Title: Developing measures of Asian American and Hispanic/Latino ethnic enclaves for five states using U.S. Census and American Community Survey data

Authors: Shariff-Marco S[1,2,3], Lin K[1,2], Meltzer D[1], Allen L[1], Boscoe F[4], Canchola AJ[1,2], Gates Kuliszewski M[5], German S[6], Guan A[1], Harris G[6], Henry K[7], Hiatt RA[1,3], Hughes AE[8,9], McGuire V[1], Oh D[1,2], Paddock LE[6], Pinheiro PS[10], Radadiya HR[8], Reyes S[8], Stroup A[6], Zhu H[8,9], Gomez SL[1,2,3], Pruitt SL[8,9].

Affiliations:

1. Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA
2. Greater Bay Area Cancer Registry, University of California San Francisco, San Francisco, CA
3. Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, CA
4. Pumphandle LLC, Camden, ME
5. New York State Cancer Registry, Albany, NY
6. Rutgers Cancer Institute of New Jersey, New Jersey State Cancer Registry, New Brunswick, NJ
7. Department of Geography, Temple University
8. Department of Population & Data Sciences, University of Texas Southwestern Medical Center, Dallas, TX
9. Harold C. Simmons Comprehensive Cancer Center, University of Texas Southwestern Medical Center, Dallas, TX
10. Department of Public Health Sciences and Sylvester Comprehensive Cancer Center, University of Miami, Miami, FL

Suggested citation: Shariff-Marco S, Lin K, Meltzer D, Allen L, Boscoe F, Canchola AJ, Gates Kuliszewski M, German S, Guan A, Harris G, Henry K, Hiatt RA, Hughes AE, McGuire V, Oh D, Paddock LE, Pinheiro PS, Radadiya HR, Reyes S, Stroup A, Zhu H, Gomez SL, Pruitt SL. Developing measures of Asian American and Hispanic/Latino ethnic enclave for five states using U.S. Census and American Community Survey data. University of California San Francisco, San Francisco, CA; October 2021. Online available at: https://cancerregistry.ucsf.edu/resources/research-tools.

Funding: R01CA237540-01A1 (MPIs: Salma Shariff-Marco, PhD (contact PI); Sandi Pruitt, PhD), Ethnicity and Nativity in Cancer – Latino & Asian Enclaves: The ENCLAVE study

For more information:
The online dashboard showing the measures described in this report is available here: https://cancerregistry.ucsf.edu/enclave.

For questions please contact Salma Shariff-Marco at salma.shariff-marco@ucsf.edu and Sandi Pruitt at Sandi.Pruitt@UTSouthwestern.edu

I. Objectives of the Study

Hispanic/Latino and Asian American populations are among the fastest growing populations in the U.S.[1-4] By 2060, more than one quarter of the U.S. population will be Hispanic/Latino and nearly one in ten will be Asian American (AA).[4] A large proportion of Hispanic/Latino and AA residents in the U.S. live in ethnic enclaves—distinct neighborhoods with high concentrations of individuals of the same ethnic origin ("co-ethnic residents") that are often characterized by recent immigration, linguistic isolation, ethnic businesses and resources, and disproportionate poverty.[5-9]

Ethnic enclaves are neighborhood environments hypothesized to contribute to outcomes across the cancer continuum through multiple pathways, some positively and some negatively. Co-ethnic residents within enclaves often maintain lifestyles, cultural norms, and behaviors (e.g., diet and physical activity, social networks, social cohesion) that are health-promoting. Enclaves may facilitate communication and information sharing due to greater access to linguistically concordant resources, including a higher percentage of residents who speak languages other than English and the presence of businesses and /or community organizations providing services and resources using languages other than English. Enclaves may also improve health by reducing or buffering exposure to racism and discrimination, which are linked to unhealthy coping behaviors (e.g., smoking, drinking), psychological responses such as stress, maladaptive physiological responses, and lower individual and collective resilience.[10-15] In contrast, socioeconomic and structural features of enclaves may contribute to unhealthy behaviors and worse health through pathways associated with low SES such as low walkability (resulting from high traffic density, poor street conditions, and low safety), poor food environments (with high concentrations of fast foods, tobacco outlets or liquor stores), or environmental hazards (poor air quality, proximity to toxic waste/landfills).

Cancer is the top cause of death among Hispanic/Latino and AA populations; yet it is generally unknown whether and how enclave residence may play a role in shaping cancer outcomes. There is no gold standard method to define ethnic enclaves and use of different measures and methods is one potential explanation for the mixed findings in the published literature about the association between enclave residence and cancer outcomes. Two reviews on the impact of ethnic enclave residence and outcomes across the cancer continuum demonstrate inconsistent associations for breast, cervical, and colorectal cancer (CRC). These three cancers have established evidence for effective early detection practices and are the focus of the ENCLAVE study (Ethnicity and Nativity in Cancer - Latino & Asian Enclaves).[5,6] The ENCLAVE Study (R01CA237540-01A1; MPIs Shariff-Marco and Pruitt) will pool cancer registry data across 5 states (CA, FL, NJ, NY and TX) containing the majority of these populations and link to contextual data to determine which patients live in ethnic enclaves. We will study the relationship among ethnic enclaves, patient nativity (U.S.- or foreign-born) and cancer incidence, stage at diagnosis, and survival for three cancer types (breast, cervix, colorectal) for Asian American and Latino populations, with attention to the 9 major ethnic groups (Cuban, Mexican, Puerto Rican, Chinese, Filipino, Japanese, Korean, South Asian and Vietnamese). Here we report on the development of multidimensional ethnic enclave measures at the census tract level for Asian American and Hispanic/Latino communities.

II. Methods for 5-State Pooled Ethnic Enclave Measures

We used data from the U.S. Census 2000 Summary File 3 (tables P6, P7, P19, P20, PCT63D, PCT63H) and the 2008-2012 American Community Survey (tables B02001, B03002, B06004D,

B06004I, B16002, B16005) to develop census tract measures for 2000 and 2010 geographies. We started with a broad set of variables that were previously applied in the ethnic enclave literature (see Appendix A). We noted that some variables were defined generally, without regard to specific race/ethnicity nor country of origin. For example, previous studies often measured the percent of census tract residents that were foreign born, without regard to country or region of birth. To ensure that our measures were relevant to our populations of interest, we opted to apply more specific measures relevant to each racial/ethnic group—these included % of residents Asian American or Hispanic/Latino, % of residents foreign-born Asian American or Hispanic/Latino, % of residents with limited English speaking Asian/Pacific Islander (API) languages or Spanish language, % of linguistically isolated households speaking API languages or Spanish language.

