Correlation between housing assistance and distance to service stations and their related polluted areas

Marco Ruas[1]

[1]EPFL

January 31, 2019

Ruas M.

Ecole Polytechnique Fédérale de Lausanne

Introduction

In urban areas transportation and road networks have become a major source of concern for public health. Yet within a city, not all populations bear the same share of environmental pollution.

Many studies have been published, which relate various levels of social vulnerability with road pollution, mainly air quality assessments. In the UK a study (Namdeo and Stringer, 2008) has explored the impacts road user charging had on health, based on NO_2 levels and the difference in exposure between populations. In another paper (Carrier et al., 2014) correlation between NO_2 exposure and schools showing different deprivation levels was examined. In both cases the results showed a strong positive correlation between vulnerability among populations and health risks associated with road networks.

One of the elements associated with road networks often not considered is gas stations. These usually appear as a source of pollution, which also can have health impacts. One paper (Terrés et al., 2010) investigated the influence gas stations had on air in their surroundings due to VOCs emanating from them. Air measurements were taken and a model based on concentration ratios was developed to conduct the study. The paper concluded stations had an influence on close surroundings with high concentrations of benzene and n-hexane being measured. This influence was shown to depend highly on the characteristic of the area in the direct vicinity of the stations. When buildings were located close by the dispersion of VOCs was hindered, thus their concentration increased.

The present paper aims to enlarge the scope of analysis described previously mixing both approaches. Indeed, the focus will be on the correlation, which exists between socioeconomically vulnerable populations and their proximity to gas stations and areas polluted by those. The initial statement is that deprived populations are more inclined to be located closer to gas stations and their pollution related sites. Socioeconomic vulnerability was here considered in terms of housing assistance data available, which was then correlated with distances to verify the above mentioned hypothesis. The area of study is the municipality of Vernier (Geneva, Switzerland).

Data

In order to perform the study several raster and vector layers were used.

Demographic data of Switzerland was collected from the Swiss Federal Office of Statistics (OFS). This demographic data contained point coordinates, which enabled to build an hectometric raster grid.

To assess vulnerability, housing assistance data in the municipality of Vernier was used. This data also came from the Swiss Federal Office of Statistics and presented itself as a table containing the detailed numbers of housing assistance.

Then, data collected from the open data collection from the Geneva Territory Information System (SITG) was also used. One point shapefile contained the location of gas stations within the entire canton of Geneva. Another shapefile, of polygons this time, was used to locate the registered polluted areas.

Finally, the boundaries of the municipality of Vernier stored in a polygon shapefile were also used. Those also came from the OFS.

All vector and raster layers used were projected according to the Swiss coordinates system : EPSG21781.

Methods

The results needed to realize the study were produced using both QGIS and GeoDa softwares. Those were used following their respective "QGIS User Guide"(Athan, 2017) and "GeoDa User Guide" (Anselin, 2003).

The main analysis tool used was a 100x100m grid with its extent limited within the boundaries of Vernier. To create this grid the hectometric point coordinates contained in the Swiss demographic data file were necessary. These coordinates allowed to locate the center of the grid cells although a 50m both longitudinal and latitudinal shift was necessary in order to match the used data.

Using the "Spatial Query" tool available in QGIS enabled to confine the grid inside the municipality of Vernier. It is worth noticing that the grid did not cover the entirety of the municipality but only its populated areas.

After that, the vector "Analysis tool" in QGIS made possible to count the number of housing assistance provided in each hectometric cell. This proved useful to then categorize the grid according to a vulnerability graduation.

Both the service stations point shapefile and the polluted areas polygon shapefile were also restricted to the municipality of Vernier thanks again to the "Spatial Query" tool. Further work was performed on the polluted areas shapefile restraining its elements to the ones having an activity type of either "service station" or "fuel related businesses".

Finally, minimal distances from the center of the grid cells to service stations and polluted areas were computed. To do this, layers of polygon centroids were created. The results were then calculated using the "Distance matrix" contained in the vector analysis tools.

In GeoDa the grid was exploited to extract box maps, residual maps and scatter plots. Those allowed to establish comparisons between the different values of vulnerability and distance. They were also of interest to verify the spatial correlation between the short distances to polluted areas and high housing assistance numbers, which made possible the verification of the hypothesis. In addition, two models of multivariate regression (with and without a spatially dependent weighted variable) were performed in order to assess their performance with respect to real data.

