Cities 126 (2022) 103708

Contents lists available at ScienceDirect

Cities

journal homepage: www.elsevier.com/locate/cities

Measuring the process of urban gentrification: A composite measure of the gentrification process in Toronto

Jason Hawkins [a,b,c,*], Usman Ahmed [b], Matthew Roorda [b], Khandker Nurul Habib [b]

[a] Department of Civil & Environmental Engineering, University of Nebraska - Lincoln
[b] Department of Civil & Mineral Engineering, University of Toronto, Canada
[c] Department of Civil, Environmental & Architectural Engineering, University of Texas at Austin, United States of America

ARTICLE INFO

Keywords:
Gentrification
Spatial statistics
Multi-criteria analysis
GIS, clustering

ABSTRACT

Gentrification is being experienced by cities around the world. Its drivers and characteristic features are complex and diverse, ranging from the displacement of low-income households to the redevelopment of commercial districts. This paper combines multiple data sources to explore the coevolution of gentrification in the residential market and developments in non-residential sectors in the Canadian city of Toronto. Analysis starts from a max-p-regions clustering based on a composite measure of the gentrification process, which includes measures of household income, educational attainment, building permits, and the composition of establishments. The variables that describe each region are examined for patterns of gentrification. We then develop a Bayesian hierarchical spatial (BHS) model to describe the change in residential property prices over the five-year period between 2011 and 2016. The results show that establishment entropy and composition are found to influence residential property prices, with significant spatial variation in its effect. The combination of endogenous region definitions and hierarchical modeling is strongly supported by model results.

1. Introduction

Gentrification is a neighborhood process that has been applied in different contexts. Therefore, it is defined differently, and the precise definition is difficult to pinpoint. Gentrification is often described as the change in the socioeconomics of a neighborhood, such as an increase in average neighborhood income (Lee, 2010) or change of socioeconomic indicators (such as income and education) of the neighborhood faster relative to the city (Grube-Cavers & Patterson, 2015). There are several geographic implications associated with the gentrification process, including the relocation of low-income households into areas of poor access to transit and employment (Skaburskis & Nelson, 2014) and the segregation of retail from residential land uses due to higher land prices incentivizing the scale economies of big box stores (Chapple & Jacobus, 2009). Gentrification is also defined as the replacement of low-value businesses by high-value businesses (Ferm, 2016). However, these dynamics have not received the same level of attention in the literature as residential change (Curran, 2007; Ferm, 2016; Lens & Meltzer, 2016; Meltzer, 2016; Parker, 2018).

It is important to quantify gentrification in urban neighborhoods. Quantification helps to identify the factors driving the gentrification process and assess the scale of gentrification in a neighborhood. Policies that aim to curb gentrification can benefit from such quantification. The definition of neighborhoods is usually based on census boundaries which are not suitable for gentrification analysis due to large variation in neighborhood size ((Brown-Saracino, 2017). The analysis unit, referred to as regions, for gentrification studies requires endogenous definition based on composite measures of gentrification to eliminate aggregation bias. These regions could be built up from the smallest available spatial units available for a city.

The method is applied to Toronto, one of the strongest real estate markets in North America. Canadian cities are among those that have undergone the gentrification process within their neighborhoods. For instance, (Grube-Cavers & Patterson, 2015) studied the relationship of gentrification to transit in the three largest Canadian cities (Toronto, Montreal, and Vancouver) and found that Toronto and Montreal have undergone gentrification with a positive relationship to urban rail transit stations. Despite efforts to address housing affordability, the cost of living was among the most important issue to Canadians according to a poll in the run-up to the fall 2019 federal election (Hall, 2019). Toronto home prices rose by 10% (month-over-month) in June 2019, and many of these sales were condominiums in gentrifying neighborhoods

* Corresponding author at: Department of Civil & Environmental Engineering, University of Nebraska - Lincoln.
E-mail address: jason.hawkins@unl.edu (J. Hawkins).

https://doi.org/10.1016/j.cities.2022.103708
Received 6 June 2021; Received in revised form 6 April 2022; Accepted 16 April 2022
Available online 30 April 2022

expanding out from the city core (The Canadian Press, 2019). Similarly, rising commercial property taxes have led to the replacement of independent stores by larger firms with the capital to bear the increasing cost of tenancy. Being the largest city in Canada, Toronto also represents a good example of the economic, social, and spatial patterns of US cities – particularly older northeast cities with strong financial and academic institutions such as Boston and New York.

Our region definition incorporates a variety of factors identified in the literature as being associated with gentrification. We integrate these gentrification measures within a hedonic price model (Rosen, 1974). Our dependent variable is the change in dwelling price per unit area between 2011 and 2016 (the previous two census years in Canada). Housing prices are determined by a combination of intrinsic factors (i.e., building features) and extrinsic factors (i.e., community features) (D'Acci, 2019). The proposed clustering approach allows for explicit community classification according to gentrification metrics found in the literature. The effects of sociodemographic, retail composition, and other gentrification indicators are then be systematically considered via a hierarchical regression model.

This paper examines gentrification as a process of spatial change involving both residential and commercial dynamics, as well as their interactions. Section 2 describes the gentrification literature, with a focus on Toronto. Section 3 sets up the joint clustering and modeling framework. The clustering and variable inputs are provided in section 4 for the Toronto case study. Section 5 presents the Bayesian hierarchical spatial (BHS) model results and illustrates the integration with clustering via parameter shrinkage plots. Finally, section 6 presents the discussion and conclusions.

2. Literature review

2.1. The gentrification process

There have been several recent efforts to categorize methods of gentrification analysis. Many studies measure gentrification as variation in either household income or residential property prices using longitudinal data (Ding et al., 2016). Preis et al. (2020) provide a summary of map-based methods implemented by planning agencies in four United States cities - Seattle, Los Angeles, Portland, and Philadelphia - and apply each method to Boston. They find different levels of gentrification for communities in Boston between the four methods. Only seven common census tracts are identified to be at a high risk of gentrification and displacement according to all four methods. The outcome of the gentrification analysis depends on the measure of gentrification used to quantify it and the specific definition and thresholds of the variables used in the model. Reades et al. (2019) highlight the shortcomings of purely rule-based maps of gentrification. As argued by Easton et al. (2020), there are often complex interactions between demographic and built form factors, requiring the application of probabilistic and model-based methods. However, even these multi-dimensional studies typically focus on residential factors - income, residential property prices, educational attainment, etc. (Easton et al., 2020; Freeman, 2005; Preis et al., 2020).

The findings of Preis et al. (2020) support the need for statistical models that captures the multi-faceted nature of gentrification. The measures developed in each of their four cities are quantifiable, in so far as most provide a numerical indicator, but lack a robust basis. In Seattle, a composite Displacement Risk Index is based on a weighted total of 14 variables. It is unclear how the variables should be chosen, how the weights should be determined, or at what level to set the cut-offs (for variables based on a threshold indicator). The same limitations exist for the measures proposed for Los Angeles, Portland, and Philadelphia. In addition, most of the indicators focus on household demographics and rely on decadal census data.

The mix of commercial establishments in an area can significantly influence its appeal and the uplift of residential gentrification (Lees et al., 2008). Lin and Yang (2019) conduct a study in Taipei on commercial gentrification near metro stations. The property units are classified into three categories and assessed for gentrification before and after the construction of metro stations. Commercial gentrification is identified as a binary variable and given a value of one if the unit changes its status from the lower category to the top-most category. A binary logit model is developed to identify the factors that affect commercial gentrification and proximity to metro stations is one of the factors. The results show that commercial gentrification tends to decrease with distance from metro stations. Other important variables include property rent, road width, income, and education level. Taxation and economic databases are used which are not publicly available for North American cities.

