Jolyon Miles-Wilson

02/11/2021

Table of Contents

[Overview 2](#_Toc86921567)

[R packages used 2](#_Toc86921568)

[Data 2](#_Toc86921569)

[Children achieving Good Level of Development 4](#_Toc86921570)

[Pupils achieving at least Level 4 in GCSE Maths and English 4](#_Toc86921571)

[Children looked after by the local authority 5](#_Toc86921572)

[Employment 5](#_Toc86921573)

[Rate of Police Stop and Search 6](#_Toc86921574)

[Households that are statutorily homeless or at risk of becoming homeless 7](#_Toc86921575)

[Analysis and Results 7](#_Toc86921576)

[Birmingham 7](#_Toc86921577)

[Children achieving Good Level of Development 7](#_Toc86921578)

[Pupils achieving at least Level 4 in GCSE Maths and English 10](#_Toc86921579)

[Children looked after by the local authority 13](#_Toc86921580)

[Employment 16](#_Toc86921581)

[Rate of Police Stop and Search 18](#_Toc86921582)

[Households that are statutorily homeless or at risk of becoming homeless 21](#_Toc86921583)

[Summary 24](#_Toc86921584)

[Haringey 26](#_Toc86921585)

[Children achieving Good Level of Development 26](#_Toc86921586)

[Pupils achieving at least Level 4 in GCSE Maths and English 29](#_Toc86921587)

[Children looked after by the local authority 32](#_Toc86921588)

[Employment 35](#_Toc86921589)

[Rate of Police Stop and Search 37](#_Toc86921590)

[Households that are statutorily homeless or at risk of becoming homeless 40](#_Toc86921591)

[Summary 43](#_Toc86921592)

[Limitations 45](#_Toc86921593)

# Overview

This document provides a description of the procedures undertaken to produce the first versions of Black Thrive’s Shared Measurement System dashboards for Birmingham and Haringey.

# R packages used

library(kableExtra)  
library(ukpolice)  
library(leaflet)  
library(gmodels)  
library(ggplot2)  
library(tidyverse)

# Data

I aimed to produce for Birmingham and Haringey the same set of indicators implemented in the current Lambeth dashboard. It has not yet been possible to acquire data for Birmingham and Haringey relating to the mental health indicators that are included in the Lambeth dashboard. The table below lists the indicators included in the Haringey and Birmingham dashboards, and the sources of their data.

Indicator

Source

Date

Source link

Good level of development

Early Years Foundation Stage Profile results

2018-2019 academic year

<https://www.gov.uk/government/statistics/early-years-foundation-stage-profile-results-2018-to-2019>

Level 4+ GCSE Maths & English

Key Stage 4 performance

2019-2020 academic year

<https://explore-education-statistics.service.gov.uk/find-statistics/key-stage-4-performance-revised#dataDownloads-1>

Children looked after by local authority

Children looked after in England including adoption

2018-2019 financial year

<https://www.gov.uk/government/statistics/children-looked-after-in-england-including-adoption-2018-to-2019>

16-64 year olds in employment

NOMIS Annual Population Survey

12 months to March 2021

<https://www.nomisweb.co.uk/datasets/apsnew>

Individuals subject to Stop and Search: Birmingham

Police API

12 months to end of August 2021

<https://data.police.uk>

Individuals subject to Stop and Search: Haringey

Metropolitan Police Stop and Search Dashboard Data

12 months to end of August 2021

<https://data.london.gov.uk/dataset/mps-stop-and-search-public-dashboard-data>

Homeless or at risk of becoming homeless

Live tables on homelessness; Detailed local authority level tables

2020-2021 financial year

<https://www.gov.uk/government/statistical-data-sets/live-tables-on-homelessness#homelessness-summary-local-authority-level-tables>

Adult population

Annual Population Survey

2021

<https://www.nomisweb.co.uk/datasets/apsnew>

Child population

Census 2011

2011

<https://www.nomisweb.co.uk/census/2011/dc2101ew>

Below I provide a brief description of each indicator and how the data were sourced and organised for each.

## Children achieving Good Level of Development

This indicator describes the number of children aged 4-5 years old who reach a “Good Level of Development” (GLD), defined as the achievement of expected level of development for Early Learning Goals and literacy and mathematics.

Data for this indicator were acquired from the Early Years Foundation Stage Profile results: 2018-2019. This data provides the number of children reaching GLD and the total number of pupils according to the ethnic categories Asian, Black, Chinese, Mixed, and White. Subsets of this dataset were produced for Haringey and Birmingham by filtering the data based on the following criteria:

Variable name

Value

la\_name

Birmingham | Haringey

characteristic

Ethnicity

time\_period

201819

## Pupils achieving at least Level 4 in GCSE Maths and English

This indicator describes the number of pupils achieving at least Level 4 in GCSE Maths and English. Data were acquired from Key Stage 4 performance 2019/20, the variables of interest of which were *t\_pupils* (i.e., total number of pupils), *t\_l2basics\_94* (i.e. total number achieving at least Level 4), and *pt\_l2basics\_94* (i.e. percentage achieving at least Level 4). Subsets of this data were produced for Haringey and Birmingham by filtering the data according to the following criteria:

Variable name

Value

la\_name

Birmingham | Haringey

breakdown

Ethnic major

## Children looked after by the local authority

This indicator describes the number of children looked after by the local authority. Data were acquired from *Children looked after in England including adoption: 2018 to 2019*. The following filter was applied to acquire data specific to Haringey/Birmingham:

|  |  |
| --- | --- |
|  |  |

Variable name

Value

geog\_n

Birmingham | Haringey

The variables of interest in this dataset were *CLA\_Black* and *CLA\_White*, which correspond to the number of children looked after by the local authority who are Black and White, respectively. In order to calculate the rate at which children were looked after by the local authority, these data were supplemented by child population statistics from the 2011 Census (*Census 2011: DC2101EW*). The data selected from this latter dataset were:

Variable name

Value

Geography

Birmingham | Haringey

Age

0-17 years old

Ethnicity

Black, White

Sex

All

## Employment

This indicator describes the number of 16-64 year olds in employment. The data for this indicator were acquired from the *Annual Population Survey* by querying the data on the NOMIS website. The query paramters used were:

Variable name

Value

Geography

Birmingham | Haringey

Date

12 months to March 2021

Employment rate

aged 16-64 employment rate for all Black or Black British, aged 16-64 employment rate - White

This data provides a *numerator*, which is the total number in employment, and a *denominator*, which is the population estimate for the ethnicity.

## Rate of Police Stop and Search

The procedure for Stop and Search data differed by local authority. For Haringey, Stop and Search data were acquired via the Metropolitan Police Stop and Search Dashboard data. Data for just Haringey was acquired by filtering the dataset by the variable *Borough of Stop*. For the purposes of the present analysis, an individual’s ethnicity was taken as the ethnicity attributed to the individual by the officer (variable name *EA Group*).

For Birmingham, I used the Police API to query Stop and Search data. To do this, it was first necessary to define the area of Birmingham using latitude and longitude coordinates. I used coordinates of the city boundary acquired from <https://mapit.mysociety.org/area/2514.html>. This provided 3634 coordinate points that define the area of Birmingham. Because the Police API cannot accept such a large number of coordinates, I reduced this to 104 coordinate points as below:

# Get coordinates of Birmingham area  
file <- jsonify::from\_json("../data/birm\_geo.json") # read geojson defining Birmingham  
df <- as.data.frame(file[["coordinates"]][[1]]) # make dataframe  
colnames(df) <- c("long","lat") # name cols  
  
# coordinates need to be reduced because there are too many for ukpolice to handle  
reduction\_factor <- 3530 # maximum size possible for api is 104 (i.e. 3634 - 3530)  
f <- round(seq(2,nrow(df),length = reduction\_factor),0) # create list of rows to remove  
reduced\_df <- df[-c(f),] # remove listed rows  
reduced\_df\_2 <- subset(reduced\_df, select = c("lat","long")) # swap lat and long columns (for ukpolice)

This produced the below area as an estimation of the geographic limits of Birmingham:

I then used the R package *ukpolice* to search the Police API for Stop and Search records within this area for the period September 2020 - August 2021. In this dataset, the variable *officer\_defined\_ethnicity* records the ethnicity attributed to the individual by the officer, and this was used as the ethnicity variable in the present analysis.