For both 2000 and 2010, we performed a principal components analysis (PCA) using 4 individual variables (hereafter variables) for each enclave measure (Asian American, Hispanic/Latino) pooling data across the 5 states in the ENCLAVE study. PCA is a data-driven data reduction technique aimed at reducing variables into a composite measure(s) while retaining as much of the variance in the data as possible. Each created component is a linear combination or weighted average of the variables. The first component is optimally weighted so that it accounts for the maximum amount of variance in the data. The second component accounts for the maximum variance not accounted for in the first component while being uncorrelated with the first component, etc. The eigenvalue of each component gives the amount of variance accounted for by the component out of the total variance. The decision of how many components to retain can be based on project goals as well as the variance explained, plots, and eigenvalues >=1. Eigenvector values for each variable reflect the weight assigned to each component variable in the linear combination.[16] In the ENCLAVE study, the goal was to develop a single measure, thus the analysis in this report and the future studies is focused on only the first component. Census tracts with small numerator counts were suppressed by Census and those with zero population or household denominator counts were excluded because composite measures could not be calculated (see Appendix B). We report the following PCA results: Eigenvalue, percent of variance explained, eigenvector values.

We categorized ethnic enclaves using the 5-state pooled data (hereafter “pooled” data) in two ways: quintiles and a dichotomous measure. Quintiles were used to identify less culturally/ethnically distinct neighborhoods from more culturally/ethnically distinct neighborhoods. In addition to quintiles, we also developed a dichotomous measure to categorize census tracts as ethnic enclave or non-enclave neighborhoods. Census tracts were classified as enclaves if they met either of the following criteria: 1) quintile 5 with >250 residents of the respective racial/ethnic group (Asian American or Hispanic/Latino); or 2) quintile 4 with >250 residents of the respective racial/ethnic group (Asian American or Hispanic/Latino) *and* spatially adjacent to a quintile 5 enclave census tract. On average, census tracts have 4,000 residents and 250 is approximately 6% reflecting the national percent of Asian American residents in the U.S. Census tracts were considered spatially adjacent following the Queen’s definition of adjacency, i.e., when one or more points of a census tract boundary touch one or more points of another census tract boundary.

Appendix C reports on methods and results for state-specific indices.

III. Results of 5-State Pooled Ethnic Enclave Measures

Results from the pooled PCA are shown below in Tables 1 and 2. The eigenvalue for the first component gives the amount of variation accounted for by the first component, out of a total variation of

4. Because our goal was to define one ethnic enclave measure and because the percent of variation explained by the first component was high (for both Asian Americans and Hispanic/Latino and both years), we retained only the first component as our enclave measure. The eigenvector values for the first component across the 4 variables were similar (for both Asian American and Hispanic/Latino enclaves and for both years), which means the 4 variables were relatively equally weighted in the linear combination.

Table 3 shows the enclave indices distribution by state. Across all states, approximately 25% of census tracts were identified as Asian American enclaves and 31% of census tracts were identified as Hispanic/Latino enclaves. California had the highest proportion of Asian American enclaves and Florida had the lowest. California followed by Texas had the highest proportion of Hispanic/Latino enclaves. Tables 4 and 5 show the distributions of the 4 variables included in the respective Asian American and Hispanic/Latino ethnic enclave indices/measures by year and state. Proportions in Q5 and enclaves were fairly consistent across states. Asian American enclaves had a higher proportion of foreign-born residents (68% of Asian American residents in Asian American enclaves were foreign-born in 2000 and 66% in 2010) than Hispanic/Latino enclaves (41% of Hispanic/Latino residents in Hispanic/Latino enclaves were foreign-born in 2000 and 39% in 2010). Appendix D includes maps that show the Asian American and Hispanic/Latino enclave pooled index for 2010 census tracts separately by state. Please see https://cancerregistry.ucsf.edu/enclave for interactive maps and additional data.

Table 1. Principal Components Analysis Results for Pooled Asian American Ethnic Enclave Index, 2000 and 2010 Census Tracts in California, Florida, New Jersey, New York, and Texas.

Asian American Ethnic Enclave Index	**2000**	**2010**
Eigenvalue (first component)	3.48	3.35
Variance explained (first component)	87.0%	83.6%
Eigenvector values (first component)		
% Asian American (Asian American count/total population)	0.51	0.51
% Foreign-Born Asian American (Asian American who are foreign-born/total population)	0.51	0.52
% Limited English: Asian/Pacific Islander (API) languages (People who speak English “not well” or “not at all” and speak API languages/total population)	0.49	0.49
% Linguistically Isolated: API languages (Households that are linguistically isolated and speak API languages/total households)	0.49	0.48

Table 2. Principal Components Analysis Results for Pooled Hispanic/Latino Ethnic Enclave Index, 2000 and 2010 Census Tracts in California, Florida, New Jersey, New York, and Texas.

Hispanic/Latino Ethnic Enclave Index	**2000**	**2010**
Eigenvalue (first component)	3.59	3.49
Variance explained (first component)	89.8%	87.3%
Eigenvector values (first component)		
% Hispanic/Latino (Hispanic or Latino count/total population)	0.50	0.49

% Foreign-Born Hispanic/Latino (Hispanic/Latino who are foreign-born/total population)	0.50	0.51
% Limited English: Spanish (People who speak English "not well" or "not at all" and speak Spanish/total population)	0.51	0.51
% Linguistically Isolated: Spanish (Households that are linguistically isolated and speak Spanish/total households)	0.50	0.49

Table 3. Percent of all census tracts classified by pooled Asian American and Hispanic/Latino ethnic enclave indices, by state, 2000 and 2010 census tracts.

