Grids - the future of small area statistical geographies?

Gale, C

Office for National Statistics

GISRUK 2023

Summary

ONS Geospatial have developed a method of creating small area geographies built around record level data using nested grids. The Python package 'gridgranulator' allows non-disclosive outputs to be produced using a mixture of 125m x 125m, 250m x 250m, 500m x 500m or 1 kilometre grid cells. Building a geography around the data allows for as spatially detailed outputs as possible, whilst the use of grids means the outputs better integrate with other data sources that are more commonly disseminated as raster layers, such as environmental or climate data.

KEYWORDS: 2021 Census, ONS, Grids, Statistical Geographies

1. Introduction

Current small area geographies used in the UK, such as Output Areas, are designed around the requirement to disseminate data that is not disclosive. They are pre-built to meet minimum population and household thresholds and data (such as that from the 2021 Census) are aggregated to them.

The pre-built nature of these geographies however means in some cases they will contain population counts many times greater than the minimum required to meet statistical disclosure requirements. As such, ONS Geospatial have developed a method of creating a small area geography that is built around the data itself.

This method uses a nested grids approach where individual level data (such as at households) can be aggregated to varying-sized nested grids based on a pre-defined rule set. The method produces grid cells that range from 1km x 1km (most sparsely populated areas) to 125m x 125m (most densely populated areas) that all meet minimum thresholds to negate the risk of identifying individuals. This approach of building a geography around the data allows for as spatially detailed outputs as possible, meaning greater insights are possible at the local level.

2. Methodology

The concept of a nested grid geography is to disseminate univariate statistics using either 125m x 125m, 250m x 250m, 500m x 500m or 1 kilometre grid cells based on minimum thresholds being met. ONS Geospatial have developed a Python package called 'gridgranulator' that is available on GitHub (https://github.com/ONSgeo/gridgranulator). The package takes point data (such as that at Unique Property Reference Number level) to be aggregated, along with rules upon which the nested grids will be based. The rules aim to produce the most granular output possible that conform to the set disclosure thresholds. This also includes record swapping between cells (constrained to their 1km parent grid cell) – so cells that are a few records short of meeting threshold requirements can be moved from adjacent cells to prevent aggregation of that cell to its parent grid cell.

This methodology when applied to a dataset will produce a unique combination of cells sizes that maximises the granularity of the output. However, this means outputs using this method can not be easily compared with each other due to different make-up of grid cell sizes. As such, the methodology also produces a 1km output that can be used for direct comparisons between datasets.

chris.gale@ons.gov.uk

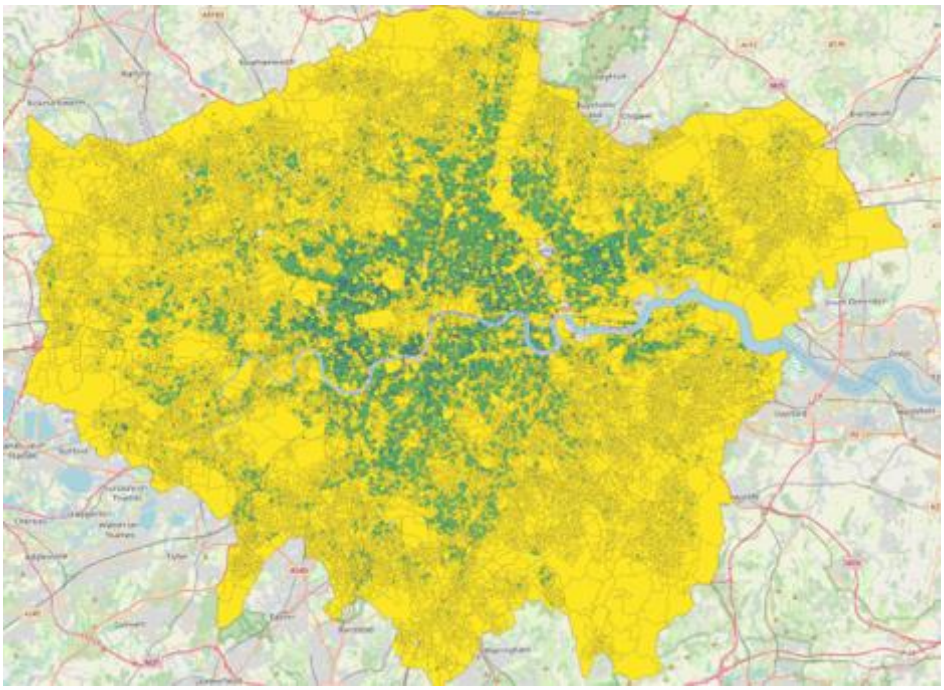
3. Results

Using a synthetic record level dataset, the methodology was applied in multiple areas and the total number of grid cells generated compared to the number of 2021 Output Areas for the corresponding extent. However, to demonstrate the advantages of the flexible nature of the nested grid approach, a minimum threshold of 50 people and 20 households were used (instead of the minimum threshold of 100 people and 40 households used for English and Welsh Output Areas). While this means the results are not directly comparable, Table 1 showcases the increased granularity possible using the nested grids methodology when adopting the same thresholds as Scottish Output Areas. In addition, this is achieved in a matter of minutes compared the months it takes to produce an updated Output Area geography.

