Analysis of Internal Migration in the UK 2011 and 2021: A Visual Analytics Approach

Student ID: 2031077

N.B. I compiled my dahsboards into a Story. This has been saved as a packaged workbook. For best results please view only the Story in Presentation mode.

Abstract

This paper characterises the design choices made throughout my Tableau dashboards. I use Munzner's Visual Analysis and Design as my main source of reference. Initially, I introduce my investigation on internal migration in England and Wales, before defining the specific tasks for which I have made the visualisations. I then elaborate on my design choices, supported by *vis* theory. Finally, I critically evaluate my project and conclude with findings from my research.

1 Introduction

This report provides a comprehensive analysis into the patterns of internal migration within England and Wales, at multiple levels of granularity. Understanding the dynamics and characteristics of changing areas, moves beyond the static picture of the make-up of an area. Understanding how an area is changing, the numbers and types of people moving in or out and the places migrants come and go, are all vital in creating policies for urban regeneration and maintenance.

The census data provides a unique opportunity for analysing migration patterns, as it is the only data source that questions the entire population as to their address one year ago. Whilst other sources capture some of this data, it is often based on representative samples or administrative data, neither of which scale accurately to an entire population.

Migration data can be used to highlight a wide array of pressing issues underlying the population. For example, Shuttleworth et al. explore the overall decrease in migration rates over time [1], whilst Simpson et al. undertake a demographic analysis of the data, to support discussions in ethnic variations in migration patterns [2]. Many illustrious researchers have further demonstrated the practicality of visualising census data effectively, something I wish to echo in my dashboards. In People and Places [3], Dorling illustrates over 500 maps, illuminating over 125 different aspects of the 2001 census data. It is therefore my aim to derive both valuable insights into within-England-and-Wales migration and display these insights in a digestible manner, such that social scientists and casual viewers alike can both draw insights from the data visualisations I am providing.

Both the 2011 and 2021 censuses define a migrant as someone with a different place of residence one year before the census, regardless of how far they moved [4]. Whilst the census data is elucidating, it has shortcomings. Data are only collected decennially, meaning only a snapshot of the bigger picture is provided and other data sources must be used to provide a picture of the intercensal years. Therefore, where necessary I have used data sources extending beyond the census, to provide the full picture. For example, the Office for National Statistics (ONS) provides annual mid-year data on internal migration moves into and out of each local authority in England and Wales, which can be used in conjunction with the census data. Census data also lacks the information to investigate motivation for migration [5]. We have to look elsewhere to discern the *why* people moved. For instance, Thomas et al. analyse migration motives related to distance using the UK Household Longitudinal Study, along with similar studies from Australia, Sweden, and the US . Their findings indicate that housing concerns primarily drive short-distance relocations, whereas employment opportunities are a significant factor for longer-distance moves [6].

Regarding pitfalls of the census data itself, approximately 9% of all moves are not captured in the census. This typically occurs when individuals move multiple times within a year or pass away before the census is conducted [7]. Furthermore, about 10% of responses to the migration question are imputed. This is most common among individuals in their early 20s, very young children (ages

0-4), and the oldest age groups. This reliance on imputation means the quality of the methodology is crucial for accurate data representation. While studies using census microdata show similar determinants of migration for both actual and imputed data [8], this does not guarantee that imputed records accurately reflect what the missing responses would have been.

The rest of this report details the logical process to create my visualisations. I have followed Munzner's "What, Why, How" framework for optimal *vis*.

2 Definitions

Definition 2.1: National Statistics Socio-economic Classification (NS-SeC): a framework for classifying the socio-economic status of people in the UK. For my purposes, I have used the 8-class version [1], whereby class 1 is seen as the UK's elite and class 8 is those most disadvantaged [9]. The breakdown of NS-SeC categories is shown in Figure 1.

Definition 2.2: Indices of Multiple Derpivation (IMD): IMD in England and Wales are a set of comprehensive measures that assess relative deprivation across small areas or neighborhoods. The indices are based on multiple dimensions of deprivation, including income, employment, health, education, crime, housing, and living environment. This is given as ordered, ordinal data where each entry is a number between 1 and 10, 1 being the most deprived and 10 the least [10, 11].

Figure 1: NS-SeC Origin Classes [12]

3 What: Data Preparation and Abstraction

The data sources for this project are outlined below:

- 2011 Census Table: UKMIG007
- 2011 Census Table: KS101EW
- 2011 Census Table: MM01CUK_ALL
- 2011 Census Table: UKMIG001

[1] NS-SeC categorisation also extends to one more class: "L15 Full-time students", which I have included in my analysis.

