**Author: Isaiah Bowen-Karlyn**

**Date: 29/11/24**

**1. Introduction**

The World Health Organisation (WHO, 2018) identifies noise as the top environmental risk to health. In western Europe, noise from road traffic results in the loss of 1.6 million healthy years of life. Physiologically, excessive or prolonged exposure to noise leads to hearing loss and tinnitus. Psychologically, noise is a stressor, disturbing homeostasis and increasing allostatic load (WHO 2018; Basner et al., 2014; Muzet 2007). As Dreger et al. (2019) note, noise can contribute to social health inequalities through uneven distributions of exposure. Based on the concept that social circumstances impact where and how people can afford to live, socially deprived people are more likely in lower quality environments. This report aims to assess the relationship between noise from road traffic and indicators of social deprivation in Glasgow using a geographically weighted regression (GWR) analysis.

Some studies have found positive relationships between lower socioeconomic characteristics and exposure to noise. In Montreal, for example, Dale et al. (2015) found significant negative relationships between noise exposure and key indicators of socioeconomic deprivation, such as median household income, the proportion of people who spend over 30 per cent of their income on housing, and the proportion of people below a low income boundary. In the United Kingdom, a Birmingham study found weak but significant negative relationships between ethnicity, indicators of socioeconomic deprivation, and noise exposure (Brainard et al., 2004). In London, higher noise exposure was linked to unemployment, but it was dependent on the source of noise with train-related noise as the only significant source (Xie and Kang, 2010).

However, the relationship between noise exposure and lower socioeconomic status is not ubiquitously true. For example, in Paris and Marseilles, a higher socioeconomic status is linked with higher exposures to road traffic noise (Havard et al., 2011; Bocquier et al., 2012). Indeed, as meta-analyses of the literature suggest, while the relationship between deprivation and high noise levels generally is a positive one, it is extremely context-specific. Dreger et al. (2019) point towards several reasons for this disparity. In the United States, there is often an association between the city centre (which is usually louder) and low socioeconomic status, in part due to suburbanisation. However, in Europe, these same movements did not occur and the city centre is not necessarily associated with lower socioeconomic status. In Paris and Marseilles, noise is one factor of many when deciding on a place of residence; as Dreger et al. (2019) note, the most desirable places to live for commuting reasons may also be associated with higher noise levels. Moreover, Drager at al. (2019) argue a key factor in environmental health inequalities is vulnerability to negative health effects of noise, such as pre-existing health conditions. More affluent households may have access to resources to protect them from negative effects of noise exposure, such as better constructed housing.

Scotland’s draft National Planning Framework 4 (NPF4) Integrated Impact Assessment identifies that deprived communities are more exposed to higher levels of noise than those in less deprived areas. Indeed, as reviewed above, the literature on noise levels and deprivation suggest that there is a higher likelihood of occurrence, but these results are extremely dependent on the local context of the study areas. Leiper and Hood (2023), in analysing the relationship between noise and deprivation in four Scottish cities (Glasgow, Edinburgh, Aberdeen, and Dundee) find that with the exception of Glasgow, there is no significant relationship between noise and the Scottish Index of Multiple Deprivation (SIMD). However, by comparing the relationship between noise and deprivation *between* Scottish cities, Leiper and Hood (2023) do not address how the relationship between noise and deprivation varies *within* these cities. This report aims to address this gap by assessing the spatially variant impact of noise from road traffic on deprivation indicators within Glasgow using a Geographically Weighted Regression (GWR) analysis.

**2. Data and Study Area**

*2.1 Study Area*

Glasgow, as the largest city by population in Scotland, is a major transportation node. Unlike Edinburgh and Aberdeen, it has large motorways that cut through the city (M8, M73, M74, M77 and M80). It boasts extensive networks of bus routes and the UK’s second largest suburban commuter rail network. According to the Scottish Government’s Noise Action Plan (NAP) Report for Glasgow, it is one of four cities in Scotland (including Edinburgh, Aberdeen, and Dundee) required to develop a NAP after the adoption of the European Directive for Management of Environmental Noise 2002/49/EC legislation (commonly referred to as END). These NAPs required the establishment of Noise Management and Quiet Areas. Therefore, Glasgow was selected as the study area due to the recent focus on the impact of noise in Glasgow and its unique characteristics as a busy, industrial city.

|  |
| --- |
| A map of a city  Description automatically generated  Figure 1: Map of the Study Area with larger roads overlaid atop the datazones in Glasgow (n=954). Datazones are a spatial level of analysis that is equal to about 500 to 1000 persons per zone. This was selected to avoid biases involved in the measurement of noise on residential versus industrial or commercial areas. |

*2.2 Data*

The data on noise from traffic was generated on the researcher’s desktop using the method developed by Bocher et al. (2019). The NoiseModelling library is a java-based, open source tool to model noise propagation based on the French standard method for noise emissions from roads and the NMPB method for noise propagation, which considers the impact of buildings around roads. This is a two step process: first is the estimation of traffic noise emission over the road network and second is the calculation of sound levels over a grid of receivers. The data for the road network and the average daily traffic volume were obtained from OpenStreetMap. The buildings layer and the digital elevation model were obtained from the Urban Big Data Centre (UBDC).

|  |
| --- |
| A map of a city  Description automatically generated  Figure 2: Map displaying the results of the noise modelling clipped to the extent of the study area. Noise levels are measured in *Lden*, which is a logarithmic composite of the day, evening, and night noise levels but with 5 dB being added to the evening value and 10 dB being added to the night value. Decibels are units that measure the intensity of a sound on a logarithmic scale. Using zonal statistics in QGIS, a mean noise value was created for each datazone in Glasgow. |

|  |
| --- |
| Figure 3: Below is a map of the average *Lden* noise level in decibels for each datazone with equal interval classes (n=7)  A map of a city  Description automatically generated |

**3. Methodology**

This study employed a well-established spatial statistical data analysis methodology (Fotheringham, Kelly and Charlton, 2012) which consists of three steps: 1) a literature based variable selection, 2) data-driven model optimisation, and 3) calibration of the best global and local models. In step 1), 17 variables were selected as potential explanatory variables for measuring socioeconomic deprivation against noise pollution, and in step 2) the number of variables was reduced to eight based on a correlation analysis and a model quality optimisation. These eight variables were inputted into a global and GWR model in step 3). The GWR model was run in R studio using the *GWmodel* and *AICcmodavg* libraries.

*3.1  Literature-Based Variable Selection*

Based on the literature, 17 variables from the 2020 Scottish Index of Multiple Deprivation (SIMD) and the 2011 Census were most likely to be relevant measures of socioeconomic deprivation. The SIMD combines proxy indicators and weighs them to provide a ranking of deprivation. The categories and their weights for these variables are: income (28%) , employment (28%), education (14%), health (14%) , access to services (9%), crime (5%) and housing (2%). These same categories were used to group the variables selected. Similar to the study by Xie and Kang (2010), other variables from the 2011 Census were added to the model, such as the self-reported health of people, households with a person with a long-term health problem, and the percentage of unemployment.

