Cities 95 (2019) 102482

Contents lists available at ScienceDirect

Cities

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

A comparison of the approaches for gentrification identification

Cheng Liu[a,b], Yu Deng[c,*], Weixuan Song[d], Qiyan Wu[e,f], Jian Gong[a]

[a] School of Public Administration, China University of Geosciences, No. 388 Lumo Road, Wuhan, China
[b] School of Environment, The University of Auckland, Private Bag 920190, Auckland, New Zealand
[c] Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, No. 11A, Datun Road, Chaoyang District, Beijing 100101, China
[d] Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing 210008, China
[e] School of Public Policy and Administration, Xi'an Jiaotong University, Xi'an, China
[f] Urban Studies Program, Simon Fraser University, Suite 2100, 515 W Hastings St, Vancouver, BC V6B 5K3, Canada

ARTICLE INFO

Keywords:
Gentrification identification
Threshold
K-means clustering
Housing reinvestment
Displacement

ABSTRACT

Gentrification can be identified via a threshold-based method and/or a machine-learning approach. The former, which is simple and theoretically sound, is complementary to the latter, which is objective. In view of a lack of research on exploiting the strengths of both approaches, this study compares a threshold-based method to *K*-means clustering. Using the city of Auckland as a case study, we find that both approaches are in accord with each other. The maximum degrees of similarity (falling in the range 0–1) between the identification results of both approaches are 0.80 and 0.56 for binary and three-level identification, respectively. By comparison, it is evident that the threshold-based set of identification rules delineates gentrification more accurately. For example, a census tract with a confluence of housing reinvestment and at least one aspect of social upgrading is more likely to be identified as gentrified. Moreover, gentrification in Auckland assumes various appearances. Retaining a simple and universal conceptual and analytical framework for gentrification helps us focus on the essentials of this urban phenomenon: reinvestment and displacement.

1. Introduction

As Smith successfully predicted in his seminal work (2002), the advent of the globalization of gentrification has exerted a profound influence on global economic and social inequalities, despite appearing in different guises and in various cultures (Ley & Teo, 2014). Rampant planetary gentrification (Lees, Shin, & López-Morales, 2016) calls for research on the "geography of gentrification" (Ley, 1996) from the perspective of comparative urbanism (Lees, 2012), *i.e.*, the "systematic study of similarity and difference among cities or urban processes" (Nijman, 2007, p. 1). As a popular area of research in urban geography, comparative approaches have the potential to yield some fresh insights into gentrification, and to promote social justice (González, 2016). However, the methodologies employed in comparative analysis are not straightforward (Lees, 2018). Hitherto, a cornucopia of comparative strategies and approaches has been proposed (Lagendijk et al., 2014; Phillips & Smith, 2018; Robinson, 2015; Tilly, 1984), attempting to address different issues in a global urban context. A highly-pertinent tactic for analyzing commonality and variation is discussed by Schmid et al. (2017). First, urbanization processes (*e.g.* gentrification) are defined to detect common underlying mechanisms and traits, and consequently, can be employed to identify an identical problem across various cases. Second, a twofold transductive research strategy is adopted to address the specificities developing in different urban territories. The first step is a synchronic (horizontal) analysis investigating the expansion and interweaving of urban processes and their resultant patterns of urban configurations. The second step is a diachronic/vertical analysis following the contemporary and future pathway/trajectory of urbanization, and reconstructing its history. Schmid's viewpoint on the significance of defining and identifying urban processes dovetails with that of Nijman: among the fundamental theoretical questions of comparative urbanism, the first is the spatial identification of the city itself (Nijman, 2007). This can be extended into the "geography of gentrification", as an issue of gentrification identification (Van Criekingen & Decroly, 2003).

As gentrification unfolds globally, gentrification identification has proven to be one of the most important topics in urban geography, and has attracted widespread attention. Unfortunately, gentrification measuring is fraught with stumbling blocks and controversies, owing to collective uncertainty regarding the conceptualization and operationalization of gentrification (Brown-Saracino, 2017). The former is caused by the ambiguity of the definition, whereas the latter results from a lack of consensus on, *e.g.*, indicators and methods of measurement. Although an unbiased identification of gentrification requires juggling precise and unambiguous concept,

* Corresponding author.
E-mail addresses: chengliu85@hotmail.com (C. Liu), dengy@igsnrr.ac.cn (Y. Deng).

https://doi.org/10.1016/j.cities.2019.102482
Received 29 October 2018; Received in revised form 19 August 2019; Accepted 14 October 2019
Available online 03 November 2019
0264-2751/

reliable indicators, and objective approaches, an appropriate ontological conceptualization can inform the selection of algorithm and input variables. On the flip side, strict operationalization is likely to improve rigor in conceptualization. However, the synergy between conceptualization and operationalization is barely examined systematically. Hence, this study attempts to identify and fill gaps in the preceding research on gentrification measurement.

2. Literature review

Owing to the heterogeneity of gentrification, the definitions and measurements of gentrification are highly debatable. This section provides overviews of both, respectively.

2.1. Conceptualization of gentrification

When gentrification travels temporally, the concept is altered, leading to a wealth of different definitions, such as rural gentrification (Phillips, 1993), super gentrification (Lees, 2000), new-build gentrification (Davidson & Lees, 2005), studentification (Smith, 2005), and tourism gentrification (Gotham, 2005). Moreover, gentrification transcends the Global North and assumes different appearances in different political and social settings in the Global East (Hyun Bang, Lees, & Ernesto, 2015) and the Global South (López-Morales, 2015), *e.g.*, state-sponsored gentrification (He, 2007) and jiaoyufication (Wu, Edensor, & Cheng, 2018). Hence, the scope of gentrification has been extended in various localities (Paul, 2015).

To account for the mutations owing to the temporal and/or social diversifications (*i.e.,* the sub-concepts of gentrification) and spatial heterogeneity (*i.e.,* gentrification outside the Global North) (Maloutas, 2011), the concept of gentrification has significantly increased its breadth of content. After comparing Neil Smith's definition of gentrification in 1996 (Smith, 1996) to that in 1982 (Smith, 1982), Boddy (2007) finds that "redevelopment" or "new buildings on previously developed land", phrases excluded from the previous version, are subsequently encompassed under the umbrella of gentrification. In fact, Smith's definition in 1996 confines gentrification to a central city, neglecting rural gentrification (Phillips, 1993). To stretch the term sufficiently wide for similar processes, Hackworth (2002) conceptualizes gentrification as "the production of urban space for progressively more-affluent users". Consequently, gentrification becomes a "far broader process linked to the profound transformation in advanced capitalism" (Sassen, 1991, p. 255).

Rose (1984) and Beauregard (1986) criticize that gentrification is a nebulous and chaotic concept, as it combines necessary tendencies with contingent conditions (such as particular gentrifiers at a particular time), and obscures the multiplicity of the process. That diverse processes in a distinct context are subsumed under the same rubric of a term is possibly problematic: (1) local and individual particularities and variance are prone to being ignored (Mosselson, 2016), and (2) the ability to realize new insights is undermined (Bondi, 1999). Some even contend that the term has been stretched "beyond the point at which it remains useful and credible as a means of understanding the processes at work" (Boddy, 2007, p. 86), and that gentrification should be treated as simply a topic of research, rather than any type of explanatory theory (Smart & Smart, 2017). Some even argue that gentrification is doomed (Bourne, 1993).

To relieve or remedy the definitional overload, some researchers (Bondi, 1999; Rose, 1984) suggest that the term should be allowed to disaggregate under the weight of its own burden. Other students claim that alternate terms or phrases such as "reurbanization" (Lambert & Boddy, 2002) or "hegemonic-cum-alienated redevelopment" (Tang, 2017) should be used in its place. Specifically, stretching gentrification to encompass different socio-political contexts at different stages draws similar skepticism (Álvarez-Rivadulla, 2007; Cartier, 2017; Lui, 2017; Yip & Tran, 2015) against the backdrop of the globalization of gentrification.

Claiming that the disintegration of gentrification would simply cause another form of chaotic conception, Clark (2005) opines that "[t]here is nothing chaotic about gentrification in inner cities and in rural areas, in neighbourhoods and in non-residential areas, through rehabilitation and through demolition/new construction. There is, however, something chaotic about conceptualising gentrification according to these aspects, since none of them stands in a necessary relation to its occurrence" (p. 264). Moreover, scholars (Butler, 2007; Lees, Slater, & Wyly, 2008) argue for retaining the label "gentrification", as its enormous political clout successfully draws international media attention to class-based displacements and social inequalities. This explains governments preferring to use "redevelopment", "revitalization" or "regeneration" to glamorize gentrification (Ley & Teo, 2014). Retaining gentrification entails conceptual conflation (Clark, 2005), or alternatively, conceptual dilution (Butler, 2007), potentially jeopardizing its usefulness. Maloutas (2011, 2017) debunks Clark's generic definition (Clark, 2015) by maintaining that the broad and ill-defined boundary of gentrification removes its indelible contextual attachments such as contextual causality. Although conceptual stretching is an inevitable risk (Davidson & Lees, 2005), it does not completely preclude spatiotemporal contextual sensitivity to gentrification. On the contrary, the analysis framework for comparative planetary gentrification (Atkinson & Bridge, 2005; Lees et al., 2016) endeavors to manage the side effects of stretching the definition by scrutinizing the commonality and dissimilarity of gentrification in different spatial and/or temporal contexts, thereby providing a useful point of departure for comparative analysis (Forrest, 2015). For instance, Mosselson (2016) argues that gentrification should be discussed concurrently with vernacular narratives and expositions to provide a better understanding of gentrification in inner-city Johannesburg.

Based on the discussion above, this study selects an "elastic yet targeted definition of gentrification" (Clark, 2005) for the ensuing identification and analysis, as follows:

> Gentrification is a process involving a change in the population of land-users such that the new users are of a higher socio-economic status than the previous users, together with an associated change in the built environment through a reinvestment in fixed capital. The greater the difference in socio-economic status, the more noticeable the process, not least because the more powerful the new users are, the more marked will be concomitant change in the built environment. It does not matter where, and it does not matter when. Any process of change fitting this description is, to my understanding, gentrification (p. 263).

By defining gentrification by its outcomes instead of by its causes or everyday character (Brown-Saracino, 2010), Clark's definition of gentrification has several major advantages in gentrification identification. First, it articulates two core components (*i.e.,* displacement of poor and working-class residents, and reinvestment). Second, it has the fewest contingent causes and relations (*i.e.,* gentrification is independent of place and time), rendering the definition less mutable, and providing conceptual clarity and theoretical rigor (Maloutas, 2011).

2.2. Operationalization of gentrification

There is a myriad of research identifying gentrification, both qualitatively and quantitatively. A comparison between both approaches in gentrification research is provided by Brown-Saracino (2017). Focusing on quantitative measures, previous gentrification efforts are classified into three categories, based on Clark's definition of gentrification: improvements to the built environment (*e.g.,* renovation), social upgrading of neighborhoods, and a combination of both. The first perspective generally measures capital reinvestment, *e.g.*, housing price increases (Figueroa, 1995) or property tax arrears (Smith & DeFilippis, 1999). However, renovation-oriented operationalization is uncommon. *Per contra*, neighborhood changes and the associated displacement of low-income residents are more prevalent in mapping gentrification (Atkinson, 2000) owing to easier and more comprehensive data availability, and the overriding importance of socio-demographic change (Slater, 2006). Freeman (2005) points out that displacement becomes virtually synonymous with gentrification. For example, Gould Ellen and O'Regan (2008) claim that neighborhoods with a substantial increase in relative income in a decade are deemed to have experienced a

substantial economic gain. Landis (2016) identifies gentrifying tracts as those with an upward shift of median household income by more than two deciles, and those currently within the first four income deciles of their respective metropolitan area. Other scholars (Ding, Hwang, & Divringi, 2016; Galster & Peacock, 1986) measure gentrification by merging factors of reinvestment of capital with those of socio-demographic change. Moreover, owing to data limitations, researchers augment regularly-used indicators with other measures, such as field survey data (Hammel & Wyly, 1996; Wyly & Hammel, 1998), the number of coffee shops (Papachristos, Smith, Scherer, & Fugiero, 2011), and newspaper coverage (Barton, 2016). With the advent of big data, the data sources have further diversified. For example, scholars (Hwang & Sampson, 2014; Ilic, Sawada, & Zarzelli, 2019; Liu, Silva, Wu, & Wang, 2017; Naik, Kominers, Raskar, Glaeser, & Hidalgo, 2017) use a streetscape to capture a visible change of gentrification.