	CA		FL		NJ		NY		TX	
	2000	2010	2000	2010	2000	2010	2000	2010	2000	2010
Asian American enclave										
Quintiles	Percent of all tracts in the state[1]									
Q1 (low enclave):	7.42	7.19	26.26	28.86	15.38	14.33	21.97	19.61	35.55	35.08
Q2	11.22	11.74	32.21	31.43	20.30	19.89	21.43	20.99	23.62	22.63
Q3	17.03	17.98	27.00	22.83	23.20	24.55	20.12	21.37	18.20	17.82
Q4	26.25	25.82	13.20	14.35	25.74	23.20	18.02	19.32	14.51	15.02
Q5 (high enclave)	38.08	37.26	1.34	2.52	15.38	18.04	18.46	18.72	8.13	9.46
Dichotomous										
Enclave	46.05	48.40	1.78	3.85	24.08	27.25	18.35	23.19	10.25	13.62
Non-enclave	53.95	51.60	98.22	96.15	75.92	72.75	81.65	76.81	89.75	86.38
Hispanic/Latino enclave										
Quintiles										
Q1 (low enclave)	8.37	10.32	25.28	24.78	33.61	30.01	38.56	39.05	8.45	9.53
Q2	18.14	18.49	26.04	23.07	25.27	26.00	19.12	19.92	17.28	17.64
Q3	20.93	20.27	21.69	21.27	17.24	17.64	15.56	16.48	23.39	22.73
Q4	23.61	23.13	16.28	18.00	12.69	14.88	14.94	13.70	25.68	24.61
Q5 (high enclave)	28.94	27.78	10.72	12.88	11.19	11.47	11.82	10.86	25.20	25.49
Dichotomous										
Enclave	44.28	42.42	18.35	20.84	18.07	18.74	18.27	16.97	39.30	38.65
Non-enclave	55.72	57.58	81.65	79.16	81.93	81.26	81.73	83.03	60.70	61.35

1. Tracts with missing data due to low/zero populations were excluded

Table 4. Median values for variables included in the PCA by pooled Asian American enclave index quintiles, 2000 and 2010 census tracts.

	Pooled data		**CA**		**FL**		**NJ**		**NY**		**TX**	
	2000	2010	2000	2010	2000	2010	2000	2010	2000	2010	2000	2010
Asian American Enclave quintiles												
Proportion of Asian American residents												
Q1 (low enclave)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Q2	0.008	0.010	0.011	0.013	0.008	0.011	0.009	0.011	0.008	0.009	0.007	0.009
Q3	0.023	0.033	0.028	0.037	0.019	0.030	0.026	0.036	0.024	0.032	0.019	0.029
Q4	0.049	0.065	0.056	0.073	0.036	0.047	0.054	0.073	0.047	0.064	0.042	0.056
Q5 (high enclave)	0.161	0.197	0.173	0.210	0.074	0.102	0.152	0.211	0.161	0.212	0.115	0.145
Proportion of foreign-born Asian American residents												
Q1 (low enclave)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Q2	0.005	0.007	0.005	0.007	0.006	0.007	0.006	0.008	0.006	0.007	0.005	0.006
Q3	0.015	0.022	0.016	0.021	0.014	0.022	0.018	0.025	0.017	0.022	0.013	0.020
Q4	0.033	0.042	0.034	0.043	0.028	0.033	0.038	0.049	0.034	0.044	0.031	0.039
Q5 (high enclave)	0.111	0.131	0.115	0.134	0.057	0.075	0.116	0.151	0.121	0.147	0.089	0.106

Proportion of Limited English-speaking residents who speak API languages												
Q1 (low enclave)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Q2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Q3	0.002	0.000	0.002	0.001	0.002	0.000	0.002	0.000	0.002	0.000	0.002	0.001
Q4	0.006	0.007	0.006	0.007	0.006	0.008	0.005	0.006	0.006	0.007	0.006	0.008
Q5 (high enclave)	0.026	0.030	0.028	0.032	0.020	0.027	0.015	0.019	0.031	0.039	0.019	0.024
Proportion of Linguistically Isolated households who speak API languages												
Q1 (low enclave)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Q2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Q3	0.002	0.000	0.001	0.000	0.003	0.000	0.000	0.000	0.001	0.000	0.003	0.000
Q4	0.009	0.010	0.009	0.010	0.009	0.010	0.007	0.009	0.009	0.010	0.010	0.011
Q5 (high enclave)	0.036	0.042	0.039	0.047	0.027	0.030	0.023	0.028	0.036	0.044	0.029	0.033
Asian American enclave (dichotomous)												
Proportion of Asian American residents												
Enclave	0.136	0.161	0.144	0.170	0.066	0.096	0.116	0.164	0.159	0.185	0.103	0.124
Non-enclave	0.012	0.015	0.026	0.028	0.010	0.012	0.016	0.021	0.012	0.015	0.006	0.006
Proportion of foreign-born Asian American residents												
Enclave	0.093	0.106	0.094	0.107	0.051	0.066	0.085	0.112	0.120	0.128	0.077	0.088
Non-enclave	0.008	0.010	0.015	0.016	0.007	0.008	0.012	0.015	0.008	0.010	0.004	0.004
Proportion of Limited English-speaking residents who speak API languages												
Enclave	0.019	0.021	0.022	0.024	0.009	0.014	0.010	0.013	0.025	0.024	0.014	0.015
Non-enclave	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Proportion of Linguistically Isolated households who speak API languages												
Enclave	0.027	0.028	0.030	0.032	0.015	0.018	0.015	0.019	0.032	0.029	0.022	0.023
Non-enclave	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 5. Median values for variables included in the PCA by pooled Hispanic/Latino enclave index quintiles, 2000 and 2010 census tracts.

	Pooled data		**CA**		**FL**		**NJ**		**NY**		**TX**	
	2000	2010	2000	2010	2000	2010	2000	2010	2000	2010	2000	2010
Hispanic/Latino enclave quintiles												
Proportion of Hispanic/Latino residents												
Q1 (low enclave)	0.022	0.035	0.048	0.063	0.021	0.030	0.023	0.035	0.015	0.024	0.033	0.059
Q2	0.061	0.095	0.080	0.119	0.045	0.076	0.047	0.075	0.057	0.083	0.067	0.113
Q3	0.118	0.174	0.139	0.206	0.096	0.140	0.095	0.145	0.105	0.147	0.125	0.196
Q4	0.262	0.339	0.282	0.372	0.216	0.282	0.221	0.272	0.239	0.288	0.282	0.361
Q5 (high enclave)	0.638	0.713	0.635	0.718	0.615	0.674	0.539	0.628	0.592	0.632	0.742	0.784
Proportion of foreign-born Hispanic/Latino residents												
Q1 (low enclave)	0.003	0.004	0.007	0.008	0.004	0.005	0.004	0.005	0.001	0.002	0.004	0.005
Q2	0.016	0.022	0.018	0.025	0.014	0.021	0.014	0.021	0.016	0.021	0.013	0.020
Q3	0.036	0.051	0.040	0.057	0.036	0.050	0.035	0.052	0.035	0.048	0.030	0.045
Q4	0.096	0.120	0.109	0.130	0.102	0.125	0.090	0.126	0.090	0.117	0.079	0.105
Q5 (high enclave)	0.283	0.298	0.315	0.313	0.400	0.395	0.263	0.309	0.241	0.278	0.241	0.265
Proportion of Limited English-speaking residents who speak Spanish												
Q1 (low enclave)	0.001	0.000	0.001	0.000	0.001	0.000	0.002	0.000	0.001	0.000	0.002	0.000
Q2	0.006	0.006	0.005	0.005	0.007	0.006	0.006	0.006	0.006	0.006	0.006	0.006
Q3	0.017	0.021	0.015	0.020	0.018	0.022	0.018	0.022	0.017	0.021	0.018	0.019
Q4	0.051	0.062	0.052	0.061	0.049	0.061	0.052	0.066	0.053	0.062	0.049	0.061
Q5 (high enclave)	0.177	0.193	0.188	0.196	0.183	0.195	0.167	0.206	0.158	0.185	0.167	0.189
Proportion of Linguistically Isolated households who speak Spanish												
Q1 (low enclave)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Q2	0.005	0.005	0.004	0.000	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.004
Q3	0.016	0.020	0.013	0.016	0.018	0.025	0.017	0.023	0.018	0.025	0.017	0.021
Q4	0.047	0.058	0.041	0.049	0.050	0.064	0.050	0.067	0.054	0.072	0.050	0.059
Q5 (high enclave)	0.168	0.186	0.159	0.167	0.188	0.214	0.174	0.221	0.185	0.198	0.169	0.193
Hispanic/Latino enclave (dichotomous)												
Proportion of Hispanic/Latino residents												