Changes in sociodemographics lead to the redevelopment of an area, inducing an upward valuation of dwellings (Chapple et al., 2017; Zapatka & Beck, 2020). The existing population is unable to pay the higher price of occupancy and must relocate. By using price, we can examine the changes in amenities enjoyed by in-migrating populations (Walks & Maaranen, 2008). We extend this hypothesis to consider that the in-migration of middle-class residents is likely preceded by, or coincident with, a change in the commercial composition of the region. That is, there are amenities in the region that are attractive to higher-income households. Papachristos et al. (2011) characterize this change in demographics as being associated with an introduction of "lifestyle amenities that appeal to the tastes - and meets the demands of - wealthier residents" (p. 217).

2.2. Gentrification in Toronto

Gentrification is not a recent phenomenon in Toronto. Walks and Maaranen (2008) found 36% of inner-city prewar neighborhoods in the city had undergone some form of gentrification as of 2008. Early examples are Yorkville, Don Vale/Cabbagetown, and Riverdale (Dantas, 1988; Sabourin, 1994). Skaburskis (2012) compares the socioeconomic characteristics of men and women in Toronto following the gentrification of the 1970s. He finds it is a common characteristic of gentrification that low-income multi-earner households are replaced by single-person professional households. By considering household income, the researcher may misinterpret the replacement of larger, older, households - with higher living expenses and often lower future earning potential - by younger persons with lower living expenses and higher earning potential. Skaburskis (2012) uses a division of household income by the square root of household size to correct for its correlation with household size. Other authors use various normalizations of income, typically referencing it to a larger geographic mean or median income statistic (Chapple & Jacobus, 2009; González et al., 2019; Walks & Maaranen, 2008). We outline our approach in a subsequent section.

Gentrification in Toronto has been associated with a financialization of the residential real estate market (August & Walks, 2018) and a transition from a manufacturing- to a service-based economy (Skaburskis & Nelson, 2014). There has been a shift, away from the replacement of rental by owner-occupied units, and towards the replacement of low-income rental by high-income rental units. Real estate investment and asset management corporations seek to maximize return on their investment in the expensive Toronto market. August and Walks (2018) identify two strategies by these corporations, which they term *squeezing* and *gentrification-by-upgrading*. The first consists of making improvements to buildings to raise the cost of living beyond the means of existing tenants. This strategy is typically employed in suburban communities. Gentrification-by-upgrading consists of targeting neglected properties in the inner city and making upgrades to the building with the intention to sell.

Immigration and globalization have also contributed to the gentrification of Toronto (Joy & Vogel, 2015). Immigrants tend to settle in a small number of major Canadian cities (Ley & Germain, 2000). According to Ley et al. (2002), net immigration had a correlation of 0.81

with average housing prices in Toronto from 1971 to 1996. They suggest that some of the patterns observed in the Toronto real estate are associated with socioeconomic polarization rather than gentrification. We capture this effect in our model by referencing increases in average income in Toronto against the national mean income. This referencing to the Canadian mean helps to address potential income polarization between Toronto and other areas of the country. (Specifications with income referenced to metropolitan and provincial averages were also tested). We further build on previous research in Toronto by taking advantage of more detailed real estate statistics: individual sales compared with Toronto Real Estate Board (TREB) averages used in the work of Ley et al. (Ley et al., 2002).

We make several contributions to the gentrification literature in Toronto (and in general). We first observe that most studies rely on preexisting census boundaries to define *neighborhoods* (regions) - the standard spatial unit of analysis in the gentrification literature (Chapple & Jacobus, 2009; Owens, 2012; Skaburskis & Nelson, 2014). This deficiency was previously identified by Brown-Saracino (2017), who finds that most studies use census tracts that "do not align with recognized neighborhoods or do not coincide with the areas studied by qualitative scholars because they are either too broad or too narrow" (p. 528). We systematize the process of defining spatial units through the adoption of a spatial clustering algorithm. Rather than using statistics provided for spatial units defined by statistical agencies for consistency between census years, we make the definition endogenous to the modeling process as suggested by Sergio Rey (Rey, 2019). Each region represents a set of underlying spatial units (in our case dissemination areas (DA)) at a detailed spatial resolution - 3702 DAs for the City of Toronto means each spatial unit is approximately 17 ha, or 6 city blocks using the average block size for Toronto (Siksna, 1997). These regions are defined by a composite gentrification measure, overcoming the limitations of previous work that used either existing neighborhood boundaries or ad-hoc clustering (Papachristos et al., 2011; Sampson et al., 1997). Our clustering of spatial units includes several features of establishments, which we find to precede the gentrification of the residential market in some cases. We then fuse annually collected administrative data and continuously collected real estate data to estimate a series of regression models that provide a robust quantification of the gentrification process. The regressions are specified as Bayesian hierarchical models, which have several statistical advantages, as we outline below.

3. Model setup

The model is defined in two steps. We begin by defining a spatial typology through the application of the max-p-regions clustering algorithm based on both residential and non-residential data sources. These *clusters* (hereafter referred to as *regions*) provide a composite measure of gentrification, identified by the underlying characteristics of each region. The use of an endogenous definition of regions mitigates the bias introduced by representing point data (i.e., individual dwelling prices) through aggregation into zones, known as the modifiable areal unit problem (MAUP). The data sources include firmographic data maintained by InfoCanada (InfoCanada, 2017), residential property sales, and Statistics Canada establishment counts by dissemination area (DA). Second, a Bayesian hierarchical spatial (BHS) model is developed based on the results of the clustering to draw out the effect of local variation in both demographic and firmographic attributes through the specification of hierarchical parameters.

3.1. *Max-P-regions for transition between 2011 and 2016*

The first step of our analysis is the clustering of spatial units using the max-p-regions algorithm proposed by Duque et al. (Duque et al., 2012). This proceeds from the definition by Fischer (Fischer, 1980) that a region is defined by spatially contiguous units that are homogeneous in their characteristics (e.g. average income, mix of industries, mix of dwellings). The objective function of the max-p-regions algorithm seeks to maximize the dissimilarity between clusters of spatial units, where dissimilarity is based on a vector of variables representing regional characteristics. The algorithm endogenously defines regions using a constraint on the minimum bound for one of the input variables (in our case population per region). It is further differentiated from other algorithms by including a spatial contiguity constraint. This means that regions on opposite sides of the study area could have similar characteristics and differ only in spatial location. The advantage of the max-p-regions approach is that it allows us to control the definition of regions such that they are based on a known set of gentrification measures.

3.2. *Bayesian hierarchical spatial (BHS) model*

The model is a hierarchical linear model, wherein the hierarchy is defined by the nested geography: dissemination areas (DAs) defined by Statistics Canada and regions defined by max-p-regions clustering (model structure illustrated in Fig. 1). We use a residential sales dataset in the model, which includes additional attributes of the property: number of bedrooms, number of washrooms, parking type, structure type, floorspace area, days listed on the market, and year of sale. Floorspace area is used to normalize sales prices and average sales prices are calculated in each of 2011 and 2016 by DA. The hierarchical structure avoids the issue of overfitting associated with a standard linear model. Rather than specifying individual (unpooled) parameters for each region, a hierarchical structure allows us to specify population parameters that also account for dependencies within the population. The model determines the relative weights of global (pooled) and local (unpooled) effects. Uncertainties due to sample sizes in each region are considered by *shrinking* local estimators to the global mean.