*3.2 Step 1) literature-based variable selection*

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 1:** All 17 independent variables selected based on the literature and multicollinearity tests. Variables that were too highly correlated were left out of the GWR model, and the variable with a \* was further excluded after AICc optimisation. Bold indicates they were used in the GWR model. | | | |
| **Grouping** | **Name** | **Variable Description** | **Highly Correlated \*\* with** |
| 1 Housing | **HouseOCrat** | Percentage of people in households that are overcrowded |  |
| **HouseNCrat** | Percentage of people in households without central heating |  |
| 2 Education | **LowerEd\_Pct** | Percentage of adults over the age of 16 with an educational attainment level of an apprenticeship, upper school qualifications, and/or lower school qualifications | Negative\_Hlth, Pct\_LTH\_prob, EduAttain, EduAttend |
| EduAttain | Attainment of school leavers | LowerEd\_Pct |
| EduAttend | School pupil attendance | LowerEd\_Pct |
| 3 Crime | **CrimeRate** | Recorded crimes of violence, sexual offences, domestic housebreaking, vandalism, drugs offences, and common assault per 10,000 population |  |
| 4 Health | HlthAlcSR | Hospital stays related to alcohol use | IncRate, Vigintilv2, Pct\_Unemployed, LowerEd\_Pct, Pct\_LTH\_prob, Negative\_Hlth, EmpRate, HlthDprsPc, HlthCIF |
| HlthCIF | Comparative illness factor | IncRate, Vigintilv2, Pct\_Unemployed, Pct\_LTH\_prob, Negative\_Hlth, EmpRate, HlthDprsPc, HlthAlcSR |
| HlthDprsPc | Proportion of population being prescribed drugs for anxiety, depression or psychosis | IncRate, Vigintilv2, Pct\_Unemployed, LowerEd\_Pct, Pct\_LTH\_prob, Negative\_Hlth, EmpRate, HlthCIF |
| **HlthSMR** | Indirectly age-sex standardised ratio for deaths of all ages registered from all causes. |  |
| Negative\_Hlth\* | Percentage of people who report having “bad” or “very bad” health | HlthCIF, HlthDprsPc, Vigintilv2, Pct\_LTH\_prob, IncRate, EmpRate, LowerEd\_Pct, HlthAlcSR, HlthDprsPc, IncRate, EmpRate |
| **Pct\_LTH\_prob** | Percentage of households with one or more persons who have a long-term health problem or disability | IncRate, EmpRate, Negative\_Hlth, HlthCIF |
| 5 Income | IncRate | Percentage of people who are income deprived | EmpRate, Vigintilv2, Pct\_Unemployed, LowerEd\_Pct, Pct\_LTH\_prob, Negative\_Hlth, HlthDprsPc, HlthCIF, HlthAlcSR |
| 6  Employment | EmpRate | Percentage of working age people who are employment deprived | IncRate, Vigintilv2, Pct\_Unemployed, LowerEd\_Pct, Pct\_LTH\_prob, Negative\_Hlth, HlthDprsPc, HlthCIF, HlthAlcSR |
| **Pct\_Unemployed** | Percentage of households with no adults in employment | IncRate, EmpRate, HlthDprsPc, HlthCIF |
| 7  Access to services | GAccRank | Geographic access to services such as petrol stations, hospitals, schools, and retail centres domain rank |  |
| 8  Deprivation | **Vigintilv2** | The deprivation rank of datazones in Glasgow on a scale of 1 to 20, with 1 indicating the highest deprivation and 20 indicating the lowest | EmpRate, IncRate, HlthDprsPc, HlthCIF, Negative\_Hlth, HlthAlcSR |
| \*\*High correlation means the absolute value of the correlation coefficient between the two variables is more than 0.7  \*Indicates variable that was further excluded from the GWR model after the AICc Optimisation | | | |

*3.21 Group 1 Housing*

HouseOCrat and HouseNCrat address characteristics of housing that might leave socioeconomically deprived persons more vulnerable to noise pollution. This is an important factor Dreger et al. (2019) identify, and it is a consistent proxy for deprivation throughout past surveys in Scotland. Based on the literature, it is expected that overcrowding and homes with the worst heating infrastructure will be negatively correlated with noise levels.

*3.22 Group 2 Education*

Numerous authors have found opposing results and often non-significant ones in the relationship between educational attainment and noise exposure (Xie and Kang, 2010; Dreger et al., 2019). Rather than measuring material factors associated with where people can live, educational attainment tends to measure more behavioural factors, such as cultural norms around schooling. Thus, while it remains an important proxy of deprivation, it is not expected that there will be a strong relationship with noise levels. Moreover, according to the technical notes of the 2020 SIMD data and the household deprivation index metadata, educational attainment is measured as “low” with level 2 qualifications. To keep consistent with the data, LowerEd\_Pct was selected over EduAttain and EduAttend for this reason.

*3.23 Group 3 Crime*

In the SIMD, crime rate per 10,000 population is the only standardised measure. Crime rates are often used as an indicator in calculating deprivation indices (Kawachi, Kennedy and Wilkinson, 1999), and the dearth of highly correlated variables suggests an improvement on the model. However, like education, crime does not necessarily reflect material factors, and thus the relationship between crime and noise is expected to have a low, but positive correlation.

*3.24 Group 4 Health*

Out of the variables tested, HlthSMR, Negative\_Hlth, and Pct\_LTH\_prob had the fewest correlations with other variables. Specifically for noise pollution, as explored earlier in Section 1, there are both psychological and physiological impacts on health from noise suggesting a negative relationship.

*3.25 Group 5 Income*

Income deprivation is highly correlated with other variables and was excluded from the GWR model.

*3.26 Group 6 Employment*

Employment deprivation was highly correlated with other variables and removed. The percentage of unemployment may be a proxy for the ability to move from adverse living environments; it is expected there will be a negative relationship with employment and noise levels.

*3.27 Group 7 Access to Services*

While access to services is an important indicator for measuring deprivation, it was removed from the GWR model. This is due to the overlap between roads as noise emission sources and higher geographic access ranks.

*3.28 Group 8 Deprivation*

The deprivation rank (Vigintilv2) measures six other variables quite well and thus was left in the model. In section 1, the relationship between deprivation indices and noise levels was found to be context-specific, and it was selected considering the significant positive relationship between deprivation and noise levels that Leiper and Hood (2023) found in Glasgow

*3.3 Step 2) Model Optimisation*

17 variables were thus reduced to nine to be used in the GWR model optimisation. The methodology associated with step 2) involved running a stepwise-AIC (Akaike Information Criterion) procedure to assess the impact of each independent variable on the goodness-of-fit of the GWR model (Fotheringham, Kelly and Charlton, 2012).

|  |
| --- |
| A diagram of a diagram with colorful dots and numbers  Description automatically generated with medium confidence  Figure 4: Stepwise AICc radial model with variables ordered based on the least contribution to the AICc value. See the table above for a description of the variables in the legend. |

|  |
| --- |
| A graph of the selected variables  Description automatically generated  Figure 5: AICc optimisation with variables in order of their improvement to the AICc value. The first step is to consider, out of all the variables, which one fits the data best. This process is continued iteratively until there is no improvement in the AICc value. The cut-off value for the AICc optimisation is three; the difference between HlthSMR and Negative\_Hlth is -0.251. Therefore, Negative\_Hlth was removed and the remaining eight variables were used in the GWR model. |

*3.31 Spatial Autocorrelation Analysis*

To run a GWR model a neighbourhood conceptualisation is needed. Queen’s contiguity (first order), inverse distance-based weighing, and the 5-nearest neighbours (KNN) tested; rook’s case and distance band weighting were rejected *a priori* due to uneven datazone polygons in Glasgow.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 2:** Table Displaying the Results of Global Moran’s I for Each Weight | | | |
| Method | Global Moran’s I Index | Z Score | P-Value |
| Queen’s Order 1 | 0.606 | 31.0 | 0.001 |
| Inverse Distance Based Weighing (1550 m) | .353 | 45.45 | 0.001 |
| KNN (5 neighbours) | .558 | 29.15 | 0.001 |

According to Table 1, queen’s contiguity is most highly spatially autocorrelated (0.606). KNN skews the results because it selects random neighbours when there are fewer than five neighbours. Considering noise levels decrease significantly over distance and were measured with a max buffer of 250 metres, the high bandwidth of 1550 metres for inverse distance weight is much higher than the limit. Thus, due to the constraints of the noise model, defining the neighbourhood as edges and corners is most appropriate. A potential drawback is that the river Clyde bisects Glasgow and this is not addressed with the queen’s contiguity method.

