Habitat International 108 (2021) 102319

Contents lists available at ScienceDirect

Habitat International

journal homepage: http://www.elsevier.com/locate/habitatint

Machine learning for analysis of wealth in cities: A spatial-empirical examination of wealth in Toronto

Eric Vaz [a,*], Fernando Bação [b], Bruno Damásio [b], Malik Haynes [c], Elissa Penfound [c]

[a] Ryerson University Department of Geography and Environmental Studies, Toronto, ON, Canada
[b] NOVA University Lisbon NOVA Information Management School, Lisbon, Portugal
[c] Ryerson UniversityYeates School of Graduate Studies, Toronto, ON, Canada

ARTICLE INFO

Keywords:
GIS
Regional analysis
Spatial analysis
Wealth habitats
Regional disparities
Toronto

ABSTRACT

Wealth in the Greater Toronto Area (GTA) continues to grow each year as Toronto's consumer market and population increase. Using a machine learning segmentation based on self-organizing maps, this paper examines the demographics, socioeconomics, and expenditure consumption patterns of the GTA's consumers. The results suggest that SOM may contribute to efficient spatial delimitation tools, enhancing the spatial patterns of clusters in the city of Toronto. The relation to urban areas displays locational neighbourhood characteristics, where the accumulation of wealth is present, pointing out a striking spatial-morphological division between census regions and geographical distribution of wealth in Toronto. In this sense, concerning regional and urban habitats, SOM position themselves as promising tools to measure wealth within highly dense urban cores with significant demographic diversity. While cities that have witnessed rapid urbanization and population growth, such as Toronto, may benefit from integrative methods that use machine learning and spatial analysis to monitor regional and urban disparities.

1. Introduction

While the geography of economic and social wellbeing and resulting inequality has had a long standing in the literature of regional and urban studies (Glaeser, Resseger, & Tobio, 2009; Vaz, Anthony, & McHenry, 2017; Wei, 2017 The specificities of the geography of wealth is often an overlooked topic within regional science and economic geography (Sachs, Mellinger, & Gallup, 2001). Urban regions across the world are emerging and many individuals and families are situating themselves in large cities or as close to large cities as possible (Glaeser & Kahn, 2004). The unaffordability of living in Toronto, for instance, has led to an excess commuting throughout the Greater Toronto Area (Buliung & Kanaroglou, 2002). (Tables 1–5, Figs. 6–8)

Several cities are becoming too expensive to live in, and as a result, only wealthy individuals and families are in a position to occupy large cities (Akbar, Rolfe, Hossain et al., 2017; Deng, Quin, & Wu, 2019). Maximization of wealth and the diminishing middle-class seems to be an ongoing concern within urban cores (Condon, 2020). Wealthy consumer behaviour is worth analyzing in a region that is witnessing unprecedented growth, such as the Greater Toronto Area (Vaz & Jokar Arsanjani, 2015). The Geography of its wealth underlies certain economic drivers relating to a strong pattern of skilled immigration, as well as abundance of resources and a hotspot for business and entrepreneurship as incentives (Buliung, Hernandez, & Mitchell, 2007). The City of Toronto (Lemon, 1991), North America's fourth largest market, has become a place with an accumulation of economic growth and increasing availability of big-box stores (Jones & Doucet, 2000). Consumer behaviour has always been something compelling and vital to the world of retail, marketing, and services across all industries, albeit seldom assessed within the larger extent of the dynamics of machine learning and cities. Through machine learning, a series of clustering methods integrating SOMs will be used to compare and understand the clusters shaped in the Greater Toronto Area. The applications of SOM are countless, with numerous variants and extensions. In 2001 more than 4000 research articles (Kohonen, 2001, p. 501) were published, making it, by far, the most popular unsupervised neural network. It is difficult to overstate the utility, flexibility and impact of this tool, in science and industry alike. From the interpretation of patterns of gene expression, to the classification of extreme climate events, organization of large collections of documents, water quality monitoring, the SOM has had an important role in improving our ability to understand and make sense of many complex datasets and problems. Fields like statistics, financial

* Corresponding author.
E-mail address: evaz@ryerson.ca (E. Vaz).

https://doi.org/10.1016/j.habitatint.2021.102319
Received 3 June 2020; Received in revised form 21 November 2020; Accepted 5 January 2021
Available online 21 January 2021
0197-3975/

Table 1
Visible minorities by cluster.

Visible Minority	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Arab	1.42	0.03	1.38	0.00	0.54	0.23
Black	11.39	0.11	4.59	0.00	1.61	0.70
Chinese	9.57	19.15	8.68	0.00	5.47	11.83
Filipino	5.75	0.05	2.64	0.00	2.37	1.89
Korean	0.56	0.00	1.26	0.00	1.09	0.44
Latin American	3.41	0.00	1.61	0.00	1.02	0.56
South Asian	18.59	0.00	10.47	0.00	3.30	4.20
Southeast Asian	2.26	0.32	1.18	0.00	0.40	0.27
West Asian	1.61	0.03	1.76	0.00	1.39	3.04
Other	2.11	0.24	0.96	0.00	0.23	0.28

Table 2
Immigration by cluster.

Immigration	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Immigration Before 2001	29.38	39.35	24.32	0.00	21.28	20.11
Immigration Between 2001 and 2005	7.81	6.27	4.78	0.00	3.22	2.69
Immigration Between 2006 and 2011	8.45	4.29	3.99	0.00	3.80	3.88
Immigration 2012 to Present	7.56	9.27	6.32	0.00	4.92	7.40
Immigrant	53.21	59.19	39.41	0.00	33.22	34.08

analysis and medicine are amongst the areas in which SOM has had a significant impact in tackling the challenges presented by high-dimensional nonlinear datasets (Qi, Liu, Liu, & Zhang, 2019). Its ability to cluster and explore multi-dimensional distributions to uncover the underlying data structures, along with the associated visualization tools has led to its application in almost every scientific field (Abarca-Alvarez, Navarro-Ligero, Valenzuela-Montes, & Campos-Sánchez, 2019; Chen, Chen, Ma, & Chen, 2019; Kala, Atkinson, & Tiwari, 2020; Wickramasinghe, Amarasinghe, & Manic, 2019).

