

Geographically Weighted Regression

Chelsey-Ann Cu

32482135

GEOB 479 L2A

University of British Columbia

Dr. Brian Klinkenberg

9 February 2018

I. What is Geographically Weighted Regression?

Regression consists of a large range of modeling spatial relationships between independent and dependent variables. In a regression analysis, one would normally make the assumption that the relationship that is being modelled is uniform throughout the study area. This entails that the dependent variable (y) is directly related to the independent variable (x) within the estimated parameters (β_0 and β_1) ($y = \beta_0 + \beta_1 x_1 + \varepsilon$) (Charlton, Fotheringham, & Brunson, 2009). This is estimated in a way that to minimize the sum of the residuals for the observations. This model is usually fitted using an Ordinary Least Squares (OLS) procedure. OLS is a global method of linear regression that generates predictions or models a dependent variable in terms of its relationship to a set of explanatory variables. An OLS estimator works with the formula: $\beta = (X^T X)^{-1} X^T y$, “where β is the vector of the estimated parameters, X is the designated matrix which contains the values of the independent variables and a column of 1s, y is the vector of observed values, and $(X^T X)^{-1}$ is the inverse of the variance-covariance matrix” (Charlton, Fotheringham, & Brunson, 2009, p. 1). A goodness of fit test is usually used in this method to check the ability of a model to replicate the observed values. This is expressed as R^2 , a value which goes from 0 to 1, measuring the proportion of variation in the observed values which is accounted for by the variation in the model. This basic regression model makes assumptions that observations are independent of one another.

Unfortunately, this idealized situation is not always the case. With geography, there is an assumption that phenomena will vary across space. Tobler expressed that “Everything is related to everything else, but near things are more related than distant things” (Tobler, 1970, p. 236). This means that both variables and residuals in a given model could show spatial dependence. As a result, parameters that are estimated with linear models of regression could be insufficient and

biased if there is a spatial aspect involved. Another issue with linear regression analysis is that it assumes the relationship being modelled is the same throughout the study area. In fact, it may be the case that the processes that create these relationships that one is interested in may change with space.

Geographically Weighted Regression (GWR) is a method for exploratory spatial data analysis (Charlton, 2009). With GWR the relationship takes into account the location (u) in the study area as an input parameter and takes on spatial coordinates with the data points. This creates a new formula: $y(u) = \beta_0(u) + \beta_1 x_1(u) + \varepsilon(u)$ (Charlton, Fotheringham, & Brunsdon, 2009). Here, the locations are assumed to be where the data is collected from, allowing a separate estimate of the parameters to be mapped. With GWR taking into account local geography, it ensures that observation points that are near to each other have a greater effect (weight) in the estimation than points that are further away. Weights are for a computation of a weighing scheme (also known as a kernel) which have bandwidths that grow larger as the GWR model approaches the OLS model (Charlton, Fotheringham, & Brunsdon, 2009).

GWR analysis is a useful analysis tool that can be used for a variety of different purposes. It can be used to better understand phenomenon through looking at how changes in one variable can cause changes in another. Moreover, it can create a consistent and accurate model which is useful in predicting phenomenon. However, it is important to remember that with GWR, all independent variables that affect the dependent variable must be taken into account, otherwise the model produced in the analysis will be an inaccurate representation. The accuracy of a GWR model can be determined by looking at the residuals, like in an OLS model.

II. Case Study: Vancouver Children's Social Scores

This study takes into account various neighbourhood factors and their relationship to the development of children's social skills in Vancouver. The analysis consisted of using explanatory regression to determine the variables that affect social skills the most. The results identified that gender, language and income were the best variables to use. Then an Ordinary Least Squares analysis was performed to determine the level of correlation in the absence of spatial information (see Figure 1). Finally, a Geographically Weighted Regression analysis was done to determine the spatial correlation between children's social skills and the given parameters (See Figure 2).

