The Age Demographics of Populations and Land Use

Executive Summary

Demographic factors including age may impact land use, and land use may affect the age profile of a population. This study considers two questions involving the age profile of populations:

1. How much do the age profiles of different regions of England and Wales differ?
2. Are there a higher proportion of under 35-year-olds in more urban areas of England compared with more rural areas?

A dataset from the 2021 census of the usual resident population by sex and five-year age group of local authorities (LAs) in England and Wales was obtained from the ONS. This contained estimates for LAs within England and Wales, which were aggregated into regions. The estimated populations are rounded to the nearest 100 anonymising the data.

The second dataset used came from the Ministry for Housing, Communities & Local Government and contained estimates in hectares of the land use in LAs in England in 2018. Land uses are classified across categories, broken up into groups split between developed and non-developed uses.

The census data was used to produce population pyramids. Examination of these pyramids revealed that while there are variations in the age profiles of the regions and Wales they are overall fairly similar with the exception of London. Several of the areas examined (East Midlands, North East, South West and Wales) had shapes indicating they had a higher proportion of older residents than the other regions. London's population was more weighted towards younger age groups.

The census and land use data were combined to answer question 2. The degree of urbanisation was determined by the proportion of developed land in an LA and its population density (estimated number of people per hectare). A clustering k-means algorithm was used to classify the LAs into four levels of urbanisation and these were compared against whether an LA had a proportion of its population under 35 which was above or below the mean proportion. Results showed evidence that more urban areas are likely to have a higher proportion of under-35-year-olds compared with more rural areas.

Limitations include: the land use data was older than the census data so there may have been changes to the land uses by 2021 but these are likely to be minor so do not negate the findings; the 2021 census may have been affected by the covid-19 pandemic which may mean findings for question 2 are underestimated.

Author: K. Carpenter 26/05/23

Aims and Objectives

This report looks at the relationship between the age demographics of a population and land use in the UK. Two datasets were combined. The first was population data for age and sex in LAs and regions in England and Wales in 2021. The second was of estimated land use according to land use categories in English LAs from 2018.

Two questions are explored: how much do the regions of England and Wales differ in terms of the age profile of their populations, and do more urbanized areas in England tend to have a higher proportion of the population aged under 35 compared with more rural areas?

Background

The links between demographic factors and land use are well-established (Hoffmann, 2020). This report focuses on the ages of a population. Age may affect land use, for example where there is a large population aged 0-16 more land may be needed for education purposes or where there is a large elderly population more land may be needed for healthcare. Land uses of an area may be one factor that determines whether people of certain ages move to or away from that location.

A limitation is that the census data on age is from 2021 while the land use data was recorded in 2018. There may have been changes to the land use data in the intervening years to 2021 but given this is a short period this is unlikely to have been significant enough to affect my findings.

Data sources

The population data was gathered during the 2021 census and is available from the ONS website . The land use data comes from what was the Ministry of Housing, Communities & Local Government and is available from the gov.uk website. Both datasets are published under the Open Government Licence v3.0 . This license allows anyone to "copy, publish, distribute and transmit the Information; adapt the Information; exploit the Information commercially and non-commercially" if the source is acknowledged and a link to the license is provided where possible. All analysis done for this report is within the terms of the license.

The land use dataset covers land use of 326 LAs and 9 regions in England and the total for England. The land use was split into 28 categories in 13 different groups and estimations of the number of hectares for each usage were provided. More details about how this data was cleaned and processed are available in TMA 02 (Carpenter, 2022). The data's source is Ordnance Survey Ltd and it was assessed against UK Statistics Authority assurance standards.

The population data contains estimates of the usual resident population of England and Wales broken down by sex and five-year age groups. It contains details for regions within England (9), and 331 LAs in England and Wales and totals for England and Wales. The data contains no missing values.

The data was gathered as part of the 2021 census conducted by the ONS. I cannot independently verify it, but the methodology used for collecting English and Welsh census data is well-established and robust with the census return rate being 97% (ONS, Date unknown, *About the Census*). The census was conducted during the covid-19 pandemic which may have affected some peoples' area of usual

residence particularly "students and people in urban areas" (ONS, Date Unknown, *Quality and methodology information for Census 2021*). This may have some impact on my ability to answer the question whether urban areas tend to have a higher proportion of the population aged under 35, as the majority of students are under 35 and I am looking at urban areas in particular. So, any findings here may be an underestimate due to the effects of the pandemic.

The population data is rounded to the nearest 100. This prevents it being used to identify individuals in small populations and so the data is anonymised. In theory the land use data with other searches could identify a person who owns a particular piece of land, but this is not a problem as ownership of property is on public record through HM Land Registry. Combining the population data with the land use data should not carry any risks of identifying anybody due to the rounding in the population data.

Analysis pipeline

Both the land use data and the population data were cleaned using pandas before being imported into and stored in MongoDB. Data was retrieved from MongoDB and processed using pandas.

Census Data

Regions and totals of LAs summed when grouped into regions didn't always equal each other due to rounding. I used LA totals aggregated to find regional totals. I decided to treat Wales as an area comparable to English regions due to similar number of LAs.

For the census data I only needed data for LAs so I removed other categories which aggregate the LAs, prior to importing it into MongoDB; removing the Metropolitan Counties, Counties, Regions, Inner and Outer London, and countries.

The LAs vary quite a lot in size of population. In order that they could be compared I calculated proportions of each age group of each sex relative to overall population in each LA.

Land Data

For the land data I did not need the regional totals, so retained only the data for LAs. I calculated the proportion of developed land use in each LA as a percentage of the total land use.

Combined Data

Combining the population data with the land data was straightforward as most LAs in England were present in both datasets. I excluded the Welsh LAs as the land use data only covered England. The boundaries of several LAs had changed between 2018 and 2021, with smaller LAs being combined into one LA. To ensure the land use data reflected these changes I added together the land use statistics for the smaller LAs and created the new LA from these totals. I do not think this will cause any major problems with the analysis.

In order to classify LAs into more and less urbanized areas I used two indicators: the proportion of developed land use in an LA, and the population density of each LA. The population density was calculated by dividing the total population for an LA by its total land area and is a measure of the number of estimated people per hectare of land.

In order to see if the LAs which were more urbanised also correlated with the LAs which had a greater proportion of their population aged under 35 I produced three choropleth maps using Folium. One

each for my two defining measurements of urbanisation, proportion of developed land as a percentage and population density. Then another showing the proportion of population aged 0-34.

To classify the LAs into groups of more or less urbanised areas I used a k-means clustering algorithm. The input to the algorithm was the proportion of developed land use in an LA along with its population density. I tried a number of clusters (3-5) The mean silhouette statistics were:

Number of clusters	**Mean Silhouette Statistic**
3	0.6946518626290309
4	0.637747008242218
5	0.6364939661203194

Table 1: Mean Silhouette Statistics

While three clusters had the highest value it did not differentiate sufficiently between the LAs with the lowest proportion of developed land use and population density which were the vast majority of LAs. Thus, I determined 4 clusters worked the best for my purposes and accurately represented the data as can be seen from the silhouette plot.

Figure 1: Silhouette plot for k=4

The four clusters can be seen in this scatterplot also indicating four clusters are appropriate for the data.