As an umbrella term, gentrification involves multifaceted complex socio-economic phenomena and processes. Because of the multidimensionality of gentrification, there appears to be a growing consensus that changes across several indicators can measure gentrification without oversimplification (Anguelovski, Connolly, Masip, & Pearsall, 2018). Bostic and Martin (2003) go even further, arguing that using a single variable to identify the multidimensionality of gentrification is almost certain to fail. As number of indicators increases, a different problem arises: how to address multiple variables simultaneously in operationalization? A straightforward approach is to compare several indicators (*e.g.,* socio-economic variables and/or human-derived ratings for the perception of a streetscape's safety) to predefined thresholds, to determine whether gentrification has transpired. Generally, the thresholds of each variable are set by a rule of thumb, such as higher-than-metropolitan-median/-average change (Choi, Van Zandt, & Matarrita-Cascante, 2018; Lipton, 1977) or threshold quantiles (Freeman, 2005). To avert multiple thresholds, a composite index is compiled via linear combination of socio-economic variables (Hee-Jung, 2013; Hwang & Lin, 2016; Ley, 1986, 1992; Timberlake & Johns-Wolfe, 2016) or principal component analysis (PCA) (Meligrana & Skaburskis, 2005; Owens, 2012; Reades, De Souza, & Hubbard, 2018; Walks & Maaranen, 2008a, 2008b). In a similar vein, when a change in the index over time exceeds a threshold, gentrification occurs. Sometimes, single indicators and composite indices are employed simultaneously to measure gentrification (Holm & Schulz, 2018; Van Criekingen & Decroly, 2003).

Taking advantage of the data pattern, another approach, *i.e.*, a classification algorithm, categorizes neighborhoods as "gentrified" or not without selection of thresholds. For example, a K -means clustering algorithm can be adopted for identifying gentrification (Delmelle, 2016; Podagrosi, Vojnovic, & Pigozzi, 2011; Ye, Vojnovic, & Chen, 2015). As compared to the descriptive statistics (*e.g.*, averages, percentage variations over time) commonly employed in prior studies, more sophisticated statistical techniques (such as K -means clustering) are rarely used (Schuler, Kent, & Monroe, 1992).

It is noteworthy that most methods of gentrification identification use a geographic information system (GIS) analytically and/or cartographically, which is the epitome of "critical GIS" (O'Sullivan, 2006): engagement of social theory with GIS and mathematics. Critical GIS not only maps the presence of gentrification in the landscape and positions it in space, but also helps to interrogate the process, exploring the gentrification theories or considering alternatives to what might be assumed as inevitable (O'Sullivan, Bergmann, & Thatcher, 2018). Two cases are highlighted to flesh out this argument. The first case concerns how community members from the "South of Market" neighborhood (hereafter abbreviated to SoMa) in San Francisco resisted gentrification via GIS-based living neighborhood map (Parker & Pascual, 2002). As a haven for start-up businesses and blue-collar workers, SoMa was under mounting pressure from gentrification, owing to strong housing demand from well-paid high-tech young professionals from Silicon Valley. To confront the city Planning Department regarding the conflict between small business and residents, the existing inhabitants formed a coalition and pinpointed and mapped displaced businesses, using phone or in-person interviews. By developing an "ontological power" (Pavlovskaya, 2009, p. 162) of gentrification, the map educated community members and provided political ammunition to local residents who voiced their concerns based on concrete evidence, rather than on emotion and/or hearsay. Eventually, the city Planning Department allowed a compromise, by approving an "Interim Industrial Protection Zone" to control gentrification. As a second example, Charles (2014) creates a heat map of redeveloped housing stock via Moran's I statistics, and visually identifies gentrification by overlaying the heat map upon that of socio-economic variables. Charles finds that the spatial clustering of teardowns in Chicago was initiated by homeowners who chose to rebuild larger houses for themselves, and who were succeeded by profit-seeking and risk-averse developers who speculated in adjacent areas. Charles' study brings freeholders as pioneer gentrifiers to light, challenging stereotypes of the spearhead of gentrification as low-income and risk-oblivious citizens, and indicating how critical GIS can potentially produce fresh insights into gentrification theories.

The threshold-based methods and K -means clustering have been proposed to tackle the challenges of the conceptualization and operationalization of gentrification identification, respectively. From the perspective of gentrification theory, a threshold-based method identifies gentrification via changes in socio-economic factors, which are simple, convenient, and theoretically sound. However, the threshold is more or less arbitrary (if logical) (Freeman, 2009). As a possible remedy for the arbitrariness of thresholds, the classification approach is based on a theory of inferential statistics, mainly including supervised and unsupervised classification. The supervised classification is guided by the presence of an outcome variable (*e.g.,* gentrified or not/level of gentrification), whereas the unsupervised classification only concerns the feature(s) (*e.g.,* the socio-economic factors of a census tract), and does not record or address measurements (*i.e.,* interpretations) of the outcome (Hastie, Tibshirani, & Friedman, 2009). For supervised gentrification identification, gathering sufficient outcome samples for training relies on a large-scale field survey, which is time-consuming and labor-intensive. Therefore, unsupervised classification is generally preferred. Although statistical methods increase the objectivity in gentrification identification, the interpretation of the results is less intuitive and more challenging, owing to the lack of a direct connection between the data pattern and the gentrification variables. An unsupervised classification algorithm cannot tell which combination of variables is necessary and/or sufficient to define "gentrification" (Boddy, 2007).

Recently, scholars have made efforts to contrast different identification methods. After comparing three different strategies (two census-data based and one employing anecdotal evidence from The New York Times), Barton (2016) concludes that the number and distribution of gentrified neighborhoods identified by each strategy vary significantly. Loukaitou-Sideris, Gonzalez, and Ong (2017) discover more consistency among assessments of the presence of gentrification by using a census-data based method, visual surveys of Google Street View, and interviews. However, a comparison of identification outcomes is insufficient for contextualizing the rules of each method and grasping the process of gentrification identification. The path to this goal requires an explicit interpretation of an identification approach in terms of gentrification variables, which is rarely studied. Thus, this study aims to find a way to link the data-based method (*i.e.*, the unsupervised classification) with the theory-based approach (*i.e.*, the threshold-based method). Instead of exhausting every possible method, we explore the threshold-based method and the K -means clustering to demonstrate their relationship, to investigate their exactitude, and to (potentially) exploit their strengths. More importantly, both approaches, albeit simple, shed a light on the applicability of theories of gentrification and statistics for gentrification identification. Thus, they are beneficial to later research on other advanced methods.

The remainder of the paper is organized as follows. Section Three

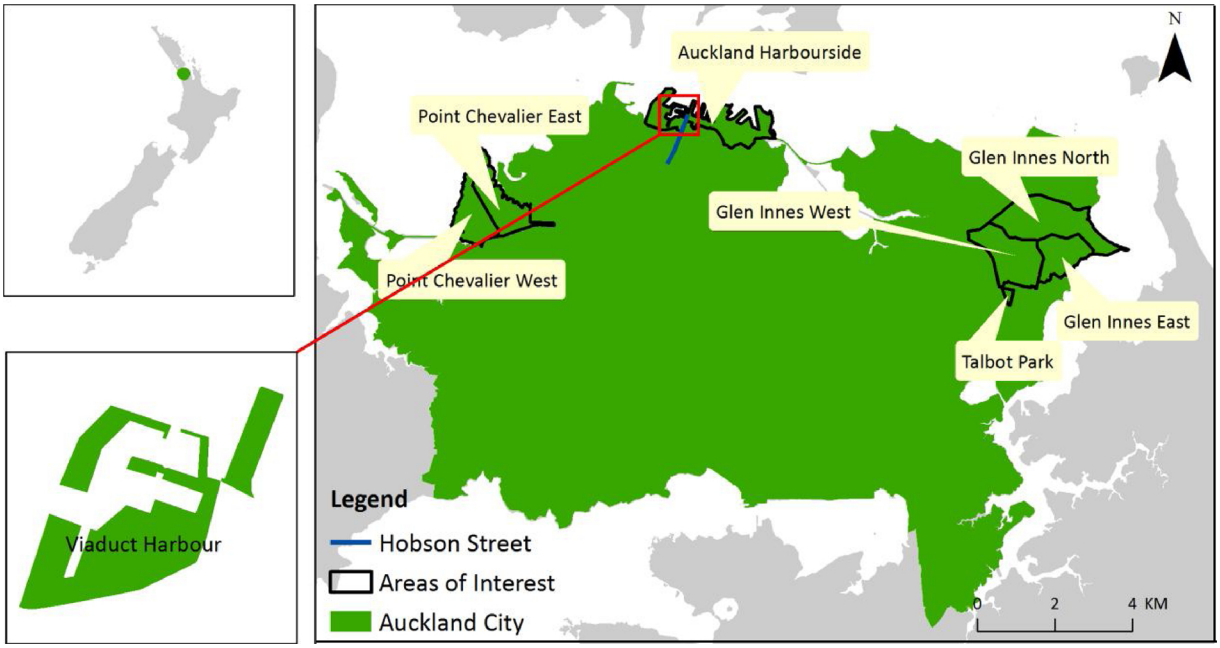

Fig. 1. Location of Auckland City. The green dot in the upper-left inset map is the Auckland City in New Zealand, and the lower-left inset map is for Viaduct Harbour in the Auckland Harbourside.

introduces the background of the study area and the empirical data used. Sections Four and Five identify and compare gentrification via the K -means clustering and the threshold-based approaches, respectively. Section Six justifies the continued use of gentrification, and Section Seven draws some conclusions and pursues directions of future work.

3. Study area and data

Situated in New Zealand's North Island (Fig. 1), the study area (Auckland city) is the largest urban area in the country. Since the 2000s, Auckland has witnessed a booming housing market, whose causes are manifold: first, Auckland is renowned for a high level of livability, attracting international tourists and investment and yielding considerable revenue; second, Auckland accepted a large number of immigrants (particularly Asian) owing to favorable immigration policy; third, owing to residing on a narrow isthmus, the city has a lack of land supply, and is unable to sustain the burgeoning housing demand. Meanwhile, new urban governance structures, in conjunction with more "entrepreneurial" state activities, have facilitated redevelopment in Auckland (Murphy, 2008). Additionally, to ease the 2008 financial crisis, all major banks in New Zealand substantially lowered their mortgage rates, rendering home buying more affordable. Accordingly, there are sufficient gentrification samples in the study area. For instance, gentrification has been studied or recorded in five areas since the 2000s (Fig. 1 and Table 1): Viaduct Harbour, Point Chevalier (East and West), Hobson Street, Talbot Park and Glen Innes (East and West). These cases will be canvassed later. The diversified gentrification, copious possibly-gentrified areas, and continuous research make Auckland an appropriate case for this research.

