UCL DEPARTMENT OF POLITICAL SCIENCE

SCHOOL OF PUBLIC POLICY

Essay Front Page

Please complete this sheet electronically and attach it as the front page

for each essay submitted.

Candidate Number: QHLB4

Essay Title, Question Number/s or Text of Question/s you have answered:

Data Analysis assessment II

Essay Number: 2

Module Code: POLS0010

Module Title: Data Analysis

Submission Date: 12/01/2025

Word Count: 1490

The essay cover sheet does not count towards the word count on your submission.

Please tick this box if you **do not** agree to your essay being used anonymously in future teaching ☐

Spatial analysis report on Lambeth

2025-01-06

Executive summary

The study finds a significant clustering pattern of high unemployment in north-east Lambeth, while low unemployment clusters exist in western Lambeth. The negative correlation between educational attainment (with a higher rank indicating less educated) and the unemployment rate is stronger in mid-Lambeth, coupled with a slight decline in absolute coefficient in southern areas. Therefore, to tackle the unemployment issue, it is recommended that councils target specific economic and educational policies in south-west and north-west areas of the borough. Applicable policies could include reduced educational expenditure and enhanced educational resources, which will incentivise citizens to access education.

The dataset:

This report focuses on the unemployment distribution in Lambeth using 2021 census data. Unemployment (Y) is defined by metadata.csv as 'being economically active but unemployed', and the data has been filtered to be Lambeth-specific. Initial unemployment distribution is displayed in the study, and Local Indicators of Spatial Association (LISA) are used for determining clustering patterns, exploring whether unemployment is spatially relevant. Given that people with poor educational qualifications are more likely to experience unemployment in their adulthood (Barry, 2019), this report also aims to discuss the correlation between unemployment rate (Y) and educational attainment. Education and Skill Deprivation (X) is selected from IMD2019 data, assessing the qualifications attained by under-18s and adults. As the scores of education attainments do not directly relate to the proportion of the population experiencing deprivation, rank is used as the indicator of education, with a higher rank meaning more educationally deprived. Geographically Weighted Regression (GWR) is adopted to analyse the local relationship between X and Y. Initial education attainments distribution and its coefficient map are also included, showing the strength of the correlation across Lambeth.

