TheData Open

SUMMER INVITATIONAL DATA OPEN

VIRTUAL DATATHON 2020

The Socioeconomic Impact of the 2012 Olympics in London.

Team 21:
Alp Aribal
Bertie Ellison-Wright
Pamelpreet Jhinger
Evonne Lee

Date: July 19, 2020

1 Introduction

1.1 Background

Hosting an Olympic games is often seen as an investment in the host city due to the increased spending infrastructure, expected boosts to tourism and the economy. Hence, due to the perceived benefits, hosting the Olympics Games became highly competitive, with many host cities promising extravagant infrastructure to secure their bids. The soaring costs, coupled with scandals in past Olympic has led to public skepticism about hosting the Olympic games. Thus, in this report, we will analyse data about Olympic host city/country, in combination with supplementary data sets, in order to increase our understanding of how hosting the Olympics has impacted the respective hosts and we will also dive into its impact on London's house pricing.

1.2 Our question

The UK stated in their bid for hosting the 2012 Olympics that they would benefit London's economy by boosting employment. We took a look at the London data sets provided relating to earnings, economic activity and infrastructure spending and found the data sets limited the scope of specific hypotheses we could hone in on.

After looking at supplementary data provided on `https://data.london.gov.uk/`, we found a rich data set created from the Annual Population Survey (APS), produced by the Office for National Statistics, providing data from 2004 to 2019 on the ethnicities of the working age population as well as their employment status within the UK, the major countries and regions of the UK and the London boroughs.

This led to the question we will investigate in this report:

> **Did hosting the London 2012 Olympics have a lasting impact on the employment of ethnic minorities?**

In particular, we are interested in whether the mean population and percentage employment of ethnic minorities in London has increased as a result of hosting the London 2012 Olympics.

2   Technical exposition

2.1   London House Pricing

The Mayor of London, Boris Johnson, described the area surrounding the Olympic Park as 'London's single most important regeneration project for the next 25 years'. The Mayor committed to delivering sustainable economic and social development throughout east London, and closing the deprivation gap between the six Olympic host boroughs and the rest of the city. These six boroughs, Barking and Dagenham, Greenwich, Hackney, Newham, Tower Hamlets and Waltham Forest, which we will refer to as the 'host boroughs' henceforth, saw significant investment into infrastructure in the years leading up to the Olympics.

We see that in 2012, these host boroughs had the 10th, 18th, 7th, 1st, 3rd and 9th highest percentages of BAME people in their population respectively compared to the other boroughs (33 in total), as shown in Fig. 1, and made up 24.6% of the total BAME population in London, compared to 18.6% of the general London population.

We then looked at house prices in each borough, in order to ascertain the impact of the Olympics on the cost of living in these six boroughs. In mid-2005, when London won its bid to host the Olympics, the host boroughs had the 1st, 7th, 12th, 5th, 22nd and 8th lowest house prices respectively out of all the boroughs in London, suggesting that in general, these areas were more low-income than elsewhere in the city.

Fig. 2 shows the percent increase in house price from mid-2005 until 2018, with the six host boroughs shown in thick black. Fig. 3 shows the percent increase in house price from 2009 to 2015 (three years before the Olympics to three years after) on a map of London, with Olympic venues shown as pink dots, which shows how it is generally the more central boroughs that saw the greatest increase in house price. Fig. 6 in the appendix shows the same visualisation, with only the host boroughs.

In order to determine whether the trend in change in house price was different in the host boroughs compared to the rest of London, we calculated the slope of the linear fit of the house price data between mid-2005 and 2012, and compared it to the slope of the linear fit of the data from 2012 until 2018. The slopes of the trend before and after the Olympics in each borough are shown in figures 4 and 5.

2.2   Employment of Ethnic Minorities

2.2.1   Features

- year: Year of observed ethnicities employed per borough.

- `area`: Area corresponding to where employees of a particular ethnicity reside. Includes London's boroughs, major regions of England and countries of the UK.

- `ethnic_group`: Broad ethnic category in which employees fall under.

    - `White`: White British and/or European.
    - `Minority`: Total of all minority groups, including `Mixed`, `Indian`, `Black`, `Pakistani/Bangladeshi` and `Other` ethnicities. `Minority` is NOT mutually exclusive to the ethnic groups, it is a union of them all.

