The Structure and Growth of Ethnic Neighborhoods

Tianran Dai Nathan Schiff

Shanghai University of Finance and Economics

Duke Kunshan University
September 15, 2023

Ethnic Neighborhood Examples: “Chinatowns” in New York City

In 2010 over 50% of Chinese-born population in New York City (425,210 people) lived in 207 of the city’s 4661 tracts (4.5%)

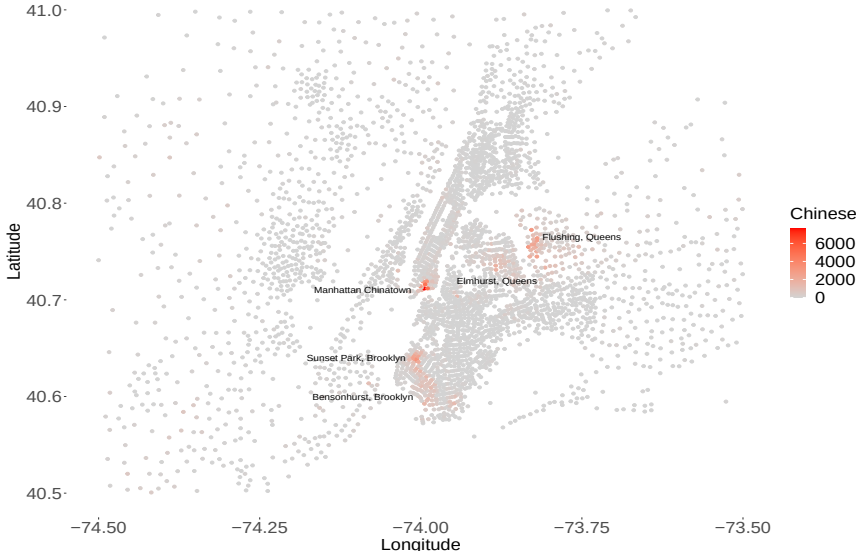
A single cluster of 26 contiguous tracts accounted for 9.2% of entire Chinese population

Within New York City, several of these neighborhoods (Manhattan Chinatown, Flushing, Elmhurst, Sunset Park, Bensonhurst) are known as “Chinatowns”

Concentration far exceeds that of US-born racial groups, but is similar to that of other foreign-born populations

Ethnic neighborhoods are a widely recognized feature of North American cities: “Chinatown, ”Little Italy,” “Japantown,” ”Historic Filipinotown,” ”Little Havana,” ”Greektown,” ”Little Saigon,” “Tehrangles”,“Olvera Street,” “Punjabi Market,” etc...

New York City Chinese Population: 2010



Ethnic neighborhoods have important causal effects...

Similar to recent literature on neighborhoods (Chetty, Hendren, and Katz 2016; Chetty and Hendren 2018) but place effects transmitted through ethnicity

- educational attainment (Borjas 1995, Aslund et. al. 2011)
- earnings and English language ability (Edin et. al. 2003, Cutler et. al. 2008, Damm 2009)
- job networks: (Hellerstein, Kutzbach, and Neumark 2014, Munshi 2003, Ioannides and Loury 2004)
- many health outcomes (ex: diet, smoking, diabetes, hypertension, breast cancer)

Additionally, long-standing policy concern about whether ethnic neighborhoods help or hinder assimilation

Countries may try to disperse refugees (ex: Germany in 2016) or use housing policy to prevent the formation of enclaves (ex: Denmark 2019)

...yet we lack a systematic understanding of these neighborhoods

How common are ethnic neighborhoods? What percentage of immigrants live in ethnic neighborhoods?

How big are ethnic neighborhoods? How long do ethnic neighborhoods last? How do they grow?

Sociology literature suggests geographic concentration of immigrants may be less important than in past: technological improvements in communications have decreased benefits of concentration, recent immigrants are more skilled, settle in a wider set of locations (Gold 2015)

Other papers suggest immigrants may be moving out of cities and concentrating in suburbs or "ethnoburbs" (Li 1998, Logan et al. 2002, Zhou et al. 2008)

Difficult questions to quantify without systematic definition of ethnic neighborhood

This paper

Three main contributions:

1. Provide a statistical definition of ethnic neighborhoods based on a choice model and show this method better predicts common ethnic neighborhood outcomes
2. Provide first comprehensive description of ethnic neighborhoods in the US, using constant-boundary census tract data from 1970-2010, and answer several fundamental questions from US immigration literature
3. Examine the structure and growth of ethnic neighborhoods—an area with no previous work—and find striking similarities to empirical patterns documented in cities

▸ Why define spatial boundaries?

Preview of results

Ethnic neighborhoods are

1. *pervasive and growing*: 43% of foreign-born population lived in ethnic neighborhoods in 1970, increasing to 67.2% by 2010
2. *decentralizing*: 56% increase in average distance to city center from 1970-2010.
3. *very large and very small*: neighborhood size follows a power law, similar to population distribution across cities (but not Zipf)
4. *mostly short-lived*: 2/3 of ethnic tracts disappear after a decade, but tracts in bigger neighborhoods persist much longer.
5. *growing through spatial expansion*: have significant population gradients declining from oldest tracts in center

Defining Neighborhoods

Definition of an Ethnic Neighborhood

No accepted definition: generally high ethnic population concentrations (residence, workplace)

"Spatial assimilation theory": ethnic neighborhoods help new immigrants—who often have few assets, difficulties speaking local language, and little knowledge of the new country—find jobs, housing, social support, and goods/services from the home country (Massey 1985)

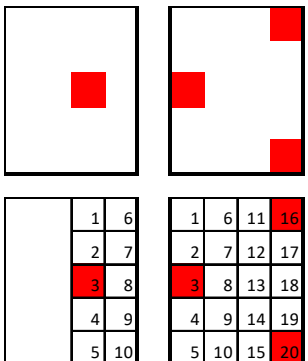
Our approach: define neighborhoods statistically as places with unusually large concentrations

We motivate our approach with a choice model and our statistical method adjusts for group size, housing capacity, and multiple hypothesis testing

A seminal paper by Logan, Zhang, Alba (2002) also uses a statistical approach (spatial correlation), but differs significantly in theoretical motivation, method, and inference

Simple choice model over housing

- We model choice of housing because geographic units for population data vary in size
- Household i in group g must choose a single house h from all (identical) houses in the J census tracts of a city
- Indirect utility: $V_{igjh} = ln(\theta_{gj}) + \epsilon_{igjh}$
- $ln(\theta_{gj})$: group level utility of *any* house in tract j (preferences and constraints)
- ϵ_{igjh}: distributed type 1 extreme value, i.i.d. across all houses in the city (individual-house match)
- Household chooses house with highest utility; probability P_{gjh} is usual logit expression of relative utilities



Choice-model based measure of concentration

Probability of choosing tract j is sum of identical probabilities of all h in j: $P_{gj} = H_j * P_{gjh}$; location shares represent utility and housing capacity

Concentration measure: share of a group g choosing a location j deflated by share of housing in a location

$$\frac{n_{g,j}/N_g}{H_j/H} = \frac{P_{gj}}{H_j/H} = \frac{P_{gjh}}{1/H}$$

- N_g is group size, $n_{g,j}$ is group pop. in j, H_j is housing in j, H is total housing, P_{gj} is probability g type chooses j, P_{gjh} is probability g type chooses h in j