This data was supplemented by population estimates data for Haringey and Birmingham from the 2021 *Annual Population Survey*.

## Households that are statutorily homeless or at risk of becoming homeless

This indicator describes the number of households who are owed prevention or relief duty, meaning that they are statutorily homeless or at risk of being so. The data were acquired from *Live tables on homelessness; Detailed authority level tables: Financial year 2020-21*. The subset used for the current analysis is *Table A8: Ethnicity of main applicants assessed as owed a prevention or relief duty by local authority 2020-21 financial year*. From this table, I selected the data corresponding to the Black and White populations of Haringey and Birmingham.

# Analysis and Results

I used Chi-square and Fisher exact tests to analyse the data. For each indicator, I first created a crosstabulation containing the frequency of Black and White individuals for which the indicator outcome was true and the frequency for which it was false. Using this crosstabulation, I then tested whether the odds of the indicator outcome were significantly different for the Black population compared to the White population. This was done using the *CrossTable* function from the *gmodels* package in R. This function provides statistics for the Chi-square and Fisher exact tests as well an estimate of the odds ratio and its confidence interval. Finally, where appropriate, odds ratios were converted into values of relative risks using the *riskratio* function from the *epitools* package.

## Birmingham

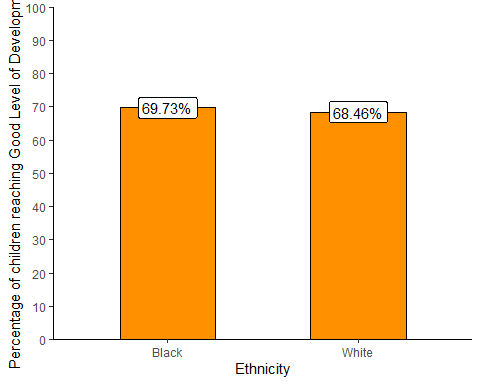
I first report on the results of the Birmingham indicators. For each indicator, I provide the code for preparing and analysing the data, followed by a written report of the findings.

### Children achieving Good Level of Development

## good development  
data <- read.csv("../data/birmingham\_good\_development\_2018\_2019\_v2.csv")  
  
total\_data <- data %>%  
 subset(gender == "Total") %>%  
 subset(ethnicity != "Total")  
  
row.names(total\_data) <- total\_data$ethnicity # set row names for indexing  
  
# prepare data frame for crosstabs  
  
black <- data.frame("black" = c("gld\_number" = total\_data["Black", "gld\_number"], "not\_gld\_number" = total\_data["Black", "number\_of\_pupils"] - total\_data["Black", "gld\_number"]))  
  
white <- data.frame("white" = c("gld\_number" = total\_data["White", "gld\_number"], "not\_gld\_number" = total\_data["White", "number\_of\_pupils"] - total\_data["White", "gld\_number"]))  
  
bw\_mat <- as.matrix(cbind(black, white)) # combine black and white into one matrix  
  
xtab <- CrossTable(bw\_mat, fisher = T, chisq = T, expected = T) # save xtab

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | Expected N |  
## | Chi-square contribution |  
## | N / Row Total |  
## | N / Col Total |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 7027   
##   
##   
## |   
## | black | white | Row Total |   
## ---------------|-----------|-----------|-----------|  
## gld\_number | 1168 | 3664 | 4832 |   
## | 1151.786 | 3680.214 | |   
## | 0.228 | 0.071 | |   
## | 0.242 | 0.758 | 0.688 |   
## | 0.697 | 0.685 | |   
## | 0.166 | 0.521 | |   
## ---------------|-----------|-----------|-----------|  
## not\_gld\_number | 507 | 1688 | 2195 |   
## | 523.214 | 1671.786 | |   
## | 0.502 | 0.157 | |   
## | 0.231 | 0.769 | 0.312 |   
## | 0.303 | 0.315 | |   
## | 0.072 | 0.240 | |   
## ---------------|-----------|-----------|-----------|  
## Column Total | 1675 | 5352 | 7027 |   
## | 0.238 | 0.762 | |   
## ---------------|-----------|-----------|-----------|  
##   
##   
## Statistics for All Table Factors  
##   
##   
## Pearson's Chi-squared test   
## ------------------------------------------------------------  
## Chi^2 = 0.9593996 d.f. = 1 p = 0.3273382   
##   
## Pearson's Chi-squared test with Yates' continuity correction   
## ------------------------------------------------------------  
## Chi^2 = 0.901141 d.f. = 1 p = 0.342476   
##   
##   
## Fisher's Exact Test for Count Data  
## ------------------------------------------------------------  
## Sample estimate odds ratio: 1.061335   
##   
## Alternative hypothesis: true odds ratio is not equal to 1  
## p = 0.3338691   
## 95% confidence interval: 0.94079 1.198239   
##   
## Alternative hypothesis: true odds ratio is less than 1  
## p = 0.8436899   
## 95% confidence interval: 0 1.175317   
##   
## Alternative hypothesis: true odds ratio is greater than 1  
## p = 0.1712876   
## 95% confidence interval: 0.9588271 Inf   
##   
##   
##

The results indicate that there was no significant difference between Black children and White children in the odds of achieving a Good Level of Development at age 5, (1, *N* = 7027) = 0.96, *p* = 0.327. The plot below shows that 69.73% of Black children and 68.46% of White children reached a Good Level of Development in 2018-2019.

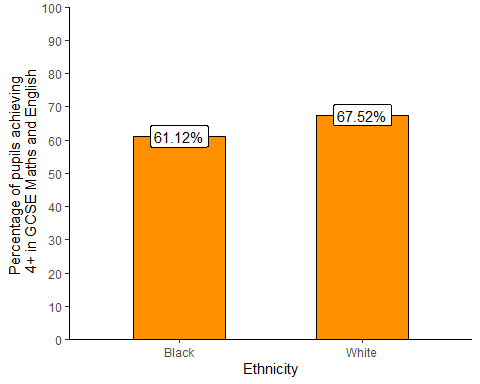


### Pupils achieving at least Level 4 in GCSE Maths and English

data <- read.csv("../data/birmingham\_4\_plus\_gcse\_2019\_2020\_v2.csv")  
  
# subset and rename  
total\_data <- data %>%  
 subset(gender == "Total") %>%  
 subset(ethnicity == "BLACK" | ethnicity == "WHITE") %>%  
 rename(total\_4\_plus = t\_l2basics\_94,  
 percent\_4\_plus = pt\_l2basics\_94)  
  
# rename ethnicity levels and row names  
total\_data$ethnicity <- as.factor(total\_data$ethnicity)  
levels(total\_data$ethnicity) <- c("Black","White")  
row.names(total\_data) <- total\_data$ethnicity  
  
total\_data <- total\_data[,-c(1)] # remove unnecessary column  
  
# calculate number who didn't achieve level 4 plus  
total\_data <- total\_data %>%  
 mutate(  
 not\_gcse = t\_pupils - total\_4\_plus  
 )  
  
# prepare data frame for xtab  
black <- data.frame("black" = c("gcse" = total\_data["Black", "total\_4\_plus"], "not\_gcse" = total\_data["Black", "not\_gcse"]))  
white <- data.frame("white" = c("gcse" = total\_data["White", "total\_4\_plus"], "not\_gcse" = total\_data["White", "not\_gcse"]))  
bw\_mat <- as.matrix(cbind(black, white))  
  