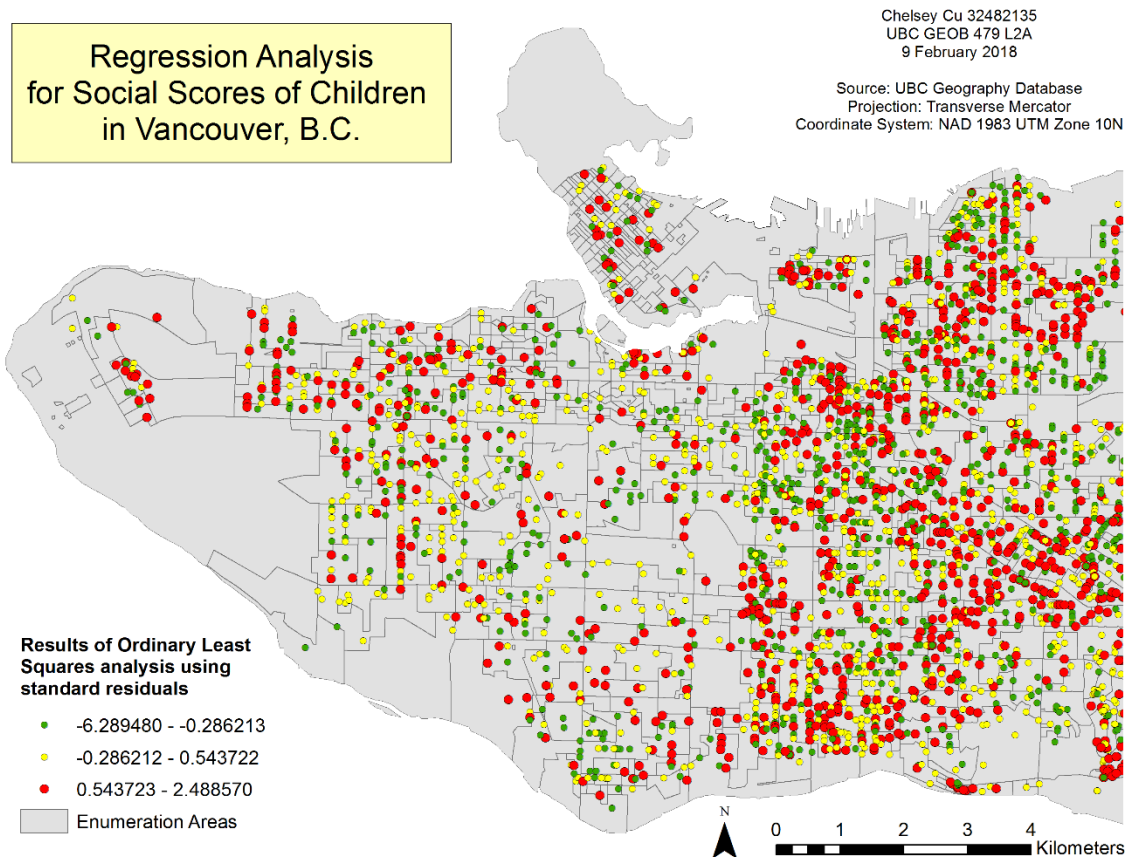


Figure 1. Regression Analysis for Social Scores of Children in Vancouver, BC. This map shows the standard residuals of the OLS analysis results. The adjusted R^2 value is 0.374.

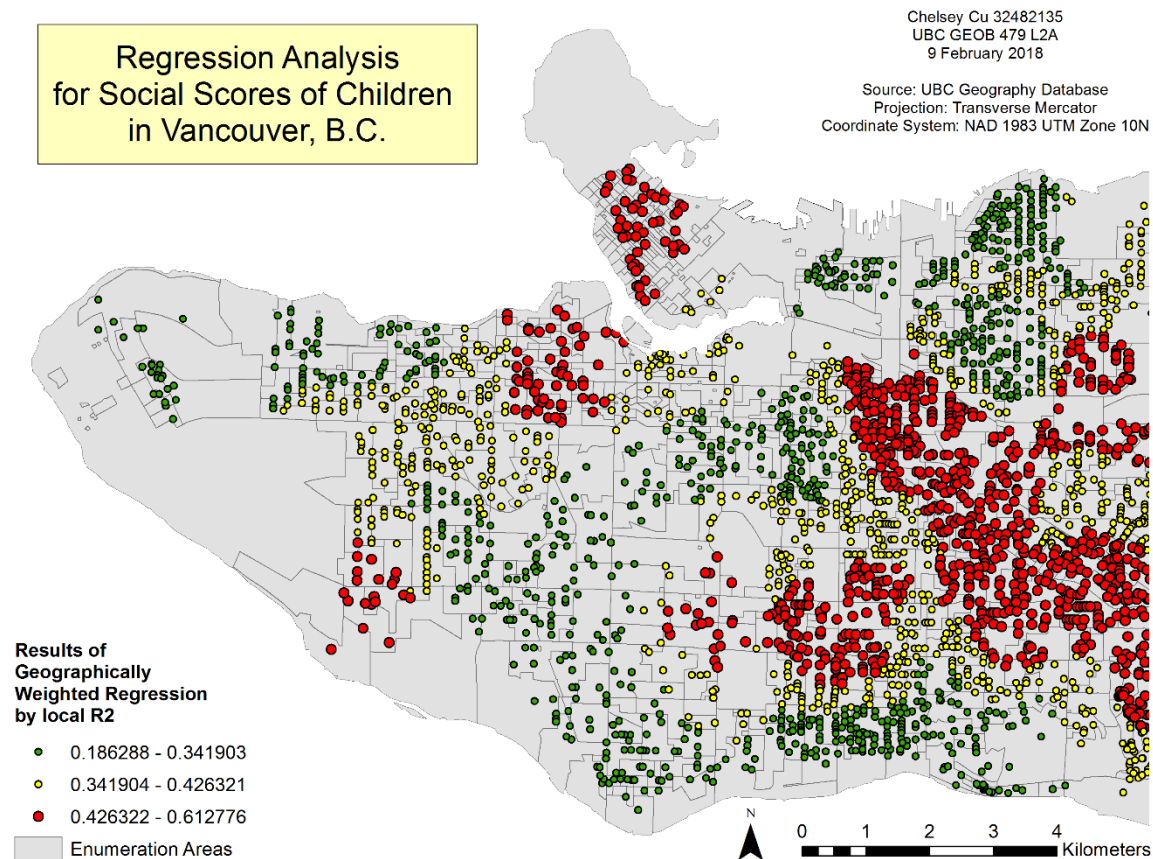


Figure 2. Regression Analysis for Social Scores of Children in Vancouver, BC. This map shows the results of a GWR analysis based on the local R^2 value to show the accuracy of the predicted model. The most accurate points are in red, while the least accurate are in green.

The results of both analyses were then compiled and using the absolute value of the differences in predicted values. A map was created to determine the difference in regression between the two types of analyses (see Figure 3). The figure shows that mostly there is a large discrepancy between OLS and GWR results in the east side of Vancouver, where the majority of the red points are. On the west side of the map, there is less discrepancy, showing that the two models fit well with one another in this particular instance. This demonstrates that the OLS

model and the GWR model are fairly similar so long as there is no relevant spatial factor in the analysis.