Provides a way to compare shares of different groups, with different group sizes, across different locations in a city:

- If $\frac{n_{A,j}/N_A}{H_j/H} > \frac{n_{B,k}/N_B}{H_k/H}$ it implies the relative utility of an A type living in a house in j is greater than the relative utility of a B type in k

Implementation: Location quotient

We have limited data on housing capacity, proxy housing stock with total population

Common proxy: count of housing units is nearly perfectly correlated with population (Glaeser Gyourko 2005); theoretical housing supply models often assume housing units equivalent to population (ex: Saiz 2010)

Our measure becomes share of total ethnic population in tract j, $s_{g,j}$, divided by the share of total population in tract j, s_j

We refer to this as the "location quotient" of a tract: $LQ_{gj} = \frac{n_{gj}/N_g}{n_j/N} = \frac{s_{gj}}{s_j}$

Location quotient (LQ) often used in studies of industrial concentration, where denominator is used to deflate location share of an industry by all economic activity

Group data are simple counts; we divide by 2 to approximate households (av. immigrant household size between 2-3 people)

Choosing a threshold

Many papers test for concentration by comparing a granular group share to the overall aggregate share (Ellison and Glaeser 1997, Duranton and Overman 2005, Billings and Johnson 2012)

These papers typically test whether $s_{gj} = s_j$ ($LQ_{gj} = 1$) by comparing each n_{gj} to a counter-factual where the N_g members of group g are distributed according to aggregate shares s_j

Problem in our application: overall population is a poor predictor of location choices for any group (ex: NYC Chinese), benchmark is too easily passed.

Ethnic tracts should represent unusual concentrations resulting from factors specific to immigrant group (language, ethnic retail, cultural institutions, etc...)

Using overall population as benchmark, 50% of tracts would be classified as ethnic tracts for the *native* population

Natives as reference group

Since we study immigrant group concentration, native-born population is a natural reference group

However, native preferences/constraints for particular locations are unlikely to be similar to those of an ethnic group *g*

Instead, we compare *distribution* of preferences by assuming percentiles of each group's location quotient distribution are equal

For example: we assume that the 80th percentile location quotient for an ethnic group is equal to the 80th percentile LQ for natives

Thus we assume that preferences for both groups are equally strong compared to random assignment of houses, but do not assume each group has the same ranking over locations

One issue: we are pooling (obviously) heterogeneous natives. However, native groups are still far less concentrated than foreign-born.

Definition of an ethnic tract

We then define location j as an ethnic location for group g if LQ_{gj} is statistically greater than 99th percentile of LQ distribution of natives ▸ multiple hypothesis testing

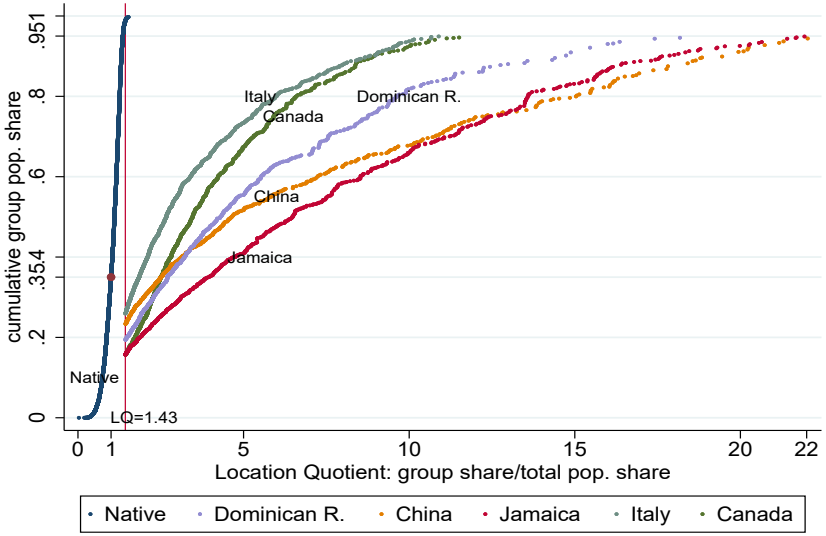
Interpretation: location j offers an unusually large relative utility to group g even when compared to the very high end of the native preference distribution

Note that all locations in a city are compared to the same threshold: we compare LQ_{gj} for all j to $LQ_{0,99}$

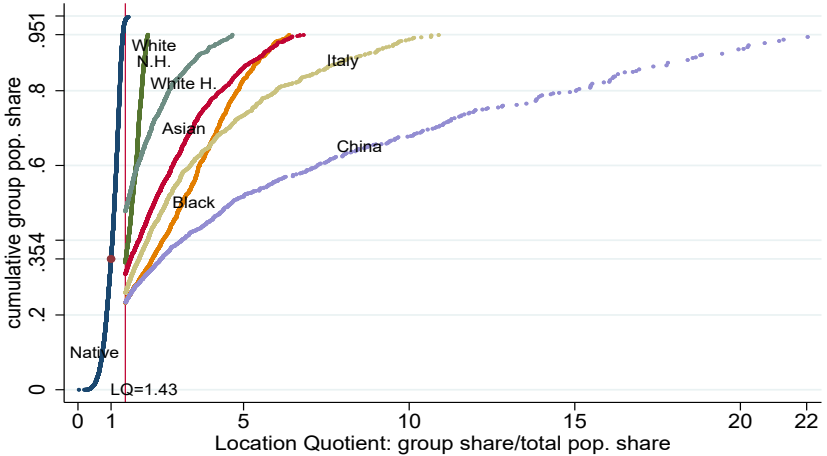
- If preferences are identical to natives, then we would find 1% of tracts are ethnic tracts. This is not the case (next slide). We use top 1% of native tracts as a placebo in some analyses.

Our method results in significantly fewer ethnic tracts than comparison to aggregate population ($LQ_{gj} = 1$): for NYC we classify 513 tracts as Chinese compared to 624 using other method

Cumulative population share by tract LQ, NYC 2010

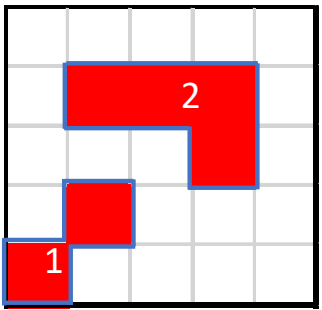


Cumul. pop. share by tract LQ, NYC 2010, native groups



Clustering tracts into neighborhoods

- Ethnic tract definition applied to each tract individually
- However, clearly highly concentrated tracts spatially cluster
- We define ethnic neighborhoods as clusters of spatially contiguous tracts; two tracts are contiguous if they share any part of their borders
- Implies that if tracts j and k are in the same ethnic neighborhood, then one can walk from j to k walking only through ethnic tracts



Neighborhood 1 has 2 tracts; neighborhood 2 has 4 tracts

Data, Maps, and Validating our Neighborhood Definition

Outline of Results Section

Exhibits and intended interpretation:

1. Show maps demonstrating that our neighborhood definition is reasonable: it captures known neighborhoods and allows us to look at neighborhoods cross-sectionally and over time
   - More formal validation in paper: compare efficacy of our method in predicting language usage and growth to simpler neighborhood definitions
2. Describe characteristics of neighborhoods *for average foreign-born resident* and how these have changed over time
3. Show exhibits on spatial structure and distribution (power law)
4. Summarize statistics on neighbor dynamics: attrition and growth

