GEDI Guidance: Methods and Measures

This document provides key information on methods for:

- Measuring ethnic diversity and segregation (Reciprocal Diversity Index (RDI) and Index of Dissimilarity (*D*))
- Areal weighting to transfer 1991-2001-2011 data into 2021 LSOAs

Use of this resource should be acknowledged as follows:
Catney, G., Lloyd, C.D., Ellis, M., Wright, R., Finney, N., Jivraj, S., Manley, D. and Nishikido, M. (2024) *GEDI Guidance: Ethnic Groups in the Census: Categorisation and Change*. Geographies of Ethnic Diversity and Inequalities (GEDI). www.gedi.ac.uk

Read more on the growth of ethnic diversity and decreasing ethnic residential segregation in England and Wales in *The Geographical Journal:* https://rgs-ibg.onlinelibrary.wiley.com/doi/10.1111/geoj.12507

This guidance document accompanies the GEDI Census briefing: Catney, G., Lloyd, C.D., Ellis, M., Wright, R., Finney, N., Jivraj, S., Manley, D. and Nishikido, M. (2024) *England and Wales are Becoming Increasingly Ethnically Diverse and Less Residentially Segregated*. GEDI Evidence from the 2021 Census Briefing Series: Issue 1. Geographies of Ethnic Diversity and Inequalities (GEDI). www.gedi.ac.uk

Geographies for analysis

In the GEDI project, the main geographical unit of analysis are Lower Layer Super Output Areas (LSOAs), which we refer to as neighbourhoods. Our analysis also makes use of Local Authority Districts. In the 2021 Census, there were 331 districts, with an average population of 180,053 people, and 35,672 neighbourhoods, with an average of 1,671 people.

Methods for measuring ethnic diversity and segregation

Reciprocal Diversity Index

The Reciprocal Diversity Index (RDI) is one way to measure the ethnic diversity of an area and allows for the comparison of ethnic diversity levels over time[1,2].

RDI is computed as follows:

$$RDI_i = 1/\sum_{m=1}^{M}\left(\frac{N_{im}}{N_i}\right)^2$$

Take as an example 16 comparable ethnic groups from 2001 to 2021 (see *Ethnic Groups in the Census: Categorisation and Change*):

Where there are M ethnic groups (in this case, 16), N_{im} is the number of people in group m in area i (e.g., LSOA) and there are N_i people in total in area i. RDI can be standardised to range from zero to one by subtracting one and dividing by M-1 [2]. To ease interpretation, the output can be multiplied by 100. This allows RDI scores to range from zero (only one ethnic group in an area) to 100 (each group makes up an equal share of the total population – for instance, 1/16th of the total population).

In 2001, the RDI For England and Wales was 2.02 and has gradually increased over time to 3.56 in 2011 and 5.14 in 2021.

Index of Dissimilarity

The Index of Dissimilarity (*D*) captures the unevenness of the spatial distribution of two groups and is the most commonly used measure of segregation[3]. An ethnic group is often compared with all others (e.g., the Pakistani ethnic group is compared with all ethnic groups other than Pakistani) but can also be compared with another group. Values of *D* range from zero to one, where zero indicates a completely even distribution (all areas have the same proportion of the two groups compared) and one indicates a completely uneven distribution (all areas comprise 100% of one group *or* the other). *D* is computed with:

$$D = 0.5 * \sum_{i=1}^{N_I} \left| \frac{N_{im}}{N_m} - \frac{N_{in}}{N_n} \right|$$

Where N_I are the number of zones (in our case, 2021 LSOAs), N_{im} are the number of people in group *m* in LSOA *i* and there are N_m people in group *m* in England and Wales. N_{in} and N_n are the equivalent for group *n*.

Areal weighting method to transfer 1991-2001-2011 data into 2021 LSOAs

The data for the Censuses of 1991 (for which the smallest available output zones are Enumeration Districts; EDs), 2001 (Lower Layer Super Output Areas; LSOAs) and 2011 (LSOAs) can be transferred to 2021 LSOAs using areal weighting[4]. The product of this weighting method is used in the research article: *Ethnic diversification and neighbourhood mixing: A rapid response analysis of the 2021 Census of England and Wales*.

With this approach, the source geography (e.g., 1991 EDs) and target geographies (e.g., 2021 LSOAs) are spatially overlaid, producing a new dataset which shows which parts of each source and target zone overlap. Postcodes are then overlaid on these overlapping segments. The area of each overlapping zone segment is then multiplied by the number of postcodes in each segment (area × pc count). The product is then aggregated by source zone code (sum(area × pc count)). The estimated population is derived with: (area × pc count)/(sum(area × pc count)) × source zone population count. The product is aggregated by target zone code to give an estimate of the population of the earlier period (1991 in this example) on the most recent geography (here, 2021 LSOAs). This approach was used to obtain population count estimates for 2021 LSOAs from 1991 EDs and 2001 LSOAs. For 2011 LSOAs, the approach was slightly different: The 2011–21 ONS lookup table was used to determine LSOAs unchanged between 2011 and 2021. Unchanged LSOAs were removed, and areal weighing was applied to all LSOAs which had changed between 2011 and 2021. The estimates and values for unchanged zones were then merged.

Note that the sum of ethnic group population counts derived from LSOAs (2001, 2011, 2021) differ slightly from published counts[5] due to small cell adjustment procedures intended to protect confidentiality where counts are small[6]. Ethnic group population counts for 1991 used for calculating the Index of Dissimilarity are derived from EDs. The sums of these differ from published totals as not all EDs for which data are published link to physical ED zones[7].

References

Additional helpful resources can be found in the 'Related Articles' section of the GEDI website.

[1] Catney, G. (2016) Exploring a decade of small area ethnic (de-) segregation in England and Wales. *Urban Studies*, 53(8), 1691–1709. Available from: https://doi.org/10.1177/0042098015576855

[2] Simpson, L. (2007) Ghettos of the mind: The empirical behaviour of indices of segregation and diversity. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170(2), 405–424. Available from: https://doi.org/10.1111/j.1467-985X.2007.00465.x

[3] Massey, D.S. & Denton, N.A. (1988) The dimensions of residential segregation. *Social Forces*, 67(2), 281–315. Available from: https://doi.org/10.1093/sf/67.2.281

[4] Catney, G., Lloyd, C. D., Ellis, M., Wright, R., Finney, N., Jivraj, S., & Manley, D. (2023) Ethnic diversification and neighbourhood mixing: A rapid response analysis of the 2021 Census of England and Wales. *The Geographical Journal*, 189(1), 63-77. Available from: https://doi.org/10.1111/geoj.12507

[5] ONS. (2022a) *Ethnic group, England and Wales: Census 2021*. Office for National Statistics. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity/bulletins/ethnicgroupenglandandwales/census2021

[6] ONS. (2022b) *Quality Assurance of Census 2021: How we check the quality of the data*. Office for National Statistics. Available from: https://www.ons.gov.uk/census/planningforcensus2021/qualityassuranceofcensus2021

[7] Majeed, F.A., Cook, D.G., Poloniecki, J. & Martin, D. (1995) Using data from the 1991 census. *British Medical Journal*, 310(6993), 1511–1514. Available from: https://doi.org/10.1136/bmj.310.6993.1511