Findings

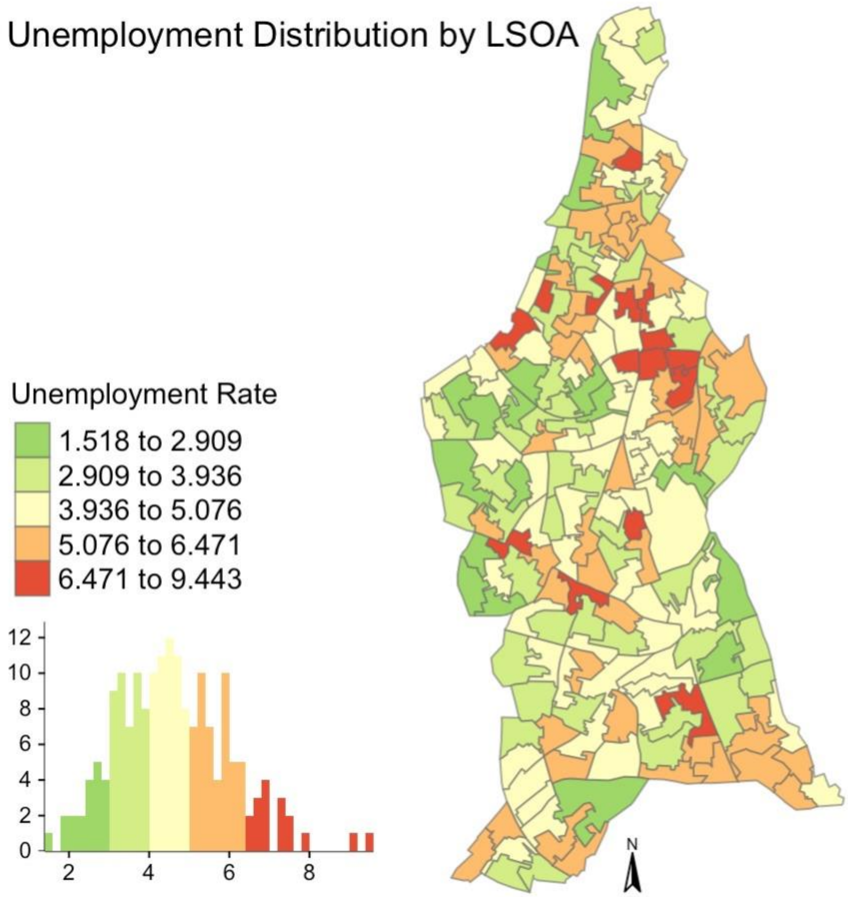

Figure 1: Unemployment Distribution Map by LSOA

The unemployment rate ranges from 1.52% to 9.44% and is divided into five groups, with the darkest red representing unemployment rates of 6.47% to 9.44%, which signify relatively severe unemployment status. Conversely, the lightest red represents unemployment rates of 1.52% to 2.90%, highlighting relatively lower unemployment. The histogram shows a roughly normal distribution of unemployment rates, while north-east regions exhibit notable higher unemployment rates of over 5.08%, and the overall margins of Lambeth show relevant lower unemployment rates up to 3.94%.

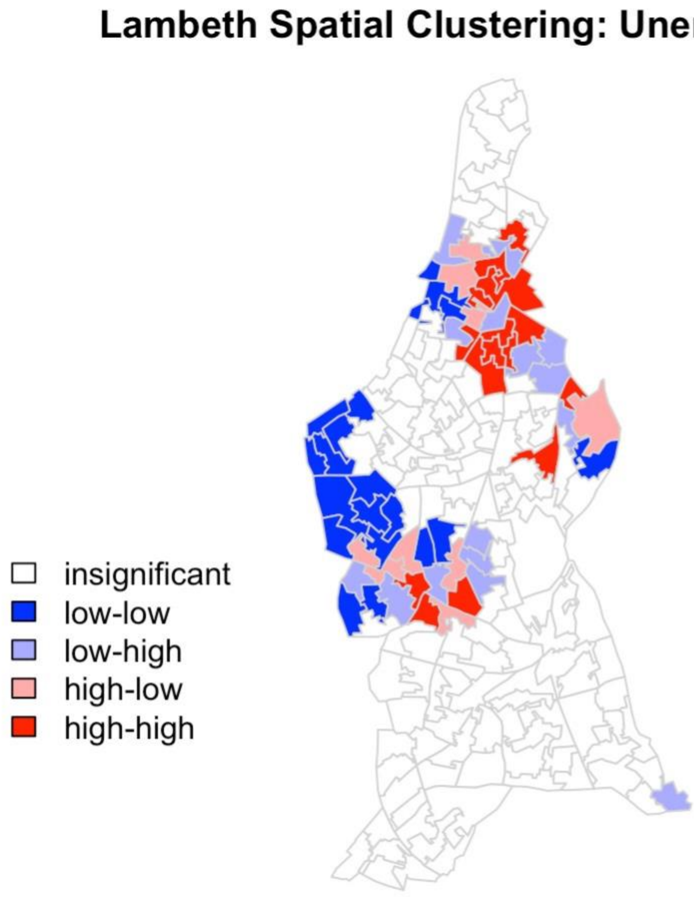

Figure 2: LISA Cluster Map of Spatial Autocorrelation

The unemployment rate ranges from 1.52% to 9.44% and is divided into five groups, with the darkest red representing unemployment rates of 6.47% to 9.44%, which signify relatively severe unemployment status. Conversely, the lightest red represents unemployment rates of 1.52% to 2.90%, highlighting relatively lower unemployment. The histogram shows a roughly normal distribution of unemployment rates, while north-east regions exhibit notable higher unemployment rates of over 5.08%, and the overall margins of Lambeth show relevant lower unemployment rates up to 3.94%.

High-High cluster, indicated by dark red, represents areas with high unemployment that are geographically concentrated and adjacent. In the map, north-east Lambeth displays prevalent high-high clusters, while some areas in mid-Lambeth show the same characteristics, corresponding to the previous unemployment distribution map. Conversely, low-low clusters are highlighted in dark blue, representing closely located low unemployment areas. On the map, western Lambeth exhibits significant low-low areas.

Low-high and high-low clusters represent areas with both low and high unemployment rate that are more difficult to explain. The map shows notable low-high and high-low clusters in north-west and north-east Lambeth, indicated in light blue and light red, signifying a lower unemployed region bounded by higher unemployed areas and vice versa. The rest of the map is dominated by insignificant clustering, indicating a more randomly dispersed unemployment distribution.

