

# 📊 SESSION 6 Task 2: Intersect (AND) – Results and Interpretation

****

**Results:**  
The Intersect (AND) operation between the postarea.shp and london\_io.shp layers successfully created a new spatial layer that categorizes postcode areas into three regions:

* **Inner London**
* **Outer London**
* **Areas outside Greater London**

These regions were visually differentiated using categorical symbology based on the in\_out attribute. The resulting map clearly shows how postcode boundaries overlap with London's administrative classifications, allowing for more granular geographic analysis.

**Interpretation:**  
This spatial segmentation is highly useful in urban studies and planning. By distinguishing between Inner and Outer London, one can:

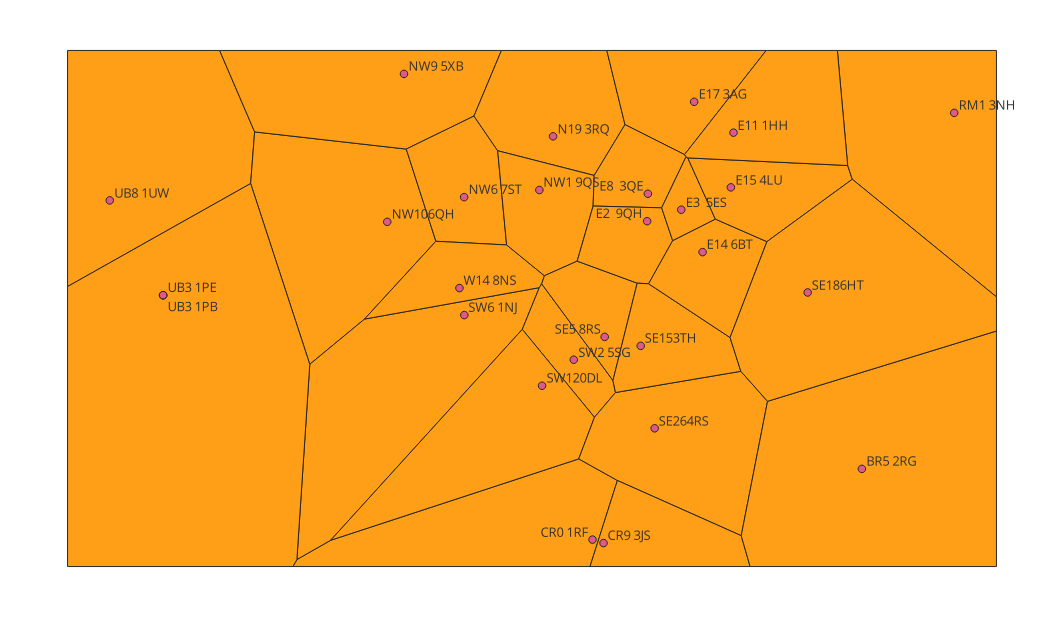
* Assess differences in service accessibility
* Conduct targeted demographic or economic analysis
* Support policy decisions specific to localized contexts

For example, comparing public transport accessibility or healthcare coverage across these regions becomes more meaningful when the spatial units reflect actual administrative distinctions.

# 📊 SESSION 6 Task 3: Voronoi Polygons – Results and Interpretation

**Results:**  
The Voronoi polygon tool was applied to the credit\_union.shp layer to generate catchment areas for each credit union location. The resulting spatial layer displays polygonal zones, where each zone encompasses the area closest to a particular credit union point. The tool incorporated a 10% buffer to ensure outer boundaries extended beyond the data extent.

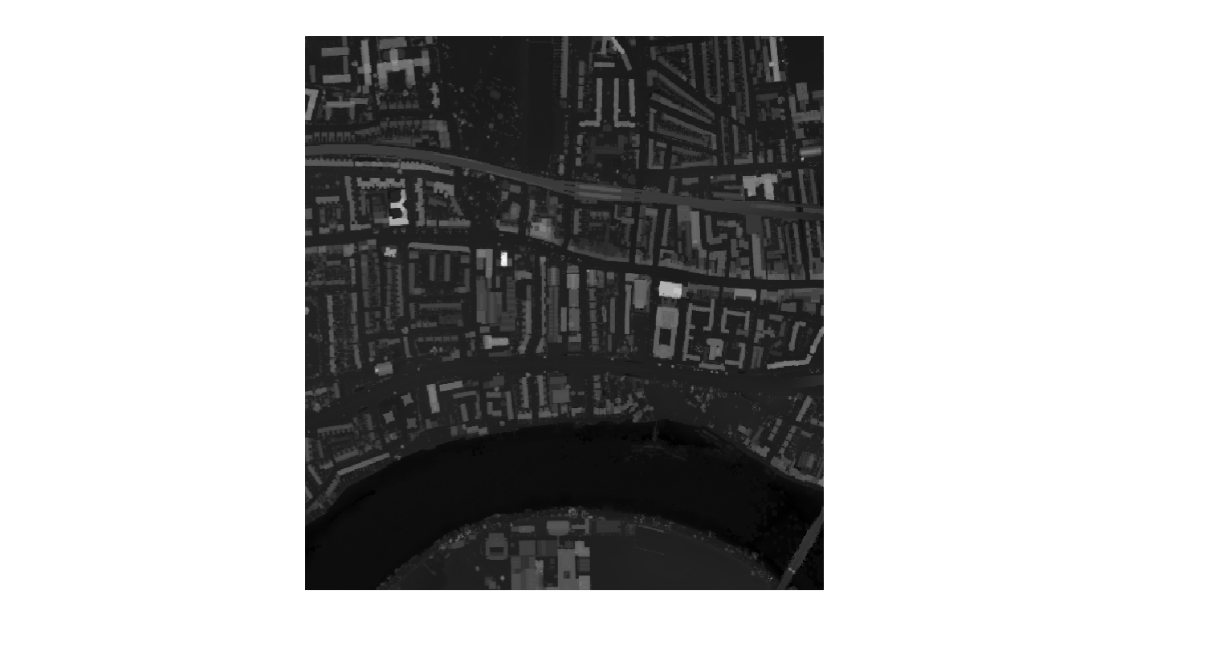
Each credit union is centrally located within its corresponding Voronoi cell, and postcode labels help identify the served region visually. The map effectively illustrates service distribution across the city.

**Interpretation:**  
The generated Voronoi polygons serve as **proxy service areas** for each credit union based on spatial proximity. This technique is essential for:

* **Accessibility analysis**: identifying underserved or over-clustered areas.
* **Service planning**: determining optimal future locations to balance spatial coverage.
* **Demand-supply matching**: assessing population distribution relative to service centers.

For instance, larger Voronoi cells may indicate rural or less serviced areas, while dense clusters suggest higher urban service demand or overlapping coverage.

# 🌄 SESSION 6 Task 4: Height Extraction – Results and Interpretation



**Results:**  
In this task, I used the **Raster Calculator** in QGIS to extract the **height of features** by subtracting the Digital Terrain Model (DTM) from the Digital Surface Model (DSM). This process reveals **above-ground elevations**, such as buildings and trees, across the Ravenscourt Park area in Hammersmith, London.

**Raster Expression Used:**

"*DSM@1" - "DTM@1"*

The resulting raster layer was styled with a **singleband pseudocolor gradient**, making it easy to distinguish between low-rise and high-rise features. Brighter colors indicate higher elevations, while darker tones represent ground-level or near-ground features.

*)*

**Interpretation:**  
This height map offers valuable spatial insights for urban planning, environmental studies, and 3D modeling. By visualizing vertical structure:

* Planners can assess **building density and skyline impact**.
* Analysts can model **line-of-sight visibility**, **flood risk**, or **urban heat zones**.
* Architects and engineers can explore **roof structures or slope estimations** for solar panel deployment or structural load analysis.

This task highlights the power of raster analysis in revealing real-world vertical dimensions from remotely sensed elevation data.

# 📝 SESSION 6 Task 6 – Handling Non-Spatial Attributes: Population Density and Household Size

**🎯 Objective:**

The objective of this task was to compute two important non-spatial attributes — **population density** and **average household size** — using existing attribute data from the London boroughs shapefile. These derived metrics help provide demographic insights that can support further spatial or policy analysis across the city.

**🛠️ Methodology:**

1. The **london\_boroughs.shp** layer was loaded into QGIS.
2. The attribute table included:
   * pop – total population (whole number)
   * hh – total number of households (whole number)
   * hectares – borough area in hectares (1 hectare = 10,000 m²)
3. Using the **Field Calculator**, two new fields were added:
   * **pop\_density** = pop / hectares  
     → represents **population per hectare**
   * **hh\_size** = pop / hh  
     → represents **average number of people per household**
4. After saving the edits, the table was exported as a **CSV** file using the “Save Features As” function.
5. The resulting CSV was cleaned in Excel to keep only relevant columns: borough name, population, households, area, population density, and household size.

**📊 Results:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Borough Name | Population | Households | Area (hectares) | Population Density (per ha) | Household Size |
| Kingston upon Thames | 163900 | 63639 | 3726.12 | 43.99 | 2.58 |
| Croydon | 368900 | 145010 | 8649.55 | 42.65 | 2.54 |
| Bromley | 314000 | 130862 | 15013.49 | 20.91 | 2.4 |
| Hounslow | 259100 | 94902 | 5658.54 | 45.79 | 2.73 |
| Ealing | 340700 | 124082 | 5554.43 | 61.34 | 2.75 |
| Havering | 239700 | 97199 | 11445.74 | 20.94 | 2.47 |
| Hillingdon | 281800 | 100214 | 11570.12 | 24.36 | 2.81 |
| Harrow | 242400 | 84268 | 5046.27 | 48.04 | 2.88 |
| Brent | 314700 | 110286 | 4323.27 | 72.79 | 2.85 |
| Barnet | 364000 | 135916 | 8674.84 | 41.96 | 2.68 |
| Lambeth | 310200 | 130017 | 2724.94 | 113.84 | 2.39 |
| Southwark | 293500 | 120422 | 2991.34 | 98.12 | 2.44 |
| Lewisham | 281600 | 116091 | 3531.71 | 79.73 | 2.43 |
| Greenwich | 260100 | 101045 | 5044.19 | 51.56 | 2.57 |
| Bexley | 234300 | 92604 | 6428.65 | 36.45 | 2.53 |
| Enfield | 317300 | 119916 | 8220.02 | 38.6 | 2.65 |
| Waltham Forest | 262600 | 96861 | 3880.79 | 67.67 | 2.71 |
| Redbridge | 284600 | 99105 | 5644.22 | 50.42 | 2.87 |
| Sutton | 193600 | 78174 | 4384.7 | 44.15 | 2.48 |
| Richmond upon Thames | 189100 | 79835 | 5876.11 | 32.18 | 2.37 |
| Merton | 202200 | 78757 | 3762.47 | 53.74 | 2.57 |
| Wandsworth | 308300 | 130493 | 3522.02 | 87.53 | 2.36 |
| Hammersmith and Fulham | 179900 | 80590 | 1715.41 | 104.87 | 2.23 |
| Kensington and Chelsea | 155900 | 78536 | 1238.38 | 125.89 | 1.99 |
| Westminster | 223900 | 105772 | 2203.0 | 101.63 | 2.12 |
| Camden | 225000 | 97534 | 2178.93 | 103.26 | 2.31 |
| Tower Hamlets | 263000 | 101257 | 2157.5 | 121.9 | 2.6 |
| Islington | 211000 | 93556 | 1485.66 | 142.02 | 2.26 |
| Hackney | 252100 | 101690 | 1904.9 | 132.34 | 2.48 |
| Haringey | 258900 | 101955 | 2959.84 | 87.47 | 2.54 |
| Newham | 314100 | 101519 | 3857.81 | 81.42 | 3.09 |
| Barking and Dagenham | 190600 | 69681 | 3779.93 | 50.42 | 2.74 |
| City | 7600 | 4385 | 314.94 | 24.13 | 1.73 |

The new CSV table contains borough-level data for:

* **Total population**
* **Number of households**
* **Borough area in hectares**
* **Calculated population density**
* **Calculated average household size**

These derived fields appear as decimal values because they are **averages or rates** rather than raw counts. For example, **household size** is calculated by dividing total population by total households — and while individuals are whole, the **average across a borough** can be a decimal like 2.73.