- `number`: Number of employees (of working age) falling under the `ethnic_group` specified working in `area`.

- `denominator`: Total working age population of falling under the `ethnic_group` specified within `area`.

- `percent`: `number` / `denominator` i.e. the percentage of the working age population of `area` of `ethnic_group` descent that are employed.

2.2.2   Data manipulation

The employment data for the major ethnic categories, white and ethnic minority, were fully populated however the specific ethnicities that the ethnic minority consisted of were not. We made the choice not to linearly interpolate this data since we were not particularly interested in a specific ethnicity's working age population, however this led to rendering line graphs that were missing points for some ethnicities.

2.2.3   Exploratory Data Analysis

Fig. 7 shows the percentage employment per ethnic group in the hosting Olympic boroughs and Fig. 8 shows the population employed by ethnic group within the same region. We can observe that pre-2012, the percentage employment of ethnic minorities sporadically but insignificantly changes and so stays around about the same but starts to pick up around 2012 and then continues to increase after the year of the Games.

We can also see a similar pattern in the rest of London, excluding the hosting boroughs in Fig. 9 and Fig. 10 where employment of ethnic minorities increased by around 10%.

When we compare this against the employment in the rest of the UK, excluding London, in Fig. 11, we can see that, although the trend is the same, the employment of ethnic minorities only went up by around 5%. This suggests a potential correlation

between the attractiveness of London in particular within the UK, pushed through Olympic campaigns.

2.3   Hypothesis Testing

Table 1: Sample statistics generated from the employment ethnicity data.

Year range	Sample mean $\overline{x}$	Sample variance s_{n-1}^2	Sample standard deviation s_{n-1}
2004 to 2011	$1,094,012.5$	$14,319,044,106.1$	$119,662.2$
2012 to 2019	$1,488,712.5$	$13,553,304,106.1$	$116,418.7$

We only have 8 samples (years) of data before 2012 and 8 samples after (and including) 2012 and we want to compare the population means of these two sets, in particular whether the population mean after 2012 was higher. This is an opportunity for a one-sided hypothesis test looking for a difference in population mean.

We shall assume that the years 2004 to 2011 and 2012 to 2019 are two independent samples of independent and identically distributed variables, the population of ethnic minorities for each year.

The Annual Population Survey is a sampled survey, therefore we do not know the true population mean before (μ_{before}) and after (μ_{after}) nor the true sample variance before ($s_{n-1\text{ before}}^2$) and after ($s_{n-1\text{ after}}^2$).

The null hypothesis $H_0 : \mu_{\text{before}} = \mu_{\text{after}}$ for this hypothesis test is that the population mean of ethnic minorities employed in London has not changed significantly from before 2012 to after 2012.

The alternative hypothesis $H_1 : \mu_{\text{before}} < \mu_{\text{after}}$ suggests that the population mean of ethnic minorities employed in London has increased significantly from before 2012 to after 2012.

We will conduct the hypothesis test with a Type I error rate of 0.01 (significance level of 1%) and using the Student's t-distribution. Using this distribution for a hypothesis test investigating for a difference in population means requires a bias-corrected sample pooled variance $s_{n_x+n_y-2}^2$.

$$s_{n_x+n_y-2}^2 = \frac{n_x-1}{n_x+n_y-2}s_{n_x-1}^2 + \frac{n_y-1}{n_x+n_y-2}s_{n_y-1}^2 = 13,936,174,106.1$$

Assuming H_0, applying the Student's t-distribution to the data along with the bias-corrected sample pooled variance gives

$$T = \frac{(\overline{X}-\overline{Y})}{s_{n_x+n_y-2}\sqrt{\frac{1}{n_x}+\frac{1}{n_y}}} \sim \Theta$$

The rejection region for the null hypothesis H_0 of a one-sided hypothesis test at the 10% significance level for two population means with sample sizes 8 and 8 is

$$R = \{t \mid |t| > t_{14,0.99}\} = \{t \mid |t| > 2.624494\}$$

The test statistic (with the appropriate parameters we have calculated inputted)

$$t = \frac{\overline{x} - \overline{y}}{s_{n_x + n_y - 2}\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} = -6.686907880637657$$

We can easily see that $6.69 > 2.62$ therefore we reject the null hypothesis H_0 at the 1% significance level and accept the alternative hypothesis H_1.

