# Using Google Trends to Measure Ethnic and Religious Identity in sub-Saharan Africa: Potentials and Limitations

*Google Trends has already been used by social science researchers to measure racism within a community. In this article LSE's Elliott Green demonstrates how this online tool is useful in measuring other personal attributes that can be challenging to assess.*

One of the more interesting online tools that has become available to social science researchers in recent years is Google Trends (GT), which shows how often a given search-term is entered relative to the total volume of searches in a particular context.  GT often gives great insight into people's private behaviour on the internet which would not otherwise be easy to measure: in one well-known example, shows that internet searches for the racial slur word N*****R in any particular state is correlated with a decline in voting for Barack Obama in US Presidential elections, relative to other Democratic candidates.  What makes this study useful is that measuring racism through surveys or other direct means can be very difficult inasmuch as individuals do not want to admit to socially unacceptable behaviour; in contrast, searching on Google is done privately and therefore can be more accurate in understanding underlying beliefs.

GT has potential to measure other personal attributes that are not otherwise easy to record.  One such measure is ethnic identity, which is often not asked on census forms in more ethnically diverse countries due to its divisive nature. Yet there is a great deal of evidence that Presidents tend to favour their ethnic kin, especially in non-democracies and in Africa; yet without detailed ethnic demographic data researchers are often forced to use residence by region as a proxy for ethnic identity .

The utility of using GT for measuring ethnic identity relies upon the assumption that people search for terms that are important to them, whether it is a racial slur or their own ethnic or religious group.  Indeed, ( research by Seth Stephens-Davidowitz shows that there is a very strong correlation between searching for "Jewish" or "Hispanic" across American states and that group's proportion of the state population.  As such I use GT here to measure ethnic and religious identity in Nigeria, which is a useful case study for this exercise in several ways.  First, both ethnicity and religion are extremely relevant in Nigeria, which has a long history of tribal and communal conflict including a major civil war in the 1960s and decades of political instability since then.  Second, Nigeria has had many problems measuring ethnic and religious identity in its censuses, such that the 1962 and 1973 results were considered by many to have been falsified, and the 1991 census was annulled after it found more than 30 million fewer Nigerians than World Bank projections; as a result the 2006 census avoided controversy by not asking questions about ethnic or religious identity.  Third, GT data is available at the state level in Nigeria, which yields 37 observations, or more than any other country in Africa.  Finally, Nigeria has the largest population of internet users of any country in Africa (63 million), which means that its search data can be considered relatively representative of the broader population.

To assess the accuracy of GT data on ethnic identity I plot below the GT results for each major ethnic/religious group in Nigeria against ethnic/religious data from the most recent Nigerian Demographic and Health Survey (DHS) in 2013, which surveyed over 56,000 people.  The GT state-level data is measured as a percentage of the state with the highest relative search rate over the past five years (given at 100 and identified by name in each Figure); in contrast, the DHS data is the actual proportion of the survey respondents who identified with the relevant ethnic/religious identity.  In each chart I report the coefficient of determination ($R^2$) from a regression of DHS data on GT data to show how well the data are correlated.

Africa at LSE: Using Google Trends to Measure Ethnic and Religious Identity in sub-Saharan Africa: Potentials and Limitations

