Data Visualization Assignment

Ethnicity

Riyad Hasan

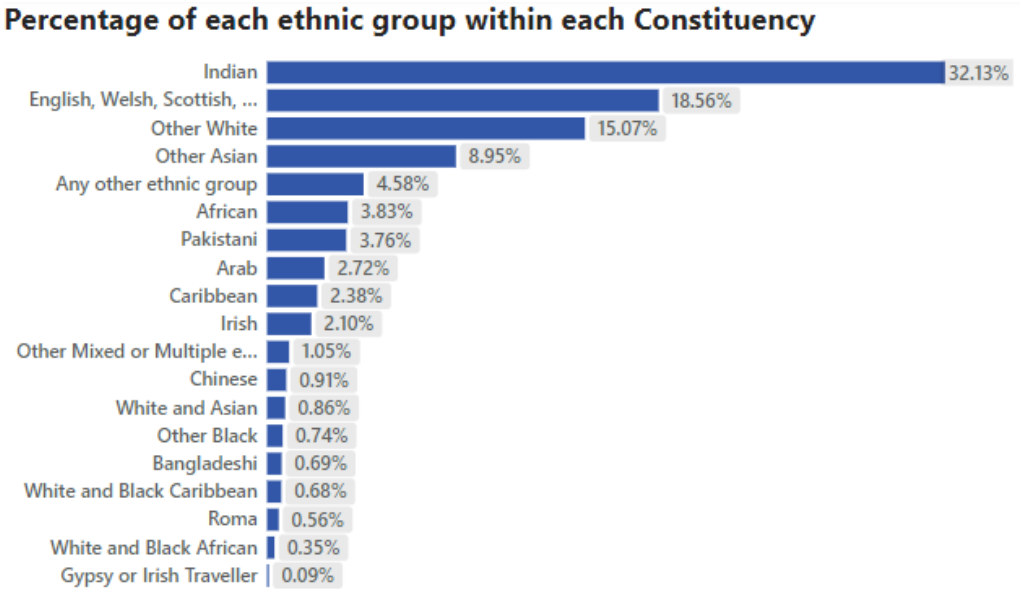
Introduction

This dataset is about the distribution of ethnic groups of every constituency within the United Kingdom. The dataset also includes broad ethnic groups and regional breakdowns allowing for further insights.

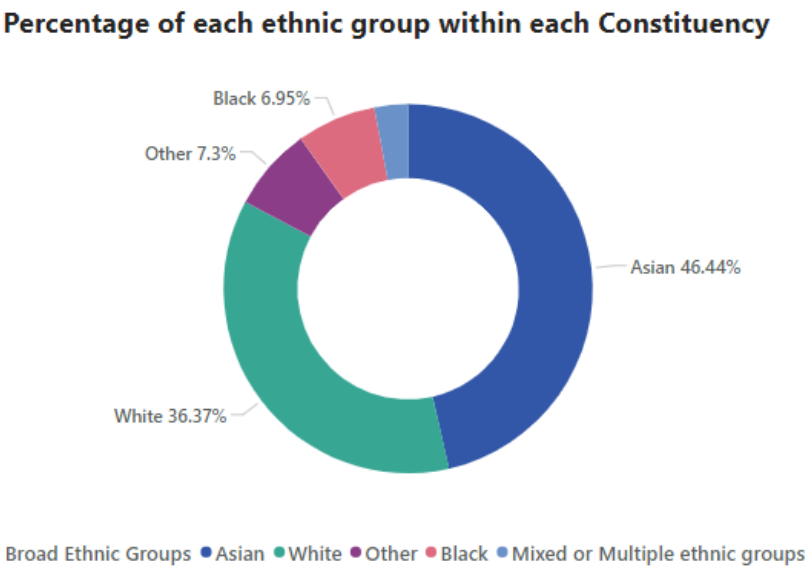
Demographics Distribution

1. What is the ethnic composition of each constituency?

Power BI was used to visualise the ethnic group breakdown within individual constituencies. A bar chart was used as it clearly displays all ethnic groups. Because there are so many constituencies, a filter by constituency was applied to the visual allowing the user to select which constituency to apply to the visual. The visual below for Harrow East has been shown.



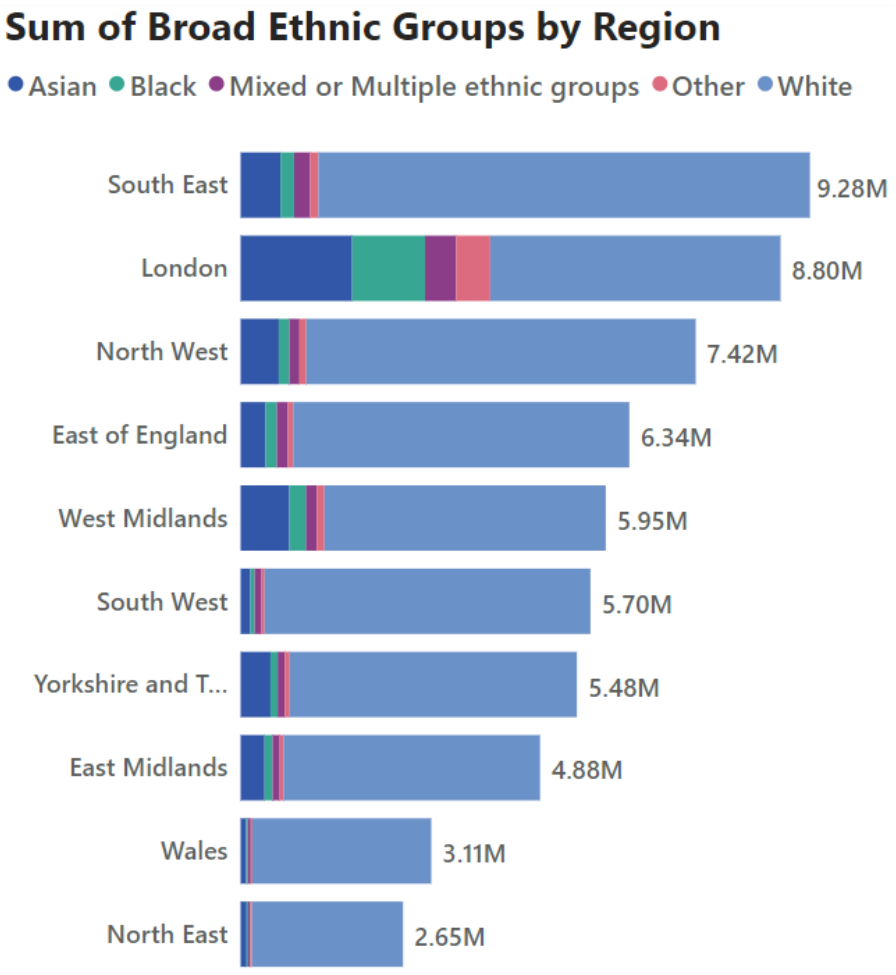
A donut chart was also used, however due to the number of categories it does not clearly display all ethnic groups. As such, for the donut chart, the broader ethnic groups category was used (E.g. Indian and Bangladeshi are two separate ethnic groups but both belong under the broader ethnic group of Asian in this dataset).