Therefore, Lambeth Council should target specific policies to tackle those high-unemployment clusters, including per-unit business subsidies or improved transportation, to ensure the area is worth investing in and creating further job opportunities. However, Munyoro (2017) argues that unemployment benefits may lead businesses to be less willing to invest in assisted areas, as there may be fewer skilled workers due to being deskilled or being unemployed for long periods. Thus, he recommends reducing unemployment benefits in order to encourage the unemployed to take low paid jobs. For low-high and high-low regions, further research should be focused on the characteristics that cause such divergent employment rates in closely related areas.

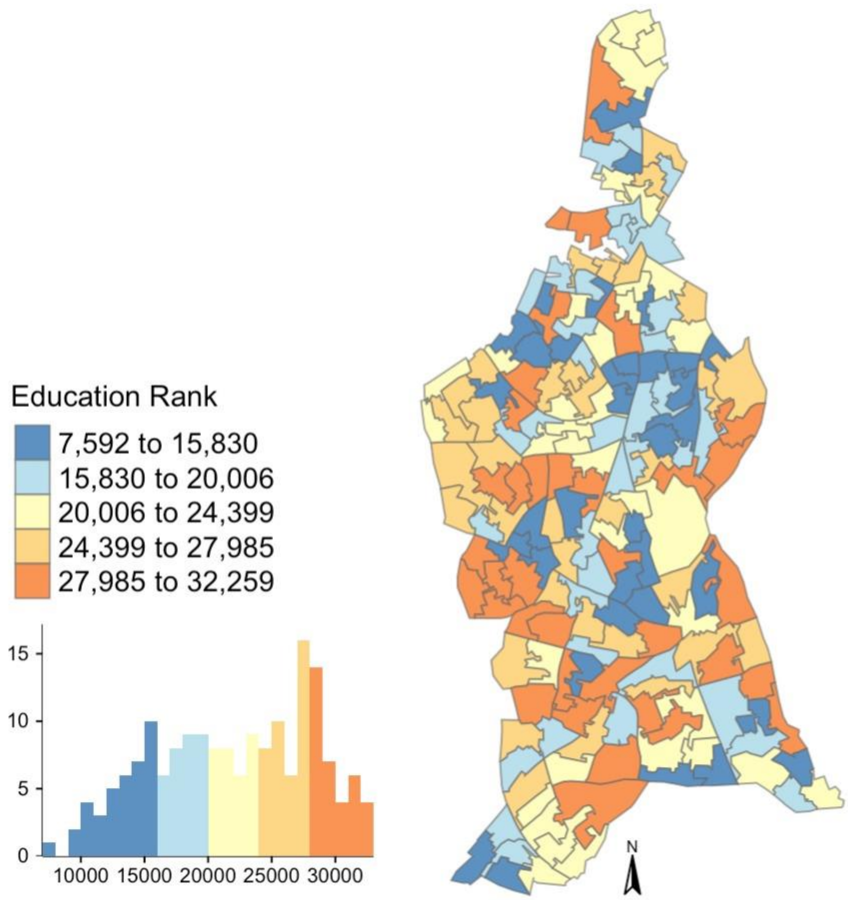

Figure 3: Education distribution and GWR coefficient of education

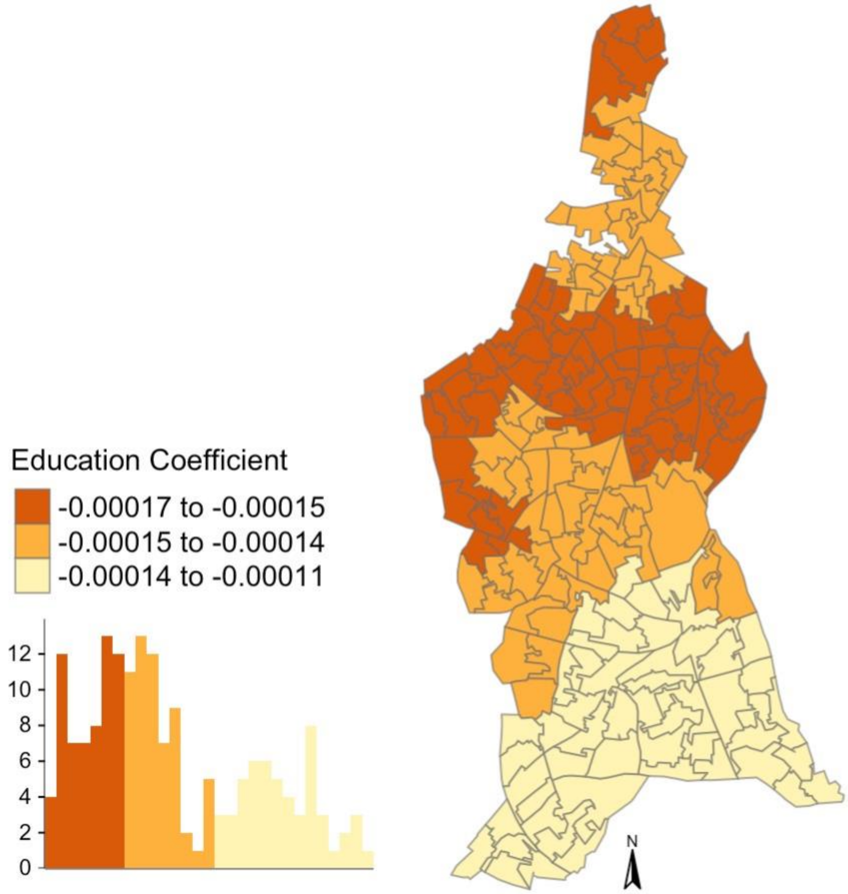

Figure 3: Education distribution and GWR coefficient of education

The distribution of education attainments (X) shows that the areas of north-east and northern Lambeth are highly educationally deprived, with most people ranked from 7592nd to 20006th, while south-west and western Lambeth are relatively less deprived, with most people ranked from 20006th to 32259th. This result corresponds to Y output, which indicates a higher unemployment rate in north-east Lambeth. The education coefficient map also demonstrates some significant outputs, with the west-mid-east and upper north regions having a stronger coefficient of -0.017% to -0.015%, meaning that one unit higher in the educational deprived rank is correlated with 0.017% to 0.015% increase in unemployment. Although the strength in the northern and southern regions is slightly weaker, the coefficients are still significant, from -0.015% to - 0.011%.

This report has demonstrated that in London, particularly Lambeth, high levels of educational deprivation are coupled with high levels of unemployment rates. These findings align with other studies elsewhere, for example, graduates in Paris are more likely to be employed than individuals with educational deprivation (OECD, 2000). The findings relating to Lambeth can be partially explained by the signaling theory that employers see education qualification as an indicator of the accumulation of capital and productivity, and people with lower educational attainment are deemed as unqualified and rejected in job interviews.