Human and social sciences are no exception with many different applications, where SOM's have proven to be an invaluable tool to improve our understanding of complex multidimensional problems such as the structure of welfare and poverty between nations (Kaski & Kohonen 1996), the comparison of the profile of smart cities (Kourtit, Nijkamp, & Arribas, 2012), to analyze spatial interaction (Yan & Thill, 2009) or in geographical information science problems. Today, SOM's continue to be a valuable tool to shed light on complex problems like the analysis of wealth in cities, and giving us the opportunity to consider the multidimensional nature of such problems.

The choice of combining spatial clustering to machine-learning leads to a series of research questions that pave the framework of this paper within three key silos:

(i) Heterogeneity: Are city cores such as Toronto heterogeneous, and leading to less diversified geodemographic landscapes?
(ii) Econocentrism: Does the existing economic disparities directly link to environmental injustice in the Greater Toronto Area?
(iii) Regional intelligence: Is there a gap between artificial intelligence, local spatial decision support systems (Vaz, 2020) where applied urban decisions can be addressed through a geodemographic approach with machine learning and urban decision making through ubiquitous datasets (Vaz, 2020)?

These research questions were answered by combining geodemographic segmentation system involving a series of steps as outlined in the methods section. Combining geodemographic segmentation with integrative methods to examine wealth is of utmost importance for demographic studies, as wealth is defined by the overall assets of Canadian families, and therefore an indirect measure of economic prosperity, equity, and livability (Sharanjit & LaRochelle-Côté, 2015). While the literature of geography of wealth continues to grow, and has attracted the interests of many geographers, marketers, and those in the location intelligence and retail geography sector at city scale, limited work has been assessed using machine learning techniques. The following section defines the study area of the Greater Toronto Area, followed by an exploratory approach to the accumulation of wealth in the GTA. Following the study area section, the paper introduces a combined

Table 4
Dwelling type by cluster.

Dwelling Type	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
High Rise	37.39	20.00	15.60	0.00	18.01	17.74
Low Rise	11.44	0.00	8.67	0.00	10.90	5.46
Detached Duplex	5.99	0.00	3.51	0.00	2.56	2.17
Other	0.14	0.00	0.20	0.00	0.00	0.00
Semi Detached	7.63	16.19	7.85	0.00	2.60	0.44
Single Owned	28.63	54.29	53.63	0.00	64.37	72.91
Tenure Owned	56.96	62.88	79.03	0.00	77.98	89.39

Table 5
Children at home by cluster.

Children at Home	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
0–4 Years Old	17.70	39.06	16.51	0.00	13.55	8.81
10–14 Years Old	15.21	18.24	16.67	0.00	20.41	24.54
15–19 Years Old	16.06	10.85	17.36	0.00	20.13	23.29
20–24 Years Old	15.75	0.69	17.04	0.00	18.49	19.89
25+ Years Old	18.36	13.62	16.02	0.00	11.61	12.32

Table 3
Education by cluster.

Education	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Apprenticeship, Trades, College	6.40	24.38	5.47	0.00	2.61	0.21
High School Certificate or Equivalent	27.09	9.86	23.54	0.00	16.32	15.06
No Certificate, Diploma, Degree	23.45	19.08	14.83	0.00	8.46	9.15
Bachelor's Degree	14.01	16.04	21.32	0.00	32.81	36.33
Above Bachelor's	8.38	8.61	12.88	0.00	24.87	27.38

machine-learning approach through SOM that distills clusters for wealth in the urban core of Toronto. This leads to the discussion section, where the identified clusters are discussed in regard to their demographic characteristics. Finally, the paper ends with a conclusion section offering insights on the implementation of SOM in line with the location dynamics of heterogeneous urban profiles and regions.

2. Study area

2.1. Potential growth of the Greater Toronto Area

The Greater Toronto Area (GTA) is the most significant and populated region within Canada, and is a financial hub occupied by different types of economic wealth within the city nexus. Toronto is one of Canada's major urban centers and is the financial and cultural capital of the country (Harris, 2012). The GTA corresponds to Canada's largest market, and is known as a global city-region that possesses the flows of international economics, culture, and migration-all of which play a vital role in the city's diversity (Rosen & Walks, 2015). The GTA is home to approximately 2.6 million individuals and happens to be one of the most accessible cities in North America, located only a 1-h drive away for five million Canadians, and only a 90-min flight for sixty percent of North America's population. Although the GTA is a diverse region, from ethnicity, income, languages, and more, the main difference between types of neighbourhoods such as, gated and non-gated is between high-income households (Walks, 2014). Suburbs in the outskirts of the City of Toronto have become more diverse in social composition and income levels (Moos & Mendez, 2015). For the sake of studying the Wealth of Geography within the GTA, typically, the neighbourhood level is often analyzed, such as census tracts (CT). Neighbourhoods within the GTA vary significantly in size and demographic profiles.

2.2. Minorities and immigration

The Toronto region is home to most of Canada's visible minority groups as it is a place that offers a great deal of opportunities, followed by other popular destinations for immigrants such as, Vancouver and Montreal. Statistics demonstrate that there has been a large growth in visible minorities and immigrants in the Toronto Census Metropolitan Area (CMA), which is Canada's largest CMA[1] (Gyimah, Walters, & Phythian, 2005). Of the 1,266,005 immigrants living in Toronto in 2016, 187,945, or 14.8 per cent were recent immigrants, having landed in Canada in the previous five years (from 2011 to May 2016). Over the recent years, Toronto has become home to the largest number of recent immigrants of any Canadian city. In 2016, Toronto was home to 17.5 per cent of all recent immigrants to Canada, while Toronto comprises 7.8 per cent of the country's population. Many immigrants, visible minority groups, as well as Canadian-born citizens are in the upper five and one percent income groups. In the GTA, there are several young millionaires and wealthy individuals. Wealth is commonly known as net worth, as well as financial assets, savings, stocks, bonds, and nonfinancial assets including real estate, vehicles, and more (Maroto & Aylsworth, 2016). However, most wealthy households within the GTA are composed of middle-aged individuals. Wealth varies over an individual's life cycle, and increases as savings are accumulated, which is why there is a peak in those in the age bracket of 55–64 who are wealthy (Daniel & Bright, 2011). Other studies agree that top income earners in Canada tend to be in the age brackets of 37–49 and 50–64 and top income earners in the 50-64-year-old age bracket have increased significantly over the past 20 years, where in the year 2006 the average household income was $155,000 with a mean income of $344,000 (Breau, 2014). A strong correlation between earning and education-attainment is present. Incomes ranging from $150,000 to $250,000 are considered high-income earners (Milligan & Smart, 2015). In the study, $150,000 per household income is the minimum threshold which will be used to analyze the wealth within the GTA). In research conducted many years ago, authors such as, Stanley, considers individuals earning six-figure incomes, and those with one million dollars or more in network, or both, as high-income individuals and consumers (1988). Canada's wealthiest individuals increased their total share of national income between 1993 and 2008.