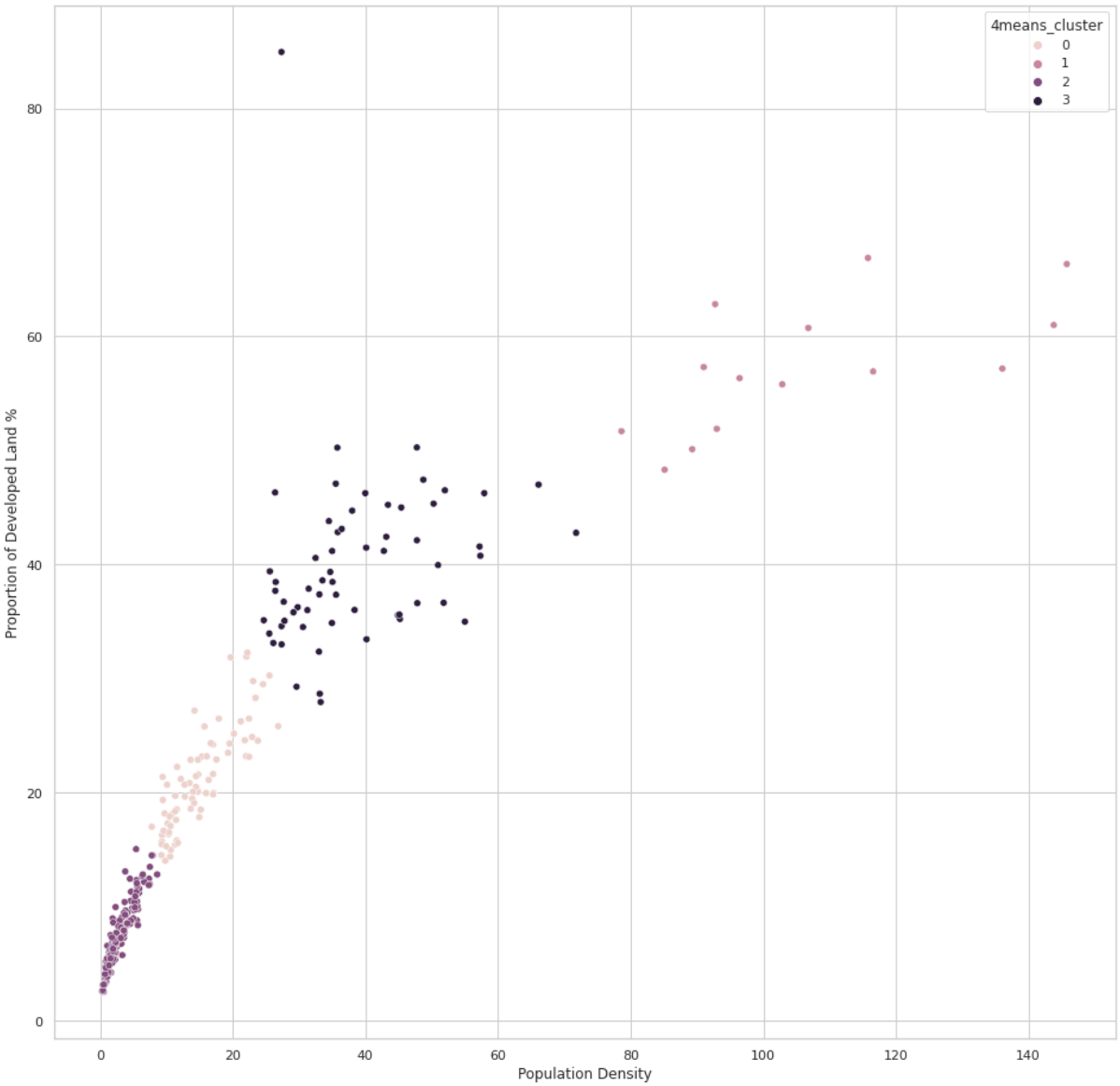


Figure 2: Scatterplot of LAs by population density and proportion developed land.

The clusters indicate the least urbanised (cluster 2) to most urbanised areas (cluster 1).

I then determined whether each LA had a proportion of population of younger people, which I determined as being aged 0-34, equal to or above the mean of 0.4111009398543328 or below the average. Thus, splitting the LAs into 2 groups. In order to see whether more urbanised areas were more likely to have a higher proportion of younger people than less urbanised areas I performed a χ^2 test to test the null hypothesis that the degree of urbanisation and whether the proportion of the population aged 0-34 is above or below average are unrelated.

Findings

Regional Differences in Age Profiles of Populations

In order to examine the age distribution of populations in the regions of England and Wales I produced population pyramids showing the proportion of each five-year age group broken down by sex.