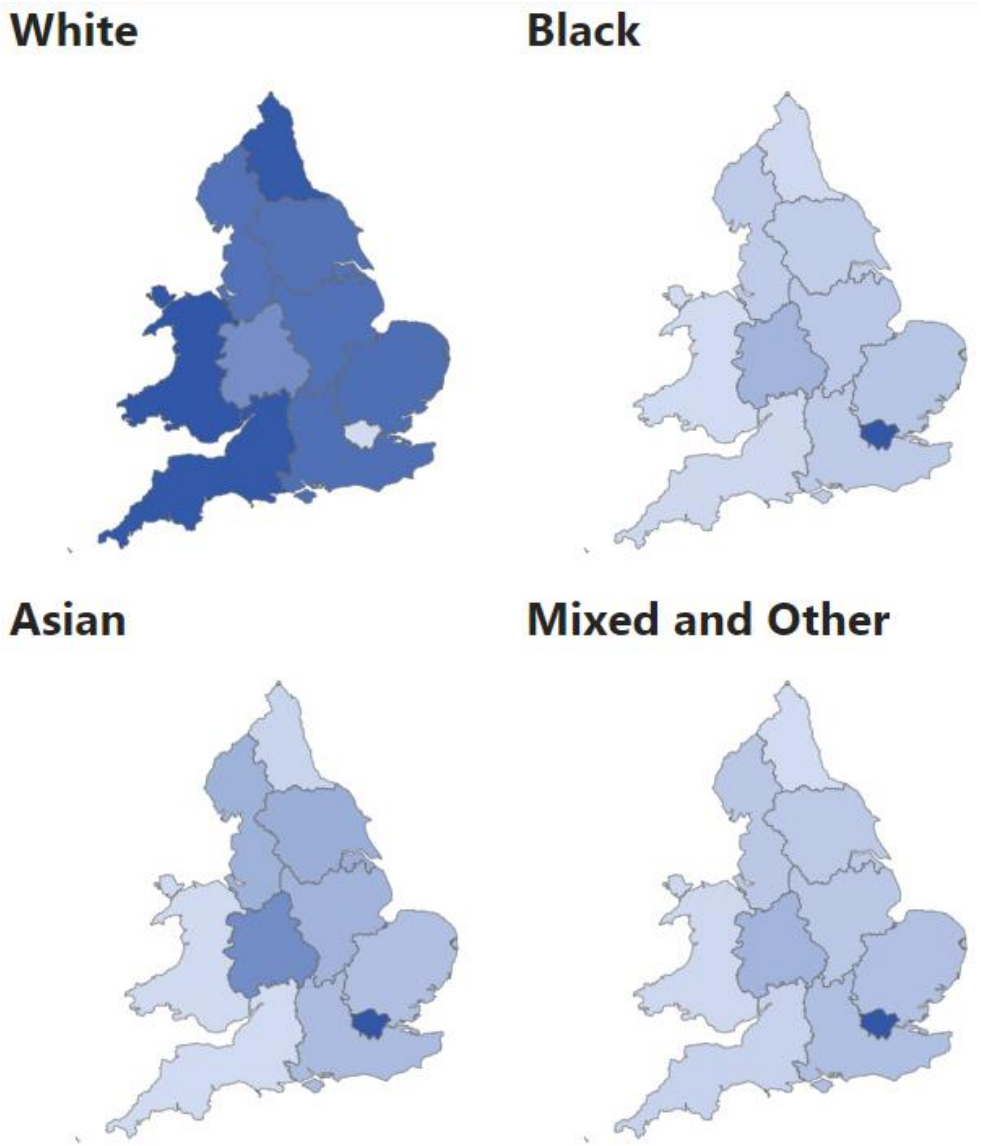
We can observe that in Harrow East, Indian is the predominant ethnic group with almost a third of the ethnic makeup (32.13%), and Asian is the predominant broad ethnic group with almost half of the ethnic makeup (46.44%). By trying different filters on the Power BI file, we can see that Harrow East is relatively more ethnically diverse than other constituencies.

2. How does ethnic composition vary across different regions?

A stacked bar chart was used to show the breakdown of broad ethnic groups across regions.



A heat shape map was also used. A UK regional map was obtained online from ONSvisual's topoJSON repository (github.com/ONSvisual/topojson_boundaries). This was used to create the map and related to the original dataset, as you can see below.