2.4   Interpretation of Hypothesis Test Result

The results tell us that there has been a significant increase in the population of ethnic minorities employed in London after 2012. This suggests a correlation between London hosting the Olympics in 2012 and the increase in employment of ethnic minorities since the data used was separated into two sets dependent on the time of the Olympics.

However, we cannot ignore the fact that these results do not directly suggest a causal relationship between London hosting the 2012 Olympics and an increase in employment of ethnic minorities.

Without the pretense that the Olympics was hosted in 2012, the data itself is blind to any economic activities that have occurred since 2004. The data itself was also aggregated into records defining geographic regions, potentially losing some useful attributes such as individual reasoning for employment and whether those employed are immigrants.

Unfortunately there is not a data set available to the public that has the level of depth we are looking for, however if there was one, we believe we could prove a causal relationship between the two variables in question.
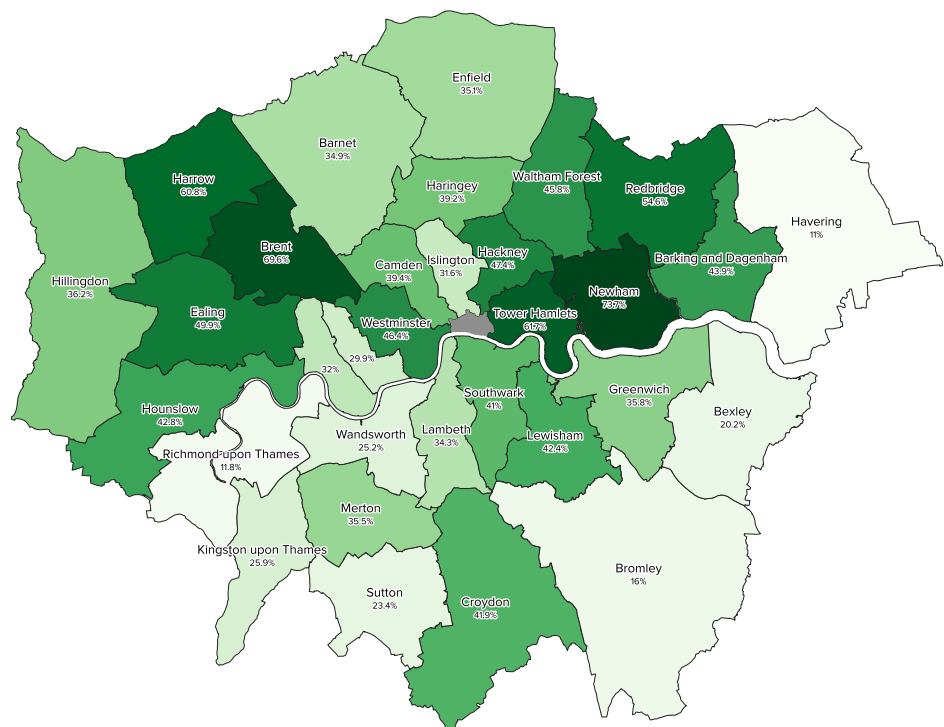
3   Appendix