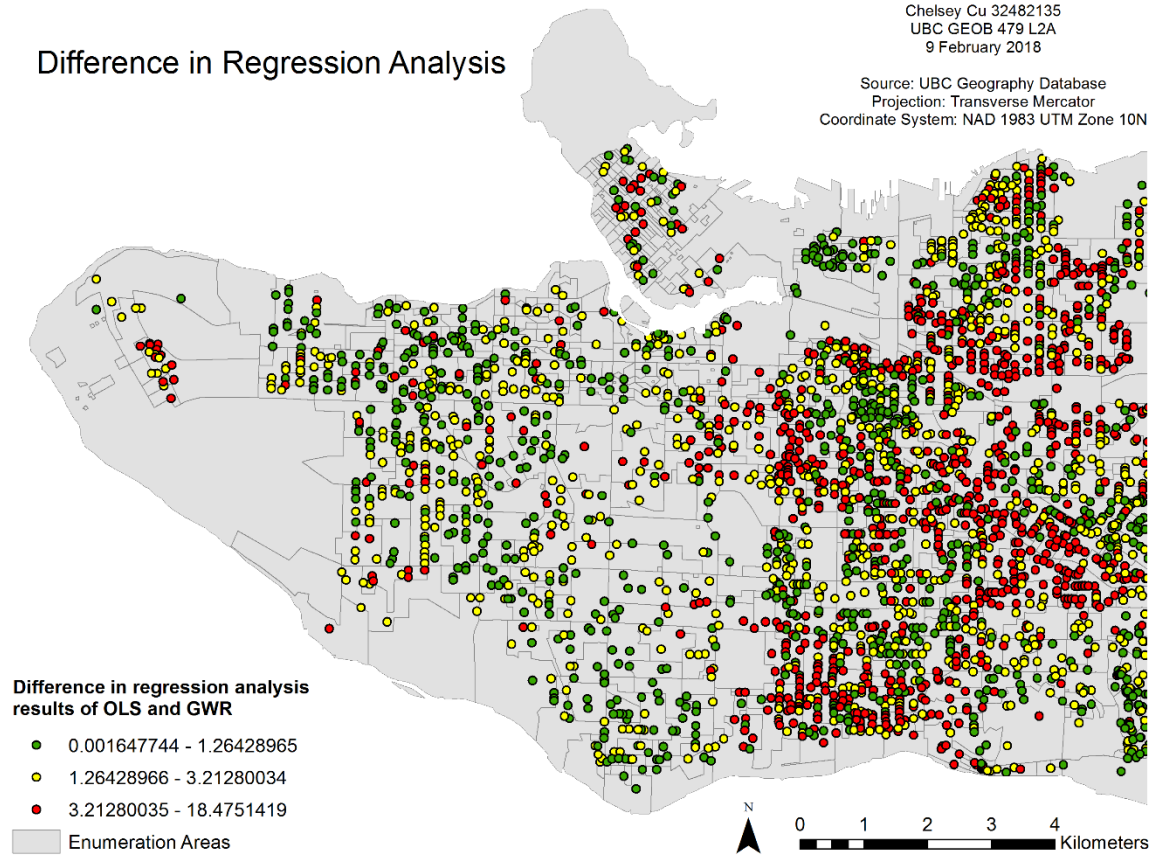


Figure 3. The Difference in Regression Analysis. This was computed by taking the absolute value of the difference between the estimated/predicted values of the GWR and OLS that was calculated for each child.

In order to look further into the spatial contexts and the given parameters, a grouping analysis was performed. This took natural clusters in the data and clumped them together to form 4 groups. They are as followed:

Table 1. Grouping Analysis for Enumeration Areas.

	Income (\$)	Neighbourhood families with > 4 members (%)	Children spend > 30 hours in childcare (%)	Neighbourhood families that are single parents (%)	Neighbourhood immigrants that have spent < 5 years in Canada (%)
<i>Group 1</i>	High	High	Average	Low	Average
<i>Group 2</i>	Low	High	High	High	Average
<i>Group 3</i>	Average	Low	Low	Low	Low
<i>Group 4</i>	Average	Average	Average	Average	High

Through grouping similar neighbourhood characteristics, the variation in social skills in terms of another factor can be more easily pinpointed. Figure 4 shows the relationship between 5 different factors and their impacts on the social scores of children. Here, one can see that the east side of Vancouver is predominantly Group 2 while the downtown area is dominated by Group 3 and the west side is a mixture of Group 1 and 3. For the most part, children living in the red areas (Group 2) experience larger negative impacts to social scores with the different variables (income, gender, and language scores). This can be seen in the coming analysis.