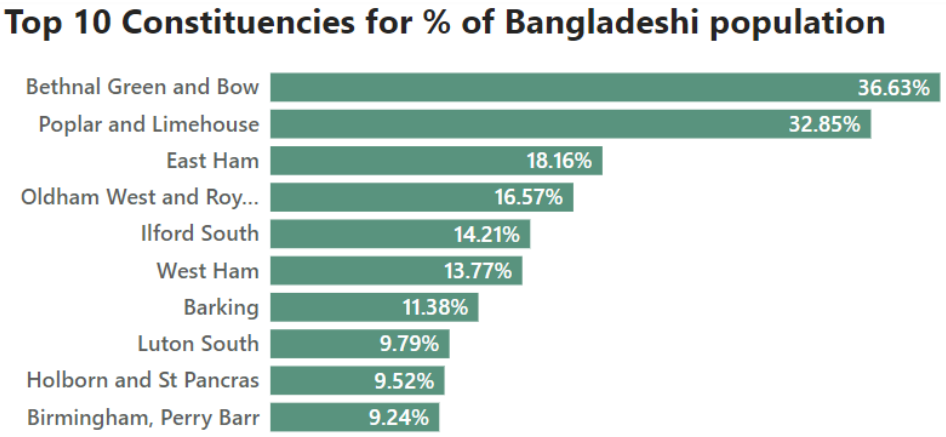


As we can see, London is the most diverse region where there is a relatively higher distribution of Asian, Black, Mixed and Other broad ethnic groups than in other regions. Conversely, London has a relatively lower distribution of White ethnic group than in other regions. This observation is reversed when it comes to the North East. South West and Wales regions. After London, the West Midlands is the region with the highest distribution of Asian, Black, Mixed and Other broad ethnic groups. Therefore London is the most diverse region whilst North East. South West and Wales are the most ethnically homogenous regions. We can also see that South East and London are the most populous regions whilst North East and Wales are the least.

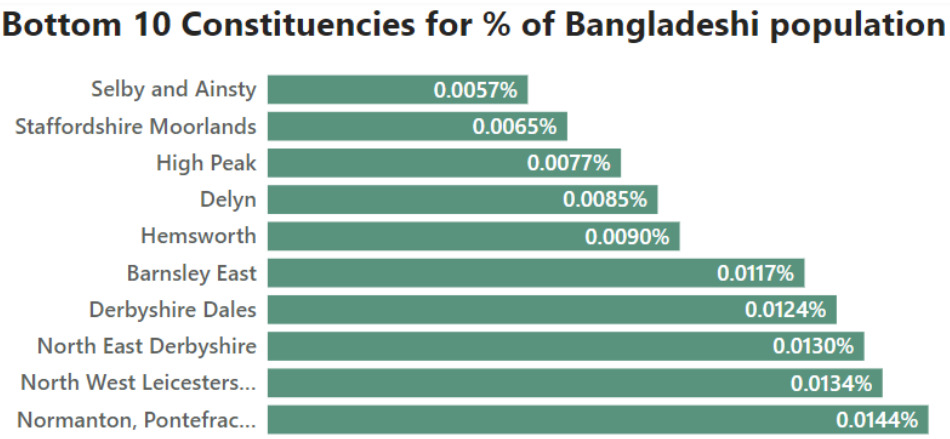
Comparative Analysis

3. Which constituencies have the highest and lowest percentages of specific ethnic groups?

For this question, the Bangladesh ethnic group will be used.



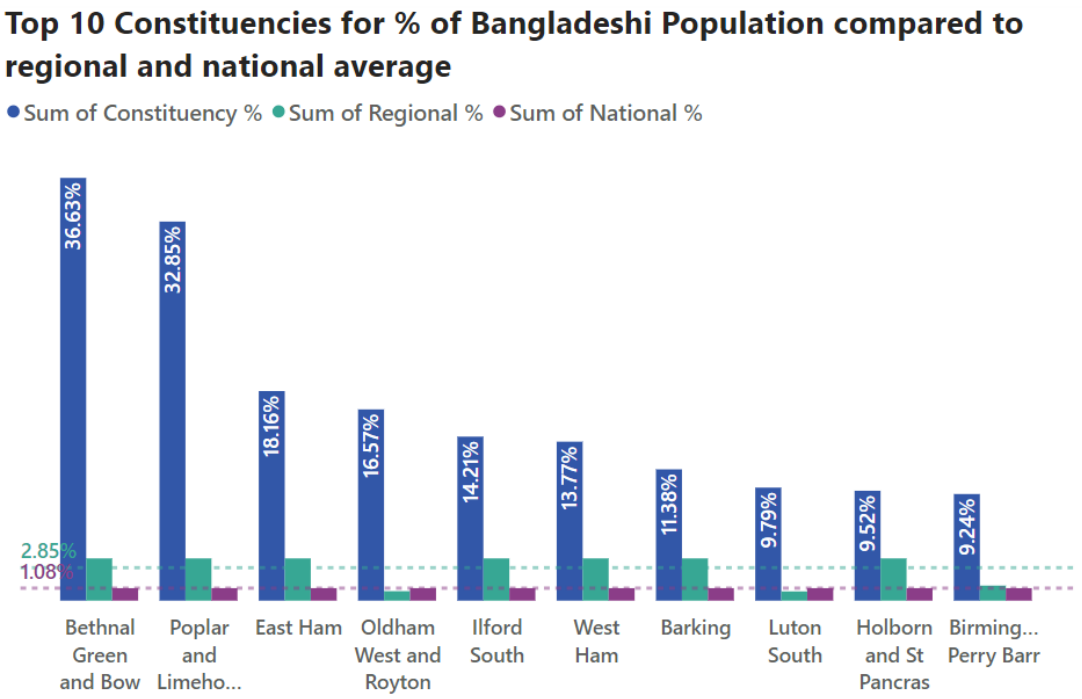
This column chart was created using constituency on the y-axis and constituency % as the x-axis. Two filters were added: one filter was applied to only show Bangladesh ethnic group and the second was to only show the top 10 results. As we can see Bethnal Green and Bow, and Poplar and Limehouse are the constituencies with the highest distribution of Bangladeshi population. Additionally, the filter was modified to only show the lowest 10 results.



This shows Selby and Ainsty, and Staffordshire Moorlands have the lowest distribution of Bangladeshi ethnic group.