Enclave	0.497	0.591	0.508	0.605	0.394	0.493	0.408	0.502	0.467	0.541	0.556	0.652
Non-enclave	0.073	0.112	0.113	0.160	0.051	0.086	0.044	0.073	0.047	0.066	0.106	0.163
Proportion of foreign-born Hispanic/Latino residents												
Enclave	0.202	0.233	0.236	0.248	0.211	0.256	0.184	0.215	0.177	0.215	0.171	0.206
Non-enclave	0.019	0.027	0.029	0.039	0.016	0.025	0.012	0.019	0.010	0.014	0.023	0.034
Proportion of Limited English-speaking residents who speak Spanish												
Enclave	0.122	0.140	0.128	0.144	0.112	0.129	0.121	0.149	0.118	0.137	0.117	0.133
Non-enclave	0.008	0.008	0.010	0.010	0.007	0.008	0.006	0.006	0.005	0.004	0.013	0.013
Proportion of Linguistically Isolated households who speak Spanish												
Enclave	0.113	0.131	0.104	0.117	0.109	0.145	0.120	0.146	0.129	0.148	0.117	0.139
Non-enclave	0.007	0.007	0.008	0.008	0.008	0.008	0.005	0.005	0.004	0.000	0.012	0.013

The 2010 enclave indices were derived using survey data from the American Community Survey with smaller sample sizes than Census long form (a questionnaire on household characteristics that was collected on a subsample of the population in decennial census years and discontinued in 2010), and thus have greater potential for error; we used 5-year estimates to reduce the potential for error.[17] To measure the amount of error, we calculated the coefficient of variation (COV) for each variable and show the results in Appendix E by enclave status. The coefficient of variation represents the error for each variable; variables with COV of 40% or higher are considered unreliable. While those census tracts classified as non-enclaves have higher COVs for each of the variables, census tracts classified as enclaves have lower COVs indicating more reliable data. This reflects the fact that enclaves include larger populations of these groups (e.g., Asian American residents) compared to non-enclave census tracts.

IV. Considerations, Caveats, and Best Practices

For multistate analyses using all 5 states, we recommend using the pooled data (CA, FL, NJ, NY, TX). However, for single state studies, we recommend using the state-specific measures. In the future, the ENCLAVE team may create enclave measures for other states; in the meantime, studies using data from other states or for more recent years may apply this methodology to create their own measures.

The quintiles will be useful for projects in which the investigators want to consider how residence in more culturally distinct neighborhoods may impact health outcomes; however, the dichotomous variable will be more useful in comparing the impact of residence in enclave vs. non-enclave neighborhoods.

Limitations. The enclave measures reported here face several limitations. For example, these measures are relative to the distribution in all census tracts in all 5 states included in our study (pooled measures) or are relative to the distribution in all census tracts within individual states (state-specific measures). Developing a single measure to reflect the multidimensional nature of ethnic enclaves is a challenging task and no single measure can fully capture the complexity of diverse neighborhood environments. The measures reported here only take into account population density of co-ethnic residents, foreign-born status, and linguistic factors; they do not account for other aspects of enclaves such as businesses and cultural centers, or specific birthplace country or recency of immigration or other neighborhood social, physical and built environment features of ethnic enclaves.

Access to Data: Calculate the ethnic enclave measures. Investigators interested in applying these measures of ethnic enclaves are welcome to apply the methodology described in the report to U.S. Census/American Community Survey data. Please cite this ENCLAVE report when the measures are used in published manuscripts, grants, and reports:

Citation: Shariff-Marco S, Lin K, Meltzer D, Allen L, Boscoe F, Canchola AJ, Gates Kuliszewski M, German S, Guan A, Harris G, Henry K, Hiatt RA, Hughes AE, McGuire V, Oh D, Paddock LE, Pinheiro PS, Radadiya HR, Reyes S, Stroup A, Zhu H, Gomez SL, Pruitt SL. Developing measures of Asian American and Hispanic/Latino ethnic enclave for five states using U.S. Census and American Community Survey data. University of California San Francisco, San Francisco, CA; October 2021. Online available at: https://cancerregistry.ucsf.edu/resources/research-tools.

Access to Data: Obtain computed ethnic enclave measures. Investigators interested in using our computed measures of ethnic enclaves in other studies may follow these steps to obtain the measures described in this report:

1. Provide the following information to Drs. Salma Shariff-Marco at salma.shariff-marco@ucsf.edu and Sandi Pruitt at Sandi.Pruitt@UTSouthwestern.edu for approval: Prepare a 1-2 page summary that includes PI contact information & study team, timeline, and funding; and brief description of proposed analyses that includes: background, purpose/overall aims, methods, and impact/dissemination.
2. Once approved, cite the funding source in abstracts, presentations, reports, and manuscripts: National Cancer Institute (NCI) R01 CA23750 (MPI: Shariff-Marco, Pruitt)
3. Once approved, please cite this ENCLAVE report when the measures are used in published manuscripts, grants, and reports:

   Citation: Shariff-Marco S, Lin K, Meltzer D, Allen L, Boscoe F, Canchola AJ, Gates Kuliszewski M, German S, Guan A, Harris G, Henry K, Hiatt RA, Hughes AE, McGuire V, Oh D, Paddock LE, Pinheiro PS, Radadiya HR, Reyes S, Stroup A, Zhu H, Gomez SL, Pruitt SL. Developing measures of Asian American and Hispanic/Latino ethnic enclave for five states using U.S. Census and American Community Survey data. University of California San Francisco, San Francisco, CA; October 2021. Online available at: https://cancerregistry.ucsf.edu/resources/research-tools.