4. Application to Toronto, Canada

4.1. *Composite gentrification measure development*

We develop a composite measurement of gentrification based on the above literature and data available for the study area of Toronto, Canada. The standard application of max-p-regions is to a cross-section of spatial units, whereby each unit is classified based on an indicator vector for the data collection year. However, we are interested in the transition (gentrification) over the five years between 2011 and 2016. To capture this effect with max-p-regions, we set the indicator vector equal to the minimum of the populations in each of the analysis years. This specification allows us to maintain a consistent set of regions and compare indicator values between years.

The clustering variables are the change in highest educational attainment distribution (less than high school, high school, diploma, bachelor's degree, or advanced degree) between 2011 and 2016; the change in median household income between the analysis years; and a building permit count. Rather than using a ranking system for income (e.g., movement of a DA from the 10th highest income to the 5th highest income), we define a numerical measure as deviations from an overall median income. Rather than measuring deviations against the value for the City of Toronto, we benchmark the measure against a larger geographic region. This approach helps to address the fact that Toronto exists in a larger spatial market. The entire City of Toronto may be subject to an uplift in median income that would not be captured by a comparison of relative positions of DAs between years. That is, households may choose to relocate from the surrounding suburbs, another part of Ontario, or even Canada. We define three measures of income change with reference to the median for the Greater Toronto Area, Ontario, and Canada. The results are then compared to determine the sensitivity of the regionalization to the assumption of the reference income (see Fig. 2). The permit count variable is filtered to include only those permits issued for retail, accommodations, and food establishments. This measure works in the spirit of the coffee shop count of Papachristos et al. (2011) or the renovation count of Laska et al. (1982) in

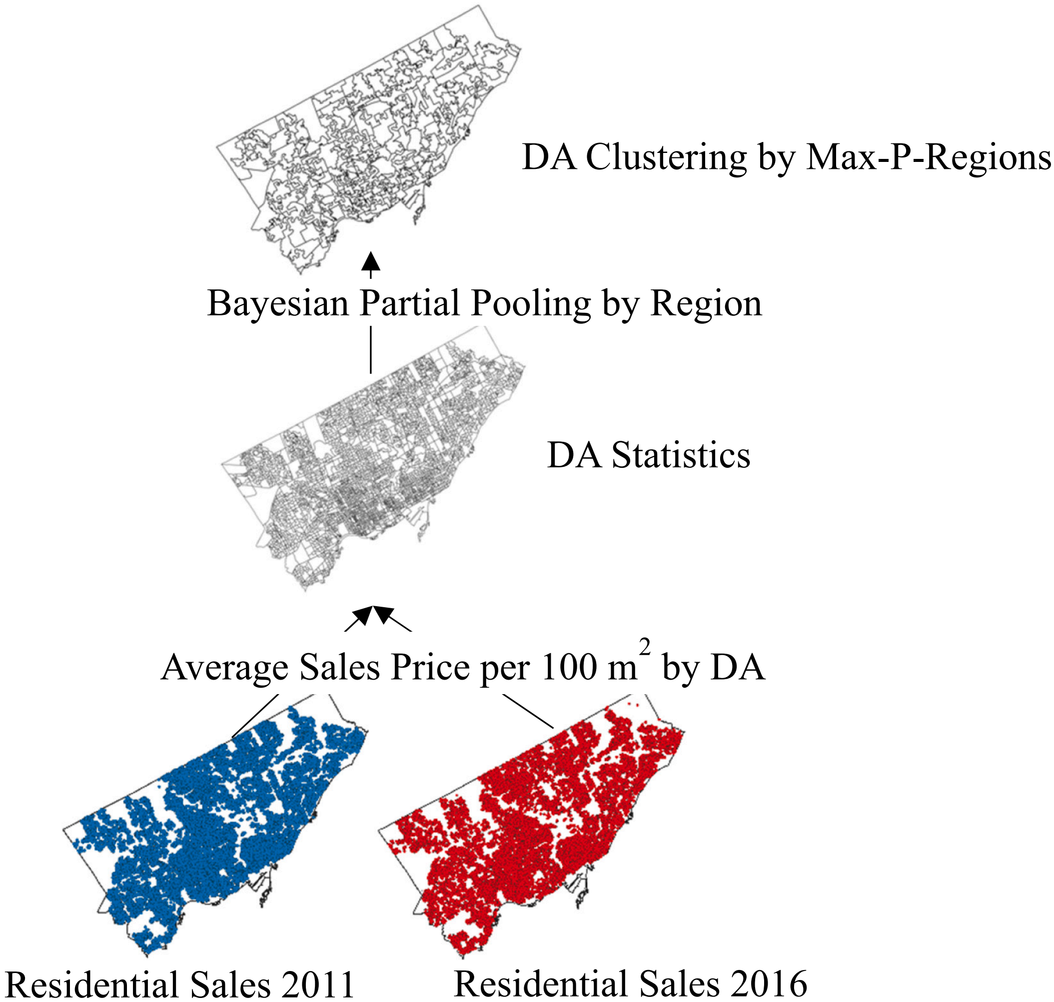


Fig. 1. BHS model structure.

that it provides a measure of transition in the commercial space of a region.

The use of a 5% threshold to construct Fig. 2 gives a small enough population of regions to easily compare the distribution of variables input to the clustering algorithm.

Breakout plots for each of the variables included in the composite measure are provided in supplemental Figs. S1-S5. To check that the composite measure is capturing multiple dimensions of gentrification, we run the max-p-regions algorithm with only a subset of variables and compare the correlation in the DA clustered into regions. The results of this analysis are shown in Fig. S5, which suggests the composite measure is capturing a range of factors contributing to variation in DA composition.

For use in the statistical model, we set a lower threshold of 0.5% of the total population for each region. This threshold gives 162 regions, which closely matches the 140 neighborhoods defined by the City of Toronto. A comparison between our regions and city-defined neighborhoods is given in Fig. 3.

4.2. Model variables development

A range of data sources can be found in the gentrification literature, from in-depth interviews (Sutton, 2010; Zukin et al., 2009) to census data and business registers (Curran, 2007; Owens, 2012; Yoon & Currid-Halkett, 2015). We obtain individual property sale prices for the study area from Toronto Real Estate Board (TREB) listings in each of the two previous census years: 2011 and 2016. Statistics Canada provides the number of establishments for each DA, which are categorized by their employment range (8 categories) and North American Industry Classification System (NAICS) code. From these data, we can define a measure of entropy given by

$$e_i = -\sum_{ij} p_{ij} ln\left(p_{ij}\right) \tag{1}$$

where e_i is a measure of the mix of sectors in DA i and p_{ij} is the proportion of establishments in DA i within sector j. We focus our entropy measure on a subset of sectors and use the more detailed six-digit industry NAICS code. Establishments within 126 industries are included in the entropy measure and are taken from retail trade, arts, accommodations, and other services - shown to be key sectors identifying gentrification (Yoon & Currid-Halkett, 2015). The change in entropy is illustrated in Fig. 4a below. Araldi and Fusco (2019) demonstrate how such mappings of establishment mix (entropy) can be used to identify commercial corridors and characterize the functional makeup of commercial districts.