Page 2 of 6

In Figures 1-3 I start with the three largest groups of Hausa, Igbo and Yoruba.  In all three cases there is a tight fit, especially with the Igbo and Yoruba.  I then examine Christians and Muslims in Figures 4 and 5, where the fit is not as strong.  (I use the search term "Islam" instead of "Muslim" as there is more search data in the former case.)  In both of these cases there are a lot of states with no results (which I coded as a zero), which could arise from the fact that data is reported only when the absolute level of search is below a certain threshold, and across Nigeria there are more searches for "Hausa" and "Yoruba" than for "Christian" or "Islam."  (There are more searches for "Christian" and "Islam" than for "Igbo.")  However, if we use "Jesus" as search term instead of "Christian," inasmuch as there are 40 per cent more searches for the former than the latter, then the fit with percentage Christian is much better (as seen in Figure 6).
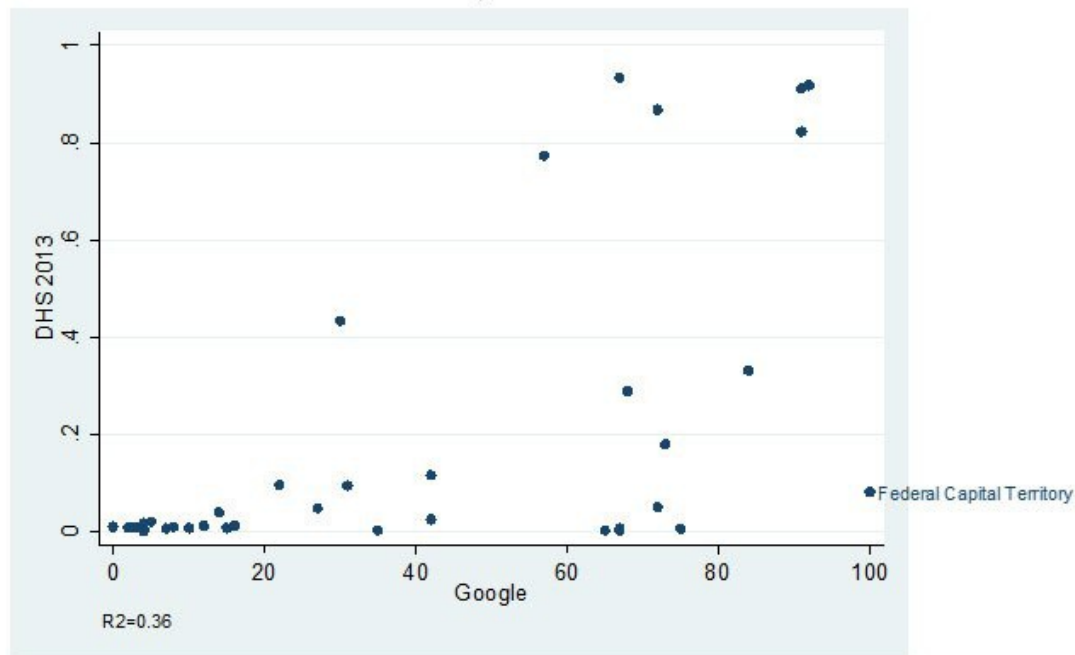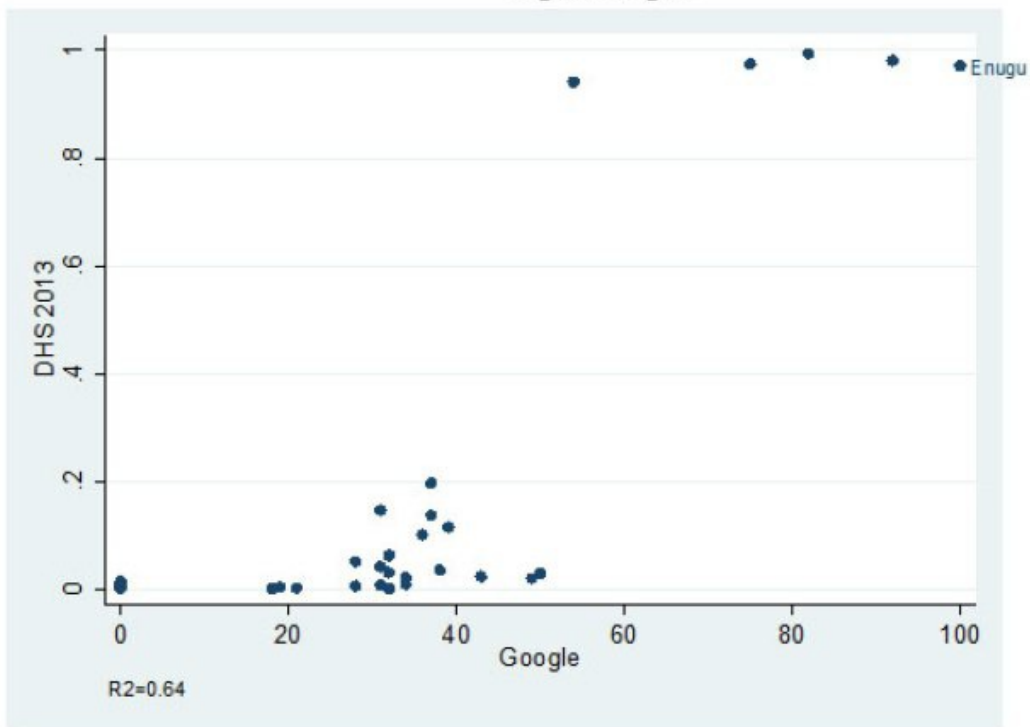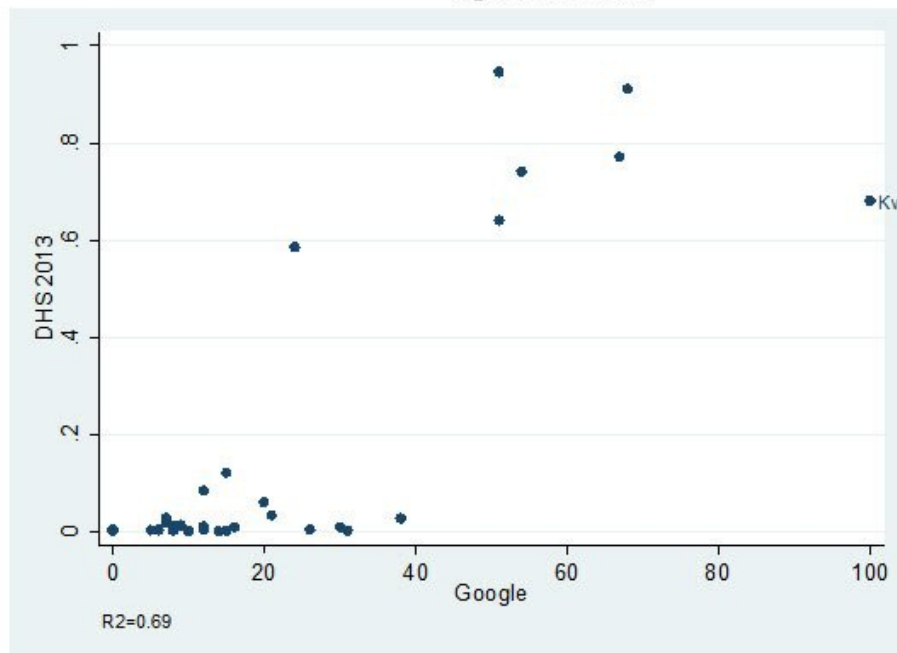