Data and Maps

Data from 1970, 1980, 1990, 2000 US Censuses; ACS 2006-2010 for year 2010

Population counts are summary level data–estimates from sample (not full enumeration)

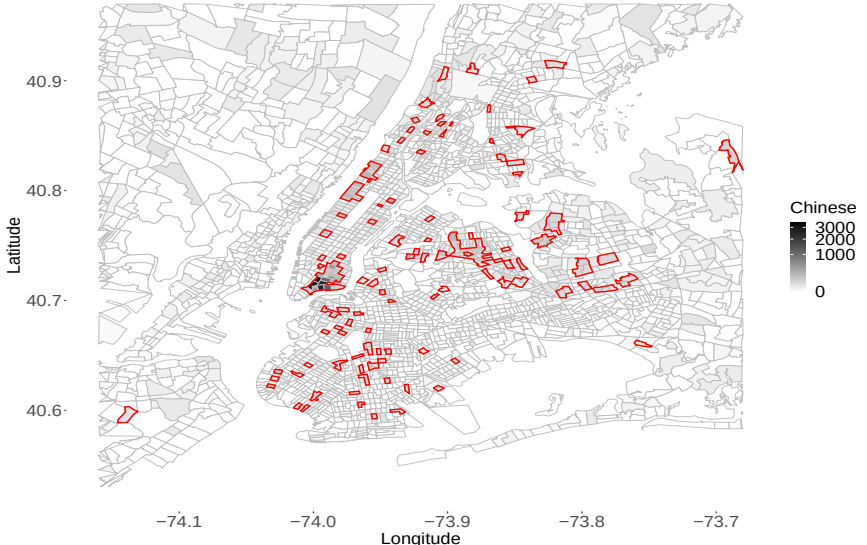
US Census records different birth countries in different census years (later Census years have much larger set of countries)

We reconcile all tract boundaries to 2010 spatial definitions using crosswalks from Logan, Xu, Stults (2012)

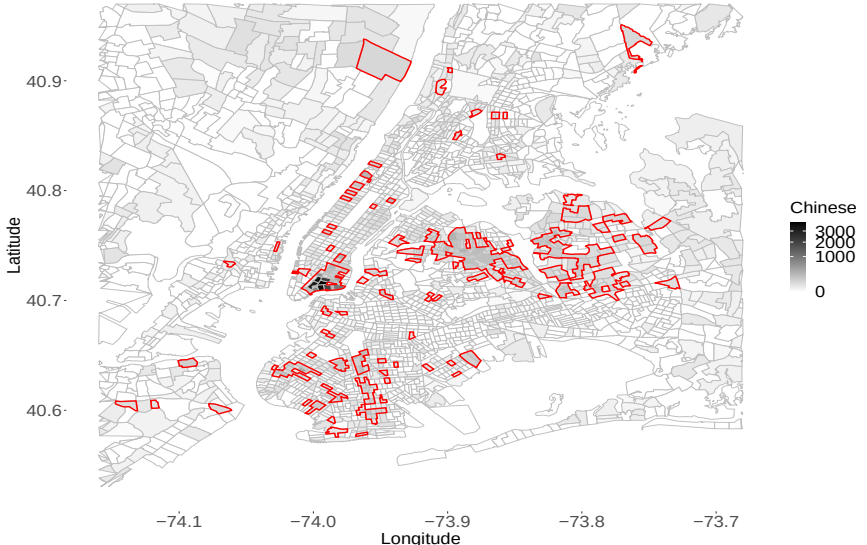
We limit our study to cities with at least 100 census tracts