# run stats  
xtab <- CrossTable(bw\_mat, fisher = T, chisq = T, expected = T)

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | Expected N |  
## | Chi-square contribution |  
## | N / Row Total |  
## | N / Col Total |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 6352   
##   
##   
## |   
## | black | white | Row Total |   
## -------------|-----------|-----------|-----------|  
## gcse | 1154 | 3014 | 4168 |   
## | 1238.851 | 2929.149 | |   
## | 5.812 | 2.458 | |   
## | 0.277 | 0.723 | 0.656 |   
## | 0.611 | 0.675 | |   
## | 0.182 | 0.474 | |   
## -------------|-----------|-----------|-----------|  
## not\_gcse | 734 | 1450 | 2184 |   
## | 649.149 | 1534.851 | |   
## | 11.091 | 4.691 | |   
## | 0.336 | 0.664 | 0.344 |   
## | 0.389 | 0.325 | |   
## | 0.116 | 0.228 | |   
## -------------|-----------|-----------|-----------|  
## Column Total | 1888 | 4464 | 6352 |   
## | 0.297 | 0.703 | |   
## -------------|-----------|-----------|-----------|  
##   
##   
## Statistics for All Table Factors  
##   
##   
## Pearson's Chi-squared test   
## ------------------------------------------------------------  
## Chi^2 = 24.05154 d.f. = 1 p = 0.000000937914   
##   
## Pearson's Chi-squared test with Yates' continuity correction   
## ------------------------------------------------------------  
## Chi^2 = 23.76892 d.f. = 1 p = 0.000001086225   
##   
##   
## Fisher's Exact Test for Count Data  
## ------------------------------------------------------------  
## Sample estimate odds ratio: 0.756425   
##   
## Alternative hypothesis: true odds ratio is not equal to 1  
## p = 0.000001184091   
## 95% confidence interval: 0.675428 0.8472768   
##   
## Alternative hypothesis: true odds ratio is less than 1  
## p = 0.0000006140868   
## 95% confidence interval: 0 0.8321658   
##   
## Alternative hypothesis: true odds ratio is greater than 1  
## p = 0.9999995   
## 95% confidence interval: 0.6876366 Inf   
##   
##   
##

The results indicate that in the 2019-2020 academic year, there was a significant difference between Black pupils and White pupils in the odds of achieving at least a Level 4 in GCSE Maths and English, (1, *N* = 6352) = 24.05, *p* < .001. The odds of a Black pupil achieving at least Level 4 in Maths and English were 0.76 times the odds for a White pupil (95% CI = 0.68 - 0.85).The plot below shows that 61.12% of Black pupils achieved at least Level 4 in English and Maths, compared to 67.52% of White pupils.

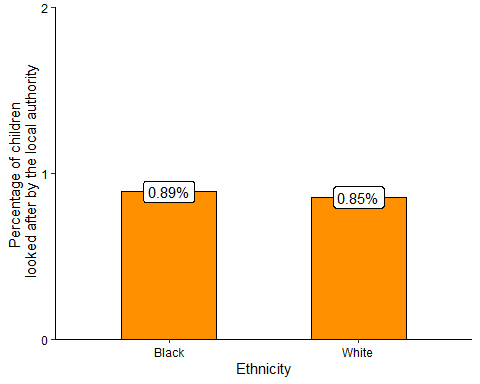


### Children looked after by the local authority

# children looked after  
data <- read.csv("../data/birmingham\_cla\_2018\_2019\_v2.csv")  
  
# prepare data for analysis  
total\_data <- data %>%  
 subset(ethnicity != "Total") %>%  
 mutate(  
 not\_cla = pop - cla,  
 percentage = as.numeric(percentage)  
 )  
  
total\_data <- total\_data[1:nrow(total\_data) - 1,] # drop unknown  
row.names(total\_data) <- total\_data$ethnicity  
  
# prepare xtab  
black <- data.frame("black" = c("cla" = total\_data["Black", "cla"], "not\_cla" = total\_data["Black", "not\_cla"]))  
white <- data.frame("white" = c("cla" = total\_data["White", "cla"], "not\_cla" = total\_data["White", "not\_cla"]))  
bw\_mat <- as.matrix(cbind(black, white))  
  
# run analysis  
xtab <- CrossTable(bw\_mat, fisher = T, chisq = T, expected = T)

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | Expected N |  
## | Chi-square contribution |  
## | N / Row Total |  
## | N / Col Total |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 144468   
##   
##   
## |   
## | black | white | Row Total |   
## -------------|-----------|-----------|-----------|  
## cla | 264 | 981 | 1245 |   
## | 256.156 | 988.844 | |   
## | 0.240 | 0.062 | |   
## | 0.212 | 0.788 | 0.009 |   
## | 0.009 | 0.009 | |   
## | 0.002 | 0.007 | |   
## -------------|-----------|-----------|-----------|  
## not\_cla | 29460 | 113763 | 143223 |   
## | 29467.844 | 113755.156 | |   
## | 0.002 | 0.001 | |   
## | 0.206 | 0.794 | 0.991 |   
## | 0.991 | 0.991 | |   
## | 0.204 | 0.787 | |   
## -------------|-----------|-----------|-----------|  
## Column Total | 29724 | 114744 | 144468 |   
## | 0.206 | 0.794 | |   
## -------------|-----------|-----------|-----------|  
##   
##   
## Statistics for All Table Factors  
##   
##   
## Pearson's Chi-squared test   
## ------------------------------------------------------------  
## Chi^2 = 0.305031 d.f. = 1 p = 0.5807455   
##   
## Pearson's Chi-squared test with Yates' continuity correction   
## ------------------------------------------------------------  
## Chi^2 = 0.2673821 d.f. = 1 p = 0.6050933   
##   
##   
## Fisher's Exact Test for Count Data  
## ------------------------------------------------------------  
## Sample estimate odds ratio: 1.039213   
##   
## Alternative hypothesis: true odds ratio is not equal to 1  
## p = 0.5733443   
## 95% confidence interval: 0.9031517 1.192428   
##   
## Alternative hypothesis: true odds ratio is less than 1  
## p = 0.7230483   
## 95% confidence interval: 0 1.167166   
##   
## Alternative hypothesis: true odds ratio is greater than 1  
## p = 0.3008236   
## 95% confidence interval: 0.9236985 Inf   
##   
##   
##

The results indicate that, in the 2018-2019 financial year, there was no significant relationship between child ethnicity and the odds of being looked after the local authority, (1, *N* = 144468) = 0.31, *p* = 0.581. The plot below shows that 0.89% of Black children and 0.85% of White children were looked after by the local authority in the 2018-2019 financial year.

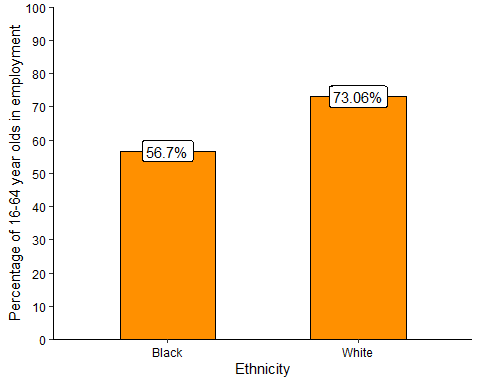


### Employment

# employment  
data <- read.csv("../data/birmingham\_employment\_rate\_2020\_2021\_v2.csv")  
  
# subset and organise data  
total\_data <- data %>%  
 subset(gender == "all") %>%  
 mutate(  
 percentage = as.numeric(percentage),  
 numerator = as.numeric(numerator),  
 denominator = as.numeric(denominator),  
 not\_in\_employment = denominator - numerator,  
 )  
   
row.names(total\_data) <- total\_data$ethnicity  
  
# prepare xtab  
black <- data.frame("black" = c("employed" = total\_data["black", "numerator"], "not\_employed" = total\_data["black", "not\_in\_employment"]))  
white <- data.frame("white" = c("employed" = total\_data["white", "numerator"], "not\_employed" = total\_data["white", "not\_in\_employment"]))  
bw\_mat <- as.matrix(cbind(black, white))  
  
# run analysis  
xtab <- CrossTable(bw\_mat, fisher = T, chisq = T, expected = T)

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | Expected N |  
## | Chi-square contribution |  
## | N / Row Total |  
## | N / Col Total |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 491600   
##   
##   
## |   
## | black | white | Row Total |   
## -------------|-----------|-----------|-----------|  
## employed | 51200 | 293200 | 344400 |   
## | 63261.432 | 281138.568 | |   
## | 2299.634 | 517.461 | |   
## | 0.149 | 0.851 | 0.701 |   
## | 0.567 | 0.731 | |   
## | 0.104 | 0.596 | |   
## -------------|-----------|-----------|-----------|  
## not\_employed | 39100 | 108100 | 147200 |   
## | 27038.568 | 120161.432 | |   
## | 5380.394 | 1210.689 | |   
## | 0.266 | 0.734 | 0.299 |   
## | 0.433 | 0.269 | |   
## | 0.080 | 0.220 | |   
## -------------|-----------|-----------|-----------|  
## Column Total | 90300 | 401300 | 491600 |   
## | 0.184 | 0.816 | |   
## -------------|-----------|-----------|-----------|  
##   
##   
## Statistics for All Table Factors  
##   
##   
## Pearson's Chi-squared test   
## ------------------------------------------------------------  
## Chi^2 = 9408.178 d.f. = 1 p = 0   
##   
## Pearson's Chi-squared test with Yates' continuity correction   
## ------------------------------------------------------------  
## Chi^2 = 9407.398 d.f. = 1 p = 0   
##   
##   
## Fisher's Exact Test for Count Data  
## ------------------------------------------------------------  
## Sample estimate odds ratio: 0.4827818   
##   
## Alternative hypothesis: true odds ratio is not equal to 1  
## p = 0   
## 95% confidence interval: 0.4756259 0.4900722   
##   
## Alternative hypothesis: true odds ratio is less than 1  
## p = 0   
## 95% confidence interval: 0 0.488874   
##   
## Alternative hypothesis: true odds ratio is greater than 1  
## p = 1   
## 95% confidence interval: 0.4767621 Inf   
##   
##   
##