2.3. Accumulation of wealth in the Greater Toronto Area

The GTA has many immigrants and visible minority groups who have migrated from several countries around the world. For these immigrants as well as Canadian-born families, homeownership is important. In Canadian society, homeownership demonstrates a cultural importance and provides opportunities for accumulating wealth. Homeownership in Canada, especially in the GTA is an important component of financial success, as many homes within the GTA sell for millions of dollars. Many expensive homes within the GTA are concentrated in Chinese and Italian neighbourhoods such as Little Italy, the Bridle Path-as well as the City of Vaughan, which has many Italian neighbourhoods, and Scarborough and Markham, which have many Chinese neighbourhoods. Studies have shown that within the Toronto CMA, Chinese and Italians have the highest rates of homeownership, meanwhile Blacks and Aboriginals have the least (Gyimah et al., 2005). High-income neighbourhoods within the City of Toronto are mainly found in areas that are well-established within the inner-city, as well as gentrifying and regentrifying neighbourhoods (Breau, Shin, & Burkhart, 2018). Most high-income neighbourhoods within the GTA consists of single-detached homes, as these are the most expensive homes purchased within Toronto in neighbourhoods such as, the Bridle Path, Forest Hill, Rosedale, and others. However, in the past decade, there has been a rapid increase of condominiums built throughout the GTA. Some of the City of Toronto's wealthiest residential condominiums is Yorkville, where the average price per square foot is $1000, as well as other residential condominiums in downtown core of the city. Residential condominiums continue to emerge within the GTA, particularly in the City of Toronto. A large proportion of the condos in the GTA are purchased by immigrants and foreign investors who remain in other countries. The Geography of Wealth within the Greater Toronto Area demonstrates diversity among homeownership and Canadian-born citizens, immigrants and more. Fig. 1 shows a map of the population density per CT for the study area.

Fig. 1. Population density of the greater Toronto area (Km^2).

[1] According to Stats Canada, a CMA consists of one or more neighbouring municipalities situated around a core. A census metropolitan area must have a total population of at least 100,000 of which 50,000 or more live in the core.

For visualization purposes, a representation of quantile classification was chosen to allow equal representation on the map and easier visualization of the difference between Toronto's downtown core and the spread of population distribution.

3. Methods

3.1. Data

The data collected for the geodemographic segmentation system was derived from SimplyAnalytics.[2] The data source is used for creating a sophisticated geodemographic segmentation system as it provides valid variables that will allow for a strong analysis and results that are both reliable and accurate. The following groups of data were used with corresponding subcategories as explored in the result section: (1) Total population, (2) Visible Minorities, (3) Household Income, (4) Educational attainment, (5) Occupation, (6) Immigration, (7) Total number of households, (8) Children at home, (9) Marital status, (10) Dwelling type, (11) Home ownership, (12) Average Total Expenditure, (13) Average Total Expenditure per expenditure type, (14) Employment status, (15) Official language spoken.

For this study, a geodemographic segmentation system is conducted using CTs as the level of geography. CTs within the Greater Toronto Area allow to assess the similarities and disparities of the spatially-defined clusters. Canadian census tracts (CT) are small geographical units that are used to represent neighbourhoods and communities within a census metropolitan area and have a population within a range of 2500–8000 individuals (CMA) (Chen et al., 2011). The source data includes a total of eighty-three variables encompassing the following topics: (i) Minorities, (ii) Income, (iii) Education, (iv) Occupation, (v) Immigration, (vi) Children at home, (vii) Marital status, (viii) Dwelling type (ix) Average food expenditure, (x) Average household expenditure, (xi) Total households.

3.2. Self-organizing maps

For many years, artificial neural networks have been studied and used to construct information processing systems, inspired by natural biological neural structures. Among the various existing neural network architectures and learning algorithms, Kohonen's self-organizing map (SOM) is one of most popular neural network models. Teuvo Kohonen proposed the SOM in the beginning of the 1980s (Kohonen, 1982) as a result of his work on associative memory and vector quantization.

One of SOM's objectives is to extract information and allow the visualization of the fundamental structures in a dataset, through a map, resulting from an unsupervised learning process (Kaski, Honkela, Lagus, & Kohonen, 1998). Due to its ability to preserve the topological relations, the SOM is frequently used as a tool for mapping high-dimensional data into two-dimensional feature maps. The most significant advantage of the SOM is that it allows for insight on the underlying structure of the data, through the analysis of the resulting map. Preserving the topology, the map allows the extraction of the latent structure of the input space, which is valuable in many tasks (e.g. dimensionality reduction, data visualization, clustering and classification).

Various extensions of the SOM have been devised to extend its application to a wide range of applications. In particular, SOM have been used extensively in geospatial problems, and a good overview of these is presented in (Agarwal & Skupin, 2008). Openshaw was one of the first geographers to point out the applicability of SOM in geography, namely for clustering (Openshaw, Blake, & Wymer, 1995). Other geospatial clustering applications of SOM include (Koua & Kraak, 2004; Spielman & Thill, 2008). SOMs (Allouche & Moulin, 2005; Sester, 2005) are used for cartographic generalization. SOMs have also been used as supervised classification tools for geospatial problems, for example in (Wan & Fraser, 1993). In (Spielman & Thill, 2008) SOM are used for data reduction, allowing the detection of spatial patterns in a socio-demographic analysis of New York City census data. Similar uses include the analysis of airline passenger flows (Yan & Thill, 2008) and linguistic variations (Yan & Thill, 2009). More recently (Lee & Rinner, 2015), used the SOM to identified trends Toronto's neighbourhood diversity and urban social change dynamics in Toronto.