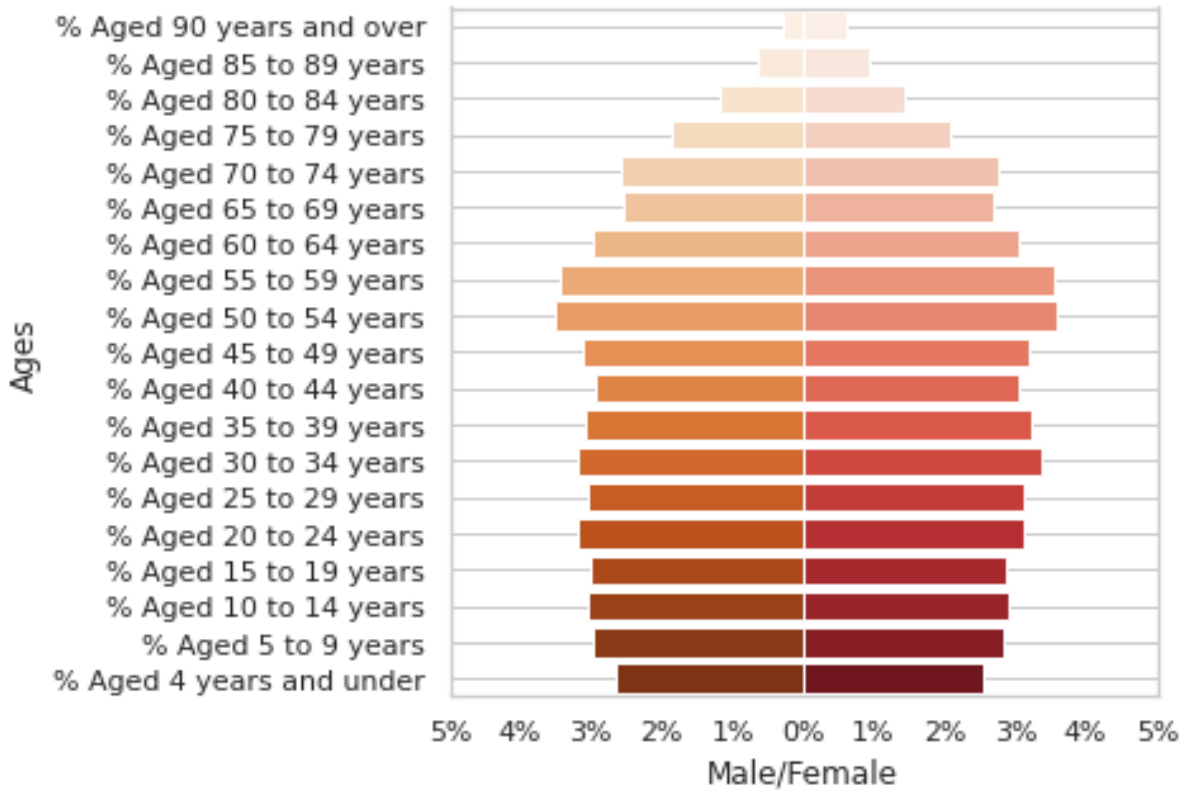


Figure 3: Population pyramid for East Midlands

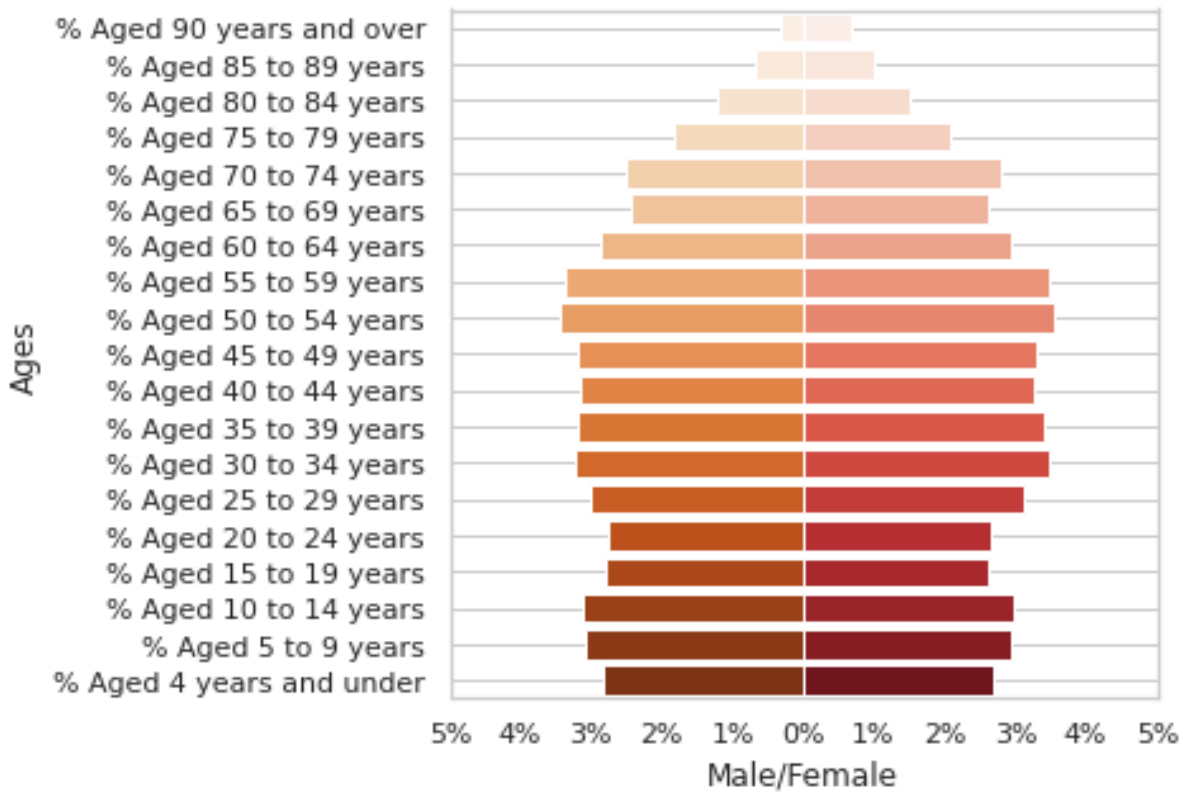


Figure 4: Population pyramid for East of England

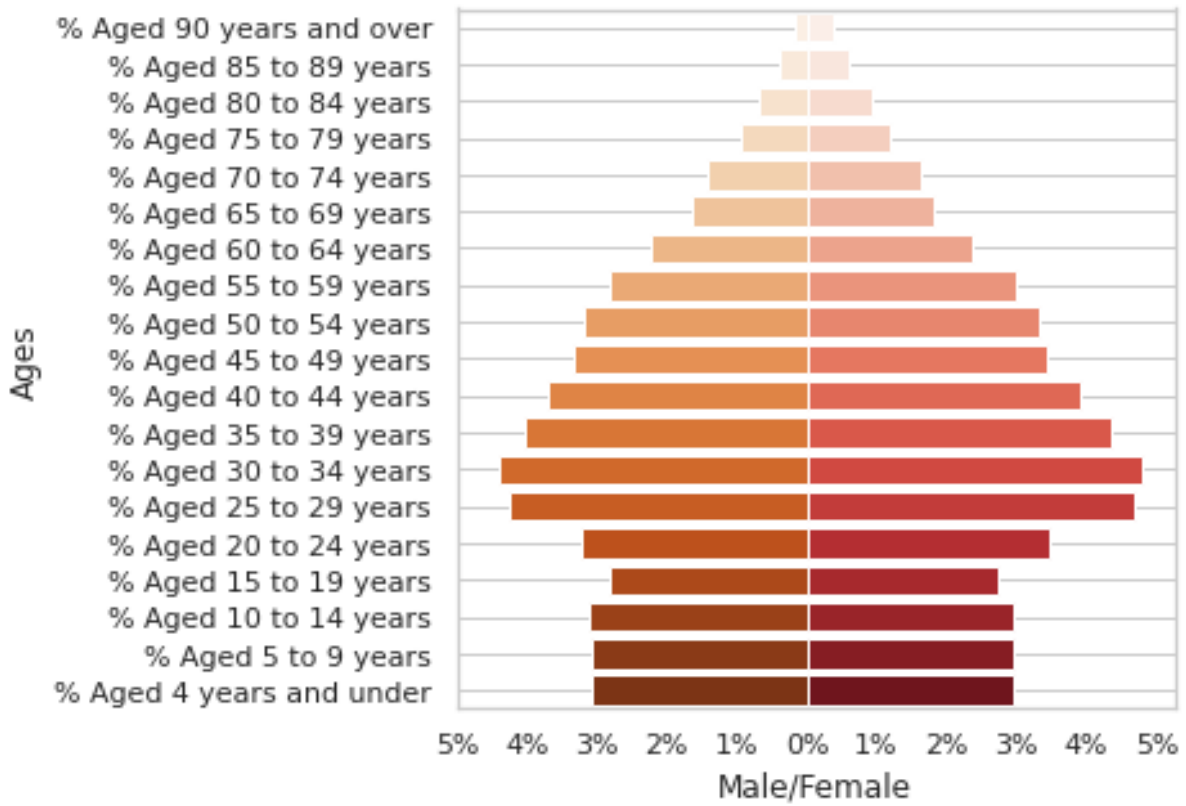


Figure 5: Population pyramid for London

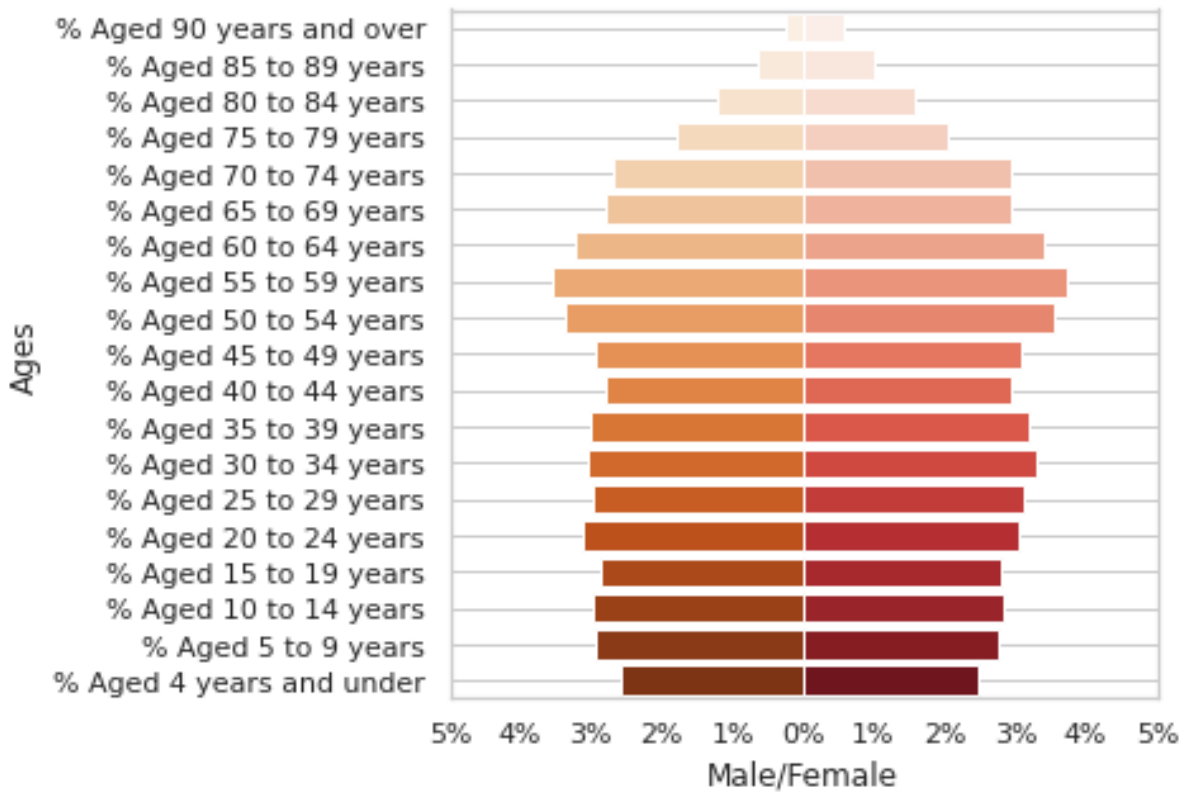


Figure 6: Population pyramid for North East

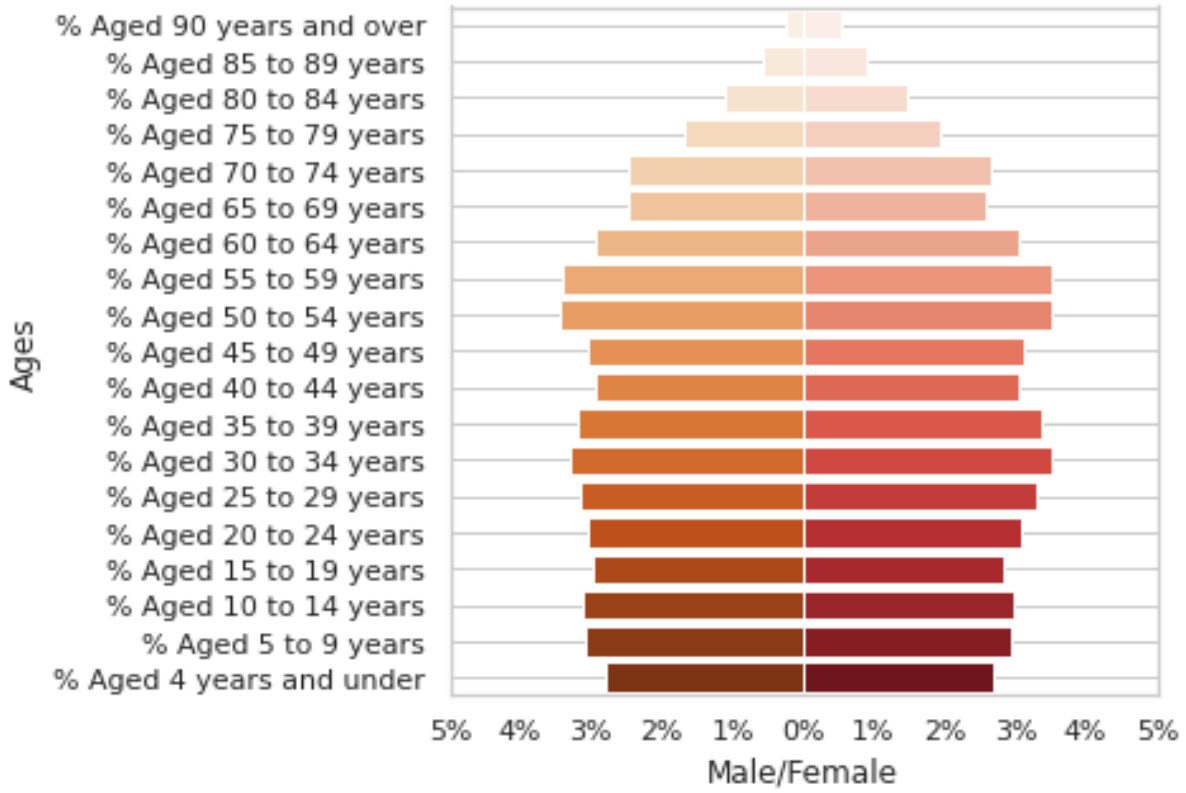


Figure 7: Population pyramid for North West

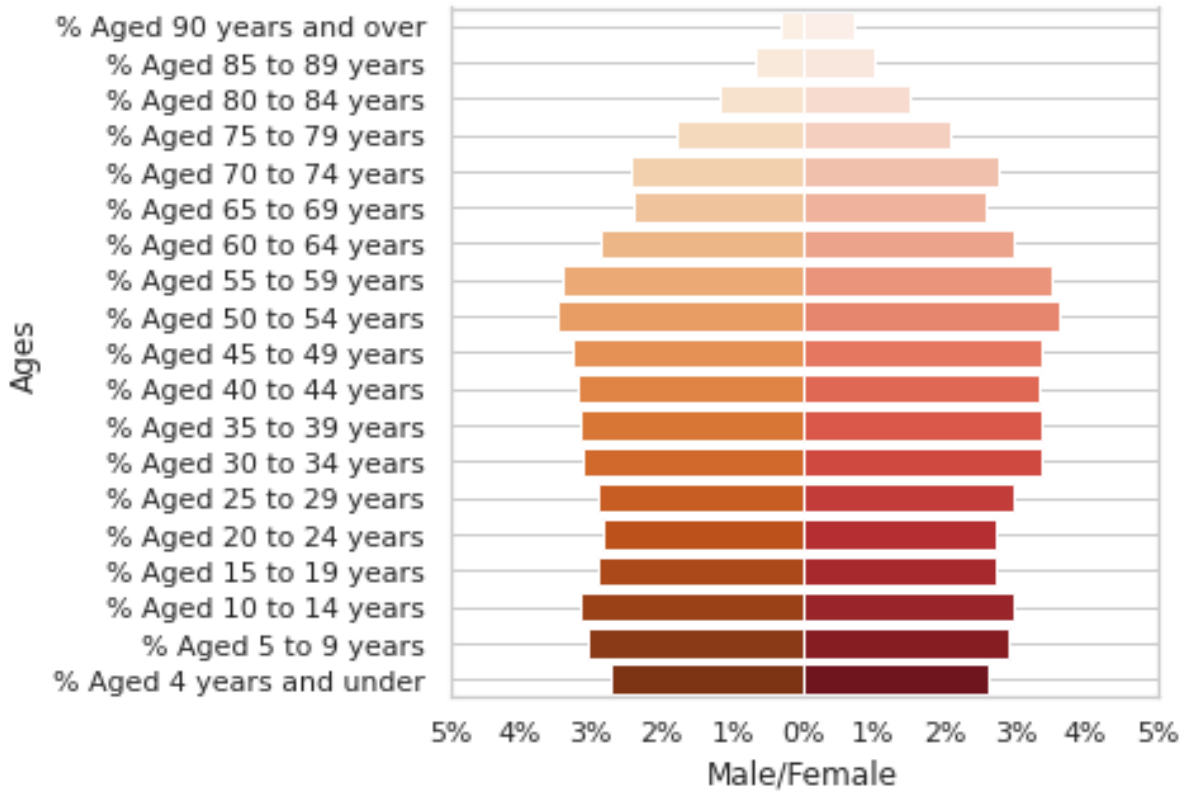


Figure 8: Population pyramid for South East

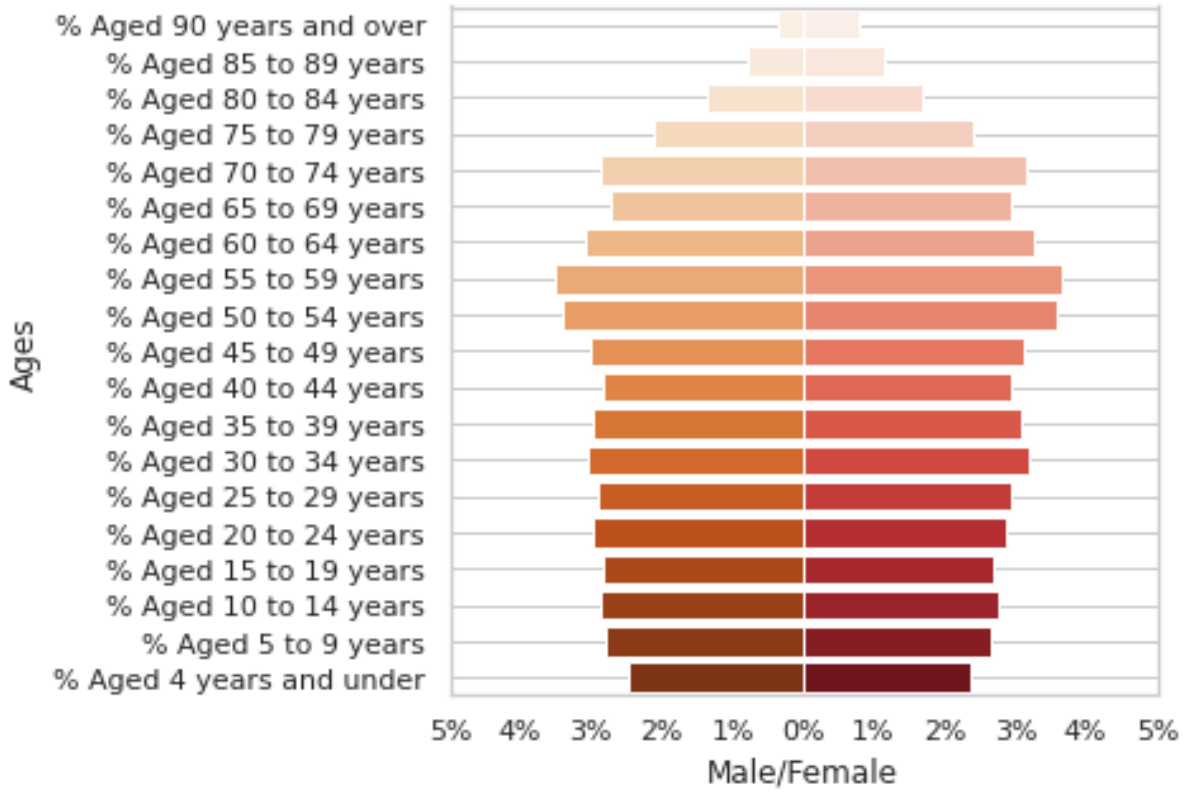


Figure 9: Population pyramid for South West

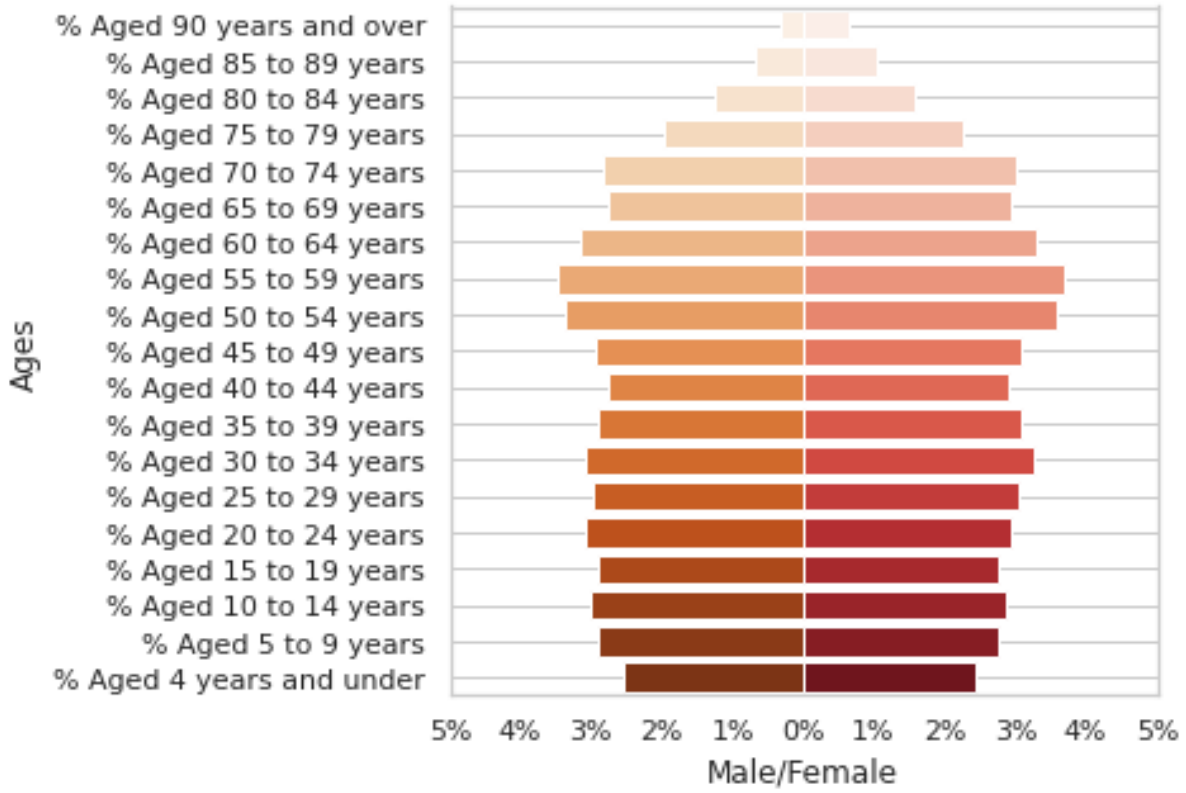


Figure 10: Population pyramid for Wales

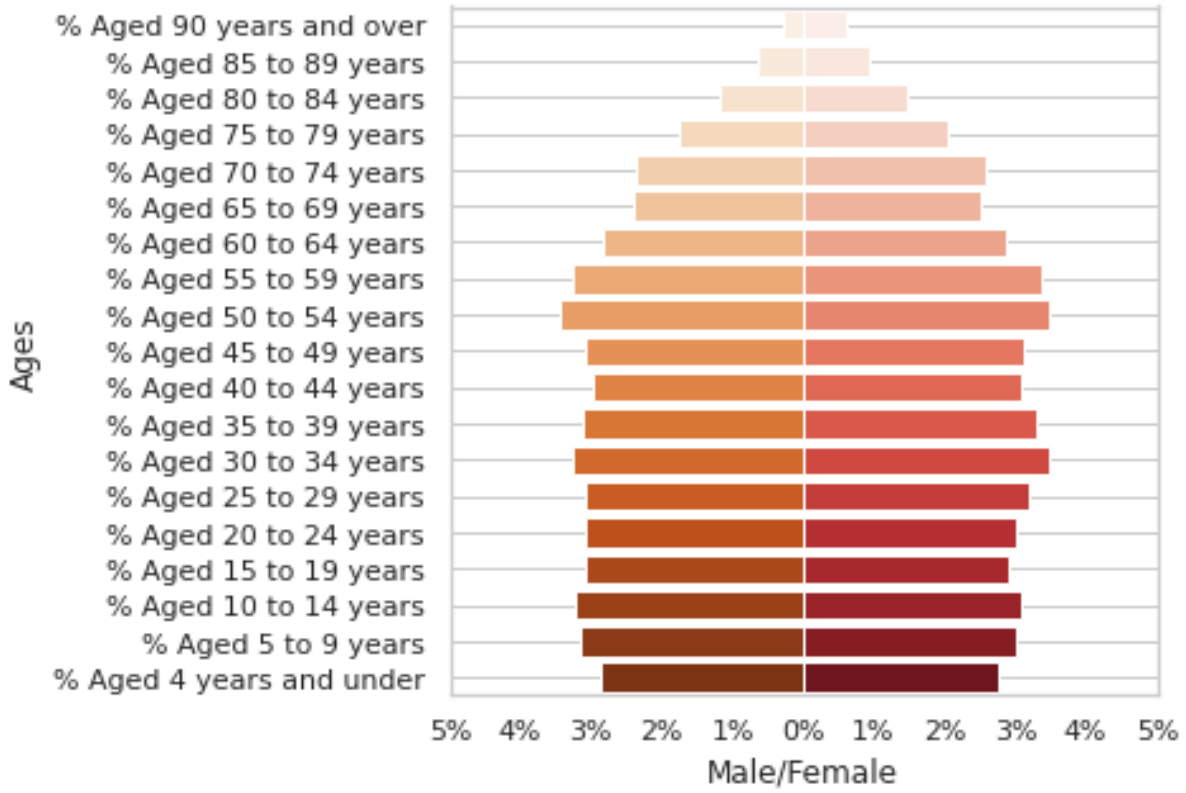


