

[Home](#) ▾ [Parenting, childcare and children's services](#) ▾ [Childcare and early years](#)  
▾ [Providing childcare](#) ▾ [Changes in access to childcare in England](#)



Research and analysis

# Methodology: Changes in access to childcare in England

Published 16 October 2024

## Applies to England

Contents

[Sequence and clustering analysis](#)

[Data](#)

[Data table for figure](#)

## Sequence and clustering analysis

Sequence and clustering analysis is a useful tool to examine patterns within and across categorical time series. It does this by identifying groups with similar experiences. Using this analysis to examine childcare accessibility can show how different populations experience access to childcare and how these patterns change over time.

# Sequence analysis

Sequence analysis uses data over time to examine patterns within and across categorical time series. This analysis is suitable for applying to series of categorical data in order to identify potential patterns, to group participants based on the similarity of their patterns, and to examine differences across pattern groups.[\[footnote 1\]](#)

## Sequencing the data

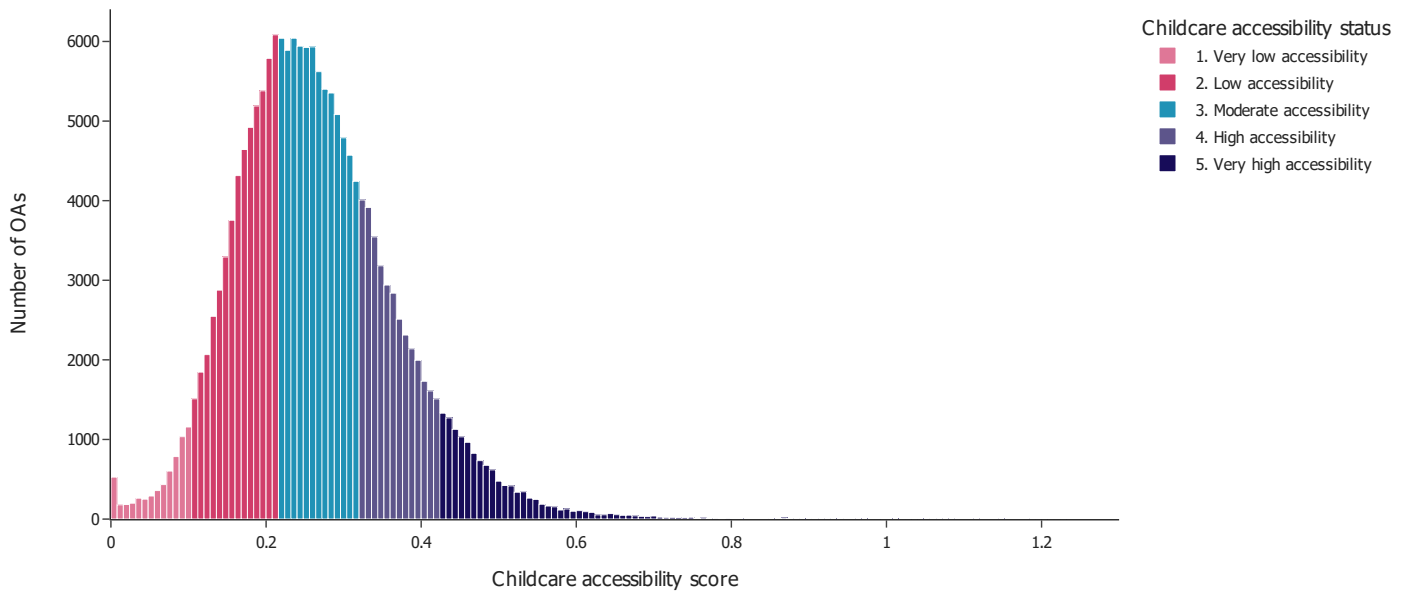
Sequence analysis requires categorical data. To produce the categories, we classified the accessibility ratios of each output area (OA) as at 31 March 2020 into 5 distinct groups. We defined the groups as:

- very low accessibility: 0 to 0.10 places per child
- low accessibility: 0.10 to 0.21 places per child
- moderate accessibility: 0.21 to 0.32 places per child
- high accessibility: 0.32 to 0.43 places per child
- very high accessibility: over 0.43 places per child

The groups were defined this way according to the spread of data and to ensure that outliers did not skew how the groups were classified.

We also tried using quintiles to categorise the data, but this led to very few OAs in the very high accessibility group due to the large positive tail in the data and we did not feel this was representative of the spread of data.

## Figure 1: Distribution of accessibility scores (31 March 2020)



View [data in an accessible table format](#).

## Optimal matching

Sequence analysis relies on an optimal matching algorithm. Optimal matching quantifies the dissimilarity between sequences. The aim is to measure how 'similar' or 'dissimilar' 2 sequences are by calculating the 'cost' of transforming one sequence into another through a series of operations.