Figure 1: Hausa

**Figure 2: Igbo**



R2=0.64

**Figure 3: Yoruba**



R2=0.69

Africa at LSE: Using Google Trends to Measure Ethnic and Religious Identity in sub-Saharan Africa: Potentials and Limitations

Page 4 of 6

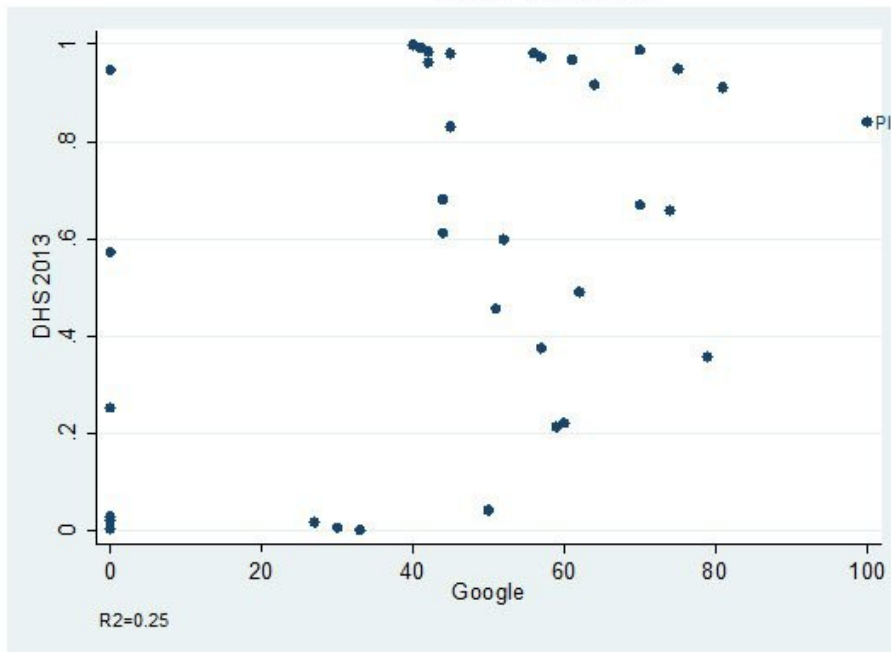## Figure 4: Christian


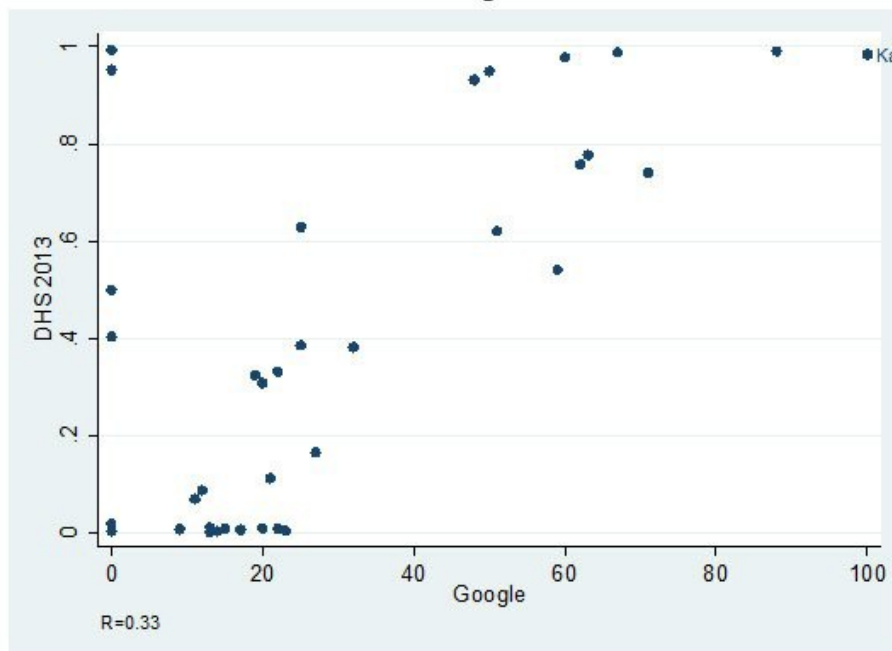
R2=0.25

## Figure 5: Islam



R=0.33

Figure 6: Jesus (DHS=% Christian)
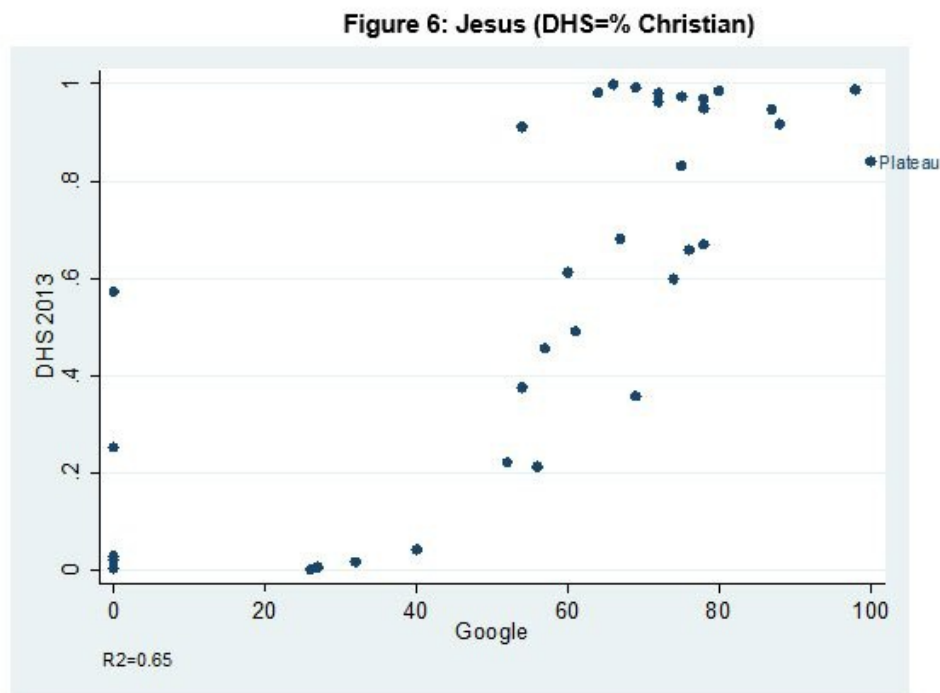
The same problem can apply to smaller ethnic groups such as the Fulani, which has an average search volume only 22 per cent that of the Igbo over the past five years.  As seen in Figure 5 the correlation between the DHS and Google data is very low, largely due to missing data.  However, this problem does not arise from even smaller groups such as the Bini, Ibibio, Ijaw, Tiv and Urhobo, where there is a perfect match between the state that has the highest relative hit rate and the state with the highest percentage of the relevant ethnic group (namely Edo, Akwa Ibom, Bayelsa, Benue and Delta, respectively).  The inaccuracy associated with the Fulani could be a result of the fact that they have multiple names across Nigeria and West Africa in general, including the Fula, Fulbe and Peul, which could lead to problems in assessing search data.