The results indicate that in the 12 months to March 2021, there was a significant difference between Black and White 16-64 year olds in the odds of being employed, (1, *N* = 491600) = 9408.18, *p* < .001. The odds of Black 16-64 year olds being employed were 0.48 times the odds of White 16-64 year olds being employed (95% CI = 0.48 - 0.49). The plot below shows that 56.7% of Black 16-64 year olds were employed in the 12 months to March 2021, compared to 73.06% of White 16-64 year olds.

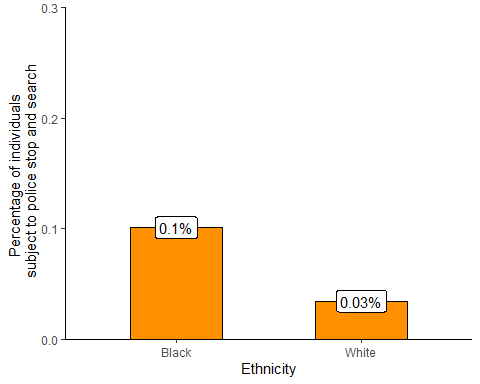


### Rate of Police Stop and Search

# stop and search  
data <- read.csv("../data/birmingham\_stop\_search\_2020\_09\_2021\_08.csv")  
  
# officer-defined ethnicity  
appearance\_freq\_data <- data %>%  
 group\_by(officer\_defined\_ethnicity) %>%  
 summarise(  
 n = n() # count frequencies  
 )   
  
appearance\_freq\_data$officer\_defined\_ethnicity[6] <- "Unknown"  
appearance\_freq\_data <- as.data.frame(appearance\_freq\_data)  
row.names(appearance\_freq\_data) <- appearance\_freq\_data$officer\_defined\_ethnicity  
  
# add population estimates  
# population estimates taken from APS apr2020-mar2021   
# Asian population = sum of Indian and Pakistani/Bangladeshi pops  
appearance\_freq\_data <- appearance\_freq\_data %>%  
 mutate(  
 pop = c(176300, 97100, 12400, 66800, 509700, NA),  
 percentage = 100 \* (n/pop),  
 not\_ss = pop - n  
 ) %>%  
 rename(  
 ethnicity = officer\_defined\_ethnicity  
 )  
  
# prepare xtabs  
black <- data.frame("Black" = c("ss" = appearance\_freq\_data["Black", "n"], "not\_ss" = appearance\_freq\_data["Black", "not\_ss"]))  
white <- data.frame("White" = c("ss" = appearance\_freq\_data["White", "n"], "not\_ss" = appearance\_freq\_data["White", "not\_ss"]))  
bw\_mat <- as.matrix(cbind(black, white))  
  
# run analysis  
xtab <- CrossTable(bw\_mat, fisher = T, chisq = T, expected = T)

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | Expected N |  
## | Chi-square contribution |  
## | N / Row Total |  
## | N / Col Total |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 606800   
##   
##   
## |   
## | Black | White | Row Total |   
## -------------|-----------|-----------|-----------|  
## ss | 98 | 176 | 274 |   
## | 43.845 | 230.155 | |   
## | 66.888 | 12.742 | |   
## | 0.358 | 0.642 | 0.000 |   
## | 0.001 | 0.000 | |   
## | 0.000 | 0.000 | |   
## -------------|-----------|-----------|-----------|  
## not\_ss | 97002 | 509524 | 606526 |   
## | 97056.155 | 509469.845 | |   
## | 0.030 | 0.006 | |   
## | 0.160 | 0.840 | 1.000 |   
## | 0.999 | 1.000 | |   
## | 0.160 | 0.840 | |   
## -------------|-----------|-----------|-----------|  
## Column Total | 97100 | 509700 | 606800 |   
## | 0.160 | 0.840 | |   
## -------------|-----------|-----------|-----------|  
##   
##   
## Statistics for All Table Factors  
##   
##   
## Pearson's Chi-squared test   
## ------------------------------------------------------------  
## Chi^2 = 79.66605 d.f. = 1 p = 0.0000000000000000004433542   
##   
## Pearson's Chi-squared test with Yates' continuity correction   
## ------------------------------------------------------------  
## Chi^2 = 78.20175 d.f. = 1 p = 0.0000000000000000009303591   
##   
##   
## Fisher's Exact Test for Count Data  
## ------------------------------------------------------------  
## Sample estimate odds ratio: 2.924804   
##   
## Alternative hypothesis: true odds ratio is not equal to 1  
## p = 0.00000000000000197176   
## 95% confidence interval: 2.260614 3.765891   
##   
## Alternative hypothesis: true odds ratio is less than 1  
## p = 1   
## 95% confidence interval: 0 3.622053   
##   
## Alternative hypothesis: true odds ratio is greater than 1  
## p = 0.000000000000001395464   
## 95% confidence interval: 2.354597 Inf   
##   
##   
##

The results indicate that in the 12 months to August 2021, there was a significant difference between Black and White individuals in the odds of being stopped and searched by Police, (1, *N* = 606800) = 79.67, *p* < .001. The odds of a Black individual being stopped and searched by police were 2.92 times the odds of a White individual being stopped and searched (95% CI = 2.26 - 3.77). The plot below shows that 0.1% of Black individuals in Birmingham were subject to Police Stop and Search, compared to 0.03% of White individuals.

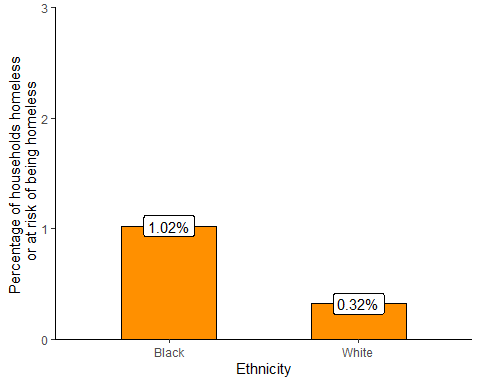


### Households that are statutorily homeless or at risk of becoming homeless

# homelessness  
data <- read.csv("../data/birmingham\_stat\_homeless\_2020\_2021\_v2.csv")  
  
total\_data <- data %>%  
 subset(ethnicity == "Asian" | ethnicity == "Black" | ethnicity == "Mixed/multiple" | ethnicity == "Other" | ethnicity == "White")  
  
total\_data <- total\_data[,-c(4)]  
total\_data <- total\_data[order(total\_data$ethnicity),]  
total\_data$ethnicity <- as.factor(total\_data$ethnicity)  
total\_data$number\_homeless <- as.numeric(total\_data$number\_homeless)  
row.names(total\_data) <- total\_data$ethnicity  
  
  
# in homelessness dataset Chinese is included in Asian  
# in population dataset it isn't. Need to account for this by taking chinese  
# from asian category and adding them to other category  
total\_data["Asian","number\_homeless"] <- total\_data["Asian","number\_homeless"] - 2  
total\_data["Other", "number\_homeless"] <- total\_data["Other", "number\_homeless"] + 2  
  
total\_data <- total\_data %>%  
 mutate(  
 percentage = 100\*(number\_homeless/population),  
 number\_not\_homeless = population - number\_homeless  
 )  
  
# prepare xtab  
black <- data.frame("black" = c("homeless" = total\_data["Black", "number\_homeless"], "not\_homeless" = total\_data["Black", "number\_not\_homeless"]))  
white <- data.frame("white" = c("homeless" = total\_data["White", "number\_homeless"], "not\_homeless" = total\_data["White", "number\_not\_homeless"]))  
bw\_mat <- as.matrix(cbind(black, white))  
  