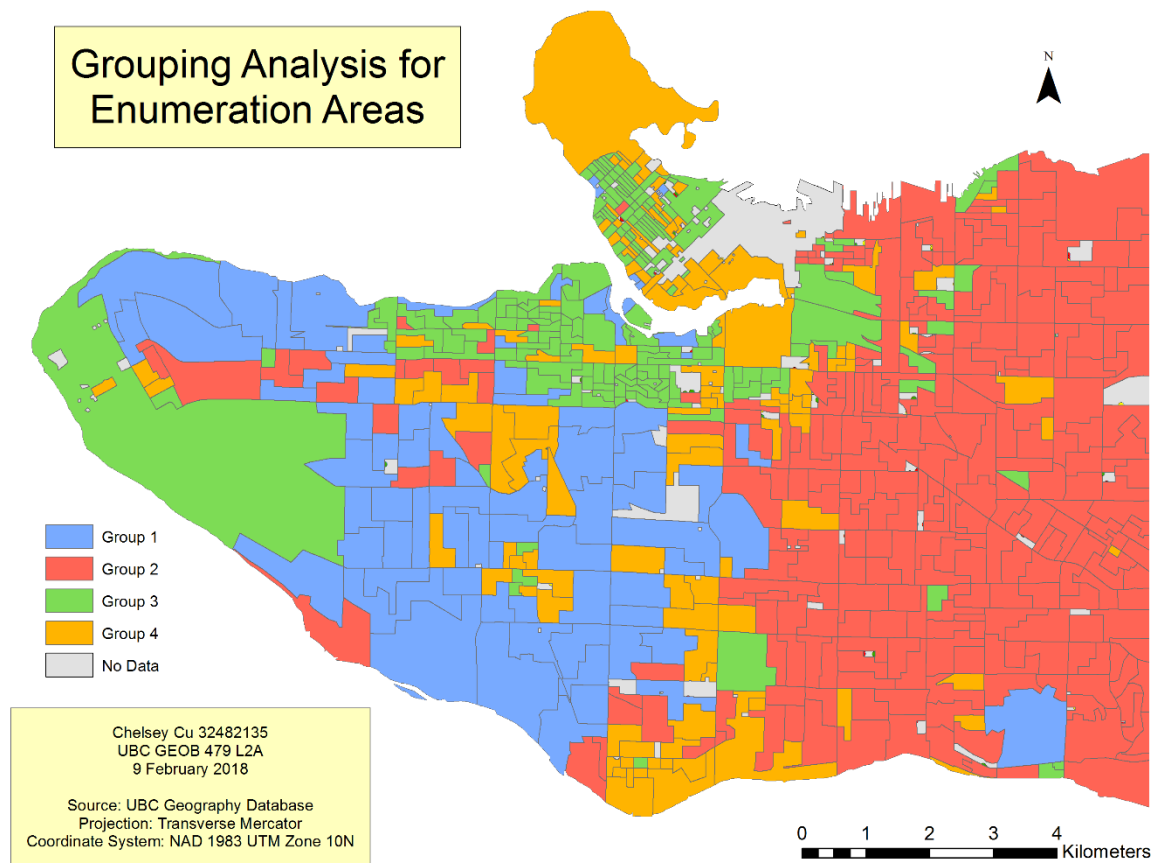


Figure 4. Grouping Analysis for Enumeration Areas. This map corresponds with the categories laid out in Table 1. The analysis shows which neighbourhoods have similar characteristics in terms of: income, family size, childcare, single parents, and immigrants. Areas that are similar in variables are grouped together.

For the purposes of this analysis, the points showing the differences between OLS and GWR will overlay the data to show the difference in spatial and aspatial analysis (Figure 5, 7, 9). The maps showing their results of the GWR by the Local R^2 values overlay the layers show the accuracy of the results (Figure 6, 8, 10). Each of the maps look at a particular parameter given the assumption that the others do not affect children's social scores.

The following map indicates that there is a connection between income and a child's social skills (see Figure 5). For this parameter, every increase in \$1000 will positively or negatively affect children's social scores by either decreasing their score by 2 points (red areas) or increasing their score by 2 points (green areas). Income has a lower impact on a child's social skills in the east side of Vancouver as opposed to the west side of Vancouver. This could be because with higher income, children can afford additional opportunities to socialize, such as extra-curricular activities (Richard & Dodge, 1982). In opposition, children living in lower income areas are not afforded the same opportunities due to limits in factors such as family funds or parental time constraints. In certain areas, like the east side of Vancouver, where income varies greatly among neighbours, social scores decrease substantially with change in income.

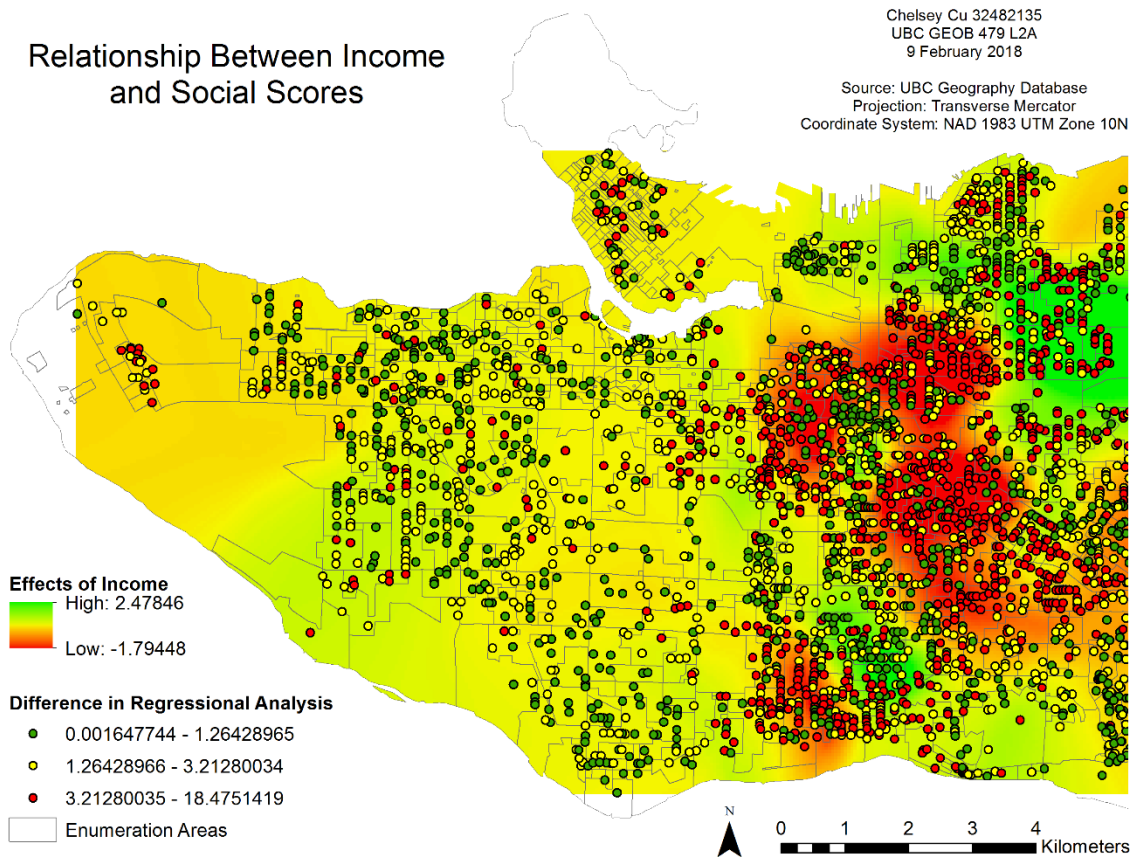


Figure 5. Relationship between Income and Social Scores with the difference in regression analysis. For every thousand dollars of increase in income, there is the above represented change in social scores.