Mapping website: https://nathanschiff.shinyapps.io/ee_maps_deployed/

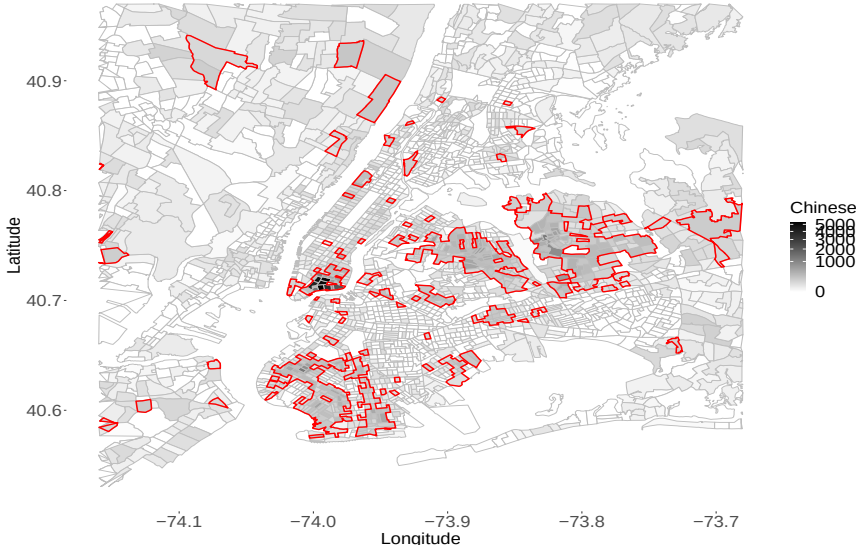
Chinese ethnic neighborhoods, NYC 1970



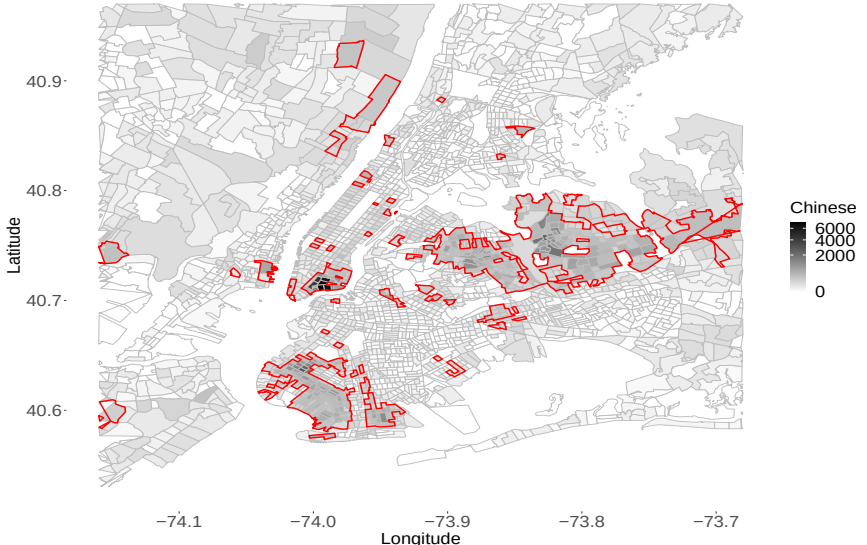
Chinese ethnic neighborhoods, NYC 1980



Chinese ethnic neighborhoods, NYC 1990

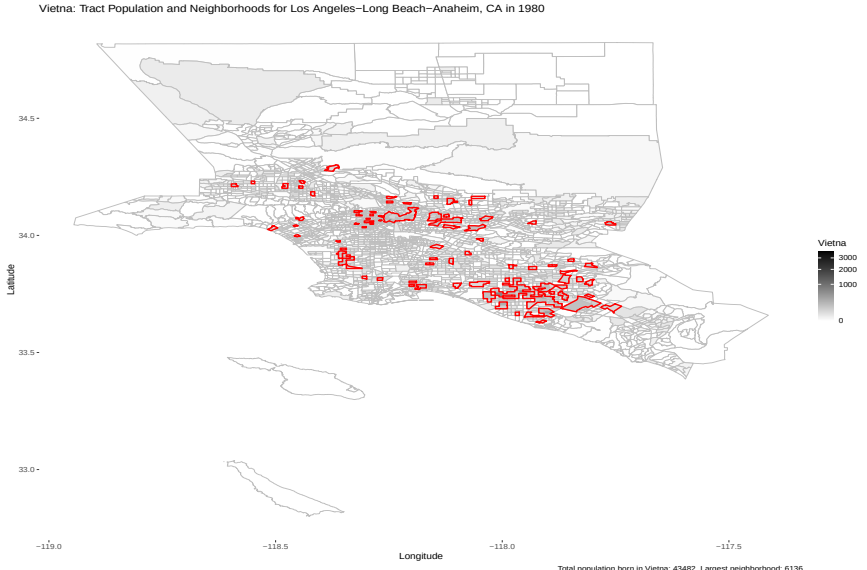


Chinese ethnic neighborhoods, NYC 2000

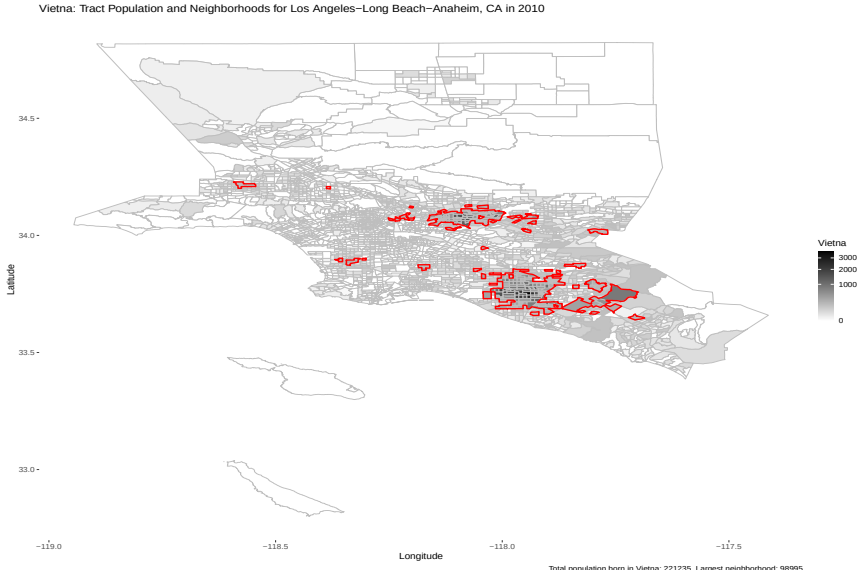


Chinese ethnic neighborhoods, NYC 2010

Vietnamese ethnic neighborhoods, LA 1980



Vietnamese ethnic neighborhoods, LA 2010



Characteristics of Neighborhoods

Characteristics of neighborhoods

We calculate the percentage of the foreign born in ethnic neighborhoods in each decade 1970-2010, as well as statistics for the average foreign-born resident in an ethnic neighborhood

Neighborhoods are getting larger and decentralizing, but natives are also decentralizing

Ethnic neighborhoods have higher shares of rental housing, older housing, commuting without a car, but this difference is decreasing over time (see paper)

HH incomes are lower (-$9000) and roughly constant over time; housing values and rent also lower (see paper)

Geographic Concentration and Decentralization of Ethnic Tracts

Population Statistics

Year	Pct. in E. Tracts	Av. Ethnic Pop.	Average LQ	Av. Ethnic Fraction
1970	43.1%	384	10.7	0.088
1980	49.6%	616	10.1	0.142
1990	57.0%	641	10.7	0.133
2000	60.1%	825	9.0	0.159
2010	67.2%	683	11.3	0.131

Decentralization

Year	Av. E. Tract Density	Av. Native Density	Av. E. Tract CBD Dist.	Av. Native CBD Dist.
1970	9329	3969	14.0	20.5
1980	7412	2897	16.1	23.1
1990	7498	2580	17.9	25.4
2000	6979	2522	19.7	26.5
2010	5590	2390	21.8	27.3

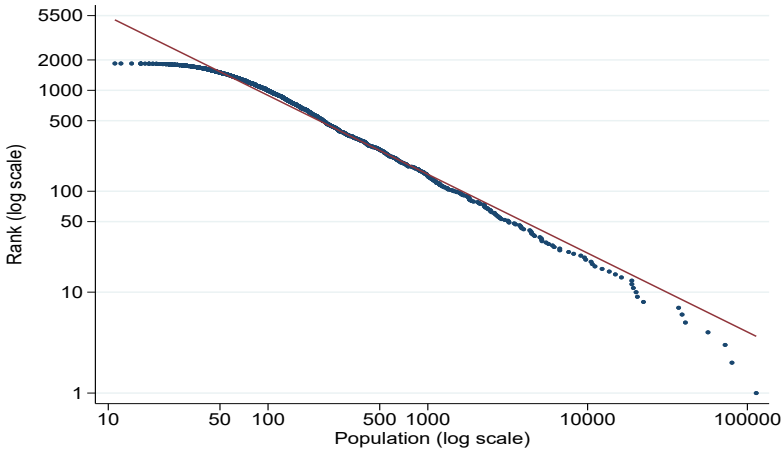
Highly Skewed Distribution of Neighborhood Sizes

Statistics for the average foreign born resident living in an ethnic neighborhood

Year	Mean Population	Median Population	Mean Tract Ct.	Median Tract Ct.
1970	18,310	1,858	37	13
1980	60,211	4,265	68	19
1990	59,847	4,895	63	20
2000	62,300	9,399	61	27
2010	42,425	4,032	44	12

Means are much larger than medians—why?

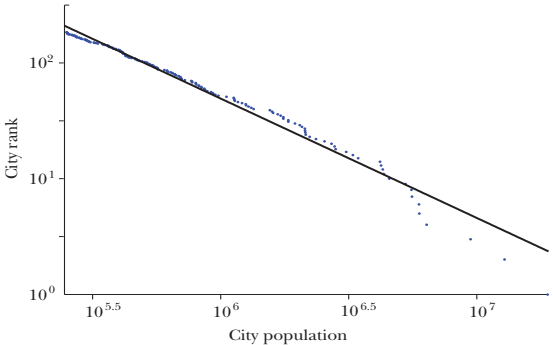
Chinese Neighborhoods 2010: Pop. Rank Pop. Size Plot



Power law for neighborhoods: $ln(rank) = 10.39 - 0.78 * ln(pop)$, $R^2 = 0.98$

Zipf's Law in US: Gabaix 2016

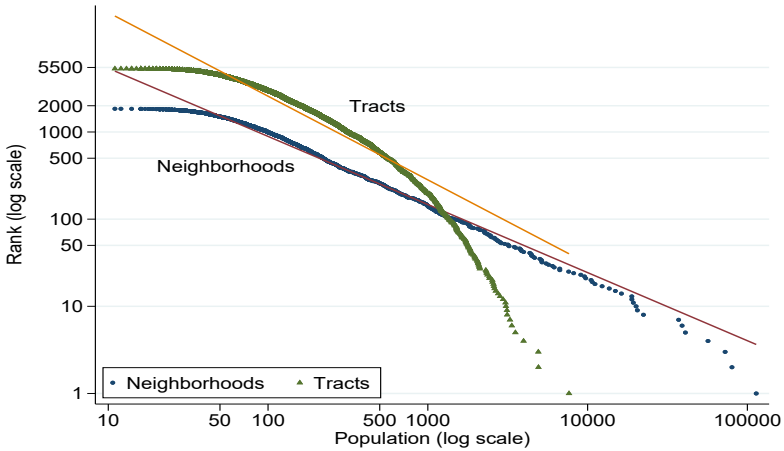
A Plot of City Rank versus Size for all US Cities with Population over 250,000 in 2010



Source: Author, using data from the *Statistical Abstract of the United States* (2012).
Notes: The dots plot the empirical data. The line is a power law fit ($R^2 = 0.98$), regressing ln *Rank* on ln *Size*. The slope is -1.03, close to the ideal Zipf's law, which would have a slope of -1.