Figure 11: Population pyramid for West Midlands

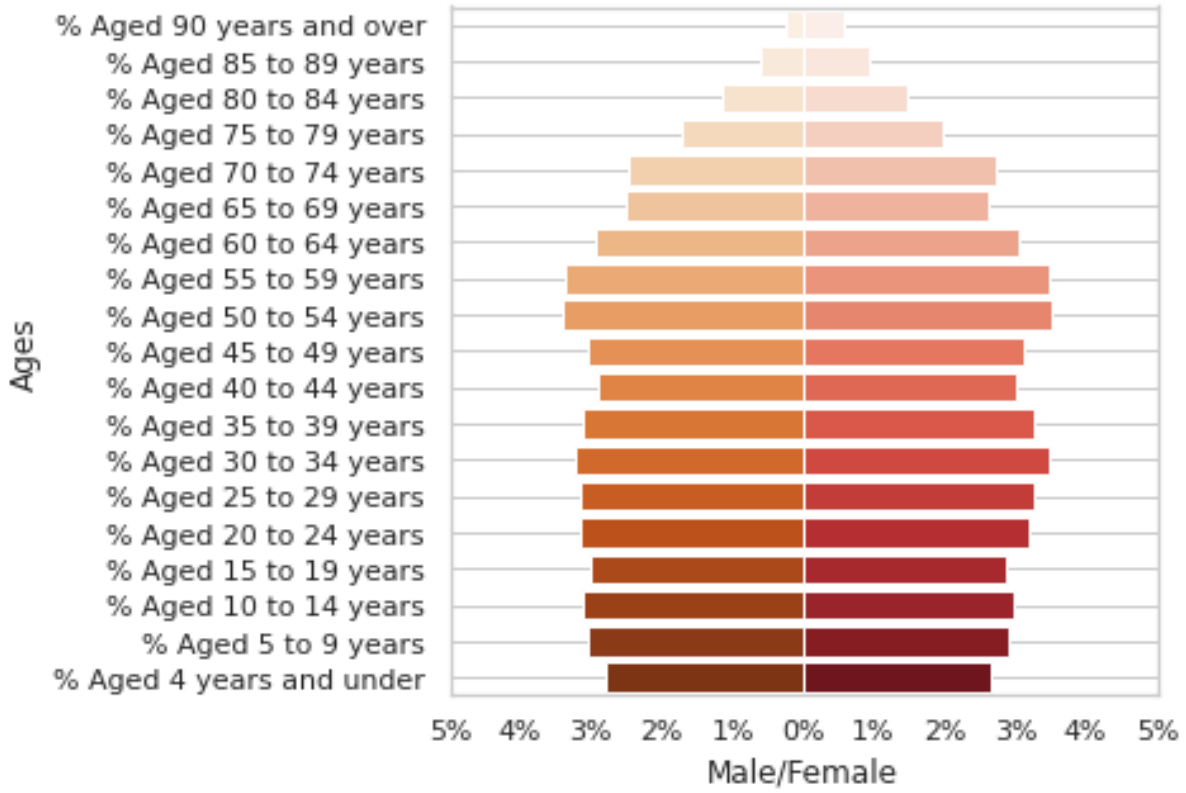


Figure 12: Population pyramid for Yorkshire and the Humber

The population pyramids for much of the regions are quite similar apart from London. They show the proportions of children seems to be between 4 and 6% of the overall population. There is often a small bulge around ages 10-14 presumably reflecting a small “baby boom”. Some regions show a small dip in proportion of people aged 40-49. There is a bulge for all regions except London at either age 50-54 or 55-59 before the proportions steadily decline as the age groups increase.

The population pyramids for the East Midlands, North East, South West and Wales all show a more pronounced “arrowhead” shape