Lab 3

Geographically Weighted Regression

A Case Study of Children's Social Skill Scores in Vancouver, Canada

Camille Potier
GEOB479
February, 2018

Introduction

The following report is a review of the use of geographically weighted regression (GWR). The first section will introduce what GWR is, the second part will evaluate the use of GWR in respect to the case study of explaining social skills score of children in Vancouver, Canada, and the final part will provide other examples of when GWR can be useful to use in order to explain spatial relationships.

Geographically Weighted Regression - An Introduction

Geographically Weighted Regression (GWR) is a popular method used within the field of Geographic Information Science that explores spatial data analysis, and models spatial relationships.[1] The foundational idea behind GWR is to explore the relationship between a dependent variable (Y), and a single or multiple independent variables (X), as it varies across the landscape. [2]

Regression analysis enables one to "model, examine, and explore spatial relationships and can help explain the factors behind observed spatial patterns." Such models can also be used to predict future patterns. The Ordinary Least Squares (OLS) regression is the most well-known technique. It calculates a **global** model for the variable you are trying to understand; only one equation is generated for the entire study area. However, another spatial regression technique which is increasingly being used is the Geographically weighted regression which provides a **local** model of the dependent variable to be explained - in such a technique a regression equation is calculated for every feature point in the data set - thus taking into accounts each feature's closest neighbors.[3] Unlike, Ordinary Least Squares Regression analysis the GWR analysis looks for geographical differences and looks at spatial variations in the relationship between the dependent variable and the independent variables. [4]

The geographically weighted regression is an extension of the linear model which allows for its analysis/model **to vary over space.** From its output it is therefore possible to find areas where the independent variables have a positive relationship with the dependent variable, whilst in other places it may be negative.

---

[1] "GWR Explained." Geographically Weighted Modelling. Accessed February 2018. http://gwr.maynoothuniversity.ie/what-is-gwr/.

[2] "Introduction to Geographically Weighted Regression." 2009. https://www.bristol.ac.uk/media-library/sites/cmpo/migrated/documents/gwr.pdf.

[3] "Regression Model and Spatial Autocorrelation." Geographic Information Systems Stack Exchange. Accessed February 2018. https://gis.stackexchange.com/questions/27245/regression-model-and-spatial-autocorrelation.

[4] Legg, Robert, and Tia Bowe. "Applying Geographically Weighted Regression to a Real Estate Problem." 2009. Accessed February 2018. http://www.esri.com/news/arcuser/0309/files/re_gwr.pdf.

By exploring spatial heterogeneity GWR addresses the geographical thinking assumption that spatial phenomenon vary across a landscape. The model is not looking at variation over the overall data space. Instead, it is used using a "weighted window" over the data, analyzing values and estimating coefficients at specific points by looking at the surrounding neighbors.[5] Typical regression-based models, such as Ordinary Least Squares ignore that assumption and thus provides a less accurate explanation of spatially varying relationships. [6] This is not to say that the OLS regression is not appropriate and accurate. Indeed, aspatial models can, in some cases, lead to high correlation between the model and estimated values from the independent variables. Nevertheless, in most cases, and while analyzing geographically sensitive topics, the GWR model will increase the accuracy of the model and in general have a higher fitness between the model and reality. Consequently, geographically weighted regressions can be seen as an improvement over using regressions such as OLS. Ordinary least squares regressions model a global relationship whilst GWR use neighboring data values to estimate spatial relationships and thus computes more accurate predictions.[7]

To provide a local model for the explanatory variables, the GWR will fit a regression equation to every feature within the same dataset. The output of this regression can provide reliable and relatively accurate statistics for estimating and exploring linear relationships. Linear relationships being either positive or negative. A linear relationship will be positive if an independent variable increasing will increase the dependent variable.[8] GWR results in output maps which enables scientists and researchers to visualize how each independent variable impacts the dependent variable spatially across the landscape (positively or negatively) and by how much. An example of GWR results will be discussed in the section below.