- 2021 Census Table: MIG001
- 2021 Census Table: MIG007EW
- English Indices of Deprivation 2015
- Welsh Indices of Deprivation 2015
- Internal migration: by local authority and region, age and sex (Years: 2002-2021)

Using these sources I was able to see migration through the lenses of age, sex and economic status for different area output levels. This involves creating a **table**, where the attributes are net migration values for different NS-SeC categories and also various derived attributes based on two or more of the columns combined. Each **item** would be a different local authority, identifiable by a unique place name and area code; the unique area name and code also act as **keys** for the table. The value each cell contains varies depending on the attribute it's representing. For example, if the column heading ends with *Net: Value*, then this implies the value in each cell is ordered, **quantitative** data, displaying how many people have migrated in an area. If the heading is *Average IMD decile*, this will describe ordered, **ordinal** data where each entry is a number between 1 and 10, 1 being the most deprived and 10 the least.

My first step was to make the dataset I have described above for easy analysis of Net Migration, extracted from the 2011 Census Table: UKMIG007. As the downloaded data comprises all of the UK, it was necessary (and this applies to all datasets I created) to filter out any local authorities from Scotland and Northern Ireland. As my initial data from 2011 for Migration by NS-SeC was concerned with breakdowns of migration for individual local authorities, and not statistics at a national level, I could simply remove the local authorities of Scotland and N. Ireland without having to adjust Net Values. These areas were easy to identify from unique geography codes that identify each nation comprising the UK; whereby England starts the code with *E* and Wales with *W*.

As I was concerned with the net values of migration for each local authority, I removed all columns that didn't contain the terminology *Net migration within the UK* excepting the *geography* and *geography code* columns. This left me with a reduced dataset displaying the net values of migration for each local authority in England and Wales, categorised by NS-SeC level.

Next, data was imported from the 2011 census table: KS101EW, which contained information on the local population for each local authority and the number of residents living in a communal establishment. This was subsequently merged with the new table I had created for Migration by NS-SeC.I next imputed the *Indices of Deprivation* for both England and Wales onto this new table.

Now that this information was all together, I created new columns detailing various exclusions and proportions such as *Net migration excluding student population* and *Proportion: Net migration as a proportion of Local Authority residents*. The full dataset can be viewed in Appendix Figure 3. This table served as the source for Dashboard's 1 & 4.

The raw data for Dashboard 6 was in an unusable state. It was given in multidimensional, tabular form; as an origin-destination matrix. With rows and columns correspond to origins and destinations (given my local authority names), respectively. Each cell, then, contained the number of migrants that corresponds to the interaction between the row and column identifiers. Adding to the multi-dimesnionality, there are multiple tables/matrices each with slightly differing semantics. Each table is segmented into specific age categories and sex. As I am not interested in the demographic information, it was my first task to combine all multidimensional matrices, into one matrix displaying the origin and destination figures across all ages and sexes.

My first step was to manipulate the data manually within excel to create a new sheet that combined the data to cover all ages and both sexes. This gave me one large csv, that I processed in Python. From here I *melted* the data. This transformed each cell in the matrix into a row. In this process, each row represents a unique origin-destination pair (key identifier), along with the corresponding count of migration. This meant it was now in a suitable state for use in tableau.

Dashboard 3 is concerned with migration by age and sex between 2011 and 2021. This meant I required combined data, extracted from 2011 and 2021 sources. From the census tables displaying Migration by sex and age for both 2011 and 2021, I was able to create a joint dataset. The table displays multidimensional, temporal, and categorical data representing migration patterns within the UK. Each age group's data point encapsulates several dimensions— proportion of residence change, both within and outside local authority boundaries, across two census years (2011 and 2021), reflecting geo-spatial movement trends. Quantitative data in the form of proportions and absolute numbers are used to quantify these trends.

I then repeated this task replacing the categorisation into age and sex categories with NS-SeC categories, for Dashboard 2.

Dashboard 7 is concerned with the inflows and outflows of migrants from local authorities in 2021. I found that whilst the 2021 census data was in a hard to understand form, if I was merely presenting the tabular data, it was nonetheless in a usable format for tableau. The dataset in the table consists of individual records detailing residential movement within a specific geographic area, coded by identifiers for each area. It categorizes people by categorical variables such as area name and gender, and by ordinal variables representing age groups. Quantitative data is also present, as counts of individuals associated with each movement event. Each entry delineates whether the individual has moved within the area, signifying an inflow, or moved out, indicating an outflow, further classified demographically. Whilst the data is coded geographically, this cannot be used alone as a key identifier, as the same geographic location is listed multiple times. Therefore, a combination of all attributes is required to distinguish between each cell.