at the top of the population pyramids for ages 50-90+. This shows they have a larger proportion of the population over 50 compared with the other regions.

London with its “Christmas tree” shape is different from the other regions. There is fairly large proportion of children then a big jump in the proportions of 25-34 years old (presumably reflecting people moving to London for work or studying) which then declines quite quickly as the age groups increase. Thus, London looks to have a younger population on average compared to the other regions.

Degree of Urbanisation and Proportion of Population under 35

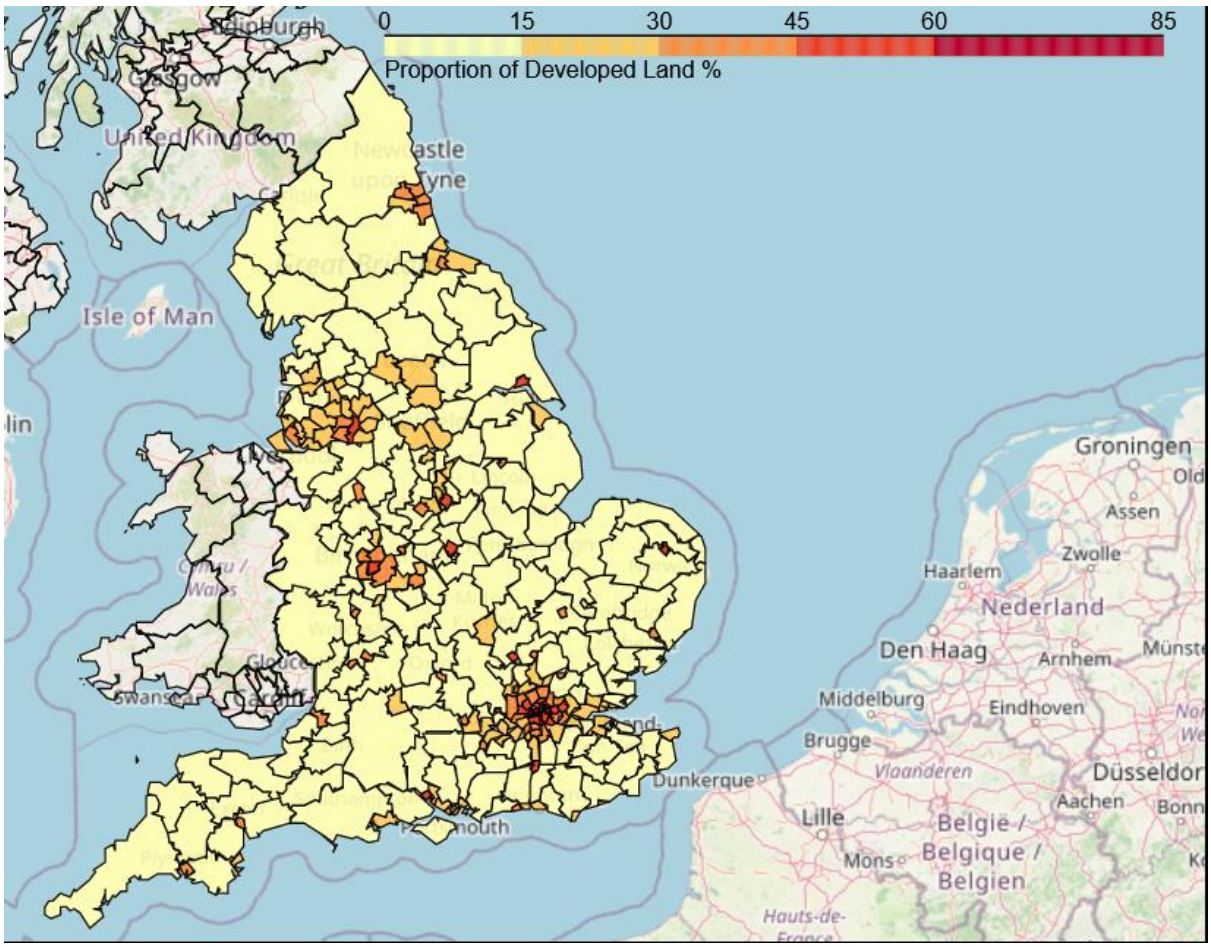


Figure 13: Choropleth map of England showing proportion of developed land as % by LA.

