

HCP Data Methodology

Data Sources

Council-held data as well as publicly available data (open data) were used to perform an analysis of the housing conditions in Birmingham local authority area. The analysis was performed against different criteria that affect living conditions such as crime, anti-social behaviour (ASB) and deprivation.

Council-held data were extracted from Birmingham City Council's (BCC) data warehouses and open data were downloaded from web sources. These data were used to create data models and drive interactive dashboards in a Business Intelligence software.

Details about the data sources are presented in Table 1 and 2 below. The Dashboard section also provides information about which data sources are used in each dashboard.

LSOAs

In order to effectively compare different geographical areas in Birmingham with each other, as well as with other geographical areas in England, the statistical Lower-layer Super Output Areas (LSOAs) were used. LSOAs were produced by the Office for National Statistics (ONS) for the reporting of small area statistics and are areas designed to be of a similar population size, with an average of approximately 1,500 residents or 650 households¹. There are 32,844 LSOAs in England and 639 in the Birmingham LA area. Some of the datasets that were used for the analysis already contained data aggregated at LSOA level. For datasets that contained data at street or postcode level, aggregations were performed to calculate the values for the LSOA level.

Geography data model

A geography data model was created to relate LSOAs to higher level geographical areas such as wards and parliamentary constituencies, based on the ONS Postcode Directory (ONSPD) dataset. The ONSPD dataset relates postcodes with different geographical areas (including wards and constituencies) and therefore, the resulting data model can provide LSOA to ward and parliamentary constituency lookups.

By using the geography data model, the data at LSOA level can be aggregated up to ward and constituency levels. However, there's a minor caveat; as seen in Figure 1 below, the LSOA boundaries do not match exactly with the ward boundaries which means that an LSOA can be part of one, two, or more wards. Therefore, when performing aggregations for a metric from LSOA to ward or constituency levels, the value of the metric for the LSOA is included as whole in the calculation of the aggregated value for all wards/constituencies to which the LSOA belongs partially. A more specific example is provided in the Dashboard section below.

¹ MHCLG Indices of Multiple Deprivation Notes



Figure 1. LSOA boundaries (black outline) do not match perfectly with ward boundaries (red outline)

National averages calculation

For the purposes of the analysis, the values for deprivation and crime in Birmingham area were compared with the national average. Following is a description of the methodology used to calculate the national average values for deprivation and crime in England.

Deprivation

The “MHCLG Indices of Multiple Deprivation in England” dataset ranks all the LSOAs in England from 1 to 32,844 based on their score in seven domains of deprivation. The LSOA with rank 1 is the most deprived and the LSOA with rank 32,844 is the least deprived². The national average was calculated as the middle value of the rank which is 16,422. Therefore, an LSOA with rank less than 16,422 is more deprived than the national average and one with rank higher than 16,422 is less deprived.

Crime

To compare the crime rate of Birmingham areas with the national average, the average crime count per LSOA was used. Datasets that contain crime data for all the police forces in England were downloaded from the UK police website for the years 2018 and 2019. The datasets were aggregated and the total number of crimes per crime category for each year was calculated. Then, the total number of crimes for each crime category is divided by the total number of LSOAs in England (32,844) to calculate the average number of crimes for each category per LSOA. The averages are calculated separately for each year (2018 and 2019). When comparing an area in Birmingham with the national average, the number of crimes for that area is divided by the number of LSOAs within the area to calculate the area’s average number of crimes per LSOA.

² MHCLG Indices of Multiple Deprivation Notes

Tenure types

Housing stock in the UK can be categorised in four different tenure types: Owner-occupied, Private Rented Sector (PRS), Housing Association (HA) and Local Authority Housing (LAH). Another important category of houses that was used in the analysis is the Houses in Multiple Occupation (HMO), which is a subset of PRS. Following is the methodology that was used to calculate the counts of houses for the different tenure types at LSOA level.

HMO, LAH and HA calculation

Datasets with information about HMO, LAH and HA houses at property level were used to calculate the number of houses for each of these three tenure types per LSOA. Using the geography data model, the data were aggregated at LSOA level and then the calculated results were merged in a single dataset. In addition, BCC provided a dataset with the total houses per LSOA that was created using data from Local Land and Property Gazetteer (LLPG) database.