The Local R^2 values from the GWR analysis show that the results around the Downtown, Kitsilano and East Vancouver region area are the most accurate (see Figure 6). This is because R^2 is the proportion of variation in the dependent variable (social skills) that is explained by the model. It is measured from 0 to 1 with a number closer to one showing the higher accuracy in data. In the case of Figure 6, the accurate points are shown in red. With green being points with higher uncertainty.

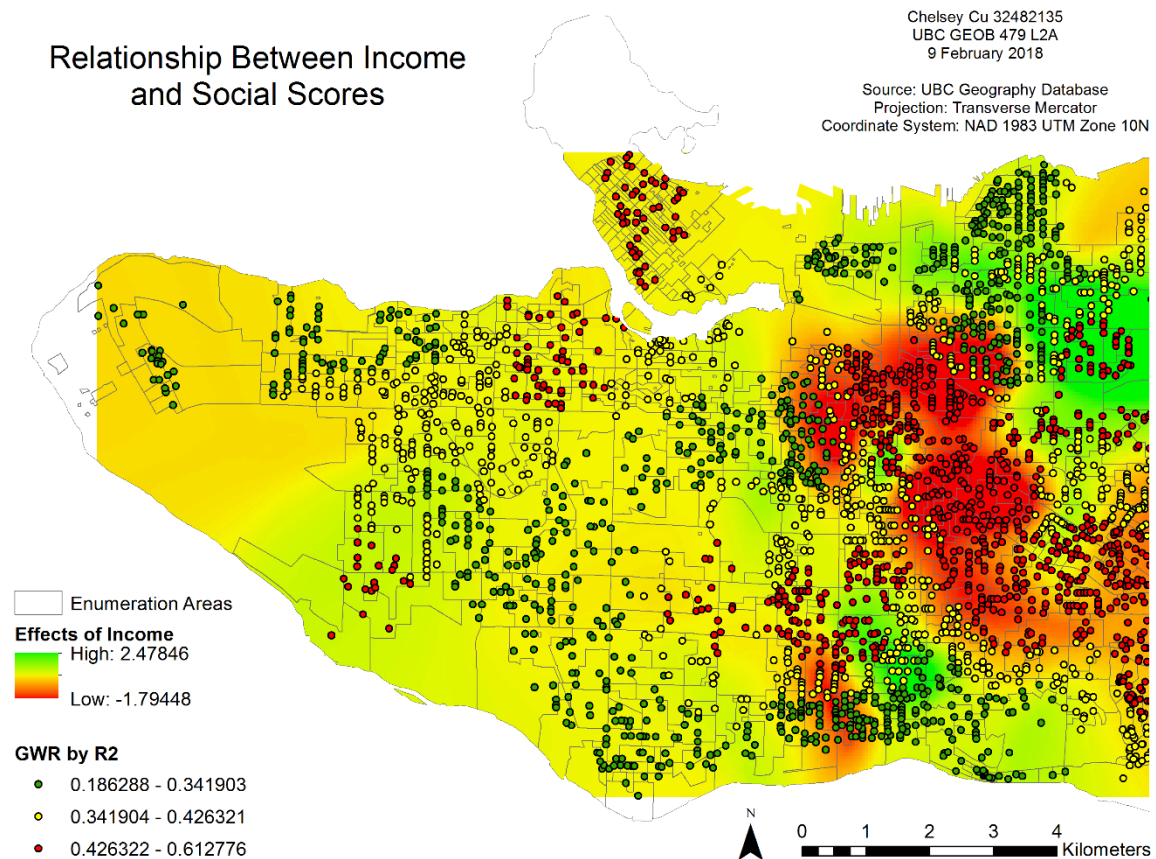


Figure 6. Relationship between Income and Social Scores with the GWR by local R^2 .

Furthermore, analysis showed that there is a relation between the social scores and language scores of children (see Figure 7). In the east side of Vancouver, language scores seemed to have a higher impact on social skills than it does on the west side. This could be attributed to the social factors of grouped areas that were previously discussed. The map below shows that there are clusters of areas where slightly higher language scores correlate to higher social scores especially in mid-eastern Vancouver. These are areas with lower incomes, larger families and larger amounts of single parents (see Figure 4). Also, in mid-west Vancouver, there is a large area where language has a low effect on children's social skills. This could be

attributed to less accurate data in general (see Figure 8), or a shift in population dynamics such as a community that has more social interaction with neighbours. It is also important to note that there is overall very little variation in language scores affecting social scores, as seen on the legend scale numbers on the map (see Figure 7, 8).