---

[5] Bivand, Roger. "Geographically Weighted Regression." October 2017. https://cran.r-project.org/web/packages/spgwr/vignettes/GWR.pdf.
[6] Legg, Robert, and Tia Bowe. "Applying Geographically Weighted Regression to a Real Estate Problem." 2009. Accessed February 2018. http://www.esri.com/news/arcuser/0309/files/re_gwr.pdf.
[7] "Regression Model and Spatial Autocorrelation." Geographic Information Systems Stack Exchange. Accessed February 2018. https://gis.stackexchange.com/questions/27245/regression-model-and-spatial-autocorrelation.
[8] "Regression Analysis Basics." ArcGis Pro. Accessed February 2018. http://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/regression-analysis-basics.htm.

Geographically Weighted Regression - Social Skill Score of Children, Vancouver

A geographically weighted regression was carried out for enumeration areas census data regarding the social skill score of children in Vancouver. The objective of the spatial analysis was to explore the relation between a child's social skills and a small set of variables related to the child and to their neighborhood. Using the explanatory regression analysis tool in ArcMap, the analysis picked out a small set of explanatory variables, to be used in carrying out the geographically weighted regression, which when combined resulted in the best fitness of the model. The variables of income, language and gender were selected to carry out an OLS output report and a GWR output of each of the parameter's estimated influence on the social skill score.

Map 1 (Appendix A) portrays the GWR results for the predicted local impact of income on the child social score skill as it varies over the enumeration areas of Vancouver. The areas in red shows where the variable of income has a negative impact on the social score whilst areas in green represent areas where income has a higher positive impact on the score. The estimated influence results are accompanied by the r^2 values of the GWR analysis, which show the 3 levels of correlation, or the levels of "fitness" of the predicted model compared to the observed values. Areas with dark blue dots have high correlation, and in such areas it can be said that the GWR model worked well. On the other hand, areas with light blue dots have a lower correlation and fit the model less accurately.

Map 2 (Appendix A) shows the local impact of language abilities on child social score. Areas with dark purple represent areas where language abilities have a higher positive impact on the score. Light purple areas represent spatial locations where the impact of language abilities is still positive yet less strong. R^2 values, seem to indicate that, like the income predictions, areas with higher positive impact have a more positive correlation between the estimated model and observed values.

In order to support the understanding of the outcomes of the GWR analysis, a Grouping analysis was further carried out using the Mapping Clusters tool within ArcMap. The grouping analysis included more variables: time spent on childcare, families of 4 or more, lone parent, recent immigrant and income. The output of the analysis resulted in 4 different groups which are portrayed on **map 3** (Appendix A). The results from the grouping analysis seem to support the fact that areas where income has a negative effect (map 1) are within group 4 which has the lowest income. Nevertheless, it is also intriguing to observe that while group 4 has the lowest income, it also has the highest childcare time spent. Therefore, even though the explanatory regression analysis tool did not present time spent on childcare as the most explanatory model, it could nevertheless be interesting to include that variable in the GWR to analyze the local spatial variation and impact of childcare on the social skill score across Vancouver.

The last map, **map 4** (Appendix A), shows the absolute difference between the estimates of the social skills scores calculated from the OLS and the GWR. It is possible to observe that areas with the biggest difference between the regressions' prediction are in East Vancouver. This implies that where the absolute difference is the biggest, the OLS regression was least successful in predicting the social skill score. A possible explanation for this is that for those feature with high absolute differences, the OLS global model may have failed to account for geographic variation and for the impact of neighbors around the features. Furthermore, by looking at the average of the predicted social scores by the OLS (77.26/100) and GWR (78.98/100) and comparing it to the average of the observed values (78.86), it can be seen that the GWR's predictions were more accurate than those of the OLS. Hence supporting the fact that the geographically weighted regression model is a more accurate predictor of the relationship between the dependent variable and the independent variables as it varies over the study area.

Applications of Geographically Weighted Regressions

As previously mentioned, geographically weighted regression can prove to be a useful tool to model, explain, and analyse spatial linear relationships between a dependent and independent variables. Most especially, GWR is useful since it accounts for spatial heterogeneity across a landscape. This method can be used to model and predict several relationships ranging from socioeconomic variables to more physical relationships, as illustrated in the following examples.