4. How do regional and national percentages of ethnic groups compare to constituency percentages?

A stacked bar chart was used to display the regional and national percentages alongside the top 10 constituencies for % of Bangladeshi population.



This green line shows the average regional % for Bangladeshi population whilst the purple line shows it for national %. As we can see, the top 10 constituencies vastly deviate from the regional and national averages.

Trend Analysis

5. What are the trends in ethnic diversity over time (if historical data is available)?

The dataset provided only covers a single year, as such no historical data is available.

Socioeconomic Correlations

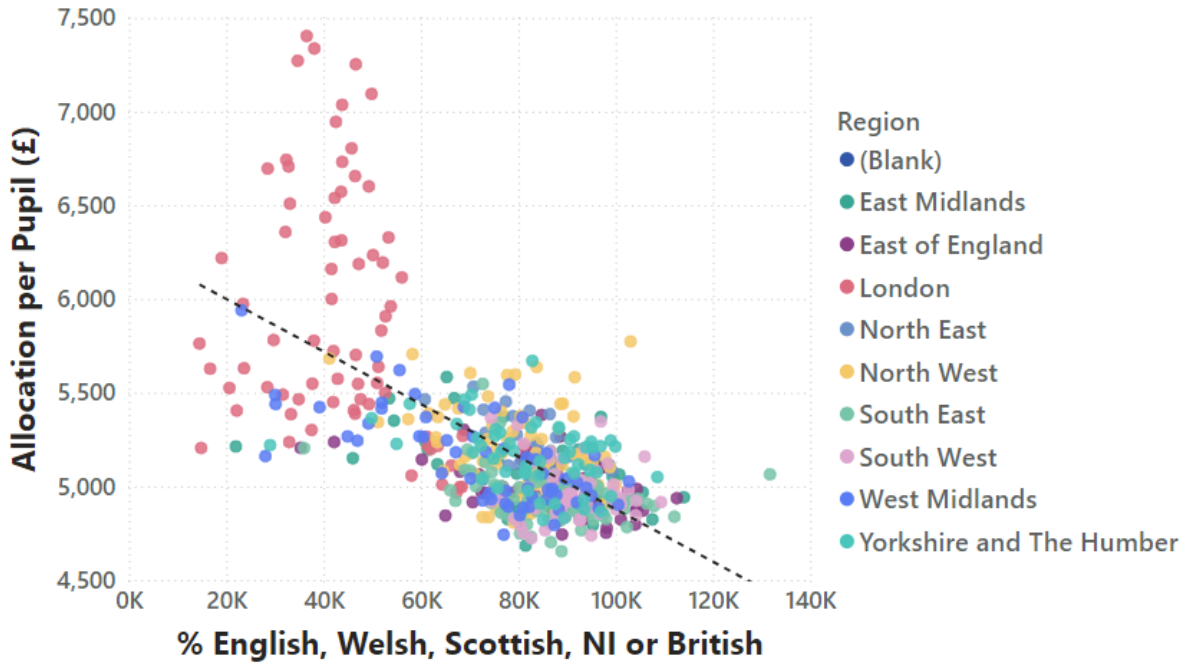
6. Is there a correlation between ethnic diversity and socioeconomic indicators (e.g., income, education)?

To answer this question, the school funding dataset was obtained from the government website. By creating a relation between the two datasets, we can investigate if there is a correlation between ethnic diversity and school funding.

Firstly, the school funding dataset was cleaned and transformed in Power BI. Only years 21-22 were kept as this Is the year for the ethnicity dataset. A many to one relationship between the constituency name columns in the ethnicity dataset and constituency name column in the school funding dataset.

To estimate ethnic diversity, the % of English, Welsh, Scottish, NI or British ethnic group in each constituency was used. The logic is that the lower this value, the more minority presence hence a higher degree of ethnic diversity.

A scatter plot with this in the x-axis and the funding allocation per pupil was created to investigate whether the degree of ethnic diversity of a constituency had any correlation with the amount of school funding that constituency received. It is shown below.



The scatter plot shows that as the % distribution of English, Welsh, Scottish, NI or British rises (so as the ethnic diversity decreases) the lower the school funding allocation.

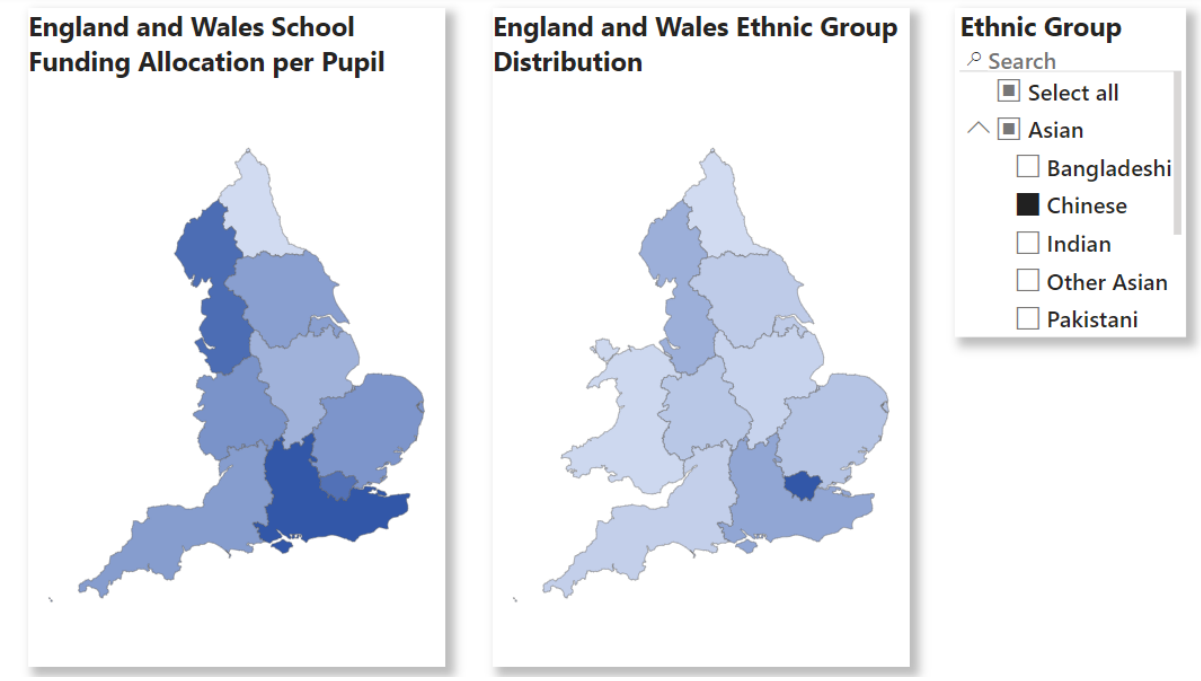
However, the region breakdown was added as a legend which shows that the majority of the highest school funding constituencies were from London. As we know from earlier analysis, London is the region with the highest ethnic diversity. As such, this scatter plot does not necessarily show a correlation, so further analysis must be performed to isolate the regional bias. Also, further analyses can be performed by relating the education attainment dataset to see whether ethnic diversity, school funding and attainment are interlinked

