Total number of iterations: 15000

Burn-in period: 5000

Thinning: 10

Kmax = 50

Dirichlet prior with gamma = 1/Kmax

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Estimated posterior distribution of the number of clusters for each survey wave. | | | | | | | | |
| **ACS Survey Year** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** |
| 2006-2010 | 0 | 0 | 0 | **0.397** | 0.383 | 0.158 | 0.055 | 0.007 |
| 2011-2015 | 0 | 0.001 | 0.279 | **0.387** | 0.206 | 0.074 | 0.039 | 0.014 |
| 2015-2019 | 0.124 | 0.132 | **0.269** | 0.258 | 0.137 | 0.063 | 0.014 | 0.003 |

Mixing weights

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 2006-2010 | | | | 2011-2015 | | | | 2015-2019 | | | |
|  | **Mean** | **SD** | **2.50%** | **97.50%** | **Mean** | **SD** | **2.50%** | **97.50%** | **Mean** | **SD** | **2.50%** | **97.50%** |
| p.1 | 0.07 | 0.02 | 0.04 | 0.12 | 0.043 | 0.027 | 0.011 | 0.113 | 0.281 | 0.014 | 0.252 | 0.307 |
| p.2 | 0.21 | 0.03 | 0.16 | 0.26 | 0.037 | 0.046 | 0.000 | 0.126 | 0.035 | 0.017 | 0.000 | 0.063 |
| p.3 | 0.08 | 0.02 | 0.04 | 0.12 | 0.052 | 0.015 | 0.031 | 0.087 | 0.112 | 0.014 | 0.084 | 0.139 |
| p.4 | 0.11 | 0.03 | 0.07 | 0.16 | 0.254 | 0.022 | 0.207 | 0.292 | 0.248 | 0.017 | 0.214 | 0.282 |
| p.5 | 0.08 | 0.01 | 0.06 | 0.12 | 0.094 | 0.019 | 0.060 | 0.132 | 0.012 | 0.011 | 0.000 | 0.036 |
| p.6 | 0.17 | 0.02 | 0.12 | 0.21 | 0.210 | 0.036 | 0.144 | 0.274 | 0.109 | 0.016 | 0.082 | 0.145 |
| p.7 | 0.07 | 0.02 | 0.04 | 0.11 | 0.092 | 0.018 | 0.055 | 0.127 | 0.115 | 0.014 | 0.089 | 0.143 |
| p.8 | 0.14 | 0.01 | 0.11 | 0.16 | 0.123 | 0.013 | 0.100 | 0.150 | 0.087 | 0.013 | 0.062 | 0.116 |
| p.9 | 0.06 | 0.02 | 0.03 | 0.09 | 0.095 | 0.016 | 0.067 | 0.130 |  |  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  | **Assignment probability (ECR algorithm) \*** | |
|  | N (original) | N (relabeled) | Mean (SD) | Median [Min, Max] |
| **ACS 2006-2010** |  |  |  |  |
| Cluster 1 | 109 |  | 0.682 (0.162) | 0.705 [0.336, 0.958] |
| Cluster 2 | 343 |  | 0.806 (0.173) | 0.873 [0.175, 0.978] |
| Cluster 3 | 110 |  | 0.746 (0.176) | 0.773 [0.161, 0.993] |
| Cluster 4 | 137 |  | 0.755 (0.133) | 0.777 [0.472, 0.986] |
| Cluster 5 | 140 |  | 0.711 (0.194) | 0.660 [0.343, 0.987] |
| Cluster 6 | 274 |  | 0.798 (0.181) | 0.856 [0.167, 0.994] |
| Cluster 7 | 85 |  | 0.773 (0.180) | 0.828 [0.372, 0.996] |
| Cluster 8 | 201 |  | .916 (0.139 | 0.976 [0.302, 0.999] |
| Cluster 9 | 79 |  | 0.695 (0.176) | 0.724 [0.381, 0.929] |
| **ACS 2011-2015\*\*** |  |  |  |  |
| Cluster 1 | 23 | 23 | 0.588 (0.105) | 0.570 [0.425, 0.817] |
| Cluster 2 | 2 | 73 | 0.713 (0.189) | 0.689 [0.333, 0.982] |
| Cluster 3 | 73 | 416 | 0.865 (0.170) | 0.960 [0.242, 0.991] |
| Cluster 4 | 416 | 158 | 0.660 (0.205) | 0.737 [0.247, 0.898] |
| Cluster 5 | 156 | 355 | 0.817 (0.190) | 0.896 [0.218, 0.980] |
| Cluster 6 | 355 | 129 | 0.763 (0.148) | 0.782 [0.297, 0.951] |
| Cluster 7 | 129 | 180 | 0.914 (0.130) | 0.975 [0.400, 0.992] |
| Cluster 8 | 180 | 144 | 0.834 (0.179) | 0.899 [0.288, 0.998] |
| Cluster 9 | 144 |  |  |  |
| **ACS 2015-2019\*\*** |  |  |  |  |
| Cluster 1 | 432 | 432 | 0.937 (0.136) | 0.994 [0.223, 0.999] |
| Cluster 2 | 40 | 40 | 0.560 (0.140) | 0.545 [0.310, 0.785] |
| Cluster 3 | 174 | 180 | 0.845 (0.180) | 0.927 [0.289, 0.996] |
| Cluster 4 | 371 | 375 | 0.920 (0.134) | 0.968 [0.219, 0.986] |
| Cluster 5 | 11 | 159 | 0.846 (0.163) | 0.923 [0.277, 0.990] |
| Cluster 6 | 158 | 176 | 0.884 (0.148) | 0.962 [0.351, 0.989] |
| Cluster 7 | 176 | 116 | 0.853 (0.170) | 0.935 [0.329, 0.994] |
| Cluster 8 | 116 |  |  |  |
| N(original): Estimated number of observations per cluster conditionally on **mapK** (3 label switching algorithms); however, the model outputs the posterior mean of the probability of success per feature and cluster using the ECR algorithm.  N(relabeled): same as N(original), except for ACS 2011-2015 and 2015-2019 we relabeled clusters 2 and cluster 5. This was based on the second-highest assignment probability.  \*Assignment probabilities after relabeling  \*\* For ACS 2011-2015, Cts in cluster 2 moved to cluster 5, while Cts in cluster 5 in ACS 2015-2019 were re-clustered to clusters 3, 4 and 6. | | | | |