We specified the substitution-cost matrix so that the cost depends on which states the OA is moving between. For example, moving from very low accessibility to very high accessibility is more difficult in practice than moving from very low accessibility to low accessibility. We have defined the substitution-cost matrix we used below:

	<b>Very low accessibility</b>	<b>Low accessibility</b>	<b>Moderate accessibility</b>	<b>High accessibility</b>	<b>Very high accessibility</b>
<b>Very low accessibility</b>	0	2	4	8	16
<b>Low accessibility</b>	2	0	2	4	8
<b>Moderate accessibility</b>	4	2	0	2	4

High accessibility	8	4	2	0	2
Very high accessibility	16	8	4	2	0

---

## Clustering analysis

We then used the distance matrix to determine an appropriate number of clusters, using a partitioning around medoids (PAM) clustering method. PAM clustering is similar to k-means clustering.<sup>[footnote 2]</sup> However, it uses medoids (the most centrally located data point within a cluster) instead of centroids (a data point that represents the centre of the cluster - the mean - and it might not necessarily be a member of the dataset) to represent the clusters. K-means clustering can be skewed by extreme values, and therefore was less appropriate for our analysis due to the presence of outliers.

In each step of the clustering analysis:

1. PAM starts by selecting k medoids.
2. Each sequence is assigned to the nearest medoid based on a chosen dissimilarity measure.
3. The algorithm iteratively refines the choice of medoids by swapping one of the medoids with a non-medoid observation if the sum of dissimilarities within clusters can be reduced.
4. The process continues until no further improvements can be made, resulting in stable clusters.

You can read more about the [PAM clustering method](#).

### Weighted PAM clustering

Normally, with PAM clustering, each sequence is assigned to the cluster with the closest medoid and the distance between the points and medoids are treated equally for all points. Instead of this, we adjusted the distance calculation by the

weights of the sequence, so that sequences with higher weights have a greater impact on determining which medoid is chosen, and thus on the formation of the clusters. We determined the weighting by counting the number of OAs within each sequence.

In our context, there were more OAs assigned to sequences reflecting very low accessibility, so we wanted to assign more weight to these.

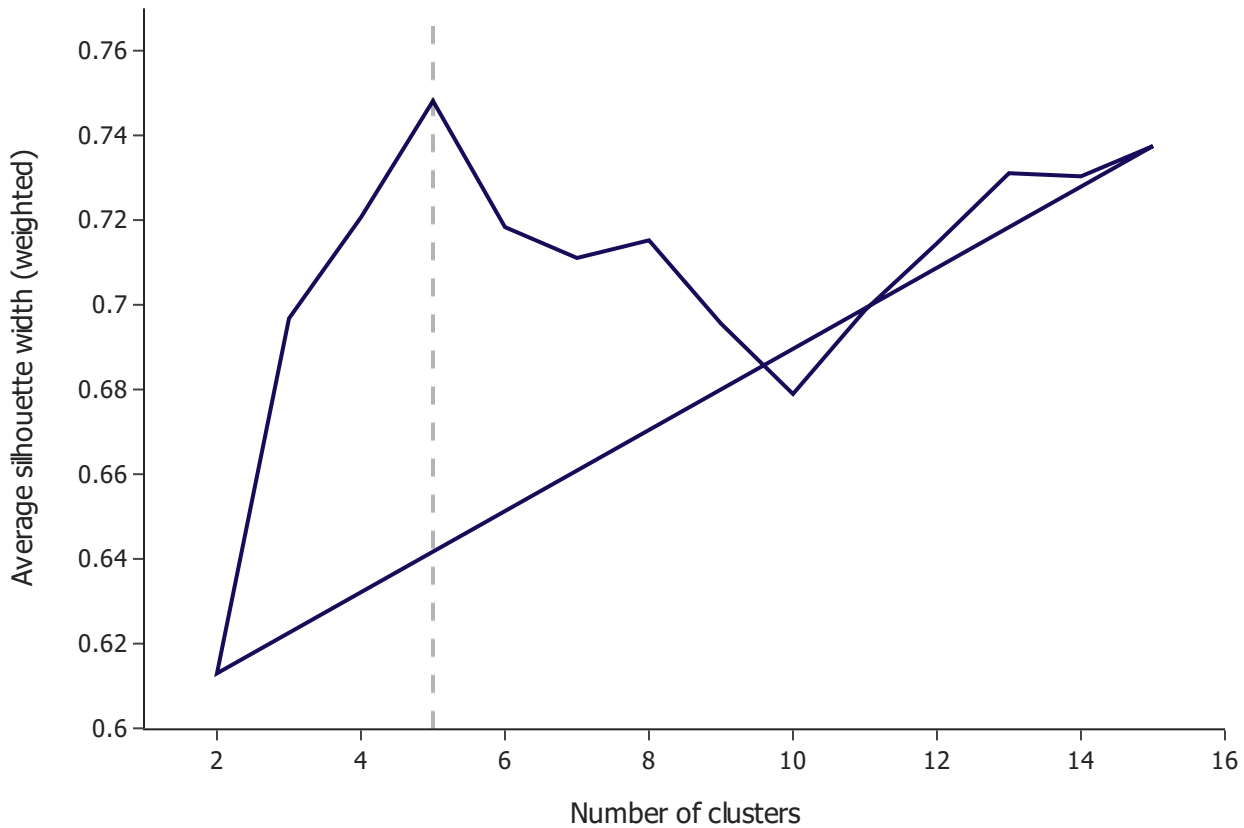
The benefit of weighting the data is that high-weight sequences have a larger impact on the clustering outcome, resulting in clusters that reflect the true data structure more closely. The weighting also helps to stabilise the clusters by emphasising reliable data points.