A detailed explanation of the SOM algorithm is outside the scope of this paper and the reader is referred to (Kohonen, 2013) for more details. Nevertheless, it is important to provide a high-level description of the SOM algorithms. The basic idea of a SOM is to map the data patterns onto an n-dimensional grid of units, also called neurons. The grid is usually known as output space, as opposed to the input space that is the original space of the data patterns. The mapping process tries to preserve topological relations in the data, i.e. patterns that are close in the input space will be mapped to neighbouring units in the output space and vice versa. The output space is usually two-dimensional and most of the implementations use a rectangular or hexagonal grid of units. Each unit, being an input layer unit, has as many weights as the input patterns and can thus be regarded as a vector in the same space of the patterns. During the training of a SOM with a given input pattern, the distance between that pattern and every unit in the network is calculated. The closest unit to the input pattern is selected as the winning unit, also known as the best matching unit, and that input pattern is mapped onto that unit. Then patterns that are close in the input space will be mapped to that unit, or to other units that are close in the output space. Thus, allowing for the "topology preserving" property, in the sense that input space neighbourhoods are preserved in the mapping process.

For our purposes the SOM can be a very useful tool, as it allows for extracting information about the manifold structure of highly dimensional input space we are using and, in the process, identifying a large variety of clusters. Additionally, the SOM topological map produced by the algorithms allows for the exploration of the results through visual analysis.

4. Discussion

The results point to a disaggregation between affluent and less-affluent areas. While the GTA strives for a homogeneous distribution of wealth, machine learning can support the detection of spatially-explicit wealth and poverty patterns among large data aggregations within city cores. A total of six clusters were identified. Georeferencing of these clusters allowed to explore the spatial distribution within the Greater Toronto Area (Fig. 2) (Fig. 3).

The results of this analysis indicate that the vast majority of the population reside in areas comprised of low-income (cluster 1) and mid-income (cluster 3) earners. Locations with low-income populations are predominantly located in the east and west peripheries of the City of Toronto and the locations with mid-income populations are located throughout the downtown core of the City of Toronto, north of the City of Toronto (Vaughan and Richmond Hill) and throughout the periphery of the Toronto Census Metropolitan Area. Locations with a relatively high-income population (cluster 5) are predominantly located in mid-to north-central City of Toronto with some census tracts located in Oakville, Mississauga and Caledon. Relatively low-income (cluster 2), and high-income (cluster 6) populations are located in a low number of census tracts, relatively low-income populations are found in eastern Oakville, western Vaughan and north-east Toronto, and high-income populations are found in central Toronto. An additional mid-income cluster (cluster 4) represents a small portion of the population and is only represented in one census tract (located in south-western Toronto). Generally, both the highest-income and lowest-income populations are

[2] SimplyAnalytics is a web-based mapping, analytics, and data visualization application that makes it easy for anyone to create interactive maps, charts, and reports.

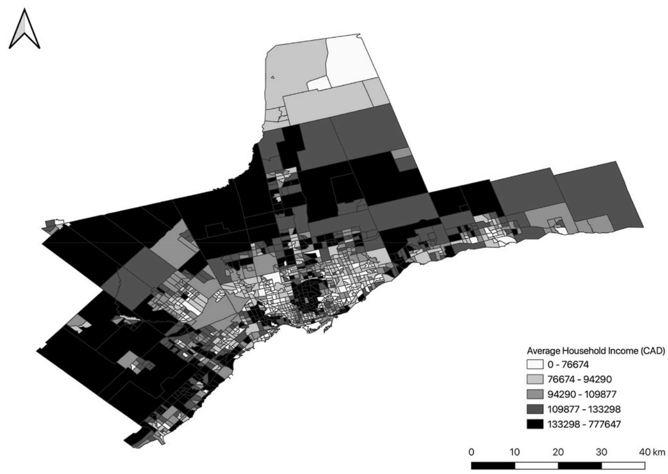

Fig. 2. Average household income in the greater Toronto area (CAD).

Fig. 3. Immigrants in the greater Toronto area (% of population).

found within the City of Toronto with the high-income populations located within central Toronto, definable as Toronto Urbanites (Fig. 4), and the low-income populations located throughout the peripheries of Toronto (Fig. 5).

The intrinsic functionalism found within the Toronto Urbanites,

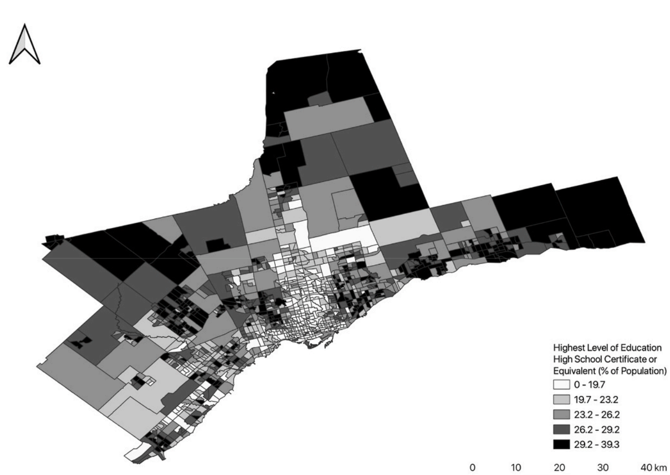

Fig. 4. Highest level of education high school certificate or equivalent in the greater Toronto area (% of population).

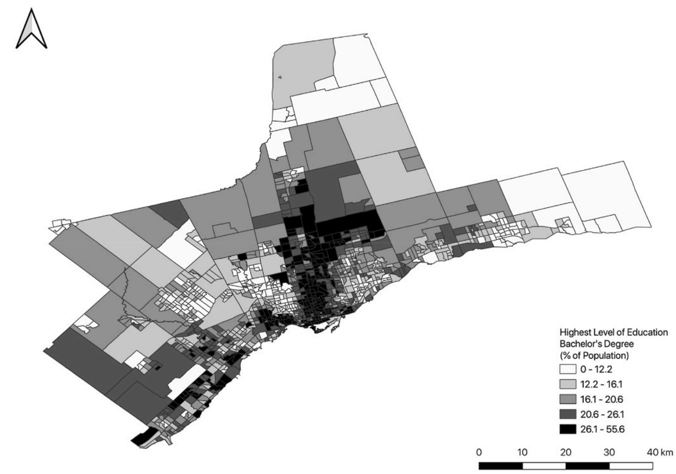

Fig. 5. Highest level of education Bachelor's degree in the greater Toronto area (% of population).

Fig. 6. Tenure owned households in the greater Toronto area (% of population).