Appendix A: Prior indicator variables considered in ethnic enclave indices

Ethnic enclave measures	Indicator variables
Asian American ethnic enclave	% recent immigrants % of API language-speaking households that are linguistically isolated % of API language-speakers with limited English proficiency % API
Hispanic/Latino ethnic enclave	% foreign-born % recent immigrants % households that are linguistically isolated % of Spanish language-speaking households that are linguistically isolated % of all language speakers with limited English proficiency % of Spanish language-speakers with limited English proficiency % Hispanic/Latino

Appendix B. Distribution of missing/excluded census tracts in PCA for enclave indices by state

(Census tracts with small numerator counts were suppressed by Census and those with zero population or household denominator counts were excluded).

State	Year	Total census tracts in state	Tracts with enclave indices	Tracts missing enclave indices
CA	2000	7049	7017	32
	2010	8057	7980	77
FL	2000	3154	3145	9
	2010	4245	4161	84
NJ	2000	1950	1931	19
	2010	2010	1996	14
NY	2000	4907	4806	101
	2010	4919	4825	94
TX	2000	4388	4369	19
	2010	5265	5214	51

Appendix C: State-specific ethnic enclave indices

Methods for State-Specific Ethnic Enclave Measures

The ENCLAVE study was designed to examine cancer outcomes among residents of all 5 states combined. As other investigators may be interested in measuring state-specific ethnic enclaves, we also calculated state-specific ethnic enclave quintiles and dichotomous measures. These analyses entailed repeating the same methods used in the pooled analysis, separately for each of the 5 states.

Results for State-Specific Ethnic Enclave Measures

Appendix Table C1 illustrates distributions using the state-specific enclave indices. Additional information about the state-specific indices can be located here: https://cancerregistry.ucsf.edu/enclave.

Additionally, we compared concordance/discordance between the pooled and state-specific measures. Appendix Table C2 and C3 shows the impact of using pooled- vs. state-specific measures. We reported the % concordance/discordance of pooled vs. state-specific measures of enclave status (dichotomous) separately by state, using only 2010 data.

For the Asian American enclave index in 2010 overall, we observed that 91.5% of census tracts are concordant and 8.5% are discordant, with CA having the highest proportion of discordant pairs. For the Hispanic/Latino enclave index in 2010 overall, we observed that 89.7% of census tracts are concordant and 10.3% are discordant, with NJ and NY having the highest proportion of discordant pairs.

Table C1. PCA Results for state-specific ethnic enclave indices, 2000 and 2010 census tracts.

	CA		**FL**		**NJ**		**NY**		**TX**	
	2000	2010	2000	2010	2000	2010	2000	2010	2000	2010
Asian American Enclave										
Eigenvalue (1st component)	3.54	3.40	2.92	2.78	3.29	3.14	3.42	3.31	3.34	3.22
Variance explained (1st component)	88.5%	85.0%	73.0%	69.5%	82.3%	78.4%	85.4%	82.9%	83.6%	80.6%
% Asian American (Eigenvector)	0.51	0.51	0.53	0.54	0.51	0.52	0.51	0.51	0.51	0.52
% Foreign-Born Asian American (Eigenvector)	0.51	0.52	0.54	0.55	0.52	0.53	0.51	0.52	0.52	0.52
% Limited English: API languages (Eigenvector)	0.50	0.49	0.47	0.46	0.48	0.47	0.49	0.49	0.48	0.48
% Linguistically Isolated: API languages (Eigenvector)	0.49	0.48	0.46	0.44	0.48	0.48	0.48	0.48	0.48	0.48
Hispanic/Latino enclave										
Eigenvalue (1st component)	3.65	3.55	3.60	3.46	3.52	3.46	3.51	3.43	3.54	3.42
Variance explained (1st component)	91.2%	88.7%	90.0%	86.6%	88.0%	86.6%	87.8%	85.8%	88.6%	85.5%
% Hispanic/Latino (Eigenvector)	0.49	0.49	0.50	0.50	0.50	0.50	0.50	0.50	0.49	0.48

% Foreign-Born Hispanic/Latino (Eigenvector)	0.51	0.51	0.50	0.51	0.50	0.50	0.50	0.50	0.50	0.50
% Limited English: Spanish (Eigenvector)	0.51	0.51	0.50	0.50	0.50	0.51	0.50	0.51	0.51	0.51
% Linguistically Isolated: Spanish (Eigenvector)	0.49	0.49	0.49	0.49	0.50	0.49	0.49	0.49	0.50	0.50

Table C2. Concordance/discordance for Asian American enclave measures using pooled & state-specific dichotomous measures in 2010

	Defined as enclave (dichotomous measure in 2010) using POOLED data	Defined as enclave (dichotomous measure in 2010) using STATE-SPECIFIC data	Concordant: Pooled enclave =yes & state-specific enclave =yes	Discordant: Pooled enclave =yes & state-specific enclave =no	Discordant: Pooled enclave =no & state-specific enclave =yes	Concordant: Pooled enclave =no & state-specific enclave =no
	N (%)[a]	N (%)[a]	N (Row %)	N (Row %)	N (Row %)	N (Row %)
All states	6395 (26.45)	5634 (23.30)	4989 (20.64)	1406 (5.82)	645 (2.67)	17136 (70.88)
CA	3862 (48.40)	2463 (30.86)	2463 (30.86)	1399 (17.53)	0 (0.00)	4118 (51.60)
FL	160 (3.85)	485 (11.66)	160 (3.85)	0 (0.00)	325 (7.81)	3676 (88.34)
NJ	544 (27.25)	548 (27.45)	537 (26.90)	7 (0.35)	11 (0.55)	1441 (72.19)
NY	1119 (23.19)	1164 (24.12)	1119 (23.19)	0 (0.00)	45 (0.93)	3661 (75.88)
TX	710 (13.62)	974 (18.68)	710 (13.62)	0 (0.00)	264 (5.06)	4240 (81.32)

[a] N is number of census tracts. % is percent of total (non-missing) census tracts in that state.

