Country: **Rwanda**

Year: **2019**

Number of Observations: **12949**

Number of Clusters: **5**

Number of variables used: **28**

Distance used: **Hamming**

**Variables used in the algorithm:**

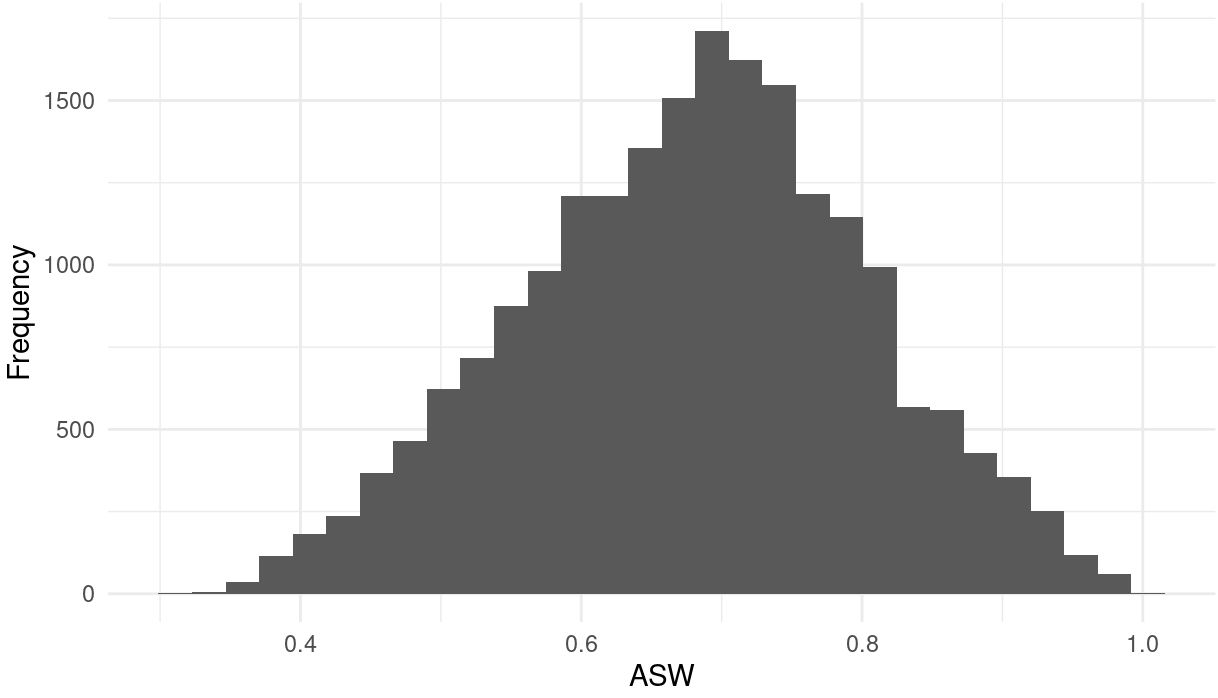
hv206,hv207,hv208,hv210,hv225,hv227,hv237,hv243a,hv243b,hv244,hv246,hv247,hv252,sh121g,sh121h,sh121i,sh121j,sh121k,sh121l,sh121n,sh121p,sh144a,water,toilet,floor,roof,cookfuel, wall

Excluded hv237a,hv246a:i,sh112aa:sh112ac for collinearity, excluded any binary variables with class imbalance of 10% or less

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

| **Cluster #** | **ASW** | **Variable 1** | **Variable 2** | **Variable 3** | **Variable 4** |
| --- | --- | --- | --- | --- | --- |
| **1** | 0.993035477118192 | hv208 | sh121g | sh121n | roof |
| **2** | 0.992241720075418 | hv208 | sh121g | sh121k | roof |
| **3** | 0.992171414087217 | hv208 | sh121g | sh121p | roof |
| **4** | 0.991464194684687 | hv208 | hv243a | sh121g | roof |
| **5** | 0.990560396923147 | hv243a | sh121g | sh121n | roof |
| **6** | 0.990444405988535 | sh121g | sh121k | sh121n | roof |
| **7** | 0.9900996723882 | hv208 | hv243a | sh121n | roof |
| **8** | 0.98991538490863 | sh121g | sh121n | sh121p | roof |
| **9** | 0.989431198745119 | sh121g | sh121j | sh121n | roof |
| **10** | 0.989172506720772 | hv208 | sh121g | sh121j | roof |

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

****

**Marginal Distributions**

| Variable | Description | % time in top clusters | Distribution |
| --- | --- | --- | --- |
| hv208 | Has television | 60% | Binary,X% 1s (or yes) |
| hv243a | Has mobile telephone | 30% | Binary,X% 1s (or yes) |
| sh121g | mattress | 90% | Binary,X% 1s (or yes) |
| sh121k | sofa | 20% | Binary,X% 1s (or yes) |
| sh121n | cupboard | 60% | Binary,X% 1s (or yes) |
| sh121j | table | 20% | Binary,X% 1s (or yes) |
| sh121p | Iron machine | 20% | Binary,X% 1s (or yes) |
| roof | Type of roof in dwelling | 100% | Binary,X% 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** | **var1** | **var2** | **var3** | **var4** | **Proportion** |
| --- | --- | --- | --- | --- | --- | --- |
| **1** | 1 | 0 | 1 | 1 | 1 | 4.53% |
| **2** | 1 | 0 | 1 | 1 | 2 |
| **3** | 2 | 0 | 0 | 0 | 0 | 31.03% |
| **4** | 2 | 0 | 0 | 0 | 1 |
| **5** | 2 | 0 | 0 | 0 | 2 |
| **6** | 2 | 0 | 0 | 0 | 3 |
| **7** | 2 | 0 | 0 | 1 | 2 |
| **8** | 2 | 1 | 0 | 0 | 2 |
| **9** | 3 | 1 | 1 | 0 | 2 | 5.82% |
| **10** | 4 | 0 | 1 | 0 | 0 | 50.82% |
| **11** | 4 | 0 | 1 | 0 | 1 |
| **12** | 4 | 0 | 1 | 0 | 2 |
| **13** | 5 | 1 | 0 | 1 | 2 | 7.81% |
| **14** | 5 | 1 | 1 | 1 | 2 |

**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.