# run analysis  
xtab <- CrossTable(bw\_mat, fisher = T, chisq = T, expected = T)

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | Expected N |  
## | Chi-square contribution |  
## | N / Row Total |  
## | N / Col Total |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 606800   
##   
##   
## |   
## | black | white | Row Total |   
## -------------|-----------|-----------|-----------|  
## homeless | 994 | 1633 | 2627 |   
## | 420.372 | 2206.628 | |   
## | 782.757 | 149.119 | |   
## | 0.378 | 0.622 | 0.004 |   
## | 0.010 | 0.003 | |   
## | 0.002 | 0.003 | |   
## -------------|-----------|-----------|-----------|  
## not\_homeless | 96106 | 508067 | 604173 |   
## | 96679.628 | 507493.372 | |   
## | 3.404 | 0.648 | |   
## | 0.159 | 0.841 | 0.996 |   
## | 0.990 | 0.997 | |   
## | 0.158 | 0.837 | |   
## -------------|-----------|-----------|-----------|  
## Column Total | 97100 | 509700 | 606800 |   
## | 0.160 | 0.840 | |   
## -------------|-----------|-----------|-----------|  
##   
##   
## Statistics for All Table Factors  
##   
##   
## Pearson's Chi-squared test   
## ------------------------------------------------------------  
## Chi^2 = 935.9275 d.f. = 1 p = 1.519756e-205   
##   
## Pearson's Chi-squared test with Yates' continuity correction   
## ------------------------------------------------------------  
## Chi^2 = 934.2967 d.f. = 1 p = 3.437901e-205   
##   
##   
## Fisher's Exact Test for Count Data  
## ------------------------------------------------------------  
## Sample estimate odds ratio: 3.217879   
##   
## Alternative hypothesis: true odds ratio is not equal to 1  
## p = 3.910051e-161   
## 95% confidence interval: 2.969872 3.485088   
##   
## Alternative hypothesis: true odds ratio is less than 1  
## p = 1   
## 95% confidence interval: 0 3.441255   
##   
## Alternative hypothesis: true odds ratio is greater than 1  
## p = 3.023572e-161   
## 95% confidence interval: 3.008277 Inf   
##   
##   
##

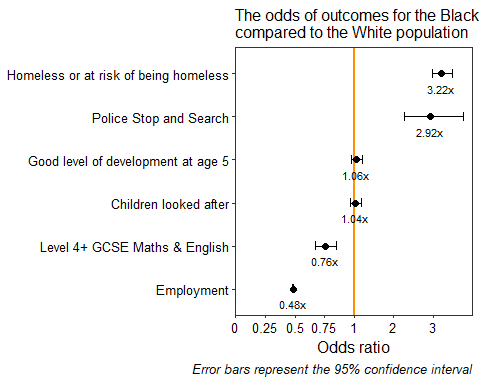
The results indicate that in the 2020-2021 financial year, there was a significant difference in the odds of being homeless or at risk of being homeless between Black and White households, (1, *N* = 606800) = 935.93, *p* < .001. The odds of a Black household being owed prevention or relief duty were 3.22 times the odds for a White household (95% CI = 2.97 - 3.49). The plot below shows that 1.02% of Black households in Birmingham were homeless or at risk of becoming homeless, compared to 0.32% of White households.



### Summary

The analyses presented highlight four areas where the Black population of Birmingham experience worse outcomes than the White population. The odds of achieving Level 4 in GCSE Maths and English and of being employed are lower for Black people compared to White people. The odds of a Black household being homeless or at risk of being homeless are three times greater than the odds for a White household, and the odds of a Black individual being subject to Police Stop and Search are nearly three times greater than the odds for a White individual.

In contrast, there appears to be no difference between Black and White children in the odds of being looked after by the local authority or of achieving a Good Level of Development at age 5.



Odds ratios and confidence intervals corresponding to the above plot

Indicator

Odds Ratio

Lower confidenceinterval (95%)

Upper confidenceinterval (95%)

Homeless or at risk of being homeless

3.22

2.97

3.49

Police Stop and Search

2.92

2.26

3.77

Good level of development at age 5

1.06

0.94

1.20

Children looked after

1.04

0.90

1.19

Level 4+ GCSE Maths & English

0.76

0.68

0.85

Employment

0.48

0.48

0.49

## Haringey

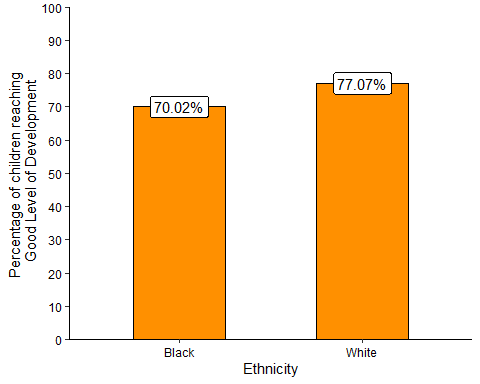
I next report the Haringey results. As above, for each indicator I provide the code for preparing and analysing the data, followed by a written account of the findings.

### Children achieving Good Level of Development

## good development  
data <- read.csv("../data/haringey\_good\_development\_2018\_2019\_v2.csv")  
  
total\_data <- data %>%  
 subset(gender == "Total") %>%  
 subset(characteristic\_type != "Total")  
  
row.names(total\_data) <- total\_data$characteristic\_type # set row names for indexing  
  
total\_data <- total\_data[,-c(1:3)] # get rid of unnecessary cols  
  
# prepare data frame for crosstabs  
  
black <- data.frame("black" = c("gld\_number" = total\_data["Black", "gld\_number"], "not\_gld\_number" = total\_data["Black", "number\_of\_pupils"] - total\_data["Black", "gld\_number"]))  
  
white <- data.frame("white" = c("gld\_number" = total\_data["White", "gld\_number"], "not\_gld\_number" = total\_data["White", "number\_of\_pupils"] - total\_data["White", "gld\_number"]))  
  
bw\_mat <- as.matrix(cbind(black, white)) # combine black and white into one matrix  
  
xtab <- CrossTable(bw\_mat, fisher = T, chisq = T, expected = T) # save xtab

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | Expected N |  
## | Chi-square contribution |  
## | N / Row Total |  
## | N / Col Total |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 2103   
##   
##   
## |   
## | black | white | Row Total |   
## ---------------|-----------|-----------|-----------|  
## gld\_number | 425 | 1153 | 1578 |   
## | 455.466 | 1122.534 | |   
## | 2.038 | 0.827 | |   
## | 0.269 | 0.731 | 0.750 |   
## | 0.700 | 0.771 | |   
## | 0.202 | 0.548 | |   
## ---------------|-----------|-----------|-----------|  
## not\_gld\_number | 182 | 343 | 525 |   
## | 151.534 | 373.466 | |   
## | 6.125 | 2.485 | |   
## | 0.347 | 0.653 | 0.250 |   
## | 0.300 | 0.229 | |   
## | 0.087 | 0.163 | |   
## ---------------|-----------|-----------|-----------|  
## Column Total | 607 | 1496 | 2103 |   
## | 0.289 | 0.711 | |   
## ---------------|-----------|-----------|-----------|  
##   
##   
## Statistics for All Table Factors  
##   
##   
## Pearson's Chi-squared test   
## ------------------------------------------------------------  
## Chi^2 = 11.47561 d.f. = 1 p = 0.0007051559   
##   
## Pearson's Chi-squared test with Yates' continuity correction   
## ------------------------------------------------------------  
## Chi^2 = 11.10203 d.f. = 1 p = 0.0008623304   
##   
##   
## Fisher's Exact Test for Count Data  
## ------------------------------------------------------------  
## Sample estimate odds ratio: 0.6947852   
##   
## Alternative hypothesis: true odds ratio is not equal to 1  
## p = 0.0008428681   
## 95% confidence interval: 0.5596396 0.8639332   
##   
## Alternative hypothesis: true odds ratio is less than 1  
## p = 0.0004877179   
## 95% confidence interval: 0 0.8347879   
##   
## Alternative hypothesis: true odds ratio is greater than 1  
## p = 0.9996708   
## 95% confidence interval: 0.5788231 Inf   
##   
##   
##

The results indicate that in Haringey there was a significant difference between Black children and White children in the odds of achieving a Good Level of Development at age 5, (1, *N* = 2103) = 11.48, *p* < .001. The odds of a Black child reaching a Good Level of Development were 0.69 times the odds for a White child (95% CI = 0.56 - 0.86).The plot below shows that 70.02% of Black children and 77.07% of White children reached a Good Level of Development in 2018-2019.