Cities in many countries follow a power law distribution with a coefficient close to -1

Chinese Neighborhoods 2010: Pop. Rank Pop. Size Plot

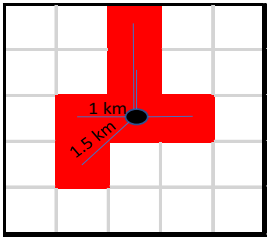


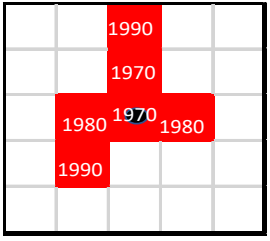
Power law only appears for neighborhoods, not tracts; not due to city size.

Spatial Structure and Growth

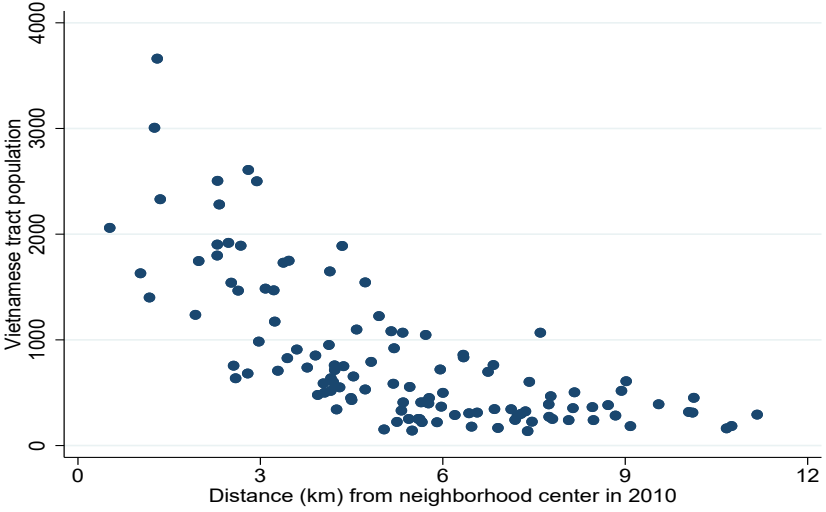
Neighborhood spatial structure

- How do neighborhood characteristics change with distance to center? We examine population, year tract joined neighborhood, age of housing stock, and rental percentage.
- We define the center of a neighborhood as the population weighted average of component tract centroids
- We then run gradient regressions with specification: $y_{jt,b} = \beta * ln(dist)_{jt,b} + \mu_b + \epsilon_{jt,b}$
- Second set of regressions replaces distance with join year
- Sample in all regressions limited to 2010 neighborhoods with at least five tracts

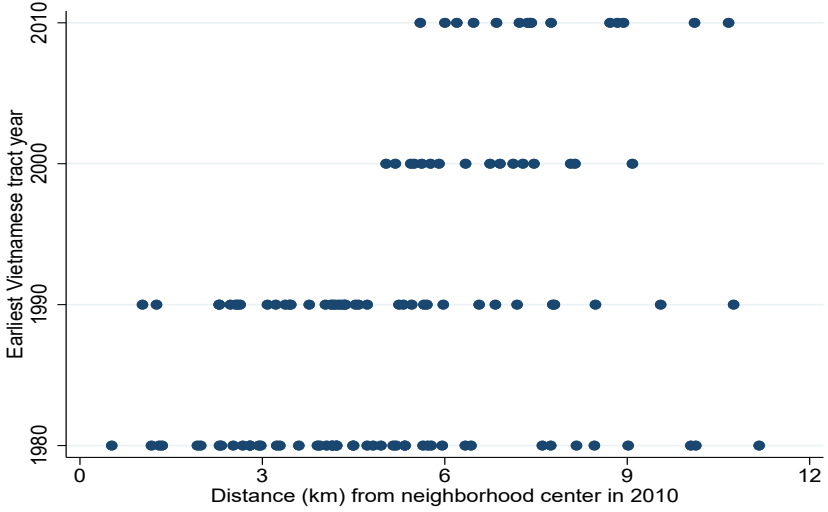




Vietnamese neighborhood: tract population by distance to center



Vietnamese neighborhood: tract status year



Log Ethnic Population by Distance

	(1) foreign	(2) Canada	(3) China	(4) Cuba	(5) Domin. R.	(6) India	(7) Italy	(8) Jamaica
log distance	-0.12***	-0.10***	-0.24***	-0.32***	-0.28***	-0.17***	-0.22***	-0.29***
	(0.01)	(0.01)	(0.02)	(0.04)	(0.02)	(0.01)	(0.02)	(0.02)
Observations	32992	3316	9069	4318	3235	8234	6237	4351
Clusters	1455	383	618	254	194	651	452	238
CI_lb	-0.04	-0.08	-0.09	-0.10	-0.10	-0.10	-0.07	-0.09
CI_ub	-0.01	-0.02	-0.04	-0.01	-0.01	-0.04	-0.02	-0.01

Note: By defining the neighborhood center using tract populations, a negative population gradient could result mechanically (ex: population weighted center is likely to be closer to larger tracts). We use a permutation procedure to adjust for mechanical gradients by randomly shuffling the tracts within a neighborhood, defining the center among the shuffled tracts, and re-running the regression.

Ethnic Tract Year by Distance

	(1) foreign	(2) Canada	(3) China	(4) Cuba	(5) Domin. R.	(6) India	(7) Italy	(8) Jamaica	
log distance	3.67***	1.42**	2.82***	6.19***	4.01***	2.35***	4.92***	4.72***	4
	(0.37)	(0.66)	(0.41)	(1.01)	(0.48)	(0.33)	(0.73)	(0.55)	
Observations	7256	1003	2989	777	1042	3016	822	1240	
Clusters	351	120	196	57	60	223	70	71	
Pred. Yr. Range	7.86	2.42	5.52	12.09	9.15	4.58	9.26	9.37	

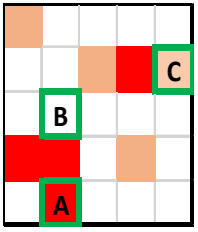
The last row of this panel shows the predicted age difference between the closest tract and further tract of a neighborhood, averaged across neighborhoods.