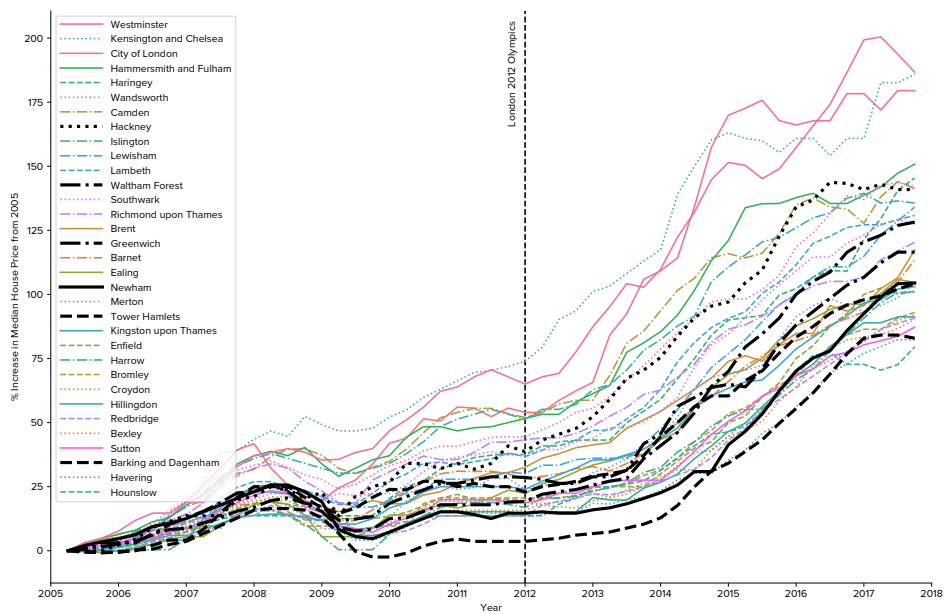**Figure 1:** Percentage of BAME in each borough in 2012.

Figure 2: Percent increase in house price in London boroughs from 2005 to 2018.
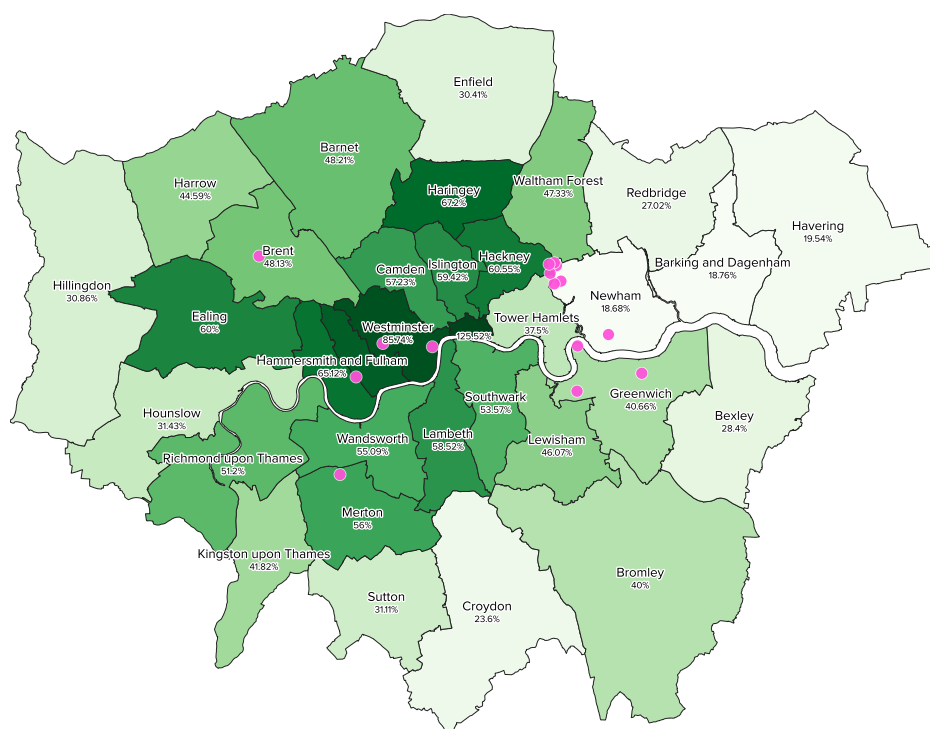
Figure 3: Map of London showing the percent increase in house price in each borough from 2009 to 2015, with Olympic venues represented as pink dots.
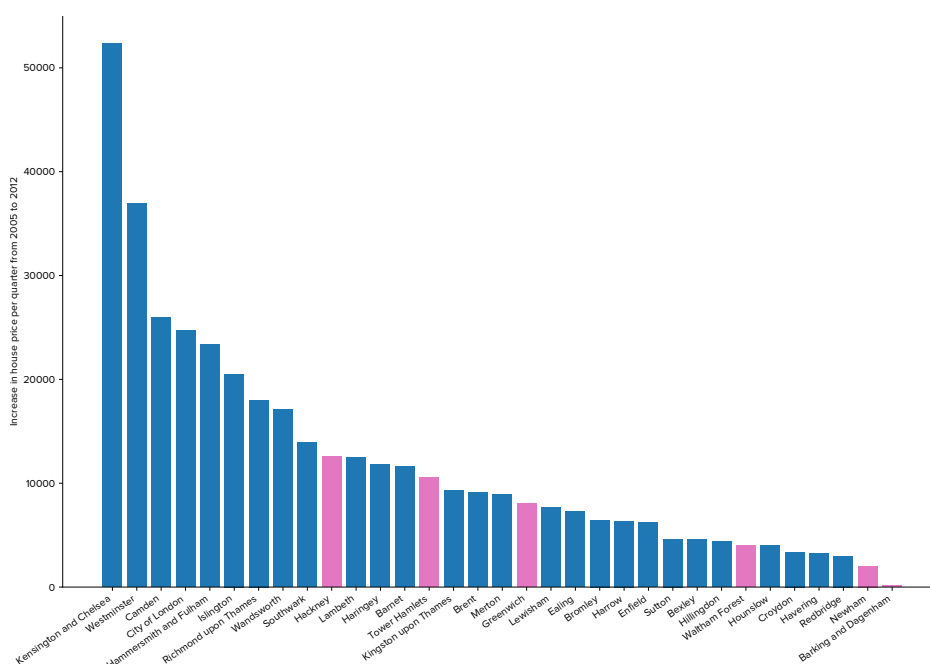