After creating these multiple datasets, it should be noted that subsets were created for some specific purposes. This meant further abstracting newly formatted tables, to contain less but more specific information; pertaining to visualisation in tableau. The subset creation was trivial and does not require explanation, as the tasks above do.

4 Why: Task Definition

This section defines the visualisation techniques used in the Tableau dashboards developed for analysing internal migration within the UK, according to Munzner's task taxonomy. Each dashboard is categorised into specific **actions** aimed at specific **targets**, enhancing user understanding and interaction.

4.1 Dashboard 1: Choropleth Map and Bar Charts

- **Action: Present**
  - **Target: Trends** – Visualising migration trends by displaying net migration across various NS-SeC categories.
  - **Target: Extremes** – Highlighting the top and bottom local authorities in terms of migration figures.
- **Action: Compare**
  - **Target: Attributes** – Enables comparison of net migration figures among different NS-SeC categories within and across local authorities.
- **Action: Summarise**
  - **Target: Derived Proportions** – Aggregates data to depict overall migration patterns.

4.2 Dashboard 6 & 7: Inflows and Outflows

- **Action: Present**
  - **Target: Trends** – Displays general migration inflows and outflows, including internal movements within local authorities.
- **Action: Compare**
  - **Target: Top and Bottom** – Identifies significant migration hubs and sinks by comparing local authorities with the highest and lowest flows.
- **Action: Locate**
  - **Target: Specific Areas** – Allows users to focus on and retrieve detailed migration data for specific local authorities.

4.3 Dashboards 2, 3 & 5: Comparative Migration Analysis

- **Action: Present**
  - **Target: Temporal Changes** – Shows how migration patterns have evolved from 2011 to 2021 across age, sex, NS-SeC and region.
- **Action: Compare**
  - **Target: Temporal and Demographic** – Compares migration trends across different demographic groups and census periods.
- **Action: Summarise**
  - **Target: Aggregates by Demographics** – Provides summaries of migration data by demographic variables to highlight broader trends.

4.4 Dashboard 4: Dimensionality Reduction and Clustering

- **Action: Discover**
  - **Target: Patterns** – Utilises PCA, t-SNE, and k-means clustering to uncover hidden patterns in the NS-SeC dataset.
- **Action: Categorise**
  - **Target: Cluster Groups** – Groups local authorities based on their NS-SeC profiles using cluster analysis.

5 How: Visualisation Justification

Once preprocessing was complete and the tasks for each dashboard were established, it was vital to follow Munzner's detailed guidelines on information visualisation (info-vis) to build suitable dashboards.

5.1 Arrange Tabular Data

Munzner describes the in-depth principles of info-vis, commencing with marks and channels. A mark is an n-dimensional graphical element in an image known as a geometric primitive [13, p.95]. It can take the form of a point, a line, an area and more. A channel controls the nature of these marks, encoding size, colour, direction, angle and more. As Munzner says, "All channels are not equal", and hence the choices I have made in my visualisations, have been thoroughly curated to present my information in enticing and informative ways.

5.1.1 Bar Charts

Across my dashboards is the use of bar charts to encode quantitative values. The bar charts I implement can be categorised as aligned, rectilinear and space-filling; as they receive their information from tabular data.

In Dashboard 1,the bar charts use line marks and encode the selected attributes with a single spatial position channel. These attributes are always ordered, showing either quantitative, net values for migration or different proportion statistics; or ordinal data. The other attribute displayed is categorical, displaying the local authority. The bar charts are all aligned by a common frame, such that the highest-accuracy aligned position channel is utilised. The regions (bars) are ordered by their value on the y-axis. That means they are ordered by vertical height such that trends are distinguishable.

As scalability is an issue with bar charts, my Dashboard displays two separate charts, one showing the top five values and one showing the bottom five values for the selected attribute. This allows me to show the most significant entries at both extremes, without showing one graph with many hundreds of bars and confusing the visualisation. The limitation of five bars to each chart also means there is enough white space between the marks, such that they are completely distinguishable.

In Dashboard 6, the bar charts remain rectilinear, however, the axis has been shifted 90°. This shift is done purposefully, such that the information is perceived as flowing 'into and out-to' the local authority selected. The bars express the value attribute with their aligned vertical position, such that they are ordered by greatest inflows on one chart and outflows on the other chart; with the vertical axis containing their key attribute of local authority. Again, for scalability the graphs are limited to the top ten, to preserve white-space between bars. I also use arrow glyphs as the bars, to highlight sequential data.

In one particular visualisation showing the change in proportion of inflows from 2011 to 2021 for regional data, bars are aligned along a common horizontal scale to show changes, with negative changes (declines) shown in one color and positive changes (increases) shown in another. This arrangement effectively facilitates comparison of the magnitude of change across different regions.