Table 1 Number of Output Areas and Grid Cells generated

Area	Number of Output Areas	Number of Grid Cells
Greater London	25,993	41,062
Cardiff	1,109	2,028
Southampton	783	1,400
Exeter	403	739
Birmingham	3,320	5,972
New Forest	591	1,380
Cornwall	1,836	5,214
Dorset	1,272	3,706

Another advantage of the nested grid approach is cells with zero counts are permitted. When comparing Figure 1, showing Output Area level data for Greater London with Figure 2 that shows a nested grids output, it clearly shows large greenspaces and natural corridors (like those along the River Lee in the North East). As such, outputs generated using the nested grid methodology better integrate with other data, such as environmental, where they are more likely to be captured or modelled as grid cells and/or disseminated as raster layers.



chris.gale@ons.gov.uk

Figure 1 Synthetic data assigned to 2021 Output Areas in Greater London.

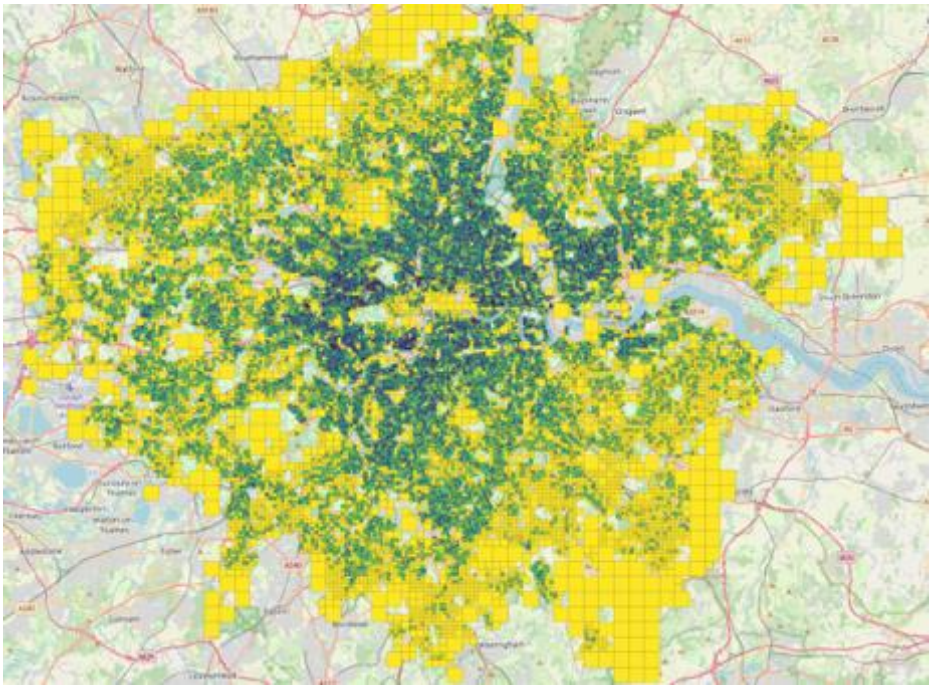


Figure 2 Synthetic data assigned to nested grids in Greater London.

4. Conclusion

The creation and testing of a nested grid methodology by ONS Geospatial demonstrates the feasibility of creating on-the-fly small area geographies around record level data. This is a significant deviation away from existing small area geographies used to report small area statistics in the UK, where the geographies are pre-defined and data aggregated to them at a later stage. The creation of the 'gridgranulator' Python package, its continual development and sharing via GitHub means anyone can adopt the methodology for their own use. While the bespoke nature of each output has potential drawbacks, the benefits of creating the most granular non-disclosive outputs and enhanced interoperability with other data sources mean nested grids can play an important role in the future dissemination of small area statistics.

Biography

Chris Gale is Head of Geospatial Analysis and Innovation at the Office for National Statistics. His work focuses on the championing of geospatial data, techniques and standards as part of the production of official statistics. Previously he has held research roles at the University of Southampton as part of the Administrative Data Research Centre for England (ADRC-E) and at UCL. He also created the 2011 Output Area Classification as part of PhD.

chris.gale@ons.gov.uk