Table C3. Concordance/discordance for Hispanic/Latino enclave measures using pooled & state-specific dichotomous measures in 2010

	Defined as enclave (dichotomous measure in 2010) using POOLED data	Defined as enclave (dichotomous measure in 2010) using STATE-SPECIFIC data	Concordant: Pooled enclave =yes & state-specific enclave =yes	Discordant: Pooled enclave =yes & state-specific enclave =no	Discordant: Pooled enclave =no & state-specific enclave =yes	Concordant: Pooled enclave =no & state-specific enclave =no
	N (%)[a]	N (%)[a]	N (Row %)	N (Row %)	N (Row %)	N (Row %)
All states	7460 (30.86)	7569 (31.31)	6266 (25.92)	1194 (4.94)	1303 (5.39)	15413 (63.75)
CA	3385 (42.42)	2590 (32.46)	2590 (32.46)	795 (9.96)	0 (0.00)	4595 (57.58)
FL	867 (20.84)	1264 (30.38)	867 (20.84)	0 (0.00)	397 (9.54)	2897 (69.62)
NJ	374 (18.74)	624 (31.26)	374 (18.74)	0 (0.00)	250 (12.53)	1372 (68.74)
NY	819 (16.97)	1475 (30.57)	819 (16.97)	0 (0.00)	656 (13.60)	3350 (69.43)
TX	2015 (38.65)	1616 (30.99)	1616 (30.99)	399 (7.65)	0 (0.00)	3199 (61.35)

[a] N is number of census tracts. % is percent of total (non-missing) census tracts in that state.

Appendix D. Maps of Asian American and Hispanic/Latino enclaves in select metropolitan areas in CA, FL, NJ, NY, and TX, 2010 census tracts (See https://cancerregistry.ucsf.edu/enclave for interactive maps)

Figure 1. Asian American enclaves in select metropolitan areas in CA, FL, NJ, NY, and TX, 2010 census tracts

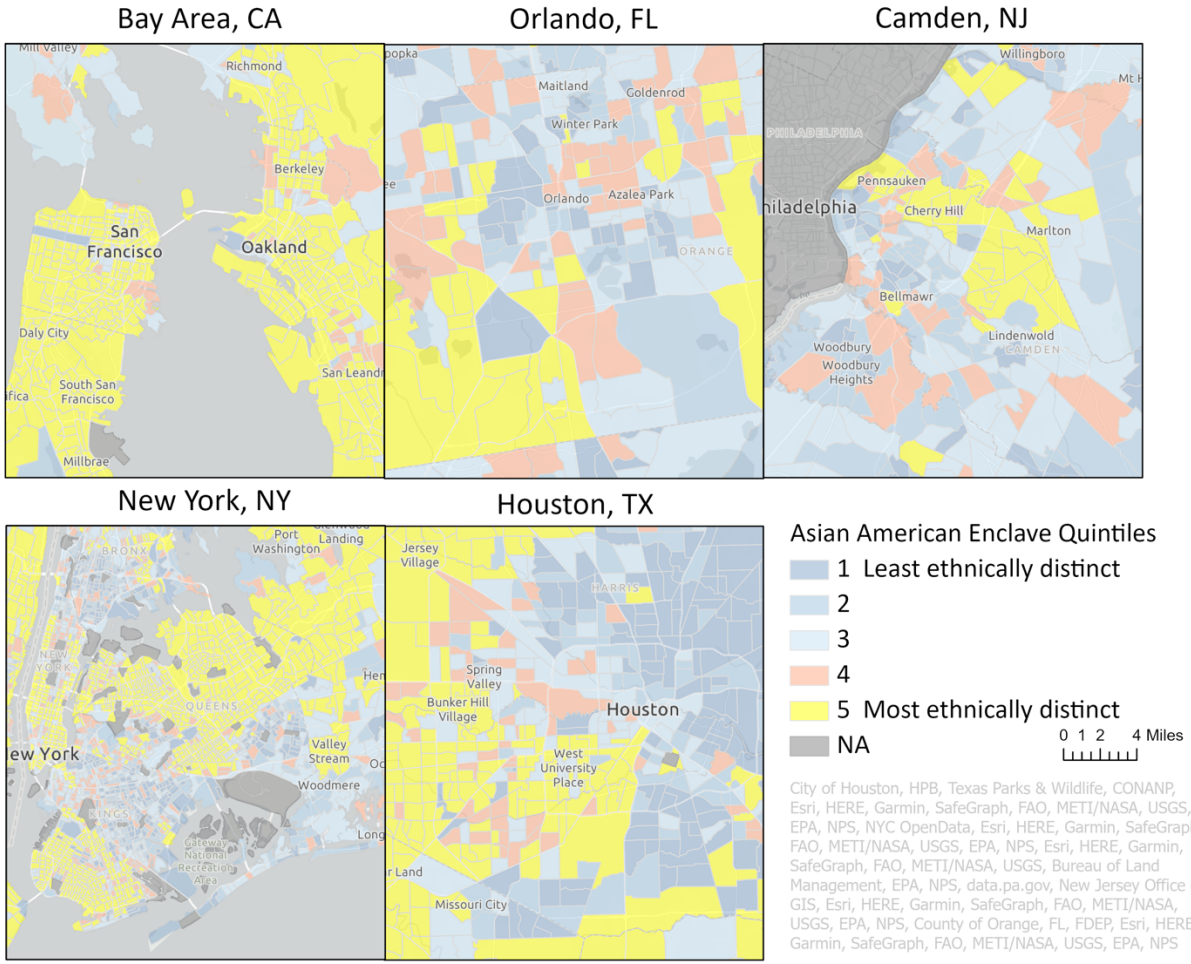
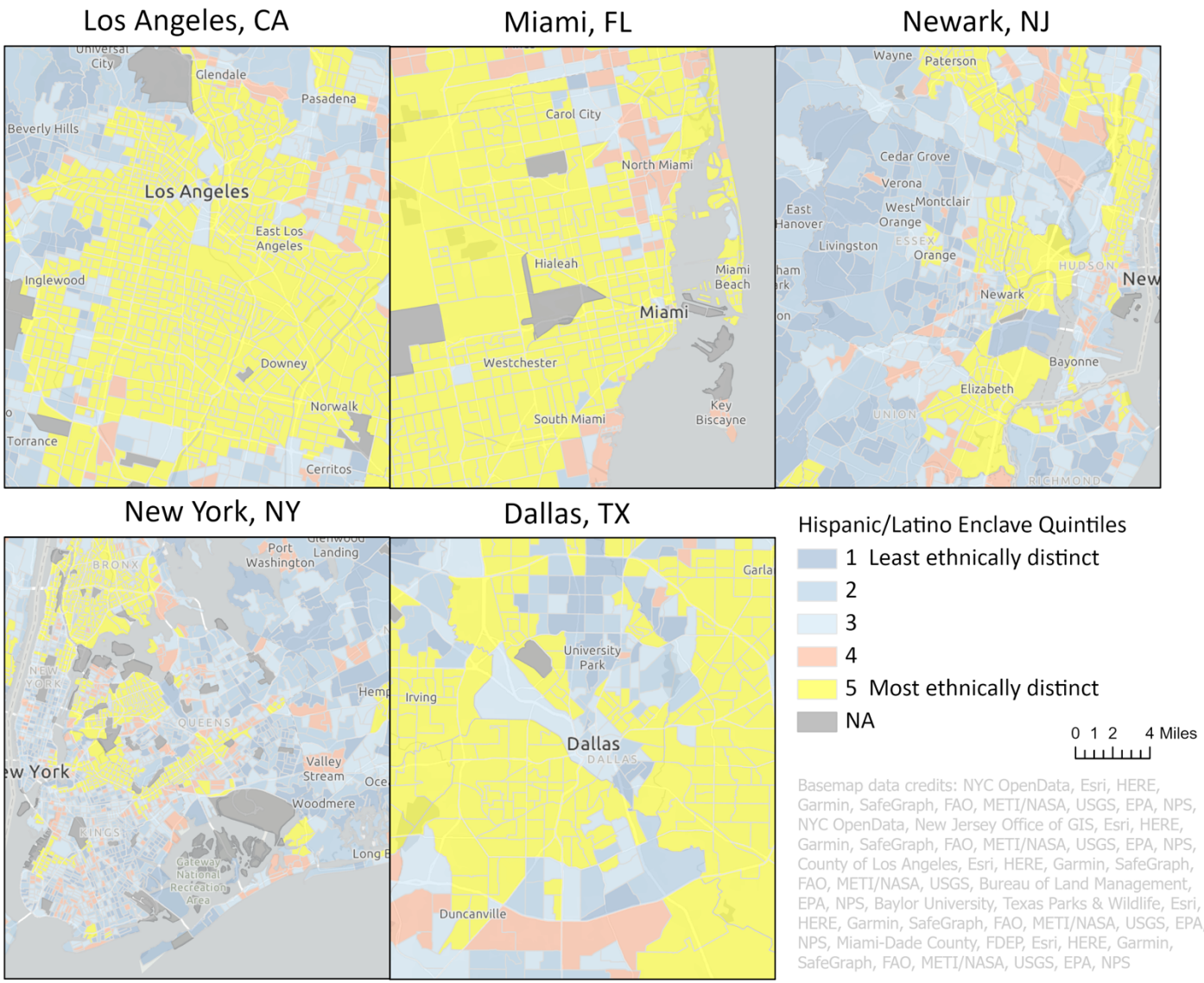