# Cluster descriptions for ACS 2006-2010

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Cluster | Cluster descriptions and distribution of age across census tracts  For this, we are also going to look at the distribution of age and race/ethnicity across census tracts within the clusters | | |  |
|  | **Description** | **Age distribution** | **Race/ethnicity composition of CTs in the cluster** | **Conclusion/Summary**  **\*Also looked at the distribution of language**  **\*all at the census tract level** |
| 1 | High success probabilities for:   * Renter * White collar occupation * < HS * >= Bachelor’s * TPPR   Low house ownership, no vehicle, poverty line | Avg median age is 33.2.  Higher average proportions in the age groups:   * 20-24 * 25-34 * 35-44 * 45-54 | * Lowest avg proportion of NHW compared to other clusters.   Highest avg proportion of Hispanic or Latinos compared to other clusters.   * Second highest avg proportion of NHB | Educated  Low income (at least compared to the state median)  Multi-lingual/multi-cultural (based on distribution of language spoken at home other than EN or speak EN not “very well”  Household crowding (Crowding among housing units) |
| 2 | High success probabilities for:   * Renter * TPPR * Lack of complete plumbing + < HS   Low probabilities for all other variables ranging from 0.3-0.04. | Avg median age is 34.4.  Higher average proportions in the age groups:   * 25-34 * 35-44 * 45-54 | * Second Highest avg proportion of Hispanic or Latinos compared to other clusters. * Highest avg proportion of NHB | Low education  Low income  Multi-lingual/multi-cultural (based on distribution of language spoken at home other than EN or speak EN not “very well”  Household crowding |
| 3 | High success probabilities for:   * Owner   Medium probabilities for:   * < HS * SNAP benefits * Female household | Avg median age is 41.4.  Higher average proportions in the age groups:   * 35-44 * 45-54 | Not very diverse, on average the proportion of NHW is 85.4 | Low income  Female households  Older than above |
| 4 | High success probabilities for:   * Owner * >= Bachelors * >= HS * Median income * White collar occupation | Avg median age is 43.0.  Higher average proportions in the age groups:   * 35-44 * 45-54 | Similar distribution to the above, majority NHW (89.6) | Older in comparison to previous cluster, but seem to be more affluent or higher SES level based on the avg proportion of house ownership. |
| 5 | High probabilities for:   * Owner * >= HS * Median Income   Low for:   * >= Bachelor’s degree * White collar occupation * Renter | Avg median age is 42.6  Higher average proportions in the age groups:   * 35-44 * 45-54 | Similar distribution to the above, majority NHW (93.1) | High income  At least HS educated |
| 6 | High probabilities for:   * Owner * >= Bachelor's degree * >= HS * white collar occupation | Avg median age is 42.9  Higher average proportions in the age groups:   * 35-44 * 45-54 | Similar distribution to the above, majority NHW | Highly educated.  Medium-high income  House ownership |
| 7 | High probabilities for:   * Renter * < HS * Female household * SNAP * Below poverty | Avg median age is 37.9.  Higher average proportions in the age groups:   * 25-34 * 35-44 * 45-54 | Lower avg proportion of NHW compared to previous clusters except cluster 1  Decent avg proportion of NHB and Hisp | High unemployment (proportions compared to the median for the state)  Low ed  Female household  Higher avg of gov assistantship (SNAP)  Decent multi-language/ multi-culture  BIG renters |
| 8 | High probabilities for:   * Renter * >= Bachelor’s degree * White collar occupation * >= HS * Median income * TPPR | Avg median age is 35.9.  Higher average proportions in the age groups:   * 25-34 * 35-44 * 45-54 | Higher avg proportion of NHA than any other minority group | 2nd highest renters  Highly educated and occupation  Medium-high income  Multi-lingual (based on avg proportion of language spoken at home other than EN) |
| 9 | High probabilities for:   * Owner * < HS * Median income * unemployment | Avg median age is 40.8.  Higher average proportions in the age groups:   * 25-34 * 35-44 * 45-54 | majority NHW (84.9) | Low ed  House ownership  High unemployment proportions |