Local Moran’s I was selected over Getis-Ord Gi\* for the local spatial autocorrelation method. This is because it identifies spatial outliers as well as clusters of high-high and low-low noise.

|  |
| --- |
| A map of a city with green and white squares  Description automatically generated  Figure 6: Significance map of global moran’s I of mean noise using queen’s contiguity. Significant clusters appear to be isolated to a pattern moving from central southwest Glasgow to the northeast, as well as some on the outskirts of the city towards the north, south, and east. Outside of the city centre but not on the outskirts noise is largely insignificant. |

|  |
| --- |
| A map of a city  Description automatically generated  Figure 7: Annotated cluster and outlier map of mean noise levels from road traffic using global Moran’s I with queen’s contiguity. Green lines indicate motorways (source OpenStreetMap). Less dense areas, as evidenced by the greater datazone area, are also typically clusters of low-low average noise. Denser areas in the city centre are more likely to be high-high clusters. There are also very few (n=11) datazones that are low values near high values and vice versa. |

Figures 6 and 7 reveal a general pattern of high-high clusters along the main motorway that bisects Glasgow. Less dense datazones on the outskirts of the city that do not have main motorways through them are often quieter.

Based on the result of the AICc model optimisation in Figure 5, a GWR model was calibrated using the remaining eight variables. The following general formula for local GWR models was used:

yᵢ = β₀(i)+β₁(i)Xᵢ₁+β₂(i)Xᵢ₂+...βₙ(i)Xᵢₙ

where  yᵢ is the dependent variable measured at location i,  Xᵢ are the explanatory variables, and β reflects the parameters that explain how a change in X affects y. Each value is dependent on the location of the local regression (Fotheringham, Kelly and Charlton, 2012, p. 230).

**4. Results**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table 3**: Results of the Global Regression and summary statistics for GWR parameter estimates. | | | | | | |
| Variable Name | Global Regression | | Local Regression | | | |
| Estimate | p-value | Min | Max | Mean | St Dev |
| HouseOCrat | 21.0 | **6.05e-06** | -28.9 | 60.0 | 12.5 | 18.4 |
| Pct\_LTH\_Prob | -2.21 | 0.723 | -56.6 | 79.5 | -3.51 | 21.6 |
| CrimeRate | 0.000237 | 0.641 | -0.0305 | 0.0110 | 0.00117 | 0.00439 |
| LowerEd\_Pct | -5.69 | 0.105 | -82.8 | 53.6 | -2.83 | 17.9 |
| Vigintilv2 | -0.139 | 0.200 | -1.44 | 0.768 | 0.0125 | 0.331 |
| HouseNCrat | 12.2 | 0.321 | -96.5 | 257 | 3.69 | 53.8 |
| Pct\_Unemployed | -1.26 | 0.789 | -61.4 | 41.8 | 2.18 | 16.9 |
| HlthSMR | 0.00185 | 0.693 | -0.0670 | 0.0919 | -0.000752 | 0.0204 |

*4.1 Global Regression Model*

The global regression indicated only overcrowding was a globally positive significant variable with average noise levels. This supports the hypothesis made in section 3.21. The relationships were aligned with the hypotheses for Pct\_Unemployed, Vigintilv2, the housing variables, and crime (see section 3.2). However, the variables measuring health and education did not fully support the hypotheses. Notably, the parameter estimate values in Table 3 have maximum positive and minimum negative values; this suggests the relationship between noise levels and these independent variables is negative in some places and positive in others.

*4.2 GWR Model*

The GWR model was run using a bi-square (Gaussian) kernel with an adaptive bandwidth, which resulted in an optimal bandwidth of 138 nearest neighbours out of 954 total, about 14%. The bi-square kernel was selected because of the decreasing weight of observations with distance, and the adaptive bandwidth enables the analysis of an unevenly distributed dataset. Furthermore, between the global and local regression, the r-squared value increased from 0.0712 to 0.395 and the AICc decreased from 6904 to 6610, a significant improvement.

|  |
| --- |
| A screenshot of a map  Description automatically generated  Figure 8: Maps displaying the positive or negative GWR local parameter estimates for a) crime rates per 10,000 people, b) the percentage of people with lower educational attainment (<level 2), c) the percentage of households with no adults in employment, and d) deprivation rank. Non-significant areas are greyed out. |

|  |
| --- |
| A screenshot of a map  Description automatically generated  Figure 9: Maps displaying the positive or negative GWR local parameter estimates for a) overcrowding, b) households without central heating, c) the percentage of households where one or more persons has a long-term health problem or disability, and d) mortality rates. Non-significant areas are greyed out. |

Figures 8a, b, and c are significant in the outskirts of Glasgow, with positive relationships between crime, noise, and lower educational attainment in the northeast. Figure 8d displays a positive relationship between higher deprivation and higher noise exposure along the M8 and M77 motorway. This supports the hypothesis made in section 3.2. However, in most datazones this relationship is not significant.

Figure 9a and b provide insight into the relationship between housing and noise levels. In the periphery there are strong positive relationships between noise levels and overcrowding (Fig 9a), but the negative cluster in the centre indicates that noise levels are not uniformly related to worse housing. This is further supported by the insignificant relationships in Figure 9b. Figures 9c and d suggest there is an inconsistent relationship between poor health and noise, but the clusters of negative relationships in Fig 9c indicate that noise is negatively associated with health issues south of the Clyde.

|  |
| --- |
| A map of a city  Description automatically generated  Figure 10: Map of Local R squared values in Glasgow. Model performs the best in the outskirts to the north and south of the city as well as just south of the city centre across the river Clyde. The model fits poorly in the southeast, southwest, and northwest of the city. |

The centre and north/south peripheries of Glasgow had high r-squared values (>0.6), but the model fits worse in most other areas (Figure 10). Notable is the gradual decline in explanatory power further from the centre. The strength of the model is not strongly related to the major motorways in Glasgow.

|  |
| --- |
| A screenshot of a map  Description automatically generated  Figure 11: Map of a) standardised global residuals from the global model and b) standardised local residuals from the GWR model. Residuals are defined as the difference between the observed Y value and the predicted value. In a GWR model, the local residuals are expected to exhibit a more spatially random distribution than the global model (Mansley and Demšar, 2015). |

On visual inspection, it is difficult to see a difference between Figure 11a and b. However, Table 4 confirms the GWR residuals are more randomly distributed. Noise levels are clearly a non-stationary phenomenon and this GWR analysis reveals there is a notable relationship between indicators of deprivation and noise exposure due to the lower Moran’s I Index value.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 4:** Assessing the level of spatial autocorrelation in standardised global and local residuals using Moran’s I Index with queen’s contiguity. Moran’s I values are higher and positive for the global model and decrease somewhat for the local model, indicating that the GWR model resulted in a more random spatial pattern. | | | |
| Residuals | Moran's I Index | Z-Value | P-Value |
| Global | 0.594 | 28.49 | 0.001 |
| Local | 0.307 | 15.8 | 0.001 |

**5. Conclusions and Discussion**

This analysis considered the explanatory significance of socioeconomic indicators on exposure to noise levels from traffic in Glasgow using a GWR. While this method has been employed in other contexts, to date there is little to no research conducted using this method in Scottish cities. This model explains up to 79 per cent of the variation in noise levels using a set of socioeconomic variables, but it performs less well in the peripheries towards the southeast, southwest, and northwest (Fig 10). The first insight gleaned from the analysis was the selection of variables in the AICc optimisation (Fig 5) and the collinearity analysis (Table 1). Contrary to the findings of Xie and Kang (2010) in London, income was not a unique or significant explanatory variable of exposure to noise levels. The findings of this study also contradict the literature in that the reporting of negative health did not significantly improve the model fit for noise levels.