Geospatial Analysis

7. How is ethnic diversity distributed geographically?

The shape map chart is used for the geospatial analysis. The same regional map in Question 2 used to generate the shape map is used here. Two charts will be made, one will visualise the ethnic group distribution across regions and the other will visualise the school funding allocation per pupil, from the relational dataset covered in the previous question, across regions. A slicer for ethnic group is used so the distribution of each ethnic group across regions can be observed. Note that the ethnic group slicer has no effect on the school funding allocation map because the school funding dataset is not separated by ethnic group.

An example with the slicer set to Chinese ethnic group is shown below. Note that Wales is not featured in the school funding allocation chart because it was not included in the school funding dataset.



The darker the colour coding, the higher the school funding allocation per pupil and the higher the ethnic group distribution of a specific ethnic group. As we can see, the South East region has the higher school funding per pupil whilst the North East region has the lowest.

The ethnic group distribution map shows that the Chinese ethnic group is most represented in the London region, whilst it's least represented in Wales, North East and East Midlands regions.

Ranking and Comparison

8. Geographic maps displaying distribution of ethnic groups

To calculate the diversity index of each constituency/region, the following formula is used:

$$D = 1 - \frac{\sum n(n-1)}{N(N-1)}$$

Where n is the number of people of a specific ethnicity in a specific constituency/region and N is the total number of people in a specific constituency/region.

To perform this in Power BI, two columns were created: (n-1) which is the constituency total subtracted by 1, and n(n-1) which is the constituency total multiplied by (n=1). These were given the whole number data type so we can use the SUM function in calculated measures. Then, calculated measures were used to define Dnum i.e. the numerator of the D expression where it was defined as follows:

```
Dnum = SUM(Ethnicity_1[n(n-1)])
```

And calculated measure was also used to define Dden i.e. the denominator of the D expression, defined as follows:

```
Dden = CALCULATE(SUM(Ethnicity_1[Constituency Total]) *
SUM(Ethnicity_1[n-1]))
```

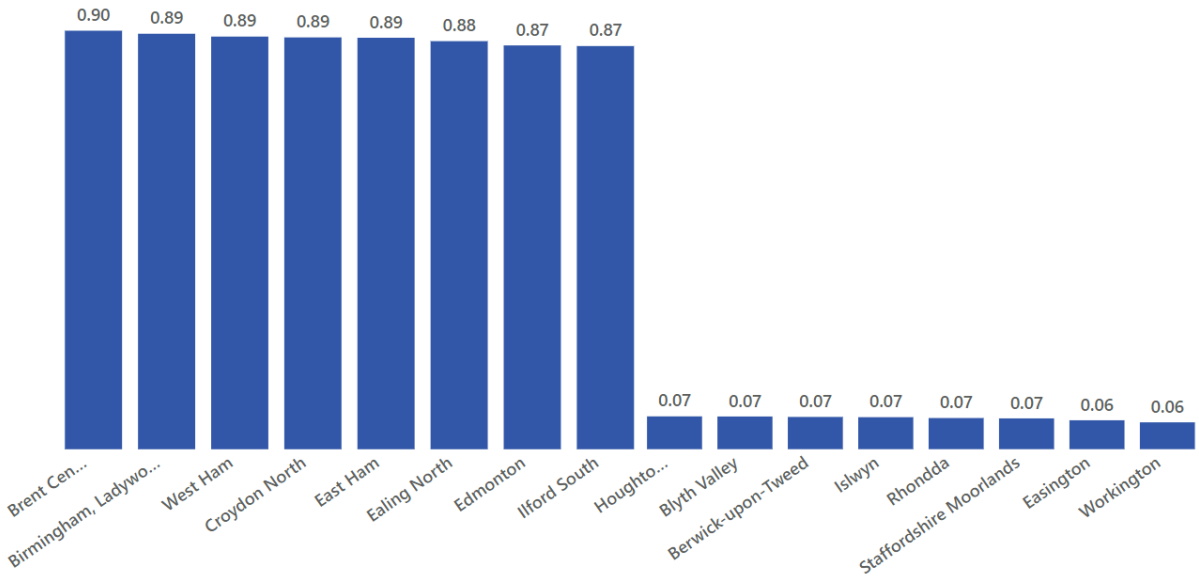
Finally, a calculated measure was used to to define D according to the expression for the diversity index defined above, where in was written in Power BI as follows:

```
D = 1 - ([Dnum] / [Dden])
```

Power BI automatically groups the columns and calculated measures according to the selected axis and filters, so using this calculated measure we can determine the diversity index D for every constituency or region.