Throughout the discussion of results, abbreviated demographic variable names are used as defined in Table 1 and summary statistics are given in Table 2.

In addition to demographic variables, firmographic data from InfoCanada (InfoCanada, 2017) and DMTI (DMTI, 2016) are used to develop variables that characterize changes in the non-residential sector. Several studies have shown the efficacy of these datasets in research applications (Bader et al., 2010; Carroll & Torfason, 2011; Joe Schlichtman & Patch, 2008; Kubrin et al., 2011; Small & McDermott, 2006; Yoon & Currid-

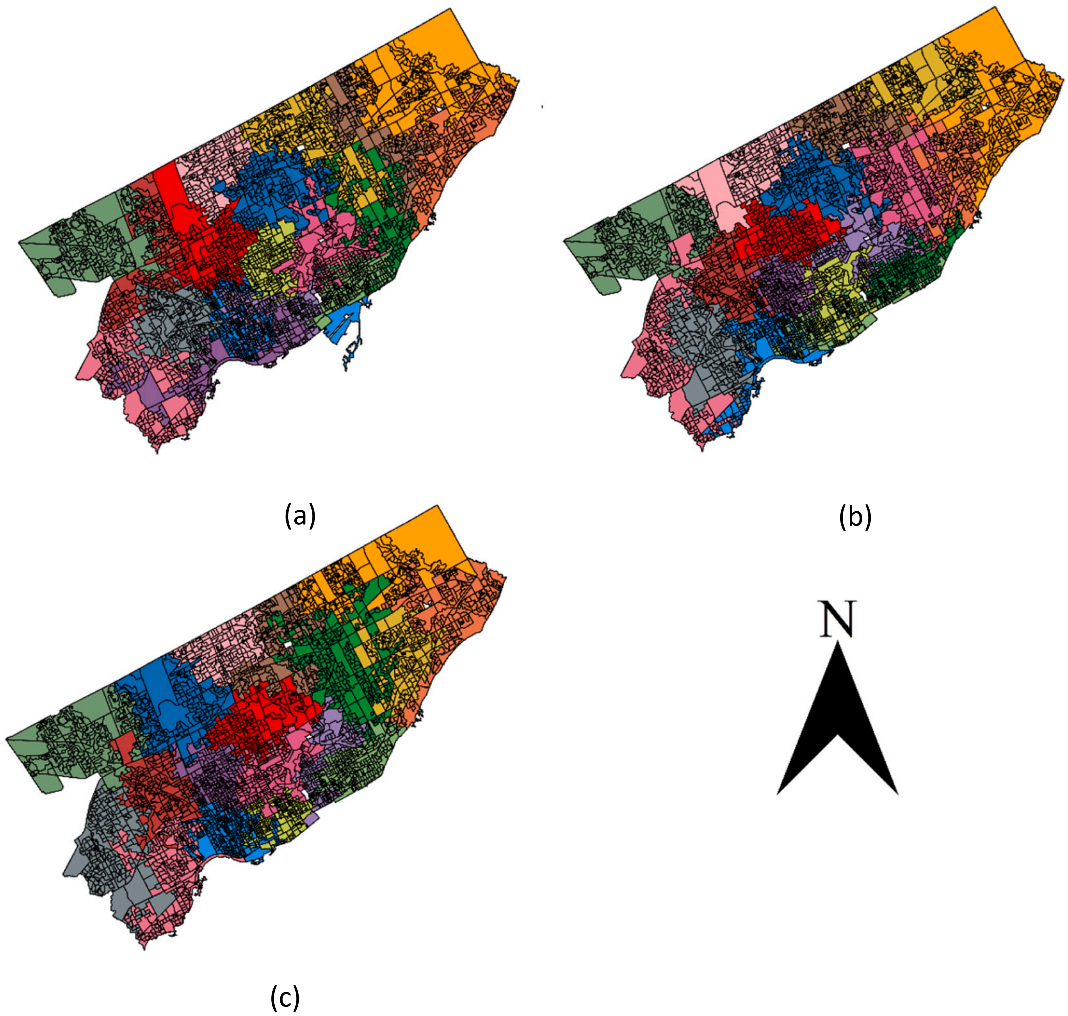


Fig. 2. Comparison of 5% population max-p-regions clustering using (a) Canada median income, (b) Greater Toronto Area median income, and (c) Ontario median income.

Halkett, 2015). The single-location establishment and entropy variable are motivated by the findings that gentrifying regions tend to attract multi-establishment businesses (i.e., chains) and lower-income neighborhoods tend to have fewer, or less diverse, retail establishments (Block et al., 2004; Lewis et al., 2005; Meltzer, 2016; Meltzer & Schuetz, 2012; Zukin et al., 2009). Gonzalez et al. (González et al., 2019) follow a similar logic to develop a composite commercial gentrification index that includes establishment churn, minority-owned establishment share, and non-chain establishment share. The establishment retirement and location type variables are mapped in Fig. 4b and c.

5. Bayesian hierarchical spatial (BHS) Model

5.1. Model specification

Eq. 2 shows the structure of the model. All variables are specified as first differences between 2016 and 2011. Region-specific fixed effects (a_{1r}), proportion of single-location establishments, business entropy, and business retirements parameters are specified conditional upon the regions defined by the max-p-regions clustering. Spatial autocorrelation is addressed through the inclusion of a conditional autoregressive (CAR) correlation structure, which captures the potential for autocorrelation between observations allocated to adjacent regions (as measured by a weight matrix, W). Lym (2021) uses a Bayesian spatial model with CAR priors to explore regional shrinkage in Ohio and is similarly motivated by the need for multi-dimensional approaches in regional studies. The dependent variable in the BHS model is the change in price per square meter for individual dwellings sold in 2011 and 2016 in the City of Toronto. Entropy and P(BR) are assumed to spatially vary across regions.

The model is estimated using the brms interface to Stan (Bürkner, 2017). We assume N(0,2.5) priors for the hyperparameters on the non-

$$\begin{aligned}\text{price} = \; & a_o + Population\ density + Pr(African) + Pr(Asian) + Pr(LatinAmerican) + \\ & Pr(Canadian/European) + Ethic\ mix + Pr(single - location) + Business\ entropy + \\ & Business\ retirements + \left(a_{1r} + Pr(single - location)_r + Business\ entropy_r + \right. \\ & \left. Business\ retirements_r | r\right) + \mathrm{CAR}(\mathrm{W}, \mathrm{group} = \mathrm{region})\end{aligned} \tag{2}$$