In conclusion, northeast Lambeth is determined as highly unemployed and is geographically close to other regions with high unemployment rates. In addition, north-east Lambeth is found to have low education levels, with a strong correlation with its high unemployment rates. Thus, Lambeth Council should focus on this area, invest in its education infrastructure, and create work opportunities, including providing subsidies for local businesses to increase occupations and increase access to devices to enhance individual study efficiency and ensure higher education quality. Further research should investigate the characteristics and causes of eastern Lambeth, where a low-unemployment region is mixed with a high-unemployment region.

Technical Appendix

Classification: Quantile and Jenks

In this study, both Quantile and Jenks methods are used to classify different variables. For education ranks with continuous values, quantile classification is adopted to group the values into five equal groups, ensuring the comparison among each group is meaningful without concerning data skewness. For Moran statistics, local R2, and Education coefficient maps, Jenks method is used to highlight the breaks with different groups. Therefore, variance within classes is minimized and variance between classes is maximized (Jenks,1967).

Clustering: Moran's Test

Moran's I autocorrelation is used to identify clustering patterns in Lambeth. Unlike traditional point-pattern analysis, which requires specific household coordinates, Moran's I is better suited for this analysis as it relies on the output areas provided in the data.

The study used Rook's case and Queen's case definition of Neighbors at the beginning. After comparing the average links and plotting Moran's I local statistics map, results are found to be unrealistic in that some value exceeds +1 or -1. Therefore, both Rook's case and Queen's case are not applicable, and distance-based neighbourhood definition is adopted. By comparing Moran's global statistics and p-value, the optimal distance threshold is set as 2000m.