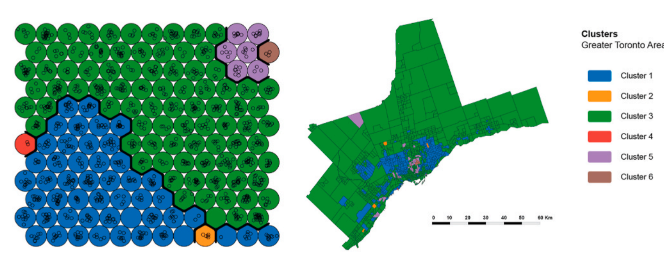

Fig. 7. Spatial distribution of SOM clusters in Toronto.

corroborates with the notion of functional cities, and the primary intention of limited spatial units, but with an increasing amount of accessibility (Guest & Lee, 1984). This has been the tendency of a growing amount of smart cities in Europe and North America, however, this brings some concerns for the distribution of wealth, and the potential of creating added environmental injustice (Vaz et al., 2017).

Visible minorities, to varying degrees, are present in all clusters (except cluster 4), however the highest total population of visible minorities are represented in cluster 1 (low-income).

Additionally, when viewed as a percentage of the overall population per cluster, visible minorities comprise a much higher percentage of the population in low-income clusters compared to high-income clusters.

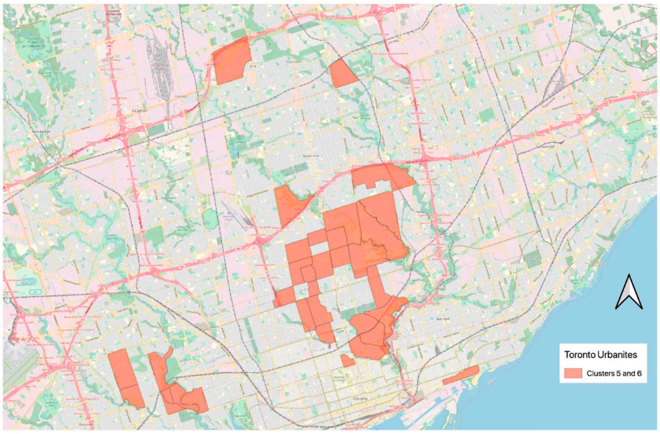

Fig. 8. Census tracts of high-income clusters 5 and 6.

Like visible minorities, immigrants are present in all clusters (except cluster 4), and also have the highest total population represented in low-income clusters (clusters 1 and 2). Again, when viewed as a percentage of the overall population per cluster, immigrants comprise a much higher percentage of the population in low-income clusters compared to high-income clusters. It should be noted that immigrants arriving in the Toronto Census Metropolitan Area prior to 2001 appear in much higher population numbers in the high-income clusters than immigrants arriving after 2001. However, immigrants arriving prior to 2001 appear in much higher population numbers in all clusters compared to immigrants arriving after 2001. Indicating that immigrants arriving prior to 2001 don't necessarily have greater economic success overall compared to immigrants arriving after 2001, there simply are just higher numbers of them. These findings are consistent with the findings from a 2018 study which analyzed the spatial dimensions of income imbalance and showed that visible minority and recent immigrant populations were more likely to be found in low-income neighbourhoods (Breau et al., 2018). These indicate that of the visible minorities included in this study, Chinese were more likely to be found in high-income clusters compared to other visible minorities and Blacks and Southeast Asians were more likely to be found in low-income areas compared to other visible minorities.

Perhaps not surprisingly, high-income cluster populations have more occurrences of higher educational degrees (bachelor's or above bachelor's), while low-income cluster populations fewer occurrences of higher educational degrees, with increased occurrences of high school certificate holders and apprenticeships/trades college degrees. Higher education shares a link with higher income levels through enhanced workforce opportunities and access to careers with higher incomes, and the findings in this study are consistent with this link.

Regarding the structure of the dwellings that are found in each cluster, high-rise and low-rise dwellings are more likely to be found in low-to mid-income clusters, specifically in areas that are within or on the periphery of the City of Toronto. Whereas single detached homes were more likely to be found in high-income clusters. Additionally, home-ownership was much more prevalent in mid-to high-income clusters compared to low-income clusters. Indicating that mid-to high-income populations are more likely to attain home ownership than low-income populations.

Regarding the age of children living at home, low-to mid-income clusters have higher numbers of children aged 0–4 living at home, whereas mid-to high-income clusters have higher numbers of children aged 10–24 living at home. This indicates that the low-to mid-income clusters represent a younger population compared to mid-to high-income clusters which represent an older population. Interestingly, older children (25+) living at home are found higher numbers in low-income clusters compared to mid-to high-income clusters. Indicating that low-income families are more likely to have adult children with less financial independence than mid-to high-income families and/or that mid-to high income families are better equipped to financially support their adult children.

An additional aspect of this study was to analyze the expenditure of the population by cluster to characterize how households are spending money. Perhaps not surprisingly, high-income clusters are spending more money overall compared to low- and mid-income clusters, and low-income clusters are spending less money overall compared to mid- and high-income clusters. Low- mid- and high-income clusters all have higher expenditures on food, shelter and transportation, with notably low expenditure on restaurant (breakfast), mortgage, and daycare expenses. This finding indicates that all populations are spending more money on essential expenses that are necessary for day to day life. Interestingly, there isn't much variance between clusters on phone service, games of chance, and rent expenditures. This indicates that, in regard to phone service and games of change, level of income does not seem to alter the likelihood of an individual obtaining these services. In regard to rent, the results indicate that low-income individuals are more likely to rent than own a property compared to mid-to high-income individuals. In addition to expenditures on food, shelter and transportation, high-income clusters are also spending money on restaurants (dinner and lunches), vacation homes, clothing, education, furniture, furnaces, household equipment, personal care products, recreation, alcohol and tobacco. Indicating that overall high-income individuals are spending more money on non-essential items and services.

Overall, the results of this study indicate that generally there are socio-economic trends that can be found in low- mid- and high-income clusters. Low-income clusters can be generally categorized as having; higher numbers of visible minorities (with Blacks and Southeast Asians comprising the highest numbers of visible minorities); higher numbers of recent immigrants; lower levels of education; a higher likelihood of living in high- and low-rise dwellings and are less likely to own their home; a higher likelihood of having young children as well as young adults living at home; and an overall smaller expenditure compared to mid-to high-income clusters. Mid-income clusters can be generally categorized as having; moderate numbers of visible minorities (with Chinese and South Asians comprising the highest numbers of visible minorities); moderate numbers of recent immigrants; higher levels of education (at least a bachelor's degree); a higher likelihood of home ownership; a moderate likelihood of having children of any age (0–25+) living at home, and an overall moderate expenditure compared to high-income clusters. High-income clusters can be generally categorized as having; lower numbers of visible minorities (with Chinese comprising the highest numbers of visible minorities); moderate numbers of recent immigrants; higher levels of education (bachelor's degree and above bachelor's degree); a very high likelihood home ownership; and a higher likelihood of having children age 10–24 living at home and an overall higher expenditure compared to low-to mid-income clusters.