A typical example of how GWR can be used is when analysis of housing prices in cities. For instance, it is possible to explain spatial variations in home prices by looking at a range of real estate market data (Living area, land area, quality, structure type, renovation, garage, green space, traffic condition, view, or age) which are the independent variable. With the GWR output result, it is possible to model how each independent variable impacts housing prices across the study area, from a negative impact to a positive impact, and from a moderate to high impact. In this study using OLS would result in some neighborhoods being overestimated in price, whilst others would be underestimated. Using GWR results in the model predictions fitting the observed housing prices more accurately than an OLS output. [9]

Another example of the use of GWR is its use in studying and modelling socioeconomic trends. For instance, GWR was used to study the effects of population characteristics on fertility preferences in Egypt's eastern delta Governorates. The regression used socioeconomic characteristics (independent variables) such as household wealth status and woman educational levels to explain how fertility preferences varied across the region to reveal heterogeneity in fertility preferences in the study area. [10] The use of GWR can be critical concerning health or medical issues where spatial variation and heterogeneity are key.

---

[9] Huang, Bo. "Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices." *International Journal of Geographical Information Science* 4 (2010): 383-401. http://www.tandfonline.com/doi/abs/10.1080/13658810802672469.

[10] Mansour, Shawky. "The effects of Population Characteristics on Fertility Preferences in Eastern Delta Governorates, Egypt: A GIS Based of Spatial Local Modelling." *American Journal of Geographic Information System*4, no. 3 (2015): 105-20. http://article.sapub.org/10.5923.j.ajgis.20150403.03.html.

Other contexts in which GWR can be used is to predict the sources of some issues such as cholera occurrences. GWR is useful in explaining relationships but it can also be a good predictor in finding the underlying source of an issue - which can be very helpful for policy makers. A study made by Felix Ndidi Nkeki for the Department of Geography and Regional Planning, University of Benin, Benin City, Nigeria, revealed through the use of GWR that cholera occurrence was strongly associated with household sources of water supply. By using an exploratory analysis taking into account for heterogeneity in relationships across geographic space, it was possible to isolate the greatest source for cholera occurrence and thus to narrow down potential solutions for the outbreak. [11]

Furthermore, GWR can also be used to explain relationships for more physical geography issues. For example, GWR was used to analyze the factors which drive afforestation in Northern Vietnam. The study used locations of timber demand as dependent variable and heterogeneity such as population, distance, industry and forest as independent variables.[12]

Conclusion

In conclusion, the geographically weighted regression is a spatial extension of aspatial regression. Unlike an OLS, this regression goes beyond generating a global model and estimates local predictions of relationships between the independent variables and the dependent variables by considering the neighbors around each field within the area of study. GWR is useful is evaluating spatial heterogeneity across landscapes to model relationships such as health problems and the impacts of range of socio economic variables on such issues; GWR often leads to higher r^2 values/higher fitness of the model and its output and predictions are more accurate.

[11] Nkeki, Felix, and Animam Osirike. "GIS-Based Local Spatial Statistical Model of Cholera Occurrence: Using Geographically Weighted Regression." *Journal of Geographic Information System,*5 (2013). http://file.scirp.org/Html/2-8401301_40630.htm.
[12] Jamhuri et al, J. "GWR-PM - Spatial variation relationship analysis with Geographically Weighted Regression (GWR) - An application at Peninsular Malaysia." *IOP Conference Series: Earth and Environmental Science*27 (2016). http://iopscience.iop.org/article/10.1088/1755-1315/37/1/012032/pdf.

Works Cited

Bivand, Roger. "Geographically Weighted Regression." October 2017. https://cran.r-project.org/web/packages/spgwr/vignettes/GWR.pdf.

"GWR Explained." Geographically Weighted Modelling. Accessed February 2018. http://gwr.maynoothuniversity.ie/what-is-gwr/.

Huang, Bo. "Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices." *International Journal of Geographical Information Science* 4 (2010): 383-401. http://www.tandfonline.com/doi/abs/10.1080/13658810802672469.

"Introduction to Geographically Weighted Regression." 2009. https://www.bristol.ac.uk/media-library/sites/cmpo/migrated/documents/gwr.pdf.

Jamhuri et al, J. "GWR-PM - Spatial variation relationship analysis with Geographically Weighted Regression (GWR) - An application at Peninsular Malaysia." *IOP Conference Series: Earth and Environmental Science*27 (2016). http://iopscience.iop.org/article/10.1088/1755-1315/37/1/012032/pdf.