By setting the distance threshold as 2000m and the confidence level at 0.95, the Moran's I global statistics return a Z-score of 2.8574 and a p-value of 0.0021, rejecting the null hypothesis that Lambeth is randomly clustered. Moran I statistics of 0.032 signifies a positive spatial autocorrelation, though the strength is relatively weak. A Local Indicators of Spatial Association map (LISA) is adopted to inform Lambeth Council that highlighted regions with clustering, warranting targeted policies and further investigation.

Correlation: GWR Geographically Weighted Regression is adopted to analyze the correlation between unemployment and educational attainments, allowing for comparison in a visualized map. An adaptive bandwidth is used to balance data density as it shrinks in denser areas and stretches in less populated areas, avoiding misleading results due to data skewness (Comber, 2023). The local R2 map shows a moderate explanatory power of educational attainments, ranging from 0.594 to 0.310. Mid-Lambeth shows stronger explanatory strength of educational attainments of 0.423 to 0.594, while upper-north and south Lambeth exhibits weaker explanatory power around 0.310 to 0.372. Initial educational rank distribution and its coefficient are selected to identify the regions with higher education deprivation coupling with high coefficients, signifying council that specific policy to implement.

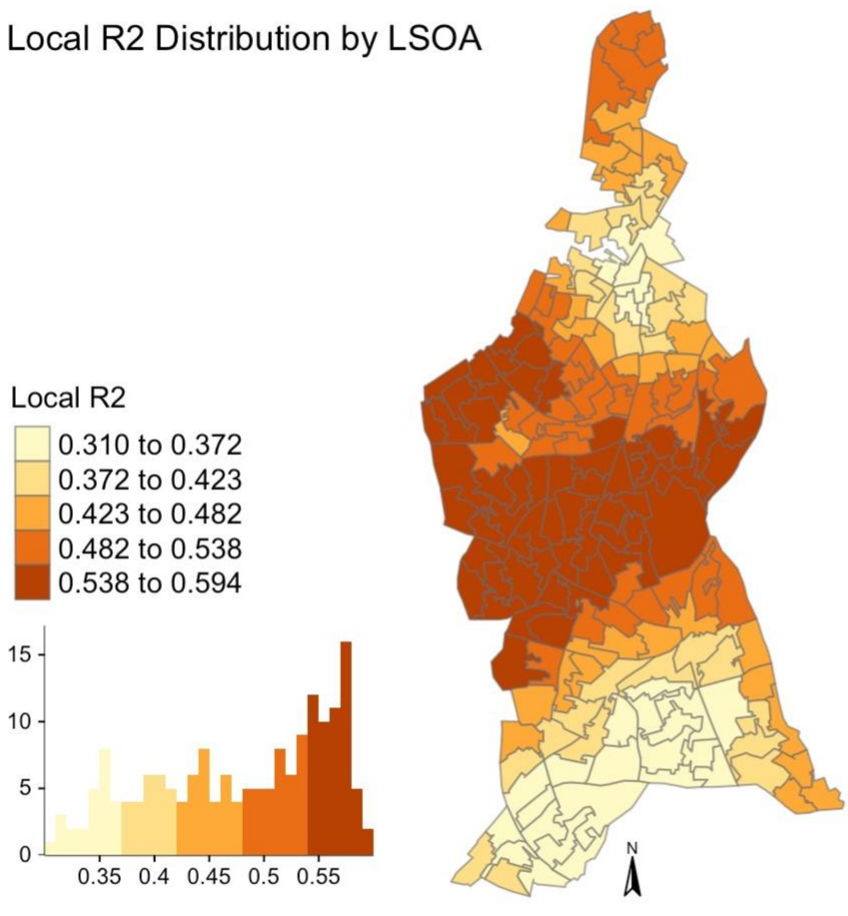

figure 4: Local R2 distribution

Limitation: Since unemployment data is formatted as OA units, converting to LSOA units requires manipulation of data, and this study simply computes the mean of various OAs into one LSOA unemployment value. Though it is reliable to convert from smaller to larger units, demonstrated by Norman (2012), it may incur biases as the mean can be greatly influenced by outliers in an LSOA unit. In addition, to explore unemployment's relationship with educational attainments, the IMD2019 dataset is selected with census 2021. In recent years, the job environment has been detrimentally affected by COVID-19 (Tetlow, 2020), and the coefficient of educational attainment can be over-magnified. Education rank distribution adopts quantile classification, but as the Lambeth data is extracted from UK data, education rank is not in Lambeth order and may confuse interpretation. Future research is recommended to search for updated 2021 IMD LSOAs data and subset Lambeth data and causes of low-high and high-low regions should be further investigated to target relevant policies.

