Country: **Nigeria**

Year: **2018**

Number of Observations: **40,427**

Number of Clusters: **5**

Number of variables used: **26**

Distance used: **Hamming**

**Variables used in the algorithm:**

hv206,hv207,hv208,hv209,hv210,hv211,hv227,hv243a,hv243b,hv244,hv246,hv247,hv252,sh121g,sh121h,sh121j,sh121k,sh121m,sh121n,sh121o,water,toilet,floor,roof,cookfuel,wall

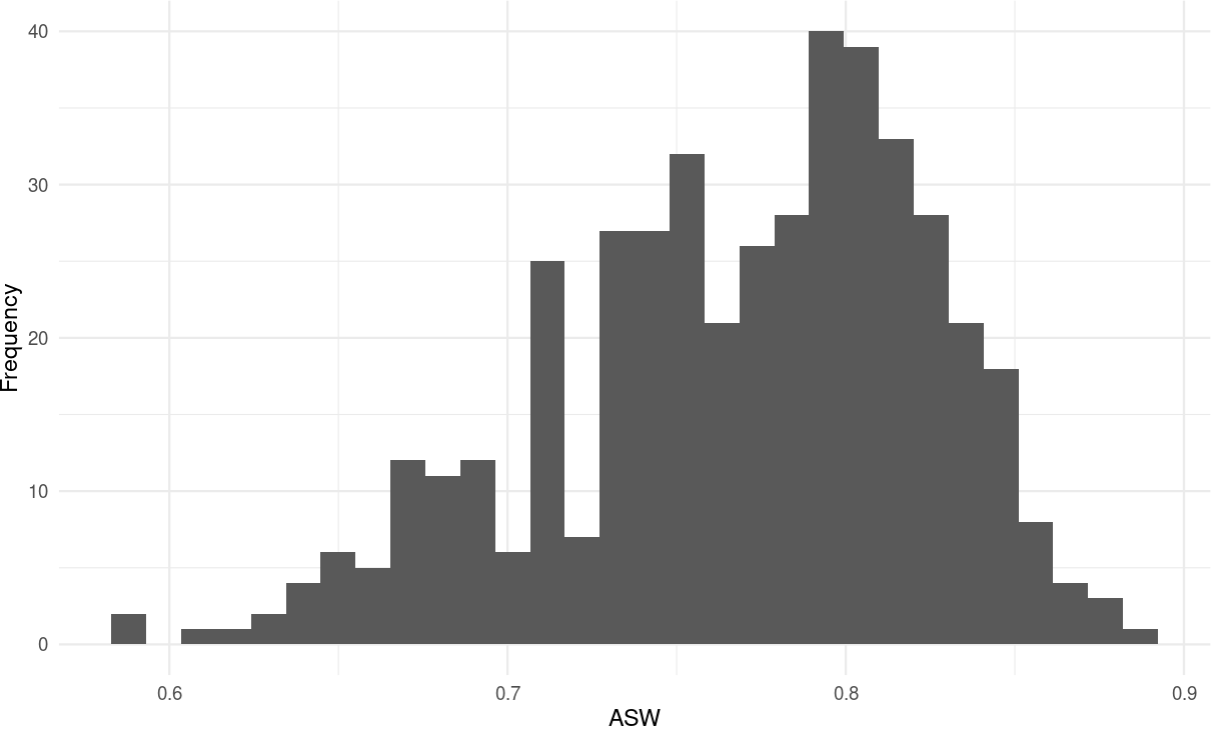
Removed hv246a:hv246g (specific animal ownership) because they were collinear with hv246 (owns any animals at all)

The dataset was partitioned into 4 parts because it was too large to run. Then, the algorithm was rerun on the entire dataset with any variables that made it in the top 10 clusters in the 4 partitions. These are the results from that.