Economic exclusion is a factor that prevents upward economic mobility in society, and it prevents and perpetuates social and economic divides. In Canada, length of residency and visible minority status are some of the factors that can be considered indicators of economic exclusion (Lightman & Gingrich, 2012). This is consistent with the findings of this study which has shown that low-income clusters are more likely to have higher numbers of visible minorities and recent immigrants.

In addition to the issue of economic exclusion is the issue of income segregation. Income segregation is referred to as the concentration of low-income households into low-income neighbourhoods and high-income households into high-income neighbourhoods. Income segregation can perpetuate issues such as low levels of education, high levels of crime and poorer health of low-income neighbourhoods (Chen, Myles, & Picot, 2010). Between the years 1980–2012 the City of Toronto has experienced an increase in income segregation at a rate that is higher than other Canadian cities. In the City of Toronto this can in part be attributed to the high number of rental units and social housing properties coupled with highly wealthy neighbourhoods (Walks Dinca-Panaitescu, & Simone 2016). Additionally, increasing home ownership and real-estate values have allowed for greater benefits to high income earners, as low-mid income earners have taken on higher debts to enter the real estate market. This disparity has further increased income segregation as the benefits of increasing real estate values are spatially unequal (Walks, 2016).

Increases in inequality often lead to elevated intergenerational inequality, children raised by high-income parents are more likely to become high earners themselves compared to children raised by low-income parents (Corak, 2013). This concept coupled with the issues of economic exclusion and income segregation highlights the importance of identifying the spatial characteristics of wealth in a major urban center like Toronto. As previously mentioned, income segregation in Toronto has increased over the past several decades, and this increase is likely contributing to the perpetuating cycle of low-income individuals and families being unable to attain income equality. Additionally, these low earners are more likely to be comprised of recent immigrants and visible minorities compared to high earners. The low-income clusters and CT's identified in this study are important to highlight as they are more likely to have populations that will struggle to attain higher earnings, higher levels of education, home ownership and attain a greater ability to purchase non-essential items and services.

5. Conclusions

The GTA continues to be Canada's largest consumer market, which is growing rapidly and becoming wealthier over time (Vaz, Shaker, & Cusimano, 2020). Geodemographic segmentation systems allow for a sophisticated analysis of consumer behaviour within the GTA. Over the past twenty-five years, geodemographics has proven itself as a valuable technique used for customer analysis, and market planning, which involves consumer behaviour and location assessment. Geodemographic classifications are categorical measures created through algorithms in computer software programs, which group small geographic areas into clusters based on similar characteristics from data (Rae & Singleton, 2015). Geodemographic segmentation systems is based primarily on cluster analysis, which allows small neighbourhoods and areas to be classified into groups that share characteristics which are more similar than dissimilar. Geodemographics analyzes people according to where they live and makes it possible to be able to say something about the characteristics of individuals. Geodemographics segmentation systems are essential tools for classifying small areas using a variety of socioeconomic data, such as lifestyle and consumer behaviour (Abbas, Ojo, & Orange, 2009).

Geodemographic segmentation systems are applied across many industries and sectors, which need to understand their target market, or potential target market.

Within a geodemographic segmentation system, clusters are identified from a distinctive set of attributes, for example wealthy neighbourhoods in an area of study. Geodemographics assumes that neighbourhoods that share the same cluster consist of like-minded individuals, which share both socioeconomic and demographic characteristics. Geodemographics provides a sophisticated technique for understanding the demographic characteristics and behaviours of individuals who are classified in the same cluster (Xiang, Stillwell, Burns, Heppenstall, & Norman, 2018). The use of spatial analysis and statistical techniques are part of geodemographic classifications, which have become firmly established in the field of geodemographic analysis (Páez, Trépanier, & Morency, 2011).

Geodemographic techniques are often applied to understanding consumer behaviour which can allow users to know key characteristics and motives of their target market. Machine learning allied to SOM in the GTA demonstrate a vast variety of demographics, socioeconomics, and expenditures, all of which play a vital role for developing insights on the different groups of wealthy consumers throughout the GTA. The themes and variables chosen for the geodemographic segmentation system allow for an insightful analysis on the wealthy consumer groups throughout the GTA. The geodemographic segmentation system created in the study allowed for the research questions to be answered, successfully.

The GTA is a diverse consumer market when analyzing demographics and socioeconomics in comparison to many other places in Canada. The GTA is diverse in household characteristics, such as average household income, visible minority groups, education, and more. The top three wealthiest clusters in the GTA is Extravagant Life-Style Consumers, Luxurious Spenders, and Flourishing Consumers. These three clusters all have an annual average household income of $150,000 and greater. The Extravagant Life-Style Consumers has the small number of census tracts, households, and population out of all clusters in the geodemographic segmentation system. The Extravagant Life-Style Consumers tend to spend large amounts on products, services, and living expenses, which is demonstrated in the results and discussion of the expenditure table for the cluster group. The Extravagant Life-Style Consumers is a highly educated consumer group, with an average household income of $652,680.81. The second wealthiest cluster is the Luxurious Spenders, which is also a highly educated consumer group, and is much larger than the Extravagant Life-Style Consumers, with a total of 20 census tracts, and a larger population and number of households. The Flourishing Consumers is the third wealthiest consumer group and is the largest of the top three consumer groups. The Flourishing Consumers have an average household income slightly greater than $200,000 and are a diverse group of consumers which have a high education attainment level throughout the cluster's population. The summary table of the wealthy consumer clusters created in the geodemographic segmentation system provide basic but important information on each cluster group, such as, average household income, average total expenditures, number of census tracts, and more. The cluster indices tables for each cluster group provided insightful findings and statistics on the demographics and socioeconomic variables for each cluster. The cluster indices tables for each cluster compares the clusters demographic and socioeconomic variables to that of the total GTA. The cluster indices conducted in the study demonstrate whether the clusters variables are under-represented or over-represented in comparison to the total GTA. The indices allow for the clusters to be easily compared to the entire GTA as well as the other clusters. The cluster indices are an important component of the study, as it conveys a great amount of information for better understanding the characteristics of each cluster. The map of each cluster shows the spatial component of the wealthy consumer behaviours as to where these households are located within the GTA. The maps make the findings clearer and are an important part of the study. The wealthscapes in the GTA continue to be a compelling group for understanding wealthy consumer behaviour. Wealth in the GTA demonstrates diversity of demographics and socioeconomics, which the geodemographic segmentation system was able to determine and the increasing role of centrality within the city core as driver of the wealth settlement within the city.