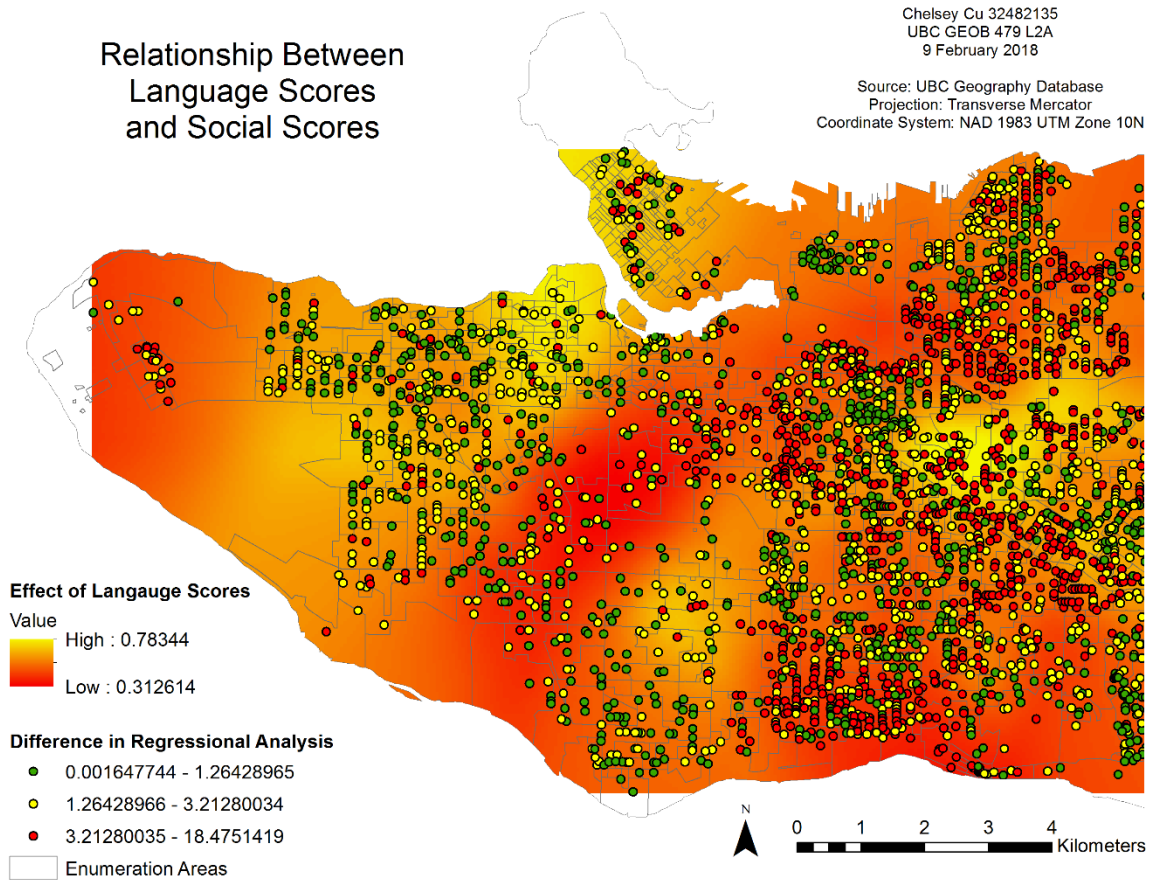


Figure 7. Relationship between Language Scores and Social Scores with difference in regression analysis. For every one unit of increase in language score, there is the above represented change in social scores.

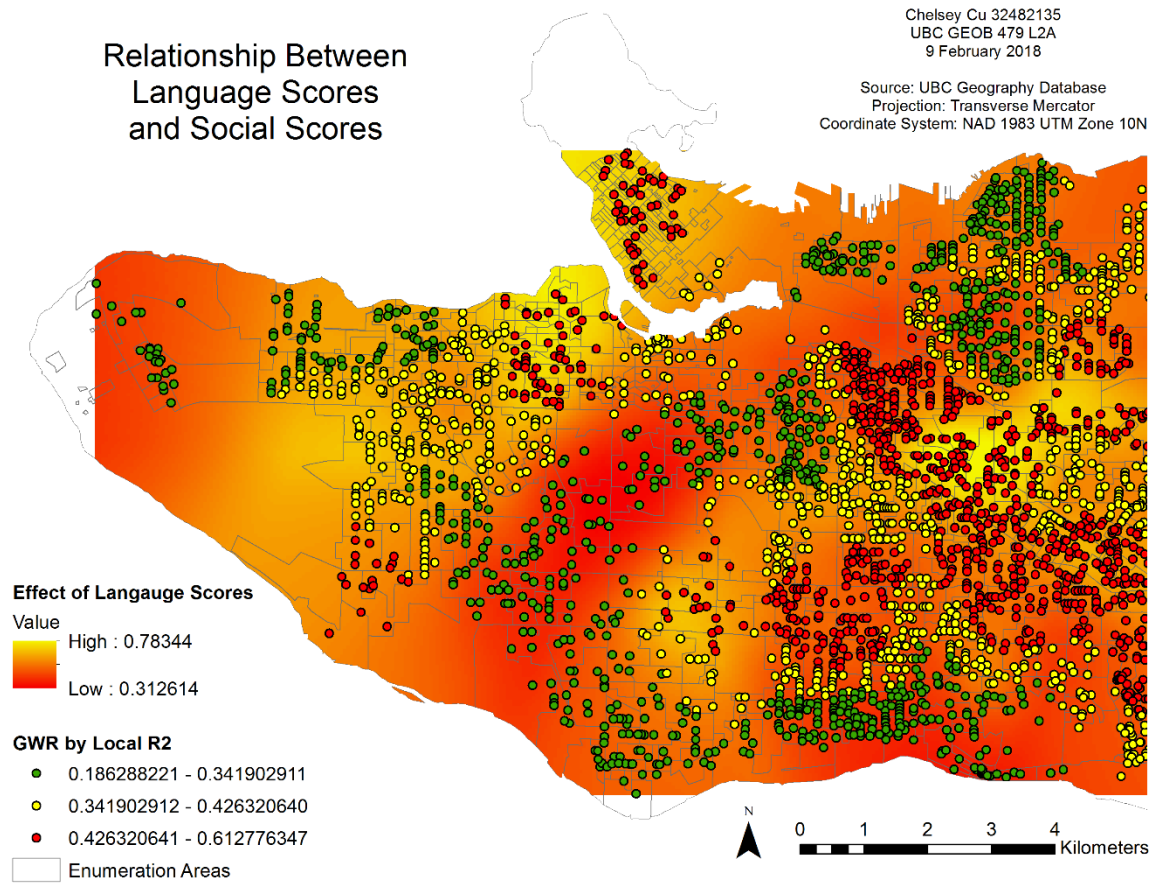


Figure 8. Relationship between Language Scores and Social Scores with the GWR by local R^2 .