Another interesting result is the strong relationship between housing characteristics and noise levels. This was both supported in the literature and found to hold true in Glasgow. However, in Figure 9a, overcrowding had a bipolar relationship to high noise levels. In the centre of the city, where the model fits well, there is a negative relationship between overcrowding and high noise levels, contrary to the hypothesis made in section 3.2. This could suggest that noise is indeed one factor of many when it comes to deciding where to live; as Dreger et al. (2019) find, affluent people – who are less likely to live in overcrowded homes – are possibly prioritising commute or location to services over quietness. Moreover, it suggests that with space being such a premium, this measure of overcrowding may not accurately reflect deprivation levels in the centre of Glasgow. Nevertheless, the positive relationship between overcrowding and high noise is globally positively significant.

Furthermore, the percentage of crime in Figure 8a does not have a strong relationship with noise levels. The positive relationship with noise levels from traffic could instead be due to the role of motorways in facilitating crime reporting and occurrence, for example.

Lower educational attainment, the percentage of unemployment, and overcrowding (Fig 8b, 8c, and 9a) is also interesting when read together; there is a cluster of negative relationships with these variables and high noise levels in an area south of the Clyde in central Glasgow. This suggests that more deprived populations are clustered in this intersection of the M77 and M8, and thus are exposed to more noise pollution. Indeed, the combined indicator of deprivation (Fig 8d) supports the conclusion that deprivation and noise is clustered along the M8 and M77. Moreover, the unipolar distribution of deprivation (Fig 8d) indicates that there is not a strong relationship between high socioeconomic status and higher noise levels. This is contrary to the findings of Bocquier et al. (2012) and Havard et al. (2011) in Marseilles and Paris respectively. It also indicates that the statement in Scotland’s draft NPF4 relating higher deprivation to higher noise levels is supported by the evidence in at least one area of Glasgow. This finding thus can guide Glasgow's NAP by suggesting that this area of Glasgow should be considered as a candidate for a Noise Management area.

The insights into the relationship between health and noise levels are relatively limited. The strong positive and negative clusters in Figure 9c oppose the hypothesis made in section 3.2; health problems are less likely in the louder centre of the city and more likely in the outskirts. In part, this could be due to demographic differences between the centre and the periphery; younger people with fewer health issues overall might be drawn to the commercial or financial centre of Glasgow and thus are exposed to higher noise levels, for example. Mortality is not significantly related to noise levels (Fig 9d).

However, the average noise level (Lden) included in this study does not measure multiple exposures at home, at work, and during leisure time activities. It also does not incorporate when and how much time people spend at home, which differs greatly between social groups. This study measured only noise emissions from vehicle traffic; air and rail noise emissions were excluded. Moreover, the data used for the socioeconomic indicators was compiled from different censuses. The timing of the 2011 census is particularly notable; these data were collected prior to the adoption of the NAP in 2014, meaning that any action to mitigate noise pollution was not fully captured.

While Table 4 shows the Moran’s I index decreased by 0.3 between the two models, there is still spatial autocorrelation present in the GWR model. This suggests that the socioeconomic variables chosen do not explain the distribution of noise levels completely. The implications are twofold. First, it suggests that other socioeconomic indicators should be selected for future studies. Second, it reveals that not all spatial variation in noise can be explained by socioeconomic characteristics; noise from traffic is dependent on infrastructure that can be entirely unrelated to deprivation indicators.

The major contribution of this analysis is the insight into the spatial variability of the impact of deprivation indicators on noise levels as seen in Figure 10. If this is applied to other cities in Scotland, it can provide a strong justification for selecting candidates for Noise Management Areas in accordance with the adopted environmental noise directives (see section 2.1). Additionally, Leiper and Hood (2023) find that there is no globally significant relationship between noise and deprivation for Dundee, Aberdeen, and Edinburgh. This study complicates their conclusions because it suggests that the level of analysis for deprivation and noise needs to be finer, at an intra-city level, as indicated by the map of local r-squared values (Figure 10). This study found that deprivation does explain some of the spatial variability of high noise levels from road traffic in Glasgow and was most significantly positively correlated in the centre of Glasgow. The variables chosen and their relationships (or lack thereof) with noise can also serve to inform a set of hypotheses for further studies that can accommodate noise from rail and air traffic sources, for example. It finally supports the conclusion that the relationship between noise and deprivation is context-specific and should be investigated accordingly.

**Works Cited**

Basner, M., Babisch, W., Davis, A., Brink, M., Clark, C., Janssen, S. and Stansfeld, S. (2014). Auditory and non-auditory effects of noise on health. *The Lancet*, [online] 383(9925), pp.1325–1332. doi:<https://doi.org/10.1016/s0140-6736(13)61613-x>.

Bocher, E., Guillaume, G., Picaut, J., Petit, G. and Fortin, N. (2019). NoiseModelling: An Open Source GIS Based Tool to Produce Environmental Noise Maps. *ISPRS International Journal of Geo-Information*, 8(3), p.130. doi:<https://doi.org/10.3390/ijgi8030130>.

Bocquier, A., Cortaredona, S., Boutin, C., David, A., Bigot, A., Chaix, B., Gaudart, J. and Verger, P. (2012). Small-area analysis of social inequalities in residential exposure to road traffic noise in Marseilles, France. *European Journal of Public Health*, 23(4), pp.540–546. doi:<https://doi.org/10.1093/eurpub/cks059>.

Brainard, J.S., Jones, A.P., Bateman, I.J. and Lovett, A.A. (2004). Exposure to Environmental Urban Noise Pollution in Birmingham, UK. *Urban Studies*, 41(13), pp.2581–2600. doi:<https://doi.org/10.1080/0042098042000294574>.

Dale, L.M., Goudreau, S., Perron, S., Ragettli, M.S., Hatzopoulou, M. and Smargiassi, A. (2015). Socioeconomic status and environmental noise exposure in Montreal, Canada. *BMC Public Health*, [online] 15(1). doi:<https://doi.org/10.1186/s12889-015-1571-2>.

Dreger, S., Schüle, S., Hilz, L. and Bolte, G. (2019). Social Inequalities in Environmental Noise Exposure: A Review of Evidence in the WHO European Region. *International Journal of Environmental Research and Public Health*, 16(6), p.1011. doi:<https://doi.org/10.3390/ijerph16061011>.

Fotheringham, A.S., Kelly, M.H. and Charlton, M. (2012). The Demographic Impacts of the Irish famine: Towards a Greater Geographical Understanding. *Transactions of the Institute of British Geographers*, 38(2), pp.221–237. doi:<https://doi.org/10.1111/j.1475-5661.2012.00517.x>.

Havard, S., Reich, B.J., Bean, K. and Chaix, B. (2011). Social inequalities in residential exposure to road traffic noise: An environmental justice analysis based on the RECORD Cohort Study. *Occupational and Environmental Medicine*, [online] 68(5), pp.366–374. doi:<https://doi.org/10.1136/oem.2010.060640>.

Kawachi, I., Kennedy, B.P. and Wilkinson, R.G. (1999). Crime: social disorganization and relative deprivation. *Social Science & Medicine*, [online] 48(6), pp.719–731. doi:<https://doi.org/10.1016/s0277-9536(98)00400-6>.

Leiper, A. and Hood, A. (2023). Noise and Deprivation in Scotland’s Four Largest Cities: Glasgow, Edinburgh, Aberdeen and Dundee. *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, 265(5), pp.2980–2989. doi:<https://doi.org/10.3397/in_2022_0419>.

Mansley, E. and Demšar, U. (2015). Space matters: Geographic variability of electoral turnout determinants in the 2012 London mayoral election. *Electoral Studies*, 40, pp.322–334. doi:<https://doi.org/10.1016/j.electstud.2015.10.003>.

Muzet, A. (2007). Environmental noise, sleep and health. *Sleep Medicine Reviews*, 11(2), pp.135–142. doi:<https://doi.org/10.1016/j.smrv.2006.09.001>.

