

Ethnicity Estimator

Software for Inferring
Ethnicity from People's
Names

Ethnicity Estimator

Predicting Ethnicity from Family Names

The **Ethnicity Estimator** classifier software is an outcome of a research project which uses family name data assembled by the Consumer Data Research Centre (CDRC). The data are taken from secure commercial consumer data sources and from the Office for National Statistics (ONS), which securely host data for England & Wales.

Ethnicity Estimator software

The underpinning research (doi.org/10.1371/journal.pone.0201774) enables estimates of the ethnic distribution from datasets which contain names, using a peer reviewed methodology. Prospective users can apply to access their own free copy of the Ethnicity Estimator software. This provides aggregate classifications reporting on estimated population for each of the standard ONS ethnicity groups. Accepted applications will allow users to apply the software to their own lists of names pursuant to the public good, and applications are welcome from the academic, government or industry sectors.



Bangladeshi populations in London, 1997



Bangladeshi populations in London, 2016

Changing Distributions of London's Ethnic Minority Communities

Comparing Bangladeshi populations in London, between 1997 and 2016 using Ethnicity Estimator data.

This method makes it possible to predict probable ethnicity at the level of the individual. The CDRC uses nationwide lists of names and addresses, obtained from administrative and consumer sources, to predict ethnicity, and then aggregate the results to smaller scales. This allows us to create highly granular, annually updated measures of neighbourhood composition where examples of change are very difficult to infer from Census data.