Figure 4: Increase in house price per quarter for each borough from 2005 to 2012, in the build-up to the games, based on a linear fit of the data, with host boroughs in pink.
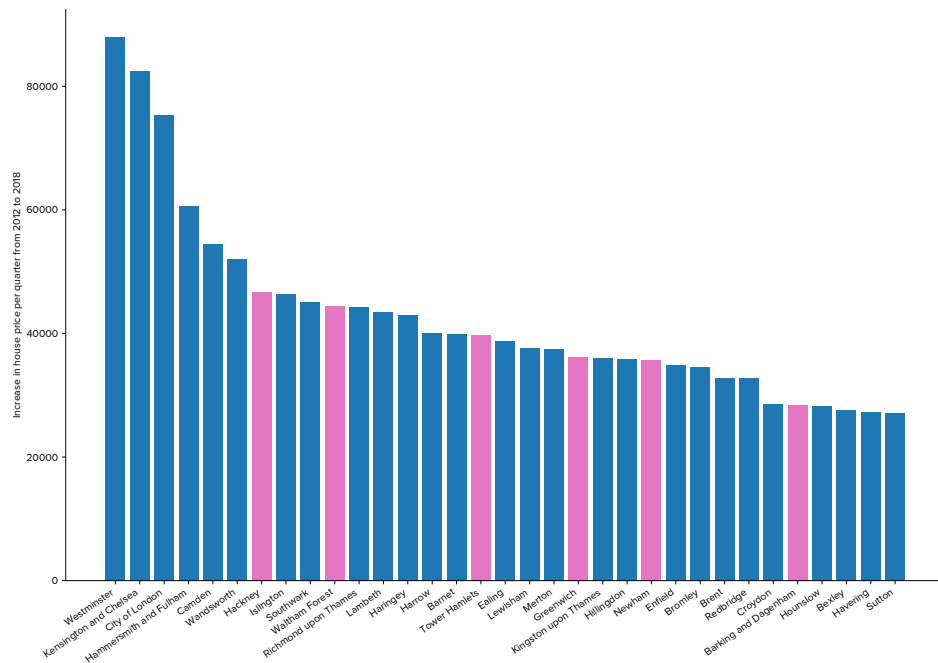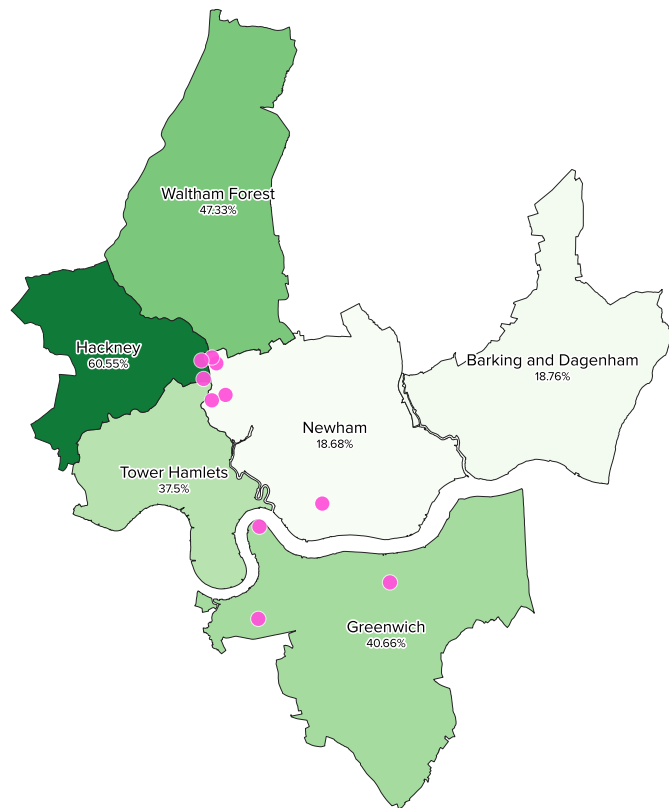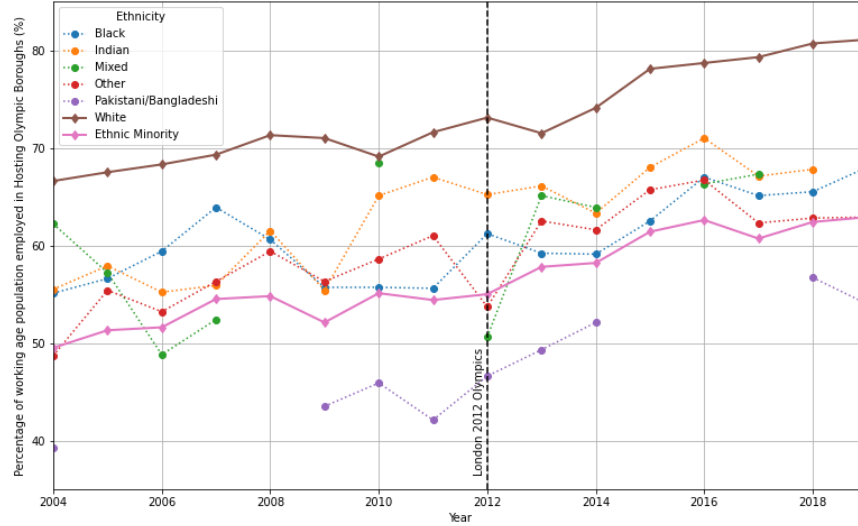
Figure 5: Increase in house price per quarter for each borough from 2012 to 2018, after the games, based on a linear fit of the data, with host boroughs in pink.