Figure 7: Fulani
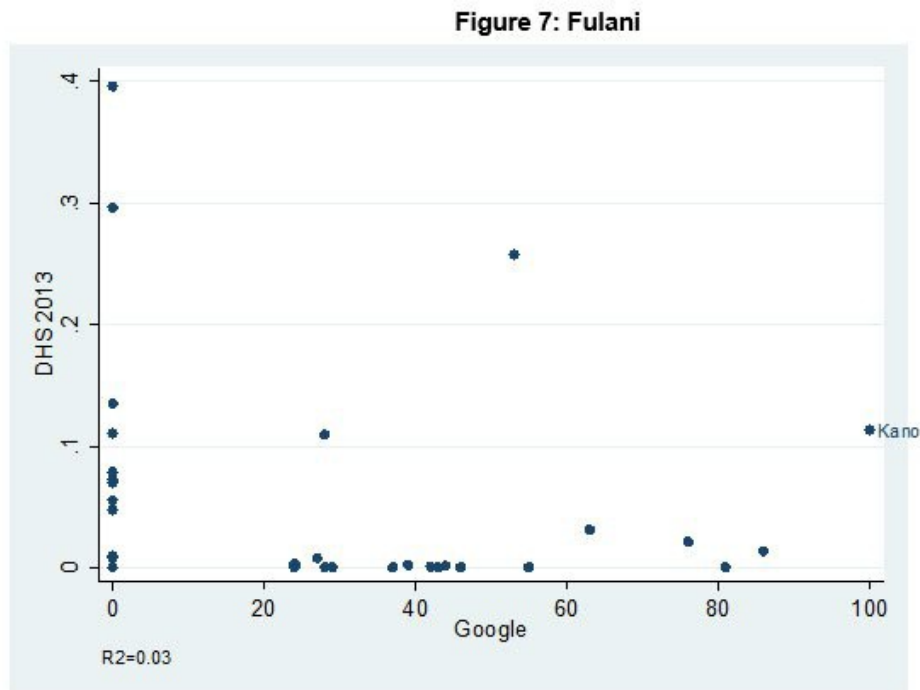
Finally, I should note that GT is not useful for comparing the proportions of groups to each other. Indeed, the DHS suggests that ratio of Hausa to Yoruba, for instance is 1.7 to 1; in contrast the ratio of searches for "Hausa" to "Yoruba" across the past five years is 7.7 to 1.

To conclude, GT can be a useful way to calculate the relative proportion of a given ethnic or religious group across a country, albeit with the caveat that search terms that have smaller absolute volumes can lead to inaccuracies, as with the Fulani and Christians and Muslims above. As Google continues to refine its search data, it is certainly possible that the possibilities for using GT for this purpose will grow in the future.

*Featured Image Credit: pjotter05 via Flickr CC BY-NC-ND 2.0*

**Dr Elliott Green** is Associate Professor in Development Studies in LSE's Department of International Development.

**The views expressed in this post are those of the author and in no way reflect those of the Africa at LSE blog, the Firoz Lalji Centre for Africa or the London School of Economics and Political Science.**