Both methods are expressed as follows:

Ordinary Linear Regression (OLR):

$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i$

Spatially weighted regression:

$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \rho \, \Sigma w_j y_j + \epsilon_i$

In both equations the β_i represent the regression coefficients, y_i the dependent variable (housing assistance number), x_1 (independent) is the minimum distance from the center of a grid cell towards the closest service station, x_2 (independent) the minimum distance towards the center of a polluted area and ϵ_i the error. For the spatially weighted regression, ρ is the spatial lag and w_j the weight of spatial unit j relative to the spatial unit i.

Results

Figure 1 represents the obtained grid and its extent within the municipality of Vernier. Cells are divided in five housing assistance ranges by natural breaks. Housing assistance numbers serve here as an indicator of vulnerability. Darker cells depict a higher number of housing assistance with results showing only a few colored clusters. The majority of cells are represented in white, which are related to regions with no housing assistance provided.

The map also shows that polluted areas are mostly present in the upper-eastern part of the municipality, whereas service stations are well spread.

In order to analyze the dependency between housing assistance and distance to both service stations and polluted areas, two scatter plots were produced. They are represented in figures 2 and 3. These two scatter plots are standardized representations showing the correlation between the dependent variable (SUM_ALLOC: total number of housing assistance per grid cell) and the independent variables (MIN_NEW: minimal distance from the center of grid cell to the center of polluted area, DIST_MINSE: minimal distance from the center of grid cell to the closest service station).

The purple line is the regression line and shows in this case a negative autocorrelation between the values (slope b = -0.235 and -0.221).

Discussion

As indicated in the results section, both scatterplots in figures 2 and 3 show a negative autocorrelation between the housing assistance and the distances. This tends to confirm the original hypothesis stating that more vulnerable people are located closer to service stations and their related polluted areas. This class of people seems therefore more prone to pollution, as some buildings are located within a 75m radius from service stations, mentioned in the paper (Terrés et al., 2010) as the region of direct pollution impact due to the stations' emissions.

This comes as an interesting result and adds up to the fact that this group of population usually resides closer to road networks. Aggravated health risk factors could then be considered. Nevertheless, this conclusion is to be tempered. Indeed, the determination coefficient (R^2) is close to zero in both cases : 0.0533 and 0.0488. This indicates the regression line does not fit the data very well and high residuals are to be expected.

This bad fitting is emphasized by the vulnerability grids obtained when applying the non-spatially and spatially weighted regression. The results are presented in the figures 4 and 5. Both figures present the same kind of results. It is worth noticing that the regression models have a hard time approximating the higher housing assistance values as mentioned above.

This could be due to a number of factors but probably the main cause here is the high quantity of cells with a zero value (315) in comparison to the other ones. With both methods all values of interest come out as having high residuals. This tends to explain even more why both models have issues with the fitting.

The residuals for the non-spatially weighted regression are presented in figures 6 and 7. These are almost similar to the ones obtained with the spatially weighted method, which does not represent an improvement over the OLR here. This is why only one map of residuals is presented. Figure 7 clearly shows all non-zero values for housing assistance present high residuals.

One could then imagine eliminating all zero values from the analysis in order to improve the fitting. Figures 8 and 9 show the scatter plots obtained selecting only non-zero values. As can be noticed the negative autocorrelation increases since the smoothing effect of all the zero values is not present.

Nonetheless, an improvement in the regression fitting is not obtained, since the determination coefficients remain almost the same as before and are even slightly worse ($R^2 = 0.0312$ and 0.0363). Also, these zero data prove valuable to the analysis as they allow to establish the difference between cells showing no vulnerability and the others. It would not make much sense in the scope of this study to establish a comparison and perform regression fitting only on data, which corresponds to the more deprived people.

Conclusion

The purpose of this paper was to look into the relationship between vulnerability and closeness to service stations and their related polluted areas. This consisted in an interesting exercise, since the correlation between both has not been studied in as much depth as is the case for pollution emanating directly from road networks. The results appear to confirm the initial hypothesis that more deprived people find themselves in areas closer to gas stations and their pollution. This could come as another aggravating factor for their health, as this category of people are usually already the most affected by transportation pollution. To further enlarge this study it would have been interesting to develop other models that would better approximate the data and show clearer results. Also studying larger data, for example over the entire city of Geneva, would have proved valuable in order to confirm the results of this small pool of experience.