5.1.2 Line and Dot Charts

In my further dashboards, where I wish to convey a change over time in proportions of migration for different sex, age and NS-SeC categories, I employ the idioms of dot and line charts. The charts displaying the proportions of migration, categorised by sex and age, display a lot of information onto a seemingly simple plot. Munzner states that dot charts typically show one value attribute and one key-attribute with a rectilinear spatial layout, often with augmentations to show a second categorical attribute. That is precisely the case here. The vertical, spatial channel displays the attribute value of proportion, whilst the horizontal spatial channel displays two categorical attributes of age and sex. The dots are further coloured based the categorical attribute of year, to distinguish between 2011 and 2021. They are also joined by lines, to highlight trends. I employ all of these methods to easily distil the multidimensional tabular data onto and easy to process dot chart, whereby clear trends can be seen for the year attribute and clear comparisons can be made between the categorical attributes of age and sex.

5.1.3 Packed Bubbles

I also utilise packed bubbles charts, where each bubble represents a local authority with its size corresponding to the value of migration. This employs the "juxtapose" method. This method involves displaying elements side-by-side for comparison without using a common scale, where each bubble's size provides a visual cue to the data it represents. The primary visual encoding is size of the bubble. The dense layout uses space effectively to represent each data point (local authority in this case) as a discrete unit, allowing viewers to visually compare the relative size of migration values across different authorities.This type of visualisation is more effective for providing a general sense of distribution and relative magnitude among the data points.

5.2 Arrange Spatial Data

5.2.1 Choropleth Map

I use three choropleth maps across my dashboards. In Dashboard 1 the map is used to show a quantitative attribute encoded as colour over regions delimited as area marks, as per the Munzner definition. My region shapes are determined using given geometry, where each region boundary is that of the local authority boundary for the 2011 census. In Dashboard 3 there is no colour-encoding, as the map is used simply to display the location of points on other scatter plots on that same dashboard. In Dashboard 5, I employ a satellite image of the choropleth map. As I am showing regions not local authorities on this Dashboard, I do away with the area marks on the previous Dashboards. Tableau does not geocode well for the regions, so I instead highlight each region with a coloured circle and the name of the region. On the other two dashboards I opt for a dark map, with England and Wales coloured in. This allows the use of white lines to mark out local authority regions and the high contrast between the dark background and white lines makes the boundaries very clear. As Dashboard 5 is not concerned with these small boundaries, I employ the satellite image so as not to distract the user from the key information in the visualisation and rather highlight the general area the issue pertains to. Where before I used a colour-map gradient to distinguish quantitative attributes on the map itself, I instead use geodesic lines between regions, which have their own colour-map.

The primary concern for the choropleth maps is the colour-map used. I will discuss the colour choices in detail in the next section.

5.3 Colour

In keeping with the principles laid down by Munzner, I will structure my analyses around luminance, saturation and hue; the three constituents of colour. Naturally, colour is used in all my dashboards to either distinguish between categorical attributes or to encode ordered attributes. To distinguish further, luminance and saturation are magnitutude channels, whilst hue is an identity channel.

5.3.1 Proposed Color Palette

Special attention is required for **color blindness**, ensuring that chosen hues are distinguishable to all viewers, including those with common forms of color vision deficiencies. A palette is suggested, combining blue, orange and grey; each chosen for specific attributes they convey in terms of neutrality, attention, and semantic representation. Furthermore, the high contrast between the colours used and the dark backgrounds ensures visibility and differentiation across various data sets, whilst being accessible to the majority of those with vision impairments.

5.3.2 Examples

For all visualisations where I wish to highlight ordinal data, I use segmented colour-maps to emphasise its discrete nature. This is seen in my t-SNE and PCA plots in Dashboard 4, where the categorical cluster attributes are labelled with different hues.

In Dashboard 1, entitled " Internal Migration Trends by NS-Sec for Local Authorities", we see a choropleth map. The map uses a white-to-blue colourmap, with a sequence of five levels, with monotonically decreasing luminanace. This gives perceptual linearity, so viewers can easily distinguish between areas with high or low migration values. Furthermore, high, medium and low-level structure is preserved.

In Dashboard 6: " Inflows and Outflows for Local Authorities, 2011", I use the size and shape channels to signify greater and smaller migration flows. Munzner states that our judgement of lengths is very accurate [13, p.236], therefore, to demonstrate the larger flows I use an arrow glyph to signify a flow and change the size accordingly; the bigger the arrow the larger the flow. Furthermore, all the arrows are in aligned planar positions, the most accurate vision channel. In Dashboard 5 the thickness of the line on the choropleth map is size-coded to show sum of migration between the regions.