Using the calculated measure D, column charts displaying the top 8 and bottom 8 constituencies via diversity index were created

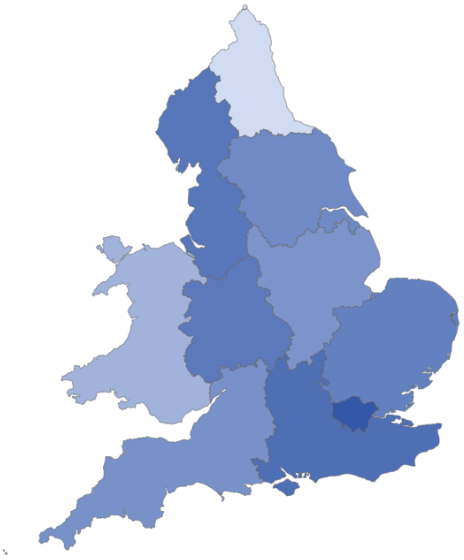
Top 8 and Bottom 8 Constituencies via Diversity Index



This chart shows that Brent Central is the most diverse constituency in the UK with a diversity index of 0.9, whilst Workington is the least diverse constituency with a diversity index of 0.06.

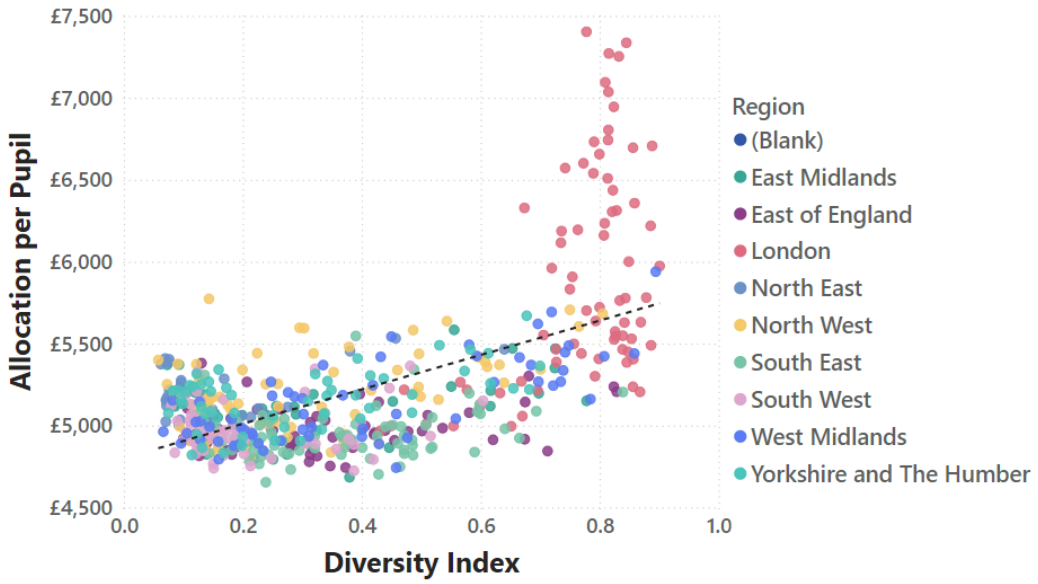
Next, we seek to evaluate the diversity index by region. Once again, the same expressions and calculated measures are used as Power BI can automatically group by when we select the category. A shape map divided by each region of England where the darker the shading the higher the regional diversity index was used to display this, which is shown as follows.

England and Wales Regional Diversity Index



From this shape map, we can see that London is the most diverse region as it has the darkest shading, whilst North East is the least diverse region as it has the lightest shading.

Furthermore, the correlation chart from Question 6 is remade with the x-axis using Diversity Index instead, as shown below.



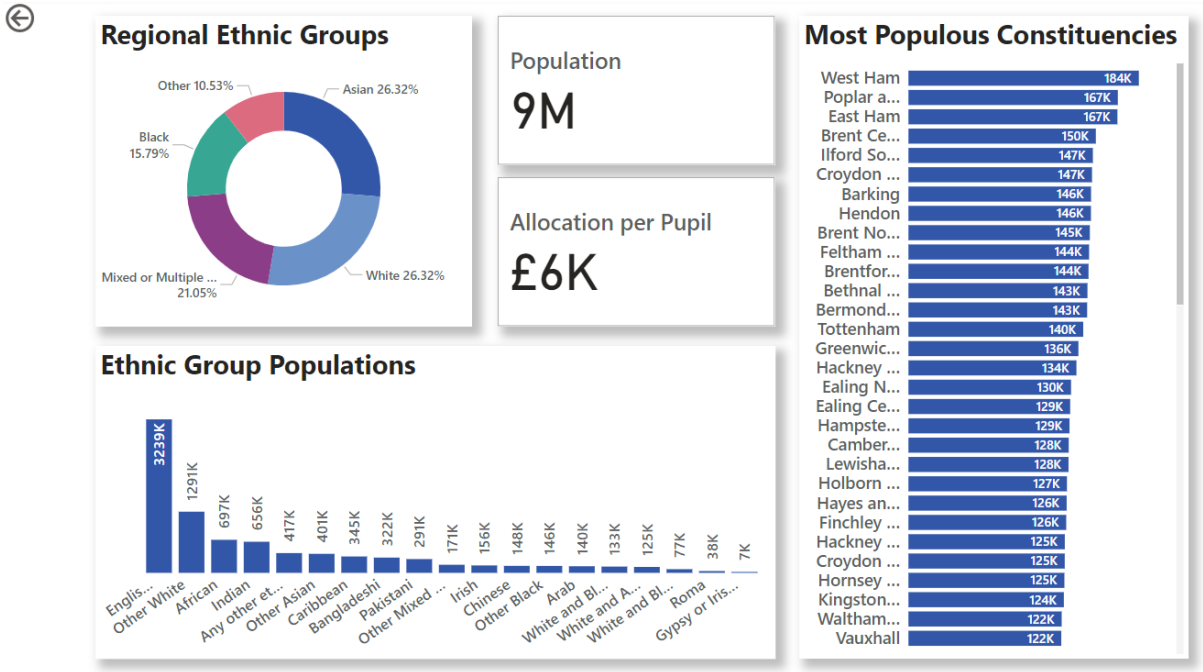
This shows the same conclusion as outlined before but uses the newly defined Diversity Index. It shows that as Diversity Index of a constituency increases, the school funding per pupil of that constituency increases. But the legend shows that most of the constituencies with the highest Diversity Index are London, so this could just mean that London constituencies receive the highest school funding per pupil.