PRS calculation

The number of PRS houses was calculated by extrapolation in 2 steps:

1. First, a growth rate of the PRS houses in Birmingham area was estimated based on a publicly available dataset that contains the number of private rented dwellings having tenancies with secured deposits for the period 2010 to 2017. Based on 2010-2017 values, the values for the years 2018 to 2020 were predicted by extrapolation and the growth rate of private rented tenancies with secure deposits from 2011 to 2020 was calculated.
2. Then, the estimated count of PRS houses per LSOA for 2020 were estimated by multiplying the growth rate calculated in the previous step, with the total number of PRS houses per LSOA recorded in the 2011 census dataset.

For 5 (out of 639) LSOAs, the sum of LAH, HA and estimated PRS houses was higher than the total houses per LSOA given by the "LLPG Total Houses" dataset (i.e. $LAH + HA + PRS > \text{total houses}$). This means that the estimated value of PRS houses in these LSOAs is too high. To mitigate this issue, the number of PRS houses for these five LSOAs was capped to the maximum possible based on the total of properties minus the LAH and HA (i.e. $PRS = \text{total houses} - LAH - HA$).

Owner-Occupied calculation

Given the count of houses per LSOA for 3 out of 4 tenure types (PRS, LAH, HA), as well as the total number of houses per LSOA, the count of owner-occupied houses was calculated as the difference of the total houses minus the sum of the counts for the other three tenure types, i.e. $\text{Owner-Occupied} = \text{total houses} - (PRS + LAH + HA)$.

Table 1: Open Data

Dataset Description	Year	Source and Licence
<p>MHCLG Indices of Multiple Deprivation for England: the dataset contains the ranks and deciles for the Index of Multiple Deprivation (IMD) and seven domains at LSOA level. The seven domains are: Income, Employment, Education Skills and Training, Health Deprivation and Disability, Crime, Barriers to Housing and Services, and Living Environment.</p> <p>The IMD is an overall relative measure of deprivation constructed by combining the seven domains of deprivation according to weights.</p>	At 2015 and at 2019	<p>URL: https://www.gov.uk/government/collections/english-indices-of-deprivation</p> <p>Licence: Open Government License v3.0*</p>
<p>ONS Mean house prices by LSOA (HPSSA dataset 47): this dataset is part of the House Price Statistics for Small Areas (HPSSAs) release, produced by ONS. They are calculated using open data from the HM Land Registry, a source of comprehensive record level administrative data on property transactions.</p> <p>The dataset contains the mean price paid for residential properties sold and registered in a given period of four consecutive quarters. <i>We extracted attributes labelled as “Dec 2017”, “Dec 2018” and “Dec 2019”, accounting for the 4 quarters of years 2017, 2018, and 2019.</i></p>	2017 to 2019	<p>URL: https://www.ons.gov.uk/peoplepopulationandcommunity/housing/datasets/meanpricepaidbylowerlayersuperoutputareahpssadataset47 (Source: Office for National Statistics)</p> <p>Licence: Open Government Licence v3.0*</p>
<p>ONS Residential property sales by LSOA (HPSSA dataset 41): this dataset is part of the House Price Statistics for Small Areas (HPSSAs) release, produced by ONS.</p> <p>The dataset contains the number of residential properties sold and registered in a given period of four consecutive quarters. <i>We extracted attributes labelled as “Dec 2017”, “Dec 2018” and “Dec 2019”,</i></p>	2017 to 2019	<p>URL: https://www.ons.gov.uk/peoplepopulationandcommunity/housing/datasets/numberofresidentialpropertysalesbylowerlayersuperoutputareahpssadataset41 (Source: Office for National Statistics)</p> <p>Licence: Open Government Licence v3.0*</p>