Scottish Government (2014). *Glasgow Agglomeration: Noise Action Plan*. [online] Scottish Government, pp.1–33. Available at: <https://www.gov.scot/publications/glasgow-agglomeration-noise-action-plan/documents/> [Accessed 27 Nov. 2024].

Scottish Government (2021). *Scotland 2045 - Our Fourth National Planning Framework: Draft*. Edinburgh.

World Health Organsation (2018). *Environmental noise guidelines for the European Region*. [online] Regional Office For Europe. Available at: <https://iris.who.int/handle/10665/279952> [Accessed 27 Nov. 2024].

Xie, H. and Kang, J. (2010). On the Relationships Between Environmental Noise and Socio-Economic Factors in Greater London. *Acta Acustica united with Acustica*, [online] 96(3), pp.472–481. doi:<https://doi.org/10.3813/aaa.918300>.

**Appendix for R Code**

|  |
| --- |
| ##-----Code for Final Lab--------- # 200024980 # Reading the necessary libraries library(sf) library(tmap) library(RColorBrewer) library(grid) library(gridExtra) library(GWmodel) library(AICcmodavg)  #----------Importing the data previously cleaned in QGIS---------- dataGWR <- sf::read\_sf("Lden\_ScotGov\_SIMD\_Roads.shp") # Reading as an sf  colnames(dataGWR)  # Plotting the average noise levels for each datazone with equal intervals  tm\_shape(dataGWR) + tm\_fill("\_mean", palette = "Reds", style = "equal", n = 7)  head(dataGWR)  # Exploring the data  head(dataGWR$DataZone) # Datazones are lines drawn where number of people is between 500 and 1000  #-------Cleaning the general health data from the 2011 census-------  # Reading General Health CSV general\_health <- read.csv("health\_census\_2011.csv", header = TRUE, skip = 10)# skipping rows because there is some metadata at first   str(general\_health) # Checking data types   # converting all people from chr to int  general\_health$All.people <- as.integer(general\_health$All.people)  sum(is.na(general\_health$All.people)) # Looks like there are 8 datazones where we dont have data for the health  # Calculate percentages for 'Bad' and 'Very.bad' by dividing bad health by the total  general\_health$percent\_Bad <- general\_health$Bad / general\_health$All.people general\_health$percent\_VeryBad <- general\_health$Very.bad / general\_health$All.people  sum(is.na(general\_health$percent\_VeryBad)) # Just checking, not too many   head(general\_health) # Subset the data --> weirdly general health is the title of the datazones column, so will need to rename that soon  general\_health <- general\_health[, c("General.health", "percent\_Bad", "percent\_VeryBad")]  # now lets combine percent of bad and very bad health to get a negative health rating  # Create a new column 'percent\_NegativeHealth' that sums 'percent\_Bad' and 'percent\_VeryBad' general\_health$percent\_NegativeHealth <- general\_health$percent\_Bad + general\_health$percent\_VeryBad # Check the result head(general\_health) # Rename column "General health" to "DataZone" colnames(general\_health)[colnames(general\_health) == "General.health"] <- "DataZone" colnames(general\_health)[colnames(general\_health) == "percent\_Bad"] <- "Bad\_Hlth" colnames(general\_health)[colnames(general\_health) == "percent\_VeryBad"] <- "VeryBad\_Hlth" colnames(general\_health)[colnames(general\_health) == "percent\_NegativeHealth"] <- "Negative\_Hlth"  #----------Merge the shapefile with the health data------------ dataGWR <- merge(dataGWR, general\_health, by = "DataZone")  head(dataGWR)  sum(is.na(dataGWR$Negative\_Hlth)) # Okay, so 3 na values, we will  # just remove these datazones later on, because 3 is hardly significant out of around 1000   tm\_shape(dataGWR) + tm\_fill("Negative\_Hlth", palette = "Reds", na.color = "blue")  # We can see some of the NA values in the map, again, not that important considering size of the data   #--------Cleaning the Employment / Dependency Data from 2011 Census-------------  #First step is to read it  emp\_dep\_csv <- read.csv("employment\_and\_dependency\_2011\_census.csv", header = TRUE, skip = 10)  sum(is.na(emp\_dep\_csv$All.households)) names(emp\_dep\_csv)  # Get the sum of unemployed adults and the sum of people with long term health problems  emp\_dep\_csv$Unemployed <- emp\_dep\_csv$No.adults.in.employment.in.household..With.dependent.children + emp\_dep\_csv$No.adults.in.employment.in.household..No.dependent.children  emp\_dep\_csv$LTH\_prob <- emp\_dep\_csv$One.or.more.persons.in.household.with.a.long.term.health.problem.or.disability..With.dependent.children + emp\_dep\_csv$One.or.more.persons.in.household.with.a.long.term.health.problem.or.disability..No.dependent.children  # Calculate percentages against the total  emp\_dep\_csv$Pct\_Unemployed <- emp\_dep\_csv$Unemployed / emp\_dep\_csv$All.households emp\_dep\_csv$Pct\_LTH\_prob <- emp\_dep\_csv$LTH\_prob / emp\_dep\_csv$All.households   emp\_dep\_csv <- na.omit(emp\_dep\_csv)  summary(emp\_dep\_csv) #Renaming columns  colnames(emp\_dep\_csv)[colnames(emp\_dep\_csv) == "Households.with.no.adults.in.employment.and.dependent.children.status.then.Dependent.children.in.household.then.Dependent.children.in.household.then.Persons.in.household.with.long.term.health.problem.or.disability.and.dependent.chilfren.status.then.Household.size"] <- "DataZone" #Subsetting emp\_dep\_csv <- emp\_dep\_csv[, c("DataZone", "Pct\_Unemployed", "Pct\_LTH\_prob")]  head(emp\_dep\_csv)  #----------Adding Employment/Health Problems data to the shapefile-------------------- dataGWR <- merge(dataGWR, emp\_dep\_csv, by = "DataZone")  colnames(dataGWR)   #--------Cleaning the Education Data from 2011 Census-------------- # While the SIMD data does have indicators for education, such as EduNoQual, I decided to also calculate this based on the metadata for how they caluclated household deprivation stats in the 2011 census # This means measuring lower educational attainment as below a level 2 qualification  # https://www.scotlandscensus.gov.uk/metadata/household-deprivation-classification/#panel-1 educsv <- read.csv("Education\_Census\_2011.csv", header = TRUE, skip = 10)  head(educsv)  # Calculating the Lower Education indicator by adding together the lower school quals, upper school quals, and apprenticeship quals  # This excluded those have a college degree or other higher education  educsv$LowerEd <- educsv$No.qualifications + educsv$Lower.school.qualifications + educsv$Upper.school.qualifications +educsv$Apprenticeship.qualifications  # calculating the percentage of this against the total no of people educsv$All.people.aged.16.and.over <- as.integer(educsv$All.people.aged.16.and.over) educsv$LowerEd\_Pct <- educsv$LowerEd / educsv$All.people.aged.16.and.over  # Checking for NAs  sum(is.na(educsv$LowerEd\_Pct)) educsv <- na.omit(educsv) sum(is.na(educsv$LowerEd\_Pct))  min(educsv$LowerEd\_Pct) # This is good that it is a value of 0.13; it means that all na values are cleaned and we would expect the minimum to be around here   # Subsetting and renaming the data  head(educsv)  # renaming the messy title of datazone to then merge later on  colnames(educsv)[colnames(educsv) == "Highest.level.of.qualification"] <- "DataZone"  educsv <- educsv[, c("DataZone", "LowerEd\_Pct")]  #----------Adding educsv to the shapefile---------------  dataGWR <- merge(dataGWR, educsv, by = "DataZone")  colnames(dataGWR)  #-------Subsetting the SIMD and the Extra Data We Just Cleaned From 2011 Census----------  # Subset columns from dataGWR with the variables from SIMD and the census, identified from the literature that might be relevant  dataGWR <- dataGWR[, c("DataZone", "Vigintilv2", "IncRate", "EmpRate", "HlthDprsPc",   "HlthAlcSR", "HlthSMR", "HlthCIF", "EduAttend", "EduAttain",   "CrimeRate", "HouseOCrat", "HouseNCrat", "Negative\_Hlth", "Pct\_Unemployed", "Pct\_LTH\_prob", "LowerEd\_Pct", "geometry",   "\_mean", "\_stdev", "\_min", "\_max", "Shape\_Leng", "Shape\_Area", "EduNoQuals",  "GAccRank")]   # Renaming the noise columns  colnames(dataGWR)[colnames(dataGWR) == "\_mean"] <- "mean\_noise" colnames(dataGWR)[colnames(dataGWR) == "\_stdev"] <- "stdev\_noise" colnames(dataGWR)[colnames(dataGWR) == "\_min"] <- "min\_noise" colnames(dataGWR)[colnames(dataGWR) == "\_max"] <- "max\_noise"  colnames(dataGWR)  summary(dataGWR)  # Let's have a quick look at the data # Standard Deviation of the noise levels  tm\_shape(dataGWR) + tm\_fill("Vigintilv2", palette = "Reds")  # Mean noise Map tm\_shape(dataGWR) + tm\_fill("mean\_noise", palette = "Reds")  # Percent Unemployed  tm\_shape(dataGWR) + tm\_fill("Pct\_Unemployed", palette = "Reds")   dev.off()  #-----Cleaning the Data, Converting to Percentages ----------  str(dataGWR) # So clearly, the data from the SIMD is not all an integer or percentage like I need it to be.  #Instead, some of the percentages have the symbol next to the number (%2), when I need it to be 0.02.   # I found the documentation for using gsub here  # https://www.statology.org/gsub-r/ # Basically, I wanted to search the chr for % and replace it with nothing (""), and then divide the number that remained by 100   dataGWR$IncRate <- as.numeric(gsub("%", "", dataGWR$IncRate)) / 100 dataGWR$EmpRate <- as.numeric(gsub("%", "", dataGWR$EmpRate)) / 100 dataGWR$HlthDprsPc <- as.numeric(gsub("%", "", dataGWR$HlthDprsPc)) / 100 dataGWR$EduAttend <- as.numeric(gsub("%", "", dataGWR$EduAttend)) / 100 dataGWR$HouseOCrat <- as.numeric(gsub("%", "", dataGWR$HouseOCrat)) / 100 dataGWR$HouseNCrat <- as.numeric(gsub("%", "", dataGWR$HouseNCrat)) / 100  # Check the structure of the modified columns str(dataGWR)  dataGWR <- na.omit(dataGWR)  sum(is.na(dataGWR)) # great, there are 0 left   #--------Correlation Matrix for the Identified Variables-------------------  # First I need to convert dataGWR into a dataframe, not a spatial dataframe   dataGWR\_df <- as.data.frame(dataGWR) # Converting this to a dataframe to avoid issues with the geometry column  names(dataGWR\_df) # Now I can run the correlation test with all the potential variables  correlation\_matrix <- cor(dataGWR\_df[, c("mean\_noise", "IncRate","EmpRate", "HlthDprsPc", "HlthAlcSR" , "HlthSMR", "HlthCIF", "Negative\_Hlth",   "HouseOCrat","Pct\_LTH\_prob" , "EduAttain" ,"EduAttend", "CrimeRate",  "HouseNCrat","LowerEd\_Pct" , "Pct\_Unemployed", "Vigintilv2", "GAccRank" )])  print(correlation\_matrix)  # I want to be able to highlight numbers that are above .7 # As such, I can visualise this easily in Excel # Here is the documentation for where I found the library to download the xlsx package  # https://cran.r-project.org/web/packages/xlsx/readme/README.html  #install.packages("openxlsx") library(openxlsx) write.xlsx(correlation\_matrix, file = "correlation\_matrix\_test.xlsx", rowNames = TRUE)  #--------Variables selected from the correlation analysis------------- # HlthSMR # Negative\_Hlth # HouseOCrat # Pct\_LTH\_prob # CrimeRate # HouseNCrat # LowerEd\_Pct # Pct\_Unemployed # Vigintilv2  #----------Mapping the dependent variable----------  # Mean Noise is the dependent variable colnames(dataGWR)  tm\_shape(dataGWR) + tm\_fill("mean\_noise", palette = "Reds", n = 7, style = "equal")  #When using GeoDa to test the different neighborhood definitions,  # I know I'm going to need a Poly\_ID - a unique values for each entry in the datazone  # As such, I used seq\_len function to do this easily in r, as demonstrated here:  # https://www.rpubs.com/Mentors\_Ubiqum/seq\_len  dataGWR$POLY\_ID <- seq\_len(nrow(dataGWR))# This creates a sequence that starts at 1 and ends at a specified value -I set this specified value to the number of rows in dataGWR  # Did this work?  colnames(dataGWR) # Great, now we can save to bring into GeoDa st\_write(dataGWR, "Lden\_ScotGov\_SIMD\_Combined\_PolyID.shp", append=FALSE)# Now I am ready to calculate weights and conduct the global spatial autocorrelation analysis   #------------AICc Optimisation-----------------  # First, as with any GWmodel call, we need to convert data into a spatial data frame dataGWRspatial <- as\_Spatial(dataGWR) names(dataGWR)  DeVar <- "mean\_noise"  # Again, here are the variables we identified  # HlthSMR # Negative\_Hlth # HouseOCrat # Pct\_LTH\_prob # CrimeRate # HouseNCrat # LowerEd\_Pct # Pct\_Unemployed # Vigintilv2  # Defining independent variables  InDeVar <- c( "Vigintilv2","Negative\_Hlth", "CrimeRate", "HouseOCrat",   "HouseNCrat","Pct\_Unemployed", "Pct\_LTH\_prob", "LowerEd\_Pct",  "HlthSMR")  # Run GWR with the specified independent variables optimalBW <- bw.gwr(mean\_noise ~ Vigintilv2+ Negative\_Hlth +   CrimeRate + HouseOCrat +   HouseNCrat+ Pct\_Unemployed + Pct\_LTH\_prob + LowerEd\_Pct+HlthSMR,   data = dataGWRspatial,   approach = "AICc",   kernel = "bisquare",   adaptive = TRUE)  # Approach is AICc, kernel is bisquare, adaptive is true  # KNN = 138, AICc is 6526.232 # Running the Model Selection  modelSel <- model.selection.gwr(DeVar, InDeVar, data=dataGWRspatial, kernel="bisquare", adaptive=TRUE, bw=optimalBW)  # Extract list of models from the results, this creates a list of the order in which they were generated sortedModels <- model.sort.gwr(modelSel, numVars <- length(InDeVar), ruler.vector = modelSel[[2]][,2]) modelList <- sortedModels[[1]]  model.view.gwr(DeVar, InDeVar, model.list=modelList) # Viewing the radial model   # Export in a figure #png(filename="ModelSelection\_RadialView\_Final.png", width=800, height = 800) #model.view.gwr(DeVar, InDeVar, model.list=modelList) dev.off()  #----------1.3 Determining the impact of variables on AICc---------------------------------------  #Prepping the graph that shows the variables with the lowest impact on AICc  # number of independent variables n <- length(InDeVar)  # Export list of AICc values from the sorted models AICcList <- sortedModels[[2]][,2]  indices <- rep(n, n) # initialise a list of indicies as n values of n, the first position is already correct  # Based on