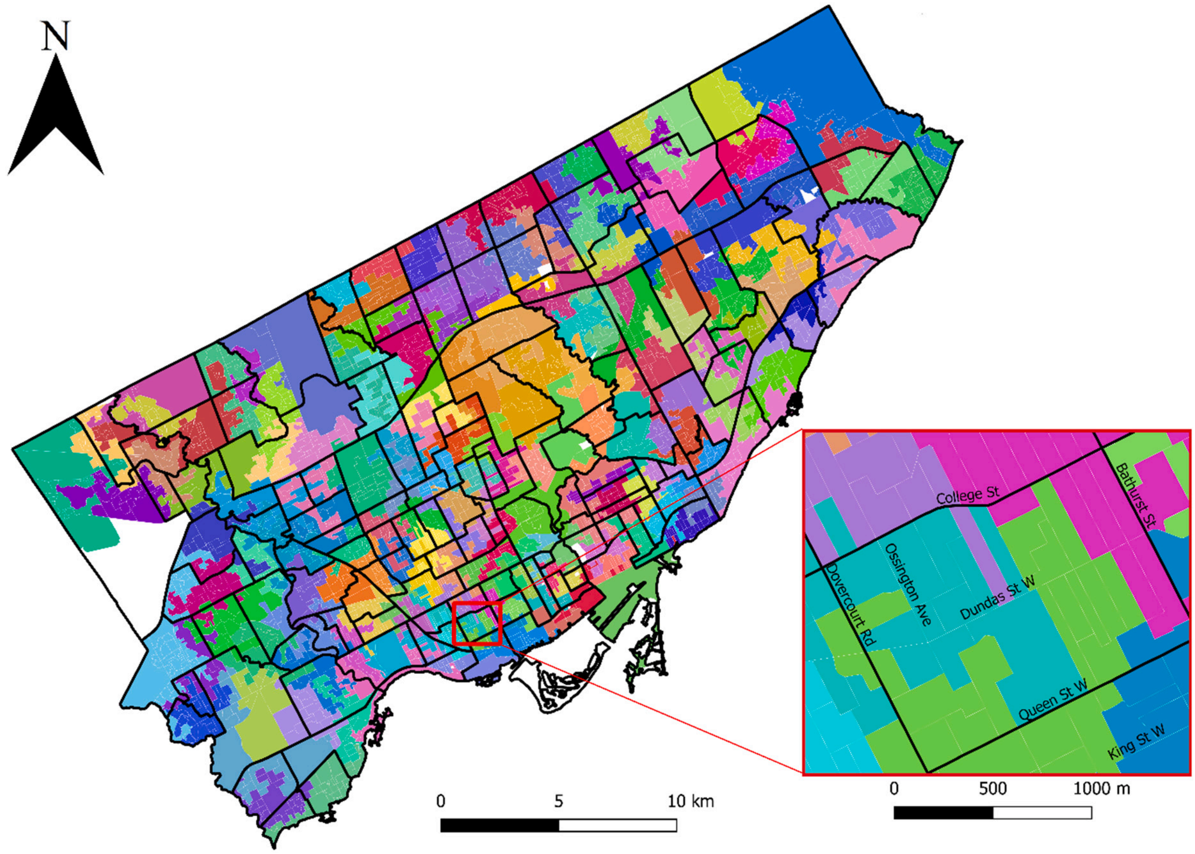


Fig. 3. Comparison of City of Toronto defined neighborhoods with max-p-regions results (black lines denote City of Toronto neighborhood boundaries).

hierarchical parameters, a N(0.5, 0.5) prior for the CAR parameter, and the default priors (generally t-distributed) provided by the software. Correlation is included between the hierarchical parameters.

5.2. Model results

Table 3 presents the results of the model estimation. We present the highest posterior density (HPD) region, which we assume corresponds to the central posterior interval and is analogous to inference on a t-statistic (Gelman et al., 2020).

Hierarchical parameters highlight the statistical strength of the regions resulting from the max-p-regions procedure. Significant variation in parameter values is observed for all region-level variables that would not be captured by a single population-level parameter. For example, the population-level entropy parameter is small and insignificant, but region-level variation suggests significant variation and differences in signs between some regions. A similar observation can be made for the single-location and business retirement variables.

5.3. Shrinkage as a cluster strength metric

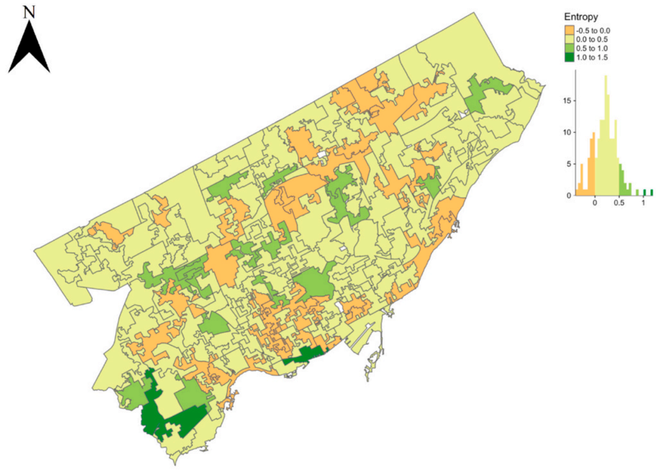
Shrinkage of random intercept (or regional fixed effects) between the unpooled and partially pooled model indicates the strength of the clustering on dwelling price change. A large amount of shrinkage would indicate that the unpooled parameter estimate variation was more likely due to sample sizes within each region rather than a true variation due to clustering variables (i.e., gentrification). Fig. 5 displays no shrinkage with partial pooling, suggesting that the variation in fixed effects is due to structural variation between regions. This result supports the need for both residential and non-residential measures in the regionalization. The findings for single location establishments support the findings of Gonzalez et al. (González et al., 2019) that variation exists in the class of establishments (i.e., chain vs. non-chain stores) in gentrifying neighborhoods.

The CAR parameter helps to control for other spatial autocorrelation. Its significant moderately large value suggests that there is spatial correlation in the residuals, which would bias parameter estimates if this parameter was excluded from the model.

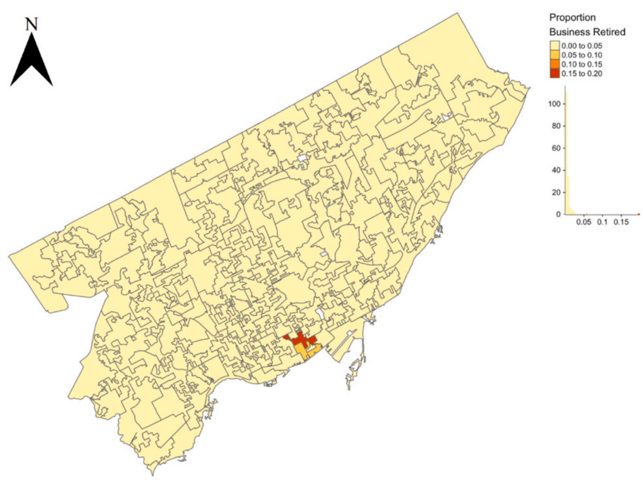
6. Discussion and conclusions

In this study, we present a case study of the spatial evolution of gentrification in the Canadian city of Toronto. We begin from a clustering to characterize the spatial units of the study area along multiple dimensions that influence gentrification: income, composition of ethnic origins, and building permits. We then apply the clustering results to a statistical model to draw out quantitative insights. While previous studies have examined real estate and social drivers of gentrification in Toronto, they have lacked a discussion of how these dynamics interact and define the spatial structure of the city. Our results suggest a strong influence of firmography on residential property price, a standard gentrification measure, that has not been extensively examined in the existing literature. The endogenous determination of spatial units strengthens the ability of the analyst to examine such patterns by mitigating the influence of MAUP and addressing spatial autocorrelation. We find the BHS model has appealing features, specifically its ability to represent flexible data generation processes while maintaining aspects of inferential statistics. The model structure allows us to play with demographic and firmographic effects to consider their spatial variation.