Legg, Robert, and Tia Bowe. "Applying Geographically Weighted Regression to a Real Estate Problem." 2009. Accessed February 2018. http://www.esri.com/news/arcuser/0309/files/re_gwr.pdf.

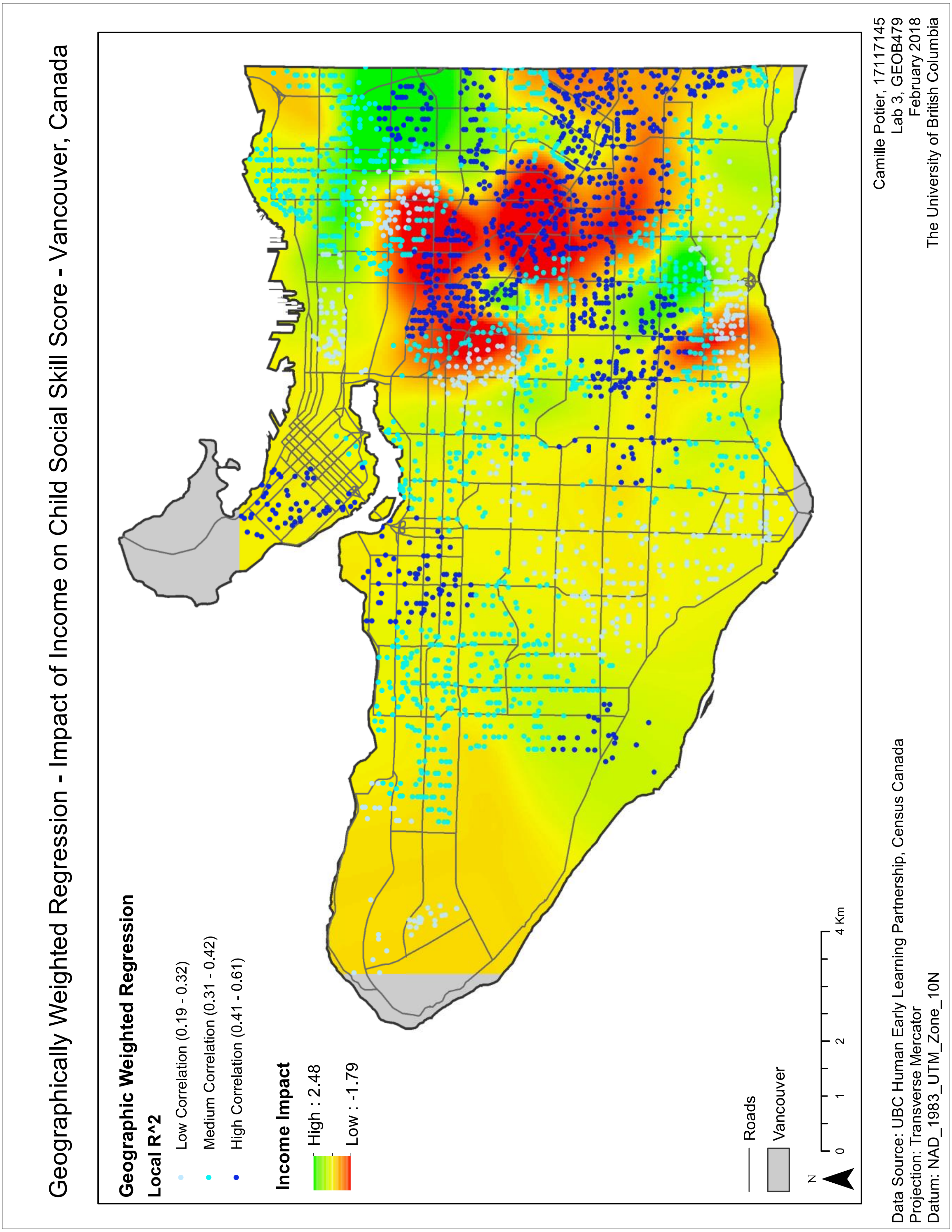
Mansour, Shawky. "The effects of Population Characteristics on Fertility Preferences in Eastern Delta Governorates, Egypt: A GIS Based of Spatial Local Modelling." *American Journal of Geographic Information System*4, no. 3 (2015): 105-20. http://article.sapub.org/10.5923.j.ajgis.20150403.03.html.