Figure 2. Hispanic/Latino enclaves in select metropolitan areas in CA, FL, NJ, NY, and TX, 2010 census tracts



Appendix E.

Table E1. Coefficient of variation for variables included in the PCA of the Asian American enclave measure by enclave status, American Community Survey 2008-2012.

	Enclave					Non-enclave				
	Total tracts	Coefficient of variation				Total tracts	Coefficient of variation			
		N/A, Est=0	0% - <12%	12% - <40%	40%+		N/A, Est=0	0% - <12%	12% - <40%	40%+
	N	%	%	%	%	N	%	%	%	%
All states										
% Asian American	6395	0	22.3	72.5	5.3	17781	20.3	0.1	15.9	63.7
% Foreign-Born Asian American	6395	0	15.0	76.4	8.6	17781	23.8	0.0	12.2	64.0
% Limited English: API languages	6395	4.0	1.4	34.6	60.0	17781	65.4	0.0	0.3	34.3
% Linguistically Isolated: API languages	6395	6.2	0.6	28.5	64.7	17781	70.7	0	0.3	29.0
CA										
% Asian American	3862	0	27.6	68.8	3.6	4118	8.1	0.2	29.7	62.0
% Foreign-Born Asian American	3862	0	17.9	74.8	7.4	4118	12.2	0	20.1	67.7
% Limited English: API languages	3862	2.7	1.6	39.7	56.0	4118	49.6	0	0.6	49.8
% Linguistically Isolated: API languages	3862	5.6	0.6	32.9	60.9	4118	59.2	0	0.8	40.0
FL										
% Asian American	160	0	1.3	72.5	26.3	4001	24.7	0.0	8.1	67.2
% Foreign-Born Asian American	160	0	1.3	63.1	35.6	4001	27.7	0	7.0	65.3
% Limited English: API languages	160	10.6	0	10.6	78.8	4001	74.3	0	0.1	25.6
% Linguistically Isolated: API languages	160	7.5	0	4.4	88.1	4001	76.5	0	0.0	23.5
NJ										
% Asian American	544	0	21.3	73.3	5.3	1452	14.2	0.1	20.1	65.6
% Foreign-Born Asian American	544	0	16.9	75.4	7.7	1452	17.6	0	17.7	64.7
% Limited English: API languages	544	7.4	0	17.5	75.2	1452	64.8	0	0	35.2
% Linguistically Isolated: API languages	544	6.3	0	16.7	77.0	1452	70.7	0	0.1	29.2
NY										
% Asian American	1119	0	13.3	80.3	6.4	3706	18.5	0.1	14.7	66.6
% Foreign-Born Asian American	1119	0	9.4	81.9	8.7	3706	21.4	0.1	12.3	66.2
% Limited English: API languages	1119	5.2	2.8	35.6	56.5	3706	64.7	0.0	0.4	34.9
% Linguistically Isolated: API languages	1119	7.7	1.2	28.3	62.8	3706	70.5	0	0.2	29.2
TX										
% Asian American	710	0	12.5	79.4	8.0	4504	31.0	0.1	9.9	59.0
% Foreign-Born Asian American	710	0	9.9	80.0	10.1	4504	34.7	0.0	7.8	57.4

% Limited English: API languages	710	4.6	0	23.9	71.4	4504	72.8	0	0.3	26.9
% Linguistically Isolated: API languages	710	7.2	0	19.4	73.4	4504	76.1	0	0.2	23.7

Table E2. Coefficient of variation for variables included in the PCA of the Hispanic/Latino enclave measure by enclave status, American Community Survey 2008-2012.