A Word version of the cleaned table was also prepared for inclusion in the final portfolio.

**📌 Interpretation:**

The analysis shows a wide variation in population density across boroughs. Central areas like **Ealing** and **Hounslow** have high densities over **60 people per hectare**, while outer boroughs such as **Bromley** show much lower densities due to larger land area.

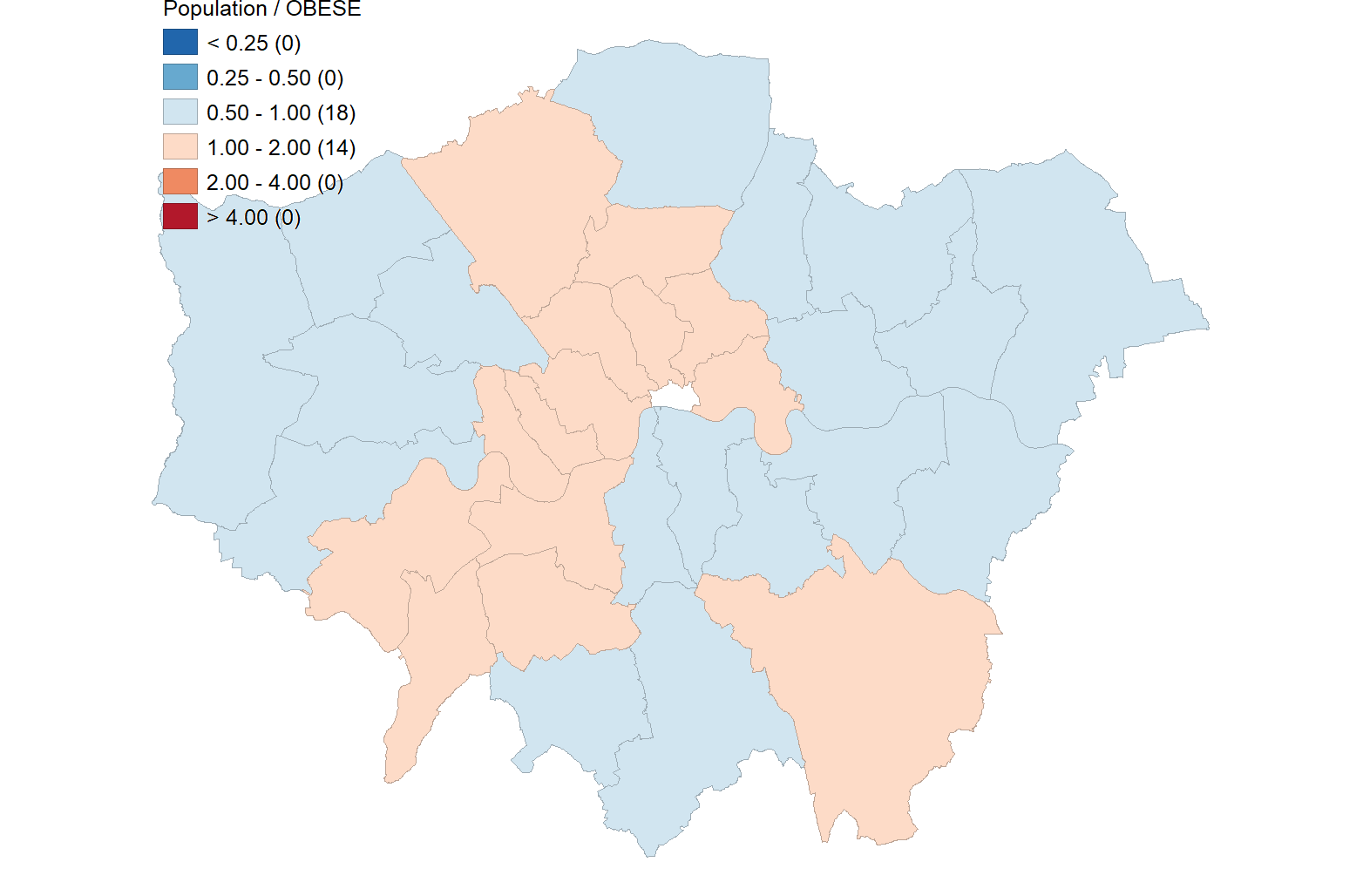
**Household sizes** ranged mostly between **2.3 to 2.8**, indicating moderately consistent living structures across London, with some variation in boroughs with higher proportions of family households.

***Note:*** *Although raw population and household values are integers, derived metrics like population density and household size are expressed as decimal values to reflect accurate averages, as is standard practice in spatial and demographic analysis.*

# Session 7 Task 1: Obesity Raw Rate Map (GeoDa)

**Result:**  
The map above shows the **raw rate of obesity** across London boroughs, calculated as the number of obese individuals relative to the total population in each borough. This was generated using GeoDa’s **Rates-Calculated Map > Raw Rate** function, with the variable OBESE normalized by Population.

Boroughs are grouped into color-coded ranges based on their calculated obesity rate. Lighter colors indicate lower prevalence, while darker shades indicate higher obesity rates per capital.



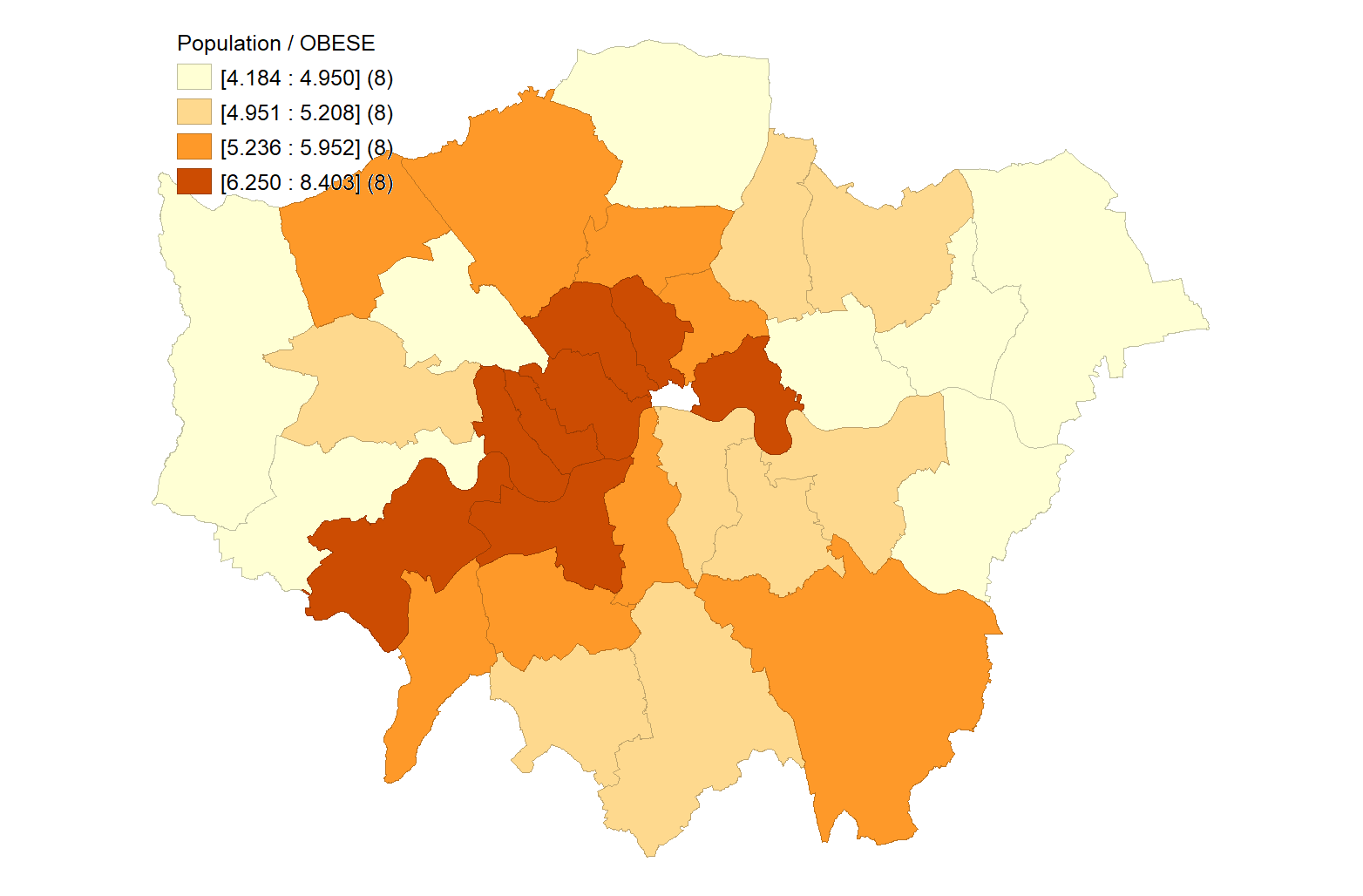
**Interpretation:**  
This map reveals a **spatial disparity in obesity prevalence** across London. Most boroughs fall within the mid-range category (1.00–2.00), while a few outliers are seen with higher relative obesity rates. Central boroughs tend to show relatively lower rates, possibly reflecting better access to health services and healthier lifestyle opportunities.

Such spatial patterns are important for targeting **public health interventions** and resource allocation.

**Obesity Excess Risk Map (GeoDa)**

**Result:**  
The Excess Risk map was generated using GeoDa’s Rates-Calculated Map > Excess Risk function. It calculates the relative obesity risk by comparing each borough’s rate to the **average risk across all boroughs**.