Nkeki, Felix, and Animam Osirike. "GIS-Based Local Spatial Statistical Model of Cholera Occurrence: Using Geographically Weighted Regression." *Journal of Geographic Information System,*5 (2013). http://file.scirp.org/Html/2-8401301_40630.htm.

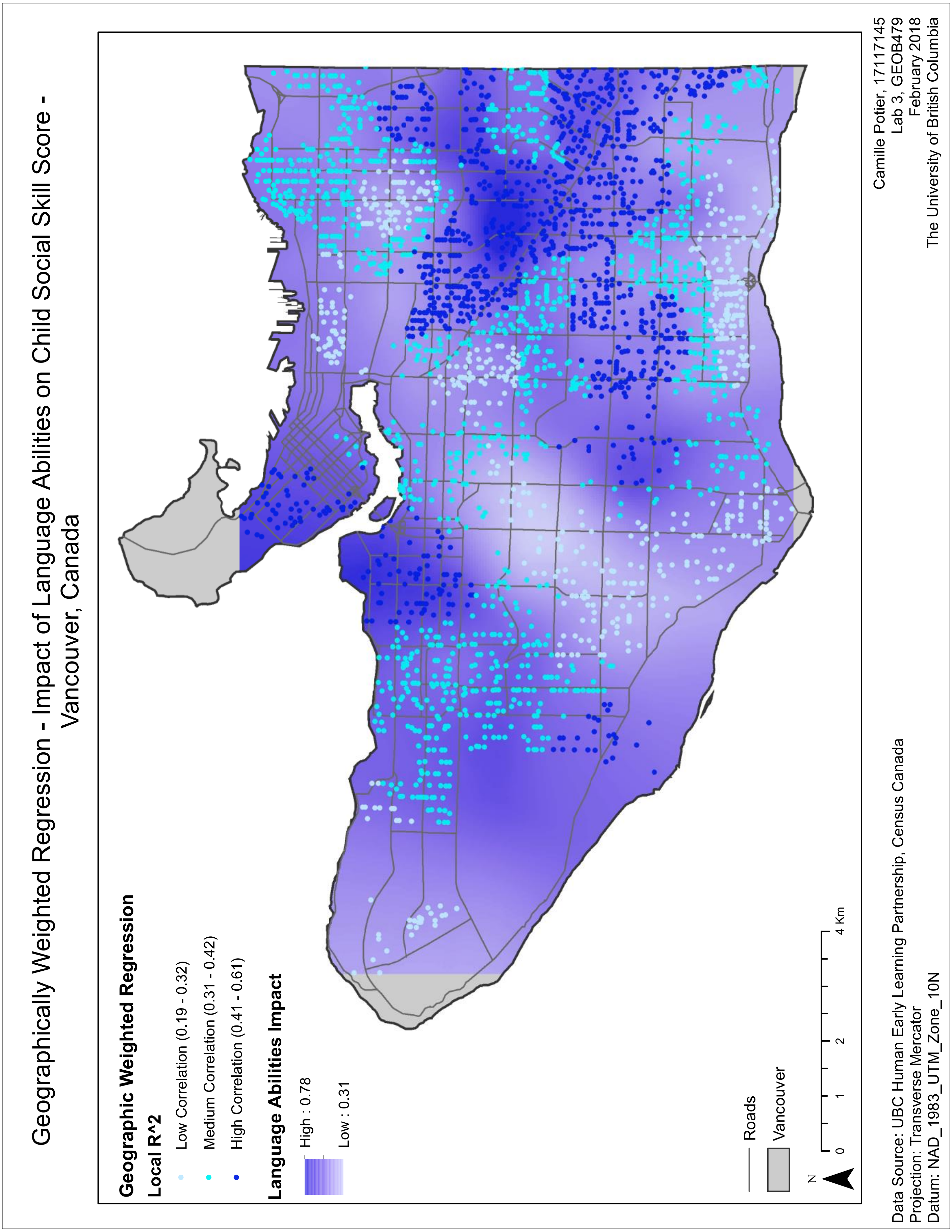
"Regression Analysis Basics." ArcGis Pro. Accessed February 2018. http://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/regression-analysis-basics.htm.