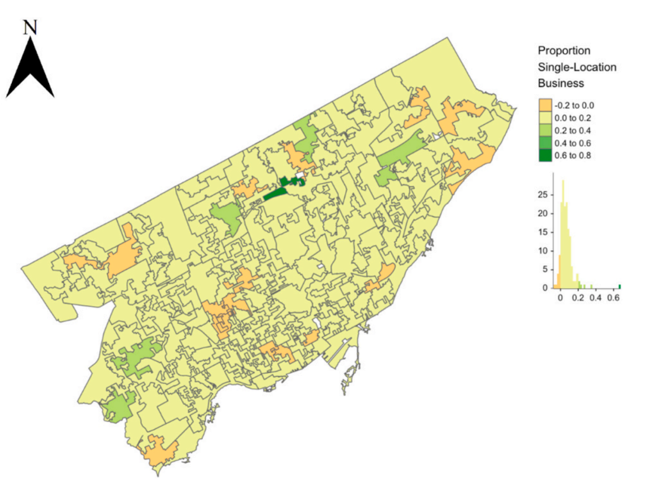
The approach presented herein is readily generalizable to other North American cities. For example, Preis et al. (2020) highlight the lack of consistency between gentrification metrics independently developed for Seattle, Los Angeles, Portland, and Philadelphia. In addition, these metrics rely on existing spatial units that are not based on gentrification-



(a)



(b)



(c)

Fig. 4. Distribution of region-specific variable values for a) establishment entropy, b) business retirements, and c) proportion of single location establishments (measured as changes from 2011 to 2016)

relevant measures (e.g., firmographic and sociodemographic differences). Our approach could be uniformly applied across these cities (and any number of other Canadian and American cities) to define spatial units based on uniform and gentrification-relevant measures and avoid aggregation bias. The shrinkage plot in Fig. 5 illustrates that such an endogenous spatial unit definition draws out significant variation between regions. Gentrification modeling and/or metrics (as used in the Preis et al. (2020) case studies) would benefit from such uniformity for inter-city comparison. The endogenous max-p-regions clustering algorithm could also form the basis for national gentrification analysis, wherein a consistent set of metrics could be used to assess patterns

Table 1
Summary of demographic variables and their descriptions.

Variable	Description
Price	Price per m^2 (×100)
Population density	Population per hectare (×1000)
Income	Average household income in DA (×\$100,000)
Pr(African)	Proportion of African ethnicity
Pr(Asian)	Proportion of Asian ethnicity
Pr(Latin American)	Proportion of Latin American ethnicity
Pr(Canadian/European)	Proportion of Canadian or European ethnicity
Ethnic mix	Ethnic mix in DA (entropy measure)

Table 2
Summary of demographic variable statistics.

	Min.	1st Quart.	Median	Mean	3rd Quart.	Max.
Price	−2.41	1.16	2.01	3.58	4.29	16.31
Population	−6.93	0.00	0.00	0.00	0.00	0.52
Income	−7.06	−0.38	−0.18	−0.32	−0.05	0.70
Pr(African)	-0.36	0.00	0.01	0.01	0.03	0.51
Pr(Asian)	−0.50	−0.05	0.01	0.01	0.07	0.61
Pr(Latin American)	−0.28	0.00	0.01	0.01	0.03	0.24
Pr(Canadian/ Europe)	−0.67	−0.10	−0.04	−0.04	0.02	0.45
Ethnic mix	−1.27	0.00	0.13	0.14	0.28	1.30

Table 3
Summary of BHS model results for change in average dwelling price per 100 m^2.

	Estimate	Est. Error	Lower HPD	Upper HPD
Population-level effects ($N = 3683$)				
Intercept	3.66	0.15	3.42	3.90
Population	−0.65	0.37	−1.26	−0.05
DA income	0.48	0.12	0.29	0.67
Pr(African)	−0.82	1.28	−2.91	1.28
Pr(Asian)	0.65	0.90	−0.83	2.12
Pr(Latin American)	−1.30	1.34	−3.51	0.92
Pr(Canadian/European)	0.57	0.95	−1.00	2.13
Ethnic mix	0.91	0.34	0.35	1.47
Pr(single-location)	−1.32	1.16	−3.21	0.59
Business entropy	0.03	0.38	−0.60	0.66
Business retirements	−0.65	2.32	−4.47	3.17
Region-level effects ($N = 162$)[2]				
Regional fixed effects	0.43	0.24	0.05	0.81
Pr(single-location)	1.39	1.13	0.10	3.60
Business entropy	1.17	0.56	0.23	2.10
Business retirements	2.41	2.44	0.16	6.76
cor(Intercept, Pr(single-location))	−0.10	0.44	−0.78	0.67
cor(Intercept, Business entropy)	0.11	0.40	−0.59	0.75
cor(Pr(single-location), business entropy))	0.01	0.44	−0.71	0.72
cor(Intercept, business retirements)	−0.01	0.45	−0.74	0.72
cor(Pr(single-location), business retirements)	0.00	0.45	−0.73	0.73
cor(business entropy, business retirements)	−0.02	0.44	−0.74	0.71
Correlation structure				
CAR	0.48	0.25	0.07	0.90
Std. Dev. CAR	1.14	0.48	0.26	1.85

[2] Standard deviations are shown for region-level effects.

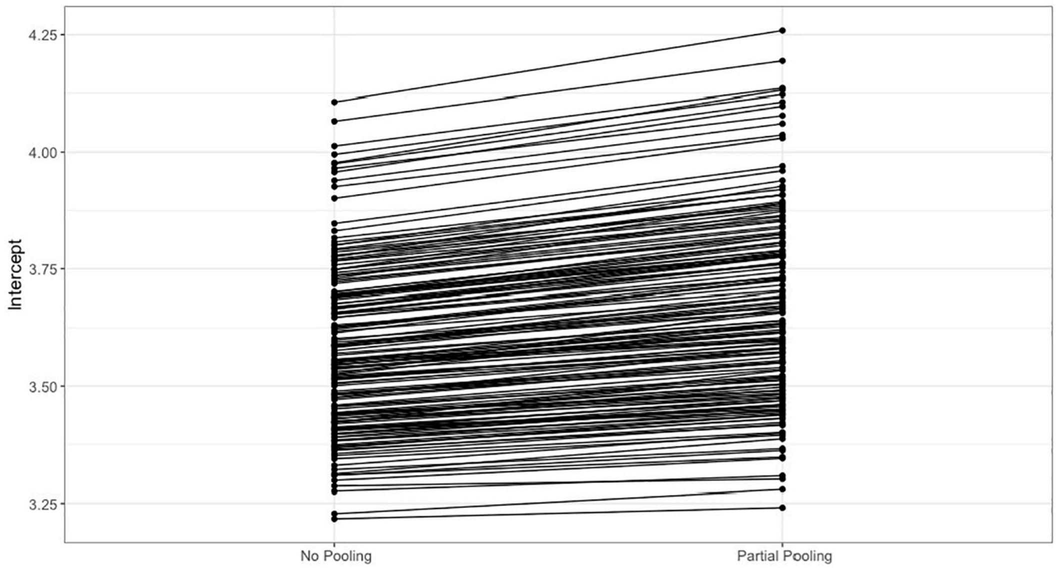


Fig. 5. Shrinkage of regional fixed effects with partial pooling

across regions, countries, or even across national borders (assuming consistency among the data inputs). For example, national housing policy could be assessed by identifying commonalities among *regions* in different states that are identified by the clustering algorithm, where the *regions* are defined by an agreed upon set of gentrification metrics.