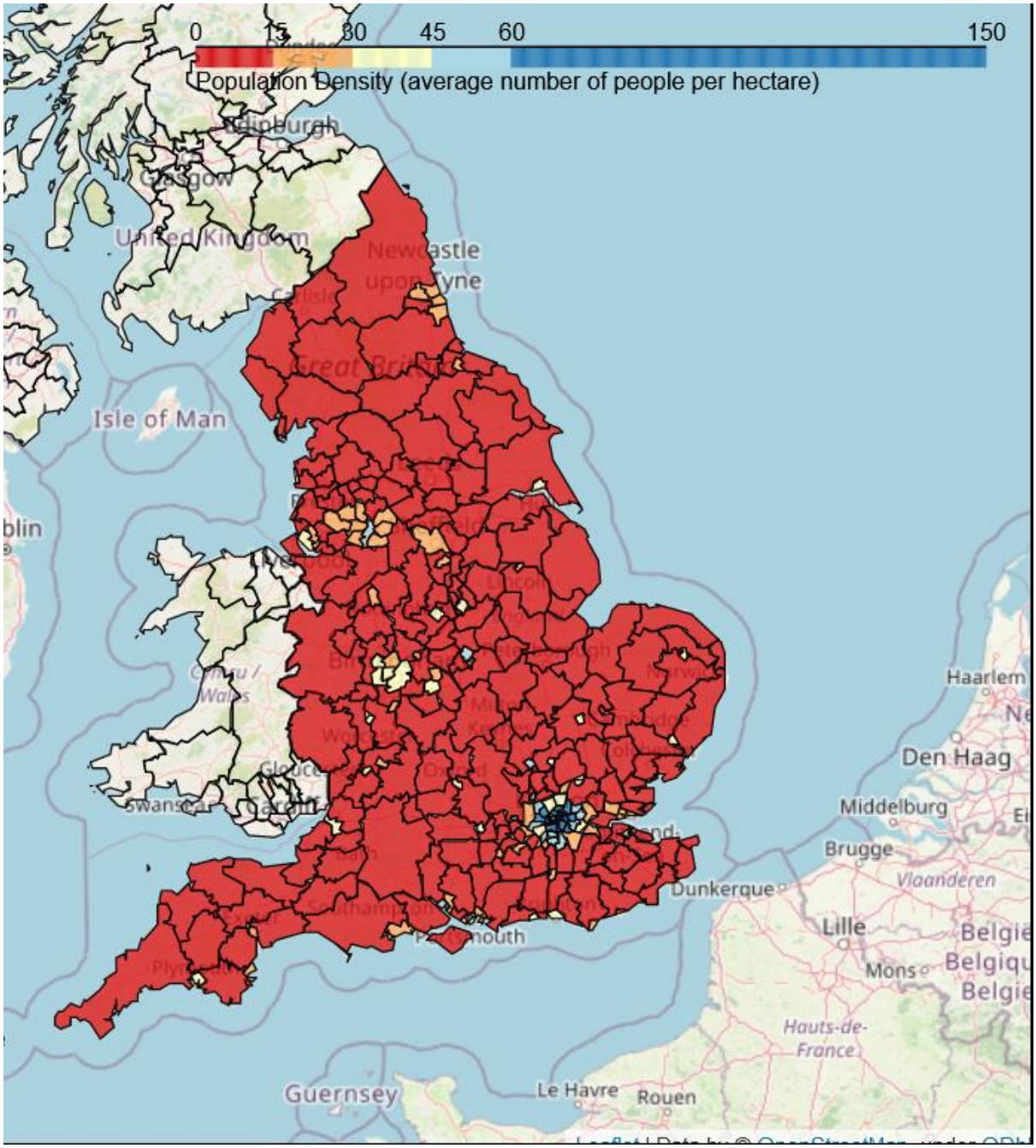


Figure 14: Choropleth map of England showing population density by LA.

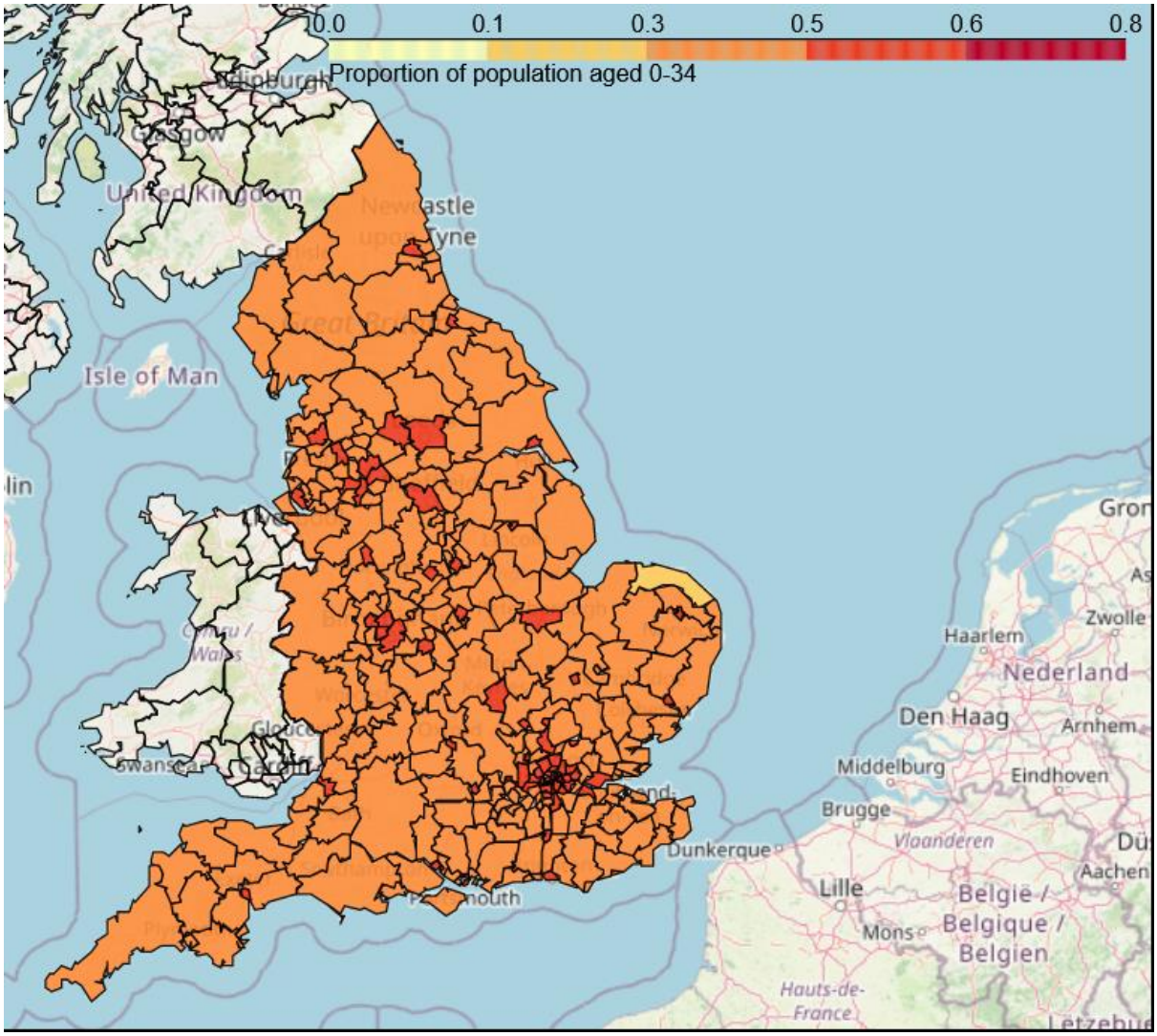


Figure 15: Choropleth map of England showing proportion of population aged 0-34 by LA.

From the choropleth maps above you can see that LAs that have a high proportion of the population under 35 do seem to correlate quite closely to LAs which have a high population density and also to LAs with high proportions of developed land. This seems to indicate that more urbanised areas (according to my definition) do have a higher proportion of their population under 35 compared with less urbanised areas.

When the LAs were classified into four different clusters according to the degree of urbanisation using the k-means clustering algorithm, and each LAs was split into those with proportions of population aged 0-35 above (and equal to) or below the average proportion a χ^2 test was performed. The outcome was a χ^2 statistic of 118.958 and a p-value of $1.2936252717544955\times10^{(-25)}$. The p-value is very small which offers some evidence against the null hypothesis that the degree of urbanisation and whether the proportion of the population aged 0-34 is above or below average are unrelated.

Conclusions

In answer to the question how much the regions of England and Wales differ in terms of the age profile of their populations the population pyramids show some variation in age demographics across the regions of England, and Wales. The variations are relatively minor except in the case of London where the age profile is distinctly different compared with the other regions. Further research is needed to determine how much variation there is between LAs in terms of age profiles of their populations.