For variables: lack of plumbing and two or more rooms, 0 means there is no lack and no households with two or more ppr, respectively.

# Cluster descriptions for ACS 2011-2015

About the survey:

* population grew 3.5% from 2006-2010 to 2011-2015.
* The two tracts in cluster 2: 25013812902 (Census Tract 8129.02, Hampden County, Massachusetts) & 25021415102 (Census Tract 4151.02, Norfolk County, Massachusetts)
* Looking at these in previous ACS, they were classified as cluster 4 and cluster 8, respectively.
* <https://www.census.gov/programs-surveys/acs/guidance/comparing-acs-data/2015/5-year-comparison.html>
* **Change in geographic boundaries** - The 2006-2010 ACS 5-year estimates used legal boundaries as of January 1, 2010. The 2011-2015 ACS 5-year estimates use legal boundaries as of January 1, 2015.—I think this has been adjusted from what I’ve seen so far
* **Also changes in the questionnaire or coding—this is a strength of our study because we are looking at each survey separately!!**
* ACTION: Moved 2 census tracts from cluster 2 in 2011-2015 to cluster 5 based on second highest assignment probability

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Cluster | Cluster descriptions and distribution of age across census tracts  For this, we are also going to look at the distribution of age and race/ethnicity across census tracts within the clusters | | |  |
|  | **Description** | **Age distribution** | **Race/ethnicity composition of CTs in the cluster** | **Conclusion** |
| 1  7 | High success probabilities for:   * Renter * Female household   Medium probabilities for:   * TPPR * < HS * Unemployment * Below poverty line | Avg median age is 40.4.  Higher average proportions in the age groups:   * 25-34 * 35-44 * 45-54 | majority NHW, and not very diverse | Renters  Female households  Low education  Crowded housing |
| ~~2~~ | ~~Medium probabilities for all variables but highest for~~   * ~~>= HS~~ * ~~Owner~~ * ~~>=Bachelors~~ * ~~Median income~~ | ~~Avg median age is 41.6.~~  ~~Higher average proportions in the age groups:~~   * ~~25-34~~ * ~~35-44~~ * ~~45-54~~ | ~~majority NHW, and not very diverse~~ | ~~Higher Ed~~  ~~high Income~~  ~~Note that there are only 2 census tracts in this cluster, and based on distribution of main variables I think they can be combined with cluster 6~~ |
| 3  2  4 | High success probabilities for:   * Owner * <HS * >= Bachelor’s degree * < HS * Median income | Avg median age is 43.2.  Same majority age groups as above | majority NHW, and not very diverse | House ownership  High ed  High income |
| 4  3  2 | High success probabilities for:   * Renter * Unemployment * < HS * TPPR | Avg median age is 35.4.  Same majority age groups as above | * Highest avg proportion of Hispanic or Latinos compared to other clusters. * Highest avg proportion of NHB | Low ed  Crowded housing  Low income  Multi-language/multi-cultural (EN not very well and other language other than EN)  Similar to clusters 1 &2 in 2006-2010 survey—further exploration through cross-tabulations |
| 5  4  5 | High probabilities for:   * Owner * >= HS | Avg median age is 44.6.  Same majority age groups as above | Similar distribution to the above, majority NHW (91.3) | At least HS educated.  House ownership |
| 6  5  6 | High probabilities for:   * Owner * >=bachelor’s degree * White collar * >= HS | Avg median age is 44.4.  Same majority age groups as above | Similar distribution to the above, majority NHW | Higher ed  House ownership |
| 7  6  1 | High probabilities for:   * >=Bachelor’s degree * Renter * White collar occupation * < HS * TPPR | Avg median age is 35.5.  Same majority age groups as above | 2nd highest avg proportion of Hispanic or Latinos and NHB compared to other clusters. | Higher ed  Renters  Multi-lingual  Crowding among housing units |
| 8  7  8 | High probabilities for:   * Renter * >= Bachelor’s degree * White collar occupation * >= HS * TPPR | Avg median age is 35.7.  Same majority age groups as above | Higher avg proportion of NHA than any other minority group | Multi-lingual  Crowding  High education  Med-high income |
| 9  8  3 | High probabilities for:   * Owner * < HS | Avg median age is 42.6.  Same majority age groups as above | majority NHW (82.5) | House ownership  Low ed  Low income |