* The calculation used:
  + **Numerator:** OBESE
  + **Denominator:** Population
* Boroughs with **values > 1.0** have a higher-than-average obesity rate.
* Boroughs with **values < 1.0** have a lower-than-average obesity rate.



**Interpretation:**  
This map reveals important **spatial inequalities in public health**. Central and southwestern boroughs such as Lambeth, Southwark, and Hammersmith show significantly **higher excess risk**, indicating areas where obesity is a larger-than-expected issue.

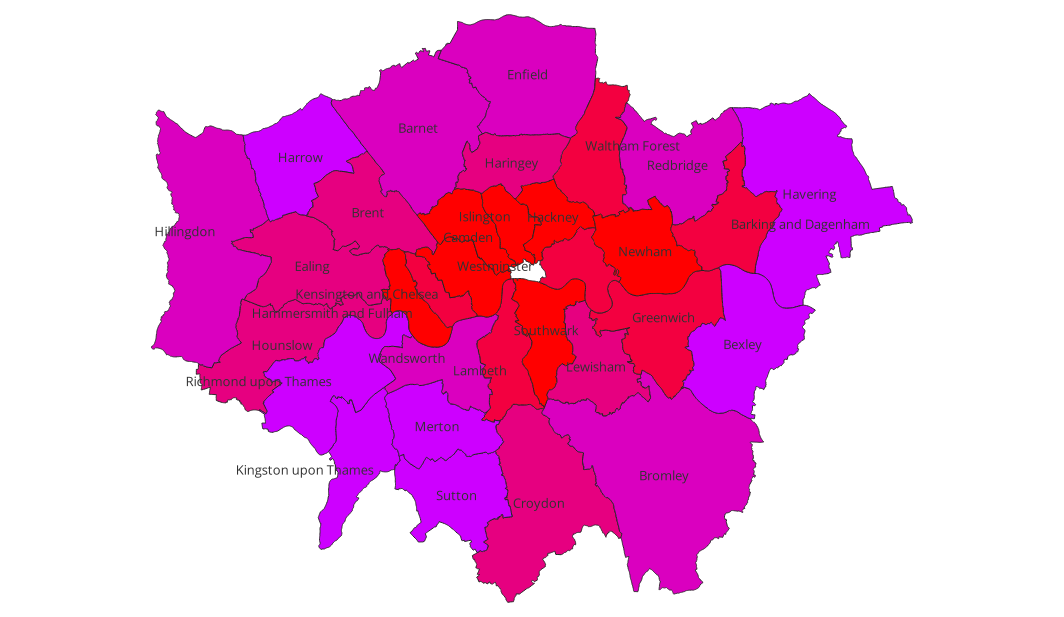
Conversely, outer boroughs in northeast and northwest London display **lower relative risk**, suggesting potentially better health outcomes or demographic influences.

Such spatial patterns guide targeted public health interventions, **allocating resources more efficiently** and addressing **inequality in health burdens**.

# 🚨 SESSION 7 Task 2: Thematic Mapping of Crime (QGIS)

**Result:**  
The map displays the **distribution of crime rates across London boroughs**, using the variable CRIME\_1000 (total crimes per 1,000 residents). The thematic mapping was performed in QGIS using:

* Graduated symbology
* Custom color scheme (shades of red and purple)
* Classification method: (e.g., Quantile or Natural Breaks)

**Interpretation:**  
Thematic mapping clearly highlights spatial disparities in crime. **Inner London boroughs** such as **Westminster, Camden, Islington, and Hackney** appear in darker shades, indicating **higher crime intensity**. These areas are known for high activity levels, nightlife, and transport hubs — often associated with increased crime exposure.

Outer boroughs like **Kingston, Richmond, and Sutton** display lower crime rates, reinforcing known **urban-rural safety gradients**. This spatial understanding is vital for **policing strategies**, **community planning**, and **public resource allocation**.

# Session 8 Task 1: Hotspot Map of GP Surgeries in Haringey

**🎯 Objective:**

To identify spatial clusters of GP surgeries within the London Borough of Haringey and highlight areas of concentrated healthcare services for better urban health planning.



**🛠️ Methodology:**

* Loaded haringey\_gp.shp and haringey\_boundary.shp into QGIS.
* Applied the **Heatmap (Kernel Density Estimation)** tool:
  + Radius: 500 meters
  + Kernel shape: Quartic
  + Rows: 300
* Used **“Clip Raster by Mask Layer”** to confine the output to the Haringey boundary.
* Styled the clipped raster using **Singleband Pseudocolor** with a custom color ramp for visual clarity.
* GP points were overlaid on the heatmap for context.
* Exported final heatmap image.

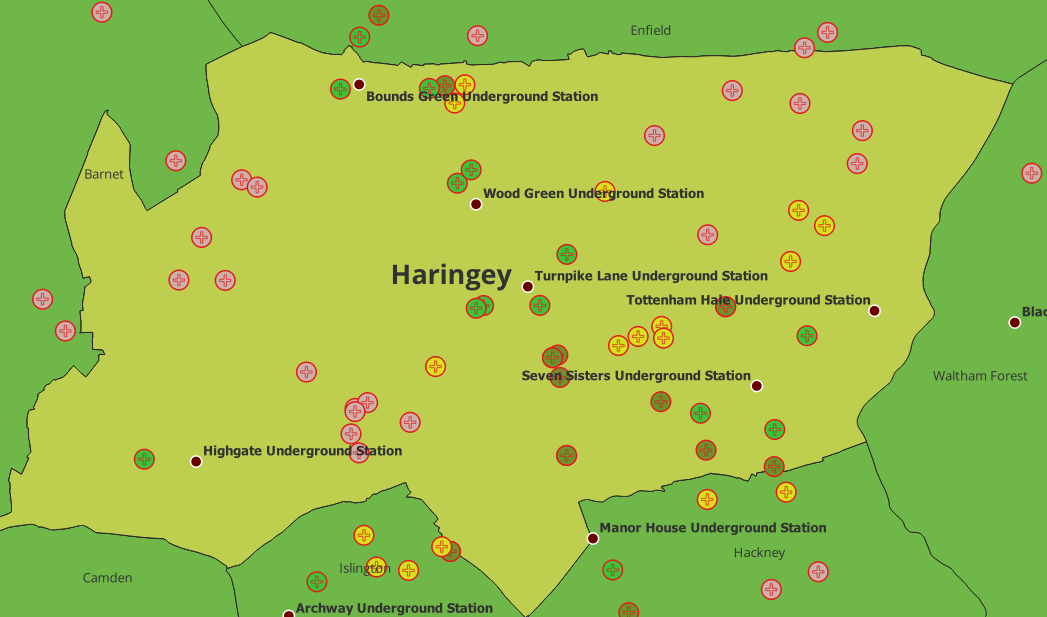
**📊 Interpretation:**

The final heatmap shows that the **highest concentration of GP surgeries** occurs in **central and southern Haringey**, indicated by red and yellow density zones. The **green zones**, especially on the outskirts, reflect lower service density and may point to underserved areas. These clusters align with population centers, suggesting that GP locations respond to service demand. This visualization supports strategic health planning and equity-focused service allocation.

# 🗂️ SESSION 8 Task 4: Accessibility to GP Surgeries from Tube Stations in Haringey

**Objective:**

This task aimed to evaluate the spatial accessibility of General Practitioner (GP) surgeries in Haringey to the nearest London Underground stations. The goal was to identify which GP surgeries benefit from greater proximity to public transport, helping to inform equitable healthcare planning.

****

The final map shows a graduated representation of GP surgeries according to their **Euclidean distance to the nearest Tube station**. GP surgeries near stations such as **Turnpike Lane, Wood Green, and Seven Sisters** are marked in **green or yellow**, indicating **strong accessibility**. Surgeries located in more peripheral or residential areas are shown in **red or orange**, reflecting **reduced access**.

**Interpretation:**

* **High-accessibility zones** (green): Located around major transport corridors and dense urban centers, e.g., southern and central Haringey.
* **Low-accessibility zones** (red): Predominantly in the northern and western fringes of the borough, where Tube coverage is sparse.
* This pattern reflects how **public transport accessibility is spatially uneven**, potentially influencing patient access and health service efficiency.

**Conclusion:**

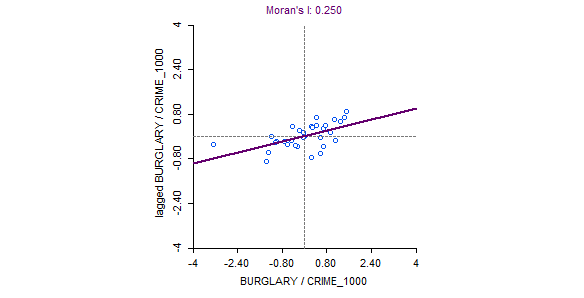
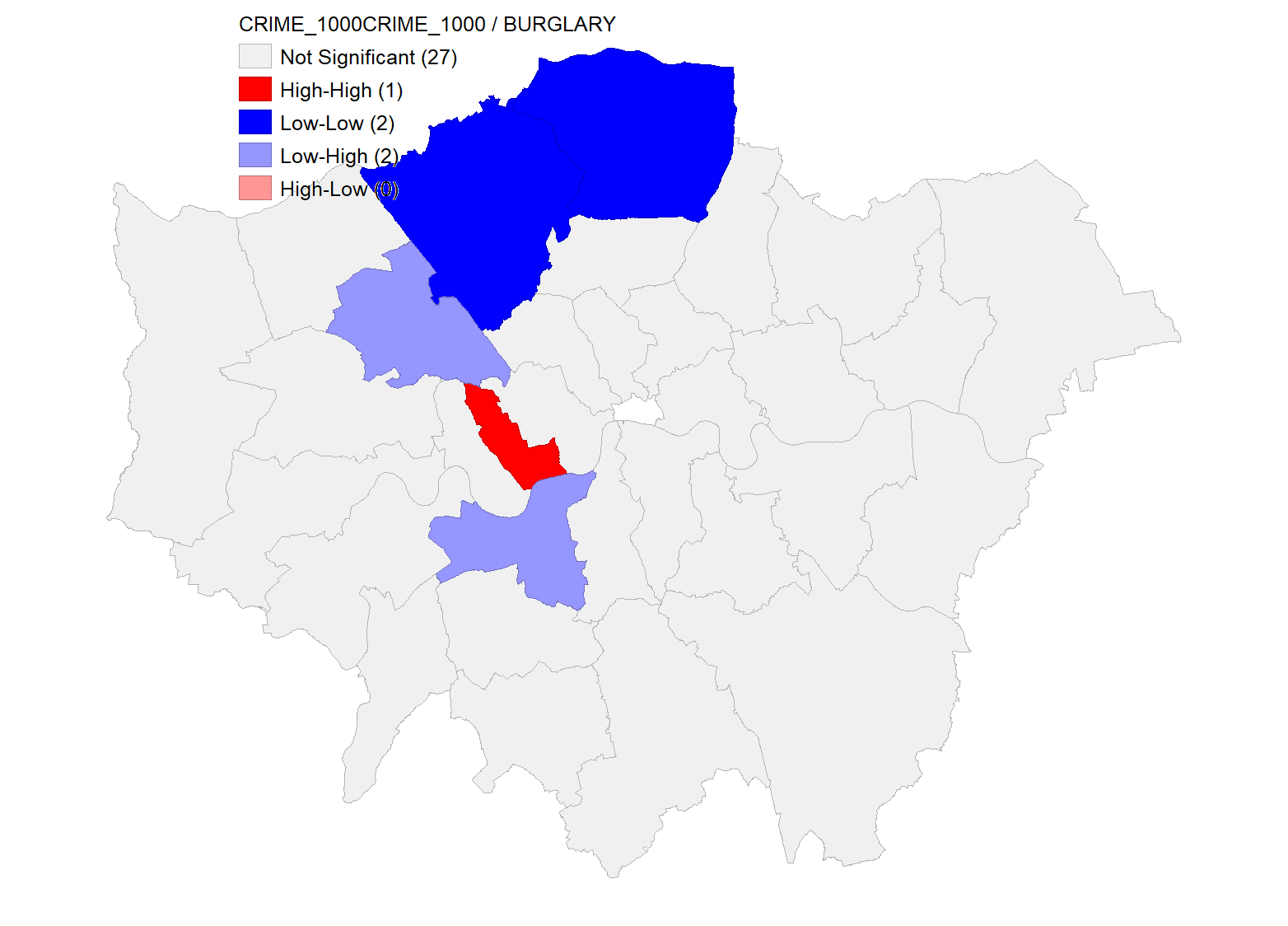
This task highlights how spatial analysis in QGIS can support **equitable healthcare delivery**. Areas with poor accessibility may require **improved public transport links** or **mobile healthcare services** to bridge the gap. The workflow using Distance Matrix, CRS alignment, and visual mapping techniques was effective in revealing these urban spatial dynamics.