Finally, the relationship social scores and gender was also analyzed (see Figure 9). On the map, red indicates areas where being a female negatively impacts social scores by decreasing their scores up to 17 points. Green indicates areas where being female positively impacts social score by 2 points. This map can be analyzed next to the income map (Figure 5) as showing counteracting variables. For instance, in areas where being female increases social scores, the effects of a thousand dollars difference in income largely decreases social scores. This can also be compared to language scores, were in the same area that being female positively impacts

social scores, high language scores also increase social scores. One can assume that these areas with extreme values have data points that are more susceptible to variability in parameters.

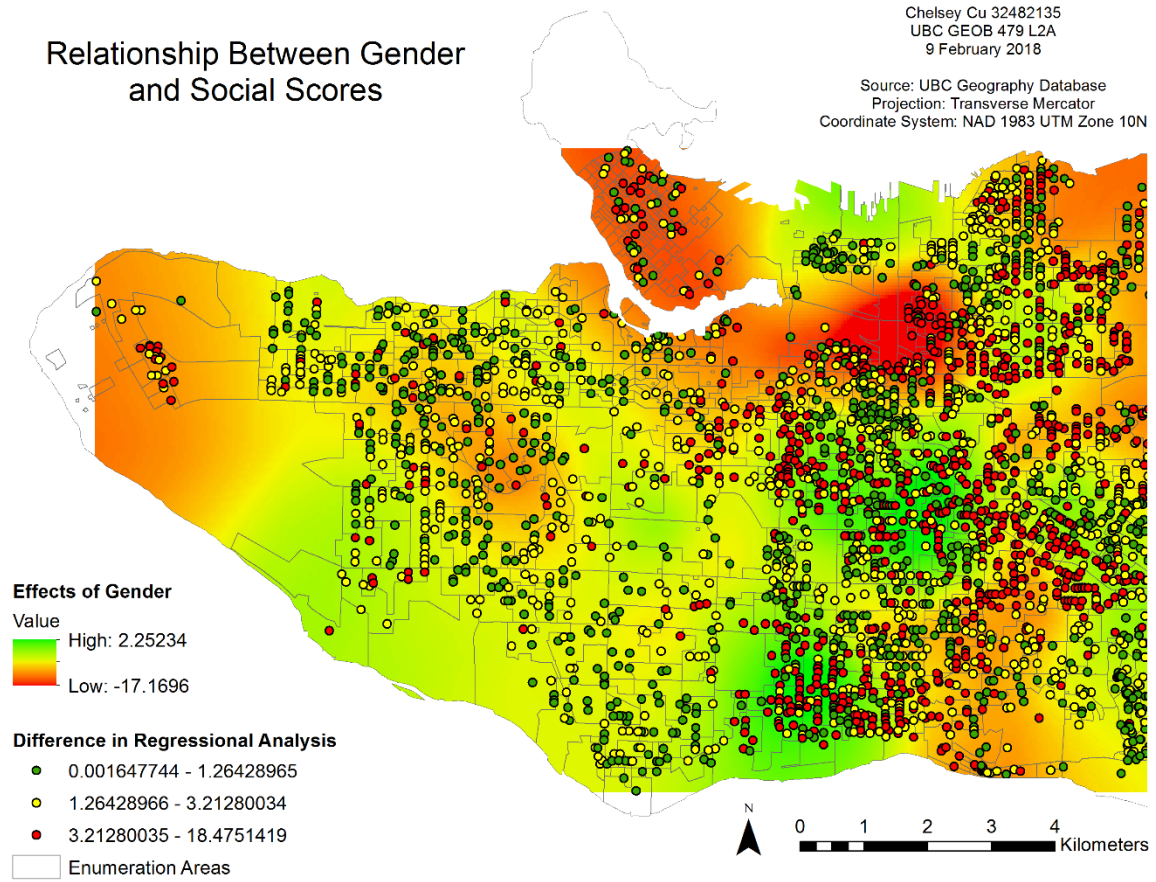


Figure 9. Relationship between Gender and Social Scores with difference in regression analysis.

This map represents a change in scores if the child is female versus male and the corresponding shift in social score due to gender.

