Country: **Zambia**

Year: **2018**

Number of Observations: **12831**

Number of Clusters: **5**

Number of variables used: **25**

Distance used: **Hamming**

**Variables used in the algorithm:**

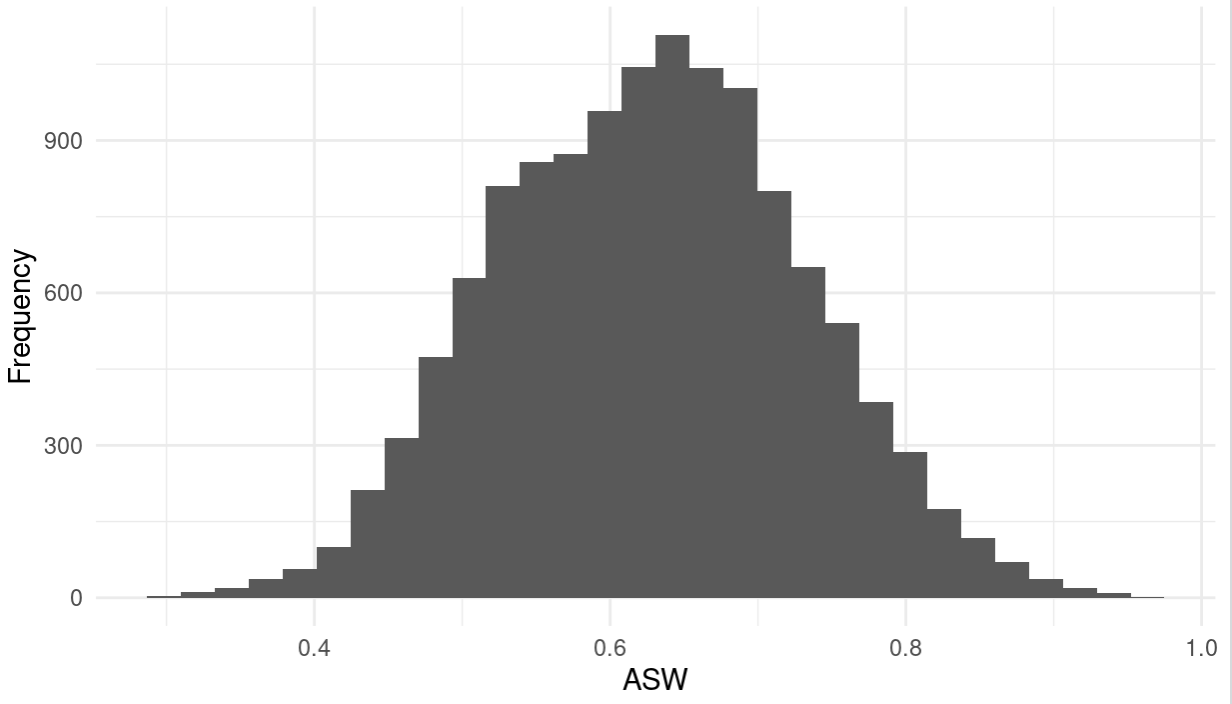
vv206,hv207,hv208,hv209,hv210,hv227,hv237,hv243a,hv243b,hv243e,hv244,hv246,hv247,hv252,hv253,sh121g,sh121h,sh121i,sh121p,water,toilet,floor,roof,cookfuel,wall

Excluded any binary variables with class imbalance of 7% or less, excluded sh119,hv237a:b,hv246e:i, hv253a:hv253z for collinearity

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

| **Cluster #** | **ASW** | **Variable 1** | **Variable 2** | **Variable 3** | **Variable 4** |
| --- | --- | --- | --- | --- | --- |
| **1** | 0.95761768160043 | hv206 | hv209 | hv243e | roof |
| **2** | 0.950346798179905 | hv206 | hv209 | hv243e | sh121p |
| **3** | 0.945460964886123 | hv206 | hv209 | hv243a | hv243e |
| **4** | 0.945327080046562 | hv208 | hv209 | hv243e | roof |
| **5** | 0.943938703164132 | hv208 | hv209 | hv243a | hv243e |
| **6** | 0.942372034130731 | hv206 | hv209 | hv243e | sh121g |
| **7** | 0.941720685968272 | hv208 | hv209 | hv243e | sh121g |
| **8** | 0.940578994318486 | hv209 | hv243e | floor | roof |
| **9** | 0.939801295878396 | hv206 | hv209 | hv243e | floor |
| **10** | 0.932522096000288 | hv206 | hv243e | floor | roof |

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



**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:

**Marginal Distributions**

| Variable | Description | % time in top clusters | Distribution |
| --- | --- | --- | --- |
| hv206 | Has electricity | 60% | Binary,28.8% 1s (or yes) |
| hv208 | Has television | 20% | Binary,31.9% 1s (or yes) |
| hv209 | Has refrigerator | 90% | Binary,19.0% 1s (or yes) |
| hv243a | Has mobile telephone | 20% | Binary,70.6% 1s (or yes) |
| hv243e | Has computer | 100% | Binary,7.2% 1s (or yes) |
| sh121g | Has bed | 20% | Binary,67.7% 1s (or yes) |
| sh121p | Has plough | 10% | Binary,10.1% 1s (or yes) |
| floor | Description of floor in dwelling | 30% | Categorical |
| roof | Description of roof in dwelling | 40% | Categorical |

**Cluster #1 Configuration**

| **Config#** | **Node** | **hv206** | **hv209** | **hv243e** | **roof** | **Proportion** |
| --- | --- | --- | --- | --- | --- | --- |
| **1** | 1 | 1 | 0 | 0 | 1 | 10.0% |
| **2** | 1 | 1 | 0 | 0 | 2 |
| **3** | 1 | 1 | 0 | 0 | 3 |
| **4** | 1 | 1 | 0 | 1 | 2 |
| **5** | 2 | 0 | 0 | 0 | 1 | 35.5% |
| **6** | 2 | 0 | 0 | 0 | 2 |
| **7** | 2 | 0 | 0 | 0 | 3 |
| **8** | 2 | 0 | 0 | 1 | 2 |
| **9** | 2 | 0 | 1 | 0 | 2 |
| **10** | 3 | 1 | 1 | 0 | 0 | 12.5% |
| **11** | 3 | 1 | 1 | 0 | 1 |
| **12** | 3 | 1 | 1 | 0 | 2 |
| **13** | 4 | 0 | 1 | 1 | 2 | 6.0% |
| **14** | 4 | 0 | 1 | 1 | 3 |
| **15** | 4 | 1 | 1 | 1 | 0 |
| **16** | 4 | 1 | 1 | 1 | 1 |
| **17** | 4 | 1 | 1 | 1 | 2 |
| **18** | 5 | 0 | 0 | 0 | 0 | 36.1% |
| **19** | 5 | 0 | 0 | 1 | 0 |
| **20** | 5 | 0 | 1 | 0 | 0 |
| **21** | 5 | 1 | 0 | 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.