To reveal the dynamics of housing reinvestment and social upgrading, property data from the Auckland Council and census data from Statistics New Zealand are used in this study, respectively. The former includes data on land value (LV), improvement value (IV, value of any buildings or other structures on the land), and capital value (CV, sum of LV and IV) for 2002, 2005, and 2011 (all in New Zealand dollars). As the outputs of a property-value-appraisal model from a mass survey process, each property value is calculated by comparing recent sales in the area with the property being valued and by considering the property type, location, land size, zoning, floor area, consented building work (such as renovations), and many other factors. Following the approach of Liu, O'Sullivan, and Perry (2018), housing reinvestment is measured by the proportional difference in IV ($IVPD$, Eq. 1 and Table 2):

$$\boldsymbol{IVPD} = \frac{\boldsymbol{IV}_2}{\boldsymbol{CV}_2} - \frac{\boldsymbol{IV}_1}{\boldsymbol{CV}_1} \tag{1}$$

Here, the IV_i and CV_i are the improvement and capital values of a property at the i th measurement time (i = 1,2), respectively. The proportion of the improvement value $\left(\frac{IV}{CV}\right)$ offsets the systematic errors in the valuation model and inflation, improving the measurement reliability in housing restoration. $IVPD$ has proven to be a reliable proxy for housing reinvestment, as demonstrated by Liu et al. (2018).

When operationalizing gentrification (Table 2), four aspects of the census data from 2001, 2006, and 2013[1] are examined: occupation, income, tenure, and ethnicity, based on the usually resident population[2] . Nine original occupational categories are reclassified into three groups: professional occupations, skilled occupations, and elementary occupations. The professional and elementary occupations are considered for measuring changes in occupational structure. Instead of a change in median personal income, a rank change is adopted to obviate the adverse impacts of inflation. Tenure indicates whether or not a person owns the dwelling he or she lives in. Two ethnic groups are explored: Europeans and Polynesians (mainly Māori and Pacific Peoples).

Of the five scales of census tracts, the finest (*i.e.*, meshblock) are used to provide detailed and precise delineation and identification of gentrification in local areas (Figueroa, 1995), avoiding the ecological fallacy (O'Sullivan & Unwin, 2010)—average conditions of an area are assigned to an entire population. Hedin, Clark, Lundholm, and Malmberg (2012) made a valuable contribution by mapping gentrification in three Swedish cities with a resolution of 300 m × 300 m.

[1] Although collections of census data and property data do not coincide completely, the three time stamps of the former are close to those of the latter. Thus, it is safe to pair both data, and to investigate the neighborhood change in two phases: Phase I (from 2001/2002 to 2005/2006) and Phase II (from 2005/2006 to 2011/2013).

[2] Usually resident population is "a count of all people who usually live, and are present, in New Zealand on a given census night" (Statistics New Zealand, 2019).

Table 1
Gentrification cases studied/recorded in Auckland since the 2000s

Place of Gentrification	Time of Gentrification	Author(s)
Viaduct Harbour in Auckland Harbourside	Before 2005	Murphy (2008)
Point Chevalier (East and West)	Mostly after 2005	Liu et al. (2018)
Hobson Street	Approximately 2006	Collins (2010)
Talbot Park	Approximately 2007	Ministry for the Environment (2008)
Glen Innes (East and West)	After 2010	Gordon et al. (2017)

Table 2
Variables for gentrification identification

Components	Variables	Descriptions
Housing Reinvestment (Property Data)[a]		
Housing	*IVPD*	Proportional difference (*PD*) in improvement value (*IV*)
Social Upgrading (Census Data)[b]		
Occupation[c]	*ProfPD*	Proportional difference in the population of professional occupations, measuring changes in high-status occupations, including legislators, administrators, managers, and professionals.
	ElementPD	Proportional difference in the population of elementary occupations measures changes in low-status occupations, including those other than the professional and skilled occupations.
Income[c]	*MIRD*	Rank difference in median personal income
Tenure[c]	*NOwnPD*	Proportional difference in the population who do not own their usual residence
Ethnicity	*EuropPD*	Proportional difference in the European population
	PolyPD	Proportional difference in the Polynesian population

[a] *IVPD* between 2005 and 2002 and between 2011 and 2005 are pooled into one variable. The *IVPD* of a census tract is the average of *IVPD* of properties within the tract.
[b] Variables derived from the census data between 2006 and 2001 and between 2013 and 2006 are pooled into a whole dataset.
[c] Only those aged 15 years and over are counted.

4. Gentrification identification based on *K* -means clustering

4.1. Principle of *K* -means clustering

To investigate the consistency and accuracy of a statistical approach to gentrification identification, we adopt a *K* -means clustering algorithm. The algorithm finds clusters (levels of gentrification) in a set of unlabeled data/census tracts, so as to minimize the total within-cluster variance (Hastie et al., 2009) based on gentrification variables (Table 2). Generally, the cluster variance is measured by the squared

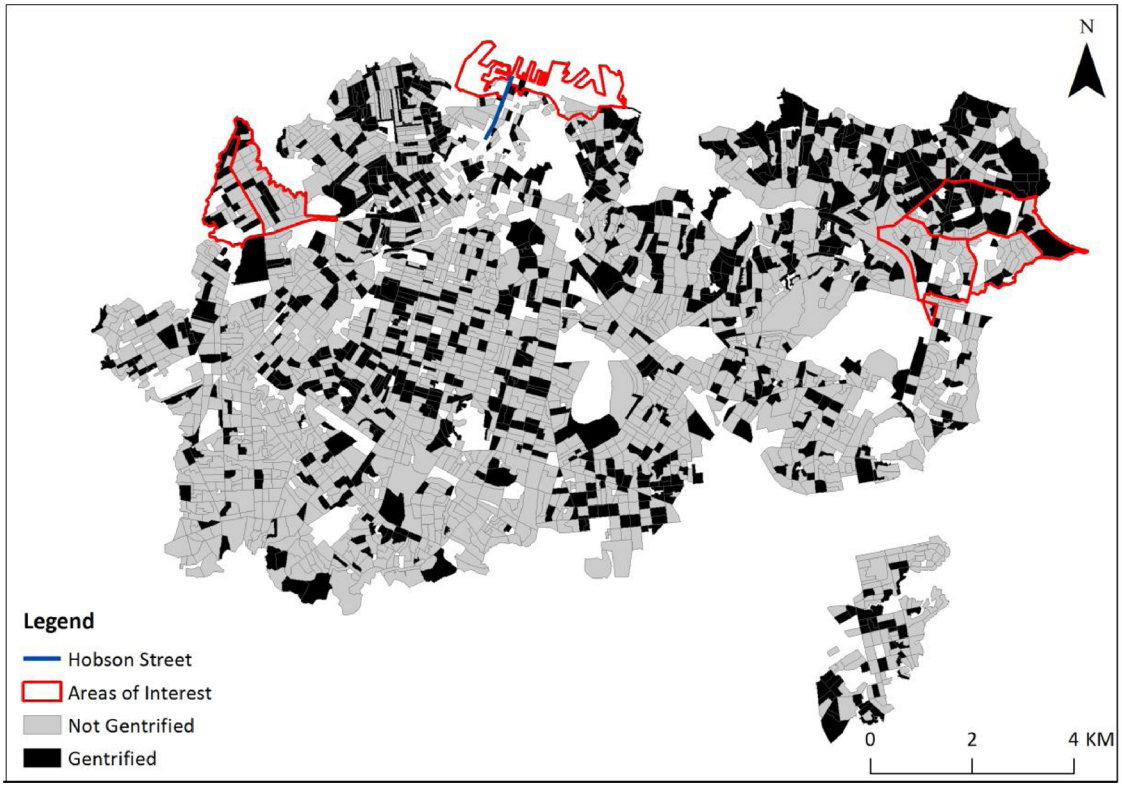

Fig. 2. Binary gentrification identification via *K* -means clustering in Phase I. Meshblocks without construction land (such as parks and roads) or without census data are not considered, forming empty spaces in the map. The empty spaces in Auckland Harbourside are from meshblocks whose census data are kept confidential.

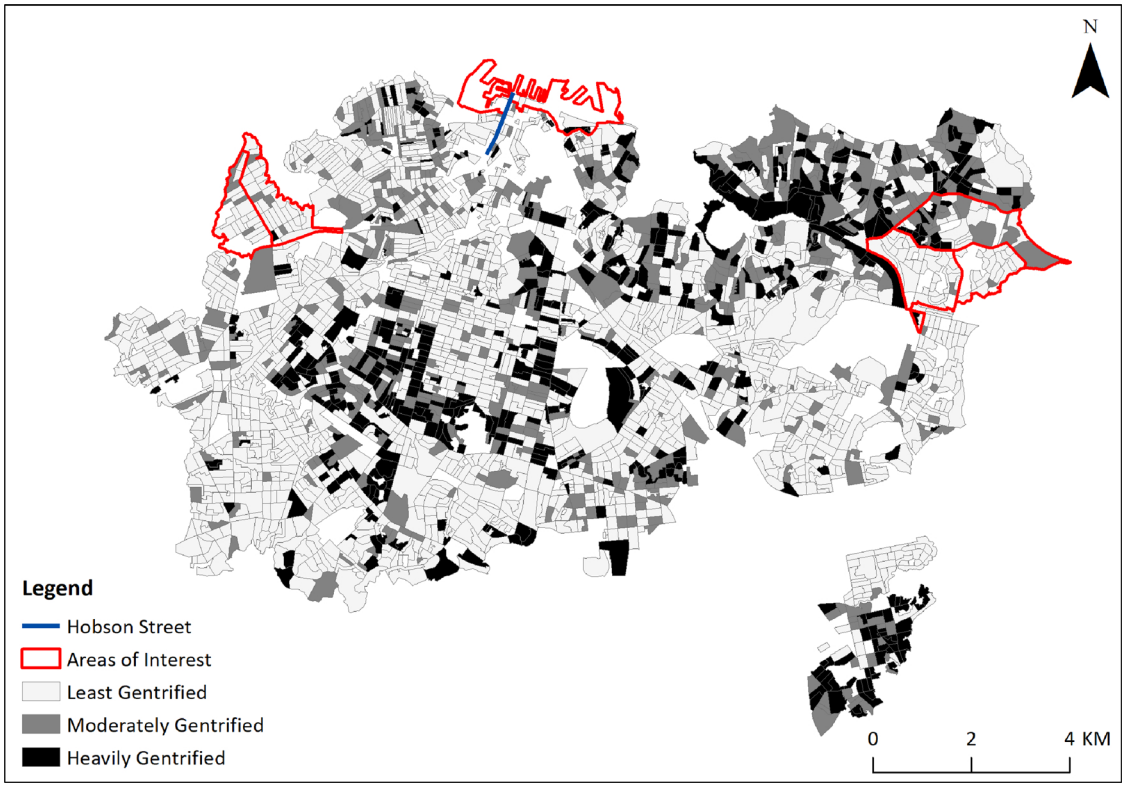

Fig. 3. Three-level gentrification identification via K -means clustering in Phase I.