We also find that housing stock age and percentage rental decline from the center

Neighborhood growth

- First calculate transition matrix: if a tract is an ethnic tract in t, how likely is it to be a tract in $t + p$?
- Next run series of growth regressions looking at effects of ethnic tract status, neighborhood size, and adjacent tract status and adjacent neighborhood size:
- $\triangle_{(t+1)} n_{cj} = \beta_1 * ethtract_{cjt} + \beta_2 * adj.ethtract_{cjt} + \beta_3 * nbpop_{c,b(jt)} + \beta_4 * adj.nbpop_{c,a(jt)} + \mu_{ct} + \epsilon_{cjt}$
- Paper appendix: controlling for tract ethnic population, ethnic tract status and nbhd. pop. still predict growth (magnitudes smaller)

City with 25 tracts



A — Ethnic tract in nbhd w 3 eth. trcts

B — Adjacent to eth tract, adj. nbhd of 3 tracts

C — Adjacent to eth tract, adj. nbhd of 1 tract

Growth and Dynamics: Results

- **Duration:** 2/3 tracts disappear after a decade, but more populous tracts last longer ▸ transition matrix
- **Ethnic Tract Growth:** ethnic tract population predicts future population. Ex: a Chinese tract gains 27 more Chinese residents in next decade than average tract in same city; tracts adjacent to Chinese tracts gain 11.7 more Chinese residents. ▸ ethnic tracts growth
- **Ethnic Neighborhood Growth:** controlling for tract ethnic population, neighborhood population can predict growth ▸ ethnic neighborhood growth
- **Growth Decomposition:** partition population growth of individual neighborhoods into intensive (existing tracts) and extensive (new tracts) margins. Find most growth through new tracts (spatial expansion) ▸ growth decomposition

Summary: a description of US ethnic neighborhoods

1. Ethnic neighborhoods exist for practically every immigrant group and over half of the foreign born population lives in these neighborhoods
2. Neighborhoods are generally poorer and denser than other places where the same group lives, but they are moving away from the center of the city and becoming more car-oriented and owner-occupied
3. The size distribution of ethnic neighborhoods follows a power law: most neighborhoods are small, but the biggest neighborhoods contain large shares of the national group population
4. Most ethnic neighborhoods disappear after a decade, but bigger neighborhoods persist longer
5. Neighborhoods have a consistent spatial structure: the highest density of group population is in the center with older rental housing. These are also the first tracts in the neighborhood with spatial expansion over time.
6. Existing neighborhoods attract additional immigrants; most population growth occurs through spatial expansion.

Conclusion

We provide a statistical definition of an ethnic neighborhood that can be used across groups, over time, and with different spatial units

This definition of an ethnic neighborhood may be useful (we hope!) for many questions on the causal effects of ethnic neighborhoods

We use our method to provide the first systematic description of ethnic neighborhoods in the US

Some governments are concerned about ethnic neighborhoods and are trying to use housing policy to discourage their growth (see Economist magazine “Denmark wants to break up ethnic enclaves. What is wrong with them?”)

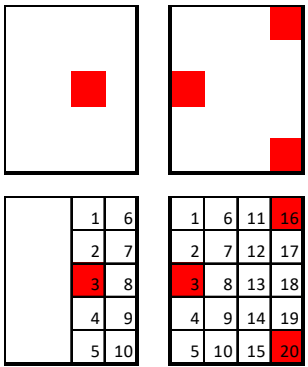
We find that ethnic neighborhoods have specific housing and commuting characteristics, and existing neighborhoods attract new immigrants.

These results *suggest* housing policy could play an important role in formation and growth

Thank you!

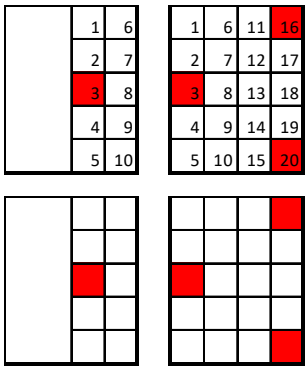
Simple choice model to explain location patterns

- Geographic units for population data often vary in size. We model the choice of housing, rather than location.
- Household i in group g must choose a single house h from all the housing in the J census tracts of a city
- Indirect utility: $V_{igjh} = ln(\theta_{gj}) + \epsilon_{igjh}$
- $ln(\theta_{gj})$: group level utility of *any* house in tract j (interpreted as preferences and constraints, including discrimination)
- ϵ_{igjh}: distributed type 1 extreme value, i.i.d. across all houses in the city (individual-house match; ex: basement has a built-in bar)



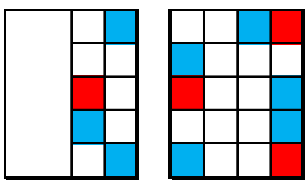
Probability of choosing a house; probability of choosing a location

- We don't observe individual housing choices, assume all houses in location j are identical with total stock in j: H_j
- Probability of choosing one of the identical houses is then:
- $$P_{gjh} = \frac{exp(ln(\theta_{gj}))}{\sum_{k=1}^{J}\sum_{h=1}^{H_k} exp(ln(\theta_{gk}))} = \frac{\theta_{gj}}{\sum_{k=1}^{J} H_k \theta_{gk}}$$
- Probability of choosing tract j is sum of identical probabilities of all h in j: $P_{gj} = H_j * P_{gjh}$



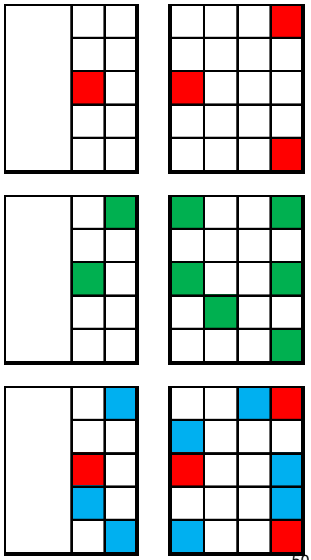
Ethnic group tract choice shares

- Can estimate P_{gj} with share of group g choosing location j: $s_{gj} = n_{gj}/N_g$
- If $s_{Aj} > s_{Bk}$ implies group A has greater relative utility for location j than B for k, or there are more houses in j than k, or both (ex: $s_{R,2} > s_{B,1}$)
- Can adjust for this issue by dividing a share s_{gj} by the share of total housing in the location, H_j/H
- This yields a ratio equivalent to comparing the probability a household in g chooses house h to a uniform distribution over all houses:
- $\frac{s_{g,j}}{H_j/H} = \frac{P_{gj}}{H_j/H} = \frac{P_{gjh}}{1/H}$



The ratio of ethnic group shares to housing shares

- The ratio $\frac{s_{g,j}}{H_j/H}$ provides a simple way to adjust for underlying capacity (housing)
- Unlike the proportion of a location's population in a group, $\frac{s_{g,j}}{H_j/H}$ is not a function of total group population
- Provides a way to compare shares of different groups across different locations in a city: if $\frac{s_{R,j}}{H_j/H} > \frac{s_{B,k}}{H_k/H}$ then $P_{Rjh} > P_{Bkh}$
- In words: the probability group R selects a randomly drawn house from location j is greater than the probability group B selects a house from k
- In picture, $\frac{s_{R,2}}{H_2/H} = \frac{s_{B,1}}{H_1/H} = 1.125$



Hypothesis testing

Billings and Johnson (2012): testing J locations for concentration represents J hypothesis tests, need to adjust p-values

If we have J independent tests and use a 5% significance level then even if null hypothesis is always true we would falsely reject $0.05 * J$ tests

Our case is different; even under null hypothesis we would reject fewer than $0.05 * J$, but still an issue

Following BJ, we use "family wise error rate" (FWER) concept: adjust p-values so that probability of making one or more false rejections under null hypothesis is only 5%

Say a city has no ethnic tracts and we test our null hypothesis on all J tracts. If at least a single tract passes, we say the city has ethnic tracts. We set the FWER so that if we tested the city 100 times, in at most five of those times would we falsely conclude the city has ethnic tracts.