Our a priori assumption was that increases in the purchase price of a dwelling are indicative of gentrification, but that dwelling prices and the desirability of an area are also influenced by the mix of establishments. In line with Papachristos et al. (2011), Chapple et al. (2017), and Zapatka and Beck (2020), we find that income tends to be positively associated with dwelling price change. Areas with higher proportions of African and Latin American immigrants tend to have lower prices. Conversely, and in line with previous findings in Toronto (Ley et al., 2002; Ley & Germain, 2000), areas with higher proportions of Asian immigrants tend to have larger positive dwelling price changes. We find that business entropy has a positive, but insignificant, influence on prices at the population level but there is significant variation in its influence across regions. A similar pattern is observed for business retirements. In both cases, the influence of these variables on prices varies across regions defined based on known measures of gentrification. This observation suggests that there is unobserved attributes variation between establishments, likely due to the quality of establishments being retired (e.g., fast food chains being replaced by upscale restaurants). Interestingly, ethnic mix is correlated with higher prices, suggesting that gentrifying households value diversity in their region, or the diversity of amenities associated with such regions. The significant role of the entropy and establishment type measures support our hypothesis that gentrification includes both residential and non-residential components. Establishment entropy is valued by those looking to live in Toronto, but the effect exhibits significant spatial variation, with negligible residual effects being captured in the population-level parameter. The models developed in this study can be used for policy analysis focused on countering gentrification. For instance, land use polices that influence the mix of businesses in a region. However, our results suggest a need for additional details to better understand the relationship between residential and non-residential drivers of gentrification. Business characteristics and whether they sell luxury goods (i.e., the price for products) would aid in policy analysis.

It is evident from our analysis that the question of gentrification is a complex, but important topic, requiring further study. Leveraging a variety of novel data sources, we were able to integrate demography and residential real estate dynamics with firmographic changes in the region. A longer time series for the data would be useful to examine the process of gentrification. With such data, we could examine the timing of changes (i.e., do changes in land use entropy precede redevelopment). Our intuition is that the reuse of existing buildings by new establishments makes this a faster transition, whereas residential redevelopment often requires new construction. The InfoCanada data on establishment type are not available prior to 2011. In addition, we initially hoped to obtain more detailed characteristics of retail establishments. Access to such data would allow us to consider the transition of retail establishments in an area towards upscale alternatives. This can be a driver of gentrification and pricing out of the existing population, which requires additional focus. Measuring gentrification is further complicated by data aggregation because we are unable to determine whether lower-income households are moving out of the area. It may be the case that high-income households are moving into an area without displacing the existing low-income population. Finally, although the hierarchical structure captures spatial effects, there remains residual spatial autocorrelation (captured through a CAR correlation structure).

CRediT authorship contribution statement

Jason Hawkins: Conceptualization, Methodology, Data curation, Software, Writing - Original draft preparation. **Usman Ahmed:** Methodology, Data curation, Software, Writing - Original draft preparation. **Matthew Roorda:** Supervision, Writing - Reviewing and editing. **Khandker Nurul Habib**: Supervision, Writing - Reviewing and editing.

Declaration of competing interest

The authors declare no conflicts of interests. They are grateful to the attendees of the 2019 NARSC annual meeting and several reviewers through multiple iterations of this paper for their valuable comments.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cities.2022.103708.

References

Araldi, A., & Fusco, G. (2019). Retail fabric assessment: Describing retail patterns within urban space. *Cities, 85*, 51–62. https://doi.org/10.1016/j.cities.2018.11.025

August, M., & Walks, A. (2018). Gentrification, suburban decline, and the financialization of multi-family rental housing: The case of Toronto. *Geoforum, 89*, 124–136. https://doi.org/10.1016/j.geoforum.2017.04.011

Bader, M. D. M., Ailshire, J. A., Morenoff, J. D., & House, J. S. (2010). Measurement of the local food environment: A comparison of existing data sources. *American Journal of Epidemiology, 171*, 609–617. https://doi.org/10.1093/aje/kwp419

Block, J. P., Scribner, R. A., & Desalvo, K. B. (2004). Fast food, race/ethnicity, and income: A geographic analysis. *American Journal of Preventive Medicine, 27*, 211–217. https://doi.org/10.1016/j.amepre.2004.06.007

Brown-Saracino, J. (2017). Explicating divided approaches to gentrification and growing income inequality. *Annual Review of Sociology, 43*, 515–539. https://doi.org/10.1146/annurev-soc-060116-053427

Bürkner, P. C. (2017). Brms: An R package for bayesian multilevel models using Stan. *Journal of Statistical Software, 80*. https://doi.org/10.18637/jss.v080.i01

Carroll, G. R., & Torfason, M. T. (2011). Restaurant organizational forms and community in the U.S. In 2005. *City & Community, 10*, 1–24. https://doi.org/10.1111/j.1540-6040.2010.01350.x

Chapple, K., & Jacobus, R. (2009). Retail trade as a route to neighborhood revitalization. *Urban and Regional Policy and its Effects, 2*, 19–68.

Chapple, K., Loukaitou-Sideris, A., González, S. R., Kadin, D., & Poirier, J. (2017). *Transit-oriented development & commercial gentrification: Exploring the linkages. Report prepared for UC CONNECT*, 112p.

Curran, W. (2007). "From the frying pan to the oven": Gentrification and the experience of industrial displacement in Williamsburg, Brooklyn. *Urban Studies, 44*, 1427–1440. https://doi.org/10.1080/00420980701373438

D'Acci, L. (2019). Quality of urban area, distance from city centre, and housing value. Case study on real estate values in Turin. *Cities, 91*, 71–92. https://doi.org/10.1016/j.cities.2018.11.008

Dantas, A. (1988). In *, 31. Overspill as an alternative style of gentrification: The case of Riverdale, Toronto* (pp. 73–86). Publication Series - University of Waterloo, Department of Geography.

Ding, L., Hwang, J., & Divringi, E. (2016). Gentrification and residential mobility in Philadelphia. *Regional Science and Urban Economics, 61*, 38–51.

DMTI. (2016). *Enhanced points of interest (EPOI) data*.

Duque, J. C., Anselin, L., & Rey, S. J. (2012). The max-p-regions problem. *Journal of Regional Science, 52*, 397–419. https://doi.org/10.1111/j.1467-9787.2011.00743.x

Easton, S., Lees, L., Hubbard, P., & Tate, N. (2020). Measuring and mapping displacement: The problem of quantification in the battle against gentrification. *Urban Studies, 57*, 286–306. https://doi.org/10.1177/0042098019851953

Ferm, J. (2016). Preventing the displacement of small businesses through commercial gentrification: Are affordable workspace policies the solution? *Planning Practice and Research, 31*, 402–419. https://doi.org/10.1080/02697459.2016.1198546

Fischer, M. M. (1980). Regional taxonomy: A comparison of some hierarchic and non-hierarchic strategies. *Regional Science and Urban Economics, 10*, 503–537. https://doi.org/10.1016/0166-0462(80)90015-0

Freeman, L. (2005). Displacement or succession? *Urban Affairs Review, 40*, 463–491. https://doi.org/10.1177/1078087404273341

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2020). *Bayesian data analysis* (3rd ed.). Boca Raton, FL, FL: CRC Press. https://doi.org/10.1109/5992.931908

González, S. R., Loukaitou-Sideris, A., & Chapple, K. (2019). Transit neighborhoods, commercial gentrification, and traffic crashes: Exploring the linkages in Los Angeles and the Bay Area. *Journal of Transport Geography, 77*, 79–89. https://doi.org/10.1016/j.jtrangeo.2019.04.010

Grube-Cavers, A., & Patterson, Z. (2015). Urban rapid rail transit and gentrification in Canadian urban centres: A survival analysis approach. *Urban Studies, 52*, 178–194. https://doi.org/10.1177/0042098014524287

Hall, C. (2019). *Who's got the election message Canadians want to hear?* CBC News.