"Regression Model and Spatial Autocorrelation." Geographic Information Systems Stack Exchange. Accessed February 2018. https://gis.stackexchange.com/questions/27245/regression-model-and-spatial-autocorrelation.

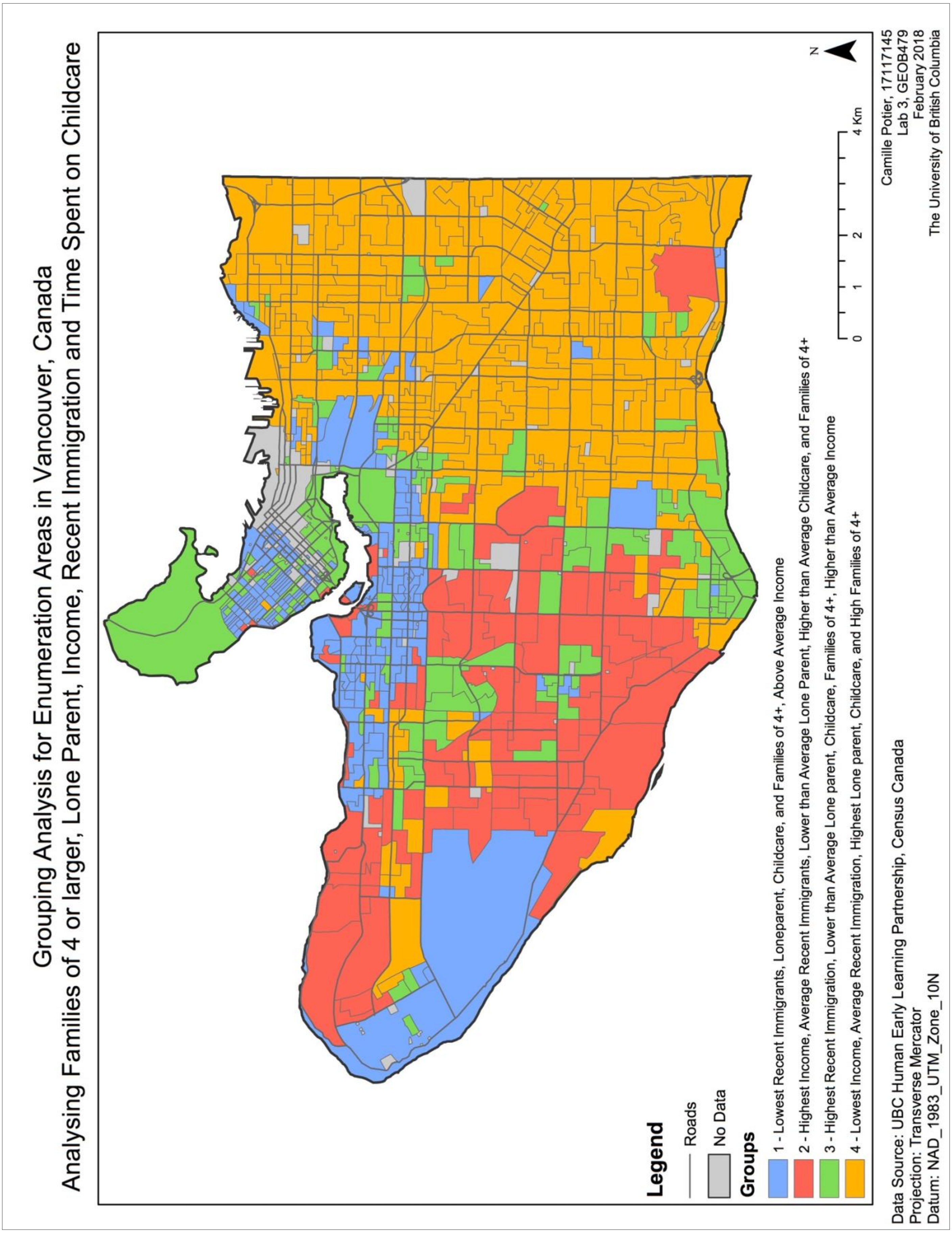
Appendix A – Map 1



Appendix A – Map 2



Appendix A – Map 3



Appendix A – Map 4