	Enclave					Non-enclave				
	Total tracts	Coefficient of variation				Total tracts	Coefficient of variation			
		N/A, Est=0	0% - <12%	12% - <40%	40%+		N/A, Est=0	0% - <12%	12% - <40%	40%+
	N	%	%	%	%	N	%	%	%	%
All states										
% Hispanic/Latino	7460	0	69.7	30.0	0.3	16716	1.1	6.2	64.5	28.2
% Foreign-Born Hispanic/Latino	7460	0	26.7	69.9	3.3	16716	8.3	0.1	32.5	59.0
% Limited English: Spanish	7460	0	10.9	80.2	8.9	16716	23.0	0.0	12.7	64.3
% Linguistically Isolated: Spanish	7460	0.1	2.6	71.5	25.8	16716	38.8	0.0	5.1	56.0
CA										
% Hispanic/Latino	3385	0	75.5	24.4	0.1	4595	0.3	9.2	79.1	11.4
% Foreign-Born Hispanic/Latino	3385	0	32.6	65.8	1.5	4595	2.4	0.2	46.5	50.9
% Limited English: Spanish	3385	0	13.1	79.6	7.4	4595	18.0	0.0	15.5	66.5
% Linguistically Isolated: Spanish	3385	0.1	1.5	68.7	29.7	4595	36.0	0	4.0	60.0
FL										
% Hispanic/Latino	867	0	60.1	39.2	0.7	3294	1.7	3.3	55.3	39.7
% Foreign-Born Hispanic/Latino	867	0	34.9	60.4	4.6	3294	9.3	0.2	27.4	63.1
% Limited English: Spanish	867	0	11.6	74.6	13.7	3294	24.3	0.0	10.5	65.1
% Linguistically Isolated: Spanish	867	0	8.2	69.2	22.6	3294	36.9	0.1	5.6	57.4
NJ										
% Hispanic/Latino	374	0	51.9	47.9	0.3	1622	0.9	1.0	60.9	37.2
% Foreign-Born Hispanic/Latino	374	0	19.3	75.7	5.1	1622	8.6	0	24.5	67.0
% Limited English: Spanish	374	0	7.8	83.2	9.1	1622	26.8	0	8.3	64.9
% Linguistically Isolated: Spanish	374	0	3.2	74.9	21.9	1622	45.1	0	4.3	50.7
NY										
% Hispanic/Latino	819	0	54.3	44.8	0.9	4006	2.2	1.3	50.3	46.2
% Foreign-Born Hispanic/Latino	819	0	11.0	84.0	5.0	4006	18.0	0.0	18.6	63.4
% Limited English: Spanish	819	0	4.9	83.3	11.8	4006	33.0	0.0	7.5	59.5
% Linguistically Isolated: Spanish	819	0	2.2	72.8	25.0	4006	50.7	0.0	4.3	44.9
TX										
% Hispanic/Latino	2015	0	73.7	26.0	0.2	3199	0.3	13.7	72.6	13.4

% Foreign-Born Hispanic/Latino	2015	0	21.1	74.1	4.8	3199	3.7	0.2	39.1	57.0
% Limited English: Spanish	2015	0	9.9	81.9	8.2	3199	14.2	0	19.6	66.2
% Linguistically Isolated: Spanish	2015	0.1	2.0	76.1	21.8	3199	26.9	0	7.6	65.5

References:

1. U.S. Hispanic population growth has leveled off. http://www.pewresearch.org/fact-tank/2017/08/03/u-s-hispanic-population-growth-has-leveled-off/ Available at: Accessed on 5-14-2018. Accessed.
2. Brown A. U.S. Hispanic and Asian populations growing, but for different reasons. Pew Research Center. June 16, 2014. http://pewrsr.ch/1nKqvSO In.
3. Center PR. *The rise of Asian Americans: Technical Report.* Washington, D.C.: Pew Research Center;2012.
4. Colby SL, Ortman JM. *Projections of the size and composition of the U.S. population: 2014 to 2060. Population estimates and projectsions. Current Population Reports, P25-1143. March 2015. U.S. Department of Commerce Economics and Statistics Administration, U.S. Census Bureau, Washington DC.*
5. Gomez SL, Shariff-Marco S, DeRouen M, et al. The impact of neighborhood social and built environment factors across the cancer continuum: Current research, methodological considerations, and future directions. *Cancer.* 2015;121(14):2314-2330.
6. Fang CY, Tseng M. Ethnic density and cancer: A review of the evidence. *Cancer.* 2018;124(9):1877-1903.
7. Gomez SL, Glaser SL, McClure LA, et al. The California Neighborhoods Data System: a new resource for examining the impact of neighborhood characteristics on cancer incidence and outcomes in populations. *Cancer Causes Control.* 2011;22(4):631-647.
8. Logan JR, Alba RD, Zhang WQ. Immigrant enclaves and ethnic communities in New York and Los Angeles. *Am Sociol Rev.* 2002;67(2):299-322.
9. Lim S, Yi SS, Lundy De La Cruz N, Trinh-Shevrin C. Defining Ethnic Enclave and Its Associations with Self-Reported Health Outcomes Among Asian American Adults in New York City. *Journal of immigrant and minority health.* 2017;19(1):138-146.
10. Roux AVD, Mair C. Neighborhoods and health. *Ann Ny Acad Sci.* 2010;1186:125-145.
11. Yen IH, Michael YL, Perdue L. Neighborhood Environment in Studies of Health of Older Adults A Systematic Review. *Am J Prev Med.* 2009;37(5):455-463.
12. Meijer M, Rohl J, Bloomfield K, Grittner U. Do neighborhoods affect individual mortality? A systematic review and meta-analysis of multilevel studies. *Soc Sci Med.* 2012;74(8):1204-1212.
13. Krieger N. Theories for social epidemiology in the 21st century: an ecosocial perspective. *International journal of epidemiology.* 2001;30(4):668-677.
14. Gallo LC, Penedo FJ, de los Monteros KE, Arguelles W. Resiliency in the Face of Disadvantage: Do Hispanic Cultural Characteristics Protect Health Outcomes? *J Pers.* 2009;77(6):1707-1746.
15. Williams DR, Mohammed SA. Racism and Health I: Pathways and Scientific Evidence. *Am Behav Sci.* 2013;57(8).
16. Hatcher L. *A step-by-step approach to using the SAS system for factor analysis and structural equation modeling.* Cary, NC: SAS Institute; 1994.
17. Spielman SE, Filch D, Nagle N. Patterns of causes of uncertainty in the American Community Survey. *Applied Geography.* 2014;46:147-157.