EconoClusters

Efficient Definition of SES in SSA using only 4 assets

# Introduction

The table below links the clustering results from 33 countries, as detailed in our accompanying paper (link to come). The project's overall goal was to derive meaningful socio-economic status (SES) clusters of households using the DHS data. In addition, we did so by using a minimal set (4 or 5) of SES/asset variables rather than the number of relevant variables collected in the DHS survey (> 30). The impetus for this comes from our studies of trauma and injury in low-resource settings in sub-Saharan Africa. We have found (REF) that SES status predicts health outcomes post-injury. Still, collecting data for research purposes on trauma victims is challenging and time-constrained, and assessing SES status must be very efficient (can only ask 4-5 questions about SES to the injury victim or their family members in the context of, particularly, severe trauma).

The user has access to several resources, including

1. The country-level clustering of 33 countries here.
2. An example of more extensive output by country with annotation for help in interpretation here.
3. Access to code was used to process the original DHS data and run the clustering analyses.

# Methods

The EconomicClusters algorithm was applied to 33 SSA countries with available Demographic Health Survey (DHS) data from 2010 or later. The EconomicClusters ran weighted k-medoids clustering on all combinations of four asset variables that characterized each country's most distinct economic groups of five. Cluster optimality was based on the maximum average silhouette width (ASW), an average measure of each observation's similarity within their cluster versus neighboring clusters**.** We validated this by analyzing cluster associations with established DHS proxies correlated with SES, including child mortality, public/private healthcare reliance, and women’s educational attainment. These evaluations are included in the linked documents.

# Key Messages

The DHS data provides a rich source of information about SES, but having available algorithms to define SES groups parsimoniously has been underdeveloped. Studies of acute trauma and injury require parsimonious characterization of SES, preventing extensive questionnaires such as the DHS. We proposed an updated, generalizable process for identifying asset variables that classify households into ranked asset groups. This framework permits the succinct identification of important asset variables in trauma registry data collection that will be used and continually updated to identify and support at-risk groups for trauma and injury.