To conclude, every method is carefully selected to meet the requirements and address unique characteristics at spatial analysis.

Reference list

Barry, F. and Hannan, A., 2019. Education, deprivation, hysteresis, unemployment. Unemployment in Ireland, pp.75-86.

Jenks, G.F., 1967. The data model concept in statistical mapping. International yearbook of cartography, 7, pp.186-190.

MUNYORO, G., 2017. THE SIGNIFICANCE OF REGIONAL POLICY TO SCOTTISH ECONOMY.

Lu, B., Hu, Y., Yang, D., Liu, Y., Liao, L., Yin, Z., Xia, T., Dong, Z., Harris, P., Brunsdon, C. and Comber, L., 2023. GWmodelS: A software for geographically weighted models. SoftwareX, 21, p.101291.

Norman, P. and Riva, M., 2012. Population health across space and time: the geographical harmonisation of the Office for National Statistics Longitudinal Study for England and Wales. Population, Space and Place, 18(5), pp.483-502.

Organisation for Economic Co-operation and Development. (2000). From initial education to working life: Making transitions work. Paris: OECD.

Tetlow, G., Pope, T. and Dalton, G., 2020. Coronavirus and unemployment. Institute for Government Insight.

Appendix

```
setwd("~/Desktop/Code/R studio")

# Load required libraries
library(sf)
library(spatstat)
library(tmap)
library(dplyr)
library(spgwr)
library(spdep)
library(grid)
library(gridExtra)

# Read input data
census_data <- read.csv("~/Downloads/Worksheet_Data/eng_wales_practical_data.csv")
lsoa_shapefile <- st_read(dsn = "~/Downloads/worksheet_data/Census_OA_Shapefile.geojso
n")
oa_shapefile <- st_read("~/Downloads/Worksheet_Data/OA_2021_EW_BGC_V2.shp")

# Process centroids for LSOA
lsoa_centroids <- st_centroid(lsoa_shapefile)
lsoa_lambeth <- oa_shapefile[grepl('Lambeth', oa_shapefile$LSOA21NM),]

# Merge census data with the LSOA shapefile based on OA code
lambeth_census <- merge(lsoa_lambeth, census_data, by.x ="OA21CD", by.y = "OA")
lambeth_census_lsoa <- lambeth_census[, -1] # Remove redundant column
st_write(lambeth_census, dsn = "~/Downloads/Worksheet_Data/Lambeth_Census_Shapefile.geoj
son", driver="GeoJSON")

# Create OA to LSOA mapping
oa_to_lsoa <- lsoa_lambeth[, c("OA21CD", "LSOA21CD")]

# Merge census data with LSOA mapping based on OA code
merged_data <- merge(census_data, oa_to_lsoa, by.x ="OA", by.y = "OA21CD")

# Aggregate data by LSOA (e.g., calculate average unemployment rate)
aggregated_data <- merged_data %>%
  group_by(LSOA21CD) %>%
  summarise(Unemployment_Rate = mean(Unemployed, na.rm =TRUE))

# Merge aggregated data with LSOA shapefile
lsoa_areas <- st_read("~/Downloads/Worksheet_Data/Lower_layer_Super_Output_Areas_2021_EW
_BGC_V3.geojson")
lsoa_merged <- merge(lsoa_areas, aggregated_data, by.x ="LSOA21CD", by.y = "LSOA21CD")



tm_shape(lsoa_merged) +
  tm_fill("Unemployment_Rate", legend.hist = TRUE, palette = "-RdYlGn" , style = "jenk
s", n = 5, title = "Unemployment Rate") +
  tm_borders(alpha = .6) +
```

```
  tm_compass(size = 2, fontsize = 0.5,position = c("center","bottom"))+

  tm_layout(title = "Unemployment Distribution by LSOA",legend.text.size = 0.9,frame
= F
ALSE,asp = 0,legend.position = c("left","bottom"),legend.hist.width = 0.3,title.size
= 1.1)

# Spatial neighbors for poly2nb and different types of adjacency neighbours <-
poly2nb(lsoa_merged) neighbours_rook <- poly2nb(lsoa_merged, queen = FALSE)

# Plot both types of neighbor relationships par(mfrow = c(1, 2)) plot(neighbours,
st_geometry(lsoa_merged), col = 'red') plot(neighbours_rook,
st_geometry(lsoa_merged), col = 'blue')

# Create spatial weight list for Moran's test listw <- nb2listw(neighbours_rook,
style = "W")

# Run Moran's I test for spatial autocorrelation
moran.test(lsoa_merged$Unemployment_Rate, listw)

# Calculate lagged unemployment rate and plot lagged_rate <- lag.listw(listw,
lsoa_merged$Unemployment_Rate) unemployment_rate <- lsoa_merged$Unemployment_Rate
plot(unemployment_rate, lagged_rate) abline(h = mean(unemployment_rate), lty = 2)
abline(v = mean(lagged_rate), lty = 2)

# Recenter the data recentered_uep <- unemployment_rate - mean(unemployment_rate)
recentered_lag <- lagged_rate - mean(lagged_rate) plot(recentered_uep,
recentered_lag) abline(h = 0, lty = 2) abline(v = 0, lty = 2)

# Linear model for spatial lag lm_model <- lm(recentered_lag ~ recentered_uep)
abline(lm_model, lty = 3)

# Local Moran's I local_moran <- localmoran(x = lsoa_merged$Unemployment_Rate, listw
= nb2listw(neighbours