I maintain a constant, filled texture on all of my graphs and visualisations. This ensures that there is no loss of effect when I do use some textural effects such as stippling. For example, Dashboard 2 shows a line graph. The line graph is the main object of the visualisation and is a filled line with some thickness, of constant, low luminance, highly saturated, hue. I then use stippling to show a regression line. This ensures, the main visualisation is not detracted from, whilst adding useful information to the graph.

5.4 Manipulating View

5.4.1 Selecting Elements

Almost all of my dashboards have some element of selection. In order to view a different attribute with respect to the overriding theme of the dashboard, the user can select from a list of attributes to change the view. Furthermore, in Dashboard's 5 and 7 the user can select multiple different combinations of attributes such as *region and year*, or *local authority, sex and migration category*; to completely change the view. This is closely related to attribute filtering, discussed in Section 5.6.1. Furthermore, the hover idiom is utilised in all my dashboards, such that when hovering over an item on a *vis*, the tool-tip pops up with meaningful information pertaining to that *vis*.

5.4.2 Highlighting

In the majority of my dashboards I have also implemented highlighting, whereby a visual pop out effect is achieved by sufficiently tampering with the saturation contrast. This means when a user hovers or clicks on certain items, the item will pop-out and so will all other items related to that item. This means users are provided with immediate visual feedback, to confirm the results of their target intention.

5.4.3 Zooming

Whilst none of the dashboards offer the ability to switch between area views, such as an option to change the granularity from local authority to region, I have designed them such that adjacent dashboards build on the previous one, by showing more/less detailed information than the last at different area granularity. For example, Dashboard 5 shows regional migration, whilst Dashboard 6 expands on this to break it down by local authority.

5.4.4 Constrained Navigation

Every dashboard that features a map of England and Wales, exhibits constrained navigation. I have fixed the view so that the user does not move the viewpoint out of the area of interest. This way the visualisation is always optimal and how it was intended.

5.5 Facet

5.5.1 Shared Encoding

Every dashboard has some form of multiform views. This means that different visualisations are juxtaposed next to each other. To link these views I have employed same colour coding, so there is shared encoding through the colour channel. Further to this, I use the principle of linked highlighting, whereby, a contiguous region in one view is distributed across another, enhancing the understanding of data relationships. This means the user can easily isolate specific items by hovering or clicking, as aforementioned in Section 5.4.

5.5.2 Shared Data

In Dashboard's 1, 3 and 4; I use the overview-detail principle. This means that I display the overarching information through the use of choropleth maps or an age and sex pyramid, and then juxtaposed with this are visualisations displaying detailed subsets of the data, such as the top ten local authorities with the highest migration, or a box plot showing the spread of data in particular attributes. I implement this by making the overview visualisation larger than the detailed view, as suggested by Munzner.

5.5.3 Static Layering

All of my visualisations use static layering. I have designed the dashboards to show specific information, therefore the user can change which attribute they are visualising but not the layers they see. The user differentiates between information by visual attention. None of the visualisations contain too many items, such that static superimposition becomes too crowded.

5.6 Reduce

5.6.1 Attribute Filtering

Many of my dashboards have drop-down menus that allow the user to filter the attribute being visualised. This means only information pertinent to the users' specific task is displayed on screen, keeping the visualisation clear.

5.6.2 Aggregation

My largest dataset has 21 attributes. My goal was to visualise this dataset with a 2D scatterplot matrix (SPLOMS), using only two of these attributes to preserve some meaningful structure. I used two dimesnionality reduction techniques and compared their efficacy. The first technique was principal component analysis (PCA). PCA converts the data into a series of principal components, which are new variables ranked according to the variance they capture from the data. These components, derived from linear combinations of the original attributes, are identified through the eigen decomposition of the data's covariance matrix, with the first principal component capturing the greatest variance. Typically, you select a number of principle components that explains 90% of the data. The second technique I used was t-SNE, a nonlinear dimensionality reduction tool that embeds high-dimensional data into two or three dimensions for visualisation. As t-SNE is typically for visualisation, I expected it to perform better than PCA. After both techniques were implemented, I used k-means as a way of clustering similar points. I also added the local authority of each point as a label. This meant I could utilise item aggregation through boxplots, alongside the SPLOMS. I gave boxplots of each of the three k-means clusters. These boxplots can be filtered by attribute to see summary statistics for each cluster for the selected attribute, with the added granularity of each point being a local authority. This helps to derive meaning from the clusters found in the dimensionally-reduced data.