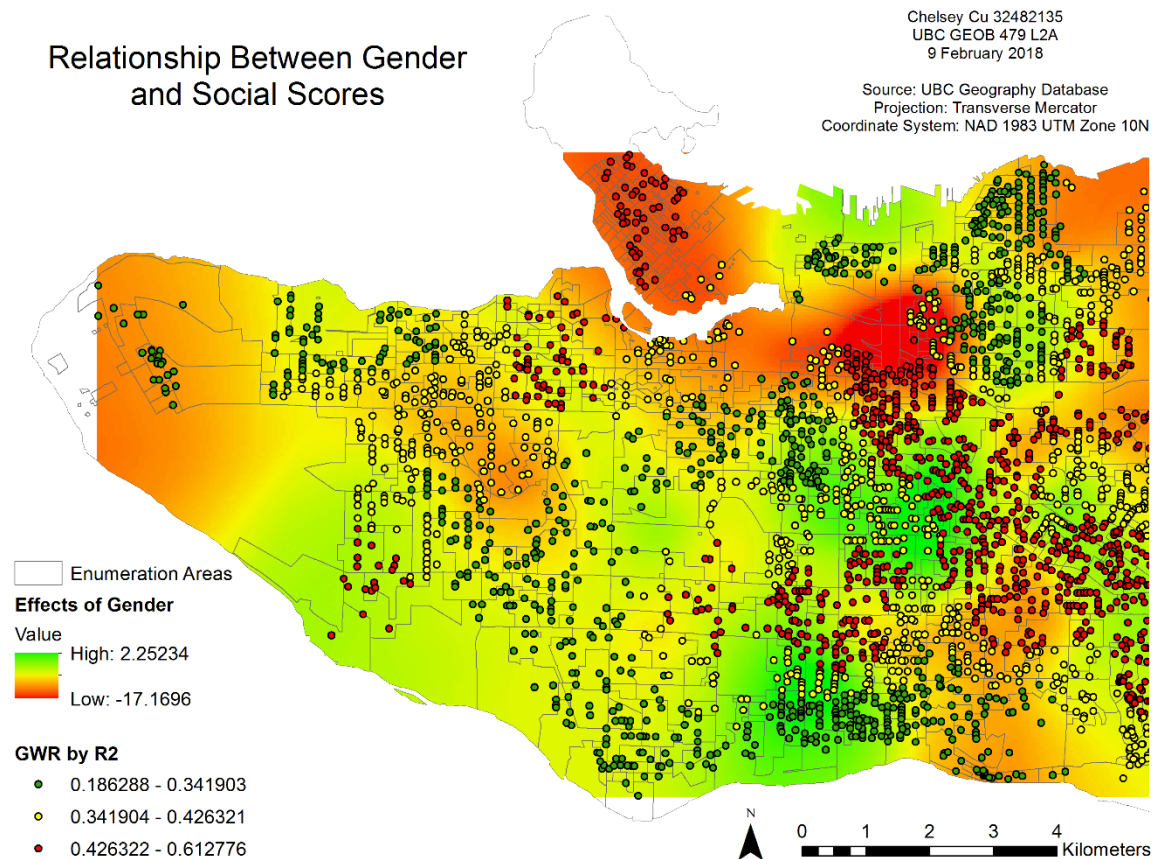


Figure 10. Relationship between Gender and Social Scores with the GWR by local R^2 .

III. Other Applications of GWR

GWR is a valuable analysis tool in many other situations than the example highlighted in the case study. Since most variables that are analyzed for correlation represent things in the real world which are inherently spatial, GWR adds the missing spatial factor to the analysis to enhance accuracy. It is only reasonable to assume that any set of variables, which exist with real world locations, are impacted by location as a factor and therefore, it should be included as a parameter in the analysis. Below are several examples in of applications in GWR which further highlights the benefits of implementing the model.

A. *Housing Costs*

An example of GWR at work, is with real estate pricing. The limitations with linear models in this field is that they often over- or underestimate the asking prices in some neighbourhoods (Legg & Bowe, 2009). In order to improve on the model accuracy, GWR is used to eliminate some of the residual errors. Legg and Bowe (2009) created a study in which took 93 homes that were listed on the market and applied a linear regression analysis and GWR based on location, and square footage. With a GWR model, the measured R^2 had higher value meaning an increase in accuracy. Their results showed that “lot value coefficients indicate that as lots are located nearer the urban core and farther from the rural townships, lot square footage price increases. In contrast, coefficients suggest that the larger the house, the less it contributes to the listing price” (Legg & Bowe, 2009, p. 45).

B. *Crime*

Other phenomenon, such as crime distribution, across cities can also be analyzed with a GWR model. Cameron *et al.*, (2016) look at the relationship between the location of alcohol outlets and the location of violent crimes in New Zealand. Here, they used data on police-attended violent incidents at the census area level. Their results confirmed that there are more violent events that occur closer to establishments that sell liquor. Through a spatial analysis they were able to identify that places that certain places, such as bars or night clubs, are associated with an increase in violence annually. Their findings are important because they show areas where intervention is most crucial to minimizing alcohol-related violence.

C. Social Justice

Further applications of GWR can be seen in differential access to resources. Tsiko (2016), explores the spatial variation of factors that affect women's access to land in Africa. The study found that less educated or HIV-positive women are more likely to be given access to own land rather than educated women. Also, an increase in population density, negatively affected women's ability to access to family land. These factors were investigated further to try to identify how they affected landownership.

IV. Conclusion

As identified by the case study and the further examples given, Geographically Weighted Regression is a useful spatial analysis method that can be used in a variety of contexts. It is important to account for spatial variability in analyzing and explaining phenomenon, given that patterns vary with geography. In other words, the nearness of an object oftentimes affects how and what results will occur.

References

- Cameron, M. P., *et al.* (2016). Alcohol Outlet Density and Violence: A Geographically Weighted Regression Approach. *Drug and Alcohol Review*, 35(3), 280-288. doi:10.1111/dar.12295
- Charlton, M. (2009). *GWR Explained*. Retrieved from Geographically Weighted Modelling: <http://gwr.maynoothuniversity.ie/what-is-gwr/>
- Charlton, M., Fotheringham, S., & Brunson, C. (2009). Geographically weighted regression. *White paper. National Centre for Geocomputation. National University of Ireland Maynooth.*
- Legg, R. J., & Bowe, T. (2009). Applying geographically weighted regression to a real estate problem, an example from Marquette, Michigan. *ArcUser*, 44-45.
- Richard, B. A., & Dodge, K. A. (1982). Social maladjustment and problem solving in school-aged children. *Journal of Consulting and Clinical Psychology*, 50(2), 226.
- Tsiko, R. (2016). Geographically Weighted Regression of Determinants Affecting Women's Access to Land in Africa. *Geosciences*, 6(1), 16. doi:10.3390/geosciences6010016
- Tobler, WR, 1970, A computer movie simulating urban growth in the Detroit region, *Economic Geography*, 46(2), 234-240.