**Results Notes**

1. **Pairwise correlations:**

It looks like there are strong positive correlations between renter-occupied housing with below 150% poverty and with no vehicle in the household; below 150% poverty and SNAP benefits; higher median income with a bachelor's degree; limited English proficiency with less than high school education and with Hispanic.   
  
Also, there is a moderately strong positive correlation between female households, those less than high school education, those below 150% poverty, those with SNAP benefits, and Hispanics.  
  
Conversely, the lack of complete plumbing and NH Asian do not seem to be strongly correlated with any variables, perhaps due to a smaller sample size.

1. **Justification for reclassification of census tracts (cluster pruning)**

We performed additional post-processing steps to reclassify census tracts that were part of small clusters (defined as a cluster that had at least 5% of the total number of census tracts). This was done to obtain meaningful/ interpretable clusters. This reclassification was done based on their second-highest posterior probability since it wasn’t significantly higher.

Also, as shown in the boxplot, some cluster assignment probabilities are low, which might indicate there is:

1. uncertainty about the assignment
2. Too many clusters might spread the data thinly across them, leading to these low assignment probabilities
3. **Posterior probability Boxplot**

* Posterior assignment probabilities, in general, seem moderately high, with a median probability above 0.8 for all clusters; however, most clusters have many outliers with lower assignment probability.
* Clusters 1 and 3, which are incidentally also the largest clusters, have the highest assignment probabilities, whereas clusters 2 and 7 have the most spread and heterogeneity in assignment probabilities among census tracts. Clusters 1 and 3 also correspond to the affluent and at-risk clusters, respectively, indicating that the differences between these two clusters may be the most pronounced

1. **Barplot summary of my inputs and Stephanie’s help on labeling**

**(OLD)**

Figure xx shows the the estimated probability of high exposure to NSDoH variables for a census tract given its assignment to a NSDoH profile/cluster. Cluster 1 was labeled as **“affluent, high education, home ownership”** census tracts based on county composition (add citation: [HD*Pulse*: An Ecosystem of Minority Health and Health Disparities Resources. National Institute on Minority Health and Health Disparities. Created 8/12/2024. Available from [https://hdpulse.nimhd.nih.gov](https://hdpulse.nimhd.nih.gov/)]) and NSDoH composition and represented 27% of MA census tracts. This cluster is characterized by a combination of probability values that indicated a high level of exposure to NSDoH variables such as owner-occupied housing, median income above the median for the state of MA, education higher than HS, and low-level exposure to working-class employment, and state assistantships programs (SNAP benefits).

Profile/cluster 2 was labeled “socioeconomic disadvantage - moderate ethnic minority” and was characterized by a high level of exposure to NSDoH variables such as less than HS education, below poverty line, state assistantships programs (SNAP benefits), renter-occupied households, working class, and low-level of exposure higher median income in the household. In terms of ethnic minorities, this profile has a medium probability of high level of exposure values for NHB and Hispanic populations above the median for the state.

Profile 3 represented the second largest cluster containing 25% of MA census tracts. We labeled this cluster as **“multi-Ethnic, socioeconomically vulnerable single female-parent and immigrant”.** This cluster had the highest probability of households with limited English proficiency, single-parent-female households, and crowding in the household. Additionally, there is a high probability of exposure to lack of plumbing, renter-occupied, no vehicle access, working class, and ethnic minority groups, especially NHB and Hispanics. This cluster is also characterized as having the lowest education levels when compared to all other NSDoH profiles. Profile 4 was labeled **“multi-ethnic, immigrant ??”**; this profile had a high probability of ethnic minorities and limited English proficiency, renter-occupied housing, no vehicle in the household, and SNAP benefits. [I agree with Stephanie’s comment about this cluster containing households of recent immigrants, given the probability of English proficiency (i.e., the households ranked their EN proficiency as “less than well”, who have different educational backgrounds, hence the possible discrepancies in these probabilities where <HS and >= bachelor’s degrees seem to be more prominent.] Profile 5 is characterized by a high probability of owner-occupied housing, < HS education, working class, ethnic minorities, and low probabilities for other household-related variables such as lack of plumbing and no vehicle. We labeled this cluster as **“multi-ethnic, working class, home ownership”**. Profile 6 had some similarities with NSDoH profile 5, except for education variables, which had slightly higher probabilities and lower probabilities for ethnic minorities and limited English proficiency, which may be an indication of a lower proportion of immigrant households residing in census tracts assigned to this profile. We labeled this cluster as **“--”.** Profile 7 had some similarities with NSDoH profile 1, given the low probabilities for ethnic minorities, and household variables except for owner-occupied housing. This cluster has high probabilities for working class and very low probability for limited English proficiency. Based on these characteristics we labeled this cluster as “”. Profile 8 is labeled as **“high earning/high education -multi-ethnic”.** This profile is characterized by a high probability of a high level of exposure to NSDoH variables such as renter-occupied housing, education >= HS and >= bachelor’s degree, median income, and ethnic minorities, especially NHA. Additionally, there are low probabilities for variables such as SNAP benefits, poverty line, and working class.