the lab: for each position we will take the number from previous step (i-1) and add (n-i), then correct  # by adding another 1, because we started at i=2 for (i in 2:n) {  indices[i]=indices[i-1]+((n-i)+1) } # Checking what this looks like, seems to match with radial plot:  indices  # Now let's find AICc values for models at these indices AICcBestModelValues <- AICcList[indices]  AICcBestModelValues # Result: [1] 6405.122 6379.580 6354.013 6339.046 6322.967 6313.754 6306.834 6298.617 6294.297 6285.674  # To be able to plot the AICc optimisation plot, we need to find out which variable was the one that was added to each best model - we do this by selecting model descriptions from the model selection result. BestModels <- sortedModels[[1]][indices] BestModels  # The last model has variables listed in the order of addition, and we will need this for our AICc plot. BestModels[n]  # With the correct order  variablesAsAdded <- c("HouseOCrat", "Pct\_LTH\_prob" , "CrimeRate" , "LowerEd\_Pct","Vigintilv2", "HouseNCrat" , "Pct\_Unemployed", "HlthSMR", "Negative\_Hlth")  #-----------1.31 Plotting the graph with the AICc impact--------------------------------- par(mar = c(8, 4, 4, 2)) # Increase bottom margin (first value) # Documentation for this margin increase was found here:  # https://www.r-bloggers.com/2010/06/setting-graph-margins-in-r-using-the-par-function-and-lots-of-cow-milk/ plot(cbind(1:9,AICcBestModelValues), col = "black", pch = 20, lty = 10,   main = "AICc optimisation", ylab = "AICc", type = "b", axes=FALSE) par(las=2) # This will rotate labels on x axis for 90 degrees axis(1, at=1:9, labels=variablesAsAdded) # this plots variable names as labels on x axis, changed to 9 axis(2, at=NULL, labels=TRUE) # this plots numbers on y axis dev.off()   # And export the plot as a figure; I left this hashed out on subsequent runs of the data  # png(filename="AICcOptimisationFinal.png", width=1024) # plot(cbind(1:9,AICcBestModelValues), col = "black", pch = 20, lty = 5,  # main = "AICc optimisation", ylab = "AICc", type = "b", axes=FALSE) # par(las=2) # This will rotate labels on x axis for 90 degrees # axis(1, at=1:9, labels=variablesAsAdded) # this plots variable names as labels on x axis # axis(2, at=NULL, labels=TRUE) # this plots numbers on y axisdev.off() # dev.off() #------------1.32 Calculating the AICc differences between the variables---------------------- #Difference between two consecutive AICc: Takes the first n-1 elements (1:(n-1)) and the last n-1 elements (2:n) and subtracts the second list from the first list AICcDifference <- AICcBestModelValues[1:(n-1)]-AICcBestModelValues[2:n] # Check how this looks AICcDifference #Result:  #[1] 25.6252613 28.5750158 15.4282422 13.1605979 9.5677228 7.0950348 4.7815375 -0.2510507 # This means that the last variable should be removed because it does not improve the model fit, the AICc is less than 3  #----------------2.1 Creating the new optimised GWR model using 8 variables------------------ # Optimised bandwidth: we take the bisquare adaptive kernel and use AICc for identification of the optimal bandwith optimalBW7 <- bw.gwr(mean\_noise ~ HouseOCrat+Pct\_LTH\_prob+CrimeRate+LowerEd\_Pct+Vigintilv2+HouseNCrat+Pct\_Unemployed+HlthSMR, data=dataGWRspatial, approach="AICc", kernel="bisquare", adaptive=TRUE)  # Run the GWR model (basic) gwrmodel <- gwr.basic(mean\_noise ~ HouseOCrat+Pct\_LTH\_prob+CrimeRate+LowerEd\_Pct+Vigintilv2+HouseNCrat+Pct\_Unemployed+HlthSMR, data=dataGWRspatial, bw=optimalBW7, kernel="bisquare", adaptive=TRUE)  # As a note, it is not necessary to run the global linear model: the GWR result below gives both the local and global gwr results.   print(gwrmodel) capture.output(gwrmodel, file="GWRmodel\_descriptiveResult.txt", append = TRUE)  #----------------2.2 Calculating Global Standardised Residuals ----------------------------------  #----------------2.21 Calculate global residuals in three steps--------------------------------  # Step 1: calculate predicted mean noise  # Here is the actual equation with the values taken from the results of the GWR  dataGWR$predictedmean\_noise <- 52.1872976 + 21.0141953 \* dataGWR$HouseOCrat + -2.2071603 \* dataGWR$Pct\_LTH\_prob + 0.0002369 \* dataGWR$CrimeRate + -5.6855664 \* dataGWR$LowerEd\_Pct + -0.1387607 \* dataGWR$Vigintilv2 + 12.1745695\* dataGWR$HouseNCrat + -1.2567261\* dataGWR$Pct\_Unemployed + 0.0018505\* dataGWR$HlthSMR  # Check what this did: head(dataGWR)  # Step 2: Calculate global residuals by subtracting the predicted value from the actual value  dataGWR$globalRes <- dataGWR$mean\_noise - dataGWR$predictedmean\_noise # Check what this did: head(dataGWR)  #Step 3: Rescale the global residuals to the 0-1 range using mean and sd  m <- mean(dataGWR$globalRes) sd <- sd(dataGWR$globalRes)  #Calculating standardised global residuals  dataGWR$stGlobalRes <- (dataGWR$globalRes-m)/sd  #------------------2.3 Creating the map of GWR by joining with the data-------------------------- # Turning the gwrmodel into a dataframe  results <- as.data.frame(gwrmodel$SDF)  names(results) head(results)  #To get the stats on the local regression summary(results)  # Appending the GWR data and the map data and then renaming variables  mapGWR <- cbind(dataGWR, as.matrix(results)) head(mapGWR) names(mapGWR)  # Renaming them from .1 to \_beta names(mapGWR)[which(names(mapGWR)=="Intercept")] <- "Intercept\_beta" names(mapGWR)[which(names(mapGWR)=="HouseOCrat.1")] <- "HouseOCrat\_beta" names(mapGWR)[which(names(mapGWR)=="CrimeRate.1")] <- "CrimeRate\_beta" names(mapGWR)[which(names(mapGWR)=="LowerEd\_Pct.1")] <- "LowerEd\_Pct\_beta" names(mapGWR)[which(names(mapGWR)=="HouseNCrat.1")] <- "HouseNCrat\_beta" names(mapGWR)[which(names(mapGWR)=="Pct\_Unemployed.1")] <- "Pct\_Unemployed\_beta" names(mapGWR)[which(names(mapGWR)=="Pct\_LTH\_prob.1")] <- "Pct\_LTH\_prob\_beta" names(mapGWR)[which(names(mapGWR)=="HlthSMR.1")] <- "HlthSMR\_beta" names(mapGWR)[which(names(mapGWR)=="Vigintilv2.1")] <- "Vigintilv2\_beta"  names(mapGWR)  #------------------------2.31 Setting Up the Bounding Boxes--------------------------------------  # Get the bounding boxes of the parameter est maps  bbox1 <- st\_bbox(mapGWR) # Range of x and y values xrange <- bbox1$xmax - bbox1$xmin # range of x values yrange <- bbox1$ymax - bbox1$ymin # range of y values # Extend the right dimension by 25% more space bbox1[3] <- bbox1[3] + (0.25 \* xrange) # Extend the bottom dimension by 25% more space bbox1[2] <- bbox1[2] - (0.25 \* yrange) # Convert bounding box to a simple feature geometry bbox1 <- bbox1 %>% st\_as\_sfc()  # Setting up a second bounding box for the local r2, etc.  # It's the same but not adjusted on the y axis  # Get the bounding boxes of the parameter est maps  bbox2 <- st\_bbox(mapGWR) # Range of x and y values xrange <- bbox2$xmax - bbox2$xmin # range of x values yrange <- bbox2$ymax - bbox2$ymin # range of y values # Extend the right dimension by 25% more space bbox2[3] <- bbox2[3] + (0.25 \* xrange) # Convert bounding box to a simple feature geometry bbox2 <- bbox2 %>% st\_as\_sfc()  #---------------2.4 Mapping parameter estimates and T values for each of the 8 variables---------  # Based off the loops we learned in the second part of the model, I created loops  # for each parameter estimate map   # Here I created a list called variables that includes both the names of the variables  # and assigned them a color map that matches what group they are in - for example,  # the two health variables have a red blue color scheme.   variables <- list(  HouseOCrat = "BrBG",  Vigintilv2 = "BrBG",  Pct\_LTH\_prob = "RdBu",  CrimeRate = "PiYG",  HouseNCrat = "BrBG",  LowerEd\_Pct = "PiYG",  HlthSMR = "RdBu",  Pct\_Unemployed = "PiYG" )   for (var in names(variables)) { # names(variables) means that this loop will iterate over the names of the variables   # that I just defined (so leaves out the colormap)  # Dynamically create column names  t\_value\_col <- paste0(var, "\_TV") # This creates an empty t value column that has the name of the variable with \_TV appended to it (which we already defined in the GWR model)   # I was able to figure out how to use paste0 from this documentation  #https://www.digitalocean.com/community/tutorials/paste-in-r  beta\_col <- paste0(var, "\_beta") # This does the same for beta   beta\_sig\_col <- paste0(var, "\_beta\_sig") # Now for beta sig     # Identify non-significant areas --> because we defined t value col as an empty list that will be filled in iteratively for each variable name, we only need to write this once   whereNonSig <- which(mapGWR[[t\_value\_col]] > -1.96 & mapGWR[[t\_value\_col]] < 1.96) # 1.96 is the cut off four significant t values     # Copy original parameter estimates  mapGWR[[beta\_sig\_col]] <- mapGWR[[beta\_col]]  # --> it's the same step as here:   # mapGWR$Pct\_LTH\_prob\_beta\_sig <- mapGWR$Pct\_LTH\_prob\_beta, but instead it has double brackets in order to reference something in a list   # the documentation for this double brackets was found here:   #https://www.dataquest.io/blog/for-loop-in-r/    # Set non-significant areas to NA  mapGWR[[beta\_sig\_col]][whereNonSig] <- NA  # Same as this line but for the empty list of beta sig col and with double brackets as above:  # mapGWR$Pct\_LTH\_prob\_beta\_sig[whereNonSig\_Pct\_LTH\_prob] <- NA # Set non-significant areas to NA   } # End for   # Creation of maps for loop maps <- list() # This is an empty list that will store all the maps, otherwise they would delete as soon as we made them  for (var in names(variables)) { # Beginning of second for loop, same concept as with the first loop   beta\_sig\_col <- paste0(var, "\_beta\_sig") # finds the variable name with the \_beta\_sig\_col appended to it   palette <- variables[[var]] # Earlier, we defined the colormap in the same list as the variables, so now instead of calling name(variables) we can just call variables to get the values we assigned to them     # Generate map; this is the same process as we did in Assignment 1, but I've replaced the names of each map with maps[[var]], which means it will create the maps and store them in the list for each variable   maps[[var]] <- tm\_shape(mapGWR, bbox = bbox1) +  tm\_fill(beta\_sig\_col, style = "equal", palette = palette, # Palette and beta\_sig\_col we have defined earlier   colorNA = "lightgray", textNA = "Non-significant") + # Setting the color of NA to light gray so that we can use a bicolor map  tm\_borders() } # End for loop   Map\_HouseOCrat\_beta\_final <- maps[["HouseOCrat"]]# I've renamed every map in the maps list by assigning the individual map name to its own variable # This helps with plotting them on a grid  Map\_Vigintilv2\_beta\_final <- maps[["Vigintilv2"]] Map\_Pct\_LTH\_prob\_beta\_final <- maps[["Pct\_LTH\_prob"]] Map\_CrimeRate\_beta\_final <- maps[["CrimeRate"]] Map\_HouseNCrat\_beta\_final <- maps[["HouseNCrat"]] Map\_LowerEd\_Pct\_beta\_final <- maps[["LowerEd\_Pct"]] Map\_HlthSMR\_beta\_final <- maps[["HlthSMR"]] Map\_Pct\_Unemployed\_beta\_final <- maps[["Pct\_Unemployed"]]   # Create a 2x2 grid for the first three maps grid.newpage() pushViewport(viewport(layout=grid.layout(2, 2))) # 2 rows, 2 columns  # Place the maps in the grid layout print(Map\_HouseOCrat\_beta\_final, vp=viewport(layout.pos.col=1, layout.pos.row=1)) print(Map\_HouseNCrat\_beta\_final, vp=viewport(layout.pos.col=2, layout.pos.row=1)) print(Map\_Pct\_LTH\_prob\_beta\_final, vp=viewport(layout.pos.col=1, layout.pos.row=2)) print(Map\_HlthSMR\_beta\_final, vp=viewport(layout.pos.col=2, layout.pos.row=2))   dev.off()  # Create a 2x2 grid for the remaining four maps grid.newpage() pushViewport(viewport(layout=grid.layout(2, 2))) # 2 rows, 2 columns  # Place the maps in the grid layout print(Map\_CrimeRate\_beta\_final, vp=viewport(layout.pos.col=1, layout.pos.row=1)) print(Map\_LowerEd\_Pct\_beta\_final, vp=viewport(layout.pos.col=2, layout.pos.row=1)) print(Map\_Pct\_Unemployed\_beta\_final, vp=viewport(layout.pos.col=1, layout.pos.row=2)) print(Map\_Vigintilv2\_beta\_final, vp=viewport(layout.pos.col=2, layout.pos.row=2)) dev.off()  #-------------2.5 Mapping Local R 2------------ tm\_shape(mapGWR, bbox=bbox2) + tm\_fill("Local\_R2", style="equal", n=7, palette="Greens")+tm\_borders() # These values were already calculated when we ran the gwrmodel   # \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_2.6 Mapping Local Residual values with Global Residuals\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  ming <- min(mapGWR$stGlobalRes) maxg <- max(mapGWR$stGlobalRes) # Finding the min and max of the global standardised residuals ming maxg  breaksGlRes <- c(ming, -2.58, -1.96, 0, 1.96, 2.58, maxg)  head(mapGWR$stGlobalRes)   # Sort the breaks in ascending order breaksGlRes <- sort(breaksGlRes, na.last = TRUE) # This sorts the breaks and removes NAs at the end  globRes<-tm\_shape(mapGWR, bbox=bbox1) + tm\_fill("stGlobalRes", style="fixed", breaks=breaksGlRes, palette="RdBu") +tm\_borders+tm\_borders()  # Check if breaks are sorted and contain no NAs print(breaksGlRes) # Make sure it is sorted and no NAs  print(globRes)  #Local Residual, which is already calculated for us in the GWRMODEL minl <- min(mapGWR$Stud\_residual) maxl <- max(mapGWR$Stud\_residual)  breaksLRes <- c(minl, -2.58, -1.96, 0, 1.96, 2.58, maxl)  locRes <- tm\_shape(mapGWR, bbox=bbox1) + tm\_fill("Stud\_residual", style="fixed", breaks=breaksLRes, palette="RdBu") +tm\_borders()  grid.newpage() # Plotting on a grid to have them side by side  pushViewport(viewport(layout=grid.layout(1,2))) print(globRes, vp=viewport(layout.pos.col = 1, layout.pos.row =1)) print(locRes, vp=viewport(layout.pos.col = 2, layout.pos.row =1)) dev.off() #------Saving to then calculate local residuals in GeoDa----------- st\_write(mapGWR,"GG\_Noise\_GWR\_results\_Roads\_Scotgov.shp", append=FALSE) |