6 Evaluation

To evaluate my visualisations, it was necessary to characterise my project as problem-driven work. Hence, I will be evaluating from the top-down. My overall validation methods are shown at each level in Figure 2.

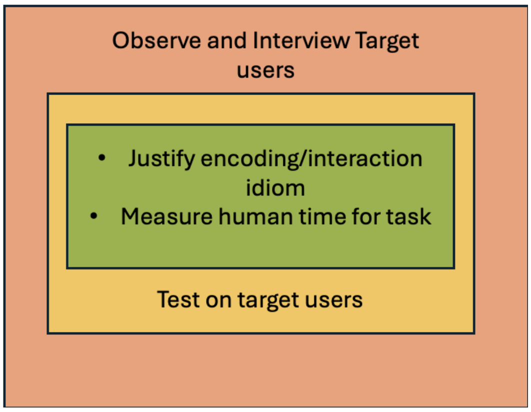


Figure 2: Validation methods by validation level. The outermost (orange) layer is the domain situation, the yellow layer is abstraction and the innermost layer is idiom validation.

6.1 Domain Situation

The goal of the project was present raw census data on internal migration, in a way that was easily digestible to the user, allowing insights to be drawn easily from otherwise cryptic data. I, therefore, implemented a field study through a questionnaire. This meant I could challenge users to derive conclusions from the data, which I had already discerned myself, to test the usefulness of my *vis*.

The two questions I asked were:

- `Identify the Local Authority with the highest net positive value for migration of` *`NS-SeC 8: never worked and long term unemployed`* `migration and then identify how many people in total moved within this local authority (internal migration).`
- `Did a higher proportion of L1,L2 & L3 workers live elsewhere one year ago, but within the same local authority in 2011 or 2021?`

The answers to the two questions respectively were *Enfield*, *331* & *2011*. All participants in my study correctly gave these answers.

6.2 Abstraction Validation

To identify whether the designed *vis*, solves the problem of my user, I gave a test group access to my dashboards and asked for qualitative feedback. This meant, they could use the dashboards freely and then report back to me on any positives and negatives they found. I asked the user's:

- `How intuitive were the dashboard interfaces?`
- `How easy was it to find and interpret migration data?`

Overall the responses highlighted ease-of-use of the dashboards. User's were happy with the comprehensiveness of information condensed into them, without sacrificing clarity. User's commented that the abstraction techniques were very useful, such as a good use of bar charts, colour scheme and clear legends. The main criticism was the need for clearer tool-tips, to make navigating the information quicker and easier. I incorporated this feedback by adjusting the tool-tips accordingly.

6.3 Idiom Validation

All of my design choices were purposeful and have been documented thoroughly throughout this report. I have made clear justifications for the specific encoding and data abstraction choices. To make sure that the encoding and interaction idioms were suitable, I timed the user's whilst they carried out the tasks detailed in Section 6.1. It took the user's an average of 34 seconds to complete task 1 and 12 seconds to complete task 2. This compared well to my timings of 27 seconds and 8 seconds, respectively, as I designed the dashboards so had added familiarity.

Overall the evaluation procedure acted as a strong feedback-loop and impacted the finished visualisations positively. As the purpose of the project was not algorithm development, there was no need for algorithmic validation. The dashboards worked well and did not take overly long to refresh.

7 Conclusion

The visualisations I have created have elucidated much on internal migration. The overall trend I have taken away is that migration levels are down in 2021 compared to 2011. Whilst this is a generalisation, it is something that seems to appear as a trend throughout most of the migration categories I have analysed. For example, whilst the age-sex distribution in 2021 was similar to that in 2011 for those migrating; the proportion decreased across all age groups for both males and females. I also found, quite interestingly, that population flows did not serve to reinforce area deprivation in 2011, unlike Bailey and Livingston's findings in 2001 [8]. Instead the public were moving into more deprived areas and out of less-deprived areas, perhaps highlighting the fact that house prices have dramatically risen in the ten years between each census. Another key finding, was that London is an outlier from other British regions when analysing migration data. Whilst people leaving regions decreased from 2011 to 2021, London's increased; being the only region in 2021 with a net outflow, which increased from 0.5% net outflow in 2011 to 2.1% in 2021. I have learnt, that whilst socio-economic status is a major factor in determining migration habits, age is the main deciding factor. I have also learnt, through my derived attributes, that if student populations are removed from local authority data, the landscape of net migration changes completely. Areas with large student populations such as Manchester, Oxford and Birmingham typically dominate the net migration. However, with their student population removed other local authorities come to the forefront. This would suggest that migration policy should always make special considerations as to the transient nature of certain groups in their population, as an area with constantly high net migration could be skewed by a high student population.