Detailed Ethnic Group Analysis

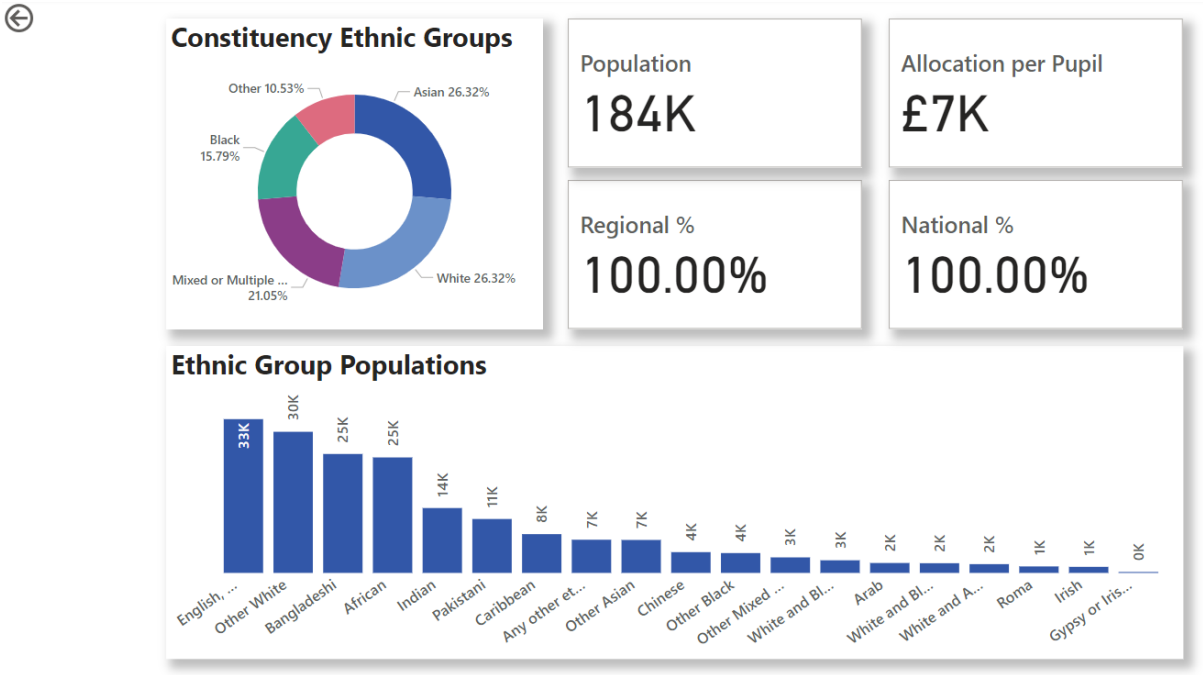
9. What are the detailed demographic characteristics of specific ethnic groups within constituencies?

Using the drill-through function, two details pages for regional breakdown and constituency breakdown were created. This allows users to go deeper and view more details for specific regions or constituencies when navigating the main dashboard (see Question 11).

By right-clicking one of the regions in the slicer or shape map of the main dashboard, the user can drill through to the regional breakdown page and see detailed information for the selected region including a bar chart of the most populous constituencies in the region, a column chart of the ethnic group populations, a donut chart of the broad ethnic group distribution and the total population and average school funding allocation per pupil The drill through page for regions with London as the drilled through region is shown below.



Next, a drill through details page for constituencies is created. By right clicking on a constituency in the main dashboard or the regional breakdown drill through page, the user can drill through to the constituency breakdown page. They can view even more detailed information for each constituency including a column chart of the ethnic group populations, a donut chart of the broad ethnic group distribution, the population and the school funding allocation per pupil. Furthermore, the user can filter by broad ethnic group and ethnic group on this page. The constituency breakdown drill through page with West Ham as the selected constituency is displayed below.



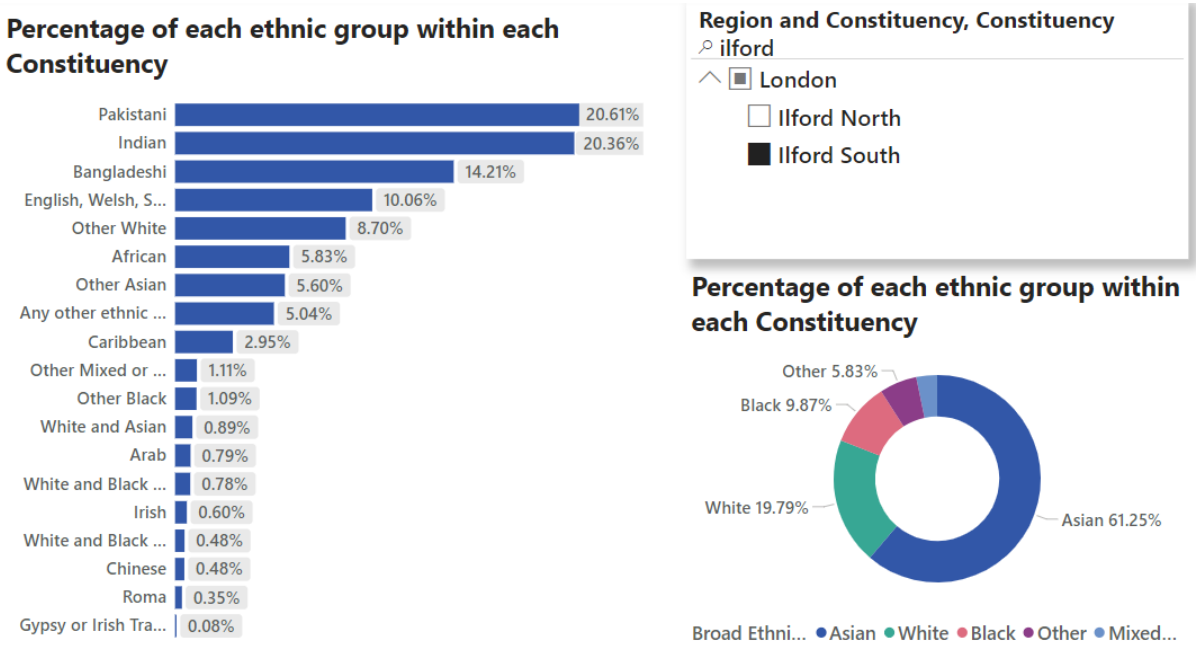
By creating relationships with other datasets, these drill through pages can include even more detailed information such as the constituency voting breakdown, the sitting MP of the constituency, regional unemployment rates and more.