Figure 6: Map showing the percent increase in house price in each host borough from 2009 to 2015, with Olympic venues represented as pink dots.

Figure 7: Percentage of working age population employed in the hosting Olympic boroughs by ethnic group yearly.
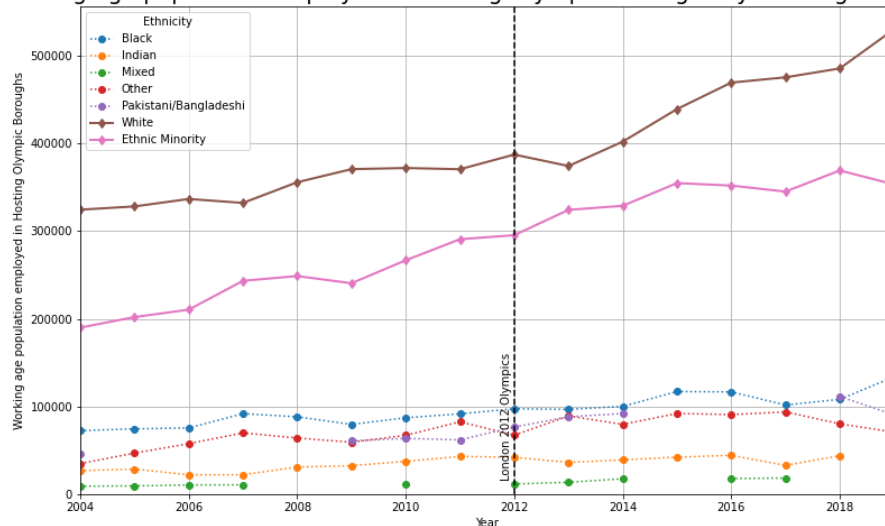
Figure 8: Population of working age population employed in the hosting Olympic boroughs by ethnic group yearly.