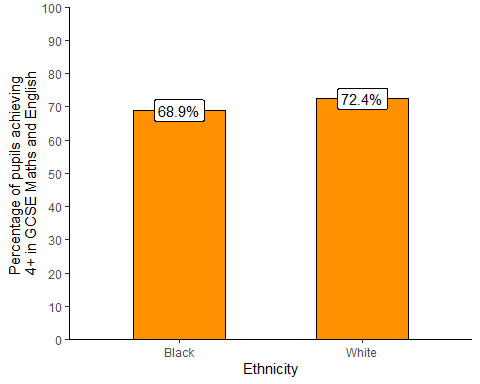


### Pupils achieving at least Level 4 in GCSE Maths and English

data <- read.csv("../data/haringey\_4\_plus\_gcse\_english\_maths\_2019\_2020\_v2.csv")  
  
# subset and rename  
total\_data <- data %>%  
 subset(gender == "Total") %>%  
 subset(ethnicity == "BLACK" | ethnicity == "WHITE") %>%  
 rename(total\_4\_plus = t\_l2basics\_94,  
 percent\_4\_plus = pt\_l2basics\_94)  
  
# rename ethnicity levels and row names  
total\_data$ethnicity <- as.factor(total\_data$ethnicity)  
levels(total\_data$ethnicity) <- c("Black","White")  
row.names(total\_data) <- total\_data$ethnicity  
  
total\_data <- total\_data[,-c(1)] # remove unnecessary column  
  
# calculate number who didn't achieve level 4 plus  
total\_data <- total\_data %>%  
 mutate(  
 not\_gcse = total\_pupils - total\_4\_plus  
 )  
  
# prepare data frame for xtab  
black <- data.frame("black" = c("gcse" = total\_data["Black", "total\_4\_plus"], "not\_gcse" = total\_data["Black", "not\_gcse"]))  
white <- data.frame("white" = c("gcse" = total\_data["White", "total\_4\_plus"], "not\_gcse" = total\_data["White", "not\_gcse"]))  
bw\_mat <- as.matrix(cbind(black, white))  
  
# run stats  
xtab <- CrossTable(bw\_mat, fisher = T, chisq = T, expected = T)

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | Expected N |  
## | Chi-square contribution |  
## | N / Row Total |  
## | N / Col Total |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 1753   
##   
##   
## |   
## | black | white | Row Total |   
## -------------|-----------|-----------|-----------|  
## gcse | 425 | 823 | 1248 |   
## | 439.256 | 808.744 | |   
## | 0.463 | 0.251 | |   
## | 0.341 | 0.659 | 0.712 |   
## | 0.689 | 0.724 | |   
## | 0.242 | 0.469 | |   
## -------------|-----------|-----------|-----------|  
## not\_gcse | 192 | 313 | 505 |   
## | 177.744 | 327.256 | |   
## | 1.143 | 0.621 | |   
## | 0.380 | 0.620 | 0.288 |   
## | 0.311 | 0.276 | |   
## | 0.110 | 0.179 | |   
## -------------|-----------|-----------|-----------|  
## Column Total | 617 | 1136 | 1753 |   
## | 0.352 | 0.648 | |   
## -------------|-----------|-----------|-----------|  
##   
##   
## Statistics for All Table Factors  
##   
##   
## Pearson's Chi-squared test   
## ------------------------------------------------------------  
## Chi^2 = 2.478448 d.f. = 1 p = 0.1154161   
##   
## Pearson's Chi-squared test with Yates' continuity correction   
## ------------------------------------------------------------  
## Chi^2 = 2.307645 d.f. = 1 p = 0.128739   
##   
##   
## Fisher's Exact Test for Count Data  
## ------------------------------------------------------------  
## Sample estimate odds ratio: 0.8419365   
##   
## Alternative hypothesis: true odds ratio is not equal to 1  
## p = 0.1221563   
## 95% confidence interval: 0.6758453 1.050048   
##   
## Alternative hypothesis: true odds ratio is less than 1  
## p = 0.06471937   
## 95% confidence interval: 0 1.014238   
##   
## Alternative hypothesis: true odds ratio is greater than 1  
## p = 0.9480283   
## 95% confidence interval: 0.6993654 Inf   
##   
##   
##

The results indicate that in the 2019-2020 academic year, there was no statistically significant difference between Black pupils and White pupils in the odds of achieving at least a Level 4 in GCSE Maths and English, (1, *N* = 1753) = 2.48, *p* = 0.115. The plot below shows that 68.9% of Black pupils achieved at least Level 4 in English and Maths, compared to 72.4% of White pupils.

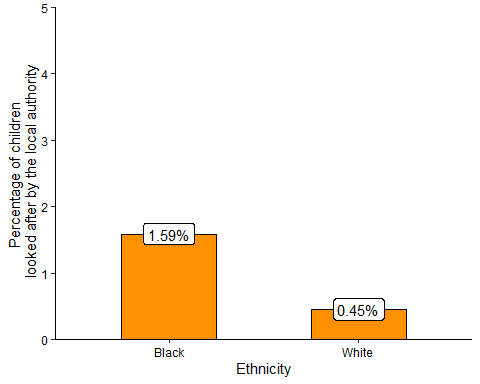


### Children looked after by the local authority

# children looked after  
data <- read.csv("../data/haringey\_children\_looked\_after\_2018\_2019\_v2.csv")  
  
# prepare data for analysis  
total\_data <- data %>%  
 subset(ethnicity == "Black" | ethnicity == "White") %>%  
 mutate(  
 not\_cla = pop - cla,  
 percentage = as.numeric(percentage)  
 )  
  
row.names(total\_data) <- total\_data$ethnicity  
  
# prepare xtab  
black <- data.frame("black" = c("cla" = total\_data["Black", "cla"], "not\_cla" = total\_data["Black", "not\_cla"]))  
white <- data.frame("white" = c("cla" = total\_data["White", "cla"], "not\_cla" = total\_data["White", "not\_cla"]))  
bw\_mat <- as.matrix(cbind(black, white))  
  
# run analysis  
xtab <- CrossTable(bw\_mat, fisher = T, chisq = T, expected = T)

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | Expected N |  
## | Chi-square contribution |  
## | N / Row Total |  
## | N / Col Total |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 42046   
##   
##   
## |   
## | black | white | Row Total |   
## -------------|-----------|-----------|-----------|  
## cla | 231 | 124 | 355 |   
## | 122.890 | 232.110 | |   
## | 95.108 | 50.355 | |   
## | 0.651 | 0.349 | 0.008 |   
## | 0.016 | 0.005 | |   
## | 0.005 | 0.003 | |   
## -------------|-----------|-----------|-----------|  
## not\_cla | 14324 | 27367 | 41691 |   
## | 14432.110 | 27258.890 | |   
## | 0.810 | 0.429 | |   
## | 0.344 | 0.656 | 0.992 |   
## | 0.984 | 0.995 | |   
## | 0.341 | 0.651 | |   
## -------------|-----------|-----------|-----------|  
## Column Total | 14555 | 27491 | 42046 |   
## | 0.346 | 0.654 | |   
## -------------|-----------|-----------|-----------|  
##   
##   
## Statistics for All Table Factors  
##   
##   
## Pearson's Chi-squared test   
## ------------------------------------------------------------  
## Chi^2 = 146.7013 d.f. = 1 p = 0.0000000000000000000000000000000009120818   
##   
## Pearson's Chi-squared test with Yates' continuity correction   
## ------------------------------------------------------------  
## Chi^2 = 145.3475 d.f. = 1 p = 0.000000000000000000000000000000001803015   
##   
##   
## Fisher's Exact Test for Count Data  
## ------------------------------------------------------------  
## Sample estimate odds ratio: 3.559091   
##   
## Alternative hypothesis: true odds ratio is not equal to 1  
## p = 0.0000000000000000000000000000001648642   
## 95% confidence interval: 2.846492 4.468195   
##   
## Alternative hypothesis: true odds ratio is less than 1  
## p = 1   
## 95% confidence interval: 0 4.309389   
##   
## Alternative hypothesis: true odds ratio is greater than 1  
## p = 0.00000000000000000000000000000009285345   
## 95% confidence interval: 2.946491 Inf   
##   
##   
##

The results indicate that, in the 2018-2019 financial year, there was a significant relationship between child ethnicity and the odds of being looked after the local authority, (1, *N* = 42046) = 146.7, *p* < .001. The odds of Black children being looked after by the local authority were 3.56 times the odds for White children (95% CI = 2.85 - 4.47).The plot below shows that 1.59% of Black children, compared to 0.45% of White children, were looked after by the local authority in the 2018-2019 financial year.