With regard to what I have learnt about data visualisation in general, I would start with planning. When dealing with masses of data it is easy to jump in and get ahead of yourself. This project has taught me that meticulous planning, following the guidelines and principles of Munzner specifically, led me to avoid unnecessary hassle. I was able to predetermine what I wanted to gain from my visualisations, which in turn meant mapping optimal routes to get there. This avoided unnecessary processing steps, downloading of extraneous data and time wasted on unfruitful visualisations. I was able to concentrate more on crafting well-executed ideas. Furthermore, I have never before considered colour theory when putting together visualisation, normally being satisfied with the defaults of the programme I am using. This has dramatically altered how I select my colour-schemes, making sure they are suitable for all vision types. Also, understanding the theory behind data abstraction techniques, such as when certain graphs are more applicable than others, means that I no longer waste time picking a graph that 'looks best'. Instead, I can easily determine which abstraction type is the most suitable for the job and conveys my data most efficiently.

References

[1] I. Shuttleworth, T. Cooke, and T. Champion. Why did fewer people change address in england and wales in the 2000s than in the 1970s? evidence from an analysis of the ons longitudinal

study. *Population, Space and Place*, 25(2):e2167, 2019.

[2] L. Simpson and N. Finney. Spatial patterns of internal migration: Evidence for ethnic groups in britain. *Population, Space and Place*, 2009. Query date: 2016-11-15.

[3] Daniel Dorling and Bethan Thomas. *People and places: A 2001 Census atlas of the UK*. The Policy Press, 2001.

[4] Office for National Statistics. Migrant indicator - variables by topic: Census 2021. `https://www.ons.gov.uk/census/census2021dictionary/variablesbytopic/internationalmigrationvariablescensus2021/migrantindicator`, 2021. Accessed: 2023-04-23.

[5] Nik Lomax. Internal migration: census data as a gold standard for insight and visualisation. *Journal of Maps*, 18(3):543–550, 2022.

[6] M. Thomas, B. Gillespie, and N. Lomax. Variations in migration motives over distance. *Demographic Research*, 40:1097–1110, 2019.

[7] P. Rees, F. Thomas, and O. Duke-Williams. Migration data from the census. In P. Rees, D. Martin, and P. Williamson, editors, *The Census Data System*, pages 245–268. John Wiley and Sons, Chichester, 2002.

[8] Nick Bailey and Mark Livingston. Selective migration and neighbourhood deprivation: Evidence from 2001 census migration data for england and scotland. *Urban Studies*, 45(4):943–961, 2008.

[9] Office for National Statistics. The national statistics socio-economic classification (ns-sec) rebased on soc 2010, 2023. Accessed: 2024-04-24.

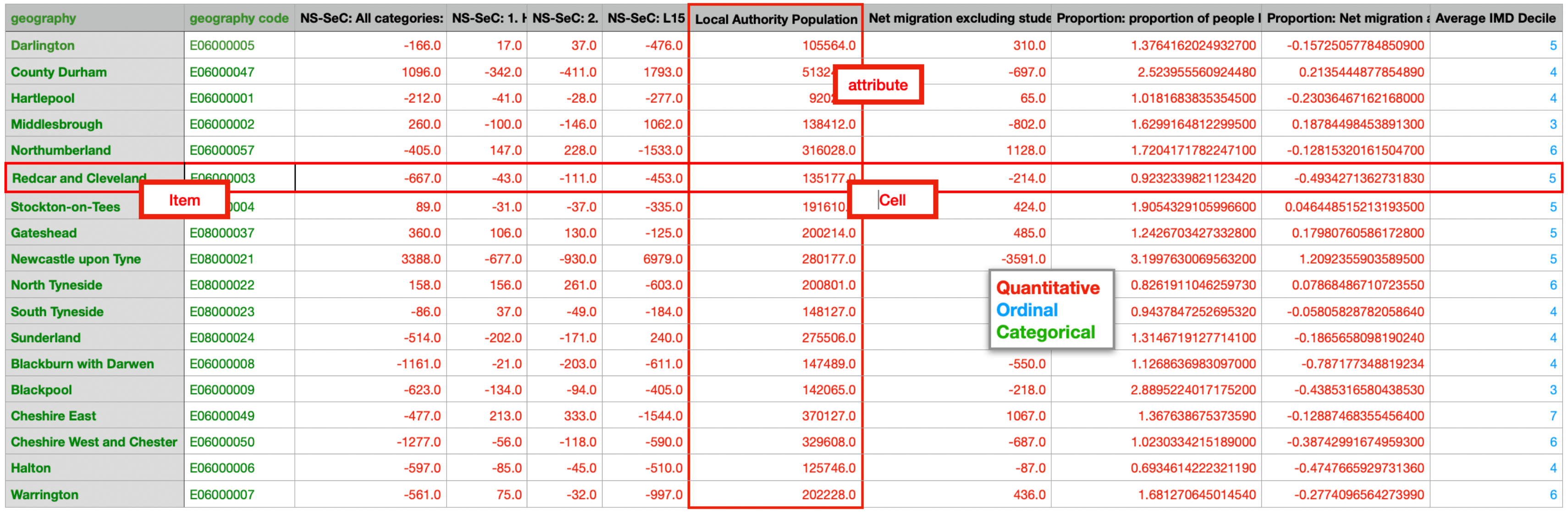
[10] Ministry of Housing, Communities and Local Government. English indices of deprivation 2021, 2021. Accessed: 2024-04-24.