accounting for the 4 quarters of years 2017, 2018, and 2019.		
<p>Police Crime: this dataset contains individual crime and anti-social behaviour (ASB) incidents, including street-level location information.</p> <p><i>This dataset was aggregated at LSOA level to allow comparisons of LSOA figures in Birmingham with averages of LSOAs in England.</i></p>	2018 to 2019	<p>URL: https://data.police.uk/data/</p> <p>Licence: Open Government Licence v3.0*</p>
<p>VOA Council tax stock of properties (table CTSOP4.1) this dataset contains number of properties and property build period at LSOA level.</p> <p>The counts are calculated from domestic property data for England and Wales extracted from the Valuation Office Agency's (VOA) administrative database.</p>	At 31st March 2019	<p>URL: https://www.gov.uk/government/statistics/council-tax-stock-of-properties-2019</p> <p>Licence: Open Government Licence v3.0*</p>
<p>ONS Number of tenancies with secured deposits, England: this dataset contains the number of privately rented dwellings that have tenancies with secured deposits held. It is at region, local authority, and LSOA levels in England.</p>	2010 to 2017	<p>URL: https://www.ons.gov.uk/peoplepopulationandcommunity/housing/datasets/numberoftenancieswithsecureddepositsengland (Source: Office for National Statistics)</p> <p>Licence: Open Government Licence v3.0*</p>
<p>ONS Local area migration indicators, UK: this dataset includes migration flows and population estimates at local authority level.</p> <p><i>The dataset used was published in August 2020 and contains estimates for 2009 to 2018.</i></p>	2009 to 2018	<p>URL: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/migrationwithintheuk/datasets/localareamigrationindicatorsunitedkingdom (Source: Office for National Statistics)</p> <p>Licence: Open Government Licence v3.0*</p>
<p>ONS Population projections: this dataset includes migration and population projections at local authority level. Projections are 2018-based.</p> <p><i>The dataset used was published in March 2020 and only projections for 2019 and 2020 were extracted from it.</i></p>	2019 to 2020	<p>URL: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationprojections/datasets/components-of-changebirthsdeathsandmigrationforregionsandlocalauthoritiesinenglandtable5 (Source: Office for National Statistics)</p>

		<p>Licence: Open Government Licence v3.0*</p>
<p>ONS Postcode directory:</p> <p>This dataset includes lookups between postcodes, LSOA, wards, constituencies, and other geographical levels.</p>	<p>At May 2020</p>	<p>URL: https://geoportal.statistics.gov.uk/data-sets/ons-postcode-directory-may-2020</p> <p>Licences:</p> <ul style="list-style-type: none"> . Contains OS data © Crown copyright and database right 2020 . Contains Royal Mail data © Royal Mail copyright and database right 2020 . Source: Office for National Statistics licensed under the Open Government Licence v3.0*
<p>ONS Lower Layer Super Output Areas (LSOA) December 2011 Boundaries: the dataset contains the digital vector boundaries for LSOAs as at 31 December 2011 in England.</p> <p><i>These boundaries are used in every map of the dashboards.</i></p>	<p>Set on 2011, still up-to-date at August 2020.</p>	<p>URL: https://geoportal.statistics.gov.uk/data-sets/lower-layer-super-output-areas-december-2011-boundaries-ew-bgc</p> <p>Licences:</p> <ul style="list-style-type: none"> . Office for National Statistics licensed under the Open Government Licence v3.0* . Contains OS data © Crown copyright and database right 2020

*Open Government License v3.0:

<http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>

Table 2: Council Data

Dataset Description	Year	Source
HMOs: This dataset contains information about the HMOs in Birmingham at property level as at August 2020.	August 2020	BCC data warehouse
LAH: This dataset contains information about the houses owned by BCC at property level as at August 2020.	August 2020	BCC data warehouse
HA: This dataset contains information about Housing Association properties in Birmingham at property level as at August 2020.	August 2020	BCC data warehouse
LLPG Total Houses: This dataset is based on Local Land and Property Gazetteer (LLPG) data and contains the number of residential properties in Birmingham per LSOA for August 2020.	August 2020	LLPG
<p>PRS and Owner Occupied: This dataset contains the number of PRS and Owner-occupied houses in Birmingham per LSOA for August 2020.</p> <p><i>The attributes of this dataset were calculated by extrapolation. See Tenure types section below for the methodology that was used.</i></p>	August 2020	Calculated dataset (cf. Tenure types section above)
<p>CRM Tickets: This dataset contains tickets as recorded in BCC's Customer Relationships Management (CRM) system for the years 2018 and 2019.</p> <p><i>The dataset was filtered for ticket categories relevant to ASB, animal related problems, graffiti, flyposting, noise nuisance, illegal drug taking or dealing.</i></p>	2018 to 2019	BCC CRM (extracted from BCC Data Warehouse)

Dashboards

Four dashboards were built based on housing conditions criteria. They offer an overview of the living conditions across Birmingham using BCC and open data.