Implementation of FWER

Our assumption is that the distribution of ethnic group LQs (percentiles) is the same as the native LQs

Implementation:

1. draw N_g people from native distribution across tracts with replacement
2. calculate minimum p-value from testing whether drawn tract population is greater than threshold, but only among first 99 percentiles
3. repeat procedure 50,000 times (50,000 counterfactual cities) to get 50,000 minimum p-values, use 5th percentile from 50,000 minimums as adjusted p-value

Lastly: we have data on individuals but model describes households; we assume household size of two by dividing tract populations by two (LQ is same, but tract p-values will be larger) ▸ defining ethnic tracts

Housing, Income, and Commuting

Year	Rental Fraction	Cmt. No Car Fraction	H. Age. Gt. 30 Yrs. Frac.	Med. H.H. Income	Med. H. Rent	Med. H. Value
1970	0.127	0.094	0.151	-8.816	-0.817	-0.096
	(0.007)	(0.009)	(0.009)	(0.515)	(0.085)	(0.016)
1980	0.107	0.077	0.122	-6.972	-0.751	-0.152
	(0.007)	(0.007)	(0.009)	(0.698)	(0.112)	(0.047)
1990	0.095	0.064	0.073	-9.258	-0.833	-0.178
	(0.006)	(0.005)	(0.007)	(0.733)	(0.091)	(0.081)
2000	0.086	0.046	0.057	-8.674	-0.658	-0.188
	(0.006)	(0.006)	(0.009)	(0.684)	(0.098)	(0.044)
2010	0.081	0.029	0.070	-9.625	-0.752	-0.251
	(0.006)	(0.005)	(0.011)	(0.796)	(0.095)	(0.051)

Power Law Regressions: Log Rank on Log Population

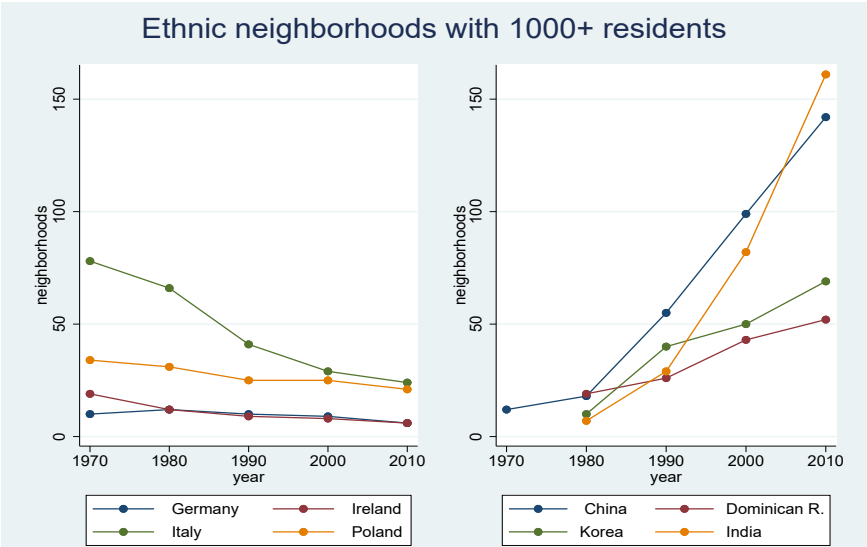
$$ln(nbrank_n - 1/2) = \alpha + \zeta ln(nbpop_n) + \epsilon_n$$

	(1) Mexico	(2) China	(3) Philip	(4) India	(5) Vietna	(6) ElSalv	(7) Korea	(8) Cuba	Ca
log pop	-0.70***	-0.86***	-0.95***	-0.92***	-0.97***	-0.93***	-0.99***	-1.07***	-1.
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0
Observations	1466	1378	1771	1427	1477	1363	1445	1312	1
R^2	0.971	0.996	0.989	0.991	0.996	0.989	0.994	0.987	0.

Power law test (Gabaix and Ibragimov 2011) fails to reject a power law for 7 of 10 groups

Sample restricted to neighborhoods larger than 25th percentile.

Count of large neighborhoods across US, selected groups



Ethnic Tract Transition Matrix

Initial Year	Type	1980	1990	2000	2010
1970	unweighted	0.37	0.27	0.23	0.18
	weighted	0.66	0.54	0.46	0.38
1980	unweighted		0.39	0.33	0.27
	weighted		0.72	0.64	0.57
1990	unweighted			0.35	0.28
	weighted			0.71	0.62
2000	unweighted				0.32
	weighted				0.71

Weighted: weight each tract by share of tot. pop. in ethnic tracts in initial year. In other words, for an ethnic resident in a tract in t, what is likelihood tract of residence remains ethnic tract in $t + p$?

Ethnic tracts and growth

	(1) Native	(2) China	(3) India	(4) Italy	(5) Jamaica	(6) Mexico
ethtract	-269.64***	27.30***	15.57***	-43.27***	0.65	112.18***
	(24.46)	(1.77)	(1.44)	(0.73)	(1.20)	(2.65)
adjacent to ethtract	-110.65***	11.77***	11.74***	-0.80***	6.37***	45.66***
	(12.04)	(0.30)	(0.31)	(0.13)	(0.28)	(1.01)
Observations	244268	244292	192238	244292	192238	244292

Interpretation example: A Chinese tract in a city gains 27 more Chinese residents in the following decade than the average tract in the same city. Tracts adjacent to Chinese tracts gain 11.7 more Chinese residents.

Ethnic neighborhoods (pop in 000's) and growth

	(1) Native	(2) China	(3) India	(4) Italy	(5) Jamaica	(6) Mexico
ethtract	-101.18***	11.43***	5.41***	-38.35***	-3.75***	124.92***
	(36.49)	(1.32)	(1.22)	(0.69)	(1.06)	(3.01)
adjacent to ethtract	-66.81***	10.46***	10.12***	-0.77***	5.20***	41.91***
	(16.46)	(0.29)	(0.29)	(0.13)	(0.24)	(0.98)
neighborhood pop	-42.78***	2.25***	4.84***	-1.44***	0.71***	-0.23***
	(9.02)	(0.16)	(0.53)	(0.14)	(0.13)	(0.02)
adjacent neighborhood pop	-12.33***	0.52***	1.65***	-0.08	0.84***	0.21***
	(3.56)	(0.06)	(0.17)	(0.06)	(0.10)	(0.02)
Observations	244268	244292	192238	244292	192238	244292
Neigh. Pop Mean	3.94	7.29	2.18	3.60	6.84	56.11
Adj. Neigh. Pop Mean	3.54	1.94	0.94	1.17	1.41	17.62

Additional results on neighborhood growth

We partition population growth of individual neighborhoods into intensive (existing tracts) and extensive margins (new tracts)

Find that most population growth occurs through new tracts joining neighborhood (spatial expansion)

	(1) Canada	(2) China	(3) Cuba	(4) India	(5) Italy	(6) Jamaica	(7) Mexico	(8) Vietnam
Ext. Share	1.03	0.80	1.92	0.74	4.37	0.88	0.61	0.87
	(0.04)	(0.10)	(0.44)	(0.05)	(2.65)	(0.14)	(0.03)	(0.05)
N	297	559	159	568	152	184	801	433

Notes: Extensive margin share can be greater than one if intensive margin tracts are losing population on average. Intensive share and extensive share always add to one. Sample limited to neighborhoods with at least five tracts and positive growth.