Euclidean distance between a vector and its associated cluster mean:

$$\boldsymbol{d}(\boldsymbol{x_i}, \boldsymbol{x_{i'}}) = \sum_{j=1}^{p} (\boldsymbol{x_{ij}} - \bar{\boldsymbol{x}}_{\boldsymbol{j}})^2 = \|\boldsymbol{x_i} - \bar{\boldsymbol{x}}^{(\boldsymbol{k})}\|^2 \quad (2)$$

In the above, x_i is a vector with p dimensions (*i.e.*, a meshblock with p gentrification variables), and $\bar{x}^{(k)}$ corresponds to the mean vector of the k th cluster associated with x_i. The total within-cluster variance can be written as:

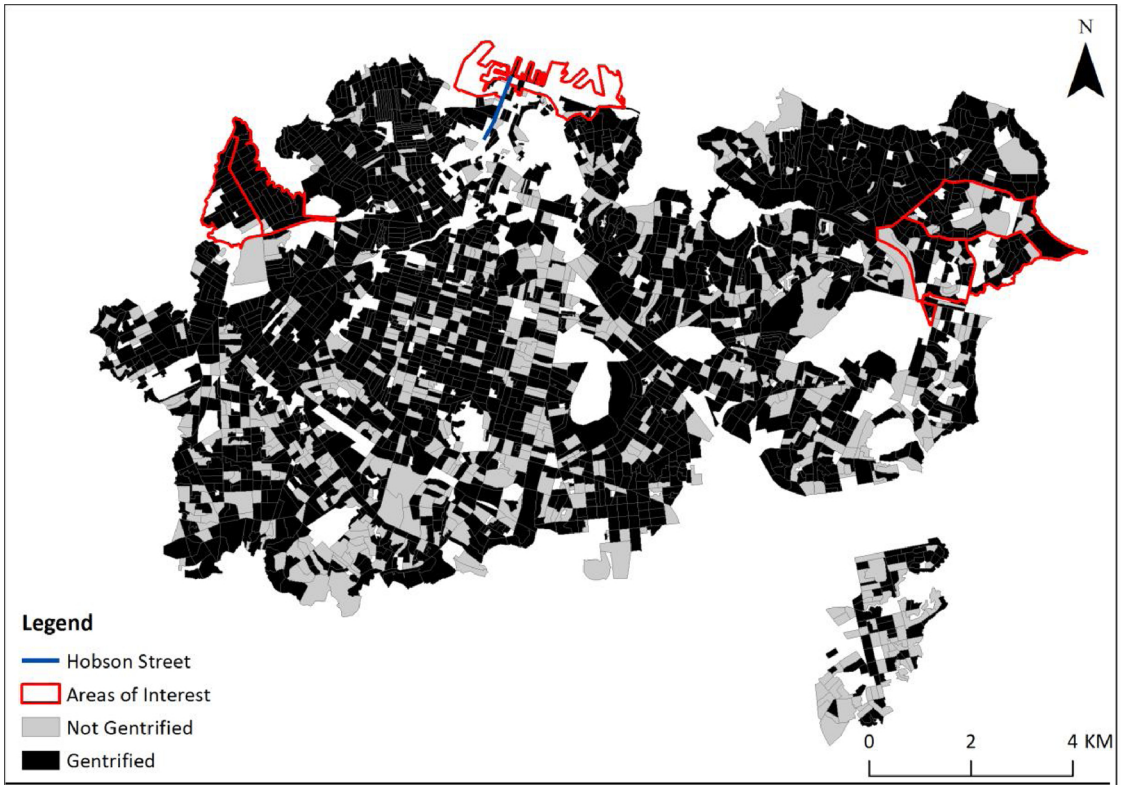

Fig. 4. Binary gentrification identification via K -means clustering in Phase II.

Fig. 5. Three-level gentrification identification via K -means clustering in Phase II.

$$V = \sum_{k=1}^{K} \sum_{i=1, i \in k}^{N} \|x_i - \bar{x}^{(k)}\|^2 \tag{3}$$

Here, $i \in k$ represents that x_i belongs to the k th cluster, and K represents the total number of clusters. When $K = 2$, meshblocks will be categorized into two opposite groups (gentrified or not); $K \geq 3$ indicates that the gentrification has different levels of intensity. To explore the effects of different intensity levels on gentrification identification, the results of $K = 2$ and $K = 3$ are examined. Without loss of generality and for ease of interpretation, $K = 3$ is selected to represent multiple intensity levels of gentrification: heavily gentrified, moderately gentrified, and least gentrified. A heavily-gentrified neighborhood undergoes drastic socio-economic changes with possible class antagonism in a revanchist city, such as New York (Smith, 1998) or Mumbai (Whitehead, 2008). Communities with mild and often peaceful social uplift are labeled as moderately-gentrified, reminiscent of Eric Clark's response when Neil Smith asked regarding the battlefields of gentrification in Malmö: "[c]onflicting interests, displacement, personal tragedies, yes, but not the desperation behind battlefields" (Clark, 2005, p. 28). Finally, those with little social upgrading are categorized as the least gentrified. The K -means clustering algorithm labels each meshblock without specifying their meaning, which is manually interpreted after the classification. The three categories are further contextualized in Section Six.

Cluster variance is susceptible to range of gentrification variables, which are therefore normalized prior to the K -means clustering process:

$$x' = \frac{x - \bar{x}}{\sigma} \tag{4}$$

In Eq. 4, x' is a normalized gentrification variable, x represents the raw data, $\bar{x}$ represents the mean of x, and σ represents the standard deviation of x.

4.2. Results of K -means clustering

Figs. 2, 3, 4 and 5 show the outcomes of the binary and three-level identification in Phases I and II (Footnote 1), respectively. For binary identification (Figs. 2 and 4), the Auckland Harbourside (Fig. 1) is identified as gentrified during Phase I, and continues to be gentrified during Phase II. This is consistent with Murphy's findings (2008) (Table 1): from 2005 onwards, Viaduct Harbour has been regenerated as a high-quality urban design precinct, and has encouraged subsequent gentrification of the waterfront. Thus, Viaduct Harbour causes gentrification straddling the two phases. Meanwhile, the majority of the areas of Point Chevalier and Glen Innes are labeled as gentrified only in Phase II, agreeing with scholars' observations (Table 1) (Gordon, Collins, & Kearns, 2017; Liu et al., 2018). In particular, both demonstrate large-scale state-led gentrification writ large. Additionally, studentification in Hobson Street and gentrification in Talbot Park (Fig. 1 and Table 1) are successfully detected in Phase II. In this regard, K -means clustering can identify gentrification relatively satisfactorily.

For the three-level identification (Figs. 3 and 5), the "patchwork" quality of gentrification (Heidkamp & Lucas, 2006; London, Lee, & Lipton, 1986) becomes bolder. The pattern, also reported in Global West (Podagrosi et al., 2011) and Global East (Ye et al., 2015), reveals how gentrification develops spatially in Auckland. During Phase I, gentrification started to emerge in Point Chevalier (East and West) and Glen Innes (East and West). Owing to the desirability of sea views, the removal of oyster shells, and rocks and sand replenishment at Point Chevalier beach (Orsman, 2009), the gentrification of Point Chevalier in Phase I was supposed to be concentrated in the northern tip, and spread into its vicinity during Phase II. Meanwhile, after the redevelopment plan was proposed in 2009, state-housing tenants in Glen Innes (East and West) were evicted, and these ex-state houses were either demolished or sold (Gordon et al., 2017). The observations suggest that there is little gentrification in the lower part of Point Chevalier (East and West) and in Glen Innes (East and West) in Phase I, which is in accord with Fig. 3 (*cf.* Fig. 2). Hence, we believe that three-

Table 3
Rules for threshold-based gentrification identification.

Rule Number	Rules[a]	Descriptions
Binary		
1	**Gentrified:** *IVPD ≥ a and MIRD ≤ b and ProfPD ≥ c and ElementPD ≤ d and NOwnPD ≤ e and EuropPD ≥ f and PolyPD ≤ g* **Not Gentrified:** *rest of others*	The most stringent rules for the occurrence of gentrification: all conditions of housing reinvestment and social upgrading should be satisfied.
2	**Gentrified:** *IVPD ≥ a and (MIRD ≤ b or ProfPD ≥ c or ElementPD ≤ d or NOwnPD ≤ e or EuropPD ≥ f or PolyPD ≤ g)* **Not Gentrified:** *rest of others*	One condition of housing reinvestment and at least one condition of social upgrading should be satisfied for the occurrence of gentrification.
3	**Gentrified:** *IVPD ≥ a or MIRD ≤ b or ProfPD ≥ c or ElementPD ≤ d or NOwnPD ≤ e or EuropPD ≥ f or PolyPD ≤ g* **Not Gentrified:** *rest of others*	The least stringent rules: at least one condition of housing reinvestment or social upgrading should be satisfied for the occurrence of gentrification.
Three-Level		
1	**Heavily Gentrified:** *IVPD ≥ i and MIRD ≤ j and ProfPD ≥ m and ElementPD ≤ n and NOwnPD ≤ p and EuropPD ≥ s and PolyPD ≤ t* **Moderately Gentrified:** *h ≤ IVPD ≤ i and j ≤ MIRD ≤ k and l ≤ ProfPD ≤ m and n ≤ ElementPD ≤ o and p ≤ NOwnPD ≤ q and r ≤ EuropPD ≤ s and t ≤ PolyPD ≤ u* **Least Gentrified:** *rest of others*	
2	**Heavily Gentrified:** *IVPD ≥ i and (MIRD ≤ j or ProfPD ≥ m or ElementPD ≤ n or NOwnPD ≤ p or EuropPD ≥ s or PolyPD ≤ t)* **Moderately Gentrified:** *h ≤ IVPD ≤ i and (j ≤ MIRD ≤ k or l ≤ ProfPD ≤ m or n ≤ ElementPD ≤ o or p ≤ NOwnPD ≤ q or r ≤ EuropPD ≤ s or t ≤ PolyPD ≤ u)* **Least Gentrified:** *rest of others*	
3	**Heavily Gentrified:** *IVPD ≥ i or MIRD ≤ j or ProfPD ≥ m or ElementPD ≤ n or NOwnPD ≤ p or EuropPD ≥ s or PolyPD ≤ t (at least four of the conditions are satisfied)* **Moderately Gentrified:** *h ≤ IVPD ≤ i or j ≤ MIRD ≤ k or l ≤ ProfPD ≤ m or n ≤ ElementPD ≤ o or p ≤ NOwnPD ≤ q or r ≤ EuropPD ≤ s or t ≤ PolyPD ≤ u (at least four of the conditions are satisfied)* **Least Gentrified:** *rest of others*	The minimum four satisfied conditions prevent meshblocks from being categorized into more than one group.

[a] The gentrification variables are explained in Table 2. a, b, c, d, e, f, and g are the binary thresholds of each gentrification variable, and (h, i), (j, k), (l, m), (n, o), (p, q), (r, s) and (t, u) are the three-level thresholds of each gentrification variable, respectively ($h < i, j < k, l < m, n < o, p < q, r < s, t < u$). The rules for three-level identification are formulated to be analogous and comparable to those of binary identification, so their relationship can be examined.

level identification is more appropriate than binary identification.

5. Threshold-based gentrification identification

5.1. Principle of threshold-based identification

Techniques for dimensional reduction (*e.g.*, PCA) are prevalent when addressing the multidimensionality of gentrification. However, instead of a linear combination of the gentrification variables, such as in PCA, this study employs structured query language (SQL)-like identification rules (Table 3) to express the threshold-based identification rules. This choice is made for two reasons. First, the results of K -means clustering from Section Four can be converted into individual thresholds for ease of explicit explanation of the nexus between both methods in terms of gentrification theory. In contrast, PCA creates a composite index, whose thresholds are challenging to be contextualized. Second, after PCA transformation of the seven gentrification variables (Table 2), the proportion variances absorbed by each component are relatively close (in descending order: 23.46%, 17.15%, 14.81%, 14.40%, 11.59%, 10.22%, and 8.37%), suggesting a weak correlation among the gentrification variables. In this section, different deciles are explored for each gentrification variable, and gentrification is identified under different combinations of thresholds (*i.e.*, deciles) of all variables. As the selection of variables materially affects gentrification identification (Meltzer & Ghorbani, 2017), three sets of identification rules are considered, according to two core components of gentrification: housing reinvestment, and social upgrading (see Clark's definition of gentrification in Section 2.1).