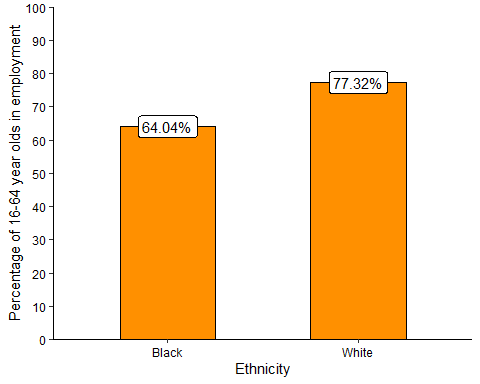


### Employment

# employment  
data <- read.csv("../data/haringey\_employment\_rate\_2020\_2021\_v3.csv")  
  
# subset and organise data  
total\_data <- data %>%  
 subset(gender == "all") %>%  
 subset(ethnicity == "black" | ethnicity == "white") %>%  
 mutate(  
 percentage = as.numeric(percentage) \* 100,  
 numerator = as.numeric(numerator),  
 denominator = as.numeric(denominator),  
 not\_in\_employment = denominator - numerator,  
 )  
   
row.names(total\_data) <- total\_data$ethnicity  
  
# prepare xtab  
black <- data.frame("black" = c("employed" = total\_data["black", "numerator"], "not\_employed" = total\_data["black", "not\_in\_employment"]))  
white <- data.frame("white" = c("employed" = total\_data["white", "numerator"], "not\_employed" = total\_data["white", "not\_in\_employment"]))  
bw\_mat <- as.matrix(cbind(black, white))  
  
# run analysis  
xtab <- CrossTable(bw\_mat, fisher = T, chisq = T, expected = T)

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | Expected N |  
## | Chi-square contribution |  
## | N / Row Total |  
## | N / Col Total |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 169200   
##   
##   
## |   
## | black | white | Row Total |   
## -------------|-----------|-----------|-----------|  
## employed | 14600 | 113200 | 127800 |   
## | 17221.277 | 110578.723 | |   
## | 398.988 | 62.138 | |   
## | 0.114 | 0.886 | 0.755 |   
## | 0.640 | 0.773 | |   
## | 0.086 | 0.669 | |   
## -------------|-----------|-----------|-----------|  
## not\_employed | 8200 | 33200 | 41400 |   
## | 5578.723 | 35821.277 | |   
## | 1231.660 | 191.816 | |   
## | 0.198 | 0.802 | 0.245 |   
## | 0.360 | 0.227 | |   
## | 0.048 | 0.196 | |   
## -------------|-----------|-----------|-----------|  
## Column Total | 22800 | 146400 | 169200 |   
## | 0.135 | 0.865 | |   
## -------------|-----------|-----------|-----------|  
##   
##   
## Statistics for All Table Factors  
##   
##   
## Pearson's Chi-squared test   
## ------------------------------------------------------------  
## Chi^2 = 1884.602 d.f. = 1 p = 0   
##   
## Pearson's Chi-squared test with Yates' continuity correction   
## ------------------------------------------------------------  
## Chi^2 = 1883.883 d.f. = 1 p = 0   
##   
##   
## Fisher's Exact Test for Count Data  
## ------------------------------------------------------------  
## Sample estimate odds ratio: 0.5221785   
##   
## Alternative hypothesis: true odds ratio is not equal to 1  
## p = 0   
## 95% confidence interval: 0.5068666 0.5379775   
##   
## Alternative hypothesis: true odds ratio is less than 1  
## p = 0   
## 95% confidence interval: 0 0.5354366   
##   
## Alternative hypothesis: true odds ratio is greater than 1  
## p = 1   
## 95% confidence interval: 0.5092835 Inf   
##   
##   
##

The results indicate that in the 12 months to March 2021, there was a significant difference between Black and White 16-64 year olds in the odds of being employed, (1, *N* = 169200) = 1884.6, *p* < .001. The odds of Black 16-64 year olds being employed were 0.52 times the odds of White 16-64 year olds being employed (95% CI = 0.51 - 0.54). The plot below shows that 64.04% of Black 16-64 year olds were employed in the 12 months to March 2021, compared to 77.32% of White 16-64 year olds.

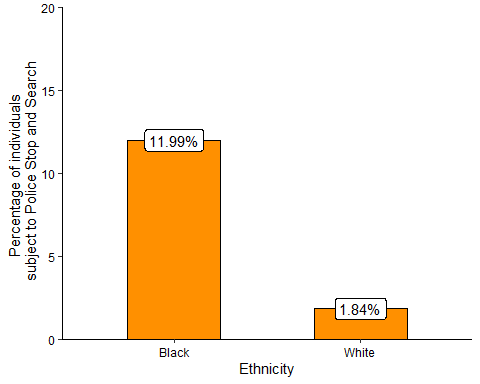


### Rate of Police Stop and Search

# stop and search  
data <- read.csv("../data/haringey\_stop\_and\_search\_2020\_2021\_v2.csv")  
  
# officer-defined ethnicity  
appearance\_freq\_data <- data %>%  
 group\_by(EA.Group) %>%  
 summarise(  
 n = n() # count frequencies  
 ) %>%  
 rename(  
 ethnicity = EA.Group  
 ) %>%  
 mutate(  
 # add population estimates  
 # population estimates taken from APS apr2020-mar2021   
 # Asian population = sum of Indian and Pakistani/Bangladeshi pops  
 pop = c(5800, 29200, 18400, 161400),  
 percentage = 100 \* (n/pop),  
 not\_ss = pop - n  
 )  
   
appearance\_freq\_data <- as.data.frame(appearance\_freq\_data)  
row.names(appearance\_freq\_data) <- appearance\_freq\_data$ethnicity  
  
# prepare xtabs  
black <- data.frame("Black" = c("ss" = appearance\_freq\_data["Black", "n"], "not\_ss" = appearance\_freq\_data["Black", "not\_ss"]))  
white <- data.frame("White" = c("ss" = appearance\_freq\_data["White", "n"], "not\_ss" = appearance\_freq\_data["White", "not\_ss"]))  
bw\_mat <- as.matrix(cbind(black, white))  
  
# run analysis  
xtab <- CrossTable(bw\_mat, fisher = T, chisq = T, expected = T)

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | Expected N |  
## | Chi-square contribution |  
## | N / Row Total |  
## | N / Col Total |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 190600   
##   
##   
## |   
## | Black | White | Row Total |   
## -------------|-----------|-----------|-----------|  
## ss | 3502 | 2971 | 6473 |   
## | 991.666 | 5481.334 | |   
## | 6354.734 | 1149.679 | |   
## | 0.541 | 0.459 | 0.034 |   
## | 0.120 | 0.018 | |   
## | 0.018 | 0.016 | |   
## -------------|-----------|-----------|-----------|  
## not\_ss | 25698 | 158429 | 184127 |   
## | 28208.334 | 155918.666 | |   
## | 223.401 | 40.417 | |   
## | 0.140 | 0.860 | 0.966 |   
## | 0.880 | 0.982 | |   
## | 0.135 | 0.831 | |   
## -------------|-----------|-----------|-----------|  
## Column Total | 29200 | 161400 | 190600 |   
## | 0.153 | 0.847 | |   
## -------------|-----------|-----------|-----------|  
##   
##   
## Statistics for All Table Factors  
##   
##   
## Pearson's Chi-squared test   
## ------------------------------------------------------------  
## Chi^2 = 7768.231 d.f. = 1 p = 0   
##   
## Pearson's Chi-squared test with Yates' continuity correction   
## ------------------------------------------------------------  
## Chi^2 = 7765.137 d.f. = 1 p = 0   
##   
##   
## Fisher's Exact Test for Count Data  
## ------------------------------------------------------------  
## Sample estimate odds ratio: 7.266425   
##   
## Alternative hypothesis: true odds ratio is not equal to 1  
## p = 0   
## 95% confidence interval: 6.906245 7.648413   
##   
## Alternative hypothesis: true odds ratio is less than 1  
## p = 1   
## 95% confidence interval: 0 7.585146   
##   
## Alternative hypothesis: true odds ratio is greater than 1  
## p = 0   
## 95% confidence interval: 6.961793 Inf   
##   
##   
##

The results indicate that in the 12 months to August 2021, there was a significant difference between Black and White individuals in the odds of being stopped and searched by Police, (1, *N* = 190600) = 7768.23, *p* < .001. The odds of a Black individual being stopped and searched by police were 7.27 times the odds of a White individual being stopped and searched (95% CI = 6.91 - 7.65). The plot below shows that 11.99% of Black individuals in Haringey were subject to Police Stop and Search, compared to 1.84% of White individuals.