### **Choosing the optimum number of clusters**

To choose the optimum number of clusters, we produced the silhouette score for each number from 1 to 16. The silhouette score is used to evaluate the quality of clusters. It measures how similar an object is to its own cluster (cohesion) compared with other clusters (separation). This provides an indication of how well each data point fits within the assigned cluster and how distinct the clusters are from one another.

We chose 5 as the optimum number of clusters, as this had the highest silhouette score, as shown below. We also applied personal judgement and deemed that 5 was a suitable number of clusters considering the data that we had and how childcare access tends to vary across the country.

### **Figure 2: Silhouette plot from clustering analysis**



View [data in an accessible table format](#).

## Code

We developed this analysis using code published by the University of Liverpool. You can view the [Github repository](#).

## Data

We produced annual childcare accessibility ratios from 31 March 2020 to 31 March 2024 for each OA in England. The ratios were calculated using a 2-step floating catchment method, which is outlined in our [methodology](#).

We did not expand the time series further as the recording of places changed in 2018 when we introduced Ofsted's new administrative system, Cygnum.

We used population data from Census 2021 for 2019 onwards, as we felt this more accurately represented the populations living in each OA at the time than using Census 2011 data.

## Urban-rural data

We used data from the [Global Human Settlement Layer](#) to classify each OA in England into different groups according to its degree of urbanisation:

- dense urban cluster grid cell
- low density rural grid cell
- rural cluster grid cell
- semi-dense urban cluster grid cell
- suburban or peri-urban grid cell
- urban centre grid cell
- very low density rural grid cell

We then linked this data using OA code to determine the degree of urbanisation of each local area.

## Census 2021 data

### Economic inactivity

We used the [Census 2021](#) data to gather data on the number of females who are economically inactive because of looking after home or family, in households that have a dependent child aged 0 to 4 years. This data is only available at middle layer super output area (MSOA) level. We therefore used MSOA figures as a proxy for the OA figures. This allowed us to determine the number of such females.

### Highest level of qualification

We used the Census 2021 data to gather data on the number of females who hold Level 4 qualifications or higher, in households that have a dependent child aged 0 to 4 years. This data is only available at MSOA level. We therefore used the MSOA figures as a proxy for the OA figures. This allowed us to determine the number of

such females.

## Deprivation

We used the Census 2021 data to gather data on deprivation for households that have a dependent child aged 0 to 4 years. We looked at households that are deprived in all 4 dimensions (education, employment, health and housing).<sup>[footnote 3]</sup> This data is only available at MSOA level. We therefore used the MSOA figures as a proxy for the OA figures.

## Output area classification

To contextualise the childcare deserts and oases, we determined which groups from the [Consumer Data Research Center \(CDRC\)](#) data were most prevalent or disproportionately found within each group. We then used information from the CDRC [Pen Portraits](#) to summarise some common features that fall within the deserts and oases.

## Data table for figure

### Data for figure 2: Silhouette plot from clustering analysis

Number of clusters	Average silhouette width (weighted)
2	0.612985
3	0.696777
4	0.720659
5	0.748138
6	0.71834
7	0.711055
8	0.715227



9	0.695569
10	0.678944
11	0.698725
12	0.714537
13	0.731068
14	0.730333
15	0.737479

See [Figure 2](#).

1. [‘Sequence Analysis: Clustering on Edit Distance’](#), PennState, 2017. ↵
2. [‘K-Means Clustering- Introduction’](#), Analytics Vidya, 2024. ↵
3. [‘Household deprivation variable: Census 2021’](#), Office for National Statistics, 2023. ↵

[↑ Back to top](#)

## Help us improve GOV.UK

To help us improve GOV.UK, we’d like to know more about your visit today. [Please fill in this survey \(opens in a new tab\)](#).

**Services and information**

**Government  
activity**

[Benefits](#)

[Births, death, marriages and care](#)

[Business and self-employed](#)

[Childcare and parenting](#)

[Citizenship and living in the UK](#)

[Crime, justice and the law](#)

[Disabled people](#)

[Driving and transport](#)

[Education and learning](#)

[Employing people](#)

[Environment and countryside](#)

[Housing and local services](#)

[Money and tax](#)

[Passports, travel and living abroad](#)

[Visas and immigration](#)

[Working, jobs and pensions](#)

[Departments](#)

[News](#)

[Guidance and regulation](#)

[Research and statistics](#)

[Policy papers and consultations](#)

[Transparency](#)

[How government works](#)

[Get involved](#)

---

[Help](#) [Privacy](#) [Cookies](#) [Accessibility statement](#) [Contact](#)

[Terms and conditions](#) [Rhestr o Wasanaethau Cymraeg](#)

[Government Digital Service](#)

**OGI**

All content is available under the [Open Government Licence v3.0](#), except where otherwise stated



© Crown copyright