Figure 9: Percentage of working age population employed in London excluding the Olympic boroughs by ethnic group yearly.
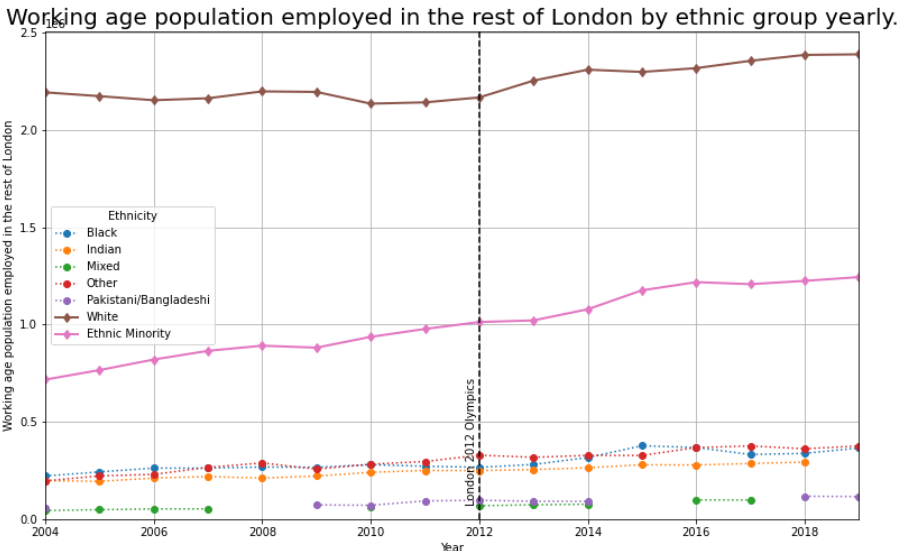

Figure 10: Population (in millions) of working age population employed in London excluding the Olympic boroughs by ethnic group yearly.
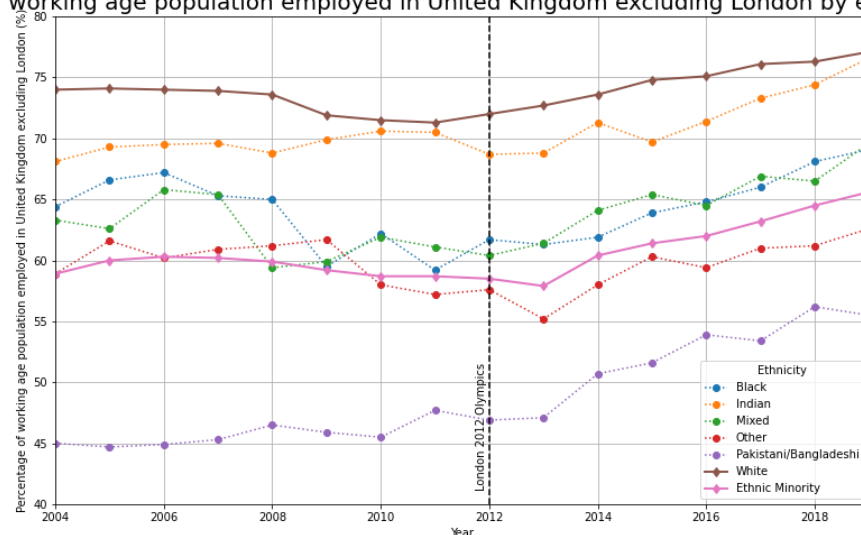
Figure 11: Percentage of working age population employed in the UK excluding London by ethnic group yearly.