For the question do more urbanized areas in England tend to have a higher proportion of the population aged under 35 compared with more rural areas, my findings do suggest this is likely to be true at LA level. The choropleth maps show correlations between LAs which have a high proportion of the population under 35 with LAs with more developed land and higher population density. When I classified the LAs into four groups of increasing urbanisation and two groups of according to proportion of younger people and performed a χ^2 the outcome suggested that the two factors were related. The impact of covid pandemic of the 2021 census may mean that these results may be slightly understated as students and normally urban-dwelling people may have been temporarily located elsewhere at the time the census was taken.

One confounding factor relating to younger people and urban areas may be the presence of universities. Universities contain many young people and in the UK are often located in urban areas. It may be that the findings here are not replicated if you look at more urbanised LAs in the UK which do not have universities.

Reflection

I didn't encounter any issues when performing the analysis. The data required minimal cleaning. I chose to store the data in MongoDB rather than a non-relational database. This choice was made to retain the features of the original data and because it allowed data for each local authority to be stored in a single document which could then be easily aggregated into regions. Combining the two datasets was straightforward except where some LAs had changed by 2021 but here I aggregated and summed the land use data so that the two datasets were comparable.

In terms of differing age profiles between the regions I was surprised to see how small the differences were between regions (except for London) but perhaps smaller differences were tricky to see using population pyramids. To see if there were differences between areas it would have been good to look at LA level. One way would be to calculate the median age of population for LAs, but this would probably not be possible with the census data used as it would only give you which five-year range the median age sat in.

The population pyramids showed how different London was from the other regions. In addition, when looking at more or less urbanised areas all the LAs in the most urbanised category were in London. Given London's large population it would be interesting to see whether the finding that more urbanized areas in England tend to have a higher proportion of the population aged under 35 compared with more rural areas still holds true if you exclude London from the analysis.

For the k-means analysis I did not choose the statistically optimum number of clusters as I wanted to counter the fact a lot of LAs have similarly low levels of population density and developed land. I also had some qualms with my use of the χ^2 test. I wanted a way to test whether seeming correlation

observed in choropleth maps was statistically significant. The test is suitable for categorical variables but the category definitions (more/less urbanised, above/below average proportion of under 35s) were all defined by me and the data comes from different sources which may render it less robust as statistical analysis.

It might be good to produce a single measure of urbanization (combining population density and proportion of developed land) and use this to measure correlation with proportion of population under 35. This might be more statistically robust than my k-means analysis and χ^2 test.

[2999 words]

References

Carpenter, K. (2022) *TM351 TMA 02*, submitted to The Open University as part of TM351 assessment.

Department for Levelling Up, Housing and Communities and Ministry of Housing, Communities & Local Government (2020) Live tables on land use. Available at: https://www.gov.uk/government/statistical-data-sets/live-tables-on-land-use (Accessed: 26th May 2023).

Hoffmann, J. (2020) 'Demographic Change and Land Use', in T. Weith, T. Barkmann, N. Gaasch, S. Rogga, C. Strauß, J. Zscheischler, (eds.) *Sustainable Land Management in a European Context*. Cham, Switzerland: Springer, p.63.

Ministry for Housing, Communities & Local Government (2020), Land Use in England, 2018. Available at: https://www.gov.uk/government/collections/land-use-in-england (Accessed: 26th May 2023).

ONS (Date unknown) *About the Census*. Available at: https://www.ons.gov.uk/census/aboutcensus/aboutthecensus (Accessed: 26th May 2023).

ONS (Date unknown) Population and household estimates, England and Wales: Census 2021. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationandhouseholdestimatesenglandandwalescensus2021(Accessed: 26th May 2023).

ONS (Date Unknown) Quality and methodology information for Census 2021
Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/methodologies/qualityandmethodologyinformationqmiforcensus2021 (Accessed: 26th May 2023).

Appendix 1: Notebooks

The notebooks below contain the details of data imports, cleaning, processing and analysis.

Notebook	**Contents**
H2632804_q1_TMA02	Importing of land use data into MongoDB Cleaning of land use data Calculating a proportion of developed land as a percentage of the total land area
H2632804_q2a_lab_notebook-TMA02	Data analysis for appendix 3 Report on Land Use
EMA Project Diary	Importing of census data into MongoDB Cleaning of census data Aggregation of census data into regions Producing population pyramids Aggregating land use data where local authority boundaries had changed Combining land use dataset and census dataset Calculating population density Using k-means clustering algorithm to classify LAs as more or less urbanised Performing χ^2 test on clusters of more or less urbanised LAs to see if independent from proportion of population under 35.

Appendix 2: Data catalogue

The data used to write this report is available under the Open Government Licence v3.0 which can be found at: https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/

The 2021 census data the usual resident population by sex and five-year age group of local authorities (LAs) in England and Wales can be found at: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationandhouseholdestimatesenglandandwalescensus2021 [Accessed 04/06/2023].

The land use data from 2018 can be found at: https://www.gov.uk/government/statistical-data-sets/live-tables-on-land-use [Accessed 04/06/2023].

Appendix 3: Report written for TMA 02 Question 2

Residential Land Use in England in 2018

Aims and Objectives

This report looks at residential land usage in Local Authorities (LAs) in England. It considers how the proportion of residential land usage varies across LAs and whether there is a high proportion of residential land use in large cities, specifically London. It investigates the relationship between the amount of land used for residential purposes and in residential gardens.

Background

The UK is divided into Local Authorities for administrative purposes (Wikipedia, 2022). LAs vary quite significantly in size. The UK government publish data every few years relating to land usage including land for residential purposes.

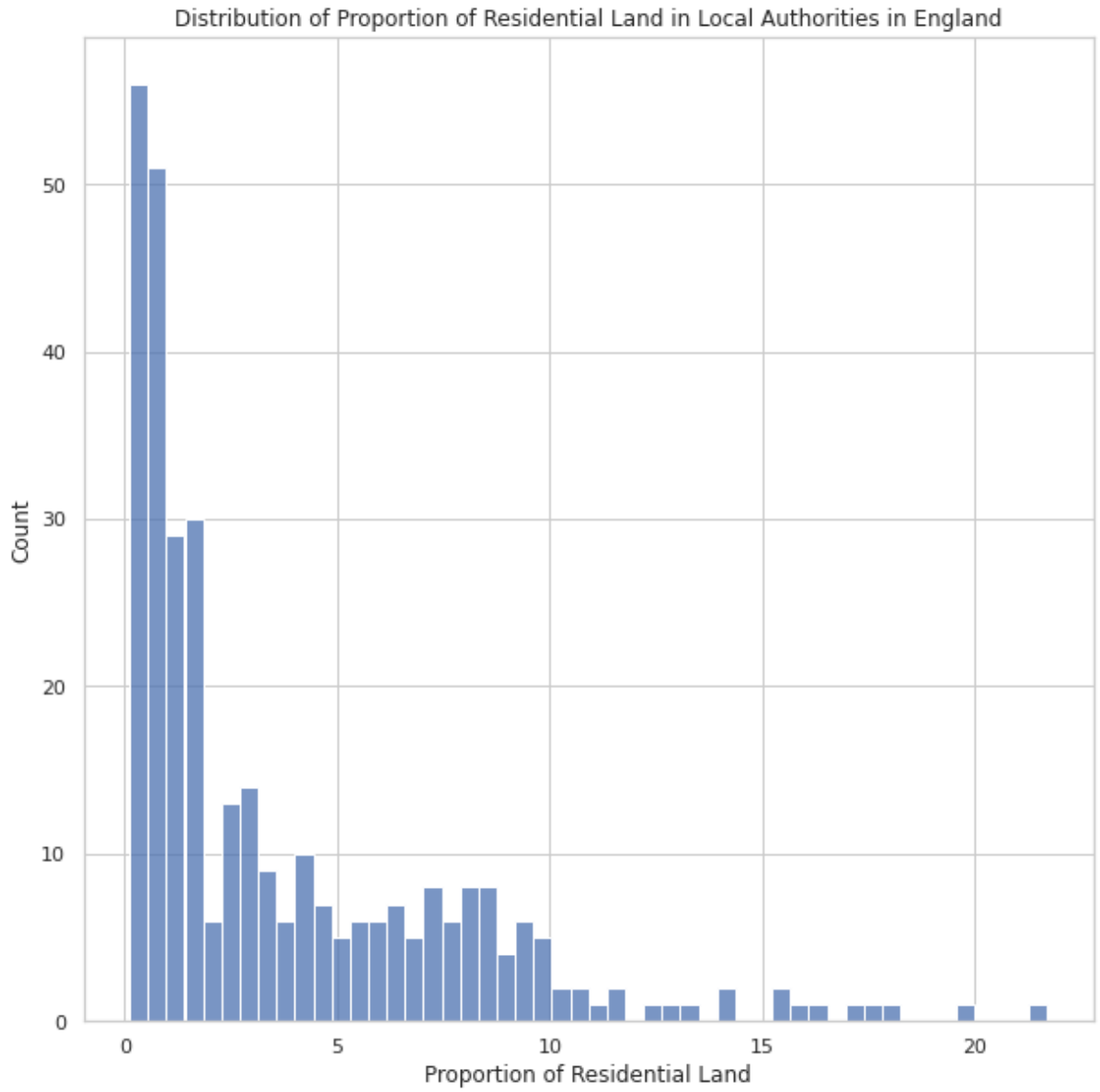
Sources of Data

The data used comes from the Department for Levelling Up, Housing and Communities and Ministry of Housing, Communities & Local Government hosted on the government website and available under the Open Government Licence v3.0. All the analysis done in this report is allowed under the conditions of the licence.