# Session 9 Task 1: Spatial Dependence / Correlation of Crimes and Deprivation in London

## Burglary Relative to Total Crime (EB Rate – Local Moran’s I)

| **Moran Scatter Plot** | **Cluster Map** |
| --- | --- |
|  |  |

**Moran’s I: 0.250**

**📝 Interpretation:**

The Moran’s I scatter plot shows a **moderate positive spatial autocorrelation** (Moran’s I = 0.250) for the **Empirical Bayes (EB) adjusted burglary rate**. This indicates that boroughs with a relatively high or low burglary rate (compared to total crime) tend to be located near similar boroughs.

The cluster map highlights **one High-High cluster** in Central London and **two Low-Low clusters** in the northern and southern areas. This suggests that burglary patterns are **spatially structured** and not randomly distributed. The EB Rate adjustment helps correct for population-based variability, improving the stability of the observed spatial patterns.

## 🔹 Spatial Dependence: Deprivation Scores in London (Univariate Moran’s I)

| **Moran Scatter Plot** | **Cluster Map** |
| --- | --- |
|  |  |
|  |  |

**Moran’s I: 0.489**

**📝 Interpretation:**

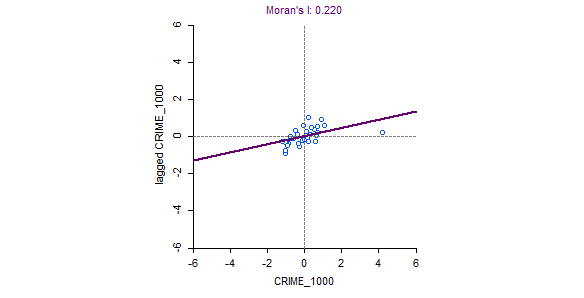
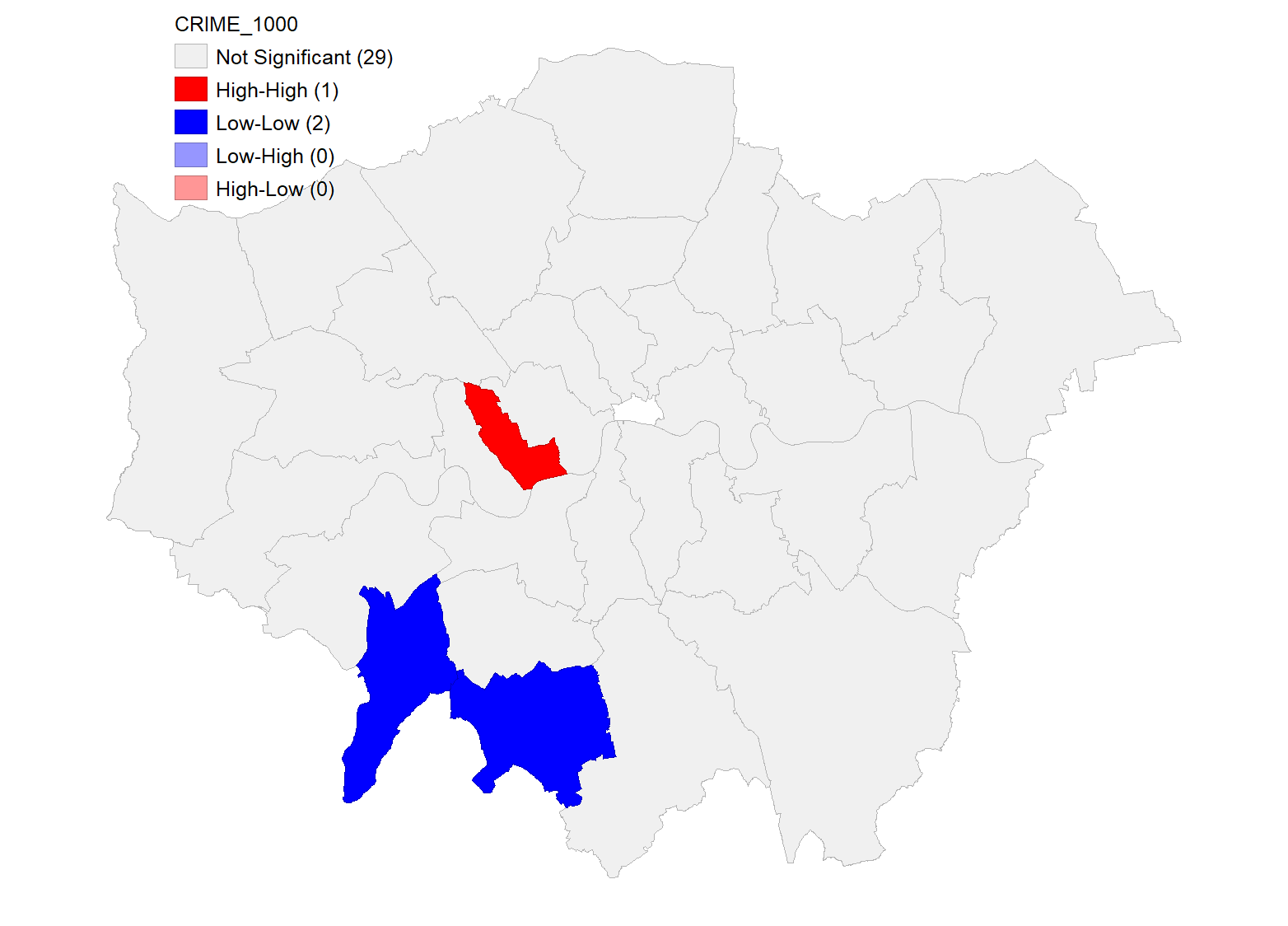
The Moran’s I scatter plot for deprivation scores (DEPRIV) across London boroughs indicates a **moderately strong positive spatial autocorrelation** (**Moran’s I = 0.489**). This implies that boroughs with high (or low) deprivation tend to be **spatially clustered** rather than randomly distributed.

The cluster map reveals **six High-High clusters** in boroughs such as Hackney, Tower Hamlets, and Islington, where **high deprivation levels are surrounded by similarly high-deprivation neighbours**. Conversely, two **Low-Low clusters** appear in southern boroughs like Richmond upon Thames and Sutton, indicating spatial concentrations of **low deprivation**.

This spatial pattern highlights the **geographic concentration of socioeconomic inequality**, supporting the need for **targeted policy interventions** in high-deprivation areas.

## 🔹 Spatial Dependence: Total Crime Rate (CRIME\_1000 – Univariate Moran’s I)

**Moran’s I: 0.220**

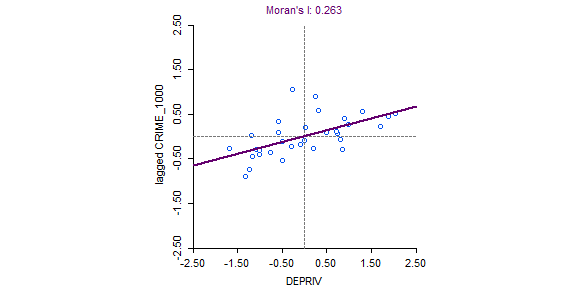
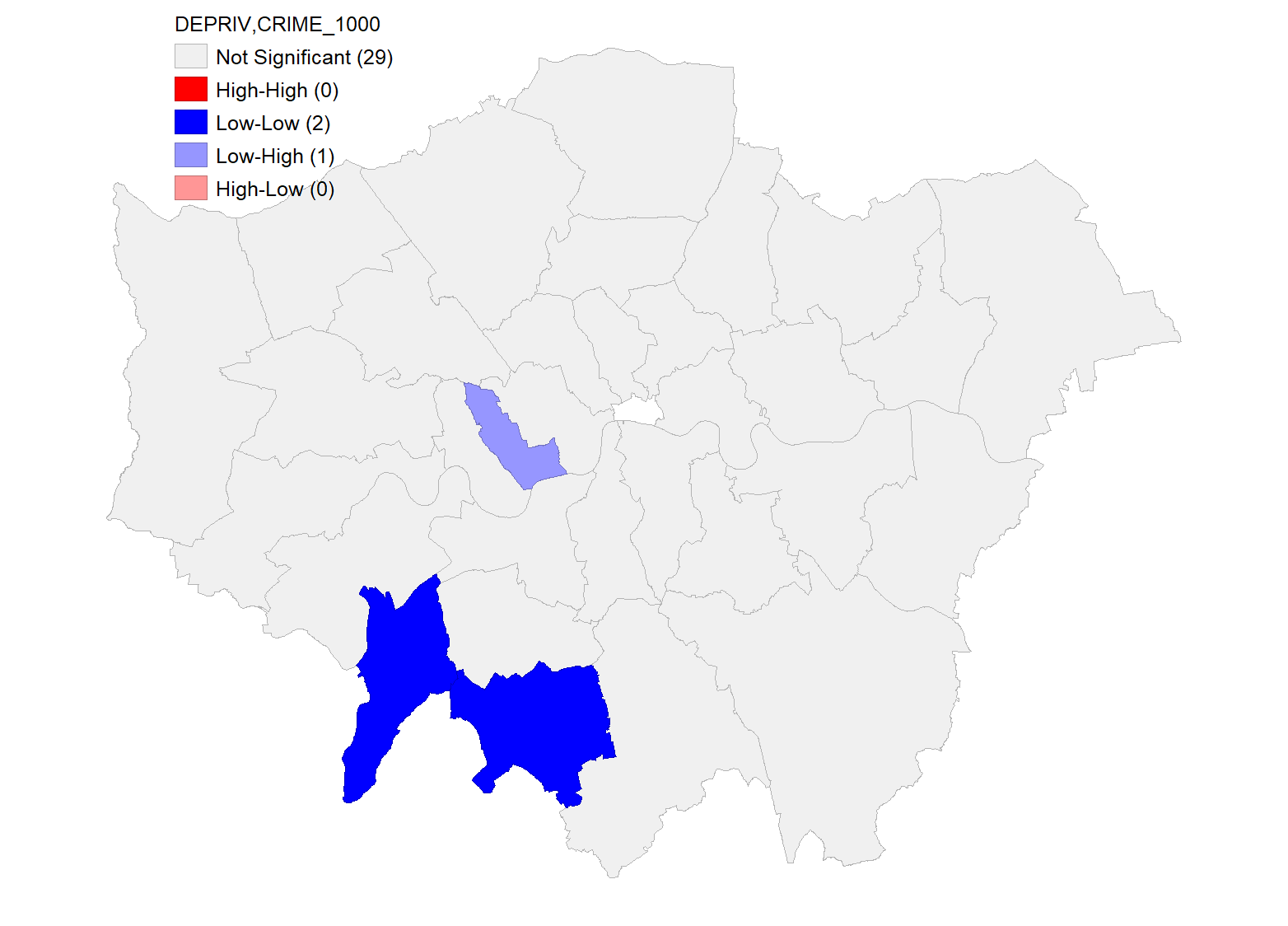
**📝 Interpretation:**