**Summary of Top Cluster Configurations (ranked by ASW)**

| **Cluster #** | **ASW** | **Variable 1** | **Variable 2** | **Variable 3** | **Variable 4** |
| --- | --- | --- | --- | --- | --- |
| **1** | 0.887062747971098 | hv206 | sh121h | sh121o | roof |
| **2** | 0.887062747971098 | hv210 | hv243a | sh121h | roof |
| **3** | 0.88076945808317 | hv206 | sh121g | sh121o | roof |
| **4** | 0.88076945808317 | hv210 | hv243a | hv252 | floor |
| **5** | 0.877098827734894 | hv206 | hv210 | hv243a | sh121o |
| **6** | 0.877098827734894 | hv209 | hv211 | sh121o | floor |
| **7** | 0.87659428105753 | hv206 | hv210 | hv243a | floor |
| **8** | 0.87659428105753 | hv209 | hv243a | hv252 | sh121g |
| **9** | 0.87136471304451 | hv206 | sh121g | sh121n | floor |
| **10** | 0.87136471304451 | hv210 | hv243a | hv252 | roof |

**Distributions of ASW values in all clusters in Nigeria**

****

**Marginal Distributions**

| Variable | Description | % time in top clusters | Distribution |
| --- | --- | --- | --- |
| hv206 | Has electricity | 50% | Binary,55.3% 1s (or yes) |
| hv209 | Has refrigerator | 20% | Binary,20.4% 1s (or yes) |
| hv210 | Has bicycle | 50% | Binary,12.6% 1s (or yes) |
| hv211 | Has motorcycle/scooter | 10% | Binary,29.2% 1s (or yes) |
| hv243a | Has mobile telephone | 60% | Binary,87.0% 1s (or yes) |
| hv252 | Frequency household members smoke inside the house | 20% | Categorical, values 0-4 |
| sh121g | Has table | 30% | Binary,61.2% 1s (or yes) |
| sh121h | Has chair | 10% | Binary,80.2% 1s (or yes) |
| sh121n | Has generator | 10% | Binary,27.7% 1s (or yes) |
| sh121o | Has fan | 30% | Binary,49.1% 1s (or yes) |
| roof | Type of roof in dwelling | 40% | Categorical, 0-3 |
| floor | Type of floor in dwelling | 40% | Categorical, 0-3 |

**Summary of variable distributions in top clusters**

Currently our method is choosing five distinct clusters of individuals within each cluster variable configuration. Here are the medioids for each of these five clusters:

**Cluster #1 Configuration**

| **Config#** | **Node** | **hv206** | **sh121h** | **sh121o** | **roof** | **Proportion** |
| --- | --- | --- | --- | --- | --- | --- |
| **1** | 1 | 0 | 1 | 1 | 0 | 7.01% |
| **2** | 1 | 0 | 1 | 1 | 1 |
| **3** | 1 | 0 | 1 | 1 | 2 |
| **4** | 1 | 0 | 1 | 1 | 3 |
| **5** | 2 | 0 | 0 | 0 | 0 | 13.26% |
| **6** | 2 | 0 | 0 | 0 | 1 |
| **7** | 2 | 0 | 0 | 0 | 2 |
| **8** | 2 | 0 | 0 | 0 | 3 |
| **9** | 2 | 0 | 0 | 1 | 0 |
| **10** | 2 | 0 | 0 | 1 | 1 |
| **11** | 2 | 0 | 0 | 1 | 2 |
| **12** | 3 | 0 | 1 | 0 | 0 | 24.41% |
| **13** | 3 | 0 | 1 | 0 | 1 |
| **14** | 3 | 0 | 1 | 0 | 2 |
| **15** | 3 | 0 | 1 | 0 | 3 |
| **16** | 4 | 1 | 0 | 1 | 0 | 41.67% |
| **17** | 4 | 1 | 0 | 1 | 1 |
| **18** | 4 | 1 | 0 | 1 | 2 |
| **19** | 4 | 1 | 0 | 1 | 3 |
| **20** | 4 | 1 | 1 | 1 | 0 |
| **21** | 4 | 1 | 1 | 1 | 1 |
| **22** | 4 | 1 | 1 | 1 | 2 |
| **23** | 4 | 1 | 1 | 1 | 3 |
| **24** | 5 | 1 | 0 | 0 | 0 | 13.65% |
| **25** | 5 | 1 | 0 | 0 | 1 |
| **26** | 5 | 1 | 0 | 0 | 2 |
| **27** | 5 | 1 | 1 | 0 | 0 |
| **28** | 5 | 1 | 1 | 0 | 1 |
| **29** | 5 | 1 | 1 | 0 | 2 |
| **30** | 5 | 1 | 1 | 0 | 3 |

**Validation of Top Cluster with Chi-Sq Distribution**

Note that in the tables, the rows are the validation variables, and the columns are the node number from the cluster distributions in the table above.