_rook, style = "W")) moran_map <- cbind(lsoa_merged, local_moran)
```

```
# Plot Local Moran's I map tm_shape(moran_map) +   tm_fill(col = "Ii", style = "jenks",
title = "Local Moran's I statistics")

# Nearest neighbors based on distance (2 km) distance_neighbours <-
dnearneigh(st_centroid(lsoa_merged), 0, 2000) moran.test(lsoa_merged$Unemployment_Rate,
listw = nb2listw(distance_neighbours, style = "W"))
```

```
# Local Moran's I with distance-based neighbors local_moran_distance <- localmoran(x
= lsoa_merged$Unemployment_Rate, listw = nb2listw(d istance_neighbours, style = "W"))
moran_map_distance <- cbind(lsoa_merged, local_moran_distance)

# Plot Local Moran's I with distance-based neighbors tm_shape(moran_map_distance) +
tm_fill(col = "Ii", style = "jenks", title = "Local Moran's I statistic (Distance-
Base d)")

# Define quadrants based on recentered data quadrant <- numeric(length =
nrow(local_moran_distance)) significance_threshold <- 0.05 quadrant[recentered_uep <
0 & recentered_lag < 0] <- 1 quadrant[recentered_uep < 0 & recentered_lag > 0] <- 2
quadrant[recentered_uep > 0 & recentered_lag < 0] <- 3 quadrant[recentered_uep > 0 &
recentered_lag > 0] <- 4 quadrant[local_moran_distance[, 5] > significance_threshold]
<- 0

# Define color scheme for quadrant visualization color_breaks <- c(0, 1, 2, 3, 4)
colors <- c("white", "blue", rgb(0, 0, 1, alpha = 0.4), rgb(1, 0, 0, alpha = 0.4),
"re d")

# Plot the spatial clustering map plot(lsoa_merged[1], border = "lightgrey",      col
= colors[findInterval(quadrant, color_breaks, all.inside = FALSE)],      main =
"Lambeth Spatial Clustering: Unemployment") legend("bottomleft", legend =
c("insignificant", "low-low", "low-high", "high-low", "hig h-high"), fill = colors,
bty = "n")

# Read IMD data and merge with census data imd_data <- st_read(dsn =
"~/Downloads/English IMD 2019/IMD_2019.dbf") imd_lambeth <- imd_data[grepl('Lambeth',
imd_data$lsoa11nm),]
merged_imd_census <- merge(imd_lambeth, aggregated_data, by.x = "lsoa11cd", by.y =
"LSOA 21CD")

# Fit linear model for unemployment rate and education rank lm_model_imd <-
lm(merged_imd_census$Unemployment_Rate ~ merged_imd_census$EduRank)
summary(lm_model_imd)
```

```
# Plot model diagnostics plot(lm_model_imd, which = 3)

# Residuals and mapping residuals_data <- residuals(lm_model_imd) map_residuals <-
cbind(merged_imd_census, residuals_data) qtm(map_residuals, fill = "residuals_data")

# GWR model imd_census_spatial <- as(merged_imd_census, "Spatial") gwr_bandwidth <-
gwr.sel(imd_census_spatial$Unemployment_Rate ~ imd_census_spatial$EduRa nk, data =
imd_census_spatial, adapt = TRUE) gwr_model <- gwr(imd_census_spatial$Unemployment_Rate
~ imd_census_spatial$EduRank, data
= imd_census_spatial, adapt = gwr_bandwidth, hatmatrix = TRUE, se.fit = TRUE)

# GWR results gwr_results <-
as.data.frame(gwr_model$SDF)
gwr_map <- cbind(imd_census_spatial, as.matrix(gwr_results))

# Plot GWR Local R2 distribution map tm_shape(gwr_map) +   tm_fill("localR2", style
= "jenks", legend.hist = TRUE, n = 5, title = "Local R2") +   tm_borders(alpha =
0.6) +   tm_compass(size = 2, fontsize = 0.5,position = c("center","bottom"))+
  tm_layout(title = "Local R2 Distribution by LSOA", legend.text.size = 0.9,frame =
FALS E,asp = 0,legend.position = c("left","bottom"),legend.hist.width = 0.3,title.size
= 1.1)

# Plot education rank and coefficient maps
map1 <- tm_shape(gwr_map) +
  tm_fill("EduRank", n = 5, style = "quantile", title = "Education Rank", palette = "-
Rd YlBu", legend.hist = TRUE) +   tm_borders(alpha = 0.6) +   tm_compass(size = 2,
fontsize = 0.5,position = c("center","bottom"))+   tm_layout(legend.text.size =
0.9,frame = FALSE,asp = 0,legend.position = c("left","bot tom"),legend.hist.width =
0.3,title.size = 1.1)

map2 <- tm_shape(gwr_map) +   tm_fill("imd_census_spatial.EduRank", n = 3, style =
"jenks", title = "Education Coeff icient", legend.hist = TRUE) +
  tm_layout(legend.text.size = 0.9,frame = FALSE,asp = 0,legend.position =
c("left","bot tom"),legend.hist.width = 0.3,title.size = 1.1) +   tm_compass(size = 2,
fontsize = 0.5,position = c("center","bottom"))+   tm_borders(alpha = 0.6)
```

```
# Display both maps

print(map1)

print(map2)


# Correlation and scatter plot for GWR results

round(cor(gwr_results[, c(3, 5, 7, 8)], use = "complete.obs"), 2)

pairs(gwr_results[, c(3, 5, 7, 8)], pch = ".")


library(psych)

describe(lsoa_merged$Unemployment_Rate)

describe(merged_imd_census$EduRank)
```