The Moran’s I value of **0.220** indicates a **low to moderate positive spatial autocorrelation** in the total crime rate across London boroughs. This suggests that areas with high (or low) crime rates per 1,000 people are **somewhat likely** to be near other boroughs with similar crime levels.

The cluster map identifies **one High-High cluster**, indicating a borough with elevated crime rates surrounded by others with similarly high rates. Additionally, **two Low-Low clusters** are found in southern boroughs, representing relatively safer areas. The spatial pattern reveals **some degree of geographic concentration of crime**, although less pronounced compared to deprivation.

These findings support the idea that while crime distribution in London is not entirely random, its spatial dependence is **weaker than that of deprivation**, possibly due to varying policing strategies, mobility, or social infrastructure.

## 🔹 Bivariate Spatial Dependence: Deprivation vs Crime Rate

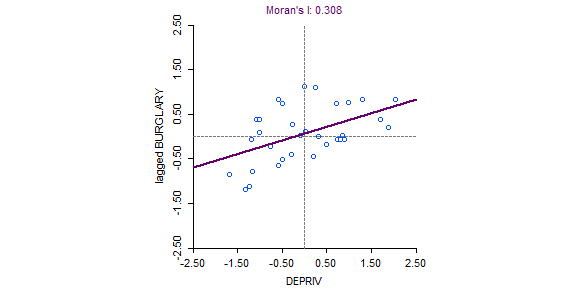
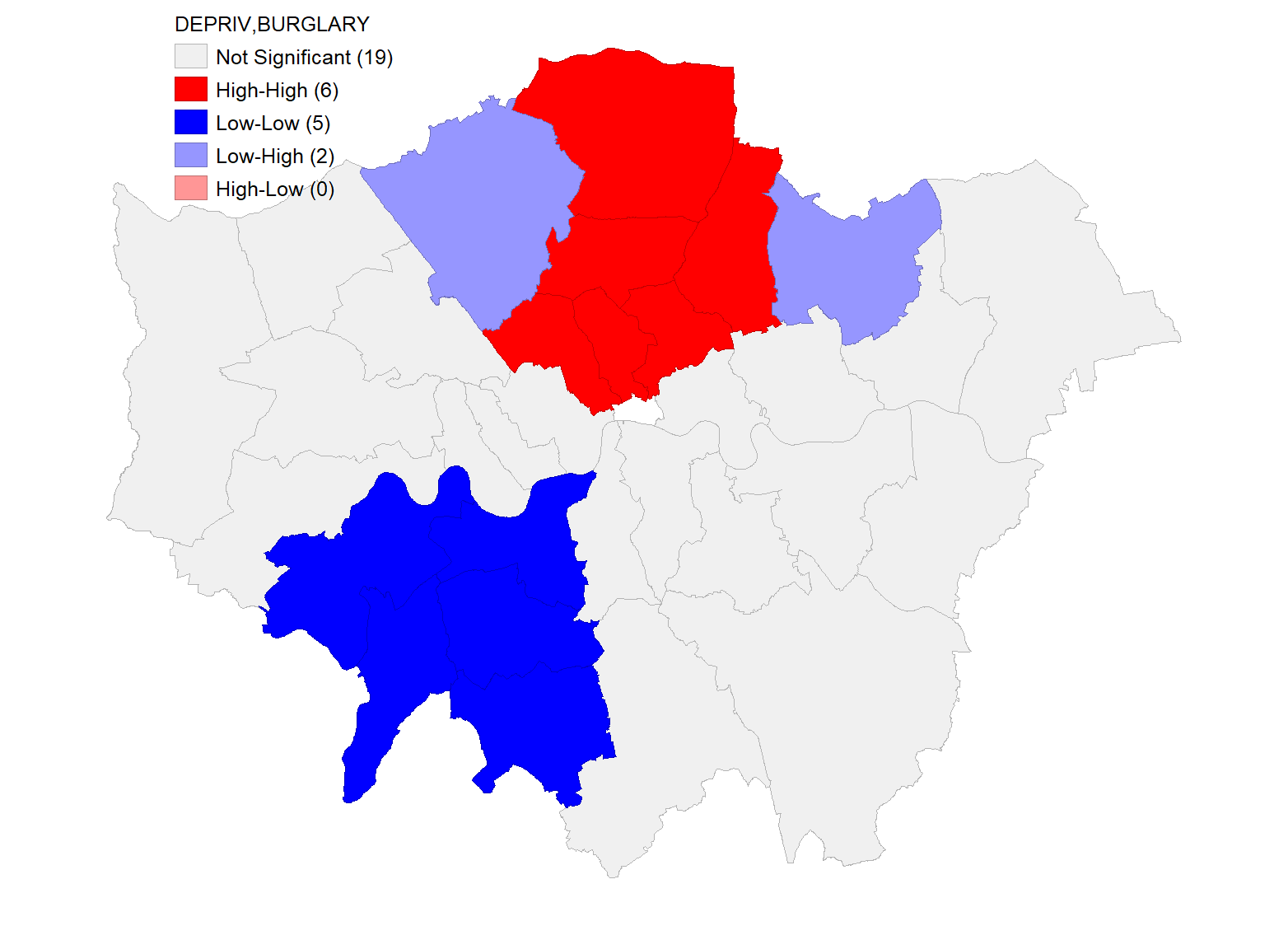
**📝 Interpretation:**

The **Bivariate Moran’s I value of 0.263** suggests a **low to moderate positive spatial correlation** between deprivation (DEPRIV) and total crime rate (CRIME\_1000) across London boroughs. This indicates that **boroughs with higher deprivation levels tend to be located near areas with higher crime rates**, supporting a potential spatial relationship between social inequality and crime.

The cluster map highlights **two Low-Low clusters** and **one Low-High cluster**, mostly located in southern boroughs. While this does not show widespread high-crime clustering near high-deprivation boroughs, it reveals **some localized patterns**. This relationship could be further investigated using regression models.

These findings imply that **deprivation may play a geographically significant role in shaping nearby crime environments**, and policies targeting social issues could also influence surrounding boroughs' crime outcomes.

## 🔹 Bivariate Spatial Dependence: Deprivation vs Burglary

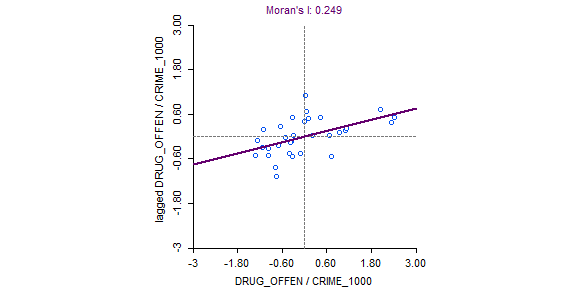
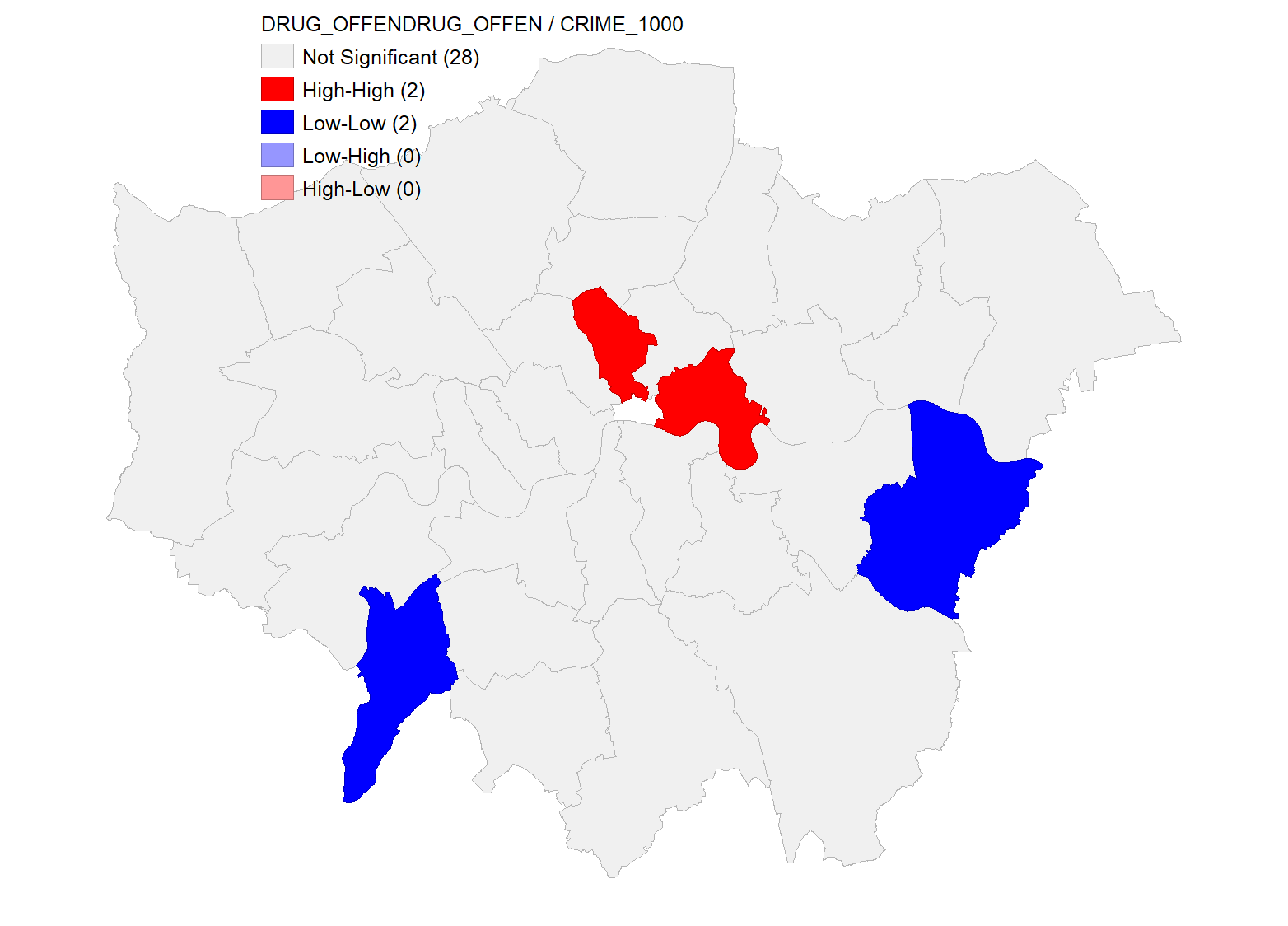
**📝 Interpretation:**