# Cluster descriptions for ACS 2015-2019

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Cluster | Cluster descriptions and distribution of age across census tracts.  For this, we are also going to look at the distribution of age and race/ethnicity across census tracts within the clusters | | |  |
|  | **Description** | **Age distribution** | **Race/ethnicity composition of CTs in the cluster** | **Conclusion** |
| 1 | High success probabilities for:   * Renter   Medium probs for:   * TPPR * <HS * Unemployment | Avg median age is 36.1  Higher average proportions in the age groups:   * 25-34 * 35-44 * 45-54 | * Highest avg proportion of Hispanic or Latinos compared to other clusters. * Highest avg proportion of NHB | Multi-cultural/ multi-lingual  Renters  Low education  Low income |
| ~~2~~ | ~~High success probabilities for:~~   * ~~Renter~~ * ~~TPPR~~ * ~~Lack of complete plumbing + < HS~~   ~~Low probabilities for all other variables ranging from 0.3-0.04.~~ | ~~Avg median age is 38.6~~  ~~Higher average proportions in the age groups:~~   * ~~20-24~~ * ~~25-34~~ * ~~35-44~~ * ~~45-54~~ | * ~~Second Highest avg proportion of Hispanic or Latinos and NHB compared to other clusters.~~ | ~~Multi-cultural~~  ~~Crowding among housing units~~  ~~Low education~~ |
| 3  2 | High success probabilities for:   * Owner * < HS   Medium probabilities for:   * Female household * Below poverty * unemployment | Avg median age is 43.7  Higher average proportions in the age groups:   * 25-34 * 35-44 * 45-54 | Clusters 3-7 have similar distributions the one thing to notice | House ownership  Low education  Low income |
| 4 | High success probabilities for:   * Owner * >= Bachelors * >= HS * White collar occupation   More affluent census tracts? | Avg median age is 44.6  Higher average proportions in the age groups:   * 35-44 * 45-54 | Majority NHW not much diversity | House ownership  High education  High income  (similar to cluster 4 and 6 in 2006-2010)  Affluent CTs |
| ~~5~~ | ~~High probabilities for:~~   * ~~Owner~~ * ~~>= bachelor’s~~ * ~~Female household~~ * ~~White collar occupation~~ * ~~Below poverty~~ | ~~Avg median age is 40.1~~  ~~Higher average proportions in the age groups:~~   * ~~25-34~~ * ~~35-44~~ * ~~45-54~~ | ~~Decent proportion of NHB~~ | ~~Higher education~~  ~~Female households~~  ~~Only 11 Cts here and all have avg proportion below poverty line >= median value for the state.~~ |
| 6  5 | High probabilities for:   * Owner * >= HS   Low for everything else | Avg median age is 46.7  Higher average proportions in the age groups:   * 35-44 * 45-54 | Majority NHW not much diversity | Not multi-lingual (based on distribution of language)  House ownership  At least HS education |
| 7  6 | High probabilities for:   * >= Bachelors * White collar occupation * >= HS * Renter | Avg median age is 35.6  Higher average proportions in the age groups:   * 20-24 * 25-34 * 35-44 * 45-54 | Decent proportion of Hispanics | Renters  High education  Multi-lingual |
| 8  7 | High probabilities for:   * Renter * >= Bachelor’s degree * White collar occupation * < HS * Median income * TPPR | Avg median age is 35.8  Higher average proportions in the age groups:   * 25-34 * 35-44 * 45-54 | * Second Highest avg proportion of Hispanic or Latinos and NHB compared to other clusters. | Multi-lingual (similar distribution of language as cluster 1)  Combination of high ed and < HS  Crowding |

**Relabeling clusters to match a reference dataset in this case we use 2006-2010.**

Original

A number of clusters on a white background

Description automatically generated

After relabeling based on cluster description similarities

A number of clusters with black text

Description automatically generated with medium confidence

Original

A number of clusters with black text

Description automatically generated with medium confidence