Country: **Niger**

Year: **2012**

Number of Observations: **10750**

Number of Clusters: **5**

Number of variables used: **26**

Distance used: **Hamming**

**Variables used for algorithm:**

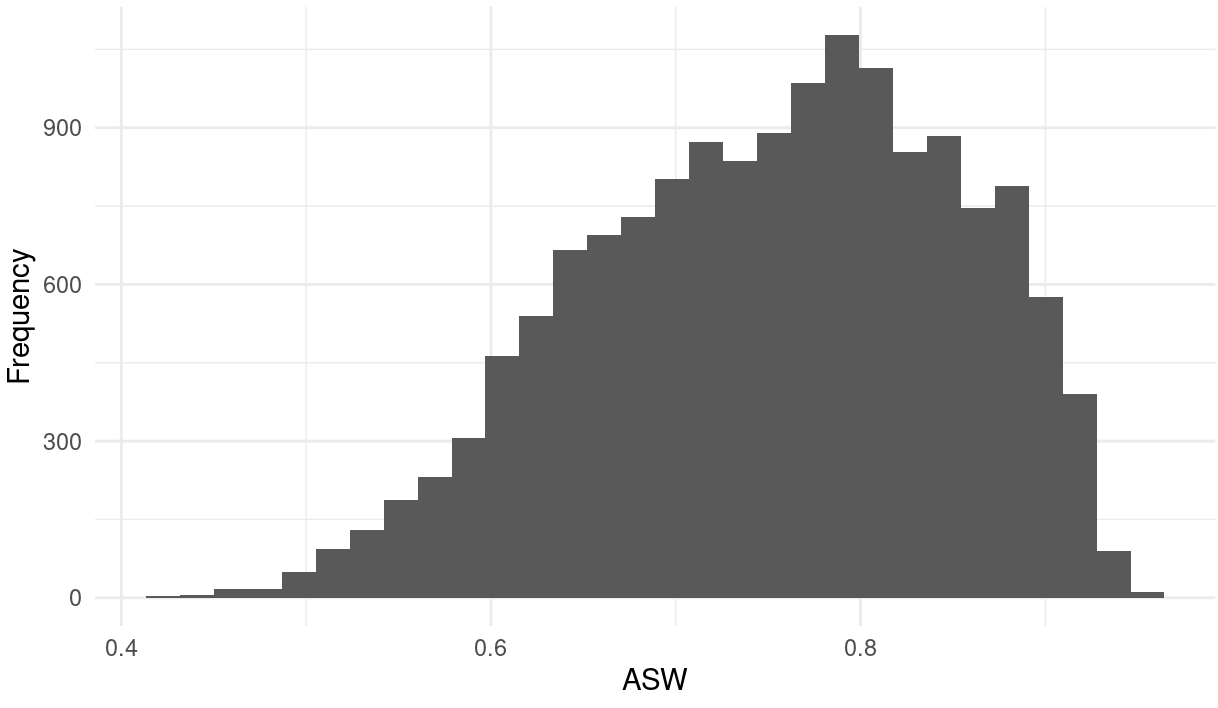
hv206,hv207,hv208,hv210,hv211,hv227,hv237,hv243a,hv243b,hv243c,hv244,hv246,hv246c,hv246g,hv246h,hv253,sh106b,sh110d,sh110m,sh141,water,toilet,floor,roof,cookfuel,wall

Excluded any binary variables with 10% class imbalance or less, and hv237c,hv253a:z for collinearity.

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

| **Cluster #** | **ASW** | **Variable 1** | **Variable 2** | **Variable 3** | **Variable 4** |
| --- | --- | --- | --- | --- | --- |
| **1** | 0.955313279408841 | hv208 | hv237 | hv253 | sh110d |
| **2** | 0.954788347270733 | hv206 | hv208 | hv237 | hv253 |
| **3** | 0.952947181625097 | hv208 | hv253 | sh110d | sh110m |
| **4** | 0.951872731614797 | hv206 | hv237 | hv253 | sh110d |
| **5** | 0.950970994592701 | hv244 | hv246 | hv253 | sh110m |
| **6** | 0.950908480937549 | hv244 | hv253 | sh110d | sh110m |
| **7** | 0.950165362553401 | hv208 | hv244 | hv253 | sh110m |
| **8** | 0.949694895314331 | hv208 | hv237 | hv243a | hv253 |
| **9** | 0.949137253717585 | hv208 | hv243a | hv253 | sh110d |
| **10** | 0.948606257375536 | hv237 | hv253 | sh110d | sh110m |

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



**Marginal Distributions**

| Variable | Description | % time in top clusters | Distribution |
| --- | --- | --- | --- |
| hv206 | Has electricity | 20% | Binary,22.9% 1s (or yes) |
| hv208 | Has television | 50% | Binary,18.8% 1s (or yes) |
| hv237 | Anything done to make water safe to drink | 50% | Binary,13.8% 1s (or yes) |
| hv243a | Has mobile telephone | 20% | Binary,55.5% 1s (or yes) |
| hv244 | Owns land useable for agriculture | 30% | Binary,71.7% 1s (or yes) |
| hv246 | Owns livestock, herds or farm animals | 10% | Binary,7.1% 1s (or yes) |
| hv253 | Has dwelling been sprayed against mosquitoes in last 12 months | 100% | Binary,1.0% 1s (or yes) |
| sh110d | VCR/DVD | 60% | Binary,14.2% 1s (or yes) |
| sh110m | Plow | 50% | Binary,9.5% 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** | **hv208** | **hv237** | **hv253** | **sh110d** | **Proportion** |
| --- | --- | --- | --- | --- | --- | --- |
| **1** | 1 | 1 | 0 | 0 | 1 | 13.49% |
| **2** | 1 | 1 | 0 | 1 | 1 |
| **3** | 1 | 1 | 1 | 0 | 1 |
| **4** | 1 | 1 | 1 | 1 | 1 |
| **5** | 2 | 0 | 0 | 0 | 1 | 13.02% |
| **6** | 2 | 0 | 0 | 1 | 1 |
| **7** | 3 | 1 | 0 | 0 | 0 | 0.90% |
| **8** | 3 | 1 | 0 | 1 | 0 |
| **9** | 4 | 0 | 0 | 0 | 0 | 5.43% |
| **10** | 4 | 0 | 0 | 1 | 0 |
| **11** | 5 | 0 | 1 | 0 | 0 | 67.17% |
| **12** | 5 | 0 | 1 | 0 | 1 |
| **13** | 5 | 0 | 1 | 1 | 0 |
| **14** | 5 | 1 | 1 | 0 | 0 |

**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.