Policy and Planning

10. How can this data inform policymaking and resource allocation?

To identify constituencies with significant ethnic minority populations, the diversity index value is utilised. The higher the diversity index D of a constituency, the more ethnic minorities are represented in the ethnic group distribution of the constituency.

In Question 8, the top 8 constituencies with the highest diversity index were found. Using the column and donut charts in Question 1 along with a Constituency slicer, these 8 constituencies were investigated to determine the exact ethnic group breakdowns. A sample of this with Ilford South as the sliced constituency is shown below.



For 2 of these constituencies, East Ham and Ilford South, there is a supermajority (>50%) of the Asian broad ethnic group at 51.54% and

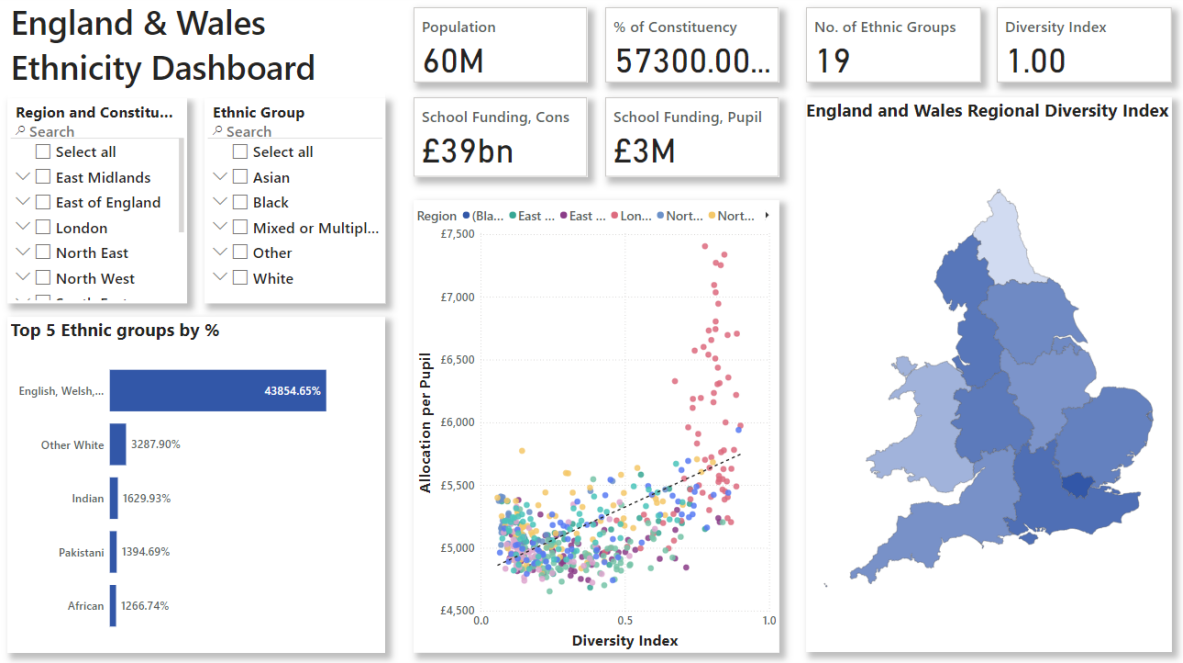
The other 6 constituencies have no supermajority i.e. the most populous of the broad ethnic groups is less than 50% of the total distribution

By using these two, the Diversity Index and the ethnic distribution dashboard, policy makers will be able to identify constituencies with a high degree of diversity and quickly identify the ethnic breakdown of these constituencies. Then, depending on the ethnic breakdown, they can come up with policy recommendations. For example, in East Ham and Ilford South, the Asian broad ethnic group which is legally and governmentally recognised as a minority ethnic group is the majority ethnic group within these specific constituencies. As such, nationwide policy for minority ethnic groups may not be applicable or recommendable for these constituencies in particular. On the other hand, the other 6 constituencies have equally high diversity indexes but have a more balanced ethnic distribution. This shows that single data values do not describe the full situation so policymakers should consider multiple different options via a dashboard to enact appropriate policies.

Interactive Dashboards

11. Can we create interactive dashboards for real-time data exploration?

The interactive dashboard can be accessed via the submitted Power BI file. A screenshot of the interactive dashboard is shown below.



This dashboard allows the user to filter by region and constituencies within the region, broad ethnic group and ethnic groups within the broad ethnic group. There is a shape map displaying diversity index by region, which can also be interacted with to filter by region. It features numerous cards to display key information such as total population, school funding and the diversity index. When no filters are selected, it shows the figures for the entirety of England and Wales, and these cards will update dynamically as filters are applied. It also features two charts, one showing the correlation between school funding and diversity index for each constituency and the other showing the top 5 most populous ethnic groups in the constituency.

Drill down capabilities were added to this dashboard, including drill through to a regional breakdown and constituency breakdown. These can be reached by right clicking a region name in the slicer or shape map, or the constituency name in the slicer, and selecting drill through to the respective details page. Refer to Question 9 to see an outline of the drill through pages.