InfoCanada. (2017). *Greater golden horseshoe 2017 business list*.

Joe Schlichtman, J., & Patch, J. (2008). Contextualizing impressions of neighborhood change: Linking business directories to ethnography. *City & Community, 7*, 273–293. https://doi.org/10.1111/j.1540-6040.2008.00261.x

Joy, M., & Vogel, R. K. (2015). Toronto's governance crisis: A global city under pressure. *Cities, 49*, 35–52. https://doi.org/10.1016/j.cities.2015.06.009

Kubrin, C. E., Squires, G. D., Graves, S. M., & Ousey, G. C. (2011). Does fringe banking exacerbate neighborhood crime rates? *Criminology & Public Policy, 10*, 437–466. https://doi.org/10.1111/j.1745-9133.2011.00719.x

Laska, S. B., Seaman, J. M., & McSeveney, D. R. (1982). Inner-city reinvestment: Neighborhood characteristics and spatial patterns over time (New Orleans). *Urban Studies, 19*, 155–165. https://doi.org/10.1080/00420988220080281

Lee, Y. Y. (2010). Gentrification and crime: Identification using the 1994 Northridge earthquake in Los Angeles. *Journal of Urban Affairs, 32*, 549–577. https://doi.org/10.1111/j.1467-9906.2010.00506.x

Lees, L., Slater, T., Wyly, E., & Taylor, R. (2008). *Gentrification*. London: Routledge.

Lens, M. C., & Meltzer, R. (2016). Is crime bad for business? Crime and commercial property values in New York city. *Journal of Regional Science, 56*, 442–470. https://doi.org/10.1111/jors.12254

Lewis, L. V. B., Sloane, D. C., Nascimento, L. M., Diamant, A. L., Guinyard, J. J., Yancey, A. K., & Flynn, G. (2005). African Americans' access to healthy food options in South Los Angeles restaurants. *American Journal of Public Health, 95*, 668–673. https://doi.org/10.2105/AJPH.2004.050260

Ley, D., & Germain, A. (2000). Immigration and the changing social geography of large Canadian cities. *Plan Canada, 40*, 29–32. https://doi.org/10.25316/IR-236

Ley, D., Tutchener, J., & Cunningham, G. (2002). Immigration, polarization, or gentrification? Accounting for changing house prices and dwelling values in gateway cities. *Urban Geography, 23*, 703–727. https://doi.org/10.2747/0272-3638.23.8.703

Lin, J.-J., & Yang, S.-H. (2019). Proximity to metro stations and commercial gentrification. *Transport Policy, 77*, 79–89.

Lym, Y. (2021). Exploring dynamic process of regional shrinkage in Ohio: A Bayesian perspective on population shifts at small-area levels. *Citieis, 115*.

Meltzer, R. (2016). Gentrification and small business: Threat or opportunity? *Cityscape: A Journal of Policy Development and Research, 18*, 57–86.

Meltzer, R., & Schuetz, J. (2012). Bodegas or bagel shops? Neighborhood differences in retail and household services. *Economic Development Quarterly, 26*, 73–94. https://doi.org/10.1177/0891242411430328

Owens, A. (2012). Neighborhoods on the rise: A typology of neighborhoods experiencing socioeconomic ascent. *City and Community, 11*, 345–369. https://doi.org/10.1111/j.1540-6040.2012.01412.x

Papachristos, A. V., Smith, C. M., Scherer, M. L., & Fugiero, M. A. (2011). More coffee, less crime? The relationship between gentrification and neighborhood crime rates in Chicago, 1991 to 2005. *City and Community, 10*, 215–240. https://doi.org/10.1111/j.1540-6040.2011.01371.x

Parker, J. N. (2018). Negotiating the space between avant-garde and "hip enough": Businesses and commercial gentrification in Wicker Park. *City and Community, 17*, 438–460. https://doi.org/10.1111/cico.12294

Preis, B., Janakiraman, A., Bob, A., & Steil, J. (2020). Mapping gentrification and displacement pressure: An exploration of four distinct methodologies. *Urban Studies*. https://doi.org/10.1177/0042098020903011

Reades, J., De Souza, J., & Hubbard, P. (2019). Understanding urban gentrification through machine learning. *Urban Studies Journal Limited, 56*, 922–942. https://doi.org/10.1177/0042098018789054

Rey, S. (2019). Geographical analysis: Reflections of a recovering editor. *Geographical Analysis*. https://doi.org/10.1111/gean.12193

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy, 82*, 34–55.

Sabourin, J. M. (1994). The process of gentrification: Lessons from an inner-city neighbourhood. In F. Fisken (Ed.), *The changing Canadian metropolis: A public policy perspective* (pp. 259–292). Toronto: Canadian Urban Institute.

Sampson, R. J., Raudenbush, S. W., & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science, 277*, 918–924. https://doi.org/10.1126/science.277.5328.918

Siksna, A. (1997). The effect of block size and form in North American and Australian city centres. *Urban Morphology, 1*, 19–33.

Skaburskis, A. (2012). Gentrification and Toronto's changing household characteristics and income distribution. *Journal of Planning Education and Research, 32*, 191–203. https://doi.org/10.1177/0739456X11428325

Skaburskis, A., & Nelson, K. (2014). Filtering and gentrifying in Toronto: Neighbourhood transitions in and out from the lowest income decile between 1981 and 2006. *Environment and Planning A, 46*, 885–900. https://doi.org/10.1068/a4666

Small, M. L., & McDermott, M. (2006). The presence of organizational resources in poor urban neighborhoods: An analysis of average and contextual effects. *Social Forces, 84*, 1697–1724. https://doi.org/10.1353/sof.2006.0067

Sutton, S. A. (2010). Rethinking commercial revitalization: A neighborhood small business perspective. *Economic Development Quarterly, 24*, 352–371. https://doi.org/10.1177/0891242410370679

The Canadian Press. (2019). In *Toronto home sales up 10% for the month of June, TREB says* (p. 2). CBC News.

Walks, R. A., & Maaranen, R. (2008). *The timing, patterning, & forms of gentrification & neighbourhood upgrading in Montreal, Toronto, & Vancouver, 1961 to 2001*. Centre for Urban and Community Studies, Cities Centre, University of Toronto.

Yoon, H., & Currid-Halkett, E. (2015). Industrial gentrification in West Chelsea, New York: Who survived and who did not? Empirical evidence from discrete-time survival analysis. *Urban Studies, 52*, 20–49. https://doi.org/10.1177/0042098014536785

Zapatka, K., & Beck, B. (2020). Does demand lead supply? Gentrifiers and developers in the sequence of gentrification, New York City 2009–2016. *Urban Studies*. https://doi.org/10.1177/0042098020940596

Zukin, S., Trujillo, V., Frase, P., Jackson, D., Recuber, T., & Walker, A. (2009). New retail capital and neighborhood change: Boutiques and gentrification in New York city. *City and Community, 8*, 47–64. https://doi.org/10.1111/j.1540-6040.2009.01269.x