In conclusion, the results obtained here show the subject is worth exploring. This work would be well suited as a complement to a study on road networks pollution levels and their correlation with deprivation to better encompass the entirety of the issue.

References

Anselin. GeoDaTM 0.9 User's Guide. Technical report, 2003.

Athan. QGIS User Guide. Technical report, 2017.

Mathieu Carrier, Philippe Apparicio, Anne-Marie Séguin, and Dan Crouse. Ambient air pollution concentration in Montreal and environmental equity: Are children at risk at school? *Case Studies on Transport Policy*, 2(2):61–69, sep 2014. doi: 10.1016/j.cstp.2014.06.003. URL https://doi.org/10.1016%2Fj.cstp.2014.06.003.

Anil Namdeo and Claire Stringer. Investigating the relationship between air pollution health and social deprivation in Leeds, UK. *Environment International*, 34(5):585–591, jul 2008. doi: 10.1016/j.envint.2007.12.015. URL https://doi.org/10.1016%2Fj.envint.2007.12.015.

Isabel M. Morales Terrés, Marta Doval Miñarro, Enrique González Ferradas, Antonia Baeza Caracena, and Jonathan Barberá Rico. Assessing the impact of petrol stations on their immediate surroundings. *Journal of Environmental Management*, 91(12):2754–2762, dec 2010. doi: 10.1016/j.jenvman.2010.08.009. URL https://doi.org/10.1016%2Fj.jenvman.2010.08.009.

Figure Captions

Figure 1. Vulnerability, service stations and polluted areas

Figure 2. Scatter plot - Housing assistance numbers, distance to service stations

Figure 3. Scatter plot - Housing assistance numbers, distance to polluted areas

Figure 4. OLR model approximation

Figure 5. Spatially weighted model approximation

Figure 6. Map of residuals for OLR

Figure 7. Map of residuals for OLR with all non-zero values for housing assistance selected.

Figure 8. Scatter plot - zero values of housing assistance not selected (Service stations)

Figure 9. Scatter plot - zero values of housing assistance not selected (Polluted areas)

Figures

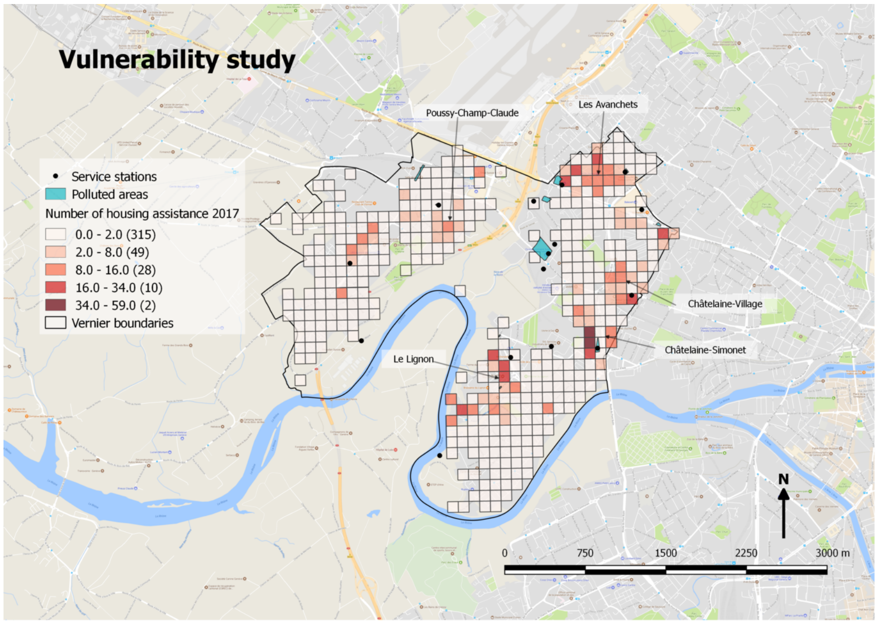
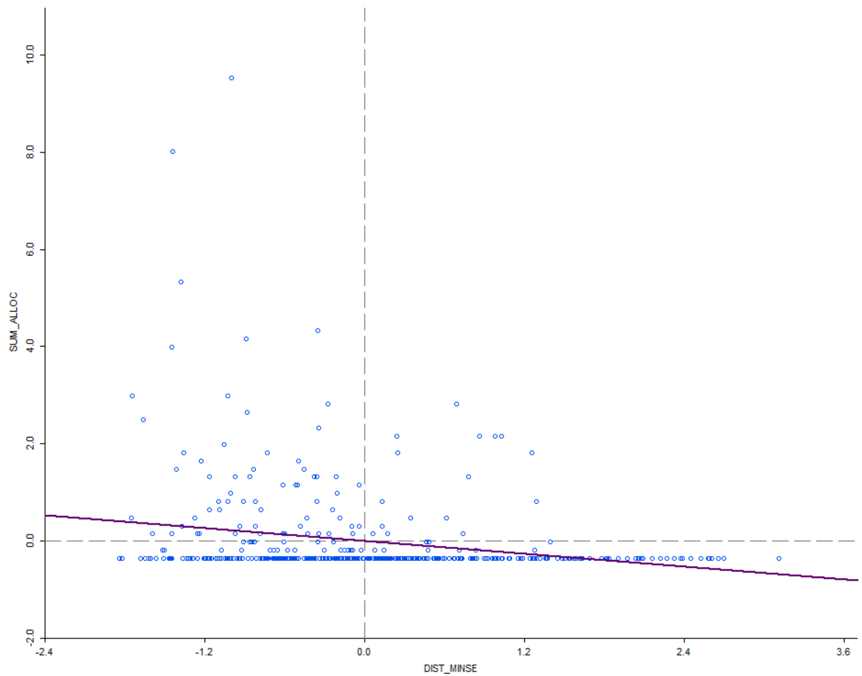


