Country: **Mali**

Year: **2018**

Number of Observations: **9510**

Number of Clusters: **5**

Number of variables used: **25**

Distance used: **Hamming**

**Variables used in the algorithm:**

hv206,hv207,hv208, hv209,hv210, hv211, hv227,hv237,hv243a,hv243c,hv244,hv246,hv247,

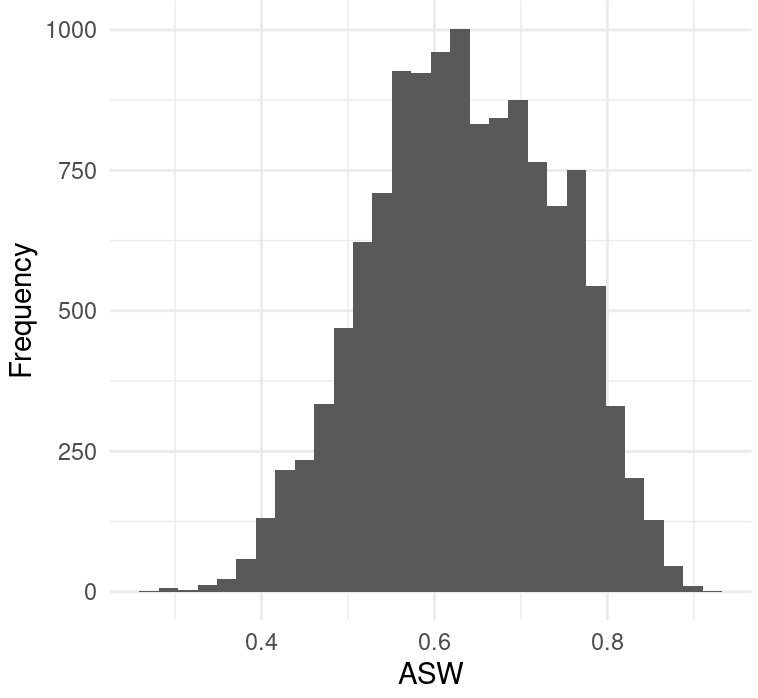
Hv252,sh121g,sh121h,sh121ksh121m,sh122g,water,toilet,floor,roof,cookfuel,wall

Removed any binary variables 10% class imbalance or less, removed sh145a,hv237b, and hv237c for collinearity

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

| **Cluster #** | **ASW** | **Variable 1** | **Variable 2** | **Variable 3** | **Variable 4** |
| --- | --- | --- | --- | --- | --- |
| **1** | 0.921222427493205 | hv209 | hv227 | hv243a | sh121h |
| **2** | 0.901255554545916 | hv208 | hv209 | hv227 | hv243a |
| **3** | 0.900866177648387 | hv209 | hv227 | hv243a | sh121m |
| **4** | 0.900473032315363 | hv227 | hv243a | hv247 | sh121m |
| **5** | 0.897501600660236 | hv227 | hv243a | hv247 | sh121h |
| **6** | 0.895732763110734 | hv227 | hv243a | sh121h | sh121m |
| **7** | 0.895529652014523 | hv206 | hv208 | hv209 | hv227 |
| **8** | 0.894187789349362 | hv209 | hv227 | hv243a | hv247 |
| **9** | 0.893987428497675 | hv208 | hv209 | hv243a | sh121h |
| **10** | 0.893577352236288 | hv206 | hv209 | hv227 | hv243a |

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

****

**Marginal Distributions**

| Variable | Description | % time in top clusters | Distribution |
| --- | --- | --- | --- |
| hv206 | Has electricity | 10% | Binary,47.1% 1s (or yes) |
| hv208 | Has radio | 20% | Binary,63.78% 1s (or yes) |
| hv209 | Has refrigerator | 60% | Binary,11.98% 1s (or yes) |
| hv227 | Has mosquito bed net for sleeping | 70% | Binary,85.9% 1s (or yes) |
| hv243a | Has mobile telephone | 90% | Binary,85.5% 1s (or yes) |
| hv247 | Has bank account | 30% | Binary,14.0% 1s (or yes) |
| sh121h | wardrobe/bookcase | 40% | Binary,21.4% 1s (or yes) |
| sh121m | internet | 30% | Binary,17.0% 1s (or yes) |

**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** | **hv210** | **hv237** | **hv243c** | **sh121k** | **Proportion** |
| --- | --- | --- | --- | --- | --- | --- |
| **1** | 1 | 0 | 0 | 0 | 0 | 14.05% |
| **2** | 1 | 0 | 1 | 0 | 0 |
| **3** | 1 | 1 | 0 | 0 | 0 |
| **4** | 1 | 1 | 1 | 0 | 0 |
| **5** | 2 | 0 | 0 | 1 | 0 | 7.51% |
| **6** | 2 | 1 | 0 | 1 | 0 |
| **7** | 3 | 1 | 0 | 0 | 1 | 8.48% |
| **8** | 3 | 1 | 0 | 1 | 1 |
| **9** | 3 | 1 | 1 | 0 | 1 |
| **10** | 3 | 1 | 1 | 1 | 1 |
| **11** | 4 | 0 | 1 | 1 | 0 | 57.09% |
| **12** | 4 | 1 | 1 | 1 | 0 |
| **13** | 5 | 0 | 0 | 0 | 1 | 12.88% |
| **14** | 5 | 0 | 0 | 1 | 1 |
| **15** | 5 | 0 | 1 | 0 | 1 |
| **16** | 5 | 0 | 1 | 1 | 1 |

**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.