[11] Welsh Government. Welsh index of multiple deprivation (wimd) 2021, 2021. Accessed: 2024-04-24.

[12] Daniel Laurison and Sam Friedman. The class ceiling: Social mobility and why it pays to be privileged. `https://ukdataservice.ac.uk/case-study/the-class-ceiling-social-mobility-and-why-it-pays-to-be-privileged/`, June 2019. Accessed: [your date of access here].

[13] Tamara Munzner. *Visualization Analysis and Design*. CRC Press, 2014.

8 Appendices



geography	geography code	NS-SeC: All categories:	NS-SeC: 1. I	NS-SeC: 2.	NS-SeC: L15	Local Authority Population	Net migration excluding stude	Proportion: proportion of people I	Proportion: Net migration a	Average IMD Decile
Darlington	E06000005	-166.0	17.0	37.0	-476.0	105564.0	310.0	1.3764162024932700	-0.15725057784850900	5
County Durham	E06000047	1096.0	-342.0	-411.0	1793.0	5132[illegible]	-697.0	2.523955560924480	0.2135444877854890	4
Hartlepool	E06000001	-212.0	-41.0	-28.0	-277.0	920[illegible]	65.0	1.0181683835354500	-0.23036467162168000	4
Middlesbrough	E06000002	260.0	-100.0	-146.0	1062.0	138412.0	-802.0	1.6299164812299500	0.18784498453891300	3
Northumberland	E06000057	-405.0	147.0	228.0	-1533.0	316028.0	1128.0	1.7204171782247100	-0.12815320161504700	6
Redcar and Cleveland	E06000003	-667.0	-43.0	-111.0	-453.0	135177.0	-214.0	0.9232339821123420	-0.4934271362731830	5
Stockton-on-Tees	[illegible]004	89.0	-31.0	-37.0	-335.0	191610[illegible]	424.0	1.9054329105996600	0.046448515213193500	5
Gateshead	E08000037	360.0	106.0	130.0	-125.0	200214.0	485.0	1.2426703427332800	0.17980760586172800	5
Newcastle upon Tyne	E08000021	3388.0	-677.0	-930.0	6979.0	280177.0	-3591.0	3.1997630069563200	1.2092355903589500	5
North Tyneside	E08000022	158.0	156.0	261.0	-603.0	200801.0	[illegible]	0.8261911046259730	0.07868486710723550	6
South Tyneside	E08000023	-86.0	37.0	-49.0	-184.0	148127.0	[illegible]	0.9437847252695320	-0.05805828782058640	4
Sunderland	E08000024	-514.0	-202.0	-171.0	240.0	275506.0	[illegible]	1.3146719127714100	-0.1865658098190240	4
Blackburn with Darwen	E06000008	-1161.0	-21.0	-203.0	-611.0	147489.0	-550.0	1.1268636983097000	-0.787177348819234	4
Blackpool	E06000009	-623.0	-134.0	-94.0	-405.0	142065.0	-218.0	2.8895224017175200	-0.4385316580438530	3
Cheshire East	E06000049	-477.0	213.0	333.0	-1544.0	370127.0	1067.0	1.367638675373590	-0.12887468355456400	7
Cheshire West and Chester	E06000050	-1277.0	-56.0	-118.0	-590.0	329608.0	-687.0	1.0230334215189000	-0.38742991674959300	6
Halton	E06000006	-597.0	-85.0	-45.0	-510.0	125746.0	-87.0	0.6934614222321190	-0.4747665929731360	4
Warrington	E06000007	-561.0	75.0	-32.0	-997.0	202228.0	436.0	1.681270645014540	-0.2774096564273990	6

Figure 3: Head of dataframe, built from combining multiple census tables for analysis (columns have been cut for brevity).