Combining spatial clustering and machine learning has allowed for this study to draw links between the geodemographic landscape of Toronto and the heterogeneity, econocentrism and regional intelligence of the city (Vaz, Zhao, & Cusimano, 2016). Regarding heterogeneity, this paper illustrates that throughout Toronto there is a wide diversity of geodemographic landscapes. However, there is greater wealth disparity when comparing individual census tracts, indicating that within census tracts the geodemographic landscape is becoming less diverse.

Regarding econocentrism and the link between economic disparities and environmental injustice, this study indicates that, due to the lack of income diversity within individual census tracts, environmental injustice is likely more prominent in low income CT's which tend to have higher numbers of immigrants and visible minorities. Typically, low income, and minority populations are disproportionately impacted by environmental injustice, and this injustice can impact the prosperity and health of these populations (Vaz et al., 2017).

Finally, regarding regional intelligence, this paper has, through the novel approach of combining clustering and machine learning, shed new light on the composition of the geodemographic landscape of Toronto. Through the identification of socio-economic trends that can be found in low-, mid- and high-income clusters, this paper has identified demographic groups that are disproportionately impacted by economic exclusion, income segregation and environmental injustice. This approach and the results of this paper indicate that the application of artificial intelligence and machine learning in urban decision making can aid in understanding the geographic composition of wealth in urban areas and lead to the identification disproportionately impacted demographic groups.

Authors statement

Eric Vaz: Writing – original draft, Writing – review & editing, Formal analysis, Data curation, Conceptualization, Fernando Bação: Formal analysis, Conceptualization, Bruno Damásio: Writing – review & editing, Formal analysis, Conceptualization, Data curation, Malik Haynes: Writing – original draft, Formal analysis, Data curation, Elissa Penfound: Formal analysis, Writing – original draft, Writing – review & editing.

Acknowledgements

The authors would like to thank the reviewers for their excellent comments, that greatly enhanced this article.