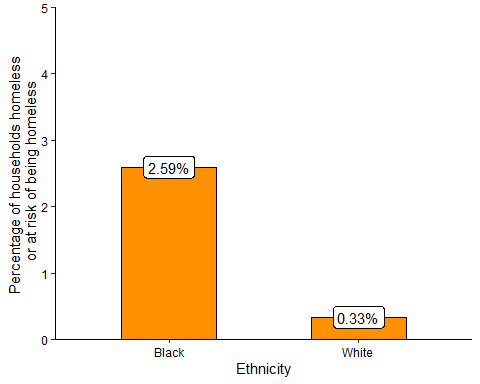


### Households that are statutorily homeless or at risk of becoming homeless

# homelessness  
data <- read.csv("../data/haringey\_stat\_homeless\_2020\_2021\_v2.csv")  
  
total\_data <- data %>%  
 subset(ethnicity == "Asian" | ethnicity == "Black" | ethnicity == "Mixed/multiple" | ethnicity == "Other" | ethnicity == "White")  
  
total\_data <- total\_data[order(total\_data$ethnicity),]  
total\_data$ethnicity <- as.factor(total\_data$ethnicity)  
total\_data$number\_homeless <- as.numeric(total\_data$number\_homeless)  
row.names(total\_data) <- total\_data$ethnicity  
  
# in homelessness dataset Chinese is included in Asian  
# in population dataset it isn't. Need to account for this by taking chinese  
# from asian category and adding them to other category  
total\_data["Asian","number\_homeless"] <- total\_data["Asian","number\_homeless"] - 7  
total\_data["Other", "number\_homeless"] <- total\_data["Other", "number\_homeless"] + 7  
  
  
total\_data <- total\_data %>%  
 mutate(  
 pop = c(5800, 29200, 4400, 18400, 161400), # haringey population estimates taken from APS apr2020-mar2021  
 percentage = 100 \* (number\_homeless/pop),  
 number\_not\_homeless = pop - number\_homeless  
 )  
# prepare xtab  
black <- data.frame("black" = c("homeless" = total\_data["Black", "number\_homeless"], "not\_homeless" = total\_data["Black", "number\_not\_homeless"]))  
white <- data.frame("white" = c("homeless" = total\_data["White", "number\_homeless"], "not\_homeless" = total\_data["White", "number\_not\_homeless"]))  
bw\_mat <- as.matrix(cbind(black, white))  
  
# run analysis  
xtab <- CrossTable(bw\_mat, fisher = T, chisq = T, expected = T)

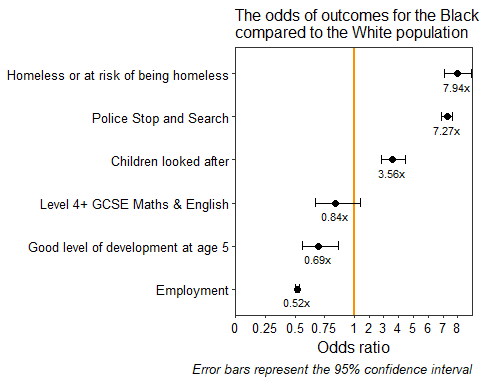
##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | Expected N |  
## | Chi-square contribution |  
## | N / Row Total |  
## | N / Col Total |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 190600   
##   
##   
## |   
## | black | white | Row Total |   
## -------------|-----------|-----------|-----------|  
## homeless | 757 | 539 | 1296 |   
## | 198.548 | 1097.452 | |   
## | 1570.750 | 284.175 | |   
## | 0.584 | 0.416 | 0.007 |   
## | 0.026 | 0.003 | |   
## | 0.004 | 0.003 | |   
## -------------|-----------|-----------|-----------|  
## not\_homeless | 28443 | 160861 | 189304 |   
## | 29001.452 | 160302.548 | |   
## | 10.754 | 1.946 | |   
## | 0.150 | 0.850 | 0.993 |   
## | 0.974 | 0.997 | |   
## | 0.149 | 0.844 | |   
## -------------|-----------|-----------|-----------|  
## Column Total | 29200 | 161400 | 190600 |   
## | 0.153 | 0.847 | |   
## -------------|-----------|-----------|-----------|  
##   
##   
## Statistics for All Table Factors  
##   
##   
## Pearson's Chi-squared test   
## ------------------------------------------------------------  
## Chi^2 = 1867.625 d.f. = 1 p = 0   
##   
## Pearson's Chi-squared test with Yates' continuity correction   
## ------------------------------------------------------------  
## Chi^2 = 1864.282 d.f. = 1 p = 0   
##   
##   
## Fisher's Exact Test for Count Data  
## ------------------------------------------------------------  
## Sample estimate odds ratio: 7.942744   
##   
## Alternative hypothesis: true odds ratio is not equal to 1  
## p = 1.237517e-278   
## 95% confidence interval: 7.098034 8.892578   
##   
## Alternative hypothesis: true odds ratio is less than 1  
## p = 1   
## 95% confidence interval: 0 8.735868   
##   
## Alternative hypothesis: true odds ratio is greater than 1  
## p = 1.237517e-278   
## 95% confidence interval: 7.223928 Inf   
##   
##   
##

The results indicate that in the 2020-2021 financial year, there was a significant difference in the odds of being homeless or at risk of being homeless between Black and White households, (1, *N* = 190600) = 1867.62, *p* < .001. The odds of a Black household being owed prevention or relief duty were 7.94 times the odds for a White household (95% CI = 7.1 - 8.89). The plot below shows that 2.59% of Black households in Birmingham were homeless or at risk of becoming homeless, compared to 0.33% of White households.



### Summary

The results of the analysis for Haringey indicate five areas where the Black population experience worse outcomes than the White population. Black individauls have half the odds of White individuals to be employed. Black children have lower odds of reaching a Good Level of Development at age 5 compared to White children, and the odds of Black children being looked after by the local authority are over three times greater than the odds for White children. The odds of a Black individual being stopped and searched by Police are over seven times greater than the odds for a White individual. Finally, Black households in Haringey have nearly eight times greater odds of being homeless or at risk of being homeless, compared to White households.



Odds ratios and confidence intervals corresponding to the above plot

Indicator

Odds Ratio

Lower confidenceinterval (95%)

Upper confidenceinterval (95%)

Homeless or at risk of being homeless

7.94

7.10

8.89

Police Stop and Search

7.27

6.91

7.65

Children looked after

3.56

2.85

4.47

Level 4+ GCSE Maths & English

0.84

0.68

1.05

Good level of development at age 5

0.69

0.56

0.86

Employment

0.52

0.51

0.54

# Limitations

A notable limitation of the present analyses is the coarseness of the ethnicity categories. Nearly all of the data considered have aggregated ethnicity in broad categories and it has not been possible to examine how outcomes vary between different Black populations, such as Black African and Black Caribbean populations. This is a limitation that needs to be addressed in future versions of the Shared Measurement System as a matter of urgency because aggregated data can only provide a vague and under-specified picture of the experiences of Black people in Britain.

All the analyses I have presented do not control for other factors that may be important for outcomes, such as age, sex, and income. Future versions of the Shared Measurement System should aim to take account of these potentially relevant factors in order to provide a more accurate picture of the inequality experienced by the Black population.