The results of the different combinations of thresholds and rules are compared to those of K -means clustering, via a similarity index:

$$s = \frac{n}{N} \tag{5}$$

Here, s is the similarity index ($0 \le s \le 1$), n is the number of meshblocks whose labels (gentrified or not for binary identification and heavily/moderately/least gentrified for three-level identification) are identical in both methods (K-means clustering or threshold-based identification), and N is the total number of census tracts. The thresholds and sets of rules for maximizing s are regarded as the equivalent of a K -means clustering method. For binary identification, with seven gentrification variables and nine possible thresholds (from first to ninth deciles) for each variable, there are 9^7 (4,782,969) combinations for each set of rules. However, owing to the excessive computational workload in three-level identification, only five deciles (*i.e.*, first, third, fifth, seventh, and ninth deciles) are investigated for each gentrification variables. As $h < i, j < k, l < m, n < o, p < q, r < s, t < u$ (Table 3), there are $(4 + 3 + 2 + 1)^7 = 10^7$ combinations for each set of rules (*cf.* $(8 + 7 + 6 + 5 + 4 + 3 + 2 + 1)^7 = 36^7$ for nine deciles), requiring approximately 54 hours with an E5-2640 CPU using parallel computing. An increasing level of gentrification identification incurs an exponential expansion of the parameter space for optimal threshold searching. This can be mitigated by a dual-step searching strategy: first, search coarsely to narrow down the possible parameter space, and then pinpoint the optimum thresholds within the new parameter space. Owing to space limitations in this paper, a more thorough investigation of the three-level threshold space is not reported here, but remains for future work.

5.2. Results of threshold-based identification

The thresholds for maximizing s are listed in Table 4. The results show that the K -means clustering method is mostly consistent with the second set of rules for both binary and three-level identifications (Figs. 6, 7, 8 and 9). For example, a census tract with a confluence of housing renovation and at least one aspect of social upgrading is more likely to be identified as the gentrified. The second set of gentrification rules suggests that gentrification is a socio-economic processes

Table 4
Outcome of threshold searching.

Rule Number	Maximum s	Thresholds Maximizing s[a]
Binary		
1	0.78	a = −0.17 (1), b = 2681 (9), c = −0.09 (1), d = 0.08 (9), e = 0.12 (9), f = −0.06 (4), g = 0.07 (9)
2	0.80	a = −0.12 (2), b = −2631 (1), c = 0.11 (8), d = −0.09 (1), e = −0.09 (1), f = −0.00 (6), g = −0.08 (1)
3	0.78	a = 0.03 (9), b = −2631 (1), c = 0.16 (9), d = −0.09 (1), e = −0.09 (1), f = −0.00 (6), g = −0.08 (1)
Three-Level		
1	0.44	(h, i) = (−0.17, −0.09) (1, 3), (j, k) = (794,2681) (7, 9), (l, m) = (−0.09, −0.01) (1, 3), (n, o) = (0.02,0.08) (7, 9), (p, q) = (0.05,0.12) (7, 9), (r, s) = (−0.18, −0.09) (1, 3), (t, u) = (0.03,0.07) (7, 9)
2	0.56	(h, i) = (−0.09,0.03) (3, 9), (j, k) = (794,2681) (7, 9), (l, m) = (−0.09, −0.01) (1, 3), (n, o) = (0.00,0.02) (5, 7), (p, q) = (0.05,0.12) (7, 9), (r, s) = (−0.09,0.11) (3, 9), (t, u) = (0.00,0.03) (5, 7)
3	0.33	(h, i) = (−0.17, −0.09) (1, 3), (j, k) = (−2631, −805) (1, 3), (l, m) = (−0.09, −0.01) (1, 3), (n, o) = (−0.09, −0.03) (1, 3), (p, q) = (−0.09, −0.03) (1, 3), (r, s) = (−0.18, −0.09) (1, 3), (t, u) = (−0.08, −0.03) (1, 3)

[a] Integers in parentheses are thresholds' corresponding deciles.

triggered by simultaneous housing reinvestment and social upgrading. However, the latter could be unfolded in diverse ways: Point Chevalier (Liu et al., 2018) and Talbot Park (Ministry for the Environment, 2008) witness an influx of well-off groups; the studentification of Hobson Street (Collins, 2010) focuses on tenure change; the gentrification of Auckland Harbourside (Murphy, 2008) revolves around occupational and/or income changes; and the gentrification of Glen Innes (Gordon et al., 2017) leads to ethnic and/or income variation. It is the diversified social upgrading that generates the variegated gentrification. As relevant to one or more aspects of social upgrading, these examples correspond to the second set of rules, further vindicating the K -means clustering method.

The maximum s between the second set of rules and the K -means clustering declines when the level of gentrification identification increases (0.80 and 0.56 for the binary and the three-level modes, respectively), signifying amplified discrepancies between both approaches. For the three-level identification, the second set of rules identifies a pattern of gentrification more compact than that of the K -means clustering: the heavily- and moderately-gentrified areas are closer to each other, and are separate from the least-gentrified areas. Meanwhile, the second set of rules labels most of Point Chevalier and Glen Innes as moderately and heavily gentrified, respectively, consistent with our observation that the gentrification of the latter is more confrontational than that of the former (refer to the definitions of gentrification level in Section 4.1). Both the level of gentrification and the various aspects of social upgrading contribute to the mutation of gentrification from previous research (Lees et al., 2008). Additionally, the threshold of *IVPD* for binary identification is -0.12, matching up with the optimal threshold (-0.125) found in the research of Liu et al. (2018). Thus, the second set of rules seems to delineate gentrification in Auckland more accurately than K -means clustering. This finding suggests that we should harness theories of statistics and the social theory in tandem, as the former reduces the arbitrariness in identification, and the latter gives us *a priori* knowledge of socio-economic settings. A better identification outcome can be achieved using a

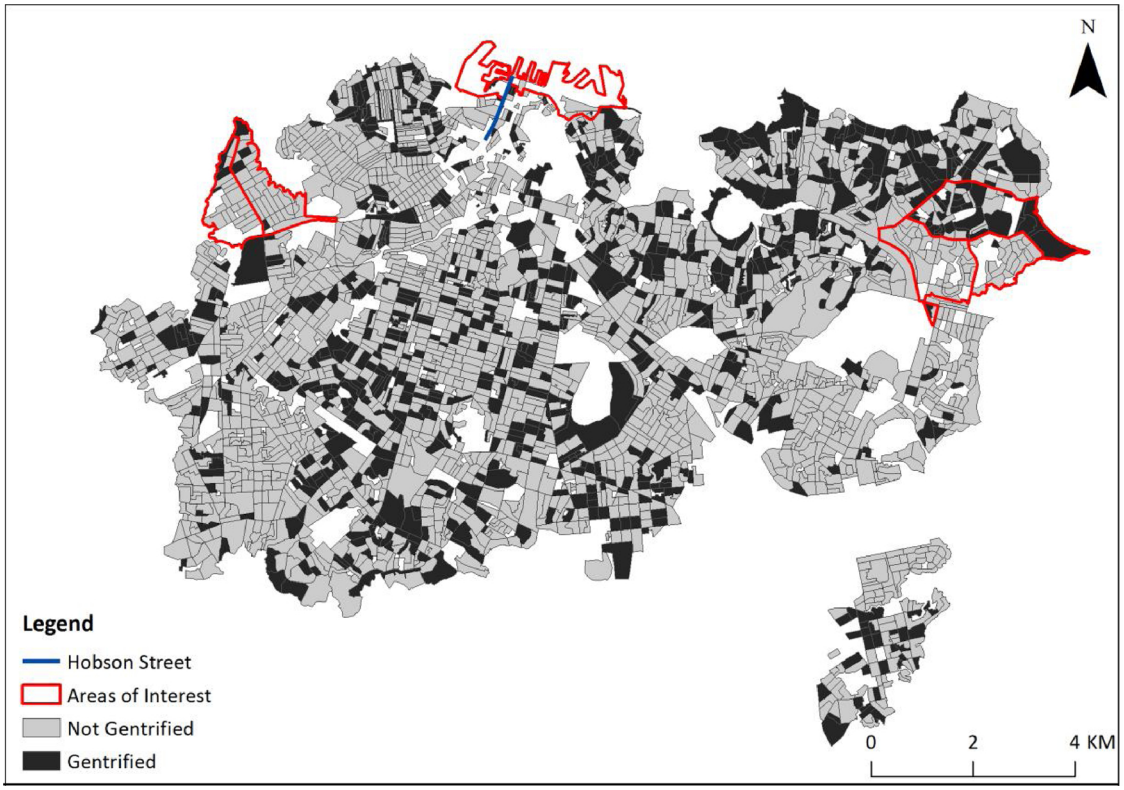

Fig. 6. Binary gentrification identification via the second set of rules (Tables 3 and 4) in Phase I.

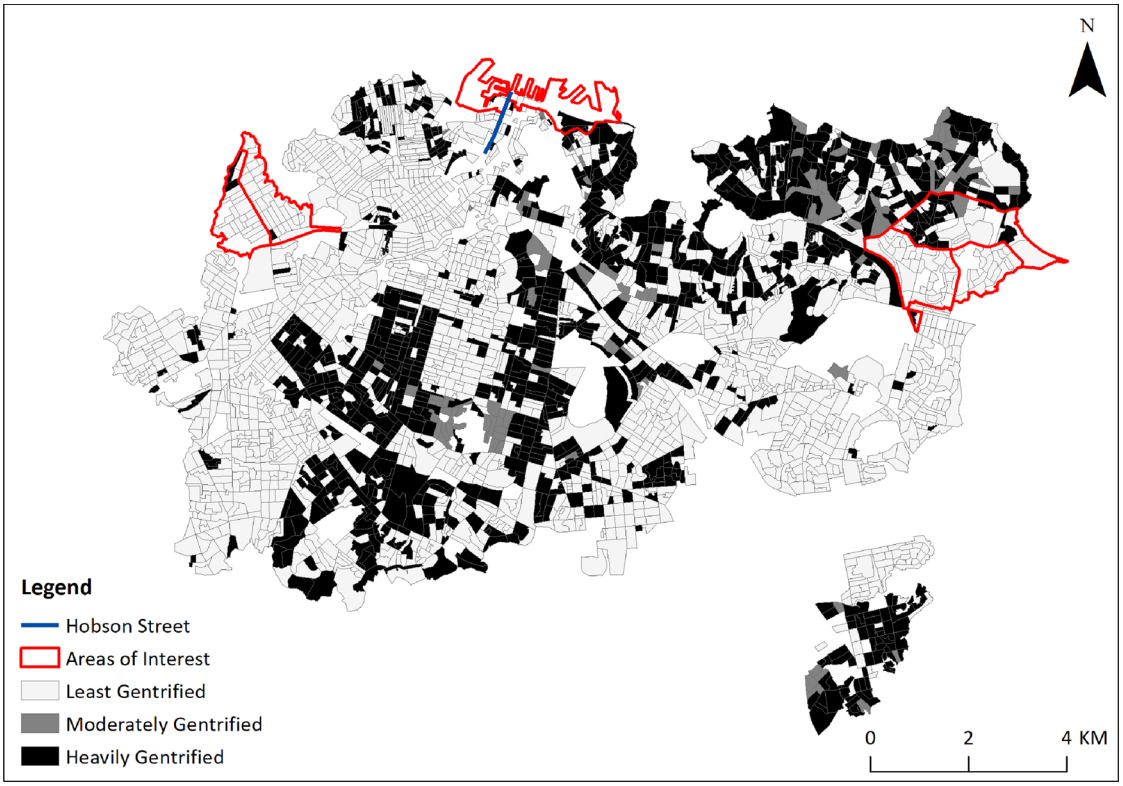

Fig. 7. Three-level gentrification identification via the second set of rules in Phase I.