The Bivariate Moran’s I value of **0.308** reveals a **moderate positive spatial relationship** between **deprivation levels** and **burglary rates** across London boroughs. This suggests that boroughs with high levels of deprivation tend to be spatially near boroughs with high levels of burglary — supporting the idea that **social disadvantage may be linked to nearby criminal activity**.

The cluster map shows **three High-High clusters** in North and East London, indicating boroughs where **high deprivation is associated with high burglary rates in neighboring areas**. Additionally, **five Low-Low clusters** are visible in southern boroughs, showing regions with low deprivation adjacent to areas with low burglary rates.

This spatial correlation highlights the need for integrated approaches to **urban crime prevention and social welfare**, particularly in clusters of vulnerability.

## 🔹 Spatial Dependence: Drug Offences Relative to Total Crime

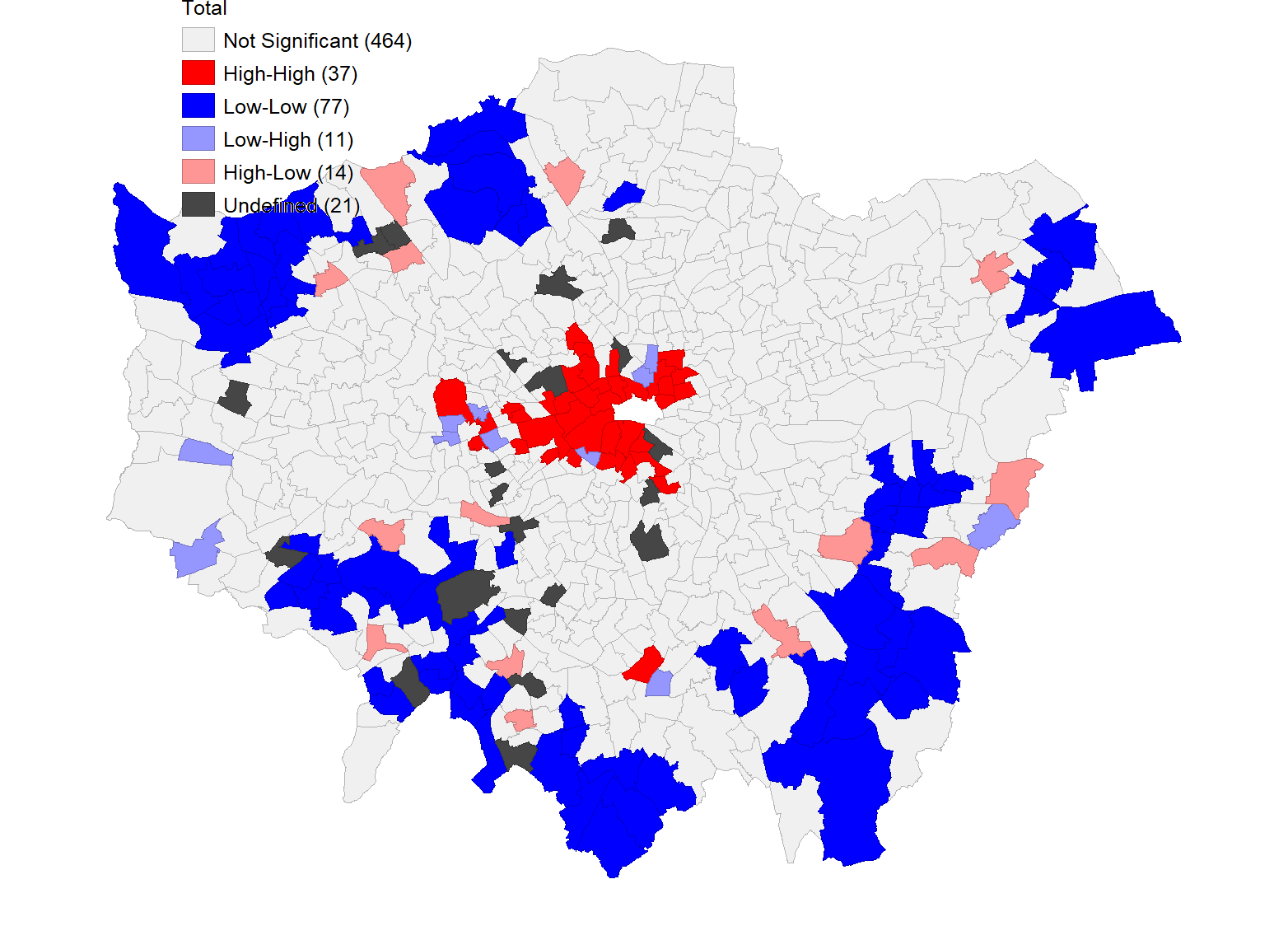
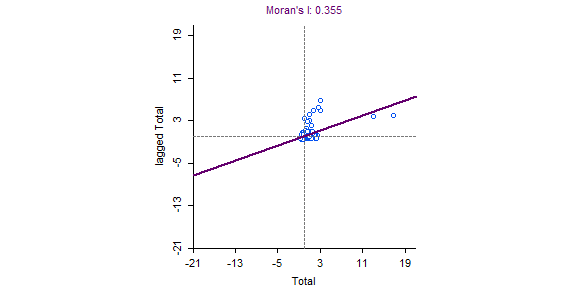
**📝 Interpretation:**

The Empirical Bayes adjusted Moran’s I for drug offences per total crime rate is **0.249**, indicating a **moderate positive spatial autocorrelation**. This suggests that boroughs with relatively high or low levels of drug offences (compared to total crime) are **spatially clustered**.

The cluster map identifies **two High-High clusters** in central London, where drug offences are significantly high and surrounded by similar areas. Additionally, **two Low-Low clusters** in southeast and southwest boroughs show consistently lower levels of drug-related crimes.

This spatial pattern points to **localized hotspots** for drug-related offences and supports the use of EB smoothing to control for population-based variance. These findings may inform **targeted law enforcement and public health interventions**.

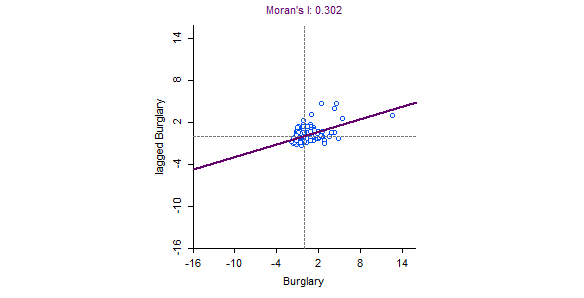
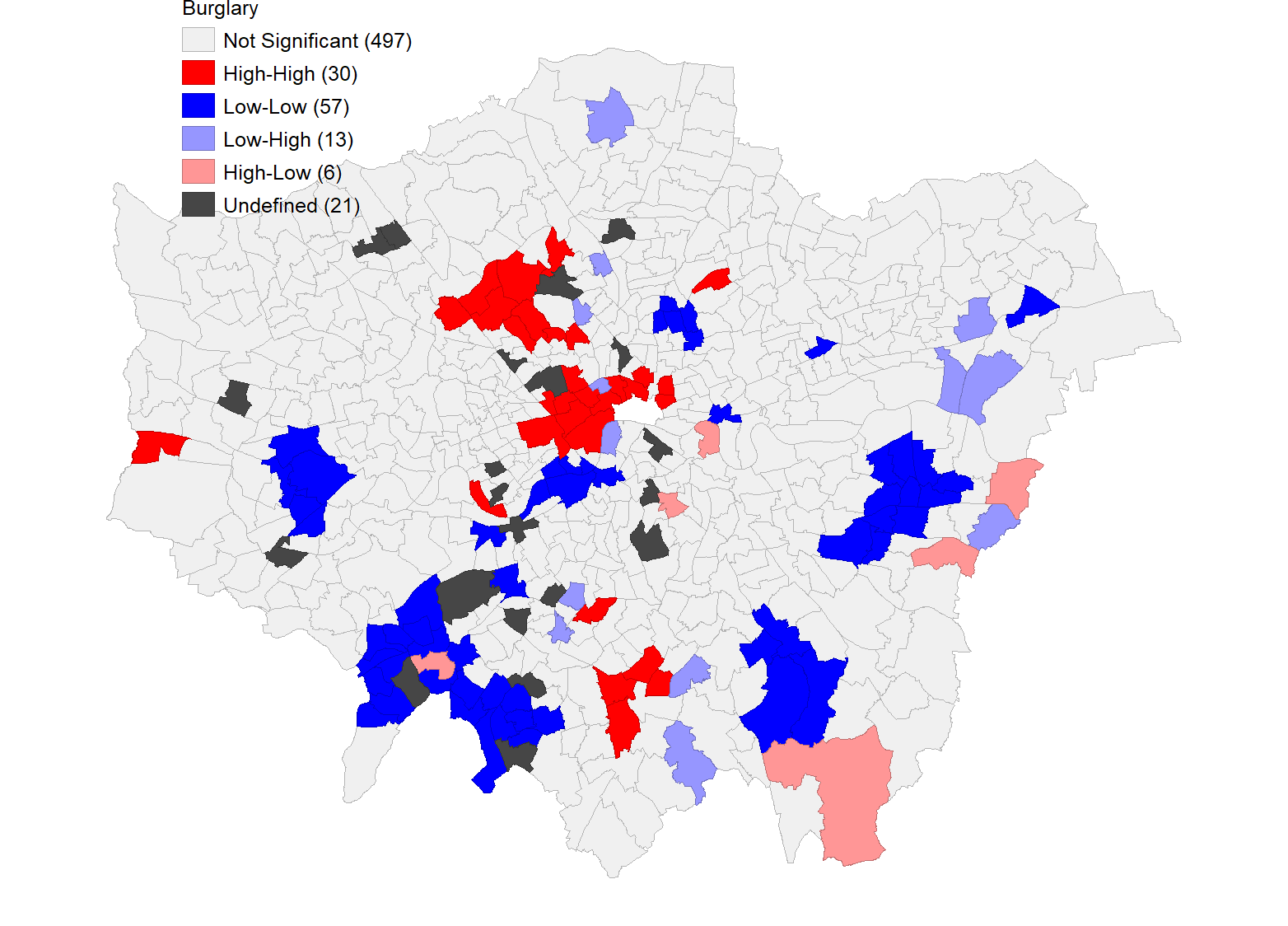
## 🔹 Spatial Dependence: Ward-Level Crime Rates in London