**We take on ideas from qualitative research analysis and conduct a “Top-down” or deductive approach to facilitate a more meaningful and systematic interpretation of the clustering analysis.** Based on the selected ACS neighborhood-level variables and previous research, we identified four (4) main thematic “codes”. (cite ADI paper and Kolak)

1. **Wealth or Financial strength** (economic health of the household; well-being)
   1. <https://www.census.gov/programs-surveys/sipp/information/sipp-survparticp-brochures/wealth-assets-ownership.html>
   2. <https://www.sciencedirect.com/topics/social-sciences/wealth>
   3. <https://www.hsph.harvard.edu/pgda/wp-content/uploads/sites/1288/2013/10/PGDA_WP_97.pdf>
   4. Proxy variables: home tenure, vehicle in the household, complete plumbing in the household, crowding in the household

In the context of our measure of NSDoH, we relate household wealth to the economic health of the household, and therefore, this thematic code is composed of the indicator variables, home tenure (owner-occupied vs. renter-occupied housing), no vehicle, lack of complete plumbing, and crowding. Household crowding, for example, is considered a marker of poverty and has been associated with adverse health outcomes. (cite: <https://www.ncbi.nlm.nih.gov/books/NBK535299/>). Additionally, studies have reported associations between homeownership and improved health outcomes, and overall economic well-being(<https://www.mdpi.com/1660-4601/14/10/1098>, <https://www.tandfonline.com/doi/full/10.1080/19371918.2015.1137518>, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7725589/>). Therefore, we rank neighborhoods within a cluster as having low wealth if the cluster is characterized by a combination of probability values that indicate a high level of exposure to NSDoH variables such as no vehicle, renter-occupied housing, lack of complete plumbing, and crowded household.

1. Economic Security (Monetary indicators or possible indicators of economic hardship or inequality like the ADI-3—I wanted to include variables that could “measure” the ability of individual households to meet basic needs sustainable like food)
   1. Proxies: Median income, poverty, SNAP benefits, unemployment and working-class status

The second thematic code we use to interpret our clustering analysis is economic security. We characterized economic security through proxy indicators of possible monetary hardship or inequality, namely median family income, poverty level, food stamps/SNAP benefits, unemployment, and working-class status.

1. Educational Attainment
   1. Education variables

The third thematic code, educational attainment, is commonly used as an indicator of socioeconomic status (cite), and in the context of NSDoH cluster profile characterization, it could fall under any of the thematic above; however, education can influence outcomes beyond SES, and thus we examine as a separate attribute. Similarly, we rank NSDoH clusters as having low education attainment based on the probability values for less than HS versus >= HS, and >= Bachelor.

1. Cultural Advantage
   1. Limited EN proficiency
   2. Ethnic minorities
   3. Female household

Cultural advantage is defined as a dynamic interplay between multi-ethnic communities and the social structures that govern daily life, such as family roles. Previous literature has shown associations between single-parent households, specifically female households, socioeconomic status disadvantage, and poor health outcomes. A potential mechanism for these differences could be that, especially in neighborhoods dominated by immigrant communities, single mothers might work in occupations with low earning potential and have limited social networks to help balance responsibilities. (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7880085/>, find one more). Additionally, limited English proficiency, a key barrier to health care, alone and across multi-ethnic groups, is associated with low health literacy and poor health. (<https://www.tandfonline.com/doi/full/10.1080/10810730.2012.712621>, <https://pubmed.ncbi.nlm.nih.gov/38338249/>) We rank neighborhoods within a cluster as having low cultural advantage if the cluster is characterized by a combination of probability values that indicate a high level of exposure to NSDoH variables such as limited English proficiency, female households, and at least 2 ethnic minorities.

We then used these codes to rank clusters based on the distribution of indicators that define the codes.

**Limitations of these results**

* We observed some heterogeneity within clusters
* Low posterior assignment probabilities leading to uncertainty in cluster assignment
* Too many clusters; we do not have a large sample size (# of census tracts) and thus as the data is spread thin across so many clusters, we ended up with small clusters and clusters with higher within-cluster variability, and therefore more difficult to interpret
* Other variables to consider for future models:
  + Nativity—this along with language could help us disentangle clusters that are clearly made of census tracts with high immigrant density
  + Other language variables. The language variable currently used in the model measures language spoken at home and ability to speak English ranked for Persons 5 Years and Over for the United States (and the rank included is “Speak English less than "very well") .

**Future directions and current work**

* Add more SDoH variables: insurance, other ethnic groups, gender

**Results of optimal care analysis**