Fig. 8. Binary gentrification identification via the second set of rules in Phase II.

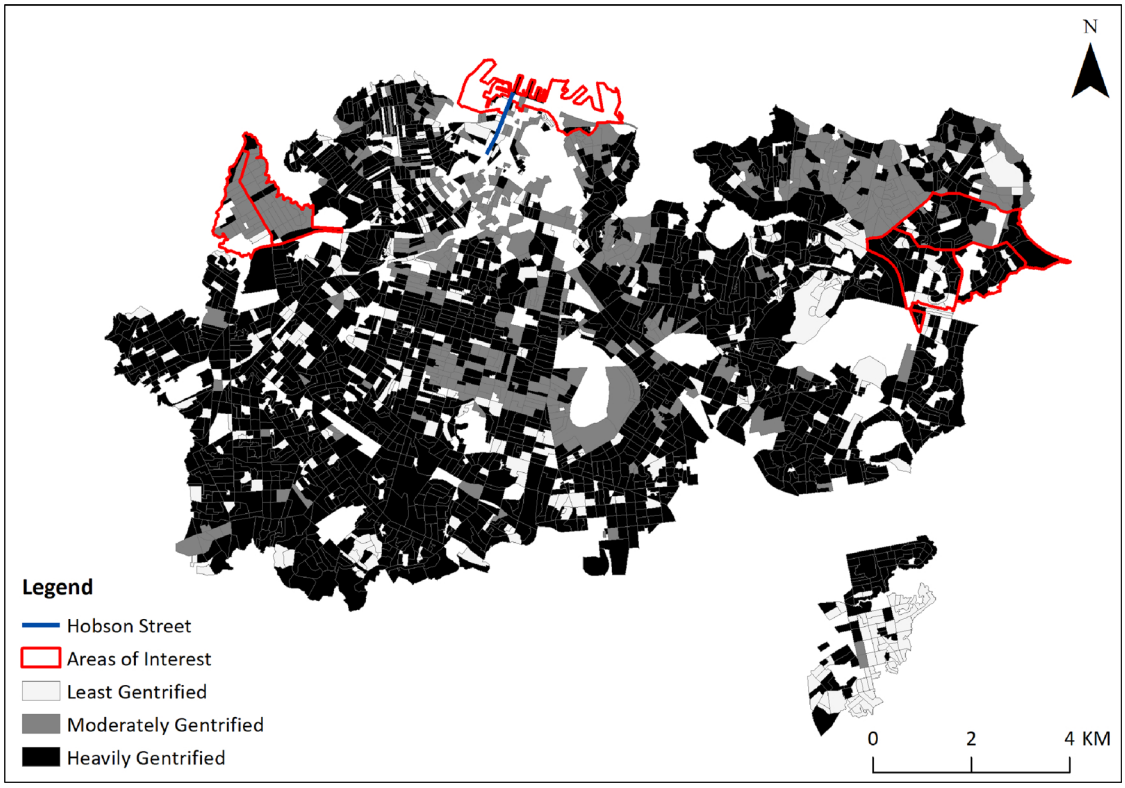

Fig. 9. Three-level gentrification identification via the second set of rules in Phase II.

conjunction of both theories.

6. Retaining the concept of "gentrification"

Actually, all thresholds in the second set of rules reasonably concentrate on both ends of the deciles (Table 4), confirming that gentrification has the potential to profoundly change neighborhoods both esthetically and socially, such as via housing reinvestment and displacement. However, the extent of the manifestation of gentrification can vary dramatically, owing to the degree of social and economic polarization, the nature and traditions of urban design, etc. (Butler, 2007). As an idle industrial land prior to gentrification, Viaduct Harbour was not gentrified until 2000, when the America's Cup was hosted in Auckland. Aiming to provide immediate and long-term benefits to the people of Auckland, the redevelopment of Viaduct Harbour pulled off a major coup (Murphy, 2008). Viaduct Harbour became a high-end commercial, residential, and recreational site, was a shining example of urban design, and was well-received by the public. Nevertheless, possible side effects of this project (such as the upsurge of rent and its spillover effects) are barely discussed. Another example is Talbot Park on the south of Glen Innes, which was renewed by governmental organizations in or near 2007. The government depicted the project as an excellent model of addressing social exclusion and fostering strong sustainable communities (Ministry for the Environment, 2008). Ironically, the tenant turnover per annum at the outset of the project (Ministry for the Environment, 2008), at approximately 50 percent, shows the impacts of increased rent on low-income tenants.

Gentrification in Glen Innes recounts a radically-different story. To provide more housing, Housing New Zealand decided to evict low-income tenants, and to redevelop this area by demolishing or selling state housing with profits. This project provoked massive and even violent demonstrations (Strongman, 2017). A documentary titled "A Place to Call Home" (2015) recorded the ordeals of evicted tenants, profit-seeking developers, and even a Māori trust, who purchased some of the state houses to help poor Māori families in a remote rural place. The case of Glen Innes reveals the controversy of gentrification: not only displacement of the underprivileged, but also the concept of "taking housing of the poor to help the poorer".

As compared to the gentrification in Viaduct Harbour and Glen Innes, which received public attention (either wanted or unwanted), the gentrification in Point Chevalier was relatively low-profile, and ostensibly a more peaceful process. Prior to gentrification, Point Chevalier was an affordable place for young families and students, who were susceptible to the rent increases incurred from gentrification. However, it was possible to displace the tenants without sparking public outcry afterwards (Liu et al., 2018). The studentification of Hobson Street down to Viaduct Harbour (Collins, 2010) bears a resemblance to that of Point Chevalier. A massive influx of international students (particularly Asian students) to the central business district (CBD) of Auckland stimulated the demand for apartments. Spotting the niche in the rental market, property developers invested in several large apartment complexes near Hobson Street, transforming the region from an outlying area of the CBD to a densely-populated zone. Reports on forced uprooting in the process were scarce.

Based on the identification results in this study, we contend that the cases discussed herein should be subsumed under the rubric of gentrification. The continued use of "gentrification" furnishes a common conceptual and analytical framework for the ensuing study of the "geography of gentrification". Moreover, a broad definition of gentrification accommodates changes in the future. Based on an "elastic yet targeted" definition of gentrification, our identification strategy successfully detects the occurrence and intensity of different types of gentrification in Auckland (Fig. 9), despite their bewildering variety of appearances. Gentrification in Point Chevalier, Hobson Street, and Viaduct Harbour is labeled as moderate owing to relatively mild social and economic change, whereas Glen Innes and Talbot Park are labeled as heavily-gentrified, owing to the large-scale displacement of low-income ethnic minorities. In addition to classic gentrification, studentification, and new-build gentrification in different localities of the city of Auckland, our method is potentially suitable for identifying any

gentrification with social and economic changes, *e.g.*, super gentrification, rural gentrification, and tourism gentrification, regardless of peculiarities (*e.g.*, site). This is because Clark's definition (2005) strips various manifestations of gentrification down to two defining characteristics: housing reinvestment, and direct/indirect social upgrading/displacement of the poor. One possible exception to the identification approach herein is the gentrification esthetic (Ley, 2003). A representative example is "loft living" in the warehouses of a former industrial zone. Zukin (1989) expounds on how investors use cultural industry as a tool to initiate gentrification of dilapidated manufacturing spaces in New York, and to extract profits. On the bottom rung of the career ladder and with fledgling artistic businesses, the majority of the avant-garde are also economically underprivileged. The replacement of the original inhabitants with the avant-garde is challenging to identify from a socio-economic perspective. Nonetheless, the gentrification esthetic is not an "end-stage, static phenomenon" (Lees et al., 2008, p. 117) as low-income pioneer gentrifiers (*e.g.*, artists, students, and "hippies") are eventually coerced out by the associated economic valorization of land, despite their endeavors to attract a subsequent influx of investment capital and affluent middle-class professionals. Moreover, the following displacement of the disadvantaged pioneer gentrifiers can be detected by our approach.

More importantly, gentrification is a complex and multifarious process. Rather than being distracted by a welter of the particularities of gentrification, we should focus on polarization and class struggles (Slater, 2006). There are two implications for the urban practice and policy of gentrification. First, based on a simple and universal framework of gentrification, our identification method helps to map the presence of gentrification in space via GIS, possibly educating and empowering the marginalized when fighting gentrification, as reported in the gentrification of SoMa in San Francisco (Parker & Pascual, 2002) (Section 2.2). Second, the displacement of gentrification is multi-dimensional (*e.g.*, occupation, income, tenure, and ethnicity). The second set of rules (Table 3) further confirms that a significant change against the disadvantaged in any aspect(s) can potentially cause social inequality. This research provides a springboard for convincing policy-makers to protect those most threatened by gentrification (Freeman, 2008) as The Real Estate Board of New York (1985) conceded in an advertorial in The New York Times: "We believe that whatever displacement gentrification causes, though, must be dealt with with [*sic*] public policies that promote low- and moderate-income housing construction and rehabilitation, and in zoning revisions that permit retail uses in less expensive, side street locations" (Quoted from Smith, 1996, p.29).

7. Discussion and conclusions

Gentrification can be identified via a threshold-based method (Van Criekingen & Decroly, 2003) and/or a machine-learning approach (Reades et al., 2018), such as K -means clustering (Ye et al., 2015). From the perspective of gentrification theory, the former identifies gentrification via changes of socio-economic factors, which are simple, convenient, and theoretically sound. However, the threshold-based method is criticized for the arbitrariness of the thresholds (Freeman, 2009). As a possible remedy, the classification approach employs the theory of inferential statistics to identify gentrification based on a data pattern. Meanwhile, the threshold-based method is likely to inform the selection of the algorithm and input variables, as well as the explicit interpretation of the identification outcome of the statistical approach. Although a combination of both approaches is beneficial to unbiased gentrification identification, there is little research on exploiting the strengths of both.

Using Auckland as the study area, this study explored two methods of gentrification identification: K -means clustering, and a threshold-based approach. The former labels census tracts according to a data pattern, avoiding the arbitrariness of the threshold-based method. However, the K -means clustering is less intuitive, and is challenging to explicitly interpret in terms of gentrification variables. Thus, the identification results of the threshold-based method are compared to those of K -means clustering. The thresholds and sets of identification rules with optimal matches are selected, serving to interpret the outcome of K -means clustering. The research proffers a potential solution in the furtherance of the objectivity and accuracy of ensuing gentrification identification.

We find that both methods accord with each other, to a certain extent. The maximum degrees of similarity (falling in the range 0–1) between both are 0.80 and 0.56 for the dualistic and the three-level gentrification identification, respectively. Moreover, the K -means clustering method is mostly consistent with a set of threshold-based rules: a census tract with a confluence of housing renovation and at least one aspect of social upgrading is more likely to be identified as gentrified. The identification strategy of K -means clustering revealed by the threshold-based rules agrees with Clark's definition: housing reinvestment and displacement/social upgrading are indispensable components of gentrification. In addition, the social upgrading of gentrification in the optimal set of rules barely covers the full spectrum of this multifaceted process, conforming to our observation: some areas witness an influx of well-off groups; some focus on a tenure change; some revolve around occupational and/or income changes; and others on ethnic and/or income variation. All these are at the expense of the underprivileged, including low-income families, students, blue-collar workers, and racial minorities.