Table 8: Partitioning neighborhood growth

Table 8: Share of Neighborhood Growth in New Ethnic Tracts

	(1) Canada	(2) China	(3) Cuba	(4) India	(5) Italy	(6) Jamaica	(7) Mexico	(8) Vietnam
Ext. Share	1.03	0.80	1.92	0.74	4.37	0.88	0.61	0.87
	(0.04)	(0.10)	(0.44)	(0.05)	(2.65)	(0.14)	(0.03)	(0.05)
N	297	559	159	568	152	184	801	433

Note: We partition a neighborhood's tracts in decade *t* into those that were already ethnic tracts in the previous decade (intensive margin) and those that became an ethnic tract in *t* (extensive margin). We sum the ethnic population change between periods among all neighborhood tracts and then calculate the share of the neighborhood change in extensive margin tracts ("Ext. Share"). We use all years of data but limit the sample to neighborhoods with at least five tracts and an increase in total population. We then calculate the mean over all neighborhoods—each neighborhood is one observation—and put the standard error in parentheses. For some groups, intensive margin tracts lose population on average, and thus the extensive share is larger than one (see text).

Table 13a: Count of language speakers

	(1) China	(2) China	(3) Italy	(4) Italy	(5) Korea	(6) Korea	(7) Vietnam	(8) Vietnam
ethnic population	1.233***	1.231***	1.283***	1.310***	1.050***	1.114***	1.123***	1.078***
	(0.026)	(0.008)	(0.034)	(0.065)	(0.033)	(0.016)	(0.023)	(0.026)
non-ethnic population	0.001***	0.002***	0.001***	0.001***	0.000	0.000	0.001***	0.002***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
ethnic tract	-8.062		-9.872***		-8.093***		-20.852**	
	(5.676)		(2.030)		(2.665)		(7.960)	
eth. pop. X eth. tract	0.061**		0.215***		0.100***		-0.009	
	(0.030)		(0.024)		(0.027)		(0.025)	
non-eth. pop. X eth. tract	0.001*		-0.001		0.000		0.004***	
	(0.001)		(0.000)		(0.000)		(0.001)	
adj. eth. pop. 000's		6.384*		50.441***		9.151***		3.797**
		(3.477)		(6.222)		(1.868)		(1.476)
neigh. pop. 000's		1.028***		1.445***		-0.042		0.357**
		(0.392)		(0.380)		(0.298)		(0.139)
Observations	20345	20345	9844	9844	15550	15550	13474	13474
Clusters	130	130	45	45	105	105	103	103
dep-var mean	121	121	52	52	65	65	87	87
nb-var mean	242.31	10.81	69.89	0.88	148.47	4.97	173.05	6.51
Adj. R2	0.96	0.96	0.77	0.78	0.96	0.96	0.92	0.92

Table 13b: Fraction with low level English

	(1) China	(2) China	(3) Italy	(4) Italy	(5) Korea	(6) Korea	(7) Vietnam	(8) Vietnam
language speakers 000's	0.026	0.043**	-0.578***	-0.031	-0.408***	-0.057***	-0.068	0.020
	(0.033)	(0.018)	(0.153)	(0.040)	(0.116)	(0.016)	(0.120)	(0.021)
ethnic tract	0.057***		0.090***		0.023**		0.081***	
	(0.006)		(0.014)		(0.009)		(0.008)	
l. speakers X eth. tract	0.013		0.505***		0.411***		0.081	
	(0.038)		(0.173)		(0.113)		(0.118)	
adj. eth. pop. 000's		0.001		0.024		0.013**		0.023*
		(0.005)		(0.023)		(0.005)		(0.012)
neigh. pop. 000's		0.001***		0.028***		0.003***		0.000
		(0.000)		(0.002)		(0.001)		(0.000)
Observations	17866	17866	8564	8564	11850	11850	11144	11144
Clusters	130	130	45	45	105	105	103	103
dep-var mean	0.50	0.50	0.32	0.32	0.56	0.56	0.58	0.58
nb-var mean	0.32	11.05	0.11	0.95	0.18	5.42	0.20	6.93
Adj. R2	0.02	0.01	0.03	0.01	0.01	0.01	0.02	0.01

(b) Fraction of Language Speakers with Low Level English

Table 15: Rental percentage and two decade lagged housing age

Table 15: Housing Age and Rental Percentage

	(1) Rental Pct	(2) Rental Pct
H. Age Pct: under 5y	-0.12***	-0.10***
	(0.02)	(0.02)
H. Age Pct: 5-10y	0.07***	0.11***
	(0.02)	(0.02)
H. Age Pct: 10-20y	0.02	0.02
	(0.02)	(0.02)
H. Age Pct: 20-30y	0.02	-0.00
	(0.02)	(0.03)
H. Age Pct: over 30y	0.31***	0.34***
	(0.02)	(0.03)
Observations	186363	139826
Lag	1	2
Clusters	114	114
Adj. R2	0.19	0.19

Table 16: Growth and rental percentage, first stage

Table 16: Neighborhood Expansion and Rental Percentage: First Stage

	(1) Canada	(2) China	(3) Cuba	(4) India	(5) Italy	(6) Jamaica	(7) Mexico	(8) Vietnam
diff_rpl2	0.10***	0.07***	0.07***	0.10***	0.12***	0.10***	0.15***	0.09***
	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)
ethnic pop.	-0.00***	0.00***	-0.00	0.00***	-0.00***	0.00***	0.00**	0.00***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
total pop.	0.01*	-0.00	-0.02***	-0.02***	0.01***	-0.02**	-0.01**	-0.01
	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)
median price	-0.18***	-0.18***	-0.16***	-0.23***	-0.22***	-0.16***	-0.08***	-0.20***
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)
median rent	-0.13***	-0.16***	-0.18***	-0.14***	-0.14***	-0.17***	-0.18***	-0.20***
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Observations	17266	18135	9053	20740	12613	8709	19202	17475
Clusters	2522	2202	1234	2570	1702	1132	2243	2216

Note: Dependent variable is rental percentage of tract housing stock. All specifications include adjacent neighborhood fixed effects, standard errors are clustered at adjacent neighborhood level.

Why define spatial boundaries?

Can study ethnic concentration on outcomes using continuous measures of population—why discretize a continuous variable?

Advantage 1: neighborhood definitions allow simple comparisons of high and low ethnic population concentrations on additional variables

- Jewish enclaves in New York City (Abramitzky, Boustan, Connor 2020), ethnic enclaves in Sweden (Edin, Fredriksson, Aslund 2003), Black American ghettos in the US (Cutler and Glaeser 1997), health outcomes in many public health papers

Advantage 2: neighborhoods can capture non-linearity, otherwise difficult to model

- Effect of ethnic population size on outcomes is non-linear, spatial effects are non-linear: not simple functions of distance, but complicated shapes
- Very difficult to study formation and growth of concentrated populations without a discrete boundary (spatial delineation)

▸ This Paper