References

Abarca-Alvarez, F. J., Navarro-Ligero, M. L., Valenzuela-Montes, L. M., & Campos-Sánchez, F. S. (2019). European strategies for adaptation to climate change with the mayors adapt initiative by self-organizing maps. *Applied Sciences, 9*(18), 3859.
Abbas, J., Ojo, A., & Orange, S. (2009). Geodemographics–a tool for health intelligence? *Public Health, 123*(1), e35–e39.
Agarwal, P., & Skupin, A. (2008). *Self-organising maps: Applications in geographic information science*. John Wiley & Sons.
Akbar, D., Rolfe, J., Hossain, R., et al. (2017). Perceived risks to housing unaffordability in resource-led regional communities: Lessons for australian regional citie. *Australasian Journal of Regional Studies, 23*(1), 2.
Allouche, M. K., & Moulin, B. (2005). Amalgamation in cartographic generalization using Kohonen's feature nets. *International Journal of Geographical Information Science, 19*(8–9), 899–914.
Breau, S. (2014). The occupy movement and the top 1% in Canada. *Antipode, 46*(1), 13–33.
Breau, S., Shin, M., & Burkhart, N. (2018). Pulling apart: New perspectives on the spatial dimensions of neighbourhood income disparities in Canadian cities. *Journal of Geographical Systems, 20*(1), 1–25.
Buliung, R., Hernandez, T., & Mitchell, J. E. (2007). Exploring the relationship between Big-Box retail and consumer travel demand in the Greater Toronto Area. *Canadian Journal of Transportation, 1*(2).
Buliung, R. N., & Kanaroglou, P. S. (2002). Commute minimization in the greater Toronto area: Applying a modified excess commute. *Journal of Transport Geography, 10*(3), 177–186.
Chen, N., Chen, L., Ma, Y., & Chen, A. (2019). Regional disaster risk assessment of China based on self-organizing map: Clustering, visualization and ranking. *International Journal of Disaster Risk Reduction, 33*, 196–206.
Chen, W.-H., Myles, J., & Picot, G. (2012). Why have poorer neighbourhoods stagnated economically while the richer have flourished? Neighbourhood income inequality in Canadian cities. *Urban Studies, 49*(4), 877–896.
Condon, P. M. (2020). *Five rules for tomorrow's cities: Design in an age of urban migration, demographic change, and a disappearing middle class*. Island Press.
Corak, M. (2013). Income inequality, equality of opportunity, and intergenerational mobility. *The Journal of Economic Perspectives, 27*(3), 79–102.
Daniel, E., & Bright, G. (2011). Exploring the geographical distribution of wealth using the output area classification. *Economic & Labour Market Review, 5*(1), 59–78.
Deng, Y., Qin, Y., & Wu, J. (2019). *Superstar cities and the globalization pressures on affordability*.
Glaeser, E. L., & Kahn, M. E. (2004). Sprawl and urban growth. In *, Vol. 4. Handbook of regional and urban economics* (pp. 2481–2527). Elsevier.
Glaeser, E. L., Resseger, M., & Tobio, K. (2009). Inequality in cities. *Journal of Regional Science, 49*(4), 617–646.
Guest, A. M., & Lee, B. A. (1984). How urbanites define their neighborhoods. *Population and Environment, 7*(1), 32–56.
Gyimah, S. O., Walters, D., & Phythian, K. L. (2005). Ethnicity, immigration and housing wealth in Toronto. *Canadian Journal of Urban Research*, 338–363.
Harris, R. (2012). *The city resurgent*. London, England: SAGE Publications Sage UK.
Jones, K., & Doucet, M. (2000). Big-box retailing and the urban retail structure: The case of the Toronto area. *Journal of Retailing and Consumer Services, 7*(4), 233–247.
Kala, A. K., Atkinson, S. F., & Tiwari, C. (2020). Exploring the socio-economic and environmental components of infectious diseases using multivariate geovisualization: West Nile Virus. *PeerJ, 8*, Article e9577.
Kaski, S., Honkela, T., Lagus, K., & Kohonen, T. (1998). WEBSOM–self-organizing maps of document collections. *Neurocomputing, 21*(1–3), 101–117.
Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics, 43*(1), 59–69.
Kohonen, T. (2001). *Self-organizing maps* (" 3rd ed.). Berlin, Heidelberg, New York: Springer-Verlag.
Kohonen, T. (2013). Essentials of the self-organizing map. *Neural Networks, 37*, 52–65.
Koua, E. L., & Kraak, M.-J. (2004). Geovisualization to support the exploration of large health and demographic survey data. *International Journal of Health Geographics, 3*(1), 12.
Kourtit, K., Nijkamp, P., & Arribas, D. (2012). Smart cities in perspective – a comparative European study by means of self-organizing maps, Innovation. *The European Journal of Social Science Research, 25*(2), 229–246. https://doi.org/10.1080/13511610.2012.660330
Lee, A. C.-D., & Rinner, C. (2015). Visualizing urban social change with self-organizing maps: Toronto neighbourhoods, 1996–2006. *Habitat International, 45*, 92–98.
Lemon, J. (1991). *Toronto. Cities, 8*(4), 258–266. https://doi.org/10.1016/0264-2751(91)90041-O
Lightman, N., & Gingrich, L. G. (2012). The intersecting dynamics of social exclusion: Age, gender, race and immigrant status in canadas labour market. *Canadian Ethnic Studies Journal, 44*(3), 121–145.
Maroto, M., & Aylsworth, L. (2016). Catching up or falling behind? Continuing wealth disparities for immigrants to Canada by region of origin and cohort. *Canadian Review of Sociology/Revue Canadienne de Sociologie, 53*(4), 374–408.
Milligan, K., & Smart, M. (2015). Taxation and top incomes in Canada. *Canadian Journal of Economics/Revue Canadienne d'économique, 48*(2), 655–681.
Moos, M., & Mendez, P. (2015). Suburban ways of living and the geography of income: How homeownership, single-family dwellings and automobile use define the metropolitan social space. *Urban Studies, 52*(10), 1864–1882.
Openshaw, S., Blake, M., & Wymer, C. (1995). Using neurocomputing methods to classify. *Innovations in GIS*, 97.
Páez, A., Trépanier, M., & Morency, C. (2011). Geodemographic analysis and the identification of potential business partnerships enabled by transit smart cards. *Transportation Research Part A: Policy and Practice, 45*(7), 640–652.
Qi, J., Liu, H., Liu, X., & Zhang, Y. (2019). Spatiotemporal evolution analysis of time-series land use change using self-organizing map to examine the zoning and scale effects. *Computers, Environment and Urban Systems, 76*, 11–23.
Rae, A., & Singleton, A. (2015). *Putting big data in its place: A regional studies and regional science perspective*. Taylor & Francis.
Rosen, G., & Walks, A. (2015). Castles in Toronto's sky: Condo-ism as urban transformation. *Journal of Urban Affairs, 37*(3), 289–310.
Sachs, J. D., Mellinger, A. D., & Gallup, J. L. (2001). The geography of poverty and wealth. *Scientific American, 284*(3), 70–75.
Sester, M. (2005). Optimization approaches for generalization and data abstraction. *International Journal of Geographical Information Science, 19*(8–9), 871–897.
Sharanjit, U., & LaRochelle-Côté, S. (2015). *Changes in wealth across the income distribution, 1999 to 2012. Insights on Canadian Society. June. Statistics Canada Catalogue no. 75-006-X*.
Spielman, S. E., & Thill, J.-C. (2008). Social area analysis, data mining, and GIS. *Computers, Environment and Urban Systems, 32*(2), 110–122.
Vaz, E. (2020). *Regional intelligence*.
Vaz, E., Anthony, A., & McHenry, M. (2017). The geography of environmental injustice. *Habitat International, 59*. https://doi.org/10.1016/j.habitatint.2016.12.001
Vaz, E., & Jokar Arsanjani, J. (2015). Predicting urban growth of the greater Toronto area - coupling a markov cellular automata with document meta-analysis. *Journal of Environmental Informatics, 25*(2), 71–80. https://doi.org/10.3808/jei.201500299
Vaz, E., Shaker, R. R., & Cusimano, M. D. (2020). A geographical exploration of environmental and land use characteristics of suicide in the greater Toronto area. *Psychiatry Research*, Article 112790.
Vaz, E., Zhao, Y., & Cusimano, M. (2016). Urban habitats and the injury landscape. *Habitat International, 56*, 52–62.
Walks, A. (2014). Gated communities, neighbourhood selection and segregation: The residential preferences and demographics of gated community residents in Canada. *Town Planning Review, 85*(1), 39–66.
Walks, A. (2016). Homeownership, asset-based welfare and the neighbourhood segregation of wealth. *Housing Studies, 31*(7), 755–784.
Walks, A., Dinca-Panaitescu, M., & Simone, D. (2016). *Income inequality and polarization in the city of Toronto and York region Part I: Examining levels and trends from spatial and non-spatial perspectives neighbourhood change. Research partnership factor-inwentash faculty of social work*. University of Toronto.
Wan, W., & Fraser, D. (1993). M2dSOMAP: Clustering and classification of remotely sensed imagery by combining multiple kohonen self-organizing maps and associative memory. In *, Vol. 3. Proceedings of 1993 international conference on neural networks (IJCNN-93-Nagoya, Japan)* (pp. 2464–2467).
Wei, Y. D. (2017). Geography of inequality in asia. *Geographical Review, 107*(2), 263–275.

Wickramasinghe, C. S., Amarasinghe, K., & Manic, M. (2019). Deep self-organizing maps for unsupervised image classification. *IEEE Transactions on Industrial Informatics, 15*(11), 5837–5845.

Xiang, L., Stillwell, J., Burns, L., Heppenstall, A., & Norman, P. (2018). A geodemographic classification of sub-districts to identify education inequality in Central Beijing. *Computers, Environment and Urban Systems, 70*, 59–70.

Yan, J., & Thill, J.-C. (2009). Visual data mining in spatial interaction analysis with self-organizing maps. *Environment and Planning B: Planning and Design, 36*(3), 466–486.