After evaluating both methods against five cases of gentrification in Auckland, we discover that the K -means clustering method identifies gentrification relatively satisfactorily. By comparison, the corresponding threshold-based set of identification rules delineates gentrification more precisely in terms of a compact gentrification pattern and the accuracy of gentrification levels, possibly owing to the synergy between the clear and simple rules of identification and the K -means clustering method. In addition, the thresholds of the gentrification variables reasonably concentrate on both ends of the deciles, agreeing with previous research. The comparison between both approaches seems to indicate that although new technologies are instrumental in gentrification identification, they are supposed to be applied with the caveat that the socio-economic meaning of a method and its variables should be properly understood. The socio-economic context can not only inform the selection of algorithm and input variables, but also improves the performance of the mathematical models.

The gentrification identified in Auckland in this research exhibits common characteristics with instances previously studied elsewhere: first, owing to peculiarities in the local social milieu, gentrification assumes various appearances in terms of levels and types of gentrification, and leads to different consequences. For instance, the slum clearance in east Auckland (Gordon et al., 2017) incurs more confrontational gentrification than that of others, such as the studentification (Collins, 2010) and new-build gentrification (Murphy, 2008) in the Auckland city center. Second, the "patchwork" quality of gentrification reported in Global West (*e.g.*, Houston) (Podagrosi et al., 2011) and Global East (*e.g.* Hong Kong) (Ye et al., 2015) can be identified, particularly via the threshold-based approach. The close affinity between gentrification in Auckland and previously-studied cases echoes the perspective of retaining the simple and universal conceptual and analytical framework of gentrification, which helps us focus on the essentials of the urban phenomenon: reinvestment and displacement. There are several evident advantages of a coherent framework of gentrification. First, gentrification identification becomes more consistent

and convenient when based on the two core components (*i.e.*, re-investment and displacement). Therefore, the combination of both *K*-means clustering and thresholding has wide applicability to the identification of various types of gentrification, which is conducive to the study of comparative gentrification (Lees, 2012) at the global scale, and has the potential to open up new insights. Second, the framework of gentrification provokes a sobering thought regarding its detrimental effects on the underprivileged, regardless of its sundry manifestations, and takes "a knife to the soft underbelly" of the phenomenon via critical research, persuading policymakers to address the issue. A case in point is critical GIS (O'Sullivan, 2006) (discussed in Section 2.2), *i.e.*, mapping the presence of gentrification in space via GIS. A coherent framework of gentrification facilitates identification of gentrification, and educates and empowers the marginalized when fighting gentrification (Parker & Pascual, 2002).

There might be several directions for future study. First, both identification approaches employed in this study are relatively simple, and more advanced statistical algorithms should be investigated to improve the precision and objectivity of gentrification identification. Second, despite the fact that unsupervised classification algorithms can alleviate the arbitrariness of threshold-based methods by using a data pattern, they barely use *a priori* knowledge (*e.g.*, socio-economic context), which is likely to enhance the accuracy of classification. Moreover, the meaning of the consequent categories is unclear, and depends on human interpretation. Possibly overcoming these drawbacks, supervised classification algorithms are becoming more applicable as samples of gentrification identification accumulate. Third, gentrification from other cities worldwide is needed to evaluate different identification approaches, *i.e.*, estimating the error of identification across various social environments and providing a more solid basis for the comparative planetary gentrification.

8. Declaration of Competing Interest

None.

Acknowledgements

This work was supported by the National Social Science Fund of China (grant number 14BJY057), the National Natural Science Foundation of China (grant numbers 41801166, 41771184, 41671155, 41271176, 41701629, 41871172 and 41601164), the Programme of Excellent Young Scientists of the Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences (grant number 2017RC202), the Cradle Project of China University of Geosciences (Wuhan) (grant number CUGL170408), the Fundamental Research Funds for the Central Universities, China University of Geosciences (Wuhan) (grant number CUGW170813) and Youth Innovation Promotion Association CAS (grant number 2019055). We are grateful to the editor and anonymous reviewers for their constructive comments.

References

Álvarez-Rivadulla, M. J. (2007). Golden ghettos: Gated communities and class residential segregation in Montevideo, Uruguay. *Environment and Planning A: Economy and Space, 39*(1), 47–63.

Anguelovski, I., Connolly, J. J. T., Masip, L., & Pearsall, H. (2018). Assessing green gentrification in historically disenfranchised neighborhoods: a longitudinal and spatial analysis of Barcelona. *Urban Geography, 39*(3), 458–491.

Atkinson, R. (2000). Measuring gentrification and displacement in greater London. *Urban Studies, 37*(1), 149–165.

Atkinson, R., & Bridge, G. (2005). *Gentrification in a global context: The new urban colonialism.* London: Routledge.

Barton, M. (2016). An exploration of the importance of the strategy used to identify gentrification. *Urban Studies, 53*(1), 92–111.

Beauregard, R. A. (1986). The chaos and complexity of gentrification. In N. Smith, & P. Williams (Eds.). *Gentrification of the City* (pp. 35–55). Boston: Allen & Unwin.

Boddy, M. (2007). Designer neighbourhoods: New-build residential development in nonmetropolitan UK Cities—The case of Bristol. *Environment and Planning A: Economy and Space, 39*(1), 86–105.

Bondi, L. (1999). Between the woof and the weft: A response to Loretta Lees. *Environment and Planning D: Society and Space, 17*(3), 253–255.

Bostic, R. W., & Martin, R. W. (2003). Black home-owners as a gentrifying force? Neighbourhood dynamics in the context of minority home-ownership. *Urban Studies, 40*(12), 2427–2449.

Bourne, L. S. (1993). The demise of gentrification? A commentary and prospective view. *Urban Geography, 14*(1), 95–107.

Brown-Saracino, J. (2017). Explicating divided approaches to gentrification and growing income inequality. *Annual Review of Sociology, 43*(1), 515–539.

Brown-Saracino, J. (Ed.). (2010). *The Gentrification Debates: A reader*. New York: Routledge.

Butler, T. (2007). For gentrification? *Environment and Planning A: Economy and Space, 39*(1), 162–181.

Cartier, C. (2017). Contextual urban theory and the 'appeal' of gentrification: Lost in transposition? *International Journal of Urban and Regional Research, 41*(3), 466–477.

Charles, S. L. (2014). The spatio-temporal pattern of housing redevelopment in suburban Chicago, 2000–2010. *Urban Studies, 51*(12), 2646–2664.

Choi, M., Van Zandt, S., & Matarrita-Cascante, D. (2018). Can community land trusts slow gentrification? *Journal of Urban Affairs, 40*(3), 394–411.

Clark, E. (2005). The order and simplicity of gentrification: A political challenge. In R. Atkinson, & G. Bridge (Eds.). *Gentrification in a global context: The new urban colonialism* (pp. 24–29). London: Routledge.

Clark, E. (2015). Afterword: The adventure of generic gentrification. In L. Lees, S. Hyun Bang, & L.-M. Ernesto (Eds.). *Global gentrifications: Uneven development and displacement* (pp. 453–456). Bristol: Policy Press.

Collins, F. L. (2010). International students as urban agents: International education and urban transformation in Auckland, New Zealand. *Geoforum, 41*(6), 940–950.

Davidson, M., & Lees, L. (2005). New-build' gentrification' and London's riverside renaissance. *Environment and Planning A, 37*(7), 1165–1190.

Delmelle, E. C. (2016). Mapping the DNA of urban neighborhoods: Clustering longitudinal sequences of neighborhood socioeconomic change. *Annals of the American Association of Geographers, 106*(1), 36–56.

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

Figueroa, R. A. (1995). A housing-based delineation of gentrification: a small area analysis of regina. *Canada. Geoforum, 26*(2), 225–236.

Forrest, R. (2015). Commentary: Variegated gentrification? *Urban Studies, 53*(3), 609–614.

Freeman, L. (2005). Displacement or succession?: Residential mobility in gentrifying neighborhoods. *Urban Affairs Review, 40*(4), 463–491.

Freeman, L. (2008). Comment on 'The eviction of critical perspectives from gentrification research. *International Journal of Urban and Regional Research, 32*(1), 186–191.

Freeman, L. (2009). Neighbourhood diversity, metropolitan segregation and gentrification: What are the links in the US? *Urban Studies, 46*(10), 2079–2101.

Galster, G. C., & Peacock, S. (1986). Urban gentrification: Evaluating alternative indicators. *Social Indicators Research, 18*(3), 321–337.

González, S. (2016). Looking comparatively at displacement and resistance to gentrification in Latin American cities. *Urban Geography, 37*(8), 1245–1252.

Gordon, R., Collins, F. L., & Kearns, R. (2017). It is the people that have made Glen Innes': State-led gentrification and the reconfiguration of urban life in Auckland. *International Journal of Urban and Regional Research, 41*(5), 767–785.

Gotham, K. F. (2005). Tourism gentrification: The case of New Orleans' Vieux Carre (French Quarter). *Urban Studies, 42*(7), 1099–1121.

Gould Ellen, I., & O'Regan, K. (2008). Reversal of fortunes? Lower-income urban neighbourhoods in the US in the 1990s. *Urban Studies, 45*(4), 845–869.

Hackworth, J. (2002). Postrecession gentrification in New York City. *Urban Affairs Review, 37*(6), 815–843.

Hammel, D. J., & Wyly, E. (1996). A model for identifying gentrified areas with census data. *Urban Geography, 17*(3), 248–268.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Second ed.). Springer.

He, S. (2007). State-sponsored gentrification under market transition the case of Shanghai. *Urban Affairs Review, 43*(2), 171–198.

Hedin, K., Clark, E., Lundholm, E., & Malmberg, G. (2012). Neoliberalization of Housing in Sweden: Gentrification, Filtering, and Social Polarization. *Annals of the Association of American Geographers, 102*(2), 443–463.

Hee-Jung, J. (2013). Determinants of neighborhood change: A multilevel analysis. *Urban Affairs Review, 49*(3), 319–352.

Heidkamp, C. P., & Lucas, S. (2006). Finding the gentrification frontier using census data: The case of Portland, Maine. *Urban Geography, 27*(2), 101–125.

Holm, A., & Schulz, G. (2018). GentriMap: A model for measuring gentrification and displacement. In I. Helbrecht (Ed.). *Gentrification and resistance: Researching displacement processes and adaption strategies* (pp. 251–277). Wiesbaden: Springer Fachmedien Wiesbaden.

Hwang, J., & Lin, J. (2016). What Have We Learned About the Causes of Recent Gentrification? *Cityscape, 18*(3), 9–26.

Hwang, J., & Sampson, R. J. (2014). Divergent pathways of gentrification: Racial inequality and the social order of renewal in Chicago neighborhoods. *American Sociological Review, 79*(4), 726–751.

Hyun Bang, S., Lees, L., & Ernesto, L.-M. (2015). Introduction: Locating gentrification in the Global East. *Urban Studies, 53*(3), 455–470.

Ilic, L., Sawada, M., & Zarzelli, A. (2019). Deep mapping gentrification in a large Canadian city using deep learning and Google Street View. *Plos One, 14*(3), e0212814.

Lagendijk, A., van Melik, R., de Haan, F., Ernste, H., Ploegmakers, H., & Kayasu, S. (2014). Comparative approaches to gentrification: A research framework. *Tijdschrift voor economische en sociale geografie, 105*(3), 358–365.