Dataset description

Object name	Column code	Description	Source *The main source used is the 2021 census and IMD_2019.dbf datasets
Unemployment rate	OA	Lower-layer super output area	"Table_Metadata.csv" is used to derive the names and columns of the 2021 census data, and OA column is used to combine with LSOA shaoefile data. Educational attainment's names and columns are derived from File_7_-_All_IoD2019_Scores__Ranks__Deciles_and_Population_Denominators_3.csv.
	ts0660013	Economically active (excluding full-time students): Unemployed	
	ts0660001	All usual residents aged 16 years and over	
Educational attainment	lsoa11cd	LSOA code (2011)	
	EduRank	Education, Skills and Training Rank (where 1 is most deprived)	

Variable	Description	Summary Statistics Mean (SD)
Unemployment_Rate	Contains the percentage of unemployed of total economically active people, calculated from divided ts0660001 by ts0660013, column for each OA level and multiply the result by a 100. This was transformed to LSOA level by calculating the mean of included OA area.	4.6%(1.39%)
EduRank	Contains the ranks of education deprivation in LSOA level	21,951.32(6,147.31)

Other Dataframes	Description	Source
agregated_data	This is created by merging census_data with LSOA mappings and then calculating the average unemployment rate (Unemployed) for each LSOA (LSOA21CD), resulting in a data frame containing LSOA21CD and the average	Derived(NA)
lsoa_merged	This spatial dataset merges the LSOA shapefile (lsoa_areas) with the aggregated_data, resulting in a file containing both LSOA geometries and the average unemployment rate for each LSOA.	
merged_imd_census	Although not shown in the code, merged_imd would likely combine an IMD (Index of Multiple Deprivation) dataset with the lsoa_merged, adding deprivation scores to each LSOA's unemployment data.	

table