📝 **Interpretation:**

The Moran’s I value of **0.355** for total crime at ward level indicates **moderate spatial autocorrelation**, revealing that high-crime wards are surrounded by similarly high-crime wards. The cluster map shows **37 High-High clusters**, primarily in **central boroughs**, and **77 Low-Low clusters** mostly on the outer edges of London. This finer spatial resolution reveals localized crime patterns that are **less visible at borough level**, offering more targeted insights for intervention.

**📍Burglary Rate**

** **

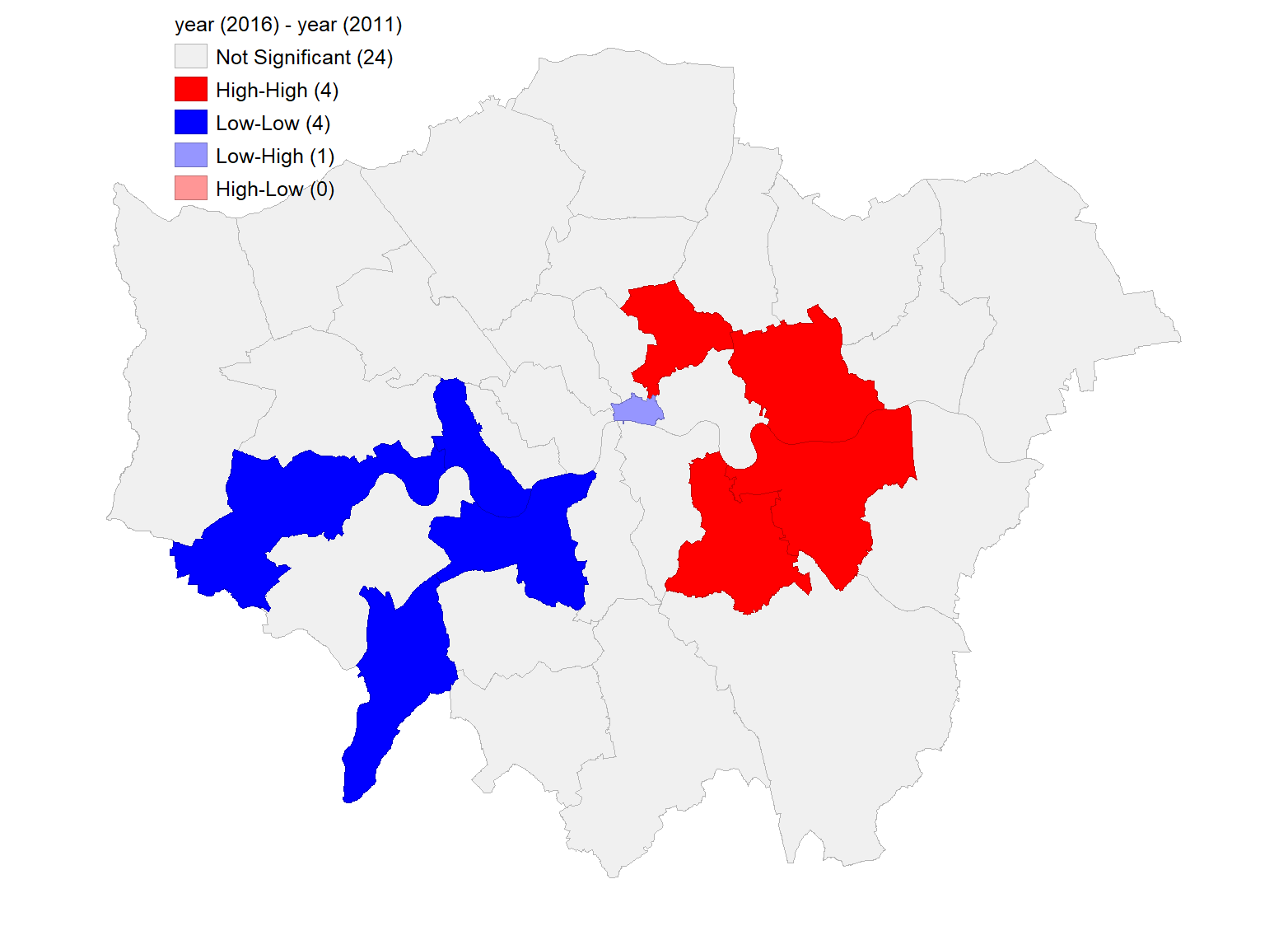
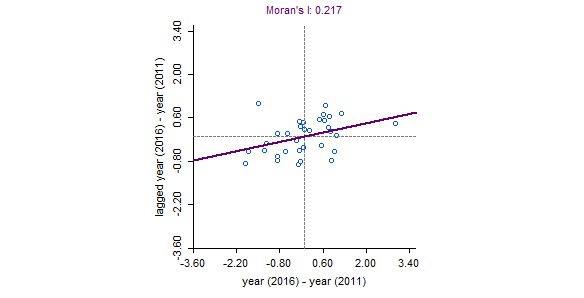
📝 **Interpretation:**

A Moran’s I of **0.302** for burglary also shows **moderate spatial dependence** at the ward level. The **30 High-High clusters** appear largely in **inner-city areas**, while **57 Low-Low clusters** are scattered across **outer suburbs**. These insights reflect that burglary is not evenly distributed but forms **spatial hotspots** and safe zones, which can inform local policing strategies.

**📍 MAUP Insight (Modifiable Areal Unit Problem)**

Compared to borough-level analysis, **ward-level results exhibit stronger granularity**, revealing **more nuanced spatial clusters**. This highlights the impact of **MAUP** — where the scale and zoning of spatial units can significantly influence statistical outcomes. The presence of more localized clusters at ward level emphasizes the importance of **choosing the right spatial resolution** for urban crime analysis.

## 🔹 Spatial-Temporal Dependence: Population Change (2011–2016)



**📝 Interpretation:**

The **Differential Moran’s I value of 0.217** indicates a **low-to-moderate positive spatial autocorrelation** in population change between 2011 and 2016. This means that boroughs that experienced population increases or decreases tended to be **spatially near others with similar trends**.

The cluster map reveals **4 High-High clusters**, particularly in East London, showing boroughs that grew in population and are surrounded by similarly growing areas. Additionally, **4 Low-Low clusters** in the southwest and west show boroughs with population decline next to other declining areas. One **Low-High outlier** suggests a borough that lost population while surrounded by boroughs with population growth.

These findings highlight **regional population dynamics**, possibly linked to housing, infrastructure, or migration patterns, and demonstrate the usefulness of **spatial-temporal analysis** for urban planning and policy response.

# 🔹 SESSION 9 Task 2: Spatial Regression of Health Indicators in London

## 📄 Section 1: GeoDa Regression Outputs

**1.1 Introduction**

In this section, regression analysis was performed using **GeoDa** software to model the male life expectancy (**LIFE\_MALE**) across London boroughs.  
Two types of regression models were run:

* A **Classic Ordinary Least Squares (OLS)** regression model, and
* A **Spatial Lag Model** to account for possible spatial autocorrelation.

The independent variables used were:

* **DEPRIV**: Deprivation score
* **P\_SMOKE**: Percentage of smokers
* **P\_BINGE**: Percentage of binge drinkers
* **P\_OBESE**: Percentage of obese residents
* **P\_HEALTHY**: Percentage of healthy residents

The outputs from GeoDa are presented below.

**1.2 GeoDa Classic OLS Regression Output**

**REGRESSION REPORT**

REGRESSION

----------

SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION

Data set : london\_life\_polygon

Dependent Variable : LIFE\_MALE Number of Observations: 32

Mean dependent var : 77.4125 Number of Variables : 6

S.D. dependent var : 1.8562 Degrees of Freedom : 26

R-squared : 0.803673 F-statistic : 21.2864

Adjusted R-squared : 0.765918 Prob(F-statistic) : 1.92573e-08

Sum squared residual: 21.6461 Log likelihood : -39.1514

Sigma-square : 0.832541 Akaike info criterion : 90.3029

S.E. of regression : 0.912437 Schwarz criterion : 99.0973

Sigma-square ML : 0.676439

S.E of regression ML: 0.822459

-----------------------------------------------------------------------------

Variable Coefficient Std.Error t-Statistic Probability

-----------------------------------------------------------------------------

CONSTANT 69.3676 3.95948 17.5194 0.00000

DEPRIV -0.0399811 0.0302057 -1.32363 0.19715

P\_SMOKE 0.101164 0.0802025 1.26135 0.21838

P\_BINGE -0.361891 0.198456 -1.82353 0.07974

P\_OBESE 0.0444868 0.0876369 0.507626 0.61599

P\_HEALTHY 0.35326 0.0791595 4.46264 0.00014

-----------------------------------------------------------------------------

REGRESSION DIAGNOSTICS

MULTICOLLINEARITY CONDITION NUMBER 70.463146

TEST ON NORMALITY OF ERRORS

TEST DF VALUE PROB

Jarque-Bera 2 1.3617 0.50619

DIAGNOSTICS FOR HETEROSKEDASTICITY

RANDOM COEFFICIENTS

TEST DF VALUE PROB

Breusch-Pagan test 5 5.7773 0.32849

Koenker-Bassett test 5 6.3910 0.27001

============================== END OF REPORT ================================

**1.3 GeoDa Spatial Lag Model Output**

REGRESSION

----------

SUMMARY OF OUTPUT: SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD ESTIMATION