Dashboard pages

The four dashboards contain a total of seven pages:

- Deprivation (one page):
 - **Deprivation**: comparison of the Index of Multiple Deprivation and its domains between Birmingham and England; information about Birmingham tenure types.
- Crime_ASB (two pages):
 - **Crime**: comparison of the police recorded crimes between Birmingham and England LSOAs; information about Birmingham tenure types.
 - **CRM_ASB**: comparison of the ASB-related tickets recorded by BCC between an area that can be selected and the Birmingham LSOA average; information about Birmingham tenure types.
- Housing_Demand-Profile-Conditions (three pages):
 - **Housing_Demand**: comparison of the number of residential properties sales and mean price paid for residential properties between Birmingham wards, constituencies, and LSOAs; information about Birmingham tenure types.
 - **Housing_Profile**: distribution of Birmingham tenure types that can be visualised on a map and broken down by constituencies, wards, and LSOAs.
 - **Housing_Conditions**: distribution of properties age bands in Birmingham LSOAs; information about Birmingham tenure types.
- Population Migration (one page):
 - **Population_Migration**: evolution of Birmingham population and migration (internal and international) between 2009 and 2020 (at LA level).

Relation between dashboard pages and datasets

Below are lists of datasets used in each page of the dashboards. These datasets refer to the ones described in the “Table 1: Open Data” and “Table 2: Council Data” sections.

All pages use (with the exception of Population_Migration page):

- Open data: “ONS Postcode directory” and “ONS Lower Layer Super Output Areas (LSOA) December 2011 Boundaries”. They are referred as “Geo data”
- Council data: “LLPG Total Houses”, “HMOs”, “LAH”, “HA”, “PRS and Owner Occupied”. They are referred as “Birmingham Tenure Profile”

The **Deprivation** page uses:

- Open data: “MHCLG Indices of Multiple Deprivation for England”, “Geo data”.
- Council data: “Birmingham Tenure Profile”

The **Crime** page uses:

- Open data: “Police Crime”, “Geo data”
- Council data: “Birmingham Tenure Profile”

The **CRM_ASB** page uses:

- Open data: “Geo data”
- Council data: “CRM Tickets”, “Birmingham Tenure Profile”

The **Housing Demand** page uses:

- Open data: “ONS Mean house prices by LSOA (HPSSA dataset 47)”, “ONS Residential property sales by LSOA (HPSSA dataset 41)”, “Geo data”
- Council data: “Birmingham Tenure Profile”

The **Housing_Profile** page uses:

- Open data: “Geo data”
- Council data: “Birmingham Tenure Profile”

The **Housing_Conditions** page uses:

- Open data: “VOA Council Tax stock of properties”, “Geo data”
- Council data: “Birmingham Tenure Profile”

The **Population_Migration** page uses:

- Open data: “ONS Local area migration indicators, UK”, “ONS Population projections”

Aggregation of LSOA data

As mentioned earlier, approximative aggregations of LSOA data are performed to provide data at Ward and Constituency levels.

In every page of the dashboard with values at Constituency and Ward levels, these values are calculated by aggregating LSOA values based on the geographical boundaries of both the LSOA and the Constituency or Ward. However, as illustrated previously in Figure 1 in the Geography data model section, there is a mismatch of the boundaries of different geographical areas. Therefore, even if one LSOA only **partially** belongs to a Ward or Constituency, the values related to this LSOA are added to the total values of the Ward or Constituency.

For example, when summing up all the PRS Houses of the Wards of a Constituency, the sum can be greater than the Constituency sum of PRS Houses. This is because one LSOA can belong **partially** to one, two, or more Wards of a Constituency, and therefore the count of PRS Houses of a LSOA can be added to multiple Wards. Whereas the Constituency total is just the total of PRS Houses for all LSOAs in the Constituency, it does not add twice or several times the values of LSOAs that can be at the same time in two or more Wards. This is illustrated in Figure 2 below, where the sum of the PRS Houses for all Wards adds up to more than the constituency total ($31962 > 23641$), because some LSOAs are in more than one Ward.

Constituency/Ward/LSOA	PRS Houses
<input type="checkbox"/> Birmingham, Ladywood	23641
<input type="checkbox"/> Ladywood	11816
<input type="checkbox"/> Soho & Jewellery Quarter	6651
<input type="checkbox"/> Nechells	1660
<input type="checkbox"/> Newtown	2044
<input type="checkbox"/> Bordesley & Highgate	3404
<input type="checkbox"/> Aston	2121
<input type="checkbox"/> North Edgbaston	2148
<input type="checkbox"/> Bordesley Green	1042
<input type="checkbox"/> Lozells	466
<input type="checkbox"/> Holyhead	610

Figure 2. LSOA-Ward-Constituency aggregation example for Ladywood Constituency and Wards