Figure 1: Vulnerability, service stations and polluted areas



#obs	R^2	const a	std-err a	t-stat a	p-value a	slope b	std-err b	t-stat b	p-value b
404	0.0488	0	0.0486	0	1	-0.221	0.0486	-4.54	7.41e-006
0	0	0	0	0	0	0	0	0	0
404	0.0488	0	0.0486	0	1	-0.221	0.0486	-4.54	7.41e-006

Figure 2: Scatter plot - Housing assistance numbers, distance to service stations

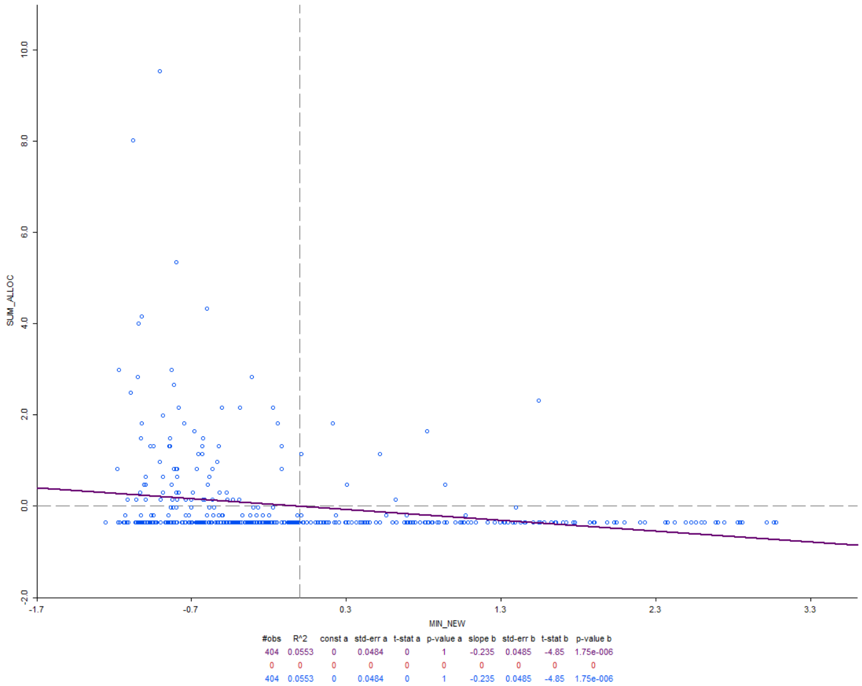