The dataset covers 326 LAs in England and contains estimated hectares of land usage under different categories as it stood in April 2018 and was provided by Ordnance Survey Ltd.

Analysis Pipeline

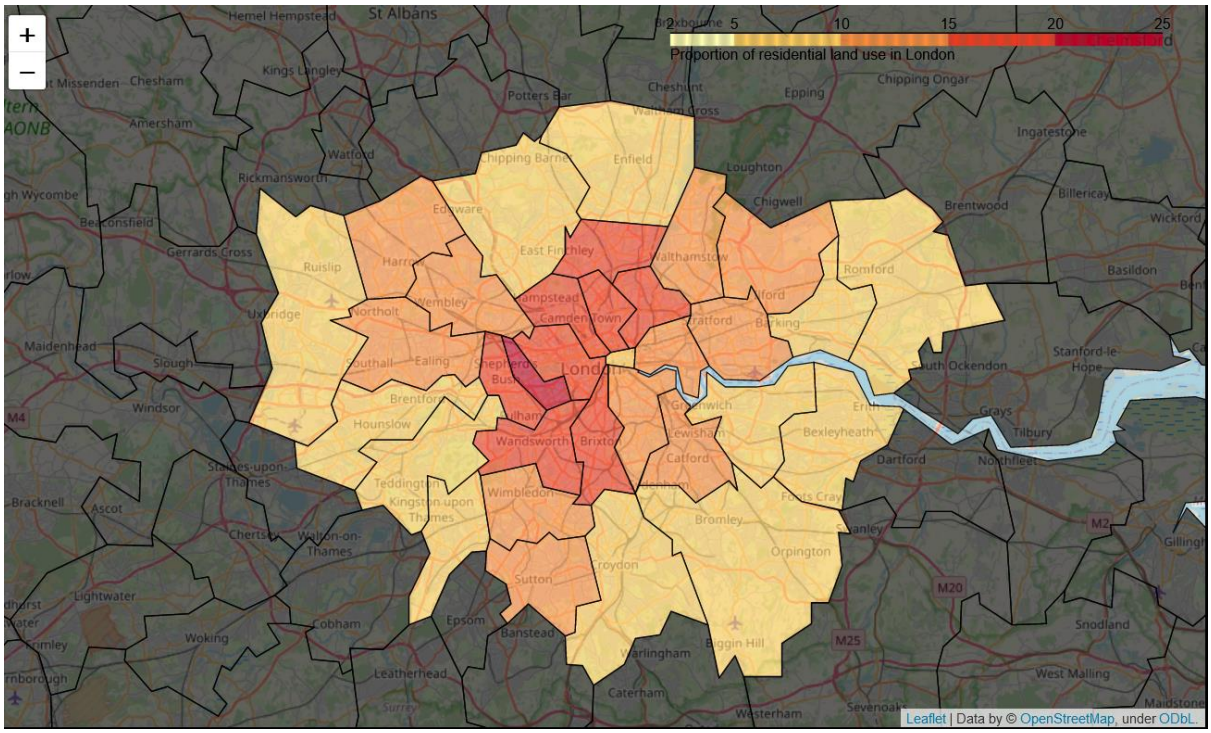
The land use data was obtained from the website above and imported and cleaned in MongoDB. More information about the import and processing of the data can be found in H2632804_q1_TMA02.ipynb and H2632804_q2a_lab_notebook.ipynb (Carpenter, 2023)



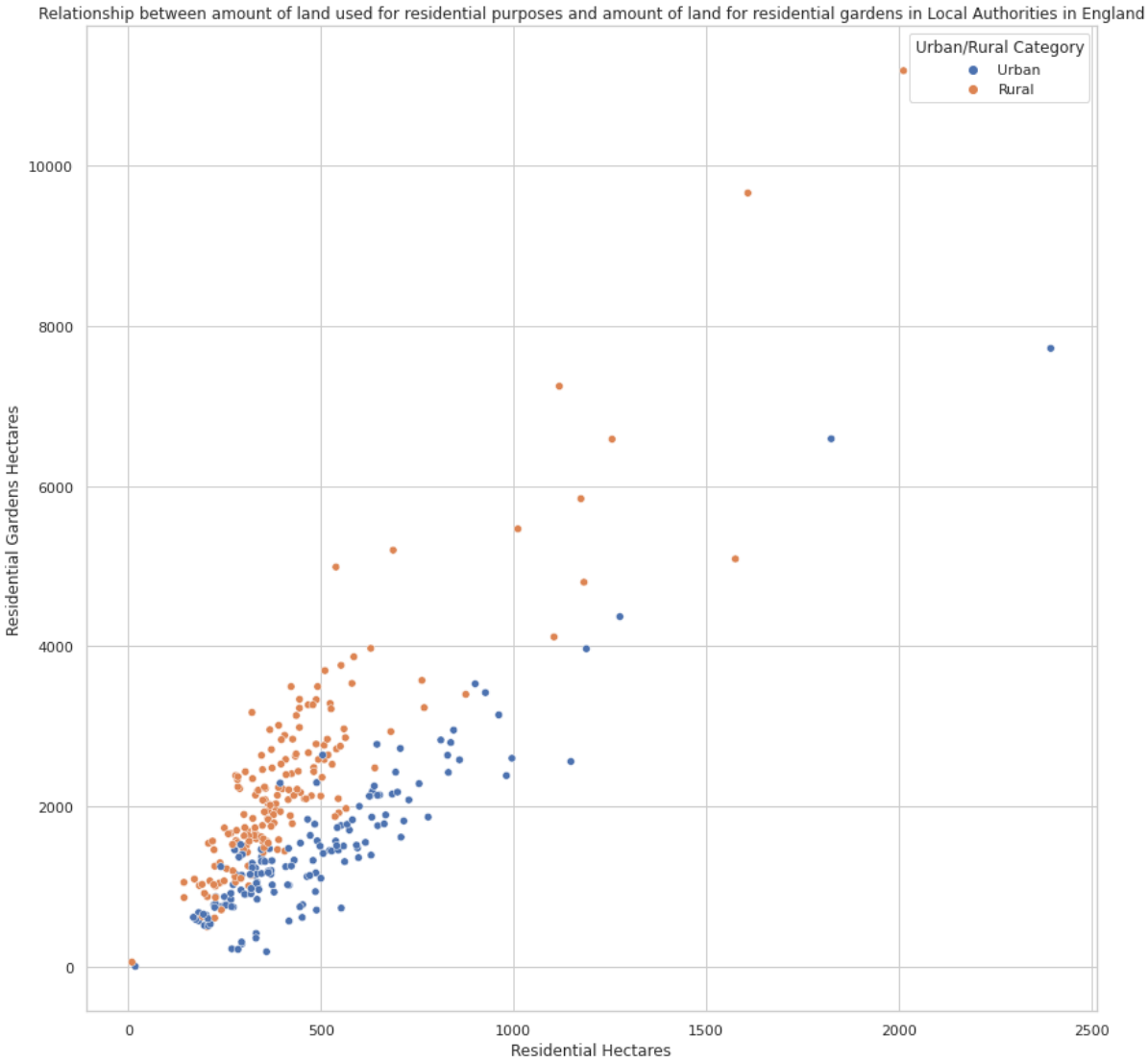
The proportion of residential land use ranges from close to 0% to approximately 22%. The majority of LAS have a small proportion of their land use as residential. Of the 326 LAS approximately 170 LAs have a proportion of residential areas of less than around 2%. A relatively small number of LAs have a proportion of over 10%.

In order to determine whether large built-up and populous cities are likely to have a high proportion of residential land usage I produced a choropleth map of London boroughs and the City of London showing the proportion of residential land use.

Proportion of residential land use in London boroughs



The map demonstrates some London LAs do have higher proportions of residential land use. This lends some weight to my hypothesis that areas with higher residential usage are likely to be found in built-up cities with large populations. There seems to be a trend of higher proportions of residential land use towards the middle of the London map. There is one possible outlier right in the centre of London which only has between 5-10% proportion of residential land use. This is presumably the City of London area which is the main CBD of London and may have a comparatively lower residential land as much of it is made up of commercial and institutional buildings.



The relationship between residential and residential garden land use is positive and linear (r coefficient 0.80, p-value 1.0478×10^{-74}). Rural areas tend to have higher garden usage than urban areas with similar amounts of residential hectares.

Conclusions

The proportion of residential land use varies quite significantly across England although the majority of LAs have less than 2% of land used for residential purposes. Large cities such as London are likely to have areas of high residential land use although there may be caveats to this. The relationship between the area used for residential purposes and residential gardens is strong, positive and linear with some differences between rural and urban areas.

References

UK Government (2020) 'Land Use in England, 2018' [Online]. Available at https://www.gov.uk/government/collections/land-use-in-england (Accessed 27 February 2023).

Wikipedia (2022) ‘Subdivisions of England’ [Online]. Available at https://en.wikipedia.org/wiki/Subdivisions_of_England (Accessed 27 February 2023).

Carpenter, K. (2023) TM351 TMA02 submitted to the OU as part of TM352 assessment.

[600 words]