Lambert, C., & Boddy, M. (2002). *Transforming the city : post-recession gentrification and re-urbanisation.* Paper presented at the Upward Neighbourhood Trajectories: Gentrification in the New CenturyScotland: University of Glasgow.

Landis, J. D. (2016). Tracking and explaining neighborhood socioeconomic change in U.S. metropolitan areas between 1990 and 2010. *Housing Policy Debate, 26*(1), 2–52.

Lees, L. (2000).). A reappraisal of gentrification: Towards a 'geography of gentrification'. *Progress in Human Geography, 24*(3), 389–408.

Lees, L. (2012). The geography of gentrification: Thinking through comparative urbanism. *Progress in Human Geography, 36*(2), 155–171.

Lees, L. (2018). Comparative urbanism in gentrification studies: fashion or progress? In L. Lees, & M. Phillips (Eds.). *Handbook of Gentrification Studies*. Edward Elgar.

Lees, L., Shin, H. B., & López-Morales, E. (2016). *Planetary gentrification.* Cambridge: Polity Press.

Lees, L., Slater, T., & Wyly, E. (2008). *Gentrification.* New York: Routledge.

Ley, D. (1986). Alternative explanations for inner-city gentrification: A Canadian assessment. *Annals of the Association of American Geographers, 76*(4), 521–535.

Ley, D. (1992). Gentrification in recession: Social change in six Canadian inner cities, 1981-1986. *Urban Geography, 13*(3), 230–256.

Ley, D. (1996). *The new middle class and the remaking of the central city.* Oxford: Oxford University Press.

Ley, D. (2003). Artists, aestheticisation and the field of gentrification. *Urban Studies, 40*(12), 2527–2544.

Ley, D., & Teo, S. Y. (2014). Gentrification in Hong Kong? Epistemology vs. Ontology. *International Journal of Urban and Regional Research, 38*(4), 1286–1303.

Lipton, S. G. (1977). Evidence of central city revival. *Journal of the American Institute of Planners, 43*(2), 136–147.

Liu, C., O'Sullivan, D., & Perry, G. L. W. (2018). The rent gap revisited: Gentrification in Point Chevalier, Auckland. *Urban Geography, 39*(9), 1300–1325.

Liu, L., Silva, E. A., Wu, C., & Wang, H. (2017). A machine learning-based method for the large-scale evaluation of the qualities of the urban environment. *Computers, Environment and Urban Systems, 65*(Supplement C), 113–125.

London, B., Lee, B. A., & Lipton, S. G. (1986). The determinants of gentrification in the United States: A city-level analysis. *Urban Affairs Quarterly, 21*(3), 369–387.

López-Morales, E. (2015). Gentrification in the Global South. *City, 19*(4), 564–573.

Loukaitou-Sideris, A., Gonzalez, S., & Ong, P. (2017). Triangulating neighborhood knowledge to understand neighborhood change: Methods to study gentrification. *Journal of Planning Education and Research, 39*(2), 227–242.

Lui, T.-L. (2017). Beneath the appearance of gentrification: Probing local complexities. *International Journal of Urban and Regional Research, 41*(3), 478–486.

Maloutas, T. (2011). Contextual diversity in gentrification research. *Critical Sociology, 38*(1), 33–48.

Maloutas, T. (2017). Travelling concepts and universal particularisms: A reappraisal of gentrification's global reach. *European Urban and Regional Studies, 25*(3), 250–265.

March, B. (2015). *A Place to Call Home.* Retrieved fromhttps://www.nzonscreen.com/title/a-place-to-call-home-2015.

Meligrana, J., & Skaburskis, A. (2005). Extent, location and profiles of continuing gentrification in Canadian Metropolitan Areas, 1981-2001. *Urban Studies, 42*(9), 1569–1592.

Meltzer, R., & Ghorbani, P. (2017). Does gentrification increase employment opportunities in low-income neighborhoods? *Regional Science and Urban Economics, 66*, 52–73.

Ministry for the Environment (2008). *Community renewal – Housing New Zealand Corporation, Talbot Park, Auckland.* Retrieved fromhttp://www.mfe.govt.nz/publications/towns-and-cities/urban-design-case-studies-local-government/community-renewal-housing.

Mosselson, A. (2016). Joburg has its own momentum': Towards a vernacular theorisation of urban change. *Urban Studies, 54*(5), 1280–1296.

Murphy, L. (2008). Third-wave gentrification in New Zealand: The case of Auckland. *Urban Studies, 45*(12), 2521–2540.

Naik, N., Kominers, S. D., Raskar, R., Glaeser, E. L., & Hidalgo, C. A. (2017). Computer vision uncovers predictors of physical urban change. *Proceedings of the National Academy of Sciences*.

Nijman, J. (2007). Introduction—Comparative urbanism. *Urban Geography, 28*(1), 1–6.

O'Sullivan, D., Bergmann, L., & Thatcher, J. E. (2018). Spatiality, maps, and mathematics in critical human geography: Toward a repetition with difference. *The Professional Geographer, 70*(1), 129–139.

O'Sullivan, D., & Unwin, D. (2010). *Geographic information analysis* (Second ed.). New Jersey: John Wiley & Sons, Inc.

O'Sullivan, D. (2006). Geographical information science: Critical GIS. *Progress in Human Geography, 30*(6), 783–791.

Orsman, B. (2009). *Resanding Auckland's beaches. The New Zealand Herald.* Retrieved fromhttp://www.nzherald.co.nz/nz/news/article.cfm?c_id=1&objectid=10551753.

Owens, A. (2012). Neighborhoods on the rise: A typology of neighborhoods experiencing socioeconomic ascent. *City & Community, 11*(4), 345–369.

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 & Community, 10*(3), 215–240.

Parker, C., & Pascual, A. (2002). A voice that could not be ignored: Community GIS and gentrification battles in San Francisco. In W. J. Craig, T. M. Harris, & D. Weiner (Eds.). *Community participation and geographic information systems* (pp. 55–64). London and New York: Taylor & Francis.

Paul, W. (2015). Speaking gentrification in the languages of the Global East. *Urban Studies, 53*(3), 615–625.

Pavlovskaya, M. (2009). Feminist visualization. In R. Kitchin, & N. Thrift (Eds.). *International encylopedia of human geography* (pp. 157–164). Amsterdam: Elsevier.

Phillips, M. (1993). Rural gentrification and the processes of class colonisation. *Journal of Rural Studies, 9*(2), 123–140.

Phillips, M., & Smith, D. P. (2018). Comparative approaches to gentrification: Lessons from the rural. *Dialogues in Human Geography, 8*(1), 3–25.

Podagrosi, A., Vojnovic, I., & Pigozzi, B. (2011). The diversity of gentrification in Houston's urban renaissance: From cleansing the urban poor to supergentrification. *Environment and Planning A: Economy and Space, 43*(8), 1910–1929.

Reades, J., De Souza, J., & Hubbard, P. (2018). Understanding urban gentrification through machine learning. *Urban Studies, 56*(5), 922–942.

Robinson, J. (2015). Thinking cities through elsewhere: Comparative tactics for a more global urban studies. *Progress in Human Geography, 40*(1), 3–29.

Rose, D. (1984). Rethinking gentrification: Beyond the uneven development of Marxist urban theory. *Environment and Planning D: Society and Space, 2*(1), 47–74.

Sassen, S. (1991). *The global city.* New York, London, Tokyo: Princeton University Press.

Schmid, C., Karaman, O., Hanakata, N. C., Kallenberger, P., Kockelkorn, A., Sawyer, L., ... Wong, K. P. (2017). Towards a new vocabulary of urbanisation processes: A comparative approach. *Urban Studies, 55*(1), 19–52.

Schuler, J. R., Kent, R. B., & Monroe, C. B. (1992). Neighborhood gentrification: A discriminant analysis of a historic district in Cleveland, Ohio. *Urban Geography, 13*(1), 49–67.

Slater, T. (2006). The eviction of critical perspectives from gentrification research. *International Journal of Urban and Regional Research, 30*(4), 737–757.

Smart, A., & Smart, J. (2017). Ain't talkin' 'bout gentrification: The erasure of alternative idioms of displacement resulting from Anglo-American academic hegemony. *International Journal of Urban and Regional Research, 41*(3), 518–525.

Smith, D. (2005). 'Studentification': the gentrification factory? In R. Atkinson, & G. Bridge (Eds.). *Gentrification in a global context: the new urban colonialism*. Routledge.

Smith, N. (1982). Gentrification and uneven development. *Economic Geography, 58*(2), 139–155.

Smith, N. (1996). *The new urban frontier: Gentrification and the revanchist city*. London and New York: Routledge.

Smith, N. (1998). Giuliani Time: The revanchist 1990s. *Social Text,* (57), 1–20.

Smith, N. (2002). New globalism, new urbanism: Gentrification as global urban strategy. *Antipode, 34*(3), 427–450.

Smith, N., & DeFilippis, J. (1999). The reassertion of economics: 1990s gentrification in the lower east side. *International Journal of Urban and Regional Research, 23*(4), 638–653.

Statistics New Zealand (2019). *Census usually resident population count.* Retrieved fromhttp://datainfoplus.stats.govt.nz/item/nz.govt.stats/05ae5a6c-05c1-4834-a28f-4837150984d6/2/#.

Strongman, S. (2017). *Niki's eviction: appeal dismissed.* Retrieved fromhttp://thewireless.co.nz/articles/niki-s-eviction-appeal-lost.

Tang, W.-S. (2017). Beyond gentrification: Hegemonic redevelopment in Hong Kong. *International Journal of Urban and Regional Research, 41*(3), 487–499.

The Real Estate Board of New York (1985). Is gentrification a dirty word. *The New York Times*.

Tilly, C. (1984). *Big structures, large processes, huge comparisons.* New York: Russell Sage Found.

Timberlake, J. M., & Johns-Wolfe, E. (2016). Neighborhood ethnoracial composition and gentrification in Chicago and New York, 1980 to 2010. *Urban Affairs Review, 53*(2), 236–272.

Van Criekingen, M., & Decroly, J.-M. (2003). Revisiting the diversity of gentrification: Neighbourhood renewal processes in Brussels and Montreal. *Urban Studies, 40*(12), 2451–2468.

Walks, R. A., & Maaranen, R. (2008a). Gentrification, social mix, and social polarization: Testing the linkages in large Canadian cities. *Urban Geography, 29*(4), 293–326.

Walks, R. A., & Maaranen, R. (2008b). *The timing, patterning, and forms of community gentrification and neighborhood change in Montreal, Toronto, and Vancouver, 1961 to 2001.* Research PaperToronto, Ontario, Canada: Centre for Urban and Community Studies.

Whitehead, J. (2008). Rent gaps, revanchism and regimes of accumulation in Mumbai. *Anthropologica, 50*(2), 269–282.

Wu, Q., Edensor, T., & Cheng, J. (2018). Beyond space: Spatial (re)production and middle-class remaking driven by Jiaoyufication in Nanjing City, China. *International Journal of Urban and Regional Research, 42*(1), 1–19.

Wyly, E. K., & Hammel, D. J. (1998). Modeling the context and contingency of gentrification. *Journal of Urban Affairs, 20*(3), 303–326.

Ye, M., Vojnovic, I., & Chen, G. (2015). The landscape of gentrification: exploring the diversity of "upgrading" processes in Hong Kong, 1986–2006. *Urban Geography, 36*(4), 471–503.

Yip, N. M., & Tran, H. A. (2015). Is 'gentrification' an analytically useful concept for Vietnam? A case study of Hanoi. *Urban Studies, 53*(3), 490–505.

Zukin, S. (1989). *Loft living: Culture and capital in urban change.* Rutgers University Press.