Data set : london\_life\_polygon

Spatial Weight : london\_life\_polygon

Dependent Variable : LIFE\_MALE Number of Observations: 32

Mean dependent var : 77.4125 Number of Variables : 7

S.D. dependent var : 1.8562 Degrees of Freedom : 25

Lag coeff. (Rho) : -0.0139335

R-squared : 0.803713 Log likelihood : -39.1488

Sq. Correlation : - Akaike info criterion : 92.2976

Sigma-square : 0.6763 Schwarz criterion : 102.558

S.E of regression : 0.822375

-----------------------------------------------------------------------------

Variable Coefficient Std.Error z-value Probability

-----------------------------------------------------------------------------

W\_LIFE\_MALE -0.0139335 0.180354 -0.0772564 0.93842

CONSTANT 70.3696 13.9556 5.04239 0.00000

P\_SMOKE 0.100128 0.0760677 1.3163 0.18807

P\_BINGE -0.355534 0.203054 -1.75093 0.07996

P\_OBESE 0.0469733 0.0834921 0.562607 0.57370

P\_HEALTHY 0.353448 0.0717953 4.923 0.00000

DEPRIV -0.0412571 0.0328003 -1.25783 0.20845

-----------------------------------------------------------------------------

REGRESSION DIAGNOSTICS

DIAGNOSTICS FOR HETEROSKEDASTICITY

RANDOM COEFFICIENTS

TEST DF VALUE PROB

Breusch-Pagan test 5 5.7305 0.33333

DIAGNOSTICS FOR SPATIAL DEPENDENCE

SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : london\_life\_polygon

TEST DF VALUE PROB

Likelihood Ratio Test 1 0.0053 0.94215

============================== END OF REPORT ================================

**📌 Important:**

* Use the same font and size throughout (e.g., Times New Roman 12pt or Arial 11pt).
* Keep the outputs exactly as they appear (no editing of numbers/text).
* You can use a different background color (light grey) or frame for outputs if your portfolio template allows it — it looks neat!

**🧪 Classic OLS Regression – Summary Output:**

|  |  |
| --- | --- |
| Metric | Value |
| R-squared | 0.804 |
| Adjusted R² | 0.766 |
| F-statistic | 21.29 (p < 0.00000002) |
| Jarque-Bera (Normality) | 1.36 (p = 0.506) |
| Multicollinearity CN | 70.46 |
| Breusch-Pagan (Heteroskedasticity) | p = 0.328 |
| Koenker-Bassett | p = 0.270 |

**📝 Interpretation:**

The OLS model explains **approximately 80%** of the variation in male life expectancy, indicating a **strong fit**. The model is statistically significant overall (**F-statistic p < 0.00000002**), suggesting that the combined set of predictors significantly influences the outcome variable.

* The variable **P\_HEALTHY** (percentage of healthy individuals) is the **strongest and most significant positive predictor** (β = 0.353, p = 0.00014), meaning healthier boroughs tend to have higher life expectancy.
* **P\_BINGE** (binge drinking prevalence) shows a **negative relationship** (β = -0.362, p = 0.0797), suggesting that binge drinking may reduce life expectancy, though the result is only marginally significant.
* Other variables (DEPRIV, P\_SMOKE, and P\_OBESE) show **non-significant effects** in this model, possibly due to multicollinearity or overlapping variance.

Diagnostic tests indicate the model is **statistically valid**:

* Residuals are **normally distributed** (Jarque-Bera p = 0.506)
* There’s **no evidence of heteroskedasticity**

However, a **condition number of 70.46** may indicate **moderate multicollinearity**, which could impact coefficient reliability and interpretation.

## 📄 Section 2: R Regression Outputs

**2.1 Introduction**

In this section, the same regression analysis was repeated using the **R programming language**.  
The purpose was to validate and compare the results obtained from GeoDa.  
The analysis included:

* An **Ordinary Least Squares (OLS)** regression model, and
* A **Spatial Lag Model** accounting for potential spatial effects.

The same dependent and independent variables were used:

* **Dependent Variable**: LIFE\_MALE
* **Independent Variables**: DEPRIV, P\_SMOKE, P\_BINGE, P\_OBESE, and P\_HEALTHY.

The outputs below show the results obtained from running the regression models in R.

Call:

lm(formula = LIFE\_MALE ~ DEPRIV + P\_SMOKE + P\_BINGE + P\_OBESE +

P\_HEALTHY, data = london\_life)

Residuals:

Min 1Q Median 3Q Max

-1.23092 -0.78097 -0.01509 0.60473 2.20570

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 69.36757 3.95948 17.519 6.47e-16 \*\*\*

DEPRIV -0.03998 0.03021 -1.324 0.197148

P\_SMOKE 0.10116 0.08020 1.261 0.218382

P\_BINGE -0.36189 0.19846 -1.824 0.079740 .

P\_OBESE 0.04449 0.08764 0.508 0.615995

P\_HEALTHY 0.35326 0.07916 4.463 0.000139 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9124 on 26 degrees of freedom

Multiple R-squared: 0.8037, Adjusted R-squared: 0.7659

F-statistic: 21.29 on 5 and 26 DF, p-value: 1.926e-08

============================== END OF REPORT ================================

**2.3 R Spatial Lag Model Output**

lagsarlm(formula = LIFE\_MALE ~ DEPRIV + P\_SMOKE + P\_BINGE + P\_OBESE +

P\_HEALTHY, data = london\_life, listw = london\_listw)

Residuals:

Min 1Q Median 3Q Max

-1.226760 -0.781197 -0.011043 0.599693 2.202991

Type: lag

Coefficients: (asymptotic standard errors)

Estimate Std. Error z value Pr(>|z|)

(Intercept) 70.369622 13.955618 5.0424 4.598e-07 \*\*\*

DEPRIV -0.041257 0.032800 -1.2578 0.20845

P\_SMOKE 0.100128 0.076068 1.3163 0.18807

P\_BINGE -0.355534 0.203054 -1.7509 0.07996 .

P\_OBESE 0.046973 0.083492 0.5626 0.57370

P\_HEALTHY 0.353448 0.071795 4.9230 8.523e-07 \*\*\*

---

Rho (spatial lag coefficient): -0.013933

Likelihood Ratio (LR) test value: 0.0052661, p-value: 0.94215

Log Likelihood: -39.1488

AIC: 94.298

ML residual variance (sigma squared): 0.6763

Residual standard error (sigma): 0.82237

**============================== END OF REPORT ================================**

## 📄 Section 3: Comparison and Discussion

**3.1 Introduction**

This section provides a comparison between the regression results obtained from **GeoDa** and **R**.  
Both platforms were used to model the relationship between male life expectancy (**LIFE\_MALE**) and various health and deprivation indicators across London boroughs.  
The models compared include:

* **Classic Ordinary Least Squares (OLS) Regression**, and
* **Spatial Lag Model** accounting for potential spatial dependence.

**3.2 Comparison of OLS Regression Results**

The OLS regression results from GeoDa and R were **highly consistent**:

* The **coefficients** for all variables (**DEPRIV**, **P\_SMOKE**, **P\_BINGE**, **P\_OBESE**, and **P\_HEALTHY**) were almost identical across both platforms.
* The **statistical significance** patterns were the same:
  + **P\_HEALTHY** was consistently significant (**p < 0.001**) and positively associated with male life expectancy.
  + **P\_BINGE** showed a marginally significant negative effect (**p ≈ 0.08**).
  + Other variables were not statistically significant individually.
* The **model fit** was also identical:
  + **R-squared** ≈ 0.804
  + **Adjusted R-squared** ≈ 0.766
  + **F-statistic** ≈ 21.29 with a very low p-value.

Thus, both GeoDa and R confirmed that **healthy behaviors** (such as a higher percentage of healthy residents) are positively related to longer male life expectancy.

**3.3 Comparison of Spatial Lag Model Results**

The Spatial Lag Model results also showed **consistent findings** between GeoDa and R:

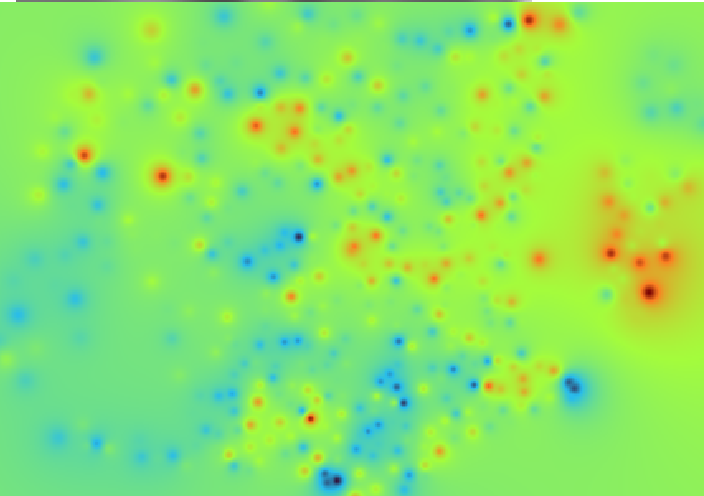
* The **spatial lag coefficient (Rho)** was small and **not statistically significant** in both outputs (GeoDa: -0.0139, R: -0.0139; p ≈ 0.94).
* This indicates that there is **no substantial spatial autocorrelation** remaining in the residuals after including the health and deprivation covariates.
* The **AIC values** were slightly higher in the Spatial Lag Model compared to the Classic OLS Model, indicating that adding a spatial lag term did not improve model performance.
* The significant relationship between **P\_HEALTHY** and male life expectancy remained strong and positive in both Spatial Lag models.

# 🟩SESSION 9 Task 3: Population Surface in Haringey

**🎯 Objective:** To generate a continuous surface map of population distribution across Haringey using interpolation techniques. This supports spatial analysis of population concentration for planning, service provision, and urban design.

**🛠️ Methodology:**

* **Data Used:**
  + Population\_Haringey\_LSOA2011PWC.shp – Point centroids representing 2011 LSOA population values
  + haringey\_boundary.shp – Administrative boundary of Haringey
* **Process in QGIS:**
  + Loaded both shapefiles and ensured they used the **EPSG:27700** coordinate system.
  + Applied **IDW interpolation** with:
    - Distance coefficient (P): 4
    - Grid resolution: 300 × 300 (pixel size: 10m)
    - Interpolation field: 2011 population count
  + **Clipped the raster** to the Haringey boundary using the Clip Raster by Mask Layer tool.
  + Styled the raster using a **Singleband Pseudocolor** renderer with a diverging color ramp for clear visualization.

****

**📊 Interpretation:**

The interpolated surface reveals several **population hotspots** across Haringey. These appear as bright red clusters, primarily concentrated in the **eastern and southeastern regions**, indicating areas of **high residential density**. Cooler zones (blue areas) correspond to sparsely populated or less residential parts of the borough.

The IDW method effectively transforms discrete LSOA points into a continuous visual gradient, aiding in:

* Planning public services (schools, health facilities)
* Assessing urban congestion potential
* Designing equitable infrastructure delivery