Figure 3: Scatter plot - Housing assistance numbers, distance to polluted areas

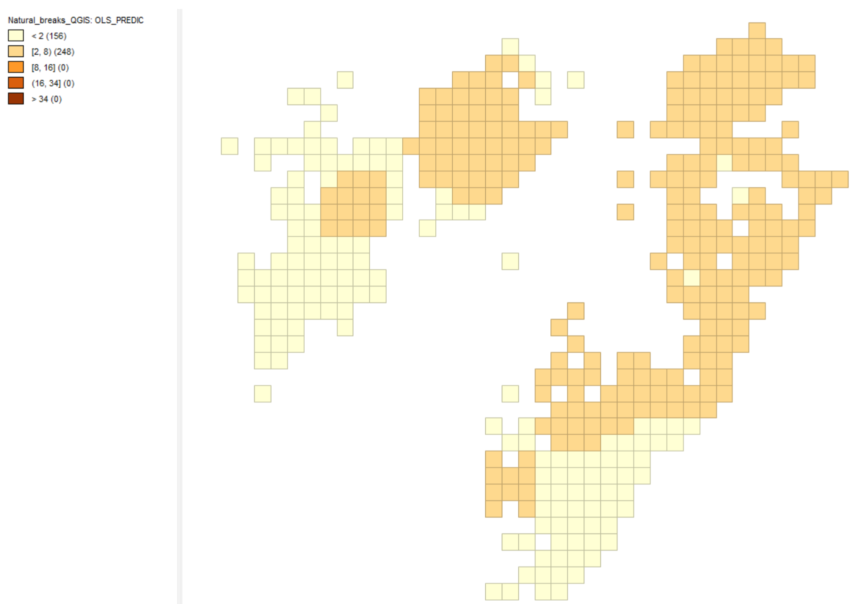


Figure 4: OLR model approximation

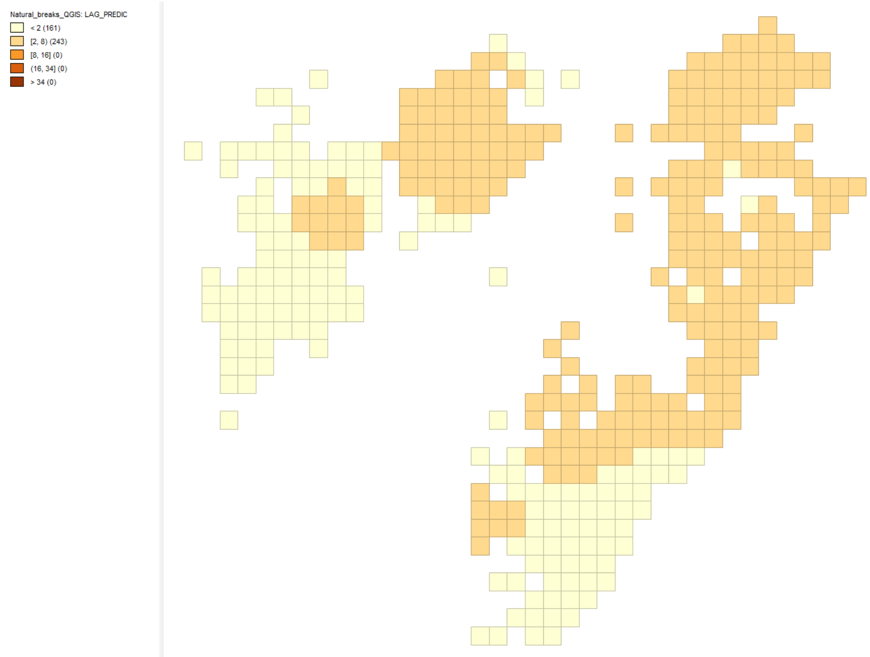


Figure 5: Spatially weighted model approximation

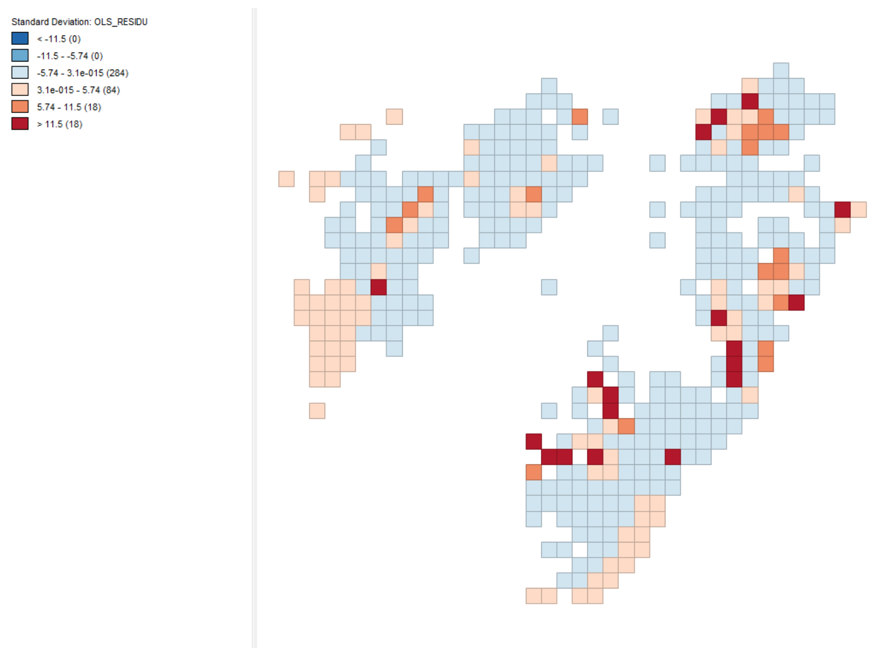


Figure 6: Map of residuals for OLR

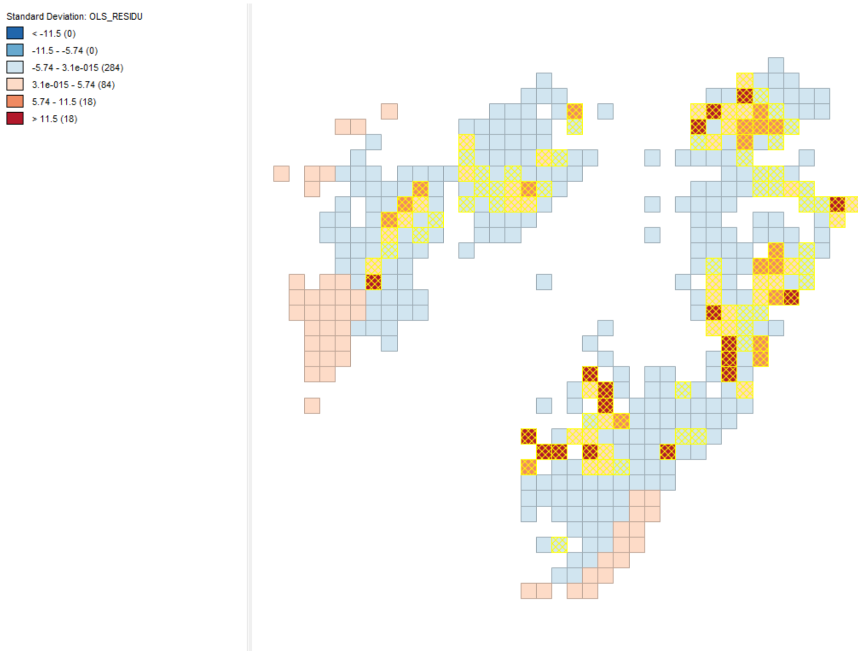
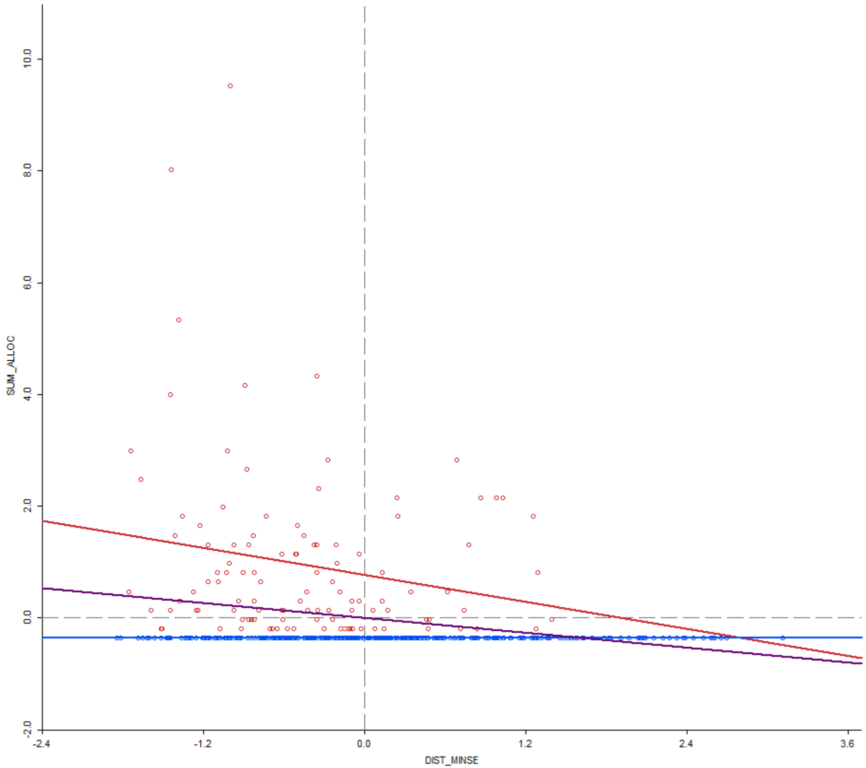
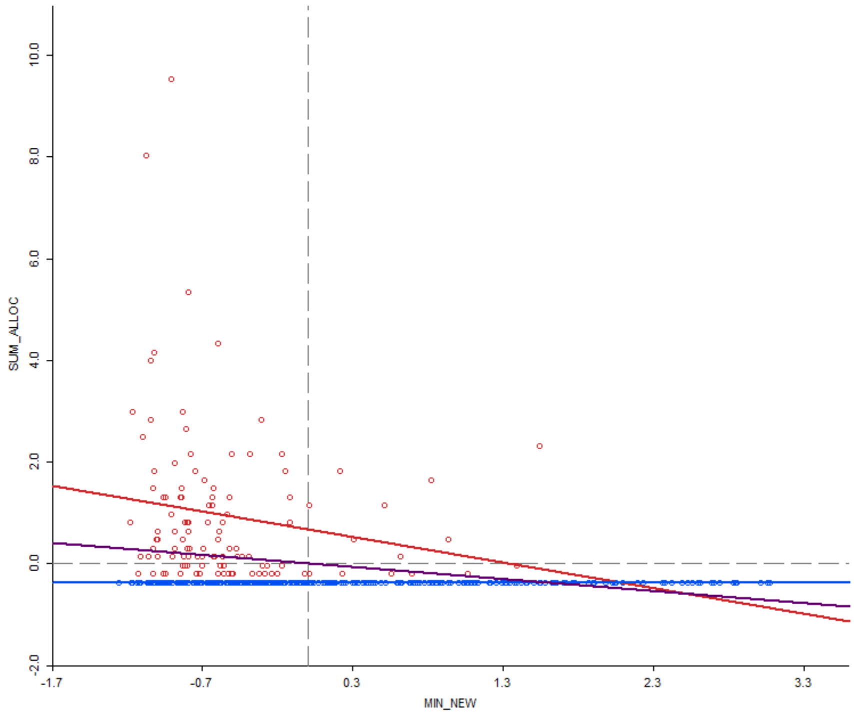


Figure 7: Map of residuals for OLR with all non-zero values for housing assistance selected.



#obs	R^2	const a	std-err a	t-stat a	p-value a	slope b	std-err b	t-stat b	p-value b
404	0.0488	0	0.0486	0	1	-0.221	0.0486	-4.54	7.41e-006
112	0.0363	0.771	0.168	4.58	1.22e-005	-0.402	0.197	-2.04	0.0441
292	1	-0.364	1.37e-016	-2.66e+015	0	-3.62e-016	1.31e-016	-2.77	0.00588

Figure 8: Scatter plot - zero values of housing assistance not selected (Service stations)



#obs	R^2	const a	std-err a	t-stat a	p-value a	slope b	std-err b	t-stat b	p-value b
404	0.0553	0	0.0484	0	1	-0.235	0.0485	-4.85	1.75e-006
112	0.0312	0.679	0.203	3.34	0.00115	-0.506	0.269	-1.88	0.0626
292	1	-0.364	1.37e-016	-2.66e+015	0	-3.47e-016	1.27e-016	-2.73	0.00664

Figure 9: Scatter plot - zero values of housing assistance not selected (Polluted areas)